

EASYDL 文档



版权所有©百度在线网络技术(北京)有限公司、北京百度网讯科技有限公司。未经本公司书面许可,任何单位和个人不得 擅自摘抄、复制、传播本文档内容,否则本公司有权依法追究法律责任。

【商标声明】

🗘 百度智能云

和其他百度系商标,均为百度在线网络技术(北京)有限公司、北京百度网讯科技有限公司的商标。本文档涉及的第三方商标,依法由相关权利人所有。未经商标权利人书面许可,不得擅自对其商标进行使用、复制、修改、传播等行为。

【免责声明】

由于产品版本升级或其他原因,本文档内容会不定期进行更新。除非另有约定,本文档仅作为使用指导。如您购买本文档介 绍的产品、服务,您的权利与义务将依据百度智能云产品服务合同条款予以具体约定。本文档内容不作任何明示或暗示的保 证。

目录	2
平台简介	9
什么是EasyDL	
EasyDL产品体系	
EasyDL产品优势	
EasyDL常用概念	
AI开发基础知识	
文心大模型	18
新手指南	19
EasyDL图像-图像分类快速开始	19
EasyDL图像-物体检测快速开始	31
EasyDL文本-文本分类单标签快速开始	
EasyDL零售行业版快速开始	54
EasyDL图像SDK集成快速开始	61
通用设备端Android ARM	61
通用设备端Linux ARM	
通用设备端Windows x86加速版	83
服务器端Linux GPU 加速版	88
专项适配硬件EdgeBoard(FZ)	94
专项适配硬件Jetson	100
价权沿田	104
EasvDL图像价格说明	104
EasyDL图像本地服务器部署价格说明	104
EasyDL图像软硬一体方案价格说明	105
EasyDL图像通用小型设备部署价格说明	105
EasyDL图像价格常见问题	105
EasyDL图像公有云API价格说明	106
EasyDL图像价格整体说明	108
EasyDL图像算力资源价格说明	109
FasyDI 文本价格说明	110
文本私有服务器部署价格说明	110
EasyDL文本公有云API价格说明	110
EasyDL文本价格整体说明	112
EasyDL文本算力价格说明	113
	112
EasyDL结构化数据公有云API价格说明	113
表格预测算力资源价格说明	
EasyDL结构化数据价格整体说明	115
EasyDL优频价格记明 FasyDI 视频公有云API价格说明	116
ี่ และการราช การราช	TTO

EasyDL视频本地服务器部署价格说明	117
EasyDL视频设备端部署价格说明	118
EasyDL视频软硬一体方案价格说明	118
EasyDL视频价格整体说明	118
EasyDL视频算力资源价格说明	119
EasyDL语音价格说明	120
EasyDL语音公有云API价格说明	120
EasyDL语音本地服务器部署价格说明	121
EasyDL语音本地设备端部署价格说明	122
EasyDL语音价格整体说明	122
EasyDL语音算力资源价格说明	123
EasyDL零售行业版价格说明	123
价格整体说明	123
公有云API价格说明	124
本地部署价格说明	129
公有云API价格说明	129
价格整体说明	135
EasyDL跨模态价格说明	135
EasyDL跨模态公有云API价格说明	135
EasyDL跨模态算力资源价格说明	137
	100
EasyDL 图像介绍	138
图像分类	140
整体介绍	140
数据准备	141
模型训练	156
模型发布	166
端云协同服务	167
常见问题	346
<i>භ್/木</i> 本2回	3/18
整体介绍	348
数据准备	349
模型训练	364
模型发布	374
端云协同服务	374
常见问题	567
图像分割	570
철종가리 整体介绍	570
数据准备	570
模型训练	587
模型发布	594

端云协同服务	594
常见问题	711
EasyDL 文本使用说明	713
EasyDL文本介绍	713
文本分类-单标签	715
整体介绍	715
数据准备	715
模型训练	731
模型部署	735
常见问题	748
文本分举-多标签	749
整体介绍	749
数据准备	750
模型训练	761
模型部署	765
情感倾问分析	
亚[叶] 扣	776
x认加/止田	700
候坐训练	100
快 尘前者	/91
文本实体关系抽取	801
整体介绍	801
数据准备	801
模型训练	806
模型部署	808
文本实体抽取	818
整体介绍	818
数据准备	818
模型训练	829
模型部署	832
短文本相似度	838
整体介绍	838
数据准备	839
模型训练	844
模型部署	848
评论观点抽取	858
整体介绍	858
数据准备	859
ᇰᇦᆺᆂᇏ᠅ᅷᅕᄴᇃ	900
2. 可八个你/工义什双角 模型训练	860

Baidu 百度智能云文档	目录
模型发布	868
文本创作 (已下线)	880
文本创作介绍	880
数据准备	880
模型训练	884
模型发布	885
FasvDI 语音使用说明	891
EasyDL语音介绍	891
语音识别	892
语音识别介绍	892
创建模型	894
训练模型	896
上线模型	897
模型调用	897
声音分类	899
声音分类整体说明	899
数据准备	900
模型训练	909
模型发布	913
9.浏览器打开webui上传文件提示 500 internal server error	940
EasyDL 视频使用说明 FasyDI 视频介绍	941
263952 (从997) 31	042
视频分类介绍	942
创建模型	943
数据准备	944
	045
上16代观开关数据集 模型训练	945
模型发布	958
	963
日标跟踪	964
	904
입功廷1天空 	905
蚁/抗/庄奋	900
模型训练	9/1
	975
EasyDL 结构化数据使用说明	1012
EasyDL结构化数据介绍	1012
表格数据预测	1012
表格数据预测介绍	1012
简介	1012

数据准备 模型训练 模型发布 故障处理

时序预测

简介

时序预测介绍

数据准备 模型训练 模型发布 故障处理

EasyDL OCR使用说明 EasyDL OCR介绍

EasyDL 跨模态使用说明 EasyDL跨模态整体介绍

API文档 错误码

图文匹配 整体介绍 数据准备 模型训练

模型评估 模型部署

发布模型,生成在线API

EasyDL 零售行业版使用说明

零售版服务介绍

目录
 1013
 1014
 1020
 1038
1039
1039
1039
1040
1042
1047
1051
1052
 1052
1053
1054
1055
 1055
1056
1056
1057
1060
1062
1064
1064
1000
1069
1071
1078
1078
1079

购买指南	1071
场景范例	1078
拓展门店验证	1078
业务门店拜访	1079
店内陈列洞察	1081
异常拍照监测	1082
定制商品检测服务	1083
服务介绍	
购买指南	1084
快速训练一个模型	1087
模型创建	1095
数据准备	1097
数据标注	1123
模型训练	1140
柑 刑՝ 动里证: 仕	11/2

7

模型优化		1145
模型发布		1149
模型使用		1151
门店管理		1158
API文档		1165
翻拍识别服条		1176
服务介绍		1176
购买指南		1177
API文档		1178
购买指南		1181
门庄拜访らのи	(1183
服务介绍		1183
购买指南		1184
体验APP		1186
API文档		1187
SDK文档		1205
从标识别服务		1000
加金以加加分		1228
购买指南		1228
API文档		1230
<u>这书</u>		1000
串拍识 <u>加服</u> 务 服条介绍		1233
API文档		1234
雨生长少口	76	4044
苓售		1241
苓售 服 服 务 介 约	名 	1243
购头指南		1246
数据看板		1252
零售版常见问题	题	1252
EasyDL桌面版使	用说明(已下线)	1255
产品简介		1255
厂的介绍		1255
切形介绍		1255
系统文持 二共 上))) (1256
下致与激活		1257
AI开反基础为	비 <u>자</u>	1261
快速开始		1262
用零代码开发		1262
用零代码开发	友头现物体检测 	1270
用零代码开发		1278
用零代码开发	发买现语义分割	1285

Baidu 百度智能云文档	平台简介
数据管理	1290
数据导入	1290
查看数据集	1294
创建数据集	1296
数据标注	1297
开发与训练	1310
零代码开发	1310
Notebook开发	1324
模型中心	1334
模型列表	1334
模型部署	1335
服务列表	1335
智能边缘控制台	1335
离线SDK部署说明	1341
专项适配硬件离线部署	1341
服务器离线部署	1351
通用小型设备离线部署	1370
AutoDL模式算法适配硬件	1385
高级调参模式算法适配硬件	1386
常见问题	1387
分享我的模型	1388
分享EasyDL定制化模型	1388
分享EasyDL定制化API	1391
版本更新记录	1393
2021年01月	1395
常见问题	1397
常见问题	1397
智能边缘控制台	1400
智能边缘控制台-单节点版	1400
智能边缘控制台-多节点版	1411
EasyEdge 智能边缘控制台-单节点版 IEC API	1422
EasyEdge 智能边缘控制台-多节点版 IECC API	

平台简介

什么是EasyDL

目录

- 1. 产品介绍
- 2. 应用场景及案例

心 产品介绍

EasyDL从2017年11月中旬起,在国内率先推出针对AI零算法基础或者追求高效率开发的企业用户的零门槛AI开发平台,提供从数据采集、标注、清洗到模型训练、部署的一站式AI开发能力。对于各行各业有定制AI需求的企业用户来说,无论您是否具备AI基础,EasyDL设计简约,极易 理解,最快5分钟即可上手学会,15分钟完成模型训练。

您采集到的原始图片、文本、音频、视频、OCR、表格等数据,经过EasyDL加工、学习、部署后,可通过公有云API调用,或部署在本地服务 器、小型设备、软硬一体方案的专项适配硬件上,通过离线SDK或私有API进一步集成,流程如下:



根据企业用户的应用场景及深度学习的技术方向,EasyDL共推出6大通用产品及1个行业产品:

- EasyDL 图像:定制基于图像进行多样化分析的AI模型,实现图像内容理解分类、图中物体检测定位等,适用于图片内容检索、安防监控、工业质检等场景
- EasyDL 文本:定制基于文心大模型的语义理解AI模型,提供一整套文本定制与应用能力,适用于文本内容审核、文本自动生成、留言分类、 电商评价打分等场景
- EasyDL 语音:定制语音识别模型,精准识别业务专有名词,适用于数据采集录入、语音指令、呼叫中心等场景,以及定制声音分类模型,适 用于区分不同声音类别等场景
- EasyDL OCR:定制文字识别模型,结构化输出关键字段内容,满足个性化卡证票据识别需求,适用于证照电子化审批、财税报销电子化等场 景
- EasyDL 视频:定制基于视频片段内容进行分类的AI模型,适用于区分不同短视频类别等场景,以及定制目标追踪AI模型,实现跟踪视频中特定目标对象及轨迹,适用于视频内容审核、人流/车流统计、养殖场牲畜移动轨迹分析等场景
- EasyDL 结构化数据:挖掘数据中隐藏的模式,解决二分类、多分类、回归等问题,适用于客户流失预测、欺诈检测、价格预测等场景
- EasyDL 零售行业版:面向零售场景的ISV、零售行业服务商等企业用户,提供基于商品识别场景的AI服务解决方案,适用于货架巡检、自助结 算台、无人零售柜等场景

心 应用场景及案例

の 工业质检

- 产品组装合格性检查: 在流水线作业中针对组合型产品键盘可能存在的不合格情况进行图片收集,将图片完成标注后发起模型训练,从而训练出 自动判断合格或不合格的模型,辅助人工判断产品质量
- 具体案例:对键盘生产可能存在错装、漏装、合格等情况进行识别

任务类型:物体检测



商品瑕疵检测:针对商品微小瑕疵进行图片收集,并对原始图片或基于光学成像形成的图片进行瑕疵标注及训练,将模型集成在检测设备或流水 线中,辅助人工提升质检效率,降低人力成本

具体案例:针对地板质检的常见问题(例如,虫眼、毛面、棘爪等)进行检测

任务类型:物体检测



心 电商/网站图片分析

海量图片打分类标签: 电商将图片按定制标签标注并训练,构建对海量图片自动打标签模型,实现将图片面向不同C端用户的精准展示,以及基于分析用户所点击图片的内容从而进行相关图片推荐等功能

具体案例:家装网站将卧室、餐厅、厨房等图片进行内容分类

任务类型:图像分类



定制图像审核策略:根据需求制定图片审核标准,网站用EasyDL训练的模型判断UGC图片是否合规,适用于视频、新闻等内容平台定制内容审 核策略,过滤不良信息,或用于线上活动,判断C端用户提交图片的合规性。

具体案例:房产网站审核用户提交信息是房源图还是非房源图片

任务类型:图像分类



₯安防监控

生产环境安全监控:对生产环境现场进行安全性监控,对是否出现挖掘机等危险物品、工人是否佩戴安全帽/穿工作服、施工区是否有烟火等情况进行检查,辅助人工判断安全隐患并及时预警,保证生产环境安全运行

具体案例:对输电线路附近是否存在挖掘机、吊车等外部隐患物体进行检测

任务类型:物体检测



超市防损监控:在安全死角安装摄像头,将采集到图片进行标注及训练,实现实时检测图片中是否有未结算商品 具体案例:在超市结算台下层安装摄像头,将视频抽帧为图片后可判断图片中有未结算商品、无未结算商品

任务类型:物体检测



货物状态监控 根据实际业务场景在轮船内安装摄像头,采用定时抓拍或视频抽帧的方式,自动判断货物状态,提升业务运营、货品管理效率



具体案例:货船调运公司安装摄像头智能监控船只上货品状态为有货或无货

心 专业领域研究

医疗镜检识别:针对医疗检验场景中可能存在的正常或异常结果进行图片收集,并基于图片关键特征进行标注完成训练,协助医生高效完成结果 判断

具体案例:针对寄生虫卵镜检图片,判断虫卵类型从而对症下药

任务类型:物体检测



培训或科普: 将图片中专业部件进行标注,训练出专业零部件识别模型,拍照后识别出详细部件并配合提供详细的部件介绍信息,将该能力集成 在公司内部使用的培训app中,方便新人通过拍照识图快速上手业务

具体案例:汽车公司对内部人员提供车辆零部件拍图能力

任务类型:物体检测



₯零售商品识别

货架陈列合规性检查:将商品陈列图片进行采集、标注及训练,集成在app中供巡货员或者店员拍照上传自动识别,通过系统自动检查完成合规 性的准确检查

具体案例:零售快消公司在各商超的货架及货柜中拍照自动检测出第一排商品个数、位置完成自动巡店

任务类型:商品检测



电商平台商品管理:对电商商品图进行采集并标注明显特征,训练出商品识别模型,实现对商品图库的快速分类,减少人工入库的标注成本

具体案例:鞋类电商对鞋底花纹进行识别,自动判断鞋子品类

任务类型:物体检测



查看更多案例

EasyDL产品体系

目录

- 1. 通用产品
- 2. 行业产品

∞ 通用产品

根据企业用户的应用场景及深度学习的技术方向,EasyDL先后推出了以下6大通用产品,充分满足各行业的AI定制化需求

心 适合人群

对AI有需求的各行业企业客户或服务集成商

心 产品特性

- 零门槛: 无需具备算法基础,最快5分钟上手学会,15分钟完成模型训练,并将产出模型快速集成到业务产线
- 高精度: 基于百度文心大模型底座,内置百度超大规模预训练模型和自研AutoDL技术,只需少量数据就能训练出高精度模型

心 具体产品

- EasyDL 图像:定制基于图像进行多样化分析的AI模型,实现图像内容理解分类、图中物体检测定位等,适用于图片内容检索、安防监控、工业质检等场景 立即使用
- EasyDL 文本:定制基于文心大模型的语义理解AI模型,提供一整套文本定制与应用能力,适用于文本内容审核、文本自动生成、留言分类、 电商评价打分等场景 立即使用
- EasyDL 语音:定制语音识别模型,精准识别业务专有名词,适用于数据采集录入、语音指令、呼叫中心等场景,以及定制声音分类模型,适 用于区分不同声音类别等场景 立即使用
- EasyDL OCR:定制文字识别模型,结构化输出关键字段内容,满足个性化卡证票据识别需求,适用于证照电子化审批、财税报销电子化等场 景 立即使用
- EasyDL 视频:定制基于视频片段内容进行分类的AI模型,适用于区分不同短视频类别等场景,以及定制目标追踪AI模型,实现跟踪视频中特定 目标对象及轨迹,适用于视频内容审核、人流/车流统计、养殖场牲畜移动轨迹分析等场景 立即使用
- EasyDL 结构化数据:挖掘数据中隐藏的模式,解决二分类、多分类、回归等问题,适用于客户流失预测、欺诈检测、价格预测等场景 立即使用

© 行业产品

EasyDL零售行业版是专用于零售行业用户训练商品检测模型的模型训练平台,平台提供海量预置的商品图片,开放基于百度大规模零售数据的预 训练模型、及数据增强合成技术,实现低成本获得高精度商品检测AI模型服务

心 适合人群

有商品识别需求的零售行业企业或服务集成商,对货架巡检、无人货柜、无人结算台等零售场景有AI定制化需求

心 产品特性

- 高可用:针对零售场景专项算法调优,基于百度大规模零售数据的预训练模型,集成图像合成与增强技术提升模型泛化能力,模型准确率可达到97%,保证模型在生产环境中具有高可用性
- 全功能: 对货架巡检场景的业务场景提供了货架拼接SDK及API接口,功能强大,体验更优

の 具体产品

- 定制商品检测API:训练定制化商品检测模型,平台提供海量预置商品图片,开放基于百度大规模零售数据的预训练模型及数据增强合成技术,实现低成本获得高精度商品检测AI模型服务
- 标准商品检测API:无需训练即可直接使用的商品检测API,支持零售商超常见商品品类检测,针对货架合规性检查场景专项调优,适应大型商超、便利店、街边店等多种复杂货架场景接口返回商品名称、规格、品类及在图片中的位置
- 货架拼接SDK:货架拼接服务支持将多个货架局部图片或视频,组合为完整货架图片,同时支持输出在完整货架图中的商品检测结果,包含 SKU的名称、数量,适用于需要在长货架进行商品检测的业务场景

立即使用

EasyDL产品优势

目录

- 1. 零门槛
- 2. 高精度
- 3. 低成本
- 4. 广适配
- 5. 可交易

心零门槛

EasyDL提供围绕AI开发、部署的端到端一站式能力,包括数据采集、标注、清洗、模型训练、模型评估、模型部署。平台设计简约,极易理解, 最快5分钟即可上手,15分钟完成模型训练

	数据管理	模型构建		模型部署与应用				
卓越全	面的智能数据服务	零代码自动构建高精度模型		灵活选择端边云部署方案				
图片数据	<u></u> 数据采集	图像分类 物体检测		☑ 公有云部署 安	全生产			
文本数据	数据清洗	义平分突 诸百分突 视频分类 OCR		☆ 本地服务器部署	能制造 能硬件			
音频数据	🏷 数据扩充			亞 本地设备端部署	动营销			
结构化数据	○ 数据标注	预置百度超大规模预训练模型		2 软硬一体部署	消巡检			
		基于飞桨深度学习平台构建						

操作流程如下:

Step 1 创建模型 确定模型名称,可添加模型描述便于后续模型迭代管理

Step 2 上传并标注数据 上传数据后,根据不同模型类型的数据要求进行标注,如果有本地已标注的数据,也可以直接上传

Step 3 训练模型并校验效果选择算法类型、配置训练任务相关参数完成训练任务启动。模型训练完毕后支持可视化查看模型效果评估报告,也 支持通过**模型校验**功能在线上传实测数据验证模型效果

Step 4 部署模型 根据业务场景,支持将模型部署为公有云API实现在线调用,或部署在本地服务器/小型设备/软硬一体方案的专项适配硬件上,通过API/SDK集成离线应用

具体操作流程详见新手指南

心 高精度

EasyDL底层框架由百度自研的飞桨深度学习框架构建而成,内置基于百度文心大模型底座基础之上的成熟预训练模型,并结合百度自研的AutoDL 技术,助力用户基于少量数据就能获得具备出色效果与性能的模型

文心大模型 面向语言理解、语言生成等文本场景,具备超强语言理解能力以及对话生成、文学创作等能力。创新性地将大数据预训练与多源丰富 知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等方面的新知识,实现模型效果不断进化

AutoDL技术 面向视觉场景,百度研发的AutoDL Transfer模型结合了模型网络结构搜索、迁移学习技术,可针对用户数据进行算法自动优化。与 通用算法相比,更适用于细分类场景。例如,通用算法可用于区分猫与狗,但如果要区分不同品种的猫,则AutoDL效果会更好

心 低成本

数据对于模型效果至关重要,面向数据服务,EasyDL除提供基础的数据上传、存储、标注外,还提供业界领先的数据采集、智能标注、多人标 注、云服务数据回流等多种低成本数据管理服务,大幅降低开发者训练数据处理成本

端云协同数据采集 EasyDL提供便捷的本地数据采集软件,支持定时拍照、视频抽帧(支持自定义抽帧规则)多种采集方式,并将图片即时同步 到平台管理,无需摄像头数据反复下载与重新导入了解详情 **智能标注** 智能标注为一套人机交互的协作标注方式,在手工标注少量数据后,系统会从数据集所有样本中筛选出最关键的难例并提示需优先标 注。通常情况下,只需标注数据集30%左右的数据即可训练模型,与标注所有数据后训练相比,模型效果几乎等同了解详情

多人标注 通过将数据集在线共享给团队成员,实现多人分工标注、审核数据,有效降低标注成本,通过多人协作提高标注效率 了解详情

云服务数据回流 当将EasyDL训练的模型以公有云方式部署在业务场景中时,通过开通*云服务数据回流功能,可将实际调用后的业务场景真实数据及识别结果在平台中查看和管理,将识别错误的图片结果人工筛选后保存至数据集持续训练,可有效长期提升模型效果了解详情

心广适配

EasyDL模型训练阶段需要在线训练,训练完成后,可将模型部署在公有云服务器、本地服务器、小型设备、软硬一体方案专项适配硬件上,通过 API或SDK进行集成,充分满足各种业务场景对模型部署的要求

公有云API

1. 训练完成的模型存储在云端,可通过独立Rest API调用模型,实现AI能力与业务系统或硬件设备整快速集成

2. 支持在线调用的高并发请求及灵活的扩缩容策略,提高资源利用率

3. 支持查找云端模型识别错误的数据,纠正结果并将其加入模型迭代的训练集,不断优化模型效果

本地服务器部署

1. 可将训练完成的模型部署在私有CPU/GPU服务器上,支持API和SDK两种集成方式

2. 可在内网/无网环境下使用模型,确保数据隐私

本地设备端部署

1. 训练完成的模型被打包成适配智能硬件的SDK,可进行设备端离线计算,有效满足业务场景中无法联网、对数据保密性要求较高、响应时延要 求更快的需求

2. 支持iOS、Android、Linux、Windows四种操作系统,基础接口封装完善,满足灵活的应用侧二次开发

软硬一体方案

为进一步提升前端智能计算的用户体验,EasyDL推出<mark>前端智能计算-软硬一体方案</mark>,将百度推出的高性能硬件与EasyDL图像分类/物体检测模型深度适配,可应用于工业分拣、视频监控等多种设备端离线计算场景,让离线AI落地更轻松

具体部署流程详见图像分类、物体检测

心 可交易

EasyDL致力于打造全面开放的AI开发生态,与百度AI市场无缝对接,支持模型、AI服务、智能硬件等自由交易

AI模型交易售卖者层面,用户将EasyDL训练完成的模型在AI市场开放售卖,可获取现金收益或平台积分;购买者层面,用户在AI市场购买业务场 景相似的EasyDL模型,并基于已购模型再训练,仅需添加少量数据,即可快速获得高精度AI模型

AI服务交易 用户将成功发布的公有云API在AI市场开放售卖,可获取现金收益或平台积分

智能硬件交易用户可在AI市场购买EasyDL软硬一体方案,同时获得EasyDL专项适配硬件和EasyDL软件的使用授权,实现EasyDL产出的模型无缝 集成进专项适配硬件

具体交易详情详见AI市场

EasyDL常用概念

目录

- 1. 模型与模型类型
- 2. 模型训练相关
- 3. 模型效果相关
- 4. 模型部署相关
- 5. 数据相关

EasyDL支持6大技术方向,每个方向包括不同的模型类型:

- EasyDL 图像:图像分类、物体检测、图像分割
- EasyDL 文本: 文本分类-单标签、文本分类-多标签、文本实体抽取、情感倾向分析、短文本相似度
- EasyDL 语音:语音识别、声音分类
- EasyDL OCR:文字识别
- EasyDL 视频:视频分类、目标跟踪
- EasyDL 结构化数据:表格预测

心 模型训练相关

രം AutoDL Transfer

AutoDL Transfer模型是百度研发的AutoDL技术之一,结合了模型网络结构搜索、迁移学习技术、并针对用户数据进行自动优化。与通用算法相比,训练时间较长,但更适用于细分类场景。例如,通用算法可用于区分猫和狗,但如果要区分不同品种的猫,则AutoDL效果会更好

രം ERNIE

领先的语义理解技术与平台文心(ERNIE),依托飞桨打造,集先进的预训练模型、全面的NLP算法集、端到端开发套件和平台化服务于一体,提供一站式NLP开发与服务,让您更简单、高效地定制企业级文本模型。文心提供的ERNIE预训练模型,已累计学习10亿多知识,能够助力各NLP任务快速提升效果。平台内置了最新的ERNIE2.0,并提供了ERNIE2.0-Base、ERNIE2.0-Large两个版本供用户选择。

心 模型效果相关

心准确率

图像分类/文本分类/声音分类等分类模型的衡量指标,正确分类的样本数与总样本数之比,越接近1模型效果越好

രം F1-score

对某类别而言为精确率和召回率的调和平均数,对图像分类/文本分类/声音分类等分类模型来说,该指标越高效果越好

の 精确率(Precision)

对某类别而言为正确预测为该类别的样本数与预测为该类别的总样本数之比

心 石回率(Recall)

对某类别而言为正确预测为该类别的样本数与该类别的总样本数之比

രം top1、top2…top5

在查看图像分类/文本分类/声音分类/视频分类模型评估报告中,top1-top5指的是针对一个数据进行识别时,模型会给出多个结果,top1为置信度最高的结果、top2次之…正常业务场景中,我们通常会采信置信度最高的识别结果,重点关注top1的结果即可。

രം mAP

mAP(mean average precision)是物体检测(Object Detection)算法中衡量算法效果的指标。对于物体检测任务,每一类object都可以计算出其精确 率(Precision)和召回率(Recall),在不同阈值下多次计算/试验,每个类都可以得到一条P-R曲线,曲线下的面积就是average

の阈值

物体检测模型会存在一个可调节的阈值(threshold),是正确结果的判定标准,例如阈值是0.6,置信度大于0.6的识别结果会被当作正确结果返回。每个物体检测模型训练完毕后,可以在模型评估报告中查看推荐阈值,在推荐阈值下F1-score的值最高。

心 模型部署相关

心 公有云API

模型部署为Restful API,可以通过HTTP请求的方式进行调用。

心 设备端SDK

模型部署为设备端SDK,可集成在前端智能计算硬件设备中,可完全在无网环境下工作,所有数据皆在设备本地运行处理。目前支持IOS、

ANDROID、WINDOWS、LINUX四种操作系统及多款主流智能计算硬件。

心本地服务器部署

模型部署为本地服务器部署,可获得基于定制EasyDL模型封装而成的本地化部署的方案,此软件包部署包开发者本地的服务器上运行能够得到与 在线API功能完全相同的接口。

```
心 软硬一体方案
```

目前EasyDL支持两款软硬一体硬件,包括EasyDL-EdgeBoard软硬一体方案及EasyDL-十目计算卡。通过在AI市场购买,可获得硬件+专项适配硬件的设备端SDK,支持在硬件中离线计算。

心 数据相关

心 智能标注

智能标注为一套人机交互的协作标注方式,目前EasyDL物体检测训练任务支持智能标注,在手工标注少量数据后,系统会从数据集所有图片中筛 选出最关键的图片并提示需要优先标注。通常情况下,只需标注数据集30%左右的数据即可训练模型。与标注所有数据后训练相比,模型效果几 乎等同。

AI开发基础知识

目录

- 1. AI概念及基本原理
- 2. AI模型训练的基本流程介绍

^の AI概念及基本原理

人工智能(Artificial Intelligence,英文缩写为AI。它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。人工智能企图生产出一种新的能以人类智能相似的方式做出反应的智能机器,该领域的研究包括机器人、语言识别、图像识别、自然语言处理等。

在EasyDL平台背后主要使用了深度学习的技术,深度学习是机器学习(ML, Machine Learning)领域中一个新的研究方向。通过学习样本数据的内在规律和表示层次,最终目标是让机器能够像人一样具有分析学习能力,能够识别文字、图像和声音等数据。

心 AI模型训练的基本流程介绍



1. 分析业务需求

在正式启动训练模型之前,需要有效分析和拆解业务需求,明确模型类型如何选择。这里我们可以举一些实际业务场景进行分析。

举例:原始业务需求—某企业希望为某个高端小区物业做一套智能监控系统,希望对多种现象智能监控并及时预警,包括保安是否在岗、小区是否有有异常噪音、小区内各个区域的垃圾桶是否已满等多个业务功能。 针对这个原始业务需求,我们可以分析出不同的监控对象所在的位置不同、监控的数据类型不同(有的针对图片进行识别、有的针对声音进行判断),需要多个模型综合应用。 监控保安是否在岗——通过图像分类模型进行判断 监控小区是否有异常噪音——定时收集声音片段通过声音分类模型进行判断 监控小区内各个区域垃圾桶是否已满——由于监控区域采集的画面可能会存在多个垃圾桶,此处需要通过物体检测模型进行判断。

2. 采集/收集数据

在通过上述第一步分析出基本的模型类型,需要进行相应的数据收集工作。数据的主要原则为**尽可能采集真实业务场景一致的数据,并覆盖可能 有的各种情况**

3. 标注数据

采集数据后,可以通过EasyDL在线标注工具或线下其他标注工具对已有的数据进行标注。如上述保安是否在岗的图像分类模型,需要将监控视频 抽帧后的图片按照【在岗】及【未在岗】两类进行整理;小区内各个区域垃圾桶是否已满,需要将监控视频抽帧后的图片标注其中每个垃圾桶的 【空】【满】两种状态进行标注。

4. 训练模型

训练模型阶段可以将已有标注好的数据基于已经确定的初步模型类型,选择算法进行训练。通过使用EasyDL平台,可以可视化在线操作训练 任务的启停、训练任务的配置。可以大幅减少线下搭建训练环境、自主编写算法代码的相关成本。

5. 评估模型效果

训练后的模型在正式集成之前,需要评估模型效果是否可用。在这个环节上EasyDL提供了详细的模型评估报告,以及在线可视化上传数据测试模型效果的功能。

6. 部署模型

当确认模型效果可用后,可以将模型部署至生产环境中。传统的方式需要将训练出的模型文件加入工程化相关处理,通过使用EasyDL,可以便捷 地将模型部署在公有云服务器或本地设备上,通过API或SDK集成应用,或直接购买软硬一体产品,有效应对各种业务场景所需,提供效果与性能 兼具的服务。

文心大模型

∞ 文心大模型是什么

文心大模型是百度发布的产业级知识增强大模型,是千行百业AI开发的首选基座大模型。文心大模型既包含基础通用的大模型,也包含了面向重 点领域和重点任务的大模型,还提供丰富的工具与平台,支撑企业与开发者进行高效便捷的应用开发。"知识增强"是文心的核心特色,文心能够 同时从大规模知识和海量多元数据中持续学习,如同站在巨人的肩膀上,训练效率和理解准确率都得到大幅提升,并具备了更好的可解释性。目 前文心大模型包括:

^の文心·NLP大模型

面向语言理解、语言生成等NLP场景,具备超强语言理解能力以及对话生成、文学创作等能力。创新性地将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等方面的新知识,实现模型效果不断进化。

⊙文心·CV大模型

基于领先的视觉技术,利用海量的图像/视频等数据,为企业/开发者提供强大的视觉基础模型,以及一整套视觉任务定制与应用能力。

⊙ 文心·跨模态大模型

基于知识增强的跨模态语义理解关键技术,可实现跨模态检索、图文生成、图片文档的信息抽取等应用的快速搭建,落实产业智能化转型的AI助力

₯ 如何在EasyDL上使用文心大模型

の 文心·NLP大模型

文心·NLP大模型已接入EasyDL文本技术方向。通过下方简单几步即可上手使用。

①在EasyDL官网选择一个文本方向的模型类型,以文本分类-单标签为例

イジュ ご楽 EasyDL 产品介紙	3 操作平台	应用案例 使用	1文档				社区	飞桨官网	AI主站	立即使用
E	asyĽ	· 금등 산그 부분	<u>ант 42 т</u> .			1 00 44 00	使.	用文档		
				选择模型类型			×			
Easy	/DL入遗Forrest 零算法者	在线使用	- Yak			本地使用				
	高精度A	● 四城	 文本 文本分类-单标签 	🕝 ERNIE大模型创作	· 语音识别	• E218				
	【使用】	103 物体检测	文本分类-多标签	文本实体抽取	1 声音分类		386 			
and the second se	【参考】」	四條分割	短文本相似度	文本实体关系抽取						/
	。【购买】		👛 情感倾向分析				日台	前磅上线		
		 視頻 	• 结构化数据	 行业版 	• OCR		1			
		1 税期分类	表格数据预测	🗧 零售行业版	CR OCR					99
			11 时序预测							
	٢		Ş				٠			
		>	EasyDL	.文本 >		EasyDL语音 >				

②创建数据集-导入文本数据,完成数据标注

/^j/^j ど楽EasyDL	产品介绍	名 操作平台	应用案例	使用文档				社区	飞桨官网 AI主站	() 15	8K1G9x5hG
文本分类模型	=	我的数据总览									提交工的
图 总览											
① 模型中心		EasyData智	能数据服务平台已上	线,使用EasyData可	享受包括多人标注、数据清淡	先、数据采集等完整数据服务	立即前往				
我的模型		创建数据集								输入数据集制	S称或ID Q
创建模型											
训练模型		yyj_多语种单	◎标签_001 区 数据	諸集组ID: 254081					□ 新增版本	HH 全部版本	自删除
校验模型		版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作		
发布模型			070010	0	- 794	**//#	011 (0.10)		# (+=>+ ==)	10100	
☑ EasyData数据服务		VI 😊	272319	0	• Bra	又中万英	0% (0/0)	-	2×10/3 4×	ADS INS.	
数据总览		sean-test-v	2 F/ 数据集组D	254062					口新規版本	昭全部版本	白刪除
公开数据集											
标签组管理		版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作		
在线标注		V1 💬	272297	0	• 已完成	文本分类	0% (0/0)	-	多人标注 导入	删除	
智能标注											
22 模型部署		tsn_文本-多	语种文本单标签 🖸	数据集组ID: 250	023				□新増版本	88 全部版本	<u> 前間</u> 除
公有云服务		版本	秋田市の	教授品	县近号) 状态	标注类型	标注说本	诸法行太	58.45		
		788,191	MARDONUU	NA 163 MIL	THE REAL AND AND A	10/72.540382	30/T-2/072	1991/6/5/025	2275.1 [*		

③创建模型-完成训练配置-开始训练,此时将会**以文心大模型为基座开始训练模型**

アア 佐葉 EasyDL 产品	个绍 操作平台	应用案例 使用文档			社区 飞桨官网	AI主站 🕕	f58K1G9x5hG
文本分类模型 🛛 🖃	训练模型						提交工单
 ※ ※	选择模型 训练配置 部署方式 选择算法	zxy-test-single 公和云部署 EasyEdge本地部署 중 高相度 ⑦ 高性能 ⑦	如何选择邮署方式?				
发布模型 EasyData数据服务 数据总览 公开数据集	模型筛选指标 ⑦ 高加数据 添加训练数据) 模型兼题Precision和A ∨ ⑦ + 请这择	-				
标签组管理 在线标注 智能标注	由中以巡过集 @	数据集 binbin_test_0629	版本 V1	分夾款量 2	糖作		
23 模型部署 公有云服务 EasyEdge 纯离线服务	训练环境	名称 规格 O GPU P40 Teslad	GPU_P40_24G显存单卡_12核CPU_40G)	的择	算力 12 TeraFLOPS	提升速度 1	价格 免费
	开始训练	GPU V100 Testad	3PU_V100_16G显存单卡_12核CPU_56G	內存	14 TeraFLOPS	1.5	免费

④完成训练后,将模型发布为公有云服务接口,即可参考API文档调用服务

PC楽EasyDL	产品分	1名 操作平台	应用案例	使用文档		社区	飞桨官网	AI主站	158K1G9x5h
本分类模型	=	发布模型							提多
总览									
模型中心		选择模型	zxy-test-single		\sim	标准接口	N范参考	下载 API 文档	
我的模型		部署方式	公有云部署		~	标准接口调	求参考说明	1:	
J建模型						字段名称	必须	英型	说明
练模型		选择版本	V5		~	text	是	string	文本数据,UTF-8編 最大长度4096汉字
验模型	_	服务名称 *				top_num	10	number	返回分类数量, 默认 个
模型		接口地址 *	https://aip.baidub	ice.com/rpc/2.0/ai_cu	isto	4-144 AV 414 AV			
syData数据服务			m/v1/text_cls/			你准接口吗	加子校说明	3:	
暖总览						字段名称	必须	供型	说明
F数据集		其他要求	若接口无法满足加 的问题,500汉9	忽的需求,请描述希望≜ ₽以内	彩决	log_id	是	number	唯一的log id,用于 定位
资组管理						results	æ	array(object)	分类结果数组
動标注				0	/500	+ name	8	string	分类名称
能标注			提交申请			+ score	香	number	置信度
型部署									
有云服务									
EasyEdge 纯离线服务									

⁰文心·CV大模型

文心·CV大模型即将接入,敬请期待!

の文心・跨模态大模型

文心·跨模态大模型即将接入,敬请期待!

新手指南

EasyDL图像-图像分类快速开始

目录

- 1. 模型示例说明
- 2. 实现步骤
- 3. 产品特色
- 4. 更多参考

∞ 模型示例说明

图像分类模型主要用于识别一张图中是否是某类物体/状态/场景,是和图片中主体或状态单一的场景。本文以猫狗识别模型为示例演示图像分类 模型训练全过程。

心 实现步骤

只需四步即可完成自定义AI模型的训练及发布的全过程。

₯ Step1:成为百度AI开放平台的开发者

要使用百度EasyDL的模型训练能力首先需要注册成为百度AI开放平台的开发者,首先让我们用5分钟来注册百度AI开放平台的开发者(如您已经是 开发者,可直接登录使用)

先点击此处注册百度账号进入,如下图的页面快速的注册一个百度账号吧。

用户名	请设置用户名	
手机号	可用于登录和找回密码	
密 码	请设置登录密码	0
俭证码	请输入验证码	获取验证码

图像分类需要提供包含不同类别的图片并标注图片即可训练图像分类模型,自动识别图中是否包含某类物体/状态/场景,下面我们来看看这次训 练所需的猫狗图片示例:

の Step2:提前准备训练数据



图片数量越多理论上训练效果越好,图像分类的图片数量建议每个类别不低于20张图片。

₯ Step3:使用EasyDL训练图像分类模型

创建模型

进入EasyDL官方平台点击【立即使用】

L入选Forrester领导者象限	
零算法基础定制	
高精度AI模型	FasyDI 重磅上线目标跟踪模型
【了解】EasyDL简介	检测并识别视频中的运动对象 实现目标跟踪与计数
【了解】EasyDL简介 【参考】应用案例 >	检测并识别视频中的运动对象 实现目标跟踪与计数
■【了解】EasyDL简介 ■【参考】应用案例 >	

点击【**图像分类】**进入操作台

		选择	模型类型	
🔁 图像分类	<u>@</u>	物体检测	▶ 图像分割	文本分类-单标签
文本分类-多	标签	短文本相似度	👛 情感倾向分析	文本实体抽取
文本实体关系	も おおおおおおおおおおおおおおおおおおおおおおおおおおおおおおおおおおおお	语音识别	● 声音分类	1 视频分类
◎ 目标跟踪	iii	表格数据预测	1 时序预测	膏 零售行业版
CCR				

在模型中心下点击【创建模型】

Baidu 百度智能云文档

图像分类模型	*=	模型列表						操作文档 载音	产税粮 常见问】	0 <u>2</u> 2714
凸 模型中心										
我的模型		创建模型								
创建模型 训练模型		【图像分类】hyb_不平衡测证	(2 区 模型ID: 120	0101				吕 训练	④ 历史版本	回 删除
校验模型		部署方式	版本	训练状态	服务状态	模型效果	操作			
发布模型 EasyData数据服务		公有云API	٧١	训练完成	未发布	top1准确率: 95.80% top5准确率: 100.00% 完整评估结果	宣看拔本配置 申请发布 校验			
数据总览										

模型信息填写完成后点击【下一步】

模型列表 > 创建模型

模型类别	图像分类	
模型名称 *	猫狗识别test	
模型归属	公司 个人	
邮箱地址 *	*********@163.com	
联系方式 *	130*****691	0
功能描述 *	识别图片中的动物是猫还是狗	
	13/500	\$
	下一步	

模型创建完成后可在【我的模型】栏查看已创建的模型信息

模型列表	操作文档	教学视频	常见问题	提交工事
创建得加				
【图像分类】 描为识别test ☑ 模型ID: 120972		Ε	目 训练	前 删除
模型创建成功,若无数据集调先在"数据中心"创建。上传训练数据训练模型后,可以在此处重要模型的最新版本				

创建数据集

在数据总览界面点击【**创建数据集】**

图像分类模型 🔹	=	我的数据总览							提交工单
值 模型中心		Face Onto 1946			42.¢ (42)+ an470:an1+ an−				
我的模型		CasyDatatistic	KARAGHTACIA,	使用EasyData可学变色	始ラハ物は、鉄路用の、鉄	西本来等元重以后依另 五印制江			
创建模型		创建数据集							我的数据集 ~ 输入数据集名称或Ⅳ Q
训练模型			•						
校验模型		dog or cat ⊵	数据集组ID: 1836	17					[7]新编版本 281全部版本 由 删除
发布模型		版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作
回 EasyData数据服务							Decision and the		
数据总览		VI 😔	192301	101	 已完成 	图像分类	0% (0/101)	-	重看与标注 多人标注 导入 清洗 …
标签组管理		meaning too		80174849-D-1003					
在线标注		和护技术 【775		3 8X85903EID- 1031	*0				Ctaugewee DD TRANGE ED BOR
云服务数据回流		版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作
摄像头数据采集		V1 🕀	191781	5794	• 已完成	图像分类	100% (5794/5794)	-	宣看与标注 多人标注 导入 清洗 …
四 模型部署									

在数据集创建界面输入数据集名称后点击【完成】

我的数据总览 > 创建数据集

数据集名称	dog or cat	
教促米刑		
<u> </u>	Sh	
数据集版本	V1	
标注类型	图像分类	
标注模板	单图单标签	~
	完成	

数据集创建完成后可在【数据总览】查看已创建完成的数据集,点击【导入】跳转至数据导入界面

我的数据总览							提交工单
EasyData智能数	据服务平台已上线,使用Easy	/Data可享受包括多人标	注、数据清洗、数据采集等完整数据服务:	立即前往			
创建数据集							我的数据集 > 输入数据集名称或ID Q
dog or cat 🖂	数据集组ID: 183617						□新编版本 铝全部版本 由删除
版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作
ν1 Θ	192301	0	• 已完成	图像分类	0% (0/0)	-	多人标注 导入 删除 共享

数据导入支持无标注信息、有标注信息两种数据标注状态的数据以及多种导入方式,以下为无标注信息·本地导入-上传图片的导入方式示例,其 余各类型导入方式可参考图像分类页面上传数据集并在线标注

	导入数据			
	数据标注状态	● 无标注信息	○ 有标注信息	
选择数据标注状态及导入方式后点击【上 传图片】	导入方式	本地导入	~	上传图片 ~
	上传图片	<u>土</u> 上传图片		
		确认并返回		

注意:上传图片时,一定注意格式要求!





选择好图片后,点击【开始上传】



上传完成后,点击【**确认并返回】**跳转至数据总览页

导入数据 数据标注状态 无标注信息 有标注信息 导入方式 本地导入 V 上传图片 ~ 上传图片 ① 上传图片 已上传25个文件 确认并返回

在数据总览页可看到所建数据集,图片上传到平台,需要一段时间,等待片刻刷新页面后,待状态由【正在导入】转为【已完成】即为导入成 功。

提交工单

EasyData智能数据服务半台	已上线,使用EasyData可享受包	9.括多人标注、数据清洗、数据米集等	完整数据服务 立即前往			
创建数据集						我的数据集 ➤ 输入数据集名称或ⅠD Q
dog or cat 🖄 数据集组	ID: 183617					□ 新塔版本 器全部版本 直制除
版本 数据集日	D 数据量	最近导入状态	标注类型	标注状态	清洗状态	操作
V1 ⊙ 192301	101	● 已完成	图像分类	0% (0/101)	-	查看与标注 多人标注 导入 清洗 ・・・

数据标注

在数据总览页找到需要标注的数据集,点击【**查看与标注**】,跳转至标注页面

我的数据总览			提交工单									
EasyData智能型	EasyOata智能投資服务平台已上线,使用EasyOata可享受包括多人标注、数据清洗、数据清洗、数据采集等完整数据服务 立即前往											
创建数据集							我的数据集 ∨ 输入数据集名称或ⅠD Q					
dog or cat 🖂	数据集组ID: 183654						□新增版本 铝全部版本 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>					
版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作					
V1 ⊖	192344	101	● 已完成	图像分类	100% (101/101)		宣看与标注 多人标注 导入 清洗 ・・・					

在左侧标签栏下,点击【添加标签】 创建数据集标签

我的鼓摇总览 > 【图片】dog or cat/V1/宣看与标注								批量标注示例
全部 (101) 有标注信息 (0) 无标注信息 (101)							导入图片	质检报告 批量标注
筛选 ✔								- 本页全选 白 删除
标签栏 漆加标签 >			-			101	13	THE
请输入标签名称 Q					\$ 7			
根据图片内容, 选择标签	无标签	无标签	无标签	无标签	无标签	无标签	无标签	无标签
	1		F				- M	1
	无标签	无标签	无标签	无标签	无标题	无标篮	无标签	无标签
首无可用标签, 请点由上方按钮添加	3		-A.Y.		-		W	
	无标签	无标签	无标签	无标签	无标签	无标签	无标签	无标签
	Tre		N			s B		No.

分别输入dog、cat并点击【确认】添加数据标签

筛选 🗸

cat	确定 取消
-----	-------

点击图片右下角红框内图标进入到数据标注界面

全部 (101)) 有标注信息 (0) 无标注	E信息 (101)							导入图片	质检报告 批量标注
筛选 🗸										□本页全选 自删除
标签	栏 添加	₩ 1 1 1 1 1 1 1 1 1 1 1 1 1			-					
请辅	1入标签名称	Q		- 10		-	9 7			All
	根据图片内容,选择标签		无标签 区 23	无标签	无标签	无标签	无标签	无标签	无标签	无标签
标签	8	数据量』				Sec.				
cat		0		1 Carton	TOR	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	-	60	1	1. 1. 1. 1.
dog		0	J. See	-24		- 1	1212/2	(JP)		
			无标签	无标签	无标签	无标签	无标签	无标签	无标签	无标签
			A		-A.M.				Y	
			无标签	无标签	无标签	无标签	无标签	无标签	无标签	无标签

在当前图片下选择右侧标签栏内的某一类别,代表为图片打上相应的标签,点击【**保存当前图片】**或直接点击下一张图标,在保存标注结果后自 动跳转至下一张。标注完所有图片后,该数据集便可用于后续模型训练



模型训练

数据集准备完成后,点击【训练】,进入模型训练配置阶段

模型列表	操作文档	教学视频	常见问题	提交工单
创建模型				
【图像分类】 猫狗说别test 🖸 模型ID: 120920		E	昌 训练	直 删除
模型的建成功,若无数据集诵先在"数据中心"的提,上传训练数据2016模型后,可以在此处量看模型的最新版本				

根据需求选择模型各项配置后,添加训练数据集,点击【开始训练】

训练模型				操作文档	教学视频	常见问题	提交工单				
选择模型	猫狗识别test ~										
训练配置											
部署方式	公有云部署 EasyEdge本地部署 如何选择部	署方式?									
训练方式	1 「 規則法 構変 現分配置 目										
选择算法	高精度 ⑦ ○ 高性能 ⑦ ○ AutoDL Transfer ⑦ ○ Al市场已购模型										
高级训练配置											
添加数据											
添加训练数据	+ 请选择 你已经选择1个数据集的2个分类 全部清望	Ŷ									
	数据集	版本	分类数量	操作							
	dog or cat	V1	2	查看详情 清空分析	1						
	🗌 增加识别结果为【其他】的默认分类 ⑦										
自定义验证集 ⑦	OFF										
自定义测试集 ⑦	OFF										
数据增强策略 ⑦	● 默认配置 手动配置 根据所选网络、默认配置必要的数据增强策略。										

在模型列表下,可以看到处于训练状态的模型,将鼠标放置感叹号图标处,即可查看训练进度,同时若**勾选短信提醒,在模型训练完成后会以短** 信的形式通知

模型列表						操作文档 教学视频 常见问题 提交工单
创建模型						
【图像分类】猫狗识别test	应 模型ID: 1209	20				⑤ 历史版本 回 删除
部署方式	版本	训练状态	服务状态	模型效果	操作	
公有云API	V1	训练中 💷	未发布	-	宣看版本配置 停止训练	
【图像分类】hyb_不平衡测i	训 试2 区 模型	练进度: 1] 完成后短信提醒至 130****	691 🗹			吕 训练 · ⓒ 历史版本 · 回 删除

模型校验

模型训练完成后,可在模型列表下,点击【校验】

模型列表						操作文档	教学视频	常见问题	提交工单
创建模型									
【图像分类】猫狗识别test 🖸 参	輕D: 120920					日训	陈 ① 〕	万史版本	亩 删除
部署方式	版本	训练状态	服务状态	模型效果	操作				
公有云API	VI	训练完成	未发布	top1准确率: 100.00% top5准确率: 100.00% 完整评估结果	查看版本配置 申请发布 校验				

点击【**启动模型校验服务**】,需等待几分钟

图像分类模型	.≡	校验模型	
凸 模型中心			
我的模型		选择模型 猫狗识别test > 部署方式 公有云API(目前仅支 > 选择版本 V1 >	
创建模型		启动模型校验服务	
训练模型			
校验模型			
发布模型			

点击【**添加图片】**,进行模型校验



在此处可以点击【申请上线】,进行模型发布,跳转到模型发布

模型发布

模型训练完成后,点击【**申请发布**】

图像分类模型	€	模型列表						操作文档 数学视频 常见问题 握交工单				
由 模型中心												
我的模型		创建模型	建模型									
创建模型												
训练模型		LEATIN 20 Sec 1 SHIPS (2003) (C	【图像分类】 拥有说册test └/ 模型ID: 120920									
校验模型		部署方式	版本	训练状态	服务状态	模型效果	操作					
发布模型												
☑ EasyData数据服务		22 ft ZIAPI	V1	训练完成	木友巾	top5准确单: 100.00% 完整评估结果	宣看版本配直 申请发作 校运					
数据总宽												

按要求填写相应信息后,点击【**提交申请】**

注:同时可勾选下方【云服务调用数据管理】的服务条款,通过云服务调用数据反馈,可查找公有云服务模型识别错误的数据,纠正结果并将其 加入下一次用于训练模型的数据集,实现训练数据的持续丰富和模型效果的持续优化。

图像分类模型 (三	发布模型		操作	文档 教学视频	栗 常见问题 提交工单		
① 模型中心 我的模型 向建模型	选择调型 建物识别hest V	标准接口规 _{标准接口请求}	芭参考 参考说班				
训练模型	部着方式 公判広部者 >	学段名称	必须	类型	说明		
校验模型 发布模型	洗月線会 V1 V 服务名称・ (20)編和時	image	R	string	習像数据, base64編 码, 要求base64編码后 大小不超过4M, 最短辺 至少15px, 最长边最大 4096mv 支放		
回 EasyData数据服务	接口地址。 https://aip.baidubce.com/rpc/2.0/ai_custom/v1				jpg/png/bmp格式		
数据总览	(classification/ dogorcat	top_num	8	number	返回分类数量, 数13为6 个		
标签组管理 在线标注	其物要求	标准接口响应字段说明:					
云服务数据回流		字段名称	必须	実型	说明		
摄像头数据采集	0000	log_id	是	number	睡一的log id,用于问题 定位		
23 模型部署	2 四章子短条转道向净短条条数并开语和条	results	8	array(object)	分类结果数组		
纯离线服务		+ name	ă	string	分类名称		
端云协同服务 ^	提交申请	+ score	8	number	重信度		
我的部署包							
我的本地设备							
下发部署包到设备							

提交申请后跳转至【**我的模型**】栏,服务状态变为【发布中】

图作	象分类模型	•≡	模型列表		操作文档 教学视频 常见问题 提交工单								
db	模型中心												
	我的模型		创建模型										
	创建模型												
	训练模型		【出像方束】 通利证制tei	【图像分英】 獲物(UB)test 区 機型(D: 120920									
	校验模型		部署方式	版本	训练状态	服务状态	模型效果	操作					
	发布模型					_	top1准确率: 100.00%						
	EasyData数据服务		公有云API	V1	训练完成	发布中	top5准确率: 100.00% 完整评估结果	查看版本配置 校验					
	對編幕常												

等待几分钟,此状态就会变为【**已发布】**,即发布成功

图像分类模型	·Ξ	模型列表						操作文档 孝	《学视频 常见问	题 提交工单
由 模型中心										
我的模型		创建模型								
创建模型		(mache) (mache)	et E/ ###UD: 120020					C2 2014F	Generation Generation	
训练模型		10218/20 5421 38295 0729108	St 🖂 WE 10. 120920					- 10.57.	G IJER4	E1 98945
校验模型		部署方式	版本	训练状态	服务状态	模型效果	操作			
发布模型					_	top1准确本: 100.00%				
☑ EasyData数据服务		公有云API	V1	训练完成	已发布	top5准确率:100.00% 完整评估结果	查看版本配置 服务详情 校验 体验H5 ⑦			
数据总览										

体验H5

模型发布成功后可在模型列表页点击【体验H5】进行模型体验

模型列表						操作文档 教学視频 常见问题 提交工
创建模型						
【图像分类】猫狗识别te	est 🖸 模型ID: 120920					曰: 训练 ⑤ 历史版本 前 删除
部署方式	版本	训练状态	服务状态	模型效果	操作	
公有云API	V1	训练完成	已发布	top1准确率: 100.00% top5准确率: 100.00% 完整评估结果	查看版本配置 服务详情 校验 体验H5 ⑦	

选择体验H5的模型并点击下一步

体验H5									
1 体验	ѝH5说明 ② 自定义样式	3 完成							
・H5中的商品	品检测功能将使用你的APP进行调用。 ^边 测将消耗个人帐号下的调用次数								
调用APP:	猫狗识别test-APPID: 24305311	\sim							

自定义样式后点击【**生成H5】**

	体验H5		×				
① 体验H5说明 2 自定义样式 ③ 完成							
	名称	猫狗识别					
猫狗识别	模型介绍	识别猫和狗的模型					
识别猫和狗的模型	开发者署名	Jace					
开发者jace 1 ··································	H5分享文案	jacetest 8/50	2				
	生成H5						
机扫描生成的二维码即可在手机端位	本验模型效果						
	体验H5		×				
① 体验H5说明	2 自2	定义样式 —— 3 完成					

用百度或微信APP扫以下二维码,在手机端体验模型效果



の Step4:模型调用

在 EasyDL"我的模型"列表页,点击【服务详情】后,会得到接口地址

模型列表						操作文档 教学视频 常见问题 提交工单
创建模型						
【图像分类】猫狗识别tes	t 🖸 模型ID: 120920					吕 训练 ⓒ 历史版本 亩 删除
部署方式	版本	训练状态	服务状态	模型效果	操作	
公有云API	V1	训练完成	已发布	top1准确率:100.00% top5准确率:100.00% 完整评估结果	宣看版本配置 服务详情 校验 体验H5 ⑦	

此接口地址在模型调用代码中会用到。点击【立即使用】

	版为详旧
服务名称:	猫狗识别
模型版本:	V1
接口地址:	https://aip.baidubce.com/rpc/2.0/ai_custom/v1/classification/dogorcat test
,	
服务状态:	已发布

经典版		应用列表					
公有云服务	^	+ 创建应用					
 • 应用列表 		应用名称	AppID	API Key	Secret Key	创建时间	操作

填写信息后点击【**立即创建】**

* 应用名称:	猫狗识别test
* 接口选择:	部分接口免费额度还未领取,请先去领取再创建应用,确保应用可以正常调用 去领取
	勾选以下接口,使此应用可以请求已勾选的接口服务,注意EasyDL图像服务已默认勾选并不可 当
	//3 o
	您已升通付费的接口为:短语首识别-甲文普通话、短语首识别-粤语、短语首识别极速版、实时
	语音识别-中文晋通话、实时语音识别-英文、通用文字识别(高精度版)、通用文字识别(高精
	度含位置版)、网络图片文字识别、身份证识别、银行卡识别、驾驶证识别、数字识别、手写文
	字识别、火车票识别、iOCR通用版、H5视频活体检测、驾驶行为分析、人脸融合、中文词向量
	表示、词义相似度、文章分类、内容审核平台-图像、内容审核平台-文本、知识问答、通用物体
	和场景识别高级版、相同图检索-入库、相似图搜索-入库、手部关键点识别、货架拼接-货架拼
	接、黑白图像上色、人脸实名认证、人脸实名认证
	+ EasyDL
	+ 语音技术
	★ 文字识别
	+ 人脸识别
	+ 自然语言处理
	中 内容审核 Ⅰ
	+ UNIT I
	➡ 知识图谱
	● 图像识别 ●
	+ 智能呼叫中心
	+ 图像搜索
	+ 人体分析
	● 图像增强与特效 ●
	+ 智能创作平台
	+ EasyMonitor
	+ BML
	+ 机器翻译
* 应用描述:	可以快速识别猫和狗的模型

立即创建后,在应用列表页即可得到 AK SK 密钥

应用列	表					
+ 8	建应用					
	应用名称	AppID	API Key	Secret Key	创建时间	操作
1	猫狗识别test	24305311	fMq7vTINNoGqUxrlidahP1fq	显示	2021-06-03 20:16:00	报表 管理 删除

通过使用在线API测试所训练的模型效果

```
import sys
import time
import socket
import json
import base64
import requests
from datetime import datetime
print(datetime.now())
domain = "aip.baidubce.com"
myaddr = socket.getaddrinfo(domain,'https')
print(str(domain) + " = " + myaddr[0][4][0])
start = time.time()
appid = 'appid'
client_id = 'AK'
client_secret = 'SK'
host = 'https://aip.baidubce.com/oauth/2.0/token?grant_type=client_credentials' host += "&client_id=%s&client_secret=%s" %
(client_id, client_secret)
session = requests.Session()
response = session.get(host)
access_token = response.json().get("access_token")
request_url = "【模型信息-服务详情-接口地址】"
with open('image.jpg', 'rb') as f:
  image = base64.b64encode(f.read()).decode('UTF8')
headers = {
'Content-Type': 'application/json'
params = {
"image": image
request_url = request_url + "?access_token=" + access_token
response = session.post(request_url, headers=headers, json=params)
content = response.content.decode('UTF-8')
print(json.loads(content))
end = time.time()
print('耗时时长: %1.2f s'% (end-start))
```

心 产品特色

可视化操作: 无需机器学习专业知识,模型创建·数据上传-模型训练-模型发布全流程可视化便捷操作,最快15分钟即可获得一个高精度模型

高精度效果:EasyDL底层结合百度 AutoDL/AutoML技术,针对用户数据自动获得最优网络和超参组合,基于少量数据就能获得出色效果和性能的模型

端云结合:训练完成的模型可发布为云端API或离线SDK,灵活适配各种使用场景及运行环境

数据支持:全方位支持训练数据的高质量采集与高效标注,支持在模型迭代过程中不断扩充数据,助力提升模型效果

● 更多参考

如对文档说明有疑问或建议,请**微信搜索"BaiduEasyDL"**添加小助手交流 备注:文档如使用中遇到报错等问题,请在控制台中通过"工单"联系我们,售后团队为您及时解决问题

EasyDL官网入口 EasyDL开发文档 EasyDL软硬一体方案 EasyDL应用案例

EasyDL图像-物体检测快速开始

目录

- 2. 实现步骤
- 3. 产品特色
- 4. 更多参考

```
心 场景介绍
```

在工业场景下,零件分拣、零件计数是在工业生产过程中的一个常见业务需求,由于零件样式多样,市面上没有现成的零件识别服务可以直接使用,往往需要定制企业零件专用的图像识别能力。某个工业质检领域的服务商收到甲方需求,希望能在工厂中实现检测螺丝和螺母,并借助机械 臂等装置协助人工完成自动分拣。该服务商经过详细了解需求,发现螺丝与螺母由于排列密集且没有规律,如果要配合机械臂完成就需要精准定 位出图片中每个零件的名称及位置。经过调研,由于市场上并没有任何一家公司有现成的螺丝螺母识别服务,同时该服务商缺少相关AI算法工程 师及算力资源,如果筹备相关技术及资源成本较高,成为该服务商面临的一大难题。无意中了解到百度EasyDL可以灵活定制并可以快速上手获得 业务所需的高精度AI能力,刚好可以解决该服务商面临的问题。

心 实现步骤

只需四步即可完成自定义AI模型的训练及发布的全过程。

心 Step1:成为百度AI开放平台的开发者

要使用百度EasyDL的模型训练能力首先需要注册成为百度AI开放平台的开发者,首先让我们用5分钟来注册百度AI开放平台的开发者(如您已经是 开发者,可直接登录使用)。

先点击此处注册百度账号进入,如下图的页面快速的注册一个百度账号吧。

用户名	请设置用户名	
于机亏	可用于登求相找四密码	
密码	请设置登录密码	C
验证码	请输入验证码	获取验证码

₯ Step2:提前准备训练数据

物体检测需要提供包含目标物体的图片并标注物体即可训练物体检测模型,自动识别图中所有目标物体的位置、名称,下面我们来看看这次需要 计数的包含螺丝螺母的图片示例:



图片数量越多理论上训练的效果就越好,物体检测的图片数据理建议不低于20张图片 注意图片需要为业务生产的真实环境所采集的图片,与真 实场景越贴近,训练模型效果越佳

₯ Step3:使用EasyDL训练物体检测模型

创建模型

进入EasyDL官方平台 点击【立即使用】



点击【物体检测】,进入操作台

	:		
▶ 图像分类	③ 物体检测	🔼 图像分割	1 文本分类-单标签
文本分类-多标签	三 短文本相似度	👛 情感倾向分析	立 文本实体抽取
文本实体关系抽取	● 语音识别	● 声音分类	1 视频分类
◎ 目标跟踪	表格数据预测	1 时序预测	膏 零售行业版
OCR			
图像 ● 文本 ● 语音 ● 视	頃 ● 结构化数据 ● 行业版 ●	OCR	

在模型列表下点击【创建模型】

Bal 3大師 EasyDL	产品介	招 操作平台 应用案例	使用文档					社区	飞桨官网	AI主站	💽 f581	(1G9x5hG
物体检测模型	·≡	模型列表							操作	文档 数4	全視频 常见问』	题 提交工单
由 模型中心 我的模型		创建模型										
创建模型 训练模型		【物件检测】 销钉缺失2 🖄 模	أألى 119283						1	3 शाह	⑤ 历史版本	宫 删除
校验模型		部署方式	版本	训练状态	服务状态	模型效果	操作					
发布模型 ご EasyData数据服务 数据总算		公有云API	V2	训练完成	未发布	mAP: 73.91% ⑦ 精确率: 72.31% ⑦ 召回率: 72.31% ⑦ 完整评估结果	查看版本配置 申请发布 校验					
标签组管理												
在线标注		【物体检测】 铆钉缺失 🖄 模式	법ID: 119268						1	∃ i/l±5	⑤ 历史版本	宫 删除
智能标注		部署方式	版本	训练状态	服务状态	模型效果	操作					
云程务数据回流 摄像失数据采集 配 模型部署 纯高线服务 编云协问服务		公有云API	V2	训练完成	未发布	mAP: 71.00%⑦ 精确率: 79.67%⑦ 召回率: 71.08%⑦ 完整评估结果	宣看极本配置 申请发布 校验					
	^	【物体检测】标注牌识别 🖸 机	夏型ID: 119267						1	昌 训练	⑤ 历史版本	宫 删除
我的命告已		部署方式	版本	训练状态	服务状态	模型效果	操作					
下发部署包到设备		公有云API	V2	训练中止	未发布	-	宣看版本配置					

填写模型信息后,点击**【下一步】**

Baid	o大脑 ∣ EasyDL	产品介绍	操作平台	应用案例	使用文档	
物体植	佥测模型	•≡	模型列表 > 仓]]建模型		
品植	模型中心					
劧	线的模型		模型类别	物体检测		
Ê	创建模型		模型名称 *	螺丝螺母识别		
ý	练模型		模型归属	公司 个人	~	
杉	交验模型					
发	这布模型		邮箱地址 *	l*******@163.0	com	
E	asyData数据服务		联系方式 *	130*****691		0
判支	牧据总览		功能描述 *	丁业场景中螺丝.	螺母识别	
枋	示签组管理				AL	
右	E线标注					12/500
칱	冒能标注					//
z	云服务数据回流			下一步		
扳	摄像头数据采集					

模型创建完成后可在【我的模型】栏查看已创建的模型信息

模型列表	操作文档	教学视频	常见问题	提交工单
创建构变				
【物体检测】螺丝螺号识别 ^[2] 模型ID: 120958		E	副训练	直 删除
模型创建成功,若无数据集调先在"数据中心"创建。上传训练数据训练模型后,可以在此处重着模型的最新版本				

创建数据

在数据总览界面点击【**创建数据集**】

初体检测模型 1 二	我的數据認见							提父上単
由 模型中心	EasyData智能	8数据服务平台已上线,	使用EasyData可享受包括	多人标注、数据清洗、数据采集等	导完整数据服务 立即前往			
创建模型	创建数据集	ו						我的数据集 > 输入数据集名标或D Q
训练模型	模拟数据回流	■ : F/ 数振集组ID: 18	3682					口斯爆炼本 别全部版本 向删除
发布模型	版本	数据集iD	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作
☑ EasyData数据服务	V1 Θ	192376	69	• 已完成	物体检测	100% (69/69)	-	宣看与极注 多人标注 导入 清洗 ・・・
数数总0. 标签组管理								
在线标注	相對病恐短征		3020 Refere	展活版 \ 好 來	40-10-200 9 91	B-19-X	德法仲式	ftware 22 to an
智能标注 云服务数据回流	VI O	192305	20	 P支柱 	物体检测	100% (20/20)		ANIF 奇景与続注 名人経注 章人 清祥 ・・・
摄像头数据采集			-					
73 模型部署 纯离线服务	coco-dataset	t-coco2014-train 🖄	数据集组ID: 183352					□新婚岐本 器全部版本 白 删除
端云协同服务 ^	版本	数据集ID	数班量	最近导入状态	标注类型	标注状态	清洗状态	操作
我的部署包	V1 😔	192006	0	● 导入失败 ⑦	物体检测	0% (0/0)		多人标注 导入 删除 共享

在数据集创建界面输入数据集名称后点击【完成】

数据集名称	螺丝螺母识别	
数据类型	图片	
数据集版本	V1	
标注类型	物体检测	
标注横板	拍形框标注	~

数据集创建完成后可在【数据总览】查看已创建完成的数据集,点击【导入】跳转至数据导入界面

我的教戚总咒 提交工4									
EasyData智能数据	EasyGata智能数据服务平台已上组,使用EasyGata可享受包括多人标注、数据清洗、数据常集等完整数据服务 立即前往								
创建数据集							我的数据集 ∨ 输入数据集名称或ⅠD Q		
螺丝螺母识别 区	数据集组ID: 183904						□新增版本 铝全部版本 白 删除		
版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作		
V1 😔	192625	0	● 已完成	物体检测	0% (0/0)	-	多人标注 导入 删除 共享		

数据导入支持无标注信息、有标注信息两种数据标注状态的数据以及多种导入方式,以下为无标注信息·本地导入-上传图片的导入方式示例,其 余各类型导入方式可参考页面上传物体检测数据集并在线标注

	导入数据			
	数据标注状态	● 无标注信息	○ 有标注信息	
选择数据标注状态及导入方式后点击【 上传图片】	导入方式	本地导入	~	上传图片 ~
	上传图片	<u> 土</u> 上传图片		
		确认并返回		




选择好图片后,点击【开始上传】

			上传图片			×								
	alla Solo Jolo			App of	84 Jo									
	删除	删除	删除	删除	删除									
	(Mo)	5	Corgo a	s les	(F)									
	删除	删除	删除	删除	删除									
	删除 删除 删除 删除													
上传题	完成后,点击 【确认	并返回】跳转至数	据总览页											
Ę	入数据													
楼	据标注状态	● 无标注信息	🔵 有标注信											
ę	入方式	本地导入		~ 上传图片	~									
上	传图片	▲ 上传图片	已上传25个文	件										

确认并返回

在数据总览页可看到所建数据集,图片上传到平台,需要一段时间,等待片刻刷新页面后,待状态由【正在导入】转为【已完成】即为导入成 功。

Baidu 百度智能云文档

我的数据总览

新手指南

提交工单

EasyData智能数据服务平台已上线,使用EasyData可享受包括多人标注、数据采集等完整数据服务 立即前往										
创建数据集							我的数据集 >	输入数据集名	称或ID Q	
螺丝螺母识别 🖄	数据集组ID: 183904						□ 新增版本	HB 全部版本	白 删除	
版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作			
V1 💬	192625	69	●已完成	物体检测	0% (0/69)	-	查看与标注 多人	赤注 导入 清	院 …	

数据标注

在数据总览页找到需要标注的数据集,点击【**查看与标注】**,跳转至标注页面

我的数据总说											
EssyOata報能数据服务平台已上线,使用EssyOata可享受包括多人标法、数据清洗、数据采集等完整数据服务 立即前往											
创建数据集							我的数据集 ➤ 输入数据集名称或ⅠD Q				
螺丝螺母识别 🕑	ú 数据集组ID: 183904						□ 新增版本 铝全部版本 直 删除				
版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作				
V1 😔	192625	69	● 已完成	物体检测	0% (0/69)	-	查看与标注 多人标注 导入 清洗 …				

在左侧标签栏下,点击【添加标签】 创建数据集标签

说的教振总览 > 【图片】螺丝螺母识别/VI/查看与标注												
全部 (69) 有标注信息 (0) 无标注信息 (69)								导入图片质检报告				
筛选 🗸								□本页全选 □ 删除				
病益株益 添加株益 请输入标签名称 Q	(STAR	A spec	100 m		No.2	of the	() ()	a p				
	无标签	无标签	无标签	无标签	无标签	无标签	无标签	无极签				
		Jost 1	1 of o		(H)	to to		×				
	无标签	无标签	无标签	无标签	无标签	无标签	无标签	无标签				
新 无可用标签,请点击上方按钮添加						100		a dia				
	无标签	无标签	无标签	无标签	无标签	无标签	无标签	无标签				

分别输入需要创建的标签名称并点击【确认】添加数据标签

点击图片右下角红框内图标进入到数据标注界面

我的数据总览 > 【图片】螺丝	螺母识别/V1/查看与标注								
全部 (69) 有标注信息 (0)	无标注信息 (69)								导入图片质检报告
筛选 🗸									本页全选 回 删除
标签栏	漆加标签 🗸	2-180	(Bar)		Ster.	21			
请输入标签名称	Q	38	A of A	0.70		1000	of		100
标签名	标注框数に	无标签 🖸 🕄	无标签	无标签	无标签	无标签	无标签	无标签	无标篮
螺母	0								
螺丝	0	alandr alandr	N.I	1 20		2	1. St.	and a second sec	- AF
		无根签	无标签	无标签	无标签	无标签	无标签	无标签	无标签
			×				0 0 0		2 J 0
		无标签	无标签	无标签	无标签	无标签	无标签	无标签	无标签

在当前图片下分别框选出图片中的目标物体并添加标签,点击【**保存当前图片】**或直接点击下一张图标,在保存标注结果后自动跳转至下一张。 标注完所有图片后,该数据集便可用于后续模型训练



模型训练

数据集准备完成后,点击【训练】,进入模型训练配置阶段

模型列表	操作文档	教学视频	常见问题	提交工单
创建传知				
【物体检测】螺丝螺母识别 IC 模型ID: 120958		E	冒 训练	団 删除
模型创建成功,若无数据集请先在"数据中心"创建,上传训练到43 ¹ 1015月21后,可以在此处重着模型的最新版本				

根据需求选择模型各项配置后,添加训练数据集,点击【开始训练】

训练模型				操作乡	、档 教学视频	常见问题	提交工单				
选择模型	螺丝螺母识别 >										
训练配置											
部署方式	公有云部署 EasyEdge本地部署 如何选择部	署方式?									
训练方式	常規训练 精度提升配置包										
选择算法	超高精度 ⑦ (高档能 ⑦ (AI市场已购模型										
高级训练配置	OFF										
添加数据											
添加训练数据	+ 请选择 你已经选择1个数据集的2个标签 全部清	Ŷ									
	数据集	版本	标签数量	操作							
	螺丝螺母识别1	V1	2	宣看详情 清空	标签						
自定义验证集 ⑦	OFF										
自定义测试集 ⑦	OFF										
数据增强策略 ⑦	◆ 默认配置 手动配置 規模所造网络 野认配置必要的数据增强策略										

在模型列表下,可以看到处于训练状态的模型,将鼠标放置感叹号图标处,即可查看训练进度,同时若**勾选短信提醒,在模型训练完成后会以短** 信的形式通知

模型列表						操作文档 教学视频 常见问题 提交工
创建模型						
【物体检测】螺丝螺母识别	⑤ 历史版本 直 删除					
部署方式	版本	训练状态	服务状态	模型效果	操作	
公有云API	V1	训练中 🔳	未发布	-	宣看版本配置 停止训练	
	训练进度: 1%					
【物体检测】聚合物红外光	谱快速定性定!	完成后短信提醒至 130****	*691 🗹			曰 训练 〇 历史版本 前 删除

模型校验

模型训练完成后,可在模型列表下,点击【校验】

模型列表						操作文档 教学视频 常见问题 提交工单					
创建模型											
【物体检测】螺	丝螺母识别 🗹 模型ID: 120958					🖂 训练 🔿 历史版本 💼 删除					
部署方式	版本	训练状态	服务状态	模型效果	操作						
公有云API	VI	训练完成	未发布	mAP: 100.00% ⑦ 精韻率: 100.00% ⑦ 召回率: 100.00% ⑦ 完整评估结果	查看版本配置 申请发布 校验						
点击【启	点击 【启动模型校验服务】 ,需等待几分钟										
校验模型											

选择模型	螺丝螺母识别	~	部署方式	公有云API(目前仅支 🗸	选择版本	V2	~
启动模型机	交验服务						

点击【**添加图片**】,进行模型校验



在此处可以点击【申请上线】,进行模型发布,跳转到模型发布

模型发布

模型训练完成后,点击【申请发布】

物体检测模型	÷	模型列表						操作文档 執	学视频 常见间	题 提交工单	
曲 模型中心											
我的模型		创建模型									
创建模型		F45/# 16/01 #844#8 (0/00							○ Emi€±	-	
训练模型		[+0)+12/0] 38.2238-0 073	【物体检测】 糖盐糖苷试测 22 機型ID: 120958								
校验模型		部署方式	版本	训练状态	服务状态	模型效果	操作				
发布模型						mAP: 100.00% ③					
回 EasyData数据服务		公有云API	V1	训练完成	未发布	精确率: 100.00% ⑦ 召回率: 100.00% ⑦	宣看版本配置 申请发布 校验				
数据总览						完整评估结果					

按要求填写相应信息后,点击【提交申请】

注:同时可勾选下方【云服务调用数据管理】的服务条款,通过云服务调用数据反馈,可查找公有云服务模型识别错误的数据,纠正结果并将其 加入下一次用于训练模型的数据集,实现训练数据的持续丰富和模型效果的持续优化。 发布模型

)件 +及 +带 开川	바뀌스스 슈핑 드리 그 그 그 그 그 그 그 그 그 그 그 그 그 그 그 그 그 그	
远拜侯空	蠓丝蠓�识别 ◆	
部署方式	公有云部署 🗸 🗸	
选择版本	V1 ~	
服务名称 *	识别螺丝和螺母	
接口地址 *	https://aip.baidubce.com/rpc/2.0/ai_cus /detection/ jacetest	tom/v1
其他要求	若接口无法满足您的需求,请描述希望 的问题,500汉字以内	解决)/500
		11
[✔ 同意云服务数据回流服务条款并开通用	服务
[提交申请	

提交申请后跳转至**【我的模型】**栏,服务状态变为**【发布中】**

模型列表						操作文档 教学视频 常见问题 提交工单
创建模型						
【物体检测】螺丝螺母识	R别 区 模型ID: 120958					吕训练 〇 历史版本 回 删除
部署方式	版本	训练状态	服务状态	模型效果	操作	
公有云API	V1	训练完成	发布中	mAP: 100.00% ⑦ 精确率: 100.00% ⑦ 召回率: 100.00% ⑦ 完整评估结果	宣看版本配置 校验	

等待几分钟,此状态就会变为【**已发布**】,即发布成功

图像分类模型	·Ξ	模型列表						操作文档 豢	数学视频 常见	问题 提交工单
由 模型中心										
我的模型		创建模型								
创建模型		【照确公准】 SER(D2)to	+ 12、横型10:120920					24/02	() F PE	TT #4/52
训练模型		1010/09/1907	A					C2 400.	O Mickey	r El anire
校验模型		部署方式	版本	训练状态	服务状态	模型效果	操作			
发布模型						top1准确率: 100.00%				
☑ EasyData数据服务		公有云API	V1	训练完成	已发布	top5准确率: 100.00% 完整评估结果	查看版本配置 服务详情 校验 体验H5 ③			
数据总览										

体验H5

模型发布成功后可在模型列表页点击【体验H5】进行模型体验

模型列表						操作文档 教学视频 常见问题 提交工单
创建模型						
【物体检测】螺丝螺母识别 [☑ 模型ID: 120958	5				吕 训练 ① 历史版本 直 删除
部署方式	版本	训练状态	服务状态	模型效果	操作	
公有云API	V1	训练完成	已发布	mAP: 100.00% ⑦ 精确率: 100.00% ⑦ 召回率: 100.00% ⑦ 完整评估結果	查看版本配置 服务详情 校验 体验HS	

选择体验H5的模型并点击下一步

1 体验	金H5说明 —— ② 自定义样式 —— ③ 完成	
·H5中的商品	品检测功能将使用你的APP进行调用。	
·每次体验检	金测将消耗个人帐号下的调用次数	
调用APP:	螺丝螺母识别-APPID: 24311653 V	

自定义样式后点击【**生成H5】**

\sim				
(1) 体	验H5说明	2 自知	主义样式 —— ③ 完成	
		名称	螺丝螺母识别	
螺丝螺母识别		模型介绍	识别螺丝螺母	
识别螺丝螺母		开发者署名	Jace	
(1) (1) (1) (1) (1) (1) (1) (1) (1) (1)		H5分享文案	jacetest 8/	50

手机扫描生成的二维码即可在手机端体验模型效果

体验H5	×
 体验H5说明 (2) 自定义样式 (3) 完成 	
用百度或微信APP扫以下二维码,在手机端体验模型效果	
■ ●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●	
完成	

在 EasyDL"我的模型"列表页,点击【服务详情】后,会得到接口地址

模型列表							操作文档 教学	视频 常见问题	提交工单
创建模型									
【物体检测】螺丝螺母识别	区 模型ID: 120958						吕 训练	⑤ 历史版本	田 删除
部署方式	版本	训练状态	服务状态	模型效果	操作				
公有云API	V1	训练完成	已发布	mAP: 100.00% ⑦ 精确率: 100.00% ⑦ 召回率: 100.00% ⑦ 完整评估结果	宣看版本配置 服务详情 4	交验 体验H5 ⑦			
此接口地址在林	模型调用代码	码中会用到。	点击【 立 】	即使用】					
			服务	5详情			×		
服务名称	: 识别螺	丝和螺母							
模型版本	:: V1								
接口地址	: https:/	/aip.baidub	ce.com/rp	c/2.0/ai_custo	m/v1/detection/j	acetest111			
服务状态	: 已发布								
		Ţ	ℤ即使用	查看API文	档				
需要登录Easy[DL控制台中	创建一个应用	刊 , 点击【	创建应用】					
经典版	应用列表								
公有云服务 へ	+ 创建应用								
• 旭用列表	应用名利	<u>я</u>	AppID	API Key	Secret Key	创建时间	操作		

填写信息后点击【**立即创建】**

Baidu	百度智能云文材	当
-------	---------	---

创建新应用	
-------	--

• 应用名称:	螺丝螺母识别	
接口洗择:	部分接口免费额度还未领取,请先去领取再创建应用,确保应用可以正常调用	去领取
NH 21	勾选以下接口,使此应用可以请求已勾选的接口服务,注意EasyDL图像服务	已默认勾选并不可取
	消。	
	您已开通付费的接口为:短语 <mark>音识别-中文普通话、短语音识别-粤语、短语</mark> 音	f识别极速版、实时
	语音识别-中文普通话、实时语音识别-英文、通用文字识别(高精度版)、通	師文字识别(高精
	度含位置版)、网络图片文字识别、身份证识别、银行卡识别、驾驶证识别、	数字识别、手写文
	字识别、火车票识别、iOCR通用版、H5视频活体检测、驾驶行为分析、人脸	融合、中文词向量
	表示、词义相似度、文章分类、内容审核平台-图像、内容审核平台-文本、知	口识问答、通用物体
	和场景识别高级版、相同图检索-入库、相似图搜索-入库、手部关键点识别、	货架拼接-货架拼
	接、黑白图像上色、人脸实名认证、人脸实名认证	
	+ EasyDL	
	 于 语音技术 	
	 + 文字识别 	
	➡ 人脸识别	
	+ 自然语言处理	
	+ 内容审核 !	
	+ UNIT !	
	➡ 知识图谱	
	➡ 图像识别 !	
	+ 智能呼叫中心	
	➡ 图像搜索	
	+ 人体分析	
	+ 图像增强与特效	
	+ 智能创作平台	
	+ EasyMonitor	
	+ BML	
	 	
· 应用描述:	可快速识别螺丝和螺母的模型	
即创建后,在应用	用列表页即可得到 AK SK 密钥	
用列表		
+ 创建应用		
应用名称	AppID API Key Secret Key	创建时间 操作

通过使用在线API测试所训练的模型效果

```
import sys
import time
import socket
import json
import base64
import requests
from datetime import datetime
print(datetime.now())
domain = "aip.baidubce.com"
myaddr = socket.getaddrinfo(domain, 'https')
print(str(domain) + " = " + myaddr[0][4][0])
start = time.time()
appid = 'appid'
client_id = 'AK'
client_secret = 'SK'
host = 'https://aip.baidubce.com/oauth/2.0/token?grant_type=client_credentials' host += "&client_id=%s&client_secret=%s" %
(client_id, client_secret)
session = requests.Session()
response = session.get(host)
access_token = response.json().get("access_token")
request_url = "【模型信息-服务详情-接口地址】"
with open('image.jpg', 'rb') as f:
  image = base64.b64encode(f.read()).decode('UTF8')
headers = \{
'Content-Type': 'application/json'
params = {
"image": image
request_url = request_url + "?access_token=" + access_token
response = session.post(request_url, headers=headers, json=params)
content = response.content.decode('UTF-8')
print(json.loads(content))
end = time.time()
print('耗时时长: %1.2f s'% (end-start))
```

```
心 产品特色
```

可视化操作: 无需机器学习专业知识,模型创建·数据上传·模型训练·模型发布全流程可视化便捷操作,最快15分钟即可获得一个高精度模型

高精度效果:EasyDL底层结合百度 AutoDL/AutoML技术,针对用户数据自动获得最优网络和超参组合,基于少量数据就能获得出色效果和性能的模型

端云结合:训练完成的模型可发布为云端API或离线SDK,灵活适配各种使用场景及运行环境

数据支持:全方位支持训练数据的高质量采集与高效标注,支持在模型迭代过程中不断扩充数据,助力提升模型效果

● 更多参考

如对文档说明有疑问或建议,请**微信搜索"BaiduEasyDL"**添加小助手交流 备注:文档如使用中遇到报错等问题,请在控制台中通过"工单"联系我们,售后团队为您及时解决问题

EasyDL官网入口 EasyDL开发文档 EasyDL软硬一体方案 EasyDL应用案例

EasyDL文本-文本分类单标签快速开始

目录

1. 场景介绍

2. 实现步骤

- 3. 产品特色
- 4. 更多参考

心 场景介绍

目前不少互联网内容平台或者电商平台中有用户评论模块,往往都需要人工维护评论信息,如将评论信息中好的评论与坏的评论进行分类,或者 将评论信息中的广告信息能有效过滤/甄别出来,当评论内容越来越高时,人工维护评论的成本就越高。而越来越多的垂直内容平台由于评论信 息内容多样,内容不一,为实现更好的文本分类效果,往往需要定制企业专属的文本分类能力。

某个酒店信息聚合平台,希望在其官网的评论模块中增加自动分类功能,能支持将好的评价和不好的评价自动分类,方便酒店管理者在后台查 看,同时该网站缺少相关AI算法工程师及算力资源,如何能更高效更低成本的获取定制文本分类服务,成为该服务商面临的一大难题。无意中了 解到百度EasyDL可以灵活定制并可以快速上手获得业务所需的高精度AI能力,刚好可以解决该服务商面临的问题。

心 实现步骤

只需四步即可完成自定义AI模型的训练及发布的全过程。

心 Step1:成为百度AI开放平台的开发者

要使用百度EasyDL的模型训练能力首先需要注册成为百度AI开放平台的开发者,首先让我们用5分钟来注册百度AI开放平台的开发者(如您已经是 开发者,可直接登录使用)。

先点击此处注册百度账号进入,如下图的页面快速的注册一个百度账号吧。

用户名	请设置用户名	
手机号	可用于登录和找回密码	
密 码	请设置登录密码	0
佥证码	请输入验证码	获取验证码

₀ Step2:提前准备训练数据

文本分类单标签可实现文本内容的自动分类,为每个文本定义一个标签类型。数据数量越多理论上训练效果越好,本次示例以线上公开数据集为例,具体数据处理全过程可参考文本分类数据集创建

心 Step3:使用EasyDL训练文本分类

创建模型

进入EasyDL官方平台 点击【立即使用】

EasyDL零门槛		
asyDL入选Forrester领导者象限		
零算法基础定制		
高精度AI模型	EasyDL重磅上线目标跟踪模型	
【了解】EasyDL简介	检测并识别视频中的运动对象 实现目标跟踪与计数	
■【参考】应用案例 >	立即尝试	
《使用】快速上手 >		
• 【购买】平台定价 >	"万有引力"计划 万元礼包助力万家企业 EasyDL近期能力升级快报 定制OCR模型好帮手:EasyDL OCR上线毗	

点击【**文本分类-单标签】**,进入操作台

	选择机	臭型奀型	
图像分类	◎ 物体检测	🔼 图像分割	文本分类-单标签
文本分类-多标签	三 短文本相似度	👛 情感倾向分析	文本实体抽取
1 文本实体关系抽取	● 语音识别	🕕 声音分类	▶ 视频分类
) 目标跟踪	表格数据预测	1 时序预测	零 售行业版
OCR			

在模型列表下点击【**创建模型】**

文本分类模型	Æ	模型列表								提交工单
6 模型中心 我的模型		创建模型								
创建模型 训练模型		【文本分类】erniem_m_zxl 已 部要方式	5 模型ID: 10894 版本	12	服务状态	模型效果	操作	음 1115	⑤ 历史版本	□ 刪除
X布模型 ど EasyData数据服务		公有云API	vi @	训练完成	未发布	准确事: 89.84% F1-score: 0.473 完整评估结果	宣看版本配置 申请发布 校验			
数据总览 公开数据集 标签组管理		EasyEdge本地部署-服务器	VI	训练完成		准确率: 89.84% F1-score : 0.473 完整评估结果	查看版本配置 申请发布			

填写模型信息后,点击**【下一步】**

模型列表 > 创建模型



模型创建完成后可在【我的模型】栏查看已创建的模型信息

模型列表		提交工单
0184-82		
【文本分类】 酒店评分 🖸 機製IO: 121263	吕 训练	亩 删除
模型创建成功,您可以选择使用公开致露集训练模型体验分类效果,或创建数据集使用自有数据进行专属分类模型的训练,模型训练后,可以在此处查看模型的最新版本和状态		

选择数据集

在模型列表页找到新建的模型并选择【公开数据集】查看现有公开数据集

模型列表		提交工单
0.18 95.20		
【文本分类】 酒店评分 区 横型ID:121263	吕 训练	□ 删除
模型创建成功,您可以选择使 <mark>非公开款原集</mark> 邮练模型体验分类效果,成创建款原集使用自有数据进行专属分类模型的训练。模型训练后,可以在此处重要模型的最新版本和状态		

本次训练将使用chnsenticorp情感分类-评测数据集

文本分类模型	.≘	4	公开数据集											
凸 模型中心														
我的模型			chnsenticorp-情景	hmsentlcorp-情態分类-训练致感集										
创建模型			版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作				
训练模型														
校验模型			V1 🖯	20000006	8249	 已完成 	文本分类	100% (8249/8249)		立者				
发布模型		Г												
☑ EasyData数据服务			chnsenticorp-情感分类-开胞数据集											
数据总览		Н	版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作				
公开数据集		Т	VI O	20000001	1178	●已完成	文本分平	100% (1178/1178)		1 16				
标签组管理		L		2000001		- C/UW	~140	issue (in dirito)						
在线标注			emotion											
智能标注			151 木	数評価の	政府開	最近最大成本	板江市町	統注於本	请评计术	28-69F				
器 模型部署			MAT.	ALM HOLD	NA MELER	MEAL OF COMP	HIVE ALL	2010-2010	111/0-1-00	APT C				
纯离线服务			VI 😔	20000000	2710	• 已完成	文本分类	100% (2710/2710)		宣看				

在模型列表页点击【模型训练】进入到数据集选择界面

文本分类模型	÷	编数列表		提交工单
曲 模型中心				
我的模型		创建模型		
创建模型		「中中公司」 調査(調査、 1/ (通知)の: 1999年5	C3 2046	17 Dillo
训练模型		LX+7741.486/F72 (2) WEILV: 12103	C #163	TAKED ES
校验模型		模型剑建成功,您可以选择使用分开数据集训练模型体验分类效果,应剑建数据集使用白有数据进行发展分类模型的训练,模型训练员,可以在在公告看模型的最新版本和成态		
发布模型				

可诜标签:			-	
5 AL 19-14		分类名称		
	~	1		
		0		

点击【开始训练】进入到模型训练阶段

文本分类模型	۰	训练模型					握交工单
 ① 模型中心 我的模型 ①建模型 ② 建模型 及物模型 及物模型 正 EasyData政務総务 政策必义 		选择模型 训练配置 部署方式 选择算法 模型等选指标 ⑦ 添加数据	 (私口行分 ∨) (公牧云部筆) EasyEdge本地部署 (● 高林田 ①) (● 高林田 ②) (● 高林田 ③) 	如何法律部署方式7			
公开数据集 标签组管理		添加训练数据	+ 请选择 你已经选择1个数据集的;	2个分类 全部清空			
在线标注 智能标注 [2] 模型部署			数据集 chrisenticarp-情感分类-评测数据集	版本 V1	分类数量 2	操作 查看评值 清空分类	
纯漉线服务		训练环境	査様 ● GPU P40 ○ GPU P40	规格 TeslaGPU_P40_24G显存单卡_12版CPU_40G	約存	算力 12 TeraFLOPS	价格
		开始训练		1858070_7100_166至存車大_12核CPU_58G	29 14	14 reraFLOPS	光詞

在模型列表下,可以看到处于训练状态的模型,将鼠标放置感叹号图标处,即可查看训练进度,同时若**勾选短信提醒,在模型训练完成后会以短** 信的形式通知

莫型列表								提交工单
创建模型								
【文本分类】酒店评分 🖸 模型	D: 121263						⑤ 历史版本	直 删除
部署方式	版本	训练状态	服务状态	模型效果	操作			
公有云API	V1	训练中 🔳	未发布	-	查看版本配置 停止训练			
【文本分类】erniem_m_zxl 🖸	训练进度 ✓ 完成症 模型ID:	: 1% 5短信提醒至 182*****572	ß			吕 训练	⑤ 历史版本	回 删除
型 校验 型训练完成后,	,可在模型	<u>⊍</u> 列表下,点	、击【 校验】					
莫型列表								提交工单
【文本分类】酒店评分1 🖸 模型	불ID: 76016					吕 训练	◎ 历史版本	回 删除
部署方式	版本	训练状态	服务状态	模型效果	操作			
公有云API	V4 (2)	训练完成	未发布	准确率: 89.15% F1-score : 0.891 完整评估结果	查看版本配置 申请发布 校验			
	総理機構 (文本分角) 道路符分 2 (規算) 形用方式 公有分点) 道路符分 2 (規算) が不分角) 道路符分 2 (規算) (文本分角) 道路行分 2 (規算) (文本分角) 道路行介」 2 (規 (文本分角) 道路行分1 2 (規 影用方式 公有云AP)	تلاقلان المحالية ال محالية المحالية المحالي محالية المحالية المحالي محالية المحالية ا	型型列表	مال المعرفة المعالية ال المعالية المعالية	مال المعرفة المعرفة المحافة المحاف محافة المحافة الم	عنه المعنية المحمة ا محمة المحمة المح	معتبر المعنية المعن المعنية المعنية ا	موتعام العربية المعربة المحكمة

点击【**启动模型校验服务**】,需等待几分钟

校验模型			
选择模型 酒店评分1 > 部署方式 公有云API(目前仅支 > 启动模型校验服务	选择版本	V4	~
输入校验文本,进行模型校验			提交工单
送拝模型 酒店祥分 ▼ 都要方式 公有云API (目前仅支_ ▼ 送拝版本 V1 ▼ 当前模型准确事 94.63% 祥冶服告	识别结果 如何优化效果 调整阈值 ◆	0.03	
第個人性能的文化,現人口工作文 又行文本物元・00、文本式職工程力502次子(子行) 这个酒店的序項很一般,服务态度页不太好	预测分类 0 1	置信度 > 3.00% ∨ 94.68% 5.32%	
10/552	a 由海 F 4a		

在此处可以点击【申请上线】,进行模型发布,跳转到模型发布

模型发布

模型训练完成后,点击【申请发布】

【文本分类】 酒店讲分1 🖸 機動D: 76016							① 历史版本	由 删除
部署方式	版本	训练状态	服务状态	模型效果	操作			
公有云API	V4 (?)	训练完成	未发布	准确率: 89.15% F1-score: 0.891 完整评估结果	實看版本配置 申请发布 校验			

按要求填写相应信息后,点击**【提交申请】**

发布模型

选择模型	酒店评分1	\sim
部署方式	公有云部署	\checkmark
选择版本	V4	\checkmark
服务名称 *	识别酒店评论好与坏	
接口地址 *	https://aip.baidubce.con	n/rpc/2.0/ai_custom/v1
	/text_cls/ jacetest3	
其他要求	若接口无法满足您的需 的问题,500汉字以内	求,请描述希望解决 0/500
	提交申请	

【文本分类】酒店评分1	☑ 模型ID: 76016					吕 训练	⑤ 历史版本	回 删除
部署方式	版本	训练状态	服务状态	模型效果	操作			
公有云API	V4 📎	训练完成	发布中	准确率: 89.15% F1-score : 0.891 完整评估结果	宣看版本配置 校验			
等待几分钟,	此状态就会	会变为【 已发	发布】 ,即发	:布成功				
模型列表								提交工单
创建模型								

【文本分类】酒店评分 🖸 模	型ID: 121263					吕 训练	⑤ 历史版本	前删除
部署方式	版本	训练状态	服务状态	模型效果	操作			
公有云API	VI	训练完成	已发布	准确率: 94.63% F1-score: 0.946 完整评估结果	宣看版本配置 服务详情 校验			

രം Step4: 模型调用

在【我的模型】栏找到发布完成的模型点击【服务详情】

文本分类模型	·≡	模型列表						提交工单
① 模型中心 我的模型		创建模型						
创建模型		【文本分类】酒店评分	☑ 模型ID: 121263				吕 训练 · ⓒ 历史版本 [亩 删除
校验模型		部署方式	版本	训练状态	服务状态	模型效果	操作	
发布模型 EasyData数据服务		公有云API	V1	训练完成	已发布	准确率: 94.63% F1-score: 0.946 完整评估结果	查看版本配置 服务详情 校验	2

接口地址在模型调用代码中将会用到,点击【**立即使用】**跳转至EasyDL控制台

	服务详情	>
服务名称:	jiudianpingfen_v1	
模型版本:	V1	
接口地址:	https://aip.baidubce.com/rpc/2.0/ai_custom/v1/text_cls/jiudianpingfen _v1	
服务状态:	已发布	
	立即使用 查看API文档	

信息填写完成后点击【立即创建】

应用列表

+ 创建应用

EasyDL文本

公有云部署

^

* 应用名称:	酒店评论识别
* 接口选择:	<mark>部分接口免费额度还未领取,请先去领取再创建应用,确保应用可以正常调用 去领取</mark> 勾选以下接口,使此应用可以请求已勾选的接口服务,注意EasyDL文本服务已默认勾选并不可取 消。
	您已开通付费的接口为:短语音识别-中文普通话、短语音识别-粤语、短语音识别极速版、实时 语音识别-中文普通话、实时语音识别-英文、通用文字识别(高精度版)、通用文字识别(高精
	度含位置版)、网络图片文字识别、身份证识别、银行卡识别、驾驶证识别、数字识别、手写文
	字识别、火车票识别、iOCR通用版、H5视频活体检测、驾驶行为分析、人脸融合、中文词向量
	表示、词义相似度、文章分类、内容审核平台-图像、内容审核平台-文本、知识问答、通用物体
	和场景识别高级版、相同图检索-入库、相似图搜索-入库、手部关键点识别、货架拼接-货架拼
	接、黑白图像上色、人脸实名认证、人脸实名认证
	+ EasyDL
	★ 文字识别
	➡ 人脸识别
	+ 自然语言处理
	+ 内容审核 !
	+ UNIT !
	➔ 知识图谱
	➡ 图像识别 Ⅰ
	➡ 智能呼叫中心
	 ▶ 图像搜索
	+ 人体分析
	➡ 图像增强与特效
	➡ 智能创作平台
	+ EasyMonitor
	+ BML
	 1 机器翻译
* 应用描述:	识别评论内容是好评还是差评
则建完成后即刻在区	立用列表页获取到AK SK密钥
如用列表	

+ 8	維应用					
	应用名称	AppID	API Key	Secret Key	创建时间	操作
1	酒店评论识别	24313338	LKDCow2cYfBzPYcuaB43pell	显示	2021-06-04 20:09:18	报表 管理 删除

在获取到API KEY 以及 Secret KEY后,我们就可以写一个示例代码调用我们之前创建并训练完成酒店评论文本分类模型

准备开发环境

我们选择用python来快速搭建一个原型,如果没有接下来需要安装以下python。可以参考下表列出的不同操作系统的安装方法进行安装。

Python的官方下载地址:下载python

系统是Windows	系统是Linux	系统是macOS
✓ 在Python官网下载Python✓ 解压并双击exe文件安装	 执行`pythonversion`看是否输出python版本; 若回显版本则系统已预装Python; 若无版本则执行`sudo apt-get update & sudo apt-get install python3.6` 	✓ macOS一般自带了Python , 无需自行安装

Windows 快速测试包

windows平台的用户如果对上述的python安装感到困难,可以下载我们的一键测试包,下载地址:windows测试包。

解压zip文件后,双击run.bat即可测试。

编写代码

新建一个 main.py

粘贴以下内容,不要忘记替换你的 API_KEY 以及 SECRET_KEY:

coding=utf-8

import sys import json

保证兼容python2以及python3 IS_PY3 = sys.version_info.major == 3 if IS_PY3:

from urllib.request import urlopen from urllib.request import Request from urllib.error import URLError from urllib.parse import urlencode from urllib.parse import quote_plus

else:

import urllib2 from urllib import quote_plus from urllib2 import urlopen from urllib2 import Request from urllib2 import URLError from urllib import urlencode reload(sys) sys.setdefaultencoding('utf8')

防止https证书校验不正确

import ssl

 ${\tt ssl_create_default_https_context} = {\tt ssl_create_unverified_context}$

百度云控制台获取到ak,sk以及 #### EasyDL官网获取到URL

ak API_KEY = 'RgdpDFjOHmRQvphsi8bLhIYE'

sk

SECRET_KEY = 'ja1pDyGaF3vgwPNW3T0EqEkkd5hgl8ug'

url

EASYDL_TEXT_CLASSIFY_URL = "https://aip.baidubce.com/rpc/2.0/ai_custom/v1/text_cls/hotel_comment"

""" TOKEN start """ TOKEN_URL = 'https://aip.baidubce.com/oauth/2.0/token'

....

获取token
"""
def fetch_token():
 params = {'grant_type': 'client_credentials',
 'client_id': API_KEY,
 'client_secret': SECRET_KEY}

post_data = urlencode(params)
if (IS_PY3):
 post_data = post_data.encode('utf-8')
req = Request(TOKEN_URL, post_data)
try:

f = urlopen(req, timeout=5) result_str = f.read() except URLError as err: print(err) if (IS_PY3):

result_str = result_str.decode()

```
result = json.loads(result_str)
```

```
if ('access_token' in result.keys() and 'scope' in result.keys()):
     if not 'brain_all_scope' in result['scope'].split(' '):
       print ('please ensure has check the ability')
       exit()
     return result['access_token']
  else:
     print ('please overwrite the correct API_KEY and SECRET_KEY')
     exit()
....
  调用远程服务
....
def request(url, data):
  if IS_PY3:
     req = Request(url, json.dumps(data).encode('utf-8'))
  else:
     req = Request(url, json.dumps(data))
  has_error = False
  try:
    f = urlopen(req)
     result_str = f.read()
     if (IS_PY3):
       result_str = result_str.decode()
     return result_str
  except URLError as err:
     print(err)
if __name__ == '__main__':
  # 获取access token
  token = fetch_token()
  # 拼接url
  url = EASYDL_TEXT_CLASSIFY_URL + "?access_token=" + token
  # 好评
  text_good = "这个酒店不错,干净而且安静,早餐也好吃"
  # 差评
  text_bad = "不怎么干净,服务员态度也差强人意,以后不会在预订了"
  # 请求接口
  # 测试好评
  response = request(url,
     {
       'text': text_good,
       'top_num': 2
    })
  result_json = json.loads(response)
  result = result_json["results"]
  # 打印好评结果
  print(text_good)
  for obj in result:
     print(" 评论类别: " + obj['name'] + " 置信度: " + str(obj['score']))
  print("")
  # 请求接口
  # 测试差评
  response = request(url,
    {
       'text': text bad.
```

```
})
result_json = json.loads(response)
result = result_json["results"]
# 打印差评结果
print(text_bad)
for obj in result:
    print(" 评论类别: " + obj['name'] + " 置信度: " + str(obj['score']))
```

运行代码

在命令行中运行python main.py

'top_num': 2

您还可以在我们的github地址中找到main.py

结果

若代码正确运行,命令行界面上会显示出运行结果:

```
这个酒店不错,干净而且安静,早餐也好吃
评论类别:good 置信度:0.974235713482
评论类别:bad 置信度:0.0257642995566
不怎么干净,服务员态度也差强人意,以后不会在预订了
评论类别:bad 置信度:0.850781261921
评论类别:good 置信度:0.149218738079
```

结果中返回了每个待分类文本的分类以及置信度,置信度高的分类说明预测的文本属于这个分类的可能越大,这样我们就能将上述酒店评论分为 好评,差评了,详细的返回和参数文档需要参照API文档EasyDL文本分类API参考文档

心 产品特色

可视化操作: 无需机器学习专业知识,模型创建·数据上传·模型训练·模型发布全流程可视化便捷操作,最快15分钟即可获得一个高精度模型

高精度效果:EasyDL底层结合百度 AutoDL/AutoML技术,针对用户数据自动获得最优网络和超参组合,基于少量数据就能获得出色效果和性能的模型

端云结合:训练完成的模型可发布为云端API或离线SDK,灵活适配各种使用场景及运行环境

数据支持:全方位支持训练数据的高质量采集与高效标注,支持在模型迭代过程中不断扩充数据,助力提升模型效果

心 更多参考

如对文档说明有疑问或建议,请**微信搜索"BaiduEasyDL"**添加小助手交流 备注:文档如使用中遇到报错等问题,请在控制台中通过"工单"联系我们,售后团队为您及时解决问题

EasyDL官网入口 EasyDL开发文档 EasyDL软硬一体方案 EasyDL应用案例

EasyDL零售行业版快速开始

心简介

本文档介绍使用EasyDL零售版商品检测快速训练一个识别可口可乐的商品检测模型,基本流程如下:

1.创建模型

2.创建SKU

3.上传和标注训练数据

4.训练模型

5.发布模型

の 步骤1.创建模型

这个步骤将会介绍如何创建模型

心 进入创建模型页面

在EasyDL零售版商品检测产品主页点击【开始训练】按钮进入到模型训练页,下面会出现两种情况:

- 第一种,如果您没有登录百度云,则会跳转到百度云登录页面,没有百度账户的客户请先注册百度账户。登录后,会跳转到模型概览页,点击 【商品检测】卡片上的【立即定制】按钮,会跳转模型训练页面的创建模型页。
- 第二种,如果您已登录,会直接进入到【我的模型】页,该页面能够管理已经创建的模型,点击左侧列表中的【创建模型】进入创建模型页面。

心 创建模型

进入创建模型页面后你会看到如下图中展示的内容

模型中心	模型列表 > 创建模型
我的模型	模型类别: 商品检测
创建模型	
训练模型	· 機型名称:
校验模型	模型归属: 公司 个人
发布模型	请输入公司名称
数据中心	• 应用场景: 请选择应用场景 🗸
我的SKU库	不同应用场景对应不同训练算法,请根据真实应用场景选择
实景图上传/标注	• 邮箱地址:
AI服务中心	• 联系方式: [2]
服务列表	• 功能描述:
	0/500
	モーチ

需要填写的项目如下:

● 模型名称

模型的名称

- 模型归属
 模型是属于公司的,还是属于个人的,如果是前者,请填写公司名称
- 应用场景

提示:**请根据真实业务应用场景选择**,选择的场景将会关联后端数据增强算法,若不确定,请选择"其他"

可选项为普通货架/货柜、智能结算台、无人零售柜、地堆商品和其他

邮箱地址

用于联系到您的邮箱地址

• 联系方式

有效的联系方式将有助于后续模型上线的人工快速审核,以及更快的百度官方支持,推荐填写个人手机号码

功能描述

描述改模型将要应到的业务场景,详细的描述,在获取官方支持时,能帮助我们为您提供准确的使用建议

像下图展示的一样完成所有填写项后点击【下一步】按钮完成模型创建,创建完成后会跳转到【我的模型】页面。

模型列表 > 创建模型	
模型类别: 商品恰测	
- 雄刑之役・ 可口可口役別	
模型归属: 公司 个人	
百度	
· 应用场景: 普通货架/货柜 🗸	
不同应用场景对应不同训练算法,请根据真实应用场景选择	
●邮箱地址: ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●	
•联系方式: 高 mulai ii n 。 [2]	
。功能描述: 用于快速创建一个识别普通货架上可口可乐的AI模型	
	24/500
下一步	
	 棋型列表 > 幼球成型 展型 法: 高品法 ・ 供型 法: 可□可□ 功況 ・ 供型 法: 「□□□ □ 功況 ・ 使型 小 ● 広田 法: ● 一 小 ● の田 法: ● 一 一 でしましたので ● い 和 い 一 でしましたののの ● い 本 ・ = ● でしましたのの ● い 本 ・ = ● でし 本 ・ = ● でしましたの ● い 本 ・ = ● でしましたの ● い 本 ・ = ● でし 本 ・ = ● でしましたの ● い 本 ・ =

の步骤2.创建SKU

这个步骤将会介绍如何创建SKU,SKU是客户需要检测的商品,在训练品台上有两个作用,其一是"SKU名称_品牌名称_规格参数"用于标注训练数据的标签,二是SKU的单品图片用于商品增强合成技术,提高模型效果。

完成上一个步骤后,会跳转到【我的模型】页面,这时您会看到如下图展示的内容,由于模型还未训练,所以模型列表中没有显示模型的效果, 在训练模型前,需要先完成SKU的创建。

模型中心	模型列表
我的模型	创建模型
创建模型	
训练模型	【商品检测】可口可口识别 模型D: 27257
校验模型	
发布模型	模型创建成功,若无SKU请先在"数据中心"创建SKU,上传实展图训练模型后,可以在此处宣奏模型的最新版本
数据中心	
我的SKU库	
实景图上传/标注	
AI服务中心	
服务列表	
	毎页显示 4 🗸 < 1 >

点击左侧列表中的【我的SKU】进入SKU管理页面,点击【创建SKU】按钮进入创建SKU页面,您会看到如下图展示的内容

模型中心	我的SKU库 > 创	建SKU	
我的模型	创建的SKU将以	"SKU名称_品牌名称_规格参数"的形式作为识别时返回的	结果,命名时请避免重复
创建模型	, 东口夕称,		
训练模型	* 岡田右柳・	如·原咪可赤、正径斤、海蚌土抽等	
校验模型	* 品牌名称:	如:可口可乐、乐事、海天等	
发布模型	* 规格参数:	如:330ml,12粒装,120g等	
数据中心	* 商品品类:	请选择商品品类	•
我的SKU库	* 包装类型:	请选择包装类型	
实景图上传/标注			
AI服务中心	商品编号:	用于匹配商品系统中的编号,若无可不填	
服务列表	SKU单品图:	上传图片 用于有效提高模型效果 示例图片	
		创建SKU 返回	

提示:在调用API接口识别SKU时,识别结果中SKU的名字是以"SKU名称_品牌名称_规格参数"的形式返回的,所以在填写SKU名称、品牌名称 和规格参数时避免这三项内容重复。

需要填写的项目如下:

- SKU名称
 SKU的名称,可适当填入SKU细节,例如:原味可乐,番茄味薯片,奥运版纯牛奶等
- 品牌名称

SKU的品牌名称,如可口可乐,乐事,伊利等

• 规格参数

SKU的规格,如330ml,500g,20片等

• 商品品类

可选择的有饮品、药品、保健品、零食、香烟、调味品、日用品和其他

● 包装类型

可选择的有瓶装、罐装、袋装、盒装和其他

• 商品编号

如果您自身的业务系统中有现成SKU对应的商品编码,比如商品条形码,可以填在该填写框中,之后模型接口将支持返回该内容,用于您快速 匹配SKU

• SKU单品图

SKU的单品图将用于商品增强合成,拍摄角度和上传张数基本原则是覆盖实际检测场景可能出现的角度,具体请参考SKU单品图数据文档中进行单品图采集。如果不上传,将会降低模型的识别效果,可以点击页面上的【示例图片】查看SKU单品图样张。

完成填写和上传SKU单品图上传后,页面内容显示如下图所示



点击【创建SKU】按钮完成创建,点击后回到【我的SKU库】,SKU列表中的SKU图数需要大约5秒的时间进行计算,刷新页面即可显示SKU单品 图片数。

心步骤3.上传和标注训练数据

这个步骤将会介绍如何上传和标注训练数据,训练数据是SKU在货架上的实景图,需要客户从真实的业务场景中采集,这些图片在被正确标注中,可以用于训练成模型。

完成上一个步骤后,在左侧列表中点击【实景图上传/标注】进入上传和标注页面,在上传前请在实景图集选择栏内创建实景图集,如下图所示

商品枚	新建实景图集			×		
定制商品检 查,如铺货	实景图集名称:	可口可乐货架图集]	商品计数,	辅助
	选择类型:	普通货架/货柜	~			
实景图上传/标注						
* 实景图集:		创建实景图集	取消			

需要填写的项目如下:

• 实景图集名称

实景图集的名称,可适当填入SKU细节,例如:原味可乐,番茄味薯片,奥运版纯牛奶等

• 选择类型

实景图集的类型,请与创建模型时选择的应用场景保持一致,上传时只上传跟选择类型相同的实景图。可选项为普通货架/货柜、智能结算 台、无人零售柜、地堆商品和其他

完成创建实景图集后,页面显示为如下图所示的内容

模型中心	实景图上传/标注	
我的模型	。 实景图集: 可□可乐货架图集 ✓	
创建模型	图片筛选: 全部图片 未标注 已标注	
训练模型		
校验模型		+ >
发布模型	标注示例	SKU标签列表 Q
数据中心		+ 新建SKU
我的SKU库		原味可乐_可口可乐_500ml 下 🕞
实景图上传/标注	1. 同一张图中需要标注全部想要识别的目标区域; 2.标注图片需要和实际场景一致;	锁定后,后续框选目标会默认以该标
服务列表	以"检测一张图中所有的可乐"为例 以实际场景为识别出"用手机在零售货柜拍摄的图片"中 有几层货架为例	签命名
	全部恒出 ※ ● <th></th>	

点击页面上【标注】为该实景图集上传作为训练数据的实景图,点击【标注示例】右侧的加号上传实景图。

实景图基本要求如下:

实景图的详细采集要求,请参考实景图数据要求文档

- 实景图片需要是从真实业务场景中采集来的数据
- 支持上传的图片格式为jpg,png,jpeg,bmp,大小限制为4M
- 建议图片尺寸:最长不超过4096px,最小不低于30px,长宽比3:1以内

标注基本要求如下:

- 完整并仅仅框选要识别的SKU
- 标注框不要框选到其它SKU或是价目标签等非要识别的SKU的干扰信息
- 在实景图中出现的所有要识别的SKU必须全部标注,不能遗漏

完成所有实景图的标注后,返回到【我的SKU库】可以查看到SKU列表中【实景图数】列显示标注了该SKU的实景图片的数量,如下图所示

模型中心	我的SKU库						
我的模型	创建SKU					请输	入SKU名称、品牌、規格进行搜索 Q
创建模型							
训练模型	SKU名称	品牌名称	规格参数	SKU图数	实景图数	状态	操作
校验模型	原味可乐	可口可乐	500ml	12	12	正常	查看 编辑 删除
发布模型							
数据中心							
我的SKU库							
实景图上传/标注							
AI服务中心							
服务列表	您已创建了1个SKU					每页显	标 10 🗸 < 1 >

の 步骤4.训练模型

这个步骤将会介绍如何训练模型

模型中心	模型训练							
我的模型	选择模型:	可口可口说别		~				
创建模型		311 311 911 911 9						
训练模型	添加SKU:	请添加要参与训练	东的SKU	~				
校验模型		序号	SKU名称	品牌名称	规格参数	操作	实景图集	SKU数量
发布模型		1	原味可乐	可口可乐	500ml	删除	可口可乐货架图集	1
数据中心								
我的SKU库								
实景图上传/标注								
AI服务中心								
服务列表								
		开始训练						

如上面图片所示,点击左侧列表中的【训练模型】,需要先后完成下面三项选择:

1. 选择要训练的模型

2. 选择需要想要模型支持检测的SKU,选择完成后,下方左侧会显示已添加的SKU,右侧会显示包含已添加SKU的实景图集

3. 选择要参与训练的实景图集

完成选择后,点击【开始训练】按钮页面跳转至【我的模型】页面,如下图所示,可以看到模型已进入训练状态,将鼠标移至状态"训练中"右边 的小问号上,可以查看训练进度,训练进度数值只是作为参考,所以推荐打开短信通知功能,这样就第一时间知晓模型训练完成了。

模型中心	模型列表							
我的模型	创建模型							
创建模型								
训练模型	【商品检测】7	可口可口识别	模型ID: 27257					
校验模型	应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作	
发布模型	云服务	V1	训练中 💷	未申请	未发布	-	停止训练	
数据中心		训练进	主度:	11%				
我的SKU库		✔ 完	成后短信提醒至 186	5***650 🗹				
实景图上传/标注								
AI服务中心								
服务列表								
							每页显示 4 🗸 <	1 >

训练完成后,可以点击校验和申请发布。

模型中心	模型列表								
我的模型	创建模型								
创建模型									
训练模型	【商品检测】同	「ロ可ロ识别	模型ID: 27257						Î
校验模型	应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作		
发布模型						mAP 95.83% ?			
数据中心	云服务	V1	训练完成	未申请	未发布	福祉 93.30% 団 召回率 92.31% ? 完整评估结果	申请发布 校验 训练		
我的SKU库									
实景图上传/标注									
AI服务中心									
服务列表									
							毎页显示 4 🗸 <	1	>

心 步骤5.发布模型

这个步骤将会介绍如何将训练好的模型发布为服务API

模型中心	发布模型	
我的模型	1 选择模型: 饮品检测 V 标准接口规范参考	
创建模型	2 选择版本: V1 ✓	说明:
训练模型	2 四々なわ。 」 」 」 」 」 」 」 」 」	必须 类型 说明
校验模型	3 * 顺方在标: ketexuenjaance 4 • 接口地社: https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/ image 長	 留像数据,base64编码,要求 base64编码,要求 base64编码后大小不超过4M, 最短边至少15px,最长边最大
发布模型	detection/ kelexuebixianwei	4096px,支持jpg/png/bmp格式
数据中心	5 其他要求: 可乐和雪器纤维检测 threshold 音	阈值,默认为当前模型推荐阈值 否 number (0-1之间),具体值可以在我的 模型列表-模型效果查看
我的SKU库	标准接口响应字段	说明:
实景图上传/标注	9/500	次衛 米利 治阳
AI服务中心	提交申请 2000年1月1日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日日	
服务利率	如果有私有化部署需求,请点此申请	室 number 唯一的ogia,用于问题定证
DX 25 2 9 4X	error_code 2	否 number 错误码,当请求错误时返回

在模型训练好后,点击模型列表内对应模型「操作」列中的「申请发布」,或是在左侧导航栏点击「发布模型」可以进入发布模型页面,如上图 所示。在对应选项中选择和输入相应内容发起模型发布的申请:

1. 选择模型 (必选)

选择需要发布的模型,只能选择已经完成训练的模型

2. 选择版本 (必选)

选择需要发布的模型版本,只能选择完成训练且没有发布过的版本

3. 服务名称 (必填)

为发布的服务命名,**服务名称不得多于20个字符**

4. 接口地址 (必填)

自定义服务的API URL,接口地址需要多于5个字符但不能超过20个字符,仅限英文

5. 其他要求

如果有其他要求可以输入要求描述

填写完上述信息后,点击「提交申请」完成发布模型申请。提交申请后,模型列表内该模型的申请状态和服务状态为有以下几种情况:

申请状态	服务状态	状态描述
审核中	未发布	服务刚申请发布,模型在审核中
审核成功	发布中	服务通过审核,进入系统自动发布阶段
审核成功	已发布	服务发布成功
审核失败	未发布	服务未通过审核,通常为模型训练结果mAP < 0.6,如需申诉,可以加入官方QQ群(群号:1009661589)咨询群 管

提示:第一次申请发布的模型需要人工审核,通常4小时内完成,如果希望加急上线,请加入官方QQ群(群号:1009661589)咨询群管高优 审核。非第一次申请发布的模型,如果模型训练结果mAP>0.6,则会自动通过审批。审批完成后,大约需要5分钟左右自动完成发布。

发布成功后,可以点击模型列表内「操作」列中的「服务详情」获取服务API URL,点击后弹出下图所示窗口:

服务详情		×
服务名称:	kelexuebijiance	
模型版本:	V1	
接口地址:	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/detection/kelexuebixianwei	
服务状态:	已发布	
	立即使用 查看API文档	

点击「查看API文档」可以快速跳转至API文档,参考文档调用API获取商品检测AI能力。

EasyDL图像SDK集成快速开始

通用设备端Android ARM

心简介

本文档包括两个部分,分别适用于SDK测试、SDK集成。开发者可根据实际需求选择参考:

- SDK测试文档:供测试Demo时参考
- SDK集成文档:供将SDK集成进自己的代码时参考

测试前的准备

- Android系统的硬件及开发环境
 - 详情参考下方文档
- EasyDL平台的Android SDK
 - 以图像分类为例,前往操作台训练模型后,选择发布为Android系统的通用设备端SDK,发布成功后即可从平台下载
- 用于激活设备端SDK的序列号
 - 前往控制台申请用于激活通用设备端SDK的序列号

效果展示



测试前的准备

- 硬件:
- 1. 准备一台PC机
- 2. 准备一台较新款的Android 手机 不支持模拟器
- SDK
- 1. 您已经生成Android SDK 并且已经下载成功
- 2. 您已经通过扫描二维码的形式,在这台Android 手机上测试成功想要的功能。
- 生成序列号
- 1. 如果是开源模型版本,不需要序列号,
- 2. 如果测试的是"按单台设备激活",需要额外获取一个序列号
- 3. 如果需要额外测试"产品线激活"的序列号,需要再准备一个包名。demo的包名com.baidu.ai.easyaimobile.demo。
- 开发环境
- 1. 因为Android是google公司的项目,测试过程中可能需要访问国外网站下载资源。
- 2. PC机上安装较新版本的Android Studio,本文使用的是4.0.1版本。下载地址
- 3. 在这台Android手机上测试通过一个Android HelloWorld项目

Android Studio的安装及手机测试

本段简单地描述如何通过Android Studio在手机上运行一个自带的Demo,有android基础的可以跳过本段。

手机上需要打开开发者模式,您也可以下载91助手,按照软件提示连接手机。

更多Android Studio的安装测试也可以百度下【Android Studio自动生成Demo】。新建Android Studio自带的测试项目, 菜单File->New Project.., 弹框中选择"Basic Activity",点"Next",之后用默认配置,点"Finish"后项目就生成了。



导入成功后,有以下的图标:	*	<u>F</u> ile	<u>E</u> dit	<u>V</u> iew	<u>N</u> aviga	te <u>C</u> ode	Analy <u>z</u> e	<u>R</u> efactor	<u>B</u> uild	R <u>u</u> n	<u>T</u> ools	VC <u>S</u>	<u>W</u> indo
			G	← →	🏠 (🞽 арр	▼ □, X	aomi MI 8	•	¢	e, ð	E Up	0 🕉
	1	МуА	pplica	tion \rangle	app		千切口	▼ 终连接D	ì	「「			
	ē	<u>کم</u> ک	ndroi	d 🔻	-1-4 046-		G-M		🗿 Mai	nActiv	ity.java	× ni	build.c

导入官方Demo

导入项目: 菜单File->New-> Import Project..,选择PaddleOCR\deploy\android_demo目录。 注意千万不要使用菜单File->New-> New Project..



导入项目后,会触发gradle的自动同步,最终效果如下:



此时项目可以正常编译,

- 如果是开源模型版本,此时可以正常运行。
- 如果需要序列号的情况,此时会界面会报错"序列号错误"

填入序列号 如果是开源模型版本,不需要序列号,序列号保持为null即可 在MainActivity开头部分填入您的序列号



此时项目可以正常编译及运行。

-如果"产品线激活"的序列号,还需要额外修改包名

修改包名 (仅"产品线激活"需要)

如果您填入的包名是"com.baidu.ai.easyaimobile.demo" 如图修改:



此时项目可以正常编译及运行。

精简版测试

仅限通用arm的 图像分类,物体检测,文字识别。其它引擎可以参考自行写。 如果是开源模型版本,不需要序列号,序列号保持为null即可。

使用MiniActivity可以在如下情况下测试:

- 不带摄像头或官方demo运行摄像头报错的开发板
- 避免摄像头预览占用CPU导致耗时测试不准确

具体步骤如下:

A. 在infertest.MainActivity中,修改文件开始位置的序列号

Baidu 百度智能云文档



B. 修改启动Activity为infertest.MainActivity,修改AndroidManifest.xml文件。主要不要漏掉开头的"."



此时启动在logcat中会发现缺少test.jpg

C. 将你的测试图片test.jpg 放入assets目录。

EasyDL-20521-0.9.7 D:\AndroidProjects\Easy

🕨 🖿 .gradle 🕨 🖿 .idea 🔻 🐚 app 🕨 🖿 build libs 🔻 🖿 src androidTest 🔻 🖿 main assets 🕨 🖿 demo infer-classify ₽ 불 test.jpg 🕨 📄 java res 릚 AndroidManifest.xml test

此时再次运行,	点击界面上的按钮,	有如下测试成功的界
---------	-----------	-----------

15:09 編香・・	
1:mao:0.9901489	
BUTTON	

の Android SDK集成文档

这个部分以Android Studio 自带的Empty Activity 模板项目为例,展示如何集成OCR Android的代码到您自己的项目中

集成前的准备

- 1. 需要一个较新款的Android手机
- 2. 请先根据上方的测试文档配置环境及测试官方Demo
- 3. 请先根据上方的测试文档测试MiniActivity,本文以MiniActivity为模板集成

集成后的代码下载

如果觉得下面步骤有模糊的地方,可以参照修改好的代码进行下载。

链接:https://pan.baidu.com/s/1cTFxYrzb1jp8bWBs6eoF8A 提取码:u7xv

zip 文件名	说明
myedge-init.zip	初始化的"Empty Activity"模板项目
myedge-finished.zip	做完本文所有步骤后的项目

新建一个项目

新建Android Studio自带的测试项目, 菜单File->New Project.. ,弹框中最后一个项目模板 "Empty Activity",点"Next",之后用默认配置, 点"Finish"后项目就生成了。这里比如给给这个项目起名为myedge

🛎 Create New Project						×
Select a Project Ter	nplate					
Phone and Tablet Wear OS TV Auto	motive Android Th	iings				
*	•	÷		¢		
No Activity Ba	sic Activity	Bottom Naviga	tion Activity	Em	pty Activity	
e []	Ad	÷	E	~		
Empty Activity						
Creates a new empty activity.						
			Previous	<u>N</u> ext	<u>C</u> ancel	Finish
M Courte New Partient			Previous	<u>N</u> ext	<u>C</u> ancel	Finish
Create New Project	ect <u>Name</u>		Previous	Next	Cancel	Finish
Configure Your Proj	ect <u>N</u> ame myedge		Previous	Next	Cancel	Finish
Create New Project	ect Name myedge Package name com.example.	myedge	Previous	Next	<u>Cancel</u>	Finish × 其它情况
Create New Project	ect Name myedge Package name com.example. Save location	myedge	Previous	Next	Cancel	Finish × 其它情况
Create New Project	ect Name myedge Package name com.example. Save location D:\AndroidProc	myedge ijects\myedge	Previous	<u>Next</u>	Cancel	Finish × 其它情况
Create New Project	ect Name myedge Package name com.example. Save location D:\AndroidPro Language	myedge ijects\myedge	Previous	Next	Cancel 申请的包名,	Finish × 其它情况
Configure Your Proj	ect Name myedge Package name com.example. Save location D:\AndroidPro Language Java	myedge ijects\myedge	Previous	Next	Cancel (中请的包名, 》	Finish × 其它情况
Create New Project	ect Name myedge Package name com.example. Save location D:\AndroidPro Language Java Minimum SDK	myedge ijects\myedge	Previous	Next	Cancel 申请的包名,	Finish × 其它情况
Configure Your Proj	Name myedge Package name com.example. Save location D:\AndroidPro Language Java Minimum SDK Your app withen mech	myedge jects\myedge API 21: Android 5	Previous	<u>Next</u> 改这里为你	Cancel 申请的包名, 下	Finish × 其它情况
Create New Project Configure Your Proj Configure Your Proj Empty Activity Creates a new empty activity.	Name myedge Package name com.example. Save location D:\AndroidPro Language Java Minimum SDK 9 Your app W Help me cho	myedge jects\myedge API 21: Android 5 ir ran on approxima	Previous	Next 改这里为你	Cancel (Cancel) (中请的包名,) (中请的包名,) (中请的包名,)	Finish × 其它情况
Create New Project Configure Your Proj Configure Your Proj Empty Activity Creates a new empty activity. The application name for most apps begins w	Iect Name Myedge Package name Com.example. Save location D:\AndroidPro Language Java Minimum SDK Your app W Help me cho Use legacy ith an uppercase letter	myedge jects\myedge API 21: Android S thrun on approxima pose android.support lib	Previous	Next 改这里为你 vices.	Cancel (Cancel) (中请的包名,) (中请的包名,) (中请的包名,) (中请的包名,) (中请的包名,)	Finish × 其它情况
Create New Project Configure Your Proj Configure Your Proj Empty Activity Creates a new empty activity. The application name for most apps begins w	Package name myedge Package name com.example. Save location D:\AndroidPro Language Java Minimum SDK ① Your app W Help me cho Use legacy ith an uppercase letter	myedge jects\myedge API 21: Android 5 in un on approxima oose android.support lib	Previous Previous 产品线授权修 不用 0 (Lollipop) nely 34.1% of de raries ⑦ Previous	Next 改这里为你 vices. 改为SDK3	Cancel	Finish × 其它情况

查看Logcat

有Android开发经验的用户可以跳过本段。

修改MainActivity文件

Covernae
protected void onCreate(Bundle savedInstanceState) {
<pre>super.onCreate(savedInstanceState);</pre>
<pre>setContentView(R.layout.activity_main);</pre>
// 加上下面这行
Log.i("MainActivity", "SHOW in Logcat"); // 表示记录info级别的日志
// Example of a call to a native method
TextView tv = findViewByld(R.id.sample_text);
tv.setText(stringFromJNI());
}

代码会标红, 此时鼠标在红色的"Log"上点以下,会提示Alt+Enter,按下Alt+Enter, 文件的第6行左右会自动添加

import android.util.Log;	
<pre>@Override protected void onCreate(Bundle savedInstanceState) { super.onCreate(savedInstanceState); android.util.Log? Alt+Enter (R.layout.activity_main); 最标放在Log上, Alt+Enter 会自动加入缺失库</pre>	
<pre>Log.i("MainActivity", "SHOW in Logcat");</pre>	
<pre>// Example of a call to a native method TextView tv = findViewById(R.id.sample_text); tv.setText(stringFromJNI()); }</pre>	

再次运行项目,可以在界面的"Run"和Logcat里看见我们之前打印的日志 "SHOW in Logcat"

111			
5	Log	at	\$ -
1		Xiaomi MI 8 Android 9, API 2 🔻 com.example.myorcapp (31195) 🔹 Verbose 🔹 Q. 搜索的文字	□ Regex Show only selected application ▼
1	E	ogcat测试的手机名称 当前app的包名 日志级别	只看当前app,可能要再选一次生效。
2	-	at com.android.internal.os.Zygoteinit.main(Zygoteinit.java:876)	
		2020-07-24 15:22:38.043 31195-31195/com.example.myorcapp I/MainActivity: SHOW in Logcat	
6	±	2020-07-24 15:22:38.091 31195-31224/com.example.myorcapp D/libEGL: eglInitialize: enter	
	↑	2020-07-24 15:22:38.091 31195-31224/com.example.myorcapp I/Adreno: QUALCOMM build	: 8c3e34a, Ic446a5d9fc
8		Build Date : 05/13/19	
5	+	OpenGL ES Shader Compiler Version: EV031.25.03.07	
1	5	Local Branch	



🔻 📃 myedge D:\AndroidProjects\myedge	
Igradle	
▶ 🖿 .idea	
🔻 🐂 app	
🖿 libs	
V src	
androidTest	
🔻 🖿 main	
🕨 🖿 java	
res	
🚮 AndroidManifest.xml	
test	
🜏 .gitignore	
🔊 build.gradle	
🚔 proguard-rules.pro	
🕨 🖿 gradle	
损 .gitignore	
🔊 build.gradle	
📊 gradle.properties	
gradlew	
🗧 gradlew.bat	
📊 local.properties	
📗 myedge-init.zip	
🔊 settings.gradle	
IIII External Libraries	
🗞 Scratches and Consoles	

本文会操作上图中的目录及文件:

[•] app/libs 目录下放入jar库文件,也可以放so文件

- app/src/main/java java代码目录
- app/src/main/assets 目前无此目录,之后放入模型文件
- app/src/main/AndroidManiftest.xml AndroidManiftest.xml文件,设置启动Activity和权限
- app/src/main/res/layout UI布局目录
- app/build.gradle 编译配置,比如修改包名

集成代码

集成之前,请确认已经跑通官方demo的精简版,官方demo有如下界面及类似结果

15:09 解育・・		t 📚 🚥
	1:mao:0.9901489	
	BUTTON	

集成步骤:

- 1. 集成库
- 2. 集成java代码
- 3. 设置权限及配置项
- 4. 复制模型文件
- 1. 集成库

复制app/libs 目录下库文件到自己的项目中

- A. 复制easyedge-sdk.jar库文件
- 如果项目中已经有其它的jar文件,那么和这些jar文件放一起
- 如果项目中没有其它的jar文件,参照官方demo方式,复制到app/libs目录下(本文的情况),与官方demo放在相同的位置

Baidu 百度智能云文档

新手	指南

Project 🔻 🤅	Э	<u>▼</u>
🔻 뻱 myedge D:\AndroidProjects\myedg	je	
🕨 🖿 .gradle		
🕨 🖿 .idea		
🔻 📑 app		
🔻 🖿 libs		
easyedge-sdk.jar		
V src		
B. 复制so目录		

φ

需要复制官方demo的libs/arm64-v8a 及 armeabi-v7a

- 如果项目中已有so库目录,arm64-v8a 及 armeabi-v7a下的so与已有目录合并。如果比如自己项目只存在arm64-v8a目录,那么官方demo的 armeabi-v7a就不需要复制了。
- 如果项目中没有so库目录(本文情况),以下二个方式二选一 *复制arm64-v8a 及 armeabi-v7a目录 到 app/src/main/jniLibs目下*参照官方 demo方式,复制arm64-v8a 及 armeabi-v7a目录到 app/libs目录下,与官方demo放在相同的位置。并修改app/build.gradle, 设

置jniLibs.srcDirs	<u>s</u>
	图片加载失败

2. 集成java代码

复制官方demo的infertest目录到自己项目中的infertest目录下,不必修改自己项目的包名。复制layout下的activity_main_test.xml到自己的项目中。之后修改android.appcompat类为androidx.appcompat下的。

具体步骤如下:

A. 复制官方demo的infertest目录

打开官方demo,右键点击java目录下的infertest目录,点"copy"

🔻 📷 app	4	compileSdkVersion 26
🕨 🖿 build	5 🗟	defaultConfig {
V 🖿 libs	6	applicationId "com.baidu.ai.easyaimobil
🕨 🖿 armб4-v8a	7	minSdkVersion 21
armeabi-v7a	8	targetSdkVersion <mark>26</mark>
🕨 📗 easyedge-sdk.jar	9	versionCode 1
🔻 🖿 src	10	versionName "1.0"
androidTest	11 🤤	ndk {
🔻 🖿 main	12	abiFilters 'armeabi-v7a', 'arm64-v8
assets	13 🖨	}
🕨 🖿 demo	14	}
infer-classify	15 👳	buildTypes {
🖢 test.jpg 🛛 N	ew	÷
🔻 🖿 java 🛛 🛛 Li	nk C++ Project with (Gradle
🔻 🖿 com.baidu.ai.edg	ut	免so员 Ctrl+X
► Inf ertest	9 <u>9</u>	Ctrl+C
右键点击 snpetest	opy Deference	Ctul LAIt Shift C
► En test	opy Reference	Ctri+Ait+Shift+C
► 🖿 xeye	opy Path	CHAR IN THE
💿 CameraActivit 😐 🖻	aste	Ctrl+V 원so资
C MainActivity Fi	nd <u>U</u> sages	Ctrl+G
C MyApplication Fi	nd in <u>P</u> ath	Ctrl+H
► res Re	epl <u>a</u> ce in Path	rdFil
🛻 AndroidManifest.xm 🛛 🗛	naly <u>z</u> e	•
test	ofactor	N

复制到自己项目中类似位置:


B. 复制官方demo的layout下的app/src/main/res/layout/activity_main_test.xml到自己项目的同名文件

Project 👻	0 ÷ 🗢 –	ac	tivity_main.xml 🗵 🧟 MainActivity.java 👋 🖉 build.gradle (myedge) 👋 🖉 build.gradle (:app) 👋 🏭 activity_main_test.xml
🔻 🐚 myedge D:\AndroidProj	jects\myedge		
🕨 🖿 .gradle		1	<pre><?xml version="1.0" encoding="utf-8"?></pre>
🕨 🖿 .idea		2 0	<pre>candroid.support.constraint.ConstraintLayout xmlns:android="http://schemas.android.com/apk/res/android"</pre>
🔻 📷 app		3	xmlns:app="http://schemas.android.com/apk/res-auto"
libs		4	<pre>xmlns:tools="http://schemas.android.com/tools"</pre>
V src		5	android:layout width="match parent"
androidTest		6	android:layout height="match parent"
🔻 🖿 main		7	tools:context=".MainActivity">
🕨 🖿 java		8	
v res		9	<textview< td=""></textview<>
drawable	e	10	android:id="@+id/sample text"
drawable	e-v24	11	android:lavout width="wrap content"
Iayout		12	android: layout beight="wrap content"
activit	ty_main.xml	13	android:text="Hello World!"
activit	ty_main_test.xml	14	app:layout constraintBottom toBottomOf="parent"
mipmap	-anydpi-v26	15	app:layout constraintLeft toLeftOf="barent"
mipmap	-hdpi	16	app:layout constraintRight toRightOf="parent"
mipmap	-mdpi	17	app:layout constraintTop toTopOf="parent" />
mipmap	-xhdpi	18	
mipmap	-xxhdpi	19	< Button
mipmap	-xxxhdpi	20	android:id="@+id/button2"
values		21	android:lavout width="0dp"
👬 Android Mai	nifest.xml	22	android:lavout height="wrap content"
test		23	android:lavout marginBottom="8dp"
.gitignore		24	android:text="Button"
🔊 build.gradle		25	app:layout constraintBottom toBottomOf="parent"
🗑 proguard-rules.pr	ro	26	app:layout constraintHorizontal bias="0.0"
🕨 🖿 gradle		27	app:layout constraintLeft toLeftOf="parent"
🦾 .gitignore		28	app:layout constraintRight toRightOf="parent"
🔊 build.gradle		29	app:layout constraintTop toTopOf="parent"
📊 gradle.properties		30	app:layout constraintVertical bias="0.974" />
gradlew		31	
🗐 gradlew.bat		32	
📊 local.properties		33	
myedge-init zin			

此时会发现xml文件里android.support.constraint.ConstraintLayout不存在,原因是自己的项目新建时用的是androidx.appcompat

C. 修改activity_main_test.xml文件

将 android.support.constraint.ConstraintLayout 修改为androidx.constraintlayout.widget.ConstraintLayout

roject 🔻	\ominus $$	¢ —	븕 activ	ity_main.xml 👋 💿 MainActivity.java 👋 🔊 build.gradle (myedge)
myedge D:\AndroidP	rojects\myedge			
.gradle			1	<pre><?xml version="1.0" encoding="utf-8"?></pre>
idea			2 C (<pre><androidx.constraintlayout.widget.constraintlayout pre="" xmlns:<=""></androidx.constraintlayout.widget.constraintlayout></pre>
📑 app			3	<pre>xmlns:app="http://schemas.android.com/apk/res-auto"</pre>
libs			4	<pre>xmlns:tools="http://schemas.android.com/tools"</pre>
src			5	android:layout width="match parent"
androidTest			6	android:layout_height="match_parent"
🔻 🖿 main			7	<pre>tools:context=".MainActivity"></pre>
🕨 🖿 java			8	
v res			9 (<textview< td=""></textview<>
drawa	ble		10	android:id="@+id/sample_text"
drawa	ble-v24		11	android:layout_width="wrap_content"
Iayout			12	android:layout_height="wrap_content"
🛃 act	vity_main.xml		13	android:text="Hello World!"
😸 acti	vity_main_test.xml		14	app:layout_constraintBottom_toBottomOf="parent"
mipma	p-anydpi-v26		15	app:layout_constraintLeft_toLeftOf="parent"
mipma	ip-hdpi		16	app:layout_constraintRight_toRightOf="parent"
mipma	ip-mdpi		17	<pre>app:layout_constraintTop_toTopOf="parent" /></pre>
mipma	ip-xhdpi		18	
🕨 🖿 mipma	ip-xxhdpi		19 (
🕨 🖿 mipma	ip-xxxhdpi		20	android:id="@+id/button2"
values			21	android:layout_width="0dp"
Android N	1anifest.xml		22	android:layout_height="wrap_content"
test			23	android:layout_marginBottom="8dp"
.gitignore			24	android:text="Button"
			25	app:layout_constraintBottom_toBottomOf="parent"
proguard-rules.	pro		26	app:layout_constraintHorizontal_bias="0.0"
gradle			27	app:layout_constraintLeft_toLeftOf="parent"
igitignore			28	app:layout_constraintRight_toRightOf="parent"
<i>i</i> wild.gradle			29	app:layout_constraintTop_toTopOf="parent"
gradle.properties			30	<pre>app:layout_constraintVertical_bias="0.974" /></pre>
■ gradlew			31	
i≣ gradlew.bat			32	
local.properties			33	
📗 myedge-init.zip				

D. 修改infertest/MainActivity.java文件

删除飘红的类导入:



选下面的飘红的类,使用Alt+Enter自动导入缺少的类。

Baidu 百度智能云文档

$\odot - ~$		in Activity Java
D:\AndroidProjects\myedge	1	<pre>package com.example.myedge.infertest;</pre>
	2	
	3	⊖import android.Manifest;
	4	<pre>import android.app.Application;</pre>
	5	<pre>import android.content.pm.PackageManager;</pre>
	6	<pre>import android.os.AsyncTask;</pre>
androidTest	7	<pre>import android.os.Bundle;</pre>
nain	8	<pre>import android.util.Log;</pre>
🖿 java	9	<pre>import android.view.View;</pre>
🗸 🖿 com.example.myedge	10	<pre>import android.widget.Button;</pre>
🔻 🛅 infertest	11	<pre>import android.widget.TextView;</pre>
C AndroidBmpUtil	12	
Contractivity	13	
Content State S	14	<pre>import androidx.appcompat.app.AppCompatActivity;</pre>
Content of the second secon	15	<pre>import androidx.core.app.ActivityCompat;</pre>
Content State S	16	<pre>import androidx.core.content.ContextCompat;</pre>
Contractivity	17	
res	18	<pre>import com.example.myedge.R;</pre>
🕨 🖿 drawable	19	
drawable-v24	20	<pre>import java.io.IOException;</pre>
🕫 🖿 layout	21	⊖import java.util.ArrayList;
activity_main.xml	22	
activity_main_test.xml	23 🛃	<pre>public class MainActivity extends AppCompatActivity {</pre>
mipmap-anydpi-v26	24	
🕨 🖿 mipmap-hdpi	25	// 请替换为你自己的序列号
mipmap-mdpi	26	private static final String SERIAL_NUM = "8B83-E6B4-258D-8D4B"; //"XXXX-XXXX-XXXX-XXXX
🕨 🖿 minman-yhdni	27	

此时项目可以编译成功,但是不能运行

3. 设置权限及配置项

将启动的Activity改为infertest.MainActivity,并根据官方demo添加网络和外部储存权限

具体步骤如下:

A. 将启动的Activity改为infertest.MainActivity

修改app/AndroidManifest.xml文件,将启动的Activity改为从.MainActivity改为infertest.MainActivity,



B. 根据官方demo添加网络和外部储存权限

添加如下权限

- <uses-permission android:name="android.permission.ACCESS_NETWORK_STATE"/>
- <uses-permission and rold_name="android.permission.INTERNET"/>
- <uses-permission android_name="android.permission.WRITE_EXTERNAL_STORAGE"/>

	— © M	/ainActivity.java 🛛 🏭 AndroidManifest.xml 🖂
D:\AndroidProjects\myedge	1	xml version="1.0" encoding="utf-8"?
•	2	<pre><manifest <="" pre="" xmlns:android="http://schemas.android.com/apk/res/android"></manifest></pre>
	3	<pre>package="com.example.myedge"></pre>
	4	<pre><uses-permission android:name="android.permission.ACCESS_NETWORK_STATE"></uses-permission></pre>
d	5	<uses-permission android:name="android.permission.INTERNET"></uses-permission>
	6	<pre><uses-permission android:name="android.permission.WRITE_EXTERNAL_STORAGE"></uses-permission></pre>
	7	<pre><application< pre=""></application<></pre>
androidTest	8	android:allowBackup="true"
nain	9	android:icon="@mipmap/ic_launcher"
🗖 java	10	android:label="myedge"
🔻 🖿 com.example.myedge	11	android:roundIcon="@mipmap/ic_launcher_round"
🔻 🖿 infertest	12	android:supportsRtl="true"
Contraction Contractica Con	13	android:theme="@style/AppTheme">
C MainActivity	14	<pre><activity android:name=".infertest.MainActivity"></activity></pre>
Content of the state of the	15	<pre><intent-filter></intent-filter></pre>
Content State S	16	<action android:name="android.intent.action.MAIN"></action>
Content State S	17	
Contractivity	18	<category android:name="android.intent.category.LAUNCHER"></category>
res	19	<pre></pre>
🚛 AndroidManifest.xml	20	<pre></pre>
est	21	<pre></pre>
gnore	22	
d.oradle	23	<pre>/manifest></pre>

注意,android 6.0以上,需要额外代码ActivityCompat.requestPermissions申请权限,具体代码见infertest.MainActivity中initPermission方法

4. 复制模型文件

复制官方demo的app/src/main/assets目录到自己项目的同名目录。

如果已经存在,合并即可。



此时运行,可以获得和官方demo精简版一样的效果。

16:47 P P

myedge	
	1:mao:0.9901489
	BUTTON

d 🛜 🊥

修改包名 (仅"产品线激活"需要)

如果您填入的包名是"com.baidu.ai.easyaimobile.demo" 如图修改app/build.gradle:

Project 🔻 😳 🛨 🛱	- 💿 MainActivity.java 🛛 🔊 build.gradle (:app) 👋 🟭 AndroidManifest.xml 🗵
矄 myedge D:\AndroidProjects\myedge	Gradle files have changed since last project sync. A project sync may be necessary for the IDE to work
Igradle	1 apply plugin: 'com.android.application'
🕨 🖿 .idea	2
🔻 📷 app	Bandroid {
build	4 compileSdkVersion 29
libs	5 buildToolsVersion "30.0.0"
V src	6
androidTest	7 defaultConfig {
🔻 🖿 main	8 applicationId com.baidu.ai.easvaimobile.demo"
▼ assets	9 minSdkVersion 21
demo	10 targetSdkVersion 29
infer-classify	11 versionCode 1
🚽 test.jpg	12 versionName "1.0"
🕨 🖿 java	13
res	14 testInstrumentationRunner "androidx.test.runner.AndroidJUnitRunner"
AndroidManifest.xml	15 🗎 }
test	16
😹 .gitignore	17 🖯 buildTypes {
🔊 build.gradle	18 release {
🗑 proguard-rules.pro	19 minifyEnabled false
gradle	20 proguardFiles getDefaultProguardFile('proguard-android-optimize.txt
🐻 .gitignore	21 3
🔊 build.gradle	22
📊 gradle.properties	23





通用设备端Linux ARM

心简介

本文将为新手提供一个快速测试和集成EasyDL & EasyEdge的 Linux Arm SDK的图文教程。

测试前的准备

- Linux ARM的硬件及开发环境
 - 详情参考下方文档
- EasyDL平台的Linux ARM SDK
 - 以图像分类为例,前往操作台训练模型后,选择发布为Linux ARM的通用设备端SDK,发布成功后即可从平台下载
- 用于激活通用设备端SDK的序列号
 - 前往控制台申请用于激活通用设备端SDK的序列号
 - 首次使用SDK或者更换序列号、更换设备时,需要联网激活。激活成功之后,有效期内可离线使用





适用于Linux Arm操作系统,如

- ubuntu、centos等
- 树莓派Raspbian
- ...

且适用于aarch64和armv7hf的CPU架构。

用户使用以上系统和架构的硬件(如RK3399开发板、树莓派4B等)即可。 **网络环境**用户安装软件和测试SDK都需要联网,用户使用的硬件需确保有效的网络连接。

软件环境 在选定硬件之后,需要在硬件上安装以下软件和第三方库以保证SDK正常编译和运行:

- cmake 3 +
- gcc 5.4 +
- opencv3.4 (可选)

可以使用以下方法确认cmake是否满足要求:

\$ cmake --version cmake version 3.13.3

若系统提示找不到cmake命令或者cmake version 低于3.x.x,则需要安装/升级cmake。

// apt安装cmake sudo apt update sudo apt install cmake

ubuntu官方的apt源update较慢,且可能访问不了,可以替换为国内的源:https://www.cnblogs.com/yongy1030/p/10315569.html

也可以使用源码编译的方式安装,参考安装方法。

// 升级cmake sudo apt-get install software-properties-common sudo add-apt-repository ppa:george-edison55/cmake-3.x sudo apt-get update sudo apt-get upgrade

安装/升级cmake后可再次执行cmake --version确认版本。

可以使用以下方法确认gcc是否满足要求:

\$ gcc --version gcc (Ubuntu/Linaro 5.4.0-6ubuntu1~16.04.12) 5.4.0 20160609

若系统提示找不到gcc命令或者gcc version 低于5.4.0,则需要安装/升级gcc。

// 安装gcc
sudo apt update
sudo apt install build-essential
// 升级gcc
sudo add-apt-repository ppa:ubuntu-toolchain-r/test # 如果找不到add-apt-repository命令,执行: apt-get install software-propertiescommon
sudo apt-get update
sudo apt-get install -y gcc-5 g++-5

```
cd /usr/bin # 升级gcc 5之后,还需要替换原来的软链接
sudo rm -r gcc # 移除之前的软连接
sudo ln -sf gcc-5 gcc # 建立gcc5的软连接
sudo rm -r g++ # 同gcc
sudo ln -sf g++-5 g++
```

安装/升级gcc后可再次执行gcc --version确认版本。

目前没有合适的方法确认系统中是否有SDK需要的OpenCV,若用户不确定是否安装OpenCV 3.4 +,并且可以被cmake find_package到,可以手 动编译安装OpenCV 3.4,也可以在之后编译SDK时自动编译OpenCV。

若选择在下一步编译EasyEdge SDK时自动编译OpenCV,则以下编译安装OpenCV的步骤可跳过。

下载OpenCV 3.4源代码包并解压:下载地址,然后编译安装:

```
// 编译安装OpenCV
cd opencv-3.4.6
mkdir build
cd build
```

cmake .. -DBUILD_DOCS=OFF -DBUILD_EXAMPLES=OFF -DBUILD_opencv_python2=OFF -DBUILD_opencv_python3=OFF -DBUILD_WITH_DEBUG_INFO=OFF -DBUILD_PACKAGE=OFF -DBUILD_opencv_core=ON -DBUILD_opencv_imgproc=ON -DBUILD_opencv_imgcodecs=ON -DBUILD_opencv_highgui=ON -DBUILD_opencv_video=OFF -DBUILD_opencv_videoi=OFF -DBUILD_opencv_dnn=OFF -DBUILD_opencv_apps=OFF -DBUILD_opencv_flann=OFF -DBUILD_opencv_gpu=OFF -DBUILD_opencv_mI=OFF -DBUILD_opencv_legacy=OFF -DBUILD_opencv_calib3d=OFF -DBUILD_opencv_features2d=OFF -DBUILD_opencv_java=OFF -DBUILD_opencv_objdetect=OFF -DBUILD_opencv_photo=OFF -DBUILD_opencv_nonfree=OFF -DBUILD_opencv_ocl=OFF -DBUILD_opencv_stitching=OFF -DBUILD_opencv_superres=OFF -DBUILD_opencv_ts=OFF -DBUILD_opencv_videostab=OFF -DBUILD_opencv_contrib=OFF -DBUILD_SHARED_LIBS=ON -DBUILD_TESTS=OFF -DBUILD_PERF_TESTS=OFF -DBUILD_WITH_CAROTENE=OFF -DCMAKE_BUILD_TYPE:STRING=Release -DWITH_FFMPEG=OFF -DWITH_IPP=OFF -DBUILD_PNG=ON -DBUILD_JPEG=ON -DBUILD_ZLIB=ON -DBUILD_FAT_JAVA_LIB=OFF -DOPENCV_CXX11=OFF -DCMAKE_INSTALL_PREFIX:PATH=/usr/lib/aarch64-linux-gnu/

```
make # 如果有多个cpu可以用-j加快编译速度,如4个CPU用 make -j4 make install
```

心测试demo

SDK介绍 用户下载的Linux Arm SDK zip包中包含SDK动态库、模型等资源文件和测试demo.cpp。

需要将SDK zip包完整的放入Arm硬件上再进行解压,否则可能会报错:

libeasyedge.so: file format not recognized; treating as linker script

Linux下解压命令: tar -xvf xxx.tar

若用户使用的硬件的CPU架构为aarch64,则解压baidu_easyedge_linux_cpp_aarch64_ARM_gcc5.4_vx.x.x_xxxxxx.tar。

若CPU架构为armv7hf,则解压baidu_easyedge_linux_cpp_armv7hf_ARM_gcc5.4_vx.x.x_xxxxxx.tar。

RK3399等开发板一般是aarch64架构,树莓派一般是armv7hf架构(最新的4B可以刷成aarch64架构)。

可通过下面的命令确认CPU架构 (armv7l在树莓派上实际是指armv7hf) :

\$ uname -m aarch64 # 或者是armv7I

解压完对应的tar包之后的目录结构如下:



编译demo 前面安装了cmake、gcc等工具之后,可以编译SDK demo,生成测试的可执行文件。步骤如下。

```
一、将获取的序列号填入demo.cpp
```

```
$ cd demo # 进入demo文件夹
$ vi demo.cpp # 若vi未找到命令,执行 sudo apt install vim
```

在打开的代码编辑页面,找到

global_controller()->set_licence_key("set your license here");

将序列号填入引号内。如果想打印demo运行过程中的日志,找到

log_config.enable_debug = false;

将false改为true即可。

二、编译 在demo目录下执行

```
mkdir build # 创建build目录
cd build
cmake .. # 如果系统中安装了opencv3.4以上
##### 或者
cmake .. -DEDGE_BUILD_OPENCV=ON # 自动编译安装opencv
```

当出现:

-- Configuring done -- Generating done -- Build files have been written to: /xxx/demo/build

表示cmake成功。然后执行编译

make # 如果有多个cpu可以用-j加快编译速度,如4个CPU用 make -j4

当出现:

[100%] Built target easyedge_serving [100%] Built target easyedge_demo

表示编译成功,在build目录下出现编译的产物:

- easyedge_demo :测试的可执行文件
- easyedge_serving:包含http server的测试的可执行文件
- thirdparty:编译安装的opencv

测试easyedge_demo 在build目录下执行:

./easyedge_demo {模型RES文件夹} {测试图片路径}

第一个参数为包含模型的文件夹路径,第二个参数为测试的图片的路径。SDK中已经包含模型文件夹,如果用户有其他模型文件,可以指定为其路径。如:

./easyedge_demo ../../../RES /xxx/test.jpg

```
然后可以看到输出的结果:
```

```
0 build > ./easyedge_demo ~/lvxiangxiang/models/SqueezeNetV1.1-tf/ ~/lvxiangxiang/images/orange.jpg
2020-07-16 16:52:33,632 DEBUG [EasyEdge] 548062597120 The recommended threshold is 0.3
2020-07-16 16:52:33,632 DEBUG [EasyEdge] 548062597120 Rescale mode is 0, target_size: 0, max_size: 0
2020-07-16 16:52:33,632 DEBUG [EasyEdge] 548062597120 Config file read done
2020-07-16 16:52:33,636 DEBUG [EasyEdge] 548062597120 Local license is ok.
2020-07-16 16:52:33,855 DEBUG [EasyEdge] 548062597120 Image will be resized to 227,227
2020-07-16 16:52:33,986 DEBUG [EasyEdge] 548062597120 Inference costs 131.011(127.854) ms
950, orange, p:0.972394
Done
```

如果是物体检测或者图像

分割模型,可以打开生成的/xxx/test.result.cpp.jpg图片,查看检测框的效果。

测试easyedge_serving easyedge_serving会开启一个http server服务,并实现了一个简单的网页,用户可以在网页上上传图片并查看预测结果。

在build目录下执行:

./easyedge_serving {模型RES文件夹} {序列号} {主机ip, 默认 0.0.0.0} {端口, 默认 24401}

如:

./easyedge_serving ../../../RES "1111-1111-1111" 0.0.0.0 24401

若日志显示:

HTTP is now serving at 0.0.0.0:24401

表示http server启动成功。此时可以打开浏览器,输入网址http://{设备ip}:24401,上传图片来进行测试。

查看设备ip的方法:

ifconfig # 如果没有ifconfig命令,执行 sudo apt install net-tools

```
如果是有线连接,找到ethO一栏,如果是wifi连接,找到wlanO一栏。
```

注意:只有本机电脑和硬件设备的网络ip在同一网段之下,才可以通过网址访问。



の集成SDK

SDK提供了一系列模型加载、预测等接口,用户可以方便的集成进自己的程序之中。

接口说明、数据格式说明以及常见错误请参考SDK技术文档。

建议先测试Demo,以及参考demo.cpp和demo的CMakeLists.txt调用流程。如果遇到错误,优先参考文件中的注释以及日志说明。

一、导入SDK头文件和库文件

在baidu_easyedge_linux_cpp_aarch64_ARM_gcc5.4_vx.x.x_xxxxxx/include下有SDK的头文件。 在 baidu_easyedge_linux_cpp_aarch64_ARM_gcc5.4_vx.x.x_xxxxxx/lib下有SDK的库文件,包含动态库libeasyedge.so和静态库 libeasyedge_static.a。用户可选择合适的导入方式。

用户将头文件和库文件拷贝至自己的项目中,并在自己的CMakeLists.txt中引用:

```
find_package(OpenCV REQUIRED)
// 导入头文件
include_directories(
    ${OpenCV_INCLUDE_DIRS}
        ${CMAKE_SOURCE_DIR}/../include/
)
// 导入库文件
link_directories(
        ${CMAKE_SOURCE_DIR}/../lib/
)
// 链接库文件
target_link_libraries((your_executable_file) ${OpenCV_LIBS} easyedge paddle_full_api_shared)
```

二、在程序中调用SDK接口

// 引入SDK头文件
include "easyedge/easyedge.h"
// step 0: 设置序列号
global_controller()->set_licence_key("set your license here");
// step 1: 配置模型资源目录
PaddleFluidConfig config;
config.model_dir = (模型文件目录);
// step 2: 创建并初始化Predictor
auto predictor = global_controller()->CreateEdgePredictor(config);
auto img = cv::imread((测试图片路径));
// step 3: 预测图像
std::vector<EdgeResultData> result;
predictor->infer(img, result);

如果是口罩检测模型,将PaddleFluidConfig config修改为PaddleMultiStageConfig config。口罩检测模型请注意输入图片中人脸大小建议保持在8*8到96*96像素之间,可根据场景远近程度缩放图片后再传入SDK。

通用设备端Windows x86加速版

心简介

Windows CPU加速版SDK是适用于EasyDL图像模型快速部署的工具包。SDK中包含了EasyDL训练的模型资源文件、SDK和demo文件。

测试前的准备

- Windows x86的硬件及开发环境
 - 详情参考下方文档
- EasyDL平台的Windows x86 加速版SDK
 - 以图像分类为例,前往操作台训练模型后,选择发布为Windows x86的通用设备端SDK并勾选加速版,发布成功后即可从平台下载
- 用于激活通用设备端加速版SDK的序列号
 - 前往控制台申请用于激活通用设备端SDK的序列号,注意选择加速版序列号
 - 首次使用SDK或者更换序列号、更换设备时,需要联网激活。激活成功之后,有效期内可离线使用

心 安装依赖

在使用SDK之前,首先要确认自己的硬件类型和相应的依赖库安装是否已经符合要求。

硬件要求:

- Intel Xeon with AVX2 and AVX512
- Intel Core Processors with AVX2
- Intel Atom Processors with SSE

软件要求:

- 64位 Windows 10
- .NET Framework 4.5
- Visual C++ Redistributable Packages for Visual Studio 2013
- Visual C++ Redistributable Packages for Visual Studio 2015
- Openvino 2020.1

其他要求:

• 第一次使用SDK请确保联网

心 SDK结构

医 图片加载失败

其中:

- bootstrap是SDK的入口脚本。
- data/model文件夹下是EasyDL训练得到的模型资源文件。
- tools文件夹下提供的是模型更新工具,用在迭代训练模型后,直接拉取新训练的模型,而不用重新下载SDK。

运行demo

打开EasyEdge.exe,输入Serial Num

	data	2020/4/15 23:51 文件夹	
*	AI 百度EasyEdge	- 🗆 X	
* *	▲ 百度EasyEdge Model Serial Nur Host Port 服务状态	× id:Easy 127.0.0.1 服务未启动 是否开机自动启动 []	2: 点击"启动服务",等待数
		启动服务	

秒即可启动成功,本地服务默认运行在

http://127.0.0.1:24401/

服务运行成功,此时可直接在浏览器中输入http://127.0.0.1:24401,在h5中测试模型效果。

AI 百度EasyEdge	× +		
▲ 不安全	6		
	【图像分割】45274 分割-电池-设备端V1	调整阈值	—————————————————————————————————————
		label	置信度
	4.11	battery	0.998
		tomato	0.997
		tomato	0.996
		battery	0.995
		tomato	0.995
		tomato	0.993
		tomato	0.972

^の Http服务集成

服务运行成功后,除网页直接访问外,也可以通过http请求的方式执行模型的预测并获取预测结果。

图像服务调用说明

Python 使用示例代码如下

import requests

```
with open('./1.jpg', 'rb') as f:
img = f.read()
```

params 为GET参数 data 为POST Body result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1}, data=img).json()

C# 使用示例代码如下



StreamReader sr = new StreamReader(response.GetResponseStream(Console.WriteLine(sr.ReadToEnd()); sr.Close(); response.Close();

C++ 使用示例代码如下,需要安装curl

```
##### include <sys/stat.h>
##### include <curl/curl.h>
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./img.jpg";
  FILE *fp = NULL;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\" \n", post_data_filename);
     return -1;
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
     \label{eq:curl_easy_setopt} curl_easy\_setopt(curl, CURLOPT\_URL, "http://127.0.0.1:24401?threshold=0.1");
     curl_easy_setopt(curl, CURLOPT_POST, 1L);
     curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
     curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
     res = curl_easy_perform(curl);
     if (res != CURLE_OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     }
     curl_easy_cleanup(curl);
  }
  curl_global_cleanup();
  fclose(fp);
  return 0:
```

请求参数

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送图片二进制。

返回参数|字段|类型|取值|说明||--------|-----|-----|-----||confidence|float|0~1|分类或检测的置信度||label|string||分 类或检测的类别||index|number||分类或检测的类别||x1,y1|float|0~1|物体检测,矩形的左上角坐标(相对长宽的比例值)||x2,y2|float |0~1|物体检测,矩形的右下角坐标(相对长宽的比例值)|

关于矩形坐标

- x1 * 图片宽度 = 检测框的左上角的横坐标
- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2 * 图片宽度 = 检测框的右下角的横坐标
- y2 * 图片高度 = 检测框的右下角的纵坐标

图像分割 返回结果格式参考API调用文档 代码参考 https://github.com/Baidu-AIP/EasyDL-Segmentation-Demo

声音服务调用说明

Python 使用示例代码如下

import requests

```
with open('./1.mp3', 'rb') as f:
audio = f.read()
```

```
###### params 为GET参数 data 为POST Body
result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1},
data=audio).json()
```

C# 使用示例代码如下

```
FileStream fs = new FileStream("./audio.mp3", FileMode.Open);
BinaryReader br = new BinaryReader(fs);
byte[] audio = br.ReadBytes((int)fs.Length);
br.Close();
fs.Close();
string url = "http://127.0.0.1:8402?threshold=0.1";
HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url);
request.Method = "POST";
Stream stream = request.GetRequestStream();
stream.Write(audio, 0, audio.Length);
stream.Close();
WebResponse response = request.GetResponse();
StreamReader sr = new StreamReader(response.GetResponseStream());
```

StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close(); response.Close();

C++ 使用示例代码如下,需要安装curl

```
##### include <sys/stat.h>
##### include <curl/curl.h>
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./audio.mp3";
  FILE *fp = NULL;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  if (fstat(fileno(fp), &stbuf) || !S\_ISREG(stbuf.st\_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\" \n", post_data_filename);
     return -1;
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
     \label{eq:curl_easy_setopt} curl_easy\_setopt(curl, CURLOPT\_URL, "http://127.0.0.1:24401?threshold=0.1");
     curl_easy_setopt(curl, CURLOPT_POST, 1L);
     \label{eq:curl_easy_setopt} curl, \ \ CURLOPT\_POSTFIELDSIZE\_LARGE, (curl\_off\_t) stbuf.st\_size); \\
     curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
     res = curl_easy_perform(curl);
     if (res != CURLE_OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     }
     curl_easy_cleanup(curl);
  }
  curl_global_cleanup();
  fclose(fp);
  return 0:
```

请求参数

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送声音二进制。

返回参数|字段|类型|取值|说明||--------|----|-----|-----|| confidence | float | 0~1 | 分类或检测的置信度 | | label | string | | 分 类或检测的类别 | | index | number | | 分类或检测的类别 |

服务器端Linux GPU 加速版

心简介

Linux GPU加速版SDK是适用于EasyDL经典版、EasyDL专业版模型快速部署的工具包。SDK中包含了EasyDL训练的模型资源文件、SDK和demo文件。

测试前的准备

- Linux x86 GPU的硬件及开发环境
 - 详情参考下方文档

- EasyDL平台的Linux x86 GPU 加速版服务器端SDK&激活序列号
 - 以经典版图像分类为例,在操作台训练「私有服务器部署-服务器端SDK」下的模型后,前往控制台申请发布Linux x86 GPU的服务器端 SDK,发布成功后即可同时获得加速版SDK和序列号
 - 首次使用SDK或者更换序列号、更换设备时,需要联网激活。激活成功之后,有效期内可离线使用

心 安装依赖

在使用SDK之前,首先要确认自己的硬件类型和相应的依赖库安装是否已经符合要求。

硬件要求:

• Linux GPU加速版SDK支持绝大部分Nvidia显卡。

软件要求:

- gcc 5.4+ (需包含GLIBCXX_3.4.22)
- cmake 3+
- CUDA 9.0 + cuDNN 7.5 或 CUDA 10.0 + cuDNN 7.5
- TensorRT 7.0
- OpenCV 3.4

其他要求:

• 第一次使用SDK请确保联网

の SDK结构

获取到的SDK解压后的目录结构是:



其中:

- cpp文件夹下有两个压缩包分别为适配不同版本CUDA的SDK,解压缩后得到include头文件、lib库文件和demo示例代码文件。
- RES文件夹下是EasyDL训练得到的模型资源文件。
- tools文件夹下提供的是模型更新工具,用在迭代训练模型后,直接拉取新训练的模型,而不用重新下载SDK。

®编译demo

解压SDK后,进入到demo目录可以直接编译demo。编译方法如下:

```
##### 1.首先进入demo目录
##### 2. 创建build文件夹并进入
$ mkdir build && cd build
##### 3. 执行编译
$ make ..
$ make ..
$ make ..3
##### 4. 执行安装 , 把ib文件安装到系统路径 , 需要sudo权限 ; 也可以选择不执行安装 , 把ib路径加为环境变量即可
$ sudo make install
##### 5. 这时候应该会产出demo编译的可执行文件 , 直接执行可以查看需要的参数
$ ./easyedge_batch_inference {res_dir} {image/image_dir}
##### 或
$ ./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401}
```

编译产出有三个可执行文件,分别为:

[•] easyedge_batch_inference:提供了对批量图片预测的能力,预测速度快,是最推荐使用的方式。

- easyedge_multi_thread:提供了对多卡或多线程预测的支持。其中多线程的支持是针对一张显卡而言的,这种方式通常并不推荐使用,因为计 算资源的相互竞争会拉慢单次预测的速度并且预测时间不稳定。
- easyedge_serving:提供了http服务的能力。启动服务后,可以轻松把http服务接口集成到自己的应用中。

其中前两个demo示例了SDK API的集成方式,第三个demo示例了如何使用SDK创建http服务。根据个人需求选择不同集成方式即可,接下来将分别介绍这两种集成方式。

ல API集成

使用SDK的API接口方式能提供功能更丰富、预测速度更快的能力。API接口的设计尽可能的降低了调用复杂度,可以很方便的集成到自己的应用 当中。

调用流程

// 1. 设置序列号
global_controller()->set_licence_key("ABCD-ABCD-ABCD");
// 2. 配置运行选项
TensorRTConfig config:
config.model_dir = argv[1];
config.device = 0; // 设置需要使用的GPU
config.max_batch_size = 4; // 优化的模型可以支持的最大batch_size,实际单次推理的图片数不能大于此值
config.max_concurrency = 1; // 设置device对应的卡可以使用的最大线程数
config.fp16 = false; //置true开启fp16模式推理会更快,精度会略微降低,但取决于硬件是否支持fp16,不是所有模型都支持fp16,
参阅文档
// 3. 创建predictor并初始化
auto predictor = global_controller()->CreateEdgePredictor(config);
predictor->init();
// 4. 执行预测
predictor->infer(imgs, result);

当机器上存在多张显卡时,可以创建多个predictor,每一个predictor通过config配置不同的显卡id,这样就可以同时在不同的显卡上执行预测。具体使用方式可以参考demo_multi_thread.cpp的示例代码。这里贴出:

```
// 配置要使用的显卡id
std::vector<int> devices{0, 1, 2, 3};
// 每个predictor对应一张显卡
std::vector<std::unique_ptr<EdgePredictor>> predictors;
// 配置运行选项
TensorRTConfig config;
config.model_dir = model_dir;
config.max_batch_size = 8; // 单次预测可以支持的最大batch_size,多个predictor的此值要保持一致
config.max_concurrency = 1; // 单个device上的最大infer并发量,建议保持为1
config.compile_level = 1; // 编译模型的策略,如果当前设置的max_batch_size与历史编译存储的不同,则重新编译模型
config.fp16 = false; // 置true开启fp16模式推理会更快,精度会略微降低,但取决于硬件是否支持fp16,且不是所有模型都支持fp16,
参阅开发文档
for(int i = 0; i < predictors.size(); ++i) {
     config.device = devices[i]; // 设置GPU id
     // 创建predictor
     auto predictor = global_controller()->CreateEdgePredictor<TensorRTConfig>(config);
     if (predictor->init() != EDGE_OK) {
            exit(-1);
      predictors.emplace_back(std::move(predictor));
// 创建与predictor数量等同线程数的infer_task
std::vector<std::thread> threads;
std::map<int /*predictor index*/, std::vector<std::vector<EdgeResultData>>> results;
for (int i = 0; i < predictors.size(); ++i) {
     assert(split_img_files[i].size() <= config.max_batch_size);</pre>
      // infer_task的实现详见demo_multi_thread.cpp, 主要功能为执行infer预测
     threads.emplace\_back(std::thread(infer\_task, std::ref(predictors[i]), std::ref(split\_img\_files[i]), std::ref(results), i)); the std::ref(res
for (auto& t : threads) {
     if (t.joinable()) {
            t.join();
}
```

通过创建多个predictor可以将同一个模型同时部署到多张显卡上,这样可以提升执行预测的吞吐率。当然也可以在不同的显卡上部署不同的模型,通过在config中指定不同的model_dir即可。

可以看到,调用SDK的API流程很简单,并可以通过config的不同配置实现不同的能力。比如支持批量图片预测、多卡部署、多线程预测、fp16加 速等。实际集成的过程中,最需要注意的就是config的配置,一个适合自己应用场景的config参数配置才能带来最佳的预测速度。

参数配置 config的定义可以参见头文件easyedge.h

```
struct TensorRTConfig : public EdgePredictorConfig {
std::string model_filename{"model"};
 std::string params_filename{"params"};
 std::string cache_name{"m_cache"};
 /**
 * @brief GPU工作空间大小设置
 * workspace_size = workspace_prefix * (1 << workspace_offset)
 * workspace_offset: 10 = KB, 20 = MB, 30 = GB
 */
int workspace_prefix{16};
 int workspace_offset{20};
 /**
 * @brief 生成最大 batch_size 为 max_batch_size 的优化模型,单次预测图片数量可以小于或等于此值
 */
int max_batch_size{1};
 /**
 * @brief 设置使用哪张 GPU 卡
 */
int device{0};
 /**
 * @brief 模型编译等级
 * 0:无论当前设置的max_batch_size是多少,仅使用历史编译产出 (如果存在)
 *1:如果当前max_batch_size与历史编译产出的max_batch_size不相等时,则重新编译模型(推荐)
 * 2:无论历史编译产出的max_batch_size为多少,均根据当前max_batch_size重新编译模型
 */
int compile_level{1};
 /**
 * @brief 设置device对应的卡可以支持的最大并发量
 * 实际预测的时候对应卡的最大并发量不超过这里设置的范围
 */
int max_concurrency{1};
 /**
 * @brief 是否开启fp16模式预测,需要硬件支持
 */
bool fp16{false};
 /**
 * @brief 设置需要使用的DLA Core
 */
int dla_core{-1};
```

<mark>cache_name</mark>:GPU加速版SDK首次加载模型会先对模型进行编译优化,通过此值可以设置优化后的产出文件名。比如,可以根据不同的配置来 命名产出文件,从而可以在之后的运行中直接通过文件名就可以索引到编译产出,而不用再次执行优化过程。

workspace_size:设置运行时可以被用来使用的最大临时显存。通常默认即可,但当执行预测失败时,可以适当调大此值。

max_batch_size:此值用来控制批量图片预测可以支持的最大图片数,实际预测的时候单次预测图片数不可大于此值,但可以是不大于此值的任 意图片数。建议设置为与实际需要批量预测的图片数量保持一致,以节省内存资源并可获得较高预测速度。

device:设置需要使用的 GPU 卡号。

compile_level:模型编译等级。通常模型的编译会比较慢,但编译产出是可以复用的。可以在第一次加载模型的时候设置合理的 max_batch_size 并在之后加载模型的时候直接使用历史编译产出。是否使用历史编译产出可以通过此值 compile_level 来控制,当此值为 0 时,表示忽略当前设 置的 max_batch_size 而仅使用历史产出(无历史产出时则编译模型);当此值为 1 时,会比较历史产出和当前设置的 max_batch_size 是否相 等,如不等,则重新编译;当此值为 2 时,无论如何都会重新编译模型。

max_concurrency:通过此值设置单张 GPU 卡上可以支持的最大 infer 并发量,其上限取决于硬件限制。init 接口会根据此值预分配 GPU 资源, 建议结合实际使用控制此值,使用多少则设置多少。注意:此值的增加会降低单次 infer 的速度且预测速度不稳定,建议优先考虑 batch inference。

fp16:默认是 fp32 模式,置 true 可以开启 fp16 模式预测,预测速度会有所提升,但精度也会略微下降,权衡使用。注意:不是所有模型都支持 fp16 模式,也不是所有硬件都支持 fp16 模式。已知不支持fp16的模型包括:EasyDL经典版图像分类高精度模型。

运行demo 命令执行格式:

./easyedge_batch_inference {模型RES文件夹} {测试图片路径或图片文件夹路径} ./easyedge_multi_thread {模型RES文件夹} {测试图片路径或图片文件夹路径}

如:



预测效果展示:



^の Http服务集成

使用SDK的http服务API可以创建一个http服务端口,从而可以在应用里或者网页中通过http请求的方式执行模型的预测并获取预测结果。

调用流程



编译并运行后就会在控制台启动一个服务,默认地址为:http://{设备ip}:24401,通过在浏览器访问此地址可以打开一个测试页面,上传图片即 可获取识别结果。

运行demo demo命令执行格式

./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401}

如:



当出现提示:HTTP is now serving at 0.0.0.0:24401, holding 1 instances 时,表示服务已经启动成功,这时候可以在浏览器里输入地址打开一个 测试页面。如下所示:



上传图片,可以看到预测效果:



图片加载失败

也可以通过在自己的项目中请求此地址做预测,并获取预测结果。URL中的get参数:

参数	说明		默认值
threshold	阈值过滤,	0~1	如不提供,则会使用模型的推荐阈值

如使用Python的请求示例:

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
        'http://127.0.0.1:24401/',
        params={'threshold': 0.1},
        data=img).json()
```

http请求的返回格式为:

917	字段	类型说明	其他
e	error_code	Number	0为成功,非0参考message获得具体错误信息
r	esults	Array	内容为具体的识别结果。其中字段的具体含义请参考开发文档
c	cost_ms	Number	预测耗时ms,不含网络交互时间

专项适配硬件EdgeBoard(FZ)

心简介

本文将为新手提供一个快速测试和集成EasyDL & EasyEdge的 Linux EdgeBoard-FZ SDK的图文教程。

测试前的准备

- EdgeBoard(FZ)硬件及开发环境
 - 详情参考下方文档
- EasyDL平台的EdgeBoard(FZ)专用SDK
 - 以图像分类为例,前往操作台训练「专项硬件适配SDK-EdgeBoard(FZ)」下的模型并发布SDK后,即可从平台下载
- 用于激活专用SDK的序列号
 - 前往控制台申请用于激活EdgeBoard(FZ)专用SDK的序列号
 - 首次使用SDK或者更换序列号、更换设备时,需要联网激活。激活成功之后,有效期内可离线使用

效果展示

```
0 build > ./easyedge_demo ~/lvxiangxiang/models/SqueezeNetV1.1-tf/ ~/lvxiangxiang/images/orange.jpg
2020-07-16 16:52:33,632 DEBUG [EasyEdge] 548062597120 The recommended threshold is 0.3
2020-07-16 16:52:33,632 DEBUG [EasyEdge] 548062597120 Rescale mode is 0, target_size: 0, max_size: 0
2020-07-16 16:52:33,632 DEBUG [EasyEdge] 548062597120 Config file read done
2020-07-16 16:52:33,636 DEBUG [EasyEdge] 548062597120 Local license is ok.
2020-07-16 16:52:33,985 DEBUG [EasyEdge] 548062597120 Image will be resized to 227,227
2020-07-16 16:52:33,986 DEBUG [EasyEdge] 548062597120 Inference costs 131.011(127.854) ms
950, orange, p:0.972394
Done
```

【图像分类】0 AlexNet-fluid



abel	置信度
07747607 orange	0.951
07749582 lemon	0.030
n07716906 spaghetti squash	0.006
03942813 ping-pong ball	0.004
n07717556 butternut squash	0.002
n03134739 croquet ball	0.001
03929660 pick, plectrum,	
plectron	0.001
04332243 strainer	0.001
03530642 honeycomb	0.000
n12144580 com	0.000
04409515 tennis ball	0.000
n04019541 puck, hockey puck	0.000

环境准备

硬件环境 EdgeBoard-FZ系列计算盒,包括ZU3/ZU5/ZU9。

EdgeBoard-FZ系列硬件购买和详细硬件参数请前往AI市场。

使用SDK时,请保持EdgeBoard-FZ内核为最新,否则运行可能出现错误。

EdgeBoard-FZ计算盒使用手册:https://ai.baidu.com/ai-doc/HWCE/Yk3b86gvp

EdgeBoard-FZ内核更新地址:https://ai.baidu.com/ai-doc/HWCE/Yk3b95s8o **网络环境** 用户安装软件和测试SDK都需要联网,用户使用的 EdgeBoard-FZ需确保有效的网络连接。**软件环境** 需要在EdgeBoard-FZ计算盒上安装以下软件和第三方库以保证SDK正常编译和运行:

- cmake 3 +
- gcc 5.4 +
- opencv3.4 (可选)

可以使用以下方法确认cmake是否满足要求:

\$ cmake --version cmake version 3.13.3

若系统提示找不到cmake命令或者cmake version 低于3.x.x,则需要安装/升级cmake。

// apt安装cmake sudo apt update sudo apt install cmake

ubuntu官方的apt源update较慢,且可能访问不了,可以替换为国内的源:https://www.cnblogs.com/yongy1030/p/10315569.html

也可以使用源码编译的方式安装,参考安装方法。

// 升级cmake sudo apt-get install software-properties-common sudo add-apt-repository ppa:george-edison55/cmake-3.x sudo apt-get update sudo apt-get upgrade

安装/升级cmake后可再次执行cmake --version确认版本。

可以使用以下方法确认gcc是否满足要求:

\$ gcc --version gcc (Ubuntu/Linaro 5.4.0-6ubuntu1~16.04.12) 5.4.0 20160609

若系统提示找不到gcc命令或者gcc version 低于5.4.0,则需要安装/升级gcc。

// 安装gcc
sudo apt update
sudo apt install build-essential
// 升级gcc
sudo add-apt-repository ppa:ubuntu-toolchain-r/test # 如果找不到add-apt-repository命令,执行: apt-get install software-propertiescommon
sudo apt-get update
sudo apt-get install -y gcc-5 g++-5

cd /usr/bin # 升级gcc 5之后,还需要替换原来的软链接 sudo rm -r gcc # 移除之前的软连接 sudo ln -sf gcc-5 gcc # 建立gcc5的软连接 sudo rm -r g++ # 同gcc sudo ln -sf g++-5 g++ 目前没有合适的方法确认系统中是否有SDK需要的OpenCV,若用户不确定是否安装OpenCV 3.4 +,并且可以被cmake find_package到,可以手动编译安装OpenCV 3.4 ,也可以在之后编译SDK时自动编译OpenCV。

若选择在下一步编译EasyEdge SDK时自动编译OpenCV,则以下编译安装OpenCV的步骤可跳过。

下载OpenCV 3.4源代码包并解压:下载地址,然后编译安装:

```
// 编译安装OpenCV
cd opencv-3.4.6
mkdir build
cd build
```

cmake .. -DBUILD_DOCS=OFF -DBUILD_EXAMPLES=OFF -DBUILD_opencv_python2=OFF -DBUILD_opencv_python3=OFF -DBUILD_WITH_DEBUG_INFO=OFF -DBUILD_PACKAGE=OFF -DBUILD_opencv_core=ON -DBUILD_opencv_imgproc=ON -DBUILD_opencv_imgcodecs=ON -DBUILD_opencv_highgui=ON -DBUILD_opencv_video=OFF -DBUILD_opencv_videoio=OFF -DBUILD_opencv_dnn=OFF -DBUILD_opencv_apps=OFF -DBUILD_opencv_flann=OFF -DBUILD_opencv_gpu=OFF -DBUILD_opencv_mI=OFF -DBUILD_opencv_legacy=OFF -DBUILD_opencv_calib3d=OFF -DBUILD_opencv_features2d=OFF -DBUILD_opencv_java=OFF -DBUILD_opencv_objdetect=OFF -DBUILD_opencv_photo=OFF -DBUILD_opencv_nonfree=OFF -DBUILD_opencv_ol=OFF -DBUILD_opencv_stitching=OFF -DBUILD_opencv_superres=OFF -DBUILD_opencv_ts=OFF -DBUILD_opencv_videostab=OFF -DBUILD_opencv_contrib=OFF -DBUILD_OPEncv_superres=OFF -DBUILD_opEFF -DBUILD_opencv_videostab=OFF -DBUILD_opencv_contrib=OFF -DBUILD_SHARED_LIBS=ON -DBUILD_TESTS=OFF -DBUILD_PEFF_TESTS=OFF -DBUILD_WITH_CAROTENE=OFF -DCMAKE_BUILD_TYPE:STRING=Release -DWITH_FFMPEG=OFF -DWITH_IPP=OFF -DBUILD_PNG=ON -DBUILD_JPEG=ON -DBUILD_ZLIB=ON -DBUILD_FAT_JAVA_LIB=OFF -DOPENCV_CXX11=OFF -DCMAKE_INSTALL_PREFIX:PATH=/usr/lib/aarch64-linux·gnu/

make # 如果有多个cpu可以用-j加快编译速度,如4个CPU用 make -j4 make install

心启动EdgeBoard-FZ计算盒

一、将计算盒连接电源

指示灯亮起,等待约1分钟。

二、连接计算盒

参考EdgeBoard-FZ使用手册配置网口或串口连接,登录EdgeBoard-FZ计算盒。

三、加载驱动

开机加载一次即可。

insmod /home/root/workspace/driver/{zu9|zu5|zu3}/fpgadrv.ko

根据计算盒的版本(zu9/zu5/zu3)选择驱动。若未加载驱动,SDK可能报错:

Failed to to fpga device: -1

四、设置系统时间

```
系统时间必须正确。
```

date --set "2019-5-18 20:48:00"

の测试demo

SDK介绍 用户下载的Linux EdgeBoard-FZ SDK zip包中包含SDK动态库、模型等资源文件和测试demo.cpp。

需要将SDK zip包完整的放入EdgeBoard-FZ 硬件上再进行解压,否则可能会报错:

libeasyedge.so: file format not recognized; treating as linker script

新手指南

Linux下解压命令: tar -xvf xxx.tar、unzip xxx.zip

SDK zip包的目录结构如下:

解压baidu_easyedge_linux_cpp_aarch64_PADDLEMOBILE_FPGA_gcc5.4_vx.x.x_xxxxxx.tar之后的目录结构如下:



编译demo 前面安装了cmake、gcc等工具之后,可以编译SDK demo,生成测试的可执行文件。步骤如下。

```
一、将获取的序列号填入demo.cpp
```

```
cd demo # 进入demo文件夹
vi demo.cpp # 若vi未找到命令,执行 sudo apt install vim
```

在打开的代码编辑页面,找到

global_controller()->set_licence_key("set your license here");

```
将序列号填入引号内。如果想打印demo运行过程中的日志,找到
```

log_config.enable_debug = false;

将false改为true即可。

二、编译 在demo目录下执行

```
mkdir build # 创建build目录
cd build
cmake .. # 如果系统中安装了opencv3.4以上
##### 或者
cmake .. -DEDGE_BUILD_OPENCV=ON # 自动编译安装opencv
```

若用户需要自定义opencv library path、gcc路径等,修改CMakeList.txt即可。

当出现:

-- Configuring done -- Generating done -- Build files have been written to: /xxx/demo/build

表示cmake成功。然后执行编译

make # 如果有多个cpu可以用-j加快编译速度,如4个CPU用 make -j4

当出现:

[100%] Built target easyedge_serving [100%] Built target easyedge_demo

表示编译成功,在build目录下出现编译的产物:

- easyedge_demo :测试的可执行文件
- easyedge_serving:包含http server的测试的可执行文件
- thirdparty:编译安装的opencv

测试easyedge_demo 在build目录下执行:

./easyedge_demo {模型RES文件夹} {测试图片路径}

第一个参数为包含模型的文件夹路径,第二个参数为测试的图片的路径。SDK中已经包含模型文件夹,如果用户有其他模型文件,可以指定为其路径。如:

./easyedge_demo ../../../RES /xxx/test.jpg

然后可以看到输出的结果:

<pre>0 build > ./easyedge_demo ~/lvxiangxiang/models/SqueezeNetV1.1-tf/ ~/lvxiangxiang/images/orange.jpg</pre>	
2020-07-16 16:52:33,632 DEBUG [EasyEdge] 548062597120 The recommended threshold is 0.3	
2020-07-16 16:52:33,632 DEBUG [EasyEdge] 548062597120 Rescale mode is 0, target_size: 0, max_size: 0	
2020-07-16 16:52:33,632 DEBUG [EasyEdge] 548062597120 Config file read done	
2020-07-16 16:52:33,636 DEBUG [EasyEdge] 548062597120 Local license is ok.	如果是物体检测或者图像
2020-07-16 16:52:33,855 DEBUG [EasyEdge] 548062597120 Image will be resized to 227,227	
2020-07-16 16:52:33,986 DEBUG [EasyEdge] 548062597120 Inference costs 131.011(127.854) ms	
950, orange, p:0.972394	
Done	

分割模型,可以打开生成的/xxx/test.result.cpp.jpg图片,查看检测框的效果。

如果用户使用的是ZU5,且执行过程中出现内存不足:Killed。这是因为FZ5A带vcu,给它预留的内存过大导致,如果用不到VCU可以把这部分改小。修改/run/media/mmcblk1p1/uEnv.txt:

ethaddr=00:0a:35:00:00:09 uenvcmd=fatload mmc 1 0x3000000 image.ub && bootm 0x3000000

bootargs=earlycon console=ttyPS0,115200 clk_ignore_unused cpuidle.off=1 root=/dev/mmcblk1p2 rw rootwait cma=128M

注意中间空行要保留。

如果预测结果明显错误或者执行过程报错,请检查内核是否为最新版本。 测试easyedge_serving easyedge_serving会开启一个http server服务, 并实现了一个简单的网页,用户可以在网页上上传图片并查看预测结果。

在build目录下执行:

./easyedge_serving {模型RES文件夹} {序列号} {主机ip, 默认 0.0.0.0} {端口, 默认 24401}

如:

./easyedge_serving ../../../RES "1111-1111-1111-1111" 0.0.0.0 24401

若日志显示:

HTTP is now serving at 0.0.0.24401

表示http server启动成功。此时可以打开浏览器,输入网址http://{设备ip}:24401,上传图片来进行测试。

查看设备ip的方法:

ifconfig # 如果没有ifconfig命令,执行 sudo apt install net-tools

找到eth0一栏。

注意:只有本机电脑和硬件设备的网络ip在同一网段之下,才可以通过网址访问。



の 集成SDK

SDK提供了一系列模型加载、预测等接口,用户可以方便的集成进自己的程序之中。

接口说明、数据格式说明以及常见错误请参考SDK技术文档。

建议先测试Demo,以及参考demo.cpp和demo的CMakeLists.txt调用流程。如果遇到错误,优先参考文件中的注释以及日志说明。

一、导入SDK头文件和库文件

在baidu_easyedge_linux_cpp_aarch64_PADDLEMOBILE_FPGA_gcc5.4_vx.x.x_xxxxxx/include下有SDK的头文件。 在 baidu_easyedge_linux_cpp_aarch64_PADDLEMOBILE_FPGA_gcc5.4_vx.x.x_xxxxxx/lib下有SDK的库文件,包含动态库libeasyedge.so和静态库 libeasyedge_static.a。用户可选择合适的导入方式。

用户将头文件和库文件拷贝至自己的项目中,并在自己的CMakeLists.txt中引用:

```
find_package(OpenCV REQUIRED)

// 导入头文件
include_directories(
    ${OpenCV_INCLUDE_DIRS}
        ${CMAKE_SOURCE_DIR}/../include/
)

// 导入库文件
link_directories(
        ${CMAKE_SOURCE_DIR}/../lib/
)

// 链接库文件
target_link_libraries({your_executable_file} ${OpenCV_LIBS} easyedge paddle-mobile)
```

二、在程序中调用SDK接口

Baidu 百度智能云文档

// 引入SDK头文件
include "easyedge/easyedge.h"
// step 0: 设置序列号
global_controller()->set_licence_key("set your license here");
// step 1: 配置模型资源目录
PaddleFluidConfig config;
config.model_dir = {模型文件目录};
// step 2: 创建并初始化Predictor
auto predictor = global_controller()->CreateEdgePredictor(config);
auto img = cv::imread(()测试图片路径});
// step 3: 预测图像
std::vector<EdgeResultData> result;
predictor->infer(img, result);

目前EdgeBoard暂不支持并行多模型计算。

专项适配硬件Jetson

心简介

Jetson SDK是适用于EasyDL图像模型快速部署的工具包。SDK中包含了EasyDL训练的模型资源文件、SDK和demo文件。

测试前的准备

- Jetson(Nano/TX2/Xavier)硬件及开发环境
 - 详情参考下方文档
- EasyDL平台的Jetson专用SDK
 - 以图像分类为例,前往操作台训练「专项硬件适配SDK-Jetson」下的模型并发布SDK后,即可从平台下载
- 用于激活专用SDK的序列号
 - 前往控制台申请用于激活Jetson专用SDK的序列号
 - 首次使用SDK或者更换序列号、更换设备时,需要联网激活。激活成功之后,有效期内可离线使用
- 心 安装依赖

在使用SDK之前,首先要确认自己的硬件类型和相应的依赖库安装是否已经符合要求。

硬件要求:

Jetson系列开发板:

- Jetson Nano
- Jetson TX2
- Jetson Xavier NX
- Jetson Xavier

软件要求:

- JetPack4.2.2
- JetPack4.4

其他要求:

• 第一次使用SDK请确保联网

JetPack的安装需要借助SDK Manager,安装过程参考Install Jetson Software with SDK Manager。

对于Jetson Nano和Xavier NX还可以使用Etcher将系统镜像烧录到micro SD Card的形式,这种方式更简单一些。

Baidu 百度智能云文档

获取到的SDK解压后的目录结构是:



图片加载失败

其中:

- cpp文件夹下有两个压缩包分别为适配不同版本JetPackck的SDK,解压缩后得到include头文件、lib库文件和demo示例代码文件。
- RES文件夹下是EasyDL训练得到的模型资源文件。

®编译demo

解压SDK后,进入到demo目录可以直接编译demo。编译方法如下:

```
##### 1.首先进入demo目录
##### 2.创建build文件夹并进入
$ mkdir build && cd build
##### 3. 执行编译
##### 3. 执行编译
$ cmake ..
$ make -j3
##### 4. 执行安装,把ib文件安装到系统路径,需要sudo权限;也可以选择不执行安装,把ib路径加为环境变量即可
$ sudo make install
##### 5. 这时候应该会产出demo编译的可执行文件,直接执行可以查看需要的参数
$ ./easyedge_batch_inference {res_dir} {image/image_dir}
##### 或
$ ./easyedge_mult_thread {res_dir} {image/image_dir}
##### 或
$ ./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401}
```

编译产出有三个可执行文件,分别为:

- easyedge_batch_inference:提供了对批量图片预测的能力,预测速度快,是最推荐使用的方式。
- easyedge_multi_thread:提供了对多线程预测的支持。但是这种方式通常并不推荐使用,因为多线程下计算资源的相互竞争会拉慢单次预测的 速度并且预测时间不稳定,建议优先考虑使用batch inference的方式。
- easyedge_serving:提供了http服务的能力。启动服务后,可以在浏览器访问测试页面,或轻松把http服务接口集成到自己的应用中。

其中前两个demo示例了SDK API的集成方式,第三个demo示例了如何使用SDK创建http服务。根据个人需求选择不同集成方式即可,接下来将分别介绍这两种集成方式。

ல API集成

使用SDK的API接口方式能提供功能丰富、预测速度快的能力。API接口的设计尽可能的降低了调用复杂度,可以很方便的集成到自己的应用当中。

调用流程

```
// 1.设置序列号
global_controller()->set_licence_key("ABCD-ABCD-ABCD");
// 2.配置运行选项
TensorRTConfig config;
config.model_dir = argv[1];
config.max_batch_size = 4; //优化的模型可以支持的最大batch_size,实际单次推理的图片数不能大于此值
config.max_concurrency = 1; //设置device对应的卡可以使用的最大线程数
config.fp16 = false; //置true开启fp16模式推理会更快,精度会略微降低,但取决于硬件是否支持fp16,不是所有模型都支持fp16,
参阅文档
// 3. 创建predictor并初始化
auto predictor = global_controller()->CreateEdgePredictor(config);
predictor -> init();
// 4. 执行预测
predictor-> infer(imgs, result);
```

可以看到,调用SDK的API流程很简单,并可以通过config的不同配置实现不同的能力。比如支持批量图片预测、多线程预测、fp16加速等。实际 集成的过程中,最需要注意的就是config的配置,一个适合自己应用场景的config参数配置才能带来最佳的预测速度。 struct TensorRTConfig : public EdgePredictorConfig { std::string model_filename{"model"}; std::string params_filename{"params"}; std::string cache_name{"m_cache"}; /** * @brief GPU工作空间大小设置 * workspace_size = workspace_prefix * (1 << workspace_offset) * workspace_offset: 10 = KB, 20 = MB, 30 = GB */ int workspace_prefix{16}; int workspace_offset{20}; /** * @brief 生成最大 batch size 为 max batch size 的优化模型,单次预测图片数量可以小干或等干此值 */ int max batch size {1}; /** * @brief 设置使用哪张 GPU 卡 */ int device{0}: /** * @brief 模型编译等级 * 0:无论当前设置的max_batch_size是多少,仅使用历史编译产出(如果存在) * 1:如果当前max_batch_size与历史编译产出的max_batch_size不相等时,则重新编译模型(推荐) * 2:无论历史编译产出的max_batch_size为多少,均根据当前max_batch_size重新编译模型 */ int compile_level{1}; /** * @brief 设置device对应的卡可以支持的最大并发量 * 实际预测的时候对应卡的最大并发量不超过这里设置的范围 */ int max_concurrency{1}; /** * @brief 是否开启fp16模式预测,需要硬件支持 */ bool fp16{false}; /** * @brief 设置需要使用的DLA Core */ int dla core{-1};

<mark>cache_name</mark>:GPU加速版SDK首次加载模型会先对模型进行编译优化,通过此值可以设置优化后的产出文件名。比如,可以根据不同的配置来 命名产出文件,从而可以在之后的运行中直接通过文件名就可以索引到编译产出,而不用再次执行优化过程。

workspace_size:设置运行时可以被用来使用的最大临时显存。通常默认即可,但当执行预测失败时,可以适当调大此值。

max_batch_size</mark>:此值用来控制批量图片预测可以支持的最大图片数,实际预测的时候单次预测图片数不可大于此值,但可以是不大于此值的任 意图片数。建议设置为与实际需要批量预测的图片数量保持一致,以节省内存资源并可获得较高预测速度。

device:设置需要使用的 GPU 卡号。

compile_level:模型编译等级。通常模型的编译会比较慢,但编译产出是可以复用的。可以在第一次加载模型的时候设置合理的 max_batch_size 并在之后加载模型的时候直接使用历史编译产出。是否使用历史编译产出可以通过此值 compile_level 来控制,当此值为 0 时,表示忽略当前设 置的 max_batch_size 而仅使用历史产出(无历史产出时则编译模型);当此值为 1 时,会比较历史产出和当前设置的 max_batch_size 是否相 等,如不等,则重新编译;当此值为 2 时,无论如何都会重新编译模型。

max_concurrency:通过此值设置单张 GPU 卡上可以支持的最大 infer 并发量,其上限取决于硬件限制。init 接口会根据此值预分配 GPU 资源, 建议结合实际使用控制此值,使用多少则设置多少。注意:此值的增加会降低单次 infer 的速度且预测速度不稳定,建议优先考虑 batch inference。

fp16 : 默认是 fp32 模式,置 true 可以开启 fp16 模式预测,预测速度会有所提升,但精度也会略微下降,权衡使用。注意:不是所有模型都支持 fp16 模式,也不是所有硬件都支持 fp16 模式。已知不支持fp16的模型包括:EasyDL图像分类高精度模型。

运行demo 命令执行格式:

./easyedge_batch_inference {模型RES文件夹} {测试图片路径或图片文件夹路径} ./easyedge_multi_thread {模型RES文件夹} {测试图片路径或图片文件夹路径}

如:



预测效果展示:



^の Http服务集成

使用SDK的http服务API可以创建一个http服务端口,从而可以在应用里或者网页中通过http请求的方式执行模型的预测并获取预测结果。

调用流程



编译并运行后就会在控制台启动一个服务,默认地址为:http://{设备ip}:24401,通过在浏览器访问此地址可以打开一个测试页面,上传图片即 可获取识别结果。

运行demo demo命令执行格式

./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401}

如:



当出现提示:HTTP is now serving at 0.0.0.0:24401, holding 1 instances 时,表示服务已经启动成功,这时候可以在浏览器里输入地址打开一个 测试页面。如下所示:



国片加载大败

上传图片,可以看到预测效果:



图片加载失败

也可以通过在自己的项目中请求此地址做预测,并获取预测结果。URL中的get参数:

参数	診数 说明		默认值	
threshold	阈值过滤,	0~1	如不提供,则会使用模型的推荐阈值	

如使用Python的请求示例:

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
        'http://127.0.0.1:24401/',
        params={'threshold': 0.1},
        data=img).json()
```

http请求的返回格式为:

	字段	类型说明	其他
	error_code	Number	0为成功,非0参考message获得具体错误信息
	results	Array	内容为具体的识别结果。其中字段的具体含义请参考开发文档
	cost_ms	Number	预测耗时ms,不含网络交互时间

∞ 开发板使用技巧

查询L4T或JetPack版本 有时候我们可能不记得自己的板子刷的是哪个JetPack版本了,可以通过下面这条命令查询L4T的版本。

```
##### 在终端输入如下命令并回车
$ head -n 1 /etc/nv_tegra_release
##### 就会输出类似如下结果
$ # # R32 (release), REVISION: 4.3, GCID: 21589087, BOARD: t210ref, EABI: aarch64, DATE: Fri Jun 26 04:38:25 UTC 2020
```

从输出的结果来看,我的板子当前的L4T版本为R32.4.3,对应JetPack4.4。但是,L4T的版本不是JetPack的版本,不过一般可以从L4T的版本唯一对应到JetPack的版本,下面列出了最近几个版本的对应关系:

```
L4T R32.4.3 --> JetPack4.4
L4T R32.4.2 --> JetPack4.4DP
L4T R32.2.1 --> JetPack4.2.2
L4T R32.2.0 --> JetPack4.2.1
```

功率模式设置与查询不同的功率模式下,执行AI推理的速度是不一样的,如果对速度需求很高,可以把功率开到最大,但记得加上小风扇散热~

```
##### 1. 运行下面这条命令可以查询开发板当前的运行功率模式
$ sudo nypmodel -q verbose
##### $ NV Power Mode: MAXN
##### $ 0
###### 如果输出为MAXN代表是最大功率模式
###### 2. 若需要把功率调到最大,运行下面这条命令
$ sudo nypmodel -m 0
###### 如果你进入了桌面系统,也可以在桌面右上角有个按钮可以切换模式
##### 3. 查询资源利用率
$ sudo tegrastats
```

价格说明

EasyDL图像价格说明

EasyDL图像本地服务器部署价格说明

EasyDL图像支持本地服务器部署的任务类型包括:图像分类、物体检测、图像分割,只需在发布模型时提交本地服务器部署申请,通过审核后即 可获得一个月免费试用。

如免费试用期结束后希望购买可在控制台在线按设备使用年限购买授权。

Baidu 百度智能云文档

心 价格参考

部署类型	按年授权价格(每个模型每年在每个设备)	永久授权价格(每个模型在每个设备)
本地服务器SDK	10000元	50000元
本地服务器API	10000元	50000元

说明1:本地服务器SDK与本地服务器API在适配硬件上的评测信息已上线,点击查看性能对比:图像分类、物体检测、图像分割 **说明2**:本地服务器部署的授权粒度为每个模型每年每台设备,例如购买了2个模型3年在1台设备的授权,价格为2(个模型)x3(年)x1(台)x10000元=60000元。可在一台设备上激活2个有效期三年的模型。

EasyDL图像软硬一体方案价格说明

目前EasyDL图像提供基于图像分类、物体检测任务的多种软硬一体方案,请前往专题页面对比不同方案的性能与价格,选择与业务场景最匹配的 方案。

心 方案获取流程

- Step 1:在EasyDL训练专项适配硬件的图像分类/物体检测模型,迭代模型至效果满足业务需求
- Step 2:发布模型时选择专项适配硬件的专用SDK,并在AI市场购买方案
- Step 3:获得硬件和用于激活专用SDK的专用序列号,参考文档集成后,即可实现离线AI预测

如有其他硬件方案需求,请提交工单咨询。

EasyDL图像通用小型设备部署价格说明

EasyDL图像支持通用小型设备部署的任务类型包括:图像分类、物体检测、图像分割,每个模型发布设备端SDK后需通过序列号激活使用,每发 布一个模型即可申请2个测试序列号供3个月内免费试用。

如免费试用期结束后希望购买可在控制台在线按设备购买授权或按产品线购买。

心价格参考

按设备购买永久授权

授权终端设备数量	单设备授权价格(基础版SDK)	单设备授权价格 (加速版SDK)
1≤设备数≤10	300元	2000元
11≤设备数≤100	250元	1500元
101≤设备数	200元	1000元

说明1:加速版SDK目前已支持部分模型,点击查看基础版与加速版的性能对比:图像分类、物体检测、图像分割 说明2:设备授权的购买是「累计阶梯计费」,例如您购买了9个授权,因为这9个是在1~10的区间,所以付费300元/个×9个;当您继续购买 2个时,累计购买量已经达到了11个,则付费1×300+1×250,因为后1个已经在11~100的区间内

按产品线购买永久授权

单个序列号可激活设备数上限为1万台,目前已支持按年购买授权:基础版SDK每年10万元,加速版SDK每年30万元。

如有授权年限、设备上限数的更多需求,请提交工单联系我们。

说明1:按产品线授权的序列号适用于开发手机APP,序列号仅限在绑定的包名下使用 说明2:如同一模型需同时购买iOS和Android包名的授权,需按2个产品线购买

EasyDL图像价格常见问题

⊙ EasyDL图像常见问题

• 每个图像API有累计10000点的免费调用额度,如需付费使用,请在控制台进行线上购买

EasyDL图像公有云API价格说明

EasyDL图像支持发布为在线API的任务类型包括:图像分类、物体检测、图像分割

模型训练并发布为API后,可以在控制台看到已发布上线的所有公有云服务。

可以根据实际需求,开通「按量后付费」后,购买「调用点包」,高调用量下的优惠方案。也可以购买「QPS叠加包」,满足业务场景的高并发 需求。具体介绍见下方。

按量后付费

只需在智能云控制台「EasyDL图像」-「公有云服务」中找到需要付费使用的接口,点击开通付费,即可完成付费开通。立即开通

根据实际调用消耗的点数,系统每小时会对您的百度智能云账户进行扣费。1点=0.001元。

例如:物体检测-高性能模型对应发布的公有云服务,每次调用消耗5点,调用1000次对应消耗5000点,所以应付费5000*0.001=5元

免费额度

EasyDL定制化API服务都具有免费调用额度,开通付费后,免费调用额度仍保留,所有技术方向的API接口均有10000点免费额度

说明:成功调用与失败调用均消耗免费额度。

免费/付费对比

对于EasyDL各个能力,免费使用和开通付费后使用的配置有较大差异,具体对比如下:

状态	免费调用额度	超过调用额度	QPS限制
免费状态	拥有	不响应请求	不保证并发
付费状态	拥有	可继续请求	图像分类-高性能API服务保证10次并发;其余API保证4次并发

价目表

产品采用分段阶梯定价方式,调用单价按照自然月累积调用量所落阶梯区间而变化。每个月第一天清零上月累积的调用量,重新开始累积本月调用量。

未购买优惠商品的调用接口享受阶梯价格,按月计算:

月调用量(万点)	每"点"对应换算(元/点)
0<月消耗调用点量<=100	0.001
100<月消耗调用点量<=500	0.0008
500<月消耗调用点量	0.00064

不同模型单次调用消耗的点数说明:

注意:公有云服务价格由实际服务压测结果计算得出,同一种模型的不同训练配置可能导致最终价格不同。下表单价仅供参考,请以实际公 有云服务发布价格为准

模型名称	单次调用消耗点数(点/次)	对应实际价格 (元/次)
图像分类-AutoDL Transfer	6	0.006
图像分类-高性能	4	0.004
图像分类-高精度	4	0.004
物体检测-高性能	5	0.005
物体检测-高精度	16	0.016
物体检测·超高精度	17	0.017
图像分割-实例分割-高精度	14	0.014
图像分割-实例分割-高性能	60	0.060
图像分割语义分割高精度	32	0.032
图像分割语义分割高性能	8	0.008
图像分类·精度提升配置包-0-50ms	4	0.004
图像分类·精度提升配置包-50-100ms	4	0.004
图像分类·精度提升配置包-100-200ms	4	0.004
图像分类-精度提升配置包-200-300ms	4	0.005
图像分类-精度提升配置包-300-500ms	6	0.006
图像分类-精度提升配置包-500-1000ms	6	0.006
图像分类-精度提升配置包-1000+ms	10	0.01
物体检测-精度提升配置包-0-200ms	4	0.004
物体检测-精度提升配置包-200-300ms	5	0.005
物体检测-精度提升配置包-300-500ms	8	0.008
物体检测-精度提升配置包-500-1000ms	8	0.008
物体检测-精度提升配置包-1000-1500ms	12	0.012
物体检测·精度提升配置包-1500-2000ms	17	0.017
物体检测·精度提升配置包-2000-3000ms	26	0.026
物体检测-精度提升配置包-5000+ms	60	0.06

说明:调用失败不计费

费用举例

从2021-8-1至2021-8-31,本月某个公有云服务API接口的月消耗调用点量为200万点(已除去免费额度),费用如下: 前100万点落入0~100w阶梯,单价0.001元/点,费用为1000元 中间100万-200万点落入100w以上阶梯,单价0.0008元/点,费用为800元 本月费用共计:1800元

余额不足提醒与欠费处理

余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。

欠费处理

- 北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。
- 欠费后您开通付费的产品将进入欠费状态,只能使用每日的免费额度,超过额度的请求系统将不再响应,且不再保证并发处理。

心 调用点包
调用点包的本质可理解为"充值折扣",像「20万点」调用点包,按点与金钱换算规则,需要消耗200000*0.001=200元,但「20万点」调用点 包的目录价为180元。

调用点包有效期为12个月。计费调用优先抵扣调用点包额度,额度耗尽自动切换为按量后付费形式,具体计费标准如下:

调用点包点数 (万点)	价格(元)	对等折扣
20	180.00	0.90
50	430.00	0.86
100	800.00	0.80
300	2250.00	0.75
500	3500.00	0.70
1000	6500.00	0.65

_の QPS叠加包

已开通付费(或购买调用点包)后,若您有临时性的QPS高并发要求,可选择在某个时间段内叠加购买QPS。立即购买

QPS叠加包分为两种:

QPS叠加包种类	价格	说明
QPS叠加包	因公有云服务模型算法不同而不同	调用消耗仍将计算点数,对应计算费用
QPS叠加包(不限调用 量)	对比同类算法的普通QPS叠加包,价格会较高一 些	调用消耗不再计算点数,即无论调用量多少,公有云服务扣费不 变

具体计费标准可在公有云服务开通「按量后付费」后,在对应的QPS叠加包购买页面查看,不同公有云服务的QPS叠加包价格不一样。

EasyDL图像价格整体说明

本文档介绍EasyDL图像各项服务的价格

EasyDL旨在为开发者提供一站式AI开发体验,仅针对训练算力及部署两项内容收费。

算力收费

EasyDL图像提供付费算力,付费算力可用于模型训练以及批量预测功能,可根据实际需求购买算力使用时长。

价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失败、训练状态为排队中时长均不纳入收费时长。

部署收费

EasyDL图像在模型部署方面支持公有云API部署、私有服务器部署、设备端离线SDK部署、软硬一体方案四种方式,可根据业务场景需求选择具体部署方式,不同部署方式的计费说明如下:

• 公有云API计费

目前图像分类、物体检测、图像分割任务支持公有云API部署,API可在平台自助付费使用

● 私有服务器部署计费

目前服务器部署包支持线上按设备数量和模型数量及有效期购买使用授权

● 设备端离线SDK计费

目前设备端SDK已支持在线购买按设备使用授权、按产品线使用授权

• 软硬一体方案计费

目前已推出多种软硬一体方案,可在AI市场咨询购买

具体计费详情,请参考不同部署方式的计费文档。

EasyDL图像算力资源价格说明

∞ 算力资源价格说明

EasyDL提供了丰富的模型训练算法,同时在训练任务和批量预测任务上也提供多种机型自由选择。

说明:

为更好支持您的模型付费训练,平台针对您的模型创建数量、任务并行数均进行了权益升级,每位用户创建模型数量从30个提升至100个; 单模型组仅支持运行 1个训练任务提升至同时运行5个训练任务。

计费方式

计费规则如下:

- 按分钟计费,不足1分钟按百分比计算。
- 按小时扣费,即北京时间整点扣费并生成账单。出账单时间是当前计费周期结束后 1小时内。例如,10:00-11:00的账单会在12:00之前生成,具体以系统出账时间为准。
- 使用 EasyDL 前需保证账户无欠款。 计费公式 费用=计算设备单价×计算设备数×使用时长 时长计量方法:只包括模型训练时的统计时间,数 据预处理等不包括在计费时长内。

产品单价

模型训练 在EasyDL图像方向的任务配置过程中,您可以选择训练的设备以及设备数量。目前图像分类的高性能和AutoDL算法只支持单设备训 练。

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

说明1:扣费发生的时间点为任务训练结束(包含手动暂停训练或自动停止训练)后,如果因EasyDL系统异常导致训练任务运行失败,则相 应训练任务的全部耗时在账单中会做扣减,不会参与计费。

说明2:为确保训练任务的正常进行,建议您在开通付费后确保账户余额不低于100元。

说明1:扣费发生的时间点为任务结束后,如果因EasyDL系统异常导致训练任务运行失败,则相应训练任务的全部耗时在账单中会做扣减, 不会参与计费。

说明2:为确保训练任务的正常进行,建议您在开通付费后确保账户余额不低于100元。

算力资源包|可用模块(图像)|计算设备|时长|定价||--------|------|-------||批量预测|CPU4核_16G|50小时|70元/个| |批量预测|CPU_4核_16G|100小时|135元/个||批量预测|CPU_4核_16G|300小时|380元/个||模型训练、批量预测| TeslaGPU_P40_24G显存单卡_12核CPU_40G内存|50小时|1000元/个||模型训练、批量预测|TeslaGPU_P40_24G显存单卡_12核CPU_40G内 存|100小时|1900元/个||模型训练、批量预测|TeslaGPU_P40_24G显存单卡_12核CPU_40G内存|300小时|5300元/个||模型训练、批量 预测|TeslaGPU_V100_16G显存单卡_12核CPU_56G内存|50小时|1300元/个||模型训练、批量预测|TeslaGPU_V100_16G显存单卡_12核 CPU_56G内存|100小时|2400元/个||模型训练、批量预测|TeslaGPU_V100_16G显存单卡_12核CPU_56G内存|300小时|6700元/个||点击 此处](https://console.bce.baidu.com/ai/?=1612251100620&fromai=1#/ai/easydlLiteImage/buyHourPackage/index)立即前往购买

余额不足提醒与欠费处理

Baidu 百度智能云文档

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。

欠费处理

- 北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。
- 欠费后您开通付费的产品将进入欠费状态,只能使用免费额度,并且无法发起新的训练任务。

EasyDL文本价格说明

文本私有服务器部署价格说明

EasyDL文本支持本地服务器部署的任务类型包括:文本分类-单标签、文本分类-多标签、文本实体抽取、文本实体关系抽取、情感倾向分析以及 评论观点抽取,只需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用

如免费试用期结束后希望购买可在控制台在线按设备使用年限购买授权。

心 价格参考

部署类型	按年授权价格(每个模型每年在每个设备)	永久授权价格(每个模型在每个设备)
本地服务器SDK	10000元	50000元
本地服务器API	10000元	50000元

说明:本地服务器部署的授权粒度为每个模型每年每台设备,例如购买了2个模型3年在1台设备的授权,价格为2(个模型) x3(年) x1(台) x10000元=60000元。可在一台设备上激活2个有效期三年的模型。

EasyDL文本公有云API价格说明

EasyDL文本支持发布为在线API的任务类型包括:文本分类-单标签、文本分类-多标签、文本实体抽取、文本实体关系抽取、情感倾向分析、短文本相似度、评论观点抽取、ERNIE大模型创作

模型训练并发布为API后,可以在控制台看到已发布上线的所有公有云服务。

可以根据实际需求,开通「按量后付费」后,购买「调用点包」,高调用量下的优惠方案。也可以购买「QPS叠加包」,满足业务场景的高并发 需求。具体介绍见下方。

按量后付费

只需在智能云控制台「EasyDL文本」-「公有云服务」中找到需要付费使用的接口,点击开通付费,即可完成付费开通。立即开通

根据实际调用消耗的点数,系统每小时会对您的百度智能云账户进行扣费。1点=0.001元。

例如:**文本实体抽取-高精度模型**对应发布的公有云服务,每次调用消耗4点,调用1000次对应消耗4000点,所以应付费4000*0.001=4元

免费额度

EasyDL定制化API服务都具有免费调用额度,开通付费后,免费调用额度仍保留,所有技术方向的API接口均有10000点免费额度

说明:成功调用与失败调用均消耗免费额度。

免费/付费对比

对于EasyDL各个能力,免费使用和开通付费后使用的配置有较大差异,具体对比如下:

状态	免费调用额度	超过调用额度	QPS限制
免费状态	拥有	不响应请求	不保证并发
付费状态	拥有	可继续请求	文本分类API服务保证10次并发;其余文本API不保证调用并发

产品采用分段阶梯定价方式,调用单价按照自然月累积调用量所落阶梯区间而变化。月初,上月累积的调用量清零,重新开始累积本月调用量。

未购买优惠商品的调用接口享受阶梯价格,按月计算:

月调用量 (万点)	每"点"对应换算(元/点)
0<月消耗调用点量<=100	0.001
100<月消耗调用点量<=500	0.0008
500<月消耗调用点量	0.00064

不同模型单次调用消耗的点数说明:

模型名称	单次调用消耗点数(点/次)	对应实际价格 (元/次)
文本分类-单标签(短文本)-高精度	8	0.008
文本分类-单标签(短文本)-高性能	4	0.004
文本分类-单标签(多语种)-高精度	8	0.008
文本分类-多标签-高精度	8	0.008
文本分类-多标签-高性能	4	0.004
文本实体抽取-高精度	4	0.004
文本实体抽取-高性能	4	0.004
文本实体关系抽取-高精度	4	0.004
情感倾向分析-高精度	8	0.008
情感倾向分析-高性能	4	0.004
短文本相似度-高精度	8	0.008
短文本相似度-高性能	4	0.004
评论观点抽取-高精度	9	0.009
评论观点抽取-高性能	4	0.004
文本创作	2500	2.5

说明1:调用失败不计费

说明2: 文本创作的公有云服务在邀测期间单独提供200万点的免费额度

费用举例

从2021-8-1至2021-8-31,本月某个公有云服务API接口的月消耗调用点量为200万点(已除去免费额度),费用如下: 前100万点落入0~100w阶梯,单价0.001元/点,费用为1000元 中间100万-200万点落入100w以上阶梯,单价0.0008元/点,费用为800元 本月费用共计:1800元

余额不足提醒与欠费处理

余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充<u>值</u>。

欠**费处**理

- 北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。
- 欠费后您开通付费的产品将进入欠费状态,只能使用每日的免费额度,超过额度的请求系统将不再响应,且不再保证并发处理。

∞ 调用点包

开通按量后付费后即可购买调用点包。立即购买

包的目录价为180元,相当于给客户打了9折。

调用点包有效期为12个月。计费调用优先抵扣调用点包额度,额度耗尽**自动切换为按量后付费形式**,具体计费标准如下:

调用点包点数(万点)	价格 (元)	对等折扣
20	180.00	0.90
50	430.00	0.86
100	800.00	0.80
300	2250.00	0.75
500	3500.00	0.70
1000	6500.00	0.65

の QPS叠加包

已开通付费(或购买调用点包)后,若您有临时性的QPS高并发要求,可选择在某个时间段内叠加购买QPS。立即购买

QPS叠加包分为两种:

QPS叠加包种类	价格	说明
QPS叠加包	因公有云服务模型算法不同而不同	调用消耗仍将计算点数,对应计算费用
QPS叠加包(不限调用 量)	对比同类算法的普通QPS叠加包,价格会较高一 些	调用消耗不在计算点数,即无论调用量多少,公有云服务扣费不 变

具体计费标准可在公有云服务开通「按量后付费」后,在对应的QPS叠加包购买页面查看,不同公有云服务的QPS叠加包价格不一样。

EasyDL文本价格整体说明

本文档介绍EasyDL文本各项服务的价格

EasyDL旨在为开发者提供一站式AI开发体验,仅针对训练算力及部署两项内容收费。

算力收费

EasyDL文本提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。 (其中文本创作操作台作为新推出任务方向,会持 续为开发者提供免费训练体验)

其余操作台算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失败、训练状态为排队中时长均不纳入收费时长。

部署收费

EasyDL文本提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长

EasyDL文本在模型部署方面支持公有云API部署、私有服务器部署两种方式,可根据业务场景需求选择具体部署方式,不同部署方式的计费说明如下:

• 公有云API部署计费

文本分类-单标签、文本实体抽取、文本实体关系抽取、情感倾向分析、短文相似度、评论观点抽取任务支持公有云API部署,可通过公有云API可 在平台自助付费使用

• 私有服务器部署计费

目前服务器部署包支持线上按设备数量和模型数量及有效期购买使用授权

具体计费详情,请参考不同部署方式的计费文档。

EasyDL文本算力价格说明

∞ 算力资源价格说明

EasyDL提供了丰富的模型训练算法,同时在训练任务上也提供多种机型自由选择。

说明:

为更好支持您的模型付费训练,平台针对您的模型创建数量、任务并行数均进行了权益升级,每位用户创建模型数量从30个提升至100个; 单模型组仅支持运行 1个训练任务提升至同时运行5个训练任务。

计费方式

计费规则如下:

- 按分钟计费,不足1分钟按百分比计算。
- 按小时扣费,即北京时间整点扣费并生成账单。出账单时间是当前计费周期结束后 1小时内。例如,10:00-11:00的账单会在12:00之前生成,具体以系统出账时间为准。
- 使用 EasyDL 前需保证账户无欠款。 计费公式 费用=计算设备单价×计算设备数×使用时长 时长计量方法:只包括模型训练时的统计时间,数 据预处理等不包括在计费时长内。

产品单价

在EasyDL文本方向的任务配置过程中,您可以选择训练的设备以及设备数量。

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

说明1:扣费发生的时间点为任务训练结束(包含手动暂停训练或自动停止训练)后,如果因EasyDL系统异常导致训练任务运行失败,则相 应训练任务的全部耗时在账单中会做扣减,不会参与计费。

说明2:为确保训练任务的正常进行,建议您在开通付费后确保账户余额不低于100元。

算力资源包|可用模块(文本)|计算设备|时长|定价||--------|-------|--------|-|模型训练|TeslaGPU_V100_16G显存单卡_12核CPU_56G内存|50小时|1300元/个||模型训练|TeslaGPU_V100_16G显存单卡_12核CPU_56G内存|100小时|2400元/个||模型训练| TeslaGPU_V100_16G显存单卡_12核CPU_56G内存|300小时|6700元/个|

余额不足提醒与欠费处理

余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。

欠费处理

- 北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。
- 欠费后您开通付费的产品将进入欠费状态,只能使用免费额度,并且无法发起新的训练任务。

EasyDL结构化数据价格说明

EasyDL结构化数据公有云API价格说明

EasyDL结构化数据支持发布为在线API的任务类型包括:表格数据预测、时序预测

模型训练并发布为API后,可以在控制台看到已发布上线的所有公有云服务。

可以根据实际需求,开通「按量后付费」后,购买「调用点包」,高调用量下的优惠方案。也可以购买「QPS叠加包」,满足业务场景的高并发

需求。具体介绍见下方。

按量后付费

只需在智能云控制台「EasyDL结构化数据」-「公有云服务」中找到需要付费使用的接口,点击开通付费,即可完成付费开通。立即开通 根据实际调用消耗的点数,系统每小时会对您的百度智能云账户进行扣费。1点=0.001元。

例如:**结构化数据-时序预测**对应发布的公有云服务,每次调用消耗5点,调用1000次对应消耗5000点,所以应付费5000*0.001=5元

免费额度

EasyDL定制化API服务都具有免费调用额度,开通付费后,免费调用额度仍保留,所有技术方向的API接口均有10000点免费额度

注:成功调用与失败调用均消耗免费额度。

免费/付费对比

对于EasyDL各个能力,免费使用和开通付费后使用的配置有较大差异,具体对比如下:

状态	免费调用额度	超过调用额度	QPS限制
免费状态	拥有	不响应请求	不保证并发
付费状态	拥有	可继续请求	时序预测API服务保证4次并发;其余API不保证并发

价目表

产品采用分段阶梯定价方式,调用单价按照自然月累积调用量所落阶梯区间而变化。月初,上月累积的调用量清零,重新开始累积本月调用量。

月调用量(万点)	每"点"对应换算(元/点)
0<月消耗调用点量<=100	0.001
100<月消耗调用点量<=500	0.0008
500<月消耗调用点量	0.00064

不同模型单次调用消耗的点数说明

模型名称	单次调用消耗点数(点/次)	对应实际价格(元/次)
结构化数据-表格数据预测	4	0.004
结构化数据-时序预测	5	0.005

说明:调用失败不计费

费用举例

从2021-8-1至2021-8-31,本月某个公有云服务API接口的月消耗调用点量为200万点(已除去免费额度),费用如下: 前100万点落入0~100w阶梯,单价0.001元/点,费用为1000元 中间100万-200万点落入100w以上阶梯,单价0.0008元/点,费用为800元 本月费用共计:1800元

余额不足提醒与欠费处理

余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。

欠**费处**理

北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。

• 欠费后您开通付费的产品将进入欠费状态,只能使用每日的免费额度,超过额度的请求系统将不再响应,且不再保证并发处理。

心 调用点包

开通按量后付费后即可购买调用点包。立即购买

调用点包的本质可理解为"充值折扣",像「20万点」调用点包,按点与金钱换算规则,需要消耗200000*0.001=200元,但「20万点」调用点 包的目录价为180元,相当于给客户打了9折。

调用点包有效期为12个月。计费调用优先抵扣调用点包额度,额度耗尽**自动切换为按量后付费形式**,具体计费标准如下:

调用点包点数(万点)	价格 (元)	对等折扣
20	180.00	0.90
50	430.00	0.86
100	800.00	0.80
300	2250.00	0.75
500	3500.00	0.70
1000	6500.00	0.65

心 QPS叠加包

已开通付费(或购买调用点包)后,若您有临时性的QPS高并发要求,可选择在某个时间段内叠加购买QPS。立即购买

QPS叠加包分为两种:

QPS叠加包种类	价格	说明
QPS叠加包	因公有云服务模型算法不同而不同	调用消耗仍将计算点数,对应计算费用
QPS叠加包(不限调用 量)	对比同类算法的普通QPS叠加包,价格会较高一 些	调用消耗不在计算点数,即无论调用量多少,公有云服务扣费不 变

具体计费标准可在公有云服务开通「按量后付费」后,在对应的QPS叠加包购买页面查看,不同公有云服务的QPS叠加包价格不一样

表格预测算力资源价格说明

EasyDL结构化数据价格整体说明

本文档介绍EasyDL结构化数据服务的价格

EasyDL旨在为开发者提供一站式AI开发体验,仅针对训练算力及部署两项内容收费。

算力收费

EasyDL结构化数据提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失败、训练状态为排队中时长均不纳入收费时长。

部署收费

EasyDL结构化数据在模型部署上支持公有云API部署

模型效果满意后,可将模型部署在公有云服务器上,公有云部署免费调用次数为10000点/接口,调用点数可在平台自主付费购买使用

EasyDL表格预测在模型部署上还支持服务器、通用小型设备离线部署

• 私有服务器部署计费

目前服务器部署包支持线上按设备数量和模型数量及有效期购买使用授权

● 设备端离线SDK计费

目前设备端SDK已支持在线购买按设备使用授权、按产品线使用授权

具体计费详情,请参考不同部署方式的计费文档

EasyDL视频价格说明

EasyDL视频公有云API价格说明

EasyDL视频-视频分类支持发布为在线API

模型训练并发布为API后,可以在控制台看到已发布上线的所有公有云服务。

可以根据实际需求,开通「按量后付费」后,购买「调用点包」,高调用量下的优惠方案。也可以购买「QPS叠加包」,满足业务场景的高并发 需求。具体介绍见下方。

按量后付费

只需在智能云控制台「EasyDL视频」-「公有云服务」中找到需要付费使用的接口,点击开通付费,即可完成付费开通。立即开通

根据实际调用消耗的点数,系统每小时会对您的百度智能云账户进行扣费。1点=0.001元。

例如:视频分类对应发布的公有云服务,每次调用消耗15点,调用1000次对应消耗15000点,所以应付费15000*0.001=15元

免费额度

EasyDL定制化API服务都具有免费调用额度,开通付费后,免费调用额度仍保留,所有技术方向的API接口均有10000点免费额度

说明:成功调用与失败调用均消耗免费额度。

免费/付费对比

对于EasyDL各个能力,免费使用和开通付费后使用的配置有较大差异,具体对比如下:

状态	免费调用额度	超过调用额度	QPS限制
免费状态	拥有	不响应请求	不保证并发
付费状态	拥有	可继续请求	视频分类API服务保证4次并发

价目表

产品采用分段阶梯定价方式,调用单价按照自然月累积调用量所落阶梯区间而变化。每个月第一天上月累积的调用量清零,重新开始累积本月调 用量。

未购买优惠商品的调用接口享受阶梯价格,按月计算:

月调用量 (万点)	每"点"对应换算(元/点)
0<月消耗调用点量<=100	0.001
100<月消耗调用点量<=500	0.0008
500<月消耗调用点量	0.00064

不同模型单次调用消耗的点数说明

模型名称	单次调用消耗点数(点/次)	对应实际价格 (元/次)
视频分类	15	0.015

说明:调用失败不计费

费用举例

从2021-8-1至2021-8-31,本月某个公有云服务API接口的月消耗调用点量为200万点(已除去免费额度),费用如下:

前100万点落入0~100w阶梯,单价0.001元/点,费用为1000元 中间100万-200万点落入100w以上阶梯,单价0.0008元/点,费用为800元 本月费用共计:1800元

余额不足提醒与欠费处理

余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。

欠费处理

- 北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。
- 欠费后您开通付费的产品将进入欠费状态,只能使用每日的免费额度,超过额度的请求系统将不再响应,且不再保证并发处理。

心 调用点包

开通按量后付费后即可购买调用点包。立即购买

调用点包的本质可理解为"充值折扣",像「20万点」调用点包,按点与金钱换算规则,需要消耗200000*0.001=200元,但「20万点」调用点 包的目录价为180元,相当于给客户打了9折。

调用点包有效期为12个月。计费调用优先抵扣调用点包额度,额度耗尽**自动切换为按量后付费形式**,具体计费标准如下:

调用点包点数(万点)	价格(元)	对等折扣
20	180.00	0.90
50	430.00	0.86
100	800.00	0.80
300	2250.00	0.75
500	3500.00	0.70
1000	6500.00	0.65

心 QPS叠加包

已开通付费(或购买调用点包)后,若您有临时性的QPS高并发要求,可选择在某个时间段内叠加购买QPS。立即购买

QPS叠加包分为两种:

QPS叠加包种类	价格	说明
QPS叠加包	因公有云服务模型算法不同而不同	调用消耗仍将计算点数,对应计算费用
QPS叠加包(不限调用 量)	对比同类算法的普通QPS叠加包,价格会较高一 些	调用消耗不在计算点数,即无论调用量多少,公有云服务扣费不 变

具体计费标准可在公有云服务开通「按量后付费」后,在对应的QPS叠加包购买页面查看,不同公有云服务的QPS叠加包价格不一样。

EasyDL视频本地服务器部署价格说明

EasyDL视频·目标跟踪支持本地服务器部署,在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。

如免费试用期结束后希望购买可在控制台在线按设备使用年限购买授权。

心价格参考

部署类型	按年授权价格(每个模型每年在每个设备)	永久授权价格(每个模型在每个设备)
本地服务器SDK	10000元	50000元

说明:本地服务器部署的授权粒度为每个模型每年每台设备,例如购买了2个模型3年在1台设备的授权,价格为2(个模型)x3(年)x1(台)x10000元=60000元。可在一台设备上激活2个有效期三年的模型。

EasyDL视频设备端部署价格说明

EasyDL视频-目标跟踪支持将定制模型部署在设备端上,每个模型发布设备端SDK后需通过序列号激活使用,每发布一个模型即可申请2个测试序 列号供3个月内免费试用。

目前已支持在控制台在线按设备购买授权或按产品线购买。

心 价格参考

按设备购买永久授权

授权终端设备数量	单设备授权价格(基础版SDK)	单设备授权价格(加速版SDK)
1≤设备数≤10	300元	2000元
11≤设备数≤100	250元	1500元
101≤设备数	200元	1000元

说明:设备授权的购买是「累计阶梯计费」,例如您购买了9个授权,因为这9个是在1~10的区间,所以付费300元/个×9个;当您继续购买2 个时,累计购买量已经达到了11个,则付费1×300+1×250,因为后1个已经在11~100的区间内

按产品线购买永久授权

单个序列号可激活设备数上限为1万台,目前已支持按年购买授权:基础版SDK每年10万元,加速版SDK每年30万元。

如有授权年限、设备上限数的更多需求,请提交工单联系我们。

说明:按产品线授权的序列号适用于开发手机APP,序列号仅限在绑定的包名下使用

EasyDL视频软硬一体方案价格说明

目前EasyDL视频提供基于目标跟踪的软硬一体方案,请前往专题页面对比不同方案的性能与价格,选择与业务场景最匹配的方案。

っ方案获取流程

- Step 1:在EasyDL训练专项适配硬件的目标跟踪模型,迭代模型至效果满足业务需求
- Step 2:发布模型时选择专项适配硬件的专用SDK,并在AI市场购买方案
- Step 3:获得硬件和用于激活专用SDK的专用序列号,参考文档集成后,即可实现离线AI预测

如有其他硬件方案需求,请提交工单咨询。

EasyDL视频价格整体说明

本文档介绍EasyDL视频各项服务的价格

EasyDL旨在为开发者提供一站式AI开发体验,仅针对训练算力及部署两项内容收费。

算力收费

EasyDL视频提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失败、训练状态为排队中时长均不纳入收费时长。

EasyDL视频模型部署上支持公有云API部署、私有服务器部署、设备端离线SDK部署、软硬一体方案四种方式,可根据业务场景需求选择具体部署方式,不同部署方式的计费说明如下:

• 公有云API计费

EasyDL视频-视频分类任务支持公有云API部署,可在平台自助付费使用

• 私有服务器部署计费

EasyDL视频-目标跟踪任务支持服务器部署,目前服务器部署包支持线上按设备数量和模型数量及有效期购买使用授权

• 设备端离线SDK计费

EasyDL视频·目标跟踪任务支持设备端SDK部署,目前设备端SDK已支持在线购买按设备使用授权、按产品线使用授权

• 软硬一体方案计费

EasyDL视频-目标跟踪任务目前已推出多种软硬一体方案,可在AI市场咨询购买

具体计费详情,请参考不同部署方式的计费文档。

EasyDL视频算力资源价格说明

⊙ 算力资源价格说明

EasyDL提供了丰富的模型训练算法,同时在训练任务上也提供多种机型自由选择。

说明:

为更好支持您的模型付费训练,平台针对您的模型创建数量、任务并行数均进行了权益升级,每位用户创建模型数量从30个提升至100个; 单模型组仅支持运行 1个训练任务提升至同时运行5个训练任务。

计费方式

具体计费规则如下:

- 按分钟计费,不足1分钟按百分比计算。
- 按小时扣费,即北京时间整点扣费并生成账单。出账单时间是当前计费周期结束后 1小时内。例如,10:00-11:00的账单会在12:00之前生成,具体以系统出账时间为准。
- 使用 EasyDL 前需保证账户无欠款。 计费公式 费用=计算设备单价×计算设备数×使用时长 时长计量方法:只包括模型训练时的统计时间,数 据预处理等不包括在计费时长内。

产品单价

模型训练 在EasyDL视频方向的任务配置过程中,您可以选择训练的设备以及设备数量。

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	Τ	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

说明1:扣费发生的时间点为任务训练结束(包含手动暂停训练或自动停止训练)后,如果因EasyDL系统异常导致训练任务运行失败,则相应训练任务的全部耗时在账单中会做扣减,不会参与计费。

说明2:为确保训练任务的正常进行,建议您在开通付费后确保账户余额不低于100元。

余额不足提醒与欠费处理

余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。 欠**费处**理

- 北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。
- 欠费后您开通付费的产品将进入欠费状态,只能使用免费额度,并且无法发起新的训练任务。

EasyDL语音价格说明

EasyDL语音公有云API价格说明

EasyDL语音-声音分类支持发布为在线API

模型训练并发布为API后,可以在控制台看到已发布上线的所有公有云服务。

可以根据实际需求,开通「按量后付费」后,购买「调用点包」,高调用量下的优惠方案。也可以购买「QPS叠加包」,满足业务场景的高并发 需求。具体介绍见下方。

按量后付费

只需在智能云控制台「EasyDL语音」-「公有云服务」中找到需要付费使用的接口,点击开通付费,即可完成付费开通。立即开通

根据实际调用消耗的点数,系统每小时会对您的百度智能云账户进行扣费。1点=0.001元。

例如:声音分类对应发布的公有云服务,每次调用消耗4点,调用1000次对应消耗4000点,所以应付费4000*0.001=4元

免费额度

EasyDL定制化API服务都具有免费调用额度,开通付费后,免费调用额度仍保留,所有技术方向的API接口均有10000点免费额度

说明:成功调用与失败调用均消耗免费额度。

免费/付费对比

对于EasyDL各个能力,免费使用和开通付费后使用的配置有较大差异,具体对比如下:

状态	免费调用额度	超过调用额度	QPS限制
免费状态	拥有	不响应请求	不保证并发
付费状态	拥有	可继续请求	声音分类API服务保证4次并发

价目表

产品采用分段阶梯定价方式,调用单价按照自然月累积调用量所落阶梯区间而变化。月初,上月累积的调用量清零,重新开始累积本月调用量。

未购买优惠商品的调用接口享受阶梯价格,按月计算:

月调用量(万点)	每"点"对应换算(元/点)
0<月消耗调用点量<=100	0.001
100<月消耗调用点量<=500	0.0008
500<月消耗调用点量	0.00064

不同模型单次调用消耗的点数说明:

模型名称	单次调用消耗点数(点/次)	对应实际价格 (元/次	()
声音分类	4	0.004	

说明:调用失败不计费

费用举例

从2021-8-1至2021-8-31,本月某个公有云服务API接口的月消耗调用点量为200万点(已除去免费额度),费用如下: 前100万点落入0~100w阶梯,单价0.001元/点,费用为1000元 中间100万-200万点落入100w以上阶梯,单价0.0008元/点,费用为800元 本月费用共计:1800元

余额不足提醒与欠费处理

余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。

欠**费处**理

- 北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。
- 欠费后您开通付费的产品将进入欠费状态,只能使用每日的免费额度,超过额度的请求系统将不再响应,且不再保证并发处理。

心 调用点包

开通按量后付费后即可购买调用点包。立即购买

调用点包的本质可理解为"充值折扣",像「20万点」调用点包,按点与金钱换算规则,需要消耗200000*0.001=200元,但「20万点」调用点 包的目录价为180元,相当于给客户打了9折。

调用点包有效期为12个月。计费调用优先抵扣调用点包额度,额度耗尽**自动切换为按量后付费形式**,具体计费标准如下:

调用点包点数(万点)	价格 (元)	对等折扣
20	180.00	0.90
50	430.00	0.86
100	800.00	0.80
300	2250.00	0.75
500	3500.00	0.70
1000	6500.00	0.65

の QPS叠加包

已开通付费(或购买调用点包)后,若您有临时性的QPS高并发要求,可选择在某个时间段内叠加购买QPS。立即购买

QPS叠加包分为两种:

QPS叠加包种类	价格	说明
QPS叠加包	因公有云服务模型算法不同而不同	调用消耗仍将计算点数,对应计算费用
QPS叠加包(不限调用 量)	对比同类算法的普通QPS叠加包,价格会较高一 些	调用消耗不在计算点数,即无论调用量多少,公有云服务扣费不 变

具体计费标准可在公有云服务开通「按量后付费」后,在对应的QPS叠加包购买页面查看,不同公有云服务的QPS叠加包价格不一样。

EasyDL语音本地服务器部署价格说明

EasyDL语音·声音分类支持本地服务器部署,只需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。

目前已支持在控制台在线按设备使用年限购买授权。

心 价格参考

部署类型	按年授权价格(每个模型每年在每个设备)	永久授权价格(每个模型在每个设备)
本地服务器SDK	10000元	50000元
本地服务器API	10000元	50000元

说明1:本地服务器SDK与本地服务器API在适配硬件上的评测信息已上线,点击查看性能对比:声音分类 说明2:本地服务器部署的授权粒度为每个模型每年每台设备,例如购买了2个模型3年在1台设备的授权,价格为2(个模型)x3(年)x1(台)x10000元=60000元。可在一台设备上激活2个有效期三年的模型。

EasyDL语音本地设备端部署价格说明

EasyDL语音·声音分类支持通用小型设备部署,每个模型发布设备端SDK后需通过序列号激活使用,每发布一个模型即可申请2个测试序列号供3 个月内免费试用。

如免费试用期结束后希望购买可在控制台在线按设备购买授权或按产品线购买。

心 价格参考

按设备购买永久授权

授权终端设备数量	单设备授权价格 (基础版SDK)	单设备授权价格(加速版SDK)
1≤设备数≤10	300元	2000元
11≤设备数≤100	250元	1500元
101≤设备数	200元	1000元

```
说明1:加速版SDK目前已支持部分模型,点击查看基础版与加速版的性能对比:声音分类
```

说明2:设备授权的购买是「累计阶梯计费」,例如您购买了9个授权,因为这9个是在1~10的区间,所以付费300元/个×9个;当您继续购买 2个时,累计购买量已经达到了11个,则付费1×300+1×250,因为后1个已经在11~100的区间内

按产品线购买永久授权

单个序列号可激活设备数上限为1万台,目前已支持按年购买授权:基础版SDK每年10万元,加速版SDK每年30万元。

如有授权年限、设备上限数的更多需求,请提交工单联系我们。

说明1:按产品线授权的序列号适用于开发手机APP,序列号仅限在绑定的包名下使用 说明2:如同一模型需同时购买iOS和Android包名的授权,需按2个产品线购买

EasyDL语音价格整体说明

本文档介绍EasyDL语音各项服务的价格

EasyDL旨在为开发者提供一站式AI开发体验,仅针对训练算力及部署两项内容收费。

算力收费

EasyDL图像提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。 (其中,语音识别由于任务特殊性,将持续为开发者提供免费训练体验)

声音分类价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失败、训练状态为排队中时长均不纳入收费时长。

部署收费

EasyDL语音在模型部署方面支持公有云API部署、私有服务器部署、设备端离线SDK部署三种方式,可根据业务场景需求选择具体部署方式,不同部署方式的计费说明如下:

• 公有云API计费

目前声音分类API可在平台自助付费使用,语音识别在语音技术下任一接口购买量包、或者开通后付费的用户,可免费使用EasyDL语音-语音识 别训练模型 • 私有服务器部署计费

目前服务器部署包支持线上按设备数量和模型数量及有效期购买使用授权

● 设备端离线SDK计费

目前设备端SDK已支持在线购买按设备使用授权、按产品线使用授权

具体计费详情,请参考不同部署方式的计费文档

EasyDL语音算力资源价格说明

⊙ 算力资源价格说明

EasyDL提供了丰富的模型训练算法,同时在训练任务上也提供多种机型自由选择。(其中,语音识别由于任务特殊性,将持续为开发者提供免费 训练体验)

说明:

为更好支持您的模型付费训练,平台针对您的模型创建数量、任务并行数均进行了权益升级,每位用户创建模型数量从30个提升至100个; 单模型组仅支持运行 1个训练任务提升至同时运行5个训练任务。

计费方式

计费规则如下:

- 按分钟计费,不足1分钟按百分比计算。
- 按小时扣费,即北京时间整点扣费并生成账单。出账单时间是当前计费周期结束后1小时内。例如,10:00-11:00的账单会在12:00之前生成,具体以系统出账时间为准。
- 使用 EasyDL 前需保证账户无欠款。 计费公式 费用=计算设备单价×计算设备数×使用时长 时长计量方法:只包括模型训练时的统计时间,数 据预处理等不包括在计费时长内。

产品单价

模型训练 在EasyDL声音方向的任务配置过程中,您可以选择训练的设备以及设备数量。

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

说明1:扣费发生的时间点为任务训练结束(包含手动暂停训练或自动停止训练)后,如果因EasyDL系统异常导致训练任务运行失败,则相 应训练任务的全部耗时在账单中会做扣减,不会参与计费。

说明2:为确保训练任务的正常进行,建议您在开通付费后确保账户余额不低于100元。

余额不足提醒与欠费处理

余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。

欠费处理

- 北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。
- 欠费后您开通付费的产品将进入欠费状态,只能使用免费额度,并且无法发起新的训练任务。

EasyDL零售行业版价格说明 价格整体说明 心简介

本文档介绍EasyDL零售版各项服务的价格,EasyDL零售版提供两种服务,分别为定制模型服务、货架拼接服务。其中定制模型服务支持云服务 API和本地服务器部署两种服务方式,货架拼接服务支持云服务API。

云服务的计费方式包括:**按调用量后付费、调用量次数包预付费**和**QPS叠加包预付费**三种,根据实际购买的项目进行付费。其中**按调用量后付** 费方式,系统根据实际调用的次数,每小时对您的百度云账户进行扣费。如果需要本地服务器部署,请加入EasyDL零售版官方QQ群(群号: 1009661589)联系群管咨询。

心 计费方式介绍

EasyDL零售版的相关服务和接口计费方式如下:

服务名 称	服务方 式	是否计 费	计费方式	免费额度
定制模 型服务	云服务 API	计费	按调用量后付费和QPS叠加 包预付费	每个模型发布的云服务API享有1000次免费调用量,超出免费额度的每次调用量根据选择的模型服务功能收取不同的费用,详情请见价格说明文档
翻拍识 别服务	云服务 API	计费	按调用量后付费、调用量次 数包和QPS叠加包	每账号享有累计免费1000次,超出部分按调用量计费,超出免费额度的每次调用量根据 选择的模型服务功能收取不同的费用,详情请见价格说明文档
货架拼 接服务	云服务 API	计费	按任务数后付费、任务次数 包和并发任务叠加包	累计免费200次拼接任务,超出免费额度的每次调用量根据选择的模型服务功能收取不同的费用,详情请见价格说明文档

免费额度用尽后,请在百度云EasyDL零售版控制台开通计费,开通后即可开始计费使用,计费方式请参考公有云API价格说明文档。

上述服务中,定制商品检测服务支持本地服务器部署,如需要这种方式,请加入EasyDL零售版官方QQ群(群号:1009661589)联系群管咨询。

公有云API价格说明

心 开通付费及购买服务

EasyDL零售版的各项服务的「开通付费」、「购买QPS叠加包」、「购买调用量次数包」、「关闭付费」等操作均在EasyDL零售版控制台进行, 只需在相应需要付费使用的接口位置,跟随页面提示完成后续充值及付费,即可完成。

< 返回EasyDL总览	零售版概览									
零售版										
应用列表	产品介绍: (展开	产品介绍: (展开查看服务功能介绍)						展开		
用量统计	可用接口列表									
监控报表	┃ 已上线的定制接口									
定制商品检测服务 >	API	模型ID	模型类型	模型名称	模型版本	状态	调用量限制	QPS限制	开通按量后付费 ?	购买QPS叠加包 ?
标准商品检测服务 ~	111	30443	商品检测	1111	V1	● 免费使用	剩余免费1000次	不保证并发	开通	购买 配额详情
货架拼接 >	silenceteapi	27185	商品检测	silencesecond	V3	●付费使用	500次/天免费 + 超 出按量计费	4	终止付费	购买丨配额详情
	silencefirst0829	27135	商品检测	silencefirst	V3	● 免费使用	500次/天免费	不保证并发	开通	购买 配额详情
	【 饮品检测									
	API		状态		调用量限制		QPS限制		开通按量后付费 ?	
	饮品检测		● 免费使用		500次/天免费		不保证并发		免费试用	
	日化品检测									
	API		状态		调用量限制		QPS限制		开通按量后付费 ?	
	日化品检测		● 免费使用		剩余免费1000次		不保证并发		免费试用	
	商品陈列翻拍识别									
	API		状态	调用	量限制	QPS限制		开通按量后付费	?	购买QPS叠加包 ?
	商品陈列翻拍识别		● 付费使用	剩余	免费999次 + 超出按量计费	4		终止付费		购买丨配额详情

⊙ 定制商品检测服务

价目表 - 按调用量后付费

定制商品检测模型云服务API支持以下三项功能,启停服务功能的方法请见 服务功能文档。

1. 商品基本信息识别 (必选)

接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度

2. 商品陈列层数识别 (可选)

货架场景:货架、端架(小方货架);冰柜场景:立式冰柜;地堆场景:堆箱、割箱、地龙

3. 商品陈列场景识别 (可选)

接口支持识别商品陈列的场景,场景类型支持:货架、端架和立式冰柜

4. 商品排面占比统计 (可选)

接口支持统计商品排面数/占比、未识别商品数、空位数及货架利用率

付费调用

每个模型发布的云服务API享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费,开通后即可开始计费使用。

注:为保证老用户可正常使用三项服务,对于服务功能上线前已经开通付费的定制模型API,默认对各项服务功能开通付费。商品陈列层数识 别和商品陈列场景识别两项服务功能,需在EasyDL零售版模型训练后台启动后方可使用,启动方法请见 服务功能文档,启动后发生调用才会 按实际调用次数进行收费。

选择开启的功能不同,单次接口调用价格不同,详情如下:

 商品基本信息识别(必选),按月调用量阶梯计费,调用单价按照自然月累积调用量所落阶梯区间而变化。月初,上月累积的调用量清零,重 新开始累积本月调用量。

月调用量 (万次)	单次调用价格 (元)	QPS限制	说明
0<月调用量<=15	0.009	4	服务器支持每秒处理4次查询
15<月调用量<=150	0.008	4	服务器支持每秒处理4次查询
150<月调用量	0.007	4	服务器支持每秒处理4次查询

• 商品陈列层数识别 (可选),单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.04	4	服务器支持每秒处理4次查询

• 商品陈列场景识别 (可选),单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.006	4	服务器支持每秒处理4次查询

• 商品排面占比统计 (可选) , 单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.02	4	服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

EasyDL零售版各项服务都具有免费调用额度,开通付费后,免费调用额度仍保留,EasyDL零售版的定制商品检测服务免费额度如下:

服务名称	免费调用额度	QPS限制
定制商品检测服务	累计1000次	1~2,服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

费用举例

从2019-3-1至2019-3-31,**定制商品检测API**在三种服务功能都开启的情况下,月调用量为600万次(已除去免费额度),那么费用如下:

1. 商品基本信息识别的费用为43,650元,明细如下:

前15万次落入0~15w阶梯,单次调用0.009元/次,费用为1,350元;

中间15万~150万次落入15~150w阶梯,单次调用0.008元/次,费用为10,800元;

最后150万~600万次落入大于150w阶梯,单次调用0.007元/次,费用为31,500元;

共计43,650元

2. 商品陈列层数识别的费用为360,000元,明细如下:

月调用量为600万次,单次调用0.04元/次,费用为240,000元

3. 商品陈列场景识别的费用为,明细如下:

月调用量为600万次,单次调用0.006元/次,费用为36,000元

综上,三月费用合计319,650元。

价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格

按天购买 50元/天

按月购买 1050元/月

购买QPS叠加包前需保证已开通按量后付费或购买次数包

购买的QPS叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

购买QPS叠加包提升的额度对定制商品检测API的所有服务功能均有效

心 定制地堆检测服务

价目表 - 按调用量后付费

定制地堆检测模型云服务API支持以下三项功能,启停服务功能的方法请见 服务功能文档。

- 商品基本信息识别(必选)
 接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度
- 2. 商品陈列场景识别 (可选)

货架场景:货架、端架(小方货架);冰柜场景:立式冰柜;地堆场景:堆箱、割箱、地龙

付费调用

每个模型发布的云服务API享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费,开通后即可开始计费使用。

注:为保证老用户可正常使用商品陈列场景识别服务功能,对于服务功能上线前已经开通付费的定制模型API,默认对各项服务功能开通付 费。商品陈列场景识别服务功能,需在EasyDL零售版模型训练后台启动后方可使用,启动方法请见 服务功能文档,启动后发生调用才会按实 际调用次数进行收费。

选择开启的功能不同,单次接口调用价格不同,详情如下:

 商品基本信息识别(必选),按月调用量阶梯计费,调用单价按照自然月累积调用量所落阶梯区间而变化。月初,上月累积的调用量清零,重 新开始累积本月调用量。

 单次调用价格(元)
 QPS限制
 说明

 0.016
 4
 服务器支持每秒处理4次查询

[•] 商品陈列场景识别 (可选) , 单次调用额外收取费用

 单次调用价格(元)
 QPS限制
 说明

 0.006
 4
 服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

EasyDL零售版各项服务都具有免费调用额度,开通付费后,免费调用额度仍保留,EasyDL零售版的定制地堆检测服务免费额度如下:

服务名称	免费调用额度	QPS限制
定制地堆检测服务	累计1000次	1~2,服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

费用举例

从2019-3-1至2019-3-31,**定制地堆检测API**在三种服务功能都开启的情况下,月调用量为600万次(已除去免费额度),那么费用如下:

1. 商品基本信息识别的费用为96,000元,明细如下:

- 月调用量为600万次,单次调用0.016元/次,费用为96,000元
- 2. 商品陈列场景识别的费用为,明细如下:
- 月调用量为600万次,单次调用0.006元/次,费用为36,000元

综上,三月费用合计132,000元。

价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格 按天购买 60元/天 按月购买 1200元/月

购买QPS叠加包前需保证已开通按量后付费或购买次数包

购买的QPS叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

购买QPS叠加包提升的额度对定制地堆检测API的所有服务功能均有效

心 翻拍识别服务

价目表 - 按调用量后付费

付费调用

每个账户享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用, 价格如下:

 调用量
 价格
 QPS限制
 说明

 每次调用
 0.05元
 4
 服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

每个账号享有一定量免费调用额度,如下表:

云服务API	免费调用额度	QPS限制	说明
商品陈列翻拍识别	累计1000次	1~2	服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

价目表 - 调用量次数包

如果对调用次数有预估,可以选择购买单次调用价格更低的次数包,价格如下:

规格	价格	QPS限制	有效期
1万次	490 元	4	1年
10万次	4,800 元	4	1年
100万次	45,000 元	4	1年
500万次	212,500 元	4	1年
1000万次	420,000 元	4	1年
2000万次	800,000 元	4	1年

购买后不可退款,次数包使用完后,开始按调用量每次0.05元收取费用 特殊说明,此计费方式仅限于单独调用翻拍模型接口,定制商品检测服务接口中的翻拍服务的计费不适用

价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

 购买方式
 每QPS价格

 按天购买
 50元/天

 按月购买
 1200元/月

购买 QPS 叠加包需保证已开通按量后付费或购买次数包

购买的 QPS 叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

心 货架拼接服务

货架拼接服务支持按任务数后付费、任务次数包预付费和并发任务叠加包预付费三种计费方式。

价目表 - 按任务数后付费

付费调用

每个账户享有累计200次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用,价格如下:

任务数	价格 (元)	并发任务数限制	说明
每次拼接任务	0.2	1	只允许一个拼接任务在运行,超出限制的任务排队等待

注:任务启动后失败和运行前终止不计费,任务成功和运行后终止会计费用

每个账号享有一定量免费调用额度,如下表:

服务 免费任务额度 并发任务数限制 说明

货架拼接 累计200次 1 只允许一个拼接任务在运行,超出限制的任务排队等待

注:任务成功与失败调用均消耗免费额度

价目表 - 任务次数包

如果对拼接任务次数有预估,可以选择购买单次任务价格更低的次数包,价格如下:

规格	价格	并发任 务 数限制	有效期
1千次	200 元	1	1年
1万次	1,900 元	1	1年
10万次	18,000 元	1	1年
100万次	150,000 元	1	1年
500万次	600,000 元	1	1年

购买后不可退款,任务次数包使用完后,开始按调用量每个任务0.2元收取费用

价目表 - 并发任务叠加包

开通付费后,并发任务数限制为1,如果有更多的并发请求需要,可以根据业务需求按天或按月购买并发任务叠加包,价格如下:

购买方式	每并发任务价格
按天购买	2元/天

按月购买 40元/月

购买 并发任务叠加包需保证已开通按量后付费或购买任务次数包

购买的并发任务叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

心 余额不足提醒与欠费处理

余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。

欠费处理

- 北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。
- 欠费后您开通付费的产品将进入欠费状态,只能使用每日的免费额度,超过额度的请求系统将不再响应,且不再保证并发处理。

本地部署价格说明

对于调用量稳定且巨大的企业,可以选择将定制化商品检测AI模型私有化部署在企业本地服务器上,如需要这种方式,请加入EasyDL零售版官方 QQ群(群号:1009661589)联系群管咨询。

公有云API价格说明

⊙ 开通付费及购买服务

EasyDL零售版的各项服务的「开通付费」、「购买QPS叠加包」、「购买调用量次数包」、「关闭付费」等操作均在EasyDL零售版控制台进行, 只需在相应需要付费使用的接口位置,跟随页面提示完成后续充值及付费,即可完成。 T

< 返回EasyDL总宽	零售版概览											
零售版												
应用列表	产品介绍: (展开)	查看服务功能介绍)										展开
用量统计	可用接口列表											
监控报表	■ 已上线的定制接口											
定制商品检测服务 🗸	API	模型ID	模型类型	模型名称		模型版本	状态		调用量限制	QPS限制	开通按量后付费 ?	购买QPS叠加包 ?
标准商品检测服务 >	111	30443	商品检测	1111		V1	● 免费使用		剩余免费1000次	不保证并发	开通	购买 配额详情
货架拼接 >	silenceteapi	27185	商品检测	silencesecor	nd	V3	● 付费使用		500次/天免费 + 超 出按量计费	4	终止付费	购买丨配额详情
	silencefirst0829	27135	商品检测	silencefirst		V3	● 免费使用		500次/天免费	不保证并发	开通	购买 配额详情
	【 饮品检测											
	API		状态			调用量限制			QPS限制		开通按量后付费 ?	
	饮品检测		● 免费使用			500次/天免费			不保证并发		免费试用	
	┃ 日化品检测											
	API		状态			调用量限制			QPS限制		开通按量后付费 ?	
	日化品检测		● 免费使用			剩余免费1000次			不保证并发		免费试用	
	商品陈列翻拍识别											
	API		状态		调用量等	制	QPS	限制		开通按量后付费	?	购买QPS叠加包 ?
	商品陈列翻拍识别		● 付费使用		剩余免费	1999次 + 超出按量计费	4			终止付费		购买丨配额详情

心 定制商品检测服务

价目表 - 按调用量后付费

定制商品检测模型云服务API支持以下三项功能,启停服务功能的方法请见 服务功能文档。

1. 商品基本信息识别 (必选)

接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度

- 商品陈列层数识别(可选)
 货架场景:货架、端架(小方货架);冰柜场景:立式冰柜;地堆场景:堆箱、割箱、地龙
- 商品陈列场景识别(可选)
 接口支持识别商品陈列的场景,场景类型支持:货架、端架和立式冰柜
- 4. 商品排面占比统计 (可选)

接口支持统计商品排面数/占比、未识别商品数、空位数及货架利用率

付费调用

每个模型发布的云服务API享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费,开通后即可开始计费使用。

注:为保证老用户可正常使用三项服务,对于服务功能上线前已经开通付费的定制模型API,默认对各项服务功能开通付费。商品陈列层数识 别和商品陈列场景识别两项服务功能,需在EasyDL零售版模型训练后台启动后方可使用,启动方法请见服务功能文档,启动后发生调用才会 按实际调用次数进行收费。

选择开启的功能不同,单次接口调用价格不同,详情如下:

 商品基本信息识别(必选),按月调用量阶梯计费,调用单价按照自然月累积调用量所落阶梯区间而变化。月初,上月累积的调用量清零,重 新开始累积本月调用量。

月调用量(万次)	单次调用价格 (元)	QPS限制	说明
0<月调用量<=15	0.009	4	服务器支持每秒处理4次查询
15<月调用量<=150	0.008	4	服务器支持每秒处理4次查询
150<月调用量	0.007	4	服务器支持每秒处理4次查询

• 商品陈列层数识别 (可选),单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明	
0.04	4	服务器支持每秒处理4次查询	

• 商品陈列场景识别 (可选),单次调用额外收取费用

Baidu 百度智能云文档

单次调用价格 (元)	QPS限制	说明
0.02	4	服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

EasyDL零售版各项服务都具有免费调用额度,开通付费后,免费调用额度仍保留,EasyDL零售版的定制商品检测服务免费额度如下:

服务名称	免费调用额度	QPS限制
定制商品检测服务	累计1000次	1~2,服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

费用举例

从2019-3-1至2019-3-31,**定制商品检测API**在三种服务功能都开启的情况下,月调用量为600万次(已除去免费额度),那么费用如下:

1. 商品基本信息识别的费用为43,650元,明细如下:

前15万次落入0~15w阶梯,单次调用0.009元/次,费用为1,350元;

中间15万~150万次落入15~150w阶梯,单次调用0.008元/次,费用为10,800元;

最后150万~600万次落入大于150w阶梯,单次调用0.007元/次,费用为31,500元;

共计43,650元

2. 商品陈列层数识别的费用为360,000元,明细如下:

月调用量为600万次,单次调用0.04元/次,费用为240,000元

3. 商品陈列场景识别的费用为,明细如下:

月调用量为600万次,单次调用0.006元/次,费用为36,000元

综上,三月费用合计319,650元。

价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格

按天购买 50元/天 按月购买 1050元/月

购买QPS叠加包前需保证已开通按量后付费或购买次数包

购买的QPS叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

购买QPS叠加包提升的额度对定制商品检测API的所有服务功能均有效

价目表 - 按调用量后付费

定制地堆检测模型云服务API支持以下三项功能,启停服务功能的方法请见 服务功能文档。

- 商品基本信息识别(必选)
 接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度
- 2. 商品陈列场景识别 (可选)

货架场景:货架、端架(小方货架);冰柜场景:立式冰柜;地堆场景:堆箱、割箱、地龙

付费调用

每个模型发布的云服务API享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费,开通后即可开始计费使用。

注:为保证老用户可正常使用商品陈列场景识别服务功能,对于服务功能上线前已经开通付费的定制模型API,默认对各项服务功能开通付 费。商品陈列场景识别服务功能,需在EasyDL零售版模型训练后台启动后方可使用,启动方法请见服务功能文档,启动后发生调用才会按实 际调用次数进行收费。

选择开启的功能不同,单次接口调用价格不同,详情如下:

 商品基本信息识别(必选),按月调用量阶梯计费,调用单价按照自然月累积调用量所落阶梯区间而变化。月初,上月累积的调用量清零,重 新开始累积本月调用量。

单次调用价格 (元)	QPS限制	说明
0.016	4	服务器支持每秒处理4次查询

• 商品陈列场景识别 (可选),单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.006	4	服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

EasyDL零售版各项服务都具有免费调用额度,开通付费后,免费调用额度仍保留,EasyDL零售版的定制地堆检测服务免费额度如下:

服务名称	免费调用额度	QPS限制
定制地堆检测服务	累计1000次	1~2,服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

费用举例

从2019-3-1至2019-3-31,**定制地堆检测API**在三种服务功能都开启的情况下,月调用量为600万次(已除去免费额度),那么费用如下:

1. 商品基本信息识别的费用为96,000元,明细如下:

月调用量为600万次,单次调用0.016元/次,费用为96,000元

2. 商品陈列场景识别的费用为,明细如下:

月调用量为600万次,单次调用0.006元/次,费用为36,000元

综上,三月费用合计132,000元。

价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格

按天购买 60元/天 按月购买 1200元/月

购买QPS叠加包前需保证已开通按量后付费或购买次数包

购买的QPS叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

购买QPS叠加包提升的额度对定制地堆检测API的所有服务功能均有效

心 翻拍识别服务

价目表 - 按调用量后付费

付费调用

每个账户享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用, 价格如下:

调用量	价格	QPS限制	说明
每次调用	0.05元	4	服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

每个账号享有一定量免费调用额度,如下表:

云服务API	免费调用额度	QPS限制	说明
商品陈列翻拍识别	累计1000次	1~2	服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

价目表 - 调用量次数包

如果对调用次数有预估,可以选择购买单次调用价格更低的次数包,价格如下:

规格	价格	QPS限制	有效期
1万次	490 元	4	1年
10万次	4,800 元	4	1年
100万次	45,000 元	4	1年
500万次	212,500 元	4	1年
1000万次	420,000 元	4	1年
2000万次	800,000 元	4	1年

购买后不可退款,次数包使用完后,开始按调用量每次0.05元收取费用 **特殊说明**,此计费方式仅限于单独调用翻拍模型接口,定制商品检测服务接口中的翻拍服务的计费不适用

价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格

按天购买 50元/天 按月购买 1200元/月

购买 QPS 叠加包需保证已开通按量后付费或购买次数包

购买的 QPS 叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

心 货架拼接服务

货架拼接服务支持按任务数后付费、任务次数包预付费和并发任务叠加包预付费三种计费方式。

价目表 - 按任务数后付费

付费调用

每个账户享有累计200次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用,价格如下:

任务数	价格(元)	并发任务数限制	说明
每次拼接任务	0.2	1	只允许一个拼接任务在运行,超出限制的任务排队等待

注:任务启动后失败和运行前终止不计费,任务成功和运行后终止会计费用

免费额度

每个账号享有一定量免费调用额度,如下表:

服务	免费任务额度	并发任务数限制	说明	
货架拼接	累计200次	1	只允许一个拼接任务在运行	,超出限制的任务排队等待

注:任务成功与失败调用均消耗免费额度

价目表 - 任务次数包

如果对拼接任务次数有预估,可以选择购买单次任务价格更低的次数包,价格如下:

规格	价格	并发任务数限制	有效期
1千次	200 元	1	1年
1万次	1,900 元	1	1年
10万次	18,000 元	1	1年
100万次	150,000 元	1	1年
500万次	600,000 元	1	1年

购买后不可退款,任务次数包使用完后,开始按调用量每个任务0.2元收取费用

价目表 - 并发任务叠加包

开通付费后,并发任务数限制为1,如果有更多的并发请求需要,可以根据业务需求按天或按月购买并发任务叠加包,价格如下:

购买方式 每并发任务价格 按天购买 2元/天 按月购买 40元/月 购买 并发任务叠加包需保证已开通按量后付费或购买任务次数包

购买的并发任务叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

∞ 余额不足提醒与欠费处理

余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。

欠费处理

- 北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。
- 欠费后您开通付费的产品将进入欠费状态,只能使用每日的免费额度,超过额度的请求系统将不再响应,且不再保证并发处理。

价格整体说明

心简介

本文档介绍EasyDL零售版各项服务的价格,EasyDL零售版提供两种服务,分别为定制模型服务、货架拼接服务。其中定制模型服务支持云服务 API和本地服务器部署两种服务方式,货架拼接服务支持云服务API。

云服务的计费方式包括:**按调用量后付费、调用量次数包预付费**和**QPS叠加包预付费**三种,根据实际购买的项目进行付费。其中**按调用量后付** 费方式,系统根据实际调用的次数,每小时对您的百度云账户进行扣费。如果需要本地服务器部署,请加入EasyDL零售版官方QQ群(群号: 1009661589)联系群管咨询。

心 计费方式介绍

EasyDL零售版的相关服务和接口计费方式如下:

服务名 称	服务方 式	是否计 费	计费方式	免费额度
定制模	云服务	计费	按调用量后付费和QPS叠加	每个模型发布的云服务API享有1000次免费调用量,超出免费额度的每次调用量根据选
型服务	API		包预付费	择的模型服务功能收取不同的费用,详情请见 价格说明文档
翻拍识	云服务	计费	按调用量后付费、调用量次	每账号享有累计免费1000次,超出部分按调用量计费, 超出免费额度的每次调用量根据
别服务	API		数包和QPS叠加包	选择的模型服务功能收取不同的费用,详情请见价格说明文档
货架拼 接服务	云服务 API	计费	按任务数后付费、任务次数 包和并发任务叠加包	累计免费200次拼接任务,超出免费额度的每次调用量根据选择的模型服务功能收取不同的费用,详情请见价格说明文档

免费额度用尽后,请在百度云EasyDL零售版控制台开通计费,开通后即可开始计费使用,计费方式请参考公有云API价格说明文档。

上述服务中,定制商品检测服务支持本地服务器部署,如需要这种方式,请加入EasyDL零售版官方QQ群(群号:1009661589)联系群管咨询。

EasyDL跨模态价格说明

EasyDL跨模态公有云API价格说明

EasyDL跨模态支持发布为在线API的任务类型包括:图文匹配

模型训练并发布为API后,可以在控制台看到已发布上线的所有公有云服务。

可以根据实际需求,开通「按量后付费」后,购买「调用点包」,高调用量下的优惠方案。也可以购买「QPS叠加包」,满足业务场景的高并发 需求。具体介绍见下方。

按量后付费

只需在智能云控制台「EasyDL」-「公有云服务」中找到需要付费使用的接口,点击开通付费,即可完成付费开通。立即开通

根据实际调用消耗的点数,系统每小时会对您的百度智能云账户进行扣费。1点=0.001元。

例如:**图文匹配·高精度模型**对应发布的公有云服务,每次调用消耗32点,调用100次对应消耗3200点,所以应付费3200*0.001=3.2元

免费额度

EasyDL定制化API服务都具有免费调用额度,开通付费后,免费调用额度仍保留,所有技术方向的API接口均有10000点免费额度

说明:成功调用与失败调用均消耗免费额度。

免费/付费对比

对于EasyDL各个能力,免费使用和开通付费后使用的配置有较大差异,具体对比如下:

状态	免费调用额度	超过调用额度	QPS限制
免费状态	拥有	不响应请求	不保证并发
付费状态	拥有	可继续请求	图像分类-高性能API服务保证10次并发;其余API保证4次并发

价目表

产品采用分段阶梯定价方式,调用单价按照自然月累积调用量所落阶梯区间而变化。每个月第一天清零上月累积的调用量,重新开始累积本月调用量。

未购买优惠商品的调用接口享受阶梯价格,按月计算:

月调用量 (万点)	每"点"对应换算(元/点)
0<月消耗调用点量<=100	0.001
100<月消耗调用点量<=500	0.0008
500<月消耗调用点量	0.00064

模型单次调用消耗的点数说明:

模型名称	单次调用消耗点数(点/次)	对应实际价格(元/次)
图文匹配-高精度	32	0.032

说明:调用失败不计费

费用举例

从2021-8-1至2021-8-31,本月某个公有云服务API接口的月消耗调用点量为200万点(已除去免费额度),费用如下: 前100万点落入0~100w阶梯,单价0.001元/点,费用为1000元 中间100万-200万点落入100w以上阶梯,单价0.0008元/点,费用为800元 本月费用共计:1800元

余额不足提醒与欠费处理

余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。

欠**费处**理

- 北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。
- 欠费后您开通付费的产品将进入欠费状态,只能使用每日的免费额度,超过额度的请求系统将不再响应,且不再保证并发处理。

∞ 调用点包

开通按量后付费后即可购买调用点包。立即购买

调用点包的本质可理解为"充值折扣",像「20万点」调用点包,按点与金钱换算规则,需要消耗200000*0.001=200元,但「20万点」调用点 包的目录价为180元。 调用点包有效期为12个月。计费调用优先抵扣调用点包额度,额度耗尽自动切换为按量后付费形式,具体计费标准如下:

调用点包点数(万点)	价格 (元)	对等折扣
20	180.00	0.90
50	430.00	0.86
100	800.00	0.80
300	2250.00	0.75
500	3500.00	0.70
1000	6500.00	0.65

EasyDL跨模态算力资源价格说明

⊙ 算力资源价格说明

EasyDL提供了丰富的模型训练算法,同时在训练任务上也提供多种机型自由选择。(其中,语音识别由于任务特殊性,将持续为开发者提供免费 训练体验)

说明:

为更好支持您的模型付费训练,平台针对您的模型创建数量、任务并行数均进行了权益升级,每位用户创建模型数量从30个提升至100个; 单模型组仅支持运行 1个训练任务提升至同时运行5个训练任务。

计费方式

计费规则如下:

- 按分钟计费,不足1分钟按百分比计算。
- 按小时扣费,即北京时间整点扣费并生成账单。出账单时间是当前计费周期结束后1小时内。例如,10:00-11:00的账单会在12:00之前生成,具体以系统出账时间为准。
- 使用 EasyDL 前需保证账户无欠款。 计费公式 费用=计算设备单价×计算设备数×使用时长 时长计量方法:只包括模型训练时的统计时间,数 据预处理等不包括在计费时长内。

产品单价

模型训练

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

说明1:扣费发生的时间点为任务训练结束(包含手动暂停训练或自动停止训练)后,如果因EasyDL系统异常导致训练任务运行失败,则相 应训练任务的全部耗时在账单中会做扣减,不会参与计费。

说明2:为确保训练任务的正常进行,建议您在开通付费后确保账户余额不低于100元。

余额不足提醒与欠费处理

余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。

欠费处理

北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。

EasyDL 图像使用说明

EasyDL图像介绍

心 任务简介

Hi,欢迎来到百度EasyDL图像

目前EasyDL图像共支持训练3种不同应用场景的模型:

• 图像分类

识别一张图中是否是某类物体/状态/场景。可以识别图片中主体单一的场景

- 物体检测
- 在一张图包含多个物体的情况下,定制识别出每个物体的位置、数量、名称。可以识别图片中有多个主体的场景
- 图像分割

对比物体检测,支持用多边形标注训练数据,模型可像素级识别目标。适合图中有多个主体、需识别其位置或轮廓的场景

- 心 产品优势
- 心 可视化操作

无需机器学习专业知识,模型创建·数据上传·模型训练·模型发布全流程可视化便捷操作,最快15分钟即可获得一个高精度模型

心 操作步骤

Step 1 创建模型

确定模型名称,记录希望模型实现的功能

Step 2 上传并标注数据

分类功能的模型:只需按分类(如合格图片vs不合格图片)上传图片即可

检测功能的模型:上传数据后,需要在数据中标注出需要检测的具体目标

分割功能的模型:上传数据后,需要在数据中标注出需要识别物体的轮廓

Step 3 训练模型并校验效果

选择部署方式与算法,用上传的数据一键训练模型

模型训练完成后,可在线校验模型效果

Step 4 发布模型

根据训练时选择的部署方式,将模型以云端API、本地部署SDK、端云协同部署包等多种方式发布使用

更详细的操作指导,请参考各类模型的技术文档

```
心 高精度效果
```

- EasyDL图像底层结合百度 AutoDL/AutoML技术,针对用户数据能够自动获得最优模型和最优超参组合,进而基于少量数据就能获得出色性能 和模型效果
- EasyDL图像以百度独有超大规模预训练模型为基座,小量级数据进行训练也可获得高精度模型

心 高精算法

- 采用PaddlePaddle深度学习框架结合Auto Model Search,保证模型效果领先
- 训练图像分类和物体检测模型时,均支持选择多种算法,满足不同场景对性能、效果的不同需求;还有专项精度提升配置包,包含自动超参搜索、小目标检测等精度优化功能,针对优化模型效果

训练图像分类模型时,支持选择AutoDL Transfer

AutoDL Transfer模型是百度研发的AutoDL技术之一,结合了模型网络结构搜索、迁移学习技术、并针对用户数据进行自动优化。与通用算法相比,训练时间较长,但更适用于细分类场景。例如,通用算法可用于区分猫和狗,但如果要区分不同品种的猫,则AutoDL效果会更好

心 免训练极速迭代

训练图像分类模型之后,支持开启免训练极速迭代模式。该模式基于深度度量学习技术(Deep Metric Learning),模式开启后,模型的迭代添加数据仅需等待几分钟即可获得效果不错的模型,无需训练。适用于数据量大,模型迭代频繁的用户需求场景。

⊙ 丰富的部署方案

 训练完成后,可将模型部署在公有云服务器、私有服务器,封装成可离线运行的设备端SDK,或直接购买软硬一体方案,灵活适配各种使用场 景及运行环境,也可直接发布为端云协同部署包,下发至边缘设备进行应用

• 本地部署的性能评测详细信息可见模型算法推理性能大表

部署方式	支持的硬件	支持的系统	技术文档
公有云API	可集成公有云API即可	不限制	图像分类 物体检测 图像分割
私有服务器部署 [私有API]	x86-64 CPU	Linux	图像分类 物体检测 图像分割
	Nvidia GPU	Linux	图像分类 物体检测 图像分割
私有服务器部署 [服务器端SDK]	x86-64 CPU	Linux	图像分类 物体检测 图像分割
	Nvidia GPU	Linux/Windows	图像分类 物体检测 图像分割
	HUAWEI Atlas 300	Linux	图像分类 物体检测
通用设备端SDK	ARM (AArch64, ARMv7I)	Linux	图像分类 物体检测 图像分割
	Hisilicon NNIE	Linux	图像分类 物体检测
	HUAWEI Atlas 200	Linux	图像分类 物体检测
	ARM	Android/iOS	图像分类 物体检测 图像分割
	Qualcomm Snapdragon GPU/DSP	Android	图像分类 物体检测
	Hisilicon Kirin NPU	Android	图像分类 物体检测
	Apple A-Bionic	iOS	图像分类 物体检测
	x86-64 CPU	Windows	图像分类 物体检测 图像分割
	Intel Movidius NCS (MyRIAD 2/MyRIAD X)	Linux/Windows	图像分类 物体检测
专项硬件适配SDK [软硬一体方案]	Baidu-EdgeBoard(FZ)	Linux	方案介绍及对比
	Baidu-EdgeBoard(VMX)	Linux/Windows	
	Nvidia-Jetson(Nano/TX2/Xavier)	Linux	
端云协同部署	x86-64 CPU	Linux	图像分类 物体检测 图像分割
	ARM (AArch64, ARMv7I)	Linux	

心 公有云API

支持图像分类、物体检测、图像分割模型

训练完成的模型存储在云端,可通过独立Rest API调用模型,实现AI能力与业务系统或硬件设备整合

具有完善的鉴权、流控等安全机制, GPU集群稳定承载高并发请求

支持查找云端模型识别错误的数据,纠正结果并将其加入模型迭代的训练集,不断优化模型效果

心 私有服务器部署

支持图像分类、物体检测、图像分割模型

Baidu 百度智能云文档

将训练完成的模型部署在私有CPU/GPU服务器上,支持私有API和服务器端SDK两种集成方式,可在内网/无网环境下使用模型,确保数据隐私

- 私有API:将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷
- 服务器端SDK:将模型封装成适配本地服务器(支持Linux和Windows)的SDK,可集成在其他程序中运行。首次联网激活后即可纯离线运行, 占用服务器资源更少,使用方法更灵活

心 设备端SDK

支持图像分类、物体检测、图像分割模型

训练完成的模型被打包成适配智能硬件(不含服务器)的SDK,可进行设备端离线计算。满足推理阶段数据敏感性要求、更快的响应速度要求

支持iOS、Android、Linux、Windows四种操作系统,基础接口封装完善,满足灵活的应用侧二次开发

提供基础版、加速版(已支持通用x86、通用ARM芯片)两种版本,可根据业务场景需求选择。了解加速版性能:图像分类物体检测

心 软硬一体方案

支持图像分类、物体检测模型,了解更多

提供与模型深度适配的高性能硬件方案,多种算力、价位可选

可应用于工业分拣、视频监控等多种设备端离线计算场景,让离线AI落地更轻松

心 智能数据服务

全方位支持训练数据的采集、标注、质检、增强,助力提升模型效果

心 数据采集

在云服务调用数据管理中,可查找云端模型识别错误的数据,纠正结果并将其加入模型迭代的训练集

可实现训练数据的持续丰富和模型效果的持续优化

点击了解功能说明:图像分类、物体检测

⊙ 智能标注

针对物体检测模型,可通过智能标注降低标注成本

启动后,只需标注数据集30%左右的数据即可训练出同等效果的模型

在图像分割任务中还提供"自动识别轮廓标注"来自动标注目标轮廓,降低标注成本

⁰ 多人标注

训练物体检测模型前,可与其他用户共享数据集,实现多人分工标注数据后再集中训练模型

心 采集/标注支持

联合第三方数据标注合作伙伴,提供全面且高质量的训练数据采集、标注服务

可在AI市场选择合适的数据服务商

图像分类

整体介绍

```
心简介
```

Hi,您好,欢迎使用百度EasyDL图像-图像分类,请您根据实际应用场景选择模型类型。

EasyDL图像支持定制图像分类、物体检测、图像分割三类模型,三类模型的功能区别如下:

- 图像分类:识别一张图中是否是某类物体/状态/场景,适用于图片内容单一、需要给整张图片分类的场景
- 物体检测:检测图中每个物体的位置、名称。适合图中有多个主体要识别、或要识别主体位置及数量的场景
- 图像分割:对比物体检测,模型可像素级识别目标。适合图中有多个主体、需识别其位置或轮廓的场景

图像分类是AI视觉应用中最经典的能力,常见的应用场景如下:

- 图片内容检索:定制训练需要识别的各种物体,并结合业务信息展现更丰富识别结果
- 图片审核:定制图像审核规则,如训练直播场景中抽烟等违规现象
- 制造业分拣或质检:定制生产线上各种产品识别,进而实现自动分拣或者质检
- 医疗诊断:定制识别医疗图像,辅助医生肉眼诊断

心 使用流程

训练模型的基本流程如下图所示,全程可视化简易操作。在数据已经准备好的情况下,最快15分钟即可获得定制模型。

下面将详细介绍每一步的操作方式和注意事项。如果文档没有解决您的问题,请在百度智能云控制台内提交工单反馈。

数据准备

の 创建数据集

在训练之前需要在数据中心【创建数据集】,添加并标注数据

EasyDL开友平台	数据集官 理
嘂 图像分类概览 ≒	平台支持统一纳管自训练模型的数据集,并支持自主版本迭代、数据查看、导入导出和删除等操作。
数据服务	
◎ 数据总览	EasyData智能数据服务平台已上线,使用EasyData可享受包括多人标注、数据清洗、数据采集等完整数据服务 立即前往
② 数据标注 へ	+ 创建数据集
标签组管理	txfl 区 数据集组D: 13364

设计分类

首先想好分类如何设计,每个分类为你希望识别出的一种结果,如要识别水果,则可以以"apple"、"pear"等分别作为一个分类;如果是审核的场 景判断合规性,可以以"qualified"、"unqualified"设计为两类,或者"qualified"、"unqualified1"、"unqualified2"、"unqualified3"……设计为多类。

注意:目前单个模型的分类上限为1000类

准备数据

基于设计好的分类准备图片:

- 每个分类需要准备20张以上
- 如果想要较好的效果,建议每个分类准备不少于100张图片
- 如果不同分类的图片具有相似性,需要增加更多图片,尽量提升图片数据的丰富度
- 一个模型的图片总量限制10万张 (每个账户的图片数量上限为10万张)

图片格式要求:

- 1、目前支持图片类型为png、jpg、bmp、jpeg,图片大小限制在14M以内
- 2、图片长宽比在3:1以内,其中最长边小于4096px,最短边大于30px

图片内容要求:

1、训练图片和实际场景要识别的图片拍摄环境一致,举例:如果实际要识别的图片是摄像头俯拍的,那训练图片就不能用网上下载的目标正面 图片;如果是需要识别白天光照下的物体,就不能使用夜晚拍摄的图片数据

2、每个分类的图片需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强

如果训练图片场景无法全部覆盖实际场景要识别的图片:

• 如果要识别的主体在图片中占比较大,模型本身的泛化能力可以保证模型的效果不受很大影响

如果识别的主体在图片中占比较小,且实际环境很复杂无法覆盖全部的场景,建议用物体检测的模型来解决问题(物体检测可以支持将要识别的主体人训练图片中框出的方式来标注,所以能适应更泛化的场景和环境)

如果需要寻求第三方数据采集团队协助数据采集,可以在百度智能云控制台内提交工单

心 上传数据集并在线标注

在完成了设计分类与准备数据后,可以通过以下方式导入数据:

- 导入未标注/分类的数据,在线进行数据标注
- 直接导入标注/分类好的数据

导入未标注数据

本地数据

支持上传图片、压缩包,或通过API导入

已有数据集

支持选择百度云BOS导入、分享链接导入、平台已有数据集导入;支持选择线上已有的数据集,包括其他图像类模型的数据集

EasyDL开发平台	< 返回 导入数据		
嘂 图像分类概览 ⇒			
数据服务	导入配置		
■ 数据总览	数据标注状态:	● 无标注信息	有标注信息
🗵 数据标注 🔷 🔨	导入方式:	请选择	
标签组管理		本地导入	
在线标注	守八记录:	BOS目录导入	
智能标注		分享链接导入	
5 数据增强	•	公开数据集	
∂ 数据回流 へ		摄像头采集数据	
云服务回流		云服务数据回流	
摄像头采集			

在线标注

上传未标注数据后,即可进入「标注数据集」页面进行在线标注。标注的方式非常简单,只需在右侧标签栏新建并选定标签即可



标签名支持中英文数字中下划线,不超过256字符

时,推荐您批量标注:选择好属于同一类的图片后,点击对应标签即可完成标注

EasyDL开发平台	< 返回 数据集详情 ① 批量标注示例 号入 批量标注
器 图像分类概览 ⇔	▼ 全部(24) 有标注信息(24) 无标注信息(0)
数据服务	
■ 数据总览	
🗵 数据标注 🔷 🔨	
标签组管理	
在线标注	相規图片內容, 选择标签
智能标注	
□ 数据增强	BMW_1_Series_Convertible_2012 6 BMW_1_Ser BMW_1_Ser BMW_1_Ser Hyundai_T BMW_1_Ser (
② 数据回流 へ	Acura_TL_Sedan_2012 4
云服务回流 摄像头采集	Hyundai_Tucson_SUV_2012 4
模型管理	Cadillac_CTS_V_Sedan_2012 2 Hyundal_T Cadillac_C Acura_TL Acura_TL Cadillac_C
◎ 模型训练	
結刑約番	

导入已标注数据

本地数据

支持上传压缩包,或通过API导入

压缩包支持通过两种格式上传,点击「上传压缩包」,即可查看详细的格式要求

EasyDL开发平台	く 返回 导入数据		
品 图像分类概览 ≒			
数据服务	导入配置		
目 数据总览	数据标注状态:	○ 无标注信息 ● 有标注信息	
🗵 数据标注 🔷 🔨	导入方式:	本地导入 ~	请选择 ^
标签组管理	导入记录:	新士官》记录	上传压缩包
在线标注	077103	B7097703	API导入
智能标注			
□ 数据增强			
② 数据回流 へ			
云服务回流			
摄像头采集			
模型管理			
@ 模型训练			

已有数据集

EasyDL开发平台 < 返回 导入数据 嘂 图像分类概览 ≒ 导入配置 数据服务 数据标注状态: ○ 无标注信息 ● 有标注信息 目 数据总览 🗵 数据标注 $^{\sim}$ 导入方式: 请选择 本地导入 标签组管理 本地导入 导入记录: BOS目录导入 在线标注 分享链接导入 智能标注 平台已有数据集 □ 数据增强 公开数据集 ⊙ 数据回流 ~ 云服务回流 摄像头采集 模型管理 ◎ 模型训练

支持选择线上已有的数据集,仅支持选择图像分类数据集
使用智能标注功能可降低数据的标注成本。启动后,系统会从数据集所有图片中筛选出最关键的图片并提示需要优先标注。通常情况下,只需标 注数据集30%左右的数据即可训练模型。与标注所有数据后训练相比,模型效果几乎等同

整体流程以图像分类的智能标注流程为例:

创建智能标注任务

启动图像分类数据集的智能标注前,请先检查以下是否已满足以下条件:

- 所有需要识别的分类标签都已创建
- 每个标签的图片数不少于10个
- 所有需要标注的图片都已加入数据集,且所有不相关的图片都已删除

若已满足,即可从导航栏进入「数据服务」-「智能标注」,创建智能标注任务,系统会基于您选择数据类型及数据量级,自动预估任务运行时 长。

智能标注分为两种任务类型:主动学习、指定模型

主动学习适合您在首次上传数据,没有使用数据训练过模型的情况下进行。智能标注算法会通过您已标注数据的规则,去标注数据

指定模型适合您已将需要智能标注的数据进行过模型训练,且模型精度表现良好(评估指标90%以上)时选择。智能标注算法将会以指定模型为 基础去智能标注数据

主动学习 主动学习仅需选择需要智能标注的数据集即可发起智能标注任务

EasyDL开发平台	< 返回 创建智能标注
嘂 图像分类概览 ≒	
数据服务 目 数据总览	智能标注任务未结束之前,暂时无法进行数据清洗或数据增强任务。 针对主动学习任务,由于系统筛选图片需一定时间,该功能将在每个分类标签下的图片数大于10张且未标图片数大于10张时方可启动 针对指定模型任务,该功能将在未标图片数大于0时方可启动 数据集量级会影响任务时长,请参考任务时长预估表,主动学习 ① 指定模型 ②
② 数据标注 へ	
标签组管理	选择数据集: usrbos1 / V1 V
在线标注	系统将对您选择的数据集进行吸烟、请确保每个标签的图片都达到10个 您的智能标注任务预计需要 5-10分钟 完成
智能标注	任务类型: • 主动学习 ⑦ 指定模型 ⑦
□ 数据增强	

指定模型 当选择指定模型后,除了选择需要进行智能标注的数据集,还需选择智能标注算法的模型。支持从EasyDL/BML已训练的模型中选择,同时也支持从历史智能标注产生的模型中选择,建议选择所选数据集的数据训练产生的精度表现良好(评估指标90%以上)的模型

EasyDL开发平台	任务类型:	 主动学习 ⑦ 指定模型 ⑦ 			
嘂 图像分类概览 ≒	选择模型:	来自训练平台(902个) ② 来自已有智能标注任务(5个) ③			
数据服务		EasyDL BML			
◎ 数据总览		输入模型名称或ID Q			
② 数据标注 へ					
标签组管理		模型名称	模型ID	版本数量	模型来源
在线标注		〉 QA_74_公有云_高精度_	18827	{ "model_id": "18827", "cnt": "2" }	EasyDL
智能标注		> 【云智一体】 qa_imgcls_bos	18813	{ "model_id": "18813", "cnt": "1" }	EasyDL
□ 数据増强 ○ 数据回流 へ		> 【云智一体】 qa_imgcls_plat	18812	{ "model_id": "18812", "cnt": "6" }	EasyDL
云服务回流		> xsy_test11	18811	{ "model_id": "18811", "cnt": "2" }	EasyDL
摄像头采集		> showcase_图像分类	18803	{ "model_id": "18803", "cnt": "2" }	EasyDL
 ·		> I具_algo	18790	{ "model_id": "18790", "cnt": "1" }	EasyDL

系统筛选难例

系统会分批筛选出最关键需标注的图片,即难例图片。

Tips:难例筛选需要一定时间,在此期间您可以正常进行其他未标注图片的标注

EasyDL开发平台	智能标注							
器 图像分类概览 ≒								
数据服务								
目 数据总览								
② 数据标注 ^	智能标注启动中,请稍候…							
标签组管理	启动过程需要大约5分钟左右的时间,请耐心等待 您的智能标注任务预计需要5-10分钟完成	9,了解功能诌	é情					
在线标注	关闭智能标注							田古本江市石
智能标注								用尸缃队难的
□ 数据增强								
○ 数据回流 へ								
云服务回流			-		X			
摄像头采集	启动智能标注	>	系统筛选难例	>	用户标注难例	>	完成标注,启动训练	
模型管理	在每个标签下的图片(图像分类单标签)或 者标注框数(物体检测、图像分割)都达到 10个时方可启动	根 集 片	國巡时數据集情况,系统目初帰述出數据 中难例图片(最关键需优先标注的图),进行自动预标注		针对展示烦标注结果的难例,您可通过两种 方式完成标注:第一,基于单张图片进行手 动标注:第二,在预标注效果符合预期基础		18.可以根据实际情况,这择系统力加进行下一轮难例筛选(系统最多支持4轮难例筛 选),或对当前预标注效果满意进入立即训	
 					上,选择一键标注		练	

智能标注任务启动后,系统为您自动筛选难例,您可以通过总览页查看进度按钮查看当前难例筛选进度,同时,进度图中也会全局展示您处于难 例筛选的具体哪一环节,以便您的操作后续。 筛选难例完成后,绿色进度条会进展到确认难例阶段,您可以点击【确认难例】完成对预标注结果 的人工确认。

图像分类模型	=	智能标注						
88 总览								
直 模型中心		智能标注	智能标注功能可通过系	統自动筛选出数据集中难例图片(最关	长键需优先标注的	图片),用户只需标	注数据集30%左右的难例数据	即可训练模型,以减少数据标注的人力投入。
我的模型								
创建模型		创建积能标	注任条					
训练模型		GTAE IN BUID	Product Value (Prod					
校验模型		序号	数据集版本ID	数据集组名称	版本	标注类型	智能标注状态	操作
发布模型		1	276395	1202catdogs	V1	主动学习	• 运行中	查看进度 确认难例 中止任务
☑ EasyData数据服务					当前您处	于第1轮难例阶段	(共4轮)	
数据总览					已为您筛送	:出轮难例图片,请确	认该轮难例图片	
标签组管理							来达预期 下-	轮筛选
在线标注							● 效果确认	
智能标注					7年1达X里19月	2019月1月1人		动标注未标图片
云服务数据回流								知道了
摄像头数据采集								
目 公有云服务								

我们为您的人工确认提供两种模式:

- 单张确认,在该模式下支持您对预标注结果进行修正后点击保存
- 一键保存所有标注,为提升您的确认效率,默认您对难例的预标注结果全部满意,即可进入下一阶段



标注难例的预训练模型,也会对您无标注信息下的图片进行预标注结果的展示,您有余力的情况下,可以完成标注确认,确认后该张图片将升级 为已标状态,该环节并非是您进入智能标注下一阶段的必备要求。



评估难例效果,完成任务

当您对难例完成确认后,您可以根据本轮次预标注的结果是否满意,判断您是否还需要进入下一轮难例筛选阶段,如果满意本轮难例的预标注效果,系统将自动为您系统其他的未标图片打标签。

第1轮难例标注中(共4轮)

1、点击右下角【保存当前标注】该预标注结果将完成确认,支持您对标注修改后再保存
2、您只有对【待确认标注】下所有预标注结果完成确认,所有难例均升级为已标状态,才可进入下一阶段
✓ 该轮预标注难例已全部完成确认
1、若预标注效果未达预期,可通过更多轮次难例筛选实现效果优化, 进入下一轮难例筛选
2、 若预标注效果符合预期,将按照当前标注效果,对剩余未标图片进行自动标注, 确认自动完成标注

中止任务

当您在任务运行中想要中止任务时,可实时点击标注页面右上方【中止任务】按钮,任务将被提前结束。



其他操作提示

• 在智能标注任务中,有任务上限吗?

支持五条智能标注任务同时运行,超过该上限您需要中止其他任务

• 智能标注中可以增删标签吗?

暂不支持。为了保证系统智能标注的效果,建议在启动功能前就创建好所有需要识别的标签 如果确实需要增删标签,可以先结束智能标注

• 智能标注中可以增删图片吗?

暂不支持。为了保证系统智能标注的效果,建议在启动功能前上传需要标注的所有图片,并删除不相关的图片。如果确实需要增删图片,可以先 结束智能标注

• 智能标注中可以修改已标注图片的标注框吗?

可以。但为了保证智能标注的效果,建议不要大量改动。如果确实需要修改大量标注,建议先结束智能标注

• 为什么我已经人工标注了很多图片,但系统预标注依然不准?

系统预标注的结果会受以下因素影响:智能标注期间,对"已标注"图片的标签进行大量改动;曾结束智能标注,并对标签、图片进行增删

• 多个数据集是否可以同时启动智能标注?

目前每个账号同一时间仅支持对一个数据集启动智能标注

• 共享中的数据集是否可以启动智能标注?

暂不支持。智能标注中的数据集也暂不支持共享

• 智能标注失败了怎么办?

可以先尝试稍后重新启动,如多次失败请提交工单联系我们

问题反馈

您在使用EasyData过程中可以通过以下任何方式联系我们:

• 在社区咨询

在论坛发帖提交问题,也可以在论坛与其他用户一起交流。前往论坛

• 提交工单

如果使用EasyData遇到其他任何问题或任何bug,您可以点此提交工单

• 添加微信小助手留言

请在微信搜索"BaiduEasyDL",并备注暗号 "EasyData",添加小助手后留言。

の数据集管理API

本文档主要说明当您线下已有大量的已经完成分类的图片数据,如何通过调用API完成图片的便捷上传和管理。EasyDL图像数据集管理API在管理 不同模型数据类型之间是通用的。上传不同模型类型数据,只是在部分接口入参存在差异,使用及接口地址完全一致。

数据集创建API

接口描述

该接口可用于创建数据集。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL控制台-公有云部署-应用列表页面创建应用
- 2、应用详情页获取API Key和Secret Key

三 🗘 百度智	能元 控制台总览 全局 🗸	
く 返回EasyDL总览	产品服务 / EasyDL图像 - 应用列表	
EasyDL图像	应用列表	
算力资源管理	+ 创建应用	
公有云部署		AppID
• 应用列表	1 teastapp107	23502406
 权限管理 		

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/create

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset_na me	是	string	数据集名称,长度不超过20个utf-8字符

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
dataset_id	否	number	创建的数据集ID

数据集列表API

接口描述

该接口可用于查看数据集列表。返回数据集的名称、类型、状态等信息。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL控制台-公有云部署-应用列表页面创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/list

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值	

Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
start	否	number	起始序号,默认为0
num	否	number	数量,默认20,最多100

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	数据集总数
results	否	array(object)	数据集列表
+dataset_id	否	number	数据集ID
+dataset_name	否	string	数据集名称
+type	否	string	数据集类型
+status	否	string	数据集状态
+special_status	否	string	数据集特殊状态,包括shared、smart和空值,分别表示共享中、智能标注中、非特殊状态

分类 (标签)列表API

接口描述

该接口可用于查看分类(标签)。返回分类(标签)的名称、包含数据量等信息。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL控制台-公有云部署-应用列表页面创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/list

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID
start	否	number	起始序号,默认0
num	否	number	数量,默认20,最多100

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	标签总数
results	否	array(object)	标签列表
+label_id	否	string	标签/分类ID
+label_name	否	string	标签/分类名称
+entity_count	否	number	图片/声音/文本数量

添加数据API

接口描述

该接口可用于在指定数据集添加数据。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL控制台-公有云部署-应用列表页面创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/addentity

URL参数:

参数 值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID
append Label	否	boolea n	确定添加标签/分类的行为:追加(true)、替换(false)。默认为追加(true)。
entity_c ontent	是	string	type为 IMAGE_CLASSIFICATION/OBJECT_DETECTION/IMAGE_SEGMENTATION/SOUND_CLASSIFICATION时,填入图 片/声音的base64编码;type为TEXT_CLASSIFICATION时,填入utf-8编码的文本。 内容限制为:图像分类base64前 10M;物体检测base64前10M;图像分割base64前10M;声音分类base64前4M,声音时长1~15秒;文本分类 10000个汉字
entity_n ame	是	string	文件名
labels	否	array(o bject)	标签/分类数据。若为空,则只上传图片,不上传标签/分类。若不为空,则应在数组中包含以下前面带+的参数
+label_ name	是	string	标签/分类名称(由中文、数字、字母、中划线、下划线组成),长度限制20B
+left	否	number	物体检测时需给出,标注框左上角到图片左边界的距离(像素)
+top	否	number	物体检测时需给出,标注框左上角到图片上边界的距离(像素)
+width	否	number	物体检测时需给出,标注框的宽度(像素)
+height	否	number	物体检测时需给出,标注框的高度(像素)

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

数据集删除API

接口描述

该接口可用于删除数据集。

接口鉴权

同模型上线后获取的API:

1、在EasyDL控制台-公有云部署-应用列表页面创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/delete

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

分类 (标签) 删除API

接口描述

该接口可用于删除分类(标签)。

接口鉴权

同模型上线后获取的API:

1、在EasyDL控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/delete

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset_ id	是	number	数据集ID
label_na me	是	string	标签/分类名称

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求,如果持续出现此类错误,请在百度智能云控制台内提交工单反 馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度智能云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量 阶梯计费;未上线计费的接口,请在百度智能云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线计费的接口,请在百度智能云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶 梯计费;未上线计费的接口,请在百度智能云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
406000	internal server error	服务错误
406001	param[xx] invalid	参数xx不合法,请检查相关参数
406002	dataset not exist	数据集不存在
406003	dataset already exists	数据集已存在
406004	dataset can not be modified temporarily	数据集暂不可修改
406005	label not exist	标签/分类不存在
406006	no permission to modify the dataset	没有修改数据集的权限
406007	dataset cannot be modified while smart annotation is running	智能标注期间不可修改数据集
406008	quota exceeded	配额超限

心 数据质检

功能概述 该功能旨在对您数据集中的图像数据进行质量检测,通过提供客观指标,为您对数据集的下一步操作(标注、清洗等)进行参照引导。

整体质检报告将包括对原图、标注信息两个层面的指标进行统计,本期先上线原图维度的质检指标,标注层面的质检指标敬请期待。

使用流程 Step 1 功能入口

您可从数据总览页操作列点击【质检报告】或查看页面点击【质检报告】进入该功能页面

版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作
V1⊡	142909	5	● 已完成	图像分类	0% (0/5)	-	查看与标注 导出 删除 质检报告

≡⊦	我的数据总览 > 【图片】的/V1/查看	
	全部 (5) 有标注信息 (0) 无标注信息 (5)	十 导入图片 🖃 质检报告 批量标注示例
	的V1版本的图片列表 筛选 🗸	□ 本页全选 前 删除

Step 2 指标查看本期报告分为整体指标和分布指标两类。整体指标包括数据集存储大小、图片数量、破损图像数三类;分布指标包括色彩分布 空间、图像存储大小分布、高宽比分布、分辨率分布、色偏分布五类。

可以通过切换数据集版本查看不同版本下质检报告。

数据质检报告		
数据版本 V1 V 数据集大小 2.04 MB 图片数量 5 5	帐 破损图片数量 0 ⑦	
②彩分布空间 日本の5 日本の5 日本の5 日本の5 日本の5 日本の5 日本の5 日本の5	田康大小分布の (日) 3 0 0 0 2 20000 40010 50010 80010 975.5890	高記公布の (部) 4 3 2 1 0 0 0 15 2 7 3 4 4 5 5
分辨事分布 ⑦ (個) 2 15 1 05	登録分布 ① (個) 4 2 1	

Step 3 对应处理 可通过hover具体指标数值进行相关操作,以高宽比分布为例:

第一步,高宽比大于3.6的超长图hover显示有1张图片比,支持点击



第二步,点击后进入符合该指标的图片操作页,可针对筛选后图片进行删除、标注等操作

I/V1/查看 > 质检报告		
高宽比分布列表(1张) 筛选 🗸 …		二本页全选 □ 删除
0% <u>自由</u> [編編con]即可进 入"举强的片枝注明式" 无秘密 C 22		毎页显示 14 マ < 1 > 17 3.8 4.5 5
	ビ調力印 ♡	
	(30) 4 3 2	

模型训练

心 图像分类创建模型

在导航【模型训练】中,点击训练模型,填写相关信息,即可创建训练模型。

操作示例:

EasyDL开发平台	我的模型				
智能标注	十 训练模型				
②数据回流 ^	【图像分类】24	2112 🗹 模型ID: 18835			
云服务回流			横刑创建成功 若无数握集诸失去"数据总发	「 创建 上传训练数据 训练 栲	闭后 可以在此处查看模型的最新版本
摄像头采集				S DOVE , IL IS WISH SAME WISH P	
模型管理					
◎ 模型训练	《 【图像分类】2	☑ 模型ID: 18834			
EasyDL开发平台	训练模型				
智能标注			1 模型准备 —	② 数据准备	③ 训练配置
□ 数据增强	模型准备				
⊙数据回流 ^	描示此生来。	▲네Z世來G #節 开川	口方推刑		
云服务回流	1英主龙井。		口有快至		
摄像头采集	*模型名称:	请输入名称	0/20		
模型管理	*所属行业:	请选择行业	~		
@ 模型训练	4				
模型校验	*应用场景:	请选择应用场景	~		
⊘ 模型校验	*业务描述:	请输入业务描述			
模型服务					
△ 公有云服务 ^			0/500		
在线服务					
批量预测					

注:1. 创建模型后可持续新增模型版本,因此不必每次训练模型都创建模型 2.目前单个用户在每种类型的模型下最多可创建10个模型,每个模型均支持多次训练。 3.如果您是企业用户,建议您按照真实企业信息进行填写,便于EasyDL团队后续更好的为您服务

⊙ 图像分类训练操作说明

数据提交后,可以在导航中找到【训练模型】,按以下步骤操作,启动模型训练:

注:1.启动训练前请确保数据已经标注完成,否则无法启动训练

2. 下述训练功能点中,标注为星号(*)的功能为非必要选择项,可根据实际需求考虑是否使用

作文档 常见问题 新手教程 提交

EasyDL开发平台

训练模型

⊙数据回流 ^	训练配置					
云服务回流	*部署方式:	公有云部署 Easy	KBD 化 KBD N	8署方式?		
摄像头采集	*训练方式:	根时免费 精度提升配置包	常规训练			
模型管理	冻探館注。		P.CAL研训院婚刑,通用将奏 ② 一 百度文心の/ V	IMED_IEO研训技描册_行业技展		
 模型训练 	2514-94-721	1000	500 300 200 100	50		
 ● 模型校验 	选择网络:	● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●		⑦ 性能最高		
模型服务	自动超参搜索:					
△ 公有云服务 へ	训练环境:	名称	规格	算力	速度比例	价格
在线服务		GPU PA	TaelaCDII DA SC星方单十 12旅CDII AOC内方	5.5 TaraELODS	1	单卡¥4.02/小时
批量预测		0.014	10300107 4700 Elbit 171201 074001315	0.0 101010		已购小时包余额: 138.2小时*节点
ふ 本地部署 へ		GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时 您享有免费时长余额: 2小时
纯离线服务		GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时
端云协同服务	上一步 开	始训练 取消		22. 400 o.l.t.t.s.#J		
2011年1月1日			■ 〒4.02//」 単当前GPU P4卡刺奈小时包	段源:138.2小时 * P点		

① 选择模型

选择此次训练的模型

② 添加数据

添加训练数据

- 先选择数据集,再按分类选择数据集里的图片,可从多个数据集选择图片
- 训练时间与数据量大小有关,以实际训练时长为准

Tips :

- 如只有1个分类需要识别,或者实际业务场景所要识别的图片内容不可控,可以在训练前勾选"增加识别结果为[其他]的默认分类"。勾选后,模型会将与训练集无关的图片识别为"其他"
- 如果同一个分类的数据分散在不同的数据集里,可以在训练时同时从这些数据集里选择分类,模型训练时会合并分类名称相同的图片

添加自定义验证集*AI模型在训练时,每训练一批数据会进行模型效果检验,以某一张验证图片作为验证数据,通过验证结果反馈去调节训练。 可以简单地把AI模型训练理解为学生学习,训练集则为每天的上课内容,验证集即为每周的课后作业,质量更高的每周课后作业能够更好的指导 学生学习并找寻自己的不足,从而提高成绩。同理AI模型训练的验证集也是这个功效。

注:学生的课后作业应该与上课内容对应,这样才能巩固知识。因此,验证集的标签也应与训练集完全一致。

添加自定义测试集*如果学生的期末考试是平时的练习题,那么学生可能通过记忆去解题,而不是通过学习的方法去做题,所以期末考试的试题 应与平时作业不能一样,才能检验学生的学习成果。那么同理,AI模型的效果测试不能使用训练数据进行测试,应使用训练数据集外的数据测 试,这样才能真实的反映模型效果。

注:期末考试的内容属于学期的内容,但不一定需要完全包括所学内容。同理,测试集的标签是训练集的全集或者子集即可。

配置数据增强策略

深度学习模型的成功很大程度上要归功于大量的标注数据集。通常来说,通过增加数据的数量和多样性往往能提升模型的效果。当在实践中无法 收集到数目庞大的高质量数据时,可以通过配置数据增强策略,对数据本身进行一定程度的扰动从而产生"新"数据。模型会通过学习大量 的"新"数据,提高泛化能力。

你可以在「默认配置」、「手动配置」、「自动数据增强」3种方式中进行选择,完成数据增强策略的配置。

默认配置

如果你不需要特别配置数据增强策略,就可以选择默认配置。后台会根据你选择的算法,自动配置必要的数据增强策略。

手动配置

EasyDL提供了大量的数据增强算子供开发者手动配置。你可以通过下方的算子功能说明或训练页面的效果展示,来了解不同算子的功能:

算子名	功能				
ShearX	剪切图像的水平边				
ShearY	剪切图像的垂直边				
TranslateX	按指定距离(像素点个数)水平移动图像				
TranslateY	按指定距离(像素点个数)垂直移动图像				
Rotate	按指定角度旋转图像				
AutoContrast	自动优化图像对比度				
Contrast	调整图像对比度				
Invert	将图像转换为反色图像				
Equalize	将图像转换为灰色值均匀分布的图像				
Solarize	为图像中指定阈值之上的所有像素值取反				
Posterize	减少每个颜色通道的bits至指定位数				
Color	调整图像颜色平衡				
Brightness	调整图像亮度				
Sharpness	调整图像清晰度				
Cutout	通过随机遮挡增加模型鲁棒性,可设定遮挡区域的长宽比例				



自动数据增强 在训练方式选择「精度提升配置包」选项后,此处数据增强策略提供「自动数据增强」选项。自动数据增强算法会根据您数据的特性,自动选择数据增强算子。使用付费机型训练的用户请注意,自动数据增强算法可能会增加模型训练时间。

模型训练完成后,可在「我的模型-查看版本配置」中,查看配置记录:

【图像分类】zh_h	andle 模型ID	: 10064				🚍 训练 🖒 历史版本 🔲 删除
部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作
公有云API	V7	训练完成	未申请	未发布	top1准确率100.00% top5准确率100.00% 完整评估结果	<u> 查看版本配置</u> 申请发布 校验

配置建议

算子的配置建议贴合实际场景。

比如,数字识别的数据集中,因为对数字的旋转很有可能导致错误样本的产生,所以不建议对数字数据集进行旋转操作。再比如,检测数据集 中,如果标注量比较少,就可以通过随机平移的算子增强数据集,模型也更容易学习到目标物体的平移不变性。

③ 训练配置

- 可选择「公有云API」、「EasyEdge本地部署」
- 不知道如何选择?请参考如何选择部署方式

选择设备

- 如果您选择了「EasyEdge本地部署」,请根据实际部署设备选择设备
- 如果您选择了「公有云API」,则可按需选择训练方式

增量训练*增量训练:在模型迭代训练时,用户在原训练数据上增加了训练数据,可通过加载原训练数据训练的模型参数进行模型训练。这样可 让模型收敛速度变快,训练时间变短,同时在数据集质量较高的情况下,可能获得的模型效果也会更好。

- 注:1. 仅可选择同一部署方式下的训练的模型作为基准模型版本
- 2. 增量训练时所选择的标签应完全包含基准模型的标签。例如基准模型训练的标签为"A"、"B",那么增量训练时的标签至少有"A"、"B",不能 只有"A",不能为"C",可以是"A"、"B"、"C"。

训练方式 EasyDL目前提供完全免费的「常规训练」,以及限时免费的「精度提升配置包」两种选项。

- 「常规训练」包括EasyDL历史提供的「高精度」、「高性能」等模型选择,以及常规的模型训练配置
- 「精度提升配置包」选用百度自有超大规模预训练模型,让模型有更好的精度效果。并提供按云调用时延选择网络模型的形式,根据您实际应用场景需求,选择更合适的模型。另外,EasyDL会持续在「精度提升配置包」中新增提升模型精度效果的配置策略,敬请期待。

自动超参搜索*自动超参搜索目前仅在精度提升配置包的选项下提供。选择开启自动超参搜索后,算法会多次实验,自动搜寻出适合模型训练的 各种参数,来达到高精度的模型效果。

注:开启自动超参搜索后会增加3倍以上的训练时间,请根据实际需求考虑后选择

高级训练配置*高级训练配置开关默认关闭,建议对深度学习有一定了解的用户根据实际情况考虑使用。高级训练配置目前提供「输入图片分辨率」、「epoch」、「数据不平衡优化」三个配置项

- 输入图片分辨率:可以根据具体应用场景选择输入图片分辨率,如目标主体在图片中较小,就可适当增加输入图片分辨率,增强目标在数据层 面的特性。推荐值为该类算法任务输入图片分辨率普遍最优值。
- epoch:训练集完整参与训练的次数。如有训练数据集较大,模型训练不充分,模型精度较低的情况,可适当设置较大epoch值(大于100), 使模型训练更完整。
- 数据不平衡优化:适用于不同分类图片量差异较大的情况。当不同分类之间图片数量差异超过10倍以上时,建议开启。开启后可提升模型准确率及泛化能力

选择算法

不同的部署方式下,可以选择不同的算法。每个算法旁边有一个小问号,可以查看详细说明。

例如:选择「公有云API」后,可以在「高精度」、「高性能」、「AutoDL Transfer」3种算法中选择。鼠标移动到「AutoDL Transfer」右侧的问 号上,可以看到对AutoDL算法的详细说明。

- 高精度模型在识别准确率上表现较好,但在识别速度上表现较弱。高性能模型反之。
- 如果你已从AI市场购买了模型算法,也可以基于已购模型的算法训练:前往AI市场购买>
- 本地部署的用户可在算法推理性能大表中查看具体硬件上评测的性能信息

③ 添加数据

添加训练数据

- 先选择数据集,再按分类选择数据集里的图片,可从多个数据集选择图片
- 训练时间与数据量大小有关,以实际训练时长为准

Tips :

- 如只有1个分类需要识别,或者实际业务场景所要识别的图片内容不可控,可以在训练前勾选"增加识别结果为[其他]的默认分类"。勾选后,模型会将与训练集无关的图片识别为"其他"
- 如果同一个分类的数据分散在不同的数据集里,可以在训练时同时从这些数据集里选择分类,模型训练时会合并分类名称相同的图片

添加自定义验证集*AI模型在训练时,每训练一批数据会进行模型效果检验,以某一张验证图片作为验证数据,通过验证结果反馈去调节训练。 可以简单地把AI模型训练理解为学生学习,训练集则为每天的上课内容,验证集即为每周的课后作业,质量更高的每周课后作业能够更好的指导 学生学习并找寻自己的不足,从而提高成绩。同理AI模型训练的验证集也是这个功效。

注:学生的课后作业应该与上课内容对应,这样才能巩固知识。因此,验证集的标签也应与训练集完全一致。

添加自定义测试集*如果学生的期末考试是平时的练习题,那么学生可能通过记忆去解题,而不是通过学习的方法去做题,所以期末考试的试题 应与平时作业不能一样,才能检验学生的学习成果。那么同理,AI模型的效果测试不能使用训练数据进行测试,应使用训练数据集外的数据测 试,这样才能真实的反映模型效果。

注:期末考试的内容属于学期的内容,但不一定需要完全包括所学内容。同理,测试集的标签是训练集的全集或者子集即可。

配置数据增强策略

深度学习模型的成功很大程度上要归功于大量的标注数据集。通常来说,通过增加数据的数量和多样性往往能提升模型的效果。当在实践中无法 收集到数目庞大的高质量数据时,可以通过配置数据增强策略,对数据本身进行一定程度的扰动从而产生"新"数据。模型会通过学习大量 的"新"数据,提高泛化能力。

你可以在「默认配置」、「手动配置」、「自动数据增强」3种方式中进行选择,完成数据增强策略的配置。

默认配置

如果你不需要特别配置数据增强策略,就可以选择默认配置。后台会根据你选择的算法,自动配置必要的数据增强策略。

手动配置

EasyDL提供了大量的数据增强算子供开发者手动配置。你可以通过下方的算子功能说明或训练页面的效果展示,来了解不同算子的功能:

算子名	功能
ShearX	剪切图像的水平边
ShearY	剪切图像的垂直边
TranslateX	按指定距离(像素点个数)水平移动图像
TranslateY	按指定距离(像素点个数)垂直移动图像
Rotate	按指定角度旋转图像
AutoContrast	自动优化图像对比度
Contrast	调整图像对比度
Invert	将图像转换为反色图像
Equalize	将图像转换为灰色值均匀分布的图像
Solarize	为图像中指定阈值之上的所有像素值取反
Posterize	减少每个颜色通道的bits至指定位数
Color	调整图像颜色平衡
Brightness	调整图像亮度
Sharpness	调整图像清晰度
Cutout	通过随机遮挡增加模型鲁棒性,可设定遮挡区域的长宽比例



自动数据增强 在训练方式选择「精度提升配置包」选项后,此处数据增强策略提供「自动数据增强」选项。自动数据增强算法会根据您数据的特性,自动选择数据增强算子。使用付费机型训练的用户请注意,自动数据增强算法可能会增加模型训练时间。

模型训练完成后,可在「我的模型-查看版本配置」中,查看配置记录:

【图像分类】zh_h	andle 模型ID	0: 10064				吕 训练	🕒 历史版本	间 删除
部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作		
公有云API	V7	训练完成	未申请	未发布	top1准确率100.00% top5准确率100.00% 完整评估结果	查看版本配置申订	青发布 校验	

配置建议

算子的配置建议贴合实际场景。

比如,数字识别的数据集中,因为对数字的旋转很有可能导致错误样本的产生,所以不建议对数字数据集进行旋转操作。再比如,检测数据集中,如果标注量比较少,就可以通过随机平移的算子增强数据集,模型也更容易学习到目标物体的平移不变性。

④ 训练模型

点击「开始训练」,训练模型。

- 模型训练过程中,可以设置训练完成的短信提醒并离开页面。
- 平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。
- 训练时间与数据量大小有关,以实际训练时长为准。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失败、训练状态为排队中时长均不纳入收费时长。

心 图像分类模型效果评估

可通过模型评估报告或模型校验了解模型效果:

- 模型评估报告:训练完成后,可以在列表中看到模型效果,以及详细的模型评估报告。
- 模型在线校验:可以在左侧导航中找到【模型校验】,在线校验模型效果。校验功能示意图:



模型评估报告

整体评估

在这个部分可以看到模型训练整体的情况说明,包括基本结论、准确率、F1-score等。这部分模型效果的指标是基于训练数据集,随机抽出部分数据不参与训练,仅参与模型效果评估计算得来。所以**当数据量较少时(如图片数量低于100个),参与评估的数据可能不超过30个,这样得出的模型评估报告效果仅供参考,无法完全准确体现模型效果。**

查看模型评估结果时,需要思考在当前业务场景,更关注精确率与召回率哪个指标。是更希望减少误识别,还是更希望减少漏识别。前者更需要 关注精确率的指标,后者更需要关注召回率的指标。同时F1-score可以有效关注精确率和召回率的平衡情况,对于希望准确率与召回率兼具的场 景,F1-score越接近1效果越好。评估指标具体的说明如下。

F1-score: 对某类别而言为精确率和召回率的调和平均数,评估报告中指各类别F1-score的平均数

准确率: 基于随机测试集进行计算,为正确分类的样本数与总样本数之比

注意:若想要更充分了解模型效果情况,建议发布模型为API后,通过调用接口批量测试,获取更准确的模型效果。

【整体评估 test02 V2效果优异 ,建议针对识别错误的图片示	「例继续优化模型」	效果。 如何优化效果?	
准确率 ⑦ 95.3%		F1-score ⑦ 95.8%	
Top2	98.4%	精确率 ⑦	96.0%
Тор3	100.0%	김回率 ⑦	96.1%
Тор4	100.0%		
Тор5	100.0%		

top1-top5准确率

对于每一个评估的图片文件,模型会给根据置信度高低,依次给出top1-top5的识别结果,其中top1置信度最高,top5的置信度最低。那么top1的 准确率值是指对于评估标准为"top1结果识别为正确时,判定为正确"给出准确率。top2准确率值是指对于评估标准为"top1或者top2只要有一个命 中正确的结果,即判定为正确"给出的准确率。……以此类推。

模型调优建议在模型评估中,EasyDL将会通过智能算法对误识别的样本进行归因分析,可推断出误识别的样本对某个模型评估指标的具体影响 以及影响程度,并提供对应优化的方案。同时还可针对某个具体表现不好的标签进行归因分析,针对性优化识别效果

模型调优建议

归因粒度	基于整个模型	基于单个标签
------	--------	--------

序号	受影响指标	影响程度	根因分析	调优对策
1	F1-Score	Φ	"高宽比"对"F1–Score"的效果有"一 定"影响,不同特征区间的"F1–Score" 方差达到"0.0352"	在【添加数据】->【数据增强策略】 中配置"ShearX,ShearY"进行增强。
2	F1-Score	Φ	"分辨率"对"F1–Score"的效果有"一 定"影响,不同特征区间的"F1–Score" 方差达到"0.0321"	在【添加数据】->【数据增强策略】 中配置"ShearX,ShearY"进行增强。
3	F1-Score	Φ	"色偏"对"F1–Score"的效果有"一定" 影响,不同特征区间的"F1–Score"方差 达到"0.0205"	在【添加数据】->【数据增强策略】 中配置"Color,Posterize"进行增强。
4	F1-Score	Φ	"亮度"对"F1–Score"的效果有"一定" 影响,不同特征区间的"F1–Score"方差 达到"0.0188"	在【添加数据】->【数据增强策略】 中配置"Brightness"进行增强。
5	F1-Score	Φ	"饱和度"对"F1–Score"的效果有"一 定"影响,不同特征区间的"F1–Score" 方差达到"0.018"	在【添加数据】->【数据增强策略】 中配置"Color"进行增强。

详细评估

这个部分支持查看模型识别错误的图片示例,以及使用混淆矩阵定位易混淆的分类。

识别错误图片示例

通过分标签查看模型识别错误的图片,寻找其中的共性,进而有针对性的扩充训练数据。

详细评估

按分类查看错误示例

不同分类的F1-score及对应的识别错误的图片(不包含训练时可能勾选的"其他"类识别错误的图片)



例如,你训练了一个将小番茄和樱桃分类的模型。在查看小番茄分类的错误示例时,发现错误示例中有好几张图片都是带着绿色根茎的小番茄 (与樱桃比较相似)。这种情况下,就需要在小番茄分类的训练集中,多增加一些带绿色根茎的图片,让模型有足够的数据能够学习到带根茎的 小番茄和樱桃的区别。

这个例子中,我们找到的是识别错误的图片中,目标特征上的共性。除此之外,还可以观察识别错误的图片在以下维度是否有共性,比如:图片的拍摄设备、拍摄角度,图片的亮度、背景等等。



定位易混淆分类

支持按识别错误样本量的绝对数值/相对数值查看混淆矩阵,获得具体到数据级别的精度评价信息。同时支持下载完整的混淆矩阵进行更深入的 分析。

定位易混淆分类

下方是imagenet2012_6c V6模型的混淆矩阵,每一个橙色的方格都对应一组易混淆的分类(最多展示10个易混淆的分类),点击方格即可进一步分析模型在识别该组分类时,依据的关键特征。

展示了数	8标注与模型预测不符数量;	前10的标签,点击标签可在下方查看示例图,帮助您有针对性地设计特征,使得类别更具区分性。			⊻⊤	载完整混淆矩阵	i .
序号	标签名称	误识别标签TOP5及其数量	精确率	測试集数量	召回率	f1-score	
1	n02777292	[2] n03942813 [1] n03179701	100.0%	7	57%	73%	分析热力图
2	n03942813	[1] n03188531 [2] n02777292	85.0%	17	100%	92%	
3	n03188531	[1] n03942813	100.0%	18	94%	97%	
4	n03179701	[1] n02777292	97.0%	32	100%	98%	

点击混淆矩阵中带有数字的方格,可以进一步分析对应易混淆分类的示例图,非常直观地对影响模型精度的因素进行判断。



示例图共分为3组,假设选定的易混淆分类是「A分类被误识别为B分类」:

- 1、实际标注为A分类,但被模型识别为B分类的所有图片
- 2、被模型准确识别为A分类的图片
- 3、被模型准确识别为B分类的图片

3组图片均支持查看原图与热力图。其中,热力图可以进一步地解释模型的决策依据,在整图范围内给出影响模型识别结果的像素重要程度,热力图的色度如下图,颜色越靠右代表像素重要性越高。

以猫狗分类为例:

如果分类为狗,但被模型误识别为猫的示例图里出现的狗,与模型准确识别为狗的示例图的图片性状大不一致,那可以判断出数据集对某种性状的狗的识别能力不足,需要继续增加该性状的狗的数据。

如下猫狗的图片,模型给出预测结果为狗,通过热力图的查看,可以看到支持模型给出狗的预测结果的决策依据正是图中狗脸附近像素区域。



掘人工易错标图片」中即可快速检查到这些图片,并同时完成标签修改。



心 图像分类模型如何提升效果

一个模型很难一次性就训练到最佳的效果,可能需要结合模型评估报告和校验结果不断扩充数据和调优。

为此我们设计了模型迭代功能,即当模型训练完毕后,会生成一个最新的版本号,首次V1、之后V2……以此类推。可以通过调整训练数据和算法,多次训练,获得更好的模型效果。

注意:如果模型已经是上线状态(包括已付费的模型服务),依然支持模型迭代。只需要在训练完毕后发布新的版本,就可以获得更新后的模型 服务。

想要提升模型效果,可以尝试以下方法:

尝试不同的训练配置 可前往训练配置页面尝试不同的配置组合,因不同数据集在不同的算法上可能表现不一致,所以建议您多尝试不同的算法选型后综合挑选精度最高的模型使用,你可以选择如下的配置项:

- 精度提升配置包
- 平衡精度性能
- 自动超参搜索

EasyDL开发平台	训练模型					操作文档	常见问题 新手教科	
125,619/02	训练配置							
智能标注	•部署方式:	公有云部署 Eas	限时免费 yEdge本地部署 浏览器/小程序部署 如何选择部署方式?					
□ 数据增强	 训练方式: 	限时免费						
○ 数据回流 ^	01205334	精度提升配直包	帯がい時					
云服务回流	选择算法:	● 百度文心-CV VIMER-CAE预训练模型-通用场景 ② 百度文心-CV VIMER-UFO预训练模型-行业场景 ③						
摄像头采集	选择网络:	1000 ● ● ● 精度最高	500 300 200 100 50 日本 日本 日	· ② 指最高				
模型管理	白动现象地步。							
◎ 模型训练	4							
模型校验	训练环境:	名称	规格	算力	速度比例	价格		
模型校验	训练环境:	名称	规格 TeslaGPU P4 8G思存单卡 12核CPU 40G内存	算力 5.5 TeraFLOPS	速度比例	价格 单卡¥4.02/小时		
模型校验 ② 模型校验 模型服务	训练环境:	名称 ● GPU P4	规格 TeslaGPU_P4_80置存单卡_12核CPU_400内存	算力 5.5 TeraFLOPS	速度比例 1	价格 单卡¥4.02/小时 已购小时包余额:	138.2小时*节点	
 概型校验 使型校验 模型服务 公 公有云服务 ^ 	训练环境:	名称	双格 TesiaGPU_P4_80显存单卡_12核CPU_400内存 TesiaGPU_P40_240显存单卡_12核CPU_400内存	算力 5.5 TeraFLOPS 12 TeraFLOPS	速度比例 1 1.47	价格 单卡¥4.02/小时 已购小时包余额: 单卡¥21.60/小时 您享有免费时长线	138.2小时*节点 138: 2小时	

② 免训练迭代模式

℗ 免训练迭代模式

整体介绍 免训练迭代模式是EasyDL针对于"需要高频迭代模型,但模型训练时间成本太高"的用户使用场景推出的新型模型迭代模式,在常规模型 训练完成后开启免训练迭代模式,即可在「免训练模式数据底库」中通过增删数据来迭代模型的预测能力。值得一提的是,该模式下新增一类标 签数据也可以短时间内马上获得新标签预测能力的模型

注: 1. 免训练迭代模式仅适合短期内快速获得具备一定精度模型的应用场景,如需获得效果更好的模型,请提升数据量和数据丰富度后进行 模型训练

 免训练迭代模式会根据模式开启时的模型和数据生成预测基础模型,基础模型影响后续的预测能力,所以在初次打开免训练迭代模式时, 尽量保证当时的模型和数据质量

使用流程

①模型训练 请创建模型并训练训练模型,在模型训练完成后请勿删除原训练数据

模型列表					操作文档 常	见问题 新手教程 提交:	I
创建模型							
【图像分类】猫和狗test 🗹 棏	莫型ID: 148763				免训练迭代模式 now 吕 训练	④ 历史版本 茴 删除	②启动免训练迭代模式 模
部署方式	版本	训练状态	服务状态	模型效果	操作		
本地部署-通用小型设备	V3	● 训练完成	-	top1准确率: 98.84% top5准确率: 100.00% 完整评估结果	查看版本配置 申请发布		
型训练完成之后,打	T开免训练:	迭代模式开关	,生成「免」	川练模式数据底库」			
注:此过程可能需要 ^{模型列表}	要一段时间	,具体时长与	模型训练相差	差无几	操作文档 常见	问题 新手教程 提交工单	1
创建模型							
【图像分类】猫和狗test 🖸 模	型ID: 148763				免训练迭代模式 поч 吕 训练	⑤ 历史版本 前 删除	
部署方式	版本	训练状态	服务状态	模型效果	操作		③免训练极速迭代 免训练
本地部署-通用小型设备	V3	• 训练完成	-	top1准确率: 98.84% top5准确率: 100.00% 完整评估结果	查看版本配置 申请发布		
免训 免训练版本部署方式	版本	数据底库状态	服务状态	模型效果	操作		
-	FastIteration	● 生成完成	-	top1准确率: 98.52% top5准确率: 100.00% 完整评估结果	免训练模式数据底库 校验 申请发布	ا م	
模式数据底库生成完	完成后,点 ;	击进入数据底	库		操作文档 常见问	题 新手教程 提交工单	
创建模型							
【图像分类】猫和狗test 🖸 横	型ID: 148763				全训练送代模式 (new) 🗄 训练 🤇	⊙ 历史版本 □ 删除	
部署方式	版本	训练状态	服务状态	模型效果	操作		在数据底库中,您可以点
本地部署-通用小型设备	V3	• 训练完成	-	top1准确率: 98.84% top5准确率: 100.00% 完整评估结果	查看版本配置 申请发布		
(免训) 免训练版本部署方式	版本	数据底库状态	服务状态	模型效果	操作		
-	FastIteration	• 生成完成	-	top1准确率: 98.52% top5准确率: 100.00% 完整评估结果	免训练模式数据底库 校验 申请发布	(
击「扩充数据底库」	来批量添加	加新标签数据	,来新增模型	型的预测能力。也可	以通过「批量删除」操作	,来删除噪声数	据,提升模型预测精度。
模型列表 > 猫和狗test/FastItera	ation > 免训练模式	数据底库			操作文档 常见	问题 新手教程 提交工的	Ŧ



说明:在数据底库中的增删操作,不论数据量的大小都会引起模型更新,建议您确定需要调整的数据后一并操作,避免重复操作,重复等待

模型发布

②图像分类模型发布整体说明

训练完成后,可将模型部署在公有云服务器、通用小型设备、本地服务器,或直接购买软硬一体方案,灵活适配各种使用场景及运行环境

公有云在线服务

训练完成的模型存储在云端,可通过独立Rest API调用模型,实现AI能力与业务系统或硬件设备整合

具有完善的鉴权、流控等安全机制, GPU集群稳定承载高并发请求

支持查找云端模型识别错误的数据,纠正结果并将其加入模型迭代的训练集,不断优化模型效果

纯离线服务 训练完成的模型整体打包为纯离线服务,可下载在本地稳定调用。纯离线服务按部署硬件芯片不同分为本地服务器部署、通用小型设备部署。为了提供更好的算法与硬件推理效果,EasyDL提供软硬一体方案部署。纯离线服务的整体支持与评测信息可详见算法与性能评测大表

本地服务器部署

可将训练完成的模型部署在私有CPU/GPU服务器上,支持服务器API和服务器SDK两种集成方式

模型服务性能表现更好,适用于对性能要求较高的场景,例如工业质检、流水线产品分拣等

通用小型设备

训练完成的模型被打包成适配智能硬件的SDK,可进行设备端离线计算。满足推理阶段数据敏感性要求、更快的响应速度要求

支持iOS、Android、Linux、Windows四种操作系统,基础接口封装完善,满足灵活的应用侧二次开发

软硬一体方案

高性能硬件与模型深度适配,多种方案可选。可应用于工业分拣、视频监控等多种设备端离线计算场景,让离线AI落地更轻松。了解更多

端云协同服务

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

「云管理,端计算」的端云协同服务,具体包括:

在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新

断网状态下模型离线计算(http服务,可调用与公有云API功能相同的接口)

联网状态下在平台管理设备运行状态、资源利用率

心 公有云部署

⊙ 如何发布图像分类API

训练完毕后可以在左侧导航栏中找到【发布模型】,依次进行以下操作即可发布公有云API:

- 选择模型
- 选择部署方式「公有云部署」
- 选择版本
- 自定义服务名称、接口地址后缀
- 申请发布

申请发布后,通常的审核周期为T+1,即当天申请第二天可以审核完成。如果需要加急、或者遇到莫名被拒的情况,请在百度智能云控制台内提 交工单反馈。

发布模型界面示意:

模型中心	发布模型						
我的模型 创建模型	选择模型: 部署方式:	img_class_new 公有云部著	× ×	标准接口规范 标准接口请求	參考 参考说明:		
训练模型	选择版本:	V3	~	字段名称	必须	类型	说明
校验模型发布模型	 服务名称: 第日地址: 	https://ain.baidubce.com/rpc/	2 Olai custom/ut/classi	image	是	string	图像数据, base64编码, 要求 base64编码后大小不超过4M, 最 短边至少15px, 最长边最大 4096px 支持ing/pmp/kmp/km
数据中心	- 18 LI ABAL -	fication/	2.0/al_custon/vi/classi	top_num	否	number	返回分类数量,默认为6个
我的数据集	其他要求:	若接口无法满足您的需求,请 500汉字以内	描述希望解决的问题,	标准接口响应。	字段说明:		
云服务调用数据				字段名称	必须	类型	说明
			0/500	log_id	是	number	唯一的log id,用于问题定位
		✓ 同意云服务调用数据管理服务	各条款并开通服务	error_code	否	number	错误码,当请求错误时返回
		提交申请		error_msg	否	string	错误描述信息,当请求错误时返回
				results	否	array(object)	分类结果数组
				+ name	否	string	分类名称
				+ score	否	number	置信度

接口赋权

在正式使用之前,还需要做的一项工作为接口赋权,需要登录EasyDL控制台中创建一个应用,获得由一串数字组成的appid,然后就可以参考接口文档正式使用了

全局 ~							Q	合作伙伴	ŢΦ	яe	帮助文档	企业资源	财务	•
ま く 返回Eas	yDL总范	产品服务 /	EasyOL图像 - 应用列表											
➤ EasyDL图像		应用列	山表											
18 公有云部著	^	+ =	建应用											
白・広用列表		_	应用名称	AppID	API Key	Secret Key	2180	时间		操作				
 		1	识别编和的	15313490	gRHegw68kpnym0H6K7GI6wUd	<u>8</u> *	2018	-12-31 17:24	:27	投表 管	R 1878			
日 · 用量次日 日 · 用素文格		2	TRY	14676113	75a5pbP7mhpBWNION2NYoYcX	B#	2018	-11-05 11:35	54	投表 管	8 808			
 101 11 11 12 13 14 15 16 <li< th=""><td></td><td>3</td><td>皮皮野</td><td>14403281</td><td>i6HhGApGanhWuYex4Oel2uG8</td><td>······ 显示</td><td>2018</td><td>-10-11 19:55</td><td>26</td><td>报表 管</td><td>8 808</td><td></td><td></td><td></td></li<>		3	皮皮野	14403281	i6HhGApGanhWuYex4Oel2uG8	······ 显示	2018	-10-11 19:55	26	报表 管	8 808			
EasyEdge&	地球者 ~												< 1	>

同时支持在「公有云服务管理」-「权限管理」中为第三方用户配置权限

示意图如下:

Ŷ	全局 ~							
	✓ 返回EasyDL总览	产品服务 / EasyDL图像 - 权限管理						
>	EasyDL图像	权限管理	权限管理					
88	公有云部署 へ	■服务权限管理						
đĐ	• 应用列表	支持为您或者他人的AppID配置可用服务权限。						
e	 权限管理 	序号 服务名称	相关操作					
þ	。 用量统计	1 1000	配置权限					
ж	。 技术文档	2 test	配置权限					
53	• 售卖服务	3 cherry	配置权限					
Þ	EasyEdge本地部署 >	4 horse的服务名称	配置权限					
		5 sea	配置权限					
		6 手势识别	配置权限					

Ú	全局 🗸	
	✔ 返回EasyDL总览	产品服务 / EasyDL图像 - 权限管理 / 配置权限
>	EasyDL图像	配置权限
	公有云部晋 へ ・ 应用列表 ・ 权限管理 ・ 用量统计 ・ 技术文档	服务名称: 1000 可用APP ID: 14403281 皮皮⊈F × 14676113 TRY × 15313490 识别猫和狗 × 请添加 ┿
	• 吉兴版穷 EasyEdge本地部署 >	保守 取消

心 图像分类API调用文档

本文档主要说明定制化模型发布后获得的API如何使用,如有疑问可通过以下方式联系我们:

- 在百度智能云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

接口描述

基于自定义训练出的图像分类模型,实现个性化图像识别。模型训练完毕后发布可获得定制化图像分类API

接口鉴权

1、在EasyDL控制台创建应用

♥ 全局 ~								Q	₽	?	읆	۲	Q
EasyDL定制训练平台	产品服务 / EasyD	DL定制训练平台 - 应	用列表 / 创建应用										
概览	创建新应用]											
应用列表	 · 应田夕段 · 	- to at											
监控报表	177.142.62	test											
技术文档	•应用类型:	游戏娱乐					~						
云服务权限管理	*接口选择:	您可以为应用勾	选多个接口权限,使此应	用可以请求已	勾选的接口服务	277							
离线SDK管理		EasyDL	 ✓ 百美椅子训练 ✓ 识别眼睛和嘴巴 	 ✓ 椅子前 ✓ 声音分 	貢色识别 计类	✓ aa ✓ 商品检测							
		➡ 百度语音											
		➡ 文字识别											
		➡ 人脸识别											
		➡ 自然语言处:	里										
		➡ 内容审核]										
		🛨 UNIT 📘											
		🛨 知识图谱											
		🛨 图像识别											
		➡ 智能呼叫中/	Ċ.										
		➡ 图像搜索											
		➡ 人体分析											
	•应用描述:	简单描述一下 点,请控制在	您使用人工智能服务的应 500字以内	用场景,如开》		几,需要检测人制	验关键						

2、应用详情页获取AK SK

♥ 全局 ~					Q		₽	?	몲	۲	Q~
EasyDL定制训练平台	产品服务 / EasyDL定制训练平台 - 应用列表 / 应用	详情									
概览	应用详情										
应用列表	编辑 查看文档	輪組 查看文档									
监控报表	应用名称	AppID		API Key			Secr	et Key			
技术文档	test2	14293933		bZsfqO2fUdHTDflxqF1	11mkTq			*** 显示			
云服务权限管理 离线SDK管理	EasyDL定制化训练服务:										

请求说明

请求示例

请求URL:请首先进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

参数	值
input_type	当取值为 url 时,需在请求参数中传入图片的URL string
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,与其他图像识别服务不同的是定制化图像识别服务以json方式请求。

Body请求示例:

```
{
"image": "<base64数据>",
"top_num": 5
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大4096px,支持 jpg/png/bmp格式 注意请去掉头部
top_nu m	否	number	-	返回分类数量,默认为6个
url	否	string	-	如果在请求URL参数中增加"input_type=url",则该参数必传,否则"image"参数必传。参数内容为URL string,用户需确保该string是有效的图片URL,否则会下载失败

请求代码示例

提示一:使用示例代码前,请记得替换其中的示例Token、图片地址或Base64信息。

提示二:部分语言依赖的类或库,请在代码注释中查看下载地址。

РНР	
JAVA	
Python3	
C++	

php</th
/**
* 发起http post请求(REST API), 并获取REST请求的结果
* @param string \$url
* @param string \$param
* @return - http response body if succeeds, else false.
*/
<pre>function request_post(\$url = '', \$param = '')</pre>
{
if (empty(\$url) empty(\$param)) {
return false;
}
\$postUrl = \$url;
\$curlPost = \$param;
// 初始化curl
<pre>\$curl = curl_init();</pre>
curl_setopt(\$curl, CURLOPT_URL, \$postUrl);

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	分类结果数组
+name	否	string	分类名称
+score	否	number	置信度

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度智能云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求,如果持续出现此类错误,请在百度智能云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度智能云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度智能云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度智能云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度智能云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有 疑问请在百度智能云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或者代码格式有误。有疑问请在百度智能云控制台内提交工 单反馈
336003	Base64解码失败	图片/音频/文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头 部。有疑问请在百度智能云控制台内提交工单反馈
336004	输入文件大小不合法	图片超出大小限制,图片限4M以内,请根据接口文档检查入参格式,有疑问请在百度智能云控制台内提交工 单反馈
336005	图片解码失败	图片编码错误(非jpg,bmp,png等常见图片格式),请检查并修改图片格式
336006	缺失必要参数	image字段缺失 (未上传图片)
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度智能云控制台 内提交工单反馈
336006	缺失必要参数	image字段缺失 (未上传图片)

の 批量预测

批量预测

部署方式为公有云服务的模型训练完成后即可使用批量预测服务。批量预测支持您将已导入百度云BOS数据存储的图片数据进行一次预测任务, 任务完成后会将预测结果保存至百度云BOS数据存储服务中。您可按照如下步骤使用批量预测服务:

①新建批量预测任务 首次使用批量预测功能会提示您开通BOS权限,请按照说明指引开通。

图像分类模型	.≡	公有云部署 > 批	:量预测					操作文档	常见问题	新手教程	提交工单	
部 总览		批量预测是一种进	批量 茨利尼一种进行批量数据推理的方式。用户可以上传批量数据进行推理,在数据处理完成后自动停止									
前 模型中心												
我的模型		101倍及宪贺束船	\$%费策略 开递付费 操作文档									
创建模型												
训练模型		新建预测	J									
校验模型		名称		任务ID	任务状态	创建时间 👙	完成时间 👙	操作				
发布模型												
☑ EasyData数据服务												
数据总览												
标签组管理						-						
在线标注												
智能标注						暂无数据						
云服务数据回流						请 新建预测任务 后查看						
摄像头数据采集												
目 公有云服务									< 1	> 10条	i/页 ~	
在线服务												

请按照页面说明填写相关信息,选择预测模型以及所需预测的数据地址,并选择对应的资源配置。每个账户享有2个小时CPU算力的免费额度

公有云部署 > 排	北量预测 〉 新建预测					操作文	档 常见问题	新手教程	提交工单
MACT. DOING									
*名称	最多输入20个文字								
*描述	500个字以内								
任务配置									
*模型名称	请选择模型								
*模型版本	请选择模型版本								
数据集配置									
数据输入路径	选择数据输入路径		0						
结果输出路径	选择结果输出路径	D							
资源配置									
运行环境	名称	规格		算力	速度比例	价格			
	O CPU	CPU_4核_16G		-	0.7	单卡¥0.03/分钟	(2小时*节点免费)	
批量预测费用 <mark>免</mark> 提交	费 当前CPU卡剩余免费 取消	}资源:2(小时 * 节点)	由于您尚未开通付费,为保证训练任务能顺利进行	亏,建议您开通计费 ,	干通付费				

②等待任务运行完成 可根据任务状态查看当前已成功预测的图片数量

图像分类模型	Ē	公有云部署 > 批量预测					操作文档 常见问题 新手教程 提交工单
器总算 證 模型中心 我的模型		批量预测是一种进行批量数据推明 价格及免费策略 开通付费	罪的方式。用户可以上传批量数据 操作文档	进行推理,在数据处理5	电成后自动停止		
创建模型 训练模型		新建预测					
校壇模型		名称	任务ID	任务状态	创建时间 🖕	完成时间 🍦	操作
区 EasyData数据服务		test	з	 运行中 ③ 已成 	功预测图片数据: 0		详情 终止
数据总览							< 1 > 10条/页 >
标签组管理							
在线标注							
智能标注							
云服务数据回流							
摄像头数据采集							
目 公有云服务							
在线服务							
批量預測							
₿ EasyEdge本地部署							
纯高线服务							
端云协同服务	^						
我的部署包							

③任务完成,查看数据任务完成后可在任务配置时所选择的结果保存BOS地址中查看预测结果数据

の 本地服务器部署

⊙ 如何在本地服务器部署

训练完毕后,可以选择将模型通过「纯离线服务」或「端云协同服务」 部署,具体介绍如下:

纯离线服务部署

可以在左侧导航栏中找到「纯离线服务」,依次进行以下操作即可将模型部署到本地服务器:

- 选择部署方式「服务器」
- 选择集成方式
- 选择模型、版本、系统和芯片
- 点击下一步

EasyDL开发平台	〈 巡回 发布新服务	操作文档 常见问题 新手教程 提交工单
	服务器 通用小型设备 专项适配硬件 3 选择部票形式 2 追写个人信息	说明: 1本地服务器部署支持将模型部署于本地的CPU、GPU服务器 上、提供APH和SDK两特集成方式: 豊着文档
端云协同服务 浏览器/小程…	集成方式 ④ SDK (API	2.本地服务器SDK:将模型封装成适配本地服务器(支持Linux 和Windows)的SDK,可集成在其他程序中运行。首次联网激 活后即可纯属线运行,占用服务器资源更少,使用方法更更活
模型应用	选择模型 showcase_图像分类 V	3.集成步骤: ① 申请SDK并在服务详情页面下载SDK → ② 在 控制台申请激活序列号 → ③ 根据开发文档集成SDK,并联网
④ 应用接入	选择版本 V1 V	激活使用。如存在设备无法联网,需要在纯高线的环境下激活 的情况 请 提及工作 联系开印
⑧ 权限管理	 送择系统和芯片 □ Δ Linux 	
🖾 调用统计	了解对应硬件 通用x86芯片 x86+英伟达GPU 飞瞵CPU	4.1入自意的項句[CREasyDL20K]]解念, 使了启续自计按 洽,不会作为其他用途使用
AI市场		Q.
◎ 我的已购模型		C
团 售卖模型	飞腾+华为Atlas3001 x86+百度昆仑-K200	
平台管理	飞瞵+百度昆仑-K200 x86+百度昆仑XPU-R200	
◎ 算力包管理	飞耦+百度昆仑-R200 海光OCU	
囯 计费管理	飞栅+零武记MLU270	

- 填写部分信息(注:个人信息的填写仅供EasyDL团队了解您,便于后续合作接洽,不会作为其他用途使用)
- 点击发布

EasyDL开发平台	< 返回 发布新服务		
ふ 本地部署 へ	服务器 通用/	小型设备 专项适配硬件	
纯离线服务	1 选择部署形式	t 填写个人信息	
端云协同服务			
浏览器/小程…	* 联系人姓名		
模型应用	* 电话	130*****766	0
④ 应用接入	* 邮箱	h*******@email.com	
⑧ 权限管理	1994 - 1994	e endidoni	
🖾 调用统计	* 职务		
AI市场	* 公司名称		
♥ 我的已购模型	* 公司所在行业	请选择行业	
I 售卖模型			
平台管理	* 业务量级	请你简要描述,每天要处理的数据量级和所需的并发数	
合 算力包管理			
囯 计费管理			0/200

① 私有API

将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷 点击「发布」后,前往「纯离线服务」等待部署包制作完成后,下载部署包,并参考文档完成集成

② 服务器端SDK

将模型封装成适配本地服务器(支持Linux和Windows)的SDK,可集成在其他程序中运行。首次联网激活后即可纯离线运行,占用服务器资源更少,使用方法更灵活

1、点击「发布」后,前往控制台申请服务器端SDK的试用序列号

2、点击「新增测试序列号」,根据模型类型选择「序列号类型」,填写「新增设备数」(所得序列号数量),点击确定即可

服务器纯离线服务管理			
API SDK			
使用说明: 1. 通过集资适配本地服务契约SDK版结使用模型:SDK支持二次开发,按CPU/GPUi设备得权使用,操作步骤如下:	4KAB		
 ○在EasyDL申请并下载SDK→0在下方申请激活序列号→○根据开发文档集成SDK,并联网激活→0离线使用 2.序列号联网激活SDK后有效期为1个月,一个测试序列号激活模型上限数为10,如需正式购买,请点由下方购买正式规权。 			
+ 购买正式授权 新培训试序列号	开发文档 工单支持		

3、离线SDK的激活和使用,请参考文档完成集成

ai 20 大脑 AI开放平台 开放 #	能力 开发平台	台 行业应用	生态合作	AI市场	开发与养
EasyDL定制AI训练平台	芯片类型	实测硬件	高精度模型	高性能模型	
搜索本产品文档内容 Q	x86-64	Xeon E5-2650 v	/4 61ms	40ms	
	Nvidia-G	BPU Tesla P4	16ms	5ms	
> 公有云部署	Atlas 30	0 Atlas 300	12ms	9ms	
~ 私有服务器部署	激活&	使用步骤			
如何在私有服务器部署	离线SDK的	り激活与使用分以下三步	7 :		
私有API集成文档	③在经典版	版控制台下载SDK、获耳	双序列号		
◇ 服务器端SDK 服务器端SDK简介	③本地运行	亍SDK,并完成首次联网	网激活		
LinuxSDK集成文档- C++	通过左	侧导航栏查看不同操作	系统SDK的开发文	て档	
LinuxSDK集成文档- Python					
Linux(Atlas)SDK集 成文档					
WindowsSDK集成文 档	SDKR	见归赵			

端云协同服务部署

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

具体使用说明请参考端云协同服务说明

本地服务器部署价格说明

EasyDL已支持将定制模型部署在本地服务器上,只需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。

如需购买永久使用授权,服务器SDK用户请在控制台点击「购买正式授权」,并按照对应步骤激活。

服务器API用户请微信搜索"BaiduEasyDL"添加小助手咨询,通过线下签订合同购买使用。

更多参考

EasyDL官网入口 EasyDL开发文档

の 纯离线SDK说明

心 纯离线SDK简介

本文档主要说明定制化模型发布后获得的服务器端SDK如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

SDK说明

图像分类服务器端SDK支持Linux、Windows两种操作系统。以下为具体的系统、硬件环境支持:

操作系统	系统支持	硬件环境要求
Linux		Intel CPU: x86_64 NVIDIA GPU: x86_64 HUAWEI Atlas 300: x86_64
Windows	64位 Windows7 及以 上	NVIDIA GPU: x86_64 环境依赖: .NET Framework 4.5 Visual C++ Redistributable Packages for Visual Studio 2013 Visual C++ Redistributable Packages for Visual Studio 2015 GPU依赖: CUDA 9.x + cuDNN 7.x

单次预测耗时参考

根据具体设备、线程数不同,数据可能有波动,请以实测为准

在算法性能及适配硬件 页面查看评测信息表

激活&使用步骤

离线SDK的激活与使用分以下三步:

- ① 下载SDK后,在控制台获取序列号
- ② 本地运行SDK,并完成首次联网激活

通过左侧导航栏查看不同操作系统SDK的开发文档

③ 正式使用

SDK常见问题

通过左侧导航栏查看不同操作系统SDK的FAQ

以下是通用FAQ,如您的问题仍未解决,请在百度智能云控制台内提交工单反馈

1、激活失败怎么办?

①可能是当前序列号已被其他设备激活使用,请核实序列号后用未被激活的序列号重新激活

②序列号填写错误,请核实序列号后重新激活

③首次激活需要联网,网络环境不佳或无网络环境,请检查网络环境后重试

④模型发布者和序列号所属账号非同一账号,如果存在这种异常建议更换账号获取有效序列号

⑤序列号已过有效期,请更换序列号后重试

⑥如有其他异常请在百度智能云控制台内提交工单反馈

の Windows集成文档

简介

本文档介绍图像分类服务器端Windows SDK的使用方法。

硬件支持:

- NVIDIA GPU (普通版,加速版)
- 操作系统支持
 - 64位 Windows 7 及以上
 - 64位Windows Server 2012及以上
- 环境依赖 (必须安装以下版本)
 - .NET Framework 4.5
 - Visual C++ Redistributable Packages for Visual Studio 2013
 - Visual C++ Redistributable Packages for Visual Studio 2015
- GPU基础版 (EasyEdge-win-x86-nvidia-gpu) 依赖 (必须安装以下版本)
 - CUDA 11.0.x + cuDNN 8.4.x(注意参照英伟达安装文档安装Zlib:http://www.winimage.com/zLibDII/zlib123dllx64.zip, 解压后将 dll_x64/zlibwapi.dll 拷贝到cuda的bin目录下) + 硬件计算能力(https://developer.nvidia.com/cuda-gpus#compute)达6.1及以上
 - CUDA 12.0.x + cuDNN 8.9.x(注意参照英伟达安装文档安装Zlib) + 硬件计算能力达7.5及以上
- GPU加速版 (EasyEdge-win-x86-nvidia-gpu-tensorrt) 依赖 (必须安装以下版本)
 - CUDA 11.0.x + cuDNN 8.4.x(注意参照英伟达安装文档安装Zlib) + TensorRT 8.4.x.x
 - CUDA 12.0.x + cuDNN 8.9.x(注意参照英伟达安装文档安装Zlib) + TensorRT 8.6.x.x
- GPU加速版 (EasyEdge-win-x86-nvidia-gpu-paddletrt) 依赖 (必须安装以下版本)
 - CUDA 11.0.x + cuDNN 8.4.x(注意参照英伟达安装文档安装Zlib) + TensorRT 8.4.3.1 + 硬件计算能力达6.1及以上
 - CUDA 12.0.x + cuDNN 8.9.x(注意参照英伟达安装文档安装Zlib) + TensorRT 8.6.1.6 + 硬件计算能力达7.5及以上
- GPU加速版 (x86-nvidia-gpu-torch)
 - CUDA 11.0.x + cuDNN 8.0.5.x
- 协议
 - HTTP
- 更详细的环境说明可参考SDK内的README.md

Release Notes | 时间 | 版本 | 说明 | | -------- | ------ | 2023-08-30 | 1.8.3 | 新增支持按实例数鉴权 | | 2023-06-29 | 1.8.2 | 优化模型 算法 || 2023-05-17 | 1.8.1 | 预测引擎升级,修复部分网络内存泄露问题 || 2023-03-16 | 1.8.0 | 预测引擎升级 || 2022-12-29 | 1.7.2 | 预测引 擎升级 | | 2022-10-27 | 1.7.1 | GPU底层引擎升级,下线基础版CUDA10.0及以下版本支持 | | 2022-09-15 | 1.7.0 | 优化模型算法;GPU CUDA9.0 CUDA10.0 标记为待废弃状态 || 2022-07-28 | 1.6.0 | 优化模型算法 || 2022-05-27 | 1.5.1 | 新增支持BML Cloud小目标检测模型 || 2022-05-18 | 1.5.0 | 修复各别机器下程序崩溃的问题 | | 2022-04-25 | 1.4.1 | EasyDL, BML升级支持paddle2模型 | | 2022-03-25 | 1.4.0 | 优化 模型算法 || 2021-12-22 | 1.3.5 | GPU基础版推理引擎优化升级;GPU加速版支持自定义模型文件缓存路径;demo程序优化环境依赖检测 || 2021-10-20 | 1.3.4 | 修复已知问题| | 2021-08-19 | 1.3.2 | 新增支持EasyDL小目标检测,新增DEMO二进制文件 || 2021-06-29 | 1.3.1 | 预测引 擎升级 || 2021-05-13 | 1.3.0 | 模型发布新增多种加速方案选择;目标追踪支持x86平台的GPU及加速版;展示已发布模型性能评估报告 || 2021-04-08 | 1.2.3 | 支持BML平台模型仓库本地上传模型 | | 2021-03-09 | 1.2.2 | 修复已知问题 | | 2021-01-27 | 1.2.1 | 新增模型支持;性能优 化;问题修复||2020-12-18|1.2.0|推理引擎升级||2020-11-26|1.1.20|新增一些模型的加速版支持||2020-10-29|1.1.19|修复已知问 题 | | 2020-09-17 | 1.1.18| 支持更多模型 | | 2020.08.11 | 1.1.17 | 支持专业版更多模型 | | 2020.06.23 | 1.1.16 | 支持专业版更多模型 | | 2020.05.15 | 1.1.15 | 更新加速版tensorrt版本,支持高精度检测 || 2020.03.13 | 1.1.14 | 支持声音分类 || 2020.02.23 | 1.1.13 | 支持多阶段模 型 || 2020.01.16 | 1.1.12 | 预测默认使用推荐阈值|| 2019.12.26 | 1.1.11 | 支持物体检测高精度算法的CPU加速版,EasyDL 专业版支持 SDK 加速版 || 2019.12.04 | 1.1.10 | 支持图像分割 || 2019.10.21 | 1.1.9 | 支持 EasyDL 专业版 || 2019.08.29 | 1.1.8 | CPU 加速版支持 || 2019.07.19 | 1.1.7 | 提供模型更新工具 || 2019.05.16 | 1.1.3 | NVIDIA GPU 支持 || 2019.03.15 | 1.1.0 | 架构与功能完善 || 2019.02.28 | 1.0.6 | 引擎功能完善 || 2019.02.13 | 1.0.5 | paddlepaddle 支持 || 2018.11.30 | 1.0.0 | 第一版 ! |

快速开始

1. 安装依赖

安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642

Visual C++ Redistributable Packages for Visual Studio 2013

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

Visual C++ Redistributable Packages for Visual Studio 2015

https://www.microsoft.com/zh-cn/download/details.aspx?id=48145

如果使用GPU版SDK,请安装CUDA + cuDNN

https://developer.nvidia.com/cuda https://developer.nvidia.com/cudnn

如果使用GPU版加速版SDK(EasyEdge-win-x86-nvidia-gpu-tensorrt),请安装TensorRT

https://developer.nvidia.com/tensorrt

根据cuda版本下载,下载后把lib目录下的所有dll,拷贝到SDK的dll目录下

注意事项

- 1. 安装目录不能包含中文
- 2. Windows Server 请自行开启,选择"我的电脑"——"属性"——"管理"——"添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启即可。

2. 运行离线SDK

解压下载好的SDK,打开EasyEdge.exe,输入Serial Num,选择鉴权模式,点击"启动服务",等待数秒即可启动成功,本地服务默认运行在

http://127.0.0.1:24401/

其他任何语言只需通过HTTP调用即可。

如启动失败,可参考如下步骤排查:	
• 百度智能云	故障排查
零门槛AI开发 平台EasyDL Windows离线SDK(公有云发 布)•问题排查	

2.1 离线鉴权 (默认鉴权模式)

首次联网激活,后续离线使用

Al ver. 1.2.0.0 core. 1	.8.1 sec. 1.5.1 – 🗆 🗙
Model	Market Tolk and Bally A 2007
Serial Num	
Host	127.0.0.1
Port	24401
鉴权模式	 ● 离线鉴权 ○ 按实例数鉴权
服务状态	服务未启动
	启动服务
	是否开机自动启动 🗌

2.2 按实例数鉴权

周期性联网激活,离线后会释放所占用鉴权,启动时请确保心跳间隔小于等于生成序列号时填写的定期确认时间

Al ver. 1.2.0.0 core. 1	.8.1 sec. 1.5.1 — 🗆 🗙
Model	[4] A.W. W. S. M. W. M. MURNER, Math.
Serial Num	
Host	127.0.0.1
Port	24401
鉴权模式	○ 离线鉴权 ④ 按实例数鉴权 心跳间隔 20 秒 🗸
服务状态	服务未启动
	启动服务
	是否开机自动启动 🗌

基于源码集成时,若需要按实例数鉴权,需要通过代码指定使用按实例数鉴权

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2); global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 20);

或通过环境变量指定

set EDGE_CONTROLLER_KEY_AUTH_MODE=2 set EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=20

2.3 序列号激活错误码
错误码	文案	描述
4001	parameters missing. 参数缺失	
4002	parameters invalid. 参数不合法	
4003	model invalid. 模型信息不合法	
4004	no more activation left. 该序列号和该设备的激活次数超上限	
4005	the serial key is out of date. 该序列号过期	
4006	the serial key has been activated. 该序列号已被其他设备激活	该序列号已被其他设备激活,不能重复激活。
4007	account invalid. 序列号不能用于其他账号的模型	序列号不能用于其他账号的模型,只能用于绑定账号的模型。
4008	serial key invalid. 序列号不合法	序列号不存在或找不到
4009	bundle id invalid. 包名不合法	
4010	product invalid. 产品不合法	如easydl的SDK使用BML的序列号来激活,会报该错误
4011	platform invalid. 平台不合法	
4012	activate too frequent. 激活太频繁	激活太频繁,请稍后再进行激活。
4013	device type and license type not match. 硬件类型和序列号类型不匹配	如使用加速版序列号激活基础版SDK会报该错误
4014	exceed max activate device num. 超过最大授权数量	
4015	technology invalid. 技术类型不合法	
4016	exceed max activate entity num. 超过最大模型数量	
4017	device invalid. 设备不合法	
4018	model invalid. 模型不合法	

3. Demo示例(以图像服务为例)

服务运行成功,此时可直接在浏览器中输入http://127.0.0.1:24401,在h5中测试模型效果。



使用说明

调用说明

Python 使用示例代码如下

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
```

params 为GET参数 data 为POST Body

 $result = requests.post('http://127.0.0.1:24401/', \ params = \{'threshold': 0.1\},$

data=img).json()

Baidu 百度智能云文档

C# 使用示例代码如下

FileStream fs = new FileStream("./img.jpg", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close(); WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close();

response.Close();

C++ 使用示例代码如下,需要安装curl

```
**include <sys/stat.h>**
**include <curl/curl.h>**
**define S_{ISREG}(m) (((m) & 0170000) == (0100000))**
**define S_{ISDIR}(m) (((m) \& 0170000) == (0040000))**
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./img.jpg";
  FILE *fp = NULL;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  }
  if (fstat(fileno(fp), &stbuf) || !S\_ISREG(stbuf.st\_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\" \n", post_data_filename);
     return -1;
  }
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
     curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
     curl_easy_setopt(curl, CURLOPT_POST, 1L);
     curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
     curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
     res = curl_easy_perform(curl);
     if (res != CURLE_OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     }
     curl_easy_cleanup(curl);
  }
  curl_global_cleanup();
  fclose(fp);
  return 0;
```

请求参数

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送图片二进制。

返回参数 | 字段 | 类型 | 取值 | 说明 | | -------- | ----- | ----- | ----- | | confidence | float | 0~1 | 分类的置信度 | | label | string | | 分类的类 别 | | index | number | | 分类的类别 |

集成指南

基于HTTP集成

通过EasyEdge.exe启动服务后,参照上面的调用说明,通过HTTP请求集成到自己的服务中

基于c++ dll集成

集成前提

Baidu 百度智能云文档

解压开的SDK包中包含src、lib、dll、include四个目录才支持基于c++ dll集成

集成方法

参考src目录中的CMakeLists.txt进行集成

基于c# dll集成

集成前提

解压开的SDK包中包含src\demo_serving_csharp、dll两个目录才支持基于c# dll集成

集成方法

参考src\demo_serving_csharp目录中的CMakeLists.txt进行集成

FAQ

1. 服务启动失败,怎么处理?

根据SDK内的README.md检查依赖是否都已正确安装

请确保相关依赖都安装正确,版本必须如下: *.NET Framework 4.5* Visual C++ Redistributable Packages for Visual Studio 2013 * Visual C++ Redistributable Packages for Visual Studio 2015

GPU依赖,版本必须如下: * CUDA 11.0.x + cuDNN 8.4.x 或者 CUDA 11.7.x + cuDNN 8.4.x

GPU加速版(EasyEdge-win-x86-nvidia-gpu-tensort)依赖,版本必须如下: * CUDA 11.0.x + cuDNN 8.4.x + TensorRT 8.4.x.x

GPU加速版(EasyEdge-win-x86-nvidia-gpu-paddletrt)依赖,版本必须如下: * CUDA 11.0.x + cuDNN 8.4.x + TensorRT 8.4.3.1

2. 服务调用时返回为空,怎么处理?调用输入的图片必须是RGB格式,请确认是否有alpha通道。

3. 多个模型怎么同时使用? SDK设置运行不同的端口,点击运行即可。

4. JAVA、C#等其他语言怎么调用SDK?

参考 https://ai.baidu.com/forum/topic/show/943765

5. 启动失败,缺失DLL? 打开EasyEdge.log,查看日志错误,根据提示处理 缺失DLL,请使用 https://www.dependencywalker.com/ 查看相应模 块依赖DLL缺失哪些,请自行下载安装

6. 启动失败,报错NotDecrypted? Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

7. 启动失败,报错 SerialNum无效

日志显示failed to get/check device id(xxx)或者 Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 C:\Users\\${用户名}\.baidu\easyedge 目录,再重新激活。

其他问题 如果无法解决,可到论坛发帖: https://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

心 Linux集成文档-C++

简介

本文档介绍EasyEdge/EasyDL的Linux CPP SDK的使用方法。

- 网络类型支持:图像分类,物体检测,图像分割,目标追踪
- 硬件支持:
 - CPU 基础版: intel x86_64 * AMD x86_64 龙芯 loongarch64 飞腾 aarch64

- CPU 加速版 Intel Xeon with Intel®AVX2 and AVX512 Intel Core Processors with AVX2 Intel Atom Processors with SSE AMD Core Processors with AVX2
- NVIDIA GPU: x86_64 PC
- 寒武纪 Cambricon MLU270
- 比特大陆计算卡SC5+
- 百度昆仑XPU K200
 - x86_64 飞腾 aarch64 百度昆仑XPU R200
 - x86_64 飞腾 aarch64
- 华为Atlas 300
- 海光DCU: x86_64 PC
- 寒武纪 MLU370 on x86_64
- 操作系统支持:Linux

根据开发者的选择,实际下载的版本可能是以下版本之一:

- EasyDL图像
 - x86 CPU 基础版
 - x86 CPU 加速版
 - Nvidia GPU 基础版
 - Nvidia GPU 加速版
 - x86 mlu270基础版
 - x86 SC5+基础版
 - Phytium MLU270基础版
 - Phytium XPU基础版
 - Phytium Atlas300I基础版
 - Hygon DCU基础版

性能数据参考算法性能及适配硬件

*intel 官方合作,拥有更好的适配与性能表现。

Release Notes

时间	版本	说明
2023.0 8.31	1.8.3	Atlas系列Soc支持语义分割模型,Atlas Cann升级到6.0.1,昆仑XPU后端推理引擎升级
2023.0 6.29	1.8.2	模型压缩能力升级
2023.0 5.17	1.8.1	支持物体检测自定义四边形模型精度无损压缩发布x86 CPU版SDK
2023.0 3.16	1.8.0	支持图像分类精度提升包本地部署
2022.1 2.29	1.7.2	模型性能优化;推理库性能优化
2022.1 0.27	1.7.1	新增语义分割模型http请求示例;升级海光DCU SDK,需配套rocm4.3版本使用;Linux GPU基础版下线适用于CUDA10.0及以 下版本的SDK:Linux GPU加读版升级推理引擎版本

Baidu 百度智能云文档

2022.0 9.15	1.7.0	Linux GPU加速版升级预测引擎;Linux GPU加速版适用于CUDA9.0、CUDA10.0的SDK为deprecated,未来移除;新增实例分 割高性能模型离线部署;性能优化
2022.0 7.28	1.6.0	Linux CPU普通版、Linux GPU普通/加速版、Jetson新增目标追踪模型接入实时流的demo
2022.0 5.27	1.5.1	CPU、GPU普通版新增支持BML Cloud小目标检测模型
2022.0 5.18	1.5.0	GPU加速版max_batch_size参数含义变更;修复GPU加速版并发预测时部分图片结果预测错误及耗时增加问题;CPU普通版预 测引擎升级;新增版本号头文件;新增飞腾Atlas300I支持,并且在EasdDL新增多种加速版本;示例代码移除frame_buffer, 新增更安全高效的safe_queue; 新增Tensor In/Out接口和Demo
2022.0 4.25	1.4.1	EasyDL, BML升级支持paddle2模型
2022.0 3.25	1.4.0	新增支持海光服务器搭配海光DCU加速卡;
2021.1 2.22	1.3.5	GPU加速版支持自定义模型文件缓存路径;新增支持飞腾MLU270服务器、飞腾XPU服务器
2021.1 0.20	1.3.4	CPU加速版推理引擎优化升级,新增支持飞腾CPU、龙芯CPU服务器、比特大陆计算卡SC5+ BM1684、寒武纪MLU270;大幅 提升EasyDL GPU加速版有损压缩加速模型的推理速度
2021.0 8.19	1.3.2	CPU、GPU普通版及无损加速版新增支持EasyDL小目标检测,CPU普通版、GPU普通版支持检测模型的batch预测
2021.0 6.29	1.3.1	CPU普通版、GPU普通版支持分类模型的batch预测,CPU加速版支持分类、检测模型的batch预测;GPU加速版支持 CUDA11.1;视频流解析支持调整分辨率;预测引擎升级
2021.0 5.13	1.3.0	新增视频流接入支持;模型发布新增多种加速方案选择;目标追踪支持x86平台的CPU、GPU加速版;展示已发布模型性能评 估报告
2021.0 3.09	1.2.1	GPU新增目标追踪支持, http server服务支持图片通过base64格式调用,EasyDL高性能检测模型和均衡检测模型CPU加速版新 增量化压缩模型
2021.0 1.27	1.1.0	EasyDL经典版分类高性能模型升级;部分SDK不再需要单独安装OpenCV
2020.1 2.18	1.0.0	1.0版本发布!安全加固升级、性能优化、引擎升级、接口优化等多项更新
2020.1 1.26	0.5.8	EasyDL经典版分类模型CPU加速版里新增量化压缩模型
2020.1 0.29	0.5.7	新增CPU加速版支持:EasyDL经典版高精度、超高精度物体检测模型和EasyDL经典版图像分割模型
2020.0 9.17	0.5.6	性能优化,支持更多模型
2020.0 8.11	0.5.5	提升预测速度;支持百度昆仑芯片
2020.0 5.15	0.5.3	优化性能,支持专业版更多模型
2020.0 4.16	0.5.2	支持CPU加速版;CPU基础版引擎升级;GPU加速版支持多卡多线程
2020.0 3.12	0.5.0	x86引擎升级;更新本地http服务接口;GPU加速版提速,支持批量图片推理
2020.0 1.16	0.4.7	ARM引擎升级;增加推荐阈值支持
2019.1 2.26	0.4.6	支持海思NNIE
2019.1 1.02	0.4.5	移除curl依赖;支持自动编译OpenCV;支持EasyDL 专业版 Yolov3; 支持EasyDL经典版高精度物体检测模型升级
2010 1		

Baidu 百度智能云文档

0.25	0.4.4	ARM引擎升级,性能提升30%; 支持EasyDL专业版模型
2019.0 9.23	0.4.3	增加海思NNIE加速芯片支持
2019.0 8.30	0.4.2	ARM引擎升级;支持分类高性能与高精度模型
2019.0 7.25	0.4.1	引擎升级,性能提升
2019.0 7.25	0.4.0	支持Xeye, 细节完善
2019.0 6.11	0.3.3	paddle引擎升级;性能提升
2019.0 5.16	0.3.2	新增NVIDIA GPU支持;新增armv7l支持
2019.0 4.25	0.3.1	优化硬件支持
2019.0 3.29	0.3.0	ARM64 支持;效果提升
2019.0 2.20	0.2.1	paddle引擎支持;效果提升
2018.1 1.30	0.1.0	第一版!

2022-5-18:【接口变更】 PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE含义变更。 变更前:预测输入图片数不大于该值均可。 变更后:预测输入图片数需等于该值。SDK内部对该接口变更做了兼容处理,在输入图片数小于该值时依然可正常运行,但预测性能会和等于该值时一致。推荐根据实际输入图片数量需求修改该值,尽可能保持最小。

2020-12-18: 【接口升级】 参数配置接口从1.0.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。 【关于SDK包与RES模型文件夹配套使用的说明】 我们强烈建议用户使用部署tar包中配套的SDK和RES。 更新模型时,如果SDK版本号有更新,请务必同时更新SDK,旧版本的SDK可能无法 正确适配新发布出来部署包中的RES模型。

快速开始

SDK在以下环境中测试通过

- x86_64, Ubuntu 16.04, gcc 5.4
- x86_64, Ubuntu 18.04, gcc 7.4
- Tesla P4, Ubuntu 16.04, cuda 9.0, cudnn 7.5
- x86_64, Ubuntu 16.04, gcc 5.4, XTCL r1.0
- aarch64, Kylin V10, gcc 7.3
- loongarch64, Kylin V10, gcc 8.3
- Bitmain SC5+ BM1684, Ubuntu 18.04, gcc 5.4
- x86_64 MLU270 , Ubuntu 18.04, gcc 7.5
- phytium MLU270 , Kylin V10 , gcc 7.3.0
- phytium XPU , Kylin V10 , gcc 7.3.0
- hygon DCU, CentOS 7.8 gcc 7.3.0

- XPU K200, x86_64, Ubuntu 18.04
- XPU K200 aarch64, Ubuntu 18.04
- XPU R200, x86_64, Ubuntu 18.04
- XPU R200 aarch64, Ubuntu 18.04
- MLU370, x86_64, Centos7.6.1810

依赖包括

- cmake 3+
- gcc 5.4 (需包含 GLIBCXX_3.4.22), gcc / glibc版本请以实际SDK ldd的结果为准
- opencv3.4.11 (可选)
- cuda && cudnn (使用NVIDIA-GPU时必须,SDK内提供多个Cuda版本推理套件,根据需要安装依赖的Cuda和Cudnn版本)
- XTCL 1.0.0.187 (使用昆仑服务器时必须)
- Rocm4.3, Miopen 2.14(使用海光DCU服务器时必须)

1. 安装依赖

以下步骤均可选,请开发者根据实际运行环境选择安装。

(可选) 安装cuda&cudnn

在NVIDIA GPU上运行必须(包括GPU基础版,GPU加速版)

对于GPU基础版,若开发者需求不同的依赖版本,请在PaddlePaddle官网下载对应版本的libpaddle_fluid.so或参考其文档进行编译,覆盖lib文件 夹下的相关库文件。

(可选) 安装TensorRT

在NVIDIA GPU上运行GPU加速版必须

下载包中提供了对应 cuda9.0、cuda10.0、cuda10.2、cuda11.0+四个版本的 SDK, cuda9.0 和 cuda10.0 的 SDK 默认依赖的 TensorRT 版本为 TensorRT7.0.0.11, cuda10.2 及以上的 SDK 默认依赖的 TensorRT 版本为 TensorRT8.4,请在这里下载对应 cuda 版本的 TensorRT,并把其中 的lib文件拷贝到系统lib目录,或其他目录并设置环境变量。

(可选) 安装XTCL 使用昆仑服务器及对应SDK时必须请安装与1.0.0.187版本兼容的XTCL。必要时,请将运行库路径添加到环境变量。

(可选) 安装Rocm、Miopen

使用海光DCU服务器对应SDK时必须

海光DCU SDK依赖Rocm 4.3和Miopen 2.14版本,推荐使用easyedge镜像

(registry.baidubce.com/easyedge/hygon_dcu_infer:1.0.2.rocm4.3), SDK镜像内运行,镜像拉取方式(wget https://aipe-easyedge-

public.bj.bcebos.com/dcu_docker_images/hygon_dcu_rocm4.3.tar.gz && docker load -i hygon_dcu_rocm4.3.tar.gz),关于海光DCU使用更多细节 可参考paddle文档

2. 使用序列号激活 请在官网获取序列号

<mark>共興其服务状態</mark> 至冬间或最低感,所加低完成的模型即是在走地。直线调用模型,可以选择将模型即著在本地的服务器、小型设备、软硬一体方案专项适配硬件上, 进过AP、SOK进一步集成,灵活道应不同业务场景。 <u>来专家服务</u> 拉斯伯						
服务器 通用小型设备 专项适配硬件						
SDK API						
此处发布、下载的SDK为未授权SDK,需要前往控制	台获取序列号激活后才能正式使用	。SDK内附有对应版本的Demo及开发文档,开	发者可参考源代码完成开发。			
模型名称	发布版本 🗘	应用平台	模型加速	发布状态	发布时间 🗘	
		通用X86 CPU-Linux	基础版	• 已发布	2021-08-19 20:24	下载SDK
			精度无损压缩加速	 已发布 	2021-08-19 20:24	下载加速版SDK
and the first state	134318-V1	2424-001 Linux	基础版	•已发布	2021-08-19 20:35	下截SDK
am_mean	宣看性能报告	Selero on o - China	精度无损压缩加速	 已发布 	2021-08-19 20:34	下载加速版SDK

基础版

•已发布

2021-08-19 18:17

下载SDK

serial_num(或者serial_key)处填入序列号可自动完成联网激活(请确保硬件首次激活时能够连接公网,如果确实不具备联网条件,需要使用纯离 线模式激活,请下载使用百度智能边缘控制台纳管SDK)

SDK内提供的一些二进制文件,填入序列号运行可自动完成激活,以下二进制具体使用说明参考下一小节 ./edgekit_serving --cfg=./edgekit_serving.yml ./easyedge_image_inference {model_dir} {image_name or image_directory} [serial_num] ./easyedge_serving [res_dir] {serial_key} {host} {port}

如果是基于源码集成,设置序列号方法如下

global_controller()->set_licence_key("")

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的变量或者源码设置)实例数鉴权环境变量设置方法

export EDGE_CONTROLLER_KEY_AUTH_MODE=2 export EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=30

实例数鉴权源码设置方法

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2) global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 300)

3. 基于预编译二进制测试图片推理和http服务测试图片推理模型资源文件默认已经打包在开发者下载的SDK包中。

请先将tar包整体拷贝到具体运行的设备中,再解压缩编译; 在Intel CPU上运行CPU加速版,如果thirdparty里包含openvino文件夹的,必须在 编译或运行demo程序前执行以下命令: source \${cpp_kit位置路径}/thirdparty/openvino/bin/setupvars.sh 或者执行 source \${cpp_kit位置路 径}/thirdparty/openvino/setupvars.sh(openvino-2022.1+) 如果SDK内不包含setupvars.sh脚本,请忽略该提示

运行预编译图片推理二进制,依次填入模型文件路径(RES文件夹路径)、推理图片、序列号(序列号尽首次激活需要使用,激活后可不用填序列号 也能运行二进制)

./easyedge_image_inference {model_dir} {image_name or image_directory} {serial_num} LD_LIBRARY_PATH=../lib ./easyedge_image_inference ../../../RES /xxx/cat.jpeg "1111-1111-1111"

demo运行效果:



图片加载失败

> ./easyedge_image_inference ../../../RES 2.jpeg

2019-02-13 16:46:12,659 INFO [EasyEdge] [easyedge.cpp:34] 140606189016192 Baidu EasyEdge Linux Development Kit 0.2.1(20190213)

2019-02-13 16:46:14,083 INFO [EasyEdge] [paddlev2_edge_predictor.cpp:60] 140606189016192 Allocate graph success. 2019-02-13 16:46:14,326 DEBUG [EasyEdge] [paddlev2_edge_predictor.cpp:143] 140606189016192 Inference costs 168 ms 1, 1:txt_frame, p:0.994905 loc: 0.168161, 0.153654, 0.920856, 0.779621 Done

启动http服务 bin目录下提供编译好的启动http服务二进制文件,可直接运行

推荐使用 edgekit_serving 启动模型服务 LD_LIBRARY_PATH=../lib ./edgekit_serving --cfg=./edgekit_serving.yml

也可以使用 easyedge_serving 启动模型服务

./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401}

LD_LIBRARY_PATH=../lib ./easyedge_serving ../../../RES "1111-1111-1111-1111" 0.0.0.0 24401

HTTP(or Webservice) is now serving at 0.0.0.24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试,网页右侧会展示模型推理结果



对于目标追踪的模型,请选择一段视频,并耐心等待结果



```
同时,可以调用HTTP接口来访问服务。
```

请求http服务 以图像预测场景为例(非语义分割模型场景,语义分割请求方式参考后面小节详细文档),提供一张图片,请求模型服务的示例参考 如下demo

python示例代码如下

```
import requests
with open('./1.jpg', 'rb') as f:
  img = f.read()
**params 为GET参数 data 为POST Body**
result = requests.post('http://127.0.0.1:24401/', params= \{'threshold': 0.1\},
                                data=img).json()
```

C# 使用示例代码如下

FileStream fs = new FileStream("./img.jpg", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close(); WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream());

StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close(); response.Close();

C++ 使用示例代码如下,需要安装curl

```
**include <sys/stat.h>**
**include <curl/curl.h>**
**include <iostream>**
**include <string>**
**define S_ISREG(m) (((m) & 0170000) == (0100000))**
**define S_ISDIR(m) (((m) & 0170000) == (0040000))**
size_t write_callback(void *ptr, size_t size, size_t num, void *data) {
  std::string *str = dynamic_cast<std::string *>((std::string *)data);
  str->append((char *)ptr, size*num);
  return size*num:
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./img.jpg";
  FILE *fp = NULL;
  std::string response;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\"\n", post_data_filename);
     return -1;
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
     curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
    curl_easy_setopt(curl, CURLOPT_POST, 1L);
    curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
    curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
 curl_easy_setopt(curl, CURLOPT_WRITEFUNCTION, write_callback);
    curl_easy_setopt(curl, CURLOPT_WRITEDATA, &response);
     res = curl_easy_perform(curl);
     if (res != CURLE_OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     std::cout << response << std::endl; // response即为返回的json数据
     curl_easy_cleanup(curl);
  curl_global_cleanup();
  fclose(fp);
  return 0:
```

关于http接口的详细介绍参考下面集成文档http服务章节的相关内容

使用说明

使用该方式,将运行库嵌入到开发者的程序当中。 **编译demo项目** SDK src目录下有完整的demo工程,用户可参考该工程的代码实现方式将SDK 集成到自己的项目中,demo工程可直接编译运行:

```
cd src
mkdir build && cd build
cmake .. && make
./easyedge_image_inference {模型RES文件夹} {测试图片路径}
**如果是NNIE引擎,使用sudo运行**
sudo ./easyedge_image_inference {模型RES文件夹} {测试图片路径}
```

(可选) SDK包内一般自带opencv库,可忽略该步骤。如果希望SDK自动编译安装所需要的OpenCV库,修改cmake的 optionEDGE_BUILD_OPENCV为ON即可。 SDK会自动从网络下载opencv源码,并编译需要的module、链接。注意,此功能必须需联网。 cmake -DEDGE_BUILD_OPENCV=ON .. && make -j16

若需自定义library search path或者gcc路径,修改CMakeList.txt即可。

使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

// step 1: 配置模型资源目录 EdgePredictorConfig config; config.model_dir = {模型文件目录};

// step 2: 创建并初始化Predictor;在这里选择合适的引擎 auto predictor = global_controller()->CreateEdgePredictor(config);

// step 3-1: 预测图像 auto img = cv::imread({图片路径}); std::vector<EdgeResultData> results; predictor->infer(img, results);

// step 3-2: 预测视频 std::vector<EdgeResultData> results; FrameTensor frame_tensor; VideoConfig video_config; video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h video_config.source_value = video_src; /* ... more video_configs, 根据需要配置video_config的各选项 */ auto video_decoding = CreateVideoDecoding(video_config); while (video_decoding->next(frame_tensor) == EDGE_OK) { results.clear(); if (frame_tensor.is_needed) { predictor->infer(frame_tensor.frame, results); render(frame_tensor.frame, results, predictor->model_info().kind); //video_decoding->display(frame_tensor); // 显示当前frame , 需在video_config中开启配置 //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置

```
输入图片不限制大小
```

SDK参数配置 SDK的参数通过EdgePredictorConfig::set_config和global_controller()->set_config配置。set_config的所有key 在easyedge_xxxx_config.h中。其中

- PREDICTOR前缀的key是不同模型相关的配置,通过EdgePredictorConfig::set_config设置
- CONTROLLER前缀的key是整个SDK的全局配置,通过global_controller()->set_config设置

```
以序列号为例,KEY的说明如下:
```

```
/**
* @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号
* 值类型:string
* 默认值:空
*/
static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";
```

EdgePredictorConfig config; config.model_dir = ...; config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, "1DB7-1111-1111-D27D");

具体支持的运行参数配置列表可以参考开发工具包中的头文件的详细说明。

相关配置均可以通过环境变量的方法来设置,对应的key名称加上前缀EDGE_即为环境变量的key。如序列号配置的环境变量key为EDGE_PREDICTOR_KEY_SERIAL_NUM,如指定CPU线程数的环境变量key为EDGE_PREDICTOR_KEY_CPU_THREADS_NUM。 注意:通过代码设置的配置会覆盖通过环境变量设置的值。

初始化

● 接口

auto predictor = global_controller()->CreateEdgePredictor(config); predictor->init();

若返回非0,请查看输出日志排查错误原因。

预测图像

• 接口

/** * @brief * 通用接口 * @param image: must be BGR , HWC format (opencv default) * @param result * @return */ virtual int infer(cv::Mat& image, std::vector<EdgeResultData>& result) = 0; /** * @brief * 批量图片推理接口 * @param image: must be BGR , HWC format (opencv default) * @param result * @return */ virtual int infer(std::vector<cv::Mat>& image, std::vector<std::vector<EdgeResultData>>& result) = 0;

图片的格式务必为opencv默认的BGR, HWC格式。

• 返回格式

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index: // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测、图像分割时才有意义
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割的模型,该字段才有意义
    // 请注意: 图像分割时,以下两个字段会比较大,使用完成之后请及时释放EdgeResultData
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding,游程编码的mask
    // 目标追踪模型,该字段才有意义
    int trackid; // 轨迹id
    int frame; // 处于视频中的第几帧
```

```
};
```

关于矩形坐标

x1 * 图片宽度 = 检测框的左上角的横坐标

EdgeTrackStat track_stat; // 跟踪状态

- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2*图片宽度 = 检测框的右下角的横坐标
- y2 * 图片高度 = 检测框的右下角的纵坐标

关于图像分割mask

cv::Mat mask为图像掩码的二维数组 { {0,0,0,0,0,0,0,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,0,0,0,0,0,0,0}, {1,1,1,1,0,0,0,0}, {1,1,1,1,0,0,0}, {1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,0,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,0,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,0,0,0},\\{1,1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},\\{1,1,1,1,0,0,0},{1,1,1,1,0,0,0},{1,1,1,1,0,0,0},{1,1,1,1,0,0,0},{1,1,1,1,0,0,0},{1,1,1,1,0,0},{1,1,1,1,0,0},{1,1,1,0,

关于图像分割mask_rle

该字段返回了mask的游程编码,解析方式可参考 http demo

以上字段可以参考demo文件中使用opencv绘制的逻辑进行解析

预测视频

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

接口

classVideoDecoding :

/**

- * @brief 获取输入源的下一帧
- * @param frame_tensor
- * @return

*/

virtual int next(FrameTensor &frame_tensor) = 0;

/**

- * @brief 显示当前frame_tensor中的视频帧
- * @param frame_tensor
- * @return
- */

virtual int display(const FrameTensor &frame_tensor) = 0;

/**

* @brief 将当前frame_tensor中的视频帧写为本地视频文件

- * @param frame_tensor
- * @return

virtual int save(FrameTensor &frame_tensor) = 0;

/**

*/

```
/**
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
* @brief 获取视频的width属性
* @return
*/
virtual int get_width() = 0;
/**
* @brief 获取视频的height属性
```

```
* @return
*/
```

```
virtual int get_height() = 0;
```

```
struct VideoConfig
```

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type; // 输入源类型
  std::string source_value;   // 输入源地址,如视频文件路径、摄像头index、网络流地址
 int skip_frames{0};
                      // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
  int retrieve_all{false};
                     // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
                      // 在采取抽帧之前设置视频的fps
  int input_fps{0};
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
 bool enable_display{false};
  std::string window_name{"EasyEdge"};
  bool display_all{false};
                       // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                       // 是否存储所有frame, 若为false, 仅存储根据skip_frames抽取的frame
  std::map::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被

Baidu 百度智能云文档

抽取返回,以作为显示或存储用。 input_fps:用于抽帧前设置fps。 resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。 conf:高级选项。部分配置会通过该map来设置。

注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

设置序列号

请在网页控制台中申请序列号,并在init初始化前设置。 LinuxSDK 首次使用需联网授权。

EdgePredictorConfig config; config.set_config(easyedge::params::PREDICTOR_KEY_SERIAL_NUM, "this-is-serial-num");

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

```
EdgeLogConfig log_config;
log_config.enable_debug = true;
global_controller()->set_log_config(log_config);
```

http服务

1. 开启http服务 http服务的启动可以参考demo_serving.cpp文件。

/**

- * @brief 开启一个简单的demo http服务。
- * 该方法会block直到收到sigint/sigterm。
- * http服务里,图片的解码运行在cpu之上,可能会降低推理速度。
- * @tparam ConfigT
- * @param config
- * @param host
- * @param port
- * @param service_id service_id user parameter, uri '/get/service_id' will respond this value with 'text/plain'
- * @param instance_num 实例数量,根据内存/显存/时延要求调整
- * @return

```
int port,
const std::string &service_id,
int instance_num = 1);
```

2. http接口详细说明

开发者可以打开浏览器,http://{设备ip}:24401,选择图片或视频来进行测试。

http 请求方式一:无额外编码 URL中的get参数:

参数	说明		默认值
threshold	阈值过滤,	0~1	如不提供,则会使用模型的推荐阈值

Python请求示例 (图片测试,针对图像分类、物体检测、实例分割等模型)

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
        'http://127.0.0.1:24401/',
        params={'threshold': 0.1},
        data=img).json()
```

Python请求示例 (图片测试, 仅针对语义分割模型, 同其他CV模型不同, 语义分割模型输出为灰度图)

```
import requests
with open('./1.jpg', 'rb') as f:
    img_data = f.read()
    res = requests.post('http://127.0.0.1:24401/',
        data=img_data)
with open("gray_result.png", "wb") as fb:
    fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
```

Python请求示例 (视频测试, 注意:区别于图片预测,需指定Content-Type;否则会调用图片推理接口)

```
import requests
with open('./1.mp4', 'rb') as f:
video_data = f.read()
result = requests.post(
    'http://127.0.0.1:24401/',
    params={'threshold': 0.1},
    headers={'Content-Type': 'video'},
    data=video_data).json()
```

http 请求方法二:json格式,图片传base64格式字符串 HTTP方法:POST Header如下:

参数	值
Content-Type	application/json

Body请求填写:

• 图像分类网络: body 中请求示例

```
{
"image": "<base64数据>",
"top_num": 5
}
```

body中参数详情

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
top_nu m	否	number	-	返回分类数量,不填该参数,则默认返回全部分类结果

• 物体检测和实例分割网络: Body请求示例:

```
{
"image": "<base64数据>",
"threshold": 0.3
```

body中参数详情:

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
thresho Id	否	number	-	默认为推荐阈值,也可自行根据需要进行设置

• 语义分割网络: Body请求示例:

```
{
"image": "<base64数据>"
}
```

body中参数详情(语义分割由于模型特殊性,不支持设置threshold值,设置了也没有意义):

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部

Python请求示例(非语义分割模型参考如下代码)

```
import base64
import requests

def main():
    with open("1.jpg 【图片路径】", 'rb') as f:
        result = requests.post("http://{服务ip地址}:24401/", json={
            "image": base64.b64encode(f.read()).decode("utf8")
        })
        # print(result.request.body)
        # print(result.request.headers)
        print(result.content)

if __name__ == '__main__':
        main()
```

Python 请求示例 (针对语义分割模型,同其他CV模型不同,语义分割模型输出为灰度图)

```
import base64
import requests
def main():
    with open("1.jpg 【图片路径】", 'rb') as f:
        res = requests.post("http://{服务ip地址}:24401/", json=["image": base64.b64encode(f.read()).decode("utf8")})
        with open("gray_result.png", "wb") as fb:
        fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
if __name__ == '__main__':
        main()
```

http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考预测图像-返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

返回示例

```
{
    "cost_ms": 52,
    "error_code": 0,
    "results": [
        {
            "confidence": 0.94482421875,
            "index": 1,
            "label": "IronMan",
            "x1": 0.059185408055782318,
            "x2": 0.18795496225357056,
            "y1": 0.14762254059314728,
            "y2": 0.52510076761245728,
            "mask": "...", // 图像分割模型字段
            "trackld": 0, // 目标追踪模型字段
        },
        ]
}
```

其他配置

1. 日志名称、HTTP 网页标题设置

通过global_controller的set_config方法设置:

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_LOG_BRAND, "MY_BRAND");

效果如下:



图片加载失败

2. CPU线程数设置

CPU线程数可通过 EdgePredictorConfig::set_config 配置

```
EdgePredictorConfig config;
config.set_config(easyedge::params::PREDICTOR_KEY_CPU_THREADS_NUM, 4);
```

3. 批量预测设置

```
int batch_size = 2; // 使用前修改batch_size再编译、执行
while (get_next_batch(imgs, img_files, batch_size, start_index)) {
    ...
}
```

GPU 加速版 预测接口 GPU 加速版 SDK 除了支持上面介绍的通用接口外,还支持图片的批量预测,预测接口如下:

```
/**
 * @brief
* GPU加速版批量图片推理接口
* @param image: must be BGR , HWC format (opency default)
* @param result
* @return
*/
virtual int infer(
     std::vector<cv::Mat>& image
     std::vector<std::vector<EdgeResultData>>& result
) = 0:
/**
* @brief
* GPU加速版批量图片推理接口,带阈值
* @related infer(cv::Mat & image, EdgeColorFormat origin_color_format, std::vector<EdgeResultData> & result, float threshold)
*/
virtual int infer(
       std::vector<cv::Mat> &images,
       EdgeColorFormat origin_color_format,
       std::vector<std::vector<EdgeResultData>> &results
       float threshold
) = 0;
```

批量图片的预测接口的使用要求在调用 init 接口的时候设置一个有效的 PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE,其含义见下方参数配置接口的介绍。

运行参数选项 在上面的内容中我们介绍了如何使用EdgePredictorConfig进行运行参数的配置。针对GPU加速版开发工具包,目前EdgePredictorConfig的运行参数所支持的Key包括如下项:

```
/**
* @brief 当有同类型的多个设备的时候,使用哪一个设备,如:
* GPU: 使用哪张GPU卡
* EdgeBoard(VMX), Movidius NCS:使用哪一张加速卡
* 值类型: int
* 默认值:0
*/
static constexpr auto PREDICTOR_KEY_DEVICE_ID = "PREDICTOR_KEY_DEVICE_ID";
/**
* @brief 生成最大 batch_size 为 max_batch_size 的优化模型,单次预测图片数量可以小于或等于此值(推荐等于此值,见release notes)
* 值类型: int
* 默认值:4
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE = "PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE";
/**
* @brief 设置device对应的GPU卡可以支持的最大并发量
* 实际预测的时候对应GPU卡的最大并发量不超过这里设置的范围,否则预测请求会排队等待预测执行
* 值类型: int
* 默认值:1
*/
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY = "PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY";
/**
* @brief 是否开启fp16模式预测,开启后预测速度会更快,但精度会略有降低。并且需要硬件支持fp16
* 值类型: bool
* 默认值: false
*/
static constexpr auto PREDICTOR_KEY_GTURBO_FP16 = "PREDICTOR_KEY_GTURBO_FP16";
/**
* @brief 模型编译等级
*1:如果当前max_batch_size与历史编译产出的max_batch_size不相等时,则重新编译模型(推荐)
* 2:无论历史编译产出的max_batch_size为多少,均根据当前max_batch_size重新编译模型
* 值类型: int
* 默认值:1
*/
```

static constexpr auto PREDICTOR_KEY_GTURBO_COMPILE_LEVEL = "PREDICTOR_KEY_GTURBO_COMPILE_LEVEL"; /** * @brief GPU工作空间大小设置 * workspace_size = workspace_prefix * (1 << workspace_offset) * workspace_offset: 10 = KB, 20 = MB, 30 = GB * 值类型: int *默认值:WORKSPACE_PREFIX: 100, WORKSPACE_OFFSET: 20,即100MB static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX = "PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX"; static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET = "PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET"; /** * @brief 需要使用的dla core * 值类型: int * 默认值:-1(不使用) static constexpr auto PREDICTOR_KEY_GTURBO_DLA_CORE = "PREDICTOR_KEY_GTURBO_DLA_CORE"; /** * @brief 自定义缓存文件存储路径 * 值类型: string * 默认值: ~/.baidu/easyedge/mcache/{model_id * 1000000 + release_id} static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_DIR = "PREDICTOR_KEY_GTURBO_CACHE_DIR"; /** * @brief 自定义缓存文件命名,默认即可 * 值类型: string *默认值:根据配置自动生成 static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_NAME = "PREDICTOR_KEY_GTURBO_CACHE_NAME"; /** * @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号 * 值类型: string *默认值:空 static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";

PREDICTOR_KEY_GTURBO_CACHE_NAME:首次加载模型会先对模型进行编译优化,通过此值可以设置优化后的产出文件名,这在多进程加载同一个模型的时候是有用的。

PREDICTOR_KEY_GTURBO_CACHE_DIR:首次加载模型经过编译优化后,产出的优化文件会存储在这个位置,可以按需修改。

PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX、PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET:设置运行时可以被用来使用的最大临时显存。

PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE:此值用来控制批量图片预测可以支持的最大图片数,实际预测的时候单次预测图片数需等于此值。

PREDICTOR_KEY_DEVICE_ID:设置需要使用的 GPU 卡号。

PREDICTOR_KEY_GTURBO_COMPILE_LEVEL: 模型编译等级。通常模型的编译会比较慢,但编译产出是可以复用的。可以在第一次加载模型的时候设置合理的 max_batch_size 并在之后加载模型的时候直接使用历史编译产出。是否使用历史编译产出可以通过此值 compile_level 来控制。当此值为 0 时,表示忽略当前设置的 max_batch_size 而仅使用历史产出(无历史产出时则编译模型);当此值为 1 时,会比较历史产出和当前设置的 max_batch_size 是否相等,如不等,则重新编译;当此值为 2 时,无论如何都会重新编译模型。

PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY:通过此值设置单张 GPU 卡上可以支持的最大 infer 并发量,其上限取决于硬件限制。init 接口 会根据此值预分配 GPU 资源,建议结合实际使用控制此值,使用多少则设置多少。注意:此值的增加会降低单次 infer 的速度,建议优先考虑 batch inference 和 multi predictor。

PREDICTOR_KEY_GTURBO_FP16:默认是 fp32 模式,置 true 可以开启 fp16 模式预测,预测速度会有所提升,但精度也会略微下降,权衡使用。注意:不是所有模型都支持 fp16 模式。目前已知不支持fp16的模型包括:图像分类高精度模型。

多线程预测 GPU 加速版 SDK 的多线程分为单卡多线程和多卡多线程两种。 单卡多线程:创建一个 predictor,并通过 PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY 控制单卡所支持的最大并发量,只需要 init 一次,多线程调用 infer 接口。 多卡多线程:多卡的 支持是通过创建多个 predictor,每个 predictor 对应一张 GPU 卡,predictor 的创建和 init 的调用放在主线程,通过多线程的方式调用 infer 接口。

已知问题 1. 多线程时图片按线程分配不均 或 不同batch size的图片交叉调用infer接口时,部分结果错误 A: EasyDL图像分类高精度模型在有些显卡上可能存在此问题,可以考虑填充假图片数据到图片比较少的线程或batch以使得infer间的图片绝对平均。

2. 显存持续增长或遇到 terminate called after throwing an instance of 'std::runtime_error' what(): Failed to create object A: 部分显卡存在此问题, 如果遇到此问题, 请确认没有频繁调用 init 接口, 通常调用 infer 接口即可满足需求。

3. 开启 fp16 后,预测结果错误 A:不是所有模型都支持 fp16 模式。目前已知的不支持fp16的模型包括:图像分类高精度模型。目前不支持的 将会在后面的版本陆续支持。

昆仑服务器 昆仑服务器SDK支持将EasyDL的模型部署到昆仑服务器上。SDK提供的接口风格一致,简单易用,轻松实现快速部署。Demo的测试 可参考上文中的测试Demo部分。

参数配置接口 在上面的内容中我们介绍了如何使用EdgePredictorConfig进行运行参数的配置。针对昆仑服务器开发工具包,目前EdgePredictorConfig的运行参数所支持的Key包括如下项:

```
/**
* @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号
* 值类型: string
*默认值:空
*/
static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";
/**
* @brief 当有同类型的多个设备的时候,使用哪一个设备,如:
* 使用哪张加速卡
* 值类型: int
*默认值:0
*/
static constexpr auto PREDICTOR_KEY_DEVICE_ID = "PREDICTOR_KEY_DEVICE_ID";
/**
* @brief 设置需要同时预测的图片数量
* 值类型: int
*默认值:1
*/
static constexpr auto PREDICTOR_KEY_KUNLUN_BATCH_SIZE = "PREDICTOR_KEY_KUNLUN_BATCH_SIZE";
```

PREDICTOR_KEY_DEVICE_ID:设置需要使用的加速卡的卡号。

PREDICTOR_KEY_KUNLUN_BATCH_SIZE:设置单次预测可以支持的图片数量。

使用方法:

```
int batch_size = 1;
config.set_config(easyedge::params::PREDICTOR_KEY_KUNLUN_BATCH_SIZE, batch_size);
```

模型调优通过设置如下环境变量,可以在初始化阶段对模型调优,从而让预测的速度更快。

export XPU_CONV_AUTOTUNE=5

FAQ

1. 如何处理一些 undefined reference?

如: undefined reference to `curl_easy_setopt@CURL_OPENSSL_3'

方案1:通过安装libcurl3 libcurl-openssl1.0-dev来解决。 方案2:如果开发者想不想使用低版本的openssl (如Ubuntu 18.04),可以link静态 库easyedge_static.a,自己指定需要的Library的版本: find_package(CURL REQUIRED)

target_link_libraries(easyedge_demo \${OpenCV_LIBS} easyedge_static pthread \${CURL_LIBRARIES} verify_static \$(其他需要的库))

其中,其他需要的库视具体sdk中包含的库而定。

2. EasyDL SDK与云服务效果不一致,如何处理?

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

3. NVIDIA GPU预测时,报错显存不足如以下错误字样:

paddle.fluid.core.EnforceNotMet: Enforce failed. Expected allocating <= available, but received allocating:20998686233 > available:19587333888. Insufficient GPU memory to allocation. at [/paddle/paddle/fluid/platform/gpu_info.cc:170]

请根据显存大小和模型配置。调整合适的初始 fraction_of_gpu_memory。参数的含义参考这里。

4. 如何将我的模型运行为一个http服务?目前cpp sdk暂未集成http运行方式; 0.4.7版本之后,可以通过start_http_server方法开启http服务。

5. 运行NNIE引擎报permission denied 日志显示:

open sys: Permission denied open err : Permission denied open err : Permission denied

请使用sudo在root下运行。

6. 运行SDK报错 Authorization failed

情况一:日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888" ./easyedge_demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除~/.baidu/easyedge 目录,再重新激活。

7. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

8. 运行二进制时,提示 libverify.so cannot open shared object file

可能cmake没有正确设置rpath,可以设置LD_LIBRARY_PATH为sdk的lib文件夹后,再运行:

LD_LIBRARY_PATH=\$LD_LIBRARY_PATH:./lib ./easyedge_demo

9. 运行二进制时提示 libopencv_videoio.so.4.5: cannot open shared object file: No such file or directory 同上面8的问题类似,没有正确设置动态 库的查找路径,可通过设置LD_LIBRARY_PATH为sdk的thirdparty/opencv/lib文件夹解决 export LD_LIBRARY_PATH=\${LD_LIBRARY_PATH}:../thirdparty/opencv/lib (tips: 上面冒号后面接的thirdparty/opencv/lib路径以实际项目中路径为准,比如也可能是../../thirdparty/opencv/lib)

10. 编译时报错:file format not recognized 可能是因为在复制SDK时文件信息丢失。请将整个压缩包复制到目标设备中,再解压缩、编译

11. 进行视频解码时,报错符号未找到、格式不支持、解析出的图片为空、无法设置抽帧 请确保安装OpenCV时,添加了-DWITH_FFMPEG=ON选项(或者GStream选项),并且检查OpenCV的安装日志中,关于 Video I/O 段落的说明是否为YES。

- -- Video I/O: -- DC1394: YES (ver 2.2.4)
- -- FFMPEG: YES
- -- avcodec: YES (ver 56.60.100)
- -- avformat: YES (ver 56.40.101)
- -- avutil: YES (ver 54.31.100)
- -- swscale: YES (ver 3.1.101) -- avresample: NO
- -- avresample: NO -- libv4l/libv4l2: NO
- -- v4l/v4l2: linux/videodev2.h
- 如果为NO,请搜索相关解决方案,一般为依赖没有安装,以apt为例:

apt-get install yasm libjpeg-dev libjasper-dev libavcodec-dev libavformat-dev libswscale-dev libdc1394-22-dev libgstreamer0.10-dev libgstreamer0.10-dev libgstreamer-plugins-base0.10-dev libv4l-dev python-numpy libtbb-dev libqt4-dev libgtk2.0-dev libfaac-dev libmp3lame-dev libopencore-amrnb-dev libopencore-amrnb-dev libtheora-dev libtheora-dev libtvorbis-dev libxvidcore-dev x264 v4l-utils ffmpeg

12. GPU加速版运行有损压缩加速的模型,运算精度较标准模型偏低首先请保证数据集不会太小。其次可以通过将模型目录RES中的calibrationtable移除,并通过将PREDICTOR_KEY_GTURBO_FP16设置为true,使用FP16的运算精度重新评估模型效果。若依然不理想,可将calibrationtable移除并将PREDICTOR_KEY_GTURBO_FP16设置为false,从而使用更高精度的FP32的运算精度。

の Linux集成文档-Python

简介

本文档介绍 EasyDL 的 Linux Python SDK 的使用方法,适用于 EasyDL 和 BML。

EasyDL 通用版:

- 网络类型支持:图像分类,物体检测,图像分割,声音分类,表格预测
- 硬件支持:
 - Linux x86_64 CPU (基础版,加速版)
 - Linux x86_64 Nvidia GPU (基础版,加速版)
- 语言支持: Python 3.5, 3.6, 3.7, 3.8, 3.9

BML:

- 网络类型支持:图像分类,物体检测,声音分类,表格预测
- 硬件支持:
 - Linux x86_64 CPU (基础版)
 - Linux x86_64 Nvidia GPU (基础版)
- 语言支持:Python 3.5, 3.6, 3.7, 3.8, 3.9

Release Notes

Baidu 百度智能云文档

时间	版本	说明
2023-03-16	1.3.7	迭代升级,新增支持文本类模型; 新增GPU 多卡多进程推理demo
2022.10.27	1.3.5	新增华为Atlas300、飞腾Atlas300 Python SDK,支持图像分类、物体检测、人脸检测、实例分割
2022.09.15	1.3.3	EasyDL CPU普通版新增支持表格预测
2022.05.27	1.3.1	CPU、GPU普通版新增支持BML Cloud小目标检测模型
2021.12.22	1.2.7	声音分类模型升级
2021.10.20	1.2.6	CPU基础版、CPU加速版、GPU基础版推理引擎优化升级
2021.08.19	1.2.5	CPU基础版、CPU无损加速版、GPU基础版新增支持EasyDL小目标检测
2021.06.29	1.2.4	CPU、GPU新增EasyDL目标跟踪支持;新增http server服务启动demo
2021.03.09	1.2.2	EasyDL CPU加速版新增支持分类、高性能检测和均衡检测的量化压缩模型
2021.01.27	1.2.1	EasyDL经典版分类高性能模型升级;支持更多模型
2020.12.18	1.2.0	推理引擎升级;接口升级;性能优化
2020.09.17	1.1.19	支持更多模型
2020.08.11	1.1.18	性能优化
2020.06.23	1.1.17	支持更多EasyDL专业版模型
2020.04.16	1.1.15	技术优化;升级 OpenVINO 版本
2020.03.12	1.1.14	新增声音识别python sdk
2020.02.12	1.1.13	新增口罩模型支持
2020.01.16	1.1.12	预测函数默认使用推荐阈值
2019.12.26	1.1.11	EasyDL 专业版支持 SDK 加速版
2019.12.04	1.1.10	支持图像分割
2019.10.21	1.1.9	支持 EasyDL 专业版
2019.08.29	1.1.8	CPU 加速版支持
2019.07.19	1.1.7	提供模型更新工具
2019.05.16	1.1.3	NVIDIA GPU 支持
2019.03.15	1.1.0	架构与功能完善
2019.02.28	1.0.6	引擎功能完善
2019.02.13	1.0.5	paddlepaddle 支持
2018.11.30	1.0.0	第一版!

2020-12-18: 【接口升级】 序列号的配置接口从1.2.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。 请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

快速开始

1. 安装依赖

- 根据引擎的不同,SDK 依赖了不同的底层引擎。根据所需自行安装。
- 使用声音分类SDK需要安装额外依赖

* pip 安装 resampy pydub six librosa 音频默认格式支持wav文件预测,如果需要预测mp3等其他音频格式的数据需要系统额外安装ffmpeg (windows系统的ffmpeg已基在sdk中无需额外安装,linux系统需要手动安装)

• 使用表格预测SDK需要安装额外依赖

pip安装 brotlipy==0.7.0 certifi==2020.6.20 joblib==1.0.1 kaggle==1.5.12 Pillow py4j pycosat python-dateutil python-slugify ruamel_yaml textunidecode threadpoolctl flask pandas==1.0.5 scikit-learn==0.23.2 lightgbm==2.2.3 catboost==0.24.1 xgboost==1.2.0 numpy==1.19.5 scipy==1.5.2

psutil==5.7.2 pypmml==0.9.7 torch==1.8.0 jieba==0.42.1 pyod==0.8.5 pyarrow==6.0.0 scikit-optimize==0.9.0 pyspark==3.3.0 另外ml算法安装(目前只支持python3.7)

pip install BaiduAl_TabularInfer-0.0.0-cp37-cp37m-linux_x86_64.whl

安装 paddlepaddle

• 使用x86_64 CPU 基础版预测时必须安装(目标跟踪、表格预测除外):

python -m pip install paddlepaddle==2.2.2 -i https://mirror.baidu.com/pypi/simple

```
若 CPU 为特殊型号,如赛扬处理器(一般用于深度定制的硬件中),请关注 CPU 是否支持 avx 指令集。如果不支持,请在paddle官网安装 noavx 版本
```

• 使用NVIDIA GPU 基础版预测时必须安装(目标跟踪、表格预测除外):

python -m pip install paddlepaddle-gpu==2.2.2.post101 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA10.1的 PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2.post110 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.0的 PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2.post111 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.1的 PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2.post112 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.2的 PaddlePaddle

不同cuda版本的环境,请参考paddle文档安装合适的 paddle 版本。 不被 paddle 支持的 cuda 和 cudnn 版本,EasyEdge 暂不支持

安装 OpenVINO 使用 x86_64 CPU 加速版 SDK 预测时必须安装。

1) 请参考 OpenVINO toolkit 文档安装 2021.4版本, 安装时可忽略Configure the Model Optimizer及后续部分

2) 运行之前,务必设置环境变量

source /opt/intel/openvino_2021/bin/setupvars.sh

安装 cuda、cudnn

• 使用Nvidia GPU 加速版预测时必须安装。 依赖的版本为 cuda9.0、cudnn7。版本号必须正确。

安装 pytorch (torch >= 1.7.0)

- 目标跟踪模型的预测必须安装pytorch版本1.7.0及以上(包含:Nvidia GPU 基础版、x86_64 CPU 基础版)。
- 目标跟踪模型 Nvidia GPU 基础版还需安装依赖cuda、cudnn。

关于不同版本的pytorch和CUDA版本的对应关系: pytorch官网目标跟踪模型还有一些列举在requirements.txt里的依赖(包括torch >= 1.7.0),均可使用pip下载安装。

pip3 install -r requirements.txt

2. 安装 easyedge python wheel 包 安装说明

pip3 install -U BaiduAI_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_x86_64.whl

具体名称以 SDK 包中的 whl 为准。 安装说明:华为 Atlas300 除了需要安装BaiduAI_EasyEdge_SDK包,还需安装

pip3 install -U EasyEdge_Devkit_Atlas300-{版本号}-cp36-cp36m-linux_x86_64.whl

安装说明:飞腾 Atlas300 除了需要安装BaiduAl_EasyEdge_SDK包,还需安装

pip3 install -U EasyEdge_Devkit_Phytium.Atlas-{版本号}-cp36-cp36m-linux_aarch64.whl

3. 使用序列号激活

丧取序列 号							
共興貨振発说明 发布纯氟线租务,将创练现就的模型部署在本地,真线调用模型。可以选择将模型部署在本地的服务器、小型设备、软键一体方案专项适配建件上, 变几47、SCK进一步奏成,更适差应不同业务结果。 							
服务器 通用小型设备 专项适配硬件							
SDK API							
此处发布、下载的SDK为未授权SDK,需要前往排	空制台获取序列号激活后才能正式使	用。SDK内附有对应版本的Demo及开发文档,开	我者可参考源代码完成开发。				
模型名称	发布版本 🗘	应用平台	模型加速	发布状态	发布时间 🗘		
		通用X86 CPU-Linux	基础版	•已发布	2021-08-19 20:24	下载SDK	
	134318-V1 度看性能报告		精度无损压缩加速	•已发布	2021-08-19 20:24	下载加速版SDK	
		英伟达GPU-Linux	基础版	 已发布 	2021-08-19 20:35	下载SDK	
sun_())⊟ #titest			精度无损压缩加速	•已发布	2021-08-19 20:34	下载加速版SDK	
		WARDOW THE AVE	基础版	 已发布 	2021-08-19 18:17	下载SDK	

修改demo.py 填写序列号

```
pred = edge.Program()
pred.set_auth_license_key("这里填写序列号")
```

4. GPU 加速版 使用 GPU 加速版,在安装完 whl 之后,必须:

- 1. 从这里下载 TensorRT7.0.0.11 for cuda9.0,并把解压后的 lib 放到 C++ SDK 的 lib 目录或系统 lib 目录
- 2. 运行时,必须在系统库路径中包含 C++ SDK 下的lib目录。如设置LD_LIBRARY_PATH

cd \${SDK_ROOT}

1. 安装 python wheel 包 tar -xzvf python/*.tar.gz pip install -U {对应 Python 版本的 wheel 包}

2. 设置 LD_LIBRARY_PATH tar -xzvf cpp/*.tar.gz export EDGE_ROOT=\$(readlink -f \$(Is -h | grep "baidu_easyedge_linux_cpp")) export LD_LIBRARY_PATH=\$EDGE_ROOT/lib

3. 运行 demo python3 demo.py {RES文件夹路径} {测试图片路径}

如果是使用 C++ SDK 自带的编译安装的 OpenCV, LD_LIBRARY_PATH 还需要包括 C++ SDK的 build 目录下的 thirdparty/lib 目录

如果没有正确设置 LD_LIBRARY_PATH,运行时可能报错:

ImportError: libeasyedge.so.0.4.3: cannot open shared object file: No such file or directory ImportError: libopencv_core.so.3.4: cannot open shared object file: No such file or directory

5. 测试 Demo

```
5.1 图片预测
```

输入对应的模型文件夹(默认为RES)和测试图片路径,运行:

python3 demo.py {model_dir} {image_name.jpg}

测试效果:



5.2 视频预测 (适用于目标跟踪)

输入对应的模型文件夹(默认为RES)和测试视频文件路径 / 摄像头id / 网络视频流地址,运行:

```
**video_type: 输入源类型 type:int**
**1 本地视频文件**
**2 摄像头的index**
**3 网络视频流**
**video_src: 输入源地址,如视频文件路径、摄像头index、网络流地址 type: string**
python3 demo.py {model_dir} {video_type} {video_src}
```

5.3 表格预测

输入对应模型文件夹(默认为RES)和测试数据地址(csv文件地址),运行:

python3 demo.py {model_dir} {/xxx/xxx.csv}

6. 测试Demo HTTP 服务 输入对应的模型文件夹(默认为RES)、序列号、设备ip和指定端口号,运行:

 $python3 \ demo_serving.py \ \{model_dir\} \ \{serial_key\} \ \{host, \ default \ 0.0.0.0\} \ \{port, \ default \ 24401\}$

后,会显示:

Running on http://0.0.0.0:24401/

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片或者视频来进行测试。也可以参考`demo_serving.py`里 http_client_test()函数请求http服务进行推理。

	【图像分割】45274 分割-电池-设备端V1	调整阈值 ————————————————————————————————————	—————————————————————————————————————
		label	置信度
	-	battery	0.998
		tomato	0.997
		tomato	0.996
		battery	0.995
		tomato	0.995
		tomato	0.993
		tomato	0.972
1	350		

图片加载失败

使用说明

使用流程 demo.py

import BaiduAI.EasyEdge as edge

```
pred = edge.Program()
pred.set_auth_license_key("这里填写序列号")
pred.init(model_dir={RES文件夹路径}, device=edge.Device.CPU, engine=edge.Engine.PADDLE_FLUID)
pred.infer_image({numpy.ndarray的图片})
pred.close()
```

import BaiduAl EasyEdge as edge

from BaiduAI.EasyEdge.serving import Serving

 $server = Serving(model_dir=\{RES文件夹路径\}, \ license=serial_key)$

请参考同级目录下demo.py里:

 $** pred.init(model_dir=xx, \ device=xx, \ engine=xx, \ device_id=xx)**$

对以下参数device\device_id和engine进行修改

 $server.run(host=host, \ port=port, \ device=edge.Device.CPU, \ engine=edge.Engine.PADDLE_FLUID)$

初始化

• 接口

```
def init(
self,
model_dir,
device=Device.CPU,
engine=Engine.PADDLE_FLUID,
config_file="conf.json",
preprocess_file="preprocess_args.json",
model_file="model"
params_file="params"
label_file="label_list.txt"
infer_cfg_file="infer_cfg.json",
device_id=0,
thread_num=1,
 ....
Args:
   model_dir: str
   device: BaiduAI.EasyEdge.Device,比如:Device.CPU
   engine: BaiduAI.EasyEdge.Engine,比如:Engine.PADDLE_FLUID
   config_file: str
   preprocess_file: str
   model_file: str
   params_file: str
   label_file: str 标签文件
   infer_cfg_file: 包含预处理、后处理信息的文件
         device_id: int 设备ID
   thread_num: int CPU的线程数
Raises:
   RuntimeError, IOError
Returns:
```

使用 NVIDIA GPU 预测时,必须满足:

bool: True if success

- 机器已安装 cuda, cudnn
- 已正确安装对应 cuda 版本的 paddle 版本
- 通过设置环境变量FLAGS_fraction_of_gpu_memory_to_use设置合理的初始内存使用比例

使用 CPU 预测时,可以通过在 init 中设置 thread_num 使用多线程预测。如:

 $pred.init(model_dir=_model_dir,\ device=edge.Device.CPU,\ engine=edge.Engine.PADDLE_FLUID,\ thread_num=1)$

预测图像

接口

Args:	
img: n	p.ndarray or bytes
thresh	nold: float
only	y return result with confidence larger than threshold
chann	el_order: string
cha	annel order HWC or CHW
color_	format: string
colo	or format order RGB or BGR
data_t	type: string
ima	age data type
Returns:	
list	

• 返回格式: [dict1, dict2, ...]

字段	类型	取值	说明
confidence	float	0~1	分类或检测的置信度
label	string		分类或检测的类别
index	number		分类或检测的类别
x1, y1	float	0~1	物体检测,矩形的左上角坐标 (相对长宽的比例值)
x2, y2	float	0~1	物体检测,矩形的右下角坐标(相对长宽的比例值)
mask	string/numpy.ndarray	图像分割的mask	

关于矩形坐标

```
x1 * 图片宽度 = 检测框的左上角的横坐标
```

y1 * 图片高度 = 检测框的左上角的纵坐标

x2 * 图片宽度 = 检测框的右下角的横坐标

y2 * 图片高度 = 检测框的右下角的纵坐标

可以参考 demo 文件中使用 opencv 绘制矩形的逻辑。

结果示例

• i) 图像分类

{	
	"index": 736,
	"label": "table",
	"confidence": 0.9
}	

● ii) 物体检测

```
{
    "y2": 0.91211,
    "label": "cat",
    "confidence": 1.0,
    "x2": 0.91504,
    "index": 8,
    "y1": 0.12671,
    "x1": 0.21289
```

}

```
"name": "cat",
"score": 1.0,
"location": {
"left": ...,
"top": ...,
"width": ...,
"height": ...,
},
"mask": ...
```

mask字段中,data_type为numpy时,返回图像掩码的二维数组

{ { {0,0,0,0,0,0,0,0,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,0,1,1,1,0,0,0,0}, {0,0,0,0,0,0,0,0,0,0}, } 其中1代表为目标区域,0代表非目标区域

data_type为string时,mask的游程编码,解析方式可参考 demo

预测视频 (目前仅限目标跟踪模型调用)

```
● 接口
```

```
def infer_frame(self, frame, threshold=None):
"""
视频推理(抽帧之后)
:param frame:
:param threshold:
:return:
"""
```

返回格式dict

字段	类型	说明
pos	dict1	当前帧每一个类别的追踪目标的像素坐标(tlwh)
id	dict2	当前帧每一个类别的追踪目标的id
score	dict3	当前帧每一个类别的追踪目标的识别置信度
label	dict4	class_idx(int)与label(string)的对应关系
class_num	int	追踪类别数

预测声音

 使用声音分类SDK需要安装额外依赖 pip 安装 resampy pydub 音频默认格式支持wav文件预测,如果需要预测mp3等其他音频格式的数据需要 系统额外安装ffmpeg (windows系统的ffmpeg已集成在sdk中无需额外安装,linux系统需要手动安装)

● 接口

```
def infer_sound(self, sound_binary, threshold=0.3):
    """
Args:
    sound_binary: sound_binary
    threshold: confidence
Returns:
```

list

• 返回格式: [dict1, dict2, ...]

字段	类型	取值	说明
confidence	float	0~1	分类的置信度
label	string		分类的类别
index	number		分类的类别

表格预测

• 考虑到表格字段内容和长度的不固定性,我们建议您参考"校验服务"页面提供的详细信息。您可以访问该页 面:https://ai.baidu.com/easydl/app/validate/ml/models/verify,并从中复制数据请求的 Body 部分作为参考模板。这将帮助您理解如何灵 活处理各种不同的字段。

接口

```
def infer_csv(self, data):
 .....
 结构化数据推理
 Args:
   data: pd.DataFrame or list or dict
 Returns:
 .....
```

• 返回格式: list 接口直接反馈预测结果数组

升级模型 适用于经典版升级模型,执行bash update_model.sh,根据提示,输入模型路径、激活码、模型ID、模型版本,等待模型更新完毕即 可。

FAQ

Q: EasyDL 离线 SDK 与云服务效果不一致,如何处理? A: 后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

Q: 运行时报错 "非法指令" 或 " illegal instruction" A: 可能是 CPU 缺少 avx 指令集支持,请在paddle官网 下载 noavx 版本覆盖安装

Q: NVIDIA GPU预测时,报错显存不足: A: 如以下错误字样:

paddle.fluid.core.EnforceNotMet: Enforce failed. Expected allocating <= available, but received allocating:20998686233 > available:19587333888. Insufficient GPU memory to allocation. at [/paddle/paddle/fluid/platform/gpu_info.cc:170]

请在运行 Python 前设置环境变量,通过export FLAGS_fraction_of_gpu_memory_to_use=0.3来限制SDK初始使用的显存量,0.3表示初始使用 30%的显存。如果设置的初始显存较小, SDK 会自动尝试 allocate 更多的显存。

Q:我想使用多线程预测,怎么做?如果需要多线程预测,可以每个线程启动一个Progam实例,进行预测。 demo.py文件中有相关示例代码。

注意: 对于CPU预测,SDK内部是可以使用多线程,最大化硬件利用率。参考init的thread_num参数。

Q: 运行SDK报错 Authorization failed

情况一: 日志显示 Http perform failed: null respond 在新的硬件上首次运行, 必须联网激活。

SDK 能够接受 HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888" ./easyedge demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限 于)以下可能的情况:

- MAC地址变化
- 磁盘变更

● BIOS重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 ~/.baidu/easyedge 目录,再重新激活。

情况三: Atlas Python SDK日志提示ImportError: libavformat.so.58: cannot open shared object file: No such file or directory 或者其他类似so找不到 可以在LD_LIBRARY_PATH环境变量加上libs和thirdpartylibs路径,例如

export LD_LIBRARY_PATH=/xxx/libs:/xxx/thirdpartylibs:\$LD_LIBRARY_PATH # tips: 这里/xxx需要替换为真实路径,/xxx路径查找方法如下

查找安装包内libs和thirdpartylibs路径的方法如下(以华为Atlas300 SDK为例,其他SDK查找方法类似):

pip3 show EasyEdge-Devkit-Atlas300 # 结果中会显示 Location 路径,也就是包的安装路径 **libs和thirdpartylibs两个路径在 Location 所指示的路径 easyedge_CANN 子文件夹下**

心 Linux集成文档-Atlas

简介

本文档介绍EasyEdge/EasyDL的Linux Atlas SDK的使用方法。

注意Atlas有两种产品形态,Atlas 200和Atlas 300,请参见此处的文档说明

- 网络类型支持:图像分类
- 硬件支持:
 - CPU: aarch64
 - Atlas 300 卡
- 操作系统支持: Atlas指定的Linux版本, Ubuntu 16.04 x86_64 或 centos 7 x86_64, 请从Atlas文档中下载。

Release Notes

时间	版本	说明
2020.3.23	0.1	初始版本,支持图像分类

性能数据

数据仅供参考,实际数值根据使用线程数、利用率等情况可能有所波动

模型类型	模型算法	芯片类型	SDK类型	实测硬件	单次预测耗时
EasyDL 图像分类	高性能	Atlas 300	Atlas 300	Atlas 800服务器	9ms
EasyDL 图像分类	高精度	Atlas 300	Atlas 300	Atlas 800服务器	12ms
EasyDL 物体检测	高性能	Atlas 300	Atlas 300	Atlas 800服务器	11ms
EasyDL 物体检测	高精度	Atlas 300	Atlas 300	Atlas 800服务器	31ms

atlas 300 加速卡注意事项

一般服务器(HOST侧)安装多个300加速卡,,每个300加速卡有4个芯片。一个芯片(DEVICE侧)可以认为是一个单独的系统,并且不共享储存 系统。

每个芯片都有独立的device-id,可以通过命令查看: sudo npu-smi info

由于模型需要在芯片上运行。因此运行SDK前,需要手动将模型复制到每个单独芯片的储存系统上。

测试atlas 300的官方demo

环境准备

请参见此处的文档说明, 搭建环境,测试HelloDavinci demo通过后,再测试本demo

修改300加速卡SSH密码 (可选)

请在咨询华为技术人员后,修改Device登录密码

```
ssh HwHiAiUser@192.168.1.199
**登录后会强制修改密码**
ssh HwHiAiUser@192.168.1.198
```

快速开始

SDK在以下环境中测试通过

• ubuntu 16.04, Atlas 800 服务器指定版本;

Atlas DDK 的ddk_info信息:

```
{
    "VERSION": "1.3.8.B902",
    "NAME": "DDK",
    "TARGET": "ASIC"
}
```

1.安装软件

sudo apt-get install sshpass build-essential

2. 测试Demo

编译运行:

下载后,模型资源文件默认已经打包在开发者下载的SDK包中,

Step 0 : 使用HwHiAiUser登录

Step 1: 运行一次install-demo.sh脚本,会得到测试demo。

Step 2:请在官网获取序列号,填写在demo_async.cpp及demo_sync.cpp的开始处license_key字段。



图片加载失败

step3: 准备测试图片

覆盖image目录下的 1.jpg,更多图片可以用于demo中的批量测试模式

step4(可选): 修改test_300.sh下的以下开发板登录信息

export DDK_PATH=\$HOME/tools/che/ddk/ddk # ddk的安装路径

declare -a DIVICE_IPS=("192.168.1.199") # 300加速卡芯片的ip地址,device=0 对应192.168.1.199 DEVICE_PASSWORD="Huawei@SYS3" # 之前 修改300加速卡SSH密码 MAIN_CPP="demo_async.cpp" # demo_async.cpp" 异步接口, "demo_async.cpp" 同步接口

OpenCV_install_dir=/home/HwHiAiUser/opencv_x64/ # OpenCV 3.4版本,需要存在 \${OpenCV_install_dir}/share/OpenCV/OpenCVConfig.cmake文件

step5: 运行demo,会自动编译OpenCV 3.4库,如果报错请自行编译,目录设置在 OpenCV_install_dir

cd demo sh test_300.sh

图像分类demo运行效果:

[stat] [100001]image/1.jpg(4 images) time used: 41ms (at 1583765958531) total:705ms [result][100001][image/1.jpg][281470472005664] is: n07747607 orange 0.973633 950;

n07747607 orange 分类名 0.973633 分类概率 950 分类名的序号

物体检测的demo运行效果:

[stat] time used : 101ms; all time used:478 images[3] result: label:no2_ynen;prob:0.985352 loc:[(0.459961,0.839844), (0.5625,0.988281)]

no2_ynen 分类名 , 也可以获取分类名的序号 0.985352 分类概率 loc:[(0.459961,0.839844), (0.5625,0.988281)] , 检测框的位置。 (0.459961,0.839844)表示左上角的点,(0.5625,0.988281)右下角的 点; 如原始图片608, 左上角(0.459961*608,0.839844*608) , 右下角(0.5625*608,0.988281*608)

SDK接口使用

使用该方式,将运行库嵌入到开发者的程序当中。

同步接口使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

```
// step 0: 设置序列号
global_controller()->set_licence_key("set your license here");
```

// step 1: 配置模型资源目录 AtlasConfig config; config.model_dir = {模型文件目录};

// step 2: 创建并初始化Predictor ;

auto predictor = global_controller()->CreateEdgePredictor(config); int ret = predictor->init(); # 若返回非0,请查看输出日志排查错误原因。 auto img = cv::imread({图片路径}); // step 3: 预测图像 std::vector<EdgeResultData> result2; predictor->infer(img, result2); # 解析result2即可获取结果

异步接口使用流程
// step 0: 设置序列号 global_controller()->set_licence_key("set your license here");

// step 1: 配置模型资源目录 AtlasConfig config; config.model_dir = {模型文件目录};

// step 3: 创建Predictor; 这这里选择合适的引擎 auto predictor = global_controller()->CreateEdgePredictor(config);

// step 4: 设置异步回调 predictor->set_result_handler(YOUR_HANDLER);

// step 5: 初始化
int ret = predictor->init();
若返回非0,请查看输出日志排查错误原因。

// step 6: 预测图像
auto img = cv::imread({图片路径});
color_format = kBGR;
float threshold = 0.1;

uint64_t seq_id; predictor->infer_async(img, color_format, 0.1, nullptr, seq_id); **YOUR_HANDLER里面有seq_id的回调结果**

设置序列号

请在网页控制台中申请序列号,并在init初始化前设置。 LinuxSDK 首次使用需联网授权。

接口

```
virtual int set_licence_key(const std::string& license) = 0;
```

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

EdgeLogConfig log_config; log_config enable_debug = true; global_controller()->set_log_config(log_config);

日志及报错

日志

日志需要开启Atlas 的 INFO级别,/etc/slog.conf中配置关闭zip格式。清空/var/dlog 目录,运行atlas 300 官方示例代码,可以在/var/dlog目录 下看见host和device开头的2个日志文件,中间是明文的info级别的日志

日志共有3处:

- host 测的easyedge.log。当前运行目录下。
- device侧的easyedge.device.xxx.log。 device侧的日志,在芯片的同名目录下。
- /var/dlog host 与device开头的log文件,ddk运行日志,其中device侧有略微延时

通用错误码

错误码	常量	解释
100000 4	RESOURCE_LOAD_F AIL	缺少data/model/conf.json文件或者该文件以及被改动。下载包中的data/model下的所有文件都不要改动,尝试 使用默认配置。或者按照报错复制到对应目录。
700000 1	AUTH_FAILED	服务端校验序列号失败
700000 2	AUTH_LICENSE_INV ALID	校验序列号
700000 3	AUTH_LICENSE_EXP IRED	序列号过期
500001	NET_CURL_PERFOR M_FAILED	服务端校验序列号的请求因为网络原因失败
600001	GET_MACHINE_ID_F AILED	没有相关权限,请反馈

Atlas SDK 错误码

错误码 值	常量	含义	报错示例信息	示例解释及解决方式
120000 11	FILE_NOT _READABL E	资源文件不可读	data/model/params IS NOT READABLE	data/model/params,这个文件不可读。 SDK下载包中的data/model下的所有文件 都不要改动,尝试使用默认配置。或者按 照报错复制到对应目录。
120000 12	HIAI_ERRO RLIST _FILE	status.h.list不是原始文件	data/model/status.h.list IS TOO SMALL	下载包中的data/model下的所有文件都不 要改动,包括status.h.list
120001 02	PREDICTO R_NOT_INI TED	create后没有调用init()函数	please call init() first	调用infer函数前没有调用init()
120001 03	PREDICTO R_NO_HAN DLER	create后没有调用 set_result_handler()函数	please call set_result_handler() first	调用infer_async函数前没有调用 set_result_handler(),建议init前调用
120001 04	PREDICTO R_ALREAD Y_INITED	init()不管是否成功,不能连续调 用。	don't call init() more than once	如果失败,请再次新建一个Predictor
120001 05	BATCH_SIZ E	AltasConfig里的batch_size设置与 model_name不符合	model batch size is 1; your config batch size is 4	batch_size设置里4,model_name设置里 params,不对应导致报错。model_name 应该设置为params-batch4
120001 06	INPUT_WID TH	preprocess_args.json被改动	model input tensor width is 224; your config resize is 226	请勿修改preprocess_args.json
120001 07	INPUT_HEI GHT	同上	同上	同上
120002 01	BATCH_TO O_MANY_I MAGES	一次输入的图片大于batch_size	too_many_images input:2; batch_size is 1	调用infer函数,输入了2张图片,大于 batch数。如果batch=1的话,每次infer只 能传一张图。
120002 02	IMAGE_FO RMAT_CHA NNELS	infer函数输入的color_format与 cv::Mat里的channel数不匹配	EdgeColorFormat is not according to cv channels; format is 101; channels is 3; seq_id1	101表示kRGBA,cv::Mat里channel应该期 望是4。如果是直接读的图片,填kBGR。
122000 01	ENGINE_M ATRIX _COMMON	Atlas DDK Matrix部分(非 CreateGraph函数)接口报错。即 返回值HIAI_StatusT不是 HIAI_OK。具体解释见Atlas官方文 档。	hiai::Graph::ParseConfigFile (graph.prototxt) ; status Code is 16855066 ; HIAI ERROR CODE is 101 HIAI_GRAPH_PROTO_FILE _PARSE_FAILED_CODE,	调用hiai::Graph:: ParseConfigFile()返回 16855066 ,对应的status.h.list中的错误 码是101。保留日志,具体见Atlas官方文 档。
122000 02	ENGINE_AI _COMMON	Atlas DDK Device引擎部分 hiai::AlStatus 不为hiai::SUCCESS	_ai_model_manager->Process()	保留日志,具体见Atlas官方文档。
122000 03	ENGINE_M ATRIX_INIT	Atlas DDK CreateGraph() 初始化 DDK报错。具体解释见Atlas官方 文档。	hiai::Graph::CreateGraph(); data/model/graph.prototxt ; status Code is 16855190 ; HIAI ERROR CODE is 225 HIAI_FILE_NOT_EXIST_CODE,	示例为缺少libatlas_device.so导致
122000 04	EDGE <i>ATLA</i> <i>S_ENGINE</i> MATRIX_INI T_DEVICE	Atlas DDK CreateGraph() 初始化 DDK报错,这个报错很可能是 device侧出现问题	hiai::Graph::CreateGraph() ; data/model/graph.prototxt ; status Code is 16855057 ; HIAI ERROR CODE is 92 HIAI_GRAPH_ENGINE _INIT_FAILED_CODE	需要具体排查DEVICE侧日志再次找具体报 错,发现原因
122000 05	ENGINE_AR GS_NULL	内部错误		请反馈
123000 01	SYNC_INFE R_TIMEOU T	调用infer同步接口时,内部会调用 infer_async函数,这个函数超时	infer sync wait timout more than 10ms	内部会调用infer_async函数超过10ms。1. 不要并发过高 2. 超时参数略微大些。

心图像分类服务器端SDK集成文档-EdgeKitProxy

简介

本文档介绍EdgeKitProxy的使用方法。

Release Notes | 时间 | 版本 | 说明 | | ------- | ------ | ------ | | 2023-05-17 | 1.0.0 | 第一版 ! |

快速开始

二进制位置

位于SDK内bin目录中,文件名为edgekit_serving, 配套edgekit_serving.yml为默认配置文件

注意事项

请参考各SDK文档中的注意事项

使用说明

服务启动

```
usage: edgekit_serving [<flags>]
Flags:
                 显示帮助
  --help
-c, --cfg=./edgekit_serving.yml
                配置文件
-m, --model_dir=./RES 模型目录
 -s, --serial_num=ABCD-EFGH-IJKL-MNOP
               序列号
  --pool_min_size=1 预测池最小预测器个数
  --pool_max_size=1 预测池最大预测器个数
   --pool_full_interval_seconds=-1
                预测池满载多少秒进行扩容
   --pool_idle_interval_seconds=-1
                预测池未满载多少秒进行缩容
   --pool_available_device=-1 ...
              预测池可用设备列表
  d, --debug 开启debug模式
--log_to_std 日志输出至终端
--log_to_file 日志输出至文件
 -d, --debug
  --log_file=easyedge.log 日志文件名
  --log_max_size=10 日志最大大小(MB)
--log_max_age=10 日志旧文件保留天数
   --log_max_backups=100 日志旧文件保留个数
-h, --host=127.0.0.1 服务监听地址
 -p, --port=24401
                   服务监听端口
   --ws_max_handle_num=1 websocket接口最大处理请求个数
   --ws_max_handle_timeout=30
                websocket接口超时时间
```

配置文件说明

```
controller:
 serialNum: AAAA-AAAA-AAAA # 序列号
 modelDir: ../../.RES # 模型目录
predictorPool:

        minSize: 1
        # 预测池最小预测器个数

        maxSize: 3
        # 预测池最大预测器个数

 fullIntervalSeconds: 1 # 预测池满载多少秒进行扩容
 idleIntervalSeconds: 1 # 预测池未满载多少秒进行缩容
 availableDevice: [-1] # 预测池可用设备列表
serving:

        host: 0.0.0.0
        # 服务监听地址

        port: 24401
        # 服务监听端口

 enableHTTP: true # 对外开启HTTP服务
 enableWS: false # 对外开启websocket服务
 WS:
 maxHandleNum: 1 # websocket接口最大处理请求个数
maxHandleTimeout: 30 # websocket接口超时时间
logging:

    debug: true
    # 开启debug模式

    logToStd: true
    # 日志输出至终端

    logToFile: false
    # 日志输出至文件

 logFile: easyedge.log # 日志文件名

        maxSize: 10
        # 日志最大大小 (MB)

        maxAge: 10
        # 日志旧文件保留天数
```

命令行参数会覆盖配置文件中同义配置

maxBackups: 100 # 日志旧文件保留个数

服务调用

HTTP服务接口url: \${监听地址}/ HTTP服务接口url: \${监听地址}/ws

请求参数

syntax = "proto3";

package easyedge.kit.proxy;

```
enum ImageType {
Bin = 0; // 图片原始二进制内容, json格式下为base64编码后结果
 Mat = 1; // 图片Mat格式内容, json格式下为base64编码后结果
message HTTPRequest {
bytes image = 1;
ImageType image_type = 2;
int32 height = 3;
int32 width
              = 4;
int32 channel = 5;
float threshold = 6;
int32 top_num = 7;
enum CommandType {
GetInfo = 0;
InferImage = 1;
enum InfoType {
Hardware = 0;
message WebSocketRequest {
string request_id = 1;
CommandType command_type = 2;
InfoType info_type = 3;
bytes image = 4;
ImageType image_type = 5;
int32 height = 6;
int32 width
               = 7;
int32 channel = 8;
int64 frame_id = 9;
float threshold = 10;
int32 top_num = 11;
```

返回参数

```
syntax = "proto3";
package easyedge.kit.proxy;
```

```
message BasicGPUInfo {
string productName = 1;
 string memUsed = 2;
 string memTotal = 3;
 string gpuUtil = 4;
 string powerLimit = 5;
 string powerDraw = 6;
 string temperature = 7;
message DevStat {
string name = 1;
 uint64 rx = 2;
 uint64 tx = 3;
message Chip {
string name = 1;
 double powerUsed = 2;
 double powerLimit = 3;
```

من بغمينة متحديدة ما عان حام

```
double temperature = 4;
double chipUtil = 5;
int64 memoryUsed = 6;
int64 memoryTotal = 7;
message SMI {
string name
                = 1
string sdkVersion = 2;
string driverVersion = 3;
repeated Chip chips = 4;
message HInfo {
string osName
                           = 1:
                          = 2;
 string hostname
repeated string ipAddr
                           = 3;
 repeated string macAddr
                             = 4;
 uint64 bootTime
                             = 5
 int32 cpuCores
                            = 6:
double cpuMhz= 7;string cpuModelName= 8;double cpuUsage= 9;
 map<string, double> cpuUsageDetail = 10;
 uint64 memTotal = 11;
uint64 memTotalUsed= 12;double memUsage= 13;
 map<string, double> memUsageDetail = 14;
 uint64 diskTotal = 15;
uint64 diskTotalUsed= 16;double diskUsage= 17;
 map<string, double> diskUsageDetail = 18;
 string userName = 19;
 bool isInternetConnected
                             = 20;
string deviceId= 21;int64 deviceTimestamp= 22;
 map<string, DevStat> netUsageDetails = 23;
 repeated BasicGPUInfo gpuInfo = 24;
                = 25;
Fotal ____
 double gpuUtil
                            = 26;
 uint64 gpuMemTotal
 uint64 gpuMemTotalUsed
                             = 27;
 double gpuMemUsage
                                = 28:
 map<string, SMI> aiChipInfo
                                = 29:
message LocationPoint {
optional int32 x = 1;
optional int32 y = 2;
message Location {
optional int32 left
                       = 1;
optional int32 top
                       = 2;
optional int32 width
                        = 3,
 optional int32 height
                        = 4;
 repeated LocationPoint points = 5;
message Point {
optional double x = 1;
optional double y = 2;
message InferResultItem {
optional int64 index = 1;
 optional double confidence = 2;
 optional double score = 3;
 optional string label = 4;
 optional string name = 5;
 entional int20 madall/ind
```

optional m_{132} modelning = o;

// 矩形检测	
optional double x1	= 7;
optional double x2	= 8;
optional double y1	= 9;
optional double y2	= 10;
optional Location locati	ion = 11;
// 四边形检测	
repeated Point points	= 12;
// 追踪	
optional int64 trackId	= 13;
optional int64 frame	= 14;
optional double fps	= 15;
optional string mask	= 16;
}	
message HTTPResponse	} {
int64 cost_ms	= 1;
int32 error_code	= 2;
int64 frame_id	= 3;
repeated InferResultIter	m results $= 4;$
}	
message WebSocketInfe	rResponse {
string request_id	= 1;
int64 cost_ms	= 2;
int32 error_code	= 3;
int64 frame id	= 4;
repeated InferResultIte	m results $= 5$;
bytes annotated	= 6; // 渲染后的图片原始二进制内容,json格式下为base64编码后结果,目前语义分割返回这个类型
}	
message WebSocketHIn	foResponse {
string request_id	= 1;
int32 status =	2;
string msg =	3;
HInfo data =	4;
}	

其他说明

单机负载均衡

通过配置文件或命令行参数配置了预测池相关配置后,若预测池最小与最大预测器个数不同,且扩缩容配置不为-1则开启单机负载均衡,服务启 动时会创建最小数量的预测器,后续根据实际请求情况,若所有预测器均有负载的持续时间大于配置中的满载扩容时间,且预测器数量未到达最 大个数时,会自动扩容,后续若请求并发数下降,预测器池中预测器不能跑满负载时,则会自动缩容,尽可能最大化利用单机资源

⁰ 纯离线API集成说明

本文档主要说明定制化图像分类模型发布为本地服务器API(通过API部署包实现)后如何使用。如还未训练模型,请先前往EasyDL进行训练。

如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

部署包使用说明 部署方法

EasyDL定制化图像分类模型的服务器API通过EasyPack实现,目前提供单机一键部署的方式。

在EasyDL申请、下载部署包后,在本地服务器新建目录(建议目录命名规则:easyDL_服务名称_模型版本号),将软件包上传至该目录。请参 考EasyPack-单机一键部署使用python2 命令来部署,部署成功后,启动服务,即可调用与在线API功能类似的接口。 运维检查

EasyDL服务器API部署应用健康检查(或故障排查)脚本:trouble_shooting.tar

脚本能力:鉴权服务健康检测、容器状态检查、端口探活、网络联通性测试、容器关键报错日志输出等

使用方法:将脚本上传至服务器任意目录(或在服务器直接下载),并解压后运行。

```
**解压**
tar vxf trouble_shooting.tar
**执行**
bash trouble_shooting.sh
```

授权说明

部署包根据服务器硬件(CPU单机或GPU单卡)进行授权,只能在申请时提交的硬件指纹所属的硬件上使用。

部署包测试期为1个月,如需购买永久授权,可提交工单咨询

性能指标

图像分类模型可部署在CPU或GPU服务器上,单实例具体性能指标参见算法性能及适配硬件 API参考

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在EasyDL进行自定义模型训练,完成训练后申请部署包,部署成功后拼接url。

请求URL: http://{IP}:{PORT}/{DEPLOY_NAME}/ImageClassification IP:服务部署所在机器的ip地址 PORT:服务部署后获取的端口 DEPLOY_NAME:申请时填写的服务名称

Header如下:

参数 值 Content-Type application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,与其他图像识别服务不同的是定制化图像识别服务以json方式请求。

Body请求示例:

```
{
"image": "<base64数据>",
"top_num": 5
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大4096px,支持 jpg/png/bmp格式 注意请去掉头部
top_nu m	否	number	-	返回分类数量,默认为6个

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	分类结果数组
+name	否	string	分类名称
+score	否	number	置信度

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如缺少必要出入参时返回:

```
{
"error_code": 336001,
"error_msg": "Invalid Argument"
```

}

错误码	错误信息	描述
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度智能云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有疑 问请在百度智能云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数代码格式是否有误。有疑问请在百度智能云控制台内提交工单反馈
336003	Base64解码失败	图片/音频/文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。 有疑问请在百度智能云控制台内提交工单反馈
336004	输入文件大小不合法	图片超出大小限制,图片限4M以内,请根据接口文档检查入参格式,有疑问请在百度智能云控制台内提交工 单反馈
336005	图片解码失败	图片编码错误(非jpg,bmp,png等常见图片格式),请检查并修改图片格式
336006	缺失必要参数	image字段缺失(未上传图片)
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度智能云控制台内提 交工单反馈
337000	Auth check failed	离线鉴权调用失败

模型更新/回滚操作说明

模型更新

1、在EasyDL-纯离线服务发布页面,找到您的服务器API发布记录,点击【更新版本】,选择「更新包」或「完整包」来发布。

两者区别:

包类型	描述
更新包	仅包含最新的模型应用,需执行download.sh脚本下载所需镜像等依赖文件
完整包	包含模型应用和其他鉴权服务,需执行download.sh脚本下载所需完整依赖文件

2、(CPU模型可忽略)如果您训练的模型为GPU版本,系统会生成多份下载链接。请在GPU服务器执行 nvidia-smi命令,根据返回的Cuda Version来选择对应的部署包链接下载。

3、在服务器新建目录(建议标记对应模型的版本号,便于区分不同模型版本),如easydl_\${DEPLOY_NAME}_v2

\${DEPLOY_NAME}:申请时填写的服务名称

以如下场景举例说明:模型版本从V1升级至V2

```
**1.新模型准备**

**创建指定版本的目录**

mkdir easedl_${DEPLOY_NAME}_v2

cd easedl_${DEPLOY_NAME}_v2

**将部署包上传至服务器该目录并解压**

tar zvxf xx.tar.gz

**解压后,进入指定目录执行download脚本下载模型所依赖文件**

cd original && bash download.sh

**2.旧模型备份**
```

```
**历史模型备份**
cp-r /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V1
**记录当前模型的端口号**
docker ps -a |grep ${DEPLOY_NAME}
```

```
**3.模型升级**
cd package/Install
**卸载当前已安装的旧的easyDL服务:${DEPLOY_NAME},前面已备份**
python2 install.py remove ${DEPLOY_NAME}
**安装当前部署包内新的EasyDL服务:${DEPLOY_NAME}**
python2 install.py install ${DEPLOY_NAME}
```

(可选操作) 更新证书 python2 install.py lu

模型回滚

以如下场景举例说明:模型版本从V2回滚至V1

方法一:

```
**重命名当前v2模型目录名称**
mv /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V2
**使用V1版本**
cp -r /home/baidu/work/${DEPLOY_NAME}_V1 /home/baidu/work/${DEPLOY_NAME}
```

```
**停止当前模型容器**
docker ps -a |grep ${DEPLOY_NAME}
docker rm -f ${容器名}
```

```
**创建新的容器**
cd /home/baidu/work/${DEPLOY_NAME} && bash start/start-1.sh
```

** (可选操作)进入V1版本部署包所在目录执行license更新操作,假如部署包在/opt目录下,以您实际目录为准** cd /opt/easydl_\${DEPLOY_NAME}_V1/original/package/Install && python2 install.py lu

方法二:

进入模型V1所在目录,参考上述【模型更新】步骤,执行模型升级操作(即先卸载v2,后升级为v1)

- 心 通用小型设备部署
- 心 如何在通用小型设备部署

训练完毕后,可以选择将模型通过「SDK-纯离线服务」或「API-端云协同服务」部署,具体介绍如下:

纯离线服务部署

纯离线服务目前仅支持通过SDK集成,可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可发布设备端SDK:

- 选择模型
- 选择部署方式「EasyEdge本地部署」-「通用小型设备」
- 选择版本
- 选择集成方式
- 点击发布

		说明:
5择模型 男女分类	~	1. 设备端SDK支持Android、iOS、Windows、Linux操作系统,具体的系统、硬
『署方式 EasyEdge本地	部署 > 通用小型设备 >	件环境支持请参考技术文档。提供可直接体验的移动端app安装包,以及相应代码 包、说明文档,供企业用户/开发者二次开发
5择版本 V1	~	2. 如SDK生成失败,或有任何其他问题,欢迎提交工单或加入QQ群
E成方式 💿 SDK-纯离线朋	服务	(679517246) 咨询了解
发布		
A. A. A. A	先择系统与芯片	
与未发布		
型田 次 小り		
纯离线服务 > 2	发布新服务	
部署方式	🔵 服务器 🛛 💿 通用小型	21设备 🔿 专项适配硬件
选择模型	男女分类	~
选择版本	V1	~
选择系统和芯片	+ 👌 Linux	
	— H Windows	
	✓通用X86 CPU Ø	英特尔◎神经计算棒
	🕂 🖷 Android	
	+ 🖷 Android	
	+ 🗭 Android + 🕊 iOS	
模型加速:	 + ♥ Android + ● iOS ✓ 同时获取加速版 ⑦ 	
模型加速:	 + ● Android + ● iOS ✓ 同时获取加速版 ⑦ 	

也可以直接在「EasyEdge本地部署」-「纯离线服务」页面点击发布新服务,按上图所述进行申请发布

端云协同服务部署

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

具体使用说明请参考端云协同服务说明

```
の纯离线SDK说明
```

の纯离线SDK简介

本文档主要说明定制化模型发布后获得的SDK如何使用,如有疑问可以通过以下方式联系我们:

• 在百度智能云控制台内提交工单

• 前往官方论坛交流,与其他开发者进行互动

SDK说明

SDK支持iOS、Android、Linux、Windows四种操作系统。以下为具体的系统、硬件环境支持:

操作系统	系统支持	硬件环境要求
iOS	iOS 8.0 以上 (A仿生芯片版 要求11.0以上)	ARMv7 ARM64 (Starndard architectures) (暂不支持模拟器)
Android	通用ARM: Android 19以上 SNPE:Android 21以上 DDK:Android 21以上	通用ARM: 绝大部分的手机和平板 、比较耗时 SNPE:高通Soc,仅支持Qualcomm Snapdragon 450 之后发布的soc。其中 660 之后的型号可能含 有 Hexagon DSP模块,具体列表见snpe 高通骁龙引擎 DDK:CPU支持华为麒麟970N、980的arm-v8a的soc,支持的机型 mate10,mate10pro,P20,mate20等 支持armeabi-v7a arm-v8a CPU 架构, DDK仅支持 arm-v8a
Linux C++		CPU: AArch64 ARMv7I ASIC: Hisilicon NNIE1.1 on AArch64(Hi3559AV100/Hi3559CV100等) ASIC: Hisilicon NNIE1.2 on ARMv7I(Hi3519AV100/Hi3559V200等)
Linux Python		Intel Movidius MyRIAD2/MyRIAD X
Linux Ubuntu 16.04		AArch64 HUAWEI Atlas 200
Windows	64位 Windows7 及以上	Intel CPU x86_64 Intel Movidius MyRIAD2/MyRIAD X (仅支持Win10) 环境依赖: .NET Framework 4.5 Visual C++ Redistributable Packages for Visual Studio 2013 Visual C++ Redistributable Packages for Visual Studio 2015

说明

针对iOS操作系统:虽然SDK库文件很大(ipa文件很大),但最终应用在用户设备中所占用的大小会**缩小很多,如图像分类下载的ipa文件可 能会100M以上,但实际集成在设备中只有20M左右**。 这与multi architechtures、bitcode和AppStore的优化有关。

单次预测耗时参考

根据具体设备、线程数不同,数据可能有波动,请以实测为准

在算法性能及适配硬件页面查看评测信息表

自适应芯片版SDK

发布SDK时可根据实际应用时的硬件/芯片配置选择最合适的SDK。如"华为NPU版"就是针对华为NPU芯片做了适配与加速的SDK。如实际应用时 需要适配多种芯片,就可以发布"自适应芯片版"SDK,SDK被集成后会自动判断设备的芯片并运行相应的模型。

加速版SDK

发布SDK时,勾选「同时获取加速版」,就可以同时获得适配部分芯片(需选中且右侧带有加速标记)的基础版SDK和加速版SDK。



目前加速版SDK已支持Windows X86、Linux ARM、iOS ARM、Android ARM环境,加速后的SDK普遍在包大小、内存占用、识别速度等方面表现 更优,详细对比请见<u>算法性能及适配硬件</u>。

加速版SDK和基础版的测试方式类似,只需在EasyDL控制台新增「加速版」测试序列号,即可获得3个月的测试期。

激活&使用SDK

SDK的激活与使用分以下四步:

① 在【我的模型】-【服务详情】内下载SDK

我的模型	创建模型								
创建模型									
训练模型	【图像分类】 ljay-gxn	n-2flowers {	模型ID: 340					⊲ 训练	🕒 历史版本
校验模型	部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作		
发布模型 数据中心	私有服务器部署	V1	训练完成	-	-	top1准确率100.00% top5准确率100.00% 完整评估结果	申请发布		
我的数据集 云服务调用数据	【图像分类】 ljay-gjd	-2flowers 巷	莫型ID: 339					⊲ 训练	④ 历史版本
	部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作		
	私有服务器部署	V1	训练完成	-		top1准确率100.00% top5准确率100.00% 完整评估结果	申请发布		
	【图像分类】cmc-高	性能 模型ID	: 332				⊲ 训练	⊙ 历史版4	2 回删除
	部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作		
	设备端SDK	V1	训练完成	审核成功	已发布	top1准确率93.33% top5准确率100.00% 完整评估结果	服务详情申请发	ŧ	

设备端SDK下载

此处下载的SDK为未授权SDK,需要获取序列号激活后才能正式使用。SDK内附有对应版本的Demo 及开发文档,开发者可参考源代码完成开发

操作系统	操作
iOS	下载SDK 🗸 获取序列号
Android	下载SDK 🗸 获取序列号
Linux	下载SDK 🗸 获取序列号
Windows	下载SDK 🗸 获取序列号

② 在<mark>控制台</mark>获取序列号

按单台或多台设备获得授权并使用SDK时:

EasyDL开发平台	纯离线服务			
昍 图像分类概览 ⇔	我的黑线服务 服务器纯黑线服务 设备端纯黑线服务 专项硬件纯属线服务			
数据服务				
⊜ 数据总览	使用说明			
② 数据标注 ^	1. 按单台实例获得授权激活SDK, 操作步骤:			
标签组管理	■ 下载所需模型已发布的设备端SDK → ▼ 在下方购买永久投权或新增测试序列号 → ■ 设备端部署SDK,联网激活后在线使用			
在线标注	 投來例就面景與基礎服务支持定期待宣议發展否正常在线,用户可以自定义检查周期(定期输认时间),当设备损坏或者无法正常在线时当前设备视作铜毁,序列号可接入新设备。 仅支持联网激活,设备需要代证联网,定期确认状态。 特别提示,如周期同一序列号激活多台设备,请商注 使多台设备激活,页面申请;如属开发手机APP,建议选择 按产品线激活 			
智能标注	使用过程中如有其他问题,可以 提交工单咨询 或 前往官方论坛 交流			
□ 数据增强	提示:如您已完成企业资质认证,则享受该账号下每发布一个模型即可免费享受3个基础版SDK测试序列号和3个加速版SDK测试序列号, 联网激活后有效期为6个月更多企业认证权益详见。企业权益专属礼包①			
⊙数据回流 ^				
云服务回流	按单台设备激活 按多台设备激活 按产品线激活			
摄像头采集	+ 购买永久得料 新提测试序列号 管理序列号 音看线计序列号			
模型管理				
◎ 模型训练	按设备 按实例数			

包括两种激活方式:

- 1. 按设备激活:是按单台设备数授权(依据硬件指纹信息),每个序列号可支持多台设备,激活后授权永久有效,但如果硬件信息发生变更 (如拆卸网卡或后续增加网卡),或者硬件损坏,则授权不可恢复
- 2. 按实例数激活:是按单台实例数授权,每隔一次"定期确认时间",后台会确认当前设备是否在线,如果离线,当前设备解绑,新设备可以接 入,该方式仅支持联网激活,设备需要保证联网,定期确认状态。

Android或iOS操作系统的SDK可以选择按产品线激活(仅支持开发手机APP),序列号与包名(Package Name/Bundle ID)绑定:

∃	☆ 控制台总范 全局						
EasyDL开发平台 ¹ 健型训练 模型校验	纯离线服务 我的离线服务 服务器组	离线服务 设备端纯离线服务 ————————————————————————————————————	专项硬件纯离线服	(5			
 ● 模型校验 単型版券 ▲ 公有宏観券 ▲ 公有宏観券 ▲ 在线服券 此量預測 ▲ 本地部署 	使用说明 1. Android/05操作系统 下载所需模型 2. 每发布一个模型图时 特别提示:如果同一个 使用过程中的有法他问 提示:"如愿已完成金金社 取问题无否有效规为令行	LINSDK可接产品线获得接收激活 已发布的设备端SDK → 前增之个测试序列号、联网激活后有 努列导激运多台设备、编辑性复方 和优计量、期等设计器等下每发布- 有更多全级计量发音下每发布- 有更多全级过程发展了中全发布。	(仪支持开发手机APP) 育在下方购买水久 资期为3个月 台设备数据页面申请, 论坛交流 一个模型即可免费享受 给专篇礼包(②)	、序列号与包名(Package Name) 段权或新增需试序转号 → 新増序列号 *序列号类型 (请法将序)	Bundle ID) 御定、操作参察:	后高估使用	×
纯离线服务 端云协同服务 浏览器/小程…	按单台设备激活 括 < + 购买永久投权	多合设备激活 按产品线激活 新增测试序列号 管理門	子列号 查看转:	 ·操作系统 Android ·包名 请输入包括 	× .		
 (少 应用接入 ④ 权限管理 ⑤ 初限管理 	操作系统 iOS	智名 basic.aa.bb.c5	序列号类型			取消	确定
AI市地 3) 我的已购模型 国 售卖模型	Android	no	基础版	未邀活	07C9-5255-4ECF-B0 5	5 0/10000	

③ 本地运行SDK,并完成首次联网激活

通过左侧导航栏查看不同操作系统SDK的开发文档

④ 正式使用

SDK常见问题

通过左侧导航栏查看不同操作系统SDK的FAQ

以下是通用FAQ,如您的问题仍未解决,请在百度智能云控制台内提交工单反馈

1、激活失败怎么办?

按设备激活时,激活失败可能由于以下几个原因造成:

①可能是当前序列号已被其他设备激活使用,请核实序列号后用未被激活的序列号重新激活

②序列号填写错误,请核实序列号后重新激活
③同一台设备绑定同一个序列号激活次数过多(超过50次),请更换序列号后重试
④首次激活需要联网,网络环境不住或无网络环境,请检查网络环境后重试
⑤序列号已过有效期,请更换序列号后重试
⑦如有其他异常请在百度智能云控制台内提交工单反馈
按产品线激活时,激活失败可能由于以下几个原因造成:
①可能是包名填写错误,请核对与序列号绑定的包名是否与实际包名一致
②序列号填写错误,请核实序列号后重新激活
③首次激活需要联网,网络环境不佳或无网络环境,请检查网络环境后重试
④模型发布者和序列号所属账号非同一账号,如果存在这种异常建议更换账号获取有效序列号
⑤序列号已过有效期,请申请延期后重试

⑥如有其他异常请在百度智能云控制台内提交工单反馈

2、怎样申请序列号使用延期

序列号激活后有效期为三个月,可以在控制台进行申请,申请流程:

1) 填写申请信息

EasyDL定制训练平台	产品服务 / EasyDL定制训练平台 - 离线SDK管理 / 申请延期			
概览	申请延期			
应用列表	* 洗坯描刊·	2本3年4至		
监控报表	从11年1天1王。	· 用达FF	·	
技术文档	*选择版本:	请选择	~	
云服务权限管理	* 申请到期时间:	2019-03-31 💼		
离线SDK管理	* 申请延期理由:	请填写少于500字的申请延期理	8	
售卖已发布服务				
		提交审核		

2) 等待审核:审核周期通常需要1-3个工作日左右,期间会有工作人员电话回访,请填写有效的联系方式并保证手机畅通

^の Android集成文档

简介

1.1 Android SDK 硬件要求 Android 版本:支持 Android 5.0 (API 21) 及以上

硬件:支持 arm64-v8a 和 armeabi-v7a, 暂不支持模拟器

通常您下载的SDK只支持固定的某一类芯片。

- 通用ARM: 支持大部分ARM 架构的手机、平板及开发板。通常选择这个引擎进行推理。
- 通用ARM GPU:支持骁龙、麒麟、联发科等带GPU的手机、平板及开发板。

- 高端芯片AI加速模块:
 - 高通骁龙引擎SNPE: 高通骁龙高端SOC,利用自带的DSP加速。其中 660 之后的型号可能含有 Hexagon DSP模块,具体列表见snpe 高通骁龙引擎官网。
 - 华为NPU引擎DDK:华为麒麟980的arm-v8a的soc。具体手机机型为mate10,mate10pro,P20,mate20,荣耀v20等。
 - **华为达芬奇NPU引擎DAVINCI**: 华为NPU的后续版本,华为麒麟810,820,990,985的arm-v8a的soc。具体手机机型为华为 mate30,p40,nova6,荣耀v30等。

通用ARM有额外的加速版,但是有一定的精度损失。

因GPU硬件限制,通用ARM GPU物体检测模型输入尺寸较大时会运行失败,可以在训练的时候将输入尺寸设为300*300。

高端芯片AI加速模块,一般情况下推理速度较快。

运行内存不能过小,一般大于demo的assets目录大小的3倍。

1.2 功能支持 | 引擎 | 图像分类| 物体检测| 图像分割| 文字识别

只支持EasyEdge | 姿态估计 | | :-: | :-: | :-: | :-: | :-: | := | 通用ARM | √ | √ | √ | √ | ↓ | 通用ARM GPU | √ | √ | √ | ↓ | | 高通骁龙引擎SNPE | √ | √ |||| | 华为达芬奇NPU引擎DDK | √ | √ |||| | 华为达芬奇NPU引擎DVINCI | √ | √ | √ |||

1.3 Release Notes

时间	版本	说明
2024.12.13	0.10.14	优化不再读取硬件信息,去除硬件信息获取开关
2024.12.10	0.10.13	新增硬件信息获取开关;初始化方式优化
2023.08.31	0.10.12	新增支持实例数鉴权;SNPE引擎升级;迭代优化
2023.06.29	0.10.11	迭代优化
2023.05.17	0.10.10	横屏兼容;迭代优化
2023.03.16	0.10.9	达芬奇NPU支持更多模型及语义分割模型;各芯片支持更多语义分割模型;精简版代码补充;迭代优化
2022.12.29	0.10.8	ARM / ARM-GPU 引擎升级;迭代优化
2022.10.27	0.10.7	达芬奇NPU新增适配麒麟985;迭代优化
2022.09.15	0.10.6	SNPE引擎升级;迭代优化
2022.07.28	0.10.5	迭代优化
2022.06.30	0.10.4	支持Android11;支持EasyEdge语义分割模型;迭代优化
2022.05.18	0.10.3	ARM / ARM-GPU 引擎升级;支持更多加速版模型发布;迭代优化
2022.03.25	0.10.2	ARM / ARM-GPU 引擎升级;支持更多检测模型;迭代优化
2021.12.22	0.10.1	DDK不再支持Kirin 970;迭代优化
2021.10.20	0.10.0	更新鉴权;更新达芬奇NPU、SNPE、通用ARM及ARM-GPU引擎;新增达芬奇NPU对检测模型的支持;支持更多姿态估计 模型
2021.07.29	0.9.17	迭代优化
2021.06.29	0.9.16	迭代优化
2021.05.13	0.9.15	更新鉴权,更新通用arm及通用arm gpu引擎
2021.04.02	0.9.14	修正bug
2021.03.09	0.9.13	更新android arm的预处理加速
2020.12.18	0.9.12	通用ARM引擎升级;新增ARM GPU引擎
2020.10.29	0.9.10	迭代优化
2020.9.01	0.9.9	迭代优化
2020.8.11	0.9.8	更新ddk 达芬奇引擎
2020.7.14	0.9.7	支持arm版ocr模型,模型加载优化
2020.6.23	0.9.6	支持arm版fasterrcnn模型
2020.5.14	0.9.5	新增华为新的达芬奇架构npu的部分图像分类模型
2020.4.17	0.9.4	新增arm通用引擎量化模型支持
2020.1.17	0.9.3	新增arm通用引擎图像分割模型支持
2019.12.26	0.9.2	新增华为kirin麒麟芯片的物体检测支持
2019.12.04	0.9.1	使用paddleLite作为arm预测引擎
2019.08.30	0.9.0	支持EasyDL专业版
2019.08.30	0.8.2	支持华为麒麟980的物体检测模型
2019.08.29	0.8.1	修复相机在开发版调用奔溃的问题
2019.06.20	0.8.0	高通手机引擎优化
2019.05.24	0.7.0	升级引擎
2019.05.14	0.6.0	优化demo程序
2019.04.12	0.5.0	新增华为麒麟980支持
2019.03.29	0.4.0	引擎优化,支持sd卡模型读取
2019.02.28	0.3.0	引擎优化,性能与效果提升;
2018.11.30	0.2.0	第一版!

2.1 安装软件及硬件准备

扫描模型下载SDK处的网页上的二维码,无需任何依赖,直接体验

如果需要源码方式测试:

打开AndroidStudio,点击 "Import Project..."。在一台较新的手机上测试。

详细步骤如下:

1. 准备一台较新的手机,如果不是通用arm版本,请参见本文的"硬件要求",确认是否符合SDK的要求

2. 安装较新版本的AndroidStudio , 下载地址

3. 新建一个HelloWorld项目, Android Studio会自动下载依赖, 在这台较新的手机上测试通过这个helloworld项目。注意不支持模拟器。

4. 解压下载的SDK。

5. 打开AndroidStudio, 点击 "Import Project..."。即: File->New-> "Import Project...", 选择解压后的目录。

6. 此时点击运行按钮(同第3步),手机上会有新app安装完毕,运行效果和二维码扫描的一样。

7. 手机上UI界面显示后,如果点击UI界面上的"开始使用"按钮,可能会报序列号错误。请参见下文修改

2.2 使用序列号激活

如果使用的是EasyEdge的开源模型,无需序列号,可以跳过本段直接测试。

建议申请包名为"com.baidu.ai.easyaimobile.demo"的序列号用于测试。

本文假设已经获取到序列号,并且这个序列号已经绑定包名。

SDK默认使用离线激活方式,即首次联网激活,后续离线使用。SDK同时支持按实例数鉴权方式,即周期性联网激活,离线后会释放所占用设备 实例。按实例数鉴权的启用参考本节2.2.3说明

2.2.1 填写序列号 打开Android Studio的项目,修改MainActivity类的开头SERIAL_NUM字段。 MainActivity 位于 app \src \main \java \com \baidu \ai \edge \demo \MainActivity.java文件内。

```
// 请替换为您的序列号
private static final String SERIAL_NUM = "XXXX-XXXX-XXXX-XXXX"; //这里填您的序列号
```

2.2.2 修改包名

如果申请的包名为"com.baidu.ai.easyaimobile.demo",这个是demo的包名,可以不用修改

打开app/build.gradle文件,修改"com.baidu.ai.easyaimobile.demo"为申请的包名

```
defaultConfig {
    applicationId "com.baidu.ai.easyaimobile.demo" // 修改为比如"com.xxx.xxx"
}
```

修改序列号和包名后,可以运行测试,效果同扫描二维码的一致

2.2.3 按实例数鉴权 设置好序列号和包名后,调用配置类的以下方法启用并配置心跳间隔时间:

XXXConfig config = new XXXConfig(); // 启用按实例数鉴权,配置心跳间隔,单位:秒 config.setInstanceAuthMode(10000);

配置类的详细说明参考后续章节【调用流程】

2.3 测试精简版

对于通用ARM、高通骁龙引擎SNPE、华为NPU引擎DDK和达芬奇NPU引擎Davinci的常见功能,项目内自带精简版,可以忽略开发板不兼容的摄像 头。 此外,由于实时摄像开启,会导致接口的耗时变大,此时也可以使用精简版测试。

目前以下硬件环境有精简版测试:

- 通用ARM:图像分类(Classify),物体检测(Detection),文字识别(OCR),图像分割(Segmentation),姿态估计(Pose)
- 通用ARM GPU:图像分类(Classify),物体检测(Detection),图像分割(Segmentation),姿态估计(Pose)
- 高通骁龙引擎SNPE:图像分类 (Classify) ,物体检测 (Detection)
- 华为NPU引擎DDK:图像分类(Classify),物体检测(Detection)
- 华为达芬奇NPU引擎Davinci:图像分类(Classify),物体检测(Detection),图像分割(Segmentation)

具体代码分别在infertest、snpetest、ddktest和davincitest目录下。

修改方法为(以通用ARM为例):更改app/main/AndroidManifest.xml中的启动Activity。

```
<activity android:name=".infertest.MainActivity"> <!-- 原始的是".MainActivity" -->
<intent-filter>
<action android:name="android.intent.action.MAIN" />
<category android:name="android.intent.category.LAUNCHER" />
</intent-filter>
</activity>
```

开启后会自动选择图像分类(Classify),物体检测(Detection),文字识别(OCR),图像分割(Segmentation)或姿态估计(Pose)测试。

Demo APP 检测模型运行示例 精简版检测模型运行示例 Hello World! paddle2-Centernet_mbv3_small_140e < ARM Detection 4.18 person 0 Start running: 0 horse 0.40 Predict 0: {size:100, firstRe confidence:0.6314938, bo truck 0.34 181)}} per86/\$9/8 Finish running Task finished cow 0.34 识别结果 置信度 0.30 名称 置信度 序号 0.63 1 person 2 0.47 person 3 0.42 car 0.40 4 horse 5 dog 0.34 BU' 6 truck 0.34 4 ►

使用说明

3.1 代码目录结构

集成时需要"复制到自己的项目里"的目录或者文件:

2. app/src/main/assets/xxxx-xxxxx 如 app/src/main/assets/infer

+app 简单的设置,模拟用户的项目 |---+libs 实际使用时需要复制到自己的项目里 |----arm64-v8a v8a的so |----armeabi-v7a v7a的so |----easyedge-sdk.jar jar库文件 |---+src/main |---+assets |----demo demo项目的配置,实际集成不需要 |----infer 也可能是其它命名, infer表示通用arm。实际使用时可以复制到自己的项目里 |---+java/com.baidu.ai.edge/demo |---+infertest 通用Arm精简版测试,里面有SDK的集成逻辑 |--- MainActivity 通用Arm精简版启动Activity,会根据assets目录判断当前的模型类型,并运行同目录的一个Task。使用时需要修改里 面的序列号 |--- TestInferArmClassifyTask 通用Arm精简版分类 |--- TestInferArmDetectionTask 通用Arm精简版检测 |--- TestInferArmOcrTask 通用Arm精简版OCR |--- TestInferArmPoseTask 通用Arm精简版姿态 |--- TestInferArmSegmentTask 通用Arm精简版分割 |--- TestInferArmGpuClassifyTask 通用ArmGpu精简版分类 |--- TestInferArmGpuDetectionTask 通用ArmGpu精简版检测 |--- TestInferArmGpuPoseTask 通用ArmGpu精简版姿态 |--- TestInferArmGpuSegmentTask 通用ArmGpu精简版分割 |---+snpetest SNPE精简版测试 |--- MainActivity SNPE精简版启动Activity,会根据assets目录判断当前的模型类型,并运行同目录的一个Task。使用时需要修改里面 的序列号 |--- TestSnpeDspClassifyTask SNPE DSP精简版分类 |--- TestSnpeDspDetectionTask SNPE DSP精简版检测 |--- TestSnpeGpuClassifyTask SNPE Gpu精简版分类 |--- TestSnpeGpuDetectionTask SNPE Gpu精简版检测 |---+ddktest DDK精简版测试 |--- MainActivity DDK精简版启动Activity,会根据assets目录判断当前的模型类型,并运行同目录的一个Task。使用时需要修改里面的 序列号 |--- TestDDKClassifyTask DDK精简版分类 |--- TestDDKDetectionTask DDK精简版检测 |---+davincitest Davinci精简版测试 |--- MainActivity Davinci精简版启动Activity,会根据assets目录判断当前的模型类型,并运行同目录的一个Task。使用时需要修改里面 的序列号 |--- TestDavinciClassifyTask Davinci精简版分类 |--- TestDavinciDetectionTask Davinci精简版检测 |--- TestDavinciSegmentTask Davinci精简版分割 |----CameraActivity 摄像头扫描示例,里面有SDK的集成逻辑 |----MainActivity 启动Activity,使用时需要修改里面的序列号 |---- build.gradle 这里修改包名 +camera_ui UI模块,集成时可以忽略 • 1

3.2 调用流程 以通用ARM的检测模型功能为例,

代码可以参考TestInferDetectionTask

- 1. 准备配置类,如InferConfig,输入:通常为一个assets目录下的文件夹,如infer。
- 2. 初始化Manager,比如InferManager。输入:第1步的配置类和序列号
- 3. 推理图片,可以多次调用 5.1 准备图片,作为Bitmap输入 6.2 调用对应的推理方法,比如detect 7.3 解析结果,结果通常是一个List,调用结 果类的Get方法,通常能获取想要的结果
- 4. 直到长时间不再使用我们的SDK,调用Manger的destroy方法释放资源。

3.3 具体接口说明下文的示例部分以通用ARM的检测模型功能为例

即接口为InferConfig, InferManager, InferManager.detect。

其它引擎和模型调用方法类似。

下文假设已有序列号及对应的包名

- 3.3.1. 准备配置类
- INFER:通用ARM, InferConfig
- ARM GPU : ArmGpuConfig
- SNPE:高通骁龙DSP, SnpeConfig
- SNPE GPU:高通骁龙GPU, SnpeGpuConfig
- DDK:华为NPU, DDKConfig
- DDKDAVINCI:华为达芬奇NPU,DDKDaVinciConfig

InferConfig mInferConfig = new InferConfig(getAssets(), "infer"); // assets 目录下的infer , infer表示通用arm

输入:assets下的配置 输出:具体的配置类

3.3.2. 初始化Manager类

- INFER:通用ARM, InferManager
- ARM GPU:通用ARM GPU, InferManager
- SNPE:高通骁龙DSP, SnpeManager
- SNPE GPU:高通骁龙GPU, SnpeManager
- DDK:华为NPU, DDKManager
- DDKDAVINCI:华为达芬奇NPU, DavinciManager

String SERIAL_NUM = "XXXX-XXXX-XXXX-XXXX";

// InferManager 为例:

InferManager manager = new InferManager(this, config, SERIAL_NUM); // config为上一步的InferConfig

注意要点

- 1. 同一个时刻只能有唯一有效的InferManager。旧的InferManager必须调用destory后,才能新建一个new InferManager()。
- 2. InferManager的任何方法,都不能在UI线程中调用。
- 3. new InferManager()及InferManager成员方法由于线程同步数据可见性问题,都**必须在一个线程中执行**。如使用android自带的**ThreadHandler 类**。

输入: 1.配置类; 2.序列号 输出: Manager类

3.3.3.推理图片

- 接口可以多次调用,但是必须在一个线程里,**不能并发**
- confidence, 置信度[0-1], 小于confidence的结果不返回。 填confidence=0, 返回所有结果
- confidence可以不填,默认用模型推荐的。

准备图片,作为Bitmap输入,

• 输入为Bitmap,其中Bitmap的options为默认。如果强制指定的话,必须使用*Bitmap.Config.ARGB_8888*

调用对应的推理方法及结果解析 见下文的各个模型方法

3.3.4 分类Classify

public interface ClassifyInterface { List<ClassificationResultModel> classify(Bitmap bitmap, float confidence) throws BaseException;

// 如InferManger 继承 ClassifyInterface

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 ClassificationResultModel 异常:一般首次出现。可以打印出异常错误码。

ClassificationResultModel

- label:分类标签,定义在label_list.txt中

- confidence:置信度,0-1

- lableIndex:标签对应的序号

3.3.5 检测Detect

对于EasyDL口罩检测模型请注意输入图片中人脸大小建议保持在88到9696像素,可根据场景远近程度缩放图片后传入

public interface DetectInterface { List<DetectionResultModel> detect(Bitmap bitmap, float confidence) throws BaseException;

// 如InferManger 继承 DetectInterface

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 DetectionResultModel List 异常:一般首次出现。可以打印出异常错误码。

DetectionResultModel

- label:标签,定义在label_list.txt中
- confidence:置信度
- bounds: Rect, 左上角和右下角坐标

3.3.6 图像分割Segmentation

```
public interface SegmentInterface {
List<SegmentationResultModel> segment(Bitmap bitmap, float confidence) throws BaseException;
// 如InferManger 继承 SegmentInterface
```

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 SegmentationResultModel 异常:一般首次出现。可以打印出异常错误码。

SegmentationResultModel

label:标签,定义在label_list.txt中
confidence:置信度
lableIndex:标签对应的序号
box: Rect对象表示的对象框
mask:byte[]表示的原图大小的0,1掩码,绘制1的像素即可得到当前对象区域
mask 字段说明,如何绘制掩码也可参考demo工程
 101
image 110 => mask(byte[]) 101 110 011

011

3.3.7 文字识别OCR

暂时只支持通用ARM引擎,不支持其它引擎,暂时只支持EasyEdge的开源OCR模型。

```
public interface OcrInterface {
List<OcrResultModel> ocr(Bitmap bitmap, float confidence) throws BaseException;
// 如InferManger 继承 OcrInterface
```

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 OcrResultModel List,每个OcrResultModel对应结果里的一个四边形。 异常:一般首次出现。可以打印出异常错误码。

OcrResultModel

- label:识别出的文字
- confidence:置信度
- List<Point>:4个点构成四边形

3.3.8 姿态估计Pose

暂时只支持通用ARM引擎,不支持其它引擎

public interface PoseInterface { List<PoseResultModel> pose(Bitmap bitmap) throws BaseException; // 如InferManger 继承 PoseInterface

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 PoseResultModel List 异常:一般首次出现。可以打印出异常错误码。

PoseResultModel

- label : 标签,定义在label_list.txt中

- confidence:置信度

- Pair<Point, Point>: 2个点构成一条线

3.3.9 释放

释放后这个对象不能再使用,如果需要使用可以重新new一个出来。

public void destory() throws BaseException

3.3.10 整体示例

以通用ARM的图像分类预测流程为例:

```
try {
// step 1: 准备配置类
InferConfig config = new InferConfig(context.getAssets(), "infer");
// step 2: 初始化预测 Manager
InferManager manager = new InferManager(context, config, "");
// step 3: 准备待预测的图像,必须为 Bitmap.Config.ARGB_8888 格式,一般为默认格式
Bitmap image = getFromSomeWhere();
// step 4: 预测图像
List<ClassificationResultModel> results = manager.classify(image, 0.3f);
// step 5: 解析结果
for (ClassificationResultModel resultModel : results) {
 Log.i(TAG, "labelIndex=" + resultModel.getLabelIndex()
  + ", labelName=" + resultModel.getLabel()
  + ", confidence=" + resultModel.getConfidence());
}
// step 6: 释放资源。预测完毕请及时释放资源
manager.destroy();
} catch (Exception e) {
Log.e(TAG, e.getMessage());
}
```

```
"autocheck_qcom": true, // 如果改成false, sdk跳过检查手机是否是高通的Soc,非高通的Soc会奔溃直接导致app闪退
"snpe_runtimes_order": [],
// 不填写为自动,按照 {DSP, GPU, GPU_FLOAT16, CPU}次序尝试初始化,也可以手动指定如[2,1,3,0],具体数字的定义见下段

public interface SnpeRuntimeInterface {
    int CPU = 0;
    int GPU = 1;
    int DSP = 2;
    int GPU_FLOAT16 = 3;
```

// SnpeManager 中,使用public static ArrayList<Integer> getAvailableRuntimes(Context context) 方法可以获取高通SOC支持的运行方式

集成指南

}

- 1. 复制库文件libs
- 2. 添加Manifest权限
- 3. 复制模型文件
- 4. 添加调用代码(见上一步具体接口说明)
- 4.1 复制库文件libs A. 如果项目里没有自己的jar文件和so文件:

```
复制app/libs 至自己项目的app/libs目录。
参照demo的app/build.gradle 中添加
android {
....
defaultConfig {
ndk {
abiFilters 'armeabi-v7a', 'arm64-v8a'
}
}
sourceSets {
main {
jniLibs.srcDirs = ['libs']
}
}
```

B. 如果项目里有自己的jar文件,但没有so文件

```
easyedge.jar文件同自己的jar文件放一起
arm64-v8a和armeabi-v7a放到app/src/main/jniLibs目录下
参照demo的app/build.gradle 中添加
android {
....
defaultConfig {
ndk {
abiFilters 'armeabi-v7a', 'arm64-v8a'
}
}
```

C. 如果项目里有自己的jar文件和so文件

```
easyedge.jar文件同自己的jar文件放一起
arm64-v8a和armeabi-v7a取交集和自己的so放一起,交集的意思是比如自己的项目里有x86目录,必须删除x86。
参照demo的app/build.gradle 中添加
android {
....
defaultConfig {
ndk {
abiFilters 'armeabi-v7a', 'arm64-v8a' // abiFilter取交集,即只能少不能多
}
}
```

jar文件库如果没有设置成功的,编译的时候可以发现报错。

so库如果没有编译进去的话,也可以通过解压apk文件确认。运行的时候会有类似jni方法找不到的报错。

4.2 Manifest配置

参考app/src/main/AndroidManifest.xml文件,添加:

```
    <uses-permission android:name="android.permission.INTERNET" />
    <uses-permission android:name="android.permission.WRITE_EXTERNAL_STORAGE" />
    <uses-permission android:name="android.permission.ACCESS_NETWORK_STATE" />
    <uses-permission android:name="android.permission.READ_PHONE_STATE" />
    <l-- Android 11 支持 -->
    <uses-permission</li>
    android:name="android.permission.MANAGE_EXTERNAL_STORAGE"
    <l-- 高版本 Android 支持 -->
    <application</a>
```

```
android:requestLegacyExternalStorage="true"
android:usesCleartextTraffic="true">
</application>
```

4.3 混淆规则 (可选) 请不要混淆SDK里的jar文件。

```
-keep class com.baidu.ai.edge.core.*.*{ *; }
```

4.4 Android 11支持 除Manifest中必要配置外,请参考BaseActivity获取所有文件访问权限,否则可能影响SDK正常使用。

SDK 默认使用 easyedge-sdk.jar,未启用 AndroidX,若您的项目使用 AndroidX,并在集成中提示 android.support 相关错误,请参考 app/build.gradle 使用 etc/easyedge-sdk-androidx.jar 以支持 AndroidX:

```
// app/build.gradle
dependencies {
    implementation project(':camera_ui')
    implementation files('libs/easyedge-sdk-androidx.jar') // 修改 jar 包依赖
}
```

错误码 | 错误码 | 错误描述 | 详细描述及解决方法 | | --- | --- | 1001 | assets 目录下用户指定的配置文件不存在或不正确 | SDK使用assets目录下一系列文件作为配置文件。如果文件缺失或内容不正确,则有此报错 || 1002 | json格式的配置文件解析出错 | 如缺少某些字段。正常情况下,配置文件请不要修改 || 1003 | 应用缺少权限 | 请根据提示动态申请缺少的权限 || 19xx | Sdk内部错误 | 请与百度人员联系 || 2001 | XxxxMANAGER 只允许一个实例 | 如已有XxxxMANAGER对象,请调用destory方法 || 2002 | XxxxMANAGER 已经调用过destory方法 | 在一个已经调用destory方法的DETECT_MANAGER对象上,不允许再调用任何方法 || 2003 | 传入的assets下模型文件路径非法 |比如缺少模型文件,XxxxConfig.getModelFileAssetPath() 返回为null || 2012 | JNI内存错误 | heap的内存不够 || 2103 | license过期 | license失效或者系统时间有异常 || 2601/2602 | assets 目录下模型文件打开/读取失败 | 请根据报错信息检查模型文件是否存在 || 27xx | Sdk内部错误 | 请与百度人员联系 || 3003 | 过2xxx | 引擎内部错误 | 请与百度人员联系 || 29xx | Sdk内部错误 | 请与百度人员联系 || 3000 | so加载错误 | 请确认所有so文件存在于apk中 || 3001 | 模型加载错误 | 请确认模型放置于能被加载到的合法路径中,并确保配置文件正确 || 3002 | 模型卸载错误 | 请与百度人员联系 || 4011 | SDK类型 与设备硬件不匹配 | 比如适配DSP的SDK运行在麒麟芯片上会出现此报错,请在部署包支持的硬件上使用SDK || 50xx | SDK调用异常 | 请与百度 人员联系 |

报错日志收集 通常 Logcat 可以看见日志及崩溃信息,若设备无法获取日志信息,可使用 Demo 中的 xCrash 工具:

```
// 1. 引入 app/build.gradle 的 xCrash 依赖
android {
  ...
  dependencies {
     implementation 'com.iqiyi.xcrash:xcrash-android-lib:2.4.5' // 可以保存崩溃信息,默认未引入
  }
}
// 2. 启用日志收集。日志将保存在 /sdcard/<包名>/xCrash
// app/src/main/java/com.baidu.ai.edge/demo/MyApplication.java
protected void attachBaseContext(Context context) {
  // 日志保存位置
  String basePath = Environment.getExternalStorageDirectory().toString() + "/" + context.getPackageName();
  // 启用
  XCrash.InitParameters params = new XCrash.InitParameters();
  params.setAppVersion(BaseManager.VERSION);
  params.setLogDir(basePath + "/xCrash");
  XCrash.init(this, params);
}
```

心 iOS集成文档

简介

本文档描述 EasyEdge/EasyDL iOS 离线预测SDK相关功能;

目前支持EasyEdge的功能包括:

- 图像分类
- 物体检测
- 人脸检测
- 姿态估计
- 百度OCR模型

目前支持EasyDL的功能包括:

- 图像分类
- 物体检测
- 图像分割

系统支持

系统:

- 通用arm版本:iOS 9.0 以上
- A仿生芯片版: iOS 15.0 及以上
- 硬件: arm64 (Starndard architectures) (暂不支持模拟器)

内存:图像分割模型需要手机内存3GB以上,并尽量减少其他程序内存占用

离线SDK包说明

根据用户的选择,下载的离线SDK,可能包括以下类型:

- EasyEdge
 - 通用ARM版:支持iPhone5s, iOS 9.0 以上所有手机。
 - A仿生芯片版:支持iPhone5s, iOS 15.0 以上手机。充分利用苹果A系列仿生芯片优势,在iPhone 8以上机型中能有显著的速度提升。

- EasyDL 通用版/全功能AI开发平台BML (原EasyDL专业版)
 - 通用ARM版:支持iPhone5s, iOS 9.0 以上所有手机。
 - A仿生芯片版:支持iPhone5s, iOS 15.0 以上手机。充分利用苹果A系列仿生芯片优势,在iPhone 8以上机型中能有显著的速度提升。
 - 自适应芯片版:同时整合了以上两种版本,自动在iOS 15以下中使用通用ARM版,在iOS 15以上系统中使用A仿生芯片版,自适应系统,但SDK体积相对较大。
- AI市场试用版SDK

SDK大小说明

SDK库的二进制与_TEXT增量约3M。

资源文件大小根据模型不同可能有所差异。

物体检测(高性能)的DemoApp在iPhone 6, iOS 11.4下占用空间实测小于40M。

虽然SDK库文件很大(体现为SDK包文件很大,ipa文件很大),但最终应用在用户设备中所占用的大小会缩小很多。 这与multi architechtures、bitcode和AppStore的优化有关。

获取序列号 生成SDK后,点击获取序列号进入控制台获取。EasyEdge控制台、EasyDL控制台、BML控制台。

试用版SDK在SDK的RES文件夹中的SN.txt中包含试用序列号。

更换序列号、更换设备时,首次使用需要联网激活。激活成功之后,有效期内可离线使用。

Release Notes

时间	版本	说明
2023.08.31	0.7.13	新增按实例数鉴权;迭代优化
2023.06.29	0.7.12	迭代优化
2023.05.17	0.7.11	CoreML引擎升级,支持更多语义分割模型;兼容横屏;迭代优化
2023.03.16	0.7.10	支持更多语义分割模型;迭代优化
2022.12.29	0.7.9	ARM引擎升级;迭代优化
2022.10.27	0.7.8	支持更多检测模型;迭代优化
2022.09.15	0.7.7	支持更多检测模型;迭代优化
2022.07.28	0.7.6	迭代优化
2022.06.29	0.7.5	支持EasyEdge语义分割模型;CoreML引擎升级,新增EasyEdge检测模型支持;迭代优化
2022.05.18	0.7.4	ARM引擎升级;支持EasyDL物体检测超高精度模型;支持更多加速版模型发布;迭代优化
2022.03.25	0.7.3	ARM引擎升级;支持更多检测模型
2021.12.22	0.7.2	支持EasyEdge更多姿态估计模型;迭代优化
2021.10.20	0.7.1	ARM引擎升级
2021.07.29	0.7.0	迭代优化
2021.04.06	0.6.1	ARM引擎升级
2021.03.09	0.6.0	支持EasyEdge人脸检测及姿态估计模型
2020.12.18	0.5.7	ARM引擎升级
2020.09.17	0.5.6	CoreML引擎升级,支持AI市场试用版SDK
2020.08.11	0.5.5	CoreML支持EasyDL专业版模型,支持EasyEdge OCR模型
2020.06.23	0.5.4	ARM引擎升级
2020.04.16	0.5.3	ARM引擎升级;支持压缩加速版模型
2020.03.13	0.5.2	ARM引擎升级;支持图像分割模型
2020.01.16	0.5.1	ARM引擎升级;增加推荐阈值支持
2019.12.04	0.5.0	ARM引擎升级;增加coreml3的支持
2019.10.24	0.4.5	支持EasyDL专业版;ARM引擎升级
2019.08.30	0.4.4	支持EasyDL经典版图像分类高性能、高精度
2019.06.20	0.4.3	引擎优化
2019.04.12	0.4.1	支持EasyDL经典版物体检测高精度、高性能模型
2019.03.29	0.4.0	引擎优化,支持CoreML;
2019.02.28	0.3.0	引擎优化,性能与效果提升;
2018.11.30	0.2.0	第一版!

快速开始 文件结构说明

.EasyEdge-iOS-SDK ├── EasyDLDemo # Demo工程文件 └── LIB # 依赖库 ├── RES │ └── easyedge # 模型资源文件夹 │ │ └── model │ │ └── params │ │ └── label_list.txt │ │ └── conf.json └── DOC # 文档

测试Demo

按如下步骤可直接运行 SDK 体验 Demo:

步骤一:用 Xcode 打开 EasyDLDemo/EasyDLDemo.xcodeproj 步骤二:配置开发者自己的签名

步骤三:连接手机运行,不支持模拟器

检测模型运行示例:







SDK使用说明 集成指南 步骤一:依赖库集成 步骤二:import <EasyDL/EasyDL.h>, import <Vision/Vision.h>

依赖库集成

- 1. 复制 LIB 目录至项目合适的位置
- 2. 配置 Build Settings 中 Search paths: 以 SDK 中 LIB 目录路径为例
- Framework Search Paths : \${PROJECT_DIR}/../LIB/lib
- Header Search Paths : \${PROJECT_DIR}/../LIB/include
- Library Search Paths : \${PROJECT_DIR}/../LIB/lib

集成过程如出现错误,请参考 Demo 工程对依赖库的引用

使用流程

1. 生成模型,下载SDK 开发者在官网下载的SDK已经自动为开发者配置了模型文件和相关配置,开发者直接运行即可。

2. 使用序列号激活 2.1. 离线激活 (默认鉴权方式) 首次联网激活,后续离线使用

将前面申请的序列号填入:

[EasyDL setSerialNumber:@"!!!Enter Your Serial Number Here!!!"];

根据序列号类型,序列号与BundleID绑定或与BundleID+设备绑定。 请确保设备时间正确。

2.2. 按实例数激活 周期性联网激活,离线后会释放所占用鉴权,启动时请确保心跳间隔小于等于生成序列号时填写的定期确认时间

填入序列号,配置按实例数鉴权并设置心跳间隔:

```
// 设置序列号
```

[EasyDL setSerialNumber:@"!!!Enter Your Serial Number Here!!!"]; // 配置实例数鉴权及心跳间隔,单位:秒 [EasyDL setInstanceAuthMode:10000];

3. 初始化模型

EasyDLModel *_model = [[EasyDLModel alloc] initModelFromResourceDirectory:@"easyedge" withError:&err];

请注意相关资源必须以folder reference方式加入Xcode工程。也即默认的easyedge文件夹在Xcode文件列表里显示为蓝色。

4. 调用检测接口

```
UlImage *img = .....;
NSArray *result = [model detectUlImage:img withFilterScore:0 andError:&err];
```

/**

```
* 检测图像
```

- * @param image 带检测图像
- * @param score 只返回得分高于score的结果(0~1)
- * @return 成功返回识别结果,NSArray的元素为对应模型的结果类型;失败返回nil,并在err中说明错误原因
- */
- (NSArray *)detectUIImage:(UIImage *)image
 - withFilterScore:(CGFloat)score
 - andError:(NSError **)err;

返回的数组类型如下,具体可参考 EasyDLResultData.h 中的定义:

模型类型	类型
图像-图像分类	EasyDLCIassfiData
图像-物体检测/人脸检测	EasyDLObjectDetectionData
图像-实例分割/语义分割	EasyDLObjSegmentationData
图像-姿态估计	EasyDLPoseData
图像-文字识别	EasyDLOcrData

错误说明

SDK的方法会返回NSError错,直接返回的NSError的错误码定义在EEasyDLErrorCode中。NSError附带message(有时候会附带 NSUnderlyingError),开发者可根据code和message进行错误判断和处理。

FAQ

1. 如何多线程并发预测?

SDK内部已经能充分利用多核的计算能力。不建议使用并发来预测。

如果开发者想并发使用,请务必注意EasyDLModel所有的方法都不是线程安全的。请初始化多个实例进行并发使用,如

```
- (void)testMultiThread {
 Ullmage *img = [Ullmage imageNamed:@"1.jpeg"];
 NSError *err;
 EasyDLModel * model1 = [[EasyDLModel alloc] initModelFromResourceDirectory:@"easyedge" withError:&err];
 EasyDLModel * model2 = [[EasyDLModel alloc] initModelFromResourceDirectory:@"easyedge" withError:&err];
 dispatch_queue_t queue1 = dispatch_queue_create("testQueue", DISPATCH_QUEUE_CONCURRENT);
 dispatch_queue_t queue2 = dispatch_queue_create("testQueue2", DISPATCH_QUEUE_CONCURRENT);
 dispatch_async(queue1, ^{
   NSError *detectErr;
    for(int \; i = 0; \; i < 1000; \; ++i) \; \{
      NSArray * res = [model1 detectUIImage:img withFilterScore:0 andError:&detectErr];
      NSLog(@"1: %@", res[0]);
 dispatch_async(queue2, ^{
   NSError *detectErr;
    for(int i = 0; i < 1000; ++i) {
      NSArray * res = [model2 detectUlImage:img withFilterScore:0 andError:&detectErr];
      NSLog(@"2: %@", res[0]);
```

2. 编译时出现 Undefined symbols for architecture arm64: ...

- 出现 cxx11, vtable 字样 : 请引入 libc++.tbd
- 出现 cv::Mat 字样:请引入 opencv2.framework
- 出现 CoreML, VNRequest 字样:请引入CoreML.framework 并务必#import <CoreML/CoreML.h>
- 3. 运行时报错 Image not found: xxx ...

请Embed具体报错的库。 4.编译时报错:Invalid bitcode version 这个可能是开发者使用的xcode低于12导致,可以升级至12版本。

の Windows集成文档

简介

本文档介绍图像分类通用小型设备Windows SDK的使用方法。

- 硬件支持:
 - Intel CPU 普通版 * x86_64

- CPU 加速版 Intel Xeon with AVX2 and AVX512 Intel Core Processors with AVX2 Intel Atom Processors with SSE * AMD Core Processors with AVX2
- Intel Movidius MyRIAD2/MyRIAD X (仅支持Win10)
- 操作系统支持
 - 普通版:64位 Windows 7 及以上,64位Windows Server2012及以上
 - 加速版:64位 Windows 10,64位Windows Server 2019及以上
- 环境依赖 (必须安装以下版本)
 - .NET Framework 4.5
 - Visual C++ Redistributable Packages for Visual Studio 2013
 - Visual C++ Redistributable Packages for Visual Studio 2015-2019
- 协议
 - HTTP
- 更详细的环境说明可参考SDK内的README.md

*intel 官方合作,拥有更好的适配与性能表现

Release Notes | 时间 | 版本 | 说明 | | ------ | ----- | 2023-08-30 | 1.8.3 | 新增支持按实例数鉴权 || 2023-06-29 | 1.8.2 | 优化模型 算法 || 2023-05-17 | 1.8.1 | 预测引擎升级, 修复部分网络内存泄露问题 || 2023-03-16 | 1.8.0 | 预测引擎升级 || 2022-12-29 | 1.7.2 | 预测引 擎升级 || 2022-10-27 | 1.7.1 | 优化模型算法 || 2022-09-15 | 1.7.0 | 新增支持表格预测 || 2022-07-28 | 1.6.0 | 优化模型算法 || 2022-05-27 | 1.5.1 | 新增支持BML Cloud小目标检测模型 || 2022-05-18 | 1.5.0 | 修复各别机器下程序崩溃的问题 || 2022-04-25 | 1.4.1 | EasyDL, BML升级 支持paddle2模型 || 2022-03-25 | 1.4.0 | 优化模型算法 || 2021-12-22 | 1.3.5 | CPU基础版推理引擎优化升级 ; demo程序优化环境依赖检测 || 2021-10-20 | 1.3.4 | CPU加速版推理引擎优化升级 || 2021-08-19 | 1.3.2 | 新增DEMO二进制文件 || 2021-06-29 | 1.3.1 | 预测引擎升级 || 2021-05-13 | 1.3.0 | 模型发布新增多种加速方案选择 ; 目标追踪支持x86平台的GPU及加速版 ; 展示已发布模型性能评估报告 || 2021-04-08 | 1.2.3 | 支持BML平台模型仓库本地上传模型 || 2021-03-09 | 1.2.2 | CPU加速版支持int8量化模型 || 2021-01-27 | 1.2.1 | 新增模型支持 ; 性能优 化 ; 问题修复 || 2020.12.18 | 1.2.0 | 推理引擎升级 || 2020-01-26 | 1.1.20 | 新增一些模型的加速版支持 || 2020.10.29 | 1.1.20 | 修复已知问 题 || 2020.05.15 | 1.1.16 | 优化性能 , 修复已知问题 || 2020-09-17 | 1.1.19 | 支持更多模型 || 2020.04.16 | 1.1.15 | 升级引擎版本 || 2020.06.23 | 1.1.17 | 支持专业版更多模型 || 2020.05.15 | 1.1.16 | 优化性能 , 修复已知问题 || 2020.04.16 | 1.1.15 | 升级引擎版本 || 2020.03.13 | 1.1.14 | 支持EdgeBoardVMX || 2020.02.23 | 1.1.13 | 支持多阶段模型 || 2020.01.16 | 1.1.12 | 预测默认使用推荐阈值 || 2019.02.26 | 1.1.11 | CPU加速版支持物体检测高精度 || 2019.12.04 | 1.1.10 | 支持图像分割 || 2019.02.11 | 1.1.9 | 支持 EasyDL 专业版 || 2019.08.29 | 1.1.8 | CPU 加速版支持 || 2019.07.19 | 1.1.7 | 提供模型更新工具 || 2019.05.16 | 1.1.3 | NVIDIA GPU 支持 || 2019.03.15 | 1.1.0 | 架构与功能完善 || 2019.02.28 | 1.0.6 | 引擎功能完善 || 2019.02.13 | 1.0.5 | paddlepaddle 支持 || 2018.11.30 | 1.0.0 | 第一版 !

快速开始

1. 安装依赖

必须安装:

安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642

Visual C++ Redistributable Packages for Visual Studio 2013

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

Visual C++ Redistributable Packages for Visual Studio 2015-2019

https://docs.microsoft.com/en-us/cpp/windows/latest-supported-vc-redist?view=msvc-160

Openvino (仅使用Python版Intel Movidius必须)

- 使用 OpenVINO[™] toolkit 安装,请参考 OpenVINO toolkit 文档安装 2020.3.1LTS(必须)版本, 安装时可忽略Configure the Model Optimizer及 后续部分。
- 使用源码编译安装,请参考 Openvino Inference Engine文档编译安装 2020.3.1LTS (必须)版本。

安装完成后,请设置环境变量OPENVINO_HOME为您设置的安装地址,默认是C: \Program Files (x86) \IntelSWTools,并确保文件夹下的openvino 的快捷方式指到了2020.3.1LTS版本。

充属性		
·算机名 硬件 高级 系统保护 远程	xiaoyuchen 的用户受量(U)	
	变量	值
要进行大多数更成,你必须作为管理员登录。		
性能	Path	C:\Users
视觉效果,处理器计划,内存使用,以及虚拟内存	TEMP	C:\Users\ AppData\Local\Temp
设置(5)	C:\Users\www.readypData\Local\Temp
- 用户配置文件		
与登录帐户相关的桌面设置		新建(N) 编辑(E) 删除(D)
设置(E) 系统变量(S)	
	变量	值 ^
	ComSpec	C:\Windows\system32\cmd.exe
系统启动、系统政障和确试信息	INTEL DEV REDIST	C:\Windows\System32\Drivers\DriverData C:\Program Files (x86)\Common Files\Intel\Shared Libraries\
25萬(1)	MIC_LD_LIBRARY_PATH	%INTEL_DEV_REDIST%compiler\lib\intel64_win_mic
汉重(_	NUMBER_OF_PROCESSO	DRS 8
	OS	Windows_NT
		新建(W) 編輯(I) 删除(L)
備定取消	处用(A)	施 走 取消
IntelSWTools		
共享 查看		
· · · · · · · · · · · · · · · · · · ·	ntelSWTools >	∨ ひ 2 浅
□ 名称	^	修改日期
openvino		2020/4/15 23:41
openvino_2019.2.275		2020/3/10 19:34
o ★or	penvino_2020.2.117	2020/4/15 23:42

注意事项

- 1. 安装目录不能包含中文
- 2. Windows Server 请自行开启,选择"我的电脑"——"属性"——"管理"——"添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启即 可。

2. 运行离线SDK

解压下载好的SDK,打开EasyEdge.exe,输入Serial Num,选择鉴权模式,点击"启动服务",等待数秒即可启动成功,本地服务默认运行在

http://127.0.0.1:24401/

其他任何语言只需通过HTTP调用即可。

如启动失败,可参考如下步骤排查:

○ 百度智能云	故障排查
零门槛AI开发 平台EasyDL	
Windows离线SDK(公有云发 布)•问题排查	2

2.1 离线鉴权 (默认鉴权模式) 首次联网激活,后续离线使用

Al ver. 1.2.0.0 core. 1	I.8.1 sec. 1.5.1 —		×
Model	[4] A.W. Sale and M. Sale and S. Sale		
Serial Num	la se		
Host	127.0.0.1		
Port	24401		
鉴权模式	● 离线鉴权 ○ 按实例数鉴权		
服务状态	服务未启动		
	启动服务		
	是否开机器	自动启动	

2.2 按实例数鉴权 周期性联网激活,离线后会释放所占用鉴权,启动时请确保心跳间隔小于等于生成序列号时填写的定期确认时间

AI ver. 1.2.0.0 core. 1	1.8.1 sec. 1.5.1 –	×
Model	[4] A.P. Solidari, A. Salitari, J. 1997.	
Serial Num		
Host	127.0.0.1	
Port	24401	
鉴权模式	○ 离线鉴权 ④ 按实例数鉴权 心跳间隔 20 秒>	
服务状态	服务未启动	
	启动服务	
	是否开机自己	加启动 🗌

基于源码集成时,若需要按实例数鉴权,需要通过代码指定使用按实例数鉴权

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2); global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 20);

或通过环境变量指定

set EDGE_CONTROLLER_KEY_AUTH_MODE=2 set EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=20

2.3 序列号激活错误码
错误码	文案	描述
4001	parameters missing. 参数缺失	
4002	parameters invalid. 参数不合法	
4003	model invalid. 模型信息不合法	
4004	no more activation left. 该序列号和该设备的激活次数超上限	
4005	the serial key is out of date. 该序列号过期	
4006	the serial key has been activated. 该序列号已被其他设备激活	该序列号已被其他设备激活,不能重复激活。
4007	account invalid. 序列号不能用于其他账号的模型	序列号不能用于其他账号的模型,只能用于绑定账号的模型。
4008	serial key invalid. 序列号不合法	序列号不存在或找不到
4009	bundle id invalid. 包名不合法	
4010	product invalid. 产品不合法	如easydl的SDK使用BML的序列号来激活,会报该错误
4011	platform invalid. 平台不合法	
4012	activate too frequent. 激活太频繁	激活太频繁,请稍后再进行激活。
4013	device type and license type not match. 硬件类型和序列号类型不匹配	如使用加速版序列号激活基础版SDK会报该错误
4014	exceed max activate device num. 超过最大授权数量	
4015	technology invalid. 技术类型不合法	
4016	exceed max activate entity num. 超过最大模型数量	
4017	device invalid. 设备不合法	
4018	model invalid. 模型不合法	

3. Demo示例(以图像服务为例)

服务运行成功,此时可直接在浏览器中输入http://127.0.0.1:24401,在h5中测试模型效果。



使用说明

调用说明

Python 使用示例代码如下

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
```

params 为GET参数 data 为POST Body

 $result = requests.post('http://127.0.0.1:24401/', \ params = \{'threshold': 0.1\},$

data=img).json()

Baidu 百度智能云文档

C# 使用示例代码如下

FileStream fs = new FileStream("./img.jpg", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close(); WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close();

response.Close();

C++ 使用示例代码如下,需要安装curl

```
**include <sys/stat.h>**
**include <curl/curl.h>**
**include <iostream>**
**include <string>**
**define S_ISREG(m) (((m) & 0170000) == (0100000))**
**define S_ISDIR(m) (((m) & 0170000) == (0040000))**
size_t write_callback(void *ptr, size_t size, size_t num, void *data) {
  std::string *str = dynamic_cast<std::string *>((std::string *)data);
  str->append((char *)ptr, size*num);
  return size*num;
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./img.jpg";
  FILE *fp = NULL;
  std::string response;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\" \n", post_data_filename);
     return -1;
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
    curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
    curl_easy_setopt(curl, CURLOPT_POST, 1L);
    curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
    curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
 curl_easy_setopt(curl, CURLOPT_WRITEFUNCTION, write_callback);
    curl_easy_setopt(curl, CURLOPT_WRITEDATA, &response);
    res = curl_easy_perform(curl);
     if (res != CURLE OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     std::cout << response << std::endl; // response即为返回的json数据
     curl_easy_cleanup(curl);
  curl_global_cleanup();
  fclose(fp);
  return 0;
```

结果 获取的结果存储在response字符串中。请求参数

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送图片二进制。

返回参数 | 字段 | 类型 | 取值 | 说明 | | ------- | ----- | ---- | ----- | | confidence | float | 0~1 | 分类的置信度 | | label | string | | 分类的类 别 | | index | number | | 分类的类别 |

集成指南

基于HTTP集成

通过EasyEdge.exe启动服务后,参照上面的调用说明,通过HTTP请求集成到自己的服务中

Baidu 百度智能云文档

集成前提

解压开的SDK包中包含src、lib、dll、include四个目录才支持基于c++ dll集成

集成方法

参考src目录中的CMakeLists.txt进行集成

基于c# dll集成

集成前提

解压开的SDK包中包含src\demo_serving_csharp、dll两个目录才支持基于c# dll集成

集成方法

参考src\demo_serving_csharp目录中的CMakeLists.txt进行集成

FAQ

1. 服务启动失败,怎么处理?

根据SDK内的README.md检查依赖是否都已正确安装

请确保相关依赖都安装正确,版本必须如下: *.NET Framework 4.5* Visual C++ Redistributable Packages for Visual Studio 2013 * Visual C++ Redistributable Packages for Visual Studio 2015

如使用的是CPU加速版,需额外确保Openvino安装正确,版本为2020.3.1LTS版 如使用Windows Server,需确保开启桌面体验

2. 服务调用时返回为空,怎么处理?调用输入的图片必须是RGB格式,请确认是否有alpha通道。

3. 多个模型怎么同时使用? SDK设置运行不同的端口,点击运行即可。

4. JAVA、C#等其他语言怎么调用SDK?

参考 https://ai.baidu.com/forum/topic/show/943765

5. 启动失败,缺失DLL? 打开EasyEdge.log,查看日志错误,根据提示处理 缺失DLL,请使用 https://www.dependencywalker.com/ 查看相应模 块依赖DLL缺失哪些,请自行下载安装

6. 启动失败,报错NotDecrypted? Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

7. 启动失败,报错 SerialNum无效

日志显示failed to get/check device id(xxx)或者 Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 C:\Users\\${用户名}\.baidu\easyedge 目录,再重新激活。

8. 勾选"开机自动启动"后,程序闪退

一般是写注册表失败。

可以确认下HKEY_CURRENT_USER下Software \Microsoft \Windows \CurrentVersion \Run能否写入(如果不能写入,可能被杀毒软件等工具管制)。也可以尝试基于bin目录下的easyedge_serving.exe命令行形式的二进制,自行配置开机自启动。

其他问题 如果无法解决,可到论坛发帖: https://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

_のLinux集成文档-C++

简介

本文档介绍EasyEdge/EasyDL的Linux CPP SDK的使用方法。

- 网络类型支持: 图像分类 物体检测 图像分割
- 硬件支持:
 - CPU: aarch64 armv7hf
 - GPU: ARM Mali G系列
 - ASIC: Hisilicon NNIE1.1 on aarch64 (Hi3559AV100/Hi3559CV100等)
 - ASIC: Hisilicon NNIE1.2 on armv7I (Hi3519AV100/Hi3559V200等)
 - Intel Movidius MyRIAD2 / MyRIAD X on x86_64
 - Intel Movidius MyRIAD2 / MyRIAD X on armv7I
 - Intel Movidius MyRIAD2 / MyRIAD X on aarch64
 - Intel iGPU on x86_64
 - 比特大陆 Bitmain SE5 (BM1684)
 - 瑞芯微 RK3399Pro / RV1109 / RV1126 / RK3568 / RK3588
 - 华为 Atlas200
 - 晶晨 A311D
 - 寒武纪 MLU220 on aarch64
 - 英特尔 iGPU
- 操作系统支持:
 - Linux (Ubuntu, Centos, Debian等)
 - 海思HiLinux
 - 树莓派Raspbian/Debian
 - 瑞芯微Firefly

性能数据参考算法性能及适配硬件

Release Notes | 时间 | 版本 | 说明 | | --- | ---- | | 2023.08.31 | 1.8.3 | Atlas系列Socs支持语义分割模型, Atlas Cann版本升级至6.0.1 | | 2023.06.29 | 1.8.2 | 比特大陆版本升级至V23.03.01 | | 2023.05.17 | 1.8.1 | 新增支持intel iGPU + CPU异构模式 | | 2023.03.16 | 1.8.0 | 新增 支持瑞芯微RK3588 || 2022.10.27 | 1.7.1 | 新增语义分割模型http请求示例 || 2022.09.15 | 1.7.0 | 新增瑞芯微 RK3568 支持, RK3399Pro、 RV1126升级到RKNN1.7.1 || 2022.07.28 | 1.6.0 | 引擎升级;新增英特尔 iGPU 支持 || 2022.04.25 | 1.4.1 | EasyDL, BML升级支持paddle2模 型 | | 2022.03.25 | 1.4.0 | EasyDL新增上线支持晶晨A311D NPU预测引擎; Arm CPU、Arm GPU引擎升级; atlas 200在EasyDL模型增加多个量 化加速版本; || 2021.12.22 | 1.3.5 | RK3399Pro, RV1109/RV1126 SDK扩展模型压缩加速能力,更新端上推理库版本;边缘控制台IEC功能升 级,适配更多通用小型设备,NNIE 在EasyDL增加量化加速版本;Atlas200升级到Cann5.0.3 | | 2021.06.29 | 1.3.1 | 视频流解析支持调整分辨 率;预测引擎升级;设备端sdk新增支持瑞芯微RV1109、RV1126 || 2021.05.13 | 1.3.0 | 新增视频流接入支持;EasyDL模型发布新增多种加速 方案选择;目标追踪支持x86平台的CPU、GPU加速版;展示已发布模型性能评估报告 | | 2021.03.09 | 1.2.0 | http server服务支持图片通过 base64格式调用 || 2021.01.27 | 1.1.0 | EasyDL经典版分类高性能模型升级;部分SDK不再需要单独安装OpenCV;新增RKNPU预测引擎支持; 新增高通骁龙GPU预测引擎支持 || 2020.12.18 | 1.0.0 | 1.0版本发布!安全加固升级、性能优化、引擎升级、接口优化等多项更新 || 2020.10.29 | 0.5.7 | 优化多线程预测细节 | | 2020.09.17 | 0.5.6 | 支持linux aarch64架构的硬件接入intel神经计算棒预测;支持比特大陆计算盒 SE50 BM1684 || 2020.08.11 | 0.5.5 | 支持linux armv7hf架构硬件(如树莓派)接入intel神经计算棒预测 || 2020.06.23 | 0.5.4 | arm引擎升级 || 2020.05.15 | 0.5.3 | 支持EasyDL 专业版新增模型 ; 支持树莓派(armv7hf, aarch64)| | 2020.04.16 | 0.5.2 | Jetson系列SDK支持多线程infer | | 2020.02.23 | 0.5.0 | 新增支持人脸口罩模型; Jetson SDK支持批量图片推理; ARM支持图像分割 || 2020.01.16 | 0.4.7 | 上线海思NNIE1.2, 支 持EasyEdge以及EasyDL; ARM引擎升级;增加推荐阈值支持 || 2019.12.26 | 0.4.6 | 海思NNIE支持EasyDL专业版 || 2019.11.02 | 0.4.5 | 移除 curl依赖;支持自动编译OpenCV;支持EasyDL 专业版 Yolov3;支持EasyDL经典版高精度物体检测模型升级||2019.10.25|0.4.4|ARM引擎升 级,性能提升30%; 支持EasyDL专业版模型 || 2019.09.23 | 0.4.3 | 增加海思NNIE加速芯片支持 || 2019.08.30 | 0.4.2 | ARM引擎升级;支持 分类高性能与高精度模型 || 2019.07.25 | 0.4.1 | 引擎升级,性能提升 || 2019.06.11 | 0.3.3 | paddle引擎升级;性能提升 || 2019.05.16 | 0.3.2 | 新增armv7l支持 || 2019.04.25 | 0.3.1 | 优化硬件支持 || 2019.03.29 | 0.3.0 | ARM64 支持;效果提升 || 2019.02.20 | 0.2.1 | paddle 引擎支持;效果提升||2018.11.30|0.1.0|第一版!|

【1.0 接口升级】参数配置接口从1.0.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例。【关于SDK包与RES模型文件夹配套使用的说明】 我们强烈建议用 户使用部署tar包中配套的SDK和RES一起使用。更新模型时,如果SDK版本号有更新,请务必同时更新SDK,旧版本的SDK可能无法正确适 配新发布出来的RES。

快速开始

SDK在以下环境中测试通过

- aarch64(arm64), Ubuntu 16.04, gcc 5.3 (RK3399)
- Hi3559AV100, aarch64, Ubuntu 16.04, gcc 5.3
- Hi3519AV100, armv7I , HiLinux 4.9.37, (Hi3519AV100R001C02SPC020)
- armv7hf, Raspbian, (Raspberry 3b)
- aarch64, Raspbian, (Raspberry 4b)
- armv7hf, Raspbian, (Raspberry 3b+)
- armv7hf, Ubuntu 16.04, (RK3288)
- Bitmain se50 BM1684, Debian 9
- Rockchip rk3399pro, Ubuntu 18.04
- Rockchip rv1126, Debain 10
- Rockchip rk3568, Ubuntu 20.04
- Rockchip rk3588, Ubuntu 20.04
- Atlas200(华为官网指定的Ubuntu 18.04版本)
- Amlogic A311D, Ubuntu 20.04
- MLU220, aarch64, Ubuntu 18.04

安装依赖

依赖包括

- cmake 3+
- gcc 5.4 以上(需包含 GLIBCXX_3.4.22), gcc / glibc版本请以实际SDK ldd的结果为准
- opencv3.4.5 (可选)

依赖说明:树莓派 树莓派Raspberry默认为armv7hf系统,使用SDK包中名称中包含 armv7hf_ARM_的tar包。如果是aarch64系统,使用SDK包中 名称中包含 aarch64_ARM_的tar包。

在安装前可通过以下命令查看是32位还是64位 :

```
getconf LONG_BIT
32
```

依赖说明:比特大陆SE计算盒 需要安装SophonSDK V23.05.01及以上版本,SDK的默认安装位置为/opt/sophon/,如SDK安装在自定义地址, 需在CMakeList.txt中指定SDK安装地址:

```
**这里修改并填入所使用的SophonSDK路径**
set(EDGE_BMSDK_ROOT "{这里填写sdk路径}")
```

可通过命令 bm-smi 查看内部SDK和驱动的版本号(SophonSDK V23.05.01对应的内部SDK和驱动为0.4.6)。对于使用旧版BM1684 SDK或者 低版本SophonSDK的用户,可参考SophonSDK安装包中的《LIBSOPHON 使用手册》先卸载旧版BM1684 SDK,安装、升级SophonSDK。

依赖说明:海思开发板海思开发板需要根据海思SDK文档配置开运行环境和编译环境,SDK和opencv都需要在该编译环境中编译。NNIE1.2用

arm-himix200-linux交叉编译好的opencv,下载链接:https://pan.baidu.com/s/13QW0ReeWx4ZwgYg4Iretyw 密码:yq0s。下载后修改SDK CMakesList.txt

依赖说明:RK3399Pro 所有用例基于 Npu driver版本1.7.1的RK3399pro开发板测试通过,SDK采用预编译模式,请务必确保板上驱动版本为 1.7.1 查看RK3399Pro板上driver版本方法:dpkg -I | grep 3399pro

依赖说明:RV1109/RV1126 所有用例基于Rknn_server版本1.7.3的RV1126开发板测试通过,SDK采用预编译模式,请务必确保板上驱动版本为 1.7.3 查看RV1109/RV1126板上Rknn_server版本方法: strings /usr/bin/rknn_server | grep build

依赖说明:RK3568 所有用例基于Rknn_server版本1.2.0的RK3568开发板测试通过, 查看RK3568板上Rknn_server版本方法: strings /usr/bin/rknn_server | grep build

依赖说明:RK3588 RK3588开发板需要确保环境正确安装了RKNPU驱动,平台用例基于v0.8.0版本的RKNPU驱动测试通过,查看RK3588NPU驱 动版本的方法: sudo cat /sys/kernel/debug/rknpu/version

依赖说明:晶晨A311D 所有用例基于晶晨A311D开发板测试通过,需要驱动版本为 6.4.4.3(下载驱动请联系开发版厂商) 查看晶晨A311D开发 板驱动版本方法:dmesg | grep Galcore

依赖说明:英特尔iGPU用户在使用英特尔iGPU SDK前,需要根据英特尔官方文档提前安装好英特尔集成显卡驱动以及相关基础软件环境,安装 完成后通过 clinfo 指令确认OpenCL能够正常识别到集成显卡信息,正确识别集显情况下clinfo指令输出参考如下:

Number of platforms	1
Platform Name	Intel(R) OpenCL HD Graphics
Platform Vendor	Intel(R) Corporation
Platform Version	OpenCL 3.0
Platform Profile	FULL_PROFILE
Platform Extensions	cl_khr_byte_addressable_store cl_khr_device_uuid cl_khr_fp16 cl_khr_global_int32_base_atomics cl_khr_global_int32_extended_atomics cl_khr_icd cl_khr_local_int32
_base_atomics cl_khr_local_int32_extended_atomics	cl_intel_command_queue_families cl_intel_subgroups cl_intel_required_subgroup_size cl_intel_subgroups_short cl_khr_spir cl_intel_accelerator cl_intel_driver_dia
gnostics cl_khr_priority_hints cl_khr_throttle_hi	nts cl_khr_create_command_queue cl_intel_subgroups_char cl_intel_subgroups_long cl_khr_il_program cl_intel_mem_force_host_memory cl_khr_subgroup_extended_types c
<pre>l_khr_subgroup_non_uniform_vote cl_khr_subgroup_b</pre>	allot cl_khr_subgroup_non_uniform_arithmetic cl_khr_subgroup_shuffle cl_khr_subgroup_shuffle_relative cl_khr_subgroup_clustered_reduce cl_intel_device_attribute_
<pre>query cl_khr_suggested_local_work_size cl_intel_s</pre>	plit_work_group_barrier cl_khr_fp64 cl_khr_subgroups cl_intel_spirv_device_side_avc_motion_estimation cl_intel_spirv_media_block_io cl_intel_spirv_subgroups cl_k
hr_spirv_no_integer_wrap_decoration cl_intel_unif	ied_shared_memory cl_khr_mipmap_image cl_khr_mipmap_image_writes cl_intel_planar_yuv cl_intel_packed_yuv cl_intel_motion_estimation cl_intel_device_side_avc_moti
on_estimation cl_intel_advanced_motion_estimation	cl_khr_int64_base_atomics cl_khr_int64_extended_atomics cl_khr_image2d_from_buffer cl_khr_depth_images cl_khr_3d_image_writes cl_intel_media_block_io cl_intel_v
<pre>a_api_media_sharing cl_intel_sharing_format_query</pre>	cl_khr_pci_bus_info
Platform Host timer resolution	lns
Platform Extensions function suffix	INTEL
Platform Name	Intel(R) OpenCL HD Graphics
Number of devices	1
Device Name	Intel(R) UHD Graphics 630 [0x9bc8]
Device Vendor	Intel(R) Corporation
Device Vendor ID	0x8886
Device Version	OpenCL 3.0 NEO
Driver Version	22.53.25242.13
Device OpenCL C Version	OpenCL C 1.2
Device Type	690
Device Profile	FULL_PROFILE
Device Available	Yes
Compiler Available	Yes
Linker Available	Yes
Max compute units	
Max clock frequency	11500472
Device Partition	(core)

使用序列号激活 请在官网获取序列号

共義其服务院的 发布局派线路界,控制研究战的模型部署在未地、高线调明模型,可以选择转模型部署在未地的服务器、小型设备、软键一体方面专项适配硬件上。 游过API、SCXII一步集成、灵活适应不同自务转展。 至新自						
服务器 通用小型设备 专项适配硬件						
SDK API						
此处发布、下载的SDK为未授权SDK,需要前行	主控制台获取序列号激活后才能正式使用	月。SDK内附有对应版本的Demo及开发文档,开	发者可参考源代码完成开发。			
模型名称	发布版本 🗘	应用平台	模型加速	发布状态	发布时间 🗘	
		通用X86 CPU-Linux	基础版	 已发布 	2021-08-19 20:24	下载SDK
			精度无损压缩加速	•已发布	2021-08-19 20:24	下载加速版SDK
	134318-V1	W (b) b const to	基础版	•已发布	2021-08-19 20:35	下载SDK
ani//uHWirear	查看性能报告	WHERE A CONTRACT OF A CONTRACT	精度无损压缩加速	•已发布	2021-08-19 20:34	下载加速版SDK
			基础版	 已发布 	2021-08-19 18:17	下载SDK

SDK内bin目录下提供预编译二进制文件,可直接运行(二**进制运行详细说明参考下一小节**),用于图片推理和模型http服务,在二进制参数的 serial_num(或者serial_key)处填入序列号可自动完成联网激活(请确保硬件首次激活时能够连接公网,如果确实不具备联网条件,需要使用纯离 线模式激活,请下载使用百度智能边缘控制台纳管SDK)

SDK内提供的一些二进制文件,填入序列号可完成自动激活,以下二进制具体使用说明参考下一小节

./edgekit_serving --cfg=./edgekit_serving.yml

- ./easyedge_image_inference {model_dir} {image_name or image_directory} {serial_num}
- ./easyedge_serving {res_dir} {serial_key} {host} {port}

如果是基于源码集成,设置序列号方法如下

global_controller()->set_licence_key("")

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的变量或者源码设置)实例数鉴权环境变量设置方法

export EDGE_CONTROLLER_KEY_AUTH_MODE=2 export EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=30

实例数鉴权源码设置方法

```
global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2)
global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 300)
```

基于预编译二进制测试图片推理和http服务测试图片推理模型资源文件默认已经打包在开发者下载的SDK包中。

对于硬件使用为: Intel Movidius MyRIAD2 / MyRIAD X / IGPU on Linux x86_64 / armv7hf / aarch64, 在编译或运行demo程序前执行以下命 令:

source \${cpp_kit位置路径}/thirdparty/openvino/bin/setupvars.sh

或者执行

source \${cpp_kit位置路径}/thirdparty/openvino/setupvars.sh (openvino-2022.1+)

如果SDK内不包含setupvars.sh脚本,请忽略该提示

运行预编译图片推理二进制,依次填入模型文件路径(RES文件夹路径)、推理图片、序列号(序列号尽首次激活需要使用,激活后可不用填序列号 也能运行二进制)

./easyedge_image_inference {model_dir} {image_name or image_directory} {serial_num} LD_LIBRARY_PATH=../lib ./easyedge_image_inference ../../../RES /xxx/cat.jpeg "1111-1111-1111"

demo运行效果:



图片加载失败

```
    >./easyedge_image_inference ../../../RES 2.jpeg
    2019-02-13 16:46:12,659 INFO [EasyEdge] [easyedge.cpp:34] 140606189016192 Baidu EasyEdge Linux Development Kit
    0.2.1(20190213)
    2019-02-13 16:46:14,083 INFO [EasyEdge] [paddlev2_edge_predictor.cpp:60] 140606189016192 Allocate graph success.
    2019-02-13 16:46:14,326 DEBUG [EasyEdge] [paddlev2_edge_predictor.cpp:143] 140606189016192 Inference costs 168 ms
    1, 1:txt_frame, p:0.994905 loc: 0.168161, 0.153654, 0.920856, 0.779621
    Done
```

启动http服务 bin目录下提供编译好的启动http服务二进制文件,可直接运行

推荐使用 edgekit_serving 启动模型服务 LD_LIBRARY_PATH=../lib ./edgekit_serving --cfg=./edgekit_serving.yml

也可以使用 easyedge_serving 启动模型服务

./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401}

LD_LIBRARY_PATH=../lib ./easyedge_serving ../../../RES "1111-1111-1111-1111" 0.0.0.0 24401

后,日志中会显示

HTTP(or Webservice) is now serving at 0.0.0.24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试,网页右侧会展示模型推理结果



同时,可以调用HTTP接口来访问服务。

请求http服务 以图像预测场景为例(非语义分割模型场景,语义分割请求方式参考后面小节详细文档),提供一张图片,请求模型服务的示例参考如下demo

python示例代码如下

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
```

params 为GET参数 data 为POST Body result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1}, data=img).json()

C# 使用示例代码如下

FileStream fs = new FileStream("./img.jpg", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close(); WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd());

sr.Close();

 $response. \underline{Close}();$

C++ 使用示例代码如下,需要安装curl

```
**include <sys/stat.h>**
**include <curl/curl.h>**
**include <iostream>**
**include <string>**
**define S_ISREG(m) (((m) & 0170000) == (0100000))**
**define S_ISDIR(m) (((m) & 0170000) == (0040000))**
size_t write_callback(void *ptr, size_t size, size_t num, void *data) {
  std::string *str = dynamic_cast<std::string *>((std::string *)data);
  str->append((char *)ptr, size*num);
  return size*num:
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./img.jpg";
  FILE *fp = NULL;
  std::string response;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\"\n", post_data_filename);
     return -1;
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
    curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
    curl_easy_setopt(curl, CURLOPT_POST, 1L);
    curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
    curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
 curl_easy_setopt(curl, CURLOPT_WRITEFUNCTION, write_callback);
    curl_easy_setopt(curl, CURLOPT_WRITEDATA, &response);
    res = curl_easy_perform(curl);
     if (res != CURLE_OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     std::cout << response << std::endl; // response即为返回的json数据
     curl_easy_cleanup(curl);
  curl_global_cleanup();
  fclose(fp);
  return 0:
```

关于http接口的详细介绍参考下面集成文档http服务章节的相关内容

集成文档

使用该方式,将运行库嵌入到开发者的程序当中。 **编译demo项目** SDK src目录下有完整的demo工程,用户可参考该工程的代码实现方式将SDK 集成到自己的项目中,demo工程可直接编译运行:

```
cd src
mkdir build && cd build
cmake .. && make
./easyedge_image_inference {模型RES文件夹} {测试图片路径}
**如果是NNIE引擎,使用sudo运行**
sudo ./easyedge_image_inference {模型RES文件夹} {测试图片路径}
```

(可选) SDK包内一般自带opencv库,可忽略该步骤。如果希望SDK自动编译安装所需要的OpenCV库,修改cmake的 optionEDGE_BUILD_OPENCV为ON即可。 SDK会自动从网络下载opencv源码,并编译需要的module、链接。注意,此功能必须需联网。 cmake -DEDGE_BUILD_OPENCV=ON .. && make -j16

若需自定义library search path或者gcc路径,修改CMakeList.txt即可。

对于硬件使用为Intel Movidius MyRIAD2 / MyRIAD X 的,如果宿主机找不到神经计算棒Intel® Neural Compute Stick,需要执行以下命令添加 USB Rules:

cp \${cpp_kit位置路径}/thirdparty/openvino/deployment_tools/inference_engine/external/97-myriad-usbboot.rules /etc/udev/rules.d/ sudo udevadm control --reload-rules sudo udevadm trigger sudo ldconfig

使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

// step 1: 配置运行参数 EdgePredictorConfig config; config.model_dir = {模型文件目录};

// step 2: 创建并初始化Predictor; 这这里选择合适的引擎
auto predictor = global_controller()->CreateEdgePredictor(config);

// step 3-1: 预测图像 auto img = cv::imread({图片路径}); std::vector<EdgeResultData> results; predictor->infer(img, results);

```
// step 3-2: 预测视频
std::vector<EdgeResultData> results;
FrameTensor frame_tensor;
VideoConfig video_config;
video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h
video_config.source_value = video_src;
/*
... more video_configs, 根据需要配置video_config的各选项
*/
auto video_decoding = CreateVideoDecoding(video_config);
while (video_decoding->next(frame_tensor) == EDGE_OK) {
  results.clear();
  if (frame_tensor.is_needed) {
    predictor->infer(frame_tensor.frame, results);
    render(frame_tensor.frame, results, predictor->model_info().kind);
  //video_decoding->display(frame_tensor); // 显示当前frame,需在video_config中开启配置
  //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置
```

对于口罩检测模型,将 EdgePredictorConfig config修改为PaddleMultiStageConfig config即可。

口罩检测模型请注意输入图片中人脸大小建议保持在 8*8到96*96像素之间,可根据场景远近程度缩放图片后再传入SDK。

SDK参数配置 SDK的参数通过 EdgePredictorConfig::set_config和global_controller()->set_config配置。set_config的所有key 在easyedge_xxxx_config.h中。其中

- PREDICTOR前缀的key是不同模型相关的配置,通过EdgePredictorConfig::set_config设置
- CONTROLLER前缀的key是整个SDK的全局配置,通过global_controller()->set_config设置

```
以序列号为例,KEY的说明如下:
```

```
/**
* @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号
* 值类型:string
* 默认值:空
*/
static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";
```

使用方法如下:

```
EdgePredictorConfig config;
config.model_dir = ...;
config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, "1DB7-1111-1111-D27D");
```

具体支持的运行参数可以参考开发工具包中的头文件的详细说明。

初始化

● 接口

auto predictor = global_controller()->CreateEdgePredictor(config); predictor->init();

若返回非0,请查看输出日志排查错误原因。

预测图像

```
• 接口
```

```
/**
* @brief
```

```
* 通用接口
```

* @param image: must be BGR , HWC format (opencv default)

```
* @param result
```

* @return

```
*/
```

```
virtual int infer(
```

cv::Mat& image, std::vector<EdgeResultData>& result

```
) = 0;
```

图片的格式务必为opencv默认的BGR, HWC格式。

● 返回格式

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index; // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测活图像分割时才有
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割时才有
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding, 游程编码的mask
};
```

关于矩形坐标

x1 * 图片宽度 = 检测框的左上角的横坐标

y1 * 图片高度 = 检测框的左上角的纵坐标

```
x2 * 图片宽度 = 检测框的右下角的横坐标
```

y2 * 图片高度 = 检测框的右下角的纵坐标

关于图像分割mask

```
cv::Mat mask为图像掩码的二维数组
{
{0,0,0,0,0,0,0,0,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,0,0,0,0,0,0,0},
}
其中1代表为目标区域,0代表非目标区域
```

关于图像分割mask_rle

该字段返回了mask的游程编码,解析方式可参考 http demo

以上字段可以参考demo文件中使用opencv绘制的逻辑进行解析

预测视频

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

● 接口

classVideoDecoding :

```
/**
* @brief 获取输入源的下一帧
* @param frame_tensor
* @return
*/
virtual int next(FrameTensor & frame_tensor) = 0;
/**
* @brief 显示当前frame_tensor中的视频帧
* @param frame_tensor
* @return
*/
virtual int display(const FrameTensor &frame_tensor) = 0;
/**
* @brief 将当前frame_tensor中的视频帧写为本地视频文件
* @param frame_tensor
* @return
*/
virtual int save(FrameTensor &frame_tensor) = 0;
/**
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
 * @brief 获取视频的width属性
 * @return
 */
virtual int get_width() = 0;
/**
* @brief 获取视频的height属性
* @return
*/
virtual int get_height() = 0;
```

struct VideoConfig

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type;
                          // 输入源类型
  std::string source_value;
                         // 输入源地址, 如视频文件路径、摄像头index、网络流地址
  int skip_frames{0};
                       // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
  int retrieve_all{false};
                     // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is needed置为false
                       // 在采取抽帧之前设置视频的fps
  int input_fps{0};
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
  bool enable display{false};
  std::string window_name{"EasyEdge"};
  bool display_all{false};
                        // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                        // 是否存储所有frame, 若为false, 仅存储根据skip_frames抽取的frame
  std::map<std::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被 抽取返回,以作为显示或存储用。input_fps:用于抽帧前设置fps。resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。conf:高级选项。部分配置会通过该map来设置。

注意: 1.如果使用 VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。 2.使用摄像头抽帧时,如 果通过 resolution 设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

设置序列号

请在网页控制台中申请序列号,并在init初始化前设置。 LinuxSDK 首次使用需联网授权。

EdgePredictorConfig config; config.set_config(easyedge::params::PREDICTOR_KEY_SERIAL_NUM, "this-is-serial-num");

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

EdgeLogConfig log_config; log_config.enable_debug = true; global_controller()->set_log_config(log_config);

http服务

1. 开启http服务 http服务的启动参考demo_serving.cpp文件。

/** * @brief 开启一个简单的demo http服务。 * 该方法会block直到收到sigint/sigterm。 * http服务里,图片的解码运行在cpu之上,可能会降低推理速度。 * @tparam ConfigT * @param config * @param host * @param port * @param service_id service_id user parameter, uri '/get/service_id' will respond this value with 'text/plain' * @param instance_num 实例数量,根据内存/显存/时延要求调整 * @return */ template<typename ConfigT> int start_http_server(const ConfigT &config, const std::string &host int port,

2. http接口详细说明 http 请求方式一:无额外编码 URL中的get参数:

const std::string &service_id, int instance_num = 1);

参数	说明		默认值	
threshold	阈值过滤,	0~1	如不提供	,则会使用模型的推荐阈值

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

Python请求示例 (图片测试,针对图像分类、物体检测、实例分割等模型)

import requests

with open('./1.jpg', 'rb') as f: img = f.read() result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1}, data=img).json()

Python请求示例 (图片测试, 仅针对语义分割模型, 同其他CV模型不同, 语义分割模型输出为灰度图)

import requests

```
with open('./1.jpg', 'rb') as f:
    img_data = f.read()
    res = requests.post('http://127.0.0.1:24401/',
        data=img_data)
    with open("gray_result.png", "wb") as fb:
        fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
```

Python请求示例 (视频测试, 注意:区别于图片预测,需指定Content-Type;否则会调用图片推理接口)

import requests

with open('./1.mp4', 'rb') as f: video_data = f.read() result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1}, headers={'Content-Type': 'video'}, data=video_data).json()

http 请求方法二:json格式,图片传base64格式字符串 HTTP方法:POST Header如下:

参数

值 Content-Type application/json

Body请求填写:

• 图像分类网络: body 中请求示例

```
{
"image": "<base64数据>",
"top_num": 5
}
```

body中参数详情

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
top_nu m	否	number	-	返回分类数量,不填该参数,则默认返回全部分类结果

• 物体检测和实例分割网络: Body请求示例:

```
{
"image": "<base64数据>",
 "threshold": 0.3
}
```

body中参数详情:

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
thresho Id	否	number	-	默认为推荐阈值,也可自行根据需要进行设置

• 语义分割网络: Body请求示例:

```
{
"image": "<base64数据>"
}
```

body中参数详情(语义分割由于模型特殊性,不支持设置threshold值,设置了也没有意义):

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部

Python请求示例(非语义分割模型参考如下代码)

import base64

import requests
import requests
def main():
with open("1.jpg 【图片路径】", 'rb') as f:
result = requests.post("http://{服务ip地址}:24401/", json={
"image": base64.b64encode(f.read()).decode("utf8")
})
print(result.request.body)
print(result.request.headers)
print(result.content)
ifname == 'main':
main()

Python 请求示例 (针对语义分割模型,同其他CV模型不同,语义分割模型输出为灰度图)

```
import base64
import requests
def main():
    with open("1.jpg 【图片路径】", 'rb') as f:
        res = requests.post("http://{服务ip地址}:24401/", json={"image": base64.b64encode(f.read()).decode("utf8")})
        with open("gray_result.png", "wb") as fb:
        fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
iff __name__ == '__main__':
        main()
```

http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考预测图像-返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

返回示例

```
{
    "cost_ms": 52,
    "error_code": 0,
    "results": [
        {
            "confidence": 0.94482421875,
            "index": 1,
            "label": "IronMan",
            "x1": 0.059185408055782318,
            "x2": 0.18795496225357056,
            "y1": 0.14762254059314728,
            "y2": 0.52510076761245728,
            "mask": "...", // 图像分割模型字段
            "trackId": 0, // 目标追踪模型字段
        },
    ]
}
```

其他配置

1. 日志名称、HTTP 网页标题设置

通过global_controller的set_config方法设置:

效果如下:

图片加载失败

FAQ

1. 如何处理一些 undefined reference / error while loading shared libraries?

如:./easyedge_demo: error while loading shared libraries: libeasyedge.so.1: cannot open shared object file: No such file or directory 这是因为二进制运行时Id无法找到依赖的库。如果是正确cmake && make 的程序,会自动处理好链接,一般不会出现此类问题。

遇到该问题时,请找到具体的库的位置,设置LD_LIBRARY_PATH。

示例一:libverify.so.1: cannot open shared object file: No such file or directory 链接找不到libveirfy.so文件,一般可通过 export LD_LIBRARY_PATH=\${LD_LIBRARY_PATH}:../../lib 解决(实际冒号后面添加的路径以libverify.so文件所在的路径为准)

示例二:libopencv_videoio.so.4.5: cannot open shared object file: No such file or directory 链接找不到ibopencv_videoio.so文件,一般可通 过 export LD_LIBRARY_PATH=\${LD_LIBRARY_PATH}:../../thirdparty/opencv/lib 解决(实际冒号后面添加的路径以libopencv_videoio.so所在路 径为准)

2. EasyDL SDK与云服务效果不一致,如何处理?

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

3. 如何将我的模型运行为一个http服务?目前epp sdk暂未集成http运行方式; 0.4.7版本之后,可以通过start_http_server方法开启http服务。

4. 运行NNIE引擎报permission denied 日志显示:

open sys: Permission denied open err : Permission denied open err : Permission denied

请使用sudo在root下运行。

5. 运行SDK报错 Authorization failed

情况一:日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888"
./easyedge_demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限 于)以下可能的情况:

- MAC地址变化
- 磁盘变更
- BIOS重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除~/.baidu/easyedge 目录,再重新激活。

6. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

7. 运行NNIE引擎报错 std::bad_alloc 检查开发板可用内存,一些比较大的网络占用内存较多,推荐内存500M以上

8. 运行二进制时,提示 libverify.so cannot open shared object file

可能cmake没有正确设置rpath,可以设置LD_LIBRARY_PATH为sdk的lib文件夹后,再运行:

LD_LIBRARY_PATH=\$LD_LIBRARY_PATH:../lib ./easyedge_demo

9. 编译时报错:file format not recognized 可能是因为在复制SDK时文件信息丢失。请将整个压缩包复制到目标设备中,再解压缩、编译

の Linux集成文档-Python

简介

本文档介绍 EasyDL 的 Linux Python SDK 的使用方法,适用于 EasyDL 通用版和BML。

- 网络类型支持:图像分类,物体检测
- 硬件支持:
 - Intel Movidius MyRIAD2 / MyRIAD X / IGPU
 - 瑞芯微 RK3399Pro
- 语言支持: Intel Movidius MyRIAD2 / MyRIAD X / IGPU: Python 3.5, 3.6, 3.7 瑞芯微 RK3399Pro: Python 3.6

Release Notes

时间 医木 沿明

ոյլոյ	瓜牛	100-97
2022.10. 27	1.3.5	新增Armv7 CPU、Armv8 CPU、Jetson、华为昇腾Atlas开发板对应Python SDK,支持图像分类、物体检测、人脸检测、实例 分割;新增 Intel IGPU 支持
2022.05. 18	1.3.0	新增RK3399Pro NPU对应Python SDK,支持图像分类、物体检测
2021.01. 27	1.2.1	EasyDL经典版分类高性能模型升级;支持更多模型
2020.12. 18	1.2.0	推理引擎升级;接口升级;性能优化
2020.09. 17	1.1.19	支持更多模型
2020.08. 11	1.1.18	性能优化
2020.06. 23	1.1.17	支持更多EasyDL专业版模型
2020.04. 16	1.1.15	技术优化;升级 OpenVINO 版本
2020.01. 16	1.1.12	预测函数默认使用推荐阈值
2019.12. 26	1.1.11	EasyDL 专业版支持加速棒
2019.10. 21	1.1.9	支持 EasyDL 专业版
2019.07. 19	1.1.7	提供模型更新工具
2019.03. 15	1.1.0	架构与功能完善
2019.02. 28	1.0.6	引擎功能完善
2019.02. 13	1.0.5	paddlepaddle 支持
2018.11. 30	1.0.0	第一版!

2020-12-18: 【接口升级】 序列号配置接口从1.2.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请 尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

快速开始

1. 安装依赖

根据引擎的不同,SDK 依赖了不同的底层引擎。根据所需自行安装。

依赖说明:Intel Movidius 加速棒 使用Intel Movidius加速棒 SDK、Intel IGPU 预测时,必须安装 OpenVINO 预测引擎,两种方式:

- 使用 OpenVINO[™] toolkit 安装,请参考 OpenVINO toolkit 文档安装 2020.3.1LTS(必须)版本, 安装时可忽略Configure the Model Optimizer及 后续部分。
- 使用源码编译安装,请参考 Openvino Inference Engine文档编译安装 2020.3.1 (必须)版本。

安装完毕,运行之前,请按照OpenVinono的文档 设置环境变量

source /opt/intel/openvino/bin/setupvars.sh

依赖说明:RK3399Pro 所有用例基于 Npu driver版本1.7.3的RK3399pro开发板测试通过 查看RK3399Pro板上driver版本方法:运行sdk内提供

demo项目,日志里会提供API和Driver版本信息

2022-12-2	0 14:26:07,765 VERBOSE [EasyEdge] [rockchip_edge_predictor.cpp:87] 547887054864 Create predictor , 5029536
D RKNNAPI	:
D RKNNAPI	: RKNN VERSION:
D RKNNAPI	: API: 1.7.3 (0cfd4a1 build: 2022-08-15 17:10:10)
D RKNNAPI	: DRV: 1.7.3 (c4ea832 build: 2022-08-13 09:13:08)
D RKNNAPI	:

升级399Pro driver版本参考瑞芯微github: https://github.com/airockchip/RK3399Pro_npu 2. 安装 easyedge python wheel 包 安装说明: Intel

Movidius 加速棒 / Intel IGPU

pip3 install -U BaiduAI_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_x86_64.whl

具体名称以 SDK 包中的 whl 为准。

安装说明:RK3399Pro

pip3 install -U BaiduAl_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_aarch64.whl pip3 install -U EasyEdge_Devkit_RK3399Pro-{版本号}-cp36-cp36m-linux_aarch64.whl

具体名称以 SDK 包中的 whl 为准,特别注意这里要同时安装两个whl包 安装说明:ArmV7 CPU

pip3 install -U BaiduAl_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_armv7l.whl pip3 install -U EasyEdge_Devkit_ARM-{版本号}-cp36-cp36m-linux_armv7l.whl

安装说明:ArmV8 CPU (Aarch64 CPU)

pip3 install -U BaiduAl_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_aarch64.whl pip3 install -U EasyEdge_Devkit_ARM-{版本号}-cp36-cp36m-linux_aarch64.whl

安装说明: Jetson SDK

pip3 install -U BaiduAI_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_aarch64.whl pip3 install -U EasyEdge_Devkit_JetPack{版本号}-{版本号}-cp36-cp36m-linux_aarch64.whl

安装说明:华为昇腾Atlas开发板

pip3 install -U BaiduAl_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_aarch64.whl EasyEdge_Devkit_Atlas200-{版本号}-cp36-cp36m-linux_aarch64.whl

3. 使用序列号激活

获取序列号	94.萬枝服务状期 次時後期後継券,将106元成坊模型部署在本地、高 港江APA、SDK进一参集成、灵活道広不同业务结果、 東本新館券 室前 2014 84.月 84.月 84.月 85.00 API 84.26 85.00 API 84.26 85.00 API 84.26 85.00 API 84.26 85.00 86.40	建成用模型。可以选择待模型部署 自然现序列号激发后才能正式使用	在未地的服务器、小型设备、敬健一体方案专项 、SDK内附有对应版本的Demo及开发文档,开;	送都硬件上。 此者可參考團代码完成开发。			
	模型名称	发布版本 🗘	应用平台	模型加速	发布状态	发布时间 🗘	
			WIRKON ODLI LINU	基础版	 已发布 	2021-08-19 20:24	下载SDK
			WHYRE CHO-FILM	精度无损压缩加速	 已发布 	2021-08-19 20:24	下载加速版SDK
	and the terms	134318-V1	WASHINGTON	基础版	 已发布 	2021-08-19 20:35	下截SDK
	and the Winear	查看性能报告	361930/01-0-Fluinx	精度无损压缩加速	 已发布 	2021-08-19 20:34	下载加速版SDK
			Webberg and the	基础板	 已发布 	2021-08-19 18:17	下截SDK

修改demo.py 填写序列号

```
pred = edge.Program()
pred.set_auth_license_key("这里填写序列号")
```

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己 的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的设置),需要调用函数指定实例数鉴权模式,并且实例数鉴权模式下,支持指定 license证书更新时间,单位是秒,要求设置为大于20的整数,否则会采用默认的license更新时间,参考

pred.set_instance_auth_mode() pred.set_instance_update_interval(200)

4. 测试 Demo

输入对应的模型文件夹(默认为RES)和测试图片路径,运行:

python3 demo.py {model_dir} {image_name.jpg}

测试效果:



使用说明

使用流程

import BaiduAI.EasyEdge as edge

```
pred = edge.Program()
pred.set_auth_license_key("这里填写序列号")
pred.init(model_dir={RES文件夹路径}, device=edge.Device.MOVIDIUS, engine=edge.Engine.OPENVINO)
pred.infer_image({numpy.ndarray的图片})
pred.close()
```

初始化

```
● 接口
```

```
def init(self,
    model_dir,
    device=Device.LOCAL,
    engine=Engine.PADDLE_FLUID,
    config_file='conf.json',
    preprocess_file='preprocess_args.json',
    model_file='model',
    params_file='params',
    graph_file='graph.ncsmodel',
    label_file='label_list.txt',
    device_id=0
    ):
 .....
Args:
   device: Device.CPU
   engine: Engine.PADDLE_FLUID
   model_dir: str
     model dir
   preprocess_file: str
   model_file: str
   params_file: str
   graph_file: str
   label_file: str
   device_id: int
Raises:
   RuntimeError, IOError
Returns:
   bool: True if success
 .....
```

● 接口

<pre>def infer_image(self, img,</pre>
Ards
Algo.
ing: np.ndarray or bytes
threshold: float
only return result with confidence larger than threshold
channel_order: string
channel order HWC or CHW
color_format: string
color format order RGB or BGR
data_type: string
image data type
Returns:
list
нин

• 返回格式: [dict1, dict2, ...]

字段	类型	取值	说明
confidence	float	0~1	分类或检测的置信度
label	string		分类或检测的类别
index	number		分类或检测的类别
x1, y1	float	0~1	物体检测,矩形的左上角坐标 (相对长宽的比例值)
x2, y2	float	0~1	物体检测,矩形的右下角坐标(相对长宽的比例值)
mask	string/numpy.ndarray	图像分割的mask	

关于矩形坐标

- x1 * 图片宽度 = 检测框的左上角的横坐标
- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2*图片宽度 = 检测框的右下角的横坐标
- y2 * 图片高度 = 检测框的右下角的纵坐标

可以参考 demo 文件中使用 opencv 绘制矩形的逻辑。

结果示例

• i) 图像分类

```
{
    "index": 736,
    "label": "table",
    "confidence": 0.9
}
```

● ii) 物体检测

Baidu 百度智能云文档

```
"y2": 0.91211,
"label": "cat",
"confidence": 1.0,
"x2": 0.91504,
"index": 8,
"y1": 0.12671,
"x1": 0.21289
}
```

• iii) 图像分割

```
{
    "name": "cat",
    "score": 1.0,
    "location": {
        "left": ...,
        "top": ...,
        "width": ...,
        "height": ...,
    },
    "mask": ...
}
```

mask字段中,data_type为numpy时,返回图像掩码的二维数组

```
{
  {
  {
  {
  {
  {
  0.0.0.0.0.0.0.0.0.0.0.0},
  {
  0.0.0.1.1.1.0.0.0.0},
  {
  0.0.0.1.1.1.0.0.0.0},
  {
  0.0.0.1.1.1.0.0.0.0},
  {
  0.0.0.0.0.0.0.0},
  {
  0.0.0.0.0.0.0.0},
  }
  其中1代表为目标区域,0代表非目标区域
```

data_type为string时,mask的游程编码,解析方式可参考 demo

升级模型 适用于经典版升级模型,执行bash update_model.sh,根据提示,输入模型路径、激活码、模型ID、模型版本,等待模型更新完毕即 可。

FAQ

Q: 运行SDK报错 Authorization failed

情况一:日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888"
./easyedge_demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- MAC地址变化
- 磁盘变更
- BIOS重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除~/.baidu/easyedge 目录,再重新激活。

情况三: ArmV7、ArmV8 CPU、Jetson、Atlas Python SDK日志提示ImportError: libavformat.so.58: cannot open shared object file: No such file or

directory 或者其他类似so找不到可以在LD_LIBRARY_PATH环境变量加上libs和thirdpartylibs路径,例如

export LD_LIBRARY_PATH=/xxx/libs:/xxx/thirdpartylibs:\$LD_LIBRARY_PATH # tips: 这里/xxx需要替换为真实路径,/xxx路径查找方法如下

查找安装包内libs和thirdpartylibs路径的方法如下(以Atlas SDK为例,其他SDK查找方法类似):

pip3 show EasyEdge-Devkit-Atlas200 # 结果中会显示 Location 路径,也就是包的安装路径 **libs和thirdpartylibs两个路径在 Location 所指示的路径 easyedge_CANN 子文件夹下**

心 Linux集成文档-Atlas

简介

本文档介绍EasyEdge/EasyDL的Linux Atlas SDK的使用方法。

注意Atlas有两种产品形态,Atlas 200和Atlas 300,请参见此处的文档说明

- 网络类型支持:图像分类
- 硬件支持:
 - CPU: aarch64
 - Atlas 200 卡
- 操作系统支持:Atlas指定的Linux版本, Ubuntu 16.04 aarch64 ,请从Atlas文档中下载。

性能数据

数据仅供参考,实际数值根据使用线程数、利用率等情况可能有所波动

模型类型	模型算法	芯片类型	SDK类型	实测硬件	单次预测耗时
EasyDL 图像分类	高性能	Atlas 200	Atlas 200	Atlas 200DK	9ms
EasyDL 图像分类	高精度	Atlas 200	Atlas 200	Atlas 200DK	12ms
EasyDL 物体检测	高性能	Atlas 200	Atlas 200	Atlas 200DK	11ms
EasyDL 物体检测	高精度	Atlas 200	Atlas 200	Atlas 200DK	31ms

Release Notes

时间	版本	说明
2020.6.15	0.2	支持物体检测
2020.3.10	0.1	初始版本,支持图像分类

测试atlas 200的官方demo

请参见此处的文档说明, 搭建开发环境,测试atlas 200的mindstudio demo通过后,再测试

快速开始

SDK在以下环境中测试通过

- ubuntu 16.04, aarch64-linux-gnu-g++ 5.4 , 编译机
- ubuntu 16.04 , 开发板

Atlas DDK 的ddk_info信息:

```
{
    "VERSION": "1.3.T34.B891",
    "NAME": "DDK",
    "TARGET": "Atlas DK"
}
```

2. 测试Demo

编译运行: 下载后,模型资源文件默认已经打包在开发者下载的SDK包中,

Step 1: 运行一次unpack.sh脚本,会得到测试demo。

Step 2:请在官网获取序列号,填写在demo_async.cpp及demo_sync.cpp的开始处license_key字段。



图片加载失败

step3: 准备测试图片

覆盖image目录下的 1.jpg,更多图片可以用于demo中的批量测试模式

step4: 修改test_200.sh下的以下开发板登录信息

export DDK_PATH=\$HOME/tools/che/ddk/ddk # ddk的安装路径

SSH_USER=HwHiAiUser@192.168.3.25 # 200 开发板的ssh登录信息 PORT=8822 # 200 开发板的ssh登录端口

step: 运行demo, 会自动编译OpenCV 3.4库

cd demo sh test_200.sh

图像分类的demo运行效果:

```
[stat] [100001]image/1.jpg(4 images) time used: 41ms (at 1583765958531) total:705ms
[result][100001][image/1.jpg][281470472005664] is: n07747607 orange 0.973633 950;
```

n07747607 orange 分类名 0.973633 分类概率 950 分类名的序号

物体检测的demo运行效果:

```
[stat] time used : 101ms; all time used:478
images[3] result:
label:no2_ynen;prob:0.985352 loc:[(0.459961,0.839844), (0.5625,0.988281)]
no2_ynen 分类名 , 也可以获取分类名的序号
0.985352 分类概率
loc:[(0.459961,0.839844), (0.5625,0.988281)] , 检测框的位置。(0.459961,0.839844)表示左上角的点,(0.5625,0.988281)右下角的
点;
如原始图片608 , 左上角(0.459961*608,0.839844*608) , 右下角(0.5625*608,0.988281*608)
```

SDK接口使用

使用该方式,将运行库嵌入到开发者的程序当中。

同步接口使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

// step 0: 设置序列号
global_controller()->set_licence_key("set your license here");

// step 1: 配置模型资源目录 AtlasConfig config; config.model_dir = {模型文件目录};

// step 2: 创建并初始化Predictor ;
auto predictor = global_controller()->CreateEdgePredictor(config);
int ret = predictor->init();
若返回非0,请查看输出日志排查错误原因。
auto img = cv::imread((图片路径));
// step 3: 预测图像
std::vector<EdgeResultData> result2;
predictor->infer(img, result2);
解析result2即可获取结果

异步接口使用流程

```
// step 0: 设置序列号
global_controller()->set_licence_key("set your license here");
```

```
// step 1: 配置模型资源目录
AtlasConfig config;
config.model_dir = {模型文件目录};
```

// step 3: 创建Predictor; 这这里选择合适的引擎 auto predictor = global_controller()->CreateEdgePredictor(config);

// step 4: 设置异步回调 predictor->set_result_handler(YOUR_HANDLER);

```
// step 5: 初始化
int ret = predictor->init();
**若返回非0,请查看输出日志排查错误原因。**
```

// step 6: 预测图像 auto img = cv::imread({图片路径}); color_format = kBGR; float threshold = 0.1;

```
uint64_t seq_id;
predictor->infer_async(img, color_format, 0.1, nullptr, seq_id);
**YOUR_HANDLER里面有seq_id的回调结果**
```

设置序列号

请在网页控制台中申请序列号,并在init初始化前设置。 LinuxSDK 首次使用需联网授权。

• 接口

virtual int set_licence_key(const std::string& license) = 0;

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

```
EdgeLogConfig log_config;
log_config.enable_debug = true;
global_controller()->set_log_config(log_config);
```

FAQ

1. 如何处理一些 undefined reference?

如: undefined reference to `curl_easy_setopt@CURL_OPENSSL_3'

方案1:通过安装libcurl3 libcurl-openssl1.0-dev来解决。 方案2:如果开发者想不想使用低版本的openssl (如Ubuntu 18.04),可以link静态 库easyedge_static.a,自己指定需要的Library的版本:

示例:修改CMakeList.txt

find_package(CURL REQUIRED) target_link_libraries(easyedge_demo \${OpenCV_LIBS} easyedge_static pthread \${CURL_LIBRARIES} verify_static \$(其他需要的库})

其中,其他需要的库视具体sdk中包含的库而定。

2. EasyDL 离线SDK与云服务效果不一致,如何处理?

目前离线SDK与云服务的处理有些许差异,具体如下:

- 图像分类模型:离线SDK与云服务使用通用(非快速训练、非AutoDL Transfer)的效果类似
- 物体检测模型:离线SDK的高精度模型与云服务的精度较低,服务性能更佳的效果类似

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

心 端云协同服务说明

服务简介

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

「云管理,端计算」的端云协同服务,具体包括:

- 在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新
- 断网状态下模型离线计算 (http服务,可调用与公有云API功能相同的接口)
- 联网状态下在平台管理设备运行状态、资源利用率

目前通用小型设备的应用平台支持Linux-ARM,具体使用流程请参考下方文档。

使用流程

Step 1 发布端云协同部署包

在我的部署包页面点击「发布端云协同部署包」

端云协同服务 > 我的部署包

端云协同服务说明							点击收
1、在可视化界面轻松实现模型部 2、断网状态下模型离线计算(ht 3、联网状态下在平台管理设备运	署包在边缘设备 tp服务,可调用 行状态、资源利	上的集成、版本更 与公有云API功能 用率	新 目同的接口)				
具体使用流程如下:							
、在本页面发布端云协同部署包 2、在「 <mark>我的本地设备</mark> 」页面新 3、在「 下发部署包到设备 」页	,填写相关信息: 曾设备,复制激; 面,将发布成功的	并申请,并在下方 舌命令,联网激活 内部署包一键下发	列表查看部署包发布 本地设备 到已激活的设备上,	状态 即可测试或正式使用			
发布端云协同部署包							
端云协同服务名称	模型ID	设备类型	应用平台	当前可用版本	新版本发布状态	部署设备数	操作
				暂无可用数	攻据		
				请稍后再试 			6

填写服务名称,选择模型版本并提交发布

端云协同服务 >	发布端云协同部署包		
设备类型	○ 服务器 (●) 通用小型设备		
模型名称	test2021	\checkmark	
端云协同服务名称			
选择版本	V1	\checkmark	
选择系统和芯片	□ Δ Linux 通用ARM		
发布部署包			

在列表查看部署包发布状态

端云协同服务说明								点击收款
1、在可视化界面轻松实现模型部 2、断网状态下模型离线计算(ht 3、联网状态下在平台管理设备运	署包在边线 tp服务,i 行状态、	象设备上的集成、 可调用与公有云A 资源利用率	版本更新 PI功能相同的接口)					
具体使用流程如下:								
 1、在本页面发布端云协同部署包 2、在「我的本地设备」页面新加 3、在「下发部署包到设备」页目 	,填写相注 增设备,多 面,将发行	关信息并申请,并 夏制激活命令,助 币成功的部署包	并在下方列表查看部署包发 《网激活本地设备 -键下发到已激活的设备上	^{设布状态} _,即可测试或正	式使用			
发布端云协同部署包								
端云协同服务名称	模型ID	设备类型	应用平台	当前可用版本	新版本发布状态	部署设备数	操作	
ecc	264	通用小型设备	Linux-通用ARM	V1	V1-已发布	0	下发到设备 发布新版本	
从发布页过来的	246	服务器	Linux-通用X86 CPU	V2, V1	V2-已发布	1	下发到设备 发布新版本	服务详情
图像分类高精度_猫狗-265	265	通用小型设备	Linux-通用ARM	-	V1-发布中	0		
						4	每页显示 10 🗸 <	1 >

Step 2 新增设备并激活

在我的本地设备页面新增设备

我的本地设备						点击地
E本页面新增设备、 列用率等信息。	联网激活本地设备后,即	『可将「 我的部署包 」页面	发布成功的部署包一键	下发到设备上。设备联网时	1,可以查看设备上部署的服务、该	设备的运行状态、资
新增设备						

1		新增设备	×
1. N.	设备名称	20个字符以内	
ŝ	备注信息	50个字符以内,自定义备注信息	
ž		0/50	
		确认取消	

在列表中,点击设备对应的「激活设备」操作,复制激活命令并在本地设备上执行即可

图	查看激活命令	× 收起 ^
在往	复制命令,并在本地设备上执行,即可激活设备	_
端云协同服务	curl -skfL 'https://verify.baidubce.com/active/v1/active/s sh?token=e177df4ed47b2265223a333630302c226b223a226e6f64652	setup.
我的本地设备 在本页面新增6 利用率等信息。	6e223a22353730313835302d382d6539386638638302d343231612d65 12d373138392d37626666336664306331343322c226e73223a2262616 6c2d636c6f7564222c227473223a31363039393939393434327d' -osetu && sh setup.sh	537653 657479 up.sh 可以宣看设备上部署的服务、设备的运行状态、资源
新增设备	\$018-7	1间 操作
linux x86 🙂		- 删除 激活设备 1 看详情
		每页显示 10 🗸 < 1 >

Step 3 下发部署包到设备,在本地调用

端云协同朋	(务) 下发部署包到设备		
将已成功	发布的端云协共服务部署包一键下发到已激活且在连接。	的设备	
服务名称	ecc	~	
模型ID	264		
应用平台	Linux-通用ARM	~	

部署包下发成功之后,会在本地启动一个HTTP推理服务。在浏览器中输入http://{设备ip}:{服务端口,默认8080},即可预览效果:

Al EasyEdge	× +		
C ▲ 不安全 yq01-	-ns-map0402a659.yq01:8081		
	【物体检测】90719 EasyDL检测超高精度-水果	调整阈值 ┥	— 当前阈值: 0.07 修改
		label	置信度
		tomato	0.998
		tomato	0.937
		tomato	0.841
		tomato	0.781
		kiwi	0.408
		tomato	0.324
		tomato	0.251
		tomato	0.146
	LAN		

具体接口调用说明请参考文档 SDK - HTTP服务调用说明

云端管理说明

模型部署包管理

在我的部署包页面可以进行已发布的模型部署包的管理。

在下发部署包到设备页面,将发布成功的部署包一键下发到已激活的设备上,即可测试或正式使用

发布及更新模型版本

点击「发布新版本」操作即可快速发布对应模型ID下的新版本。同一模型ID下已发布的模型版本均会显示在列表的「当前可用版本」中。

端云协同服务 > 我的部署包

端云协同服务说明 点击收起 1、在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新 2、断网状态下模型离线计算(http服务,可调用与公有云API功能相同的接口) 3、联网状态下在平台管理设备运行状态、资源利用率									
具体使用流程如下:									
 1、在本页面发布端云协同部署包 2、在「我的本地设备」页面新: 3、在「下发部署包到设备」页目 	,填写相 增设备, 面,将发	关信息并申请, 夏制激活命令,〕 布成功的部署包	并在下方列表查看部署包发 联网激活本地设备 一键下发到已激活的设备」	ۇ布状态 二,即可测试或正	式使用				
发布端云协同部署包									
端云协同服务名称	模型ID	设备类型	应用平台	当前可用版本	新版本发布状态	部署设备数	操作		
ecc	264	通用小型设备	Linux-通用ARM	V1	V1-已发布	0	下发到设备	发布新版本	
10 C 10 C 10	246	服务器	Linux-通用X86 CPU	V2, V1	V2-已发布	1	下发到设备	发布新版本 服	务详情
0.000.000.0000	265	通用小型设备	Linux-通用ARM	-	V1-发布中	0			
							每页显示	10 🗸 <	1 >
			发布新版本				×		
将最新训练的模型	型版本发	发布为服务,	发布成功后,即可	从云端下发到]设备				
	服务名	称 e	ecc						
	模型ID	2	.64						
	选择新	版本	V1		\sim				
			确认 取消	¥					

新版本发布成功后,即可在「下发部署包到设备」页面或当前服务的「服务详情」页面,将新版本下发到本地设备上。

端云协同服务 > 我的部署包

端云协同服务说明

- 1、在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新
- 2、断网状态下模型离线计算(http服务,可调用与公有云API功能相同的接口)
- 3、联网状态下在平台管理设备运行状态、资源利用率

具体使用流程如下:

1、在本页面发布端云协同部署包,填写相关信息并申请,并在下方列表查看部署包发布状态

- 2、在「 我的本地设备 」页面新增设备,复制激活命令,联网激活本地设备
- 3、在「下发部署包到设备」页面,将发布成功的部署包一键下发到已激活的设备上,即可测试或正式使用

发布端云协同部署包

端云协同服务名称	模型ID	设备类型	应用平台	当前可用版本	新版本发布状态	部署设备数	操作	
ecc	264	通用小型设备	Linux-通用ARM	V1	V1-已发布	0	下发到设备	发布新版本
No. Contraction	246	服务器	Linux-通用X86 CPU	V2, V1	V2-已发布	1	下发到设备	发布新版本 服务详情
图像分类高精度_猫狗-265	265	通用小型设备	Linux-通用ARM	-	V1-发布中	0		[mail

点击收t

F

端云协同服务 >	▶ 服务	计情				
服务名称 , A要 。P只要用 设备类型 服务器 下发到更多设备	, 模型 应用	!ID 246 平台 Linux-通	用X86 CPU	当前	可用版本 V2, V1	部署设备数 1/1
设备名称	最新下发模型版本	部署到期时间	设备连接状态	服务下发状态	最新同步时间	操作
ubuntu-local-fjy 💬	V2	2021-02-06	在线	V2下发成功	2021-01-07 14:45	查看服务配置 查看设备详情 下发新版本 移除设备

管理模型已部署的设备

在上述的「服务详情」页面,可以查看并管理当前服务已部署的设备,包括移除设备、将服务下发到更多的设备等。

端云协同服务 >	> 服务	详情							
服务名称 设备类型 服务器	■ 模型	ID 246 平台 Linux-通	用X86 CPU	当前	可用版本 V2, V1	部署设	设备数 1/1		
下发到更多设备									
设备名称	最新下发模型版本	部署到期时间	设备连接状态	服务下发状态	最新同步时间	操作			
ubuntu-local-fjy 😁	V2	2021-02-06	在线	V2下发成功	2021-01-07 14:45	查看服务配置 宣	這看设备详情	下发新版本	移除设备

本地设备管理

.

在我的本地设备页面可以进行所有本地设备的管理。

查看单台设备的运行状态

点击单台设备的「服务详情」,可查看设备上运行的多个服务及设备状态:

端云协同服务 > 我的本地	也设备					
线的本地设备 E本页面新增设备、联网激活 J用率等信息。	本地设备后,即可将「 — — — — — — — — — — — — — — — — — — —	的部署包」]	页面发布成功的部署包一键下发	到设备上。设备联	网时,可以查看设备上部	点击收起 署的服务、设备的运行状态、资源
新增设备						
设备名称	创建时间	设备类型	应用平台	设备连接状态	最新同步时间	操作
123123 💬	2021-01-07 13:25	-	-	未激活	-	删除 激活设备 查看详情
showcase-test 💬	2020-12-15 17:57	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 19:22	激活设备 宣看详情
linux-x86-zqw-2 💬	2020-12-15 16:24	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 19:42	激活设备 宣看详情
linux-x86-zqw 😳	2020-12-15 15:29	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-15 16:40	激活设备 宣看详情
firefly_rk3399pro 💬	2020-12-15 14:39	-		离线 ⑦	2020-12-15 21:33	激活设备 宣看详情
ubuntu-local-fjy 😳	2020-12-14 21:38	服务器	Linux-AMD64(x86-64)	在线	2021-01-07 15:01	宣看详情
hfl-1 💬	2020-12-14 19:24	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 14:18	激活设备 宣看详情
edge新增 💮	2020-12-11 14:52	-	-	未激活	-	删除 激活设备 查看详情

设备详情会展示当前设备的最新同步时间,以及CPU使用率、内存使用率等。服务列表则展示了当前设备上部署服务的运行情况和资源占用情况

端云协同服务 > 我的本地设备 > ubuntu-local-fjy								
设备详情 20多々称 ubuntu local fix	海绵绿木 左线		空中中国第6	DITE				
设备类型 服务器	应用平台 Linux	-AMD64(x86-64)	最新同步时间 2	2021-01-07 15:00				
CPU使用率 31.1%		存使用率 5.8%						
端云协同服务详情								
服务名称	模型ID	CPU占比	内存使用情况	内存占比	操作			
1000.007	246	0.01%	156.7MB	0.93%	查看服务配置			

⊙ 软硬一体方案部署

心如何获取图像分类软硬一体产品

为进一步提升前端智能计算的用户体验,EasyDL推出了多款软硬一体方案。将高性能硬件与EasyDL图像分类/物体检测模型深度适配,可应用于 工业分拣、视频监控等多种设备端离线计算场景,让离线AI落地更轻松。了解不同方案

方案获取流程如下:

Step 1:在EasyDL训练专项适配所选硬件的图像分类/物体检测模型,迭代模型至效果满足业务要求

训练配置						
*部署方式	公有云部署 Eas	RBB集要 syEdge本地部署 浏览器/小程序部署 如何选择部署方式?				
*选择设备	○服务器	通用小型设备 专项适配硬件				
选择硬件	• Edgeboard(FZ)	◯ Edgeboard(VMX) 辨影(Alr/Pro) <mark>NEW</mark>	Jetson(Nano/TX2/X	avier) 了解	释不同方案	
*选择算法	○ 高精度 ⑦ 〇	高性能 ⑦ (AI市场已购模型				
高级训练配置						
训练环境	名称	规格	算力	速度比例	价格	
	GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	免费	
	GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥0.36/分钟	(50小时*节点免费)
	GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥0.45/分钟	(53.28小时*节点免费)
训练费用 <mark>免费</mark>						
当前GPU V100卡#	◎余免费资源:53.28(小)	、时 * 节点) 为保证训练任务顺利进行,请保证您的账户余额充足,	可点击进行充值 立即	<mark>充值</mark> ,推荐购3	8.特惠小时包,可享	受至少8 <mark>折优惠</mark> 训练 前往购买
开始训练						

Step 2:发布模型时选择对应硬件

发布模型

选择模型:	test02	\sim	
部署方式:	专项硬件适配SDK	\sim	
选择硬件:	Jetson(Nano/TX2/Xavier)	\sim	了解更多
选择版本:	V2	~	
	提交		

Step 3:在AI市场购买方案获得硬件和用于激活专用SDK的专用序列号,参考文档集成后,即可实现离线AI预测

我的模型 > test02 > V1的专项硬件适配SDK服务详情

■ 专项硬件适配SDK获取 专项硬件适配SDK是EasyDL软硬一体方案的软件部分,建议在AI市场购买整体方案,同时获得专用 SDK激活序列号及专项适配硬件 如已在其他渠道购买硬件,可点击「获取序列号」前往控制台,支持申请专用的测试序列号、购买专 用的永久有效序列号

专项硬件适配SDK	操作
EdgeBoard(VMX)专用SDK	前往AI市场购买 下载SDK 🗸 管理序列号

如有其他硬件方案需求,请在百度智能云控制台内提交工单反馈。

心图像分类EdgeBoard(FZ)专用SDK集成文档

简介

本文档介绍 EasyEdge/EasyDL在EdgeBoard®边缘计算盒/Lite计算卡上的专用软件的使用流程。

EdgeBoard系列硬件可直接应用于AI项目研发与部署,具有高性能、易携带、通用性强、开发简单等四大优点。

详细硬件参数请在AI市场浏览。

EdgeBoard产品使用手册:https://ai.baidu.com/ai-doc/HWCE/Yk3b86gvp

软核版本

Baidu 百度智能云文档

CPP-SDK版本	对应软核
1.3.2、1.3.4、1.3.5	1.8.1
1.3.0、1.3.1、1.3.2、1.3.4	1.8
0.5.7-1.2.1	1.5
0.5.2+	1.4

SDK升级需配合EdgeBoard硬件软核升级,建议升级软核为SDK对应版本,否则可能出现结果错误或者其他异常。

可以通过dmesg | grep "DRIVER Version" 命令获取EdgeBoard当前的软核版本

Release Notes 注意*:升级完成相应的软核之后需要重启机器生效。

sdk对应的软核说明: 如果客户使用的软核是mobile版本的,需要使用1.4的SDK;如果不是mobile 版本,可以选择1.5+(目前最高版本更新至 1.8.1)版本的SDK使用。

1.5+版本的软核以及sdk更新情况如下表所示:

时间	版本	说明	EdgeBoard非mobile对应的软核以及特性	
2021.1 2.20	1.3.5	升预测引擎为PaddleLite 1.8.1,推理库支持了 Ubuntu18.04文件系统	https://ai.baidu.com/ai-doc/HWCE/Fkuqounlk(含有EB升级 Ubuntu18.04系统的步骤)	
2021.1 0.15	1.3.2、 1.3.4、1.3.4	推理库支持了Ubuntu18.04文件系统	https://ai.baidu.com/ai-doc/HWCE/Fkuqounlk	
2021.0 6.29	1.3.1	视频流解析支持分辨率调整;预测引擎升级;	https://ai.baidu.com/ai-doc/HWCE/Ikqgcqt5x	
2021.0 5.14	1.3.0	新增视频流接入支持;展示已发布模型性能评 估报告	https://ai.baidu.com/ai-doc/HWCE/Ikqgcqt5x	
2021.0 5.14	1.2.1	功能无更新	https://ai.baidu.com/ai-doc/HWCE/okqiwkm32	
2020.1 0.29	0.5.7	预测引擎切换为PaddleLite 1.5	-	
2019.1 2.27	0.4.5	引擎升级,支持zu5/zu3,支持EasyDL 高精度 检测模型	-	
2019.0 7.25	0.4.0	EdgeBoard SDK Release!	-	

mobile软核以及sdk更新情况如下表所示: | 时间 | 版本 | 说明 |EdgeBoard mobile对应的软核以及特性 | | --- | ---- | ---- | | 2021.05.14|1.2.1|功能无更新|https://ai.baidu.com/ai-doc/HWCE/okqiwkm32|https://ai.baidu.com/ai-doc/HWCE/Lkqiwlziw||

快速开始

开发者从EasyEdge/EasyDL下载的软件部署包中,包含了简单易用的SDK和Demo。只需简单的几个步骤,即可快速部署运行EdgeBoard计算盒。

部署包中包含多版本SDK:

• baidueasyedge_linux_cpp_aarch64_EdgeBoardFZ1.8*:适用于EdgeBoard 1.5+软核

• baidu easyedge_linux_cpp_aarch64_EdgeBoardFZ1.4*:适用于EdgeBoard 1.4软核

SDK文件结构
bai	idu_easyedge_linux_cpp_aarch64_EdgeBoardFZ1.5_*
-	ReadMe.txt
\vdash	— bin
	easyedge_image_inference
	easyedge_serving
	easyedge_video_inference
\vdash	include
	Leasyedge
\vdash	— lib
	libeasyedge.so -> libeasyedge.so.1
	libeasyedge.so.1 -> libeasyedge.so.1.3.1
	libeasyedge.so.1.3.1
	libeasyedge_static.a
	libeasyedge_videoio.so -> libeasyedge_videoio.so.1
	libeasyedge_videoio.so.1 -> libeasyedge_videoio.so.1.3.1
	libeasyedge_videoio.so.1.3.1
	libeasyedge_videoio_static.a
	libpaddle_full_api_shared.so -> libpaddle_full_api_shared.so.1.8.0
	libpaddle_full_api_shared.so.1.8.0
	libverify.so -> libverify.so.1
	libverify.so.1 -> libverify.so.1.0.0
	Libverify.so.1.0.0
-	now_sre.log
-	src
	CMakeLists.txt
	cmake
	├─── common
	demo_image_inference
	demo_serving
	demo_video_inference
	thirdparty
1	opencv

1.1.0+的SDK自带OpenCV, src编译的时候会引用thirdparty/opencv路径下的头文件和库文件。

Demo使用流程

用户在AI市场购买计算盒之后,请参考以下步骤进行集成和试用。

1. 将计算盒连接电源

指示灯亮起,等待约1分钟。

- 参考EdgeBoard使用文档配置网口或串口连接。登录EdgeBoard计算盒。
- 加载驱动 (开机加载一次即可)。

insmod /home/root/workspace/driver/{zu9|zu5|zu3}/fpgadrv.ko

根据购买的版本,选择合适的驱动。若未加载驱动,可能报错:

Failed to to fpga device: -1

• 设置系统时间 (系统时间必须正确)

date --set "2019-5-18 20:48:00"

2. (可选) 启动HTTP服务

部署包中附带了HTTP服务功能,开发者可以进入SDK根目录,运行easyedge_serving程序启动HTTP服务。

```
**./easyedge_serving {RES目录} "" {绑定的host,默认0.0.0.} {绑定的端口,默认24401}**
cd ${SDK_ROOT}
export LD_LIBRARY_PATH=./lib
./demo/easyedge_serving ../../../RES ""
```

日志显示

2019-07-18 13:27:05,941 INFO [EasyEdge] [http_server.cpp:136] 547974369280 Serving at 0.0.0.0:24401

则启动成功。此时可直接在浏览器中输入http://{EdgeBoard计算盒ip地址}:24401/,在h5中测试模型效果。



同时,可以调用HTTP接口来访问盒子。具体参考下文接口说明。

EdgeBoard HTTP Server 目前使用的是单线程处理请求。

3. 编译运行Demo

编译:

```
cd src
mkdir build && cd build
cmake .. && make
```

运行

./easyedge_image_inference {RES资源文件夹路径} {测试图片路径}

便可看到识别结果。

使用说明

使用流程

激活成功之后,有效期内可离线使用。

- 1. 配置PaddleFluidConfig
- 2. 新建Predictor :global_controller()->CreateEdgePredictor(config);
- 3. 初始化 predictor->init()
- 4. 传入图片开始识别predictor->infer(img, ...);

目前EdgeBoard暂不支持并行多模型计算。

接口说明

设置序列号 请在网页控制台中申请序列号,并在init初始化前设置。

EdgePredictorConfig config;

 $config.set_config(easyedge::params::PREDICTOR_KEY_SERIAL_NUM, \ "this-is-serial-num");$

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己 的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的变量设置),需要设置额外的环境变量,指定CONTROLLER_KEY_AUTH_MODE为2, global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2),实例数鉴权模式下还支持指定license证书更新时间,单 位是秒,要求设置为大于20的整数,否则会采用默认的license更新时间,修改实例数鉴权license更新时间的方法参考 global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 300)

预测图片

```
/**
* @brief 同步预测接口
* inference synchronous
* Supported by most chip and engine
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @param threshold
* @return
*/
virtual int infer(
    cv::Mat &image, std::vector<EdgeResultData> &result, float threshold = 0.1
) = 0;
```

识别结果说明

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index; // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // object detection field
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
};
```

关于矩形坐标

```
x1 * 图片宽度 = 检测框的左上角的横坐标
```

y1 * 图片高度 = 检测框的左上角的纵坐标

```
x2 * 图片宽度 = 检测框的右下角的横坐标
```

y2 * 图片高度 = 检测框的右下角的纵坐标

可以参考demo文件中使用opencv绘制矩形的逻辑。

HTTP 私有服务请求说明

http 请求参数

URL中的get参数:

参数	说明		默认值
threshold	阈值过滤,	0~1	0.1

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

Python

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
    'http://127.0.0.1:24401/',
    params={'threshold': 0.1},
    data=img).json()
```

Cpp label=C#

FileStream fs = new FileStream("./img.jpg", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close();

WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close(); response.Close();

Cpp label=C++ 需要安装curl

```
#include <sys/stat.h>
 #include <curl/curl.h>
 #include <iostream>
 #include <string>
 #define S_ISREG(m) (((m) & 0170000) == (0100000))
 #define S_ISDIR(m) (((m) & 0170000) == (0040000))
 size_t write_callback(void *ptr, size_t size, size_t num, void *data) {
    std::string *str = dynamic_cast<std::string *>((std::string *)data);
    str->append((char *)ptr, size*num);
    return size*num;
 }
 int main(int argc, char *argv[]) {
    const char *post_data_filename = "./img.jpg";
    FILE *fp = NULL;
    std::string response;
    struct stat stbuf = { 0, };
    fp = fopen(post_data_filename, "rb");
    if (!fp) {
      fprintf(stderr, "Error: failed to open file "%s"
 ", post_data_filename);
      return -1;
    }
    if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
      fprintf(stderr, "Error: unknown file size "%s"
 ", post_data_filename);
      return -1;
    }
    CURL *curl;
    CURLcode res;
    curl_global_init(CURL_GLOBAL_ALL);
    curl = curl_easy_init();
    if (curl != NULL) {
      curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
      curl_easy_setopt(curl, CURLOPT_POST, 1L);
      curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
      curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
  curl_easy_setopt(curl, CURLOPT_WRITEFUNCTION, write_callback);
      curl_easy_setopt(curl, CURLOPT_WRITEDATA, &response);
      res = curl_easy_perform(curl);
      if (res != CURLE_OK) {
         fprintf(stderr, "curl_easy_perform() failed: %s
 ", curl_easy_strerror(res));
      }
      std::cout << response << std::endl; // response即为返回的json数据
      curl_easy_cleanup(curl);
    }
    curl_global_cleanup();
    fclose(fp);
    return 0;
 }
```

Java请求示例

http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考预测图像-返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

```
{
  "cost_ms": 52,
  "error_code": 0,
  "results": [
   {
      "confidence": 0.94482421875,
      "index": 1,
      "label": "IronMan",
      "x1": 0.059185408055782318,
      "x2": 0.18795496225357056,
      "y1": 0.14762254059314728,
      "y2": 0.52510076761245728
    },
    {
      "confidence": 0.94091796875,
      "index": 1,
      "label": "IronMan",
      "x1": 0.79151463508605957,
      "x2": 0.92310667037963867,
      "y1": 0.045728668570518494,
      "y2": 0.42920106649398804
}
```

预测视频

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

● 接口

classVideoDecoding :

/**

- * @brief 获取输入源的下一帧
- * @param frame_tensor
- * @return

*/

virtual int next(FrameTensor &frame_tensor) = 0;

/**

- * @brief 显示当前frame_tensor中的视频帧
- * @param frame_tensor
- * @return
- */

virtual int display(const FrameTensor &frame_tensor) = 0;

/**

* @brief 将当前frame_tensor中的视频帧写为本地视频文件

- * @param frame_tensor
- * @return

virtual int save(FrameTensor & frame_tensor) = 0;

```
/**
```

*/

```
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
* @brief 获取视频的width属性
* @return
*/
virtual int get_width() = 0;
/**
```

```
* @brief 获取视频的height属性
* @return
*/
virtual int get_height() = 0;
```

```
struct VideoConfig
```

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type; // 输入源类型
  std::string source_value;   // 输入源地址,如视频文件路径、摄像头index、网络流地址
 int skip_frames{0};
                      // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
  int retrieve_all{false};
                     // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
                      // 在采取抽帧之前设置视频的fps
  int input_fps{0};
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
 bool enable_display{false};
  std::string window_name{"EasyEdge"};
  bool display_all{false};
                       // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                       // 是否存储所有frame, 若为false, 仅存储根据skip_frames抽取的frame
  std::map::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被

Baidu 百度智能云文档

抽取返回,以作为显示或存储用。 input_fps:用于抽帧前设置fps。 resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。 conf:高级选项。部分配置会通过该map来设置。

注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

错误说明

SDK所有主动报出的错误,均覆盖在EdgeStatus枚举中。同时SDK会有详细的错误日志,开发者可以打开Debug日志查看额外说明:

EdgeLogConfig log_config; log_config.enable_debug = true; global_controller()->set_log_config(log_config);

FAQ

1. 如何处理一些 undefined reference?

如: undefined reference to `curl_easy_setopt@CURL_OPENSSL_3'

可以通过安装libcurl3 libcurl-openssl1.0-dev来解决。如果开发者想不想使用低版本的openssl(如Ubuntu 18.04),可以link静态 库easyedge_static.a,自己指定需要的Library的版本。

示例:修改CMakeList.txt

```
find_package(CURL REQUIRED)
target_link_libraries(easyedge_demo ${OpenCV_LIBS} easyedge_static pthread ${CURL_LIBRARIES} paddle-mobile)
```

2. error while loading shared libraries: libeasyedge.so.0.4.0: cannot open shared object file: No such file or directory

类似错误包括libpaddle-mobile.so找不到。

直接运行SDK自带的二进制可能会有这个问题,设置LD_LIBRARY_PATH为SDK部署包中的lib目录即可。开发者自行使用CMake编译的二进制可以 有效管理.so的依赖。

3. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

4. 预测过程中报内存不足"Killed"

此问题仅出现在ZU5,因为FZ5A带vcu,给他预留的内存过大导致,如果用不到VCU可以把这部分改小。修改/run/media/mmcblk1p1/uEnv.txt:

ethaddr=00:0a:35:00:00:09 uenvcmd=fatload mmc 1 0x3000000 image.ub && bootm 0x3000000

bootargs=earlycon console=ttyPS0,115200 clk_ignore_unused cpuidle.off=1 root=/dev/mmcblk1p2 rw rootwait cma=128M

注意中间空行要保留。

5. 预测结果异常

如果购买的计算盒较早,驱动文件较旧,而SDK比较新(或SDK比较旧,但是计算盒较新),可能出现结果异常,如结果均为空或者 nan。 请参

考"软核版本"小节更新软核和驱动版本。

6. 编译过程报错file format not recognized

libeasyedge.so: file format not recognized; treating as linker script

下载的SDK zip包需要放到板子内部后,再解压、编译。

7. 提示 driver_version(1.4.0) not match paddle_lite_version(1.5.1) 需更新驱动,否则可能导致结果异常。参考"软核版本"小节。

心图像分类EdgeBoard(VMX)专用SDK集成文档

简介

本文档旨在介绍 EasyDL在EdgeBoard USB加速卡VMX(以下简称VMX加速卡或加速卡)上的专用软件的使用流程。 EdgeBoard系列硬件适用于项 目研发与部署,具有高性能、易携带、通用性强、开发简单等四大优点。 您可在AI市场了解EdgeBoard相关系列产品,同时可以在软硬一体方 案了解性能数据。

注意:本型号主要面向产品集成和企业项目,未同时售卖散热片和外壳,部分情况下芯片温度较高,<mark>开发过程中,请勿用手触摸,谨防烫伤</mark>

硬件介绍

VMX加速卡,采用Intel® Movidus [™] 视觉 MyriadX处理器芯片,通过 USB3.0 通讯type-c接口方式,配合外围电路即可将该模组嵌入到第三方智能 化产品中,采用标准 USB通讯协议,对接简单,开发速度快,具有强大的深度学习计算功能。可通过OpenVINO [™]和OpenCV软件库工具链移植算 法,兼容百度PaddlePaddle支持Paddle2onnx和PaddleHub并集成EasyDL,使产品应用范围广,性能更稳定,增强用户体验。

VMX加速卡适用于深度学习加速,能够解决复杂的人工智能软硬件设计挑战,它可以集成基于视觉的加速器和推理引擎来实现深度边缘学习的解决方案。(3D/2D人脸识别、人头检测、人脸属性分析(性别、年龄)、人脸特征比对、手势及姿态识别、物体检测及分类、算法移植等功能。)

硬件配置与说明

核心板模块: Intel® Movidus[™] MyriadX,内置内存LP-DDR4 4GBit。

• 硬件指标

CPU

o Intel® Movidius Myriad X MA2485 Vision Processing Unit

o Total performance of over 4 trillion operations per second (TOPS) o Over 1 TOPS performance on neural network inference w/ NCE accelerator o 16 Programmable 128-bit VLIW Vector Processors o 16 Configurable MIPI Lanes w/ enhanced Vision Accelerators o 2.5 MB of Homogenous On-Chip Memory w/ 4Gbit LPDDR4

Size

o 38mm x 38mm

Interface o USB TYPE C (USB3.0) 辅助接口精简设计

Boot o USB 启动模式 - 内置 switch 缺省模式设置

Power o 平均功耗0.5W~2.2W

Security o 支持 eFuse 加密

运行说明

VMX加速卡包含独立的AI运算芯片,采用 USB Type-C通讯方式,通讯协议简单可靠,可连接不同芯片架构主机,包括 X86、ARM SOC等。加速卡运行需要通过TypeC接口连接宿主机执行,宿主机目前支持的软硬件环境包括:

- Linux: x86-64, armv7hf
- Windows: x86-64, Windows 10

使用过程中,请尽量避免直接触碰板卡元器件;或者使用防静电锡纸包裹板卡。

Baidu 百度智能云文档

开发者从EasyDL训练模型之后,下载的软件部署包中,包含了简单易用的SDK和Demo。只需简单的几个步骤,即可快速部署运行。

Release Notes

Python SDK

时间	版本	说明
2020.12.18	1.2.0	性能优化;接口优化升级;推理引擎升级
2020.09.17	1.1.19	支持更多模型与平台

Python SDK适用于Linux x86-64和Windows平台。

2020-12-18: 【接口升级】 Python SDK序列号配置接口从1.2.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

C++ SDK

时间	版本	说明
2021.06.29	1.3.1	视频流解析支持分辨率调整
2021.05.14	1.3.0	新增视频流接入支持;展示已发布模型性能评估报告
2020.12.18	1.0.0	性能优化;接口优化升级;推理引擎升级
2020.09.17	0.5.6	新增C++ SDK,支持Linux armv7hf(树莓派)架构的硬件接入VMX预测

C++ SDK适用于Linux x86-64、Linux armv7hf平台。

2020-12-18: 【接口升级】 参数配置接口从1.0.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

将加速卡连接宿主机 请使用质量合规的usb线连接。连接之后,检查设备是否被操作系统识别: Linux 通过Isusb -v 命令检查是否有 Myriad设 备:

> sudo lsusb -v	grep -C 5 Myriad
bMaxPacketSize	e0 64
idVendor	0x03e7
idProduct	0x2485
bcdDevice	0.01
iManufacturer	1 Movidius Ltd.
iProduct	2 Movidius Myriad
iSerial	3 03e72485

Windows 可以在设备管理器中查询。

如果使用 VirtualBox 之类的虚拟机,请在虚拟机加入 03e7:24 和 03e7:f63b 两个 usb 设备。

获取并安装依赖

1) 安装依赖

宿主机与sdk为以下情况: 1**□** Windows x86-64: 请参考 OpenVINO toolkit 文档安装 2020.3.1LTS 版本 2**□** Linux x86-64且使用Python SDK时 必须: 请参考 OpenVINO toolkit 文档安装 2020.3.1LTS 版本, 安装时可忽略Configure the Model Optimizer及后续部分。

安装完毕,运行之前,请按照OpenVino的文档设置环境变量

source /opt/intel/openvino/bin/setupvars.sh

获取序列号 从AI市场订单详情或者EasyDL控制台获取序列号。

更换序列号、更换设备时,首次使用需要联网激活。激活成功之后,有效期内可离线使用。

请确保激活设备时使用的 操作系统账号与后续使用时运行的账号一致,否则会造成验证失败

Python SDK

1. 安装wheel包

pip3 install -U BaiduAl_EasyEdge_SDK-{版本号}-cp37-cp37m-linux_x86_64.whl

注意,请根据python的版本选择对应的whl文件,其中,1.2.0是SDK版本号,cp37表示是python3.7版本

注意,pip安装时请添加-U参数

2. 将步骤2中获得的序列号 填入demo.py

```
pred = edge.Program()
pred.set_auth_license_key("这里填写序列号")
```

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的设置),需要调用函数指定实例数鉴权模式,并且实例数鉴权模式下,支持指定 license证书更新时间,单位是秒,要求设置为大于20的整数,否则会采用默认的license更新时间,参考

```
pred.set_instance_auth_mode()
pred.set_instance_update_interval(200)
```

3. 测试demo.py

python3 demo.py {模型资源文件夹RES路径} {待识别的图片路径}

生成的样例结果图片如下:



使用流程

import BaiduAI.EasyEdge as edge

pred = edge.Program() pred.set_auth_license_key("这里填写序列号") pred.init(model_dir={RES文件夹路径}, device=edge.Device.MOVIDIUS, engine=edge.Engine.OPENVINO) pred.infer_image({numpy.ndarray的图片}) pred.close()

接口的详细说明请主要参考 SDK 中的接口注释

接口说明

Program

● 初始化

def init(self,

model_dir, device=Device.CPU, engine=Engine.NCSDK, config_file='conf.json', preprocess_file='preprocess_args.json', model_file='model', params_file='params', graph_file='graph.ncsmodel', label_file='label_list.txt', device_id=0, **kwargs): Args: model_dir: str device: BaiduAI.EasyEdge.Device engine: BaiduAI.EasyEdge.Engine preprocess_file: str model_file: str params_file: str graph_file: str ncs的模型文件 或 PaddleV2的模型文件 label_file: str device_id: int 设备ID thread_num: int CPU的线程数

Raises:

RuntimeError, IOError Returns: bool: True if success

```
....
```

• 预测单张图像

```
def infer_image(self, img, threshold=None,
        channel_order='HWC',
        color_format='BGR',
        data_type='numpy'
        ):
 ....
 Args:
   img: np.ndarray or bytes
   channel_order(string):
     channel order: HWC or CHW
   color_format(string):
     color format order: RGB or BGR
   threshold(float):
      only return result with confidence larger than threshold
   data_type(string): 仅在图像分割时有意义。 'numpy' or 'string'
      'numpy': 返回已解析的mask
      'string': 返回未解析的mask游程编码
 Returns:
   list
 ....
```

• 返回格式: [dict1, dict2, ...]

Baidu 百度智能云文档

字段	类型	取值	说明
confidence	float	0~1	分类或检测的置信度
label	string		分类或检测的类别
index	number		分类或检测的类别
x1, y1	float	0~1	物体检测,矩形的左上角坐标 (相对长宽的比例值)
x2, y2	float	0~1	物体检测,矩形的右下角坐标(相对长宽的比例值)
mask	string/numpy.ndarray	图像分割的mask	

关于矩形坐标

x1 * 图片宽度 = 检测框的左上角的横坐标

y1 * 图片高度 = 检测框的左上角的纵坐标

x2*图片宽度 = 检测框的右下角的横坐标

y2 * 图片高度 = 检测框的右下角的纵坐标

可以参考 demo 文件中使用 opencv 绘制矩形的逻辑。

结果示例

● i) 图像分类

```
{
    "index": 736,
    "label": "table",
    "confidence": 0.9
}
```

● ii) 物体检测

```
{
    "y2": 0.91211,
    "label": "cat",
    "confidence": 1.0,
    "x2": 0.91504,
    "index": 8,
    "y1": 0.12671,
    "x1": 0.21289
}
```

• iii) 图像分割

```
{
    "name": "cat",
    "score": 1.0,
    "location": {
        "left": ...,
        "top": ...,
        "width": ...,
        "height": ...,
    },
    "mask": ...
}
```

mask字段中,data_type为numpy时,返回图像掩码的二维数组

{

```
{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0},
{0, 0, 0, 1, 1, 1, 0, 0, 0, 0},
{0, 0, 0, 1, 1, 1, 0, 0, 0, 0},
{0, 0, 0, 1, 1, 1, 0, 0, 0, 0},
{0, 0, 0, 1, 1, 1, 0, 0, 0, 0},
{0, 0, 0, 0, 0, 0, 0, 0, 0, 0},
}
其中1代表为目标区域,0代表非目标区域
```

data_type为string时,mask的游程编码,解析方式可参考 demo

C++ SDK

使用说明

模型资源文件默认已经打包在开发者下载的SDK包中。请先将SDK包整体拷贝到具体运行的宿主机设备中,再解压缩编译;

在编译或运行demo程序前执行以下命令:

source \${cpp_kit位置路径}/thirdparty/openvino/bin/setupvars.sh

如果openvino预测引擎找不到设备需要执行以下命令:

sudo cp \${cpp_kit位置路径}/thirdparty/openvino/deployment_tools/inference_engine/external/97-myriad-usbboot.rules /etc/udev/rules.d/ sudo udevadm control --reload-rules sudo udevadm trigger sudo ldconfig ```

使用流程

```
// step 1: 配置运行参数
  EdgePredictorConfig config;
 config.set_config(easyedge::params::PREDICTOR_KEY_SERIAL_NUM, "this-is-serial-num"); // 设置序列号
 config.model_dir = {模型文件目录};
  // step 2: 创建并初始化Predictor
 auto predictor = global_controller()->CreateEdgePredictor(config);
 if (predictor->init() != EDGE_OK) {
    exit(-1);
  // step 3-1: 预测图像
 auto img = cv::imread({图片路径});
 std::vector<EdgeResultData> results;
 predictor->infer(img, results);
// step 3-2: 预测视频
std::vector<EdgeResultData> results;
FrameTensor frame_tensor;
 VideoConfig video_config;
 video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h
 video_config.source_value = video_src;
  ... more video_configs, 根据需要配置video_config的各选项
  */
 auto video_decoding = CreateVideoDecoding(video_config);
  while (video_decoding->next(frame_tensor) == EDGE_OK) {
    results.clear();
    if \; (frame\_tensor.is\_needed) \; \{
      predictor->infer(frame_tensor.frame, results);
      render(frame_tensor.frame, results, predictor->model_info().kind);
    //video_decoding->display(frame_tensor); // 显示当前frame,需在video_config中开启配置
     //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置
```

}

运行参数配置 运行参数的配置通过结构体 EdgePredictorConfig 完成,其定义如下所示:

```
struct EdgePredictorConfig {
 /**
  * @brief 模型资源文件夹路径
 */
 std::string model_dir;
 std::map<std::string, std::string> conf;
 EdgePredictorConfig();
 template<typename T>
 T get_config(const std::string &key, const T &default_value);
 template<typename T = std::string>
 T get_config(const std::string &key);
 template<typename T>
 const T *get_config(const std::string &key, const T *default_value);
 template<typename T>
 void set_config(const std::string &key, const T &value);
 template<typename T>
 void set_config(const std::string &key, const T *value);
 static EdgePredictorConfig default_config();
```

```
};
```

运行参数选项的配置以key、value的方式存储在类型为std::map的conf中,并且键值对的设置和获取可以通

过EdgePredictorConfig的set_config和get_config函数完成。同时部分参数也支持以环境变量的方式设置键值对。 EdgePredictorConfig的具体使用 方法可以参考开发工具包中的demo工程。

具体支持的运行参数可以参考开发工具包中的头文件。

初始化

接口

auto predictor = global_controller()->CreateEdgePredictor(config); predictor->init();

若返回非0,请查看输出日志排查错误原因。

预测图像

```
● 接口
```

/** * @brief

- * 通用接口
- * @param image: must be BGR , HWC format (opencv default)
- * @param result
- * @return
- */

virtual int infer(

cv::Mat& image, std::vector<EdgeResultData>& result

) = 0;

图片的格式务必为opencv默认的BGR, HWC格式。

• 返回格式

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index; // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测活图像分割时才有
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割时才有
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding, 游程编码的mask
```

};

关于矩形坐标

x1 * 图片宽度 = 检测框的左上角的横坐标

- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2*图片宽度 = 检测框的右下角的横坐标
- y2 * 图片高度 = 检测框的右下角的纵坐标

```
关于图像分割mask
```

```
cv::Mat mask为图像掩码的二维数组
{
{0,0,0,0,0,0,0,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,0,0,0,0,0,0,0},
}
其中1代表为目标区域,0代表非目标区域
```

关于图像分割mask_rle

该字段返回了mask的游程编码,解析方式可参考 http demo

以上字段可以参考demo文件中使用opencv绘制的逻辑进行解析

预测视频

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

接口

classVideoDecoding :

/**

- * @brief 获取输入源的下一帧
- * @param frame_tensor
- * @return

*/

virtual int next(FrameTensor &frame_tensor) = 0;

/**

- * @brief 显示当前frame_tensor中的视频帧
- * @param frame_tensor
- * @return
- */

virtual int display(const FrameTensor &frame_tensor) = 0;

/**

* @brief 将当前frame_tensor中的视频帧写为本地视频文件

- * @param frame_tensor
- * @return

virtual int save(FrameTensor & frame_tensor) = 0;

```
/**
```

*/

```
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
* @brief 获取视频的width属性
* @return
*/
virtual int get_width() = 0;
/**
```

```
* @brief 获取视频的height属性
* @return
*/
virtual int get_height() = 0;
```

struct VideoConfig

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type; // 输入源类型
  std::string source_value;   // 输入源地址,如视频文件路径、摄像头index、网络流地址
 int skip_frames{0};
                      // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
  int retrieve_all{false};
                     // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
                      // 在采取抽帧之前设置视频的fps
  int input_fps{0};
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
 bool enable_display{false};
  std::string window_name{"EasyEdge"};
  bool display_all{false};
                       // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                       // 是否存储所有frame, 若为false, 仅存储根据skip_frames抽取的frame
  std::map::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被

Baidu 百度智能云文档

抽取返回,以作为显示或存储用。 input_fps:用于抽帧前设置fps。 resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。 conf:高级选项。部分配置会通过该map来设置。

注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

设置序列号

请在网页控制台中申请序列号,并在init初始化前设置。 LinuxSDK 首次使用需联网授权。

EdgePredictorConfig config; config.set_config(easyedge::params::PREDICTOR_KEY_SERIAL_NUM, "this-is-serial-num");

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己 的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的变量设置),需要设置额外的环境变量,指定CONTROLLER_KEY_AUTH_MODE为2, global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2),实例数鉴权模式下还支持指定license证书更新时间,单 位是秒,要求设置为大于20的整数,否则会采用默认的license更新时间,修改实例数鉴权license更新时间的方法参考 global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 300)

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

EdgeLogConfig log_config; log_config.enable_debug = true; global_controller()->set_log_config(log_config);

http服务

1. 开启http服务 http服务的启动参考demo_serving.cpp文件。

```
/**
 * @brief 开启一个简单的demo http服务。
 * 该方法会block直到收到sigint/sigterm。
 * http服务里,图片的解码运行在cpu之上,可能会降低推理速度。
 * @tparam ConfigT
 * @param config
 * @param host
 * @param port
 * @param service_id service_id user parameter, uri '/get/service_id' will respond this value with 'text/plain'
 * @param instance_num 实例数量,根据内存/显存/时延要求调整
 * @return
 */
template<typename ConfigT>
int start_http_server(
     const ConfigT & config,
     const std::string &host
     int port.
     const std::string &service_id
     int instance_num = 1);
```

2. 请求http服务

开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。

URL中的get参数:

参数	说明	默认值	
threshold	阈值过滤,0~2	山如不提供,则会使用模型的推荐阈值	

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

```
Python请求示例
```

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
        'http://127.0.0.1:24401/',
        params={'threshold': 0.1},
        data=img).json()
```

Java请求示例

• http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考预测图像·返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

返回示例

```
"cost_ms": 52,
  "error_code": 0,
  "results": [
    {
       "confidence": 0.94482421875,
       "index": 1,
       "label": "IronMan",
       "x1": 0.059185408055782318,
       "x2": 0.18795496225357056,
       "y1": 0.14762254059314728,
       "y2": 0.52510076761245728
    },
       "confidence": 0.94091796875,
      "index": 1,
       "label": "IronMan",
       "x1": 0.79151463508605957,
       "x2": 0.92310667037963867,
       "y1": 0.045728668570518494,
       "y2": 0.42920106649398804
    }
}
```

其他配置

• 1. 日志名称、HTTP 网页标题设置

通过global_controller的set_config方法设置:

 $global_controller()->set_config(easyedge::params::KEY_LOG_BRAND, "MY_BRAND");$

效果如下:

图片加载失败

Linux FAQ

1. EasyDL 离线 SDK 与云服务效果不一致,如何处理?我们会逐渐消除这部分差异,如果开发者发现差异较大,可通过工单、论坛联系我们协助处理。

2. 硬件出现问题或者出现故障怎么办? 软件使用有问题怎么处理?

- 如果持续在静电较多的环境中使用,建议使用防静电锡纸包裹板卡
- 如果硬件无法启动等故障,您可以通过商品页联系供应商处理;其它硬件问题,您可以邮件 edgeboard-vmx.com ,我们将在0-2日内处理您的问题。为加快处理进度,您在邮件中,尽量描述清楚问题或者需求细节,避免来回沟通。
- 软件使用问题,请尽量通过工单、论坛联系我们协助处理。

3. 运行时报错:NC_ERROR

Can not init Myriad device: NC_ERROR

一般是硬件没有插上,请确保Isusb能够找到该硬件。或者等待几秒后再试。

快速开始 Windows

1. 安装依赖

将操作系统升级到Windows 10

安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642

Visual C++ Redistributable Packages for Visual Studio 2013

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

Visual C++ Redistributable Packages for Visual Studio 2015

https://www.microsoft.com/zh-cn/download/details.aspx?id=48145

注意事项

- 1. 安装目录不能包含中文
- 2. Windows Server 请自行开启,选择"我的电脑"——"属性"——"管理"——"添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启即 可。

2. 运行离线SDK

解压下载好的SDK,打开EasyEdge.exe,输入Serial Num



图片加载失败

点击"启动服务",等待数秒即可启动成功,本地服务

默认运行在

http://127.0.0.1:24401/

其他任何语言只需通过HTTP调用即可。

接口调用说明

Python 使用示例代码如下

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
```

params 为GET参数 data 为POST Body result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1}, data=img).json()

C## 使用示例代码如下

FileStream fs = new FileStream("./img.jpg", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close();

WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close(); response.Close();

请求参数

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送图片二进制。

返回参数|字段|类型|取值|说明||--------|----|-----|-----|| confidence | float | 0~1 | 分类或检测的置信度 || label | string || 分 类或检测的类别 || index | number || 分类或检测的类别 || x1, y1 | float | 0~1 | 物体检测,矩形的左上角坐标 (相对长宽的比例值) || x2, y2 | float | 0~1 | 物体检测,矩形的右下角坐标 (相对长宽的比例值) |

关于矩形坐标

- x1 * 图片宽度 = 检测框的左上角的横坐标
- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2*图片宽度 = 检测框的右下角的横坐标

y2 * 图片高度 = 检测框的右下角的纵坐标

Windows FAQ

1. 服务启动失败,怎么处理?

请确保相关依赖都安装正确,版本必须如下: .NET Framework 4.5 Visual C++ Redistributable Packages for Visual Studio 2013 * Visual C++ Redistributable Packages for Visual Studio 2015

2. 服务调用时返回为空,怎么处理?调用输入的图片必须是RGB格式,请确认是否有alpha通道。

3. 多个模型怎么同时使用? SDK设置运行不同的端口,点击运行即可。

4. JAVA、C#等其他语言怎么调用SDK?

参考 http://ai.baidu.com/forum/topic/show/943765

5. 启动失败,缺失DLL? 打开EasyEdge.log,查看日志错误,根据提示处理 缺失DLL,请使用 http://www.dependencywalker.com/ 查看相应模块 依赖DLL缺失哪些,请自行下载安装

6. 启动失败,报错NotDecrypted? Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

7. 其他问题 如果无法解决,可到论坛发帖: http://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

心图像分类Jetson专用SDK集成文档

简介

本文档介绍EasyEdge/EasyDL的Jetson SDK的使用方法。Jetson SDK支持的硬件包括Jetson nano, Jetson TX2, Jetson AGX Xavier和Jetson Xavier NX。您可在AI市场了解Jetson相关系列产品,同时可以在软硬一体方案了解部署方案。

模型支持:

- EasyDL图像:图像分类高精度,图像分类高性能,物体检测高精度,物体检测均衡,物体检测高性能,目标跟踪单标签模型。
- BML :
 - 公开数据集预训练模型:SSD-MobileNetV1,YOLOv3-DarkNet,YOLOv3-MobileNetV1,ResNet50,ResNet101,SE-ResNeXt50,SE-ResNeXt101,MobileNetV2,EfficientNetB0_small,EfficientNetB4,MobileNetV3_large_x1_0,ResNet18_vd,SE_ResNet18_vd,Xception71。
 - 百度超大规模数据集预训练模型:YOLOv3-DarkNet, MobileNetV3_large_x1_0, ResNet50_vd, ResNet101_vd。
- EasyEdge: EasyEdge支持的模型较多,详见查看模型网络适配硬件。若模型不在此列表,可以尝试使用自定义网络生成端计算组件。

软件版本支持使用EasyDL的Jetson系列SDK需要安装指定版本的JetPack和相关组件。所支持的JetPack版本会随着SDK版本的升级和新版本 JetPack的推出而不断的更新。在使用SDK前请务必保证软件版本满足此处声明版本。目前所支持的JetPack版本包括:

- JetPack5.0.2
- JetPack5.0.1
- JetPack4.6
- JetPack4.5
- JetPack4.4 (deprecated,该版本SDK会在未来某个版本移除,请切换至新版本JetPack)
- JetPack4.2.2 (已移除,请切换至新版本JetPack)

安装JetPack时请务必安装对应的组件:

- 使用SDK Manager安装JetPack需要勾选TensorRT、OpenCV、CUDA、cuDNN等选项。
- 使用SD Card Image方式(仅对Jetson Nano和Jetson Xavier NX有效)则无需关心组件问题,默认会全部安装。

Release Notes | 时间 | 版本 | 说明 | | --- | --- | | 2022.12.29 | 1.7.2 | 新增支持JetPack5.0.2;缓存机制优化;模型性能优化 || 2022.07.28 | 1.6.0 | 新增支持JetPack5.0.1,新增目标追踪接入实时流的demo || 2022.05.18 | 1.5.0 | 部分模型切换格式,max_batch_size含义变更,由输入 图片数不大于该值变更为等于该值;移除适用于JetPack4.2.2的SDK;示例代码demo_stream_inference重构;示例代码移除frame_buffer,新增 更安全高效的safe_queue || 2021.12.22 | 1.3.5 | 新增支持JetPack4.6;支持在EasyEdge平台语义分割模型生成开发套件;修复缓存问题;支持 自定义缓存路径 || 2021.10.20 | 1.3.4 | 新增支持JetPack4.5;大幅提升EasyDL有损压缩加速模型的推理速度 || 2021.06.29 | 1.3.1 | 视频流支 持分辨率调整;支持将预测后的视频推流,新增推流demo || 2021.05.13 | 1.3.0 | 新增视频流接入支持;EasyDL模型发布新增多种加速方案选 择;目标追踪支持x86平台的CPU、GPU加速版;展示已发布模型性能评估报告 || 2021.03.09 | 1.2.1 | EasyEdge新增一系列模型的支持;性能 优化 || 2021.01.27 | 1.1.0 | EasyDL经典版高性能分类模型升级;

EasyDL经典版检测模型新增均衡选项;

EasyEdge平台新增Jetson系列端计算组件的生成;

问题修复||2020.12.18|1.0.0|接口升级和一些性能优化||2020.08.11|0.5.5|部分模型预测速度提升||2020.06.23|0.5.4|支持 JetPack4.4DP,支持EasyDL专业版更多模型||2020.05.15|0.5.3|专项硬件适配SDK支持Jetson系列|

2022-5-18: 【接口变更】 PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE含义变更。 变更前:预测输入图片数不大于该值均可。 变更后:预 测输入图片数需等于该值。SDK内部对该接口变更做了兼容处理,在输入图片数小于该值时依然可正常运行,但预测性能会和等于该值时一 致。推荐根据实际输入图片数量需求修改该值,尽可能保持最小。 【版本移除】 适用于JetPack4.4版本的SDK被标记为deprecated, SDK会 在未来某个版本移除,建议切换至最新版本JetPack。 适用于JetPack4.2.2版本的SDK被移除。

2020-12-18: 【接口升级】 参数配置接口从1.0.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

2021-10-20: 【版本移除】 适用于JetPack4.2.2版本的SDK被标记为deprecated,该版本代码已停止更新,SDK会在未来某个版本移除,请切 换至新版本JetPack

快速开始 安装依赖 本SDK适用于JetPack4.5、JetPack4.6、JetPack5.0系列版本,请务必安装其中之一版本,并使用对应版本的SDK。 注意在安 装JetPack时,需同时安装CUDA、cuDNN、OpenCV、TensorRT等组件。

如已安装JetPack需要查询相关版本信息,请参考下文中的开发板信息查询与设置。

使用序列号激活

首先请在官网获取序列号。



将获取到的序列号填写到demo文件中或以参数形式传入。

图片加载失败

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己 的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的变量设置),需要设置额外的环境变量,指定CONTROLLER_KEY_AUTH_MODE为2, global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2),实例数鉴权模式下还支持指定license证书更新时间,单 位是秒,要求设置为大于20的整数,否则会采用默认的license更新时间,修改实例数鉴权license更新时间的方法参考 global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 300)

编译并运行Demo 模型资源文件默认已经打包在开发者下载的SDK包中。Demo工程直接编译即可运行。

编译运行:

cd src mkdir build && cd build cmake .. make -j\$(nproc) **make install 为可选,也可将lib所在路径添加为环境变量** sudo make install sudo ldconfig ./demo_batch_inference/easyedge_batch_inference {模型RES文件夹} {测试图片路径或仅包含图片的文件夹路径} {序列号}

demo运行示例:

baidu@nano:~/ljay/easydl/sdk/demo/build\$./demo_batch_inference/easyedge_batch_inference ../../../RES/ /ljay/images/mix008.jpeg 2020-08-06 20:56:30,665 INFO [EasyEdge] 548125646864 Compiling model for fast inference, this may take a while (Acceleration) 2020-08-06 20:57:58,427 INFO [EasyEdge] 548125646864 Optimized model saved to: /home/baidu/.baidu/easyedge/jetson/mcache/24110044320/m_cache, Don't remove it Results of image /ljay/images/mix008.jpeg: 2, kiwi, p:0.997594 loc: 0.352087, 0.56119, 0.625748, 0.868399 2, kiwi, p:0.993221 loc: 0.45789, 0.0730294, 0.73641, 0.399429 2, kiwi, p:0.992884 loc: 0.156876, 0.0598725, 0.3802, 0.394706 1, tomato, p:0.992125 loc: 0.523592, 0.389156, 0.657738, 0.548069 1, tomato, p:0.991821 loc: 0.665461, 0.419503, 0.805282, 0.573558 1, tomato, p:0.989883 loc: 0.297427, 0.439999, 0.432197, 0.59325 1, tomato, p:0.981654 loc: 0.383444, 0.248203, 0.506606, 0.400926 1, tomato, p:0.971682 loc: 0.183775, 0.556587, 0.286996, 0.711361 1, tomato, p:0.968722 loc: 0.379391, 0.0386965, 0.51672, 0.209681 Done



测试Demo HTTP 服务

编译demo完成之后,会同时生成一个http服务,运行

./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401} ./easyedge_serving ../../../RES "1111-1111-1111" 0.0.0.0 24401

日志中会显示

HTTP is now serving at 0.0.0.0:24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。



同时,可以调用HTTP接口来访问服务,具体参考下文接口说明。

使用说明

使用该方式,将运行库嵌入到开发者的程序当中。

使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

```
// step 1: 配置模型运行参数
EdgePredictorConfig config;
config.model_dir = model_dir;
config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, serial_num);
config.set_config(params::PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE, 1); // 优化的模型可以支持的batch_size
config.set_config(params::PREDICTOR_KEY_GTURBO_FP16, false); // 置true开启fp16模式推理会更快,精度会略微降低,但取决于硬
件是否支持fp16,不是所有模型都支持fp16,参阅文档
config.set_config(params::PREDICTOR_KEY_GTURBO_COMPILE_LEVEL, 1); // 编译模型的策略,如果当前设置的max_batch_size与历史编
译存储的不同,则重新编译模型
// step 2: 创建并初始化Predictor
auto predictor = global_controller()->CreateEdgePredictor<EdgePredictorConfig>(config);
if (predictor->init() != EDGE_OK) {
  exit(-1);
// step 3-1: 预测图像
auto img = cv::imread({图片路径});
std::vector<EdgeResultData> results;
predictor->infer(img, results);
// step 3-2: 预测视频
std::vector<EdgeResultData> results
FrameTensor frame_tensor;
VideoConfig video_config;
video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h
video_config.source_value = video_src;
/*
... more video_configs, 根据需要配置video_config的各选项
*/
auto video_decoding = CreateVideoDecoding(video_config);
while (video_decoding->next(frame_tensor) == EDGE_OK) {
 results.clear();
 if (frame_tensor.is_needed) {
    predictor->infer(frame_tensor.frame, results);
    render(frame_tensor.frame, results, predictor->model_info().kind);
  //video_decoding->display(frame_tensor); // 显示当前frame,需在video_config中开启配置
  //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置
```

初始化接口

```
auto predictor = global_controller()->CreateEdgePredictor<EdgePredictorConfig>(config);
if (predictor->init() != EDGE_OK) {
    exit(-1);
}
```

若返回非0,请查看输出日志排查错误原因。

预测接口

```
/**
* @brief
* 单图预测接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
    cv::Mat& image, std::vector<EdgeResultData>& result
) = 0;
/**
* @brief
* 批量图片预测接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
    std::vector<cv::Mat>& image,
     std::vector<std::vector<EdgeResultData>>& results
) = 0;
/**
 * @brief
* 批量图片预测接口,带阈值
 * @related infer(cv::Mat & image, EdgeColorFormat origin_color_format, std::vector<EdgeResultData> &result, float threshold)
*/
virtual int infer(
       std::vector<cv::Mat> &images,
       EdgeColorFormat origin_color_format,
       std::vector<std::vector<EdgeResultData>> &results
       float threshold
) = 0;
```

图片的格式务必为opencv默认的BGR, HWC格式。

批量图片的预测接口的使用要求在调用 init 接口的时候设置一个有效的 max_batch_size,其含义见下方参数配置接口的介绍。

参数配置接口 参数配置通过结构体 EdgePredictorConfig完成。

```
struct EdgePredictorConfig {
    /**
     * @brief 模型资源文件夹路径
    */
    std::string model_dir;
    std::map<std::string, std::string> conf;
    EdgePredictorConfig();
    template<typename T>
    T get_config(const std::string &key, const T &default_value);
    template<typename T = std::string>
    T get_config(const std::string &key);
    template<typename T>
    const T *get_config(const std::string &key, const T *default_value);
    template<typename T>
    void set_config(const std::string &key, const T &value);
    template<typename T>
    void set_config(const std::string &key, const T *value);
    static EdgePredictorConfig default_config();
  };
运行参数选项的配置以key、value的方式存储在类型为std::map的conf中,并且键值对的设置和获取可以通
```

过EdgePredictorConfig的set_config和get_config函数完成。同时也支持以环境变量的方式设置键值对。 EdgePredictorConfig的具体使用方法可以 参考开发工具包中的demo工程。

针对Jetson开发工具包,目前EdgePredictorConfig的运行参数所支持的Key包括如下项:

```
/**
* @brief 当有同类型的多个设备的时候,使用哪一个设备,如:
* GPU: 使用哪张GPU卡
* EdgeBoard(VMX), Movidius NCS:使用哪一张加速卡
* 值类型: int
* 默认值:0
*/
static constexpr auto PREDICTOR_KEY_DEVICE_ID = "PREDICTOR_KEY_DEVICE_ID";
/**
* @brief 生成 batch_size 为 max_batch_size 的优化模型,单次预测图片数量可以小于或等于此值(推荐等于此值,见release notes)
* 值类型: int
* 默认值:4
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE = "PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE";
/**
* @brief 设置device对应的GPU卡可以支持的最大并发量
* 实际预测的时候对应GPU卡的最大并发量不超过这里设置的范围,否则预测请求会排队等待预测执行
* 值类型: int
* 默认值:1
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY = "PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY";
/**
* @brief 是否开启fp16模式预测,开启后预测速度会更快,但精度会略有降低。并且需要硬件支持fp16
* 值类型: bool
* 默认值: false
*/
static constexpr auto PREDICTOR_KEY_GTURBO_FP16 = "PREDICTOR_KEY_GTURBO_FP16";
```

/**

* @brief 模型编译等级

```
*1:如果当前max_batch_size与历史编译产出的max_batch_size不相等时,则重新编译模型(推荐)
```

Baidu 百度智能云文档

*2:无论历史编译产出的max_batch_size为多少,均根据当前max_batch_size重新编译模型 * 值类型: int * 默认值:1 static constexpr auto PREDICTOR_KEY_GTURBO_COMPILE_LEVEL = "PREDICTOR_KEY_GTURBO_COMPILE_LEVEL"; /** * @brief GPU工作空间大小设置 * workspace_size = workspace_prefix * (1 << workspace_offset) * workspace_offset: 10 = KB, 20 = MB, 30 = GB * 值类型: int *默认值:WORKSPACE_PREFIX: 100, WORKSPACE_OFFSET: 20,即100MB static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX = "PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX"; static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET = "PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET"; /** * @brief 需要使用的dla core * 值类型: int * 默认值:-1(不使用) static constexpr auto PREDICTOR_KEY_GTURBO_DLA_CORE = "PREDICTOR_KEY_GTURBO_DLA_CORE"; /** * @brief 自定义缓存文件存储路径 * 值类型: string *默认值:~/.baidu/easyedge/mcache/{model_id * 1000000 + release_id} static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_DIR = "PREDICTOR_KEY_GTURBO_CACHE_DIR"; /** * @brief 自定义缓存文件命名,默认即可 * 值类型: string *默认值:根据配置自动生成 static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_NAME = "PREDICTOR_KEY_GTURBO_CACHE_NAME"; /** * @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号 * 值类型: string *默认值:空 static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";

PREDICTOR_KEY_GTURBO_CACHE_NAME:首次加载模型会先对模型进行编译优化,通过此值可以设置优化后的产出文件名。

PREDICTOR_KEY_GTURBO_CACHE_DIR:首次加载模型经过编译优化后,产出的优化文件会存储在这个位置,可以按需修改。

PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX、PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET:设置运行时可以被用来使用的最大临时显存。

PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE:此值用来控制批量图片预测可以支持的最大图片数,实际预测的时候单次预测图片数需等于此值。

PREDICTOR_KEY_DEVICE_ID:设置需要使用的 GPU 卡号,对于Jetson,此值无需更改。

PREDICTOR_KEY_GTURBO_COMPILE_LEVEL: 模型编译等级。通常模型的编译会比较慢,但编译产出是可以复用的。可以在第一次加载模型的时候设置合理的 max_batch_size 并在之后加载模型的时候直接使用历史编译产出。是否使用历史编译产出可以通过此值 compile_level 来控制,当此值为 0 时,表示忽略当前设置的 max_batch_size 而仅使用历史产出(无历史产出时则编译模型);当此值为 1 时,会比较历史产出和当前设置的 max_batch_size 是否相等,如不等,则重新编译;当此值为 2 时,无论如何都会重新编译模型。

PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY:通过此值设置单张 GPU 卡上可以支持的最大 infer 并发量,其上限取决于硬件限制。init 接口 会根据此值预分配 GPU 资源,建议结合实际使用控制此值,使用多少则设置多少。注意:此值的增加会降低单次 infer 的速度,建议优先考虑 batch inference。

PREDICTOR_KEY_GTURBO_FP16:默认是 fp32 模式,置 true 可以开启 fp16 模式预测,预测速度会有所提升,但精度也会略微下降,权衡使用。注意:不是所有模型都支持 fp16 模式,也不是所有硬件都支持 fp16 模式。已知不支持fp16的模型包括:EasyDL图像分类高精度模型。

预测视频接口

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

classVideoDecoding :

```
/**
* @brief 获取输入源的下一帧
* @param frame_tensor
* @return
*/
virtual int next(FrameTensor &frame_tensor) = 0;
/**
* @brief 显示当前frame_tensor中的视频帧
* @param frame_tensor
* @return
*/
virtual int display(const FrameTensor &frame_tensor) = 0;
/**
* @brief 将当前frame_tensor中的视频帧写为本地视频文件
* @param frame_tensor
* @return
*/
virtual int save(FrameTensor &frame_tensor) = 0;
/**
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
* @brief 获取视频的width属性
* @return
 */
virtual int get_width() = 0;
/**
* @brief 获取视频的height属性
* @return
*/
virtual int get_height() = 0;
```

struct VideoConfig

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type;
                          // 输入源类型
  std::string source_value;
                        // 输入源地址, 如视频文件路径、摄像头index、网络流地址
 int skip_frames{0};
                      // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
  int retrieve_all{false}; // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
  int input_fps{0};
                      // 在采取抽帧之前设置视频的fps
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
  bool enable_display{false};
  std::string window_name{"EasyEdge"};
 bool display_all{false};
                       // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
 bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                       // 是否存储所有frame,若为false,仅存储根据skip_frames抽取的frame
  std::map<std::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被 抽取返回,以作为显示或存储用。input_fps:用于抽帧前设置fps。resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。conf:高级选项。部分配置会通过该map来设置。

注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

EdgeLogConfig log_config; log_config.enable_debug = true; global_controller()->set_log_config(log_config);

返回格式

预测成功后,从 EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index; // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测或图像分割时才有
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割时才有
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding, 游程编码的mask
};
```

关于矩形坐标

x1 图片宽度 = 检测框的左上角的横坐标 y1 图片高度 = 检测框的左上角的纵坐标 x2 图片宽度 = 检测框的右下角的横坐标 y2 图片高度 = 检测框 的右下角的纵坐标

以上字段可以参考demo文件中使用opencv绘制的逻辑进行解析

http服务

/**

1. 开启http服务 http服务的启动参考demo_serving.cpp文件。

* http服务里,图片的解码运行在cpu之上,可能会降低推理速度。

* @brief 开启一个简单的demo http服务。 * 该方法会block直到收到sigint/sigterm。

2. 请求http服务

开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。

URL中的get参数:

参数	说明		默认值	
threshold	阈值过滤,	0~1	如不提供,则会使用模型的推荐阈值	

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

Python请求示例

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
        'http://127.0.0.1:24401/',
        params={'threshold': 0.1},
        data=img).json()
```

Java请求示例参考这里

http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考接口使用·返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

返回示例

```
"cost_ms": 52,
"error_code": 0,
"results": [
    "confidence": 0.94482421875,
    "index": 1
    "label": "IronMan",
    "x1": 0.059185408055782318,
    "x2": 0.18795496225357056,
    "y1": 0.14762254059314728,
     "y2": 0.52510076761245728
    "confidence": 0.94091796875,
    "index": 1.
    "label": "IronMan",
    "x1": 0.79151463508605957,
    "x2": 0.92310667037963867,
    "y1": 0.045728668570518494,
    "y2": 0.42920106649398804
```

多线程预测 Jetson 系列 SDK 支持多线程预测,创建一个 predictor,并通过 PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY 控制所支持的最大 并发量,只需要 init 一次,多线程调用 infer 接口。需要注意的是多线程的启用会随着线程数的增加而降低单次 infer 的推理速度,建议优先使用 batch inference 或权衡考虑使用。

已知问题 1. 多线程时图片按线程分配不均 或 不同batch size的图片交叉调用infer接口时,部分结果错误

A:EasyDL图像分类高精度模型在有些显卡上可能存在此问题,可以考虑填充假图片数据到图片比较少的线程或batch以使得infer间的图片绝对平均。

2. 显存持续增长或遇到 terminate called after throwing an instance of 'std::runtime_error' what(): Failed to create object

A:如果遇到此问题,请确认没有频繁调用 init 接口,通常调用 infer 接口即可满足需求。

3. 开启 fp16 后, 预测结果错误

A:不是所有模型都支持 fp16 模式。目前已知的不支持fp16的模型包括:EasyDL图像分类高精度模型。目前不支持的将会在后面的版本陆续支持。

4. 部分模型不支持序列化

A:针对JetPack4.4、4.5版本,部分模型无法使用序列化,如已知的BML的MobileNetV1-SSD和物体检测高性能模型。需要每次加载模型的时候

编译模型,过程会比较慢。此问题将在后续JetPack版本中修复。目前JetPack4.6版本SDK已修复该问题。

开发板信息查询与设置查询L4T或JetPack版本查询JetPack版本信息,可以通过下面这条命令先查询L4T的版本。

在终端输入如下命令并回车 \$ head -n 1 /etc/nv_tegra_release **就会输出类似如下结果** \$ # # R32 (release), REVISION: 4.3, GCID: 21589087, BOARD: t210ref, EABI: aarch64, DATE: Fri Jun 26 04:38:25 UTC 2020

从输出的结果来看,板子当前的L4T版本为R32.4.3,对应JetPack4.4。 注意,L4T的版本不是JetPack的版本,一般可以从L4T的版本唯一对应到 JetPack的版本,下面列出了最近几个版本的对应关系:

L4T R32.6.1 --> JetPack4.6 L4T R32.5.1 --> JetPack4.5.1 L4T R32.5 --> JetPack4.5 L4T R32.4.3 --> JetPack4.4 L4T R32.4.2 --> JetPack4.4DP L4T R32.2.1 --> JetPack4.2.2 L4T R32.2.0 --> JetPack4.2.1

功率模式设置与查询不同的功率模式下,执行AI推理的速度是不一样的,如果对速度需求很高,可以把功率开到最大,但记得加上小风扇散热~

```
**1. 运行下面这条命令可以查询开发板当前的运行功率模式**

$ sudo nvpmodel -q verbose

**$ NV Power Mode: MAXN**

**$ 0**

**如果输出为MAXN代表是最大功率模式**

**2. 若需要把功率调到最大,运行下面这条命令**

$ sudo nvpmodel -m 0
```

如果你进入了桌面系统,也可以在桌面右上角有个按钮可以切换模式

3. 查询资源利用率 \$ sudo tegrastats

FAQ 1. EasyDL SDK与云服务效果不一致,如何处理?

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

2. 运行SDK报错 Authorization failed 日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888" ./easyedge_demo ...

3. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

4. 运行demo时报找不到libeasyedge_extension.so

需要export libeasyedge_extension.so所在的路径,如路径为/home/work/baidu/cpp/lib,则需执行:

export LD_LIBRARY_PATH=/home/work/baidu/cpp/lib:\${LD_LIBRARY_PATH}

或者在编译完后执行如下命令将lib文件安装到系统路径:

sudo make install

如不能安装,也可手动复制lib下的文件到/usr/local/lib下。

5. 运行demo时报如下之一错误

2020-12-17 16:15:07,924 INFO [EasyEdge] 547633188880 Compiling model for fast inference, this may take a while (Acceleration) Killed

或

2020-12-17 16:15:07,924 INFO [EasyEdge] 547633188880 Build graph failed

请适当降低PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE和PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY的值后尝试。

6. 运行有损压缩加速的模型,运算精度较标准模型偏低首先请保证数据集不会太小。其次可以通过将模型目录RES中的calibrationtable移除,并通过将PREDICTOR_KEY_GTURBO_FP16设置为true,使用FP16的运算精度重新评估模型效果。若依然不理想,可将calibrationtable移除并将PREDICTOR_KEY_GTURBO_FP16设置为false,从而使用更高精度的FP32的运算精度。

心图像分类辨影专用SDK集成文档

简介

本文档介绍EasyEdge/EasyDL的辨影软硬一体方案SDK的使用方法。支持的硬件包括辨影Air、辨影Pro。您可以在软硬一体方案了解部署方案。

模型支持:

- EasyDL图像:图像分类高精度,图像分类高性能,物体检测高精度,物体检测均衡,物体检测高性能
- BML :
 - 公开数据集预训练模型:SSD-MobileNetV1,YOLOv3-DarkNet,YOLOv3-MobileNetV1,ResNet50,ResNet101,SE-ResNeXt50,SE-ResNeXt101,MobileNetV2,EfficientNetB0_small,EfficientNetB4,MobileNetV3_large_x1_0,ResNet18_vd,SE_ResNet18_vd,Xception71。
 - 百度超大规模数据集预训练模型:YOLOv3-DarkNet, MobileNetV3_large_x1_0, ResNet50_vd, ResNet101_vd。
- EasyEdge:EasyEdge支持的模型较多,详见查看模型网络适配硬件。若模型不在此列表,可以尝试使用自定义网络生成端计算组件。

Release Notes | 时间 | 版本 | 说明 | | --- | ---- | | 2022.08.01 | 1.3.5 | 新增支持辨影软硬一体方案部署 |

辨影软件接入使用SDK 辨影Air/Pro自带软件预置了大量飞桨开源模型,支持EasyDL/BML模型SDK一键导入使用,详细的辨影使用说明见购买后 获得的使用说明书

• 辨影推理主界面



• 辨影设置界面。在应用中可选预置模型能力,也可选择EasyDL/BML导入的模型SDK



快速开始 接下来的文档内容将会描述辨影SDK的集成开发教程,仅需要使用辨影自带软件的用户无需关注
首先请在EasyDL智能云官网获取序列号。

✔ 返回EasyDL总览	く 返回 专项硬件适配服务	
EasyDL图像	按单台设备激活 按多台设备激活	
算力资源管理	(赤田光明, (屋工委委编初七升、邀任七注章)	
公有云部署 ~	医历时间, (板)里有这次方式、 称(石)方达等) 提示:如您已完成企业资质认证,将享受该账号下任意部署包免费试用2个月,更多企业认证权益详见 企业权益	結专属礼包 ?
EasyEdge本地部署 ^		
• 服务器纯离线服务	购买永久授权 新增测试序列号 管理序列号 查看转让序列号 批量离线激活	
• 设备端纯离线服务		
• 专项硬件纯离线	设备名 激活状态	序列号
	自定义设备 🖉 已过期	8BA2-39AA-73F
	自定义设备 🧷 未激活	2013-2CD8-458

将获取到的序列号填写到demo文件中或以参数形式传入。



默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的变量设置),需要设置额外的环境变量,指定CONTROLLER_KEY_AUTH_MODE为2,global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2),实例数鉴权模式下还支持指定license证书更新时间,单位是秒,要求设置为大于20的整数,否则会采用默认的license更新时间,修改实例数鉴权license更新时间的方法参考 global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 300)

编译并运行Demo 模型资源文件默认已经打包在开发者下载的SDK包中。Demo工程直接编译即可运行。

编译运行:

cd src mkdir build && cd build cmake ... make **make install 为可选,也可将lib所在路径添加为环境变量** sudo make install sudo ldconfig ./demo_batch_inference/easyedge_batch_inference {模型RES文件夹} {测试图片路径或仅包含图片的文件夹路径} {序列号}

baidu@nano:~/ljay/easydl/sdk/demo/build\$./demo_batch_inference/easyedge_batch_inference ../../../RES/ /ljay/images/mix008.jpeg

2020-08-06 20:56:30,665 INFO [EasyEdge] 548125646864 Compiling model for fast inference, this may take a while (Acceleration) 2020-08-06 20:57:58,427 INFO [EasyEdge] 548125646864 Optimized model saved to:

/home/baidu/.baidu/easyedge/jetson/mcache/24110044320/m_cache, Don't remove it

Results of image /ljay/images/mix008.jpeg:

- 2, kiwi, p:0.997594 loc: 0.352087, 0.56119, 0.625748, 0.868399
- 2, kiwi, p:0.993221 loc: 0.45789, 0.0730294, 0.73641, 0.399429
- 2, kiwi, p:0.992884 loc: 0.156876, 0.0598725, 0.3802, 0.394706
- 1, tomato, p:0.992125 loc: 0.523592, 0.389156, 0.657738, 0.548069
- 1, tomato, p:0.991821 loc: 0.665461, 0.419503, 0.805282, 0.573558
- 1, tomato, p:0.989883 loc: 0.297427, 0.439999, 0.432197, 0.59325
- 1, tomato, p:0.981654 loc: 0.383444, 0.248203, 0.506606, 0.400926

demo运行示例:

^{1,} tomato, p:0.971682 loc: 0.183775, 0.556587, 0.286996, 0.711361

^{1,} tomato, p:0.968722 loc: 0.379391, 0.0386965, 0.51672, 0.209681 Done



测试Demo HTTP 服务

编译demo完成之后,会同时生成一个http服务,运行

./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401} ./easyedge_serving ../../../RES "1111-1111-1111" 0.0.0.0 24401

日志中会显示

HTTP is now serving at 0.0.0.0:24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。



同时,可以调用HTTP接口来访问服务,具体参考下文接口说明。

使用说明

使用该方式,将运行库嵌入到开发者的程序当中。

使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

```
// step 1: 配置模型运行参数
EdgePredictorConfig config;
config.model_dir = model_dir;
config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, serial_num);
config.set_config(params::PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE, 1); // 优化的模型可以支持的最大batch_size,实际单次推理的图
片数不能大于此值
config.set_config(params::PREDICTOR_KEY_GTURBO_FP16, false); // 置true开启fp16模式推理会更快,精度会略微降低,但取决于硬
件是否支持fp16,不是所有模型都支持fp16,参阅文档
config.set_config(params::PREDICTOR_KEY_GTURBO_COMPILE_LEVEL, 1); // 编译模型的策略,如果当前设置的max_batch_size与历史编
译存储的不同,则重新编译模型
// step 2: 创建并初始化Predictor
auto predictor = global_controller()->CreateEdgePredictor<EdgePredictorConfig>(config);
if (predictor->init() != EDGE_OK) {
  exit(-1);
// step 3-1: 预测图像
auto img = cv::imread({图片路径});
std::vector<EdgeResultData> results;
predictor->infer(img, results);
// step 3-2: 预测视频
std::vector<EdgeResultData> results;
FrameTensor frame_tensor;
VideoConfig video_config;
video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h
video_config.source_value = video_src;
/*
... more video_configs, 根据需要配置video_config的各选项
*/
auto video_decoding = CreateVideoDecoding(video_config);
while (video_decoding->next(frame_tensor) == EDGE_OK) {
 results.clear();
 if (frame_tensor.is_needed) {
    predictor->infer(frame_tensor.frame, results);
    render(frame_tensor.frame, results, predictor->model_info().kind);
  //video_decoding->display(frame_tensor); // 显示当前frame, 需在video_config中开启配置
  //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置
```

初始化接口



若返回非0,请查看输出日志排查错误原因。

预测接口

```
/**
* @brief
* 单图预测接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
    cv::Mat& image, std::vector<EdgeResultData>& result
) = 0;
/**
* @brief
* 批量图片预测接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
    std::vector<cv::Mat>& image,
     std::vector<std::vector<EdgeResultData>>& results
) = 0;
/**
 * @brief
* 批量图片预测接口,带阈值
 * @related infer(cv::Mat & image, EdgeColorFormat origin_color_format, std::vector<EdgeResultData> &result, float threshold)
*/
virtual int infer(
       std::vector<cv::Mat> &images,
       EdgeColorFormat origin_color_format,
       std::vector<std::vector<EdgeResultData>> &results
       float threshold
) = 0;
```

图片的格式务必为opencv默认的BGR, HWC格式。

批量图片的预测接口的使用要求在调用 init 接口的时候设置一个有效的 max_batch_size,其含义见下方参数配置接口的介绍。

参数配置接口 参数配置通过结构体 EdgePredictorConfig完成。

```
struct EdgePredictorConfig {
    /**
     * @brief 模型资源文件夹路径
    */
    std::string model_dir;
    std::map<std::string, std::string> conf;
    EdgePredictorConfig();
    template<typename T>
    T get_config(const std::string &key, const T &default_value);
    template<typename T = std::string>
    T get_config(const std::string &key);
    template<typename T>
    const T *get_config(const std::string &key, const T *default_value);
    template<typename T>
    void set_config(const std::string &key, const T &value);
    template<typename T>
    void set_config(const std::string &key, const T *value);
    static EdgePredictorConfig default_config();
  };
运行参数选项的配置以key、value的方式存储在类型为std::map的conf中,并且键值对的设置和获取可以通
```

运行参数选项的配直以key、value的方式存储在交型为std::map的cont中,并且键值对的设直和获取可以通 过EdgePredictorConfig的set_config和get_config函数完成。同时也支持以环境变量的方式设置键值对。 EdgePredictorConfig的具体使用方法可以 参考开发工具包中的demo工程。

针对Jetson开发工具包,目前EdgePredictorConfig的运行参数所支持的Key包括如下项:

```
/**
* @brief 当有同类型的多个设备的时候,使用哪一个设备,如:
* GPU: 使用哪张GPU卡
* EdgeBoard(VMX), Movidius NCS:使用哪一张加速卡
* 值类型: int
* 默认值:0
*/
static constexpr auto PREDICTOR_KEY_DEVICE_ID = "PREDICTOR_KEY_DEVICE_ID";
/**
* @brief 生成最大 batch_size 为 max_batch_size 的优化模型,单次预测图片数量可以小于或等于此值
* 值类型: int
* 默认值:4
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE = "PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE";
/**
* @brief 设置device对应的GPU卡可以支持的最大并发量
* 实际预测的时候对应GPU卡的最大并发量不超过这里设置的范围,否则预测请求会排队等待预测执行
* 值类型: int
* 默认值:1
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY = "PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY";
/**
* @brief 是否开启fp16模式预测,开启后预测速度会更快,但精度会略有降低。并且需要硬件支持fp16
* 值类型: bool
* 默认值: false
*/
static constexpr auto PREDICTOR_KEY_GTURBO_FP16 = "PREDICTOR_KEY_GTURBO_FP16";
/**
* @brief 模型编译等级
```

```
* 1:如果当前max_batch_size与历史编译产出的max_batch_size不相等时,则重新编译模型(推荐)
```

*2:无论历史编译产出的max_batch_size为多少,均根据当前max_batch_size重新编译模型 * 值类型: int * 默认值:1 static constexpr auto PREDICTOR_KEY_GTURBO_COMPILE_LEVEL = "PREDICTOR_KEY_GTURBO_COMPILE_LEVEL"; /** * @brief GPU工作空间大小设置 * workspace_size = workspace_prefix * (1 << workspace_offset) * workspace_offset: 10 = KB, 20 = MB, 30 = GB * 值类型: int *默认值:WORKSPACE_PREFIX: 100, WORKSPACE_OFFSET: 20,即100MB static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX = "PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX"; static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET = "PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET"; /** * @brief 需要使用的dla core * 值类型: int * 默认值:-1(不使用) static constexpr auto PREDICTOR_KEY_GTURBO_DLA_CORE = "PREDICTOR_KEY_GTURBO_DLA_CORE"; /** * @brief 自定义缓存文件存储路径 * 值类型: string * 默认值: ~/.baidu/easyedge/mcache/{model_id * 1000000 + release_id} static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_DIR = "PREDICTOR_KEY_GTURBO_CACHE_DIR"; /** * @brief 自定义缓存文件命名,默认即可 * 值类型: string *默认值:根据配置自动生成 static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_NAME = "PREDICTOR_KEY_GTURBO_CACHE_NAME"; /** * @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号 * 值类型: string *默认值:空 static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";

PREDICTOR_KEY_GTURBO_CACHE_NAME:首次加载模型会先对模型进行编译优化,通过此值可以设置优化后的产出文件名。

PREDICTOR_KEY_GTURBO_CACHE_DIR:首次加载模型经过编译优化后,产出的优化文件会存储在这个位置,可以按需修改。

PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX、PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET:设置运行时可以被用来使用的最大临时显存。

PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE:此值用来控制批量图片预测可以支持的最大图片数,实际预测的时候单次预测图片数不可大于此值,但可以是不大于此值的任意图片数。

PREDICTOR_KEY_GTURBO_COMPILE_LEVEL: 模型编译等级。通常模型的编译会比较慢,但编译产出是可以复用的。可以在第一次加载模型的时候设置合理的 max_batch_size 并在之后加载模型的时候直接使用历史编译产出。是否使用历史编译产出可以通过此值 compile_level 来控制,当此值为 0 时,表示忽略当前设置的 max_batch_size 而仅使用历史产出(无历史产出时则编译模型);当此值为 1 时,会比较历史产出和当前设置的 max_batch_size 是否相等,如不等,则重新编译;当此值为 2 时,无论如何都会重新编译模型。

PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY:通过此值设置单张 GPU 卡上可以支持的最大 infer 并发量,其上限取决于硬件限制。init 接口 会根据此值预分配 GPU 资源,建议结合实际使用控制此值,使用多少则设置多少。注意:此值的增加会降低单次 infer 的速度,建议优先考虑 batch inference。

PREDICTOR_KEY_GTURBO_FP16:默认是 fp32 模式,置 true 可以开启 fp16 模式预测,预测速度会有所提升,但精度也会略微下降,权衡使用。注意:不是所有模型都支持 fp16 模式,也不是所有硬件都支持 fp16 模式。已知不支持fp16的模型包括:EasyDL图像分类高精度模型。

预测视频接口

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

classVideoDecoding :

```
/**
* @brief 获取输入源的下一帧
* @param frame_tensor
* @return
*/
virtual int next(FrameTensor & frame_tensor) = 0;
/**
* @brief 显示当前frame_tensor中的视频帧
* @param frame_tensor
* @return
*/
virtual int display(const FrameTensor &frame_tensor) = 0;
/**
* @brief 将当前frame_tensor中的视频帧写为本地视频文件
* @param frame_tensor
* @return
*/
virtual int save(FrameTensor & frame_tensor) = 0;
/**
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
* @brief 获取视频的width属性
 * @return
 */
virtual int get_width() = 0;
/**
* @brief 获取视频的height属性
* @return
*/
virtual int get_height() = 0;
```

struct VideoConfig

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type;
                          // 输入源类型
  std::string source_value;
                        // 输入源地址, 如视频文件路径、摄像头index、网络流地址
 int skip_frames{0};
                      // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
  int retrieve_all{false}; // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
  int input_fps{0};
                      // 在采取抽帧之前设置视频的fps
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
  bool enable_display{false};
  std::string window_name{"EasyEdge"};
 bool display_all{false};
                       // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                       // 是否存储所有frame,若为false,仅存储根据skip_frames抽取的frame
  std::map<std::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被 抽取返回,以作为显示或存储用。input_fps:用于抽帧前设置fps。resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。conf:高级选项。部分配置会通过该map来设置。

注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

EdgeLogConfig log_config; log_config.enable_debug = true; global_controller()->set_log_config(log_config);

返回格式

预测成功后,从 EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index; // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测或图像分割时才有
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割时才有
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding, 游程编码的mask
};
```

关于矩形坐标

x1 图片宽度 = 检测框的左上角的横坐标 y1 图片高度 = 检测框的左上角的纵坐标 x2 图片宽度 = 检测框的右下角的横坐标 y2 图片高度 = 检测框 的右下角的纵坐标

以上字段可以参考demo文件中使用opencv绘制的逻辑进行解析

http服务

/**

1. 开启http服务 http服务的启动参考demo_serving.cpp文件。

* @brief 开启一个简单的demo http服务。 * 该方法会block直到收到sigint/sigterm。 * http服务里,图片的解码运行在cpu之上,可能会降低推理速度。 * @tparam ConfigT * @param config * @param host * @param port * @param service_id service_id user parameter, uri '/get/service_id' will respond this value with 'text/plain' * @param instance_num 实例数量,根据内存/显存/时延要求调整 * @return */ template<typename ConfigT> int start_http_server(const ConfigT &config, const std::string &host, int port, const std::string &service_id, int instance_num = 1);

2. 请求http服务

开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。

URL中的get参数:

参数	说明		默认值
threshold	阈值过滤,	0~1	如不提供,则会使用模型的推荐阈值

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

Python请求示例

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
        'http://127.0.0.1:24401/',
        params={'threshold': 0.1},
        data=img).json()
```

Java请求示例参考这里

http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考接口使用·返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

返回示例

```
"cost_ms": 52,
"error_code": 0,
"results": [
    "confidence": 0.94482421875,
    "index": 1
    "label": "IronMan",
    "x1": 0.059185408055782318,
    "x2": 0.18795496225357056,
    "y1": 0.14762254059314728,
     "y2": 0.52510076761245728
    "confidence": 0.94091796875,
    "index": 1.
    "label": "IronMan",
    "x1": 0.79151463508605957,
    "x2": 0.92310667037963867,
    "y1": 0.045728668570518494,
    "y2": 0.42920106649398804
```

多线程预测 辨影系列 SDK 支持多线程预测,创建一个 predictor,并通过 PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY 控制所支持的最大并 发量,只需要 init 一次,多线程调用 infer 接口。需要注意的是多线程的启用会随着线程数的增加而降低单次 infer 的推理速度,建议优先使用 batch inference 或权衡考虑使用。

已知问题 1. 多线程时图片按线程分配不均 或 不同batch size的图片交叉调用infer接口时,部分结果错误

A:EasyDL图像分类高精度模型在有些显卡上可能存在此问题,可以考虑填充假图片数据到图片比较少的线程或batch以使得infer间的图片绝对平均。

2. 显存持续增长或遇到 terminate called after throwing an instance of 'std::runtime_error' what(): Failed to create object

A:如果遇到此问题,请确认没有频繁调用 init 接口,通常调用 infer 接口即可满足需求。

3. 开启 fp16 后, 预测结果错误

A:不是所有模型都支持 fp16 模式。目前已知的不支持fp16的模型包括:EasyDL图像分类高精度模型。目前不支持的将会在后面的版本陆续支持。

4. 部分模型不支持序列化

A:针对JetPack4.4版本,部分模型无法使用序列化,如已知的BML的MobileNetV1-SSD和物体检测高性能模型。需要每次加载模型的时候编译模

型,过程会比较慢。此问题将在后续JetPack版本中修复。

开发板信息查询与设置查询L4T或JetPack版本查询JetPack版本信息,可以通过下面这条命令先查询L4T的版本。

在终端输入如下命令并回车 \$ head -n 1 /etc/nv_tegra_release **就会输出类似如下结果** \$ # # R32 (release), REVISION: 4.3, GCID: 21589087, BOARD: t210ref, EABI: aarch64, DATE: Fri Jun 26 04:38:25 UTC 2020

从输出的结果来看,板子当前的L4T版本为R32.4.3,对应JetPack4.4。 注意,L4T的版本不是JetPack的版本,一般可以从L4T的版本唯一对应到 JetPack的版本,下面列出了最近几个版本的对应关系:

L4T R32.6.1 --> JetPack4.6 L4T R32.5.1 --> JetPack4.5.1 L4T R32.5 --> JetPack4.5 L4T R32.4.3 --> JetPack4.4 L4T R32.4.2 --> JetPack4.4DP L4T R32.2.1 --> JetPack4.2.2 L4T R32.2.0 --> JetPack4.2.1

功率模式设置与查询不同的功率模式下,执行AI推理的速度是不一样的,如果对速度需求很高,可以把功率开到最大,但记得加上小风扇散热~

```
**1. 运行下面这条命令可以查询开发板当前的运行功率模式**

$ sudo nvpmodel -q verbose

**$ NV Power Mode: MAXN**

**$ 0**

**如果输出为MAXN代表是最大功率模式**

**2. 若需要把功率调到最大,运行下面这条命令**

$ sudo nvpmodel -m 0
```

如果你进入了桌面系统,也可以在桌面右上角有个按钮可以切换模式

3. 查询资源利用率 \$ sudo tegrastats

FAQ 1. EasyDL SDK与云服务效果不一致,如何处理?

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

2. 运行SDK报错 Authorization failed 日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888"
./easyedge_demo ...

3. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

4. 运行demo时报找不到libeasyedge_extension.so

需要export libeasyedge_extension.so所在的路径,如路径为/home/work/baidu/cpp/lib,则需执行:

export LD_LIBRARY_PATH=/home/work/baidu/cpp/lib:\${LD_LIBRARY_PATH}

或者在编译完后执行如下命令将lib文件安装到系统路径:

sudo make install

如不能安装,也可手动复制lib下的文件到/usr/local/lib下。

5. 运行demo时报如下之一错误

2020-12-17 16:15:07,924 INFO [EasyEdge] 547633188880 Compiling model for fast inference, this may take a while (Acceleration) Killed

或

2020-12-17 16:15:07,924 INFO [EasyEdge] 547633188880 Build graph failed

请适当降低PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE和PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY的值后尝试。

6. 运行有损压缩加速的模型,运算精度较标准模型偏低 首先请保证数据集不会太小。其次可以通过将模型目录 RES中的 calibration table 移除,并 通过将 PREDICTOR_KEY_GTURBO_FP16设置为 true,使用 FP16的运算精度重新评估模型效果。若依然不理想,可将 calibration table 移除并 将 PREDICTOR_KEY_GTURBO_FP16设置为 false,从而使用更高精度的FP32的运算精度。

- 心 端云协同部署
- 心 端云协同服务说明

服务简介

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

「云管理,端计算」的端云协同服务,具体包括:

- 在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新
- 断网状态下模型离线计算 (http服务,可调用与公有云API功能相同的接口)
- 联网状态下在平台管理设备运行状态、资源利用率

目前本地服务器的应用平台支持Linux-AMD64(x86-64),具体使用流程请参考下方文档。

使用流程

Step 1 发布端云协同部署包

在我的部署包页面点击「发布端云协同部署包」

1	图像分类 定制图像分类模型 在各分类图片之间	模型	操作文档 教号 张图整体是什么 况下,训练数据	举祝頻 常见问题 物体/状态/场景。 每类仅需20-100	援交工单 张,最快10分钟可训练	东完毕		收起 ^
模型中心	端云协同服务 > 我的部署包	L.						
我的模型	端云协同服务说明							点击收起
创建模型 训练模型	 在可視化界面轻松实现模型部 断网状态下模型离线计算(h) 联网状态下在平台管理设备运 	署包在边缘设备 tp服务,可调用 行状态、资源利	上的集成、版本更 与公有云API功能 I用率	新 相同的接口)				
校验模型	具体使用流程如下:							
发布模型	1、在本页面发布端云协同部署包	,填写相关信息	并申请,并在下方 注命令 联网谈话	列表查看部署包发布	讨法态			
EasyData数据服务	3、在「下发部署包到设备」页	面,将发布成功	的部署包一键下发	到已激活的设备上,	即可测试或正式使用			
数据总览	发布综元协同部署包	1						
标签组管理	A MARLADISHPERS	1						
在线标注	嬸云协同服务名称	模型ID	设备类型	应用平台	当前可用版本	新版本发布状态	部署设备数	操作
云服务数据回流								
EasyEdge本地部署								
纯离线服务								
端云协同服务 へ								
我的部署包					暂无可用	数据		
我的本地设备					请稍后再证			
下发部署包到设备								

填写服务名称,选择模型版本并提交发布

模型中心	端云协同服务 > 发布端云	协同部署包	
我的模型	设备类型 💿 服	务器 🗌 通用小型设备	
创建模型			
训练模型	模型名称 识别	樱桃和小眷茄	\checkmark
校验模型	端云协同服务名称		
发布模型	选择版本 V1		\checkmark
EasyData数据服务	选择系统和芯片	Linux	
数据总览		通用X86 CPU	
标签组管理			
在线标注	发布部署包		
云服务数据回流			

在列表查看部署包发布状态

5	图像分类模型 操作文档 数学视频 常见问题 提交工单 定制图像分类模型,可以识别一张图整体是什么物体/状态/场景。 在各分类图片之间差异明显的情况下,训练数据每类仅需20-100张,最快10分钟可训练完毕
模型中心	端云协同服务 > 我的部署包
我的模型	端云协同服务说明
创建模型 训练模型	1、在可视化界遍轻松突现模型部署包在边缘设备上的集成、版本更新 2、断段状态下模型高线计算(http服务,可调用与公看云API功能相同的接口) 3、联网状态下在平台管理设备运行状态、资源利用率
校验模型	具体使用流程如下:
发布模型	 在本页區及布瑞云协同部署目,填写相关信息并申请,并在下方列表責着部署包发布状态 在「我的本始设备」页面新增设备,复制激活命令, 原网激活本地设备
EasyData数据服务	3、在「下友部書包到设备」页面,将发布成功的邮書包一個下发到已激活的设备上,即可测试或正式使用
数据总览	发布瑞云协同邮署包
标签组管理	信三体回彩条文轮 排削D 边发光则 页用可石 当前可用乐士 新乐士世立学术 却是边多物 语此
在线标注	端ムが内成功立が (株正U 以前天正 広力でロ コ前り力加成中 前版中次10人の IP者及前款 第1F first 25282 服務員 Linux通知V86 CDLI V2 V2
云服务数据回流	
EasyEdge本地部署	毎页显示 10 🗸 < 1 >
纯离线服务	
端云协同服务 へ	
我的部署包	
我的本地设备	

Step 2 新增设备并激活

在我的本地设备页面新增设备

模型中心	端云协同服务 > 1	找的本地设备					
我的模型	我的本地设备						点击收制
创建模型	在本页面新增设备、I 利用率等信息。	关网激活本地设备后,即可	将「 我的部署包 」页	面发布成功的部署包一键下	*发到设备上。设备联网时,可以	查看设备上部署的服务、设行	备的运行状态、资源
训练模型							
校验模型	新增设备						
发布模型	设备名称	创建时间	设备类型	应用平台	设备连接状态	最新同步时间	操作
EasyData数据服务							
数据总览							
标签组管理							

		新增设备	×
	设备名称	20个字符以内	
4	备注信息	50个字符以内, 自定义备注信息	
c		0/50	
Ŕ		0/00/2	
		确认 取消	

在列表中,点击设备对应的「激活设备」操作,复制激活命令并在本地设备上执行即可

() () () () () () () () () () () () () (查看激活命令	×		收起 へ
在行	复制命令,并在本地设备上执行,即可激活设备			
端云协同服务	curl -skfL 'https://verify.baidubce.com/active/v1/active/s sh?token=e177df4ed47b2265223a333630302c226b223a226e6f64652;	etup.		
我的本地设备 在本页面新增设 利用率等信息。	6e223a22353730313835302d382d65393866386363302d343231612d65 12d373138392d376266663366643063313433222c226e73223a2262616 6c2d636c6f7564222c227473223a31363039393939393434327d' -osetu && sh setup.sh	37653 57479 p.sh	可以查看设备上部署的服务、	点击收起 设备的运行状态、资源
新增设备				
设备名称 linux x86 ビ	知道了 2021-01-07 14:02 - 未返活		寸间 操作 	插详情

Step 3 下发部署包到设备,在本地调用

在下发部署包到设备页面,将发布成功的部署包一键下发到已激活的设备上,即可测试或正式使用

模型中心	端云协同服务 > 下发部署包到设备
我的模型	将已成功发布的端云协共服务部署包一键下发到已激活且在连接中的设备
创建模型	
训练模型	服务名称
校验模型	模型ID 246
发布模型	应用平台 Linux-通用X86 CPU 🗸
EasyData数据服务	选择版本 V2 🗸
数据总览	12 (2) 0 x
标签组管理	JS:f≠ix™ UDUNTU-I0 ×
在线标注	服务配置 ④基础设置
云服务数据回流	服务端口 8080
EasyEdge本地部署	
纯离线服务	下发服务
端云协同服务 へ	

部署包下发成功之后,会在本地启动一个HTTP推理服务。在浏览器中输入http://{设备ip}:{服务端口,默认8080},即可预览效果:

当前阈值: 0.07 修改

置信度 0.998

0.937

0.841

0.781

0,408

0.324

0.251

▲ 不安全 yq01-ns-map0402a659.yq01:8081



具体接口调用说明请参考文档 SDK - HTTP服务调用说明

云端管理说明

模型部署包管理

在我的部署包页面可以进行已发布的模型部署包的管理。

发布及更新模型版本

点击「发布新版本」操作即可快速发布对应模型ID下的新版本。同一模型ID下已发布的模型版本均会显示在列表的「当前可用版本」中。

模	輕中心	端云协同服务 > 我的部署包	!								
我	的模型	端云协同服务说明									点击收起
Û	建模型	1、在可視化界面轻松实现模型部	署包在边	缘设备上的集成、 可调用与公布三。	版本更新						
າແ	练模型	2、断网状态下供型高线计算(h 3、联网状态下在平台管理设备运	itp服务, 行状态、	可调用与公有云# 资源利用率	(P14)用它们可以按口)						
校	验模型	具体使用流程如下:									
发	市模型	1、在本页面发布端云协同部署包 2、在「 我的本地设备 」页面新	,填写相: 增设备,复	关信息并申请,扌 夏制激活命令,耶	+在下方列表查看部署包发 K网激活本地设备	发布状态					
Ea	asyData数据服务	3、在「 下发部署包到设备 」页 	面,将发7	^{先成功的部署包}	-键下发到已激活的设备上	-,即可测试或正	式使用				
数	据总览	发布端云协同部署包									
标	签组管理										
在	线标注	端云协同服务名称	模型ID	设备类型	应用半台	当前可用版本	新版本发布状态	部署设备数	操作		
云	服务数据回流	ecc	264	通用小型设备	Linux-通用ARM	V1	V1-已发布	0	下发到设备	发布新版本	
Ea	asyEdge本地部署		246	服务器	Linux-通用X86 CPU	V2, V1	V2-已发布	1	下发到设备	h 发布新版本 [服务详情
纯	离线服务	00000000	265	通用小型设备	Linux-通用ARM	-	V1-发布中	0			
孈	云协同服务 へ								每页显示	10 🗸 <	1 >
	我的部署包										
	我的本地设备										
	下发部署包到设备										
			发布	新版本				>			
	将最新训练的模	韓型版本发布为服务,	发布质	龙功后,艮	「可从云端下发	到设备					
		服务名称 e	CC								
		模型ID 2	64								
		选择新版本	V1			\sim					
		ł	制认	1	取消						

新版本发布成功后,即可在「下发部署包到设备」页面或当前服务的「服务详情」页面,将新版本下发到本地设备上。

模型中心	端云协同服务 > 我的部署包
我的模型	端云协同服务说明 点击收款
创建模型	1、在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新
训练模型	 断网状态下模型离线计算(http服务,可调用与公有云API功能相同的接口) 联网状态下在平台管理设备运行状态、资源利用率
校验模型	具体使用流程如下:
发布模型	1、在本页面发布端云协同部署包,填写相关信息并申请,并在下方列表查看部署包发布状态
FasyData数据服务	2、任·按約本型比會」贝圓兩項设备,呈刺激活命で、時期激活命和設备 3、在「下发部署包到设备」页面,将发布成功的部署包一键下发到已激活的设备上,即可测试或正式使用

RX 306 /25 342	发布端云协问部署包
标签组管理	端云协同服务名称 模型ID 设备类型 应用平台 当前可用版本 新版本发布状态 部署设备数 操作
在线标注	ecc 264 通用小型设备 Linux-通用ARM V1 V1-已发布 0 下发到设备 发布新版本
云服务数据回流	246 服务器 Linux-通用X86 CPU V2, V1 V2-已发布 1 下发到设备发布新版本 服务详情
EasyEdge本地部署	图像分类高精度_磁狗-265 265 通用小型设备 Linux-通用ARM - V1-发布中 0
纯离线服务	
模型中心	端云协同服务 >
我的模型	
创建模型	
训练模型	
校验模型	下发到更多设备
发布模型	设备名称 最新下发模型版本 部署到期时间 设备连接状态 服务下发状态 最新同步时间 操作
EasyData数据服务	ubuntu-local-fly ⊖ V2 2021-02-06 在线 V2下发成功 2021-01-07 14:45 宣看服务配置 重看设备详情 下发新版本 移脉设备
数据总简	
经济的 第	
你亚坦管理	
在线标注	

管理模型已部署的设备

在上述的「服务详情」页面,可以查看并管理当前服务已部署的设备,包括移除设备、将服务下发到更多的设备等。

模型中心	端云协同服务 : > 服务详情
我的模型	服务名称 模型D 246 当前可用版本 V2, V1 部署设备数 1/1
创建模型	设备类型 服务器 应用平台 Linux-通用X86 CPU
训练模型	
校验模型	下及到更多议备
发布模型	设备名称 最新下发模型版本 部署到期时间 设备连接状态 服务下发状态 最新同步时间 操作
EasyData数据服务	ubuntu-local-fiy 💬 V2 2021-02-06 在线 V2下发成功 2021-01-07 14:45 宣者服务配置 宣者设备详情 下发新版本 网络设备
数据总览	
标签组管理	
在线标注	

本地设备管理

在我的本地设备页面可以进行所有本地设备的管理。

查看单台设备的运行状态

点击单台设备的「服务详情」,可查看设备上运行的多个服务及设备状态:

模型中心	端云协同服务 > 我的本地设备						
我的模型	我的本地设备						点击收起
创建模型	在本页面新增设备、联网激活 利用率等信息。	舌本地设备后,即可将「 🕇	的部署包」『	〔面发布成功的部署包一键下发	到设备上。设备联	网时,可以查看设备上部	署的服务、设备的运行状态、资源
训练模型							
校验模型	新增设备						
发布模型	设备名称	创建时间	设备类型	应用平台	设备连接状态	最新同步时间	操作
EasyData数据服务	123123 💬	2021-01-07 13:25	-		未激活	-	删除 激活设备 查看详情
数据总觉	showcase-test 💬	2020-12-15 17:57	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 19:22	激活设备 查看详情
标签组管理	linux-x86-zqw-2 😔	2020-12-15 16:24	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 19:42	激活设备 查看详情
在线标注	linux-x86-zqw 😳	2020-12-15 15:29	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-15 16:40	激活设备 查看详情
云服务数据回流	firefly_rk3399pro 💬	2020-12-15 14:39	-		离线 ⑦	2020-12-15 21:33	激活设备 宣看详情
EasyEdge本地部署	ubuntu-local-fjy 💬	2020-12-14 21:38	服务器	Linux-AMD64(x86-64)	在线	2021-01-07 15:01	查看详情
纯离线服务	hfl-1 💬	2020-12-14 19:24	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 14:18	激活设备 宣看详情
端云协同服务 ヘ	edge###	2020-12-11 14:52	-	-	未激活	-	删除 激活设备 查看详情

设备详情会展示当前设备的最新同步时间,以及CPU使用率、内存使用率等。服务列表则展示了当前设备上部署服务的运行情况和资源占用情况

模型中心	端云协同服务 > 我的本地设备 >	ubuntu-local-fjy				
我的模型	设备详情					
创建模型	设备名称 ubuntu-local-fjy	连接状态 在絕	Ę	实时刷新	OFF	
训练模型	设备类型 服务器	应用平台 Lin	ux-AMD64(x86-64)	最新同步时间 20	21-01-07 15:00	
校验模型	CPU使用率		内存使用率			
发布模型	31.1%		35.8%			
EasyData数据服务						
数据总览	端云协同服务详情					
标签组管理	服务名称	模型ID	CPU占比	内存使用情况	内存占比	操作
在线标注	2008-012	246	0.01%	156.7MB	0.93%	宣看服务配置
云服务数据回流						
EasyEdge本地部署						
纯离线服务						

申请延期端云协同部署包有效期为1个月,到期后可点击申请延期,工作人员审核后会增加测试时间

EasyEdge 🛛 😑	端云协同服务 > lzw_ResNet18_148_soc > 服务详情
^{很型中心} 全部模型 上传本地模型 体验开源模型	展务先称 Low_Readwartet4.4
與周续服务 离线计算SDK	设备农祚 最新下发模型版本 部署到期时间 设备连接状态 服务下发状态 最新同步时间 操作
端云场网座券 我的創筆包 我的本地设备 下发部署包到设备	bml-create-device 🕞 VI 2021-07-17 ③ 在线 VITF发成功 2021-10-21 14:29 室着服务危重 室着设备详例 <mark>中调成时</mark> 49转设备

心端云协同-JetsonNano部署文档

端云协同-Jetson系列硬件 端云协同支持的 Jetson 系列硬件包括 Jetson nano, Jetson TX2, Jetson AGX Xavier 和 Jetson Xavier NX。

准备环境

- 1. JetPack:安装JetPack 4.4 版本(目前端云协同仅支持 JetPack 4.4 版本),并注意在安装 JetPack 时务必安装对应的组件:
- 使用 SDK Manager 安装 JetPack 需要勾选 TensorRT、OpenCV、CUDA、cuDNN 等选项。
- 使用 SD Card Image 方式(仅对 Jetson Nano 和 Jetson Xavier NX 有效)则无需关心组件问题,默认会全部安装。
- 2. docker: 安装 nvidia-docker 或 docker(版本 >= 19.03),一般 Jetson 系列硬件自带操作系统都已包含 nvidia-docker 或 docker,可通过以下命令确认: // nvidia-docker
 nvidia-docker version
 // docker

docker version 修改启动参数 /etc/docker/daemon.json 文件,将"default-runtime"改为"nvidia",并重启 nvidia-docker 或 docker。



3. **依赖库文件**:下载文件 easyedge_runtime_j44.csv,并将该文件置于 Jetson 宿主机的 /etc/nvidia-container-runtime/host-files-forcontainer.d/ 目录内。

激活设备

- 1. 在 EasyDL/BML/EasyEdge 平台「端云协同服务」-「我的本地设备」页面新增设备。
- 2. 在设备列表中点击设备对应的「激活设备」操作,复制激活命令。
- 3. 在 Jetson 设备上,执行激活命令。激活过程中提示选择 containerd 或者 docker 时,选择 docker,如下图:

下发部署

1. 在端云协同服务-我的部署包页面点击发布端云协同部署包

图像分类模型	端云协同服务 > 我的部署包						操作文档 常见问题	新手教程 提交工单
图 总览								
◎ 模型中心	端云协同服务说明							点击收起
我的模型	1、在可视化界面轻松实现模型部	署包在边缘设备上的集	成、版本更新					
创建模型	 2、町网状心下候型高時订具(n) 3、联网状态下在平台管理设备 	mp服务,可调用与公有 i行状态、资源利用率	ZAMUN能相同的按L	IJ				
训练模型	具体使用流程如下:							
校验模型	 在本页面发布端云协同部署包 在「我的本地设备」页面新 	,填写相关信息并申请 增设备.复制激活命令	,并在下方列表查看着 ,	8署包发布状态				
发布模型	3、在「下发部署包到设备」页	面,将发布成功的部署	包一键下发到已激活的	9设备上,即可测试或正式	使用			
☑ EasyData数据服务							1	
数据总览						发布端云协同部署包	请输入服务名称	Q
标签组管理	端云协同服务名称	模型ID	设备类型	应用平台	当前可用版本	新版本发布状态	部要设备数	操作
在线标注	A0.44.09 (2006)20 (410)	DEELD.	MEAT	10470 T L1	200-770/A-+	WINK TO A TO TO BE	No. III OF IIII OF III OF IIII OF III OF III OF III OF III OF III OF IIII OF IIII OF IIII OF III OF IIII OF IIII OF IIII OF IIII OF IIII OF IIII OF III OF IIII OF IIIIII OF IIII OF IIII OF IIIII OF IIIIII OF IIII OF III	JAIF
智能标注								
云服务数据回流								
摄像头数据采集								
23 模型部署								
公有云服务					85 平 可用8	b 462		
EasyEdge 纯离线服务					M 7C PJ HI SC	1.50		
EasyEdge 端云协同服务 へ					请相后将认			
我的部署包								
我的本地设备								
下发部署包到设备								

2. 选择模型名称和版本,发布为专项适配硬件部署包

图像分类模型	端云协同服务 > 发布端云协同部署包	操作文档 常见问题 新手教程 提交工单
我的模型		
创建模型	设备类型 图务器 通用小型设备 💿 专项适配硬件	说明:
训练模型	模型名称 tost V	 在本地设备输入命令 docker version , 确认设备上是否安装Dockerd,如并未安 通动计算机的关键。
校 短 模 型 发 布 模 型	端云协同服务名称 test	2. 调整dockerd启动参数: default-
☑ EasyData数据服务	选择版本 V2 ~	runtime调整为nvdia。并重启dockerd。详 情请参考技术文档
数据总览	遗择系统 Linux专用SDK	3. 在运行激活时,选择containerd/docker 的步骤,选择docker
在线标注	发布部署包	
智能标注		
云服务数据回流		
摄像头数据采集		
12 模型部署		
公有云服务		
EasyEdge 纯离线服务		
EasyEdge 端云协同服务 へ		
我的部署包		
我的本地设备		
下发部署包到设备		
AI市场		
我的已购模型		
Ann		

3. 点击导航栏中的下发部署包到设备,即可将刚才生成的部署包下发到已经激活的设备当中

心 浏览器或小程序部署

 · 浏览器或小程序部署

浏览器或小程序部署

简介 本文档介绍EasyDL的浏览器/小程序部署SDK的使用方法,

SDK支持范围 浏览器部署

PC浏览器: Chrome、Safari、Firefox

手机浏览器: Baidu App、Safari、Chrome、UC and QQ Browser

小程序部署

小程序:百度小程序、微信小程序

支持的操作系统

系统: MacOS、Windows

demo文件结构 SDK解压缩之后,目录结构如下

public
model
model.json
chunk_n.dat
src
components
App.vue
config.json
env.d.ts
label.json
main.ts
modelInfo.json
usePredict.ts
index.html
package.json
READ.md
tsconfig.json
tsconfig.node.json
vite.config.ts
yarn.lock

demo基于vite,其中public/model下的model.json、chunk_1.dat...chunk_n.dat为模型文件,src下为业务代码,index.html为入口文件

快速开始 依赖node及npm,如果没有node,请前往node官网下载长期维护版本

启动项目:npm run dev

启动后控制台输出

vite v2.8.4 dev server running at:

> Local: http://localhost:3000/

> Network: use `--host` to expose

到浏览器打开 http://localhost:3000/ 即可体验demo

模型预测结果示例 图像分类示例

 $[0.4450492858886719,\, 0.3961234986782074,\, 0.0122891990467906,\, 0.14653800427913666]$

数组的index为对应的标签,值为置信度

物体检测示例

```
[[1, 0.2247152328491211, 0.11200979351997375, 0.07523892819881439, 0.8540866374969482, 0.5503567457199097], [2, 0.1224712328491211, 0.511200979351997375, 0.27523892819881439, 0.8540866374969482, 0.5503567457199097],...]
```

输出结果是一个二维数组,第二维的结果为:[标签,置信度,矩形框x1坐标,矩形框y1坐标,矩形框x2坐标,矩形框y2坐标]

浏览器开发

参考src/usePredict文件

```
// 加载推理引擎
import {Runner, env} from '@paddlejs/paddlejs-core';
// 使用webgl计算方案(暂不能使用wasm、webgpu等计算方案)
import '@paddlejs/paddlejs-backend-webgl';
...
// 注册引擎
const runner = new Runner({
    modelPath: '/model',
    keepRatio: config.rescale_mode === 'keep_ratio',
    mean: config.img_mean.reduce((memo, v) => [...memo, +((v / 255).toFixed(3))], [] as number[]),
    std: config.scale.reduce((memo, v) => [...memo, +((1 / 255 / v).toFixed(3))], [] as number[]),
    bgr: config.colorFormat === 'BGR',
    feedShape: {
       fw: config.resize[0],
       fh: config.resize[1]
    }
  });
...
// init runner
await runner.init();
// predict and get result
await runner.predict(img);
```

更多可参考PaddleJS工程页

小程序开发

微信小程序

微信小程序需添加 Paddle.js微信小程序插件 步骤: 小程序管理界面 --> 设置 --> 第三方设置 --> 插件管理 --> 添加插件 --> 搜索 wx7138a7bb793608c3 并添加

掌上百度小程序

手百小程序需添加paddlejs百度智能小程序动态库 引入动态库代码包 代码示例:

```
{

"dynamicLib": {

// 定义一个别名,小程序中用这个别名引用动态库。

"paddlejs": {

"provider": "paddlejs"

}

}
```

使用动态库

在使用页面的json文件里配置如下信息:

```
{
    "usingSwanComponents": {
        "paddlejs": "dynamicLib://paddlejs/paddlejs"
    }
}
```

从而页面中可以使用此组件:

```
<view class="container">
<view>下面这个自定义组件来自于动态库</view>
<paddlejs />
</view>
```

示例

index.swan

```
<view class="container">
<!--index.wxml-->
<image style="width:100%; height: 300px; " src="{{imgPath}}"></image>
<button bindtap="chooseImage">选择图片</button>
<button bindtap="chooseImage">选择图片</button>
<button bindtap="doPredict" class="btn" type="primary">新鲜度预测</button>
<l-- 返回结果 -->
<view class="result" s-if="resultType">预测结果 : {{resultType}}</view>
<view class="result" s-if="resultVal">预测可信度 : {{resultType}}</view>
<view class="result" s-if="resultVal">预测可信度 : {{resultVal}}</view>
</view class="result" s-if="resultVal">预测可信度 : {{resultVal}}</ri>
```

index.js

```
Page({
 data: {
    imgPath: '',
    content: ",
    resultType: ",
    resultVal: ",
    isShow: true,
    options: { // 模型配置项
      modelPath: 'http://localhost:3000/model',
      fileCount: 3,
      needPreheat: true,
      feedShape: {
         fw: 224,
         fh: 224
      },
      fetchShape: [1, 7, 1, 1],
      fill: [255, 255, 255, 255],
      scale: 256,
      targetSize: { height: 224, width: 224 },
      mean: [0.485, 0.456, 0.406],
      std: [0.229, 0.224, 0.225]
   },
    status: '' // 初始值为'', 变为'predict'时会触发模型预测
 },
```

/**

```
* 选择图片
   */
  chooseImage: function () {
     const me = this;
     this.setData({
        ishow: false
     });
     swan.chooseImage({
        count: 1,
        sizeType: ['original', 'compressed'],
        sourceType: ['album', 'camera'],
        success(res) {
          const path = res.tempFilePaths[0];
          swan.getFileSystemManager().readFile({
             filePath: path,
             encoding: 'base64',
             success: res => {
                me.setData({
                   imgBase64: res && res.data,
                   imgPath: path
                });
             },
             fail: res => {
                console.log(res);
             }
          });
        }
     });
  },
  predict(e) {
     const status = e && e.detail && e.detail.status;
     if (status === 'loaded') {
        this.setData({status: 'loaded', isShow: false});
     }
     else if (status === 'complete') {
        const data = e.detail.data;
        const maxItem = this.getMaxItem(data);
        this.setData({status: '', resultType: maps[maxItem.index], resultVal: maxItem.value});
     }
  },
  doPredict() {
     this.setData({status: 'predict'});
  },
  getMaxItem(datas = []) {
     let max = Math.max.apply(null, datas);
     let index = datas.indexOf(max);
     return {value: max, index};
  },
});
```

Prop

名称	类型	默认值	是否必 选	描述
options	string		是	模型配置项,参考src/usePridict
imgBase 64	string		是	要预测的图像的base64
status	string		是	当前状态,status变化触发组件调用相应的api,当status变为predict时,组件会读取imgBase64作为输入 的图像,调用模型预测APi

∞ 模型加速整体说明

功能简介 当您发布时纯离线服务时,平台已结合最新的量化、剪枝、蒸馏技术,推出丰富的模型压缩加速方案,以提高您的SDK部署效率。

覆盖范围:服务器、通用小型设备、专项适配硬件均支持该功能。

具体原理:针对目标芯片,对模型做深度优化压缩加速,加速后模型在推理速度、内存占用、体积大小等指标上表现更优。发布加速模型可能需要一段时间,同时会有微小的精度损失。发布完成后可通过性能报告对比具体加速效果。

使用流程 选择加速方式 结合选择的系统与芯片不同,分别为您提供不同的压缩方式。

纯离线服务 > 发布新服务		操作文档 教学视频 常见问题 提交工单
服务器 通用小型设备 专项适配硬件 通用小型设备 专项适配硬件 (2) 15 平今1,6 平日		记明: 1.本地服务器部署支持将模型部署于本地的CPU、GPU服务器 上,提供API和SDK两种集成万式: 查看文档
Apreeb WID2A Apreeb WID2A Apreeb WID2A Apreeb WID2A Apreeb WID2A Apreeb WID2A Apreeb WID2A		2.本地服务器SDK: 希欄型對表成适配本地服务器 (实持Linux 和Windows) 的SDK, 可集成在其他程序中运行。首次联网激 活后即可纯离线运行,占用服务器资源更少,使用方法更灵活
选择模型 dog-cat-test 选择版本 V2	× ×	3.集成步骤:①申请SDK并在服务详情页面下载SDK → ② 在 控的台申请激活序列号 → ③ 积极开发文信集成SDK, 并联网 激活使用。如存在设备无法联网,需要在纯高线的环境下激活 的情况,请提实工单联系我们。
选择系统和芯片 △ Linux 通用X86 CPU 英伟选GPU 华为 Atlas 300 百度 昆仑XPU 士 € Windows		4.个人信息的填写仅供EasyDL团队了解忽,便于后续合作接 洽,不会作为其他用途使用
模型加速 ⑦ 🗾 基础-无加速 天加速	□ 精度无损压缩加速 在積度尽可能无限的前提下加速模型	
○ 精度微揚历压缩加速-中 限时免费 在部分芯片上,内容/显存占用降低,推理 速度可以获得一定提升		
下—————————————————————————————————————		-

提示:基础SDK默认作为勾选项存在,可后续与您的加速SDK进行效果与性能比对,方便您进一步挑选

查看发布状态 点击完成发布后,将自动跳转至列表页,可分别查看不同加速方案下的模型发布进度及发布时间。

服务器 通用小型设	输入模型名称	Q							
SDK API									
此处发布、下载的SDK	此处发布、下载的SDK为未提权SDK,需要前往控制台获取序列号激活后才能正式使用。SDK内附有对应版本的Demo及开发文档,开发者可参考源代码完成开发。								
模型名称	发布版本 关	应用平台	模型加速	发布状态	发布时间 💲	操作			
	45045 1/0		基础版	•发布中	2021-05-13 20:49	下载SDK			
dog-cat-test	115215-V2	通用X86 CPU-Linux	精度无损压缩加速	•发布中	2021-05-13 20:49	下载加速版SDK			

常见问题

心数据相关问题

需要上传多少张图片才能训练出效果较好的模型?

• 每个分类至少需要准备20张以上。如果想要较好的效果,建议每个分类准备不少于100张图片。

上传图片的总量有限制吗?

• 每个账号下所有数据集的图片总数不能超过10万张。

心训练相关问题

数据处理失败或者状态异常怎么办?

 如是是图像分类模型上传处理失败,请先检查已上传的分类命名是否正确,是否存在中文命名、或者增加了空格;然后检查下数据图片量是否 超过上限(10万张);再检查图片中是否有损坏。如果自查没有发现问题,请在百度智能云控制台内提交工单反馈

模型训练失败怎么办?

• 如果遇到模型训练失败的情况,请先尝试重新训练,如多次重新训练后仍然失败,请在百度智能云控制台内提交工单反馈

已经上线的模型还可以继续优化吗?

 已经上线的模型依然可以持续优化,操作上还是按照标准流程在训练模型中-选择要优化的模型和数据完成训练,然后在模型列表中更新线上 服务,完成模型的优化

Step 1 重新训练

点击我的模型列表——找到需要重新训练的模型——点击训练,进行新版本模型训练

Step 2 重新发布

点击我的模型列表——找到新训练好的模型版本——点击申请发布

【图像分类】	百美椅子训	练 模型ID:230)			全部版本 删除
应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作
云服务	V2	训练完成	未申请	未发布	top1准确率87.61% top5准确率100.00% 完整评估效果	申请发布 校验 训练
离线SDK	V1	训练完成	未申请	未发布	top1准确率85.84% top5准确率100.00% 完整评估效果	申请发布 训练
						每页显示 12 🗸 < 1 🚿

Step 3 确认发布

在出来的弹窗中点击确定

我的模型		完整评估效果		
创建模型	申请发布		×	
训练模型	当前模型的线上服务为htt	tps://aip.baidubce.com/rpc/2.0/ai_custom/v1/classification/model_172,相应的模型版本为V	全部版本	删除
校验模型	是否确认要替换为V2			
发布模型		确定取消	训练 í	
改据中心 数据集管理	【图像分类】椅子颜色	查识别 模型ID: 2598	全部版本	删除

心 模型效果相关问题

如何通过「完整评估结果」里的错误示例优化模型?

- 错误示例中,左侧是正确的结果,右侧是模型的识别结果
- 观察模型识别有误的图片有哪些共同点,并有针对性地补充训练数据。比如:当图片比较亮的时候模型都能识别正确,但比较暗的时候模型就 识别错了。这时就需要补充比较暗的图片作为训练数据



我的数据有限,如何优化效果?

- 在训练配置页面-数据增强策略中配置更多数据增强的算子,来增加训练数据。也可在精度提升配置包-数据增强策略中选择自动数据增强策略,从而自动补充适合场景的增强数据
- 如果您是通过将模型发布为公有云服务进行应用,即可通过云服务数据管理功能,将实际调用云服务识别的图片加入训练集,不断迭代模型

实际调用服务时模型效果变差?

- 训练图片和实际场景要识别的图片拍摄环境应一致,举例:如果实际要识别的图片是摄像头俯拍的,那训练图片就不能用网上下载的目标正面 图片
- 每个标签的图片需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强
- 如果使用的是云服务,可以开通云服务数据管理功能,将实际调用云服务识别的图片加入训练集,不断迭代模型

**如果训练数据已经达到以上要求,且单个分类/标签的图片量超过200张以上,效果仍然不佳,请在百度智能云控制台内提交工单反馈

心模型上线/部署相关问题

每个账号可以上线几个模型?是否可以删除已上线的模型?

• 不限制发布模型数量,已上线模型无法删除

线上的部署方式支不支持我的硬件?

部署类型	支持的硬件示例
通用ARM	绝大多数安卓、苹果手机;瑞芯微RK32、RK32、RK35系列、树莓派等开发板
英特尔神经计算棒	NCS 1代、NCS 2代
海思NNIE	Hi3559AV100/Hi3559CV100等
华为昇腾Atlas开发板	Atlas200计算盒、Atlas300 计算卡
比特大陆SE计算盒	Bitmain SE5
通用x86CPU	绝大多数英特尔和AMD CPU
通用x86CPU加速版	英特尔志强、酷睿、凌动系列CPU
高通骁龙	骁龙660以后芯片的手机
华为NPU	mate10,mate10pro,P20,mate20,荣耀v20等
华为达芬奇NPU	mate30,p40,nova6,荣耀v30等
英伟达GPU	消费级显卡GeForce系列、RTX系列、TITAN,专业显卡Quadro、Tesla系列
英伟达Jetson	TX2、Nano、Xavier、Xavier NX

物体检测

整体介绍

心简介

Hi,您好,欢迎使用百度EasyDL图像。

EasyDL图像支持定制图像分类、物体检测、图像分割三类模型。三类模型的功能区别如下:

- 图像分类:识别一张图中是否是某类物体/状态/场景,适用于图片内容单一、需要给整张图片分类的场景
- 物体检测:检测图中每个物体的位置、名称。适合图中有多个主体要识别、或要识别主体位置及数量的场景
- 图像分割:对比物体检测,支持用多边形标注训练数据,模型可像素级识别目标。适合图中有多个主体、需识别其位置或轮廓的场景

以下是关于物体检测模型的技术文档。

∞ 应用场景

- 视频监控:如检测是否有违规物体、行为出现
- 工业质检:如检测图片里微小瑕疵的数量和位置
- 医疗诊断:如医疗细胞计数、中草药识别等

心 使用流程

训练模型的基本流程如下图所示,全程可视化简易操作。在数据已经准备好的情况下,最快几分钟即可获得定制模型。

新手教程详细介绍每一步的操作方式。如果文档没有解决您的问题,请在百度智能云控制台内提交工单反馈。



数据准备

心 创建数据集

在训练之前需要在数据中心【创建数据集】,导入并标注数据。

如果训练数据需要多人分工标注,可以创建多个数据集。将训练数据分批上传到这些数据集后,再将数据集"共享"给自己的小伙伴,同步进行标 注。

设计标签

在上传之前确定想要识别哪几种物体,并上传含有这些物体的图片。每个标签对应想要在图片中检测出的一种物体

注意:标签的上限为1000种

准备图片

基于设计好的标签准备图片:

- 每种要识别的物体在所有图片中出现的数量需要大于50
- 如果某些要区分的物体具有相似性,需要增加更多图片
- 一个模型的图片总量限制4张~10万张
- 单张图片中的目标数不能超过1000个

如有特殊需求,请提交工单联系我们

图片格式要求:

- 目前支持图片类型为png、jpg、bmp、jpeg,图片大小限制在14M以内
- 图片长宽比在3:1以内,其中最长边小于4096px,最短边大于30px

图片内容要求:

- 训练图片和实际场景要识别的图片拍摄环境一致,举例:如果实际要识别的图片是摄像头俯拍的,那训练图片就不能用网上下载的目标正面图 片;如果是需要识别白天光照下的物体,就不能使用夜晚拍摄的图片数据
- 每个标签的图片需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强

如果需要寻求第三方数据采集团队协助数据采集,请在百度智能云控制台内提交工单反馈

心上传数据集并在线标注

在完成了设计标签与准备数据后,可以通过以下方式导入数据:

- 导入未标注的数据,在线进行数据标注
- 直接导入标注好的数据

导入未标注数据

本地数据

支持上传图片、压缩包,或通过API导入

已有数据集

支持选择百度云BOS导入、分享链接导入、平台已有数据集导入;支持选择线上已有的数据集,包括其他图像类模型的数据集

EasyDL开发平台	< 返回 导入数据		
器 物体检测概览 ⇔			
数据服务			
◎ 数据总览	导入配置		
◎ 数据标注 へ	数据标注状态:	 无标注信息 有标注信息 	
标签组管理		SEVe 19	
方线标注	导人方式:	頃选择	
社会物力主	导入记录:	本地导入	
智能标注		BOS目录导入	
□ 数据增强		分享链接导入	
		平台已有数据集	
ご 数据回流 へ		公开数据集	
云服务回流		摄像头采集数据	
摄像头采集		云服务数据回流	

在线标注

上传未标注数据后,即可进入「标注数据集」页面进行在线标注

标注方法

注意:单张图片的标注框数不能超过500个,如有特殊需求,请提交工单联系我们



Step 1 首先在标注框上方找到工具栏,点击标注按钮在图片中拖动画框,圈出要识别的目标

Step 2 然后在右侧的标签栏中,增加新标签,或选择已有标签



标注技巧

• 所有图片中出现的目标物体都需要被框出 (框可以重叠)



全部框出

部分框出

• 框应包含整个物体,且尽可能不要包含多余的背景



包含整个物体

未框全或包含多余背景

- 如果图片中存在很多相同标签的目标物体,可以使用右侧的锁定按钮。锁定标签后,只需要在左侧框选目标物体即可,不用再重复选择标签
- 若需要标注的图片量较大时(如超过100张),可以启动智能标注来降低标注成本

导入已标注数据

本地数据

支持上传压缩包,或通过API导入

EasyDL开发平台	< 返回 导入数据			
嘂 物体检测概览 ≒				创建数据集
数据服务				Ŭ
■ 数据总览	导入配置			
② 数据标注 ^	数据标注状态:	○ 无标注信息 ● 有标注信息		
标签组管理	导入方式:	本地导入 🗸	上传压缩包	
在线标注				
智能标注	标注格式:	 ● 平台默认格式 ③ ○ xml (特指voc) ③ 	json (特指coco) ⑦	
□ 数据增强	上传压缩包:	土 上传压缩包		
②数据回流 ^	导入记录:	暂无导入记录		
云服务回流	4			
摄像头采集				

已有数据集

支持选择百度云BOS导入、分享链接导入、平台已有数据集导入

EasyDL开发平台	く 返回 导入数据	ŝ	
嘂 物体检测概览 ⇔			
数据服务			
◎ 数据总览	导入配置		
② 数据标注 へ	数据标注状态:	○ 无标注信息	
标签组管理	日入支式	通体探	
在线标注	47703.40	**** = 1	
智能标注	导入记录:	BOS目录导入	
		分享链接导入	
い 数据増強		平台已有数据集	
⊙数据回流 ∧		公开数据集	
云服务回流			
堪像斗亚隼			

使用智能标注功能可降低数据的标注成本。启动后,系统会从数据集所有图片中筛选出最关键的图片并提示需要优先标注。通常情况下,只需标 注数据集30%左右的数据即可训练模型。与标注所有数据后训练相比,模型效果几乎等同

整体流程以物体检测的智能标注流程为例:

创建智能标注任务

启动物体检测数据集的智能标注前,请先检查以下是否已满足以下条件:

- 所有需要识别的标签都已创建
- 每个标签的标注框数不少于10个
- 所有需要标注的图片都已加入数据集,且所有不相关的图片都已删除

若已满足,即可从导航栏进入「数据服务」-「智能标注」,创建智能标注任务,系统会基于您选择数据类型及数据量级,自动预估任务运行时长

Bai 岱大脑 EasyDa	ta	产品介绍	数据周	服务 使	使用文档	
EasyData数据服务	.≡	智能标注				
我的数据总览					创建智能标注任务	×
☞ 数据标注		智能标	注	1 由于系	统领法阁长需要一定驾徒的时间 建议在未标注阁长数十平100时启动 想塞标注效率:	右的
在线标注				2、每个标	·签的标注框数都达到10个时才可以启动功能;	
智能标注		创建智能	钻标注	3、不同的	为数据量级会影响您的任务时长,请参考 任务时长预估表。 ⑦	
多人标注 ^{new}		序号	娄灯:	数据集类型	物体检测 实例分割 文本分类单标签	
寻求标注支持		,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	2011	诜择数据集	xvf mask fake V1 V	
➡ 数据采集		1	31	A217 XX,107K	S (空) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1	武力
接入摄像头采集图片	\sim	2	27	[忽然地小小小总是当中336周围来在111次回,明晰的不够一切。但110/LT在831即公式10一, 您的智能标注任务预计需要38分钟完成	計
云服务数据回流	\sim	3	25			腹
L品 数据清洗		4	27		启动智能标注	动

系统筛选难例

系统会分批筛选出最关键需标注的图片,即难例图片。

Tips:难例筛选需要一定时间,在此期间您可以正常进行其他未标注图片的标注

曾能标注启动中,请稍候 启动过程需要大约5分转左右的时间,请耐心等待,了象 您的智能标注任务预计需要38分钟完成	珎艍 滓锖		
关闭智能标注			
1.启动智能标注	2.系统筛选难例	3.用户标注难例	4.完成标注,启动训练
在每个标签的已标注框数都达到10个后,启动智 能标注	根据您的数据集情况,系统自动筛选出数据集中 难例图片(最关键需优先标注的图片),进行自 动预标注	针对展示预标注结果的难例,您可通过两种方式 完成标注:第一,基于单张图片进行手动标注: 第二,在预标注效果符合预期基础上,选择一键 标注	您可以根据实际情况,选择系统为您进行下一轮 难例筛选(系统最多支持4轮难例筛选),或对 当前预标注效果满意进入立即训练

用户确认难例

智能标注任务启动后,系统为您自动筛选难例,您可以通过总览页查看进度按钮查看当前难例筛选进度,同时,进度图中也会全局展示您处于难 例筛选的具体哪一环节,以便您的操作后续。 筛选难例完成后,绿色进度条会进展到确认难例阶段,您可以点击【确认难例】完成对预标注结果 的人工确认。

创建智能	能标注任务						● 图像智能标注任务 ○ 文本智能标注任务
序号	数据集ID	数据集名称		版本	智能标注状态	操作	
1	3107	tj-智能标注-检测-demo		V2	已中止	重新启动	查看记录
2	2751	xyf_test_data1		V1	已中止	重新启动	查看记录
3	256	test123		V1	运行中	查看进度	难例确认 中止任务 查看记录
4	2744	zzy-测试智能标注	当前您处于第1轮难例阶段 已为您筛选出轮难例图片,请	(共4 确认该转	论) 论难例图片		查看记录
5	3111	tj-智能标注-检测-1110			• f:	E务完成	查看记录
6	2831	py3升级-智能标注-物体检测·	● ● 筛选难例 难例确认		效果评估	一轮筛选	查看记录
7	1965	物体检测-多人标注用				如満了	查看记录
8	2981	赵骞专属数据集1		••	L.70%	ᆂᆐᇛᄥ	查看记录

我们为您的人工确认提供两种模式:

- 单张确认,在该模式下支持您对预标注结果进行修正后点击保存
- 一键保存所有标注,为提升您的确认效率,默认您对难例的预标注结果全部满意,即可进入下一阶段

在线标注 > test123 > 标注			⇨ 智能标注中
有标注信息(22) 无标注信息(3) 符确认标注(5)		标签栏	Q
第1轮难例标注中(共4轮)	一键保存所有标注①		搜索标签
1、高击石下海【除今国制标运】 按预标注最黑特子规模认, 2 持您对标注得应后间除夺 2、您只有对【待确认标注】下所有预标注信果完成确认,所有难例均开级为已标状态,才可进入下一阶段	X.	smoke	
	标注结果 请在右侧选择标签	•	
	保存当前标注		

标注难例的预训练模型,也会对您无标注信息下的图片进行预标注结果的展示,您有余力的情况下,可以完成标注确认,确认后该张图片将升级 为已标状态,该环节并非是您进入智能标注下一阶段的必备要求。



评估难例效果,完成任务

当您对难例完成确认后,您可以根据本轮次预标注的结果是否满意,判断您是否还需要进入下一轮难例筛选阶段,如果满意本轮难例的预标注效果,系统将自动为您系统其他的未标图片打标签。

第1轮难例标注中(共4轮)

1、点击右下角【保存当前标注】该预标注结果将完成确认,支持您对标注修改后再保存
2、您只有对【待确认标注】下所有预标注结果完成确认,所有难例均升级为已标状态,才可进入下一阶段
🐼 该轮预标注难例已全部完成确认
1、若预标注效果未达预期,可通过更多轮次难例筛选实现效果优化,
进入下一轮难例筛选
2 芙菇坛计放用煤合颈期,修按照光前坛计放用,对剩余主坛图片进行自动标计
2、石顶标注双米行石顶期,符及照当前标注双米,对剩余水标图片近行日初标注, 确认自动完成标注

当您在任务运行中想要中止任务时,可实时点击标注页面右上方【中止任务】按钮,任务将被提前结束。

经典版-物化	确认中止该任务?	题 提交工单
定制物体检测模型,可	中止智能标注任务后,确认中的难例将不会生效。	-张图片中要识别多个物体,物体计数等场景中。
在各检测物体之间差界	确认 取消	号,训练时间可能需要30分钟以上。
线标注 > xj_智能标注性能 >	标注	中止智能标注 🗅 智能核

其他操作提示

• 在智能标注任务中,有任务上限吗?

支持五条智能标注任务同时运行,超过该上限您需要中止其他任务

• 智能标注中可以增删标签吗?

暂不支持。为了保证系统智能标注的效果,建议在启动功能前就创建好所有需要识别的标签 如果确实需要增删标签,可以先结束智能标注

• 智能标注中可以增删图片吗?

暂不支持。为了保证系统智能标注的效果,建议在启动功能前上传需要标注的所有图片,并删除不相关的图片。如果确实需要增删图片,可以先 结束智能标注

• 智能标注中可以修改已标注图片的标注框吗?

可以。但为了保证智能标注的效果,建议不要大量改动。如果确实需要修改大量标注,建议先结束智能标注

• 为什么我已经人工标注了很多图片,但系统预标注依然不准?

系统预标注的结果会受以下因素影响:智能标注期间,对"已标注"图片的标签进行大量改动;曾结束智能标注,并对标签、图片进行增删

• 多个数据集是否可以同时启动智能标注?

目前每个账号同一时间仅支持对一个数据集启动智能标注

• 共享中的数据集是否可以启动智能标注?

暂不支持。智能标注中的数据集也暂不支持共享

• 智能标注失败了怎么办?

可以先尝试稍后重新启动,如多次失败请提交工单联系我们

问题反馈

您在使用EasyData过程中可以通过以下任何方式联系我们:

● 在社区咨询

在论坛发帖提交问题,也可以在论坛与其他用户一起交流。前往论坛

• 提交工单

如果使用EasyData遇到其他任何问题或任何bug,您可以点此提交工单

• 添加微信小助手留言

请在微信搜索"BaiduEasyDL",并备注暗号 "EasyData",添加小助手后留言。 请在微信搜索"BaiduEasyDL",并备注暗号 "EasyData",添加小助手 后留言。

心 数据集多人标注

如果训练数据需要多人分工标注,可以创建多个数据集。将训练数据分批上传到这些数据集后,再将数据集"共享"给自己的小伙伴,同步进行标 注。

共享方式如下:

1. 在「数据集管理」页面,点击需要共享的数据集对应操作栏中的「共享」

模型中心	数据集管理					寻求数据支持:百度众测数据服务商
我的模型	创建数据集 数据上传API					
创建模型	ID 名称	类型	标签数	图片数	状态	操作
训练模型 校验模型	44339 MM豆 2	物体检测	1	60	正常	查看标注/上传智能标注历史制除 株字
发布模型	38867 更上层楼 🖉	物体检测	2	23	正常	查看 标注/上传 智能标注历史 删 除 共享
数据中心	27672 new Ø	物体检测	0	0	新建	查看 标注/上传 删除 共享
数据集管理	9798 super_band_2 🖉	物体检测	3	107	正常	查看 标注/上传 删除 共享
创建数据集	8994 super_band 🖉	物体检测	3	7638	正常	查看 标注/上传 删除 共享详情
数据标注/上传 云服务调用数据	8096 马 //	物体检测	2	20	正常	查看 标注/上传 删除 共享
						每页显示 10 🗸 < 1 >

2. 在共享页面,勾选被共享数据集的授权使用范围,生成共享链接。如需被共享人标注数据,则需勾选「修改」



3. 复制共享链接,并发送给小伙伴



4. 被共享人打开链接后,即可在「数据集管理」页面看到被共享的数据集,并进行被授权的操作

5. 训练模型时,在「训练模型」页面添加训练数据时,可从多个数据集(如多个被共享的数据集)选择数据

模型中心	训练模型								
我的模型	选择模型: super_band	~							
创建模型	应用类型: 🔽 云服务								
训练模型	选择算法: 〇 高	 透择算法: ○ 高精度 ⑦ ● 高性能 ⑦ ○ 高线服务 ⑦ 							
校验模型	□ 离线服务 2								
发布模型	参与训练的标签	参与训练的标签							
数据中心	数据集	标签数量	操作						
数据集管理	super_band	3	查看详情 清空标签						
创建数据集	super_band_2	3	查看详情 清空标签						
数据标注/上传	继续添加 全部清空								
云服务调用数据	你已经选择2个数据集的6个标签								
		开始训	练						

^の数据集管理API

本文档主要说明当您线下已有大量的已经完成标注的图片数据,如何通过调用API完成图片及标注的便捷上传和管理。 EasyDL图像数据集管理API 在管理不同模型数据类型之间是通用的。上传不同模型类型数据,只是在部分接口入参存在差异,使用及接口地址完全一致。

数据集创建API

接口描述

该接口可用于创建数据集。

接口鉴权

同模型上线后获取的API:

1、在EasyDL控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/create

URL参数:

参数	值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

ieddolyff i

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset_na me	是	string	数据集名称,长度不超过20个utf-8字符

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
dataset_id	否	number	创建的数据集ID

数据集列表API

接口描述

该接口可用于查看数据集列表。返回数据集的名称、类型、状态等信息。

接口鉴权

同模型上线后获取的API:

1、在EasyDL控制台-公有云部署-应用列表页面创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/list

URL参数:

		_	 	
参数	值			

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
start	否	number	起始序号,默认为0
num	否	number	数量,默认20,最多100

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	数据集总数
results	否	array(object)	数据集列表
+dataset_id	否	number	数据集ID
+dataset_name	否	string	数据集名称
+type	否	string	数据集类型
+status	否	string	数据集状态
+special_status	否	string	数据集特殊状态,包括shared、smart和空值,分别表示共享中、智能标注中、非特殊状态

分类 (标签)列表API

接口描述

该接口可用于查看分类(标签)。返回分类(标签)的名称、包含数据量等信息。

接口鉴权

同模型上线后获取的API:

1、在EasyDL控制台-公有云部署-应用列表页面创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法:POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/list

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID
start	否	number	起始序号,默认0
num	否	number	数量,默认20,最多100

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id ,用于问题定位
total_num	否	number	标签总数
results	否	array(object)	标签列表
+label_id	否	string	标签/分类ID
+label_name	否	string	标签/分类名称
+entity_count	否	number	图片/声音/文本数量

添加数据API

接口描述

该接口可用于在指定数据集添加数据。

接口鉴权

同模型上线后获取的API:

1、在EasyDL控制台-公有云部署-应用列表页面创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/addentity

URL参数:

参数 值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数
字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID
append Label	否	boolea n	确定添加标签/分类的行为:追加(true)、替换(false)。默认为追加(true)。
entity_c ontent	是	string	type为 IMAGE_CLASSIFICATION/OBJECT_DETECTION/IMAGE_SEGMENTATION/SOUND_CLASSIFICATION时,填入图 片/声音的base64编码;type为TEXT_CLASSIFICATION时,填入utf-8编码的文本。 内容限制为:图像分类base64前 10M;物体检测base64前10M;图像分割base64前10M;声音分类base64前4M,声音时长1~15秒;文本分类 10000个汉字
entity_n ame	是	string	文件名
labels	否	array(o bject)	标签/分类数据。若为空,则只上传图片,不上传标签/分类。若不为空,则应在数组中包含以下前面带+的参数
+label_ name	是	string	标签/分类名称(由中文、数字、字母、中划线、下划线组成),长度限制20B
+left	否	number	物体检测时需给出,标注框左上角到图片左边界的距离(像素)
+top	否	number	物体检测时需给出,标注框左上角到图片上边界的距离(像素)
+width	否	number	物体检测时需给出,标注框的宽度(像素)
+height	否	number	物体检测时需给出,标注框的高度(像素)

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

数据集删除API

接口描述

该接口可用于删除数据集。

接口鉴权

同模型上线后获取的API:

1、在EasyDL控制台-公有云部署-应用列表页面创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/delete

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值

Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

分类 (标签) 删除API

接口描述

该接口可用于删除分类(标签)。

接口鉴权

同模型上线后获取的API:

1、在EasyDL控制台-公有云部署-应用列表页面创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/delete

URL参数:

参数	值	

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset_ id	是	number	数据集ID
label_na me	是	string	标签/分类名称

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

```
例如Access Token失效返回:
```

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度智能云控制台内提交工单反馈
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度智能云控制台内提交工单反馈
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量 阶梯计费;未上线计费的接口,请在百度智能云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计 费;未上线计费的接口,请在百度智能云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶 梯计费;未上线计费的接口,请在百度智能云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
406000	internal server error	服务错误
406001	param[xx] invalid	参数xx不合法,请检查相关参数
406002	dataset not exist	数据集不存在
406003	dataset already exists	数据集已存在
406004	dataset can not be modified temporarily	数据集暂不可修改
406005	label not exist	标签/分类不存在
406006	no permission to modify the dataset	没有修改数据集的权限
406007	dataset cannot be modified while smart annotation is running	智能标注期间不可修改数据集
406008	quota exceeded	配额超限

心 数据质检

功能概述 该功能旨在对您数据集中的图像数据进行质量检测,通过提供客观指标,为您对数据集的下一步操作(标注、清洗等)进行参照引导。

整体质检报告将包括对原图、标注信息两个层面的指标进行统计,本期先上线原图维度的质检指标,标注层面的质检指标敬请期待。

使用流程 Step 1 功能入口

您可从数据总览页操作列点击【质检报告】或查看页面点击【质检报告】进入该功能页面

奴掂重 5	取 <u></u> <u></u> 一一一一一一一一一一一一一一一一一一一一一一一一一一一一一一一一一一	你注尖型	你注认念	府		
5 < 返回 数据集详情 ③	● 巳元成	图像分尖	0% (0/5)	-	互有与标注 守山 删除 [贞位妆古] 94 86588 8888	i±

Step 2 指标查看本期报告分为整体指标和分布指标两类。整体指标包括数据集存储大小、图片数量、破损图像数三类;分布指标包括色彩分布 空间、图像存储大小分布、高宽比分布、分辨率分布、色偏分布五类。

可以通过切换数据集版本查看不同版本下质检报告。



Step 3 对应处理 可通过hover具体指标数值进行相关操作,以高宽比分布为例:

第一步,高宽比在0.4-0.8的标注框,hover显示有6张图片,支持点击



第二步,点击后进入符合该指标的图片操作页,可针对筛选后图片进行删除、标注等操作

基本信息 数据集大小: 7.18 MB 标注框总数: 84 张 当前标注情况: 84张已标注	型片教室: 84 後 福田四片教室: 0 ① 8 平均県注意: 1 ② 已時医療合社: 100%
原图推度 标注推度	标注相高宽比分布列表 (6张) ▼ □
标注框离宽比分布 ⊙ 60	
40 框 鼓 量	< 1 > 148/A
20	

模型训练

の 物体检测创建模型

在导航【模型训练】中,点击训练模型,填写模型名称、所属行业、应用场景等信息,即可进入数据准备环节操作示例:

EasyDL开发平台	训练模型				
器 物体检测概览 ≒ 数据服务				1 模型准备	 — ③ 训练配置
日 数据总览	模型准备				
标签组管理	模型选择:	 创建新模型 已有模型 	2		
在线标注	标注模版:	矩形框 ⑦ 自定	⊻四边形 ⑦		
留能标注 口 数据增强			-		
⊙数据回流 ^	•模型名称:	请输入名称	0/20		
云服务回流 摄像头采集	•所属行业:	请选择行业			
模型管理	*应用场景:	请选择应用场景			
◎ 模型训练	•业务描述:	请输入业务描述			

注:1. 创建模型后可持续新增模型版本,因此不必每次训练模型都创建模型 2.目前单个用户在每种类型的模型下最多可创建10个模型,每个模型均支持多次训练。 3.如果您是企业用户,建议您按照真实企业信息进行填写,便于EasyDL团队后续更好的为您服务

心物体检测训练操作说明

数据提交后,可以在导航中找到【训练模型】,按以下步骤操作,启动模型训练:

注:1.启动训练前请确保数据已经标注完成,否则无法启动训练

2. 下述训练功能点中,标注为星号(*)的功能为非必要选择项,可根据实际需求考虑是否使用

物体检测模型	Ē	训练模型		操作文档
88 总宽				
凸 模型中心		边律模型	都是捕导捕导发 ∨	
我的模型		训练配置		
创建模型		部署方式	公有云部署 EasyEdge本地部署 如何选择部署方式?	
训练模型		训练方式	<mark>税募除费</mark> 常规训练 精度提升配置包	
校验模型				
发布模型		增量训练		
🖉 EasyData数	据服务		透择该模型下的已训练版本,在数据应用场景不变的情况下,扩充数据来进行增量训练,提升模型训练效率和精度 什么是增量训练	
數据总览		选择算法	○ 百度超大规模预训练模型-通用场景 ⑦	
标签组管理		选择网络 ⑦	2000 1500 1000 500 300 200 有皮最高 NEW 住能最高	
在线标注		自动超參搜索 ⑦		
云服务数据[回流	高级训练配置		
摄像头数据	采集	in the second second		
◎ 公有云服务		添加数据		
在线服务		添加数据集	+ 请选择	
批量预测			数据集 版本 标签数量	操作
Z EasyEdge本	地部署		aircraft_train V1 1	移除
纯离线服务				
端云协同服务	В ^	自定义验证集 ⑦		
我的部署自	<u>n</u>	选择验证集	+ 请选择	
我的本地讨	设备		验证集标签应与训练集完全一致。为获得效果更好的模型,验证集数据需更贴近实际应用场景	
			alight 117-1 1-56-41-0	477.74

① 选择模型

选择此次训练的模型 ② 添加数据

半监督训练*半监督深度学习是半监督学习和深度学习结合的产物,可以理解为在深度学习算法中使用无标签样本。

模型取得优异表现离不开大量有标记样本。在现实生活中,有标记样本获取代价高昂,而无标签样本却很容易获得。由此想把半监督学习引入到 深度学习中。

当打开半监督训练开关后,可以将未标注的数据添加至训练数据中参加训练。同时,这些未标注的数据在半监督训练完成后将会自动生成对应的标签信息,如在「保存自动生成标签」字段下选择了"是",则可在EasyData数据服务对应数据集中查看并确认对应的标签结果

注:开启半监督训练后会增加部分训练时间,一般不会大于对应全量标注数据训练的训练时间两倍,请根据实际需求考虑后选择。例如「80

已标注样本+20未标注样本」半监督训练与「100已标注样本」常规训练的训练时间对比,前者训练时间会更长,但不会大于后者训练时长的 两倍

*添加数据集 + 请选择

数据集 版本 分类数量 数据量 标注状态 操作 testss V1 1 1 已标注 移除 testss V1 1 38 未标注 移除						
testss V1 1 1 已标注 移除 testss V1 1 38 未标注 移除	数据集	版本	分类数量	数据量	标注状态	操作
testss V1 1 38 未标注 移除	testss	V1	1	1	已标注	移除
	testss	V1	1	38	未标注	移除

保存自动生成标签 ⑦ 📀 是 🛛 🔿

添加训练数据

- 先选择数据集,再按标签选择数据集里的图片,可从多个数据集选择图片
- 训练时间与数据量大小有关,1000张图片可能需要几个小时训练,请耐心等待

Tips :

• 如果包含同一个标签的数据分散在不同的数据集里,可以在训练时同时从这些数据集里选择,模型训练时会按标签名称合并

添加自定义验证集*AI模型在训练时,每训练一批数据会进行模型效果检验,以某一张验证图片作为验证数据,通过验证结果反馈去调节训练。 可以简单地把AI模型训练理解为学生学习,训练集则为每天的上课内容,验证集即为每周的课后作业,质量更高的每周课后作业能够更好的指导 学生学习并找寻自己的不足,从而提高成绩。同理AI模型训练的验证集也是这个功效。

注:学生的课后作业应该与上课内容对应,这样才能巩固知识。因此,验证集的标签也应与训练集完全一致。 添加自定义测试集*如果学生的期末考试是平时的练习题,那么学生可能通过记忆去解题,而不是通过学习的方法去做题,所以期末考试的 试题应与平时作业不能一样,才能检验学生的学习成果。那么同理,AI模型的效果测试不能使用训练数据进行测试,应使用训练数据集外的 数据测试,这样才能真实的反映模型效果。

注:期末考试的内容属于学期的内容,但不一定需要完全包括所学内容。同理,测试集的标签是训练集的全集或者子集即可。

配置数据增强策略

深度学习模型的成功很大程度上要归功于大量的标注数据集。通常来说,通过增加数据的数量和多样性往往能提升模型的效果。当在实践中无法 收集到数目庞大的高质量数据时,可以通过配置数据增强策略,对数据本身进行一定程度的扰动从而产生"新"数据。模型会通过学习大量 的"新"数据,提高泛化能力。

你可以在「默认配置」、「手动配置」、「自动数据增强」3种方式中进行选择,完成数据增强策略的配置。

默认配置

如果你不需要特别配置数据增强策略,就可以选择默认配置。后台会根据你选择的算法,自动配置必要的数据增强策略。

手动配置

EasyDL提供了大量的数据增强算子供开发者手动配置。你可以通过下方的算子功能说明或训练页面的效果展示,来了解不同算子的功能:

算子名	功能
ShearX_BBox	剪切图像的水平边
ShearX_Only_BBoxes	剪切标注框内图像的水平边
ShearY_BBox	剪切图像的垂直边
ShearY_Only_BBoxes	剪切标注框内图像的垂直边
TranslateX_BBox	按指定距离(像素点个数)水平移动图像及标注框
TranslateX_Only_BBoxes	按指定距离(像素点个数)水平移动标注框内的图像
TranslateY_BBox	按指定距离(像素点个数)垂直移动图像及标注框
TranslateY_Only_BBoxes	按指定距离(像素点个数)垂直移动标注框内的图像
Rotate_BBox	按指定角度旋转图像及标注框
Rotate_Only_BBoxes	按指定角度旋转标注框内的图像
AutoContrast	自动优化图像对比度
Contrast	调整图像对比度
Equalize	将图像转换为灰色值均匀分布的图像
Equalize_Only_BBoxes	将标注框内的图像转换为灰色值均匀分布的图像
Solarize	为图像中指定阈值之上的所有像素值取反
Solarize_Only_BBoxes	为标注框内的图像中指定阈值之上的所有像素值取反
Solarize_add	为图像中指定阈值之下的所有像素值加上像素偏移值
Posterize	减少每个颜色通道的bits至指定位数
Color	调整图像颜色平衡
Brightness	调整图像亮度
Sharpness	调整图像清晰度
Cutout	通过随机遮挡增加模型鲁棒性,可设定遮挡区域的长宽比例
BBox_Cutout	通过在标注框附近进行随机遮挡增加模型鲁棒性,可设定遮挡区域的长宽比例
Cutout_Only_BBoxes	只在标注框内通过随机遮挡增加模型鲁棒性,可设定遮挡区域的长宽比例
Flip_Only_BBoxes	对标注框内的图像进行左右翻转

效果展示

 \times

剪切图像的垂直边,能更好地识别发生了垂直方向形变的图像



自动数据增强

在训练方式选择「精度提升配置包」选项后,此处数据增强策略提供「自动数据增强」选项。自动数据增强算法会根据您数据的特性,自动选择 数据增强算子。使用付费机型训练的用户请注意,自动数据增强算法可能会增加模型训练时间。

模型训练完成后,可在「我的模型-查看版本配置」中,查看配置记录:

【物体检测】zh_d	etect_defaul	t 模型ID:10065				吕 训练 🕒 历史版本 🔟 删除
部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作
公有云API	V4	训练完成	未申请	未发布	mAP 80.06% ? 精确率 75.00% ? 召回率 85.71% ? 完整评估结果	<u> 宣看版本配置</u> 申请发布 校验

配置建议

算子的配置建议贴合实际场景。

比如,数字识别的数据集中,因为对数字的旋转很有可能导致错误样本的产生,所以不建议对数字数据集进行旋转操作。再比如,检测数据集中,如果标注量比较少,就可以通过随机平移的算子增强数据集,模型也更容易学习到目标物体的平移不变性。

③ 训练配置

部署方式

可选择「公有云API」、「EasyEdge本地部署」

如何选择部署方式

选择设备如果您选择了「EasyEdge本地部署」,请根据实际部署设备选择-如果您选择了「公有云API」,则可按需选择训练方式**增量训练***增 量训练:在模型迭代训练时,用户在原训练数据上增加了训练数据,可通过加载原训练数据训练的模型参数进行模型训练。这样可让模型收敛速 度变快,训练时间变短,同时在数据集质量较高的情况下,可能获得的模型效果也会更好。

注: 仅可选择同一部署方式下的训练的模型作为基准模型版本

训练方式

- 「常规训练」包括EasyDL历史提供的「高精度」、「高性能」等模型选择,以及常规的模型训练配置
- 「精度提升配置包」选用百度自有超大规模预训练模型,让模型有更好的精度效果。并提供按云调用时延选择网络模型的形式,根据您实际应用场景需求,选择更合适的模型。

自动超参搜索* 自动超参搜索目前仅在精度提升配置包的选项下提供。选择开启自动超参搜索后,算法会多次实验,自动搜寻出适合模型训练的 各种参数,来达到高精度的模型效果。

注:开启自动超参搜索后会增加3倍以上的训练时间,请根据实际需求考虑后选择

高级训练配置*高级训练配置开关默认关闭,建议对深度学习有一定了解的用户根据实际情况考虑使用。高级训练配置目前提供「输入图片分辨率」、「epoch」、两个配置项

- 输入图片分辨率:可以根据具体应用场景选择输入图片分辨率,如检测目标在图片中较小,就可适当增加输入图片分辨率,增强检测目标在数据层面的特性。推荐值为该类算法任务输入图片分辨率普遍最优值。
- epoch:训练集完整参与训练的次数。如有训练数据集较大,模型训练不充分,模型精度较低的情况,可适当设置较大epoch值(大于100), 使模型训练更完整。

选择算法

不同的部署方式下,可以选择不同的算法。每个算法旁边有一个小问号,可以查看详细说明。

例如:选择「公有云API」后,可以在「超高精度」、「高精度」、「高性能」3种算法中选择。鼠标移动到「高精度」右侧的问号上,可以看到 对高精度算法的详细说明。

イ ジレ ^ン	产品介	绍 操作平台	应用案例	使用文档
物体检测模型	€	训练模型		
品 总览				
由 模型中心		选择模型	水果检测	
我的模型		训练配置		
创建模型		部署方式	公有云部署	EasyEdge本地部署 如何选择部署方式?
训练模型		训练方式	常规训练	段时免费 精度提升配置句
校验模型			1000000	
发布模型		选择算法	● 超高精度	② 高精度 ② 高性能 ② AI市场已购模型
🖉 EasyData数据服务		高级训练配置		

通常,高精度模型在识别准确率上表现较好,但在识别速度上表现较弱。高性能模型反之。

在「精度提升配置包」中提供「小目标检测」算法供用户选择,当检测目标小于图片的5%,使用小目标检测算法可获得效果不错的模型。

注:小目标检测算法目前仅支持本地服务器部署

<i>┥</i> チーン ど楽EasyDL	产品	介绍 操作平台	应用案例	使用文档
物体检测模型	Ē	训练模型		
品 总览				
山 模型中心		选择模型	水果检测	~
我的模型		训练配置		
创建模型		部署方式	公有云部署	EasyEdge本地部署 如何选择部署方式?
训练模型		选择设备	○ 服务器	
校验模型				
发布模型		训练方式	常规训练	精度提升配置包
❷ EasyData数据服务		选择算法	小目标检测	1 2 当检测目标小于图片的5%时,使用小目标检测算法可提升模型识别效果
数据总览		添加数据		
标签组管理		添加数据集	+ 请选择	
在线标注				

此外,如果你已从AI市场购买了模型算法,也可以基于已购模型的算法训练:前往AI市场购买>

④ 训练模型

点击「开始训练」,训练模型。

- 训练时间与数据量大小有关,1000张图片可能需要几个小时训练,请耐心等待。
- 模型训练过程中,可以设置训练完成的短信提醒并离开页面。
- 平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

⊙物体检测模型效果评估

可通过模型评估报告或模型校验了解模型效果:

- 模型评估报告:训练完成后,可以在【我的模型】列表中看到模型效果,以及详细的模型评估报告。
- 模型在线校验:可以在左侧导航中找到【校验模型】,在线校验模型效果。校验功能示意图:

校验模型							
选择模型:	super_band V	应用类型: 云服	务(目前仅支持)	∨ 选择版	i本: V7	~	
当前模型mAPA	平均精度 89.13% 评估报告				识别结果 如何优化效果?		
					调整阈值	——— 当前阈值: 0.3	
2		-			预测标签	置信度>30%	
T			-		1. person	98.41%	
					2. positive	94.63%	
点击添加图	<u>н</u>				申请上线	纠正识别结果	

模型评估报告

整体评估

在这个部分可以看到模型训练整体的情况说明,包括基本结论、mAP、精确率、召回率。这部分模型效果的指标是基于训练数据集,随机抽出部 分数据不参与训练,仅参与模型效果评估计算得来。所以**当数据量较少时(如图片数量低于100个),参与评估的数据可能不超过30个,这样得** 出的模型评估报告效果仅供参考,无法完全准确体现模型效果。

注意:若想要更充分了解模型效果情况,建议发布模型为API后,通过调用接口批量测试,获取更准确的模型效果。

整体评估



查看模型评估结果时,需要思考在当前业务场景,更关注精确率与召回率哪个指标。是更希望减少误识别,还是更希望减少漏识别。前者更需要 关注精确率的指标,后者更需要关注召回率的指标。同时F1-score可以有效关注精确率和召回率的平衡情况,对于希望准确率与召回率兼具的场 景,F1-score越接近1效果越好。评估指标说明如下

F1-score: 对某类别而言为精确率和召回率的调和平均数,评估报告中指各类别F1-score的平均数

mAP: mAP(mean average precision)是物体检测(Object Detection)算法中衡量算法效果的指标。对于物体检测任务,每一类object都可以计算出 其精确率(Precision)和召回率(Recall),在不同阈值下多次计算/试验,每个类都可以得到一条P-R曲线,曲线下的面积就是average

精确率: 正确预测的物体数与预测物体总数之比。评估报告中具体指经比较F1-score最高的阈值下的结果

召回率:正确预测的物体数与真实物体数之比。评估报告中具体指经比较F1-score最高的阈值下的结果

模型调优建议在模型评估中,EasyDL将会通过智能算法对误识别的样本进行归因分析,可推断出误识别的样本对某个模型评估指标的具体影响 以及影响程度,并提供对应优化的方案。同时还可针对某个具体表现不好的标签进行归因分析,针对性优化识别效果

模型调优建议

归因粒度	基于整个模型	基于单个标签

序号	受影响指标	影响程度	根因分析	调优对策
1	F1-Score	ф	"高宽比"对"F1-Score"的效果有"一 定"影响,不同特征区间的"F1-Score" 方差达到"0.0352"	在【添加数据】->【数据增强策略】 中配置"ShearX,ShearY"进行增强。
2	F1-Score	 ф	"分辨率"对"F1-Score"的效果有"一 定"影响,不同特征区间的"F1-Score" 方差达到"0.0321"	在【添加数据】->【数据增强策略】 中配置"ShearX,ShearY"进行增强。
3	F1-Score	中	"色偏"对"F1-Score"的效果有"一定" 影响,不同特征区间的"F1-Score"方差 达到"0.0205"	在【添加数据】>【数据增强策略】 中配置"Color,Posterize"进行增强。
4	F1-Score	中	"亮度"对"F1-Score"的效果有"一定" 影响,不同特征区间的"F1-Score"方差 达到"0.0188"	在【添加数据】>【数据增强策略】 中配置"Brightness"进行增强。
5	F1-Score	ф	"饱和度"对"F1–Score"的效果有"一 定"影响,不同特征区间的"F1–Score" 方差达到"0.018"	在【添加数据】>【数据增强策略】 中配置"Color"进行增强。

详细评估

在这个部分可以看到不同阈值下的F1-score、模型识别错误的图片示例,以及使用混淆矩阵定位易混淆的标签。

详细评估

不同阈值下F1-score表现



不同标签的mAP及对应的识别错误的图片

识别错误图片示例

通过分标签查看模型识别错误的图片,直寻找其中的共性,进而有针对性的扩充训练数据;或发现是标注错误,从而直接点击修改标注来将标注 修正

如下图所示,可以通过勾选「误识别」、「漏识别」来分别查看两种错误识别的情况:

错	错误详情	×
<	原标注结果 ① 模型识别结果 ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●	
	■ 正确识别 • 误识别 • 漏识别 🗌 俗改标注 如何解读错误示例? 🗹 正确识别 🗹	误识别 ⑦ 🔽 漏识别 ⑦

• 误识别:红框内没有目标物体(准备训练数据时没有标注),但模型识别到了目标物体

观察误识别的目标有什么共性:例如,一个检测电动车的模型,把很多自行车误识别成了电动车(因为电动车和自行车外观上比较相似)。这时,就需要在训练集中为自行车特别建立一个标签,并且在所有训练集图片中,将自行车标注出来。

可以把模型想象成一个在认识世界的孩童,当你告诉他电动车和自行车分别是什么样时,他就能认出来;当你没有告诉他的时候,他就有可能把 自行车认成电动车。

• 漏识别:橙框内应该有目标物体(准备训练数据时标注了),但模型没能识别出目标物体

观察漏识别的目标有什么共性:例如,一个检测会议室参会人数的模型,会漏识别图片中出现的白色人种。这大概率是因为训练集中缺少白色人种的标注数据造成的。因此,需要在训练集中添加包含白色人种的图片,并将白色人种标注出来。

黄色人种和白色人种在外貌的差别上是比较明显的,由于几乎所有的训练数据都标注的是黄色人种,所以模型很可能认不出白色人种。需要增加 白色人种的标注数据,让模型学习到黄色人种和白色人种都属于「参会人员」这个标签。

以上例子中,我们找到的是识别错误的图片中,目标特征上的共性。除此之外,还可以观察识别错误的图片在以下维度是否有共性,比如:图片 的拍摄设备、拍摄角度,图片的亮度、背景等等。

定位易混淆标签

支持按识别错误样本量的绝对数值/相对数值查看混淆矩阵,同时支持下载完整的混淆矩阵进行更深入的分析。

定位易混沸	位易混淆标签								
下方是Aut	⇒方是Auto2022031600-32-46 VI模型的混淆矩阵 ,每一个橙色的方格都对应一组易混淆的标签(最多展示10个易混淆的标签)								
展示了数据标	展示了致趣标注与模型预测不符数量前00分标签。点击标签可在下方重看示例图、帮助您有针对性地设计特征、使得类别更真区分性。 上下载完整混								
序号	标签名称	误识别标签TOP5及其数	2	精确率	測试集数量	召回率	f1-score		
1	Cacca	[8] 背景类		100.0	% 10	20%	33%		
2	Lopi			25.0	% 2	100%	40%		
3	Caccg	[20] 背景类	[1] Clksg	93.2	% 72	76%	84%		
4	Ssdwl	[14] Ssdwr		87.7	6 83	86%	87%		
5	Clksa	[2] 背景类		100.0	% 10	80%	89%		

心物体检测模型如何提升效果

一个模型很难一次性就训练到最佳的效果,可能需要结合模型评估报告和校验结果不断扩充数据和调优。

为此我们设计了模型迭代功能,即当模型训练完毕后,会生成一个最新的版本号,首次V1、之后V2……以此类推。可以通过调整训练数据和算法,多次训练,获得更好的模型效果。

注意:如果模型已经是上线状态(包括已付费的模型服务),依然支持模型迭代。只需要在训练完毕后发布新的版本,就可以获得更新后的模型 服务。

想要提升模型效果,可以尝试以下两种方法:

检查并优化训练数据

1. 检查是否存在训练数据过少的情况,建议每个标签标注50个框以上,如果低于这个量级建议扩充。

- 2. 检查不同标签的标注框数量是否均衡,建议不同标签的标注框数数据量级相同,并尽量接近,如果有的标签框数很多,有的标签框数很少, 会影响模型整体的识别效果。
- 3. 通过**模型效果评估报告中的错误识别示例**,有针对性地扩充训练数据。
- 4. 检查测试模型的数据与训练数据的采集来源是否一致,如果设备不一致、或者采集的环境不一致,那么很可能会存在模型效果不错但实际测 试效果较差的情况。针对这种情况建议重新调整训练集,将训练数据与实际业务场景数据尽可能一致。

云服务调用数据管理

开通云服务调用数据管理功能后,可查找云服务模型识别错误的数据,纠正结果并将其加入模型迭代的训练集,实现训练数据的持续丰富和模型 效果的持续优化

具体使用流程如下:

1. 为已上线接口开通云服务调用数据服务

此接口未开通EasyDL云服务调用数据管理,是否开通? >> 功能介绍 現K模型 校验模型 技術項型 方提供本功能之気服务调用数据管理,是否开通? >> 功能介紹 方提供本功能之口服务调用数据管理,但否开通 方提供本功能之目的,您通过项口上传的数据会存储在百度EasyDL服务得到、数据将从开通本功能次日开始存储、最多存储30万的数据、超过30万的数 数据集管理 2. 百度仅提供数据存储空间服务,该数据仅用于向您提供本功能服务,不会用于其他目的。 3. 您使用本功能,应当保证在开通服务期间不得允许任何其他第三方使用以上提口,并保证对所上传的数据享有合法权益,有仅进行上述处理并承担全部法 数据标注上传 您用通本愿务限初间的同意百度以上述方式存储您上传的数据。		我的模型
按验模型 开通*EasyOL宏服务调用数据管理*(以下简称*本功能*),可意找宏服务模型识别错误的数据,纠正结果并将其加入模型迭代的训练集,实现训练数据的开 室间模型发展的持续优化。 发用型		训练模型
发布模型 â印模型效果的持续优化。 1、为提供本功能之目的,意通过接口上传的数据会存储在首度EasyOL服务器中,数据将从开通本功能次日开始存储。最多存储30天的数据,超过30天的数 初期除。 发展管理 2. 直度仅提供数据存储空间服务,该数据仅用于向虑提供本功能服务,不会用于其他目的。 发展管理 2. 直度仅提供数据存储空间服务,该数据仅用于向虑提供本功能服务,不会用于其他目的。 3. 您使用本功能,直当保证在开通服务期间不得允许任何其他第三方使用以上提口,并保证对所上传的数据享有合法权益,有权进行上述处理并承担全部活任。 发展标注上传 您开通本服务即视力感问应自度以上述方式存储您上传的数据。	8,纠正结果并将其加入模型迭代的训练集,实现训练数据的持续丰	校验模型
改題年心 动制除。 2. 百度仅提供数据存储空间服务,该数据仅用于向您提供本功履服务,不会用于其他目的。 3. 56使用本功能,应当保证在开通服务制间不得允许任何其他第三方使用以上提口,并保证对所上传的数据享有合法权益,有权进行上述处理并承担全部法任。 数据依注上作食 芯开通本服务加税为您同意百度以上述方式存储您上作的数据。	直本功能次日开始存储。最多存储30天的数据,超过30天的数据将自	发布模型
設備集管理 2. 百度仅提供数据存储空间服务,该数据仅用于向您提供本功施服务,不会用于其他目的。 3. 应使用本功能, 应当保证在开通服务期间不得允许任何其他第三方使用以上提口,并保证对所上传的数据享有合法权益,有权进行上述处理并承担全部法任。 按据榜注/上传 您开通本服务即视为您同意百度以上述方式存储您上传的数据。		改据中心
 3. 您使用本功能, 应当保证在开通服务期间不得允许任何其他第三方使用以上接口, 并保证对所上传的数据享有合法权益, 有权进行上述处理并承担全部法 创建数据集 任。 数据标注/上传 您开通本服务即视为您问题百度以上述方式存储您上传的数据。 		数据集管理
数据标注/上传 您开通本服务即视为您同意百度以上述方式存储您上传的数据。	所上传的数据享有合法权益,有权进行上述处理并承担全部法律责	创建数据集
		数据标注/上传
云服务调用数据		云服务调用数据

- 2. 通过选择调用时间、标签,并设置筛选条件,查看疑似错误识别的图片
- 注意:数据将从开通功能后开始存储,最多存储30天的数据。当天调用的数据暂不支持即时查看,可在第二天查看

模型中心	云服务调用数据管理	服务条款
我的模型	选择接口: horse V 对应模型: 物体检测 对应版本: V3	
创建模型	演用时间: 2019/7/31_2019/2/2 ♠ [5]	
训练模型	мпанана, сола/цон - сола/оса 🛲 🕅	
校验模型	选择标签: horse V	
发布模型	筛选条件: ● 图片中识别出该标签,该标签的置信度介于 0 % ~ 60 %	
数据中心	○ 图片中址约出的设体金数重	
数据集管理	<u>童</u> 看祥遗结果	
创建数据集		
数据标注/上传		
云服务调用数据		

3. 将接口识别错误的图片添加到指定数据集(建议新建数据集)并纠正结果。后续训练模型时,只需增加包含接口数据的数据集,即可提升模型效果

尝试不同的训练配置 可前往训练配置页面尝试不同的配置组合,因不同数据集在不同的算法上可能表现不一致,所以建议您多尝试不同的算法选型后综合挑选精度最高的模型使用,你可以选择如下的配置项:

- 增量训练
- 精度提升配置包
- 自动超参搜索
- 自定义验证集

- 数据增强策略
- 在高级训练配置中增加输入图片分辨率

注:如您需检测的目标在图中占比小于5%,建议您选择「精度提升配置包」中的小目标检测算法,但出于对算法性能的考虑,目前小目标检测算法仅支持本地部署

物体检测模型	÷	训练模型							操作文档	常见问题	新手教程	提交工单
第 总览												
曲 模型中心		远择模型	水果楦测									
我的模型		训练配置										
创建模型		部署方式	公有云部署 Eas	/Edge本地部署	如何选择部署方式?							
训练模型		训练方式	常规训练 精度提	限时免费 升配置包								
校验模型		洗择算法	0 招來請度 ②	高請度の一方		首 刊)						
发布模型						~.m						
🖉 EasyData数据服务		高级训练配置										
数据总览		添加数据										
标签组管理		添加数据集	+ 请选择									
在线标注		白宝义验证集团										
云服务数据回流		BRANK O										
摄像头数据采集		日走又別此来①										
◎ 公有云服务		数据增强策略 ⑦	○ 默认配置 ○	手动配置								
在线服务			积弱所远网络,默认配)	重必要的数据增强策略。								
批量预测		训练环境	名称	规格			算力	速度比例	价格			
器 EasyEdge本地部署			O GPU P4	TeslaGPU_P4_8G	显存单卡_12核CPU_40G内存		5.5 TeraFLOPS	1	免费			
纯高线服务			GPU P40	TeslaGPU_P40_24	G显存单卡_12核CPU_40G内	存	12 TeraFLOPS	1.27	单卡¥0.36/分钟(50小时*节点炙	2费)	
端云协同服务	^		0.000.000	TeeleCBL 1000 16		h Za	14 TerreELODP	1.67	6-EV0 4E/(336) /	10.10.001+32.0	5 (A (B))	
我的部署包			U GPU V100	TesiaGP0_V100_10	20単行車 ト_12核CPU_00GP	317	14 TerancOPS	1.07	平下40.40/万钟((2.10/049/17))	x()631/	
我的本地设备		开始训练										

模型发布

⑦物体检测模型发布整体说明

训练完成后,可将模型部署在公有云服务器、通用小型设备、本地服务器,或直接购买软硬一体方案,灵活适配各种使用场景及运行环境

公有云在线服务

训练完成的模型存储在云端,可通过独立Rest API调用模型,实现AI能力与业务系统或硬件设备整合

具有完善的鉴权、流控等安全机制, GPU集群稳定承载高并发请求

支持查找云端模型识别错误的数据,纠正结果并将其加入模型迭代的训练集,不断优化模型效果

纯离线服务 训练完成的模型整体打包为纯离线服务,可下载在本地稳定调用。纯离线服务按部署硬件芯片不同分为本地服务器部署、通用小型设备部署。为了提供更好的算法与硬件推理效果,EasyDL提供软硬一体方案部署。纯离线服务的整体支持与评测信息可详见算法与性能评测大表

本地服务器部署

可将训练完成的模型部署在私有CPU/GPU服务器上,支持服务器API和服务器SDK两种集成方式

模型服务性能表现更好,适用于对性能要求较高的场景,例如工业质检、流水线产品分拣等

通用小型设备

训练完成的模型被打包成适配智能硬件的SDK,可进行设备端离线计算。满足推理阶段数据敏感性要求、更快的响应速度要求

支持iOS、Android、Linux、Windows四种操作系统,基础接口封装完善,满足灵活的应用侧二次开发

软硬一体方案

高性能硬件与模型深度适配,多种方案可选。可应用于工业分拣、视频监控等多种设备端离线计算场景,让离线AI落地更轻松。了解更多

端云协同服务

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

「云管理,端计算」的端云协同服务,具体包括:

在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新

断网状态下模型离线计算(http服务,可调用与公有云API功能相同的接口)

联网状态下在平台管理设备运行状态、资源利用率

- 心 公有云部署
- ^の如何发布物体检测API

在新手教程中点击链接过来的用户请注意,您仍需要完整训练模型后,按如下操作指引,方可使用公有云服务

训练完毕后可以在左侧导航栏中找到【发布模型】,依次进行以下操作即可发布公有云API:

- 选择模型
- 选择部署方式「公有云部署」
- 选择版本
- 自定义服务名称、接口地址后缀
- 申请发布

申请发布后,通常的审核周期为T+1,即当天申请第二天可以审核完成。如果需要加急、或者遇到莫名被拒的情况,请在百度智能云控制台内提 交工单反馈

发布模型界面示意:

模型中心	发布模型							
我的模型	选择模型:	彬彬models	~		标准接口规范参	考		
创建模型	部署方式:	公有云部署	\sim		标准接口请求参	*考记明:		
训练模型	选择版本:	V1	\sim		字段名称	必须	类型	说明
校验模型	•服务名称:				image	是	string	图像数据,base64编码,要求 base64编码后大小不超过4M,
发布模型	* 接口地址:	https://aip.baidubce.com/rpd	c/2.0/ai_	custom/v1/detec				最短辺至少15px,最长辺最大 4096px,支持jpg/png/bmp格式
数据中心	耳他要求:	tion/			threshold	否	number	阈值,默认为当前模型推荐阈值 (0-1之间),具体值可以在我的 模型列表-模型效果查看
蚁 插 朱 盲 王			标准接口脑应应	2 60 32 PB -				
创建数据集					10/1212 11 11:12 7	-FX 0043 -		
数据标注/上传				0/500	字段名称	必须	类型	说明
云服务调用数据		✓ 同意云服务调用数据管理服	服务条款;	并开通服务	log_id	是	number	唯一的log id,用于问题定位
		提交申请			error_code	否	number	错误码,当请求错误时返回
					error_msg	否	string	错误描述信息,当请求错误时返 回
					results	否	array(object)	标签结果数组
					+ name	否	string	标签名称
					+ score	否	number	置信度
					+ location	否	object	

接口赋权

在正式使用之前,还需要做的一项工作为接口赋权,需要登录EasyDL版控制台中创建一个应用,获得由一串数字组成的appid,然后就可以参考 接口文档正式使用了

Ô	全局 ~						Q	合作伙伴	Тŧ	я £	帮助文档	企业资源	财务	
88	< 返回EasyDL总览	产品服务 / EasyOL图像 - 5	2用列表											
>	EasyDL图像	应用列表												
98	公有云部著 へ	+ 91853/8												
ß	 应用列表 	应用名称		AppID	API Key	Secret Key	创建的	时间		摘作				
	 权限管理 	1 记刻描和用		15313490	gRHegw68kpnym0H6K7GI6wUd	<u>B</u> R	2018	-12-31 17:24	27	11.8 W	2 873			
	 用重取订 請求文約 	2 TRY		14676113	75a5pbP7mhpBWNION2NYoYcX	显示	2018	-11-05 11:35	54	12.87 W	2 8/3			
8-6 8-6	 售卖服务 	3 皮皮虾		14403281	i6HhGApGanhWuYex4Oel2uG8	显示	2018	-10-11 19:55	26	HR W	2 8/78			
Þ	EasyEdge本地部署 ~												< 1	>

同时支持在「公有云服务管理」-「权限管理」中为第三方用户配置权限

示意图如下:

V	全局 ~	
88	< 返回EasyDL总览 [₹]	"品服务 / EasyOL图像 - 权限管理
>	EasyDL图像	权限管理
88	公有云部署 ^	服务权限管理
æ	。 应用列表	支持为您或者他人的AppID配置可用服务权限。
F	 权限管理 	序号 服务名称 相关操作
þ	。 用量统计	1 1000 配置权限
Ħ	。 技术文档	2 test 配置权限
۳đ	• 售卖服务	3 cherry 配置权限
⊳	EasyEdge本地部署 ~	4 horse的服务名称 配置权限
		5 sea 配置权限
		6 手势识别 配置 权限
Ŷ	全局 ~	
	< 返回EasyDL总览	产品服务 / EasyDL图像 - 权限管理 / 配置权限
>	EasyDL图像	配置权限
88	公有云部署	12.94 dt 12.
Ē	• 应用列表	版另合称· 1000
Ē	 权限管理 	可用APPID: 14403281 皮皮虾 ×
 雨	 用量统计 	14676113 TRY ×
	10.40.000	15313490 记别版和狗 ×
Ħ	 技术又档 	请添加 十
e t	• 售卖服务	alation 1
⊳	EasyEdge本地部署	保存 取消

^の物体检测API调用文档

本文档主要说明定制化模型发布后获得的API如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

接口描述

基于自定义训练出的物体检测模型,实现定制图像识别。

接口鉴权

1、在EasyDL控制台创建应用

♥ 全局 ∨							Q	₽	?	品	۲	
EasyDL定制训练平台	产品服务 / EasyD	L定制训练平台 - 应	用列表 / 创建应用									
概览	创建新应用	I										
应用列表	• 应田夕称•											
监控报表	YTT 143 474 00 -	test										
技术文档	•应用类型:	游戏娱乐				~						
云服务权限管理	*接口选择:	您可以为应用勾	选多个接口权限,使此应	用可以请求已勾选的接	口服务							
离线SDK管理		EasyDL	 ✓ 百美椅子训练 ✓ 识别眼睛和嘴巴 	✓ 椅子顔色识别 ✓ 声音分类	✓ aa ✓ 商品检测							
		+ 百度语音										
		🛨 文字识别										
		➡ 人脸识别										
		🛨 自然语言处理	8									
		➡ 内容审核]									
		🛨 UNIT 📘										
		+ 知识图谱										
		➡ 图像识别										
		+ 智能呼叫中/	Ь									
		➡ 图像搜索										
		➡ 人体分析										
	•应用描述:	简单描述一下: 点,请控制在	您使用人工智能服务的应用 500字以内	用场景 , 如开发一款美	颜相机,需要检测人脸	送关键						

ī.

♥ 全局 ~					Q	≣	¢	?	8	۲	Q ~
EasyDL定制训练平台	产品服务 / EasyDL定制训练平台 - 应用列表 / 应用详情										
概览	应用详情										
应用列表	· 编辑 · 查看文档										
监控报表	应用名称	AppID		API Key			Secr	et Key			
技术文档	test2	14293933		bZsfqO2fUdHTDflxqF11r	nkTq			*** 显示			
云服务权限管理	[[]] 中制化训练现象。										
离线SDK管理	EasyDL 定制1U训练服务:										

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先进行自定义模型训练,完成训练后申请上线,上线成功后可在服务列表中查看并获取url。

URL参数:

参数	值
input_type	当取值为 url 时,需在请求参数中传入图片的URL string
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,与其他图像识别服务不同的是定制化图像识别服务以json方式请求。

Body请求示例:

```
{
"image": "<base64数据>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大4096px,支持 jpg/png/bmp格式 注意请去掉头部
thresho Id	否	number	-	默认值为推荐阈值,请在我的模型列表·模型效果查看推荐阈值
url	否	string	-	如果请求URL参数中增加"input_type=url",则该参数必传,否则"image"参数必传。参数内容为URL string, 用户需确保该string是有效的图片URL,否则会下载失败

请求代码示例

提示一:使用示例代码前,请记得替换其中的示例Token、图片地址或Base64信息。

提示二:部分语言依赖的类或库,请在代码注释中查看下载地址。

РНР	
Java	
Python3	
C++	

```
<?php
/**
* 发起http post请求(REST API),并获取REST请求的结果
* @param string $url
* @param string $param
* @return - http response body if succeeds, else false.
*/
function request_post($url = '', $param = '')
{
  if (empty($url) || empty($param)) {
    return false;
  }
 $postUrl = $url;
  $curlPost = $param;
 // 初始化curl
  $curl = curl_init();
  curl_setopt($curl, CURLOPT_URL, $postUrl);
            AND OUDLODT LIEADE
```

返回说明

物体检测-矩形框标注

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	识别结果数组
+name	否	string	分类名称
+score	否	number	置信度
+location	否		
++left	否	number	检测到的目标主体区域到图片左边界的距离
++top	否	number	检测到的目标主体区域到图片上边界的距离
++width	否	number	检测到的目标主体区域的宽度
++height	否	number	检测到的目标主体区域的高度

物体检测-自定义四边形标注

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	物体检测目标信息
+name	否	string	目标物体标签
+score	否	number	置信度
+location	否	object	目标物体所在位置
++points	否	list(object)	目标物体所在四边形的顶点信息
+++x	否	number	顶点横坐标
+++y	否	number	顶点纵坐标
error_code	否	number	错误码,当请求错误时返回
error_msg	否	string	错误描述信息,当请求错误时返回

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有 疑问请在百度云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈
336003	Base64解码失败	图片/音频/文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头 部。有疑问请在百度云控制台内提交工单反馈
336004	输入文件大小不合法	图片超出大小限制,图片限4M以内,请根据接口文档检查入参格式,有疑问请在百度云控制台内提交工单反 馈
336005	图片解码失败	图片编码错误(非jpg,bmp,png等常见图片格式),请检查并修改图片格式
336006	缺失必要参数	image字段缺失(未上传图片)
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

の 本地服务器部署

∞ 如何在本地服务器部署

训练完毕后,可以选择将模型通过「纯离线服务」或「端云协同服务」部署,具体介绍如下:

纯离线服务部署

可以在左侧导航栏中找到「纯离线服务」,依次进行以下操作即可将模型部署到本地服务器:

- 选择部署方式「服务器」
- 选择集成方式
- 选择模型、版本、系统和芯片

• 点击下一步

模型中心	纯离线服务 > 发布新服务
我的模型	
创建模型	3 选择部署形式 2 填写个人信息
训练模型	部署方式 💿 服务器 🔵 通用小型设备 🔵 专项适配硬件
校验模型	
发布模型	
EasyData数据服务	选择模型 fffffff V
数据总览	這择版本 V4 🗸
在线标注	选择系统和芯片 🔄 🛆 Linux
云服务数据回流	適用X86 CPU ◎ 英伟达GPU ◎
EasyEdge本地部署	华为 Atlas 300 百度 昆仑XPU
纯离线服务	+ 📲 Windows
AI市场	欄型加速: 同时获取加速版 ③
我的已购模型	
售卖模型	下一步

• 填写部分信息(注:个人信息的填写仅供EasyDL团队了解您,便于后续合作接洽,不会作为其他用途使用)

• 点击发布

训练模型	* 联系人姓名		
校验模型	* 由迁	12244444247	0
发布模型	*C 10	132 247	0
EasyData数据服务	* 邮箱	6*********@qq.com	
数据总览	* 职务		
在线标注	* 公司名称		
云服务数据回流	* 公司所在行业	请选择行业 🗸	
EasyEdge本地部署	* 业务量级	请你简要描述,每天要处理的数据量级和所需的并发数	
纯离线服务			
AI市场		c	J/500 J
我的已购模型			
售卖模型	* 功能描述	请简要描述应用AI能力的使用场景,期望达到的效果等	
		C	0/500
	上一步发	布	

① 私有API

将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷

点击「发布」后,前往「纯离线服务」等待部署包制作完成后,下载部署包,并参考文档完成集成

② 服务器端SDK

将模型封装成适配本地服务器(支持Linux和Windows)的SDK,可集成在其他程序中运行。首次联网激活后即可纯离线运行,占用服务器资源更少,使用方法更灵活

1、点击「发布」后,前往控制台申请服务器端SDK的试用序列号

2、点击「新增测试序列号」,根据模型类型选择「序列号类型」,填写「新增设备数」(所得序列号数量),点击确定即可

服务器纯离线服务管理					
API SDK					
使用说明: 1. 通过集成适配本地服务器的SDK高结使用模型: SDK支持二次开发,按CPU/GPU设备授权使用,操作步骤如下: の 在EasyDL申请并下载SDK → © 在下方申请激活序列号 → © 根据开发文档集成SDK,并联网激活 → © 馬结使用 2.序列号联网激活SDK后有效期为1个月,一个测试序列号激活模型上限数为10,如需正式购买,请点击下方购买正式模权。	收起				
+ 购买正式授权 新增测试序列号	开发文档 工单支持				

3、离线SDK的激活和使用,请参考文档完成集成

ai 公 大脑 AI 开放平台 开放	放能力	开发平台	行业应用	生态合作	AI市场	开发与教学	Q
EasyDL定制AI训练平台		芯片类型	实测硬件	高精度模型	高性能模型		
援索本产品文档内容 Q		x86-64	Xeon E5-2650 v4	61ms	40ms		
		Nvidia-GPU	Tesla P4	16ms	5ms		
> 公有云部署		Atlas 300	Atlas 300	12ms	9ms		
~ 私有服务器部署	- E	激活&使用	月步骤				
如何在私有服务器部署	-	室绿SDK的搬迁与使田分以上二半·					
私有API集成文档		mtsJohunakh-JuthJuki-JuthJuki-					
→ 服务器端SDK							
服务器端SDK简介 LinuxSDK集成文档- C++ LinuxSDK集成文档- Python Linux(Atlas)SDK集 成文档 WindowsSDK集成文 档		◎本地运行SDK,并完成首次联网激活					
		通过左侧导航栏查看不同操作系统SDK的开发文档					
) -				
		SDK常见	可尟				

端云协同服务部署

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

具体使用说明请参考端云协同服务说明

本地服务器部署价格说明

EasyDL已支持将定制模型部署在本地服务器上,只需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。

如需购买永久使用授权,服务器SDK用户请在控制台点击「购买正式授权」,并按照对应步骤激活。

服务器API用户请微信搜索"BaiduEasyDL"添加小助手咨询,通过线下签订合同购买使用。

更多参考

EasyDL官网入口 EasyDL开发文档

心 纯离线SDK说明

^の 纯离线SDK简介

本文档主要说明定制化模型发布后获得的服务器端SDK如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

SDK说明

物体检测服务器端SDK支持Linux、Windows两种操作系统。以下为具体的系统、硬件环境支持:

操作系统	系统支持	硬件环境要求
Linux		CPU: x86_64 NVIDIA GPU: x86_64 HUAWEI Atlas 300: x86_64
Windows	64位 Windows7 及以 上	NVIDIA GPU: x86_64 环境依赖: .NET Framework 4.5 Visual C++ Redistributable Packages for Visual Studio 2013 Visual C++ Redistributable Packages for Visual Studio 2015 GPU依赖: CUDA 9.x + cuDNN 7.x

单次预测耗时参考

根据具体设备、线程数不同,数据可能有波动,请以实测为准

在算法性能及适配硬件查看评测信息表

激活&使用步骤

离线SDK的激活与使用分以下三步:

- ① 下载SDK后,在控制台获取序列号
- ② 本地运行SDK,并完成首次联网激活

通过左侧导航栏查看不同操作系统SDK的开发文档

③ 正式使用

SDK常见问题

通过左侧导航栏查看不同操作系统SDK的FAQ

以下是通用FAQ,如您的问题仍未解决,请在百度云控制台内提交工单反馈

1、激活失败怎么办?

①可能是当前序列号已被其他设备激活使用,请核实序列号后用未被激活的序列号重新激活

②序列号填写错误,请核实序列号后重新激活

③同一台设备绑定同一个序列号激活次数过多(超过50次),请更换序列号后重试

④首次激活需要联网,网络环境不佳或无网络环境,请检查网络环境后重试

⑤模型发布者和序列号所属账号非同一账号,如果存在这种异常建议更换账号获取有效序列号

⑥序列号已过有效期,请更换序列号后重试

⑦如有其他异常请在百度云控制台内提交工单反馈

の Windows集成文档

简介

本文档介绍物体检测服务器端Windows SDK的使用方法。

- 硬件支持:
 - NVIDIA GPU (普通版,加速版)
- 操作系统支持
 - 64位 Windows 7 及以上
 - 64位 Windows Server 2012及以上
- 环境依赖 (必须安装以下版本)
 - .NET Framework 4.5
 - Visual C++ Redistributable Packages for Visual Studio 2013
 - Visual C++ Redistributable Packages for Visual Studio 2015
- GPU基础版 (EasyEdge-win-x86-nvidia-gpu) 依赖 (必须安装以下版本)
 - CUDA 11.0.x + cuDNN 8.4.x(注意参照英伟达安装文档安装Zlib:http://www.winimage.com/zLibDll/zlib123dllx64.zip,解压后将 dll_x64/zlibwapi.dll 拷贝到cuda的bin目录下) + 硬件计算能力(https://developer.nvidia.com/cuda-gpus#compute)达6.1及以上
 - CUDA 12.0.x + cuDNN 8.9.x(注意参照英伟达安装文档安装Zlib) + 硬件计算能力达7.5及以上
- GPU加速版 (EasyEdge-win-x86-nvidia-gpu-tensorrt) 依赖 (必须安装以下版本)
 - CUDA 11.0.x + cuDNN 8.4.x(注意参照英伟达安装文档安装Zlib) + TensorRT 8.4.x.x
 - CUDA 12.0.x + cuDNN 8.9.x(注意参照英伟达安装文档安装Zlib) + TensorRT 8.6.x.x
- GPU加速版 (EasyEdge-win-x86-nvidia-gpu-paddletrt) 依赖 (必须安装以下版本)
 - CUDA 11.0.x + cuDNN 8.4.x(注意参照英伟达安装文档安装Zlib) + TensorRT 8.4.3.1 + 硬件计算能力达6.1及以上
 - CUDA 12.0.x + cuDNN 8.9.x(注意参照英伟达安装文档安装Zlib) + TensorRT 8.6.1.6 + 硬件计算能力达7.5及以上
- GPU加速版 (x86-nvidia-gpu-torch)
 - CUDA 11.0.x + cuDNN 8.0.5.x
- 协议
 - HTTP
- 更详细的环境说明可参考SDK内的README.md

Release Notes | 时间 | 版本 | 说明 | | -------- | ------ | -2023-08-30 | 1.8.3 | 新增支持按实例数鉴权 | | 2023-06-29 | 1.8.2 | 优化模型 算法 || 2023-05-17 | 1.8.1 | 预测引擎升级,修复部分网络内存泄露问题 || 2023-03-16 | 1.8.0 | 预测引擎升级 || 2022-12-29 | 1.7.2 | 预测引 擎升级 | | 2022-10-27 | 1.7.1 | GPU底层引擎升级,下线基础版CUDA10.0及以下版本支持 | | 2022-09-15 | 1.7.0 | 优化模型算法;GPU CUDA9.0 CUDA10.0 标记为待废弃状态 || 2022-07-28 | 1.6.0 | 优化模型算法 || 2022-05-27 | 1.5.1 | 新增支持BML Cloud小目标检测模型 || 2022-05-18 | 1.5.0 | 修复各别机器下程序崩溃的问题 | | 2022-04-25 | 1.4.1 | EasyDL, BML升级支持paddle2模型 | | 2022-03-25 | 1.4.0 | 优化 模型算法 || 2021-12-22 | 1.3.5 | GPU基础版推理引擎优化升级; GPU加速版支持自定义模型文件缓存路径; demo程序优化环境依赖检测 || 2021-10-20 | 1.3.4 | 修复已知问题| | 2021-08-19 | 1.3.2 | 新增支持EasyDL小目标检测,新增DEMO二进制文件 || 2021-06-29 | 1.3.1 | 预测引 擎升级 || 2021-05-13 | 1.3.0 | 模型发布新增多种加速方案选择;目标追踪支持x86平台的GPU及加速版;展示已发布模型性能评估报告 || 2021-04-08 | 1.2.3 | 支持BML平台模型仓库本地上传模型 | | 2021-03-09 | 1.2.2 | 修复已知问题 | | 2021-01-27 | 1.2.1 | 新增模型支持;性能优 化;问题修复||2020-12-18|1.2.0|推理引擎升级||2020-11-26|1.1.20|新增一些模型的加速版支持||2020-10-29|1.1.19|修复已知问 题 || 2020-09-17 | 1.1.18| 支持更多模型 || 2020.08.11 | 1.1.17 | 支持专业版更多模型 || 2020.06.23 | 1.1.16 | 支持专业版更多模型 || 2020.05.15 | 1.1.15 | 更新加速版tensorrt版本,支持高精度检测 | | 2020.03.13 | 1.1.14 | 支持声音分类 | | 2020.02.23 | 1.1.13 | 支持多阶段模 型 | | 2020.01.16 | 1.1.12 | 预测默认使用推荐阈值| | 2019.12.26 | 1.1.11 | 支持物体检测高精度算法的CPU加速版,EasyDL 专业版支持 SDK 加速版 || 2019.12.04 | 1.1.10 | 支持图像分割 || 2019.10.21 | 1.1.9 | 支持 EasyDL 专业版 || 2019.08.29 | 1.1.8 | CPU 加速版支持 || 2019.07.19 | 1.1.7 | 提供模型更新工具 || 2019.05.16 | 1.1.3 | NVIDIA GPU 支持 || 2019.03.15 | 1.1.0 | 架构与功能完善 || 2019.02.28 | 1.0.6 | 引擎功能完善 || 2019.02.13 | 1.0.5 | paddlepaddle 支持 || 2018.11.30 | 1.0.0 | 第一版 ! |

快速开始

1. 安装依赖

安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642

Visual C++ Redistributable Packages for Visual Studio 2013

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

Visual C++ Redistributable Packages for Visual Studio 2015

https://www.microsoft.com/zh-cn/download/details.aspx?id=48145

如果使用GPU版SDK,请安装CUDA + cuDNN

https://developer.nvidia.com/cuda https://developer.nvidia.com/cudnn

如果使用GPU版加速版SDK,请安装TensorRT

https://developer.nvidia.com/tensorrt

根据cuda版本下载,下载后把lib目录下的所有dll,拷贝到SDK的dll目录下

注意事项

- 1. 安装目录不能包含中文
- 2. Windows Server 请自行开启,选择"我的电脑"——"属性"——"管理"——"添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启即可。

2. 运行离线SDK

解压下载好的SDK,SDK默认使用cuda9版本,如果需要cuda10请运行EasyEdge CUDA10.0.bat切换到cuda10版本,之后打开EasyEdge.exe,选 择鉴权模式,输入Serial Num,点击"启动服务",等待数秒即可启动成功,本地服务默认运行在

http://127.0.0.1:24401/

其他任何语言只需通过HTTP调用即可。

如启动失败,可参考如下步骤排查:



2.1 离线鉴权 (默认鉴权模式) 首次联网激活,后续离线使用

Al ver. 1.2.0.0 core. 1.8.1 sec. 1.5.1 - X					
Model	a an an an airte a				
Serial Num]		
Host	127.0.0.1]		
Port	24401]		
鉴权模式	● 离线鉴权 ○ 按实例数鉴权				
服务状态	服务未启动				
	启动服务	是否开机自	动启动		

2.2 按实例数鉴权 周期性联网激活,离线后会释放所占用鉴权,启动时请确保心跳间隔小于等于生成序列号时填写的定期确认时间

Al ver. 1.2.0.0 core. 1	l.8.1 sec. 1.5.1 — □ ×
Model	[4] A.P. Weiner, M. Weiner, Phys. Rev. Lett. 76, 1000 (1996).
Serial Num	
Host	127.0.0.1
Port	24401
鉴权模式	○ 离线鉴权 ● 按实例数鉴权 心跳间隔 20 秒>
服务状态	服务未启动
	启动服务 是否开机自动启动 🗌

基于源码集成时,若需要按实例数鉴权,需要通过代码指定使用按实例数鉴权

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2);	
global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL	, 20);

或通过环境变量指定

set EDGE_CONTROLLER_KEY_AUTH_MODE=2 set EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=20

の 2.3 序列号激活错误码

错误码	文案	描述
4001	parameters missing. 参数缺失	
4002	parameters invalid. 参数不合法	
4003	model invalid. 模型信息不合法	
4004	no more activation left. 该序列号和该设备的激活次数超上限	
4005	the serial key is out of date. 该序列号过期	
4006	the serial key has been activated. 该序列号已被其他设备激活	该序列号已被其他设备激活,不能重复激活。
4007	account invalid. 序列号不能用于其他账号的模型	序列号不能用于其他账号的模型,只能用于绑定账号的模型。
4008	serial key invalid. 序列号不合法	序列号不存在或找不到
4009	bundle id invalid. 包名不合法	
4010	product invalid. 产品不合法	如easydl的SDK使用BML的序列号来激活,会报该错误
4011	platform invalid. 平台不合法	
4012	activate too frequent. 激活太频繁	激活太频繁,请稍后再进行激活。
4013	device type and license type not match. 硬件类型和序列号类型不匹配	如使用加速版序列号激活基础版SDK会报该错误
4014	exceed max activate device num. 超过最大授权数量	
4015	technology invalid. 技术类型不合法	
4016	exceed max activate entity num. 超过最大模型数量	
4017	device invalid. 设备不合法	
4018	model invalid. 模型不合法	

3. Demo示例(以图像服务为例)

服务运行成功,此时可直接在浏览器中输入http://127.0.0.1:24401,在h5中测试模型效果。



使用说明

调用说明

Python 使用示例代码如下

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
```

params 为GET参数 data 为POST Body

 $result = requests.post('http://127.0.0.1:24401/', params=\{'threshold': 0.1\},$

data=img).json()

C# 使用示例代码如下

FileStream fs = new FileStream("./img.jpg", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close(); WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close();

response.Close();

C++ 使用示例代码如下,需要安装curl

```
**include <sys/stat.h>**
**include <curl/curl.h>**
**define S_ISREG(m) (((m) & 0170000) == (0100000))**
**define S_{ISDIR}(m) (((m) \& 0170000) == (0040000))**
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./img.jpg";
  FILE *fp = NULL;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  }
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\"\n", post_data_filename);
     return -1;
  }
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
     curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
     curl_easy_setopt(curl, CURLOPT_POST, 1L);
     curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
     curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
     res = curl_easy_perform(curl);
     if (res != CURLE_OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     }
     curl_easy_cleanup(curl);
  curl_global_cleanup();
  fclose(fp);
  return 0;
```

请求参数

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送图片二进制。

返回参数|字段|类型|取值|说明||--------|-----|-----|-----|| confidence | float | 0~1 | 检测的置信度 || label | string || 检测的类 别 || index | number || 检测的类别 || x1, y1 | float | 0~1 | 矩形的左上角坐标 (相对长宽的比例值) || x2, y2 | float | 0~1 | 矩形的右下角坐标 (相 对长宽的比例值) |

关于矩形坐标

- x1 * 图片宽度 = 检测框的左上角的横坐标
- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2*图片宽度 = 检测框的右下角的横坐标
- y2*图片高度 = 检测框的右下角的纵坐标

集成指南

基于HTTP集成

通过EasyEdge.exe启动服务后,参照上面的调用说明,通过HTTP请求集成到自己的服务中

基于c++ dll集成

集成前提

解压开的SDK包中包含src、lib、dll、include四个目录才支持基于c++ dll集成

集成方法

参考src目录中的CMakeLists.txt进行集成

基于c# dll集成

集成前提

解压开的SDK包中包含src\demo_serving_csharp、dll两个目录才支持基于c# dll集成

集成方法

参考src\demo_serving_csharp目录中的CMakeLists.txt进行集成

FAQ

1. 服务启动失败,怎么处理?

根据SDK内的README.md检查依赖是否都已正确安装

请确保相关依赖都安装正确,版本必须如下: *.NET Framework 4.5* Visual C++ Redistributable Packages for Visual Studio 2013 * Visual C++ Redistributable Packages for Visual Studio 2015

GPU依赖,版本必须如下: * CUDA 11.0.x + cuDNN 8.4.x 或者 CUDA 11.7.x + cuDNN 8.4.x

GPU加速版(EasyEdge-win-x86-nvidia-gpu-tensorrt)依赖,版本必须如下: * CUDA 11.0.x + cuDNN 8.4.x + TensorRT 8.4.x.x

GPU加速版(EasyEdge-win-x86-nvidia-gpu-paddletrt)依赖,版本必须如下: * CUDA 11.0.x + cuDNN 8.4.x + TensorRT 8.4.3.1

2. 服务调用时返回为空,怎么处理?调用输入的图片必须是RGB格式,请确认是否有alpha通道。

3. 多个模型怎么同时使用? SDK设置运行不同的端口,点击运行即可。

4. JAVA、C#等其他语言怎么调用SDK?

参考 https://ai.baidu.com/forum/topic/show/943765

5. 启动失败,缺失DLL? 打开EasyEdge.log,查看日志错误,根据提示处理 缺失DLL,请使用 https://www.dependencywalker.com/ 查看相应模 块依赖DLL缺失哪些,请自行下载安装

6. 启动失败,报错NotDecrypted? Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

7. 启动失败,报错 SerialNum无效

日志显示failed to get/check device id(xxx)或者 Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 C:\Users\\${用户名}\.baidu\easyedge 目录,再重新激活。

其他问题 如果无法解决,可到论坛发帖: https://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

简介

本文档介绍EasyEdge/EasyDL的Linux CPP SDK的使用方法。

- 网络类型支持:图像分类,物体检测,图像分割,目标追踪
- 硬件支持:
 - CPU 基础版: intel x86_64 * AMD x86_64 龙芯 loongarch64 飞腾 aarch64
 - CPU 加速版 Intel Xeon with Intel®AVX2 and AVX512 Intel Core Processors with AVX2 Intel Atom Processors with SSE AMD Core Processors with AVX2
 - NVIDIA GPU: x86_64 PC
 - 寒武纪 Cambricon MLU270
 - 比特大陆计算卡SC5+
 - 百度昆仑XPU K200
 - x86_64 飞腾 aarch64 百度昆仑XPU R200
 - x86_64 飞腾 aarch64
 - 华为Atlas 300
 - 海光DCU: x86_64 PC
 - 寒武纪 MLU370 on x86_64
- 操作系统支持:Linux

根据开发者的选择,实际下载的版本可能是以下版本之一:

- EasyDL图像
 - x86 CPU 基础版
 - x86 CPU 加速版
 - Nvidia GPU 基础版
 - Nvidia GPU 加速版
 - x86 mlu270基础版
 - x86 SC5+基础版
 - Phytium MLU270基础版
 - Phytium XPU基础版
 - Phytium Atlas300I基础版
 - Hygon DCU基础版

性能数据参考算法性能及适配硬件

*intel 官方合作,拥有更好的适配与性能表现。

Release Notes

时间	版本	说明
2023.0 8.31	1.8.3	Atlas系列Soc支持语义分割模型,Atlas Cann升级到6.0.1,昆仑XPU后端推理引擎升级
2023.0 6.29	1.8.2	模型压缩能力升级
2023.0	101	サ本業を登してきる。1911年19月1日日本である。2011年2011

5.17		
2023.0 3.16	1.8.0	支持图像分类精度提升包本地部署
2022.1 2.29	1.7.2	模型性能优化;推理库性能优化
2022.1 0.27	1.7.1	新增语义分割模型http请求示例;升级海光DCU SDK,需配套rocm4.3版本使用;Linux GPU基础版下线适用于CUDA10.0及以 下版本的SDK;Linux GPU加速版升级推理引擎版本
2022.0 9.15	1.7.0	Linux GPU加速版升级预测引擎;Linux GPU加速版适用于CUDA9.0、CUDA10.0的SDK为deprecated,未来移除;新增实例分 割高性能模型离线部署;性能优化
2022.0 7.28	1.6.0	Linux CPU普通版、Linux GPU普通/加速版、Jetson新增目标追踪模型接入实时流的demo
2022.0 5.27	1.5.1	CPU、GPU普通版新增支持BML Cloud小目标检测模型
2022.0 5.18	1.5.0	GPU加速版max_batch_size参数含义变更;修复GPU加速版并发预测时部分图片结果预测错误及耗时增加问题;CPU普通版预 测引擎升级;新增版本号头文件;新增飞腾Atlas300l支持,并且在EasdDL新增多种加速版本;示例代码移除frame_buffer, 新增更安全高效的safe_queue; 新增Tensor In/Out接口和Demo
2022.0 4.25	1.4.1	EasyDL, BML升级支持paddle2模型
2022.0 3.25	1.4.0	新增支持海光服务器搭配海光DCU加速卡;
2021.1 2.22	1.3.5	GPU加速版支持自定义模型文件缓存路径;新增支持飞腾MLU270服务器、飞腾XPU服务器
2021.1 0.20	1.3.4	CPU加速版推理引擎优化升级,新增支持飞腾CPU、龙芯CPU服务器、比特大陆计算卡SC5+ BM1684、寒武纪MLU270;大幅 提升EasyDL GPU加速版有损压缩加速模型的推理速度
2021.0 8.19	1.3.2	CPU、GPU普通版及无损加速版新增支持EasyDL小目标检测,CPU普通版、GPU普通版支持检测模型的batch预测
2021.0 6.29	1.3.1	CPU普通版、GPU普通版支持分类模型的batch预测,CPU加速版支持分类、检测模型的batch预测;GPU加速版支持 CUDA11.1;视频流解析支持调整分辨率;预测引擎升级
2021.0 5.13	1.3.0	新增视频流接入支持;模型发布新增多种加速方案选择;目标追踪支持x86平台的CPU、GPU加速版;展示已发布模型性能评 估报告
2021.0 3.09	1.2.1	GPU新增目标追踪支持, http server服务支持图片通过base64格式调用,EasyDL高性能检测模型和均衡检测模型CPU加速版新 增量化压缩模型
2021.0 1.27	1.1.0	EasyDL经典版分类高性能模型升级;部分SDK不再需要单独安装OpenCV
2020.1 2.18	1.0.0	1.0版本发布!安全加固升级、性能优化、引擎升级、接口优化等多项更新
2020.1 1.26	0.5.8	EasyDL经典版分类模型CPU加速版里新增量化压缩模型
2020.1 0.29	0.5.7	新增CPU加速版支持:EasyDL经典版高精度、超高精度物体检测模型和EasyDL经典版图像分割模型
2020.0 9.17	0.5.6	性能优化,支持更多模型
2020.0 8.11	0.5.5	提升预测速度;支持百度昆仑芯片
2020.0 5.15	0.5.3	优化性能,支持专业版更多模型
2020.0 4.16	0.5.2	支持CPU加速版;CPU基础版引擎升级;GPU加速版支持多卡多线程
2020.0 3.12	0.5.0	x86引擎升级;更新本地http服务接口;GPU加速版提速,支持批量图片推理

2020.0 1.16	0.4.7	ARM引擎升级;增加推荐阈值支持
2019.1 2.26	0.4.6	支持海思NNIE
2019.1 1.02	0.4.5	移除curl依赖;支持自动编译OpenCV;支持EasyDL 专业版 Yolov3; 支持EasyDL经典版高精度物体检测模型升级
2019.1 0.25	0.4.4	ARM引擎升级,性能提升30%; 支持EasyDL专业版模型
2019.0 9.23	0.4.3	增加海思NNIE加速芯片支持
2019.0 8.30	0.4.2	ARM引擎升级;支持分类高性能与高精度模型
2019.0 7.25	0.4.1	引擎升级,性能提升
2019.0 7.25	0.4.0	支持Xeye,细节完善
2019.0 6.11	0.3.3	paddle引擎升级;性能提升
2019.0 5.16	0.3.2	新增NVIDIA GPU支持;新增armv7l支持
2019.0 4.25	0.3.1	优化硬件支持
2019.0 3.29	0.3.0	ARM64 支持;效果提升
2019.0 2.20	0.2.1	paddle引擎支持;效果提升
2018.1 1.30	0.1.0	第一版!

2022-5-18:【接口变更】 PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE含义变更。 变更前:预测输入图片数不大于该值均可。 变更后:预测输入图片数需等于该值。SDK内部对该接口变更做了兼容处理,在输入图片数小于该值时依然可正常运行,但预测性能会和等于该值时一致。推荐根据实际输入图片数量需求修改该值,尽可能保持最小。

2020-12-18: 【接口升级】 参数配置接口从1.0.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。 【关于SDK包与RES模型文件夹配套使用的说明】 我们强烈建议用户使用部署tar包中配套的SDK和RES。 更新模型时,如果SDK版本号有更新,请务必同时更新SDK,旧版本的SDK可能无法 正确适配新发布出来部署包中的RES模型。

快速开始

SDK在以下环境中测试通过

- x86_64, Ubuntu 16.04, gcc 5.4
- x86_64, Ubuntu 18.04, gcc 7.4
- Tesla P4, Ubuntu 16.04, cuda 9.0, cudnn 7.5
- x86_64, Ubuntu 16.04, gcc 5.4, XTCL r1.0
- aarch64, Kylin V10, gcc 7.3
- loongarch64, Kylin V10, gcc 8.3

- Bitmain SC5+ BM1684, Ubuntu 18.04, gcc 5.4
- x86_64 MLU270 , Ubuntu 18.04, gcc 7.5
- phytium MLU270 , Kylin V10 , gcc 7.3.0
- phytium XPU , Kylin V10 , gcc 7.3.0
- hygon DCU, CentOS 7.8 gcc 7.3.0
- XPU K200, x86_64, Ubuntu 18.04
- XPU K200 aarch64, Ubuntu 18.04
- XPU R200, x86_64, Ubuntu 18.04
- XPU R200 aarch64, Ubuntu 18.04
- MLU370, x86_64, Centos7.6.1810

依赖包括

- cmake 3+
- gcc 5.4 (需包含 GLIBCXX_3.4.22) , gcc / glibc版本请以实际SDK ldd的结果为准
- opencv3.4.11 (可选)
- cuda && cudnn (使用NVIDIA-GPU时必须,SDK内提供多个Cuda版本推理套件,根据需要安装依赖的Cuda和Cudnn版本)
- XTCL 1.0.0.187 (使用昆仑服务器时必须)
- Rocm4.3, Miopen 2.14(使用海光DCU服务器时必须)

1. 安装依赖

以下步骤均可选,请开发者根据实际运行环境选择安装。

(可选) 安装cuda&cudnn

在NVIDIA GPU上运行必须(包括GPU基础版,GPU加速版)

对于GPU基础版,若开发者需求不同的依赖版本,请在PaddlePaddle官网下载对应版本的libpaddle_fluid.so或参考其文档进行编译,覆盖lib文件 夹下的相关库文件。

(可选) 安装TensorRT

在NVIDIA GPU上运行GPU加速版必须

下载包中提供了对应 cuda9.0、cuda10.0、cuda10.2、cuda11.0+四个版本的 SDK, cuda9.0 和 cuda10.0 的 SDK 默认依赖的 TensorRT 版本为 TensorRT7.0.0.11, cuda10.2 及以上的 SDK 默认依赖的 TensorRT 版本为 TensorRT8.4,请在这里下载对应 cuda 版本的 TensorRT,并把其中 的lib文件拷贝到系统lib目录,或其他目录并设置环境变量。

(可选) 安装XTCL 使用昆仑服务器及对应SDK时必须</mark>请安装与1.0.0.187版本兼容的XTCL。必要时,请将运行库路径添加到环境变量。

(可选) 安装Rocm、Miopen

使用海光DCU服务器对应SDK时必须

海光DCU SDK依赖Rocm 4.3和Miopen 2.14版本,推荐使用easyedge镜像 (registry.baidubce.com/easyedge/hygon_dcu_infer:1.0.2.rocm4.3), SDK镜像内运行,镜像拉取方式(wget https://aipe-easyedgepublic.bj.bcebos.com/dcu_docker_images/hygon_dcu_rocm4.3.tar.gz && docker load -i hygon_dcu_rocm4.3.tar.gz), 关于海光DCU使用更多细节 可参考paddle文档

2. 使用序列号激活 请在官网获取序列号

纯高线服务说明								
发布将离线超感,将训练完成的模型部署在本地。离线调用模型,可以选择将模型部署在本地的服务器,小型设备、软键一体方案专项适配硬件上, 通过APA、SOK进一步集成,灵活适应不同业务场景。								
发亮彩色系 起射台								
服务器 通用小型设备 专项适配硬件								
PA XC2								
此处发布、下载的SDK为来授何SDK,需要前往拉射台获 <mark>和序列</mark> 号激活后才施正式使用。SDK内排有对应版本约Demo及开发文档,开发者可参考清代码完成开发。								
模型名称	发布版本 🗘	应用平台	模型加速	发布状态	发布时间 🗘			
		通用X86 CPU-Linux	基础版	 已发布 	2021-08-19 20:24	下载SDK		
			精度无损压缩加速	 已发布 	2021-08-19 20:24	下载加速版SDK		
and the transf	134318-V1	英伟达GPU-Linux	基础版	•已发布	2021-08-19 20:35	下载SDK		
on Child Million	查看性能报告		精度无损压缩加速	 已发布 	2021-08-19 20:34	下载加速版SDK		
		Webberg Western	基础版	 已发布 	2021-08-19 18:17	下载SDK		

SDK内bin目录下提供预编译二进制文件,可直接运行(二**进制运行详细说明参考下一小节**),用于图片推理和模型http服务,在二进制参数的 serial_num(或者serial_key)处填入序列号可自动完成联网激活(请确保硬件首次激活时能够连接公网,如果确实不具备联网条件,需要使用纯离 线模式激活,请下载使用百度智能边缘控制台纳管SDK)

```
**SDK内提供的一些二进制文件,填入序列号运行可自动完成激活,以下二进制具体使用说明参考下一小节**
./edgekit_serving --cfg=./edgekit_serving.yml
./easyedge_image_inference {model_dir} {image_name or image_directory} {serial_num}
./easyedge_serving {res_dir} {serial_key} {host} {port}
```

如果是基于源码集成,设置序列号方法如下

global_controller()->set_licence_key("")

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的变量或者源码设置)实例数鉴权环境变量设置方法

export EDGE_CONTROLLER_KEY_AUTH_MODE=2 export EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=30

实例数鉴权源码设置方法

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2) global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 300)

3. 基于预编译二进制测试图片推理和http服务测试图片推理模型资源文件默认已经打包在开发者下载的SDK包中。

请先将tar包整体拷贝到具体运行的设备中,再解压缩编译; 在Intel CPU上运行CPU加速版,如果thirdparty里包含openvino文件夹的,必须在 编译或运行demo程序前执行以下命令: source \${cpp_kit位置路径}/thirdparty/openvino/bin/setupvars.sh 或者执行 source \${cpp_kit位置路 径}/thirdparty/openvino/setupvars.sh(openvino-2022.1+) 如果SDK内不包含setupvars.sh脚本,请忽略该提示

运行预编译图片推理二进制,依次填入模型文件路径(RES文件夹路径)、推理图片、序列号(序列号尽首次激活需要使用,激活后可不用填序列号 也能运行二进制)

./easyedge_image_inference {model_dir} {image_name or image_directory} {serial_num} LD_LIBRARY_PATH=../lib ./easyedge_image_inference ../../../RES /xxx/cat.jpeg "1111-1111-1111"

demo运行效果:



```
> ./easyedge_image_inference ../../../RES 2.jpeg
  2019-02-13 16:46:12,659 INFO [EasyEdge] [easyEdge.cpp:34] 140606189016192 Baidu EasyEdge Linux Development Kit
  0.2.1(20190213)
  2019-02-13 16:46:14,083 INFO [EasyEdge] [paddlev2_edge_predictor.cpp:60] 140606189016192 Allocate graph success.
  2019-02-13 16:46:14,326 DEBUG [EasyEdge] [paddlev2_edge_predictor.cpp:143] 140606189016192 Inference costs 168 ms
  1, 1:txt_frame, p:0.994905 loc: 0.168161, 0.153654, 0.920856, 0.779621
  Done
启动http服务 bin目录下提供编译好的启动http服务二进制文件,可直接运行
```

推荐使用 edgekit_serving 启动模型服务 LD_LIBRARY_PATH=../lib ./edgekit_serving --cfg=./edgekit_serving.yml **也可以使用 easyedge_serving 启动模型服务** **./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401}**

LD_LIBRARY_PATH=../lib ./easyedge_serving ../../../RES "1111-1111-1111" 0.0.0.0 24401

后,日志中会显示

HTTP(or Webservice) is now serving at 0.0.0.0:24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试,网页右侧会展示模型推理结果



对于目标追踪的模型,请选择一段视频,并耐心等待结果



同时,可以调用HTTP接口来访问服务。

请求http服务 以图像预测场景为例(非语义分割模型场景,语义分割请求方式参考后面小节详细文档),提供一张图片,请求模型服务的示例参考 如下demo

python示例代码如下
import requests

C# 使用示例代码如下

```
FileStream fs = new FileStream("./img.jpg", FileMode.Open);
BinaryReader br = new BinaryReader(fs);
byte[] img = br.ReadBytes((int)fs.Length);
br.Close();
fs.Close();
string url = "http://127.0.0.1:8402?threshold=0.1";
HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url);
request.Method = "POST";
Stream stream = request.GetRequestStream();
stream.Write(img, 0, img.Length);
stream.Close();
WebResponse response = request.GetResponse();
StreamBeader sr = new StreamBeaderresponse GetResponse();
```

StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close(); response.Close();

C++ 使用示例代码如下,需要安装curl

```
**include <sys/stat.h>**
**include <curl/curl.h>**
**include <iostream>**
**include <string>**
**define S_ISREG(m) (((m) & 0170000) == (0100000))**
**define S_ISDIR(m) (((m) & 0170000) == (0040000))**
size_t write_callback(void *ptr, size_t size, size_t num, void *data) {
  std::string *str = dynamic_cast<std::string *>((std::string *)data);
  str->append((char *)ptr, size*num);
  return size*num:
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./img.jpg";
  FILE *fp = NULL;
  std::string response;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\"\n", post_data_filename);
     return -1;
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
    curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
    curl_easy_setopt(curl, CURLOPT_POST, 1L);
    curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
    curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
 curl_easy_setopt(curl, CURLOPT_WRITEFUNCTION, write_callback);
    curl_easy_setopt(curl, CURLOPT_WRITEDATA, &response);
    res = curl_easy_perform(curl);
     if (res != CURLE_OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     std::cout << response << std::endl; // response即为返回的json数据
     curl_easy_cleanup(curl);
  curl_global_cleanup();
  fclose(fp);
  return 0:
```

关于http接口的详细介绍参考下面集成文档http服务章节的相关内容

使用说明

使用该方式,将运行库嵌入到开发者的程序当中。 **编译demo项目** SDK src目录下有完整的demo工程,用户可参考该工程的代码实现方式将SDK 集成到自己的项目中,demo工程可直接编译运行:

```
cd src
mkdir build && cd build
cmake .. && make
./easyedge_image_inference {模型RES文件夹} {测试图片路径}
**如果是NNIE引擎,使用sudo运行**
sudo ./easyedge_image_inference {模型RES文件夹} {测试图片路径}
```

(可选) SDK包内一般自带opencv库,可忽略该步骤。如果希望SDK自动编译安装所需要的OpenCV库,修改cmake的 optionEDGE_BUILD_OPENCV为ON即可。 SDK会自动从网络下载opencv源码,并编译需要的module、链接。注意,此功能必须需联网。 cmake -DEDGE_BUILD_OPENCV=ON .. && make -j16

若需自定义library search path或者gcc路径,修改CMakeList.txt即可。

使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

// step 1: 配置模型资源目录 EdgePredictorConfig config; config.model_dir = {模型文件目录};

// step 2: 创建并初始化Predictor;在这里选择合适的引擎 auto predictor = global_controller()->CreateEdgePredictor(config);

// step 3-1: 预测图像 auto img = cv::imread({图片路径}); std::vector<EdgeResultData> results; predictor->infer(img, results);

// step 3-2: 预测视频

std::vector<EdgeResultData> results; FrameTensor frame_tensor; VideoConfig video_config; video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h video_config.source_value = video_src; /* ... more video_configs, 根据需要配置video_config的各选项 */ auto video_decoding = CreateVideoDecoding(video_config); while (video_decoding->next(frame_tensor) == EDGE_OK) { results.clear(); if (frame_tensor.is_needed) { predictor->infer(frame_tensor.frame, results); render(frame_tensor.frame, results, predictor->model_info().kind); //video_decoding->display(frame_tensor); // 显示当前frame , 需在video_config中开启配置 //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置

```
}
```

输入图片不限制大小

SDK参数配置 SDK的参数通过 EdgePredictorConfig::set_config和global_controller()->set_config配置。set_config的所有key 在easyedge_xxxx_config.h中。其中

• PREDICTOR前缀的key是不同模型相关的配置,通过EdgePredictorConfig::set_config设置

• CONTROLLER前缀的key是整个SDK的全局配置,通过global_controller()->set_config设置

```
以序列号为例,KEY的说明如下:
```

```
/**
* @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号
* 值类型:string
* 默认值:空
*/
static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";
```

EdgePredictorConfig config; config.model_dir = ...; config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, "1DB7-1111-1111-D27D");

具体支持的运行参数配置列表可以参考开发工具包中的头文件的详细说明。

相关配置均可以通过环境变量的方法来设置,对应的key名称加上前缀EDGE_即为环境变量的key。如序列号配置的环境变量key为EDGE_PREDICTOR_KEY_SERIAL_NUM,如指定CPU线程数的环境变量key为EDGE_PREDICTOR_KEY_CPU_THREADS_NUM。 注意:通过代码设置的配置会覆盖通过环境变量设置的值。

初始化

● 接口

auto predictor = global_controller()->CreateEdgePredictor(config); predictor->init();

若返回非0,请查看输出日志排查错误原因。

预测图像

• 接口

/** * @brief * 通用接口 * @param image: must be BGR , HWC format (opencv default) * @param result * @return */ virtual int infer(cv::Mat& image, std::vector<EdgeResultData>& result) = 0; /** * @brief * 批量图片推理接口 * @param image: must be BGR , HWC format (opencv default) * @param result * @return */ virtual int infer(std::vector<cv::Mat>& image, std::vector<std::vector<EdgeResultData>>& result) = 0;

图片的格式务必为opencv默认的BGR, HWC格式。

• 返回格式

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index: // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测、图像分割时才有意义
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割的模型,该字段才有意义
    // 请注意: 图像分割时,以下两个字段会比较大,使用完成之后请及时释放EdgeResultData
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding,游程编码的mask
    // 目标追踪模型,该字段才有意义
    int trackid; // 轨迹id
    int frame; // 处于视频中的第几帧
```

关于矩形坐标

};

x1*图片宽度 = 检测框的左上角的横坐标

EdgeTrackStat track_stat; // 跟踪状态

- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2*图片宽度 = 检测框的右下角的横坐标
- y2 * 图片高度 = 检测框的右下角的纵坐标

关于图像分割mask

cv::Mat mask为图像掩码的二维数组 { {0,0,0,0,0,0,0,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,0,0,0,0,0,0,0}, {1,1,1,1,0,0,0,0}, {1,1,1,1,0,0,0}, {1,1,1,1,0,0,0}, {1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,

关于图像分割mask_rle

该字段返回了mask的游程编码,解析方式可参考 http demo

以上字段可以参考demo文件中使用opencv绘制的逻辑进行解析

预测视频

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

接口

classVideoDecoding :

/**

- * @brief 获取输入源的下一帧
- * @param frame_tensor
- * @return

*/

virtual int next(FrameTensor &frame_tensor) = 0;

/**

- * @brief 显示当前frame_tensor中的视频帧
- * @param frame_tensor
- * @return
- */

virtual int display(const FrameTensor &frame_tensor) = 0;

/**

* @brief 将当前frame_tensor中的视频帧写为本地视频文件

- * @param frame_tensor
- * @return

*/

virtual int save(FrameTensor & frame_tensor) = 0;

```
/**
```

```
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
* @brief 获取视频的width属性
* @return
*/
virtual int get_width() = 0;
/**
```

```
* @brief 获取视频的height属性
* @return
*/
```

```
virtual int get_height() = 0;
```

struct VideoConfig

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
                        // 输入源类型
  SourceType source_type;
  std::string source_value;   // 输入源地址,如视频文件路径、摄像头index、网络流地址
                      // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
 int skip_frames{0};
  int retrieve_all{false};
                     // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
                      // 在采取抽帧之前设置视频的fps
  int input_fps{0};
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
 bool enable_display{false};
  std::string window_name{"EasyEdge"};
  bool display_all{false};
                       // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                       // 是否存储所有frame, 若为false, 仅存储根据skip_frames抽取的frame
  std::map::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被

Baidu 百度智能云文档

抽取返回,以作为显示或存储用。 input_fps:用于抽帧前设置fps。 resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。 conf:高级选项。部分配置会通过该map来设置。

注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

设置序列号

请在网页控制台中申请序列号,并在init初始化前设置。 LinuxSDK 首次使用需联网授权。

EdgePredictorConfig config; config.set_config(easyedge::params::PREDICTOR_KEY_SERIAL_NUM, "this-is-serial-num");

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

```
EdgeLogConfig log_config;
log_config.enable_debug = true;
global_controller()->set_log_config(log_config);
```

http服务

1. 开启http服务 http服务的启动可以参考demo_serving.cpp文件。

/**

- * @brief 开启一个简单的demo http服务。
- * 该方法会block直到收到sigint/sigterm。
- * http服务里,图片的解码运行在cpu之上,可能会降低推理速度。
- * @tparam ConfigT
- * @param config
- * @param host
- * @param port
- * @param service_id service_id user parameter, uri '/get/service_id' will respond this value with 'text/plain'
- * @param instance_num 实例数量,根据内存/显存/时延要求调整
- * @return

```
*/
template<typename ConfigT>
int start_http_server(
    const ConfigT &config,
    const std::string &host,
    int port,
    const std::string &service_id
```

int instance_num = 1);

2. http接口详细说明

开发者可以打开浏览器,http://{设备ip}:24401,选择图片或视频来进行测试。

http 请求方式一:无额外编码 URL中的get参数:

参数	说明		默认值
threshold	阈值过滤,	0~1	如不提供,则会使用模型的推荐阈值

Python请求示例 (图片测试,针对图像分类、物体检测、实例分割等模型)

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
        'http://127.0.0.1:24401/',
        params={'threshold': 0.1},
        data=img).json()
```

Python请求示例 (图片测试, 仅针对语义分割模型, 同其他CV模型不同, 语义分割模型输出为灰度图)

```
import requests
with open('./1.jpg', 'rb') as f:
    img_data = f.read()
    res = requests.post('http://127.0.0.1:24401/',
        data=img_data)
with open("gray_result.png", "wb") as fb:
    fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
```

Python请求示例 (视频测试, 注意:区别于图片预测,需指定Content-Type;否则会调用图片推理接口)

```
import requests
with open('./1.mp4', 'rb') as f:
video_data = f.read()
result = requests.post(
    'http://127.0.0.1:24401/',
    params={'threshold': 0.1},
    headers={'Content-Type': 'video'},
    data=video_data).json()
```

http 请求方法二:json格式,图片传base64格式字符串 HTTP方法:POST Header如下:

参数	值
Content-Type	application/json

Body请求填写:

• 图像分类网络: body 中请求示例

```
{
"image": "<base64数据>",
"top_num": 5
}
```

body中参数详情

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
top_nu m	否	number	-	返回分类数量,不填该参数,则默认返回全部分类结果

• 物体检测和实例分割网络: Body请求示例:

```
{
"image": "<base64数据>",
"threshold": 0.3
```

body中参数详情:

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
thresho Id	否	number	-	默认为推荐阈值,也可自行根据需要进行设置

• 语义分割网络: Body请求示例:

```
{
"image": "<base64数据>"
}
```

body中参数详情(语义分割由于模型特殊性,不支持设置threshold值,设置了也没有意义):

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部

Python请求示例(非语义分割模型参考如下代码)

```
import base64
import requests

def main():
    with open("1.jpg 【图片路径】", 'rb') as f:
        result = requests.post("http://{服务ip地址}:24401/", json={
            "image": base64.b64encode(f.read()).decode("utf8")
        })
        # print(result.request.body)
        # print(result.request.headers)
        print(result.content)

if __name__ == '__main__':
        main()
```

Python 请求示例 (针对语义分割模型,同其他CV模型不同,语义分割模型输出为灰度图)

```
import base64
import requests
def main():
    with open("1.jpg 【图片路径】", 'rb') as f:
        res = requests.post("http://{服务ip地址}:24401/", json={"image": base64.b64encode(f.read()).decode("utf8")})
        with open("gray_result.png", "wb") as fb:
        fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
if __name__ == '__main__':
        main()
```

http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考预测图像-返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

返回示例

```
{
    "cost_ms": 52,
    "error_code": 0,
    "results": [
        {
            "confidence": 0.94482421875,
            "index": 1,
            "label": "IronMan",
            "x1": 0.059185408055782318,
            "x2": 0.18795496225357056,
            "y1": 0.14762254059314728,
            "y2": 0.52510076761245728,
            "mask": "...", // 图像分割模型字段
            "trackId": 0, // 目标追踪模型字段
        },
        ]
}
```

其他配置

1. 日志名称、HTTP 网页标题设置

通过global_controller的set_config方法设置:

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_LOG_BRAND, "MY_BRAND");

效果如下:



图片加载失败

2. CPU线程数设置

CPU线程数可通过 EdgePredictorConfig::set_config 配置

```
EdgePredictorConfig config;
config.set_config(easyedge::params::PREDICTOR_KEY_CPU_THREADS_NUM, 4);
```

3. 批量预测设置

```
int batch_size = 2; // 使用前修改batch_size再编译、执行
while (get_next_batch(imgs, img_files, batch_size, start_index)) {
    ...
}
```

GPU 加速版 预测接口 GPU 加速版 SDK 除了支持上面介绍的通用接口外,还支持图片的批量预测,预测接口如下:

```
/**
 * @brief
* GPU加速版批量图片推理接口
* @param image: must be BGR , HWC format (opency default)
* @param result
* @return
*/
virtual int infer(
     std::vector<cv::Mat>& image
     std::vector<std::vector<EdgeResultData>>& result
) = 0:
/**
* @brief
* GPU加速版批量图片推理接口,带阈值
* @related infer(cv::Mat & image, EdgeColorFormat origin_color_format, std::vector<EdgeResultData> & result, float threshold)
*/
virtual int infer(
       std::vector<cv::Mat> &images,
       EdgeColorFormat origin_color_format,
       std::vector<std::vector<EdgeResultData>> &results
       float threshold
) = 0;
```

批量图片的预测接口的使用要求在调用 init 接口的时候设置一个有效的 PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE,其含义见下方参数配置接口的介绍。

运行参数选项 在上面的内容中我们介绍了如何使用EdgePredictorConfig进行运行参数的配置。针对GPU加速版开发工具包,目前EdgePredictorConfig的运行参数所支持的Key包括如下项:

```
/**
* @brief 当有同类型的多个设备的时候,使用哪一个设备,如:
* GPU: 使用哪张GPU卡
* EdgeBoard(VMX), Movidius NCS:使用哪一张加速卡
* 值类型: int
* 默认值:0
*/
static constexpr auto PREDICTOR_KEY_DEVICE_ID = "PREDICTOR_KEY_DEVICE_ID";
/**
* @brief 生成最大 batch_size 为 max_batch_size 的优化模型,单次预测图片数量可以小于或等于此值(推荐等于此值,见release notes)
* 值类型: int
* 默认值:4
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE = "PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE";
/**
* @brief 设置device对应的GPU卡可以支持的最大并发量
* 实际预测的时候对应GPU卡的最大并发量不超过这里设置的范围,否则预测请求会排队等待预测执行
* 值类型: int
* 默认值:1
*/
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY = "PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY";
/**
* @brief 是否开启fp16模式预测,开启后预测速度会更快,但精度会略有降低。并且需要硬件支持fp16
* 值类型: bool
* 默认值: false
*/
static constexpr auto PREDICTOR_KEY_GTURBO_FP16 = "PREDICTOR_KEY_GTURBO_FP16";
/**
* @brief 模型编译等级
*1:如果当前max_batch_size与历史编译产出的max_batch_size不相等时,则重新编译模型(推荐)
* 2:无论历史编译产出的max_batch_size为多少,均根据当前max_batch_size重新编译模型
* 值类型: int
* 默认值:1
*/
```

static constexpr auto PREDICTOR_KEY_GTURBO_COMPILE_LEVEL = "PREDICTOR_KEY_GTURBO_COMPILE_LEVEL"; /** * @brief GPU工作空间大小设置 * workspace_size = workspace_prefix * (1 << workspace_offset) * workspace_offset: 10 = KB, 20 = MB, 30 = GB * 值类型: int *默认值:WORKSPACE_PREFIX: 100, WORKSPACE_OFFSET: 20,即100MB static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX = "PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX"; static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET = "PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET"; /** * @brief 需要使用的dla core * 值类型: int * 默认值:-1(不使用) static constexpr auto PREDICTOR_KEY_GTURBO_DLA_CORE = "PREDICTOR_KEY_GTURBO_DLA_CORE"; /** * @brief 自定义缓存文件存储路径 * 值类型: string * 默认值: ~/.baidu/easyedge/mcache/{model_id * 1000000 + release_id} static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_DIR = "PREDICTOR_KEY_GTURBO_CACHE_DIR"; /** * @brief 自定义缓存文件命名,默认即可 * 值类型: string *默认值:根据配置自动生成 static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_NAME = "PREDICTOR_KEY_GTURBO_CACHE_NAME"; /** * @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号 * 值类型: string *默认值:空 static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";

PREDICTOR_KEY_GTURBO_CACHE_NAME:首次加载模型会先对模型进行编译优化,通过此值可以设置优化后的产出文件名,这在多进程加载同一个模型的时候是有用的。

PREDICTOR_KEY_GTURBO_CACHE_DIR:首次加载模型经过编译优化后,产出的优化文件会存储在这个位置,可以按需修改。

PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX、PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET:设置运行时可以被用来使用的最大临时显存。

PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE:此值用来控制批量图片预测可以支持的最大图片数,实际预测的时候单次预测图片数需等于此值。

PREDICTOR_KEY_DEVICE_ID:设置需要使用的 GPU 卡号。

PREDICTOR_KEY_GTURBO_COMPILE_LEVEL: 模型编译等级。通常模型的编译会比较慢,但编译产出是可以复用的。可以在第一次加载模型的时候设置合理的 max_batch_size 并在之后加载模型的时候直接使用历史编译产出。是否使用历史编译产出可以通过此值 compile_level 来控制。当此值为 0 时,表示忽略当前设置的 max_batch_size 而仅使用历史产出(无历史产出时则编译模型);当此值为 1 时,会比较历史产出和当前设置的 max_batch_size 是否相等,如不等,则重新编译;当此值为 2 时,无论如何都会重新编译模型。

PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY:通过此值设置单张 GPU 卡上可以支持的最大 infer 并发量,其上限取决于硬件限制。init 接口 会根据此值预分配 GPU 资源,建议结合实际使用控制此值,使用多少则设置多少。注意:此值的增加会降低单次 infer 的速度,建议优先考虑 batch inference 和 multi predictor。

PREDICTOR_KEY_GTURBO_FP16:默认是 fp32 模式,置 true 可以开启 fp16 模式预测,预测速度会有所提升,但精度也会略微下降,权衡使用。注意:不是所有模型都支持 fp16 模式。目前已知不支持fp16的模型包括:图像分类高精度模型。

多线程预测 GPU 加速版 SDK 的多线程分为单卡多线程和多卡多线程两种。 单卡多线程:创建一个 predictor,并通过 PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY 控制单卡所支持的最大并发量,只需要 init 一次,多线程调用 infer 接口。 多卡多线程:多卡的 支持是通过创建多个 predictor,每个 predictor 对应一张 GPU 卡,predictor 的创建和 init 的调用放在主线程,通过多线程的方式调用 infer 接口。

已知问题 1. 多线程时图片按线程分配不均 或 不同batch size的图片交叉调用infer接口时,部分结果错误 A: EasyDL图像分类高精度模型在有些显卡上可能存在此问题,可以考虑填充假图片数据到图片比较少的线程或batch以使得infer间的图片绝对平均。

2. 显存持续增长或遇到 terminate called after throwing an instance of 'std::runtime_error' what(): Failed to create object A: 部分显卡存在此问题, 如果遇到此问题, 请确认没有频繁调用 init 接口, 通常调用 infer 接口即可满足需求。

3. 开启 fp16 后,预测结果错误 A:不是所有模型都支持 fp16 模式。目前已知的不支持fp16的模型包括:图像分类高精度模型。目前不支持的 将会在后面的版本陆续支持。

昆仑服务器 昆仑服务器SDK支持将EasyDL的模型部署到昆仑服务器上。SDK提供的接口风格一致,简单易用,轻松实现快速部署。Demo的测试 可参考上文中的测试Demo部分。

参数配置接口 在上面的内容中我们介绍了如何使用EdgePredictorConfig进行运行参数的配置。针对昆仑服务器开发工具包,目前EdgePredictorConfig的运行参数所支持的Key包括如下项:

```
/**
* @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号
* 值类型: string
* 默认值:空
*/
static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";
/**
* @brief 当有同类型的多个设备的时候,使用哪一个设备,如:
* 使用哪张加速卡
* 值类型: int
*默认值:0
*/
static constexpr auto PREDICTOR_KEY_DEVICE_ID = "PREDICTOR_KEY_DEVICE_ID";
/**
* @brief 设置需要同时预测的图片数量
* 值类型: int
*默认值:1
*/
static constexpr auto PREDICTOR_KEY_KUNLUN_BATCH_SIZE = "PREDICTOR_KEY_KUNLUN_BATCH_SIZE";
```

PREDICTOR_KEY_DEVICE_ID:设置需要使用的加速卡的卡号。

PREDICTOR_KEY_KUNLUN_BATCH_SIZE:设置单次预测可以支持的图片数量。

使用方法:

```
int batch_size = 1;
config.set_config(easyedge::params::PREDICTOR_KEY_KUNLUN_BATCH_SIZE, batch_size);
```

模型调优通过设置如下环境变量,可以在初始化阶段对模型调优,从而让预测的速度更快。

export XPU_CONV_AUTOTUNE=5

FAQ

1. 如何处理一些 undefined reference?

如: undefined reference to `curl_easy_setopt@CURL_OPENSSL_3'

方案1:通过安装libcurl3 libcurl-openssl1.0-dev来解决。 方案2:如果开发者想不想使用低版本的openssl (如Ubuntu 18.04),可以link静态 库easyedge_static.a,自己指定需要的Library的版本: find_package(CURL REQUIRED)

target_link_libraries(easyedge_demo \${OpenCV_LIBS} easyedge_static pthread \${CURL_LIBRARIES} verify_static \$(其他需要的库))

其中,其他需要的库视具体sdk中包含的库而定。

2. EasyDL SDK与云服务效果不一致,如何处理?

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

3. NVIDIA GPU预测时,报错显存不足如以下错误字样:

paddle.fluid.core.EnforceNotMet: Enforce failed. Expected allocating <= available, but received allocating:20998686233 > available:19587333888. Insufficient GPU memory to allocation. at [/paddle/paddle/fluid/platform/gpu_info.cc:170]

请根据显存大小和模型配置。调整合适的初始 fraction_of_gpu_memory。参数的含义参考这里。

4. 如何将我的模型运行为一个http服务?目前cpp sdk暂未集成http运行方式; 0.4.7版本之后,可以通过start_http_server方法开启http服务。

5. 运行NNIE引擎报permission denied 日志显示:

open sys: Permission denied open err : Permission denied open err : Permission denied

请使用sudo在root下运行。

6. 运行SDK报错 Authorization failed

情况一:日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888" ./easyedge_demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除~/.baidu/easyedge 目录,再重新激活。

7. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

8. 运行二进制时,提示 libverify.so cannot open shared object file

可能cmake没有正确设置rpath,可以设置LD_LIBRARY_PATH为sdk的lib文件夹后,再运行:

LD_LIBRARY_PATH=\$LD_LIBRARY_PATH:./lib ./easyedge_demo

9. 运行二进制时提示 libopencv_videoio.so.4.5: cannot open shared object file: No such file or directory 同上面8的问题类似,没有正确设置动态 库的查找路径,可通过设置LD_LIBRARY_PATH为sdk的thirdparty/opencv/lib文件夹解决 export LD_LIBRARY_PATH=\${LD_LIBRARY_PATH}:../thirdparty/opencv/lib (tips: 上面冒号后面接的thirdparty/opencv/lib路径以实际项目中路径为准,比如也可能是../../thirdparty/opencv/lib)

10. 编译时报错:file format not recognized 可能是因为在复制SDK时文件信息丢失。请将整个压缩包复制到目标设备中,再解压缩、编译

11. 进行视频解码时,报错符号未找到、格式不支持、解析出的图片为空、无法设置抽帧 请确保安装OpenCV时,添加了-DWITH_FFMPEG=ON选项(或者GStream选项),并且检查OpenCV的安装日志中,关于 Video I/O 段落的说明是否为YES。

- -- Video I/O: -- DC1394: YES (ver 2.2.4)
- -- FFMPEG: YES
- -- avcodec: YES (ver 56.60.100)
- -- avformat: YES (ver 56.40.101)
- -- avutil: YES (ver 54.31.100)
- -- swscale: YES (ver 3.1.101) -- avresample: NO
- -- avresample: NO -- libv4l/libv4l2: NO
- -- v4l/v4l2: linux/videodev2.h

如果为NO,请搜索相关解决方案,一般为依赖没有安装,以apt为例:

apt-get install yasm libjpeg-dev libjasper-dev libavcodec-dev libavformat-dev libswscale-dev libdc1394-22-dev libgstreamer0.10-dev libgstreamer0.10-dev libgstreamer-plugins-base0.10-dev libv4l-dev python-numpy libtbb-dev libqt4-dev libgtk2.0-dev libfaac-dev libmp3lame-dev libopencore-ammb-dev libopencore-ammb-dev libtheora-dev libtheora-dev libtvidcore-dev x264 v4l-utils ffmpeg

12. GPU加速版运行有损压缩加速的模型,运算精度较标准模型偏低首先请保证数据集不会太小。其次可以通过将模型目录RES中的calibrationtable移除,并通过将PREDICTOR_KEY_GTURBO_FP16设置为true,使用FP16的运算精度重新评估模型效果。若依然不理想,可将calibrationtable移除并将PREDICTOR_KEY_GTURBO_FP16设置为false,从而使用更高精度的FP32的运算精度。

の Linux集成文档-Python

简介

本文档介绍 EasyDL 的 Linux Python SDK 的使用方法,适用于 EasyDL 和 BML。

EasyDL 通用版:

- 网络类型支持:图像分类,物体检测,图像分割,声音分类,表格预测
- 硬件支持:
 - Linux x86_64 CPU (基础版,加速版)
 - Linux x86_64 Nvidia GPU (基础版,加速版)
- 语言支持: Python 3.5, 3.6, 3.7

BML:

- 网络类型支持:图像分类,物体检测,声音分类
- 硬件支持:
 - Linux x86_64 CPU (基础版)
 - Linux x86_64 Nvidia GPU (基础版)
- 语言支持: Python 3.5, 3.6, 3.7

Release Notes

Baidu 百度智能云文档

时间	版本	说明
2023-03-16	1.3.7	迭代升级,新增支持文本类模型; 新增GPU 多卡多进程推理demo
2022.10.27	1.3.5	新增华为Atlas300、飞腾Atlas300 Python SDK,支持图像分类、物体检测、人脸检测、实例分割
2022.09.15	1.3.3	EasyDL CPU普通版新增支持表格预测
2022.05.27	1.3.1	CPU、GPU普通版新增支持BML Cloud小目标检测模型
2021.12.22	1.2.7	声音分类模型升级
2021.10.20	1.2.6	CPU基础版、CPU加速版、GPU基础版推理引擎优化升级
2021.08.19	1.2.5	CPU基础版、CPU无损加速版、GPU基础版新增支持EasyDL小目标检测
2021.06.29	1.2.4	CPU、GPU新增EasyDL目标跟踪支持;新增http server服务启动demo
2021.03.09	1.2.2	EasyDL CPU加速版新增支持分类、高性能检测和均衡检测的量化压缩模型
2021.01.27	1.2.1	EasyDL经典版分类高性能模型升级;支持更多模型
2020.12.18	1.2.0	推理引擎升级;接口升级;性能优化
2020.09.17	1.1.19	支持更多模型
2020.08.11	1.1.18	性能优化
2020.06.23	1.1.17	支持更多EasyDL专业版模型
2020.04.16	1.1.15	技术优化;升级 OpenVINO 版本
2020.03.12	1.1.14	新增声音识别python sdk
2020.02.12	1.1.13	新增口罩模型支持
2020.01.16	1.1.12	预测函数默认使用推荐阈值
2019.12.26	1.1.11	EasyDL 专业版支持 SDK 加速版
2019.12.04	1.1.10	支持图像分割
2019.10.21	1.1.9	支持 EasyDL 专业版
2019.08.29	1.1.8	CPU 加速版支持
2019.07.19	1.1.7	提供模型更新工具
2019.05.16	1.1.3	NVIDIA GPU 支持
2019.03.15	1.1.0	架构与功能完善
2019.02.28	1.0.6	引擎功能完善
2019.02.13	1.0.5	paddlepaddle 支持
2018.11.30	1.0.0	第一版!

2020-12-18: 【接口升级】 序列号的配置接口从1.2.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。 请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

快速开始

1. 安装依赖

- 根据引擎的不同, SDK 依赖了不同的底层引擎。根据所需自行安装。
- 使用声音分类SDK需要安装额外依赖*pip 安装 resampy pydub six librosa 音频默认格式支持wav文件预测,如果需要预测mp3等其他音频格式的数据需要系统额外安装ffmpeg (windows系统的ffmpeg已基在sdk中无需额外安装, linux系统需要手动安装)
- 使用表格预测SDK需要安装额外依赖 pip安装 brotlipy==0.7.0 certifi==2020.6.20 joblib==1.0.1 kaggle==1.5.12 Pillow py4j pycosat pythondateutil python-slugify ruamel_yaml text-unidecode threadpoolctl flask pandas==1.0.5 scikit-learn==0.23.2 lightgbm==2.2.3 catboost==0.24.1 xgboost==1.2.0 numpy==1.19.5 scipy==1.5.2 psutil==5.7.2 pypmml==0.9.7 torch==1.8.0 jieba==0.42.1 pyod==0.8.5 pyarrow==6.0.0 scikit-optimize==0.9.0 pyspark==3.3.0 另外ml算法 安装(目前只支持python3.7) pip install BaiduAl_TabularInfer-0.0.0-cp37-cp37m-linux_x86_64.whl 安装 paddlepaddle

• 使用x86_64 CPU 基础版预测时必须安装(目标跟踪除外):

python -m pip install paddlepaddle==2.2.2 -i https://mirror.baidu.com/pypi/simple

若 CPU 为特殊型号,如赛扬处理器(一般用于深度定制的硬件中),请关注 CPU 是否支持 avx 指令集。如果不支持,请在paddle官网安装 noavx 版本

• 使用NVIDIA GPU 基础版预测时必须安装(目标跟踪除外):

python -m pip install paddlepaddle-gpu==2.2.2.post101 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA10.1的 PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2 -i https://mirror.baidu.com/pypi/simple #CUDA10.2的PaddlePaddle
python -m pip install paddlepaddle-gpu==2.2.2.post110 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.0的
PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2.post111 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.1的 PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2.post112 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.2的 PaddlePaddle

不同cuda版本的环境,请参考paddle文档安装合适的 paddle 版本。 不被 paddle 支持的 cuda 和 cudnn 版本,EasyEdge 暂不支持**安装 OpenVINO** 使用x86_64 CPU 加速版 SDK 预测时必须安装。

1) 请参考 OpenVINO toolkit 文档安装 2021.4版本, 安装时可忽略Configure the Model Optimizer及后续部分

2) 运行之前,务必设置环境变量

source /opt/intel/openvino_2021/bin/setupvars.sh

安装 cuda、cudnn

• 使用Nvidia GPU 加速版预测时必须安装。依赖的版本为 cuda9.0、cudnn7。版本号必须正确。

安装 pytorch (torch >= 1.7.0)

- 目标跟踪模型的预测必须安装pytorch版本1.7.0及以上(包含:Nvidia GPU 基础版、x86_64 CPU 基础版)。
- 目标跟踪模型Nvidia GPU 基础版还需安装依赖cuda、cudnn。

关于不同版本的pytorch和CUDA版本的对应关系: pytorch官网目标跟踪模型还有一些列举在requirements.txt里的依赖(包括torch >= 1.7.0),均可使用pip下载安装。

pip3 install -r requirements.txt

2. 安装 easyedge python wheel 包 安装说明

pip3 install -U BaiduAI_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_x86_64.whl

具体名称以 SDK 包中的 whl 为准。 安装说明:华为 Atlas300 除了需要安装BaiduAI_EasyEdge_SDK包,还需安装

pip3 install -U EasyEdge_Devkit_Atlas300-{版本号}-cp36-cp36m-linux_x86_64.whl

安装说明:飞腾 Atlas300 除了需要安装BaiduAl_EasyEdge_SDK包,还需安装

pip3 install -U EasyEdge_Devkit_Phytium.Atlas-{版本号}-cp36-cp36m-linux_aarch64.whl

3. 使用序列号激活

作化而 (2.3.45 %) V(3) f										
	发布结果结婚的,将说的完成的情况都是有生态。我就说明我提出,可以选择将做型部确在生地的服务值,小型设备,就是一体方面专项运配操作上。 通过AFA,它XXH———————————————————————————————————									
共雨向河口	发气频应多 控制台									
	服务器 通用小型设备 专项适配操作									
	SDK API									
3大4以17ツリ 5	此处发布、下载的SOK为未很的SOK,需要能往控制台赶就序列号激活后才能正式使用,SDK内附有对应版本的Demo及开发文档,开发者可参考原代码完成开发,									
	模型名称	发布版本 关	应用平台	模型加速	发布状态	发布时间 🗘				
		134318-V1 宣看性能投告	通用X86 CPU-Linux	基础板	 已发布 	2021-08-19 20:24	下载SDK			
				精度无损压缩加速	 已发布 	2021-08-19 20:24	下载加速版SDK			
	aug du Plateteet		英伟达GPU-Linux	基础板	 已发布 	2021-08-19 20:35	下截SDK			
	on Chiching Million			精度无损压缩加速	 已发布 	2021-08-19 20:34	下载加速版SDK			
				基础版	 已发布 	2021-08-19 18:17	下號SDK			

修改demo.py 填写序列号

pred = edge.Program() pred.set_auth_license_key("这里填写序列号")

4. GPU 加速版 使用 GPU 加速版,在安装完 whl 之后,必须:

1. 从这里下载 TensorRT7.0.0.11 for cuda9.0,并把解压后的 lib 放到 C++ SDK 的 lib 目录或系统 lib 目录

2. 运行时,必须在系统库路径中包含 C++ SDK 下的lib目录。如设置LD_LIBRARY_PATH

```
cd ${SDK_ROOT}
**1. 安装 python wheel 包**
tar -xzvf python/*.tar.gz
pip install -U {对应 Python 版本的 wheel 包}
**2. 设置 LD_LIBRARY_PATH**
tar -xzvf cpp/*.tar.gz
```

export EDGE_ROOT=\$(readlink -f \$(ls -h | grep "baidu_easyedge_linux_cpp")) export LD_LIBRARY_PATH=\$EDGE_ROOT/lib

3. 运行 demo python3 demo.py {RES文件夹路径} {测试图片路径}

如果是使用 C++ SDK 自带的编译安装的 OpenCV, LD_LIBRARY_PATH 还需要包括 C++ SDK的 build 目录下的 thirdparty/lib 目录

```
如果没有正确设置 LD_LIBRARY_PATH,运行时可能报错:
```

ImportError: libeasyedge.so.0.4.3: cannot open shared object file: No such file or directory ImportError: libopencv_core.so.3.4: cannot open shared object file: No such file or directory

5. 测试 Demo

5.1 图片预测

输入对应的模型文件夹(默认为RES)和测试图片路径,运行:

python3 demo.py {model_dir} {image_name.jpg}

测试效果:



图片加载失败

5.2 视频预测 (适用于目标跟踪)

输入对应的模型文件夹(默认为RES)和测试视频文件路径/摄像头id/网络视频流地址,运行:

```
**video_type: 输入源类型 type:int**
**1 本地视频文件**
**2 摄像头的index**
**3 网络视频流**
**video_src: 输入源地址,如视频文件路径、摄像头index、网络流地址 type: string**
python3 demo.py {model_dir} {video_type} {video_src}
```

```
6. 测试Demo HTTP 服务 输入对应的模型文件夹(默认为RES)、序列号、设备ip和指定端口号,运行:
```

 $python 3 \ demo_serving.py \ \{model_dir\} \ \{serial_key\} \ \{host, \ default \ 0.0.0.0\} \ \{port, \ default \ 24401\}$

后,会显示:

Running on http://0.0.0.0:24401/

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片或者视频来进行测试。也可以参考`demo_serving.py`里 http_client_test()函数请求http服务进行推理。



使用说明

使用流程 demo.py

import BaiduAI.EasyEdge as edge

pred = edge.Program() pred.set_auth_license_key("这里填写序列号") pred.init(model_dir={RES文件夹路径}, device=edge.Device.CPU, engine=edge.Engine.PADDLE_FLUID) pred.infer_image({numpy.ndarray的图片}) pred.close()

demo_serving.py

import BaiduAl.EasyEdge as edge from BaiduAl.EasyEdge.serving import Serving

 $server = Serving(model_dir=\{RES \chi free key), \ license=serial_key)$

```
**请参考同级目录下demo.py里:**
```

```
** pred.init(model\_dir=xx, \ device=xx, \ engine=xx, \ device\_id=xx)**
```

```
**对以下参数device\device_id和engine进行修改**
```

```
server.run(host=host, \ port=port, \ device=edge.Device.CPU, \ engine=edge.Engine.PADDLE\_FLUID)
```

初始化

• 接口

```
def init(self,
    model_dir,
    device=Device.CPU,
    engine=Engine PADDLE_FLUID,
    config_file='conf.json',
    preprocess_file='preprocess_args.json',
    model_file='model',
    params_file='params'
    label_file='label_list.txt',
    infer_cfg_file='infer_cfg.json',
    device_id=0,
    thread_num=1
    ):
  ....
  Args:
    model_dir: str
     device: BaiduAI.EasyEdge.Device,比如:Device.CPU
     engine: BaiduAI.EasyEdge.Engine,比如:Engine.PADDLE_FLUID
     config_file: str
     preprocess_file: str
     model_file: str
     params_file: str
     label_file: str 标签文件
     infer_cfg_file: 包含预处理、后处理信息的文件
device_id: int 设备ID
    thread_num: int CPU的线程数
  Raises:
     RuntimeError, IOError
```

RuntimeError, IOError Returns: bool: True if success

使用 NVIDIA GPU 预测时,必须满足:

- 机器已安装 cuda, cudnn
- 已正确安装对应 cuda 版本的 paddle 版本
- 通过设置环境变量 FLAGS_fraction_of_gpu_memory_to_use 设置合理的初始内存使用比例

使用 CPU 预测时,可以通过在 init 中设置 thread_num 使用多线程预测。如:

 $pred.init(model_dir=_model_dir,\ device=edge.Device.CPU,\ engine=edge.Engine.PADDLE_FLUID,\ thread_num=1)$

预测图像

● 接口

def infer_im	age(self, img,
thr	eshold=0.3,
cha	annel_order='HWC',
col	or_format= <mark>'BGR'</mark> ,
da	ta_type='numpy'):
Args:	
img: np.	ndarray or bytes
threshol	d: float
only r	eturn result with confidence larger than threshold
channel	_order: string
chanr	el order HWC or CHW
color_fo	rmat: string
color	format order RGB or BGR
data_typ	e: string
image	e data type
Returns:	
list	

• 返回格式: [dict1, dict2, ...]

字段	类型	取值	说明
confidence	float	0~1	分类或检测的置信度
label	string		分类或检测的类别
index	number		分类或检测的类别
x1, y1	float	0~1	物体检测,矩形的左上角坐标 (相对长宽的比例值)
x2, y2	float	0~1	物体检测,矩形的右下角坐标(相对长宽的比例值)
mask	string/numpy.ndarray	图像分割的mask	

关于矩形坐标

- x1 * 图片宽度 = 检测框的左上角的横坐标
- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2 * 图片宽度 = 检测框的右下角的横坐标
- y2 * 图片高度 = 检测框的右下角的纵坐标

可以参考 demo 文件中使用 opencv 绘制矩形的逻辑。

结果示例

• i) 图像分类

```
{
    "index": 736,
    "label": "table",
    "confidence": 0.9
}
```

● ii) 物体检测

Baidu 百度智能云文档

```
"y2": 0.91211,
"label": "cat",
"confidence": 1.0,
"x2": 0.91504,
"index": 8,
"y1": 0.12671,
"x1": 0.21289
}
```

• iii) 图像分割

```
{
    "name": "cat",
    "score": 1.0,
    "location": {
        "left": ...,
        "top": ...,
        "width": ...,
        "height": ...,
    },
    "mask": ....
}
```

mask字段中,data_type为numpy时,返回图像掩码的二维数组

```
{
{0,0,0,0,0,0,0,0,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,0,0,0,0,0,0,0},
}
其中1代表为目标区域,0代表非目标区域
```

data_type为string时,mask的游程编码,解析方式可参考 demo

预测视频 (目前仅限目标跟踪模型调用)

```
• 接口
```

```
def infer_frame(self, frame, threshold=None):
"""
视频推理(抽帧之后)
:param frame:
:param threshold:
:return:
"""
```

```
● 返回格式dict
```

字段	类型	说明
pos	dict1	当前帧每一个类别的追踪目标的像素坐标(tlwh)
id	dict2	当前帧每一个类别的追踪目标的id
score	dict3	当前帧每一个类别的追踪目标的识别置信度
label	dict4	class_idx(int)与label(string)的对应关系
class_num	int	追踪类别数

预测声音

• 使用声音分类SDK需要安装额外依赖 *pip 安装 resampy pydub* 音频默认格式支持wav文件预测,如果需要预测mp3等其他音频格式的数据需要 系统额外安装ffmpeg (windows系统的ffmpeg已集成在sdk中无需额外安装,linux系统需要手动安装)

接口

• 返回格式: [dict1, dict2, ...]

字段	类型	取值	说明
confidence	float	0~1	分类的置信度
label	string		分类的类别
index	number		分类的类别

升级模型 适用于经典版升级模型,执行bash update_model.sh</mark>,根据提示,输入模型路径、激活码、模型ID、模型版本,等待模型更新完毕即 可。

FAQ

Q: EasyDL 离线 SDK 与云服务效果不一致,如何处理? A: 后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

Q: 运行时报错"非法指令"或"illegal instruction"A: 可能是 CPU 缺少 avx 指令集支持,请在paddle宫网 下载 noavx 版本覆盖安装

Q: NVIDIA GPU预测时,报错显存不足: A: 如以下错误字样:

paddle.fluid.core.EnforceNotMet: Enforce failed. Expected allocating <= available, but received allocating:20998686233 > available:19587333888.

Insufficient GPU memory to allocation. at [/paddle/paddle/fluid/platform/gpu_info.cc:170]

请在运行 Python 前设置环境变量,通过export FLAGS_fraction_of_gpu_memory_to_use=0.3来限制SDK初始使用的显存量,0.3表示初始使用 30%的显存。如果设置的初始显存较小,SDK 会自动尝试 allocate 更多的显存。

Q:我想使用多线程预测,怎么做?如果需要多线程预测,可以每个线程启动一个Progam实例,进行预测。 demo.py文件中有相关示例代码。

注意: 对于CPU预测,SDK内部是可以使用多线程,最大化硬件利用率。参考init的thread_num参数。

Q: 运行SDK报错 Authorization failed

情况一:日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888" ./easyedge_demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- MAC地址变化
- 磁盘变更
- BIOS重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 ~/.baidu/easyedge 目录,再重新激活。

情况三: Atlas Python SDK日志提示ImportError: libavformat.so.58: cannot open shared object file: No such file or directory 或者其他类似so找不到可以在LD_LIBRARY_PATH环境变量加上libs和thirdpartylibs路径,例如

export LD_LIBRARY_PATH=/xxx/libs:/xxx/thirdpartylibs:\$LD_LIBRARY_PATH # tips: 这里/xxx需要替换为真实路径,/xxx路径查找方法如下

查找安装包内libs和thirdpartylibs路径的方法如下(以华为Atlas300 SDK为例,其他SDK查找方法类似):

pip3 show EasyEdge-Devkit-Atlas300 # 结果中会显示 Location 路径,也就是包的安装路径 **libs和thirdpartylibs两个路径在 Location 所指示的路径 easyedge_CANN 子文件夹下**

ல Linux集成文档-Atlas

简介

本文档介绍EasyEdge/EasyDL的Linux Atlas SDK的使用方法。

注意Atlas有两种产品形态,Atlas 200和Atlas 300 ,请参见此处的文档说明

- 网络类型支持:图像分类
- 硬件支持:
 - CPU: aarch64
 - Atlas 300 卡
- 操作系统支持:Atlas指定的Linux版本, Ubuntu 16.04 x86_64 或 centos 7 x86_64 ,请从Atlas文档中下载。

Release Notes

时间	版本	说明
2020.3.23	0.1	初始版本,支持图像分类

性能数据

数据仅供参考,实际数值根据使用线程数、利用率等情况可能有所波动

模型类型	模型算法	芯片类型	SDK类型	实测硬件	单次预测耗时
EasyDL 图像分类	高性能	Atlas 300	Atlas 300	Atlas 800服务器	9ms
EasyDL 图像分类	高精度	Atlas 300	Atlas 300	Atlas 800服务器	12ms
EasyDL 物体检测	高性能	Atlas 300	Atlas 300	Atlas 800服务器	11ms
EasyDL 物体检测	高精度	Atlas 300	Atlas 300	Atlas 800服务器	31ms

atlas 300 加速卡注意事项

一般服务器(HOST侧)安装多个300加速卡,,每个300加速卡有4个芯片。一个芯片(DEVICE侧)可以认为是一个单独的系统,并且不共享储存 系统。

每个芯片都有独立的device-id,可以通过命令查看: sudo npu-smi info

由于模型需要在芯片上运行。因此运行SDK前,需要手动将模型复制到每个单独芯片的储存系统上。

测试atlas 300的官方demo

环境准备

请参见此处的文档说明, 搭建环境,测试HelloDavinci demo通过后,再测试本demo

修改300加速卡SSH密码 (可选)

请在咨询华为技术人员后,修改Device登录密码

ssh HwHiAiUser@192.168.1.199 **登录后会强制修改密码** ssh HwHiAiUser@192.168.1.198

快速开始

SDK在以下环境中测试通过

```
• ubuntu 16.04, Atlas 800 服务器指定版本;
```

```
Atlas DDK 的ddk_info信息:
```

```
{
    "VERSION": "1.3.8.B902",
    "NAME": "DDK",
    "TARGET": "ASIC"
}
```

1.安装软件

sudo apt-get install sshpass build-essential

2. 测试Demo

编译运行:

下载后,模型资源文件默认已经打包在开发者下载的SDK包中,

- Step 0:使用HwHiAiUser登录
- Step 1: 运行一次install-demo.sh脚本,会得到测试demo。
- Step 2:请在官网获取序列号,填写在demo_async.cpp及demo_sync.cpp的开始处license_key字段。



```
step3: 准备测试图片
```

覆盖image目录下的 1.jpg,更多图片可以用于demo中的批量测试模式

step4(可选): 修改test_300.sh下的以下开发板登录信息

export DDK_PATH=\$HOME/tools/che/ddk/ddk # ddk的安装路径

declare -a DIVICE_IPS=("192.168.1.199") # 300加速卡芯片的ip地址,device=0 对应192.168.1.199 DEVICE_PASSWORD="Huawei@SYS3" # 之前 修改300加速卡SSH密码 MAIN_CPP="demo_async.cpp" # demo_async.cpp" 异步接口, "demo_async.cpp" 同步接口

```
OpenCV_install_dir=/home/HwHiAiUser/opencv_x64/# OpenCV 3.4版本,需要存在

${OpenCV_install_dir}/share/OpenCV/OpenCVConfig.cmake文件
```

step5: 运行demo,会自动编译OpenCV 3.4库, 如果报错请自行编译,目录设置在 OpenCV_install_dir

cd demo sh test_300.sh

图像分类demo运行效果:

[stat] [100001]image/1.jpg(4 images) time used: 41ms (at 1583765958531) total:705ms [result][100001][image/1.jpg][281470472005664] is: n07747607 orange 0.973633 950;

n07747607 orange 分类名 0.973633 分类概率 950 分类名的序号

物体检测的demo运行效果:

[stat] time used : 101ms; all time used:478 images[3] result: label:no2_ynen;prob:0.985352 loc:[(0.459961,0.839844), (0.5625,0.988281)]

no2_ynen 分类名 , 也可以获取分类名的序号 0.985352 分类概率 loc:[(0.459961,0.839844), (0.5625,0.988281)] , 检测框的位置。 (0.459961,0.839844)表示左上角的点,(0.5625,0.988281)右下角的 点; 如原始图片608, 左上角(0.459961*608,0.839844*608) , 右下角(0.5625*608,0.988281*608)

SDK接口使用

使用该方式,将运行库嵌入到开发者的程序当中。

同步接口使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

```
// step 0: 设置序列号
global_controller()->set_licence_key("set your license here");
```

// step 1: 配置模型资源目录 AtlasConfig config; config.model_dir = {模型文件目录};

// step 2: 创建并初始化Predictor ;

auto predictor = global_controller()->CreateEdgePredictor(config); int ret = predictor->init(); # 若返回非0,请查看输出日志排查错误原因。 auto img = cv::imread({图片路径}); // step 3: 预测图像 std::vector<EdgeResultData> result2; predictor->infer(img, result2); # 解析result2即可获取结果

异步接口使用流程

// step 0: 设置序列号 global_controller()->set_licence_key("set your license here");

// step 1: 配置模型资源目录 AtlasConfig config; config.model_dir = {模型文件目录};

// step 3: 创建Predictor; 这这里选择合适的引擎 auto predictor = global_controller()->CreateEdgePredictor(config);

// step 4: 设置异步回调 predictor->set_result_handler(YOUR_HANDLER);

// step 5: 初始化
int ret = predictor->init();
若返回非0,请查看输出日志排查错误原因。

// step 6: 预测图像
auto img = cv::imread({图片路径});
color_format = kBGR;
float threshold = 0.1;

uint64_t seq_id; predictor->infer_async(img, color_format, 0.1, nullptr, seq_id); **YOUR_HANDLER里面有seq_id的回调结果**

设置序列号

请在网页控制台中申请序列号,并在init初始化前设置。 LinuxSDK 首次使用需联网授权。

接口

```
virtual int set_licence_key(const std::string& license) = 0;
```

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

EdgeLogConfig log_config; log_config enable_debug = true; global_controller()->set_log_config(log_config);

日志及报错

日志

日志需要开启Atlas 的 INFO级别,/etc/slog.conf中配置关闭zip格式。清空/var/dlog 目录,运行atlas 300 官方示例代码,可以在/var/dlog目录 下看见host和device开头的2个日志文件,中间是明文的info级别的日志

日志共有3处:

- host 测的easyedge.log。当前运行目录下。
- device侧的easyedge.device.xxx.log。 device侧的日志, 在芯片的同名目录下。
- /var/dlog host 与device开头的log文件,ddk运行日志,其中device侧有略微延时

通用错误码

错误码	常量	解释
100000 4	RESOURCE_LOAD_F AIL	缺少data/model/conf.json文件或者该文件以及被改动。下载包中的data/model下的所有文件都不要改动,尝试 使用默认配置。或者按照报错复制到对应目录。
700000 1	AUTH_FAILED	服务端校验序列号失败
700000 2	AUTH_LICENSE_INV ALID	校验序列号
700000 3	AUTH_LICENSE_EXP IRED	序列号过期
500001	NET_CURL_PERFOR M_FAILED	服务端校验序列号的请求因为网络原因失败
600001	GET_MACHINE_ID_F AILED	没有相关权限,请反馈

Atlas SDK 错误码

错误码 值	常量	含义	报错示例信息	示例解释及解决方式
120000 11	FILE_NOT _READABL E	资源文件不可读	data/model/params IS NOT READABLE	data/model/params,这个文件不可读。 SDK下载包中的data/model下的所有文件 都不要改动,尝试使用默认配置。或者按 照报错复制到对应目录。
120000 12	HIAI_ERRO RLIST _FILE	status.h.list不是原始文件	data/model/status.h.list IS TOO SMALL	下载包中的data/model下的所有文件都不 要改动,包括status.h.list
120001 02	PREDICTO R_NOT_INI TED	create后没有调用init()函数	please call init() first	调用infer函数前没有调用init()
120001 03	PREDICTO R_NO_HAN DLER	create后没有调用 set_result_handler()函数	please call set_result_handler() first	调用infer_async函数前没有调用 set_result_handler(),建议init前调用
120001 04	PREDICTO R_ALREAD Y_INITED	init()不管是否成功,不能连续调 用。	don't call init() more than once	如果失败,请再次新建一个Predictor
120001 05	BATCH_SIZ E	AltasConfig里的batch_size设置与 model_name不符合	model batch size is 1; your config batch size is 4	batch_size设置里4,model_name设置里 params,不对应导致报错。model_name 应该设置为params-batch4
120001 06	INPUT_WID TH	preprocess_args.json被改动	model input tensor width is 224; your config resize is 226	请勿修改preprocess_args.json
120001 07	INPUT_HEI GHT	同上	同上	同上
120002 01	BATCH_TO O_MANY_I MAGES	一次输入的图片大于batch_size	too_many_images input:2; batch_size is 1	调用infer函数,输入了2张图片,大于 batch数。如果batch=1的话,每次infer只 能传一张图。
120002 02	IMAGE_FO RMAT_CHA NNELS	infer函数输入的color_format与 cv::Mat里的channel数不匹配	EdgeColorFormat is not according to cv channels; format is 101; channels is 3; seq_id1	101表示kRGBA,cv::Mat里channel应该期 望是4。如果是直接读的图片,填kBGR。
122000 01	ENGINE_M ATRIX _COMMON	Atlas DDK Matrix部分(非 CreateGraph函数)接口报错。即 返回值HIAI_StatusT不是 HIAI_OK。具体解释见Atlas官方文 档。	hiai::Graph::ParseConfigFile (graph.prototxt) ; status Code is 16855066 ; HIAI ERROR CODE is 101 HIAI_GRAPH_PROTO_FILE _PARSE_FAILED_CODE,	调用hiai::Graph:: ParseConfigFile()返回 16855066 ,对应的status.h.list中的错误 码是101。保留日志,具体见Atlas官方文 档。
122000 02	ENGINE_AI _COMMON	Atlas DDK Device引擎部分 hiai::AlStatus 不为hiai::SUCCESS	_ai_model_manager->Process()	保留日志,具体见Atlas官方文档。
122000 03	ENGINE_M ATRIX_INIT	Atlas DDK CreateGraph() 初始化 DDK报错。具体解释见Atlas官方 文档。	hiai::Graph::CreateGraph(); data/model/graph.prototxt ; status Code is 16855190 ; HIAI ERROR CODE is 225 HIAI_FILE_NOT_EXIST_CODE,	示例为缺少libatlas_device.so导致
122000 04	EDGE <i>ATLA</i> <i>S_ENGINE</i> MATRIX_INI T_DEVICE	Atlas DDK CreateGraph() 初始化 DDK报错,这个报错很可能是 device侧出现问题	hiai::Graph::CreateGraph() ; data/model/graph.prototxt ; status Code is 16855057 ; HIAI ERROR CODE is 92 HIAI_GRAPH_ENGINE _INIT_FAILED_CODE	需要具体排查DEVICE侧日志再次找具体报 错,发现原因
122000 05	ENGINE_AR GS_NULL	内部错误		请反馈
123000 01	SYNC_INFE R_TIMEOU T	调用infer同步接口时,内部会调用 infer_async函数,这个函数超时	infer sync wait timout more than 10ms	内部会调用infer_async函数超过10ms。1. 不要并发过高 2. 超时参数略微大些。

心图像分类服务器端SDK集成文档-EdgeKitProxy

简介

本文档介绍EdgeKitProxy的使用方法。

Release Notes | 时间 | 版本 | 说明 | | ------- | ------ | ------ | | 2023-05-17 | 1.0.0 | 第一版 ! |

快速开始

二进制位置

位于SDK内bin目录中,文件名为edgekit_serving, 配套edgekit_serving.yml为默认配置文件

注意事项

请参考各SDK文档中的注意事项

使用说明

服务启动

```
usage: edgekit_serving [<flags>]
Flags:
                 显示帮助
  --help
-c, --cfg=./edgekit_serving.yml
               配置文件
-m, --model_dir=./RES 模型目录
 -s, --serial_num=ABCD-EFGH-IJKL-MNOP
               序列号
  --pool_min_size=1 预测池最小预测器个数
  --pool_max_size=1 预测池最大预测器个数
   --pool_full_interval_seconds=-1
                预测池满载多少秒进行扩容
   --pool_idle_interval_seconds=-1
                预测池未满载多少秒进行缩容
   --pool_available_device=-1 ...
              预测池可用设备列表
  d, --debug 开启debug模式
--log_to_std 日志输出至终端
--log_to_file 日志输出至文件
 -d, --debug
  --log_file=easyedge.log 日志文件名
  --log_max_size=10 日志最大大小(MB)
--log_max_age=10 日志旧文件保留天数
   --log_max_backups=100 日志旧文件保留个数
-h, --host=127.0.0.1 服务监听地址
 -p, --port=24401
                   服务监听端口
   --ws_max_handle_num=1 websocket接口最大处理请求个数
   --ws_max_handle_timeout=30
                websocket接口超时时间
```

配置文件说明

```
controller:
 serialNum: AAAA-AAAA-AAAA # 序列号
 modelDir: ../../.RES # 模型目录
predictorPool:

        minSize: 1
        # 预测池最小预测器个数

        maxSize: 3
        # 预测池最大预测器个数

 fullIntervalSeconds: 1 # 预测池满载多少秒进行扩容
 idleIntervalSeconds: 1 # 预测池未满载多少秒进行缩容
 availableDevice: [-1] # 预测池可用设备列表
serving:

        host: 0.0.0.0
        # 服务监听地址

        port: 24401
        # 服务监听端口

 enableHTTP: true # 对外开启HTTP服务
 enableWS: false # 对外开启websocket服务
 WS:
 maxHandleNum: 1 # websocket接口最大处理请求个数
maxHandleTimeout: 30 # websocket接口超时时间
logging:

    debug: true
    # 开启debug模式

    logToStd: true
    # 日志输出至终端

    logToFile: false
    # 日志输出至文件

 logFile: easyedge.log # 日志文件名

        maxSize: 10
        # 日志最大大小 (MB)

        maxAge: 10
        # 日志旧文件保留天数
```

命令行参数会覆盖配置文件中同义配置

maxBackups: 100 # 日志旧文件保留个数

服务调用

HTTP服务接口url: \${监听地址}/ HTTP服务接口url: \${监听地址}/ws

请求参数

syntax = "proto3";

package easyedge.kit.proxy;

```
enum ImageType {
Bin = 0; // 图片原始二进制内容, json格式下为base64编码后结果
 Mat = 1; // 图片Mat格式内容, json格式下为base64编码后结果
message HTTPRequest {
bytes image = 1;
ImageType image_type = 2;
int32 height = 3;
int32 width
              = 4;
int32 channel = 5;
float threshold = 6;
int32 top_num = 7;
enum CommandType {
GetInfo = 0;
InferImage = 1;
enum InfoType {
Hardware = 0;
message WebSocketRequest {
string request_id = 1;
CommandType command_type = 2;
InfoType info_type = 3;
bytes image = 4;
ImageType image_type = 5;
int32 height = 6;
                = 7;
int32 width
int32 channel = 8;
int64 frame_id = 9;
float threshold = 10;
int32 top_num = 11;
```

返回参数

```
syntax = "proto3";
package easyedge.kit.proxy;
```

```
message BasicGPUInfo {
string productName = 1;
 string memUsed = 2;
 string memTotal = 3;
 string gpuUtil = 4;
 string powerLimit = 5;
 string powerDraw = 6;
 string temperature = 7;
message DevStat {
string name = 1;
 uint64 rx = 2;
 uint64 tx = 3;
message Chip {
string name = 1;
 double powerUsed = 2;
 double powerLimit = 3;
```

dauble temperature

1

```
double temperature = 4;
double chipUtil = 5;
int64 memoryUsed = 6;
int64 memoryTotal = 7;
message SMI {
string name
                = 1
string sdkVersion = 2;
string driverVersion = 3;
repeated Chip chips = 4;
message HInfo {
string osName
                           = 1:
                          = 2;
 string hostname
repeated string ipAddr
                           = 3;
 repeated string macAddr
                             = 4;
 uint64 bootTime
                             = 5
                           = 6;
 int32 cpuCores
double cpuMhz= 7;string cpuModelName= 8;double cpuUsage= 9;
 map<string, double> cpuUsageDetail = 10;
 uint64 memTotal = 11;
uint64 memTotalUsed= 12;double memUsage= 13;
 map<string, double> memUsageDetail = 14;
 uint64 diskTotal = 15;
uint64 diskTotalUsed= 16;double diskUsage= 17;
 map<string, double> diskUsageDetail = 18;
 string userName = 19;
 bool isInternetConnected
                             = 20;
string deviceId= 21;int64 deviceTimestamp= 22;
 map<string, DevStat> netUsageDetails = 23;
 repeated BasicGPUInfo gpuInfo = 24;
                = 25;
Fotal ____
 double gpuUtil
                            = 26;
 uint64 gpuMemTotal
 uint64 gpuMemTotalUsed
                             = 27;
 double gpuMemUsage
                                = 28:
 map<string, SMI> aiChipInfo
                                = 29:
message LocationPoint {
optional int32 x = 1;
optional int32 y = 2;
message Location {
optional int32 left
                       = 1;
optional int32 top
                       = 2;
optional int32 width
                        = 3,
 optional int32 height
                        = 4;
 repeated LocationPoint points = 5;
message Point {
optional double x = 1;
optional double y = 2;
message InferResultItem {
optional int64 index = 1;
 optional double confidence = 2;
 optional double score = 3;
 optional string label = 4;
 optional string name = 5;
 entional int20 madall/ind
```

optional m_{132} modelning = o;

// 矩形检测	
optional double x1	= 7;
optional double x2	= 8;
optional double y1	= 9;
optional double y2	= 10;
optional Location locati	ion = 11;
// 四边形检测	
repeated Point points	= 12;
// 追踪	
optional int64 trackId	= 13;
optional int64 frame	= 14;
optional double fps	= 15;
optional string mask	= 16;
}	
message HTTPResponse	9 {
int64 cost_ms	= 1;
int32 error code	= 2;
int64 frame_id	= 3;
repeated InferResultIte	m results = 4:
}	
message WebSocketInfe	erResponse {
string request id	= 1:
int64 cost ms	= 2:
int32 error code	= 3:
int64 frame_id	= 4
repeated InferResultIte	m results = 5:
bytes annotated	= 6: // 渲染后的图片原始 ^一 进制内容,ison格式下为base64编码后结果,目前语义分割返回这个举型
}	
J	
message WebSocketHIn	foResponse {
string request id	= 1;
int32 status =	2:
string msg =	3:
Hinfo data =	4:
}	.,

其他说明

单机负载均衡

通过配置文件或命令行参数配置了预测池相关配置后,若预测池最小与最大预测器个数不同,且扩缩容配置不为-1则开启单机负载均衡,服务启 动时会创建最小数量的预测器,后续根据实际请求情况,若所有预测器均有负载的持续时间大于配置中的满载扩容时间,且预测器数量未到达最 大个数时,会自动扩容,后续若请求并发数下降,预测器池中预测器不能跑满负载时,则会自动缩容,尽可能最大化利用单机资源

⁰ 纯离线API集成说明

本文档主要说明定制化物体检测模型发布为服务器API(通过部署包实现)后如何使用。如还未训练模型,请先前往EasyDL进行训练。

如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

部署包使用说明

部署方法

EasyDL定制化物体检测模型的服务器API通过EasyPack实现,目前提供单机一键部署的方式。

在EasyDL申请、下载部署包后,在本地服务器新建目录(建议目录命名规则:easyDL_服务名称_模型版本号),将软件包上传至该目录。请参 考EasyPack-单机一键部署使用python2版本来部署,部署成功后,启动服务,即可调用与在线API功能类似的接口。

运维检查

EasyDL服务器API部署应用健康检查(或故障排查)脚本:trouble_shooting.tar

脚本能力:鉴权服务健康检测、容器状态检查、端口探活、网络联通性测试、容器关键报错日志输出等

使用方法:将脚本上传至服务器任意目录(或在服务器直接下载),并解压后运行。

```
**解压**
tar vxf trouble_shooting.tar
**执行**
bash trouble_shooting.sh
```

授权说明

API部署包根据服务器硬件(CPU单机或GPU单卡)进行授权,只能在申请时提交的硬件指纹所属的硬件上使用。

部署包测试期为1个月,如需购买永久授权,可提交工单咨询

性能指标

物体检测模型可部署在CPU或GPU服务器上,单实例具体性能指标参见算法性能及适配硬件 API参考

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在EasyDL进行自定义模型训练,完成训练后申请部署包,部署成功后拼接url。

请求URL: http://{IP}:{PORT}/{DEPLOY_NAME}/ObjectDetection IP:服务部署所在机器的ip地址 PORT:服务部署后获取的端口 DEPLOY_NAME: 申请时填写的服务名称

Header如下:

```
参数 值
Content-Type application/json
```

注意:如果出现336001的错误码很可能是因为请求方式错误,与其他图像识别服务不同的是定制化图像识别服务以json方式请求。

Body请求示例:

```
{
"image": "<base64数据>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大4096px,支持 jpg/png/bmp格式 注意请去掉头部
thresho Id	否	number	-	默认值为0.3,请在我的模型列表-模型效果查看推荐阈值

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	识别结果数组
+name	否	string	分类名称
+score	否	number	置信度
+location	否		
++left	否	number	检测到的目标主体区域到图片左边界的距离
++top	否	number	检测到的目标主体区域到图片上边界的距离
++width	否	number	检测到的目标主体区域的宽度
++height	否	number	检测到的目标主体区域的高度

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如缺少必要出入参时返回:

{ "error_code": 336001, "error_msg": "Invalid Argument"

}

错误码	错误信息	描述
336000	Internal error	服务器内部错误,请再次请求 , 如果持续出现此类错误,请在百度智能云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有疑 问请在百度智能云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数代码格式是否有误。有疑问请在百度智能云控制台内提交工单反馈
336003	Base64解码失败	图片/音频/文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。 有疑问请在百度智能云控制台内提交工单反馈
336004	输入文件大小不合法	图片超出大小限制,图片限4M以内,请根据接口文档检查入参格式,有疑问请在百度智能云控制台内提交工 单反馈
336005	图片解码失败	图片编码错误(非jpg,bmp,png等常见图片格式),请检查并修改图片格式
336006	缺失必要参数	image字段缺失(未上传图片)
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请请在百度智能云控制台 内提交工单反馈
337000	Auth check failed	离线鉴权调用失败

模型更新/回滚操作说明

模型更新

1、在EasyDL-纯离线服务发布页面,找到您的服务器API发布记录,点击【更新版本】,选择「更新包」或「完整包」来发布。

两者区别:

包类型	描述
更新包	仅包含最新的模型应用,需执行download.sh脚本下载所需镜像等依赖文件
完整包	包含模型应用和其他鉴权服务,需执行download.sh脚本下载所需完整依赖文件

2、(CPU模型可忽略)如果您训练的模型为GPU版本,系统会生成多份下载链接。请在GPU服务器执行 nvidia-smi命令,根据返回的Cuda Version来选择对应的部署包链接下载。
3、在服务器新建目录(建议标记对应模型的版本号,便于区分不同模型版本),如easydl_\${DEPLOY_NAME}_v2

\${DEPLOY_NAME}:申请时填写的服务名称

以如下场景举例说明:模型版本从V1升级至V2

```
**1.新模型准备**
**创建指定版本的目录**
mkdir easedl_${DEPLOY_NAME}_v2
cd easedI_${DEPLOY_NAME}_v2
**将部署包上传至服务器该目录并解压**
tar zvxf xx.tar.gz
**解压后,进入指定目录执行download脚本下载模型所依赖文件**
cd original && bash download.sh
**2.旧模型备份**
**历史模型备份**
cp -r /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V1
**记录当前模型的端口号**
docker ps -a |grep ${DEPLOY_NAME}
**3.模型升级**
cd package/Install
**卸载当前已安装的旧的easyDL服务: ${DEPLOY_NAME},前面已备份**
```

```
python2 install.py remove ${DEPLOY_NAME}
**安装当前部署包内新的EasyDL服务: ${DEPLOY_NAME}**
python2 install.py install ${DEPLOY_NAME}
```

(可选操作) 更新证书 python2 install.py lu

模型回滚

以如下场景举例说明:模型版本从V2回滚至V1

方法一:

```
***重命名当前v2模型目录名称**
mv /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V2
***使用V1版本**
cp -r /home/baidu/work/${DEPLOY_NAME}_V1 /home/baidu/work/${DEPLOY_NAME}
***停止当前模型容器**
docker ps -a |grep ${DEPLOY_NAME}
docker rm -f ${容器名}
***创建新的容器**
cd /home/baidu/work/${DEPLOY_NAME} && bash start/start-1.sh
*** (可选操作)进入V1版本部署包所在目录执行license更新操作,假如部署包在/opt目录下,以您实际目录为准**
```

cd /opt/easydl_\${DEPLOY_NAME}_V1/original/package/Install && python2 install.py lu

方法二:

进入模型V1所在目录,参考上述【模型更新】步骤,执行模型升级操作(即先卸载v2,后升级为v1)

心 端云协同服务说明

服务简介

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

「云管理,端计算」的端云协同服务,具体包括:

- 在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新
- 断网状态下模型离线计算 (http服务,可调用与公有云API功能相同的接口)

• 联网状态下在平台管理设备运行状态、资源利用率

目前本地服务器的应用平台支持Linux-AMD64(x86-64),具体使用流程请参考下方文档。

使用流程

Step 1 发布端云协同部署包

在我的部署包页面点击「发布端云协同部署包」

模型中心	端云协同服务 > 我的部署包							
我的模型	端云协同服务说明							点击收
创建模型	1、在可视化界面轻松实现模型部	書包在边缘设备.	上的集成、版本更	新				
训练模型	 2、断网状态下模型离线计算(ht 3、联网状态下在平台管理设备运 	tp服务,可调用 行状态、资源利	与公有云API功能村 用率	目同的接口)				
校验模型	具体使用流程如下:							
发布模型	 在本页面发布端云协同部署包 在「我的本地设备」页面新 	填写相关信息; 普设备、复制激;	^{并申请,并在下方} 5命令,联网激活	列表查看部署包发有 本地设备	币状态			
EasyData数据服务	3、在「下发部署包到设备」页[面,将发布成功的	的部署包一键下发	到已激活的设备上,	即可测试或正式使用			
数据总览	发布端云协同部署包							
标签组管理								
在线标注	端云协同服务名称	模型ID	设备类型	应用平台	当前可用版本	新版本发布状态	部署设备数	操作
云服务数据回流								
EasyEdge本地部署								
纯离线服务								
端云协同服务 へ								
我的部署包					暂无可用数	化据		
我的本地设备					请稍后再试			
下发部署句到设备								

填写服务名称,选择模型版本并提交发布

模型中心	端云协同服务 > 发	布端云协同部署包	
我的模型	设备类型	● 服务器 ○ 通用小型设备	
创建模型	描刊々わ	2001-189-140-01-1	
训练模型	快坐古你	况为外经物地们的普刀口	v
校验模型	端云协同服务名称		
发布模型	选择版本	V1	\sim
EasyData数据服务	选择系统和芯片	Δ Linux	
数据总览		通用X86 CPU	
标签组管理			
在线标注	发布部署包		
云服务数据回流			

在列表查看部署包发布状态

模型中心	端云协同服务 > 我的部署包
我的模型	端云协问服务说明
创建模型	1、在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新
训练模型	2、前周状态;操型黑线灯算(Iftitp提供,可调用与这指式APH加能相同的接口) 3、联局状态下在平台管理设备运行状态、资源利用率
校验模型	具体使用流程如下:
发布模型	 在本页面发布能否协同部署包,填写相关信息并申请,并在下方列表置看部署包发布状态 在「我的本地设备」页面新增设备,复制激活命令,联网激活本地设备
EasyData数据服务	3、在「下发部署包到设备」页面,将发布成功的部署包一键下发到已激活的设备上,即可测试或正式使用
数据总览	发布端五阶间影景台
标签组管理	
在线标注	端云协同服务名称 模型ID 设备类型 应用平台 当前可用版本 新版本发布状态 部署设备数 操作
云服条数据回流	ecc 264 通用小型设备 Linux-通用ARM V1 V1-已发布 0 下发到设备 发布新版本
	从发布页过来的 246 服务器 Linux-通用X86 CPU V2, V1 V2-已发布 1 下发到设备 发布新版本 服务详情
EasyEdge本地部署	图像分类高精度 猫狗-265 265 通用小型设备 Linux-通用ARM - V1-发布中 0
纯离线服务	
端云协同服务 ヘ	每页显示 10 🗸 < 1 >
我的部署包	
我的本地设备	

Step 2 新增设备并激活

在我的本地设备页面新增设备

模型中心	端云协同服务 > 我的	内本地设备					
我的模型	我的本地设备						点击收起
创建模型	在本页面新增设备、联M 利用率等信息。	刚激活本地设备后,即可?	将「 我的部署包 」页面	发布成功的部署包一键一	下发到设备上。设备联网时,可以	查看设备上部署的服务、设备的	运行状态、资源
训练模型							
校验模型	新增设备						
发布模型	设备实际	创建时间	设备类型	应用平台	设备连接状态	最新同步时间	操作
EasyData数据服务	KA 88 1419	D3YE*31-3	MBXE	20012	K BALLSONG	NUMBER OF STREET	JATE
数据总览							
标签组管理							

	新增设备	×
2 [设备名称	20个字符以内	
备注信息	50个字符以内, 自定义备注信息	
5		
	0/50	
	确认 取消	

在列表中,点击设备对应的「激活设备」操作,复制激活命令并在本地设备上执行即可

Re the second se	查看激活命令	×	收起 ^
在	复制命令,并在本地设备上执行,即可激活设备		
端云协同服务 我的本地设备 在本页面新增设 利用率等信息。	curl -skfL 'https://verify.baidubce.com/active/vl/active/setu sh?token=177df4ed47b2265223a333630302c226b233226eff6465222c2 6e223a22353730313835302d382d65393866363302d343231612d653765 12d373138392d376266663366643063313433222c226e73223a2262616574 6c2d636c6f7564222c227473223a313630393939393434327d' -osetup.sh && sh setup.sh	22 33 79 1	点击收起 可以宣着设备上部署的服务、设备的运行状态、资源
设备名称	知道了		1间 操作
II/IUX X86 G	2021-01-07 14-02 - 不適/治	1	翻ALAUCUU 14年7月 每页显示 10 ✔ < 1 >

Step 3 下发部署包到设备,在本地调用

在下发部署包到设备页面,将发布成功的部署包一键下发到已激活的设备上,即可测试或正式使用

模型中心	端云协同服	务 > 下发部署包到设备	
我的模型	将已成功	发布的端云协共服务部署包一键下发到已激活且在连接中的设备	
创建模型			
训练模型	服务名称	and address	\checkmark
校验模型	模型ID	246	
发布模型	应用平台	Linux-通用X86 CPU	\checkmark
EasyData数据服务	选择版本	V2	\checkmark
数据总览	(生物)人名		
标签组管理	近伴议面	ubuntu-Io ×	
在线标注	服务配置	● 基础设置	
云服务数据回流		服务端口 8080	
EasyEdge本地部署			
纯离线服务	下发服务		
端云协同服务 へ			

部署包下发成功之后,会在本地启动一个HTTP推理服务。在浏览器中输入http://{设备ip}:{服务端口,默认8080},即可预览效果:

【物体检测】S		调整阈值 🔷	当前阈值: 0.07 修改
		label	置信度
		tomato	0.998
		tomato	0.937
III. D		tomato	0.841
		tomato	0.781
		kiwi	0.408
		tomato	0.324
	The second secon	tomato	0.251
		tomato	0.146

具体接口调用说明请参考文档 SDK - HTTP服务调用说明

模型部署包管理

在我的部署包页面可以进行已发布的模型部署包的管理。

发布及更新模型版本

点击「发布新版本」操作即可快速发布对应模型ID下的新版本。同一模型ID下已发布的模型版本均会显示在列表的「当前可用版本」中。

模型中心		端云协同服务 > 我的部署包	1								
我的模型		端云协同服务说明									点击收起
创建模型		1、在可视化界面轻松实现模型部	『署包在边线	缘设备上的集成、 可调用与公室三。	版本更新						
训练模型		 2、町网仏芯下侯聖高线计算(n) 3、联网状态下在平台管理设备返 	a行状态、	可调用与公有云A 资源利用率	1413/月8代日日15分女山)						
校验模型		具体使用流程如下:									
发布模型		 1、在本页面发布端云协同部署包 2、在「我的本地设备」页面新 	2, 填写相注 i增设备, \$	关信息并申请,主 夏制激活命令,即	‡在下方列表宣看部署包》 《网激活本地设备	发布状态					
EasyData数	文据服务	3、在「 下发部署包到设备 」页	〔面,将发行	^{午成功的部署包一}	-键下发到已激活的设备」	E,即可测试或正	式使用				
数据总览		发布端云协同部署包									
标签组管理											
在线标注		端云协同服务名称	模型ID	设备类型	应用半台	当前可用版本	新版本发布状态	部署设备数	操作	a utari	
云服务数据回]流	ecc	264	通用小型设备	Linux-通用ARM	V1	V1-已友布	0	下发到设	备 反布新版本	AV 141.10
EasyEdge4	本地部署		246	服労益	Linux-通用X86 CPU	V2, V1	V2-已友巾	1	下发到设	畲 友巾莉啟本 服	穷许丽
纯离线服务		and the second second	200	旭 用小望设留	LINUX-週月ARM	-	1-及中中	0			
端云协同服务									每页显示	10 🗸 <	1 >
我的部署	2										
我的本地	设备										
下发部署	包到设备										
	_								_		
			₩±	- xc luc +				,			
1			反竹	小似个	•				^		
将	最新训练的模	型版本发布为服务,	发布质	成功后,即	可从云端下发	到设备					
		服务文称	~~								
2		1005 TO 101 C									
		模型ID 2	64								
		选择新版本	V1			\sim					
1											
1		ł	确认]	取消						

新版本发布成功后,即可在「下发部署包到设备」页面或当前服务的「服务详情」页面,将新版本下发到本地设备上。

模型中心	端云协同服务 > 我的部署包									
我的模型	端云协同服务说明									点击收起
创建模型	 在可视化界面轻松实现模型部 断网状态下模型离线计算(bt) 	署包在边缘设备 to III 条 可调用	6上的集成、	版本更新						
训练模型	3、联网状态下在平台管理设备运	行状态、资源和	利用率	- 140 BE10140 13 2 CL /						
校验模型	具体使用流程如下:									
发布模型	 在本页面发布端云协同部署包, 2、在「我的本地设备」页面新¹ 	,填写相关信息 曾设备,复制激	ll并申请,并 ll活命令,联	存下方列表查看部署包发 网激活本地设备	这布状态					
EasyData数据服务	3、在「下发部署包到设备」页[面,将发布成功	的部署包一	·键下发到已激活的设备上	,即可测试或正	式使用				
数据总览	发布端云协同部署包									
标签组管理		10 Tolor 37 d	Ne A40 W10					473 (8-		
在线标注	靖云伊问服务名称	模型ID 设置	音奕型	应用半台	当刖可用版本	新酿本友伟状态	部者设备数	操作		
云服冬数据回道	ecc	264 通用	 引小型设备	Linux-通用ARM	V1	V1-已发布	0	下发到设备	发布新版本	
EAUX 25 SAUELEINE	the second second	246 服务	555	Linux-通用X86 CPU	V2, V1	V2-已发布	1	下发到设备	发布新版4	服务详情
EasyEdge本地部署	图像分类高精度_猫狗-265	265 通用	月小型设备	Linux-通用ARM	-	V1-发布中	0			
纯离线服务										

模型中心	端云协同服务 〉
我的模型	服务名称 , , , , , , , , , , , , , , , , , , ,
创建模型	设备类型 服务器 应用平台 Linux-通用X86 CPU
训练模型	下业和画名论名
校验模型	TASKYWE
发布模型	设备名称 最新下发模型版本 部署到期时间 设备连接状态 服务下发状态 最新同步时间 操作
EasyData数据服务	ubuntu-local-fjy ⊖ V2 2021-02-06 在线 V2下发成功 2021-01-07 14:45 宣看服务配置 宣看设备详续 下发新版本 移除设备
数据总览	
标签组管理	
在线标注	

管理模型已部署的设备

在上述的「服务详情」页面,可以查看并管理当前服务已部署的设备,包括移除设备、将服务下发到更多的设备等。

模型中心	端云协同服务 〉 / X							
我的模型								
创建模型	10.75177 10 10 10 10 10 10 10 10 10 10 10 10 10							
训练模型	下北利田克仍多							
校验模型								
发布模型	设备名称 最新下发模型版本 部署到期时间 设备连接状态 服务下发状态 最新同步时间 操作							
EasyData数据服务	ubuntu-local-fly 〇 V2 2021-02-06 在线 V2下发成功 2021-01-07 14:45 宣者祖务配置 宣者设备详情 下发新版本 B接设备							
数据总览								
标签组管理								
在线标注								

本地设备管理

0

在我的本地设备页面可以进行所有本地设备的管理。

查看单台设备的运行状态

点击单台设备的「服务详情」,可查看设备上运行的多个服务及设备状态:

模型中心	端云协问服务 > 我的本地设备						
我的模型	我的本地设备						点击收起
创建模型	在本页面新增设备、联网激活 利用率等信息。	本地设备后,即可将「 数	的部署包」了	页面发布成功的部署包一键下发	到设备上。设备联	网时,可以查看设备上部	署的服务、设备的运行状态、资源
训练模型							
校验模型	新增设备						
发布模型	设备名称	创建时间	设备类型	应用平台	设备连接状态	最新同步时间	操作
EasyData数据服务	123123 💬	2021-01-07 13:25	-		未激活		删除 激活设备 查看详情
数据总览	showcase-test 💮	2020-12-15 17:57	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 19:22	激活设备 查看详情
标签组管理	linux-x86-zqw-2 😳	2020-12-15 16:24	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 19:42	激活设备 查看详情
在线标注	linux-x86-zqw ⊙	2020-12-15 15:29	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-15 16:40	激活设备 宣看详情
云服务数据回流	firefly_rk3399pro 💮	2020-12-15 14:39	-	-	离线 ⑦	2020-12-15 21:33	激活设备 查看详情
EasyEdge本地部署	ubuntu-local-fjy 😳	2020-12-14 21:38	服务器	Linux-AMD64(x86-64)	在线	2021-01-07 15:01	查看详情
纯离线服务	hfl-1 💬	2020-12-14 19:24	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 14:18	激活设备 查看详情
端云协问服务 ヘ	edge新增 ☉	2020-12-11 14:52	-		未激活		删除 激活设备 查看详情

设备详情会展示当前设备的最新同步时间,以及CPU使用率、内存使用率等。服务列表则展示了当前设备上部署服务的运行情况和资源占用情况

模型中心	端云协同服务 > 我的本地设备 > ubuntu-local-fjy					
我的模型	设备详情					
创建模型	设备名称 ubuntu-local-fjy	连接状态 在	线	突时刷新	DFF	
训练模型	设备类型 服务器	应用半台 Lir	nux-AMD64(x86-64)	最新同步时间 2021	-01-07 15:00	
校验模型	CPU使用率		内存使用率			
发布模型	31.1%		35.8%			
EasyData数据服务						
数据总览	端云协同服务详情					
标签组管理	服务名称	模型ID	CPU占比	内存使用情况	内存占比	操作
在线标注	300.00	246	0.01%	156.7MB	0.93%	宣看服务配置
云服务数据回流						
EasyEdge本地部署						
纯离线服务						

心 通用小型设备部署

∞ 如何在通用小型设备部署

训练完毕后,可以选择将模型通过「SDK-纯离线服务」或「API-端云协同服务」部署,具体介绍如下:

纯离线服务部署

纯离线服务目前仅支持通过SDK集成,可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可发布设备端SDK:

- 选择模型
- 选择部署方式「EasyEdge本地部署」-「通用小型设备」
- 选择版本
- 选择集成方式
- 点击发布

模型中心	发布模型
我的模型	说明:
创建模型	选择模型 男女分类 V 1. 设备端SDK支持Android、iOS、Windows、Linux操作系统,具体的系统、硬
训练模型	部署方式 EasyEdge本地部署
校验模型	选择版本 V1
发布模型	(679617246) 咨询了解 集成方式 ④ SDK-线离线服务
EasyData数据服务	
数据总览	发布
在线标注	
云服务数据回流	
EasyEdge本地部署	
纯离线服务	

- 再根据实际使用设备选择系统与芯片
- 点击发布

模型中心	纯离线服务 > 发布新服务			
我的模型				
创建模型	部者方式			
训练模型	這擇模型 男女分类 🗸			
校验模型	邊撐版本 V1 ✓			
发布模型	选择系统和芯片 → Δ Linux			
EasyData数据服务	Uindows			
数据总览	─ 通用X86 CPU ● 英特尔◎神经计算棒			
在线标注	🕂 🖷 Android			
云服务数据回流	🕂 🗯 ios			
EasyEdge本地部署	模型加速: 🗹 同时获取加速版 🕥			
纯离线服务				
AI市场	发布			

也可以直接在「EasyEdge本地部署」-「纯离线服务」页面点击发布新服务,按上图所述进行申请发布

端云协同服务部署

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

具体使用说明请参考端云协同服务说明

の 纯离线SDK说明

心纯离线SDK简介

本文档主要说明定制化模型发布后获得的SDK如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单
- 前往官方论坛交流,与其他开发者进行互动

SDK说明

SDK支持iOS、Android、Linux、Windows四种操作系统。以下为具体的系统、硬件环境支持:

EasyDL 图像使用说明

Baidu 百度智能云文档

操作系统	系统支持	硬件环境要求
iOS	iOS 8.0 以上 (A仿生芯片版 要求11.0以上)	ARMv7 ARM64 (Starndard architectures)(暂不支持模拟器)
Android	通用ARM: Android 19以上 SNPE:Android 21以上 DDK:Android 21以上	通用ARM: 绝大部分的手机和平板 、比较耗时 SNPE:高通Soc,仅支持Qualcomm Snapdragon 450 之后发布的soc。其中 660 之后的型号可能含 有 Hexagon DSP模块,具体列表见snpe 高通骁龙引擎 DDK:CPU支持华为麒麟970N、980的arm-v8a的soc,支持的机型 mate10,mate10pro,P20,mate20等 支持armeabi-v7a arm-v8a CPU 架构, DDK仅支持 arm-v8a
Linux C++		CPU: AArch64 ARMv7I ASIC: Hisilicon NNIE1.1 on AArch64(Hi3559AV100/Hi3559CV100等) ASIC: Hisilicon NNIE1.2 on ARMv7I(Hi3519AV100/Hi3559V200等)
Linux Python		Intel Movidius MyRIAD2/MyRIAD X
Linux Ubuntu 16.04		AArch64 HUAWEI Atlas 200
Windows	64位 Windows7 及以上	Intel CPU x86_64 Intel Movidius MyRIAD2/MyRIAD X (仅支持Win10) 环境依赖: .NET Framework 4.5 Visual C++ Redistributable Packages for Visual Studio 2013 Visual C++ Redistributable Packages for Visual Studio 2015

说明

针对iOS操作系统:虽然SDK库文件很大(ipa文件很大),但最终应用在用户设备中所占用的大小会**缩小很多,如图像分类下载的ipa文件可 能会100M以上,但实际集成在设备中只有20M左右**。 这与multi architechtures、bitcode和AppStore的优化有关。

单次预测耗时参考

根据具体设备、线程数不同,数据可能有波动,请以实测为准

在算法性能及适配硬件页面查看评测信息表

自适应芯片版SDK

发布SDK时可根据实际应用时的硬件/芯片配置选择最合适的SDK。如"华为NPU版"就是针对华为NPU芯片做了适配与加速的SDK。如实际应用时 需要适配多种芯片,就可以发布"自适应芯片版"SDK,SDK被集成后会自动判断设备的芯片并运行相应的模型。

加速版SDK

发布SDK时,勾选「同时获取加速版」,就可以同时获得适配部分芯片(需选中且右侧带有加速标记)的基础版SDK和加速版SDK。



目前加速版SDK已支持Windows X86、Linux ARM、iOS ARM、Android ARM环境,加速后的SDK普遍在包大小、内存占用、识别速度等方面表现 更优,详细对比请见<u>算法性能及适配硬件</u>。

加速版SDK和基础版的测试方式类似,只需在EasyDL控制台新增「加速版」测试序列号,即可获得3个月的测试期。

激活&使用SDK

SDK的激活与使用分以下四步:

① 在【我的模型】-【服务详情】内下载SDK

我的模型	创建模型								
创建模型									
训练模型	【图像分类】 ljay-gxn	n-2flowers {	模型ID: 340					⊲ 训练	🕒 历史版本
校验模型	部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作		
发布模型 数据中心	私有服务器部署	V1	训练完成	-	-	top1准确率100.00% top5准确率100.00% 完整评估结果	申请发布		
我的数据集 云服务调用数据	【图像分类】 ljay-gjd	-2flowers 巷	莫型ID: 339					⊲ 训练	④ 历史版本
	部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作		
	私有服务器部署	V1	训练完成	-		top1准确率100.00% top5准确率100.00% 完整评估结果	申请发布		
	【图像分类】cmc-高	性能 模型ID	: 332				⊲ 训练	⊙ 历史版4	2 回删除
	部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作		
	设备端SDK	V1	训练完成	审核成功	已发布	top1准确率93.33% top5准确率100.00% 完整评估结果	服务详情申请发	ŧ	

设备端SDK下载

此处下载的SDK为未授权SDK,需要获取序列号激活后才能正式使用。SDK内附有对应版本的Demo 及开发文档,开发者可参考源代码完成开发

操作系统	操作
iOS	下载SDK 🗸 获取序列号
Android	下载SDK 🗸 获取序列号
Linux	下载SDK 🗸 获取序列号
Windows	下载SDK 🗸 获取序列号

② 在<mark>控制台</mark>获取序列号

按单台设备获得授权并使用SDK时:

✓ 返回EasyDL总览	产品服务 / EasyDL定制训练平台 - 通用设备编服务
经典版	通用设备端服务
公有云服务 > 私有服务器部署服务	按单台设备激活 按多台设备激活 按产品线激活
通用设备端服务	使用说明
专项硬件适配服务	1. 按单台设备获得模状激活SDK,操作步骤: ◎ 下载所需模型已发布的设备铺SDK → ◎ 在下方购买永久模权或新增测试序列号 → ◎ 设备锚部署SDK,联网激活后离线使用
	2. 序列号类型: 授权序列号分为基础版、加速版两种,加速版SDK (仅支持通用ARM、通用X86芯片) 需用加速版序列号激活了解详情 3. 测试序列号:每发布一个模型即可斯增2个基础版测试序列号和2个加速版测试序列号,联网激活后有效服为3个月 特别提示:如需用同一序列号激活多台设备,请前往按多台设备激活页面申请:如需开发APP,建议选择按产品线激活 使用过程中如有其他问题,可以提交工单咨询或前往官方论坛交流

Android或iOS操作系统的SDK可以选择按产品线激活(仅支持开发手机APP),序列号与包名(Package Name/Bundle ID)绑定:

< 返回EasyDL总览	产品服务 / EasyDL定制训练平台 - 通用设备端	服务			
经典版	通用设备端服务				
公有云服务 ~	按单台设备激活 按多台设备激活	按产品线激活			
通用设备端服务 专项硬件适配服务	使用说明 1. Android/IOS操作系统的SDK可按产品 ① 下载所需模型已发布的设备端SD4	品线获得授权激活,序列号与包名(P K	ID)绑定,操作步骤: 昏端部署SDK,联网激活后离线使用		
	 2.每发布一个模型即可新增2个测试序列 如需购买永久授权或使用过程中有任何问 4. 新聞序列条 	则号,联网激活后有效期为3个月 问题,可以提交工单咨询或前往1	新增序列号	to total	×
	操作系统 the formation of	包名	*包名	Android	
				 序列号将与包名绑定,请确保填写 每发布一个模型即可新增2个序列; 询 	正确 号,如需新增更多序列号,可 <mark>提交工单</mark> 咨
					确定 取消

③ 本地运行SDK,并完成首次联网激活

通过左侧导航栏查看不同操作系统SDK的开发文档

④ 正式使用

SDK常见问题

通过左侧导航栏查看不同操作系统SDK的FAQ

以下是通用FAQ,如您的问题仍未解决,请在百度云控制台内提交工单反馈

1、激活失败怎么办?

按设备激活时,激活失败可能由于以下几个原因造成:

①可能是当前序列号已被其他设备激活使用,请核实序列号后用未被激活的序列号重新激活

②序列号填写错误,请核实序列号后重新激活

③同一台设备绑定同一个序列号激活次数过多(超过50次),请更换序列号后重试

④首次激活需要联网,网络环境不佳或无网络环境,请检查网络环境后重试

⑤模型发布者和序列号所属账号非同一账号,如果存在这种异常建议更换账号获取有效序列号

⑥序列号已过有效期,请更换序列号后重试

⑦如有其他异常请在百度云控制台内提交工单反馈

按产品线激活时,激活失败可能由于以下几个原因造成:

①可能是包名填写错误,请核对与序列号绑定的包名是否与实际包名一致

②序列号填写错误,请核实序列号后重新激活

③首次激活需要联网,网络环境不佳或无网络环境,请检查网络环境后重试

④模型发布者和序列号所属账号非同一账号,如果存在这种异常建议更换账号获取有效序列号

⑤序列号已过有效期,请申请延期后重试

⑥如有其他异常请在百度云控制台内提交工单反馈

2、怎样申请序列号使用延期

序列号激活后有效期为三个月,可以在控制台进行申请,申请流程:

1) 填写申请信息

2) 等待审核:审核周期通常需要1-3个工作日左右,期间会有工作人员电话回访,请填写有效的联系方式并保证手机畅通

の Android集成文档

简介

1.1 Android SDK 硬件要求 Android 版本:支持 Android 5.0 (API 21) 及以上

硬件:支持 arm64-v8a 和 armeabi-v7a,暂不支持模拟器

通常您下载的SDK只支持固定的某一类芯片。

- 通用ARM: 支持大部分ARM 架构的手机、平板及开发板。通常选择这个引擎进行推理。
- 通用ARM GPU:支持骁龙、麒麟、联发科等带GPU的手机、平板及开发板。
- 高端芯片AI加速模块:
 - 高通骁龙引擎SNPE: 高通骁龙高端SOC,利用自带的DSP加速。其中 660 之后的型号可能含有 Hexagon DSP模块,具体列表见snpe 高通骁龙引擎官网。
 - 华为NPU引擎DDK:华为麒麟980的arm-v8a的soc。具体手机机型为mate10,mate10pro,P20,mate20,荣耀v20等。
 - **华为达芬奇NPU引擎DAVINCI**: 华为NPU的后续版本,华为麒麟810,820,990,985的arm-v8a的soc。具体手机机型为华为 mate30,p40,nova6,荣耀v30等。

通用ARM有额外的加速版,但是有一定的精度损失。

因GPU硬件限制,通用ARM GPU物体检测模型输入尺寸较大时会运行失败,可以在训练的时候将输入尺寸设为300*300。

高端芯片AI加速模块,一般情况下推理速度较快。

运行内存不能过小,一般大于demo的assets目录大小的3倍。

1.2 功能支持 | 引擎 | 图像分类| 物体检测| 图像分割| 文字识别

只支持EasyEdge | 姿态估计 | | :-: | :-: | :-: | :-: | :-: | :=: | 通用ARM | √ | √ | √ | √ | ↓ | 通用ARM GPU | √ | √ | √ | ↓ | 高通骁龙引擎SNPE | √ | √ |||| | 华为NPU引擎DDK | √ | √ |||| | 华为达芬奇NPU引擎DAVINCI | √ | √ | √ |||

1.3 Release Notes

时间	版本	说明
2023.08.31	0.10.12	新增支持实例数鉴权;SNPE引擎升级;迭代优化
2023.06.29	0.10.11	迭代优化
2023.05.17	0.10.10	横屏兼容;迭代优化
2023.03.16	0.10.9	达芬奇NPU支持更多模型及语义分割模型;各芯片支持更多语义分割模型;精简版代码补充;迭代优化
2022.12.29	0.10.8	ARM / ARM-GPU 引擎升级;迭代优化
2022.10.27	0.10.7	达芬奇NPU新增适配麒麟985;迭代优化
2022.09.15	0.10.6	SNPE引擎升级;迭代优化
2022.07.28	0.10.5	迭代优化
2022.06.30	0.10.4	支持Android11;支持EasyEdge语义分割模型;迭代优化
2022.05.18	0.10.3	ARM / ARM-GPU 引擎升级;支持更多加速版模型发布;迭代优化
2022.03.25	0.10.2	ARM / ARM-GPU 引擎升级;支持更多检测模型;迭代优化
2021.12.22	0.10.1	DDK不再支持Kirin 970;迭代优化
2021.10.20	0.10.0	更新鉴权;更新达芬奇NPU、SNPE、通用ARM及ARM-GPU引擎;新增达芬奇NPU对检测模型的支持;支持更多姿态估计 模型
2021.07.29	0.9.17	迭代优化
2021.06.29	0.9.16	迭代优化
2021.05.13	0.9.15	更新鉴权,更新通用arm及通用arm gpu引擎
2021.04.02	0.9.14	修正bug
2021.03.09	0.9.13	更新android arm的预处理加速
2020.12.18	0.9.12	通用ARM引擎升级;新增ARM GPU引擎
2020.10.29	0.9.10	迭代优化
2020.9.01	0.9.9	迭代优化
2020.8.11	0.9.8	更新ddk 达芬奇引擎
2020.7.14	0.9.7	支持arm版ocr模型,模型加载优化
2020.6.23	0.9.6	支持arm版fasterrcnn模型
2020.5.14	0.9.5	新增华为新的达芬奇架构npu的部分图像分类模型
2020.4.17	0.9.4	新增arm通用引擎量化模型支持
2020.1.17	0.9.3	新增arm通用引擎图像分割模型支持
2019.12.26	0.9.2	新增华为kirin麒麟芯片的物体检测支持
2019.12.04	0.9.1	使用paddleLite作为arm预测引擎
2019.08.30	0.9.0	支持EasyDL专业版
2019.08.30	0.8.2	支持华为麒麟980的物体检测模型
2019.08.29	0.8.1	修复相机在开发版调用奔溃的问题
2019.06.20	0.8.0	高通手机引擎优化
2019.05.24	0.7.0	升级引擎
2019.05.14	0.6.0	优化demo程序
2019.04.12	0.5.0	新增华为麒麟980支持
2019.03.29	0.4.0	引擎优化,支持sd卡模型读取
2019.02.28	0.3.0	引擎优化,性能与效果提升;
2018.11.30	0.2.0	第一版!

快速开始

2.1 安装软件及硬件准备

如果需要源码方式测试:

打开AndroidStudio , 点击 "Import Project..."。在一台较新的手机上测试。

详细步骤如下:

1. 准备一台较新的手机,如果不是通用arm版本,请参见本文的"硬件要求",确认是否符合SDK的要求

2. 安装较新版本的AndroidStudio , 下载地址

- 3. 新建一个HelloWorld项目, Android Studio会自动下载依赖, 在这台较新的手机上测试通过这个helloworld项目。注意不支持模拟器。
- 4. 解压下载的SDK。
- 5. 打开AndroidStudio,点击 "Import Project..."。即:File->New-> "Import Project...",选择解压后的目录。
- 6. 此时点击运行按钮(同第3步),手机上会有新app安装完毕,运行效果和二维码扫描的一样。
- 7. 手机上UI界面显示后,如果点击UI界面上的"开始使用"按钮,可能会报序列号错误。请参见下文修改

2.2 使用序列号激活

如果使用的是EasyEdge的开源模型,无需序列号,可以跳过本段直接测试。

建议申请包名为"com.baidu.ai.easyaimobile.demo"的序列号用于测试。

本文假设已经获取到序列号,并且这个序列号已经绑定包名。

SDK默认使用离线激活方式,即首次联网激活,后续离线使用。SDK同时支持按实例数鉴权方式,即周期性联网激活,离线后会释放所占用设备 实例。按实例数鉴权的启用参考本节2.2.3说明

2.2.1 填写序列号 打开Android Studio的项目,修改MainActivity类的开头SERIAL_NUM字段。 MainActivity 位于 app \src \main \java \com \baidu \ai \edge \demo \MainActivity.java文件内。

// 请替换为您的序列号 private static final String SERIAL_NUM = "XXXX-XXXX-XXXX-XXXX"; //这里填您的序列号

2.2.2 修改包名

如果申请的包名为"com.baidu.ai.easyaimobile.demo",这个是demo的包名,可以不用修改

打开app/build.gradle文件,修改"com.baidu.ai.easyaimobile.demo"为申请的包名

```
defaultConfig {
    applicationId "com.baidu.ai.easyaimobile.demo" // 修改为比如"com.xxx.xxx"
}
```

修改序列号和包名后,可以运行测试,效果同扫描二维码的一致

2.2.3 按实例数鉴权 设置好序列号和包名后,调用配置类的以下方法启用并配置心跳间隔时间:

```
XXXConfig config = new XXXConfig();
// 启用按实例数鉴权,配置心跳间隔,单位:秒
config.setInstanceAuthMode(10000);
```

配置类的详细说明参考后续章节【调用流程】

2.3 测试精简版

对于通用ARM、高通骁龙引擎SNPE、华为NPU引擎DDK和达芬奇NPU引擎Davinci的常见功能,项目内自带精简版,可以忽略开发板不兼容的摄像 头。

此外,由于实时摄像开启,会导致接口的耗时变大,此时也可以使用精简版测试。

目前以下硬件环境有精简版测试:

Baidu 百度智能云文档

- 通用ARM:图像分类(Classify),物体检测(Detection),文字识别(OCR),图像分割(Segmentation),姿态估计(Pose)
- 通用ARM GPU:图像分类(Classify),物体检测(Detection),图像分割(Segmentation),姿态估计(Pose)
- 高通骁龙引擎SNPE:图像分类(Classify),物体检测(Detection)
- 华为NPU引擎DDK:图像分类(Classify),物体检测(Detection)
- 华为达芬奇NPU引擎Davinci:图像分类(Classify),物体检测(Detection),图像分割(Segmentation)

具体代码分别在infertest、snpetest、ddktest和davincitest目录下。

修改方法为(以通用ARM为例):更改app/main/AndroidManifest.xml中的启动Activity。

```
<activity android:name=".infertest.MainActivity"> <!-- 原始的是".MainActivity" -->
<intent-filter>
<action android:name="android.intent.action.MAIN" />
<category android:name="android.intent.category.LAUNCHER" />
</intent-filter>
```

</activity>

开启后会自动选择图像分类(Classify),物体检测(Detection),文字识别(OCR),图像分割(Segmentation)或姿态估计(Pose)测试。

Demo APP 检测模型运行示例 精简版检测模型运行示例 Hello World! paddle2-Centernet_mbv3_small_140e < ARM Detection 4.18 person 0 Start running: 0 horse 0.40 Predict 0: {size:100, firstRe confidence:0.6314938, bo truck 0.34 181)}} per86/\$9/8 Finish running Task finished cow 0.34 识别结果 置信度 0.30 名称 置信度 序号 0.63 1 person 2 0.47 person 3 0.42 car 0.40 4 horse 5 dog 0.34 BU' 6 truck 0.34 4 ►

使用说明

3.1 代码目录结构

集成时需要"复制到自己的项目里"的目录或者文件:

2. app/src/main/assets/xxxx-xxxxx 如 app/src/main/assets/infer

+app 简单的设置,模拟用户的项目 |---+libs 实际使用时需要复制到自己的项目里 |----arm64-v8a v8a的so |----armeabi-v7a v7a的so |----easyedge-sdk.jar jar库文件 |---+src/main |---+assets |----demo demo项目的配置,实际集成不需要 |----infer 也可能是其它命名, infer表示通用arm。实际使用时可以复制到自己的项目里 |---+java/com.baidu.ai.edge/demo |---+infertest 通用Arm精简版测试,里面有SDK的集成逻辑 |--- MainActivity 通用Arm精简版启动Activity,会根据assets目录判断当前的模型类型,并运行同目录的一个Task。使用时需要修改里 面的序列号 |--- TestInferArmClassifyTask 通用Arm精简版分类 |--- TestInferArmDetectionTask 通用Arm精简版检测 |--- TestInferArmOcrTask 通用Arm精简版OCR |--- TestInferArmPoseTask 通用Arm精简版姿态 |--- TestInferArmSegmentTask 通用Arm精简版分割 |--- TestInferArmGpuClassifyTask 通用ArmGpu精简版分类 |--- TestInferArmGpuDetectionTask 通用ArmGpu精简版检测 |--- TestInferArmGpuPoseTask 通用ArmGpu精简版姿态 |--- TestInferArmGpuSegmentTask 通用ArmGpu精简版分割 |---+snpetest SNPE精简版测试 |--- MainActivity SNPE精简版启动Activity,会根据assets目录判断当前的模型类型,并运行同目录的一个Task。使用时需要修改里面 的序列号 |--- TestSnpeDspClassifyTask SNPE DSP精简版分类 |--- TestSnpeDspDetectionTask SNPE DSP精简版检测 |--- TestSnpeGpuClassifyTask SNPE Gpu精简版分类 |--- TestSnpeGpuDetectionTask SNPE Gpu精简版检测 |---+ddktest DDK精简版测试 |--- MainActivity DDK精简版启动Activity,会根据assets目录判断当前的模型类型,并运行同目录的一个Task。使用时需要修改里面的 序列号 |--- TestDDKClassifyTask DDK精简版分类 |--- TestDDKDetectionTask DDK精简版检测 |---+davincitest Davinci精简版测试 |--- MainActivity Davinci精简版启动Activity,会根据assets目录判断当前的模型类型,并运行同目录的一个Task。使用时需要修改里面 的序列号 |--- TestDavinciClassifyTask Davinci精简版分类 |--- TestDavinciDetectionTask Davinci精简版检测 |--- TestDavinciSegmentTask Davinci精简版分割 |----CameraActivity 摄像头扫描示例,里面有SDK的集成逻辑 |----MainActivity 启动Activity,使用时需要修改里面的序列号 |---- build.gradle 这里修改包名 +camera_ui UI模块,集成时可以忽略 • 1

3.2 调用流程 以通用ARM的检测模型功能为例,

代码可以参考TestInferDetectionTask

- 1. 准备配置类,如InferConfig,输入:通常为一个assets目录下的文件夹,如infer。
- 2. 初始化Manager,比如InferManager。输入:第1步的配置类和序列号
- 3. 推理图片,可以多次调用 3.1 准备图片,作为Bitmap输入 3.2 调用对应的推理方法,比如detect 3.3 解析结果,结果通常是一个List,调用结 果类的Get方法,通常能获取想要的结果
- 4. 直到长时间不再使用我们的SDK,调用Manger的destroy方法释放资源。

3.3 具体接口说明 下文的示例部分以通用ARM的检测模型功能为例

即接口为InferConfig, InferManager, InferManager.detect。

其它引擎和模型调用方法类似。

下文假设已有序列号及对应的包名

- 3.3.1. 准备配置类
- INFER:通用ARM, InferConfig
- ARM GPU : ArmGpuConfig
- SNPE:高通骁龙DSP, SnpeConfig
- SNPE GPU:高通骁龙GPU, SnpeGpuConfig
- DDK:华为NPU, DDKConfig
- DDKDAVINCI:华为达芬奇NPU,DDKDaVinciConfig

InferConfig mInferConfig = new InferConfig(getAssets(), "infer"); // assets 目录下的infer , infer表示通用arm

输入:assets下的配置 输出:具体的配置类

3.3.2. 初始化Manager类

- INFER:通用ARM, InferManager
- ARM GPU:通用ARM GPU, InferManager
- SNPE:高通骁龙DSP, SnpeManager
- SNPE GPU:高通骁龙GPU, SnpeManager
- DDK:华为NPU,DDKManager
- DDKDAVINCI:华为达芬奇NPU, DavinciManager

String SERIAL_NUM = "XXXX-XXXX-XXXX-XXXX";

```
// InferManager 为例:
new InferManager(this, config, SERIAL_NUM)); // config为上一步的InferConfig
```

注意要点

- 1. 同一个时刻只能有唯一有效的InferManager。旧的InferManager必须调用destory后,才能新建一个new InferManager()。
- 2. InferManager的任何方法,都不能在UI线程中调用。
- 3. new InferManager()及InferManager成员方法由于线程同步数据可见性问题,都**必须在一个线程中执行**。如使用android自带的**ThreadHandler** 类。

输入: 1.配置类; 2.序列号 输出: Manager类

3.3.3.推理图片

- 接口可以多次调用,但是必须在一个线程里,**不能并发**
- confidence, 置信度[0-1], 小于confidence的结果不返回。 填confidence=0, 返回所有结果
- confidence可以不填,默认用模型推荐的。

准备图片,作为Bitmap输入,

• 输入为Bitmap,其中Bitmap的options为默认。如果强制指定的话,必须使用*Bitmap.Config.ARGB_8888*

调用对应的推理方法及结果解析 见下文的各个模型方法

3.3.4 分类Classify

public interface ClassifyInterface { List<ClassificationResultModel> classify(Bitmap bitmap, float confidence) throws BaseException;

// 如InferManger 继承 ClassifyInterface

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 ClassificationResultModel 异常:一般首次出现。可以打印出异常错误码。

ClassificationResultModel

- label:分类标签,定义在label_list.txt中

- confidence:置信度,0-1

- lableIndex:标签对应的序号

3.3.5 检测Detect

对于EasyDL口罩检测模型请注意输入图片中人脸大小建议保持在88到9696像素,可根据场景远近程度缩放图片后传入

public interface DetectInterface { List<DetectionResultModel> detect(Bitmap bitmap, float confidence) throws BaseException;

// 如InferManger 继承 DetectInterface

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 DetectionResultModel List 异常:一般首次出现。可以打印出异常错误码。

DetectionResultModel

- label:标签,定义在label_list.txt中
- confidence:置信度
- bounds: Rect, 左上角和右下角坐标

3.3.6 图像分割Segmentation

```
public interface SegmentInterface {
List<SegmentationResultModel> segment(Bitmap bitmap, float confidence) throws BaseException;
// 如InferManger 继承 SegmentInterface
```

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 SegmentationResultModel 异常:一般首次出现。可以打印出异常错误码。

SegmentationResultModel

label:标签,定义在label_list.txt中
confidence:置信度
lableIndex:标签对应的序号
box: Rect对象表示的对象框
mask:byte[]表示的原图大小的0,1掩码,绘制1的像素即可得到当前对象区域
mask 字段说明,如何绘制掩码也可参考demo工程
 101
image 110 => mask(byte[]) 101 110 011

011

3.3.7 文字识别OCR

暂时只支持通用ARM引擎,不支持其它引擎,暂时只支持EasyEdge的开源OCR模型。

```
public interface OcrInterface {
List<OcrResultModel> ocr(Bitmap bitmap, float confidence) throws BaseException;
// 如InferManger 继承 OcrInterface
```

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 OcrResultModel List,每个OcrResultModel对应结果里的一个四边形。 异常:一般首次出现。可以打印出异常错误码。

OcrResultModel

- label:识别出的文字
- confidence:置信度
- List<Point>:4个点构成四边形

3.3.8 姿态估计Pose

暂时只支持通用ARM引擎,不支持其它引擎

public interface PoseInterface { List<PoseResultModel> pose(Bitmap bitmap) throws BaseException; // 如InferManger 继承 PoseInterface

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 PoseResultModel List 异常:一般首次出现。可以打印出异常错误码。

PoseResultModel

- label : 标签,定义在label_list.txt中

- confidence:置信度

- Pair<Point, Point>: 2个点构成一条线

3.3.9 释放

释放后这个对象不能再使用,如果需要使用可以重新new一个出来。

public void destory() throws BaseException

3.3.10 整体示例

以通用ARM的图像分类预测流程为例:

```
try {
// step 1: 准备配置类
InferConfig config = new InferConfig(context.getAssets(), "infer");
// step 2: 准备预测 Manager
InferManager manager = new InferManager(context, config, "");
// step 3: 准备待预测的图像,必须为 Bitmap.Config.ARGB_8888 格式,一般为默认格式
Bitmap image = getFromSomeWhere();
// step 4: 预测图像
List<ClassificationResultModel> results = manager.classify(image, 0.3f);
// step 5: 解析结果
for (ClassificationResultModel resultModel : results) {
 Log.i(TAG, "labelIndex=" + resultModel.getLabelIndex()
  + ", labelName=" + resultModel.getLabel()
  + ", confidence=" + resultModel.getConfidence());
}
// step 6: 释放资源。预测完毕请及时释放资源
manager.destroy();
} catch (Exception e) {
Log.e(TAG, e.getMessage());
}
```

```
"autocheck_qcom": true, // 如果改成false, sdk跳过检查手机是否是高通的Soc,非高通的Soc会奔溃直接导致app闪退
"snpe_runtimes_order": [],
// 不填写为自动,按照 {DSP, GPU, GPU_FLOAT16, CPU}次序尝试初始化,也可以手动指定如[2,1,3,0],具体数字的定义见下段

public interface SnpeRuntimeInterface {
    int CPU = 0;
    int GPU = 1;
    int DSP = 2;
    int GPU_FLOAT16 = 3;
```

// SnpeManager 中,使用public static ArrayList<Integer> getAvailableRuntimes(Context context) 方法可以获取高通SOC支持的运行方式

集成指南

}

- 1. 复制库文件libs
- 2. 添加Manifest权限
- 3. 复制模型文件
- 4. 添加调用代码(见上一步具体接口说明)
- 4.1 复制库文件libs A. 如果项目里没有自己的jar文件和so文件:

```
复制app/libs 至自己项目的app/libs目录。
参照demo的app/build.gradle 中添加
android {
....
defaultConfig {
ndk {
abiFilters 'armeabi-v7a', 'arm64-v8a'
}
}
sourceSets {
main {
jniLibs.srcDirs = ['libs']
}
}
```

B. 如果项目里有自己的jar文件,但没有so文件

```
easyedge.jar文件同自己的jar文件放一起
arm64-v8a和armeabi-v7a放到app/src/main/jniLibs目录下
参照demo的app/build.gradle 中添加
android {
....
defaultConfig {
ndk {
abiFilters 'armeabi-v7a', 'arm64-v8a'
}
}
```

C. 如果项目里有自己的jar文件和so文件

```
easyedge.jar文件同自己的jar文件放一起
arm64-v8a和armeabi-v7a取交集和自己的so放一起,交集的意思是比如自己的项目里有x86目录,必须删除x86。
参照demo的app/build.gradle 中添加
android {
....
defaultConfig {
ndk {
abiFilters 'armeabi-v7a', 'arm64-v8a' // abiFilter取交集,即只能少不能多
}
```

jar文件库如果没有设置成功的,编译的时候可以发现报错。

so库如果没有编译进去的话,也可以通过解压apk文件确认。运行的时候会有类似jni方法找不到的报错。

4.2 Manifest配置

参考app/src/main/AndroidManifest.xml文件,添加:

```
    <uses-permission android:name="android.permission.INTERNET" /><uses-permission android:name="android.permission.WRITE_EXTERNAL_STORAGE" /><uses-permission android:name="android.permission.ACCESS_NETWORK_STATE" /><uses-permission android:name="android.permission.READ_PHONE_STATE" />Android 11 支持 --><uses-permission android.permission.MANAGE_EXTERNAL_STORAGE" tools:ignore="scopedStorage" />高版本 Android 支持 --><application</li>
```

```
android:requestLegacyExternalStorage="true"
android:usesCleartextTraffic="true">
</application>
```

4.3 混淆规则 (可选) 请不要混淆SDK里的jar文件。

```
-keep class com.baidu.ai.edge.core.*.*{ *; }
```

4.4 Android 11支持 除Manifest中必要配置外,请参考BaseActivity获取所有文件访问权限,否则可能影响SDK正常使用。

SDK 默认使用 easyedge-sdk.jar,未启用 AndroidX,若您的项目使用 AndroidX,并在集成中提示 android.support 相关错误,请参考 app/build.gradle 使用 etc/easyedge-sdk-androidx.jar 以支持 AndroidX:

```
// app/build.gradle
dependencies {
    implementation project(':camera_ui')
    implementation files('libs/easyedge-sdk-androidx.jar') // 修改 jar 包依赖
}
```

错误码 | 错误码 | 错误描述 | 详细描述及解决方法 | | --- | --- | 1001 | assets 目录下用户指定的配置文件不存在或不正确 | SDK使用assets目录下一系列文件作为配置文件。如果文件缺失或内容不正确,则有此报错 || 1002 | json格式的配置文件解析出错 | 如缺少某些字段。正常情况下,配置文件请不要修改 || 1003 | 应用缺少权限 | 请根据提示动态申请缺少的权限 || 19xx | Sdk内部错误 | 请与百度人员联系 || 2001 | XxxxMANAGER 只允许一个实例 | 如已有XxxxMANAGER对象,请调用destory方法 || 2002 | XxxxMANAGER 已经调用过destory方法 | 在一个已经调用destory方法的DETECT_MANAGER对象上,不允许再调用任何方法 || 2003 | 传入的assets下模型文件路径非法 |比如缺少模型文件, XxxxConfig.getModelFileAssetPath() 返回为null || 2012 | JNI内存错误 | heap的内存不够 || 2103 | license过期 | license失效或者系统时间有异常 || 2601/2602 | assets 目录下模型文件打开/读取失败 | 请根据报错信息检查模型文件是否存在 || 27xx | Sdk内部错误 | 请与百度人员联系 || 4011 | SDK类型 | 28xx | 引擎内部错误 | 请与百度人员联系 || 29xx | Sdk内部错误 | 请与百度人员联系 || 3003 | 域型加载错误 | 请确认模型放置于能被加载到的合法路径中,并确保配置文件正确 || 3002 | 模型卸载错误 | 请与百度人员联系 || 4011 | SDK类型 与设备硬件不匹配 | 比如适配DSP的SDK运行在麒麟芯片上会出现此报错,请在部署包支持的硬件上使用SDK || 50xx | SDK调用异常 | 请与百度 人员联系 |

报错日志收集 通常 Logcat 可以看见日志及崩溃信息,若设备无法获取日志信息,可使用 Demo 中的 xCrash 工具:

```
// 1. 引入 app/build.gradle 的 xCrash 依赖
android {
  ...
  dependencies {
     implementation 'com.iqiyi.xcrash:xcrash-android-lib:2.4.5' // 可以保存崩溃信息,默认未引入
  }
}
// 2. 启用日志收集。日志将保存在 /sdcard/<包名>/xCrash
// app/src/main/java/com.baidu.ai.edge/demo/MyApplication.java
protected void attachBaseContext(Context context) {
  // 日志保存位置
  String basePath = Environment.getExternalStorageDirectory().toString() + "/" + context.getPackageName();
  // 启用
  XCrash.InitParameters params = new XCrash.InitParameters();
  params.setAppVersion(BaseManager.VERSION);
  params.setLogDir(basePath + "/xCrash");
  XCrash.init(this, params);
}
```

の iOS集成文档

简介

本文档描述 EasyEdge/EasyDL iOS 离线预测SDK相关功能;

目前支持EasyEdge的功能包括:

- 图像分类
- 物体检测
- 人脸检测
- 姿态估计
- 百度OCR模型

目前支持EasyDL的功能包括:

- 图像分类
- 物体检测
- 图像分割

系统支持

系统:

- 通用arm版本:iOS 9.0 以上
- A仿生芯片版: iOS 15.0 及以上
- 硬件: arm64 (Starndard architectures) (暂不支持模拟器)

内存:图像分割模型需要手机内存3GB以上,并尽量减少其他程序内存占用

离线SDK包说明

根据用户的选择,下载的离线SDK,可能包括以下类型:

- EasyEdge
 - 通用ARM版:支持iPhone5s, iOS 9.0 以上所有手机。
 - A仿生芯片版:支持iPhone5s, iOS 15.0 以上手机。充分利用苹果A系列仿生芯片优势,在iPhone 8以上机型中能有显著的速度提升。

- EasyDL 通用版/全功能AI开发平台BML (原EasyDL专业版)
 - 通用ARM版:支持iPhone5s, iOS 9.0 以上所有手机。
 - A仿生芯片版:支持iPhone5s, iOS 15.0 以上手机。充分利用苹果A系列仿生芯片优势,在iPhone 8以上机型中能有显著的速度提升。
 - 自适应芯片版:同时整合了以上两种版本,自动在iOS 15以下中使用通用ARM版,在iOS 15以上系统中使用A仿生芯片版,自适应系统,但SDK体积相对较大。
- AI市场试用版SDK

SDK大小说明

SDK库的二进制与_TEXT增量约3M。

资源文件大小根据模型不同可能有所差异。

物体检测(高性能)的DemoApp在iPhone 6, iOS 11.4下占用空间实测小于40M。

虽然SDK库文件很大(体现为SDK包文件很大,ipa文件很大),但最终应用在用户设备中所占用的大小会缩小很多。 这与multi architechtures、bitcode和AppStore的优化有关。

获取序列号 生成SDK后,点击获取序列号进入控制台获取。EasyEdge控制台、EasyDL控制台、BML控制台。

试用版SDK在SDK的RES文件夹中的SN.txt中包含试用序列号。

更换序列号、更换设备时,首次使用需要联网激活。激活成功之后,有效期内可离线使用。

Release Notes

时间	版本	说明
2023.08.31	0.7.13	新增按实例数鉴权;迭代优化
2023.06.29	0.7.12	迭代优化
2023.05.17	0.7.11	CoreML引擎升级,支持更多语义分割模型;兼容横屏;迭代优化
2023.03.16	0.7.10	支持更多语义分割模型;迭代优化
2022.12.29	0.7.9	ARM引擎升级;迭代优化
2022.10.27	0.7.8	支持更多检测模型;迭代优化
2022.09.15	0.7.7	支持更多检测模型;迭代优化
2022.07.28	0.7.6	迭代优化
2022.06.29	0.7.5	支持EasyEdge语义分割模型;CoreML引擎升级,新增EasyEdge检测模型支持;迭代优化
2022.05.18	0.7.4	ARM引擎升级;支持EasyDL物体检测超高精度模型;支持更多加速版模型发布;迭代优化
2022.03.25	0.7.3	ARM引擎升级;支持更多检测模型
2021.12.22	0.7.2	支持EasyEdge更多姿态估计模型;迭代优化
2021.10.20	0.7.1	ARM引擎升级
2021.07.29	0.7.0	迭代优化
2021.04.06	0.6.1	ARM引擎升级
2021.03.09	0.6.0	支持EasyEdge人脸检测及姿态估计模型
2020.12.18	0.5.7	ARM引擎升级
2020.09.17	0.5.6	CoreML引擎升级,支持AI市场试用版SDK
2020.08.11	0.5.5	CoreML支持EasyDL专业版模型,支持EasyEdge OCR模型
2020.06.23	0.5.4	ARM引擎升级
2020.04.16	0.5.3	ARM引擎升级;支持压缩加速版模型
2020.03.13	0.5.2	ARM引擎升级;支持图像分割模型
2020.01.16	0.5.1	ARM引擎升级;增加推荐阈值支持
2019.12.04	0.5.0	ARM引擎升级;增加coreml3的支持
2019.10.24	0.4.5	支持EasyDL专业版;ARM引擎升级
2019.08.30	0.4.4	支持EasyDL经典版图像分类高性能、高精度
2019.06.20	0.4.3	引擎优化
2019.04.12	0.4.1	支持EasyDL经典版物体检测高精度、高性能模型
2019.03.29	0.4.0	引擎优化,支持CoreML;
2019.02.28	0.3.0	引擎优化,性能与效果提升;
2018.11.30	0.2.0	第一版!

快速开始 文件结构说明

测试Demo

按如下步骤可直接运行 SDK 体验 Demo:

步骤一:用 Xcode 打开 EasyDLDemo/EasyDLDemo.xcodeproj 步骤二:配置开发者自己的签名

步骤三:连接手机运行,不支持模拟器

检测模型运行示例:







SDK使用说明 集成指南 步骤一:依赖库集成 步骤二:import <EasyDL/EasyDL.h>, import <Vision/Vision.h>

依赖库集成

- 1. 复制 LIB 目录至项目合适的位置
- 2. 配置 Build Settings 中 Search paths: 以 SDK 中 LIB 目录路径为例
- Framework Search Paths : \${PROJECT_DIR}/../LIB/lib
- Header Search Paths : \${PROJECT_DIR}/../LIB/include
- Library Search Paths : \${PROJECT_DIR}/../LIB/lib

集成过程如出现错误,请参考 Demo 工程对依赖库的引用

使用流程

1. 生成模型,下载SDK 开发者在官网下载的SDK已经自动为开发者配置了模型文件和相关配置,开发者直接运行即可。

2. 使用序列号激活 2.1. 离线激活 (默认鉴权方式) 首次联网激活,后续离线使用

将前面申请的序列号填入:

[EasyDL setSerialNumber:@"!!!Enter Your Serial Number Here!!!"];

根据序列号类型,序列号与BundleID绑定或与BundleID+设备绑定。 请确保设备时间正确。

2.2. 按实例数激活 周期性联网激活,离线后会释放所占用鉴权,启动时请确保心跳间隔小于等于生成序列号时填写的定期确认时间

填入序列号,配置按实例数鉴权并设置心跳间隔:

```
// 设置序列号
```

[EasyDL setSerialNumber:@"!!!Enter Your Serial Number Here!!!"]; // 配置实例数鉴权及心跳间隔,单位:秒 [EasyDL setInstanceAuthMode:10000];

3. 初始化模型

EasyDLModel *_model = [[EasyDLModel alloc] initModelFromResourceDirectory:@"easyedge" withError:&err];

请注意相关资源必须以folder reference方式加入Xcode工程。也即默认的easyedge文件夹在Xcode文件列表里显示为蓝色。

4. 调用检测接口

UlImage *img =; NSArray *result = [model detectUIImage:img withFilterScore:0 andError:&err];

/**

```
* 检测图像
```

```
* @param image 带检测图像
```

- * @param score 只返回得分高于score的结果(0~1)
- * @return 成功返回识别结果,NSArray的元素为对应模型的结果类型;失败返回nil,并在err中说明错误原因

*/

- (NSArray *)detectUIImage:(UIImage *)image

```
withFilterScore:(CGFloat)score
```

```
andError:(NSError **)err;
```

返回的数组类型如下,具体可参考 EasyDLResultData.h 中的定义:

模型类型	类型
图像-图像分类	EasyDLClassfiData
图像-物体检测/人脸检测	EasyDLObjectDetectionData
图像-实例分割/语义分割	EasyDLObjSegmentationData
图像-姿态估计	EasyDLPoseData
图像-文字识别	EasyDLOcrData

错误说明

SDK的方法会返回NSError错,直接返回的NSError的错误码定义在EEasyDLErrorCode中。NSError附带message(有时候会附带 NSUnderlyingError),开发者可根据code和message进行错误判断和处理。

FAQ

1. 如何多线程并发预测?

SDK内部已经能充分利用多核的计算能力。不建议使用并发来预测。

如果开发者想并发使用,请务必注意EasyDLModel所有的方法都不是线程安全的。请初始化多个实例进行并发使用,如

```
- (void)testMultiThread {
 UlImage *img = [UlImage imageNamed:@"1.jpeg"];
 NSError *err;
 EasyDLModel * model1 = [[EasyDLModel alloc] initModelFromResourceDirectory:@"easyedge" withError:&err];
 EasyDLModel * model2 = [[EasyDLModel alloc] initModelFromResourceDirectory:@"easyedge" withError:&err];
 dispatch_queue_t queue1 = dispatch_queue_create("testQueue", DISPATCH_QUEUE_CONCURRENT);
 dispatch_queue_t queue2 = dispatch_queue_create("testQueue2", DISPATCH_QUEUE_CONCURRENT);
 dispatch_async(queue1, ^{
    NSError *detectErr;
    for(int \; i = 0; \; i < 1000; \; ++i) \; \{
      NSArray * res = [model1 detectUIImage:img withFilterScore:0 andError:&detectErr];
      NSLog(@"1: %@", res[0]);
 dispatch_async(queue2, ^{
   NSError *detectErr;
    for(int i = 0; i < 1000; ++i) {
      NSArray * res = [model2 detectUlImage:img withFilterScore:0 andError:&detectErr];
      NSLog(@"2: %@", res[0]);
```

2. 编译时出现 Undefined symbols for architecture arm64: ...

- 出现 cxx11, vtable 字样:请引入 libc++.tbd
- 出现 cv::Mat 字样:请引入 opencv2.framework
- 出现 CoreML, VNRequest 字样:请引入CoreML.framework 并务必#import <CoreML/CoreML.h>
- 3. 运行时报错 Image not found: xxx ...

请Embed具体报错的库。 4.编译时报错:Invalid bitcode version 这个可能是开发者使用的xcode低于12导致,可以升级至12版本。

の Windows集成文档

简介

本文档介绍物体检测通用小型设备Windows SDK的使用方法。

- 硬件支持:
 - Intel CPU 普通版 * x86_64

- CPU 加速版 Intel Xeon with AVX2 and AVX512 Intel Core Processors with AVX2 Intel Atom Processors with SSE * AMD Core Processors with AVX2
- Intel Movidius MyRIAD2/MyRIAD X (仅支持Win10)
- 操作系统支持
 - 普通版:64位 Windows 7 及以上,64位Windows Server2012及以上
 - 加速版:64位 Windows 10,64位Windows Server 2019及以上
- 环境依赖 (必须安装以下版本)
 - .NET Framework 4.5
 - Visual C++ Redistributable Packages for Visual Studio 2013
 - Visual C++ Redistributable Packages for Visual Studio 2015-2019
- 协议
 - HTTP
- 更详细的环境说明可参考SDK内的README.md

*intel 官方合作,拥有更好的适配与性能表现

Release Notes | 时间 | 版本 | 说明 | | ------ | ----- | 2023-08-30 | 1.8.3 | 新增支持按实例数鉴权 || 2023-06-29 | 1.8.2 | 优化模型 算法 || 2023-05-17 | 1.8.1 | 预测引擎升级, 修复部分网络内存泄露问题 || 2023-03-16 | 1.8.0 | 预测引擎升级 || 2022-12-29 | 1.7.2 | 预测引 擎升级 || 2022-10-27 | 1.7.1 | 优化模型算法 || 2022-09-15 | 1.7.0 | 新增支持表格预测 || 2022-07-28 | 1.6.0 | 优化模型算法 || 2022-05-27 | 1.5.1 | 新增支持BML Cloud小目标检测模型 || 2022-05-18 | 1.5.0 | 修复各别机器下程序崩溃的问题 || 2022-04-25 | 1.4.1 | EasyDL, BML升级 支持paddle2模型 || 2022-03-25 | 1.4.0 | 优化模型算法 || 2021-12-22 | 1.3.5 | CPU基础版推理引擎优化升级 ; demo程序优化环境依赖检测 || 2021-10-20 | 1.3.4 | CPU加速版推理引擎优化升级 || 2021-08-19 | 1.3.2 | 新增DEMO二进制文件 || 2021-06-29 | 1.3.1 | 预测引擎升级 || 2021-05-13 | 1.3.0 | 模型发布新增多种加速方案选择 ; 目标追踪支持x86平台的GPU及加速版 ; 展示已发布模型性能评估报告 || 2021-04-08 | 1.2.3 | 支持BML平台模型仓库本地上传模型 || 2021-03-09 | 1.2.2 | CPU加速版支持int8量化模型 || 2021-01-27 | 1.2.1 | 新增模型支持 ; 性能优 化 ; 问题修复 || 2020.12.18 | 1.2.0 | 推理引擎升级 || 2020-11-26 | 1.1.20 | 新增一些模型的加速版支持 || 2020.10.29 | 1.1.20 | 修复已知问 题 || 2020.05.15 | 1.1.16 | 优化性能 , 修复已知问题 || 2020-09-17 | 1.1.19 支持更多模型 || 2020.08.11 | 1.1.18 | 支持专业版更多模型 || 2020.06.23 | 1.1.17 | 支持专业版更多模型 || 2020.02.23 | 1.1.13 | 支持多阶段模型 || 2020.04.16 | 1.1.15 | 升级引擎版本 || 2020.03.13 | 1.1.14 | 支持EdgeBoardVMX || 2020.02.23 | 1.1.13 | 支持多阶段模型 || 2020.01.16 | 1.1.12 | 预测默认使用推荐阈值 || 2019.02.9 | 1.1.8 | CPU 加速版支持物体检测高精度 || 2019.12.04 | 1.1.10 | 支持图像分割 || 2019.01.21 | 1.1.9 | 支持 EasyDL 专业版 || 2019.08.29 | 1.1.8 | CPU 加速版支持 || 2019.07.19 | 1.1.7 | 提供模型更新工具 || 2019.05.16 | 1.1.3 | NVIDIA GPU 支持 || 2019.03.15 | 1.1.0 | 架构与功能完善 || 2019.02.28 | 1.0.6 | 引擎功能完善 || 2019.02.13 | 1.0.5 | paddlepaddle 支持 || 2018.11.30 | 1.0.0 | 第一版 !

快速开始

1. 安装依赖

必须安装:

安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642

Visual C++ Redistributable Packages for Visual Studio 2013

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

Visual C++ Redistributable Packages for Visual Studio 2015-2019

https://docs.microsoft.com/en-us/cpp/windows/latest-supported-vc-redist?view=msvc-160

Openvino (仅使用Python Intel Movidius必须)

- 使用 OpenVINO[™] toolkit 安装,请参考 OpenVINO toolkit 文档安装 2020.3.1LTS(必须)版本, 安装时可忽略Configure the Model Optimizer及 后续部分。
- 使用源码编译安装,请参考 Openvino Inference Engine文档编译安装 2020.3.1LTS (必须)版本。

安装完成后,请设置环境变量OPENVINO_HOME为您设置的安装地址,默认是C: \Program Files (x86) \IntelSWTools,并确保文件夹下的openvino 的快捷方式指到了2020.3.1LTS版本。

充属性			
-算机名 硬件 高级 系统保护 远程	xiaoyuchen 的用户变量(U)		
	变量	值	
要进行大多数更改,你必须作为管理员登录。		Cilliser (OpeDrive	
性能	Path	C:\Program Files (x86)\IntelSWTools	asoft\WindowsApp
视觉效果,处理器计划,内存使用,以及虚拟内存	TEMP	C:\Users' AppData\Local\Temp	o
设置(5)	тмр	C:\Users\www.jeAppData\Local\Temp	
用户配置文件			
与登录帐户相关的桌面设置		新建(N) 编	攝(E) 删除(D)
设置(E)	系统变量(S)		
	变量	值	^
	ComSpec	C:\Windows\system32\cmd.exe	
条统启动、条统改晫和调试信息	INTEL DEV REDIST	C:\Windows\system32\Drivers\DriverData C:\Program Files (x86)\Common Files\Inte	\Shared Libraries\
公署(工)	MIC_LD_LIBRARY_PATH	%INTEL_DEV_REDIST%compiler\lib\intel64	win_mic
	NUMBER_OF_PROCESSORS	8	
	OS Dath	Windows_NT	in a la' Anna lination au C
环境变量(<u>N</u>)	Path	C:\Program Files (xoo)\Parallels\Parallels	oois (Applications; c v
		新建(W) 编	辑(I) 删除(L)
确定 取消 应用(A)			
		lē	腚 取消
IntelSWTools			
共享 查看			
🔜 « Program Files (x86) > IntelS	WTools →	5 v	⊘ 搜索
	^		
		修改日期	胡
💦 openvi	no	2020/4	/15 23:41
openvi	no_2019.2.275	2020/3	/10 19:34
* openvi	no 2020 2 117	2020/4	/15 22.42
	10_2020.2.117	2020/4	15 23.42

注意事项

- 1. 安装目录不能包含中文
- 2. Windows Server 请自行开启,选择"我的电脑"——"属性"——"管理"——"添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启即 可。

2. 运行离线SDK

解压下载好的SDK,打开EasyEdge.exe,输入Serial Num,选择鉴权模式,点击"启动服务",等待数秒即可启动成功,本地服务默认运行在

http://127.0.0.1:24401/

其他任何语言只需通过HTTP调用即可。

如启动失败,可参考如下步骤排查:

🗘 百度智能云	故障排查
零门槛AI开发	
平台EasyDL	2 M
	2.
Windows离线SDK(公有云发)	
布)●问题排查	
1	

2.1 离线鉴权 (默认鉴权模式) 首次联网激活,后续离线使用

Al ver. 1.2.0.0 core. 1	1.8.1 sec. 1.5.1 — □ >	<
Model	[4] A.R. Talkinson, Mathematical Society, 199	
Serial Num	The second se	
Host	127.0.0.1	
Port	24401	
鉴权模式	● 离线鉴权 ○ 按实例数鉴权	
服务状态	服务未启动	
	启动服务	
	是否开机自动启动 [

2.2 按实例数鉴权 周期性联网激活,离线后会释放所占用鉴权,启动时请确保心跳间隔小于等于生成序列号时填写的定期确认时间

AI ver. 1.2.0.0 core. 1	1.8.1 sec. 1.5.1 —	×
Model	[4] A.P. Salara and R.A.M. A. Marini	
Serial Num		
Host	127.0.0.1	
Port	24401	
鉴权模式	○ 离线鉴权 ④ 按实例数鉴权 心跳间隔 20 秒 ->	
服务状态	服务未启动	
	启动服务	
	是否开机自己	动启动 🗌

基于源码集成时,若需要按实例数鉴权,需要通过代码指定使用按实例数鉴权

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2); global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 20);

或通过环境变量指定

set EDGE_CONTROLLER_KEY_AUTH_MODE=2 set EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=20

心 2.3 序列号激活错误码

错误码	文案	描述
4001	parameters missing. 参数缺失	
4002	parameters invalid. 参数不合法	
4003	model invalid. 模型信息不合法	
4004	no more activation left. 该序列号和该设备的激活次数超上限	
4005	the serial key is out of date. 该序列号过期	
4006	the serial key has been activated. 该序列号已被其他设备激活	该序列号已被其他设备激活,不能重复激活。
4007	account invalid. 序列号不能用于其他账号的模型	序列号不能用于其他账号的模型,只能用于绑定账号的模型。
4008	serial key invalid. 序列号不合法	序列号不存在或找不到
4009	bundle id invalid. 包名不合法	
4010	product invalid. 产品不合法	如easydl的SDK使用BML的序列号来激活,会报该错误
4011	platform invalid. 平台不合法	
4012	activate too frequent. 激活太频繁	激活太频繁,请稍后再进行激活。
4013	device type and license type not match. 硬件类型和序列号类型不匹配	如使用加速版序列号激活基础版SDK会报该错误
4014	exceed max activate device num. 超过最大授权数量	
4015	technology invalid. 技术类型不合法	
4016	exceed max activate entity num. 超过最大模型数量	
4017	device invalid. 设备不合法	
4018	model invalid. 模型不合法	

3. Demo示例(以图像服务为例)

服务运行成功,此时可直接在浏览器中输入http://127.0.0.1:24401,在h5中测试模型效果。



使用说明

调用说明 使用示例如下:

python	
с#	
C++	
java	

import requests

```
with open('./1.jpg', 'rb') as f:
img = f.read()
```

```
**params 为GET参数 data 为POST Body**
result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1},
data=img).json()
```

 $\frac{\text{print}(\text{result})}{}$

结果 获取的结果存储在response字符串中。请求参数

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送图片二进制。

返回参数|字段|类型|取值|说明||--------|-----|-----|-----|| confidence|float|0~1|检测的置信度||label|string||检测的类 别||index|number||检测的类别||x1, y1|float|0~1|矩形的左上角坐标(相对长宽的比例值)||x2, y2|float|0~1|矩形的右下角坐标(相 对长宽的比例值)|

关于矩形坐标

- x1 * 图片宽度 = 检测框的左上角的横坐标
- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2 * 图片宽度 = 检测框的右下角的横坐标

y2 * 图片高度 = 检测框的右下角的纵坐标

集成指南

基于HTTP集成

通过EasyEdge.exe启动服务后,参照上面的调用说明,通过HTTP请求集成到自己的服务中

基于c++ dll集成

集成前提

解压开的SDK包中包含src、lib、dll、include四个目录才支持基于c++ dll集成

集成方法

参考src目录中的CMakeLists.txt进行集成

基于c# dll集成

集成前提

解压开的SDK包中包含src\demo_serving_csharp、dll两个目录才支持基于c# dll集成

集成方法

参考src\demo_serving_csharp目录中的CMakeLists.txt进行集成

FAQ

1. 服务启动失败,怎么处理?

根据SDK内的README.md检查依赖是否都已正确安装

请确保相关依赖都安装正确,版本必须如下: *.NET Framework 4.5* Visual C++ Redistributable Packages for Visual Studio 2013 * Visual C++ Redistributable Packages for Visual Studio 2015

如使用的是Python Intel Movidius版,需额外确保Openvino安装正确,版本为2020.3.1LTS版 如使用Windows Server,需确保开启桌面体验

2. 服务调用时返回为空,怎么处理?调用输入的图片必须是RGB格式,请确认是否有alpha通道。

3. 多个模型怎么同时使用? SDK设置运行不同的端口,点击运行即可。

4. JAVA、C#等其他语言怎么调用SDK?

参考 https://ai.baidu.com/forum/topic/show/943765

5. 启动失败,缺失DLL? 打开EasyEdge.log,查看日志错误,根据提示处理 缺失DLL,请使用 https://www.dependencywalker.com/ 查看相应模 块依赖DLL缺失哪些,请自行下载安装

6. 启动失败,报错NotDecrypted? Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

7. 启动失败,报错 SerialNum无效

日志显示failed to get/check device id(xxx)或者 Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 C:\Users\\${用户名}\.baidu\easyedge 目录,再重新激活。

8. 勾选"开机自动启动"后,程序闪退

一般是写注册表失败。

可以确认下HKEY_CURRENT_USER下Software \Microsoft \Windows \CurrentVersion \Run能否写入(如果不能写入,可能被杀毒软件等工具管制)。也可以尝试基于bin目录下的easyedge_serving.exe命令行形式的二进制,自行配置开机自启动。

其他问题 如果无法解决,可到论坛发帖: https://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

_のLinux集成文档-C++

简介

本文档介绍EasyEdge/EasyDL的Linux CPP SDK的使用方法。

- 网络类型支持: 图像分类 物体检测 图像分割
- 硬件支持:
 - CPU: aarch64 armv7hf
 - GPU: ARM Mali G系列
 - ASIC: Hisilicon NNIE1.1 on aarch64 (Hi3559AV100/Hi3559CV100等)
 - ASIC: Hisilicon NNIE1.2 on armv7I (Hi3519AV100/Hi3559V200等)
 - Intel Movidius MyRIAD2 / MyRIAD X on x86_64
 - Intel Movidius MyRIAD2 / MyRIAD X on armv7I
 - Intel Movidius MyRIAD2 / MyRIAD X on aarch64
 - Intel iGPU on x86_64
 - 比特大陆 Bitmain SE5 (BM1684)

- 瑞芯微 RK3399Pro / RV1109 / RV1126 / RK3568 / RK3588
- 华为 Atlas200
- 晶晨 A311D
- 寒武纪 MLU220 on aarch64
- 英特尔 iGPU
- 操作系统支持:
 - Linux (Ubuntu, Centos, Debian等)
 - 海思HiLinux
 - 树莓派Raspbian/Debian
 - 瑞芯微Firefly

性能数据参考算法性能及适配硬件

Release Notes | 时间 | 版本 | 说明 | | --- | ---- | | 2023.08.31 | 1.8.3 | Atlas系列Socs支持语义分割模型, Atlas Cann版本升级至6.0.1 | | 2023.06.29 | 1.8.2 | 比特大陆版本升级至V23.03.01 | | 2023.05.17 | 1.8.1 | 新增支持intel iGPU + CPU异构模式 | | 2023.03.16 | 1.8.0 | 新增 支持瑞芯微RK3588 || 2022.10.27 | 1.7.1 | 新增语义分割模型http请求示例 || 2022.09.15 | 1.7.0 | 新增瑞芯微 RK3568 支持, RK3399Pro、 RV1126升级到RKNN1.7.1 || 2022.07.28 | 1.6.0 | 引擎升级;新增英特尔 iGPU 支持 || 2022.04.25 | 1.4.1 | EasyDL, BML升级支持paddle2模 型 | | 2022.03.25 | 1.4.0 | EasyDL新增上线支持晶晨A311D NPU预测引擎;Arm CPU、Arm GPU引擎升级;atlas 200在EasyDL模型增加多个量 化加速版本; || 2021.12.22 | 1.3.5 | RK3399Pro, RV1109/RV1126 SDK扩展模型压缩加速能力,更新端上推理库版本;边缘控制台IEC功能升 级,适配更多通用小型设备,NNIE 在EasyDL增加量化加速版本;Atlas200升级到Cann5.0.3 | | 2021.06.29 | 1.3.1 | 视频流解析支持调整分辨 率;预测引擎升级;设备端sdk新增支持瑞芯微RV1109、RV1126 || 2021.05.13 | 1.3.0 | 新增视频流接入支持;EasyDL模型发布新增多种加速 方案选择;目标追踪支持x86平台的CPU、GPU加速版;展示已发布模型性能评估报告 | | 2021.03.09 | 1.2.0 | http server服务支持图片通过 base64格式调用 || 2021.01.27 | 1.1.0 | EasyDL经典版分类高性能模型升级;部分SDK不再需要单独安装OpenCV;新增RKNPU预测引擎支持; 新增高通骁龙GPU预测引擎支持 || 2020.12.18 | 1.0.0 | 1.0版本发布!安全加固升级、性能优化、引擎升级、接口优化等多项更新 || 2020.10.29 | 0.5.7 | 优化多线程预测细节 | | 2020.09.17 | 0.5.6 | 支持linux aarch64架构的硬件接入intel神经计算棒预测;支持比特大陆计算盒 SE50 BM1684 || 2020.08.11 | 0.5.5 | 支持linux armv7hf架构硬件(如树莓派)接入intel神经计算棒预测 || 2020.06.23 | 0.5.4 | arm引擎升级 || 2020.05.15 | 0.5.3 | 支持EasyDL 专业版新增模型 ; 支持树莓派(armv7hf, aarch64)| | 2020.04.16 | 0.5.2 | Jetson系列SDK支持多线程infer | | 2020.02.23 | 0.5.0 | 新增支持人脸口罩模型; Jetson SDK支持批量图片推理; ARM支持图像分割 || 2020.01.16 | 0.4.7 | 上线海思NNIE1.2, 支 持EasyEdge以及EasyDL; ARM引擎升级;增加推荐阈值支持 || 2019.12.26 | 0.4.6 | 海思NNIE支持EasyDL专业版 || 2019.11.02 | 0.4.5 | 移除 curl依赖;支持自动编译OpenCV;支持EasyDL 专业版 Yolov3;支持EasyDL经典版高精度物体检测模型升级||2019.10.25|0.4.4|ARM引擎升 级,性能提升30%;支持EasyDL专业版模型||2019.09.23|0.4.3|增加海思NNIE加速芯片支持||2019.08.30|0.4.2|ARM引擎升级;支持 分类高性能与高精度模型 || 2019.07.25 | 0.4.1 | 引擎升级,性能提升 || 2019.06.11 | 0.3.3 | paddle引擎升级;性能提升 || 2019.05.16 | 0.3.2 | 新增armv7l支持 || 2019.04.25 | 0.3.1 | 优化硬件支持 || 2019.03.29 | 0.3.0 | ARM64 支持;效果提升 || 2019.02.20 | 0.2.1 | paddle 引擎支持;效果提升||2018.11.30|0.1.0|第一版!|

【1.0 接口升级】参数配置接口从1.0.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例。【关于SDK包与RES模型文件夹配套使用的说明】 我们强烈建议用 户使用部署tar包中配套的SDK和RES一起使用。更新模型时,如果SDK版本号有更新,请务必同时更新SDK,旧版本的SDK可能无法正确适 配新发布出来的RES。

快速开始

SDK在以下环境中测试通过

- aarch64(arm64), Ubuntu 16.04, gcc 5.3 (RK3399)
- Hi3559AV100, aarch64, Ubuntu 16.04, gcc 5.3
- Hi3519AV100, armv7I , HiLinux 4.9.37, (Hi3519AV100R001C02SPC020)
- armv7hf, Raspbian, (Raspberry 3b)
- aarch64, Raspbian, (Raspberry 4b)
- armv7hf, Raspbian, (Raspberry 3b+)
- armv7hf, Ubuntu 16.04, (RK3288)
- Bitmain se50 BM1684, Debian 9
- Rockchip rk3399pro, Ubuntu 18.04
- Rockchip rv1126, Debain 10
- Rockchip rk3568, Ubuntu 20.04
- Rockchip rk3588, Ubuntu 20.04
- Atlas200(华为官网指定的Ubuntu 18.04版本)
- Amlogic A311D, Ubuntu 20.04
- MLU220, aarch64, Ubuntu 18.04

安装依赖

依赖包括

- cmake 3+
- gcc 5.4 以上(需包含 GLIBCXX_3.4.22) , gcc / glibc版本请以实际SDK ldd的结果为准
- opencv3.4.5 (可选)

依赖说明:树莓派树莓派Raspberry默认为armv7hf系统,使用SDK包中名称中包含armv7hf_ARM_的tar包。如果是aarch64系统,使用SDK包中名称中包含aarch64_ARM_的tar包。

在安装前可通过以下命令查看是32位还是64位:

getconf LONG_BIT 32

依赖说明:比特大陆SE计算盒 需要安装SophonSDK V23.05.01及以上版本,SDK的默认安装位置为/opt/sophon/,如SDK安装在自定义地址, 需在CMakeList.txt中指定SDK安装地址:

```
**这里修改并填入所使用的SophonSDK路径**
set(EDGE_BMSDK_ROOT "{这里填写sdk路径}")
```

可通过命令 bm-smi 查看内部SDK和驱动的版本号(SophonSDK V23.05.01对应的内部SDK和驱动为0.4.6)。对于使用旧版BM1684 SDK或者 低版本SophonSDK的用户,可参考SophonSDK安装包中的《LIBSOPHON 使用手册》先卸载旧版BM1684 SDK,安装、升级SophonSDK。

依赖说明:海思开发板 海思开发板需要根据海思SDK文档配置开运行环境和编译环境,SDK和opencv都需要在该编译环境中编译。NNIE1.2用 arm-himix200-linux交叉编译好的opencv,下载链接:https://pan.baidu.com/s/13QWOReeWx4ZwgYg4Iretyw 密码:yqOs。下载后修改SDK CMakesList.txt

依赖说明:RK3399Pro 所有用例基于 Npu driver版本1.7.1的RK3399pro开发板测试通过,SDK采用预编译模式,请务必确保板上驱动版本为 1.7.1 查看RK3399Pro板上driver版本方法:dpkg -I | grep 3399pro

依赖说明:RV1109/RV1126 所有用例基于Rknn_server版本1.7.3的RV1126开发板测试通过,SDK采用预编译模式,请务必确保板上驱动版本为 1.7.3 查看RV1109/RV1126板上Rknn_server版本方法:strings /usr/bin/rknn_server | grep build

依赖说明:RK3568 所有用例基于Rknn_server版本1.2.0的RK3568开发板测试通过, 查看RK3568板上Rknn_server版本方法: strings /usr/bin/rknn_server | grep build

依赖说明:RK3588 RK3588开发板需要确保环境正确安装了RKNPU驱动,平台用例基于v0.8.0版本的RKNPU驱动测试通过,查看RK3588NPU驱 动版本的方法: sudo cat /sys/kernel/debug/rknpu/version

依赖说明:晶晨A311D所有用例基于晶晨A311D开发板测试通过,需要驱动版本为 6.4.4.3(下载驱动请联系开发版厂商) 查看晶晨A311D开发 板驱动版本方法:dmesg | grep Galcore

依赖说明:英特尔iGPU用户在使用英特尔iGPU SDK前,需要根据英特尔官方文档提前安装好英特尔集成显卡驱动以及相关基础软件环境,安装 完成后通过 clinfo 指令确认OpenCL能够正常识别到集成显卡信息,正确识别集显情况下clinfo指令输出参考如下:

root@baidu-QiTianM430-N000:~# clinfo	
Number of platforms	1
Platform Name	Intel(R) OpenCL HD Graphics
Platform Vendor	Intel(R) Corporation
Platform Version	OpenCL 3.0
Platform Profile	FULL_PROFILE
Platform Extensions	cl_khr_byte_addressable_store cl_khr_device_uuid cl_khr_fp16 cl_khr_global_int32_base_atomics cl_khr_global_int32_extended_atomics cl_khr_icd cl_khr_local_int32
_base_atomics cl_khr_local_int32_extended_atomics	cl_intel_command_queue_families cl_intel_subgroups cl_intel_required_subgroup_size cl_intel_subgroups_short cl_khr_spir cl_intel_accelerator cl_intel_driver_dia
gnostics cl_khr_priority_hints cl_khr_throttle_hi	nts cl_khr_create_command_queue cl_intel_subgroups_char cl_intel_subgroups_long cl_khr_il_program cl_intel_mem_force_host_memory cl_khr_subgroup_extended_types c
l_khr_subgroup_non_uniform_vote cl_khr_subgroup_b	allot cl_khr_subgroup_non_uniform_arithmetic cl_khr_subgroup_shuffle cl_khr_subgroup_shuffle_relative cl_khr_subgroup_clustered_reduce cl_intel_device_attribute_
query cl_khr_suggested_local_work_size cl_intel_s	plit_work_group_barrier cl_khr_fp64 cl_khr_subgroups cl_intel_spirv_device_side_avc_motion_estimation cl_intel_spirv_media_block_io cl_intel_spirv_subgroups cl_k
hr_spirv_no_integer_wrap_decoration cl_intel_unif	ied_shared_memory cl_khr_mipmap_image cl_khr_mipmap_image_writes cl_intel_planar_yuv cl_intel_packed_yuv cl_intel_motion_estimation cl_intel_device_side_avc_moti
on_estimation cl_intel_advanced_motion_estimation	cl_khr_int64_base_atomics cl_khr_int64_extended_atomics cl_khr_image2d_from_buffer cl_khr_depth_images cl_khr_3d_image_writes cl_intel_media_block_io cl_intel_v
a_api_media_sharing cl_intel_sharing_format_query	cl_khr_pci_bus_info
Platform Host timer resolution	ins
Platform Extensions function suffix	INTEL
Platform Name	Intel(R) OpenCL HD Graphics
Number of devices	1
Device Name	Intel(R) UHD Graphics 630 [0x9bc8]
Device Vendor	Intel(R) Corporation
Device Vendor ID	0×8086
Device Version	OpenCL 3.0 NEO
Driver Version	22.53.25242.13
Device OpenCL C Version	OpenCL C 1.2
Device Type	GPU
Device Profile	FULL_PROFILE
Device Available	Yes
Compiler Available	Yes
Linker Available	Yes
Max compute units	24
Max clock frequency	1150MHz
Device Partition	(case)

使用序列号激活 请在官网获取序列号

纯离线服务说明 发布纯高线服务,将训练完成的模型部署在本地、高线调用模型。可以选择将模型部署在本地的服务器、小型设备、软硬一体方案专项适配硬件上, 通过API、SDK进一步集成,灵活适应不同业务场景。 发布新服务 控制台 服务器 通用小型设备 专项适配硬件 SDK API 此处发布、下载的SDK为未授权SDK,需要前往控制台获取序列号激活后才能正式使用,SDK内附有对应版本的Demo及开发文档,开发者可参考源代码完成开发, 模型名称 发布版本 () 应用平台 模型加速 发布状态 发布时间 () •已发布 基础版 2021-08-19 20:24 下载SDK 通用X86 CPU-Linux 精度无损压缩加速 下载加速版SDI 2021-08-19 20:24 基础版 已发布 2021-08-19 20:35 下载SDK 134318-V1 查看性能报告 英伟达GPU-Linux 精度无损压缩加速 sun_小目标test

基础版

SDK内bin目录下提供预编译二进制文件,可直接运行(二进制运行详细说明参考下一小节),用于图片推理和模型http服务,在二进制参数的 serial_num(或者serial_key)处填入序列号可自动完成联网激活(请确保硬件首次激活时能够连接公网,如果确实不具备联网条件,需要使用纯离 线模式激活,请下载使用百度智能边缘控制台纳管SDK)

• 已发布

已发布

2021-08-19 20:34

2021-08-19 18:17

下数加速版SDK

下载SDK

SDK内提供的一些二进制文件,填入序列号可完成自动激活,以下二进制具体使用说明参考下一小节 ./edgekit_serving --cfg=./edgekit_serving.yml ./easyedge_image_inference {model_dir} {image_name or image_directory} {serial_num} ./easyedge_serving {res_dir} {serial_key} {host} {port}

如果是基于源码集成,设置序列号方法如下

global_controller()->set_licence_key("")

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己 的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的变量或者源码设置) 实例数鉴权环境变量设置方法

export EDGE_CONTROLLER_KEY_AUTH_MODE=2 export EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=30

实例数鉴权源码设置方法

 $global_controller() -> set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2)$ global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 300)

基于预编译二进制测试图片推理和http服务测试图片推理模型资源文件默认已经打包在开发者下载的SDK包中。

对于硬件使用为: Intel Movidius MyRIAD2 / MyRIAD X / IGPU on Linux x86_64 / armv7hf / aarch64, 在编译或运行demo程序前执行以下命 会: source \${cpp_kit位置路径}/thirdparty/openvino/bin/setupvars.sh 或者执行 source \${cpp_kit位置路径}/thirdparty/openvino/setupvars.sh (openvino-2022.1+) 如果SDK内不包含setupvars.sh脚本,请忽略该提示

也能运行二进制)

./easyedge_image_inference {model_dir} {image_name or image_directory} {serial_num} LD_LIBRARY_PATH=../lib ./easyedge_image_inference ../.././RES /xxx/cat.jpeg "1111-1111-1111-1111" demo运行效果: 图片加载失败 > ./easyedge_image_inference ../../../../RES 2.jpeg 2019-02-13 16:46:12,659 INFO [EasyEdge] [easyedge.cpp:34] 140606189016192 Baidu EasyEdge Linux Development Kit 0.2.1(20190213) 2019-02-13 16:46:14,083 INFO [EasyEdge] [paddlev2_edge_predictor.cpp:60] 140606189016192 Allocate graph success. 2019-02-13 16:46:14,326 DEBUG [EasyEdge] [paddlev2_edge_predictor.cpp:143] 140606189016192 Inference costs 168 ms 1, 1:txt_frame, p:0.994905 loc: 0.168161, 0.153654, 0.920856, 0.779621 Done 启动http服务 bin目录下提供编译好的启动http服务二进制文件,可直接运行 **推荐使用 edgekit_serving 启动模型服务** LD_LIBRARY_PATH=../lib ./edgekit_serving --cfg=./edgekit_serving.yml **也可以使用 easyedge_serving 启动模型服务** **./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401} ** **LD_LIBRARY_PATH=../lib ./easyedge_serving ../../../RES "1111-1111-1111-1111" 0.0.0.0 24401**

后,日志中会显示

HTTP(or Webservice) is now serving at 0.0.0.0:24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试,网页右侧会展示模型推理结果



同时,可以调用HTTP接口来访问服务。

请求http服务 以图像预测场景为例(非语义分割模型场景,语义分割请求方式参考后面小节详细文档),提供一张图片,请求模型服务的示例参考 如下demo

python	
c#	
C++	

Baidu 百度智能云文档

java

import requests

```
with open('./1.jpg', 'rb') as f:
img = f.read()
```

```
**params 为GET参数 data 为POST Body**
result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1},
data=img).json()
```

 $\frac{\text{print}(\text{result})}{\text{result}}$

关于http接口的详细介绍参考下面集成文档http服务章节的相关内容

集成文档

使用该方式,将运行库嵌入到开发者的程序当中。 编译demo项目 SDK src目录下有完整的demo工程,用户可参考该工程的代码实现方式将SDK 集成到自己的项目中,demo工程可直接编译运行:

cd src mkdir build && cd build cmake .. && make ./easyedge_image_inference {模型RES文件夹} {测试图片路径} **如果是NNIE引擎,使用sudo运行** sudo ./easyedge_image_inference {模型RES文件夹} {测试图片路径}

(可选) SDK包内一般自带opencv库,可忽略该步骤。如果希望SDK自动编译安装所需要的OpenCV库,修改cmake的 optionEDGE_BUILD_OPENCV为ON即可。 SDK会自动从网络下载opencv源码,并编译需要的module、链接。注意,此功能必须需联网。

cmake -DEDGE_BUILD_OPENCV=ON .. && make -j16

若需自定义library search path或者gcc路径,修改CMakeList.txt即可。

对于硬件使用为Intel Movidius MyRIAD2 / MyRIAD X 的,如果宿主机找不到神经计算棒Intel® Neural Compute Stick,需要执行以下命令添加 USB Rules:

cp \${cpp_kit位置路径}/thirdparty/openvino/deployment_tools/inference_engine/external/97-myriad-usbboot.rules /etc/udev/rules.d/ sudo udevadm control --reload-rules sudo udevadm trigger sudo ldconfig

使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

// step 1: 配置运行参数

EdgePredictorConfig config config.model_dir = {模型文件目录}; // step 2: 创建并初始化Predictor; 这这里选择合适的引擎 auto predictor = global_controller()->CreateEdgePredictor(config); // step 3-1: 预测图像 auto img = cv::imread({图片路径}); std::vector<EdgeResultData> results; predictor->infer(img, results); // step 3-2: 预测视频 std::vector<EdgeResultData> results; FrameTensor frame_tensor; VideoConfig video_config; video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h video_config.source_value = video_src; /* ... more video_configs, 根据需要配置video_config的各选项 */ auto video_decoding = CreateVideoDecoding(video_config); while (video_decoding->next(frame_tensor) == EDGE_OK) { results.clear(); if (frame_tensor.is_needed) { predictor->infer(frame_tensor.frame, results); render(frame_tensor.frame, results, predictor->model_info().kind); //video_decoding->display(frame_tensor); // 显示当前frame,需在video_config中开启配置 //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置

对于口罩检测模型,将 EdgePredictorConfig config修改为PaddleMultiStageConfig config即可。

口罩检测模型请注意输入图片中人脸大小建议保持在 88至19696像素之间,可根据场景远近程度缩放图片后再传入SDK。

SDK参数配置 SDK的参数通过EdgePredictorConfig::set_config和global_controller()->set_config配置。set_config的所有key 在easyedge_xxxx_config.h中。其中

- PREDICTOR前缀的key是不同模型相关的配置,通过EdgePredictorConfig::set_config设置
- CONTROLLER前缀的key是整个SDK的全局配置,通过global_controller()->set_config设置

以序列号为例,KEY的说明如下:

```
/**
* @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号
* 值类型:string
* 默认值:空
*/
static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";
```

```
使用方法如下:
```

```
EdgePredictorConfig config;
config.model_dir = ...;
config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, "1DB7-1111-1111-D27D");
```

具体支持的运行参数可以参考开发工具包中的头文件的详细说明。

初始化

接口

```
auto predictor = global_controller()->CreateEdgePredictor(config);
predictor->init();
```

若返回非0,请查看输出日志排查错误原因。

预测图像

```
• 接口
```

```
/**
* @brief
* 通用接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
    cv::Mat& image, std::vector<EdgeResultData>& result
) = 0;
```

图片的格式务必为opencv默认的BGR, HWC格式。

• 返回格式

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index: // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测活图像分割时才有
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割时才有
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding, 游程编码的mask
};
```

关于矩形坐标

```
x1 * 图片宽度 = 检测框的左上角的横坐标
```

- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2*图片宽度 = 检测框的右下角的横坐标
- y2 * 图片高度 = 检测框的右下角的纵坐标

关于图像分割mask

```
cv::Mat mask为图像掩码的二维数组
{
{0,0,0,0,0,0,0,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,0,0,0,0,0,0,0},
}
其中1代表为目标区域,0代表非目标区域
```

关于图像分割mask_rle

该字段返回了mask的游程编码,解析方式可参考 http demo

Baidu 百度智能云文档

预测视频

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

● 接口

classVideoDecoding :

```
/**
* @brief 获取输入源的下一帧
* @param frame_tensor
* @return
*/
virtual int next(FrameTensor &frame_tensor) = 0;
/**
* @brief 显示当前frame_tensor中的视频帧
* @param frame_tensor
* @return
*/
virtual int display(const FrameTensor &frame_tensor) = 0;
/**
* @brief 将当前frame_tensor中的视频帧写为本地视频文件
* @param frame_tensor
* @return
*/
virtual int save(FrameTensor &frame_tensor) = 0;
/**
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
 * @brief 获取视频的width属性
 * @return
*/
virtual int get_width() = 0;
/**
* @brief 获取视频的height属性
* @return
*/
```

virtual int get_height() = 0;

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type;
                          // 输入源类型
  std::string source_value;
                         // 输入源地址, 如视频文件路径、摄像头index、网络流地址
 int skip_frames{0};
                       // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
  int retrieve_all{false};
                     // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
  int input_fps{0};
                       // 在采取抽帧之前设置视频的fps
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
  bool enable_display{false};
  std::string window_name{"EasyEdge"};
  bool display_all{false};
                        // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                        // 是否存储所有frame, 若为false, 仅存储根据skip_frames抽取的frame
  std::map<std::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被 抽取返回,以作为显示或存储用。input_fps:用于抽帧前设置fps。resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。conf:高级选项。部分配置会通过该map来设置。

注意: 1.如果使用 VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。 2.使用摄像头抽帧时,如 果通过 resolution 设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

设置序列号

请在网页控制台中申请序列号,并在init初始化前设置。 LinuxSDK 首次使用需联网授权。

EdgePredictorConfig config; config.set_config(easyedge::params::PREDICTOR_KEY_SERIAL_NUM, "this-is-serial-num");

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

```
EdgeLogConfig log_config;
log_config.enable_debug = true;
global_controller()->set_log_config(log_config);
```

http服务

1. 开启http服务 http服务的启动参考demo_serving.cpp文件。

/** * @brief 开启一个简单的demo http服务。 * 该方法会block直到收到sigint/sigterm。 * http服务里,图片的解码运行在cpu之上,可能会降低推理速度。 * @tparam ConfigT * @param config * @param host * @param port * @param service_id service_id user parameter, uri '/get/service_id' will respond this value with 'text/plain' * @param instance_num 实例数量,根据内存/显存/时延要求调整 * @return */ template<typename ConfigT> int start_http_server(const ConfigT &config, const std::string &host int port,

2. http接口详细说明 http 请求方式一:无额外编码 URL中的get参数:

const std::string &service_id, int instance_num = 1);

参数	说明		默认值	
threshold	阈值过滤,	0~1	如不提供,	则会使用模型的推荐阈值

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

Python请求示例 (图片测试,针对图像分类、物体检测、实例分割等模型)

import requests

with open('./1.jpg', 'rb') as f: img = f.read() result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1}, data=img).json()

Python请求示例 (图片测试, 仅针对语义分割模型, 同其他CV模型不同, 语义分割模型输出为灰度图)

import requests

```
with open('./1.jpg', 'rb') as f:
    img_data = f.read()
    res = requests.post('http://127.0.0.1:24401/',
        data=img_data)
    with open("gray_result.png", "wb") as fb:
        fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
```

Python请求示例 (视频测试, 注意:区别于图片预测,需指定Content-Type;否则会调用图片推理接口)

import requests

with open('./1.mp4', 'rb') as f: video_data = f.read() result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1}, headers={'Content-Type': 'video'}, data=video_data).json()

http 请求方法二:json格式,图片传base64格式字符串 HTTP方法:POST Header如下:

参数

值 Content-Type application/json

Body请求填写:

• 图像分类网络: body 中请求示例

```
{
"image": "<base64数据>",
"top_num": 5
}
```

body中参数详情

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
top_nu m	否	number	-	返回分类数量,不填该参数,则默认返回全部分类结果

• 物体检测和实例分割网络: Body请求示例:

```
{
"image": "<base64数据>",
 "threshold": 0.3
}
```

body中参数详情:

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
thresho Id	否	number	-	默认为推荐阈值,也可自行根据需要进行设置

• 语义分割网络: Body请求示例:

```
{
"image": "<base64数据>"
}
```

body中参数详情(语义分割由于模型特殊性,不支持设置threshold值,设置了也没有意义):

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部

Python请求示例(非语义分割模型参考如下代码)

import base64

import requests
def main():
with open("1.jpg 【图片路径】", 'rb') as f:
result = requests.post("http://{服务ip地址}:24401/", json={
"image": base64.b64encode(f.read()).decode("utf8")
})
print(result.request.body)
print(result.request.headers)
print(result.content)
ifname == 'main':
main()

Python 请求示例 (针对语义分割模型,同其他CV模型不同,语义分割模型输出为灰度图)

```
import base64
import requests
def main():
    with open("1.jpg 【图片路径】", 'rb') as f:
        res = requests.post("http://{服务ip地址}:24401/", json={"image": base64.b64encode(f.read()).decode("utf8")})
        with open("gray_result.png", "wb") as fb:
        fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
iff __name__ == '__main__':
        main()
```

http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考预测图像-返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

返回示例

```
{
    "cost_ms": 52,
    "error_code": 0,
    "results": [
        {
            "confidence": 0.94482421875,
            "index": 1,
            "label": "IronMan",
            "x1": 0.059185408055782318,
            "x2": 0.18795496225357056,
            "y1": 0.14762254059314728,
            "y2": 0.52510076761245728,
            "mask": "...", // 图像分割模型字段
            "trackId": 0, // 目标追踪模型字段
        },
    ]
}
```

其他配置

1. 日志名称、HTTP 网页标题设置

通过global_controller的set_config方法设置:

效果如下:

图片加载失败

FAQ

1. 如何处理一些 undefined reference / error while loading shared libraries?

如:./easyedge_demo: error while loading shared libraries: libeasyedge.so.1: cannot open shared object file: No such file or directory 这是因为二进制运行时Id无法找到依赖的库。如果是正确cmake && make 的程序,会自动处理好链接,一般不会出现此类问题。

遇到该问题时,请找到具体的库的位置,设置LD_LIBRARY_PATH。

示例一:libverify.so.1: cannot open shared object file: No such file or directory 链接找不到libveirfy.so文件,一般可通过 export LD_LIBRARY_PATH=\${LD_LIBRARY_PATH}:../../lib 解决(实际冒号后面添加的路径以libverify.so文件所在的路径为准)

示例二:libopencv_videoio.so.4.5: cannot open shared object file: No such file or directory 链接找不到ibopencv_videoio.so文件,一般可通 过 export LD_LIBRARY_PATH=\${LD_LIBRARY_PATH}:../../thirdparty/opencv/lib 解决(实际冒号后面添加的路径以libopencv_videoio.so所在路 径为准)

2. EasyDL SDK与云服务效果不一致,如何处理?

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

3. 如何将我的模型运行为一个http服务?目前cpp sdk暂未集成http运行方式; 0.4.7版本之后,可以通过start_http_server方法开启http服务。

4. 运行NNIE引擎报permission denied 日志显示:

open sys: Permission denied open err : Permission denied open err : Permission denied

请使用sudo在root下运行。

5. 运行SDK报错 Authorization failed

情况一:日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888"
./easyedge_demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限 于)以下可能的情况:

- MAC地址变化
- 磁盘变更
- BIOS重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 ~/.baidu/easyedge 目录,再重新激活。

6. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

7. 运行NNIE引擎报错 std::bad_alloc 检查开发板可用内存,一些比较大的网络占用内存较多,推荐内存500M以上

8. 运行二进制时,提示 libverify.so cannot open shared object file

可能cmake没有正确设置rpath,可以设置LD_LIBRARY_PATH为sdk的lib文件夹后,再运行:

LD_LIBRARY_PATH=\$LD_LIBRARY_PATH:../lib ./easyedge_demo

9. 编译时报错:file format not recognized 可能是因为在复制SDK时文件信息丢失。请将整个压缩包复制到目标设备中,再解压缩、编译

の Linux集成文档-Python

简介

本文档介绍 EasyDL 的 Linux Python SDK 的使用方法,适用于 EasyDL 通用版和BML。

- 网络类型支持:图像分类,物体检测
- 硬件支持:
 - Intel Movidius MyRIAD2 / MyRIAD X / IGPU
 - 瑞芯微 RK3399Pro
- 语言支持: Intel Movidius MyRIAD2 / MyRIAD X / IGPU: Python 3.5, 3.6, 3.7 瑞芯微 RK3399Pro: Python 3.6

Release Notes

时间 医木 沿明

ոյլոյ	瓜牛	100-97
2022.10. 27	1.3.5	新增Armv7 CPU、Armv8 CPU、Jetson、华为昇腾Atlas开发板对应Python SDK,支持图像分类、物体检测、人脸检测、实例 分割;新增 Intel IGPU 支持
2022.05. 18	1.3.0	新增RK3399Pro NPU对应Python SDK,支持图像分类、物体检测
2021.01. 27	1.2.1	EasyDL经典版分类高性能模型升级;支持更多模型
2020.12. 18	1.2.0	推理引擎升级;接口升级;性能优化
2020.09. 17	1.1.19	支持更多模型
2020.08. 11	1.1.18	性能优化
2020.06. 23	1.1.17	支持更多EasyDL专业版模型
2020.04. 16	1.1.15	技术优化;升级 OpenVINO 版本
2020.01. 16	1.1.12	预测函数默认使用推荐阈值
2019.12. 26	1.1.11	EasyDL 专业版支持加速棒
2019.10. 21	1.1.9	支持 EasyDL 专业版
2019.07. 19	1.1.7	提供模型更新工具
2019.03. 15	1.1.0	架构与功能完善
2019.02. 28	1.0.6	引擎功能完善
2019.02. 13	1.0.5	paddlepaddle 支持
2018.11. 30	1.0.0	第一版!

2020-12-18: 【接口升级】 序列号配置接口从1.2.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请 尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

快速开始

1. 安装依赖

根据引擎的不同,SDK 依赖了不同的底层引擎。根据所需自行安装。

依赖说明:Intel Movidius 加速棒 使用Intel Movidius加速棒 SDK、Intel IGPU 预测时,必须安装 OpenVINO 预测引擎,两种方式:

- 使用 OpenVINO[™] toolkit 安装,请参考 OpenVINO toolkit 文档安装 2020.3.1LTS(必须)版本, 安装时可忽略Configure the Model Optimizer及 后续部分。
- 使用源码编译安装,请参考 Openvino Inference Engine文档编译安装 2020.3.1 (必须)版本。

安装完毕,运行之前,请按照OpenVinono的文档 设置环境变量

source /opt/intel/openvino/bin/setupvars.sh

依赖说明:RK3399Pro 所有用例基于 Npu driver版本1.7.3的RK3399pro开发板测试通过 查看RK3399Pro板上driver版本方法:运行sdk内提供

demo项目,日志里会提供API和Driver版本信息

2022-12-20) 14:26:07,765 VERBOSE [EasyEdge] [rockchip_edge_predictor.cpp:87] 547887054864 Create predictor , 5029536
D RKNNAPI:	
D RKNNAPI:	RKNN VERSION:
D RKNNAPI:	API: 1.7.3 (0cfd4a1 build: 2022-08-15 17:10:10)
D RKNNAPI:	DRV: 1.7.3 (c4ea832 build: 2022-08-13 09:13:08)
D RKNNAPI:	

升级399Pro driver版本参考瑞芯微github: https://github.com/airockchip/RK3399Pro_npu 2. 安装 easyedge python wheel 包 安装说明: Intel

Movidius 加速棒 / Intel IGPU

pip3 install -U BaiduAI_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_x86_64.whl

具体名称以 SDK 包中的 whl 为准。

安装说明:RK3399Pro

pip3 install -U BaiduAl_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_aarch64.whl pip3 install -U EasyEdge_Devkit_RK3399Pro-{版本号}-cp36-cp36m-linux_aarch64.whl

具体名称以 SDK 包中的 whl 为准,特别注意这里要同时安装两个whl包 安装说明:ArmV7 CPU

pip3 install -U BaiduAl_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_armv7l.whl pip3 install -U EasyEdge_Devkit_ARM-{版本号}-cp36-cp36m-linux_armv7l.whl

安装说明:ArmV8 CPU (Aarch64 CPU)

pip3 install -U BaiduAl_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_aarch64.whl pip3 install -U EasyEdge_Devkit_ARM-{版本号}-cp36-cp36m-linux_aarch64.whl

安装说明: Jetson SDK

pip3 install -U BaiduAI_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_aarch64.whl pip3 install -U EasyEdge_Devkit_JetPack{版本号}-{版本号}-cp36-cp36m-linux_aarch64.whl

安装说明:华为昇腾Atlas开发板

pip3 install -U BaiduAl_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_aarch64.whl EasyEdge_Devkit_Atlas200-{版本号}-cp36-cp36m-linux_aarch64.whl

3. 使用序列号激活

获取序列号	安美林振身状却 文秀美林振身、村川振気成的度型部署在本地。高鉄現用模型。可以选择种模型部署在本地的服务器、小型设备、软硬一件方案等项温配硬件上、 活びみ、50以一一型風点、展活面不得业务局展。 東京市市場局 聖田谷 「「「」」」」 「「」」」 「「」」」 「」」							
	模型名称	发布版本 🗘	应用平台	模型加速	发布状态	发布时间 🗘		
		134318V1 查看性能报告	通用XB6 CPU-Linux	基础板	 已发布 	2021-08-19 20:24	TUSDK	
				精度无损压缩加速	 已发布 	2021-08-19 20:24	下载加速版SDK	
			英伟达GPU-Linux	基础板	 已发布 	2021-08-19 20:35	下観SDK	
	sun_()r⊟ #ritest			精度无损压缩加速	 已发布 	2021-08-19 20:34	下载加速版SDK	
			WARLEY'S CONTRACTOR	基础版	 已发布 	2021-08-19 18:17	下载SDK	

修改demo.py 填写序列号

```
pred = edge.Program()
pred.set_auth_license_key("这里填写序列号")
```

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己 的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的设置),需要调用函数指定实例数鉴权模式,并且实例数鉴权模式下,支持指定 license证书更新时间,单位是秒,要求设置为大于20的整数,否则会采用默认的license更新时间,参考 pred.set_instance_auth_mode()
pred.set_instance_update_interval(200)

4. 测试 Demo

输入对应的模型文件夹(默认为RES)和测试图片路径,运行:

python3 demo.py {model_dir} {image_name.jpg}

测试效果:



图片加载失败

使用说明

使用流程

import BaiduAI.EasyEdge as edge

```
pred = edge.Program()
pred.set_auth_license_key("这里填写序列号")
pred.init(model_dir={RES文件夹路径}, device=edge.Device.MOVIDIUS, engine=edge.Engine.OPENVINO)
pred.infer_image({numpy.ndarray的图片})
pred.close()
```

初始化

```
• 接口
```

```
def init(self,
    model_dir,
    device=Device.LOCAL,
    engine=Engine.PADDLE_FLUID,
    config_file='conf.json',
    preprocess_file='preprocess_args.json',
    model_file='model',
    params_file='params',
    graph_file='graph.ncsmodel',
    label_file='label_list.txt',
    device_id=0
    ):
.....
Args:
   device: Device.CPU
   engine: Engine.PADDLE_FLUID
   model_dir: str
     model dir
   preprocess_file: str
   model_file: str
   params_file: str
   graph_file: str
   label_file: str
   device_id: int
Raises:
   RuntimeError, IOError
Returns:
   bool: True if success
 .....
```

● 接口

<pre>def infer_image(self, img,</pre>
Args:
img: np.ndarray or bytes
threshold: float
only return result with confidence larger than threshold
channel order: string
channel_order. String
color format order RGB or BGR
data_type: string
image data type
Returns:
list

• 返回格式: [dict1, dict2, ...]

字段	类型	取值	说明
confidence	float	0~1	分类或检测的置信度
label	string		分类或检测的类别
index	number		分类或检测的类别
x1, y1	float	0~1	物体检测,矩形的左上角坐标 (相对长宽的比例值)
x2, y2	float	0~1	物体检测,矩形的右下角坐标(相对长宽的比例值)
mask	string/numpy.ndarray	图像分割的mask	

关于矩形坐标

```
x1 * 图片宽度 = 检测框的左上角的横坐标
```

- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2 * 图片宽度 = 检测框的右下角的横坐标
- y2 * 图片高度 = 检测框的右下角的纵坐标

可以参考 demo 文件中使用 opencv 绘制矩形的逻辑。

结果示例

• i) 图像分类

```
{
    "index": 736,
    "label": "table",
    "confidence": 0.9
}
```

● ii) 物体检测

Baidu 百度智能云文档

```
"y2": 0.91211,
"label": "cat",
"confidence": 1.0,
"x2": 0.91504,
"index": 8,
"y1": 0.12671,
"x1": 0.21289
}
```

• iii) 图像分割

```
{
    "name": "cat",
    "score": 1.0,
    "location": {
        "left": ...,
        "top": ...,
        "width": ...,
        "height": ...,
    },
    "mask": ...
}
```

mask字段中,data_type为numpy时,返回图像掩码的二维数组

```
{
  {
  {
  {
  {
  {
  0.0.0.0.0.0.0.0.0.0.0.0},
  }
  {
  0.0.0.1.1.1.0.0.0.0},
  {
  0.0.0.1.1.1.0.0.0.0},
  {
  0.0.0.1.1.1.0.0.0.0},
  {
  0.0.0.0.0.0.0.0},
  {
  0.0.0.0.0.0.0.0},
  }
  其中1代表为目标区域,0代表非目标区域,0代表非目标区域
  }
}
```

data_type为string时,mask的游程编码,解析方式可参考 demo

升级模型 适用于经典版升级模型,执行bash update_model.sh,根据提示,输入模型路径、激活码、模型ID、模型版本,等待模型更新完毕即 可。

FAQ

Q: 运行SDK报错 Authorization failed

情况一: 日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888"
./easyedge_demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- MAC地址变化
- 磁盘变更
- BIOS重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除~/.baidu/easyedge 目录,再重新激活。

情况三: ArmV7、ArmV8 CPU、Jetson、Atlas Python SDK日志提示ImportError: libavformat.so.58: cannot open shared object file: No such file or

directory 或者其他类似so找不到可以在LD_LIBRARY_PATH环境变量加上libs和thirdpartylibs路径,例如

export LD_LIBRARY_PATH=/xxx/libs:/xxx/thirdpartylibs:\$LD_LIBRARY_PATH # tips: 这里/xxx需要替换为真实路径,/xxx路径查找方法如下

查找安装包内libs和thirdpartylibs路径的方法如下(以Atlas SDK为例,其他SDK查找方法类似):

pip3 show EasyEdge-Devkit-Atlas200 # 结果中会显示 Location 路径,也就是包的安装路径 **libs和thirdpartylibs两个路径在 Location 所指示的路径 easyedge_CANN 子文件夹下**

心 Linux集成文档-Atlas

简介

本文档介绍EasyEdge/EasyDL的Linux Atlas SDK的使用方法。

注意Atlas有两种产品形态,Atlas 200和Atlas 300,请参见此处的文档说明

- 网络类型支持:图像分类
- 硬件支持:
 - CPU: aarch64
 - Atlas 200 卡
- 操作系统支持:Atlas指定的Linux版本, Ubuntu 16.04 aarch64 ,请从Atlas文档中下载。

性能数据

数据仅供参考,实际数值根据使用线程数、利用率等情况可能有所波动

模型类型	模型算法	芯片类型	SDK类型	实测硬件	单次预测耗时
EasyDL 图像分类	高性能	Atlas 200	Atlas 200	Atlas 200DK	9ms
EasyDL 图像分类	高精度	Atlas 200	Atlas 200	Atlas 200DK	12ms
EasyDL 物体检测	高性能	Atlas 200	Atlas 200	Atlas 200DK	11ms
EasyDL 物体检测	高精度	Atlas 200	Atlas 200	Atlas 200DK	31ms

Release Notes

时间	版本	说明
2020.6.15	0.2	支持物体检测
2020.3.10	0.1	初始版本,支持图像分类

测试atlas 200的官方demo

请参见此处的文档说明, 搭建开发环境,测试atlas 200的mindstudio demo通过后,再测试

快速开始

SDK在以下环境中测试通过

- ubuntu 16.04, aarch64-linux-gnu-g++ 5.4 , 编译机
- ubuntu 16.04 , 开发板

Atlas DDK 的ddk_info信息:

```
{
    "VERSION": "1.3.T34.B891",
    "NAME": "DDK",
    "TARGET": "Atlas DK"
}
```

2. 测试Demo

编译运行: 下载后,模型资源文件默认已经打包在开发者下载的SDK包中,

Step 1: 运行一次unpack.sh脚本,会得到测试demo。

Step 2:请在官网获取序列号,填写在demo_async.cpp及demo_sync.cpp的开始处license_key字段。



图片加载失败

step3: 准备测试图片

覆盖image目录下的 1.jpg,更多图片可以用于demo中的批量测试模式

step4: 修改test_200.sh下的以下开发板登录信息

export DDK_PATH=\$HOME/tools/che/ddk/ddk # ddk的安装路径

SSH_USER=HwHiAiUser@192.168.3.25 # 200 开发板的ssh登录信息 PORT=8822 # 200 开发板的ssh登录端口

step: 运行demo, 会自动编译OpenCV 3.4库

cd demo sh test_200.sh

图像分类的demo运行效果:

```
[stat] [100001]image/1.jpg(4 images) time used: 41ms (at 1583765958531) total:705ms
[result][100001][image/1.jpg][281470472005664] is: n07747607 orange 0.973633 950;
```

n07747607 orange 分类名 0.973633 分类概率 950 分类名的序号

物体检测的demo运行效果:

```
[stat] time used : 101ms; all time used:478
images[3] result:
label:no2_ynen;prob:0.985352 loc:[(0.459961,0.839844), (0.5625,0.988281)]
no2_ynen 分类名 , 也可以获取分类名的序号
0.985352 分类概率
loc:[(0.459961,0.839844), (0.5625,0.988281)] , 检测框的位置。(0.459961,0.839844)表示左上角的点,(0.5625,0.988281)右下角的
点;
如原始图片608 , 左上角[0.459961*608,0.839844*608) , 右下角(0.5625*608,0.988281*608)
```

SDK接口使用

使用该方式,将运行库嵌入到开发者的程序当中。

同步接口使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

// step 0: 设置序列号
global_controller()->set_licence_key("set your license here");

// step 1: 配置模型资源目录 AtlasConfig config; config.model_dir = {模型文件目录};

// step 2: 创建并初始化Predictor ;
auto predictor = global_controller()->CreateEdgePredictor(config);
int ret = predictor->init();
若返回非0,请查看输出日志排查错误原因。
auto img = cv::imread((图片路径));
// step 3: 预测图像
std::vector<EdgeResultData> result2;
predictor->infer(img, result2);
解析result2即可获取结果

异步接口使用流程

```
// step 0: 设置序列号
global_controller()->set_licence_key("set your license here");
```

```
// step 1: 配置模型资源目录
AtlasConfig config;
config.model_dir = {模型文件目录};
```

// step 3: 创建Predictor;这这里选择合适的引擎 auto predictor = global_controller()->CreateEdgePredictor(config);

// step 4: 设置异步回调 predictor->set_result_handler(YOUR_HANDLER);

```
// step 5: 初始化
int ret = predictor->init();
**若返回非0,请查看输出日志排查错误原因。**
```

// step 6: 预测图像 auto img = cv::imread({图片路径}); color_format = kBGR; float threshold = 0.1;

```
uint64_t seq_id;
predictor->infer_async(img, color_format, 0.1, nullptr, seq_id);
**YOUR_HANDLER里面有seq_id的回调结果**
```

设置序列号

请在网页控制台中申请序列号,并在init初始化前设置。 LinuxSDK 首次使用需联网授权。

● 接口

virtual int set_licence_key(const std::string& license) = 0;

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

```
EdgeLogConfig log_config;
log_config.enable_debug = true;
global_controller()->set_log_config(log_config);
```

FAQ

1. 如何处理一些 undefined reference?

如: undefined reference to `curl_easy_setopt@CURL_OPENSSL_3'

方案1:通过安装libcurl3 libcurl-openssl1.0-dev来解决。 方案2:如果开发者想不想使用低版本的openssl (如Ubuntu 18.04),可以link静态 库easyedge_static.a,自己指定需要的Library的版本:

示例:修改CMakeList.txt

find_package(CURL REQUIRED) target_link_libraries(easyedge_demo \${OpenCV_LIBS} easyedge_static pthread \${CURL_LIBRARIES} verify_static \${其他需要的库})

其中,其他需要的库视具体sdk中包含的库而定。

2. EasyDL 离线SDK与云服务效果不一致,如何处理?

目前离线SDK与云服务的处理有些许差异,具体如下:

- 图像分类模型:离线SDK与云服务使用通用(非快速训练、非AutoDL Transfer)的效果类似
- 物体检测模型:离线SDK的高精度模型与云服务的精度较低,服务性能更佳的效果类似

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

心 端云协同服务说明

服务简介

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

「云管理,端计算」的端云协同服务,具体包括:

- 在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新
- 断网状态下模型离线计算(http服务,可调用与公有云API功能相同的接口)
- 联网状态下在平台管理设备运行状态、资源利用率

目前通用小型设备的应用平台支持Linux-ARM,具体使用流程请参考下方文档。

使用流程

Step 1 发布端云协同部署包

在我的部署包页面点击「发布端云协同部署包」

侯空中心	端云协同服务 > 我的部署包							
我的模型	端云协同服务说明							点击收
创建模型	1、在可视化界面轻松实现模型部	署包在边缘设备.	上的集成、版本更	ព				
训练模型	 2、断网状态下模型离线计算(ht 3、联网状态下在平台管理设备运 	tp服务,可调用 行状态、资源利	与公有云API功能标 用率	目同的接口)				
校验模型	具体使用流程如下:							
发布模型	 在本页面发布端云协同部署包 在「我的本地设备」页面新 	,填写相关信息: ^{例设备} 复制游	并申请,并在下方: 5念今 联网激活:	列表宣看部署包发有 ttm设备	状态			
EasyData数据服务	3、在「下发部署包到设备」页	面,将发布成功的	的部署包一键下发	到已激活的设备上,	即可测试或正式使用			
数据总览	发布端云协同部署包							
标签组管理								
在线标注	端云协同服务名称	模型ID	设备类型	应用平台	当前可用版本	新版本发布状态	部署设备数	操作
云服务数据回流								
EasyEdge本地部署								
纯离线服务								
端云协同服务 へ								
我的部署包					暂无可用数	据		
					请稍后再试			

填写服务名称,选择模型版本并提交发布

模型中心	端云协同服务 > 发布端云协同部署包		
我的模型	设备类型 🔷 服务器 🔘 遥	用小型设备	
创建模型			
训练模型	模型省标 test2021	~	
校验模型	端云协同服务名称		
发布模型	选择版本 V1	~	
EasyData数据服务	选择系统和芯片 🔄 🛆 Linux		
数据总览	通用ARM		
标签组管理			
在线标注	发布部署包		
云服务数据回流			
EasyEdge本地部署			
纯离线服务			

在列表查看部署包发布状态

補助する	
候型中心	端云协同服务 > 我的部署包
我的模型	端云协问服务说明
创建模型	1、在可视化界面轻松突现模型部署包在边缘设备上的集成、版本更新
训练模型	2、断网状态下模型黑线计算(http:服务,可调用与公布云API功能相同的接口) 3、联网状态下在平台管理设备运行状态、资源利用率
校验模型	具体使用流程如下:
发布模型	 在本页面发布瑞云协同影響包,填写相关信息并申请,并在下方列表置看影響包发布状态 在「我約本地设备」页面前增设备,复则激活命令,联网激活本地设备
EasyData数据服务	3、在「下发邮署包到设备」页面,将发布成功的邮署包一键下发到已激活的设备上,即可测试或正式使用
数据总览	发布施云协问部署包
标签组管理	
在线标注	端云协同服务名称 模型ID 设备类型 应用平台 当前可用版本 新版本发布状态 部署设备数 操作
	ecc 264 通用小型设备 Linux-通用ARM V1 V1-已发布 0 下发到设备 发布新版本
<u> </u>	从发布页过来的 246 服务器 Linux-通用X86 CPU V2, V1 V2-已发布 1 下发到设备发布新版本 服务详情
EasyEdge本地部署	图像分类高精度_ 猫狗-265 265 通用小型设备 Linux-通用ARM - V1-发布中 0
纯离线服务	
端云协同服务 へ	每页显示 10 🗸 < 1 >
我的部署包	
我的本地设备	

Step 2 新增设备并激活

在我的本地设备页面新增设备

模型中心	端云协同服务 > 我的]本地设备					
我的模型	我的本地设备						点击收起
创建模型	在本页面新增设备、联网 利用率等信息。	激活本地设备后,即可料	将「 我的部署包 」页面发	这布成功的部署包一键下	发到设备上。设备联网时,可以还	查看设备上部署的服务、设备的; ;	运行状态、资源
训练模型							
校验模型	新增设备						
发布模型							
EasyData数据服务	设备名称	创建时间	设备类型	应用平台	设备连接状态	最新同步时间	操作
数据总览							
标签组管理							

•		新增设备	×
	设备名称	20个字符以内	
	备注信息	50个字符以内, 自定义备注信息	
c		0/50	
ŧ			
		确认 取消	

在列表中,点击设备对应的「激活设备」操作,复制激活命令并在本地设备上执行即可

图	查看激活命令	×		收起 へ
在往	复制命令,并在本地设备上执行,即可激活设备			
端云协同服务	curl -skfL 'https://verify.baidubce.com/act sh?token=e177df4ed47b2265223a333630302c226b2	ive/v1/active/setup.		
我的本地设备 在本页面新增设	6e223a22353730313835302d382d65393866386383 12d373138392d376266663366643063313433222c22 6c2d636c6f7564222c227473223a313630393939393	02d343231612d6537653 5e73223a226261657479 434327d' -osetup.sh	可以查看设备上部署的服务、设备的运行状	点击收起 态、资源
利用华夺信息。	éé sh setup.sh			
新增设备				
设备名称	知道了		寸间 操作	
linux x86 🙂	2021-01-07 14:02	木淑石 -	删版 激活设备 <mark>算</mark> 看详情	

Step 3 下发部署包到设备,在本地调用

在下发部署包到设备页面,将发布成功的部署包一键下发到已激活的设备上,即可测试或正式使用

模型中心	端云协同服务 > 下发部署包到设备					
我的模型	将已成功发布的端云协共服务部署包一键下发到已激活且在连接中的设备					
创建模型						
训练模型	服务名称	ecc	\sim			
校验模型	模型ID	264				
发布模型	应用平台	Linux-通用ARM	\sim			
EasyData数据服务						
数据总览						
标签组管理						

部署包下发成功之后,会在本地启动一个HTTP推理服务。在浏览器中输入http://{设备ip}:{服务端口,默认8080},即可预览效果:

当前阈值: 0.07 修改

置信度 0.998

0.937

0.841

0.781

0,408

0.324

0.251

▲ 不安全 yq01-ns-map0402a659.yq01:8081



具体接口调用说明请参考文档 SDK - HTTP服务调用说明

云端管理说明

模型部署包管理

在我的部署包页面可以进行已发布的模型部署包的管理。

发布及更新模型版本

点击「发布新版本」操作即可快速发布对应模型ID下的新版本。同一模型ID下已发布的模型版本均会显示在列表的「当前可用版本」中。

模型中心	s	端云协同服务 > 我的部署包	1								
我的模型		端云协同服务说明									点击收起
创建模型		1、在可視化界面轻松实现模型部	署包在边线	缘设备上的集成、 	版本更新						
训练模型		 2、町网状态下模型高线订算(nt) 3、联网状态下在平台管理设备运 	tp服务, 行状态、	可调用与公有云A 资源利用率	(PIJ)能怕问的按口)						
校验模型		具体使用流程如下:									
发布模型		 在本页面发布端云协同部署包 2、在「我的本地设备」页面新: 	,填写相: 增设备,1	关信息并申请,扌 夏制激活命令,彫	‡在下方列表宣看部署包发 K网激活本地设备	发布状态					
EasyData	a数据服务	3、在「下发部署包到设备」页	面,将发7	^{午成功的部署包一}	-键下发到已激活的设备上	1,即可测试或正	式使用				
数据总览		发布端云协同部署包									
标签组管理	Ŧ										
在线标注		端云协同服务名称	模型ID	设备类型	应用平台	当前可用版本	新版本发布状态	部署设备数	操作		
云服务数据	医回流	ecc	264	通用小型设备	Linux-通用ARM	V1	V1-已发布	0	下发到设备	发布新版本	
EasyEdg	le本地部署		246	服务器	Linux-通用X86 CPU	V2, V1	V2-已发布	1	下发到设备	· 发布新版本	服务详情
纯离线服务	K 5	100 M 100 M 100	265	通用小型设备	Linux-通用ARM	-	V1-发布中	0			
端云协同朋	服务 へ								每页显示	10 🗸 <	1 >
我的部	著包										
我的本	地设备										
下发部	署包到设备										
_	_								_		
			发布	5新版本	Ţ.			>			
*	将最新训练的棒	莫型版本发布为服务,	发布质	成功后,艮	「可从云端下发	到设备					
		00 47 47 55									
		服务省称 60	CC								
		模型ID 20	64								
		选择新版本	V1			\sim					
		6	角认	1	取消						

新版本发布成功后,即可在「下发部署包到设备」页面或当前服务的「服务详情」页面,将新版本下发到本地设备上。

模型中心	端云协同服务 > 我的部署包
我的模型	端云协同服务说明 点击收款
创建模型	1、在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新
训练模型	 断网状态下模型离线计算(http服务,可调用与公有云API功能相同的接口) 联网状态下在平台管理设备运行状态、资源利用率
校验模型	具体使用流程如下:
发布模型	1、在本页面发布端云协同部署包,填写相关信息并申请,并在下方列表查看部署包发布状态
FasyData数据服务	2、任·按約本型比會」贝圓兩項设备,呈刺激活命で、時期激活命和設备 3、在「下发部署包到设备」页面,将发布成功的部署包一键下发到已激活的设备上,即可测试或正式使用

RX 306 /25 342	发布端云协问部署包
标签组管理	端云协同服务名称 模型ID 设备类型 应用平台 当前可用版本 新版本发布状态 部署设备数 操作
在线标注	ecc 264 通用小型设备 Linux-通用ARM V1 V1-已发布 0 下发到设备 发布新版本
云服务数据回流	246 服务器 Linux-通用X86 CPU V2, V1 V2-已发布 1 下发到设备发布新版本 服务详情
EasyEdge本地部署	图像分类高精度_磁狗-265 265 通用小型设备 Linux-通用ARM - V1-发布中 0
纯离线服务	
模型中心	端云协同服务 >
我的模型	
创建模型	
训练模型	
校验模型	下发到更多设备
发布模型	设备名称 最新下发模型版本 部署到期时间 设备连接状态 服务下发状态 最新同步时间 操作
EasyData数据服务	ubuntu-local-fly ⊖ V2 2021-02-06 在线 V2下发成功 2021-01-07 14:45 宣看服务配置 重看设备详情 下发新版本 移脉设备
数据总简	
经济的 第	
你亚坦管理	
在线标注	

管理模型已部署的设备

在上述的「服务详情」页面,可以查看并管理当前服务已部署的设备,包括移除设备、将服务下发到更多的设备等。

模型中心	端云协同服务 : > 服务详情
我的模型	服务名称 模型D 246 当前可用版本 V2, V1 部署设备数 1/1
创建模型	设备类型 服务器 应用平台 Linux-通用X86 CPU
训练模型	
校验模型	下发到更多设备
发布模型	设备名称 最新下发模型版本 部署到期时间 设备连接状态 服务下发状态 最新同步时间 操作
EasyData数据服务	ubuntu-local-fiy 💬 V2 2021-02-06 在线 V2下发成功 2021-01-07 14:45 宣者服务配置 宣者设备详情 下发新版本 网络设备
数据总览	
标签组管理	
在线标注	

本地设备管理

在我的本地设备页面可以进行所有本地设备的管理。

查看单台设备的运行状态

点击单台设备的「服务详情」,可查看设备上运行的多个服务及设备状态:

模型中心	端云协同服务 > 我的本地设备						
我的模型	我的本地设备						点击收起
创建模型	在本页面新增设备、联网激活 利用率等信息。	舌本地设备后,即可将「 \$	的部署包」了	瓦发布成功的部署包一键下发	到设备上。设备联	网时,可以查看设备上部	署的服务、设备的运行状态、资源
训练模型							
校验模型	新增设备						
发布模型	设备名称	创建时间	设备类型	应用平台	设备连接状态	最新同步时间	操作
EasyData数据服务	123123 😳	2021-01-07 13:25		-	未激活	-	删除 激活设备 查看详情
数据总览	showcase-test 💬	2020-12-15 17:57	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 19:22	激活设备 查看详情
标签组管理	linux-x86-zqw-2 😁	2020-12-15 16:24	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 19:42	激活设备 查看详情
在线标注	linux-x86-zqw 😳	2020-12-15 15:29	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-15 16:40	激活设备 宣看详情
云服务数据回流	firefly_rk3399pro 💬	2020-12-15 14:39	-		离线 ⑦	2020-12-15 21:33	激活设备 查看详情
EasyEdge本地部署	ubuntu-local-fjy 💬	2020-12-14 21:38	服务器	Linux-AMD64(x86-64)	在线	2021-01-07 15:01	查看详情
纯离线服务	hfl-1 💬	2020-12-14 19:24	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 14:18	激活设备 宣看详情
端云协同服务 ヘ	edge新博 💬	2020-12-11 14:52	-	-	未激活	-	删除 激活设备 查看详情

设备详情会展示当前设备的最新同步时间,以及CPU使用率、内存使用率等。服务列表则展示了当前设备上部署服务的运行情况和资源占用情况

5						
模型中心	端云协同服务 > 我的本地设备 >	ubuntu-local-fjy				
我的模型	设备详情					
创建模型	设备名称 ubuntu-local-fjy	连接状态 在线	ê	实时刷新	OFF	
训练模型	设备类型 服务器	应用平台 Linu	JX-AMD64(x86-64)	最新同步时间 202	1-01-07 15:00	
校验模型	CPU使用率		内存使用率			
发布模型	31.1%		35.8%			
EasyData数据服务						
数据总览	端云协同服务详情					
标签组管理	服务名称	模型ID	CPU占比	内存使用情况	内存占比	操作
在线标注	2008-02	246	0.01%	156.7MB	0.93%	宣看服务配置
云服务数据回流						
EasyEdge本地部署						
纯离线服务						

```
 ⊙ 软硬一体方案部署
```

心 如何获取物体检测软硬一体产品

为进一步提升前端智能计算的用户体验,EasyDL推出了多款软硬一体方案。将高性能硬件与EasyDL图像分类/物体检测模型深度适配,可应用于 工业分拣、视频监控等多种设备端离线计算场景,让离线AI落地更轻松。了解不同方案

方案获取流程如下:

Step 1:在EasyDL训练专项适配所选硬件的图像分类/物体检测模型,迭代模型至效果满足业务要求

训练配置						
*部署方式	公有云部署 Eas	R时免费 xyEdge本地部署 浏览器/小程序部署 如何选择部署方式?				
*选择设备	◎服务器	通用小型设备				
选择硬件	 Edgeboard(FZ) 	○ Edgeboard(VMX) ○ 辨影(Air/Pro) NEW ○	Jetson(Nano/TX2/X	avier) 71	解不同方案	
*选择算法	○ 高精度 ⑦ 🔹 💿	高性能 ⑦ 🔹 📄 AI市场已购模型				
高级训练配置						
训练环境	名称	规格	算力	速度比例	价格	
	GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	免费	
	GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥0.36/分钟	(50小时*节点免费)
	GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥0.45/分钟	(53.28小时*节点免费)
训练费用 免费						
当前GPU V100卡师	◎余免费资源:53.28(小	时 * 节点) 为保证训练任务顺利进行,请保证您的账户余额充足,	可点击进行充值 立即	<mark>充值</mark> ,推荐购到	实特惠小时包,可享受	受至少8 <mark>折优惠</mark> 训练 前往购买
开始训练						

Step 2:发布模型时选择对应硬件

模型中心	发布模型			
我的模型	选择模型:	test02	~	
创建模型	部署方式:	专项硬件适配SDK	~	
训练模型	选择硬件:	Jetson(Nano/TX2/Xavier)	~	了解更多
校验模型	选择版本:	V2	~	
发布模型		提交		
EasyData数据服务				
我的数据集				
创建数据集				
标注数据集				
管理云服务调用数据				

Step 3:在AI市场购买方案获得硬件和用于激活专用SDK的专用序列号,参考文档集成后,即可实现离线AI预测

模型中心	我的模型 > test02 > V1的专项硬件适配SDK服务详情				
我的模型	专项硬件适配SDK获取				
创建模型	专项硬件适配SDK是EasyDL软硬一体方案的软件部分,建议在AI市场购买整体方案,同时获得专用 SDK激活序列号及专项适配硬件				
训练模型	如已在其他渠道购买硬件,可点击「获取序列号」前往控制台,支持申请专用的测试序列号、购买专 用的永久有效序列号				
校验模型	专项硬件适配SDK 操作				
发布模型	EdgeBoard(VMX)专用SDK 前往AI市场购买 下载SDK V 管理序列号				
EasyData数据服务					
我的数据集					
创建数据集					
标注数据集					
管理云服务调用数据					

如有其他硬件方案需求,请在百度智能云控制台内提交工单反馈。

心物体检测EdgeBoard(FZ)专用SDK集成文档

简介

本文档介绍 EasyEdge/EasyDL在EdgeBoard®边缘计算盒/Lite计算卡上的专用软件的使用流程。

EdgeBoard系列硬件可直接应用于AI项目研发与部署,具有高性能、易携带、通用性强、开发简单等四大优点。

详细硬件参数请在AI市场浏览。

EdgeBoard产品使用手册:https://ai.baidu.com/ai-doc/HWCE/Yk3b86gvp

软核版本

Baidu 百度智能云文档

CPP-SDK版本	对应软核
1.3.2、1.3.4、1.3.5	1.8.1
1.3.0、1.3.1、1.3.2、1.3.4	1.8
0.5.7-1.2.1	1.5
0.5.2+	1.4

SDK升级需配合EdgeBoard硬件软核升级,建议升级软核为SDK对应版本,否则可能出现结果错误或者其他异常。

可以通过dmesg | grep "DRIVER Version" 命令获取EdgeBoard当前的软核版本

Release Notes 注意*:升级完成相应的软核之后需要重启机器生效。

sdk对应的软核说明: 如果客户使用的软核是mobile版本的,需要使用1.4的SDK;如果不是mobile 版本,可以选择1.5+(目前最高版本更新至 1.8.1)版本的SDK使用。

1.5+版本的软核以及sdk更新情况如下表所示:

时间	版本	说明	EdgeBoard非mobile对应的软核以及特性	
2021.1 2.20	1.3.5	升预测引擎为PaddleLite 1.8.1,推理库支持了 Ubuntu18.04文件系统	https://ai.baidu.com/ai-doc/HWCE/Fkuqounlk(含有EB升级 Ubuntu18.04系统的步骤)	
2021.1 0.15	1.3.2、 1.3.4、1.3.4	推理库支持了Ubuntu18.04文件系统	https://ai.baidu.com/ai-doc/HWCE/Fkuqounlk	
2021.0 6.29	1.3.1	视频流解析支持分辨率调整;预测引擎升级;	https://ai.baidu.com/ai-doc/HWCE/Ikqgcqt5x	
2021.0 5.14	1.3.0	新增视频流接入支持;展示已发布模型性能评 估报告	https://ai.baidu.com/ai-doc/HWCE/Ikqgcqt5x	
2021.0 5.14	1.2.1	功能无更新	https://ai.baidu.com/ai-doc/HWCE/okqiwkm32	
2020.1 0.29	0.5.7	预测引擎切换为PaddleLite 1.5	-	
2019.1 2.27	0.4.5	引擎升级,支持zu5/zu3,支持EasyDL 高精度 检测模型	-	
2019.0 7.25	0.4.0	EdgeBoard SDK Release!	-	

mobile软核以及sdk更新情况如下表所示: | 时间 | 版本 | 说明 |EdgeBoard mobile对应的软核以及特性 | | --- | ---- | ---- | | 2021.05.14|1.2.1|功能无更新|https://ai.baidu.com/ai-doc/HWCE/okqiwkm32|https://ai.baidu.com/ai-doc/HWCE/Lkqiwlziw||

快速开始

开发者从EasyEdge/EasyDL下载的软件部署包中,包含了简单易用的SDK和Demo。只需简单的几个步骤,即可快速部署运行EdgeBoard计算盒。

部署包中包含多版本SDK:

• baidu easyedge_linux_cpp_aarch64_EdgeBoardFZ1.8*:适用于EdgeBoard 1.5+软核

• baidu easyedge_linux_cpp_aarch64_EdgeBoardFZ1.4*:适用于EdgeBoard 1.4软核

SDK文件结构

bai	du_easyedge_linux_cpp_aarch64_EdgeBoardFZ1.5_*
\vdash	— ReadMe.txt
	— bin
	easyedge_image_inference
	easyedge_serving
	easyedge_video_inference
	include
	L easyedge
	— lib
	libeasyedge.so -> libeasyedge.so.1
	libeasyedge.so.1 -> libeasyedge.so.1.3.1
	libeasyedge.so.1.3.1
	libeasyedge_static.a
	libeasyedge_videoio.so -> libeasyedge_videoio.so.1
	libeasyedge_videoio.so.1 -> libeasyedge_videoio.so.1.3.1
	libeasyedge_videoio.so.1.3.1
	libeasyedge_videoio_static.a
	libpaddle_full_api_shared.so -> libpaddle_full_api_shared.so.1.8.0
	libpaddle_full_api_shared.so.1.8.0
	libverify.so -> libverify.so.1
	libverify.so.1 -> libverify.so.1.0.0
	Libverify.so.1.0.0
	now_sre.log
	src
	CMakeLists.txt
	cmake
	common
	demo_image_inference
	demo_serving
	Lefter demo_video_inference
L	— thirdparty
l	opencv

1.1.0+的SDK自带OpenCV, src编译的时候会引用thirdparty/opencv路径下的头文件和库文件。

Demo使用流程

用户在AI市场购买计算盒之后,请参考以下步骤进行集成和试用。

1. 将计算盒连接电源

指示灯亮起,等待约1分钟。

- 参考EdgeBoard使用文档配置网口或串口连接。登录EdgeBoard计算盒。
- 加载驱动 (开机加载一次即可)。

insmod /home/root/workspace/driver/{zu9|zu5|zu3}/fpgadrv.ko

根据购买的版本,选择合适的驱动。若未加载驱动,可能报错:

Failed to to fpga device: -1

• 设置系统时间 (系统时间必须正确)

date --set "2019-5-18 20:48:00"

2. (可选) 启动HTTP服务

部署包中附带了HTTP服务功能,开发者可以进入SDK根目录,运行easyedge_serving程序启动HTTP服务。

```
**./easyedge_serving {RES目录} "" {绑定的host,默认0.0.0.} {绑定的端口,默认24401}**
cd ${SDK_ROOT}
export LD_LIBRARY_PATH=./lib
./demo/easyedge_serving ../../../RES ""
```

日志显示

2019-07-18 13:27:05,941 INFO [EasyEdge] [http_server.cpp:136] 547974369280 Serving at 0.0.0.0:24401

则启动成功。此时可直接在浏览器中输入http://{EdgeBoard计算盒ip地址}:24401/,在h5中测试模型效果。



同时,可以调用HTTP接口来访问盒子。具体参考下文接口说明。

EdgeBoard HTTP Server 目前使用的是单线程处理请求。

3. 编译运行Demo

编译:

```
cd src
mkdir build && cd build
cmake .. && make
```

运行

./easyedge_image_inference {RES资源文件夹路径} {测试图片路径}

便可看到识别结果。

使用说明

使用流程

激活成功之后,有效期内可离线使用。

- 1. 配置PaddleFluidConfig
- 2. 新建Predictor :global_controller()->CreateEdgePredictor(config);
- 3. 初始化 predictor->init()
- 4. 传入图片开始识别predictor->infer(img, ...);

目前EdgeBoard暂不支持并行多模型计算。

接口说明

预测图片

/**

- * @brief 同步预测接口
- * inference synchronous
- * Supported by most chip and engine
- * @param image: must be BGR , HWC format (opencv default)
- * @param result
- * @param threshold
- * @return
- */
- virtual int infer(

cv::Mat &image, std::vector<EdgeResultData> &result, float threshold = 0.1

) = **O**;

识别结果说明

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index; // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // object detection field
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
};
```

关于矩形坐标

```
x1 * 图片宽度 = 检测框的左上角的横坐标
```

y1 * 图片高度 = 检测框的左上角的纵坐标

```
x2*图片宽度 = 检测框的右下角的横坐标
```

y2 * 图片高度 = 检测框的右下角的纵坐标

可以参考demo文件中使用opencv绘制矩形的逻辑。

HTTP 私有服务请求说明

http 请求参数

URL中的get参数:

参数	说明		默认值
threshold	阈值过滤,	0~1	0.1

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

Python

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
    'http://127.0.0.1:24401/',
    params={'threshold': 0.1},
    data=img).json()
```

Cpp label=C#

FileStream fs = new FileStream("./img.jpg", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close();

WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close(); response.Close();

Cpp label=C++ 需要安装curl

```
#include <sys/stat.h>
 #include <curl/curl.h>
 #include <iostream>
 #include <string>
 #define S_ISREG(m) (((m) & 0170000) == (0100000))
 #define S_ISDIR(m) (((m) & 0170000) == (0040000))
 size_t write_callback(void *ptr, size_t size, size_t num, void *data) {
    std::string *str = dynamic_cast<std::string *>((std::string *)data);
    str->append((char *)ptr, size*num);
    return size*num;
 }
 int main(int argc, char *argv[]) {
    const char *post_data_filename = "./img.jpg";
    FILE *fp = NULL;
    std::string response;
    struct stat stbuf = { 0, };
    fp = fopen(post_data_filename, "rb");
    if (!fp) {
      fprintf(stderr, "Error: failed to open file "%s"
 ", post_data_filename);
      return -1;
    }
    if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
      fprintf(stderr, "Error: unknown file size "%s"
 ", post_data_filename);
      return -1;
    }
    CURL *curl;
    CURLcode res;
    curl_global_init(CURL_GLOBAL_ALL);
    curl = curl_easy_init();
    if (curl != NULL) {
      curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
      curl_easy_setopt(curl, CURLOPT_POST, 1L);
      curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
      curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
  curl_easy_setopt(curl, CURLOPT_WRITEFUNCTION, write_callback);
      curl_easy_setopt(curl, CURLOPT_WRITEDATA, &response);
      res = curl_easy_perform(curl);
      if (res != CURLE_OK) {
         fprintf(stderr, "curl_easy_perform() failed: %s
 ", curl_easy_strerror(res));
      }
      std::cout << response << std::endl; // response即为返回的json数据
      curl_easy_cleanup(curl);
    }
    curl_global_cleanup();
    fclose(fp);
    return 0;
 }
```

Java请求示例

http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考预测图像-返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

```
{
  "cost_ms": 52,
  "error_code": 0,
  "results": [
   {
      "confidence": 0.94482421875,
      "index": 1,
      "label": "IronMan",
      "x1": 0.059185408055782318,
      "x2": 0.18795496225357056,
      "y1": 0.14762254059314728,
      "y2": 0.52510076761245728
    },
    {
      "confidence": 0.94091796875,
      "index": 1,
      "label": "IronMan",
      "x1": 0.79151463508605957,
      "x2": 0.92310667037963867,
      "y1": 0.045728668570518494,
      "y2": 0.42920106649398804
}
```

预测视频

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

● 接口

classVideoDecoding :

/**

- * @brief 获取输入源的下一帧
- * @param frame_tensor
- * @return

*/

virtual int next(FrameTensor &frame_tensor) = 0;

/**

- * @brief 显示当前frame_tensor中的视频帧
- * @param frame_tensor
- * @return
- */

virtual int display(const FrameTensor &frame_tensor) = 0;

/**

* @brief 将当前frame_tensor中的视频帧写为本地视频文件

- * @param frame_tensor
- * @return

*/

virtual int save(FrameTensor & frame_tensor) = 0;

```
/**
```

```
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
* @brief 获取视频的width属性
* @return
*/
virtual int get_width() = 0;
/**
```

```
* @brief 获取视频的height属性
* @return
*/
virtual int get_height() = 0;
```

struct VideoConfig

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type; // 输入源类型
  std::string source_value;   // 输入源地址,如视频文件路径、摄像头index、网络流地址
 int skip_frames{0};
                      // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
  int retrieve_all{false};
                     // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
                      // 在采取抽帧之前设置视频的fps
  int input_fps{0};
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
 bool enable_display{false};
  std::string window_name{"EasyEdge"};
  bool display_all{false};
                       // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                       // 是否存储所有frame, 若为false, 仅存储根据skip_frames抽取的frame
  std::map::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被
Baidu 百度智能云文档

抽取返回,以作为显示或存储用。 input_fps:用于抽帧前设置fps。 resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。 conf:高级选项。部分配置会通过该map来设置。

注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

错误说明

SDK所有主动报出的错误,均覆盖在EdgeStatus枚举中。同时SDK会有详细的错误日志,开发者可以打开Debug日志查看额外说明:

EdgeLogConfig log_config; log_config.enable_debug = true; global_controller()->set_log_config(log_config);

FAQ

1. 如何处理一些 undefined reference?

如: undefined reference to `curl_easy_setopt@CURL_OPENSSL_3'

可以通过安装libcurl3 libcurl-openssl1.0-dev来解决。如果开发者想不想使用低版本的openssl(如Ubuntu 18.04),可以link静态 库easyedge_static.a,自己指定需要的Library的版本。

示例:修改CMakeList.txt

```
find_package(CURL REQUIRED)
target_link_libraries(easyedge_demo ${OpenCV_LIBS} easyedge_static pthread ${CURL_LIBRARIES} paddle-mobile)
```

2. error while loading shared libraries: libeasyedge.so.0.4.0: cannot open shared object file: No such file or directory

类似错误包括libpaddle-mobile.so找不到。

直接运行SDK自带的二进制可能会有这个问题,设置LD_LIBRARY_PATH为SDK部署包中的lib目录即可。开发者自行使用CMake编译的二进制可以 有效管理.so的依赖。

3. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

4. 预测过程中报内存不足"Killed"

此问题仅出现在ZU5,因为FZ5A带vcu,给他预留的内存过大导致,如果用不到VCU可以把这部分改小。修改/run/media/mmcblk1p1/uEnv.txt:

ethaddr=00:0a:35:00:00:09 uenvcmd=fatload mmc 1 0x3000000 image.ub && bootm 0x3000000

bootargs=earlycon console=ttyPS0,115200 clk_ignore_unused cpuidle.off=1 root=/dev/mmcblk1p2 rw rootwait cma=128M

注意中间空行要保留。

5. 预测结果异常

如果购买的计算盒较早,驱动文件较旧,而SDK比较新(或SDK比较旧,但是计算盒较新),可能出现结果异常,如结果均为空或者 nan。 请参

考"软核版本"小节更新软核和驱动版本。

6. 编译过程报错file format not recognized

libeasyedge.so: file format not recognized; treating as linker script

下载的SDK zip包需要放到板子内部后,再解压、编译。

7. 提示 driver_version(1.4.0) not match paddle_lite_version(1.5.1) 需更新驱动,否则可能导致结果异常。参考"软核版本"小节。

心物体检测EdgeBoard(VMX)专用SDK集成文档

简介

本文档旨在介绍 EasyDL在EdgeBoard USB加速卡VMX(以下简称VMX加速卡或加速卡)上的专用软件的使用流程。 EdgeBoard系列硬件适用于项 目研发与部署,具有高性能、易携带、通用性强、开发简单等四大优点。 您可在AI市场了解EdgeBoard相关系列产品,同时可以在软硬一体方 案了解性能数据。

注意:本型号主要面向产品集成和企业项目,未同时售卖散热片和外壳,部分情况下芯片温度较高,<mark>开发过程中,请勿用手触摸,谨防烫伤</mark>

硬件介绍

VMX加速卡,采用Intel® Movidus[™] 视觉 MyriadX处理器芯片,通过 USB3.0 通讯type-c接口方式,配合外围电路即可将该模组嵌入到第三方智能 化产品中,采用标准 USB通讯协议,对接简单,开发速度快,具有强大的深度学习计算功能。可通过OpenVINO[™]和OpenCV软件库工具链移植算 法,兼容百度PaddlePaddle支持Paddle2onnx和PaddleHub并集成EasyDL,使产品应用范围广,性能更稳定,增强用户体验。

VMX加速卡适用于深度学习加速,能够解决复杂的人工智能软硬件设计挑战,它可以集成基于视觉的加速器和推理引擎来实现深度边缘学习的解决方案。(3D/2D人脸识别、人头检测、人脸属性分析(性别、年龄)、人脸特征比对、手势及姿态识别、物体检测及分类、算法移植等功能。)

硬件配置与说明

核心板模块: Intel® Movidus[™] MyriadX,内置内存LP-DDR4 4GBit。

• 硬件指标

CPU

o Intel® Movidius Myriad X MA2485 Vision Processing Unit

o Total performance of over 4 trillion operations per second (TOPS) o Over 1 TOPS performance on neural network inference w/ NCE accelerator o 16 Programmable 128-bit VLIW Vector Processors o 16 Configurable MIPI Lanes w/ enhanced Vision Accelerators o 2.5 MB of Homogenous On-Chip Memory w/ 4Gbit LPDDR4

Size

o 38mm x 38mm

Interface o USB TYPE C (USB3.0) 辅助接口精简设计

Boot o USB 启动模式 - 内置 switch 缺省模式设置

Power o 平均功耗0.5W~2.2W

Security o 支持 eFuse 加密

运行说明

VMX加速卡包含独立的AI运算芯片,采用 USB Type-C通讯方式,通讯协议简单可靠,可连接不同芯片架构主机,包括 X86、ARM SOC等。加速卡运行需要通过TypeC接口连接宿主机执行,宿主机目前支持的软硬件环境包括:

- Linux: x86-64, armv7hf
- Windows: x86-64, Windows 10

使用过程中,请尽量避免直接触碰板卡元器件;或者使用防静电锡纸包裹板卡。

Baidu 百度智能云文档

开发者从EasyDL训练模型之后,下载的软件部署包中,包含了简单易用的SDK和Demo。只需简单的几个步骤,即可快速部署运行。

Release Notes

Python SDK

时间	版本	说明
2020.12.18	1.2.0	性能优化;接口优化升级;推理引擎升级
2020.09.17	1.1.19	支持更多模型与平台

Python SDK适用于Linux x86-64和Windows平台。

2020-12-18: 【接口升级】 Python SDK序列号配置接口从1.2.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

C++ SDK

时间	版本	说明
2021.06.29	1.3.1	视频流解析支持分辨率调整
2021.05.14	1.3.0	新增视频流接入支持;展示已发布模型性能评估报告
2020.12.18	1.0.0	性能优化;接口优化升级;推理引擎升级
2020.09.17	0.5.6	新增C++ SDK,支持Linux armv7hf(树莓派)架构的硬件接入VMX预测

C++ SDK适用于Linux x86-64和Linux armv7hf平台。

2020-12-18: 【接口升级】 参数配置接口从1.0.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

将加速卡连接宿主机 请使用质量合规的usb线连接。连接之后,检查设备是否被操作系统识别: Linux 通过Isusb -v 命令检查是否有 Myriad设 备:

> sudo lsusb -v	grep -C 5 Myriad
bMaxPacketSiz	e0 64
idVendor	0x03e7
idProduct	0x2485
bcdDevice	0.01
iManufacturer	1 Movidius Ltd.
iProduct	2 Movidius Myriad
iSerial	3 03e72485

Windows 可以在设备管理器中查询。

如果使用 VirtualBox 之类的虚拟机,请在虚拟机加入 03e7:24 和 03e7:f63b 两个 usb 设备。

获取并安装依赖

1) 安装依赖

宿主机与sdk为以下情况: 1□ Windows x86-64: 请参考 OpenVINO toolkit 文档安装 2022.1.1LTS 版本 2□ Linux x86-64: 请参考 OpenVINO toolkit 文档安装 2022.1.1LTS 版本, 安装时可忽略Configure the Model Optimizer及后续部分。

注:1. ebvpu+arm 组合产品正在下架过渡期,后续不再维护 armv7 版本的 sdk;2. 当前问题可以升级 openvino 版本解决,但无法保证 armv7 环境下的稳定性;

安装完毕,运行之前,请按照OpenVino的文档 设置环境变量

source /opt/intel/openvino/bin/setupvars.sh

2) 从EasyDL 控制台获取SDK 在任意位置解压缩。

获取序列号 从AI市场订单详情或者EasyDL控制台获取序列号。

更换序列号、更换设备时,首次使用需要联网激活。激活成功之后,有效期内可离线使用。

请确保激活设备时使用的 操作系统账号与后续使用时运行的账号一致,否则会造成验证失败

Python SDK

1. 安装wheel包

pip3 install -U BaiduAI_EasyEdge_SDK-{版本号}-cp37-cp37m-linux_x86_64.whl

注意,请根据python的版本选择对应的whl文件,其中,1.2.0是SDK版本号,cp37表示是python3.7版本

--

注意,pip安装时请添加-U参数

2. 将步骤2中获得的序列号 填入demo.py

pred = edge.Program() pred.set_auth_license_key("这里填写序列号")

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己 的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的设置),需要调用函数指定实例数鉴权模式,并且实例数鉴权模式下,支持指定 license证书更新时间,单位是秒,要求设置为大于20的整数,否则会采用默认的license更新时间,参考

pred.set_instance_auth_mode()
pred.set_instance_update_interval(200)

3. 测试demo.py

python3 demo.py {模型资源文件夹RES路径} {待识别的图片路径}

生成的样例结果图片如下:



使用流程

import BaiduAI.EasyEdge as edge

pred = edge.Program() pred.set_auth_license_key("这里填写序列号") pred.init(model_dir={RES文件夹路径}, device=edge.Device.MOVIDIUS, engine=edge.Engine.OPENVINO) pred.infer_image({numpy.ndarray的图片}) pred.close()

接口的详细说明请主要参考 SDK 中的接口注释

接口说明

Program

● 初始化

def init(self,

model_dir, device=Device.CPU, engine=Engine.NCSDK, config_file='conf.json', preprocess_file='preprocess_args.json', model_file='model', params_file='params', graph_file='graph.ncsmodel', label_file='label_list.txt', device_id=0, **kwargs): Args: model_dir: str device: BaiduAI.EasyEdge.Device engine: BaiduAI.EasyEdge.Engine preprocess_file: str model_file: str params_file: str graph_file: str ncs的模型文件 或 PaddleV2的模型文件 label_file: str device_id: int 设备ID thread_num: int CPU的线程数

Raises:

RuntimeError, IOError Returns: bool: True if success

```
....
```

• 预测单张图像

```
def infer_image(self, img, threshold=None,
        channel_order='HWC',
        color_format='BGR',
        data_type='numpy'
        ):
 ....
 Args:
   img: np.ndarray or bytes
   channel_order(string):
     channel order: HWC or CHW
   color_format(string):
     color format order: RGB or BGR
   threshold(float):
      only return result with confidence larger than threshold
   data_type(string): 仅在图像分割时有意义。 'numpy' or 'string'
      'numpy': 返回已解析的mask
      'string': 返回未解析的mask游程编码
 Returns:
   list
 ....
```

• 返回格式: [dict1, dict2, ...]

Baidu 百度智能云文档

字段	类型	取值	说明
confidence	float	0~1	分类或检测的置信度
label	string		分类或检测的类别
index	number		分类或检测的类别
x1, y1	float	0~1	物体检测,矩形的左上角坐标 (相对长宽的比例值)
x2, y2	float	0~1	物体检测,矩形的右下角坐标(相对长宽的比例值)
mask	string/numpy.ndarray	图像分割的mask	

关于矩形坐标

x1 * 图片宽度 = 检测框的左上角的横坐标

y1 * 图片高度 = 检测框的左上角的纵坐标

x2 * 图片宽度 = 检测框的右下角的横坐标

y2 * 图片高度 = 检测框的右下角的纵坐标

可以参考 demo 文件中使用 opencv 绘制矩形的逻辑。

结果示例

● i) 图像分类

```
{
    "index": 736,
    "label": "table",
    "confidence": 0.9
}
```

● ii) 物体检测

```
{
    "y2": 0.91211,
    "label": "cat",
    "confidence": 1.0,
    "x2": 0.91504,
    "index": 8,
    "y1": 0.12671,
    "x1": 0.21289
}
```

• iii) 图像分割

```
{
    "name": "cat",
    "score": 1.0,
    "location": {
        "left": ...,
        "top": ...,
        "width": ...,
        "height": ...,
    },
    "mask": ...
}
```

mask字段中,data_type为numpy时,返回图像掩码的二维数组

{

```
{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0},
{0, 0, 0, 1, 1, 1, 0, 0, 0, 0},
{0, 0, 0, 1, 1, 1, 0, 0, 0, 0},
{0, 0, 0, 1, 1, 1, 0, 0, 0, 0},
{0, 0, 0, 1, 1, 1, 0, 0, 0, 0},
{0, 0, 0, 0, 0, 0, 0, 0, 0, 0},
}
其中1代表为目标区域,0代表非目标区域
```

data_type为string时,mask的游程编码,解析方式可参考 demo

C++ SDK

使用说明

模型资源文件默认已经打包在开发者下载的SDK包中。请先将SDK包整体拷贝到具体运行的宿主机设备中,再解压缩编译;

在编译或运行demo程序前执行以下命令:

source \${cpp_kit位置路径}/thirdparty/openvino/bin/setupvars.sh

如果openvino预测引擎找不到设备需要执行以下命令:

sudo cp \${cpp_kit位置路径}/thirdparty/openvino/deployment_tools/inference_engine/external/97-myriad-usbboot.rules /etc/udev/rules.d/ sudo udevadm control --reload-rules sudo udevadm trigger sudo ldconfig ```

使用流程

```
// step 1: 配置运行参数
 EdgePredictorConfig config;
 config.set_config(easyedge::params::PREDICTOR_KEY_SERIAL_NUM, "this-is-serial-num"); // 设置序列号
 config.model_dir = {模型文件目录};
  // step 2: 创建并初始化Predictor
 auto predictor = global_controller()->CreateEdgePredictor(config);
 if (predictor->init() != EDGE_OK) {
    exit(-1);
  // step 3-1: 预测图像
 auto img = cv::imread({图片路径});
 std::vector<EdgeResultData> results;
 predictor->infer(img, results);
// step 3-2: 预测视频
std::vector<EdgeResultData> results;
FrameTensor frame_tensor;
 VideoConfig video_config;
 video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h
 video_config.source_value = video_src;
  ... more video_configs, 根据需要配置video_config的各选项
  */
 auto video_decoding = CreateVideoDecoding(video_config);
  while (video_decoding->next(frame_tensor) == EDGE_OK) {
    results.clear();
    if \; (frame\_tensor.is\_needed) \; \{
      predictor->infer(frame_tensor.frame, results);
      render(frame_tensor.frame, results, predictor->model_info().kind);
    //video_decoding->display(frame_tensor); // 显示当前frame,需在video_config中开启配置
     //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置
```

}

运行参数配置 运行参数的配置通过结构体 EdgePredictorConfig 完成,其定义如下所示:

```
struct EdgePredictorConfig {
 /**
  * @brief 模型资源文件夹路径
 */
 std::string model_dir;
 std::map<std::string, std::string> conf;
 EdgePredictorConfig();
 template<typename T>
 T get_config(const std::string &key, const T &default_value);
 template<typename T = std::string>
 T get_config(const std::string &key);
 template<typename T>
 const T *get_config(const std::string &key, const T *default_value);
 template<typename T>
 void set_config(const std::string &key, const T &value);
 template<typename T>
 void set_config(const std::string &key, const T *value);
 static EdgePredictorConfig default_config();
```

```
};
```

运行参数选项的配置以key、value的方式存储在类型为std::map的conf中,并且键值对的设置和获取可以通

过EdgePredictorConfig的set_config和get_config函数完成。同时部分参数也支持以环境变量的方式设置键值对。 EdgePredictorConfig的具体使用 方法可以参考开发工具包中的demo工程。

具体支持的运行参数可以参考开发工具包中的头文件。

初始化

接口

auto predictor = global_controller()->CreateEdgePredictor(config); predictor->init();

若返回非0,请查看输出日志排查错误原因。

预测图像

```
● 接口
```

/** * @brief * 通用接口

- * @param image: must be BGR , HWC format (opencv default)
- * @param result
- * @return

*/

virtual int infer(

cv::Mat& image, std::vector<EdgeResultData>& result

) = 0;

图片的格式务必为opencv默认的BGR, HWC格式。

• 返回格式

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index; // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测活图像分割时才有
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割时才有
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding, 游程编码的mask
```

};

关于矩形坐标

x1 * 图片宽度 = 检测框的左上角的横坐标

- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2*图片宽度 = 检测框的右下角的横坐标
- y2 * 图片高度 = 检测框的右下角的纵坐标

关于图像分割mask

```
cv::Mat mask为图像掩码的二维数组
{
{0,0,0,0,0,0,0,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,0,0,0,0,0,0,0},
}
其中1代表为目标区域,0代表非目标区域
```

关于图像分割mask_rle

该字段返回了mask的游程编码,解析方式可参考 http demo

以上字段可以参考demo文件中使用opencv绘制的逻辑进行解析

预测视频

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

接口

classVideoDecoding :

/**

- * @brief 获取输入源的下一帧
- * @param frame_tensor
- * @return

*/

virtual int next(FrameTensor &frame_tensor) = 0;

/**

- * @brief 显示当前frame_tensor中的视频帧
- * @param frame_tensor
- * @return
- */

virtual int display(const FrameTensor &frame_tensor) = 0;

/**

* @brief 将当前frame_tensor中的视频帧写为本地视频文件

- * @param frame_tensor
- * @return

virtual int save(FrameTensor &frame_tensor) = 0;

```
/**
```

*/

```
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
* @brief 获取视频的width属性
* @return
*/
virtual int get_width() = 0;
/**
* @brief 获取视频的height属性
```

```
* @return
*/
```

```
virtual int get_height() = 0;
```

struct VideoConfig

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type; // 输入源类型
  std::string source_value;   // 输入源地址,如视频文件路径、摄像头index、网络流地址
  int skip_frames{0};    // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
int retrieve_all{false};   // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
                         // 在采取抽帧之前设置视频的fps
  int input_fps{0};
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
  bool enable_display{false};
  std::string window_name{"EasyEdge"};
  bool display_all{false};
                         // 是否显示所有frame,若为false,仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                           // frame存储为视频文件的路径
  bool save_all{false};
                          // 是否存储所有frame, 若为false, 仅存储根据skip_frames抽取的frame
  std::map::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被

Baidu 百度智能云文档

抽取返回,以作为显示或存储用。 input_fps:用于抽帧前设置fps。 resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。 conf:高级选项。部分配置会通过该map来设置。

注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

设置序列号

请在网页控制台中申请序列号,并在init初始化前设置。 LinuxSDK 首次使用需联网授权。

EdgePredictorConfig config; config.set_config(easyedge::params::PREDICTOR_KEY_SERIAL_NUM, "this-is-serial-num");

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己 的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的变量设置),需要设置额外的环境变量,指定CONTROLLER_KEY_AUTH_MODE为2, global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2),实例数鉴权模式下还支持指定license证书更新时间,单 位是秒,要求设置为大于20的整数,否则会采用默认的license更新时间,修改实例数鉴权license更新时间的方法参考 global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 300)

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

EdgeLogConfig log_config; log_config enable_debug = true; global_controller()->set_log_config(log_config);

http服务

1. 开启http服务 http服务的启动参考demo_serving.cpp文件。

```
/**
 * @brief 开启一个简单的demo http服务。
 * 该方法会block直到收到sigint/sigterm。
 * http服务里,图片的解码运行在cpu之上,可能会降低推理速度。
 * @tparam ConfigT
 * @param config
 * @param host
 * @param port
 * @param service_id service_id user parameter, uri '/get/service_id' will respond this value with 'text/plain'
 * @param instance_num 实例数量,根据内存/显存/时延要求调整
 * @return
 */
template<typename ConfigT>
int start_http_server(
     const ConfigT & config,
     const std::string &host
     int port,
     const std::string &service_id
     int instance_num = 1);
```

● 2. 请求http服务

开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。

URL中的get参数:

参数	说明	默认值		
threshold	阈值过滤,0~2	山如不提供,则会使用模型的推荐阈值		

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

```
Python请求示例
```

```
import requests
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
        'http://127.0.0.1:24401/',
        params=['threshold': 0.1],
        data=img).json()
```

Java请求示例

• http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考预测图像·返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

返回示例

```
"cost_ms": 52,
  "error_code": 0,
  "results": [
    {
       "confidence": 0.94482421875,
       "index": 1,
       "label": "IronMan",
       "x1": 0.059185408055782318,
       "x2": 0.18795496225357056,
       "y1": 0.14762254059314728,
       "y2": 0.52510076761245728
    },
       "confidence": 0.94091796875,
      "index": 1,
       "label": "IronMan",
       "x1": 0.79151463508605957,
       "x2": 0.92310667037963867,
       "y1": 0.045728668570518494,
       "y2": 0.42920106649398804
}
```

其他配置

• 1. 日志名称、HTTP 网页标题设置

通过global_controller的set_config方法设置:

global_controller()->set_config(easyedge::params::KEY_LOG_BRAND, "MY_BRAND");

效果如下:

图片加载失败

Linux FAQ

1. EasyDL 离线 SDK 与云服务效果不一致,如何处理?我们会逐渐消除这部分差异,如果开发者发现差异较大,可通过工单、论坛联系我们协助处理。

2. 硬件出现问题或者出现故障怎么办? 软件使用有问题怎么处理?

- 如果持续在静电较多的环境中使用,建议使用防静电锡纸包裹板卡
- 如果硬件无法启动等故障,您可以通过商品页联系供应商处理;其它硬件问题,您可以邮件 edgeboard-vmx.com ,我们将在0-2日内处理您的问题。为加快处理进度,您在邮件中,尽量描述清楚问题或者需求细节,避免来回沟通。
- 软件使用问题,请尽量通过工单、论坛联系我们协助处理。

3. 运行时报错:NC_ERROR

Can not init Myriad device: NC_ERROR

一般是硬件没有插上,请确保Isusb能够找到该硬件。或者等待几秒后再试。

快速开始 Windows

1. 安装依赖

将操作系统升级到Windows 10

安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642

Visual C++ Redistributable Packages for Visual Studio 2013

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

Visual C++ Redistributable Packages for Visual Studio 2015

https://www.microsoft.com/zh-cn/download/details.aspx?id=48145

注意事项

- 1. 安装目录不能包含中文
- 2. Windows Server 请自行开启,选择"我的电脑"——"属性"——"管理"——"添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启即 可。

2. 运行离线SDK

解压下载好的SDK,打开EasyEdge.exe,输入Serial Num



图片加载失败

点击"启动服务",等待数秒即可启动成功,本地服务

默认运行在

http://127.0.0.1:24401/

其他任何语言只需通过HTTP调用即可。

接口调用说明

Python 使用示例代码如下

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
```

params 为GET参数 data 为POST Body result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1}, data=img).json()

C## 使用示例代码如下

FileStream fs = new FileStream("./img.jpg", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close();

WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close(); response.Close();

请求参数

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送图片二进制。

返回参数|字段|类型|取值|说明||--------|----|----|----||confidence|float|0~1|分类或检测的置信度||label|string||分 类或检测的类别||index|number||分类或检测的类别||x1,y1|float|0~1|物体检测,矩形的左上角坐标(相对长宽的比例值)||x2,y2|float |0~1|物体检测,矩形的右下角坐标(相对长宽的比例值)|

关于矩形坐标

- x1 * 图片宽度 = 检测框的左上角的横坐标
- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2 * 图片宽度 = 检测框的右下角的横坐标

y2 * 图片高度 = 检测框的右下角的纵坐标

Windows FAQ

1. 服务启动失败,怎么处理?

请确保相关依赖都安装正确,版本必须如下: .NET Framework 4.5 Visual C++ Redistributable Packages for Visual Studio 2013 * Visual C++ Redistributable Packages for Visual Studio 2015

2. 服务调用时返回为空,怎么处理?调用输入的图片必须是RGB格式,请确认是否有alpha通道。

3. 多个模型怎么同时使用? SDK设置运行不同的端口,点击运行即可。

4. JAVA、C#等其他语言怎么调用SDK?

参考 http://ai.baidu.com/forum/topic/show/943765

5. 启动失败,缺失DLL? 打开EasyEdge.log,查看日志错误,根据提示处理 缺失DLL,请使用 http://www.dependencywalker.com/ 查看相应模块 依赖DLL缺失哪些,请自行下载安装

6. 启动失败,报错NotDecrypted? Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

7. 其他问题 如果无法解决,可到论坛发帖: http://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

心物体检测Jetson专用SDK集成文档

简介

本文档介绍EasyEdge/EasyDL的Jetson SDK的使用方法。Jetson SDK支持的硬件包括Jetson nano, Jetson TX2, Jetson AGX Xavier和Jetson Xavier NX。您可在AI市场了解Jetson相关系列产品,同时可以在软硬一体方案了解部署方案。

模型支持:

- EasyDL图像:图像分类高精度,图像分类高性能,物体检测高精度,物体检测均衡,物体检测高性能,目标跟踪单标签模型。
- BML :
 - 公开数据集预训练模型:SSD-MobileNetV1,YOLOv3-DarkNet,YOLOv3-MobileNetV1,ResNet50,ResNet101,SE-ResNeXt50,SE-ResNeXt101,MobileNetV2,EfficientNetB0_small,EfficientNetB4,MobileNetV3_large_x1_0,ResNet18_vd,SE_ResNet18_vd,Xception71。
 - 百度超大规模数据集预训练模型:YOLOv3-DarkNet, MobileNetV3_large_x1_0, ResNet50_vd, ResNet101_vd。
- EasyEdge: EasyEdge支持的模型较多,详见查看模型网络适配硬件。若模型不在此列表,可以尝试使用自定义网络生成端计算组件。

软件版本支持使用EasyDL的Jetson系列SDK需要安装指定版本的JetPack和相关组件。所支持的JetPack版本会随着SDK版本的升级和新版本 JetPack的推出而不断的更新。在使用SDK前请务必保证软件版本满足此处声明版本。目前所支持的JetPack版本包括:

- JetPack5.0.2
- JetPack5.0.1
- JetPack4.6
- JetPack4.5
- JetPack4.4 (deprecated,该版本SDK会在未来某个版本移除,请切换至新版本JetPack)
- JetPack4.2.2 (已移除,请切换至新版本JetPack)

安装JetPack时请务必安装对应的组件:

- 使用SDK Manager安装JetPack需要勾选TensorRT、OpenCV、CUDA、cuDNN等选项。
- 使用SD Card Image方式(仅对Jetson Nano和Jetson Xavier NX有效)则无需关心组件问题,默认会全部安装。

Release Notes | 时间 | 版本 | 说明 | | --- | --- | | 2022.12.29 | 1.7.2 | 新增支持JetPack5.0.2;缓存机制优化;模型性能优化 || 2022.07.28 | 1.6.0 | 新增支持JetPack5.0.1,新增目标追踪接入实时流的demo || 2022.05.18 | 1.5.0 | 部分模型切换格式,max_batch_size含义变更,由输入 图片数不大于该值变更为等于该值;移除适用于JetPack4.2.2的SDK;示例代码demo_stream_inference重构;示例代码移除frame_buffer,新增 更安全高效的safe_queue || 2021.12.22 | 1.3.5 | 新增支持JetPack4.6;支持在EasyEdge平台语义分割模型生成开发套件;修复缓存问题;支持 自定义缓存路径 || 2021.10.20 | 1.3.4 | 新增支持JetPack4.5;大幅提升EasyDL有损压缩加速模型的推理速度 || 2021.06.29 | 1.3.1 | 视频流支 持分辨率调整;支持将预测后的视频推流,新增推流demo || 2021.05.13 | 1.3.0 | 新增视频流接入支持;EasyDL模型发布新增多种加速方案选 择;目标追踪支持x86平台的CPU、GPU加速版;展示已发布模型性能评估报告 || 2021.03.09 | 1.2.1 | EasyEdge新增一系列模型的支持;性能 优化 || 2021.01.27 | 1.1.0 | EasyDL经典版高性能分类模型升级;

EasyDL经典版检测模型新增均衡选项;

EasyEdge平台新增Jetson系列端计算组件的生成;

问题修复||2020.12.18|1.0.0|接口升级和一些性能优化||2020.08.11|0.5.5|部分模型预测速度提升||2020.06.23|0.5.4|支持 JetPack4.4DP,支持EasyDL专业版更多模型||2020.05.15|0.5.3|专项硬件适配SDK支持Jetson系列|

2022-5-18: 【接口变更】 PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE含义变更。 变更前:预测输入图片数不大于该值均可。 变更后:预 测输入图片数需等于该值。SDK内部对该接口变更做了兼容处理,在输入图片数小于该值时依然可正常运行,但预测性能会和等于该值时一 致。推荐根据实际输入图片数量需求修改该值,尽可能保持最小。 【版本移除】 适用于JetPack4.4版本的SDK被标记为deprecated, SDK会 在未来某个版本移除,建议切换至最新版本JetPack。 适用于JetPack4.2.2版本的SDK被移除。

2020-12-18: 【接口升级】 参数配置接口从1.0.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

2021-10-20: 【版本移除】 适用于JetPack4.2.2版本的SDK被标记为deprecated,该版本代码已停止更新,SDK会在未来某个版本移除,请切 换至新版本JetPack

快速开始 安装依赖 本SDK适用于JetPack4.5、JetPack4.6、JetPack5.0系列版本,请务必安装其中之一版本,并使用对应版本的SDK。 注意在安 装JetPack时,需同时安装CUDA、cuDNN、OpenCV、TensorRT等组件。

如已安装JetPack需要查询相关版本信息,请参考下文中的开发板信息查询与设置。

使用序列号激活

首先请在官网获取序列号。



将获取到的序列号填写到demo文件中或以参数形式传入。

图片加载失败

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己 的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的变量设置),需要设置额外的环境变量,指定CONTROLLER_KEY_AUTH_MODE为2, global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2),实例数鉴权模式下还支持指定license证书更新时间,单 位是秒,要求设置为大于20的整数,否则会采用默认的license更新时间,修改实例数鉴权license更新时间的方法参考 global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 300)

编译并运行Demo 模型资源文件默认已经打包在开发者下载的SDK包中。Demo工程直接编译即可运行。

编译运行:

cd src mkdir build && cd build cmake ... make -j\$(nproc) **make install 为可选,也可将lib所在路径添加为环境变量** sudo make install sudo ldconfig ./demo_batch_inference/easyedge_batch_inference {模型RES文件夹} {测试图片路径或仅包含图片的文件夹路径} {序列号}

demo运行示例:

baidu@nano:~/ljay/easydl/sdk/demo/build\$./demo_batch_inference/easyedge_batch_inference ../../../RES/ /ljay/images/mix008.jpeg 2020-08-06 20:56:30,665 INFO [EasyEdge] 548125646864 Compiling model for fast inference, this may take a while (Acceleration) 2020-08-06 20:57:58,427 INFO [EasyEdge] 548125646864 Optimized model saved to: /home/baidu/.baidu/easyedge/jetson/mcache/24110044320/m_cache, Don't remove it Results of image /ljay/images/mix008.jpeg: 2, kiwi, p:0.997594 loc: 0.352087, 0.56119, 0.625748, 0.868399 2, kiwi, p:0.993221 loc: 0.45789, 0.0730294, 0.73641, 0.399429 2, kiwi, p:0.992884 loc: 0.156876, 0.0598725, 0.3802, 0.394706 1, tomato, p:0.992125 loc: 0.523592, 0.389156, 0.657738, 0.548069 1, tomato, p:0.991821 loc: 0.665461, 0.419503, 0.805282, 0.573558 1, tomato, p:0.989883 loc: 0.297427, 0.439999, 0.432197, 0.59325 1, tomato, p:0.981654 loc: 0.383444, 0.248203, 0.506606, 0.400926 1, tomato, p:0.971682 loc: 0.183775, 0.556587, 0.286996, 0.711361 1, tomato, p:0.968722 loc: 0.379391, 0.0386965, 0.51672, 0.209681 Done



测试Demo HTTP 服务

编译demo完成之后,会同时生成一个http服务,运行

./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401} ./easyedge_serving ../../../RES "1111-1111-1111" 0.0.0.0 24401

日志中会显示

HTTP is now serving at 0.0.0.0:24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。



同时,可以调用HTTP接口来访问服务,具体参考下文接口说明。

使用说明

使用该方式,将运行库嵌入到开发者的程序当中。

使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

```
// step 1: 配置模型运行参数
EdgePredictorConfig config;
config.model_dir = model_dir;
config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, serial_num);
config.set_config(params::PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE, 1); // 优化的模型可以支持的batch_size
config.set_config(params::PREDICTOR_KEY_GTURBO_FP16, false); // 置true开启fp16模式推理会更快,精度会略微降低,但取决于硬
件是否支持fp16,不是所有模型都支持fp16,参阅文档
config.set_config(params::PREDICTOR_KEY_GTURBO_COMPILE_LEVEL, 1); // 编译模型的策略,如果当前设置的max_batch_size与历史编
译存储的不同,则重新编译模型
// step 2: 创建并初始化Predictor
auto predictor = global_controller()->CreateEdgePredictor<EdgePredictorConfig>(config);
if (predictor->init() != EDGE_OK) {
  exit(-1);
// step 3-1: 预测图像
auto img = cv::imread({图片路径});
std::vector<EdgeResultData> results;
predictor->infer(img, results);
// step 3-2: 预测视频
std::vector<EdgeResultData> results
FrameTensor frame_tensor;
VideoConfig video_config;
video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h
video_config.source_value = video_src;
/*
... more video_configs, 根据需要配置video_config的各选项
*/
auto video_decoding = CreateVideoDecoding(video_config);
while (video_decoding->next(frame_tensor) == EDGE_OK) {
 results.clear();
 if (frame_tensor.is_needed) {
    predictor->infer(frame_tensor.frame, results);
    render(frame_tensor.frame, results, predictor->model_info().kind);
  //video_decoding->display(frame_tensor); // 显示当前frame,需在video_config中开启配置
  //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置
```

初始化接口

```
auto predictor = global_controller()->CreateEdgePredictor<EdgePredictorConfig>(config);
if (predictor->init() != EDGE_OK) {
    exit(-1);
}
```

若返回非0,请查看输出日志排查错误原因。

预测接口

```
/**
* @brief
* 单图预测接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
    cv::Mat& image, std::vector<EdgeResultData>& result
) = 0;
/**
* @brief
* 批量图片预测接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
    std::vector<cv::Mat>& image,
     std::vector<std::vector<EdgeResultData>>& results
) = 0;
/**
 * @brief
* 批量图片预测接口,带阈值
 * @related infer(cv::Mat & image, EdgeColorFormat origin_color_format, std::vector<EdgeResultData> &result, float threshold)
*/
virtual int infer(
       std::vector<cv::Mat> &images,
       EdgeColorFormat origin_color_format,
       std::vector<std::vector<EdgeResultData>> &results
       float threshold
) = 0;
```

图片的格式务必为opencv默认的BGR, HWC格式。

批量图片的预测接口的使用要求在调用 init 接口的时候设置一个有效的 max_batch_size,其含义见下方参数配置接口的介绍。

参数配置接口 参数配置通过结构体 EdgePredictorConfig完成。

```
struct EdgePredictorConfig {
    /**
     * @brief 模型资源文件夹路径
    */
    std::string model_dir;
    std::map<std::string, std::string> conf;
    EdgePredictorConfig();
    template<typename T>
    T get_config(const std::string &key, const T &default_value);
    template<typename T = std::string>
    T get_config(const std::string &key);
    template<typename T>
    const T *get_config(const std::string &key, const T *default_value);
    template<typename T>
    void set_config(const std::string &key, const T &value);
    template<typename T>
    void set_config(const std::string &key, const T *value);
    static EdgePredictorConfig default_config();
 };
运行参数选项的配置以key、value的方式存储在类型为std::map的conf中,并且键值对的设置和获取可以通
过EdgePredictorConfig的set_config和get_config函数完成。同时也支持以环境变量的方式设置键值对。 EdgePredictorConfig的具体使用方法可以
参考开发工具包中的demo工程。
```

针对Jetson开发工具包,目前EdgePredictorConfig的运行参数所支持的Key包括如下项:

```
/**
* @brief 当有同类型的多个设备的时候,使用哪一个设备,如:
* GPU: 使用哪张GPU卡
* EdgeBoard(VMX), Movidius NCS:使用哪一张加速卡
* 值类型: int
* 默认值:0
*/
static constexpr auto PREDICTOR_KEY_DEVICE_ID = "PREDICTOR_KEY_DEVICE_ID";
/**
* @brief 生成 batch_size 为 max_batch_size 的优化模型,单次预测图片数量可以小于或等于此值(推荐等于此值,见release notes)
* 值类型: int
* 默认值:4
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE = "PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE";
/**
* @brief 设置device对应的GPU卡可以支持的最大并发量
* 实际预测的时候对应GPU卡的最大并发量不超过这里设置的范围,否则预测请求会排队等待预测执行
* 值类型: int
* 默认值:1
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY = "PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY";
/**
* @brief 是否开启fp16模式预测,开启后预测速度会更快,但精度会略有降低。并且需要硬件支持fp16
* 值类型: bool
* 默认值: false
*/
static constexpr auto PREDICTOR_KEY_GTURBO_FP16 = "PREDICTOR_KEY_GTURBO_FP16";
```

/**

* @brief 模型编译等级

```
*1:如果当前max_batch_size与历史编译产出的max_batch_size不相等时,则重新编译模型(推荐)
```

Baidu 百度智能云文档

*2:无论历史编译产出的max_batch_size为多少,均根据当前max_batch_size重新编译模型 * 值类型: int * 默认值:1 static constexpr auto PREDICTOR_KEY_GTURBO_COMPILE_LEVEL = "PREDICTOR_KEY_GTURBO_COMPILE_LEVEL"; /** * @brief GPU工作空间大小设置 * workspace_size = workspace_prefix * (1 << workspace_offset) * workspace_offset: 10 = KB, 20 = MB, 30 = GB * 值类型: int *默认值:WORKSPACE_PREFIX: 100, WORKSPACE_OFFSET: 20,即100MB static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX = "PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX"; static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET = "PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET"; /** * @brief 需要使用的dla core * 值类型: int * 默认值:-1(不使用) static constexpr auto PREDICTOR_KEY_GTURBO_DLA_CORE = "PREDICTOR_KEY_GTURBO_DLA_CORE"; /** * @brief 自定义缓存文件存储路径 * 值类型: string *默认值:~/.baidu/easyedge/mcache/{model_id * 1000000 + release_id} static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_DIR = "PREDICTOR_KEY_GTURBO_CACHE_DIR"; /** * @brief 自定义缓存文件命名,默认即可 * 值类型: string *默认值:根据配置自动生成 static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_NAME = "PREDICTOR_KEY_GTURBO_CACHE_NAME"; /** * @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号 * 值类型: string *默认值:空 static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";

PREDICTOR_KEY_GTURBO_CACHE_NAME:首次加载模型会先对模型进行编译优化,通过此值可以设置优化后的产出文件名。

PREDICTOR_KEY_GTURBO_CACHE_DIR:首次加载模型经过编译优化后,产出的优化文件会存储在这个位置,可以按需修改。

PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX、PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET:设置运行时可以被用来使用的最大临时显存。

PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE:此值用来控制批量图片预测可以支持的最大图片数,实际预测的时候单次预测图片数需等于此值。

PREDICTOR_KEY_DEVICE_ID:设置需要使用的 GPU 卡号,对于Jetson,此值无需更改。

PREDICTOR_KEY_GTURBO_COMPILE_LEVEL: 模型编译等级。通常模型的编译会比较慢,但编译产出是可以复用的。可以在第一次加载模型的时候设置合理的 max_batch_size 并在之后加载模型的时候直接使用历史编译产出。是否使用历史编译产出可以通过此值 compile_level 来控制,当此值为 0 时,表示忽略当前设置的 max_batch_size 而仅使用历史产出(无历史产出时则编译模型);当此值为 1 时,会比较历史产出和当前设置的 max_batch_size 是否相等,如不等,则重新编译;当此值为 2 时,无论如何都会重新编译模型。

PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY:通过此值设置单张 GPU 卡上可以支持的最大 infer 并发量,其上限取决于硬件限制。init 接口 会根据此值预分配 GPU 资源,建议结合实际使用控制此值,使用多少则设置多少。注意:此值的增加会降低单次 infer 的速度,建议优先考虑 batch inference。

PREDICTOR_KEY_GTURBO_FP16:默认是 fp32 模式,置 true 可以开启 fp16 模式预测,预测速度会有所提升,但精度也会略微下降,权衡使用。注意:不是所有模型都支持 fp16 模式,也不是所有硬件都支持 fp16 模式。已知不支持fp16的模型包括:EasyDL图像分类高精度模型。

预测视频接口

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

classVideoDecoding :

```
/**
* @brief 获取输入源的下一帧
* @param frame_tensor
* @return
*/
virtual int next(FrameTensor &frame_tensor) = 0;
/**
* @brief 显示当前frame_tensor中的视频帧
* @param frame_tensor
* @return
*/
virtual int display(const FrameTensor &frame_tensor) = 0;
/**
* @brief 将当前frame_tensor中的视频帧写为本地视频文件
* @param frame_tensor
* @return
*/
virtual int save(FrameTensor &frame_tensor) = 0;
/**
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
* @brief 获取视频的width属性
* @return
 */
virtual int get_width() = 0;
/**
* @brief 获取视频的height属性
* @return
*/
virtual int get_height() = 0;
```

struct VideoConfig

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type;
                          // 输入源类型
  std::string source_value;
                         // 输入源地址, 如视频文件路径、摄像头index、网络流地址
 int skip_frames{0};
                      // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
  int retrieve_all{false}; // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
  int input_fps{0};
                      // 在采取抽帧之前设置视频的fps
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
  bool enable_display{false};
  std::string window_name{"EasyEdge"};
  bool display_all{false};
                       // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                       // 是否存储所有frame,若为false,仅存储根据skip_frames抽取的frame
  std::map<std::string, std::string> conf;
};
```

source_type: 输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value: 若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被 抽取返回,以作为显示或存储用。input_fps:用于抽帧前设置fps。resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。conf:高级选项。部分配置会通过该map来设置。

注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

```
EdgeLogConfig log_config;
log_config.enable_debug = true;
global_controller()->set_log_config(log_config);
```

返回格式

预测成功后,从 EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index; // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测或图像分割时才有
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割时才有
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding, 游程编码的mask
};
```

关于矩形坐标

x1 图片宽度 = 检测框的左上角的横坐标 y1 图片高度 = 检测框的左上角的纵坐标 x2 图片宽度 = 检测框的右下角的横坐标 y2 图片高度 = 检测框 的右下角的纵坐标

以上字段可以参考demo文件中使用opencv绘制的逻辑进行解析

http服务

1. 开启http服务 http服务的启动参考demo_serving.cpp文件。

/** * @brief 开启一个简单的demo http服务。 * 该方法会block直到收到sigint/sigterm。 * http服务里,图片的解码运行在cpu之上,可能会降低推理速度。 * @tparam ConfigT * @param config * @param host * @param port * @param service_id service_id user parameter, uri '/get/service_id' will respond this value with 'text/plain' * @param instance_num 实例数量,根据内存/显存/时延要求调整 * @return */ template<typename ConfigT> int start_http_server(const ConfigT &config, const std::string &host, int port, const std::string &service_id,

int instance_num = 1);

2. 请求http服务

开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。

URL中的get参数:

参数	说明		默认值
threshold	阈值过滤,	0~1	如不提供,则会使用模型的推荐阈值

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

python	
с#	
C++	
java	

import requests

```
with open('./1.jpg', 'rb') as f:
img = f.read()
```

```
**params 为GET参数 data 为POST Body**
result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1},
data=img).json()
```

 $\frac{\text{print}(\text{result})}{\text{result}}$

http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考接口使用·返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

返回示例

```
"cost_ms": 52,
  "error_code": 0,
  "results": [
    {
       "confidence": 0.94482421875,
       "index": 1,
       "label": "IronMan",
       "x1": 0.059185408055782318,
       "x2": 0.18795496225357056,
       "y1": 0.14762254059314728,
       "y2": 0.52510076761245728
       "confidence": 0.94091796875,
       "index": 1,
       "label": "IronMan",
       "x1": 0.79151463508605957,
       "x2": 0.92310667037963867,
       "y1": 0.045728668570518494,
       "y2": 0.42920106649398804
}
```

多线程预测 Jetson 系列 SDK 支持多线程预测,创建一个 predictor,并通过 PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY 控制所支持的最大 并发量,只需要 init 一次,多线程调用 infer 接口。需要注意的是多线程的启用会随着线程数的增加而降低单次 infer 的推理速度,建议优先使用 batch inference 或权衡考虑使用。

已知问题 1. 多线程时图片按线程分配不均 或 不同batch size的图片交叉调用infer接口时,部分结果错误

A: EasyDL图像分类高精度模型在有些显卡上可能存在此问题,可以考虑填充假图片数据到图片比较少的线程或batch以使得infer间的图片绝对平均。

2. 显存持续增长或遇到 terminate called after throwing an instance of 'std::runtime_error' what(): Failed to create object

A:如果遇到此问题,请确认没有频繁调用 init 接口,通常调用 infer 接口即可满足需求。

3. 开启 fp16 后, 预测结果错误

A:不是所有模型都支持 fp16 模式。目前已知的不支持fp16的模型包括:EasyDL图像分类高精度模型。目前不支持的将会在后面的版本陆续支持。

4. 部分模型不支持序列化

A:针对JetPack4.4、4.5版本,部分模型无法使用序列化,如已知的BML的MobileNetV1-SSD和物体检测高性能模型。需要每次加载模型的时候 编译模型,过程会比较慢。此问题将在后续JetPack版本中修复。目前JetPack4.6版本SDK已修复该问题。

开发板信息查询与设置 查询L4T或JetPack版本 查询JetPack版本信息,可以通过下面这条命令先查询L4T的版本。

在终端输入如下命令并回车 \$ head -n 1 /etc/nv_tegra_release **就会输出类似如下结果** \$ # # R32 (release), REVISION: 4.3, GCID: 21589087, BOARD: t210ref, EABI: aarch64, DATE: Fri Jun 26 04:38:25 UTC 2020

从输出的结果来看,板子当前的L4T版本为R32.4.3,对应JetPack4.4。 注意,L4T的版本不是JetPack的版本,一般可以从L4T的版本唯一对应到 JetPack的版本,下面列出了最近几个版本的对应关系:

L4T R32.6.1 --> JetPack4.6 L4T R32.5.1 --> JetPack4.5.1 L4T R32.5 --> JetPack4.5 L4T R32.4.3 --> JetPack4.4 L4T R32.4.2 --> JetPack4.4DP L4T R32.2.1 --> JetPack4.2.2 L4T R32.2.0 --> JetPack4.2.1

功率模式设置与查询不同的功率模式下,执行AI推理的速度是不一样的,如果对速度需求很高,可以把功率开到最大,但记得加上小风扇散热~

1. 运行下面这条命令可以查询开发板当前的运行功率模式 \$ sudo nvpmodel -q verbose **\$ NV Power Mode: MAXN** **\$ 0** **如果输出为MAXN代表是最大功率模式** **2. 若需要把功率调到最大,运行下面这条命令**

\$ sudo nvpmodel -m 0

如果你进入了桌面系统,也可以在桌面右上角有个按钮可以切换模式

3. 查询资源利用率 \$ sudo tegrastats

FAQ 1. EasyDL SDK与云服务效果不一致,如何处理?

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

2. 运行SDK报错 Authorization failed 日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888"
./easyedge_demo ...

3. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

4. 运行demo时报找不到libeasyedge_extension.so

需要export libeasyedge_extension.so所在的路径,如路径为/home/work/baidu/cpp/lib,则需执行:

export LD_LIBRARY_PATH=/home/work/baidu/cpp/lib:\${LD_LIBRARY_PATH}

或者在编译完后执行如下命令将lib文件安装到系统路径:

sudo make install

如不能安装,也可手动复制lib下的文件到/usr/local/lib下。

5. 运行demo时报如下之一错误

2020-12-17 16:15:07,924 INFO [EasyEdge] 547633188880 Compiling model for fast inference, this may take a while (Acceleration) Killed

或

2020-12-17 16:15:07,924 INFO [EasyEdge] 547633188880 Build graph failed

请适当降低PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE和PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY的值后尝试。

6. 运行有损压缩加速的模型,运算精度较标准模型偏低首先请保证数据集不会太小。其次可以通过将模型目录RES中的calibrationtable移除,并通过将PREDICTOR_KEY_GTURBO_FP16设置为true,使用FP16的运算精度重新评估模型效果。若依然不理想,可将calibrationtable移除并将PREDICTOR_KEY_GTURBO_FP16设置为false,从而使用更高精度的FP32的运算精度。

心物体检测辨影专用SDK集成文档

简介

本文档介绍EasyEdge/EasyDL的辨影软硬一体方案SDK的使用方法。支持的硬件包括辨影Air、辨影Pro。您可以在软硬一体方案了解部署方案。

模型支持:

• EasyDL图像:图像分类高精度,图像分类高性能,物体检测高精度,物体检测均衡,物体检测高性能

• BML :

- 公开数据集预训练模型:SSD-MobileNetV1,YOLOv3-DarkNet,YOLOv3-MobileNetV1,ResNet50,ResNet101,SE-ResNeXt50,SE-ResNeXt101,MobileNetV2,EfficientNetB0_small,EfficientNetB4,MobileNetV3_large_x1_0,ResNet18_vd,SE_ResNet18_vd,Xception71。
- 百度超大规模数据集预训练模型:YOLOv3-DarkNet, MobileNetV3_large_x1_0, ResNet50_vd, ResNet101_vd。
- EasyEdge:EasyEdge支持的模型较多,详见查看模型网络适配硬件。若模型不在此列表,可以尝试使用自定义网络生成端计算组件。

Release Notes | 时间 | 版本 | 说明 | | --- | ---- | | 2022.08.01 | 1.3.5 | 新增支持辨影软硬一体方案部署 |

辨影软件接入使用SDK 辨影Air/Pro自带软件预置了大量飞桨开源模型,支持EasyDL/BML模型SDK一键导入使用,详细的辨影使用说明见购买后 获得的使用说明书

• 辨影推理主界面



• 辨影设置界面。在应用中可选预置模型能力,也可选择EasyDL/BML导入的模型SDK



快速开始 接下来的文档内容将会描述辨影SDK的集成开发教程,仅需要使用辨影自带软件的用户无需关注

首先请在EasyDL智能云官网获取序列号。

✔ 返回EasyDL总览	く 返回 专项硬件适配服务	
EasyDL图像	按单台设备激活 按多台设备激活	
算力资源管理	(赤田光明, (屋工委委城初七升、浙江七注堂)	
公有云部署 ~	医形成吗。 (Ref) 量有这次方式、家店方店等) 提示:如您已完成企业资质认证,将享受该账号下任意部署包免费试用2个月,更多企业认证权益详见 企业权益	专属礼包 ??
EasyEdge本地部署 ^		
• 服务器纯离线服务	购买永久授权 新增测试序列号 管理序列号 查看转让序列号 批量离线激活	
• 设备端纯离线服务		
• 专项硬件纯离线	设备名 激活状态	序列号
	自定义设备 🖉 已过期	8BA2-39AA-73F
	自定义设备 🧷 未激活	2013-2CD8-458

将获取到的序列号填写到demo文件中或以参数形式传入。



默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己 的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的变量设置),需要设置额外的环境变量,指定CONTROLLER_KEY_AUTH_MODE为2, global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2),实例数鉴权模式下还支持指定license证书更新时间,单 位是秒,要求设置为大于20的整数,否则会采用默认的license更新时间,修改实例数鉴权license更新时间的方法参考 global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 300)

编译并运行Demo 模型资源文件默认已经打包在开发者下载的SDK包中。Demo工程直接编译即可运行。

编译运行:

cd src mkdir build && cd build cmake ... make **make install 为可选,也可将lib所在路径添加为环境变量** sudo make install sudo ldconfig ./demo_batch_inference/easyedge_batch_inference {模型RES文件夹} {测试图片路径或仅包含图片的文件夹路径} {序列号}

baidu@nano:~/ljay/easydl/sdk/demo/build\$./demo_batch_inference/easyedge_batch_inference ../../../RES/ /ljay/images/mix008.jpeg

2020-08-06 20:56:30,665 INFO [EasyEdge] 548125646864 Compiling model for fast inference, this may take a while (Acceleration) 2020-08-06 20:57:58,427 INFO [EasyEdge] 548125646864 Optimized model saved to:

/home/baidu/.baidu/easyedge/jetson/mcache/24110044320/m_cache, Don't remove it

Results of image /ljay/images/mix008.jpeg:

- 2, kiwi, p:0.997594 loc: 0.352087, 0.56119, 0.625748, 0.868399
- 2, kiwi, p:0.993221 loc: 0.45789, 0.0730294, 0.73641, 0.399429
- 2, kiwi, p:0.992884 loc: 0.156876, 0.0598725, 0.3802, 0.394706
- 1, tomato, p:0.992125 loc: 0.523592, 0.389156, 0.657738, 0.548069
- 1, tomato, p:0.991821 loc: 0.665461, 0.419503, 0.805282, 0.573558
- 1, tomato, p:0.989883 loc: 0.297427, 0.439999, 0.432197, 0.59325
- 1, tomato, p:0.981654 loc: 0.383444, 0.248203, 0.506606, 0.400926
- 1, tomato, p:0.971682 loc: 0.183775, 0.556587, 0.286996, 0.711361
- 1, tomato, p:0.968722 loc: 0.379391, 0.0386965, 0.51672, 0.209681 Done

demo运行示例:



测试Demo HTTP 服务

编译demo完成之后,会同时生成一个http服务,运行

./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401} ./easyedge_serving ../../../RES "1111-1111-1111" 0.0.0.0 24401

日志中会显示

HTTP is now serving at 0.0.0.0:24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。



同时,可以调用HTTP接口来访问服务,具体参考下文接口说明。

使用说明

使用该方式,将运行库嵌入到开发者的程序当中。

使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

Baidu 百度智能云文档

```
// step 1: 配置模型运行参数
EdgePredictorConfig config;
config.model_dir = model_dir;
config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, serial_num);
config.set_config(params::PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE, 1); // 优化的模型可以支持的最大batch_size,实际单次推理的图
片数不能大于此值
config.set_config(params::PREDICTOR_KEY_GTURBO_FP16, false); // 置true开启fp16模式推理会更快,精度会略微降低,但取决于硬
件是否支持fp16,不是所有模型都支持fp16,参阅文档
config.set_config(params::PREDICTOR_KEY_GTURBO_COMPILE_LEVEL, 1); // 编译模型的策略,如果当前设置的max_batch_size与历史编
译存储的不同,则重新编译模型
// step 2: 创建并初始化Predictor
auto predictor = global_controller()->CreateEdgePredictor<EdgePredictorConfig>(config);
if (predictor->init() != EDGE_OK) {
  exit(-1);
// step 3-1: 预测图像
auto img = cv::imread({图片路径});
std::vector<EdgeResultData> results;
predictor->infer(img, results);
// step 3-2: 预测视频
std::vector<EdgeResultData> results;
FrameTensor frame_tensor;
VideoConfig video_config;
video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h
video_config.source_value = video_src;
/*
... more video_configs, 根据需要配置video_config的各选项
*/
auto video_decoding = CreateVideoDecoding(video_config);
while (video_decoding->next(frame_tensor) == EDGE_OK) {
 results.clear();
 if (frame_tensor.is_needed) {
    predictor->infer(frame_tensor.frame, results);
    render(frame_tensor.frame, results, predictor->model_info().kind);
  //video_decoding->display(frame_tensor); // 显示当前frame, 需在video_config中开启配置
  //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置
```

初始化接口

```
auto predictor = global_controller()->CreateEdgePredictor<EdgePredictorConfig>(config);
if (predictor->init() != EDGE_OK) {
    exit(-1);
}
```

若返回非0,请查看输出日志排查错误原因。

预测接口

```
/**
* @brief
* 单图预测接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
    cv::Mat& image, std::vector<EdgeResultData>& result
) = 0;
/**
* @brief
* 批量图片预测接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
    std::vector<cv::Mat>& image,
     std::vector<std::vector<EdgeResultData>>& results
) = 0;
/**
 * @brief
* 批量图片预测接口,带阈值
 * @related infer(cv::Mat & image, EdgeColorFormat origin_color_format, std::vector<EdgeResultData> &result, float threshold)
*/
virtual int infer(
       std::vector<cv::Mat> &images,
       EdgeColorFormat origin_color_format,
       std::vector<std::vector<EdgeResultData>> &results,
       float threshold
) = 0;
```

图片的格式务必为opencv默认的BGR, HWC格式。

批量图片的预测接口的使用要求在调用 init 接口的时候设置一个有效的 max_batch_size,其含义见下方参数配置接口的介绍。

参数配置接口 参数配置通过结构体 EdgePredictorConfig完成。

```
struct EdgePredictorConfig {
    /**
     * @brief 模型资源文件夹路径
    */
    std::string model_dir;
    std::map<std::string, std::string> conf;
    EdgePredictorConfig();
    template<typename T>
    T get_config(const std::string &key, const T &default_value);
    template<typename T = std::string>
    T get_config(const std::string &key);
    template<typename T>
    const T *get_config(const std::string &key, const T *default_value);
    template<typename T>
    void set_config(const std::string &key, const T &value);
    template<typename T>
    void set_config(const std::string &key, const T *value);
    static EdgePredictorConfig default_config();
  };
运行参数选项的配置以key、value的方式存储在类型为std::map的conf中,并且键值对的设置和获取可以通
```

过EdgePredictorConfig的set_config和get_config函数完成。同时也支持以环境变量的方式设置键值对。 EdgePredictorConfig的具体使用方法可以 参考开发工具包中的demo工程。

针对Jetson开发工具包,目前EdgePredictorConfig的运行参数所支持的Key包括如下项:

```
/**
* @brief 当有同类型的多个设备的时候,使用哪一个设备,如:
* GPU: 使用哪张GPU卡
* EdgeBoard(VMX), Movidius NCS:使用哪一张加速卡
* 值类型: int
* 默认值:0
*/
static constexpr auto PREDICTOR_KEY_DEVICE_ID = "PREDICTOR_KEY_DEVICE_ID";
/**
* @brief 生成最大 batch_size 为 max_batch_size 的优化模型,单次预测图片数量可以小于或等于此值
* 值类型: int
* 默认值:4
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE = "PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE";
/**
* @brief 设置device对应的GPU卡可以支持的最大并发量
* 实际预测的时候对应GPU卡的最大并发量不超过这里设置的范围,否则预测请求会排队等待预测执行
* 值类型: int
* 默认值:1
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY = "PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY";
/**
* @brief 是否开启fp16模式预测,开启后预测速度会更快,但精度会略有降低。并且需要硬件支持fp16
* 值类型: bool
* 默认值: false
*/
static constexpr auto PREDICTOR_KEY_GTURBO_FP16 = "PREDICTOR_KEY_GTURBO_FP16";
/**
* @brief 模型编译等级
```

*1:如果当前max_batch_size与历史编译产出的max_batch_size不相等时,则重新编译模型(推荐)

Baidu 百度智能云文档

*2:无论历史编译产出的max_batch_size为多少,均根据当前max_batch_size重新编译模型 * 值类型: int * 默认值:1 static constexpr auto PREDICTOR_KEY_GTURBO_COMPILE_LEVEL = "PREDICTOR_KEY_GTURBO_COMPILE_LEVEL"; /** * @brief GPU工作空间大小设置 * workspace_size = workspace_prefix * (1 << workspace_offset) * workspace_offset: 10 = KB, 20 = MB, 30 = GB * 值类型: int *默认值:WORKSPACE_PREFIX: 100, WORKSPACE_OFFSET: 20,即100MB static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX = "PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX"; static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET = "PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET"; /** * @brief 需要使用的dla core * 值类型: int * 默认值:-1(不使用) static constexpr auto PREDICTOR_KEY_GTURBO_DLA_CORE = "PREDICTOR_KEY_GTURBO_DLA_CORE"; /** * @brief 自定义缓存文件存储路径 * 值类型: string * 默认值: ~/.baidu/easyedge/mcache/{model_id * 1000000 + release_id} static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_DIR = "PREDICTOR_KEY_GTURBO_CACHE_DIR"; /** * @brief 自定义缓存文件命名,默认即可 * 值类型: string *默认值:根据配置自动生成 static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_NAME = "PREDICTOR_KEY_GTURBO_CACHE_NAME"; /** * @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号 * 值类型: string *默认值:空 static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";

PREDICTOR_KEY_GTURBO_CACHE_NAME:首次加载模型会先对模型进行编译优化,通过此值可以设置优化后的产出文件名。

PREDICTOR_KEY_GTURBO_CACHE_DIR:首次加载模型经过编译优化后,产出的优化文件会存储在这个位置,可以按需修改。

PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX、PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET:设置运行时可以被用来使用的最大临时显存。

PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE:此值用来控制批量图片预测可以支持的最大图片数,实际预测的时候单次预测图片数不可大于此值,但可以是不大于此值的任意图片数。

PREDICTOR_KEY_GTURBO_COMPILE_LEVEL: 模型编译等级。通常模型的编译会比较慢,但编译产出是可以复用的。可以在第一次加载模型的时候设置合理的 max_batch_size 并在之后加载模型的时候直接使用历史编译产出。是否使用历史编译产出可以通过此值 compile_level 来控制,当此值为 0 时,表示忽略当前设置的 max_batch_size 而仅使用历史产出(无历史产出时则编译模型);当此值为 1 时,会比较历史产出和当前设置的 max_batch_size 是否相等,如不等,则重新编译;当此值为 2 时,无论如何都会重新编译模型。

PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY:通过此值设置单张 GPU 卡上可以支持的最大 infer 并发量,其上限取决于硬件限制。init 接口 会根据此值预分配 GPU 资源,建议结合实际使用控制此值,使用多少则设置多少。注意:此值的增加会降低单次 infer 的速度,建议优先考虑 batch inference。

PREDICTOR_KEY_GTURBO_FP16:默认是 fp32 模式,置 true 可以开启 fp16 模式预测,预测速度会有所提升,但精度也会略微下降,权衡使用。注意:不是所有模型都支持 fp16 模式,也不是所有硬件都支持 fp16 模式。已知不支持fp16的模型包括:EasyDL图像分类高精度模型。

预测视频接口

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

classVideoDecoding :

```
/**
* @brief 获取输入源的下一帧
* @param frame_tensor
* @return
*/
virtual int next(FrameTensor & frame_tensor) = 0;
/**
* @brief 显示当前frame_tensor中的视频帧
* @param frame_tensor
* @return
*/
virtual int display(const FrameTensor &frame_tensor) = 0;
/**
* @brief 将当前frame_tensor中的视频帧写为本地视频文件
* @param frame_tensor
* @return
*/
virtual int save(FrameTensor &frame_tensor) = 0;
/**
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
* @brief 获取视频的width属性
 * @return
 */
virtual int get_width() = 0;
/**
* @brief 获取视频的height属性
* @return
*/
virtual int get_height() = 0;
```

struct VideoConfig
```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type;
                          // 输入源类型
  std::string source_value;
                         // 输入源地址, 如视频文件路径、摄像头index、网络流地址
 int skip_frames{0};
                      // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
  int retrieve_all{false}; // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
  int input_fps{0};
                      // 在采取抽帧之前设置视频的fps
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
  bool enable_display{false};
  std::string window_name{"EasyEdge"};
 bool display_all{false};
                       // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                       // 是否存储所有frame,若为false,仅存储根据skip_frames抽取的frame
  std::map<std::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被 抽取返回,以作为显示或存储用。input_fps:用于抽帧前设置fps。resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。conf:高级选项。部分配置会通过该map来设置。

注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

EdgeLogConfig log_config; log_config.enable_debug = true; global_controller()->set_log_config(log_config);

返回格式

预测成功后,从 EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index; // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测或图像分割时才有
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割时才有
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding, 游程编码的mask
};
```

关于矩形坐标

x1 图片宽度 = 检测框的左上角的横坐标 y1 图片高度 = 检测框的左上角的纵坐标 x2 图片宽度 = 检测框的右下角的横坐标 y2 图片高度 = 检测框 的右下角的纵坐标

以上字段可以参考demo文件中使用opencv绘制的逻辑进行解析

http服务

/**

1. 开启http服务 http服务的启动参考demo_serving.cpp文件。

* @brief 开启一个简单的demo http服务。 * 该方法会block直到收到sigint/sigterm。

* http服务里,图片的解码运行在cpu之上,可能会降低推理速度。 * @tparam ConfigT * @param config * @param host * @param port * @param service_id service_id user parameter, uri '/get/service_id' will respond this value with 'text/plain' * @param instance_num 实例数量,根据内存/显存/时延要求调整 * @return */ template<typename ConfigT> int start_http_server(const ConfigT &config, const std::string &host, int port, const std::string &service_id, int instance_num = 1);

2. 请求http服务

开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。

URL中的get参数:

参数	说明		默认值
threshold	阈值过滤,	0~1	如不提供,则会使用模型的推荐阈值

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

Python请求示例

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
        'http://127.0.0.1:24401/',
        params={'threshold': 0.1},
        data=img).json()
```

Java请求示例参考这里

http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考接口使用·返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

返回示例

```
"cost_ms": 52,
"error_code": 0,
"results": [
    "confidence": 0.94482421875,
    "index": 1
    "label": "IronMan",
    "x1": 0.059185408055782318,
    "x2": 0.18795496225357056,
    "y1": 0.14762254059314728,
     "v2": 0.52510076761245728
    "confidence": 0.94091796875,
    "index": 1.
    "label": "IronMan",
    "x1": 0.79151463508605957,
    "x2": 0.92310667037963867,
    "y1": 0.045728668570518494,
     "y2": 0.42920106649398804
```

多线程预测 辨影系列 SDK 支持多线程预测,创建一个 predictor,并通过 PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY 控制所支持的最大并 发量,只需要 init 一次,多线程调用 infer 接口。需要注意的是多线程的启用会随着线程数的增加而降低单次 infer 的推理速度,建议优先使用 batch inference 或权衡考虑使用。

已知问题 1. 多线程时图片按线程分配不均 或 不同batch size的图片交叉调用infer接口时,部分结果错误

A:EasyDL图像分类高精度模型在有些显卡上可能存在此问题,可以考虑填充假图片数据到图片比较少的线程或batch以使得infer间的图片绝对平均。

2. 显存持续增长或遇到 terminate called after throwing an instance of 'std::runtime_error' what(): Failed to create object

A:如果遇到此问题,请确认没有频繁调用 init 接口,通常调用 infer 接口即可满足需求。

3. 开启 fp16 后, 预测结果错误

A:不是所有模型都支持 fp16 模式。目前已知的不支持fp16的模型包括:EasyDL图像分类高精度模型。目前不支持的将会在后面的版本陆续支持。

4. 部分模型不支持序列化

A:针对JetPack4.4版本,部分模型无法使用序列化,如已知的BML的MobileNetV1-SSD和物体检测高性能模型。需要每次加载模型的时候编译模

型,过程会比较慢。此问题将在后续JetPack版本中修复。

开发板信息查询与设置查询L4T或JetPack版本查询JetPack版本信息,可以通过下面这条命令先查询L4T的版本。

在终端输入如下命令并回车 \$ head -n 1 /etc/nv_tegra_release **就会输出类似如下结果** \$ # # R32 (release), REVISION: 4.3, GCID: 21589087, BOARD: t210ref, EABI: aarch64, DATE: Fri Jun 26 04:38:25 UTC 2020

从输出的结果来看,板子当前的L4T版本为R32.4.3,对应JetPack4.4。 注意,L4T的版本不是JetPack的版本,一般可以从L4T的版本唯一对应到 JetPack的版本,下面列出了最近几个版本的对应关系:

L4T R32.6.1 --> JetPack4.6 L4T R32.5.1 --> JetPack4.5.1 L4T R32.5 --> JetPack4.5 L4T R32.4.3 --> JetPack4.4 L4T R32.4.2 --> JetPack4.4DP L4T R32.2.1 --> JetPack4.2.2 L4T R32.2.0 --> JetPack4.2.1

功率模式设置与查询不同的功率模式下,执行AI推理的速度是不一样的,如果对速度需求很高,可以把功率开到最大,但记得加上小风扇散热~

```
**1. 运行下面这条命令可以查询开发板当前的运行功率模式**

$ sudo nvpmodel -q verbose

**$ NV Power Mode: MAXN**

**$ 0**

**如果输出为MAXN代表是最大功率模式**

**2. 若需要把功率调到最大,运行下面这条命令**

$ sudo nvpmodel -m 0
```

如果你进入了桌面系统,也可以在桌面右上角有个按钮可以切换模式

3. 查询资源利用率 \$ sudo tegrastats

FAQ 1. EasyDL SDK与云服务效果不一致,如何处理?

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

2. 运行SDK报错 Authorization failed 日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888"
./easyedge_demo ...

3. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

4. 运行demo时报找不到libeasyedge_extension.so

需要export libeasyedge_extension.so所在的路径,如路径为/home/work/baidu/cpp/lib,则需执行:

export LD_LIBRARY_PATH=/home/work/baidu/cpp/lib:\${LD_LIBRARY_PATH}

或者在编译完后执行如下命令将lib文件安装到系统路径:

sudo make install

如不能安装,也可手动复制lib下的文件到/usr/local/lib下。

5. 运行demo时报如下之一错误

2020-12-17 16:15:07,924 INFO [EasyEdge] 547633188880 Compiling model for fast inference, this may take a while (Acceleration) Killed

或

2020-12-17 16:15:07,924 INFO [EasyEdge] 547633188880 Build graph failed

请适当降低PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE和PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY的值后尝试。

6. 运行有损压缩加速的模型,运算精度较标准模型偏低首先请保证数据集不会太小。其次可以通过将模型目录RES中的calibrationtable移除,并 通过将PREDICTOR_KEY_GTURBO_FP16设置为true,使用FP16的运算精度重新评估模型效果。若依然不理想,可将calibrationtable移除并 将PREDICTOR_KEY_GTURBO_FP16设置为false,从而使用更高精度的FP32的运算精度。

⊙ 浏览器或小程序部署

心 浏览器或小程序部署

浏览器或小程序部署

简介 本文档介绍EasyDL的浏览器/小程序部署SDK的使用方法,

SDK支持范围 浏览器部署

PC浏览器: Chrome、Safari、Firefox 手机浏览器: Baidu App、Safari、Chrome、UC and QQ Browser

小程序部署

小程序:百度小程序、微信小程序

支持的操作系统

系统: MacOS、Windows

demo文件结构 SDK解压缩之后,目录结构如下

public
model
model.json
chunk_n.dat
src
components
App.vue
config.json
env.d.ts
label.json
main.ts
modelInfo.json
usePredict.ts
index.html
package.json
READ.md
tsconfig.json
tsconfig.node.json
vite.config.ts
varn.lock

demo基于vite,其中public/model下的model.json、chunk_1.dat...chunk_n.dat为模型文件,src下为业务代码,index.html为入口文件

快速开始 依赖node及npm,如果没有node,请前往node官网下载长期维护版本

安装依赖:npm install

启动项目:npm run dev

vite v2.8.4 dev server running at:

- > Local: http://localhost:3000/
- > Network: use `--host` to expose

到浏览器打开 http://localhost:3000/ 即可体验demo

模型预测结果示例 图像分类示例

 $[0.4450492858886719,\, 0.3961234986782074,\, 0.0122891990467906,\, 0.14653800427913666]$

数组的index为对应的标签,值为置信度

物体检测示例

[[1, 0.2247152328491211, 0.11200979351997375, 0.07523892819881439, 0.8540866374969482, 0.5503567457199097], [2, 0.1224712328491211, 0.511200979351997375, 0.27523892819881439, 0.8540866374969482, 0.5503567457199097],...]

输出结果是一个二维数组,第二维的结果为:[标签 , 置信度 , 矩形框x1坐标, 矩形框y1坐标, 矩形框x2坐标, 矩形框y2坐标]

浏览器开发

参考src/usePredict文件

```
// 加载推理引擎
import {Runner, env} from '@paddlejs/paddlejs-core';
// 使用webgl计算方案(暂不能使用wasm、webgpu等计算方案)
import '@paddlejs/paddlejs-backend-webgl';
...
// 注册引擎
const runner = new Runner({
     modelPath: '/model',
     keepRatio: config.rescale_mode === 'keep_ratio',
     mean: config.img_mean.reduce((memo, v) => [...memo, +((v / 255).toFixed(3))], [] as number[]),
     std: config.scale.reduce((memo, v) => [...memo, +((1 / 255 / v).toFixed(3))], [] as number[]),
     bgr: config.colorFormat === 'BGR',
     feedShape: {
       fw: config.resize[0],
       fh: config.resize[1]
     }
  });
...
// init runner
await runner.init();
// predict and get result
await runner.predict(img);
```

更多可参考PaddleJS工程页

小程序开发

微信小程序

```
微信小程序需添加 Paddle.js微信小程序插件
步骤:
小程序管理界面 --> 设置 --> 第三方设置 --> 插件管理 --> 添加插件 --> 搜索 wx7138a7bb793608c3 并添加
```

掌上百度小程序

手百小程序需添加paddlejs百度智能小程序动态库**引入动态库代码包** 代码示例:

```
{

"dynamicLib": {

// 定义一个别名,小程序中用这个别名引用动态库。

"paddlejs": {

"provider": "paddlejs"

}

}
```

使用动态库

在使用页面的json文件里配置如下信息:

```
{
    "usingSwanComponents": {
        "paddlejs": "dynamicLib://paddlejs/paddlejs"
    }
}
```

从而页面中可以使用此组件:

```
<view class="container">
<view>下面这个自定义组件来自于动态库</view>
<paddlejs />
</view>
```

示例

index.swan

```
<view class="container">
<!--index.wxml-->
<image style="width:100%; height: 300px; " src="{{imgPath}}"></image>
<button bindtap="chooselmage">选择图片</button>
<button bindtap="doPredict" class="btn" type="primary">新鲜度预测</button>
<!-- 返回结果 -->
<view class="result" s-if="resultType">预测结果 : {{resultType}}</view>
<view class="result" s-if="resultVal">预测可信度 : {{resultType}}</view>
<view class="result" s-if="resultVal">预测可信度 : {{resultType}}</view>
</view class="result" s-if="resultVal">预测可信度 : {{resultVal}}</view>
</view class="result" s-if="resultVal">{view>
</view class="result" s-if="resultVal">{view>
</view>
```

index.js

```
Page({
 data: {
    imgPath: '',
    content: ",
    resultType: ",
    resultVal: ",
    isShow: true,
    options: { // 模型配置项
      modelPath: 'http://localhost:3000/model',
      fileCount: 3,
      needPreheat: true,
      feedShape: {
         fw: 224,
         fh: 224
      },
      fetchShape: [1, 7, 1, 1],
      fill: [255, 255, 255, 255],
      scale: 256,
      targetSize: { height: 224, width: 224 },
      mean: [0.485, 0.456, 0.406],
      std: [0.229, 0.224, 0.225]
   },
    status: '' // 初始值为'', 变为'predict'时会触发模型预测
 },
```

/**

```
* 选择图片
   */
  chooseImage: function () {
     const me = this;
     this.setData({
        ishow: false
     });
     swan.chooseImage({
        count: 1,
        sizeType: ['original', 'compressed'],
        sourceType: ['album', 'camera'],
        success(res) {
           const path = res.tempFilePaths[0];
           swan.getFileSystemManager().readFile({
             filePath: path,
             encoding: 'base64',
             success: res => {
                me.setData({
                   imgBase64: res && res.data,
                   imgPath: path
                });
             },
             fail: res => {
                console.log(res);
             }
          });
        }
     });
  },
  predict(e) {
     const status = e && e.detail && e.detail.status;
     if (status === 'loaded') {
        this.setData({status: 'loaded', isShow: false});
     }
     else if (status === 'complete') {
        const data = e.detail.data;
        const maxItem = this.getMaxItem(data);
        this.setData({status: '', resultType: maps[maxItem.index], resultVal: maxItem.value});
     }
  },
  doPredict() {
     this.setData({status: 'predict'});
  },
  getMaxItem(datas = []) {
     let max = Math.max.apply(null, datas);
     let index = datas.indexOf(max);
     return {value: max, index};
  },
});
```

Prop

名称	类型	默认值	是否必 选	描述
options	string		是	模型配置项,参考src/usePridict
imgBase 64	string		是	要预测的图像的base64
status	string		是	当前状态,status变化触发组件调用相应的api,当status变为predict时,组件会读取imgBase64作为输入 的图像,调用模型预测APi

⊙ 智能边缘控制台-单节点版

₯ EasyEdge 智能边缘控制台-单节点版 IEC

EasyEdge Intelligent Edge Console (以下简称IEC) 是EasyEdge推出的边缘设备管理的本地化方案。可以运行于多种架构、多系统、多类型的终端之上。通过IEC,用户可以方便地在本地进行

- EasyDL/BML/EasyEdge的SDK的 离线 / 在线激活,服务管理
- 接入本地和远程摄像头,网页中实时预览
- 自动监控和记录相关事件
- 硬件信息的可视化查看

支持的系统+CPU架构包括:

- Windows x86_64 (Windows 7 ~ Windows 10, 暂不支持Windows 11)
- Linux x86_64 / arm32 / arm64

支持各类常见的AI加速芯片,包括:

- NVIDIA GPU / Jetson 系列
- Baidu EdgeBoard FZ系列
- 比特大陆 Bitmain SC / SE 系列
- 华为 Atlas 系列
- 寒武纪 MLU 系列
- 其他EasyDL/EasyEdge/BML支持的AI芯片

完整列表可参考<mark>这里</mark>

Release Note

注意:2.0.0之后,默认以系统服务形式安装iec,无法兼容1.x版本的iec

版本号	发布时间	更新说明
2.2.0	2022-10-27	新增onvif/gb28181支持;完善端云通信逻辑
2.0.0	2022-03-22	支持连接中心节点IECC;支持以系统服务安装
1.0.2	2021-12-22	更新视频预览推流库;新增若干AI芯片支持;支持多种芯片温度、功耗展示;多项性能优化
1.0.0	2021-09-16	IEC 第一版!

快速开始

从这里选择您需要的操作系统和CPU架构下载:

- Windows amd64 : intel、AMD的64位x86_84 CPU
- Linux amd64: intel、AMD的64位x86_84 CPU
- Linux arm : 树莓派等32位的ARM CPU
- Linux arm64: RK3399、飞腾等64位的ARM CPU

或者从纯离线服务管理页可下载智能边缘控制台

<i>と</i> 楽EasyDL	产品介绍	操作平台	应用案例	使用文档
物体检测模型	∢☴ 纯蔷	线服务		
创建模型				
训练模型	Ŕ	吨离线服务说明		
校验模型	3 11	ξ布纯离线服务,料 测过API.SDK进一	将训练完成的相 步集成,灵活	具型部署在本地,离线调用模型。可以选择将模型部署在本地的服务器、小型设备、软硬一体方案专项适配硬件上, 适应不同业务结署:您可下载EasyEdge智能边缘控制台,便接管理多个本地模型服务。
发布模型	^	45 ± 06 82 de	1000-1000 1000-1000	
☑ EasyData数据服务		及扣制服务	省市公式工作	
数据总览				
标签组管理	服	务器 通用小型	型设备 专	项适配硬件
在线标注		SDK API		
智能标注				
云服务数据回流	此	处发布、下载的S	DK为未授权S	DK,需要前往控制台 <mark>获取序列号</mark> 激活后才能正式使用。SDK内附有对应版本的Demo及开发文档,开发者可参考源代码完成开发,

您也可以通过先安装多节点版本IECC,通过中心节点来自动连接安装边缘节点。

Linux 安装 解压缩之后,目录结构如下

0 EasyEdge-IEC-v2.0.0-linux-amd64 > tree .
easyedge-iec
easyeuge-lec-setup.sh
easyedge-iec.service-conf.init.d easyedge-iec.service-conf.systemd
easyedge-iec.service-conf.upstart
easyedge-lec.service.ymi
readme.txt
1 directory, 8 files
以系统服务形式安装(推荐)以root用户运行./easyedge-iec-setup.sh install即可
[setup]: sudo could not be found

- [setup]: Start to install IEC ... [setup]: + bash -c "cp easyedge-iec /usr/sbin/easyedge-iec" [setup]: + bash -c "chmod +x /usr/sbin/easyedge-iec" [setup]: + bash -c "cp etc/easyedge-iec.service.yml /etc/easyedge-iec/easyedge-iec.yml" [setup]: + bash -c "cp etc/easyedge-iec.service-conf.init.d /etc/init.d/easyedge-iec" [setup]: + bash -c "chmod +x /etc/init.d/easyedge-iec" [setup]: Install IEC success! [setup]: + bash -c "service easyedge-iec start" Starting easyedge-iec: success [setup]: Start to check IEC status... [setup]: + bash -c "curl -s 127.0.0.1:8702 >/dev/null" [setup]: IEC status: OK! [easyedge-iec]: default configure file: /etc/easyedge-iec/easyedge-iec.yml [easyedge-iec]: default log file: /var/log/easyedge-iec/easyedge-iec.log [easyedge-iec]: service usage: service easyedge-iec { start | stop } [setup]: Done!
- 日志: /var/log/easyedge-iec/easyedge-iec.log
- 系统配置: /etc/easyedge-iec/easyedge-iec.yml
- 服务启动/停止: service easyedge-iec { start | stop } (不同操作系统内可能不同,具体命令参考安装日志)

自定义安装 (不推荐) 自定义安装方法仅限于 安装脚本无法识别的情况。

- 拷贝 ./EasyEdge-IEC-v2.0.0/ 整个目录至自定义文件夹,如/opt/EasyEdge-IEC
- 进入到 /opt/EasyEdge-IEC
- 通过 nohup 等方法运行 ./easyedge-iec-linux-{您的系统架构} amd64: intel、AMD的64位x86_84 CPU arm:树莓派等32位的ARM CPU * arm64: RK3399、飞腾等64位的ARM CPU
- 日志: ./log/easyedge-iec.log
- 系统配置: ./easyedge-iec.yml

Windows 安装

解压缩之后,安装目录如下所示:

0 tmp2 > tree EasyEdge-IEC-v2.0.0-windows-amd64
EasyEdge-IEC-v2.0.0-windows-amd64
easyedge-iec.exe
easyedge-iec-setup.bat
├─── etc
easyedge-iec.yml
└─── readme.txt

- 1 directory, 4 files
- 打开命令行(非powershell)运行 easyedge-iec-setup.bat install。

A-IX	ijnang _{1±}			叩令门 肌直 应列之口	,1177/パリル3省	とり りりり	http://(区面的:070	2/easyeuge/lec kp
Admini	istrator: C:\\	Windows\sy	vstem32\cmd	.exe		(四)		>
crosoft) 2018 \Users\	t Windows Microsof \eval>_	s [Versio Ft Corpor	n 10.0.17 ation。保	763.2145] 留所有权利。	-	2005((R) 移动(M) 大小(S) 最小化(N) 最大化(X) 关闭(C) 編輯(E) 默认值(D) 属性(P)	古 键	
"C:\V	Windows	system	32\cmd.e	xe"		×		
远坝	字体	布局	颜色					
光 () () () ()	际大小 小(S) 中(M) 大(L)			命令记录 缓冲区大小(B): 缓冲区数量(N):	50 ×			
	構选项 快速編辑 插入模式 启用 Ctrl 粘贴时筛	模式(Q) (I) 键快捷方 选剪贴板((K) 内容(F)	取消勾选				
	+	SHITC / V	/101 H25090	(THATHS KALERE(C)				
	⊕_22]¥ 启用对选 扩展的文	定内容自: 本选择键(动换行(E) (X)					
当前	前代码页							
65	001 (UTF	F-8)						
□ 使 了 了解	5月旧版招 了解有关 更多有关	注制台(U)(日版控制台 新控制台	需要重新。 <u>合模式</u> 的详 功能的信息	自动,影响所有控制台) 细信息				
2 /21.	20 B/							

启动之后,打开浏览器,访问 http://{设备ip}:8702/easyedge 即可:



默认用户名密码为 admin / easyedge

EasyEdge 智能边	缘控制;	台			
总览					
③ 本地模型服务 添加模型服务		┃基本信息 本和会ainm12_tig	海你系统 linux MAC18414	6-92-5648-82-47 IDHH	⊮ 10.232.101.23.10
模型服务管理		王······ · · · · · · · · · · · · · · · ·	指纹文件 下载指纹文件 所属用户		
② 我的摄像头 添加摄像头		┃ 资源利用情况			
摄像头管理		F 00/	ET ON	EE ON	
本地事件 本地事件记录		5.9 % CPU利用率	57.8% _{运行内存}	55.8% 储存空间	
 ② 设置 系统设置 		CPU型号 Intel(R) X CPU核数 48核	运行内存 125.64GB 已用内存 72.60GB	总空间 16.39TB 已用空间 9.15TB	
运行日志		┃ 本地项目管理			
		模型服务管理 >	序列号管理 >	摄像头管理	本地事件记录
		3 0 数量 运行中	0 0 数量 有效期内	2 2 数量 运行中	16742

功能使用说明

①添加模型服务 首先,点击导航栏的「本地模型服务」-「添加模型服务」。在页面中定义服务名称后,将已经下载好的Linux/Windows版本的 SDK与IEC关联。关联完毕后可按两种激活方式,激活使用SDK。

EasyEdge 智能边缘	腔制	台				操作指南	a) admin 🗸
总览		添加模型服务					
⑦本地模型服务 添加模型服务		■新建模型服务 *服务名称	9 请输入名称				
模型服务管理 ② 我的摄像头 添加摄像头		* SDK ② 激活方式	土 选择SDK 联网激活	高线激活			
摄像头管理			请先上传SDK				
🖾 本地事件		服务配置					
本地事件记录							
 ② 设置 系统设置 			参数配置				
运行日志			添加并激活				
新大昌·1/100							

部分SDK需要提前安装系统依赖,如TRT等,具体请参考EasyDL/BML/EasyEdge SDK使用文件中的环境依赖安装说明

联网激活

1. 在关联SDK完成后,需要在百度智能云控制台对应部署方式管理页中新增测试序列号或购买正式序列号。(图中以服务器版SDK为例)

✓ 返回EasyDL总览	产品服务 / EasyDL图像 -	品服务 / EasyOL图像 - 服务器纯属线服务										
EasyDL图像	服务器纯离线服务	服务器纯属线服务管理										
公有云部署 🗸	API SDR	c										
EasyEdge本地部署 へ												
• 服务器结束线服务	使用说明: (展开查	2看部署介绍及操作步骤等	0						◇展开			
• 设备储结离线服务	十 购买正式授权	新培测试序列号 ?							开发文档 工单支持			
。 专项硬件纯离线	设备名	技术方向	模型ID	可激活模型数	序列号类型	序列号	激活时间	到期日期	操作			
	自定义设备 🖉	四倍		3	试用		未激活	2021-12-31	如何激活			
	自定义设备 🖉	图像		3	试用		未激活	2021-12-31	如何激活			
	自定义设备 🖉	田僚		3	试用		未激活	2021-12-31	如何激活			

2. 再在IEC中填入所申请的序列号

总宽	添加模型服务	
③ 本地模型服务	新建模型服务	
添加模型服务	*服务名称	zrf 蜀誌
模型服务管理	* 80K @	* 12X007/
◎ 我的摄像头	- 304 0	iii Lesefage-Linux-m462366.zip
添加摄像头	****	
摄像头管理	AC/11/130	
🖆 本地事件	_	Model ID 46266
本地事件记录		* 序列号
© 设置		
系统设置	服务配置	*Host 0.0.0
运行日志		*Port 24401
		9%E1
		<i>"</i>
		<i>添加井客</i> 活
	•	
版本号: V1.0.0		

- 3. 配置服务,在服务端口不冲突占用的情况下,使用默认即可
- 4. 添加并激活

离线激活

1. 在IEC总览页面下载「指纹文件」

EasyEdge 智能边缘把	制台	操作指南	admin ~
总览			
① 本地模型服务 へ 添加模型服务 模型服务管理	基本信息 主約名 路作系统 Inux MAC地址		
② 我的摄像头 ∧ 添加摄像头	资源利用情况		
摄像头管理 □ 本地事件 ^	6.4% счиляя ціляя (1.6%)		
★地事件に来 ○ 设置 へ 系统设置	CPU型号 Intel(Pi)X 运行內存 251.3156 总空间 13.87718 • CPU核数 28號 已用內存 11.5608 已用空间 10.23718		

2. 在百度智能云的控制台中找到SDK对应的管理列表,图中以服务器SDK为例。申请序列号后,点击对应序列号尾部的「离线激活」操作,按指

	< 返回EasyDL总宽	产品服务 / EasyDL图像 - 版	务器纯离线服务								
	EasyDL图像	服务器纯离线服务	管理								
	公有云部署 >	API SDK									
	EasyEdge本地部署 へ										
	• 服务器纯高线服务	使用说明: (展开查:	音部署介绍及操作步骤等)								√ 展开
	• 设备端纯离线服务	十 购买正式授权	新增测试序列号 ?							开发文档	工单支持
	• 专项硬件纯离线	设备名	技术方向	模型ID	可激活模型数	序列号类型	序列号	激活时间	到期日期	操作	
己的注		自定义设备 🧷	25:10		1	试用		2021-03-09	永久	申请延期	
51/啟/占		自定义设备 🗷	图像		10	试用		2021-09-16	2021-10-16	申请延期	
		自定义设备 🧷	田像		10	试用	1.	未激活	激活后30天内有效	联网激活 离约	\$.敢活
		自定义设备 🧷	图像		10	试用		未激活	激活后30天内有效	联网激活丨离约	教活
		自定义设备 🖉	图像			试用		未激活	激活后31天内有效	联网激活丨南纬	教派
		自定义设备 🗷	田像			试用		未激活	激活后31天内有效	联网激活丨离约	\$.数活
									每页显示	10 🗸 < 1	•
											\odot
											-

3. 在IEC的添加模型服务页面,上传下载好的授权文件,完成激活

E	EasyEdge 智能泛	边缘控制	旧台					操作指南	a admin 🗸
	总览		添加模型服务						
¢	本地模型服务		新建模型服务						
			*服务名称						
	模型服务管理								
Q	我的摄像头			EasyEdge-Linux-m462x86.zip					
	添加摄像头		激活方式	联网激活 离线激活					
	摄像头管理			服务器版本激活流程					
Ē	本地事件			1. 在当前页面下载"指纹文件"并	2.在百度智能云控制台上传指的 文件,并填写Model ID和	3.在百度智能云控制台下载授权	4.在当前页面上传"授权文件"并 洲运用到924		
	本地事件记录			C30MODel ID-51Melease ID	Release ID		80.10 90.12 10 37		
0	设置			指纹文件 下载指纹文件					
	系统设置			Release ID 102355 0					
	运行日志			*授权文件 上选择授权文件					
			服务配置						
				参数配置					
				漆加并激活					

激活完成后即可在「模型服务管理」列表中启动服务,使用后续的操作栏功能。

体验本地demo

```
点击「本地demo体验」即可在立即上传图片进行预测
```

【物体检测】97741 e7.5检测高性能–Auto分辨率V1		
	label	置信度
2		
		请上传文件后查看识别结果

接入摄像头

使用接入摄像头功能首先需要添加摄像头,请参考第②步,完成后按照第③步操作 **注:服务启动后也可参考「模型发布」模块的技术文档进行开** 发使用,本文档主要介绍IEC使用功能

EasyEdge 智能记	四緣控制	制台						操作指	南 🧧 admin 🗸
总览			模型服务管理						
③ 本地模型服务									
添加模型服务			新编版务 宣信序列号						
			服务名称	模型名称	应用平台	服务状态	服务开关	操作	
◎ 我的摄像头			晓宇加速版	e7.5检测高性能–Auto分 耕率V1	Linux-通用X86 CPU加速	运行中		查看详情 体验本地demo 接入摄作	铁 删除
添加摄像头			zyf_胡乱测试模型	测试20210902V1	Linux-通用X86 CPU	运行中			
摄像头管理									
🗈 本地事件									
© 设置									
系统设置									
运行日志									
版本号: V1.0.0									

激活错误码

错误码	文案	描述
4001	parameters missing. 参数缺失	
4002	parameters invalid. 参数不合法	
4003	model invalid. 模型信息不合法	
4004	no more activation left. 该序列号和该设备的激活次数超上限	
4005	the serial key is out of date. 该序列号过期	
4006	the serial key has been activated. 该序列号已被其他设备激活	该序列号已被其他设备激活,不能重复激活。
4007	account invalid. 序列号不能用于其他账号的模型	序列号不能用于其他账号的模型,只能用于绑定账号的模型。
4008	serial key invalid. 序列号不合法	序列号不存在或找不到
4009	bundle id invalid. 包名不合法	
4010	product invalid. 产品不合法	如easydl的SDK使用BML的序列号来激活,会报该错误
4011	platform invalid. 平台不合法	
4012	activate too frequent. 激活太频繁	激活太频繁,请稍后再进行激活。
4013	device type and license type not match. 硬件类型和序列号类型不匹配	如使用加速版序列号激活基础版SDK会报该错误
4014	exceed max activate device num. 超过最大授权数量	
4015	technology invalid. 技术类型不合法	
4016	exceed max activate entity num. 超过最大模型数量	
4017	device invalid. 设备不合法	
4018	model invalid. 模型不合法	

②添加摄像头 导航栏点击「我的摄像头」-「添加摄像头」,定义摄像头名称、备注后即可添加摄像头。支持本地摄像头和网络摄像头。 摄像头 添加成功后即可设置摄像头的运行时间和频率

EasyEdge 智能边缘控制	制台	操作指南	a admin ~
总宽	汤加温囊头		
 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	* 福佳头名称		
受 我的摄像头 ∧ 恣加摄像头 恣加摄像头	摄像头类型 ◆ 本地描像头 ○ 网络摄像头 请选择细像头 ◇		
捕像头茸茸 本地事件 へ 本地事件記录 (0) 设置 へ 系统设置 运行日志	* 紙行時詞 000000 至 000000 + 非項ー項 通動的率 個項 1 0 0 0 相取 1 0 温馨提示:1 這行時间回動的原準的回動服務方法接通像央流, 常介推塑結果的运行方式, 推理結果可在進等中记录中違客。 2. 此处护设置不影响服像失实时预定服务通道。 捕象头突时预定可在描像头管理模块宣音		
版本号: V1.0.0			

③摄像头接入模型服务预测点击「本地模型服务」-「模型服务管理」中,所需接入预测的服务的「接入摄像头」

EasyEdge 智能边	2缘控制台	3					操作指南	a admin ~	
总宽		模型服务管理							
③ 本地模型服务 添加模型服务		新增服务 查看序列号							
模型服务管理		服务名称	模型名称	应用平台	服务状态	服务开关	操作		在弹出
◎ 我的摄像头		晓宇加速版	e7.5检测高性能Auto分 耕率V1	Linux-通用X86 CPU加速	运行中		查看详情 体验本地demo 接入摄像头		
添加摄像头		zyf_胡乱测试模型	测试20210902V1	Linux-通用X86 CPU	运行中				
摄像头管理									
🗈 本地事件									

在弹出的弹窗中选择第②

步中添加的摄像头,此时点击确认即可在「摄像头管理」中的实时预览功能中查看摄像头预测结果,识别结果默认不保存。如需保存识别结果, 可设置对应的「本地事件触发条件」,根据标签和置信度,将识别结果保存至本地事件记录当中。设置多个标签条件时,IEC会以"或"的逻辑来将 所有满足条件的识别结果保存

接入摄像头		×
某个ip-camera	~ 添加	
摄像头名称	本地时间触发条件 52%	操作
某个ip-camera	标签 tomato 置信度 52 52 52 52 52 52 52 52 52 52 53 53 53 53 53 53 54 55 54 55	%~100% × 取消接入
	确定取消	

④本地事件 点击导航栏「本地事件记录」,可通过服务名称、摄像头名称、事件记录的时间、标签及置信度来筛选识别结果查看,多个标签及置 信度同样也是"或"的逻辑记录。如有想要删除的事件数据可选择后删除,全选为本页全选。





与中心节点连接之后,边缘节点主程序版本会自动随控制中心版本升级。 (>2.0.0)

• Step 1 在IECC中添加边缘节点,选择「边缘节点已安装IEC」,并记录IP地址与Token

EasyEdge 智能边缘控制	16-3节点版
总览	边缘节点 > 添加边缘节点
② 模型服务中心 模型服务管理 下发模型服务	添加边缘节点 ② 注册边缘节点 2 意适边缘节点
	激活方式 ○ 边缘带点来安裂EC
	请复制中心节点IP地址、Token密码、在边缘设备安装的"IEC-系统设置页面"进行激活:操作完成后请在下方进行校验、校验通过即成功激活边缘节点。
摄像头设备 回 事件中心 事件记录	□ IP地址 10.232.191.23.8602
	Token b00258b3-6c04-4696-b4dc-e9dd83de941d
系统设置	上一步 校验漱活结果

• Step 2 在IEC的系统设置中打开多节点模式,并填入刚才记录的IP地址与Token,点击建立连接

EasyEdge 智能边缘控制	制中心
总览	系统设置
⑦ 本地模型服务 添加模型服务 模型服务管理	*事件保存时限 30 <a>>>> ▼ ★ *文件保存路径 /opt/baidu/easyedge/iec/events
	系统时间显示 ③
添加摄像头	多节点模式 ③
摄像头管理 闫 本地事件 本地事件记录	中心节点连接状态 • 未连接 (IP地址和Token, 请在"多节点版"的"添加边缘节点"第二步中查看) 输入IP地址
运行日志	

• 连接完成后即可在中心节点IECC去监控/管理/应用在边缘节点上的IEC

Baidu 百度智能云文档

配置项*

listonHost: 127 0 0 1

配置文件etc/easyedge-iec.yml中有关于IEC的各项配置说明,一般无需修改,请确保理解配置项含义之后,再做修改。

```
###### IEC系统配置
###### ------ 高级配置一般无需修改 -------
#######!!!注意!!!请确保理解配置项含义后再做修改
version: 3
com:
# hub: 作为中心节点模式启动。 edge: 作为子节点启动
# role: edge
#硬件利用率刷新时间间隔: 过低的刷新间隔可能会消耗CPU资源。
hardwareUsageRefreshSecond: 60
#事件监测触发扫描周期
eventTriggerIntervalSecond: 10
# IEC保存SDK等资源的路径:可填写 "default", 也可以直接填写绝对路径
appDataFolder: default
# 是否开启DEBUG模式:开启之后,将会打印大量日志,便于追踪系统问题
debug: no
# 是否需要用户名/密码登陆,强烈建议打开!(默认用户名/密码为 admin/easyedge)
loginEnabled: yes
#登录有效时间,单位秒
sessionMaxAge: 3600
logging:
# 是否把日志打印到控制台
toStd: no
# 是否把日志输出到文件。关闭后,将无法在页面中正确显示系统日志
toFile: yes
loggingFile: /var/log/easyedge-iec/easyedge-iec.log
# 0:info; -1:debug; -2:verbose
level: -1
webservice
# WEB服务的监听端口
listenPort: 8702
listenHost: 0.0.0.0
sdk:
# GPU SDK所使用的cuda版本:9/10/10.2/11.0/11.1。请安装完cuda之后,这设置正确的版本号。
cudaVersion: 10.2
# AI服务启动时,额外配置的 LD_LIBRARY_PATH(linux) 或者 PATH(windows)
libPath: ./
# AI服务启动时,额外配置的其他环境变量。
ENVs:
 EDGE_CONTROLLER_KEY_LOG_BRAND: EasyEdge
###### EDGE_CONTROLLER_KEY_XXX: XXXX
commu:
# 普通消息等待respond的超时时间
respondWaitTimeoutSecond: 2
###### 数据库相关配置
db:
sqliteDbFile: /var/lib/easyedge-iec/easyedge-iec.db
hubDbFile: /var/lib/easyedge-iec/easyedge-iec.hub.db
eventDbFile: /var/lib/easyedge-iec/easyedge-event.db
fileServerDbFile: /var/lib/easyedge-iec/easyedge-fileserver.hub.db
nodeMonitorDbFile: /var/lib/easyedge-iec/easyedge-nodemonitor.hub.db
###### 推流相关配置
mediaserver:
flvPort: 8715
rtmpPort: 8716
###### 视频流相关配置
edgestream:
logLevel: -1
```

1131011103L 121.0.0.1

listenPort: 8710 #摄像头预览:识别结果绘制延迟消失 renderExtendFrames: 10 # 预测队列大小: 如果设置为60,当摄像头fps=30时,视频延迟约为2秒。 降低inferenceQueueSize可以降低预览延迟,但是根据硬件的算 力情况,可能导致模型推理速度跟不上,没有识别结果,不建议设置太低 inferenceQueueSize: 60 videoEncodeBitRate: 400000 #视频采样 & 视频实时预览分辨率设置 # 0: auto, 1: 1080p, 2: 720p, 3: 480p, 4: 360p, 5: 240p resolution: 0 # 内置多媒体服务配置 # port设为0表示关闭 mediaServerHost: 127.0.0.1 mediaServerFlvPort: 8713 mediaServerRtmpPort: 8714 mediaServerRtspPort: 0

FAQ 启动服务后,进程中出现两个easyedge-iec进程这是正常现象,IEC通过守护进程的方式来完成更新等操作。

启动服务时,显示端口被占用port already been used 通过修改 easyedge-iecc.yml文件的配置后,再重新启动服务。

安装服务时,报错permission denied 请以管理员身份运行安装程序。

中心节点重启后,边缘节点IEC 一直离线中心节点短时间的离线,边缘节点会自动重连。如果中心节点已经恢复在线,边缘节点长时间未自动连接上,可通过边缘节点iec的方法来重新连接(右上角 admin - 重启系统)

IEC 是否有Android / iOS 版本 我们将会在近期发布对Android操作系统的支持

添加SDK时,报错 SDK不支持该硬件。 SDK not supported by this device 一般是因为使用的SDK跟硬件不匹配,如 GPU的SDK,硬件没有GPU 卡。 对于Jetson,也可能是Jetpack版本不支持,可以通过查看 本机Jetpack版本和SDK支持的Jetpack版本列表(cpp文件中的文件名来查看)来 匹配。

⊙ 智能边缘控制台-多节点版

₯ EasyEdge 智能边缘控制台——多节点版

整体介绍

智能边缘控制台 - 多节点版(EasyEdge Intelligent EdgeConsole Center 以下简称IECC),是EasyEdge推出的边缘资源管理、服务应用与管理一站 式本地化方案。

通过IECC,用户可以方便地在中心节点管理子节点:

- 边缘硬件资源的管理与监控
- EasyDL/BML/EasyEdge的SDK的 离线 / 在线激活,服务管理
- 视频流解析, 接入本地和远程摄像头, 网页中实时预览
- 自动监控和记录相关视频流推理事件

支持的系统+CPU架构包括:

- Windows x86_64 (Windows 7 ~ Windows 10, 暂不支持Windows 11)
- Linux x86_64 / arm32 / arm64

支持各类常见的AI加速芯片,包括:

- NVIDIA GPU / Jetson 系列
- Baidu EdgeBoard FZ系列
- 比特大陆 Bitmain SC / SE 系列
- 华为 Atlas 系列
- 寒武纪 MLU 系列
- 其他EasyDL/EasyEdge/BML支持的AI芯片

连接说明 以下为 中心节点 (控制中心),边缘节点/子节点,摄像头的连接示意:



其中:

- 控制中心需要有固定IP,而边缘节点可以处于多级子网之下,只需IEC能够主动访问到控制中心节点即可
- 模型服务均运行于各边缘节点之上
- 摄像头均与边缘节点相连

Release Note

版本号	发布时间	更新说明
2.2.0	2022-10-27	边缘节点新增Android支持;新增onvif/gb28181支持;优化端云通信通道安全
2.0.0	2022-03-25	多节点版上线!
1.0.2	2021-12-22	更新视频预览推流库;新增若干AI芯片支持;支持多种芯片温度、功耗展示;多项性能优化
1.0.0	2021-09-16	智能边缘控制台 - 单节点版 IEC 第一版!

安装从这里选择您需要的操作系统和CPU架构下载:

- Windows amd64 : intel、AMD的64位x86_84 CPU
- Linux amd64: intel、AMD的64位x86_84 CPU
- Linux arm : 树莓派等32位的ARM CPU
- Linux arm64: RK3399、飞腾等64位的ARM CPU

或者从纯离线服务管理页可下载智能边缘控制台

ど桨EasyDL ご桨EasyDL	产品介绍 操作平台 应用案例 使用文档
物体检测模型	三 纯嘉线服务
创建模型	
训练模型	纯离线服务说明
校验模型	发布纯高线服务,将训练完成的模型部署在本地,高线调用模型。可以选择将模型部署在本地的服务器,小型设备、软硬一体方案专项适配硬件上, 通过API、SDK讲一步集成,显示该应不同业务场局、您可下着EasyEcton智能均缘控制点,便遵管理多个太抽模型服务。
发布模型	
☑ EasyData数据服务	
数据总览	
标签组管理	服务器 通用小型设备 专项适配硬件
在线标注	SDK API
智能标注	
云服务数据回流	此处发布、下载的SDK为未授权SDK,需要前往控制台获取序列号激活后才能正式使用。SDK内附有对应版本的Demo及开发文档,开发者可参考源代码完成开发。

以Linux为例,解压缩后目录结构如下所示:

./EasyEdge-IECC-v{版本号}/

- |-- easyedge-iecc
- |-- easyedge-iecc-setup.sh
- |-- etc/
- |-- etc/easyedge-iec.yml
- -- readme.txt

Linux 系统

通过系统服务形式安装 (推荐)

以管理员运行bash easyedge-iecc-setup.sh install 即可。

0 EasyEdge-IEC-v2.0.0 > bash ./easyedge-iecc-setup.sh install [setup]: sudo could not be found [setup]: Start to install IECC... [setup]: + bash -c "cp easyedge-iec-linux-amd64 /usr/sbin/easyedge-iecc" [setup]: + bash -c "chmod +x /usr/sbin/easyedge-iecc" [setup]: + bash -c "cp easyedge-iec-* /var/lib/easyedge-iecc/fs/tmp" [setup]: + bash -c "cp etc/easyedge-iecc.service.yml /etc/easyedge-iecc/easyedge-iecc.yml" [setup]: + bash -c "cp etc/easyedge-iecc.service-conf.init.d /etc/init.d/easyedge-iecc" [setup]: + bash -c "chmod +x /etc/init.d/easyedge-iecc" [setup]: Install IECC success! [setup]: + bash -c "service easyedge-iecc start" Starting easyedge-iecc: success [setup]: Start to check IECC status... [setup]: + bash -c "curl -s 127.0.0.1:8702 >/dev/null" [setup]: IECC status: OK! [easyedge-iecc]: default configure file: /etc/easyedge-iecc/easyedge-iecc.yml [easyedge-iecc]: default log file: /var/log/easyedge-iecc/easyedge-iecc.log [easyedge-iecc]: service usage: service easyedge-iecc { start | stop } [setup]: Done!

出现sucess字样,表示安装成功。

- 日志: /var/log/easyedge-iecc/easyedge-iecc.log
- 系统配置: /etc/easyedge-iecc/easyedge-iecc.yml
- 服务启动/停止: service easyedge-iecc { start | stop } (不同操作系统内可能不同,具体命令参考安装日志)
- 配置服务自启动:可根据不同操作系统参考这里进行对应配置

可通过 bash easyedge-iecc-setup.sh uninstall 来卸载,以及bash easyedge-iecc-setup.sh upgrade来升级为当前安装包的版本

自定义安装 (不推荐)

自定义安装仅限于 安装脚本无法识别您的操作系统的情况。

- 拷贝 ./EasyEdge-IEC-v2.0.0/ 整个目录至自定义文件夹,如/opt/EasyEdge-IEC
- 进入到 /opt/EasyEdge-IEC
- 通过 nohup 等方法运行 ./easyedge-iec-linux-{您的系统架构} --com.role=hub amd64: intel、AMD的64位x86_84 CPU arm:树莓派等32位的 ARM CPU * arm64: RK3399、飞腾等64位的ARM CPU
- 日志: ./log/easyedge-iecc.log
- 系统配置: ./easyedge-iecc.yml

Windows 系统 打开命令行(非powershell)运行 easyedge-iecc-setup.bat install。

注:如果遇到hang住的情况,可修改命令行配置

Х

rosoft	strator: C:\l	Nindows\sy	/stem32\cm	d.exe				A	- L
crosoft Windows [Version 10.0.17763.2145]) 2018 Microsoft Corporation。保留所有权利。						还原(R) 移动(M)		右键	
Users\	eval>_				_	天小(S) 最小化(N)			
					•	最大化(X)			
					×	关闭(C)			
						编辑(E) 默认值(D)	>		
						属性(P)	-		
"C:\\	Vindows	system	32\cmd.e	exe" 属性		×			
选项	字体	布局	颜色						
	-+								
70/	***			もより発		-			
•	小(S)			缓冲区大小(B):	50				
0	∓(M) ±(I)			缓冲区数量(N):	4				
0.				□丢弃旧的副本(D)		_			
编辑	副洗顶								
		#-+*(O)	_	取消勾选					
	犬迷珊瑚	関エレ(Q) 、							
		() (紳士)津古	=t(k)						
		REINGLE/J	内容(F)						
	钻贴时篮	洗胆素板							
	粘贴时筛) 将 Ctrl+S	选剪贴板 hift+C/V	/用作复制	刂/粘贴的快捷键(C)					
	粘贴时筛) 将 Ctrl+S t洗择	选剪贴板 hift+C/V	/用作复制	引/粘贴的快捷键(C)					
	粘贴时筛) 将 Ctrl+S 本选择 自用对洗:	选剪炻板 hift+C/V 完内窗户	/ 用作复制	刂/粘贴的快捷键(C)					
	站贴时筛) 将 Ctrl+S 体选择 自用对选 扩展的文	选剪贴板 ihift+C/V 定内容自: 本洗择键	/ 用作复制 动换行(E) (X)	训/粘贴的快捷键(C)					
	結贴时筛) 将 Ctrl+S 体选择 自用对选 扩展的文	选购贴板 Shift+C/V 定内容自: 本选择键	/ 用作复制 动换行(E) (X)	刂/粘贴的快速键(C)					

验证安装: 启动之后,打开浏览器,访问 http://{设备ip}:8602/easyedge 即可:



更新服务:关闭服务,下载最新的安装包,重新执行安装流程即可。

注:1.中心节点更新到新版之后,已连接的边缘节点会自动跟随中心节点,自我升级到同样的版本。

2. 报错: Text file busy. 一般是因为服务没有停止。

使用流程 Step 1 注册并激活边缘节点

• 在IECC导航栏中点击边缘节点,点击页面中的添加边缘节点按钮

EasyEdge智能边缘控制	EasyEdge智能边缘控制台-多节点版								操作指南 🧧 adm
总宽	边缘节点								
	已添加的边缘节点							添加边缘节点	
下发模型服务	边缘节点名称	节点状态	节点标识	CPU	AI芯片	模型服务数量	摄像头数量	服务添加时间	操作
⊜ 边缘资源中心	x86-test	 在线 		Intel(R) Core(T					
边缘节点 摄像头设备		• 离线	dry_persona L_cpu	Intel(R) Xeon(R				2022-03-23 19:30:00	

- 注册边缘节点,填写基本信息
- 激活边缘节点,根据边缘节点上是否安装智能边缘控制台-单节点版(IEC)分两种激活方式
 - 边缘节点未安装IEC:复制提供的命令,在边缘节点的终端中输入执行(命令会自动在当前目录,下载单节点版IEC并注册到控制中心)。终端命令执行完成后,在下方校验激活结果,如结果通过即可完成边缘节点的激活

添加边缘节	ち点	
> 注册边约	%节点 2 激活边缘节点	
激活方式	● 边缘节点未安装IEC) 边缘节点已安装IEC	
	请复制以下命令,粘贴在边缘设备terminal执行进行激活;在边缘节点操作完成后可在下方进行校验,校验通过即成功激活边缘	节点。
	Windows Linux	
	上一步 校验激活结果	

• 边缘节点已安装IEC:记录页面中提供的IP地址和Token

添加边缘节		
✓ 注册边维	8 节点 2 激活边缘节点	
激活方式	○ 边缘节点未安装IEC ○ 边缘节点已安装IEC	
	请复制中心节点IP地址、Token密码,在边缘设备安装的"IEC~系统设置页面"进行激活;操作完成后请在下方进行校验,校验	通过即成功激活边缘节点。
	IP地址 10.187.123.18:8702 Token 94766edf-868c-46a1-9549-bc27b760b17a	0
	上一步 校验激活结果	

• 在边缘节点的IEC-系统设置中,打开多节点模式开关,将刚才记录的IP地址和Token填入其中,建立连接

总览	系统设置
☞ 本地模型服务	
添加模型服务 模型服务管理	*文件保存路径 /opt/baidu/easyedge/iec/events
	系统时间显示 ⑦
添加摄像头	多节点模式 ③
摄像头管理	
	中心节点连接状态 • 未连接 (IP地址和Token, 请在"多节点版"的"添加边缘节点"第二步中查看)
本地事件记录	输入IP地址 10.232.191.23:8602
② 设置	Token •••••
系统设置	建立连接
运行日志	

• 成功激活后可在边缘节点页面中看到一行状态为在线的记录

EasyEdge智能边缘控制	台–多节点版								操作指南	a) admin v
总览	边缘节点									
② 模型服务中心 模型服务管理	已添加的边缘节点							添加边缘节点		
下发模型服务	边缘节点名称	节点状态	节点标识	CPU	AI芯片	模型服务数量	摄像头数量	服务添加时间	操作	
目 边缘资源中心		 在 结 		Intel(R) Xeon(R						

Step 2 上传并下发模型服务

• 在模型服务管理-已添加的模型服务页面中点击添加模型服务

EasyEdge智能边缘控制	台多节点版						操作指南	a admin ~
总宽	模型服务管理							
◎ 模型服务中心 模型服务管理	已下发的模型服务 💽	已添加的模型服务				添加模型服务		
下发模型服务 〇 边缘资源中心	服务名称	模型名称	应用平台	已部署节点数量	添加时间	操作		
	cpu 加速版模型	e7.5检测高性能-Aut						

• 上传来自于EasyDL/BML的SDK,目前仅支持Windows/Linux的SDK

EasyEdge智能边缘控制台–多节点版							
总览	模型服务管理	> 新建模型服务					
✿ 模型服务中心	添加模型服績	务					
模型服务管理	★叩夕勾狗	建态》之为					
下发模型服务	^ 服穷名称	「					
🖯 边缘资源中心	* SDK ?	♪ 选择SDK					
边缘节点		添加	取消				
摄像头设备							

- 添加成功后可在已添加的模型服务页面查看添加的模型服务SDK
- 在模型服务SDK上传成功以及边缘节点也添加激活过后,即可将模型服务下发至边缘。点击导航栏-下发模型服务,选择已添加的模型服务, 选择下发的目标节点(支持多节点批量下发)进行模型服务下发

EasyEdge智能边缘控制	11台-多节点版	操作指南	a) admin 🗸
总苋	8673221 2) 10.623221		
⑦ 模型服务中心 模型服务管理	这样明显思考		
下发模型服务 曰 边缘资源中心	Myfeith PAR CPU Linux 前通版 入解研探索等点 Q 已逸节点(0)		
边缘节点	市点名称 连接状态 标识 市点名称 连接状态 标识 のcharge=08 本 広体		
摄像头设备 (1) 事件中心			
事件记录			
◎ 设置 系统设置			

确定下发配置后,填入模型服务在边缘节点联网激活运行的序列号(支持批量导入)即可完成模型服务下发,序列号可在智能云控制台获取。
 离线激活的过程可参考IECC中的具体指引

EasyEdge智能边缘控制台—多节点版						
总览	模型服务管理 > 下发模型服务					
② 模型服务中心 模型服务管理 下发模型服务	下发模型服务					
目 边缘资源中心	激活方式 💿 联网激活 🔿 高线激活	批量导入序列号 @				
边缘节点 摄像头设备	边缘节点1: docker-x86	请输入序列号激活				
E) 事件中心 事件记录						

• 完成上述流程后即可在模型服务管理-已下发的模型服务列表中查看记录,并进行下一步应用功能体验

注:完成此步骤后即可在边缘节点进行二次集成已下发的模型服务,具体的集成方式可在文档·某图像任务类型-模型发布中查找对应的 SDK开发文档进行集成开发

EasyEdge智能边缘控制	台-多节点版							操作指南	a) admin 🗸
总览	模型服务管理								
○ 模型服务中心 模型服务管理	已下发的模型服务	已添加的模型服务						下发模型服务 请输入模型服务名称	
下发模型服务	服务名称	模型名称	服务所在边缘节点	应用平台	服务状态	漆加时间	服务开关	操作	
送 辺藤贡源中心 边織苔貞		e7.5检测高性能			 通行中 				

下发时可以通过高级配置设置服务运行的host和port。若不设置,默认host为0.0.0.0,port为系统随机分配的可用端口

Step 3 配置摄像头

Step 3 - 5 描述的是如何使用IECC可视化进行视频流式推理与应用,对此有需求的用户建议详细查看后续步骤内容。如仅需对下发的模型服务进行二次集成的用户无需进行后续操作,参考SDK对应的开发文档进行集成即可

• 首先需要确定边缘节点已经接入物理摄像头,可通过USB插口接入,也可通过RTSP/RTMP流式协议接入。在摄像头设备页面点击添加摄像头 按钮,填写对应的信息添加摄像头。支持设置摄像头的运行时间以及摄像头的抽帧频率

EasyEdge智能边缘控制	台多节点版		操作指南	a admin v
总宽	摄像头设备 > 添加摄1	像 头		
☆ 模型服务中心	添加摄像头			
模型服务管理				
ト友検望服务				
边缘节点				
摄像头设备				
當 事件中心				
事件记录		○ 本地頭像头 ● 网络摄像头		
© 1211	-	★接续说他是 支持RTSP、RTMP、如://sp///admin:1234		
系统设置				
运行日志				
		油碱铁平 每期11 1 帧		
		溫馨提示: 1. 造行时间和抽帧频率的设置抽题务连接册像头后,保存推理给果的运行方式。推理结果可在事件记录中重着。 2. 此处的设置不影响描像头突时预定服务推理,抽像头突时预定可在抽像头管理模块重着		
		(1) (1) (1) (1) (1) (1) (1) (1) (1) (1)		

• 添加完成后可在摄像头设备页面查看记录

EasyEdge智能边缘控制	台-多节点版					操作指南	a) admin ∨
	摄像头设备						
⑦ 模型服务中心 模型服务管理	已添加的摄像头				添加損錄头	请输入摄像头名称	
下发模型服务	摄像头名称	摄像头标识	连接状态	所在节点名称	摄像头开关	操作	

• 点击预览可查看摄像头预览画面



Baidu 百度智能云文档

• 模型服务可接入摄像头直接进行预测,并可同时设置告警规则,出发告警条件的结果将会以事件的形式保存至IECC中。

品山侯至加劳官·庄贝间十列应加劳的按八旗像大保下									
EasyEdge智能边缘控制台	台-多节点版							操作指南	a) admin 🗸
总览	模型服务管理								
◎ 模型服务中心	口下令的道利服务								
模型服务管理		CANADIMENS							
下发模型服务	服务名称	模型名称	服务所在边缘节点	应用平台	服务状态	添加时间	服务开关	操作	
□ 边缘资源中心 边缘劳点	CPU Linux 普	e7.5检测高性能						查看详情 体验demo 接入摄像头 删除	
ALL NO IS THE									

 将已添加至IECC的摄像头与模型服务关联,并在下方设置对应的事件告警条件。告警规则通过标签阈值的方式来建立,例如设置"猕猴桃"标签 阈值80%-100%,则大于80%置信度的"猕猴桃"识别结果将会保存至事件记录中

已下发的模型服	接入摄像头			
服务名称	请选择摄像头	· 添加 這應: 仅可选择模型服务所在边缘节点的摄像头		
CPU Linux 普.	ip cam		操作	
		标签 kiwi ∨ 0 % ~ 100% ×		
	ip cam	标签 请选择标签 ∨ 置信度 ● 0 % ~ 100% ×		
		确定	取 消	

• 也可在摄像头设备页面-实时预览中查看实时的模型服务预测结果

摄像头实时预览				
原始摄像头 CPU Linux 普通版				
		预测标签	置信度 🍦	
	2022-03-23 12:52:14	2. tomato	96.74%	
197377 0.660349	3.	3. kiwi	96.73%	
tomato: 0.71534 (matabili 0.727542)		1. 荔枝	96.23%	
		2. tomato	95.24%	
		2. tomato	94.90%	
		2. tomato	91.59%	
		2. tomato	90.72%	



• 可在事件中心-事件记录中查看满足时间告警条件的图片记录

EasyEdge智能辺缘控制	台−多节点版	操作指南 🦲 admin 🗸
总览	本地事件记录	
⑦ 模型服务中心 模型服务管理	法释边情节点 dockerx86 ◇ 选择组务 请选择服务 ◇ 选择组像头 请选择服像头 ◇ 时间范围 开始 -	
下发模型服务	- 标志消遣 ~	
目 边缘资源中心	事件概范 事件自动前断 ② 🦰	全选 首 删除
边缘节点		
摄像头设备		
□ 事件中心		
事件记录	新无数联	

高级配置说明 在系统设置 - 高级,可以修改控制中心的高级系统配置

IECc系统配置 version: 3 com: # hub: 作为中心节点模式启动。 edge: 作为子节点启动 role: hub #硬件利用率刷新时间间隔: 过低的刷新间隔可能会消耗CPU资源。 hardwareUsageRefreshSecond: 60 # IECC保存SDK等资源的路径:可填写 "default",也可以直接填写绝对路径 appDataFolder: default # 是否开启DEBUG模式:开启之后,将会打印大量日志,便于追踪系统问题 debug: no # 是否需要用户名/密码登陆,强烈建议打开!(默认用户名/密码为 admin/easyedge) loginEnabled: yes #登录有效时间,单位秒 sessionMaxAge: 3600 logging: # 是否把日志打印到控制台 toStd: no # 是否把日志输出到文件。关闭后,将无法在页面中正确显示系统日志 toFile: yes loggingFile: /var/log/easyedge-iecc/easyedge-iecc.log # 0:info; -1:debug; -2:verbose level: -1 webservice: #WEB服务的监听端口 listenPort: 8602 listenHost: 0.0.0.0 commu: mqServer: host: 0.0.0.0 port: 8632 HTTPPort: 8620 maxPayload: 8388608 pingIntervalSecond: 30 # 普通消息等待respond的超时时间 respondWaitTimeoutSecond: 2 nodeRefreshIntervalSecond: 30 ###### ------- 以下高级配置一般无需修改 -----------###### !!!注意!!! 请确保理解配置项含义后再做修改 ###### 数据库相关配置 db: sqliteDbFile: /var/lib/easyedge-iecc/easyedge-iecc.db hubDbFile: /var/lib/easyedge-iecc/easyedge-iecc.hub.db eventDbFile: /var/lib/easyedge-iecc/easyedge-event.db fileServerDbFile: /var/lib/easyedge-iecc/easyedge-fileserver.hub.db nodeMonitorDbFile: /var/lib/easyedge-iecc/easyedge-nodemonitor.hub.db ###### 推流相关配置 mediaserver: flvPort: 8613 rtmpPort: 8614 ###### 文件服务器相关配置

fileserver: root: /var/lib/easyedge-iecc/fs

FAQ

启动服务后,进程中出现两个easyedge-iec进程这是正常现象,IEC通过守护进程的方式来完成更新等操作。

启动服务时,显示端口被占用port already been used 通过修改 easyedge-iecc.yml文件的配置后,再重新启动服务。

安装服务时,报错permission denied 请以管理员身份运行安装程序。

添加SDK时,报错 SDK不支持该硬件。 SDK not supported by this device 一般是因为使用的SDK跟硬件不匹配,如 GPU的SDK,硬件没有GPU 卡。 对于Jetson,也可能是Jetpack版本不支持,可以通过查看 本机Jetpack版本和SDK支持的Jetpack版本列表(cpp文件中的文件名来查看)来 匹配。

∞ 模型加速整体说明

功能简介 当您发布时纯离线服务时,平台已结合最新的量化、剪枝、蒸馏技术,推出丰富的模型压缩加速方案,以提高您的SDK部署效率。

覆盖范围:服务器、通用小型设备、专项适配硬件均支持该功能。

具体原理:针对目标芯片,对模型做深度优化压缩加速,加速后模型在推理速度、内存占用、体积大小等指标上表现更优。发布加速模型可能需要一段时间,同时会有微小的精度损失。发布完成后可通过性能报告对比具体加速效果。

使用流程 选择加速方式 结合选择的系统与芯片不同,分别为您提供不同的压缩方式。

纯离线服务 >	发布新服务	操作文档 教学视频 常见问题 提交工单
服务器 ji 1 选择部署开	 適用小型设备 专项适配硬件 形式 (2) 填写个人信息 	说明: 1.本地服务冠部署支持该借型部署于本地約CPU、GPU服务器 上,提供APRISOK两种集成方式: 查看文档
集成方式	SDK API	2.本地服务器SDK: 将模型封装成适配本地服务器 (支持Linux 和Windows)的SDK,可集成在其他程序中运行。首次联网激 活后即可纯离线运行,占用服务器资源更少,使用方法更灵活
选择模型 选择版本	dog-cat-test V V2 V	3.集成步骤:①申请SDK并在服务详情页面下载SDK →②在 控制台申请邀活序列号 →③根据开发文信集成SDK,并联网 激活使用。如存在设备无法联网,需要在纯高线的环境下激活 的情况,请提交工单联系到()。
选择系统和芯片	 ▲ Linux 通用X86 CPU 延行为 Atlas 300 百度 昆仑XPU 雪く Windows 	4.个人信息的填写仅供EasyDL团队了解忽,便于后续合作接 治,不会作为其他用途使用
模型加速 ⑦	◎ 基础-无加速 无加速 充加速	
	☐ 精度做损压缩加速-中 施时免费 在部分达片上,内容/指律台周围拖低,推理 速度可以获得一定进升	
	<u></u> ⊤–步	

提示:基础SDK默认作为勾选项存在,可后续与您的加速SDK进行效果与性能比对,方便您进一步挑选

查看发布状态 点击完成发布后,将自动跳转至列表页,可分别查看不同加速方案下的模型发布进度及发布时间。

服务器 通用小型设	输入模型名称	Q							
SDK API									
此处发布、下载的SDK	此处发布、下载的SDK为未授权SDK,需要前往控制台获取序列与激活后才能正式使用。SDK内附有对应版本的Demo及开发文档,开发者可参考源代码完成开发。								
模型名称	发布版本 关	应用平台	模型加速	发布状态	发布时间 关	操作			
	115215-V2		基础版	•发布中	2021-05-13 20:49	下载SDK			
dog-cat-test		精度无损压缩加速	•发布中	2021-05-13 20:49	下载加速版SDK				

常见问题

心数据相关问题

需要上传多少张图片才能训练出效果较好的模型?

• 每种要识别的物体在所有图片中出现的数量需要大于50。如果某些要区分的物体具有相似性,需要增加更多图片。

上传图片的总量有限制吗?

• 每个账号下所有数据集的图片总数不能超过10万张。

心 训练相关问题

数据处理失败或者状态异常怎么办?

 如是是图像分类模型上传处理失败,请先检查已上传的分类命名是否正确,是否存在中文命名、或者增加了空格;然后检查下数据图片量是否 超过上限(10万张);再检查图片中是否有损坏。如果自查没有发现问题请在百度智能云控制台内提交工单反馈

模型训练失败怎么办?

• 如果遇到模型训练失败的情况,请在百度智能云控制台内提交工单反馈

已经上线的模型还可以继续优化吗?

 已经上线的模型依然可以持续优化,操作上还是按照标准流程在训练模型中-选择要优化的模型和数据完成训练,然后在模型列表中更新线上 服务,完成模型的优化

```
点击我的模型列表——找到新训练好的模型版本——点击申请发布
```

【图像分类】百美椅子训练 模型ID: 230 全部版本 删除							
应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作	
云服务	V2	训练完成	未申请	未发布	top1准确率87.61% top5准确率100.00% 完整评估效果	申请发布 校验 训练	
离线SDK	V1	训练完成	未申请	未发布	top1准确率85.84% top5准确率100.00% 完整评估效果	申请发布 训练	
						每页显示 12 🗸 < 1	

在出来的弹窗中点击确定

我的模型	完整评估效果	
创建模型	申请发布 ×	
训练模型 校验模型	① 当前模型的线上服务为https://aip.baidubce.com/rpc/2.0/ai_custom/v1/classification/model_172,相应的模型版本为V1。 是否确认要替换为V2	全部版本删除
发布模型	爾定取消	川练
改据中心 数据集管理	【图像分类】椅子颜色识别 模型ID:2598	全部版本 删除

∞ 模型效果相关问题

物体检测模型如何正确标注?

- 所有图片中出现的目标物体都需要被框出(框可以重叠)
- 框应包含整个物体,且尽可能不要包含多余的背景
- 如果图片中存在很多相同标签的目标物体,可以使用右侧的锁定按钮。锁定标签后,只需要在左侧框选目标物体即可,不用再重复选择标签

如何通过「完整评估结果」里的错误示例优化模型?

- 错误示例中, 左侧是正确的结果, 右侧是模型的识别结果
- 观察模型识别有误的图片有哪些共同点,并有针对性地补充训练数据。比如:当图片比较亮的时候模型都能识别正确,但比较暗的时候模型就 识别错了。这时就需要补充比较暗的图片作为训练数据

我的数据有限,如何优化效果?

• 先申请发布模型,并备注说明希望通过云服务数据管理功能,将实际调用云服务识别的图片加入训练集,不断迭代模型

实际调用服务时模型效果变差?

- 训练图片和实际场景要识别的图片拍摄环境应一致,举例:如果实际要识别的图片是摄像头俯拍的,那训练图片就不能用网上下载的目标正面 图片
- 每个标签的图片需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强
- 如果使用的是云服务,可以开通云服务数据管理功能,将实际调用云服务识别的图片加入训练集,不断迭代模型

如果训练数据已经达到以上要求,且单个分类/标签的图片量超过200张以上,效果仍然不佳,请在百度智能云控制台内提交工单反馈

心 智能标注相关问题

智能标注功能目前已对物体检测模型开放,了解功能详情

"一键标注"和"立即训练"要如何选择?

• 当系统推荐"立即训练",且系统预标注的框确实已非常精准时,可以不用标注剩余数据,直接开始模型训练。此时,仅用当前已标注图片训练

的模型,与标注所有数据后训练的模型相比,效果几乎等同

• 如果系统预标注的框还有些不精准,可以启动一键标注,人工确认系统标注的标注框后,再开始训练

选择了"立即训练"之后是否还可以"一键标注"?

- 选择"立即训练"之后,系统默认为您结束此次智能标注
- 再次启动智能标注后,您可以通过以下方式进行一键标注:
 - 根据系统提示,进入一键标注
 - 查看系统对"未标注[优先]"图片的预标注,点击"满意预标注结果"后,进入一键标注

智能标注结束后,又往数据集上传了新图片,是否可以直接"一键标注"新图片?

- 如果您创建了新的标签、或新上传的图片场景和之前的图片场景差异较大,建议不要使用一键标注,而是从头开始智能标注(即再次筛选关键 图片)
- 如果不是以上情况,再次启动智能标注后,可以通过以下方式进行一键标注:
 - 根据系统提示,进入一键标注
 - 查看系统对"未标注[优先]"图片的预标注,点击"满意预标注结果"后,进入一键标注

智能标注中可以增删标签吗?

- 暂不支持。为了保证系统智能标注的效果,建议在启动功能前就创建好所有需要识别的标签
- 如果确实需要增删标签,可以先结束智能标注

智能标注中可以增删图片吗?

- 暂不支持。为了保证系统智能标注的效果,建议在启动功能前上传需要标注的所有图片,并删除不相关的图片
- 如果确实需要增删图片,可以先结束智能标注

智能标注中可以修改已标注图片的标注框吗?

- 可以。但为了保证智能标注的效果,建议不要大量改动
- 如果确实需要修改大量标注,建议先结束智能标注

为什么我已经人工标注了很多图片,但系统预标注依然不准?

- 系统预标注的结果会受以下因素影响:
 - 智能标注期间,对"已标注"图片的标签进行大量改动
 - 曾结束智能标注,并对标签、图片进行增删
- 如果您没有进行以上操作,系统标注结果依然不理想,请在百度智能云控制台内提交工单反馈

多个数据集是否可以同时启动智能标注?

• 目前每个账号同一时间仅支持对一个数据集启动智能标注

共享中的数据集是否可以启动智能标注?

• 暂不支持。智能标注中的数据集也暂不支持共享,如有疑问请在百度智能云控制台内提交工单反馈

智能标注失败了怎么办?

- 可以先尝试稍后重新启动
- 若再次遇到问题,请在百度智能云控制台内提交工单反馈

☑ 模型上线相关问题

希望加急上线怎么处理?

• 请在百度智能云控制台内提交工单反馈

每个账号可以上线几个模型?是否可以删除已上线的模型?

• 每个账号最多申请发布十个模型,已上线模型无法删除

申请发布模型审核不通过都是什么原因?

 可能原因有,1、经过电话沟通当前模型存在一些问题或者不再使用,如训练数据异常、数据量不够、不想再继续使用等原因,沟通达成一致 拒绝。2、电话未接通且模型效果较差,会直接拒绝。如果需要申诉,请在百度智能云控制台内提交工单反馈

の 离线SDK发布问题

发布SDK时可以选择的操作系统、芯片和加速类型与什么有关?

 与训练时选择的算法有关,以「物体检测-矩形框」为例,可以通过该链接查看适配的硬件以及不同算法在不同硬件下的性能表现:算法性能和 适配硬件,同时也可观看视频介绍:

•	
🗘 百度智能云	产品介绍
零门槛AI开发 💦 🖳	
平台FasyDI	1.
	\supset .
纯离线SDK发布可选类型介绍	

图像分割

整体介绍

心简介

Hi,您好,欢迎使用百度EasyDL图像。

EasyDL图像支持定制图像分类、物体检测、图像分割三类模型。三类模型的功能区别如下:

- 图像分类:识别一张图中是否是某类物体/状态/场景,适用于图片内容单一、需要给整张图片分类的场景
- 物体检测:检测图中每个物体的位置、名称。适合图中有多个主体要识别、或要识别主体位置及数量的场景
- 图像分割:对比物体检测,支持用多边形标注训练数据,模型可像素级识别目标。适合图中有多个主体、需识别其位置或轮廓的场景

以下是关于图像分割模型的技术文档。

∞ 应用场景

- 专业检测:应用于专业场景的图像分析,比如在卫星图像中识别建筑、道路、森林,或在医学图像中定位病灶、测量面积等
- 智能交通:识别道路信息,包括车道标记、交通标志等

心 使用流程

训练模型的基本流程如下图所示,全程可视化简易操作。在数据已经准备好的情况下,最快几分钟即可获得定制模型。

下面将详细介绍每一步的操作方式和注意事项。如果文档没有解决您的问题,请在百度智能云控制台内提交工单反馈。



数据准备

心 创建数据集

在训练之前需要在数据中心【创建数据集】。

如果训练数据需要多人分工标注,可以创建多个数据集。将训练数据分批上传到这些数据集后,再将数据集"共享"给自己的小伙伴,同步进行标 注。

设计标签

在上传之前确定想要识别哪几种物体,并上传含有这些物体的图片。每个标签对应想要在图片中识别出的一种物体

注意:标签的上限为1000种

准备图片

基于设计好的标签准备图片:

- 每种要识别的物体在所有图片中出现的数量需要大于50
- 如果某些标签的图片具有相似性,需要增加更多图片
- 一个模型的图片总量限制4张~10万张

图片格式要求:

- 1、目前支持图片类型为png、jpg、bmp、jpeg,图片大小限制在14M以内
- 2、图片长宽比在3:1以内,其中最长边小于4096px,最短边大于30px

图片内容要求:

1、训练图片和实际场景要识别的图片拍摄环境一致,举例:如果实际要识别的图片是摄像头俯拍的,那训练图片就不能用网上下载的目标正面 图片

2、每个标签的图片需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强

如果需要寻求第三方数据采集团队协助数据采集,请在百度智能云控制台内提交工单反馈

心上传数据集并在线标注

在完成了设计标签与准备数据后,可以通过以下方式导入数据:

- 导入未标注的数据,在线进行数据标注
- 直接导入标注好的数据

导入未标注数据

本地数据

支持上传图片、压缩包,或通过API导入

已有数据集

支持选择百度云BOS导入、分享链接导入、平台已有数据集导入;支持选择线上已有的数据集,包括其他图像类模型的数据集

1	ど楽EasyDL	产品介绰	召 操作平台	应	用案例 使用	文档	
物体	本检测模型	ŧ	我的数据总览 >	水果测	l试/V1/导入		
6.0	总览 模型中心 我的模型		┃ 创建信息 ~ 数据集ⅠD 备注	28742	1	版本号	V1
	创建模型 训练模型 校验模型 发布模型		┃ 标注信息 ∨ 标注类型 数据总量	物体检 19	测	标注模板 已标注	矩形框标注 5(进度26.32%)
M	EasyData数据服务 数据总览		标签个数 待确认	2 0		标注框数 大小	8 0.5M
	标签组管理 在线标注 云服务数据回流		 Ⅰ数据清洗 暫未做过数据清 Ⅰ导入数据 数据标注状态 	洗任务	● 无标注信息	○ 有标注信息	
	摄 医大致 括木 果 公有 云 服 务 在 线 服 务 批量 预 测		导入方式 一键导入 Label	me 근체	请选择 BOS目录导入 分享链接导入		^
	EasyEdge本地部署 纯离线服务 端云协同服务	^			平台已有数据集 摄像头采集数据 云服务回流数据		

在线标注

标注方式

在【数据标注/上传】页面上传并在线标注图片:

Step 1选择数据集

Step 2上传已准备好的图片

Step 3在标注区域内进行标注

首先在标注框上方找到工具栏,点击标注按钮在图片中拖动画框,圈出要识别的目标

然后在右侧的标签栏中,增加新标签,或选择已有标签



自动识别轮廓标注 推荐使用自动识别轮廓工具进行标注,鼠标左键点击目标即可自动出现标注区,鼠标右键点击误识别的区域可取消误识别区域 的标注,反复操作即可精炼出十分准确的标注结果



图片标注Tips

• 所有图片中出现的目标物体都需要被框出 (框可以重叠)



• 框应包含整个物体,且尽可能不要包含多余的背景



包含整个物体

- 未框全或包含多余背景
- 如果图片中存在很多相同标签的目标物体,可以使用右侧的锁定按钮。锁定标签后,只需要在左侧框选目标物体即可,不用再重复选择标签

の 数据集智能标注

使用智能标注功能可降低数据的标注成本。启动后,系统会从数据集所有图片中筛选出最关键的图片并提示需要优先标注。通常情况下,只需标 注数据集30%左右的数据即可训练模型。与标注所有数据后训练相比,模型效果几乎等同

整体流程以物体检测的智能标注流程为例:

创建智能标注任务

启动物体检测数据集的智能标注前,请先检查以下是否已满足以下条件:

- 所有需要识别的标签都已创建
- 每个标签的标注框数不少于10个
- 所有需要标注的图片都已加入数据集,且所有不相关的图片都已删除

若已满足,即可从导航栏进入「数据服务」-「智能标注」,创建智能标注任务,系统会基于您选择数据类型及数据量级,自动预估任务运行时长

Bai 認大脑 EasyData 产品介绍 数据服务 使用文档								
EasyData数据服务 《三	智能标注							
我的数据总览		创建智能标注任务	×					
☞ 数据标注	智能标注	1、由于系统筛选图片需要一定等待的时间,建议在未标注图片数大于100时启动,提高标注效率;	右的难					
在线标注		2、每个标签的标注框数都达到10个时才可以启动功能;						
智能标注	创建智能标注	3、不同的数据量级会影响您的任务时长,请参考 任务时长预估表。 🕜						
多人标注 ^{new}	序号 数	数据集类型 物体检测 实例分割 文本分类单标签						
寻求标注支持	1 2'	选择数据集 xyf_mask_fake	a⊊h ata					
☞ 数据采集		系统将对您选择的数据集进行校验,请确保每个标签的标注框数都达到10个;	14/J 🖻					
接入摄像头采集图片 🛛 🗸 🗸	2 23	您的智能标注任务预计需要38分钟完成	动查					
云服务数据回流 🗸 🗸	3 25		接 难					
数据清洗 清洗任务管理	4 23	启动智能标注	动 查					

系统筛选难例

系统会分批筛选出最关键需标注的图片,即难例图片。

```
Tips:难例筛选需要一定时间,在此期间您可以正常进行其他未标注图片的标注
```

智能标注启动中, 请稍候 启动过程需要大约5分钟左右的时间,请耐心等待,了解 您的智能标注任务预计需要38分钟完成	功能详情			
关闭智能标注				
	ج ج ا	>		
1.启动智能标注	2.系统筛选难例		3.用户标注难例	4.完成标注,启动训练
在每个标签的已标注框数都达到10个后,启动智能标注	根据您的数据集情况,系统自动筛选出数据集中 难例图片(最关键需优先标注的图片),进行自 动预标注		针对展示预标注结果的难例,您可通过两种方式 完成标注:第一,基于单张图片进行手动标注: 第二,在预标注效果符合预期基础上,选择一键 标注	您可以根据实际情况,选择系统为您进行下一轮 难例筛选(系统最多支持4轮难例筛选),或对 当前预标注效果满意进入立即训练

用户确认难例

智能标注任务启动后,系统为您自动筛选难例,您可以通过总览页查看进度按钮查看当前难例筛选进度,同时,进度图中也会全局展示您处于难 例筛选的具体哪一环节,以便您的操作后续。 筛选难例完成后,绿色进度条会进展到确认难例阶段,您可以点击【确认难例】完成对预标注结果 的人工确认。

创建智能标注任务 ○ 文本智能标注任务 ○ 文本智能标注任务									
序号	数据集ID	数据集名称		版本	智能标注状态	操作			
1	3107	tj-智能标注-检测-demo		V2	已中止	重新启动	查看记录		
2	2751	xyf_test_data1		V1	已中止	重新启动	查看记录		
3	256	test123		V1	运行中	查看进度	难例确认 中止任务 重	建 看记录	
4	2744	zzy-测试智能标注	当前您处于第1轮难例阶段(共4轮) 已为您筛选出轮难例图片,请确认该轮难例图片			查看记录			
5	3111	tj-智能标注-检测-1110			f	▶ 任务完成	查看记录		
6	2831	py3升级-智能标注-物体检测·	● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●		效果评估	下一轮筛选	查看记录		
7	1965	物体检测-多人标注用					查看记录		
8	2981	赵骞专属数据集1		* 1	L.70%	二人が注し	查看记录		

我们为您的人工确认提供两种模式:

- 单张确认,在该模式下支持您对预标注结果进行修正后点击保存
- 一键保存所有标注,为提升您的确认效率,默认您对难例的预标注结果全部满意,即可进入下一阶段


标注难例的预训练模型,也会对您无标注信息下的图片进行预标注结果的展示,您有余力的情况下,可以完成标注确认,确认后该张图片将升级 为已标状态,该环节并非是您进入智能标注下一阶段的必备要求。



评估难例效果,完成任务

当您对难例完成确认后,您可以根据本轮次预标注的结果是否满意,判断您是否还需要进入下一轮难例筛选阶段,如果满意本轮难例的预标注效果,系统将自动为您系统其他的未标图片打标签。

第1轮难例标注中(共4轮)

1、点击右下角【保存当前标注】该预标注结果将完成确认,支持您对标注修改后再保存
2、您只有对【待确认标注】下所有预标注结果完成确认,所有难例均升级为已标状态,才可进入下一阶段
✓ 该轮预标注难例已全部完成确认
1、若预标注效果未达预期,可通过更多轮次难例筛选实现效果优化, 进入下一轮难例筛选
2、 若预标注效果符合预期,将按照当前标注效果,对剩余未标图片进行自动标注, 确认自动完成标注

中止任务

当您在任务运行中想要中止任务时,可实时点击标注页面右上方【中止任务】按钮,任务将被提前结束。

经典版-物(确认中止该任务?	题 提交工单
定制物体检测模型,可	中止智能标注任务后,确认中的难例将不会生效。	-张图片中要识别多个物体,物体计数等场景中。
在各检测物体之间差异	确认 取消	+,训练时间可能需要30分钟以上。
线标注 > xj_智能标注性能 > 材	范注	中止智能标注 智能标

其他操作提示

• 在智能标注任务中,有任务上限吗?

支持五条智能标注任务同时运行,超过该上限您需要中止其他任务

• 智能标注中可以增删标签吗?

暂不支持。为了保证系统智能标注的效果,建议在启动功能前就创建好所有需要识别的标签 如果确实需要增删标签,可以先结束智能标注

• 智能标注中可以增删图片吗?

暂不支持。为了保证系统智能标注的效果,建议在启动功能前上传需要标注的所有图片,并删除不相关的图片。如果确实需要增删图片,可以先 结束智能标注

• 智能标注中可以修改已标注图片的标注框吗?

可以。但为了保证智能标注的效果,建议不要大量改动。如果确实需要修改大量标注,建议先结束智能标注

• 为什么我已经人工标注了很多图片,但系统预标注依然不准?

系统预标注的结果会受以下因素影响:智能标注期间,对"已标注"图片的标签进行大量改动;曾结束智能标注,并对标签、图片进行增删

• 多个数据集是否可以同时启动智能标注?

目前每个账号同一时间仅支持对一个数据集启动智能标注

• 共享中的数据集是否可以启动智能标注?

暂不支持。智能标注中的数据集也暂不支持共享

• 智能标注失败了怎么办?

可以先尝试稍后重新启动,如多次失败请提交工单联系我们

问题反馈

您在使用EasyData过程中可以通过以下任何方式联系我们:

● 在社区咨询

在论坛发帖提交问题,也可以在论坛与其他用户一起交流。前往论坛

● 提交工单

如果使用EasyData遇到其他任何问题或任何bug,您可以点此提交工单

• 添加微信小助手留言

请在微信搜索"BaiduEasyDL",并备注暗号 "EasyData",添加小助手后留言。

心 数据集多人标注

如果训练数据需要多人分工标注,可以创建多个数据集。将训练数据分批上传到这些数据集后,再将数据集"共享"给自己的小伙伴,同步进行标 注。

共享方式如下:

1. 在「数据集管理」页面,点击需要共享的数据集对应操作栏中的「共享」

模型中心	数据集管理	l					寻求数据支持: 百度众测 数据服务商
我的模型	创建数据	集 数据上传API					
创建模型	ID	名称	类型	标签数	图片数	状态	操作
训练模型	44339	мм豆 ∥	物体检测	1	60	正常	查看 标注/上传 智能标注历史 删
校验模型							帝善 标注/上供 知能标注压由 副
发布模型	38867	更上层楼 ∥	物体检测	2	23	正常	量值 (0.12)上18 首都的社会文 and 除 共享
数据中心	27672	new 🖉	物体检测	0	0	新建	查看 标注/上传 删除 共享
数据集管理	9798	super_band_2 🖉	物体检测	3	107	正常	查看 标注/上传 删除 共享
创建数据集	8994	super_band 🖉	物体检测	3	7638	正常	查看 标注/上传 删除 共享详情
如18107.11/上下 一服条调用数据	8096	马 /	物体检测	2	20	正常	查看 标注/上传 删除 共享
ム版力時内以信							毎页显示 10 🗸 < 1 >

^{2.} 在共享页面,勾选被共享数据集的授权使用范围,生成共享链接。如需被共享人标注数据,则需勾选「修改」

黄刑中心	Sheet
关至中心	效抵集管理 〉 共享
我的模型	温馨提示:为了更有序地进行标注。每个数据集仅支持共享给一位用户。若一批训练数据需要多人共同标注,请先将数据拆分并上传到不同的数据集。训练时可从多个
创建模型	数据集选择数据进行训练。
训练模型	共享数据集: MM豆 🗸
校验模型	授予范囲: 🗾 查看 🗌 修改 🗌 使用
发布模型	
据中心	生成共享销程 返回
数据集管理	
创建数据集	
数据标注/上传	
云服务调用数据	

3. 复制共享链接,并发送给小伙伴



4. 被共享人打开链接后,即可在「数据集管理」页面看到被共享的数据集,并进行被授权的操作

5. 训练模型时,在「训练模型」页面添加训练数据时,可从多个数据集(如多个被共享的数据集)选择数据

模型中心	训练模型					
我的模型	选择模型: super band	~				
创建模型	应用类型: ✔ 云服务					
训练模型	选择算法: 〇	选择算法: 〇 高積度 ⑦ ④ 高性能 ⑦				
校验模型	□ 离线服务 ⑦					
发布模型	参与训练的标签					
数据中心	数据集	标签数量	操作			
数据集管理	super_band	3	查看详情 清空标签			
创建数据集	super_band_2	3	查看详情 清空标签			
数据标注/上传	继续添加 全部清空					
云服务调用数据						
	你已经选择2个数据集的6个标签					
		开始训练				

^の数据集管理API

本文档主要说明当您线下已有大量的图片数据,如何通过调用API完成图片的便捷上传和管理。EasyDL图像数据集管理API在管理不同模型数据类 型之间是通用的。上传不同模型类型数据,只是在部分接口入参存在差异,使用及接口地址完全一致。

数据集创建API

接口描述

该接口可用于创建数据集。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL控制台-公有云部署-应用列表页面创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/create

URL参数:

参数	值

```
access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"
```

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset_na me	是	string	数据集名称,长度不超过20个utf-8字符

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
dataset_id	否	number	创建的数据集ID

数据集列表API

接口描述

该接口可用于查看数据集列表。返回数据集的名称、类型、状态等信息。

接口鉴权

同模型上线后获取的API:

1、在EasyDL控制台-公有云部署-应用列表页面创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/list

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

字段	必选	类型	说明	
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类	
start	否	number	起始序号,默认为0	
num	否	number	数量,默认20,最多100	

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	数据集总数
results	否	array(object)	数据集列表
+dataset_id	否	number	数据集ID
+dataset_name	否	string	数据集名称
+type	否	string	数据集类型
+status	否	string	数据集状态
+special_status	否	string	数据集特殊状态,包括shared、smart和空值,分别表示共享中、智能标注中、非特殊状态

分类 (标签)列表API

接口描述

该接口可用于查看分类(标签)。返回分类(标签)的名称、包含数据量等信息。

接口鉴权

同模型上线后获取的API:

1、在EasyDL控制台-公有云部署-应用列表页面创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法:POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/list

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID
start	否	number	起始序号,默认0
num	否	number	数量,默认20,最多100

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id ,用于问题定位
total_num	否	number	标签总数
results	否	array(object)	标签列表
+label_id	否	string	标签/分类ID
+label_name	否	string	标签/分类名称
+entity_count	否	number	图片/声音/文本数量

添加数据API

接口描述

该接口可用于在指定数据集添加数据。

接口鉴权

同模型上线后获取的API:

1、在EasyDL控制台-公有云部署-应用列表页面创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/addentity

URL参数:

参数 值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID
append Label	否	boolea n	确定添加标签/分类的行为:追加(true)、替换(false)。默认为追加(true)。
entity_c ontent	是	string	type为 IMAGE_CLASSIFICATION/OBJECT_DETECTION/IMAGE_SEGMENTATION/SOUND_CLASSIFICATION时,填入图 片/声音的base64编码;type为TEXT_CLASSIFICATION时,填入utf-8编码的文本。 内容限制为:图像分类base64前 10M;物体检测base64前10M;图像分割base64前10M;声音分类base64前4M,声音时长1~15秒;文本分类 10000个汉字
entity_n ame	是	string	文件名

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id ,用于问题定位

数据集删除API

接口描述

该接口可用于删除数据集。

接口鉴权

同模型上线后获取的API:

1、在EasyDL控制台-公有云部署-应用列表页面创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/delete

URL参数:

参数 值	参数
---------	----

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

分类 (标签) 删除API

接口描述

该接口可用于删除分类(标签)。

接口鉴权

同模型上线后获取的API:

1、在EasyDL控制台-公有云部署-应用列表页面创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/delete

URL参数:

参数	值			
----	---	--	--	--

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset_ id	是	number	数据集ID
label_na me	是	string	标签/分类名称

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id ,用于问题定位

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度智能云控制台内提交工单反 馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求,如果持续出现此类错误,请在百度智能云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量 阶梯计费;未上线计费的接口,请在百度智能云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线计费的接口,请在百度智能云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶 梯计费;未上线计费的接口,请在百度智能云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
406000	internal server error	服务错误
406001	param[xx] invalid	参数xx不合法,请检查相关参数
406002	dataset not exist	数据集不存在
406003	dataset already exists	数据集已存在
406004	dataset can not be modified temporarily	数据集暂不可修改
406005	label not exist	标签/分类不存在
406006	no permission to modify the dataset	没有修改数据集的权限
406007	dataset cannot be modified while smart annotation is running	智能标注期间不可修改数据集
406008	quota exceeded	配额超限

心 数据质检

功能概述 该功能旨在对您数据集中的图像数据进行质量检测,通过提供客观指标,为您对数据集的下一步操作(标注、清洗等)进行参照引导。

整体质检报告将包括对原图、标注信息两个层面的指标进行统计,本期先上线原图维度的质检指标,标注层面的质检指标敬请期待。

使用流程 Step 1 功能入口

您可从数据总览页操作列点击【质检报告】或查看页面点击【质检报告】进入该功能页面

版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作	
V1⊙	142909	5	● 已完成	图像分类	0% (0/5)	-	查看与标注 导出 删除 质检报告	
Ē	我的数据总	览〉【图	片】的/V1/查看					
	全部 (5)	有标注信	息 (0) 无标注信,	急 (5)		+	- 导入图片 三 质检报告 批量标注示例	
	的V1版本的图片列表 筛选 🗸 🗌 本页全选							
	而成计算时间定		E D Anno C	60 (m)	ne setter Terreration and setter			

Step 2 指标查看本期报告分为整体指标和分布指标两类。整体指标包括数据集存储大小、图片数量、破损图像数三类;分布指标包括色彩分布 空间、图像存储大小分布、高宽比分布、分辨率分布、色偏分布五类。

可以通过切换数据集版本查看不同版本下质检报告。

数据供检报告 数据版本 V1 ✓ 数据集大小 2.04 MB 图片数量	5 张 碱抗图片数量 0 ⑦	
色彩分布空间	図書大小分布 〇 (B) 2 0 0 0 2 0 0 2 0 0	高度比分布 ⊙ (3) 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5 4 4 5 4 5 5 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
3)相当分布 ① (30) 2 15 1 3 3 5	88976 ℃ (8) 4 	

Step 3 对应处理 可通过hover具体指标数值进行相关操作,以高宽比分布为例:

第一步,高宽比大于3.6的超长图hover显示有1张图片比,支持点击



第二步,点击后进入符合该指标的图片操作页,可针对筛选后图片进行删除、标注等操作

//V1/查看 > 质检报告			
高宽比分布列表(1张) 筛选 ~ …			□本页全选 □ 删除
0%- 34. [\$\$\$\$[con] 即可进 入"母强国共和注意式" 天标签 ② 23		每页显示	
	(派)		
	4 • • • • • • • • • • • • • • • • • • •		

模型训练

心 图像分割创建模型

在导航【创建模型】中,选择任务场景,填写模型名称、联系方式、功能描述等信息,即可创建模型。

其中任务场景分为实例分割和语义分割

语义分割: 图像分割指将每个像素点归属为对象类的过程。其中,语义分割适用于分割目标主体单一的场景,简单举例来说语义分割能够识别出 图片中哪些像素是归属于"人"的标签,但无法区分"不同的人"

实例分割:图像分割指将每个像素点归属为对象类的过程。其中,实例分割会先定位目标再进行分割,简答举例来说,相比语义分割,实例分割 能够识别出"人"的同时,还能区分"不同的人"

图	象分割模型		模型列表	> 创建模型
88	总览 模型中心		模型类别	图像分割
	我的模型		任务场景 🔸	实例分割 ⑦ 语义分割 ⑦
	创建模型			and the second sec
	训练模型			
	校验模型 发布模型		模型名称 🗲	
<u>//</u>	EasyData数据服务		您的身份	企业管理者 企业员工 学生 教师
	数据总览		公司名称,	▲ 清输入公司名称 点击完成企业认证即可获得企业专属权益礼包 ⑦
	在线标注 智能标注		所属行业,	正並いに加性校民、いビン性中心り並然回避保当, 元和冶水市会自知向少小心, オカ心及(双上並代証化已主頭) ・ - - - - - - - - - -
	云服务数据回流		应用场景 🔸	请选择应用场景 >
	摄像头数据采集		邮箱地址 🔺	z*****@baidu.com
	公有云服务 在线服务		联系方式 🔸	135*****919
	批量预测		功能描述 🔸	k
	EasyEdge本地部署			
	纯离线服务			0/500
	端云协同服务	^		完成
	我的部署包			
	我的本地设备			
	下发部署包到设备			

模型创建成功后,可以在【我的模型】中看到刚刚创建的模型。

注:1. 创建模型后可持续新增模型版本,因此不必每次训练模型都创建模型 2.目前单个用户在每种类型的模型下最多可创建10个模型,每个模型均支持多次训练。 3.如果您是企业用户,建议您按照真实企业信息进行填写,便于EasyDL团队后续更好的为您服务

② 图像分割训练操作说明

数据提交后,可以在导航中找到【训练模型】,按以下步骤操作,启动模型训练:

注:1.启动训练前请确保数据已经标注完成,否则无法启动训练

2. 下述训练功能点中,标注为星号(*)的功能为非必要选择项,可根据实际需求考虑是否使用

图像分割模型	$i \equiv 1$	训练模型						操
器 总览 ① 模型中心		选择模型	sun_test_1214					
我的模型		训练配置						
创建模型		部署方式	公有云部署 Eas	/Edge本地部署 如	何选择部署方式?			
训练模型		增量训练						
校验模型			选择该模型下的已训练用	反本,在数据应用场景不3	变的情况下,扩充数据来进行增量	量训练,提升模型训练效率和邾	1度 什么是增量训练	
发布模型		选择算法	 默认 					
☑ EasyData数据服务		高级训练配置						
数据总览								
标签组管理		添加数据						
在线标注		添加数据集	+ 请选择					
智能标注			数据集		版本	标签数量		扬
云服务数据回流			分割猫猫		V1	2		
摄像头数据采集								
◎ 公有云服务		自定义验证集 ⑦						
在线服务		自定义测试集 ⑦						
批量预测		训练环境	名称	规格		算力	速度比例	价格
記 EasyEdge本地部署			GPU P4	TeslaGPU P4 8G显着	≅单卡 12核CPU 40G内存	5.5 TeraELOPS	\$ 1	免费
纯离线服务			0.014					
端云协同服务	^		GPU P40	TeslaGPU_P40_24G	显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥0.36/分钟
我的部署包			GPU V100	TeslaGPU_V100_16G	显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥0.45/分钟

① 选择模型

选择此次训练的模型

② 添加训练数据

先选择数据集,再按标签选择数据集里的图片。可从多个数据集选择图片(相同标签的训练图片会被合并)

训练时间与数据量大小有关,1000张图片可能需要几个小时训练,请耐心等待

图像分割模型	\in	训练模型							操作文档	常见问题
部 总览 山 模型中心		选择模型	afasdfasdfasdfads							
我的模型		训练配置								
创建模型		部署方式	公有云部署 Eas	yEdge本地部署	如何选择部署方式?					
训练模型		选择算法	○ 默认							
校验模型		高级训练配置								
发布模型		添加数据								
☑ EasyData数据服务										
数据总宽		添加数据集	+ 请选择							
标签组管理			数据集		版本	标签	E数量		操作	
在线标注			seg_diff		V7	5			移除	
智能标注										
云服务数据回流		目定又描址集 ②								
摄像头数据采集		自定义测试集 ⑦								
目 公有云服务		训练环境	名称	规格		算力	速度比例	价格		
在线服务			O GPU P4	TeslaGPU_P4_8	IG显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	免费		
批量预测			0011040	T		10 TEL ODD	1.47	# Javo og (/ \##	(0.00++++++++++++++++++++++++++++++++++	+++++++++++++++++++++++++++++++++++++++
認 EasyEdge本地部署			GP0 P40	TestaGP0_P40_	246亚仔单卡_12核CP0_406内仔	12 TeranLors	1.47	率下40.30/万世	(20)89*0 M(96)8 + 500)8	9 ° 17 # 7 49 E
纯离线服务			GPU V100	TeslaGPU_V100	_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥0.45/分钟		
端云协同服务 我的部署包	^	开始训练								

③ 选择部署方式

可选择「公有云API」、「EasyEdge本地部署」

如何选择部署方式

增量训练*增量训练:在模型迭代训练时,用户在原训练数据上增加了训练数据,可通过加载原训练数据训练的模型参数进行模型训练。这样可 让模型收敛速度变快,训练时间变短,同时在数据集质量较高的情况下,可能获得的模型效果也会更好。

注: 仅可选择同一部署方式下的训练的模型作为基准模型版本

当前仅支持选择默认算法

⑤ 训练模型

点击「开始训练」,训练模型。

- 训练时间与数据量大小有关,1000张图片可能需要几个小时训练,请耐心等待。
- 模型训练过程中,可以设置训练完成的短信提醒并离开页面。
- 平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

优惠政策:

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有图像分割 操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。

心 图像分割模型效果评估

可通过模型评估报告或模型校验了解模型效果:

- 模型评估报告:训练完成后,可以在【我的模型】列表中看到模型效果,以及详细的模型评估报告。
- 模型在线校验:可以在左侧导航中找到【校验模型】,在线校验模型效果。校验功能示意图:

校验模型							
选择模型: 楼楼的脑袋瓜	✔ 应用类型:	云服务(目前仅支持)	~ 选择)	版本:	V1	~	
当前模型mAP平均精度 100.00% 评估持	长告			识别结果 女	0何优化效果?		
).3
1 Kar	M.	11 miles	11 march 100			置信度>30%	
				1. head		79.28%	
点击添加图片				ŧ	请上线		

模型评估报告

整体评估

在这个部分可以看到模型训练整体的情况说明,包括基本结论、mAP、精确率、召回率。这部分模型效果的指标是基于训练数据集,随机抽出部 分数据不参与训练,仅参与模型效果评估计算得来。所以**当数据量较少时(如图片数量低于100个),参与评估的数据可能不超过30个,这样得** 出的模型评估报告效果仅供参考,无法完全准确体现模型效果。

注意:若想要更充分了解模型效果情况,建议发布模型为API后,通过调用接口批量测试,获取更准确的模型效果。

实例分割

整体评估

fenge01 V14效果优异,建议针对识别错误的图片示例继续优化模型效果。由于目前训练集数据量较少,该结论仅供参考,建议扩充训练集得到更准确的评估效果。 如何优 化效果?







查看模型评估结果时,需要思考在当前业务场景,更关注精确率与召回率哪个指标。是更希望减少误识别,还是更希望减少漏识别。前者更需要 关注精确率的指标,后者更需要关注召回率的指标。同时F1-score可以有效关注精确率和召回率的平衡情况,对于希望准确率与召回率兼具的场 景,F1-score越接近1效果越好。评估指标说明如下

F1-score: 对某类别而言为精确率和召回率的调和平均数,评估报告中指各类别F1-score的平均数

mAP: mAP(mean average precision)是物体检测(Object Detection)算法中衡量算法效果的指标。对于物体检测任务,每一类object都可以计算出 其精确率(Precision)和召回率(Recall),在不同阈值下多次计算/试验,每个类都可以得到一条P-R曲线,曲线下的面积就是average

精确率: 正确预测的物体数与预测物体总数之比。评估报告中具体指经比较F1-score最高的阈值下的结果

召回率:正确预测的物体数与真实物体数之比。评估报告中具体指经比较F1-score最高的阈值下的结果

语义分割

整体评估

dxu_bisenet_01 V2整体效果欠佳,建议针对识别错误的图片示例继续优化模型效果。由于目前训练集数据量较少,该结论仅供参考,建议扩充训练集得到更准确的评估效 果。 如何优化效果?





Kappa系数 ⑦: 0.9147

mloU: mloU(mean intersection over union) 是语义分割常用评价指标,mean是对于类别的平均,即每个类别loU的平均值。某个类别的loU计算 方式如下为,(当前类别预测正确像素点个数)/(当前类别预测正确像素点个数 + 本属于当前类却被预测为其他类像素点个数 + 本属于其他类 却被预测为当前类的像素点个数)。由于mloU是像素级别的交并比评估,数值上会略低于mAP,不会影响实际使用效果。

准确率:指类别预测正确的像素占总像素的比例,准确率越高模型质量越好。

Kappa系数:用于一致性检验的指标,可以用于衡量分类的效果。取值为-1到1之间,通常大于0,Kappa系数越高模型质量越好。

详细评估

在这个部分可以看到不同阈值下的F1-score,以及模型识别错误的图片示例。



不同阈值下F1-score表现



不同标签的mAP及对应的识别错误的图片



识别错误图片示例

通过分标签查看模型识别错误的图片,寻找其中的共性,进而有针对性的扩充训练数据。

如下图所示,可以通过勾选「误识别」、「漏识别」来分别查看两种错误识别的情况:



• 误识别:红色遮盖内没有目标物体(准备训练数据时没有标注),但模型识别到了目标物体

观察误识别的目标有什么共性:例如,一个检测电动车的模型,把很多自行车误识别成了电动车(因为电动车和自行车外观上比较相似)。这时,就需要在训练集中为自行车特别建立一个标签,并且在所有训练集图片中,将自行车标注出来。

可以把模型想象成一个在认识世界的孩童,当你告诉他电动车和自行车分别是什么样时,他就能认出来;当你没有告诉他的时候,他就有可能把 自行车认成电动车。

• 漏识别:橙色遮盖内应该有目标物体(准备训练数据时标注了),但模型没能识别出目标物体

观察漏识别的目标有什么共性:例如,一个检测会议室参会人数的模型,会漏识别图片中出现的白色人种。这大概率是因为训练集中缺少白色人种的标注数据造成的。因此,需要在训练集中添加包含白色人种的图片,并将白色人种标注出来。

黄色人种和白色人种在外貌的差别上是比较明显的,由于几乎所有的训练数据都标注的是黄色人种,所以模型很可能认不出白色人种。需要增加 白色人种的标注数据,让模型学习到黄色人种和白色人种都属于「参会人员」这个标签。

以上例子中,我们找到的是识别错误的图片中,目标特征上的共性。除此之外,还可以观察识别错误的图片在以下维度是否有共性,比如:图片 的拍摄设备、拍摄角度,图片的亮度、背景等等。

心 图像分割模型如何提升效果

一个模型很难一次性就训练到最佳的效果,可能需要结合模型评估报告和校验结果不断扩充数据和调优。

为此我们设计了模型迭代功能,即当模型训练完毕后,会生成一个最新的版本号,首次V1、之后V2……以此类推。可以通过调整训练数据和算法,多次训练,获得更好的模型效果。

注意:如果模型已经是上线状态(包括已付费的模型服务),依然支持模型迭代。只需要在训练完毕后发布新的版本,就可以获得更新后的模型 服务。

想要提升模型效果,可以尝试以下两种方法:

检查并优化训练数据

- 1. 检查是否存在训练数据过少的情况,建议每个标签标注50个目标以上,如果低于这个量级建议扩充。
- 2. 检查不同标签的标注目标数是否均衡,建议不同标签的标注目标数数据量级相同,并尽量接近,如果有的标签标注的很多,有的标签标注的 很少,会影响模型整体的识别效果。
- 3. 通过模型效果评估报告中的错误识别示例,有针对性地扩充训练数据。
- 4. **检查测试模型的数据与训练数据的采集来源是否一致**,如果设备不一致、或者采集的环境不一致,那么很可能会存在模型效果不错但实际测 试效果较差的情况。针对这种情况建议**重新调整训练集,将训练数据与实际业务场景数据尽可能一致**。

云服务调用数据管理

开通云服务调用数据管理功能后,可查找云服务模型识别错误的数据,纠正结果并将其加入模型迭代的训练集,实现训练数据的持续丰富和模型 效果的持续优化

具体使用流程如下:

1. 为已上线接口开通云服务调用数据服务

模型中心	云服务调用数据管理						
我的模型	选择接口: superband V						
创建模型	⊬担口中平通Facu们 干服系通图数据随道 显示平通2 >> 11能介绍						
训练模型							
校验模型	开通"EasyDL云服务调用数据管理"(以下简称"本功能"),可查找云服务模型识别错误的数据,纠正结果并将其加入模型迭代的训练集,实现训练数据的持续丰						
发布模型	富和模型效果的持续优化。 1 も現代また部分目的 の添け原目上在効果服合序時本古田Econの1 服系現白 教授夜川正通また部分日正処存得 目文を時20年的時度 短过20年分務度夜白						
数据中心	1. プロピアサイがあた。ロロサ、 ゆんほうエススレース」(キロコのスカロス 17 MEX El Dok.04 J Ok.1837 56177、 XX36177771 スページのかく レブトロオブ No. XI 27 TY Neu Ox (L) TX326, KBL2 U OX (L) XX361751 动圏除。						
数据集管理	2. 百度仅提供数据存储空间服务,该数据仅用于向虑提供本功能服务,不会用于其他目的。						
创建数据集	3. 您使用本功能,应当保证在开通服务期间不得允许任何其他第三方使用以上接口,并保证对所上倍的数据享有合法权益,有权进行上述处理并承担全部法律责任。						
数据标注/上传	您开通本服务即视为您同意百度以上述方式存储您上传的数据。						
云服务调用数据							
	□ 我已仔细问读并接受以上条款						
	开通服务						

2. 通过选择调用时间、标签,并设置筛选条件,查看疑似错误识别的图片

注意:数据将从开通功能后开始存储,最多存储30天的数据。当天调用的数据暂不支持即时查看,可在第二天查看

云服务调用数据管理	服务条款
遗择接口: horse ✔ 对应模型: 物体检测 对应版本: V3	
· 通用时间: 2010/2/01 2010/2/00 种 (5)	
watanalma. 2018/121 - 2018/012 🛄 🚺	
选择师篮: horse V	
筛选条件: ● 图片中识别出读标签,读标签的置信度介于 0 % ~ 60 %	
○ 图片中识别出的该标金数量	
童看筛选结果	
	五服务调用数据管理 通用时回: 2019/7/31 - 2019/6/29 會 ? 透择码至: horse 施工 ● 四片中识别出该标签,该标签的置信度介于 0 % ~ 60 % ● 四片中识别出的技标签数量

3. 将接口识别错误的图片添加到指定数据集(建议新建数据集)并纠正结果。后续训练模型时,只需增加包含接口数据的数据集,即可提升模型效果

尝试不同的训练配置 可前往训练配置页面尝试不同的配置组合,因不同数据集在不同的算法上可能表现不一致,所以建议您多尝试不同的算法选型后综合挑选精度最高的模型使用,你可以选择如下的配置项:

- 增量训练
- 在高级训练配置中增加输入图片分辨率

图像分割模型 → 三	训练模型		操作文档 常见问题 新手教程 摆交工单
部 总览			
▲ 模型中心	选择模型	調试 >	
我的模型	训练配置		
创建模型	部署方式	公有云部署 EasyEdgo本地部署 如何选择部署方式?	
训练模型	选择设备) 服务器 〇 通用小型设备	
校验模型	选择算法	】 對认 音響算法性能及话配硬件	
发布模型	100.000.000		
☑ EasyData数据服务	/6500.120 III	」 同步支持公有云部著 (2)	
数据总览	高级训练配置		
标签组管理	添加数据		
在线标注	添加数据集	+ 请选择	
智能标注		数据集 版本 标签数量	操作
云服务数据回流		±77	70.00
摄像头数据采集			移标
目 公有云服务	自定义验证集 ⑦		
在线服务	自定义测试集 ⑦		
批量预测	训练环境	名称 规格 算力 速度比例	价格
器 EasyEdge本地部署		-	
纯高线服务		GPU P4 IestaGPU_P4_8G亟存单卡_12%CPU_40G内存 5.5 TeraFLOPS 1	犯责
端云协同服务 へ		GPU P40 TeslaGPU_P40_24G量存单卡_12核CPU_40G内存 12 TeraFLOPS 1.47	单卡¥0.36/分钟 (50小时*节点免费)
我的部署包		○ GPU V100 TesiaGPU_V100_16G显存单卡_12核CPU_56G内存 14 TeraFLOPS 3.66	单卡¥0.45/分钟 (72.16小时*节点免费)

模型发布

图像分割模型发布整体说明

训练完成后,可将模型部署在公有云服务器、通用小型设备、本地服务器,或直接购买软硬一体方案,灵活适配各种使用场景及运行环境

公有云在线服务

训练完成的模型存储在云端,可通过独立Rest API调用模型,实现AI能力与业务系统或硬件设备整合

具有完善的鉴权、流控等安全机制, GPU集群稳定承载高并发请求

支持查找云端模型识别错误的数据,纠正结果并将其加入模型迭代的训练集,不断优化模型效果

纯离线服务 训练完成的模型整体打包为纯离线服务,可下载在本地稳定调用。纯离线服务按部署硬件芯片不同分为本地服务器部署、通用小型设备部署。为了提供更好的算法与硬件推理效果,EasyDL提供软硬一体方案部署。纯离线服务的整体支持与评测信息可详见算法与性能评测大表

本地服务器部署

可将训练完成的模型部署在私有CPU/GPU服务器上,支持服务器API和服务器SDK两种集成方式

模型服务性能表现更好,适用于对性能要求较高的场景,例如工业质检、流水线产品分拣等

通用小型设备

训练完成的模型被打包成适配智能硬件的SDK,可进行设备端离线计算。满足推理阶段数据敏感性要求、更快的响应速度要求

支持iOS、Android、Linux、Windows四种操作系统,基础接口封装完善,满足灵活的应用侧二次开发

端云协同服务

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

「云管理,端计算」的端云协同服务,具体包括:

在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新

断网状态下模型离线计算(http服务,可调用与公有云API功能相同的接口)

联网状态下在平台管理设备运行状态、资源利用率

心 公有云部署

の 如何发布API

训练完毕后可以在左侧导航栏中找到【发布模型】,依次进行以下操作即可发布公有云API:

- 选择模型
- 选择部署方式「公有云部署」

- 选择版本
- 自定义服务名称、接口地址后缀
- 申请发布

申请发布后,通常的审核周期为T+1,即当天申请第二天可以审核完成。如果需要加急、或者遇到莫名被拒的情况,请在百度智能云控制台内提 交工单反馈

发布模型界面示意:

模型中心	发布模型					
我的模型 创建模型	选择模型: 部署方式:	cxy-edge4.3-分割:服务器部署❤ 公有云部署 ❤	标准接口规范参 标准接口请求参	*考 *考说明:		
训练模型	选择版本:	V1 ¥	字段名称	必须	类型	说明
校验模型 发布模型	 服务名称: 接口地址: 	test_txfg	image	是	string	图像数据,base64编码,要求 base64编码后大小不超过4M, 最短边至少15px,最长边最大 4096px支持ipa(nng/bmp&式
数据中心 数据集管理	其他要求:	entation/ test_txfg 若接口无法满足您的需求,请指述希望解决的问题。	threshold	否	number	阔值,默认为当前模型推荐阈值 (0-1之间),具体值可以在我的 模型列表-模型效果查看
创建数据集			标准接口响应字	"段说明:		
数据标注/上传		0/500	字段名称	必须	类型	说明
云服务调用数据		提交申请	log_id	是	number	唯一的log id,用于问题定位
			error_code	否	number	错误码,当请求错误时返回
			error_msg	否	string	错误描述信息,当请求错误时返 回
			results	否	array(object)	标签结果数组
			+ name	否	string	标签名称
			+ score	否	number	置信度
			+ location	否	object	

接口赋权

在正式使用之前,还需要做的一项工作为接口赋权,需要登录EasyDL控制台中创建一个应用,获得由一串数字组成的appid,然后就可以参考接口文档正式使用了

ô	全局 ~						Q	合作伙伴	ŢФ	ns ^e	帮助文档	企业资源	财务	•
88	< 返回EasyDL总览	产品服务 /	EasyOL图像 - 应用列表											
>	EasyDL图像	应用列	间表											
98	公有云部著 ^	+ 3	建合用											
±	 应用列表 	_	应用名称	AppID	API Key	Secret Key	创建的	时间		捕作				
2	 权限管理 - 四冊信注 	1	识别摄和胸	15313490	gRHegw68kpnym0H6K7Gl6wUd	显示	2018	-12-31 17:24	27	11.8 M	2 803			
9-12	 技术文档 	2	TRY	14676113	75a5pbP7mhpBWNION2NYoYcX	B A	2018	-11-05 11:35	54	11.8 ¥	2 853			
833	• 售卖服务	3	皮皮肝	14403281	i6HhGApGanhWuYex4Oel2uG8	显示	2018	-10-11 19:55	26	11.8 ¥	2 803			
Þ	EasyEdge本地部署 ~												< 1	2

同时支持在「公有云服务管理」-「权限管理」中为第三方用户配置权限

示意图如下:

Ŷ	全局 ~		
88	✓返回EasyDL总览	产品服务 / EasyDL图像 - 权限管理	
>	EasyDL图像	权限管理	
88	公有云部署 ^	■服务权限管理	
đ	• 应用列表	支持为您或者他人的AppID配置可用服务权限。	
e	 权限管理 	序号 服务名称	相关操作
þ	• 用量統计	1 1000	配置权限
ж	。 技术文档	2 test	配置权限
E11	• 售卖服务	3 cherry	配置权限
Þ	EasyEdge本地部署 ~	4 horse的服务名称	配置权限
		5 sea	配置权限
		6 手势识别	配置权限

V	全局 ~	
	✔ 返回EasyDL总览	产品服务 / EasyDL图像 - 权限管理 / 配置权限
>	EasyDL图像	配置权限
88	公有云部署 ^	服务名称: 1000
ß	 应用列表 約700% 	可用APP ID: 14403281 皮皮虾 ×
ē ģ	• 用量统计	14676113 TRY ×
표	。 技术文档	15313490 识别猫和狗 ×
e 1	• 售卖服务	请添加 +
Þ	EasyEdge本地部署 ~	保存 取消

_の API调用文档

本文档主要说明定制化模型发布后获得的API如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

接口描述

基于自定义训练出的图像分割模型,实现定制图像识别。

接口鉴权

1、在EasyDL控制台创建应用

♥ 全局 ~								Q	≣	₽	?	品	۲	Q
EasyDL定制训练平台	产品服务 / EasyDl	L定制训练平台 - 应	用列表 / 创建应用											
概览	创建新应用													
应用列表	• 应田夕称·	- to at												
监控报表	12,713,121,101,1	test												
技术文档	•应用类型:	游戏娱乐					~							
云服务权限管理	*接口选择:	您可以为应用勾	选多个接口权限,使此应	用可以请求已勾选的	内接口服务	div.								
离线SDK管理		EasyDL	 ✓ 百美椅子训练 ✓ 识别眼睛和嘴巴 	 ✓ 椅子顔色び ✓ 声音分类 	マ另川	✓ aa ✓ 商品检测								
		+ 百度语音												
		➡ 文字识别												
		➡ 人脸识别												
		🛨 自然语言处理	2											
		➡ 内容审核]											
		🕂 UNIT 📘												
		🛨 知识图谱												
		🛨 图像识别												
		➡ 智能呼叫中/	b.											
		➡ 图像搜索												
		➡ 人体分析												
	•应用描述:	简单描述一下: 点,请控制在	您使用人工智能服务的应 500字以内	用场景,如开发一家	次美颜相析	1,需要检测人脸	关键							

2、应用详情页获取AK SK

♥ 全局 ~					Q	≣	₽	0	놂	۲	Q~
EasyDL定制训练平台	产品服务 / EasyDL定制训练平台 - 应用列表 / 应用	详情									
概览	应用详情	应用详情									
应用列表	编辑 查看文档	<u>編輯</u> _ 查看文档									
监控报表	应用名称	AppiD		API Key			Secr	et Key			
技术文档	test2	14293933		bZsfqO2fUdHTDflxqF1	11mkTq			*** 显示			
云服务权限管理 离线SDK管理	EasyDL定制化训练服务:									_	

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先进行自定义模型训练,完成训练后申请上线,上线成功后可在服务列表中查看并获取url。

URL参数:

参数	值
input_type	当取值为 url 时,需在请求参数中传入图片的URL string
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,与其他图像识别服务不同的是定制化图像识别服务以json方式请求。

Body请求示例:

```
{
"image": "<base64数据>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大4096px,通用 算法训练的模型发布API支持jpg/png/bmp格式。肺炎CT影像识别专用算法训练发布的API在此基础上支持 dicom格式,限制4M以内。 注意请去掉头部
thresho Id	否	number	-	默认值为推荐阈值,请在我的模型列表·模型效果查看推荐阈值
url	否	string	-	如果在请求URL参数中增加"input_type=url",则该参数必传,否则"image"参数必传。参数内容为URL string,用户需确保该string是有效的图片URL,否则会下载失败

请求代码示例

Python3 -.... EasyDL 物体检测 调用模型公有云API Python3实现 import json import base64 import requests 使用 requests 库发送请求 使用 pip (或者 pip3) 检查我的 python3 环境是否安装了该库,执行命令 pip freeze | grep requests 若返回值为空,则安装该库 pip install requests **目标图片的本地文件路径,支持jpg/png/bmp格式** IMAGE_FILEPATH = "【您的测试图片地址,例如:./example.jpg】" -

返回说明

实例分割返回参数

字段	类型	说明
log_id	number	唯一的log id,用于问题定位
results	array(ob ject)	识别结果数组
+name	string	分类名称
+score	number	置信度
+locati		
on		
++left	number	检测到的目标主体区域到图片左边界的距离
++top	number	检测到的目标主体区域到图片上边界的距离
++width	number	检测到的目标主体区域的宽度
++heig ht	number	检测到的目标主体区域的高度
+mask	string	基于游程编码的字符串,编码内容为和原图宽高相同的布尔数组:若数组值为0,代表原图此位置像素点不属于检测目标,若 为1,代表原图此位置像素点属于检测目标。下载解码SDK

语义分割返回参数

字段	类型	说明
log_id	number	唯一的log id,用于问题定位
results	array(ob ject)	识别结果数组
+name	string	分类名称
+mask	string	基于游程编码的字符串,编码内容为和原图宽高相同的布尔数组:若数组值为0,代表原图此位置像素点不属于检测目标,若 为1,代表原图此位置像素点属于检测目标。下载解码SDK

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

```
例如Access Token失效返回:
```

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述			
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。			
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。			
3	Unsupported openapi method	用的API不存在,请检查后重新尝试			
4	Open api request limit reached	集群超限额			
6	No permission to access data	无权限访问该用户数据			
13	Get service token failed	获取token失败			
14	IAM Certification failed	IAM鉴权失败			
15	app not exists or create failed	应用不存在或者创建失败			
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈			
18	Open api qps request limit reached	PS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 十费的接口,请在百度云控制台内提交工单反馈			
19	Open api total request limit reached	青求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈			
100	Invalid parameter	无效的access_token参数,请检查后重新尝试			
110	Access token invalid or no longer valid	access_token无效			
111	Access token expired	access token过期			
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈			
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有 疑问请在百度云控制台内提交工单反馈			
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈			
336003	Base64解码失败	图片/音频/文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头 部。有疑问请在百度云控制台内提交工单反馈			
336004	输入文件大小不合法	图片超出大小限制,图片限4M以内,请根据接口文档检查入参格式,有疑问请在百度云控制台内提交工单反 馈			
336005	图片解码失败	图片编码错误(非jpg,bmp,png等常见图片格式),请检查并修改图片格式			
336006	缺失必要参数	image字段缺失(未上传图片)			
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈			

の 本地服务器部署

训练完毕后,可以选择将模型通过「纯离线服务」或「端云协同服务」部署,具体介绍如下:

纯离线服务部署

可以在左侧导航栏中找到「纯离线服务」,依次进行以下操作即可将模型部署到本地服务器:

- 选择部署方式「服务器」
- 选择集成方式
- 选择模型、版本、系统和芯片
- 点击下一步

の 如何在服务器部署

模型中心	纯离线服务 > 发布新服务		
我的模型			
创建模型	① 选择部署形式 ② 填写个人信息		
训练模型	部署方式 💿 服务器 🔿 通用小型设备 🔿 专项适配硬件		
校验模型			
发布模型			
EasyData数据服务	选择模型 ttttttt V		
数据总览	违持版本 V4 🗸		
在线标注	选择系统和芯片 🔄 🛆 Linux		
云服务数据回流	通用X86 CPU () 英伟达GPU ()		
EasyEdge本地部署	华为 Atlas 300 百度 昆仑XPU		
纯离线服务	🕂 🚝 Windows		
AI市场	模型加速: 同时获取加速版 ③		
我的已购模型			
售卖模型	<u>स—न</u>		

- 填写部分信息(注:个人信息的填写仅供EasyDL团队了解您,便于后续合作接洽,不会作为其他用途使用)
- 点击发布

训练模型	* 联系人姓名	
校验模型	1. at 14	
发布模型	* 电诂	132*****247
FasyData数据服务	* 邮箱	6******@qq.com
	* 即务	
(据总览		
线标注	* 公司名称	
云服务数据回流	* 公司所在行业	请选择行业 🗸
asyEdge本地部署		
电离线服务	* 业务重级	请你简要描述,每天要处理的數据量级和所需的并发数
山市场		0/500
的已购模型	* 功能描述	请简要描述应用AI能力的使用话号。期望达到的效果等
与卖模型		
		0/500
	上一步	发布

① 私有API

将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷

点击「发布」后,前往「纯离线服务」等待部署包制作完成后,下载部署包,并参考文档完成集成

② 服务器端SDK

将模型封装成适配本地服务器(支持Linux和Windows)的SDK,可集成在其他程序中运行。首次联网激活后即可纯离线运行,占用服务器资源更少,使用方法更灵活

- 1、点击「发布」后,前往控制台申请服务器端SDK的试用序列号
- 2、点击「新增测试序列号」,根据模型类型选择「序列号类型」,填写「新增设备数」(所得序列号数量),点击确定即可

服务器纯离线服务管理	
API SDK	
使用说明: 1. 通过集成适配本地服务器的SDK高线使用模型:SDK支持二次开发,按CPU/GPU设备授权使用,操作步骤如下: © 在EasyDL申请并下载SDK → © 在下方申请激活序列号 → © 根据开发文档集成SDK,并联网激活 → © 真线使用 2.序列号联网激活SDK后有效期为1个月,一个测试序列导激活模型上限数为10,如需正式购买,请点击下方购买正式授权。	收起
+ 购买正式接权 新增测试序列号	开发文档 工单支持

3、离线SDK的激活和使用,请参考文档完成集成

Bai 公 大脑 AI 开放平台 开放能力	开发平台	行业应用	生态合作	AI市场	开发与教学	Q
EasyDL定制Al训练平台	芯片类型	实测硬件	高精度模型	高性能模型		
援索本产品文档内容Q	x86-64	Xeon E5-2650 v4	61ms	40ms		
	Nvidia-GPU	Tesla P4	16ms	5ms		
> 公有云部署	Atlas 300	Atlas 300	12ms	9ms		
 ✓ 私有服务器部署 如何在私有服务器部署 私有API集成文档 ✓ 服务器端SDK 服务器端SDK简介 	激活&使月 离线SDK的激流 ③ 在经典版控制 ③ 本地运行SD	目步骤 ⁶⁵ 与使用分以下三步: 9台下载SDK、获取序 K,并完成首次联网激	序列号 效活			
LinuxSDK集成文档- C++ LinuxSDK集成文档- Python	通过左侧导	航栏查看不同操作系	统SDK的开发了	文档		
Linux(Atlas)SDK集 成文档 WindowsSDK集成文 档	^{③正式使用} SDK常见	问题				

端云协同服务部署

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

具体使用说明请参考端云协同服务说明

本地服务器部署价格说明

EasyDL已支持将定制模型部署在本地服务器上,只需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。

如需购买永久使用授权,服务器SDK用户请在控制台点击「购买正式授权」,并按照对应步骤激活。

服务器API用户请微信搜索"BaiduEasyDL"添加小助手咨询,通过线下签订合同购买使用。

更多参考

EasyDL官网入口 EasyDL开发文档

^の纯离线SDK说明

心纯离线SDK简介

本文档主要说明定制化模型发布后获得的服务器端SDK如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

SDK说明

图像分割服务器端SDK支持Linux、Windows两种操作系统。以下为具体的系统、硬件环境支持:

操作系统	系统支持	硬件环境要求
Linux		CPU: x86_64 NVIDIA GPU: x86_64
		NVIDIA GPU: x86_64
Windows	64位 Windows7 及以 上	环境依赖: .NET Framework 4.5 Visual C++ Redistributable Packages for Visual Studio 2013 Visual C++ Redistributable Packages for Visual Studio 2015 GPU依赖:
		CUDA 9.x + cuDNN 7.x

根据具体设备、线程数不同,数据可能有波动,请以实测为准

在算法性能及适配硬件页面查看评测信息表。

激活&使用步骤

离线SDK的激活与使用分以下三步:

- ① 下载SDK后,在控制台获取序列号
- ② 本地运行SDK,并完成首次联网激活

通过左侧导航栏查看不同操作系统SDK的开发文档

③ 正式使用

SDK常见问题

通过左侧导航栏查看不同操作系统SDK的FAQ

以下是通用FAQ,如您的问题仍未解决,请在百度智能云控制台内提交工单反馈

1、激活失败怎么办?

①可能是当前序列号已被其他设备激活使用,请核实序列号后用未被激活的序列号重新激活

②序列号填写错误,请核实序列号后重新激活

③同一台设备绑定同一个序列号激活次数过多(超过50次),请更换序列号后重试

④首次激活需要联网,网络环境不佳或无网络环境,请检查网络环境后重试

⑤模型发布者和序列号所属账号非同一账号,如果存在这种异常建议更换账号获取有效序列号

⑥序列号已过有效期,请更换序列号后重试

⑦如有其他异常请在百度智能云控制台内提交工单反馈

心 Windows集成文档

简介

本文档介绍图像分割服务器端Windows SDK的使用方法。

- 硬件支持:
 - NVIDIA GPU (普通版,加速版)
- 操作系统支持
 - 64位 Windows 7 及以上
 - 64位 Windows Server 2012及以上
- 环境依赖 (必须安装以下版本)
 - .NET Framework 4.5
 - Visual C++ Redistributable Packages for Visual Studio 2013
 - Visual C++ Redistributable Packages for Visual Studio 2015
- GPU基础版 (EasyEdge-win-x86-nvidia-gpu) 依赖 (必须安装以下版本)
 - CUDA 11.0.x + cuDNN 8.4.x(注意参照英伟达安装文档安装Zlib:http://www.winimage.com/zLibDII/zlib123dllx64.zip,解压后将 dll_x64/zlibwapi.dll 拷贝到cuda的bin目录下) + 硬件计算能力(https://developer.nvidia.com/cuda-gpus#compute)达6.1及以上

- CUDA 12.0.x + cuDNN 8.9.x(注意参照英伟达安装文档安装Zlib) + 硬件计算能力达7.5及以上
- GPU加速版 (EasyEdge-win-x86-nvidia-gpu-tensorrt) 依赖 (必须安装以下版本)
 - CUDA 11.0.x + cuDNN 8.4.x(注意参照英伟达安装文档安装Zlib) + TensorRT 8.4.x.x
 - CUDA 12.0.x + cuDNN 8.9.x(注意参照英伟达安装文档安装Zlib) + TensorRT 8.6.x.x
- GPU加速版 (EasyEdge-win-x86-nvidia-gpu-paddletrt) 依赖 (必须安装以下版本)
 - CUDA 11.0.x + cuDNN 8.4.x(注意参照英伟达安装文档安装Zlib) + TensorRT 8.4.3.1 + 硬件计算能力达6.1及以上
 - CUDA 12.0.x + cuDNN 8.9.x(注意参照英伟达安装文档安装Zlib) + TensorRT 8.6.1.6 + 硬件计算能力达7.5及以上
- GPU加速版 (x86-nvidia-gpu-torch)
 - CUDA 11.0.x + cuDNN 8.0.5.x
- 协议
 - HTTP
- 更详细的环境说明可参考SDK内的README.md

Release Notes | 时间 | 版本 | 说明 | | -------- | ------ | -2023-08-30 | 1.8.3 | 新增支持按实例数鉴权 | | 2023-06-29 | 1.8.2 | 优化模型 算法 || 2023-05-17 | 1.8.1 | 预测引擎升级,修复部分网络内存泄露问题 || 2023-03-16 | 1.8.0 | 预测引擎升级 || 2022-12-29 | 1.7.2 | 预测引 擎升级 | | 2022-10-27 | 1.7.1 | GPU底层引擎升级,下线基础版CUDA10.0及以下版本支持 | | 2022-09-15 | 1.7.0 | 优化模型算法;GPU CUDA9.0 CUDA10.0 标记为待废弃状态 || 2022-07-28 | 1.6.0 | 优化模型算法 || 2022-05-27 | 1.5.1 | 新增支持BML Cloud小目标检测模型 || 2022-05-18 | 1.5.0 | 修复各别机器下程序崩溃的问题 | | 2022-04-25 | 1.4.1 | EasyDL, BML升级支持paddle2模型 | | 2022-03-25 | 1.4.0 | 优化 模型算法 || 2021-12-22 | 1.3.5 | GPU基础版推理引擎优化升级;GPU加速版支持自定义模型文件缓存路径;demo程序优化环境依赖检测 || 2021-10-20 | 1.3.4 | 修复已知问题|| 2021-08-19 | 1.3.2 | 新增支持EasyDL小目标检测,新增DEMO二进制文件 || 2021-06-29 | 1.3.1 | 预测引 擎升级 || 2021-05-13 | 1.3.0 | 模型发布新增多种加速方案选择;目标追踪支持x86平台的GPU及加速版;展示已发布模型性能评估报告 || 2021-04-08 | 1.2.3 | 支持BML平台模型仓库本地上传模型 | | 2021-03-09 | 1.2.2 | 修复已知问题 | | 2021-01-27 | 1.2.1 | 新增模型支持;性能优 化;问题修复||2020-12-18|1.2.0|推理引擎升级||2020-11-26|1.1.20|新增一些模型的加速版支持||2020-10-29|1.1.19|修复已知问 题 || 2020-09-17 | 1.1.18| 支持更多模型 || 2020.08.11 | 1.1.17 | 支持专业版更多模型 || 2020.06.23 | 1.1.16 | 支持专业版更多模型 || 2020.05.15 | 1.1.15 | 更新加速版tensorrt版本,支持高精度检测 | | 2020.03.13 | 1.1.14 | 支持声音分类 | | 2020.02.23 | 1.1.13 | 支持多阶段模 型 | | 2020.01.16 | 1.1.12 | 预测默认使用推荐阈值| | 2019.12.26 | 1.1.11 | 支持物体检测高精度算法的CPU加速版,EasyDL 专业版支持 SDK 加速版 || 2019.12.04 | 1.1.10 | 支持图像分割 || 2019.10.21 | 1.1.9 | 支持 EasyDL 专业版 || 2019.08.29 | 1.1.8 | CPU 加速版支持 || 2019.07.19 | 1.1.7 | 提供模型更新工具 || 2019.05.16 | 1.1.3 | NVIDIA GPU 支持 || 2019.03.15 | 1.1.0 | 架构与功能完善 || 2019.02.28 | 1.0.6 | 引擎功能完善 | | 2019.02.13 | 1.0.5 | paddlepaddle 支持 | | 2018.11.30 | 1.0.0 | 第一版 ! |

快速开始

1. 安装依赖

安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642

Visual C++ Redistributable Packages for Visual Studio 2013

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

Visual C++ Redistributable Packages for Visual Studio 2015

https://www.microsoft.com/zh-cn/download/details.aspx?id=48145

如果使用GPU版SDK,请安装CUDA+cuDNN

https://developer.nvidia.com/cuda https://developer.nvidia.com/cudnn

如果使用GPU版加速版SDK,请安装TensorRT

https://developer.nvidia.com/tensorrt

根据cuda版本下载,下载后把lib目录下的所有dll,拷贝到SDK的dll目录下

注意事项

1. 安装目录不能包含中文

2. Windows Server 请自行开启,选择"我的电脑"——"属性"——"管理"——"添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启即可。

2. 运行离线SDK

解压下载好的SDK,SDK默认使用cuda9版本,如果需要cuda10请运行EasyEdge CUDA10.0.bat切换到cuda10版本,之后打开EasyEdge.exe,输 入Serial Num,选择鉴权模式,点击"启动服务",等待数秒即可启动成功,本地服务默认运行在

http://127.0.0.1:24401/

其他任何语言只需通过HTTP调用即可。

如启动失败,可参考如下步骤排查:	
🗘 百度智能云	故障排查
零门槛AI开发	
平台EasyDL	
Windows离线SDK(公有云发	
五〕●问题排查	

2.1 离线鉴权 (默认鉴权模式) 首次联网激活,后续离线使用

Al ver. 1.2.0.0 core. 1	.8.1 sec. 1.5.1 -	- [×
Model	Mark Market Mark 19	i i		
Serial Num	and the second sec			
Host	127.0.0.1			
Port	24401			
鉴权模式	● 离线鉴权 ○ 按实例数鉴权			
服务状态	服务未启动			
	启动服务			
	是2	研机自动	动启动	

2.2 按实例数鉴权 周期性联网激活,离线后会释放所占用鉴权,启动时请确保心跳间隔小于等于生成序列号时填写的定期确认时间

Al ver. 1.2.0.0 core. 1.8.1 sec. 1.5.1	– 🗆 X
Model	
Serial Num	
Host 127.0.0.1	
Port 24401	
鉴权模式 ○ 离线鉴权 ④ 按实例数鉴权 心跳间隔 20	秒 ~
服务状态 服务未启动	
启动服务	是否开机自动启动 🗔

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2); global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 20);

或通过环境变量指定

set EDGE_CONTROLLER_KEY_AUTH_MODE=2 set EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=20

の 2.3 序列号激活错误码

错误码	文案	描述
4001	parameters missing. 参数缺失	
4002	parameters invalid. 参数不合法	
4003	model invalid. 模型信息不合法	
4004	no more activation left. 该序列号和该设备的激活次数超上限	
4005	the serial key is out of date. 该序列号过期	
4006	the serial key has been activated. 该序列号已被其他设备激活	该序列号已被其他设备激活,不能重复激活。
4007	account invalid. 序列号不能用于其他账号的模型	序列号不能用于其他账号的模型,只能用于绑定账号的模型。
4008	serial key invalid. 序列号不合法	序列号不存在或找不到
4009	bundle id invalid. 包名不合法	
4010	product invalid. 产品不合法	如easydl的SDK使用BML的序列号来激活,会报该错误
4011	platform invalid. 平台不合法	
4012	activate too frequent. 激活太频繁	激活太频繁,请稍后再进行激活。
4013	device type and license type not match. 硬件类型和序列号类型不匹配	如使用加速版序列号激活基础版SDK会报该错误
4014	exceed max activate device num. 超过最大授权数量	
4015	technology invalid. 技术类型不合法	
4016	exceed max activate entity num. 超过最大模型数量	
4017	device invalid. 设备不合法	
4018	model invalid. 模型不合法	

3. Demo示例(以图像服务为例)

v .

四 五度FeeuFdee

服务运行成功,此时可直接在浏览器中输入http://127.0.0.1:24401,在h5中测试模型效果。

【图像分割】45274 分割电池设备端V1	调整阈值〇	当前阈值: 0.5 <u>修改</u>
	label	置信度
	battery	0.998
	tomato	0.997
	tomato	0.996
	battery	0.995
	tomato	0.995
	tomato	0.993
	tomato	0.972

Python 使用示例代码如下

import requests

```
with open('./1.jpg', 'rb') as f: 
 img = f.read()
```

```
**params 为GET参数 data 为POST Body**
result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1},
data=img).json()
```

C# 使用示例代码如下

FileStream fs = new FileStream("./img.jpg", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close(); WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close();

response.Close();

C++ 使用示例代码如下,需要安装curl

```
**include <sys/stat.h>**
**include <curl/curl.h>**
**define S_ISREG(m) (((m) & 0170000) == (0100000))**
**define S_{ISDIR}(m) (((m) \& 0170000) == (0040000))**
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./img.jpg";
  FILE *fp = NULL;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  }
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\"\n", post_data_filename);
     return -1;
  }
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
     \label{eq:curl_easy_setopt} curl_easy\_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
     curl_easy_setopt(curl, CURLOPT_POST, 1L);
     curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
     curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
     res = curl_easy_perform(curl);
     \text{if }(\text{res }!=\text{CURLE_OK}) \ \{
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     }
     curl_easy_cleanup(curl);
  }
  curl_global_cleanup();
  fclose(fp);
  return 0;
```

请求参数

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送图片二进制。

返回参数|字段|类型|取值|说明||--------|-----|-----|-----|| confidence | float | 0~1 | 分割的置信度 || label | string || 分割的类 别 || index | number || 分割的类别 || mask | string || 游程编码的mask | 代码参考 https://github.com/Baidu-AIP/EasyDL-Segmentation-Demo

集成指南

基于HTTP集成

通过EasyEdge.exe启动服务后,参照上面的调用说明,通过HTTP请求集成到自己的服务中

基于c++ dll集成

集成前提

解压开的SDK包中包含src、lib、dll、include四个目录才支持基于c++ dll集成

集成方法

参考src目录中的CMakeLists.txt进行集成

基于c# dll集成

集成前提

解压开的SDK包中包含src\demo_serving_csharp、dll两个目录才支持基于c# dll集成

集成方法

参考src\demo_serving_csharp目录中的CMakeLists.txt进行集成

FAQ

1. 服务启动失败,怎么处理?

根据SDK内的README.md检查依赖是否都已正确安装

请确保相关依赖都安装正确,版本必须如下: *.NET Framework 4.5* Visual C++ Redistributable Packages for Visual Studio 2013 * Visual C++ Redistributable Packages for Visual Studio 2015

GPU依赖,版本必须如下: * CUDA 11.0.x + cuDNN 8.4.x 或者 CUDA 11.7.x + cuDNN 8.4.x

GPU加速版(EasyEdge-win-x86-nvidia-gpu-tensort)依赖,版本必须如下: * CUDA 11.0.x + cuDNN 8.4.x + TensorRT 8.4.x.x

GPU加速版(EasyEdge-win-x86-nvidia-gpu-paddletrt)依赖,版本必须如下: * CUDA 11.0.x + cuDNN 8.4.x + TensorRT 8.4.3.1

2. 服务调用时返回为空,怎么处理?调用输入的图片必须是RGB格式,请确认是否有alpha通道。

3. 多个模型怎么同时使用? SDK设置运行不同的端口,点击运行即可。

4. JAVA、C#等其他语言怎么调用SDK?

参考 https://ai.baidu.com/forum/topic/show/943765

5. 启动失败,缺失DLL? 打开EasyEdge.log,查看日志错误,根据提示处理 缺失DLL,请使用 https://www.dependencywalker.com/ 查看相应模 块依赖DLL缺失哪些,请自行下载安装

6. 启动失败,报错NotDecrypted? Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

7. 启动失败,报错 SerialNum无效

日志显示failed to get/check device id(xxx)或者 Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 C:\Users\\${用户名}\.baidu\easyedge 目录,再重新激活。

其他问题 如果无法解决,可到论坛发帖: https://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

心 Linux集成文档-C++

简介

本文档介绍EasyEdge/EasyDL的Linux CPP SDK的使用方法。

- 网络类型支持:图像分类,物体检测,图像分割,目标追踪
- 硬件支持:
 - CPU 基础版: intel x86_64 * AMD x86_64 龙芯 loongarch64 飞腾 aarch64

- CPU 加速版 Intel Xeon with Intel®AVX2 and AVX512 Intel Core Processors with AVX2 Intel Atom Processors with SSE AMD Core Processors with AVX2
- NVIDIA GPU: x86_64 PC
- 寒武纪 Cambricon MLU270
- 比特大陆计算卡SC5+
- 百度昆仑XPU K200
 - x86_64 飞腾 aarch64 百度昆仑XPU R200
 - x86_64 飞腾 aarch64
- 华为Atlas 300
- 海光DCU: x86_64 PC
- 寒武纪 MLU370 on x86_64
- 操作系统支持:Linux

根据开发者的选择,实际下载的版本可能是以下版本之一:

- EasyDL图像
 - x86 CPU 基础版
 - x86 CPU 加速版
 - Nvidia GPU 基础版
 - Nvidia GPU 加速版
 - x86 mlu270基础版
 - x86 SC5+基础版
 - Phytium MLU270基础版
 - Phytium XPU基础版
 - Phytium Atlas300I基础版
 - Hygon DCU基础版

性能数据参考算法性能及适配硬件

*intel 官方合作,拥有更好的适配与性能表现。

Release Notes

时间	版本	说明
2023.0 8.31	1.8.3	Atlas系列Soc支持语义分割模型,Atlas Cann升级到6.0.1,昆仑XPU后端推理引擎升级
2023.0 6.29	1.8.2	模型压缩能力升级
2023.0 5.17	1.8.1	支持物体检测自定义四边形模型精度无损压缩发布x86 CPU版SDK
2023.0 3.16	1.8.0	支持图像分类精度提升包本地部署
2022.1 2.29	1.7.2	模型性能优化;推理库性能优化
2022.1 0.27	1.7.1	新增语义分割模型http请求示例;升级海光DCU SDK,需配套rocm4.3版本使用;Linux GPU基础版下线适用于CUDA10.0及以 下版本的SDK:Linux GPU加读版升级推理引擎版本

2022.0 9.15	1.7.0	Linux GPU加速版升级预测引擎;Linux GPU加速版适用于CUDA9.0、CUDA10.0的SDK为deprecated,未来移除;新增实例分 割高性能模型离线部署;性能优化
2022.0 7.28	1.6.0	Linux CPU普通版、Linux GPU普通/加速版、Jetson新增目标追踪模型接入实时流的demo
2022.0 5.27	1.5.1	CPU、GPU普通版新增支持BML Cloud小目标检测模型
2022.0 5.18	1.5.0	GPU加速版max_batch_size参数含义变更;修复GPU加速版并发预测时部分图片结果预测错误及耗时增加问题;CPU普通版预 测引擎升级;新增版本号头文件;新增飞腾Atlas300I支持,并且在EasdDL新增多种加速版本;示例代码移除frame_buffer, 新增更安全高效的safe_queue; 新增Tensor In/Out接口和Demo
2022.0 4.25	1.4.1	EasyDL, BML升级支持paddle2模型
2022.0 3.25	1.4.0	新增支持海光服务器搭配海光DCU加速卡;
2021.1 2.22	1.3.5	GPU加速版支持自定义模型文件缓存路径;新增支持飞腾MLU270服务器、飞腾XPU服务器
2021.1 0.20	1.3.4	CPU加速版推理引擎优化升级,新增支持飞腾CPU、龙芯CPU服务器、比特大陆计算卡SC5+ BM1684、寒武纪MLU270;大幅 提升EasyDL GPU加速版有损压缩加速模型的推理速度
2021.0 8.19	1.3.2	CPU、GPU普通版及无损加速版新增支持EasyDL小目标检测,CPU普通版、GPU普通版支持检测模型的batch预测
2021.0 6.29	1.3.1	CPU普通版、GPU普通版支持分类模型的batch预测,CPU加速版支持分类、检测模型的batch预测;GPU加速版支持 CUDA11.1;视频流解析支持调整分辨率;预测引擎升级
2021.0 5.13	1.3.0	新增视频流接入支持;模型发布新增多种加速方案选择;目标追踪支持x86平台的CPU、GPU加速版;展示已发布模型性能评 估报告
2021.0 3.09	1.2.1	GPU新增目标追踪支持, http server服务支持图片通过base64格式调用,EasyDL高性能检测模型和均衡检测模型CPU加速版新 增量化压缩模型
2021.0 1.27	1.1.0	EasyDL经典版分类高性能模型升级;部分SDK不再需要单独安装OpenCV
2020.1 2.18	1.0.0	1.0版本发布!安全加固升级、性能优化、引擎升级、接口优化等多项更新
2020.1 1.26	0.5.8	EasyDL经典版分类模型CPU加速版里新增量化压缩模型
2020.1 0.29	0.5.7	新增CPU加速版支持:EasyDL经典版高精度、超高精度物体检测模型和EasyDL经典版图像分割模型
2020.0 9.17	0.5.6	性能优化,支持更多模型
2020.0 8.11	0.5.5	提升预测速度;支持百度昆仑芯片
2020.0 5.15	0.5.3	优化性能,支持专业版更多模型
2020.0 4.16	0.5.2	支持CPU加速版;CPU基础版引擎升级;GPU加速版支持多卡多线程
2020.0 3.12	0.5.0	x86引擎升级;更新本地http服务接口;GPU加速版提速,支持批量图片推理
2020.0 1.16	0.4.7	ARM引擎升级;增加推荐阈值支持
2019.1 2.26	0.4.6	支持海思NNIE
2019.1 1.02	0.4.5	移除curl依赖;支持自动编译OpenCV;支持EasyDL 专业版 Yolov3; 支持EasyDL经典版高精度物体检测模型升级
2010 1		

0.25	0.4.4	ARM引擎升级,性能提升30%; 支持EasyDL专业版模型
2019.0 9.23	0.4.3	增加海思NNIE加速芯片支持
2019.0 8.30	0.4.2	ARM引擎升级;支持分类高性能与高精度模型
2019.0 7.25	0.4.1	引擎升级,性能提升
2019.0 7.25	0.4.0	支持Xeye, 细节完善
2019.0 6.11	0.3.3	paddle引擎升级;性能提升
2019.0 5.16	0.3.2	新增NVIDIA GPU支持;新增armv7l支持
2019.0 4.25	0.3.1	优化硬件支持
2019.0 3.29	0.3.0	ARM64 支持;效果提升
2019.0 2.20	0.2.1	paddle引擎支持;效果提升
2018.1 1.30	0.1.0	第一版!

2022-5-18:【接口变更】 PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE含义变更。 变更前:预测输入图片数不大于该值均可。 变更后:预测输入图片数需等于该值。SDK内部对该接口变更做了兼容处理,在输入图片数小于该值时依然可正常运行,但预测性能会和等于该值时一致。推荐根据实际输入图片数量需求修改该值,尽可能保持最小。

2020-12-18: 【接口升级】 参数配置接口从1.0.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。 【关于SDK包与RES模型文件夹配套使用的说明】 我们强烈建议用户使用部署tar包中配套的SDK和RES。 更新模型时,如果SDK版本号有更新,请务必同时更新SDK,旧版本的SDK可能无法 正确适配新发布出来部署包中的RES模型。

快速开始

SDK在以下环境中测试通过

- x86_64, Ubuntu 16.04, gcc 5.4
- x86_64, Ubuntu 18.04, gcc 7.4
- Tesla P4, Ubuntu 16.04, cuda 9.0, cudnn 7.5
- x86_64, Ubuntu 16.04, gcc 5.4, XTCL r1.0
- aarch64, Kylin V10, gcc 7.3
- loongarch64, Kylin V10, gcc 8.3
- Bitmain SC5+ BM1684, Ubuntu 18.04, gcc 5.4
- x86_64 MLU270 , Ubuntu 18.04, gcc 7.5
- phytium MLU270 , Kylin V10 , gcc 7.3.0
- phytium XPU , Kylin V10 , gcc 7.3.0
- hygon DCU, CentOS 7.8 gcc 7.3.0
- XPU K200, x86_64, Ubuntu 18.04
- XPU K200 aarch64, Ubuntu 18.04
- XPU R200, x86_64, Ubuntu 18.04
- XPU R200 aarch64, Ubuntu 18.04
- MLU370, x86_64, Centos7.6.1810

依赖包括

- cmake 3+
- gcc 5.4 (需包含 GLIBCXX_3.4.22), gcc / glibc版本请以实际SDK ldd的结果为准
- opencv3.4.11 (可选)
- cuda && cudnn (使用NVIDIA-GPU时必须,SDK内提供多个Cuda版本推理套件,根据需要安装依赖的Cuda和Cudnn版本)
- XTCL 1.0.0.187 (使用昆仑服务器时必须)
- Rocm4.3, Miopen 2.14(使用海光DCU服务器时必须)

1. 安装依赖

以下步骤均可选,请开发者根据实际运行环境选择安装。

(可选) 安装cuda&cudnn

在NVIDIA GPU上运行必须(包括GPU基础版,GPU加速版)

对于GPU基础版,若开发者需求不同的依赖版本,请在PaddlePaddle官网下载对应版本的libpaddle_fluid.so或参考其文档进行编译,覆盖lib文件 夹下的相关库文件。

(可选) 安装TensorRT

在NVIDIA GPU上运行GPU加速版必须

下载包中提供了对应 cuda9.0、cuda10.0、cuda10.2、cuda11.0+四个版本的 SDK, cuda9.0 和 cuda10.0 的 SDK 默认依赖的 TensorRT 版本为 TensorRT7.0.0.11, cuda10.2 及以上的 SDK 默认依赖的 TensorRT 版本为 TensorRT8.4,请在这里下载对应 cuda 版本的 TensorRT,并把其中 的lib文件拷贝到系统lib目录,或其他目录并设置环境变量。

(可选) 安装XTCL 使用昆仑服务器及对应SDK时必须请安装与1.0.0.187版本兼容的XTCL。必要时,请将运行库路径添加到环境变量。

(可选) 安装Rocm、Miopen

使用海光DCU服务器对应SDK时必须

海光DCU SDK依赖Rocm 4.3和Miopen 2.14版本,推荐使用easyedge镜像

(registry.baidubce.com/easyedge/hygon_dcu_infer:1.0.2.rocm4.3), SDK镜像内运行,镜像拉取方式(wget https://aipe-easyedge-

public.bj.bcebos.com/dcu_docker_images/hygon_dcu_rocm4.3.tar.gz && docker load -i hygon_dcu_rocm4.3.tar.gz),关于海光DCU使用更多细节 可参考paddle文档

2. 使用序列号激活 请在官网获取序列号

<mark>純高技服务が明</mark> 変形成高线服务,将30%完成的模型部署在本地。高技调用模型,可以选择将模型部署在本地的服务器、小型设备、数硬一体方案专项适配被件上。 通过APL、SOX田一步集成、灵浪运应不同业务结果。 <u>業件解解条</u> 控制台						
副条数 通田小型设备 专项话即硬件						
DROPER ARTOVALING IN VOLUMENCIT						
SDK API						
此处发布、下载的SDK为未授权SDK,需要前往控制	自获取序列号激活后才能正式使用	I。SDK内附有对应版本的Demo及开发文档,开	发者可参考源代码完成开发。			
模型名称	发布版本 💲	应用平台	模型加速	发布状态	发布时间 🗘	
	134318V1 查看性能报告	通用X86 CPU-Linux	基础版	• 已发布	2021-08-19 20:24	下载SDK
			精度无损压缩加速	 已发布 	2021-08-19 20:24	下载加速版SDK
and the first state		2010-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0-0	基础版	•已发布	2021-08-19 20:35	下截SDK
sur_ore wrest		261912/01-0-LTHUX	精度无损压缩加速	 已发布 	2021-08-19 20:34	下载加速版SDK

基础版

SDK内bin目录下提供预编译二进制文件,可直接运行(二进制运行详细说明参考下一小节),用于图片推理和模型http服务,在二进制参数的

•已发布

2021-08-19 18:17

下载SDK

serial_num(或者serial_key)处填入序列号可自动完成联网激活(请确保硬件首次激活时能够连接公网,如果确实不具备联网条件,需要使用纯离 线模式激活,请下载使用百度智能边缘控制台纳管SDK)

SDK内提供的一些二进制文件,填入序列号运行可自动完成激活,以下二进制具体使用说明参考下一小节 ./edgekit_serving --cfg=./edgekit_serving.yml ./easyedge_image_inference {model_dir} {image_name or image_directory} [serial_num} ./easyedge_serving [res_dir] {serial_key} {host} {port}

如果是基于源码集成,设置序列号方法如下

global_controller()->set_licence_key("")

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的变量或者源码设置)实例数鉴权环境变量设置方法

export EDGE_CONTROLLER_KEY_AUTH_MODE=2 export EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=30

实例数鉴权源码设置方法

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2) global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 300)

3. 基于预编译二进制测试图片推理和http服务测试图片推理模型资源文件默认已经打包在开发者下载的SDK包中。

请先将tar包整体拷贝到具体运行的设备中,再解压缩编译; 在Intel CPU上运行CPU加速版,如果thirdparty里包含openvino文件夹的,必须在 编译或运行demo程序前执行以下命令: source \${cpp_kit位置路径}/thirdparty/openvino/bin/setupvars.sh 或者执行 source \${cpp_kit位置路 径}/thirdparty/openvino/setupvars.sh(openvino-2022.1+) 如果SDK内不包含setupvars.sh脚本,请忽略该提示

运行预编译图片推理二进制,依次填入模型文件路径(RES文件夹路径)、推理图片、序列号(序列号尽首次激活需要使用,激活后可不用填序列号 也能运行二进制)

./easyedge_image_inference {model_dir} {image_name or image_directory} {serial_num} LD_LIBRARY_PATH=../lib ./easyedge_image_inference ../../../RES /xxx/cat.jpeg "1111-1111-1111"

demo运行效果:



图片加载失败

> ./easyedge_image_inference ../../../RES 2.jpeg

2019-02-13 16:46:12,659 INFO [EasyEdge] [easyedge.cpp:34] 140606189016192 Baidu EasyEdge Linux Development Kit 0.2.1(20190213)

2019-02-13 16:46:14,083 INFO [EasyEdge] [paddlev2_edge_predictor.cpp:60] 140606189016192 Allocate graph success. 2019-02-13 16:46:14,326 DEBUG [EasyEdge] [paddlev2_edge_predictor.cpp:143] 140606189016192 Inference costs 168 ms 1, 1:txt_frame, p:0.994905 loc: 0.168161, 0.153654, 0.920856, 0.779621 Done

启动http服务 bin目录下提供编译好的启动http服务二进制文件,可直接运行

推荐使用 edgekit_serving 启动模型服务 LD_LIBRARY_PATH=../lib ./edgekit_serving --cfg=./edgekit_serving.yml

也可以使用 easyedge_serving 启动模型服务

./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401}

LD_LIBRARY_PATH=../lib ./easyedge_serving ../../../RES "1111-1111-1111-1111" 0.0.0.0 24401

HTTP(or Webservice) is now serving at 0.0.0.24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试,网页右侧会展示模型推理结果



对于目标追踪的模型,请选择一段视频,并耐心等待结果



```
同时,可以调用HTTP接口来访问服务。
```

请求http服务 以图像预测场景为例(非语义分割模型场景,语义分割请求方式参考后面小节详细文档),提供一张图片,请求模型服务的示例参考 如下demo

python示例代码如下

```
import requests
with open('./1.jpg', 'rb') as f:
  img = f.read()
**params 为GET参数 data 为POST Body**
result = requests.post('http://127.0.0.1:24401/', params= \{'threshold': 0.1\},
                                data=img).json()
```

C# 使用示例代码如下

FileStream fs = new FileStream("./img.jpg", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close(); WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream());

StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close(); response.Close();

C++ 使用示例代码如下,需要安装curl

```
**include <sys/stat.h>**
**include <curl/curl.h>**
**include <iostream>**
**include <string>**
**define S_ISREG(m) (((m) & 0170000) == (0100000))**
**define S_ISDIR(m) (((m) & 0170000) == (0040000))**
size_t write_callback(void *ptr, size_t size, size_t num, void *data) {
  std::string *str = dynamic_cast<std::string *>((std::string *)data);
  str->append((char *)ptr, size*num);
  return size*num:
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./img.jpg";
  FILE *fp = NULL;
  std::string response;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\"\n", post_data_filename);
     return -1;
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
    curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
    curl_easy_setopt(curl, CURLOPT_POST, 1L);
    curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
    curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
 curl_easy_setopt(curl, CURLOPT_WRITEFUNCTION, write_callback);
    curl_easy_setopt(curl, CURLOPT_WRITEDATA, &response);
    res = curl_easy_perform(curl);
     if (res != CURLE_OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     std::cout << response << std::endl; // response即为返回的json数据
     curl_easy_cleanup(curl);
  curl_global_cleanup();
  fclose(fp);
  return 0:
```

关于http接口的详细介绍参考下面集成文档http服务章节的相关内容

使用说明

使用该方式,将运行库嵌入到开发者的程序当中。 **编译demo项目** SDK src目录下有完整的demo工程,用户可参考该工程的代码实现方式将SDK 集成到自己的项目中,demo工程可直接编译运行:

```
cd src
mkdir build && cd build
cmake .. && make
./easyedge_image_inference {模型RES文件夹} {测试图片路径}
**如果是NNIE引擎,使用sudo运行**
sudo ./easyedge_image_inference {模型RES文件夹} {测试图片路径}
```

(可选) SDK包内一般自带opencv库,可忽略该步骤。如果希望SDK自动编译安装所需要的OpenCV库,修改cmake的 optionEDGE_BUILD_OPENCV为ON即可。 SDK会自动从网络下载opencv源码,并编译需要的module、链接。注意,此功能必须需联网。 cmake -DEDGE_BUILD_OPENCV=ON .. && make -j16

若需自定义library search path或者gcc路径,修改CMakeList.txt即可。

使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

// step 1: 配置模型资源目录 EdgePredictorConfig config; config.model_dir = {模型文件目录};

// step 2: 创建并初始化Predictor;在这里选择合适的引擎 auto predictor = global_controller()->CreateEdgePredictor(config);

// step 3-1: 预测图像 auto img = cv::imread({图片路径}); std::vector<EdgeResultData> results; predictor->infer(img, results);

// step 3-2: 预测视频
std::vector<EdgeResultData> results;
FrameTensor frame_tensor;
VideoConfig video_config;
video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h
video_config.source_value = video_src;
/*
... more video_configs, 根据需要配置video_config的各选项
*/
auto video_decoding = CreateVideoDecoding(video_config);

```
while (video_decoding->next(frame_tensor) == EDGE_OK) {
```

```
results.clear();
if (frame_tensor.is_needed) {
    predictor->infer(frame_tensor.frame, results);
    render(frame_tensor.frame, results, predictor->model_info().kind);
}
//video_decoding->display(frame_tensor); // 显示当前frame , 需在video_config中开启配置
//video_decoding->save(frame_tensor); // 存储当前frame到视频 , 需在video_config中开启配置
```

```
}
```

输入图片不限制大小

SDK参数配置 SDK的参数通过EdgePredictorConfig::set_config和global_controller()->set_config配置。set_config的所有key 在easyedge_xxxx_config.h中。其中

• PREDICTOR前缀的key是不同模型相关的配置,通过EdgePredictorConfig::set_config设置

• CONTROLLER前缀的key是整个SDK的全局配置,通过global_controller()->set_config设置

```
以序列号为例,KEY的说明如下:
```

```
/**
* @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号
* 值类型:string
* 默认值:空
*/
static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";
```

EdgePredictorConfig config; config.model_dir = ...; config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, "1DB7-1111-1111-D27D");

具体支持的运行参数配置列表可以参考开发工具包中的头文件的详细说明。

相关配置均可以通过环境变量的方法来设置,对应的key名称加上前缀EDGE_即为环境变量的key。如序列号配置的环境变量key为EDGE_PREDICTOR_KEY_SERIAL_NUM,如指定CPU线程数的环境变量key为EDGE_PREDICTOR_KEY_CPU_THREADS_NUM。 注意:通过代码设置的配置会覆盖通过环境变量设置的值。

初始化

● 接口

auto predictor = global_controller()->CreateEdgePredictor(config); predictor->init();

若返回非0,请查看输出日志排查错误原因。

预测图像

• 接口

```
/**
* @brief
* 通用接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
     cv::Mat& image, std::vector<EdgeResultData>& result
) = 0;
/**
* @brief
* 批量图片推理接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
     std::vector<cv::Mat>& image, std::vector<std::vector<EdgeResultData>>& result
) = 0;
```

图片的格式务必为opencv默认的BGR, HWC格式。

• 返回格式

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index: // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测、图像分割时才有意义
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割的模型,该字段才有意义
    // 请注意: 图像分割时,以下两个字段会比较大,使用完成之后请及时释放EdgeResultData
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding,游程编码的mask
    // 目标追踪模型,该字段才有意义
    int trackid; // 轨迹id
    int frame; // 处于视频中的第几帧
```

关于矩形坐标

};

x1*图片宽度 = 检测框的左上角的横坐标

EdgeTrackStat track_stat; // 跟踪状态

- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2*图片宽度 = 检测框的右下角的横坐标
- y2 * 图片高度 = 检测框的右下角的纵坐标

关于图像分割mask

cv::Mat mask为图像掩码的二维数组 { {0,0,0,0,0,0,0,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,1,1,1,0,0,0,0}, {0,0,0,0,0,0,0,0,0,0}, {1,1,1,1,0,0,0,0}, {1,1,1,1,0,0,0}, {1,1,1,1,0,0,0}, {1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0,0},{1,1,1,1,0,0,0},\\{1,1,1,0,0,0,0},{1,1,1,1,0,0,0},\\{1,1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,1,0,0,0},{1,1,1,0,0},{1,1,1,0,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,1,0,0},{1,1,1,0

关于图像分割mask_rle

该字段返回了mask的游程编码,解析方式可参考 http demo

以上字段可以参考demo文件中使用opencv绘制的逻辑进行解析

预测视频

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

接口

classVideoDecoding :

/**

- * @brief 获取输入源的下一帧
- * @param frame_tensor
- * @return

*/

virtual int next(FrameTensor &frame_tensor) = 0;

/**

- * @brief 显示当前frame_tensor中的视频帧
- * @param frame_tensor
- * @return
- */

virtual int display(const FrameTensor &frame_tensor) = 0;

/**

* @brief 将当前frame_tensor中的视频帧写为本地视频文件

- * @param frame_tensor
- * @return

virtual int save(FrameTensor &frame_tensor) = 0;

```
/**
```

*/

```
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
* @brief 获取视频的width属性
* @return
*/
virtual int get_width() = 0;
/**
```

```
* @brief 获取视频的height属性
* @return
*/
virtual int get_height() = 0;
```

struct VideoConfig

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type; // 输入源类型
  std::string source_value;   // 输入源地址,如视频文件路径、摄像头index、网络流地址
 int skip_frames{0};
                      // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
  int retrieve_all{false};
                     // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
                      // 在采取抽帧之前设置视频的fps
  int input_fps{0};
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
 bool enable_display{false};
  std::string window_name{"EasyEdge"};
  bool display_all{false};
                       // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                       // 是否存储所有frame, 若为false, 仅存储根据skip_frames抽取的frame
  std::map::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被

Baidu 百度智能云文档

抽取返回,以作为显示或存储用。 input_fps:用于抽帧前设置fps。 resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。 conf:高级选项。部分配置会通过该map来设置。

注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

设置序列号

请在网页控制台中申请序列号,并在init初始化前设置。 LinuxSDK 首次使用需联网授权。

EdgePredictorConfig config; config.set_config(easyedge::params::PREDICTOR_KEY_SERIAL_NUM, "this-is-serial-num");

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

```
EdgeLogConfig log_config;
log_config.enable_debug = true;
global_controller()->set_log_config(log_config);
```

http服务

1. 开启http服务 http服务的启动可以参考demo_serving.cpp文件。

/**

- * @brief 开启一个简单的demo http服务。
- * 该方法会block直到收到sigint/sigterm。
- * http服务里,图片的解码运行在cpu之上,可能会降低推理速度。
- * @tparam ConfigT
- * @param config
- * @param host
- * @param port
- * @param service_id service_id user parameter, uri '/get/service_id' will respond this value with 'text/plain'
- * @param instance_num 实例数量,根据内存/显存/时延要求调整
- * @return

```
*/
template<typename ConfigT>
int start_http_server(
    const ConfigT &config,
    const std::string &host,
    int port,
    const std::string &service_id,
    int instance_num = 1);
```

2. http接口详细说明

开发者可以打开浏览器,http://{设备ip}:24401,选择图片或视频来进行测试。

http 请求方式一:无额外编码 URL中的get参数:

参数	说明		默认值		
threshold	阈值过滤,	0~1	如不提供,则会使用模型的推荐阈值		

Python请求示例 (图片测试,针对图像分类、物体检测、实例分割等模型)

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
        'http://127.0.0.1:24401/',
        params={'threshold': 0.1},
        data=img).json()
```

Python请求示例 (图片测试, 仅针对语义分割模型, 同其他CV模型不同, 语义分割模型输出为灰度图)

```
import requests
with open('./1.jpg', 'rb') as f:
    img_data = f.read()
    res = requests.post('http://127.0.0.1:24401/',
        data=img_data)
with open("gray_result.png", "wb") as fb:
    fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
```

Python请求示例 (视频测试, 注意:区别于图片预测,需指定Content-Type;否则会调用图片推理接口)

```
import requests
with open('./1.mp4', 'rb') as f:
video_data = f.read()
result = requests.post(
    'http://127.0.0.1:24401/',
    params={'threshold': 0.1},
    headers={'Content-Type': 'video'},
    data=video_data).json()
```

http 请求方法二:json格式,图片传base64格式字符串 HTTP方法:POST Header如下:

参数	值
Content-Type	application/json

Body请求填写:

• 图像分类网络: body 中请求示例

```
{
"image": "<base64数据>",
"top_num": 5
}
```

body中参数详情

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
top_nu m	否	number	-	返回分类数量,不填该参数,则默认返回全部分类结果

• 物体检测和实例分割网络: Body请求示例:

```
{
"image": "<base64数据>",
"threshold": 0.3
```

body中参数详情:

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
thresho Id	否	number	-	默认为推荐阈值,也可自行根据需要进行设置

• 语义分割网络: Body请求示例:

```
{
"image": "<base64数据>"
}
```

body中参数详情(语义分割由于模型特殊性,不支持设置threshold值,设置了也没有意义):

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部

Python请求示例(非语义分割模型参考如下代码)

```
import base64
import requests

def main():
    with open("1.jpg 【图片路径】", 'rb') as f:
        result = requests.post("http://{服务ip地址}:24401/", json={
            "image": base64.b64encode(f.read()).decode("utf8")
        })
        # print(result.request.body)
        # print(result.request.headers)
        print(result.content)

if __name__ == '__main__':
        main()
```

Python 请求示例 (针对语义分割模型,同其他CV模型不同,语义分割模型输出为灰度图)

```
import base64
import requests
def main():
    with open("1.jpg 【图片路径】", 'rb') as f:
        res = requests.post("http://{服务ip地址}:24401/", json={"image": base64.b64encode(f.read()).decode("utf8")})
        with open("gray_result.png", "wb") as fb:
        fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
if __name__ == '__main__':
        main()
```

http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考预测图像-返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

返回示例

```
{
    "cost_ms": 52,
    "error_code": 0,
    "results": [
        {
            "confidence": 0.94482421875,
            "index": 1,
            "label": "IronMan",
            "x1": 0.059185408055782318,
            "x2": 0.18795496225357056,
            "y1": 0.14762254059314728,
            "y2": 0.52510076761245728,
            "mask": "...", // 图像分割模型字段
            "trackId": 0, // 目标追踪模型字段
        },
        ]
}
```

其他配置

1. 日志名称、HTTP 网页标题设置

通过global_controller的set_config方法设置:

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_LOG_BRAND, "MY_BRAND");

效果如下:



图片加载失败

2. CPU线程数设置

CPU线程数可通过 EdgePredictorConfig::set_config 配置

```
EdgePredictorConfig config;
config.set_config(easyedge::params::PREDICTOR_KEY_CPU_THREADS_NUM, 4);
```

3. 批量预测设置

```
int batch_size = 2; // 使用前修改batch_size再编译、执行
while (get_next_batch(imgs, img_files, batch_size, start_index)) {
    ...
}
```

GPU 加速版 预测接口 GPU 加速版 SDK 除了支持上面介绍的通用接口外,还支持图片的批量预测,预测接口如下:

```
/**
 * @brief
* GPU加速版批量图片推理接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
     std::vector<cv::Mat>& image
     std::vector<std::vector<EdgeResultData>>& result
) = 0:
/**
* @brief
* GPU加速版批量图片推理接口,带阈值
* @related infer(cv::Mat & image, EdgeColorFormat origin_color_format, std::vector<EdgeResultData> & result, float threshold)
*/
virtual int infer(
       std::vector<cv::Mat> &images,
       EdgeColorFormat origin_color_format,
       std::vector<std::vector<EdgeResultData>> &results
       float threshold
) = 0;
```

批量图片的预测接口的使用要求在调用 init 接口的时候设置一个有效的 PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE,其含义见下方参数配置接口的介绍。

运行参数选项 在上面的内容中我们介绍了如何使用EdgePredictorConfig进行运行参数的配置。针对GPU加速版开发工具包,目前EdgePredictorConfig的运行参数所支持的Key包括如下项:

```
/**
* @brief 当有同类型的多个设备的时候,使用哪一个设备,如:
* GPU: 使用哪张GPU卡
* EdgeBoard(VMX), Movidius NCS:使用哪一张加速卡
* 值类型: int
* 默认值:0
*/
static constexpr auto PREDICTOR_KEY_DEVICE_ID = "PREDICTOR_KEY_DEVICE_ID";
/**
* @brief 生成最大 batch_size 为 max_batch_size 的优化模型,单次预测图片数量可以小于或等于此值(推荐等于此值,见release notes)
* 值类型: int
* 默认值:4
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE = "PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE";
/**
* @brief 设置device对应的GPU卡可以支持的最大并发量
* 实际预测的时候对应GPU卡的最大并发量不超过这里设置的范围,否则预测请求会排队等待预测执行
* 值类型: int
* 默认值:1
*/
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY = "PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY";
/**
* @brief 是否开启fp16模式预测,开启后预测速度会更快,但精度会略有降低。并且需要硬件支持fp16
* 值类型: bool
* 默认值: false
*/
static constexpr auto PREDICTOR_KEY_GTURBO_FP16 = "PREDICTOR_KEY_GTURBO_FP16";
/**
* @brief 模型编译等级
*1:如果当前max_batch_size与历史编译产出的max_batch_size不相等时,则重新编译模型(推荐)
* 2:无论历史编译产出的max_batch_size为多少,均根据当前max_batch_size重新编译模型
* 值类型: int
* 默认值:1
*/
```

static constexpr auto PREDICTOR_KEY_GTURBO_COMPILE_LEVEL = "PREDICTOR_KEY_GTURBO_COMPILE_LEVEL"; /** * @brief GPU工作空间大小设置 * workspace_size = workspace_prefix * (1 << workspace_offset) * workspace_offset: 10 = KB, 20 = MB, 30 = GB * 值类型: int *默认值:WORKSPACE_PREFIX: 100, WORKSPACE_OFFSET: 20,即100MB static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX = "PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX"; static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET = "PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET"; /** * @brief 需要使用的dla core * 值类型: int * 默认值:-1(不使用) static constexpr auto PREDICTOR_KEY_GTURBO_DLA_CORE = "PREDICTOR_KEY_GTURBO_DLA_CORE"; /** * @brief 自定义缓存文件存储路径 * 值类型: string * 默认值: ~/.baidu/easyedge/mcache/{model_id * 1000000 + release_id} static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_DIR = "PREDICTOR_KEY_GTURBO_CACHE_DIR"; /** * @brief 自定义缓存文件命名,默认即可 * 值类型: string *默认值:根据配置自动生成 static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_NAME = "PREDICTOR_KEY_GTURBO_CACHE_NAME"; /** * @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号 * 值类型: string *默认值:空 static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";

PREDICTOR_KEY_GTURBO_CACHE_NAME:首次加载模型会先对模型进行编译优化,通过此值可以设置优化后的产出文件名,这在多进程加载同一个模型的时候是有用的。

PREDICTOR_KEY_GTURBO_CACHE_DIR:首次加载模型经过编译优化后,产出的优化文件会存储在这个位置,可以按需修改。

PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX、PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET:设置运行时可以被用来使用的最大临时显存。

PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE:此值用来控制批量图片预测可以支持的最大图片数,实际预测的时候单次预测图片数需等于此值。

PREDICTOR_KEY_DEVICE_ID:设置需要使用的 GPU 卡号。

PREDICTOR_KEY_GTURBO_COMPILE_LEVEL: 模型编译等级。通常模型的编译会比较慢,但编译产出是可以复用的。可以在第一次加载模型的时候设置合理的 max_batch_size 并在之后加载模型的时候直接使用历史编译产出。是否使用历史编译产出可以通过此值 compile_level 来控制。当此值为 0 时,表示忽略当前设置的 max_batch_size 而仅使用历史产出(无历史产出时则编译模型);当此值为 1 时,会比较历史产出和当前设置的 max_batch_size 是否相等,如不等,则重新编译;当此值为 2 时,无论如何都会重新编译模型。

PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY:通过此值设置单张 GPU 卡上可以支持的最大 infer 并发量,其上限取决于硬件限制。init 接口 会根据此值预分配 GPU 资源,建议结合实际使用控制此值,使用多少则设置多少。注意:此值的增加会降低单次 infer 的速度,建议优先考虑 batch inference 和 multi predictor。

PREDICTOR_KEY_GTURBO_FP16:默认是 fp32 模式,置 true 可以开启 fp16 模式预测,预测速度会有所提升,但精度也会略微下降,权衡使用。注意:不是所有模型都支持 fp16 模式。目前已知不支持fp16的模型包括:图像分类高精度模型。

多线程预测 GPU 加速版 SDK 的多线程分为单卡多线程和多卡多线程两种。 单卡多线程:创建一个 predictor,并通过 PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY 控制单卡所支持的最大并发量,只需要 init 一次,多线程调用 infer 接口。 多卡多线程:多卡的 支持是通过创建多个 predictor,每个 predictor 对应一张 GPU 卡,predictor 的创建和 init 的调用放在主线程,通过多线程的方式调用 infer 接口。

已知问题 1. 多线程时图片按线程分配不均 或 不同batch size的图片交叉调用infer接口时,部分结果错误 A: EasyDL图像分类高精度模型在有些显卡上可能存在此问题,可以考虑填充假图片数据到图片比较少的线程或batch以使得infer间的图片绝对平均。

2. 显存持续增长或遇到 terminate called after throwing an instance of 'std::runtime_error' what(): Failed to create object A: 部分显卡存在此问题, 如果遇到此问题, 请确认没有频繁调用 init 接口, 通常调用 infer 接口即可满足需求。

3. 开启 fp16 后,预测结果错误 A:不是所有模型都支持 fp16 模式。目前已知的不支持fp16的模型包括:图像分类高精度模型。目前不支持的 将会在后面的版本陆续支持。

昆仑服务器 昆仑服务器SDK支持将EasyDL的模型部署到昆仑服务器上。SDK提供的接口风格一致,简单易用,轻松实现快速部署。Demo的测试 可参考上文中的测试Demo部分。

参数配置接口 在上面的内容中我们介绍了如何使用EdgePredictorConfig进行运行参数的配置。针对昆仑服务器开发工具包,目前EdgePredictorConfig的运行参数所支持的Key包括如下项:

```
/**
* @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号
* 值类型: string
*默认值:空
*/
static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";
/**
* @brief 当有同类型的多个设备的时候,使用哪一个设备,如:
* 使用哪张加速卡
* 值类型: int
*默认值:0
*/
static constexpr auto PREDICTOR_KEY_DEVICE_ID = "PREDICTOR_KEY_DEVICE_ID";
/**
* @brief 设置需要同时预测的图片数量
* 值类型: int
*默认值:1
*/
static constexpr auto PREDICTOR_KEY_KUNLUN_BATCH_SIZE = "PREDICTOR_KEY_KUNLUN_BATCH_SIZE";
```

PREDICTOR_KEY_DEVICE_ID:设置需要使用的加速卡的卡号。

PREDICTOR_KEY_KUNLUN_BATCH_SIZE:设置单次预测可以支持的图片数量。

使用方法:

```
int batch_size = 1;
config.set_config(easyedge::params::PREDICTOR_KEY_KUNLUN_BATCH_SIZE, batch_size);
```

模型调优通过设置如下环境变量,可以在初始化阶段对模型调优,从而让预测的速度更快。

export XPU_CONV_AUTOTUNE=5

FAQ

1. 如何处理一些 undefined reference?

如: undefined reference to `curl_easy_setopt@CURL_OPENSSL_3'

方案1:通过安装libcurl3 libcurl-openssl1.0-dev来解决。 方案2:如果开发者想不想使用低版本的openssl (如Ubuntu 18.04),可以link静态 库easyedge_static.a,自己指定需要的Library的版本:

示例:修改CMakeList.txt

find_package(CURL REQUIRED)

target_link_libraries(easyedge_demo \${OpenCV_LIBS} easyedge_static pthread \${CURL_LIBRARIES} verify_static \$(其他需要的库))

其中,其他需要的库视具体sdk中包含的库而定。

2. EasyDL SDK与云服务效果不一致,如何处理?

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

3. NVIDIA GPU预测时,报错显存不足如以下错误字样:

paddle.fluid.core.EnforceNotMet: Enforce failed. Expected allocating <= available, but received allocating:20998686233 > available:19587333888. Insufficient GPU memory to allocation. at [/paddle/paddle/fluid/platform/gpu_info.cc:170]

请根据显存大小和模型配置。调整合适的初始 fraction_of_gpu_memory。参数的含义参考这里。

4. 如何将我的模型运行为一个http服务?目前cpp sdk暂未集成http运行方式; 0.4.7版本之后,可以通过start_http_server方法开启http服务。

5. 运行NNIE引擎报permission denied 日志显示:

open sys: Permission denied open err : Permission denied open err : Permission denied

请使用sudo在root下运行。

6. 运行SDK报错 Authorization failed

情况一:日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888" ./easyedge_demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除~/.baidu/easyedge 目录,再重新激活。

7. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

8. 运行二进制时,提示 libverify.so cannot open shared object file

可能cmake没有正确设置rpath,可以设置LD_LIBRARY_PATH为sdk的lib文件夹后,再运行:

LD_LIBRARY_PATH=\$LD_LIBRARY_PATH:./lib ./easyedge_demo

9. 运行二进制时提示 libopencv_videoio.so.4.5: cannot open shared object file: No such file or directory 同上面8的问题类似,没有正确设置动态 库的查找路径,可通过设置LD_LIBRARY_PATH为sdk的thirdparty/opencv/lib文件夹解决 export LD_LIBRARY_PATH=\${LD_LIBRARY_PATH}:../thirdparty/opencv/lib (tips: 上面冒号后面接的thirdparty/opencv/lib路径以实际项目中路径为准,比如也可能是../../thirdparty/opencv/lib)

10. 编译时报错:file format not recognized 可能是因为在复制SDK时文件信息丢失。请将整个压缩包复制到目标设备中,再解压缩、编译

11. 进行视频解码时,报错符号未找到、格式不支持、解析出的图片为空、无法设置抽帧 请确保安装OpenCV时,添加了-DWITH_FFMPEG=ON选项(或者GStream选项),并且检查OpenCV的安装日志中,关于 Video I/O 段落的说明是否为YES。

- -- Video I/O:-- DC1394: YES (ver 2.2.4)
- -- FFMPEG: YES
- -- avcodec: YES (ver 56.60.100)
- -- avformat: YES (ver 56.40.101)
- -- avutil: YES (ver 54.31.100) -- swscale: YES (ver 3.1.101)
- -- swscale: YES (ver 3 -- avresample: NO
- -- libv4l/libv4l2: NO
- -- v4l/v4l2: linux/videodev2.h

如果为NO,请搜索相关解决方案,一般为依赖没有安装,以apt为例:

apt-get install yasm libjpeg-dev libjasper-dev libavcodec-dev libavformat-dev libswscale-dev libdc1394-22-dev libgstreamer0.10-dev libgstreamer0.10-dev libgstreamer-plugins-base0.10-dev libv4l-dev python-dev python-numpy libtbb-dev libqt4-dev libgtk2.0-dev libfaac-dev libmp3lame-dev libopencore-amrnb-dev libopencore-amrnb-dev libtheora-dev libtheora-dev libtvorbis-dev libtvidcore-dev x264 v4l-utils ffmpeg

12. GPU加速版运行有损压缩加速的模型,运算精度较标准模型偏低首先请保证数据集不会太小。其次可以通过将模型目录RES中的calibrationtable移除,并通过将PREDICTOR_KEY_GTURBO_FP16设置为true,使用FP16的运算精度重新评估模型效果。若依然不理想,可将calibrationtable移除并将PREDICTOR_KEY_GTURBO_FP16设置为false,从而使用更高精度的FP32的运算精度。

の Linux集成文档-Python

简介

本文档介绍 EasyDL 的 Linux Python SDK 的使用方法,适用于 EasyDL 和 BML。

EasyDL 通用版:

- 网络类型支持:图像分类,物体检测,图像分割,声音分类,表格预测
- 硬件支持:
 - Linux x86_64 CPU (基础版,加速版)
 - Linux x86_64 Nvidia GPU (基础版,加速版)
- 语言支持: Python 3.5, 3.6, 3.7

BML:

- 网络类型支持:图像分类,物体检测,声音分类
- 硬件支持:
 - Linux x86_64 CPU (基础版)
 - Linux x86_64 Nvidia GPU (基础版)
- 语言支持: Python 3.5, 3.6, 3.7

Release Notes

Baidu 百度智能云文档

时间	版本	说明					
2023-03-16	1.3.7	迭代升级,新增支持文本类模型; 新增GPU 多卡多进程推理demo					
2022.10.27	1.3.5	新增华为Atlas300、飞腾Atlas300 Python SDK,支持图像分类、物体检测、人脸检测、实例分割					
2022.09.15	1.3.3	EasyDL CPU普通版新增支持表格预测					
2022.05.27	1.3.1	CPU、GPU普通版新增支持BML Cloud小目标检测模型					
2021.12.22	1.2.7	声音分类模型升级					
2021.10.20	1.2.6	CPU基础版、CPU加速版、GPU基础版推理引擎优化升级					
2021.08.19	1.2.5	CPU基础版、CPU无损加速版、GPU基础版新增支持EasyDL小目标检测					
2021.06.29	1.2.4	CPU、GPU新增EasyDL目标跟踪支持;新增http server服务启动demo					
2021.03.09	1.2.2	EasyDL CPU加速版新增支持分类、高性能检测和均衡检测的量化压缩模型					
2021.01.27	1.2.1	EasyDL经典版分类高性能模型升级;支持更多模型					
2020.12.18	1.2.0	推理引擎升级;接口升级;性能优化					
2020.09.17	1.1.19	支持更多模型					
2020.08.11	1.1.18	生能优化					
2020.06.23	1.1.17	支持更多EasyDL专业版模型					
2020.04.16	1.1.15	支术优化;升级 OpenVINO 版本					
2020.03.12	1.1.14	新增声音识别python sdk					
2020.02.12	1.1.13	新增口罩模型支持					
2020.01.16	1.1.12	须测函数默认使用推荐阈值					
2019.12.26	1.1.11	EasyDL 专业版支持 SDK 加速版					
2019.12.04	1.1.10	支持图像分割					
2019.10.21	1.1.9	支持 EasyDL 专业版					
2019.08.29	1.1.8	CPU 加速版支持					
2019.07.19	1.1.7	提供模型更新工具					
2019.05.16	1.1.3	NVIDIA GPU 支持					
2019.03.15	1.1.0	架构与功能完善					
2019.02.28	1.0.6	引擎功能完善					
2019.02.13	1.0.5	paddlepaddle 支持					
2018.11.30	1.0.0	第一版!					

2020-12-18: 【接口升级】 序列号的配置接口从1.2.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。 请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

快速开始

1. 安装依赖

- 根据引擎的不同, SDK 依赖了不同的底层引擎。根据所需自行安装。
- 使用声音分类SDK需要安装额外依赖*pip 安装 resampy pydub six librosa 音频默认格式支持wav文件预测,如果需要预测mp3等其他音频格式的数据需要系统额外安装ffmpeg (windows系统的ffmpeg已基在sdk中无需额外安装, linux系统需要手动安装)
- 使用表格预测SDK需要安装额外依赖 pip安装 brotlipy==0.7.0 certifi==2020.6.20 joblib==1.0.1 kaggle==1.5.12 Pillow py4j pycosat pythondateutil python-slugify ruamel_yaml text-unidecode threadpoolctl flask pandas==1.0.5 scikit-learn==0.23.2 lightgbm==2.2.3 catboost==0.24.1 xgboost==1.2.0 numpy==1.19.5 scipy==1.5.2 psutil==5.7.2 pypmml==0.9.7 torch==1.8.0 jieba==0.42.1 pyod==0.8.5 pyarrow==6.0.0 scikit-optimize==0.9.0 pyspark==3.3.0 另外ml算法 安装(目前只支持python3.7) pip install BaiduAl_TabularInfer-0.0.0-cp37-cp37m-linux_x86_64.whl 安装 paddlepaddle
- 使用x86_64 CPU 基础版预测时必须安装(目标跟踪除外):

python -m pip install paddlepaddle==2.2.2 -i https://mirror.baidu.com/pypi/simple

若 CPU 为特殊型号,如赛扬处理器(一般用于深度定制的硬件中),请关注 CPU 是否支持 avx 指令集。如果不支持,请在paddle官网安装 noavx 版本

• 使用NVIDIA GPU 基础版预测时必须安装(目标跟踪除外):

python -m pip install paddlepaddle-gpu==2.2.2.post101 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA10.1的 PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2 -i https://mirror.baidu.com/pypi/simple #CUDA10.2的PaddlePaddle
python -m pip install paddlepaddle-gpu==2.2.2.post110 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.0的
PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2.post111 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.1的 PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2.post112 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.2的 PaddlePaddle

不同cuda版本的环境,请参考paddle文档安装合适的 paddle 版本。 不被 paddle 支持的 cuda 和 cudnn 版本,EasyEdge 暂不支持**安装 OpenVINO** 使用x86_64 CPU 加速版 SDK 预测时必须安装。

1) 请参考 OpenVINO toolkit 文档安装 2021.4版本, 安装时可忽略Configure the Model Optimizer及后续部分

2) 运行之前,务必设置环境变量

source /opt/intel/openvino_2021/bin/setupvars.sh

安装 cuda、cudnn

• 使用Nvidia GPU 加速版预测时必须安装。依赖的版本为 cuda9.0、cudnn7。版本号必须正确。

安装 pytorch (torch >= 1.7.0)

- 目标跟踪模型的预测必须安装pytorch版本1.7.0及以上(包含:Nvidia GPU 基础版、x86_64 CPU 基础版)。
- 目标跟踪模型Nvidia GPU 基础版还需安装依赖cuda、cudnn。

关于不同版本的pytorch和CUDA版本的对应关系: pytorch官网目标跟踪模型还有一些列举在requirements.txt里的依赖(包括torch >= 1.7.0),均可使用pip下载安装。

pip3 install -r requirements.txt

2. 安装 easyedge python wheel 包 安装说明

pip3 install -U BaiduAI_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_x86_64.whl

具体名称以 SDK 包中的 whl 为准。 安装说明:华为 Atlas300 除了需要安装BaiduAI_EasyEdge_SDK包,还需安装

pip3 install -U EasyEdge_Devkit_Atlas300-{版本号}-cp36-cp36m-linux_x86_64.whl

安装说明:飞腾 Atlas300 除了需要安装BaiduAl_EasyEdge_SDK包,还需安装

pip3 install -U EasyEdge_Devkit_Phytium.Atlas-{版本号}-cp36-cp36m-linux_aarch64.whl

3. 使用序列号激活

	- 天東高坂第710時 - 天南京坂第710時 - 安市高泉市、将国佐見広竹模型部署在木地、高) - 遊行為早に成立。 天市高田島 - 佐知白	发给装量越感,将刘浩完起的微型基础在生地,虽然周用模型,可以选择将模型基础在本地的服务器、小型设备、软硬一体方案专项适配操作上, 通过Am, SDA进一步集成,民活适应不同业务结果。 										
获取序列号												
	模型名称	发布版本 关	应用平台	模型加速	发布状态	发布时间 🗘						
		124318V1 查看性能报告	通用X86 CPU-Linux	基础版	 已发布 	2021-08-19 20:24	下载SDK					
				精度无损压缩加速	 已发布 	2021-08-19 20:24	下载加速版SDK					
	oup du El Etast		英伟达GPU-Linux	基础板	 已发布 	2021-08-19 20:35	下载SDK					
	anu Crum Allinear			精度无损压缩加速	 已发布 	2021-08-19 20:34	下载加速版SDK					
				基础板	 已发布 	2021-08-19 18:17	下截SDK					

修改demo.py 填写序列号

pred = edge.Program() pred.set_auth_license_key("这里填写序列号")

4. GPU 加速版 使用 GPU 加速版,在安装完 whl 之后,必须:

1. 从这里下载 TensorRT7.0.0.11 for cuda9.0,并把解压后的 lib 放到 C++ SDK 的 lib 目录或系统 lib 目录

2. 运行时,必须在系统库路径中包含 C++ SDK 下的lib目录。如设置LD_LIBRARY_PATH

```
cd ${SDK_ROOT}
**1. 安装 python wheel 包**
tar -xzvf python/*.tar.gz
pip install -U (对应 Python 版本的 wheel 包}
**2. 设置 LD_LIBRARY_PATH**
tar -xzvf cpp/*.tar.gz
```

export EDGE_ROOT=\$(readlink -f \$(ls -h | grep "baidu_easyedge_linux_cpp")) export LD_LIBRARY_PATH=\$EDGE_ROOT/lib

3. 运行 demo python3 demo.py {RES文件夹路径} {测试图片路径}

如果是使用 C++ SDK 自带的编译安装的 OpenCV, LD_LIBRARY_PATH 还需要包括 C++ SDK的 build 目录下的 thirdparty/lib 目录

```
如果没有正确设置 LD_LIBRARY_PATH,运行时可能报错:
```

ImportError: libeasyedge.so.0.4.3: cannot open shared object file: No such file or directory ImportError: libopencv_core.so.3.4: cannot open shared object file: No such file or directory

5. 测试 Demo

5.1 图片预测

输入对应的模型文件夹(默认为RES)和测试图片路径,运行:

python3 demo.py {model_dir} {image_name.jpg}

测试效果:



图片加载失败

5.2 视频预测 (适用于目标跟踪)

输入对应的模型文件夹(默认为RES)和测试视频文件路径/摄像头id/网络视频流地址,运行:

```
**video_type: 输入源类型 type:int**
**1 本地视频文件**
**2 摄像头的index**
**3 网络视频流**
**video_src: 输入源地址,如视频文件路径、摄像头index、网络流地址 type: string**
python3 demo.py {model_dir} {video_type} {video_src}
```

```
6. 测试Demo HTTP 服务 输入对应的模型文件夹(默认为RES)、序列号、设备ip和指定端口号,运行:
```

 $python 3 \ demo_serving.py \ \{model_dir\} \ \{serial_key\} \ \{host, \ default \ 0.0.0.0\} \ \{port, \ default \ 24401\}$

后,会显示:

Running on http://0.0.0.0:24401/

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片或者视频来进行测试。也可以参考`demo_serving.py`里 http_client_test()函数请求http服务进行推理。



使用说明

使用流程 demo.py

import BaiduAI.EasyEdge as edge

pred = edge.Program() pred.set_auth_license_key("这里填写序列号") pred.init(model_dir={RES文件夹路径}, device=edge.Device.CPU, engine=edge.Engine.PADDLE_FLUID) pred.infer_image({numpy.ndarray的图片}) pred.close()

demo_serving.py

import BaiduAl.EasyEdge as edge from BaiduAl.EasyEdge.serving import Serving

 $server = Serving(model_dir=\{RES文件 夹路径\}, \ license=serial_key)$

```
**请参考同级目录下demo.py里:**
```

```
** pred.init(model\_dir=xx, \ device=xx, \ engine=xx, \ device\_id=xx)**
```

对以下参数device\device_id和engine进行修改

 $server.run(host=host, \ port=port, \ device=edge.Device.CPU, \ engine=edge.Engine.PADDLE_FLUID)$

初始化

• 接口

```
def init(self,
    model_dir,
    device=Device.CPU,
    engine=Engine PADDLE_FLUID,
    config_file='conf.json',
    preprocess_file='preprocess_args.json',
    model_file='model',
    params_file='params'
    label_file='label_list.txt',
    infer_cfg_file='infer_cfg.json',
    device_id=0,
    thread_num=1
    ):
   ....
  Args:
    model_dir: str
     device: BaiduAI.EasyEdge.Device,比如:Device.CPU
     engine: BaiduAI.EasyEdge.Engine,比如:Engine.PADDLE_FLUID
     config_file: str
     preprocess_file: str
     model_file: str
     params_file: str
     label_file: str 标签文件
     infer_cfg_file: 包含预处理、后处理信息的文件
device_id: int 设备ID
    thread_num: int CPU的线程数
  Raises:
     RuntimeError, IOError
```

RuntimeError, IOError Returns: bool: True if success

使用 NVIDIA GPU 预测时,必须满足:

- 机器已安装 cuda, cudnn
- 已正确安装对应 cuda 版本的 paddle 版本
- 通过设置环境变量 FLAGS_fraction_of_gpu_memory_to_use 设置合理的初始内存使用比例

使用 CPU 预测时,可以通过在 init 中设置 thread_num 使用多线程预测。如:

 $pred.init (model_dir=_model_dir,\ device=edge.Device.CPU,\ engine=edge.Engine.PADDLE_FLUID,\ thread_num=1)$

预测图像

● 接口

def infer_image(self, img,
threshold=0.3,
channel_order='HWC',
color_format='BGR',
data_type='numpy'):
Args:
img: np.ndarray or bytes
threshold: float
only return result with confidence larger than threshold
channel_order: string
channel order HWC or CHW
color_format: string
color format order RGB or BGR
data_type: string
image data type
Returns:
list

• 返回格式: [dict1, dict2, ...]

字段	类型	取值	说明
confidence	float	0~1	分类或检测的置信度
label	string		分类或检测的类别
index	number		分类或检测的类别
x1, y1	float	0~1	物体检测,矩形的左上角坐标 (相对长宽的比例值)
x2, y2	float	0~1	物体检测,矩形的右下角坐标(相对长宽的比例值)
mask	string/numpy.ndarray	图像分割的mask	

关于矩形坐标

```
x1 * 图片宽度 = 检测框的左上角的横坐标
```

- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2 * 图片宽度 = 检测框的右下角的横坐标
- y2 * 图片高度 = 检测框的右下角的纵坐标

可以参考 demo 文件中使用 opencv 绘制矩形的逻辑。

结果示例

• i) 图像分类

```
{
"index": 736,
"label": "table",
"confidence": 0.9
}
```

● ii) 物体检测

Baidu 百度智能云文档

```
"y2": 0.91211,
"label": "cat",
"confidence": 1.0,
"x2": 0.91504,
"index": 8,
"y1": 0.12671,
"x1": 0.21289
}
```

• iii) 图像分割

```
{
    "name": "cat",
    "score": 1.0,
    "location": {
        "left": ...,
        "top": ...,
        "width": ...,
        "height": ...,
    },
    "mask": ....
}
```

mask字段中,data_type为numpy时,返回图像掩码的二维数组

```
{
{0,0,0,0,0,0,0,0,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,0,0,0,0,0,0,0},
}
其中1代表为目标区域,0代表非目标区域
```

data_type为string时,mask的游程编码,解析方式可参考 demo

预测视频 (目前仅限目标跟踪模型调用)

```
• 接口
```

```
def infer_frame(self, frame, threshold=None):
"""
视频推理(抽帧之后)
:param frame:
:param threshold:
:return:
"""
```

```
● 返回格式dict
```

字段	类型	说明
pos	dict1	当前帧每一个类别的追踪目标的像素坐标(tlwh)
id	dict2	当前帧每一个类别的追踪目标的id
score	dict3	当前帧每一个类别的追踪目标的识别置信度
label	dict4	class_idx(int)与label(string)的对应关系
class_num	int	追踪类别数

预测声音

• 使用声音分类SDK需要安装额外依赖 *pip 安装 resampy pydub* 音频默认格式支持wav文件预测,如果需要预测mp3等其他音频格式的数据需要 系统额外安装ffmpeg (windows系统的ffmpeg已集成在sdk中无需额外安装,linux系统需要手动安装)

接口

def infer_sound(self, sound_binary, threshold=0.3):
Args:
sound_binary: sound_binary
threshold: confidence
Returns:
list

• 返回格式: [dict1, dict2, ...]

字段	类型	取值	说明
confidence	float	0~1	分类的置信度
label	string		分类的类别
index	number		分类的类别

升级模型 适用于经典版升级模型,执行bash update_model.sh</mark>,根据提示,输入模型路径、激活码、模型ID、模型版本,等待模型更新完毕即 可。

FAQ

Q: EasyDL 离线 SDK 与云服务效果不一致,如何处理? A: 后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

Q: 运行时报错"非法指令"或"illegal instruction"A: 可能是 CPU 缺少 avx 指令集支持,请在paddle宫网 下载 noavx 版本覆盖安装

Q: NVIDIA GPU预测时,报错显存不足: A: 如以下错误字样:

paddle.fluid.core.EnforceNotMet: Enforce failed. Expected allocating <= available, but received allocating:20998686233 > available:19587333888.

Insufficient GPU memory to allocation. at [/paddle/paddle/fluid/platform/gpu_info.cc:170]

请在运行 Python 前设置环境变量,通过export FLAGS_fraction_of_gpu_memory_to_use=0.3来限制SDK初始使用的显存量,0.3表示初始使用 30%的显存。如果设置的初始显存较小,SDK 会自动尝试 allocate 更多的显存。

Q:我想使用多线程预测,怎么做?如果需要多线程预测,可以每个线程启动一个Progam实例,进行预测。 demo.py文件中有相关示例代码。

注意: 对于CPU预测,SDK内部是可以使用多线程,最大化硬件利用率。参考init的thread_num参数。

Q: 运行SDK报错 Authorization failed

情况一:日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888"
./easyedge_demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- MAC地址变化
- 磁盘变更
- BIOS重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 ~/.baidu/easyedge 目录,再重新激活。

情况三: Atlas Python SDK日志提示ImportError: libavformat.so.58: cannot open shared object file: No such file or directory 或者其他类似so找不到可以在LD_LIBRARY_PATH环境变量加上libs和thirdpartylibs路径,例如

export LD_LIBRARY_PATH=/xxx/libs:/xxx/thirdpartylibs:\$LD_LIBRARY_PATH # tips: 这里/xxx需要替换为真实路径,/xxx路径查找方法如下

查找安装包内libs和thirdpartylibs路径的方法如下(以华为Atlas300 SDK为例,其他SDK查找方法类似):

pip3 show EasyEdge-Devkit-Atlas300 # 结果中会显示 Location 路径,也就是包的安装路径 **libs和thirdpartylibs两个路径在 Location 所指示的路径 easyedge_CANN 子文件夹下**

心 纯离线API集成说明

本文档主要说明定制化图像分割模型发布为本地服务器API(通过API部署包实现)后如何使用。如还未训练模型,请先前往EasyDL进行训练。

如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

部署包使用说明

部署方法

EasyDL定制化图像分割模型的服务器API通过EasyPack实现,目前提供单机一键部署的方式。

在EasyDL申请、下载部署包后,在本地服务器新建目录(建议目录命名规则:easyDL_服务名称_模型版本号),将软件包上传至该目录。请参考EasyPack-单机一键部署使用python2版本来部署,部署成功后,启动服务,即可调用与在线API功能类似的接口。

运维检查

EasyDL服务器API部署应用健康检查(或故障排查)脚本:trouble_shooting.tar

脚本能力:鉴权服务健康检测、容器状态检查、端口探活、网络联通性测试、容器关键报错日志输出等

使用方法:将脚本上传至服务器任意目录(或在服务器直接下载),并解压后运行。

解压 tar vxf trouble_shooting.tar **执行** bash trouble_shooting.sh

授权说明

部署包根据服务器硬件(CPU单机或GPU单卡)进行授权,只能在申请时提交的硬件指纹所属的硬件上使用。

部署包测试期为1个月,如需购买永久授权,可提交工单咨询

性能指标

图像分割模型可部署在CPU或GPU服务器上,单实例具体性能指标参见算法性能及适配硬件

接口描述

```
请求说明
```

```
请求示例
```

HTTP 方法: POST

请求URL:请首先在EasyDL进行自定义模型训练,完成训练后申请部署包,部署成功后拼接url。

请求URL: http://{IP}:{PORT}/{DEPLOY_NAME}/ImageSegmentation IP:服务部署所在机器的ip地址 PORT:服务部署后获取的端口 DEPLOY_NAME:申请时填写的服务名称

Header如下:

参数	值
о т	and the attack for

Content-Type application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,与其他图像识别服务不同的是定制化图像识别服务以json方式请求。

Body请求示例:

```
{
"image": "<base64数据>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大4096px,支持 jpg/png/bmp格式 注意请去掉头部
thresho Id	否	number	-	默认值为推荐阈值,请在我的模型列表-模型效果查看推荐阈值

返回说明

返回参数

字段	是否必 选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(o bject)	识别结果数组
+name	否	string	分类名称
+score	否	number	置信度
+locati on	否		
++left	否	number	检测到的目标主体区域到图片左边界的距离
++top	否	number	检测到的目标主体区域到图片上边界的距离
++width	否	number	检测到的目标主体区域的宽度
++heig ht	否	number	检测到的目标主体区域的高度
+mask	否	array	基于游程编码的字符串,编码内容为和原图宽高相同的布尔数组:若数组值为0,代表原图此位置像素点不属于检测 目标,若为1,代表原图此位置像素点属于检测目标。查看解码示例

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 336001,
    "error_msg": "Invalid Argument"
}
```

错误码	错误信息	描述
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度智能云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有疑 问请请在百度智能云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度智能云控制台内提交工单反馈
336003	Base64解码失败	图片/音频/文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。 有疑问请在百度智能云控制台内提交工单反馈
336004	输入文件大小不合法	图片超出大小限制,图片限4M以内,请根据接口文档检查入参格式,有疑问请在百度智能云控制台内提交工 单反馈
336005	图片解码失败	图片编码错误(非jpg,bmp,png等常见图片格式),请检查并修改图片格式
336006	缺失必要参数	image字段缺失(未上传图片)
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度智能云控制台内提 交工单反馈
337000	Auth check failed	离线鉴权调用失败,

模型更新/回滚操作说明

模型更新

1、在EasyDL-纯离线服务发布页面,找到您的服务器API发布记录,点击【更新版本】,选择「更新包」或「完整包」来发布。

两者区别:

包类型	描述	
更新包	仅包含最新的模型应用,需执行download.sh脚本下载所需镜像等依赖文件	
完整包	包含模型应用和其他鉴权服务,需执行download.sh脚本下载所需完整依赖文件	

2、(CPU模型可忽略)如果您训练的模型为GPU版本,系统会生成多份下载链接。请在GPU服务器执行 nvidia-smi命令,根据返回的Cuda Version来选择对应的部署包链接下载。

3、在服务器新建目录(建议标记对应模型的版本号,便于区分不同模型版本),如easydl_\${DEPLOY_NAME}_v2

\${DEPLOY_NAME}:申请时填写的服务名称

以如下场景举例说明:模型版本从V1升级至V2

```
**1.新模型准备**
**创建指定版本的目录**
mkdir easedI_${DEPLOY_NAME}_v2
cd easedI_${DEPLOY_NAME}_v2
**将部署包上传至服务器该目录并解压**
tar zvxf xx.tar.gz
**解压后,进入指定目录执行download脚本下载模型所依赖文件**
cd original && bash download.sh
**2.旧模型备份**
**历史模型备份**
cp -r /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V1
**记录当前模型的端口号**
docker ps -a |grep ${DEPLOY_NAME}
**3.模型升级**
cd package/Install
**卸载当前已安装的旧的easyDL服务: ${DEPLOY_NAME},前面已备份**
python2 install.py remove ${DEPLOY_NAME}
**安装当前部署包内新的EasyDL服务: ${DEPLOY_NAME}**
python2 install.py install ${DEPLOY_NAME}
```

(可选操作) 更新证书 python2 install.py lu Baidu 百度智能云文档

模型回滚

以如下场景举例说明:模型版本从V2回滚至V1

方法一:

```
**重命名当前v2模型目录名称**
mv /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V2
**使用V1版本**
cp -r /home/baidu/work/${DEPLOY_NAME}_V1 /home/baidu/work/${DEPLOY_NAME}
**停止当前模型容器**
docker ps -a |grep ${DEPLOY_NAME}
docker rm -f ${容器名}
**创建新的容器**
cd /home/baidu/work/${DEPLOY_NAME} && bash start/start-1.sh
*** (可选操作)进入V1版本部署包所在目录执行license更新操作,假如部署包在/opt目录下,以您实际目录为准**
cd /opt/easydl_${DEPLOY_NAME}_V1/original/package/Install && python2 install.py lu
```

方法二:

进入模型V1所在目录,参考上述【模型更新】步骤,执行模型升级操作(即先卸载v2,后升级为v1)

心 端云协同服务说明

服务简介

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

「云管理,端计算」的端云协同服务,具体包括:

- 在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新
- 断网状态下模型离线计算(http服务,可调用与公有云API功能相同的接口)
- 联网状态下在平台管理设备运行状态、资源利用率

目前本地服务器的应用平台支持Linux-AMD64(x86-64),具体使用流程请参考下方文档。

使用流程

Step 1 发布端云协同部署包

在我的部署包页面点击「发布端云协同部署包」

ED Adv 199 Titl		_						
我时候型	端云协同服务说明							点击收起
创建模型	1、在可视化界面轻松实现模型部	8署包在边缘设备。 ************************************	上的集成、版本更	新				
训练模型	 2、前网状态下候望离线计算(3、联网状态下在平台管理设备; 	a行状态、资源利	ラ公有 ZAPIJJilet 用率	데이미/1☆니)				
校验模型	具体使用流程如下:							
发布模型	1、在本页面发布端云协同部署	回,填写相关信息; 5 ⁴⁴³⁰ 条 复制激;	井申请,并在下方 王念念 - 昭岡激汗	列表查看部署包发行	市状态			
EasyData数据服务	2、在·我的本地设备」贝固制 3、在「下发部署包到设备」『	(咱议田,复制淑) 〔面,将发布成功的	6 9 7 , 获网激活 的部署包一键下发:	平地设面 到已激活的设备上,	即可测试或正式使用			
数据总览		1						
	发布端云协同部署包							
防亚组首理	端云协同服务名称	模型ID	设备类型	应用平台	当前可用版本	新版本发布状态	部署设备数	操作
在线标注								
云服务数据回流								
EasyEdge本地部署								
纯离线服务								
端云协同服务 ヘ								
我的部署包					暂无可用	数据		
带的主体记名					请稍后再证			

填写服务名称,选择模型版本并提交发布

模型中心	端云协同服务 > 发布端云协同部署包	
我的模型	设备类型 💿 服务器 🔵 通用小型设备	
创建模型		
训练模型	模型名称 识别模桃和小番茄	
校验模型	端云协同服务名称	
发布模型	选择版本 V1 💙	
EasyData数据服务	选择系统和芯片 🔄 🛆 Linux	
数据总览	通用X86 CPU	
标签组管理		
在线标注	发布部署包	
云服务数据回流		

在列表查看部署包发布状态

模型中心	端云协同服务 > 我的部署包
我的模型	端云协同服务说明
创建模型	1、在可視化界區轻松实現模型認著包在边緣设备上的集成、版本更新
训练模型	 新网状态下模型属线计算(http服务,可调用与公有云API功能相同的接口) 联网状态下在平台管理设备运行状态、资源利用率
校验模型	具体使用流程如下:
发布模型	 在本页面发布端云协网部署包,填写铝关信息并申请,并在下方列表查看部署包发布状态 在「影約本地984」页面断境设备、暂间联天命令、联网联系本地984。
EasyData数据服务	3.在「 大发部者包到议备 」页面,将发布成功的部署包一幅下发到已激活的设备上。即可测试成正式使用
数据总览	发布端云协同影客包
标签组管理	
在线标注	端云协同服务名称 模型ID 设备类型 应用平台 当前可用版本 新版本发布状态 部署设备数 操作
一服务数据回道	ecc 264 通用小型设备 Linux-通用ARM V1 V1-E发布 0 下发到设备 发布新版本
LURD KARLEINS	从发布页过来的 246 服务器 Linux-通用X86 CPU V2, V1 V2-已发布 1 下发到设备 发布新版本 服务详情
EasyEdge本地部署	图像分类高精度 猫狗-265 265 通用小型设备 Linux-通用ARM - V1-发布中 0
纯离线服务	
端云协同服务 へ	毎页显示 10 🗸 < 1 >
我的部署包	
我的本地设备	

Step 2 新增设备并激活

在我的本地设备页面新增设备



		新增设备	×
	设备名称	20个字符以内	
4	备注信息	50个字符以内, 自定义备注信息	
c		0/50	
Ŕ		0/00/2	
		确认 取消	

在列表中,点击设备对应的「激活设备」操作,复制激活命令并在本地设备上执行即可

() () () () () () () () () () () () () (查看激活命令	×		收起 へ
在行	复制命令,并在本地设备上执行,即可激活设备			
端云协同服务	curl -skfL 'https://verify.baidubce.com/active/v1/active/s sh?token=e177df4ed47b2265223a333630302c226b223a226e6f64652;	etup.		
我的本地设备 在本页面新增设 利用率等信息。	6e223a22353730313835302d382d65393866386363302d343231612d65 12d373138392d376266663366643063313433222c226e73223a2262616 6c2d636c6f7564222c227473223a31363039393939393434327d' -osetu && sh setup.sh	37653 57479 p.sh	可以查看设备上部署的服务、	点击收起 设备的运行状态、资源
新增设备				
设备名称 linux x86 ビ	知道了 2021-01-07 14:02 - 未返活		寸间 操作 	插详情

Step 3 下发部署包到设备,在本地调用

在下发部署包到设备页面,将发布成功的部署包一键下发到已激活的设备上,即可测试或正式使用

模型中心	端云协同服务 > 下发部署包到设备
我的模型	将已成功发布的端云协共服务部署包一键下发到已激活且在连接中的设备
创建模型	
训练模型	服务名称
校验模型	模型ID 246
发布模型	应用平台 Linux-通用X86 CPU 🗸
EasyData数据服务	选择版本 V2 🗸
数据总览	28 (2).0 42
标签组管理	JS:f≠ix™ ubuntu-Io ×
在线标注	服务配置 ④基础设置
云服务数据回流	服务端口 8080
EasyEdge本地部署	
纯离线服务	下发服务
端云协同服务 へ	

部署包下发成功之后,会在本地启动一个HTTP推理服务。在浏览器中输入http://{设备ip}:{服务端口,默认8080},即可预览效果:

当前阈值: 0.07 修改

置信度 0.998

0.937

0.841

0.781

0,408

0.324

0.251

▲ 不安全 yq01-ns-map0402a659.yq01:8081



具体接口调用说明请参考文档 SDK - HTTP服务调用说明

云端管理说明

模型部署包管理

在我的部署包页面可以进行已发布的模型部署包的管理。

发布及更新模型版本

点击「发布新版本」操作即可快速发布对应模型ID下的新版本。同一模型ID下已发布的模型版本均会显示在列表的「当前可用版本」中。

	模型中心	端云协同服务 > 我的部署包									
1	我的模型	端云协同服务说明									点击收起
	创建模型	1、在可视化界面轻松实现模型部	署包在边线	缘设备上的集成、 	版本更新						
	训练模型	 2、町网状念下候型高线订算(nt) 3、联网状态下在平台管理设备运 	tp服务, 行状态、	可调用与公有云A 资源利用率	PIJI能怕问的按口)						
-	校验模型	具体使用流程如下:									
	发布模型	 1、在本页面发布端云协同部署包 2、在「我的本地设备」页面新 	,填写相: 增设备, \$	关信息并申请,扌 夏制激活命令,彫	+在下方列表宣看部署包发 (网激活本地设备	定布状态					
	EasyData数据服务	3、在「 下发部署包到设备 」页	面,将发7	^{午成功的部署包一}	-键下发到已激活的设备上	,即可测试或正	式使用				
	数据总宽	发布端云协同部署包									
	标签组管理										
	在线标注	端云协同服务名称	模型ID	设备类型	应用平台	当前可用版本	新版本发布状态	部署设备数	操作		
	云服务数据回流	ecc	264	通用小型设备	Linux-通用ARM	V1	V1-已发布	0	下发到设备发	(布新版本	
	EasyEdge本地部署		246	服务器	Linux-通用X86 CPU	V2, V1	V2-已发布	1	下发到设备发	就在新版本 服	务详情
	纯离线服务	0.000.000.0000	265	通用小型设备	Linux-通用ARM	-	V1-发布中	0			
	端云协同服务 へ							:	每页显示 10	\mathbf{v} <	1 >
	我的部署包										
	我的本地设备										
	下发部署包到设备										
			安寿	亲版本				,			
			X1	7 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	•				Ì		
	将最新训练的模]型版本发布为服务,	发布原	成功后, 艮	1可从云端下发	到设备					
		服务之称 向									
		100000000000000000000000000000000000000									
		模型ID 20	64								
		选择新版本	V1			\sim					
		E	制认	1	取消						
				-							

新版本发布成功后,即可在「下发部署包到设备」页面或当前服务的「服务详情」页面,将新版本下发到本地设备上。

模型中心	端云协同服务 > 我的部署包
我的模型	端云协同服务说明 点击收起
创建模型	1、在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新
训练模型	 断网状态下模型离线计算(http服务,可调用与公有云API功能相同的接口) 联网状态下在平台管理设备运行状态、资源利用率
校验模型	具体使用流程如下:
发布模型	1、在本页面发布端云协同部署包,填写相关信息并申请,并在下方列表查看部署包发布状态
FasyData数据服务	 在 · 其的本地设备 」贝固斯增设备, 复制激活命令, 联网激活本地设备 3、在 「下发部署包到设备 」页面,将发布成功的部署包一键下发到已激活的设备上,即可测试或正式使用
奴姑忌见	发布端云协同部署包
标签组管理	罐云协同服务名称 模型ID 设备类型 应用平台 当前可用版本 新版本发布状态 部署设备数 操作
在线标注	ecc 264 通用小型设备 Linux-通用ARM V1 V1-已发布 0 下发到设备 发布新版本
云服务数据回流	246 服务器 Linux-通用X86 CPU V2, V1 V2-已发布 1 下发到设备 发布新版本 服务详情
EasyEdge本地部署	图像分类高精度_磁狗-265 265 通用小型设备 Linux-通用ARM - VI-发布中 0
纯离线服务	
模型中心	端云协同服务 >>
我的模型	22220
创建模型	施労合称 //被求 // 「
训练模型	
校验模型	下发到更多设备
发布模型	设备名称 最新下发模型版本 部署到期时间 设备连接状态 服务下发状态 最新同步时间 操作
EasvData数据服务	ubuntu-local-fly ⊖ V2 2021-02-06 在线 V2下发成功 2021-01-07 14:45 宣君服务配置 宣君设备详情 下发新版本 移稳设备
教振总览	
xx30175335	
标立坦官理	
在线标注	

管理模型已部署的设备

在上述的「服务详情」页面,可以查看并管理当前服务已部署的设备,包括移除设备、将服务下发到更多的设备等。

模型中心	端云协同服务 > > 服务详情
我的模型	服务名称 模型D 246 当前可用版本 V2, V1 部署设备数 1/1
创建模型	设备类型 服务器 应用平台 Linux-通用X86 CPU
训练模型	THE DIS AND
校验模型	₩¥2424
发布模型	设备名称 最新下发模型版本 部署到期时间 设备连接状态 服务下发状态 最新同步时间 操作
EasyData数据服务	ubuntu-local-fiy 〇 V2 2021-02-06 在线 V2下发成功 2021-01-07 14:45 宣看服务配置 宣看设备详情 下发新版本 移続设备
数据总览	
标签组管理	
在线标注	

本地设备管理

在我的本地设备页面可以进行所有本地设备的管理。

查看单台设备的运行状态

点击单台设备的「服务详情」,可查看设备上运行的多个服务及设备状态:

模型中心	端云协同服务 > 我的本	地设备					
我的模型	我的本地设备						点击收起
创建模型	在本页面新增设备、联网激活 利用率等信息。	5本地设备后,即可将「 🕽	的部署包」『	页面发布成功的部署包一键下发	到设备上。设备联	网时,可以查看设备上部	署的服务、设备的运行状态、资源
训练模型							
校验模型	新増设备						
发布模型	设备名称	创建时间	设备类型	应用平台	设备连接状态	最新同步时间	操作
EasyData数据服务	123123 💬	2021-01-07 13:25	-	-	未激活	-	删除 激活设备 查看详情
数据总觉	showcase-test 💬	2020-12-15 17:57	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 19:22	激活设备 宣看详情
标签组管理	linux-x86-zqw-2 💬	2020-12-15 16:24	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 19:42	激活设备 宣看详情
在线标注	linux-x86-zqw 😳	2020-12-15 15:29	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-15 16:40	激活设备 查看详情
云服务数据回流	firefly_rk3399pro Θ	2020-12-15 14:39	-	-	离线 ⑦	2020-12-15 21:33	激活设备 查看详情
EasyEdge本地部署	ubuntu-local-fjy 💬	2020-12-14 21:38	服务器	Linux-AMD64(x86-64)	在线	2021-01-07 15:01	查看详情
纯离线服务	hfi-1 💬	2020-12-14 19:24	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 14:18	激活设备 查看详情
端云协同服务 ヘ	edge新增 💬	2020-12-11 14:52	-	-	未激活		删除 激活设备 查看详情

设备详情会展示当前设备的最新同步时间,以及CPU使用率、内存使用率等。服务列表则展示了当前设备上部署服务的运行情况和资源占用情况

模型中心	端云协同服务 > 我的本地设备	> ubuntu-local-fjy				
我的模型	设备详情					
创建模型	设备名称 ubuntu-local-fjy	连接状态 在約	Ę	实时刷新	OFF	
训练模型	设备类型 服务器	应用平台 Linu	ux-AMD64(x86-64)	最新同步时间 202	1-01-07 15:00	
校验模型	CPU使用率		内存使用率			
发布模型	31.1%		35.8%			
EasyData数据服务						
数据总览	端云协同服务详情					
标签组管理	服务名称	模型ID	CPU占比	内存使用情况	内存占比	操作
在线标注	2010/02	246	0.01%	156.7MB	0.93%	宣看服务配置
云服务数据回流						
EasyEdge本地部署						
纯离线服务						

心 通用小型设备部署

⊙ 如何在通用小型设备部署

训练完毕后,可以选择将模型通过「SDK-纯离线服务」或「API-端云协同服务」部署,具体介绍如下:

纯离线服务部署

纯离线服务目前仅支持通过SDK集成,可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可发布设备端SDK:

- 选择模型
- 选择部署方式「EasyEdge本地部署」-「通用小型设备」
- 选择版本
- 选择集成方式
- 点击发布

Baidu 百度智能云文档

模型中心	发布模型					
我的模型						设明、
创建模型	选择模型	男女分类	\sim			1 30.50
训练模型	部署方式	EasyEdge本地部署	\checkmark	通用小型设备	\checkmark	- (g) 件环境: 包、说
校验模型	选择版本	V1	~			2. 如SI
发布模型	集成方式	● SDK-纯离线服务				(6795
EasyData数据服务						
数据总览	发布					
在线标注		-				
云服务数据回流						
EasyEdge本地部署						
纯离线服务						

1. 设备镭SDK支持Android、IOS、Windows、Linux操作系统,具体的系统、硬 件环境支持诸参考技术文档。提供可直接体验的移动罐app安装包,以及相应代码 包、说明文档,供企业用户/开发者二次开发

如SDK生成失败,或有任何其他问题,欢迎<mark>提交工单</mark>或加入QQ群 679517246)咨询了解

- 再根据实际使用设备选择系统与芯片
- 点击发布

模型中心	纯离线服务 > 2	 定 有 新服务
我的模型		
创建模型	部署方式	○ 服务器 ● 通用小型设备 ○ 专项适配硬件
of the first of the	许择增利	田 ケ 公米
训练模型	751+19535	<u>лул</u> у V
校验模型	选择版本	V1 ~
发布模型	No. 177 177 (do 170 do 11	
	选择系统和芯片	+ Δ Linux
EasyData数据服务		- K Windows
数据总览		●通用X86 CPU ● 英特尔®神经计算棒
在线标注		🛨 🖷 Android
一眼冬粉绿同冻		+ 🗯 ios
乙胺对氯胺四加		
EasyEdge本地部署	模型加速:	✔ 同时获取加速版 ②
纯离线服务		
A selector		
AI巾助	发布	

也可以直接在「EasyEdge本地部署」-「纯离线服务」页面点击发布新服务,按上图所述进行申请发布

端云协同服务部署

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

具体使用说明请参考端云协同服务说明

心 纯离线SDK说明

^の 纯离线SDK简介

本文档主要说明定制化模型发布后获得的SDK如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

SDK说明

目前已支持Windows、Android、iOS、Linux四种操作系统,更多硬件支持敬请期待。
Baidu 百度智能云文档

操作系统	系统支持	硬件环境要求
Windows	64位 Windows7 及以上	Intel CPU x86_64 环境依赖: .NET Framework 4.5 Visual C++ Redistributable Packages for Visual Studio 2013 Visual C++ Redistributable Packages for Visual Studio 2015
Android	通用ARM: Android 19以上	绝大部分的手机和平板 、比较耗时。支持armeabi-v7a arm-v8a CPU 架 构
iOS	iOS 8.0 以上	ARMv7 ARM64 (Starndard architectures) (暂不支持模拟器)
Linux		aarch64 armv7l

单次预测耗时参考

根据具体设备、线程数不同,数据可能有波动,请以实测为准

在算法性能及适配硬件页面查看评测信息表。

激活&使用SDK

SDK的激活与使用分以下四步:

① 在【我的模型】-【服务详情】内下载SDK

创建模型						
【图像分类】ljay-j	gxn-2flowers	s 模型ID: 340				♂ 训练 ⑤ 历史版本
部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作
私有服务器部署	V1	训练完成	-	-	top1准确率100.00% top5准确率100.00% 完整评估结果	申请发布
【图像分类】ljay-i	gjd-2flowers	:模型ID: 339				⊲ 训练 O 历史版本
部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作
私有服务器部署	V1	训练完成	-	-	top1准确率100.00% top5准确率100.00% 完整评估结果	申请发布
【图像分类】cmc·	高性能 模型	ID: 332				♂ 训练 ④ 历史版本 前 删除
部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作
设备端SDK	V1	训练完成	审核成功	已发布	top1准确率93.33% top5准确率100.00% 完整评估结果	服务详情中请发布
	創建模型 【回像分类】 ijay-f 部署方式 私有限务器部署 【回像分类】 ijay-f 部署方式 【回像分类】 ijay-f 部署方式 【回像分类】 ijay-f 部署方式 【回像分类】 ijay-f 部署方式 【回像分类】 cmc- 部署方式 (資告端SDK	台建模型 「国像分类」」ay-gx-2flower 部署方式 版本 私有服务登部署 V1 「国像分类」」ay-gx-2flower 部署方式 版本 「国像分类」Concetter 版本 「国像分类」Concetter 版本 「国像分类」Concetter 版本 「国像分类」Concetter 版本 「国像分类」Concetter 版本	会議委監 【団像分楽】 jay-JJI-2Flowers JUDE: 340 御商方式 版本 训练状态 私有服务器部書 V1 训练完成 「団像分楽】 jay-JJI-2Flowers JUDE: 349 「個法状态 「団像分楽】 jay-JJI-2Flowers JUDE: 349 「個法状态 「団像分楽] 「山田、 「山岳洗成 「団像分楽] 「山田、 「山岳洗成 「団像分楽] 「山田、 「山岳洗成 「田電方式 版本 训练状态 「山田、 「山岳洗成 「山岳洗成 「山田、 「山岳洗成 「山岳洗成	台建模型 「田像分身」 ay-安二/bower-安川D: 340 御宿方式 版本 训练状态 申请状态 私有服务器部番 V1 训练完成 合. 「田像分身」 ay-安二/bower-安川D: 339 二 二 「田像分身」 ay-安二/bower-安川D: 339 二 一 「田像分身」 ay-安二/bower-安川D: 339 二 一 「田像分身」 ay-安二/bower-安川D: 339 二 一 「田像方式 版本 训练状态 申请状态 「田像分身」 cm-安二/bower-安川D: 339 二 二 「田像方式 版本 训练死点 中请状态 「田像分算」 cm-安二/bower-安川D: 339 二 二 二 「田家方式 版本 训练死点 申请状态 「田像分算」 cm-安二/bower-安川D: 349 二 二 二 「田像分子」 cm-安二/bower-安川D: 359 二 二 二 「田像分子」 cm-安二/bower-set/bo	AIE 4 ALL CIERE-Syal 版本 回路状态 申请状态 風务状态 極有方式 版本 问题状态 中请状态 風分、 紅石服务会部書 V1 问题状态 中请状态 見の、 「四田小女川」 シントン・シントン・ シントン・ シントン・ シントン・ 「昭田小女川」 レントン・ シントン・ シントン・ シントン・ 「田田小女川」 レントン・ シントン・ シントン・ シントン・ シントン・ 「田田小女川」 レントン・ レーン・ シントン・ シントン・ シントン・ シントン・ 「田田小女川」 山田小女川」 山田小女川」 山田小女川」 シントン・ シントン・ シントン・ シントン・ シントン・ シントン・ シントン・ シントン・ シントン・	日日間 少昇】 lay-gu-2flowers USD: 340 単確状态 服务状态 模型效果 部署方式 版本 単体状态 服务状态 模型效果 私有服务器部署 V1 以抵完成 Cop120連書100.00% Cop2進書100.00% SPEi中信点果 配層分支】 lay-gu-2flowers USD: 339 Cop120連書100.00% Cop2進書100.00% SPEi中信点果 配層方式 版本 ULK状态 申请状态 服务状态 種型效果 配層分支】 cop120連書100.00% SPEi中信点果 Cop120連書100.00% SPEi中信点果 配層分支】 cop120連書100.00% SPEi中信点果 . Cop120連書100.00% Cop20進書100.00% SPEi中信点果 配層分支】 cop120連書100.00% SPEi中信点果 配層分支】 cop120連書10.00% SPEi中信点果 配層分支】 cop120連書10.00% SPEi中信点果 配層方式 版本 ULK状态 申请状态 服务状态 . . 配合式 Lat 次本 ULK状态 申請状态 服务状态 . . ULG 満分気

② 在控制台获取序列号

按单台设备获得授权并使用SDK:

序列号管理			
按单台设备激活	按产品线激活		
使用说明 1.按单台设备获得损 ① 下载所需模型已 2.每发布一个模型团 使用过程中有问题可	权并使用SDK(如需开发APP,建议选择按产品线激 发布的离线SDK → ① 设备强部署离线SDK,并联网 问新增2个序列号,联网激活后有效期为3个月,可非 以提交工单或咨询EasyDL官方OQ群(群号:18528	活),操作步骤: 激活 → ① 高线使用 3谙延期 9236)	
+ 新增序列号		新增序列号	×
设备名	激活状态	*新增设备数	请输入新增数量
试用版设备 🧷	未激活		 每个序列号只可在一台设备上激活SDK 每发布一个模型即可新增2个序列号,如需新增更多序列号,可提交工单答
试用版设备 🖉	未激活		询
试用版设备 🖉	未激活		确定取消
试用版设备 🗷	未激活		

③ 本地运行SDK,并完成首次联网激活

④ 正式使用

SDK常见问题

以下是通用FAQ,如您的问题仍未解决,请在百度智能云控制台内提交工单反馈

1、激活失败怎么办?

按设备激活时,激活失败可能由于以下几个原因造成:

①可能是当前序列号已被其他设备激活使用,请核实序列号后用未被激活的序列号重新激活

②序列号填写错误,请核实序列号后重新激活

③同一台设备绑定同一个序列号激活次数过多(超过50次),请更换序列号后重试

④首次激活需要联网,网络环境不佳或无网络环境,请检查网络环境后重试

⑤模型发布者和序列号所属账号非同一账号,如果存在这种异常建议更换账号获取有效序列号

⑥序列号已过有效期,请更换序列号后重试

⑦如有其他异常请在百度智能云控制台内提交工单反馈

2、怎样申请序列号使用延期

序列号激活后有效期为三个月,可以在控制台进行申请,申请流程:

1) 填写申请信息

2) 等待审核:审核周期通常需要1-3个工作日左右,期间会有工作人员电话回访,请填写有效的联系方式并保证手机畅通

の Android集成文档

简介

1.1 Android SDK 硬件要求 Android 版本:支持 Android 5.0 (API 21) 及以上

硬件:支持 arm64-v8a 和 armeabi-v7a,暂不支持模拟器

通常您下载的SDK只支持固定的某一类芯片。

- 通用ARM: 支持大部分ARM 架构的手机、平板及开发板。通常选择这个引擎进行推理。
- 通用ARM GPU:支持骁龙、麒麟、联发科等带GPU的手机、平板及开发板。
- 高端芯片AI加速模块:
 - 高通骁龙引擎SNPE: 高通骁龙高端SOC,利用自带的DSP加速。其中 660 之后的型号可能含有 Hexagon DSP模块,具体列表见snpe 高通骁龙引擎官网。

- 华为NPU引擎DDK:华为麒麟980的arm-v8a的soc。具体手机机型为mate10,mate10pro,P20,mate20,荣耀v20等。
- **华为达芬奇NPU引擎DAVINCI**: 华为NPU的后续版本,华为麒麟810,820,990,985的arm-v8a的soc。具体手机机型为华为 mate30,p40,nova6,荣耀v30等。

通用ARM有额外的加速版,但是有一定的精度损失。

因GPU硬件限制,通用ARM GPU物体检测模型输入尺寸较大时会运行失败,可以在训练的时候将输入尺寸设为300*300。

高端芯片AI加速模块,一般情况下推理速度较快。

运行内存不能过小,一般大于demo的assets目录大小的3倍。

1.2 功能支持 | 引擎 | 图像分类| 物体检测| 图像分割| 文字识别

只支持EasyEdge | 姿态估计 | | :-: | :-: | :-: | :-: | :-: | :=: | 通用ARM | √ | √ | √ | √ | ↓ | 通用ARM GPU | √ | √ | √ | ↓ | 高通骁龙引擎SNPE | √ | √ |||| | 华为达芬奇NPU引擎DDK | √ | √ |||| | 华为达芬奇NPU引擎DVINCI | √ | √ | ↓ |||

1.3 Release Notes

时间	版本	说明
2023.08.31	0.10.12	新增支持实例数鉴权;SNPE引擎升级;迭代优化
2023.06.29	0.10.11	迭代优化
2023.05.17	0.10.10	横屏兼容;迭代优化
2023.03.16	0.10.9	达芬奇NPU支持更多模型及语义分割模型;各芯片支持更多语义分割模型;精简版代码补充;迭代优化
2022.12.29	0.10.8	ARM / ARM-GPU 引擎升级;迭代优化
2022.10.27	0.10.7	达芬奇NPU新增适配麒麟985;迭代优化
2022.09.15	0.10.6	SNPE引擎升级;迭代优化
2022.07.28	0.10.5	迭代优化
2022.06.30	0.10.4	支持Android11;支持EasyEdge语义分割模型;迭代优化
2022.05.18	0.10.3	ARM / ARM-GPU 引擎升级;支持更多加速版模型发布;迭代优化
2022.03.25	0.10.2	ARM / ARM-GPU 引擎升级;支持更多检测模型;迭代优化
2021.12.22	0.10.1	DDK不再支持Kirin 970;迭代优化
2021.10.20	0.10.0	更新鉴权;更新达芬奇NPU、SNPE、通用ARM及ARM-GPU引擎;新增达芬奇NPU对检测模型的支持;支持更多姿态估计 模型
2021.07.29	0.9.17	迭代优化
2021.06.29	0.9.16	迭代优化
2021.05.13	0.9.15	更新鉴权,更新通用arm及通用arm gpu引擎
2021.04.02	0.9.14	修正bug
2021.03.09	0.9.13	更新android arm的预处理加速
2020.12.18	0.9.12	通用ARM引擎升级;新增ARM GPU引擎
2020.10.29	0.9.10	迭代优化
2020.9.01	0.9.9	迭代优化
2020.8.11	0.9.8	更新ddk 达芬奇引擎
2020.7.14	0.9.7	支持arm版ocr模型,模型加载优化
2020.6.23	0.9.6	支持arm版fasterrcnn模型
2020.5.14	0.9.5	新增华为新的达芬奇架构npu的部分图像分类模型
2020.4.17	0.9.4	新增arm通用引擎量化模型支持
2020.1.17	0.9.3	新增arm通用引擎图像分割模型支持
2019.12.26	0.9.2	新增华为kirin麒麟芯片的物体检测支持
2019.12.04	0.9.1	使用paddleLite作为arm预测引擎
2019.08.30	0.9.0	支持EasyDL专业版
2019.08.30	0.8.2	支持华为麒麟980的物体检测模型
2019.08.29	0.8.1	修复相机在开发版调用奔溃的问题
2019.06.20	0.8.0	高通手机引擎优化
2019.05.24	0.7.0	升级引擎
2019.05.14	0.6.0	优化demo程序
2019.04.12	0.5.0	新增华为麒麟980支持
2019.03.29	0.4.0	引擎优化,支持sd卡模型读取
2019.02.28	0.3.0	引擎优化,性能与效果提升;
2018.11.30	0.2.0	第一版!

快速开始

2.1 安装软件及硬件准备

如果需要源码方式测试:

打开AndroidStudio,点击 "Import Project..."。在一台较新的手机上测试。

详细步骤如下:

1. 准备一台较新的手机,如果不是通用arm版本,请参见本文的"硬件要求",确认是否符合SDK的要求

2. 安装较新版本的AndroidStudio , 下载地址

- 3. 新建一个HelloWorld项目, Android Studio会自动下载依赖, 在这台较新的手机上测试通过这个helloworld项目。注意不支持模拟器。
- 4. 解压下载的SDK。
- 5. 打开AndroidStudio, 点击 "Import Project..."。即: File->New-> "Import Project...", 选择解压后的目录。
- 6. 此时点击运行按钮(同第3步),手机上会有新app安装完毕,运行效果和二维码扫描的一样。
- 7. 手机上UI界面显示后,如果点击UI界面上的"开始使用"按钮,可能会报序列号错误。请参见下文修改

2.2 使用序列号激活

如果使用的是EasyEdge的开源模型,无需序列号,可以跳过本段直接测试。

建议申请包名为"com.baidu.ai.easyaimobile.demo"的序列号用于测试。

本文假设已经获取到序列号,并且这个序列号已经绑定包名。

SDK默认使用离线激活方式,即首次联网激活,后续离线使用。SDK同时支持按实例数鉴权方式,即周期性联网激活,离线后会释放所占用设备 实例。按实例数鉴权的启用参考本节2.2.3说明

2.2.1 填写序列号 打开Android Studio的项目,修改MainActivity类的开头SERIAL_NUM字段。 MainActivity 位于 app \src \main \java \com \baidu \ai \edge \demo \MainActivity.java文件内。

// 请替换为您的序列号 private static final String SERIAL_NUM = "XXXX-XXXX-XXXX-XXXX"; //这里填您的序列号

2.2.2 修改包名

如果申请的包名为"com.baidu.ai.easyaimobile.demo",这个是demo的包名,可以不用修改

打开app/build.gradle文件,修改"com.baidu.ai.easyaimobile.demo"为申请的包名

```
defaultConfig {
    applicationId "com.baidu.ai.easyaimobile.demo" // 修改为比如"com.xxx.xxx"
}
```

修改序列号和包名后,可以运行测试,效果同扫描二维码的一致

2.2.3 按实例数鉴权 设置好序列号和包名后,调用配置类的以下方法启用并配置心跳间隔时间:

```
XXXConfig config = new XXXConfig();
// 启用按实例数鉴权,配置心跳间隔,单位:秒
config.setInstanceAuthMode(10000);
```

配置类的详细说明参考后续章节【调用流程】

2.3 测试精简版

对于通用ARM、高通骁龙引擎SNPE、华为NPU引擎DDK和达芬奇NPU引擎Davinci的常见功能,项目内自带精简版,可以忽略开发板不兼容的摄像 头。

此外,由于实时摄像开启,会导致接口的耗时变大,此时也可以使用精简版测试。

目前以下硬件环境有精简版测试:

Baidu 百度智能云文档

- 通用ARM:图像分类(Classify),物体检测(Detection),文字识别(OCR),图像分割(Segmentation),姿态估计(Pose)
- 通用ARM GPU:图像分类(Classify),物体检测(Detection),图像分割(Segmentation),姿态估计(Pose)
- 高通骁龙引擎SNPE:图像分类(Classify),物体检测(Detection)
- 华为NPU引擎DDK:图像分类(Classify),物体检测(Detection)
- 华为达芬奇NPU引擎Davinci:图像分类(Classify),物体检测(Detection),图像分割(Segmentation)

具体代码分别在infertest、snpetest、ddktest和davincitest目录下。

修改方法为(以通用ARM为例):更改app/main/AndroidManifest.xml中的启动Activity。

```
<activity android:name=".infertest.MainActivity"> <!-- 原始的是".MainActivity" -->
<intent-filter>
<action android:name="android.intent.action.MAIN" />
<category android:name="android.intent.category.LAUNCHER" />
</intent-filter>
```

</activity>

开启后会自动选择图像分类(Classify),物体检测(Detection),文字识别(OCR),图像分割(Segmentation)或姿态估计(Pose)测试。

Demo APP 检测模型运行示例 精简版检测模型运行示例 Hello World! paddle2-Centernet_mbv3_small_140e < ARM Detection 4.18 person 0 Start running: 0 horse 0.40 Predict 0: {size:100, firstRe confidence:0.6314938, bo truck 0.34 181)}} per86/\$9/8 Finish running Task finished cow 0.34 识别结果 置信度 0.30 名称 置信度 序号 0.63 1 person 2 0.47 person 3 0.42 car 0.40 4 horse 5 dog 0.34 BU' 6 truck 0.34 4 ►

使用说明

3.1 代码目录结构

集成时需要"复制到自己的项目里"的目录或者文件:

2. app/src/main/assets/xxxx-xxxxx 如 app/src/main/assets/infer

+app 简单的设置,模拟用户的项目 |---+libs 实际使用时需要复制到自己的项目里 |----arm64-v8a v8a的so |----armeabi-v7a v7a的so |----easyedge-sdk.jar jar库文件 |---+src/main |---+assets |----demo demo项目的配置,实际集成不需要 |----infer 也可能是其它命名, infer表示通用arm。实际使用时可以复制到自己的项目里 |---+java/com.baidu.ai.edge/demo |---+infertest 通用Arm精简版测试,里面有SDK的集成逻辑 |--- MainActivity 通用Arm精简版启动Activity,会根据assets目录判断当前的模型类型,并运行同目录的一个Task。使用时需要修改里 面的序列号 |--- TestInferArmClassifyTask 通用Arm精简版分类 |--- TestInferArmDetectionTask 通用Arm精简版检测 |--- TestInferArmOcrTask 通用Arm精简版OCR |--- TestInferArmPoseTask 通用Arm精简版姿态 |--- TestInferArmSegmentTask 通用Arm精简版分割 |--- TestInferArmGpuClassifyTask 通用ArmGpu精简版分类 |--- TestInferArmGpuDetectionTask 通用ArmGpu精简版检测 |--- TestInferArmGpuPoseTask 通用ArmGpu精简版姿态 |--- TestInferArmGpuSegmentTask 通用ArmGpu精简版分割 |---+snpetest SNPE精简版测试 |--- MainActivity SNPE精简版启动Activity,会根据assets目录判断当前的模型类型,并运行同目录的一个Task。使用时需要修改里面 的序列号 |--- TestSnpeDspClassifyTask SNPE DSP精简版分类 |--- TestSnpeDspDetectionTask SNPE DSP精简版检测 |--- TestSnpeGpuClassifyTask SNPE Gpu精简版分类 |--- TestSnpeGpuDetectionTask SNPE Gpu精简版检测 |---+ddktest DDK精简版测试 |--- MainActivity DDK精简版启动Activity,会根据assets目录判断当前的模型类型,并运行同目录的一个Task。使用时需要修改里面的 序列号 |--- TestDDKClassifyTask DDK精简版分类 |--- TestDDKDetectionTask DDK精简版检测 |---+davincitest Davinci精简版测试 |--- MainActivity Davinci精简版启动Activity,会根据assets目录判断当前的模型类型,并运行同目录的一个Task。使用时需要修改里面 的序列号 |--- TestDavinciClassifyTask Davinci精简版分类 |--- TestDavinciDetectionTask Davinci精简版检测 |--- TestDavinciSegmentTask Davinci精简版分割 |----CameraActivity 摄像头扫描示例,里面有SDK的集成逻辑 |----MainActivity 启动Activity,使用时需要修改里面的序列号 |---- build.gradle 这里修改包名 +camera_ui UI模块,集成时可以忽略 • 1

3.2 调用流程 以通用ARM的检测模型功能为例,

代码可以参考TestInferDetectionTask

- 1. 准备配置类,如InferConfig,输入:通常为一个assets目录下的文件夹,如infer。
- 2. 初始化Manager,比如InferManager。输入:第1步的配置类和序列号
- 3. 推理图片,可以多次调用 3.1 准备图片,作为Bitmap输入 3.2 调用对应的推理方法,比如detect 3.3 解析结果,结果通常是一个List,调用结 果类的Get方法,通常能获取想要的结果
- 4. 直到长时间不再使用我们的SDK,调用Manger的destroy方法释放资源。

3.3 具体接口说明下文的示例部分以通用ARM的检测模型功能为例

即接口为InferConfig, InferManager, InferManager.detect。

其它引擎和模型调用方法类似。

下文假设已有序列号及对应的包名

- 3.3.1. 准备配置类
- INFER:通用ARM, InferConfig
- ARM GPU : ArmGpuConfig
- SNPE:高通骁龙DSP, SnpeConfig
- SNPE GPU:高通骁龙GPU,SnpeGpuConfig
- DDK:华为NPU, DDKConfig
- DDKDAVINCI:华为达芬奇NPU,DDKDaVinciConfig

InferConfig mInferConfig = new InferConfig(getAssets(), "infer"); // assets 目录下的infer , infer表示通用arm

输入:assets下的配置 输出:具体的配置类

3.3.2. 初始化Manager类

- INFER:通用ARM, InferManager
- ARM GPU:通用ARM GPU, InferManager
- SNPE:高通骁龙DSP, SnpeManager
- SNPE GPU:高通骁龙GPU, SnpeManager
- DDK:华为NPU,DDKManager
- DDKDAVINCI:华为达芬奇NPU, DavinciManager

String SERIAL_NUM = "XXXX-XXXX-XXXX-XXXX";

```
// InferManager 为例:
new InferManager(this, config, SERIAL_NUM)); // config为上一步的InferConfig
```

注意要点

- 1. 同一个时刻只能有唯一有效的InferManager。旧的InferManager必须调用destory后,才能新建一个new InferManager()。
- 2. InferManager的任何方法,都不能在UI线程中调用。
- 3. new InferManager()及InferManager成员方法由于线程同步数据可见性问题,都**必须在一个线程中执行**。如使用android自带的**ThreadHandler** 类。

输入: 1.配置类; 2.序列号 输出: Manager类

3.3.3.推理图片

- 接口可以多次调用,但是必须在一个线程里,**不能并发**
- confidence, 置信度[0-1], 小于confidence的结果不返回。 填confidence=0, 返回所有结果
- confidence可以不填,默认用模型推荐的。

准备图片,作为Bitmap输入,

• 输入为Bitmap,其中Bitmap的options为默认。如果强制指定的话,必须使用*Bitmap.Config.ARGB_8888*

调用对应的推理方法及结果解析 见下文的各个模型方法

3.3.4 分类Classify

public interface ClassifyInterface { List<ClassificationResultModel> classify(Bitmap bitmap, float confidence) throws BaseException;

// 如InferManger 继承 ClassifyInterface

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 ClassificationResultModel 异常:一般首次出现。可以打印出异常错误码。

ClassificationResultModel

- label:分类标签,定义在label_list.txt中

- confidence:置信度,0-1

- lableIndex:标签对应的序号

3.3.5 检测Detect

对于EasyDL口罩检测模型请注意输入图片中人脸大小建议保持在88到9696像素,可根据场景远近程度缩放图片后传入

public interface DetectInterface { List<DetectionResultModel> detect(Bitmap bitmap, float confidence) throws BaseException;

// 如InferManger 继承 DetectInterface

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 DetectionResultModel List 异常:一般首次出现。可以打印出异常错误码。

DetectionResultModel

- label:标签,定义在label_list.txt中
- confidence:置信度
- bounds:Rect,左上角和右下角坐标

3.3.6 图像分割Segmentation

```
public interface SegmentInterface {
List<SegmentationResultModel> segment(Bitmap bitmap, float confidence) throws BaseException;
// 如InferManger 继承 SegmentInterface
```

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 SegmentationResultModel 异常:一般首次出现。可以打印出异常错误码。

SegmentationResultModel

label:标签,定义在label_list.txt中
confidence:置信度
lableIndex:标签对应的序号
box: Rect对象表示的对象框
mask:byte[]表示的原图大小的0,1掩码,绘制1的像素即可得到当前对象区域
mask 字段说明,如何绘制掩码也可参考demo工程
 101
image 110 => mask(byte[]) 101 110 011

011

3.3.7 文字识别OCR

暂时只支持通用ARM引擎,不支持其它引擎,暂时只支持EasyEdge的开源OCR模型。

```
public interface OcrInterface {
List<OcrResultModel> ocr(Bitmap bitmap, float confidence) throws BaseException;
// 如InferManger 继承 OcrInterface
```

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 OcrResultModel List,每个OcrResultModel对应结果里的一个四边形。 异常:一般首次出现。可以打印出异常错误码。

OcrResultModel

- label:识别出的文字
- confidence:置信度
- List<Point>:4个点构成四边形

3.3.8 姿态估计Pose

暂时只支持通用ARM引擎,不支持其它引擎

public interface PoseInterface { List<PoseResultModel> pose(Bitmap bitmap) throws BaseException; // 如InferManger 继承 PoseInterface

输入 Bitmap 默认格式或者指定Bitmap.Config.ARGB_8888 输出 PoseResultModel List 异常:一般首次出现。可以打印出异常错误码。

PoseResultModel

- label : 标签,定义在label_list.txt中

- confidence:置信度

- Pair<Point, Point>: 2个点构成一条线

3.3.9 释放

释放后这个对象不能再使用,如果需要使用可以重新new一个出来。

public void destory() throws BaseException

3.3.10 整体示例

以通用ARM的图像分类预测流程为例:

```
try {
// step 1: 准备配置类
InferConfig config = new InferConfig(context.getAssets(), "infer");
// step 2: 准备预测 Manager
InferManager manager = new InferManager(context, config, "");
// step 3: 准备待预测的图像,必须为 Bitmap.Config.ARGB_8888 格式,一般为默认格式
Bitmap image = getFromSomeWhere();
// step 4: 预测图像
List<ClassificationResultModel> results = manager.classify(image, 0.3f);
// step 5: 解析结果
for (ClassificationResultModel resultModel : results) {
 Log.i(TAG, "labelIndex=" + resultModel.getLabelIndex()
  + ", labelName=" + resultModel.getLabel()
  + ", confidence=" + resultModel.getConfidence());
}
// step 6: 释放资源。预测完毕请及时释放资源
manager.destroy();
} catch (Exception e) {
Log.e(TAG, e.getMessage());
}
```

```
"autocheck_qcom": true, // 如果改成false, sdk跳过检查手机是否是高通的Soc,非高通的Soc会奔溃直接导致app闪退
"snpe_runtimes_order": [],
// 不填写为自动,按照 {DSP, GPU, GPU_FLOAT16, CPU}次序尝试初始化,也可以手动指定如[2,1,3,0],具体数字的定义见下段

public interface SnpeRuntimeInterface {
    int CPU = 0;
    int GPU = 1;
    int DSP = 2;
    int GPU_FLOAT16 = 3;
```

// SnpeManager 中,使用public static ArrayList<Integer> getAvailableRuntimes(Context context) 方法可以获取高通SOC支持的运行方式

集成指南

}

- 1. 复制库文件libs
- 2. 添加Manifest权限
- 3. 复制模型文件
- 4. 添加调用代码(见上一步具体接口说明)
- 4.1 复制库文件libs A. 如果项目里没有自己的jar文件和so文件:

```
复制app/libs 至自己项目的app/libs目录。
参照demo的app/build.gradle 中添加
android {
.....
defaultConfig {
ndk {
abiFilters 'armeabi-v7a', 'arm64-v8a'
}
}
sourceSets {
main {
jniLibs.srcDirs = ['libs']
}
}
```

B. 如果项目里有自己的jar文件,但没有so文件

```
easyedge.jar文件同自己的jar文件放一起
arm64-v8a和armeabi-v7a放到app/src/main/jniLibs目录下
参照demo的app/build.gradle 中添加
android {
....
defaultConfig {
ndk {
abiFilters 'armeabi-v7a', 'arm64-v8a'
}
}
```

C. 如果项目里有自己的jar文件和so文件

```
easyedge.jar文件同自己的jar文件放一起
arm64-v8a和armeabi-v7a取交集和自己的so放一起,交集的意思是比如自己的项目里有x86目录,必须删除x86。
参照demo的app/build.gradle 中添加
android {
....
defaultConfig {
ndk {
abiFilters 'armeabi-v7a', 'arm64-v8a' // abiFilter取交集,即只能少不能多
}
}
```

jar文件库如果没有设置成功的,编译的时候可以发现报错。

so库如果没有编译进去的话,也可以通过解压apk文件确认。运行的时候会有类似jni方法找不到的报错。

4.2 Manifest配置

参考app/src/main/AndroidManifest.xml文件,添加:

```
    <uses-permission android:name="android.permission.INTERNET" />
    <uses-permission android:name="android.permission.WRITE_EXTERNAL_STORAGE" />
    <uses-permission android:name="android.permission.ACCESS_NETWORK_STATE" />
    <uses-permission android:name="android.permission.READ_PHONE_STATE" />
    <l-- Android 11 支持 -->
    <uses-permission</li>
    android:name="android.permission.MANAGE_EXTERNAL_STORAGE"
    <l-- 高版本 Android 支持 -->
    <application</a>
```

```
android:requestLegacyExternalStorage="true"
android:usesCleartextTraffic="true">
</application>
```

4.3 混淆规则 (可选) 请不要混淆SDK里的jar文件。

```
-keep class com.baidu.ai.edge.core.*.*{ *; }
```

4.4 Android 11支持 除Manifest中必要配置外,请参考BaseActivity获取所有文件访问权限,否则可能影响SDK正常使用。

SDK 默认使用 easyedge-sdk.jar,未启用 AndroidX,若您的项目使用 AndroidX,并在集成中提示 android.support 相关错误,请参考 app/build.gradle 使用 etc/easyedge-sdk-androidx.jar 以支持 AndroidX:

```
// app/build.gradle
dependencies {
    implementation project(':camera_ui')
    implementation files('libs/easyedge-sdk-androidx.jar') // 修改 jar 包依赖
}
```

错误码 | 错误码 | 错误描述 | 详细描述及解决方法 | | --- | --- | 1001 | assets 目录下用户指定的配置文件不存在或不正确 | SDK使用assets目录下一系列文件作为配置文件。如果文件缺失或内容不正确,则有此报错 || 1002 | json格式的配置文件解析出错 | 如缺少某些字段。正常情况下,配置文件请不要修改 || 1003 | 应用缺少权限 | 请根据提示动态申请缺少的权限 || 19xx | Sdk内部错误 | 请与百度人员联系 || 2001 | XxxxMANAGER 只允许一个实例 | 如已有XxxxMANAGER对象,请调用destory方法 || 2002 | XxxxMANAGER 已经调用过destory方法 | 在一个已经调用destory方法的DETECT_MANAGER对象上,不允许再调用任何方法 || 2003 | 传入的assets下模型文件路径非法 |比如缺少模型文件, XxxxConfig.getModelFileAssetPath() 返回为null || 2012 | JNI内存错误 | heap的内存不够 || 2103 | license过期 | license失效或者系统时间有异常 || 2601/2602 | assets 目录下模型文件打开/读取失败 | 请根据报错信息检查模型文件是否存在 || 27xx | Sdk内部错误 | 请与百度人员联系 || 4011 | SDK类型 | 28xx | 引擎内部错误 | 请与百度人员联系 || 29xx | Sdk内部错误 | 请与百度人员联系 || 3003 | 减型加载错误 | 请确认模型放置于能被加载到的合法路径中,并确保配置文件正确 || 3002 | 模型卸载错误 | 请与百度人员联系 || 4011 | SDK类型 与设备硬件不匹配 | 比如适配DSP的SDK运行在麒麟芯片上会出现此报错,请在部署包支持的硬件上使用SDK || 50xx | SDK调用异常 | 请与百度 人员联系 |

报错日志收集 通常 Logcat 可以看见日志及崩溃信息,若设备无法获取日志信息,可使用 Demo 中的 xCrash 工具:

```
// 1. 引入 app/build.gradle 的 xCrash 依赖
android {
  ...
  dependencies {
     implementation 'com.iqiyi.xcrash:xcrash-android-lib:2.4.5' // 可以保存崩溃信息,默认未引入
  }
}
// 2. 启用日志收集。日志将保存在 /sdcard/<包名>/xCrash
// app/src/main/java/com.baidu.ai.edge/demo/MyApplication.java
protected void attachBaseContext(Context context) {
  // 日志保存位置
  String basePath = Environment.getExternalStorageDirectory().toString() + "/" + context.getPackageName();
  // 启用
  XCrash.InitParameters params = new XCrash.InitParameters();
  params.setAppVersion(BaseManager.VERSION);
  params.setLogDir(basePath + "/xCrash");
  XCrash.init(this, params);
}
```

心 iOS集成文档

简介

本文档描述 EasyEdge/EasyDL iOS 离线预测SDK相关功能;

目前支持EasyEdge的功能包括:

- 图像分类
- 物体检测
- 人脸检测
- 姿态估计
- 百度OCR模型

目前支持EasyDL的功能包括:

- 图像分类
- 物体检测
- 图像分割

系统支持

系统:

- 通用arm版本:iOS 9.0 以上
- A仿生芯片版: iOS 15.0 及以上
- 硬件: arm64 (Starndard architectures) (暂不支持模拟器)

内存:图像分割模型需要手机内存3GB以上,并尽量减少其他程序内存占用

离线SDK包说明

根据用户的选择,下载的离线SDK,可能包括以下类型:

- EasyEdge
 - 通用ARM版:支持iPhone5s, iOS 9.0 以上所有手机。
 - A仿生芯片版:支持iPhone5s, iOS 15.0 以上手机。充分利用苹果A系列仿生芯片优势,在iPhone 8以上机型中能有显著的速度提升。

- EasyDL 通用版/全功能AI开发平台BML (原EasyDL专业版)
 - 通用ARM版:支持iPhone5s, iOS 9.0 以上所有手机。
 - A仿生芯片版:支持iPhone5s, iOS 15.0 以上手机。充分利用苹果A系列仿生芯片优势,在iPhone 8以上机型中能有显著的速度提升。
 - 自适应芯片版:同时整合了以上两种版本,自动在iOS 15以下中使用通用ARM版,在iOS 15以上系统中使用A仿生芯片版,自适应系统,但SDK体积相对较大。
- AI市场试用版SDK

SDK大小说明

SDK库的二进制与_TEXT增量约3M。

资源文件大小根据模型不同可能有所差异。

物体检测(高性能)的DemoApp在iPhone 6, iOS 11.4下占用空间实测小于40M。

虽然SDK库文件很大(体现为SDK包文件很大,ipa文件很大),但最终应用在用户设备中所占用的大小会缩小很多。 这与multi architechtures、bitcode和AppStore的优化有关。

获取序列号 生成SDK后,点击获取序列号进入控制台获取。EasyEdge控制台、EasyDL控制台、BML控制台。

试用版SDK在SDK的RES文件夹中的SN.txt中包含试用序列号。

更换序列号、更换设备时,首次使用需要联网激活。激活成功之后,有效期内可离线使用。

Release Notes

时间	版本	说明
2023.08.31	0.7.13	新增按实例数鉴权;迭代优化
2023.06.29	0.7.12	迭代优化
2023.05.17	0.7.11	CoreML引擎升级,支持更多语义分割模型;兼容横屏;迭代优化
2023.03.16	0.7.10	支持更多语义分割模型;迭代优化
2022.12.29	0.7.9	ARM引擎升级;迭代优化
2022.10.27	0.7.8	支持更多检测模型;迭代优化
2022.09.15	0.7.7	支持更多检测模型;迭代优化
2022.07.28	0.7.6	迭代优化
2022.06.29	0.7.5	支持EasyEdge语义分割模型;CoreML引擎升级,新增EasyEdge检测模型支持;迭代优化
2022.05.18	0.7.4	ARM引擎升级;支持EasyDL物体检测超高精度模型;支持更多加速版模型发布;迭代优化
2022.03.25	0.7.3	ARM引擎升级;支持更多检测模型
2021.12.22	0.7.2	支持EasyEdge更多姿态估计模型;迭代优化
2021.10.20	0.7.1	ARM引擎升级
2021.07.29	0.7.0	迭代优化
2021.04.06	0.6.1	ARM引擎升级
2021.03.09	0.6.0	支持EasyEdge人脸检测及姿态估计模型
2020.12.18	0.5.7	ARM引擎升级
2020.09.17	0.5.6	CoreML引擎升级,支持AI市场试用版SDK
2020.08.11	0.5.5	CoreML支持EasyDL专业版模型,支持EasyEdge OCR模型
2020.06.23	0.5.4	ARM引擎升级
2020.04.16	0.5.3	ARM引擎升级;支持压缩加速版模型
2020.03.13	0.5.2	ARM引擎升级;支持图像分割模型
2020.01.16	0.5.1	ARM引擎升级;增加推荐阈值支持
2019.12.04	0.5.0	ARM引擎升级;增加coreml3的支持
2019.10.24	0.4.5	支持EasyDL专业版;ARM引擎升级
2019.08.30	0.4.4	支持EasyDL经典版图像分类高性能、高精度
2019.06.20	0.4.3	引擎优化
2019.04.12	0.4.1	支持EasyDL经典版物体检测高精度、高性能模型
2019.03.29	0.4.0	引擎优化,支持CoreML;
2019.02.28	0.3.0	引擎优化,性能与效果提升;
2018.11.30	0.2.0	第一版!

快速开始 文件结构说明

.EasyEdge-iOS-SDK ├── EasyDLDemo # Demo工程文件 └── LIB # 依赖库 ├── RES │ └── easyedge # 模型资源文件夹 │ │ └── model │ │ └── params │ │ └── label_list.txt │ │ └── conf.json └── DOC # 文档

测试Demo

按如下步骤可直接运行 SDK 体验 Demo:

步骤一:用 Xcode 打开 EasyDLDemo/EasyDLDemo.xcodeproj 步骤二:配置开发者自己的签名

步骤三:连接手机运行,不支持模拟器

检测模型运行示例:







SDK使用说明 集成指南 步骤一:依赖库集成 步骤二:import <EasyDL/EasyDL.h>, import <Vision/Vision.h>

依赖库集成

- 1. 复制 LIB 目录至项目合适的位置
- 2. 配置 Build Settings 中 Search paths: 以 SDK 中 LIB 目录路径为例
- Framework Search Paths : \${PROJECT_DIR}/../LIB/lib
- Header Search Paths : \${PROJECT_DIR}/../LIB/include
- Library Search Paths : \${PROJECT_DIR}/../LIB/lib

集成过程如出现错误,请参考 Demo 工程对依赖库的引用

使用流程

1. 生成模型,下载SDK 开发者在官网下载的SDK已经自动为开发者配置了模型文件和相关配置,开发者直接运行即可。

2. 使用序列号激活 2.1. 离线激活 (默认鉴权方式) 首次联网激活,后续离线使用

将前面申请的序列号填入:

[EasyDL setSerialNumber:@"!!!Enter Your Serial Number Here!!!"];

根据序列号类型,序列号与BundleID绑定或与BundleID+设备绑定。 请确保设备时间正确。

2.2. 按实例数激活 周期性联网激活,离线后会释放所占用鉴权,启动时请确保心跳间隔小于等于生成序列号时填写的定期确认时间

填入序列号,配置按实例数鉴权并设置心跳间隔:

```
// 设置序列号
```

[EasyDL setSerialNumber:@"!!!Enter Your Serial Number Here!!!"]; // 配置实例数鉴权及心跳间隔,单位:秒 [EasyDL setInstanceAuthMode:10000];

3. 初始化模型

EasyDLModel *_model = [[EasyDLModel alloc] initModelFromResourceDirectory:@"easyedge" withError:&err];

请注意相关资源必须以folder reference方式加入Xcode工程。也即默认的easyedge文件夹在Xcode文件列表里显示为蓝色。

4. 调用检测接口

UIImage *img =; NSArray *result = [model detectUIImage:img withFilterScore:0 andError:&err];

/**

```
* 检测图像
```

```
* @param image 带检测图像
```

- * @param score 只返回得分高于score的结果(0~1)
- * @return 成功返回识别结果,NSArray的元素为对应模型的结果类型;失败返回nil,并在err中说明错误原因
- */

```
- (NSArray *)detectUIImage:(UIImage *)image
```

```
withFilterScore:(CGFloat)score
```

andError:(NSError **)err;

返回的数组类型如下,具体可参考 EasyDLResultData.h 中的定义:

模型类型	类型
图像-图像分类	EasyDLClassfiData
图像-物体检测/人脸检测	EasyDLObjectDetectionData
图像-实例分割/语义分割	EasyDLObjSegmentationData
图像-姿态估计	EasyDLPoseData
图像-文字识别	EasyDLOcrData

错误说明

SDK的方法会返回NSError错,直接返回的NSError的错误码定义在EEasyDLErrorCode中。NSError附带message(有时候会附带 NSUnderlyingError),开发者可根据code和message进行错误判断和处理。

FAQ

1. 如何多线程并发预测?

SDK内部已经能充分利用多核的计算能力。不建议使用并发来预测。

如果开发者想并发使用,请务必注意EasyDLModel所有的方法都不是线程安全的。请初始化多个实例进行并发使用,如

```
- (void)testMultiThread {
 UIImage *img = [UIImage imageNamed:@"1.jpeg"];
 NSError *err;
 EasyDLModel * model1 = [[EasyDLModel alloc] initModelFromResourceDirectory:@"easyedge" withError:&err];
 EasyDLModel * model2 = [[EasyDLModel alloc] initModelFromResourceDirectory:@"easyedge" withError:&err];
 dispatch_queue_t queue1 = dispatch_queue_create("testQueue", DISPATCH_QUEUE_CONCURRENT);
 dispatch_queue_t queue2 = dispatch_queue_create("testQueue2", DISPATCH_QUEUE_CONCURRENT);
 dispatch_async(queue1, ^{
   NSError *detectErr;
   for(int i = 0; i < 1000; ++i) {
      NSArray * res = [model1 detectUlImage:img withFilterScore:0 andError:&detectErr];
      NSLog(@"1: %@", res[0]);
 dispatch_async(queue2, ^{
   NSError *detectErr;
    for(int i = 0; i < 1000; ++i) {
      NSArray * res = [model2 detectUlImage:img withFilterScore:0 andError:&detectErr];
      NSLog(@"2: %@", res[0]);
```

2. 编译时出现 Undefined symbols for architecture arm64: ...

- 出现 cxx11, vtable 字样 : 请引入 libc++.tbd
- 出现 cv::Mat 字样:请引入 opencv2.framework
- 出现 CoreML, VNRequest 字样:请引入CoreML.framework 并务必#import <CoreML/CoreML.h>
- 3. 运行时报错 Image not found: xxx ...

请Embed具体报错的库。 4.编译时报错:Invalid bitcode version 这个可能是开发者使用的xcode低于12导致,可以升级至12版本。

の Windows集成文档

简介

本文档介绍图像分割服务器端Windows SDK的使用方法。

- 硬件支持:
 - Intel CPU 普通版 * x86_64

- CPU 加速版 Intel Xeon with AVX2 and AVX512 Intel Core Processors with AVX2 Intel Atom Processors with SSE * AMD Core Processors with AVX2
- Intel Movidius MyRIAD2/MyRIAD X (仅支持Win10)
- 操作系统支持
 - 普通版:64位 Windows 7 及以上,64位Windows Server2012及以上
 - 加速版:64位 Windows 10,64位Windows Server 2019及以上
- 环境依赖 (必须安装以下版本)
 - .NET Framework 4.5
 - Visual C++ Redistributable Packages for Visual Studio 2013
 - Visual C++ Redistributable Packages for Visual Studio 2015-2019
- 协议
 - HTTP
- 更详细的环境说明可参考SDK内的README.md

*intel 官方合作,拥有更好的适配与性能表现

Release Notes | 时间 | 版本 | 说明 | | ------ | ----- | 2023-08-30 | 1.8.3 | 新增支持按实例数鉴权 || 2023-06-29 | 1.8.2 | 优化模型 算法 || 2023-05-17 | 1.8.1 | 预测引擎升级, 修复部分网络内存泄露问题 || 2023-03-16 | 1.8.0 | 预测引擎升级 || 2022-12-29 | 1.7.2 | 预测引 擎升级 || 2022-10-27 | 1.7.1 | 优化模型算法 || 2022-09-15 | 1.7.0 | 新增支持表格预测 || 2022-07-28 | 1.6.0 | 优化模型算法 || 2022-05-27 | 1.5.1 | 新增支持BML Cloud小目标检测模型 || 2022-05-18 | 1.5.0 | 修复各别机器下程序崩溃的问题 || 2022-04-25 | 1.4.1 | EasyDL, BML升级 支持paddle2模型 || 2022-03-25 | 1.4.0 | 优化模型算法 || 2021-12-22 | 1.3.5 | CPU基础版推理引擎优化升级 ; demo程序优化环境依赖检测 || 2021-10-20 | 1.3.4 | CPU加速版推理引擎优化升级 || 2021-08-19 | 1.3.2 | 新增DEMO二进制文件 || 2021-06-29 | 1.3.1 | 预测引擎升级 || 2021-05-13 | 1.3.0 | 模型发布新增多种加速方案选择 ; 目标追踪支持x86平台的GPU及加速版 ; 展示已发布模型性能评估报告 || 2021-04-08 | 1.2.3 | 支持BML平台模型仓库本地上传模型 || 2021-03-09 | 1.2.2 | CPU加速版支持int8量化模型 || 2021-01-27 | 1.2.1 | 新增模型支持 ; 性能优 化 ; 问题修复 || 2020.12.18 | 1.2.0 | 推理引擎升级 || 2020-01-26 | 1.1.20 | 新增一些模型的加速版支持 || 2020.10.29 | 1.1.20 | 修复已知问 题 || 2020.05.15 | 1.1.16 | 优化性能 , 修复已知问题 || 2020-09-17 | 1.1.19 | 支持更多模型 || 2020.04.16 | 1.1.15 | 升级引擎版本 || 2020.06.23 | 1.1.17 | 支持专业版更多模型 || 2020.05.15 | 1.1.16 | 优化性能 , 修复已知问题 || 2020.04.16 | 1.1.15 | 升级引擎版本 || 2020.03.13 | 1.1.14 | 支持EdgeBoardVMX || 2020.02.23 | 1.1.13 | 支持多阶段模型 || 2020.01.16 | 1.1.12 | 预测默认使用推荐阈值 || 2019.02.26 | 1.1.11 | CPU加速版支持物体检测高精度 || 2019.12.04 | 1.1.10 | 支持图像分割 || 2019.02.11 | 1.1.9 | 支持 EasyDL 专业版 || 2019.08.29 | 1.1.8 | CPU 加速版支持 || 2019.07.19 | 1.1.7 | 提供模型更新工具 || 2019.05.16 | 1.1.3 | NVIDIA GPU 支持 || 2019.03.15 | 1.1.0 | 架构与功能完善 || 2019.02.28 | 1.0.6 | 引擎功能完善 || 2019.02.13 | 1.0.5 | paddlepaddle 支持 || 2018.11.30 | 1.0.0 | 第一版 !

快速开始

1. 安装依赖

必须安装:

安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642

Visual C++ Redistributable Packages for Visual Studio 2013

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

Visual C++ Redistributable Packages for Visual Studio 2015-2019

https://docs.microsoft.com/en-us/cpp/windows/latest-supported-vc-redist?view=msvc-160

Openvino (仅使用Python Intel Movidius必须)

- 使用 OpenVINO[™] toolkit 安装,请参考 OpenVINO toolkit 文档安装 2020.3.1LTS(必须)版本, 安装时可忽略Configure the Model Optimizer及 后续部分。
- 使用源码编译安装,请参考 Openvino Inference Engine文档编译安装 2020.3.1LTS (必须)版本。

安装完成后,请设置环境变量OPENVINO_HOME为您设置的安装地址,默认是C: \Program Files (x86) \IntelSWTools,并确保文件夹下的openvino 的快捷方式指到了2020.3.1LTS版本。

充属性			
算机名 硬件 高级 系统保护 远程	xiaoyuchen 的用户变量(U)		_
	变量	值	
要进行大多数更改,你必须作为管理员登录。		Cillier ConeDrive	
性能	Path	C:\UsersNAppData\Local\Microsoft\WindowsApp	
视觉效果,处理器计划,内存使用,以及虚拟内存	TEMP	C:\Users\ AppData\Local\Temp	
设置(5)	тмр	C:\Users\xxxxyz	
用户配置文件			
与登录帐户相关的桌面设置		新建(N) 编辑(E) 删除(D)	
设置(£)	系统变量(S)		
	变量	值	^
启动和政障恢复	ComSpec	C:\Windows\system32\cmd.exe	
系统启动、系统故障和调试信息	INTEL DEV REDIST	C:\Windows\System32\Drivers\DriverData	
公開の	MIC_LD_LIBRARY_PATH	%INTEL_DEV_REDIST%compiler\lib\intel64_win_mic	
	NUMBER_OF_PROCESSORS	8	
	OS	Windows_NT	
环境变量(<u>N</u>)	Path	C:\Program Files (x86)\Parallels\Parallels Tools\Applications;C	~
		新建(W) 编辑(I) 删除(L)	
确定 取消 应用(A)			
		确 定 取消	
IntelSWTools			
共享 查着			
Register And American Stress (x86) → IntelS	WTools →	✓ ⁰ < 損	螦
□ 夕称	^	修改日期	
L 1113		TPRAH70	
💦 🔪 openvir	no	2020/4/15 23:41	
openvir	no_2019.2.275	2020/3/10 19:34	
* openvir	no 2020 2 117	2020/4/15 23:42	
		2020/4/10 23.42	

注意事项

- 1. 安装目录不能包含中文
- 2. Windows Server 请自行开启,选择"我的电脑"——"属性"——"管理"——"添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启即 可。

2. 运行离线SDK

解压下载好的SDK,打开EasyEdge.exe,输入Serial Num,选择鉴权模式,点击"启动服务",等待数秒即可启动成功,本地服务默认运行在

http://127.0.0.1:24401/

其他任何语言只需通过HTTP调用即可。

如启动失败,可参考如下步骤排查:

○百度智能云	故障排查
零门槛AI开发 平台EasyDL	
Windows离线SDK(公有云发 布)•问题排查	<

2.1 离线鉴权 (默认鉴权模式) 首次联网激活,后续离线使用

Al ver. 1.2.0.0 core. 1	1.8.1 sec. 1.5.1 —	×
Model	Market States and South Contraction	
Serial Num		
Host	127.0.0.1	
Port	24401	
鉴权模式	● 离线鉴权 ○ 按实例数鉴权	
服务状态	服务未启动	
	启动服务	
	是否开机自动启动	

2.2 按实例数鉴权 周期性联网激活,离线后会释放所占用鉴权,启动时请确保心跳间隔小于等于生成序列号时填写的定期确认时间

AI ver. 1.2.0.0 core. 1	1.8.1 sec. 1.5.1 —	×
Model	[4] A.P. Sales and R. Barter, Phys. Rev. B 10, 1998.	
Serial Num		
Host	127.0.0.1	
Port	24401	
鉴权模式	○ 离线鉴权 ④ 按实例数鉴权 心跳间隔 20 秒 🗸	
服务状态	服务未启动	
	启动服务	
	是否开机自己	动启动 🗌

基于源码集成时,若需要按实例数鉴权,需要通过代码指定使用按实例数鉴权

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2); global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 20);

或通过环境变量指定

set EDGE_CONTROLLER_KEY_AUTH_MODE=2 set EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=20

心 2.3 序列号激活错误码

错误码	文案	描述
4001	parameters missing. 参数缺失	
4002	parameters invalid. 参数不合法	
4003	model invalid. 模型信息不合法	
4004	no more activation left. 该序列号和该设备的激活次数超上限	
4005	the serial key is out of date. 该序列号过期	
4006	the serial key has been activated. 该序列号已被其他设备激活	该序列号已被其他设备激活,不能重复激活。
4007	account invalid. 序列号不能用于其他账号的模型	序列号不能用于其他账号的模型,只能用于绑定账号的模型。
4008	serial key invalid. 序列号不合法	序列号不存在或找不到
4009	bundle id invalid. 包名不合法	
4010	product invalid. 产品不合法	如easydl的SDK使用BML的序列号来激活,会报该错误
4011	platform invalid. 平台不合法	
4012	activate too frequent. 激活太频繁	激活太频繁,请稍后再进行激活。
4013	device type and license type not match. 硬件类型和序列号类型不匹配	如使用加速版序列号激活基础版SDK会报该错误
4014	exceed max activate device num. 超过最大授权数量	
4015	technology invalid. 技术类型不合法	
4016	exceed max activate entity num. 超过最大模型数量	
4017	device invalid. 设备不合法	
4018	model invalid. 模型不合法	

3. Demo示例(以图像服务为例)

服务运行成功,此时可直接在浏览器中输入http://127.0.0.1:24401,在h5中测试模型效果。



使用说明

调用说明

使用示例代码如下

python	
c#	
C++	
java	

import requests

```
with open('./1.jpg', 'rb') as f:
img = f.read()
```

```
**params 为GET参数 data 为POST Body**
result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1},
data=img).json()
```

 $\frac{\text{print}(\text{result})}{\text{result}}$

结果 获取的结果存储在response字符串中。请求参数

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送图片二进制。

返回参数|字段|类型|取值|说明||--------|----|-----|-----|| confidence | float | 0~1 | 分割的置信度 || label | string || 分割的类 别|| index | number || 分割的类别 || mask | string || 游程编码的mask | 代码参考 https://github.com/Baidu-AIP/EasyDL-Segmentation-Demo

集成指南

基于HTTP集成

通过EasyEdge.exe启动服务后,参照上面的调用说明,通过HTTP请求集成到自己的服务中

基于c++ dll集成

集成前提

解压开的SDK包中包含src、lib、dll、include四个目录才支持基于c++ dll集成

集成方法

参考src目录中的CMakeLists.txt进行集成

基于c# dll集成

集成前提

解压开的SDK包中包含src\demo_serving_csharp、dll两个目录才支持基于c# dll集成

集成方法

参考src\demo_serving_csharp目录中的CMakeLists.txt进行集成

FAQ

1. 服务启动失败,怎么处理?

根据SDK内的README.md检查依赖是否都已正确安装

请确保相关依赖都安装正确,版本必须如下: .NET Framework 4.5 Visual C++ Redistributable Packages for Visual Studio 2013 * Visual C++ Redistributable Packages for Visual Studio 2015

如使用的是CPU加速版,需额外确保Openvino安装正确,版本为2020.3.1LTS版 如使用Windows Server,需确保开启桌面体验

2. 服务调用时返回为空,怎么处理?调用输入的图片必须是RGB格式,请确认是否有alpha通道。

3. 多个模型怎么同时使用? SDK设置运行不同的端口,点击运行即可。

4. JAVA、C#等其他语言怎么调用SDK?

参考 https://ai.baidu.com/forum/topic/show/943765

5. 启动失败,缺失DLL? 打开EasyEdge.log,查看日志错误,根据提示处理 缺失DLL,请使用 https://www.dependencywalker.com/ 查看相应模 块依赖DLL缺失哪些,请自行下载安装

6. 启动失败,报错NotDecrypted? Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

7. 启动失败,报错 SerialNum无效

日志显示failed to get/check device id(xxx)或者 Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 C:\Users\\${用户名}\.baidu\easyedge 目录,再重新激活。

8. 勾选"开机自动启动"后,程序闪退

一般是写注册表失败。

可以确认下HKEY_CURRENT_USER下Software \Microsoft \Windows \CurrentVersion \Run能否写入(如果不能写入,可能被杀毒软件等工具管制)。也可以尝试基于bin目录下的easyedge_serving.exe命令行形式的二进制,自行配置开机自启动。

其他问题 如果无法解决,可到论坛发帖: https://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

^の Linux集成文档-C++

简介

本文档介绍EasyEdge/EasyDL的Linux CPP SDK的使用方法。

- 网络类型支持: 图像分类 物体检测 图像分割
- 硬件支持:
 - CPU: aarch64 armv7hf
 - GPU: ARM Mali G系列
 - ASIC: Hisilicon NNIE1.1 on aarch64 (Hi3559AV100/Hi3559CV100等)
 - ASIC: Hisilicon NNIE1.2 on armv7I (Hi3519AV100/Hi3559V200等)
 - Intel Movidius MyRIAD2 / MyRIAD X on x86_64
 - Intel Movidius MyRIAD2 / MyRIAD X on armv7I
 - Intel Movidius MyRIAD2 / MyRIAD X on aarch64
 - Intel iGPU on x86_64
 - 比特大陆 Bitmain SE50 (BM1684)
 - 瑞芯微 RK3399Pro / RV1109 / RV1126 / RK3568 / RK3588
 - 华为 Atlas200
 - 晶晨 A311D
 - 寒武纪 MLU220 on aarch64
 - 英特尔 iGPU
- 操作系统支持:

- Linux (Ubuntu, Centos, Debian等)
- 海思HiLinux
- 树莓派Raspbian/Debian
- 瑞芯微Firefly

性能数据参考算法性能及适配硬件

Release Notes | 时间 | 版本 | 说明 | | --- | ---- | | 2023.08.31 | 1.8.3 | Atlas系列Socs支持语义分割模型, Atlas Cann版本升级至6.0.1 | | 2023.06.29 | 1.8.2 | 比特大陆版本升级至V23.03.01 | | 2023.05.17 | 1.8.1 | 新增支持intel iGPU + CPU异构模式 | | 2023.03.16 | 1.8.0 | 新增 支持瑞芯微RK3588 || 2022.10.27 | 1.7.1 | 新增语义分割模型http请求示例 || 2022.09.15 | 1.7.0 | 新增瑞芯微 RK3568 支持, RK3399Pro、 RV1126升级到RKNN1.7.1 || 2022.07.28 | 1.6.0 | 引擎升级;新增英特尔 iGPU 支持 || 2022.04.25 | 1.4.1 | EasyDL, BML升级支持paddle2模 型 | | 2022.03.25 | 1.4.0 | EasyDL新增上线支持晶晨A311D NPU预测引擎;Arm CPU、Arm GPU引擎升级;atlas 200在EasyDL模型增加多个量 化加速版本; || 2021.12.22 | 1.3.5 | RK3399Pro, RV1109/RV1126 SDK扩展模型压缩加速能力,更新端上推理库版本;边缘控制台IEC功能升 级,适配更多通用小型设备,NNIE 在EasyDL增加量化加速版本;Atlas200升级到Cann5.0.3 | | 2021.06.29 | 1.3.1 | 视频流解析支持调整分辨 率;预测引擎升级;设备端sdk新增支持瑞芯微RV1109、RV1126 || 2021.05.13 | 1.3.0 | 新增视频流接入支持;EasyDL模型发布新增多种加速 方案选择;目标追踪支持x86平台的CPU、GPU加速版;展示已发布模型性能评估报告 | | 2021.03.09 | 1.2.0 | http server服务支持图片通过 base64格式调用 || 2021.01.27 | 1.1.0 | EasyDL经典版分类高性能模型升级;部分SDK不再需要单独安装OpenCV;新增RKNPU预测引擎支持; 新增高通骁龙GPU预测引擎支持 | | 2020.12.18 | 1.0.0 | 1.0版本发布!安全加固升级、性能优化、引擎升级、接口优化等多项更新 | | 2020.10.29 | 0.5.7 | 优化多线程预测细节 | | 2020.09.17 | 0.5.6 | 支持linux aarch64架构的硬件接入intel神经计算棒预测;支持比特大陆计算盒 SE50 BM1684 || 2020.08.11 | 0.5.5 | 支持linux armv7hf架构硬件(如树莓派)接入intel神经计算棒预测 || 2020.06.23 | 0.5.4 | arm引擎升级 || 2020.05.15 | 0.5.3 | 支持EasyDL 专业版新增模型 ; 支持树莓派(armv7hf, aarch64)| | 2020.04.16 | 0.5.2 | Jetson系列SDK支持多线程infer | | 2020.02.23 | 0.5.0 | 新增支持人脸口罩模型; Jetson SDK支持批量图片推理; ARM支持图像分割 || 2020.01.16 | 0.4.7 | 上线海思NNIE1.2, 支 持EasyEdge以及EasyDL;ARM引擎升级;增加推荐阈值支持||2019.12.26|0.4.6|海思NNIE支持EasyDL专业版||2019.11.02|0.4.5|移除 curl依赖;支持自动编译OpenCV;支持EasyDL 专业版 Yolov3;支持EasyDL经典版高精度物体检测模型升级||2019.10.25|0.4.4|ARM引擎升 级,性能提升30%; 支持EasyDL专业版模型 || 2019.09.23 | 0.4.3 | 增加海思NNIE加速芯片支持 || 2019.08.30 | 0.4.2 | ARM引擎升级;支持 分类高性能与高精度模型||2019.07.25|0.4.1|引擎升级,性能提升||2019.06.11|0.3.3|paddle引擎升级;性能提升||2019.05.16| 0.3.2 | 新增armv7l支持 || 2019.04.25 | 0.3.1 | 优化硬件支持 || 2019.03.29 | 0.3.0 | ARM64 支持;效果提升 || 2019.02.20 | 0.2.1 | paddle 引擎支持;效果提升||2018.11.30|0.1.0|第一版!|

【1.0 接口升级】参数配置接口从1.0.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例。【关于SDK包与RES模型文件夹配套使用的说明】 我们强烈建议用 户使用部署tar包中配套的SDK和RES一起使用。更新模型时,如果SDK版本号有更新,请务必同时更新SDK,旧版本的SDK可能无法正确适 配新发布出来的RES。

快速开始

SDK在以下环境中测试通过

- aarch64(arm64), Ubuntu 16.04, gcc 5.3 (RK3399)
- Hi3559AV100, aarch64, Ubuntu 16.04, gcc 5.3
- Hi3519AV100, armv7I , HiLinux 4.9.37, (Hi3519AV100R001C02SPC020)
- armv7hf, Raspbian, (Raspberry 3b)
- aarch64, Raspbian, (Raspberry 4b)
- armv7hf, Raspbian, (Raspberry 3b+)
- armv7hf, Ubuntu 16.04, (RK3288)
- Bitmain se50 BM1684, Debian 9
- Rockchip rk3399pro, Ubuntu 18.04
- Rockchip rv1126, Debain 10
- Rockchip rk3568, Ubuntu 20.04
- Rockchip rk3588, Ubuntu 20.04

- Atlas200(华为官网指定的Ubuntu 18.04版本)
- Amlogic A311D, Ubuntu 20.04
- MLU220, aarch64, Ubuntu 18.04

安装依赖

依赖包括

- cmake 3+
- gcc 5.4 以上(需包含 GLIBCXX_3.4.22), gcc / glibc版本请以实际SDK ldd的结果为准
- opencv3.4.5 (可选)

依赖说明:树莓派 树莓派Raspberry默认为armv7hf系统,使用SDK包中名称中包含 armv7hf_ARM_的tar包。如果是aarch64系统,使用SDK包中 名称中包含 aarch64_ARM_的tar包。

在安装前可通过以下命令查看是32位还是64位 :

getconf LONG_BIT 32

依赖说明:比特大陆SE计算盒 需要安装SophonSDK V23.05.01及以上版本,SDK的默认安装位置为 /opt/sophon/,如SDK安装在自定义地址, 需在CMakeList.txt中指定SDK安装地址:

这里修改并填入所使用的SophonSDK路径 set(EDGE_BMSDK_ROOT "{这里填写sdk路径}")

可通过命令 bm-smi 查看内部SDK和驱动的版本号(SophonSDK V23.05.01对应的内部SDK和驱动为0.4.6)。对于使用旧版BM1684 SDK或者 低版本SophonSDK的用户,可参考SophonSDK安装包中的《LIBSOPHON 使用手册》先卸载旧版BM1684 SDK,安装、升级SophonSDK。

依赖说明:海思开发板 海思开发板需要根据海思SDK文档配置开运行环境和编译环境,SDK和opencv都需要在该编译环境中编译。NNIE1.2用 arm-himix200-linux交叉编译好的opencv,下载链接:https://pan.baidu.com/s/13QWOReeWx4ZwgYg4Iretyw 密码:yq0s。下载后修改SDK CMakesList.txt

依赖说明:RK3399Pro 所有用例基于 Npu driver版本1.7.1的RK3399pro开发板测试通过,SDK采用预编译模式,请务必确保板上驱动版本为 1.7.1 查看RK3399Pro板上driver版本方法:dpkg -I | grep 3399pro

依赖说明:RV1109/RV1126 所有用例基于Rknn_server版本1.7.3的RV1126开发板测试通过,SDK采用预编译模式,请务必确保板上驱动版本为 1.7.3 查看RV1109/RV1126板上Rknn_server版本方法:strings /usr/bin/rknn_server | grep build

依赖说明:RK3568 所有用例基于Rknn_server版本1.2.0的RK3568开发板测试通过, 查看RK3568板上Rknn_server版本方法: strings /usr/bin/rknn_server | grep build

依赖说明:RK3588 RK3588开发板需要确保环境正确安装了RKNPU驱动,平台用例基于v0.8.0版本的RKNPU驱动测试通过,查看RK3588NPU驱 动版本的方法: sudo cat /sys/kernel/debug/rknpu/version

依赖说明:晶晨A311D所有用例基于晶晨A311D开发板测试通过,需要驱动版本为 6.4.4.3(下载驱动请联系开发版厂商) 查看晶晨A311D开发 板驱动版本方法:dmesg | grep Galcore

依赖说明:英特尔iGPU用户在使用英特尔iGPU SDK前,需要根据英特尔官方文档提前安装好英特尔集成显卡驱动以及相关基础软件环境,安装 完成后通过 clinfo 指令确认OpenCL能够正常识别到集成显卡信息,正确识别集显情况下clinfo指令输出参考如下:

rootebalaa-Qlillanm+30-N0001~# Clinto	
Number of platforms	
Platform Name	Intel(R) OpenCL HD Graphics
Platform Vendor	Intel(R) Corporation
Platform Version	OpenCL 3.0
Platform Profile	FULL_PROFILE
Platform Extensions	cl_khr_byte_addressable_store cl_khr_device_uuid cl_khr_fp16 cl_khr_global_int32_base_atomics cl_khr_global_int32_extended_atomics cl_khr_icd cl_khr_local_int32
_base_atomics cl_khr_local_int32_extended_atomics	cl_intel_command_queue_families cl_intel_subgroups cl_intel_required_subgroup_size cl_intel_subgroups_short cl_khr_spir cl_intel_accelerator cl_intel_driver_dia
<pre>gnostics cl_khr_priority_hints cl_khr_throttle_hi</pre>	nts cl_khr_create_command_queue cl_intel_subgroups_char cl_intel_subgroups_long cl_khr_il_program cl_intel_mem_force_host_memory cl_khr_subgroup_extended_types c
l_khr_subgroup_non_uniform_vote cl_khr_subgroup_b	allot cl_khr_subgroup_non_uniform_arithmetic cl_khr_subgroup_shuffle cl_khr_subgroup_shuffle_relative cl_khr_subgroup_clustered_reduce cl_intel_device_attribute_
query cl_khr_suggested_local_work_size cl_intel_s	plit_work_group_barrier cl_khr_fp64 cl_khr_subgroups cl_intel_spirv_device_side_avc_motion_estimation cl_intel_spirv_media_block_io cl_intel_spirv_subgroups cl_k
hr_spirv_no_integer_wrap_decoration cl_intel_unif	ied_shared_memory cl_khr_mipmap_image cl_khr_mipmap_image_writes cl_intel_planar_yuv cl_intel_packed_yuv cl_intel_motion_estimation cl_intel_device_side_avc_moti
on_estimation cl_intel_advanced_motion_estimation	cl_khr_int64_base_atomics cl_khr_int64_extended_atomics cl_khr_image2d_from_buffer cl_khr_depth_images cl_khr_3d_image_writes cl_intel_media_block_io cl_intel_v
a_api_media_sharing cl_intel_sharing_format_query	cl_khr_pci_bus_info
Platform Host timer resolution	ins
Platform Extensions function suffix	INTEL
Platform Name	Intel(R) OpenCL HD Graphics
Number of devices	
Device Name	Intel(R) UHD Graphics 630 [0x9bc8]
Device Vendor	Intel(R) Corporation
Device Vendor ID	0x8086
Device Version	OpenCL 3.0 NEO
Driver Version	22.53.25242.13
Device OpenCL C Version	OpenCL C 1.2
Device Type	GPU
Device Profile	FULL_PROFILE
Device Available	Yes
Compiler Available	Yes
Linker Available	Yes
Max compute units	24
Max clock frequency	1150MHz
Device Partition	(core)

使用序列号激活 请在官网获取序列号

纯高线服务说明 发觉你就最短期,将到底完成的模型图里在本地。面线调明模型,可以选择环模型部署在本地的服务器、小型设备、软键一体方面专项运配键件上, 通过API、SCK进一步集成,灵活运道不同业务场器。 							
服务器 通用小型设备 专项适配硬件							
SDK API							
此处发布、下载的SDK为未授你SDK,需要前往拉勒台段取序列号激活后才能正式使用。SDK的附有对应版本的Demo及开发文性,开发者可参考演代码完成开发。							
模型名称	发布版本 🗘	应用平台	模型加速	发布状态	发布时间 🗘		
	134318-V1 查看性能报告	通用X86 CPU-Linux	基础版	• 已发布	2021-08-19 20:24	下载SDK	
sun_小目标test			精度无损压缩加速	 已发布 	2021-08-19 20:24	下载加速版SDK	
		英伟达GPU-Linux	基础版	•已发布	2021-08-19 20:35	下载SDK	
			精度无损压缩加速	 已发布 	2021-08-19 20:34	下载加速版SDK	
			基础版	 已发布 	2021-08-19 18:17	下载SDK	

SDK内bin目录下提供预编译二进制文件,可直接运行(二**进制运行详细说明参考下一小节**),用于图片推理和模型http服务,在二进制参数的 serial_num(或者serial_key)处填入序列号可自动完成联网激活(请确保硬件首次激活时能够连接公网,如果确实不具备联网条件,需要使用纯离 线模式激活,请下载使用百度智能边缘控制台纳管SDK)

SDK内提供的一些二进制文件,填入序列号可完成自动激活,以下二进制具体使用说明参考下一小节 ./edgekit_serving --cfg=./edgekit_serving.yml ./easyedge_image_inference {model_dir} {image_name or image_directory} {serial_num} ./easyedge_serving {res_dir} {serial_key} {host} {port}

如果是基于源码集成,设置序列号方法如下

global_controller()->set_licence_key("")

默认情况下(联网激活或者离线激活的场景),按照上述说明正确设置序列号即可,如果是实例数鉴权模式(请在百度智能云控制台再次确认自己的序列号是实例数鉴权模式,仅实例数鉴权需要进行下面的变量或者源码设置)实例数鉴权环境变量设置方法

export EDGE_CONTROLLER_KEY_AUTH_MODE=2 export EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=30

实例数鉴权源码设置方法

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2) global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 300)

基于预编译二进制测试图片推理和http服务测试图片推理模型资源文件默认已经打包在开发者下载的SDK包中。

对于硬件使用为:Intel Movidius MyRIAD2 / MyRIAD X / IGPU on Linux x86_64 / armv7hf / aarch64,在编译或运行demo程序前执行以下命 令: source \${cpp_kit位置路径}/thirdparty/openvino/bin/setupvars.sh 或者执行

source \${cpp_kit位置路径}/thirdparty/openvino/setupvars.sh (openvino-2022.1+) 如果SDK内不包含setupvars.sh脚本,请忽略该提示

运行预编译图片推理二进制,依次填入模型文件路径(RES文件夹路径)、推理图片、序列号(序列号尽首次激活需要使用,激活后可不用填序列号 也能运行二进制)

./easyedge_image_inference {model_dir} {image_name or image_directory} {serial_num}
LD_LIBRARY_PATH=../lib ./easyedge_image_inference ../../../RES /xxx/cat.jpeg "1111-1111-1111"

demo运行效果:



```
>./easyedge_image_inference ../../../RES 2.jpeg
2019-02-13 16:46:12,659 INFO [EasyEdge] [easyedge.cpp:34] 140606189016192 Baidu EasyEdge Linux Development Kit
0.2.1(20190213)
2019-02-13 16:46:14,083 INFO [EasyEdge] [paddlev2_edge_predictor.cpp:60] 140606189016192 Allocate graph success.
2019-02-13 16:46:14,326 DEBUG [EasyEdge] [paddlev2_edge_predictor.cpp:143] 140606189016192 Inference costs 168 ms
1, 1:txt_frame, p:0.994905 loc: 0.168161, 0.153654, 0.920856, 0.779621
Done
```

推荐使用 edgekit_serving 启动模型服务 LD_LIBRARY_PATH=../lib ./edgekit_serving --cfg=./edgekit_serving.yml **也可以使用 easyedge_serving 启动模型服务** **./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401}**

LD_LIBRARY_PATH=../lib ./easyedge_serving ../../../RES "1111-1111-1111" 0.0.0.0 24401

后,日志中会显示

HTTP(or Webservice) is now serving at 0.0.0.0:24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试,网页右侧会展示模型推理结果



同时,可以调用HTTP接口来访问服务。

请求http服务 以图像预测场景为例(非语义分割模型场景,语义分割请求方式参考后面小节详细文档),提供一张图片,请求模型服务的示例参考如下demo

python	
c#	
C++	
java	

import requests

```
with open('./1.jpg', 'rb') as f:
img = f.read()
```

```
**params 为GET参数 data 为POST Body**
result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1},
data=img).json()
```

 $\frac{\text{print}(\text{result})}{\text{result}}$

关于http接口的详细介绍参考下面集成文档http服务章节的相关内容

集成文档

使用该方式,将运行库嵌入到开发者的程序当中。 编译demo项目 SDK src目录下有完整的demo工程,用户可参考该工程的代码实现方式将SDK 集成到自己的项目中,demo工程可直接编译运行:

cd src mkdir build && cd build cmake .. && make ./easyedge_image_inference {模型RES文件夹} {测试图片路径} **如果是NNIE引擎,使用sudo运行** sudo ./easyedge_image_inference {模型RES文件夹} {测试图片路径}

(可选) SDK包内一般自带opencv库,可忽略该步骤。如果希望SDK自动编译安装所需要的OpenCV库,修改cmake的 optionEDGE_BUILD_OPENCV为ON即可。 SDK会自动从网络下载opencv源码,并编译需要的module、链接。注意,此功能必须需联网。

cmake -DEDGE_BUILD_OPENCV=ON .. && make -j16

若需自定义library search path或者gcc路径,修改CMakeList.txt即可。

对于硬件使用为Intel Movidius MyRIAD2 / MyRIAD X 的,如果宿主机找不到神经计算棒Intel® Neural Compute Stick,需要执行以下命令添加 USB Rules:

cp \${cpp_kit位置路径}/thirdparty/openvino/deployment_tools/inference_engine/external/97-myriad-usbboot.rules /etc/udev/rules.d/ sudo udevadm control --reload-rules sudo udevadm trigger sudo ldconfig

使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

// step 1: 配置运行参数

EdgePredictorConfig config config.model_dir = {模型文件目录}; // step 2: 创建并初始化Predictor; 这这里选择合适的引擎 auto predictor = global_controller()->CreateEdgePredictor(config); // step 3-1: 预测图像 auto img = cv::imread({图片路径}); std::vector<EdgeResultData> results; predictor->infer(img, results); // step 3-2: 预测视频 std::vector<EdgeResultData> results; FrameTensor frame_tensor; VideoConfig video_config; video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h video_config.source_value = video_src; /* ... more video_configs, 根据需要配置video_config的各选项 */ auto video_decoding = CreateVideoDecoding(video_config); while (video_decoding->next(frame_tensor) == EDGE_OK) { results.clear(); if (frame_tensor.is_needed) { predictor->infer(frame_tensor.frame, results); render(frame_tensor.frame, results, predictor->model_info().kind); //video_decoding->display(frame_tensor); // 显示当前frame,需在video_config中开启配置 //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置

对于口罩检测模型,将 EdgePredictorConfig config修改为PaddleMultiStageConfig config即可。

口罩检测模型请注意输入图片中人脸大小建议保持在 88到9696像素之间,可根据场景远近程度缩放图片后再传入SDK。

SDK参数配置 SDK的参数通过EdgePredictorConfig::set_config和global_controller()->set_config配置。set_config的所有key 在easyedge_xxxx_config.h中。其中

- PREDICTOR前缀的key是不同模型相关的配置,通过EdgePredictorConfig::set_config设置
- CONTROLLER前缀的key是整个SDK的全局配置,通过global_controller()->set_config设置

以序列号为例,KEY的说明如下:

```
/**
* @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号
* 值类型:string
* 默认值:空
*/
static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";
```

```
使用方法如下:
```

```
EdgePredictorConfig config;
config.model_dir = ...;
config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, "1DB7-1111-1111-D27D");
```

具体支持的运行参数可以参考开发工具包中的头文件的详细说明。

初始化

接口

```
auto predictor = global_controller()->CreateEdgePredictor(config);
predictor->init();
```

若返回非0,请查看输出日志排查错误原因。

预测图像

```
• 接口
```

```
/**
* @brief
* 通用接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
    cv::Mat& image. std::vector<EdgeResultData>& result
) = 0;
```

图片的格式务必为opencv默认的BGR, HWC格式。

• 返回格式

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index: // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测活图像分割时才有
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割时才有
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding, 游程编码的mask
};
```

关于矩形坐标

```
x1 * 图片宽度 = 检测框的左上角的横坐标
```

- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2*图片宽度 = 检测框的右下角的横坐标

```
y2 * 图片高度 = 检测框的右下角的纵坐标
```

关于图像分割mask

```
cv::Mat mask为图像掩码的二维数组
{
{0,0,0,0,0,0,0,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,0,0,0,0,0,0,0},
}
其中1代表为目标区域,0代表非目标区域
```

关于图像分割mask_rle

该字段返回了mask的游程编码,解析方式可参考 http demo

Baidu 百度智能云文档

预测视频

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

● 接口

classVideoDecoding :

```
/**
* @brief 获取输入源的下一帧
* @param frame_tensor
* @return
*/
virtual int next(FrameTensor &frame_tensor) = 0;
/**
* @brief 显示当前frame_tensor中的视频帧
* @param frame_tensor
* @return
*/
virtual int display(const FrameTensor &frame_tensor) = 0;
/**
* @brief 将当前frame_tensor中的视频帧写为本地视频文件
* @param frame_tensor
* @return
*/
virtual int save(FrameTensor &frame_tensor) = 0;
/**
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
 * @brief 获取视频的width属性
 * @return
*/
virtual int get_width() = 0;
/**
* @brief 获取视频的height属性
* @return
*/
```

struct VideoConfig

virtual int get_height() = 0;

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type;
                          // 输入源类型
  std::string source_value;
                         // 输入源地址, 如视频文件路径、摄像头index、网络流地址
 int skip_frames{0};
                       // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
  int retrieve_all{false};
                     // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
  int input_fps{0};
                       // 在采取抽帧之前设置视频的fps
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
  bool enable_display{false};
  std::string window_name{"EasyEdge"};
  bool display_all{false};
                        // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                        // 是否存储所有frame, 若为false, 仅存储根据skip_frames抽取的frame
  std::map<std::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被 抽取返回,以作为显示或存储用。input_fps:用于抽帧前设置fps。resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。conf:高级选项。部分配置会通过该map来设置。

注意: 1.如果使用 VideoConfig的 display 功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。 2.使用摄像头抽帧时,如 果通过 resolution 设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

设置序列号

请在网页控制台中申请序列号,并在init初始化前设置。 LinuxSDK 首次使用需联网授权。

EdgePredictorConfig config; config.set_config(easyedge::params::PREDICTOR_KEY_SERIAL_NUM, "this-is-serial-num");

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

```
EdgeLogConfig log_config;
log_config.enable_debug = true;
global_controller()->set_log_config(log_config);
```

http服务

1. 开启http服务 http服务的启动参考demo_serving.cpp文件。

/** * @brief 开启一个简单的demo http服务。 * 该方法会block直到收到sigint/sigterm。 * http服务里,图片的解码运行在cpu之上,可能会降低推理速度。 * @tparam ConfigT * @param config * @param host * @param port * @param service_id service_id user parameter, uri '/get/service_id' will respond this value with 'text/plain' * @param instance_num 实例数量,根据内存/显存/时延要求调整 * @return */ template<typename ConfigT> int start_http_server(const ConfigT &config, const std::string &host int port,

2. http接口详细说明 http 请求方式一:无额外编码 URL中的get参数:

const std::string &service_id, int instance_num = 1);

参数	说明		默认值	
threshold	阈值过滤,	0~1	如不提供,	则会使用模型的推荐阈值

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

Python请求示例 (图片测试,针对图像分类、物体检测、实例分割等模型)

import requests

with open('./1.jpg', 'rb') as f: img = f.read() result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1}, data=img).json()

Python请求示例 (图片测试, 仅针对语义分割模型, 同其他CV模型不同, 语义分割模型输出为灰度图)

import requests

```
with open('./1.jpg', 'rb') as f:
    img_data = f.read()
    res = requests.post('http://127.0.0.1:24401/',
        data=img_data)
    with open("gray_result.png", "wb") as fb:
        fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
```

Python请求示例 (视频测试, 注意:区别于图片预测,需指定Content-Type;否则会调用图片推理接口)

import requests

with open('./1.mp4', 'rb') as f: video_data = f.read() result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1}, headers={'Content-Type': 'video'}, data=video_data).json()

http 请求方法二:json格式,图片传base64格式字符串 HTTP方法:POST Header如下:

参数 值 Content-Type application/json

Body请求填写:

• 图像分类网络: body 中请求示例

```
{
"image": "<base64数据>",
"top_num": 5
}
```

body中参数详情

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
top_nu m	否	number	-	返回分类数量,不填该参数,则默认返回全部分类结果

• 物体检测和实例分割网络: Body请求示例:

```
{
"image": "<base64数据>",
"threshold": 0.3
}
```

```
body中参数详情:
```

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
thresho Id	否	number	-	默认为推荐阈值,也可自行根据需要进行设置

• 语义分割网络: Body请求示例:

```
{
"image": "<base64数据>"
}
```

body中参数详情(语义分割由于模型特殊性,不支持设置threshold值,设置了也没有意义):

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部

Python请求示例 (非语义分割模型参考如下代码)
import base64

import requests
def main(): with open("1.jpg 【图片路径】", 'rb') as f: result = requests.post("http://{服务ip地址}:24401/", json={ "image": base64.b64encode(f.read()).decode("utf8")
})
<pre># print(result.request.body)</pre>
print(result.request.headers)
print(result.content)
ifname == 'main':
main()

Python 请求示例 (针对语义分割模型,同其他CV模型不同,语义分割模型输出为灰度图)

```
import base64
import requests
def main():
    with open("1.jpg 【图片路径】", 'rb') as f:
        res = requests.post("http://{服务ip地址}:24401/", json={"image": base64.b64encode(f.read()).decode("utf8")})
        with open("gray_result.png", "wb") as fb:
        fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
if __name__ == '__main__':
        main()
```

http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考预测图像-返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

返回示例

```
{
    "cost_ms": 52,
    "error_code": 0,
    "results": [
        {
            "confidence": 0.94482421875,
            "index": 1,
            "label": "IronMan",
            "x1": 0.059185408055782318,
            "x2": 0.18795496225357056,
            "y1": 0.14762254059314728,
            "y2": 0.52510076761245728,
            "mask": "...", // 图像分割模型字段
            "trackId": 0, // 目标追踪模型字段
        },
    ]
}
```

其他配置

1. 日志名称、HTTP 网页标题设置

通过global_controller的set_config方法设置:

效果如下:

图片加载失败

FAQ

1. 如何处理一些 undefined reference / error while loading shared libraries?

如:./easyedge_demo: error while loading shared libraries: libeasyedge.so.1: cannot open shared object file: No such file or directory 这是因为二进制运行时Id无法找到依赖的库。如果是正确cmake && make 的程序,会自动处理好链接,一般不会出现此类问题。

遇到该问题时,请找到具体的库的位置,设置LD_LIBRARY_PATH。

示例一:libverify.so.1: cannot open shared object file: No such file or directory 链接找不到libveirfy.so文件,一般可通过 export LD_LIBRARY_PATH=\${LD_LIBRARY_PATH}:../../lib 解决(实际冒号后面添加的路径以libverify.so文件所在的路径为准)

示例二:libopencv_videoio.so.4.5: cannot open shared object file: No such file or directory 链接找不到ibopencv_videoio.so文件,一般可通 过 export LD_LIBRARY_PATH=\${LD_LIBRARY_PATH}:../../thirdparty/opencv/lib 解决(实际冒号后面添加的路径以libopencv_videoio.so所在路 径为准)

2. EasyDL SDK与云服务效果不一致,如何处理?

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

3. 如何将我的模型运行为一个http服务?目前epp sdk暂未集成http运行方式; 0.4.7版本之后,可以通过start_http_server方法开启http服务。

4. 运行NNIE引擎报permission denied 日志显示:

open sys: Permission denied open err : Permission denied open err : Permission denied

请使用sudo在root下运行。

5. 运行SDK报错 Authorization failed

情况一:日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888"
./easyedge_demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限 于)以下可能的情况:

- MAC地址变化
- 磁盘变更
- BIOS重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除~/.baidu/easyedge 目录,再重新激活。

6. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

7. 运行NNIE引擎报错 std::bad_alloc 检查开发板可用内存,一些比较大的网络占用内存较多,推荐内存500M以上

8. 运行二进制时,提示 libverify.so cannot open shared object file

可能cmake没有正确设置rpath,可以设置LD_LIBRARY_PATH为sdk的lib文件夹后,再运行:

LD_LIBRARY_PATH=\$LD_LIBRARY_PATH:../lib ./easyedge_demo

9. 编译时报错:file format not recognized 可能是因为在复制SDK时文件信息丢失。请将整个压缩包复制到目标设备中,再解压缩、编译

心 端云协同服务说明

服务简介

EasyDL端云协同服务由EasyEdge端与边缘AI服务平台提供、基于百度智能边缘构建,能够便捷地将EasyDL定制模型的推理能力拓展至应用现场, 提供临时离线、低延时的计算服务。

「云管理,端计算」的端云协同服务,具体包括:

- 在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新
- 断网状态下模型离线计算(http服务,可调用与公有云API功能相同的接口)
- 联网状态下在平台管理设备运行状态、资源利用率

目前通用小型设备的应用平台支持Linux-ARM,具体使用流程请参考下方文档。

使用流程

Step 1 发布端云协同部署包

在我的部署包页面点击「发布端云协同部署包」

模型中心	端云协同服务 > 我的部署包							
我的模型	端云协同服务说明							
创建模型	1、在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新							
训练模型	2、断网状态下模型震线计算(http强务,可调用与公有云API功能相同的接口) 3、联网状态下在平台曾理设备运行状态、资源利用率							
校验模型	具体使用流程如下:							
发布模型	 在本页面发布端云协同部署包,填写相关信息并申请,并在下方列表宣看部署包发布状态 在「我的本地设备」页面斯增设备、复则激活命令、既则激活本地设备 							
EasyData数据服务	3、在「下发部著包制设备」页面,将发布成功的部署包一键下发到已激活的设备上,即可测试或正式使用							
数据总览	业态操示协同部署的							
标签组管理								
在线标注	端云协问服务名称 模型D 设备类型 应用平台 当前可用版本 新版本发布状态 部署设备数	操作						
云服务数据回流								
EasyEdge本地部署								
纯离线服务								
端云协同服务へ	^							
我的部署包	★ ★ 新聞							
我的本地设备	。							
下发部署包到设备								

填写服务名称,选择模型版本并提交发布

模型中心	端云协同服务 > 发布端云协同部署包						
我的模型	设备类型 🔷 服务器 💿 通用小型设备						
创建模型							
训练模型	模型名称 test2021 V						
校验模型	端云协同服务名称						
发布模型	选择版本 V1 🗸						
EasyData数据服务	选择系统和芯片 🔄 🛆 Linux						
数据总览	通用ARM						
标签组管理							
在线标注	发布部署包						
云服务数据回流							
EasyEdge本地部署							
纯离线服务							

在列表查看部署包发布状态

模型中心	端云协同服务 > 我的部署包								
我的模型	端云协同服务说明								
创建模型	1、在可视化界面轻松爽現模型部署包在边缘设备上的集成、版本更新								
训练模型	2、断网状态下模型离线计算(http温务,可调用与公有云API功能相同的接口) 3、联网状态下在平台管理设备运行状态、资源利用率								
校验模型	具体使用流程如下:								
发布模型	 在本页面发布融合协同影響包,填写相关信息并申请,并在下方列表查看影響包发布状态 在「我的本地设备」页面新增设备,复制激活命令,联网激活本地设备 								
EasyData数据服务	3、在「下发邮書包到设备」页面,将发布成功的邮署包一键下发到已激活的设备上,即可测试或正式使用								
数据总览	关布施云协问部署包								
标签组管理									
在线标注	端云协同服务名称 模型D 设备类型 应用平台 当前可用版本 新版本发布状态 部署设备数 操作								
一服务数据回流	ecc 264 通用小型设备 Linux-通用ARM V1 V1-已发布 0 下发到设备 发布新版本								
Add 75 KAIG (2016)	从发布页过来的 246 服务器 Linux-通用X86 CPU V2, V1 V2-已发布 1 下发到设备 发布新版本 服务详情								
EasyEdge本地部署	图像分类高精度_猫狗-265 265 通用小型设备 Linux-通用ARM - V1-发布中 0								
纯离线服务									
端云协同服务 へ	毎页显示 10 🗸 🤇 1 >								
我的部署包									
我的本地设备									

Step 2 新增设备并激活

在我的本地设备页面新增设备

模型中心	端云协同服务 > 我的]本地设备					
我的模型	我的本地设备						点击收起
创建模型	在本页面新增设备、联网 利用率等信息。	激活本地设备后,即可料	将「 我的部署包 」页面发	这布成功的部署包一键下	发到设备上。设备联网时,可以还	查看设备上部署的服务、设备的; ;	运行状态、资源
训练模型							
校验模型	新增设备						
发布模型							
EasyData数据服务	设备名称	创建时间	设备类型	应用平台	设备连接状态	最新同步时间	操作
数据总览							
标签组管理							

		新增设备	×
i i	备名称	20个字符以内	
备	注信息	50个字符以内,自定义备注信息	
0		0.50	
K		0/50	
		确认取消	

在列表中,点击设备对应的「激活设备」操作,复制激活命令并在本地设备上执行即可

	查看激活命令	× 收起 ^
在	复制命令,并在本地设备上执行,即可激活设备	
端云协同服务	<pre>curl -skfL 'https://verify.baidubce.com/active/v1/active/s sh?token=e177df4ed47b2265223a333630302c226b223a226e6f646522</pre>	/setup.
我的本地设备 在本页面新增设 利用率等信息。	6e223a22353730313835302d382d65393866386363302d343231612d653 12d373138392d3762666633666430663313433222c226e73223a22626165 6c2d636c6f7564222c227473223a31363039393939393434327d' -osetu 66 sh setup.sh	6537653 点击收起 1657479 tup.sh 可以重看设备上部署的服务、设备的运行状态、资源
新增设备		1词 场代
linux x86 🐨	知道了 2021-01-07 14:02 木政活	

Step 3 下发部署包到设备,在本地调用

在下发部署包到设备页面,将发布成功的部署包一键下发到已激活的设备上,即可测试或正式使用

模型中心	端云协同服务 > 下发部署包到设备					
我的模型	将已成功发	这布的端云协共服务部署包一键下发到已激活且在连接中的设计	<u>e</u>			
创建模型						
训练模型	服务名称	ecc	~			
校验模型	模型ID	264				
发布模型	应用平台	Linux-通用ARM	~			
EasyData数据服务						
数据总览						
标签组管理						

部署包下发成功之后,会在本地启动一个HTTP推理服务。在浏览器中输入http://{设备ip}:{服务端口,默认8080},即可预览效果:

当前阈值: 0.07 修改

置信度 0.998

0.937

0.841

0.781

0,408

0.324

0.251

▲ 不安全 yq01-ns-map0402a659.yq01:8081



具体接口调用说明请参考文档 SDK - HTTP服务调用说明

云端管理说明

模型部署包管理

在我的部署包页面可以进行已发布的模型部署包的管理。

发布及更新模型版本

点击「发布新版本」操作即可快速发布对应模型ID下的新版本。同一模型ID下已发布的模型版本均会显示在列表的「当前可用版本」中。

	模型中心 端云协同服务 > 我的部署包									
	我的模型 端云协同服务说明									点击收起
	创建模型 1、在可视化双面轻松实现模型部署包在边缘设备上的集成、版本更新									
	2、断网状态下模型属线计算(http混务,可调用与公有云API功能相同的接口) 调练模型 3、联网状态下在平台管理设备运行状态、资源利用率									
	校验模型 具体使用流程如下:									
	发布模型 1、在本项面波布端云协同部署包、填写相关信息并申请、并在下方列表世看部署包发布状态 2、在「契約本地设备」页面新增设备、复制激活命令、联网激活本地设备									
	EasyData数据服务	3、在「 下发部署包到设备 」页	面,将发7	6成功的部署包-	-键下发到已激活的设备上	,即可测试或正	式使用			
	数据总览	发布端云协同部署包								
	标签组管理									
	在线标注	端云协同服务名称	模型ID	设备类型	应用平台	当前可用版本	新版本发布状态	部署设备数	操作	
	云服务数据回流	ecc	264	通用小型设备	Linux-通用ARM	V1	V1-已友布	0	卜友到设备 友布新版本	
	EasyEdge本地部署		246	服労器	Linux-通用X86 CPU	V2, V1	V2-已友布	1	卜友到设备 友布新版本 日	服务详情
	纯离线服务	100 C 100 C	265	通用小型设备	LINUX-週用ARM	-	VI-友巾屮	U		
	端云协同服务 へ								毎页显示 10 🗸 <	1 >
	我的部署包									
	我的本地设备									
	下发部署包到设备									
_	_								_	
			发布	新版本	t			>	ĸ	
	将最新训练的横	莫型版本发布为服务,	发布质		「可从云端下发	到设备				
		服冬夕称								
		100.25 1010								
		模型ID 2	64							
		选择新版本	V1			\sim				
		6	制认		取消					

新版本发布成功后,即可在「下发部署包到设备」页面或当前服务的「服务详情」页面,将新版本下发到本地设备上。

模型中心	端云协同服务 > 我的部署包								
我的模型	端云协同服务说明 点击收表								
创建模型	1、在可视化界面轻松实现模型部署包在边缘设备上的集成、版本更新								
训练模型	2、断网状态下模型离线计算(http服务,可调用与公有云API功能相同的接口) 3. 联网状态下在平台管理设备运行状态。 意源利用率								
校验模型									
发布槽刑	1、在本页面发布端云协同部署包,填写相关信息并申请,并在下方列表查看部署包发布状态								
及印度至	2、在「我約本地设备」页面新增设备,复制激活命令,联网激活本地设备 3、在「下发邮票包到设备」页面,将发布成功的图看包一键下发到已激活的设备上,即可测试或正式使用								
EasyData数据服务									
数据总览	发布履云协同部署包								
标签组管理									
在线标注	端云协问服务名称 模型D 设备类型 应用平台 当前可用版本 新版本发布状态 部署设备数 操作								
云服务数据回流	ecc 264 通用小型设备 Linux-通用ARM V1 V1-已发布 0 下发到设备 发布新版本								
Facu Felge 大地 部 网	246 服务器 Linux-通用X86 CPU V2, V1 V2-已发布 1 下发到设备 发布新版本 服务详情								
EasyEdge本地部者	图像分类高精度_猫狗-265 265 通用小型设备 Linux-通用ARM - V1-发布中 0								
纯离线服务									
模型中心	端云协同服务 >Jummu,服务详情								
我的模型	昭永之松 超刊ID 246 当前可用近太 V2 V1 如居心冬秋 1/1								
创建模型									
训练模型									
校验模型	下发到更多设备								
岩布積刑	边条交投 网络下世旗网络士 如果树脂时间 边条体接住木 服务工业计术 网络回电时间 编作								
EasyData数据服务	ubuntu-local-TJY ▽ v2 2021-02-06 住地 V2ト及加切 2021-01-07 14:45 宣看被急能置 宣看设备详情 ト友熟成本 移族设备								
数据总览									
标签组管理									
在线标注									

管理模型已部署的设备

在上述的「服务详情」页面,可以查看并管理当前服务已部署的设备,包括移除设备、将服务下发到更多的设备等。

模型中心	端云协同服务 〉 》 服务详情						
我的模型	服务名称	模型ID 246	当前可用版本 V2, V1	部署设备数 1/1			
创建模型	设备类型 服务器	应用平台 Linux-通用X86 CPU					
训练模型	工业对市会活体						
校验模型	下及到更多以留						
发布模型	设备名称 最新下发模型员	版本 部署到期时间 设备连接状态	服务下发状态 最新同步时间	操作			
EasyData数据服务	ubuntu-local-fjy 💬 V2	2021-02-06 在线	V2下发成功 2021-01-07 14:45	查看服务配置 查看设备详情 下发新版本 修除设备			
数据总览							
标签组管理							
在线标注							

本地设备管理

在我的本地设备页面可以进行所有本地设备的管理。

查看单台设备的运行状态

点击单台设备的「服务详情」,可查看设备上运行的多个服务及设备状态:

模型中心	端云协同服务 > 我的本地设备								
我的模型	我的本地设备						点击收起		
创建模型	在本页面新增设备、联网激; 利用率等信息。	舌本地设备后,即可将「 <mark>我</mark> 的	的部署包」了	〔面发布成功的部署包一键下发	到设备上。设备联	网时,可以查看设备上部	署的服务、设备的运行状态、资源		
训练模型									
校验模型	新增设备								
发布模型	设备名称	创建时间	设备类型	应用平台	设备连接状态	最新同步时间	操作		
EasyData数据服务	123123 ©	2021-01-07 13:25	-		未激活		删除 激活设备 查看详情		
数据总览	showcase-test 💬	2020-12-15 17:57	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 19:22	激活设备 查看详情		
标签组管理	linux-x86-zqw-2 💬	2020-12-15 16:24	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 19:42	激活设备 查看详情		
在线标注	linux-x86-zqw 💮	2020-12-15 15:29	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-15 16:40	激活设备 宣看详情		
云服务数据回流	firefly_rk3399pro 💬	2020-12-15 14:39	-		离线 ⑦	2020-12-15 21:33	激活设备 查看详情		
EasyEdge本地部署	ubuntu-local-fjy 💬	2020-12-14 21:38	服务器	Linux-AMD64(x86-64)	在线	2021-01-07 15:01	查看详情		
纯离线服务	hfi-1 Θ	2020-12-14 19:24	服务器	Linux-AMD64(x86-64)	离线 ⑦	2020-12-16 14:18	激活设备 查看详情		
端云协同服务 ヘ	edge#f#	2020-12-11 14:52			去激活		删除 激沃设备 查看详情		

设备详情会展示当前设备的最新同步时间,以及CPU使用率、内存使用率等。服务列表则展示了当前设备上部署服务的运行情况和资源占用情况

0							
模型中心		端云协同服务 > 我的本地设备	> ubuntu-local-fjy				
我的模型		设备详情					
创建模型		设备名称 ubuntu-local-fjy	连接状态 在約	ŧ	实时刷新	OFF	
训练模型		设备类型 服务器	应用平台 Lin	ux-AMD64(x86-64)	最新同步时间 20	021-01-07 15:00	
校验模型		CPU使用率		內存使用率			
发布模型		31.1%		35.8%			
EasyData数i	据服务						
数据总览		端云协同服务详情					
标签组管理		服务名称	模型ID	CPU占比	内存使用情况	内存占比	操作
在线标注		1000.00	246	0.01%	156.7MB	0.93%	宣看服务配置
云服务数据回溯	£						
EasyEdge本	地部署						
纯离线服务							

₯智能边缘控制台-单节点版

₯ EasyEdge 智能边缘控制台-单节点版 IEC

EasyEdge Intelligent Edge Console (以下简称IEC) 是EasyEdge推出的边缘设备管理的本地化方案。可以运行于多种架构、多系统、多类型的终端之上。通过IEC,用户可以方便地在本地进行

- EasyDL/BML/EasyEdge的SDK的 离线 / 在线激活,服务管理
- 接入本地和远程摄像头,网页中实时预览
- 自动监控和记录相关事件
- 硬件信息的可视化查看

支持的系统+CPU架构包括:

- Windows x86_64 (Windows 7 ~ Windows 10, 暂不支持Windows 11)
- Linux x86_64 / arm32 / arm64

支持各类常见的AI加速芯片,包括:

- NVIDIA GPU / Jetson 系列
- Baidu EdgeBoard FZ系列
- 比特大陆 Bitmain SC / SE 系列
- 华为 Atlas 系列
- 寒武纪 MLU 系列

Baidu 百度智能云文档

• 其他EasyDL/EasyEdge/BML支持的AI芯片

完整列表可参考<mark>这</mark>里

Release Note

注意:2.0.0之后,默认以系统服务形式安装iec,无法兼容1.x版本的iec

版本号	发布时间	更新说明
2.2.0	2022-10-27	新增onvif/gb28181支持;完善端云通信逻辑
2.0.0	2022-03-22	支持连接中心节点IECC;支持以系统服务安装
1.0.2	2021-12-22	更新视频预览推流库;新增若干AI芯片支持;支持多种芯片温度、功耗展示;多项性能优化
1.0.0	2021-09-16	IEC 第一版!

快速开始

从这里选择您需要的操作系统和CPU架构下载:

- Windows amd64 : intel、AMD的64位x86_84 CPU
- Linux amd64: intel、AMD的64位x86_84 CPU
- Linux arm : 树莓派等32位的ARM CPU
- Linux arm64: RK3399、飞腾等64位的ARM CPU

或者从纯离线服务管理页可下载智能边缘控制台

达桨EasyDL ^خ ار <i>ز</i>	▲品介绍 操作平台 应用案例 使用文档
物体检测模型	三 纯离线服务
创建模型	
训练模型	纯离线服务说明
校验模型	发布纯高线服务,将训练完成的模型部署在本地,高线调用模型。可以选择将模型部署在本地的服务器,小型设备、软硬一体方案专项适配硬件上, 通过APL,SDK词一步集成,灵活该所不同业务场景;您可下载EasyEdoe智能边缘控制点,便像管理多个本地模型服务。
发布模型	世元·新冠来 20年二前制备 ↓ 工業20年前後的制备 ①
☑ EasyData数据服务	
数据总览	
标签组管理	服务器 通用小型设备 专项适配硬件
在线标注	SDK API
智能标注	
云服务数据回流	此处发布、下载的SDK为未授权SDK,需要前往控制台获取序列号激活后才能正式使用。SDK内附有对应版本的Demo及开发文档,开发者可参考源代码完成开发

您也可以通过先安装多节点版本IECC,通过中心节点来自动连接安装边缘节点。

Linux 安装 解压缩之后,目录结构如下

0 EasyEdge-IEC-v2.0.0-linux-amd64 > tree .	
├─── easyedge-iec └─── easyedge-iec-setup.sh └─── etc └─── easyedge-iec_service-conf init d	
i i easyedge-iec.service-conf.systemd i i easyedge-iec.service-conf.upstart i i easyedge-iec.service.yml i easyedge-iec.yml i readme.txt	
1 directory, 8 files	

以系统服务形式安装(推荐) 以root用户运行./easyedge-iec-setup.sh install 即可

[setup]: sudo could not be found [setup]: Start to install IEC... [setup]: + bash -c "cp easyedge-iec /usr/sbin/easyedge-iec" [setup]: + bash -c "chmod +x /usr/sbin/easyedge-iec" [setup]: + bash -c "cp etc/easyedge-iec.service.yml /etc/easyedge-iec/easyedge-iec.yml" [setup]: + bash -c "cp etc/easyedge-iec.service-conf.init.d /etc/init.d/easyedge-iec" [setup]: + bash -c "chmod +x /etc/init.d/easyedge-iec" [setup]: Install IEC success! [setup]: + bash -c "service easyedge-iec start" Starting easyedge-iec: success [setup]: Start to check IEC status... [setup]: + bash -c "curl -s 127.0.0.1:8702 >/dev/null" [setup]: IEC status: OK! [easyedge-iec]: default configure file: /etc/easyedge-iec/easyedge-iec.yml [easyedge-iec]: default log file: /var/log/easyedge-iec/easyedge-iec.log [easyedge-iec]: service usage: service easyedge-iec { start | stop } [setup]: Done!

- 日志: /var/log/easyedge-iec/easyedge-iec.log
- 系统配置: /etc/easyedge-iec/easyedge-iec.yml
- 服务启动/停止: service easyedge-iec { start | stop } (不同操作系统内可能不同,具体命令参考安装日志)

自定义安装 (不推荐) 自定义安装方法仅限于 安装脚本无法识别的情况。

- 拷贝 ./EasyEdge-IEC-v2.0.0/ 整个目录至自定义文件夹,如/opt/EasyEdge-IEC
- 进入到 /opt/EasyEdge-IEC
- 通过 nohup 等方法运行 ./easyedge-iec-linux-{您的系统架构} amd64: intel、AMD的64位x86_84 CPU arm:树莓派等32位的ARM CPU * arm64: RK3399、飞腾等64位的ARM CPU
- 日志: ./log/easyedge-iec.log
- 系统配置: ./easyedge-iec.yml

Windows 安装

解压缩之后,安装目录如下所示:

0 tmp2 > tree EasyEdge-IEC-v2.0.0-windows-amd64	
EasyEdge-IEC-v2.0.0-windows-amd64	
easyedge-iec.exe	
easyedge-iec-setup.bat	
etc	
easyedge-iec.yml	
└─── readme.txt	

1 directory, 4 files

打开命令行(非powershell)运行 easyedge-iec-setup.bat install。

如果遇到hang住的情况,可修改命令行配置 启动之后,打开	F浏览器,访问 http	://{设备ip}:8702	/easyedge/iec 即可:
ໝ Administrator: C:\Windows\system32\cmd.exe Microsoft Windows [Version 10.0.17763.2145] (c) 2018 Microsoft Corporation。保留所有权利。 C:\Users\eval>_	 还原(R) 移动(M) 大小(S) 最小化(N) 最大化(X) 	↑ 右键	×
	x 关闭(C) 编辑(E) > 默认值(D) 属性(P)		

3							
光标大小		命令记录					
● 小(S)		缓冲区大小(B):	50 🗘				
〇中(M)		經)由区数母(NI)·	4				
○大(L)		发/中凸数里(N);	· ·				
		□ 丢弃旧的副本(D)					
编辑选项							
□快速编辑模式	t(Q) 🔶 I	取消勾选					
☑插入模式(I)							
☑ 启用 Ctrl 键	夬捷方式(K)						
□ 粘贴时控洗剪	则标内容(F)						
	□将 Ctrl+Shift+C/V 用作复制/粘贴的快捷键(C)						
□将 Ctrl+Shift	t+C/V 用作复制	/粘贴的快捷键(C)					
□将 Ctrl+Shift 文本选择	t+C/V 用作复制	/粘贴的快捷键(C)					
 □将 Ctrl+Shiff 文本选择 ☑ 启用对选定内 	t+C/V 用作复制	/粘贴的快捷键(C)					
 □ 将 Ctrl+Shift 文本选择 ☑ 启用对选定内 ☑ 扩展的文本送 	t+C/V 用作复制 容自动换行(E) 5择键(X)	/粘贴的快捷键(C)					
□ 将 Ctrl+Shif 文本选择 ☑ 启用对选定内 ☑ 扩展的文本选 当前代码页	++C/V 用作复制 日容自动换行(E) 5择键(X)	/粘贴的快捷键(C)					
 ○ 将 Ctrl+Shift 文本选择 ○ 启用对选定内 ○ 扩展的文本送 当前代码页 65001 (UTF-8) 	++C/V 用作复制 回容自动换行(E) 起择键(X)	/粘贴的快捷键(C)					
 ○ 相次(1) 5022 g ○ 将 Ctrl+Shift 文本选择 ○ 启用对选定内 ○ 扩展的文本送 当前代码页 65001 (UTF-8) ○ 使用问题: 	++C/V 用作复制 四容自动换行(E) 5择键(X)	/粘贴的快捷键(C)					
 ○ 内容 Ctrl+Shift 文本选择 ○ 启用对选定内 ○ 扩展的文本送 当前代码页 65001 (UTF-8) 〕使用旧版控制: 	++C/V 用作复制]容自动换行(E) 5择键(X) 台(U)(需要重新版	/粘贴的快捷键(C) 自动, 影响所有控制台) 94年自					
 ○ 内容 Ctrl+Shift 文本选择 ○ 启用对选定内 ○ 扩展的文本送 当前代码页 65001 (UTF-8) 〕使用旧版控制: 了解有关归版 	++C/V 用作复制 的容自动换行(E) 站择键(X) 台(U)(需要重新派 控制台模式的详	/粘贴的快捷键(C) 自动,影响所有控制台) 细信息					
 ○ 内名 (1982年) ○ 将 Ctrl+Shift 文本选择 ○ 启用对选定内 ○ 扩展的文本送 当前代码页 65001 (UTF-8) 〕使用旧版控制: 了解有关目版 7解更多有关部: 	++C/V 用作复制 印容自动换行(E) 5择键(X) 台(U)(需要重新成 控制台模式的详	/粘贴的快捷键(C) 13动,影响所有控制台) 细信息					
 ○ 内名 1982年 ○ 将 Ctrl+Shift 文本选择 ○ 启用对选定内 ○ 扩展的文本送 当前代码页 65001 (UTF-8) 〕使用旧版控制: 了解有关回版 7解更多有关新生 	k+C/V 用作复制 如容自动换行(E) 好容自动换行(E) 好键(X) 台(U)(需要重新局 台(U)(需要重新局 拉制台模式的详	/粘贴的快捷键(C) 音动,影响所有控制台) 细信息					

启动之后,打开浏览器,访问 http://{设备ip}:8702/easyedge 即可:

О БЯВЖЕ I Ва ¹ ВХАВ		
EasyEdge 智能边缘控制台 EmpEdge和E快速要要4模型,可诺化完成模型服务domod存设、储备头车对指理、识别结果事件记录、超速完成本地数要与模型年联,快速体验模型效果	本地管理员登录	
	初始後号: admin 初始密码: assyndge 물국	
	¥hr-min	修改密码
相关产品: EanyOL率门临A开发平台 BAA全功能A开发平台 EanyEdga随与边缘A服务平台		版本号: V1.0.0

默认用户名密码为 admin / easyedge

EasyEdge 智能边线	缘控制 [·]	台			
总览					
⑦ 本地模型服务		┃ 基本信息			
添加模型服务		主机名 szth-aip-m12-tia	操作系统 linux MAC:	地址 6c:92:bf:46:82:47 P地均	10.232.191.23, 19
模型服务管理		联网状态 在线			
◎ 我的摄像头		┃ 资源利用情况			
添加摄像头					
摄像头管理					
🖆 本地事件		5.9%	57.8%	55.8%	
本地事件记录			1217 MA		
◎ 设置		CPU型号 Intel(R) X	运行内存 125.64GB	总空间 16.39TB	
系统设置		CPU核数 48核	已用内存 72.60GB	已用空间 9.15TB	
运行日志					
		▲ 本地项目管理			
		横型服务管理	▶ 序列号管理 >	摄像头管理	本地事件记录 >
		3 0 数量 运行中	0 0 _{数量 有效期内}	2 2 _{数量 运行中}	16742 ***

功能使用说明

①添加模型服务 首先,点击导航栏的「本地模型服务」-「添加模型服务」。在页面中定义服务名称后,将已经下载好的Linux/Windows版本的 SDK与IEC关联。关联完毕后可按两种激活方式,激活使用SDK。

EasyEdge 智能边缘控制台	操作指南	admin 🗸
总览 溺加棋型服务		
 ○ 木地機型服务 ▲ ● 新建模型服务 ● 振速模型服务 ● 服务名称 清编人名称 		
 4型服务管理 - SOK ◎ 1 法排SOK ● 数的描像头 ▲ 法第50K ● 数的描像头 ● 数的描像头 		
摄像关管理 强务上传SDK 服务发管理 最多配置 stort 0000		
本地事件记录 *Port 24401		
道行日志 和20H 8.4		
版本号:V1.0.0		

部分SDK需要提前安装系统依赖,如TRT等,具体请参考EasyDL/BML/EasyEdge SDK使用文件中的环境依赖安装说明

联网激活

1. 在关联SDK完成后,需要在百度智能云控制台对应部署方式管理页中新增测试序列号或购买正式序列号。(图中以服务器版SDK为例)

✓ 返回EasyDL总览	产品服务 / EasyDL图像 -	产品服务 / EasyOL图像 - 服务器线集线服务									
EasyDL图像	服务器纯离线服务	服务器纯真线服务管理									
公有云部署 🗸 🗸	API SD	ĸ									
EasyEdge本地部署 <											
• 服务器纯高线服务	使用说明: (展开)	近看部署介绍及操作步骤等)							×	✓ ₩.并	
• 设备端纯真线服务	+ 购买正式授权	新增测试序列号 ?							开发文档	工单支持	
。 专项硬件纯离线	设备名	技术方向	模型ID	可激活模型数	序列号类型	序列号	激活时间	到期日期	銀作		
	自定义设备 🧷	田僚		3	试用		未激活	2021-12-31	如何激活		
	自定义设备 之	图像		3	试用		未激活	2021-12-31	如何激活		
	自定义设备 🖉	图像		3	试用		未激活	2021-12-31	如何激活		

2. 再在IEC中填入所申请的序列号

总宽	波加坡型服务
③ 本地模型服务	新建模型服务
添加模型服务	* 服务名称 zy_测试
模型服务管理	* STW 0
◎ 我的摄像头	EssyEdge-Linux-m462366.zip
添加摄像头	※任先式 国の単写 家は筆述
摄像头管理	
🖹 本地事件	Model ID 46266
本地事件记录	
 设置 	
系统设置	· 服务配置 * Host 0.0.00
运行日志	
	*\$2.2
	质加升激活
版本号: V1.0.0	

- 3. 配置服务,在服务端口不冲突占用的情况下,使用默认即可
- 4. 添加并激活

离线激活

1. 在IEC总览页面下载「指纹文件」

EasyEdge 智能	边缘控制	冶			操作指南	admin ~
总览						
① 本地模型服务 添加模型服务 模型服务管理		基本信息 主机名 ———————— ————————————————————————————	操作系统 linux 指纹文件 下载指纹文件	MAC地址 所属用户 root		
② 我的摄像头 添加摄像头		资源利用情况				
摄像头管理 画 本地事件		6.4% ^{CPU利用集}	4.6% 通行内存	73.8%		
◆地學什比來 ② 设置 系统设置		CPU競号 Intel(R) X CPU核数 28核	运行内存 251.31GB 已用内存 11.56GB	总空间 13.87TB 已用空间 10.23TB		
运行日志						

2. 在百度智能云的控制台中找到SDK对应的管理列表,图中以服务器SDK为例。申请序列号后,点击对应序列号尾部的「离线激活」操作,按指

	< 返回EasyDL总宽	产品服务 / EasyDL图像 - 服	务器纯离线服务									
	EasyDL图像	服务器纯离线服务管	管理									
	公有云部署 ~	API SDK										
	EasyEdge本地部署 へ											
	• 服务器线离线服务	使用说明:(展开查看	f部署介绍及操作步骤等)								◇ 展开	
	。 设备端纯离线服务	+ 购买正式授权	新增测试序列号 ?							开发文档	工单支持	
	。 专项硬件纯高线	设备名	技术方向	模型ID	可激活模型数	序列号类型	序列号	激活时间	到期日期	操作		
可以此注		自定义设备 🖉	122-180		1	试用		2021-03-09	永久	申请延期		
51/啟/占		自定义设备 🖉	2518		10	试用		2021-09-16	2021-10-16	申请延期		
		自定义设备 🧷	田稼		10	试用	$(1,1) \in \mathcal{M}(\mathbb{R}^n)$	未激活	激活后30天内有效	联网激活 別	线激活	
		自定义设备 🖉	图像		10	试用		未激活	激活后30天内有效	联网激活 7	线激活	
		自定义设备 🖉	图像			试用		未激活	激活后31天内有效	联网激活丨制	线激活	
		自定义设备 🧷	田像			试用		未激活	激活后31天内有效	联网激活!#	1线激活	
									每页显示	10 💙 <	1 >	
												$^{\circ}$
												\sim

3. 在IEC的添加模型服务页面,上传下载好的授权文件,完成激活

EasyEdge 智能	边缘控制	旧台		操作指南	a) admin 🗸
总览		添加模型服务			
② 本地模型服务		新建模型服务	8		
添加模型服务		*服务名称			
模型服务管理					
◎ 我的摄像头			EasyEdge-Linux-m462x88.zip		
添加摄像头		激活方式	联网激活 高线激活		
摄像头管理			服务器版本激活流程		
🖾 本地事件			1. 在当前页面下载"挑放文件"并 2. 在百度智能完洁影纷上作指纹 3. 在百度智能完洁影纷上作指纹 3. 在百度智能完洁影纷上下数模纹 4. 在当前页面上语"操纹文件"并 つ思い上述の第一次第一次第一次第一次第一次第一次第一次第一次第一次第一次第一次第一次第一次第		
本地事件记录			LCRANOOR UNITHERSE ID Release ID X11 X10 (Kat 197)		
③ 设置			指設文件 下载相线文件 Mordal ID 46966 CI		
系统设置			Release ID 102355 🗇		
运行日志			* 授权文件 1. 选择授权文件		
		服务配置			
전상 1월 11일 - 12일 - 12일 2월 12일 - 1					
			参数配置		
			添加并撤送		

激活完成后即可在「模型服务管理」列表中启动服务,使用后续的操作栏功能。

体验本地demo

点击「本地demo体验」即可在立即上传图片进行预测

【物体检测】97741 e7.5检测高性能–Auto分辨率V1

正正を定める

接入摄像头

使用接入摄像头功能首先需要添加摄像头,请参考第②步,完成后按照第③步操作 **注:服务启动后也可参考「模型发布」模块的技术文档进行开** 发使用,本文档主要介绍IEC使用功能

EasyEdge 智能说	边缘控制	台					操作指南 🛛 admin 🗸
总览		模型服务管理					
③ 本地模型服务		新增服务 查看序列号					
添加模型服务							
模型服务管理		服务名称	模型名称	应用平台	服务状态	服务开关	操作
◎ 我的摄像头		晓宇加速版	e7.5检测高性能–Auto分 辨率V1	Linux-通用X86 CPU加速	运行中		查看详情 体验本地demo 接入摄像头 顯除
添加摄像头		zyf_胡乱测试模型	测试20210902V1	Linux-通用X86 CPU	运行中		
摄像头管理							
🖾 本地事件							
③ 设置							
系统设置							
运行日志							
版本号: V1.0.0							

激活错误码

错误码	文案	描述
4001	parameters missing. 参数缺失	
4002	parameters invalid. 参数不合法	
4003	model invalid. 模型信息不合法	
4004	no more activation left. 该序列号和该设备的激活次数超上限	
4005	the serial key is out of date. 该序列号过期	
4006	the serial key has been activated. 该序列号已被其他设备激活	该序列号已被其他设备激活,不能重复激活。
4007	account invalid. 序列号不能用于其他账号的模型	序列号不能用于其他账号的模型,只能用于绑定账号的模型。
4008	serial key invalid. 序列号不合法	序列号不存在或找不到
4009	bundle id invalid. 包名不合法	
4010	product invalid. 产品不合法	如easydl的SDK使用BML的序列号来激活,会报该错误
4011	platform invalid. 平台不合法	
4012	activate too frequent. 激活太频繁	激活太频繁,请稍后再进行激活。
4013	device type and license type not match. 硬件类型和序列号类型不匹配	如使用加速版序列号激活基础版SDK会报该错误
4014	exceed max activate device num. 超过最大授权数量	
4015	technology invalid. 技术类型不合法	
4016	exceed max activate entity num. 超过最大模型数量	
4017	device invalid. 设备不合法	
4018	model invalid. 模型不合法	

②添加摄像头 导航栏点击「我的摄像头」-「添加摄像头」,定义摄像头名称、备注后即可添加摄像头。支持本地摄像头和网络摄像头。 摄像头 添加成功后即可设置摄像头的运行时间和频率

EasyEdge 智能边缘搭	空制台	操作指南	a admin ~
总宽	酒加酒像头		
 ○ 本地模型服务 > 添加模型服务 - 模型服务管理 	• 通春头会称 请输入团像头合称 佛社 课师入周僚头告注		
○ 我的摄像头 へ 添加摄像头	▲ 國像头类型 ● 本地攝像兵 ○ 网络摄像兵 请这时想像头		
振程头管理 / 《 本地事件记录 / 》 。 设置 / 《 无机电件记录 / 《 无机设置 / 《 近行日志			
版本号: V1.0.0			

③摄像头接入模型服务预测点击「本地模型服务」-「模型服务管理」中,所需接入预测的服务的「接入摄像头」

EasyEdge 智能边	边缘控制;						操作指南	a admin ~	
总宽		模型服务管理							
③ 本地模型服务		新提服务							
添加模型服务									
模型服务管理		服务名称	模型名称	应用平台	服务状态	服务开关	操作		在弹出的弹窗中选择第2
◎ 我的摄像头		晓宇加速版	e7.5检测高性能Auto分 辨率V1	Linux-通用X86 CPU加速	运行中		查看详情 体验本地demo 接入摄像头		
添加摄像头		zyf_胡乱测试模型	测试20210902V1	Linux-通用X86 CPU	运行中				
摄像头管理									
🗈 本地事件									

步中添加的摄像头,此时点击确认即可在「摄像头管理」中的实时预览功能中查看摄像头预测结果,识别结果默认不保存。如需保存识别结果, 可设置对应的「本地事件触发条件」,根据标签和置信度,将识别结果保存至本地事件记录当中。设置多个标签条件时,IEC会以"或"的逻辑来将 所有满足条件的识别结果保存

接入摄像头			
某个ip-camera	~ 添加		
摄像头名称	本地时间触发条件	52%	操作
某个ip-camera	标签 tomato ✓ 置信度 + 新増一项	52 % ~ 100% ×	取消接入
	确定	取消	

④本地事件 点击导航栏「本地事件记录」,可通过服务名称、摄像头名称、事件记录的时间、标签及置信度来筛选识别结果查看,多个标签及置 信度同样也是"或"的逻辑记录。如有想要删除的事件数据可选择后删除,全选为本页全选。

EasyEdge 智能边缘控制	旧台				操作指南	admin ~
总览	本地事件记录					
⑦ 本地模型服务 ^ 添加模型服务 模型服务管理	选择服务 请选择服务 >>	选择摄像头 请选择摄像头	✓ 时间按图 开始	★ 结束		
 ② 我的摄像头 ^ 添加摄像头 摄像头管理 二 本地事件 ^ 	第件概范 事件自动制新 ○ ●	Sec.	00		**	(1) (1) (1) (1) (1) (1) (1) (1) (1) (1)
本地事件记录 ③ 设置	除于-加速版 - 某个ip-camera 荔枝 3 4 2021-09-16 13:30:17	晚宇-加速版 - 其个ip-camera 荔枝 2 2021-09-16 13:39:06	终于—加过版 — 某个ip-camera 荔枝 1 2021-09-16 13:38:56	院宇-加退版 - 某个Ip-camera 蘇枝 2 2021-09-16 13:38:47	戦学-加速版 - 某个ip-camara 	
运行日志	(株子-323版 - 東个)a-camera 高板 5		成子- 加速版 - 米个ら-camera 展在 3			
版本号: V1.0.0						46 >

⑤ 连接到智能边缘控制台-多节点版(IECC)

与中心节点连接之后,边缘节点主程序版本会自动随控制中心版本升级。(>2.0.0)

● Step 1 在IECC中添加边缘节点,选择「边缘节点已安装IEC」,并记录IP地址与Token

EasyEdge 智能边缘控制	1台—多节点版
总览	边缘节点 > 添加边缘节点
◎ 模型服务中心 模型服务管理 下发模型服务	添加边缘节点 ◇ 注册边缘节点 2 激活边缘节点
	激活方式 ○ 边缘节点未安装IEC
摄像头设备	请复制中心节点IP地址、Token密码,在迎嫁设备安装的 IEU—系统设置贝面 进行激活,操作元成后端在下力进行极短,极强通过即成功激活边缘节点。
回 事件中心 事件记录	IP地址 10.232.191.23.8602
	10k8n DUU25803-6044-4696-6400-680003089410
系统设置	・ 上一步 校验激活結果
运行日志	

• Step 2 在IEC的系统设置中打开多节点模式,并填入刚才记录的IP地址与Token,点击建立连接

EasyEdge 智能边缘控制	则中心
总览	系统设置
✿ 本地模型服务 添加模型服务 模型服务管理	*事件保存时限 30 <a>>> → 天 *文件保存路径 /opt/baidu/easyedge/iec/events
	系统时间显示 ③
添加摄像头 摄像头管理	多节点模式 ③
三本地事件 本地事件记录	中心节点连接状态 • 未连接 (IP地址和Token,请在"多节点版"的"添加边缘节点"第二步中宣看) 输入IP地址
运行日志	

• 连接完成后即可在中心节点IECC去监控/管理/应用在边缘节点上的IEC

配置项*

配置文件etc/easyedge-iec.yml中有关于IEC的各项配置说明,一般无需修改,请确保理解配置项含义之后,再做修改。

```
###### IEC系统配置
###### ------ 高级配置一般无需修改 -----
###### !!!注意!!! 请确保理解配置项含义后再做修改
version: 3
com:
# hub: 作为中心节点模式启动。 edge: 作为子节点启动
# role: edge
#硬件利用率刷新时间间隔: 过低的刷新间隔可能会消耗CPU资源。
hardwareUsageRefreshSecond: 60
#事件监测触发扫描周期
eventTriggerIntervalSecond: 10
# IEC保存SDK等资源的路径:可填写 "default", 也可以直接填写绝对路径
appDataFolder: default
# 是否开启DEBUG模式:开启之后,将会打印大量日志,便于追踪系统问题
debug: no
# 是否需要用户名/密码登陆,强烈建议打开!(默认用户名/密码为 admin/easyedge)
loginEnabled: yes
#登录有效时间,单位秒
sessionMaxAge: 3600
logging:
# 是否把日志打印到控制台
toStd: no
# 是否把日志输出到文件。关闭后,将无法在页面中正确显示系统日志
toFile: yes
loggingFile: /var/log/easyedge-iec/easyedge-iec.log
# 0:info; -1:debug; -2:verbose
level: -1
webservice:
# WEB服务的监听端口
listenPort: 8702
listenHost: 0.0.0.0
sdk:
# GPU SDK所使用的cuda版本:9/10/10.2/11.0/11.1。请安装完cuda之后,这设置正确的版本号。
cudaVersion: 10.2
# AI服务启动时,额外配置的 LD_LIBRARY_PATH(linux) 或者 PATH(windows)
libPath: ./
# AI服务启动时,额外配置的其他环境变量。
ENVs:
 EDGE_CONTROLLER_KEY_LOG_BRAND: EasyEdge
```

FDGE CONTROLLER KEY XXX XXXX

commu: # 普通消息等待respond的超时时间 respondWaitTimeoutSecond: 2 ###### 数据库相关配置 db: sqliteDbFile: /var/lib/easyedge-iec/easyedge-iec.db hubDbFile: /var/lib/easyedge-iec/easyedge-iec.hub.db eventDbFile: /var/lib/easyedge-iec/easyedge-event.db fileServerDbFile: /var/lib/easyedge-iec/easyedge-fileserver.hub.db nodeMonitorDbFile: /var/lib/easyedge-iec/easyedge-nodemonitor.hub.db ###### 推流相关配置 mediaserver: flvPort: 8715 rtmpPort: 8716 ###### 视频流相关配置 edgestream: logLevel: -1 listenHost: 127.0.0.1 listenPort: 8710 #摄像头预览:识别结果绘制延迟消失 renderExtendFrames: 10 # 预测队列大小: 如果设置为60,当摄像头fps=30时,视频延迟约为2秒。 降低inferenceQueueSize可以降低预览延迟,但是根据硬件的算 力情况,可能导致模型推理速度跟不上,没有识别结果,不建议设置太低 inferenceQueueSize: 60 videoEncodeBitRate: 400000 #视频采样 & 视频实时预览分辨率设置 # 0: auto, 1: 1080p, 2: 720p, 3: 480p, 4: 360p, 5: 240p resolution: 0 # 内置多媒体服务配置 # port设为0表示关闭 mediaServerHost: 127.0.0.1 mediaServerFlvPort: 8713 mediaServerRtmpPort: 8714 mediaServerRtspPort: 0

FAQ 启动服务后,进程中出现两个easyedge-iec进程这是正常现象,IEC通过守护进程的方式来完成更新等操作。

启动服务时,显示端口被占用port already been used 通过修改 easyedge-iecc.yml文件的配置后,再重新启动服务。

安装服务时,报错permission denied 请以管理员身份运行安装程序。

中心节点重启后,边缘节点IEC 一直离线中心节点短时间的离线,边缘节点会自动重连。如果中心节点已经恢复在线,边缘节点长时间未自动连接上,可通过边缘节点iec的方法来重新连接(右上角 admin - 重启系统)

IEC 是否有Android / iOS 版本 我们将会在近期发布对Android操作系统的支持

添加SDK时,报错 SDK不支持该硬件。 SDK not supported by this device 一般是因为使用的SDK跟硬件不匹配,如 GPU的SDK,硬件没有GPU 卡。 对于Jetson,也可能是Jetpack版本不支持,可以通过查看 本机Jetpack版本和SDK支持的Jetpack版本列表(cpp文件中的文件名来查看)来 匹配。

⊙ 智能边缘控制台-多节点版

₯ EasyEdge 智能边缘控制台──多节点版

整体介绍

智能边缘控制台 - 多节点版(EasyEdge Intelligent EdgeConsole Center 以下简称IECC),是EasyEdge推出的边缘资源管理、服务应用与管理一站 式本地化方案。

通过IECC,用户可以方便地在中心节点管理子节点:

- 边缘硬件资源的管理与监控
- EasyDL/BML/EasyEdge的SDK的 离线 / 在线激活,服务管理

Baidu 百度智能云文档

- 视频流解析,接入本地和远程摄像头,网页中实时预览
- 自动监控和记录相关视频流推理事件

支持的系统+CPU架构包括:

- Windows x86_64 (Windows 7 ~ Windows 10,暂不支持Windows 11)
- Linux x86_64 / arm32 / arm64

支持各类常见的AI加速芯片,包括:

- NVIDIA GPU / Jetson 系列
- Baidu EdgeBoard FZ系列
- 比特大陆 Bitmain SC / SE 系列
- 华为 Atlas 系列
- 寒武纪 MLU 系列
- 其他EasyDL/EasyEdge/BML支持的AI芯片



连接说明 以下为 中心节点 (控制中心),边缘节点/子节点,摄像头的连接示意:

其中:

- 控制中心需要有固定IP,而边缘节点可以处于多级子网之下,只需IEC能够主动访问到控制中心节点即可
- 模型服务均运行于各边缘节点之上
- 摄像头均与边缘节点相连

Release Note

版本号	发布时间	更新说明
2.2.0	2022-10-27	边缘节点新增Android支持;新增onvif/gb28181支持;优化端云通信通道安全
2.0.0	2022-03-25	多节点版上线!
1.0.2	2021-12-22	更新视频预览推流库;新增若干AI芯片支持;支持多种芯片温度、功耗展示;多项性能优化
1.0.0	2021-09-16	智能边缘控制台 - 单节点版 IEC 第一版!

安装 从这里选择您需要的操作系统和CPU架构下载:

- Windows amd64 : intel、AMD的64位x86_84 CPU
- Linux amd64: intel、AMD的64位x86_84 CPU

- Linux arm:树莓派等32位的ARM CPU
- Linux arm64: RK3399、飞腾等64位的ARM CPU

或者从纯离线服务管理页可下载智能边缘控制台

<i>ギキレ</i> ズ楽EasyDL	产品介绍 操作平台 应用案例 使用文档	
物体检测模型		
创建模型		
训练模型	纯离线服务说明	
校验模型	发布纯高线服务,将训练完成的模型部署在本地,高线调用模型。可以选择将模型部署在本地的服务器、小型设备、软硬一体方案专项适配硬件上, 通过ABL、CPU装,使用。目示还由工作从各场局、使可工作LeanGatementshiphing的一种。使得使用多个大地使用服务。	
发布模型	通过APT、3UT发生少来成,火焰迫进个问业分词原,影响「Micasycouyotalept发体性的日,使其管理多一种必须呈成分。	
🖾 EasyData数据服务	发布新融资 督能式控制台 业 下就省能比排控制台 ①	
数据总览		
标签组管理	服务器 通用小型设备 专项适配硬件	
在线标注	SDK API	
智能标注		
云服务数据回流	此处发布、下载的SDK为未侵权SDK,需要前往控制台获取序列号激活后才能正式使用。SDK内附有对应版本的Demo及开发文档,开发者可参考源代码完成开发,	

以Linux为例,解压缩后目录结构如下所示:

- ./EasyEdge-IECC-v{版本号}/
- -- easyedge-iecc
- |-- easyedge-iecc-setup.sh
- |-- etc/
- |-- etc/easyedge-iec.yml
- |-- readme.txt

Linux 系统

通过系统服务形式安装 (推荐)

以管理员运行bash easyedge-iecc-setup.sh install 即可。



出现sucess字样,表示安装成功。

- 日志: /var/log/easyedge-iecc/easyedge-iecc.log
- 系统配置: /etc/easyedge-iecc/easyedge-iecc.yml
- 服务启动/停止: service easyedge-iecc { start | stop } (不同操作系统内可能不同,具体命令参考安装日志)
- 配置服务自启动:可根据不同操作系统参考这里进行对应配置

可通过 bash easyedge-iecc-setup.sh uninstall 来卸载,以及bash easyedge-iecc-setup.sh upgrade来升级为当前安装包的版本

自定义安装 (不推荐)

自定义安装仅限于 安装脚本无法识别您的操作系统的情况。

- 拷贝 ./EasyEdge-IEC-v2.0.0/ 整个目录至自定义文件夹,如/opt/EasyEdge-IEC
- 进入到 /opt/EasyEdge-IEC
- 通过 nohup 等方法运行 ./easyedge-iec-linux-{您的系统架构} --com.role=hub amd64: intel、AMD的64位x86_84 CPU arm:树莓派等32位的 ARM CPU * arm64: RK3399、飞腾等64位的ARM CPU
- 日志: ./log/easyedge-iecc.log
- 系统配置: ./easyedge-iecc.yml

Windows 系统 打开命令行(非powershell)运行 easyedge-iecc-setup.bat install。

注:如果遇到hang住的情况,可修改命令行配置

	strator: C:\\	Windows\sy	stem32\cm	d.exe				-	×
rosoft) 2018	t Windows Microsof \eval>_	[Versio t Corpor	on 10.0.1 Pation。 🖗	7763.2145] 张留所有权利。	 还原(F 移动(N 大小(S 最小(A) 	R) M) 5) E(N)	右键		
					□ 最大化 x 关闭((编辑(E 默认值 属性(F	と(X) C) E) > 直(D)			
"C:\\	Nindows	\system	32\cmd.	exe" 属性	>	×			
匹坝	子体	布向	颜色						
光	辰大小			命令记录					
•	小(S)			缓冲区大小(B):	50 🔹				
0	中(M) 大(L)			缓冲区数量(N):	4 🔶				
0.				□ 丢弃旧的副本(D)					
编 (1)	貴选项 快速编辑	莫式(Q)	←	取消勾选					
	插入模式(中田 Ctrl	(I) *******	=t(k)						
	粘贴时筛	先剪贴板	内容(F)						
	将 Ctrl+S	hift+C/	/用作复制	引/粘贴的快捷键(C)					
文	本选择								
	启用对选	定内容自	动换行(E)						
	扩展的文	本选择键	(X)						
1	前代码贝	. 01							
65	前代码贝 001 (UTF	-8)							
口使	E用旧版控 了解有关L	制台(U)(1版控制台	(需要重新 合模式的词	启动, 影响所有控制台) キ细信息					
了解	更多有关	新控制台	功能的信						
了解	更多有关	新控制台	功能的信	ê.					

验证安装:启动之后,打开浏览器,访问 http://{设备ip}:8602/easyedge 即可:



更新服务:关闭服务,下载最新的安装包,重新执行安装流程即可。

- 注:1.中心节点更新到新版之后,已连接的边缘节点会自动跟随中心节点,自我升级到同样的版本。
- 2. 报错: Text file busy. 一般是因为服务没有停止。

使用流程 Step 1 注册并激活边缘节点

• 在IECC导航栏中点击边缘节点,点击页面中的添加边缘节点按钮

EasyEdge智能边缘控制台-多节点版								操作指南 🧧 admi	
总览	边缘节点								
⑦ 模型服务中心 模型服务管理	已添加的边缘节点							添加边缘节点	请输入边缘节点名称
下发模型服务	边缘节点名称	节点状态	节点标识	CPU	AI芯片	模型服务数量	摄像头数量	服务添加时间	操作
◎ 边缘资源中心	x86-test	 在线 		Intel(R) Core(T					
边缘节点 摄像头设备		● 离线	dry_persona L_cpu	Intel(R) Xeon(R				2022-03-23 19:30:00	查看详情 删除

- 注册边缘节点,填写基本信息
- 激活边缘节点,根据边缘节点上是否安装智能边缘控制台-单节点版(IEC)分两种激活方式
 - 边缘节点未安装IEC:复制提供的命令,在边缘节点的终端中输入执行(命令会自动在当前目录,下载单节点版IEC并注册到控制中心)。终端命令执行完成后,在下方校验激活结果,如结果通过即可完成边缘节点的激活

添加边缘节点	
✓ 注册边缘节点 2 激活边缘节点	
激活方式 ○ 边缘节点未安装IEC ○ 边缘节点已安装IEC	
请复制以下命令,粘贴在边缘设备terminal执行进行激活;在边缘节点操作完成后可在下方进行校验,校验通过即成功激活边线	象节点。
Windows Linux powershell (new-object Net.WebClient).DownloadFile('http://10.187.123.18:8702/easyedge/fileserver/v1/download/e	٥
asyedge-iec-setup.bat','.\easyedge-iec-setup.bat') && .\easyedge-iec-setup.bat install_from_iecc 10.187.123.18.8 702 94766edf-868c-46a1-9549-bc27b760b17a	
上一步 校验激活结果	

• 边缘节点已安装IEC:记录页面中提供的IP地址和Token

添加边缘节	i.a.	
>> 注册边线	8 节点 ————————————————————————————————————	
激活方式	○ 边缘节点未安装IEC ● 边缘节点已安装IEC	
	请复制中心节点IP地址、Token密码,在边缘设备安装的"IEC-系统设置页面"进行激活;操作完成后请在下方进行校验,校验过	通过即成功激活边缘节点。
		a
	IP地址 10.187.123.18:8702	
	Token 94766edf-868c-46a1-9549-bc27b760b17a	
	上一步 校验激活结果	

• 在边缘节点的IEC·系统设置中,打开多节点模式开关,将刚才记录的IP地址和Token填入其中,建立连接

总览	系统设置
⑦ 本地模型服务 添加模型服务	
模型服务管理	*文件保存路径 /opt/baidu/easyedge/iec/events
	系统时间显示 ③
添加摄像头	多节点模式 🔍 💽
摄像头管理	
	中心节点连接状态 • 未连接 (IP地址和Token, 请在"多节点版"的"添加边缘节点"第二步中查看)
本地事件记录	· 输入ⅣP地址 10.232.191.23:8602
	Token 🔍
系统设置	建立连接
运行日志	

• 成功激活后可在边缘节点页面中看到一行状态为在线的记录

EasyEdge智能边缘控制	台–多节点版								操作指南	admin 🗸
总览	边缘节点									
○ 模型服务中心 模型服务管理	已添加的边缘节点							添加边缘节点		
下发模型服务	边缘节点名称	节点状态	节点标识	CPU	AI芯片	模型服务数量	摄像头数量	服务添加时间	操作	
🖯 边缘资源中心	docker-x86	• 在线		Intel(R) Xeon(R		3	1	2022-03-23 19:25:36	查看详情 删除	t

Step 2 上传并下发模型服务

• 在模型服务管理-已添加的模型服务页面中点击添加模型服务

EasyEdge智能边缘控制	EasyEdge智能边缘控制台-多节点版 境作指摘							
总览	模型服务管理							
⑦ 模型服务中心 模型服务管理	已下发的模型服务 💽	已添加的模型服务				添加模型服务		
下发模型服务	服务名称	模型名称	应用平台	已部署节点数量	添加时间	操作		
日辺線資源中心 1000年上	cpu 加速版模型	e7.5检测高性能Aut	Linux	1	2022/03/23 19:24:49	下发至边缘 删除		

● 上传来自于EasyDL/BML的SDK,目前仅支持Windows/Linux的SDK

EasyEdg	EasyEdge智能边缘控制台–多节点版								
总览		模型服务管理	> 新建模型服务						
☆ 模型服务		添加模型服务	<u>z</u>						
模型服务	管理	★ 肥冬夕む	违 输) 夕称						
下发模型	服务								
🖯 边缘资源		* SDK 🕐	♪ 选择SDK						
边缘节点			添加	取消					
摄像头设	备								

- 添加成功后可在已添加的模型服务页面查看添加的模型服务SDK
- 在模型服务SDK上传成功以及边缘节点也添加激活过后,即可将模型服务下发至边缘。点击导航栏-下发模型服务,选择已添加的模型服务, 选择下发的目标节点(支持多节点批量下发)进行模型服务下发

EasyEdge智能边缘控制	11台-多节点版	操作指南 🔹 admin 🗸
总览	• EFTRER (?) ILFRIGHT	
⑦ 模型服务中心 模型服务管理 下发模型服务	 · (pu 加速板模型 へ) · (pu 加速板模型 へ) · (pu 加速板模型 · (pu 加速板 · (pu 加速 · (pu mu ·	
□ 边缘资源中心 边缘节点 摄像头设备	节点名称 接接状态 标识 节点名称 连接状态 桥识 docker-x80 - 石线 -	
□ 事件中心 事件记录 ○ 设置 5.6600零		

确定下发配置后,填入模型服务在边缘节点联网激活运行的序列号(支持批量导入)即可完成模型服务下发,序列号可在智能云控制台获取。
 离线激活的过程可参考IECC中的具体指引

EasyEdge智能边缘控制	台-多节点版	操作指南 🛛 admin 🗸
总宽	模型服务管理 > 下发模型服务	
Q 模型服务中心 模型服务管理 下发模型服务	下发模型服务 ② 私先下发配置 3 服务表示配置	
□ 边缘资源中心	激活方式 O 联网激活 〇 高线激活	
边缘节点 摄像头设备	边缘节点1: docker-x86	
回 事件中心		

• 完成上述流程后即可在模型服务管理-已下发的模型服务列表中查看记录,并进行下一步应用功能体验

注:完成此步骤后即可在边缘节点进行二次集成已下发的模型服务,具体的集成方式可在文档-某图像任务类型-模型发布中查找对应的 SDK开发文档进行集成开发

EasyEdge智能边缘控制台	計─多节点版								操作指南	a) admin 🗸
	模型服务管理									
	已下发的模型服务	已添加的模型服务						下发模型服务		
下发模型服务	服务名称	模型名称	服务所在边缘节点	应用平台	服务状态	添加时间	服务开关	操作		
 边缘资源中心 边缘节点 	CPU Linux 普	e7.5检测高性能			 运行中 					

下发时可以通过高级配置设置服务运行的host和port。若不设置,默认host为0.0.0.0,port为系统随机分配的可用端口

Step 3 配置摄像头

Step 3 - 5 描述的是如何使用IECC可视化进行视频流式推理与应用,对此有需求的用户建议详细查看后续步骤内容。如仅需对下发的模型服务进行二次集成的用户无需进行后续操作,参考SDK对应的开发文档进行集成即可

• 首先需要确定边缘节点已经接入物理摄像头,可通过USB插口接入,也可通过RTSP/RTMP流式协议接入。在摄像头设备页面点击添加摄像头 按钮,填写对应的信息添加摄像头。支持设置摄像头的运行时间以及摄像头的抽帧频率

EasyEdge智能边缘控制	台多节点版		操作指南	ə admin v						
	摄像头设备 > 添加摄像	描像头设备 > 添加播像头								
 ○ 根型服务中心 根型服务管理 下发模型服务 □ 边缘资源中心 辺缘等点 損優先设备 	添加摄像头 * 攝像头名称 * 选择所属边缘节点 绑定标识									
四 事件中心. 事件记录	备注摄像头类型	(1)输入通费头面注 ○ 本地類像头 ● 网络摄像头								
○ 设置 系统:包置 运行日志		- 従其遺地社 支持FFSP, FTMP, 22 mp/ladmin:1234g 150 - 运行时间 000000 単 000000 + 前編-現 描述频率 毎月 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1								

• 添加完成后可在摄像头设备页面查看记录

EasyEdge智能边缘控制	台-多节点版					操作指南	a admin ∨
总览	摄像头设备						
② 模型服务中心 模型服务管理	已添加的摄像头				添加摄像头		
下发模型服务	摄像头名称	摄像头标识	连接状态	所在节点名称	摄像头开关	操作	
⊜ 边缘资源中心	ip cam		 运行中 	docker-x86		实时预览 查看详情 本地	事件记录 删除

• 点击预览可查看摄像头预览画面

摄像头实时预览			×
原始摄像头			
2022-03-23 42 .3 7:36	预测标签	置信度 🍦	
	_		
		-	
		己置模型后查看	

Step 4 模型服务接入视频流预测

模型服务可接入摄像头直接进行预测,并可同时设置告警规则,出发告警条件的结果将会以事件的形式保存至IECC中。
 点击模型服务管理页面中对应服务的接入摄像头操作

EasyEdge智能边缘控制	台–多节点版							操作指南 🛛 ə admin 🗸
总览	模型服务管理							
⑦ 模型服务中心 模型服务管理	已下发的模型服务	已添加的模型服务						下发模型服务 训输入模型服务名称 Q
下发模型服务	服务名称	模型名称	服务所在边缘节点	应用平台	服务状态	漆加时间	服务开关	操作
 边缘节点 	CPU Linux 普	e7.5检测高性能						查看详情 体验demo 接入摄像头 删除

 将已添加至IECC的摄像头与模型服务关联,并在下方设置对应的事件告警条件。告警规则通过标签阈值的方式来建立,例如设置"猕猴桃"标签 阈值80%-100%,则大于80%置信度的"猕猴桃"识别结果将会保存至事件记录中

	接入摄像头			
服务名称	请选择摄像头	· 添加 注意: 仅可选择模型服务所在边缘节点的摄像头		
CPU Linux 普.	ip cam		操作	
		标签 kiwi ◇ 【信度 ● 0 % ~ 100% ×		
	ip cam	标签 请选择标签 ∨ 置信度 ● 0 % ~ 100% ×		
		· · · · · · · · · · · · · · · · · · ·	取 消	

• 也可在摄像头设备页面-实时预览中查看实时的模型服务预测结果

摄像头	摄像头实时预览				
称	原始摄像头 CPU Linux 普通版				
			预测标签	置信度 🍦	
	Lamater Distance	2022-03-23 12:52:14	2. tomato	96.74%	
	TITET & BACKET		3. kiwi	96.73%	
	tomato: 0,715030 tomato: 0,891233 tomato: 0,891233		1. 荔枝	96.23%	
	PETTER AUSTRAL		2. tomato	95.24%	
			2. tomato	94.90%	
			2. tomato	91.59%	
			2. tomato	90.72%	

Step 5 视频事件告警

• 可在事件中心-事件记录中查看满足时间告警条件的图片记录

EasyEdge智能边缘控制	台-多节点版	操作指南	a) admin 🗸
总览	本地事件记录		
⑦ 模型服务中心 模型服务管理	选择边接节点 dooker-x86 ~ 选择服务 请选择服务 ~ 选择服务 , 请选择服务夫 ~ 时间范围 开始 -	结束	
下发模型服务	标签课选 ~		
□ 边缘资源中心	事件概変 事件自动時所 の 💽	🗌 全选	₫ 删除
边缘节点 摄像头设备			
回事件中心			
事件记录			

高级配置说明 在系统设置 - 高级,可以修改控制中心的高级系统配置

IECc系统配置 version: 3 com: # hub: 作为中心节点模式启动。 edge: 作为子节点启动 role: hub #硬件利用率刷新时间间隔: 过低的刷新间隔可能会消耗CPU资源。 hardwareUsageRefreshSecond: 60 # IECC保存SDK等资源的路径:可填写 "default",也可以直接填写绝对路径 appDataFolder: default # 是否开启DEBUG模式:开启之后,将会打印大量日志,便于追踪系统问题 debug: no # 是否需要用户名/密码登陆,强烈建议打开!(默认用户名/密码为 admin/easyedge) loginEnabled: yes #登录有效时间,单位秒 sessionMaxAge: 3600 logging: # 是否把日志打印到控制台 toStd: no # 是否把日志输出到文件。关闭后,将无法在页面中正确显示系统日志 toFile: yes loggingFile: /var/log/easyedge-iecc/easyedge-iecc.log # 0:info; -1:debug; -2:verbose level: -1 webservice: #WEB服务的监听端口 listenPort: 8602 listenHost: 0.0.0.0 commu: mqServer: host: 0.0.0.0 port: 8632 HTTPPort: 8620 maxPayload: 8388608 pingIntervalSecond: 30 # 普通消息等待respond的超时时间 respondWaitTimeoutSecond: 2 nodeRefreshIntervalSecond: 30 ###### ------- 以下高级配置一般无需修改 -----------###### !!!注意!!! 请确保理解配置项含义后再做修改 ###### 数据库相关配置 db: sqliteDbFile: /var/lib/easyedge-iecc/easyedge-iecc.db hubDbFile: /var/lib/easyedge-iecc/easyedge-iecc.hub.db eventDbFile: /var/lib/easyedge-iecc/easyedge-event.db fileServerDbFile: /var/lib/easyedge-iecc/easyedge-fileserver.hub.db nodeMonitorDbFile: /var/lib/easyedge-iecc/easyedge-nodemonitor.hub.db ###### 推流相关配置 mediaserver: flvPort: 8613 rtmpPort: 8614 ###### 文件服务器相关配置

fileserver: root: /var/lib/easyedge-iecc/fs

FAQ

启动服务后,进程中出现两个easyedge-iec进程这是正常现象,IEC通过守护进程的方式来完成更新等操作。

启动服务时,显示端口被占用port already been used 通过修改 easyedge-iecc.yml文件的配置后,再重新启动服务。

安装服务时,报错permission denied 请以管理员身份运行安装程序。

添加SDK时,报错 SDK不支持该硬件。 SDK not supported by this device 一般是因为使用的SDK跟硬件不匹配,如 GPU的SDK,硬件没有GPU 卡。对于Jetson,也可能是Jetpack版本不支持,可以通过查看 本机Jetpack版本和SDK支持的Jetpack版本列表(cpp文件中的文件名来查看)来 匹配。

常见问题

心数据相关问题

需要上传多少张图片才能训练出效果较好的模型?

• 每种要识别的物体在所有图片中出现的数量需要大于50。如果某些要区分的物体具有相似性,需要增加更多图片。

上传图片的总量有限制吗?

- 每个账号下所有数据集的图片总数不能超过10万张。
- 心训练相关问题

数据处理失败或者状态异常怎么办?

 如是是图像分类模型上传处理失败,请先检查已上传的分类命名是否正确,是否存在中文命名、或者增加了空格;然后检查下数据图片量是否 超过上限(10万张);再检查图片中是否有损坏。如果自查没有发现问题,请在百度智能云控制台内提交工单反馈

模型训练失败怎么办?

• 如果遇到模型训练失败的情况,请在百度智能云控制台内提交工单反馈

已经上线的模型还可以继续优化吗?

 已经上线的模型依然可以持续优化,操作上还是按照标准流程在训练模型中-选择要优化的模型和数据完成训练,然后在模型列表中更新线上 服务,完成模型的优化

【图像分类】	百美椅子训	练 模型ID:23()			全部版本 删除
应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作
云服务	V2	训练完成	未申请	未发布	top1准确率87.61% top5准确率100.00% 完整评估效果	申请发布 校验 训练
离线SDK	V1	训练完成	未申请	未发布	top1准确率85.84% top5准确率100.00% 完整评估效果	申请发布 训练
						每页显示 12 🗸 🖌 1

点击我的模型列表——找到新训练好的模型版本——点击申请发布





② 模型效果相关问题

图像分割模型如何正确标注?

- 所有图片中出现的目标物体都需要被标出 (标注可以重叠)
- 标注应包含整个物体,且尽可能不要包含多余的背景
- 如果图片中存在很多相同标签的目标物体,可以使用右侧的锁定按钮。锁定标签后,只需要在左侧标注目标物体即可,不用再重复选择标签

如何通过「完整评估结果」里的错误示例优化模型?

- 错误示例中,左侧是正确的结果,右侧是模型的识别结果
- 观察模型识别有误的图片有哪些共同点,并有针对性地补充训练数据。比如:当图片比较亮的时候模型都能识别正确,但比较暗的时候模型就 识别错了。这时就需要补充比较暗的图片作为训练数据

我的数据有限,如何优化效果?

• 先申请发布模型,并备注说明希望通过云服务数据管理功能,将实际调用云服务识别的图片加入训练集,不断迭代模型

实际调用服务时模型效果变差?

- 训练图片和实际场景要识别的图片拍摄环境应一致,举例:如果实际要识别的图片是摄像头俯拍的,那训练图片就不能用网上下载的目标正面 图片
- 每个标签的图片需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强
- 如果使用的是云服务,可以开通云服务数据管理功能,将实际调用云服务识别的图片加入训练集,不断迭代模型

如果训练数据已经达到以上要求,且单个分类/标签的图片量超过200张以上,效果仍然不佳,请在百度智能云控制台内提交工单反馈

⊙ 智能标注相关问题

智能标注功能目前已对图像分割模型开放,了解功能详情

"一键标注"和"立即训练"要如何选择?

- 当系统推荐"立即训练",且系统预标注确实已非常精准时,可以不用标注剩余数据,直接开始模型训练。此时,仅用当前已标注图片训练的模型,与标注所有数据后训练的模型相比,效果几乎等同
- 如果系统预标注还有些不精准,可以启动一键标注,人工确认系统的标注后,再开始训练

选择了"立即训练"之后是否还可以"一键标注"?

- 选择"立即训练"之后,系统默认为您结束此次智能标注
- 再次启动智能标注后,您可以通过以下方式进行一键标注:
 - 根据系统提示,进入一键标注
 - 查看系统对"未标注[优先]"图片的预标注,点击"满意预标注结果"后,进入一键标注

智能标注结束后,又往数据集上传了新图片,是否可以直接"一键标注"新图片?

- 如果您创建了新的标签、或新上传的图片场景和之前的图片场景差异较大,建议不要使用一键标注,而是从头开始智能标注(即再次筛选关键 图片)
- 如果不是以上情况,再次启动智能标注后,可以通过以下方式进行一键标注:
 - 根据系统提示,进入一键标注
 - 查看系统对"未标注[优先]"图片的预标注,点击"满意预标注结果"后,进入一键标注

智能标注中可以增删标签吗?

- 暂不支持。为了保证系统智能标注的效果,建议在启动功能前就创建好所有需要识别的标签
- 如果确实需要增删标签,可以先结束智能标注

智能标注中可以增删图片吗?

- 暂不支持。为了保证系统智能标注的效果,建议在启动功能前上传需要标注的所有图片,并删除不相关的图片
- 如果确实需要增删图片,可以先结束智能标注

智能标注中可以修改已标注图片的标注吗?

- 可以。但为了保证智能标注的效果,建议不要大量改动
- 如果确实需要修改大量标注,建议先结束智能标注

为什么我已经人工标注了很多图片,但系统预标注依然不准?

- 系统预标注的结果会受以下因素影响:
 - 智能标注期间,对"已标注"图片的标签进行大量改动
 - 曾结束智能标注,并对标签、图片进行增删
- 如果您没有进行以上操作,系统标注结果依然不理想,请在百度智能云控制台内提交工单反馈

多个数据集是否可以同时启动智能标注?

• 目前每个账号同一时间仅支持对一个数据集启动智能标注

共享中的数据集是否可以启动智能标注?

• 暂不支持。智能标注中的数据集也暂不支持共享,如有疑问,请在百度智能云控制台内提交工单反馈

智能标注失败了怎么办?

- 可以先尝试稍后重新启动
- 若再次遇到问题,请在百度智能云控制台内提交工单反馈

∞ 模型上线相关问题

希望加急上线怎么处理?

• 请在百度智能云控制台内提交工单反馈

每个账号可以上线几个模型?是否可以删除已上线的模型?

• 每个账号最多申请发布十个模型,已上线模型无法删除

申请发布模型审核不通过都是什么原因?

 可能原因有,1、经过电话沟通当前模型存在一些问题或者不再使用,如训练数据异常、数据量不够、不想再继续使用等原因,沟通达成一致 拒绝。2、电话未接通且模型效果较差,会直接拒绝。如果需要申诉,请在百度智能云控制台内提交工单反馈

EasyDL 文本使用说明

EasyDL文本介绍

の概述

Hi,您好,欢迎使用百度EasyDL定制化训练和服务平台。

EasyDL平台的自然语言处理模型定制能力,基于文心·NLP大模型领先的语义理解技术,为企业/开发者提供一整套NLP定制与应用能力。

当前EasyDL平台提供了8种模型定制能力:

- 文本分类-单标签:定制分类标签实现文本内容的自动分类,每个文本仅属于一种标签类型
- 文本分类-多标签:定制分类标签实现文本内容的自动分类,每个文本可同时属于多个分类标签
- 情感倾向分析:定制情感倾向分析模型,可实现文本按情感的正向(positive)和负向(negative)做自动分类
- 短文本相似度:定制短文本相似度模型,是基于深度学习技术,可实现对两个文本进行相似度的比较计算
- 文本实体抽取:定制文本实体抽取模型,实现对文本进行内容抽取,并识别为自定义的实体类别
- 文本实体关系抽取:定制实体关系抽取模型,是指从文本中抽取出预定义的实体类型及实体间的关系类型,得到包含语义信息的实体关系三元组,每个实体关系三元组由两个实体及其关系构成
- 评论观点抽取:定制评论观点抽取模型,实现从文本中抽取评价片段、评价维度、评价观点,并判断评价情感倾向
- 大模型创作:定制文本创作模型,基于ERNIE 3.0大模型实现对输入文本内容进行创作和续写

心 产品优势

の 可视化操作

无需机器学习专业知识,通过模型创建-数据上传-模型训练-模型发布全流程可视化便捷操作,最快15分钟即可获得一个高精度模型。

心 操作步骤

Step 1 创建模型

确定模型名称,记录希望模型实现的功能。

Step 2 上传并标注数据

不同类型的任务对应的数据格式不一致,您可以上传未标注数据并使用平台提供的标注工具进行标注。或直接上传各任务的标注数据。

Step 3 训练模型并校验效果

选择部署方式与算法,用上传的数据一键训练模型。

模型训练完成后,可在线校验模型效果。

Step 4 发布模型

根据训练时选择的部署方式,将模型以云端API、设备端私有API等多种方式发布使用

更详细的操作指导,请参考各类模型的技术文档

心高精度效果

EasyDL文本任务内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等方面的新知识,实现模型效果不断进化。

文心大模型是百度发布的产业级知识增强大模型,是千行百业AI开发的首选基座大模型。文心大模型既包含基础通用的大模型,也包含了面向重 点领域和重点任务的大模型,还提供丰富的工具与平台,支撑企业与开发者进行高效便捷的应用开发。"知识增强"是文心的核心特色,文心能够 同时从大规模知识和海量多元数据中持续学习,如同站在巨人的肩膀上,训练效率和理解准确率都得到大幅提升,并具备了更好的可解释性。

◎ 丰富的部署方案

训练完成后,可将模型部署在公有云服务器、私有服务器。

部署方式	支持的硬件	支持的 系统	技术文档
公有云API	无需关注硬件,发 布到公有云	不限制	文本分类-单标签、文本分类-多标签、情感倾向分析、短文本相似度、文本实体抽 取、文本实体关系抽取、评论观点抽取、大模型创作
EasyEdge本地-私有服务 器部署[私有API]	x86-64 CPU	Linux	文本分类-单标签、文本分类-多标签、情感倾向分析、短文本相似度、文本实体抽取、文本实体关系抽取、评论观点抽取
EasyEdge本地-私有服务 器部署[私有API]	Nvidia GPU	Linux	文本分类-单标签、文本分类-多标签、情感倾向分析、短文本相似度、文本实体抽取、文本实体关系抽取、评论观点抽取

心 公有云API

已全面支持文本分类(单标签、多标签)、情感倾向分析、短文本相似度、文本实体抽取、文本实体关系抽取、评论观点抽取、大模型创作。

训练完成的模型存储在云端,可通过独立Rest API调用模型,实现AI能力与业务系统整合。

具有完善的鉴权、流控等安全机制,GPU集群稳定承载高并发请求。

₯私有服务器部署

已全面支持文本分类(单标签、多标签)、情感倾向分析、短文本相似度、文本实体抽取、文本实体关系抽取、评论观点抽取。

EasyDL文本任务使用EasyEdge本地部署服务,支持**私有服务器部署API的本地化部署**。私有API的特点如下:

- 将训练完成的模型部署在私有CPU/GPU服务器上,可在内网/无网环境下使用模型,确保数据隐私
- 将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口
- 可纯离线完成部署,服务调用便捷

私有API支持的Linux发行版本如下:

• Ubuntu: 14、16、18

- Centos: 7.0及以上
- redhat: 7.2以上
- suse 12

```
文本分类-单标签
<sup>整体介绍</sup>
```

心简介

Hi,您好,欢迎使用百度EasyDL定制化训练和服务平台。

定制文本分类的模型,是基于自建分类体系的机器学习方法,可实现文本按内容类型做自动分类。平台目前提供的文本分类模型包括:文本分类 (单标签)和文本分类(多标签)两种模型类型,请您根据自己的业务场景来选择合适的模型。本文介绍的是关于**文本分类(单标签)**的模型介 绍。

文本分类(单标签)场景:如您对网络文章进行舆情分析,判断舆情是正向评价还是负向评价,即每条文本**仅有一个**分类标准,此问题属于单标 签的文本分类场景;

文本分类(多标签)场景:如您对网络文章进行板块划分,即每条文本有**两个及以上**分类标准,文章可能属于娱乐、国际、生活等多个标签,则 可使用多标签的文本分类模型;

更多详情访问: EasyDL自然语言处理方向

₯应用场景

- 1、投诉信息分类:训练客服投诉信息的自动分类,将每个用户投诉的内容进行分类管理,节省大量客服人力
- 2、媒体文章分类:训练网络媒体文章的自动分类,进而实现各类文章的自动分类
- 3、文本审核:定制训练文本审核的模型,如训练文本中是否含有违规/偏激性质的描述
- 4、其他:尽情脑洞大开,训练你希望实现的文本分类(单标签)模型

心 技术特色

文本分类模型内置<u>文心大模型</u>,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等方面的新知识,实现模型效果不断进化。

文心大模型是百度发布的产业级知识增强大模型,是千行百业AI开发的首选基座大模型。文心大模型既包含基础通用的大模型,也包含了面向重 点领域和重点任务的大模型,还提供丰富的工具与平台,支撑企业与开发者进行高效便捷的应用开发。"知识增强"是文心的核心特色,文心能够 同时从大规模知识和海量多元数据中持续学习,如同站在巨人的肩膀上,训练效率和理解准确率都得到大幅提升,并具备了更好的可解释性。

心 使用流程

训练模型的基本流程如下图所示,全程可视化简易操作,在数据已经准备好的情况下,最快15分钟即可获得定制模型。



数据准备

心 创建数据集并导入

创建数据集

在训练模型之前,需要创建数据集。需输入数据集名称、选择相应的标注模版、选择数据去重策略,即可创建一个空数据集。

<i>┿</i> ジ ^ン ど桨EasyDL	产品介	绍 操作平台	应用案例	使用文档		
文本分类模型	<u>ا</u>	我的数据总览				
部 总览						
由 模型中心		EasyData智能数	据服务平台已上线	,使用EasyData可享受包括多人标	注、数据清洗、数据采集等完整数	双据服务 立即前往
我的模型		创建数据集				
创建模型)			
训练模型		pf文本分类测试	2 I 数据集组IC	311642		
校验模型		版本	数据集ID	数据量	最近导入状态	标注类型
发布模型						
☑ EasyData数据服务		V1 😳	336508	1018	 已完成 	文本分类
数据总览		~***		- 211620		
公开数据集		PIX本方尖测试	「ビ」 奴 坊 果 狙 ロ	- 311038		

数据自动去重即平台对您上传的数据进行重复样本的去重。建议创建数据集时选择「数据自动去重」

如果待导入数据集是中文简体/繁体,请选择『短文本单标签』;如果待导入数据集是非中文的其他语言,请选择『多语种文本单标签』,点 击可查看支持的全部语言种类。

导入数据 创建数据集后,在「数据总览」页面中,找到该数据集,点击右侧操作列下的「导入」,即可进入导入数据页面。

1	ど桨EasyDL	产品介	·绍 操作平台	应用案例	使用文	て档		
文ズ	本分类模型	Ē	我的数据总览 >	pf文本分类测	试2/V1/导入			
88	总览							
Ē	模型中心		┃创建信息 ∨					
	我的模型		数据集ID	336508		版本号	V1	
	创建模型		备注					
	训练模型		┃标注信息 ∨					
	校验模型		标注类型	文本分类		标注模板	短文本单标签	
	发布模型		数据总量	1018		已标注	1018(进度100.00%)
2	EasyData数据服务		标签个数	2		待确认	0	您可以使用4种方案上传文
	数据总览		大小	0.11M				
	公开数据集		■数据清洗					
	标签组管理		暂未做过数据滞	情洗任务				
	在线标注		┃ 导入数据					
	智能标注		数据标注状态	○ 无标	注信息	○ 有标注信息		
8	公有云服务		导入方式	请选择	L		^	
	在线服务			本地导	A			
L.	EasyEdge本地部署			BOSE	录导入			
	纯离线服务			分享链	接导入			
				+80	可如活来			
木公米	的数据 公别为·							
インス								

- 本地导入
- BOS目录导入
- 分享链接导入
- 平台已有数据集

1-	ど桨EasyDL	产品介绍	操作平台	应	用案例	使用文	挡			
文才	本分类模型	Ē	我的数据总览 >	pf文本	≤分类测试2/V	'1/导入				
	总览									
đ	模型中心		┃ 创建信息 >							
	我的模型		数据集ID	33650	08		版本号	V1		
	创建模型		备注	ß						
	训练模型		┃标注信息 >							
	校验模型		标注类型	文本分	类		标注模板	短文本单	单标签	
	发布模型		数据总量	1018			已标注	1018 (រ៉ូ	进度100.00%)	
<u>~</u>	EasyData数据服务		标签个数	2			待确认	0		
	数据总览		大小	0.11M						
	公开数据集		数据清洗							
	标签组管理		暂未做过数据清	情洗任务						
	在线标注		┃ 导入数据							
	智能标注		数据标注状态		○ 无标注信!	息	○ 有标注信息			
8	公有云服务		导入方式		本地导入			~	请选择 ^	
	在线服务				确认并返回				上传Excel文	件
12	EasyEdge本地部署								上传TXT文本	:
	纯离线服务								上传压缩包	

- 以压缩包的方式上传
- 以TXT文本文件方式上传
- 以Excel文件的方式上传 以压缩包方式上传

如果您想上传的数据为压缩包,请根据您的数据是否已标注,按照以下格式要求完成数据上传。

无标注数据

- 压缩包内包含上传的所有文本数据,每一个文本文件将作为一个样本上传,详见示例压缩包
- 压缩包格式为.zip格式,压缩包内文件类型支持txt,文件编码仅支持UTF-8

有标注数据

- 压缩包格式为.zip格式,同时压缩包大小在5GB以内,文本编码仅支持UTF-8,每个文本文件最长不能超过4096个字符
- 标注文件中标签由数字、中英文、中/下划线组成,长度上限256字符。

有标注数据可以使用以下两种格式组织压缩包的内容:

- 以文件夹命名样本的标签:压缩包内按照文本类别数量分为多个文件夹,以文件夹的名称作为文本类别标签,文件夹下的所有txt文件作为样本,详细请见示例压缩包
- 用json文件标记分类:压缩包内仅支持单个文本文件(txt)及同名的json格式标注文件的上传,可传多组样本,详细请见示例压缩包

以TXT文本文件上传

- 无标注数据文本文件内数据格式要求为"文本内容\n"(即每行一个未标注样本,使用回车换行),详见数据样例。有标注数据中文本文件内数据格式要求为"文本内容\t标签\n"(即每行一个标注样本,使用tab键将文本内容与标签分开,使用回车换行),详见数据样例。每一行表示一组数据,每组数据的字符数建议不超过4096个字符,超出将被截断;训练的字符数不超过512个字符,超出的字符可正常保存,但不参与训练。
- 文本文件类型支持txt,编码仅支持UTF-8,单次上传限制100个文本文件,最多可上传100万个文本文件

以Excel文件上传

- Excel文件内数据格式要求为:每行是一个样本详见数据样例,如果您上传的为有标注数据,则每行的样本包含两列,第一列为数据文本内容,第二列为文本对应标签,详见数据样例;如果您上传的为无标注数据,则每行样本仅包含第一列数据文本内容,每个数据样本文本内容的字符数建议不超过4096个,超出将被截断。
- 文件类型支持xlsx格式,单次上传文件个数上限为100个
- 请确保您上传的样本在sheet1中,注意,首行作为表头将被系统忽略

BOS目录导入 需选择Bucket地址与对应的文件夹地址。

请确保将全部文本已通过txt文件保存至同一层文件目录,该层目录下子文件目录及非相关内容(包括压缩包格式等)不导入。

分享链接导入 需输入链接地址。分享链接导入的要求如下:

• 仅支持来自百度BOS、阿里OSS、腾讯COS、华为OBS的共享链接

平台已有数据集

- 导入无标注数据时,选择需要导入的数据集名称,可导入其不带标注的全部数据,或未标注的数据
- 导入已标注数据时,选择需要导入的数据集名称,可导入其某个或全部标签下的数据

准备数据集的技巧 文本分类任务中,可参考以下准备数据集的技巧:设计分类

首先想好分类如何设计,每个分类为你希望识别出的一种结果,如要识别新闻的内容类型,则可以以"科技"、"体育"、"农业"等分别作为一个分类 标准;如果审核场景中通过文本判断是否出现广告,可以设计为两类设计为"正常"、"不正常"两类,或者"正常"、"异常原因一"、"异常原因二"、 "异常原因三"等多类。

注意:目前单个模型的上限为1000类,暂不支持扩容

数据量

基于设计好的分类准备文本数据,每个分类建议至少需要准备50个文本文件以上,如果想要较好的效果,建议文件1000个起,如果某些分类的 文本具有相似性,需要增加更多文本。

文本的基本格式要求: 目前文本文件类型支持txt,文本文件大小限制长度最大4096,格式为UTF-8字符。一个模型的文本总量限制10万个文本 文件。

数据分布

- 训练集文本需要和实际场景要识别的文本环境一致
- 考虑实际应用场景的种种可能性,每个分类的文本需要覆盖实际场景里面存在的可能性,训练集若能覆盖的场景越多,模型的泛化能力则越强

可能的疑问

- 如果训练文本数据无法全部覆盖实际场景要识别的文本,怎么办?
- 答:训练的模型算法会有一定的泛化能力,尽可能覆盖即可。
- 多语种模型支持全球94种语言:

南非语,阿姆哈拉语,阿拉伯语,阿萨姆语,阿塞拜疆语,白俄罗斯语,保加利亚语,孟加拉语,孟加拉语(拉丁化),布列塔尼语,波斯尼亚语,加泰隆语, 捷克语,威尔士语,丹麦语,德语,希腊语,英语,世界语,西班牙语,爱沙尼亚语,巴斯克语,波斯语,芬兰语,法语,弗里斯兰语,爱尔兰语,苏格兰盖 尔语,加利西亚语,古吉拉特语,希伯来语,印地语,印地语(拉丁化),克罗地亚语,匈牙利语,亚美尼亚语,印尼语,冰岛语,意大利语,日语,爪哇语, 格鲁吉亚语,哈萨克语,高棉语,康纳达语,韩语,库尔德语,柯尔克孜语,拉丁语,老挝语,立陶宛语,拉脱维亚语,马拉加斯语,马其顿语,马拉亚拉 姆语,蒙古语,马拉提语,马来语,缅甸语,尼泊尔语,荷兰语,挪威语,奥里亚语,旁遮普语,巴利语,普什图语,葡萄牙语,罗马尼亚语,俄语,梵语,信 德语,僧伽罗语,斯洛伐克语,斯洛文尼亚语,索马里语,阿尔巴尼亚语,塞尔维亚语,巽他语,瑞典语,斯瓦希里语,泰米尔语,泰米尔语(拉丁化),泰 卢固语,泰卢固语(拉丁化),泰语,他加禄语,土耳其语,维吾尔语,乌克兰语,乌尔都语,乌尔都语(拉丁化),乌兹别克斯坦语,越南语,意第绪语。

如果需要寻求第三方数据采集团队协助数据采集,请在百度云控制台内提交工单反馈

心 在线标注

在线标注

Step 1 进入标注页面 上传未标注的数据后,可以通过以下两个方式进入标注页面:

• 在「数据总览」页面,该数据集对应的操作列下,点击「标注」,即可进入标注页面

- 在「在线标注」页面,选择该数据集,即可进入标注页面
- **Step 2 进行文本标注** 针对尚未进行标注的数据,通过以下方式进行标注:
- 在右边标签栏中添加标签
- 针对文本内容,选择其对应的标签
- 点击下一篇,此篇文本的内容即可进行自动保存,且您将开始对下一篇文本进行标注

针对已进行标注的数据,通过以下方式进行标注修改:

- 进入需修改标签的文本的标注页面,选择右边标签栏中的标签
- 点击下一篇,对此篇文本标签的修改即可进行自动保存

全部(10) 天街注信息(0) 有街注信息(0) 标语注意(0)	标签栏	満加特盤 〜
	请输入标签名称	ı Q
1. 白色的,很漂亮,做工还可以; 2. 网上的软件资源非常非富,这是铁实它的最主要原因; 3. 电地不错,昨天从下午两点到晚上十点还有25分钟的刺你时间(关闭泪像头,无线和蓝牙)主要拷贝东西,看起来正常使用小小时左右	根据文本内在	3、选择唯一标签
没问题: 4. 散热不循,chut核心不过40~55度,很多小本要上到00度了; 5. 变压器很小巧,很多小本的电源都用的是大本的电源,本刻是很轻,可做行重量还是比较重。	标签D	
	标签C	快捷鍵 2
	标签B	KER 3
	标签A	快捷键 ④

标注技巧:每个标签的已标注文本需大于100条,且"无标注数据"大于0条,即可启动智能标注功能,提高标注效率

Step 3 查看标注信息 通过以下方式查看已标注的文本信息:

- 在「数据总览」页面,该数据集对应的操作列下,点击「查看」,进入查看标注页面后,点击「有标注信息」
- 通过选择左侧标签中的不同标签名称,即可查看不同标签下的文本数据

tj-yunfan-textV1	版本的文本列表 👘	<i>۱</i> ۷			回非	批量删除
全部标签 (6)	输入标签	名称 Q	序号	文本內容損要	操作	
标签名称	数据量	操作	1	说能很不错: 周边环境不错,紧邻金沙滩; 推荐 海里房 不足:比较偏: 吃饭不放便; 酒店自带的餐厅径价让太差,价格大贵,口味不好;	宣看 删除	13
test_5	2 1178	删除 导入	2	图它小巧带着方便 发拾量不差很大 有点珍珠白的颜色,很喜欢。 罐盘按起来很舒服,触摸板挺人性化的。	查看 删除	18
test_4	2 1178	删除 导入	3	硬盘声音大,好像是deli说批产品都有这个问题,E6400和E5400都有碳盘声音大的问题。	宣看 删约	kt
test_3	2 1178	删除 导入	4	做工不错,主要事价结实在,2400XT,看了网上的笔记半显卡用行,还是比较靠前的,跟9300有的比,于是就选了它,买之前看了3星的,3400带9200GS独显的那款,要3999,只多了…	查看 删除	ki
test_2	1178	删除 导入	5	仔细回想这本文集,发现自己喜欢的只是写《教室朝前,没有风事》的麻宁,不知道是她成长了还是自己止步不前了,没有感觉了。或许那是她对生活的睿智,但却不能让我感动了,觉得…	查看 删除	19
test_1	1178	删除 导入	6	我昨天咬到的书,质量还不错,而且配有光盘,光盘质量也还较清晰,我以筋也学过一点响强,看了书中描写,还以前学得意不多,应该是比较适合初学者的,总体难度不是很大,光盘的…	查看 删除	13
test_0	1178	删除 导入	7	性价比非常好的一款,推荐购买! 自己又添了一根20的条子,用起来超興!	查看 删除	18
			8	住这个酒店实在是太享受了.不仅可以使用五彩缤纷的白毛巾.还可以免费听赏别人KTV包厢里的高音演奏以下是某某歌奏穿的话谢谢 谢谢大家 首先我更感谢我的父母 还要感谢背后支	查看 删除	18
			9	我家宝贝的爸爸长期在外工作,很少在家,宝宝一岁半了,他好像还没找到做爸爸的感觉,我买这本书就是想教宝宝要学会和爸爸相处,数地爸爸学一点"理论",让他知道他这个爸爸的角…	查看 删除	18
	< 1 / 1 >		10	自己马上就有宝宝了,期待着宝宝蹲临人世,所以提前看看着就方面的书籍。看了这本书,真的很好,不但丛中可以学到做背孩子的方法,自己也丛中收获很多。犹太民族起历了那么多约…	查看 删除	18

当前数据集标注模板:短文本单标签,该标签有标注信息文本:1178个

心 数据去重

重复样本的定义

一个样本包括文本内容和标签。重复样本的定义,是指您上传的数据中,存在两个样本的文本内容完全一致。则被判定为两个样本是重复样本。 例如:

文本内容	标签
今天北京的空气不错	weather
今天北京的空气不错	weather
今天北京的空气不错	local

上表三个样本均为重复样本,后两个样本虽然标签不一,但文本内容一致,也为重复样本。根据文本出现的顺序,最后一次的重复样本将代替之前的重复样本。

小Tips:"如何利用好重复样本" 如果您的数据存在样本种类不均衡的现象,您可以通过将重复样本数量小的那一类,使其样本数量增加到与数据量大的那一类样本数量相近,以提高模型训练的效果,这种方法也称为"上采样"。
平台去重策略

平台提供了可去重的数据集,即对您上传的数据进行重复样本的去重。

当您创建了一个去重的数据集时,在后续上传数据的过程中,平台可通过检验您当前上传的样本与已上传到此数据集下的样本是否相同,如果相同,则会使用新的样本替代旧的样本。此时分为几种情况,如下:

- 数据集中有未标注样本,上传重复的已标注样本,此时未标注样本将被覆盖
- 数据集中有已标注样本,上传重复的未标注样本,此时已标注样本将被覆盖
- 数据集中有已标注样本,上传不同标注的已标注样本,此时已有的标注样本将被覆盖

の API上传

本文档主要说明当您线下已有大量的已经完成分类整理的文本数据,如何通过调用API完成文本数据的便捷上传和管理。

EasyDL数据集管理API在管理不同模型数据类型之间是通用的。上传不同模型类型数据,只是在部分接口入参存在差异,使用及接口地址完全一致。

数据集创建API

接口描述

该接口可用于创建数据集。

接口鉴权

同发布模型后获取的API鉴权方式:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/create

URL参数:

参数 值 access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset_na me	是	string	数据集名称,长度不超过20个utf-8字符

若上传声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
dataset_id	否	number	创建的数据集ID

查看数据集列表API

接口描述

该接口可用于查看数据集列表。返回数据集的名称、类型、状态等信息。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/list

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
start	否	number	起始序号,默认为0
num	否	number	数量,默认20,最多100

若查看声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明		
log_id	是	number	唯一的log id,用于问题定位		
total_num	否	number	数据集总数		
results	否	array(object)	数据集列表		
+dataset_id	否	number	数据集ID		
+dataset_name	否	string	数据集名称		
+type	否	string	数据集类型		
+status	否	string	数据集状态		
+special_status	否	string	数据集特殊状态,包括shared、smart和空值,分别表示共享中、智能标注中、非特殊状态		

查看分类(标签)列表API

接口描述

该接口可用于查看分类(标签)。返回分类(标签)的名称、包含数据量等信息。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/list

URL参数:

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID
start	否	number	起始序号,默认0
num	否	number	数量,默认20,最多100

若查看声音分类的全部分类,在type参数应传「SOUND_CLASSIFICATION」

返回说明

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	标签总数
results	否	array(object)	标签列表
+label_id	否	string	标签/分类ID
+label_name	否	string	标签/分类名称
+entity_count	否	number	图片/声音/文本数量

添加数据API

接口描述

该接口可用于在指定数据集添加数据。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法:POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/addentity

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明	
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类	
dataset _id	是	number	数据集ID	
append Label	否	boolea n	确定添加标签/分类的行为:追加(true)、替换(false)。默认为追加(true)。	
entity_c ontent	是	string	type为 IMAGE_CLASSIFICATION/OBJECT_DETECTION/IMAGE_SEGMENTATION/SOUND_CLASSIFICATION时,填入图 片/声音的base64编码;type为TEXT_CLASSIFICATION时,填入utf-8编码的文本。内容限制为:图像分类base64前 10M;物体检测base64前10M;图像分割base64前10M;声音分类base64前4M,声音时长1~15秒;文本分类 10000个汉字	
entity_n ame	是	string	文件名	
labels	是	array(o bject)	际签/分类数据	
+label_ name	是	string	标签/分类名称(由数字、字母、中划线、下划线组成),长度限制20B	
+left	否	number	物体检测时需给出,标注框左上角到图片左边界的距离(像素)	
+top	否	number	物体检测时需给出,标注框左上角到图片上边界的距离(像素)	
+width	否	number	物体检测时需给出,标注框的宽度(像素)	
+height	否	number	物体检测时需给出,标注框的高度(像素)	

若上传声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

数据集删除API

接口描述

该接口可用于删除数据集。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/delete

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID

若刪除声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

分类 (标签) 删除API

接口描述

该接口可用于删除分类(标签)。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/delete

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset_ id	是	number	数据集ID
label_na me	是	string	标签/分类名称

若删除声音分类的子类,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
   "error_code": 110,
   "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量 阶梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计 费;未上线计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶 梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
406000	internal server error	服务错误
406001	param[xx] invalid	参数xx不合法,请检查相关参数
406002	dataset not exist	数据集不存在
406003	dataset already exists	数据集已存在
406004	dataset can not be modified temporarily	数据集暂不可修改
406005	label not exist	标签/分类不存在
406006	no permission to modify the dataset	没有修改数据集的权限
406007	dataset cannot be modified while smart annotation is running	智能标注期间不可修改数据集
406008	quota exceeded	配额超限

心 创建数据集并导入

创建数据集

在训练模型之前,需要创建数据集。需输入数据集名称、选择相应的标注模版、选择数据去重策略,即可创建一个空数据集。

<i>ᅻ</i>	产品介绍	3 操作平台	应用案例	使用文档		
文本分类模型	÷	我的数据总览				
品 总览						
曲 模型中心		EasyData智能数	居服务平台已上线,	,使用EasyData可享受包括多人标注、	、数据清洗、数据采集等完整数据服务	立即前往
我的模型		创建数据集				
创建模型						
训练模型		pf文本分类测试2	☑ 数据集组Ⅳ	: 311642		
校验模型		版本	数据集ID	数据量	最近导入状态	标注类型
发布模型						
☑ EasyData数据服务		V1 ⊙	336508	1018	● 已完成	文本分类
数据总览						
公开数据集		pf又本分类测试1		311639		

数据自动去重即平台对您上传的数据进行重复样本的去重。建议创建数据集时选择「数据自动去重」

如果待导入数据集是中文简体/繁体,请选择『短文本单标签』;如果待导入数据集是非中文的其他语言,请选择『多语种文本单标签』,点 击可查看支持的全部语言种类。

导入数据 创建数据集后,在「数据总览」页面中,找到该数据集,点击右侧操作列下的「导入」,即可进入导入数据页面。

<i>ذرذ</i> ر	ど 楽 EasyDL	产品介	个绍 操作平台	应用案例	使用文档		
文本分	分类模型	۹ <u>–</u>	我的数据总览	> pf文本分类测试	2/V1/导入		
品模	说 ^{建型中心}		┃ 创建信息 ~				
我	的模型 J建模型		数据集ID 备注	336508	版本号	V1	
训	练模型		■标注信息 ~ 标注类型	文本分类	标注模板	短文本单标签	
发	(布模型		数据总量	1018 2	已标注	1018(进度100.00%) 0	
₩ Ei	asyData致掂服务 X据总览		大小	0.11M			您可以使用4种方案上传文
公标	\开数据集 i签组管理		■数据清洗 暫未做过数排	請洗任务			
在	5线标注 3能标注		 导入数据 数据标注状系	5 💿 无标注	E信息 〇 有标注	信息	
6 公	有云服务		导入方式	请选择		^	
在 記 Ei	E 线服务 asyEdge本地部署			本地导入 BOS目录 分享链接	、 t导入 t导入		
纯	加离线服务			平台已有	ī数据集		

本分类的数据,分别为:

- 本地导入
- BOS目录导入
- 分享链接导入
- 平台已有数据集

本地导入您可以通过以下三种方式进行本地数据的导入:

1	ど桨EasyDL	产品介绍	操作平台	应用	用案例	使用文	当				
文才	本分类模型	Ē	我的数据总览 >	pf文本	分类测试2/V	1/导入					
	总览										
Ð	模型中心		┃创建信息 >								
	我的模型		数据集ID	33650	8		版本号	V1			
	创建模型		备注	Ľ							
	训练模型		┃标注信息 >								
	校验模型		标注类型	文本分	类		标注模板	短文本	单杉	示签	
	发布模型		数据总量	1018			已标注	1018 (进度	度100.00%)	
<u>~</u>	EasyData数据服务		标签个数	2			待确认	0			
	数据总览		大小	0.11M							
	公开数据集		数据清洗								
	标签组管理		暂未做过数据清	青洗任务							
	在线标注		导入数据								
	智能标注		数据标注状态		无标注信息	<u>.</u>	○ 有标注信息				
8	公有云服务		导入方式		本地导入			,	~	请选择 へ	_
	在线服务				确认并返回				ſ	上传Excel文	件
2	EasyEdge本地部署									上传TXT文本	ŝ
	纯离线服务								l	上传压缩包	

- 以压缩包的方式上传
- 以TXT文本文件方式上传
- 以Excel文件的方式上传 以压缩包方式上传

如果您想上传的数据为压缩包,请根据您的数据是否已标注,按照以下格式要求完成数据上传。

无标注数据

- 压缩包内包含上传的所有文本数据,每一个文本文件将作为一个样本上传,详见示例压缩包
- 压缩包格式为.zip格式,压缩包内文件类型支持txt,文件编码仅支持UTF-8

有标注数据

- 压缩包格式为.zip格式,同时压缩包大小在5GB以内,文本编码仅支持UTF-8,每个文本文件最长不能超过4096个字符
- 标注文件中标签由数字、中英文、中/下划线组成,长度上限256字符。

有标注数据可以使用以下两种格式组织压缩包的内容:

- 以文件夹命名样本的标签:压缩包内按照文本类别数量分为多个文件夹,以文件夹的名称作为文本类别标签,文件夹下的所有txt文件作为样本,详细请见示例压缩包
- 用json文件标记分类:压缩包内仅支持单个文本文件(txt)及同名的json格式标注文件的上传,可传多组样本,详细请见示例压缩包

以TXT文本文件上传

- 无标注数据文本文件内数据格式要求为"文本内容\n"(即每行一个未标注样本,使用回车换行),详见数据样例。有标注数据中文本文件内数据格式要求为"文本内容\t标签\n"(即每行一个标注样本,使用tab键将文本内容与标签分开,使用回车换行),详见数据样例。每一行表示一组数据,每组数据的字符数建议不超过4096个字符,超出将被截断;训练的字符数不超过512个字符,超出的字符可正常保存,但不参与训练。
- 文本文件类型支持txt,编码仅支持UTF-8,单次上传限制100个文本文件,最多可上传100万个文本文件

以Excel文件上传

- Excel文件内数据格式要求为:每行是一个样本详见数据样例,如果您上传的为有标注数据,则每行的样本包含两列,第一列为数据文本内容,第二列为文本对应标签,详见数据样例;如果您上传的为无标注数据,则每行样本仅包含第一列数据文本内容,每个数据样本文本内容的字符数建议不超过4096个,超出将被截断。
- 文件类型支持xlsx格式,单次上传文件个数上限为100个
- 请确保您上传的样本在sheet1中,注意,首行作为表头将被系统忽略

BOS目录导入需选择Bucket地址与对应的文件夹地址。

请确保将全部文本已通过txt文件保存至同一层文件目录,该层目录下子文件目录及非相关内容(包括压缩包格式等)不导入。

分享链接导入 需输入链接地址。分享链接导入的要求如下:

• 仅支持来自百度BOS、阿里OSS、腾讯COS、华为OBS的共享链接

平台已有数据集

- 导入无标注数据时,选择需要导入的数据集名称,可导入其不带标注的全部数据,或未标注的数据
- 导入已标注数据时,选择需要导入的数据集名称,可导入其某个或全部标签下的数据

准备数据集的技巧 文本分类任务中,可参考以下准备数据集的技巧:设计分类

首先想好分类如何设计,每个分类为你希望识别出的一种结果,如要识别新闻的内容类型,则可以以"科技"、"体育"、"农业"等分别作为一个分类 标准;如果审核场景中通过文本判断是否出现广告,可以设计为两类设计为"正常"、"不正常"两类,或者"正常"、"异常原因一"、"异常原因二"、 "异常原因三"等多类。

注意:目前单个模型的上限为1000类,暂不支持扩容

数据量

基于设计好的分类准备文本数据,每个分类建议至少需要准备50个文本文件以上,如果想要较好的效果,建议文件1000个起,如果某些分类的 文本具有相似性,需要增加更多文本。

文本的基本格式要求: 目前文本文件类型支持txt,文本文件大小限制长度最大4096,格式为UTF-8字符。一个模型的文本总量限制10万个文本 文件。

数据分布

- 训练集文本需要和实际场景要识别的文本环境一致
- 考虑实际应用场景的种种可能性,每个分类的文本需要覆盖实际场景里面存在的可能性,训练集若能覆盖的场景越多,模型的泛化能力则越强

可能的疑问

• 如果训练文本数据无法全部覆盖实际场景要识别的文本,怎么办?

答:训练的模型算法会有一定的泛化能力,尽可能覆盖即可。

• 多语种模型支持全球94种语言:

南非语,阿姆哈拉语,阿拉伯语,阿萨姆语,阿塞拜疆语,白俄罗斯语,保加利亚语,孟加拉语,孟加拉语(拉丁化),布列塔尼语,波斯尼亚语,加泰隆语, 捷克语,威尔士语,丹麦语,德语,希腊语,英语,世界语,西班牙语,爱沙尼亚语,巴斯克语,波斯语,芬兰语,法语,弗里斯兰语,爱尔兰语,苏格兰盖 尔语,加利西亚语,古吉拉特语,希伯来语,印地语,印地语(拉丁化),克罗地亚语,匈牙利语,亚美尼亚语,印尼语,冰岛语,意大利语,日语,爪哇语, 格鲁吉亚语,哈萨克语,高棉语,康纳达语,韩语,库尔德语,柯尔克孜语,拉丁语,老挝语,立陶宛语,拉脱维亚语,马拉加斯语,马其顿语,马拉亚拉 姆语,蒙古语,马拉提语,马来语,缅甸语,尼泊尔语,荷兰语,挪威语,奥里亚语,旁遮普语,巴利语,普什图语,葡萄牙语,罗马尼亚语,俄语,梵语,信 德语,僧伽罗语,斯洛伐克语,斯洛文尼亚语,索马里语,阿尔巴尼亚语,塞尔维亚语,巽他语,瑞典语,斯瓦希里语,泰米尔语,泰米尔语(拉丁化),泰 卢固语,泰卢固语(拉丁化),泰语,他加禄语,土耳其语,维吾尔语,乌克兰语,乌尔都语,乌尔都语(拉丁化),乌兹别克斯坦语,越南语,意第绪语。

如果需要寻求第三方数据采集团队协助数据采集,请在百度云控制台内提交工单反馈

心 数据去重

重复样本的定义

一个样本包括文本内容和标签。重复样本的定义,是指您上传的数据中,存在两个样本的文本内容完全一致。则被判定为两个样本是重复样本。 例如:

文本内容	标签
今天北京的空气不错	weather
今天北京的空气不错	weather
今天北京的空气不错	local

上表三个样本均为重复样本,后两个样本虽然标签不一,但文本内容一致,也为重复样本。根据文本出现的顺序,最后一次的重复样本将代替之前的重复样本。

小Tips:"如何利用好重复样本"如果您的数据存在样本种类不均衡的现象,您可以通过将重复样本数量小的那一类,使其样本数量增加到与数据量大的那一类样本数量相近,以提高模型训练的效果,这种方法也称为"上采样"。

平台去重策略

平台提供了可去重的数据集,即对您上传的数据进行重复样本的去重。

当您创建了一个去重的数据集时,在后续上传数据的过程中,平台可通过检验您当前上传的样本与已上传到此数据集下的样本是否相同,如果相同,则会使用新的样本替代旧的样本。此时分为几种情况,如下:

- 数据集中有未标注样本,上传重复的已标注样本,此时未标注样本将被覆盖
- 数据集中有已标注样本,上传重复的未标注样本,此时已标注样本将被覆盖
- 数据集中有已标注样本,上传不同标注的已标注样本,此时已有的标注样本将被覆盖

模型训练

心 创建模型

步骤 Step 1 创建模型

在【模型中心】或者【模型中心-我的模型】点击创建模型; *数据集是中文数据集,任务场景请选择『短文本分类任务』;* 数据集是非中文数据 集,任务场景请选择『多语种文本分类任务』,模型共支持全球94种语言,点击可查阅

Step 2 填写基本信息 选择模型类型、提交模型名称、模型描述、联系方式即可创建模型。

Step 3 查看已创建的模型 模型创建成功后,可以在【我的模型】中看到刚刚创建的模型,操作示例见下图。

文本分类模型	模型列表 > 创建模型	Ł交工单
△ 模型中心 我的模型	模型笑刻 文本公共-奉柳签	
创建模型	任务结果 * 多语种文本分类任务 > ②	
训练模型 校验模型		
发布模型		
∠ EasyData数据服务	test	
数据总览	所属行业 * 教育培训 ~	
公开:XX编集 标签组管理	应用场景 * 请选择应用场景 ∨	
在线标注	邮箱抢战 * z*********@163.com	
智能标注 23 模型印音	联系方式。 153****472 ③	
纯面线服务	动能描述。 多语种分类任务支持94种语言	
	14000	
	T-#	

- 1. 创建模型后可持续新增模型版本,因此不必每次训练模型都创建模型
- 目前单个用户在每种类型的模型下最多可创建10个模型,每个模型均支持多次训练,若需要创建超过10个以上的模型,请在百度云控制 台内提交工单反馈。
- 3. 如果您是企业用户,建议您按照真实企业信息进行填写,便于EasyDL团队后续更好的为您服务
- 支持语言清单:

南非语, 阿姆哈拉语, 阿拉伯语, 阿萨姆语, 阿塞拜疆语, 白俄罗斯语, 保加利亚语, 孟加拉语, 孟加拉语(拉丁化), 布列塔尼语, 波斯尼亚语, 加泰隆语,

捷克语,威尔士语,丹麦语,德语,希腊语,英语,世界语,西班牙语,爱沙尼亚语,巴斯克语,波斯语,芬兰语,法语,弗里斯兰语,爱尔兰语,苏格兰盖 尔语,加利西亚语,古吉拉特语,希伯来语,印地语,印地语(拉丁化),克罗地亚语,匈牙利语,亚美尼亚语,印尼语,冰岛语,意大利语,日语,爪哇语, 格鲁吉亚语,哈萨克语,高棉语,康纳达语,韩语,库尔德语,柯尔克孜语,拉丁语,老挝语,立陶宛语,拉脱维亚语,马拉加斯语,马其顿语,马拉亚拉 姆语,蒙古语,马拉提语,马来语,缅甸语,尼泊尔语,荷兰语,挪威语,奥里亚语,旁遮普语,巴利语,普什图语,葡萄牙语,罗马尼亚语,俄语,梵语,信 德语,僧伽罗语,斯洛伐克语,斯洛文尼亚语,索马里语,阿尔巴尼亚语,塞尔维亚语,巽他语,瑞典语,斯瓦希里语,泰米尔语,泰米尔语(拉丁化),泰 卢固语,泰卢固语(拉丁化),泰语,他加禄语,土耳其语,维吾尔语,乌克兰语,乌尔都语,乌尔都语(拉丁化),乌兹别克斯坦语,越南语,意第绪语。

心 效果优化

通过模型迭代、检查并优化训练数据、选择高精度模型等方法,能够提升模型效果。**模型迭代**

一个模型很难一次性就训练到最佳的效果,通常会需要结合模型评估报告和校验结果不断扩充数据和调优。

为此平台提供了模型迭代功能,即当模型训练完毕后,会生成一个最新的版本号,首次V1、之后V2……以此类推。可以通过调整训练数据和算法,多次训练,以获得适合业务需求的模型效果。

如果模型已经是上线状态,依然支持模型迭代,只是需要在训练完毕后更新线上服务接口,在接口地址不变的情况下可以持续优化效果。

检查并优化训练数据

- 检查是否存在训练数据过少的情况,建议每个类别的文本数量不少于1000个,如果低于这个量级建议扩充
- 检查不同类别的数据量是否均衡,建议不同分类的数据量级相同,并尽量接近,如果有的类别数据量很高,有的类别数据量较低,会影响模型 整体的识别效果
- 通过模型效果评估报告中的各分类的详细评估指标,有针对性地扩充训练数据,比如下图中的评估报告显示,标签为1的分类精确值数据表现 较标签为0的精确值要差,即可考虑从两个方向进行针对性优化,一是增加标签为1的数据集样本数据量,二是检查现有标签为1数据集中是否 存在定义模糊的情况,提升标签为1的数据集质量,优化模型效果。

文本分类模型 ← 我的模型 > pr文本测试2-V3模型评估报告 BB 总览 邮署方式 公有云API 反本 V3 ✓ 我的模型 文本数 1018 分类数 2 训练时长 7分钟	ビ来 EasyDL 产品介绍	操作平台 应用案例 使用文档
設置 協選 部署方式 公有云API 版本 V3 我的模型 文本数 1018 分类数 2 训练时长 7分钟	文本分类模型 (三 我的	的模型 > pf文本测试2-V3模型评估报告
	 28 总发 通 模型中心 38 約約模型 文本 	居方式 公有云API ∨ 版本 V3 ∨
创建模型 训练模型 算整体评估 校验模型 pf文本测试2 V3效果优异。如何优化效果? 发布模型 ど EasyData数据服务 91.1%	- 创建模型 训练模型	整体评估 文本测试2 V3效果优异。 如何优化效果?
数据总览 详细评估 公开数据集 详细评估 标签组管理 评估样本具体数据情况	数据总览 公开数据集 详 标签组管理 ;	羊细评估
在线标注 智能标注 公 有 云 服 多 在 线 服 多 か の で ま の で の の の の の の の の の の の の の の の	在线标注 智能标注 目 公有云服务 在线服务	随机测试集 近确数量: 278 错误数量: 27
□ EasyEdge本地部署 純高线服务	記 EasyEdge本地部署 纯离线服务	各分类的精确値 各分类的F1-Score 0 93.4% 0 91.3% 1 89.0% 1 91.0% 各分类的召回値 93.2% 1 93.2%

 检查测试模型的数据与训练数据的文本类型与风格是否一致,如果不一致,那么很可能会存在模型效果不错但实际测试效果较差的情况。针对 这种情况建议重新调整训练集,将训练数据与实际业务场景数据尽可能一致

选择高精度模型

在训练模型时,选择高精度的模型,将提升模型的预测准确率。

「高精度」算法内置<u>文心大模型</u>,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等方 面的新知识,实现模型效果不断进化。

心 评估效果

模型评估 模型训练完成后,可以在「我的模型」列表中查看该模型的效果,以及完整评估结果。

文本分类模型 (三	我的模型 > 文本标题分类VI模型评估报告			提交工单
28 总筑				
直 模型中心	部署方式 EasyEdge撮务器部署 > 版本 VI	~		
我的模型	文本数 19430 分类数 5 训练时长 2/	时20分钟		
创建模型				
训练模型	整体评估			
校验模型	文本标题分类 VI及果优异。 如何优化效果?			
发布模型	() 准确率 ()	F1-score @	() 精助率 ()	日日本の
回 EasyData数据服务	96.8%	96.7%	96.7%	96.8%
数据总支				
公开数据集	详细评估			
移签坦管理	评估样本具体数据情况			
在线标注	随机测试集			
智能标注	正确数量: 2821			
日 公有云服务	预测表现 错误数量: 93			
在线服务 PT Excelence 本地対理	告分类的精确值	各分类的F1-Score		
Paratico de la construction de	819 97.1%	87.1%		
106(0.00)	87. 99.0%	97.2%		
	教育 97.2%	NT 98.1%		
	81te 97.7%	First 96.1%		
	212 940%	85 Std		
	各分类的召田值			
	B110 97.1%			
	B/* \$6.4%			
	RT 98.9%			
	科技 95.0%			
	10 00.0%			

「完整评估结果」页面中将记录整体评估与详细评估的报告,包括该模型整体的准确率、F1-score、精确率、召回率,以及评估样本具体数据情况,各分类的精确值、F1-Score等指标。

整体评估中,各指标的释义如下:

- 准确率:正确分类的样本数与总样本数之比
- F1-score:给每个类别相同的权重,计算每个类别的F1-score,然后求平均值
- 精确率:给每个类别相同的权重,计算每个类别的精确率,然后求平均值
- 召回率:给每个类别相同的权重,计算每个类别的召回率,然后求平均值

如果单个标签的文本量在100条以内,会影响评估指标的科学有效性,请确保提交的训练数据中每个标签的数据量

模型校验

在完成训练后,发布模型前,可以先进行模型校验,以确保模型在实际环境中能获得预期的性能。操作方法如下:

- 1. 在左侧「模型中心」目录中点击「校验模型」,进入校验模型页面
- 2. 选择需要校验能力的模型、部署方式、版本,点击「启动模型校验服务」

3. 校验服务启动后,在左侧输入文本,点击「校验」后,右侧识别结果栏将输出预测结果

- 校验数据支持两种输入方式,直接输入文本或上传txt格式文本,文本长度上限为512汉字
- 在识别结果栏中,可进行阈值的调整。置信度在阈值以下的预测结果将不予显示。例如,阈值调整为0.85时,置信度小于85%的预测结果将不予显示。各类标签的置信度均小于85%时,识别结果栏将显示为:"没有满足条件的识别结果"
- 4. 若您对预测结果满意,可点击「申请上线」,进行模型的发布

如果单个标签的文本量在100条以内,会影响评估指标的科学有效性,请确保提交的训练数据中每个标签的数据量

心 发起训练

训练模型

完成数据的标注,或提交已标注的数据后,即可在「模型中心」目录中点击「训练模型」,开始模型的训练。

按以下步骤操作,启动模型训练:

文本分类模型 《二 训练模型

38 SE						
自 根型中心	选择模型	文本标题分类				
我的模型	训练配置					
创建模型	部署方式	公有云部署 EasyEdge3	地部署 如何选择部署方式?			
(1185-18) 22	选择算法	◎ 高精度 ② ○ 高性能	0			
校验模型	<i>曝形调读指任(</i> ?)	編列集研Precision和B V	0			
发布模型	1 Martin C	5-12-5-50 ⁻				
☑ EasyData数据服务	Statutes					
数据总定	添加数据集	+ 请选择				
公开数据集	自定义测试集 🗇					
标篮细管理	31100 TT 10	A110	171.47	***	THE REPORT OF ANY	0.10
在线标注	111.121.171.08	-10 ft	现怕	解刀	35.56.07.99	0740
智能标注		GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1	免费
白 公有云服务		O GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	1.5	单卡¥0.45/分钟 (348.62小时*节点小时包余额)
在线服务	TTAGUEAR					
22 EasyEdge本地部署	97923135					
纯高线服务						

Step 1 选择模型 选择此次训练的模型 Step 2 训练配置

部署方式

可选择「公有云部署」、「EasyEdge本地部署」。

如何选择部署方式

选择设备

- 如果您选择了「EasyEdge本地部署」,请根据实际部署设备选择
- 如果您选择了「公有云部署」,无需选择设备

选择算法

您可以根据训练的需求,选择「高精度」或「高性能」算法。不同的算法将影响训练时间、预测速度与模型准确率。

- 如果您选择了高精度的模型,模型预测准确率将更高。如果您手中的标注数据集样本较少(例如少于1000条),可选择「高精度」的算法。
 使用高精度的算法训练模型将会耗时更久,实验环境下1000个样本,预计在20分钟左右完成训练
- 如果您选择了高性能的模型,相同训练数据量的情况下,训练耗时更短,模型预测速度更快。使用10000条训练样本,将在15min内完成训练.
 同样的数据量情况下,效果比高精度的模型4-5%

「高精度」算法内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义 等方面的新知识,实现模型效果不断进化

模型筛选指标

选择不同的模型选择方式,对应的模型各项效果指标将有所不同。如果没有特殊场景的要求,使用默认即可(兼顾Precision精确度和Recall召回率)。有以下指标可供选择:

- 模型兼顾Precision和Recall:挑选模型时,兼顾Precision精确度和Recall召回率,如场景中没有对精度或召回的特别要求,建议您使用此默认 指标
- Precision最高的模型:挑选模型时,优先挑选Precision精度最高的模型作为部署模型
- Recall最多的模型:挑选模型时,优先选择召回率最高的模型作为部署模型
- ACC最大的模型:挑选模型时,优先挑选预测样本数最多的模型作为部署模型
- Loss最小的模型:挑选模型时,优先挑选预测偏差最小的模型

Step 3 添加数据

添加训练数据

- 先选择数据集,再按标签选择数据集里的文本,可从多个数据集选择文本。注意,文本分类模型至少需要选择2个及以上分类
- 训练时间与数据量大小、选择的算法、训练环境有关。在选择GPU P40、高性能的模型时,10000条训练样本,将在15min内完成训练

添加自定义测试集

上传不包含在训练集的测试数据,可获得更客观的模型效果评估结果。

添加自定义测试集的目的:

如果学生的期末考试是平时的练习题,那么学生可能通过记忆去解题,而不是通过学习的方法去做题,所以期末考试的试题应与平时作业不能一样,才能检验学生的学习成果。那么同理,AI模型的效果测试不能使用训练数据进行测试,应使用训练数据集外的数据测试,这样才能 真实的反映模型效果

期末考试的内容属于学期的内容,但不一定需要完全包括所学内容。同理,测试集的标签是训练集的全集或者子集即可

Step 4 训练模型

点击「开始训练」,即可开始训练模型。

- 训练时间与数据量大小、选择的算法、训练环境有关
- 模型训练过程中,可以设置训练完成的短信提醒并离开页面
- 平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

优惠政策:

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有声音分类操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。可参考价格说明

模型部署

心 整体介绍

训练完成后,可将模型部署在公有云服务器、私有化服务器上,通过API进行调用。 公有云API

- 模型训练完毕后,为了更方便企业用户一站式完成AI模型应用,文本分类模型支持将模型发布成为在线的restful API接口,可以参考示例文 档通过HTTP请求的方式进行调用,快速集成在业务中进行使用。
- 具有完善的鉴权、流控等安全机制,GPU集群稳定承载高并发请求

相关费用 将模型发布为公有云API后,将获得一部分免费调用次数,超出免费额度将根据调用次数进行收费。详见EasyDL价格文档。

私有服务器部署

支持将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便 捷。适用于对数据敏感度、隐私性要求较高、在线离线均有调用需求的企业场景。相关费用 EasyDL已支持将定制模型部署在本地服务器上,只 需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。如需购买永久使用授权,请提交工单咨询。

心 公有云API

の 发布API

发布模型,生成在线API

训练完毕后可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可发布公有云API:

- 1. 选择模型
- 2. 选择部署方式「公有云部署」

- 3. 选择版本
- 4. 自定义服务名称、接口地址后缀
- 5. 申请发布

	文Z	本分类模型	•≡	发布模型			
		总览					
	Ē	模型中心		选择模型	酒店评分1 ~		
		我的模型			公有云部署 >>		
		创建模型					
		训练模型		选择版本	V3 ~		
		校验模型		服务名称 *	识别酒店评论好与坏		
	[3] (II)	发布模型		接口地址 * 其他要求	https://aip.baidubce.com/rpc/2.0/ai_custo		
		EasyData数据服务			m/v1/text_cls/ jacetest3		
发布模型界面示意:		数据总览					
		公开数据集			右接口无法满足恋的需求,请抽述希望解决 的问题,500汉字以内		
		标签组管理					
		在线标注			0/500		
		智能标注			提交申请		
		公有云服务					
		在线服务					
		EasyEdge本地部署					
		纯离线服务					

发布完成 申请发布后,通常的审核周期为T+1,即当天申请第二天可以审核完成。

如果需要加急、或者遇到莫名被拒的情况,请在百度云控制台内提交工单反馈。

申请发布通过后,界面和状态示意:

【文本分类】	┃ 新闻 模型	≝ID: 19900					全部版本	删除
应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作		
云服务	V1	训练完成	未申请	未发布	top1准确率100.00% top5准确率100.00% 完整评估结果	申请发布校验	训练	

つ。调用API

接口描述

基于自定义训练出的文本分类模型,实现个性化文本识别。模型训练完毕后发布可获得定制化文本分类API 详情访问:EasyDL 文本分类·单标签 进行训练。

更多训练模型过程中的常见问题请查看常见问题文档。

如有其它问题,请在百度云控制台内提交工单反馈。

请求说明

请求示例

```
HTTP 方法: POST
```

请求URL:请首先在EasyDL 文本分类-单标签进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值

Content-Type application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,定制化文本分类服务以json方式请求。

Body请求示例:

```
{
"text": "<UTF-8编码文本>",
"top_num": 6
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text	是	string	-	文本数据,UTF-8编码。最大长度4096汉字
top_num	否	number	-	返回分类数量,默认为6个

请求示例代码

Python3	
""" EasyDL 文本分类单标签 调用模型公有云API Python3实现 """	
import json import base64 import requests	
使用 requests 库发送请求 使用 pip(或者 pip3)检查我的 python3 环境是否安装了该库,执行命令 pip freeze grep requests 若返回值为空,则安装该库 pip install requests	
目标文本的 本地文件路径,UTF-8编码,最大长度4096汉字 TEXT_FILEPATH = "【您的测试文本地址,例如:./example.txt】"	

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	分类结果数组
+name	否	string	分类名称
+score	否	number	置信度

-

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或在百度 云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问 请在百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大4096 UTF-8字符。
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云 控制台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

の调用API

接口描述

基于自定义训练出的文本分类模型,实现个性化文本识别。模型训练完毕后发布可获得定制化文本分类API 详情访问:EasyDL 文本分类·单标签 进行训练。

更多训练模型过程中的常见问题请查看常见问题文档。

如有其它问题,请在百度云控制台内提交工单反馈。

请求说明

- 请求示例
- HTTP 方法: POST

请求URL:请首先在EasyDL 文本分类-单标签进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值

Content-Type application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,定制化文本分类服务以json方式请求。

Body请求示例:

```
{
"text": "<UTF-8编码文本>",
"top_num": 6
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text	是	string	-	文本数据,UTF-8编码。最大长度4096汉字
top_num	否	number	-	返回分类数量,默认为6个

请求示例代码

Python3	
нин	
EasyDL 文本分类单标签 调用模型公有云API Python3实现 """	
import json	
import base64	
import requests	
""	
使用 requests 库发送请求	
使用 pip(或者 pip3)检查我的 python3 环境是否安装了该库,执行命令	
pip freeze grep requests	
若返回值为空,则安装该库	
nip install requests	
目标文本的 本地文件路径,UTF-8编码,最大长度4096汉字	

TEXT_FILEPATH = "【您的测试文本地址,例如:./example.txt】"

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	分类结果数组
+name	否	string	分类名称
+score	否	number	置信度

-

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复 制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或在百度 云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问 请在百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大4096 UTF-8字符。
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云 控制台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

の 离线API

の 发布API

在训练模型时,您需要选择「EasyEdge本地部署」的训练方式,才能发布本地部署的私有API。

私有API介绍

将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷

发布私有API的流程训练完毕后,您可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可将模型部署到私有服务器:

1. 在「发布模型」页面中,选择模型及模型的版本,选择部署方式为「EasyEdge本地部署」、集成方式为「API-纯离线服务」。点击「发布」, 即可跳转至「发布新服务」页面

	文本分类模型	发布模型		
	品 总览			
	<u>前</u> 模型中心	选择模型	文本标题分类	\sim
	我的模型	部署方式	EasyEdge本地部署	~
	创建模型			
	训练模型	选择版本	V1	~
发布模型页面示意:	校验模型	集成方式	○ API纯离线服务	
	发布模型			
	☑ EasyData数据服务	发布		
	数据总览			
	公开数据集			
	标签组管理			
	在线标注			
	智能标注			

2. 在「发布新服务」页面,选择部署类型,填写服务名称、证书生效时间等信息,选择对应的系统和芯片。

- 部署类型可支持单模型部署和增量部署
- 增量部署申请,指需要在一台服务器上部署多个模型部署包时使用。进行增量部署时,需在「已部署服务」选择同台服务器历史中最近 部署的部署包,此步骤用来关联不同部署包中的license文件
- 3. 上传指纹文件。详细操作见指纹提取工具说明,可通过指纹工具进行指纹的提取
- 4. 点击下一步,填写个人详细信息后即可发布。发布完成后,即可在服务器目录下看到发布处于审核中的状态

个人信息的填写仅供EasyDL团队了解您,便于后续合作接洽,不会作为其他用途使用

5. 等待审核通过,前往「纯离线服务」等待部署包制作完成后,下载部署包,并参考文档完成集成

发布新服务页面示意:

文本分类模型	纯离线服务 > 发布新服务
品 总览	
<u>伯</u> 模型中心	服务器
我的模型	
创建模型	
训练模型	集成方式 O API
校验模型	部署类型 💿 单模型部署 🔘 增量部署
发布模型	
☑ EasyData数据服务	证书 测试证书 ~
数据总览	选择模型 文本标题分类 >>
公开数据集	选择版本 V1 ~
标签组管理	
在线标注	服务名称 *
智能标注	证书生效时间 * 2022/3/31
目 公有云服务	指纹文件 * 上传文件 ③指纹提取工具下载 指纹提取工具使用说明
在线服务	
記 EasyEdge本地部署	选择系统和芯片 👔 🗘 Linux 了解对应硬件
纯离线服务	下一步

价格说明 EasyDL已支持将定制模型部署在本地服务器上,只需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。

如需购买永久使用授权,请提交工单咨询。

つ。调用API

本文档主要说明定制化模型本地部署后,如何使用本地API。如还未训练模型,请先前往EasyDL 文本分类-单标签 进行训练。

如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:868826008)联系群管

部署包使用说明 部署方法

EasyDL定制化文本分类模型的本地部署通过EasyPack实现,目前提供单机一键部署的方式。

在EasyDL申请、下载部署包后,在本地服务器新建目录(建议目录命名规则:easyDL_服务名称_模型版本号),将软件包上传至该目录。请参 考EasyPack-单机一键部署使用python2版本来部署,部署成功后,启动服务,即可调用与在线API功能类似的接口。

运维检查

EasyDL服务器API部署应用健康检查(或故障排查)脚本:trouble_shooting.tar

脚本能力:鉴权服务健康检测、容器状态检查、端口探活、网络联通性测试、容器关键报错日志输出等

使用方法:将脚本上传至服务器任意目录(或在服务器直接下载),并解压后运行。

```
**解压**
tar vxf trouble_shooting.tar
**执行**
bash trouble_shooting.sh
```

本地部署包根据服务器硬件(CPU单机或GPU单卡)进行授权,只能在申请时提交的硬件指纹所属的硬件上使用。

部署包测试期为1个月,如需购买永久授权,可提交工单咨询

API参考

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在EasyDL进行自定义模型训练,完成训练后申请本地部署,本地部署成功后拼接url。

请求URL: http://{IP}:{PORT}/{DEPLOY_NAME}/TextClassification

- IP:服务本地部署所在机器的ip地址
- PORT:服务部署后获取的端口
- DEPLOY_NAME:申请时填写的本地服务名称

Header如下:

参数 值

Content-Type application/json

Body请求示例:

```
{
"text": "<UTF-8编码数据>",
"top_num": 5
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text	是	string	-	文本数据,UTF-8编码。最大长度4096汉字
top_num	否	number	-	返回分类数量,默认为6个

返回说明

```
返回参数
```

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	分类结果数组
+name	否	string	分类名称
+score	否	number	置信度

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如缺少必要出入参时返回:

}

{ "error_code": 336001,

"error_msg": "Invalid Argument"

错误码	错误信息	描述
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(868826008)或工单联系技术支持团队
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或工单联系技 术支持团队
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数代码格式是否有误。有疑问请通过QQ群(868826008)或工单联系 技术支持团队
336003	Base64解码失败	图片/音频/文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有 疑问请通过QQ群(868826008)或工单联系技术支持团队
336004	输入文件大小不合 法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大1024 UTF-8字符。有疑问请通过 QQ群(868826008)或工单联系技术支持团队
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云控制 台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请通过QQ群(868826008) 或工单联系技术支持团队
337000	Auth check failed	离线鉴权调用失败

模型更新/回滚操作说明 模型更新 1、在EasyDL-纯离线服务发布页面,找到您的服务器API发布记录,点击【更新版本】,选择「更新包」或 「完整包」来发布。 两者区别:

包类型	描述
更新包	仅包含最新的模型应用,需执行download.sh脚本下载所需镜像等依赖文件
完整包	包含模型应用和其他鉴权服务,需执行download.sh脚本下载所需完整依赖文件

2、(CPU模型可忽略)如果您训练的模型为GPU版本,系统会生成多份下载链接。请在GPU服务器执行 nvidia-smi命令,根据返回的Cuda Version来选择对应的部署包链接下载。

3、在服务器新建目录(建议标记对应模型的版本号,便于区分不同模型版本),如easydl_\${DEPLOY_NAME}_v2

\${DEPLOY_NAME}:申请时填写的服务名称

以如下场景举例说明:模型版本从V1升级至V2

```
**1.新模型准备**
**创建指定版本的目录**
mkdir easedI_${DEPLOY_NAME}_v2
cd easedI_${DEPLOY_NAME}_v2
**将部署包上传至服务器该目录并解压**
tar zvxf xx.tar.gz
**解压后,进入指定目录执行download脚本下载模型所依赖文件**
cd original && bash download.sh
**2.旧模型备份**
**历史模型备份**
cp -r /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V1
**记录当前模型的端口号**
docker ps -a grep ${DEPLOY_NAME}
**3.模型升级**
cd package/Install
**卸载当前已安装的旧的easyDL服务: ${DEPLOY_NAME},前面已备份**
python2 install.py remove ${DEPLOY_NAME}
**安装当前部署包内新的EasyDL服务: ${DEPLOY_NAME}**
python2 install.py install ${DEPLOY_NAME}
**(可选操作)更新证书**
python2 install.py lu
```

方法一:

重命名当前v2模型目录名称 mv /home/baidu/work/\${DEPLOY_NAME} /home/baidu/work/\${DEPLOY_NAME}_V2 **使用V1版本** cp-r /home/baidu/work/\${DEPLOY_NAME}_V1 /home/baidu/work/\${DEPLOY_NAME} **停止当前模型容器** docker ps -a |grep \${DEPLOY_NAME} docker rm -f \${容器名} **创建新的容器** cd /home/baidu/work/\${DEPLOY_NAME} && bash start/start-1.sh ** (可选操作)进入V1版本部署包所在目录执行license更新操作,假如部署包在/opt目录下,以您实际目录为准** cd /opt/easydl_\${DEPLOY_NAME}_V1/original/package/Install && python2 install.py lu

方法二:

进入模型V1所在目录,参考上述【模型更新】步骤,执行模型升级操作(即先卸载v2,后升级为v1)

の 发布API

在训练模型时,您需要选择「EasyEdge本地部署」的训练方式,才能发布本地部署的私有API。

私有API介绍

将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷

发布私有API的流程训练完毕后,您可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可将模型部署到私有服务器:

1. 在「发布模型」页面中,选择模型及模型的版本,选择部署方式为「EasyEdge本地部署」、集成方式为「API-纯离线服务」。点击「发布」, 即可跳转至「发布新服务」页面

	文Z	本分类模型	€	发布模型		
	 品 協 模型中心 我的模型 	总览				
			选择模型	文本标题分类	\sim	
		我的模型		部署方式	EasyEdge本地部署	~
		创建模型				
		训练模型		选择版本	V1	~
发布模型页面示意:		校验模型		集成方式	○ API纯离线服务	
		发布模型				
	<u></u>	EasyData数据服务		发布		
		数据总览				
		公开数据集				
		标签组管理				
		在线标注				
		智能标注				

- 2. 在「发布新服务」页面,选择部署类型,填写服务名称、证书生效时间等信息,选择对应的系统和芯片。
 - 部署类型可支持单模型部署和增量部署
 - 增量部署申请,指需要在一台服务器上部署多个模型部署包时使用。进行增量部署时,需在「已部署服务」选择同台服务器历史中最近 部署的部署包,此步骤用来关联不同部署包中的license文件
- 3. 上传指纹文件。详细操作见指纹提取工具说明,可通过指纹工具进行指纹的提取

4. 点击下一步,填写个人详细信息后即可发布。发布完成后,即可在服务器目录下看到发布处于审核中的状态

个人信息的填写仅供EasyDL团队了解您,便于后续合作接洽,不会作为其他用途使用

5. 等待审核通过,前往「纯离线服务」等待部署包制作完成后,下载部署包,并参考文档完成集成

发布新服务页面示意:

文本分类模型	纯离线服务 > 发布新服务
品 描照中心	服条哭
我的模型	
创建模型	
训练模型	集成方式 O API
校验模型	部署类型 💿 单模型部署 💿 增量部署
友布模型 図 EasyData数据服务	证书 测试证书 ~
数据总览	选择模型 文本标题分类 🗸
公开数据集	选择版本 V1 ~
标签组管理	服务名称 *
在线标注智能标注	证书生效时间 * 2022/3/31 mm
目 公有云服务	指纹文件 * 上传文件 ③ 指纹提取工具下载 指纹提取工具使用说明
在线服务	
22 EasyEdge本地部署	ZEJŦ示SUTAUC/T 上 ▲▲ Linux 了解对应硬件
纯离线服务	下一步

价格说明 EasyDL已支持将定制模型部署在本地服务器上,只需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。

如需购买永久使用授权,请提交工单咨询。

常见问题

心训练相关问题

数据处理失败或者状态异常怎么办?

如是是文本分类模型上传处理失败,请先检查已上传的分类命名是否正确,是否存在特殊字符、或者增加了空格(标签仅支持中英文数字及下划 线,长度不超过256字符);然后检查上传的数据文本量是否超过上限(10万);再检查文本是否有损坏。如果自查没有发现问题请在百度云控 制台内提交工单反馈。

模型训练失败怎么办?

如果遇到模型训练失败的情况,请在百度云控制台内提交工单反馈。

已经上线的模型还可以继续优化吗?

已经上线的模型依然可以持续优化,操作上还是按照标准流程在训练模型中-选择要优化的模型和数据完成训练,然后在模型列表中更新线上服务,完成模型的优化

Step 1 重新训练 点击我的模型列表——找到需要重新训练的模型——点击训练,进行新版本模型训练

Step 2 重新发布 点击我的模型列表——找到新训练好的模型版本——点击申请发布

【图像分类】	百美椅子训	练 模型ID:230)			全部版本 删除
应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作
云服务	V2	训练完成	未申请	未发布	top1准确率87.61% top5准确率100.00% 完整评估效果	申请发布校验训练
离线SDK	V1	训练完成	未申请	未发布	top1准确率85.84% top5准确率100.00% 完整评估效果	申请发布 训练
						每页显示 12 🗸 < 1 >

Step 3 确认发布 在出现的弹窗中点击确定

我的模型	完整评估效果	
创建模型	申请发布 ×	
训练模型	(1) 当前模型的线上服务为https://aip.baidubce.com/rpc/2.0/ai_custom/v1/classification/model_172,相应的模型版本为V1。	全部版本 删除
校验模型	是否确认要替换为V2	
发布模型	砌定取消	练
次据中 心	【図ゆ公米】 埼子語会辺別 埼井川り・2508	全部版本 删除
数据集管理		T There we wanted

∞ 模型效果相关问题

模型效果怎么调优?

如果对模型效果不满意,您可以先查看训练数据是否和实际场景要分类的文本一致,以及训练数据量是否太少。

如果训练数据量已经达到一定丰富度(如单个分类/标签的文本量超过2000条以上),效果仍然不佳,请在百度云控制台内提交工单反馈。

○ 模型上线相关问题

希望加急上线怎么处理?

请在百度云控制台内提交工单反馈。

每个账号可以上线几个模型?是否可以删除已上线的模型?

不限制发布模型数量,已上线模型无法删除

心 模型部署相关问题

模型发布公有云部署后是否收费?调用量不够怎么办?

目前EasyDL全部文本任务均提供部分免费调用额度,并支持付费购买额外的调用额度。详情请参考:EasyDL文本价格说明

模型能否支持私有化部署?

目前EasyDL已支持将定制模型部署在私有服务器上,只需在发布模型时提交私有服务器部署申请,通过审核后即可获得一个月免费试用,并支持 在控制台在线按设备使用年限购买授权。详情请参考:文本私有服务器部署价格说明

申请发布模型审核不通过都是什么原因?

可能的原因包括:

1、经过电话沟通当前模型存在一些问题或者不再使用,如训练数据异常、数据量不够、不想再继续使用等原因,沟通达成一致拒绝。

2、电话未接通且模型效果较差,会直接拒绝。

如果需要申诉,请在百度云控制台内提交工单反馈。

文本分类-多标签

整体介绍

心简介

Hi,您好,欢迎使用百度EasyDL定制化训练和服务平台。

定制文本分类的模型,是基于自建分类体系的机器学习方法,可实现文本按内容类型做自动分类。平台目前提供的文本分类模型包括:文本分类 (单标签)和文本分类(多标签)两种模型类型,请您根据自己的业务场景来选择合适的模型。本文介绍的是关于**文本分类(多标签)**的模型介 绍。

文本分类(单标签)场景:如您对网络文章进行舆情分析,判断舆情是正向评价还是负向评价,即每条文本**仅有一个**分类标准,此问题属于单标 签的文本分类场景;

文本分类(多标签)场景:如您对网络文章进行板块划分,即每条文本有**两个及以上**分类标准,文章可能属于娱乐、国际、生活等多个标签,则 可使用多标签的文本分类模型

更多详情访问: EasyDL自然语言处理方向

心应用场景

1、新闻分类:定制训练媒体文章文本的自动分类,识别文章所属的一个或多个领域标签

2、商品名称分类:定制训练商品名称的分类模型,识别商品所属的一个或多个品类

3、其他:尽情脑洞大开,训练你希望实现的文本分类多标签的模型

心 技术特色

文本分类模型内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等方面 的新知识,实现模型效果不断进化。

文心大模型是百度发布的产业级知识增强大模型,是千行百业AI开发的首选基座大模型。文心大模型既包含基础通用的大模型,也包含了面向重 点领域和重点任务的大模型,还提供丰富的工具与平台,支撑企业与开发者进行高效便捷的应用开发。"知识增强"是文心的核心特色,文心能够 同时从大规模知识和海量多元数据中持续学习,如同站在巨人的肩膀上,训练效率和理解准确率都得到大幅提升,并具备了更好的可解释性。

心 使用流程

训练模型的基本流程如下图所示,全程可视化简易操作,在数据已经准备好的情况下,最快15分钟即可获得定制模型。



数据准备

心 创建数据集并导入

创建数据集

在训练模型之前,需要创建数据集。需输入数据集名称、选择相应的标注模版、选择数据去重策略,即可创建一个空数据集。

ーー・ ど桨EasyDL	产品介绍	3 操作平台	应用案例	使用文档			
文本分类模型	÷	我的数据总览					
88 总览							
凸 模型中心		EasyData智能数	据服务平台已上约	线,使用EasyData可	享受包括多人标注、数据	清洗、数据采集等完整数据服务 立即前往	
我的模型		创建数据集					
创建模型							
训练模型		pf文本分类测试	区 数据集组ID	: 306469			
校验模型		版本	数据集ID		数据量	最近导入状态	标注类型
发布模型			000707				
☑ EasyData数据服务		VI 😁	330727		0	● 守八大政 ⑦	义平万尖
数据总览		ten 立木_名标签	「二 数据集组Ⅱ	D: 250032			
标签组管理		131_大华-罗彻亚	山城市未知	200032			
在线标注		版本	数据集ID		数据量	最近导入状态	标注类型

数据自动去重即平台对您上传的数据进行重复样本的去重。建议创建数据集时选择「数据自动去重」

导入数据 创建数据集后,在「数据总览」页面中,找到该数据集,点击右侧操作列下的「导入」,即可进入导入数据页面。

<i>- خابخ</i> ど桨EasyDL	产品介绍	绍 操作平台	应用案例 使用	文档	
文本分类模型	€	我的数据总览 >	pf文本分类测试/V1/导》	λ.	
88 总览					
曲 模型中心		┃创建信息 ∨			
我的模型		数据集ID	330727	版本号	V1
创建模型		备注			
训练模型		┃标注信息 >			
校验模型		标注类型	文本分类	标注模板	短文本多标签
发布模型		数据总量	0	已标注	0
☑ EasyData数据服务		标签个数	0	待确认	0
数据总览		大小	0M		
标签组管理		数据清洗			
在线标注		暂未做过数据清	洗任务		
目 公有云服务		导入数据			
在线服务		数据标注状态	无标注信息	○ 有标注信息	Į.
ิ EasyEdge本地部署		导入方式	请选择		^
纯离线服务			本地导入		
			BOS目录导入		
			分享链接导入		
			平台已有数据集		

您可以使用4种方案上传文本分类的数据,分别为:

- 本地导入
- BOS目录导入
- 分享链接导入
- 平台已有数据集

本地导入您可以通过以下三种方式进行本地数据的导入:

1	ど桨EasyDL	产品介绍	操作平台	应用案例	使用文档	
文Z	本分类模型	Ē	我的数据总览	〉 pf文本分类测试/V1	/导入	
	总览					
đ	模型中心		┃ 创建信息 ∨			
	我的模型		数据集ID	330727	版本号	V1
	创建模型		备注			
	训练模型		┃标注信息 〜			
	校验模型		标注类型	文本分类	标注模板	短文本多标签
	发布模型		数据总量	0	已标注	0
27	EasyData数据服务		标签个数	0	待确认	0
	数据总览		大小	0M		
	标签组管理		┃数据清洗			
	在线标注		暫未做过数据》	青洗任务		
8	公有云服务		┃ 导入数据			
	在线服务		数据标注状态	 无标注信 	息 🦳 🦳 有标注信	息
12	EasyEdge本地部署		导入方式	本地导入		~ 请选择 ^
	纯离线服务			确认并返回	2	上传Excel文件
						上传TXT文本
						上传压缩包

- 以压缩包的方式上传
- 以TXT文本文件方式上传
- 以Excel文件的方式上传 以压缩包方式上传
- 文本文件的编码方式: UTF-8, 每个文本文件最长不能超过4096个汉字(字符)
- 压缩包仅支持zip格式;大小需要在5GB以内;

注意,如果您上传的数据是带有标注信息的数据,则需要在压缩包里的创建文件夹,文件夹名即是标签名,只能包含数字/字母/下划线,一个样本有多个标签,则从属于多个文件夹。例如"北京明天气温骤降,请注意保暖"的文本文件同时存在于"北京本地"和"天气"两个文件夹下。

以TXT文本文件上传

- 每行样本最长不能超过4096个汉字(字符),文件编码方式:UTF-8
- txt文件内的标注数据格式要求为"文本内容\t标注标签\t...标注标签\t\n"(\t代表tab制表符,\n代表回车换行),如果是无标注信息的数据,则每行只有文本内容即可

以Excel文件上传

- Excel文件内数据格式要求为:使用第一列作为待标注文本,第二列作为标注信息列(此列仅支持数字或字母),每行是一组样本,每组数据 文本内容的字符数建议不超过4096,超出将被截断。
- 文件类型支持xlsx格式,单次上传限制100个文件

BOS目录导入 需选择Bucket地址与对应的文件夹地址。

请确保将全部文本已通过txt文件保存至同一层文件目录,该层目录下子文件目录及非相关内容(包括压缩包格式等)不导入。

分享链接导入需输入链接地址。分享链接导入的要求如下:

• 仅支持来自百度BOS、阿里OSS、腾讯COS、华为OBS的共享链接

平台已有数据集

• 导入无标注数据时,选择需要导入的数据集名称,可导入其不带标注的全部数据,或未标注的数据

• 导入已标注数据时,选择需要导入的数据集名称,可导入其某个或全部标签下的数据

准备数据集的技巧 文本分类任务中,可参考以下准备数据集的技巧:设计分类

设计整个数据集的分类体系,即抽象出文本所需识别的标签,标签也是你希望识别出的结果。例如娱乐新闻的内容类型,则可以以"男星"、"大陆"、"港台"、"童星"等分别作为标签体系。

注意:目前单个模型的上限为1000类,如果要超过这个量级请在百度云控制台内提交工单反馈

数据量

基于设计好的分类标签准备文本数据,每个标签建议至少需要准备50个以上的样本,如果想要较好的效果,建议1000-10000个文本样本,如果 某些分类的文本具有相似性,需要增加更多文本。

文本的基本格式要求: 目前文本文件类型支持txt,文本文件大小限制长度最大4096,格式为UTF-8字符。一个模型的文本总量限制100万个文本 文件。

数据分布

- 训练集文本需要和实际场景要识别的文本环境一致
- 考虑实际应用场景的种种可能性,每个分类的文本需要覆盖实际场景里面存在的可能性,训练集若能覆盖的场景越多,模型的泛化能力则越强

可能的疑问

• 如果训练文本数据无法全部覆盖实际场景要识别的文本,怎么办?

答:训练的模型算法会有一定的泛化能力,尽可能覆盖即可。

• 多语种模型支持全球94种语言:

南非语,阿姆哈拉语,阿拉伯语,阿萨姆语,阿塞拜疆语,白俄罗斯语,保加利亚语,孟加拉语,孟加拉语(拉丁化),布列塔尼语,波斯尼亚语,加泰隆语, 捷克语,威尔士语,丹麦语,德语,希腊语,英语,世界语,西班牙语,爱沙尼亚语,巴斯克语,波斯语,芬兰语,法语,弗里斯兰语,爱尔兰语,苏格兰盖 尔语,加利西亚语,古吉拉特语,希伯来语,印地语,印地语(拉丁化),克罗地亚语,匈牙利语,亚美尼亚语,印尼语,冰岛语,意大利语,日语,爪哇语, 格鲁吉亚语,哈萨克语,高棉语,康纳达语,韩语,库尔德语,柯尔克孜语,拉丁语,老挝语,立陶宛语,拉脱维亚语,马拉加斯语,马其顿语,马拉亚拉 姆语,蒙古语,马拉提语,马来语,缅甸语,尼泊尔语,荷兰语,挪威语,奥里亚语,旁遮普语,巴利语,普什图语,葡萄牙语,罗马尼亚语,俄语,梵语,信 德语,僧伽罗语,斯洛伐克语,斯洛文尼亚语,索马里语,阿尔巴尼亚语,塞尔维亚语,巽他语,瑞典语,斯瓦希里语,泰米尔语,泰米尔语(拉丁化),泰 卢固语,泰卢固语(拉丁化),泰语,他加禄语,土耳其语,维吾尔语,乌克兰语,乌尔都语,乌尔都语(拉丁化),乌兹别克斯坦语,越南语,意第绪语。

如果需要寻求第三方数据采集团队协助数据采集,请在百度云控制台内提交工单反馈

心 在线标注

在线标注

Step 1 进入标注页面 上传未标注的数据后,可以通过以下两个方式进入标注页面:

- 在「数据总览」页面,该数据集对应的操作列下,点击「标注」,即可进入标注页面
- 在「在线标注」页面,选择该数据集,即可进入标注页面
- **Step 2 进行文本标注** 针对尚未进行标注的数据,通过以下方式进行标注:
- 在右边标签栏中添加标签
- 针对文本内容,选择其对应的标签(可进行多选)
- 点击下一篇,此篇文本的内容即可进行自动保存,且您将开始对下一篇文本进行标注

针对已进行标注的数据,通过以下方式进行标注修改:

- 进入需修改标签的文本的标注页面,选择右边标签栏中的标签 (可进行多选)
- 点击下一篇,对此篇文本标签的修改即可进行自动保存

EasyDL 文本使用说明

添加标签 🔍

N7) 无标注信息(0) 有标注信息(7)	标注示例	标签栏
标签: 0, 18, 7, 3, 1 ① 删除文本 上一篇	下一篇	请输入标签名称
面密铁铁,2014年12月11日,原,被查在民股局协议周期。协议的定规量子考某某甲的监护权归被告,由其托务,然后,周期至今,李某某甲一直不愿与被查生活,由原告实际抗胃,被告未尽财任何托务义务,原告曾多次 各尽具作为父亲对子女囚忌的生活,教育联础重新议关并发付抚育贵,可说告认将种理由抵托,使法子的身心受到极大的伤害,最生子已年满十两岁,并且要未完愿意照顾原告主活,居包又有抚胃能力,现要求时令要更最生 某某年间抚养关车面描音抚养,就想每月发付抚着那心巧正是其早新十八周岁上。	歷求被 子李	8
		7
		18
		3
		1
		0

Step 3 查看标注信息 通过以下方式查看已标注的文本信息:

- 在「数据总览」页面,该数据集对应的操作列下,点击「查看」,进入查看标注页面后,点击「有标注信息」
- 通过选择左侧标签中的不同标签名称,即可查看不同标签下的文本数据

tj-yunfan-text	v1版本的文本列表 育选 🗸				回批量删除
全部标签(6	() 输入标签名称 Q	序号	文本內容損要	操作	
标签名称	数据量 操作	1	设施很不错: 周边环境不错,紧邻金沙滩: 推荐 海景房 不是:比较编: 吃饭不放便; 流后白带的银行性价比太差。价格太贵,口味不好;	查看	删除
test_5	🖄 1178 副除 导力	2	图它小巧等着方便 发热量不是很大 有点珍珠白的颜色,很喜欢, 镭盒按起来很舒服,触摸板银人性化的。	宣看	#12\$
test_4	🖂 1178 副除 导》	3	硬盒声音大,好像是deli这批产品都有这个问题,E6400RE5400都有硬盒声音大的问题。	查看	删除
test_3	🗹 1178 🏾 删除 导力	4	做工不错,主要事价钱实在、2400XT,看了网上的笔记本显卡排行,还是比较靠前的,跟9300有的比,于是就选了它。买之前看了3星的,3400带9200CS逸显的那款,要3999,只多了…	查看	#108
test_2	区 1178 副除 导入	5	仔细回想这本文集,发现自己喜欢的只是写《数室额南,没有风筝》的麻宁,不知道是地成长了还是自己止步不前了,没有感觉了,或许那是她对生活的香ヤ,但却不能让我感动了,觉得…	查看	删除
test_1	🖂 1178 🏾 删除 导力	6	我称天收到的书,质量还不错,而且配有光盘,光盘质量也还较清晰。我以前也学过一点瑜珈,看了书中描写,还以前学得差不多,但该是比较适合初学者的,总体难度不是很大。光盘的…	查看	#128
test_0	🖂 1178 删除 导之	7	性价比非常好的一款,操荐购买! 自己又添了一根2G的条子,用起来超真!	查看	删除
		8	住这个酒店夹在是太享受了,不仅可以使用五彩缤纷的白毛巾,还可以免费听赏别人KTV包厢里的高音演奏… 以下是某某歌奏家的话谢谢 谢谢大家 首先我要感谢我的父母 还要感谢背后支	宣看	#12\$
		9	我家宝贝的窖窖长期在外工作,很少在家,宝宝一岁半了,他好像还没找到窖窖窑的感觉。我买这本书就最想救宝宝要学会和窖窖相处,载地窖窖带一点"爆论",让他知道他这个窖窖的角…	宣看	删除
	< 1 / 1 >	10	自己马上就有宝宝了,期待看宝宝婵临人世,所以提脱看着家数方面的书籍。看了这本书,真的很好。不但丛中可以学到数算孩子的方法,自己也丛中收获很多,犹太民族经历了那么多的…	查看	#178

收据集标注模板:短文本单标签,该标签有标注信息文本:1178个

心 数据去重

重复样本的定义

一个样本包括文本内容和标签。重复样本的定义,是指您上传的数据中,存在两个样本的文本内容完全一致。则被判定为两个样本是重复样本。 例如:

文本内容	标签
未来的学和教正在改变,学生将会在家里学习,机器人将走上讲台。	education/science
未来的学和教正在改变,学生将会在家里学习,机器人将走上讲台。	education/science
未来的学和教正在改变,学生将会在家里学习,机器人将走上讲台。	Al/robot

上表三个样本均为重复样本,后两个样本虽然标签不一,但文本内容一致,也为重复样本。

Tips: "如何利用好重复样本",如果您在模型训练过程中,需要通过增加某个类别标签的预测权重,可以通过增加此标签的重复样本来达到此目标。

平台去重策略

平台提供了可去重的数据集,即对您上传的数据进行重复样本的去重。注意:当您确定了数据集为去重或非去重的属性后,便不可修改。

当您创建了一个去重的数据集时,在后续上传数据的过程中,平台可通过检验您当前上传的样本与已上传到此数据集下的样本是否相同,如果相同,则会使用新的样本替代旧的样本。此时分为几种情况,如下:

- 1. 数据集中有未标注样本,上传重复的已标注样本,此时未标注样本将被覆盖
- 2. 数据集中有已标注样本,上传重复的未标注样本,此时已标注样本将被覆盖
- 3. 数据集中有已标注样本,上传不同标注的已标注样本,此时已有的标注样本将被覆盖

の API上传

本文档主要说明当您线下已有大量的已经完成分类整理的文本数据,如何通过调用API完成文本数据的便捷上传和管理。

EasyDL数据集管理API在管理不同模型数据类型之间是通用的。上传不同模型类型数据,只是在部分接口入参存在差异,使用及接口地址完全一致。

数据集创建API

接口描述

该接口可用于创建数据集。

接口鉴权

同发布模型后获取的API鉴权方式:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/create

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION,TEXT_CLASSIFICATION_MUL 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类(单标签) 、文本分类(多标签)
dataset_ name	是	string	数据集名称,长度不超过20个utf-8字符

若上传声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
dataset_id	否	number	创建的数据集ID

查看数据集列表API

接口描述

该接口可用于查看数据集列表。返回数据集的名称、类型、状态等信息。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key
请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/list

URL参数:

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION, TEXT_CLASSIFICATION_MUL 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类(单标签)、文本分类(多标签)
start	否	number	起始序号,默认为0
num	否	number	数量,默认20,最多100

若查看声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	数据集总数
results	否	array(object)	数据集列表
+dataset_id	否	number	数据集ID
+dataset_name	否	string	数据集名称
+type	否	string	数据集类型
+status	否	string	数据集状态
+special_status	否	string	数据集特殊状态,包括shared、smart和空值,分别表示共享中、智能标注中、非特殊状态

查看分类 (标签)列表API

接口描述

该接口可用于查看分类(标签)。返回分类(标签)的名称、包含数据量等信息。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/list

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:



Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION, TEXT_CLASSIFICATION_MUL 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类(单标签)、文本分类(多标签)
dataset _id	是	number	数据集ID
start	否	number	起始序号,默认O
num	否	number	数量,默认20,最多100

若查看声音分类的全部分类,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	标签总数
results	否	array(object)	标签列表
+label_id	否	string	标签/分类ID
+label_name	否	string	标签/分类名称
+entity_count	否	number	图片/声音/文本数量

添加数据API

接口描述

该接口可用于在指定数据集添加数据。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/addentity

URL参数:

参数 值 access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION, TEXT_CLASSIFICATION_MUL 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类(单标签)、文本分类(多标签)
dataset _id	是	number	数据集ID
append Label	否	boolea n	确定添加标签/分类的行为:追加(true)、替换(false)。默认为追加(true)。
entity_c ontent	是	string	type为 IMAGE_CLASSIFICATION/OBJECT_DETECTION/IMAGE_SEGMENTATION/SOUND_CLASSIFICATION时,填入图 片/声音的base64编码;type为TEXT_CLASSIFICATION或 TEXT_CLASSIFICATION_MUL时,填入utf-8编码的文本。内容 限制为:图像分类base64前10M;物体检测base64前10M;图像分割base64前10M;声音分类base64前4M,声音 时长1~15秒;文本分类单标签和文本分类多标签4096个汉字
entity_n ame	是	string	文件名
labels	是	array(o bject)	标签/分类数据
+label_ name	是	string	标签/分类名称(由数字、字母、中划线、下划线组成),长度限制20B
+left	否	number	物体检测时需给出,标注框左上角到图片左边界的距离(像素)
+top	否	number	物体检测时需给出,标注框左上角到图片上边界的距离(像素)
+width	否	number	物体检测时需给出,标注框的宽度(像素)
+height	否	number	物体检测时需给出,标注框的高度(像素)

若上传声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

数据集删除API

接口描述

该接口可用于删除数据集。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/delete

URL参数:

参数	值
----	---

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

application/jeen

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION, TEXT_CLASSIFICATION_MUL 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类(单标签)、文本分类(多标签)
dataset _id	是	number	数据集ID

若刪除声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id ,用于问题定位

分类 (标签) 删除API

接口描述

该接口可用于删除分类(标签)。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/delete

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION, TEXT_CLASSIFICATION_MUL 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类(单标签)、文本分类(多标签)
dataset _id	是	number	数据集ID
label_n ame	是	string	标签/分类名称

若删除声音分类的子类,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求 ,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量 阶梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计 费;未上线计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶 梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access token参数,请检查后重新尝试
100	· · [· · · · · ·	
110	Access token invalid or no longer valid	access_token无效
110 111	Access token invalid or no longer valid Access token expired	access_token无效 access token过期
110 111 406000	Access token invalid or no longer valid Access token expired internal server error	access_token无效 access token过期 服务错误
110 111 406000 406001	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数
110 111 406000 406001 406002	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid dataset not exist	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数 数据集不存在
110 111 406000 406001 406002 406003	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid dataset not exist dataset already exists	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数 数据集不存在 数据集已存在
110 111 406000 406001 406002 406003 406004	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid dataset not exist dataset already exists dataset can not be modified temporarily	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数 数据集不存在 数据集已存在 数据集暂不可修改
110 111 406000 406001 406002 406003 406004 406005	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid dataset not exist dataset already exists dataset can not be modified temporarily label not exist	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数 数据集不存在 数据集已存在 数据集暂不可修改 标签/分类不存在
110 111 406000 406001 406002 406003 406004 406005	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid dataset not exist dataset already exists dataset can not be modified temporarily label not exist no permission to modify the dataset	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数 数据集不存在 数据集智不可修改 标签/分类不存在 没有修改数据集的权限
110 111 406000 406001 406002 406003 406004 406005 406006	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid dataset not exist dataset already exists dataset can not be modified temporarily label not exist no permission to modify the dataset dataset cannot be modified while smart annotation is running	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数 数据集不存在 数据集营不可修改 标签/分类不存在 没有修改数据集的权限 智能标注期间不可修改数据集

模型训练

の 效果优化

通过模型迭代、检查并优化训练数据、选择高精度模型等方法,能够提升模型效果。**模型迭代**

一个模型很难一次性就训练到最佳的效果,通常会需要结合模型评估报告和校验结果不断扩充数据和调优。

为此平台提供了模型迭代功能,即当模型训练完毕后,会生成一个最新的版本号,首次V1、之后V2……以此类推。可以通过调整训练数据和算法,多次训练,以获得适合业务需求的模型效果。

如果模型已经是上线状态,依然支持模型迭代,只是需要在训练完毕后更新线上服务接口,在接口地址不变的情况下可以持续优化效果。

检查并优化训练数据

- 检查是否存在训练数据过少的情况,建议每个类别的文本数量不少于1000个,如果低于这个量级建议扩充
- 检查不同类别的数据量是否均衡,建议不同分类的数据量级相同,并尽量接近,如果有的类别数据量很高,有的类别数据量较低,会影响模型 整体的识别效果
- 通过模型效果评估报告中的各分类的详细评估指标,有针对性地扩充训练数据。例如下图中评估报告显示有20个标签数,但部分标签(19、 17、15等)精确值极低,即可考虑从两个方向进行针对性优化,一是增加相应标签的数据集样本数据量,二是检查现有相应标签数据集中是

否存在定义模糊的情况,提升标签数据集质量,以此达到优化模型的效果。

ナシン・ご茶 EasyDL 产品介	公 操作平台	应用案例 使用文	【档			
文本分类模型 4 三	我的模型 > pf文	本分类多标签测试–V1模型	』评估报告			
88 总览	文本数 1611	标签数 20	训练时长	9分钟		
曲 模型中心						
我的模型	整体评估					
创建模型	- pf文本分类多标签测	则试 V1整体效果欠佳。	如何优化效果?			
训练模型	F1-score	0				精确率 ⑦
校验模型	45.7%	6				49.3%
发布模型						
☑ EasyData数据服务	详细评估					
数据总览	评估样本top5标	签的召回分布				
标签组管理						
在线标注		随机测试集				
目 公有云服务		9: 22				
在线服务		6: 32				
ิ EasyEdge本地部署	召回表现	0: 244				
纯离线服务		3: 46				
		5: 42				
	冬公米的精确值			冬公米的F1	Score	
			100.0%	10	1	0.0%
	19		0.0%	17		0.0%
	8		92.9%	15		0.0%
	10		0.0%	13		0.0%
	17		0.0%	16		0.0%
	15		0.0%	11		25.0%

 检查测试模型的数据与训练数据的文本类型与风格是否一致,如果不一致,那么很可能会存在模型效果不错但实际测试效果较差的情况。针对 这种情况建议重新调整训练集,将训练数据与实际业务场景数据尽可能一致

选择高精度模型

在训练模型时,选择高精度的模型,将提升模型的预测准确率。

「高精度」算法内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等方 面的新知识,实现模型效果不断进化。

心 评估效果

模型评估模型训练完成后,可以在「我的模型」列表中查看该模型的效果,以及完整评估结果。

文本分类模型 (三	我的模型 > 文本标题分类VI模型评估报告			报交工单
28 BX				
直 模型中心	部署方式 EasyEdge復身發酵署 > 版本 >	n ~		
我的模型	文本数 19430 分类数 5 训练时	2小时20分钟		
创建模型				
训练模型	整体评估			
校验模型	文本标题分类 VI效果优异。 如何优化效果?			
发布模型	() 准确率 ()	FI-score ®	() HINE ()	() 2回車 ⑦
☑ EasyData数据服务	96.8%	96.7%	96.7%	96.8%
数据总定				
公开数据集	详细评估			
标近组管理	评估样本具体数据情况			
在线标注	随机测试集			
智能标注	正确数量: 2821			
B 公有云服务	情误数量: 93			
在线板的 図 EasyEdge本地部署	各分类的精确值	告分类的F1-Score		
经累结服务	封经 97.9%	财经 97.1%		
	B7* 98.0%	房产 97.2%		
	教育 97.2%	80 B		
	Nite 97.2%	移放 96.1%		
	社会 94.0%	社会 95.3%		
	各分类的召田值			
	R18 97.1%			
	R/* 96.4%			
	RW 98.9%			
	NIE 96.0%			
	N.O			

「完整评估结果」页面中将记录整体评估与详细评估的报告,包括该模型整体的准确率、F1-score、精确率、召回率,以及评估样本具体数据情况,各分类的精确值、F1-Score等指标。

整体评估中,各指标的释义如下:

- 准确率:正确分类的样本数与总样本数之比
- F1-score:给每个类别相同的权重,计算每个类别的F1-score,然后求平均值
- 精确率:给每个类别相同的权重,计算每个类别的精确率,然后求平均值
- 召回率:给每个类别相同的权重,计算每个类别的召回率,然后求平均值

如果单个标签的文本量在100条以内,会影响评估指标的科学有效性,请确保提交的训练数据中每个标签的数据量

模型校验

在完成训练后,发布模型前,可以先进行模型校验,以确保模型在实际环境中能获得预期的性能。操作方法如下:

- 1. 在左侧「模型中心」目录中点击「校验模型」,进入校验模型页面
- 2. 选择需要校验能力的模型、部署方式、版本,点击「启动模型校验服务」
- 3. 校验服务启动后,在左侧输入文本,点击「校验」后,右侧识别结果栏将输出预测结果
 - 校验数据支持两种输入方式,直接输入文本或上传txt格式文本,文本长度上限为512汉字
 - 在识别结果栏中,可进行阈值的调整。置信度在阈值以下的预测结果将不予显示。例如,阈值调整为0.85时,置信度小于85%的预测结果将不予显示。各类标签的置信度均小于85%时,识别结果栏将显示为:"没有满足条件的识别结果"
- 4. 若您对预测结果满意,可点击「申请上线」,进行模型的发布

如果单个标签的文本量在100条以内,会影响评估指标的科学有效性,请确保提交的训练数据中每个标签的数据量

⊙发起训练

训练模型

完成数据的标注,或提交已标注的数据后,即可在「模型中心」目录中点击「训练模型」,开始模型的训练。

按以下步骤操作,启动模型训练:

文本分类模型 《三	训练模型								
38 A.C.									
由 模型中心	选择模型 文本标题分类 ~								
我的模型	10(682)置								
创建模型	部署方式 公告云部署 ExeyEdge本地錄著 如何选择起著方式?								
0185-40世	选择算法 💿 高精度 🛈 🔷 高性能 🕲	选择测法 ● 离魏度 ◎ 〕 激性能 ◎							
校验模型	模型等选指标 ③ 模型兼顾Precision和R > ③								
发布模型	添加数据								
☑ EasyData数据服务									
数据总定	添加款据集 + 请选择	点如资源集 + 诵选择							
公开数据集	自定义测试集 🗇 📃								
标签细管理	训练环境 杂胎 塑棉	算力	速度比例	价格					
在线标注									
智能标注	 GPU P40 TeslaGPU_P40_24G显存单 	■卡_12核CPU_40G内存 12 TeraFLOPS	5 1	免费					
日 公有云服务	O GPU V100 TeslaGPU_V100_16G显存해	单卡_12核CPU_56G内存 14 TeraFLOPS	1.5	单卡¥0.45/分钟 (348.62小时*节点小时包余额)					
在线服务	37.56-3142								
盟 EasyEdge本地部署	() FALMEND								
纯高线服务									

Step 1 选择模型 选择此次训练的模型 Step 2 训练配置

部署方式

可选择「公有云部署」、「EasyEdge本地部署」。

如何选择部署方式

选择设备

- 如果您选择了「EasyEdge本地部署」,请根据实际部署设备选择
- 如果您选择了「公有云部署」,无需选择设备

选择算法

您可以根据训练的需求,选择「高精度」或「高性能」算法。不同的算法将影响训练时间、预测速度与模型准确率。

- 如果您选择了高精度的模型,模型预测准确率将更高。如果您手中的标注数据集样本较少(例如少于1000条),可选择「高精度」的算法。
 使用高精度的算法训练模型将会耗时更久,实验环境下1000个样本,预计在20分钟左右完成训练
- 如果您选择了高性能的模型,相同训练数据量的情况下,训练耗时更短,模型预测速度更快。使用10000条训练样本,将在15min内完成训练.
 同样的数据量情况下,效果比高精度的模型4-5%

「高精度」算法内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义 等方面的新知识,实现模型效果不断进化

模型筛选指标

选择不同的模型选择方式,对应的模型各项效果指标将有所不同。如果没有特殊场景的要求,使用默认即可(兼顾Precision精确度和Recall召回率)。有以下指标可供选择:

- 模型兼顾Precision和Recall:挑选模型时,兼顾Precision精确度和Recall召回率,如场景中没有对精度或召回的特别要求,建议您使用此默认指标
- Precision最高的模型:挑选模型时,优先挑选Precision精度最高的模型作为部署模型
- Recall最多的模型:挑选模型时,优先选择召回率最高的模型作为部署模型
- ACC最大的模型:挑选模型时,优先挑选预测样本数最多的模型作为部署模型
- Loss最小的模型:挑选模型时,优先挑选预测偏差最小的模型

Step 3 添加数据

添加训练数据

- 先选择数据集,再按标签选择数据集里的文本,可从多个数据集选择文本。注意,文本分类多标签模型至少需要选择2个及以上分类
- 训练时间与数据量大小和您选择的模型类型有关,如果您选择的是高性能的模型,使用10000条训练样本将在5min内完成训练;如果您选择的是高精度的模型,使用10000条训练样本,将在60-100min完成训练

添加自定义测试集

上传不包含在训练集的测试数据,可获得更客观的模型效果评估结果。

添加自定义测试集的目的:

如果学生的期末考试是平时的练习题,那么学生可能通过记忆去解题,而不是通过学习的方法去做题,所以期末考试的试题应与平时作业不能一样,才能检验学生的学习成果。那么同理,AI模型的效果测试不能使用训练数据进行测试,应使用训练数据集外的数据测试,这样才能 真实的反映模型效果

期末考试的内容属于学期的内容,但不一定需要完全包括所学内容。同理,测试集的标签是训练集的全集或者子集即可

Step 4 训练模型

点击「开始训练」,即可开始训练模型。

- 训练时间与数据量大小、选择的算法、训练环境有关
- 模型训练过程中,可以设置训练完成的短信提醒并离开页面
- 平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	/	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

优惠政策:

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有声音分类操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。可参考价格说明

模型部署

心 整体介绍

训练完成后,可将模型部署在公有云服务器、私有化服务器上,通过API进行调用。 公有云API

- 模型训练完毕后,为了更方便企业用户一站式完成AI模型应用,文本分类模型支持将模型发布成为在线的restful API接口,可以参考示例文 档通过HTTP请求的方式进行调用,快速集成在业务中进行使用。
- 具有完善的鉴权、流控等安全机制,GPU集群稳定承载高并发请求

相关费用 将模型发布为公有云API后,将获得一部分免费调用次数,超出免费额度将根据调用次数进行收费。详见EasyDL价格文档。

私有服务器部署

支持将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便 捷。适用于对数据敏感度、隐私性要求较高、在线离线均有调用需求的企业场景。相关费用 EasyDL已支持将定制模型部署在本地服务器上,只 需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。如需购买永久使用授权,请提交工单咨询。

心 公有云API

^の发布API

发布模型,生成在线API

训练完毕后可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可发布公有云API:

- 1. 选择模型
- 2. 选择部署方式「公有云部署」
- 3. 选择版本
- 4. 自定义服务名称、接口地址后缀
- 5. 申请发布

发布模型界面示意:

文本分类模型	=	发布模型	
器 总览			
植型中心		选择模型	酒店评分1 ~
我的模型		部署方式	公有云部署 >>
创建模型			
训练模型		选择版本	V3 ~
校验模型		服务名称 *	识别酒店评论好与坏
发布模型		接□地北 ★	https://aip.baidubce.com/rpc/2.0/ai_custo
🖉 EasyData数据周	服务		m/v1/text_cls/ jacetest3
数据总览		甘州西北	
公开数据集		共祀安水	右接口无法满足您的需求,请描述希望解决 的问题,500汉字以内
标签组管理			
在线标注			0/500
智能标注			提交申请
目 公有云服务			
在线服务			
盟 EasyEdge本地	部署		
纯离线服务			

发布完成 申请发布后,通常的审核周期为T+1,即当天申请第二天可以审核完成。

如果需要加急、或者遇到莫名被拒的情况,请在百度云控制台内提交工单反馈。

申请发布通过后,界面和状态示意:

【文本分类】	】新闻 模型	젵ID: 19900					全部版本	删除
应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作		
云服务	V1	训练完成	未申请	未发布	top1准确率100.00% top5准确率100.00% 完整评估结果	申请发布校验	训练	

つ。调用API

接口描述

基于自定义训练出的文本分类模型,实现个性化文本识别。模型训练完毕后发布可获得定制化文本分类API 详情访问:EasyDL 文本分类-多标签 进行训练。

更多训练模型过程中的常见问题请查看常见问题文档。

如有其它问题,请在百度云控制台内提交工单反馈。

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在定制化训练平台进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

-



Header如下:

参数	值
Content-Type	application/jsor

注意:如果出现336001的错误码很可能是因为请求方式错误,定制化文本分类服务以json方式请求。

```
Body请求示例:
```

```
{
"text": "<UTF-8编码文本>",
"threshold": 0.6
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text	是	string	-	文本数据,UTF-8编码。最大长度4096汉字
threshold	否	number	-	对识别的文本标签进行阈值条件的筛选,默认阈值为0.5

请求示例代码

Python3
""" EasyDL 文本分类多标签 调用模型公有云API Python3实现 """
import json import base64 import requests
使用 requests 库发送请求 使用 pip(或者 pip3)检查我的 python3 环境是否安装了该库,执行命令 pip freeze grep requests 若返回值为空,则安装该库 pip install requests
目标文本的 本地文件路径,UTF-8编码,最大长度4096汉字 TEXT_FILEPATH = "【您的测试文本数据地址,例如:./example.txt】"

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id ,用于问题定位
results	否	array(object)	分类结果数组
+name	否	string	分类名称
+score	否	number	置信度

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求 ,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或工单联 系技术支持团队,在百度云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问 请在百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大4096 UTF-8字符。
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云 控制台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

つ调用API

接口描述

基于自定义训练出的文本分类模型,实现个性化文本识别。模型训练完毕后发布可获得定制化文本分类API 详情访问:EasyDL 文本分类-多标签 进行训练。

更多训练模型过程中的常见问题请查看常见问题文档。

如有其它问题,请在百度云控制台内提交工单反馈。

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在定制化训练平台进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,定制化文本分类服务以json方式请求。

Body请求示例:

```
{
"text": "<UTF-8编码文本>",
"threshold": 0.6
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text	是	string	-	文本数据,UTF-8编码。最大长度4096汉字
threshold	否	number	-	对识别的文本标签进行阈值条件的筛选,默认阈值为0.5

请求示例代码

Python3

•

```
"""
EasyDL 文本分类多标签 调用模型公有云API Python3实现
"""
import json
import base64
import requests
"""
使用 requests 库发送请求
使用 pip (或者 pip3) 检查我的 python3 环境是否安装了该库,执行命令
pip freeze | grep requests
若返回值为空,则安装该库
pip install requests
"""
**目标文本的 本地文件路径,UTF-8编码,最大长度4096汉字**
```

```
TEXT_FILEPATH = "【您的测试文本数据地址,例如:./example.txt】"
```

```
返回说明
```

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id ,用于问题定位
results	否	array(object)	分类结果数组
+name	否	string	分类名称
+score	否	number	置信度

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
   "error_code": 110,
   "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或工单联 系技术支持团队,在百度云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问 请在百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大4096 UTF-8字符。
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云 控制台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

の 离线API

心 私有部署服务-文本分类多标签

训练完毕后可以在左侧导航栏中找到【发布模型】,依次进行以下操作即可将模型部署到私有服务器:

模型中心	发布模型			
我的模型	选择模型:	老市民	~	说明
创建模型	部署方式:	私有服务器部署	~	私有服务器部署支持将模型部署于本地的CPU、GPU服务器上,提供API和SDK两 种集成方式: 宣看文档
训练模型	集成方式:	● 私有API ─ 服务器站	#SDK	私有API: 將模型以Docker形式在本地服务器(仅支持Linux)上部署为http服
校验模型	选择版本:	V1	\checkmark	务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷
发布模型				服务器端5DK: 将模型封装成适配本地服务器(支持Linux和Windows)的5DK, 可集成在其他程序中运行。首次联网激活后即可纯离线运行,占用服务器资源更
数据中心		提交申请		少,使用方法更灵活
我的数据集				
云服务调用数据				

- 选择模型
- 选择部署方式「私有服务器部署」
- 选择集成方式「私有API」

将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷

点击「提交申请」后,前往控制台申请私有部署包。并参考文档完成集成

✓返回EasyDL总览	产品服务 / EasyOL定则训练平台 - 私有服务器部署管理								
经典版	私有服务器部署管理								
公有云服务管理 >	私有API集成	服务器端SDK集成							
私有服务器部署管理									
设备端服务管理	使用说明:(展开查看部署介绍及操作步骤等)								
数据管理服务	+ 申请私有部署包					请输入私有服务名称	Q,	开发文档	
	申请ID	私有服务名称	模型ID	模型版本	申请日期	证书到期日期	状态	操作	
	1008	test1	17282	V5	2019-09-12	2019-10-05	● 审核通过	查看申请信 下载更新包 更新版本	
	1000	new	35261	V1	2019-09-09	2019-10-09	● 流程关闭, 重新申请	查看申请信	
	989	test1	17282	V5	2019-09-05	2019-10-05	● 审核中	查看申请信	
	988	test1	17282	V5	2019-09-05	2019-10-05	● 审核通过	查看申请信 下载完整部 更新版本	
	987	test	6793	V2	2019-09-05	2019-10-05	● 审核中	查看申请信	
							每页显示	10 🗸 <	

私有服务器部署价格说明

EasyDL经典版已支持将定制模型部署在私有服务器上,只需在发布模型时提交私有服务器部署申请,通过审核后即可获得**一个月免费试用**。如需 购买**永久使用授权**,请微信搜索"BaiduEasyDL"添加小助手咨询,通过线下签订合同购买使用。

心私有部署服务API调用说明-文本分类多标签

本文档主要说明定制化模型本地部署后,如何使用本地API。如还未训练模型,请先前往EasyDL自然语言处理方向进行训练。

如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:868826008)联系群管

部署包使用说明

部署方法

EasyDL定制化文本分类模型的本地部署通过EasyPack实现,目前提供单机一键部署的方式。

在EasyDL申请、下载部署包后,在本地服务器新建目录(建议目录命名规则:easyDL_服务名称_模型版本号),将软件包上传至该目录。请参 考EasyPack-单机一键部署使用python2版本来部署,部署成功后,启动服务,即可调用与在线API功能类似的接口。

运维检查

EasyDL服务器API部署应用健康检查(或故障排查)脚本:trouble_shooting.tar

脚本能力:鉴权服务健康检测、容器状态检查、端口探活、网络联通性测试、容器关键报错日志输出等

使用方法: 将脚本上传至服务器任意目录(或在服务器直接下载),并解压后运行。

```
**解压**
tar vxf trouble_shooting.tar
**执行**
bash trouble_shooting.sh
```

授权说明

本地部署包根据服务器硬件(CPU单机或GPU单卡)进行授权,只能在申请时提交的硬件指纹所属的硬件上使用。

部署包测试期为1个月,如需购买永久授权,可提交工单咨询

API参考

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在EasyDL自然语言处理方向进行自定义模型训练,完成训练后申请本地部署,本地部署成功后拼接url。

请求URL: http://{IP}:{PORT}/{DEPLOY_NAME}/TextClassification

- IP:服务本地部署所在机器的ip地址
- PORT:服务部署后获取的端口
- DEPLOY_NAME:申请时填写的本地服务名称

Header如下:

```
参数 值
Content-Type application/json
```

Body请求示例:

```
{
"text": "<UTF-8编码数据>",
"top_num": 5
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text	是	string	-	文本数据,UTF-8编码。最大长度4096汉字
top_num	否	number	-	返回分类数量,默认为6个

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	分类结果数组
+name	否	string	分类名称
+score	否	number	置信度

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如缺少必要出入参时返回:

{ "error_code": 336001,

"error_msg": "Invalid Argument"

}

错误码	错误信息	描述
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(868826008)或工单联系技术支持团队
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或工单联系技 术支持团队
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数代码格式是否有误。有疑问请通过QQ群(868826008)或工单联 系技术支持团队
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问请在 百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大4096 UTF-8字符。有疑问请通过 QQ群(868826008)或工单联系技术支持团队
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云控制 台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请通过QQ群(868826008) 或工单联系技术支持团队
337000	Auth check failed	离线鉴权调用失败

模型更新/回滚操作说明 模型更新 1、在EasyDL-纯离线服务发布页面,找到您的服务器API发布记录,点击【更新版本】,选择「更新包」或 「完整包」来发布。 两者区别:

包类型	描述
更新包	仅包含最新的模型应用,需执行download.sh脚本下载所需镜像等依赖文件
完整包	包含模型应用和其他鉴权服务,需执行download.sh脚本下载所需完整依赖文件

2、(CPU模型可忽略)如果您训练的模型为GPU版本,系统会生成多份下载链接。请在GPU服务器执行 nvidia-smi命令,根据返回的Cuda Version来选择对应的部署包链接下载。

3、在服务器新建目录(建议标记对应模型的版本号,便于区分不同模型版本),如easydl_\${DEPLOY_NAME}_v2

\${DEPLOY_NAME}:申请时填写的服务名称

以如下场景举例说明:模型版本从V1升级至V2

```
**1.新模型准备**
**创建指定版本的目录**
mkdir easedl_${DEPLOY_NAME}_v2
cd easedI_${DEPLOY_NAME}_v2
**将部署包上传至服务器该目录并解压**
tar zvxf xx.tar.gz
**解压后,进入指定目录执行download脚本下载模型所依赖文件**
cd original && bash download.sh
**2.旧模型备份**
**历史模型备份**
cp -r /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V1
**记录当前模型的端口号**
docker ps -a |grep ${DEPLOY_NAME}
**3.模型升级**
cd package/Install
**卸载当前已安装的旧的easyDL服务: ${DEPLOY_NAME},前面已备份**
python2 install.py remove ${DEPLOY_NAME}
**安装当前部署包内新的EasyDL服务: ${DEPLOY_NAME}**
python2 install.py install ${DEPLOY_NAME}
**(可选操作) 更新证书**
python2 install.py lu
```

模型回滚 以如下场景举例说明:模型版本从V2回滚至V1

方法一:

```
**重命名当前v2模型目录名称**
mv /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V2
**使用V1版本**
cp -r /home/baidu/work/${DEPLOY_NAME}_V1 /home/baidu/work/${DEPLOY_NAME}
**停止当前模型容器**
docker ps -a |grep ${DEPLOY_NAME}
docker ms -f ${容器名}
**创建新的容器**
cd /home/baidu/work/${DEPLOY_NAME} && bash start/start-1.sh
** (可选操作)进入V1版本部署包所在目录执行license更新操作,假如部署包在/opt目录下,以您实际目录为准**
cd /opt/easydl_${DEPLOY_NAME}_V1/original/package/Install && python2 install.py lu
```

方法二:

进入模型V1所在目录,参考上述【模型更新】步骤,执行模型升级操作(即先卸载v2,后升级为v1)

情感倾向分析

整体介绍

心 任务简介

定制情感倾向分析模型,可实现文本按情感的正向(positive)和负向(negative)做自动分类。您只需提供正向和负向标签对应的训练数据,即 可训练获得情感倾向分析模型。

更多详情访问: EasyDL自然语言处理方向

⊙应用场景

- 1. 电商评论分类:可对商品的评论信息进行分类,将不同用户对同一商品的评论内容按情感极性予以分类展示
- 2. 商品舆情监控:通过对产品多维度评论观点进行倾向性分析,给用户提供该产品全方位的评价,方便用户进行决策
- 3. 舆情分类:通过对需要舆情监控的实时文字数据流进行情感倾向性分析,把握用户对热点信息的情感倾向性变化
- 4. 其他:尽情脑洞大开,训练你希望实现的情感倾向分析的模型

の技术特色

情感倾向分析模型内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等 方面的新知识,实现模型效果不断进化。

文心大模型是百度发布的产业级知识增强大模型,是千行百业AI开发的首选基座大模型。文心大模型既包含基础通用的大模型,也包含了面向重 点领域和重点任务的大模型,还提供丰富的工具与平台,支撑企业与开发者进行高效便捷的应用开发。"知识增强"是文心的核心特色,文心能够 同时从大规模知识和海量多元数据中持续学习,如同站在巨人的肩膀上,训练效率和理解准确率都得到大幅提升,并具备了更好的可解释性。

心 使用流程

训练模型的基本流程如下图所示,全程可视化简易操作,在数据已经准备好的情况下,最快15分钟即可获得定制模型。



数据准备

の 创建数据集并导入

创建数据集

在训练模型之前,需要先在数据总览【创建数据集】。只需输入数据集名称、选择数据去重策略,即可创建一个空数据集。

<i>ー・シン</i> 迷楽EasyDL	产品介绍	吕 操作平台	应用案例	使用文档			
情感倾向分析	Æ	我的数据总览					
品 总览							
由 模型中心		EasyData智能	数据服务平台已」	上线,使用Eas	syData可享受包括多人	标注、数据清洗、数据采集等完整	整数据服务 立即前往
我的模型		创建数据集					
创建模型			,				
训练模型		pf情感倾向分析	测试 🖸 数据	集组ID: 3116:	38		
校验模型		版本数	据集ID	数据量	最近导入状态	标注类型	标注状态
发布模型							
🖉 EasyData数据服务		V1 ⊙ 33	36503	0	● 已完成	情感倾向分析	0% (0/0)
数据总览		我我2121 17	数据年纪ID: 20	99966			
在线标注		34342121	xAJICITESEID' 20				

数据自动去重即平台对您上传的数据进行重复样本的去重。建议创建数据集时选择「数据自动去重」

导入数据创建数据集后,在「数据总览」页面中,找到该数据集,点击右侧操作列下的「导入」,即可进入导入数据页面。

1	ど桨EasyDL	产品介	绍	操作平台	应	用案例	使用文材	当		
情愿	感倾向分析	Ē	Ŧ	我的数据总览 >	1111111	I/V1/导入				
	总览									
Ē	模型中心		I	创建信息 🗸						
	我的模型			数据集ID	34683	9		版本号	V1	
	创建模型			备注	Ľ					
	训练模型		I	标注信息 🗸						
	校验模型			标注类型	情感倾	向分析		标注模板	情感倾向	同分析
	发布模型			数据总量	0			已标注	0	
<u></u>	EasyData数据服务		;	标签个数	2			待确认	0	
	数据总览			大小	0M					
	在线标注		I	数据清洗						
	公有云服务		:	暂未做过数据清	洗任务					
	在线服务		I.	导入数据						
멉	EasyEdge本地部署			数据标注状态	0	○ 无标注信息	3	○ 有标注信	息	
	纯离线服务			导入方式		请选择			^	
~	AI市场					本地导入				
	我的已购模型					BOS目录导)	λ.			
	售卖模型					分享链接导 <i>)</i> 平台已有数排	、 居集			

您可以使用4种方案上传情感倾向分析的数据,分别为:

- 本地导入
- BOS目录导入
- 分享链接导入
- 平台已有数据集

不论您上传无标注信息的数据或有标注信息的数据,都需要以下述格式要求进行上传。区别在于上传有标注信息的数据时,需要针对正向文本内 容与负向文本内容分别进行上传。每个数据集里面默认包含正向(positive)标签和负向(negative)标签。

本地导入您可以通过以下三种方式进行本地数据的导入:

- 以压缩包的方式上传
- 以TXT文本文件方式上传
- 以Excel文件的方式上传 以压缩包方式上传
- 一个文本文件保存一个样本,文本文件的编码方式:UTF-8,每个文本文件最长不能超过512个汉字(字符)
- 压缩包仅支持zip格式;大小需要在5GB以内

以TXT文本文件上传

- 一个文本文件包含多个样本,文本文件中每行为一个样本
- 一个文本文件包含一个样本,单次上传限制100个文件,最多可上传100万个文本文件
- 每行样本最长不能超过512个汉字(字符),文件编码方式: UTF-8

以Excel文件上传

• Excel文件上传数据格式为每行是一个样本,每个数据文本内容的字符数建议不超过512个,超出将被截断

- 文件类型支持xlsx格式,单次上传限制100个文件
- 需确保上传的样本在sheet1中,且数据都在首列

BOS目录导入 需选择Bucket地址与对应的文件夹地址。

请确保将全部文本已通过txt文件保存至同一层文件目录,该层目录下子文件目录及非相关内容(包括压缩包格式等)不导入。

分享链接导入需输入链接地址。分享链接导入的要求如下:

• 仅支持来自百度BOS、阿里OSS、腾讯COS、华为OBS的共享链接

平台已有数据集

- 导入无标注数据时,选择需要导入的数据集名称,可导入其不带标注的全部数据,或未标注的数据
- 导入已标注数据时,选择需要导入的数据集名称,可导入其某个或全部标签下的数据

准备数据集的技巧 情感倾向分析任务中,可参考以下准备数据集的技巧:设计分类

情感倾向分析的数据集,默认只使用正向和负向两种标签的数据,所以数据集中无需创建标签,您只需准备对应情感倾向的标签数据即可。

数据量

每个标签建议至少需要准备50个以上的样本,如果想要较好的效果,建议准备1000-10000个文本样本,如果某些分类的文本具有相似性,需要 增加更多文本。**数据分布**

- 训练集文本需要和实际场景要识别的文本内容的业务范围一致,且标签对应文本的数量分布一致。如训练集的业务范围是图书商品的情感倾向 分析,而预计线上对应的场景或业务是电子产品的情感倾向分析,此时两者不一致,将会导致模型实际应用效果不佳
- 考虑实际应用场景有多种可能性,每个场景都需要准备相对应的训练数据,训练集若能覆盖的场景越多,模型的泛化能力则越强
- 建议对高频的业务场景尽量做到覆盖,并通过线上bad case来进行训练数据的优化

如果需要寻求第三方数据采集团队协助数据采集,请在百度云控制台内提交工单反馈

心 在线标注

在线标注

Step 1 进入标注页面 上传未标注的数据后,可以通过以下两个方式进入标注页面:

• 在「数据总览」页面,该数据集对应的操作列下,点击「标注」,即可进入标注页面

<i>-لغابخ</i> K桨EasyDL	产品介绍	操作平台	应用案例	使用文	档			社区	飞桨官网	AI主站	() f5i	BK1G9x5hG
情感倾向分析	÷	我的数据总览										提交工单
18 总览												
由 模型中心		EasyData智能	能数据服务平台已。	上线,使用B	EasyData可享受包括多人核	示注、数据清洗、数据采集	等完整数据服务 立即前往					
我的模型		创建数据集									输入数据集合	S称或ID Q
创建模型			-									
训练模型		測试情感傾向	N 区 数据集组ID	275312						[]新增版本	믬 全部版本	回 删除
校验模型		版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作			
发布模型		V2 G	206177	100	• 卫宫成	德威倾向分析	100% (100/100)		-	家人标注	RA HEL	1
☑ EasyData数据服务		¥2 ()	200177	100	• LINDA	14/2/14/19/19/19/17	10070 (1007100)			SPICIOUL	ANY MOLT	J
数据总览		V1 💬 👘	296016	112	 已完成 	情感倾向分析	100% (112/112)	-	查看	多人标注	导入 标注	
在线标注												
目 公有云服务		binbin_test_1	1127 🖸 数据集组	∄ID: 2539	50					[异新増版本	铝 全部版本	直 删除

• 在「在线标注」页面,选择该数据集,即可进入标注页面

チジン・ ど楽 EasyDL 产品が	1-组 操作平台 应用案例 使用文档	社区 飞桨官网	AI圭站 f58K1G9x5hG	1
情感倾向分析	在线标注 > tj不去重文本 > 标注			-
器 总览	全部(9136) 无标注信息(8236) 有标注信息(900)	板注示例	标签栏	
山 模型中心				
我的模型	标签: 请在右侧选择标签	自 删除文本 上一篇 下一篇	请输入标签名称 Q	
创建模型			根据文本内容,选择标签	
训练模型	1、做工精致; 2、开机运行时间差不多4个小时,基本够用了。不过要注意头几次充电量 级的机器卖占了,另外送的申助原确实好便,就是小了一占占, 1	要充分,以激活电池; 3、其他的就是优派自己介		**Step 2
校验模型	HER FORMULATING FF FOR A COMPANY HER COLOUR THE FORMULA FF THE FORMULA		positive 快捷键 1	
发布模型			negative 供捷键 2	
図 EasyData数据服务				
数据总览				
在线标注		Ш		
目 公有云服务				
左线服务				

针对尚未进行标注的数据,通过以下方式进行标注:

- 选择右边标签栏中的标签。若该文本内容为积极情感,则选择positive;若该文本内容为消极情感,则选择negative。系统默认选择上方的标签
- 点击下一篇,此篇文本的内容即可进行自动保存,且您将开始对下一篇文本进行标注

针对已进行标注的数据,通过以下方式进行标注修改:

- 进入需修改标签的文本的标注页面,选择右边标签栏中的标签
- 点击下一篇,对此篇文本标签的修改即可进行自动保存

全部(9136) 无标注信息(8237) 有标注信息(899)	标注示例	标签栏	
标签:请在有限选择特征 ① 图除文:	第一子 第一上 #	请输入标签名称	
性价化不错,交通方便,行政指层感觉很好,只是早上8点楼上装楼,好渺。 中餐厅档次大纸,虽然便宜,但是和酒店档次不相配。 1		根据文本内容,	选择标签
		positive	0.882
		negative	快速電 2

标注技巧:对于批量文本需要标注为同一个标签的情况,可以在右侧标签区域将标签置顶,进而提高标注效率**Step 3 查看标注信息** 通过以下方式查看已标注的文本信息:

- 在「数据总览」页面,该数据集对应的操作列下,点击「查看」,进入查看标注页面后,点击「有标注信息」
- 通过选择左侧标签中的不同标签名称,即可查看不同标签下的文本数据

我的数据总览 〉【文本】 宣看								
全面 (0130) 用級主体意 (009) 天板注信意 (0227) 年初 日本								
▶ 的文本列表 第迭 ∨				回批量删除				
全部标签 (2)	序号	文本內容勝要	揚作					
标签名称 数据量 操作	1	看过该书,感觉中医暂时不会消亡,尚有一、二十株老树活着,还有毛以林、黄熜、刘力红等一批有一定临床经验的博士在,还可以延缓其生命,但是不改变目前中医院校约办学模式,改…	章者	删除				
positive 41 清空	2	·最看的时候觉得很多方面说的不错,但是仔细做下来,就觉得一些东东也是作者东拼西凑过来,扣来就很少翻了, 0	章看	#123				
negative 858 清空	3	同事们都开始读了,并实践在孩子身上,效果很明显,真是本好书! 1	宣看	#103				
	4	不知道什么原因。这个酒店目前的价格做了下调,而且用末最然有说锅价。不错! 这次入住的最大床房,依旧最紧促的布局,舒服的床,总体来说,还是很不错的! 1	宣看	删除				
	5	內存数量配置偏低 内存插槽于掌托下,撕拆卸安装,不方便 蓝牙模块采用软件控制 0	章者	删除				
	6	外观还是比较漂亮大方的,价格也很公道,DDR3的内存,领先不少台式机。 1	章者	删除				
	7	价格还可以,我是6768家下的(用了一张500波20的券),第二天就逢回到6909了。送的正品泰格斯包很漂亮,也很扎实。 1	登看	#128				
	8	这书我看他的丰强时就感觉它是给我一个希望的书,可一看和我的爆反了,没什么帮助的,就是觉的失败 0	童看	#103				
	9	一家很好的度假型混在环境一流,尤其是户外的自然环境很好,房间的设施也不错,难一有点不足的是混在里的银厅价格高了一点已和香港4星混运内的价格基不多了.温泉一般,比南京汤山的额	童有	删除				
< 1 / 1 >	10	当年在《参考讯息》上曾经拜读过步廷敏的"文明冲突"文章摘要,正式出版的这本书第一次读别,还没看完。他可真是当代伟大的政治学家,一是系统科学地总结了世界文明的发展历史和…	童箱	删除				

当前数据集标注模板:情愿倾向分析,该标签有标注信息文本:41个

の 数据去重

重复样本的定义

一个样本包括文本内容和标签。重复样本的定义,是指您上传的数据中,存在两个样本的文本内容完全一致。则被判定为两个样本是重复样本。 例如:

文本内容	标签
这个酒店的地段不错,交通方便	1
这个酒店的地段不错,交通方便	1
这个酒店的地段不错,交通方便	0

上表三个样本均为重复样本,后两个样本虽然标签不一,但文本内容一致,也为重复样本。

Tips:如果您在模型训练过程中,需要通过增加某个类别标签的预测权重,可以通过增加此标签的重复样本来达到此目标

平台去重策略

平台提供了可去重的数据集,即对您上传的数据进行重复样本的去重。

注意:当您确定了数据集为去重或非去重的属性后,便不可修改

当您创建了一个去重的数据集时,在后续上传数据的过程中,平台可通过检验您当前上传的样本与已上传到此数据集下的样本是否相同,如果相同,则会使用新的样本替代旧的样本。此时分为几种情况,如下:

- 数据集中有未标注样本,上传重复的已标注样本,此时未标注样本将被覆盖
- 数据集中有已标注样本,上传重复的未标注样本,此时已标注样本将被覆盖
- 数据集中有已标注样本,上传不同标注的已标注样本,此时已有的标注样本将被覆盖

^の API上传

本文档主要说明当您线下已有大量的已经完成分类整理的文本数据,如何通过调用API完成文本数据的便捷上传和管理。

EasyDL经典版数据集管理API在管理不同模型数据类型之间是通用的。上传不同模型类型数据,只是在部分接口入参存在差异,使用及接口地址 完全一致。

数据集创建API

接口描述

该接口可用于创建数据集。

接口鉴权

同发布模型后获取的API鉴权方式:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/create

URL参数:



access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

content-type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION,TEXT_CLASSIFICATION_MUL 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类(单标签) 、文本分类(多标签)
dataset_ name	是	string	数据集名称,长度不超过20个utf-8字符

若上传声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
dataset_id	否	number	创建的数据集ID

查看数据集列表API

接口描述

该接口可用于查看数据集列表。返回数据集的名称、类型、状态等信息。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/list

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION, TEXT_CLASSIFICATION_MUL, SENTI_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类(单标签)、文本分类(多标签)、情感倾向分析
start	否	number	起始序号,默认为0
num	否	number	数量,默认20,最多100

若查看声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	数据集总数
results	否	array(object)	数据集列表
+dataset_id	否	number	数据集ID
+dataset_name	否	string	数据集名称
+type	否	string	数据集类型
+status	否	string	数据集状态
+special_status	否	string	数据集特殊状态,包括shared、smart和空值,分别表示共享中、智能标注中、非特殊状态

查看分类 (标签)列表API

接口描述

该接口可用于查看分类(标签)。返回分类(标签)的名称、包含数据量等信息。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/list

URL参数:

参数	值			
----	---	--	--	--

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION, TEXT_CLASSIFICATION_MUL, SENTI_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类(单标签)、文本分类(多标签)、情感倾向分析
dataset _id	是	number	数据集ID
start	否	number	起始序号,默认0
num	否	number	数量,默认20,最多100

若查看声音分类的全部分类,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	标签总数
results	否	array(object)	标签列表
+label_id	否	string	标签/分类ID
+label_name	否	string	标签/分类名称
+entity_count	否	number	图片/声音/文本数量

添加数据API

接口描述

该接口可用于在指定数据集添加数据。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/addentity

URL参数:

参数 值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值

Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION, TEXT_CLASSIFICATION_MUL, SENTI_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类(单标签)、文本分类(多标签)、情感倾向分析
dataset _id	是	number	数据集ID
entity_c ontent	是	string	type为 IMAGE_CLASSIFICATION/OBJECT_DETECTION/IMAGE_SEGMENTATION/SOUND_CLASSIFICATION时,填入图 片/声音的base64编码;type为TEXT_CLASSIFICATION或 TEXT_CLASSIFICATION_MUL时,填入utf-8编码的文本。内容 限制为:图像分类base64前10M;物体检测base64前10M;图像分割base64前10M;声音分类base64前4M,声音 时长1~15秒;文本分类单标签和多标签为4096个汉字;情感倾向分析为512个汉字
entity_n ame	是	string	文件名
labels	是	array(o bject)	标签/分类数据
+label_ name	是	string	标签/分类名称(由数字、字母、中划线、下划线组成),长度限制20B
+left	否	number	物体检测时需给出,标注框左上角到图片左边界的距离(像素)
+top	否	number	物体检测时需给出,标注框左上角到图片上边界的距离(像素)
+width	否	number	物体检测时需给出,标注框的宽度(像素)
+height	否	number	物体检测时需给出,标注框的高度(像素)

若上传声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

数据集删除API

接口描述

该接口可用于删除数据集。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/delete

URL参数:

参数 值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
a .	

Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION, TEXT_CLASSIFICATION_MUL 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类(单标签)、文本分类(多标签)
dataset _id	是	number	数据集ID

若刪除声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

分类 (标签) 删除API

情感倾向分析,不可删除分类标签。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量 阶梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶 梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
110 111	Access token invalid or no longer valid Access token expired	access_token无效 access token过期
110 111 406000	Access token invalid or no longer valid Access token expired internal server error	access_token无效 access token过期 服务错误
110 111 406000 406001	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数
110 111 406000 406001 406002	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid dataset not exist	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数 数据集不存在
110 111 406000 406001 406003	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid dataset not exist dataset already exists	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数 数据集不存在 数据集已存在
110 111 406000 406001 406002 406003	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid dataset not exist dataset already exists dataset can not be modified temporarily	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数 数据集不存在 数据集已存在 数据集暂不可修改
110 111 406000 406001 406003 406004 406005	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid dataset not exist dataset already exists dataset can not be modified temporarily label not exist	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数 数据集不存在 数据集已存在 数据集暂不可修改
110 111 406000 406001 406003 406004 406005	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid dataset not exist dataset already exists dataset can not be modified temporarily label not exist no permission to modify the dataset	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数 数据集不存在 数据集已存在 数据集暂不可修改 标签/分类不存在 没有修改数据集的权限
110 111 406000 406001 406003 406004 406005 406006	Access token invalid or no longer valid Access token expired internal server error param[xx] invalid dataset not exist dataset already exists dataset can not be modified temporarily label not exist no permission to modify the dataset dataset cannot be modified while smart annotation is running	access_token无效 access token过期 服务错误 参数xx不合法,请检查相关参数 参数xx不合法,请检查相关参数 数据集不存在 数据集已存在 数据集暂不可修改 标签/分类不存在 没有修改数据集的权限 智能标注期间不可修改数据集

模型训练

の 创建模型

创建模型

在模型中心目录中选择「创建模型」,填写模型名称、模型归属、所属行业、应用场景、邮箱地址、联系方式、功能描述等信息,即可创建模型。

目前单个用户在每种类型的模型下最多可创建**10个**模型,每个模型均支持多次训练,若需要创建超过10个以上的模型,请在百度云控制台内提 交工单反馈。

情	感倾向分析		模型列表 > 6	刘建模型	
	总览				
ſŦ	模型中心		模型类别	情感倾向分析	
Г	我的模型		模型名称 *		
L	创建模型		模型归属	公司 个人	
Т	训练模型				
Т	校验模型			请输入公司名称	
L	发布模型		所属行业 *	请选择行业 ~	
11	EasyData数据服务	5	应用场景 *	请选择应用场景 >	
	数据总览				
	在线标注		邮箱地址 *	z*******@baidu.com	
	公有云服务		联系方式 *	135****919	0
	在线服务		功能描述 *		
	EasyEdge本地部署	2			
	纯离线服务			0/500	
~	AI市场			0/300	4
	我的已购模型			完成	
	售卖模型				

模型创建成功后,即可在「我的模型」中看到刚刚创建的模型。

注:

1. 创建模型后可持续新增模型版本,因此不必每次训练模型都创建模型。

2.目前单个用户在每种类型的模型下最多可创建10个模型,每个模型均支持多次训练。

3.如果您是企业用户,建议您按照真实企业信息进行填写,便于EasyDL团队后续更好的为您服务。

心 评估效果

模型评估模型训练完成后,可以在「我的模型」列表中查看该模型的效果,以及完整评估结果。

情懸傾向分析	我的局型 > EmotionalTest-VI購取评估股份
88 总览 益 模型中心 我的模型	部策方式 公衛吾API 〜 総本 VI 〜 文本鉄 9137 分売数 2 領部所长 9小時1565時
台建模型 山防模型 校紀模型 发布模型 I EaryCotat数或语例	Тякняй Environment feet Visitageta, safet (KR333) 3 ### 0 90,0% 47.4%
数据总克 在线标注	详细评估
 公有云银务 在线服务 EasyEdge本地部署 纯离线服务 八市场 	· 特任年4月時間構成 · 時間間は第 · 時間間には第 · 日帰時間 · 208
我的已购模型 售卖模型	63次約時間道 65次約月-Score pol/line 86.0% pol/line 84.7% 63分素が目前通 pol/line 84.7%

「完整评估结果」页面中将记录整体评估与详细评估的报告,包括该模型整体的准确率、F1-score、精确率、召回率,以及评估样本具体数据情况,各分类的精确值、F1-Score、召回值等指标。

整体评估中,各指标的释义如下:

- 准确率:正确分类的样本数与总样本数之比
- F1-score:给每个类别相同的权重,计算每个类别的F1-score,然后求平均值
- 精确率:给每个类别相同的权重,计算每个类别的精确率,然后求平均值
- 召回率:给每个类别相同的权重,计算每个类别的召回率,然后求平均值

如果单个标签的文本量在100条以内,会影响评估指标的科学有效性,请确保提交的训练数据中每个标签的数据量

模型校验

在完成训练后,发布模型前,可以先进行模型校验,以确保模型在实际环境中能获得预期的性能。操作方法如下:

- 1. 在左侧「模型中心」目录中点击「校验模型」,进入校验模型页面
- 2. 选择需要校验能力的模型、部署方式、版本,点击「启动模型校验服务」
- 3. 校验服务启动后,在左侧输入文本,点击「校验」后,右侧识别结果栏将输出预测结果
 - 校验数据支持两种输入方式,直接输入文本或上传txt格式文本,文本长度上限为512汉字
 - 在识别结果栏中,可进行阈值的调整。置信度在阈值以下的预测结果将不予显示。例如,阈值调整为0.85时,置信度小于85%的预测结果将不予显示。当positive与negative两类标签的置信度均小于85%时,识别结果栏将显示为:"没有满足条件的识别结果"
- 4. 若您对预测结果满意,可点击「申请上线」,进行模型的发布

心 效果优化

通过模型迭代、检查并优化训练数据、选择高精度模型等方法,能够提升模型效果。****模型迭代****

一个模型很难一次性就训练到最佳的效果,通常会需要结合模型评估报告和校验结果不断扩充数据和调优。

为此平台提供了模型迭代功能,即当模型训练完毕后,会生成一个最新的版本号,首次V1、之后V2……以此类推。可以通过调整训练数据和算法,多次训练,以获得适合业务需求的模型效果。

如果模型已经是上线状态,依然支持模型迭代,只是需要在训练完毕后更新线上服务接口,在接口地址不变的情况下可以持续优化效果。

检查并优化训练数据

- 检查是否存在训练数据过少的情况,建议每个类别的文本数量不少于1000个,如果低于这个量级建议扩充
- 检查不同类别的数据量是否均衡,建议不同分类的数据量级相同,并尽量接近,如果有的类别数据量很高,有的类别数据量较低,会影响模型 整体的识别效果
- 通过模型效果评估报告中的各分类的详细评估指标,有针对性地扩充训练数据,比如下图中的评估报告显示,负向分类精确值数据表现较差, 即可考虑从两个方向进行针对性优化,一是增加负向数据集样本数据量,二是检查现有负向数据集中是否存在定义模糊的情况,提升负向数据 集质量,优化模型效果。

我的模型 〉 sur	n_无自定义V2模型评估报告	Ī			
文本数 1185	分类数 2	训练时长	1小时19分钟		
整体评估					
sun_无自定义 Vi	2效果较好。 如何优化效果	2			
う ^{准确率} 61.2	@ %		F1-score @ 38.4%		〇 〇 〇 〇 〇 〇 〇 〇 〇 〇 〇 〇 〇 〇
详细评估					
评估样本具体	数据情况				
	随机测试集				
775-06-08-153	正确数量: 377				
1303940042	错误数量: 239				
各分类的精确	值		各分类的F1-Score		
negative	_	50.0%	negative	0.8%	
positive		61.2%	positive	75.9%	
各分类的召回	值				
negative	1	0.4%			
positive		99.7%			

 检查测试模型的数据与训练数据的文本类型与风格是否一致,如果不一致,那么很可能会存在模型效果不错但实际测试效果较差的情况。针对 这种情况建议重新调整训练集,将训练数据与实际业务场景数据尽可能一致

选择高精度模型

在训练模型时,选择高精度的模型,将提升模型的预测准确率。

「高精度」算法内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等方 面的新知识,实现模型效果不断进化

の发起训练

训练模型

完成数据的标注,或提交已标注的数据后,即可在「模型中心」目录中点击「训练模型」,开始模型的训练。

按以下步骤操作,启动模型训练:

情感倾向分析 《三	训练模型							
88 BZ								
直 模型中心	选择模型	选择模型 sun_无指定义 ~						
我的模型	训练配置	(約62) 面						
创建模型	部署方式	13式 公有2回答 EasyEdga本地解释 如何指导部署为15?						
训练模型	选择算法	及岸前注 ○ 高橋泉 ○ ○ 周恒能 ○ ○ A市场已的模型						
校验模型	#图带造指标 ② 標整兼Wreadsban和Rea · · ③							
发布模型	添加数据							
☑ EasyData数据服务								
数据总范	107.0093.0034	+ 1812539						
在线标注		数据集	版本	分类数量		操作		
自 公有云服务		tj-不去重文本	V3	2		80		
在线服务		vd 10.0 tl V1 2 Rtd						
器 EasyEdge本地部署		Afterior and a transformed a t						
纯高线服务		多人称注-188	¥1	2		1918		
□ AI市场	自定义测试集 🕲							
我的已购模型	训练环境	名称	规格	算力	速度比例	价格		
售卖模型		 GPU P40 	TeslaGPU_P40_24G盟存单卡_12核CPU_40G内存	12 TeraFLOPS	1	免费		
		GPU V100	TeslaGPU_V100_163显存单卡_12核CPU_56G内存	14 TeraFLOPS	1.5	单卡¥0.45/分钟 (348.95小时*节点小时包余额)		
	FF \$12 10 10 15							

Step 1 选择模型 选择此次训练的模型 Step 2 训练配置

部署方式

可选择「公有云部署」、「EasyEdge本地部署」。

如何选择部署方式

选择设备

- 如果您选择了「EasyEdge本地部署」,请根据实际部署设备选择
- 如果您选择了「公有云部署」,无需选择设备

您可以根据训练的需求,选择「高精度」或「高性能」算法。不同的算法将影响训练时间、预测速度与模型准确率。

- 如果您选择了高精度的模型,模型预测准确率将更高。使用10000条训练样本,将在60-100min完成训练
- 如果您选择了高性能的模型,相同训练数据量的情况下,训练耗时更短,模型预测速度更快。使用10000条训练样本,将在5min内完成训练

如果您已从AI市场购买了模型算法,也可以基于已购模型的算法训练:前往AI市场购买

「高精度」算法内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义 等方面的新知识,实现模型效果不断进化

模型筛选指标

选择不同的模型选择方式,对应的模型各项效果指标将有所不同。如果没有特殊场景的要求,使用默认即可。有以下指标可供选择:

- 模型兼顾Precision和Recall:挑选模型时,兼顾Precision精确度和Recall召回率,如场景中没有对精度或召回的特别要求,建议您使用此默认指标
- Precision最高的模型:挑选模型时,优先挑选Precision精度最高的模型作为部署模型
- Recall最多的模型:挑选模型时,优先选择召回率最高的模型作为部署模型
- ACC最大的模型:挑选模型时,优先挑选预测样本数最多的模型作为部署模型
- Loss最小的模型:挑选模型时,优先挑选预测偏差最小的模型

Step 3 添加数据

添加训练数据

- 先选择数据集,再按标签 (positive、negative) 选择数据集里的文本,可从多个数据集选择文本
- 训练时间与数据量大小、选择的算法、训练环境有关。在选择GPU P40、高性能的模型时,10000条训练样本,将在5min内完成训练

添加自定义测试集

上传不包含在训练集的测试数据,可获得更客观的模型效果评估结果。

添加自定义测试集的目的:

如果学生的期末考试是平时的练习题,那么学生可能通过记忆去解题,而不是通过学习的方法去做题,所以期末考试的试题应与平时作业不能一样,才能检验学生的学习成果。那么同理,AI模型的效果测试不能使用训练数据进行测试,应使用训练数据集外的数据测试,这样才能 真实的反映模型效果

期末考试的内容属于学期的内容,但不一定需要完全包括所学内容。同理,测试集的标签是训练集的全集或者子集即可

Step 4 训练模型

点击「开始训练」,即可开始训练模型。

- 训练时间与数据量大小、选择的算法、训练环境有关
- 模型训练过程中,可以设置训练完成的短信提醒并离开页面
- 平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有声音分类操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。可参考价格说明

模型部署

心 整体介绍

训练完成后,可将模型部署在公有云服务器、私有化服务器上,通过API进行调用。 公有云API

模型训练完毕后,为了更方便企业用户一站式完成AI模型应用,情感倾向分析模型支持将模型发布成为在线的restful API接口,可以参考示例文档通过HTTP请求的方式进行调用,快速集成在业务中进行使用。适用于绝大多数通用情感倾向分析场景,如电商评论分析、商品舆情监控等。

相关费用 将模型发布为公有云API后,将获得一部分免费调用次数,超出免费额度将根据调用次数进行收费。详见EasyDL价格文档。

私有服务器部署

支持将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便 捷。适用于对数据敏感度、隐私性要求较高的企业场景,如医美平台用户评论分析。 **相关费用** EasyDL已支持将定制模型部署在本地服务器上, 只需在发布模型时提交本地服务器部署申请,通过审核后即可获得**一个月**免费试用。如需购买永久使用授权,请提交工单咨询。

心 公有云API

の 发布API

发布模型,生成在线API

训练完毕后可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可发布公有云API:

- 1. 选择模型
- 2. 选择部署方式「公有云部署」
- 3. 选择版本
- 4. 自定义服务名称、接口地址后缀
- 5. 申请发布

发布模型界面示意:
情愿	感倾向分析	 发布模型	
	总览		
Ē	模型中心	选择模型	sun_无自定义 V
	我的模型	部署方式	公有云部署
	创建模型		
	训练模型	选择版本	V1 ~
	校验模型	服务名称 *	
	发布模型	接口地址 *	https://aip.baidubce.com/rpc/2.0/ai_custo
<u></u>	EasyData数据服务		m/v1/sentiment_cls/
	数据总览		
	在线标注	具他要求	若接口无法满足您的需求,请描述希望解决 的问题,500汉字以内
	公有云服务		
	在线服务		0/500
	EasyEdge本地部署		提交申请
	纯离线服务		
	AI市场		
	我的已购模型		
	售卖模型		

发布完成 申请发布后,通常的审核周期为T+1,即当天申请第二天可以审核完成。

如果需要加急、或者遇到莫名被拒的情况,请在百度云控制台内提交工单反馈。

申请发布通过后,界面和状态示意:

【情感倾向分析】gq_sen_调参后高精度 模型ID: 10125							릚 训练	① 历史版本	直 删除
部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作			
公有云API	V1	训练完成	审核成功	已发布	准确率: 92.85% F1-score : 0.93 完整评估结果	查看版本配	置服务	详情 校验	

心调用API

接口描述

基于自定义训练出的情感倾向分析模型,实现个性化情感倾向分析。模型训练完毕后发布可获得定制化情感倾向分析API。详情访问:EasyDL 情感倾向分析进行训练。

如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:868826008)联系群管理员

API参考 请求说明

请求示例

HTTP 方法:POST

请求URL:请首先在EasyDL 情感倾向分析进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,定制化文本分类服务以json方式请求。

Body请求示例:

```
{
"text": "<UTF-8编码文本>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明	
text	是	string	-	文本数据,UTF-8编码。	最大长度4096汉字

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	分类结果数组
+name	否	string	分类名称
+score	否	number	置信度

在线调试 EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回 结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

```
例如Access Token失效返回:
```

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或工单联系 技术支持团队,在百度云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问 请在百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大4096 UTF-8字符。
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云 控制台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

の调用API

接口描述

基于自定义训练出的情感倾向分析模型,实现个性化情感倾向分析。模型训练完毕后发布可获得定制化情感倾向分析API。详情访问:EasyDL 情感倾向分析进行训练。

如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:868826008)联系群管理员

API参考 请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在EasyDL 情感倾向分析进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,定制化文本分类服务以json方式请求。

Body请求示例:

```
{
"text": "<UTF-8编码文本>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明	
text	是	string	-	文本数据,UTF-8编码。	最大长度4096汉字

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	分类结果数组
+name	否	string	分类名称
+score	否	number	置信度

在线调试 EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回 结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或工单联系 技术支持团队,在百度云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问 请在百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大4096 UTF-8字符。
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云 控制台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

の 离线API

の 发布API

在训练模型时,您需要选择「EasyEdge本地部署」的训练方式,才能发布本地部署的私有API。

私有API介绍

将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷

发布私有API的流程训练完毕后,您可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可将模型部署到私有服务器:

1. 在「发布模型」页面中,选择模型及模型的版本,选择部署方式为「EasyEdge本地部署」、集成方式为「API-纯离线服务」。点击「发布」, 即可跳转至「发布新服务」页面

发布模型页面示意:

情愿	感倾向分析	 发布模型		
	总览			
	模型中心	选择模型	edge_test_情感倾向分析	\sim
	我的模型	部署方式	EasyEdge本地部署	\sim
	创建模型			
	训练模型	选择版本	V2	\sim
	校验模型	集成方式	● API-纯离线服务	
	发布模型			
<u>//</u>	EasyData数据服务	发布		
	数据总览	及中		
	在线标注			
	公有云服务			
	在线服务			
립	EasyEdge本地部署			
	纯离线服务			
	AI市场			
	我的已购模型			
	售卖模型			

2. 在「发布新服务」页面,选择部署类型,填写服务名称、证书生效时间等信息,选择对应的系统和芯片。

- 部署类型可支持单模型部署和增量部署
- 增量部署申请,指需要在一台服务器上部署多个模型部署包时使用。进行增量部署时,需在「已部署服务」选择同台服务器历史中最近 部署的部署包,此步骤用来关联不同部署包中的license文件

3. 上传指纹文件。详细操作见指纹提取工具说明,可通过指纹工具进行指纹的提取

4. 点击下一步,填写个人详细信息后即可发布。发布完成后,即可在服务器目录下看到发布处于审核中的状态

个人信息的填写仅供EasyDL团队了解您,便于后续合作接洽,不会作为其他用途使用

5. 等待审核通过,前往「纯离线服务」等待部署包制作完成后,下载部署包,并参考文档完成集成

发布新服务页面示意:

情感倾向分析	纯离线服务 > 发布新服务
品 总览	
<u>前</u> 模型中心	服务器
我的模型	1 洗择部署形式 (2) 填写个人信息
创建模型	
训练模型	集成方式 O API
校验模型	部署类型 💿 单模型部署 💿 增量部署
发布模型	
🖉 EasyData数据服务	业书 测试址书 🗸
数据总览	选择模型 edge_test_情感倾向分析 V
在线标注	选择版本 V3 ~
目 公有云服务	服务名称 *
在线服务	服务名称不能为空
22 EasyEdge本地部署	证书生效时间 * 2022/3/22
纯离线服务	
🖸 AI市场	指纹文件 * 上传文件 ⑦ 指纹提取工具下载 指纹提取工具使用说明
我的已购模型	指状又叶个能为圣
售卖模型	选择系统和心方 Ⅰ ↓ ↓ Linux
	选择系统和芯片不能为空

价格说明 EasyDL已支持将定制模型部署在本地服务器上,只需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。

如需购买永久使用授权,请提交工单咨询。

の调用API

接口描述 本文档主要说明定制化模型本地部署后,如何使用本地API。如还未训练模型,请先前往EasyDL 情感倾向分析进行训练。

如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:868826008)联系群管理员

部署包使用说明 部署方法

EasyDL情感倾向分析模型的本地部署通过EasyPack实现,目前提供单机一键部署的方式。

在EasyDL申请、下载部署包后,在本地服务器新建目录(建议目录命名规则:easyDL_服务名称_模型版本号),将软件包上传至该目录。请参 考EasyPack-单机一键部署使用python2版本来部署,部署成功后,启动服务,即可调用与在线API功能类似的接口。

运维检查

EasyDL服务器API部署应用健康检查(或故障排查)脚本:trouble_shooting.tar

脚本能力:鉴权服务健康检测、容器状态检查、端口探活、网络联通性测试、容器关键报错日志输出等

使用方法:将脚本上传至服务器任意目录(或在服务器直接下载),并解压后运行。

授权说明

本地部署包根据服务器硬件(CPU单机或GPU单卡)进行授权,只能在申请时提交的硬件指纹所属的硬件上使用。

部署包测试期为1个月,如需购买永久授权,可提交工单咨询。

API参考

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在EasyDL自然语言处理方向进行自定义模型训练,完成训练后申请本地部署,本地部署成功后拼接url。

请求URL: http://{IP}:{PORT}/{DEPLOY_NAME}/SentimentClassification

- IP:服务本地部署所在机器的ip地址
- PORT:服务部署后获取的端口
- DEPLOY_NAME:申请时填写的本地服务名称

Header如下:

参数 值 Content-Type application/json

Body请求示例:

```
{
"text": "<UTF-8编码数据>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明		
text	是	string	-	文本数据,UTF-8编码。	最大长度512个汉字,	超出将被截断

返回说明

```
返回参数
```

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	分类结果数组
+name	否	string	分类名称
+score	否	number	置信度

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如缺少必要出入参时返回:

{ "error_code": 336001,

"error_msg": "Invalid Argument" }

错误码	错误信息	描述
336000	Internal error	服务器内部错误,请再次请求 ,如果持续出现此类错误,请通过QQ群 (868826008) 或工单联系技术支持团队
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或工单联系技 术支持团队
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数代码格式是否有误。有疑问请通过QQ群(868826008)或工单联 系技术支持团队
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问请在 百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大4096 UTF-8字符。有疑问请通过 QQ群(868826008)或工单联系技术支持团队
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到清重试,如反复失败,请在百度云控制 台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请通过QQ群(868826008) 或工单联系技术支持团队
337000	Auth check failed	离线鉴权调用失败

模型更新/回滚操作说明 模型更新 1、在EasyDL-纯离线服务发布页面,找到您的服务器API发布记录,点击【更新版本】,选择「更新包」或 「完整包」来发布。 两者区别:

包类型	描述
更新包	仅包含最新的模型应用,需执行download.sh脚本下载所需镜像等依赖文件
完整包	包含模型应用和其他鉴权服务,需执行download.sh脚本下载所需完整依赖文件

2、(CPU模型可忽略)如果您训练的模型为GPU版本,系统会生成多份下载链接。请在GPU服务器执行 nvidia-smi命令,根据返回的Cuda Version来选择对应的部署包链接下载。

3、在服务器新建目录(建议标记对应模型的版本号,便于区分不同模型版本),如easydl_\${DEPLOY_NAME}_v2

\${DEPLOY_NAME}:申请时填写的服务名称

以如下场景举例说明:模型版本从V1升级至V2

```
**1.新模型准备**
**创建指定版本的目录**
mkdir easedI_${DEPLOY_NAME}_v2
cd easedI_${DEPLOY_NAME}_v2
**将部署包上传至服务器该目录并解压**
tar zvxf xx.tar.gz
**解压后,进入指定目录执行download脚本下载模型所依赖文件**
cd original && bash download.sh
**2.旧模型备份**
**历史模型备份**
cp -r /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V1
**记录当前模型的端口号**
docker ps -a grep ${DEPLOY_NAME}
**3.模型升级**
cd package/Install
**卸载当前已安装的旧的easyDL服务: ${DEPLOY_NAME},前面已备份**
python2 install.py remove ${DEPLOY_NAME}
**安装当前部署包内新的EasyDL服务: ${DEPLOY_NAME}**
python2 install.py install ${DEPLOY_NAME}
**(可选操作)更新证书**
python2 install.py lu
```

方法一:

```
**重命名当前v2模型目录名称**
mv /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V2
**使用V1版本**
cp -r /home/baidu/work/${DEPLOY_NAME}_V1 /home/baidu/work/${DEPLOY_NAME}
**停止当前模型容器**
docker ps -a |grep ${DEPLOY_NAME}
docker mn -f ${容器名}
**创建新的容器**
cd /home/baidu/work/${DEPLOY_NAME} && bash start/start-1.sh
** (可选操作)进入V1版本部署包所在目录执行license更新操作,假如部署包在/opt目录下,以您实际目录为准**
cd /opt/easydl_${DEPLOY_NAME}_V1/original/package/Install && python2 install.py lu
```

```
方法二:
```

进入模型V1所在目录,参考上述【模型更新】步骤,执行模型升级操作(即先卸载v2,后升级为v1)

文本实体关系抽取

整体介绍

心简介

定制实体关系抽取模型,是指从文本中抽取出预定义的实体类型及实体间的关系类型,得到包含语义信息的实体关系三元组,每个实体关系三元组的两个实体及其关系构成。

更多详情访问: EasyDL自然语言处理方向

心 应用场景

对内容中的关键实体及实体间的关系类型进行识别和抽取,如:

- 金融研报信息识别
- 法律案件抽取
- 其他:尽情脑洞大开,训练你希望实现的文本实体关系抽取的模型

```
心 技术特色
```

文本实体关系抽取模型内置<u>文心大模型</u>,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语 义等方面的新知识,实现模型效果不断进化,当前模型主要应用于中文环境下的语义分析。

文心大模型是百度发布的产业级知识增强大模型,是千行百业AI开发的首选基座大模型。文心大模型既包含基础通用的大模型,也包含了面向重 点领域和重点任务的大模型,还提供丰富的工具与平台,支撑企业与开发者进行高效便捷的应用开发。"知识增强"是文心的核心特色,文心能够 同时从大规模知识和海量多元数据中持续学习,如同站在巨人的肩膀上,训练效率和理解准确率都得到大幅提升,并具备了更好的可解释性。

心 使用流程

训练模型的基本流程如下图所示,全程可视化简易操作,在数据已经准备好的情况下,最快15分钟即可获得定制模型。



数据准备

₯ 创建数据集并导入

创建数据集

在训练模型之前,需要在【数据总览】里面"创建数据集"。需输入数据集名称、选择相应的标注模版、选择数据去重策略,即可创建一个空数据 集。

ーナン^j ፚ桨EasyDL	产品介绍	操作平台	应用案例	使用文档		
文本实体关系抽取		我的数据总览				
总览						
曲 模型中心		EasyData智能	数据服务平台已上约	线,使用EasyData可享受	包括多人标注、数据清洗、数	据采集等完整数据服务 立即前往
我的模型	ſ	创建数据集	ן			
创建模型	L		J			
训练模型		pf文本实体抽取	1测试任务1 🖸 🚦	数据集组ID: 311654		
校验模型		版本	数据集ID	数据量	最近导入状态	标注类型
发布模型						
☑ EasyData数据服务		V1 💬	336521	0	● 导入失败 ⑦	文本实体关系抽取
数据总览		1		007000		
在线标注		zzy-多人短证[┘	数据集组ID: 207090		

数据自动去重即平台对您上传的数据进行重复样本的去重。建议创建数据集时选择「数据自动去重」

****导入无标注数据**** 创建数据集后,在「数据总览」页面中,找到该数据集,点击右侧操作列下的「导入」,即可进入导入数据页面。

ィ ジジンズ楽EasyDL 产	品介绍 操作平台	应用案例 使用	文档	
文本实体关系抽取	我的数据总览	〉 pf文本实体抽取测试任务	§1/V1/导入	
88 总览				
曲 模型中心	┃ 创建信息 ∨			
我的模型	数据集ID	336521	版本号	V1
创建模型	备注			
训练模型	┃标注信息 〜			
校验模型	标注类型	文本实体关系抽取	标注模板	文本实体关系抽取
发布模型	数据总量	0	已标注	0
🖉 EasyData数据服务	标签个数	0	待确认	0
数据总览	大小	OM		
在线标注	■数据清洗			
目 公有云服务	暂未做过数据》	清洗任务		
在线服务	导入数据			
鋁 EasyEdge本地部署	数据标注状态	● 无标注信息	○ 有标注信	息
纯离线服务	导入方式	请选择		^
		本地导入		
		BOS目录导入		
		分享链接导入		

您可以使用3种方案上传数据,分别为:

- 本地导入
- BOS目录导入
- 分享链接导入

本地导入您可以通过以下三种方式进行本地数据的导入:

7-	アズ楽EasyDL	产品介绍	操作平台	应用案例 使用	用文档	
文Z	本实体关系抽取	Ť	我的数据总览	> pf文本实体抽取测试任	务1/V1/导入	
	总览		创建信良 ∨			
đ	模型中心		数据集ID	336521	版本号	V1
	我 时候 型		备注	ß		
	训练模型		┃标注信息 〜			
	校验模型		标注类型	文本实体关系抽取	标注模板	文本实体关系抽取
	发布模型		数据总量	0	已标注	0
~	EasyData数据服务		标签个数	0	待确认	0
	数据总览		大小	OM		
	在线标注		■数据清洗	=)+ / <i>r f</i> r		
8	公有云服务		智木做过数据?	育洗仕夯		
	在线服务		导入数据		_ <u>+</u> =\	
12	EasyEdge本地部署		<u> </u>	 九标注信息 	○ 有标注信点	<u>z</u>
	纯离线服务		导入方式	本地导入		✓ 请洗择 ∧
				确认并返回		上传Excel文件 上传TXT文本 上传压缩包

- 以压缩包的方式上传
- 以TXT文本文件方式上传
- 以Excel文件的方式上传

通过压缩包上传时,需注意:

- 压缩包内每一个txt文件为一个样本,文本文件编码须为UTF-8,每个样本字符数不得超过512个字符(包括汉字、数字、符号等),超出将被 截断
- 压缩包的格式为zip;压缩包最大不超过5G;详见数据样例

通过TXT文本上传时,需注意:

- 文本实体关系抽取数据txt文件中,每一行为一个样本,文本文件的编码格式须为UTF-8,每个样本字符数不得超过512个字符(包括汉字、数字、符号等),超出将被截断,详见数据样例。
- 文件格式支持txt格式,单次可上传100个文件,最多可上传100万个文件。

通过Excel文件上传时,需注意:

- 如果您上传的文本实体关系抽取数据未Excel文件,那么要求您的Excel文件每行为一个样本,每个样本字符数不得超过512个字符(包括汉字、数字、符号等),超出将被截断。注意,表头作为首行将被系统忽略。
- 文件格式支持xlsx格式,单次可上传100个文件,详见数据样例。

BOS目录导入

需选择Bucket地址与对应的文件夹地址。

请确保将全部文本已通过txt文件保存至同一层文件目录,该层目录下子文件目录及非相关内容(包括压缩包格式等)不导入。

分享链接导入需输入链接地址。分享链接导入的要求如下:

仅支持来自百度BOS、阿里OSS、腾讯COS、华为OBS的共享链接 导入有标注数据 创建数据集后,在「数据总览」页面中,找到该数据集,点击 右侧操作列下的「导入」,即可进入导入数据页面。 已标注数据上传方式:针对已标注的文本实体关系抽取数据集上传目前本平台仅支持Excel文件格式上传。

以Excel文件导入

- 要求上传的Excel文件,首行为表头,表头表示每一列代表的数据类型,依次为"文本内容、实体关系1、实体关系2、.."。其中实体关系内格式为:{实体1位置,实体1类别},{实体2位置,实体2类别},实体关系。每个标注内均以英文逗号间隔,且内容顺序不可变。详见数据样例。
- 第二行起每行为一个样本,每个样本文本内容字符数不得超过512个字符(包括汉字、符号、数字等),超出将报错;
- 目前Excel文件格式支持xlsx格式,单次可上传100个文件;文本样例如下。

文本内容	实体关系1
今年年初,党中央、国务院根据国内外经济形势的变化,及时作出扩大内需、保持经济持续快速增长的重大决	{[5,7],ORG},
策。	{[9,11],ORG},lead

准备数据集的技巧在每个数据集项目中可以包含多个实体及其关系的文本数据,每个文本数据的实体数量以及关系数据可以不同。以下是文本实 体关系抽取任务的小tips,请您查收:

- 思考实体类型:根据您所需要的具体场景,来考虑您的文本数据中包含的实体类型数量
- 思考实体关系类型:根据您已有的实体,考虑各实体之间的关系

可能的疑问

• 什么是实体关系抽取?

答:实体关系抽取是指从文本中抽取出预定义的实体类型及实体间的关系类型,得到包含语义信息的实体关系三元组,每个实体关系三元组由两 个实体及其关系构成,如<实体1,实体2,实体关系>

- 如果训练文本数据无法全部覆盖实际场景要识别的文本,怎么办?
- 答:训练的模型算法会有一定的泛化能力,尽可能覆盖即可

心 在线标注

文本实体关系抽取数据标注

如果您上传的是文本实体关系抽取数据集为无标注数据集,为了整个模型的正确运行,您可以点击【去标注】根据以下两步操作完成实体关系抽取的数据标注工作。首先您需要创建实体类别,并设置实体关系,第二步您需要根据您上传的文本实体关系抽取数据,对每个样本选择实体,标 注实体类别和关系。下面将详细介绍以上两个步骤。 **创建实体类别,设置实体关系**首先,我们需要对上传的无标注数据进行实体类型的设置与 实体关系类型的设置。如下图所示,首先点击右上角的【添加/修改】功能。

在弹出来的弹框中,创建实体类别标签,实体类别选择时相互独立。在创建关系标签时,关系标签目录下,第一个空格为关系类型,后两个空格 为实体类别,点击空格后在下拉框中选择相应的实体类型,即第二个空格为实体1,第三个空格为实体2。需要注意的是,在关系类型的创建中, 与实体类型相组合可以明确实体关系的方向,如【因果关系】中,如果实体1为原因,实体2为结果,则关系方向为【因果关系,实体1,实体 2】,反之则为【因果关系,实体2,实体1】。

Bai 公大脑 EasyDL 产品介绍	操作平台	立用案例 使用文档		社区 飞桨官网 AI主站
		 添加关系标签 ⑦ 理理实件类别称签 1 □ 	×	定义的实体类型及实体间的关 收起 ^ "可以抽取出⊀王鲁纯-配音-杠
模型中心	在线标注 > ;	谢输入文本标签名称		zzy-test-relation V1 ▼
我必須轉型」	无极注信	● 創建关系标签		- 标签栏 + 添加修改
训修养模型	定位义	123 1 2 2	Ē	请输入关系标签名称 Q
校验機型 发布機型	地理位置	請給入关系标签名称 请指定文本标签 《 请指定文本标签 《 请指定文本标签 》 请指定文本标签 》	+=	123
EasyData数据服务 数据总统 在线线注	也就十多.	- Materia		

在创建实体类别与关系类别的时候,需注意以下几点:

• 创建的实体类别标签不能相同,同理创建的关系标签也不能相同。

• 每个实体类别名称不超过10个字,每个实体关系名称也得在10个字以内。

选择实体,标注实体类别和关系在上述创建完实体类别,设置实体关系后,针对上传的未标注数据,对数据中每个样本进行实体标注以及关系标 注。 **实体标注** 对每个样本实体标注时,点击鼠标左键,覆盖您想要标注的实体字符,在弹出的选项框中选中您所要的实体关系类别,在一个文本中,可以存在多个实体。采样以上方法,根据您的具体样本,可进行详细标注。

Bai也大脑	EasyDL 产品介绍	操作平台 应用案例 使用文档	社区 飞桨官网 AI主站
	创建模型	大阪江信息(1)	1990日 · 1980年1994ス 遺論入关系が送名称 Q
	vinee便型 校验模型		123
	发布模型 EasyData数据服务	(」IIIIのAGGI / ITTINAGU () 定田(オレイロの) / III(U) (7, SRRT,) PER① 郑権夷、PER① 中朝朝、PER① 事業 百日〇	nation
	数据总览在线标注	Act 知識集会 知識集会 認識非会体後期販売: Act 五星大阪 School School Skipc,	Graduate birthday
	EasyEdge本地部署 纯离线服务	Text 百里大阪 TWorks Act - 五里大阪	Act Direct
		PER 刘纪期为 Direct 五期大祝	
		Act	

关系标注

当您完成一个文本的实体类别标注后,您可以对实体间的关系类型进行标注。选中实体1,点击鼠标左键,移动鼠标到实体2,鼠标点击左键则两 个实体之间建立联系,此时在弹出的实体关系标签中选中所需标签完成关系标注。点击右键此时产生一条曲线,连接实体1和实体2。当标注完成 后,可以看到标注结果栏产生刚刚标注完成的实体关系。任意点击标注栏的实体关系,可在文本内容上看到对应的关系。如您想删除已标注的关 系,可在标注结果部分选中关系点击删除即可。

模型中心	在线标注 > binbin_test_1208 > 标注		binbin_test_1208 V1
我的模型			
创建模型	无标注信息(0) 有标注信息(2000)	查看标注说明	标签栏 + 添加/
训练模型	定位文本内容	回 删除文本 上一篇 下一篇	请输入关系标签名称
校验模型			103
发布模型	(五星大版店) TVWorks ⑦ 中市 NiAladith PEP の P. 2660年、	标注结果	nation
EasyData教伊服务	PER① 郑柔美、PER① 4 Act 李	Act	nationality
*####	泰、PER① 字崇霄联 PER Direct	郑柔美、	Graduate
Triblert		Act 五星大饭	birthday
EssyEdge本地報題		Act - - 五星大坂…	Act
LasyLuge本地回答		Act 一 曹曦文、	Direct
10.000		五星大坂	
		Direct	
		五星大饭	
大聽 │EasyDL 产品介绍	操作平台 应用案例 使用文档 大标注信息(0) 有标注信息(2000)	Direct 五星大阪 	社区飞梁百网
大廳 EasyDL 产品介绍 創建模型	操作平台 应用案例 使用文档 无标注信息(0) 日報社信息(2000)		<u>社区 飞来官网</u> (5)12112 (7)191 (2010)、主衆国の学校会
大脑 EasyDL 产品介绍 创建模型 训练模型	操作平台 应用案例 使用文档 末続注意思(0) 存転注意思(2000) 定位文本内容 Q	五星大弦 重書縣注(3)明	社区 毛袋首网 1952日 1953 開始人关系研究名称
大廳 EasyDL 产品介绍 創建模型 同時模型 校验模型	操作平台 成用案例 使用文档 大統正信息(0) 存転注信息(2000) 定位文本内容 Q (石豊大坂内) WWMM (6) 由a 30 / 988 / PER (6) 国 31 / 988	五星大卷 	<u>北区 飞楽官网</u> Winitia (1997) 環路入美原原語名称 123
大簡 EasyDL 产品介绍 部連規型 初連規型 初時規型 技能構型 技能構型 发売機型	操作平台 应用案例 使用文档 未続注信息(0) 448注信息(2000) 定位文本内吉 Q (五重大版法) TVWorks (① 由曲文化/分明, PER(① 事) 数数字, PER(① 郑葉集, PER(① 牛菊病, PER(① 雪蔵、PER(① 事)	Date: 五星大饭 重書板注説明 回 删除文本 上一篇 下一篇 标注結果 Act	計区 飞梁首网 Windtha Paper 開始入关系研究会成 123 nation
大師 EasyDL 产品介绍 創建構型 閲想構型 校論構型 校論構型 发佈構型 EasyData数据服务	提作平台 应用室例 使用文档 大标注信息() 有标注信息(2000) 定位文本内容 Q 「正二大版品」 TWMorks (6) 是由文化/朝林、PER() 年 张敏宁、 PER() 发発美、PER() 牛萌病、PER() 雪道文、PER() 非 要 変、PER() 牛萌病、PER() 雪道文、PER() 非 変、PER() 学動調整 PER() 性主派的部件 PER() 情報局	五星大弦 五星大弦 重響時注説明 自删除文本 上一篇 下一篇 标注结果 标注结果 Ad 五星大饭 云素問試 云素問試	社区 飞来首网 with L ・ with 調査人主奏感感会像 123 nation nationality Graduate
大館 EasyDL 产品介绍 分離構型 の健構型 回離構型 短途構型 反応構型 反応構型 反応構型 反応構型 反応構型 反応構型 反応構型 反応構型 反応構型	操作平台 広用案例 使用文档 大級注意思(の) 有以社位起(2000) 定位文木内容 Q (五星大坂治) TWildes @ 由曲 知心得掛, PER ① 等, 孫敏宁, PER ① 郑墨美、PER ① 牛萌病、PER ① 雪載文、PER ① 李 奏、PER ① 李號電戰 PER ① 快主演的部中 PER ① 傳感副	Dalk Dalk	出区 飞梁首网 INAL 1992年 1993 123 nation nationality Graduate birthday
大館 EasyDL 产品介绍 御建模型 回修模型 授始模型 授始模型 授始模型 授始模型 授始模型 授始模型 授始模型 授始模型 授物理表型 授物理表型 授助理表型 希望表示 在代新主	操作平台 広用案例 使用文档 末期注意意() 有級注信に(2000) 定位文本内容 Q (五重大版四) 「WWorks () 是由 別心得助, PER () 事, 张敏宇, PER () 茅業美、PER () 牛胡病、PER () 雪道文、PER () 幸 素、PER () 李就質戰 PER () 姓主演的部本 PER () 博感副	Diece 五星大阪 五星大阪 五星大阪 五星大阪 本 本 香藤桃注湖朝 南 南京注結果 和 石屋大阪 和 石屋大坂	社区 飞楽首网 With Parts 第級人关系称称的名称 123 nation nationality Graduate bithday Act
大館 EasyDL	操作平台 放用条例 使用文档 大部注信息(0) 存品注信息(2000) 大部注信息(2) 存品注信息(2000) 定位文本内答 Q (五重大版品) 10000 mm (五重大版品) 10000 mm (五重大版品) 10000 mm (平民①) 5年美、PER① 牛胡鳴、PER① 電気、PER① 季 家、PER① 牛胡電歌 PER① 株主演的部市 PER① 構築器	五星大阪 五星大阪 一一二二二 一二二二 一二二二 「「二二二」 「二二二」 「二二」	社区 2次官府 P 2012年1日 P 2014 123 nation nationality Graduate birthday Act Direct
大館 EasyDL	操作平台 放用案例 使用文档 大統計信号(0) 合統社公共(2000) 定位文本内否 Q (五重大版石) (Works ⑥ 由面 知)分辨为 FER ① 事 张敏宁, FER ① 邓承美、FER ① 牛萌娘、FER ① 零載文 FER ① 零 表、FER ① 李號電戰 FER ① 快主演的部中 FER ① 傳感副	Dallaction 五星大阪 由制除文本 上一篇 下一篇 修建编集 一 季素用联 五星大坂 Ad - 季泰、 Ad - 万星大坂	計区 2次首内 P 994 1994年 P 994 123 nation nationality Graduate brithday Act Direct
大郎 EasyDL	操作平台 広用案例 使用文档 大級注意思() 有以社会局(2000) 定位文木内容 Q 定位文木内容 Q (五星大坂法) 「Wildows @ 黒田知心研細, PER ① 町, 孫敏宁, PER① 邦泰美、PER① 中柳菊、PER① 町載之、PER① 季 案、PER① 亨振爾戰 PER① 社主演的部中 PER① 傳感副	Discet 五星大阪 五星大阪 重着新注決時 南 删除文本 上一篇 下一篇 标注結果 Act 「五星大阪 Act 「五二 「五星大阪 Act 「二 二 「二 二 二 二 二 二	社区 C装白网 World Press 開始人主用級語名称 123 nation nationality Graduate brithday Act Direct

根据您上传的数据样本,每个样本中的实体数与关系数可以不同。

为两个样本是重复样本。例如:

文本内容	实体关系1
今年年初,党中央、国务院根据国内外经济形势的变化,及时作出扩大内需、保持经济持续快速增长的重大决策。	{[5,7],ORG},{[9,11],ORG},lead
今年年初,党中央、国务院根据国内外经济形势的变化,及时作出扩大内需、保持经济持续快速增长的重大决策。	{[5,7],ORG}, {[9,11],ORG},friends
今年年初,党中央、国务院根据国内外经济形势的变化,及时作出扩大内需、保持经济持续快速增长的重大决策。	{[5,7],ORG}, {[9,11],ORG},friends

上表三个样本均为重复样本,前两个样本虽然实体关系不同,但文本内容一致,为重复样本,后两个样本的文本内容与实体关系都一致,则也为重复样本。根据文本出现的顺序,最后一次的重复样本将代替之前的重复样本。

小Tips:"如何利用好重复样本"如果您的数据存在样本种类不均衡的现象,您可以通过将重复样本数量小的那一类,使其样本数量增加到与数据 量大的那一类样本数量相近,以提高模型训练的效果,这种方法也称为"上采样"。

平台去重策略 平台提供了可去重的数据集,即对您上传的数据进行重复样本的去重。注意:当您确定了数据集为去重或非去重的属性后,便不可 修改。

当您创建了一个去重的数据集时,在后续上传数据的过程中,平台可通过检验您当前上传的样本与已上传到此数据集下的样本是否相同,如果相同,则会使用新的样本替代旧的样本。此时分为几种情况,如下:

- 数据集中有未标注样本,上传重复的已标注样本,此时未标注样本将被覆盖
- 数据集中有已标注样本,上传重复的未标注样本,此时已标注样本将被覆盖
- 数据集中有已标注样本,上传不同标注的已标注样本,此时已有的标注样本将被覆盖

模型训练

₯ 创建模型

步骤 Step 1 创建模型

在【模型中心】或者【模型中心-我的模型】点击创建模型。

Step 2 填写基本信息 选择模型类型、提交模型名称、模型描述、联系方式即可创建模型。

Step 3 查看已创建的模型 模型创建成功后,可以在【我的模型】中看到刚刚创建的模型。

1.创建模型后可持续新增模型版本,因此不必每次训练模型都创建模型

2.目前单个用户在每种类型的模型下最多可创建**10个**模型,每个模型均支持多次训练,若需要创建超过10个以上的模型,请在百度云控制台 内提交工单反馈。

3.如果您是企业用户,建议您按照真实企业信息进行填写,便于EasyDL团队后续更好的为您服务

心 评估效果

模型评估 模型训练完成后,可以在「我的模型」列表中查看该模型的效果。

「完整评估结果」页面中将记录整体评估的报告,包括该模型整体的F1-score、精确率、召回率。可以切换查看训练集与自定义测试集的效果评 估报告。



整体评估中,各指标的释义如下:

- F1-score:给每个类别相同的权重,计算每个类别的F1-score,然后求平均值
- 精确率:给每个类别相同的权重,计算每个类别的精确率,然后求平均值
- 召回率:给每个类别相同的权重,计算每个类别的召回率,然后求平均值
- **模型校验**

在完成训练后,发布模型前,可以先进行模型校验,以确保模型在实际环境中能获得预期的性能。操作方法如下:

- 1. 在左侧「模型中心」目录中点击「校验模型」,进入校验模型页面
- 2. 选择需要校验能力的模型、部署方式、版本,点击「启动模型校验服务」
- 3. 校验服务启动后,在左侧输入文本,点击「校验」后,右侧识别结果栏将输出预测结果,您就可以看见识别出的实体关系以及实体关系数
 - 校验数据支持两种输入方式,直接输入文本或上传txt格式文本,文本长度上限为512汉字
- 4. 若您对预测结果满意,可点击「申请上线」,进行模型的发布

心训练模型

训练模型

完成数据的标注,或提交已标注的数据后,即可在「模型中心」目录中点击「训练模型」,开始模型的训练。

按以下步骤操作,启动模型训练:

文本实体关系抽取 《三	训练模型							
認 总览								
① 模型中心	送择模型							
我的模型	训练配置							
创建模型	部署方式	公有云部署 EasyEdge本3	8部署 如何选择部署方式?					
训练模型	选择算法	 高精度 ⑦ 						
校验模型	模型等选指标 ⑦	模型兼顾Precision和R V	Ø					
发布模型	添加数据							
☑ EasyData数据服务								
数据总算	添加数据集	+ 请选择						
在线标注		数据集		版本	实体关系数量		操作	
自 公有云服务		tj-文本实体关系抽取-不去重		V6	5		移動	
在线服务								
题 EasyEdge本地部署	自定义测试集 ①							
纯高线服务	训练环境	名称	现格		算力	速度比例	价格	
		GPU P40	TeslaGPU_P40_24G显存单卡_	12核CPU_40G内存	12 TeraFLOPS	1	免费	
		GPU V100	TeslaGPU_V100_16G显存单卡	12核CPU_56G内存	14 TeraFLOPS	1.5	单卡¥0.45/分钟(348.62小时*节点小时包余档	Ø
	开始训练							

Step 1 选择模型 选择此次训练的模型 Step 2 训练配置

部署方式

可选择「公有云部署」、「EasyEdge本地部署」。

如何选择部署方式

选择设备

- 如果您选择了「EasyEdge本地部署」,请根据实际部署设备选择
- 如果您选择了「公有云部署」,无需选择设备

选择算法

提供「高精度」算法。

• 1000条标注数据,在P40机器上预计在10分钟左右完成训练

「高精度」算法内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义 等方面的新知识,实现模型效果不断进化 选择不同的模型选择方式,对应的模型各项效果指标将有所不同。如果没有特殊场景的要求,使用默认即可(兼顾Precision精确度和Recall召回率)。有以下指标可供选择:

- 模型兼顾Precision和Recall:挑选模型时,兼顾Precision精确度和Recall召回率,如场景中没有对精度或召回的特别要求,建议您使用此默认指标
- Precision最高的模型:挑选模型时,优先挑选Precision精度最高的模型作为部署模型
- Recall最多的模型:挑选模型时,优先选择召回率最高的模型作为部署模型
- ACC最大的模型:挑选模型时,优先挑选预测样本数最多的模型作为部署模型
- Loss最小的模型:挑选模型时,优先挑选预测偏差最小的模型

Step 3 添加数据

添加训练数据

- 需选择1个数据集。建议您使用的每个实体关系的样本数应达到1000个以上,再启动训练,如果您提交的训练数据中,每个实体关系的样本数 不足1000,可能会影响模型的训练效果
- 训练时间与数据量大小和您选择的模型类型有关

添加自定义测试集

上传不包含在训练集的测试数据,可获得更客观的模型效果评估结果。

添加自定义测试集的目的:

如果学生的期末考试是平时的练习题,那么学生可能通过记忆去解题,而不是通过学习的方法去做题,所以期末考试的试题应与平时作业不能一样,才能检验学生的学习成果。那么同理,AI模型的效果测试不能使用训练数据进行测试,应使用训练数据集外的数据测试,这样才能 真实的反映模型效果

期末考试的内容属于学期的内容,但不一定需要完全包括所学内容。同理,测试集的标签是训练集的全集或者子集即可

Step 4 训练模型

点击「开始训练」,即可开始训练模型。

- 训练时间与数据量大小、选择的算法、训练环境有关 (https://ai.baidu.com/ai-doc/EASYDL/pki8mxr4m)
- 模型训练过程中,可以设置训练完成的短信提醒并离开页面
- 平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

优惠政策:

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有声音分类 操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。

模型部署

心 整体介绍

训练完成后,可将模型部署在公有云服务器、私有化服务器上,通过API进行调用。 公有云API

- 模型训练完毕后,为了更方便企业用户一站式完成AI模型应用,文本实体关系抽取模型支持将模型发布成为在线的restful API接口,可以参考示例文档通过HTTP请求的方式进行调用,快速集成在业务中进行使用。
- 具有完善的鉴权、流控等安全机制,GPU集群稳定承载高并发请求

相关费用 将模型发布为公有云API后,将获得一部分免费调用次数,超出免费额度将根据调用次数进行收费。详见EasyDL价格文档。

私有服务器部署

支持将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便 捷。适用于对数据敏感度、隐私性要求较高、在线离线均有调用需求的企业场景。相关费用 EasyDL已支持将定制模型部署在本地服务器上,只 需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。如需购买永久使用授权,请提交工单咨询。

心 公有云API

の发布API

发布模型,生成在线API

训练完毕后可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可发布公有云API:

- 1. 选择模型
- 2. 选择部署方式「公有云部署」
- 3. 选择版本
- 4. 自定义服务名称、接口地址后缀
- 5. 申请发布

发布完成 申请发布后,通常的审核周期为T+1,即当天申请第二天可以审核完成。

如果需要加急、或者遇到莫名被拒的情况,请在百度云控制台内提交工单反馈。

つ调用API

接口描述

基于自定义训练出的文本实体关系抽取模型,实现个性化实体关系识别。模型训练完毕后发布可获得定制化实体关系抽取API。

更多训练模型过程中的常见问题请查看常见问题文档。

如有其它问题,请在百度云控制台内提交工单反馈。

请求说明 请求示例

HTTP 方法:POST

请求URL:请首先在定制化训练平台进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

```
参数 值
```

Content-Type application/json

值

注意:如果出现336001的错误码很可能是因为请求方式错误,服务以json方式请求。

Body请求示例:

```
{
"text": "<UTF-8编码文本>"
}
```

请求参数

参数	是否必选	类型	可选值范围	说明
text	是	string	-	文本数据,UTF-8编码。最大长度512个字符最大长度512个字符

注意:通过API接口预测时,模型仅接收512个字符(包括汉字和标点符号)内的文本,超出将被截断。

返回说明

返回参数

字段	是否必 选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
error_co de	否	number	错误码,当请求错误时返回
error_ms g	否	string	错误描述信息,当请求错误时返回
content	是	string	预测的文本内容(最多仅支持512个字符,包括汉字、字母、符号等)
entities	是	array	此为实体数组,可包含多个实体
+alias	是	string	每个实体的识别信息,建议按顺序1、2、3、依次梳理标注的实体值,其中1代表第一个实体,2代表第二个实体,依次类推
+offset	是	array	实体所在的位置
+span	是	string	实体内容
+tag	是	string	实体的类别
relations	是	array	实体关系的数据,可以包括多个实体关系
+alias	是	string	每个实体关系识别信息,建议按顺序1、2、3、依次梳理标注的实体关系值,其中1代表第一个实体关系,2代表 第一个实体关系,依次类推
+predica te	是	string	实体关系类型
+from	是	string	实体关系的主体实体
+to	是	string	实体关系的目标实体

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有 疑问请在百度云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问 请在百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大512 UTF-8字符。
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云 控制台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

の调用API

接口描述

基于自定义训练出的文本实体关系抽取模型,实现个性化实体关系识别。模型训练完毕后发布可获得定制化实体关系抽取API。

更多训练模型过程中的常见问题请查看常见问题文档。

如有其它问题,请在百度云控制台内提交工单反馈。

请求说明 请求示例

HTTP 方法:POST

请求URL:请首先在定制化训练平台进行自定义模型训练,完成训练后可在服务列表中查看并获取url。



Header如下:

参数 值 Content-Type application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,服务以json方式请求。

```
Body请求示例:
```

```
{
"text": "<UTF-8编码文本>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text	是	string	-	文本数据,UTF-8编码。最大长度512个字符最大长度512个字符

注意:通过API接口预测时,模型仅接收512个字符(包括汉字和标点符号)内的文本,超出将被截断。

返回说明

返回参数

字段	是否必 选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
error_co de	否	number	错误码,当请求错误时返回
error_ms g	否	string	错误描述信息,当请求错误时返回
content	是	string	预测的文本内容(最多仅支持512个字符,包括汉字、字母、符号等)
entities	是	array	此为实体数组,可包含多个实体
+alias	是	string	每个实体的识别信息,建议按顺序1、2、3、依次梳理标注的实体值,其中1代表第一个实体,2代表第二个实体,依次类推
+offset	是	array	实体所在的位置
+span	是	string	实体内容
+tag	是	string	实体的类别
relations	是	array	实体关系的数据,可以包括多个实体关系
+alias	是	string	每个实体关系识别信息,建议按顺序1、2、3、依次梳理标注的实体关系值,其中1代表第一个实体关系,2代表 第一个实体关系,依次类推
+predica te	是	string	实体关系类型
+from	是	string	实体关系的主体实体
+to	是	string	实体关系的目标实体

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有 疑问请在百度云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问 请在百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大512 UTF-8字符。
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云 控制台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

の 发布API

在训练模型时,您需要选择「EasyEdge本地部署」的训练方式,才能发布本地部署的私有API。

私有API介绍

将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷

发布私有API的流程训练完毕后,您可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可将模型部署到私有服务器:

1. 在「发布模型」页面中,选择模型及模型的版本,选择部署方式为「EasyEdge本地部署」、集成方式为「API-纯离线服务」。点击「发布」, 即可跳转至「发布新服务」页面

发布模型页面示意:

文才	▶实体关系抽取	发布模型	
	总览		
1	模型中心	选择模型	zlx文本实体关系抽取自定义测试集
	我的模型	部署方式	公有云部署 🗸
	创建模型		
	训练模型	选择版本	V1 ~
	校验模型	服务名称 *	
	发布模型	接口地址 *	https://aip.baidubce.com/rpc/2.0/ai_custo
<u></u>	EasyData数据服务		m/v1/entity_relationship_xtr/
	数据总览		
	在线标注	其他要求	若接口无法满足您的需求,请描述希望解决
	公有云服务		的问题,500汉字以内
	在线服务		
F1	EasyEdge本地部署		0/500
	纯离线服务		提交申请

2. 在「发布新服务」页面,选择部署类型,填写服务名称、证书生效时间等信息,选择对应的系统和芯片。

- 部署类型可支持单模型部署和增量部署
- 增量部署申请,指需要在一台服务器上部署多个模型部署包时使用。进行增量部署时,需在「已部署服务」选择同台服务器历史中最近 部署的部署包,此步骤用来关联不同部署包中的license文件

3. 上传指纹文件。详细操作见指纹提取工具说明,可通过指纹工具进行指纹的提取

4. 点击下一步,填写个人详细信息后即可发布。发布完成后,即可在服务器目录下看到发布处于审核中的状态

个人信息的填写仅供EasyDL团队了解您,便于后续合作接洽,不会作为其他用途使用

5. 等待审核通过,前往「纯离线服务」等待部署包制作完成后,下载部署包,并参考文档完成集成

价格说明 EasyDL已支持将定制模型部署在本地服务器上,只需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。

如需购买永久使用授权,请提交工单咨询。

の调用API

本文档主要说明定制化模型本地部署后,如何使用本地API。如还未训练模型,请先前往训练。

如有疑问可以通过以下方式联系我们:

• 在百度云控制台内提交工单

- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:868826008)联系群管

部署包使用说明 部署方法

EasyDL定制化文本实体关系抽取模型的本地部署通过EasyPack实现,目前提供单机一键部署的方式。

在EasyDL申请、下载部署包后,在本地服务器新建目录(建议目录命名规则:easyDL_服务名称_模型版本号),将软件包上传至该目录。请参考EasyPack-单机一键部署使用python2版本来部署,部署成功后,启动服务,即可调用与在线API功能类似的接口。

运维检查

EasyDL服务器API部署应用健康检查(或故障排查)脚本:trouble_shooting.tar

脚本能力:鉴权服务健康检测、容器状态检查、端口探活、网络联通性测试、容器关键报错日志输出等

使用方法:将脚本上传至服务器任意目录(或在服务器直接下载),并解压后运行。

解压 tar vxf trouble_shooting.tar **执行** bash trouble_shooting.sh

授权说明

本地部署包根据服务器硬件(CPU单机或GPU单卡)进行授权,只能在申请时提交的硬件指纹所属的硬件上使用。

部署包测试期为1个月,如需购买永久授权,可提交工单咨询

API参考

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在EasyDL进行自定义模型训练,完成训练后申请本地部署,本地部署成功后拼接url。

请求URL: http://{IP}:{PORT}/{DEPLOY_NAME}/EntityRelationExtraction

- IP:服务本地部署所在机器的ip地址
- PORT:服务部署后获取的端口
- DEPLOY_NAME:申请时填写的本地服务名称

Header如下:

参数 值 Content-Type application/json

Body请求示例:

```
{
```

```
"text": "<UTF-8编码数据>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text	是	string	-	文本数据,UTF-8编码。最大长度512个字符(包括汉字、字母、数字和符号)

返回参数

字段	是否必 选	类型	说明		
log_id	是	number	唯一的log id,用于问题定位		
error_co de	否	number	错误码,当请求错误时返回		
error_ms g	否	string	错误描述信息,当请求错误时返回		
content	是	string	预测的文本内容(最多仅支持512个字符,包括汉字、字母、符号等)		
entities	是	array	此为实体数组,可包含多个实体		
+alias	是	string	每个实体的识别信息,建议按顺序1、2、3、依次梳理标注的实体值,其中1代表第一个实体,2代表第二个实体,依次类推		
+offset	是	array	实体所在的位置		
+span	是	string	实体内容		
+tag	是	string	实体的类		
relations	是	array	实体关系的数据,可以包括多个实体关系		
+alias	是	string	每个实体关系识别信息,建议按顺序1、2、3、依次梳理标注的实体关系值,其中1代表第一个实体关系,2代表 第一个实体关系,依次类推		
+predica te	是	string	实体关系类型		
+from	是	string	实体关系的主体实体		
+to	是	string	实体关系的目标实体		

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如缺少必要出入参时返回:

```
{
    "error_code": 336001,
    "error_msg": "Invalid Argument"
}
```

出现的 出现合自

+#++

阳庆归	相伏向心	通灯
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(868826008)或工单联系技术支持团队
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或工单联系技 术支持团队
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数代码格式是否有误。有疑问请通过QQ群(868826008)或工单联 系技术支持团队
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问请在 百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大512 UTF-8字符。有疑问请通过 QQ群(868826008)或工单联系技术支持团队
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到清重试,如反复失败,请在百度云控制 台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请通过QQ群(868826008) 或工单联系技术支持团队
337000	Auth check failed	离线鉴权调用失败

模型更新/回滚操作说明 模型更新 1、在EasyDL-纯离线服务发布页面,找到您的服务器API发布记录,点击【更新版本】,选择「更新包」或 「完整包」来发布。 两者区别:

包类型	描述
更新包	仅包含最新的模型应用,需执行download.sh脚本下载所需镜像等依赖文件
完整包	包含模型应用和其他鉴权服务,需执行download.sh脚本下载所需完整依赖文件

2、(CPU模型可忽略)如果您训练的模型为GPU版本,系统会生成多份下载链接。请在GPU服务器执行 nvidia-smi命令,根据返回的Cuda Version来选择对应的部署包链接下载。

3、在服务器新建目录(建议标记对应模型的版本号,便于区分不同模型版本),如easydl_\${DEPLOY_NAME}_v2

\${DEPLOY_NAME}:申请时填写的服务名称

以如下场景举例说明:模型版本从V1升级至V2

```
**1.新模型准备**
**创建指定版本的目录**
mkdir easedI_${DEPLOY_NAME}_v2
cd easedl_{DEPLOY_NAME}_v2
**将部署包上传至服务器该目录并解压**
tar zvxf xx.tar.gz
**解压后,进入指定目录执行download脚本下载模型所依赖文件**
cd original && bash download.sh
**2.旧模型备份**
**历史模型备份**
cp -r /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V1
**记录当前模型的端口号**
docker ps -a |grep ${DEPLOY_NAME}
**3.模型升级**
cd package/Install
**卸载当前已安装的旧的easyDL服务: ${DEPLOY_NAME},前面已备份**
python2 install.py remove ${DEPLOY_NAME}
**安装当前部署包内新的EasyDL服务: ${DEPLOY_NAME}**
python2 install.py install ${DEPLOY_NAME}
**(可选操作)更新证书**
python2 install.py lu
```

模型回滚 以如下场景举例说明:模型版本从V2回滚至V1

重命名当前v2模型目录名称
mv /home/baidu/work/\${DEPLOY_NAME} /home/baidu/work/\${DEPLOY_NAME}_V2
使用V1版本
cp -r /home/baidu/work/\${DEPLOY_NAME}_V1 /home/baidu/work/\${DEPLOY_NAME}
停止当前模型容器
docker ps -a |grep \${DEPLOY_NAME}
docker rm -f \${容器名}
创建新的容器
cd /home/baidu/work/\${DEPLOY_NAME} && bash start/start-1.sh
** (可选操作) 进入V1版本部署包所在目录执行license更新操作,假如部署包在/opt目录下,以您实际目录为准**
cd /opt/easydl_\${DEPLOY_NAME}_V1/original/package/Install && python2 install.py lu

方法二:

进入模型V1所在目录,参考上述【模型更新】步骤,执行模型升级操作(即先卸载v2,后升级为v1)

文本实体抽取

整体介绍

心简介

定制文本实体抽取模型,实现对文本进行内容抽取,并识别为自定义的实体类别。平台提供文本实体标注的工具,您可在平台上传文档,完成标注后可直接进行模型训练。

更多详情访问: EasyDL自然语言处理方向

∞ 应用场景

对内容中的关键实体进行识别和抽取,如:

- 金融研报信息识别
- 法律案件文书实体抽取
- 医疗病例实体抽取
- 其他:尽情脑洞大开,训练你希望实现的文本实体抽取的模型

心 技术特色

文本实体抽取模型内置<u>文心大模型</u>,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等 方面的新知识,实现模型效果不断进化。

文心大模型是百度发布的产业级知识增强大模型,是千行百业AI开发的首选基座大模型。文心大模型既包含基础通用的大模型,也包含了面向重 点领域和重点任务的大模型,还提供丰富的工具与平台,支撑企业与开发者进行高效便捷的应用开发。"知识增强"是文心的核心特色,文心能够 同时从大规模知识和海量多元数据中持续学习,如同站在巨人的肩膀上,训练效率和理解准确率都得到大幅提升,并具备了更好的可解释性。

```
∞ 使用流程
```

训练模型的基本流程如下图所示,全程可视化简易操作,在数据已经准备好的情况下,最快15分钟即可获得定制模型。



数据准备

の API上传

本文档主要说明当您线下已有大量的已经完成整理的文本数据,如何通过调用API完成文本数据的便捷上传和管理。

EasyDL数据集管理API在管理不同模型数据类型之间是通用的。上传不同模型类型数据,只是在部分接口入参存在差异,使用及接口地址完全一

致。

数据集创建API

接口描述

该接口可用于创建数据集。

接口鉴权

同发布模型后获取的API鉴权方式:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/create

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,TEXT_ENTITY_EXTRACTION
dataset_name	是	string	数据集名称,长度不超过20个utf-8字符

若上传声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
dataset_id	否	number	创建的数据集ID

查看数据集列表API

接口描述

该接口可用于查看数据集列表。返回数据集的名称、类型、状态等信息。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/list

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:



Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,TEXT_ENTITY_EXTRACTION
start	否	number	起始序号,默认为0
num	否	number	数量,默认20,最多100

返回说明

返回参数

字段	必选	类型	说明	
log_id	是	number	唯一的log id,用于问题定位	
total_num	否	number	数据集总数	
results	否	array(object)	数据集列表	
+dataset_id	否	number	数据集ID	
+dataset_name	否	string	数据集名称	
+type	否	string	数据集类型	
+status	否	string	数据集状态	
+special_status	否	string	数据集特殊状态,包括shared、smart和空值,分别表示共享中、智能标注中、非特殊状态	

查看实体类别列表API

接口描述

该接口可用于查看数据集的实体类别。返回实体类别的名称、包含数据量等信息。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/list

URL参数:

参数 值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,TEXT_ENTITY_EXTRACTION
dataset_id	是	number	数据集ID
start	否	number	起始序号,默认0
num	否	number	数量,默认20,最多100

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	实体类别总数
results	否	array(object)	实体类别列表
+label_id	否	string	实体类别ID
+label_name	否	string	实体类别名称
+entity_count	否	number	样本数量

添加数据API

接口描述

该接口可用于在指定数据集添加数据。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法:<mark>POST</mark>

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/addentity

URL参数:

参数	值		

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,TEXT_ENTITY_EXTRACTION
dataset_id	是	number	数据集ID
appendLab el	否	boolean	确定添加标签/分类的行为:追加(true)、替换(false)。默认为追加(true)。
entity_cont ent	是	string	type为TEXT_ENTITY_EXTRACTION时,填入utf-8编码的文本。文本实体抽取限制512个字符(包括汉字、数字、 字母)
entity_nam e	是	string	文件名
labels	是	array(obje ct)	实体类别数据
+label_nam e	是	string	实体类别名称(由数字、字母、中划线、下划线组成),长度限制20B
+offset	是	array	文本实体抽取任务需要给出,是抽取的具体实体内容的位置,从entity_content中,第一个字符记为0起算,以数组"[n,m]"的形式填入

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

数据集删除API

接口描述

该接口可用于删除数据集。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/delete

URL参数:

参数	值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

字段	必选	类型	说明
type	是	string	数据集类型,TEXT_ENTITY_EXTRACTION
dataset_id	是	number	数据集ID

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

实体类别删除API

接口描述

该接口可用于删除实体类别。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/delete

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取

Header如下:

参数

值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,TEXT_ENTITY_EXTRACTION
dataset_id	是	number	数据集ID
label_name	是	string	实体类别名称

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

• error_code:错误码。

• error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
"error_code": 110,
"error_msg": "Access token invalid or no longer valid"
```

}

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求 ,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量 阶梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计 费;未上线计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶 梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
406000	internal server error	服务错误
406001	param[xx] invalid	参数xx不合法,请检查相关参数
406002	dataset not exist	数据集不存在
406003	dataset already exists	数据集已存在
406004	dataset can not be modified temporarily	数据集暂不可修改
406005	label not exist	标签/分类/实体类别不存在
406006	no permission to modify the dataset	没有修改数据集的权限
406007	dataset cannot be modified while smart annotation is running	智能标注期间不可修改数据集
406008	quota exceeded	配额超限

心 创建数据集并导入

创建数据集

在训练模型之前,需要在【数据总览】里面"创建数据集"。需输入数据集名称、选择相应的标注模版、选择数据去重策略,即可创建一个空数据 集。

1	ど楽EasyDL	产品介绍	绍 操作平台	应用案例	使用文档				
文Z	本实体抽取	÷	我的数据总览						
	总览								
ß	模型中心		EasyData智能数	据服务平台已上线	,使用EasyDat	a可享受包括多人标注、	数据清洗、数据采集等完整数据服务	立即前往	
	我的模型		创建数据集						
	创建模型								
	训练模型		智能标注实体抽取	マロン 数据集组	D: 312626				
	校验模型		版本	数据集ID		数据量	最近导入状态		标注类型
	发布模型								
~	EasyData数据服务		V1 😔	337627		0	● 导入失败 ②		又本实体抽取
	数据总览								
	标签组管理		pt又本实体抽取测	リロロン 数据集	戦回し: 311647				

数据自动去重即平台对您上传的数据进行重复样本的去重。建议创建数据集时选择「数据自动去重」

导入无标注数据创建数据集后,在「数据总览」页面中,找到该数据集,点击右侧操作列下的「导入」,即可进入导入数据页面。

1	ど桨EasyDL	产品介	诏	操作平台	应	用案例	使用文材	Ш.		
文Z	本实体抽取	Ē		我的数据总览 >	智能相	示注实体抽取八	/1/导入			
	总览			47本片白						
Ē	模型中心			刨建信息 ∨						
	我的模型			数据集ID	33762	27		版本号	V1	
	创建模型			备注	Ľ					
	训练模型		I	标注信息 🗸						
	校验模型			标注类型	文本实	6体抽取		标注模板	文本实体抽取	
	发布模型			数据总量	0			已标注	0	
22	EasyData数据服务			标签个数	0			待确认	0	
	数据总览			大小	0M					
	标签组管理		I	数据清洗						
	在线标注			暂未做过数据清	洗任务	5				
	智能标注		I	导入数据						
8	公有云服务			数据标注状态		○ 无标注信息	3	○ 有标注信/	息	
	在线服务			导入方式		请选择			^	
H	EasyEdge本地部署				ſ	本地导入				
	纯离线服务					BOS目录导)				
						分享链接导)	<			

您可以使用3种方案上传数据,分别为:

- 本地导入
- BOS目录导入
- 分享链接导入

本地导入您可以通过以下三种方式进行本地数据的导入:

文本实体抽取 ● 我的数据总览 〉 智能标注实体抽取/V//导入 B 总览 ● 根型中心 我的模型 数据集D 337627 版本号 V1 台建模型 数据集D 337627 版本号 V1 台建模型 前线模型 备注 ① 日本文体集中取 台建模型 1 标注信息 ~ 校验模型 ● 大小 日标注 0 交布模型 0 日标注 受加合数 0 日标注 0 放射磁型 0 日标注 0 砂糖合型 6 0 日标注	<i>ᡶ^ᡃᡄ</i> 桨EasyDL	产品介绍	沼 操作平台	应用案例	使用文档		
····································	文本实体抽取	÷	我的数据总览	> 智能标注实体抽耳	双/V1/导入		
● 健型中心 ● 創題権息 > ● 教的模型 数据集印 337627 版本号 V1 ● 創建模型 ● 注 ● 三 ● ○ ●	品 总览						
数約模型 数据集D 337627 版本号 V1 通建模型 	曲 模型中心		┃创建信息 ∨				
	我的模型		数据集ID	337627	版本号	V1	
训练模型I 标注信息 ~校验模型标注失型尔本实体抽取标注 板文本文体抽取发布模型数据总量0日标注0方小0行60标签组管理大小0存长场注音相优过数据标注音卡.依过数据标注长智能标注15資格成路055資格就路為05資源公有云服务資高以后向資源EasyEdge本地部署 術家服务資源<たる法定の資源近資源たは資源たは日日日日日日日日日日日日 </th <th>创建模型</th> <th></th> <th>备注</th> <th></th> <th></th> <th></th> <th></th>	创建模型		备注				
校验模型 「标注类型 文本文体抽取 师记 中 公式 ○	训练模型		┃标注信息 〜				
发布模型 数据总量 ○ 已标注 ○ basyData数据服务 ○ 行确认 ○ 数据总览 ○ 行确认 ○ 数据总览 ○ 行确认 ○ 数据总路 ○ ○ ○ ○ 校弦相管理 日数据清洗 ○ ○ ○ ○ 存线标注 日 State □ ○ ○ ○ 管能标注 日 State □ ○ ○ ○ 資格成方 ○ 万 ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	校验模型		标注类型	文本实体抽取	标注模板	文本实体	抽取
○ EasyData数据服务 「标签个数 0 放据总文 ○ ○ 行确认 0 太小 ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	发布模型		数据总量	0	已标注	0	
数据总览 大小 OM 标签组管理 I 数据清洗 在线标注	☑ EasyData数据服务		标签个数	0	待确认	0	
标签组管理 I 数据清洗 在线标注 暂未做过数据清洗任务 查能标注 I 导入数据 算能标注 J 导入数据 公有云服务 数据标注状态 在线服务 导入方式 Axbept 小市公社信息 第3次Edge本地部署 小商线服务 近期 近日	数据总览		大小	0M			
在线标注 暂未做过数据清洗任务 冒能标注] 导入数据 公有云服务 数据标注状态 五标注信息 有标注信息 在线服务 导入方式 本地导入 请选择 へ 認 EasyEdge本地部署	标签组管理		数据清洗				
智能标注 J 导入数据 数据标注状态 五标注信息 方式 本地导入 在线服务 导入方式 本地导入 「请选择 へ 資源或服务 「小商以并返回	在线标注		暂未做过数据	清洗任务			
目 公有云服务 数据标注状态 ① 无标注信息 有标注信息 在线服务 导入方式 本地导入 请选择 ^ 22 EasyEdge本地部署	智能标注		┃ 导入数据				
在线服务 导入方式 本地导入 请选择 ^ 認 EasyEdge本地部署	目 公有云服务		数据标注状态	 无标注 	信息 🦳 有标注信息	3	
記 EasyEdge本地部署 強高线服务 か した 使 た の に の の の の の の の の の の の の の の の の の	在线服务		导入方式	本地导入		~	请选择 ^
纯离线服务 上传TXT文本 上传压缩包 上传压缩包	鋁 EasyEdge本地部署			确认并透	ž 🗆		上传Excel文件
	纯离线服务						上传TXT文本 上传压缩包

- 以压缩包的方式上传
- 以TXT文本文件方式上传
- 以Excel文件的方式上传

通过压缩包上传时,需注意:

- 压缩包内的一个文本文件将作为一个样本上传。压缩包格式为.zip格式,压缩包内文件类型支持txt,编码仅支持UTF-8
- 每组数据的数建议不超过512个字符,超出将被截断

通过TXT文本上传时,需注意:

- 文本文件内数据格式要求为"文本内容\n"(即每行一个样本,使用回车换行),每一行表示一组数据,每组数据的数建议不超过512个字符, 超出将被截断
- 文本文件类型支持txt,编码仅支持UTF-8,单次上传限制100个文本文件,最多可上传100万个文件

通过Excel文件上传时,需注意:

- Excel文件内首行为表头,每行为一个样本,每个样本字符数不得超过512个字符,超出将被截断
- 文件格式支持xlsx格式,单次可上传100个文件

BOS目录导入需选择Bucket地址与对应的文件夹地址。

请确保将全部文本已通过txt文件保存至同一层文件目录,该层目录下子文件目录及非相关内容(包括压缩包格式等)不导入。

分享链接导入 需输入链接地址。分享链接导入的要求如下:

• 仅支持来自百度BOS、阿里OSS、腾讯COS、华为OBS的共享链接

导入有标注数据创建数据集后,在「数据总览」页面中,找到该数据集,点击右侧操作列下的「导入」,即可进入导入数据页面。

您可以使用本地上传的方案上传数据。 您可以通过以下三种方式进行本地数据的导入:

• 以Excel文件的方式上传

● 以API的形式导入

通过Excel文件上传时,需注意:

- Excel文件内首行为表头,每行为一个样本,每个样本字符数不得超过512个字符,超出将被截断
- 文件格式支持xlsx格式,单次可上传100个文件

通过API上传时,需注意:

• 可参考以下文档:实体抽取API数据管理

什么是实体类别?

实体类别(Entity Type)是指某类事物的集合,每一类数据对象的个体称为实体,如人/角色(例如学生),对象(例如发票),概念(例如 简介)或事件(例如交易)。实体类别名标签名由数字、中英文、中/下划线组成,长度上限256字符

心 在线标注

通过平台导入「无标注信息」的数据集后,可对无标注数据进行标注操作。

创建标签

进入到待标注的数据集,您需要在右侧的标签栏中创建标签,点击「添加/搜索标签」后,即可输入标签名称,注意平台仅支持数字和字母的标 签名

Q

标次程						
你立动						



根据文本内容,选择唯一标签

添加完标签后,可以添加标签的备注信息,如下图:

	标签栏				0
	根据文本内容,选择标签	(标签	内容不	「支持续	扁辑)
	maxmaxss_1	$\overline{\uparrow}$	٢	•	Ū
填写备注	信息,最多20字符			确认	

标注实体

您可以在文中划选需要标注的文本,然后在弹出的下落标签中选择需要标注的标签,如下图。也可以在划选文本后,在右侧的标签区域点击选择 标签

定位内容	Q	
文本内容文本内容文本内容文本内容; 本内容,文本内容文本内容文本内容	文本内容,文本内容文本内容3 maxmaxss_1maxmaxss_1…	文本内容文本内容文本 :内容,文本内容文本
本内容文本内容文本内容文本内容	maxm	
	maxmaxss	
	maxmaxss_1max	

您也可以通过点击某个标签的「连续标注」功能(如图中红框标记),在文中通过连续划选文本内容来进行连续标注。如下图:
无标注信息(19)	有标注信息(0)				标签栏	Q
	Q	直 删除文本	上一篇 下一篇	PLANT	添加/搜索标签	A ii
15款花蔬交织的创 园艺中愈来愈; 思和创意的设计, F 单纯花语 营造 1.龟背竹有着宽	意花艺(组图) 流行的"厨房花园",在日常生活中也可以用花艺 形成了精美实用且极具亲和力的花蔬组合。 诗意 宽厚硕大的叶片,通常它为花材起到了烘托主题 4州了不全感常先士的句法与短社	艺来阐述。那些常见的花材与天然 1的作用。 <mark>黄色马蹄莲花朵</mark> PLAN	\$食材,经过巧妙的构 T 有着强烈的视觉效	FOOD DATE OBJ PER		
 米, 电背竹恰为已 花材: 马蹄莲 制作: 将马蹄 	是供了不会喧兵夺王的包装与陪衬。 PLANT 与 <mark>龟背叶 PLANT</mark> 兰扎成花束,加几片叶材在底部,插入瓶中即ī	J.				

在您需要进行在文中进行关键词检索时,可以在文本上方的「定位内容」区域,输入文本内容,便可在文本区域高亮出检索的关键词,以方便您 定位内容和标注。

定位内容	Q	<u></u>
由于许多编码只能表达有有有限的字 准,Unicode字符集非常大,它囊括 ASCII文本文件和UTF–8文本文件完	检测视图 了大多数已 È一致	maxmaxss_1maxmaxss_1通常它们只能用于表达几种语言。Unicode制定文本内容了一种试图能够表达所有已知语言的标 印的字符集。Unicode有多种字符 编码,其中最文本内容常见的是UTF–8,这种编码能够向后兼容ASCII,相同内容的的

查看标注信息

您可以在标签中,点击「高亮」图标按钮,来达到让文本区域高亮显示标注信息,以方便您的查看标注情况。图标按钮如下图所示:

	\$	忝加/搜索标签
定位内容	FOOD	_
に動花茶六(4)的(4)音花芋(4)の)	PLANT	▼ ◎ ⊂ 6 i
园艺中愈来愈流行的"厨房花园",在日常生活中也可以用花艺来阐述。那些常见的花材与天然食材,经过巧妙的构	DATE	
思和创意的设计,形成了精美实用且极具亲和力的花蔬组合。 单纯花语 营造语意	OBJ	
1.龟背竹有着宽厚硕大的叶片,通常它为花材起到了烘托主题的作用。 <mark>黄色马蹄莲花朵</mark> PLANT <mark>有着强烈的视觉效</mark>	PER	
果,龟背竹恰为它提供了不会喧宾夺主的包装与陪衬。		
花材: 马蹄莲 <mark>PLANT</mark> 与 <mark>角</mark> 背叶 PLANT		
制作:将马蹄兰扎成花束,加几片叶材在底部,插入瓶中即可。		

心 数据去重

重复样本的定义

一个文本实体抽取的样本包括文本内容和实体类别。重复样本的定义,是指您上传的数据中,存在两个样本的文本内容完全一致。则被判定为两 个样本是重复样本。例如:

以下3条都是文本实体抽取任务的重复样本,样本示例如下:

文本内容	实体类别
今天北京的空气不错	北京:loc
今天北京的空气不错	今天:date
今天北京的空气不错	北京:local

上述两个表中,都代表三个样本均为重复样本,后两个样本虽然标签不一,但文本内容一致,也为重复样本。

Tips: "如何利用好重复样本",如果您在模型训练过程中,需要通过增加某个类别标签的预测权重,可以通过增加此标签的重复样本来达到此目标。

平台去重策略

平台提供了可去重的数据集,即对您上传的数据进行重复样本的去重。注意:当您确定了数据集为去重或非去重的属性后,便不可修改。

当您创建了一个去重的数据集时,在后续上传数据的过程中,平台可通过检验您当前上传的样本与已上传到此数据集下的样本是否相同,如果相同,则会使用新的样本替代旧的样本。此时分为几种情况,如下:

- 1. 数据集中有未标注样本,上传重复的已标注样本,此时未标注样本将被覆盖
- 2. 数据集中有已标注样本,上传重复的未标注样本,此时已标注样本将被覆盖
- 3. 数据集中有已标注样本,上传不同标注的已标注样本,此时已有的标注样本将被覆盖

模型训练

心 创建模型

步骤 Step 1 创建模型

在【模型中心】或者【模型中心-我的模型】点击创建模型。

Step 2 填写基本信息 选择模型类型、提交模型名称、模型描述、联系方式即可创建模型。

Step 3 查看已创建的模型 模型创建成功后,可以在【我的模型】中看到刚刚创建的模型。

1.创建模型后可持续新增模型版本,因此不必每次训练模型都创建模型

2.目前单个用户在每种类型的模型下最多可创建**10个**模型,每个模型均支持多次训练,若需要创建超过10个以上的模型,请在百度云控制台 内提交工单反馈。

3.如果您是企业用户,建议您按照真实企业信息进行填写,便于EasyDL团队后续更好的为您服务

心 评估效果

模型评估 模型训练完成后,可以在「我的模型」列表中查看该模型的效果。

「完整评估结果」页面中将记录整体评估的报告,包括该模型整体的F1-score、精确率、召回率。可以切换查看训练集与自定义测试集的效果评 估报告。

部署方式 公有云API >> 版本 V2	~	
训练集 自定义测试集 自定义测试集 · · · · · · · · · · · · · · · · · · ·	2011年春秋得至北方大抗士发展,40秒得小岛号光常是很热客	
洞路約187012307, 当初加速时1050米021009年0070021380本来的508 训练时长 21分钟	NUTRADIC CHURACH, MADYBICTHRON,	
整体评估 zix文本实体抽取自定义测试集 V2效果优异。 如何优化效果?		
94.3%	● 92.6%	26■ ♥ 96.1%

整体评估中,各指标的释义如下:

• F1-score:给每个类别相同的权重,计算每个类别的F1-score,然后求平均值

- 精确率:给每个类别相同的权重,计算每个类别的精确率,然后求平均值
- 召回率:给每个类别相同的权重,计算每个类别的召回率,然后求平均值

如果单个文本实体类别的文本量在100条以内,评估指标的计算将出现误差,建议每个文本实体类别提供超过1000个实体样本

模型校验

在完成训练后,发布模型前,可以先进行模型校验,以确保模型在实际环境中能获得预期的性能。操作方法如下:

- 1. 在左侧「模型中心」目录中点击「校验模型」,进入校验模型页面
- 2. 选择需要校验能力的模型、部署方式、版本,点击「启动模型校验服务」
- 3. 校验服务启动后,在左侧输入文本,点击「校验」后,右侧识别结果栏将输出预测结果
 - 校验数据支持两种输入方式,直接输入文本或上传txt格式文本,文本长度上限为512汉字
- 4. 若您对预测结果满意,可点击「申请上线」,进行模型的发布

如果单个文本实体类别的文本量在100条以内,评估指标的计算将出现误差,建议每个文本实体类别提供超过1000个实体样本

心 效果优化

通过模型迭代、检查并优化训练数据、选择高精度模型等方法,能够提升模型效果。****模型迭代****

一个模型很难一次性就训练到最佳的效果,通常会需要结合模型评估报告和校验结果不断扩充数据和调优。

为此平台提供了模型迭代功能,即当模型训练完毕后,会生成一个最新的版本号,首次V1、之后V2……以此类推。可以通过调整训练数据和算法,多次训练,以获得适合业务需求的模型效果。

如果模型已经是上线状态,依然支持模型迭代,只是需要在训练完毕后更新线上服务接口,在接口地址不变的情况下可以持续优化效果。

检查并优化训练数据

- 检查是否存在训练数据过少的情况,建议每个文本实体类别提供超过1000个实体,如果低于这个量级建议扩充
- 检查测试模型的数据与训练数据的文本类型与风格是否一致,如果不一致,那么很可能会存在模型效果不错但实际测试效果较差的情况。针对 这种情况建议重新调整训练集,将训练数据与实际业务场景数据尽可能一致

选择高精度模型

在训练模型时,选择高精度的模型,将提升模型的预测准确率。

「高精度」算法内置<u>文心大模型</u>,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等方 面的新知识,实现模型效果不断进化。

心训练模型

训练模型

完成数据的标注,或提交已标注的数据后,即可在「模型中心」目录中点击「训练模型」,开始模型的训练。

按以下步骤操作,启动模型训练:

文本实体推取 (三	训练模型					
88 总览						
山 模型中心	选择模型	zbx文本实体抽取无自定 ~				
我的模型	005.62.22					
创建模型	部署方式	公有云部署 EasyEdge本:	由部署 如何选择部署方式?			
训练模型	选择算法	○ 高精度 ⑦ ○ 高性能 0	1			
校验模型	模型筛选指标 ②	模型兼顾Precision和R V	Ø			
发布模型	添加数据					
◎ EasyOata数据服务	(5 to \$2 12 #	1.200440				
数据总览	20.0120.4899	+ 14)514				
标签组管理		数据集	版本	实体类别数量		操作
在线标注		测试一下	V1	13		移移
智能标注		测试实体抽取	VI	6		移除
□ 公有云服务	navara A					
在残服务	BACAMMIN O					
認 EasyEdge本地部署	训练环境	名称	规格	算力	速度比例	价格
纯离线服务		GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1	免费
		O GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	1.5	单卡¥0.45/分钟 (348.62小时*节点小时包余额)
	开始训练					

Step 1 选择模型 选择此次训练的模型 Step 2 训练配置

部署方式

可选择「公有云部署」、「EasyEdge本地部署」。

如何选择部署方式

选择设备

- 如果您选择了「EasyEdge本地部署」,请根据实际部署设备选择
- 如果您选择了「公有云部署」,无需选择设备

选择算法

您可以根据训练的需求,选择「高精度」或「高性能」算法。不同的算法将影响训练时间、预测速度与模型准确率。

- 如果您选择了高精度的模型,模型预测准确率将更高。如果您手中的标注数据集样本较少(例如少于1000条),可选择「高精度」的算法。
 使用高精度的算法训练模型将会耗时更久,实验环境下1000个样本,预计在20分钟左右完成训练
- 如果您选择了高性能的模型,相同训练数据量的情况下,训练耗时更短,模型预测速度更快。使用10000条训练样本,将在15min内完成训练.
 同样的数据量情况下,效果比高精度的模型4-5%

「高精度」算法内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义 等方面的新知识,实现模型效果不断进化

模型筛选指标

选择不同的模型选择方式,对应的模型各项效果指标将有所不同。如果没有特殊场景的要求,使用默认即可(兼顾Precision精确度和Recall召回率)。有以下指标可供选择:

- 模型兼顾Precision和Recall:挑选模型时,兼顾Precision精确度和Recall召回率,如场景中没有对精度或召回的特别要求,建议您使用此默认 指标
- Precision最高的模型:挑选模型时,优先挑选Precision精度最高的模型作为部署模型
- Recall最多的模型:挑选模型时,优先选择召回率最高的模型作为部署模型
- ACC最大的模型:挑选模型时,优先挑选预测样本数最多的模型作为部署模型
- Loss最小的模型:挑选模型时,优先挑选预测偏差最小的模型

Step 3 添加数据

添加训练数据

- 可选择多个数据集。注意,文本实体抽取模型至少需要有1个及以上实体类别,样本数建议超过1000
- 训练时间与数据量大小和您选择的模型类型有关

添加自定义测试集

上传不包含在训练集的测试数据,可获得更客观的模型效果评估结果。

添加自定义测试集的目的:

如果学生的期末考试是平时的练习题,那么学生可能通过记忆去解题,而不是通过学习的方法去做题,所以期末考试的试题应与平时作业不能一样,才能检验学生的学习成果。那么同理,AI模型的效果测试不能使用训练数据进行测试,应使用训练数据集外的数据测试,这样才能 真实的反映模型效果

期末考试的内容属于学期的内容,但不一定需要完全包括所学内容。同理,测试集的标签是训练集的全集或者子集即可

Step 4 训练模型

点击「开始训练」,即可开始训练模型。

- 训练时间与数据量大小、选择的算法、训练环境有关
- 模型训练过程中,可以设置训练完成的短信提醒并离开页面
- 平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	7	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

优惠政策:

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有声音分类操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。可参考价格说明 模型部署

心 整体介绍

训练完成后,可将模型部署在公有云服务器、私有化服务器上,通过API进行调用。 公有云API

- 模型训练完毕后,为了更方便企业用户一站式完成AI模型应用,文本实体抽取模型支持将模型发布成为在线的restful API接口,可以参考示例 文档通过HTTP请求的方式进行调用,快速集成在业务中进行使用。
- 具有完善的鉴权、流控等安全机制,GPU集群稳定承载高并发请求

相关费用 将模型发布为公有云API后,将获得一部分免费调用次数,超出免费额度将根据调用次数进行收费。详见EasyDL价格文档。

私有服务器部署

支持将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便 捷。适用于对数据敏感度、隐私性要求较高、在线离线均有调用需求的企业场景。相关费用 EasyDL已支持将定制模型部署在本地服务器上,只 需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。如需购买永久使用授权,请提交工单咨询。

```
心 公有云API
```

```
の发布API
```

发布模型,生成在线API

训练完毕后可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可发布公有云API:

- 1. 选择模型
- 2. 选择部署方式「公有云部署」
- 3. 选择版本
- 4. 自定义服务名称、接口地址后缀
- 5. 申请发布

发布完成 申请发布后,通常的审核周期为T+1,即当天申请第二天可以审核完成。

如果需要加急、或者遇到莫名被拒的情况,请在百度云控制台内提交工单反馈。

```
の调用API
```

接口描述

基于自定义训练出的文本实体抽取模型,实现个性化文本实体识别。模型训练完毕后发布可获得定制化实体抽取API 详情访问:定制化训练和服 务平台进行训练。

更多训练模型过程中的常见问题请查看常见问题文档。

如有其它问题,请在百度云控制台内提交工单反馈。

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在定制化训练平台进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

参数	值			
----	---	--	--	--

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/jsor

注意:如果出现336001的错误码很可能是因为请求方式错误,服务以json方式请求。

Body请求示例:

```
{
"text": "<UTF-8编码文本>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明	
text	是	string	-	文本数据,UTF-8编码。	最大长度512个字符

注意:通过API接口预测时,模型仅接收512个字符(包括汉字和标点符号)内的文本,超出将被截断。

请求示例代码

Python3	
	_
==== EasyDL 文本实体抽取 调用模型公有云API Python3实现 ====	
import json	
import base64	
import requests	
使用 requests 库发送请求	
使用 pin (或者 pin3) 检查我的 python3 环境是否安装了该库,执行命令	
pip freeze grep requests	
若返回值为空,则安装该库	
pip install requests	
11 H	
**目标立本的 本地文件路径	
TEXT_FILEPATH = "【您的测试文本数据地址,例如:./example.txt】"	-

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码,当请求错误时返回
error_msg	否	string	错误描述信息,当请求错误时返回
content	是	string	预测的文本内容(最多仅支持512个字)
results	否	array(object)	分类结果数组
+span	否	string	抽取的具体实体内容
+offset	否	array(number)	数组由两个元素组成,分别是实体的起始位置和终止位置,从entity_content中,第一个字符记为0起算
+tag	否	string	识别对应span的实体类别

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述	
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。	
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。	
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试	
4	Open api request limit reached	集群超限额	
6	No permission to access data	无权限访问该用户数据	
13	Get service token failed	获取token失败	
14	IAM Certification failed	IAM鉴权失败	
15	app not exists or create failed	应用不存在或者创建失败	
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈	
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈	
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈	
100	Invalid parameter	无效的access_token参数,请检查后重新尝试	
110	Access token invalid or no longer valid	access_token无效	
111	Access token expired	access token过期	
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈	
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有 疑问请在百度云控制台内提交工单反馈	
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈	
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问 请在百度云控制台内提交工单反馈	
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大512 UTF-8字符。	
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云 控制台内提交工单反馈	
336006	缺失必要参数	未上传文本文件	
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈	

在训练模型时,您需要选择「EasyEdge本地部署」的训练方式,才能发布本地部署的私有API。

私有API介绍

将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷

发布私有API的流程训练完毕后,您可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可将模型部署到私有服务器:

1. 在「发布模型」页面中,选择模型及模型的版本,选择部署方式为「EasyEdge本地部署」、集成方式为「API-纯离线服务」。点击「发布」, 即可跳转至「发布新服务」页面

发布模型页面示意:

文本实体抽取		发布模型			
BB 总览					
一 模型中心		选择模型	test_edge_文本实体抽取	\sim	
我的模型		部署方式	EasyEdge本地部署	\sim	
创建模型					
训练模型		选择版本	V2	\checkmark	
校验模型		集成方式	○ API纯离线服务		
发布模型					
🖉 EasyData数	舌服务	发布			
数据总览					
标签组管理					
在线标注					
智能标注					
目 公有云服务					
在线服务					
鋁 EasyEdge本	也部署				
纯离线服务					

2. 在「发布新服务」页面,选择部署类型,填写服务名称、证书生效时间等信息,选择对应的系统和芯片。

- 部署类型可支持单模型部署和增量部署
- 增量部署申请,指需要在一台服务器上部署多个模型部署包时使用。进行增量部署时,需在「已部署服务」选择同台服务器历史中最近 部署的部署包,此步骤用来关联不同部署包中的license文件
- 3. 上传指纹文件。详细操作见指纹提取工具说明,可通过指纹工具进行指纹的提取

4. 点击下一步,填写个人详细信息后即可发布。发布完成后,即可在服务器目录下看到发布处于审核中的状态

个人信息的填写仅供EasyDL团队了解您,便于后续合作接洽,不会作为其他用途使用

5. 等待审核通过,前往「纯离线服务」等待部署包制作完成后,下载部署包,并参考文档完成集成

价格说明 EasyDL已支持将定制模型部署在本地服务器上,只需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。

如需购买永久使用授权,请提交工单咨询。

つ调用API

本文档主要说明定制化模型本地部署后,如何使用本地API。如还未训练模型,请先前往训练。

如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:868826008)联系群管

部署包使用说明

部署方法

EasyDL定制化文本分类模型的本地部署通过EasyPack实现,目前提供单机一键部署的方式。

在EasyDL申请、下载部署包后,在本地服务器新建目录(建议目录命名规则:easyDL_服务名称_模型版本号),将软件包上传至该目录。请参考EasyPack-单机一键部署使用python2版本来部署,部署成功后,启动服务,即可调用与在线API功能类似的接口。

运维检查

```
EasyDL服务器API部署应用健康检查(或故障排查)脚本:trouble_shooting.tar
```

脚本能力:鉴权服务健康检测、容器状态检查、端口探活、网络联通性测试、容器关键报错日志输出等

使用方法:将脚本上传至服务器任意目录(或在服务器直接下载),并解压后运行。

解压 tar vxf trouble_shooting.tar **执行** bash trouble_shooting.sh

授权说明

本地部署包根据服务器硬件(CPU单机或GPU单卡)进行授权,只能在申请时提交的硬件指纹所属的硬件上使用。

部署包测试期为1个月,如需购买永久授权,可提交工单咨询

API参考

请求说明

请求示例

HTTP 方法: POST

```
请求URL:请首先在EasyDL自然语言处理方向进行自定义模型训练,完成训练后申请本地部署,本地部署成功后拼接url。
```

请求URL: http://{IP}:{PORT}/{DEPLOY_NAME}/EntityExtraction

- IP:服务本地部署所在机器的ip地址
- PORT:服务部署后获取的端口
- DEPLOY_NAME:申请时填写的本地服务名称

Header如下:

参数 值 Content-Type application/json

Body请求示例:

```
{
"text": "<UTF-8编码数据>"
}
```

```
Body中放置请求参数,参数详情如下:
```

请求参数

参数	是否必选	类型	可选值范围	说明
text	是	string	-	文本数据,UTF-8编码。最大长度512个字符(包括汉字、字母、数字和符号)

返回说明

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码,当请求错误时返回
error_msg	否	string	错误描述信息,当请求错误时返回
content	是	string	预测的文本内容 (最多仅支持512个字)
results	否	array(object)	分类结果数组
+span	否	string	抽取的具体实体内容
+offset	否	array(number)	数组由两个元素组成,分别是实体的起始位置和终止位置,从entity_content中,第一个字符记为0起算
+tag	否	string	识别对应span的实体类别

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如缺少必要出入参时返回:

```
{
"error_code": 336001,
"error_msg": "Invalid Argument"
```

}

错误码	错误信息	描述
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(868826008)或工单联系技术支持团队
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或工单联系技 术支持团队
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数代码格式是否有误。有疑问请通过QQ群(868826008)或工单联 系技术支持团队
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问请在 百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大512 UTF-8字符。有疑问请通过 QQ群(868826008)或工单联系技术支持团队
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云控制 台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请通过QQ群(868826008) 或工单联系技术支持团队
337000	Auth check failed	离线鉴权调用失败

模型更新/回滚操作说明 模型更新 1、在EasyDL-纯离线服务发布页面,找到您的服务器API发布记录,点击【更新版本】,选择「更新包」或 「完整包」来发布。 两者区别:

包类型	描述
更新包	仅包含最新的模型应用,需执行download.sh脚本下载所需镜像等依赖文件
完整包	包含模型应用和其他鉴权服务,需执行download.sh脚本下载所需完整依赖文件

2、(CPU模型可忽略)如果您训练的模型为GPU版本,系统会生成多份下载链接。请在GPU服务器执行 nvidia-smi命令,根据返回的Cuda Version来选择对应的部署包链接下载。

3、在服务器新建目录(建议标记对应模型的版本号,便于区分不同模型版本),如easydl_\${DEPLOY_NAME}_v2

以如下场景举例说明:模型版本从V1升级至V2

```
**1.新模型准备**
**创建指定版本的目录**
mkdir easedI_${DEPLOY_NAME}_v2
cd easedI_${DEPLOY_NAME}_v2
**将部署包上传至服务器该目录并解压**
tar zvxf xx.tar.gz
**解压后,进入指定目录执行download脚本下载模型所依赖文件**
cd original && bash download.sh
**2.旧模型备份**
**历史模型备份**
cp -r /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V1
**记录当前模型的端口号**
docker ps -a |grep ${DEPLOY_NAME}
**3.模型升级**
cd package/Install
**卸载当前已安装的旧的easyDL服务: ${DEPLOY_NAME},前面已备份**
python2 install.py remove ${DEPLOY_NAME}
**安装当前部署包内新的EasyDL服务: ${DEPLOY_NAME}**
python2 install.py install ${DEPLOY_NAME}
**(可选操作)更新证书**
python2 install.py lu
```

模型回滚 以如下场景举例说明:模型版本从V2回滚至V1

方法一:

```
**重命名当前v2模型目录名称**
mv /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V2
**使用V1版本**
cp -r /home/baidu/work/${DEPLOY_NAME}_V1 /home/baidu/work/${DEPLOY_NAME}
**停止当前模型容器**
docker ps -a ]grep ${DEPLOY_NAME}
docker rm -f ${容器名}
**创建新的容器**
cd /home/baidu/work/${DEPLOY_NAME} && bash start/start-1.sh
*** (可选操作)进入V1版本部署包所在目录执行license更新操作,假如部署包在/opt目录下,以您实际目录为准**
cd /opt/easydl_${DEPLOY_NAME}_V1/original/package/Install && python2 install.py lu
```

```
方法二:
```

进入模型V1所在目录,参考上述【模型更新】步骤,执行模型升级操作(即先卸载v2,后升级为v1)

短文本相似度

整体介绍

心简介

Hi,您好,欢迎使用百度EasyDL定制化训练和服务平台。

定制短文本相似度模型,是基于深度学习技术,可实现对两个文本进行相似度的比较计算。

更多详情访问: EasyDL自然语言处理方向

⊙应用场景

- 1. 搜索场景下的搜索信息匹配
- 2. 新闻媒体场景下的新闻推荐
- 3. 新闻媒体场景下的标题去重
- 4. 客户场景下的问题匹配

心 技术特色

方面的新知识,实现模型效果不断进化。

文心大模型是百度发布的产业级知识增强大模型,是千行百业AI开发的首选基座大模型。文心大模型既包含基础通用的大模型,也包含了面向重 点领域和重点任务的大模型,还提供丰富的工具与平台,支撑企业与开发者进行高效便捷的应用开发。"知识增强"是文心的核心特色,文心能够 同时从大规模知识和海量多元数据中持续学习,如同站在巨人的肩膀上,训练效率和理解准确率都得到大幅提升,并具备了更好的可解释性。

心 使用流程

训练模型的基本流程如下图所示,全程可视化简易操作,在数据已经准备好的情况下,最快15分钟即可获得定制模型。



数据准备

◎ 上传数据集

您可以上传带有标注信息的数据,和无标注信息的数据。您可以根据自己的情况,选择上传方式,目前平台提供上传方式有:

- 上传Excel文件
- 上传TXT文本
- 上传压缩包
- 通过API导入

下面分别为您介绍几种上传方式

以Excel文件上传

- Excel文件内数据格式要求为:每行是一个样本,使用第一列和第二列分别作为需要计算相似度的两个文本,第三列为相似度标签(如果导入 无标注数据,此列无数据)。第一列和第二列的文本内容的字符数建议不超过512个,超出将被截断。
- 文件类型支持xlsx格式,单次上传限制100个文件
- 请确保您上传的样本在sheet1中,且数据都在首列。注意,首行作为表头将被系统忽略

以压缩包方式上传

- 压缩包格式为.zip格式,单个压缩包限制5G以内
- 压缩包内文本文件类型为txt,每个txt每行数据格式要求为"文本内容1\t文本内容2\t标注结果\n",标注结果仅用1/0表示,1代表相似,0代表 不相似。一行表示一组数据,每个文本可以有多行短文本组数据,每组数据字符数建议不超过1024个字符(约512个汉字)

以TXT文本文件上传

- 支持文本文件类型为txt,编码仅支持UTF-8,单次上传限制100个文本文件。
- 短文本相似度的数据格式要求为"文本内容1\t文本内容2\t标注结果\n",一行表示一组数据,每组数据字符数建议不超过1024个字符(约512 个汉字),可上传多个文本文件

通过API方式导入

您可以通过API导入文档,查看上传数据的方式

ல API上传

本文档主要说明当您线下已有大量的已经完成分类整理的文本数据,如何通过调用API完成文本数据的便捷上传和管理。

EasyDL文本数据集管理API在管理不同模型数据类型之间是通用的。上传不同模型类型数据,只是在部分接口入参存在差异,使用及接口地址完全一致。

同发布模型后获取的API鉴权方式:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

数据集创建API

接口描述

该接口可用于创建数据集。

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/create

URL参数:

参数 值	
------	--

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,TEXT_MATCHING 短文本相似度
dataset_name	是	string	数据集名称,长度不超过20个utf-8字符

若上传声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
dataset_id	否	number	创建的数据集ID

查看数据集列表API

接口描述

该接口可用于查看数据集列表。返回数据集的名称、类型、状态等信息。

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/list

URL参数:

Baidu 百度智能云文档

参数 值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,短文本相似度为TEXT_MATCHING
start	否	number	起始序号,默认为0
num	否	number	数量,默认20,最多100

若查看声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明		
log_id	是	number	惟一的log id,用于问题定位		
total_num	否	number	发据集总数		
results	否	array(object)	数据集列表		
+dataset_id	否	number	数据集ID		
+dataset_name	否	string	数据集名称		
+type	否	string	数据集类型		
+status	否	string	数据集状态		
+special_status	否	string	数据集特殊状态,包括shared、smart和空值,分别表示共享中、智能标注中、非特殊状态		

查看分类 (标签)列表API

接口描述

该接口可用于查看分类(标签)。返回分类(标签)的名称、包含数据量等信息。短文本相似度标签仅存在0和1两个标签,其中0代表不相似,1 代表相似。

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/list

URL参数:

参数

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

值

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,短文本相似度为TEXT_MATCHING
dataset_id	是	number	数据集ID
start	否	number	起始序号,默认0
num	否	number	数量,默认2,最多2

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	标签总数
results	否	array(object)	标签列表
+label_id	否	string	标签/分类ID
+label_name	否	string	标签/分类名称
+entity_count	否	number	图片/声音/文本数量

添加数据API

接口描述

该接口可用于在指定数据集添加数据。

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/addentity

URL参数:

参数 值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,短文本相似度为TEXT_MATCHING
dataset_id	是	number	数据集ID
appendLabel	否	boolean	确定添加标签/分类的行为:追加(true)、替换(false)。默认为追加(true)。
entity_content	是	string	type为TEXT_MATCHING时,填入utf-8编码的文本。每行文本内容的格式为"文本内容1\t文本内容2"。 内容限制为:文本相似度中每个文本不得超过512个字(包括标点)
entity_name	是	string	文件名
labels	是	array(object)	标签/分类数据
+label_name	是	string	标签/分类名称:仅可为0或1,其中0代表不相似,1代表相似

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

数据集删除API

接口描述

该接口可用于删除数据集。

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/delete

URL参数:

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
а т	

Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,短文本相似度为TEXT_MATCHING
dataset_id	是	number	数据集ID

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

分类 (标签) 删除API

接口描述

该接口可用于删除分类(标签),短文本相似度不支持删除标签。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
  "error_code": 110,
  "error_msg": "Access token invalid or no longer valid"
}
```

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求 ,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量 阶梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计 费;未上线计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶 梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
406000	internal server error	服务错误
406001	param[xx] invalid	参数xx不合法,请检查相关参数
406002	dataset not exist	数据集不存在
406003	dataset already exists	数据集已存在
406004	dataset can not be modified temporarily	数据集暂不可修改
406005	label not exist	标签/分类不存在
406006	no permission to modify the dataset	没有修改数据集的权限
406007	dataset cannot be modified while smart annotation is running	智能标注期间不可修改数据集
406008	quota exceeded	配额超限

模型训练

の 创建模型

步骤 Step 1 创建模型

在【模型中心】或者【模型中心-我的模型】点击创建模型。

Step 2 填写基本信息、选择模型类型、提交模型名称、模型描述、联系方式即可创建模型。

Step 3 查看已创建的模型 模型创建成功后,可以在【我的模型】中看到刚刚创建的模型。

短文本相似度模型 ∢─	模型列表 > 创建模型	
品 总览		
fb 模型中心	模型类别短文本相似度	
我的模型	模型名称 *	
创建模型	模型归属 公司 个人	
训练模型		
校验模型	请输入公司名称	
发布模型	所属行业 * 请选择行业 ~	
🖉 EasyData数据服务	应田杨县 →	
数据总览	应用均录 " 用处并应用均录 *	
公开数据集	邮箱地址 * z******@baidu.com	
在线标注	联系方式 * 135*****919	0
目 公有云服务	功能描述 *	
在线服务		
	0/500	
	完成	

1.创建模型后可持续新增模型版本,因此不必每次训练模型都创建模型

2.目前单个用户在每种类型的模型下最多可创建**10个**模型,每个模型均支持多次训练,若需要创建超过10个以上的模型,请在百度云控制台 内提交工单反馈。

3.如果您是企业用户,建议您按照真实企业信息进行填写,便于EasyDL团队后续更好的为您服务

の 效果优化

通过模型迭代、检查并优化训练数据、选择高精度模型等方法,能够提升模型效果。**模型迭代**

一个模型很难一次性就训练到最佳的效果,通常会需要结合模型评估报告和校验结果不断扩充数据和调优。

为此平台提供了模型迭代功能,即当模型训练完毕后,会生成一个最新的版本号,首次V1、之后V2……以此类推。可以通过调整训练数据和算法,多次训练,以获得适合业务需求的模型效果。

如果模型已经是上线状态,依然支持模型迭代,只是需要在训练完毕后更新线上服务接口,在接口地址不变的情况下可以持续优化效果。

检查并优化训练数据

- 检查是否存在训练数据过少的情况,建议文本数量不少于1000个,如果低于这个量级建议扩充
- 通过模型效果评估报告中的详细评估指标,有针对性地扩充训练数据。例如下图评估报告中显示"0"(不相似)的精确值较低,可考虑从两方面进行优化,一是适当增加"0"不相似数据集样本量,而是检查当前数据集中"0"不相似数据集质量,是否有定义模糊的情况等,以此达到模型优化的效果。

^ナ ビ桨EasyDL 产	品介绍 操作平台 应用案例 使用文档
短文本相似度模型	我的模型 > pf短文相似度测试Vt模型评估报告
器总览 ① 模型中心 我的模型	部署方式 公有云API(目前仅 > 版本 V1 > 训练时长 24小时7分钟
创建模型	
训练模型	整体评估
发布模型	pf短又相似度测试 因训练时间过长,触发了训练时长保护机制,模型训练不完整。 如何优化效果? 准确率 ⑦ FI-score ⑦
☑ EasyData数据服务	92.7%
数据总览 公开数据集	祥细评估
在线标注	各分类的精确值 各分类的F1-Score
目 公有云服务 在线服务	0 91.7% 0 91.2%
	1 93.3% 1 93.7% 各分类的召回值 93.7% 1 93.7%
	0 90.7%
	1 94.1%

 检查测试模型的数据与训练数据的文本类型与风格是否一致,如果不一致,那么很可能会存在模型效果不错但实际测试效果较差的情况。针对 这种情况建议重新调整训练集,将训练数据与实际业务场景数据尽可能一致

选择高精度模型

在训练模型时,选择高精度的模型,将提升模型的预测准确率。

「高精度」算法内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等方 面的新知识,实现模型效果不断进化。

心 评估效果

模型评估 模型训练完成后,可以在「我的模型」列表中查看该模型的效果,以及完整评估结果。

短文本相似度模型 《三	我的模型 > text-xsy1-VI模型评估报告			
22 単変 合 模型中心 和約模型	部署方式 公有云API(目前仅 > 消防時日 47分钟	版本 VI ~		
創建模型	1 軟体運仕			
校验模型 发布模型	text-xsyl VI放果优异。 加何优化放果?	FI-score @	() 精錬率 ()	
回 EasyData数据服务 数据总算	94.2%	94.0%	93.8%	94.1%
公开数据集 在线标注 == 公布元服务	各分类的精确值	备分类的FI-Score		
在线服务	1 6 各分类的召回值	91.6% 1 92. 95.9% 0 95.	7% 1%	
	1	93.9% 94.4%		

「完整评估结果」页面中将记录整体评估与详细评估的报告,包括该模型整体的准确率、F1-score、精确率、召回率,以及评估样本具体数据情况,各分类的精确值、F1-Score等指标。

整体评估中,各指标的释义如下:

- 准确率:正确分类的样本数与总样本数之比
- F1-score:给每个类别相同的权重,计算每个类别的F1-score,然后求平均值
- 精确率:给每个类别相同的权重,计算每个类别的精确率,然后求平均值
- 召回率:给每个类别相同的权重,计算每个类别的召回率,然后求平均值

如果在训练阶段,使用的数据集中,相似或不相似的文本量在100条以内,训练出来的模型的效果评估报告的参考价值较小,建议您训练时 数据量准备充足 **模型校验**

在完成训练后,发布模型前,可以先进行模型校验,以确保模型在实际环境中能获得预期的性能。操作方法如下:

1. 在左侧「模型中心」目录中点击「校验模型」,进入校验模型页面

2. 选择需要校验能力的模型、部署方式、版本,点击「启动模型校验服务」

3. 校验服务启动后,在左侧输入两对文本,点击「校验」后,右侧识别结果栏将输出预测结果

如果在训练阶段,使用的数据集中,相似或不相似的文本量在100条以内,训练出来的模型的效果评估报告的参考价值较小,建议您训练 时数据量准备充足

の 发起训练

训练模型

完成数据的标注,或提交已标注的数据后,即可在「模型中心」目录中点击「训练模型」,开始模型的训练。

按以下步骤操作,启动模型训练:

Step 1 选择模型 选择此次训练的模型 Step 2 训练配置

部署方式

可选择「公有云部署」。

选择算法

您可以根据训练的需求,选择「高精度」或「高性能」算法。不同的算法将影响训练时间、预测速度与模型准确率。

- 如果您选择了高精度的模型,模型预测准确率将更高。如果您手中的标注数据集样本较少(例如少于1000条),可选择「高精度」的算法。
 使用高精度的算法训练模型将会耗时更久,实验环境下1000个样本,预计在20-60分钟左右完成训练
- 如果您选择了高性能的模型,相同训练数据量的情况下,训练耗时更短,模型预测速度更快。使用10000条训练样本,将在10min内完成训练.
 同样的数据量情况下,效果比高精度的模型4-5%

「高精度」算法内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义 等方面的新知识,实现模型效果不断进化

模型筛选指标

选择不同的模型选择方式,对应的模型各项效果指标将有所不同。如果没有特殊场景的要求,使用默认即可(兼顾Precision精确度和Recall召回率)。有以下指标可供选择:

- 模型兼顾Precision和Recall:挑选模型时,兼顾Precision精确度和Recall召回率,如场景中没有对精度或召回的特别要求,建议您使用此默认指标
- Precision最高的模型:挑选模型时,优先挑选Precision精度最高的模型作为部署模型
- Recall最多的模型:挑选模型时,优先选择召回率最高的模型作为部署模型
- ACC最大的模型:挑选模型时,优先挑选预测样本数最多的模型作为部署模型
- Loss最小的模型:挑选模型时,优先挑选预测偏差最小的模型

Step 3 添加数据

添加训练数据

- 可选择多个数据集
- 训练时间与数据量大小和您选择的模型类型有关,如果您选择的是高性能的模型,使用10000条训练样本将在10min内完成训练;如果您选择的是高精度的模型,使用10000条训练样本,将在20-60min完成训练

添加自定义测试集

上传不包含在训练集的测试数据,可获得更客观的模型效果评估结果。

添加自定义测试集的目的:

如果学生的期末考试是平时的练习题,那么学生可能通过记忆去解题,而不是通过学习的方法去做题,所以期末考试的试题应与平时作业不能一样,才能检验学生的学习成果。那么同理,AI模型的效果测试不能使用训练数据进行测试,应使用训练数据集外的数据测试,这样才能 真实的反映模型效果

期末考试的内容属于学期的内容,但不一定需要完全包括所学内容。同理,测试集的标签是训练集的全集或者子集即可

Step 4 训练模型

点击「开始训练」,即可开始训练模型。

- 训练时间与数据量大小、选择的算法、训练环境有关
- 模型训练过程中,可以设置训练完成的短信提醒并离开页面
- 平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

优惠政策:

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有声音分类 操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。可参考价格说明

模型部署

- 心 公有云API
- ^の发布API

发布模型,生成在线API

训练完毕后可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可发布公有云API:

- 1. 选择模型
- 2. 选择部署方式「公有云部署」
- 3. 选择版本
- 4. 自定义服务名称、接口地址后缀
- 5. 申请发布

发布模型界面示意:

发布模型	
选择模型	text-xsy1 V
部署方式	公有云部署 🗸 🗸 🗸
选择版本	V1 ~
服务名称 *	
接口地址 *	https://aip.baidubce.com/rpc/2.0/ai_custo
	m/v1/text_match/
其他要求	若接口无法满足您的需求,请描述希望解决 的问题,500汉字以内
	0/500
	提交申请
	发布模型 选择模型 部署方式 选择版本 服务名称 * 接口地址 * 其他要求

发布完成 申请发布后,通常的审核周期为T+1,即当天申请第二天可以审核完成。

如果需要加急、或者遇到莫名被拒的情况,请在百度云控制台内提交工单反馈。

の调用API

本文档主要说明定制化模型发布后获得的API如何使用,如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:868826008)联系群管

接口描述

基于自定义训练出的短文本匹配模型,实现个性化短文本相似度计算。模型训练完毕后发布可获得定制API

接口鉴权

- 1、在EasyDL——控制台创建应用
- 2、应用列表页获取AK SK

Û	Global 🗸						Q	工单	消息	帮助文档	企业组织	财务	中文	() ~
88	✓ 返回EasyDL总览	产品服务 /												
>	专业版	应用列	表											
88	公有云服务管理 ^	+ 0(3	建成用											
₫	• 应用列表		应用名称	AppID	API Key	Secret Key			创建的	时间	操	作		
ē	 权限管理 用量统计 	1	ff_easydl_test1	148266060 2	AzSZnqX2DXGv0pxX1qZZE5X m	****** 显示			2019	-10-21 16:20:	03 👭	表 管理	删除	
Ħ	 监控报表 	2	test_double_app	148265769 7	sif9YPOK6Uf4f4zIHddxwGxp	•••••• 显示			2019	-08-23 15:17	15 报	表 管理	删除	
55 E1	 技术文档 本地部署包管理 	3	wahaha	148265769 6	WriypRCqjKHmKYW4uWXqfa20	****** 显示			2019	-08-23 15:16	:01 👭	表 管理	删除	
Þ	设备端服务管理	4	hjpj_test_zbs	148265767 3	NdvOkPPI949FoiBrW6eOoDL1	•••••• 显示			2019	-08-22 18:18	47 报	表 管理	删除	
44		5	hjpj_test_zbs	148265767 2	GkMah53SGTAUjLS2Tb7bRa0A	•••••• 显示			2019	-08-22 18:18	32 #	表 管理	删除	
		6	hjpj_test_zb	148265767 1	scko2V4AH1Iya5wjS4sD5tXS	****** 显示			2019	-08-22 18:16	52 👭	表 管理	删除	
		7	hjpj_test_zb	148265767 0	Gz3LQLGKAvI3DSTRxeMXY1UL	•••••• 显示			2019	-08-22 18:16	37 措	表 管理	删除	
		8	hjpj_test	148265766 9	jNrZSjIDa5y2OAt2MsLYX0B9	****** 显示			2019	-08-22 18:16	06 👭	表 管理	删除	6
		9	hjpj_test	148265766 8	KHbADxabR9zr0GiTwKK6IFLK	••••• 显示			2019	-08-22 18:15	51 报	表 管理	删除	

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

参数	值		

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,需以json方式请求。

Body请求示例:

```
{
"text_a": "<UTF-8编码数据>",
"text_b": "<UTF-8编码数据>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text_a	是	string	-	文本数据,UTF-8编码。最大长度512个字符(包括汉字、字母、数字和符号)
text_b	是	string	-	文本数据,UTF-8编码。最大长度512个字符(包括汉字、字母、数字和符号)

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
score	是	number	文本相似度,值的范围[0,1],相似度递增

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
   "error_code": 110,
   "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

Baidu 百度智能云文档

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(649285136)或工单联系技术支持团 队。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请通过QQ群(649285136)或工单联系技术支持团 队。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请通过QQ群(649285136)联系群管手动提额
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线计 费的接口,请通过QQ群(649285136)联系群管手动提额
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上 线计费的接口,请通过QQ群(649285136)联系群管手动提额
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(868826008)或工单联系技术支持团 队
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本编码错误等等,可检查下文本编码、代码格式是否有误。有疑问请通 过QQ群(868826008)或工单联系技术支持团队
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或者代码格式有误。有疑问请通过QQ群(868826008)或工单 联系技术支持团队
336003	Base64解码失败	文本格式有误。有疑问请通过QQ群(868826008)或工单联系技术支持团队
336004	输入本文大小不合法	文本超出大小限制,每个文本限制512个字符(包括汉字、字符、数字或字母),有疑问请通过QQ群 (868826008)或工单联系技术支持团队
336005	文本解码失败	文本编码错误,请检查并修改文本格式
336006	缺失必要参数	text字段缺失(未上传文本)
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请通过QQ群 (868826008)或工单联系技术支持团队

の调用API

本文档主要说明定制化模型发布后获得的API如何使用,如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:868826008)联系群管

接口描述

基于自定义训练出的短文本匹配模型,实现个性化短文本相似度计算。模型训练完毕后发布可获得定制API

接口鉴权

- 1、在EasyDL——控制台创建应用
- 2、应用列表页获取AK SK

)	Global V						Q	工单	消息	帮助文档	企业组织	财务	中文	()
8	< 返回EasyDL总览	产品服务 /	EasyDL定制训练平台 - 应用列表											
•	专业版	应用列	列表											
3	公有云服务管理 ^	+ 0	1建应用											
9	• 应用列表	-	应用名称	AppID	API Key	Secret Key			创建	时间	採	作		
) D	 权限管理 用量统计 	1	ff_easydl_test1	148266060 2	AzSZnqX2DXGv0pxX1qZZE5X m	****** 显示			2019	-10-21 16:20:	03 报	表 管理	删除	
6	• 监控报表	2	test_double_app	148265769 7	sif9YPOK6Uf4f4zIHddxwGxp	•••••• 显示			2019	-08-23 15:17	15 报	表 管理	删除	
5	 技术文档 本地部署包管理 	3	wahaha	148265769 6	WriypRCqjKHmKYW4uWXqfa20	****** 显示			2019	-08-23 15:16	01 报	表 管理	删除	
]	设备端服务管理	4	hjpj_test_zbs	148265767 3	NdvOkPPI949FoiBrW6eOoDL1	•••••• 显示			2019	-08-22 18:18	47 18	表 管理	删除	
5		5	hjpj_test_zbs	148265767 2	GkMah53SGTAUjLS2Tb7bRa0A	•••••• 显示			2019	-08-22 18:18	32 报	表 管理	HIRR	
		6	hjpj_test_zb	148265767 1	scko2V4AH1Iya5wjS4sD5tXS	•••••• 显示			2019	-08-22 18:16	52 报	表 管理	删除	
		7	hjpj_test_zb	148265767 0	Gz3LQLGKAvI3DSTRxeMXY1UL	•••••• 显示			2019	-08-22 18:16	37 措	表 管理	删除	
		8	hjpj_test	148265766 9	jNrZSjiDa5y2OAt2MsLYX0B9	****** 显示			2019	-08-22 18:16	06 18	表 管理	删除	6
		9	hjpj_test	148265766 8	KHbADxabR9zr0GiTwKK6IFLK	••••• 显示			2019	-08-22 18:15	51 报	表 管理	删除	

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

```
参数 值
access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"
```

Header如下:

参数 值 Content-Type application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,需以json方式请求。

Body请求示例:

```
{
"text_a": "<UTF-8编码数据>",
"text_b": "<UTF-8编码数据>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text_a	是	string	-	文本数据,UTF-8编码。最大长度512个字符(包括汉字、字母、数字和符号)
text_b	是	string	-	文本数据,UTF-8编码。最大长度512个字符(包括汉字、字母、数字和符号)

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
score	是	number	文本相似度,值的范围[0,1],相似度递增

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
   "error_code": 110,
   "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

Baidu 百度智能云文档

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(649285136)或工单联系技术支持团 队。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请通过QQ群(649285136)或工单联系技术支持团 队。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请通过QQ群(649285136)联系群管手动提额
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线计费的接口,请通过QQ群(649285136)联系群管手动提额
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上 线计费的接口,请通过QQ群(649285136)联系群管手动提额
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(868826008)或工单联系技术支持团 队
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本编码错误等等,可检查下文本编码、代码格式是否有误。有疑问请通 过QQ群(868826008)或工单联系技术支持团队
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或者代码格式有误。有疑问请通过QQ群(868826008)或工单 联系技术支持团队
336003	Base64解码失败	文本格式有误。有疑问请通过QQ群(868826008)或工单联系技术支持团队
336004	输入本文大小不合法	文本超出大小限制,每个文本限制512个字符(包括汉字、字符、数字或字母),有疑问请通过QQ群 (868826008)或工单联系技术支持团队
336005	文本解码失败	文本编码错误,请检查并修改文本格式
336006	缺失必要参数	text字段缺失(未上传文本)
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请通过QQ群 (868826008)或工单联系技术支持团队

心 发布私有部署服务器

⊙ 发布私有部署服务器

训练完毕后可以在左侧导航栏中找到【发布模型】,依次进行以下操作即可将模型部署到私有服务器:

模型中心	发布模型			
我的模型	选择模型:	老市民	~	说明
创建模型	部署方式:	私有服务器部署	~	私有服务器邮署支持将模型邮署于本地的CPU、GPU服务器上,提供API和SDK两 种集成方式:查看文档
训练模型	集成方式:	私有API 图 服务器端	SDK	私有API: 將模型以Docker形式在本地服务器(仅支持Linux)上部署为http服
校验模型	选择版本:	V1	~	务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷
发布模型				服务器端SDK:将模型封装成适配本地服务器(支持Linux和Windows)的SDK, 可集成在其他程序中运行。首次联网激活后即可纯离线运行,占用服务器资源更
数据中心		提交申请		少,使用方法更灵活
我的数据集				
云服务调用数据				

- 选择模型
- 选择部署方式「私有服务器部署」
- 选择集成方式「私有API」(当前仅支持此方式)

将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷

点击「提交申请」后,前往控制台申请私有部署包。并参考文档完成集成

✓返回EasyDL总览	产品服务 / EasyDL定制训	练平台 - 私有服务器部署管理						
经典版	私有服务器部署管	理						
公有云服务管理 >	私有API集成	服务器端SDK集成						
私有服务器部署管理								
设备谐服务管理	使用说明: (展开1	查看部署介绍及操作步骤等)						
数据管理服务	+ 申请私有部署包					请输入私有服务名称	R Q	开发文档
	申请ID	私有服务名称	模型ID	模型版本	申请日期	证书到期日期	状态	操作
	1008	test1	17282	V5	2019-09-12	2019-10-05	 审核通过 	查看申请信 下载更新包 更新版本
	1000	new	35261	V1	2019-09-09	2019-10-09	● 流程关闭,重新申请	查看申请信
	989	test1	17282	V5	2019-09-05	2019-10-05	● 审核中	查看申请信
	988	test1	17282	V5	2019-09-05	2019-10-05	● 审核通过	查看申请信 下载完整部 更新版本
	987	test	6793	V2	2019-09-05	2019-10-05	● 审核中	查看申请信
							每页显示	10 🗸 <

私有服务器部署价格说明

EasyDL经典版已支持将定制模型部署在私有服务器上,只需在发布模型时提交私有服务器部署申请,通过审核后即可获得**一个月免费试用**。如需 购买**永久使用授权**,请微信搜索"BaiduEasyDL"添加小助手咨询,通过线下签订合同购买使用。

心 私有部署服务API调用说明文档

本文档主要说明定制化模型本地部署后,如何使用本地API。如还未训练模型,请先前往EasyDL进行训练。

如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:868826008)联系群管

部署包使用说明

部署方法

EasyDL定制化短文本相似度模型的本地部署通过EasyPack实现,目前提供单机一键部署的方式。

在EasyDL申请、下载部署包后,在本地服务器新建目录(建议目录命名规则:easyDL_服务名称_模型版本号),将软件包上传至该目录。请参 考EasyPack-单机一键部署使用python2版本来部署,部署成功后,启动服务,即可调用与在线API功能类似的接口。

运维检查

EasyDL服务器API部署应用健康检查(或故障排查)脚本:trouble_shooting.tar

脚本能力:鉴权服务健康检测、容器状态检查、端口探活、网络联通性测试、容器关键报错日志输出等

使用方法: 将脚本上传至服务器任意目录(或在服务器直接下载),并解压后运行。

```
**解压**
tar vxf trouble_shooting.tar
**执行**
bash trouble_shooting.sh
```

授权说明

本地部署包根据服务器硬件(CPU单机或GPU单卡)进行授权,只能在申请时提交的硬件指纹所属的硬件上使用。

部署包测试期为1个月,如需购买永久授权,可提交工单咨询

API参考

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先进行自定义模型训练,完成训练后申请本地部署,本地部署成功后拼接url。

请求URL:http://{IP}:{PORT}/{DEPLOY_NAME}/TextMatching

- IP:服务本地部署所在机器的ip地址
- PORT:服务部署后获取的端口
- DEPLOY_NAME:申请时填写的本地服务名称

Header如下:

```
参数 值
Content-Type application/json
```

Body请求示例:

```
{
"text_a": "<UTF-8编码数据>",
"text_b": "<UTF-8编码数据>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text_a	是	string	-	文本数据,UTF-8编码。最大长度512个字符(包括汉字、字母、数字和符号)
text_b	是	string	-	文本数据,UTF-8编码。最大长度512个字符(包括汉字、字母、数字和符号)

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
score	是	number	文本相似度,从0-1,相似度递增

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如缺少必要出入参时返回:

	{
	"error_code": 336001,
	"error_msg": "Invalid Argument"
}	

错误码	错误信息	描述
336000	Internal error	服务器内部错误,请再次请求 , 如果持续出现此类错误,请通过QQ群 (868826008) 或工单联系技术支持团队
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或工单联系技 术支持团队
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数代码格式是否有误。有疑问请通过QQ群(868826008)或工单联系 技术支持团队
336003	Base64解码失败	图片/音频/文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有 疑问请通过QQ群(868826008)或工单联系技术支持团队
336004	输入文件大小不合 法	文本超出大小限制,文本限4M以内,请根据接口文档检查入参格式,有疑问请通过QQ群(868826008)或工单 联系技术支持团队
336005	文本解码失败	文本编码错误,请检查并修改文本格式
336006	缺失必要参数	文本字段内容缺失
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请通过QQ群(868826008) 或工单联系技术支持团队
337000	Auth check failed	离线鉴权调用失败

模型更新/回滚操作说明 模型更新 1、在EasyDL-纯离线服务发布页面,找到您的服务器API发布记录,点击【更新版本】,选择「更新包」或 「完整包」来发布。 两者区别:

包类型	描述
更新包	仅包含最新的模型应用,需执行download.sh脚本下载所需镜像等依赖文件
完整包	包含模型应用和其他鉴权服务,需执行download.sh脚本下载所需完整依赖文件

2、(CPU模型可忽略)如果您训练的模型为GPU版本,系统会生成多份下载链接。请在GPU服务器执行 nvidia-smi命令,根据返回的Cuda Version来选择对应的部署包链接下载。

3、在服务器新建目录(建议标记对应模型的版本号,便于区分不同模型版本),如easydl_\${DEPLOY_NAME}_v2

\${DEPLOY_NAME}:申请时填写的服务名称

以如下场景举例说明:模型版本从V1升级至V2

```
**1.新模型准备**
**创建指定版本的目录**
mkdir easedl_${DEPLOY_NAME}_v2
cd easedI_${DEPLOY_NAME}_v2
**将部署包上传至服务器该目录并解压**
tar zvxf xx.tar.gz
**解压后,进入指定目录执行download脚本下载模型所依赖文件**
cd original && bash download.sh
**2.旧模型备份**
**历史模型备份**
cp -r /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V1
**记录当前模型的端口号**
docker ps -a |grep ${DEPLOY_NAME}
**3.模型升级**
cd package/Install
**卸载当前已安装的旧的easyDL服务: ${DEPLOY_NAME},前面已备份**
python2 install.py remove ${DEPLOY_NAME}
**安装当前部署包内新的EasyDL服务: ${DEPLOY_NAME}**
python2 install.py install ${DEPLOY_NAME}
**(可选操作)更新证书**
python2 install.py lu
```

模型回滚 以如下场景举例说明:模型版本从V2回滚至V1

方法一:

```
**重命名当前v2模型目录名称**
mv /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V2
**使用V1版本**
cp -r /home/baidu/work/${DEPLOY_NAME}_V1 /home/baidu/work/${DEPLOY_NAME}
**停止当前模型容器**
docker ps -a |grep ${DEPLOY_NAME}
docker ms -f ${容器名}
**创建新的容器**
cd /home/baidu/work/${DEPLOY_NAME} && bash start/start-1.sh
** (可选操作)进入V1版本部署包所在目录执行license更新操作,假如部署包在/opt目录下,以您实际目录为准**
cd /opt/easydl_${DEPLOY_NAME}_V1/original/package/Install && python2 install.py lu
```

方法二:

进入模型V1所在目录,参考上述【模型更新】步骤,执行模型升级操作(即先卸载v2,后升级为v1)

评论观点抽取

整体介绍

心简介

Hi,您好,欢迎使用百度EasyDL定制化训练和服务平台。

定制评论观点抽取的模型,是基于自建抽取体系的机器学习方法,可实现从评论文本中抽取评价片段、评价维度、评价观点、并判断评论的情感倾向。

更多详情访问:EasyDL自然语言处理方向

⊙应用场景

评论数据中隐藏着大量的商业价值:

- 新品用户分析:商家很难单纯从新产品消费额、出货量等结构化数据中找到是否满足消费者需求的答案,可通过新产品的用户评论,完成用户 分析。
- 挖掘产品问题:挖掘出有价值的信息,比如产品的问题、用户的偏好、竞品的差异性,甚至洞察出新的商业机会。
- 辅助消费决策:通过对比同一类型产品不同商品或商家的评论观点信息,可以辅助用户进行消费决策。
- 互联网舆情分析:商家可通过对评论及其情感倾向的分析,监控品牌和商品的舆情信息变化。

∞ 技术特色

评论观点抽取模型内置<u>文心大模型</u>,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等 方面的新知识,实现模型效果不断进化。

文心大模型是百度发布的产业级知识增强大模型,是千行百业AI开发的首选基座大模型。文心大模型既包含基础通用的大模型,也包含了面向重 点领域和重点任务的大模型,还提供丰富的工具与平台,支撑企业与开发者进行高效便捷的应用开发。"知识增强"是文心的核心特色,文心能够 同时从大规模知识和海量多元数据中持续学习,如同站在巨人的肩膀上,训练效率和理解准确率都得到大幅提升,并具备了更好的可解释性。

```
っ 使用流程
```

训练模型的基本流程如下图所示,全程可视化简易操作,在数据已经准备好的情况下,最快15分钟即可获得定制模型。



数据准备

☑ 创建数据集并导入

1.创建数据集

1. 选择【EasyDate数据服务】目录下数据总览,点击"创建数据集"。

ビ桨EasyDL ご浆EasyDL	产品介绍	3 操作平台	应用案例	使用文档		
评论观点抽取	Æ	我的数据总览				
88 总览						
面 模型中心		EasyData智能数	据服务平台已上线	,使用EasyData可享受包括多人标注。	、数据清洗、数据采集等完整数据服务	立即前往
我的模型		创建数据集				
创建模型						
训练模型		酒店评论观点抽耳	2 [公 数据集组]	D: 310391		
校验模型		版本	数据集ID	数据量	最近导入状态	标注类型
发布模型		100	225087	22		2022년 10년 11년 11년 11년 11년 11년 11년 11년 11년 11
☑ EasyData数据服务		VI 💬	335087	22	■ 匚元风	许论观点拙取
数据总览		八百法古法中代	-/	200020		
在线标注		公开闾佔数据集	<u>小</u> 奴姑果珇ID.	203020		

3. 输入数据集名称,选择数据集属性:是否对数据进行去重操作,详细方法见数据去重策略。

4. 点击完成,在数据总览目录下可以看到生成一个空数据集项目。

评论观点抽取 🛛 🚽 🖂	我的数据总	览								提交工单
 器 总览 凸 模型中心 我的模型 	EasyData 创建数据	ə智能数据服务平台已 集	上线,使用Easy	Data可享受包括多人标注、	数据清洗、数据采集等完整数	据服务 立即前往			输入数据集合	S称或ID Q
创建模型 训练模型	111111 🖂	数据集组ID: 8944	1					□ 新増版本	日日 全部版本	白 删除
校验模型	版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作		
发布模型 図 EasyData数据服务	V1 Θ	11618	0	• 已完成	评论观点抽取	0% (0/0)	-	导入 删除		
数据总览	comm-n	区 数据集组ID:8	943					口新增版本	铝 全部版本	回 删除
在线标注 公有云服务	版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作		
在线服务	∨4 ⊙	11617	11	• 已完成	评论观点抽取	0% (0/11)	-	查看 导入 标	主 清洗 ・・・	
43 EasyEoge本地部會 纯离线服务	V3 💬	11616	9	• 已完成	评论观点抽取	44% (4/9)	-	查看 导入 标	主 删除	
	qa-测试参	勿删0307 🖂 数据1	戦组ID: 8942					口新编版本	田 全部版本	白 删除
	版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作		
	∨1 ⊙	11610	10	• 已完成	评论观点抽取	0% (0/10)	-	查看导入标	主清洗 …	
	ssssassa	区 数据集组ID:8	3940					□ 新増版本	铝 全部版本	前 删除
	版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作		

2.导入未标注文本数据

点击【导入】进入到新创建的评论观点抽取数据集中,平台暂只支持上传无标注信息的数据。

<i>- ビ</i> 桨EasyDL	产品介绍	操作平台	应用案例	使用文档	
评论观点抽取	Ē	我的数据总览	> 酒店评论观点抽取八	/1/导入	
88 总览					
曲 模型中心		┃创建信息 >			
我的模型		数据集ID	335087	版本号	V1
创建模型		备注			
训练模型		┃标注信息 >			
校验模型		标注类型	评论观点抽取	标注模板	评论观点抽取
发布模型		数据总量	22	已标注	2(进度9.09%)
@ EasyData数据服务		大小	0.01M		
数据总览		数据清洗			
在线标注		暂未做过数据》	清洗任务		
目 公有云服务		导入数据			
在线服务		数据标注状态	○ 无标注信息	ļ,	
記 EasyEdge本地部署		导入方式	请选择		^
纯离线服务			本地导入		
			BOS目录导)	\	
			分享链接导入		
			半台匕有数据	5来	

导入方式,分别为「本地导入」、「BOS目录导入」、「分享链接导入」、「平台已有数据集」;

通过本地导入时,可通过excel文件、TXT文件、压缩包形式上传

• 通过文本上传时,需注意:

EasyDL	产品介	1绍 操作平台	应用案例	使用文档				社区	飞桨官网	AI主站	
评论观点抽取	ŧ	我的数据总览	> 111111/V1/导入								
88 总宽											
由 模型中心		┃创建信息 ∨									
我的模型		数据集ID	11618	版本号	V1						
创建模型		备注									
训练模型		┃标注信息 〜									
校验模型		标注类型	评论观点抽取	标注模板	评论观点	抽取					
发布模型		数据总量	0	已标注	0						
☑ EasyData数据服务		大小	OM								
数据总览		数据清洗									
在线标注		智禾做过数据	清洗仕务								
目 公有云服务		导入数据									
在线服务		数据标注状态	○ 无标注信	息							
習 EasyEdge本地部署		导入方式	本地导入		~	请选择 <					
纯离线服务			确认并返	Ð		上传Excel文件					
						上传TXT文本					
						上传压缩包					

文本文件内数据格式要求为"文本内容\n"(即每行一个样本,使用回车换行),每一行表示一组数据,每组数据的数建议不超过512个字符,超 出将被截断

文本文件类型支持txt,编码仅支持UTF-8,单次上传限制100个文本文件,最多可上传100万个文本文件。

• 通过压缩包上传时,需注意:

压缩包内的一个文本文件将作为一个样本上传。压缩包格式为.zip格式,压缩包内文件类型支持txt,编码仅支持UTF-8。

每组数据的字数建议不超过512个字符,超出将被截断。

• 通过Excel文件上传时,需注意:

Excel文件内首行为表头,每行为一个样本,每个样本字符数不得超过512个字符,超出将被截断

文件格式支持xlsx格式,单次可上传100个文件。

通过BOS目录导入格式要求

请确保将全部文本已通过txt文件保存至同一层文件目录,该层目录下子文件目录及非相关内容(包括压缩包格式等)不导入

分享链接导入格式要求

请确保将全部文本文件保存至同一压缩包,压缩包仅支持zip格式,压缩前源文件大小限制5G以内;仅支持来自百度BOS、阿里OSS、腾讯COS、 华为OBS的共享链接

通过平台已有数据集导入

直接点选您需要的数据集即可导入。

其他:暂不支持API接口上传服务

心 数据去重策略

重复样本的定义

重复样本的定义,是指您上传的数据中,存在两个样本的文本内容完全一致。则被判定为两个样本是重复样本。例如:

文本内容

理发师的手艺真不错

理发师的手艺真不错

平台去重策略

平台提供了可去重的数据集,即对您上传的数据进行重复样本的去重。注意:当您确定了数据集为去重或非去重的属性后,便不可修改。

当您创建了一个去重的数据集时,在后续上传数据的过程中,平台可通过检验您当前上传的样本与已上传到此数据集下的样本是否相同,如果相同,则会使用新的样本替代旧的样本。

通过平台导入「无标注信息」的数据集后,可对无标注数据进行标注操作。

EasyDL	产品介绍	召 操作平	台 应用案例	使用文档				社区	飞桨官网	AI主站	😝 百里幕蕊
评论观点抽取	÷	我的数据总测	5								提交工单
Ⅲ 总览											
凸 模型中心		EasyData	智能数据服务平台已。	上线,使用Easy[Data可享受包括多人标注、数	y据清洗、数据采集等完整数	据服务 立即前往				
我的模型		创建数据集	R.							输入数据	集名称或ID Q
创建模型											
训练模型		111111	数据集组ID: 8944						口新增度	本 铝全部港	本 自删除
校验模型		版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作		
发布模型 図 EasyData数据服务		V1 ⊡	11618	10	• 已完成	评论观点抽取	0% (0/10)	-	查看 导入	标注 清洗	
数据总览											
在线标注		comm-n	☑ 数据集组ⅠD:89	943					□2 新增度	本 铝全部槽	本自删除
自 公有云服务		版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作		
在线服务		V4 ☉	11617	11	• 已完成	评论观点抽取	0% (0/11)	-	查看 导入	标注 清洗	
認 EasyEdge本地部署 纯离线服务		V3 💬	11616	9	• 已完成	评论观点抽取	44% (4/9)	-	查看 导入	标注删除	
		qa—测试勿	1删0307 🖸 数据集	[組ID: 8942					口新增数	本 铝全部槽	本 宜删除
		版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作		
		vi Θ	11610	10	● 已完成	评论观点抽取	0% (0/10)	-	查看 导入	标注 清洗	
		ssssassa	☑ 数据集组Ⅳ:8	940					口新增度	本 铝全部港	本 白删除



评价对象 (必填)

画选评价片段后,会出现弹窗如下,请激活图中按钮划选评价对象或手动输入评价对象。

这款斜挎包	回的整体质感太好了,配备	色非常漂声	则百搭。即用日日日
* 评价对象	请输入或划取片段中的评价对象	т] 🖂 🖂	DIRITER, RE
自动预标	自动标出评价维度、评价观点、情	重倾向	·····································
评价维度	请划取片段中的评价维度	[T]	
评价观点	请划取片段中的评价观点	53	
情感倾向	🔾 正向 🔹 负向		
	确认	取消	

• 自动预标注可在您输入评价对象之后,自动为您标注该评价片段中针对该评价对象的评价维度、评价观点,以及标注情感倾向。

```
在线标注 > 111111 > 标注
```

全部(10) 无标注信息(0) 有标注信息(10)		简约模式 🕐 💽 🔬 标注说明	
定位文本内容	□ 标为负例(②	
这款斜挎包的整体质感太好 斜挎包质感好正向 了,配色非常漂亮 斜挎包配色-漂亮正向 ,款式特别百搭 斜挎包-数式-特别百搭正向 。斜挎包链条的五金用料很好,包身面料的做工特别上档次。斜挎角度的设计超楼,背起来非常轻松。			
	* 评价对象 斜挎包五金 ~~	11 D	
	自动预标 🔺 自动标出评价维度、评价观点、情感候	þ	
	以下三个字段,如无可标注信息,可直接保存	_	
	评价维度 用料	T. D	
	- 起始位置 + 结尾位置 +	-	
	请划取片段中的评价维度	T ii	
	评价观点好	92 D	
	— 起始位置 十 — 结尾位置 十		
	情感倾向正向负向		
	确认	取消	

- 预标注完成后您也可以激活划选按钮对预标注的评论维度进行增加、删除、修改等操作,评价观点暂不支持多个。
- 如不使用自动预标功能,您也可以激活划选按钮自行选择文字内容。
- 对于评价维度、评价观点,您还可以点击加减号左右移动选择文本范围。

```
在线标注 > 111111 > 标注
```

全部(10) 无标注信息(0) 有标注信息(10)		简约模式 ⑦ 💽 🔬 标注说明	
定位文本内容 Q	□ 标为负例 ⑦	回 删除文本 上一篇 下一篇	
这款斜挎包的整体质感太好 斜挎包-质感-好-正	向了,配色非常漂亮 斜挎包配色-漂亮正向,款式	特别百搭	
斜挎包款式特别百搭正向 。斜挎包链条的五金用料很好,包身面料的做工特别上档次。斜挎角度的设计超棒,背起来非常轻松。 			
	* 评价对象 斜持包五金 V	に 面	
	自动预标 A 自动标出评价维度、评价观点、情感倾向		
	以下三个字段,如无可标注信息,可直接保存		
	评价维度 用料	17 D	
	一 起始位置 + - 结尾位置 +		
	请划取片段中的评价维度	行 前	
	评价观点 好	22 亩	
	- 起始位置 十 - 结尾位置 十		
	情感倾向 正向 负向		
	确认即	2消	

4. 完成标注 完成标注后效果如图:
| 评论观点抽取 🛛 🖃 | 在线标注 > 111111 > 标注 | 111111 / VI | \sim |
|---------------------------------------|--|--------------|--------|
| 88 总览 | 全部(10) 无标注信息(10) 有标注信息(0) 荷约模式 ⑦ 示标注信息(0) | 评价对象列表 | |
| 由 模型中心 | | 第 《秋章 | ~ |
| 我的模型 | 定位文本内容 Q □ 标为负例 ⑦ □ 删除文本 上一篇 下一篇 | 正向 1 | |
| 创建模型 | | 评价片段1 | |
| 训练模型 | 影响的外观磁频,优发效果和观影体磁都个错。影响一环境-磁页一上问, 交通非常方便,地图和地下的停车场车位都推扰走的,激秀人民也都
 | 评价维度 环境 | |
| 校验模型 | 102/02101494.o | 评价观点 超赞 | |
| 友布検型 | | 1689月9 止回 | |
| EasyDatagedallers | | | |
| 友经标注 | | L | |
| 日 公有云服务 | | | |
| 在线服务 | | | |
| 謵 EasyEdge本地部署 | | | |
| 纯离线服务 | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

确认标注无误,可点击下一篇(即为保存)。

5.标为负例 若您认为此篇文本无任何评论观点,则可支持选择【标为负例】,此类样本将不会参与训练,后续模型遇到此类样本也无需预测。

6.简约模式开启简约模式后,会隐藏掉详细的标注结果(评价对象-评价维度-评价观点·情感倾向),方便您阅读文本内容,如图为简约模式与普通模式对比:

全部(10) 无标注信息(3) 有标注信息(7)		简约模式 🕐 开 🔵	标注说明
定位文本内容Q	🗌 标为负例 곗 🖻	『删除文本 上一篇	下一篇
全屋定制这个价格有点贵了 片段1 ,从网上无意间发现然后联系到最后的安装: 好,沟通也很到位,唯一遗憾的是货品在安装过程中有损坏,来回换货又经历了 确实有些异味,能接受,所有新柜子都会有味的,放一下就好啦~总体感觉还不管	完毕,经历了近两个多月的时间 近一个月的时间,不清楚板材; 昔,设计测量的尺寸很完美。	司,整体感觉人员服务意访 是否为进口的板材,但刚:	R比较 安装后
全部(10) 无标注信息(3) 有标注信息(7)	(简约模式 🕐 🔵 🔬	标注说明
定位文本内容 Q	🗌 标为负例 ⊘ 🛛	回 删除文本 上一篇	下一篇
全屋定制这个价格有点费了 全屋定制价格贵-负向,,从网上无意间发现然和 人员服务意识比较好,沟通也很到位,唯一遗憾的是货品在安装过程中有损坏, 板材,但刚安装后确实有些异味,能接受,所有新柜子都会有味的,放一下就如	后联系到最后的安装完毕,经历 来回换货又经历了近一个月的 子啦∼总体感觉还不错,设计测扩	5了近两个多月的时间,整 1时间,不清楚板材是否为 量的尺寸很完美。	体感觉 进口的
模型训练			

の创建模型

步骤

- 1. 在【模型中心-我的模型】点击创建模型或直接进入到【模型中心-创建模型】;
- 2. 选择模型类型、提交模型名称、模型描述、联系方式即可创建模型。
- 3. 模型创建成功后,可以在【我的模型】中看到刚刚创建的模型。

->EasyDL	产品)	介绍 操作平台	应用案例	使用文档				社区	飞桨官网	AI主站	I	百里慕蕊
评论观点抽取	Ē	模型列表										提交工具
题总 题												
曲 模型中心		创建模型										
我的模型		CORPORT AS FEMALE									2111.04*	107 2019 0
创建模型		L PT IS AUAR MEAN	1 计化观点信收例	R田−8 区 18日10: 19449							10115125	
训练模型				模型创建成功。	若无数据集请先在"数据中心"俞	建 、上传训练数据 训练 模型后	可以在此处查看模	型的最新問	本			
校验模型												
发布模型												

1.创建模型后可持续新增模型版本,因此不必每次训练模型都创建模型

2.目前单个用户在每种类型的模型下最多可创建**10个**模型,每个模型均支持多次训练,若需要创建超过10个以上的模型,请在百度云控制台 内提交工单反馈。

3.如果您是企业用户,建议您按照真实企业信息进行填写,便于EasyDL团队后续更好的为您服务

∞ 效果评估

模型评估报告

训练完成后,可以在【我的模型】列表中看到模型效果,以及详细的模型评估报告。

「完整评估结果」页面中将记录整体评估与详细评估的报告,包括该模型整体的准确率、F1-score、精确率、召回率,以及评估样本具体数据情况,各分类的精确值、F1-Score、召回值等指标。

整体评估中,各指标的释义如下:

- F1-score:给每个类别相同的权重,计算每个类别的F1-score,然后求平均值
- 精确率:给每个类别相同的权重,计算每个类别的精确率,然后求平均值
- 召回率:给每个类别相同的权重,计算每个类别的召回率,然后求平均值

【评论观点抽取】测试v100 ^[2] 模型ID: 15444 日 训练 ^[2] 历史版本 回 删除								
部署方式	版本	训练状态	服务状态	模型效果	操作			
公有云API	V2	● 训练完成	已发布	召回率:52.94% 精确度:71.08% F1-score:0.599 完整评估结果	查看版本配置	服务详情 校验		
评论观点抽取 🛛 🕀	我的模型 > 测词	v100V2模型评估报告						
田 总览 品 模型中心	部署方式 公	r有云API 〜	版本 V2	~				
我的模型	文本数 14	标注数 90	训练时长 9分4	¢				
创建模型	軟体運住							
校验模型	测试v100 V2整体	效果欠佳。由于目前训练集数;	8量较少,该结论仅供参考	5,建议扩充训练集得到更准确的评估	效果。			
发布模型								
☑ EasyData数据服务	各评价字段指	标						
数据总览	情感傾向							
在线标注	F1-score)		○ 精确率				
目 公有云服务 方任服务	100.0)%		100.0%		100.0%		
記 EasyEdge本地部署	评价片段							
纯离线服务	F1-score	3		精确率		日回率		
	36.49	%		50.0%		28.6%		
	评价维度							
	57.1%	• ⁄o		〇 ^{精确率} 66.7%		O ^{召回率} 50.0%		
	评价观点词							
	F1-score 54.59	, %		〇 ^{精确率} 75.0%		日 日 日 日 単 42.9%		

实际效果可以在左侧目录中找到【模型校验】功能进行校验,或者发布为接口后测试。

模型校验操作步骤:

上传文本/自行输入文本--输入评价对象和勾选抽取内容--校验按钮--右侧返回json结果。

1. 在上传文本时,只支持上传格式为txt,文本文件内数据格式要求为"文本内容 \n"(即每行一个样本,使用回车换行),每一行表示一组数

据,每组数据的数建议不超过512个字符,超出将被截断。

- 2. 您可以手动输入您想要抽取的评价对象,也支持不输入评价对象进行抽取。
- 3. 选择您想要抽取的评价字段(希望抽取评价片段、评价维度、评价观点,还是判断评论情感倾向,此处只支持单选)
- 4. 校验过程需要等待数秒,返回结果会在右侧以json形式显示出来,支持下载查看json文档

模型校验示意图:

评论观点抽取 🛛 🕣	校验模型	提交工单
田 总览 凸 模型中心 我的模型	透择模型 選ばv100 > 部署方式 公布云API > 选择版本 V2 > 当期模型構确率 71.0% 评估报告 辺刻Laon档果	
创建模型	请输入校验的文本,或 <u>点击上传文本</u> 支持文本格式:txt,文本长度上限为512汉字(字符)	山 下载json文件
训练模型 校验模型	这家理发店的服务一点都不好 [{	
发布模型	prob: 0.98446452817645 text: 负问	
☑ EasyData数据服务)	
数据总览		
在线标注		
自 公有云服务		
在线服务		
II EasyEdge本地部署	13/512	
纯离线服务	评价对象 理发店 评价片段 评价维度 评价观点词 • 情感频向	
	· 徐敏	

心 优化效果

通过模型迭代、检查并优化训练数据、选择高精度模型等方法,能够提升模型效果。**模型迭代**

一个模型很难一次性就训练到最佳的效果,通常会需要结合模型评估报告和校验结果不断扩充数据和调优。

为此平台提供了模型迭代功能,即当模型训练完毕后,会生成一个最新的版本号,首次V1、之后V2……以此类推。可以通过调整训练数据和算法,多次训练,以获得适合业务需求的模型效果。

如果模型已经是上线状态,依然支持模型迭代,只是需要在训练完毕后更新线上服务接口,在接口地址不变的情况下可以持续优化效果。

检查并优化训练数据

- 新增训练数据
- 通过模型效果评估报告中的各分类的详细评估指标,有针对性地扩充训练数据
- 检查测试模型的数据与训练数据的文本类型与风格是否一致,如果不一致,那么很可能会存在模型效果不错但实际测试效果较差的情况。针对 这种情况建议重新调整训练集,将训练数据与实际业务场景数据尽可能一致

选择高精度模型

在训练模型时,选择高精度的模型,将提升模型的预测准确率。

「高精度」算法内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等方 面的新知识,实现模型效果不断进化

心 发起训练

训练模型

完成数据的标注,即可在「模型中心」目录中点击「训练模型」,开始模型的训练。

按以下步骤操作,启动模型训练:

评论观点抽取	训练模型					提交工单
品 感觉	选择模型	评论观点抽取模型8 💚				
我的模型	训练配置					
创建模型	部署方式	公有云部署 EasyEdge本地部署	如何选择部署方式?			
训练模型	选择算法	○ 高精度 ⑦ ○ 高性能 ⑦				
校验模型	添加数据					
发布模型	添加数据集	+ 请选择				
 EasyData数据服务 数据总览 		数据集	版本		操作	
在线标注		勿动_沙盒测试	V2		移除	
由 公有云服务	自定义测试集 ⑦					
在线服务						
器 EasyEdge本地部署	训练环境	名称 规格		算力	价格	
纯离线服务		O GPU P40 TeslaGPU,	_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	免费	
		GPU V100 TeslaGPU	_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	单卡¥0.45/分钟(50小时*节点小时包余额)	
	开始训练					

Step 1 选择模型 选择此次训练的模型 Step 2 训练配置

部署方式

可选择「公有云部署」、「EasyEdge本地部署」。

如何选择部署方式

选择设备

- 如果您选择了「EasyEdge本地部署」,请根据实际部署设备选择
- 如果您选择了「公有云部署」,无需选择设备

选择算法

您可以根据训练的需求,选择「高精度」或「高性能」算法。不同的算法将影响训练时间、预测速度与模型准确率。如果您手中的标注数据集样 本较少(例如少于1000条),可选择「高精度」的算法;如果您手中有充足的数据集,您可选择「高性能」的算法。

- 高精度:预测准确率效果更高,训练时长与训练文本的长度和数量成正比,1000个样本预计在20分钟完成训练
- 高性能:在相同续联数据量的情况下,有着更快的预测速度,但准确度效果平均损失1~4%

「高精度」算法内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义 等方面的新知识,实现模型效果不断进化

Step 3 添加数据

添加训练数据

- 先选择数据集,再按标签(评价观点词、评价维度、情感倾向、评价片段)选择数据集里的文本,可从多个数据集选择文本
- 训练时间与数据量大小、选择的算法、训练环境有关

添加自定义测试集

上传不包含在训练集的测试数据,可获得更客观的模型效果评估结果。

添加自定义测试集的目的:

如果学生的期末考试是平时的练习题,那么学生可能通过记忆去解题,而不是通过学习的方法去做题,所以期末考试的试题应与平时作业不能一样,才能检验学生的学习成果。那么同理,AI模型的效果测试不能使用训练数据进行测试,应使用训练数据集外的数据测试,这样才能 真实的反映模型效果

期末考试的内容属于学期的内容,但不一定需要完全包括所学内容。同理,测试集的标签是训练集的全集或者子集即可

Step 4 训练模型

点击「开始训练」,即可开始训练模型。

- 训练时间与数据量大小、选择的算法、训练环境有关
- 模型训练过程中,可以设置训练完成的短信提醒并离开页面
- 平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

优惠政策:

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有声音分类 操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。可参考价格说明

模型发布

心 整体说明

训练完成后,可将模型部署在公有云服务器、私有化服务器上,通过API进行调用。 公有云API

- 模型训练完毕后,为了更方便企业用户一站式完成AI模型应用,评论观点抽取模型支持将模型发布成为在线的restful API接口,可以参考示例 文档通过HTTP请求的方式进行调用,快速集成在业务中进行使用。
- 具有完善的鉴权、流控等安全机制,GPU集群稳定承载高并发请求

相关费用 将模型发布为公有云API后,将获得一部分免费调用次数,超出免费额度将根据调用次数进行收费。详见EasyDL价格文档。

私有服务器部署

支持将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便 捷。适用于对数据敏感度、隐私性要求较高、在线离线均有调用需求的企业场景。相关费用 EasyDL已支持将定制模型部署在本地服务器上,只 需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。如需购买永久使用授权,请提交工单咨询。

心 公有云API

の 发布API

发布模型生成在线API

·训练完毕后就可以在左侧目录栏中找到【发布模型】,发布模型表单页面需要自定义接口地址后缀、服务名称,即可申请发布。

发布模型界面示意:

评词	伦观点抽取	发布模型		
	总览			
Ē	模型中心	选择模型	测试v100	\checkmark
	我的模型	部署方式	公有云部署	\checkmark
	创建模型			
	训练模型	选择版本	V2	\checkmark
	校验模型	服务名称 *	商品评论观点抽取模型	
	发布模型	接口地址 *	https://aip.baidubce.com/	rpc/2.0/ai_custo
<u></u>	EasyData数据服务		m/v1/comment_analyse/	opinion_extrac
	数据总览	甘他亜求	せやっては法ロのの高さ	>=+++、
	在线标注	ALCAN	右接口无法满足您的需求, 的问题,500汉字以内	谊抽还布 望胜/犬
	公有云服务			
	在线服务			0/500
	EasyEdge本地部署		提交申请	
	纯离线服务			

或者,在我的模型列表——找到新训练好的模型版本——点击申请发布

【评论观点抽取】测试v100 🗌	模型ID: 15444				吕 训练 ① 历史版本 🔟 删除
部署方式	版本	训练状态	服务状态	模型效果	操作
公有云API	V2	● 训练完成	未发布	召回率: 52.94% 精确度: 71.08% F1-score: 0.599 完整评估结果	查看版本記錄 申请发布 校验

申请发布后,通常的审核周期为T+1,即当天申请第二天可以审核完成,如果需要加急、或者遇到莫名被拒的情况,请在百度云控制台内提交工 单反馈

申请发布通过后,界面和状态示意:

服务名称	模型名称	模型版本	服务状态	调用单价 ⑦	更新时间	操作
评论观点抽取模型	测试v100	V2	● 发布中 ⑦	9点/次	2022-03-10 16:31	-
111111111	test-xsy	V7	● 已发布	9点/次	2022-03-09 11:30	服务详情 更新版本
沙盒模型测试	沙盒测试勿动	V3	● 已发布	4点/次	2022-03-07 16:07	服务详情 更新版本

の调用API

接口描述

基于自定义训练出的评论观点抽取模型,实现对评论文本中评价维度、评价观点的抽取以及对评价观点的情感倾向的判断。模型训练完毕后发布 可获得定制化评论观点抽取API 详情访问:定制化训练和服务平台进行训练。

更多训练模型过程中的常见问题请查看常见问题文档。

如有其它问题,请在百度云控制台内提交工单反馈。

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在定制化训练平台进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

参数 值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

```
参数 值
Content-Type application/json
```

注意:如果出现336001的错误码很可能是因为请求方式错误,定制化文本分类服务以json方式请求。

Body请求示例:

```
{
    "text": "电影故事结构很松散,但是小丽的演技一如既往的不错,配合她的男演员挺好的,比如说小杰,他在对手戏中有很强的互动性",
    "analyse_object": "小丽",
    "analyse_type": 1
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text	是	string	-	超过512个token将被截断
analyse_object	否	string	-	评论实体对象
analyse_type	是	int	1/2/3/4	只能选择枚举值的一个

请求示例代码

```
Python3
                                                                                                                                 *
 **coding=utf-8**
                                                                                                                                 import sys
import json
**保证兼容python2以及python3**
IS_PY3 = sys.version_info.major == 3
if IS_PY3:
  from urllib.request import urlopen
  from urllib.request import Request
  from urllib.error import URLError
  from urllib.parse import urlencode
  from urllib.parse import quote_plus
else:
  import urllib2
  from urllib import quote_plus
  from urllib2 import urlopen
  from urllib2 import Request
                                                                                                                                 \mathbf{T}
```

返回说明

字段	是否必 选	类型	说明
analyse_t ype	是	int	1代表评价片段,2代表评价维度,3代表评价观点词,4代表评价情感倾向;示例:analyse_type:"1"或 analyse_type:"4",每次请求仅为单值;
log_id	是	number	唯一的log id,用于问题定位
result	是	array(obj ect)	需要计算的评价对象数组,元素为字典
+start_off set	是	int	开始位置
+prob	是	float	置信度
+end_offs et	是	int	结束位置
+text	是	string	抽取文本
text	是	string	输入文本,超过512个token将被截断

返回示例

```
{
    "analyse_type":1,
    "log_id":7000918336750814129,
    "result":[
    {
        "result":[
        {
        "start_offset": 10,
        "prob": 0.4721265733242,
        "end_offset": 24,
        "text": "但是小丽的演技一如既往的不错"
        }
        ],
        "text":"电影故事结构很松散,但是小丽的演技一如既往的不错,配合她的男演员挺好的,比如说小杰,他在对手戏中有很强的互动性"
}
```

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
  "error_code": 110,
  "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或在百度 云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问 请在百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大512 UTF-8字符。
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云 控制台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

の调用API

接口描述

基于自定义训练出的评论观点抽取模型,实现对评论文本中评价维度、评价观点的抽取以及对评价观点的情感倾向的判断。模型训练完毕后发布 可获得定制化评论观点抽取API 详情访问:定制化训练和服务平台进行训练。

更多训练模型过程中的常见问题请查看 常见问题文档。

如有其它问题,请在百度云控制台内提交工单反馈。

请求说明

- 请求示例
- HTTP 方法: POST

请求URL:请首先在定制化训练平台进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

参数	值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,定制化文本分类服务以json方式请求。

Body请求示例:

```
{
    "text": "电影故事结构很松散,但是小丽的演技一如既往的不错,配合她的男演员挺好的,比如说小杰,他在对手戏中有很强的互动性",
    "analyse_object": "小丽",
    "analyse_type": 1
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text	是	string	-	超过512个token将被截断
analyse_object	否	string	-	评论实体对象
analyse_type	是	int	1/2/3/4	只能选择枚举值的一个

请求示例代码

Python3	
coding=utf-8	_
import sys	
import json	
保证兼容python2以及python3	
IS_PY3 = sys.version_info.major == 3	
if IS_PY3:	
from urllib.request import urlopen	
from urllib.request import Request	
from urllib parse import urlencode	
from urllib.parse import quote plus	
else:	
import urllib2	
from urllib import quote_plus	
from urllib2 import urlopen	
from urllib2 import Request	

返回说明

字段	是否必 选	类型	说明
analyse_t ype	是	int	1代表评价片段,2代表评价维度,3代表评价观点词,4代表评价情感倾向;示例:analyse_type:"1"或 analyse_type:"4",每次请求仅为单值;
log_id	是	number	唯一的log id,用于问题定位
result	是	array(obj ect)	需要计算的评价对象数组,元素为字典
+start_off set	是	int	开始位置
+prob	是	float	置信度
+end_offs et	是	int	结束位置
+text	是	string	抽取文本
text	是	string	输入文本,超过512个token将被截断

返回示例

```
{
    "analyse_type":1,
    "log_id":7000918336750814129,
    "result":[
    {
        "start_offset": 10,
        "prob": 0.4721265733242,
        "end_offset": 24,
        "text": "但是小丽的演技一如既往的不错"
    }
    ],
        "text":"电影故事结构很松散,但是小丽的演技一如既往的不错,配合她的男演员挺好的,比如说小杰,他在对手戏中有很强的互动性"
}
```

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
  "error_code": 110,
  "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或在百度 云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问 请在百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大512 UTF-8字符。
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云 控制台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

心 纯离线服务

の 发布API

在训练模型时,您需要选择「EasyEdge本地部署」的训练方式,才能发布本地部署的私有API。

私有API介绍

将模型以Docker形式在本地服务器(仅支持Linux)上部署为http服务,可调用与公有云API功能相同的接口。可纯离线完成部署,服务调用便捷

发布私有API的流程训练完毕后,您可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可将模型部署到私有服务器:

1. 在「发布模型」页面中,选择模型及模型的版本,选择部署方式为「EasyEdge本地部署」、集成方式为「API-纯离线服务」。点击「发布」, 即可跳转至「发布新服务」页面

发布模型页面示意:

评计	论观点抽取	 纯离线服务 ≥ 及	5 仲 新 服 务			
88	总览 模型中心	服务器				
	我的模型 创建模型	1 选择部署形	式	- 2 填写个	人信息	
	训练模型	集成方式				
	校验模型	部署类型	● 单模型部署	○ 増量部署		
<u>//</u>	及中候望 EasyData数据服务	证书	测试证书		\sim	
	数据总览	选择模型	同时训练测试		\sim	
	在线标注	选择版本	V2		\checkmark	
	公有云服务	服务名称 *	extract			
	EasyEdge本地部署	证书生效时间 *	2022/3/10	ŝ		
	纯离线服务	指纹文件 *	上传文件	② 指纹提取工具	下载 指纹提取	工具使用说明
		选择系统和芯片 了解对应硬件				

2. 在「发布新服务」页面,选择部署类型,填写服务名称、证书生效时间等信息,选择对应的系统和芯片。

- 部署类型可支持单模型部署和增量部署
- 增量部署申请,指需要在一台服务器上部署多个模型部署包时使用。进行增量部署时,需在「已部署服务」选择同台服务器历史中最近 部署的部署包,此步骤用来关联不同部署包中的license文件
- 3. 上传指纹文件。详细操作见指纹提取工具说明,可通过指纹工具进行指纹的提取
- 4. 点击下一步,填写个人详细信息后即可发布。发布完成后,即可在服务器目录下看到发布处于审核中的状态

个人信息的填写仅供EasyDL团队了解您,便于后续合作接洽,不会作为其他用途使用

- 5. 等待审核通过,前往「纯离线服务」等待部署包制作完成后,下载部署包,并参考文档完成集成
- 价格说明 EasyDL已支持将定制模型部署在本地服务器上,只需在发布模型时提交本地服务器部署申请,通过审核后即可获得一个月免费试用。

如需购买永久使用授权,请提交工单咨询。

つ。调用API

本文档主要说明定制化模型本地部署后,如何使用本地API。如还未训练模型,请先前往EasyDL进行训练。

如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:868826008)联系群管

部署包使用说明 部署方法

EasyDL定制化文本分类模型的本地部署通过EasyPack实现,目前提供单机一键部署的方式。

在EasyDL申请、下载部署包后,在本地服务器新建目录(建议目录命名规则:easyDL_服务名称_模型版本号),将软件包上传至该目录。请参考EasyPack-单机一键部署使用python2版本来部署,部署成功后,启动服务,即可调用与在线API功能类似的接口。

运维检查

EasyDL服务器API部署应用健康检查(或故障排查)脚本:trouble_shooting.tar

脚本能力:鉴权服务健康检测、容器状态检查、端口探活、网络联通性测试、容器关键报错日志输出等

使用方法:将脚本上传至服务器任意目录(或在服务器直接下载),并解压后运行。

```
**解压**
tar vxf trouble_shooting.tar
**执行**
bash trouble_shooting.sh
```

授权说明

本地部署包根据服务器硬件(CPU单机或GPU单卡)进行授权,只能在申请时提交的硬件指纹所属的硬件上使用。

部署包测试期为1个月,如需购买永久授权,可提交工单咨询

API参考

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在EasyDL进行自定义模型训练,完成训练后申请本地部署,本地部署成功后拼接url。

请求URL: http://{IP}:{PORT}/{DEPLOY_NAME}/TextClassification

- IP:服务本地部署所在机器的ip地址
- PORT:服务部署后获取的端口
- DEPLOY_NAME:申请时填写的本地服务名称

Header如下:

参数	值
Content-Type	application/json

Body请求示例:

```
{
    "text": "电影故事结构很松散,但是小丽的演技一如既往的不错,配合她的男演员挺好的,比如说小杰,他在对手戏中有很强的互动性",
    "analyse_object": "小丽 ",
    "analyse_type ":1
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
text	是	string	-	超过512个token将被截断
analyse_object	否	string	-	评论实体对象
analyse_type	是	int	1/2/3/4	只能选择枚举值的一个

返回说明

返回参数

字段	是否必 选	类型	说明
analyse_t ype	是	int	1代表评价片段,2代表评价维度,3代表评价观点词,4代表评价情感倾向;示例:analyse_type:"1"或 analyse_type:"4",每次请求仅为单值;
log_id	是	number	唯一的log id,用于问题定位
result	是	array(obj ect)	需要计算的评价对象数组,元素为字典
+start_off set	是	int	开始位置
+prob	是	float	置信度
+end_offs et	是	int	结束位置
+text	是	string	抽取文本
text	是	string	输入文本,超过512个token将被截断

返回示例

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如缺少必要出入参时返回:

```
{
    "error_code": 336001,
    "error_msg": "Invalid Argument"
}
```

错误码 错误信息

烘沫

336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(868826008)或工单联系技术支持团队
336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本的编码UTF-8等问题。有疑问请通过QQ群(868826008)或工单联系技 术支持团队
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数代码格式是否有误。有疑问请通过QQ群(868826008)或工单联系 技术支持团队
336003	Base64解码失败	图片/音频/文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有 疑问请通过QQ群(868826008)或工单联系技术支持团队
336004	输入文件大小不合 法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大1024 UTF-8字符。有疑问请通过 QQ群(868826008)或工单联系技术支持团队
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云控制 台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请通过QQ群(868826008) 或工单联系技术支持团队
337000	Auth check failed	离线鉴权调用失败

模型更新/回滚操作说明 模型更新 1、在EasyDL-纯离线服务发布页面,找到您的服务器API发布记录,点击【更新版本】,选择「更新包」或 「完整包」来发布。 两者区别:

包类型	描述
更新包	仅包含最新的模型应用,需执行download.sh脚本下载所需镜像等依赖文件
完整包	包含模型应用和其他鉴权服务,需执行download.sh脚本下载所需完整依赖文件

2、(CPU模型可忽略)如果您训练的模型为GPU版本,系统会生成多份下载链接。请在GPU服务器执行 nvidia-smi命令,根据返回的Cuda Version来选择对应的部署包链接下载。

3、在服务器新建目录(建议标记对应模型的版本号,便于区分不同模型版本),如easydl_\${DEPLOY_NAME}_v2

\${DEPLOY_NAME}:申请时填写的服务名称

以如下场景举例说明:模型版本从V1升级至V2

```
**1.新模型准备**
**创建指定版本的目录**
mkdir easedI_${DEPLOY_NAME}_v2
cd easedl_{DEPLOY_NAME}_v2
**将部署包上传至服务器该目录并解压**
tar zvxf xx.tar.gz
**解压后,进入指定目录执行download脚本下载模型所依赖文件**
cd original && bash download.sh
**2.旧模型备份**
**历史模型备份**
cp -r /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V1
**记录当前模型的端口号**
docker ps -a |grep ${DEPLOY_NAME}
**3.模型升级**
cd package/Install
**卸载当前已安装的旧的easyDL服务: ${DEPLOY_NAME},前面已备份**
python2 install.py remove ${DEPLOY_NAME}
**安装当前部署包内新的EasyDL服务: ${DEPLOY_NAME}**
python2 install.py install ${DEPLOY_NAME}
**(可选操作)更新证书**
python2 install.py lu
```

模型回滚 以如下场景举例说明:模型版本从V2回滚至V1

```
**重命名当前v2模型目录名称**

mv /home/baidu/work/${DEPLOY_NAME} /home/baidu/work/${DEPLOY_NAME}_V2

**使用V1版本**

cp -r /home/baidu/work/${DEPLOY_NAME}_V1 /home/baidu/work/${DEPLOY_NAME}

**停止当前模型容器**

docker ps -a |grep ${DEPLOY_NAME}

docker ms -f ${容器名}

**创建新的容器**

cd /home/baidu/work/${DEPLOY_NAME} && bash start/start-1.sh

*** (可选操作)进入V1版本部署包所在目录执行license更新操作,假如部署包在/opt目录下,以您实际目录为准**

cd /opt/easydl_${DEPLOY_NAME}_V1/original/package/Install && python2 install.py lu
```

方法二:

进入模型V1所在目录,参考上述【模型更新】步骤,执行模型升级操作(即先卸载v2,后升级为v1)

文本创作 (已下线)

```
文本创作介绍
```

简介

Hi,您好,欢迎使用百度EasyDL定制化训练和服务平台。

定制文本创作模型,基于ERNIE 3.0大模型实现对输入文本内容进行创作和续写。平台提供标注工具,您可在平台上传文档,完成标注后可直接进行模型训练。

更多详情访问: EasyDL自然语言处理方向

应用场景 1、新闻资讯创作:训练新闻资讯文本的自动续写和创作模型,节省大量编辑人力

2、文章摘要:训练对网络媒体等文章的自动摘要,进而实现各类文章的自动摘要

3、诗歌对联创作、专业文本续写:定制训练文学创作的模型,对诗词、对联和专业文本进行创作和续写

4、营销广告文案创作:定制训练广告文案创作的模型,给广告营销带来新的创作灵感

5、其他:尽情脑洞大开,训练你希望实现的创作模型 技术特色 文本创作模型内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等方面的新知识,实现模型效果不断进化。

文心大模型是百度发布的产业级知识增强大模型,是千行百业AI开发的首选基座大模型。文心大模型既包含基础通用的大模型,也包含了面向重 点领域和重点任务的大模型,还提供丰富的工具与平台,支撑企业与开发者进行高效便捷的应用开发。"知识增强"是文心的核心特色,文心能够 同时从大规模知识和海量多元数据中持续学习,如同站在巨人的肩膀上,训练效率和理解准确率都得到大幅提升,并具备了更好的可解释性。

使用流程训练模型的基本流程如下图所示,全程可视化简易操作,在数据已经准备好的情况下,最快15分钟即可获得定制模型。

Ē	<u>ت</u> ب	==	ক
<u>o</u>		0	0
1. 创建模型	2. 上传并标注数据	3.训练模型并校验效果	4. 上线模型获取API或离线SDK

数据准备

心 创建数据集并导入

1. 创建数据集

您可以在左侧导航栏中中,选择"数据总览"并点击主内容区域的按钮「创建数据集」,默认数据类型为"文本",标注类型为"文本创作"。

文本创作		<u>ا</u>	我的数据总览 >	创建数据集
品 党党				
曲 模型	中心		数据集名称	限制50个汉字内
我的标	莫型		数据类型	文本
创建植	莫 <u>型</u>		数据集版本	V1
训练	莫型			
校验机	莫型		标汪类型	文本创作
发布核	莫型			
🖉 Easyl	Data数据服务			
数据	总览		标注模板	 文本创作
在线机	示注			
目 公有;	云服务		数据集属性	 数据自动去重 ⑦ 数据不去重 ⑦
在线题	服务			完成

2.导入文本数据

进入到新创建的文本创作数据集中。您可以在文本创作任务的数据集中,上传带有标注信息的数据,和无标注信息的数据。

^{チナ} ズ楽EasyDL 产	*品介绍 操作平台 应用案例 使用文档	社区 飞桨官网 AI主站 🕰 liys29
ERNIE大模型创作	夏 我的鼓摇总览 > zl_test/V3/导入	
器 总览		常见问题
□ 模型中心		
我的模型	数据集D 111291 版本号 V3	1、样本的格式与字符限制 >
创建模型	备注 🖸	2、标注数据说明 >>
训练模型	Ⅰ 标注信息 ∨	
校验模型	标注类型 大模型创作 标注模板 大模型创作	
发布模型	数据总量 10000 已标注 10000 (进度100.00%)	
☑ EasyData数据服务	待确认 0 大小 0.08M	
数据总览	数据清洗	
在线标注	暂未做过数据清洗任务	
12 模型部署	导入数据	
公有云服务	数据标注状态 ● 无标注信息 ○ 有标注信息	
	导入方式 本地导入 ~	
	上代Excel文件 ① 上代Excel文件	
	綿认并返回	

在数据导入方式选择本地数据集,根据您已有的数据存储格式,选择上传格式。目前对未标注数据和已标注数据都仅提供了Excel文件上传方式。

2.1未标注数据上传方式:

以Excel文件上传

- 第一列作为原始文本,每行是一组样本,首行为表头默认将被忽略,每组数据文本内容的字符数不超过512个字符(包括中英文、数字、符号等),超出的字符可正常保存,但可能无法参与训练。详见平台导入数据处的数据样例。
- 2. 文件类型支持xlsx格式,单次上传限制100个文件;文件格式示意图如下:

文本创作	我的数据总览 > 111111111/V1/导入						
88 总览							
曲 模型中心	┃ 创建信息 ~						
我的模型	数据集ID 339442 版本号 V1						
创建模型	备注						
训练模型	↓标注信息 ~						
校验模型	标注类型 文本创作 标注模板 文本创作						
发布模型	数据总量 98 已标注 98 (进度100.00%)						
☑ EasyData数据服务	待确认 0 大小 0.01M						
数据总览	┃ 数据清洗						
在线标注	暂未做过数据清洗任务						
目 公有云服务	导入数据						
在线服务	数据标注状态 💿 无标注信息 🦳 有标注信息						
	导入方式 本 地导入 ~						
	上传Excel文件						
	确认并返回						

示例:例如在歌词创作场景中,希望用户输入歌名,由模型创作歌词,则上传文本为:"歌名:晴天"。

请注意,"歌名:"作为样本的前缀,需要固定在每一个样本中添加,不固定的前缀,将影响模型效果;

示例样本请详见平台导入数据处的数据样例。

2.2已标注数据上传方式:

以Excel文件导入

- 1. Excel文件内数据格式要求为:首行为表头,将不录入数据集中,第一列和第二列分别作为模型输入文本和模型输出文本
- 每行是一组样本,输入文本不超过512个字符,输出文本不超过128个字符,超出的字符可正常保存,但可能无法参与训练。(字符包括中英文、数字、符号等)
- 3. 文件类型支持xlsx格式,单次上传限制100个文本文件;文件格式示意图如下:

输入文本内容	输出文本内容
文本内容a1	文本内容a2
文本内容a1	文本内容b2
文本内容a3	文本内容c1

示例:例如在歌词创作场景中,希望用户输入歌名,由模型创作歌词,此场景的标注数据形式可有多种:

数据格式一:

- 输入文本内容:"歌名:晴天;歌词:"
- 输出文本内容:"故事的小黄花 从出生那年就飘着 童年的荡秋千"

在数据格式一中,输入到模型的文本是"歌名:晴天;歌词:","歌名:"作为模型输入的前缀,需要固定在每一个样本中添加,不固定的前缀,将 影响模型效果;"歌词:"作为模型输出样本的后缀,需要固定在每一个样本中添加,不固定的后缀,将影响模型效果;

您需要在模型预测阶段,确保回传模型输入包含了用户输入的内容,并且拼接了前缀"歌名:"和后缀"歌词:",在模型服务返回内容时,则直接 输出歌词。 数据格式二:

- 输入文本内容:"歌名:晴天;"
- 输出文本内容:"歌词:故事的小黄花 从出生那年就飘着 童年的荡秋千"

在数据格式二中,前缀"歌名:"和后缀"歌词:"分别在模型的输入和输出中。"歌名:"作为输入文本的前缀,需要固定在每一个样本中添加,不固定的前缀,将影响模型效果;"歌词:"作为输出样本的前缀,需要固定在每一个样本中添加,不固定的后缀,将影响模型效果;

您需要确保预测推理阶段,模型输入包含了用户输入的内容,并且拼接了前缀"歌名:"

数据格式三:

- 输入文本内容:"晴天"
- 输出文本内容:"故事的小黄花 从出生那年就飘着 童年的荡秋千"

在数据格式三中,前缀"歌名:"和后缀"歌词:"都不存在模型的训练数据,则所有样本都不要添加前缀和后缀。则您需要确保在预测推理阶段, 模型输入仅有歌名内容。

详见平台导入数据处的数据样例。

4. 上传时,单个数据集总量仅支持上传10000条样本(包括标注数据和未标注数据)。

② 文本创作数据标注

文本创作数据标注

1. 图中所示,模型输出文本框中,您可直接编辑模型输出文本,如图所示:

文本创作	在线标注 > 111111111 > 标注	11111111 / VI ~
彩 总览	金额(98) 天栅注痛息(1) 有标注信息(97)	模型输出文本
凸 模型中心		
我的模型	模型输入文本 自 删除文本 上一派 下一篇	请输入模型生成的文本
创建模型		
训练模型	歌名: 大吾大利今晚纪鸿 歌词:	
校验模型		
发布模型		
☑ EasyData数据服务		
数据总览		
在线标注		
回 公有云服务		
在线服务		
		0/128

2. 编辑完成后,点击'下一篇'按钮自动保存并进入下一篇数据

心 文本创作数据集去重策略说明

重复样本的定义一个样本包括文本内容和标签。重复样本的定义,是指您上传的数据中,存在两个样本的文本内容完全一致。则被判定为两个样本是重复样本。例如:

模型输入内容	模型输出内容
歌名:晴天;歌词:	故事的小黄花 从出生那年就飘着 童年的荡秋千
歌名:晴天;歌词:	故事的小黄花
歌名:晴;歌词:	故事的小黄花 从出生那年就飘着 童年的荡秋千

上表三个样本均为重复样本,前两个样本虽然标签不一,但文本内容一致,也为重复样本。

Tips: "如何利用好重复样本",如果您在模型训练过程中,需要通过增加某个类别标签的预测权重,可以通过增加此标签的重复样本来达到此目标。

平台去重策略 平台提供了可去重的数据集,即对您上传的数据进行重复样本的去重。注意:当您确定了数据集为去重或非去重的属性后,便不可 修改。

当您创建了一个去重的数据集时,在后续上传数据的过程中,平台可通过检验您当前上传的样本与已上传到此数据集下的样本是否相同,如果相同,则会使用新的样本替代旧的样本。此时分为几种情况,如下:

数据集中有未标注样本,上传重复的已标注样本,此时未标注样本将被覆盖

数据集中有已标注样本,上传重复的未标注样本,此时已标注样本将被覆盖

数据集中有已标注样本,上传不同标注的已标注样本,此时已有的标注样本将被覆盖

模型训练

の 模型创建

创建模型 在左侧目录【模型中心】下点击创建模型,按照指示,填入相应的模型名称,选择模型归属、所属行业、应用场景,以及您的下关系信息。填写完成后点击下一步,即可我完成模型的创建。

文本创作	 模型列表 > 仓	刘建模型
部 总览 山 模型中心	模型类别	文本创作
我的模型	模型名称 *	
创建模型	您的身份	企业管理者 企业员工 学生 教师
训练模型 校验模型	公司名称 *	百度
发布模型		企业认证流程较快,认证过程中您可继续创建模型,完成后系统会自动同步状态,
🖉 EasyData数据服务	所属行业 *	请选择行业 🗸
数据总览	应用场景 *	请选择应用场景 🗸
在线标注	邮箱地址 *	r*****@163.com
目 公有云服务 在线服务	联系方式 *	186****356 Ø
	功能描述 *	
		0/500
		完成

模型创建完成后,在左侧目录【模型中心】下点击我的模型,可以看到您已经创建成功的模型。

1.创建模型后可持续新增模型版本,因此不必每次训练模型都创建模型

2.目前单个用户在每种类型的模型下最多可创建10个模型,每个模型均支持多次训练,若需要创建超过10个以上的模型,请在百度云控制台 内提交工单反馈。

3.如果您是企业用户,建议您按照真实企业信息进行填写,便于EasyDL团队后续更好的为您服务

の 模型训练

创建任务

当您的模型以及数据集创建完成后,您可以点击左边目录导航栏中的【训练模型】,创建模型或选择您已经创建的模型,再添加您要使用的训练数据集,建议您使用的每个文本创作数据的样本数应达到1000个以上,再启动训练.

训练环境 平台为您提供了GPU算力机器,TeslaGPU*V100_32G显存单卡* 80核CPU_640G内存,训练设备数默认为8(暂不支持增删机器)。

添加数据集以及完成配置后,点击开始训练即可启动训练。

注意:文本创作任务,每次最高支持1万条样本的训练,训练时间最长约1个小时。在您提交任务后,需要与平台其他用户任务排队等待算力机器,此时间由排队任务数决定。

心 模型效果评估

模型评估报告

- 1. 校验指标:仅提供 BLEU-4指标 : Bilingual evaluation understudy, BLEU 的分数取值范围是 0~100%, 分数越接近100%, 说明生成的句子 质量越高。
- 2. 训练完成后,可以在【我的模型】列表中看到模型效果,以及详细的模型评估报告。

模型校验

实际效果可以在左侧目录中找到【模型校验】功能进行校验,或者发布为接口后测试。

校验过程中,您只需要输入或者上传校验文本(文本上限长度为512字符),即可进行校验,输出仅支持52token。

省前模型精确率 81.97% 评估报告		识别结果 如何优化效果?
请输入校验的文本,或 点击上传文本	支持文本格式:txt,文本长度上限为512汉字(字符)	输出文本 仅支持出书52个token
请输入文本		
		没有满足条件的识别结果
	0/512	le le
校验 再次上传文本		申请上线

模型发布

の 模型发布

发布模型生成在线API

模型训练完毕后,为了更方便企业用户一站式完成AI模型应用,支持将模型发布成为在线的restful API接口,可以参考示例文档通过HTTP请求的 方式进行调用,快速集成在业务中进行使用。具有完善的鉴权、流控等安全机制,GPU集群稳定承载高并发请求。

左侧目录栏中找到【模型部署-公有云服务-发布新服务】,填写发布模型表单页面,自定义接口地址后缀、服务名称,即可申请发布,平台上发 布模型流程完全一致。

发布模型界面示意如下:

文本创作	÷	公有四部署								提交工单
88 总筑 通 模型中心 我的模型 创建模型		公有芸藝券映期 1. 年約24百三級券、将別紙次成約線型部署在百度三級券員、通过AP4線21 国用模型。 初何捐機型就作成於公有三級券化及現而代码示例件考, 参考文档 2. 客个公有三級券條整遇器種副起還本何间,參次項用所消耗的点就也不同。具体的身份可以在下方服券利款"调用身份"中容量,換算規則: 1.6=0.001元。 3. 每个24百运路发展点成之后享有10000点克量曲版,仅作为该重券用用消耗,如需指定使用,就仅在过3条位开油付值								点击收起
训练模型 校验模型		发布新服务	控制台						输入模型名称	
发布模型		服务名称	模型名称	模型版本	服务状态	调用单价 ⑦	更新时间	操作		
☑ EasyData数据服务 数据总览		emie3v1	binbin_test_1109	VI	● 已发布	-	2021-11-09 21:52	服务详情 更新经	反本	
在线标注										
□ 公有云服务										
在线服务										

文本创作	4≣	发布模型	
品 总览			
曲 模型中心		选择模型	AI写周报
我的模型		部署方式	公有云部署 >>
创建模型			nu 13 nu HL. Ni
训练模型		选择版本	V4 ~
校验模型		服务名称 *	
发布模型		接口地址 *	https://aip.baidubce.com/rpc/2.0/ai_custo
☑ EasyData数据服务			m/v1/text_gen/
数据总览		甘州西北	
在线标注		兵他安水	若接口无法满足您的需求,请描述希望解决 的问题,500汉字以内
目 公有云服务			
在线服务			0/500
			提交申请

或者,在模型中心——我的模型——选择模型发布。

文本创作	<=	模型列表							
88 总宽									
凸 模型中心		创建模型							
我的模型		【なまの後】 小宮田坂	「二 株面につ: 104400						
创建模型		X401FI Al-DATE	☑ 18(12)10-184429						
训练模型		部署方式	版本	训练状态	服务状态	模型效果	操作		
校验模型		公有云API	V4	• 训练完成	未发布	BLEU-4: 5.90% 完整评估结果	查看版本配置 申请发布 校验		
反巾模型									
 EasyData数据服务 数据总览 		【文本创作】asadadas	【文本创作】 asadadasdad 🖸 模型ID: 175281						
在线标注		部署方式	版本	训练状态	服务状态	模型效果	操作		
目 公有云服务		公有云API	V1	• 训练终止	未发布	-	查看版本配置		
在线服务									
发布后:									
模型列表 > binbin_tes	st_1109 > \	V1							

I	基础信息		
	任务开始时间	2021-11-09 16:07 任务时长 ③ 5小时44分钟 训练时长 ③ 4小时47分钟	
	训练算法	公有云APIERNIE3.0知识增强大模型算法	
I	训练数据集		
	No. 数据集名称		操作
	1	binbin_test_1109V1	查看详情

申请发布后,通常的审核周期为T+1天,即当天申请第二天可以审核完成,如果需要加急、或者遇到莫名被拒的情况,请在百度云控制台内提交 工单反馈

^の文本创作-调用API文档

本文档主要说明定制化模型发布后获得的API如何使用,如有疑问可以通过以下方式联系我们:

在百度云控制台内提交工单

进入EasyDL社区交流,与其他开发者进行互动

加入EasyDL官方QQ群(群号:868826008)联系群管

接口描述 基于自定义训练出的创作模型,实现基于输入文本内容的个性化创作。模型训练完毕后发布可获得定制API。 请求说明 HTTP 方法: POST

请求URL:请首先在定制化训练平台进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,定制化文本分类服务以json方式请求。

Body请求示例:

```
{
.
"text": "<UTF-8编码文本>",
"max_gen_len": "64"
}
```

Body中放置请求参数,参数详情请看模型请求参数。

模型请求参数:

参数	是否必 选	类型	可选值范围	说明
text	是	string	512token以 内	输入文本,超过512个token将被截断
max_gen_le n	否	int	0-128token	生成时返回字符数,可选0-128,可按需设置,通常生成字符数越少,用户等待时间越少。默认取 值为:64

请求示例:

###### coding=utf-8	
mport sys	
mport json	
###### 保证兼容python2以及python3	
S_PY3 = sys.version_info.major == 3	
f IS_PY3:	
from urllib.request import urlopen	
from urllib.request import Request	
from urllib.error import URLError	
from urllib.parse import urlencode	
from urllib.parse import quote_plus	
else:	
import urllib2	
from urllib import quote_plus	
from urllib2 import urlopen	
from urllib2 import Request	
from urllib2 import URLError	
from urllib import urlencode	
reload(sys)	
sys.setdefaultencoding('utf8')	
##### 防止https证书校验不正确	
mport ssl	
ssl_create_default_https_context = sslcreate_unverified_context	
###### 百度云控制台获取到ak,sk以及	
###### EasyDL官网获取到URL	
##### ak	

```
API_KEY = 'kQWXQ8oe5G5T7ATzXXXXXXX'
###### sk
SECRET_KEY = 'Y30GtHsKzyH6fUUsQI32GvoBXXXXXXX'
###### url
EASYDL_TEXT_CLASSIFY_URL = "https://aip.baidubce.com/rpc/2.0/ai_custom/v1/text_gen/lirics_gen"
""" TOKEN start """
TOKEN_URL = 'https://aip.baidubce.com/oauth/2.0/token'
.....
  获取token
....
def fetch_token():
  params = {'grant_type': 'client_credentials',
         'client_id': API_KEY,
         'client_secret': SECRET_KEY}
  post_data = urlencode(params)
  if (IS_PY3):
     post_data = post_data.encode('utf-8')
  req = Request(TOKEN\_URL, \ post\_data)
  try:
     f = urlopen(req, timeout=5)
     result\_str = f.read()
     print('success')
  except URLError as err:
     print(err)
  if (IS_PY3):
     result_str = result_str.decode()
  result = json.loads(result_str)
  if ('access_token' in result.keys() and 'scope' in result.keys()):
     if not 'brain_all_scope' in result['scope'].split(' '):
       print('please ensure has check the ability')
       exit()
     return result['access_token']
  else:
     print('please overwrite the correct API_KEY and SECRET_KEY')
     exit()
.....
  调用远程服务
.....
def request(url, data):
  if IS_PY3:
     req = Request(url, json.dumps(data).encode('utf-8'))
  else:
     req = Request(url, json.dumps(data))
  has_error = False
  try:
     f = urlopen(req)
     result_str = f.read()
     if (IS_PY3):
       result\_str = result\_str.decode()
     return result str
  except URLError as err:
     print(err)
if __name__ == '__main__':
  # 获取access token
```

token = fetch_token()

模型返回参数:

参数	是否必选	类型	可选值范围	说明
log_id	是	number	-	唯一的log id,用于问题定位
+content	否	string	-	返回的生成结果
+is_truncate	否	boolean	0或1	返回的生成结果是否被截断,1为被截断,0为没被截断,与设置的max_gen_len的token数有关

示例样本1:

以歌词创作场景为例,模型输入(入参)为歌词名称,模型输出(出参)为歌词内容,假设训练数据中,输入文本的数据模板的为:"歌名: xxx;歌词:",输出文本的数据模板为:"xxxx"(xxxx代表生成歌词内容)。

当用户输入:"夏日的海边"

- 入参text字段为:"歌名:夏日的海边;歌词:"
- 出参content为:"故事的小黄花 从出生那年就飘着 童年的荡秋千"

其中,"歌名:"作为样本的前缀,需要固定添加在每一次的请求中,不固定的前缀,将影响模型效果;"歌词:"作为样本的后缀,需要固定添加 在每一次的请求中,不固定的后缀,将影响模型效果;

示例输入:

```
{
"text": "歌名:夏日的海边;歌词:",
"max_gen_len": "64"
}
```

示例返回:

```
{
"log_id": "123456",
"result": [{
"content": "故事的小黄花 从出生那年就飘着 童年的荡秋千",
"is_truncate": 0
}]
}
```

示例样本2:以旅行问答场景为例:

- 用户输入:"十月去青海应该带什么?",则入参text字段为:"问题是:十月去青海应该带什么?答案是:"
- 输出:"带个男朋友",则content为"带个男朋友"

其中,"问题是:"是前缀、"答案是:"是后缀,前缀避免改为"问题:"或"题目是:"等相关词组,后缀避免改为"答案:"或"回答:"等相关词组;前 后缀都需要固定添加在每一次的请求中,否则将影响模型效果。

示例输入:

```
EasyDL 语音使用说明
```

```
{
"text": "问题是:十月去青海应该带什么?答案是:",
"max_gen_len": "64"
}
```

示例返回:

```
{
"log_id": "123456",
"result": [{
"content": "带个男朋友",
"is_truncate": 0
}]
}
```

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

error_code:错误码。

error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有 疑问请在百度云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈
336003	Base64解码失败	文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问 请在百度云控制台内提交工单反馈
336004	输入文件大小不合法	文本大小不合法,目前支持文本文件类型为支持txt,文本文件大小限制长度最大512 UTF-8字符。
336005	解码失败	文本编码错误(不是utf-8),目前支持文本文件类型为支持txt。如果遇到请重试,如反复失败,请在百度云 控制台内提交工单反馈
336006	缺失必要参数	未上传文本文件
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

EasyDL 语音使用说明

EasyDL语音介绍

心 功能介绍

EasyDL语音,包含语音识别和声音分类两种训练能力,零代码自助训练语音识别语言模型,声音分类模型。提升业务领域专有名词识别准确率, 区分不同声音类别,广泛适用于行业数据采集录入、语音指令、呼叫中心、声音类型检测等应用场景。

• 科学评估,提供多维报告

上传业务场景音频和标注文本,系统自动评估语音识别基础模型得到基线准确率,输出字准、句准、核心词准等多维度评估结果报告

• 上传语料,深度训练模型

选择基础模型上传业务场景相关文本训练语料即可自助训练语言模型,支持词汇、长文本等多种训练方式

• 迭代优化,获取最佳模型

可多次上传文本数据迭代训练,每次训练后系统自动评估训练结果,训练效果精准提升,直观可视

• 自动上线,模型专属使用

训练效果满意后,无需复杂操作,通过申请流程即可将模型上线使用,模型专属使用

∞ 特色优势

零门槛操作

一站式自动化训练,上传文件即可最快10分钟训练优化语言模型。

高精准评估

系统自动评估多种基础模型,推荐最优模型进行训练。训练前后均提供字准、句准、核心词准等多维度评估结果报告,

强训练效果

预置百度超大规模预训练模型,多个基础模型支持多行业多业务场景。支持词汇、长段文本等多种训练方式。支持多次上传训练文本,迭代训练 不断优化模型,平均5%-25%识别准确率提升。

超灵活部署

模型通过申请流程即可自动上线,用户账号专属使用。支持在线API,websocket API,多种操作系统的SDK,适配多种终端的使用需求

⊙应用场景

训练语音识别模型可以在如下的应用场景中获得更好的识别效果

- 语音对话:APP语音助手,金融、医疗、航空公司智能机器人对话等短语音交互场景,使用领域中的专业术语进行训练,提高对话精准度
- 语音指令: 智能硬件语音控制、 app内语音搜索关键词、语音红包等场景,训练固定搭配的指令内容,让控制更精确
- 语音录入:农业采集、工业质检、物流快递单录入、餐厅下单、电商货品清点等业务信息语音录入场景,训练业务中的常用词,录入的结果更加有效
- 电话客服:运营商、金融、地产销售等电话客服业务场景,使用领域中的专业术语进行训练,提高对话精准度

训练声音分类模型可以在如下的应用场景中定制区分不同的声音类型

- 安防监控: 定制识别不同的异常或正常的声音, 进而用于突发状况预警
- 科学研究:定制识别同一物种的不同个体的声音、或者不同物种的声音,协助野外作业研究

心 快速使用

语音技术下任一接口进行付费,即可免费训练语音识别模型,不收取额外的训练费用。

语音识别

语音识别介绍

HI,您好,欢迎使用EasyDL语音识别。

原语音自训练平台即日已结束公测正式上线,品牌升级更名为"EasyDL语音识别",平台和语音识别通用接口全面打通,语音技术下任一接口开通 付费即可免费训练语音识别模型,无需额外费用。

如果您在调用通用语音识别模型时遇到如下困难:

1、在垂直业务领域下通用语音识别模型准确率不满足需求,语音识别应用的场景专业词汇较集中,如医疗词汇、金融词汇、教育用语、交通地 名、人名等,识别结果存在"同音不同字"的情况。例如"虹桥机场"识别为"红桥机场";"债券"识别为"在劝"。

2、语音识别结果不准带来更高的后处理成本,并且语音识别模型针对性优化训练存在技术门槛、成本高、训练周期长。

欢迎使用EasyDL语音识别,可以通过自助训练语言模型的方式有效提升您业务场景下的识别准确率。

平台使用的基本流程如下图所示,全程可视化简易操作,在数据已经准备好的情况下,最快一天内即可获得专属模型。

 01
 →
 02
 →
 03
 →
 04

 创建模型
 系统评估
 训练模型
 上线模型

根据业务需求, 创建并选择 想要训练的接口类型 杀统计估
上传音频测试集,系统自动
评估基础模型识别率

上传业务文本,一键启动模型训练,可迭代训练

自动上线模型, 语音识别接 口配置参数即可调用

1、**创建模型**:选择您需要训练的语音识别接口,目前支持训练**短语音识别·中文普通话、短语音识别极速版、实时语音识别·中文、呼叫中心语音** 解决方案接口。填写基础信息为您的模型进行命名和功能描述,并留下您的联系方式以便于我们和您联系。

2、**系统评估**:上传您业务场景中的**真实音频和对应的正确标注文本**(尽可能覆盖全部的场景),**客观科学地评估基础模型的识别率。**根据评估 结果,系统自动推荐最佳的基础模型,您可以选择任一基础模型进行训练。

3、**训练模型**:上传您**业务场景中出现的高频词汇或者是长句文本**,可以有效提升业务用语的识别率;并可以迭代训练,持续优化。

5、**上线模型**:得到满意的训练模型即可申请上线,审批通过自动上线模型。模型上线后,在语音识别的接口中配置模型参数即可使用训练后的 效果。

开始使用平台前,先了解以下您需要提前准备的物料及准备建议:

1、【测试集(包括业务音频+准确100%的标注文本)】,用于评估基础模型识别率和训练后模型识别率,相当于准备一份"标准答案"。如 果模型使用业务范围较广(例如某行业领域模型),建议测试集在1000-3000条左右评估会相对客观;如果是针对某些特定场景训练,可只 提供该场景的音频测试集几十条-几百条均可,包含希望评估的业务内容即可。

2、【训练集(投入平台进行训练的文本)】,用于语言模型训练,建议文本要和测试集的内容强相关。训练文本可以放置希望提升识别效 果的词汇,如业务上的固定搭配和业务关键词等,或者可以将某个词汇放在不同句式的句子中,高频出现。**影响训练效果的关键因素 为"文本出现的频率"和"上下文的句意理解"等**。无需重复提交大量文本,少量关键文本即可有训练效果。

进入EasyDL语音识别

输入用户名及密码,点击"登录",进入EasyDL语音识别。可以看到整体训练流程,点击创建模型可以直接进行模型创建,点击模型中心可以进入 到模型列表页面。

<i>-داذا</i> ک	asyDL	产品介绍	操作平台	应用案例	使用文档						f58K1G9x5hG	
总览												
			•		0			٠			•	
		创建模型			Ŷ		● 训练模型			上线模型		
	1.创建模型,	选择训练的语音识	只别接口	2.上传测词	【音频和标注文本评估基础模型	3.上传业务词》	C或长文本自动训练棋		4.上线模	型,语音识别接口配置	参数使用	
					创建模型	模型中心						

整体训练流程将按照目录栏的顺序依次操作即可。

模型中心
我的模型
创建模型
训练模型
上线模型
模型调用

下面将详细介绍每一步的操作方式和注意事项。若遇到的问题在此文档没有找到答案,可以加入官方QQ群(群号:686267521)咨询群管。

创建模型

在导航栏【模型中心】-【我的模型】页中可以点击【创建模型】按钮;也可以直接点击左侧导航菜单中的【创建模型】进入创建模型步骤。目前 一个账号下支持创建10个模型,模型可删除。

在创建模型步骤中,需要进行"基础信息填写""上传测试集""选择基础模型"三个环节完成创建。

测试集的作用为通过上传音频和正确的标注文本评估基础模型的识别率,根据基础模型识别率选择最合适的基础模型进行训练。等模型训练后系 统自动使用该测试集评估得到训练后模型的识别率,可以直观的查看训练提升效果。

ど桨EasyD	L 产品介绍	操作平台	应用案例 使用文档	8	f58K1G9x5hG
		7 7 相	吾音识别模型 语音识别接口进行语言模型 传业务音频和标注文本用于 传文本语和对基础模型进行 送	操作文档 数学视频 提交工单 出的训练。有效提升业务场景下语音识别准编率5-25% 于评估基础模型及训练后模型准确率。别注集建议取自业务真实音频,测试集越丰富评估结果越 于训练优化,训练集建议包含业务场景的常用语,以及测试中识别有误的词句指配,且与测试集6	- 敬起へ 密現 力容
1	模型中心	创建模型			
	我的模型				
	创建模型		(1) 基础信息	(2) 上性海滨菜 (3) 选择基础模型	
	训练模型	接口类型:	请选择接口类型	*	
	上线模型	道刑女内.	(市台) (岡田) 化平		
	模型调用	快迎石桥:	调制八模型名称		
		公司/个人	🔾 公司 🥥 个人		
		所属行业	教育培训	✓	
		应用场景	智能客服	~	
		应用设备	手机	~	
		功能描述:	描述不得超过200个字	符	

- 1. 基础信息:包括接口类型、模型名称、公司/个人、所属行业、应用场景、应用设备、功能描述、邮箱地址、联系方式
 - 接口类型:包括短语音识别(支持16K采样率音频)、实时语音识别(支持16K采样率音频)和呼叫中心场景(支持8K采样率音频)3 种,用户可以基于应用场景和音频采样率来进行选择。
 - 模型名称:用户可自行填写模型名称,可支持中文、英文、数字、下划线.+#*()^-
 - 公司/个人:模型归属企业则需要填写企业名称,归属个人则不需填写
 - 所属行业:企业业务或个人应用所属的行业信息
 - 应用场景:语音识别模型应用落地的业务场景
 - 应用设备:业务中使用语音技术的录音设备终端
 - 功能描述:描述模型应用的场景,有助于上线审核哦
 - 邮箱地址:填写联系人的邮箱地址,用于模型上线等信息的通知
 - 联系方式:第一个模型需要用户填写联系方式,后面的模型系统自动复制第一个模型的联系方式(可修改)其中,公司/个人、所属行业、应用场景、应用设备、功能描述、邮箱地址、联系方式在第一个模型中的填写信息会重复使用,后面创建的模型不用重复填写,但可修改信息
- 2. 上传测试集:包括填写测试集名称、上传语音文件、上传标注文件
 - 上传测试集:用户可自行填写测试集名称,可支持中文、英文、数字、下划线.+#*()^-
 - 上传语音文件:上传音频压缩zip文件(请将所有音频文件直接压缩,请勿将音频存放在文件夹内再压缩),格式要求:
 - 16k 16bit单声道pcm/wav文件
 - 8k 16bit 单声道pcm/wav文件(客服场景);
 - 音频文件名请不要包含中文、特殊符号、空格等字符;
 - 所有音频需直接打包压缩为zip文件格式后上传,zip大小不超过100M,解压后单个音频大小不超过150M
 - 上传标注文件:上传音频的标注文本txt文件,格式要求:
 - 标注文件内容应与音频文件相对应的内容一致(单条音频对应文本长度不超过5000字);
 - 标注文件格式应为txt格式,GBK编码;

• 标注txt文本中,由音频名称、标注内容两部分构成,用"tab"区隔,带后缀或不带后缀均可;

上传完语音文件及标注文件,点击【开始评估】,后台进入评估状态,此时弹窗提示评估完毕时间,并自动跳转回【我的模型】。**一个账号只能** 同时评估一个模型。待模型评估完毕后通过【我的模型】可以点击进入"选择基础模型"

提示	×
预计2019-03-20 13:00:41完成预估,数据在后 3秒后,弹窗自动关闭	台持续评估,您可在列表中查看状态
	返回模型列表

3. 选择基础模型:系统根据基础模型的识别率自动推荐适合训练的基础模型,基础模型识别率超过50%才可选择进行训练。

相	莫型中心	创建模型		
	我的模型			
	创建模型		 ① 基础信息 	
	训练模型	评估结果:	推荐使用"输入法模型"进行训练,该模型识别率为98.14% 宣誓评估详情	
	上线模型			
	模型调用		输入法模型98.14% 推荐	
			报索模型96.65%	
			开始训练	

若基础模型识别率未达到50%,请检查语音文件和标注文件内容是否匹配,若不匹配,训练结果无意义。若检查标注文件无误后识别率仍旧过低,可以加入官方QQ群进行咨询:686267521

- 短语音识别产品类型中目前支持对短语音识别极速版进行训练
- 实时语音识别产品类型中目前支持对实时语音识别的中文普通话模型进行训练
- 呼叫中心产品类型中目前支持对呼叫中心语音解决方案进行训练

选择基础模型后点击【开始训练】即可在该模型上进行模型训练。

点击"查看评估详情"可以查看测试集在基础模型上的具体识别结果,评估详情包括:字准率,句准率,插入错误,删除错误,替换错误5个指标, 以及在该测试集上的具体识别结果与标注结果的对比,根据识别错误信息可以更加精准地准备训练文本。

- ビ菜 EasyDL 产品介绍	操作平台 应	可用案例 使用文档			O fi	58K1G9x5hG
	西 7 7 7 7	音识别模型 操作文档 音识别接口进行语言模型的训练。 4 业务音频和标注文本用于评估基础相 文本语科对基础模型进行训练优化。	1 数学视频 提交工单 有效提升业务场景下语音识别准确率5 增起及训练后模型准确率。测试集建设 训练集建议包含业务场景的常用语。	-25% 观自业务真实音频,测试集延丰富评估结果越客观 以及测试中识别有误的词句搭配,且与测试集内容	收起へ	
模型中心	训练模型					
我的模型	选择模型:	abc_reading 🗸 🗸	查看模型评估详情	常见问题		
创建模型	选择版本:	基础模型 🗸 🗸		1、训练语料的格式要求是什么	~	
加1564至 上线模型 模型调用	训练方式:	可选择基础模型动模型训练后版本进行 以下训练方式,您可以选择一种上传交 上传热词 格式要求[2]	7送代训练 "件,也可以全部上传			
		上传句篇 格式要求[] 上传句式 格式要求[] 开始训练 日本				

13137118

7-7- CREasyDE / m/m	床下十百 四用来列	使用文档			U ISSNIG
		使型 操作文档 数学规划 提交 行音音模型的训练。有效提升业务场展下语 注文本用于评估基础模型及训练后模型准确 础模型进行训练优化。训练集建议包含业务	工单 音识别准确带5-25% 事。测试集建议取自业务真实音频,测试集 汤量的常用语,以及测试中识别有误的词句	^{线越丰} 富评估结果越客观 收起 1搭配,且与测试集内容	^
模型中心	创建模型 > 查看评估详情				
我的模型					
创建模型					
训练模型	(只利昭来) 字准: 97.03%	蓉换错误率: 1.86	% 插入错误率:	0%	
上线模型	句准: 80%	删除错误率: 1.12	%		
模型调用	【 评估详情			न	·载
	文件名称	识别结果对比			
	1m-lj_1.pcm	文件标注: 附近的照相馆 评估结果: 附近的照相馆			
	1m-lj_10.pcm	文件标注: 我 附 近 有 大 超 市 吗 评估结果: 我 附 近 有 大 超 市 吗			
	1m-li 11.pcm	文件标注: 高上海大学最近的 ፆ	: *		

在"查看评估详情"页点击"返回上一步"或"创建模型"可返回选择基础模型

训练模型

可以在【创建模型】-"选择基础模型"页点击【开始训练】按钮进入【训练模型】

也可以在【我的模型】列表页选择已创建完成的模型点击操作栏中的"开始训练"进入【训练模型】;

也可以直接在左侧导航栏中点击【训练模型】,进入【训练模型】

在训练模型步骤中,选择需要训练的模型,并上传训练文本。目前有两种训练方式可以选择,可以上传热词,或者是长段文本,也可以两种均上 传进行训练。

梼	極中心	训练模型					
	我的模型	选择模型:	百度AI	~	查看模型评估详情	常见问题	
ř.	创建模型	选择版本:	基础模型	~		1、训练语料的格式要求是什么	
I.	训练模型		可选择基础模型或	花模型训练后版本 进	行迭代训练		
	模型调用	训练方式:	以下训练方式,	您可以选择一种上	传文件,也可以全部上传		
			上传热词	格式要求?			
			上传句篇	格式要求?			
			开始训练				

热词文本格式要求:热词训练支持上传热词txt文件进行训练,每个词之间需要换行,txt格式要求gbk编码,大小不超过5M

句篇文本格式要求:句篇训练支持上传多行单句或一整段篇章(一段文字且需要符号)txt文件进行训练,txt格式要求gbk编码,大小不超过5M

建议您上传与您所需模型内容相关度较高的文本或关键词,以便最大程度提高您的模型识别率

上传训练文本成功之后点击【开始训练】,后台进入模型训练状态,此时弹窗提示评估完毕时间,并自动跳转回【我的模型】。**一个账号下同时 只能训练一个模型。**待模型训练完毕后生成新的模型版本,在【我的模型】列表页可以查看模型训练结果。

提示		:	×
预计2019.12.20 13: 30完成训练, 5秒后,弹窗自动关闭	数据在后台持续评估,	您可在列表中查看状态	
		返回模型列表	

在【我的模型】列表,

模型ID	模型名称	当前版本	训练状态	基础模型准确率及模型效果	操作
1019	qwert	∨1	训练完成	输入法模型:98.14% 当前版本:99.26% 训练结果详情	历史版本 申请上线 迭代训练 下载 删除

可以查看基础模型的识别率,和当前版本的识别率,了解训练提升效果

- 训练结果详情:可以查看训练后模型在测试集上的识别详情,包括:字准率,句准率,插入错误,删除错误,替换错误5个指标,以及在测试 集上的具体识别详情。可以进行操作
 - 历史版本:可以查看历史训练的所有记录并进行操作
 - 申请上线:对当前模型训练结果较为满意,可以点击申请上线,跳转至上线模型步骤
 - 迭代训练:当前模型训练结果不满意,可以在当前版本基础上或者基础模型上继续添加新的训练语料,进行迭代训练获得新的模型版本
 - 下载:可以下载评估模型上传的测试集和训练模型的训练集
 - 删除:可以删除整个模型(包括所有历史版本),删除后不能恢复

上线模型

可以在【我的模型】选择要上线的模型,在操作栏点击"申请上线"

或者在左侧导航栏中点击【上线模型】,选择要上线的模型和版本进行上线(只有模型训练成功生成版本号才可上线)

模型中心	我的模型 > 历	史版本									
我的模型	■历史版本	历史版本									
创建模型	当前版本	训练状态	基础模型准确率	当前版本准确率	操作						
训练模型	V4	训练完成	87.84%	92.37%	申请上线 送代训练 下载 训练结果详情						
上线模型 横型调用	V3	训练完成	87.84%	92.37%	申请上线 送代训练 下载 训练结果详情						
	V2	训练完成	87.84%	92.46%	申请上线 迭代训练 下载 训练结果详情						
	V1	训练完成	87.84%	92.55%	申请上线 迭代训练 下载 训练结果详情						

一个账号下最多只能上线3个模型。申请上线后需要后台管理员进行审核,1-3天内会有审核结果,可在【我的模型】中查看审核状态。若对审核 过程有任何问题可以加入官方QQ群(群号:686267521)咨询群管。

• 审核中:可以查看历史版本训练情况;可以取消申请,取消后方可继续训练

2529	模型名称1	V2	上线审核中 ?	搜索模型: 75.66% 当前版本: 75.66% 训练结果详情	历史版本 取消甲	申请	
审核失败	: 问号可查看)	审核失败原	因;可以查看」	历史版本训练情况; 可	J以迭代训练或	重新训练;可以重新	间请上约
2530 模型	型名称模型名 V2	上线审核失	输入法模型: 75.6 2 当前版本: 75.66 说练结果详情	36% 历史版本 申请上线 选 	10018 1205018 T 42 BI		
审核通过	:审核通过则	后台自动上	线,上线时间,	需要1-3天,上线过程	中模型不可以做	故任何操作	
2580	模型名称3	V2	上线中	输入法模型: 75.66% 当前版本: 75.66% 训练结果详情	。正在上线	线中	
上线完成	: 上线完成的	模型可以正	式调用				
252925299	模型名称1	V2	已上线	输入法模型: 75.66% 当前版本: 75.66% 训练结果 详情	历史版本 SDK下载	制除	

模型调用

上线通过的模型,在【我的模型】可以点击"模型调用",查看如何使用模型

也可以在左侧导航栏中点击【模型调用】

中文普通话模型:87.18% V1 已上线 实时语音识别 当前版本:90.67% 历史版本 模型调用 下载 删除 训练结果详情

模型中心	模型调用									
我的模型	选择模型:	实时来一个	~							
创建模型										
训练模型		第一步: 创建语言 第二步: 获取专用	皆技术应用(若已创建 属模型参数 模型ID: 4	可直接使用),获取鉴权参 1690 基础模型pid: 1537	数AppID, API Key, Sec	ret Key,立即创建				
上线模型		第三步:根据业务情况,选择合适的调用方式,配置鉴权参数和专属模型参数即可使用。若您已经使用语音识别的服务,只需要在您的接口中补充模型参 影响回求职调味后的识别效率								
模型调用		该基础模型支持在以下产品中使用:实时语音识别,该产品目前为邀测状态,请点击下方合作咨询进行业务开通								
		产品米刑	其政權刑	调田方式	这 权余物	去属描刑参数	18.00			
)叩天室	2011年1月1日	99/H1/J1C	1111A 397 RA	5 10 10 10 19 19 19 19 19 19 19 19 19 19 19 19 19	1170			
				WebSocket API	API Key, Secret Key	lm_id=4690, dev_pid=1537	技术文档			
		实时语音识别	中文普通话模型	Android SDK	AppID, API Key, Secret Key	LM_ID=4690, PID=1537	SDK下载 技术文档			
				IOS SDK			SDK下载 技术文档			
				Linux SDK			SDK下载 技术文档			

选择您需要上线的模型 (训练完成的模型才可申请上线)

短语音识别:

step1:创建语音技术应用(若已创建可直接使用),获取鉴权参数AppID, API Key, Secret Key。立即创建

step2:获取专属模型参数 模型ID: xxxx 基础模型pid: xxxx

step3:配置鉴权参数和专属模型参数即可使用

短语音识别极速版支持API方式调用,具体使用方法详见技术文档

实时语音识别:

step1:创建语音技术应用(若已创建可直接使用),获取鉴权参数AppID, API Key, Secret Key。立即创建

step2:获取专属模型参数 模型ID: xxxx 基础模型pid: xxxx

step3:根据业务情况,选择合适的调用方式,配置**鉴权参数**和**专属模型参数**即可使用

实时语音识别支持Websocket API,Android、iOS、Linux SDK方式调用 该接口目前处于邀测阶段,权限开通请点击合作咨询或者加入官方QQ 群:588369236 获取技术文档和demo下载地址。

呼叫中心模型-V1:

- step1:创建智能呼叫中心应用(若已创建可直接使用),获取鉴权参数AppID,API Key,Secret Key。立即创建
- step2:获取专属模型参数 模型ID: xxxx

step3:下载SDK,配置**鉴权参数**和**专属模型参数**即可使用

呼叫中心模型-V1支持C++ SDK、JAVA SDK、MRCP server三种调用方式,下载百度智能呼叫中心SDK即可使用。 立即下载 具体使用方法详 见技术文档

呼叫中心模型-V2:

step1:创建语音技术应用(若已创建可直接使用),获取鉴权参数AppID,API Key,Secret Key。立即创建

step2:获取专属模型参数 模型ID: xxxx

step3:下载SDK,配置鉴权参数和专属模型参数即可使用

呼叫中心模型-V2支持MRCP server调用方式,具体使用方法详见技术文档

语音自训练平台公测期间,为了帮助客户验证线上效果,每个账号支持上线3个模型,每个账号累计有50000次免费调用量。正式商用后,免费 资源可能会有所调整。 公测期QPS限额:个人未认证账户2QPS;个人认证账户3QPS;企业认证账户5QPS。 公测期间如需更多资源,欢迎商务 合作咨询。

声音分类

声音分类整体说明

心 什么是声音分类模型

声音分类是指可以定制识别出当前音频是哪种声音,或者是什么状态/场景的声音。

EasyDL声音分类可以定制的模型更多可以区分出不同物种发出的声音,如果希望定制声纹识别模型(如区分出当前音频是谁的声音),目前用 EasyDL声音分类暂时无法解决。

目前声音分类使用EasyDL支持对最长15s左右的音频进行处理,在正式使用EasyDL声音分类模型之前,需要将已有的数据进行分段处理。



⊙ 声音分类的典型应用场景

- 安防监控:定制识别不同的异常或正常的声音,进而用于突发状况预警。比如监控在工业生产场景中监控是否出现了异常噪音,从而辅助人工 测试的时候判断是否出现bug。
- 科学研究:定制识别同一物种的不同个体的声音、或者不同物种的声音,协助野外作业研究。比如动物研究机构从野外采集的声音,借助于 EasyDL声音分类模型,判断当前音频属于什么物种。
- 其他:尽情脑洞大开,训练你希望实现的声音分类模型。

心 定制声音分类模型的整体流程

定制声音分类模型基本流程如下图所示,全程可视化简易操作,在数据已经准备好的情况下,最快几分钟即可获得定制模型。



分析业务需求

这里我们详细介绍下,在使用EasyDL平台之前首先需要分析业务需求。这一步主要将实际业务需求转换为模型设计,在声音分类场景中,首先需 要明确的问题为**业务场景可能出现的全部声音类型有哪些?**,这里很多企业开发者往往会主要关注**业务场景中需要重点识别出的异常声音分类**, 而忽略了**正常的声音也是一种分类**。

以**某服务商接到项目,需要判断出小区附近是否存在较大噪音**为例,综合考虑小区附近可能有的声音类型,在这个场景需要定制声音分类模型能 有效区分**正常无噪音、正常噪音如救护车、警车声音、异常噪音,如汽车大声按喇叭等三**类状态。那么在后续的准备数据阶段,也需要能有效准 备这三类声音。
数据准备

心 创建数据集

在训练之前需要在数据中心【创建数据集】

イン・ビ業 EasyDL 产品介	1名 操作平台	应用案例 使用3	文档				社区 飞桨官网	AI主站	658K1G9x5hG
声音分类模型 →三	我的数据总览								提交工单
88 总览									
山 模型中心	EasyData智能算	取据服务平台已上线,使用	JEasyData可享受包括	多人标注、数据清洗、数据采集等完整	國務總服务 立即前往				
我的模型	创建数据集							输入	.数据集名称或D Q
创建模型	11111 52 858							54096-+ 00-4	carrie d'iane
训练模型							L.	- 00 -	
2000 (1000) 发布模型	版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	操作		
回 EasyData数据服务	V1 ⊖	309243	0	• 已完成	音頻分类	0% (0/0)	多人标注	导入删除	
数据总览	蜂鸣器大数据	区 数据集组ID: 25548	19				R	新婚版本 铝金	全部版本 白 影除
标签组管理	版本	教掘集の	数据量	最近导入状态	板注慮型	标注状态	操作		
住魏称注 曰 公有云服务	_								
在线服务	V1 Θ	274020	885	 已完成 	音频分类	100% (885/885)	宣看多	人标注 导入	标注 …
盟 EasyEdge本地部署	蜂鸣器 区 製	友据集组ⅠD:255374						新增版本 铝合	2部版本 白 删除
纯高线服务	版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	操作		
	V1 💬	273870	171	• 已完成 ③	音频分类	100% (171/171)	直看 多	人标注 导入	标注 …
	debug 🗹 👮	数据集组ID:207484					G	所增版本 品名	2部版本 由删除
	拓大	864240 m	彩燈器	黑灰色为中央	经注册制	经注仲成	45.85		
声音分类模型 🐳	我的数据	总览 > 创建数据集							
88 总览									
① 模型中心	数据集名	▶ 限制50个汉章	字内				常见问题		
我的模型	数据类型	音频					1、如何设计分类		~
创建模型	数据编版:	本 V1					2、音频的具体格;	式要求	~
训练模型							3、批量上传的压结	宿包格式要求	~
校验模型	标注类型	音频分类	(letti				4、BOS目录导入	各式要求	~
发布模型		16	24				5、分享链接导入村	各式要求	~
☑ EasyData数据服务		1 Las							
数据总览	标注模板	○ 短音频单标	ī资						
标签组管理		完成							
在线标注									
目 公有云服务									
在线服务									
器 EasyEdge本地部署									
纯离线服务									

设计分类

- 每个标签就是对这个音频希望识别出的全部结果。标签的上限为1000种。
- 标签名由数字、中英文、中/下划线组成,长度上限256字符。

音频的具体格式要求

- 训练集音频需要和实际场景要识别的音频环境一致,举例:如果实际场景要识别的音频都是手机摄录的,那训练的音频也需要同样的场景获得,而不要采用网上随便下载的音频。
- 每个标签的音频需要覆盖实际场景里面的可能性,如不同环境下,训练集覆盖的场景越多,模型的泛化能力越强。
- 如果需要寻求第三方数据采集团队协助数据采集,可以加入官方QQ群(群号:679517246)联系群管咨询了解。
- 音频支持mp3, m4a, wav格式,单个音频大小在4M内且时长小于15s。

◎ 上传数据集

上传数据要求说明

这里我们对上传数据的要求不仅是格式上的要求,更重要的是介绍怎样的数据可以更有效提升模型效果

设计分类

首先想好分类如何设计,每个分类为你希望识别出的一种结果,如要识别猫狗的叫声,则可以以"猫"、"狗"等分别作为一个分类;如果安防监控通 过声音判断是否出现异常状态,可以以"正常""不异常"设计为两类,或者"正常""异常原因一"、"异常原因二"、"异常原因三"……设计为多类。

准备数据

基于设计好的分类准备音频数据,每个分类需要准备50个音频文件以上,如果想要较好的效果,建议100个起音频文件,如果某些分类的声音具 有相似性,需要增加更多音频。

音频的基本格式要求: 目前支持音频文件 类型为支持wav,mp3,m4a ,音频文件大小限制在4M以内。一个模型的音频总量限制**10万个**音频文件。

注意1:训练集音频需要和实际场景要识别的音频环境一致,举例:如果实际场景要识别的声音都是手机采集的,那训练的音频文件也需要同 样的场景获得,而不要采用网上随便下载的音频 注意2:考虑实际应用场景可能有的种种可能性,每个分类的音频需要覆盖实际场景里面可能有的可能性,如噪音干扰、多种可能的采集设 备,训练集覆盖的场景越多,模型的泛化能力越强。 注意3:如果需要寻求第三方数据采集团队协助数据采集,请在百度云控制台内提交工单反馈

注息3·如木而女寸水东二刀数ය木朱凶队切动数ය木朱,明住日反厶住何口的提义工平区的

你可能会有的问题:如果训练音频数据无法全部覆盖实际场景要识别的音频,怎么办?

答:本身模型算法会有一定的泛化能力,尽可能覆盖即可。

导入未标注数据

本地数据

1	旦	λ	쐐	据
	4	~	玖	5/古

数据标注状态	● 无标注信息	○ 有标注信息	
导入方式	本地导入		\sim
上传压缩包	<u> </u> 上传压缩包		

已有数据集支持选择百度云BOS导入、分享链接导入、平台已有数据集导入;支持选择线上已有的数据集,包括其他语音类模型的数据集

┃导入数据

数据标注状态	● 无标注信息 ○ 有标注信息	
导入方式	请选择	^
	本地导入	
	BOS目录导入	
	分享链接导入	
	平台已有数据集	

导入已标注数据

本地数据

マハ奶油	

数据标注状态	○ 无标注信息	● 有标注信息		
导入方式	本地导入		\sim	请选择 へ
	确认并返回			上传压缩包
				API导入

已有数据集支持选择百度云BOS导入、分享链接导入、平台已有数据集导入;支持选择线上已有的数据集,包括其他语音类模型的数据集

┃导入数据

数据标注状态	○ 无标注信息	● 有标注信息
导入方式	请选择	
	本地导入	

BOS目录导	λ	
分享链接导	Л	
平台已有数	据集	

^の数据集管理API

本文档主要说明当您线下已有大量的已经完成分类整理的音频数据,如何通过调用API完成音频数据的便捷上传和管理。

^

数据集创建API

接口描述

该接口可用于创建数据集。

接口鉴权

同发布模型后获取的API鉴权方式:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/create

URL参数:

参数 值 access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset_na me	是	string	数据集名称,长度不超过20个utf-8字符

若上传声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
dataset_id	否	number	创建的数据集ID

查看数据集列表API

接口描述

该接口可用于查看数据集列表。返回数据集的名称、类型、状态等信息。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/list

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
start	否	number	起始序号,默认为0
num	否	number	数量,默认20,最多100

若查看声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明		
log_id	是	number	唯一的log id,用于问题定位		
total_num	否	number	数据集总数		
results	否	array(object)	数据集列表		
+dataset_id	否	number	数据集ID		
+dataset_name	否	string	数据集名称		
+type	否	string	数据集类型		
+status	否	string	数据集状态		
+special_status	否	string	数据集特殊状态,包括shared、smart和空值,分别表示共享中、智能标注中、非特殊状态		

查看分类 (标签)列表API

接口描述

该接口可用于查看分类(标签)。返回分类(标签)的名称、包含数据量等信息。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/list

URL参数:

参数 值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID
start	否	number	起始序号,默认0
num	否	number	数量,默认20,最多100

若查看声音分类的全部分类,在type参数应传「SOUND_CLASSIFICATION」

返回说明

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	标签总数
results	否	array(object)	标签列表
+label_id	否	string	标签/分类ID
+label_name	否	string	标签/分类名称
+entity_count	否	number	图片/声音/文本数量

添加数据API

接口描述

该接口可用于在指定数据集添加数据。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/addentity

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID
append Label	否	boolea n	确定添加标签/分类的行为:追加(true)、替换(false)。默认为追加(true)。
entity_c ontent	是	string	type为 IMAGE_CLASSIFICATION/OBJECT_DETECTION/IMAGE_SEGMENTATION/SOUND_CLASSIFICATION时,填入图 片/声音的base64编码;type为TEXT_CLASSIFICATION时,填入utf-8编码的文本。内容限制为:图像分类base64前 10M;物体检测base64前10M;图像分割base64前10M;声音分类base64前4M,声音时长1~15秒;文本分类 10000个汉字
entity_n ame	是	string	文件名
labels	是	array(o bject)	标签/分类数据
+label_ name	是	string	标签/分类名称(由数字、字母、中划线、下划线组成),长度限制20B
+left	否	number	物体检测时需给出,标注框左上角到图片左边界的距离(像素)
+top	否	number	物体检测时需给出,标注框左上角到图片上边界的距离(像素)
+width	否	number	物体检测时需给出,标注框的宽度(像素)
+height	否	number	物体检测时需给出,标注框的高度(像素)

若上传声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id ,用于问题定位

数据集删除API

接口描述

该接口可用于删除数据集。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/delete

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取

参数	值
а т	

Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset _id	是	number	数据集ID

若刪除声音分类数据集,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

分类 (标签) 删除API

接口描述

该接口可用于删除分类(标签)。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/delete

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、文本分类
dataset_ id	是	number	数据集ID
label_na me	是	string	标签/分类名称

若删除声音分类的子类,在type参数应传「SOUND_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
   "error_code": 110,
   "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求 ,如果持续出现此类错误,请在百度云控制台内提交工单反馈
2	Service temporarily unavailable	服务暂不可用,请再次请求,如果持续出现此类错误,请在百度云控制台内提交工单反馈
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量 阶梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶 梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
406000	internal server error	服务错误
406001	param[xx] invalid	参数xx不合法,请检查相关参数
406002	dataset not exist	数据集不存在
406003	dataset already exists	数据集已存在
406004	dataset can not be modified temporarily	数据集暂不可修改
406005	label not exist	标签/分类不存在
406006	no permission to modify the dataset	没有修改数据集的权限
406007	dataset cannot be modified while smart annotation is running	智能标注期间不可修改数据集
406008	quota exceeded	配额超限

模型训练

心 声音分类训练操作说明

数据提交后,可以在导航中找到【训练模型】,按以下步骤操作,启动模型训练:

注意:启动训练前请确保数据已经标注完成,否则无法启动训练

<i>ーージ^ン达</i> 楽EasyDL	产品	介绍 操作平台	应用案例	使用文档		社	☑ 飞桨官网	AI主站	0 f58K10	39x5hG
声音分类模型	Ē	训练模型					操作	文档 常见问题	教学视频	提交工单
器 总览										
凸 模型中心		选择模型	蜂鸣器测试xj	~						
我的模型		训练配置								
创建模型		部署方式	公有云部署	EasyEdge本地部署	如何选择部署方式?					
训练模型		选择设备	○ 服务器	○ 通用小型设备						
校验模型		2年18月11日	B B Z 2 1	大乐物:十分台口:二百百百分						
发布模型		2014种723	U MAN	宣有异本性能及迫略使件						
☑ EasyData数据服务		高级设置	同步支持公	有云部署 ②						
数据总览		添加数据								
标签组管理		添加数据集	+ 请选择							
在线标注			数据集		版本	分类数量		操作		
目 公有云服务			蜂鸣器大数排	居	V1	2		移除		
在线服务										
盟 EasyEdge本地部署		开始训练								
纯离线服务										

step1:选择模型

选择此次训练的模型。

step2:训练配置

部署方式

● 可选择「公有云部署」、「EasyEdge本地部署」

选择设备

• 如果您选择了「EasyEdge本地部署」,请根据实际部署设备选择设备。

选择算法

• 当前语音分类仅支持默认算法,点击可查看算法性能及适配硬件。

高级训练配置

• 如果您选择了「EasyEdge本地部署」,选择【同步支持公有云部署】,训练完成后模型可部署到百度云上进行使用。

step3:添加数据

- 先选择数据集,再按分类选择数据集里的音频,可从多个数据集选择音频
- 声音分类模型至少需要选择2个及以上分类

step4:训练模型

点击「开始训练」,训练模型。

- 训练时间与数据量大小有关,1000个音频文件大约可以在30min内训练完成
- 模型训练过程中,可以设置训练完成的短信提醒并离开页面,在模型训练完毕即可收到短信通知。

平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

优惠政策:

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有声音分类操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。

注意:如果遇到模型超过2天停留在训练中的状态,或者遇到训练失败的情况,请在百度云控制台内提交工单反馈

心 声音分类模型效果评估报告说明

模型评估报告内容说明

模型训练完成后我们可以在模型列表中看到模型效果及查看模型评估报告的入口。进入模型评估报告页面,我们可以看到整体报告内容中包含以 下几个区域内容:

模型基本信息

在这个部分可以选择应用类型(声音分类目前仅支持云服务)、训练版本、相应版本提交的音频数量、相应版本提交的分类数量。

整体评估

在这个部分可以看到模型训练整体的情况说明,包括基本结论、准确率、F1-score、精确率、召回率。这部分模型效果的结果内容是基于训练数 据集,随机抽出部分数据不参与训练,仅参与模型效果评估计算得来。所以**当数据量较少时(如音频数量低于100个),参与评估的数据可能不 超过30个音频,这样得出的模型评估报告效果仅供参考,无法完全准确体现模型效果。**

注意:若想要更充分了解模型效果情况,建议发布模型为API后,通过调用接口测试批量音频数据获取更准确的模型效果。

详细评估

在这个部分可以看到上述训练效果背后的原始评估数据。以及不同top结果的准确率效果情况,下面为相关名词解释。



准确率

准确率含义为正确分类的样本数与总样本数之比,这里指的总样本是指**从总训练数据中随机抽取部分数据参与模型评估的总样本**,在上文截图中,参与训练的音频数200个,实际参与评估的音频为下面详细评估预测表现表格的数据总和,即50个。那么准确率为参与评估的正确数量 46/50,结果为92.0%

F1-Score

F1-score是指对某类别而言为精确率和召回率的调和平均数,此处为各类别F1-score的平均数。 就某类而言,精确率和召回率体现了该分类的精确率及召回率的平衡情况:

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

精确率

精确率是对某类别而言为正确预测为该类别的样本数与预测为该类别的总样本数之比,此处为各类别精确率的平均数。如果精确率比较低,有一 定可能原因为不同类别效果存在差异所致,请检查下不同类型样本量是否均衡。

召回率

召回率是指对某类别而言为正确预测为该类别的样本数与该类别的总样本数之比,此处为各类别召回率的平均数。

top1、top2、top…5

是指对于每一个评估的音频文件,模型会给根据置信度高低,依次给出top1-top5的识别结果,其中top1置信度最高,top5的置信度最低。那么 top1的准确率值是指对于评估标准为"top1结果识别为正确时,判定为正确"给出准确率。top2准确率值是指对于评估标准为"top1或者top2只要有 一个命中正确的结果,即判定为正确"给出的准确率。……以此类推。

如何解读模型效果

在看模型评估报告结果中,首要需要关注下详细评估中的预测表现,这里可以看到所有评估报告的数据是基于什么量级进行计算的。当整体参与 评估的数量较少时,所有数值可能无法真实反映模型效果。

评估样本具体		
预测主知	正确数量: 217	
顶侧衣环	错误数量:3	

在查看模型评估结果可能需要思考在当前业务场景精确率与召回率更关注哪个指标,是更希望减少误识别,还是更希望减少误召回。前者更需要 关注召回率的指标,后者更需要关注精确率的指标。同时F1-SCORE可以有效关注精确率和召回率的平衡情况,对于希望召回与识别效果兼具的 场景,F1-Score越接近1效果越好。

心 声音分类训练时长说明

训练时长与数据量、所选算法紧密相关。

目前**声音分类的训练时长主要影响因素为数据量**,以下为内部测试的数据量与训练时长的对应关系,供参考:

 数据量
 训练时长

 数十个音频
 60min左右

 数百个音频
 90min左右

 数千个音频
 120min左右

 数万个音频
 150min以上

心 如何提升模型效果

如何充分测试模型效果

模型校验

在查看模型评估报告基础上,首先使用模型校验功能测试**未参与过训练**的音频数据进行模型训练,在这一步尽量上传不同类别的数据充分测试, 并在测试过程中线下记录识别错误的音频。在测试过程中需要关注以下内容:

1. 不同分类的准确率是否存在明显差异

2. 识别错误的音频是否存在一些共性?比如设备相似、音调相似、环境相似等等

3. 识别错误的音频人耳是否能明显分辨

<i>۴ ک</i> ¥EasyDL خ ار ب	计招 操作平台 应用案例 使用文档 社区	区 飞桨官网 AI主站 f58K1G9x5hG
声音分类模型 🛛 🖃	校验模型	操作文档 常见问题 教学视频 提交工单
田 总览 山 模型中心 我的模型	选择模型 13.5声音验证 > 部署方式 公有云API > 选择版本 V2 当前模型准确事 100.00% 评估报告 识	✓ >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
创建模型	3Q	整磷值 ○ 0.03 预测分类 置信度 > 3.00% ∨
训练模型 校验模型		
发布模型 ☑ EasyData数据服务	R	
数据总览	《法法》加音频 或直接指线文件互此处	没有满足条件的识别结果
在线标注	1.支持音频格式: mp3, m4a, wav, 大小限制在4M内, 2.单个音频长度在15a以内	
公有云服务 在线服务		
記 EasyEdge本地部署 如离线服务		
2 A 1999 (1999) (1999 (1999 (1999 (1999 (1999 (1999 (1999 (1999 (1999 (1999 (1		申请上线

发布模型为API通过调用接口充分测试

将声音分类模型发布为API后,调用接口进行批量测试,在测试过程中同样重点关注上述三点内容。

如何提升模型效果

在充分测试模型效果基础上,如果发现模型效果欠佳,建议根据以下顺序分析并提升模型效果。

检查并优化训练数据

首先**检查目前欠佳的模型是否存在训练数据过少**的情况,建议每个类别的音频量不少于200个,如果低于这个量级建议扩充。

在扩充数据中需要一并**检查不同类别的数据量是否均衡**,建议**不同分类的数据量级相同,并尽量接近**,如果有的类别数据量很高,有的类别数据 量较低,那么可能会存在不同类别的准确率不同,同时低准确率的分类会拉低整体模型效果。

另外需要检查测试模型的音频数据与训练数据采集来源是否一致,如果设备不一致、或者采集的环境不一致(录音室环境及实际生产环境的差异),那么很可能会存在模型效果不错但实际测试效果较差的情况。针对这种情况建议重新调整训练集,将训练数据与实际业务场景数据尽可能 一致

最后也请确认**识别错误的音频人耳是否能清晰分辨**,模型效果很难超越人耳的识别精度效果,这种情况,请在百度云控制台内提交工单反馈。

联系EasyDL团队提升模型效果

在完成上述检查并优化训练数据的工作后,仍然发现了显著的模型效果低的情况,请在百度云控制台内提交工单反馈。

模型发布

心 声音分类模型发布整体说明

训练完成后,可将模型部署在公有云服务器,发布成为在线的restful API接口,参考示例文档通过HTTP请求的方式进行调用。或将模型通过服务 器端SDK部署在私有服务器,

公有云API

将模型发布为API后,将获得一部分免费调用次数,超出免费额度将根据调用次数进行收费。

私有服务器部署

可将训练完成的模型部署在私有CPU/GPU服务器上,在内网/无网环境下使用模型,确保数据隐私。

心 公有云部署

心 声音分类模型如何发布为API

声音分类模型训练完毕后就可以在左侧目录栏中找到【发布模型】,发布模型表单页面需要自定义接口地址后缀及服务名称,即可申请发布。

ビ来EasyDL ^と	产品介绍	操作平台 应用案	例 使用文档			社区 飞桨官网	AI主站	f58K1G9x5hG
声音分类模型	÷	模型列表				操作文	1档 常见问题 著	牧学视频 提交工单
器 总览								
由 模型中心		创建模型						
我的模型		「雪立八米」 终时限302年:	「/ 横形に、170000					
创建模型		【严百万尖】如吻留测现对	四 候至10:172996					加和
训练模型		部署方式	版本	训练状态	服务状态	模型效果	操作	
校验模型						top1准确率: 98.64%	查看版本配置	服条详情
发布模型		本地部署-服务器	V1	 训练完成 	-	top5准确率: 100.00% 完整评估结果	申请发布	
☑ EasyData数据服务							7	
数据总览		【声音分类】13.5声音验证	☑ 模型ID: 172973				吕 训练 🕓 历史	版本
标签组管理								
在线标注		部署方式	版本	训练状态	服务状态	模型效果	操作	
目 公有云服务				Not following with	- 0 -	top1)准确率: 100.00%	查看版本配置	申请发布
在线服务		公有云API	V2	• 训练死战	木友巾	top5准佣率:100.00% 完整评估结果	校验	
認 EasyEdge本地部署								
纯离线服务		【声音分类】西门子西伯乐	斯声音分类_win 🖸	模型ID: 168883			8	训练 亩 删除
			検到会びを成功した工	動使物連た方の動使中心	·* 6424 L4±20145-5649 201	体 推測后 一司以右山从本重推测的	日 年(16)十	

申请发布后,通常的审核周期为T+0,即申请当天可以审核完成,如遇到周末及其他法定假日,将顺延至节后第一天完成审核,如果需要加急、 或者遇到莫名被拒的情况,请在百度云控制台内提交工单反馈。

<i>ナ・</i> ど業EasyDL 产品:	介绍 操作平台 应用案例 使用文档	社区 飞桨官网 AI主站 🚺 158K1G9x5hG
声音分类模型 🛛 🕁	发布模型	操作文档 常见问题 教学视频 提交工单
田 总览 凸 模型中心 我的模型	选择模型 13.5声音皱证 ✓ 邮看方式 公有云邮署 ✓	标准接口规范参考 下载 API 文档 标准接口请求参考说明:
创建模型 训练模型 校验模型	选择版本 V2 ~ V 服务名称 •	字段名称 必须 类型 说明 sound 是 string 常规, base64编码,要 求bbase04编码后大小不 超过4A,发表tmp3, m4A, wavHsT,
发布模型 図 EasyData数据服务	接口地址 * https://aip.baidubce.com/rpc/2.0/ai_custo m/vt/sound_cis/	top_num 否 number 适回分类数量、数认为6 个 标准接口调应字段说明:
数据总览 标签组管理 在线标注	其他要求 若接口无法满足您的需求,请描述希望解决 的问题。500以字以内	字段名称 必須 类型 说明 log_id 是 number 譜一句log id, 用于问题 定位
 公有云服务 在线服务 超 EasyEdge本地部署 	0/560	results 否 anny(object) 分类结果要组 + name 否 string 分类名称 + score 否 number 宽信度
纯离线服务		

模型发布成功后,可以在模型列表跟踪审核状态,在审核通过后,审核发布中预计等待15分钟左右时间,API将完成上线,并在模型列表相应模型位置查看到【服务详情】入口,点击可查看接口地址,调用API请参考API调用文档。

^の声音分类API调用文档

接口描述

声音分类模型完成模型发布后可以获得定制的声音分类API,实现基于定制的声音分类模型,调用API输入15s以内的音频数据,返回自定义的分类结果。

接口鉴权

创建应用

进入EasyDL控制台概览页,在已上线的定制接口下,可以看到已发布上线的声音分类API。在正式调用之前,首先点击**创建应用**,定义应用名称、应用类型、应用描述等信息,完成应用创建。

全局 ~							Q I#	消息 帮助文档	企业组织 财务	Q
EasyDL定制训练平台	应用			用量				请选择时间	段 2019-10-18 - 2019-	10-18 🕅
概览				API	调用量	调用纬	:败	失败率	详细统计	
应用列表	F	建应用:4个		test	0	0		0%	查看	l
监控报表		管理应用			0	0		0%	查看	
技术文档		创建应用		test	0	0		0%	查看	
云服务权限管理			-	商品检测	0	0		0%	宣看	
离线服务管理本地部署包管理	可用服务列表									
售卖已发布服务	已上线的定制接口	数据管理接口	货架拼接接口							
	API	模型ID	模型类型	模型名称	模型版本	状态	调用量限制	QPS限制	付费开通	
	test	10808	声音分类	识别声音-动物	V1	● 免费使用	500次/天免费	不保证并发	开通付费	
	tiantianquan	36773	图像分割	检测甜甜圈	V1	● 免费使用	500次/天免费	不保证并发	免费试用	
	test	18597	图像分类	识别月季和玫 瑰	V1	● 付费使用	500次/天免费 + 超出按量计 费	10	已开通	
	商品检测	13260	物体检测	TC商品检测	V1	● 免费使用	500次/天免费	不保证并发	开通付费	
	声音分类	10789	声音分类	动物声音分类	V3	● 免费使用	500次/天免费	不保证并发	开通付费	
	识别眼睛和嘴 巴	8086	物体检测	识别眼睛和嘴 巴	V3	● 付费使用	500次/天免费 + 超出按量计 费	10	已开通	

应用创建后,在应用详情页获取API Key、Secret Key。

EasyDL定制训练平台	产品服务 / EasyDL定	制训练平台 - 应用引	列表 / 应用详情						
概览	应用详情								
应用列表	编辑 查:	看文档							
监控报表	应用名称		AppID		API Key		Secret Key		
技术文档	动物声音识别。	小程序	17554456		xN9yOF678YGGsrMe44O9B	7g5	****** 显示		
云服务权限管理									1
离线服务管理	可用服务列表	 数据管理接口 	1 後架拼接接口						
本地部署包管理									
售卖已发布服务	模型ID	模型类型	服务名称	状态	请求地址	调用量限制		QPS限制	
	10808	声音分类	test	● 免费使用	https://aip.baidubce.com/rpc/2 .0/ai_custom/v1/sound_cls/ani mal0000	500次/天免费		不保证并发	

请求说明

请求示例

HTTP 方法:POST

请求URL:请在应用详情页面获取接口地址,与发布模型时定义的url一致。

全局 ~					Q	工单	消息	帮助文档	企业组织	财务	Q
EasyDL定制训练平台	产品服务 / EasyDL定制训练平台 - 应	^告 品服务 / EasyDL定制训练平台 - 应用列表 / 应用详情									
概览	应用详情										
应用列表	编辑 查看文档										
监控报表	应用名称	AppID	AppID API Key					Secret Key			
技术文档	动物声音识别小程序	17554456	1456 xN9yOF678YGGsrMe44O		e4409B	7g5		****** 显示			
云服务权限管理 离线服务管理	可用服务列表:										
本地部署包管理	已上线的定制接口 数据管理	發口 货架拼接接口									
售卖已发布服务	模型ID 模型类型	服务名称	状态	请求地址		调用量降	艮制		QPS限制		
	10808 声音分类	test	● 免费使用	https://aip.baidubce.com/ .0/ai_custom/v1/sound_cls 	rpc/2 :/ani	500次/3	天免费		不保证并发		

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body请求示例:

```
EasyDL 语音使用说明
```

300y 请水亦例:

```
{
    "sound": "<base64数据>",
    "top_num": 6
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必 选	类型	可选值 范围	说明
sound	是	string	-	音频,base64编码,要求base64编码后大小不超过4M,支持mp3、m4a、wav格式,单个文件时间长度不 超过15s 注意需要去掉编码头后再进行urlencode。
top_nu m	否	number	-	返回分类数量,默认为6个

请求代码示例

提示一:使用示例代码前,请记得替换其中的示例Token、音频地址或Base64信息。

提示二:部分语言依赖的类或库,请在代码注释中查看下载地址。

PHP代码示例

```
<?php
/**
* 发起http post请求(REST API),并获取REST请求的结果
* @param string $url
* @param string $param
* @return - http response body if succeeds, else false.
*/
function request_post($url = '', $param = '')
{
  if (empty($url) || empty($param)) {
    return false;
  }
  $postUrl = $url;
  $curlPost = $param;
  // 初始化curl
  $curl = curl_init();
  curl_setopt($curl, CURLOPT_URL, $postUrl);
  curl_setopt($curl, CURLOPT_HEADER, 0);
  // 要求结果为字符串且输出到屏幕上
  curl_setopt($curl, CURLOPT_RETURNTRANSFER, 1);
  curl_setopt($curl, CURLOPT_SSL_VERIFYPEER, false);
  // post提交方式
  curl_setopt($curl, CURLOPT_POST, 1);
  curl_setopt($curl, CURLOPT_POSTFIELDS, $curlPost);
  // 运行curl
  $data = curl_exec($curl);
  curl_close($curl);
  return $data;
}
$token = '[调用鉴权接口获取的token]';
$url = '【接口地址】?access_token='.$token;
$bodys = "{\"sound\":\"sfasq35sadvsvqwr5q...\"}"
$res = request_post($url, $bodys);
var_dump($res);
```

```
.....
EasyDL 声音分类 调用模型公有云API Python3实现
import json
import base64
import requests
....
使用 requests 库发送请求
使用 pip(或者 pip3)检查我的 python3 环境是否安装了该库,执行命令
pip freeze | grep requests
若返回值为空,则安装该库
pip install requests
....
**目标音频的本地文件路径,支持mp3/m4a/wav格式**
SOUND_FILEPATH = "【您的测试音频地址,例如./example.mp3】"
**可选的请求参数**
**top_num: 返回的分类数量,不声明的话默认为 6 个**
PARAMS = {"top_num": 3}
**服务详情中的接口地址**
MODEL_API_URL = "【您的API地址】"
**调用 API 需要 ACCESS_TOKEN。若已有 ACCESS_TOKEN 则于下方填入该字符串**
**否则,留空 ACCESS_TOKEN,于下方填入 该模型部署的 API_KEY 以及 SECRET_KEY,会自动申请并显示新 ACCESS_TOKEN**
ACCESS_TOKEN = "【您的ACCESS_TOKEN】"
API_KEY = "【您的API_KEY】"
SECRET_KEY = "【您的SECRET_KEY】"
print("1. 读取目标音频 '{}'".format(SOUND_FILEPATH))
with open(SOUND_FILEPATH, 'rb') as f:
 base64_data = base64.b64encode(f.read())
 base64_str = base64_data.decode('UTF8')
print("将 BASE64 编码后音频的字符串填入 PARAMS 的 'sound' 字段")
PARAMS["sound"] = base64_str
if not ACCESS_TOKEN:
  print("2. ACCESS_TOKEN 为空,调用鉴权接口获取TOKEN")
  auth_url = "https://aip.baidubce.com/oauth/2.0/token?grant_type=client_credentials"
        "&client_id={}&client_secret={}".format(API_KEY, SECRET_KEY)
 auth_resp = requests.get(auth_url)
  auth_resp_json = auth_resp.json()
 ACCESS_TOKEN = auth_resp_json["access_token"]
  print("新 ACCESS_TOKEN: {}".format(ACCESS_TOKEN))
else:
  print("2. 使用已有 ACCESS_TOKEN")
print("3. 向模型接口 'MODEL_API_URL' 发送请求")
request_url = "{}?access_token={}".format(MODEL_API_URL, ACCESS_TOKEN)
response = requests.post(url=request_url, json=PARAMS)
response_json = response.json()
response_str = json.dumps(response_json, indent=4, ensure_ascii=False)
print("结果:\n{}".format(response_str))
```

JAVA代码示例

package com.baidu.ai.aip;

import com.baidu.ai.aip.utils.HttpUtil; import com.baidu.ai.aip.utils.GsonUtils;

import java.util.*;

```
/**
* easydl声音分类
```

*/

```
public class EasydlSoundClassify {
```

/**

```
* 重要提示代码中所需工具类
```

- * FileUtil,Base64Util,HttpUtil,GsonUtils请从
- * https://ai.baidu.com/file/658A35ABAB2D404FBF903F64D47C1F72
- * https://ai.baidu.com/file/C8D81F3301E24D2892968F09AE1AD6E2
- * https://ai.baidu.com/file/544D677F5D4E4F17B4122FBD60DB82B3
- * https://ai.baidu.com/file/470B3ACCA3FE43788B5A963BF0B625F3

```
* 下载
```

```
*/
```

public static String easydlSoundClassify() {

```
// 请求url
String url = "【接口地址】";
```

System.out.println(result);

public static void main(String[] args) {

EasydlSoundClassify.easydlSoundClassify();

return result;
} catch (Exception e) {
 e.printStackTrace();

try {

Map<String, Object> map = new HashMap<>();

map.put("sound", "sfasq35sadvsvqwr5q...");

String param = GsonUtils.toJson(map);

// 注意这里仅为了简化编码每一次请求都去获取access_token,线上环境access_token有过期时间,客户端可自行缓存,过期后重 新获取。 String accessToken = "[调用鉴权接口获取的token]"; String result = HttpUtil.post(url, accessToken, "application/json", param);

```
}
```

}

}

}

return null;

C++代码示例

```
**include <iostream>**
**include <curl/curl.h>**
// libcurl库下载链接:https://curl.haxx.se/download.html
// jsoncpp库下载链接: https://github.com/open-source-parsers/jsoncpp/
const static std::string request_url = "【接口地址】";
static std::string easydlSoundClassify_result;
/**
* curl发送http请求调用的回调函数,回调函数中对返回的json格式的body进行了解析,解析结果储存在全局的静态变量当中
* @param 参数定义见libcurl文档
* @return 返回值定义见libcurl文档
*/
static size_t callback(void *ptr, size_t size, size_t nmemb, void *stream) {
  // 获取到的body存放在ptr中,先将其转换为string格式
  easydlSoundClassify_result = std::string((char *) ptr, size * nmemb);
  return size * nmemb;
}
/**
* easydl声音分类
* @return 调用成功返回0,发生错误返回其他错误码
*/
int easydlSoundClassify(std::string &json_result, const std::string &access_token) {
  std::string url = request_url + "?access_token=" + access_token;
  CURL *curl = NULL;
  CURLcode result_code;
  int is_success;
  curl = curl_easy_init();
  if (curl) {
    curl_easy_setopt(curl, CURLOPT_URL, url.data());
    curl_easy_setopt(curl, CURLOPT_POST, 1);
    curl_slist *headers = NULL;
    headers = curl_slist_append(headers, "Content-Type:application/json;charset=UTF-8");
    curl_easy_setopt(curl, CURLOPT_HTTPHEADER, headers);
    curl_easy_setopt(curl, CURLOPT_POSTFIELDS, "{\"sound\":\"sfasq35sadvsvqwr5q...\"}");
    result_code = curl_easy_perform(curl);
    if (result_code != CURLE_OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n",
            curl_easy_strerror(result_code));
       is_success = 1;
       return is_success;
    }
    json_result = easydlSoundClassify_result;
    curl_easy_cleanup(curl);
     is_success = 0;
  } else {
     fprintf(stderr, "curl_easy_init() failed.");
     is_success = 1;
  }
  return is_success;
}
```

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	分类结果数组
+name	否	string	分类名称
+score	否	number	置信度

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有 疑问请在百度云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请在百度云控制台内提交工单反馈
336003	Base64解码失败	图片/音频/文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头 部。有疑问请在百度云控制台内提交工单反馈
336004	输入文件大小不合法	音频大小超出限制,请根据接口文档检查入参格式,音频文件大小应控制在4M以内,有疑问请在百度云控制 台内提交工单反馈
336005	音频解码失败	音频解码失败失败,音频文件非所限定的mp3, m4a, wav格式
336006	缺失必要参数	未上传音频文件,请补充必要参数后重新请求
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

心 私有服务器部署

の LinuxSDK集成文档

简介

本文档介绍 EasyDL 的 Linux Python SDK 的使用方法,适用于 EasyDL 和 BML。

EasyDL 通用版:

- 网络类型支持:图像分类,物体检测,图像分割,声音分类,表格预测
- 硬件支持:
 - Linux x86_64 CPU (基础版,加速版)
 - Linux x86_64 Nvidia GPU (基础版,加速版)
- 语言支持: Python 3.5, 3.6, 3.7

BML :

- 网络类型支持:图像分类,物体检测,声音分类
- 硬件支持:
 - Linux x86_64 CPU (基础版)
 - Linux x86_64 Nvidia GPU (基础版)
- 语言支持: Python 3.5, 3.6, 3.7

Release Notes

Baidu 百	度智能云文档
---------	--------

版本	说明
1.3.5	新增华为Atlas300、飞腾Atlas300 Python SDK,支持图像分类、物体检测、人脸检测、实例分割
1.3.3	EasyDL CPU普通版新增支持表格预测
1.3.1	CPU、GPU普通版新增支持BML Cloud小目标检测模型
1.2.7	声音分类模型升级
1.2.6	CPU基础版、CPU加速版、GPU基础版推理引擎优化升级
1.2.5	CPU基础版、CPU无损加速版、GPU基础版新增支持EasyDL小目标检测
1.2.4	CPU、GPU新增EasyDL目标跟踪支持;新增http server服务启动demo
1.2.2	EasyDL CPU加速版新增支持分类、高性能检测和均衡检测的量化压缩模型
1.2.1	EasyDL经典版分类高性能模型升级;支持更多模型
1.2.0	推理引擎升级;接口升级;性能优化
1.1.19	支持更多模型
1.1.18	性能优化
1.1.17	支持更多EasyDL专业版模型
1.1.15	技术优化;升级 OpenVINO 版本
1.1.14	新增声音识别python sdk
1.1.13	新增口罩模型支持
1.1.12	预测函数默认使用推荐阈值
1.1.11	EasyDL 专业版支持 SDK 加速版
1.1.10	支持图像分割
1.1.9	支持 EasyDL 专业版
1.1.8	CPU 加速版支持
1.1.7	提供模型更新工具
1.1.3	NVIDIA GPU 支持
1.1.0	架构与功能完善
1.0.6	引擎功能完善
1.0.5	paddlepaddle 支持
1.0.0	第一版!
	版本 1.3.5 1.3.3 1.3.1 1.2.7 1.2.6 1.2.5 1.2.4 1.2.2 1.2.2 1.2.1 1.2.0 1.1.19 1.1.13 1.1.13 1.1.13 1.1.13 1.1.13 1.1.13 1.1.13 1.1.13 1.1.13 1.1.13 1.1.13 1.1.3

2020-12-18: 【接口升级】 序列号的配置接口从1.2.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。 请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

快速开始

1. 安装依赖

- 根据引擎的不同, SDK 依赖了不同的底层引擎。根据所需自行安装。
- 使用声音分类SDK需要安装额外依赖*pip 安装 resampy pydub six librosa 音频默认格式支持wav文件预测,如果需要预测mp3等其他音频格式的数据需要系统额外安装ffmpeg (windows系统的ffmpeg已基在sdk中无需额外安装, linux系统需要手动安装)
- 使用表格预测SDK需要安装额外依赖 pip安装 brotlipy==0.7.0 certifi==2020.6.20 joblib==1.0.1 kaggle==1.5.12 Pillow py4j pycosat pythondateutil python-slugify ruamel_yaml text-unidecode threadpoolctl flask pandas==1.0.5 scikit-learn==0.23.2 lightgbm==2.2.3 catboost==0.24.1 xgboost==1.2.0 numpy==1.19.5 scipy==1.5.2 psutil==5.7.2 pypmml==0.9.7 torch==1.8.0 jieba==0.42.1 pyod==0.8.5 pyarrow==6.0.0 scikit-optimize==0.9.0 pyspark==3.3.0 另外ml算法 安装(目前只支持python3.7) pip install BaiduAI_TabularInfer-0.0.0-cp37-cp37m-linux_x86_64.whl 安装 paddlepaddle
- 使用x86_64 CPU 基础版预测时必须安装(目标跟踪除外):

若 CPU 为特殊型号,如赛扬处理器(一般用于深度定制的硬件中),请关注 CPU 是否支持 avx 指令集。如果不支持,请在paddle官网安装 noavx 版本

• 使用NVIDIA GPU 基础版预测时必须安装(目标跟踪除外):

python -m pip install paddlepaddle-gpu==2.2.2.post101 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA10.1的 PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2 -i https://mirror.baidu.com/pypi/simple #CUDA10.2的PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2.post110 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.0的 PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2.post111 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.1的 PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2.post112 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.2的 PaddlePaddle

不同cuda版本的环境,请参考paddle文档安装合适的 paddle 版本。 不被 paddle 支持的 cuda 和 cudnn 版本,EasyEdge 暂不支持**安装 OpenVINO** 使用x86_64 CPU 加速版 SDK 预测时必须安装。

1) 请参考 OpenVINO toolkit 文档安装 2021.4版本, 安装时可忽略Configure the Model Optimizer及后续部分

2) 运行之前,务必设置环境变量

source /opt/intel/openvino_2021/bin/setupvars.sh

安装 cuda、cudnn

• 使用Nvidia GPU 加速版预测时必须安装。 依赖的版本为 cuda9.0、cudnn7。版本号必须正确。

安装 pytorch (torch >= 1.7.0)

- 目标跟踪模型的预测必须安装pytorch版本1.7.0及以上(包含:Nvidia GPU 基础版、x86_64 CPU 基础版)。
- 目标跟踪模型Nvidia GPU 基础版还需安装依赖cuda、cudnn。

关于不同版本的pytorch和CUDA版本的对应关系: pytorch官网目标跟踪模型还有一些列举在requirements.txt里的依赖(包括torch >= 1.7.0),均可使用pip下载安装。

pip3 install -r requirements.txt

2. 安装 easyedge python wheel 包 安装说明

pip3 install -U BaiduAI_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_x86_64.whl

具体名称以 SDK 包中的 whl 为准。 安装说明:华为 Atlas300 除了需要安装BaiduAI_EasyEdge_SDK包,还需安装

pip3 install -U EasyEdge_Devkit_Atlas300-{版本号}-cp36-cp36m-linux_x86_64.whl

安装说明:飞腾 Atlas300 除了需要安装BaiduAl_EasyEdge_SDK包,还需安装

pip3 install -U EasyEdge_Devkit_Phytium.Atlas-{版本号}-cp36-cp36m-linux_aarch64.whl

3. 使用序列号激活

获取序列号	中具其服务功時 死何其親議務, 特別認知識的理解者在未地的服务部,小型设备,软硬一体方案专项适应硬件上,						
	模型名称	发布版本 🗘	应用平台	模型加速	发布状态	发布时间 🗘	
	sun_/\⊟filtest	134318-V1 宣看性能报告	通用X86 CPU-Linux	基础板	 已发布 	2021-08-19 20:24	下载SDK
				精度无损压缩加速	 已发布 	2021-08-19 20:24	下载加速版SDK
			英伟达GPU-Linux	基础板	 已发布 	2021-08-19 20:35	下截SDK
				精度无损压缩加速	 已发布 	2021-08-19 20:34	下载加速版SDK
				基础板	 已发布 	2021-08-19 18:17	下截SDK

修改demo.py 填写序列号

pred = edge.Program() pred.set_auth_license_key("这里填写序列号")

4. GPU 加速版 使用 GPU 加速版,在安装完 whl 之后,必须:

1. 从这里下载 TensorRT7.0.0.11 for cuda9.0,并把解压后的 lib 放到 C++ SDK 的 lib 目录或系统 lib 目录

2. 运行时,必须在系统库路径中包含 C++ SDK 下的lib目录。如设置LD_LIBRARY_PATH

```
cd ${SDK_ROOT}
**1. 安装 python wheel 包**
tar -xzvf python/*.tar.gz
pip install -U {对应 Python 版本的 wheel 包}
**2. 设置 LD_LIBRARY_PATH**
tar -xzvf cpp/*.tar.gz
```

```
export EDGE_ROOT=$(readlink -f $(Is -h | grep "baidu_easyedge_linux_cpp"))
export LD_LIBRARY_PATH=$EDGE_ROOT/lib
```

3. 运行 demo python3 demo.py {RES文件夹路径} {测试图片路径}

如果是使用 C++ SDK 自带的编译安装的 OpenCV, LD_LIBRARY_PATH 还需要包括 C++ SDK的 build 目录下的 thirdparty/lib 目录

```
如果没有正确设置 LD_LIBRARY_PATH,运行时可能报错:
```

ImportError: libeasyedge.so.0.4.3: cannot open shared object file: No such file or directory ImportError: libopencv_core.so.3.4: cannot open shared object file: No such file or directory

5. 测试 Demo

5.1 图片预测

输入对应的模型文件夹(默认为RES)和测试图片路径,运行:

python3 demo.py {model_dir} {image_name.jpg}

测试效果:



图片加载失败

5.2 视频预测 (适用于目标跟踪)

输入对应的模型文件夹(默认为RES)和测试视频文件路径/摄像头id/网络视频流地址,运行:

```
**video_type: 输入源类型 type:int**
**1 本地视频文件**
**2 摄像头的index**
**3 网络视频流**
**video_src: 输入源地址,如视频文件路径、摄像头index、网络流地址 type: string**
python3 demo.py {model_dir} {video_type} {video_src}
```

6. 测试Demo HTTP 服务 输入对应的模型文件夹(默认为RES)、序列号、设备ip和指定端口号,运行:

 $python 3 \ demo_serving.py \ \{model_dir\} \ \{serial_key\} \ \{host, \ default \ 0.0.0.0\} \ \{port, \ default \ 24401\}$

后,会显示:

Running on http://0.0.0.0:24401/

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片或者视频来进行测试。也可以参考`demo_serving.py`里 http_client_test()函数请求http服务进行推理。



使用说明

使用流程 demo.py

import BaiduAI.EasyEdge as edge

pred = edge.Program() pred.set_auth_license_key("这里填写序列号") pred.init(model_dir={RES文件夹路径}, device=edge.Device.CPU, engine=edge.Engine.PADDLE_FLUID) pred.infer_image({numpy.ndarray的图片}) pred.close()

demo_serving.py

import BaiduAl.EasyEdge as edge from BaiduAl.EasyEdge.serving import Serving

server = Serving(model_dir={RES文件夹路径}, license=serial_key)

```
**请参考同级目录下demo.py里:**
```

```
**pred.init(model_dir=xx, device=xx, engine=xx, device_id=xx)**
```

对以下参数device\device_id和engine进行修改

server.run(host=host, port=port, device=edge.Device.CPU, engine=edge.Engine.PADDLE_FLUID)

初始化

• 接口

```
def init(self,
    model_dir,
    device=Device.CPU,
    engine=Engine PADDLE_FLUID,
    config_file='conf.json',
    preprocess_file='preprocess_args.json',
    model_file='model',
    params_file='params'
    label_file='label_list.txt',
    infer_cfg_file='infer_cfg.json',
    device_id=0,
    thread_num=1
    ):
   ....
  Args:
    model_dir: str
     device: BaiduAI.EasyEdge.Device,比如:Device.CPU
     engine: BaiduAI.EasyEdge.Engine,比如:Engine.PADDLE_FLUID
     config_file: str
     preprocess_file: str
     model_file: str
     params_file: str
     label_file: str 标签文件
     infer_cfg_file: 包含预处理、后处理信息的文件
device_id: int 设备ID
    thread_num: int CPU的线程数
  Raises:
     RuntimeError, IOError
```

Returns: bool: True if success

使用 NVIDIA GPU 预测时,必须满足:

- 机器已安装 cuda, cudnn
- 已正确安装对应 cuda 版本的 paddle 版本
- 通过设置环境变量 FLAGS_fraction_of_gpu_memory_to_use 设置合理的初始内存使用比例

使用 CPU 预测时,可以通过在 init 中设置 thread_num 使用多线程预测。如:

 $pred.init(model_dir=_model_dir,\ device=edge.Device.CPU,\ engine=edge.Engine.PADDLE_FLUID,\ thread_num=1)$

预测图像

● 接口

def infer_image(self, img,	
threshold=0.3,	
channel_order='HWC',	
color_format='BGR',	
data_type='numpy'):	
Args:	
img: np.ndarray or bytes	
threshold: float	
only return result with confidence larger than threshold	
channel_order: string	
channel order HWC or CHW	
color_format: string	
color format order RGB or BGR	
data_type: string	
image data type	
Returns:	
list	

• 返回格式: [dict1, dict2, ...]

字段	类型	取值	说明
confidence	float	0~1	分类或检测的置信度
label	string		分类或检测的类别
index	number		分类或检测的类别
x1, y1	float	0~1	物体检测,矩形的左上角坐标 (相对长宽的比例值)
x2, y2	float	0~1	物体检测,矩形的右下角坐标(相对长宽的比例值)
mask	string/numpy.ndarray	图像分割的mask	

关于矩形坐标

- x1 * 图片宽度 = 检测框的左上角的横坐标
- y1 * 图片高度 = 检测框的左上角的纵坐标
- x2*图片宽度 = 检测框的右下角的横坐标
- y2 * 图片高度 = 检测框的右下角的纵坐标

可以参考 demo 文件中使用 opencv 绘制矩形的逻辑。

结果示例

• i) 图像分类

```
{
    "index": 736,
    "label": "table",
    "confidence": 0.9
}
```

● ii) 物体检测

```
"y2": 0.91211,
"label": "cat",
"confidence": 1.0,
"x2": 0.91504,
"index": 8,
"y1": 0.12671,
"x1": 0.21289
}
```

• iii) 图像分割

```
{
    "name": "cat",
    "score": 1.0,
    "location": {
        "left": ...,
        "top": ...,
        "width": ...,
        "height": ...,
    },
    "mask": ...
}
```

mask字段中,data_type为numpy时,返回图像掩码的二维数组

```
{
{0,0,0,0,0,0,0,0,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,0,0,0,0,0,0,0},
}
其中1代表为目标区域,0代表非目标区域
```

data_type为string时,mask的游程编码,解析方式可参考 demo

预测视频 (目前仅限目标跟踪模型调用)

```
• 接口
```

```
def infer_frame(self, frame, threshold=None):
"""
视频推理(抽帧之后)
:param frame:
:param threshold:
:return:
"""
```

```
● 返回格式dict
```

字段	类型	说明
pos	dict1	当前帧每一个类别的追踪目标的像素坐标(tlwh)
id	dict2	当前帧每一个类别的追踪目标的id
score	dict3	当前帧每一个类别的追踪目标的识别置信度
label	dict4	class_idx(int)与label(string)的对应关系
class_num	int	追踪类别数

预测声音

• 使用声音分类SDK需要安装额外依赖 *pip 安装 resampy pydub* 音频默认格式支持wav文件预测,如果需要预测mp3等其他音频格式的数据需要 系统额外安装 ffmpeg (windows系统的ffmpeg已集成在sdk中无需额外安装,linux系统需要手动安装)

接口

def infer_sound(self, sound_binary, threshold=0.3):
Args:
sound_binary: sound_binary
threshold: confidence
Returns:
list

• 返回格式: [dict1, dict2, ...]

字段	类型	取值	说明
confidence	float	0~1	分类的置信度
label	string		分类的类别
index	number		分类的类别

升级模型 适用于经典版升级模型,执行bash update_model.sh</mark>,根据提示,输入模型路径、激活码、模型ID、模型版本,等待模型更新完毕即 可。

FAQ

Q: EasyDL 离线 SDK 与云服务效果不一致,如何处理? A: 后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

Q: 运行时报错"非法指令"或"illegal instruction"A: 可能是 CPU 缺少 avx 指令集支持,请在paddle宫网 下载 noavx 版本覆盖安装

Q: NVIDIA GPU预测时,报错显存不足: A: 如以下错误字样:

paddle.fluid.core.EnforceNotMet: Enforce failed. Expected allocating <= available, but received allocating:20998686233 > available:19587333888.

Insufficient GPU memory to allocation. at [/paddle/paddle/fluid/platform/gpu_info.cc:170]

请在运行 Python 前设置环境变量,通过export FLAGS_fraction_of_gpu_memory_to_use=0.3来限制SDK初始使用的显存量,0.3表示初始使用 30%的显存。如果设置的初始显存较小,SDK 会自动尝试 allocate 更多的显存。

Q:我想使用多线程预测,怎么做?如果需要多线程预测,可以每个线程启动一个Progam实例,进行预测。 demo.py文件中有相关示例代码。

注意: 对于CPU预测,SDK内部是可以使用多线程,最大化硬件利用率。参考init的thread_num参数。

Q: 运行SDK报错 Authorization failed

情况一:日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888" ./easyedge_demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- MAC地址变化
- 磁盘变更
- BIOS重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除<mark>~/.baidu/easyedge 目录</mark>,再重新激活。

情况三: Atlas Python SDK日志提示ImportError: libavformat.so.58: cannot open shared object file: No such file or directory 或者其他类似so找不到可以在LD_LIBRARY_PATH环境变量加上libs和thirdpartylibs路径,例如

export LD_LIBRARY_PATH=/xxx/libs:/xxx/thirdpartylibs:\$LD_LIBRARY_PATH # tips: 这里/xxx需要替换为真实路径,/xxx路径查找方法如下

查找安装包内libs和thirdpartylibs路径的方法如下(以华为Atlas300 SDK为例,其他SDK查找方法类似):

pip3 show EasyEdge-Devkit-Atlas300 # 结果中会显示 Location 路径,也就是包的安装路径 **libs和thirdpartylibs两个路径在 Location 所指示的路径 easyedge_CANN 子文件夹下**

心 WindowsSDK集成文档

简介

本文档介绍声音分类服务器端Windows SDK的使用方法。

- 操作系统支持
 - 64位 Windows 7 及以上
 - 64位Windows Server 2012及以上
- 环境依赖 (必须安装以下版本)
 - .NET Framework 4.5
 - Visual C++ Redistributable Packages for Visual Studio 2013
 - Visual C++ Redistributable Packages for Visual Studio 2015
- 协议
 - HTTP
- 更详细的环境说明可参考SDK内的README.md

Release Notes | 时间 | 版本 | 说明 | | -------- | ------ | -2023-08-30 | 1.8.3 | 新增支持按实例数鉴权 | | 2023-06-29 | 1.8.2 | 优化模型 算法 || 2023-05-17 | 1.8.1 | 预测引擎升级,修复部分网络内存泄露问题 || 2023-03-16 | 1.8.0 | 预测引擎升级 || 2022-12-29 | 1.7.2 | 预测引 擎升级 | | 2022-10-27 | 1.7.1 | GPU底层引擎升级,下线基础版CUDA10.0及以下版本支持 | | 2022-09-15 | 1.7.0 | 优化模型算法;GPU CUDA9.0 CUDA10.0 标记为待废弃状态 || 2022-07-28 | 1.6.0 | 优化模型算法 || 2022-05-27 | 1.5.1 | 新增支持BML Cloud小目标检测模型 || 2022-05-18 | 1.5.0 | 修复各别机器下程序崩溃的问题 || 2022-04-25 | 1.4.1 | EasyDL, BML升级支持paddle2模型 || 2022-03-25 | 1.4.0 | 优化 模型算法 || 2021-12-22 | 1.3.5 | GPU基础版推理引擎优化升级; GPU加速版支持自定义模型文件缓存路径; demo程序优化环境依赖检测 || 2021-10-20 | 1.3.4 | 修复已知问题| | 2021-08-19 | 1.3.2 | 新增支持EasyDL小目标检测,新增DEMO二进制文件 || 2021-06-29 | 1.3.1 | 预测引 擎升级 || 2021-05-13 | 1.3.0 | 模型发布新增多种加速方案选择;目标追踪支持x86平台的GPU及加速版;展示已发布模型性能评估报告 || 2021-04-08 | 1.2.3 | 支持BML平台模型仓库本地上传模型 | | 2021-03-09 | 1.2.2 | 修复已知问题 | | 2021-01-27 | 1.2.1 | 新增模型支持;性能优 化;问题修复||2020-12-18|1.2.0|推理引擎升级||2020-11-26|1.1.20|新增一些模型的加速版支持||2020-10-29|1.1.19|修复已知问 题 || 2020-09-17 | 1.1.18| 支持更多模型 || 2020.08.11 | 1.1.17 | 支持专业版更多模型 || 2020.06.23 | 1.1.16 | 支持专业版更多模型 || 2020.05.15 | 1.1.15 | 更新加速版tensorrt版本,支持高精度检测 | | 2020.03.13 | 1.1.14 | 支持声音分类 | | 2020.02.23 | 1.1.13 | 支持多阶段模 型|| 2020.01.16 | 1.1.12 | 预测默认使用推荐阈值|| 2019.12.26 | 1.1.11 | 支持物体检测高精度算法的CPU加速版,EasyDL 专业版支持 SDK 加速版 || 2019.12.04 | 1.1.10 | 支持图像分割 || 2019.10.21 | 1.1.9 | 支持 EasyDL 专业版 || 2019.08.29 | 1.1.8 | CPU 加速版支持 || 2019.07.19 | 1.1.7 | 提供模型更新工具 || 2019.05.16 | 1.1.3 | NVIDIA GPU 支持 || 2019.03.15 | 1.1.0 | 架构与功能完善 || 2019.02.28 | 1.0.6 | 引擎功能完善 | | 2019.02.13 | 1.0.5 | paddlepaddle 支持 | | 2018.11.30 | 1.0.0 | 第一版 ! |

快速开始

1. 安装依赖

安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

Visual C++ Redistributable Packages for Visual Studio 2015

https://www.microsoft.com/zh-cn/download/details.aspx?id=48145

如果使用GPU版SDK,请安装CUDA+cuDNN

https://developer.nvidia.com/cuda https://developer.nvidia.com/cudnn

声音分类依赖

·安装six	
--------	--

打开cmd,进入sdk包所在目录。执行EasyEdge-win-mXXXX-x86-nvidia-gpu\python37\python.exe -m pip install -U six -i http://mirrors.aliyun.com/pypi/simple/ --trusted-host mirrors.aliyun.com

·安装librosa

打开cmd,进入sdk包所在目录。执行EasyEdge-win-mXXXX-x86-nvidia-gpu\python37\python.exe -m pip install -U librosa -i http://mirrors.aliyun.com/pypi/simple/ --trusted-host mirrors.aliyun.com

注意事项

- 1. 安装目录不能包含中文
- 2. Windows Server 请自行开启,选择"我的电脑"——"属性"——"管理"——"添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启即 可。

2. 运行离线SDK

解压下载好的SDK,SDK默认使用cuda9版本,如果需要cuda10请运行EasyEdge CUDA10.0.bat切换到cuda10版本,之后打开EasyEdge.exe,选

择鉴权模式,输入Serial Num

LC.

图片加载失败

点击"启动服务",等待数秒即可启动成功,本地服务默认运行在

http://127.0.0.1:24401/

其他任何语言只需通过HTTP调用即可。

2.1	离线鉴权	(默认鉴权模式)	首次联网激活,	后续离线使用	

Al ver. 1.2.0.0 core. 1	1.8.1 sec. 1.5.1			-		Х
Model	bl. alt	Tel ^t and	10-00-0	210	:	
Serial Num						
Host	127.0.0.1					
Port	24401					
鉴权模式	● 离线鉴权 (○ 按实例数鉴权				
服务状态	服务未启动					
		启动服务				
				是否开机	自动启动	

2.2 按实例数鉴权 周期性联网激活,离线后会释放所占用鉴权,启动时请确保心跳间隔小于等于生成序列号时填写的定期确认时间

AI ver. 1.2.0.0 core. 1	.8.1 sec. 1.5.1	-		×
Model	la ser de la ser de la ser en			
Serial Num				
Host	127.0.0.1			
Port	24401]	
鉴权模式	○ 离线鉴权 ④ 按实例数鉴权 心跳间隔 20 [秒 ~	·	
服务状态	服务未启动			
	启动服务	是否开机的	自动启动	

基于源码集成时,若需要按实例数鉴权,需要通过代码指定使用按实例数鉴权

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2); global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 20);

或通过环境变量指定

set EDGE_CONTROLLER_KEY_AUTH_MODE=2 set EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=20

3. Demo示例

服务运行成功,此时可直接在浏览器中输入http://127.0.0.1:24401,在h5中测试模型效果。

$\leftrightarrow \ \ni \ {\tt G}$	① 127.0.0.1:24401	le ☆	

```
【声音分类】82450 _____-短音频单标签-5种声音分类V1
```



使用说明

调用说明

Python 使用示例代码如下



C# 使用示例代码如下

FileStream fs = new FileStream("./audio.mp3", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] audio = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(audio, 0, audio.Length); stream.Close(); WebResponse response = request.GetResponse();

StreamReader sr = new StreamReader(response.GetResponseStream());
Console.WriteLine(sr.ReadToEnd());
sr.Close();
response.Close();

C++ 使用示例代码如下,需要安装curl

```
**include <sys/stat.h>**
**include <curl/curl.h>**
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./audio.mp3";
  FILE *fp = NULL;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\"\n", post_data_filename);
     return -1;
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
     curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
     curl_easy_setopt(curl, CURLOPT_POST, 1L);
     curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
     curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
     res = curl_easy_perform(curl);
     if (res != CURLE_OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     curl_easy_cleanup(curl);
  curl_global_cleanup();
  fclose(fp);
  return 0;
```

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送声音二进制。

返回参数 | 字段 | 类型 | 取值 | 说明 | | -------- | ----- | ----- | ----- | | confidence | float | 0~1 | 分类的置信度 | | label | string | | 分类的类别 | 别 | | index | number | | 分类的类别 |

集成指南

基于HTTP集成

通过EasyEdge.exe启动服务后,参照上面的调用说明,通过HTTP请求集成到自己的服务中

基于c++ dll集成

集成前提

解压开的SDK包中包含src、lib、dll、include四个目录才支持基于c++ dll集成

集成方法

参考src目录中的CMakeLists.txt进行集成

基于c# dll集成

集成前提

解压开的SDK包中包含src\demo_serving_csharp、dll两个目录才支持基于c# dll集成

集成方法

参考src\demo_serving_csharp目录中的CMakeLists.txt进行集成

FAQ

1. 服务启动失败,怎么处理?

请确保相关依赖都安装正确,版本必须如下: *.NET Framework 4.5* Visual C++ Redistributable Packages for Visual Studio 2013 * Visual C++ Redistributable Packages for Visual Studio 2015

2. 服务调用时返回为空,怎么处理?调用输入的音频不为空。

3. 多个模型怎么同时使用? SDK设置运行不同的端口,点击运行即可。

4. JAVA、C#等其他语言怎么调用SDK?

参考 https://ai.baidu.com/forum/topic/show/943765

5. 启动失败,缺失DLL? 打开EasyEdge.log,查看日志错误,根据提示处理 缺失DLL,请使用 https://www.dependencywalker.com/ 查看相应模 块依赖DLL缺失哪些,请自行下载安装

6. 启动失败,报错NotDecrypted? Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

7. 启动失败,报错 SerialNum无效

日志显示failed to get/check device id(xxx)或者 Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 C:\Users\\${用户名}\.baidu\easyedge 目录,再重新激活。

其他问题 如果无法解决,可到论坛发帖: https://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

心 通用小型设备部署

^のWindows集成文档

简介

本文档介绍Windows CPU SDK的使用方法。

- 网络类型支持:图像分类,物体检测,图像分割,声音分类
- 硬件支持:
 - Intel CPU 普通版 * x86_64
 - CPU 加速版 Intel Xeon with AVX2 and AVX512 Intel Core Processors with AVX2 Intel Atom Processors with SSE * AMD Core Processors with AVX2
 - Intel Movidius MyRIAD2/MyRIAD X (仅支持Win10)
- 操作系统支持
 - 普通版:64位 Windows 7 及以上,64位Windows Server2012及以上
 - 加速版:64位 Windows 10,64位Windows Server 2019及以上
- 环境依赖 (必须安装以下版本)
 - .NET Framework 4.5
 - Visual C++ Redistributable Packages for Visual Studio 2013
 - Visual C++ Redistributable Packages for Visual Studio 2015-2019
- 协议
 - HTTP
- 更详细的环境说明可参考SDK内的README.md

*intel 官方合作,拥有更好的适配与性能表现

Release Notes |时间 |版本 |说明 | | ------- | ----- || 2023-08-30 | 1.8.3 |新增支持按实例数鉴权 || 2023-06-29 | 1.8.2 | 优化模型 算法 || 2023-05-17 | 1.8.1 |预测引擎升级,修复部分网络内存泄露问题 || 2023-03-16 | 1.8.0 |预测引擎升级 || 2022-12-29 | 1.7.2 | 预测引 擎升级 || 2022-10-27 | 1.7.1 |优化模型算法 || 2022-09-15 | 1.7.0 | 新增支持表格预测 || 2022-07-28 | 1.6.0 | 优化模型算法 || 2022-05-27 | 1.5.1 | 新增支持BML Cloud小目标检测模型 || 2022-05-18 | 1.5.0 | 修复各别机器下程序崩溃的问题 || 2022-04-25 | 1.4.1 | EasyDL, BML升级 支持paddle2模型 || 2022-03-25 | 1.4.0 | 优化模型算法 || 2021-12-22 | 1.3.5 | CPU基础版推理引擎优化升级 ; demo程序优化环境依赖检测 || 2021-10-20 | 1.3.4 | CPU加速版推理引擎优化升级 || 2021-08-19 | 1.3.2 | 新增DEMO二进制文件 || 2021-06-29 | 1.3.1 | 预测引擎升级 || 2021-05-13 | 1.3.0 | 模型发布新增多种加速方案选择 ; 目标追踪支持x86平台的GPU及加速版 ; 展示已发布模型性能评估报告 || 2021-04-08 | 1.2.3 | 支持BML平台模型仓库本地上传模型 || 2021-03-09 | 1.2.2 | CPU加速版支持int8量化模型 || 2021-01-27 | 1.2.1 | 新增模型支持 ; 性能优 化 ; 问题修复 || 2020.12.18 | 1.2.0 | 推理引擎升级 || 2020-11-26 | 1.1.20 | 新增一些模型的加速版支持 || 2020.10.29 | 1.1.20 | 修复已知问 题 || 2020.05.15 | 1.1.16 | 优化性能,修复已知问题 || 2020-09-17 | 1.1.19 | 支持更多模型 || 2020.04.16 | 1.1.15 | 升级引擎版本 || 2020.06.23 | 1.1.17 | 支持专业版更多模型 || 2020.02.15 | 1.1.16 | 优化性能,修复已知问题 || 2020.04.16 | 1.1.15 | 升级引擎版本 || 2020.03.13 | 1.1.14 | 支持EdgeBoardVMX || 2020.02.23 | 1.1.13 | 支持多阶段模型 || 2020.04.16 | 1.1.19 | 支持 EasyDL 专业版 || 2019.02.26 | 1.1.11 | CPU加速版支持物体检测高精度 || 2019.12.04 | 1.1.10 | 支持图像分割 || 2019.01.21 | 1.1.9 | 支持 EasyDL 专业版 || 2019.08.29 | 1.1.8 | CPU 加速版支持 || 2019.07.19 | 1.1.7 | 提供模型更新工具 || 2019.05.16 | 1.1.3 | NVIDIA GPU 支持 || 2019.03.15 | 1.1.0 | 架构与功能完善 || 2019.02.28 | 1.0.6 | 引擎功能完善 || 2019.02.13 | 1.0.5 | paddlepaddle 支持 || 2018.11.30 | 1.0.0 | 第一版 !|

快速开始

1. 安装依赖

必须安装:

安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642
Visual C++ Redistributable Packages for Visual Studio 2013

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

Visual C++ Redistributable Packages for Visual Studio 2015-2019

https://docs.microsoft.com/en-us/cpp/windows/latest-supported-vc-redist?view=msvc-160

可选安装:

Openvino (仅使用Intel Movidius必须)

- 使用 OpenVINO[™] toolkit 安装,请参考 OpenVINO toolkit 文档安装 2020.3.1LTS(必须)版本,安装时可忽略Configure the Model Optimizer及 后续部分。
- 使用源码编译安装,请参考 Openvino Inference Engine文档编译安装 2020.3.1LTS (必须)版本。

安装完成后,请设置环境变量OPENVINO_HOME为您设置的安装地址,默认是C:\Program Files (x86)\IntelSWTools,并确保文件夹下的openvino 的快捷方式指到了2020.3.1LTS版本。

充厚性	xiaovuchen 的用户变量(U)		
·算机名 硬件 高级 系统保护 远程	· · · · · · · · · · · · · · · · · · ·	值	
要进行大多数更改,你必须作为管理员登录。	OneDrive	Cilliser OneDrive	
一性能	OPENVINO_HOME	C:\Program Files (x86)\IntelSWT	ools
视觉效果,处理器计划,内存使用,以及虚拟内存 设置(S)…	Path TEMP TMP	C:\Users\AppData\L C:\Users\AppData\L C:\Users\AppData\L	ocal\Microsoft\WindowsApp ocal\Temp ocal\Temp
用戶配置文件 与登录帐户相关的桌面设置		新建(N)	编辑(E) 删除(D)
设置(E)	系统变量(S)		
	变量	值	^
启动和故障恢复	ComSpec	C:\Windows\system32\cmd.exe	
系统启动、系统故障和调试信息 设置①	DriverData INTEL_DEV_REDIST MIC_LD_LIBRARY_PATH NUMBER_OF_PROCESSO OS	C:\Windows\System32\Drivers\E C:\Program Files (x86)\Common %INTEL_DEV_REDIST%compiler\ RS 8 Windows_NT)riverData Files\Intel\Shared Libraries\ \lib\intel64_win_mic
确定 取消 应用	Α)	新建(W)	编辑(I) 删除(L) 删除(L) 微定 取消
共享 查看			
✓ Program Files (x86) → International Action of the In	elSWTools >	~	○ 夕 搜索
□ 名称	^	修	改日期
💙 📊 ope	nvino	20	20/4/15 23:41
ope	nvino_2019.2.275	20	20/3/10 19:34
ope	nvino_2020.2.117	20	20/4/15 23:42
注意事项			
1. 安装目录不能包含中文			

2. Windows Server 请自行开启,选择"我的电脑"——"属性"——"管理"——"添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启即可。

2. 运行离线SDK

解压下载好的SDK,打开EasyEdge.exe,输入Serial Num,选择鉴权模式,点击"启动服务",等待数秒即可启动成功,本地服务默认运行在

http://127.0.0.1:24401/

其他任何语言只需通过HTTP调用即可。

2.1 离线鉴权 (默认鉴权模式) 首次联网激活,后续离线使用

Al ver. 1.2.0.0 core. 1	l.8.1 sec. 1.5.1	-		×
Model	Mark Street Selection	200		
Serial Num				
Host	127.0.0.1			
Port	24401			
鉴权模式	● 离线鉴权 ○ 按实例数鉴权		_	
服务状态	服务未启动			
	启动服务	是否开机	自动启动	

2.2 按实例数鉴权 周期性联网激活,离线后会释放所占用鉴权,启动时请确保心跳间隔小于等于生成序列号时填写的定期确认时间

Al ver. 1.2.0.0 core. 1	.8.1 sec. 1.5.1	– 🗆 X
Model	al and the second sec	V A REPAIR
Serial Num		
Host	127.0.0.1	
Port	24401	
鉴权模式	○ 离线鉴权 ④ 按实例数鉴权 → 跳间隔	暑 20 秒 ~
服务状态	服务未启动	
	启动服务	
		是否开机自动启动 🗌

基于源码集成时,若需要按实例数鉴权,需要通过代码指定使用按实例数鉴权

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2); global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 20);

或通过环境变量指定

set EDGE_CONTROLLER_KEY_AUTH_MODE=2 set EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=20

3. Demo示例

服务运行成功,此时可直接在浏览器中输入http://127.0.0.1:24401,在h5中测试模型效果。

፼ ☆ □ ≗

← → C ③ 127.0.0.1:24401

【声音分类】82450 _____-短音频单标签-5种声音分类V1



如果上传音频文件后,弹窗报错 500 internal server error,可参看该解决方案:点击这里跳转

使用说明

声音服务调用说明

Python 使用示例代码如下

```
import requests
```

```
with open('./1.mp3', 'rb') as f:
audio = f.read()
```

```
**params 为GET参数 data 为POST Body**
result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1},
data=audio).json()
```

C# 使用示例代码如下

```
FileStream fs = new FileStream("./audio.mp3", FileMode.Open);
BinaryReader br = new BinaryReader(fs);
byte[] audio = br.ReadBytes((int)fs.Length);
br.Close();
fs.Close();
string url = "http://127.0.0.1:8402?threshold=0.1";
HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url);
request.Method = "POST";
Stream stream = request.GetRequestStream();
stream.Write(audio, 0, audio.Length);
stream.Close();
WebResponse response = request.GetResponse();
StreamReader sr = new StreamReader(response.GetResponseStream());
Console.WriteLine(sr.ReadToEnd());
sr.Close();
response.Close();
```

C++ 使用示例代码如下,需要安装curl

```
**include <sys/stat.h>**
**include <curl/curl.h>**
**define S_ISREG(m) (((m) & 0170000) == (0100000))**
**define S_ISDIR(m) (((m) & 0170000) == (0040000))**
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./audio.mp3";
  FILE *fp = NULL;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
    fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  }
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\" \n", post_data_filename);
     return -1;
  }
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
     struct curl_slist* headers = NULL;
     // 根据您的音频格式,选择对应的Content-Type
     //MP3: audio/mpeg
     //Ogg: audio/ogg
     //Mav: audio/mav
    headers = curl_slist_append(headers, "Content-Type: audio/mpeg");
     //set headers
     curl_easy_setopt(curl, CURLOPT_HTTPHEADER, headers);
    curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
     curl_easy_setopt(curl, CURLOPT_POST, 1L);
     curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
     curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
     res = curl_easy_perform(curl);
     if (res != CURLE_OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     curl_easy_cleanup(curl);
  }
  curl_global_cleanup();
  fclose(fp);
  return 0;
```

请求参数

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送声音二进制。

返回参数|字段|类型|取值|说明||--------|----|-----|-----|| confidence | float | 0~1 | 分类或检测的置信度 || label | string || 分 类或检测的类别 || index | number || 分类或检测的类别 |

Baidu 百度智能云文档

基于HTTP集成

通过EasyEdge.exe启动服务后,参照上面的调用说明,通过HTTP请求集成到自己的服务中

基于c++ dll集成

集成前提

解压开的SDK包中包含src、lib、dll、include四个目录才支持基于c++ dll集成

集成方法

参考src目录中的CMakeLists.txt进行集成

基于c# dll集成

集成前提

解压开的SDK包中包含src\demo_serving_csharp、dll两个目录才支持基于c# dll集成

集成方法

参考src\demo_serving_csharp目录中的CMakeLists.txt进行集成

FAQ

1. 服务启动失败,怎么处理?

请确保相关依赖都安装正确,版本必须如下: *.NET Framework 4.5* Visual C++ Redistributable Packages for Visual Studio 2013 * Visual C++ Redistributable Packages for Visual Studio 2015

如使用的是CPU加速版,需额外确保Openvino安装正确,版本为2020.3.1LTS版 如使用Windows Server,需确保开启桌面体验

2. 服务调用时返回为空,怎么处理?调用输入的图片必须是RGB格式,请确认是否有alpha通道。

3. 多个模型怎么同时使用? SDK设置运行不同的端口,点击运行即可。

4. JAVA、C#等其他语言怎么调用SDK?

参考 https://ai.baidu.com/forum/topic/show/943765

5. 启动失败,缺失DLL? 打开EasyEdge.log,查看日志错误,根据提示处理 缺失DLL,请使用 https://www.dependencywalker.com/ 查看相应模 块依赖DLL缺失哪些,请自行下载安装

6. 启动失败,报错NotDecrypted? Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

7. 启动失败,报错 SerialNum无效

日志显示failed to get/check device id(xxx)或者 Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

```
以及系统相关信息。
```

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 C:\Users\\${用户名}\.baidu\easyedge 目录,再重新激活。

8. 勾选"开机自动启动"后,程序闪退

一般是写注册表失败。

可以确认下HKEY_CURRENT_USER下Software \Microsoft \Windows \CurrentVersion \Run能否写入(如果不能写入,可能被杀毒软件等工具管制)。也可以尝试基于bin目录下的easyedge_serving.exe命令行形式的二进制,自行配置开机自启动。

9.浏览器打开webui上传文件提示 500 internal server error



查看根目录下EasyEdge.log

1、如果报错日志为: "ModuleNotFoundError: No module named 'librosa'"

解决方案:

1) 打开cmd命令行终端,进入SDK根目录下 python37文件夹

2) 使用python37文件夹内自带python解释器安装缺失的依赖项

python.exe -m pip install librosa

2、如果报错: Microsoft Visual C++ 14.0 is required. Get it with "Microsoft Visual C++ Build Tools"

或者 Failed to build soxr ERROR: Could not build wheels for soxr which use PEP 517 and cannot be installed directly

解决方案:本质原因是缺少数学库的头文件,需下载Microsoft Visual C++ Build Tools: https://blog.csdn.net/qq_42685893/article/details/129459771

其他问题 如果无法解决,可到论坛发帖: https://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

EasyDL 视频使用说明

EasyDL视频介绍

心简介

```
Hi,您好,欢迎使用百度EasyDL视频
```

目前EasyDL视频支持训练以下模型:

• 视频分类

分析短视频的内容,识别出视频内人体做的是什么动作,物体/环境发生了什么变化

• 目标跟踪

对视频流中的特定运动对象检测识别,获取目标的运动参数,从而实现对后续视频帧该对象的运动预测(轨迹、速度等),实现对运动目标的行 为理解

心 可视化操作

无需深度学习专业知识,模型创建-数据上传-模型训练-模型发布全流程可视化便捷操作,最快15分钟即可获得一个高精度模型

心 操作步骤

Step 1 创建模型

确定模型名称,记录希望模型实现的功能

Step 2 上传并标注数据

前往数据总览页面上传数据,并在线标注数据

Step 3 训练模型并校验效果

选择部署方式与算法,用上传的数据一键训练模型

模型训练完成后,可在线校验模型效果

Step 4 发布模型

根据训练时选择的部署方式,将模型以云端API的方式发布使用

更详细的操作指导,请参考各类模型下的技术文档

视频分类

视频分类介绍

心简介

EasyDL视频分类是针对视频内容识别推出的一个定制化训练平台。定制视频分类模型,可以用于分析短视频的内容,识别出视频内人体做的是什么动作,物体/环境发生了什么变化。

心 产品功能

● AI模型训练平台

专门用于训练视频内容分析相关业务场景下的高精度AI模型

• 定制化识别

客户可自定义要识别的场景,充分满足客户定制化需求

• 全可视化操作

所有模型训练相关的操作都可以在网页上进行,无需编程,仅需四步即可部署定制化AI模型

∞ 产品优势

• 可即用

业务流程极简,全可视化界面操作,无需深度学习基础,仅需四步即可部署定制化AI模型

更轻快

训练数据每类仅需50个短视频文件,最快15分钟可训练完毕

高精度

超过三分之二的模型准确率>90%

强安全

数据加密与隔离,完善的服务调用鉴权,为客户的数据和模型提供企业级安全保障

心 适用场景

• 人体动作监控

定制监控人体特殊动作,比如特殊手势,工地/后厨人员行为等

• 环境变化监控

定制监控环境变化,比如山体塌方,泥石流等

• 视频内容分析

快速分析视频内容,可用于短视频APP和直播平台中

- 物体状态变化监控
 定制识别特定物体的移动方向、形态变化等
- 其他
 尽情脑洞大开,训练你希望实现的视频分类模型

心 如何访问EasyDL视频分类

• 产品首页

请访问:EasyDL视频分类

模型训练

登录百度云后,请访问:视频分类模型训练,可进行模型训练和部署。

控制台

登录百度云后,请访问:百度云控制台,可进行应用创建和模型云服务管理

创建模型

进入创建模型页面

在EasyDL视频分类产品主页点击【开始训练】按钮进入到模型训练页,下面会出现两种情况:

第一种,如果您没有登录百度云,则会跳转到百度云登录页面,没有百度账户的客户请先注册百度账户。登录后,会跳转到模型概览页,点击 「视频分类」卡片上的「点击前往」按钮,会跳转模型训练页面的创建模型页。

第二种,如果您已登录,会直接进入到我的模型页,该页面能够管理已经创建的模型,点击左侧列表中的创建模型进入创建模型页面。

创建模型 进入创建模型页面后你会看到如下图中展示的内容

视频	频分类模型	€	模型列表 > 仓	刘建模型	
63	总览 模型中心 我的模型		模型类别 模型名称 *	视频分类	
	创建模型		模型归属	公司 个人	
	校验模型		邮箱地址 *	z******@baidu.com	
	发布模型		联系方式 *	135****919	0
	EasyData数据服务 数据总览		功能描述 *		
	标签组管理			0/500	
	在线标注				
	公有云服务			完成	
	在线服务				

需要填写的项目如下:

● 模型名称

模型的名称

• 模型归属

模型是属于公司的,还是属于个人的,如果是前者,请填写公司名称

● 选择行业

请选择您公司所属的行业

- 应用场景
 请选择模型将会被应用于的业务场景
- 邮箱地址
 用于联系到您的邮箱地址
- 联系方式

有效的联系方式将有助于后续模型上线的人工快速审核,以及更快的百度官方支持,推荐填写个人手机号码

• 功能描述

描述改模型将要应到的业务场景,详细的描述,在获取官方支持时,能帮助我们为您提供准确的使用建

如下图所示,完成所有填写项后点击【完成】按钮完成模型创建,创建完成后会跳转到我的模型页面。

视频分类模型 ∢──	模型列表 > 创建模型	
	模型类别 视频分类	
fb 模型中心	模型名称 * 手语识别	
我的模型		
创建模型	模型归属 公司 个人	
训练模型	百度	
校验模型	所属行业 * 智能硬件 >	
发布模型		
🖉 EasyData数据服务	应用场景 * 人体动作监控 ~	
数据总览	邮箱地址 * z*******@baidu.com	
标签组管理	联系方式 * 135*****919 ⑦	
在线标注		
目 公有云服务	功能描述 * 用于智能硬件中的手语识别功能	
在线服务		
	14/500	
	完成	

数据准备

心 创建数据集

创建数据集完成模型创建后,会跳转到我的模型页面,这时您会看到如下图展示的内容,由于模型还未训练,所以模型列表中没有显示模型的效果,在训练模型前,需要先完成创建数据集。

视频分类模型	\in	模型列表					操作文档	常见问题	教学视频	提交工单
88 总览										
6 模型中心		创建模型								
我的模型		【御師公米】王语识别	乙 横刑10:100094					ç	三 训练	前 期除
创建模型										
训练模型			模型创建成功,若	无数据集请先在"数据中心	"创建,上传训练数据:	训练 模型后,可以在此处查看	模型的最新版:	*		
校验模型										
发布模型										
☑ EasyData数据服务		【视频分类】jay_test_1	3.5 🖸 模型ID: 172966				음 训	练③月	历史版本	直 删除
数据总览		部署方式	版本	训练状态	服务状态	模型效果		操作		
标签组管理						top1准确率: 70.59%				
在线标注		公有云API	V3	• 训练完成	未发布	top5准确率: 100.00%	%	宣看版本配: 校验	直 甲请发	ti
目 公有云服务						Some of Highland				
在线服务										

点击模型列表内的上传或是左侧栏数据中心下的创建数据集可以进入到创建、导入数据集页面,如下图所示

视频分类模型 ∢──	我的数据总览	> 创建数据集
品 总览		
曲 模型中心	数据集名称	限制50个汉字内
我的模型	数据类型	视频
创建模型	数据集版本	V1
训练模型		
校验模型	标注类型	视频分类
发布模型		1 mm
☑ EasyData数据服务		
数据总览	标注模板	● 短视频单标签
标签组管理		中 由
在线标注		70%
目 公有云服务		
在线服务		

心 上传数据集

上传视频分类数据集

已标注数据上传目前支持本地导入、BOS目录导入、分享链接导入、平台已有数据集导入,4种导入方式。 支持的标注格式有文件夹命名分类和 json平台通用两种。



• 以文件夹命名分类

导入要求如下:

- 1. 上传已标注文件要求格式为zip格式压缩包,同时压缩前源文件大小在5GB以内
- 2. 压缩包内支持一个文件夹的名称作为标签,文件夹下的所有视频文件作为样本
- 3. 标签由数字、中英文、中/下划线组成,长度上限256字符
- 4. 单视频文件类型要求为mp4/mov,单文件大小限制在4M内,视频码率不超过3Mbps,长度限制10s以内
- 5. 您的账户下数据集数量限制为20G视频,如果需要提升数据额度,可在平台提交工单
- json平台通用



导入要求如下:

- 1. 上传已标注文件要求格式为zip格式压缩包,同时压缩前源文件大小在5GB以内
- 2. 压缩包内需要包括视频源文件 (mp4/mov) 及同名的json格式标注文件
- 3. 标注文件中标签由数字、中英文、中/下划线组成,长度上限256字符
- 4. 单视频文件类型要求为mp4/mov,单文件大小限制在4M内,视频码率不超过3Mbps,长度限制10s以内
- 5. 您的账户下数据集数量限制为20G视频,如果需要提升数据额度,可在平台提交工单

视频内容要求:

1、训练视频和实际场景要识别的视频拍摄环境一致,举例:如果实际要识别的视频是摄像头俯拍的,那训练视频就不能用网上下载的目标正面 视频

2、每个视频需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强

未标注数据上传目前支持本地导入、BOS目录导入、分享链接导入、平台已有数据集导入,4种导入方式。 导入数据的要求为:

- 1. 压缩包仅支持zip格式,压缩前源文件大小限制5GB以内
- 2. 单视频文件类型要求为mp4/mov,单次上传限制10个文件
- 3. 单个视频文件大小限制在4M内,视频码率不超过3Mbps,长度限制120min
- 4. 分辨率大于1080P的视频会被压缩至1080P,编码格式不是h264格式的视频会被转为h264格式
- 5. 您的账户下数据集数量限制为20G视频,如果需要提升数据额度,可在平台提交工单

导入完成后可在平台完成在线数据标注。如有任何问题,请提交工单联系我们

视频内容要求:

1、训练视频和实际场景要识别的视频拍摄环境一致,举例:如果实际要识别的视频是摄像头俯拍的,那训练视频就不能用网上下载的目标正面 视频

2、每个视频需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强

心 视频分类数据集管理API

目录

- 1. 数据集管理API介绍
- 2. 数据集创建API
- 3. 查看数据集列表API
- 4. 查看分类 (标签) 列表API
- 5. 添加数据API
- 6. 数据集删除API
- 7. 分类 (标签) 删除API

8. 错误码

数据集管理API介绍

本文档主要说明当您线下已有大量的已经完成分类整理的视频数据时,如何通过调用API完成视频数据的便捷上传和管理。 EasyDL数据集管理API 在管理不同模型数据类型之间是通用的。上传不同模型类型数据,只是在部分接口入参存在差异,使用及接口地址完全一致。

数据集创建API

接口描述

该接口可用于创建数据集。

接口鉴权

同发布模型后获取的API鉴权方式:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/create

URL参数:

参数	值	
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"	

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, VIDEO_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、视频分类、文本分类
dataset_ name	是	string	数据集名称,长度不超过20个utf-8字符

若上传视频分类数据集,在type参数应传「VIDEO_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id ,用于问题定位
dataset_id	否	number	创建的数据集ID

查看数据集列表API

接口描述

该接口可用于查看数据集列表。返回数据集的名称、类型、状态等信息。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/list

URL参数:

参数		值
access	token	通过API Kev和Secret Kev获取的access token.参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, VIDEO_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、视频分类、文本分类
start	否	number	起始序号,默认为0
num	否	number	数量,默认20,最多100

若查看视频分类数据集,在type参数应传「VIDEO_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	数据集总数
results	否	array(object)	数据集列表
+dataset_id	否	number	数据集ID
+dataset_name	否	string	数据集名称
+type	否	string	数据集类型
+status	否	string	数据集状态
+special_status	否	string	数据集特殊状态,包括shared、smart和空值,分别表示共享中、智能标注中、非特殊状态

查看分类(标签)列表API

接口描述

该接口可用于查看分类(标签)。返回分类(标签)的名称、包含数据量等信息。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/list

URL参数:

参数	值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, VIDEO_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、视频分类、文本分类
dataset _id	是	number	数据集ID
start	否	number	起始序号,默认0
num	否	number	数量,默认20,最多100

若查看视频分类的全部分类,在type参数应传「VIDEO_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
total_num	否	number	标签总数
results	否	array(object)	标签列表
+label_id	否	string	标签/分类ID
+label_name	否	string	标签/分类名称
+entity_count	否	number	图片/声音/文本数量

添加数据API

接口描述

该接口可用于在指定数据集添加数据。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/addentity

URL参数:

参数	值

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, VIDEO_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、视频分类、文本分类
dataset _id	是	number	数据集ID
append Label	否	boolea n	确定添加标签/分类的行为:追加(true)、替换(false)。默认为追加(true)。
entity_c ontent	是	string	type为 IMAGE_CLASSIFICATION/OBJECT_DETECTION/IMAGE_SEGMENTATION/SOUND_CLASSIFICATION/VIDEO_CLASSIFICA TION时,填入图片/声音/视频的base64编码;type为TEXT_CLASSIFICATION时,填入utf-8编码的文本。内容限制为: 图像分类base64前10M;物体检测base64前10M;图像分割base64前10M;声音分类base64前4M,声音时长 1~15秒;视频分类base64前4M,时长10秒内;文本分类10000个汉字
entity_n ame	是	string	文件名
labels	是	array(o bject)	标签/分类数据
+label_ name	是	string	标签/分类名称(由数字、字母、中划线、下划线组成),长度限制20B
+left	否	number	物体检测时需给出,标注框左上角到图片左边界的距离(像素)
+top	否	number	物体检测时需给出,标注框左上角到图片上边界的距离(像素)
+width	否	number	物体检测时需给出,标注框的宽度(像素)
+height	否	number	物体检测时需给出,标注框的高度(像素)

若上传视频分类数据集,在type参数应传「VIDEO_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

数据集删除API

接口描述

该接口可用于删除数据集。

接口鉴权

同模型上线后获取的API:

- 1、在EasyDL——控制台创建应用
- 2、应用详情页获取API Key和Secret Key

请求说明

请求示例

HTTP 方法:POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/dataset/delete

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, VIDEO_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、视频分类、文本分类
dataset _id	是	number	数据集ID

若刪除视频分类数据集,在type参数应传「VIDEO_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

分类 (标签) 删除API

接口描述

该接口可用于删除分类(标签)。

接口鉴权

同模型上线后获取的API:

1、在EasyDL——控制台创建应用

2、应用详情页获取API Key和Secret Key

请求说明

请求示例

```
HTTP 方法: POST
```

Baidu 百度智能云文档

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/label/delete

URL参数:

参数	值
	济社ADLK, 和O

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

字段	必选	类型	说明
type	是	string	数据集类型,可包括: IMAGE_CLASSIFICATION, OBJECT_DETECTION, IMAGE_SEGMENTATION, SOUND_CLASSIFICATION, VIDEO_CLASSIFICATION, TEXT_CLASSIFICATION 分别对应:图像分类、物体检测、图像分割、声音分类、视频分类、文本分类
dataset _id	是	number	数据集ID
label_n ame	是	string	标签/分类名称

若删除视频分类的子类,在type参数应传「VIDEO_CLASSIFICATION」

返回说明

返回参数

字段	必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
   "error_code": 110,
   "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

Baidu 百度智能云文档

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求 ,如果持续出现此类错误,请在百度云控制台内提交工单反馈
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量 阶梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶 梯计费;未上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
406000	internal server error	服务错误
406001	param[xx] invalid	参数xx不合法,请检查相关参数
406002	dataset not exist	数据集不存在
406003	dataset already exists	数据集已存在
406004	dataset can not be modified temporarily	数据集暂不可修改
406005	label not exist	标签/分类不存在
406006	no permission to modify the dataset	没有修改数据集的权限
406007	dataset cannot be modified while smart annotation is running	智能标注期间不可修改数据集
406008	quota exceeded	配额超限

模型训练

の 模型训练操作说明

在完成上传数据集后,在左侧列表中点击**训练模型**进入训练模型页面,如下图所示

视频分类模型 ▲	训练模型
品 总览	
由 模型中心	选择模型
我的模型	训练配置
创建模型	部署方式 公有云部署
训练模型	选择算法 • 默认
校验模型	添加数据
发布模型	添加数据集 + 请选择
🖉 EasyData数据服务	
数据总览	开始训练
标签组管理	
在线标注	
目 公有云服务	
在线服务	

选择上传的数据集和相应的标签,如下图所示

视频分类模型		添	加数据	集		×	常见问题	教学视频		
器 总览										
曲 模型中心	可选项 ②	全部添加		已选项	全部删除					
我的模型	▲ 手语识别 V1	~		へ 手语识别 V1	×					
创建模型	Sci-Fi	~		Sci-Fi	×					
2014年1月1	 Thriller	~		Thriller	×					
WILKIN (B) LEE	Crime	~		Crime	×					
校验模型	Drama	~		Drama	×					
发布模型	✓ jay_test_video_cls V1	+	>							
☑ EasyData数据服务	✓ zyh-视频分类 V1	+								
教理总监	✓ dxu_video_test V1									
W NUMBER OF	✓ zxx-test V3									
标签组管理	✓ zxx-test V1									
在线标注	✓ video V2									
1 公有云服务	✓ yhh V1									
+400 A	✓ yyw-video V1	+								
住残服労					你已选择1个数据集的4个标	笠				
		12-2-2-2** 1 × X38-第2** 1 * 5-35 協定 取消								



視頻分类模型 《	≡	训练模型				操作文档	常见问题	教学视频	提交工单
HH 总览									
曲 模型中心		选择模型	手语识别						
我的模型		训练配置							
创建模型		部署方式	公有云部署						
训练模型		选择算法	○ 默认						
校验模型		添加数据							
发布模型		添加数据集	上 请洗择						
☑ EasyData数据服务		10000 SA 311 PR	1 HUKEDT						
数据总览			数据集	版本	分类数量		操作		
标签组管理			手语识别	V1	4		移除		
在线标注		22.64.30145							
目 公有云服务		21 XEL UTSAK							
在线服务									

点击开始训练后自动跳转到我的模型开始训练,训练过程中推荐打开短信通知,如下图所示,这样模型训练好后我们将第一时间以短信的方式告

	视频分类模型	=	模型列表					操作文档	常见问题	教学视频	提交工单
知您。	88 总览										
	曲 模型中心		创建模型								
	我的模型		【御師公类】王语识别	(3) 历史版本			面 删除				
知您。	创建模型								C Michael E and		
知您。	训练模型		部署方式	版本	训练状态	服务状态	模型效果		操作		
	校验模型		公有云API	V1	•训练中 💶	未发布	-		查看版本配	置 终止训	练
	发布模型			บ	练进度:	1%					
	☑ EasyData数据服务		【视频分类】jay_test_13	8.5 🖸 模型ID: 17	完成后短信提醒至 135**	***919		8 i	训练 🕒 🛙	历史版本	面 删除
	8611月1日15										

训练完成后,如下图所示,可以查看模型的完整评估报告,上传一些视频在线校验模型。

模型中心	模型列表								
我的模型	创建模型	创建模型							
创建模型									
训练模型	【視頻分类】:	【视频分类】手语识别 模型D: 111							
校验模型	应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作		
发布模型	一服务	1/1	训练地成	土由法	主告布	top1准确率94.67%	电速发车 姑恐 训练		
数据中心	2110,95	vi	witsp.76.06	小中市	AP 96.10	完整评估结果	中间及中口发展 机线		

平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

优惠政策:

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有视频分类操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。

∞ 模型效果评估报告

简介

模型训练完成后,模型列表中可以看到模型的结果,包括两个指标:top1准确率和top5准确率,也可以点击「完整评估报告」查看更为详细的模型表现,包括准确率、F1-score、精确率和召回率,本文档会介绍如何解读模型的各项指标。

模型训练结果

模型的训练结果是如何得到的?

所有训练数据中,系统会随机抽取70%的标注数据作为训练数据,剩余的30%作为测试数据,训练数据训练出的模型去对测试数据进行检测,检 测得到的结果跟人为标注的结果进行比对,得到准确率、F1-score、精确率和召回率。

提示:训练数据,即上传的视频越接近真实业务里需要预测的视频,模型训练结果越具有参考性。

在查看模型评估结果可能需要思考在当前业务场景精确率与召回率更关注哪个指标,是更希望减少误识别,还是更希望减少误召回。前者更需要 关注召回率的指标,后者更需要关注精确率的指标。同时F1-SCORE可以有效关注精确率和召回率的平衡情况,对于希望召回与识别效果兼具的 场景,F1-Score越接近1效果越好。

完整评估报告

如果需要了解更为详细的模型效果表现,可以在模型列表中点击三项指标下方的「完整评估报告」,完整评估报告页面如下图所示:

Baidu 百度智能云文档

模型管理 > 手语识别1
▌ 手语识别1评估报告
应用类型: 云服务(目前仅支持) 🗸 版本: V1 🗸 视频数: 557 分类数: 3
▶ 整体评估
手请识别1 V1效果优异 如何优化效果?
准确率 ⑦ F1-score ⑦ 精确率 ⑦ 召回率 ⑦
98.2% 97.1% 98.4% 96.1%
1 洋细亚/士
评估样本具体数据情况
随机测试集
正确数量: 327
预测表现 错误数量: 6
冬个记到任里姝施家效里
top1 98%
top2 99%
top3 100%
评估报告
如下图所示:
▌ 手语识别1评估报告
应用类型: 云服务(目前仅支持) 💙 版本: V1 🗸 视频数: 557 分类数: 3
在这部分可以选择模型的版本,以及看到每个版本参与训练的视频数和分类数。
整体评估
如下图所示:
1 49 / t+ 30 / t+
准确率 ⑦ F1-score ⑧ F1-score ⑦ F
3 0.270 9 .170 9 .170 9 .170
在这部分 四项指标的会议加下 ·
● 准确率
准确率 Accuracy = 模型正确预测所有分类的数量/所有分类客观存在的数据总数
对于一个分类模型而言,准确率表示这个模型中所有分类的综合识别效果。 如果准确率为1,说明所有分类在测试数据中都被正确识别

• F1-score

F1-score是模型中一个分类的精确率和召回率的调和平均数,对于希望召回与识别效果兼具的场景,F1-Score越接近1效果越好。

● 精确率

精确率 Precision = 模型正确预测为该分类的数量/模型预测为该分类的总数

对于一个分类而言,精确率越高,说明模型识别出是这个分类的所有结果中,正确数量的占比越高。**如果精确率为1,说明识别出的所有结果都**

是对的,但不说明该分类全部都被识别出来了,可能会存在漏识别。

 - 召回率

召回率 Recall = 模型正确预测为该分类的数量/该分类客观存在的数据总数

对于一个分类而言, 召回率越高, 说明模型越完整地识别出这个分类。**如果召回率为1, 说明这个分类全部都别模型识别出来了, 但不表示识别** 出是这个分类的结果都是对的, 可能会存在误识别。

如果对模型的效果有疑问,可以点击「如何优化效果」查看模型效果不佳的原因,如下图所示:

咨询如何优化效果		×
* 问题名称:	部分分类效果较差 🗸	
	提交咨询前,建议先查看 效果优化参考文档	
其他要求:	请输入其他要求	
	0/2000	1.
	提交	

将您想咨询的问题描述填写后提交,我们线下会有专员联系到您帮您解决问题。

详细评估

如下图,这里可以看到所有评估报告的数据是基于什么量级的数据进行计算的,当整体参与评估的数量较少时,所有数值可能无法真实反映模型 效果。同时,可以看到模型多个识别结果时的准确率。

| 详细评估

评估样本具体数据情况

	随机测试集
预测主现	正确数量: 327
"则测衣现	错误数量:6

多个识别结果准确率效果



模型发布

の 模型发布整体说明

训练完成后,可将模型部署在公有云服务器上,通过API进行调用。

公有云API

- 训练完成的模型存储在云端,可通过独立Rest API调用模型,实现AI能力与业务系统或硬件设备整合
- 具有完善的鉴权、流控等安全机制,GPU集群稳定承载高并发请求

相关费用

将模型发布为API后,将获得一部分免费调用次数,如需更多调用量,请在百度云控制台内提交工单反馈。

^の 发布为API

^の如何发布为API

在完成模型训练后,可将训练好的模型发布为可调用的服务API。

模型中心	发布模型		
我的模型	1 选择模型:	手语识别	~
创建模型	 应用类型: 	云服务	\checkmark
训练模型	3 选择版本:	V1	\sim
校验模型	4 * 服务名称:		
发布模型	5 * 接口地址:	https://aip.baidubce	e.com/rpc/2.0/ai_custom/
数据中心		vi/ video_cis/	
我的数据集	6 其他要求:	若接口无法满足您 决的问题,500汉	的需求,请描述希望解 字以内
			0/500
		提交申请	

点击模型列表内对应模型「操作」列中的「申请发布」,或是在左侧导航栏点击发布模型可以进入发布模型页面,如上图所示。在对应选项中选 择和输入相应内容发起模型发布的申请:

1. 选择模型 (必选)

选择需要发布的模型,只能选择已经完成训练的模型

2. 选择服务

视频分类仅支持发布为云服务API

3. 选择版本 (必选)

选择需要发布的模型版本,只能选择完成训练且没有发布过的版本

4. 服务名称 (必填)

为发布的服务命名,服务名称不得多于20个字符

- 5. 接口地址(必填) 自定义服务的API URL, 接口地址需要多于5个字符但不能超过20个字符,仅限英文
- 6. 其他要求
 - 如果有其他要求可以输入要求描述

填写完上述信息后,点击「提交申请」完成发布模型申请。提交申请后,模型列表内该模型的申请状态和服务状态为有以下几种情况:

申请状态	服务状态	状态描述
审核中	未发布	服务刚申请发布,模型在审核中
审核成功	发布中	服务通过审核,进入系统自动发布阶段
审核成功	已发布	服务发布成功
审核失败	未发布	服务未通过审核,通常为模型训练结果mAP < 0.6,如需申诉,请在百度云控制台内提交工单反 馈

提示:第一次申请发布的模型需要人工审核,通常4小时内完成,如果希望加急上线,请在百度云控制台内提交工单反馈。非第一次申请发布的模型,如果模型训练结果mAP>0.6,训练数据量大于20,则会自动通过审批。审批完成后,大约需要5分钟左右自动完成发布。

EasyDL 视频使用说明

Baidu 百度智能云文档

服务详情			
服务名称:	silence手语识别01		
模型版本:	V1		
接口地址:	https://aip.baidubce.com/rpc/2.0/	ai_custom/v1/video	_cls/silence_shouyu1
服务状态:	已发布		
	立即使用	查看API文档	

点击「查看API文档」可以快速跳转至API文档,参考文档调用API获取视频分类AI能力。

心 视频分类API调用文档

简介

本文档主要说明视频分类模型发布后获得的API如何使用,如有疑问可以通过以下方式联系我们:

• 在百度云控制台内提交工单,咨询问题类型请选择人工智能服务

• 进入EasyDL社区交流,与其他开发者进行互动

接口鉴权

1、在EasyDL——控制台创建应用

♥ 全局 ~									Q	=	₽	?	몲	۲
EasyDL定制训练平台	产品服务 / EasyDL定制训	练平台 - 概览												
概览	概览													
应用列表	应用			用量							法法保险	10760 200	19 00 26	2018 00 26
监控报表											明石作职	TINIEX 20	18-09-26	- 2018-09-26
技术文档				API		调用量		调用失	敗	失	敗率		详细绣	51+
云服务权限管理				百美椅子训练	ξ.	0		0		0%			查看	
嘉純SDK管理	已刻	些回应用:0个	:0个		J	0		0		0%			查看	
(1) (0) - 0 - 12 - 12 - 12 - 12 - 12 - 12 - 12	前定应用			aa		0		0		0%			查看	
				识别眼睛和瞬	巴	0		0		0%			查看	
				声音分类		0		0		0%			查看	
	可用服务列表													
	API	模型ID	模型	类型	模型名称		模型版本		状态		调用量	限制	QP	S限制
	百美椅子训练	230	图像	分类	百美椅子	训练	V1		● 免费使用		500次/	/天免费	不住	呆证并发
	椅子颜色识别	2598	图像:	分类	椅子颜色	识别	V1		● 免费使用		500次/	天免费	不信	呆证并发
	aa	3025	图像	分类	test0208	1	V3		● 暫未配置权 ?	限	500次/	天免费	不住	呆证并发
	识别眼睛和嘴巴	8086	物体	检测	识别眼睛	和嘴巴	V3		● 免费使用		500次/	天免费	不信	呆证并发
	声音分类	10789	声音	分类	动物声音	分类	V3		● 暫未配置权	R	500次/	天免费	不住	呆证并发

Baidu 百度智能云文档

1

♥ 全局 ∨							Q	⊞	ø	?	8	۲	Q
EasyDL定制训练平台	产品服务 / EasyDI	L定制训练平台 - 应	用列表 / 创建应用										
概览	创建新应用												
应用列表	• 应田夕称•												
监控报表	775143371494	test											
技术文档	•应用类型:	游戏娱乐				~							
云服务权限管理	•接口选择:	您可以为应用勾	选多个接口权限,使此	应用可以请求已勾选的排	接口服务								
离线SDK管理		EasyDL	✓ 百美椅子训练✓ 识别眼睛和嘴目	 ✓ 椅子顔色识易 巴 ✓ 声音分类 		aa 商品检测							
		➡ 百度语音											
		文字识别											
		➡ 人脸识别											
		➡ 自然语言处理	8										
		➡ 内容审核]										
		🕂 UNIT 📘											
		✤ 知识图谱											
		+ 图像识别											
		+ 智能呼叫中/	D										
		 ・ ・											
		土 人体分析											
	•应用描述:	简单描述一下: 点,请控制在	您使用人工智能服务的 500字以内	应用场景,如开发一款;	美颜相机,需	要检测人脸关键							

2、应用详情页获取AK SK

♥ 全局 ~					Q	≣	₽	?	8	۲	Q ~
EasyDL定制训练平台	产品服务 / EasyDL定制训练平台 - 应用列表 / 应F	◎品服务 / EasyDL定制训练平台-应用列表 / 应用详情									
概览	应用详情	应用详情									
应用列表	编辑 查看文档										
监控报表	应用名称	AppID		API Key			Secr	et Key			
技术文档	test2	14293933		bZsfqO2fUdHTDflxqF1	1mkTq			*** 显示			
云服务权限管理 离线SDK管理	EasyDL定制化训练服务:										

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在定制视频分类训练平台进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

参数	值	
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"	

Header如下:

参数 值 Content-Type application/json

注意:如果出现336001和336002的错误码很可能是因为请求方式错误,与其他图像识别服务不同的是定制化图像识别服务以json方式请求。

Body请求示例:

٤	
	"video": " <base64数据>",</base64数据>
	"top_num": 6
ł	

Body中放置请求参数,参数详情如下:

参数	是否必 选	类型	可选值范 围	说明
video	是	string	-	视频,base64编码,建议视频码率不超过3Mbps,长度不超过10s,base64编码后大小不超过4M,支持 mp4, mov格式
top_nu m	否	number	-	返回分类数量,默认为6个

请求代码示例

提示一:使用示例代码前,请记得替换其中的示例Token、视频地址或Base64信息。

```
提示二:部分语言依赖的类或库,请在代码注释中查看下载地址。
```

Python3	
	-
EasyDL 视频分类 调用模型公有云API Python3实现 """	
import json	
import base64	<u> </u>
import requests	

使用 requests 库发送请求	
使用 pip(或者 pip3)检查我的 python3 环境走谷安装了这阵,执行命令	
pip freeze grep requests	
关诉同位为空间安排这座	
石层凸值为工,对实权以片	
pip install requests	
non	
日标视频的 本地文件路径,支持mp4_moy格式	
VIDEO_HILEPATH = "【您的测试优频地址,例如:./example.mp4"	_
	•

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码,当请求错误时返回
error_msg	否	string	错误描述信息,当请求错误时返回
results	否	array(object)	分类结果数组
+name	否	string	分类名称
+score	否	number	置信度

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

{
 "error_code": 110,

- "error_msg": "Access token invalid or no longer valid"
- }

需要重新获取新的Access Token再次请求即可。

1 Unkno 2 Service unavai	wn error e temporarily ilable	服务器内部错误,请再次请求 ,如果持续出现此类错误,请在百度云控制台内提交工单反馈。 服务暂不可用,请再次请求 ,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2 2 unavai	e temporarily ilable	服务暂不可用,请再次请求,如果持续出现此类错误,请在百度云控制台内提交工单反馈。
	norted energy	
3 3 methor	id	调用的API不存在,请检查后重新尝试
4 Open a reache	api request limit ed	集群超限额
6 No per data	rmission to access	无权限访问该用户数据
13 Get se	ervice token failed	获取token失败
14 IAM Ce	ertification failed	IAM鉴权失败
15 failed	ot exsits or create	应用不存在或者创建失败
0pen a 17 limit re	api daily request eached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18 Iimit re	api qps request eached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19 limit re	api total request eached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100 Invalid	l parameter	无效的access_token参数,请检查后重新尝试
110 Access no long	s token invalid or ger valid	access_token无效
111 Access	s token expired	access token过期
336000 Interna	al error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001 Invalid	l Argument	入参格式有误,比如缺少必要参数、视频base64编码错误等等,可检查下视频编码、代码格式是否有误。有 疑问请在百度云控制台内提交工单反馈
336002 JSON7	不合法	入参格式或调用方式有误,比如缺少必要参数代码格式是否有误。有疑问请在百度云控制台内提交工单反馈
336003 Base6	34解码失败	图片/音频/文本/视频格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去 掉头部。有疑问请在百度云控制台内提交工单反馈
336004 输入文	[件大小不合法	视频超出大小限制,图片限20M以内,请根据接口文档检查入参格式,有疑问请在百度云控制台内提交工 单反馈
336005 视频解	码失败	视频编码错误(非MP4/MOV视频格式),请检查并修改视频格式
336006 缺失必	多多数	video字段缺失(未上传视频)
336100 model unavai	temporarily ilable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

常见问题

心训练相关问题

数据处理失败或者状态异常怎么办?

请检查已上传的分类命名是否正确,例如是否存在中文命名、或者有空格,如果自查没有发现问题请在百度云控制台内提交工单反馈。

模型训练失败怎么办?

如果遇到模型训练失败的情况,请在百度云控制台内提交工单反馈。

已经上线的模型还可以继续优化吗?

已经上线的模型依然可以持续优化,操作上还是按照标准流程在训练模型中-选择要优化的模型和数据完成训练,然后在模型列表中更新线上服务,完成模型的优化。

心 模型效果相关问题

模型效果怎么调优?

当效果不满意时,请查看您的训练数据是否和实际场景中要识别的图片一致,以及训练数据量是否太少。如果训练数据量已经达到一定丰富度, 例如单个分类/标签的视频超过50个,效果却仍然不佳,请在百度云控制台内提交工单反馈。

∞ 模型上线相关问题

希望加急上线怎么处理?

请在百度云控制台内提交工单反馈。

每个账号可以上线几个模型?是否可以删除已上线的模型?

每个账号最多申请发布10个模型,已上线模型无法删除。

心 收费相关问题

接口上线后是否收费?调用量不够怎么办?

目前是限量免费使用的原则,上线模型后可免费获得500次/天,qps=1的调用限额。**如有超过这个量的需求,请在百度云控制台内提交工单反** 馈。费用问题不用太过担心,如果所需要的量级非常高,可能会基于实际要求适当收费。价格可以根据其他已推出的图像等能力的价格作为参 考。

心 其他问题

模型能否支持私有化部署?

目前我们提供的方案支持公有云API在线调用,尚不支持离线SDK和服务端的私有化部署。

申请发布模型审核不通过都是什么原因?

可能为以下原因: 1、当前您的模型存在一些问题,如训练数据异常、数据量不够,或者您不想再继续使用等一系列原因,我们会和您通过电话 沟通,沟通达成一致后再拒绝。 2、您的电话未接通且模型效果较差,会直接拒绝。如果需要申诉,请在百度云控制台内提交工单反馈。

目标跟踪

目标跟踪介绍

心 功能介绍

目标跟踪是指对视频流中的特定运动对象检测识别,获取目标的运动参数,从而实现对后续视频帧该对象的运动预测(轨迹、速度等),实现对 运动目标的行为理解。

⊙ 应用场景

目标计数:流水线上特定产品的数量统计;商场、旅游景点的人流统计等

智能化交通:人流、车流分析;行人运动轨迹预测;交通违规抓拍等

人/动物的轨迹分析:监控摄像下的行人可疑移动轨迹分析;养殖场动物移动轨迹监测等

应用示例1,飞机轨迹跟踪:



应用示例2,生猪行为分析:



の 开始使用

Step 1 创建模型

确定模型名称,记录希望模型实现的功能

Step 2 上传并标注数据

前往数据总览页面上传数据,并在线标注数据

Step 3 训练模型并校验效果

选择部署方式与算法,用上传的数据一键训练模型

模型训练完成后,可在线校验模型效果

Step 4 发布模型

将模型以本地部署的方式发布使用

更详细的操作指导,请参考左侧导航栏中各步骤的操作文档

创建模型

进入创建模型页面

在EasyDL视频点击【立即使用】按钮后,选择目标跟踪到目标跟踪模型训练页,下面会出现两种情况:

第一种,如果您没有登录百度云,则会跳转到百度云登录页面,没有百度账户的客户请先注册百度账户。登录后,会跳转到模型概览页,点击 「视频分类」卡片上的「点击前往」按钮,会跳转模型训练页面的创建模型页。

第二种,如果您已登录,会直接进入到我的模型页,该页面能够管理已经创建的模型,点击左侧列表中的创建模型进入创建模型页面。

创建模型 进入创建模型页面后你会看到如下图中展示的内容

目相	示跟踪 ∢──	模型列表 > 创	建模型		操作文档
88	总览 横型中心	模型类别	目标跟踪		
	我的模型	模型名称 *			
	创建模型	模型归属	公司 个人		
	训练模型		请输入公司名称		
	校验模型	所属行业 *	请选择行业		
	发布模型	7111415	17/2)T () ±		
11	EasyData数据服务	应用场景 🔹	请选择应用场景 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>		
	数据总览	邮箱地址 🔹	z******@baidu.com		
	在线标注	联系方式 *	135****919	0	
13	EasyEdge本地部署	Th的描述 。			
	纯离线服务	-7JRC10020			
			0/500		
			07500		
			完成		

需要填写的项目如下:

模型名称

模型的名称

模型归属
 模型是属于公司的,还是属于个人的,如果是前者,请填写公司名称

Baidu 百度智能云文档

所属行业
 请选择您公司所属的行业

● 应用场景

请选择模型将会被应用于的业务场景

• 邮箱地址

用于联系到您的邮箱地址

• 联系方式

有效的联系方式将有助于后续模型上线的人工快速审核,以及更快的百度官方支持,推荐填写个人手机号码

• 功能描述

描述改模型将要应到的业务场景,详细的描述,在获取官方支持时,能帮助我们为您提供准确的使用建

像下图展示的一样完成所有填写项后点击【下一步】按钮完成模型创建,创建完成后会跳转到我的模型页面。

目标跟踪 ▲	模型列表 > 创建模型
器 总览	模型类别 目标跟踪
<u></u> 位 模型中心	模型名称 * 车辆识别跟踪
我的模型	
创建模型	模型归属 公司 个人
训练模型	百度
校验模型	所属行业 * 交通出行 >
发布模型	
☑ EasyData数据服务	应用场景 * 视频流中的目标移动轨迹分析 >
数据总览	邮箱地址 * z******@baidu.com
在线标注	联系方式 * 135*****919 ⑦
記 EasyEdge本地部署	
纯离线服务	功能描述 * 识别交通出行场景中的违章车辆并跟踪轨迹
	19/500
	完成

数据准备

心 创建数据集

创建数据集 在训练之前需要在导航栏的「数据总览」页面【创建数据集】,如下图所示。

目村	示跟踪		我的数据总览 >	创建数据集
88	总览 模型中心		数据集名称	车辆跟踪
	我的模型		数据类型	视频
	创建模型		数据集版本	V1
	校验模型		标注类型	目标跟踪
	发布模型 EasyData数据服务			2 Contraction of the second se
	数据总览			
	在线标注			
멉	EasyEdge本地部署			
	纯离线服务			
点击	【完成】,成功创建	数据集		

目标跟踪	<=	我的数据总》	g							提交工单
88 总览										
曲 模型中心		EasyData	智能数据服务平台已	上线,使用EasyDa	ta可享受包括多人标注、	数据清洗、数据采集等完整数	据服务 立即前往			
我的模型		创建数据制	5						输入数据集4	S称或ID Q
创建模型										
训练模型		车辆跟踪	公 数据集组ID: 3	800280				□ 新增版本	믬 全部版本	<u> 創除</u>
校验模型		版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	操作		
发布模型							001 (0 (0)			
☑ EasyData数据服务		VI 💬	323894	U	● ビ元成	日标跟踪	0% (0/0)	守入 删除		

心 上传数据集

上传目标跟踪数据集已标注数据上传基于CVAT标注好的数据以MOT1.1的数据集形式导出,上传数据压缩包:

- 压缩包仅支持zip格式,大小限制5GB以内
- 压缩包内单个视频长度限制在10分钟内,至少应上传4个视频标注压缩包
- 从CVAT导出的标注数据压缩包可多次上传一起导入数据集组。也可通过本地解压再添加到同一个文件夹后压缩上传

我的奴据总觉。	1111111/V1/导入				
┃ 创建信息 〜				常见问题	
数据集ID	4420	版本号	V1	1、如何进行数据标注?	\sim
备注	ß			2、已标注数据的上传压缩包格式要求	\sim
┃标注信息 〜					
标注类型	目标追踪	标注模板	目标追踪		
数据总量	0	已标注	0		
标签个数	0	目标数	0		
待确认	0	大小	ОМ		
数据清洗 暂未做过数据清: 导入数据 数据标注状态 导入方式 上传压缩包	先任务 ① 有标注信息 本地导入 ① 上传压缩包 ④ MOT16-02_split. ④ MOT16-05_split. ④ MOT16-11_split.; ④ MOT16-11_split.; 通认并返回	已上传3个文件 zip X -0.zip X zip X	~		

如有任何问题,请提交工单联系我们

视频内容要求:

1、训练视频和实际场景要识别的视频拍摄环境一致,举例:如果实际要识别的视频是摄像头俯拍的,那训练视频就不能用网上下载的目标正面 视频

2、每个视频需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强

如果需要寻求第三方数据采集团队协助数据采集,请在百度云控制台内提交工单反馈

未标注数据上传目前支持本地导入、BOS目录导入、分享链接导入、平台已有数据集导入,4种导入方式本地导入的要求为

- 1. 压缩包仅支持zip格式,压缩前源文件大小限制5GB以内
- 2. 单视频文件类型要求为mp4/mov,单次上传限制10个文件
- 3. 单个视频文件大小限制在3G内,视频码率不超过3Mbps,长度限制120min
- 4. 分辨率大于1080P的视频会被压缩至1080P,编码格式不是h264格式的视频会被转为h264格式
- 5. 您的账户下数据集数量限制为20G视频,如果需要提升数据额度,可在平台提交工单

导入完成后可在平台完成在线数据标注。如有任何问题,请提交工单联系我们

视频内容要求:

1、训练视频和实际场景要识别的视频拍摄环境一致,举例:如果实际要识别的视频是摄像头俯拍的,那训练视频就不能用网上下载的目标正面 视频

2、每个视频需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强

心 在线标注

目标跟踪在线标注 在创建好数据集,并导入视频数据后。可点击数据总览页面,上传数据对应的「查看与标注」操作开始标注任务。

Baidu 百度智能云文档

目标跟踪 → 三	我的数据总览						提交工单
凸 模型中心							
我的模型	EasyData智行	能数据服务平台已上线,傅	用EasyData可享受包括	多人标注、数据清洗、数据采集等	完整数据服务 立即前往		
创建模型	创建数据集						输入数据集名称或ID Q
训练模型							
校验模型	ztf 区 数	居集组ID:157573					□新埔版本 铝全部版本 向無除
发布模型	版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	操作
🖾 EasyData数据服务	V1 ⊖	162703	0	 已完成 	目标追踪	0% (0/0)	导入 删除
<u> </u>							
在线标注	sdasdasd 🖸	」 数据集组ID: 157566					□新增版本 記全部版本 □删除
13 模型部署	版本	数据集ID	数据量	晨近星) 好态	标注类刑	标注试态	海作
纯高线服务	104.4-		100, per 200.	ARK 107 10700	1074674.00	10726-0730	, MIE
	V1 ⊖	162695	7	 已完成 	目标追踪	100% (7/7)	查看与标注 导入 删除
							1
							< 1 > 10条/页 ~
						,	< 1 > 10 条页 >

在线标注视频示意

在线标注图例

1.添加标签

点击标注页面右上角的「添加标签」,输入标签名称来添加标签。标签添加成功后可在右侧标签栏查看所添加的标签,并提供「修改标签名称」、「连续标注」的功能

在线标注 > sdasdasd > 标注					
全部(8) 无标注信息	(1) 有标注信息(7)		查看标注说明	标签栏	添加标签 🗸
pexels-kelly-lacy-5473	765.mp4		保存当前标注(S)	people	快援键 1
			>	标注结果	
00:00/00:21	◀ ▶ │ ◀◀ ▶ ▶▶ │ 報該	文 1 <u>~</u> / 642	倍速 1X 📢		
			+ 淡加文件		

2.标注目标第一关键帧

在视频画面中框选出所要识别的目标,并选择标签



3.标注目标第二关键帧

在标注完目标第一关键帧之后,播放视频。待目标移动一段距离后,暂停视频并选定标注框,将标注框拖动至目标当前位置,标注目标第二关键 帧。此时即完成了目标第一关键帧到第二关键帧之间帧的全部标注,可拖动进度条查看。重复关键帧的标注操作,来完成目标出现在视频画面中 的全部标注。

全部(;	8) 无标注信息(1)	有标注信息(7)			查看标注说明	标签栏	添加标签
pexels	s-kelly-lacy-5473765.	mp4		ß	限存当前标注(S)	car	快捷键
						people	快總羅〔
<					>	标注结果 car_1	1 A ⊚ (
00:	01/00:21		● 較数 45 💮 / 64	2	倍速 1X ■)		

注:如目标移动轨迹不规律、过长等情况下,需要分段标注多个关键帧,来保证目标关键帧之间标注准确性

4.标注「目标消失帧」

目标即将要消失在画面中时,需要标注目标消失帧,从而完成目标的整个标注

注:如目标一直存在在视频画面中,则无需进行目标消失操作



模型训练

心 模型训练操作说明

数据提交后,可以在导航中找到【训练模型】,选择此次训练的模型,并按以下步骤操作,启动模型训练:

目标跟踪	.≡	训练模型				操作文档	常见问题	教学視频	提交工单	
器 总览										
曲 模型中心		选择模型	车辆识别跟踪							
我的模型		训练配置								
创建模型		部署方式	EasyEdge本地部署							
训练模型		选择设备	● 服务器 ○ 通用小型	g备 💿 专项适配硬件						
校验模型		VH #17 007 2+								
发布模型		迈拜昇法	U BAIA							
☑ EasyData数据	服务	添加数据								
数据总览		添加数据集	+ 请选择							
在线标注			数据集	版本	标签数量	1	操作			
記 EasyEdge本地	部署		car_detect	V1	1		移除			
纯离线服务										
		开始训练								

添加数据:

目标跟踪	·=	添加数据集						首 常见问题	教学视频	提交工单
器 总览										
曲 模型中心		可选项 ②	全部添加		已选项	全部删除				
我的模型		✓ 測试目标跟踪 V1	+		∧ car_detect V1	×				
017meth.ml		✓ 示例数据集 V1	+		car	×				
刨建候型		✓ mot15 V1	+							
训练模型		✓ eagle V1	+							
校验模型		✓ car_mot V1	+							
发布模型		✓ visdrone V2	+							
		✓ car_detect V2	+	>						
EasyData数	据服务	✓ car_detect V1	~							
数据总览		✓ motor-179115 V1	+							
在线标注		✓ motor2 V1	+					操作		
認 EasyEdge本地部署	计由感觉	✓ motor1 V1	+							
	- AC-CP-14	✓ car V1	+					移除		
纯离线服务		✓ xyf_dataset_object_test1 V3	+							
		ve und dataset ablant tost1 1/9	±							
	你已选择1个数据集的1									
備定取消										

训练配置

部署方式

目前仅支持「EasyEdge本地部署」
目前支持「服务器部署」、「通用小型设备」以及「专项适配硬件」,其中专项适配硬件支持Jetson(Nano/TX2/Xavier)了解不同方案

选择算法

当前仅支持选择默认算法

添加数据

添加训练数据

添加某单个标签的不同视频数据,在视频数据量为50以上时,模型可以训练充分,得到效果不错的模型。

训练模型

点击「开始训练」,训练模型。

- 训练时间与数据量大小有关
- 模型训练过程中,可以设置训练完成的短信提醒并离开页面,如下图所示

目标跟踪	\in	模型列表				ł	操作文档 常见问题 教	学视频 提交工单
器 总览								
自 模型中心		创建模型						
我的模型		【日标明瞭】车辆识别明瞭	7. 横刑(D: 19028	5			 Б. Б. Ф. Б. 	5本 南 删除
创建模型			C RED. 10020				0 1/12/1	
训练模型		部署方式	版本	训练状态	服务状态	模型效果	操作	
校验模型		本地部署-服务器	V1	• 训练中 🔢	-	-	查看版本配置	终止训练
发布模型			i	川练进度:	1%			
☑ EasyData数据服	务	【目标跟踪】cow 🖸 模	D: 181121	完成后短信提醒至 135	*****919		吕 训练 ③ 历史間	反本 面 删除
数据总览 在线标注		部署方式	版本	训练状态	服务状态	模型效果	操作	
記 EasyEdge本地部 纯离线服务	3署	本地部署-服务器	V1	● 训练完成	-	MOTA: 65.30% MOTP: 68.85% 召回率: 83.40% 完整评估结果	查看版本配置 (校验	申请发布

平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

优惠政策:

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有目标跟踪操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。

⊙ 模型效果评估报告

简介

模型训练完成后,模型列表中可以看到模型的结果,包括三个指标:MOTA、MOTP、召回率,也可以点击「完整评估报告」查看更为详细的模型 表现,本文档会介绍如何解读模型的各项指标。

模型训练结果

模型的训练结果是如何得到的?

所有训练数据中,系统会随机抽取70%的标注数据作为训练数据,剩余的30%作为测试数据,训练数据训练出的模型去对测试数据进行检测,检测得到的结果跟人为标注的结果进行比对,得到MOTA、MOTP、IDF1-score和召回率。

提示:训练数据,即上传的视频越接近真实业务里需要预测的视频,模型训练结果越具有参考性。

Baidu 百度智能云文档

在查看模型评估结果可能需要思考在当前业务场景MOTP与召回率更关注哪个指标,是更希望减少误识别,还是更希望减少误召回。前者更需要 关注召回率的指标,后者更需要关注MOTP的指标。同时IDF1-Score可以有效关注MOTP和召回率的平衡情况,对于希望召回与识别效果兼具的场 景,IDF1-Score越接近1效果越好。

完整评估报告

如果需要了解更为详细的模型效果表现,可以在模型列表中点击三项指标下方的「完整评估报告」,完整评估报告页面如下图所示:



评估报告

如下图所示:



在这部分可以选择模型的版本,以及看到每个版本参与训练的视频数。

整体评估

如下图所示:

▲ 整体评估

手语识别1 V1效果优异 如何优化效果?

准确率 🤉

98.2%







在这部分,四项指标的含义如下:

• MOTA

目标跟踪任务中的MOTA指标,指除误报、丢失目标、ID异常切换情况的正确预测样本占所有样本的比率

对于一个模型而言,MOTA表示这个模型中所有标签的综合识别效果。因效果较差的模型可能存在ID异常切换的情况多,大于样本总数的情况, 所以MOTA的取值可能为负,MOTA∈(-∞, 1]。**如果MOTA为1,说明所有样本在测试数据中都被正确识别** 目标跟踪任务中的MOTP指标,指各个阈值都为默认值0.5的情况下正确预测的目标数与预测目标总数之比

对于一个标签而言,MOTP越高,说明模型识别出是这个标签的所有结果中,正确数量的占比越高。**如果MOTP为1,说明识别出的所有结果都是** 对的,但可能会存在漏识别。

召回率 Recall = 模型正确预测为该标签的ID数量/该标签真实存在的ID总数

召回率越高,说明模型越完整地识别出这个标签。

详细评估

IDF1-score

IDF1-score代表该模型的综合评测效果,越高效果越好。此处为默认平均阈值为0.5时的IDF1-Score

IDSW

代表检测目标的ID异常切换的次数。该值越低,模型鲁棒性越好

┃详细评估

IDF1-score ⑦: 0.59340778488561

ID发生切换的次数

代表检测目标的ID异常切换的次数。该值越低,模型鲁棒性越好

IDSW: 753

按标签查看错误示例

不同标签的MOTA及对应的识别错误的视频

peo 66%	people的错误结果示例 如何解读错误示例?

• 错误示例

可查看按照默认阈值下被判定为错误识别的视频片段样本示例,点击左下角的筛选项可查看正确识别、误识别、漏识别的各个情况



模型发布

心 模型发布整体说明

训练完成后,可将模型部署在本地服务器,通用小型设备以及专项适配硬件

本地服务器部署

- 可将训练完成的模型部署在本地GPU服务器上,支持服务器SDK的集成方式
- 可在内网/无网环境下使用模型,确保数据隐私

通用小型设备

- 通用小型设备部署支持将模型部署在本地的小型计算设备上,提供SDK的集成方式
- 通用小型设备SDK-纯离线服务:支持Windows操作系统,具体的系统、硬件环境支持请参考技术文档。提供相应代码包、说明文档,供企业用 户/开发者二次开发
- 如存在设备无法联网,需要在纯离线的环境下激活的情况,或SDK生成失败等任何其他问题,欢迎提交工单或加入QQ群(679517246)咨询 了解

专项适配硬件

- 为加速开发者们落地离线AI项目,EasyDL提供多种高性价比的软硬一体方案,支持在AI市场直接购买了解更多
- 专项硬件适配SDK和激活序列号是EasyDL软硬一体方案的软件部分。如已在其他渠道购买专项适配硬件,也可在此发布为专用SDK,并单独购 买激活序列号
- の 服务器端SDK
- 心 目标跟踪服务器端SDK简介

本文档主要说明定制化模型发布后获得的服务器端SDK如何使用,如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

SDK说明

目标跟踪服务器端SDK支持Linux操作系统。以下为具体的系统、硬件环境支持:

操作系统 系统支持 硬件环境要求

	CPU: x86 64
	- NVIDIA GPU: x86 64
Linux	HUAWEI Atlas 300:
	x86_64

单次预测时延根据具体设备、线程数不同,数据可能有波动,请以实测为准

激活&使用步骤

离线SDK的激活与使用分以下三步:

① 下载SDK后,在控制台获取序列号

② 本地运行SDK,并完成首次联网激活

通过左侧导航栏查看不同操作系统SDK的开发文档

③ 正式使用

SDK常见问题

通过左侧导航栏查看不同操作系统SDK的FAQ

以下是通用FAQ,如您的问题仍未解决,请在百度云控制台内提交工单反馈

1、激活失败怎么办?

①可能是当前序列号已被其他设备激活使用,请核实序列号后用未被激活的序列号重新激活

②序列号填写错误,请核实序列号后重新激活

③同一台设备绑定同一个序列号激活次数过多(超过50次),请更换序列号后重试

④首次激活需要联网,网络环境不佳或无网络环境,请检查网络环境后重试

⑤模型发布者和序列号所属账号非同一账号,如果存在这种异常建议更换账号获取有效序列号

⑥序列号已过有效期,请更换序列号后重试

⑦如有其他异常请在百度云控制台内提交工单反馈

心目标跟踪服务器端SDK集成文档-Linux-C++

简介

本文档介绍EasyEdge/EasyDL的Linux CPP SDK的使用方法。

- 网络类型支持:图像分类,物体检测,图像分割,目标追踪
- 硬件支持:
 - CPU 基础版: intel x86_64 * AMD x86_64 龙芯 loongarch64 飞腾 aarch64
 - CPU 加速版 Intel Xeon with Intel®AVX2 and AVX512 Intel Core Processors with AVX2 Intel Atom Processors with SSE AMD Core Processors with AVX2
 - NVIDIA GPU: x86_64 PC
 - 寒武纪 Cambricon MLU270
 - 比特大陆计算卡SC5+
 - 百度昆仑XPU K200
 - x86_64 飞腾 aarch64 百度昆仑XPU R200
 - x86_64 飞腾 aarch64
 - 华为Atlas 300
 - 海光DCU: x86_64 PC
 - 寒武纪 MLU370 on x86_64
- 操作系统支持:Linux

根据开发者的选择,实际下载的版本可能是以下版本之一:

- EasyDL图像
 - x86 CPU 基础版
 - x86 CPU 加速版
 - Nvidia GPU 基础版
 - Nvidia GPU 加速版
 - x86 mlu270基础版
 - x86 SC5+基础版
 - Phytium MLU270基础版

- Phytium XPU基础版
- Phytium Atlas300I基础版
- Hygon DCU基础版

性能数据参考算法性能及适配硬件

*intel 官方合作,拥有更好的适配与性能表现。

Release Notes

时间	版本	说明
2022.1 2.29	1.7.2	模型性能优化;推理库性能优化
2022.1 0.27	1.7.1	新增语义分割模型http请求示例;升级海光DCU SDK,需配套rocm4.3版本使用;Linux GPU基础版下线适用于CUDA10.0及以 下版本的SDK;Linux GPU加速版升级推理引擎版本
2022.0 9.15	1.7.0	Linux GPU加速版升级预测引擎;Linux GPU加速版适用于CUDA9.0、CUDA10.0的SDK为deprecated,未来移除;新增实例分 割高性能模型离线部署;性能优化
2022.0 7.28	1.6.0	Linux CPU普通版、Linux GPU普通/加速版、Jetson新增目标追踪模型接入实时流的demo
2022.0 5.27	1.5.1	CPU、GPU普通版新增支持BML Cloud小目标检测模型
2022.0 5.18	1.5.0	GPU加速版max_batch_size参数含义变更;修复GPU加速版并发预测时部分图片结果预测错误及耗时增加问题;CPU普通版预 测引擎升级;新增版本号头文件;新增飞腾Atlas300I支持,并且在EasdDL新增多种加速版本;示例代码移除frame_buffer, 新增更安全高效的safe_queue; 新增Tensor In/Out接口和Demo
2022.0 4.25	1.4.1	EasyDL, BML升级支持paddle2模型
2022.0 3.25	1.4.0	新增支持海光服务器搭配海光DCU加速卡;
2021.1 2.22	1.3.5	GPU加速版支持自定义模型文件缓存路径;新增支持飞腾MLU270服务器、飞腾XPU服务器
2021.1 0.20	1.3.4	CPU加速版推理引擎优化升级,新增支持飞腾CPU、龙芯CPU服务器、比特大陆计算卡SC5+ BM1684、寒武纪MLU270;大幅 提升EasyDL GPU加速版有损压缩加速模型的推理速度
2021.0 8.19	1.3.2	CPU、GPU普通版及无损加速版新增支持EasyDL小目标检测,CPU普通版、GPU普通版支持检测模型的batch预测
2021.0 6.29	1.3.1	CPU普通版、GPU普通版支持分类模型的batch预测,CPU加速版支持分类、检测模型的batch预测;GPU加速版支持 CUDA11.1;视频流解析支持调整分辨率;预测引擎升级
2021.0 5.13	1.3.0	新增视频流接入支持;模型发布新增多种加速方案选择;目标追踪支持x86平台的CPU、GPU加速版;展示已发布模型性能评 估报告
2021.0 3.09	1.2.1	GPU新增目标追踪支持, http server服务支持图片通过base64格式调用,EasyDL高性能检测模型和均衡检测模型CPU加速版新 增量化压缩模型
2021.0 1.27	1.1.0	EasyDL经典版分类高性能模型升级;部分SDK不再需要单独安装OpenCV
2020.1 2.18	1.0.0	1.0版本发布!安全加固升级、性能优化、引擎升级、接口优化等多项更新
2020.1 1.26	0.5.8	EasyDL经典版分类模型CPU加速版里新增量化压缩模型
2020.1 0.29	0.5.7	新增CPU加速版支持:EasyDL经典版高精度、超高精度物体检测模型和EasyDL经典版图像分割模型
2020.0 9.17	0.5.6	性能优化,支持更多模型

Baidu 百度智能云文档

2020.0 8.11	0.5.5	提升预测速度;支持百度昆仑芯片
2020.0 5.15	0.5.3	优化性能,支持专业版更多模型
2020.0 4.16	0.5.2	支持CPU加速版;CPU基础版引擎升级;GPU加速版支持多卡多线程
2020.0 3.12	0.5.0	x86引擎升级;更新本地http服务接口;GPU加速版提速,支持批量图片推理
2020.0 1.16	0.4.7	ARM引擎升级;增加推荐阈值支持
2019.1 2.26	0.4.6	支持海思NNIE
2019.1 1.02	0.4.5	移除curl依赖;支持自动编译OpenCV;支持EasyDL 专业版 Yolov3; 支持EasyDL经典版高精度物体检测模型升级
2019.1 0.25	0.4.4	ARM引擎升级,性能提升30%; 支持EasyDL专业版模型
2019.0 9.23	0.4.3	增加海思NNIE加速芯片支持
2019.0 8.30	0.4.2	ARM引擎升级;支持分类高性能与高精度模型
2019.0 7.25	0.4.1	引擎升级,性能提升
2019.0 7.25	0.4.0	支持Xeye, 细节完善
2019.0 6.11	0.3.3	paddle引擎升级;性能提升
2019.0 5.16	0.3.2	新增NVIDIA GPU支持;新增armv7l支持
2019.0 4.25	0.3.1	优化硬件支持
2019.0 3.29	0.3.0	ARM64 支持;效果提升
2019.0 2.20	0.2.1	paddle引擎支持;效果提升
2018.1 1.30	0.1.0	第一版!

2022-5-18:【接口变更】 PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE含义变更。 变更前:预测输入图片数**不大于**该值均可。 变更后:预 测输入图片数需等于该值。SDK内部对该接口变更做了兼容处理,在输入图片数小于该值时依然可正常运行,但预测性能会和等于该值时一 致。推荐根据实际输入图片数量需求修改该值,尽可能保持最小。

2020-12-18: 【接口升级】 参数配置接口从1.0.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。 【关于SDK包与RES模型文件夹配套使用的说明】 我们强烈建议用户使用部署tar包中配套的SDK和RES。 更新模型时,如果SDK版本号有更新,请务必同时更新SDK,旧版本的SDK可能无法 正确适配新发布出来部署包中的RES模型。

快速开始

SDK在以下环境中测试通过

• x86_64, Ubuntu 16.04, gcc 5.4

- x86_64, Ubuntu 18.04, gcc 7.4
- Tesla P4, Ubuntu 16.04, cuda 9.0, cudnn 7.5
- x86_64, Ubuntu 16.04, gcc 5.4, XTCL r1.0
- aarch64, Kylin V10, gcc 7.3
- loongarch64, Kylin V10, gcc 8.3
- Bitmain SC5+ BM1684, Ubuntu 18.04, gcc 5.4
- x86_64 MLU270 , Ubuntu 18.04, gcc 7.5
- phytium MLU270 , Kylin V10 , gcc 7.3.0
- phytium XPU , Kylin V10 , gcc 7.3.0
- hygon DCU, CentOS 7.8 gcc 7.3.0
- XPU K200, x86_64, Ubuntu 18.04
- XPU K200 aarch64, Ubuntu 18.04
- XPU R200, x86_64, Ubuntu 18.04
- XPU R200 aarch64, Ubuntu 18.04
- MLU370, x86_64, Centos7.6.1810

依赖包括

- cmake 3+
- gcc 5.4 (需包含 GLIBCXX_3.4.22) , gcc / glibc版本请以实际SDK ldd的结果为准
- opencv3.4.11 (可选)
- cuda9.0_cudnn7 (使用NVIDIA-GPU时必须)
- XTCL 1.0.0.187 (使用昆仑服务器时必须)
- Rocm4.3, Miopen 2.14(使用海光DCU服务器时必须)

1. 安装依赖

以下步骤均可选,请开发者根据实际运行环境选择安装。

(可选) 安装cuda&cudnn

在NVIDIA GPU上运行必须(包括GPU基础版,GPU加速版)

对于GPU基础版,若开发者需求不同的依赖版本,请在PaddlePaddle官网下载对应版本的libpaddle_fluid.so或参考其文档进行编译,覆盖lib文件 夹下的相关库文件。

(可选) 安装TensorRT

在NVIDIA GPU上运行GPU加速版必须

下载包中提供了对应 cuda9.0、cuda10.0、cuda10.2、cuda11.0+四个版本的 SDK, cuda9.0 和 cuda10.0 的 SDK 默认依赖的 TensorRT 版本为 TensorRT7.0.0.11, cuda10.2 及以上的 SDK 默认依赖的 TensorRT 版本为 TensorRT8.4,请在这里下载对应 cuda 版本的 TensorRT,并把其中 的lib文件拷贝到系统lib目录,或其他目录并设置环境变量。

(可选) 安装XTCL 使用昆仑服务器及对应SDK时必须</mark>请安装与1.0.0.187版本兼容的XTCL。必要时,请将运行库路径添加到环境变量。

(可选) 安装Rocm、Miopen

使用海光DCU服务器对应SDK时必须

海光DCU SDK依赖Rocm 4.3和Miopen 2.14版本,推荐使用easyedge镜像 (registry.baidubce.com/easyedge/hygon_dcu_infer:1.0.2.rocm4.3), SDK镜像内运行,镜像拉取方式(wget https://aipe-easyedgepublic.bj.bcebos.com/dcu_docker_images/hygon_dcu_rocm4.3.tar.gz && docker load -i hygon_dcu_rocm4.3.tar.gz),关于海光DCU使用更多细节 可参考paddle文档

2. 使用序列号激活 在控制台获取的序列号请通过参数配置结构体EdgePredictorConfig的成员函数set_config(easyedge::params::PREDICTOR_KEY_SERIAL_NUM, "this-is-serial-num")设置。

具体请参考SDK自带的Demo.cpp文件的使用方法。

3. 测试Demo

模型资源文件默认已经打包在开发者下载的SDK包中。Demo工程直接编译即可运行。

请先将tar包整体拷贝到具体运行的设备中,再解压缩编译; 在Intel CPU上运行CPU加速版,如果thirdparty里包含openvino文件夹的,必须在 编译或运行demo程序前执行以下命令: source \${cpp_kit位置路径}/thirdparty/openvino/bin/setupvars.sh 或者执行 source \${cpp_kit位置路 径}/thirdparty/openvino/setupvars.sh(openvino-2022.1+)

请在官网获取序列号,填写在demo.cpp中



部分SDK中已经包含预先编译的二进制,如 bin/easyedge_demo, bin/easyedge_serving,配置LD_LIBRARY_PATH后,可直接运行: LD_LIBRARY_PATH=../lib ./bin/easyedge_serving

编译运行:

```
cd src
mkdir build && cd build
cmake .. && make
./easyedge_image_inference {模型RES文件夹} {测试图片路径}
**如果是NNIE引擎,使用sudo运行**
sudo ./easyedge_image_inference {模型RES文件夹} {测试图片路径}
```

如果希望SDK自动编译安装所需要的OpenCV库,修改cmake的optionEDGE_BUILD_OPENCV为ON即可。 SDK会自动从网络下载opencv源码,并 编译需要的module、链接。注意,此功能必须需联网。

cmake -DEDGE_BUILD_OPENCV=ON

若需自定义library search path或者gcc路径,修改CMakeList.txt即可。

demo运行效果:



图片加载失败

>./easyedge_image_inference ../../../RES 2.jpeg
2019-02-13 16:46:12,659 INFO [EasyEdge] [easyedge.cpp:34] 140606189016192 Baidu EasyEdge Linux Development Kit
0.2.1(20190213)
2019-02-13 16:46:14,083 INFO [EasyEdge] [paddlev2_edge_predictor.cpp:60] 140606189016192 Allocate graph success.
2019-02-13 16:46:14,326 DEBUG [EasyEdge] [paddlev2_edge_predictor.cpp:143] 140606189016192 Inference costs 168 ms
1, 1:txt_frame, p:0.994905 loc: 0.168161, 0.153654, 0.920856, 0.779621

对于支持批量预测的模型和SDK,可在使用前修改demo_image_inference或demo_batch_inference里的batch_size再编译、执行。

详情请参考下方<mark>使用说明-其他配置</mark>部分

4. 测试Demo HTTP 服务

编译demo完成之后,会同时生成一个http服务运行

./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401} ./easyedge_serving ../../../RES "1111-1111-1111" 0.0.0.0 24401

后,日志中会显示

HTTP is now serving at 0.0.0.0:24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。



对于目标追踪的模型,请选择一段视频,并耐心等待结果



图片加载失败

同时,可以调用HTTP接口来访问服务,具体参考下文接口说明。

使用说明

使用该方式,将运行库嵌入到开发者的程序当中。

使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

// step 1: 配置模型资源目录
EdgePredictorConfig config;

```
config.model_dir = {模型文件目录};
  // step 2: 创建并初始化Predictor;在这里选择合适的引擎
 auto predictor = global_controller()->CreateEdgePredictor(config);
 // step 3-1: 预测图像
 auto img = cv::imread({图片路径});
 std::vector<EdgeResultData> results;
 predictor->infer(img, results);
// step 3-2: 预测视频
std::vector<EdgeResultData> results;
FrameTensor frame_tensor;
 VideoConfig video_config;
 video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h
 video_config.source_value = video_src;
 /*
 ... more video_configs, 根据需要配置video_config的各选项
 */
 auto video_decoding = CreateVideoDecoding(video_config);
 while (video_decoding->next(frame_tensor) == EDGE_OK) {
    results.clear();
    if (frame_tensor.is_needed) {
      predictor->infer(frame_tensor.frame, results);
      render(frame_tensor.frame, results, predictor->model_info().kind);
    //video_decoding->display(frame_tensor); // 显示当前frame,需在video_config中开启配置
    //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置
```

输入图片不限制大小

SDK参数配置 SDK的参数通过EdgePredictorConfig::set_config和global_controller()->set_config配置。set_config的所有key 在easyedge_xxxx_config.h中。其中

- PREDICTOR前缀的key是不同模型相关的配置,通过EdgePredictorConfig::set_config设置
- CONTROLLER前缀的key是整个SDK的全局配置,通过global_controller()->set_config设置

以序列号为例,KEY的说明如下:

```
/**
* @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号
* 值类型:string
* 默认值:空
*/
static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";
```

```
使用方法如下:
```

```
EdgePredictorConfig config;
config.model_dir = ...;
config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, "1DB7-1111-1111-D27D");
```

具体支持的运行参数配置列表可以参考开发工具包中的头文件的详细说明。

相关配置均可以通过环境变量的方法来设置,对应的key名称加上前缀EDGE_即为环境变量的key。如序列号配置的环境变量key为EDGE_PREDICTOR_KEY_SERIAL_NUM,如指定CPU线程数的环境变量key为EDGE_PREDICTOR_KEY_CPU_THREADS_NUM。 注意:通过代 码设置的配置会覆盖通过环境变量设置的值。

Baidu 百度智能云文档

初始化

• 接口

auto predictor = global_controller()->CreateEdgePredictor(config); predictor->init();

若返回非0,请查看输出日志排查错误原因。

预测图像

● 接口

/** * @brief * 通用接口 * @param image: must be BGR , HWC format (opencv default) * @param result * @return */ virtual int infer(cv::Mat& image, std::vector<EdgeResultData>& result) = 0; /** * @brief * 批量图片推理接口 * @param image: must be BGR , HWC format (opencv default) * @param result * @return */ virtual int infer(std::vector<cv::Mat>& image, std::vector<std::vector<EdgeResultData>>& result

) = 0;

图片的格式务必为opencv默认的BGR, HWC格式。

• 返回格式

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index; // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测、图像分割时才有意义
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割的模型,该字段才有意义
    // 请注意: 图像分割时,以下两个字段会比较大,使用完成之后请及时释放EdgeResultData
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding,游程编码的mask
    // 目标追踪模型,该字段才有意义
    int trackid; // 轨迹id
    int frame; // 处于视频中的第几帧
    EdgeTrackStat track_stat; // 跟踪状态
};
```

关于矩形坐标

x1 * 图片宽度 = 检测框的左上角的横坐标

y1 * 图片高度 = 检测框的左上角的纵坐标

x2 * 图片宽度 = 检测框的右下角的横坐标

y2 * 图片高度 = 检测框的右下角的纵坐标

关于图像分割mask

```
cv::Mat mask为图像掩码的二维数组
{
{0,0,0,0,0,0,0,0,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,0,0,0,0,0,0,0},
}
其中1代表为目标区域,0代表非目标区域
```

关于图像分割mask_rle

该字段返回了mask的游程编码,解析方式可参考 http demo

以上字段可以参考demo文件中使用opencv绘制的逻辑进行解析

预测视频

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

● 接口

classVideoDecoding :

```
/**
* @brief 获取输入源的下一帧
* @param frame_tensor
* @return
*/
virtual int next(FrameTensor & frame_tensor) = 0;
/**
* @brief 显示当前frame_tensor中的视频帧
* @param frame_tensor
* @return
*/
virtual int display(const FrameTensor &frame_tensor) = 0;
/**
* @brief 将当前frame_tensor中的视频帧写为本地视频文件
* @param frame_tensor
* @return
*/
virtual int save(FrameTensor &frame_tensor) = 0;
/**
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
 * @brief 获取视频的width属性
 * @return
 */
virtual int get_width() = 0;
/**
* @brief 获取视频的height属性
* @return
*/
virtual int get_height() = 0;
```

struct VideoConfig

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type;
                          // 输入源类型
  std::string source_value;
                         // 输入源地址, 如视频文件路径、摄像头index、网络流地址
  int skip_frames{0};
                       // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
  int retrieve_all{false};
                     // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
  int input_fps{0};
                       // 在采取抽帧之前设置视频的fps
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
  bool enable display{false};
  std::string window_name{"EasyEdge"};
  bool display_all{false};
                        // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                        // 是否存储所有frame, 若为false, 仅存储根据skip_frames抽取的frame
  std::map::string, std::string> conf;
};
```

source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value:若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被 抽取返回,以作为显示或存储用。input_fps:用于抽帧前设置fps。resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。conf:高级选项。部分配置会通过该map来设置。

注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

设置序列号

请在网页控制台中申请序列号,并在init初始化前设置。 LinuxSDK 首次使用需联网授权。

EdgePredictorConfig config; config.set_config(easyedge::params::PREDICTOR_KEY_SERIAL_NUM, "this-is-serial-num");

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

EdgeLogConfig log_config; log_config.enable_debug = true; global_controller()->set_log_config(log_config);

http服务

1. 开启http服务 http服务的启动可以参考demo_serving.cpp文件。

/** * @brief 开启一个简单的demo http服务。 * 该方法会block直到收到sigint/sigterm。 * http服务里,图片的解码运行在cpu之上,可能会降低推理速度。 * @tparam ConfigT * @param config * @param host * @param port * @param service_id service_id user parameter, uri '/get/service_id' will respond this value with 'text/plain' * @param instance_num 实例数量,根据内存/显存/时延要求调整 * @return */ template<typename ConfigT> int start_http_server(const ConfigT &config, const std::string &host int port, const std::string &service_id, int instance_num = 1);

2. 请求http服务

开发者可以打开浏览器,http://{设备ip}:24401,选择图片或视频来进行测试。

http 请求方式一:无额外编码

• 图片测试:不使用图片base64格式

URL中的get参数:

参数	说明		默认值		
threshold	阈值过滤,	0~1	如不提供,则会使用模型的推荐阈值		

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

Python请求示例 (针对非语义分割模型)

import requests

```
with open('./1.jpg', 'rb') as f:
    img_data = f.read()
    result = requests.post(
        'http://127.0.0.1:24401/',
        params={'threshold': 0.1},
        data=img_data).json()
```

Python请求示例 (针对语义分割模型)

import requests

```
with open('./1.jpg', 'rb') as f:
img_data = f.read()
res = requests.post('http://127.0.0.1:24401/',
data=img_data)
with open("gray_result.png", "wb") as fb:
fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
```

Java请求示例

• 视频测试

Python请求示例 (注意:区别于图片预测,需指定Content-Type;否则会调用图片推理接口)

import requests

```
with open('./1.mp4', 'rb') as f:
video_data = f.read()
result = requests.post(
    'http://127.0.0.1:24401/',
    params={'threshold': 0.1},
    headers={'Content-Type': 'video'},
    data=video_data).json()
```

http 请求方法二:图片使用base64格式 HTTP方法:POST Header如下:

参数	值				
Content-Type	application/json				
Body请求填写:					

• 分类网络: body 中请求示例

```
{
"image": "<base64数据>"
"top_num": 5
}
```

body中参数详情

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
top_nu m	否	number	-	返回分类数量,不填该参数,则默认返回全部分类结果

• 检测和分割网络: Body请求示例:

```
{
"image": "<base64数据>"
}
```

body中参数详情:

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px, 支持jpg/png/bmp格式 注意去掉头部
thresho Id	否	number	-	默认为推荐阈值,也可自行根据需要进行设置

```
http 返回数据
```

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考预测图像·返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

请求示例 (针对非语义分割模型)

import base64
import requests

def main():
 with open("1.jpg 【图片路径】", 'rb') as f:
 result = requests.post("http://{服务ip地址}:24401/", json={
 "image": base64.b64encode(f.read()).decode("utf8")
 })

```
# print(result.request.body)
# print(result.request.headers)
print(result.content)
```

if __name__ == '__main__': main()

请求示例 (针对语义分割模型)

import base64 import requests

def main():

```
with open("1.jpg 【图片路径】", 'rb') as f:
res = requests.post("http://{服务ip地址}:24401/", json={"image": base64.b64encode(f.read()).decode("utf8")})
with open("gray_result.png", "wb") as fb:
fb.write(res.content) # 语义分割模型是像素点级别输出,可将api返回结果保存为灰度图,每个像素值代表该像素分类结果
```

if __name__ == '__main__':
 main()

返回示例

```
{
 "cost_ms": 52,
 "error_code": 0,
 "results": [
   {
      "confidence": 0.94482421875,
      "index": 1,
      "label": "IronMan",
      "x1": 0.059185408055782318,
      "x2": 0.18795496225357056,
      "y1": 0.14762254059314728,
      "y2": 0.52510076761245728
      "confidence": 0.94091796875,
      "index": 1,
      "label": "IronMan",
      "x1": 0.79151463508605957,
      "x2": 0.92310667037963867,
      "y1": 0.045728668570518494,
      "y2": 0.42920106649398804
```

其他配置

1. 日志名称、HTTP 网页标题设置

通过global_controller的set_config方法设置:

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_LOG_BRAND, "MY_BRAND");

效果如下:

图片加载失败

2. CPU线程数设置

CPU线程数可通过 EdgePredictorConfig::set_config 配置

```
EdgePredictorConfig config;
config.set_config(easyedge::params::PREDICTOR_KEY_CPU_THREADS_NUM, 4);
```

3. 批量预测设置

int batch_size = 2; // 使用前修改batch_size再编译、执行
while (get_next_batch(imgs, img_files, batch_size, start_index)) {
 ...
}

GPU 加速版 预测接口 GPU 加速版 SDK 除了支持上面介绍的通用接口外,还支持图片的批量预测,预测接口如下:

```
/**
 * @brief
 * GPU加速版批量图片推理接口
 * @param image: must be BGR , HWC format (opencv default)
 * @param result
 * @return
*/
virtual int infer(
     std::vector<cv::Mat>& image
     std::vector<std::vector<EdgeResultData>>& result
) = 0;
/**
 * @brief
 * GPU加速版批量图片推理接口,带阈值
 * @related infer(cv::Mat & image, EdgeColorFormat origin_color_format, std::vector<EdgeResultData> &result, float threshold)
*/
virtual int infer(
       std::vector<cv::Mat> &images,
       EdgeColorFormat origin_color_format,
       std::vector<std::vector<EdgeResultData>> &results
       float threshold
) = 0;
```

批量图片的预测接口的使用要求在调用 init 接口的时候设置一个有效的 PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE , 其含义见下方参数配置接口的介绍。

运行参数选项 在上面的内容中我们介绍了如何使用EdgePredictorConfig进行运行参数的配置。针对GPU加速版开发工具包,目前EdgePredictorConfig的运行参数所支持的Key包括如下项:

```
/**
* @brief 当有同类型的多个设备的时候,使用哪一个设备,如:
* GPU: 使用哪张GPU卡
* EdgeBoard(VMX), Movidius NCS : 使用哪一张加速卡
* 值类型 : int
* 默认值 : 0
*/
static constexpr auto PREDICTOR_KEY_DEVICE_ID = "PREDICTOR_KEY_DEVICE_ID";
/**
* @brief 生成最大 batch_size 为 max_batch_size 的优化模型,单次预测图片数量可以小于或等于此值(推荐等于此值,见release notes)
* 值类型: int
* 默认值 : 4
*/
static constexpr auto PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE = "PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE";
```

/** * @brief 设置device对应的GPU卡可以支持的最大并发量 * 实际预测的时候对应GPU卡的最大并发量不超过这里设置的范围,否则预测请求会排队等待预测执行 * 值类型: int * 默认值:1 */ static constexpr auto PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY = "PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY"; /** * @brief 是否开启fp16模式预测,开启后预测速度会更快,但精度会略有降低。并且需要硬件支持fp16 * 值类型: bool * 默认值: false */ static constexpr auto PREDICTOR_KEY_GTURBO_FP16 = "PREDICTOR_KEY_GTURBO_FP16"; /** * @brief 模型编译等级 *1:如果当前max_batch_size与历史编译产出的max_batch_size不相等时,则重新编译模型(推荐) *2:无论历史编译产出的max_batch_size为多少,均根据当前max_batch_size重新编译模型 * 值类型: int * 默认值:1 */ static constexpr auto PREDICTOR_KEY_GTURBO_COMPILE_LEVEL = "PREDICTOR_KEY_GTURBO_COMPILE_LEVEL"; /** * @brief GPU工作空间大小设置 * workspace_size = workspace_prefix * (1 << workspace_offset) * workspace_offset: 10 = KB, 20 = MB, 30 = GB * 值类型: int *默认值:WORKSPACE_PREFIX:100,WORKSPACE_OFFSET:20,即100MB */ static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX = "PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX"; static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET = "PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET"; /** * @brief 需要使用的dla core * 值类型: int * 默认值:-1(不使用) */ static constexpr auto PREDICTOR_KEY_GTURBO_DLA_CORE = "PREDICTOR_KEY_GTURBO_DLA_CORE"; /** * @brief 自定义缓存文件存储路径 * 值类型: string *默认值:~/.baidu/easyedge/mcache/{model_id * 1000000 + release_id} */ static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_DIR = "PREDICTOR_KEY_GTURBO_CACHE_DIR"; /** * @brief 自定义缓存文件命名,默认即可 * 值类型: string * 默认值: 根据配置自动生成 */ static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_NAME = "PREDICTOR_KEY_GTURBO_CACHE_NAME"; /** * @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号 * 值类型: string *默认值:空 */ static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";

PREDICTOR_KEY_GTURBO_CACHE_NAME:首次加载模型会先对模型进行编译优化,通过此值可以设置优化后的产出文件名,这在多进程加载同一个模型的时候是有用的。

PREDICTOR_KEY_GTURBO_CACHE_DIR:首次加载模型经过编译优化后,产出的优化文件会存储在这个位置,可以按需修改。

PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX、PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET:设置运行时可以被用来使用的最大临时显

存。

PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE:此值用来控制批量图片预测可以支持的最大图片数,实际预测的时候单次预测图片数需等于此值。

PREDICTOR_KEY_DEVICE_ID:设置需要使用的 GPU 卡号。

PREDICTOR_KEY_GTURBO_COMPILE_LEVEL: 模型编译等级。通常模型的编译会比较慢,但编译产出是可以复用的。可以在第一次加载模型的时候设置合理的 max_batch_size 并在之后加载模型的时候直接使用历史编译产出。是否使用历史编译产出可以通过此值 compile_level 来控制。当此值为 0 时,表示忽略当前设置的 max_batch_size 而仅使用历史产出(无历史产出时则编译模型);当此值为 1 时,会比较历史产出和当前设置的 max_batch_size 是否相等,如不等,则重新编译;当此值为 2 时,无论如何都会重新编译模型。

PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY:通过此值设置单张 GPU 卡上可以支持的最大 infer 并发量,其上限取决于硬件限制。init 接口 会根据此值预分配 GPU 资源,建议结合实际使用控制此值,使用多少则设置多少。注意:此值的增加会降低单次 infer 的速度,建议优先考虑 batch inference 和 multi predictor。

PREDICTOR_KEY_GTURBO_FP16:默认是 fp32 模式,置 true 可以开启 fp16 模式预测,预测速度会有所提升,但精度也会略微下降,权衡使用。注意:不是所有模型都支持 fp16 模式。目前已知不支持fp16的模型包括:图像分类高精度模型。

多线程预测 GPU 加速版 SDK 的多线程分为单卡多线程和多卡多线程两种。 单卡多线程:创建一个 predictor,并通过 PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY 控制单卡所支持的最大并发量,只需要 init 一次,多线程调用 infer 接口。 多卡多线程:多卡的 支持是通过创建多个 predictor,每个 predictor 对应一张 GPU 卡,predictor 的创建和 init 的调用放在主线程,通过多线程的方式调用 infer 接口。

已知问题 1. 多线程时图片按线程分配不均 或 不同batch size的图片交叉调用infer接口时,部分结果错误 A: EasyDL图像分类高精度模型在有些显卡上可能存在此问题,可以考虑填充假图片数据到图片比较少的线程或batch以使得infer间的图片绝对平均。

2. 显存持续增长或遇到 terminate called after throwing an instance of 'std::runtime_error' what(): Failed to create object A: 部分显卡存在此问题, 如果遇到此问题, 请确认没有频繁调用 init 接口, 通常调用 infer 接口即可满足需求。

3. 开启 fp16 后,预测结果错误 A:不是所有模型都支持 fp16 模式。目前已知的不支持fp16的模型包括:图像分类高精度模型。目前不支持的 将会在后面的版本陆续支持。

昆仑服务器 昆仑服务器SDK支持将EasyDL的模型部署到昆仑服务器上。SDK提供的接口风格一致,简单易用,轻松实现快速部署。Demo的测试 可参考上文中的测试Demo部分。

参数配置接口 在上面的内容中我们介绍了如何使用EdgePredictorConfig进行运行参数的配置。针对昆仑服务器开发工具包,目前EdgePredictorConfig的运行参数所支持的Key包括如下项:

```
/**
* @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号
* 值类型: string
*默认值:空
static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";
/**
* @brief 当有同类型的多个设备的时候,使用哪一个设备,如:
* 使用哪张加速卡
* 值类型: int
* 默认值:0
static constexpr auto PREDICTOR_KEY_DEVICE_ID = "PREDICTOR_KEY_DEVICE_ID";
/**
* @brief 设置需要同时预测的图片数量
* 值类型: int
* 默认值:1
static constexpr auto PREDICTOR_KEY_KUNLUN_BATCH_SIZE = "PREDICTOR_KEY_KUNLUN_BATCH_SIZE";
```

PREDICTOR_KEY_DEVICE_ID:设置需要使用的加速卡的卡号。

PREDICTOR_KEY_KUNLUN_BATCH_SIZE:设置单次预测可以支持的图片数量。

使用方法:

int batch_size = 1;

config.set_config(easyedge::params::PREDICTOR_KEY_KUNLUN_BATCH_SIZE, batch_size);

模型调优 通过设置如下环境变量,可以在初始化阶段对模型调优,从而让预测的速度更快。

export XPU_CONV_AUTOTUNE=5

FAQ

1. 如何处理一些 undefined reference?

如: undefined reference to `curl_easy_setopt@CURL_OPENSSL_3'

方案1:通过安装libcurl3 libcurl-openssl1.0-dev来解决。 方案2:如果开发者想不想使用低版本的openssl (如Ubuntu 18.04),可以link静态 库easyedge_static.a,自己指定需要的Library的版本:

示例:修改CMakeList.txt

find_package(CURL REQUIRED) target_link_libraries(easyedge_demo \${OpenCV_LIBS} easyedge_static pthread \${CURL_LIBRARIES} verify_static \${其他需要的库})

其中,其他需要的库视具体sdk中包含的库而定。

2. EasyDL SDK与云服务效果不一致,如何处理?

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

3. NVIDIA GPU预测时,报错显存不足如以下错误字样:

paddle.fluid.core.EnforceNotMet: Enforce failed. Expected allocating <= available, but received allocating:20998686233 > available:19587333888. Insufficient GPU memory to allocation. at [/paddle/paddle/fluid/platform/gpu_info.cc:170]

请根据显存大小和模型配置。调整合适的初始 fraction_of_gpu_memory。参数的含义参考这里。

4. 如何将我的模型运行为一个http服务?目前cpp sdk暂未集成http运行方式; 0.4.7版本之后,可以通过start_http_server方法开启http服务。

5. 运行NNIE引擎报permission denied 日志显示:

open sys: Permission denied open err : Permission denied open err : Permission denied

请使用sudo在root下运行。

6. 运行SDK报错 Authorization failed

情况一:日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888" ./easyedge_demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更

● bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 ~/.baidu/easyedge 目录,再重新激活。

7. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

8. 运行二进制时,提示 libverify.so cannot open shared object file

可能cmake没有正确设置rpath,可以设置LD_LIBRARY_PATH为sdk的lib文件夹后,再运行:

LD_LIBRARY_PATH=\$LD_LIBRARY_PATH:../lib ./easyedge_demo

9. 运行二进制时提示 libopencv_videoio.so.4.5: cannot open shared object file: No such file or directory 同上面8的问题类似,没有正确设置动态 库的查找路径,可通过设置LD_LIBRARY_PATH为sdk的thirdparty/opencv/lib文件夹解决

export LD_LIBRARY_PATH=\${LD_LIBRARY_PATH}:../thirdparty/opencv/lib (tips: 上面冒号后面接的thirdparty/opencv/lib路径以实际项目中路径为准,比如也可能是../../thirdparty/opencv/lib)

10. 编译时报错:file format not recognized 可能是因为在复制SDK时文件信息丢失。请将整个压缩包复制到目标设备中,再解压缩、编译

11. 进行视频解码时,报错符号未找到、格式不支持、解析出的图片为空、无法设置抽帧 请确保安装OpenCV时,添加了-DWITH_FFMPEG=ON选项(或者GStream选项),并且检查OpenCV的安装日志中,关于Video I/O段落的说明是否为YES。

- -- Video I/O:
- -- DC1394: YES (ver 2.2.4)
- -- FFMPEG: YES
- -- avcodec: YES (ver 56.60.100)
- -- avformat: YES (ver 56.40.101)
- -- avutil: YES (ver 54.31.100)
- -- swscale: YES (ver 3.1.101)
- -- avresample: NO -- libv4l/libv4l2: NO
- -- libv4l/libv4l2: NO-- v4l/v4l2: linux/videodev2.h

如果为NO,请搜索相关解决方案,一般为依赖没有安装,以apt为例:

apt-get install yasm libjpeg-dev libjasper-dev libavcodec-dev libavformat-dev libswscale-dev libdc1394-22-dev libgstreamer0.10-dev libgstreamer0.10-dev libgstreamer-plugins-base0.10-dev libv4l-dev python-dev python-numpy libtbb-dev libgt4-dev libgtk2.0-dev libfaac-dev libmp3lame-dev libopencore-amrnb-dev libopencore-amrnb-dev libtheora-dev libtvorbis-dev libtvordcore-dev x264 v4l-utils ffmpeg

12. GPU加速版运行有损压缩加速的模型,运算精度较标准模型偏低首先请保证数据集不会太小。其次可以通过将模型目录RES中的calibrationtable移除,并通过将PREDICTOR_KEY_GTURBO_FP16设置为true,使用FP16的运算精度重新评估模型效果。若依然不理想,可将calibrationtable移除并将PREDICTOR_KEY_GTURBO_FP16设置为false,从而使用更高精度的FP32的运算精度。

心 通用小型设备

心目标跟踪WindowsSDK集成文档

简介

本文档介绍目标跟踪通用小型设备Windows SDK的使用方法。

- 硬件支持:
 - Intel CPU 普通版 * x86_64
 - CPU 加速版 Intel Xeon with AVX2 and AVX512 Intel Core Processors with AVX2 Intel Atom Processors with SSE
 - Intel Movidius MyRIAD2/MyRIAD X (仅支持Win10)

- 操作系统支持
 - 普通版: 64位 Windows 7 及以上
 - 加速版:64位 Windows 10
- 环境依赖 (必须安装以下版本)
 - .NET Framework 4.5
 - Visual C++ Redistributable Packages for Visual Studio 2013
 - Visual C++ Redistributable Packages for Visual Studio 2015
- 协议
 - HTTP

Release Notes | 时间 | 版本 | 说明 | | -------- | ------ | -2023-08-30 | 1.8.3 | 新增支持按实例数鉴权 | | 2023-06-29 | 1.8.2 | 优化模型 算法 || 2023-05-17 | 1.8.1 | 预测引擎升级,修复部分网络内存泄露问题 || 2023-03-16 | 1.8.0 | 预测引擎升级 || 2022-12-29 | 1.7.2 | 预测引 擎升级 | | 2022-10-27 | 1.7.1 | GPU底层引擎升级,下线基础版CUDA10.0及以下版本支持 | | 2022-09-15 | 1.7.0 | 优化模型算法;GPU CUDA9.0 CUDA10.0 标记为待废弃状态 || 2022-07-28 | 1.6.0 | 优化模型算法 || 2022-05-27 | 1.5.1 | 新增支持BML Cloud小目标检测模型 || 2022-05-18 | 1.5.0 | 修复各别机器下程序崩溃的问题 | | 2022-04-25 | 1.4.1 | EasyDL, BML升级支持paddle2模型 | | 2022-03-25 | 1.4.0 | 优化 模型算法 || 2021-12-22 | 1.3.5 | GPU基础版推理引擎优化升级;GPU加速版支持自定义模型文件缓存路径;demo程序优化环境依赖检测 || 2021-10-20 | 1.3.4 | 修复已知问题| | 2021-08-19 | 1.3.2 | 新增支持EasyDL小目标检测,新增DEMO二进制文件 | | 2021-06-29 | 1.3.1 | 预测引 擎升级 || 2021-05-13 | 1.3.0 | 模型发布新增多种加速方案选择;目标追踪支持x86平台的GPU及加速版;展示已发布模型性能评估报告 || 2021-04-08 | 1.2.3 | 支持BML平台模型仓库本地上传模型 | | 2021-03-09 | 1.2.2 | CPU加速版支持int8量化模型 | | 2021-01-27 | 1.2.1 | 新增模型 支持;性能优化;问题修复||2020.12.18|1.2.0|推理引擎升级||2020-11-26|1.1.20|新增一些模型的加速版支持||2020.10.29|1.1.20 |修复已知问题||2020.05.15|1.1.16|优化性能,修复已知问题||2020-09-17|1.1.19|支持更多模型||2020.08.11|1.1.18|支持专业版 更多模型 || 2020.06.23 | 1.1.17 | 支持专业版更多模型 || 2020.05.15 | 1.1.16 | 优化性能,修复已知问题 || 2020.04.16 | 1.1.15 | 升级引擎 版本 || 2020.03.13 | 1.1.14| 支持EdgeBoardVMX || 2020.02.23 | 1.1.13| 支持多阶段模型 || 2020.01.16 | 1.1.12 | 预测默认使用推荐阈值|| 2019.12.26 | 1.1.11 | CPU加速版支持物体检测高精度 | | 2019.12.04 | 1.1.10 | 支持图像分割 | | 2019.10.21 | 1.1.9 | 支持 EasyDL 专业版 | | 2019.08.29 | 1.1.8 | CPU 加速版支持 | | 2019.07.19 | 1.1.7 | 提供模型更新工具 | | 2019.05.16 | 1.1.3 | NVIDIA GPU 支持 | | 2019.03.15 | 1.1.0 | 架构与功能完善 | | 2019.02.28 | 1.0.6 | 引擎功能完善 | | 2019.02.13 | 1.0.5 | paddlepaddle 支持 | | 2018.11.30 | 1.0.0 | 第一版 ! |

快速开始

1. 安装依赖

必须安装:

安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642

Visual C++ Redistributable Packages for Visual Studio 2013

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

Visual C++ Redistributable Packages for Visual Studio 2015

https://www.microsoft.com/zh-cn/download/details.aspx?id=48145

可选安装:

Openvino (仅使用Python Intel Movidius必须)

- 使用 OpenVINO[™] toolkit 安装,请参考 OpenVINO toolkit 文档安装 2020.3.1LTS(必须)版本,安装时可忽略Configure the Model Optimizer及 后续部分。
- 使用源码编译安装,请参考 Openvino Inference Engine文档编译安装 2020.3.1LTS (必须)版本。

安装完成后,请设置环境变量OPENVINO_HOME为您设置的安装地址,默认是C:\Program Files (x86)\IntelSWTools,并确保文件夹下的openvino的快捷方式指到了2020.3.1LTS版本。

Baidu 百度智能云文档

充属性		
算机名 硬件 高级 系统保护 远程	xiaoyuchen 的用户受重(U)	
要进行大多数重改,你必须作为管理员登录。	变量	值 Chilsen LongDrive
WEINARDER, WEINFICHERER.	OPENVINO_HOME	C:\Program Files (x86)\IntelSWTools
如常效果 处理器计划 内方使田 以及虚拟内方	Path	C:\UsersAppData\Local\Microsoft\WindowsApp
NUSCHOR, XLEERING, PHTCH, KAXDENPHT	TEMP	C:\Users
设置(5)	TMP	C:\Users\www.me.eppUata\Loca\\iemp
用户配置文件		
与登录帐户相关的桌面设置		新建(N) 编辑(E) 删除(D)
设置(E)	系统变量(S)	
	变量	值 ^
启动和故障恢复	ComSpec	C:\Windows\system32\cmd.exe
系统启动、系统故障和调试信息	INTEL DEV REDIST	C:\Windows\System32\Drivers\DriverData C:\Program Files (x86)\Common Files\Intel\Shared Libraries\
20年の1	MIC_LD_LIBRARY_PATH	%INTEL_DEV_REDIST%compiler\lib\intel64_win_mic
	NUMBER_OF_PROCESSORS	5 8
	OS	Windows_NT
环境变量(<u>N</u>)	Patri	C:(Program riles (xoo)/Parallels/Parallels Tools/Applications;c
		新建(W) 编辑(I)
确定 取消 应用(A)		禍定 取消
IntelSWTools		
共享 查看		
Program Files (x86) > IntelSV	NIOOIS >	▼ O 没 搜索
□ 名称	^	修改日期
	00	2020/4/15 23:41
*		2020/ 1/ 10 20111
openvir	no_2019.2.275	2020/3/10 19:34
openvi	no 2020.2.117	2020/4/15 23:42
> *		2020/4/10 20112
注意事项		
1. 安装目录不能包含中文		
2. Windows Server 请自行开启,选择"我的电脑" 可。	——"属性"——"管理"-	—— "添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启即
2. 运行离线SDK		
解压下载好的SDK,打开EasyEdge.exe,输入Seri	al Num,选择鉴权模封	式,点击"启动服务",等待数秒即可启动成功,本地服务默认运行在
http://127.0.0.1:24401/		
其他任何语言只需通过HTTP调用即可。		

2.1 离线鉴权(默认鉴权模式) 首次联网激活,后续离线使用

—		
Al ver. 1.2.0.0 core. 1	1.8.1 sec. 1.5.1 –	×
Model	March 2012 and a start of a start in	
Serial Num	The second se	
Host	127.0.0.1	
Port	24401	
鉴 权模式	● 离线鉴权 ○ 按实例数鉴权	
服务状态	服务未启动	
	启动服务	
	是否开机自己	动启动 🗌

2.2 按实例数鉴权 周期性联网激活,离线后会释放所占用鉴权,启动时请确保心跳间隔小于等于生成序列号时填写的定期确认时间

AI ver. 1.2.0.0 core. 1	1.8.1 sec. 1.5.1 —		×
Model	[4] A.W. W. W. A. W.		
Serial Num	(
Host	127.0.0.1		
Port	24401		
鉴权模式	○ 离线鉴权 ● 按实例数鉴权 心跳间隔 20 秒 ~		
服务状态	服务未启动		
	启动服务		
	是否开机自治	动启动	

基于源码集成时,若需要按实例数鉴权,需要通过代码指定使用按实例数鉴权

global_controller()->set_config(easyedge::params::CONTROLLER_KEY_AUTH_MODE, 2); global_controller()->set_config(easyedge::params::CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL, 20);

或通过环境变量指定

set EDGE_CONTROLLER_KEY_AUTH_MODE=2 set EDGE_CONTROLLER_KEY_INSTANCE_UPDATE_INTERVAL=20

3. Demo示例(以图像服务为例)

服务运行成功,此时可直接在浏览器中输入http://127.0.0.1:24401,在h5中测试模型效果。



使用说明

图像服务调用说明

Python 使用示例代码如下

import requests

```
with open('./1.mp4', 'rb') as f:
img = f.read()
```

params 为GET参数 data 为POST Body result = requests.post('http://127.0.0.1:24401/', params={'threshold': 0.1}, data=img).json() FileStream fs = new FileStream("./1.mp4", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close(); WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd());

C++ 使用示例代码如下,需要安装curl

sr.Close();
response.Close();

```
**include <sys/stat.h>**
**include <curl/curl.h>**
**include <iostream>**
**include <string>**
**define S_ISREG(m) (((m) & 0170000) == (0100000))**
**define S_ISDIR(m) (((m) & 0170000) == (0040000))**
size_t write_callback(void *ptr, size_t size, size_t num, void *data) {
  std::string *str = dynamic_cast<std::string *>((std::string *)data);
  str->append((char *)ptr, size*num);
  return size*num;
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./1.mp4";
  FILE *fp = NULL;
  std::string response;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\" \n", post_data_filename);
     return -1;
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
    curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
    curl_easy_setopt(curl, CURLOPT_POST, 1L);
    curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
    curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
 curl_easy_setopt(curl, CURLOPT_WRITEFUNCTION, write_callback);
    curl_easy_setopt(curl, CURLOPT_WRITEDATA, &response);
     res = curl_easy_perform(curl);
     if (res != CURLE OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     std::cout << response << std::endl; // response即为返回的json数据
     curl_easy_cleanup(curl);
  curl_global_cleanup();
  fclose(fp);
  return 0;
```

结果 获取的结果存储在response字符串中。请求参数

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送图片二进制。

返回参数|字段|类型|取值|说明||--------|-----|-----|-----|| confidence|float|0~1|追踪的置信度||label|string||追踪的类 别||index|number||追踪的类别||x1, y1|float|0~1|矩形的左上角坐标(相对长宽的比例值)||x2, y2|float|0~1|矩形的右下角坐标(相 对长宽的比例值)||trackId|int||轨迹id||frame|int||帧号|

关于矩形坐标

- x1 * 图片宽度 = 检测框的左上角的横坐标
- y1 * 图片高度 = 检测框的左上角的纵坐标

x2 * 图片宽度 = 检测框的右下角的横坐标

y2 * 图片高度 = 检测框的右下角的纵坐标

FAQ

1. 服务启动失败,怎么处理?

根据SDK内的README.md检查依赖是否都已正确安装

请确保相关依赖都安装正确,版本必须如下: .NET Framework 4.5 Visual C++ Redistributable Packages for Visual Studio 2013 * Visual C++ Redistributable Packages for Visual Studio 2015

如使用的是Python Intel Movidius版,需额外确保Openvino安装正确,版本为2020.3.1LTS版 如使用Windows Server,需确保开启桌面体验

2. 服务调用时返回为空,怎么处理?调用输入的图片必须是RGB格式,请确认是否有alpha通道。

3. 多个模型怎么同时使用? SDK设置运行不同的端口,点击运行即可。

4. JAVA、C#等其他语言怎么调用SDK?

参考 https://ai.baidu.com/forum/topic/show/943765

5. 启动失败,缺失DLL? 打开EasyEdge.log,查看日志错误,根据提示处理 缺失DLL,请使用 https://www.dependencywalker.com/ 查看相应模 块依赖DLL缺失哪些,请自行下载安装

6. 启动失败,报错NotDecrypted? Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

7. 启动失败,报错 SerialNum无效

日志显示failed to get/check device id(xxx)或者 Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 C: \Users \\${用户名} \.baidu \easyedge 目录,再重新激活。

其他问题 如果无法解决,可到论坛发帖: https://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

心 软硬一体方案

心 目标跟踪Jetson专用SDK集成文档

简介

本文档介绍EasyEdge/EasyDL的Jetson SDK的使用方法。Jetson SDK支持的硬件包括Jetson nano, Jetson TX2, Jetson AGX Xavier和Jetson Xavier NX。您可在AI市场了解Jetson相关系列产品,同时可以在软硬一体方案了解部署方案。

模型支持:

• EasyDL图像:图像分类高精度,图像分类高性能,物体检测高精度,物体检测均衡,物体检测高性能,目标跟踪单标签模型。

• BML :

- 公开数据集预训练模型:SSD-MobileNetV1,YOLOv3-DarkNet,YOLOv3-MobileNetV1,ResNet50,ResNet101,SE-ResNeXt50,SE-ResNeXt101,MobileNetV2,EfficientNetB0_small,EfficientNetB4,MobileNetV3_large_x1_0,ResNet18_vd,SE_ResNet18_vd,Xception71。
- 百度超大规模数据集预训练模型:YOLOv3-DarkNet, MobileNetV3_large_x1_0, ResNet50_vd, ResNet101_vd。
- EasyEdge: EasyEdge支持的模型较多,详见查看模型网络适配硬件。若模型不在此列表,可以尝试使用自定义网络生成端计算组件。

软件版本支持使用EasyDL的Jetson系列SDK需要安装指定版本的JetPack和相关组件。所支持的JetPack版本会随着SDK版本的升级和新版本 JetPack的推出而不断的更新。在使用SDK前请务必保证软件版本满足此处声明版本。目前所支持的JetPack版本包括:

JetPack5.0.2

Baidu 百度智能云文档

- JetPack5.0.1
- JetPack4.6
- JetPack4.5
- JetPack4.4 (deprecated,该版本SDK会在未来某个版本移除,请切换至新版本JetPack)
- JetPack4.2.2 (已移除,请切换至新版本JetPack)

安装JetPack时请务必安装对应的组件:

- 使用SDK Manager安装JetPack需要勾选TensorRT、OpenCV、CUDA、cuDNN等选项。
- 使用SD Card Image方式(仅对Jetson Nano和Jetson Xavier NX有效)则无需关心组件问题,默认会全部安装。

Release Notes | 时间 | 版本 | 说明 | | --- | --- | | 2022.12.29 | 1.7.2 | 新增支持JetPack5.0.2 ; 缓存机制优化;模型性能优化 || 2022.07.28 | 1.6.0 | 新增支持JetPack5.0.1 , 新增目标追踪接入实时流的demo || 2022.05.18 | 1.5.0 | 部分模型切换格式,max_batch_size含义变更,由输入 图片数不大于该值变更为等于该值;移除适用于JetPack4.2.2的SDK;示例代码demo_stream_inference重构;示例代码移除frame_buffer,新增 更安全高效的safe_queue || 2021.12.22 | 1.3.5 | 新增支持JetPack4.6 ; 支持在EasyEdge平台语义分割模型生成开发套件;修复缓存问题;支持 自定义缓存路径 || 2021.10.20 | 1.3.4 | 新增支持JetPack4.5 ; 大幅提升EasyDL有损压缩加速模型的推理速度 || 2021.06.29 | 1.3.1 | 视频流支 持分辨率调整;支持将预测后的视频推流,新增推流demo || 2021.05.13 | 1.3.0 | 新增视频流接入支持;EasyDL模型发布新增多种加速方案选 择;目标追踪支持x86平台的CPU、GPU加速版;展示已发布模型性能评估报告 || 2021.03.09 | 1.2.1 | EasyEdge新增一系列模型的支持;性能 优化 || 2021.01.27 | 1.1.0 | EasyDL经典版高性能分类模型升级;

EasyDL经典版检测模型新增均衡选项;

EasyEdge平台新增Jetson系列端计算组件的生成;

问题修复||2020.12.18|1.0.0|接口升级和一些性能优化||2020.08.11|0.5.5|部分模型预测速度提升||2020.06.23|0.5.4|支持 JetPack4.4DP,支持EasyDL专业版更多模型||2020.05.15|0.5.3|专项硬件适配SDK支持Jetson系列|

2022-5-18:【接口变更】 PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE含义变更。 变更前:预测输入图片数不大于该值均可。 变更后:预测输入图片数需等于该值。SDK内部对该接口变更做了兼容处理,在输入图片数小于该值时依然可正常运行,但预测性能会和等于该值时一致。推荐根据实际输入图片数量需求修改该值,尽可能保持最小。 【版本移除】适用于JetPack4.4版本的SDK被标记为deprecated,SDK会在未来某个版本移除,建议切换至最新版本JetPack。 适用于JetPack4.2.2版本的SDK被移除。

2020-12-18: 【接口升级】 参数配置接口从1.0.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

2021-10-20: 【版本移除】 适用于JetPack4.2.2版本的SDK被标记为deprecated,该版本代码已停止更新,SDK会在未来某个版本移除,请切 换至新版本JetPack

快速开始 安装依赖 本SDK适用于JetPack4.5、JetPack4.6、JetPack5.0系列版本,请务必安装其中之一版本,并使用对应版本的SDK。 注意在安装JetPack时,需同时安装CUDA、cuDNN、OpenCV、TensorRT等组件。

如已安装JetPack需要查询相关版本信息,请参考下文中的开发板信息查询与设置。

使用序列号激活

首先请在官网获取序列号。



将获取到的序列号填写到demo文件中或以参数形式传入。



图片加载失败

编译并运行Demo 模型资源文件默认已经打包在开发者下载的SDK包中。Demo工程直接编译即可运行。

Baidu 百度智能云文档

编译运行:

demo运行示例:

baidu@nano:~/ljay/easydl/sdk/demo/build\$./demo_batch_inference/easyedge_batch_inference ../../../RES/ /ljay/images/mix008.jpeg 2020-08-06 20:56:30,665 INFO [EasyEdge] 548125646864 Compiling model for fast inference, this may take a while (Acceleration) 2020-08-06 20:57:58,427 INFO [EasyEdge] 548125646864 Optimized model saved to: /home/baidu/.baidu/easyedge/jetson/mcache/24110044320/m_cache, Don't remove it Results of image /ljay/images/mix008.jpeg: 2, kiwi, p:0.997594 loc: 0.352087, 0.56119, 0.625748, 0.868399 2, kiwi, p:0.993221 loc: 0.45789, 0.0730294, 0.73641, 0.399429 2, kiwi, p:0.992884 loc: 0.156876, 0.0598725, 0.3802, 0.394706 1, tomato, p:0.992125 loc: 0.523592, 0.389156, 0.657738, 0.548069 1, tomato, p:0.991821 loc: 0.665461, 0.419503, 0.805282, 0.573558 1, tomato, p:0.989883 loc: 0.297427, 0.439999, 0.432197, 0.59325 1, tomato, p:0.981654 loc: 0.383444, 0.248203, 0.506606, 0.400926 1, tomato, p:0.971682 loc: 0.183775, 0.556587, 0.286996, 0.711361 1, tomato, p:0.968722 loc: 0.379391, 0.0386965, 0.51672, 0.209681 Done

检测结果展示:



图片加载失败

测试Demo HTTP 服务

编译demo完成之后,会同时生成一个http服务,运行

./easyedge_serving {res_dir} {serial_key} {host, default 0.0.0.0} {port, default 24401} ./easyedge_serving ../../../RES "1111-1111-1111" 0.0.0.0 24401

日志中会显示

HTTP is now serving at 0.0.0.0:24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。



同时,可以调用HTTP接口来访问服务,具体参考下文接口说明。

使用说明

使用该方式,将运行库嵌入到开发者的程序当中。

使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

```
// step 1: 配置模型运行参数
EdgePredictorConfig config;
config.model_dir = model_dir;
config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, serial_num);
config.set_config(params::PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE, 1); // 优化的模型可以支持的batch_size
config.set_config(params::PREDICTOR_KEY_GTURBO_FP16, false); // 置true开启fp16模式推理会更快,精度会略微降低,但取决于硬
件是否支持fp16,不是所有模型都支持fp16,参阅文档
config.set_config(params::PREDICTOR_KEY_GTURBO_COMPILE_LEVEL, 1); // 编译模型的策略,如果当前设置的max_batch_size与历史编
译存储的不同,则重新编译模型
// step 2: 创建并初始化Predictor
auto predictor = global_controller()->CreateEdgePredictor<EdgePredictorConfig>(config);
if (predictor->init() != EDGE_OK) {
  exit(-1);
// step 3-1: 预测图像
auto img = cv::imread({图片路径});
std::vector<EdgeResultData> results;
predictor->infer(img, results);
// step 3-2: 预测视频
std::vector<EdgeResultData> results
FrameTensor frame_tensor;
VideoConfig video_config;
video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h
video_config.source_value = video_src;
/*
... more video_configs, 根据需要配置video_config的各选项
*/
auto video_decoding = CreateVideoDecoding(video_config);
while (video_decoding->next(frame_tensor) == EDGE_OK) {
 results.clear();
 if (frame_tensor.is_needed) {
    predictor->infer(frame_tensor.frame, results);
    render(frame_tensor.frame, results, predictor->model_info().kind);
  //video_decoding->display(frame_tensor); // 显示当前frame,需在video_config中开启配置
  //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置
```

初始化接口

```
auto predictor = global_controller()->CreateEdgePredictor<EdgePredictorConfig>(config);
if (predictor->init() != EDGE_OK) {
    exit(-1);
}
```

若返回非0,请查看输出日志排查错误原因。

预测接口

```
/**
* @brief
* 单图预测接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
     cv::Mat& image, std::vector<EdgeResultData>& result
) = 0;
/**
* @brief
* 批量图片预测接口
* @param image: must be BGR , HWC format (opencv default)
* @param result
* @return
*/
virtual int infer(
    std::vector<cv::Mat>& image,
     std::vector<std::vector<EdgeResultData>>& results
) = 0;
/**
 * @brief
* 批量图片预测接口,带阈值
 * @related infer(cv::Mat & image, EdgeColorFormat origin_color_format, std::vector<EdgeResultData> &result, float threshold)
*/
virtual int infer(
       std::vector<cv::Mat> &images,
       EdgeColorFormat origin_color_format,
       std::vector<std::vector<EdgeResultData>> &results
       float threshold
) = 0;
```

图片的格式务必为opencv默认的BGR, HWC格式。

批量图片的预测接口的使用要求在调用 init 接口的时候设置一个有效的 max_batch_size,其含义见下方参数配置接口的介绍。

参数配置接口 参数配置通过结构体 EdgePredictorConfig完成。

```
struct EdgePredictorConfig {
    /**
    * @brief 模型资源文件夹路径
    */
    std::string model_dir;
    std::map<std::string, std::string> conf;
    EdgePredictorConfig();
    template<typename T>
    T get_config(const std::string &key, const T &default_value);
    template<typename T = std::string>
    T get_config(const std::string &key);
    template<typename T>
    const T *get_config(const std::string &key, const T *default_value);
    template<typename T>
   void set_config(const std::string &key, const T &value);
   template<typename T>
   void set_config(const std::string &key, const T *value);
    static EdgePredictorConfig default_config();
 };
运行参数选项的配置以key、value的方式存储在类型为std::map的conf中,并且键值对的设置和获取可以通
过EdgePredictorConfig的set_config和get_config函数完成。同时也支持以环境变量的方式设置键值对。 EdgePredictorConfig的具体使用方法可以
参考开发工具包中的demo工程。
针对Jetson开发工具包,目前EdgePredictorConfig的运行参数所支持的Key包括如下项:
```

/** * @brief 当有同类型的多个设备的时候,使用哪一个设备,如: * GPU: 使用哪张GPU卡 * EdgeBoard(VMX), Movidius NCS:使用哪一张加速卡 * 值类型: int * 默认值:0 */ static constexpr auto PREDICTOR_KEY_DEVICE_ID = "PREDICTOR_KEY_DEVICE_ID"; /** * @brief 生成 batch_size 为 max_batch_size 的优化模型,单次预测图片数量可以小于或等于此值(推荐等于此值,见release notes) * 值类型: int * 默认值:4 */ static constexpr auto PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE = "PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE"; /** * @brief 设置device对应的GPU卡可以支持的最大并发量 * 实际预测的时候对应GPU卡的最大并发量不超过这里设置的范围,否则预测请求会排队等待预测执行 * 值类型: int * 默认值:1 static constexpr auto PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY = "PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY"; /** * @brief 是否开启fp16模式预测,开启后预测速度会更快,但精度会略有降低。并且需要硬件支持fp16 * 值类型: bool * 默认值: false */ static constexpr auto PREDICTOR_KEY_GTURBO_FP16 = "PREDICTOR_KEY_GTURBO_FP16";

/**

* @brief 模型编译等级

```
*1:如果当前max_batch_size与历史编译产出的max_batch_size不相等时,则重新编译模型(推荐)
```

Baidu 百度智能云文档

*2:无论历史编译产出的max_batch_size为多少,均根据当前max_batch_size重新编译模型 * 值类型: int * 默认值:1 static constexpr auto PREDICTOR_KEY_GTURBO_COMPILE_LEVEL = "PREDICTOR_KEY_GTURBO_COMPILE_LEVEL"; /** * @brief GPU工作空间大小设置 * workspace_size = workspace_prefix * (1 << workspace_offset) * workspace_offset: 10 = KB, 20 = MB, 30 = GB * 值类型: int *默认值:WORKSPACE_PREFIX: 100, WORKSPACE_OFFSET: 20,即100MB static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX = "PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX"; static constexpr auto PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET = "PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET"; /** * @brief 需要使用的dla core * 值类型: int * 默认值:-1(不使用) static constexpr auto PREDICTOR_KEY_GTURBO_DLA_CORE = "PREDICTOR_KEY_GTURBO_DLA_CORE"; /** * @brief 自定义缓存文件存储路径 * 值类型: string *默认值: ~/.baidu/easyedge/mcache/{model_id * 1000000 + release_id} static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_DIR = "PREDICTOR_KEY_GTURBO_CACHE_DIR"; /** * @brief 自定义缓存文件命名,默认即可 * 值类型: string *默认值:根据配置自动生成 static constexpr auto PREDICTOR_KEY_GTURBO_CACHE_NAME = "PREDICTOR_KEY_GTURBO_CACHE_NAME"; /** * @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号 * 值类型: string *默认值:空 static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";

PREDICTOR_KEY_GTURBO_CACHE_NAME:首次加载模型会先对模型进行编译优化,通过此值可以设置优化后的产出文件名。

PREDICTOR_KEY_GTURBO_CACHE_DIR:首次加载模型经过编译优化后,产出的优化文件会存储在这个位置,可以按需修改。

PREDICTOR_KEY_GTURBO_WORKSPACE_PREFIX、PREDICTOR_KEY_GTURBO_WORKSPACE_OFFSET:设置运行时可以被用来使用的最大临时显存。

PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE:此值用来控制批量图片预测可以支持的最大图片数,实际预测的时候单次预测图片数需等于此值。

PREDICTOR_KEY_DEVICE_ID:设置需要使用的 GPU 卡号,对于Jetson,此值无需更改。

PREDICTOR_KEY_GTURBO_COMPILE_LEVEL: 模型编译等级。通常模型的编译会比较慢,但编译产出是可以复用的。可以在第一次加载模型的时候设置合理的 max_batch_size 并在之后加载模型的时候直接使用历史编译产出。是否使用历史编译产出可以通过此值 compile_level 来控制,当此值为 0 时,表示忽略当前设置的 max_batch_size 而仅使用历史产出(无历史产出时则编译模型);当此值为 1 时,会比较历史产出和当前设置的 max_batch_size 是否相等,如不等,则重新编译;当此值为 2 时,无论如何都会重新编译模型。

PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY:通过此值设置单张 GPU 卡上可以支持的最大 infer 并发量,其上限取决于硬件限制。init 接口 会根据此值预分配 GPU 资源,建议结合实际使用控制此值,使用多少则设置多少。注意:此值的增加会降低单次 infer 的速度,建议优先考虑 batch inference。

PREDICTOR_KEY_GTURBO_FP16:默认是 fp32 模式,置 true 可以开启 fp16 模式预测,预测速度会有所提升,但精度也会略微下降,权衡使用。注意:不是所有模型都支持 fp16 模式,也不是所有硬件都支持 fp16 模式。已知不支持fp16的模型包括:EasyDL图像分类高精度模型。

预测视频接口

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

classVideoDecoding :

```
/**
* @brief 获取输入源的下一帧
* @param frame_tensor
* @return
*/
virtual int next(FrameTensor &frame_tensor) = 0;
/**
* @brief 显示当前frame_tensor中的视频帧
* @param frame_tensor
* @return
*/
virtual int display(const FrameTensor &frame_tensor) = 0;
/**
* @brief 将当前frame_tensor中的视频帧写为本地视频文件
* @param frame_tensor
* @return
*/
virtual int save(FrameTensor &frame_tensor) = 0;
/**
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
* @brief 获取视频的width属性
 * @return
 */
virtual int get_width() = 0;
/**
* @brief 获取视频的height属性
* @return
*/
virtual int get_height() = 0;
```

struct VideoConfig
```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type;
                          // 输入源类型
  std::string source_value;
                        // 输入源地址, 如视频文件路径、摄像头index、网络流地址
 int skip_frames{0};
                      // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
 int retrieve_all{false};    // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
  int input_fps{0};
                       // 在采取抽帧之前设置视频的fps
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
  bool enable_display{false};
  std::string window_name{"EasyEdge"};
 bool display_all{false};
                        // 是否显示所有frame, 若为false, 仅显示根据skip_frames抽取的frame
  bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
  bool save_all{false};
                        // 是否存储所有frame,若为false,仅存储根据skip_frames抽取的frame
  std::map<std::string, std::string> conf;
};
```

source_type: 输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source_value: 若source_type为视频文件,该值为 指向视频文件的完整路径;若source_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source_type为网络 视频流,则为该视频流的完整地址。skip_frames:设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true,标记为 is_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被 抽取返回,以作为显示或存储用。input_fps:用于抽帧前设置fps。resolution:设置摄像头采样的分辨率,其值请参考easyedge_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。conf:高级选项。部分配置会通过该map来设置。

注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo_video_inference。

日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

```
EdgeLogConfig log_config;
log_config.enable_debug = true;
global_controller()->set_log_config(log_config);
```

返回格式

预测成功后,从 EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index; // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测或图像分割时才有
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割时才有
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding, 游程编码的mask
};
```

关于矩形坐标

x1 图片宽度 = 检测框的左上角的横坐标 y1 图片高度 = 检测框的左上角的纵坐标 x2 图片宽度 = 检测框的右下角的横坐标 y2 图片高度 = 检测框 的右下角的纵坐标

以上字段可以参考demo文件中使用opencv绘制的逻辑进行解析

http服务

1. 开启http服务 http服务的启动参考demo_serving.cpp文件。

/** * @

- * @brief 开启一个简单的demo http服务。 * 该方法会block直到收到sigint/sigterm。
- * http服务里,图片的解码运行在cpu之上,可能会降低推理速度。
- * @tparam ConfigT
- * @param config
- * @param host
- * @param port
- * @param service_id service_id user parameter, uri '/get/service_id' will respond this value with 'text/plain'
- * @param instance_num 实例数量,根据内存/显存/时延要求调整
- * @return

*/

template<typename ConfigT>
int start_http_server(
 const ConfigT &config,
 const std::string &host,
 int port,
 const std::string &service_id,
 int instance_num = 1);

2. 请求http服务

开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。

URL中的get参数:

参数	说明		默认值
threshold	阈值过滤,	0~1	如不提供,则会使用模型的推荐阈值

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

Python请求示例

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
        'http://127.0.0.1:24401/',
        params={'threshold': 0.1},
        data=img).json()
```

Java请求示例参考这里

http 返回数据

字段	类型说明	其他
error_code	Number	0为成功,非0参考message获得具体错误信息
results	Array	内容为具体的识别结果。其中字段的具体含义请参考接口使用·返回格式一节
cost_ms	Number	预测耗时ms,不含网络交互时间

返回示例

```
"cost_ms": 52,
"error_code": 0,
"results": [
    "confidence": 0.94482421875,
    "index": 1
    "label": "IronMan",
    "x1": 0.059185408055782318,
    "x2": 0.18795496225357056,
    "y1": 0.14762254059314728,
     "y2": 0.52510076761245728
    "confidence": 0.94091796875,
    "index": 1.
    "label": "IronMan",
    "x1": 0.79151463508605957,
    "x2": 0.92310667037963867,
    "y1": 0.045728668570518494,
    "y2": 0.42920106649398804
```

多线程预测 Jetson 系列 SDK 支持多线程预测,创建一个 predictor,并通过 PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY 控制所支持的最大 并发量,只需要 init 一次,多线程调用 infer 接口。需要注意的是多线程的启用会随着线程数的增加而降低单次 infer 的推理速度,建议优先使用 batch inference 或权衡考虑使用。

已知问题 1. 多线程时图片按线程分配不均 或 不同batch size的图片交叉调用infer接口时,部分结果错误

A:EasyDL图像分类高精度模型在有些显卡上可能存在此问题,可以考虑填充假图片数据到图片比较少的线程或batch以使得infer间的图片绝对平均。

2. 显存持续增长或遇到 terminate called after throwing an instance of 'std::runtime_error' what(): Failed to create object

A:如果遇到此问题,请确认没有频繁调用 init 接口,通常调用 infer 接口即可满足需求。

3. 开启 fp16 后, 预测结果错误

A:不是所有模型都支持 fp16 模式。目前已知的不支持fp16的模型包括:EasyDL图像分类高精度模型。目前不支持的将会在后面的版本陆续支持。

4. 部分模型不支持序列化

A:针对JetPack4.4、4.5版本,部分模型无法使用序列化,如已知的BML的MobileNetV1-SSD和物体检测高性能模型。需要每次加载模型的时候

编译模型,过程会比较慢。此问题将在后续JetPack版本中修复。目前JetPack4.6版本SDK已修复该问题。

开发板信息查询与设置查询L4T或JetPack版本查询JetPack版本信息,可以通过下面这条命令先查询L4T的版本。

在终端输入如下命令并回车 \$ head -n 1 /etc/nv_tegra_release **就会输出类似如下结果** \$ # # R32 (release), REVISION: 4.3, GCID: 21589087, BOARD: t210ref, EABI: aarch64, DATE: Fri Jun 26 04:38:25 UTC 2020

从输出的结果来看,板子当前的L4T版本为R32.4.3,对应JetPack4.4。 注意,L4T的版本不是JetPack的版本,一般可以从L4T的版本唯一对应到 JetPack的版本,下面列出了最近几个版本的对应关系:

L4T R32.6.1 --> JetPack4.6 L4T R32.5.1 --> JetPack4.5.1 L4T R32.5 --> JetPack4.5 L4T R32.4.3 --> JetPack4.4 L4T R32.4.2 --> JetPack4.4DP L4T R32.2.1 --> JetPack4.2.2 L4T R32.2.0 --> JetPack4.2.1

功率模式设置与查询不同的功率模式下,执行AI推理的速度是不一样的,如果对速度需求很高,可以把功率开到最大,但记得加上小风扇散热~

```
**1. 运行下面这条命令可以查询开发板当前的运行功率模式**

$ sudo nvpmodel -q verbose

**$ NV Power Mode: MAXN**

**$ 0**

**如果输出为MAXN代表是最大功率模式**

**2. 若需要把功率调到最大,运行下面这条命令**

$ sudo nvpmodel -m 0
```

如果你进入了桌面系统,也可以在桌面右上角有个按钮可以切换模式

3. 查询资源利用率 \$ sudo tegrastats

FAQ 1. EasyDL SDK与云服务效果不一致,如何处理?

后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

2. 运行SDK报错 Authorization failed 日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888"
./easyedge_demo ...

3. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl_slist_append(headers, "Expect:");

4. 运行demo时报找不到libeasyedge_extension.so

需要export libeasyedge_extension.so所在的路径,如路径为/home/work/baidu/cpp/lib,则需执行:

export LD_LIBRARY_PATH=/home/work/baidu/cpp/lib:\${LD_LIBRARY_PATH}

或者在编译完后执行如下命令将lib文件安装到系统路径:

sudo make install

如不能安装,也可手动复制lib下的文件到/usr/local/lib下。

5. 运行demo时报如下之一错误

2020-12-17 16:15:07,924 INFO [EasyEdge] 547633188880 Compiling model for fast inference, this may take a while (Acceleration) Killed

或

2020-12-17 16:15:07,924 INFO [EasyEdge] 547633188880 Build graph failed

请适当降低PREDICTOR_KEY_GTURBO_MAX_BATCH_SIZE和PREDICTOR_KEY_GTURBO_MAX_CONCURRENCY的值后尝试。

6. 运行有损压缩加速的模型,运算精度较标准模型偏低首先请保证数据集不会太小。其次可以通过将模型目录RES中的calibrationtable移除,并通过将PREDICTOR_KEY_GTURBO_FP16设置为true,使用FP16的运算精度重新评估模型效果。若依然不理想,可将calibrationtable移除并将PREDICTOR_KEY_GTURBO_FP16设置为false,从而使用更高精度的FP32的运算精度。

EasyDL 结构化数据使用说明

EasyDL结构化数据介绍

心简介

Hi,您好,欢迎使用百度EasyDL结构化数据

目前EasyDL结构化数据支持训练以下模型:

• 表格数据预测

通过机器学习技术从表格化数据中发现潜在规律,从而创建机器学习模型,并基于机器学习模型处理新的数据,为业务应用生成预测结果

• 时序预测

通过机器学习技术从历史数据中发现潜在规律,从而对未来的变化趋势进行预测。

心 可视化操作

无需机器学习专业知识,模型创建-数据上传-模型训练-模型发布全流程可视化便捷操作,最快15分钟即可获得一个高精度模型

心 操作步骤

Step 1 创建模型

确定模型名称,记录希望模型实现的功能

Step 2 上传并标注数据

Step 3 训练模型并校验效果

选择部署方式与算法,用上传的数据一键训练模型

模型训练完成后,可在线校验模型效果

Step 4 发布模型

根据训练时选择的部署方式,将模型以云端API的方式发布使用

更详细的操作指导,请参考各类模型的技术文档

表格数据预测

表格数据预测介绍

简介

Hi,您好,欢迎使用百度EasyDL定制化训练和服务平台。

定制表格数据预测模型,旨在帮助用户通过机器学习技术从表格化数据中发现潜在规律,从而创建机器学习模型,并基于机器学习模型处理新的 数据,为业务应用生成预测结果。本文介绍**表格数据预测**模型,根据预测数据的不同,可以分为如下几种类型:

- 回归:目标列是连续的实数范围,或者属于某一段连续的实数区间。如在销量预测场景中,销量值可能是某个取值范围内的任意值,解决该问 题的模型属于回归模型。
- 二分类:目标列是离散值,且只有两种可能的取值。如在精准营销场景中预测一个用户是否为潜在购买用户,其目标列仅存 在"True"和"False"两种取值,解决该问题的模型属于二分类模型。
- 多分类:目标列是离散值,并具有有限的可能取值。如在用户分类场景中,根据用户的历史消费数据,将用户划分到不同消费偏好的类别中, 解决该问题的模型属于多分类模型。

以下是关于表格数据预测模型的技术文档。

∞ 应用场景

- 精准营销:从客户消费记录中挖掘客户群的共有特征,分析出客户的购物偏好,从而实现广告的精准投放
- 信用评分:金融公司分析客户的历史行为数据,建立用户信用模型,从而确定贷款额度等
- 价格预测:从历史数据中发现商品的变化规律以及影响价格的因素,从而为未来的商业行为提供支持
- 客户流失预测:根据客户历史数据获得数据挖掘模型,从而生成客户流失预测列表,为市场营销策略提供有价值的业务洞察力。
- 使用流程

训练模型的基本流程如下图所示,全程可视化简易操作。在数据已经准备好的情况下,最快几分钟即可获得定制模型。

下面将详细介绍每一步的操作方式和注意事项。如果文档没有解决您的问题,请在百度云控制台内提交工单反馈。

Ē	L	E	ୢୄ
<u> </u>	0	0	
1. 创建模型	2. 上传数据	3. 训练模型并校验效果	4. 上线模型获取API

数据准备

心 表格数据集介绍

表格数据介绍

训练数据的质量决定了训练所得模型效果可达到的上限。数据上传后无法修改其内容。如果在导入训练数据后需要对其进行更改,必须重新导 入。

数据要求

数据文件格式要求:

- 目前仅支持CSV格式的数据文件
- 一次仅能上传一个文件,可以是一个CSV文件或由多个CSV文件压缩成的zip包
- 单个上传文件大小不能超过5GB
- 一个数据集包含的总文件大小不能超过20GB

数据文件内容要求:

- 当数据文件包含列名时,列名称可以包含字母、数字和下划线(_),但不能以下划线开头。
- 文件内容以换行符(即字符"\n",或称为LF)分隔各行,行内容以英文逗号(即字符",")分隔各列
- 必须包含要预测的值即目标列,且目标列的数据类型会决定模型的类型。
- 文件中文本列取值长度不能超过4096个字符。
- 必须至少包含两列,且不得超过1000列。
- 数据集的总行数不能超过1000万行。

- zip包中的多个CSV文件必须使用相同的编码格式,都包含列名或都不包含列名;且列的顺序必须保持一致
- 在扩充数据集时,新导入数据文件的首行与数据集的列名相同时,将被视为列名,否则将被视作数据
- 心 创建表格数据集

在EasyDL经典版中,您可以在"数据总览"页面,完成数据集创建、数据导入等操作,为模型构建准备好数据。

创建数据集

数据集需要先定义,然后再导入数据。

- 1. 单击"数据总览",进入数据集列表页面。
- 2. 单击"创建数据集",进入数据集创建页面。
- 输入数据集名称,单击"完成"结束创建。
 完成创建后,可以在数据集列表中查看新建的数据集。系统默认生成V1版本,当前数据内容为空,可以通过导入的方式向其中添加数据。

导入数据

通过导入的方式可以向数据集中添加或追加数据。

- 1. 单击"数据总览",进入数据集列表页面。
- 2. 单击待导入数据集所在行的"导入"按钮,进入数据集导入页面。
- 3. 导入数据文件。

导入的数据文件可以是CSV文件或由CSV文件组成的压缩包文件。 如果导入的是CSV文件,支持数据预览,如果是压缩包格式,则不支持预览。

- 4. 根据数据文件的实际情况进行列名设置。
 - 设置首行为列名:将导入的数据文件中的首行作为列名。
 - 设置首行非列名:此时系统会自动生成列名,而将首行作为数据。
- 5. 单击"确认并返回"完成导入操作。

模型训练

₯表格预测模型介绍

表格数据预测模型介绍

表格数据预测模型是基于结构化数据进行建模,系统会基于用户上传的数据使用预置算法进行模型构建与训练。表格数据预测模型目前支持回归和分类两种类型的模型,其中分类模型包括二分类和多分类模型。

回归

回归模型通常用来预测一个数值,其反映的是变量或属性间的依赖关系,建模过程即求解将一个或多个变量映射到一个实数值的函数。它可以应 用到市场营销的各个方面,如销量预测、价格预测等场景中。

分类

分类是找出一组对象的共同或差异点以将其划分为相同或不同的类,其目的是通过分类模型,将数据项映射到某个给定的类别。它可以应用到客 户分群、客户行为预测、客户满意度分析等场景中。

其中二分类模型是指预测值包括两种类别,多分类模型是指预测值包括多种类别。

心 创建模型

在EasyDL经典版中,您可以在"模型中心"进行模型的创建。在EasyDL中模型可以包括多个版本,每次训练会生成一个版本。各个版本的模型之间 相互独立,可以分别进行版本发布等操作。

创建模型

模型需要先创建,然后才能进行训练。

1. 单击"创建模型",进入模型创建页面。

2. 填写模型创建信息,如下图所示。

·	经典版-表格数据预测模型 提文工单 定制基于素格数据的模型。可实现表格中某列类则或数值的预测,该类模型可应用在销量预测、接信评估等场景。此类模型训练 ^{收起 个} 速度较快,具有X个特征列的Y条样本的表格数据,一般可在30分钟内训练完毕。	
模型中心	模型列表 > 创建模型	
我的模型	模型关列: 表格数据预则	
创建模型		
训练模型	* 模型名称:	
校验模型	機型归屬: 公司 个人	
发布模型	谢输入公司各称	
EasyData数据服务	 所届行业: 请选择行业 	
数据总览	• 应用场最: 请选择应用场最 V	
	● 邮箱地址: *****@34.com	
	• 联系方式: 111****111 ②	
	* 功能描述:	
	0/500	
	ت. اور م	

- 模型名称:指定模型的名称
- 模型归属:公司或个人,并输入相关名称
- 所属行业:请根据实际情况进行选择
- 应用场景:请根据实际情况进行选择
- 邮箱地址:请根据实际情况进行设置
- 联系方式:请根据实际情况进行设置
- 功能描述:用于记录模型创建的背景、用途等方面的信息。
- 3. 单击"下一步"完成模型创建。

完成创建后,可以在模型列表中查看新建的模型。新建的模型不包含任何版本的模型,在训练后会生成新的模型版本。

心训练模型

EasyDL经典版提供的的机器学习算法不仅性能高、可扩展,还针对速度、规模和准确性进行了优化,可以在大规模数据集上进行训练。

训练模型

在准备好数据集并创建模型后,可以创建训练任务。

- 1. 单击"训练模型",进入模型训练页面。
- 2. 填写模型训练信息,如下图所示。

预算中。	定制基于 速度较的	且版-表格数据预测 F表格数调的模型,可实现表格中某 e,具有X个特征列的Y条样本的表标	模型 ^{(列类別或数} ^(数据, 一)	<mark>提交工单</mark> 效值的预测,该类模型可应用在销量预测、授信评估等场景。此类模型训练 般可在30分钟内训练完毕。	收起 ^
模型中心	训练模型				
我的模型	选择模型	diamonds价格预测	~		
创建模型	选择数据集	diamonds V1	~		
训练模型	选择目标列 2	price	~		
校验模型	算法类型 ?	· 自动 () 二分类 () 多分类			
发布模型	部署方式	公 公有云API			
EasyData数据服务		开始训练			
数据总览					

- 选择模型:选择要训练的模型
- 选择数据集:选择训练模型使用的数据集
- 选择目标列:从数据集中选择一列作为预测列
- 算法类型:包括二分类、多分类和回归,也可以选择自动,此时系统会根据数据集以及选择的目标列进行判断。
- 部署方式:当前仅支持公有云API方式。
- 3. 单击"开始训练"启动训练任务。

启动训练任务后,系统会在模型下的列表中创建一个新的模型版本,新建的版本处于"训练中"的状态,当处于"训练完成"状态时表示模型已完成训练。

4. 平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

ロスチバル伯知し・	各类算力价格如下	:
-----------	----------	---

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

优惠政策:

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有表格预测 操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。

⁰ 查看模型

训练任务结束后,可以查看模型的各项指标,以确定模型是否满足要求。不同类型的模型包含了不同的指标,用户可以根据实际的业务场景查看 关键指标是否满足要求。

查看模型

对于"训练完成"的模型,可以查看其评估结果。

- 1. 单击"我的模型",进入模型列表页面。
- 2. 单击待查看模型的"历史版本",进入模型版本列表页面。
- 3. 点击待查看模型版本所在行的"完整评估结果",系统展示评估结果页面。
 - 回归模型的评估结果包括各项常用评估指标以及特征重要性,如下所示:

模型管理 > diamonds价格预测	
diamonds价格预测评估报告 部署方式: 公有云API(目前仅支持) ∨ 版本: V1 ∨	
整体评估 特征 10列 目标列 price 算法类型 回归	
diamonds价格预测 V1整体效果欠佳 ,建议针对识别错误的样本示例继续优化模型效果。如何优化效果?	
名称	数值
MAE (Mean Absolute Error) 平均绝对误差 ⑦	6.192
MSE (Mean Squared Error) 均方误差 ⑦	1519.512
MAPE (Mean Absolute Percentage Error) 平均绝对百分比误差 ⑦	0.73%
R2 Score (决定系数) 回归得分函数 ⑦	1.000
详细评估	
特征重要性 ③ 最多展示重要性Top15的特征	

• 二分类模型的评估结果包括混淆矩阵、F1-Score阈值曲线、KS曲线、ROC曲线、P-R曲线、Lift曲线、Gain曲线和特征重要性,如下所示:

	部署方式 公有 训练时长 7分钟	云API(目前仅 ~	版本 V1	~					
	整体评估 特征 20列 目标列 TARGET_5Yrs 算法类型 二分类								
整体评估:	以 1 v 为正例 阈值 ⑦ 准確率 ⑦	值 ⑦	E1-score @	0.5 〇 恢复推荐阈值	精确率の	日本の			
	 70.5% 混淆矩阵 ⑦ 预测值 		67.8%	0	71.7%	83.3%			
	真实值	1 83.33%	0 16.67%						
	0	48.47%	51.53%						









KS曲线 ⑦

TPR-FPR

0.8

0.7

0.6

0.4

0.3

0.1

0

• TPR • FPR

最大差值0.38







0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1



详细评估:

阈值

特征重要性⑦ 最多展示重要性Top15的特征



• 多分类模型的评估结果包括混淆矩阵、ROC曲线、P-R曲线和特征重要性,如下所示:



ROC曲线 ⑦ (AUC ⑦ 值为 0.998)

P-R曲线 ⑦





特征重要性⑦ 最多展示重要性Top15的特征



の 校验模型

在完成训练后,发布模型前,可以先进行模型校验,以确保模型在实际环境中能获得预期的性能。

校验模型

在训练任务成功完成后,即可使用实际数据进行校验。

- 1. 单击"校验模型",进入模型校验页面。
- 2. 选择要校验的模型及其版本。
- 3. 单击"启动模型校验服务"。
 - 校验数据支持两种输入方式,表单方式或json格式,并支持切换
 - 系统自动填充了校验数据,用户可以直接使用预置的数据进行预测,也可以修改后再进行预测。
- 4. 单击"预测",可以在右侧结果面板中查看预测结果。

```
校验示例如下所示:
```

校验模型	
选择模型: iris分类	✓ 部署方式: 公有云API(目前仅支持) ✓ 选择版本: V1 ✓ □PII(結果 0nG(けん)効果2
预测数据	□ 受制 之切换son视图 国 预则 1 { 2 "spec: "setosa"
字段名 类型	取值 3 }
sepal_length 数值	5.2
sepal_width 数值	3.5
petal_length 数值	1.4
petal_width 数值	0.2
预测	申请上线

模型发布

训练完成后,可将模型部署在公有云服务器上,通过API进行调用。

公有云API

- 训练完成的模型存储在云端,可通过独立Rest API调用模型,实现AI能力与业务系统或硬件设备整合
- 具有完善的鉴权、流控等安全机制,GPU集群稳定承载高并发请求
- 一次API请求中最多可以包含100条预测数据

相关费用

将模型发布为API后,将获得1000次免费调用次数,如需更多调用量,请在百度云控制台内提交工单反馈。

心 公有云部署

^の如何发布表格数据预测API

训练完毕后可以在左侧导航栏中找到【发布模型】,依次进行以下操作即可发布公有云API:

- 选择模型
- 选择部署方式「公有云部署」
- 选择版本
- 自定义服务名称、接口地址后缀
- 申请发布

发布模型

申请发布后,通常的审核周期为T+1,即当天申请第二天可以审核完成。如果需要加急、或者遇到莫名被拒的情况,请在百度云控制台内提交工 单反馈。

发布模型界面示意:

选择模型:	diamonds价格预测 V		标准接口规?	范参考 _{考说明:}			
即曾万式;	公有云API V						
选择版本:	V1 ~		字段名称	必须	类型	说明	
* 服务名称: * 接囗地址:	https://aip.baidubce.com/rpc/2.0/ai_cust	tom/v1/table_infe	include_req	否	boolean	返回结果是否包含特征数据: false,不包含; true,包含,默认 为false	
其他要求:	1/ 若接口无法满足您的需求,请描述希望	是解决的问题,	data	是	array(object)	待预测数据,每条待预测数据是 由各个特征及其取值构成的键值 对的集合	
	500汉字以内		标准接口响应字段说明:				
		0/500	字段名称	必须	类型	说明	
	提交申请		log_id	是	number	唯一的log id,用于问题定位	
			error_code	否	number	错误码,当请求错误时返回	
			error_msg	否	string	错误描述信息, 当请求错误时返 回	
			results	否	array(object)	预测结果数组	

接口赋权

在正式使用之前,还需要做的一项工作为接口赋权,需要登录EasyDL经典版控制台中创建一个应用,获得由一串数字组成的appid,然后就可以 参考接口文档正式使用了

✓ 返回EasyDL总览	产品服务	/ EasyDL定制训练平台 - 应用列表						
经典版	应用	列表						
公有云服务管理 ^	+ (射跳应用						
• 应用列表		应用名称	AppID	API Key	Secret Key	创建时间	操作	
* 权限管理	1	识别猫和狗	15313490	gRHegw68kpnym0H6K7GI6wUd	•••••• 显示	2018-12-31 17:24:27	报表管理制除	
 用量统计 	2	TRY	14676113	75a5pbP7mhpBWNION2NYoYcX	••••••• 显示	2018-11-05 11:35:54	报表 管理 删除	
 监控报表 计未支撑 	3	皮皮虾	14403281	i6HhGApGanhWuYex4Oel2uG8	•••••• 显示	2018-10-11 19:55:26	报表管理删除	
 技术又档 售卖服务 								< 1 >
本地部署服务管理								
设备端服务管理								
数据管理服务								

同时支持在「公有云服务管理」-「权限管理」中为第三方用户配置权限

示意图如下:

✔ 返回EasyDL总览	产品服务 / EasyDL定制训练平台 - 4	权限管理		
经典版	权限管理			
公有云服务管理 へ ・ 应用列表	■ 服务权限管理 支持为您或者他人的AppID配置	重可用服务权限。		
权限管理	序号 服务名称		相关操作	
▫ 用量统计	1 test		配置权限	
◎ 监控报表	2 cherry		配置权限	
 技术文档 	3 horse的服务名称	X	配置权限	
 售卖服务 	4 sea		配置权限	
本地部署服务管理	5 手势识别		配置权限	
设备端服务管理				
数据官理服务				
≮ 返回EasyDL总览	产品服务 / EasyDL定制训练 ^I	平台 - 权限管理 / 配置权限		
经典版	配置权限			
公有云服务管理 ^	呢怎么你, ta	set		
• 应用列表	יתיום פעאת	51		
 权限管理 	可用APP ID:	14403281 皮皮虾 ×		
▫ 用量统计		14676113 TRY ×		
◎ 监控报表		15313490 识别猫和狗 ×		
◎ 技术文档		请添加 🕇		
◎ 售卖服务		保存取消		
本地部署服务管理				
本地部署服务管理 设备端服务管理				
本地部署服务管理 设备端服务管理 数据管理服务				
本地部署服务管理 设备端服务管理 数据管理服务				

^の API调用文档

本文档主要说明定制化模型发布后获得的API如何使用,如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

接口描述

基于自定义训练出的表格数据预测模型,实现表格数据预测。模型训练完毕后发布可获得定制化表格数据预测API

接口鉴权

1、在EasyDL经典版控制台创建应用

Baidu 百度智能云文档

♀ 全局 ∨						Q	₽	?	R	۲	Q
EasyDL定制训练平台	产品服务 / EasyD	DL定制训练平台-应	2用列表 / 创建应用								
概览	创建新应用]									
应用列表	• 应用字段·	. to at									
监控报表	122,713,123,101,1	test									
技术文档	•应用类型:	游戏娱乐			×						
云服务权限管理	•接口选择:	您可以为应用勾)选多个接口权限,使此	应用可以请求已勾选的接	口服务						
离线SDK管理		EasyDL	 ✓ 百美椅子训练 ✓ 识别眼睛和嘴目 	 ✓ 椅子颜色识别 巴 ✓ 声音分类 	✓ aa ✓ 商品检测						
		➡ 百度语音									
		➡ 文字识别									
		➡ 人脸识别									
		➡ 自然语言处	理								
		+ 内容审核									
		+ UNIT									
		+ 知识图谱									
		▲ 图像识别	2								
		★ 智能呼叫中	0								
		■ 国際技術									
		100000									
	•应用描述:	简单描述一下 点,请控制在	您使用人工智能服务的 500字以内	应用场景 , 如开发一款美	鏡相机,需要检測人脸关键						
1											

2、应用详情页获取AK SK

♥ 全局 ~					Q	≣	¢	?	8	۲	Q ~
EasyDL定制训练平台	产品服务 / EasyDL定制训练平台 - 应用列表 / 应用	日详情									
概览	应用详情	应用详情									
应用列表	编辑 查看文档	集 田 查看文档									
监控报表	应用名称	AppID		API Key			Secr	et Key			
技术文档	test2	14293933		bZsfqO2fUdHTDflxqF1	1mkTq			** 显示			
云服务权限管理 离线SDK管理	EasyDL定制化训练服务:									_	

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,与其他图像识别服务不同的是定制化图像识别服务以json方式请求。

Body请求接口:

{	
	"include_req": false,
	"data": <待预测数组>
}	

考虑到表格字段内容和长度的不固定性,我们建议您参考"校验服务"页面提供的详细信息。您可以访问该页

面:https://ai.baidu.com/easydl/app/validate/ml/models/verify, 启动校验服务后,点击"复制"按钮从中复制数据请求的 Body 部分作为参考模

板。这将帮助您理解如何灵活处理各种不同的字段。

以如下数据特征列为例,请求body格式为:

```
{
"include_req": false,
"data": [{
 "sepal_length": 5.1,
 "sepal_width": 3.5,
 "petal_length": 1.4,
 "petal_width": 0.2
 },
 {
 "sepal_length": 5.3,
 "sepal_width": 3.4,
 "petal_length": 1.2,
 "petal_width": 0.3
}
]
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必选	类型	可选值范围	说明
include_req	否	boolean	-	返回结果是否包含特征数据:false,不包含;true,包含,默认为false
include_proba	否	boolean	-	返回结果是否包含分类概率:false,不包含;true,包含,默认为false
data	是	array	-	待预测数据,每条待预测数据是由各个特征及其取值构成的键值对的集合

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码,当请求错误时返回
error_msg	否	string	错误描述信息,当请求错误时返回
batch_result	否	array(object)	预测结果数组

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

```
例如Access Token失效返回:
```

```
{
   "error_code": 110,
   "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

Baidu 百度智能云文档

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、data格式错误等等,可检查下请求数据格式是否有误。有疑问请在百度 云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或者代码格式有误。有疑问请在百度云控制台内提交工单反馈
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

⊙ 服务器部署

の Linux集成文档-Python

简介

本文档介绍 EasyDL 的 Linux Python SDK 的使用方法,适用于 EasyDL 和 BML。

EasyDL 通用版:

- 网络类型支持:图像分类,物体检测,图像分割,声音分类,表格预测
- 硬件支持:
 - Linux x86_64 CPU (基础版,加速版)
 - Linux x86_64 Nvidia GPU (基础版,加速版)
- 语言支持: Python 3.5, 3.6, 3.7, 3.8, 3.9

BML :

- 网络类型支持:图像分类,物体检测,声音分类
- 硬件支持:
 - Linux x86_64 CPU (基础版)

- Linux x86_64 Nvidia GPU (基础版)
- 语言支持: Python 3.5, 3.6, 3.7, 3.8, 3.9

Release Notes

时间	版本	说明
2022.10.27	1.3.5	新增华为Atlas300、飞腾Atlas300 Python SDK,支持图像分类、物体检测、人脸检测、实例分割
2022.09.15	1.3.3	EasyDL CPU普通版新增支持表格预测
2022.05.27	1.3.1	CPU、GPU普通版新增支持BML Cloud小目标检测模型
2021.12.22	1.2.7	声音分类模型升级
2021.10.20	1.2.6	CPU基础版、CPU加速版、GPU基础版推理引擎优化升级
2021.08.19	1.2.5	CPU基础版、CPU无损加速版、GPU基础版新增支持EasyDL小目标检测
2021.06.29	1.2.4	CPU、GPU新增EasyDL目标跟踪支持;新增http server服务启动demo
2021.03.09	1.2.2	EasyDL CPU加速版新增支持分类、高性能检测和均衡检测的量化压缩模型
2021.01.27	1.2.1	EasyDL经典版分类高性能模型升级;支持更多模型
2020.12.18	1.2.0	推理引擎升级;接口升级;性能优化
2020.09.17	1.1.19	支持更多模型
2020.08.11	1.1.18	性能优化
2020.06.23	1.1.17	支持更多EasyDL专业版模型
2020.04.16	1.1.15	技术优化;升级 OpenVINO 版本
2020.03.12	1.1.14	新增声音识别python sdk
2020.02.12	1.1.13	新增口罩模型支持
2020.01.16	1.1.12	预测函数默认使用推荐阈值
2019.12.26	1.1.11	EasyDL 专业版支持 SDK 加速版
2019.12.04	1.1.10	支持图像分割
2019.10.21	1.1.9	支持 EasyDL 专业版
2019.08.29	1.1.8	CPU 加速版支持
2019.07.19	1.1.7	提供模型更新工具
2019.05.16	1.1.3	NVIDIA GPU 支持
2019.03.15	1.1.0	架构与功能完善
2019.02.28	1.0.6	引擎功能完善
2019.02.13	1.0.5	paddlepaddle 支持
2018.11.30	1.0.0	第一版!

2020-12-18: 【接口升级】 序列号的配置接口从1.2.0版本开始已升级为新接口,以前的方式被置为deprecated,并将在未来的版本中移除。 请尽快考虑升级为新的接口方式,具体使用方式可以参考下文介绍以及demo工程示例,谢谢。

快速开始

1. 安装依赖

- 根据引擎的不同,SDK 依赖了不同的底层引擎。根据所需自行安装。
- 使用声音分类SDK需要安装额外依赖

* pip 安装 resampy pydub six librosa 音频默认格式支持wav文件预测,如果需要预测mp3等其他音频格式的数据需要系统额外安 装ffmpeg (windows系统的ffmpeg已基在sdk中无需额外安装,linux系统需要手动安装)

• 使用表格预测SDK需要安装额外依赖 pip安装 brotlipy==0.7.0 certifi==2020.6.20 joblib==1.0.1 kaggle==1.5.12 Pillow py4j pycosat pythondateutil python-slugify ruamel_yaml text-unidecode threadpoolctl flask pandas==1.0.5 scikit-learn==0.23.2 lightgbm==2.2.3 catboost==0.24.1 xgboost==1.2.0 numpy==1.19.5 scipy==1.5.2

psutil==5.7.2 pypmml==0.9.7 torch==1.8.0 jieba==0.42.1 pyod==0.8.5 pyarrow==6.0.0 scikit-optimize==0.9.0 pyspark==3.3.0 另外ml算法安装(目前只支持python3.7)

pip install BaiduAl_TabularInfer-0.0.0-cp37-cp37m-linux_x86_64.whl 安装 paddlepaddle

• 使用x86_64 CPU 基础版预测时必须安装(目标跟踪、表格预测除外):

python -m pip install paddlepaddle==2.2.2 -i https://mirror.baidu.com/pypi/simple

若 CPU 为特殊型号,如赛扬处理器(一般用于深度定制的硬件中),请关注 CPU 是否支持 avx 指令集。如果不支持,请在paddle官网安装 noavx 版本

• 使用NVIDIA GPU 基础版预测时必须安装(目标跟踪、表格预测除外):

python -m pip install paddlepaddle-gpu==2.2.2.post101 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA10.1的 PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2 -i https://mirror.baidu.com/pypi/simple #CUDA10.2的PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2.post110 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.0的 PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2.post111 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.1的 PaddlePaddle

python -m pip install paddlepaddle-gpu==2.2.2.post112 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html #CUDA11.2的 PaddlePaddle

不同cuda版本的环境,请参考paddle文档安装合适的 paddle 版本。 不被 paddle 支持的 cuda 和 cudnn 版本,EasyEdge 暂不支持

安装 OpenVINO 使用 x86_64 CPU 加速版 SDK 预测时必须安装。

1) 请参考 OpenVINO toolkit 文档安装 2021.4版本, 安装时可忽略Configure the Model Optimizer及后续部分

2) 运行之前,务必设置环境变量

source /opt/intel/openvino_2021/bin/setupvars.sh

安装 cuda、cudnn

• 使用Nvidia GPU 加速版预测时必须安装。依赖的版本为 cuda9.0、cudnn7。版本号必须正确。

安装 pytorch (torch >= 1.7.0)

- 目标跟踪模型的预测必须安装pytorch版本1.7.0及以上(包含:Nvidia GPU 基础版、x86_64 CPU 基础版)。
- 目标跟踪模型Nvidia GPU 基础版还需安装依赖cuda、cudnn。

关于不同版本的pytorch和CUDA版本的对应关系:pytorch官网目标跟踪模型还有一些列举在requirements.txt里的依赖(包括torch >= 1.7.0),均可使用pip下载安装。

pip3 install -r requirements.txt

2. 安装 easyedge python wheel 包 安装说明

pip3 install -U BaiduAI_EasyEdge_SDK-{版本号}-cp36-cp36m-linux_x86_64.whl

具体名称以 SDK 包中的 whl 为准。 安装说明:华为 Atlas300 除了需要安装BaiduAl_EasyEdge_SDK包,还需安装

pip3 install -U EasyEdge_Devkit_Atlas300-{版本号}-cp36-cp36m-linux_x86_64.whl

pip3 install -U EasyEdge_Devkit_Phytium.Atlas-{版本号}-cp36-cp36m-linux_aarch64.whl

3. 使用序列号激活

丧取序列号						
转集线服务说明 发布纯黑线服务,若训练现就种模型跟著在本地,真线调用模型,可以选择将模型跟著在本地的服务器、小型设备、软键一体方来专项还配建件上, 进口AP、SCK是一步集成,更适差应不同位步结像, 						
服务器 通用小型设备 专项适配硬件						
SDK API						
此处发布、下载的SDK为未授权SDK,需要前往控	制台获取序列号激活后才能正式使用	。SDK内附有对应版本的Demo及开发文档,开	发者可参考源代码完成开发。			
模型名称	发布版本 🗘	应用平台	模型加速	发布状态	发布时间 🗘	
		通用X86 CPILL inux	基础版	 已发布 	2021-08-19 20:24	下载SDK
		ENIVER CHO-FUNDY	精度无损压缩加速	 已发布 	2021-08-19 20:24	下载加速版SDK
	134318-V1	2010-24-000 L L Inco	基础版	•已发布	2021-08-19 20:35	下截SDK
ani_inear	宣看性能报告	94912/0FO-CITUX	精度无损压缩加速	• 已发布	2021-08-19 20:34	下载加速版SDK
		Webberger and the states	基础版	 已发布 	2021-08-19 18:17	下截SDK

修改demo.py 填写序列号

```
pred = edge.Program()
pred.set_auth_license_key("这里填写序列号")
```

4. GPU 加速版 使用 GPU 加速版,在安装完 whl 之后,必须:

1. 从这里下载 TensorRT7.0.0.11 for cuda9.0,并把解压后的 lib 放到 C++ SDK 的 lib 目录或系统 lib 目录

2. 运行时,必须在系统库路径中包含 C++ SDK 下的lib目录。如设置LD_LIBRARY_PATH

cd \${SDK_ROOT}

1. 安装 python wheel 包 tar -xzvf python/*.tar.gz pip install -U {对应 Python 版本的 wheel 包}

2. 设置 LD_LIBRARY_PATH tar -xzvf cpp/*.tar.gz export EDGE_ROOT=\$(readlink -f \$(Is -h | grep "baidu_easyedge_linux_cpp")) export LD_LIBRARY_PATH=\$EDGE_ROOT/lib

3. 运行 demo python3 demo.py {RES文件夹路径} {测试图片路径}

如果是使用 C++ SDK 自带的编译安装的 OpenCV,LD_LIBRARY_PATH 还需要包括 C++ SDK的 build 目录下的 thirdparty/lib 目录

如果没有正确设置 LD_LIBRARY_PATH,运行时可能报错:

ImportError: libeasyedge.so.0.4.3: cannot open shared object file: No such file or directory ImportError: libopencv_core.so.3.4: cannot open shared object file: No such file or directory

5. 测试 Demo

5.1 表格预测 输入对应模型文件夹(默认为RES)和测试数据地址(csv文件地址),运行:

python3 demo.py {model_dir} {/xxx/xxx.csv}

6. 测试Demo HTTP 服务 输入对应的模型文件夹(默认为RES)、序列号、设备ip和指定端口号,运行:

 $python 3 \ demo_serving.py \ \{model_dir\} \ \{serial_key\} \ \{host, \ default \ 0.0.0.0\} \ \{port, \ default \ 24401\} \ (host, \ default \ 0.0.0.0) \ (host, \ default \ 24401) \ (host, \ default \ 0.0.0.0) \ (host, \ default \ 24401) \ (host, \ default \ 2$

后,会显示:

Running on http://0.0.0.0:24401/

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片或者视频来进行测试。也可以参考`demo_serving.py`里 http_client_test()函数请求http服务进行推理。

AI 百度EasyEdge	× +		
℃ ▲ 不安全	6		
	【图像分割】45274 分割-电池-设备端V1	调整阈值	────────────────────────────────────
		label	置信度
	- Han	battery	0.998
		tomato	0.997
		tomato	0.996
		battery	0.995
		tomato	0.995
		tomato	0.993
		tomato	0.972



图片加载失败

使用说明

使用流程 demo.py

import BaiduAI.EasyEdge as edge

```
pred = edge.Program()
pred.set_auth_license_key("这里填写序列号")
pred.init(model_dir={RES文件夹路径}, device=edge.Device.CPU, engine=edge.Engine.PADDLE_FLUID)
pred.infer_image({numpy.ndarray的图片})
pred.close()
```

demo_serving.py

```
import BaiduAl.EasyEdge as edge
from BaiduAl.EasyEdge.serving import Serving
```

server = Serving(model_dir={RES文件夹路径}, license=serial_key)

请参考同级目录下demo.py里:

pred.init(model_dir=xx, device=xx, engine=xx, device_id=xx)

对以下参数device\device_id和engine进行修改

server.run(host=host, port=port, device=edge.Device.CPU, engine=edge.Engine.PADDLE_FLUID)

初始化

● 接口

def init(self, model_dir, device=Device.CPU, engine=Engine.PADDLE_FLUID, config_file="conf.json", preprocess_file="preprocess_args.json", model_file="model", params_file="params", label_file="label_list.txt", infer_cfg_file="infer_cfg.json", device_id=0, thread_num=1, Args: model_dir: str device: BaiduAI.EasyEdge.Device,比如:Device.CPU engine: BaiduAI.EasyEdge.Engine,比如:Engine.PADDLE_FLUID config_file: str preprocess_file: str model_file: str params_file: str label_file: str 标签文件 infer_cfg_file: 包含预处理、后处理信息的文件 device_id: int 设备ID thread_num: int CPU的线程数 Raises: RuntimeError, IOError Returns: bool: True if success

使用 NVIDIA GPU 预测时,必须满足:

- 机器已安装 cuda, cudnn
- 已正确安装对应 cuda 版本的 paddle 版本
- 通过设置环境变量FLAGS_fraction_of_gpu_memory_to_use设置合理的初始内存使用比例

使用 CPU 预测时,可以通过在 init 中设置 thread_num 使用多线程预测。如:

 $pred.init (model_dir=_model_dir,\ device=edge.Device.CPU,\ engine=edge.Engine.PADDLE_FLUID,\ thread_num=1)$

预测图像

```
    接口
```

```
def infer_image(self, img, threshold=0.3, channel_order="HWC", color_format="BGR", data_type="numpy"):
"""
Args:
    img: np.ndarray or bytes
    threshold: float
        only return result with confidence larger than threshold
    channel_order: string
        channel order HWC or CHW
    color_format: string
        color format order RGB or BGR
    data_type: string
    image data type
Returns:
    list
"""
```

[•] 返回格式: [dict1, dict2, ...]

字段	类型	取值	说明
confidence	float	0~1	分类或检测的置信度
label	string		分类或检测的类别
index	number		分类或检测的类别
x1, y1	float	0~1	物体检测,矩形的左上角坐标 (相对长宽的比例值)
x2, y2	float	0~1	物体检测,矩形的右下角坐标(相对长宽的比例值)
mask	string/numpy.ndarray	图像分割的mask	

关于矩形坐标

x1 * 图片宽度 = 检测框的左上角的横坐标

y1 * 图片高度 = 检测框的左上角的纵坐标

x2*图片宽度 = 检测框的右下角的横坐标

y2 * 图片高度 = 检测框的右下角的纵坐标

可以参考 demo 文件中使用 opencv 绘制矩形的逻辑。

结果示例

● i) 图像分类

```
{
    "index": 736,
    "label": "table",
    "confidence": 0.9
}
```

● ii) 物体检测

```
{
    "y2": 0.91211,
    "label": "cat",
    "confidence": 1.0,
    "x2": 0.91504,
    "index": 8,
    "y1": 0.12671,
    "x1": 0.21289
}
```

• iii) 图像分割

```
{
    "name": "cat",
    "score": 1.0,
    "location": {
        "left": ...,
        "top": ...,
        "width": ...,
        "height": ...,
    },
    "mask": ....
}
```

mask字段中,data_type为numpy时,返回图像掩码的二维数组

```
{
  {
  {
  {
  {
  {
  0.0.0.0.0.0.0.0.0.0.0.0},
  },
  {
  0.0.0.1.1.1.0.0.0.0},
  {
  0.0.0.1.1.1.0.0.0.0},
  {
  0.0.0.1.1.1.0.0.0.0},
  {
  0.0.0.1.1.1.0.0.0.0},
  {
  0.0.0.0.0.0.0.0},
  {
  0.0.0.0.0.0.0.0},
  }
  其中1代表为目标区域,0代表非目标区域,0代表非目标区域
  }
```

data_type为string时,mask的游程编码,解析方式可参考 demo

预测视频 (目前仅限目标跟踪模型调用)

● 接口

def infer_frame(self, frame, threshold=None): """ 视频推理(抽帧之后) :param frame: :param threshold: :return: """

● 返回格式dict

字段	类型	说明
pos	dict1	当前帧每一个类别的追踪目标的像素坐标(tlwh)
id	dict2	当前帧每一个类别的追踪目标的id
score	dict3	当前帧每一个类别的追踪目标的识别置信度
label	dict4	class_idx(int)与label(string)的对应关系
class_num	int	追踪类别数

预测声音

• 使用声音分类SDK需要安装额外依赖 *pip 安装 resampy pydub* 音频默认格式支持wav文件预测,如果需要预测mp3等其他音频格式的数据需要 系统额外安装ffmpeg (windows系统的ffmpeg已集成在sdk中无需额外安装,linux系统需要手动安装)

● 接口

```
def infer_sound(self, sound_binary, threshold=0.3):
"""
Args:
    sound_binary: sound_binary
    threshold: confidence
Returns:
    list
"""
```

• 返回格式: [dict1, dict2, ...]

字段	类型	取值	说明
confidence	float	0~1	分类的置信度
label	string		分类的类别
index	number		分类的类别

表格预测

 考虑到表格字段内容和长度的不固定性,我们建议您参考"校验服务"页面提供的详细信息。您可以访问该页 面:https://ai.baidu.com/easydl/app/validate/ml/models/verify,并从中复制数据请求的 Body 部分作为参考模板。这将帮助您理解如何灵 活处理各种不同的字段。 接口

```
def infer_csv(self, data):
"""
结构化数据推理
Args:
data: pd.DataFrame or list or dict
Returns:
```

• 返回格式: list 接口直接反馈预测结果数组

升级模型 适用于经典版升级模型,执行bash update_model.sh,根据提示,输入模型路径、激活码、模型ID、模型版本,等待模型更新完毕即 可。

FAQ

Q: EasyDL 离线 SDK 与云服务效果不一致,如何处理? A: 后续我们会消除这部分差异,如果开发者发现差异较大,可联系我们协助处理。

Q: 运行时报错"非法指令"或"illegal instruction"A: 可能是 CPU 缺少 avx 指令集支持,请在paddle宫网 下载 noavx 版本覆盖安装

Q: NVIDIA GPU预测时,报错显存不足: A: 如以下错误字样:

paddle.fluid.core.EnforceNotMet: Enforce failed. Expected allocating <= available, but received allocating:20998686233 > available:19587333888. Insufficient GPU memory to allocation. at [/paddle/paddle/fluid/platform/gpu_info.cc:170]

请在运行 Python 前设置环境变量,通过export FLAGS_fraction_of_gpu_memory_to_use=0.3来限制SDK初始使用的显存量,0.3表示初始使用 30%的显存。如果设置的初始显存较小,SDK 会自动尝试 allocate 更多的显存。

Q:我想使用多线程预测,怎么做?如果需要多线程预测,可以每个线程启动一个Progam实例,进行预测。 demo.py文件中有相关示例代码。

注意:对于CPU预测,SDK内部是可以使用多线程,最大化硬件利用率。参考init的thread_num参数。

Q: 运行SDK报错 Authorization failed

情况一:日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP_PROXY 的环境变量通过代理处理自己的网络请求。如

export HTTP_PROXY="http://192.168.1.100:8888"
./easyedge_demo ...

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- MAC地址变化
- 磁盘变更
- BIOS重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 ~/.baidu/easyedge 目录,再重新激活。

情况三: Atlas Python SDK日志提示ImportError: libavformat.so.58: cannot open shared object file: No such file or directory 或者其他类似so找不到 可以在LD_LIBRARY_PATH环境变量加上libs和thirdpartylibs路径,例如

export LD_LIBRARY_PATH=/xxx/libs:/xxx/thirdpartylibs:\$LD_LIBRARY_PATH # tips: 这里/xxx需要替换为真实路径,/xxx路径查找方法如下

查找安装包内libs和thirdpartylibs路径的方法如下(以华为Atlas300 SDK为例,其他SDK查找方法类似):

pip3 show EasyEdge-Devkit-Atlas300 # 结果中会显示 Location 路径,也就是包的安装路径 **libs和thirdpartylibs两个路径在 Location 所指示的路径 easyedge_CANN 子文件夹下**

- ⊙通用小型设备部署
- 心 Windows集成文档

简介

本文档介绍表格预测通用小型设备Windows SDK的使用方法。

- 硬件支持:
 - Intel CPU 普通版 * x86_64
 - CPU 加速版 Intel Xeon with AVX2 and AVX512 Intel Core Processors with AVX2 Intel Atom Processors with SSE * AMD Core Processors with AVX2
- 操作系统支持
 - 普通版:64位 Windows 7 及以上,64位Windows Server2012及以上
 - 加速版:64位 Windows 10,64位Windows Server 2019及以上
- 环境依赖 (必须安装以下版本)
 - .NET Framework 4.5
 - Visual C++ Redistributable Packages for Visual Studio 2013
 - Visual C++ Redistributable Packages for Visual Studio 2015-2019
- 协议
 - HTTP
- 更详细的环境说明可参考SDK内的README.md

*intel 官方合作,拥有更好的适配与性能表现

Release Notes | 时间 | 版本 | 说明 | | ------ | ----- | 2023-06-29 | 1.8.2 | 优化模型算法 || 2023-05-17 | 1.8.1 | 预测引擎升级 , 修 复部分网络内存泄露问题 || 2023-03-16 | 1.8.0 | 预测引擎升级 || 2022-12-29 | 1.7.2 | 预测引擎升级 || 2022-10-27 | 1.7.1 | GPU底层引擎升 级 , 下线基础版CUDA10.0及以下版本支持 || 2022-09-15 | 1.7.0 | 新增支持表格预测 || 2022-07-28 | 1.6.0 | 优化模型算法 || 2022-05-27 | 1.5.1 | 新增支持BML Cloud小目标检测模型 || 2022-05-18 | 1.5.0 | 修复各别机器下程序崩溃的问题 || 2022-04-25 | 1.4.1 | EasyDL, BML升级 支持paddle2模型 || 2022-03-25 | 1.4.0 | 优化模型算法 || 2021-12-22 | 1.3.5 | CPU基础版推理引擎优化升级 ; demo程序优化环境依赖检测 || 2021-10-20 | 1.3.4 | CPU加速版推理引擎优化升级 || 2021-08-19 | 1.3.2 | 新增DEMO二进制文件 || 2021-06-29 | 1.3.1 | 预测引擎升级 || 2021-05-13 | 1.3.0 | 模型发布新增多种加速方案选择 ; 目标追踪支持x86平台的GPU及加速版 ; 展示已发布模型性能评估报告 || 2021-04-08 | 1.2.3 | 支持BML平台模型仓库本地上传模型 || 2021-03-09 | 1.2.2 | CPU加速版支持int8量化模型 || 2021-01-27 | 1.2.1 | 新增模型支持 ; 性能优 化 ; 问题修复 || 2020.12.18 | 1.2.0 | 推理引擎升级 || 2020-11-26 | 1.1.20 | 新增一些模型的加速版支持 || 2020.10.29 | 1.1.20 | 修复已知问 题 || 2020.05.15 | 1.1.16 | 优化性能 , 修复已知问题 || 2020-09-17 | 1.1.19 | 支持更多模型 || 2020.04.16 | 1.1.18 | 支持专业版更多模型 || 2020.06.23 | 1.1.17 | 支持专业版更多模型 || 2020.02.23 | 1.1.13 | 支持零阶段模型 || 2020.04.16 | 1.1.12 | 预测默认使用推荐阀值 || 2019.02.26 | 1.1.11 | CPU加速版支持物体检测高精度 || 2019.12.04 | 1.1.10 | 支持图像分割 || 2019.02.11 | 1.1.9 | 支持 EasyDL 专业版 || 2019.08.29 | 1.1.8 | CPU 加速版支持 /| 2019.07.19 | 1.1.7 | 提供模型更新工具 || 2019.05.16 | 1.1.3 | NVIDIA GPU 支持 || 2019.03.15 | 1.1.0 | 架构与功能完善 || 2019.02.28 | 1.0.6 | 引擎功能完善 || 2019.02.13 | 1.0.5 | paddlepaddle 支持 || 2018.11.30 | 1.0.0 | 第一版 !|

快速开始

1. 安装依赖

必须安装:

安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642

Visual C++ Redistributable Packages for Visual Studio 2013

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

Visual C++ Redistributable Packages for Visual Studio 2015-2019

https://docs.microsoft.com/en-us/cpp/windows/latest-supported-vc-redist?view=msvc-160

表格预测依赖

brotlipy==0.7.0 certifi==2020.6.20 joblib==1.0.1 kaggle==1.5.12 Pillow py4j pycosat python-dateutil python-slugify ruamel_yaml textunidecode threadpoolctl flask pandas==1.0.5 scikit-learn==0.23.2 lightgbm==2.2.3 catboost==0.24.1 xgboost==1.2.0 numpy==1.19.5 scipy==1.5.2

psutil==5.7.2 pypmml==0.9.7 torch==1.8.0 jieba==0.42.1 pyod==0.8.5 pyarrow==6.0.0 scikit-optimize==0.9.0 pyspark==3.3.0

注意事项

1. 安装目录不能包含中文

2. Windows Server 请自行开启,选择"我的电脑"——"属性"——"管理"——"添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启即 可。

2. 运行离线SDK

解压下载好的SDK,打开EasyEdge.exe,输入Serial Num

Ľ

点击"启动服务",等待数秒即可启动成功,本地服务

图片加载失败

默认运行在

http://127.0.0.1:24401/

其他任何语言只需通过HTTP调用即可。

🔄 data	2020/4/15 23:51	文件	+夹	
AI 百度EasyEdge		_		×
Model	id : Easy	18		
Serial Nur				
Host	127.0.0.1			
Port	24401			
服务状态	服务未启动			
	是否开机自动,	自动 🗌		
	自动服务			
	/H4/JJJK 7)			

使用说明

调用说明

Python 使用示例代码如下

json结构说明

考虑到表格字段内容和长度的不固定性,我们建议您参考SDK内置demo页面展示的json示例,这将帮助您理解如何灵活处理各种不同的字段。

import requests

C# 使用示例代码如下



StreamReader sr = new StreamReader(response.GetResponseStream()
Console.WriteLine(sr.ReadToEnd());
sr.Close();
response.Close();

C++ 使用示例代码如下,需要安装curl

```
**include <sys/stat.h>**
**include <curl/curl.h>**
**include <iostream>**
**include <string>**
**define S_ISREG(m) (((m) & 0170000) == (0100000))**
**define S_{ISDIR}(m) (((m) \& 0170000) == (0040000))**
size_t write_callback(void *ptr, size_t size, size_t num, void *data) {
  std::string *str = dynamic_cast<std::string *>((std::string *)data);
  str->append((char *)ptr, size*num);
  return size*num:
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./1.json";
  FILE *fp = NULL;
  std::string response;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\"\n", post_data_filename);
     return -1;
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
    curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
    curl_easy_setopt(curl, CURLOPT_POST, 1L);
    curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
    curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
 curl_easy_setopt(curl, CURLOPT_WRITEFUNCTION, write_callback);
    curl_easy_setopt(curl, CURLOPT_WRITEDATA, &response);
    res = curl_easy_perform(curl);
     if (res != CURLE_OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     std::cout << response << std::endl; // response即为返回的json数据
     curl_easy_cleanup(curl);
  curl_global_cleanup();
  fclose(fp);
  return 0:
```

结果 获取的结果存储在response字符串中。请求参数

字段	类型	取值	说明
threshold	float	0~1	置信度阈值

HTTP POST Body直接发送图片二进制。

返回参数

预测结果数组

集成指南

基于HTTP集成

通过EasyEdge.exe启动服务后,参照上面的调用说明,通过HTTP请求集成到自己的服务中

Baidu 百度智能云文档

基于c++ dll集成

集成前提

解压开的SDK包中包含src、lib、dll、include四个目录才支持基于c++ dll集成

集成方法

参考src目录中的CMakeLists.txt进行集成

基于c# dll集成

集成前提

解压开的SDK包中包含src\demo_serving_csharp、dll两个目录才支持基于c# dll集成

集成方法

参考src\demo_serving_csharp目录中的CMakeLists.txt进行集成

FAQ

1. 服务启动失败,怎么处理?

根据SDK内的README.md检查依赖是否都已正确安装

请确保相关依赖都安装正确,版本必须如下: *.NET Framework 4.5* Visual C++ Redistributable Packages for Visual Studio 2013 * Visual C++ Redistributable Packages for Visual Studio 2015

如使用的是CPU加速版,需额外确保Openvino安装正确,版本为2020.3.1LTS版 如使用Windows Server,需确保开启桌面体验

2. 服务调用时返回为空,怎么处理?调用输入的图片必须是RGB格式,请确认是否有alpha通道。

3. 多个模型怎么同时使用? SDK设置运行不同的端口,点击运行即可。

4. JAVA、C#等其他语言怎么调用SDK?

参考 https://ai.baidu.com/forum/topic/show/943765

5. 启动失败,缺失DLL? 打开EasyEdge.log,查看日志错误,根据提示处理 缺失DLL,请使用 https://www.dependencywalker.com/ 查看相应模 块依赖DLL缺失哪些,请自行下载安装

6. 启动失败,报错NotDecrypted? Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

7. 启动失败,报错 SerialNum无效

日志显示failed to get/check device id(xxx)或者 Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 C:\Users\\${用户名}\.baidu \easyedge 目录,再重新激活。

8. 勾选"开机自动启动"后,程序闪退

一般是写注册表失败。

可以确认下HKEY_CURRENT_USER下Software \Microsoft \Windows \CurrentVersion \Run能否写入(如果不能写入,可能被杀毒软件等工具管制)。也可以尝试基于bin目录下的easyedge_serving.exe命令行形式的二进制,自行配置开机自启动。

其他问题 如果无法解决,可到论坛发帖: https://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

故障处理

训练相关问题

数据处理失败或者状态异常怎么办?

- 如果是首次导入数据失败,请检查数据文件编码格式是否为UTF-8或GBK
- 如果是非首次导入数据失败,请检查新导入文件的首行是否与首次导入时一致,即都为列名或都为数据
- 如果自查没有发现问题,请在百度云控制台内提交工单反馈

模型训练失败怎么办?

• 如果遇到模型训练失败的情况,请在百度云控制台内提交工单反馈

模型效果相关问题

实际调用服务时模型效果变差?

- 在实际业务场景中,数据的分布可能会发生变化,即使用历史数据训练的模型不能正确对新的数据进行预测,此时需要收集新的数据并重新进行模型训练
- 如果训练模型使用的数据量太小,导致训练数据不能正确反映全部的数据特征,也会导致上线模型的效果变差。
- **如果线上请求数据与训练数据未发生显著的分布变化,请在百度云控制台内提交工单反馈

模型上线相关问题

希望加急上线怎么处理?

• 请在百度云控制台内提交工单反馈

每个账号可以上线几个模型?是否可以删除已上线的模型?

• 每个账号最多申请发布十个模型,已上线模型无法删除

申请发布模型审核不通过都是什么原因?

 可能原因有,1、经过电话沟通当前模型存在一些问题或者不再使用,如训练数据异常、数据量不够、不想再继续使用等原因,沟通达成一致 拒绝。2、电话未接通且模型效果较差,会直接拒绝。如果需要申诉,请在百度云控制台内提交工单反馈

心 训练任务失败错误排查

表格预测一般是由于任务类型选择错误引起的。当训练任务失败时,请您检查选择的任务类型与目标列是否匹配,以下为各个任务类型对应的目 标:

算法类 型	目标列
二分类	目标列是离散值,且只有两种可能的取值。如在精准营销场景中预测一个用户是否为潜在购买用户,其目标列仅存在"True"和"False"两 种取值,解决该问题的模型属于二分类模型。
多分类	目标列是离散值,并具有有限的可能取值。如在用户分类场景中,根据用户的历史消费数据,将用户划分到不同消费偏好的类别中, 解决该问题的模型属于多分类模型。 不建议将重复率很小的值或时间列作为目标列。
回归	目标列是连续的实数范围,或者属于某一段连续的实数区间。如在销量预测场景中,销量值可能是某个取值范围内的任意值,解决该 问题的模型属于回归模型。 目标列不能包含大量无法转成数值的异常值

若您还是无法判断算法类型,请选择默认的自动。

时序预测

时序预测介绍

简介

Hi,您好,欢迎使用百度EasyDL定制化训练和服务平台。

定制时序预测模型,旨在帮助用户通过机器学习技术从历史数据中发现潜在规律,从而对未来的变化趋势进行预测。本文介绍**时序预测**模型:相 较于表格数据预测使用的分类或回归模型,时序预测模型使用的训练数据中必须包含有效时序的特征,一般时序具有固定的频率,且在连续时间 范围内的每个时间点上都有一个值。 以下是关于时序预测模型的技术文档。

∞ 应用场景

- 销量预测:基于历史销量数据预测当期的销售量,进而帮助厂商制定更合理的生产或备货计划,从而提高利润
- 交通流量预测:基于给定路段的历史交通量数据推测未来的交通量,为交通运输规划与研究提供决策依据
- 价格预测:从历史数据中发现商品的变化规律以及影响价格的因素,从而为未来的商业行为提供支持

心 使用流程

训练模型的基本流程如下图所示,全程可视化简易操作。在数据已经准备好的情况下,最快几分钟即可获得定制模型。
下面将详细介绍每一步的操作方式和注意事项。如果文档没有解决您的问题,请在百度云控制台内提交工单反馈。



数据准备

心 时序数据集介绍

时序数据介绍

时序数据包含有时序特征,常规时序数据是具有一定频率的并且在连续时间范围内的每个采样点上都有一个值。

一个时序数据集可以包含一个或多个时间序列,如下数据集包含一个品牌在A、B两个地区的每日销售数据:

datetime,area,sales_quantity 9/3/2018,A,2000 9/3/2018,B,600 9/4/2018,A,2300 9/4/2018,B,550 9/5/2018,A,2100 9/5/2018,B,650 9/6/2018,B,700 9/7/2018,A,2450 9/7/2018,B,650

上述数据内容可以分为A地区销量时序:

datetime,area,sales_quantity 9/3/2018,A,2000 9/4/2018,A,2300 9/5/2018,A,2100 9/6/2018,A,2400 9/7/2018,A,2450

B地区销量时序:

datetime,area,sales_quantity 9/3/2018,B,600 9/4/2018,B,550 9/5/2018,B,650 9/6/2018,B,700 9/7/2018,B,650

一个时序数据集除具有时间特征以及一个标量特征外,还可以具有其它影响标量取值的特征,如在销量数据场景下,当天的气温、是否节假日等 因素也会影响销售数据: datetime,is_holiday,sales_quantity 9/3/2018,Y,600 9/4/2018,N,550 9/5/2018,N,650 9/6/2018,Y,700 9/7/2018,N,650

数据要求

- 目前仅支持CSV格式的数据文件
- 一次仅能上传一个文件,可以是一个CSV文件或由多个CSV文件压缩成的zip包
- 单个上传文件大小不能超过5GB
- 一个数据集包含的总文件大小不能超过20GB

数据文件内容要求

- 当数据文件包含列名时,列名称可以包含字母、数字和下划线(_),但不能以下划线开头。
- 文件内容以换行符(即字符"\n",或称为LF)分隔各行,行内容以英文逗号(即字符",")分隔各列
- 必须包含要预测的值即目标列,且目标列的数据类型会决定模型的类型。
- 文件中文本列取值长度不能超过4096个字符。
- 文件必须至少包含两列,并至少包含一个日期列,总列数不得超过1000列。
- 数据集的总行数不能超过1000万行。
- zip包中的多个CSV文件必须使用相同的编码格式,都包含列名或都不包含列名;且列的顺序必须保持一致
- 在扩充数据集时,新导入数据文件的首行与数据集的列名相同时,将被视为列名,否则将被视作数据。

の 创建时序数据集

在EasyDL经典版中,您可以在"数据总览"页面,完成数据集创建、数据导入等操作,为模型构建准备好数据。

创建数据集

数据集需要先定义,然后再导入数据。

- 1. 单击"数据总览",进入数据集列表页面。
- 2. 单击"创建数据集",进入数据集创建页面。
- 输入数据集名称,单击"完成"结束创建。
 完成创建后,可以在数据集列表中查看新建的数据集。系统默认生成V1版本,当前数据内容为空,可以通过导入的方式向其中添加数据。

导入数据

通过导入的方式可以向数据集中添加或追加数据。

- 1. 单击"数据总览",进入数据集列表页面。
- 2. 单击待导入数据集所在行的"导入"按钮,进入数据集导入页面。
- 3. 导入数据文件。

导入的数据文件可以是CSV文件或由CSV文件组成的压缩包文件。 如果导入的是CSV文件,支持数据预览,如果是压缩包格式,则不支持预览。

- 4. 根据数据文件的实际情况进行列名设置。
 - 设置首行为列名:将导入的数据文件中的首行作为列名。
 - 设置首行非列名:此时系统会自动生成列名,而将首行作为数据。
- 5. 单击"确认并返回"完成导入操作。

模型训练

心时序预测模型介绍

时序预测模型介绍

时序预测模型是基于包含时间特征的结构化数据进行建模,系统会基于用户上传的数据使用预置算法进行模型构建与训练。当完成模型训练后,系统不仅提供了常见的评估指标而且会生成可视化的预测序列效果图,帮助用户检查模型的好坏。对于达到业务要求的时序预测模型,可以部署为在线服务,通过远程调用的方式对新的时间数据进行预测。

应用场景

- 销量预测:基于历史销量数据预测当期的销售量,进而帮助厂商制定更合理的生产或备货计划,从而提高利润
- 交通流量预测:基于给定路段的历史交通量数据推测未来的交通量,为交通运输规划与研究提供决策依据
- 价格预测:从历史数据中发现商品的变化规律以及影响价格的因素,从而为未来的商业行为提供支持

心 创建模型

在EasyDL中,您可以在"模型中心"进行模型的创建。在EasyDL中模型可以包括多个版本,每次训练会生成一个版本。各个版本的模型之间相互独立,可以分别进行版本发布等操作。

创建模型

模型需要先创建,然后才能进行训练。

- 1. 单击"创建模型",进入模型创建页面。
- 2. 填写模型创建信息,如下图所示。

	时序预测模型 提文工单 定定制基于时序数据的模型,根据历史数据对未来的变化趋势进行预测,该类模型可应用在销量预测、交通流量预测等场景。此 类模型训练速度较快,具有TODO个特征列的TODO条样本的来格数据,一般可在TODO分钟内训练完毕。
模型中心模型	型列表 > 创建模型
我的模型	
创建模型	*************************************
训练模型	
校验模型	Deale-Line 244 17.
发布模型 FasyData数据服务	
数据总院	
	即時地址 * z*******@baidu.com
	联系方式 * 135****919 ⑦
	功能推进。
	0/500
	世一 7

- 模型名称:指定模型的名称
- 模型归属:公司或个人,并输入相关名称
- 所属行业:请根据实际情况进行选择
- 应用场景:请根据实际情况进行选择
- 邮箱地址:请根据实际情况进行设置
- 联系方式:请根据实际情况进行设置

• 功能描述:用于记录模型创建的背景、用途等方面的信息。

3. 单击"下一步"完成模型创建。

完成创建后,可以在模型列表中查看新建的模型。新建的模型不包含任何版本的模型,在训练后会生成新的模型版本。

の训练模型

EasyDL提供的的时序预测算法不仅性能高、可扩展,还针对速度、规模和准确性进行了优化,可以在大规模数据集上进行训练。

训练模型

在准备好数据集并创建模型后,可以创建训练任务。

- 1. 单击"训练模型",进入模型训练页面。
- 2. 填写模型训练信息,如下图所示。

如果选择的数据集中仅包含时间和数值两列,则训练参数配置如下所示:

训练模型

选择模型	mymodel01	~
选择数据集	timeseries V1	~
选择时间列 🕜	timestamp	\sim
选择时间间隔 🕜	5 今钟	~
选择目标列 🕜	value	∨ 预览序列
滑动窗口大小 🕜	30 🔨	
预测长度 🕐	10	
部署方式	☑ 公有云API	
	开始训练	

- 选择模型:选择要训练的模型
- 选择数据集:选择训练模型使用的数据集
- 选择目标列:从数据集中选择表示时序的日期时间列
- 选择时间间隔:请根据序列中相邻两个样本点的时间间隔进行设置
- 选择目标列:需要被预测的随时间而变化的列
- 滑动窗口大小:表示使用多久的历史数据生成预测数据
- 预测长度:要预测的序列的长度,该长度因小于滑动窗口大小

如果选择的数据集中除时间和数值外还有其它列,则训练参数配置如下所示:
训练模型

选择模型	销量预测	\sim	
选择数据集	dominicks_OJ V1	\sim	
选择时间列⑦	WeekStarting	\sim	
选择分组字段 ⑦	Store x Brand x		
选择时间间隔 ⑦	1	~	
选择目标列 ⑦	Quantity	~	预览序列
滑动窗口大小 ⑦	30		
预测长度 ⑦	10		
部署方式	公 公有云API		
	开始训练		

相比于仅包含一个序列的训练任务,其需要配置分组字段:

- 分组字段用于将数据集划分为具有相同时间戳的多个时序
- 最迟支持设置两个分组字段
- 分组字段只能为类别列

3. 单击"开始训练"启动训练任务。

启动训练任务后,系统会在模型下的列表中创建一个新的模型版本,新建的版本处于"训练中"的状态,当处于"训练完成"状态时表示模型已完成训练。

4. 平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

优惠政策:

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有时序预测 操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。

の 查看模型

训练任务结束后,可以查看模型的各项指标,以确定模型是否满足要求。不同类型的模型包含了不同的指标,用户可以根据实际的业务场景查看 关键指标是否满足要求。

查看模型

对于"训练完成"的模型,可以查看其评估结果。

- 1. 单击"我的模型",进入模型列表页面。
- 2. 单击待查看模型的"历史版本",进入模型版本列表页面。
- 3. 点击待查看模型版本所在行的"完整评估结果",系统展示评估结果页面。

预测结果包括整体评估以及详细评估:

- 整体评估:提供了包括MAE、MSE、MAPE、R2 Score四项数值指标
- 详细评估:则给出了预测结果的可视化表示,其中预测序列包含了预测部分的对比以及差值情况;完整序列则不仅仅包含了生成的预测 序列部分,也包含了用于生成预测序列的已知序列部分。

我的模型 > mymodel01模型评估报告	
部署方式 公有云API (目前仅支持) 💙 版本 V1 🗸 🗸	
训练时长 8分钟	
整体评估 特征 1列 目标列 value 算法类型 时序预则	
mymodel01 V1整体效果欠佳。 如何优化效果?	
名称	数直
MAE (Mean Absolute Error) 平均绝对误差 ⑦	6.530
MSE (Mean Squared Error) 均方误差 ⑦	72.380
MAPE (Mean Absolute Percentage Error) 平均绝对百分比误差 ⑦	12.36%
R2 Score (决定系数) 回归得分函数 ⑦	0.476

详细评估







完整序列



心 校验模型

在完成训练后,发布模型前,可以先进行模型校验,以确保模型在实际环境中能获得预期的性能。

在训练任务成功完成后,即可使用实际数据进行校验。

- 1. 单击"校验模型",进入模型校验页面。
- 2. 选择要校验的模型及其版本。
- 3. 单击"启动模型校验服务"。
 - 对于单序列模型,系统会自动生成校验数据
 - 对于多序列模型,可以通过上传CSV文件来填充测试数据,但每次测试时只能包含一个序列的数据
- 4. 单击"预测",可以在右侧结果面板中查看预测结果。

校验示例如下所示:

模型列表

选择模型 mymodel01 V	部署方式 公有云API (目前仅支持) > 选择版	本 V1 V
当前模型MAE(平均绝对误差) 6.530 评估报告		识别结果 如何优化效果?
预测数据	土 上传 csv 文件 📋 复制son格式 📃 预测	预测结果
timestamp	value	1 {
日期	数值	2 "tim: [3 "2015-09-09 15:33:00",
1 2015-09-08 11:39:00	73	4 "2015-09-09 15:38:00", 5 "2015-09-09 15:43:00",
2 2015-09-08 11:44:00	62	6 "2015-09-09 15:48:00", 7 "2015-09-09 15:53:00",
3 2015-09-08 11:59:00	66	8 "2015-09-09 15:58:00", 9 "2015-09-09 16:03:00",
4 2015-09-08 12:19:00	69	10 "2015-09-09 16:08:00", 11 "2015-09-09 16:13:00",
5 2015-09-08 12:24:00	65	12 "2015-09-09 16:18:00", 13]
6 2015-09-08 12:27:00	76	14 "val: [15 "CC 057045020410"
预测		申请上线

模型发布

心 模型发布整体说明

训练完成后,可将模型部署在公有云服务器上,通过API进行调用。

公有云API

- 训练完成的模型存储在云端,可通过独立Rest API调用模型,实现AI能力与业务系统或硬件设备整合
- 具有完善的鉴权、流控等安全机制,GPU集群稳定承载高并发请求

相关费用

将模型发布为API后,将获得1000次免费调用次数,如需更多调用量,请在百度云控制台内提交工单反馈。

心 公有云部署

^の如何发布时序预测API

训练完毕后可以在左侧导航栏中找到【发布模型】,依次进行以下操作即可发布公有云API:

- 选择模型
- 选择部署方式「公有云部署」
- 选择版本
- 自定义服务名称、接口地址后缀
- 申请发布

<mark>单</mark>反馈。

发布模型界面示意:

发布模型					
选择模型	mymodel01 V	标/ 注接口规 : 标准接口请求参	范参考 _{考说明:}		
HHE 73 24		字段名称	必须	类型	说明
选择版本 服务名称*		data	毘	object	待预测数据,待预测数据是由各 个特征序列组成的对象。若训练 时设置了分组字段,一次API请求 中的数据必须同一个分组的序 列。
接口地址 *	https://aip.baidubce.com/rpc/2.0/ai_custom/v1/time_seri	标准接口响应字	段说明:		
其他要求	若接口无法满足您的需求,请描述希望解决的问题,	字段名称	必须	类型	说明
	500汉字以内	log_id	是	number	唯一的log id,用于问题定位
	0/500	error_code	否	number	错误码, 当请求错误时返回
	提交申请	error_msg	柘	string	错误描述信息,当请求错误时返 回
		result	否	object	预测结果

接口赋权

在正式使用之前,还需要做的一项工作为接口赋权,需要登录EasyDL控制台中创建一个应用,获得由一串数字组成的appid,然后就可以参考接口文档正式使用了

Ŷ	全局 ~					Q	合作伙伴	工単	ï,¢	帮助文档	企业资源	财务	W ~
	✓ 返回EasyDL总览	产品服务 /	EasyDL定制训练平台 - 应用列	表									
>	通用版	应用列	则表										
88	公有云服务 ^	+ 创	建应用										
ጭ	 应用列表 		应用名称	AppID	API Key	Secr	ret Key		创建时间	9	操作		
ß	 权限管理 田島(G)) 	1	app001	19762766	fh8fPHX3EQ3tpuj8lzPGI mZm	*****	* 显示		2020-05	-07 17:05:16	报表	管理 删除	
e m	 用型就订 收放起主 				zCYKCnAxyUuXZmnVY5								
949 949	- 西田成城	2	电影打标签	19732557	6j4n3K	*****	** 显示		2020-05	-05 19:08:47	报表	管理制除	
50 E3	 售卖服务 	3	dogcat分类	19692977	Mmojm5Ks7GE9dkgGW5 3dDPoA	*****	** 显示		2020-05	-02 09:59:47	报表	管理 删除	
Þ	EasyEdge本地部署 🗸	4	文本分类	19678854	VndCBIKvmXj9htnZ5EqA zsUH	*****	* 显示		2020-04	-30 19:16:53	报表	管理制除	
												< 1	>
ĥ													

同时支持在「公有云服务管理」-「权限管理」中为第三方用户配置权限

示意图如下:

Ŷ	全局 ~			Q	合作伙伴	工単	消息	帮助文档	企业资源	财务	W ~
88	✔ 返回EasyDL总览	产品服务 / Eas	yDL定制训练平台 - 权限管理								
>	通用版	权限管理									
88	公有云服务 ^	服务权	艰管理								
ф	■ 应用列表	支持为您回	始者他人的AppID配置可用服务权限。								
œ	• 权限管理	序号	服务名称	相关操	作						
e	▪ 用量统计	1	iris	配置权	限						
(ii)	• 监控报表	2	wht0001	配置权	限						
Ħ	▪ 技术文档	3	文本分类	配置权	隈						
E 1	• 售卖服务	4	102-objectcategories	配置权	限						
Þ	EasyEdge本地部署 🗸										
											ቆ
=											

Ŷ	全局 ~				Q	合作伙伴	工単	消息	帮助文档	企业资源	财务	W ~
88	✓ 返回EasyDL总览	产品服务 / EasyDL定制	前训练平台 - 权限管理 /	配置权限								
>	通用版	配置权限										
	公有云服务 ^	服务名称:	wht0001									
ቍ	■ 应用列表											
B	 权限管理 	可用APP ID:	19678854	文本分类 ×								
A	• 用悬统计		19692977	dogcat分类 ×								
	1000000		19732557	电影打标签 ×								
Ņ	■ 监控报表		19762766	app001 ×								
ж	・ 技术文档		请添加 +									
E 1	 售卖服务 											
₽	EasyEdge本地部署 🗸		保存取消									
												<u>_</u>
==												

ゐ API调用文档

本文档主要说明定制化模型发布后获得的API如何使用,如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

接口描述

基于自定义训练出的表格数据预测模型,实现表格数据预测。模型训练完毕后发布可获得定制化表格数据预测API

接口鉴权

1、在EasyDL控制台创建应用

♥ 全局 ∨									Q	≣	₽	?	8	۲	Q
EasyDL定制训练平台	产品服务 / EasyD	L定制训练平台 - 店	2用列表 / 创建应用												
概览	创建新应用	I													
应用列表	• 应田之称•														
监控报表	177113-12101	test													
技术文档	•应用类型:	游戏娱乐					×	~							
云服务权限管理	•接口选择:	您可以为应用勾)选多个接口权限,使此应)	可以	请求已勾选的接口	1服务									
离线SDK管理		EasyDL	 ✓ 百美椅子训练 ✓ 识别眼睛和嘴巴 	\sim	椅子顔色识别 声音分类	\checkmark	aa 商品检测								
		➡ 百度语音													
		➡ 文字识别													
		➡ 人脸识别													
		• 自然语言处	王璧												
		➡ 内容审核													
		🕂 UNIT 📘													
		+ 知识图谱													
		➡ 图像识别													
		➡ 智能呼叫中。	ŵ.												
		+ 图像搜索													
		+ 人体分析													
	•应用描述:	简单描述一下 点,请控制在	您使用人工智能服务的应序 500字以内	目场景	, 如开发一款美颜	阴相机 ,	需要检测人脸关键	建							

2、应用详情页获取AK SK

♥ 全局 ~					Q	⊞	P	0	몲	۲	Q~
EasyDL定制训练平台	产品服务 / EasyDL定制训练平台 - 应用列表 / 应用	详情									
概览	应用详情										
应用列表	编辑 查看文档										
监控报表	应用名称	AppID	_	API Key			Secr	et Key			
技术文档	test2	14293933		bZsfqO2fUdHTDflxqF11m	nkTq			** 显示			
因服务权限管理 离线SDK管理	EasyDL定制化训练服务:										

请求说明

请求示例

请求URL:请首先进行自定义模型训练,完成训练后可在服务列表中查看并获取url。

URL参数:

参数	值 值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/json

注意:如果出现336001的错误码很可能是因为请求方式错误,与其他图像识别服务不同的是定制化图像识别服务以json方式请求。

Body请求接口:

```
{
"include_req": false,
"data": <待预测数组>
}
```

具体示例如下:

```
{
   "data": {
    "datetime":
    [ "2015-09-09 15:33:00", "2015-09-09 15:38:00", "2015-09-09 15:43:00"],
    "sales_quantity":
    [ "10", "15" , "20"]
   }
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必 选	类型	可选值 范围	说明
data	是	object	-	待预测数据,待预测数据是由各个特征序列组成的对象。若训练时设置了分组字段,一次API请求中的数据 必须同一个分组的序列。

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码,当请求错误时返回
error_msg	否	string	错误描述信息,当请求错误时返回
results	否	object	预测结果数组

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
"error_code": 110,
```

```
"error_msg": "Access token invalid or no longer valid"
```

```
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请在百度云控制台内提交工单反馈
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请在百度云控制台内提交工单反馈
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请在百度云控制台内提交工单反馈
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在百度云控制台内提交工单反馈
336001	Invalid Argument	入参格式有误,比如缺少必要参数、data格式错误等等,可检查下请求数据格式是否有误。有疑问请在百度 云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或者代码格式有误。有疑问请在百度云控制台内提交工单反馈
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请在百度云控制台内提 交工单反馈

故障处理

心训练任务失败错误排查

时序预测任务失败,一般主要是由于时间列格式错误或配置错误导致的,您可以参照如下内容检测您的数据或配置是否正确:

时间列要求

在创建时序预测任务时,所使用的数据集的时间列应为Date类型并以**yyyy-MM-dd**开头,否则会造成训练任务失败. 如 "yyyy-MM-dd'T'HH:mm:ssX" "yyyy-MM-dd HH:mm:ss" "yyyy-MM-dd" 等等

时间间隔

选择时间列与目标列后,系统会根据数据的前100行去推断时间间隔,但是难免会推断错误。

预处理阶段会根据选择的时间间隔对数据进行重新采样,若选择了过大的时间间隔,降采样会导致数据过短。反之,过小的时间间隔会导致数据

分布异常,合理的选择时间间隔是保证预测质量的重要环节之一。

时间序列长度

时序预测通常会采用历史的一段时间的数据作为特征,因此数据长度必须满足一定要求才能保证有足够的训练与评估样本。

当时序数据为分组数据时,每个分组会作为一个单独的时间序列,因此需要保证每个分组的时间序列足够长。

另外选取了不适当的时间间隔,降采样后也会导致数据量不够的情况。

分组字段

为了确保每个时间列有充分的数据,限制了平均每个分组至少包含20条数据,若数据中存在大量超短时序数据,请您提前做好筛选。

窗口长度与预测长度

时序模型的输入为窗口长度中的各个特征,输出为预测长度的目标列。因此当数据中的时间序列较短的时候,请您**合理地降低窗口长度和预测长** 度,有助于模型有充足的样本进行预测。

EasyDL OCR使用说明

EasyDL OCR介绍

心 功能介绍

EasyDL文字识别,可定制识别图片中的文字信息,结构化输出关键字段内容,极大提升OCR模型训练效率,满足个性化卡证票据识别需求

• 数据标注

创建数据集并上传真实图片,定义数据识别字段作为标注标签,在图片中框选对应的 Key/Value 内容区域,自动识别框选区域内容完成转写, 标注人员对识别结果进行查验纠正即可完成标注

• 数据生成

基于已标注数据,将图中已框选 Value 区内容进行抹除,选择对应的字体、字号、颜色,并根据该字段的内容选择相匹配的语料库,即可完成 虚拟数据生成底板的创建,并基于此底板生成任意张版式相同内容不同的虚拟数据,快速扩充数据集规模,结合真实数据一同用作模型训练集

• 模型训练与管理

支持根据使用场景需求创建多个的识别模型,选择包含已标注数据及虚拟数据的数据集进行训练,即可自动排队完成训练,同时输出预测准确率供参考;也可扩充数据集对现有模型进行迭代训练,产出新版本

• 服务部署

对训练完成的模型可上传真实数据进行模型校验,效果满意后即可一键发布上线,自动分配机器资源完成部署,并生成标准API接口供业务调用

心 特色优势

⁰零门槛操作

提供一站式流程化训练,并预置最佳预训练模型及训练参数,无需算法基础、无需关注算法细节即可完成模型训练

心高精度效果

基于百度丰富的商用模型实训经验,预置最佳实践产出的预训练模型,并基于百度自研的 EnDet 实体检测模型进行训练,模型平均准确率可达 90% 以上

心 低成本数据

提供可视化数据管理平台,对上传图片进行智能预标注,仅需核对修改即可完成标注,并可基于一张标注图片批量生成虚拟数据,快速扩充训练 集,启动模型训练

心 超灵活部署

支持多种部署方式,公有云服务可一键部署,即刻生成 Restful API,毫秒级调用响应,高并发承载;同时,完整平台支持私有化部署,可用于搭 建企业内部 AI 中台;也可支持产出模型容器化打包进行本地部署,快速完成项目交付

心 应用场景

• 证照电子化审批:对政府、金融、企业等审批流程中涉及到的各种证照,如食品/药品经营许可证、特种设备审批证等,进行定制训练,快速

提取关键信息完成线上审批,实现 7*24 小时无间断服务

- 财税报销电子化:对不同金融或税务机构发型的各类财务发票、银行单据进行定制训练,快速实现财税凭证的录入,大幅度节约凭证邮寄、录入成本,实现线上电子化报税报销
- 保险智能理赔:对不同版式的保单或不同地区、不同医疗系统开具的医疗票据进行定制训练,实现保险理赔相关材料的快速录入,降低人力成本,提升保险理赔的业务安全性及快捷性

API文档

心 请求说明

请求示例

HTTP 方法:POST

请求URL: https://aip.baidubce.com/rest/2.0/ai_custom/v1/ocr

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数	值
Content-Type	application/x-www-form-urlencoded

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必 选	类型	可选值 范围	说明
image	和 url 二选一	string	-	图像数据,base64编码后进行urlencode,要求base64编码和urlencode后大小不超过4M,最短边至少 15px,最长边最大4096px,支持jpg/jpeg/png/bmp格式
url	和 image 二选一	string	-	图片完整URL,URL长度不超过1024字节,URL对应的图片base64编码后大小不超过4M,最短边至少 15px,最长边最大4096px,支持jpg/jpeg/png/bmp格式,当image字段存在时url字段失效,不支持https的图 片链接。当image字段存在时 url 字段失效
modell d	是	string	-	模型 ID,自训练产出模型的唯一标示,可用于调用指定的已发布模型进行结构化识别,可在「我的模型」页查看并复制使用
detect_ directio n	是	string	true/fal se	是否开启图像方向矫正功能,可选值有: - true:开启图像方向矫正功能,可自动矫正不同旋转角度的图片进行识别; - false:关闭图像矫正功能,如要识别的内容均为正向图片,建议可关闭此功能避免误矫正。

の 返回说明

返回参数

字段	类型	说明
log_id	int	调用请求的唯一日志id,如需技术支持进行问题排查请反馈此id以快速进行问题定位
error_code	int	0代表成功,如果有错误码返回可以参考错误码列表排查问题
error_msg	string	如果error_code具体的失败信息,可以参考下方错误码列表排查问题
result	Object	识别返回的结果,每一个key代表识别字段名称,对应的value为该字段的识别结果
+ "key"	Array	识别内容,"key"为数据标注时创建的字段名称,将不同版式的内容进行归一化输出
++ probability	Object	字段的置信度,包括平均和最小置信度
+++ average	int	字段的平均置信度
+++ min	int	字段的最小置信度
++ location	Object	字段在原图上对应的矩形框位置
+++ top	int	字段文本框左上角点的上边距
+++ left	int	字段文本框左上角点的左边距
+++ width	int	字段文本框的宽度
+++ height	int	字段文本框的高度
++ word	string	字段识别结果

返回示例

```
{
  "log_id":"161648117706127",
  "result":{
    "收款卡号":[
      {
         "probability":{
           "average":0.9972677230834961,
           "min":0.6964160077398716
         },
         "location":{
           "top":404,
           "left":451,
           "width":303,
           "height":61
        },
         "word":"119086830765501"
      }
    ],
     "日期":[
      {
         "probability":{
           "average":0,
           "min":0
         "location":{
           "top":-1,
           "left":-1,
           "width":0,
           "height":0
         "word":"20180330"
      }
  }
}
```

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

• error_code : 错误码。

• error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请在控制台提交工单联系技术支持团队
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请在控制台提交工单联系技术支持团队
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
14	IAM Certification failed	IAM鉴权失败,建议用户参照文档自查生成sign的方式是否正确,或换用控制台中ak sk的方式调用
17	Open api daily request limit reached	每天请求量超限额
18	Open api qps request limit reached	QPS超限额
19	Open api total request limit reached	请求总量超限额
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
216100	invalid param	请求中包含非法参数,请检查后重新尝试
216102	service not support	请求了不支持的服务,请检查调用的url
216110	appid not exist	appid不存在,请确保调用接口所使用的应用归属于创建模型的百度云账号,且已开通该接口权限
216200	empty image	图片为空,请检查后重新尝试
216201	image format error	上传的图片格式错误,现阶段我们支持的图片格式为:PNG、JPG、JPEG、BMP,请进行转码或更换图片
216202	image size error	上传的图片大小错误,现阶段我们支持的图片大小为:base64编码后小于4M,分辨率不高于 4096*4096,请重新上传图片
216630	recognize error	识别错误,请再次请求,如果持续出现此类错误,请在控制台提交工单联系技术支持团队
282000	internal error	服务器内部错误,可能是图片尺寸过大文字太多识别超时,如果持续出现此类错误,请在控制台提交工单 联系技术支持团队
336100	model temporarily unavailable	模型长期未调用需激活,遇到该错误码请等待5分钟后再次请求,即可恢复正常,若反复重试依然报错或有 疑问请在百度智能云控制台内提交工单反馈

EasyDL 跨模态使用说明

EasyDL跨模态整体介绍

Hi,您好,欢迎使用百度EasyDL定制化训练和服务平台。

EasyDL平台的跨模态模型定制能力,基于文心·跨模态大模型的领先语义理解技术,为企业/开发者提供一整套跨模态定制与应用能力。

当前EasyDL平台提供了1种模型定制能力:

• 图文匹配:定制图文匹配模型,对文本及图片信息进行深度理解,计算两者的匹配度

EasyDL平台后续将提供更多类型的跨模态模型定制能力。

心 产品优势

心 可视化操作

无需深度学习专业知识,通过模型创建·数据上传·模型训练·模型发布全流程可视化便捷操作,最快15分钟即可获得一个高精度模型。

操作步骤

Step 1 创建模型 确定模型名称,记录希望模型实现的功能。

Step 2 上传并标注数据 不同类型的任务对应的数据格式不一致,您可以上传未标注数据并使用平台提供的标注工具进行标注。或直接上传各任 务的标注数据。

Step 3 训练模型并校验效果 选择部署方式与算法,用上传的数据一键训练模型。

模型训练完成后,可在线校验模型效果。

Step 4 发布模型 根据训练时选择的部署方式,将模型以云端API的方式发布使用

更详细的操作指导,请参考各类模型的技术文档

心 高精度效果

EasyDL跨模态任务内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等 方面的新知识,实现模型效果不断进化。

文心大模型是百度发布的产业级知识增强大模型,是千行百业AI开发的首选基座大模型。文心大模型既包含基础通用的大模型,也包含了面向重 点领域和重点任务的大模型,还提供丰富的工具与平台,支撑企业与开发者进行高效便捷的应用开发。"知识增强"是文心的核心特色,文心能够 同时从大规模知识和海量多元数据中持续学习,如同站在巨人的肩膀上,训练效率和理解准确率都得到大幅提升,并具备了更好的可解释性。

心 灵活的部署方案

训练完成后,可将模型部署在公有云服务器上。

- 训练完成的模型存储在云端,可通过独立Rest API调用模型,实现AI能力与业务系统整合
- 具有完善的鉴权、流控等安全机制,GPU集群稳定承载高并发请求

图文匹配

整体介绍

心 任务简介

定制图文匹配模型,可实现视觉、文本跨模态理解能力,计算图文匹配度。您只需提供图片及文本的训练数据,即可训练获得图文匹配模型。

更多详情访问:EasyDL跨模态方向

₯应用场景

- 内容质量评论:计算互联网内容中图片与文案的匹配程度,进而量化评价内容质量,可应用于电商、广告营销、互联网社区等领域
- 图文素材推荐:基于对已有图文素材的匹配度分析,在用户进行内容创作时,推荐相匹配的图片或文案素材
- 其他:尽情脑洞大开,训练你希望实现的图文匹配模型

心 技术特色

图文匹配模型内置<u>文心大模型</u>,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本、图像数据中词汇、结构、语义 等方面的新知识,实现模型效果不断进化。

文心大模型是百度发布的产业级知识增强大模型,是千行百业AI开发的首选基座大模型。文心大模型既包含基础通用的大模型,也包含了面向重 点领域和重点任务的大模型,还提供丰富的工具与平台,支撑企业与开发者进行高效便捷的应用开发。"知识增强"是文心的核心特色,文心能够 同时从大规模知识和海量多元数据中持续学习,如同站在巨人的肩膀上,训练效率和理解准确率都得到大幅提升,并具备了更好的可解释性。

心 使用流程

训练模型的基本流程如下图所示,全程可视化简易操作,在数据已经准备好的情况下,最快15分钟即可获得定制模型。



数据准备

```
心 创建数据集并导入
```

创建数据集在训练模型之前,需要先在数据总览【创建数据集】。输入数据集名称(限制50汉字),默认生成数据集版本V1,标注类型为图文 匹配,配置后点击"完成",成功创建一条空的图文匹配数据集。

图	文匹配模型	∢≡	我的数据总	览					
	总览								
đ	模型中心		EasyData	智能数据服务平	台已上线,使	用EasyData可享受包括	5多人标注、数据清淡	先、数据采集等完整	数据服务
	我的模型		创建数据	集					
	创建模型								
	训练模型		ууvv0516	「 数据集组	ID: 9489				
	校验模型		版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗
	发布模型			10011	2		四六匹西	00((0 (0)	
2	EasyData数据服务		V2 😁	12211	3	● 巳元成	图义匹配	0% (0/3)	_
	数据总览		V1 💬	12210	9	● 已完成	图文匹配	0% (0/9)	-
	公开数据集								
8	公有云服务		为为为为	为为 🖸 数据	集组ID: 9486				
	在线服务		版本	粉捉住い	粉捉景	是近日)壮本	标注米刑	标注出本	法法

导入数据 创建数据集后,在「数据总览」页面中,找到该数据集,点击右侧操作列下的「导入」,即可进入导入数据页面,可以通过以下方式导入数据:

- 导入未标注的数据,在线进行数据标注
- 直接导入标注好的数据

不论您上传无标注信息的数据或有标注信息的数据,都需要以下述格式要求进行上传。 同时目前 有标注信息 上传格式仅支持 json (平台通用)

导入未标注的数据 本地导入

支持上传图片、压缩包

- 目前支持图片类型为jpg, png, bmp, jpeg, 图片大小限制在14M以内。
- 图片长宽比在3:1以内,其中最长边小于4096px,最短边大于30px。
- 上传已标注文件要求格式为zip格式压缩包

已有数据集支持选择百度云 BOS 导入、分享链接导入、平台已有数据集导入;支持选择线上已有的数据集,包括其他图像类模型的数据集

- BOS目录导入格式要求:请确保将全部图片已保存至同一层文件目录,该层目录下子文件目录及非相关内容(包括压缩包格式等)不导入
- 分享链接导入请确保将全部图片已保存至同一压缩包,压缩包仅支持zip格式,压缩前源文件大小限制5G以内;仅支持来自百度BOS、阿里 OSS、腾讯COS、华为OBS的共享链接
- 图片类型支持jpg/png/bmp/jpeg,单图需小于14M,长宽比小于 3:1,其中最长边需要小于4096px,最短边需要大于30px
- 您的账户下图片数据集大小限制为10万张图片,如果需要提升数据额度,可在平台提交工单

图2	文匹配模型	 我的数据总览 >	yyvv0	516/V2/导入				
	总览	┃ 创建信息 ~						
1	模型中心	数据集ID	12211		版本号	V2		
	我的模型	备注						
	创建模型							
	训练模型	┃ 标注信息 ∨						
	校验模型	标注类型	图文匹	他心	标注模板	图文[心配	
	发布模型	数据总量	3		已标注	0(进	<u>}</u> 度0.00%)	
~	EasyData数据服务	大小	0.26M					
	数据总览	┃数据清洗		本地导入				
	公开数据集	暂未做过数据清	青洗任务	BOS目录导入				
		┃ 导入数据		刀子 挺按守八 平台已有数据集				
	公有乙服务	数据标注状态		摄像头采集数据				
	在线服务	导入方式		请洗择			~	
				确认并返回				

导入已标注的数据本地导入上传压缩包,标注格式仅支持 json (平台通用)

- 上传已标注文件要求格式为zip格式压缩包,同时压缩前源文件大小在5GB以内
- 压缩包内需要包括图片源文件 (jpg/png/bmp/jpeg) 及同名的json格式标注文件,详细请见示例压缩包 已有数据集 支持选择百度云BOS导 入、分享链接导入、平台已有数据集导入,标注格式仅支持 json (平台通用)
- BOS目录导入格式要求:请确保将全部图片已保存至同一层文件目录,该层目录下子文件目录及非相关内容(包括压缩包格式等)不导入
- 分享链接导入请确保将全部图片已保存至同一压缩包,压缩包仅支持zip格式,压缩前源文件大小限制5G以内;仅支持来自百度BOS、阿里 OSS、腾讯COS、华为OBS的共享链接
- 图片类型支持jpg/png/bmp/jpeg,单图需小于14M,长宽比小于 3:1,其中最长边需要小于4096px,最短边需要大于30px
- 您的账户下图片数据集大小限制为10万张图片,如果需要提升数据额度,可在平台提交工单

图2	文匹配模型	.≡	我的数据总览	> yyvv(0516/V2/导入			
	总览		数据集ID	12211		版本号	V2	
đ	模型中心		备注					
	我的模型		┃标注信息 〜					
	创建模型		标注类型	图文图	匹配	标注模板	图文匹配	
	训练模型		数据总量	3		已标注	0(进度0.00	%)
	校验模型		大小	0.26	M			
	发布模型		┃数据清洗					
22	EasyData数据服务		暂未做过数据》	青洗任爹	_务 本地导入			
	数据总览		┃导入数据		BOS目录导入			
	公开数据集		数据标注状态		7字证按守八 平台已有数据集			
	公有云服务		导入方式		本地导入		~	
	在线服务		上传压缩包		<u></u> 1 上传压缩包			
					确认并返回			

心 在线标注

在线标注 Step 1 进入标注页面 上传未标注的数据后,可以通过以下方式进入标注页面:

• 在「数据总览」页面,该数据集对应的操作列下,点击「查看与标注」,即可进入标注页面

图文匹配模型	÷	我的数据总	览								提交工单
88 总览											
曲 模型中心		EasyDat	a智能数据服务P	平台已上线,傅	用EasyData可享受包持	括多人标注、数据清洗	、数据采集等完整	逐数据服务 立即前行	ŧ		
我的模型		创建数据	集							输入数据集;	名称或ID Q
创建模型											
训练模型		yyvv0516	I 数据集组	∃ID: 9489					□ 新増版2	18 全部版本	回 删除
校验模型		版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作		
发布模型 図 EasyData数据服务		V2 😳	12211	з	• 已完成	图文匹配	0% (0/3)	-	查看 多人标注	三导入 标注	
数据总览		V1 ⊙	12210	9	● 已完成	图文匹配	0% (0/9)	-	查看 多人标注	- 导入 标注	
公开数据集											
目 公有云服务		为为为为	为为 🖸 数据	s集组ID: 948	6				□ 新増版2		□ 删除
在线服务		版本	数据集ID	数据量	最近导入状态	标注类型	标注状态	清洗状态	操作		

Step 2 进行文本标注

针对尚未进行标注的数据,通过以下方式进行标注:

- 选中右侧文本框输入描述文本,文本长度限制在32个汉字。
- 点击保存文字描述,点击下一张图片,您将开始对下一张图片进行文本描述。

针对已进行标注的数据,通过以下方式进行标注修改:

- 进入需修改标签的文本标注页面,选中右边文本框重新编辑描述文本,文本长度限制在32个汉字。
- 点击保存文字描述,点击下一张图片,您将开始对下一张图片进行文本描述。



Step 3 查看标注信息 通过以下方式查看已标注的文本信息:

• 在「数据总览」页面,该数据集对应的操作列下,点击「查看」,进入查看标注页面后,点击「有标注信息」,即可查看标注图文的相应情况

图文匹配模型 ∢──	我的数据总览 > 【跨模态】γγwv0516/V2/查看
器 总览 晶 模型中心	全部 (3) 有标注信息 (0) 无标注信息 (3) 导入图片 标注图片
我的模型 创建模型	切換展示模式 ⑦ 筛选 ∨ □本页全选 白 删除
训练模型	
校強模型	
☑ EasyData数据服务	无描述文本 无描述文本 无描述文本
数据总览	
公开数据集	
目 公有云服务	
在线服务	
	〈 1 〉 42 条/页 丶

• 点击图文视角切换,转换为文字列表信息,点击查看,可查看标注图文的相应情况

图文匹配模型 √三	我的数据总览 > 【跨模态】yyyv0516/V2/查看	
88 总览	◆第 (2)	
曲 模型中心	主命(5) 行列に目記(10) 元列に目記(13)	守入国方 杨注国方
我的模型	切换展示模式 ⑦ 筛选 ✔	本页全选 回 删除
创建模型		
训练模型	序号 文本内容摘要	操作
校验模型	1	查看 删除
发布模型		
☑ EasyData数据服务	2	查看 删除
数据总览	3	查看 删除
公开数据集		
目 公有云服务		
在线服务		
		〈 1 〉 42 条/页 >

模型训练

の 创建模型

创建模型 在模型中心目录中选择「创建模型」,填写模型名称、模型归属、所属行业、应用场景、邮箱地址、联系方式、功能描述等信息,即可 创建模型。

目前单个用户在每种类型的模型下最多可创建10个模型,每个模型均支持多次训练,若需要创建超过10个以上的模型,请在百度云控制台内提 交工单反馈。

图文	て匹配模型	€	模型列表	> ê	刘建模型	
	总览					
Ē	模型中心		模型类别		图文匹配	
	我的模型		模型名称	*		
	创建模型		您的身份		企业管理者 企业员工	
	训练模型		八司夕秒	4	mellit	口中武众小江江
	校验模型		公司石朴	^	moint	口元成正业以证,
	发布模型		所属行业	*	请选择行业	
<u>//</u>	EasyData数据服务		应用场景	*	请选择应用场景 ~	
	数据总览		邮箱地址	*	z*****@baidu.com	
	公开数据集					
	公有云服务		联系方式	*	636****816	0
	在线服务		功能描述	*		
					0/500	
					完成	

模型创建成功后,即可在「我的模型」中看到刚刚创建的模型。

注:

1.创建模型后可持续新增模型版本,因此不必每次训练模型都创建模型。

2.目前单个用户在每种类型的模型下最多可创建10个模型,每个模型均支持多次训练。

3.如果您是企业用户,建议您按照真实企业信息进行填写,便于EasyDL团队后续更好的为您服务。

训练模型

完成数据的标注,或提交已标注的数据后,即可在「模型中心」目录中点击「训练模型」,开始模型的训练。

按以下步骤操作,启动模型训练:

图文匹配模	₩ (三)	训练模型				
88 总览		V# 19 28 88	Not had been tool	827 JF		
由 模型中心		达洋模型	pm测114保型	<u>弗一</u> 亚		
我的模型	₽	训练配置				
创建模型	型	部署方式	公有云部署	_		
训练模型	型	选择算法	◎ 高精度 ⑦ 💿 高性能 ⑦	第二步		
校验模型	坦	添加数据		5		
发布模型	민	添加数据集	+ 请选择 第三步			
🖉 EasyDat	ita数据服务		854048	15-1-	48.0-	
数据总划	览		双括来	版中	3wTF	
公开数据	据集		为为为为为为	V1	移除	
自 公有云服	服务	自定义测试集 ⑦				
在线服务	务	20100-272-00	am 第四步	40.14	** +	10.18
		111123;37:492	- CAN	规始	<u>я</u> л	DINE
			O GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	免费
			 GPU V100 	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	单卡¥0.45/分钟
		训练费用 ¥0.45/ 开始训练	分钟 为保证训练任务顺利进行,请任	驱您的账户余额充足,可点击进行充值 立即充值		

Step 1 选择模型 选择此次训练的模型 Step 2 训练配置

部署方式

可选择「公有云部署」。

选择算法

您可以根据训练的需求,选择「高精度」或「高性能」算法。不同的算法将影响训练时间、预测速度与模型准确率。

- 如果您选择了高精度的模型,模型预测准确率更高,少于1000条样本同样有很好的效果。使用高精度的算法训练模型将会耗时更久,实验环境下1000个样本,预计在20-60分钟左右完成训练
- 高性能算法即将对外提供。相同训练数据量的情况下,训练耗时更短,模型预测速度更快。使用10000条训练样本,将在10min内完成训练.同样的数据量情况下,效果比高精度的模型4-5%

「高精度」算法内置文心大模型,将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义 等方面的新知识,实现模型效果不断进化

Step 3 添加数据

添加训练数据

- 可选择多个数据集
- 训练时间与数据量大小和您选择的模型类型有关,如果您选择的是高性能的模型,使用10000条训练样本将在10min内完成训练;如果您选择的是高精度的模型,使用10000条训练样本,将在20-60min完成训练

添加自定义测试集

上传不包含在训练集的测试数据,可获得更客观的模型效果评估结果。

添加自定义测试集的目的:

如果学生的期末考试是平时的练习题,那么学生可能通过记忆去解题,而不是通过学习的方法去做题,所以期末考试的试题应与平时作业不能一样,才能检验学生的学习成果。那么同理,AI模型的效果测试不能使用训练数据进行测试,应使用训练数据集外的数据测试,这样才能 真实的反映模型效果

期末考试的内容属于学期的内容,但不一定需要完全包括所学内容。同理,测试集的标签是训练集的全集或者子集即可

Step 4 训练模型

点击「开始训练」,即可开始训练模型。

- 训练时间与数据量大小、选择的算法、训练环境有关
- 训练环境可选择GPU P40或GPU V100。其中GPU P40可以免费使用;GPU V100训练速度更快,需要付费使用,可参考价格说明
- 模型训练过程中,可以设置训练完成的短信提醒并离开页面

平台提供付费算力,付费算力可用于模型训练,可根据实际需求购买算力使用时长。

各类算力价格如下:

算力环境	规格	算力	速度比例	价格
CPU	CPU_16核_64G	1	1	单卡¥4.02/小时
GPU P4	TeslaGPU_P4_8G显存单卡_12核CPU_40G内存	5.5 TeraFLOPS	1	单卡¥4.02/小时
GPU P40	TeslaGPU_P40_24G显存单卡_12核CPU_40G内存	12 TeraFLOPS	1.47	单卡¥21.60/小时
GPU V100	TeslaGPU_V100_16G显存单卡_12核CPU_56G内存	14 TeraFLOPS	3.66	单卡¥27.00/小时

优惠政策:

为回馈开发者长期以来对EasyDL平台的大力支持,训练算力将针对单账户 x 单操作台粒度提供5小时免费训练时长(例如,每账户享有跨模态图 文匹配操作台5小时免费训练时长)。

同时,用户此前购买的算力小时包仍生效使用,支持算力小时包和储值两种付费方式。算力按分钟计费,账单金额精确至小数点后2位。训练失 败、训练状态为排队中时长均不纳入收费时长。

心 评估效果

模型评估

可通过模型评估报告或模型校验了解模型效果:

• 模型评估报告:训练完成后,可以在【我的模型】列表中看到模型效果,以及详细的模型评估报告。

图:	文匹配模型	我的模型 > pm测试模型-VI模型评估报告
88	总览 模型中心 我的模型	部署方式 公有云API (目前仅、> 版本 VI > 图片数 10 文本数 10 训练时长 7分钟
	训练模型 校验模型	整体评估 pm测试模型 VI效果优异。由于目前训练集数螺赢较少,该结论仅供参考,建议扩充训练集得到更准确的评估效果。
2	发布模型 EasyData数据服务 数据总览	Mean Recall® Recall®1 Recall®1 Recall®1 Recall®1 Recall®10 Recall@10 <th< th=""></th<>
	公开数据集	
8	公有云服务	
	在线服务	

• 模型在线校验:可以在左侧导航中找到【校验模型】,在线校验模型效果。

图文匹配模型	Ē	校验模型	
器 总览 山 模型中心		选择模型 pm測试模型 ✓ 部署方式 公有云API ✓ 选择版本	× 1V
我的模型 创建模型		当前候型Mean Hecall 100.00% IPTA版目 step01: 上传需校验的图片	step02:输入需校验的文本
训练模型 校 <u>验模型</u>			点击上传文本
发布模型			请输入文本
些 EasyData数据服务 数据总览		<u>``</u>	
公开数据集 目 公有云服务		点击添加图片 或拖绳图片至此处	
在线服务		图片类型为jpg, png, bmp, jpog, 图片大小限制在4M内。	
		开始校验	

「完整评估结果」页面中将记录整体评估与详细评估的报告,包括该模型整体的Mean Recall、Recall@1 、Recall@5、Recall@10等指标。

整体评估中,各指标的释义如下:

- Mean Recall:通过模型计算得到top1、top5、top10匹配结果的召回率平均值,该数值越大表明模型效果越好
- Recall@1:通过模型计算得到top1匹配结果的召回率平均值,该数值越大表明模型效果越好
- Recall@5:通过模型计算得到top5匹配结果的召回率平均值,该数值越大表明模型效果越好
- Recall@10:通过模型计算得到top5匹配结果的召回率平均值,该数值越大表明模型效果越好

如果在训练阶段,使用的数据集中,数据集总量在100条以内,训练出来的模型的效果评估报告的参考价值较小,建议您训练时数据量准备 充足

模型校验

在完成训练后,发布模型前,可以先进行模型校验,以确保模型在实际环境中能获得预期的性能。操作方法如下:

- 1. 在左侧「模型中心」目录中点击「校验模型」,进入校验模型页面
- 2. 选择需要校验能力的模型、部署方式、版本,点击「启动模型校验服务」
- 3. 校验服务启动后,在左侧上传图片,右侧文本框内输入文本,点击「校验」后,识别结果栏将输出匹配度结果,您可参考匹配度说明了解匹 配度的分析方法。

心 匹配度说明

心如何运用匹配度分析?

图文匹配旨在为您选出"图片—文本"匹配度最优组合。因此,您可通过"一张图片 vs 多段文本"或"多张图片 vs 一段文本"进行多次校验模型,选出 最满意的图文匹配组合。

例:

• 通过上传一张图片,计算一张图片与多段文本的匹配度。经分析,上传的图片和5段文本匹配度大小不一,您可以根据模型给出的匹配度结合 自身感受选出最满意的组合;

如下表所示,前三段文字描述包含了图片主体"猫"和"蝴蝶",以及主体关系"追",因此图文匹配度较高;第四段文字仅包含主体,未说明关系,第 五段文字仅抽象的描述了主体属性,因此导致图文匹配度较低。

上传图片	匹配文本	匹配度
	草地上一只小猫在追一只蝴蝶	66.66%
and the second second	一只小猫为了追蝴蝶跳起来了	61.07%
, The second	一只跳起来的小猫在追空中的蝴蝶	60.76%
	草地·猫·蝴蝶	47.58%
	可爱跳脱萌宠	24.73%

• 也可以针对同一段文本,计算文本与多张图片的匹配度,通过排序得到最相关的图片。

如下图所示,前三张图片符合"水墨画"特点,同时图片主体包含"熊猫"和"竹子",以及主体关系"吃",因此图文匹配度较高;后两张不符合"水墨 画"特点,因此导致图文匹配度较低。



心 效果优化

通过模型迭代、检查并优化训练数据,能够提升模型效果。**模型迭代**

一个模型很难一次性就训练到最佳的效果,通常会需要结合模型评估报告和校验结果不断扩充数据和调优。

为此平台提供了模型迭代功能,即当模型训练完毕后,会生成一个最新的版本号,首次V1、之后V2……以此类推。可以通过调整训练数据和算法,多次训练,以获得适合业务需求的模型效果。

如果模型已经是上线状态,依然支持模型迭代,只是需要在训练完毕后更新线上服务接口,在接口地址不变的情况下可以持续优化效果。

检查并优化训练数据

- 检查是否存在训练数据过少的情况,建议图文标注匹配数量不少于1000个,如果低于这个量级建议扩充
- 检查测试模型的数据与训练数据的采集来源是否一致,如果设备不一致、或者采集的环境不一致,那么很可能会存在模型效果不错但实际测试 效果较差的情况。针对这种情况建议重新调整训练集,将训练数据与实际业务场景数据尽可能一致。

模型部署

^の发布公有云API

发布模型,生成在线API

训练完毕后可以在左侧导航栏中找到「发布模型」,依次进行以下操作即可发布公有云API:

- 1. 选择模型
- 2. 选择部署方式「公有云部署」
- 3. 选择版本
- 4. 自定义服务名称、接口地址后缀
- 5. 申请发布

发布模型界面示意:

图文匹配模型	 发布模型	
品 总览		
fb 模型中心	选择模型	pm测试模型 ~
我的模型	部署方式	公有云部署
创建模型		
训练模型	选择版本	V1 ~
校验模型	服务名称 *	
发布模型	接口地址 *	https://aip.baidubce.com/rpc/2.0/ai_custo
EasyData数据服务		m/v1/VQA/
数据总览	甘他亜戈	
公开数据集	共旧安水	若接口无法满足您的需求,请描述希望解决 的问题,500汉字以内
目 公有云服务		
在线服务		0/500
		提交申请

发布完成 申请发布后,通常的审核周期为T+1,即当天申请第二天可以审核完成。

如果需要加急、或者遇到莫名被拒的情况,请在百度云控制台内提交工单反馈。

心 调用公有云API

本文档主要说明定制化模型发布后获得的API如何使用,如有疑问可以通过以下方式联系我们:

- 在百度云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:868826008)联系群管

接口描述

基于自定义训练出的图文匹配模型,实现个性化图文匹配相似度计算。模型训练完毕后发布可获得定制API

接口鉴权

1、在EasyDL——控制台创建应用

2、应用列表页获取AK SK

♥ 全局 ~					Q	≣	¢	?	몹	۲	Q ~
EasyDL定制训练平台	产品服务 / EasyDL定制训练平台 - 应用列表 / 应用详情										
概览	应用详情										
应用列表	编辑 查看交档										
监控报表	应用名称	AppID		API Key			Secre	et Key			
技术文档	test2	14293933		bZsfqO2fUdHTDflxqF1	1mkTq			*** 显示			
云服另KK管理 离线SDK管理	EasyDL定制化训练服务:	L									

请求说明

HTTP 方法: POST

请求URL:请首先进行自定义模型训练,完成训练后可在服务列表中查看并获取url。 URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

```
参数 值
Content-Type application/json
```

注意:如果出现336001的错误码很可能是因为请求方式错误,需以json方式请求。

Body请求示例:

{ "text": "<UTF-8编码文本>", "image": "<base64数据>" }

Body中放置请求参数,参数详情如下:

请求参数

字段名 称	必须	类型	说明
text	是	string	文本,utf-8编码,支持txt格式,文本长度限制为32个字
image	是	string	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大4096px,支持 jpg/png/bmp格式 注意请去掉头部

请求示例代码

..... EasyDL 图文匹配 调用模型公有云API Python3实现 import json import base64 import requests 使用 requests 库发送请求 使用 pip (或者 pip3) 检查我的 python3 环境是否安装了该库,执行命令 pip freeze | grep requests 若返回值为空,则安装该库 pip install requests ###### 目标文本的 本地文件路径, UTF-8编码, 最大长度4096汉字 TEXT_FILEPATH = "【您的测试文本地址,例如:./example.txt】" ###### 目标图片的 本地文件路径,支持jpg/png/bmp格式 IMAGE_FILEPATH = "【您的测试图片地址,例如:./example.jpg】" ###### 服务详情 中的 接口地址 MODEL_API_URL = "【您的API地址】" ###### 调用 API 需要 ACCESS_TOKEN。若已有 ACCESS_TOKEN 则于下方填入该字符串 ###### 否则, 留空 ACCESS_TOKEN,于下方填入 该模型部署的 API_KEY 以及 SECRET_KEY,会自动申请并显示新 ACCESS_TOKEN ACCESS_TOKEN = "【您的ACCESS_TOKEN】" API_KEY = "【您的API_KEY】" SECRET_KEY = "【您的SECRET_KEY】" PARAMS = {} print("1. 读取目标文本 '{}'".format(TEXT_FILEPATH)) with open(TEXT_FILEPATH, 'r') as f: text_str = f.read() print("将读取的文本填入 PARAMS 的 'text' 字段") PARAMS["text"] = text_str print("2. 读取目标图片 '{}'".format(IMAGE_FILEPATH)) with open(IMAGE_FILEPATH, 'rb') as f: base64_data = base64.b64encode(f.read()) base64_str = base64_data.decode('UTF8') print("将 BASE64 编码后图片的字符串填入 PARAMS 的 'image' 字段") PARAMS["image"] = base64_str if not ACCESS_TOKEN: print("3. ACCESS_TOKEN 为空,调用鉴权接口获取TOKEN") auth_url = "https://aip.baidubce.com/oauth/2.0/token?grant_type=client_credentials" "&client_id={}&client_secret={}".format(API_KEY, SECRET_KEY) auth_resp = requests.get(auth_url) auth_resp_json = auth_resp.json() ACCESS_TOKEN = auth_resp_json["access_token"] print("新 ACCESS_TOKEN: {}".format(ACCESS_TOKEN)) else: print("3. 使用已有 ACCESS_TOKEN") print("4. 向模型接口 'MODEL_API_URL' 发送请求") request_url = "{}?access_token={}".format(MODEL_API_URL, ACCESS_TOKEN) response = requests.post(url=request_url, json=PARAMS) response_json = response.json() response_str = json.dumps(response_json, indent=4, ensure_ascii=False) print("结果:\n{}".format(response_str))

返回说明

	字段名称	必须	类型	说明
	log_id	是	number	唯一的log id,用于问题定位
	error_code	否	number	错误码,当请求错误时返回
	error_msg	否	string	错误描述信息,当请求错误时返回
	score	否	number	图像与文本信息的匹配度,数值为0-1之间

在线调试

EasyDL零基础开发平台提供了 示例代码中心(API调试平台)-示例代码 ,用于帮助开发者在线调试接口,查看在线调用的请求内容和返回结果、复制和下载示例代码等功能,简单易用。

错误示例 需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(649285136)或工单联系技术支持团队。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请通过QQ群(649285136)或工单联系技术支持团队。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线计费 的接口,请通过QQ群(649285136)联系群管手动提额
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线计费的接口,请通过QQ群(649285136)联系群管手动提额
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线计费的 接口,请通过QQ群(649285136)联系群管手动提额
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期

336000 Internal error 服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(868826008)或工单联系技术支持团队

336001	Invalid Argument	入参格式有误,比如缺少必要参数、文本编码错误等等,可检查下文本编码、代码格式是否有误。有疑问请通过QQ群 (868826008)或工单联系技术支持团队;入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下 图片编码、代码格式是否有误。有疑问请在百度智能云控制台内提交工单反馈
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或者代码格式有误。有疑问请通过QQ群(868826008)或工单联系技术 支持团队
336003	Base64解码 失败	图片/音频/文本格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问 请在百度智能云控制台内提交工单反馈
336004	输入文本或图 片大小不合法	文本超出大小限制,每个文本限制512个字符(包括汉字、字符、数字或字母),有疑问请通过QQ群(868826008) 或工单联系技术支持团队 ;图片超出大小限制,图片限4M以内,请根据接口文档检查入参格式,有疑问请在百度智能 云控制台内提交工单反馈
336005	文本或图片解 码失败	文本编码错误,请检查并修改文本格式;图片编码错误(非jpg,bmp,png等常见图片格式),请检查并修改图片格式
336006	缺失必要参数	text字段缺失(未上传文本);image字段缺失(未上传图片)
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请通过QQ群(868826008)或工单 联系技术支持团队

EasyDL 零售行业版使用说明

零售版服务介绍

心简介

EasyDL是百度大脑中的一个定制化训练和服务平台,EasyDL零售版是EasyDL针对零售场景推出的行业版。EasyDL零售版提供两种服务,分别为 定制商品检测服务和货架拼接服务。

定制模型服务是EasyDL零售版的一项服务,专门用于训练货架合规性检查、自助结算台、无人零售货柜等场景下的定制化AI模型,训练出的模型 将以API的形式为客户提供服务。该服务包含以下2种定制模型:

1. 商品检测模型

- 适用场景:适用于适用于货架、端架、挂架等场景的商品陈列规范核查,支持识别商品基本信息,陈列顺序、层数、场景,统计排面数 量和占比
- 服务功能:
 - 商品基本信息识别:商品的名称、品牌、规格、编号;商品在图片中的坐标位置;商品识别的置信度
 - 商品陈列层数识别:商品陈列所在货架层数和货架总层数;商品陈列顺序;货架是否拍摄完整判断
 - 商品陈列场景识别:场景类型支持:普通货架、挂钩货架、斜口篮货架、端架、立式冰柜、卧式冰柜、冷风柜、地堆、割箱、地 龙、小端架、吧台
 - 商品排面占比统计:商品的排面数及排面占比率;每层货架可识别商品数量及未知商品数量;货架的总空位数、每层货架空位数
 及货架利用率
 - 商品陈列翻拍识别:识别商品陈列照片是对手机屏幕翻拍的可能性

2. 地堆检测模型

+适用场景:适用于堆箱、堆头、地龙等场景的商品陈列规范核查,支持识别商品基本信息,可视商品计数,纵深商品计数和占地面积+服务功能:

+ 商品基本信息识别:商品的名称、品牌、规格、编号;商品在图片中的坐标位置;商品识别的置信度;陈列顺序;可视商品计数, 纵深商品计数和占地面积

+ 商品陈列场景识别:场景类型支持:普通货架、挂钩货架、斜口篮货架、端架、立式冰柜、卧式冰柜、冷风柜、地堆、割箱、地 龙、小端架、吧台

+ 商品陈列翻拍识别:识别商品陈列照片是对手机屏幕翻拍的可能性

货架拼接服务基于百度EasyDL深度学习算法,支持将多个货架局部图片或视频,组合为完整货架图片。同时支持输出在完整货架图中的商品 检测结果,包含SKU的名称和数量,适用于需要在长货架进行商品检测的业务场景。

翻拍识别服务能够识别出通过手机翻拍出的商品陈列照片,比如商品货架陈列图片和地堆商品陈列图片,可降低人工审核人力,高效审核零售业

务中通过翻拍原有图片来造假的图片。

价签识别服务能够识别货架和促销活动中的价签信息,可识别各个价签在图片中的像素位置,以及价签内商品名称和价格,可用于洞察商品在线 下渠道分销的价格区间。

心功能介绍

心 定制商品检测服务

• AI模型训练平台

专门用于定制货架合规性检查、自助结算台、无人零售货柜等零售场景下识别商品的高精度AI模型。

• 全可视化操作

所有模型训练相关的操作都可以在网页上进行,无需编程,仅需五步即可部署定制化AI模型。

• 预置的商品库

预置近千种商品单品图可供客户在创建SKU时选择,用于合成训练数据,极大降低了训练数据采集和标注成本。

- 可自定义商品
- 客户可根据业务需求创建属于自己的商品,商品信息支持完全自定义,充分满足客户定制化需求。
- 全面的商品信息

商品基本信息识别

接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度

商品陈列层数识别

接口支持识别商品陈列所在货架层数,货架类型支持:货架、端架和立式冰柜内货架

商品陈列场景识别

接口支持识别商品陈列的场景,场景类型支持:普通货架、挂钩货架、斜口篮货架、端架、立式冰柜、地堆、割箱、地龙、小端架、吧台

商品排面占比统计

接口支持统计商品排面数/占比、未识别商品数、总空位数、每货架层的空位数及货架利用率

商品陈列翻拍识别

识别商品陈列照片是对手机屏幕翻拍的可能性

心 货架拼接服务

• 拼接和商品检测相结合

支持将多个货架局部图片或视频,组合为完整货架图片,并支持输出在完整货架图中的商品检测结果,包含SKU的名称和数量。

• 丰富的服务方式

支持三种服务方式:云服务API、完全开源的SDK以及可以直接体验的手机APP。

心 适用场景

• 货架合规性检查

精准识别出货架、冰柜和端架上陈列商品的数量和种类,为品牌商分析陈列排面占比,重点SKU分销率、缺货率、合格率提供数据支撑。

• 互动营销

训练定制的商品识别,实现对C端用户提交的商品图片进行识别,配合游戏规则完成闯关/抽奖式的互动营销。

⊙ 技术优势

• 免费训练与测试

平台提供大量免费的GPU训练资源,及每天500次免费调用量,用于模型迭代和效果验证,有效降低项目开发和测试成本。

• 高可用模型效果

针对零售场景专项算法调优,结合图像合成与增强技术提升模型泛化能力,模型准确率可达97%+,保证模型在生产环境中具有高可用性。

• 预置模型和数据

平台提供直接可用的商品检测API,覆盖常见商品品类;提供大量预置单品图数据,可用于训练定制模型,有效提升项目落地效率。

• 企业级安全保障

数据加密与隔离,完善的服务调用鉴权,为客户的数据和模型提供企业级安全保障。

• 功能完善且丰富

全面覆盖各类零售场景的商品识别需求,应对不同场景的业务需求提供多种可选服务类型。

心 与EasyDL物体检测的相同点和不同点

EasyDL零售版是EasyDL专门针对零售场景下识别商品推出的版本,相比于物体检测模型,零售版更贴合快消零售场景下的业务需求,专门用于 训练货架合规性检查、自助结算台、无人零售货柜等场景下的定制化商品检测AI模型,训练出的模型可发布成云服务API,服务支持四种功能:商 品基本信息识别、商品陈列层数识别、商品陈列场景识别和商品排面占比统计,适用于识别货架中的商品信息,商品计数和陈列顺序等,辅助货 架商品陈列合规检查,如铺货率、陈列情况等。

相同点

同为检测模型,接口支持返回目标物体的名称和物体在图片上的位置。

不同点

- 模型算法不同:零售版的模型算法专门根据零售行业的场景和业务需求做了专项优化,基于百度大脑大规模零售数据预训练,并利用商品增强 合成技术将SKU单品图合成实景货架图,有针对性的提高了训练商品检测模型的精确度。
- 训练数据不同:零售版的数据除了需要标注的实景业务图片外,支持为每个SKU标签上传单品图。SKU单品图用来降低实景图即训练数据采集和标注成本的。为了让模型能够完整地识别一个SKU,需要训练的图片中出现这个SKU的各个角度的样子,这意味着需要从实际业务场景中采集大量的图片,并且进行大量的标注工作。为了降低这部分的成本,我们通过数据合成和增强技术,只需为SKU上传各个角度的单品图,且单品图无需进行任何标注,即可让模型学习到这个SKU各个角度的样子。合成图片过程在训练阶段自动完成,无需操作操作和进行标注。EasyDL零售版平台预置了近千种商品,每个商品预置了50张左右的单品图,绝大多数情况下无需再自行上传单品图。
- 云服务API功能不同:零售版云服务API支持四种功能:商品基本信息识别、商品陈列层数识别、商品陈列场景识别和商品排面占比统计;物体 检测云服务API仅支持商品基本信息识别。

购买指南

心 开通付费及购买服务

EasyDL零售版的各项服务的「开通付费」、「购买QPS叠加包」、「购买调用量次数包」、「关闭付费」等操作均在EasyDL零售版控制台进行, 只需在相应需要付费使用的接口位置,跟随页面提示完成后续充值及付费,即可完成。

< 返回EasyDL总宽	零任	零售版概览											
零售版													
应用列表		产品介绍: (現开查看與努功統介绍) 現开											
用量统计	可用接口列表												
监控服表													
定制商品检测服务 🗸	,	API	模型ID	模型类型	模型名称		模型版本	状态		调用量限制	QPS限制	开通按量后付费 ?	购买QPS叠加包 ?
标准商品检测服务 >	1	111	30443	商品检测	1111		V1	● 免费使用	ŧ	剩余免费1000次	不保证并发	开通	购买 配额详情
货架拼接 >	5	silenceteapi	27185	商品检测	silenceseco	nd	V3	● 付费使用	ŧ	500次/天免费 + 超 出按量计费	4	终止付费	购买丨配额详情
	5	silencefirst0829	27135	商品检测	silencefirst		V3	●免费使用	ŧ	500次/天免费	不保证并发	开通	购买 配额详情
	1 8	文品检测											
	,	API		状态			调用量限制			QPS限制		开通按量后付费 ?	
	t	饮品检测		● 免费使用			500次/天免费			不保证并发		免费试用	
	1.6	日化品检测											
	,	API		状态			调用量限制			QPS限制		开通按量后付费 ?	
	8	日化品检测		● 免费使用			剩余免费1000次			不保证并发		免费试用	
	ß	商品陈列翻拍识别											
	,	API		状态		调用量的	限制	G	QPS限制		开通按量后付费	?	购买QPS叠加包 ?
	,	商品陈列翻拍识别		● 付费使用		剩余免售	费999次 + 超出按量计费	4	1		终止付费		购买丨配额详情

⑦ 定制商品检测服务

⊙价目表 - 按调用量后付费

定制商品检测模型云服务API支持以下三项功能,启停服务功能的方法请见 服务功能文档。

1. 商品基本信息识别 (必选)

接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度

- 商品陈列层数识别(可选)
 接口支持识别商品陈列所在货架层数,货架类型支持:货架、端架和立式冰柜内货架
- 商品陈列场景识别(可选)
 货架场景:货架、端架(小方货架);冰柜场景:立式冰柜;地堆场景:堆箱、割箱、地龙
- 商品排面占比统计(可选)
 接口支持统计商品排面数/占比、未识别商品数、空位数及货架利用率
- 商品陈列翻拍识别(可选)
 识别商品陈列照片是对手机屏幕翻拍的可能性

付费调用

每个模型发布的云服务API享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费,开通后即可开始计费使用。

注:为保证老用户可正常使用三项服务,对于服务功能上线前已经开通付费的定制模型API,默认对各项服务功能开通付费。商品陈列层数识 别和商品陈列场景识别两项服务功能,需在EasyDL零售版模型训练后台启动后方可使用,启动方法请见服务功能文档,启动后发生调用才会 按实际调用次数进行收费。

选择开启的功能不同,单次接口调用价格不同,详情如下:

 商品基本信息识别(必选),按月调用量阶梯计费,调用单价按照自然月累积调用量所落阶梯区间而变化。月初,上月累积的调用量清零,重 新开始累积本月调用量。

月调用量 (万次)	单次调用价格 (元)	QPS限制	说明
0<月调用量<=15	0.009	4	服务器支持每秒处理4次查询
15<月调用量<=150	0.008	4	服务器支持每秒处理4次查询
150<月调用量	0.007	4	服务器支持每秒处理4次查询

• 商品陈列层数识别 (可选),单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.04	4	服务器支持每秒处理4次查询
• 商品陈列场景识别	(可选),	单次调用额外收取费用
单次调用价格 (元)	QPS限制	说明
0.006	4	服务器支持每秒处理4次查询
• 商品排面占比统计	(可选),	单次调用额外收取费用
单次调用价格 (元)	QPS限制	说明
0.02	4	服务器支持每秒处理4次查询
 ● 商品陈列翻拍识别 	(可选)	单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.05	4	服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

EasyDL零售版各项服务都具有免费调用额度,开通付费后,免费调用额度仍保留,EasyDL零售版的定制商品检测服务免费额度如下:

服务名称	免费调用额度	QPS限制
定制商品检测服务	累计1000次	1~2,服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

费用举例

从2019-3-1至2019-3-31,**定制商品检测API**在三种服务功能都开启的情况下,月调用量为600万次(已除去免费额度),那么费用如下:

1. 商品基本信息识别的费用为43,650元,明细如下:

前15万次落入0~15w阶梯,单次调用0.009元/次,费用为1,350元;

中间15万~150万次落入15~150w阶梯,单次调用0.008元/次,费用为10,800元;

最后150万~600万次落入大于150w阶梯,单次调用0.007元/次,费用为31,500元;

共计43,650元

2. 商品陈列层数识别的费用为360,000元,明细如下:

月调用量为600万次,单次调用0.04元/次,费用为240,000元

3. 商品陈列场景识别的费用为,明细如下:

月调用量为600万次,单次调用0.006元/次,费用为36,000元

综上,三月费用合计319,650元。

心价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式	每QPS价格
按天购买	50元/天
按月购买	1050元/月

购买QPS叠加包前需保证已开通按量后付费或购买次数包

购买的QPS叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

购买QPS叠加包提升的额度对定制商品检测API的所有服务功能均有效

○ 定制地堆检测服务

心 价目表 - 按调用量后付费

定制地堆检测模型云服务API支持以下三项功能,启停服务功能的方法请见 服务功能文档。

- 商品基本信息识别(必选)
 接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度
- 2. 商品陈列场景识别 (可选)

货架场景:货架、端架(小方货架);冰柜场景:立式冰柜;地堆场景:堆箱、割箱、地龙

3. 商品陈列翻拍识别 (可选)

识别商品陈列照片是对手机屏幕翻拍的可能性

付费调用

每个模型发布的云服务API享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费,开通后即可开始计费使用。

注:为保证老用户可正常使用商品陈列场景识别服务功能,对于服务功能上线前已经开通付费的定制模型API,默认对各项服务功能开通付 费。商品陈列场景识别服务功能,需在EasyDL零售版模型训练后台启动后方可使用,启动方法请见服务功能文档,启动后发生调用才会按实 际调用次数进行收费。

选择开启的功能不同,单次接口调用价格不同,详情如下:

 商品基本信息识别(必选),按月调用量阶梯计费,调用单价按照自然月累积调用量所落阶梯区间而变化。月初,上月累积的调用量清零,重 新开始累积本月调用量。

单次调用价格 (元)	QPS限制	说明
0.016	4	服务器支持每秒处理4次查询

• 商品陈列场景识别 (可选),单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.006	4	服务器支持每秒处理4次查询

• 商品陈列翻拍识别 (可选),单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.05	4	服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

EasyDL零售版各项服务都具有免费调用额度,开通付费后,免费调用额度仍保留,EasyDL零售版的定制地堆检测服务免费额度如下:

服务名称	免费调用额度	QPS限制
定制地堆检测服务	累计1000次	1~2,服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

费用举例

从2019-3-1至2019-3-31,**定制地堆检测API**在三种服务功能都开启的情况下,月调用量为600万次(已除去免费额度),那么费用如下:

1. 商品基本信息识别的费用为96,000元,明细如下:

月调用量为600万次,单次调用0.016元/次,费用为96,000元

2. 商品陈列场景识别的费用为,明细如下:

月调用量为600万次,单次调用0.006元/次,费用为36,000元

综上,三月费用合计132,000元。

心价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格 按天购买 60元/天

按月购买 1200元/月

购买QPS叠加包前需保证已开通按量后付费或购买次数包

购买的QPS叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

购买QPS叠加包提升的额度对定制地堆检测API的所有服务功能均有效

心 翻拍识别服务

⊙价目表 - 按调用量后付费

付费调用

每个账户享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用, 价格如下:

调用量	价格	QPS限制	说明
每次调用	0.05元	4	服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

每个账号享有一定量免费调用额度,如下表:

云服务API	免费调用额度	QPS限制	说明
商品陈列翻拍识别	累计1000次	1~2	服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

心价目表 - 调用量次数包

如果对调用次数有预估,可以选择购买单次调用价格更低的次数包,价格如下:

规格	价格	QPS限制	有效期
1万次	490 元	4	1年
10万次	4,800 元	4	1年
100万次	45,000 元	4	1年
500万次	212,500 元	4	1年
1000万次	420,000 元	4	1年
2000万次	800,000 元	4	1年

购买后不可退款,次数包使用完后,开始按调用量每次0.05元收取费用

心价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格

按天购买 50元/天

按月购买 1200元/月

购买 QPS 叠加包需保证已开通按量后付费或购买次数包

购买的 QPS 叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

心 价签识别服务

⊙价目表 - 按调用量后付费

付费调用

每个账户享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用, 价格如下:

调用量	价格	QPS限制	说明
每次调用	0.05元	2	服务器支持每秒处理2次查询

注:调用失败不计费

免费额度

每个账号享有一定量免费调用额度,如下表:

云服务API 免费调用额度
 QPS限制
 说明

 价签识别
 累计1000次
 1
 服务器支持每秒处理1次查询

注:成功调用与失败调用均消耗免费额度

心 价目表 - 调用量次数包

如果业务上对调用次数有预估,可以选择购买单次调用价格更低的次数包,价格如下:

规格	价格	QPS限制	有效期
1万次	475 元	2	1年 (366天)
5万次	2,250 元	2	1年 (366天)
10万次	4,250 元	2	1年 (366天)
20万次	8,000 元	2	1年 (366天)
50万次	18,750 元	2	1年 (366天)
100万次	35,000 元	2	1年 (366天)
500万次	150,000 元	2	1年 (366天)

购买后不可退款,次数包使用完后,开始按调用量每次0.05元收取费用

唸价目表 - QPS叠加包

开通付费后,免费QPS为2,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格

按天购买 100元/天

按月购买 2000元/月

购买 QPS 叠加包需保证已开通按量后付费或购买次数包

购买的 QPS 叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

心 货架拼接服务

货架拼接服务支持按任务数后付费、任务次数包预付费和并发任务叠加包预付费三种计费方式。

心价目表 - 按任务数后付费

付费调用

每个账户享有累计200次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用,价 格如下:

任务数	价格 (元)	并发任务数限制	说明
每次拼接任务	0.2	1	只允许一个拼接任务在运行,超出限制的任务排队等待

注:任务启动后失败和运行前终止不计费,任务成功和运行后终止会计费用

免费额度

```
每个账号享有一定量免费调用额度,如下表:
```

服务	免费任务额度	并发任务数限制	说明	
货架拼接	累计200次	1	只允许一个拼接任务在运行,超出限制的任务排队等待	

注:任务成功与失败调用均消耗免费额度

心 价目表 - 任务次数包

如果对拼接任务次数有预估,可以选择购买单次任务价格更低的次数包,价格如下:

规格	价格	并发任务数限制	有效期
1千次	200 元	1	1年
1万次	1,900 元	1	1年
10万次	18,000 元	1	1年
100万次	150,000 元	1	1年
500万次	600,000 元	1	1年

购买后不可退款,任务次数包使用完后,开始按调用量每个任务0.2元收取费用

₯价目表 - 并发任务叠加包

开通付费后,并发任务数限制为1,如果有更多的并发请求需要,可以根据业务需求按天或按月购买并发任务叠加包,价格如下:

购买方式 每并发任务价格 按天购买 2元/天

按月购买 40元/月

购买 并发任务叠加包需保证已开通按量后付费或购买任务次数包

购买的并发任务叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

⊙ 余额不足提醒与欠费处理

₯余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。

₯欠费处理

- 北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。
- 欠费后您开通付费的产品将进入欠费状态,只能使用每日的免费额度,超过额度的请求系统将不再响应,且不再保证并发处理。

场景范例

拓展门店验证

⊙ 场景简介

对未覆盖铺货的待拓展门店数据,使用门脸文字识别服务确认门店是否真实存在,确认存在并转化为已铺货的业务门店后,可以在平台上移动至业务门店库,可在业务门店拜访场景时用作到访打卡验证。

心 使用步骤流程

1.创建门店库

2.导入待确认门店列表

3.使用门脸文字识别验证

4.移入业务门店

心 最佳实践

1.创建门店库

进入到EasyDL零售版门店管理页面,参考文档门店库创建 创建门店库。

2.导入待确认门店列表

创建好门店库后,进入到门店库,参考文档门店导入将本地的待确认的未铺货门店数据导入至门店库待确认门店列表。

3.使用门脸文字识别验证

为了验证收集到的未铺货门店数据是真实存在的,可以使用门脸文字识别服务进行验证,服务会在鉴权对应的门店库待确认门店列表中,根据待确认门店ID找到对应门店,通过传入的门脸图片和经纬度坐标进行真实性判断,满足判断条件即真实存在,服务或返回对应标识结果,门店会移动至待拓展门店列表中。

API URL

门店识别API URL: https://aip.baidubce.com/rpc/2.0/easydl/v1/retail/facade_v2

```
详细调用方法和参数解释,请见门脸文字识别技术文档
```

请求示例

请求API时传入tbc_store_id,调用门店识别API的请求参数示例如下:

```
{
    "image": "<base64数据>",
    "tbc_store_id":"QSDD121242331",
    "location": "116.271422,40.228393",
    "location_type": 3
}
```

返回示例

仅words_result中第一个结果会与待确认门店库中tbc_store_id对应门店信息进行真实性判断,words_result.is_match是判定标识字段,1表示拍摄的门店与门店库中tbc_store_id对应门店信息匹配,0则表示不匹配。

```
{
  "log_id": 3257574830993687599,
  "words_result_num": 2,
  "words_result": [
    {
       "brief": ""
       "score": 0.9812122343,
       "words": "Lawson",
       "is_match":1
       "type": "ocr",
       "channel": {
          "store_type": "便利店",
          "chain_type": "连锁便利店"
       "score": 0.9700000286,
       "type": "ocr"
       "brief": ""
       "words": "天天超市"
       "channel": {
          "store type": "小店",
          "chain_type": "独立超市"
}
```

4.移入业务门店

经过第3点验证通过的门店,会自动移动到待拓展门店列表中,待业务侧确认该待拓展的门店已经成功铺货后,可在门店库详情页面,将这些门店移入至业务门店列表,移入后,可使用门脸文字识别服务进行日常业务门店拜访打卡的验证,该场景范例可参考文档:业务门店拜访

业务门店拜访
对已铺货的业务门店拜访过程中,使用门脸文字识别服务验证拜访打卡的是否为当前在SFA里选择的门店。

心 使用步骤流程

1.创建门店库

2.导入业务门店列表

3.使用门脸文字识别验证

心 最佳实践

1.创建门店库

进入到EasyDL零售版门店管理页面,参考文档 门店库创建 创建门店库。

2.导入业务门店列表

创建好门店库后,进入到门店库,参考文档门店导入将本地的业务门店数据导入至门店库业务门店列表。

3.使用门脸文字识别验证

为了验证拜访打卡的门店与在SFA中选择的门店一致,可以使用门脸文字识别服务进行验证,服务会在鉴权对应的门店库业务门店列表中,通过 传入的门脸图片和经纬度坐标进行匹配度检索,检索到结果的会返回在门店库中对应门店的ID,如果ID和SFA中选择的门店ID一致,则说明当前 拍摄的门店,服务或返回对应标识结果。

API URL

门店识别API URL: https://aip.baidubce.com/rpc/2.0/easydl/v1/retail/facade_v2

详细调用方法和参数解释,请见门脸文字识别技术文档

请求示例

请求API时传入tbc_store_id,调用门店识别API的请求参数示例如下:

```
{
"image": "<base64数据>",
"location": "116.271422,40.228393",
"location_type": 3
}
```

返回示例

当在业务门店列表中检索到匹配的门店时,会通过words_result中的第一个结果返回,words_result.type是判定标识字段,如果返回"image_search",表示该结果为检索到的结果,"store_id"和"words"分别为门店库中存的门店ID和门店名称;如果返回"ocr",表示文字识别结果,即在门店库中没有检索到门店,此时返回的是OCR文字识别的结果。

```
"log_id": 3257574830993687599,
  "words_result_num": 2,
  "words_result": [
    {
       "store_id": "SFA123456",
       "brief": "",
       "score": 1.0,
       "words": "Lawson",
       "is_match":-1,
       "type": "image_search",
       "channel": {
         "store_type": "便利店",
          "chain_type": "连锁便利店"
         }
       "score": 0.970000286,
       "type": "ocr",
       "brief": "",
       "words": "天天超市",
       "channel": {
          "store_type": "小店",
          "chain_type": "独立超市"
}
```

店内陈列洞察

```
 ⁰ 场景简介
```

洞察已铺货的门店店内商品陈列情况,如识别货架中的商品信息,商品计数和陈列顺序等,辅助货架商品陈列合规检查,如铺货率、陈列情况等。

```
心 使用步骤流程
```

1.训练AI识别模型

2.调用云服务API

3.查看商品陈列信息

```
心 最佳实践
```

1.训练Al识别模型

参考文档 快速训练一个模型, 训练一个支持识别所需SKU的AI模型。

2.调用云服务API

- 参考文档 服务功能,按需求开启云服务API的服务功能。
- 参考文档 商品检测API调用方法, 按需要获取的业务指标进行字段取值。

请求示例

请求API时,请求参数image_store传入业务门店ID,请求参数示例如下:

```
{
"image": "<base64数据>",
"image_store": "SFA123456"
}
```

3.查看商品陈列信息

当调用API时,在请求参数"image_store"中传入业务门店ID,可在平台业务门店列表中的门店详情中查看到对应门店ID的商品陈列信息。

异常拍照监测

⁰ 场景简介

对拜访门店过程中拍摄的照片进行异常拍摄监测,平台目前支持识别2类异常行为:翻拍和窜拍。

- 商品陈列翻拍识别能够识别出通过手机翻拍出的商品陈列照片,比如商品货架陈列图片和地堆商品陈列图片,可降低人工审核人力,高效审核 零售业务中通过翻拍原有图片来造假的图片。

心 使用步骤流程

1.调用商品陈列翻拍识别API

2.调用窜拍识别API

3.查看翻拍信息

4.查看窜拍信息

心 最佳实践

1.调用商品陈列翻拍识别API

参考文档 翻拍识别API调用方法 调用服务API。

请求示例

请求参数示例如下:

"image": "<base64数据>" }

返回示例

设定一个判定为翻拍图片的阈值,即如果results.name为"recapture"的"score"大于这个值,则认为这张图片是翻拍。

```
{
    "log_id": 3142711278302327859,
    "results": [
        {
            "name": "original",
            "score": 0.9980294108390808
        },
        {
            "name": "recapture",
            "score": 0.001970605691894889
        }
    ]
}
```

2.调用窜拍识别API

参考文档 窜拍识别API调用方法,调用服务API,接口调用流程如下:

1) 创建图集,获得图集ID

创建图集API:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/create_dataset

2) 指定图集ID,把要识别窜拍的图片上传到该图集

上传图片API:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/upload

3) 创建任务,指定图集ID,对该图集做窜拍识别,获取任务ID

创建任务API:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/create_task

4)指定任务ID,轮询任务结果,可以每隔一定时间调用API查询任务结果,比如10s查询一次,任务完成后可以查到结果

查询结果API:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/query

另外,可通过图集列表API和任务列表API查询所有图集和任务。

3.查看翻拍信息

当调用API时,在请求参数"image_store"中传入业务门店ID,可在平台业务门店列表中的门店详情中查看到对应门店ID的商品陈列信息,其中包含翻拍信息。

4.查看窜拍信息

当调用API时,在请求参数"image_store"中传入业务门店ID,可在平台业务门店列表中的门店详情中查看到对应门店ID的窜拍信息。

定制商品检测服务

服务介绍

```
心简介
```

定制商品检测服务是EasyDL零售版的一项服务,专门用于训练货架合规性检查、自助结算台、无人零售货柜等场景下的定制化商品检测AI模型, 训练出的模型将以API的形式为客户提供服务,服务支持四种功能:商品基本信息识别、商品陈列层数识别、商品陈列场景识别和商品排面占比统 计,适用于识别货架中的商品信息,商品计数和陈列顺序等,辅助货架商品陈列合规检查,如铺货率、陈列情况等。

心 功能介绍

• AI模型训练平台

专门用于定制货架合规性检查、自助结算台、无人零售货柜等零售场景下识别商品的高精度AI模型。

• 全可视化操作

所有模型训练相关的操作都可以在网页上进行,无需编程,仅需五步即可部署定制化AI模型。

• 预置的商品库

预置近千种商品单品图可供客户在创建SKU时选择,用于合成训练数据,极大降低了训练数据采集和标注成本。

• 可自定义商品

客户可根据业务需求创建属于自己的商品,商品信息支持完全自定义,充分满足客户定制化需求。

• 全面的商品信息

商品基本信息识别

接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度

商品陈列层数识别

接口支持识别商品陈列所在货架层数,货架总层数以及商品的陈列顺序,货架类型支持:货架、端架、冷风柜和立式冰柜内货架

商品陈列场景识别

接口支持识别商品陈列的场景,场景类型支持:普通货架、挂钩货架、斜口篮货架、端架、立式冰柜、卧式冰柜、冷风柜、地堆、割箱、 地龙、小端架、吧台

商品排面占比统计

接口支持统计商品排面数/占比、未识别商品数、总空位数、每货架层的空位数及货架利用率

商品陈列翻拍识别

识别商品陈列照片是对手机屏幕翻拍的可能性

心 使用流程

- 1. 创建模型
- 2. 创建SKU
- 3. 实景图片上传和标注
- 4. 训练并校验模型
- 5. 发布模型,获得定制的商品检测API

⊙ 特色优势

• 免费训练与测试

平台提供大量免费的GPU训练资源,及每天1000次免费调用量,用于模型迭代和效果验证,有效降低项目开发和测试成本

• 高可用模型效果

针对零售场景专项算法调优,结合图像合成与增强技术提升模型泛化能力,模型准确率可达97%+,保证模型在生产环境中具有高可用性

• 预置模型和数据

平台提供直接可用的商品检测API,覆盖常见商品品类;提供大量预置单品图数据,可用于训练定制模型,有效提升项目落地效率

• 企业级安全保障

数据加密与隔离,完善的服务调用鉴权,为客户的数据和模型提供企业级安全保障

心 适用场景

• 货架合规性检查

精准识别出货架、冰柜和端架上陈列商品的数量和种类,为品牌商分析陈列排面占比,重点SKU分销率、缺货率、合格率提供数据支撑

互动营销

训练定制的商品识别,实现对C端用户提交的商品图片进行识别,配合游戏规则完成闯关/抽奖式的互动营销

购买指南

心 开通付费及购买服务

EasyDL零售版的各项服务的「开通付费」、「购买QPS叠加包」、「购买调用量次数包」、「关闭付费」等操作均在EasyDL零售版控制台进行, 只需在相应需要付费使用的接口位置,跟随页面提示完成后续充值及付费,即可完成。

✓返回EasyDL总宽	零售版概览									
零售版										
应用列表	产品介绍: (展开	查看服务功能介绍)								展开
用量统计	可用接口列表									
监控报表	■ 已上线的定制接口									
定制商品检测服务 ~	API	模型ID	模型类型	模型名称	模型版本	状态	调用量限制	QPS限制	开通按量后付费 ?	购买QPS叠加包 ?
标准商品检测服务 >	111	30443	商品检测	1111	V1	●免费使用	剩余免费1000次	不保证并发	开通	购买 配额详情
货架拼接 >	silenceteapi	27185	商品检测	silencesecond	V3	● 付费使用	500次/天免费 + 超 出按量计费	4	终止付费	购买丨配额详情
	silencefirst0829	27135	商品检测	silencefirst	V3	● 免费使用	500次/天免费	不保证并发	开通	购买 配额详情
	【 饮品检测									
	API		状态		调用量限制		QPS限制		开通按量后付费 ?	
	饮品检测		● 免费使用		500次/天免费		不保证并发		免费试用	
	┃ 日化品检测									
	API		状态		调用量限制		QPS限制		开通按量后付费 ?	
	日化品检测		● 免费使用		剩余免费1000次		不保证并发		免费试用	
	┃ 商品陈列翻拍识别									
	API		状态	30J	用量限制	QPS限制		开通按量后付费	?	购买QPS叠加包 ?
	商品陈列翻拍识别		● 付费使用	炯	余免费999次 + 超出按量计费	4		终止付费		购买丨配额详情

心 定制商品检测服务

价目表 - 按调用量后付费

定制商品检测模型云服务API支持以下三项功能,启停服务功能的方法请见 服务功能文档。

- 商品基本信息识别(必选)
 接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度
- 商品陈列层数识别(可选)
 接口支持识别商品陈列所在货架层数,货架类型支持:货架、端架和立式冰柜内货架
- 商品陈列场景识别(可选)
 货架场景:货架、端架(小方货架);冰柜场景:立式冰柜;地堆场景:堆箱、割箱、地龙
- 4. 商品排面占比统计 (可选)

接口支持统计商品排面数/占比、未识别商品数、空位数及货架利用率

付费调用

每个模型发布的云服务API享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费,开通后即可开始计费使用。

注:为保证老用户可正常使用三项服务,对于服务功能上线前已经开通付费的定制模型API,默认对各项服务功能开通付费。商品陈列层数识 别和商品陈列场景识别两项服务功能,需在EasyDL零售版模型训练后台启动后方可使用,启动方法请见 服务功能文档,启动后发生调用才会 按实际调用次数进行收费。

选择开启的功能不同,单次接口调用价格不同,详情如下:

商品基本信息识别(必选),按月调用量阶梯计费,调用单价按照自然月累积调用量所落阶梯区间而变化。月初,上月累积的调用量清零,重新开始累积本月调用量。

月调用量 (万次)	单次调用价格 (元)	QPS限制	说明
0<月调用量<=15	0.009	4	服务器支持每秒处理4次查询
15<月调用量<=150	0.008	4	服务器支持每秒处理4次查询
150<月调用量	0.007	4	服务器支持每秒处理4次查询

• 商品陈列层数识别 (可选),单次调用额外收取费用

į	单次调用价格 (元)	QPS限制	说明
C	0.04	4	服务器支持每秒处理4次查询
•	商品陈列场景识别	(可选) ,	,单次调用额外收取费用
ĺ	单次调用价格 (元)	QPS限制	说明
0	0.006	4	服务器支持每秒处理4次查询
•	商品排面占比统计	· (可选) ,	,单次调用额外收取费用
i	単次调用价格 (元)	QPS限制	说明

0.02	4	服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

EasyDL零售版各项服务都具有免费调用额度,开通付费后,免费调用额度仍保留,EasyDL零售版的定制商品检测服务免费额度如下:

 服务名称
 免费调用额度
 QPS限制

 定制商品检测服务
 累计1000次
 1~2,服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

费用举例

从2019-3-1至2019-3-31,**定制商品检测API**在三种服务功能都开启的情况下,月调用量为600万次(已除去免费额度),那么费用如下:

1. 商品基本信息识别的费用为43,650元,明细如下:

前15万次落入0~15w阶梯,单次调用0.009元/次,费用为1,350元;

中间15万~150万次落入15~150w阶梯,单次调用0.008元/次,费用为10,800元;

最后150万~600万次落入大于150w阶梯,单次调用0.007元/次,费用为31,500元;

共计43,650元

2. 商品陈列层数识别的费用为360,000元,明细如下:

月调用量为600万次,单次调用0.04元/次,费用为240,000元

3. 商品陈列场景识别的费用为,明细如下:

月调用量为600万次,单次调用0.006元/次,费用为36,000元

综上,三月费用合计319,650元。

价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

 购买方式
 每QPS价格

 按天购买
 50元/天

 按月购买
 1050元/月

购买QPS叠加包前需保证已开通按量后付费或购买次数包

购买的QPS叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

购买QPS叠加包提升的额度对定制商品检测API的所有服务功能均有效

⊙ 定制地堆检测服务

价目表 - 按调用量后付费

定制地堆检测模型云服务API支持以下三项功能,启停服务功能的方法请见 服务功能文档。

- 商品基本信息识别(必选)
 接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度
- 2. 商品陈列场景识别 (可选)
 - 货架场景:货架、端架(小方货架);冰柜场景:立式冰柜;地堆场景:堆箱、割箱、地龙

付费调用

每个模型发布的云服务API享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费,开通后即可开始计费使用。

注:为保证老用户可正常使用商品陈列场景识别服务功能,对于服务功能上线前已经开通付费的定制模型API,默认对各项服务功能开通付 费。商品陈列场景识别服务功能,需在EasyDL零售版模型训练后台启动后方可使用,启动方法请见 服务功能文档,启动后发生调用才会按实 际调用次数进行收费。 • **商品基本信息识别(必选)**,按月调用量阶梯计费,调用单价按照自然月累积调用量所落阶梯区间而变化。月初,上月累积的调用量清零,重 新开始累积本月调用量。

单次调用价格 (元)	QPS限制	说明
0.016	4	服务器支持每秒处理4次查询

• 商品陈列场景识别 (可选),单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.006	4	服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

EasyDL零售版各项服务都具有免费调用额度,开通付费后,免费调用额度仍保留,EasyDL零售版的定制地堆检测服务免费额度如下:

服务名称	免费调用额度	QPS限制		
定制地堆检测服务	累计1000次	1~2,服务器支持每秒处理1~2次查询		

注:成功调用与失败调用均消耗免费额度

费用举例

从2019-3-1至2019-3-31,**定制地堆检测API**在三种服务功能都开启的情况下,月调用量为600万次(已除去免费额度),那么费用如下:

1. 商品基本信息识别的费用为96,000元,明细如下:

月调用量为600万次,单次调用0.016元/次,费用为96,000元

2. 商品陈列场景识别的费用为,明细如下:

月调用量为600万次,单次调用0.006元/次,费用为36,000元

综上,三月费用合计132,000元。

价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格 按天购买 60元/天 按月购买 1200元/月

购买QPS叠加包前需保证已开通按量后付费或购买次数包

购买的QPS叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

购买QPS叠加包提升的额度对定制地堆检测API的所有服务功能均有效

快速训练一个模型

心 简介

本文档介绍使用EasyDL零售版商品检测快速训练一个识别可口可乐的商品检测模型,基本流程如下:

1.创建模型

2.创建SKU

3.上传和标注训练数据

4.训练模型

5.发布模型

6.使用模型API

の 步骤1.创建模型

这个步骤将会介绍如何创建模型

进入创建模型页面

在EasyDL零售版商品检测产品主页点击【开始训练】按钮进入到模型训练页,下面会出现两种情况:

- 第一种,如果您没有登录百度智能云,则会跳转到百度智能云登录页面,没有百度账户的客户请先注册百度账户。登录后,会跳转到模型概览 页,点击【商品检测】卡片上的【立即定制】按钮,会跳转模型训练页面的创建模型页。
- 第二种,如果您已登录,会直接进入到【我的模型】页,该页面能够管理已经创建的模型,点击左侧列表中的【创建模型】进入创建模型页面。

创建模型

进入创建模型页面后你会看到如下图中展示的内容

模型中心	模型列表 > 创建模	型
我的模型	模型类别:	商品检测
创建模型	• 模型名称:	
训练模型		
校验模型	模型归属:	
发布模型		请输入公司名称
数据中心	• 应用场景:	请选择应用场景 🗸
我的SKU库		不同应用场景对应不同训练算法,请根据真实应用场景选择
实景图上传/标注	* 邮箱地址:	
AI服务中心	•联系方式:	
服务列表	* 功能描述:	
		0/500
		下一步

需要填写的项目如下:

● 模型名称

模型的名称

- 模型归属
 模型是属于公司的,还是属于个人的,如果是前者,请填写公司名称
- 应用场景

提示:**请根据真实业务应用场景选择**,选择的场景将会关联后端数据增强算法,若不确定,请选择"其他"

可选项为普通货架/货柜、智能结算台、无人零售柜、地堆商品和其他

• 邮箱地址

用于联系到您的邮箱地址

- Baidu 百度智能云文档
 - 联系方式

有效的联系方式将有助于后续模型上线的人工快速审核,以及更快的百度官方支持,推荐填写个人手机号码

• 功能描述

描述改模型将要应到的业务场景,详细的描述,在获取官方支持时,能帮助我们为您提供准确的使用建议

完成所有填写项后点击【下一步】按钮完成模型创建,创建完成后会跳转到【我的模型】页面。

心 步骤2.创建SKU

这个步骤将会介绍如何创建SKU,SKU是客户需要检测的商品,在训练品台上有两个作用,其一是"SKU名称_品牌名称_规格参数"用于标注训练数据的标签,二是SKU的单品图片用于商品增强合成技术,提高模型效果。

完成上一个步骤后,会跳转到【我的模型】页面,这时您会看到如下图展示的内容,由于模型还未训练,所以模型列表中没有显示模型的效果, 在训练模型前,需要先完成SKU的创建。

模型中心	模型列表
我的模型	创建模型
创建模型	
训练模型	【商品检测】可口可口识别 模型ID: 27257 回
校验模型	
发布模型	模型创建成功,若无SKU请先在"数据中心"创建SKU,上传实展图训练模型后,可以在此处查看模型的最新版本
数据中心	
我的SKU库	
实景图上传/标注	
AI服务中心	
服务列表	
	每页显示 4 🗸 < 1 >

点击左侧列表中的【我的SKU】进入SKU管理页面,点击【创建SKU】按钮进入创建SKU页面,您会看到如下图展示的内容

模型中心	我的SKU库 > 仓	我的SKU库 > 创建SKU					
我的模型	创建的SKU将以	("SKU名称_品牌名称_规格参数"的形式作为识别时道	反回的结果,命名时请避免重复				
创建模型	* 商品名称:	加、面映可乐、正夜片、海鲜生地等					
训练模型		知, 你听到小、工会儿、闷雪上加雪					
校验模型	*品牌名称:	如:可口可乐、乐事、海天等					
发布模型	* 规格参数:	如:330ml, 12粒装, 120g等					
数据中心	* 商品品类:	请选择商品品类	~				
我的SKU库	* 包装类型:	请选择包装类型	~				
实景图上传/标注	商品编号:	用于匹配商品系统中的编号,若无可不填					
AI服务中心							
服务列表	SKU单品图:	上传图片 用于有效提高模型效果 示例图片					
		创建SKU返回					

提示:在调用API接口识别SKU时,识别结果中SKU的名字是以"SKU名称_品牌名称_规格参数"的形式返回的,所以在填写SKU名称、品牌名称和规格参数时避免这三项内容重复。

需要填写的项目如下:

• SKU名称

SKU的名称,可适当填入SKU细节,例如:原味可乐,番茄味薯片,奥运版纯牛奶等

• 品牌名称

SKU的品牌名称,如可口可乐,乐事,伊利等

• 规格参数

SKU的规格,如330ml,500g,20片等

● 商品品类

可选择的有饮品、药品、保健品、零食、香烟、调味品、日用品和其他

- 包装类型
 可选择的有瓶装、罐装、袋装、盒装和其他
- 商品编号

如果您自身的业务系统中有现成SKU对应的商品编码,比如商品条形码,可以填在该填写框中,之后模型接口将支持返回该内容,用于您快速 匹配SKU

• SKU单品图

SKU的单品图不是模型训练的必须数据,其作用为用来合成实景图,连同手工标注的实景图一起用于训练,降低实景图即训练数据采集和标注 成本。拍摄角度和上传张数基本原则是覆盖实际检测场景可能出现的角度,具体请参考SKU单品图数据文档中进行单品图采集。

当每个SKU的实景图大于20张时,可以先不上传SKU单品图进行训练,后续提升模型效果以补充实景图为主,如果无法提供足量的实景图数据,可以通过上传SKU单品图来提升模型效果。

完成填写和上传SKU单品图上传后,页面内容显示如下图所示



点击【创建SKU】按钮完成创建,点击后回到【我的SKU库】,SKU列表中的SKU图数需要大约5秒的时间进行计算,刷新页面即可显示SKU单品 图片数。

心 步骤3.上传和标注训练数据

这个步骤将会介绍如何上传和标注训练数据,训练数据是SKU在货架上的实景图,需要客户从真实的业务场景中采集,这些图片在被正确标注中,可以用于训练成模型。

完成上一个步骤后,在左侧列表中点击【实景图上传/标注】进入上传和标注页面,在上传前请在实景图集选择栏内创建实景图集,如下图所示

商品村	新建实景图集			X		
定制商品检					商品计数,	辅助
查,如铺货	实景图集名称:	可口可乐货架图集				
	选择类型:	普通货架/货柜	\checkmark			
实景图上传/标注						
*						
大水回来,		创建实景图集	取消			

需要填写的项目如下:

• 实景图集名称

实景图集的名称,可适当填入SKU细节,例如:原味可乐,番茄味薯片,奥运版纯牛奶等

• 选择类型

实景图集的类型,请与创建模型时选择的应用场景保持一致,上传时只上传跟选择类型相同的实景图。可选项为普通货架/货柜、智能结算 台、无人零售柜、地堆商品和其他

完成创建实景图集后,页面显示为如下图所示的内容

实景图上传/标注		寻求数据支持: 百度官方数据服务 第三方数据服务商
• 实景图集: 优化0827	❤ 实景图上传API 查看实景图集标注规范	
图片筛选: 全部图片(0) 无标注信息(0)	待确认(0) 有标注信息(0) 辅助标注	Ø
《 标注示例 十 点击上传图片		
标注示例		SKU标签列表
1. 哪些需要框出? 所有图片中出现的目标物体都需要被框出(框可以 重叠)	 怎么画标注框? 恆 框应包含整个物体,且尽可能不要包含多余的 育景 如果图片中存在很多相同的目标物体,可以使 用右侧的锁定按钮。锁定标签后,只需要在左侧框 选目标物体即可 	★新建SKU 茶兀他子绿菜、农夫山泉_500ml< 雪碧_可口可乐_11 可乐_可口可乐_300ml 雪碧纤维_可口可乐_500ml 可乐原味_可口可乐_500ml
▲ 部框出	包含整个物体	
保存当前图片		

点击页面上【标注】为该实景图集上传作为训练数据的实景图,点击【标注示例】右侧的上传图片上传实景图。

上传完需要标注的图片后,EasyDL零售版的在线标注工具提供了辅助标注功能,该功能可以使用平台预置模型和用户自训练定制模型预先为未标 注的图片进行预标注,来降低整体标注工作的成本,如下图所示,使用方式请参考实景图标注文档。

模型中心	实景图上传/标注 寻求数据支持:百度官方数据	屠服务 第三方数据服务商
我的模型	· 实景图集: 测试辅助标注 ✔ 实景图上传API 查看实景图集标注规范	
创建模型		
训练模型	图片筛选: 全部图片(151) 无标注信息(0) 待确认(128) 有标注信息(23) 辅助标注 ②	
校验模型		+ >
发布模型	请在去倒选择标签	标签列表
数据中心		
我的SKU库		新建SKU
⑦暑图上传/标注		nl
A. (9) 19/12		
云服务调用数据		

实景图基本要求如下:

实景图的详细采集要求,请参考实景图数据要求文档

- 实景图片需要是从真实业务场景中采集来的数据
- 支持上传的图片格式为jpg,png,jpeg,bmp,大小限制为4M
- 建议图片尺寸:最长不超过4096px,最小不低于30px,长宽比3:1以内

标注基本要求如下:

实景图的具体标注要求,请参考实景图标注规范文档

- 完整并仅仅框选要识别的SKU
- 标注框不要框选到其它SKU或是价目标签等非要识别的SKU的干扰信息
- 在实景图中出现的所有要识别的SKU必须全部标注,不能遗漏

完成所有实景图的标注后,返回到【我的SKU库】可以查看到SKU列表中【实景图数】列显示标注了该SKU的实景图片的数量,如下图所示

模型中心	我的SKU库						
我的模型	创建SKU					请	输入SKU名称、品牌、规格进行搜索 Q
创建模型							
训练模型	SKU名称	品牌名称	规格参数	SKU图数	实景图数	状态	操作
校验模型	原味可乐	可口可乐	500ml	12	12	正常	查看 编辑 删除
发布模型							
数据中心							
我的SKU库							
实景图上传/标注							
AI服务中心							
服务列表	您已创建了1个SKL	J				每页	显示 10 🗸 < 1 >

の 步骤4.训练模型

这个步骤将会介绍如何训练模型

模型中心	模型训练								
我的模型	选择模型:	可口可口识别		~					
创建模型									
训练模型	添加SKU:	请添加要参与训练	东的SKU	~					
校验模型		序号	SKU名称	品牌名称	规格参数	操作	~	实景图集	SKU数量
发布模型		1	原味可乐	可口可乐	500ml	删除	~	可口可乐货架图集	1
数据中心									
我的SKU库									
实景图上传/标注									
AI服务中心									
服务列表									
		开始训练							

如上面图片所示,点击左侧列表中的【训练模型】,需要先后完成下面三项选择:

1. 选择要训练的模型

2. 选择需要想要模型支持检测的SKU,选择完成后,下方左侧会显示已添加的SKU,右侧会显示包含已添加SKU的实景图集

3. 选择要参与训练的实景图集

完成选择后,点击【开始训练】按钮页面跳转至【我的模型】页面,如下图所示,可以看到模型已进入训练状态,将鼠标移至状态"训练中"右边 的小问号上,可以查看训练进度,训练进度数值只是作为参考,所以推荐打开短信通知功能,这样就第一时间知晓模型训练完成了。

模型中心	模型列表							
我的模型	创建模型							
创建模型								
训练模型	【商品检测】;	可口可口识别	模型ID: 27257					88
校验模型	应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作	
发布模型	云服务	V1	训练中 💶	未申请	未发布		停止训练	
数据中心		训练进	度:	11%				
我的SKU库		🖌 完	成后短信提醒至 186	***650 🗹				
实景图上传/标注								
AI服务中心								
服务列表								
							毎页显示 4 🗸 <	I >



模型中心	模型列表								
我的模型	创建模型								
创建模型									
训练模型	【商品检测】;	可口可口识别	模型ID: 27257						Ī
校验模型	应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作		
发布模型						mAP 95.83% ?			
数据中心	云服务	V1	训练完成	未申请	未发布	精绷率 95.00% ? 召回率 92.31% ? 宗務评估结果	申请发布 校验 训练		
我的SKU库						JUE HILDER			
实景图上传/标注									
AI服务中心									
服务列表									
							毎页显示 4 🗸 <	1	>

心 步骤5.发布模型

这个步骤将会介绍如何将训练好的模型发布为服务API

模型中心	发布模型					
我的模型	1 选择模型:	饮品检测 🗸	标准接口规范参	考		
创建模型	 选择版本: 	V1 🗸	标准接口请求参	9考说明:		
训练模型	 3 · 服务名称: 	kelexuebiijance	字段名称	必须	类型	说明
校验模型	4 *接口地址:	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/	image	是	string	图像数据,base64编码,要求 base64编码后大小不超过4M, 最短边至少15px,最长边最大
发布模型		detection/ kelexuebixianwei				4096px,文持Jpg/png/pmp恰北
数据中心	5 其他要求:	可乐和雪碧纤维检测	threshold	否	number	阔值,默认为当前模型推荐阈值 (0-1之间),具体值可以在我的 模型列表-模型效果查看
我的SKU库			标准接口响应等	■段说明:		
实景图上传/标注		9/500	字段名称	必须	类型	说明
AI服务中心		提交申请	log id		number	唯一的log id 用于问题定位
服务列表		如果有私有化部署需求,请点此申请	log_id	۸L T	number	
			enor_code	1	number	相厌闷, 当阴水相误时返回

在模型训练好后,点击模型列表内对应模型「操作」列中的「申请发布」,或是在左侧导航栏点击「发布模型」可以进入发布模型页面,如上图 所示。在对应选项中选择和输入相应内容发起模型发布的申请:

1. 选择模型 (必选)

选择需要发布的模型,只能选择已经完成训练的模型

2. 选择版本 (必选)

选择需要发布的模型版本,只能选择完成训练且没有发布过的版本

3. 服务名称 (必填)

为发布的服务命名,**服务名称不得多于20个字符**

4. 接口地址 (必填)

自定义服务的API URL,接口地址需要多于5个字符但不能超过20个字符,仅限英文

5. 其他要求

如果有其他要求可以输入要求描述

填写完上述信息后,点击「提交申请」完成发布模型申请。提交申请后,模型列表内该模型的申请状态和服务状态为有以下几种情况:

申请状态	服务状态	状态描述
审核中	未发布	服务刚申请发布,模型在审核中
审核成功	发布中	服务通过审核,进入系统自动发布阶段
审核成功	已发布	服务发布成功
审核失败	未发布	服务未通过审核,通常为模型训练结果mAP < 0.6,如需申诉,可以加入官方QQ群(群号:1009661589)咨询群 管

提示:第一次申请发布的模型需要人工审核,通常4小时内完成,如果希望加急上线,请加入官方QQ群(群号:1009661589)咨询群管高优 审核。非第一次申请发布的模型,如果模型训练结果mAP>0.6,则会自动通过审批。审批完成后,大约需要5分钟左右自动完成发布。

^の步骤6.使用模型API

发布成功后,可以点击模型列表内「操作」列中的「配置服务功能」,如下图:

模型中心	模型列表						
我的模型	创建模型						
创建模型							
训练模型	【商品检测】饮品	金测 模型ID:	27899				🔜 训练 🕓 历史版本 🗴 删除
校验模型	部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作
发布模型						mAP: 96.95% 🛛	
数据中心	公有云API	V1	训练完成	审核成功	已发布	精确率: 88.46% ② 召回率: 92.00% ⑦ 完整评估结果	查看版本详情 配置服务功能 校验 体验H5 回

点击后弹出下图所示窗口,可以获取模型的云服务API URL,API使用方式请参考API调用方法文档。

我的模型:	饮品检测V1的公有云API服务详情						
服务名称:	kelexuebijjance	服务功能描述: 模型服务接口使用方法请见 <mark>API文档</mark>					
候空版本· 服务状态:	已发布	服务功能	接口返回字 段	内容			
接口地址:	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/detection/kelexuebixianwe i		name	商品标签,包含商品名 称、品牌、规格			
服务功能:	已选功能不同,单次接口调用价格不同,详细收费方式请见 计费文档	商品基本信息识	sku_code	商品编号			
	基础功能服务	别	score	识别结果的置信度			
	商品基本信息识别 商品标签(名称、品牌、规格)、编号、坐标、置信度		location	商品检测框在图片上的像 素坐标			
	可选服务功能 商品陈列层数识别	商品陈列场景识 别	scenes	图片中包含的陈列场景类 型,支持货架、端架和立 式冰柜			
	商品所在货架层数及货架总层数,支持区分不同货架		scene	每个商品所在的陈列场景			
	商品陈列场景识别 商品陈列的场景类型,支持货架、端架、立式冰柜		shelf	商品所在的货架编号,从 左往右依次递增			
	商品排面占比统计 商品陈列的排面数及占比	商品陈列层数识	layer	商品所在货架层数编号, 从上往下依次递增			
		60	layer_count	统计各货架的总层数			
	提示: 升启或夫财功能后, 约5分钟后生效		layer_top	判断货架最顶层是否拍摄 完整			
				综合统计结果,包含识别			

在该页面可以为模型的云服务API配置服务功能,支持以下四项功能:

- 商品基本信息识别(必选)
 接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度
- 商品陈列层数识别(可选)

接口支持识别商品陈列所在货架层数,货架总层数以及商品的陈列顺序,货架类型支持:货架、端架和立式冰柜内货架

- 商品陈列场景识别(可选)
 接口支持识别商品陈列的场景,场景类型支持:普通货架、挂钩货架、斜口篮货架、端架、立式冰柜、地堆、割箱、地龙、小端架、吧台
- 商品排面占比统计 (可选)

接口支持统计商品排面数/占比、未识别商品数、总空位数、每货架层的空位数及货架利用率

接口单次调用的费用,根据开启的功能不同而不同,详情可见购买指南文档。

可在页面随时开启和关闭可选的功能,变更功能后约5分钟生效,生效后单次调用费用按变更后的功能计费,接口将返回变更后的功能字段,详 情可见API调用方法文档。

模型创建

○进入模型训练页面

在EasyDL零售版产品主页点击【开始训练】按钮进入到模型训练页,下面会出现两种情况:

- 第一种,如果您没有登录百度智能云,则会跳转到百度智能云登录页面,没有百度账户的客户请先注册百度账户。登录后,会跳转到模型概览 页,点击【商品检测】卡片上的【立即定制】按钮,会跳转模型训练页面的创建模型页。
- 第二种,如果您已登录,会直接进入到【我的模型】页,该页面能够管理已经创建的模型,点击左侧列表中的【创建模型】进入创建模型页面。

心 创建模型

进入创建模型页面后你会看到如下图中展示的内容:

模型列表 > 创建模型
模型类别: 商品检测
. 指則(42).
* 198.au 1999 -
模型归属: 公司 个人
请输入公司名称
◎应用场景: 请选择应用场景 ∨
不同应用场景对应不同训练算法,请根据真实应用场景选择
• 邮箱地址:
• 联系方式: [7]
• 功能描述:
0/500
下一步

需要填写的项目如下:

- 模型名称
 模型的名称
- 模型归属

模型是属于公司的,还是属于个人的,如果是前者,请填写公司名称

• 应用场景

提示:选择模型将要应用的场景,请根据真实应用场景选择,选择的场景将会关联后端数据增强算法,若不确定,请选择"其他"

可选项为普通货架/货柜、智能结算台、无人零售柜、地堆商品和其他

• 邮箱地址

用于联系到您的邮箱地址

- 联系方式
 有效的联系方式将有助于后续模型上线的人工快速审核,以及更快的百度官方支持,推荐填写个人手机号码
- 功能描述
 描述改模型将要应到的业务场景,详细的描述,在获取官方支持时,能帮助我们为您提供准确的使用建议
- 像下图展示的一样完成所有填写项后点击【下一步】按钮完成模型创建,创建完成后会跳转到【我的模型】页面。

模型中心	模型列表 > 创建模型
我的模型	模型类别: 商品检测
创建模型	- 語副ぐや・ 「 ゴロゴロ30 00
训练模型	, אראר אראר אראר אראר אראר אראר אראר אר
校验模型	模型归属: 公司 个人
发布模型	百度
数据中心	→ 应用场景: 普通貨架/货柜 ✓
我的SKU库	不同应用场景对应不同训练算法,请根据真实应用场景选择
实景图上传/标注	● 郎箱地址: ■ ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●
AI服务中心	• 联系方式: IPBLIC IIP。 [2]
服务列表	 功能描述: 用于快速创建一个识别普通货架上可口可乐的AI模型
	24/500
	下一步
	が 下一步

数据准备

の SKU创建

在Web页面创建SKU

单个创建SKU

在模型训练页面,点击左侧列表中的【我的SKU】进入SKU管理页面,点击【创建SKU】按钮进入创建SKU页面。

我的SKU库						
创建SKU 本地批量上传 API上传				请输	入SKU名利	、品牌、规格进行搜索 Q
SKU名称	品牌名称	规格参数	SKU单品图数	被标注次数	状态	操作
番茄味薯片-袋装2	乐事	75g	0	1	正常	查看 编辑 删除
可乐-瓶装1	可口可乐	300ml	0	3	新建	查看 编辑 删除

您会看到如下图展示的内容:	:
---------------	---

模型中心	我的SKU库 > 创	l建SKU	
我的模型	创建的SKU将以	"SKU名称 品牌名称 规格参数"的形式作为识别时	返回的结果,命名时请避免重复
创建模型			
训练模型	* 商品名称:	如:原味可乐、止疼片、海鲜生抽等	
校验模型	* 品牌名称:	如:可口可乐、乐事、海天等	
发布模型	* 规格参数:	如:330ml,12粒装,120g等	
数据中心	* 商品品类:	请选择商品品类	\checkmark
我的SKU库	* 句装类型:	请诜择句装类刑	×
实景图上传/标注	UNAL	HAST-GAXE	
AI服务中心	商品编号:	用于匹配商品系统中的编号,若无可不填	
服务列表	SKU单品图:	上传图片 用于有效提高模型效果 示例图	图片
		创建SKU 返回	

提示:SKU识别结果中,SKU的名字是以"SKU名称_品牌名称_规格参数"的形式返回的,在填写SKU名称、品牌名称和规格参数时,请避免这 三项内容重复。

● SKU名称

SKU的名称,可适当填入SKU细节,例如:原味可乐,番茄味薯片,奥运版纯牛奶等

- 品牌名称
 SKU的品牌名称,如可口可乐,乐事,伊利等
- 规格参数
 SKU的规格,如330ml,500g,20片等
- 商品品类

可选择的有饮品、药品、保健品、零食、香烟、调味品、日用品和其他

包装类型

可选择的有瓶装、罐装、袋装、盒装和其他

• 商品编号

如果您自身的业务系统中有现成SKU对应的商品编码,比如商品条形码,可以填在该填写框中,之后模型接口将支持返回该内容,用于您快速 匹配SKU

• SKU单品图

SKU的单品图将用于商品增强合成,拍摄角度和上传张数基本原则是覆盖实际检测场景可能出现的角度,请参考「SKU单品图数据要求」文档 中进行单品图采集。如果不上传,将会降低模型的识别效果,可以点击页面上的【示例图片】查看SKU单品图样张。

完成填写和上传SKU单品图上传后,页面内容显示如下图所示

模型中心	我的SKU库 >	创建SKU
我的模型	创建的SKU将	3以"SKU名称 <u>。品牌名称,规格参数"的形式作为识别时返回的结果,命名时请避免重复</u>
创建模型 训练模型	* 商品名称:	原妹可乐
校验模型	* 品牌名称:	可பரசு
发布模型	* 规格参数:	500ml
数据中心	* 商品品类:	牧品 🗸
我的SKU库	* 包装类型:	瓶装 イ
实景图上传/标注	商品编号:	用于匹配商品系统中的编号,若无可不填
AI服务中心 服务列表	SKU单品图:	上传图片 用于有效提高模型效果 示例图片
		您已上作12张SKU单品图
		创建SKU 送回

点击【创建SKU】按钮完成创建,点击后回到【我的SKU库】,SKU列表中的SKU图数需要大约5秒的时间进行计算,刷新页面即可显示SKU单品 图片数。

批量创建SKU

在模型训练页面,点击左侧列表中的【我的SKU】进入SKU管理页面,点击【本地批量上传】按钮进入批量上传页面。

我的SKU库						
创建SKU 本地批量上传 API上传				请辅	ì入SKU名種	尔、品牌、规格进行搜索 Q
SKU名称	品牌名称	规格参数	SKU单品图数	被标注次数	状态	操作
番茄味薯片-袋装2	乐事	75g	0	1	正常	查看编辑删除
可乐-瓶装1	可口可乐	300ml	0	3	新建	查看 编辑 删除

您会看到如下图展示的内容:

我的SKU库 >	本地批量上传	
选择本地文件:		
文件要求说明:	下载示例文件 1. 支持xlxs和xls格式的 2. 文件编码支持:UTF 3. 文件必须包含以下表	Excel文件,以及CSV文件(分隔符为英文逗号) -8,可参考 编码查看和修改方法文档 头,各列的内容规范如下:
	表格名称	内容规范
	sku_name	不得多于30个字符,支持中文、英文、数字和-&'+*()%/
	brand_name	不得多于20个字符,只支持中文、英文、数字和-&'+*()%/
	specs	不得多于20个字符,只支持中文、英文、数字和-&.'+*()%/
	product_code	不得多于40个字符,支持英文、数字和-&'+*()%/
	category	数字代表商品类别:1为饮品,2为药品,3为保健品,4为零食,5为香烟,6为调味品,7为日用品,99为其它
	package_type	数字代表商品包装类型:1为瓶装,2为罐装,3为袋装,4为盒装,5为桶装,99为其它
确认	取消	

按照要求上传SKU标签信息文件即可,**建议下载示例文件修改后上传,以保证编码为UTF-8。**如果是自建的文件,可以参考编码查看和修改方法文 档进行编码修改。

提示:将文档上传前,请确认UTF-8编码的文档内的文字没有乱码。

使用API管理SKU数据

SKU的创建、删除、查询和上传SKU单品图均可以通过调用API实现,API使用方法请参考文档SKU管理API。

心 文档编码查看和修改方法

本文档介绍如何查看一个文档的编码以及修改编码的方法。

使用示例文件

进入SKU本地批量上传页面,点击页面上的下载示例文件,下载后在源文件中删除示例SKU后,再添加进需要创建的SKU信息。



查看和修改CSV文件的编码

提示:将文档上传前,请确认UTF-8编码的文档内的文字没有乱码。

下载Visual Studio Code,在Visual Studio Code打开csv文档,右下角可看到文件的编码格式,如下图所示:



如果不是UTF-8,点击显示编码格式的区域,输入框会弹出两个选项,如下图所示:

	SI	KU本地上传示例CSV文件.csv — Bai	du_API		_			
						ĵ)	\triangleright	Ξ
	Reopen with Encoding				、地上传示例(CSV文(
id id					age_type			
		3 番茄味薯片-袋装,乐事,75g,	87373789,4, <mark>3</mark>					
ру								
4-				Spaces: 4 UTF	-8 LF Plair			

点击「Save with Encoding」,会弹出编码选择,选择「UTF-8」后即可。

	SKU本地上传示例CSV文件.csv — Baidu_API				
			ርኃ	\triangleright	
ur	UTF-8 utf8	:地上传示	₹例CSV	と件.cs	
ed	UTF-8 with BOM utf8bom	age_typ	e		
	UTF-16 LE utf16le				
	UIF-16 BE UTIEDE				
.p					
64	-bit 🛞 0 🖄 0 📄 Server not selected Ln 3, Col 29 Spaces: 4 UTF	8 LF	Plain Text	:)	ন্দ



提示:将文档上传前,请确认UTF-8编码的文档内的文字没有乱码。

由于Excel的默认编码和操作系统以及软件版本有关,建议使用示例文件修改后上传,如果没有使用,又无法正常上传至平台,可以将源文件另存为CSV文件后,参考修改CSV文件的编码的方法转换成UTF-8的CSV文件进行上传。

の SKU标签组

创建标签组

直接创建标签组

在模型训练页面,点击左侧列表中的【我的SKU】进入SKU管理页面,点击【当前标签组】下拉列表中【+创建分组】创建空白标签组,创建成功 后,可在【当前标签组】下拉列表中显示新建的标签组。

我的SKU库						操作文档	常见问题	新手引导	提交工单
当前标签组:	全部SKU库 (3)		^						
创建SKU	+创建分组 全部SKU库 (3)					请输入SK	U名称、品牌	1、规格进行搜	索 Q
	;可口可乐公司(2)		SKU单品图数	被标注次数	状态	操作		
	饮用水	农夫山泉	550ml	60(系统预置60张)	64	新建	查看 编辑	諿 删除	
	可乐	可口可乐	300ml	120(系统预置120张)	0	新建	查看编	輯 删除	
	雪碧纤维	可口可乐	500ml	60(系统预置60张)	12	新建	查看 编辑	輯 删除	

提示:此时创建的标签组内没有SKU记录。

导入SKU时创建标签组

点击【当前标签组】下拉列表中【全部SKU库】,在SKU列表中左侧勾选SKU,再点击【导入标签组】。

我的SKU	车					操作文档	常见问题	新手引导	提交工单
当前标签组	1: 全部SKU库 (3)		~						
导入标签	^{签组} 批量删除	已选择 1 个SKU							
	SKU名称	品牌名称	规格参数	SKU单品图数	被标注次数	状态	操作		
	饮用水	农夫山泉	550ml	60(系统预置60张)	64	新建	查看编	辑 删除	
	可乐	可口可乐	300ml	120(系统预置120张)	0	新建	查看编	辑 删除	
	雪碧纤维	可口可乐	500ml	60(系统预置60张)	12	新建	查看 编	辑 删除	

如果您之前还没有创建过标签组,您会看到如下图所示的内容。填写标签组名称,点击【确定】,创建新的标签组并自动导入选中SKU记录到该 标签组,成功创建的标签组可以在【当前标签组】下拉列表中显示。

我的SKU库	_				操作文档	常见问题	新手引导	提交工单
	导入	标签组	×					
当前标签组: 全部SKU库 (3) 导入标签组 批量删除	标3	签组名称 *	0/8					
SKU名称	品牌名	确定	取消	被标注次数	状态	操作		
☑ 饮用水	农夫山泉	550ml	60(系统预置60张)	64	新建	查看编	緝 删除	
可乐	可口可乐	300ml	120(系统预置120张)	0	新建	查看编	諿 删除	
雪碧纤维	可口可乐	500ml	60(系统预置60张)	12	新建	查看编	緝 删除	

如果您之前已经创建过标签组,您会看到如下图所示的内容:

我的SKU库						
		导入标签组			×	
当前标签组:	全部SKU库 (35)					
		选择标签组 *	可口可乐公司 (6)	\sim	
	且	创建新分组(OFF			
	SKU名称					被标注次数
	可乐瓶装1		确定	取消		3
	可乐–瓶装	可口可乐	300ml	0		0

点击【创建新分组】状态栏为【ON】状态,再填写标签组名称,点击【确定】,创建新的标签组并自动导入选中SKU记录到该标签组,成功创建的标签组可以在【当前标签组】下拉列表中显示。

我的SKU库	_					操作文档	常见问题	新手引导	提交工具
	导入标签组			×					
销标签组: 全部SKU库 (3 导入标签组 批量删除) 选择标签组	★ 可口可乐公		~					
■ SKU名称	创建新分组	ON *			被标注次数	状态	操作		
☑ 饮用水	10/22 25 1210		(b) 0/0		64	新建	查看 绐	罐 删除	
可乐		确定	取消		0	新建	查看 练	輪 删除	
雪碧纤维	可口可乐	500ml	60(系统预置60张)		12	新建	查看 组	輯 删除	

编辑标签组

导入SKU

点击【当前标签组】下拉列表中【全部SKU库】,在SKU列表中左侧勾选SKU,再点击【导入标签组】。

我的SKL	J库					操作文档	常见问题	新手引导	提交工单
当前标签	组: 全部SKU库 (3)	9	~] 1						
导入标	送组 批量删除 3	已选择 2 个SKU							
•	SKU名称	品牌名称	规格参数	SKU单品图数	被标注次数	状态	操作		
	饮用水	农夫山泉	550ml	60(系统预置60张)	64	新建	查看 编辑	針 删除	
	可乐	可口可乐	300ml	120(系统预置120张)	0	新建	查看 编辑	員 删除	
	2 雪碧纤维	可口可乐	500ml	60(系统预置60张)	12	新建	查看 编辑	員 删除	

选择需要导入SKU的标签组,点击【确定】,完成导入,此时点击【当前标签组】下拉列表,选中想要查看的标签组,便可以浏览该标签组中的 SKU。

我的SKU库					操作文档	常见问题	新手引导	提交工单
	导入标签组							
当前标签组: 全部SKU库 (3) 导入标签组 批量删除	选择标签组 * 可口可乐公司 (0) ~ ~ 创建新分组							
SKU名称				被标注次数	状态	操作		
饮用水		确定	取消	64	新建	查看编	遺 删除	
☑ 可乐	可口可乐	300ml	120(系统预置120张)	0	新建	查看 编辑	遺 删除	
✓ 雪碧纤维	可口可乐	500ml	60(系统预置60张)	12	新建	查看编	遺 删除	

移除SKU

移除单个SKU

点击【当前标签组】下拉列表,选中一个标签组,在SKU列表中「操作」列点击【移除】。

我的SKUM	车					操作文档	常见问题	新手引导	提交工单
当前标签组	1: 可口可乐公司	(2)	~						
创建SKI	U本地批量上	传API上传]			请输入Sk	(U名称、品牌	、规格进行排	関索 Q
	SKU名称	品牌名称	规格参数	SKU单品图数	被标注次数	状态	操作		
	可乐	可口可乐	300ml	120(系统预置120张)	0	新建	查看 编	辑 移除	
	雪碧纤维	可口可乐	500ml	60(系统预置60张)	12	新建	查看 编	辑 移除	

确认后即可将该SKU从改标签组中移除,移除SKU并不会删除SKU库中的SKU,该SKU仍然会被保留在"全部SKU库"中。

我的SKU库									操作文档	常见问题	刻刻
当前标签组:	可口可乐公司 (6)		~								
创建SKU	本地批量上传	API上传	移除SKL	J		×			可乐		
	SKU名称	品牌名利		"可乐瓶装_可口可	∵乐_300ml"将从"`	可口可乐公司"中移除	被标注次数	状态	操作		
	可乐瓶装1	可口可分	J	该SKU仍保留在"全	部SKU库"中		3	新建	查看	编辑	移除
	可乐瓶装	可口可分				确定 取消	0	新建	查看	编辑	移除
	魔爪	可口可乐		1.34kg	0		4	正常	查看	编辑	移除

批量移除SKU

同样的,您也可以选中该标签组中一个或多个SKU的,再点击批量移除,实现对批量SKU从该标签组中移除的作用。移除SKU并不会删除SKU库中的SKU,移除的SKU仍然会被保留在"全部SKU库"中。

我的SKU国	Ē					操作文档	常见问题	新手引导	提交工单
当前标签组	: 可口可乐公司	(2)	~						
批量移時	已选择 1 个SP	KU							
•	SKU名称	品牌名称	规格参数	SKU单品图数	被标注次数	状态	操作		
	可乐	可口可乐	300ml	120(系统预置120张)	0	新建	查看编	辑 移除	
	雪碧纤维	可口可乐	500ml	60(系统预置60张)	12	新建	查看编	辑 移除	

重名名标签组

点击【当前标签组】下拉列表,点击【编辑】按钮修改标签组的名称。

我的SKU库				
当前标签组:	全部SKU库 (35)	^		
创建SKU	十 创建分组 全部SKU库 (35)			
	可口可乐公司 (6)	⊡ ₫	 马称	规格参数

点击【完成】即可完成编辑,如下图所示。

我的SKU库				
当前标签组:	全部SKU库 (35)	^		
导入标签组	+ 创建分组 全部SKU库 (35)			
•	可口可乐公司	完成 马称	规格参数	SKU单品图数
	可乐瓶装1	可口可乐	300ml	0

删除标签组

点击【当前标签组】下拉列表,点击【删除】按钮删除此标签组。

当前标签	组: 全部SKU库 (35)	^	
创建S	★创建分组		
	全部SKU库 (35)		
	可口可乐公司(6)	四回 名称	规格参数

确认后删除该标签组,删除SKU标签组后,该标签组下的SKU仍然会保留在"全部SKU库"中。

我的SKU库						
当前标签组:	全部SKU库 (35)	^				
创建SKU	+创建分组	删除分约	(日)		×	
- Bittering	全部SKU库 (35)	00100707	4			
	可口可乐公司 (6)		"可口可乐公司" 将被删	除		被标注次数
	可乐-瓶装1		该标签组下的SKU仍保留在"全部SKU库			3
	可乐-瓶装			确定 取消		0
	魔爪	可口可乐	1.34kg	0		4

⊙ SKU单品图数据要求

简介

训练一个定制商品检测模型需要准备两类数据:SKU单品图片和实景图片。本文档将详细介绍训练一个定制商品检测模型的数据要求,如规格、 大小、尺寸等,并提供相应的图片样例。另外,可以参考实景图标注规范文档文档内容,了解各个业务场景的数据如何正确标注。

点击下载数据采集与标注规范长图,一张图看懂如何采集和标注数据,让您不走弯路,获得一个高精度的商品检测模型。

SKU单品图指的是单个商品的图片,**不是模型训练必须的数据**,SKU单品图的作用是用来合成实景图,连同手工标注的实景图一起用于训练,降低实景图即训练数据采集和标注成本。

当每个SKU的实景图大于20张时,可以先不上传SKU单品图进行训练,后续提升模型以补充实景图为主,如果无法提供足量的实景图数据,可以通过上传SKU单品图来提升模型效果。

为了让模型能够完整地识别一个SKU,需要训练的图片中出现这个SKU的各个角度的样子,这意味着需要从实际业务场景中采集大量的图片,并 且进行大量的标注工作。为了降低这部分的成本,我们通过数据合成和增强技术,只需为SKU上传各个角度的单品图,且单品图无需进行任何标 注,即可让模型学习到这个SKU各个角度的样子。

平台上预置了近千个SKU,每个预置的SKU已匹配了50张左右各角度的单品图,绝大多数情况下无需再自行上传单品图,可根据训练结果补齐识别效果不好角度的单品图。

格式要求

图片的格式为:jpg、jpeg、png、bmp,图片大小不超过4M。

分辨率大小要求

图片分辨率大小建议达到1920×1080以上,拍摄相机像素200W以上,保证图片上的SKU清晰不模糊。

图片内容要求

项目	要求
图中SKU数量	图片上仅可出现一个SKU
背景	纯色,且背景颜色与SKU主体有区分度
角度	覆盖到实际检测场景中SKU可能出现的差异性较大的所有角度
光照	覆盖到实际检测场景中SKU可能出现的差异性较大的光照条件,比如灯光的颜色

下面举两个例子:

1. 当要检测的SKU是罐装可口可乐时,背景不可以是非纯色或是纯红色,如下图所示:



当业务场景是货架陈列审核,且货架上的商品无确定的展示面时,单品图需要覆盖到可能在货架上出现的所有差异性较大的角度,考虑到拍摄角度,一个SKU需要覆盖到水平视角、俯视视角和仰视视角,如下图所示:



各场景单品图推荐拍摄角度和上传图片数

拍摄角度和上传张数基本原则是覆盖实际检测场景可能出现的角度,请根据实际业务场景的情况灵活调整单品图的拍摄角度和上传图片数。

场景	推荐拍摄角度	推荐上传图片数
普通货架/货柜审核	水平视角、俯视视角和仰视视角	10 张
无人零售货柜	俯视视角	10 张
智能结算台	水平视角、俯视视角,可根据实际情况增加仰视视角	每个角度各10 张
地堆商品审核	可尽量在实景图中覆盖需要检测的角度	无需上传单品图

单品图片样例

以一个货架上瓶装商品为例,如下图所示:



使用API管理线上SKU数据

SKU的创建、删除、查询和上传SKU单品图均可以通过调用API实现,API使用方法请参考文档SKU管理API。

∞ 实景图上传

使用平台在线上传图片

SKU在货架上的实景图是模型训练需要用到的训练数据,需要客户从真实的业务场景中采集,这些图片在被正确标注中,可以用于训练成模型。

在完成SKU创建后,可在模型训练页面左侧列表中点击【实景图上传/标注】进入上传和标注页面,在上传前请在实景图集选择栏内创建实景图集,如下图所示

商品村	新建实景图集			×		
定制商品检 查,如铺货	实景图集名称:	安景图集名称: 可口可乐货架图集			商品计数,	辅助
实景图上传/标注	选择类型:	普通货架/货柜	~			
* 实景图集:		创建实景图集	取消			

需要填写的项目如下:

• 实景图集名称

实景图集的名称,可适当填入SKU细节,例如:原味可乐,番茄味薯片,奥运版纯牛奶等

• 选择类型

实景图集的类型,请与创建模型时选择的应用场景保持一致,上传时只上传跟选择类型相同的实景图。可选项为普通货架/货柜、智能结算 台、无人零售柜、地堆商品和其他

完成创建实景图集后,页面显示为如下图所示的内容

模型中心	实景图上传/标注		
我的模型	- 实景图集: 可口可乐货架图集	×	
创建模型			
训练模型	图片师选: 全部图片 木标注 已标注		
校验模型	标注示例		+ >
发布模型	+=:+=./pi		
数据中心	(1)小王小·四)		SKUMETHR Q
我的SKU库			╋ 新建SKU
实景图上传/标注			原味可乐_可口可乐_500ml 下 合
AI服务中心	1. 同一张图中需要标注全部想要识别的目标区域; 以"检测一张图中所有的可乐"为例	2.标注图片需要和实际场景一致; 以实际场景为识别出"用手机在零售货柜拍摄的图片"中	锁定后,后续框选目标会默认以该标 签命名
服务列表		有几层货架为例	
	✓ ★ ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●	小学校 1000 日本 1000 1000	
	保存		

点击页面上【标注】为该实景图集上传作为训练数据的实景图,点击【标注示例】右侧的加号上传实景图。

实景图基本要求如下:

实景图的详细采集要求,请参考实景图数据要求文档

- 实景图片需要是从真实业务场景中采集来的数据
- 支持上传的图片格式为jpg,png,jpeg,bmp,大小限制为4M
- 建议图片尺寸:最长不超过4096px,最小不低于30px,长宽比3:1以内

标注基本要求如下:

标注工具的使用方式,请参考实景图标注文档,实景图的具体标注要求,请参考实景图标注规范文档

- 完整并仅仅框选要识别的SKU
- 标注框不要框选到其它SKU或是价目标签等非要识别的SKU的干扰信息
- 在实景图中出现的所有要识别的SKU必须全部标注,不能遗漏

完成所有实景图的标注后,返回到【我的SKU库】可以查看到SKU列表中【实景图数】列显示标注了该SKU的实景图片的数量,如下图所示

模型中心	我的SKU库						
我的模型	创建SKU					请辩)入SKU名称、品牌、規格进行搜索 Q
创建模型							
训练模型	SKU名称	品牌名称	规格参数	SKU图数	实景图数	状态	操作
校验模型	原味可乐	可口可乐	500ml	12	12	正常	查看 编辑 删除
发布模型							
数据中心							
我的SKU库							
实景图上传/标注							
AI服务中心							
服务列表	您已创建了1个SKU					每页题	显示 10 🗸 < 1 >

使用API线下上传图片

实景图也可以使用线下标注工具标注好后,通过API上传至EasyDL零售版训练平台,也可以通过API创建和删除实景图集,API使用方法请参考文档实景图管理API。

心 实景图数据要求

简介

训练一个定制商品检测模型需要准备两类数据:SKU单品图片和实景图片。本文档将详细介绍训练一个定制商品检测模型的数据要求,如规格、 大小、尺寸等,并提供相应的图片样例。另外,可以参考实景图标注规范文档文档内容,了解各个业务场景的实景图片如何正确标注。

点击下载数据采集与标注规范长图,一张图看懂如何采集和标注数据,让您不走弯路,获得一个高精度的商品检测模型。

实景图指的是从业务场景中采集的图片,这些图片需要手动标注,只有标注的图片才会被用于训练,下面给出一些场景中的图片样例:

格式要求

图片的格式为:jpg、jpeg、png、bmp,图片大小不超过4M。

图片内容要求

上传标注的图片内容需要跟实际业务检测图片来源一致,比如货架上商品陈列审核业务,上传标注的图片是业务员巡店时拍摄的图片;无人货柜 业务,上传标注的图片是货柜里摄像头采集的实际投放时摆放了商品的图片;智能结算台业务,上传标注的图片是结算台日常结算时拍摄的图 片。

采集设备要求

采集设备推荐与实际业务中拍摄图片的设备一致。比如,智能结算台业务场景,采集设备推荐为结算台;无人货柜业务场景,采集设备推荐为货 柜;普通货架/货柜审核业务场景,采集设备推荐为手机。

分辨率大小要求

实景图中能够清晰看清每一个要识别的SKU,各场景的推荐图片分辨率如下:

场景	推荐图片分辨率
普通货架/货柜审核	1920×1440以上
地堆商品审核	1920×1440以上
无人零售货柜	1280×720以上
智能结算台	1280×720以上

以普通货架/货柜审核场景为例,如下图所示:



拍摄角度要求

在保证清晰度的前提下,**实景图采集时的拍摄角度建议与实际检测时保持一致**。普通货架/货柜审核场景需要注意,图片尽量从正面拍摄,角度 可以少量倾斜,但不要倾斜过大,如下图所示:



拍摄角度过大,以至增加商品特征识别难度

推荐上传标注图片数

在第一次训练时,建议每个SKU至少有20张实景图,上传的实景图,只有标注过的图片会被训练,所有训练的图片中,系统会随机抽取70%作为 训练集,剩余的30%作为测试集,如果标注的训练数据不足,可能会导致某个SKU的精确度远低于其它SKU,或是训练结果出现mAP、精确率、 召回率全都为0的情况。

第一次训练后,通过调取服务接口测试模型效果,根据测试结果,不断补足识别效果达不到的需求的SKU实景图,这个过程可参考模型优化方法 文档和使用模型优化工具处理识别效果不佳的实景图片。

实景图片样例

普通货架/货柜陈列审核场景



地堆商品陈列审核场景

OULINISI DRAFT TSINGTAG Street I i. TSINGTAG VG TSINGTAO TSINGTAG INGTA

无人货柜场景



智能结算台场景



使用API管理实景图数据

实景图也可以使用线下标注工具标注好后,通过API上传至EasyDL零售版训练平台,也可以通过API创建和删除实景图集,API使用方法请参考文 档实景图管理API。

ல SKU管理API

简介

- 本文档主要说EasyDL零售版的定制商品检测服务中的SKU管理API如何使用,如有疑问可以通过以下方式联系我们:
- 在百度智能云控制台内提交工单,咨询问题类型请选择人工智能服务
- 进入EasyDL社区交流,与其他开发者进行互动

• 加入EasyDL零售版官方QQ群 (群号:1009661589) 联系群管

接口鉴权

1. 进入EasyDL零售版的百度智能云控制台应用列表页面,如下图所示:

Û	全局 ~							
	✔ 返回EasyDL总览		产品服务 /	产品服务 / EasyDL定制训练平台 - 应用列表				
>	零售版		应用列表					
ଙ	应用列表		+ 创建应用					
⊕	用量统计							
	UP 449 477 Mile			应用名称		AppID	API Key	Secret Key
T	监控按衣		1	silence123		16935698	mbqkk0SDDHxrjUuBy2pwVCGO	****** 显示
<u>[4</u>]	定制商品检测服务	~						
no	标准布里检测肥多		2	Silence-EasyDL		1563/143	ghcLQC05Qsui22LEhjvt1b48	****** 显示
80	初始的回应规规分	Ť						
Ð	货架拼接	~						

2. 如果还未创建应用,请点击「创建应用」按钮进行创建。创建应用后,参考鉴权认证机制文档,使用API Key(AK)和Secret Key(SK)获取 access_token

SKU管理API概览

SKU管理API包含以下API:

接口名称	HTTP方法	API URL	说明
SKU创建	POST	https://aip.baidubce.com/rpc/2.0/easydl/retail/sku/create	用于创建SKU
SKU更新	POST	https://aip.baidubce.com/rpc/2.0/easydl/retail/sku/update	用于更新SKU
SKU列表	POST	https://aip.baidubce.com/rpc/2.0/easydl/retail/sku/list	用于列出所有SKU,可获得SKU的所有信息
SKU删除	POST	https://aip.baidubce.com/rpc/2.0/easydl/retail/sku/delete	用于删除指定SKU
SKU数据添加API	POST	https://aip.baidubce.com/rpc/2.0/easydl/retail/sku/addentity	用于为指定SKU上传单品图片

SKU创建API

接口描述

该接口可用于创建SKU

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/retail/sku/create

URL参数:

参数	值

access_token 通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

请求参数

字段	必选	类型	说明
sku_name	是	string	SKU名称,长度限制100个字符,支持中文、英文、数字和- & '.+ * () % / #
brand_nam e	是	string	品牌名称,长度限制100个字符,支持中文、英文、数字和- & '.+ *()% / #
specs	是	string	规格参数,长度限制100个字符,支持中文、英文、数字和- & ' . + * () % / #
category	是	string	商品品类:饮品 drink,药品 medicine,保健品 healthcare products,零食 snacks,香烟 cigarette,调味品 condiment,日用品 daily necessities,其它 other
package_ty pe	是	string	包装类型:瓶装 bottled,罐装 canned,袋装 bagged,盒装 boxed,桶装 barrel,其它 other
product_co de	否	string	商品编号,长度限制100个字符,支持英文、数字和- & '.+ * () % / #
package_i mage	否	string	图片数据,将图片转化为base64编码上传,要求图片大小不超过4M,最短边至少15px,最长边最大4096px,支 持jpg/png/bmp格式, 注意请去掉头部。

返回说明

返回参数

字段	必选	类型	说明
log_id	是	int	唯一的log id,用于问题定位
sku_id	是	int	创建的SKU ID

SKU更新API

接口描述

该接口可用于更新SKU

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/retail/sku/update

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

请求参数

字段	必选	类型	说明
sku_id	是	int	SKU的ID,SKU完成创建时刻获取,或可通过SKU列表接口查询
sku_na me	是	string	SKU名称,长度限制100个字符,支持中文、英文、数字和- & ' . + * () % / #
brand_ name	是	string	品牌名称,长度限制100个字符,支持中文、英文、数字和- & ' . + * ()% / #
specs	是	string	规格参数,长度限制100个字符,支持中文、英文、数字和- & ' . + * ()% / #
categor y	是	string	商品品类:饮品 drink,药品 medicine,保健品 healthcare products,零食 snacks,香烟 cigarette,调味品 condiment,日用品 daily necessities,其它 other
packag e_type	是	string	包装类型:瓶装 bottled,罐装 canned,袋装 bagged,盒装 boxed,桶装 barrel,其它 other
product _code	否	string	商品编号,长度限制100个字符,支持英文、数字和- & ' . + * () % / #
packag e_imag e_url	否	string	和package_image二选一,当package_image字段存在时,该字段输入失效,以package_image字段为准。图片大小 不超过4M,最短边至少15px,最长边最大4096px,支持jpg/png/bmp格式。通过URL上传SKU包装图时,请确保图片 链接有外网访问权限,否则图片将会上传失败,若需要补充SKU包装图,可使用SKU更新API。
packag e_imag e	否	string	和package_image二选一,当package_image字段存在时,以package_image字段为准 。图片数据,将图片转化为 base64编码上传,要求图片大小不超过4M,最短边至少15px,最长边最大4096px,支持jpg/png/bmp格式, 注意请 去掉头部。

返回说明

返回参数

字段	必选	类型	说明
log_id	是	int	唯一的log id,用于问题定位
sku_id	是	int	创建的SKU ID

SKU列表API

接口描述

该接口可用于获取已创建的SKU列表

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/retail/sku/list

URL参数:



access_token 通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值		
Contont Tuno	opplication /ico		

Content-Type application/json

请求参数

字段	必选	类型	说明
start	是	int	起始位置
num	是	int	结果数量,最大数量为100

返回说明

返回参数

字段	必选	类型	说明
log_id	是	int	唯一的log id,用于问题定位
total_num	是	int	返回结果数
results	是	array	返回结果
+sku_id	是	int	SKU ID
+sku_name	是	string	SKU名称
+brand_name	是	string	品牌名称
+specs	是	string	规格参数
+category	是	string	商品品类
+package_type	是	string	包装类型
+product_code	否	string	商品编号
+status	是	string	SKU状态:新建 new,上传SKU单品图片中 uploading,错误 error,正常 normal,训练中 training
+entity_count	是	int	SKU单品图数量

SKU删除API

接口描述

该接口可用于删除SKU

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/retail/sku/delete

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

请求参数

字段	必选	类型	说明
sku_id	是	string	SKU的ID,SKU完成创建时刻获取,或可通过SKU列表接口查询

返回说明

返回参数

字段	必选	类型	说明
log_id	是	int	唯一的log id,用于问题定位

SKU数据添加API

接口描述

该接口可用于为SKU上传单品图片
请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/retail/sku/addentity

URL参数:

access_token 通过API Key和Secret Key获取的access_token,鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

请求参数

字段	必选	类型	说明
sku_id	是	int	SKU ID
entity_content	是	string	SKU单品图的base64编码

返回说明

返回参数

字段	必选	类型	说明
log_id	是	int	唯一的log id,用于问题定位

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(1009661589)或工单联系技术支持 团队。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请通过QQ群(1009661589)或工单联系技术支持团 队。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请通过QQ群(1009661589)联系群管手动提额
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请通过QQ群(1009661589)联系群管手动提额
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请通过QQ群(1009661589)联系群管手动提额
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
406000	internal server error	服务错误
406001	param[xx] invalid	无效的参数xx,请检查相关参数
406002	dataset not exist	实景图集不存在
406003	dataset already exists	实景图集已存在
406004	dataset can not be modified temporarily	实景图集暂时不能被修改
406006	no permission to modify the dataset	没有修改实景图集的权限
406008	[xx] quota exceeded	xx配额超限
406009	sku does not exist	SKU不存在
406010	sku already exists	SKU已存在
406011	sku cannot be modified temporarily	SKU暂时不能被修改

の 实景图管理API

简介

本文档主要说EasyDL零售版的定制商品检测服务中的实景图管理API如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单,咨询问题类型请选择人工智能服务
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL零售版官方QQ群(群号:1009661589)联系群管

1. 进入EasyDL零售版的百度智能云控制台应用列表页面,如下图所示:

Ŷ	全局 🗸							
	≮ 返回EasyDL总	览	产品服务 /	≏品服务 / EasyDL定制训练平台 - 应用列表				
>	零售版		应用列	应用列表				
ଙ	应用列表		十创建应用					
⊕	用量统计			应用名称		AppID	API Key	Secret Key
T	监控报表		1	silence123		16935698	mbqkk0SDDHxrjUuBy2pwVCGO	******* 显示
<u>[4</u>]	定制商品检测服务	~	2	Silence-EasyDL		15637143	ghcLQC05QsuiZ2LEhjvtTb48	****** <u>显</u> 示
88	标准商品检测服务	~						
Ð	货架拼接	~						

2. 如果还未创建应用,请点击「创建应用」按钮进行创建。创建应用后,参考鉴权认证机制文档,使用API Key(AK)和Secret Key(SK)获取 access_token

实景图集管理API

实景图集管理API包含以下API:

接口名称	HTTP方 法	请求 Body	API URL	说明
实景图集创建	POST	JSON	https://aip.baidubce.com/rpc/2.0/easydl/retail/dataset/cr eate	用于创建实景图集
实景图集列表	POST	JSON	https://aip.baidubce.com/rpc/2.0/easydl/retail/dataset/li st	用于列出所有实景图集,可获得实景图集的所 有信息
实景图集删除	POST	JSON	https://aip.baidubce.com/rpc/2.0/easydl/retail/dataset/d elete	用于删除指定实景图集
实景数据添加 API	POST	JSON	https://aip.baidubce.com/rpc/2.0/easydl/retail/dataset/a ddentity	用于为指定实景图集上传实景图片

实景图集创建API

接口描述

该接口可用于创建实景图集

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/retail/dataset/create

URL参数:

参数 值 access_token 通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数 值 Content-Type application/json

请求参数

字段	必选	类型	说明
type	是	string	实景图集类型,SKU_DETECTION
dataset_ name	是	string	实景图集名称,长度不超过20个utf-8字符
scene	是	string	应用场景,不同场景对应内容为:普通货架/货柜 general shelf/container,智能结算台 smart self-checkout,无人 零售柜 smart vending machine,地堆商品 type genus,其它 other

返回说明

返回参数

字段	必选	类型	说明
log_id	是	int	唯一的log id,用于问题定位
dataset_id	是	int	创建的实景图集ID

实景图集列表API

接口描述

该接口可用于获取已创建的实景图集列表

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/retail/dataset/list

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数 值 Content-Type application/json

请求参数

字段	必选	类型	说明
start	是	int	起始位置
num	是	int	结果数量
type	是	string	实景图集类型,SKU_DETECTION

返回说明

返回参数

字段	必选	类型	说明
total_num	是	int	返回结果数
results	是	array	返回结果
+dataset_id	是	int	实景图集id
+dataset_name	是	string	实景图集名称
+type	是	string	实景图集类型
+status	是	string	实景图集状态:新建 new,上传实景图片中 uploading,错误 error,正常 normal,训练中 training
+special_status	是	string	特殊状态,商品检测模型值为空
+scene	是	string	实景图集场景

实景数据添加API

接口描述

该接口可用于为实景图集上传标注好的图片

请求说明

请求示例

HTTP 方法:POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/retail/dataset/addentity

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数 值 Content-Type application/json

请求参数

字段	必选	类型	说明
dataset_id	是	int	实景图集id
type	是	string	实景图集类型,SKU_DETECTION
entity_conte nt	是	string	实景图片的base64编码
entity_name	是	string	实景图片名称
appendLabe I	否	boolea n	确定添加标签的行为:追加(true)、替换(false)。默认为追加(true)
labels	否	array	SKU标签数据, 如果不传该参数,则上传的为无标注信息的图片
+label_nam e	否	string	SKU标签名称,请先参考 <mark>SKU创建文档</mark> 完成SKU创建,格式为:SKU名称_品牌名称_规格参数,例如:雪碧_可口 可乐_500ml
+left	否	int	标注框左上角到图片左边界的距离(像素)
+top	否	int	标注框左上角到图片上边界的距离(像素)
+width	否	int	标注框的宽度(像素)
+height	否	int	标注框的高度(像素)

返回说明

 字段
 必选
 类型
 说明

 log_id
 是
 int
 唯一的log id , 用于问题定位

实景图集删除API

接口描述

该接口可用于删除实景图集

请求说明

请求示例

HTTP 方法: POST

请求URL: https://aip.baidubce.com/rpc/2.0/easydl/retail/dataset/delete

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

请求参数

字段	必选	类型	说明
dataset_id	是	string	实景图集ID
type	是	string	实景图集类型,SKU_DETECTION

返回说明

返回参数

字段	必选	类型	说明
log_id	是	int	唯一的log id,用于问题定位

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
   "error_code": 110,
   "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(1009661589)或工单联系技术支持 团队。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请通过QQ群(1009661589)或工单联系技术支持团 队。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请通过QQ群(1009661589)联系群管手动提额
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请通过QQ群(1009661589)联系群管手动提额
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请通过QQ群(1009661589)联系群管手动提额
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
406000	internal server error	服务错误
406001	param[xx] invalid	无效的参数xx,请检查相关参数
406002	dataset not exist	实景图集不存在
406003	dataset already exists	实景图集已存在
406004	dataset can not be modified temporarily	实景图集暂时不能被修改
406006	no permission to modify the dataset	没有修改实景图集的权限
406008	[xx] quota exceeded	xx配额超限
406009	sku does not exist	SKU不存在
406010	sku already exists	SKU已存在
406011	sku cannot be modified temporarily	SKU暂时不能被修改

心 数据相关常见问题

为什么建议每个SKU至少出现在20张实景图中?

上传的实景图,只有标注过的图片会被训练,所有训练的图片中,系统会随机抽取70%作为训练集,剩余的30%作为测试集,如果标注的训练数 据不足,可能会导致某个SKU的精确度远低于其它SKU,或是训练结果出现mAP、精确率、召回率全都为0的情况。

SKU单品图是用来做什么的?

SKU单品图用来降低实景图即训练数据采集和标注成本的,为了让模型能够完整地识别一个SKU,需要训练的图片中出现这个SKU的各个角度的 样子,这意味着需要从实际业务场景中采集大量的图片,并且进行大量的标注工作。为了降低这部分的成本,我们通过数据合成和增强技术,只 需为SKU上传各个角度的单品图,且单品图无需进行任何标注,即可让模型学习到这个SKU各个角度的样子。由百度提供的SKU预置了50张左右 的单品图,绝大多数情况下无需再自行上传单品图。

SKU单品图需不需要标注?

SKU单品图不要标注,只需要参考「SKU单品图数据要求」文档采集并上传至相应的SKU即可。

SKU单品图和实景图分别是什么样的图片?

SKU单品图是单个商品的摆拍图,要求背景为纯色;实景图是商品在真实业务场景里的图片,比如商品在超市货架上的图片。具体两类图片的数据要求,请参考「SKU单品图数据要求」和「实景图数据要求」文档。

每个账号允许创建多少个SKU?

每个账号默认允许创建的SKU数量为50个,如果需要增加SKU数量,请加入官方QQ群(群号:1009661589)咨询解决。

数据标注

心 在线标注工具介绍

本文档介绍EasyDL零售版在线标注工具中各个功能的简介。

在线标注工具页面布局

可以使用EasyDL零售版在线标注工具进行标注,在线标注工具提供的多种功能提高标注效率,如下图图所示,下面会对Banner区域(红)、实景 图预览区(黄)、图片标注区(蓝)、SKU标签区(绿)四个区域的功能进行介绍。



在线标注工具介绍

Banner区域

く 返回悪鬼馬 全部(3)(6) 天和注意息(2)(6) 特徴(4)(00) 有标注意思(3)
快速整核 ① 辅助标注 ② 设置 > 标注初用 切発数単無 >

从左至右依次为:

1. 返回图集库

点击后返回实景图集库,如果当前图片的标注信息未保存,将会丢失

2. 全部

该Tab显示图集下的所有图片,包含无标注信息、待确认、有标注信息的所有图片

3. 无标注信息

该Tab显示图集下的所有未标注的所有图片,该Tab下的图片不会参与模型训练

4. 待确认

该Tab显示辅助标注功能预标注后的图片,该Tab下的图片不会参与模型训练,关于辅助标注的功能详情请见产品文档

5. 有标注信息

该Tab显示手动标注和在「待确认」Tab下确认后的图片,该Tab下的图片会参与模型训练

6. 快速复核

该功能可将「有标注信息」下图片上所有标注框,按照SKU分类后汇总展示,并支持修改或删除标注信息,可用于快速复核标注是否准确,关于快速复核的功能详情请见产品文档

7. 辅助标注

该功能可预先给未标注的实景图标上无标签的橙色辅助标注框和有标签的非橙色标注框,关于辅助标注的功能详情请见产品文档

8. 设置



- 标注辅助线,勾选后,标注时会有辅助线更精确地画标注框
- 预览包装,勾选后,鼠标HoverSKU标签或是标注框时,右侧标签栏会显示SKU包装图,SKU包装图
- 极速预览,勾选后,平台对1M以上单图自动进行压缩后展示,可以一定程度提升图片加载速度
- 8. 标注说明
 - 实景图上传API,可以通过API上传未标注和已标注的实景图,点击后跳转至该说明文档
 - 实景图集标注规范,详细说明实景图如何正确标注,点击后跳转至该说明文档
 - 标注工具使用方法,详细说明在线标注工具的使用方法,点击后跳转至该说明文档
- 9. 切换数据集

可在此处切换实景图集

实景图预览区



```
从左至右依次为:
```

- 按SKU筛选
 可以按照SKU标签筛选包含该SKU标注信息的实景图
- 2. 卡片视图

默认状态,仅在页面左侧显示实景图片缩略图,点击可以切换图片,**切换时不会自动保存该图片上标注信息的修改内容**,如需保存该图片上 的标注信息,可以通过「保存」按钮或是「左右切换」按钮进行保存

3. 拓展卡片视图

点击后,可以预览所有图片,该视图下可批量删除图片,或快速定位图片,切换图片时不会自动保存该图片上标注信息的修改内容

图片标注区



支持同时选择多个标注框,点击右侧标签后,为选中的多个标注框一次性附上标签

2. 单选标注

可在图片上进行画框标注,可选择一个已存在的标注框,点击右侧标签后,为选中的一个标注框一次性附上标签

3. 清除

可清除单张图片上的橙色辅助标注框

4. 拖动

可以拖动图片,配合放大、缩小工具使用调整图片显示

5. 删除

可删除当前显示的实景图片

6. 全屏

开启全屏标注

7. 缩小

将图片缩小显示

8. 放大

将图片放大显示

9. 保存

当实景图标标注或是已有标注发生变化时,可点击,保存当前图片的标注信息,快捷键S

10. 左右切换

上一张/下一张图片,**切换图片时会自动保存该图片上标注信息的修改内容**

SKU标签区

1. 切换标签组

选择标签组后,标签栏内仅显示该标签组包含的SKU,可用于缩小SKU标签显示范围,方便查找。标签组功能详情请见产品文档



2. 创建SKU

点击后,可在当前页面创建SKU

3. 搜索栏

可以搜索包含关键字的SKU,支持按SKU名称、品牌名称、规格参数进行搜索,方便找到常用的标注标签,可配合锁定和置顶工具使用

4. SKU标签列表

标签显示优先级:锁定的标签 > 置顶的标签 > 图上有标注框的标签 > 最新创建的标签,优先级越高,展示位置越靠前。 鼠标放在标签上停留几秒可以显示完整的标签名。



锁定

用于锁定该SKU标签,锁定后,单选标注和多选标注选择中的标注框都将以锁定的标签命名,锁定后的标签,底色为蓝色,位于标签栏 最上方,优先级高于「置顶」的标签

快捷键

位于标签列表的前10个标签,快捷键依次为1至0,在画完或选中标注框后,可通过键盘上快捷键快速给标注框选上标签

● 零度可乐_零度可乐_PET500ML*24	<u></u> 1	
● 零度可乐_零度可乐_PET1dot25L	:::: 2	
● 生姜可乐_生姜可乐_PET400ML	···· 3	
● 桃味可乐_桃味可乐_PET500ML	···· 4	
● 香草可乐_香草可乐_PET500ML	···· 5	
● 咖啡可乐_咖啡可乐_PET400ML	···· 6	
● 咖啡可乐_咖啡可乐(摩登罐)_CAN3	···· 7	
● 咖啡可乐_咖啡可乐(摩登罐)_CAN3	:::: 8	
● 零度可乐_零度可乐_PET2dot0L*6	···· 9	
● 零度可乐_零度可乐_PET2dot0L	:::: 0	

● 零度可乐_零度可乐_PET1dot25L*12

• KU13423_SKU1122_1123

快捷键总览

S键:保存当前标注内容

方向左键:切换至上一张图片,并保存当前图片标注内容

方向右键:切换至下一张图片,并保存当前图片标注内容

数字键1~9:给标注框打上对应键位的标签

₯ 实景图标注

实景图的标注工作可以在EasyDL零售版平台上完成,也可以在线下完成后上传到平台上,下面将分别介绍两种方法。

在线标注-辅助标注

辅助标注介绍

EasyDL零售版的在线标注工具提供了辅助标注功能,该功能可以使用平台预置模型和用户自训练定制模型预先为未标注的图片进行预标注,来降低整体标注工作的成本。在EasyDL零售版的实景图上传/标注页面可以使用该功能进行标注,如下图所示:



辅助标注功能提供两种预标注方式的选择,一种是通用检测模型,另一种是通用检测模型+自训练定制模型,如下图所示:

启动辅助标注		×

辅助画标注框功能可以预先给未标注的实景图标上标注框



启动 取消

启动辅助标注

选项1. 通用检测模型

选择「通用检测模型」可预先给未标注的实景图标上无标签的橙色辅助标注框,如下图所示:



注:只为需要识别的SKU的所有橙色辅助标注框附上标签,不需要识别的不处理不影响训练

在参考实景图上传文档将图片上传至平台后,在页面选择需要辅助标注的图片量,由于图片量越大,完成辅助标框所需时间越长,两个选项的相应建议如下:

- 在尝试第一次标注不同包装类型的商品时(如瓶装、袋装、盒装等),建议选该选项
- 需要立刻开始标注图片时,建议选该选项

全部图片 (图片越多,完成所需时间越长)

• 未标注的图片量很大,可以等待数小时后在进行标注时,建议选该选项

注:图片量和所需完成时间的关系大致为:所需时长=未标注图片量/100*5(分钟)

选项2. 通用检测模型+自训练定制模型

选择「通用检测模型」可预先给未标注的实景图标上无标签的橙色标注框和有标签的蓝色标注框,如下图所示:



此选项需要选择自训练定制模型的版本,建议选择mAP值高的版本。

注:支持选择2019年9月1日后完成训练的模型版本,建议选择mAP数值高的模型版本

启动辅助标注	×
辅助画标注框功能可以预先给未标注的实景图标上标注框	
通用检测模型 通用检测模型 + 自训练定制模型 为图中的商品标上橙色无标签标注框 为图中自训练模型可识别的商品标上蓝色有标签标注框	
	SKU标签列
	昌特惠4块装3+1_舒肤信 0000001050100003_
	0000001050100003_
选择标注模型: 27185: silencesecond V V4 - mAP: 56.73% V 2 支持选择2019年9月1日后 议选择mAP数值高的模型	与完成训练的模型版本,建 出版本
选择标注范围: 前100张图片(低于100张则标注全部图片量,可在5分钟左右完成) 全部图片(图片越多,完成所需时间越长)	000001080300073_ 3测试_小白测试_小白测 f、可口可乐_300ml
启动取消	fc_可口可乐_2L

点击启动后,页面如下图所示,后台开始根据所选择的图片范围标注图片,此时可以去往其它页面或是退出该页面,比如标注其它实景图集中的 图片,**但是只允许同时对一个图集启动辅助标框功能**。



过程中,可以点击「终止」按钮终止辅助标框进程,终止后图片将恢复至启动前状态。

完成辅助标注

辅助标注完成后,选择到相应启动辅助标注功能的图集后,显示如下页面:



点击「前往标注」按钮后,会进入到标注页面,使用辅助标注功能处理的图片会被归类到待确认的分类Tab下,如下图所示:



可以看到图片上会出现橙色的无标签标注框,如果选择了「通用检测模型+自训练定制模型」,则部分可识别的商品会被打上蓝色的有标签标注 框。

确认辅助标注图片

对于两类辅助标注的结果,需要做如下两个操作:

1. 标注橙色标注框

对于橙色无标签的标注框,标注方式为选中一个标注框,点击右侧标签进行标注。

注:只为需要识别的SKU的所有橙色辅助标注框附上标签,不需要识别的不处理不影响训练。

2. 核对蓝色标注框

核对标注好图片后,点击「确认并保存当前图片」或「切换并保存」按钮,该张图片会被保存到「有标注信息」Tab下。

标注栏工具也提供了多选标注功能,可以选中多个标签后进行标注,详情见「标注工具的使用」部分。

异常处理:还原所有未确认图片

如果发现很多图片上大量的辅助标注框位置都不准确,可以点击「还原所有未确认图片」按钮,按钮位置如下图所示,将所有还没有确认的图片 全部还原到「未标注」的图片分类Tab下,图片上的橙色辅助标注框也将被清空,这种情况建议不使用辅助标框功能,用一般标注进行标注,也 可以加入EasyDL零售版官方QQ群(群号1009661589),将不准确的情况反馈给我们,帮助我们优化提升该功能。

* 实景图集:	茉莉园超市 🗸 🗸			实景图上	传API 查看实景图集标	ī注规范	
							_
图片筛选:	全部图片(26)	未标注(0)	待确议	人(12)	已标注(14)	还原所有未确认图片	0
							_

在线标注-手动标注

对于辅助标注功能无法满足的图片,可以参考标注工具使用方法文档,使用在线标注工具手动标注。

线下标注

线下标注数据上传

如果您打算使用线下的标注工具(如标注精灵、labelme等)标注数据或是已有一些标注好的数据,可以参考实景图管理API文档将线下标注好的数据上传至EasyDL零售版训练平台。

⊙ 货架场景实景图标注规范

简介

数据的标注质量决定了模型的效果,本文档详细介绍各个场景采集的实景图应该如何标注。可以参考实景图数据要求文档,了解各个业务场景的 数据采集规范。

点击下载数据采集与标注规范长图,一张图看懂如何采集和标注数据,让您不走弯路,获得一个高精度的商品检测模型。

实景图标注原则

标注框的要求

标注框是标注时的最小单位,需要能够完全覆盖目标SKU的最小矩形框,如下图所示:



SKU被遮挡时的标注方法

无人零售货柜和智能结算台业务场景中,只标注商品露出来的部分。普通货架/货柜审核场景,建议只标注露出部分超过70%且具备识别特征的 SKU,如下图所示:



相似SKU需要标注

普通货架/货柜审核场景中,实景图上容易出现与目标SKU很相似的SKU,这些SKU也建议进行标注,能够降低模型将这些相似的SKU识别为目标 SKU的可能性。相似SKU分为以下几种情况:

1. 同品牌不同系列、口味、包装



2. 不同规格的同款产品



3. 过于相似的竞品



避免误标注和漏标注

误标注即将非目标SKU标注为目标SKU,漏标注即实景图上目标SKU没有全部被标注出来,标注的时候尽量避免这些情况发生,较多的误标注和 漏标注会严重影响模型效果。下图为漏标情况:



心 地堆场景实景图标注规范

简介

数据的标注质量决定了模型的效果,本文档详细介绍各个场景采集的实景图应该如何标注。可以参考实景图数据要求文档,了解各个业务场景的数据采集规范。

实景**图**标注原则

以下图举例,如果需要识别的是「青岛纯生」、「青岛白啤」、「青岛经典10度」三种箱装SKU。



正确的标注方式如下图所示:



1. 区分顶箱SKU和非顶箱SKU创建标签

需要创建如下四个SKU,【SKU名称】、【品牌名称】、【规格参数】通过【_】连接:

- 1) 箱装经典10度_青岛啤酒_500ml
- 2) 箱装经典10度-**顶箱_**青岛啤酒_500ml
- 3) 箱装白啤_青岛啤酒_500ml
- 4) 箱装白啤-顶箱_青岛啤酒_500ml
- 5) 箱装纯生_青岛啤酒_500ml
- 6) 箱装纯生-顶箱_青岛啤酒_500ml

注:【SKU名称】、【品牌名称】、【规格参数】中「顶箱」或「topbox」需用符号「-」连接在文字最后,如【原味可乐-顶箱】或【原味可 乐-topbox】

2. 区分顶箱SKU和非顶箱SKU标注

3. 完全覆盖目标SKU的最小矩形框

4. 避免误标注和漏标注

误标注即将非目标SKU标注为目标SKU,漏标注即实景图上目标SKU没有全部被标注出来,标注的时候尽量避免这些情况发生,**较多的误标注和 漏标注会严重影响模型效果**。

心 快速复核标注

本文档介绍EasyDL零售版在线标注工具中快速复核功能,在完成标注后,可以使用该功能对标注的图片进行复核,检查标注的是否正确。

该功能可将【有标注信息】下图片上所有标注框,按照SKU分类后汇总展示,并支持修改或删除标注信息,可用于快速复核标注是否准确。

使用流程

1. 功能入口

进入标注页面后,在下图所示位置,点击「快速复核」

く 返回图集库	全部[396] 无标注信息(296) 待确认(100) 有标注信息(3)	快速复核 ⑦	辅助标注 ⑦	设置 > 标注说明	切换数据集 ~
V: 80 88		к.я к.я — 16% +		全部SKU库 (339) 〜	创建SKU
				请输入标签名称	Q
A STATE OF A					

2. 启动快速复核

点击「快速复核」后即启动该功能,启动后需等待一段时间,**等待过程中,不可返回实景图标注和中断流程,可以通过新开页面的方式进行其他** 操作,如等待时间超过20分钟,可以提交工单进行反馈。

く 返回实景图标注							
按标签类别查看	选择标签 🗸			✔ 本页全选 👘	5 0 0	应用所有更改	放弃更改
		当前表	#据集复核任条创建由 请超等				
		⊐HJ¥.	•				

启动完成后,您将看到如下页面,每一个标注框将以小图的形式显示。

く 返回实景图标注									
按标签类别查看	选择标签	~				本页全选		应用所有更改	放弃更改
	€			S IS	5	2	R R R R R R R R R R R R R R R R R R R	9	茶 元 7 7 7
	1 1 1 1 1 1 1 1 1 1 1 1 1 1	本 本 元 13	14	1 5	If	T T T T	茶 元 7 7 18	(茶元 [19	茶 元 ン 20
*	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	本 元 23		2) 9299 225	26	· 茶元 · · · · · · · · · · · · · · · · · · ·		- - - - - - - - - -	κ π 30
a state		(茶江 33 33	林 元 34	本 元 メ 35	and the second s	100% (100%) (10%	100%	<u>π</u> 39	茶:
							<	1 2 3 >	40 条/页 ~

3. 复核标注

Step1. 选择需要复核的标签

可以滑动或是通过搜索框查找需要复核的SKU



选择SKU标签后,下方将仅显示这个SKU的小图(即该SKU标注过的每一个标注框内容)

按标签类别查看	茶兀柠檬红茶_农夫山泉	L_500ml ∨				本页全选		应用所有更改	放弃更改	全部SKU库 (339) > 创建SKU	
				请输入标签名称 Q)						
π	π.	茶元	なって	余元	本 元	π	茶元	× π	茶元	● 零度可乐_零度可乐_PET500ML*24 □□ 1	
	1		4	5		7 3 7	в 🖓	9	10	 * 零度可乐_零度可乐_PET1dot25L □ 2 	
				100						● 生姜可乐_生姜可乐_PET400ML 🔤 3	
			The second		X			*		• 桃脉可乐_桃脉可乐_PET500ML 👘 4	
π	iπ a	2 7	iπ >> •	π	π	177	Tr S	4 77 1	π 1 - 9	 ● 香草可乐_香草可乐_PET500ML □ 5 	
	12	13	-8 14	1 15	16	17	18	19	20	• 咖啡可乐_咖啡可乐_PET400ML 📖 6	
					X					 咖啡可乐_咖啡可乐(摩登罐)_CAN3 □ 7 	
1 to 1	1	1		· K	π					● 咖啡可乐_咖啡可乐(摩登罐)_CAN3 □ 8	
12.3		1		102						 零度可乐_零度可乐_PET2dot0L*6 □ 9 	
1	1 22	23	24	25	26					● 零度可乐_零度可乐_PET2dot0L □ □ 0	
										 零度可乐_零度可乐_PET1dot25L*12 	
										 卫龙312g大面筋香辣味_卫龙_312g*12袋 	
										 KU13423_SKU1122_1123 	
										• 挂条框_卫龙_0	
								< 1 >	40 条/页 ~	 冠能狗粮小型犬幼犬粮16磅泰迪贵宾法斗 	

鼠标悬停在小图上,可以显示该小图(标注框)的SKU标签名称

茶兀柠檬红茶_农夫山泉_500ml



也可以通过点击「放大」,查看大图,以查看更多商品的细节



Step2. 修正或删除标注

1) 修正标注

选中单个或是多个同类标注错误的图片

按标签类别查看 茶兀柠檬红茶_农夫山泉_500ml >>	已造择2个 取消选择 □本页金选 □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □	全部SKU库 (339) > 创建SKU
		 清総入街签名称 八月総入街签名称 八月総長4106g 0000000015010003_膨低_个
		 000000000150100003.重小豆(个 衣夫果園復芒30.衣夫山鹿,500ml 00000000015500035,毛毛虫(个 000000000150500073.呈果可請(个 小白那近(小白那近,小白那近,小白那近
		 町長、町口町長、300ml 町長、甲口町長、2. 町長、甲口町長、300ml 第二百槍県有花県、衣丸山泉、500ml 第二指着事具花花、衣丸山泉、500ml 第二指律事具花菜、大丸山泉、500ml
	< 1 > 40 条/页 \	 茶兀蜜桃乌龙茶,农夫山泉_500ml 茶兀蚊瑰器枝红茶,农夫山泉_500ml

在右侧标签栏选择正确SKU标签,选择后,图片会变为「已修改」状态



此时,在切换到修改后的SKU标签时,可以发现有修改过来的小图(标注框)

标注说明

く 返回実景图标注

按标签类别查看	茶兀玫瑰荔枝红茶_农夫	山泉_50 ∨		- 本页全选 り ご の	应用所有更改	放弃更改	全部SKU库 (339) ~ 创建SKU
							建築人務要求的 文 の の
					< 1 >	40 条/页 ~	● 茶兀玫瑰荔枝红茶_农夫山泉_600ml

2) 删除标注

如果发现误标注且不属于需要标注的小图,可以删除标注。选中单个或多个小图,点击删除按钮删除

く 返回实景图标注						删除标注框		
按标签类别查看	茶兀玫瑰荔枝红茶_农夫山	泉_50 ~		已选择1个	取消选择 本页全选	0 5 2 0	应用所有更改	放弃更改
(HE	1			茶元の	5 T		9	本 元 10
删除后,图	片的状态变为「	已删除状态」						
く 返回实景图标注								
按标签类别查看	茶兀玫瑰荔枝红茶_农夫山	l泉_50 ∨		已选择1个	取消选择 本页全选	0 5 0	应用所有更改	放弃更改
C the	1	a a a a a a a a a a a a a a a a a a a			s 7		3	一 (本) (10
3) 一些其它	已支持的操作							
	本页全选		5 2 6		应用所有	可更改	放弃	更改
	×π 7	来 、	8		本 元 9		茶 元 10	

从左至右分别是:

- 撤回上一步:取消上一步的操作(修改、删除)
- 重做上一步:重做上一步的操作(修改、删除)
- 还原本页所有操作:取消当前页上的所有操作(修改、删除)
- 4. 完成复核

应用所有更改

当复核完所有SKU时,可以点击「应用所有更改」完成复核工作。



ľ	确认提醒					7		0	
	(!)	确认应用ヌ 更改后,4	打"茉莉园超市"实 \$ \$次标签复核结果\$	景图集的 将覆盖等	的更改吗? 实景图集原 [?]	有标签,并	1 自动同步到	削实景图集	
			硑	定	取消]			

放弃所有更改

如果不想将复核所做的修改应用到实景图中,可以点击「放弃更改」或是「返回实景图标注」。

< 返回实	景图标注	ן								
按标签类别	別查看	茶兀玫瑰荔枝红茶	_农夫山泉_50 >				本页全选	0 5 0	应用所有更改	放弃更改
	τ τ τ τ	茶元	2	4	5 5		本 元 7		o the second	
确认后	,所	有修改的内]容将会丢失。							
	5						7			8
	确	认提醒	∃ ±							
	(!	确认放弃 放弃后,	卒对''茉 您更改	莉园起 文的内 [:]	3〕市''实 容将会	景图集 丢失	的更改	玛?	
				确定	Ξ	取氵	肖			

模型训练

心 模型训练操作说明

在完成创建模型和实景图上传和标注后,即可开始训练模型。在模型训练页面,点击左侧列表中的【训练模型】进入模型训练页面,您会看到如 下图展示的内容:

模型中心	模型训练								
我的模型	选择模型:	可口可口识别		~					
创建模型									
训练模型	添加SKU:	请添加要参与训练	东的SKU	~					
校验模型		序号	SKU名称	品牌名称	规格参数	操作	~	实景图集	SKU数量
发布模型		1	原味可乐	可口可乐	500ml	删除	~	可口可乐货架图集	1
数据中心									
我的SKU库									
实景图上传/标注									
AI服务中心									
服务列表									
		开始训练							

如上面图片所示,点击左侧列表中的【训练模型】,需要先后完成下面三项选择:

1. 选择要训练的模型

2. 选择需要想要模型支持检测的SKU,选择完成后,下方左侧会显示已添加的SKU,右侧会显示包含已添加SKU的实景图集

3. 选择要参与训练的实景图集

完成选择后,点击【开始训练】按钮页面跳转至【我的模型】页面,如下图所示,可以看到模型已进入训练状态,将鼠标移至状态"训练中"右边 的小问号上,可以查看训练进度,训练进度数值只是作为参考,所以推荐打开短信通知功能,这样就第一时间知晓模型训练完成了。

模型中心	模型列表							
我的模型	创建模型							
创建模型								
训练模型	【商品检测】可	コロ可口识別:	模型ID: 27257					
校验模型	应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作	
发布模型	云服务	V1	训练中 💷	未申请	未发布		停止训练	
数据中心		训练进	度:	11%				
我的SKU库		✓ 完成	成后短信提醒至 186*	***650 🗹				
实景图上传/标注								
AI服务中心								
服务列表								
							毎页显示 4 💙 < 1	>



模型中心	模型列表								
我的模型	创建模型								
创建模型									
训练模型	【商品检测】	可口可口识别	模型ID: 27257						
校验模型	应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作		
发布模型						mAP 95.83% ?			
数据中心	云服务	V1	训练完成	未申请	未发布	相端率 93.30% 2 召回率 92.31% ? 完整评估结果	申请发布 校验 训练		
我的SKU库									
实景图上传/标注									
AI服务中心									
服务列表									
							毎页显示 4 🖌 <	1	>

心训练时长与等待时间说明

EasyDL训练平台各类模型均是使用GPU集群进行训练,一个模型训练通常需要几十分钟至几个小时不等,在EasyDL零售版中,训练时长与参与训 练的SKU单品图和实景图片数量有关,下表为各种训练数据量级所需要的大致训练时间:

训练实景图片数	SKU是否上传单品图	训练平均时长
6000以上	是	10小时以上,24小时以内
6000以上	否	10小时以上,24小时以内
4000~6000	是	10小时左右
4000~6000	否	8小时左右
2000~4000	是	6小时左右
2000~4000	否	4小时左右
1000~2000	是	5小时左右
1000~2000	否	3小时左右
500~1000	是	4小时左右
500~1000	否	2小时左右
100~500	是	3小时左右
100~500	否	1小时左右
100以内	是	3小时左右
100以内	否	1小时以内

EasyDL 零售行业版使用说明

注:上述训练时长为多次训练的平均值,仅供参考,建议开启短信通知(如下图),以便模型训练完成时,我们能够第一时间通知到您。

模型中心	模型列表							
我的模型	创建模型							
创建模型								
训练模型	【商品检測】『	可口可口识别	模型ID: 27257					
校验模型	应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作	
发布模型	云服务	V1	训练中 🔢	未申请	未发布	-	停止训练	
数据中心		训练进	度:	11%				
我的SKU库		✓ 完月	或后短信提醒至 186*	**650 🗹				
实景图上传/标注								
AI服务中心								
服务列表								
							毎页显示 4 🗸 <	1 >

模型效果评估

心简介

在参考模型训练操作说明文档完成模型训练后,可参考此文档了解模型效果。

心 模型训练结果

模型的训练结果是如何得到的?

上传的实景图,只有标注过的图片会被训练,所有训练图片中,**系统会随机抽取70%的标注数据作为训练数据,剩余的30%作为测试数据**,训练 数据训练出的模型去对测试数据进行检测,检测得到的结果跟人为标注的结果进行比对,得到页面显示的mAP,精确率和召回率。

提示:训练数据,即上传标注的实景图片越接近真实业务里需要预测的图片,模型训练结果越具有参考性。

◎ 模型版本详情

模型列表

BDERE								
【商品检测】李锦记 #	遺型ID:8853	38				〓 训练	⑤ 历史版本	面 删除
部署方式	版本	训练状态	申请状态	服务状态	SKU F1-score 分布(② 操作		
公有云API	V2	训练完成	未申请	未发布	[85%, 100%] 4个 [60%, 85%) 12个 [0%, 60%) 5个	查看版本详情	申请发布 校验	2

模型训练好后,可以在模型列表中看到SKU的F1-score分布情况,如果需要了解更为详细的模型效果表现,可以在模型列表中点击「**查看版本详 情**」。

基础信息

模型列表 > 李锦记 > V2

基础信息

```
模型ID 88538 版本 V2 图片数 85 SKU数 21
```

```
训练完成时间: 2020-12-14 18:50 训练算法: 公有云API--默认算法
```

模型版本的基础信息包含以下内容:

- 模型ID
- 训练版本
- 训练图片数
- 训练SKU数
- 训练完成时间
- 训练算法

SKU F1-score分布

SKU F1-score分布

4个		12个	5个	
100%	85%	60%		0%

直观展示模型中SKU F1-score的分布,F1-score是模型中一个SKU的精确率和召回率的调和平均数,可以作为判断模型中各SKU效果的指标,通常情况下:

- F1-score>85%时,可满足商品计数需求
- F1-score>60%时,可满足统计商品分销需求

模型整体效果

模型整体效果



页面上显示的分销准确率、mAP、召回率和精确率数值,是模型里所有SKU在建议阈值下的平均值,建议阈值可以在模型的「模型整体F1-score 走势图」中查看。四项指标的含义分别为:

• 分销准确率

按图片粒度统计分销准确率,即单张评估图片中正确识别出所有SKU种类的平均概率。

分销准确率 = 一张图片内正确识别的SKU种类数/(人工标注出的SKU种类数 U 模型识别出的SKU种类数)

• mAP

mAP在[0,1]区间,越接近1模型效果越好,mAP不高也不说明模型里所有的SKU识别效果不好。

mAP(mean average precision)是物体检测(Object Detection)算法中衡量算法效果的指标。对于物体检测任务,每一类object都可以计算出其精确 率(Precision)和召回率(Recall),在不同阈值下多次计算/试验,每个类都可以得到一条P-R曲线,曲线下的面积就是average precision(AP)的值。 "mean"的意思是对每个类的AP再求平均,得到的就是mAP的值。

• 精确率

对于一个SKU而言,精确率越高,说明模型识别出是这个SKU的所有结果中,正确数量的占比越高。如果精确率为1,说明识别出的所有结果都是 对的,但不说明该SKU全部都被识别出来了,可能会存在漏识别。

```
精确率 Precision = 模型正确预测为该SKU的数量/模型预测为该SKU的总数
```


对于一个SKU而言,召回率越高,说明模型越完整地识别出这个SKU。如果召回率为1,说明这个SKU全部都被模型识别出来了,但不表示识别出 是这个SKU的结果都是对的,可能会存在误识别。

召回率 Recall = 模型正确预测为该SKU的数量/SKU客观存在的总数

模型整体F1-score走势图

F1-score是模型中一个SKU的精确率和召回率的调和平均数,在以相同权重考虑precision和recall的情况下,用来衡量一个模型的效果。

F1-score = 2*Precision*Recall/(Precision+Recall)

模型整体F1-score走势图

不同 阈值 下 F1-score 表现



该曲线图展示了模型中各SKU在不同阈值下的F1-score平均值,根据该曲线可以得到阈值的最优值,即图中显示的「**建议阈值**」。模型列表和整体评估展示的三项模型效果指标数据,均是模型在「**建议阈值**」下的结果。另外,在模型发布后,调用服务API时可通过请求参数中threshold调 节阈值,默认为建议阈值。

阈值(threshold),是正确结果的判定标准,例如阈值是0.6,置信度大于0.6的识别结果会被当作正确结果返回。

训练及评估数据明细

训练及评估数	女据明 细					$\overline{1}$	全部明细下载
No.	名称	训练图片数	训练标注框数	评估图片数	评估标注框数	F1-score	操作
1	薄盐生抽_李锦记_500ml	30	159	13	26	92.86%	查看详情
2	薄盐生抽_李锦记_1.75L	9	19	1	3	66.67%	查看详情
3	薄盐醇味鲜酱油_李锦记_500ml	9	21	6	-11	95.65%	查看详情
4	凉拌汁_李锦记_207ml	3	8	2	6	50.00%	查看详情
5	蒸鱼豉油_李锦记_410ml	8	23	3	10	37.84%	查看详情
				每页显示	5 🗸 🕻 1	2 3 4	4 5 >

由于训练过程中,**系统会随机抽取70%的标注数据作为训练数据,剩余的30%作为测试数据,所以模型训练存在一定的随机性**。明细数据列表中 展示了该模型版本训练时的图片数据和标注框的分布情况,帮助用户更具针对性的分析各个SKU的指标对应的数据量,以便针对性补充训练数据 来优化模型。该表单支持下载,以便用作模型报告的制作。表单中的各列含义如下:



- 名称:SKU的标签名称,名称品牌规格
- 训练图片数:实际用于训练模型的各个SKU的图片数量(训练模型时选择所有图集中含有图片总量的70%)
- 训练标注框数:实际用于训练模型的图片中,各个SKU的标注框数
- 评估图片数:用于评估个SKU训练效果的图片数量(训练模型时选择所有图集中含有图片总量的30%),即用来得到模型各项指标(mAP、 Precision、Recall、F1-score)的评估集
- 评估标注框数:用于评估的图片中,各个SKU的标注框数
- F1-score:每个SKU的F1-score,F1-score是模型中一个SKU的精确率和召回率的调和平均数,可以作为判断模型中各SKU效果的指标,通常情况下:
 - F1-score>85%时,可满足商品计数需求
 - F1-score>60%时,可满足统计商品分销需求
- 操作:查看详情,可查看所有参与训练和评估的图片

模型优化

心 优化方法

简介

模型训练好后,必经的一个过程是模型优化,EasyDL零售版已专门根据零售业务场景中的数据调优了模型算法,所以优化EasyDL零售版训练的 模型,不需要理解和调优模型算法中的各种参数,仅需要优化训练数据即可。优化一个EasyDL零售版的商品检测模型,可以分为以下几个步骤进 行:

- 1. 前提 正确采集实景图和单品图,并正确标注实景图
- 2. 补充实景图 使用EasyDL零售版提供的模型优化工具 (云服务数据回流) 补充实景图
- 3. 补充SKU单品图 上传SKU识别不好的角度的单品图
- 4. 重新训练模型 补充好数据后用新旧数据一起重新训练模型
- 5. 重新发布模型 将新训练的模型版本发布为API后测试模型效果
- 6. 重复优化 根据测试结果重复2-5步骤直到模型效果可商用

1. 前提

数据质量是保证模型效果的前提,在EasyDL零售版中,数据质量涉及实景图和单品图的图片质量,以及实景图的标注质量,**开始模型优化前,请** 先学习如何采集合格的图片和进行合格的标注,各个参考文档如下:

- 实景图采集:实景图数据要求
- 实景图标注:实景图标注规范
- SKU单品图:SKU单品图数据要求

点击下载数据采集与标注规范长图,一张图看懂如何采集和标注数据,让您不走弯路,获得一个高精度的商品检测模型。

2. 补充实景图

推荐使用EasyDL零售版提供的模型优化工具 - 云服务数据回流工具,参考使用文档优化工具补充实景图来优化模型。

3. 补充SKU单品图

在上一步使用云服务数据回流工具时,发现SKU识别效果较差的角度,参考文档SKU单品图数据要求采集SKU相应角度的单品图并上传到对应 SKU。

4. 重新训练

在补充了相应数据后,在模型列表中「操作」列中点击「训练」,选择需要检测的SKU和所有需要训练的实景图集,包括上一次训练的图集和新 增的图集(如果优化模型的数据是存在新的实景图集中),确认选择无误后开始训练。 提示:重新训练之前训练或发布过的模型,训练过程中和训练后均不会影响之前训练的模型版本,训练之前发布的版本API接口依旧有效,也可以选择发布之前训练过的其他模型版本。

同一个模型所有训练的版本均可以在模型列表中进行查看,如下图所示:

模型中心	模型列表						
我的模型	创建模型						
创建模型							
训练模型	【商品检测】饮品	品检测 模型ID	27899				🖴 训练 🕒 历史版本 🔟 删除
校验模型	部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作
发布模型 数据中心 我的SKU库	公有云API	V1	训练完成	审核成功	已发布	mAP: 96.95% ? 精确率: 88.46% ? 召回率: 92.00% ? 完整评估结果	查看版本详情 配置服务功能 校验 体验H5 ⑦
实景图上传/标注	【商品检测】饮料	斗检测 模型ID	27888				😑 训练 🕓 历史版本 🔟 删除
云服务调用数据	部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作
AI服务中心 服务列表 增值服务中心	公有云API	V2	训练完成	未申请	未发布	mAP: 93.70% ? 精确率: 83.33% ? 召回率: 80.00% ? 完整评估结果	查看版本详情 申请发布 校验
货架拼接服务							每页显示 4 💙 < 1 >

在模型列表中点击「全部版本」图标后进入到模型全部版本管理页面,如下图所示,可在页面上查看所有版本的训练集和选择发布任意已经成功 训练好的版本。

模型中心	模型列表	> 饮料检测				
我的模型	■ 饮料检测	全部版本				
创建模型	版本	训练状态	申请状态	服务状态	模型效果	操作
训练模型					mAP: 93.70% ?	
校验模型	V2	训练完成	未申请	未发布	補頭率: 83.33% (7) 召回率: 80.00% (7) 完整评估结果	查看版本详情 申请发布 校验
发布模型						
数据中心 我的SKU库	V1	训练完成	未申请	未发布	mAP: 95.72% ⑦ 精确率: 85.19% ⑦ 召回率: 92.00% ⑦ 完整评估结果	查看版本详情 申请发布 校验

5. 重新发布模型

模型重新训练好后,如果之前发布过的模型版本还在测试,可以直接重新发布,用新的版本执行下一步;如果之前发布过的模型版本已经上线到 生产环境,请参考下面**注意**内容,确保不影响线上生产环境的情况下,执行重新发布。

注意:重新发布模型后,服务API URL不会变化,模型自动切换为新发布的训练版本,如果服务API已上线生产环境,发布新版本前,请先在 模型列表和完整评估报告中,确认新版本的模型指标和各SKU的精确度与已上线版本是更优或者相差不大的,以确保新发布版本的模型效果 不会对线上业务不会产生不好的影响。

6. 重复优化

根据每次优化数据后的模型测试结果,重复2-5步骤直到模型效果可商用。

心 优化工具

简介

本文档介绍如何使用云服务数据回流功能来优化商品检测模型。

云服务调用数据管理

EasyDL零售版云服务数据回流功能,可用于查找云服务模型识别错误的数据,纠正结果并将其加入模型迭代的训练集,实现训练数据的持续丰富 和模型效果的持续优化。

提示:模型发布成功后,才可以开通该功能,如果模型还未发布,可以参考文档模型发布发布一个模型。

该功能的使用步骤如下:

- 1. 开通功能
- 2. 筛选数据
- 3. 修正标注
- 4. 优化模型

步骤1. 开通功能

开通此项功能有两种方式,一是在发布模型页面,发布模型时勾选「同意云服务调用数据管理服务条款并开通服务」,发布后即可开通这项功能;二是在左侧导航栏「数据中心」点击「云服务调用数据」,在页面上选择**已发布**的定制商品检测模型接口,选择后仔细阅读服务条款,接受 后即可开通这项功能。如下两张图所示:

模型中心	云服务调用数据管理									
我的模型	选择镀囗: silencefirst0829 ✓									
创建模型	此接口未开通EasyDL云服务调用数据管理,是否开递? >> 功能介绍									
训练模型										
校验模型	开通"EasyDL云服务调用数据管理"(以下简称"本功能"),可查找云服务模型识别错误的数据,纠正结果并将其加入模型这代的训练集,实现训练数据的持续丰富和模型效果的持续优化。									
数据中心	 为提供本功能之目的,您通过接口上传的数据会存储在百度EasyDL服务器中,数据将从开通本功能次日开始存储。最多存储30天的数据,超过30天的数据将自动删除。 									
#MoVUE	2. 百度仅提供数据存储空间服务,该数据仅用于向您提供本功能服务,不会用于其他目的。									
实景图上传/标注	3. 您使用本功能,应当保证在开通服务期间不得允许任何其他第三方使用以上接口,并保证对所上传的数据享有合法权益,有权进行上述处理并承担全部法律责任。									
云服务调用数据	您开通本服务即视为您同意百度以上述方式存储您上传的数据。									
AI服务中心	▶ 我已仔细间读并接受以上条数									
服务列表										
增值服务中心										
货架拼接服务										
模型中心	「云服务调用数据」页面									
我的模型	后准指门担防急者									
创建模型	这排模型: 饮料检测 ✓ 标准接口请求参考说明:									
训练模型	应用类型: 云服务 ✔									
******	送择版本: ∨1 ∨ 7*********************************									
牧 牧 短 侯 型	ElistAtiin United and Attick ElisAtiin United Attick ElisAtiin United Attick ElisAtiin United Attick Inage E String Stri									
数据中心	detection/ 网络默认为当前模型推荐网络 threshold 否 number (0.171m) 和体值可以在转动									
我的SKU库	其他要求: 若接口无法满足您的需求,请描述希望解决的问题。 极望列表-极型效果查看 500以字中以内									
实景图上传/标注	标准接口响应字段说明:									

✓ 同意云服务调用数据管理服务条款并开通服务

服务开通后,次日服务开始生效,生效后接口调用的数据可在次日通过网页查看。

举例:如果您在2019年8月29日开通该功能,该功能将于2019年8月30日生效,如果在30日调用过接口,那么30日使用该接口识别过的图片, 将在31日0点后可以在网页上查看到。

「发布模型」页面

字段名称 必须 类型

是

error_code 否

error_msg 否

+ name 否

number

number

string

results 否 array(object) SKU结果数组

string

log_id

说明

SKU名称

唯一的log id,用于问题定位

错误码,当请求错误时返回

错误描述信息,当请求错误时返

步骤2. 筛选数据

云服务调用数据

AI服务中心

服务列表

增值服务中心

数据查看说明

货架拼接服务

模型中心	云服务调用数据管理	服务条款
我的模型	选择接口: silenceteapi V 对应模型: silencesecond 对应版本: V1	
创建模型		
训练模型	调用时间: 2019/8/2 - 2019/8/28 🎁 😨	
校验模型	选择SKU: 茶兀柚子绿茶_农夫山泉_500ml 💙	
发布模型	筛选条件: 💿 图片中识别出该标签,该标签的置信度介于 0 % ~ 60 %	
数据中心	○ 图片中识别出的该标签数量	
我的SKU库	监看筛选结果	
实景图上传/标注	筛选结果	
云服务调用数据	按调用时间排序 ✓ 批量率	重新标注
AI服务中心		
服务列表		
增值服务中心		
货架拼接服务		

如上图所示,选择想要查询的接口调用时间和要筛选的SKU标签,筛选条件可以根据模型的阈值和业务的情况而定,筛选出的图片会显示在页面上。举个例子,比如调用接口识别图片时,设置的阈值(threshold)是0.6,业务上认为置信度达到80%以上才是可以接受的结果,那么这里置信度填写的标签应该是0~80%。这样,所有调用识别的图片中,含有该SKU标签且置信度在60%~80%的图片就会被筛选出显示在页面上。

阈值(threshold),是正确结果的判定标准,例如阈值是0.6,置信度大于0.6的识别结果会被当作正确结果返回。在调用接口时,可以通过参数「threshold」设定,如果不填,则默认设置为推荐阈值,推荐阈值可以在「我的模型」页模型的「完整评估结果」里查看。

步骤3. 修正标注

		选择接口:	silenceteapi	\sim	对应模型: silencesecond	对应版本	:: V1			
创建模型	查看识别结果								×	
训练模型										
校验模型							预测SKU	置信度		
发布模型			- maan	12			1-茶兀柚子绿茶_农夫	73.96%		
数据中心				A			2-茶兀柚子绿茶_农夫	64.20%		
我的SKU库							3-茶兀柚子绿茶_农夫	59.58%		
实景图上传/		000	aa I II	D						
云服务调用数				in is						批量重新标注
AI服务中心		153	I I I I I I I I I I I I I I I I I I I							
服务列表			HQ	7						Distant and the
增值服务中心		REP		1						Real
货架拼接服务										
								重新标注	取消	

如上图所示,选择一张图片点击查看,可以看到图中三个识别结果都是误识别,点击「重新标注」后,在弹窗内选择需要将该图片添加到的实景 图集,如下图所示:

选择接口:	silenceteapi	\sim	对应模型:	silencesecond	对应版本:	V1
调用时间:	确认提示					×
选择SKU:	您可以将错误的图》 优化效果。	└保存	在实景图集中,方便	后续在数据标注—未标	注图片中重新标	注,并继续训练
筛选条件:	选择实景图集:		忙化0827 查询实景图集			新增实景图集
		17	忙化0827			取消
		汃	则试22			-17.74
		te	est2		- 1	
	■ 师选结果	+/	act		_	

0827」,表示用于放8月27日优化模型的图片,这样便于标注和在训练模型时直接勾选上这个实景图集。点击「提交」后,可以点击实景图集的 名称立刻跳转去标注页面,也可以点击「继续处理数据」留在该页面继续处理其它的图片,如下图所示:



步骤4. 迭代模型

将所有需要修正标注的图片都标注好后,去到我的模型页,在需要优化的模型表里,点击「训练」,如下图所示:

模型中心	模型列表								
我的模型	创建模型								
创建模型									
训练模型	【商品检测】	次品检测 模型	빌D: 27899						İİ
校验模型	应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作		
发布模型						mAP 96.95% ?			
数据中心	云服务	V1	训练完成	审核成功	已发布	有咖里 88.46% [7] 召回率 92.00% ? 完整评估结果	服务详情校验训练本验	H5 🤉	
我的SKU库									
实景图上传/标注	【商品检测】	次料检测 模型	빌D: 27888						Ū
云服务调用数据	应用类型	版本	训练状态	申请状态	服务状态	模型效果	操作		
AI服务中心						mAP 95.72% ?			
服务列表	云服务	V1	训练完成	未申请	未发布	有端率 83.19% [7] 召回率 92.00% [7] 完整评估结果	申请发布 校验 训练		
增值服务中心 货架拼接服务							每页显示 4 🗸 <	1	>

点击后会跳转至训练页面,如下图所示,新增勾选添加了这部分修正标注的图片的图集即可,比如如果是将这些图片放在实景图集「优化 0827」,那么勾选最初训练这个模型的实景图集的同时勾选这个新的图集,确认选择无误后,点击开始训练,训练完成后即完成了一次模型优 化。

模型中心	模型训练							
我的模型	洗择樺型:	牧品检测		~				
创建模型		10 ANN 122.00						
训练模型	添加SKU:	请添加要参与训	练的SKU	~	2			
校验模型		序号	SKU名称	品牌名称	规格参数	操作	实景图集	SKU数量
发布模型		1	茶兀柚子绿茶	农夫山泉	500ml	删除	可口可乐货架图集	3
数据中心		2	雪碧纤维	可口可乐	500ml	删除	优化0827	1
我的SKU库		3	可乐原味	可口可乐	500ml	删除		
实景图上传/标注								
云服务调用数据								
AI服务中心								
服务列表								
增值服务中心								
货架拼接服务		开始训练						

模型发布

在完成模型训练后,即可将模型发布为云服务API,发布成功后调用API即可获得模型支持的商品检测能力。模型训练完成后,可以在我的模型列表中发起模型上线申请,如下图所示:

【商品检测】饮料检测 模型ID: 27888 🛛 🔤 训练 🕓 历史版本 🔟									
部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作			
公有云API	V2	训练完成	未申请	未发布	mAP: 93.70% ? 精确率: 83.33% ? 召回率: 80.00% ? 完整评估结果	查看版本详情中请发布校验			

点击模型列表内对应模型「操作」列中的「申请发布」,或是在左侧导航栏点击「发布模型」可以进入发布模型页面,如下图所示:

模型中心	发布模型
我的模型 创建模型	1 选择模型: 饮品检测 标准接口规范参考 标准接口演求参考说明:
训练模型	2 选择版本: V1 ▼ 3 - 服务名称: kelexuebijiance
校验模型发布模型	4 * 接口地址: https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/ image 是 string 检验64编码方小不超过4M, 最短空型小50x,最长边量大 detection/ kelexuebixianwei
数据中心	5 其他要求: 可乐和雪碧纤维检测 面值,默认为当前搜型推荐阈值 (0-1之间),具体值可以在我的 提型两序,模型改革者
我的SKU库 实景图上传/标注	0/500 标准接口购应字段说明:
	接交申请 学段各称 必须 类型 说明
AI服务中心	如果有私有化認書需求,请点此申请 log_id 是 number 唯一的log id,用于问题定位
服务列表	error_code 否 number 错误码,当请求错误时返回

在对应选项中选择和输入相应内容发起模型发布的申请:

1. 选择模型 (必选)

选择需要发布的模型,只能选择已经完成训练的模型

2. 选择版本 (必选)

选择需要发布的模型版本,只能选择完成训练且没有发布过的版本

3. 服务名称 (必填)

为发布的服务命名,**服务名称不得多于20个字符**

4. 接口地址 (必填)

自定义服务的API URL,接口地址需要多于5个字符但不能超过20个字符,仅限英文

5. 其他要求

如果有其他要求可以输入要求描述

填写完上述信息后,点击「提交申请」完成发布模型申请。提交申请后,模型列表内该模型的申请状态和服务状态为有以下几种情况:

申请状态	服务状态	状态描述
审核中	未发布	服务刚申请发布,模型在审核中, 申请发布后,会自动通过审核
审核成功	发布中	服务通过审核,进入系统自动发布阶段, 约5分钟左右完成发布
审核成功	已发布	服务发布成功

发布成功后,可以点击模型列表内「操作」列中的「配置服务功能」,如下图:

模型中心	模型列表								
我的模型	创建模型								
创建模型									
训练模型	【商品检测】饮品	·检测 模型ID:	吕 训练	④ 历史版本	面 删除				
校验模型	部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作		
发布模型						mAP : 96.95% 🖓	_	_	
数据中心	公有云API	V1	训练完成	审核成功	已发布	精備率: 88.46% ? 召回率: 92.00% ? 完整评估结果	查看版本详情 配置	服务功能校验	体验H5 🛛

点击后弹出下图所示窗口,可以获取模型的云服务API URL。

我的模型 > 1111V1的公有云API服务详情

服务名称:	111	服务功能描述: 模型服务接口使用方法请师 <mark>API文档</mark>					
快至成本·		服务功能	接口返回字段	内容			
接口地址:	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/detection/silence02		name	商品标签,包含商品名 称、品牌、规格			
服务功能:	已选功能不同,单次接口调用价格不同,详细收费方式请见 计费文档	大口地上行向いつい	sku_code	商品编号			
	基础功能服务	尚而基本信息识别	score	识别结果的置信度			
	商品基本信息识别 商品标签(名称、品牌、规格)、编号、坐标、置信度		location	商品检测框在图片上的像 素坐标			
	可选服务功能			图片中包含的陈列场景类 型,支持货架、端架、立 式冰柜、地堆和割箱			
	商品陈列层数识别 商品所在倍望层数及传望总层数、支持区分不同货架	商品陈列场景识别	scenes				
			scene	每个商品所在的陈列场景			
	商品跡列场景识别 支持五类场景: 货架、端架、立式冰柜、地堆和割箱		shelf	商品所在的货架编号,从 左往右依次递增			
	商品排面占比统计 商品排面数/占比、未识别商品数、空位数及货架利用率	商品陈列层数识别	layer	商品所在货架层数编号, 从上往下依次递增			
	提示:开启或关闭功能后,约5分钟后生效		layer_count	统计各货架的总层数			
	立即使用 返回		layer_top	判断货架最顶层是否拍摄 完整			

在该页面可以为模型的云服务API配置服务功能,详情见服务功能文档

点击右侧「API文档」可以快速跳转至API文档,参考文档调用API获取商品检测AI能力。

模型使用

∞ 服务功能

在完成模型发布后,可以点击模型列表内「操作」列中的「配置服务功能」,如下图:

EasyDL零售版 《三	=	模型列表				操作文	格 常见问题	教学视频	新手引导	提交工单	专家定制
① 模型中心 我的模型		创建模型							请辑	入模型名称	
创建模型 训练模型		【商品检测】饮品Dem	ю ☑ 模型Ⅳ:3	0443						吕 训练	更多
校验模型		部署方式	版本	训练状态	服务状态	SKU F1-score 分布 ⑦	操作				
发布模型		公有云API	V1	• 训练完成	已发布	-	查看版本详情	配置服务功能	胞 校验 (本验H5 ⑦	
☑ 数据中心											

点击后弹出下图所示窗口,可以获取模型的云服务API,API使用方式请参考API调用方法文档。
我的模型 > 饮品DemoV1的公有云API服务详情 操作文档 常见问题 教学视频 新手引导 提交工单 专家定制

服务名称:	111	服务功能描述: 模型服务接口使用;	方法请见 API文档	
候皇叔平·		服务功能	接口返回字段	内容
服务状态: 接口地址:	已发布		name	商品标签,包含商品 名称、品牌、规格
	nce02	商品基本信白	sku_code	商品编号
服务功能:	已选功能不同,单次接口调用价格不同,详细收费方式请见 计费文档	识别	score	识别结果的置信度
	基础功能服务 商品基本信息识别		location	商品检测框在图片上 的像素坐标
	商品标签(名称、品牌、规格)、编号、像素坐标 可选服务功能(开关状态变更合改变APi返回参数,详细请见 API文档) 商品标列层数识别 商品所在发发思数、陈列顺序及货架总层数	商品陈列场景 识别	scenes	支持十类场景:货 架、端架、立式冰 柜、卧式冰柜、冷风 柜、挂钩货架、斜口 篮货架、地堆、割 箱、地龙
	商品陈列场景识别		scene	每个商品所在的陈列 场景
	支持十类场景: 货架、端架、立式冰柜、冷风		shelf	商品所在的货架编 号,从左往右依次递 增
	Moonf 即日しまで11 商品排面数/占比、未识别商品数、空位数及货 (N) 架利用率	商品陈列层数	layer	商品所在货架层数编 号,从上往下依次递 增
	商品陈列翻拍识别	识别	layer_count	统计各货架的总层数
	提示:开启或关闭功能后,约5分钟后生效		layer_top	判断货架最顶层是否 拍摄完整
	立即使用 返回		sku_sn	每一层货架上商品的 陈列顺序

在该页面可以为模型的云服务API配置服务功能,支持以下三项功能:

• 商品基本信息识别 (必选)

接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度

• 商品陈列层数识别 (可选)

接口支持识别商品陈列所在货架层数,货架总层数以及商品的陈列顺序,货架类型支持:货架、端架、冷风柜和立式冰柜内货架

商品陈列场景识别(可选)

接口支持识别商品陈列的场景,场景类型支持:普通货架、挂钩货架、斜口篮货架、端架、立式冰柜、卧式冰柜、冷风柜、地堆、割箱、地 龙、小端架、吧台

- 商品排面占比统计(可选)
 接口支持统计商品排面数/占比、未识别商品数、总空位数、每货架层的空位数及货架利用率
- 商品陈列翻拍识别(可选)
 识别商品陈列照片是对手机屏幕翻拍的可能性

接口单次调用的费用,根据开启的功能不同而不同,详情可见购买指南文档。

可在页面随时开启和关闭可选的功能,**变更功能后约5分钟生效,生效后单次调用费用按变更后的功能计费**,接口将返回变更后的功能字段,详 情可见API调用方法文档。

心 体验H5

简介

在完成模型发布后,可以将模型云服务API快速集成进H5页面中体验模型效果。

生成体验H5页面

可以点击模型列表内「操作」列中的「体验H5」,如下图:

×

模型中心	模型列表							
我的模型	创建模型							
创建模型								
训练模型	【商品检测】饮品	·检测 模型ID:	27899				🔤 训练 🕓 历史版本	<u>面</u> 删除
校验模型	部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作	
发布模型 数据中心	公有云API	V1	训练完成	审核成功	已发布	mAP: 96.95% 2 精确率: 88.46% 2 召回率: 92.00% 2 完整评估结果	查看版本详情 配置服务功能 校验	体验H5 🛛
<u></u> 击后弹出下图所示窗	『口,选择─	-个已创	建的APPID	授权继续。				
体验H5说明								×
H5中的定制化商	品检测功能	將使用	你的app进	! 行调用。				
调用APP: E	asyDL-API	PID: 16	6096 ×					
温馨提示:每次(体验识别将	消耗个人	人账号下的	间用次数	0			
							继续	

如果还未在百度智能云创建任何APP,请按照页面弹窗提示,前往EasyDL零售版控制台应用列表,点击「创建应用」按钮进行创建。

< 返回EasyDL总宽	产品服务	服务 / EasyOL定制训练平台 - 应用列表											
零售版	应用	应用列表											
应用列表	+	创建应用											
用量统计		应用名称	AppID	API Key	Secret Key	创建时间	操作						
监控报表	1	silence123	16935698	mbqkk0SDDHxrjUuBy2pwVCGO	•••••• 显示	2019-08-01 18:45:10	报表 管理 删除						
定制商品检測服务 >	2	Silence-EasyDL	15637143	ghcLQC05QsuiZ2LEhjvtTb48	显示	2019-02-27 11:01:26	报表 管理 删除						
你准时由拉测服务 ~ 仿架拼接 ~													



自定义H5首页样式

	模型名称:	饮品识别
饮品识别	模型介绍:	货架Demo
货架Demo	开发者署名:	Silence
开发者: Silence	H5分享文案:	饮品识别Demo 8/50
		수 라니드

完成配置后,即可扫描弹框内的二维码进行体验模型服务的效果。

体验H5

 \times

用百度或微信APP扫以下二维码,在手机端体验模型效果



修改已配置的H5页面

注:每次体验识别将消耗个人账号下的调用次数

H5页面功能介绍

扫描二维码后,可看到下图页面:

() 一一一一一一一一一一一一一一一一一一一一一一一一一一一一一一一一一一一一
货架Demo
开发者:

点击上传图片后,可以选择本地相册照片,也可以拍摄照片,上传后,将会使用发布的模型云服务API对图片进行识别,结果如下图所示:



- 1. 场景识别结果,在服务功能开启了「商品陈列场景识别」的云服务API的H5页面可以支持返回该结果
- 置信度过滤设置,根据调整的值,查看图片上的商品识别结果,比如图中设定的66%,置信度大于0.66的识别结果将返回,体现在页面上的 是会出现蓝色的检测框
- 3. 商品数量统计结果,在服务功能开启了「商品基本信息识别」的云服务API的H5页面可以支持返回该结果
- 商品数量/层数结果切换按钮,在服务功能开启了「商品陈列层数识别」的云服务API的H5页面可以支持该选项,点击「查看商品层数」后,可以查看各个商品陈列所在货架层数,显示如下图结果页面:



陈列场景:端架



の 云服务API

在完成模型发布后,即可调用云服务API获取商品检测AI能力,可以点击模型列表内「操作」列中的「服务详情」获取API URL,如下图:

模型中心	模型列表						
我的模型	创建模型						
创建模型							
训练模型	【商品检测】饮品	₁检測 模型ID:	27899				😑 训练 🕓 历史版本 🔟 删除
校验模型	部署方式	版本	训练状态	申请状态	服务状态	模型效果	操作
发布模型 数据中心	公有云API	V1	训练完成	审核成功	已发布	mAP: 96.95% ⑦ 精确率: 88.46% ⑦ 召回案: 92.00% ②	查看版本详情 配置服务功能 校验 体验H5 ⑦
我的SKU库						完整评估结果	

点击后弹出下图所示窗口,可以获取模型的云服务API URL。

我的模型 >	1111V1的公有云API服务详情			
服务名称:	111	服务功能描述: 模型服务接口使用方法	请贝 API文档	
服冬州太子		服务功能	接口返回字段	内容
接口地址:	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/detection/silence02		name	商品标签,包含商品名 称、品牌、规格
服务功能:	已选功能不同,单次接口调用价格不同,详细收费方式请见 计费文档	· · · · · · · · · · · · · · · · · · ·	sku_code	商品编号
	基础功能服务	商品基本信息识别	score	识别结果的置信度
	商品基本信息识别 商品标签(名称、品牌、规格)、编号、坐标、置信度		location	商品检测框在图片上的像 素坐标
	可选服务功能 商品陈列层数识别 商品陈列层数识别	商品陈列场景识别	scenes	图片中包含的陈列场景类 型,支持货架、端架、立 式冰柜、地堆和割箱
			scene	每个商品所在的陈列场景
	商品陈列场景识别 支持五类场景: 货架、端架、立式冰柜、地堆和割箱		shelf	商品所在的货架编号,从 左往右依次递增
	商品排面占比统计 商品排面数/占比、未识别商品数、空位数及货架利用率	商品陈列层数识别	layer	商品所在货架层数编号, 从上往下依次递增
	提示:开启或关闭功能后,约5分钟后生效		layer_count	统计各货架的总层数
	立即使用 返回		layer_top	判断货架最顶层是否拍摄 完整

API使用方式详细介绍请参考API调用方法文档

门店管理

の门店库创建

本文档介绍门店管理功能中的门店库创建流程。

在页面创建门店库

在左侧导航栏点击「我的门店库」进入门店管理页面,如下图所示

EasyDL≉ ⇒	我的门店	与库									
创建模型	_										
训练模型	+	创建门店库						请输入门店库名称或门	店库ID		
校验模型	ID		门店库名称	业务门店	待确认门店	待拓展门店	关联应用	导入状态	操作		
发布模型	31	107690	门店库9	0	0	0	387752	 新建 	立看	导入	删除
高数据中心 へ	31	107686	门店库8	61	0	0	.1079	• 导入成功	童看	导入	删除
我的SKU库	31	107684	门店库7-API	196	0	0	₩386	• 部分导入成功 ⑦	宣看	导入	删除
火水回来阵 实易图标注	31	107683	门店库6	213	0	0	1949	• 导入失败 ⑦	立看	导入	删除
我的门店库	31	107669	门店库5	9	0	0	ToJ35698	• 导入成功	皇帝	导入	删除
云服务数据	30	000051	门店库4	171	0	0	0141	• 部分导入成功 ③	宣看	导入	删除
区 AI服务中心 へ	30	1000049	门店库3	220	0	0	101940	• 部分导入成功 ⑦	立看	导入	删除
服务列表	30	000048	门店库2	7	0	0	3	• 导入成功	宣看	导入	删除

点击「创建门店库」按钮,在弹框中按要求输入信息

创建门店库	×
*门店库名称	20字以内,如可乐项目门店库
*控制台应用	请选择在百度云控制台创建的应用 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
	取消 创建

• 门店库名称:定义门店库的名称,20字以内,创建后可在门店库列表中修改

• 控制台应用:选择在百度云EasyDL零售版控制台创建的应用,用于关联门店库,如未创建,请前往创建后再进行创建

门店库关联应用说明:

一个门店库可以关联多个应用,一个应用只能关联一个底库

按要求完成填写后,点击「创建」按钮即可完成门店库的创建,创建的门店库可在门店库列表中看到。

心门店导入

本文档介绍门店管理功能中的门店导入流程。

门店导入

创建好门店库后,在列表中「操作」列点击「导入」,可进入门店导入页面。

```
    ※返回 与入门店
    「门店库信息
    「门店库名称: 「门店庫9
    「门店庫10: 31107690
    ※联应用: 32387752
    ※分门店数: -
    待确认门店数: -
    待知()「店数: -
    行工店类型
    ● 业务门店 ②
    ● 存确认门店 ③
    ● 存職()「日本 ③
    ● 存職()「日本 ③
    ● 存職()「日本 ④
```

1. 门店类型

业务门店

业务门店指的是客户业务系统中真实的业务门店,进入门店导入页面后,在「门店类型」选项中选择「业务门店」即为指定将门店导入业务门店 列表中。

待确认门店

待确认门店指的是待确认的未铺货门店,如百度智能拓店服务输出待拓展门店数据,在「门店类型」选项中选择「待确认门店」即为指定将门店 导入业务门店列表中。

2. 导入方式

批量导入

该方式适用于将门店多条门店数据导入至门店库中,在「导入方式」选项中选择「批量导入」,按页面所提示内容将本地的门店文件上传至平 台,点击「确认」即可开始导入。

< 返回 导入门店						
门店库信息 门店库名称: 「 业务门店数: -	门店库9 -		门店库ID: 待确认门店数	31107690	关联应用: 待拓展门店数:	32387752 -
导入门店						
门店类型	● 业务门店 ⑦	○ 待确认门店	5 O			
导入方式	• 批量导入	单店导入				
*选择本地文件	土 上传文件					
文件要求说明	业 下载示例文件					
	1. 支持xlxs和xls格式	的Excel文件,以	【及CSV文件(分隔符为英文逗号)			
	2. 文件编码支持:UT	F-8, 可参考 编	码查看和修改方法文档			
	3. 文件必须包含以1	「表头,各列的内	容规范如下:			
	表头名称	是否必填	内容规范			
	user_store_id	否	用户业务系统中的真实的业务门	EID,长度限制128B,支持字母、数字、下划线		
	atoro nomo		门庄夕秒 上帝阔判1001D			
确认	取消					

单店导入

该方式适用于将单条门店数据导入至门店库中,在「导入方式」选项中选择「单店导入」,在页面上按要求完成各项目填写,点击「确认」即可 开始导入。

< 返回 导入门股	ī				
门店库信息					
门店库名称: 业条门店数:	「]店库9	门店库ID: 待确认门店数:	31107690	关联应用: 待拓展门店数:	32387752
TENT MERK.		IS NOW TO DE SK.		TO TAKE THE AC	
导入门店					
门店类型	● 业务门店 ⑦ ○ 待确认门店 ⑦				
导入方式	○ 批量导入 ● 单店导入				
导入方式 业务门店ID	 批量导入 単店导入 定义门店的唯一ID,建议使用业务中的真实门店ID 		0		
导入方式 业务门店ID *门店名称	 批量导入 ● 单店导入 定义门店的唯一D,建议使用业务中的真实门店□ 门店的名称,如沃尔玛(宣武门外大街店) 		0		
导入方式 业务门店ID *门店名称 *门店经度	 批量导入 ● 单店导入 定义门店的唯一D,建议使用业务中的真实门店D 门店的名称,如沃尔玛(宣武门外大街店) 门店的经度,如113.028771 		0		
导入方式 业务门店ID *门店名称 *门店经度 *门店纬度	 批量导入 ● 单店导入 定义门店的唯一D,建议使用业务中的真实门店D 门店的名称,如沃尔玛(富武门外大街店) 门店的经度,如113.028771 门店的纬度,如40.228393 		0		

门店导入状态

在门店导入页面点击「确认」后,可在门店库列表中查看门店导入的状态。

ID	门店库名称	业务门店	待确认门店	待拓展门店	关联应用	导入状态	操作		
31107690	门店库9	0	0	0	32387752	 新建 	查看	导入	删除
31107686	门店库8	61	0	0	33941079	• 导入成功	查看	导入	删除
31107684	门店库7–API	196	0	0	33857386	• 部分导入成功 ⑦	查看	导入	删除
31107683	门店库6	213	0	0	33856949	 导入失败 ⑦ 	查看	导入	删除

状态说如下:

- 新建:新创建的门店库,未进行任何门店导入操作
- 导入中:门店导入过程中的状态
- 导入成功:所有门店都导入成功的状态
- 部分导入成功:存在部分门店导入失败的状态,可在?内图标里查看详情,也可点击下载失败详情文件到本地
- 导入失败:所有门店都导入失败的状态

○门店查看和编辑

本文档介绍门店管理功能中的门店查看功能。

门店库详情

门店导入完成后,在门店库列表「操作」列点击「查看」即可进入门店库详情页面,页面包含门店库基本信息和门店列表。

< 返回 门店库详情					导入
门店库信息 门店库名称: 门店库2 业务门店数: 7		门店库ID: 30000048 待确认门店数: -		关联应用: 15637143 待拓展(门店数: –	
业务门店 待确认门店 待拓展门店	5		全部省 / 全部市 💛 🔘 👌	全部渠道 🗸 💿 🦷 请输入门店名称调	Nijen 🛛
业务门店ID	门店名称	经度	纬度	坐标类型	操作
10002	晨晨超市	116.312803	40.047735	BD09II	查看/编辑 删除
10003	绿丰果蔬超市	118	40	BD09II	查看/编辑 删除
10004	王一烧烤	114	34	BD09II	查看/编辑 删除
10005	利华平价超市	115	35	BD09II	查看/编辑 删除
10006	盛源酒楼	116	36	BD09II	查看/编辑 删除
10007	太河全羊馆	117.1	37.2	BD09II	查看/编辑 删除
store_16869062714978_c91bd6cb 1e1b	四川特产文化而集	121.54409	31.22114	GCJ02II	查看/编辑 删除

门店信息

在门店列表的「操作」列中点击「查看/编辑」按钮,进入门店详情页。门店信息会展示门店导入时填入的内容,非必填项如果未填写,则会显示「-」。

门店信息						编辑
业务门店ID:	10002	待确认门店ID:	-	门店名称:	晨晨超市	
经度:	116.312803	纬度:	40.047735	坐标类型:	BD09II	
省份:	-	城市:	-	区县:	-	
地址:	-	渠道类型:	-	连锁类型:	-	
潜力评分:						

点击「编辑」按钮,即可以对门店信息进行修改。

编辑门	店
-----	---

 \times

门店信息

业务门店ID	10002
待确认门店ID	定义待确定门店的唯一ID
*门店名称	晨晨超市
*门店经度	116.312803
*门店纬度	40.047735
*坐标类型	BD09II
省份	门店所在省份,如江苏省
城市	门店所在城市,如南京市
区县	门店所在区/县,如秦淮区
地址	门店所在地址,如中山南路1号新百商场2楼
渠道类型	门店所属渠道类型,如 超市

门脸图片

门脸图片为对应门店的店招图,图片要求如下:

1.至少上传1张图片,最多上传5张图片,建议为多角度拍摄的门脸图片

2.支持png、jpg、jpeg格式

3.图片最短边不小于15px,最长边不大于4096px



取消

确定

点击编辑,即可对已上传的门脸图片进行修改:

编辑门店图片

EasyDL 零售行业版使用说明

 \times

门脸图片



图片要求: 1.至少上传1张图片,最多上传5张图片,建议为多角度拍摄的门脸图片 2.支持png、jpg、jpeg格式 3.图片最短边不小于15px,最长边不大于4096px

取消	确定

注意:门脸图片修改后,保存修改内容会需要一定时间,修改过程中,图片不能被修改,页面显示不会实时更新,可刷新页面观察是否完成 修改,修改完成后,可再对图片进行修改。

商品陈列信息

如果用户同时使用定制商品检测服务的云服务API,可在请求参数中传入门店ID,以建立商品检测识别到的陈列信息和门店信息的关联,建立后, 可在此部分查看该门店最近一次被拜访时的商品陈列信息。

商品陈列信息 ①	1					详情
商品分销数:	5	陈列场景:	地堆、货架、卧式冰柜	最近拜访日期:	2023-07-24	

点击「详情」按钮后,可进入详情页,支持查看最近30天内该门店调用定制商品检测服务API时上传的照片和对应的识别结果。

选择拜访日期: 2023-07-24 🗸 ①

陈列图片	< 4 /20个 > AI识别结	果		
	基本信息	B		^
	log_id:	3611443463434687000 🗖	陈列场景: 地堆	
	近常の	ά: –	货架利用率: =	
	货架空(立数: -	翻拍指数: 0%	
	上传时间	司: 2023-07-24 14:54:57		
	结果明细	8	● 商品数量	○ 所有商品 へ
	sku		数量	占比
	-		1	50.00%
		ACC AND A	1	50.00%

注意:通过定制商品检测服务的云服务API获取的图片可能会有缺失。

窜拍信息

如果用户同时使用窜拍识别服务,可在窜拍任务中定义门店ID,以建立窜拍和门店信息的关联,建立后,可在此部分查看该门店最近一次窜拍任 务所设定最大阈值的信息。

窜拍信息 ①				详情
窜拍图片数量:	20	涉及窜拍人员数: 1	最近一次窜拍任务: 120	

点击「详情」按钮后,可进入窜拍任务详情页,查看窜拍任务具体的结果。

く 返回 🍵	看结果				
基本信 审拍任务 成功图片	€ D: 120 di M: 506 ₩	書时间: 2023-06-13 23:59:32 白灰明: 区分不同人员、区分不同门店		越東封詞: 2023-08-14 00.02:16	
任务结 阈值:	 果◎ 土下航完整結果 0.8 ∨ 人员: 前沿担人员 ∨ 门店: 	10002 ~			筛选
	人员		门店		
\sim	店内测试		10002		
	重复组ID	重复组图片		操作	
	1	5		宣看图片	
	3	5		查看图片	
	2	3		查看图片	
	5	3		查看图片	
	4	2		查看图片	
	6	2		查看图片	

の门店库和门店删除

该文档介绍如何删除门店库和门店。

删除门店库

进入我的门店库页面。

我的门店库									
+ 创建门店商	ξ.					请输入门店库名称或门	店库ID		
ID	门店库名称	业务门店	待确认门店	待拓展门店	关联应用	导入状态	操作	_	
31107690	门店库9	0	0	0	32387752	 新建 	查看 导	ex I	删除
31107686	门店库8	61	0	0	33941079	• 导入成功	查看 与	KA I	删除
31107684	门店库7-API	196	0	0	33857386	• 部分导入成功 ⑦	查看 导	ex i	删除
31107683	门店库6	213	0	0	33856949	• 导入失败 ②	查看 与	ελ I	删除

在门店库列表「操作」列点击「删除」,在弹窗中点击「确认」后即可将门店库删除。

注意:门店库删除后不可恢复,请谨慎操作



删除门店

进入我的门店库页面。

我的	门店库							
1	+ 创建门店库						请输入门店库名称或门	吉库ID
	ID	门店库名称	业务门店	待确认门店	待拓展门店	关联应用	导入状态	操作
	31107690	门店库9	0	0	0	32387752	 新建 	查看 导入 删除
	31107686	门店库8	61	0	0	33941079	• 导入成功	查看 导入 删除
	31107684	门店库7–API	196	0	0	33857386	● 部分导入成功 ⑦	查看 导入 删除
	31107683	门店库6	213	0	0	33856949	• 导入失败 ⑦	查看 导入 删除

在门店库列表「操作」列点击「查看」,进入门店库详情页面。

业务门店ID	门店名称	经度	纬度	坐标类型	操作
10002	晨晨超市	116.312803	40.047735	BD09II	查看/编辑 删除
10003	绿丰果蔬超市	118	40	BD09II	查看/编辑 删除
10004	王一烧烤	114	34	BD09II	查看/编辑 删除
10005	利华平价超市	115	35	BD09II	查看/编辑 删除
10006	盛源酒楼	116	36	BD09II	查看/编辑 删除

在门店列表「操作」列点击「删除」,在弹窗中点击「确认」后即可将门店库删除。

注意:门店删除后不可恢复,请谨慎操作

•	删除门店		×
	门店删除后不可恢复,请确认是否删除		
		取消	确定

API文档

^の商品检测API调用方法

接口描述

本文档主要说明定制化商品检测模型发布后获得的API如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单,咨询问题类型请选择人工智能服务
- 进入EasyDL社区交流,与其他开发者进行互动

Baidu 百度智能云文档

接口鉴权

1. 进入EasyDL零售版的百度智能云控制台应用列表页面,如下图所示:

Û	全局 ~								
	✔ 返回EasyDL总览	产品服务)	产品服务 / EasyDL定制训练平台 - 应用列表						
>	零售版	应用多	列表						
¢	应用列表	+ ê	川建应用						
⊕	用量统计		应用名称	AppID	API Key	Secret Key			
T	监控报表	1	silence123	16935698	mbqkk0SDDHxrjUuBy2pwVCGO	******* 显示			
<u></u>	定制商品检测服务 >	2	Silence-EasyDL	15637143	ghcLQC05QsuiZ2LEhjvtTb48	₩###### 显示			
88	标准商品检测服务 🛛 🗸								
a 🖬 a	货架拼接 >								

2. 如果还未创建应用,请点击「创建应用」按钮进行创建。创建应用后,参考鉴权参考文档,使用API Key(AK)和Secret Key(SK)获取 access_token

请求说明

请求示例

HTTP 方法: POST

请求URL:请首先在EasyDL零售版进行定制商品检测模型训练,完成训练后申请上线,上线成功后可在服务列表中查看并获取url。

URL参数:

参数	值	
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档	

Header如下:

参数	值
Content-Type	application/json

注意:如果出现336001和336002的错误码很可能是因为请求方式错误,与其他图像识别服务不同的是定制化图像识别服务以json方式请求。

Body请求示例:

```
{
"image": "<base64数据>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大4096px,支持 jpg/png/bmp格式 注意请去掉头部
thresho Id	否	number	0~1	可精确到小数点后两位, 默认值为建议阈值 ,请在 我的模型列表·完整评估结果 查看推荐阈值。阈值 (threshold),是正确结果的判定标准,例如阈值是0.6,置信度大于0.6的识别结果会被当作正确结果返 回。
split_sh elf	否	boolea n	True/Fa Ise	True表示区分货架节数,False表示不区分货架节数。默认为False。当未开通 货架层数识别 服务功能时,传 该参数不生效。

提示:image参数中"去掉头部"指的是图片经base64编码后的头部信息「data:image/jpeg;base64,」,如下图所示:

点击这里选择选择要转换成Base64的图片

C:\fakepath\14458.jpg

复制 清空

返回说明

排面数:同层同列去重后的SKU检测数量

开启模型服务功能可参考服务功能文档

返回参数

返回结果为JSON格式

字段	是否必 选	类型	说明	需要开 启的模 型服务 功能
recaptu re_scor e	否	float	图片是对手机屏幕翻拍的可能性评分。翻拍判定方法:设定一个判定为翻拍图片的阈值,即如果 recapture_score大于这个值,则认为这张图片是翻拍。请结合业务实际情况和实测结果进行设定阈值	商品陈 列翻拍 识别
statistic s	否	object{}	对于整张图片的综合识别统计结果	商品排 面占比 统计
+known _sku_n um	否	int	定制模型识别的SKU总数量	商品排 面占比 统计
+unkno wn_sku _num	否	int	定制模型未识别的SKU总数量	商品排 面占比 统计
+known _sku_fa cing	否	int	定制模型识别的SKU总排面数量,如果图片为非货架陈列场景(如堆箱、割箱、地龙等),结果不具参考意 义	商品排 面占比 统计
+unkno wn_sku _facing	否	int	定制模型未识别的SKU总排面数量,如果图片为非货架陈列场景(如堆箱、割箱、地龙等),结果不具参考 意义	商品排 面占比 统计
+vacan cy_num	否	int	货架上的空位数量,如果图片为非货架陈列场景(如堆箱、割箱、地龙等),结果不具参考意义	商品排 面占比 统计
+share _of_she If	否	float	定制商品检测模型识别SKU的排面占比= (known_sku_facing)/(known_sku_facing+ unknown_sku_facing)。 如果图片为非货架陈列场景(如堆箱、割箱、地龙等),结果不具参考意义	商品排 面占比 统计
+known _sku_pr oportio n	否	float	定制商品检测模型识别SKU的数量占比= (known_sku_num)/(known_sku_num+ unknown_sku_num)	商品排 面占比 统计

+utilizat ion	否	float	货架利用率= (known_sku_num)/(known_sku_num+ unknown_sku_num+vacancy_num)。 如果图片为非货架陈列场景(如冰柜、端架、普通货架等),结果不具参考意义	商品排 面占比 统计
sku_co unt	否	array[o bject]	定制商品检测模型识别的各类SKU的总数和排面占比	商品排 面占比 统计
+name	否	string	定制商品检测模型识别的SKU标签,SKU名称_品牌名称_规格参数,为在EasyDL零售版上创建SKU时填写的 内容	商品排 面占比 统计
+sku_c ode	否	string	定制商品检测模型识别的SKU编码,为在EasyDL零售版上创建SKU时填写的内容	商品排 面占比 统计
+sku_n um	否	int	定制商品检测模型识别的各类SKU的总数	商品排 面占比 统计
+propor tion	否	float	定制商品检测模型识别的各类SKU的数量占比= sku_num/(known_sku_num+unknown_sku_num)	商品排 面占比 统计
+sku_fa cing	否	int	定制商品检测模型识别的各类SKU的总排面数	商品排 面占比 统计
+sku_s os	否	float	定制商品检测模型识别的各类SKU的排面数量占比= sku_facing/(known_sku_facing+ unknown_sku_facing)。 如果图片为非货架陈列场景(如堆箱、割箱、地龙等),结果不具参考意义	商品排 面占比 统计
shelf_in fo	否	array[o bject]	各组货架的每层货架的详细统计信息,如果图片为非货架陈列场景(如堆箱、割箱、地龙等),结果不具参 考意义	商品排 面占比 统计
+shelf	否	int	商品所在货架编号,"-1"表示未识别到货架,编号为图中货架最左从往右数依次增大	商品排 面占比 统计
+layer_i nfo	否	array[o bject]	各货架层的详细空位数信息	商品排 面占比 统计
++layer	否	int	商品所在层数编号,"-1"表示未识别到层数,编号为从图中货架最上层往下依次增大	商品排 面占比 统计
++layer _vacanc y	否	int	每一层的空位数量	商品排 面占比 统计
++layer _known _sku_n um	否	int	每一层的可识别SKU数量	商品排 面占比 统计
++layer _unkno wn_sku _num	否	int	每一层的未知SKU数量	商品排 面占比 统计
layer_c ount	否	array[o bject]	图片中,各货架的总层数,如果图片为非货架陈列场景(如堆箱、割箱、地龙等),结果不具参考意义	商品陈 列层数 识别

Baidu	百度智能云文档
-------	---------

+shelf	否	int	商品所在货架编号,"-1"表示未识别到货架,编号为图中货架最左从往右数依次增大	商品陈 列层数 识别
+layer_ num	否	int	货架的总层数,如果图片为非货架陈列场景(如冰柜、端架、普通货架等),结果不具参考意义	商品陈 列层数 识别
layer_to p	否	int	判断是否拍摄到货架最上一层,0表示未拍摄到,1表示拍摄到,-1表示图片中未识别到货架	商品陈 列层数 识别
layer_c omplete	否	int	表面货架是否拍摄完整,0表示不完整,1表示完整	商品陈 列层数 识别
scenes	否	array[st ring]	图片中包含的陈列场景类型。返回所有场景去重后的集合,"GE"表示端架,"shelf"表示货架,"freezer"表 示冰柜,"TG"表示地堆,"cutbox"表示割箱,"DL"表示地龙,"HS"表示挂钩货架,"OBS"表示斜口篮货 架,"SGE"表示小端架,"barcounter"表示吧台,"HF"表示卧式冰柜,"OACR"表示冷风柜,"HGE"表示挂钩 端架,"unknown"表示未识别到场景	商品陈 列场景 识别
log_id	是	int	唯一的log id,用于问题定位	商品基 本信息 识别
results	否	array[o bject]	图片中每个商品的详细信息	-
+name	否	string	SKU名称_品牌名称_规格参数	商品基 本信息 识别
+scene	否	string	表示该SKU所在的陈列场景。"GE"表示端架,"shelf"表示货架,"freezer"表示冰柜,"TG"表示地 堆,"cutbox"表示割箱,"DL"表示地龙,"HS"表示挂钩货架,"OBS"表示斜口篮货架,"SGE"表示小端 架,"barcounter"表示吧台,"HF"表示卧式冰柜,"OACR"表示冷风柜,"HGE"表示挂钩端架,"unknown"表 示未识别到场景	商品陈 列场景 识别
+shelf	否	int	商品所在货架编号,"-1"表示未识别到货架,编号为图中货架最左从往右数依次增大,如果图片为非货架陈 列场景(如堆箱、割箱、地龙等),结果不具参考意义	商品陈 列层数 识别
+layer	否	int	商品所在层数编号,"-1"表示未识别到层数,编号为从图中货架最上层往下依次增大,如果图片为非货架陈 列场景(如堆箱、割箱、地龙等),结果不具参考意义	商品陈 列层数 识别
+sku_s n	否	string	商品的陈列排序序号,返回"A-B"或"A-B-C",如"2-1"或"2-1-1",其中A、B、C分别为数字,A代表同一货架 层的横向顺序,从左至右依次增大;B代表同一货架的层纵向序号,从下至上依次增大;如果存在大商品上 下陈列有小商品或包含小商品的情况,会出现C,从左至右依次增大。如果图片为非货架陈列场景(如堆 箱、割箱、地龙等),结果不具参考意义	商品陈 列层数 识别
+sku_c ode	否	string	商品编号,由用户在模型训练页面创建SKU时自定义	商品基 本信息 识别
+score	否	float	置信度	商品基 本信息 识别
+locatio n	否	object{}	每个商品在图上的像素位置	商品基 本信息 识别
++left	否	int	检测到的目标主体区域到图片左边界的像素距离	商品基 本信息 识别
++top	否	int	检测到的目标主体区域到图片上边界的像素距离	商品基 本信息

Baidu 百度智能云文档

EasyDL 零售行业版使用说明

				识别
				商品基
++width	否	int	检测到的目标主体区域的像素宽度	本信息
				识别
hair				商品基
++neig	否	int	检测到的目标主体区域的像素高度	本信息
in				识别

建议翻拍判定方法

设定一个判定为翻拍图片的阈值,即如果recapture的score大于这个值,则认为这张图片是翻拍。通常有两中对应的业务模式:

注:以下数值均为建议值,实际应用的阈值请结合业务实际情况和实测结果进行设定

1. 业务里查翻拍的原则是宁可错杀一千,也不愿错放一个的,那么可以把认为是翻拍的阈值放在0.8~0.95。

2. 业务里查翻拍的原则是允许错放过一些翻拍的图片,但是查到的一定要对,那么可以把认为是翻拍的阈值放在0.98甚至0.99。

心 地堆检测API调用方法

接口描述

本文档主要说明定制化商品检测模型发布后获得的API如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单,咨询问题类型请选择人工智能服务
- 进入EasyDL社区交流,与其他开发者进行互动

接口鉴权

1. 进入EasyDL零售版的百度智能云控制台应用列表页面,如下图所示:

Û	全局 ~									
	✔ 返回EasyDL总览	Ē	产品服务 / EasyDL定制训练平台 - 应用列表							
>	零售版		应用列表							
¢	应用列表		+ ê]建应用						
⊕	用量统计			应用名称		AppID	API Key	Secret Key		
T	监控报表		1	silence123		16935698	mbqkk0SDDHxrjUuBy2pwVCGO	****** 显示		
<u>(4</u>)	定制商品检测服务 🗸		2	Silence-EasyDL		15637143	ghcLQC05QsuiZ2LEhjvtTb48	****** 显示		
88	标准商品检测服务	~								
ß	货架拼接	~								

2. 如果还未创建应用,请点击「创建应用」按钮进行创建。创建应用后,参考鉴权参考文档,使用API Key(AK)和Secret Key(SK)获取 access_token

请求说明

请求示例

HTTP 方法: POST

请求URL: 请首先在EasyDL零售版进行定制商品检测模型训练,完成训练后申请上线,上线成功后可在服务列表中查看并获取url。

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

注意:如果出现336001和336002的错误码很可能是因为请求方式错误,与其他图像识别服务不同的是定制化图像识别服务以json方式请求。

Body请求示例:

```
{
"image": "<base64数据>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大4096px,支持 jpg/png/bmp格式 注意请去掉头部
thresho Id	否	number	0~1	可精确到小数点后两位, 默认值为建议阈值 ,请在 我的模型列表·完整评估结果 查看推荐阈值。阈值 (threshold),是正确结果的判定标准,例如阈值是0.6,置信度大于0.6的识别结果会被当作正确结果返 回。

提示:image参数中"去掉头部"指的是图片经base64编码后的头部信息「data:image/jpeg;base64,」,如下图所示:

点击这里选择选择要转换成Base64的图片	C:\fakepath\14458.jpg	复制	清空
data:image/jpeg;base64, ⁹ 9j/4RCRRXhpZgA/	ATU0AKgAAAAgAEQEAAAMAAAABDYAAAAEBAAMAAAABEgAAAA	AECAAN	/IAA
AADAAAA2gEGAAMAAAABAAIAAAEOAAIAA	AAEbm9yAAEPAAIAAAAHAAAA4AEQAAIAAAAJAAAA5wESAAMAA	AABAAE	EAA
AEVAAMAAAABAAMAAAEaAAUAAAABAAAA	8AEbAAUAAAABAAAA+AEoAAMAAAABAAIAAAExAAIAAAAkAAABA	AAEyAA	IIAA
AAUAAABJAITAAMAAAABAAEAAIdpAAQAAA	AABAAABOIgIAAQAAAABAAADeAAABFgACAAIAAhIVUFXRUKASFd.	JLUFMN	/IDA
AAAr8gAAAJxAACvyAAAAnEEFkb2JIIFBob3F	Rvc2hvcCBDQyAyMDE3IChNYWNpbnRvc2gpADIwMTk6MDQ6MTG	gMjE6N	VTU
6MiaAACWCmaAFAAAAAQAAAvaCnQAFAA	AAAQAAAwKIIaADAAAAAQACAACIJwADAAAAAQCaAACQAAAHAA		VMT

返回说明

开启模型服务功能可参考服务功能文档

排面数:同层同列去重后的SKU检测数量

返回参数

字段	是否必 选	类型	说明	需要开启的 模型 服务 功 能
recapture _score	否	float	图片是对手机屏幕翻拍的可能性评分。翻拍判定方法:设定一个判定为翻拍图片的阈值,即如果 recapture_score大于这个值,则认为这张图片是翻拍。请结合业务实际情况和实测结果进行设定阈值	商品陈列翻 拍识别(可 选)
log_id	是	number	唯一的log id,用于问题定位	商品基本信 息识别 (必 选)
scenes	否	array(st ring)	图片中包含的陈列场景类型。返回所有场景去重后的集合,"GE"表示端架,"shelf"表示货架,"freezer"表示冰柜,"TG"表示地堆,"cutbox"表示割箱,"DL"表示地龙,"HS"表示挂钩货架,"OBS"表示斜口篮货架,"SGE"表示小端架,"barcounter"表示吧台,"unknown"表示未识别到场景	商品陈列场 景识别(可 选)
results	否	array(o	识别结果教组	商品基本信 息识别 (必

Baidu 百度智能云文档

	н	bject)		选)
+stack_s n	否	number	堆头的序号,从左至右依次增加。举例:如果模型检测出图片存在两个堆头,那边两个堆头从左至右 的stack_sn分别为1和2	商品基本信 息识别 (必 选)
+stack_in fo	否	array(o bject)	各个堆头的检测结果	商品基本信 息识别 (必 选)
++type	否	number	堆头陈列的类型:1为KA可识别长宽;2为非KA可计数;-1为其它(不可计数和不可计长宽)	商品基本信 息识别 (必 选)
++stack_ height	否	number	该堆头的高度(Y轴数量)	商品基本信 息识别 (必 选)
++stack_ width	否	number	该堆头的宽度(列数,X轴数量)	商品基本信 息识别 (必 选)
++stack_ depth	否	number	该堆头的纵深(Z轴数量)	商品基本信 息识别 (必 选)
++area	否	number	该堆头的占地面数量 = width * depth	商品基本信 息识别 (必 选)
++stack_ sku_num	否	number	该堆头中的可识别SKU的含纵深总数量	商品基本信 息识别 (必 选)
++stack_ sku_num _visible	否	number	该堆头中的可识别SKU的可见总数量	商品基本信 息识别 (必 选)
++stack_ sku_info	否	array(o bject)	堆箱内含有SKU的信息	商品基本信 息识别 (必 选)
+++name	否	string	SKU标签名称	商品基本信 息识别 (必 选)
+++sku_ code	否	string	SKU编码	商品基本信 息识别 (必 选)
+++num_ with_dept h	否	number	该类SKU在该堆头中的含纵深总数量	商品基本信 息识别 (必 选)
+++num_ without_d epth	否	number	该类SKU在该堆头中的可见总数量	商品基本信 息识别 (必 选)
+col_info	否	array(o bject)	各个堆头中,各列的检测结果	商品基本信 息识别 (必 选)
++col_loc _info	否	array(o bject)	各列的深(Z轴)和高(Y轴)信息	商品基本信 息识别 (必 选)
+++col_s n	否	array(o bject)	列的序号,从左至右依次增加	商品基本信 息识别(必 选)

ر ينع

+++col_d epth	否	number	列的纵深(Z轴数量)SKU数量	商品基本信 息识别 (必 选)
+++col_h eight	否	number	列的高度(Y轴数量)SKU数量	商品基本信 息识别 (必 选)
++col_nu m_info	否	array(o bject)	各列中各类SKU信息和数量结果	商品基本信 息识别 (必 选)
+++col_s n	否	number	列的序号,从左至右依次增加	商品基本信 息识别 (必 选)
+++col_s ku_num	否	number	该列的可识别SKU的含纵深总数量	商品基本信 息识别 (必 选)
+++col_s ku_num_ visible	否	number	该列的可识别SKU的可见总数量	商品基本信 息识别 (必 选)
+++col_s ku_info	否	array(o bject)	列内各类SKU的信息	商品基本信 息识别 (必 选)
++++nam e	否	string	SKU标签名称	商品基本信 息识别 (必 选)
++++sku _code	否	string	SKU编码	商品基本信 息识别 (必 选)
++++num _with_de pth	否	number	该类SKU在该列中的含纵深总数量	商品基本信 息识别 (必 选)
++++num _without_ depth	否	number	该类SKU在该列中的可见总数量	商品基本信 息识别 (必 选)
++s_col_i nfo	否	array(o bject)	各列中,各 子列 的检测结果,每个纵深(Z轴)列为一个子列	商品基本信 息识别 (必 选)
+++col_s n	否	number	列的序号,从左至右依次增加	商品基本信 息识别 (必 选)
+++s_col _sn	否	number	子列的序号,从内到外,依次增加	商品基本信 息识别 (必 选)
+++s_col _sku_nu m	否	number	该列的可识别SKU纵深总数量	商品基本信 息识别 (必 选)
+++s_col _sku_nu m_visible	否	number	该列的可识别SKU可见总数量	商品基本信 息识别(必 选)(必 选)
+++s_col _sku_info	否	array(o bject)	当前子列内包含SKU的信息	商品基本信 息识别 (必 选)

				ي تي ا
++++nam e	否	string	SKU标签名称	商品基本信 息识别 (必 选)
++++sku _code	否	string	SKU编码	商品基本信 息识别 (必 选)
++++num _with_de pth	否	number	该类SKU在该子列中的纵深总数量	商品基本信 息识别 (必 选)
++++num _without_ depth	否	number	该类SKU在该子列中的可见总数量	商品基本信 息识别 (必 选)
+sku_inf o	否	array(o bject)	该堆头中的所有商品信息	商品基本信 息识别 (必 选)
++name	否	number	SKU标签名称	商品基本信 息识别 (必 选)
++sku_c ode	否	number	SKU编码	商品基本信 息识别 (必 选)
++scene	否	string	该SKU的陈列场景,"GE"表示端架,"shelf"表示货架,"freezer"表示冰柜,"TG"表示地堆,"cutbox"表 示割箱,"DL"表示地龙,"HS"表示挂钩货架,"OBS"表示斜口篮货架,"SGE"表示小端 架,"barcounter"表示吧台,"unknown"表示未识别到场景	商品陈列场 景识别 (可 选)
++score	否	number	置信度	商品基本信 息识别 (必 选)
++locatio n	否	array(o bject)	该SKU在原图中的像素位置信息	商品基本信 息识别 (必 选)
+++left	否	number	检测到的目标主体区域到图片左边界的像素距离(px)	商品基本信 息识别 (必 选)
+++top	否	number	检测到的目标主体区域到图片上边界的像素距离(px)	商品基本信 息识别 (必 选)
+++width	否	number	检测到的目标主体区域的像素宽度(px)	商品基本信 息识别 (必 选)
+++heigh t	否	number	检测到的目标主体区域的像素高度(px)	商品基本信 息识别 (必 选)
++threed _location	否	array(o bject)	该SKU在堆头中的三维位置信息	商品基本信 息识别 (必 选)
+++sku_ height	否	number	SKU所在堆头中的3D位置,Y轴高度(从下往上数第几个SKU)	商品基本信 息识别 (必 选)
+++sku_ width	否	number	SKU所在堆头中的3D位置,X轴宽度(从左往右数第几个SKU)	商品基本信 息识别 (必 选)

Baidu 百度智能云文档

depth

息识别 (必

选)

建议翻拍判定方法

+++sku_ 否

设定一个判定为翻拍图片的阈值,即如果recapture的score大于这个值,则认为这张图片是翻拍。通常有两中对应的业务模式:

注:以下数值均为建议值,实际应用的阈值请结合业务实际情况和实测结果进行设定

1. 业务里查翻拍的原则是宁可错杀一千,也不愿错放一个的,那么可以把认为是翻拍的阈值放在0.8~0.95。

2. 业务里查翻拍的原则是允许错放过一些翻拍的图片,但是查到的一定要对,那么可以把认为是翻拍的阈值放在0.98甚至0.99。

心 错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code : 错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
   "error_code": 110,
   "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

Baidu 百度智能云文档

错误码	错误信息	描述				
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(1009661589)或工单联系技术支持 团队。				
2	Service temporarily unavailable	δ务暂不可用,请再次请求, 如果持续出现此类错误,请通过QQ群(1009661589)或工单联系技术支持团 人。				
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试				
4	Open api request limit reached	集群超限额				
6	No permission to access data	无权限访问该用户数据				
13	Get service token failed	获取token失败				
14	IAM Certification failed	IAM鉴权失败				
15	app not exists or create failed	应用不存在或者创建失败				
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请通过QQ群(1009661589)联系群管手动提额				
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线计 费的接口,请通过QQ群(649285136)联系群管手动提额				
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上 线计费的接口,请通过QQ群(649285136)联系群管手动提额				
100	Invalid parameter	无效的access_token参数,请检查后重新尝试				
110	Access token invalid or no longer valid	access_token无效				
111	Access token expired	access token过期				
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(1009661589)或工单联系技术支持 团队				
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有 疑问请通过QQ群(1009661589)或工单联系技术支持团队				
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数或代码格式有误。有疑问请通过QQ群(1009661589)或工单 联系技术支持团队				
336003	Base64解码失败	图片格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问请 通过QQ群(1009661589)或工单联系技术支持团队				
336004	输入文件大小不合法	图片或音频、文本格式有误,请根据接口文档检查入参格式,有疑问请通过QQ群(1009661589)或工单联 系技术支持团队				
336005	解码输入失败/分词错 误	图片编码错误(非jpg,bmp,png等常见图片格式),请检查并修改图片格式				
336006	缺失必要参数	image字段缺失(未上传图片)				
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请通过QQ群 (1009661589)或工单联系技术支持团队				

翻拍识别服务

服务介绍

心简介

商品陈列翻拍识别能够识别出通过手机翻拍出的商品陈列照片,比如商品货架陈列图片和地堆商品陈列图片,可降低人工审核人力,高效审核零售业务中通过翻拍原有图片来造假的图片。

适用于识别以下类型的商品陈列翻拍图片:

- 货架/货柜上商品陈列图片
- 地堆商品陈列图片

∞ 使用须知

- 服务价格详情请见购买指南
- 服务接口调用方法请见API文档

购买指南

⊙ 开通付费及购买服务

EasyDL零售版的各项服务的「开通付费」、「购买QPS叠加包」、「购买调用量次数包」、「关闭付费」等操作均在EasyDL零售版控制台进行, 只需在相应需要付费使用的接口位置,跟随页面提示完成后续充值及付费,即可完成。

≮返回EasyDL总宽		零售版概览										
零售版												
应用列表		产品介绍: (展开查看服务功能介绍)									展开	
用量统计		可用接口列表	羽娥口列表									
监控报表		┃ 已上线的定制接口										
定制商品检测服务 🗸		API	模型ID	模型类型	模型名称	模型版本	状态		调用量限制	QPS限制	开通按量后付费 ?	购买QPS叠加包 ?
标准商品检测服务 >>		111	30443	商品检测	1111	V1	● 免费使	用	剩余免费1000次	不保证并发	开通	购买 配额详情
货架拼接 >		silenceteapi	27185	商品检测	silencesecond	V3	●付费使	用	500次/天免费 + 超 出按量计费	4	终止付费	购买丨配额详情
		silencefirst0829	27135	商品检测	silencefirst	V3	● 免费使	用	500次/天免费	不保证并发	开通	购买 配额详情
		【 饮品检测										
		API		状态		调用量限制			QPS限制		开通按量后付费 ?	
		饮品检测		● 免费使用		500次/天免费			不保证并发		免费试用	
		┃ 日化品检测										
		API		状态		调用量限制			QPS限制		开通按量后付费 ?	
		日化品检测		● 免费使用		剩余免费1000次			不保证并发		免费试用	
		▮ 商品陈列翻拍识别										
		API		状态	30/	月量限制		QPS限制		开通按量后付费	?	购买QPS叠加包 ?
		商品陈列翻拍识别		● 付费使用	剩余	长免费999次 + 超出按量计数	ŧ.	4		终止付费		购买丨配额洋情

ゆけ费方式

商品陈列翻拍识别目前支持下列三种计费方式:

- 1. 按调用量后付费
- 2. 调用量次数包预付费
- 3. QPS叠加包预付费

⊙价目表 - 按调用量后付费

付费调用

每个账户享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用, 价格如下:

调用量	价格	QPS限制	说明
每次调用	0.05元	4	服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

每个账号享有一定量免费调用额度,如下表:

云服务API	免费调用额度	QPS限制	说明
商品陈列翻拍识别	累计1000次	1~2	服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

心价目表 - 调用量次数包

如果业务上对调用次数有预估,可以选择购买单次调用价格更低的次数包,价格如下:

规格	价格	QPS限制	有效期
1万次	490 元	4	1年(366天)
10万次	4,800 元	4	1年(366天)
100万次	45,000 元	4	1年(366天)
500万次	212,500 元	4	1年(366天)
1000万次	420,000 元	4	1年(366天)
2000万次	800,000 元	4	1年 (366天)

购买后不可退款,次数包使用完后,开始按调用量每次0.05元收取费用

特殊说明,此计费方式仅限于单独调用翻拍模型接口,定制商品检测服务接口中的翻拍服务的计费不适用

心价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格 按天购买 50元/天

按月购买 1200元/月

购买 QPS 叠加包需保证已开通按量后付费或购买次数包

购买的 QPS 叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

API文档

```
の API调用方法
```

简介

本文档主要说明如何使用翻拍识别API,如有疑问可以通过以下方式联系我们:

• 在百度云控制台内提交工单,咨询问题类型请选择人工智能服务

接口鉴权

1. 进入EasyDL零售版的百度云控制台应用列表页面,如下图所示:

Û	全局 ~												
	≮ 返回EasyDL总	总览 产品服务 / EasyDL定制训练平台 - 应用列表											
>	零售版		应用列	应用列表									
¢	应用列表		+ 创建应用										
⊕	用量统计			应用名称	AppID	ΑΡΙ Κον	Secret Key						
T	监控报表		1	silence123	16935698	mbakk0SDDHxriUuBv2pwVCGQ	******* 显示						
<u>66</u>	定制商品检测服务	~											
пе	标准商品检测服务		2	Silence-EasyDL	15637143	ghcLQC05QsuiZ2LEhjvtTb48	******* 显示						
80	初始的面包的股另	Ť											
Ð	货架拼接	~											

2. 如果还未创建应用,请点击「创建应用」按钮进行创建。创建应用后,参考鉴权参考文档,使用API Key(AK)和Secret Key(SK)获取 access_token

接口调用

请求说明

请求示例

HTTP 方法: POST

商品陈列翻拍识别请求URL: https://aip.baidubce.com/rpc/2.0/easydl/v1/retail/recapture

URL参数:

参数 值	
------	--

access_token 通过API Key和Secret Key获取的access_token,参考"Access Token获取"

Header如下:

参数 值 Content-Type application/json

注意:如果出现336001和336002的错误码很可能是因为请求方式错误,与其他图像识别服务不同的是定制化图像识别服务以json方式请求。

Body请求示例:

{ "image": "<base64数据>" }

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大4096px,支持 jpg/png/bmp格式 注意请去掉头部

提示:image参数中"去掉头部"指的是图片经base64编码后的头部信息「data:image/jpeg;base64,」,如下图所示:

点击这里选择选择要转换成Base64的图片	C:\fakepath\14458.jpg	复制	清空
	TUOAKgAAAAgAEQEAAAMAAAABDYAAAAEBAAMAAAABEgAAAA	ECAAN	IAA
AADAAAA2gEGAAMAAAABAAIAAAEOAAIAAA	AEbm9yAAEPAAIAAAAHAAAA4AEQAAIAAAAJAAAA5wESAAMAAA	AABAAE	AA
AEVAAMAAAABAAMAAAEaAAUAAAABAAAA8	AEbAAUAAAABAAAA+AEoAAMAAAABAAIAAAExAAIAAAAkAAABA	AAEyAA	AA
AAUAAABJAITAAMAAAABAAEAAIdpAAQAAA	ABAAABOIgIAAQAAAABAAADeAAABFgACAAIAAhIVUFXRUkASFd.	JLUFMN	1DA
AAAr8gAAAJxAACvyAAAAnEEFkb2JIIFBob3Rv 6MiaAACWCmaAFAAAAAQAAAvaCnQAFAAA	rc2nvcCBDQyAyMDE31ChNYWNpbnRvc2gpADlwMTk6MDQ6MTQ AAQAAAwKIIaADAAAAAQACAACIJwADAAAAAQCaAACQAAAHAA	gMjE6N AABDAv	/MT

PHP	
JAVA	
Python3	
C++	

٠

-

php</th
/**
*发起http post请求(REST API),并获取REST请求的结果
* @param string \$url
* @param string \$param
* @return - http response body if succeeds, else false.
*/
function request_post(\$url = '', \$param = '')
{
if (empty(\$url) empty(\$param)) {
return false;
}
\$postUrl = \$url;
\$curlPost = \$param;
// 初始化curl
<pre>\$curl = curl_init();</pre>
curl_setopt(\$curl, CURLOPT_URL, \$postUrl);
A CONTRACT OUDLAST HEADER AN

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
results	否	array(object)	分类结果数组
+name	否	string	分类名称,结果会返回"recapture"和"original"两类,recapture为翻拍,original为原图。
+score	否	number	置信度,分别返回"recapture"和"original"两类的置信度

建议翻拍判定方法

设定一个判定为翻拍图片的阈值,即如果recapture的score大于这个值,则认为这张图片是翻拍。通常有两中对应的业务模式:

注:以下数值均为建议值,实际应用的阈值请结合业务实际情况和实测结果进行设定

- 1. 业务里查翻拍的原则是宁可错杀一千,也不愿错放一个的,那么可以把认为是翻拍的阈值放在0.8~0.95。
- 2. 业务里查翻拍的原则是允许错放过一些翻拍的图片,但是查到的一定要对,那么可以把认为是翻拍的阈值放在0.98甚至0.99。

心 错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code : 错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
   "error_code": 110,
   "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(群号:1009661589)或工单联系技术 支持团队。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请通过QQ群(群号:1009661589)或工单联系技术支 持团队。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请通过QQ群(群号:1009661589)联系群管手动提额
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线计 费的接口,请通过QQ群(群号:1009661589)联系群管手动提额
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上 线计费的接口,请通过QQ群(群号:1009661589)联系群管手动提额
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(群号:1009661589)或工单联系技术 支持团队
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有疑 问请通过QQ群(群号:1009661589)或工单联系技术支持团队
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数代码格式是否有误。有疑问请通过QQ群(群号:1009661589) 或工单联系技术支持团队
336003	Base64解码失败	图片格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问请 通过QQ群(群号:1009661589)或工单联系技术支持团队
336004	输入文件大小不合法	图片或音频、文本格式有误,请根据接口文档检查入参格式,有疑问请通过QQ群(群号:1009661589)或工 单联系技术支持团队
336005	解码输入失败/分词错 误	图片编码错误(非jpg,bmp,png等常见图片格式),请检查并修改图片格式
336006	缺失必要参数	image字段缺失(未上传图片)
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请通过QQ群(群 号:1009661589)或工单联系技术支持团队

购买指南

心 开通付费及购买服务

EasyDL零售版的各项服务的「开通付费」、「购买QPS叠加包」、「购买调用量次数包」、「关闭付费」等操作均在EasyDL零售版控制台进行, 只需在相应需要付费使用的接口位置,跟随页面提示完成后续充值及付费,即可完成。

< 返回EasyDL思觉	零售版概览											
零售版												
应用列表	产品介绍: (展开5	产品介绍: (限升查看编员为能介绍) 脱升]用独口列表										
用量统计	可用接口列表											
监控报表	┃ 已上线的定制接口											
定制商品检测服务 >	API	模型ID	模型类型	模型名称	模型版本	状态	调用量限制	QPS限制	开通按量后付费 ?	购买QPS叠加包 ?		
标准商品检测服务 >	111	30443	商品检测	1111	V1	●免费使用	剩余免费1000次	不保证并发	开通	购买 配额详情		
货架拼接 >	silenceteapi	27185	商品检测	silencesecond	V3	●付费使用	500次/天免费 + 超 出按量计费	4	终止付费	购买丨配额详情		
	silencefirst0829	27135	商品检测	silencefirst	V3	●免费使用	500次/天免费	不保证并发	开通	购买 配额详情		
	(快品检测)											
	API		状态		调用量限制		QPS限制		开通按量后付费 ?			
	饮品检测		● 免费使用		500次/天免费		不保证并发		免费试用			
	日化品检测											
	API		状态		调用量限制		QPS限制		开通按量后付费 ?			
	日化品检测		● 免费使用		剩余免费1000次		不保证并发		免费试用			
	商品陈列翻拍识别											
	API		状态		调用量限制	QPS限制		开通按量后付费	?	购买QPS叠加包 ?		
	商品陈列翻拍识别		 付费使用 		剩余免费999次+超出按量计多	1 4		终止付费		购买丨配额详情		

◎ 计费方式

商品陈列翻拍识别目前支持下列三种计费方式:

- 1. 按调用量后付费
- 2. 调用量次数包预付费
- 3. QPS叠加包预付费

心价目表 - 按调用量后付费

付费调用

每个账户享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用, 价格如下:

```
        调用量
        价格
        QPS限制
        说明

        每次调用
        0.05元
        4
        服务器支持每秒处理4次查询
```

注:调用失败不计费

免费额度

每个账号享有一定量免费调用额度,如下表:

云服务API	免费调用额度	QPS限制	说明
商品陈列翻拍识别	累计1000次	1~2	服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

₯ 价目表 - 调用量次数包

如果业务上对调用次数有预估,可以选择购买单次调用价格更低的次数包,价格如下:

规格	价格	QPS限制	有效期
1万次	490 元	4	1年 (366天)
10万次	4,800 元	4	1年 (366天)
100万次	45,000 元	4	1年(366天)
500万次	212,500 元	4	1年(366天)
1000万次	420,000 元	4	1年(366天)
2000万次	800,000 元	4	1年 (366天)

购买后不可退款,次数包使用完后,开始按调用量每次0.05元收取费用 特殊说明,此计费方式仅限于单独调用翻拍模型接口,定制商品检测服务接口中的翻拍服务的计费不适用

心价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格

按天购买 50元/天 按月购买 1200元/月

购买 QPS 叠加包需保证已开通按量后付费或购买次数包

购买的 QPS 叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

门店拜访SDK (原货架拼接服务)

服务介绍

心简介

百度门店拜访SDK主要包含货架拼接和门脸文字识别两个主要功能,下面是详细介绍。

心 货架拼接

货架拼接服务基于百度EasyDL深度学习算法,支持将多个货架局部图片或视频,组合为完整货架图片。同时支持输出在完整货架图中的商品检测 结果,包含SKU的名称和数量,适用于需要在长货架进行商品检测的业务场景。

拼接方式介绍

货架拼接服务支持三种拼接方式:

- 1. 图片拼接-手机端实时拼接:拍摄图片进行拼接,可实时获得拼接结果
- 2. 图片拼接-云端非实时拼接:拍摄图片进行拼接,需要2~3分钟获得拼接结果

3. 视频拼接-云端非实时拼接:拍摄视频进行拼接,需要2~3分钟获得拼接结果

下面为货架拼接的效果图:





使用方式介绍

货架拼接服务提供以下三种使用方式:

- 云服务API,面向云端非实时拼接方式,可参考货架拼接文档API调用方法
- 可二次开发的SDK,支持iOS和Android,可参考文档iOS SDK和Android SDK
- 可以直接使用的体验APP,包含上述三种拼接方式的体验功能

使用须知

- 1. 使用货架拼接服务前请先参考文档快速训练一个模型或模型发布完成模型发布,货架拼接服务中需要使用商品检测服务输出完整货架图中的 商品检测结果
- 2. 商品检测服务接口升级更新后,货架拼接服务会自动更新
- 3. 视频/图片-非实时拼接方式,每个账号共可免费使用200次货架拼接,SDK和体验APP均会消耗账号内的调用次数
- 4. 视频/图片-非实时拼接方式,每个账号仅支持同时进行一个拼接任务,超出需排队等待

体验APP功能简述

- 1. 支持拍摄一个或多个货架局部视频/图片进行拼接
- 2. 视频/图片-非实时拼接方式支持返回拼接后的完整图片,所有图片拼接去重后的商品识别结果
- 3. 图片拼接·手机端实时拼接支持实时拼接效果预览,所有图片拼接去重后的商品识别结果,不支持输出所有图片拼接的大图
- 4. 体验APP仅支持同时进行一个非实时拼接任务,超出需要排队等待

心门脸文字识别

门脸文字识别功能支持识别图片中的门脸文字信息,包含门脸名称和描述文字。

使用须知

目前该服务处于**邀请使用**状态,请加入QQ群(群号:1009661589)联系管理员申请邀测,提供公司名称和在EasyDL零售版控制台应用列表创建 应用的APPID,如下图:

全局 ~						Q	工单	消息	帮助文档	企业组织	财务	S ~
✓ 返回EasyDL总览	产品服务	/ EasyDL定制训练平台 - 应用列表										
零售版	应用列表											
应用列表	+ 创建进用											
用量统计		应用名称	AppID	API Key	Secret Key	创建的	†(ii)		操作			
监控报表	1	silence123	16935698	mbqkk0SDDHxrjUuBy2pwVCGO	显 示	2019-	-08-01 18:	45:10	报表 管理	#10		
定制商品检测服务 >	2	Silence-EasyDL	15637143	ghcLQC05QsuiZ2LEhjvtTb48	显示	2019	02-27 11:	01:26	报表 管理	#100		
500E960E2880€99 ♥											< 1	>

使用方式介绍

门脸文字识别功能提供以下三种使用方式:

- 云服务API,需完成邀测申请
- 可二次开发的SDK,支持iOS和Android,可参考文档iOS SDK和Android SDK
- 可以直接使用的体验APP

购买指南

货架拼接服务支持按任务数后付费、任务次数包预付费和并发任务叠加包预付费三种计费方式。

心 开通付费及购买服务

货架拼接服务的「开通付费」、「购买任务次数包」、「购买并发任务叠加包」、「关闭付费」等操作均在EasyDL零售版控制台进行,只需在相应需要付费使用的接口位置,跟随页面提示完成后续充值及付费,即可完成。

< 返回EasyDL总览	商品陈列翻拍识别									
零售版	API	状态	调用量限制			QPS限制	开通按量后付费 ?	购买次数1	ä ?	购买QPS叠加包 ?
应用列表	商品陈列翻拍识别	●付费修	B使用 剩余免费988次 + 超出技		出按量计费	4	终止付费	购买		购买丨配额详情
用量统计	货架拼接									
监控报表	服务	状态		任务数限制 ?		并发任务限制 ?	开通按任务数后付费 ?	购买任务》	欠数包 ?	购买并发任务 ?
定制商品检测服务 >	货架拼接	● 免费(七用	剩余免费199次		不保证并发	开通	购买		购买 配额详情
标准商品检测服务 >										
货架拼接 ~	API		状态		调用量限制		QPS限制		开通按量后付费	
	创建任务		● 免费使用	Ð	无限制		不保证并发		免费试用	
	上传图片		● 免费使用	Ð	无限制		不保证并发		免费试用	
	启动任务		● 免费使用	Ð	无限制		不保证并发		免费试用	
	查询任务		● 免费使用	Ð	无限制		不保证并发		免费试用	
	终止任务		● 免费使用	Ð	无限制		不保证并发		免费试用	
	任务列表		● 免费使用	1	无限制		不保证并发		免费试用	

₯价目表 - 按任务数后付费

付费调用

每个账户享有累计200次免费调用额度,免费额度用尽后,请在百度智能云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用,价格如下:

任务数	价格(元)	并发任务数限制	说明
每次拼接任务	0.2	1	只允许一个拼接任务在运行,超出限制的任务排队等待

注:任务启动后失败和运行前终止不计费,任务成功和运行后终止会计费用

免费额度

每个账号享有一定量免费调用额度,如下表:

服务	免费任务额度	并发任务数限制	说明
货架拼接	累计200次	1	只允许一个拼接任务在运行,超出限制的任务排队等待

注:任务成功与失败调用均消耗免费额度

心 价目表 - 任务次数包

如果对拼接任务次数有预估,可以选择购买**单次任务价格更低**的次数包,价格如下:

规格	价格	并发任务数限制	有效期
1千次	200 元	1	1年
1万次	1,900 元	1	1年
10万次	18,000 元	1	1年
100万次	150,000 元	1	1年
500万次	600,000 元	1	1年

购买后不可退款,任务次数包使用完后,开始按调用量每个任务0.2元收取费用

₯价目表 - 并发任务叠加包

开通付费后,并发任务数限制为1,如果有更多的并发请求需要,可以根据业务需求按天或按月购买并发任务叠加包,价格如下:

购买方式 每并发任务价格

按天购买 2元/天

按月购买 40元/月

购买的并发任务叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

体验APP

心简介

本文档主要说EasyDL零售版的门店拜访体验APP(原货架拼接体验APP)如何获取和使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单,咨询问题类型请选择人工智能服务
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:1009661589)联系群管

^の 生成体验APP

在完成模型发布后,在EasyDL零售版上货架拼接服务页面上可以生成货架拼接体验APP,如下图所示:

模型中心	货架拼接服务					服务介
我的模型 创建模型	下载SDK: 体验货架拼接:	下载iOS SDK 下载Android SDK 生成体验APP	SDK开发文档 API开发文档			
训练模型		绑定的商品检测服务名称	关联的应用名称	支持检测的SKU	状态	操作
校验模型		silencefirst0829	Silence-EasyDL	查看	可用	下载
发布模型		silenceteapi	Silence-EasyDL	查看	可用	下载
数据中心 我的SKU库 实景图上传/标注 云服务调用数据				每页显示 10	v <	1 >
AI服务中心						
服务列表						
增值服务中心						
货架拼接服务						

点击「生成体验APP」按钮后弹出如下图所示弹窗:

生成体验APP		×
* 选择服务	请选择需要绑定的商品检测服务 🗸 🗸 🗸 🗸 🗸 🗸	
	选择的服务用于对拼接的图片进行商品检测,该服务更新后,货架拼接服务会自动更新	
* 选择应用	请选择在控制台创建的应用 🗸 🗸	
	选择的应用用于为体验APP鉴权,如未创建,请前往 <mark>控制台</mark> 进行创建	
	生成体验APP	

需要选择两项内容:

- 选择服务:选择的服务用于对拼接的图片进行商品检测,该服务更新后,货架拼接服务会自动更新。一个商品检测服务只能被绑定到一个货架 拼接体验APP上
- 选择应用:用于为体验APP鉴权,如未创建,请前往EasyDL零售版控制台应用列表进行创建

点击「生成体验APP」按钮后,可以看到页面上出现体验APP列表,可以在列表中看到该体验APP的状态为「生成中」,生成体验APP通常需要五 分钟左右,如下图所示

货架拼接服务						服务介绍	
下载SDK:	下载iOS SDK	下载Android SDK	SDK开发文档 API开发文档				
体验货架拼接:	生成体验APP						
	绑定的商品检测服务	务名称	关联的应用名称	支持检测的SKU	状态	操作	
	kelexuebijiance		EasyDL	查看	生成中 🙎	生成体验APP通常需要	五分钟左右

^の使用体验APP

体验APP生成后,在体验APP列表中「操作」一列点击下载后会弹出下载二维码,使用手机摄像头扫码二维码后下载到手机使用。

^の更新体验APP

如果用于绑定体验APP的应用(在百度智能云控制台创建的应用)被删除,体验APP的鉴权将会失效,体验APP列表中的状态将会变为「不可用」,体验APP也将无法使用。如需继续使用该APP,可以在体验APP列表中更新该APP,点击「更新」后在弹窗内选择新的应用即可。

API文档

^の API调用方法

简介

本文档主要说EasyDL零售版的货架拼接服务API如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单,咨询问题类型请选择人工智能服务
- 进入EasyDL社区交流,与其他开发者进行互动

API总览

接口列表

API名称	描述	API
创建任务	开始拼接整个流程	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/create
上传图片	上传货架局部图片	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/upload
开始任务	启动货架拼接离线任务	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/start
查询结果	查询任务运行状态或者结果	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/query
终止任务	终止正在进行或者等待的任务	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/terminate
任务列表	列出所有状态的任务列表	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/list

任务状态 task_status

状态	描述
Created	已创建的任务
Queued	排队中的任务
Running	正在拼接的任务
Success	拼接成功的任务
Failure	拼接失败的任务
Terminated	被手动终止的任务

启动任务后,免费阶段,无论任务成功、失败、终止均会消耗免费任务数;付费使用阶段,仅对拼接成功和手动终止的任务进行计费。

接口鉴权

1. 进入EasyDL零售版的百度智能云控制台应用列表页面,如下图所示:
| | 土向 🗸 | | | | | | |
|-------------|--------------|------|----------------------------|----------|--------------------------|------------|--|
| | ✔ 返回EasyDL总览 | 产品服务 | 产品服务 / EasyDL定制训练平台 - 应用列表 | | | | |
| > | 零售版 | 应用 | 应用列表 | | | | |
| ଙ | 应用列表 | + | + 创建应用 | | | | |
| ⊕ | 用量统计 | | 应用名称 | AppID | API Key | Secret Key | |
| T | 监控报表 | 1 | silence123 | 16935698 | mbqkk0SDDHxrjUuBy2pwVCGO | ******* 显示 | |
| <u>[4</u>] | 定制商品检测服务、 | 2 | Silence-EasyDL | 15637143 | ghcLQC05QsuiZ2LEhjvtTb48 | ****** 显示 | |
| | 标准商品检测服务 | / | | | | | |
| Ē | 货架拼接 | / | | | | | |

2. 如果还未创建应用,请点击「创建应用」按钮进行创建。创建应用后,参考鉴权参考文档,使用API Key(AK)和Secret Key(SK)获取 access_token

API使用方法

创建任务API

接口描述

创建货架拼接任务,开始整个拼接的流程。

请求说明

请求示例

HTTP 方法: POST

 $\texttt{BCURL}: \texttt{https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/create}$

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/jsor

Body中放置请求参数,参数详情如下:

请求参数

参数名称	是否必 需	参数类型	描述	参数值限制
api_url	是	string	商品检测服务的url	无
row_image_nu ms	是	array[num ber]	各行待拼接货架图片的数量,array长度为货架图片的行数,array[i]为 第i行的货架图片数量	行数不大于3,行内图片数量不大于 60
detection_thre shold	否	float	商品检测服务的阈值	默认值为商品检测服务的阈值,取值 范围[0,1]
nms_iou_thres hold	否	float	检测框矫准去重的阈值	默认值为0.3, 取值范围[0.2,0.8]

Body请求示例:

{ "api_url": "http://xxxxx",

"row_image_nums": [3, 3, 4],

"detection_threshold": 0.3,

"nms_iou_threshold" : 0.45

}

EasyDL 零售行业版使用说明

返回说明

返回参数

参数名称	是否必需	参数类型	描述
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码
error_msg	否	string	错误描述
task_id	否	string	新建任务对应的id
log_task_id	否	string	用于demo显示的任务id,用于问题定位
task_status	否	string	任务状态

Body返回示例:

```
{
  "log_id": xxxxx,
  "task_id": "xxxx,
  "log_task_id": "xxx",
  "task_status": "Created"
}
{
  "log_id": xxxx,
  "error_code": 336204,
  "error_msg": "api name authentication failed"
}
```

上传图片API

接口描述

为指定任务上传待拼接的货架图

提示:只有在Created状态的任务才可以上传图片

请求说明

请求示例

HTTP 方法: POST

 $\texttt{BCURL}: \texttt{https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/upload}$

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数值

Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名 称	是否必 需	参数类 型	描述	参数值限制	
task_id	是	string	货架拼接任务id	无	
row	是	number	图片对应行的 index	取值从0开始,需小于创建任务参数row_image_nums的长度	
column	是	number	图片在行内所在 的index	取值从0开始,需小于创建任务参数row_image_nums[row]的取值	
image	是	string	上传图片的 base64编码	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大 4096px,支持jpg/png/bmp格式 注意请去掉头部	

Body请求示例:

```
{
    "task_id": "xxxx",
    "row": 1,
    "column": 2,
    "image": "xxx=="
}
```

返回说明

接口返回参数:

参数名称	是否必需	参数类型	描述
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码
error_msg	否	string	错误描述
task_status	否	string	任务状态

Body返回示例:

```
{
    "log_id": xxxx,
    "task_status": "Created"
}
{
    "log_id": xxxxx,
    "error_code": 336201,
    "error_msg": "unknown task id"
}
```

开始任务API

接口描述

开始执行货架拼接任务

提示:只有在Created状态的任务可以启动,若启动任务数到达用户的上限(默认为1,即同时只可以启动一个拼接任务),任务进入Queued 状态。

请求说明

请求示例

HTTP 方法: POST

 $\label{eq:com_rec} \begin{array}{l} & \texttt{B} \end{tabular} \\ & \texttt{B} \end{tabular$

URL参数:

参数 值

access_token 通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/jsor

Body中放置请求参数,参数详情如下:

请求参数

参数名称	是否必需	参数类型	描述
task_id	是	string	货架拼接任务id

Body请求示例:

{
 "task_id": "xxxx"
}

返回说明

返回参数

参数名称	是否必需	参数类型	描述
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码
error_msg	否	string	错误描述
task_status	否	string	任务状态
missed_location	否	array	缺失图片对应行和列的index

Body返回示例:

```
{# 启动成功
   "log_id": xxxx,
   "task_status": "Running"
}
{# 用户已运行的货架拼接任务已达上限,排队等待
   "log_id": xxxx,
   "task_status": "Queued"
}
{# 货架图片未全部上传
   "log_id": xxxx,
   "error_code": 336211,
   "error_msg": "some images missed",
   "missed_location": [[0, 2], [1, 3]] # [[row, column]...]
}
```

查询结果API

接口描述

查询任务运行的状态或者结果信息

请求说明

请求示例

HTTP 方法: POST

接口URL:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/query

URL参数:

access_token 通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名称	是否必需	参数类型	描述	参数值限制
task_id	是	string	货架拼接任务id	无

Body请求示例:

{ "task_id": "xxxx" }

返回说明

返回参数

参数名称	是否必需	参数类型	描述
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码
error_msg	否	string	错误描述
log_task_id	否	string	用于demo显示的任务id,用于问题定位
task_status	否	string	任务状态
task_result	否	dict	任务拼接结果
+image_url	否	string	拼接成功后大图的url
+preview_image_url	否	string	拼接成功后大图预览图的url,压缩到1M以下,用于快速预览
+bbox	否	array	在拼接大图上的商品检测框
++name	否	string	商品名称
++score	否	float	检测框置信度
++sku_code	否	string	商品对应的sku code
++location	否	dict	检测框的位置
+++left	否	number	检测框的左上角像素点的横坐标
+++top	否	number	检测框的左上角像素点的纵坐标
+++height	否	number	检测框的高度
+++width	否	number	检测框的宽度
+sku_stat_info	否	array	在拼接大图上的商品检测框的统计信息
++name	否	string	商品名称
++sku_code	否	string	商品对应的sku code
++count	否	number	检测对应商品的数量
++proportion	否	float	统计商品在完整图片中的排面占比
+stitch_error_code	否	array[number]	拼接错误码
+fail_msg	否	string	拼接失败的错误信息

stitch_error_code取值

stitch_error_code	描述
0	拼接成功
100	水平矫正失败
200	竖直矫正失败
300	拼接失败,可能原因相邻图像重叠度不足30%
400	显存不足(OOM),图片数量过多
500	GPU所能分配的单张图片的显存不足,单张图片太大

Body返回示例:

```
{#任务(已创建/排队中/运行中/已取消)
  "log_id": xxxx,
"log_task_id": "xxx",
"task_status": "Created/Queued/Running/Terminated"
}
{#拼接任务运行成功
  "log_id": xxxx,
"log_task_id": "xxx",
"task_status": "Success",
"task_result": {
 "image_url": "https://xxxx",
 "preview_image_url": "https://xxxx",
 "bbox": [{
   "name": "xxx",
   "score": xxx,
  "sku_code": "xxx",
  "location": {
       "left": xxx,
       "top": xxx,
       "width": xxx,
       "height": xxx
  }
 }, ... ],
  "sku_stat_info":[{
  "sku_code": "xxx",
   "name": "xxx",
  "count": n
 }, ...],
  "stitch_error_code": [100, ...]
}
}
{#拼接任务运行失败
  "log_id": xxxx,
"log_task_id": "xxx",
  "task_status": "Failure",
  "task_result": {
   "fail_msg": "image stitch job running timeout"
}
}
{#请求错误
  "log_id": xxxx,
  "error_code": 336201,
  "error_msg": "unknown task id"
}
```

终止任务API

接口描述

终止正在进行或者排队的任务

请求示例

HTTP 方法: POST

 $\texttt{BCURL}: \texttt{https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/terminate}$

URL参数:

参数		值
acces	ss_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名称	是否必需	参数类型	描述	参数值限制
task_id	是	string	货架拼接任务id	无

Body请求示例:

{ "task_id": "xxxx" }

返回说明 返回参数

参数名称	是否必需	参数类型	描述
log_id	是	number	唯一的log id ,用于问题定位
error_code	否	number	错误码
error_msg	否	string	错误描述
task_status	否	string	任务状态

Body返回示例:

```
{
    "log_id": xxxx,
    "task_status": "Terminated"
}
```

任务列表API

接口描述

根据查询条件查询任务列表,多个条件取交集;按照创建时间倒序。

请求说明

请求示例

HTTP 方法: POST

接口URL:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/list

URL参数:

参数 值

access_token 通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名称	是否必需	参数类型	描述	参数值限制
task_ids	否	array[string]	只返回指定id的任务信息	无
begin_time	否	number	只返回begin_time以后创建的任务信息	时间戳
end_time	否	number	只返回end_time之前创建的任务信息	时间戳

Body请求示例:

```
{
"task_ids": ["xx", "xxx"],
"begin_time": 1562763431,
"end_time": 1562763842
}
```

返回说明

返回参数

参数名称	是否必需	参数类型	描述
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码
error_msg	否	string	错误描述
tasks_info	否	array	任务列表
+task_id	否	string	任务id
+log_task_id	否	string	用于demo显示的任务id,用于问题定位
+task_status	否	string	任务状态
+create_time	否	number	任务创建时间

Body返回示例:

```
{
  "log_id": xxxx,
  "tasks_info": [
  {
    "task_id": "xxx",
    "log_task_id": "xxx",
    "task_status": "Created/Queued/Running/...",
    "create_time": 1562763842
  }, ...
]
}
```

心错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code : 错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如图片大小超4M导致报错,错误返回以下内容:

```
{
    "error_code": 336210,
    "error_msg": "invalid image size"
}
```

货架拼接服务错误码如下表:

error_c ode	error_msg	描述
336200	internal error	内部错误,如果持续出现此类错误,请通过QQ群(群号:1009661589)或工单联系技术支持团队。
336201	unknown task id	未知的任务id
336202	invalid param: 'param_name'	请求参数'param_name'的参数值不合法
336203	missing param: 'param_name'	请求参数'param_name'缺失
336204	api name authentication failed	api name和app_id认证失败
336205	current task status not	当前任务状态不支持对应的操作:只有Created状态下的任务可以进行加图和启动任
000200	support {action}.	务,Created/Queued/Running状态下的任务可被终止
336206	invalid base64	加图操作:错误的base64图片编码
336207	failed loading image	加图操作:加载图片失败
336208	invalid image format	加图操作:不支持的图片格式,支持格式: bmp、jpg、jpeg、png
336209	invalid image shape	加图操作:不支持的图片形状,货架图片长宽需大于等于200
336210	invalid image size	加图操作:不支持的图片大小,图片大小不超过4M
336211	some images missed	启动拼接任务:货架图片缺失
336212	invalid json	请求数据格式不正确

心 API调用方法

简介

本文档主要说EasyDL零售版的货架拼接服务API如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单,咨询问题类型请选择人工智能服务
- 进入EasyDL社区交流,与其他开发者进行互动

API总览

接口列表

API名称	描述	API
创建任务	开始拼接整个流程	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/create
上传图片	上传货架局部图片	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/upload
开始任务	启动货架拼接离线任务	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/start
查询结果	查询任务运行状态或者结果	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/query
终止任务	终止正在进行或者等待的任务	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/terminate
任务列表	列出所有状态的任务列表	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/list

任务状态 task_status

状态	描述
Created	已创建的任务
Queued	排队中的任务
Running	正在拼接的任务
Success	拼接成功的任务
Failure	拼接失败的任务
Terminated	被手动终止的任务

启动任务后,免费阶段,无论任务成功、失败、终止均会消耗免费任务数;付费使用阶段,仅对拼接成功和手动终止的任务进行计费。

接口鉴权

1. 进入EasyDL零售版的百度智能云控制台应用列表页面,如下图所示:

Ŷ	全局 🗸							
	✔ 返回EasyDL总览	7	产品服务 / EasyDL定制训练平台 - 应用列表					
>	零售版		应用列表					
¢	应用列表		+ 创建应用					
⊕	用量统计			应用名称		AppID	API Key	Secret Key
T	监控报表		1	silence123		16935698	mbqkk0SDDHxrjUuBy2pwVCGO	******* 显示
[4]	定制商品检测服务	~	2	Silence-EasyDL		15637143	ghcLQC05QsuiZ2LEhjvtTb48	****** 显示
88	标准商品检测服务	~						
₿	货架拼接	~						

2. 如果还未创建应用,请点击「创建应用」按钮进行创建。创建应用后,参考鉴权参考文档,使用API Key(AK)和Secret Key(SK)获取 access_token

API使用方法

创建任务API

接口描述

创建货架拼接任务,开始整个拼接的流程。

请求说明

请求示例

HTTP 方法: POST

接口URL:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/create

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名称	是否必 需	参数类型	描述	参数值限制
api_url	是	string	商品检测服务的url	无
row_image_nu ms	是	array[num ber]	各行待拼接货架图片的数量,array长度为货架图片的行数,array[i]为 第i行的货架图片数量	行数不大于3,行内图片数量不大于 60
detection_thre shold	否	float	商品检测服务的阈值	默认值为商品检测服务的阈值,取值 范围[0,1]
nms_iou_thres hold	否	float	检测框矫准去重的阈值	默认值为0.3, 取值范围[0.2,0.8]

Body请求示例:

```
{
   "api_url": "http://xxxxx",
   "row_image_nums": [3, 3, 4],
   "detection_threshold": 0.3,
   "nms_iou_threshold": 0.45
}
```

返回说明

返回参数

参数名称	是否必需	参数类型	描述
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码
error_msg	否	string	错误描述
task_id	否	string	新建任务对应的id
log_task_id	否	string	用于demo显示的任务id,用于问题定位
task_status	否	string	任务状态

Body返回示例:

```
{
  "log_id": xxxxx,
  "task_id": "xxxx",
  "log_task_id": "xxxx",
  "task_status": "Created"
}
{
  "log_id": xxxx,
  "error_code": 336204,
  "error_msg": "api name authentication failed"
}
```

上传图片API

接口描述

为指定任务上传待拼接的货架图

提示:只有在Created状态的任务才可以上传图片

请求说明

请求示例

HTTP 方法: POST

 $\texttt{BCURL}: \texttt{https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/upload}$

URL参数:

蚁 但

access_token 通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名 称	是否必 需	参数类 型	描述	参数值限制
task_id	是	string	货架拼接任务id	无
row	是	number	图片对应行的 index	取值从0开始,需小于创建任务参数row_image_nums的长度
column	是	number	图片在行内所在 的index	取值从0开始,需小于创建任务参数row_image_nums[row]的取值
image	是	string	上传图片的 base64编码	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大 4096px,支持jpg/png/bmp格式 注意请去掉头部

Body请求示例:

```
{
    "task_id": "xxxx",
    "row": 1,
    "column": 2,
    "image": "xxx=="
}
```

返回说明

接口返回参数:

参数名称	是否必需	参数类型	描述
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码
error_msg	否	string	错误描述
task_status	否	string	任务状态

Body返回示例:

```
{
    "log_id": xxxx,
    "task_status": "Created"
}
{
    "log_id": xxxxx,
    "error_code": 336201,
    "error_msg": "unknown task id"
}
```

开始任务API

开始执行货架拼接任务

提示:只有在Created状态的任务可以启动,若启动任务数到达用户的上限(默认为1,即同时只可以启动一个拼接任务),任务进入Queued 状态。

请求说明

请求示例

HTTP 方法: POST

 $\texttt{BCURL}: \texttt{https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/start}$

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名称	是否必需	参数类型	描述
task_id	是	string	货架拼接任务id

Body请求示例:

```
{
"task_id": "xxxx"
}
```

返回说明

返回参数

参数名称	是否必需	参数类型	描述
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码
error_msg	否	string	错误描述
task_status	否	string	任务状态
missed_location	否	array	缺失图片对应行和列的index

Body返回示例:

```
{# 启动成功
   "log_id": xxxx,
   "task_status": "Running"
}
{# 用户已运行的货架拼接任务已达上限,排队等待
   "log_id": xxxx,
   "task_status": "Queued"
}
{# 货架图片未全部上传
   "log_id": xxxx,
   "error_code": 336211,
   "error_msg": "some images missed",
   "missed_location": [[0, 2], [1, 3]] # [[row, column]...]
}
```

查询结果API

接口描述

查询任务运行的状态或者结果信息

请求说明

请求示例

HTTP 方法: POST

接口URL:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/query

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名称	是否必需	参数类型	描述	参数值限制
task_id	是	string	货架拼接任务id	无

Body请求示例:

```
{
"task_id": "xxxx"
}
```

返回说明

返回参数

参数名称	是否必需	参数类型	描述
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码
error_msg	否	string	错误描述
log_task_id	否	string	用于demo显示的任务id,用于问题定位
task_status	否	string	任务状态
task_result	否	dict	任务拼接结果
+image_url	否	string	拼接成功后大图的url
+preview_image_url	否	string	拼接成功后大图预览图的url,压缩到1M以下,用于快速预览
+bbox	否	array	在拼接大图上的商品检测框
++name	否	string	商品名称
++score	否	float	检测框置信度
++sku_code	否	string	商品对应的sku code
++location	否	dict	检测框的位置
+++left	否	number	检测框的左上角像素点的横坐标
+++top	否	number	检测框的左上角像素点的纵坐标
+++height	否	number	检测框的高度
+++width	否	number	检测框的宽度
+sku_stat_info	否	array	在拼接大图上的商品检测框的统计信息
++name	否	string	商品名称
++sku_code	否	string	商品对应的sku code
++count	否	number	检测对应商品的数量
++proportion	否	float	统计商品在完整图片中的排面占比
+stitch_error_code	否	array[number]	拼接错误码
+fail_msg	否	string	拼接失败的错误信息

stitch_error_code取值

stitch_error_code	描述
0	拼接成功
100	水平矫正失败
200	竖直矫正失败
300	拼接失败,可能原因相邻图像重叠度不足30%
400	显存不足(OOM),图片数量过多
500	GPU所能分配的单张图片的显存不足,单张图片太大

Body返回示例:

```
{#任务(已创建/排队中/运行中/已取消)
  "log_id": xxxx,
"log_task_id": "xxx",
"task_status": "Created/Queued/Running/Terminated"
}
{#拼接任务运行成功
  "log_id": xxxx,
"log_task_id": "xxx",
"task_status": "Success",
"task_result": {
 "image_url": "https://xxxx",
 "preview_image_url": "https://xxxx",
 "bbox": [{
   "name": "xxx",
   "score": xxx,
 "sku_code": "xxx",
 "location": {
      "left": xxx,
      "top": xxx,
       "width": xxx,
       "height": xxx
  }
 }, ... ],
 "sku_stat_info":[{
 "sku_code": "xxx",
   "name": "xxx",
  "count": n
 }, ...],
 "stitch_error_code": [100, ...]
}
}
{#拼接任务运行失败
  "log_id": xxxx,
"log_task_id": "xxx",
  "task_status": "Failure",
  "task_result": {
   "fail_msg": "image stitch job running timeout"
}
}
{#请求错误
  "log_id": xxxx,
  "error_code": 336201,
  "error_msg": "unknown task id"
}
```

终止任务API

接口描述

终止正在进行或者排队的任务

请求说明

请求示例

HTTP 方法: POST

接口URL:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/terminate

URL参数:

access_token 通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名称	是否必需	参数类型	描述	参数值限制
task_id	是	string	货架拼接任务id	无

Body请求示例:

```
{
"task_id": "xxxx"
}
```

返回说明

返回参数

参数名称	是否必需	参数类型	描述
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码
error_msg	否	string	错误描述
task_status	否	string	任务状态

Body返回示例:

```
{

"log_id": xxxx,

"task_status": "Terminated"

}
```

任务列表API

接口描述

根据查询条件查询任务列表,多个条件取交集;按照创建时间倒序。

请求说明

请求示例

HTTP 方法: POST

接口URL:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_stitch/list

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名称	是否必需	参数类型	描述	参数值限制
task_ids	否	array[string]	只返回指定id的任务信息	无
begin_time	否	number	只返回begin_time以后创建的任务信息	时间戳
end_time	否	number	只返回end_time之前创建的任务信息	时间戳

Body请求示例:

```
{
"task_ids": ["xx", "xxx"],
"begin_time": 1562763431,
"end_time": 1562763842
}
```

返回说明

返回参数

参数名称	是否必需	参数类型	描述
log_id	是	number	唯一的log id,用于问题定位
error_code	否	number	错误码
error_msg	否	string	错误描述
tasks_info	否	array	任务列表
+task_id	否	string	任务id
+log_task_id	否	string	用于demo显示的任务id,用于问题定位
+task_status	否	string	任务状态
+create_time	否	number	任务创建时间

Body返回示例:

```
{
  "log_id": xxxx,
  "tasks_info": [
  {
    "task_id": "xxx",
    "log_task_id": "xxx",
    "task_status": "Created/Queued/Running/...",
    "create_time": 1562763842
  }, ...
]
}
```

SDK文档

ு SDK介绍

简介

本文档介绍门店拜访(原货架拼接)iOS、Android SDK的获取方式和系统支持,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单,咨询问题类型请选择人工智能服务
- 进入EasyDL社区交流,与其他开发者进行互动
- 加入EasyDL官方QQ群(群号:1009661589)联系群管

获取SDK

在EasyDL零售版的门店拜访SDK页面(原货架拼接服务)直接下载,如下图所示:

Ea	isyDL零售版		门店拜访SDK						操作	乍文档	常见问题
Ð] 模型中心		功能介绍								
	我的模型			92/80/11 SQ							
	创建模型		门店拜访SDK集成了适用于快消场景的AI能力,保障快消业务代表在线下拜访门店过程中的行为规范和业务需要。 1.门脸文字识别:支持识别门脸图片中门店各称和其他文字信息、适用于门店拜访签到和新门店注册验真								
	训练模型		 2.货架拼接:支持 3.拍摄行为引导: 	寺在手机端将实时拍摄的货架图片进行拼接,最终获得将重复区域去重后的商品识别结果,适用于在长货架进行商品识别的业 · 支持拍照过程中对光线、手机方向、照片是否模糊,照片是否被替换进行检测,保障拍摄的照片真实有效						场景	
	校验模型										
	发布模型		下载SDK:	下载iOS SDK 下载Andr	oid SDK SDK开发文	挡 API开发文档					
11	〕数据中心		体验货架拼接:	生成体验APP							
	我的SKU库										
	实景图集库			绑定的商品检测服务名称	关联的应用名称	版本	支持检测的SKU	状态	操作		
	实景图标注			标准服务饮品检测	Silence-EasyDL	iOS: V5.0.0	-	可用	下载		
	云服务数据回流					Android: V5.0.0					
Pl	】 AI服务中心			111	Silence-EasyDL	iOS: V1.0.0 Android: V1.0.0	查看	可用	下载 更新		
	服务列表										
Ľ	增值服务中心			silencefirst0829	Silence-EasyDL	iOS: V1.0.0 Android: V1.0.0	查看	可用	下载 更新		
	门店拜访SDK(原货架	拼接)				100.144.0.0					
				silenceteapi	Silence-EasyDL	Android: V1.0.0	查看	可用	下载 更新		
							< 1		10条/页 ~		

iOS系统支持

系统:iOS 9.0 以上

硬件:armv7 arm64 (Starndard architectures) (暂不支持模拟器)

Android系统支持

系统: Android Level 22以上

രം iOS_SDK

简介

本文档描述货架拼接iOS SDK如何使用。

系统支持

系统:iOS 9.0 以上

硬件:armv7 arm64 (Starndard architectures) (暂不支持模拟器)

Release Notes

时间	版本	说明
2022.12.22	5.0.0	更名为门店拜访SDK,新增门脸文字识别功能、防窜拍功能
2021.12.22	4.1.0	新增手机端实时拼接模糊图像检测功能
2021.10.20	4.0.0	新增手机端实时拼接功能
2021.03.09	3.0.1	新增光线和手机方向检测功能
2020.12.30	3.0.0	新增支持拍摄图片,云端拼接功能
2020.11.12	2.0.0	新增支持排面统计占比
2019.08.30	1.0.0	支持拍摄视频,端上抽帧,云端拼接

集成指南 库依赖

SDK依赖以下静态库/动态库,需正确集成至项目中并配置Framework Search Paths / Header Search Paths / Library Search Paths :

- opencv2.framework:OpenCV V4.5.2,必须引入
- libmontage_algo.a:手机端实时拼接功能库,可选引入,集成时请一并拷贝头文件目录montage_algo至项目合适路径
- libEasyDL.a:模糊图像检测引擎库,可选引入,集成时请一并拷贝头文件目录EasyDL至项目合适路径
- libpaddle_api_full_bundled.a:模糊图像检测引擎库,可选引入,集成时请一并拷贝头文件目录paddlelite至项目合适路径

集成摄像头相关逻辑

UI部分包括摄像头代码均开源,可参考以下文件,用户拷贝相关代码至项目中即可,并修改相应文件名以避免符号冲突:

- easydl-stitch-ios/ViewController/ImagePickerViewController:云端拼接拍照逻辑,重合度算法开源
- easydl-stitch-ios/ViewController/MBStitchCameraViewController:手机端实时拼接拍照逻辑,拼接等相关算法依赖 libmontage_algo.a
- easydl-stitch-ios/ViewController/VideoStitchViewController#startUlImagePicker():云端拼接视频逻辑入口,参考该方法内对系统 UlImagePickerController的使用

拍照拼接参数配置

// easydl-stitch-ios/EasyDLStitch/EasyDLStitchParams.h

define kThreshold 75 // 判断重合的阈值,0~100之间 ###### define kStrategy "phash" // 重合算法,类型:["ahash","phash","dhash"] ###### define kThetaZ 60 // 手机倾斜Z轴角度阈值 ###### define kThetaXY 20 // 手机倾斜XY轴角度阈值 ###### define kOverLapControl true // 是否与遮罩重合才可以拍照 ###### define kOrientationControl true // 是否手机持握方向符合要求才可以拍照 ###### define kSkipFrames 3 // 跳过视频帧比对的数量,比如3为每3帧比对一次 ###### define kBrightnessLow -3 // 光线强度阈值,-99~99之间 ###### define kBrightnessHigh 5 // 光线强度阈值,-99~99之间 ###### define kBrightnessControl false // 是否光线符合要求才可以拍照

参数说明:

- 获取重合度有"ahash","phash","dhash"三种算法,返回0~100之间的数值,越大表示重合度越高。不同算法返回的数值有区别,需相应调整 阈值
- 手机倾斜XY轴指左右倾斜,Z轴是前后倾斜,当倾斜角度过大会影响拼接效果
- 设置只有手机倾斜角度、待拍摄图片与上一张图片重合度、环境光线亮度等条件符合要求才拍摄图片,保证拍摄效果
- 相机默认为每秒30帧,修改kSkipFrames的值调节做重合度对比的速度,避免卡顿或拍摄状态切换过快

视频拼接参数配置

体验APP中对视频截取帧的频率为1秒1帧,由于每个视频的帧数不能大于60,所以体验APP不能拼接长度大于60s的视频。开发者可根据实际情况 调整截帧的频率,并相应限制视频长度。调整频率方法:

// easydl-stitch-ios/ViewController/StitchViewController.m

static int frameInterval = 1;//截帧间隔(秒)

并在合适的地方提示视频长度限制:

UIAlertController *actionSheet = [UIAlertController alertControllerWithTitle:@"选择对象(视频长度不能超过60s)" message:nil preferredStyle:UIAlertControllerStyleActionSheet];

SDK工程结构

```
EasyDL-Image-Stitching-iOS
|- LIB
 - include
   |- montage_algo/ // 手机端实时拼接功能库头文件
    |- EasyDL/ // 手机端实时拼接模糊图像检测引擎头文件
   |- paddlelite/ // 手机端实时拼接模糊图像检测引擎头文件
 - libs
   |- opencv2.framework // OpenCV库
|- libmontage_algo.a // 手机端实时拼接功能库
|- libEasyDL.a // 手机端实时拼接模糊图像检测引擎
    |- libpaddle_api_full_bundled.a // 手机端实时拼接模糊图像检测引擎
- EasyDLStitch
 - images/
                // 资源文件
 |- easydl-stitch-ios/ // Demo工程文件
 - RES/
    |- conf.json // API配置文件
    |- fuzzy_model/ // 模糊图像检测模型
```

SDK调用流程

获取鉴权

1. 进入EasyDL零售版的百度智能云控制台应用列表页面,如下图所示:

Û	全局 🗸							
	✔ 返回EasyDL总览	Ē	产品服务 /	EasyDL定制训练平台 - 应用列表				
>	零售版		应用歹	川表				
¢	应用列表		+ 81	建应用				
\oplus	 全局 > 《返回EasyDL总览 零售版 应用列表 用量统计 监控报表 定制商品检测服务 > 抗准商品检测服务 > 货架拼接 > 			应用名称	Appl	ID	API Key	Secret Key
T	监控报表		1	silence123	1693	35698	mbqkk0SDDHxrjUuBy2pwVCGO	****** 显示
<u>[4</u>]	定制商品检测服务	~	2	Silence-EasyDL	1563	37143	ghcLQC05QsuiZ2LEhjvtTb48	******* 显示
8	标准商品检测服务	ř						
111	贞栄拼按	~						

2. 如果还未创建应用,请点击「创建应用」按钮进行创建。创建应用后,参考鉴权参考文档,使用API Key(AK)和Secret Key(SK)获取 access_token

下载SDK包后,填写ak、sk等信息。在RES/conf.json相应位置填入:



云端非实时拼接调用流程

- 1. 创建任务:开始拼接整个流程
- 2. 加货架图:上传图片
- 3. 开始任务:启动货架拼接离线任务
- 4. 查询任务:查询拼接任务的状态和结果

```
其他:
```

- 1. 取消任务: 取消正在进行或者等待的任务
- 2. 任务列表:查询所有状态的任务列表
- 3. 终止任务:终止正在进行或者等待的任务

调用API

云端拼接API的调用逻辑已封装在EasyDLStitchApiService文件中,以下为SDK调用货架拼接API的方法说明。货架拼接API接口的返回值及其他信息 参见文档货架拼接API调用方法。

创建拼接任务

- (void)createSpliceTaskWithConfig:(NSDictionary *)config successHandler:(SuccessBlock)successHandler failHandler: (FailureBlock)failHandler;

其中config为参数,后面两个回调block。参数取值及描述:

参数名称	是否必 需	参数类型	描述	参数值限制
api_url	是	string	商品检测服务的url	无
row_image_nu ms	是	array[num ber]	各行待拼接货架图片的数量,array长度为货架图片的行数,array[i]为 第i行的货架图片数量	行数不大于3,行内图片数量不大于 60
detection_thre shold	否	float	商品检测服务的阈值	默认值为商品检测服务的阈值,取值 范围[0,1]
nms_iou_thres hold	否	float	检测框矫准去重的阈值	默认值为0.45, 取值范围[0.2,0.8]

上传图片

- (void)uploadImageWithConfig:(NSDictionary *)config successHandler:(SuccessBlock)successHandler failHandler: (FailureBlock)failHandler;

其中config为参数,后面两个回调block。参数取值及描述:

参数名称	是否必需	参数类型	描述	参数值限制
task_id	是	string	货架拼接任务id	无
row	是	number	图片对应行的index	取值从0开始,需小于创建任务参数row_image_nums的长度
column	是	number	图片在行内所在的index	取值从0开始,需小于创建任务参数row_image_nums[row]的取值
image	是	string	上传图片的base64编码	

启动拼接任务

-(void)startSpliceTaskWithConfig:(NSDictionary *)config successHandler:(SuccessBlock)successHandler failHandler: (FailureBlock)failHandler;

其中config为参数,后面两个回调block。参数取值及描述:

参数名称	是否必需	参数类型	描述	参数值限制
task_id	是	string	货架拼接任务id	无

查询任务状态

-(void)queryTaskResultWithConfig:(NSDictionary *)config successHandler:(SuccessBlock)successHandler failHandler: (FailureBlock)failHandler;

其中config为参数,后面两个回调block。参数取值及描述:

参数名称	是否必需	参数类型	描述	参数值限制
task_id	是	string	货架拼接任务id	无

查询任务列表

-(void)listTaskWithConfig:(NSDictionary *)config successHandler:(SuccessBlock)successHandler failHandler:(FailureBlock)failHandler;

其中config为参数,后面两个回调block。参数取值及描述:

参数名称	是否必需	参数类型	描述	参数值限制
task_ids	否	array[string]	只返回指定id的任务信息	无
begin_time	否	number	只返回begin_time以后创建的任务信息	时间戳
end_time	否	number	只返回end_time之前创建的任务信息	时间戳

终止任务

-(void)terminateTaskWithConfig:(NSDictionary *)config successHandler:(SuccessBlock)successHandler failHandler: (FailureBlock)failHandler;

其中config为参数,后面两个回调block。参数取值及描述:

参数名称	是否必需	参数类型	描述	参数值限制
task_id	是	string	货架拼接任务id	无

手机端实时拼接调用流程

1. 准备目录

- 2. 对比图片:新图片(now)在参与实时拼接前需先与上一张参与拼接的图片(last)进行对比,如果now与last的对比特征合法则可以成功拼接,否则 无法得到理想实时拼接结果
- 3. 实时拼接:SDK会寻找now.jpg并进行实时拼接得到新拼接结果
- 4. 上传云端,得到结果

其他:

1. 撤销拼接:撤销最后一次拼接结果

SDK的调用

手机端实时拼接的调用逻辑已封装在EasyDLMBStitchApiService文件中,包括商品检测API以及拼接算法的调用,以下为SDK调用的方法说明。

准备目录

实时拼接过程中产生的文件需保存在本地,首先初始化准备保存的目录。单个拼接任务的保存目录不可发生变化。

- (void)prepare;

```
// 调用示例
```

[[EasyDLMBStitchApiService sharedService] prepare];

对比图片

- (void)compareImage:(UIImage *)image firstFrame:(BOOL)isFirstFrame completionHandler:(CompletionHandler)completionHandler;

```
// 调用示例
```

```
[[EasyDLMBStitchApiService sharedService] compareImage:image firstFrame:_firstFrameCompare completionHandler:^(id responseObject,
NSError *error) {
    if (error) {
        NSLog(@"%@", error.localizedDescription);
    } else {
        // 回调结果
        CompareResult *compareResult = responseObject;
        // do something
    }
```

}];

参数说明:

[•] image:当前要参与对比的图片

- isFirstFrame:是否是第一帧
 - 第一帧的定义取决于上一张参与拼接的图片(last)是否已经被对比过。假设有图片A和B,先用A与last对比,且last是初次被对比,此时isFirstFrame应为true,再用B与last对比,此时isFirstFrame应为false
- completionHandler:异步回调,在主线程

```
**CompareResult**
```

```
// montage_algo/EasyDLStitchAlgo.h
// 当前图片相对上一张参与拼接的图片的方位
@property(nonatomic) ImageDirection direction;
/**
* 是否需要判断方向,如当拍摄完图片过近时,direction可能由于两图过于相似而不可靠,这种情况不需要判断方向,即该值=false
* 一般direction不可靠时,该值=false
* 对比的两张图是第一次对比时,该值=false
* 当该值=true时,请在调用实时拼接API前确认方法是否合法,否则可能导致拼接失败
*/
@property(nonatomic, getter = needCheckDirection) BOOL checkDirection;
// 对比结果中的方位是否合法,非法的方位将无法完成拼接
@property(nonatomic) BOOL directionValid;
// 两张图片重叠部分的点位
@property(nonatomic, retain) NSArray<NSValue *> *points;
// 两张图片重叠状态
@property(nonatomic) OverlapStatus overlapStatus;
// 最后一次参与拼接的图片序号,从1开始
@property(nonatomic) int lastImgIndex;
```

对比结果的合法性判断参考

```
switch (compareResult.overlapStatus) {
    case OverlapStatus_Correct:
        if (compareResult.needCheckDirection && !compareResult.directionValid) {
            // 非法,当前参与对比的图片方位不正确,无法拼接
        } else {
            // 合法
        }
        break;
    case OverlapStatus_TooFar:
        // 非法,两张图重叠度过低
        break;
    case OverlapStatus_TooClose:
        // 非法,两张图重叠度过高
        break;
}
```

实时拼接

- 建议调用拼接前参考【对比结果的合法性判断】,用不合法的对比结果进行实时拼接将无法获得正确输出
- 将要参与拼接的图片必须命名为"now.jpg"(也可使用EasyDLMBStitchApiService.FILENAME_IMAGE_NOW),并保存在 [EasyDLMBStitchApiService sharedService].currentTask.workDir指向的目录下,否则实时拼接无法正常工作。成功拼接后"now.jpg"会被SDK重 新命名为{index}.jpg,其中{index}代表图片序号。
- 为获得更快的拼接效率,建议减小参与拼接的图片尺寸;为了保证拼接效果,缩放后的图片尺寸应不小于宽648和高864

- (void)stitchImageWithCompareResult:(CompareResult *)compareResult completionHandler:(CompletionHandler)completionHandler;

// 调用示例 [[EasyDLMBStitchApiService sharedService] stitchImageWithCompareResult:_latestCompareResult completionHandler:^(id responseObject, NSError *error) { if (error) { NSLog(@"%@", error.localizedDescription); } else { // 回调结果 StitchResult *stitchResult = responseObject; // do something } }]; // 保存now.jpg示例 NSURL *workDirUrl = [EasyDLMBStitchApiService sharedService].currentTask.workDir; /* 小图为拼接,大图为获得更好的商品检测效果 */ [EasyDLFileManager saveImage:image toUrl:[workDirUrl URLByAppendingPathComponent:FILENAME_IMAGE_NOW] andResizeTo:CGSizeMake(648, 864)]; // SDK默认使用保存的一系列`{index}.jpg`调用商品检测API并取得结果。 // 由于建议减小该系列图片尺寸以获得更优的拼接效率,但同时更小尺寸的图片对商品检测精度有一定影响,因此为提高精度,建议同时保 存最佳尺寸的图片用于上传云端。 [EasyDLFileManager saveImageForBestInfer:image toUrl:[workDirUrl URLByAppendingPathComponent:[NSString

stringWithFormat:@"%@/%@", DIR_NAME_FULL_IMAGE, FILENAME_IMAGE_NOW]]];

参数说明:

- compareResult:对比图片回调返回的结果
- completionHandler:异步回调,在主线程。该回调在拼接任务全部完成后到达,如需更快获取缩略图和完整拼接图,可配置 EasyDLMBAPIStitchDelegate协议

StitchResult

// montage_algo/EasyDLStitchAlgo.h

```
    // 拼接错误码,0表示成功,其他为错误
    @property(nonatomic) int errCode;
    // 最近一张参与拼接的图片的序号
    @property(nonatomic) int latestImgIndex;
    // 缩略拼接图路径
    @property(nonatomic, retain) NSURL *thumbnailUrl;
    // 完整拼接图路径
    @property(nonatomic, retain) NSURL *fullImageUrl;
```

EasyDLMBAPIStitchDelegate

// easydl-montage-ios/EasyDLStitch/EasyDLMBStitchApiService.h

@protocol EasyDLMBAPIStitchDelegate <NSObject>

- (void)onStitchThumbnailGenerated:(NSURL *)thumbnailURL;
- (void)onStitchFullImageGenerated:(NSURL *)fullImageURL;
 @end
- // 配置示例 [EasyDLMBStitchApiService sharedService].stitchDelegate = self;

上传云端,得到结果

- (void)mergeDetectedResultsWithCompletionHandler:(CompletionHandler)completionHandler;

```
// 调用示例
[[EasyDLMBStitchApiService sharedService] mergeDetectedResultsWithCompletionHandler:^(id responseObject, NSError *error) {
    if (error) {
        NSLog(@"%@", error.localizedDescription);
    } else {
        // 回调结果
        MergeResult *mergeResult = responseObject;
        // do something
    }
]];
```

参数说明:

• completionHandler:异步回调,在主线程。该回调在检测任务全部完成后到达,如需感知检测任务的开始和进度更新,可配置 EasyDLMBAPIMergeDelegate协议

MergeResult

// montage_algo/EasyDLStitchAlgo.h

// 错误码,0表示成功,其他为错误
 @property(nonatomic) int errCode;
 // 商品检测并去重合并后的结果
 @property(nonatomic, retain) NSDictionary *correctSKUDict;

EasyDLMBAPIMergeDelegate

// easydl-montage-ios/EasyDLStitch/EasyDLMBStitchApiService.h

@protocol EasyDLMBAPIMergeDelegate <NSObject>

- (void)onMergeResultsStarted:(int)totalImageCount;

- (void)onMergeProgressUpdated:(int)totalImageCount completedCount:(int)completedCount; @end

// 配置示例 [EasyDLMBStitchApiService sharedService].mergeDelegate = self;

撤销拼接

```
SDK支持撤销最后一次拼接结果,如需撤销多张,请多次操作
```

- (void)undoLastTakenImageWithCompletionHandler:(CompletionHandler)completionHandler;

```
// 调用示例
```

 $\label{eq:linear} [[EasyDLMBStitchApiService sharedService] undoLastTakenImageWithCompletionHandler: \cite{theta} (id responseObject, NSError * error) \cite{theta} (id responseObject, NSError *$

```
if (error) {
    NSLog(@"%@", error.localizedDescription);
} else {
    // 回调结果,包含撤销后,当前最后一次参与拼接的图片信息
    lmageInfo *lastImageInfo = responseObject;
    // do something
}
}];
```

参数说明:

• completionHandler:异步回调,在主线程

ImageInfo

 $//\ easydl-montage-ios/EasyDLS titch/EasyDLMBS titchApiService.h$

```
    // 图片序号
    @property(nonatomic) int index;
    // 图片列坐标
    @property(nonatomic) int x;
    // 图片行坐标
    @property(nonatomic) int y;
```

模糊图像检测

手机端实时拼接已接入AI模型以支持模糊图像检测,除参考库依赖正确引入依赖库,需保证 RES/fuzzy_model 目录下的模型文件存在。调用示例 如下,也可参考 EasyDL-Stitch/easydl-stitch-ios/ViewController/MBStitchCameraViewController.m 文件中对 FuzzyModelProxy 的使用:

```
// 模型初始化
- (NSError *)modelInit;
// 推理图像判断是否模糊
- (void)inferImage:(UIImage *)image completionHandler:(void (^)(BOOL fuzzy, NSError *error))completionHandler;
// 调用示例
FuzzyModelProxy *fuzzyModelProxy = [[FuzzyModelProxy alloc] init];
[fuzzyModelProxy modelInit];
if (fuzzyModelProxy && fuzzyModelProxy.engineActive) {
  [fuzzyModelProxy inferImage:image completionHandler:^(BOOL fuzzy, NSError *error) {
     if (!error && !fuzzy) {
       // 图像非模糊
    } else {
       // 图像模糊或推理失败
     }
  }];
}
```

阈值设置

手机端实时拼接支持设置:

- 最小IOU置信度
- 最大IOU置信度
- NMS置信度
- 商品检测API最大重试次数

```
/**
* 最小拼接引导置信度
*/
@property(nonatomic) CGFloat minIOUThreshold;
/**
* 最大拼接引导置信度
*/
@property(nonatomic) CGFloat maxIOUThreshold;
/**
* 去重置信度
*/
@property(nonatomic) CGFloat nmsIOUThreshold;
/**
* 商品检测API重试次数
*/
@property(nonatomic) int detectRetryTimes;
/**
* 商品检测API最大并发数
*/
@property(nonatomic) int maxQPS;
```

// easydl-montage-ios/EasyDLStitch/EasyDLMBStitchApiService.h

门脸文字识别调用流程

- 1. 初始化门店定位
- 2. 门脸图片上传云端

SDK 调用

门脸文字识别流程通过 EasyDLDoorAPIService 调用,具体使用和返回参数见下

初始化门店定位

[[EasyDLDoorAPIService sharedService] startLocation];

门脸图片上传云端,获取门店检测结果

```
// easydl-montage-ios/EasyDLDoor/EasyDLDoorAPIService.h
// 开始门脸文字识别
[[EasyDLDoorAPIService sharedService] detectDoorImage:image];
// 获取门脸识别结果
[EasyDLDoorAPIService sharedService].blockNSDictionary = ^(NSDictionary * _Nonnull blockNSDictionary, NSError * _Nonnull error) {
    if(blockNSDictionary != nil && error == nil) {
        // 门脸图片识别成功
        // 门脸图上识别成功
        // 校验门店结果
    [EasyDLStitchAlgo checkDoorData:documentsDirectory ocrInfo:ocrJson];
    }
}
```

模糊图像检测

门脸文字识别已接入AI模型以支持模糊图像检测,除参考库依赖正确引入依赖库,需保证 RES/fuzzy_model 目录下的模型文件存在。 调用示例可 参考【手机端实时拼接调用流程-模糊图像检测】,也可参考【门脸文字识别调用流程】 EasyDL-Stitch/easydl-stitchios/ViewController/DoorCameraViewController.mm 文件中对 FuzzyModelProxy 的使用。

防止图片窜拍开关参数设置

```
// easydl-montage-ios/EasyDLDoor/CheckImageConfig.h
- (BOOL)getPirateImageCheck;
// 调用示例
/**
* 窜拍开关默认开启
*/
[CheckImageConfig sharedService].pirateImageCheck = true;
// 获取开关状态
[[CheckImageConfig sharedService] getPirateImageCheck];
```

错误码 以下为SDK使用的错误码,API接口错误码参见货架拼接API错误码。

错误码	说明
200002	模型配置错误,请检查传入的配置文件是否有效
100006	API,AK/SK 换取token失败
100007	API,请求 API 失败
100008	API,请求商品检测API失败
200001	手机端实时拼接 - 前端引导对比图像出错
200003	手机端实时拼接 - 撤销上一次拼接结果出错
200004	手机端实时拼接 - 商品检测+去重过程中出错
200005	手机端实时拼接 - 拼接出错
300001	手机端实时拼接 - 模糊图像检测出错

രം Android_SDK

简介

本文介绍SDK的的功能使用,即下载包中的sdk module。

SDK为货架拼接云端非实时API和手机端实时拼接的封装,无任何额外功能。如果有和API文档不符的地方,以SDK为准。

支持Android Level 22及以上编译和使用。

Release Notes

时间	版本	说明
2022.12.22	5.0.0	更名为门店拜访SDK,新增门脸文字识别功能、防窜拍功能
2021.12.22	4.1.0	新增手机端实时拼接模糊图像检测功能
2021.08.20	4.0.0	新增手机端实时拼接功能
2021.03.09	3.0.1	新增光线和手机方向检测功能
2020.12.30	3.0.0	新增支持拍摄图片,云端拼接功能
2020.11.12	2.0.0	新增支持排面统计占比
2019.08.30	1.0.0	支持拍摄视频,端上抽帧,云端拼接

测试

获取鉴权

1. 进入EasyDL零售版的百度智能云控制台应用列表页面,如下图所示:

Ŷ	全局 ~									
	≮ 返回EasyDL总货	đ	产品服务 / EasyDL定制训练平台 - 应用列表							
>	零售版		应用列表							
ଙ	应用列表		+ 创建应用							
⊕	用量统计		_	应用名称	AppID	API Key	Secret Key			
T	监控报表		1	silence123	16935698	mbqkk0SDDHxrjUuBy2pwVCGO	****** 显示			
£4]	定制商品检测服务	~	2	Silence-EasyDL	15637143	ghcLQC05QsuiZ2LEhjvtTb48	****** 显示			
88	标准商品检测服务	~								
Ē	货架拼接	~								

2. 如果还未创建应用,请点击「创建应用」按钮进行创建。创建应用后,参考鉴权参考文档,使用API Key(AK)和Secret Key(SK)获取 access_token

{ "ak": "MzOzhObvEZ6InG1K3renXXXX", // API Key的值 "sk": "188fRHYvLPmIPrNCDpBnkhL3ydXXXXX", // Secret Key的值 "apiUrl": "https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/detection/XXXX" // 定制商品检测服务API

}

正常情况下,启动的app及其功能和扫描二维码一致

demo的请求和结果会放在/sdcard/com.baidu.ai.easydl.montage中

测试云端拼接mini demo

测试app通过后,可以修改app/src/main/AndroidManifest.xml 内的启动app,修改为 "com.baidu.ai.easydl.minidemo.MiniActivity"

MiniActivity中有3个task,测试时需要填入 Appkey, AppSecret, ApiUrl信息

- ApiTestAsyncTask ,测试简单流程。
- QueryAsyncTask ,测试查询列表。
- RequestTestAsyncTask,测试assets/request下的图片输入。这个目录可以从SD卡中/sdcard/com.baidu.ai.easydl.montage/X/request复制。

手机端实时拼接调用流程

第1步:初始化

- 1) 【获取实例】
- 2) 【初始化API】

第2步:对比图片

新图片(now)在参与实时拼接前需先与上一张参与拼接的图片(last)进行对比,如果now与last的对比特征合法则可以成功拼接,否则无法得到理想 实时拼接结果

第3步:实时拼接 SDK会寻找**now.jpg**并进行实时拼接得到新结果

第4步:上传云端,得到结果

SDK 调用

手机端实时拼接各流程通过 MobileStitchAPI 调用,具体使用和返回参数见下

初始化

MobileStitchAPI不支持多线程,且仅有一个实例以保证实时拼接过程中的正确文件操作。

获取实例

/**
* 获取实例
*
* @param appKey 网页上的应用的appkey
* @param secretKey 网页上的应用的appSecret
* @param apiUrl 商品检测服务API
*/
public static MobileStitchAPI getInstance(String appKey, String secretKey, String apiUrl);
/**
* 获取实例
*
* @param appKey 网页上的应用的appkey
* @param secretKey 网页上的应用的appSecret

- * @param apiUrl 商品检测服务API
- * @param numConcurrency 同时调用商品检测API的并发数

```
*/
```

public static MobileStitchAPI getInstance(String appKey, String secretKey, String apiUrl, int numConcurrency);

初始化API

/**

- * 初始化API
- *

*/

^{* @}param workDirPath 保存实时拼接过程产生的各类文件的路径

public void init(String workDirPath);

• 异步回调: MobileStitchAPIListener.onAPIPrepared(int, int[])

对比图片

```
/**

* 对比当前图片与上一张参与拼接的图片

*

* @param currentImgBitmap 当前要参与对比的图片

* @param firstFrame 是否是第一帧

*/

public void compareImages(Bitmap currentImgBitmap, boolean firstFrame);
```

- 第一帧的定义取决于上一张参与拼接的图片(last)是否已经被对比过。假设有图片A和B,先用A与last对比,且last是初次被对比,此时 firstFrame应为true,再用B与last对比,此时firstFrame应为false。
- 异步回调: MobileStitchAPIListener.onImagesCompared(CompareResult)

CompareResult

```
// 当前图片相对上一张参与拼接的图片的方位,参考MobileStitchAPI.DIRECTION_{LEFT|UP|RIGHT|DOWN|UNKNOWN}
public int getDirection();
// 对比结果中的方位是否合法,非法的方位将无法完成拼接
public boolean isDirectionValid();
/***
* 是否需要判断方向,如当拍摄完图片过近时,direction可能由于两图过于相似而不可靠,这种情况不需要判断方向,即该值=false
* 一般direction不可靠时,该值=false
* 对比的两张图是第一次对比时,该值=false
* 当该值=true时,请在调用实时拼接API前确认方法是否合法
*/
public boolean needCheckDirection();
// 两张图片重叠部分的点位
public List<PointF> getPoints();
```

```
switch (compareResult.getOverlapStatus()) {
    case CompareResult.OVERLAP_CORRECT:
    if (compareResult.needCheckDirection() && !compareResult.isDirectionValid()) {
        // 非法,当前参与对比的图片方位不正确,无法拼接
    } else {
        // 合法
    }
    break;
    case CompareResult.OVERLAP_TOO_FAR:
    // 非法,两张图重叠度过低
    break;
    case CompareResult.OVERLAP_TOO_CLOSE:
    // 非法,两张图重叠度过高
    break;
}
```

实时拼接

```
/**
* 拼接图片
*
* @param compareResult 对比图片回调返回的结果
*/
public void stitchImage(CompareResult compareResult);
```

- 建议调用拼接前参考【对比结果的合法性判断】,用不合法的对比结果进行实时拼接将无法获得正确输出
- 将要参与拼接的图片必须命名为"now.jpg"(也可使用MobileStitchAPI.IMAGE_NAME_NOW),并保存在初始化API时的 workDirPath 目录下,否

则实时拼接无法正常工作。成功拼接后"now.jpg"会被SDK重新命名为{index}.jpg,其中{index}代表图片序号。

- 为获得更快的拼接效率,建议减小参与拼接的图片尺寸;为了保证拼接效果,缩放后的图片尺寸应不小于宽648和高864
- 异步回调:
 - 缩略拼接图生成: MobileStitchAPIListener.onStitchThumbnailGenerated(String)
 - 完整拼接图生成: MobileStitchAPIListener.onStitchFullImageGenerated(String)
 - 拼接完成:MobileStitchAPIListener.onStitchCompleted(MobileStitchResult)

MobileStitchResult

// 获取缩略拼接图路径 public String getThumbnailPath();

// 获取完整拼接图路径 public String getFullImgPath();

// 获取最近一张参与拼接的图片的序号 public int getLatestPhotoIndex();

// 拼接是否成功 public boolean isSuccess();

保存最佳尺寸的图片以提高商品检测精度

SDK默认使用以上保存的一系列{index}.jpg调用商品检测API并取得结果,由于建议减小该系列图片尺寸以获得更优的拼接效率,但同时更小尺寸的图片对商品检测精度有一定影响,因此为提高精度,建议同时保存最佳尺寸的图片用于上传云端。

```
// 1.原图
Bitmap bitmap = getFromSomewhere();
// 2.计算最佳缩放系数
float scaleFactor = calculateScaleFactor(bitmap);
// 3.缩放获得最佳尺寸的图片
Bitmap scaledBitmap = Bitmap.createScaledBitmap(bitmap,
                              (int) (bitmap.getWidth() * scaleFactor),
                               (int) (bitmap.getHeight() * scaleFactor),
                               true);
// 4.保存
String fullImgFilepath = workDirPath + "/"
               + MobileStitchAPI.DIR_NAME_FULL_IMAGE + "/"
               + MobileStitchAPI.IMAGE_NAME_NOW;
ImageUtil.saveBitmap(scaledBitmap, fullImgFilepath);
/**
* 计算最佳缩放系数
*/
private float calculateScaleFactor(Bitmap originalBitmap) {
  int longerSide;
  int shorterSide;
  if (originalBitmap.getWidth() > originalBitmap.getHeight()) {
     longerSide = originalBitmap.getWidth();
     shorterSide = originalBitmap.getHeight();
  } else {
     longerSide = originalBitmap.getHeight();
     shorterSide = originalBitmap.getWidth();
  return Math.min(1333f / IongerSide, 800f / shorterSide);
```

撤销拼接结果

//最后一张参与拼接的图片路径,workDirPath为初始化时工作目录路径 // latestPhotoIndex为最后一张参与拼接的图片序号,可在 MobileStitchAPIListener 以下回调时赋值 // 1. onAPIPrepared() - 参数takenPhotoSize // 2. onStitchCompleted() - 参数result.getLatestPhotoIndex() // 3. onDeletionConfirmed() - 参数latestPhotoIndex String filepath = workDirPath + "/" + latestPhotoIndex + ".jpg"; // 删除图片 FileUtil.deleteFile(filepath); // 通知SDK // SDK确认删除后回调 MobileStitchAPIListener.onDeletionConfirmed(int, int[]) mobileStitchAPI.notifyLatestPhotoDeletion();

上传云端,得到结果

```
/**
* 上传云端检测,并获得结果
*/
public void mergeDetectResults();
```

- 异步回调:
 - 上传进度更新: MobileStitchAPIListener.onDetectProgressUpdated(int)
 - 检测完成: MobileStitchAPIListener.onDetectedResultsMerged(MergeResult)

MergeResult

```
// 获取商品检测并去重后的结果
public String getCorrectedSKUJson();
```

MobileStitchAPIListener

手机端实时拼接通过 MobileStitchAPIListener 异步回调各函数结果,监听器可通过:

- mobileStitchAPI.registerListener()注册
- mobileStitchAPI.unRegisterListener()注销

```
/**
* API准备好时的回调
*
* @param takenPhotoSize 工作目录下已拍摄的图像数量
* @param latestPhotoPos 最新拍摄图片的坐标
*/
void onAPIPrepared(int takenPhotoSize, int[] latestPhotoPos);
/**
* 图片对比完成
*
* @param compareResult 对比结果
*/
void onImagesCompared(CompareResult compareResult);
/**
*调用拼接接口后,缩略图生成后的回调
*
* @param thumbnailName 缩略图在工作目录下的文件名
*/
void onStitchThumbnailGenerated(String thumbnailName);
/**
*调用拼接接口后,完整拼接图片生成后的回调
*
* @param fullImageName 完成拼接图片在工作目录下的文件名
*/
void onStitchFullImageGenerated(String fullImageName);
/**
* 拼接完成回调
*
* @param mobileStitchResult 拼接结果
*/
void onStitchCompleted(MobileStitchResult mobileStitchResult);
/**
* 删除确认回调
*
* @param latestPhotoIndex -1=操作失败,否则返回删除后,最新拍摄图片的下标;如删除了3.jpg,将返回2
* @param latestPhotoPos 最新拍摄图片的坐标
*/
void onDeletionConfirmed(int latestPhotoIndex, int[] latestPhotoPos);
/**
* 检测图片进度更新回调
*
* @param leftCount 剩余要处理图片的数量
*/
void onDetectProgressUpdated(int leftCount);
/**
* 商品检测并去重处理完成的回调
*
* @param mergeResult 检测并去重结果
*/
void onDetectedResultsMerged(MergeResult mergeResult);
```

模糊图像检测

手机端实时拼接已接入AI模型以支持模糊图像检测,需引入以下依赖库及模型文件:

- libedge-infer.so: 模糊图像检测引擎库
- easyedge-sdk.jar: 模糊图像检测引擎库
- sdk/src/main/assets/infer/: 模糊图像检测模型所在文件夹

以下为调用示例,也可参考 app/src/main/java/com/baidu/ai/easydl/montage/page/photo/mobilestitch/MobileStitchViewPresenter.java 类中

对 FuzzyModelProxy 的使用:

```
/* 初始化 */
FuzzyModelProxy fuzzyModelStateListener = new FuzzyModelProxy.ModelStateListener() {
  @Override
  public void onInitialized(Exception exception) {
    if (exception != null) {
       // 模糊模型初始化失败
    } else {
      // 模糊模型初始化成功
  }
  @Override
  public void onDestroyed() {
    // 模糊模型销毁回调
};
FuzzyModelProxy fuzzyModelProxy = new FuzzyModelProxy(mContext, fuzzyModelStateListener);
fuzzyModelProxy.initModel();
/* 调用示例 */
if (fuzzyModelProxy.modelEngineActivate()) {
  Bitmap bitmap = bitmapFromSomewhere();
  fuzzyModelProxy.infer(bitmap, new FuzzyModelProxy.ModelInferListener() {
    @Override
    public void onCompleted(boolean fuzzy) {
       if~(!fuzzy)~\{
         // 图像非模糊
      } else {
         // 图像模糊
    }
    @Override
    public void onException(Exception exception) {
       // 图像检测失败
\} else {
  // 模糊图像推理引擎异常
}
/* 销毁 */
if (fuzzyModelProxy != null) {
  fuzzyModelProxy.destroyModelEngine();
  fuzzyModelStateListener = null;
```

阈值的设置

手机端实时拼接支持设置:

```
1. 最小IOU置信度
```

```
2. 最大IOU置信度
```

```
3. NMS置信度
```

4. 商品检测API最大重试次数

```
/**
* 设置最低iou置信度,需在init()后调用
*
* @param threshold 在0-1范围内有效
*/
public void setMinIouThreshold(float threshold);
/**
* 设置最高iou置信度,需在init()后调用
*
* @param threshold 在0-1范围内有效
*/
public void setMaxIouThreshold(float threshold);
/**
* 设置NMS置信度,需在init()后调用
*
* @param threshold 在0-1范围内有效
*/
public void setNmsIouThreshold(float threshold);
/**
* 设置商品检测API最大重试次数
* @param maxRetryTimes 最大重试次数,<=0无效
*/
public void setMaxRetryTimes(int maxRetryTimes);
```

云端非实时拼接调用流程

第1步, 创建任务, 上传图片

```
1) 【创建任务:开始拼接整个流程】
```

- 2) 【加货架图:上传图片】
- 3) 【开始任务:启动货架拼接离线任务】

```
第2步,不定时查询结果,一般10分钟后有结果参数
【查询结果:查询任务运行状态或者结果】
```

其它可选操作: 【终止任务:终止正在进行或者等待的任务】 【任务列表:查询所有状态的任务列表】

SDK 调用

根据调用流程 , SDK有两种调用方式 :

- StitchApi, api的封装
- StitchTask, StitchApi的封装,避免taskId的传递。一个task对应一个StitchTask

返回参数以及其他信息详见文档货架拼接API调用方法。

StitchApi

```
初始化
```
```
/**
* 初始化
* @param appKey 网页上的应用的appkey
* @param secretKey 网页上的应用的appSecret
*/
public StitchApi(String appKey, String secretKey) {
    super(appKey, secretKey);
}
/**
* 初始化
* @param appKey 网页上的应用的appkey
* @param secretKey 网页上的应用的appSecret
* @param connection 自定义HTTP连接
*/
public StitchApi(String appKey, String secretKey, ISdkConnection connection) {
    super(appKey, secretKey, connection);
}
```

创建任务

public CreateStitchResponse create(CreateStitchRequest request);

// CreateStitchRequest 及 CreateStitchResponse 参数同API文档

同步上传图片

public CommonStitchResponse upload(UploadImageRequest request);

// UploadImageRequest 及 CreateStitchResponse 参数同API文档

设置图片的话,以下2个方法2选1

public void setImageFile(String imageFile) ;

public void setImageInputStream(InputStream inputStream)

异步上传图片

```
public void uploadAsync(UploadImageRequest request,
IApiResponseListener<CommonStitchResponse> listener)
// UploadImageRequest 参数同API文档
```

// IApiResponseListener<CommonStitchResponse> 接口: onSdkResponse(CommonStitchResponse response, String userDefinedRequestId) // 其中userDefinedRequestId是在UploadImageRequest 里面设置的

// 使用 clearAysncQueue()可以清空未开始的任务

开始任务

CommonStitchResponse start(CommonStitchRequest request)

查询结果

public QueryStitchReponse query(CommonStitchRequest request)

任务列表

public ListStitchResponse list(ListStitchRequest request)

Baidu 百度智能云文档

StitchTask

一个任务新建一个StitchTask 调用方式同 StitchApi ,参数中不必设置taskld

AbstractApiRequest

目前Request类的基类。

// 设置自定义请求ld , 调用异步接口的回调使用 public void setUserDefinedRequestId(String userDefinedRequestId)

// 设置是否添加debug日志 public void setEnableDebug(boolean enableDebug)

CommonStitchResponse 及 AbstractApiResponse

// 获取任务状态
public String getTaskStatus();

// 获取logld
public String getLogld();

// 获取服务端返回的原始json public JSONObject getOriginalJson();

// 获取请求 public AbstractApiRequest getRequest();

门脸文字识别调用流程

第1步:初始化

- 1) 【获取实例】
- 2) 【初始化API】

第2步:门脸图片上传云端,获取门脸文字识别结果

- 1) 【门脸文字识别】
- 2) 【释放资源】

SDK 调用

门脸文字识别流程通过 DetectionDoorAPI 调用,具体使用和返回参数见下

初始化

DetectionDoorAPI不支持多线程,且仅有一个实例。

获取实例

```
/**
* 获取实例
```

```
* @param appKey 网页上的应用的appkey
```

```
* @param secretKey 网页上的应用的appSecret
```

*/

public static DetectionDoorAPI getInstance(String appKey, String secretKey);

初始化API

```
/**
* 初始化API
* 建议传参getApplicationContext
* 初始化(文件夹/定位)
*/
public void init(Context context);
```

```
// 注册门脸文字识别监听
public void registerListener(DoorAPIListener listener);
public interface DoorAPIListener {
  /**
  * 识别成功, 返回门脸文字识别结果
  * @param responseJson
  */
 void onDetectSuccess(String responseJson);
 /**
  * 识别异常
  * @param Exception e
  */
  void onException(Exception e);
}
// 销毁监听
public void unRegisterListener();
```

门脸图片上传云端,获取门店检测结果

门脸文字识别

```
/**
* 开始门脸文字识别
* @param bitmap
*/
public void detectDoorImage(Bitmap bitmap);
```

释放资源

// 释放资源 public void destroy();

模糊图像检测

门脸文字识别已接入AI模型以支持模糊图像检测,需引入以下依赖库及模型文件:

- libedge-infer.so: 模糊图像检测引擎库
- easyedge-sdk.jar:模糊图像检测引擎库
- sdk/src/main/assets/infer/: 模糊图像检测模型所在文件夹

调用示例可参考【手机端实时拼接调用流程-模糊图像检测】,也可参考【门脸文字识别调用流程】 app/src/main/java/com/baidu/ai/easydl/montage/page/door/IDoorViewPresenter.java类中对 FuzzyModelProxy 的使用。

集成指南

添加NDK编译架构

SDK依赖OpenCV库,需添加NDK编译选项,支持常用的两个架构,可参考app/build.gradle配置

```
ndk {
abiFilters "arm64-v8a", "armeabi-v7a"
}
```

集成拍照逻辑

查看com.baidu.ai.easydl.montage.page.photo.take 包,里面均为摄像拍照逻辑。

拍照参数设置

package com.baidu.ai.easydl.montage.page.photo;

```
public interface IPhotoParam {
```

```
/**
  * 两张图片的hash算法
 */
 String IMAGE_COMPARE_HASH_METHOD = "pHash"; // pHash,dHash,ahash
 /**
  * 两张图片的hash比较值
  */
 float IMAGE_COMPARE_HASH_CONFIDENCE_THRESHOLD = 0.75f;
 /**
  * 相机的Sensor的旋转误差值, 取值为0-180, 大于180表示忽略
  */
 int SENSOR_ORIENTATION_EVENT_DELTA = 20;
 /**
  * 传感器的SensorY的旋转误差值 , 取值为0-180 , 大于180表示忽略
  */
 double SENSOR_ORIENTATION_SENSOR_Y_DELTA = Math.PI / 6;
 /**
  * 拍照建议的最低亮度值
  */
 double SENSOR_LIGHT_LUMEN_MIN = 100;
 /**
  * 拍照建议的最高亮度值
  */
 double SENSOR_LIGHT_LUMEN_MAX = 500;
 /**
  * 40%图片的透明度
 */
 float IMAGE_SLIDE_TRANSPARENT_ALPHA = 0.5f;
 /**
  * 每行货架最多的照片数量,服务端支持最大60
  */
 int SLOT_MAX_PHOTO_NUM = 60;
 // 下面的参数,请不要修改
 float IMAGE_SLIDE_CROP_RATIO = 0.4f;
 boolean IMAGE_COMPARE_HASH_DEBUG_SAVE_IMAGES = false;
}
```

防止图片窜拍开关参数设置

package com.baidu.ai.utils;

```
public class CheckImageConfig {
    /**
    * 是否开启防窜拍
    * 默认开启
    */
    private volatile boolean pirateImageCheck = true;
    public void setPirateImageCheck(boolean pirateImageCheck) {
        this.pirateImageCheck = pirateImageCheck;
    }
    public boolean getPirateImageCheck() {
        return pirateImageCheck;
    }
}
```

价签识别服务

服务介绍

心简介

价签识别可识别货架和促销活动中的价签信息,可识别各个价签在图片中的像素位置,以及价签内商品名称和价格,可用于洞察商品在线下渠道 分销的价格区间。

心 使用须知

目前该服务处于**邀请使用**状态,请加入QQ群(群号:1009661589)联系管理员或是提交合作咨询申请使用权限,提供公司名称和在EasyDL零售版控制台应用列表创建应用的APPID,如下图:

全局 ~						Q	工单	消息	帮助文档	企业组织	财务	5~
✓ 返回EasyDL总览	产品服务	£另 / EasyOL定新训练平台 - 应用列表										
零售版	应用	应用列表										
应用列表	+	创建应用										
用量统计		应用名称	AppID	API Key	Secret Key	创建时	(6)		操作			
监控报表	1	silence123	16935698	mbqkk0SDDHxrjUuBy2pwVCGO	••••••• 显示	2019-0	08-01 18:4	5:10	报表 管理	删除		
定制商品检测服务 >	2	Silence-EasyDL	15637143	ghcLQC05QsuiZ2LEhjvtTb48	••••••• 显示	2019-0	02-27 11:0	1:26	报表 管理	删除		
50×EMm152858659 ♥											< 1	>

心 相关解决方案

• 数字化访销解决方案

围绕快速消费品企业在线下渠道中的销量逻辑,提供基于AI技术的数字化访销解决方案,对访销过程的精细化管理,通过提升一线业务人员人效,最终实现销量的增长。

心 使用须知

● 服务接口调用方法请见API文档

购买指南

心 开通付费及购买服务

EasyDL零售版的各项服务的「开通付费」、「购买QPS叠加包」、「购买调用量次数包」、「关闭付费」等操作均在EasyDL零售版控制台进行, 只需在相应需要付费使用的接口位置,跟随页面提示完成后续充值及付费,即可完成。

< ™回EasyDL出现	零售版概览									
零售版										
应用列表	产品介绍: (展开)	查看服务功能介绍)								展开
用量统计	可用接口列表									
监控报表	┃ 已上线的定制接口	己上域的定制接口								
定制商品检测服务 >	API	模型ID	模型类型	模型名称	模型版本	状态	调用量限制	QPS限制	开通按量后付费 ?	购买QPS叠加包 ?
标准商品检测服务 >	111	30443	商品检测	1111	V1	●免费使用	剩余免费1000次	不保证并发	开通	购买 配额详情
货架拼接 >	silenceteapi	27185	商品检测	silencesecond	V3	● 付费使用	500次/天免费 + 超 出按量计费	4	终止付费	购买丨配额详情
	silencefirst0829	27135	商品检测	silencefirst	V3	●免费使用	500次/天免费	不保证并发	开通	购买丨配额详情
	↓ 饮品检测									
	API		状态		调用量限制		QPS限制		开通按量后付费 ?	
	饮品检测		● 免费使用		500次/天免费		不保证并发		免费试用	
	日化品检测									
	API		状态		调用量限制		QPS限制		开通按量后付费 ?	
	日化品检测		● 免费使用		剩余免费1000次		不保证并发		免费试用	
	商品陈列翻拍识别									
	API		状态	101 <i>)</i>	用量限制	QPS限制		开通按量后付费	?	购买QPS叠加包 ?
	商品陈列翻拍识别		● 付费使用	瀬は	余免费999次+超出按量计费	t 4		终止付费		购买丨配额详情

Ⅰ○ 计费方式

价签识别目前支持下列三种计费方式:

- 1. 按调用量后付费
- 2. 调用量次数包预付费
- 3. QPS叠加包预付费

∞价目表 - 按调用量后付费

付费调用

每个账户享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用, 价格如下:

调用量	价格	QPS限制	说明
每次调用	0.05元	2	服务器支持每秒处理2次查询

注:调用失败不计费

免费额度

每个账号享有一定量免费调用额度,如下表:

云服务API	免费调用额度	QPS限制	说明	
价签识别	累计1000次	1	服务器支持每秒处理1次查询	

注:成功调用与失败调用均消耗免费额度

₯ 价目表 - 调用量次数包

如果业务上对调用次数有预估,可以选择购买单次调用价格更低的次数包,价格如下:

Baidu 百度智能云文档

规格	价格	QPS限制	有效期
1万次	475 元	2	1年 (366天)
5万次	2,250 元	2	1年(366天)
10万次	4,250 元	2	1年(366天)
20万次	8,000 元	2	1年 (366天)
50万次	18,750 元	2	1年 (366天)
100万次	35,000 元	2	1年 (366天)
500万次	150,000 元	2	1年 (366天)

购买后不可退款,次数包使用完后,开始按调用量每次0.05元收取费用

心价目表 - QPS叠加包

开通付费后,免费QPS为2,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格

按天购买 100元/天

按月购买 2000元/月

购买 QPS 叠加包需保证已开通按量后付费或购买次数包

购买的 QPS 叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

API文档

^の API使用文档

简介

本文档主要说明如何使用价签识别API

获得使用权限

请提交合作咨询申请使用权限,提供公司名称和在EasyDL零售版控制台应用列表创建应用的APPID,如下图:

全局 ~						Q	工单	消息	帮助文档	企业组织	财务	S ~
✓ 返回EasyDL总览	产品服务	1号 / EaryOL定制训练平台。应用列表										
零售版	应用	应用列表										
应用列表	+	创建应用										
用量统计		应用名称	AppID	API Key	Secret Key	创建时	间		操作			
监控报表	1	silence123	16935698	mbqkk0SDDHxrjUuBy2pwVCGO	显示	2019-	08-01 18:4	5:10	报表 管理	删除		
定制商品检测服务 >	2	Silence-EasyDL	15637143	ghcLQC05QsuiZ2LEhjvtTb48	•••••• 显示	2019-	02-27 11:0	1:26	报表 管理	删除		
标准网络控制服务 >> 货架拼接 >>											< 1	>

接口鉴权

1. 进入EasyDL零售版的百度云控制台应用列表页面,如下图所示:

Baidu 百度智能云文档

	王向									
	≮ 返回EasyDL总货	ĉ ř	^全 品服务 / EasyDL定制训练平台 - 应用列表							
>	零售版		应用列表							
¢	应用列表		+ 创建应用							
€	用量统计			应用名称		AppID	API Key	Secret Key		
T	监控报表		1	silence123		16935698	mbqkk0SDDHxrjUuBy2pwVCGO	******* 显示		
[42]	定制商品检测服务	~	2	Silence-EasyDL		15637143	ghcLQC05QsuiZ2LEhjvtTb48	****** 显示		
88	标准商品检测服务	~								
Ð	货架拼接	~								

2. 如果还未创建应用,请点击「创建应用」按钮进行创建。创建应用后,参考鉴权参考文档,使用API Key(AK)和Secret Key(SK)获取 access_token

请求说明

请求示例

价签识别请求URL: https://aip.baidubce.com/rpc/2.0/easydl/v1/retail/price_tag

HTTP 方法: POST

URL参数:

参数	值				

access_token 通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

提示:如果出现336001或336002的错误码很可能是因为请求方式错误,请以json方式请求。

Body请求示例:

```
{
"image": "<base64数据>"
}
```

Body中放置请求参数,参数详情如下:

请求参数

参数	是否必 选	类型	可选值 范围	说明
image	是	string	-	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大4096px,支持 jpg/png/bmp格式 注意请去掉头部

提示:image参数中"去掉头部"指的是图片经base64编码后的头部信息「data:image/jpeg;base64,」,如下图所示:

点击这里选择选择要转换成Base64的图片 C:\fakepath\14458.jpg

a'lbâ

复制 清空

data:image/jpeg;base64, 9j/4RCRRXhpZgAATU0AKgAAAAgAEQEAAAMAAAABDYAAAAEBAAMAAAABEgAAAAECAAMAA AADAAAA2gEGAAMAAAABAAIAAAEOAAIAAAEbm9yAAEPAAIAAAAHAAAA4AEQAAIAAAAJAAAA5wESAAMAAAABBAAEAA AEVAAMAAAABAAMAAAEaAAUAAAABAAAABAAABAABAABAAAAHAAAAABAAA+AEoAAMAAAABAAIAAAEXAAIAAAAABAAEyAAIAA AAUAAABJAITAAMAAAABAAEAAIdpAAQAAAABAAABOIgIAAQAAAABAAABAAAbaAABGACAAIAAhIVUFXRUkASFdJLUFMMDA AAAr8gAAAJxAACvyAAAAnEEFkb2JIIFBob3Rvc2hvcCBDQyAyMDE3IChNYWNpbnRvc2gpADIwMTk6MDQ6MTQgMjE6MTU 6MiaAACWCmaAFAAAAAQAAAvaCnQAFAAAAAQAAAWKIIaADAAAAAQACAACIJwADAAAAQACaAACQAAAHAAAABDAvMT

返回说明

返回参数

字段	是否必选	类型	说明
log_id	是	number	唯一的log id,用于问题定位
words_result_num	否	number	检测到的价签数量
words_result	否	array(object)	识别的价签信息
+name	否	string	价签中的商品名称
+price	否	string	价签中的价格
+brief	否	string	价签中的其它文字信息
+location	否		价签的像素位置
++left	否	number	检测到的目标主体区域到图片左边界的距离
++top	否	number	检测到的目标主体区域到图片上边界的距离
++width	否	number	检测到的目标主体区域的宽度
++height	否	number	检测到的目标主体区域的高度

の错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码	错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(群号:1009661589)或工单联系技术 支持团队。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请通过QQ群(群号:1009661589)或工单联系技术支 持团队。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exsits or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请通过QQ群(群号:1009661589)联系群管手动提额
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线计 费的接口,请通过QQ群(群号:1009661589)联系群管手动提额
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上 线计费的接口,请通过QQ群(群号:1009661589)联系群管手动提额
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336000	Internal error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(群号:1009661589)或工单联系技术 支持团队
336001	Invalid Argument	入参格式有误,比如缺少必要参数、图片base64编码错误等等,可检查下图片编码、代码格式是否有误。有疑 问请通过QQ群(群号:1009661589)或工单联系技术支持团队
336002	JSON不合法	入参格式或调用方式有误,比如缺少必要参数代码格式是否有误。有疑问请通过QQ群(群号:1009661589) 或工单联系技术支持团队
336003	Base64解码失败	图片格式有误或base64编码有误,请根据接口文档检查格式,base64编码请求时注意要去掉头部。有疑问请 通过QQ群(群号:1009661589)或工单联系技术支持团队
336004	输入文件大小不合法	图片或音频、文本格式有误,请根据接口文档检查入参格式,有疑问请通过QQ群(群号:1009661589)或工 单联系技术支持团队
336005	解码输入失败/分词错 误	图片编码错误(非jpg,bmp,png等常见图片格式),请检查并修改图片格式
336006	缺失必要参数	image字段缺失 (未上传图片)
336100	model temporarily unavailable	遇到该错误码请等待1分钟后再次请求,可恢复正常,若反复重试依然报错或有疑问请通过QQ群(群 号:1009661589)或工单联系技术支持团队

窜拍识别服务

服务介绍

心简介

窜拍识别服务可对用户上传的数据进行疑似窜拍图(相似图)分组,用户可按照「人」、「门店」、「时间」定义需要识别的范围,服务按定义的范围返回有相似图存在的图片组。服务支持对上传的图片进行批量识别,窜拍识别支持两类任务:

周期任务:定期对调用定制模型云服务API的图片数据进行窜拍识别

∞ 与门店关联

当窜拍识别任务中定义的图片门店ID与门店库中的业务门店ID一致时,可以支持在门店管理的门店库中查看该门店最近一次窜拍任务的识别情况。

API文档

の API调用方法

接口总览

1.图片管理接口

接口名称	API URL	API描述
创建图集 API	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/create_d ataset	创建图集,获取图集ID
上传图片 API	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/upload	指定图集ID,将需要识别的图片上传至该 图集
图集列表 API	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/dataset_ ist	查询已创建的图集和状态

图集状态 dataset_status

状态	描述
Idle	空闲状态,可被使用的图集
Processing	正在运行或是排队中的任务涉及的图集

2.任务管理接口

接口名称	API URL	API描述
创建任务API	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/create_task	创建窜拍识别任务,获取任务ID
查询结果API	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/query	指定任务ID,查询任务状态和结果
任务列表API	https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/task_list	查询已创建的任务和状态

任务状态 task_status

状态	描述
Queued	排队中的任务
Running	正在运行的任务
Success	运行成功的任务
Failure	运行失败的任务

接口调用流程

1.创建图集,获得图集ID

创建图集API:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/create_dataset

2.指定图集ID,把要识别窜拍的图片上传到该图集

上传图片API:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/upload

3.创建任务,指定图集ID,对该图集做窜拍识别,获取任务ID

创建任务API:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/create_task

4.指定任务ID,轮询任务结果,可以每隔一定时间调用API查询任务结果,比如10s查询一次,任务完成后可以查到结果

查询结果API:https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/query

另外,可通过图集列表API和任务列表API查询所有图集和任务。

接口调用方法

创建图集API

1.接口描述

创建图集,用于存放需要识别窜拍的图片

2.请求说明

HTTP 方法: POST

接口URL: https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/create_dataset

URL参数:

参数	值
----	---

access_token 通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

请求参数

无需上传参数

3.返回说明

返回参数

参数名称	是否必须	数据类型	含义
log_id	是	int	唯一的log id,用于问题定位
dataset_id	是	int	图集ID
dataset_status	是	string	图集状态

上传图片API

1.接口描述

为指定图集上传需要识别窜拍的图片。仅有Idle状态的图集,可以上传图片,对Running状态的图集的上传图片报错。

2.请求说明

HTTP 方法: POST

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数	值
Content-Type	application/json

Body中放置请求参数,参数详情如下:

参数名称	是否必 须	数据类 型	取值范 围	含义
dataset_id	是	int	-	指定上传图片的数据集id
image	是	string	-	图像数据,base64编码,要求base64编码后大小不超过4M,最短边至少15px,最长边最大4096px,支持 jpg/png/bmp格式 注意请去掉头部
image_na me	是	string	-	图片名称或编码,请保证该字段在创建任务时定义的查窜拍范围内的唯一性
image_pe ople	是	string	-	该图片的人员信息
image_sto re	是	string	-	该图片的网点信息,如ID
image_tim e	是	string	-	该图片的拍摄时间,格式为"yyyy-MM-dd hh:mm:ss"

提示:当一张图片的image_name、image_people、image_store、image_time全都一致时,判为重复图片,后上传的直接忽略。

3.返回说明

返回参数

参数名称	是否必须	数据类型	含义
log_id	是	int	唯一的log id ,用于问题定位
dataset_status	是	string	任务状态

图集列表API

1.接口描述

按条件查询创建的图集信息

2.请求说明

HTTP 方法:POST

 $\texttt{BCURL: https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/dataset_list}$

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名称	是否必需	参数类型	描述
dataset_ids	否	array[string]	只返回指定id的任务信息,最多返回前200个dataset_id的结果
begin_time	否	number	时间戳,只返回begin_time以后创建的任务信息
end_time	否	number	时间戳,只返回end_time之前创建的任务信息

3.返回说明

参数名称	是否必需	参数类型	描述
log_id	是	int	唯一的log id,用于问题定位
dataset_info	否	array[dict]	任务列表,最多返回前200个dataset_id的结果
+dataset_id	否	string	任务id
+dataset_status	否	string	任务状态
+create_time	否	int	时间戳,图集创建时间
+image_num	否	int	图集中含有的图片数量

创建任务API

1.接口描述

创建窜拍识别任务。

2.请求说明

HTTP 方法:POST

 $\texttt{BCURL: https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/create_task}$

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名 称	是否必 须	数据类 型	取值范围	含义
dataset _id	是	int	-	指定需要进行识别窜拍的图集ID
thresho Id	否	array[fl oat]	每个元素 取值范围 为 0.60~1.00	例如:[0.8,0.95]。可精确到小数点后两位。阈值为界定两张图视为窜拍的score的阈值,当score大于 threshold时,则判定为疑似窜拍结果返回。不填,默认为0.8
task_pe ople	否	boolea n	"True"、 "False"	区分不同人员,在同一人员拍摄的图片范围内,识别窜拍。True为区分,False为不区分;不填,默认为 False
task_st ore	否	string	"undefined "、"within"	"undefined"表示不区分网点;within表示区分不同网点,在同一网点的图片范围内,识别窜拍;不填,默 认为"undefined"
task_ti me	否	string	"undefined "` "year"` "month"` "day"	"undefined"表示不区分年份;"year"表示区分不同年份,在同一年份内的图片范围内,识别窜 拍;"month"表示区分不同月份(不同年同月,算作不同月),在同一月份内的图片范围内,识别窜 拍;"day" 表示区分不同日子(不同年或不同月的同一日,算作不同日),在同一日子内的图片范围内, 识别窜拍。不填,默认为"undefined"

3.返回说明

返回参数

参数名称	是否必须	数据类型	含义
log_id	是	int	唯一的log id,用于问题定位
task_id	否	string	新建任务对应的id
task_status	否	string	任务状态

查询结果API

1.接口描述

查询指定任务的结果。

2.请求说明

HTTP 方法:POST

 $\texttt{BCURL: https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/query}$

URL参数:

参数	值
access_token	通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名称	是否必须	数据类型	取值范围	含义
task_id	是	string	-	任务对应的id

3.返回说明

返回参数

参数名称	是否必须	数据类型	含义
log_id	是	int	唯一的log id,用于问题定位
task_status	否	string	任务状态
task_results_url	否	dict	dict的key为阈值,value为结果文件URL,文件格式为json。url每次查询随机生成,有效期7天

文件内的参数

参数名 称	是否必 须	数据类 型	含义
task_res ults	否	array[di ct]	任务结果
+re_peo ple	否	string	当创建任务「task_people」为true时,返回人员信息,与用户上传时传上的「img_people」一致;当 「task_people」为false时,返回-1
+re_stor e	否	string	当创建任务「task_store」为"within"时,返回人员信息,网点信息,与用户上传时传上的「image_store」一致;当 「task_store」为"undefined"时,返回-1
+re_year	否	int	当创建任务「task_time」不为"undefined"时,返回年份,与用户上传时传上的「image_time」中的"YY"一致;当 「task_time」均为"undefined"时,返回-1
+re_mon th	否	int	当创建任务「task_time」为"month"或"day"时,返回月份,与用户上传时传上的「image_time」中的"MM"一致;当 「task_time」为"undefined"或"year"时,返回-1
+re_day	否	int	当创建任务「task_time」为"day"时,返回日信息,与用户上传时传上的「image_time」中的"DD"一致;当 「task_time」不为"day"时,返回-1
+img_gro up	否	array[di ct]	包含的疑似相似图的信息
++group _id	否	int	相似图组的组ID
++img_n ame	否	array[str ing]	相似图组内的包含的所有图片名称,与用户上传时传上的「image_name」一致

任务列表API

1.接口描述

按条件查询范围内的任务状态。

2.请求说明

HTTP 方法:POST

接口URL: https://aip.baidubce.com/rpc/2.0/ai_custom_retail/v1/tasks/image_similar/task_list

URL参数:

参数 值

access_token 通过API Key和Secret Key获取的access_token,参考鉴权认证机制文档

Header如下:

参数 值 Content-Type application/json

Body中放置请求参数,参数详情如下:

请求参数

参数名称	是否必需	参数类型	描述
task_ids	否	array[string]	只返回指定id的任务信息,最多返回前200个task_id的结果
begin_time	否	number	时间戳,只返回begin_time以后创建的任务信息
end_time	否	number	时间戳,只返回end_time之前创建的任务信息

3.返回说明

返回参数

参数名称	是否必需	参数类型	描述
log_id	是	number	唯一的log id,用于问题定位
tasks_info	否	array[dict]	任务列表,最多返回前200个task_id的结果
+task_id	否	string	任务id
+task_status	否	string	任务状态
+create_time	否	number	时间戳,任务创建时间

心 错误码

错误码

若请求错误,服务器将返回的JSON文本包含以下参数:

- error_code:错误码。
- error_msg:错误描述信息,帮助理解和解决发生的错误。

例如Access Token失效时,接口返回:

```
{
    "error_code": 110,
    "error_msg": "Access token invalid or no longer valid"
}
```

需要重新获取新的Access Token再次请求即可。

错误码列表如下:

error_cod e错误码	error_msg错误信息	描述
1	Unknown error	服务器内部错误,请再次请求, 如果持续出现此类错误,请通过QQ群(1009661589)或工单联系技术支 持团队。
2	Service temporarily unavailable	服务暂不可用,请再次请求, 如果持续出现此类错误,请通过QQ群(1009661589)或工单联系技术支持 团队。
3	Unsupported openapi method	调用的API不存在,请检查后重新尝试
4	Open api request limit reached	集群超限额
6	No permission to access data	无权限访问该用户数据
13	Get service token failed	获取token失败
14	IAM Certification failed	IAM鉴权失败
15	app not exists or create failed	应用不存在或者创建失败
17	Open api daily request limit reached	每天请求量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费; 未上线计费的接口,请通过QQ群(1009661589)联系群管手动提额
18	Open api qps request limit reached	QPS超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未上线 计费的接口,请通过QQ群(649285136)联系群管手动提额
19	Open api total request limit reached	请求总量超限额,已上线计费的接口,请直接在控制台开通计费,调用量不受限制,按调用量阶梯计费;未 上线计费的接口,请通过QQ群(649285136)联系群管手动提额
100	Invalid parameter	无效的access_token参数,请检查后重新尝试
110	Access token invalid or no longer valid	access_token无效
111	Access token expired	access token过期
336200	internal error	内部错误,如果持续出现此类错误,请通过QQ群(群号:1009661589)或工单联系技术支持团队。
336201	unknown task id	未知的任务id
336202	invalid param: 'param_name'	请求参数'param_name'的参数值不合法
336203	missing param: 'param_name'	请求参数'param_name'缺失
336206	invalid base64	加图操作:错误的base64图片编码
336207	failed loading image	加图操作:加载图片失败
336208	invalid image format	加图操作:不支持的图片格式,支持格式: bmp、jpg、jpeg、png
336209	invalid image shape	加图操作:不支持的图片形状,图片长宽需在[15, 4096]之间
336210	invalid image size	加图操作:不支持的图片大小,图片大小不超过4M
336212	invalid json	请求数据格式不正确
336215	unknown dataset id	未知的dataset id
336216	dataset is busy	dataset正在进行窜拍任务,暂时无法上传图片
336217	dataset is empty	dataset为空,需上传图片后才可触发窜拍任务

零售版常见问题

心 训练相关问题

为什么建议每个SKU至少出现在20张实景图中?

上传的实景图,只有标注过的图片会被训练,所有训练的图片中,系统会随机抽取70%作为训练集,剩余的30%作为测试集,如果标注的训练数 据不足,可能会导致某个SKU的精确度远低于其它SKU,或是训练结果出现mAP、精确率、召回率全都为0的情况。

模型的训练结果是如何得到的?

上传的实景图,只有标注过的图片会被训练,所有训练图片中,系统会随机抽取70%作为训练集,剩余的30%作为测试集,训练集训练出的模型 去对测试集进行检测,检测得到的结果跟人为标注的结果进行比对,得到页面显示的mAP,精确率和召回率。

为什么同样的数据集每次训练出来的结果会不一样?

上传的实景图,只有标注过的图片会被训练,所有训练的图片中,系统会随机抽取70%作为训练集,剩余的30%作为测试集,由于训练数据和测 试数据每次都是随机抽取的,所以同样的数据集每次训练出来的结果会不一样。

SKU单品图是用来做什么的?

SKU单品图用来降低实景图即训练数据采集和标注成本的。为了让模型能够完整地识别一个SKU,需要训练的图片中出现这个SKU的各个角度的 样子,这意味着需要从实际业务场景中采集大量的图片,并且进行大量的标注工作。为了降低这部分的成本,我们通过数据合成和增强技术,只 需为SKU上传各个角度的单品图,且单品图无需进行任何标注,即可让模型学习到这个SKU各个角度的样子。由百度提供的SKU预置了50张左右 的单品图,绝大多数情况下无需再自行上传单品图。

SKU单品图需不需要标注?

SKU单品图不要标注,只需要参考「SKU单品图数据要求」文档采集并上传至相应的SKU即可。

模型训练失败怎么办?

如果遇到模型训练失败的情况,请直接加入官方QQ群(群号:1009661589)咨询解决。

心模型校验相关问题

如何校验模型的效果?

校验模型的目的是验证模型效果是否达到业务需求和找到模型效果最优的阈值,步骤如下两点:

- 初步校验:模型训练好后,可以使用「校验模型」功能,在页面上提交几张没有被用于训练且从实际业务场景中采集到的图片,调整阈值查 看结果,找到校验结果最优的阈值范围。如果无论什么阈值都无法满足您的业务需求,则需要继续优化模型,可以针对校验中发现的问题参 考模型效果优化相关问题对模型进行调优。
- 2. 批量校验:在初步校验后,得到校验结果最优的阈值范围,申请发布模型,发布成功后调用服务接口进行批量校验,找到校验结果最优的阈 值。调用接口的时候可以通过threshold这个参数设置阈值,threshold可以精确度小数点后2位。

∞ 模型上线相关问题

希望加急上线怎么处理?

加入官方QQ群(群号:1009661589)咨询群管高优审核。

每个账号可以上线几个模型?是否可以删除已上线的模型?

每个账号最多申请发布十个模型,已上线模型无法删除。

_のSKU相关问题

每个账号允许创建多少个SKU

每个账号默认允许创建的SKU数量为50个,如果需要增加SKU数量,请加入官方QQ群(群号:1009661589)咨询解决。

心收费相关问题

接口上线后是否收费?调用量不够怎么办?

目前接口是限量免费使用的原则,上线模型后可以免费获得1000次/天,qps=2的调用限额(QPS为每秒请求数)。如需调用更多次数,请在控制台中开通付费,开通付费后无调用次数限制,按调用成功次数收费,QPS免费提高至4,费用请参考服务价格文档。

心 模型效果优化相关问题

如何正确标注

- 单独框选要识别的SKU,不可同时框选多个目标
- 完整并仅仅框选要识别的SKU

- 标注框不要框选到其它SKU或是价目标签等非要识别的SKU的干扰信息
- 在实景图中出现的所有要识别的SKU必须全部标注

部分SKU识别效果太差

上传SKU单品图能够有效提高识别效果,上传要求如下:

- 图片像素足够高,不能模糊不清
- 单品图背景颜色必须为纯色且与SKU主体颜色不相似
- 角度、光线覆盖到实际检测场景中SKU所有可能出现的情况
- 要识别的SKU建议至少出现在20张实景图中,并且正确标注

错误示例中检查出漏标的情况

- 参考模型列表中模型效果下的「完整评估结果」内的推荐阈值,校验时适当调低阈值后测试效果,找到合适的阈值,在调用API接口时用请求 参数threshold设定合适的阈值
- 在实景图中出现的该SKU必须全部标注,不能存在漏标注的情况
- 要识别的SKU建议至少出现在20张实景图中,并且正确标注

某些SKU在特定的一些角度下识别度很低

- 添加出现这些角度的SKU的实景图到训练的实景图集中,并正确标注
- 为这些SKU添加这些角度的单品图

错误示例中出现一个大框框选住多个SKU的情况

 参考模型列表中模型效果下的「完整评估结果」内的推荐阈值,校验时适当调高阈值后测试效果,找到合适的阈值,在调用API接口时用请求 参数threshold设定合适的阈值

训练结果精确度很高,但是校验或是在实际场景调用时结果不好

- 要识别的SKU建议至少出现在20张实景图中,并且正确标注
- 参与训练的实景图集一定要包含和实际商品检测场景环境一致的图片,每个标签的图片需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强
- 参考模型列表中模型效果下的「完整评估结果」内的推荐阈值,校验时适当调整阈值后测试效果,找到合适的阈值,在调用API接口时用请求 参数threshold设定合适的阈值
- 用于训练、校验、实际检测的图片,像素都要足够高,不能出现模糊不清的情况

mAP、精确率、召回率全都为0

• 要识别的各个SKU推荐至少出现在20张以上实景图中,并且正确标注

误识别到相识度极高的非目标SKU

● 提交工单或是加入官方QQ群(群号:1009661589),将这些相似度极高的SKU基本信息(SKU名称、品牌、规格、包装)反馈给我们

心 其他问题

模型能否支持私有化部署?

目前定制商品检测服务提供在线调用API,如需私有化部署,可以提交工单咨询或是加入官方QQ群(群号:1009661589)联系管理员反馈

申请发布模型审核不通过都是什么原因?

可能原因如下:

1. 经过电话沟通当前模型存在一些问题或者不再使用,如训练数据异常、数据量不够、不想再继续使用等原因,沟通达成一致拒绝

2. 电话未接通且模型效果较差,会直接拒绝。如果需要申诉,加入官方QQ群(群号:1009661589) 咨询群管解决

零售版服务介绍

心简介

EasyDL是百度大脑中的一个定制化训练和服务平台,EasyDL零售版是EasyDL针对零售场景推出的行业版。EasyDL零售版提供两种服务,分别为 定制商品检测服务和货架拼接服务。

定制模型服务是EasyDL零售版的一项服务,专门用于训练货架合规性检查、自助结算台、无人零售货柜等场景下的定制化AI模型,训练出的模型 将以API的形式为客户提供服务。该服务包含以下2种定制模型:

- 1. 商品检测模型
 - 适用场景:适用于适用于货架、端架、挂架等场景的商品陈列规范核查,支持识别商品基本信息,陈列顺序、层数、场景,统计排面数 量和占比
 - 服务功能:
 - 商品基本信息识别:商品的名称、品牌、规格、编号;商品在图片中的坐标位置;商品识别的置信度
 - 商品陈列层数识别:商品陈列所在货架层数和货架总层数;商品陈列顺序;货架是否拍摄完整判断
 - 商品陈列场景识别:场景类型支持:普通货架、挂钩货架、斜口篮货架、端架、立式冰柜、卧式冰柜、冷风柜、地堆、割箱、地 龙、小端架、吧台
 - 商品排面占比统计:商品的排面数及排面占比率;每层货架可识别商品数量及未知商品数量;货架的总空位数、每层货架空位数
 及货架利用率
 - 商品陈列翻拍识别:识别商品陈列照片是对手机屏幕翻拍的可能性

2. 地堆检测模型

+适用场景:适用于堆箱、堆头、地龙等场景的商品陈列规范核查,支持识别商品基本信息,可视商品计数,纵深商品计数和占地面积+服务功能:

- + 商品基本信息识别:商品的名称、品牌、规格、编号;商品在图片中的坐标位置;商品识别的置信度;陈列顺序;可视商品计数, 纵深商品计数和占地面积
- + 商品陈列场景识别:场景类型支持:普通货架、挂钩货架、斜口篮货架、端架、立式冰柜、卧式冰柜、冷风柜、地堆、割箱、地
- 龙、小端架、吧台
 - + 商品陈列翻拍识别:识别商品陈列照片是对手机屏幕翻拍的可能性

货架拼接服务基于百度EasyDL深度学习算法,支持将多个货架局部图片或视频,组合为完整货架图片。同时支持输出在完整货架图中的商品 检测结果,包含SKU的名称和数量,适用于需要在长货架进行商品检测的业务场景。

翻拍识别服务能够识别出通过手机翻拍出的商品陈列照片,比如商品货架陈列图片和地堆商品陈列图片,可降低人工审核人力,高效审核零售业 务中通过翻拍原有图片来造假的图片。

价签识别服务能够识别货架和促销活动中的价签信息,可识别各个价签在图片中的像素位置,以及价签内商品名称和价格,可用于洞察商品在线 下渠道分销的价格区间。

心 功能介绍

☑ 定制商品检测服务

• AI模型训练平台

专门用于定制货架合规性检查、自助结算台、无人零售货柜等零售场景下识别商品的高精度AI模型。

• 全可视化操作

所有模型训练相关的操作都可以在网页上进行,无需编程,仅需五步即可部署定制化AI模型。

• 预置的商品库

预置近千种商品单品图可供客户在创建SKU时选择,用于合成训练数据,极大降低了训练数据采集和标注成本。

• 可自定义商品

客户可根据业务需求创建属于自己的商品,商品信息支持完全自定义,充分满足客户定制化需求。

• 全面的商品信息

商品基本信息识别

接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度

商品陈列层数识别

接口支持识别商品陈列所在货架层数,货架类型支持:货架、端架和立式冰柜内货架

商品陈列场景识别

接口支持识别商品陈列的场景,场景类型支持:普通货架、挂钩货架、斜口篮货架、端架、立式冰柜、地堆、割箱、地龙、小端架、吧台

商品排面占比统计

接口支持统计商品排面数/占比、未识别商品数、总空位数、每货架层的空位数及货架利用率

商品陈列翻拍识别

识别商品陈列照片是对手机屏幕翻拍的可能性

心 货架拼接服务

• 拼接和商品检测相结合

支持将多个货架局部图片或视频,组合为完整货架图片,并支持输出在完整货架图中的商品检测结果,包含SKU的名称和数量。

• 丰富的服务方式

支持三种服务方式:云服务API、完全开源的SDK以及可以直接体验的手机APP。

心 适用场景

• 货架合规性检查

精准识别出货架、冰柜和端架上陈列商品的数量和种类,为品牌商分析陈列排面占比,重点SKU分销率、缺货率、合格率提供数据支撑。

互动营销

训练定制的商品识别,实现对C端用户提交的商品图片进行识别,配合游戏规则完成闯关/抽奖式的互动营销。

心 技术优势

• 免费训练与测试

平台提供大量免费的GPU训练资源,及每天500次免费调用量,用于模型迭代和效果验证,有效降低项目开发和测试成本。

• 高可用模型效果

针对零售场景专项算法调优,结合图像合成与增强技术提升模型泛化能力,模型准确率可达97%+,保证模型在生产环境中具有高可用性。

• 预置模型和数据

平台提供直接可用的商品检测API,覆盖常见商品品类;提供大量预置单品图数据,可用于训练定制模型,有效提升项目落地效率。

• 企业级安全保障

数据加密与隔离,完善的服务调用鉴权,为客户的数据和模型提供企业级安全保障。

• 功能完善且丰富

全面覆盖各类零售场景的商品识别需求,应对不同场景的业务需求提供多种可选服务类型。

心 与EasyDL物体检测的相同点和不同点

EasyDL零售版是EasyDL专门针对零售场景下识别商品推出的版本,相比于物体检测模型,零售版更贴合快消零售场景下的业务需求,专门用于 训练货架合规性检查、自助结算台、无人零售货柜等场景下的定制化商品检测AI模型,训练出的模型可发布成云服务API,服务支持四种功能:商 品基本信息识别、商品陈列层数识别、商品陈列场景识别和商品排面占比统计,适用于识别货架中的商品信息,商品计数和陈列顺序等,辅助货 架商品陈列合规检查,如铺货率、陈列情况等。 相同点

同为检测模型,接口支持返回目标物体的名称和物体在图片上的位置。

不同点

- 模型算法不同:零售版的模型算法专门根据零售行业的场景和业务需求做了专项优化,基于百度大脑大规模零售数据预训练,并利用商品增强 合成技术将SKU单品图合成实景货架图,有针对性的提高了训练商品检测模型的精确度。
- 训练数据不同:零售版的数据除了需要标注的实景业务图片外,支持为每个SKU标签上传单品图。SKU单品图用来降低实景图即训练数据采集和标注成本的。为了让模型能够完整地识别一个SKU,需要训练的图片中出现这个SKU的各个角度的样子,这意味着需要从实际业务场景中采集大量的图片,并且进行大量的标注工作。为了降低这部分的成本,我们通过数据合成和增强技术,只需为SKU上传各个角度的单品图,且单品图无需进行任何标注,即可让模型学习到这个SKU各个角度的样子。合成图片过程在训练阶段自动完成,无需操作操作和进行标注。EasyDL零售版平台预置了近千种商品,每个商品预置了50张左右的单品图,绝大多数情况下无需再自行上传单品图。
- 云服务API功能不同:零售版云服务API支持四种功能:商品基本信息识别、商品陈列层数识别、商品陈列场景识别和商品排面占比统计;物体 检测云服务API仅支持商品基本信息识别。

购买指南

心 开通付费及购买服务

EasyDL零售版的各项服务的「开通付费」、「购买QPS叠加包」、「购买调用量次数包」、「关闭付费」等操作均在EasyDL零售版控制台进行, 只需在相应需要付费使用的接口位置,跟随页面提示完成后续充值及付费,即可完成。

✓返回EasyDL总宽		零售版概览											
零售版													
应用列表		产品介绍: (展开查	看服务功能介绍)										展开
用量统计		可用接口列表											
监控报表		■ 已上线的定制接口											
定制商品检测服务 >>		API	模型ID	模型类型	模型名称		模型版本	状态		调用量限制	QPS限制	开通按量后付费 ?	购买QPS叠加包 ?
标准商品检测服务 >		111	30443	商品检测	1111		V1	●免费使	用	剩余免费1000次	不保证并发	开通	购买 配额详情
货架拼接 >	~	silenceteapi	27185	商品检测	silencesecon	d	V3	●付费使	用	500次/天免费 + 超 出按量计费	4	终止付费	购买丨配额详情
		silencefirst0829	27135	商品检测	silencefirst		V3	●免费使	用	500次/天免费	不保证并发	开通	购买 配额详情
		【 饮品检测											
		API		状态			调用量限制			QPS限制		开通按量后付费 ?	
		饮品检测		● 免费使用			500次/天免费			不保证并发		免费试用	
		┃ 日化品检测											
		API		状态			调用量限制			QPS限制		开通按量后付费 ?	
		日化品检测		● 免费使用			剩余免费1000次			不保证并发		免费试用	
		商品陈列翻拍识别											
		API		状态		调用量的	限制		QPS限制		开通按量后付费	?	购买QPS叠加包 ?
		商品陈列翻拍识别		● 付费使用		剩余免费	费999次 + 超出按量计费		4		终止付费		购买丨配额详情

⊙ 定制商品检测服务

₯ 价目表 - 按调用量后付费

定制商品检测模型云服务API支持以下三项功能,启停服务功能的方法请见 服务功能文档。

1. 商品基本信息识别 (必选)

接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度

2. 商品陈列层数识别 (可选)

接口支持识别商品陈列所在货架层数,货架类型支持:货架、端架和立式冰柜内货架

3. 商品陈列场景识别(可选)

货架场景:货架、端架(小方货架);冰柜场景:立式冰柜;地堆场景:堆箱、割箱、地龙

- 4. 商品排面占比统计(可选)
 接口支持统计商品排面数/占比、未识别商品数、空位数及货架利用率
- 商品陈列翻拍识别(可选)
 识别商品陈列照片是对手机屏幕翻拍的可能性

每个模型发布的云服务API享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费,开通后即可开始计费使用。

注:为保证老用户可正常使用三项服务,对于服务功能上线前已经开通付费的定制模型API,默认对各项服务功能开通付费。商品陈列层数识 别和商品陈列场景识别两项服务功能,需在EasyDL零售版模型训练后台启动后方可使用,启动方法请见服务功能文档,启动后发生调用才会 按实际调用次数进行收费。

选择开启的功能不同,单次接口调用价格不同,详情如下:

 商品基本信息识别(必选),按月调用量阶梯计费,调用单价按照自然月累积调用量所落阶梯区间而变化。月初,上月累积的调用量清零,重 新开始累积本月调用量。

月调用量 (万次)	单次调用价格 (元)	QPS限制	说明
0<月调用量<=15	0.009	4	服务器支持每秒处理4次查询
15<月调用量<=150	0.008	4	服务器支持每秒处理4次查询
150<月调用量	0.007	4	服务器支持每秒处理4次查询

• 商品陈列层数识别 (可选),单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.04	4	服务器支持每秒处理4次查询

• 商品陈列场景识别 (可选) , 单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.006	4	服务器支持每秒处理4次查询

• 商品排面占比统计 (可选),单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.02	4	服务器支持每秒处理4次查询

• 商品陈列翻拍识别 (可选),单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.05	4	服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

EasyDL零售版各项服务都具有免费调用额度,开通付费后,免费调用额度仍保留,EasyDL零售版的定制商品检测服务免费额度如下:

服务名称	免费调用额度	QPS限制
定制商品检测服务	累计1000次	1~2,服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

费用举例

从2019-3-1至2019-3-31,**定制商品检测API**在三种服务功能都开启的情况下,月调用量为600万次(已除去免费额度),那么费用如下:

1. 商品基本信息识别的费用为43,650元,明细如下:

前15万次落入0~15w阶梯,单次调用0.009元/次,费用为1,350元;

中间15万~150万次落入15~150w阶梯,单次调用0.008元/次,费用为10,800元;

最后150万~600万次落入大于150w阶梯,单次调用0.007元/次,费用为31,500元;

共计43,650元

2. 商品陈列层数识别的费用为360,000元,明细如下:

月调用量为600万次,单次调用0.04元/次,费用为240,000元

3. 商品陈列场景识别的费用为,明细如下:

月调用量为600万次,单次调用0.006元/次,费用为36,000元

综上,三月费用合计319,650元。

心价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格

按天购买 50元/天

按月购买 1050元/月

购买QPS叠加包前需保证已开通按量后付费或购买次数包

购买的QPS叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

购买QPS叠加包提升的额度对定制商品检测API的所有服务功能均有效

⊙ 定制地堆检测服务

⊙价目表 - 按调用量后付费

定制地堆检测模型云服务API支持以下三项功能,启停服务功能的方法请见 服务功能文档。

- 商品基本信息识别(必选)
 接口支持识别商品信息(商品名称、品牌、规格)、编号和置信度
- 商品陈列场景识别(可选)
 货架场景:货架、端架(小方货架);冰柜场景:立式冰柜;地堆场景:堆箱、割箱、地龙
- 3. 商品陈列翻拍识别 (可选)

识别商品陈列照片是对手机屏幕翻拍的可能性

付费调用

每个模型发布的云服务API享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费,开通后即可开始计费使用。

注:为保证老用户可正常使用商品陈列场景识别服务功能,对于服务功能上线前已经开通付费的定制模型API,默认对各项服务功能开通付 费。商品陈列场景识别服务功能,需在EasyDL零售版模型训练后台启动后方可使用,启动方法请见服务功能文档,启动后发生调用才会按实 际调用次数进行收费。

选择开启的功能不同,单次接口调用价格不同,详情如下:

 商品基本信息识别(必选),按月调用量阶梯计费,调用单价按照自然月累积调用量所落阶梯区间而变化。月初,上月累积的调用量清零,重 新开始累积本月调用量。

 单次调用价格(元)
 QPS限制
 说明

 0.016
 4
 服务器支持每秒处理4次查询

Baidu 百度智能云文档

• 商品陈列场景识别 (可选) ,单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.006	4	服务器支持每秒处理4次查询
• 商品陈列翻拍识别	(可选),	单次调用额外收取费用

单次调用价格 (元)	QPS限制	说明
0.05	4	服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

EasyDL零售版各项服务都具有免费调用额度,开通付费后,免费调用额度仍保留,EasyDL零售版的定制地堆检测服务免费额度如下:

服务名称	免费调用额度	QPS限制
定制地堆检测服务	累计1000次	1~2,服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

费用举例

从2019-3-1至2019-3-31,**定制地堆检测API**在三种服务功能都开启的情况下,月调用量为600万次(已除去免费额度),那么费用如下:

1. 商品基本信息识别的费用为96,000元,明细如下:

月调用量为600万次,单次调用0.016元/次,费用为96,000元

2. 商品陈列场景识别的费用为,明细如下:

月调用量为600万次,单次调用0.006元/次,费用为36,000元

综上,三月费用合计132,000元。

心价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格 按天购买 60元/天

按月购买 1200元/月

购买QPS叠加包前需保证已开通按量后付费或购买次数包

购买的QPS叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

购买QPS叠加包提升的额度对定制地堆检测API的所有服务功能均有效

心 翻拍识别服务

心价目表 - 按调用量后付费

付费调用

每个账户享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用, 价格如下:

Baidu 百度智能云文档

EasyDL 零售行业版使用说明

 调用量
 价格
 QPS限制
 说明

 每次调用
 0.05元
 4
 服务器支持每秒处理4次查询

注:调用失败不计费

免费额度

每个账号享有一定量免费调用额度,如下表:

云服务API	免费调用额度	QPS限制	说明
商品陈列翻拍识别	累计1000次	1~2	服务器支持每秒处理1~2次查询

注:成功调用与失败调用均消耗免费额度

心 价目表 - 调用量次数包

如果对调用次数有预估,可以选择购买单次调用价格更低的次数包,价格如下:

规格	价格	QPS限制	有效期
1万次	490 元	4	1年
10万次	4,800 元	4	1年
100万次	45,000 元	4	1年
500万次	212,500 元	4	1年
1000万次	420,000 元	4	1年
2000万次	800,000 元	4	1年

购买后不可退款,次数包使用完后,开始按调用量每次0.05元收取费用

心 价目表 - QPS叠加包

开通付费后,免费QPS为4,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式 每QPS价格 按天购买 50元/天 按月购买 1200元/月

购买 QPS 叠加包需保证已开通按量后付费或购买次数包

购买的 QPS 叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

心 价签识别服务

心价目表 - 按调用量后付费

付费调用

每个账户享有累计1000次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用, 价格如下:

调用量	价格	QPS限制	说明
每次调用	0.05元	2	服务器支持每秒处理2次查询

注:调用失败不计费

免费额度

每个账号享有一定量免费调用额度,如下表:

云服务API	免费调用额度	QPS限制	说明
价签识别	累计1000次	1	服务器支持每秒处理1次查询

注:成功调用与失败调用均消耗免费额度

心 价目表 - 调用量次数包

如果业务上对调用次数有预估,可以选择购买单次调用价格更低的次数包,价格如下:

规格	价格	QPS限制	有效期
1万次	475 元	2	1年 (366天)
5万次	2,250 元	2	1年 (366天)
10万次	4,250 元	2	1年 (366天)
20万次	8,000 元	2	1年 (366天)
50万次	18,750 元	2	1年 (366天)
100万次	35,000 元	2	1年 (366天)
500万次	150,000 元	2	1年 (366天)

购买后不可退款,次数包使用完后,开始按调用量每次0.05元收取费用

心价目表 - QPS叠加包

开通付费后,免费QPS为2,如果有更多的并发请求需要,可以根据业务需求按天或按月购买QPS叠加包,价格如下:

购买方式	每QPS价格
按天购买	100元/天
按月购买	2000元/月

购买 QPS 叠加包需保证已开通按量后付费或购买次数包

购买的 QPS 叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

心 货架拼接服务

货架拼接服务支持按任务数后付费、任务次数包预付费和并发任务叠加包预付费三种计费方式。

⊙价目表 - 按任务数后付费

付费调用

每个账户享有累计200次免费调用额度,免费额度用尽后,请在百度云EasyDL零售版控制台开通计费后继续使用,开通后即可开始计费使用,价 格如下:

任务数	价格 (元)	并发任务数限制	说明
每次拼接任务	0.2	1	只允许一个拼接任务在运行,超出限制的任务排队等待

免费额度

每个账号享有一定量免费调用额度,如下表:

	服务	免费任务额度	并发任务数限制	说明
--	----	--------	---------	----

货架拼接 累计200次 1 只允许一个拼接任务在运行,超出限制的任务排队等待

注:任务成功与失败调用均消耗免费额度

心价目表 - 任务次数包

如果对拼接任务次数有预估,可以选择购买单次任务价格更低的次数包,价格如下:

规格	价格	并发任务数限制	有效期
1千次	200 元	1	1年
1万次	1,900 元	1	1年
10万次	18,000 元	1	1年
100万次	150,000 元	1	1年
500万次	600,000 元	1	1年

购买后不可退款,任务次数包使用完后,开始按调用量每个任务0.2元收取费用

心价目表 - 并发任务叠加包

开通付费后,并发任务数限制为1,如果有更多的并发请求需要,可以根据业务需求按天或按月购买并发任务叠加包,价格如下:

购买方式	每并发任务价格
按天购买	2元/天
按月购买	40元/月

购买 并发任务叠加包需保证已开通按量后付费或购买任务次数包

购买的并发任务叠加包自起始日0点生效,至停止日24点失效。若起始日为本日,则是从下单成功时当即生效

心 余额不足提醒与欠费处理

心 余额不足提醒

根据您历史的账单金额,判断您的账户余额(含可用代金券)是否足够支付未来的费用,若不足以支付,系统将在欠费前三天、两天、一天发送 续费提醒短信,请您收到短信后及时前往控制台财务中心充值。

の 欠费处理

- 北京时间整点检查您的账户余额是否足以支付本次账单的费用(如北京时间11点整检查账户余额是否足以支付10点至11点的账单费用),若
 不足以支付,即为欠费,欠费时系统会发送欠费通知。
- 欠费后您开通付费的产品将进入欠费状态,只能使用每日的免费额度,超过额度的请求系统将不再响应,且不再保证并发处理。

数据看板

新增门店数据

零售版常见问题

为什么建议每个SKU至少出现在20张实景图中?

上传的实景图,只有标注过的图片会被训练,所有训练的图片中,系统会随机抽取70%作为训练集,剩余的30%作为测试集,如果标注的训练数 据不足,可能会导致某个SKU的精确度远低于其它SKU,或是训练结果出现mAP、精确率、召回率全都为0的情况。

模型的训练结果是如何得到的?

上传的实景图,只有标注过的图片会被训练,所有训练图片中,系统会随机抽取70%作为训练集,剩余的30%作为测试集,训练集训练出的模型 去对测试集进行检测,检测得到的结果跟人为标注的结果进行比对,得到页面显示的mAP,精确率和召回率。

为什么同样的数据集每次训练出来的结果会不一样?

上传的实景图,只有标注过的图片会被训练,所有训练的图片中,系统会随机抽取70%作为训练集,剩余的30%作为测试集,由于训练数据和测 试数据每次都是随机抽取的,所以同样的数据集每次训练出来的结果会不一样。

SKU单品图是用来做什么的?

SKU单品图用来降低实景图即训练数据采集和标注成本的。为了让模型能够完整地识别一个SKU,需要训练的图片中出现这个SKU的各个角度的 样子,这意味着需要从实际业务场景中采集大量的图片,并且进行大量的标注工作。为了降低这部分的成本,我们通过数据合成和增强技术,只 需为SKU上传各个角度的单品图,且单品图无需进行任何标注,即可让模型学习到这个SKU各个角度的样子。由百度提供的SKU预置了50张左右 的单品图,绝大多数情况下无需再自行上传单品图。

SKU单品图需不需要标注?

SKU单品图不要标注,只需要参考「SKU单品图数据要求」文档采集并上传至相应的SKU即可。

模型训练失败怎么办?

如果遇到模型训练失败的情况,请直接加入官方QQ群(群号:1009661589)咨询解决。

心 模型校验相关问题

如何校验模型的效果?

校验模型的目的是验证模型效果是否达到业务需求和找到模型效果最优的阈值,步骤如下两点:

- 初步校验:模型训练好后,可以使用「校验模型」功能,在页面上提交几张没有被用于训练且从实际业务场景中采集到的图片,调整阈值查 看结果,找到校验结果最优的阈值范围。如果无论什么阈值都无法满足您的业务需求,则需要继续优化模型,可以针对校验中发现的问题参 考模型效果优化相关问题对模型进行调优。
- 批量校验:在初步校验后,得到校验结果最优的阈值范围,申请发布模型,发布成功后调用服务接口进行批量校验,找到校验结果最优的阈值。调用接口的时候可以通过threshold这个参数设置阈值,threshold可以精确度小数点后2位。

⊙ 模型上线相关问题

希望加急上线怎么处理?

加入官方QQ群(群号:1009661589)咨询群管高优审核。

每个账号可以上线几个模型?是否可以删除已上线的模型?

每个账号最多申请发布十个模型,已上线模型无法删除。

の SKU相关问题

每个账号允许创建多少个SKU

每个账号默认允许创建的SKU数量为50个,如果需要增加SKU数量,请加入官方QQ群(群号:1009661589)咨询解决。

心 收费相关问题

接口上线后是否收费?调用量不够怎么办?

目前接口是限量免费使用的原则,上线模型后可以免费获得1000次/天,qps=2的调用限额(QPS为每秒请求数)。如需调用更多次数,请在控制台中开通付费,开通付费后无调用次数限制,按调用成功次数收费,QPS免费提高至4,费用请参考服务价格文档。

心 模型效果优化相关问题

如何正确标注

- 单独框选要识别的SKU,不可同时框选多个目标
- 完整并仅仅框选要识别的SKU
- 标注框不要框选到其它SKU或是价目标签等非要识别的SKU的干扰信息
- 在实景图中出现的所有要识别的SKU必须全部标注

部分SKU识别效果太差

- 上传SKU单品图能够有效提高识别效果,上传要求如下:
- 图片像素足够高,不能模糊不清
- 单品图背景颜色必须为纯色且与SKU主体颜色不相似
- 角度、光线覆盖到实际检测场景中SKU所有可能出现的情况
- 要识别的SKU建议至少出现在20张实景图中,并且正确标注

错误示例中检查出漏标的情况

- 参考模型列表中模型效果下的「完整评估结果」内的推荐阈值,校验时适当调低阈值后测试效果,找到合适的阈值,在调用API接口时用请求 参数threshold设定合适的阈值
- 在实景图中出现的该SKU必须全部标注,不能存在漏标注的情况
- 要识别的SKU建议至少出现在20张实景图中,并且正确标注

某些SKU在特定的一些角度下识别度很低

- 添加出现这些角度的SKU的实景图到训练的实景图集中,并正确标注
- 为这些SKU添加这些角度的单品图

错误示例中出现一个大框框选住多个SKU的情况

 参考模型列表中模型效果下的「完整评估结果」内的推荐阈值,校验时适当调高阈值后测试效果,找到合适的阈值,在调用API接口时用请求 参数threshold设定合适的阈值

训练结果精确度很高,但是校验或是在实际场景调用时结果不好

- 要识别的SKU建议至少出现在20张实景图中,并且正确标注
- 参与训练的实景图集一定要包含和实际商品检测场景环境一致的图片,每个标签的图片需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强
- 参考模型列表中模型效果下的「完整评估结果」内的推荐阈值,校验时适当调整阈值后测试效果,找到合适的阈值,在调用API接口时用请求 参数threshold设定合适的阈值
- 用于训练、校验、实际检测的图片,像素都要足够高,不能出现模糊不清的情况

mAP、精确率、召回率全都为0

• 要识别的各个SKU推荐至少出现在20张以上实景图中,并且正确标注

误识别到相识度极高的非目标SKU

● 提交工单或是加入官方QQ群(群号:1009661589),将这些相似度极高的SKU基本信息(SKU名称、品牌、规格、包装)反馈给我们

心 其他问题

模型能否支持私有化部署?

目前定制商品检测服务提供在线调用API,如需私有化部署,可以提交工单咨询或是加入官方QQ群(群号:1009661589)联系管理员反馈

申请发布模型审核不通过都是什么原因?

可能原因如下:

1. 经过电话沟通当前模型存在一些问题或者不再使用,如训练数据异常、数据量不够、不想再继续使用等原因,沟通达成一致拒绝

2. 电话未接通且模型效果较差,会直接拒绝。如果需要申诉,加入官方QQ群(群号:1009661589)咨询群管解决

EasyDL桌面版使用说明(已下线)

产品简介

产品介绍

飞桨EasyDL桌面版是百度针对客户端开发的零门槛AI开发平台,可在离线状态通过本地资源完成包括数据管理与数据标注、模型训练、模型部署的一站式AI开发流程。

无需机器学习专业知识,通过 模型创建→数据上传→模型训练→模型发布 全流程可视化便捷操作,最快15分钟即可获得一个高精度模型。

针对有一定AI模型开发基础的开发者,飞桨EasyDL同时还提供了预置模型调参、Notebook建模两种建模方式,开发者可根据自身经验进行调整, 以获得更适合特定场景的模型。

目前已支持训练图像分类、物体检查、实例分割、语义分割4种不同应用场景的模型。

⊙应用场景

图像分类:

- 图片内容检索:定制训练需要识别的各种物体,并结合业务信息展现更丰富识别结果
- 图片审核:定制图像审核规则,如训练直播场景中抽烟等违规现象
- 制造业分拣或质检:定制生产线上各种产品识别,进而实现自动分拣或者质检
- 医疗诊断:定制识别医疗图像,辅助医生肉眼诊断

物体检测:

- 视频监控:如检测是否有违规物体、行为出现
- 工业质检:如检测图片里微小瑕疵的数量和位置
- 医疗诊断:如医疗细胞计数、中草药识别等

实例分割:

- 专业检测:应用于专业场景的图像分析,比如在卫星图像中识别建筑、道路、森林,或在医学图像中定位病灶、测量面积等
- 智能交通:识别道路信息,包括车道标记、交通标志等

语义分割:

- 自动驾驶:识别道路障碍、分割线等,辅助驾驶决策
- 智能分拣:包括工业零部件分拣、垃圾分拣等

功能介绍

飞桨EasyDL桌面版提供数据处理、模型训练、模型部署全流程的模型生产能力。

数据处理:

提供针对图像的成熟标注模板及工具,便捷的为AI开发准备高质量训练数据,提供可视化管理能力,支持不同数据格式的导入、导出、查看

AotuDL建模:

为零AI开发基础的用户提供的建模方式,内置基于百度文心大模型的成熟预训练模型,可针对用户数据进行算法自动优化,助用户使用少量数据 也能获得具备出色效果与性能的模型

预置模型调参建模:

为有一定AI开发基础的开发者提供预置模型调参建模方式,涵盖ResNet、YOLO、PicoDet、MaskRCNN等近30种网络类型,适配大部分场景,开发者只需选择合适的预训练模型以及网络,根据自身经验进行调整,以获得更适合特定场景的模型

Notebook建模:

集成了包括PaddleX、PaddleDetection、PaddleSeg、PaddleClas等端到端开发套件的轻量级IDE,用户可在该模块内进行代码编辑、调试等开发 工作,快速高效的完成各类任务的实现,可对预置模型调参中的模型进行代码级优化

模型部署:

训练完成的模型可发布为在服务器、小型设备、专项适配硬件上直接部署的SDK,覆盖主流芯片与操作系统,充分满足不同业务场景对模型部署 的要求

心不同版本功能对比

~~~~ ~~~~

产品分为标准版与高级版,标准版提供图像分类、物体检测、实例分割、语义分割场景下完整的模型生产能力,高级版在标准版的基础上在各环 节进一步提供了便捷化的应用工具。

每位新用户下载并成功激活飞桨EasyDL后,将专享30天高级版免费试用权益

| 模块  | 功能点        | 功能明细                                                   | 标准版 | 高级版 |
|-----|------------|--------------------------------------------------------|-----|-----|
|     | 数据集创建      | 创建图像分类、物体检测、实例分割、语义分割数据集                               | 1   | 1   |
|     | 数据导入       | 导入图像分类、物体检测、实例分割、语义分割已标注/未标注数据集                        | 1   | 1   |
| 数据  |            | 图像分类提供常规标注能力、批量标注能力                                    | 1   | 1   |
|     | 数据标注       | 物体检测提供常规标注能力                                           | 1   | 1   |
|     |            | 实例分割、语义分割提供常规标注能力,自动识别轮廓标注能力                           | 1   | 1   |
|     |            | 发起图像分类、物体检测、实例分割、语义分割任务训练                              | 1   | 1   |
|     | AutoDI 柑弌  | 选择训练算法                                                 | 1   | 1   |
|     | AutoDL模式   | 根据部署环境提供针对性训练算法                                        | 1   | 1   |
|     |            | 手动/自动设置数据增强策略                                          |     | 1   |
|     | 预置模型调参模式   | 发起图像分类、物体检测、实例分割、语义分割任务训练                              |     | 1   |
|     |            | 选择网络、预训练模型                                             |     | 1   |
|     |            | 手动/自动设置模型训练参数                                          |     | 1   |
| 开发  |            | 手动/自动设置数据增强策略                                          |     | 1   |
|     | Notebook模式 | 预置PaddleX、PaddleDetection、PaddleSeg、PaddleClas等端到端开发套件 | 1   | 1   |
|     |            | 零代码开发模型转Notebook优化                                     | 1   | 1   |
|     |            | 支持代码高亮,自动补全                                            | 1   | 1   |
|     |            | 可管理代码、数据、模型等类型的文件                                      | 1   | 1   |
|     |            | 实时可视化CPU、内存、显卡、硬盘的运行数据                                 | 1   | 1   |
|     | 评估报告       | 训练完成后自动生成模型评估报告                                        | 1   | 1   |
|     | 任务校验       | 训练完成后支持发起模型校验                                          | 1   | 1   |
| 档刑  | 柑刑如翌       | 导出模型源文件                                                | 1   | 1   |
| 1天空 | 1天王山石      | 导出可直接用于服务器、小型设备、专项适配硬件上直接部署的SDK,覆盖主流芯片与操作系统            |     | 1   |

系统支持

## 心客户端安装系统要求

- Windows: Windows 10及以上版本 (64 bit)
- Mac: macOS 10.11及以上版本 (IntelCPU)
- Linux: Ubuntu 18.04及以上版本 (64 bit)

の CPU训练环境要求

请确保CPU芯片支持AVX指令集

如果您的计算机有NVIDIA® GPU,且需要使用GPU环境进行训练,请确保满足以下条件:

- Windows 10/11: 驱动版本需527.41及以上,需安装 CUDA 12.0 与 cuDNN v8.9.0
- Ubuntu 18.04/20.04: 驱动版本需525.60.13及以上,需安装 CUDA 12.0 与 cuDNN v8.9.0

の CUDA、cuDNN安装指南

您可参考NVIDIA官方文档了解CUDA和CUDNN的安装流程和配置方法,详见:

- CUDA 安装指南: https://docs.nvidia.com/cuda/
- cuDNN安装指南: https://docs.nvidia.com/deeplearning/cudnn/install-guide/index.html

下载与激活

心下载

访问飞桨EasyDL桌面版官网:https://ai.baidu.com/easydl/paddle ,根据您的系统,下载对应的客户端。

心激活

每位新用户下载并成功激活飞桨EasyDL后,将专享30天高级版免费试用权益。高级版体验到期后,若仍需要使用高级版特有功能,可点击进入「购买高级版」。

完成高级版购买后,在客户端内输入获得的高级版序列号并联网激活,即可继续使用高级版功能。

购买入口1

...



# 飞桨EasyDL

零门槛AI开发平台

|          |                |                                  |                                        |                            |                         |            | 购买入□  |
|----------|----------------|----------------------------------|----------------------------------------|----------------------------|-------------------------|------------|-------|
| 使用基础版    |                |                                  |                                        |                            |                         |            |       |
|          |                |                                  | 激活高级版                                  |                            |                         |            |       |
|          | 1.首<br>太       | 訂次激活基础間                          | 反/高级版时请确保您的                            | 设备处于耶                      | 送网状                     |            |       |
|          | 心<br>2.基<br>式A | 基础版提供数排<br>AI开发能力,请              | 据标注、模型训练与源<br>高级版在此基础上提供)              | 文件导出等<br>适配多种剖             | 译一站<br>邓署场              |            |       |
|          | 景的提示           | 的SDK 了解基示:您的高级版                  | 基础版/高级版<br>试用权益已到期, 您可                 | 可以继续使                      | 用基                      |            |       |
|          | 石出 片           | 版或激活高级                           | 版,高级版序列号请                              | 列买高级版                      | 获取                      |            |       |
| • 订阅     |                | く 〉 产品总览                         |                                        |                            |                         |            | 0 0   |
| ,"" 订阅购3 | 买列表            | 欢迎订阅飞桨EasyDL高级版                  | 发现订阅飞桨EasyDL高级版                        |                            |                         |            |       |
|          |                | 免费试用:有效期剩余 3天 (如                 | 1需延期,请点击右侧购买高级版)                       |                            |                         |            | 购买高级版 |
| 5        |                |                                  | ŀ                                      | 98                         |                         |            |       |
| (据       |                | 灵活的高阶调参训练方式                      | 高效的数据增强策略                              | 便捷的模型部署                    | 嗜服务                     |            |       |
| ž        |                | 支持手动、自动参数设置,系统<br>动设置参数训练完成后可视化展 | 自 有效提高数据利用率,助力训练出<br>示 精度更高、泛化能力更强的模型  | 模型一键发布为离线SE<br>广,最快5分钟即可完成 | DK,适配<br>成业务集成          |            |       |
| ·        |                | 图像分类 物体检测 实例分割                   | 1) 语义分割                                |                            |                         |            |       |
| (务       |                | AutoDL模式预训练模型下载列制                | 专:基于不同部署环境下的专属预训练模型,购买高级版后             | 可根据实际需求联网下载(               | 吏用                      |            |       |
| 個        |                | 适配部署方式                           | 介绍                                     | 预训练模型 ⑦                    | 下载状态                    | 发布时间       | 操作    |
|          |                | 专项适配硬件SDK ②                      | 基于英伟达Jetson系列硬件提供深度适配预训练模型             | 高性能                        | •已下载                    | 2022-04-20 |       |
|          |                |                                  |                                        | 高精度                        | ● 已下载                   | 2022-04-20 |       |
|          |                | 服务器SDK                           | 适配于部署环境为服务器的预训练模型,模型精度更<br>高,对预测资源要求更高 | 高性能                        | ● 已下载                   | 2022-04-20 |       |
|          |                |                                  |                                        | 27 吉祉44                    |                         | 0000 00 10 |       |
|          |                |                                  |                                        | ALIEN 111H6                |                         | 2023-02-16 |       |
|          |                |                                  |                                        | 高精度                        | <ul> <li>已下载</li> </ul> | 2023-02-16 |       |

|                 | 订阅     | く 〉 产品总览                       |                                   |                                    |            | 00           |
|-----------------|--------|--------------------------------|-----------------------------------|------------------------------------|------------|--------------|
| EntyDL          | 订阅购买列表 | 欢迎订阅飞桨EasyDL高级版                |                                   |                                    |            | 帮助中心         |
|                 |        | 免费试用:有效期剩余 4天 (如需延期,           | 请点击右侧购买高级版)                       |                                    |            | 问题反馈<br>案例中心 |
| AN .            |        | 6                              |                                   | 90                                 | 需联网并登录百度账户 | 购买高级版        |
| 数据              |        | 灵活的高阶调参训练方式                    | 高效的数据增强策略                         | 便捷的模型部署服务                          |            | 合作咨询<br>新手教程 |
| <b>≪≫</b><br>开发 |        | 支持手动、自动参数设置,系统自动设置参数训练完成后可视化展示 | 有效提高数据利用率,助力训练出<br>精度更高、泛化能力更强的模型 | 模型一键发布为离线SDK,适配<br>广,最快5分钟即可完成业务集成 |            |              |
| ÷.              |        |                                |                                   |                                    |            |              |

## 点击高级版购买入口跳转到云端登录个人账号

| ○ 百度智能云                                                   | 语言 ~                                          |
|-----------------------------------------------------------|-----------------------------------------------|
| 百度智能云 云智一体深入产业                                            | 百度账号 云账号 ③                                    |
| 聚焦行业核心场景、打造行业标杆应用                                         | 扫码登录 账号登录 手机号登录                               |
| 全线自研的AI大底座,满足产业对智算基础设施的需求<br>先进的技术和丰富的解决方案,与实体经济深度融合,共创价值 | ■ 2014年2月1日<br>第11号                           |
|                                                           | 1000000000000000000000000000000000000         |
| 618新客体验馆                                                  | 2.2                                           |
| 热销云产品限时尝鲜售完即止                                             | 我已阅读并同意 百度用户协议 和 隐私政策<br>使用百度智能云APP/百度APP归码登录 |
|                                                           | 账号与百度APP、贴吧、网盘等产品通用 注册云账号 子用户登录   忘记密码   忘记账号 |
|                                                           |                                               |

## 根据需求选购并点击下一步继续购买操作

| ◎ 购买EasyDL身                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 電版                                                                                                                    |      |    |          |                       |                       |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|------|----|----------|-----------------------|-----------------------|
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                       |      |    |          |                       |                       |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 1 选择产品                                                                                                                | 2 确认 | 丁单 | 3 在线支付   | 4 支付成功                |                       |
| I买EasyDL桌面版                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                       |      |    |          |                       |                       |
| 权年限: 按年                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 腰权 🗸                                                                                                                  |      |    |          |                       |                       |
| <b>≞</b> : −                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 1 +                                                                                                                   |      |    |          |                       |                       |
| N号: 新序                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 列号 🗸                                                                                                                  |      |    |          |                       |                       |
| 买费用: ¥ 200                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 00                                                                                                                    |      |    |          |                       |                       |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                       |      |    |          |                       |                       |
| 买说明:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                       |      |    |          |                       |                       |
| 支持购买新序列号、延+<br>序列号价格参考如下:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 5.已购序列号的有效期。                                                                                                          |      |    |          |                       |                       |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                       |      |    |          |                       |                       |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 价格                                                                                                                    |      |    |          |                       |                       |
| 桌面版(按年授权)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 价格<br>20000元                                                                                                          |      |    |          |                       |                       |
| 盧面飯(按年段(2)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 价格<br>20000元<br>50000元<br>¥20000 重機详続                                                                                 |      |    |          |                       |                       |
| 梁墨版(按年授代)     梁墨版(永久受代)     梁墨版(永久受代)     マン     マン | ☆林橋<br>20000元<br>50000元<br>橋 ¥4橋                                                                                      |      |    |          |                       |                       |
| <ul> <li>▲画版(按年段权)</li> <li>▲画版(永久段权)</li> <li>歩</li> <li>取</li> <li>支付</li> <li>支付</li> <li>您的订单已提</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | <ul> <li>价格</li> <li>20000元</li> <li>50000元</li> <li>4</li> <li>¥20000 或者详绪</li> <li>交, 请支付。</li> </ul>               |      |    |          | 原价:¥20,000.00         | 待支付总金额:¥20,000C       |
| <ul> <li></li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | <ul> <li>价格</li> <li>20000元</li> <li>50000元</li> <li>¥20000 章者详情</li> <li>X20000 章者详情</li> <li>X20000 章者详情</li> </ul> | 配置列表 | 政策 | 时长 代金券金額 | 原价: ¥ 20,000.00<br>新和 | 待支付总金额:¥20,000<br>業支付 |

支付成功后会在购买记录里显示已购买产品信息和序列号信息

确认支付 ¥20,000.00 ¥20,000.00
| 三 〇 〇 百度智能 元 | 控制台总览 全局 🗸                           |                         | Q         |     |   | Ħ | 备案 | Τ¢  | 文档  | 企业 | 支持   | 财务   | 生态    | English | <b>t</b> |
|--------------|--------------------------------------|-------------------------|-----------|-----|---|---|----|-----|-----|----|------|------|-------|---------|----------|
| < 返回EasyDL总览 | < 返回 EasyDL 桌面版                      |                         |           |     |   |   |    |     |     |    |      |      |       |         |          |
| EasyDL桌面版    | 产品介绍:<br>提供本地一站式模型训练与邮署能力,支持<br>了解评情 | 零基础定制高精度模型,本地高效管理数据与模型, | ,一键安装即可使用 | 8   |   |   |    |     |     |    |      |      |       |         | 收起       |
|              | 购买产品列表<br>高级版                        |                         |           |     |   |   |    |     |     |    |      |      |       |         |          |
|              | 购买日期                                 | 激活日期                    |           | 失效日 | 阴 |   |    |     |     | 序列 | E)   |      |       |         |          |
|              |                                      |                         | 没有到       | 文据  |   |   |    |     |     |    |      |      |       |         |          |
|              |                                      |                         |           |     |   |   |    | 共 C | 康 〈 | 1  | > 10 | )条/页 | ✓ 跳转3 | ĒG      | 0        |

## 返回EasyDL桌面版客户端输入序列号完成高级版激活

| É      | 飞桨 EasyDL 操作 | 管理 窗口                      |             |              |  |  |  |  |  |
|--------|--------------|----------------------------|-------------|--------------|--|--|--|--|--|
| • • •  |              | 查看序列号信息                    |             | 飞桨 EasyDL    |  |  |  |  |  |
| EasyDL | 产品介绍         | 激活高级版                      |             |              |  |  |  |  |  |
|        | 数据看板         | <b>获取/延期序列号</b><br>激活序列号功能 |             |              |  |  |  |  |  |
|        |              |                            |             |              |  |  |  |  |  |
| 总览     |              |                            |             |              |  |  |  |  |  |
| C      |              | <b>0</b> ↑                 | 14.29 %     | 暂无数据         |  |  |  |  |  |
| 数据     |              | 训练中的任务数 ②                  | CPU使用率 ⑦    | GPU使用率 ⑦     |  |  |  |  |  |
| <>     |              |                            |             |              |  |  |  |  |  |
| 开发     |              | ■零代码任务总览 〇                 | ■零代码任务总览 C  |              |  |  |  |  |  |
|        |              |                            |             |              |  |  |  |  |  |
| 模型     |              | ▶ 图像分类                     | ♥ 物体检测-     | 矩形框          |  |  |  |  |  |
| 100    |              |                            |             |              |  |  |  |  |  |
| 服务     |              | 0 3                        | 3 0         | 3 3          |  |  |  |  |  |
| 35     |              | 训练中 训练完成                   | 任务总数    训练中 | 训练完成    任务总数 |  |  |  |  |  |
| 订阅     |              |                            |             |              |  |  |  |  |  |
|        |              | ■零代码任务进展 С                 |             |              |  |  |  |  |  |
|        |              | 任务名称                       | 训练方式 🖓 版2   | 5 创建时间       |  |  |  |  |  |
|        |              |                            |             |              |  |  |  |  |  |
|        |              |                            |             |              |  |  |  |  |  |



## AI开发基础知识

### ₯ AI概念及基本原理

人工智能(Artificial Intelligence,英文缩写为AI。它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。人工智能企图生产出一种新的能以人类智能相似的方式做出反应的智能机器,该领域的研究包括机器人、语言识别、图像识别、自然语言处理等。

在EasyDL平台背后主要使用了深度学习的技术,深度学习是机器学习(ML, Machine Learning)领域中一个新的研究方向。通过学习样本数据的内在规律和表示层次,最终目标是让机器能够像人一样具有分析学习能力,能够识别文字、图像和声音等数据。

心 AI模型开发的基本流程介绍



#### \*\*step1:分析业务需求\*\*

在正式启动训练模型之前,需要有效分析和拆解业务需求,明确模型类型如何选择。这里我们可以举一些实际业务场景进行分析。

**举例:原始业务需求—某企业希望为某个高端小区物业做一套智能监控系统,希望对多种现象智能监控并及时预警,包括保安是否在岗、小区是否有有异常噪音、小区内各个区域的垃圾桶是否已满等多个业务功能。**针对这个原始业务需求,我们可以分析出不同的监控对象所在的位置不同、监控的数据类型不同(有的针对图片进行识别、有的针对声音进行判断),需要多个模型综合应用。**监控保安是否在岗**——通过图像分类模型进行判断 **监控小区是否有异常噪音**——定时收集声音片段通过声音分类模型进行判断 **监控小区内各个区域垃圾桶是否已满**——由于监控区域采集的画面可能会存在多个垃圾桶,此处需要通过物体检测模型进行判断。

#### \*\*step2:采集/收集数据\*\*

在通过上述第一步分析出基本的模型类型,需要进行相应的数据收集工作。数据的主要原则为**尽可能采集真实业务场景一致的数据,并覆盖可能 有的各种情况** 

#### \*\*step3:标注数据\*\*

采集数据后,可以通过EasyDL在线标注工具或线下其他标注工具对已有的数据进行标注。如上述保安是否在岗的图像分类模型,需要将监控视频 抽帧后的图片按照【在岗】及【未在岗】两类进行整理;小区内各个区域垃圾桶是否已满,需要将监控视频抽帧后的图片标注其中每个垃圾桶的 【空】【满】两种状态进行标注。

#### \*\*step4:训练模型\*\*

训练模型阶段可以将已有标注好的数据基于已经确定的初步模型类型,选择算法进行训练。通过使用EasyDL平台,可以可视化在线操作训练任务的启停、训练任务的配置。可以大幅减少线下搭建训练环境、自主编写算法代码的相关成本。

#### \*\*step5:评估模型效果\*\*

训练后的模型在正式集成之前,需要评估模型效果是否可用。在这个环节上EasyDL提供了详细的模型评估报告,以及在线可视化上传数据测试模型效果的功能。

#### \*\*step6:部署模型\*\*

当确认模型效果可用后,可以将模型部署至生产环境中。传统的方式需要将训练出的模型文件加入工程化相关处理,通过使用EasyDL,可以便捷 地将模型部署在公有云服务器或本地设备上,通过API或SDK集成应用,或直接购买软硬一体产品,有效应对各种业务场景所需,提供效果与性能 兼具的服务。

## 快速开始

#### 用零代码开发实现图像分类

∞ 示例说明

图像分类模型主要用于识别一张图中是否是某类物体/状态/场景,适合图片中主体或状态单一的场景。本文以害虫识别模型在macOS客户端中的 使用为示例演示图像分类模型训练全过程。

#### 心 实现步骤

只需八步即可完成自定义AI模型的训练及发布的全过程。

#### Step1:提前准备训练数据

图像分类需要提供包含不同类别的图片并标注,完成后即可训练对应图像分类模型,自动识别图中是否包含某类物体/状态/场景,下面我们来看 看这次训练所需的昆虫图片示例:

图片数量越多理论上训练效果越好,图像分类的图片数量建议每个类别不低于20张图片



## Step2:创建数据集

在数据总览界面点击【创建数据集】

| • • •              | 数据           | く > 我的数据总览        | ? ©          |
|--------------------|--------------|-------------------|--------------|
| EnsyDL<br>490      | 数据总览<br>数据标注 | 创建放掘集             | 输入数据集名称或ID Q |
| 总览<br>【<br>】<br>数据 |              |                   |              |
| 开发<br>研发<br>模型     |              | 还没有创建数据集<br>创建数据集 |              |
| 部署<br>33           | •            |                   |              |
| 订阅                 |              |                   |              |
|                    |              |                   |              |
|                    |              |                   |              |
|                    |              |                   |              |

## 在数据集创建界面输入数据集名称、选择标注类型后点击【完成】

く > 创建数据集

| 数据集名称 | 粉蝶or蝗虫 |                                        |      |
|-------|--------|----------------------------------------|------|
| 数据类型  | 图片     |                                        |      |
| 标注类型  | 图像分类   | 物体检测                                   | 实例分割 |
|       |        | 750<br>7<br>7<br>7<br>7<br>7<br>7<br>7 |      |
|       | 完成     |                                        |      |

## Step3 : 导入数据

数据集创建完成后可在【数据总览】查看已创建完成的数据集,点击【导入】跳转至数据导入界面

| 粉蝶or蝗虫 🖸 |     |      |          |      | 回 删除 |
|----------|-----|------|----------|------|------|
| 数据集ID    | 数据量 | 标注类型 | 标注状态     | 操作   |      |
| 20 💮     | 0   | 图像分类 | 0% (0/0) | 导入导出 |      |

数据导入支持无标注信息、有标注信息两种数据标注状态的数据以及多种导入方式,以下为无标注信息图片的导入为示例,其余各类型导入方式 可参考 导入图像数据选择数据标注状态与文件路劲



在数据总览页找到需要标注的数据集,点击【查看与标注】,跳转至标注页面

\_

| < > 我的数据 | 舌忌览 |      |           |          | (?)        | 0 |
|----------|-----|------|-----------|----------|------------|---|
| 创建数据集    |     |      |           |          | 输入数据集名称或ID | Q |
| 粉蝶or蝗虫 🖸 |     |      |           |          | 靣 删除       |   |
| 数据集ID    | 数据量 | 标注类型 | 标注状态      | 操作       |            |   |
| 20 💮     | 40  | 图像分类 | 0% (0/40) | 查看与标注 导入 | 导出         |   |

### 在左侧标签栏下,点击【添加标签】创建数据集标签



分别输入粉蝶、蝗虫并点击【确认】添加数据标签

# 筛选 🗸

|                       | 粉蝣    |                         | 确定取消 |
|-----------------------|-------|-------------------------|------|
| 点击图片                  | 右下角红框 | 内图标进入到数据标注界面            |      |
| <ul> <li>数</li> </ul> | 掘     | く 〉 【图片】粉蝶or蝗虫/V1/查看与标注 | o •  |



在当前图片下选择右侧标签栏内的某一类别,代表为图片打上相应的标签,点击【保存当前图片】或直接点击下一张图标,在保存标注结果后自 动跳转至下一张。标注完所有图片后,该数据集便可用于后续训练任务



### Step5:创建训练任务

### 在任务总览界面点击【创建任务】

| 收起<br>凯练参数灵活性更高,该两种训练模式可灵活选择。 对于训           |
|---------------------------------------------|
| LUIDAYYAXALLIXID, XAYTTUDAYXAYYXAXD+° XJJUI |
|                                             |
|                                             |
| 4.发布为模型                                     |
| 效果满足业务需求的任务可<br>发布为模型,进行邮署操作                |
| 输入任务名称或ID Q                                 |
| 吕 训练 ⑤ 历史版本 直 删除                            |
| 操作                                          |
| 查看版本配置 校验<br>在notebook打开 ⑦                  |
|                                             |
|                                             |
|                                             |
|                                             |

## 在训练任务创建界面输入任务名称、选择任务类型后点击【创建任务】



任务创建完成后,点击【训练】,进入训练配置阶段

| 【图像分类】害虫识别 🖄 任务ID: 13 |                                                  | 吕 训练 面 删除 |
|-----------------------|--------------------------------------------------|-----------|
|                       | 任务创建成功,若无数据集请先在"数据中心"创建 ,上传训练数据 训练 后,可以在此处查看最新版本 |           |

### 根据需求选择各项训练配置后,添加训练数据集,点击【开始训练】

| •••        | 开发          | く 〉 训练模型           |          |                                  |         | (  | 0 | 9 |
|------------|-------------|--------------------|----------|----------------------------------|---------|----|---|---|
| EasyDL     | 零代码开发       |                    |          |                                  |         |    |   |   |
|            | Notebook 开发 | 训练方式               | AutoDL模式 | 高 高级调参模式 ⑦                       |         |    |   |   |
| 急災         |             | ▌ 训练配置             |          | 高级版专享                            |         |    |   |   |
| C          |             | 导出类型               | ○ 仅导出源文件 | ● 导出源文件与离线SDK ⑦                  |         |    |   |   |
| 26.58      |             | 部署方式               | 服务器SDK   | ● 通用小型设备SDK ● 专项适配硬件SD           | к @     |    |   |   |
| 开发         |             | 选择算法 ⑦             | ● 超高性能 ⑦ | ○ 高性能 ⑦ ○ 高精度 ⑦ 查看算法性的           | 也及适配硬件  |    |   |   |
| <b>秋</b> 田 |             | 高级训练配置             |          |                                  |         |    |   |   |
|            |             | 训练完成后同<br>步发布为模型 ⑦ | ○ 是 ○ 否  |                                  |         |    |   |   |
|            |             |                    | *模型名称    | 害虫识别                             |         |    |   |   |
| 订阅         |             |                    | 模型类型     | 图像分类                             |         |    |   |   |
|            |             |                    | 模型版本     | V2                               |         |    |   |   |
|            |             |                    | 版水描述1    |                                  |         |    |   |   |
|            |             |                    | 加大中加上口   |                                  |         |    |   |   |
|            | i           |                    |          | 0 / 100                          |         |    |   |   |
|            |             |                    |          |                                  |         |    |   |   |
|            |             | ▶ 添加数据             |          |                                  |         |    |   |   |
|            |             | * 添加数据集            | + 请选择    |                                  |         |    |   |   |
|            |             |                    | 数据集      | 版本                               | 分类数量    | 操作 |   |   |
|            |             |                    | 害虫识别     | V1                               | 2       | 移除 |   |   |
|            |             | 自定义验证集 ⑦           |          |                                  |         |    |   |   |
|            |             | 自定义测试集 ⑦           |          |                                  |         |    |   |   |
|            |             | 高*数据增强策略(          | ⑦ 🗿 默认配置 | ○ 手动配置                           |         |    |   |   |
|            |             | 训练环境               | 了解详情     |                                  |         |    |   |   |
|            |             |                    |          |                                  |         |    |   |   |
|            |             |                    | 名称       | 规格                               |         |    |   |   |
|            |             |                    | O CPU    | Intel(R) Core(TM) i5-8257U CPU @ | 1.40GHz |    |   |   |
|            |             | TT 66420164        |          |                                  |         |    |   |   |
|            |             | 开始训练               |          |                                  |         |    |   |   |

在任务总览页任务列表下,可以看到处于训练状态的训练任务,将鼠标放置感叹号图标处,即可查看训练进度

| くう任多                                                                        | <b></b> 身总览 |             |                    |      |      |        |           | ? (0) |
|-----------------------------------------------------------------------------|-------------|-------------|--------------------|------|------|--------|-----------|-------|
| 零代码开发介绍<br>面向应用开发者提供AutoDL训练模式及高阶调参训练模式,AutoDL训练模式自动化程度更高,高阶调参训练模式训练参数设置灵活。 |             |             |                    |      |      |        |           |       |
| 创建任务                                                                        |             |             |                    |      |      |        | 输入任务名称或ID | Q     |
| 【图像分类】                                                                      | 害虫识别 🖄 任务   | D: 13       |                    |      |      |        | ④ 历史版本 位  | 「删除   |
| 任务版本                                                                        | 训练方式        | 训练算法        | 训练状态               | 模型效果 | 对应模型 | 操作     |           |       |
| V1                                                                          | AutoDL模式    | 仅导出源文件-自行部署 | _高性能 ● 训练中 🔢       | -    |      | 查看版本配算 | t 删除 终止训练 |       |
|                                                                             |             |             | 训练进度:<br>剩余时间:12分钟 | 45%  |      |        |           |       |

### Step6:模型校验

训练完成后,可在任务列表下,点击【校验】

## Baidu 百度智能云文档

| 〈 〉 任务                                                                      | 务总览         |               |        |                                                |      |              |         | 0   | $\odot$ |
|-----------------------------------------------------------------------------|-------------|---------------|--------|------------------------------------------------|------|--------------|---------|-----|---------|
| 零代码开发介绍<br>面向应用开发者提供AutoDL训练模式及高阶调参训练模式,AutoDL训练模式自动化程度更高,高阶调参训练模式训练参数设置灵活。 |             |               |        |                                                |      |              |         |     | 展开      |
| 创建任务                                                                        |             |               |        |                                                |      |              | 输入任务名称或 | ÎD  | Q       |
| 【图像分类】                                                                      | 害虫识别 🖸 任务ID | : 13          |        |                                                |      | 吕 训练         | ④ 历史版本  | 面 删 | 除       |
| 任务版本                                                                        | 训练方式        | 训练算法          | 训练状态   | 模型效果                                           | 对应模型 | 操作           |         |     |         |
| V1                                                                          | AutoDL模式    | 仅导出源文件自行部署高性能 | ● 训练完成 | top1准确率: 100.00%<br>top5准确率: 100.00%<br>完整评估结果 |      | 查看版本配置<br>删除 | 1 发布为模型 | 校验  |         |

## 点击【添加图片】,进行模型校验

| く > 校验  | <b>☆模型</b>   |         |        |              |               | 0 | 0 |
|---------|--------------|---------|--------|--------------|---------------|---|---|
| 选择任务    | 害虫识别 >       | 选择版本 V1 | ∨ 训练算法 | 仅导出源文件自行部署高性 | 能             |   |   |
| 当前模型准确率 | 100.00% 评估报告 |         |        | 调整阈值 🔹       | 0.03          |   |   |
|         |              |         | 识别结果:  | 预测分类         | 置信度 > 3.00% ~ |   |   |
|         |              |         | Ĥ      | 839          | 99.98%        |   |   |

点击添加图片

## Step7:发布为模型

### 确认模型效果满意后,可在任务列表下,点击【发布为模型】

| く > 任約 | ( ) 任务总缆                                        |                                           |                      |                                                |      |              |        |     |    |
|--------|-------------------------------------------------|-------------------------------------------|----------------------|------------------------------------------------|------|--------------|--------|-----|----|
| 零代码开发介 | 绍                                               | で、「「「「「」」、「」、「」、「」、「」、「」、「」、「」、「」、「」、「」、「 | 満ず白み少程度東京 京約3        | 用条训练结计训练条数沿带寻汗                                 |      |              |        |     | 展开 |
| 创建任务   | 1/2/1/2/19/19/19/19/19/19/19/19/19/19/19/19/19/ |                                           | 天56日40101至12(又同),同初日 | 99 9 MISINTR⊁VMISN 9 XA KA≣KALI                |      |              | 输入任务名称 | 或ID | Q  |
| 【图像分类】 | 害虫识别 🖸 任务IC                                     | D: 13                                     |                      |                                                |      | 吕 训练         | ⑤ 历史版本 | Ū ł | 删除 |
| 任务版本   | 训练方式                                            | 训练算法                                      | 训练状态                 | 模型效果                                           | 对应模型 | 操作           |        |     |    |
| V1     | AutoDL模式                                        | 仅导出源文件自行部署高性能                             | ● 训练完成               | top1准确率: 100.00%<br>top5准确率: 100.00%<br>完整评估结果 |      | 查看版本配置<br>删除 | 发布为模型  | 校验  |    |

输入模型名称、版本描述后点击【确认】正式发布未模型

发布至模型仓库

| *模型名称 | 害虫识别    |
|-------|---------|
| 模型类型  | 图像分类    |
| 模型版本  | V1      |
| 版本描述  |         |
|       | 0 / 100 |
|       |         |
|       | 取 消 确 定 |

## Step8:导出模型文件或部署SDK

根据您在训练时选择的部署方式,可在模型总览版本列表中,点击【导出模型文件】或【部署】

|   |                    | 模型   |   | < > 版    | 本列表         |          |            |    |                     |   |               | 0 0 |
|---|--------------------|------|---|----------|-------------|----------|------------|----|---------------------|---|---------------|-----|
|   | EntyOL             | 模型列表 |   | 模型名称: 害: | 由识别         |          | 模型ID: 5    |    |                     |   |               |     |
|   |                    | 模型部署 |   | 模型类型: 图  | <b>黎</b> 分类 |          | 模型来源:零代码开发 |    |                     |   |               |     |
|   | az<br>C            |      |   | 版本       | 对应任务        | 训练方式     |            | 描述 | 导入时间 🌲              |   | 操作            |     |
|   | 数据                 |      |   | V2       | 10-V2       | AutoDL模式 |            |    | 2021-11-22 08:31:21 | [ | 导出模型文件 ⑦ 部署 💬 |     |
|   | <b>≮&gt;</b><br>开发 |      |   | V1       | 9-V1        | AutoDL模式 |            |    | 2021-11-22 07:42:03 |   | 导出模型文件 ⊙      |     |
| I |                    |      |   |          |             |          |            |    |                     |   |               |     |
|   |                    |      |   |          |             |          |            |    |                     |   |               |     |
|   | 8748<br>313        |      | 4 |          |             |          |            |    |                     |   |               |     |
|   | 订阅                 |      |   |          |             |          |            |    |                     |   |               |     |
|   |                    |      |   |          |             |          |            |    |                     |   |               |     |
|   |                    |      |   |          |             |          |            |    |                     |   |               |     |
|   |                    |      |   |          |             |          |            |    |                     |   |               |     |
|   |                    |      |   |          |             |          |            |    |                     |   |               |     |
|   |                    |      |   |          |             |          |            |    |                     |   |               |     |
|   |                    |      |   |          |             |          |            |    |                     |   |               |     |
|   |                    |      |   |          |             |          |            |    |                     |   |               |     |
|   |                    |      |   |          |             |          |            |    |                     |   |               |     |

 $\times$ 

点击部署进入服务发布界面,选择模型部署设备的芯片型号,点击【发布】,跳转至服务列表页

く 〉 模型部署

| 选择模型    | 害虫识别                                                                                         | $\sim$ |
|---------|----------------------------------------------------------------------------------------------|--------|
| 模型版本    | V2                                                                                           | $\sim$ |
| 训练方式    | 零代码开发–AutoDL模式                                                                               |        |
| 训练算法    | 服务器–高性能                                                                                      |        |
| 选择部署环境  | 服务器                                                                                          |        |
| 选择系统和芯片 | ● 本地发布                                                                                       |        |
|         | <ul> <li>☐ ↓ Linux</li> <li>通用X86 CPU 英伟达GPU</li> <li>一 € Windows</li> <li>英伟达GPU</li> </ul> |        |

发布

在服务列表中选择需要导出的模型SDK,点击【导出SDK】,选择存储位置后即可完成模型SDK的导出

|   |                  | 部署   | く 〉 服务总览    |              |                 |                         |            |                  |        | 0 0 |
|---|------------------|------|-------------|--------------|-----------------|-------------------------|------------|------------------|--------|-----|
|   | EasyOL<br>AP     | 服务列表 | 飞桨EasyDL推出智 | 1能边缘控制台、助力本: | 地部署场景用户高效管理端侧预  | <b>〕</b> 测设备与服务,极速完     | 成本地数据与模型串联 | 、提高模型落地效率。了解     | 洋情 ×   |     |
|   | 总览               |      | 服务器 通用小     | 型设备  专项适配硬作  | ¥ @             |                         |            |                  | 输入模型名称 | Q   |
|   | ひ数据              |      | 模型名称        | 发布版本 📏       | 应用平台            | 发布状态                    | 发布方式       | 发布时间 关           | 操作     |     |
|   | <b>*&gt;</b>     |      | 害虫识别        | 5-V2         | 通用X86 CPU-Linux | <ul> <li>已发布</li> </ul> | 本地部署       | 2021-11-22 20:51 | 导出SDK  |     |
|   |                  |      | 猫狗分类        | 1-V1         | 英伟达GPU-Windows  | • 已发布                   | 本地部署       | 2021-09-27 17:34 | 导出SDK  |     |
|   | 模型               |      |             |              |                 |                         |            |                  |        |     |
| 1 | 部署               |      |             |              |                 |                         |            |                  |        |     |
|   | <b>3</b> 6<br>订阅 | •    |             |              |                 |                         |            |                  |        |     |
|   |                  |      |             |              |                 |                         |            |                  |        |     |
|   |                  |      |             |              |                 |                         |            |                  |        |     |
|   |                  |      |             |              |                 |                         |            |                  |        |     |
|   |                  |      |             |              |                 |                         |            |                  |        |     |
|   |                  |      |             |              |                 |                         |            |                  |        |     |
|   |                  |      |             |              |                 |                         |            |                  |        |     |
|   |                  |      |             |              |                 |                         |            |                  |        |     |
|   |                  |      |             |              |                 |                         |            |                  |        |     |

## 用零代码开发实现物体检测

心 示例说明

物体检测模型主要用于检测图中每个物体的位置、类型。适合图中有多个主体要识别、或要识别主体位置及数量的场景。本文以螺丝螺母识别模型在macOS客户端中的使用为示例演示物体检测模型训练全过程。

#### 心 实现步骤

只需八步即可完成自定义AI模型的训练及发布的全过程。

#### Step1:提前准备训练数据

物体检测需要提供包含目标物体的图片并标注物体即可训练物体检测模型,自动识别图中所有目标物体的位置、名称,下面我们来看看这次需要计数的包含螺丝螺母的图片示例:

图片数量越多理论上训练效果越好,物体检测的图片数量建议每个类别不低于20张图片 注意图片需要为业务生产的真实环境所采集的图片, 与真实场景越贴近,训练模型效果越佳



Step2:创建数据集

在数据总览界面点击【创建数据集】

| •••                | 数据     | < > 我的数据总览        | 0          | 0 |
|--------------------|--------|-------------------|------------|---|
| EasyDL.            | 数据总览   | 创建数据集             | 输入数据集名称或ID | Q |
| 急览                 | 蚁城(标)土 |                   |            |   |
| <b>し</b><br>数据     |        |                   |            |   |
| <b>∢&gt;</b><br>开发 |        |                   |            |   |
| ↓ 検型               |        | 还没有创建数据集<br>创建数据集 |            |   |
| 部署                 | ٩      |                   |            |   |
| <b>38</b><br>订阅    |        |                   |            |   |
|                    |        |                   |            |   |
|                    |        |                   |            |   |
|                    |        |                   |            |   |
|                    |        |                   |            |   |
|                    |        |                   |            |   |

在数据集创建界面输入数据集名称、选择标注类型后点击【完成】

く > 创建数据集

| 数据集名称 | 螺母and螺帽 |      |             |
|-------|---------|------|-------------|
| 数据类型  | 图片      |      |             |
| 标注类型  | 图像分类    | 物体检测 | <b>实例分割</b> |
|       | 完成      |      |             |

Step3 : 导入数据

数据集创建完成后可在【数据总览】查看已创建完成的数据集,点击【导入】跳转至数据导入界面

| 螺母and螺帽 🖸 | 螺母and螺帽 🖸 |      |          |      |  |  |
|-----------|-----------|------|----------|------|--|--|
| 数据集ID     | 数据量       | 标注类型 | 标注状态     | 操作   |  |  |
| 22        | 0         | 物体检测 | 0% (0/0) | 导入导出 |  |  |

数据导入支持无标注信息、有标注信息两种数据标注状态的数据以及多种导入方式,以下为无标注信息图片的导入为示例,其余各类型导入方式 可参考 导入图像数据选择数据标注状态与文件路劲



| 螺母and螺帽 🗌 | 螺母and螺帽 🖸 |      |           |             |  |  |
|-----------|-----------|------|-----------|-------------|--|--|
| 数据集ID     | 数据量       | 标注类型 | 标注状态      | 操作          |  |  |
| 22 😳      | 69        | 物体检测 | 0% (0/69) | 查看与标注 导入 导出 |  |  |

在左侧标签栏下,点击【添加标签】创建数据集标签

| •••                      | 数据   | く 〉 【图片】螺母and螺帽/V1/查看与标注            |           |          |       | 0 0              |
|--------------------------|------|-------------------------------------|-----------|----------|-------|------------------|
| EasyDL<br>H <sup>J</sup> | 数据总览 | <b>全部 (69)</b> 有标注信息 (0) 无标注信息 (69) |           |          |       | 导入图片标注图片         |
| 总览                       | 数据标注 | 筛选 🗸                                |           |          |       | 全选 💿 删除 🗈 删除失效图片 |
| C<br>数据                  |      | 标签栏 添加标签                            |           |          |       |                  |
| <b>≪≫</b><br>开发          |      | 请输入标签名称 Q                           | Contra la | ( ) / s) | 0000  |                  |
| <b>使</b> 型               |      |                                     | 无标签       | 无标签      | 无标篮   | 无标篮              |
| 部署<br>100                | 4    |                                     | (BAR)     |          | and a | Log2             |
|                          |      |                                     | 无标签       | 无标签      | 无标签   | 无标签              |
|                          |      | 暂无可用标签 ,请点击上方按钮添加                   |           |          |       | at to            |
|                          |      |                                     | 无标签       | 无标签      | 无标签   | 无标签              |
|                          |      |                                     |           |          |       |                  |

分别输入螺丝、螺母并点击【确认】添加数据标签

筛选 🗸



#### 点击图片右下角红框内图标进入到数据标注界面

| •••             | 数据   | く > 【图片】螺母and螺帽   | /V1/查看与标注  |                           |               |         | 00               |
|-----------------|------|-------------------|------------|---------------------------|---------------|---------|------------------|
| EasyDL .        | 数据总览 | 全部 (69) 有标注信息 (0) | 无标注信息 (69) |                           |               |         | 导入图片标注图片         |
| 急览              | 数据标注 | 筛选 🗸              |            |                           |               |         | 全选 💿 删除 💿 删除失效图片 |
| C<br>数据         |      | 标签栏               | 添加标签       |                           |               |         |                  |
| <b>《》</b><br>开发 |      | 请输入标签名称           | Q          | a day                     |               | 679     | 310              |
| ● 模型            |      | 标签名               | 标注框数.11    | 无标签 🖸 🔀                   | 无标签。          | 无标签     | 无标签              |
|                 |      | 螺母                | 0          | 点击【编辑icon】即<br>入"单张图片标注模式 | 可进            |         |                  |
| 88              | 4    | 螺丝                | 0          | ( toy E                   | (Page)        | 140     | 10 % z           |
| 订间              |      |                   |            |                           |               |         |                  |
|                 |      |                   |            | 无标签                       | 无标签           | 无标签     | 无标签              |
|                 |      |                   |            |                           |               |         |                  |
|                 |      |                   |            | (                         | Carlos Carlos | A a a a | Cry C            |
|                 |      |                   |            | 无标签                       | 无标签           | 无标签     | 无标签              |
|                 |      |                   |            |                           |               |         |                  |

在当前图片下选择右侧标签栏内的某一类别,代表为图片打上相应的标签,点击【保存当前图片】或直接点击下一张图标,在保存标注结果后自 动跳转至下一张。标注完所有图片后,该数据集便可用于后续训练任务



Step5:创建训练任务

在任务总览界面点击【创建任务】

| •••             | 开发                   | く 〉 任务总览                                      |                                   |                        |                                                |                |                              |                        | 00           |
|-----------------|----------------------|-----------------------------------------------|-----------------------------------|------------------------|------------------------------------------------|----------------|------------------------------|------------------------|--------------|
| Casy/DC<br>AV   | 零代码开发<br>Notebook 开发 | 零代码开发介绍<br>面向应用开发者提供AutoDL划<br>练完成的任务可通过完整评估: | ll练模式及高阶调参训练模式,<br>报告、校验来验证效果,满足」 | AutoDL训练模式<br>业务应用需求的任 | 自动化程度更高,高级调参<br>务可发布为模型进行后续部署                  | 莫式支持设置<br>针操作。 | 训练参数灵活性更高,该两种                | 中训练模式可灵活选持             | 收起<br>2。 对于训 |
| じ数据             |                      | ~1                                            | Ę                                 |                        |                                                |                |                              |                        |              |
| <b>(</b>        |                      | 1.发起训练                                        | 2.选择数据                            | ł                      | 3.评估校验                                         |                | 4.发布为模型                      |                        |              |
| <b>开发</b><br>御刊 |                      | 选择训练方式并完成训练参<br>数配置                           | > 选择训练数据集, 自<br>动配置数据增强策略         | 动或手 )<br>              | 任务训练完成后,通过评<br>报告,校验验证任务效果                     | ± →            | 效果满足业务需求的任务可<br>发布为模型,进行部署操作 |                        |              |
| 服务              | •                    | 创建任务                                          |                                   |                        |                                                |                |                              | 输入任务名称或ID              | Q            |
| 36              |                      | 【图像分类】害虫识别 🖸                                  | 任务ID: 36                          |                        |                                                |                | 吕 训练                         | ⑤ 历史版本                 | 1 删除         |
| 1.1 (4)         |                      | 任务版本 训练方式                                     | 训练算法                              | 训练状态                   | 模型效果                                           | 对应模型           | 操作                           |                        |              |
|                 |                      | V1 高级调参模式                                     | MobileNet-MobileNetV2             | ● 训练完成                 | top1准确率: 100.00%<br>top5准确率: 100.00%<br>完整评估结果 | 害虫识别           | 查看<br>在nc                    | 版本配置 校验<br>btebook打开 ⑦ |              |
|                 |                      |                                               |                                   |                        |                                                |                |                              |                        |              |
|                 |                      |                                               |                                   |                        |                                                |                |                              |                        |              |
|                 |                      |                                               |                                   |                        |                                                |                |                              |                        |              |

在训练任务创建界面输入任务名称、选择任务类型后点击【创建任务】



## 任务创建完成后,点击【训练】,进入训练配置阶段

| 【物体检测】螺丝螺母识别 [2] 任务ID:14                         | 吕 训练 | 面 删除 |
|--------------------------------------------------|------|------|
| 任务创建成功,若无数据集请先在"数据中心"创建, 上传训练数据 训练 后,可以在此处查看最新版式 |      |      |

## 根据需求选择各项训练配置后,添加训练数据集,点击【开始训练】

|   | •••            | 开发          | く 〉 训练模型                            |          |                     |                       |      |    | 0 |
|---|----------------|-------------|-------------------------------------|----------|---------------------|-----------------------|------|----|---|
|   | Easy DL.       | 零代码开发       |                                     |          |                     |                       |      |    |   |
|   | 总览             | Notebook 开发 | <ul><li>训练方式</li><li>训练配置</li></ul> | AutoDL模式 | 高 高级调参模式 ⑦          |                       |      |    |   |
|   | C              |             | 导出类型                                | ○ 仅导出源文件 | □ 与出源文件与离线SDK       | 0                     |      |    |   |
|   | 数据             |             | 选择算法 ⑦                              | ● 超高性能 ⑦ | ○ 高性能 ⑦ ○ 高精        | 腹 ⑦ 查看算法性能及适配         | 硬件   |    |   |
| 1 | 开发             |             | 高级训练配置                              |          |                     |                       |      |    |   |
|   | <b>拿</b><br>模型 |             | 训练完成后同<br>步发布为模型 ⑦                  | 🔾 是 💽 否  |                     |                       |      |    |   |
|   | 服务             |             | ▮ 添加数据                              |          |                     |                       |      |    |   |
|   |                | 4           | * 添加数据集                             | + 请选择    |                     |                       |      |    |   |
|   | 13 181         |             |                                     | 数据集      | 版本                  |                       | 标签数量 | 操作 |   |
|   |                |             |                                     | 螺丝螺母检测   | V1                  |                       | 2    | 移除 |   |
|   |                |             | 自定义验证集 ⑦                            |          |                     |                       |      |    |   |
|   |                |             | 自定义测试集 ⑦                            |          |                     |                       |      |    |   |
|   |                |             | 高 *数据增强策略                           | ② 💽 默认配置 | ○ 手动配置              |                       |      |    |   |
|   |                |             | 训练环境                                | 了解详情     |                     |                       |      |    |   |
|   |                |             |                                     | 名称       | 规格                  |                       |      |    |   |
|   |                |             |                                     | O CPU    | Intel(R) Core(TM) i | 5-8257U CPU @ 1.40GHz |      |    |   |
|   |                |             | 开始训练                                |          |                     |                       |      |    |   |
|   |                |             | 71 81 81 50                         |          |                     |                       |      |    |   |

在任务总览页任务列表下,可以看到处于训练状态的训练任务,将鼠标放置感叹号图标处,即可查看训练进度

| 【物体检测】 | 螺丝螺母识别 🖂 1 | 任务ID: 14 |             |      |      | ⑤ 历史版本 前 删除      |
|--------|------------|----------|-------------|------|------|------------------|
| 任务版本   | 训练方式       | 训练算法     | 训练状态        | 模型效果 | 对应模型 | 操作               |
| V1     | AutoDL模式   | 服务器高性能   | • 训练中 💶     | -    |      | 查看版本配置 删除 终止训练   |
|        |            |          | 训练进度:       | 1%   |      |                  |
| 【图像分类】 | 害虫识别 🖸 任务I | D: 13    | 剩余时间:1小时9分钟 |      |      | 😑 训练 🕑 历史版本 🔟 删除 |

Step6:模型校验

## 训练完成后,可在任务列表下,点击【校验】

| 〈 〉 任師                                                                                                              | 〈 > 任务总宽   |                 |        |                                                |      |              |         |      |   |  |  |
|---------------------------------------------------------------------------------------------------------------------|------------|-----------------|--------|------------------------------------------------|------|--------------|---------|------|---|--|--|
| 零代码开发介绍<br>面向应用开发者提供AutoDL训练模式及高阶调参训练模式,AutoDL训练模式自动化程度更高,高阶调参训练模式训练参数设置灵活。                                         |            |                 |        |                                                |      |              |         |      |   |  |  |
| 副内国在ガイス省当進 FYAULOOL WIESSINGLY, AND MISSINGLY, AULOOL WIESSINGLY 日本化した AC K (M) (M) (M) (M) (M) (M) (M) (M) (M) (M) |            |                 |        |                                                |      |              |         |      | 2 |  |  |
| 【物体检测】                                                                                                              | 螺丝螺母识别 🖸 任 | 务ID:14          |        |                                                |      | 吕 训练         | ⑤ 历史版本  | 面 删除 | f |  |  |
| 任务版本                                                                                                                | 训练方式       | 训练算法            | 训练状态   | 模型效果                                           | 对应模型 | 操作           |         |      |   |  |  |
| V1                                                                                                                  | AutoDL模式   | 仅导出源文件—自行部署—高性能 | ● 训练完成 | top1准确率: 100.00%<br>top5准确率: 100.00%<br>完整评估结果 |      | 查看版本配量<br>删除 | 置 发布为模型 | 校验   |   |  |  |

点击【添加图片】,进行模型校验



| 识别结果 如何优化效果? |                |
|--------------|----------------|
| 周整阈值         | 0.3            |
| 预测标签         | 置信度 > 30.00% ~ |
| 7. 螺丝        | 99.96%         |
| 6. 螺丝        | 99.90%         |
| 5. 螺丝        | 99.84%         |
| 4. 螺丝        | 99.77%         |
| 3. 螺母        | 99.66%         |
| 2. 螺母        | 99.59%         |
|              |                |

## 点击添加图片

Step7:发布为模型

## 确认模型效果满意后,可在任务列表下,点击【发布为模型】

| < > 任务总览                                                                                                                                   |          |               |        |                                                |      |              |                       |          |  |  |
|--------------------------------------------------------------------------------------------------------------------------------------------|----------|---------------|--------|------------------------------------------------|------|--------------|-----------------------|----------|--|--|
| 零代码开发介绍         展行           面向应用开发者提供AutoDL训练模式及高阶调参训练模式, AutoDL训练模式自动化程度更高, 高阶调参训练模式训练参数设置灵活。            创建任务         输入任务名称或D         Q |          |               |        |                                                |      |              |                       |          |  |  |
| 创建任务<br>【物体检测】 螺丝螺母识别 [Z] 任务[D]: 14 二 二 二 二 二 二 二 二 二 二 二 二 二 二 二 二 二 二                                                                    |          |               |        |                                                |      |              | 输入任务名称或ID<br>⑤ 历史版本 前 | <br>[ 删除 |  |  |
| 任务版本                                                                                                                                       | 训练方式     | 训练算法          | 训练状态   | 模型效果                                           | 对应模型 | 操作           |                       | /        |  |  |
| V1                                                                                                                                         | AutoDL模式 | 仅导出源文件自行部署高性能 | ● 训练完成 | top1准确率: 100.00%<br>top5准确率: 100.00%<br>完整评估结果 |      | 查看版本配]<br>删除 | 置 发布为模型 校验            | Þ        |  |  |

## 输入模型名称、版本描述后点击【确认】正式发布未模型

| 发布至模型 | 仓库      |
|-------|---------|
| *模型名称 | 螺丝螺母识别  |
| 模型类型  | 物体检测    |
| 模型版本  | V1      |
| 版本描述  | 0 / 100 |
|       | 取 消 确 定 |

## Step8:导出模型文件或部署SDK

根据您在训练时选择的部署方式,可在模型总览版本列表中,点击【导出模型文件】或【部署】

## Baidu 百度智能云文档

|                   | 模型   | < > 1    | 版本列表   |            |            |    |                     |             | 0 | 0 |
|-------------------|------|----------|--------|------------|------------|----|---------------------|-------------|---|---|
| EnsyOL            | 模型列表 | 模型名称: 蚊  | 累丝螺母识别 |            | 模型ID: 5    |    |                     |             |   |   |
|                   | 模型部署 | 模型类型: \$ | 物体检测   |            | 模型来源:零代码开发 |    |                     |             |   |   |
| 总宽                |      | 版本       | 对应任务   | 训练方式       |            | 描述 | 导入时间 👙              | 操作          |   |   |
| ひ数据               |      | V2       | 10-V2  | AutoDL模式   |            |    | 2021-11-22 08:31:21 |             |   |   |
| $\langle \rangle$ |      |          | 0.10   | Aut-DI ### |            |    | 0001 11 00 07.40.00 | Buittente O |   |   |
| 开发                |      | VI       | 5-01   | Autoblight |            |    | 2021-11-22 07.42.03 | GUREXIL Q   |   |   |
| 模型                |      |          |        |            |            |    |                     |             |   |   |
|                   |      |          |        |            |            |    |                     |             |   |   |
| 32                | 4    |          |        |            |            |    |                     |             |   |   |
| 订阅                |      |          |        |            |            |    |                     |             |   |   |
|                   |      |          |        |            |            |    |                     |             |   |   |
|                   |      |          |        |            |            |    |                     |             |   |   |
|                   |      |          |        |            |            |    |                     |             |   |   |
|                   |      |          |        |            |            |    |                     |             |   |   |
|                   |      |          |        |            |            |    |                     |             |   |   |
|                   |      |          |        |            |            |    |                     |             |   |   |
|                   |      |          |        |            |            |    |                     |             |   |   |
|                   |      |          |        |            |            |    |                     |             |   |   |
|                   |      |          |        |            |            |    |                     |             |   |   |

点击部署进入服务发布界面,选择模型部署设备的芯片型号,点击【发布】,跳转至服务列表页

| 选择模型<br>莫型版本 | 螺丝螺母识别 >                                                                                                   |
|--------------|------------------------------------------------------------------------------------------------------------|
| 莫型版本         |                                                                                                            |
|              | V2 ~                                                                                                       |
| 川练方式         | 零代码开发–AutoDL模式                                                                                             |
| 川练算法         | 服务器高性能                                                                                                     |
| 先择部署环境       | 服务器                                                                                                        |
| 选择系统和芯片      | <ul> <li>◆ 本地发布</li> <li>☆ Linux</li> <li>通用X86 CPU 英伟达GPU</li> <li>■ € Windows</li> <li>英伟达GPU</li> </ul> |

发布

在服务列表中选择需要导出的模型SDK,点击【导出SDK】,选择存储位置后即可完成模型SDK的导出

## Baidu 百度智能云文档

|                 | 部署   | く 〉 服务总览   |              |                 |                         |            |                  |        | 0 0 |
|-----------------|------|------------|--------------|-----------------|-------------------------|------------|------------------|--------|-----|
| EasyCL.         | 服务列表 | 飞桨EasyDL推出 | 出智能边缘控制台,助力本 | 地部署场景用户高效管理端侧预  | 顺测设备与服务,极速完             | 成本地数据与模型串联 | ,提高模型落地效率。了      | 解详情 ×  |     |
| 总变              |      | 服务器 通用     | 小型设备 专项适配硬   | 件 ⑦             |                         |            |                  | 输入模型名称 | Q   |
| C               |      | 模型名称       | 发布版本 🗘       | 应用平台            | 发布状态                    | 发布方式       | 发布时间 🗘           | 操作     |     |
| <>              |      | 螺丝螺母识别     | 5-V2         | 通用X86 CPU-Linux | <ul> <li>已发布</li> </ul> | 本地部署       | 2021-11-22 20:51 | 导出SDK  |     |
| 开发<br><b>()</b> |      | 猫狗分类       | 1-V1         | 英伟达GPU-Windows  | •已发布                    | 本地部署       | 2021-09-27 17:34 | 导出SDK  |     |
| 模型              |      |            |              |                 |                         |            |                  |        |     |
| 88 <b>8</b>     |      |            |              |                 |                         |            |                  |        |     |
| 订阅              |      |            |              |                 |                         |            |                  |        |     |
|                 |      |            |              |                 |                         |            |                  |        |     |
|                 |      |            |              |                 |                         |            |                  |        |     |
|                 |      |            |              |                 |                         |            |                  |        |     |
|                 |      |            |              |                 |                         |            |                  |        |     |
|                 |      |            |              |                 |                         |            |                  |        |     |
|                 |      |            |              |                 |                         |            |                  |        |     |
|                 |      |            |              |                 |                         |            |                  |        |     |

## 用零代码开发实现实例分割

### っ 示例说明

对比物体检测,实例分割支持用多边形标注训练数据,且模型可像素级识别目标。适合图中有多个主体、需识别其位置或轮廓的场景。本文以工件分割模型在macOS客户端中的使用为示例演示实例分割模型训练全过程。

#### 心 实现步骤

只需八步即可完成自定义AI模型的训练及发布的全过程。

#### Step1:提前准备训练数据

实例分割需要提供包含目标物体的图片并标注物体即可训练分割模型,自动识别图中所有目标物体的位置、轮廓、名称,下面我们来看看这次需 要分割的包含螺丝螺母的图片示例:

图片数量越多理论上训练效果越好,物体检测的图片数量建议每个类别不低于20张图片 注意图片需要为业务生产的真实环境所采集的图片, 与真实场景越贴近,训练模型效果越佳



Step2:创建数据集

在数据总览界面点击【创建数据集】

|   | •••             | 数据   | < > 我的数据总缆 | 0          | 0 |
|---|-----------------|------|------------|------------|---|
|   | EasyDL<br>AN    | 数据总览 | 创建数据集      | 输入数据集名称或ID | Q |
|   | 急览              | 数据标注 |            |            |   |
| l | <b>じ</b><br>数据  |      |            |            |   |
|   | <b>《》</b><br>开发 |      |            |            |   |
|   |                 |      | 还没有创建数据集   |            |   |
|   | 模型              |      | 创建数据集      |            |   |
|   | 部署              | •    |            |            |   |
|   | 86              |      |            |            |   |
|   | 1][4]           |      |            |            |   |
|   |                 |      |            |            |   |
|   |                 |      |            |            |   |
|   |                 |      |            |            |   |
|   |                 |      |            |            |   |
|   |                 |      |            |            |   |
|   |                 |      |            |            |   |

在数据集创建界面输入数据集名称、选择标注类型后点击【完成】

| く ) 创建数 | <b>文据集</b> |      |                                                     |
|---------|------------|------|-----------------------------------------------------|
| 数据集名称   | 工件分割       |      |                                                     |
| 数据类型    | 图片         |      |                                                     |
| 标注类型    | 图像分类       | 物体检测 | <b>实例分割</b><br>■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■ |
|         | 完成         |      |                                                     |

## Step3 : 导入数据

数据集创建完成后可在【数据总览】查看已创建完成的数据集,点击【导入】跳转至数据导入界面

| 工件分割 🖸 |     |      |          |      | 回 删除 |
|--------|-----|------|----------|------|------|
| 数据集ID  | 数据量 | 标注类型 | 标注状态     | 操作   |      |
| 24 💮   | 0   | 实例分割 | 0% (0/0) | 导入导出 |      |

数据导入支持无标注信息、有标注信息两种数据标注状态的数据以及多种导入方式,以下为无标注信息图片的导入为示例,其余各类型导入方式 可参考 导入图像数据选择数据标注状态与文件路劲

|   | 导入数据          |               |            |                   |
|---|---------------|---------------|------------|-------------------|
|   | 数据标注状态        | ● 无标注信息       | ○ 有标注信息    |                   |
|   | 导入方式          | <u>个</u> 选择目录 |            |                   |
| ( | 🐌 提示:1.导入后请避负 | 免改动本地该数据,     | 以免影响数据标注、楷 | 莫型训练功能正常使用        |
|   | 2.每次导入仅3      | 支持选择唯一目录,     | 如您想快速体验一站式 | 弌功能,可联网下载已标训练数据样例 |
|   | 图像分类训练数       | 数据集(coco格式)   |            |                   |
|   |               |               |            |                   |

| 3  | 导入   | 格式要求 |
|----|------|------|
| 0、 | 77/1 | 山山安小 |

#### 图片格式要求

目前支持图片类型为jpg, png, bmp, jpeg, 图 片大小限制在14M以内。

图片长宽比在3:1以内,其中最长边小于 4096px,最短边大于30px。

## 导入路径要求

无标注信息:导入请确保将全部图片保存至同 一层文件目录。

有标注信息:导入请确保将全部图片与对应标 注信息保存至同一层文件目录。该目录下子文 件目录及非相关内容(例如压缩包)不导入。

## 完成后,点击【确认并返回】跳转至数据总览页

| ┃ 导入数据         |               |                                  |
|----------------|---------------|----------------------------------|
| 数据标注状态         | ● 无标注信息       | ○ 有标注信息                          |
| 导入方式           | <u>个</u> 重新选择 | 上次导入路径:/Users/heyun02/Desktop/数据 |
| 🔒 提示: 1.导入后请避约 | 免改动本地该数据,     | 以免影响数据标注、模型训练功能正常使用              |
| 2.每次导入仅3       | 支持选择唯一目录,     | 如您想快速体验一站式功能,可联网下载已标训练数据样例       |
| 图像分类训练         | 数据集(coco格式)   |                                  |
| 确认并返回          |               |                                  |

#### Step4:标注数据

.

在数据总览页找到需要标注的数据集,点击【查看与标注】,跳转至标注页面

| 工件分割 🖸 |     |      |           |             | 回 删除 |
|--------|-----|------|-----------|-------------|------|
| 数据集ID  | 数据量 | 标注类型 | 标注状态      | 操作          |      |
| 24 💮   | 57  | 实例分割 | 0% (0/57) | 查看与标注 导入 导出 |      |

### 在左侧标签栏下,点击【添加标签】创建数据集标签

| •••            | 数据     | く 〉 【图片】工件分割/V1/查看与标注        |     |     |     | 0 0                                                   |
|----------------|--------|------------------------------|-----|-----|-----|-------------------------------------------------------|
| EasyDL<br>vid  | 数据总览   | 全部 (57) 有标注信息 (0) 无标注信息 (57) |     |     |     | 导入图片标注图片                                              |
| 急災             | 如3店村小土 | 筛选 🗸                         |     |     |     | 全选 🖮 删除 🗈 删除失效图片                                      |
| ● 数据 ● 数据 ● 新聞 |        | ■ ●                          | 元标签 | 无标签 | 无标签 | 无标签                                                   |
| 部署             |        |                              | 无标签 | 无标签 | 无标签 | <b>1</b><br>7<br>7<br>6<br>7<br>6<br>7<br>6<br>7<br>6 |
|                |        | 暂无可用标签 ,请点击上方按钮添加            | 无标签 | 无标签 | T标签 | <b>来</b><br>死标签                                       |

输入螺丝钉并点击【确认】添加数据标签

筛选 🗸



#### 点击图片右下角红框内图标进入到数据标注界面



在当前图片下选择右侧标签栏内的某一类别,代表为图片打上相应的标签,点击【保存当前图片】或直接点击下一张图标,在保存标注结果后自 动跳转至下一张。标注完所有图片后,该数据集便可用于后续训练任务



#### Step5:创建训练任务

在任务总览界面点击【创建任务】

| • | •••         | 开发                   | く〉任务                                | 总览                                 |                                |                         |                                              |                |                             |                          | 0 0           |
|---|-------------|----------------------|-------------------------------------|------------------------------------|--------------------------------|-------------------------|----------------------------------------------|----------------|-----------------------------|--------------------------|---------------|
|   | EssyDL<br>↔ | 零代码开发<br>Notebook 开发 | <b>零代码开发介</b><br>面向应用开发者<br>练完成的任务可 | <b>绍</b><br>指提供AutoDL训练<br>通过完整评估报 | 拣模式及高阶调参训练模式,<br>告、校验来验证效果,满足」 | AutoDL训练模式<br>业务应用需求的任: | 自动化程度更高,高级调参标<br>务可发布为模型进行后续部署               | 莫式支持设置<br>针操作。 | 训练参数灵活性更高,该两                | 种训练模式可灵活送                | 收起<br>轻择。 对于训 |
|   | じ数据         |                      | 4                                   | 1                                  | Ę                              |                         |                                              |                |                             |                          |               |
| L | $\diamond$  |                      | 1.发起i                               | 训练                                 | 2.选择数据                         | ŝ                       | 3.评估校验                                       |                | 4.发布为模型                     |                          |               |
|   | 开发<br>御刊    |                      | 选择训练方式并<br>数配置                      | 完成训练参                              | > 选择训练数据集, 自<br>动配置数据增强策略      | 1动或手 >                  | 任务训练完成后,通过评f<br>报告,校验验证任务效果                  | ± >            | 效果满足业务需求的任务<br>发布为模型,进行部署操( | ग<br>'F                  |               |
|   | 服务          | •                    | 创建任务                                |                                    |                                |                         |                                              |                |                             | 输入任务名称或                  | ID Q          |
|   |             |                      | 【图像分类】                              | 害虫识别 🖸 f                           | E务ID: 36                       |                         |                                              |                | 吕 训练                        | ⑤ 历史版本                   | □ 删除          |
|   | 11.00       |                      | 任务版本 i                              | 训练方式                               | 训练算法                           | 训练状态                    | 模型效果                                         | 对应模型           | 操                           | 作                        |               |
|   |             |                      | V1 i                                | 高级调参模式                             | MobileNet-MobileNetV2          | ● 训练完成                  | top1准确率:100.00%<br>top5准确率:100.00%<br>完整评估结果 | 害虫识别           | 查                           | 看版本配置 校验<br>hotebook打开 ⑦ |               |
|   |             |                      |                                     |                                    |                                |                         |                                              |                |                             |                          |               |

在训练任务创建界面输入任务名称、选择任务类型后点击【创建任务】



#### 任务创建完成后,点击【训练】,进入训练配置阶段

| 【实例分割】工件分割 🖸 任务ID:40                             | : | 训练 | 面 删除 |
|--------------------------------------------------|---|----|------|
| 任务创建成功,若无数据集请先在"数据中心"创建 ,上传训练数据 训练 后,可以在此处查看最新版本 |   |    |      |

#### 根据需求选择各项训练配置后,添加训练数据集,点击【开始训练】

|                                                                    | 开发    | く > 训练任务                                                        |                                                                   |                                                                              |                   |               | 0 0 |
|--------------------------------------------------------------------|-------|-----------------------------------------------------------------|-------------------------------------------------------------------|------------------------------------------------------------------------------|-------------------|---------------|-----|
| (1)<br>(1)<br>(1)<br>(1)<br>(1)<br>(1)<br>(1)<br>(1)<br>(1)<br>(1) | 零代码开发 | 训练方式 //<br>  训练配置<br>                                           | AutoDL模式 高结<br>Q导出源文件<br>服务器SDK                                   | 臨島後年99<br>(清参様式 ⑦<br>高島後年98<br>の 号出源文件与高統SDK ⑦<br>④ 満用小型设备SDK ②  参切送配硬件SDK ⑨ |                   |               |     |
| <b>开发</b><br>模型<br>部署                                              | ł     | 送择算法<br>高级训练配置<br>训练完成后同 ②<br>步发布为模型<br><b>》 添加防闭</b><br>添加训练数据 | <ul> <li>● 默认</li> <li>● 是</li> <li>● 否</li> <li>+ 请选择</li> </ul> |                                                                              |                   |               |     |
|                                                                    |       | 自定义验证集 ⑦<br>自定义测试集 ⑦<br>训练环境                                    | 数据集<br>工件分割<br>OFF<br>OFF<br>7解详情<br>名称<br>CFU                    | 版本<br>VI<br>规格<br>Intel(R) ContrTM ) 15-8257U CPU & 1.40GHz                  | 标 <b>公数量</b><br>1 | 路作<br>宣者详情 移移 |     |
|                                                                    |       | 开始训练                                                            |                                                                   |                                                                              |                   |               |     |

在任务总览页任务列表下,可以看到处于训练状态的训练任务,将鼠标放置感叹号图标处,即可查看训练进度

| 【实例分割】 | 】工件检测 🖸 f | 壬务ID: 15  |               |      |      | ① 历史版本    | 面 删除 |
|--------|-----------|-----------|---------------|------|------|-----------|------|
| 任务版本   | 训练方式      | 训练算法      | 训练状态          | 模型效果 | 对应模型 | 操作        |      |
| V1     | AutoDL模式  | 通用小型设备默。  | ▶ ● 训练中 💶     | -    |      | 查看版本配置 删除 | 终止训练 |
|        |           |           | 训练进度:         | 1%   |      |           |      |
| 【物体检测】 | 】螺丝螺母识别 [ | Z」任务ID:14 | 剩余时间: 1小时10分钟 |      |      | 吕 训练      | 面 删除 |

#### Step6:模型校验

## 训练完成后,可在任务列表下,点击【校验】

| 〈 〉 任务                                                                             | 5总览         |               |        |                                                |      |              |         | ?   | 0  |
|------------------------------------------------------------------------------------|-------------|---------------|--------|------------------------------------------------|------|--------------|---------|-----|----|
| <b>零代码开发介绍</b><br>面向应用开发者提供AutoDL训练模式及高阶调参训练模式,AutoDL训练模式自动化程度更高,高阶调参训练模式训练参数设置灵活。 |             |               |        |                                                |      |              |         |     |    |
| 创建任务                                                                               |             |               |        |                                                |      |              | 输入任务名称司 | ΣID | Q  |
| 【实例分割】                                                                             | 工件检测 🖸 任务IC | 0: 15         |        |                                                |      | 吕 训练         | ④ 历史版本  | 面册  | 削除 |
| 任务版本                                                                               | 训练方式        | 训练算法          | 训练状态   | 模型效果                                           | 对应模型 | 操作           |         |     |    |
| V1                                                                                 | AutoDL模式    | 仅导出源文件自行部署高性能 | ● 训练完成 | top1准确率: 100.00%<br>top5准确率: 100.00%<br>完整评估结果 |      | 查看版本配置<br>删除 | 1 发布为模型 | 校验  |    |

## 点击【添加图片】,进行模型校验



点击添加图片

Step7:发布为模型

### 确认模型效果满意后,可在任务列表下,点击【发布为模型】

| 〈 〉 任务                                                                                             | 扬总览         |                 |        |                                                |      |              |         | ?   | 0  |
|----------------------------------------------------------------------------------------------------|-------------|-----------------|--------|------------------------------------------------|------|--------------|---------|-----|----|
| 零代码开发介绍         展开           面向应用开发者提供AutoDL训练模式及高阶调参训练模式自动化程度更高,高阶调参训练模式训练参数设置灵活。            创稿任条 |             |                 |        |                                                |      | 展开           |         |     |    |
| 创建任务                                                                                               |             |                 |        |                                                |      |              | 输入任务名称司 | ΣID | Q  |
| 【实例分割】                                                                                             | 工件检测 🖸 任务IC | ): 15           |        |                                                |      | 吕 训练         | ⑤ 历史版本  | ₫ # | 删除 |
| 任务版本                                                                                               | 训练方式        | 训练算法            | 训练状态   | 模型效果                                           | 对应模型 | 操作           |         |     |    |
| V1                                                                                                 | AutoDL模式    | 仅导出源文件-自行部署-高性能 | ● 训练完成 | top1准确率: 100.00%<br>top5准确率: 100.00%<br>完整评估结果 |      | 查看版本配置<br>删除 | 置 发布为模型 | 校验  |    |

输入模型名称、版本描述后点击【确认】正式发布未模型

| 发布至模型 | 仓库      |
|-------|---------|
| *模型名称 | 工件检测    |
| 模型类型  | 实例分割    |
| 模型版本  | V1      |
| 版本描述  | 0 / 100 |
|       | 取 消 确 定 |

## Step8:导出模型文件或部署SDK

根据您在训练时选择的部署方式,可在模型总览版本列表中,点击【导出模型文件】或【部署】

|   |                       | 模型   |   | く > 版2   | 本列表   |          |            |    |                     |               | 0 | 0 |
|---|-----------------------|------|---|----------|-------|----------|------------|----|---------------------|---------------|---|---|
| ( | EonyOL<br>AP          | 模型列表 | ŧ | 莫型名称: 工作 | +检测   |          | 模型ID: 5    |    |                     |               |   |   |
|   |                       | 模型部署 | 相 | 莫型类型: 实例 | 利分割   |          | 模型来源:零代码开发 |    |                     |               |   |   |
|   | AR<br>A               |      |   | 版本       | 对应任务  | 训练方式     |            | 描述 | 导入时间 🍦              | 操作            |   |   |
|   | 数据                    |      |   | V2       | 10-V2 | AutoDL模式 |            |    | 2021-11-22 08:31:21 | 导出模型文件 ⑦ 部署 💬 |   |   |
|   | <b>&lt;&gt;</b><br>开发 |      |   | V1       | 9-V1  | AutoDL模式 |            |    | 2021-11-22 07:42:03 | 导出模型文件 ⊙      |   |   |
| I | Ŷ                     |      |   |          |       |          |            |    |                     |               |   |   |
|   |                       |      |   |          |       |          |            |    |                     |               |   |   |
|   | 878<br>00             |      |   |          |       |          |            |    |                     |               |   |   |
|   | <b>百日</b><br>订回       | (    |   |          |       |          |            |    |                     |               |   |   |
|   |                       |      |   |          |       |          |            |    |                     |               |   |   |
|   |                       |      |   |          |       |          |            |    |                     |               |   |   |
|   |                       |      |   |          |       |          |            |    |                     |               |   |   |
|   |                       |      |   |          |       |          |            |    |                     |               |   |   |
|   |                       |      |   |          |       |          |            |    |                     |               |   |   |
|   |                       |      |   |          |       |          |            |    |                     |               |   |   |
|   |                       |      |   |          |       |          |            |    |                     |               |   |   |
|   |                       |      |   |          |       |          |            |    |                     |               |   |   |

 $\times$ 

点击部署进入服务发布界面,选择模型部署设备的芯片型号,点击【发布】,跳转至服务列表页

## Baidu 百度智能云文档

| 选择模型    | 工件检测             | $\sim$ |
|---------|------------------|--------|
| 模型版本    | V2               | $\sim$ |
| 训练方式    | 零代码开发–AutoDL模式   |        |
| 训练算法    | 服务器高性能           |        |
| 选择部署环境  | 服务器              |        |
| 选择系统和芯片 | ● 本地发布           |        |
|         | 📃 🛆 Linux        |        |
|         | 通用X86 CPU 英伟达GPU |        |
|         | - Kindows        |        |
|         | 英伟达GPU           |        |
|         |                  |        |

#### 发布

## 在服务列表中选择需要导出的模型SDK,点击【导出SDK】,选择存储位置后即可完成模型SDK的导出

|                    | 部署   | く 〉 服务总览   |             |                 |                   |            |                  |        | 0 0 |
|--------------------|------|------------|-------------|-----------------|-------------------|------------|------------------|--------|-----|
| EasyOL             | 服务列表 | 飞桨EasyDL推出 | 智能边缘控制台,助力本 | 地部署场景用户高效管理端侧剂  | <b>测设备与服务,极速完</b> | 成本地数据与模型串联 | ,提高模型落地效率。了觸     | 详情 ×   |     |
| 总变                 |      | 服务器 通用小    | 型设备 专项适配硬   | 件 ②             |                   |            |                  | 输入模型名称 | Q   |
| C<br>BUR           |      | 模型名称       | 发布版本 📏      | 应用平台            | 发布状态              | 发布方式       | 发布时间 🖒           | 操作     |     |
| <b>«&gt;</b><br>π% |      | 工件检测       | 5-V2        | 通用X86 CPU-Linux | •已发布              | 本地部署       | 2021-11-22 20:51 | 导出SDK  |     |
| ŵ                  |      | 猫狗分类       | 1-V1        | 英伟达GPU-Windows  | •已发布              | 本地部署       | 2021-09-27 17:34 | 导出SDK  |     |
| 模型                 |      |            |             |                 |                   |            |                  |        |     |
| 部署                 |      |            |             |                 |                   |            |                  |        |     |
| 36<br>170          |      | •          |             |                 |                   |            |                  |        |     |
|                    |      |            |             |                 |                   |            |                  |        |     |
|                    |      |            |             |                 |                   |            |                  |        |     |
|                    |      |            |             |                 |                   |            |                  |        |     |
|                    |      |            |             |                 |                   |            |                  |        |     |
|                    |      |            |             |                 |                   |            |                  |        |     |
|                    |      |            |             |                 |                   |            |                  |        |     |
|                    |      |            |             |                 |                   |            |                  |        |     |
|                    |      |            |             |                 |                   |            |                  |        |     |
|                    |      |            |             |                 |                   |            |                  |        |     |

## 用零代码开发实现语义分割

っ 示例说明

对比实例分割,语义分割指将每个像素点归属为对象类的过程。适用于分割目标主体单一的场景,简单举例来说语义分割能够识别出图片中哪些像素是归属于"人"的标签,但无法区分"不同的人"。本文以在macOS客户端中的使用为示例演示实例分割模型训练全过程。

## 心 实现步骤

只需八步即可完成自定义AI模型的训练及发布的全过程。

### Step1:提前准备训练数据

语义分割需要提供包含目标物体的图片并标注物体即可训练分割模型,,下面我们来看看语义分割的城市建筑及道路物体分割图片示例:

我的数据总览 > 【图片】语义分割test/V3/查看与标注

| 全部 (845)   有标注信息 (<br> | 845) 无标注信息 (0) |                           |          |          | 导入图片质    | 金报告 标注图)   |
|------------------------|----------------|---------------------------|----------|----------|----------|------------|
| 筛选 🗸                   |                |                           |          |          |          | □本页全选 □ 删除 |
| 标签栏                    | 添加标签 🗸 🗸       |                           |          | j -      |          |            |
| 请输入标签名称                | Q              |                           |          |          |          |            |
| 标签名                    | 数据量↓           | building                  | building | building | building |            |
| window                 | 828            |                           |          |          | T- 🚿     |            |
| vegetation             | 630            |                           |          |          | - N. 500 |            |
| sky                    | 845            |                           |          |          |          |            |
| road                   | 703            | building 🗹 🖸              | building | building | building |            |
| pavement               | 614            | 点击【编辑icon】即<br>入"单张图片标注模式 | 可进<br>t" |          | 14666 f  |            |
| door                   | 412            |                           |          |          |          |            |

Step2:创建数据集

在数据总览界面点击【创建数据集】

| •••     | 数据   | < > 我的数据总览 | 3                    | 0 |
|---------|------|------------|----------------------|---|
| EasyDL. | 数据总览 | 创建数据集      | 输入数据集名称可D            | 0 |
|         | 数据标注 |            | ANY CREATE A LOOK OF |   |
| 总览      |      |            |                      |   |
| C       |      |            |                      |   |
| 数据      |      |            |                      |   |
| 开发      |      |            |                      |   |
| ÷.      |      | 还没有创建数据集   |                      |   |
| 模型      |      | 创建数据集      |                      |   |
| 部署      | 4    |            |                      |   |
| 31      |      |            |                      |   |
| 订间      |      |            |                      |   |
|         |      |            |                      |   |
|         |      |            |                      |   |
|         |      |            |                      |   |
|         |      |            |                      |   |
|         |      |            |                      |   |
|         |      |            |                      |   |
|         |      |            |                      |   |

在数据集创建界面输入数据集名称、选择标注类型后点击【完成】



数据集创建完成后可在【数据总览】查看已创建完成的数据集,点击【导入】跳转至数据导入界面

数据导入支持无标注信息、有标注信息两种数据标注状态的数据以及多种导入方式,以下为无标注信息图片的导入为示例,其余各类型导入方式 可参考 导入图像数据选择数据标注状态与文件路劲 上传图片时,请注意格式要求!

#### 3、导入格式要求

^

## 图片格式要求

目前支持图片类型为jpg, png, bmp, jpeg, 图 片大小限制在14M以内。

图片长宽比在3:1以内,其中最长边小于 4096px,最短边大于30px。

### 导入路径要求

无标注信息:导入请确保将全部图片保存至同 一层文件目录。

有标注信息:导入请确保将全部图片与对应标 注信息保存至同一层文件目录。该目录下子文 件目录及非相关内容(例如压缩包)不导入。

#### 完成后,点击【确认并返回】跳转至数据总览页

#### Step4:标注数据

在数据总览页找到需要标注的数据集,点击【查看与标注】,跳转至标注页面

#### 在左侧标签栏下,点击【添加标签】创建数据集标签

我的数据总览 > 【图片】语义分割test/V3/查看与标注

**全部 (845)** 有标注信息 (845) 无标注信息 (0)

#### 筛选 🗸



#### 输入并点击【确认】添加数据标签





点击图片右下角编辑图标即可进入到数据标注界面,在当前图片下选择右侧标签栏内的某一类别,代表为图片打上相应的标签,点击【保存当前 图片】或直接点击下一张图标,在保存标注结果后自动跳转至下一张。标注完所有图片后,该数据集便可用于后续训练任务

| 我的數                            | 数据总览 > 【图片】语义分割test/V3/查看与标注 > 标注                                                                                       | 语义分割test / V3 | ~      |              |
|--------------------------------|-------------------------------------------------------------------------------------------------------------------------|---------------|--------|--------------|
|                                | 全部(845) 无标注信息(0) 有标注信息(845) ② 自动识别轮廓工具示例 ▲ ◎ 辅助设置                                                                       | 标签栏           | 添加标签 🗸 |              |
|                                | 使用说明: 1.通过绘制矩形定义目标所在范围,左击目标中心位置即可识别目标轮廓: 2.若识别结果不精准,可通过添加正负点来调整。左击增加"漏识别区 域",右击可减去"误识别区域"; 3. 对识别提果满意之后,点击"确认轮廓"按钮并添加标签 | 请输入标签名称       | Q      |              |
| <b>F T</b>                     |                                                                                                                         | tree          | 快捷鍵 1  |              |
| <ul> <li></li> <li></li> </ul> |                                                                                                                         | window        | 快捷键 2  |              |
| 0                              |                                                                                                                         | vegetation    | 快捷鍵 3  |              |
| 0                              |                                                                                                                         | sky           | 快捷键 4  |              |
| 0                              |                                                                                                                         | road          | 快捷鍵 5  | Step5:创建训练任务 |
| $\diamond$                     |                                                                                                                         | pavement      | 快捷键 6  |              |
|                                | ding                                                                                                                    | door          | 快捷键 7  |              |
|                                |                                                                                                                         | car           | 快捷鍵 8  |              |
| Ð                              | window                                                                                                                  | building      | 快捷键 9  |              |
| Θ                              |                                                                                                                         |               |        |              |
|                                | taftit!                                                                                                                 |               |        |              |
| ⊻″                             |                                                                                                                         |               |        |              |
|                                |                                                                                                                         |               |        |              |

## 在任务总览界面点击【创建任务】



#### 在训练任务创建界面输入任务名称、选择任务类型后点击【创建任务】



| •••               | 开发                   | く 〉 任务总览                                           |                                           |                                                                |                                         | 0           | 2 0       |
|-------------------|----------------------|----------------------------------------------------|-------------------------------------------|----------------------------------------------------------------|-----------------------------------------|-------------|-----------|
| EasyDL<br>70      | 零代码开发<br>Notebook 开发 | 零代码开发介绍<br>面向应用开发者提供AutoDL训练模式<br>练完成的任务可通过完整评估报告、 | 。及高阶调参训练模式,AutoDL训练都<br>绞验来验证效果,满足业务应用需求的 | <sup>我</sup> 式自动化程度更高,高级调参模式支持设 <u>1</u><br>3任务可发布为模型进行后续部署操作。 | 置训练参数灵活性更高,该两种                          | m训练模式可灵活选择。 | 收起<br>对于训 |
| し数据               |                      |                                                    | Ð                                         | (J)                                                            |                                         |             |           |
| ↔<br>开发<br>模型     |                      | 1.发起训练<br>选择训练方式并完成训练参 ><br>数配置                    | 2.选择数据<br>选择训练数据集,自动或手动配置数据增强策略           | 3.评估校验 任务训练完成后,通过评估 》 报告,校验验证任务效果                              | 4.发布为模型<br>效果满足业务需求的任务可<br>发布为模型,进行部署操作 |             |           |
| 服务                |                      | 创建任务                                               |                                           |                                                                |                                         | 输入任务名称或ID   | Q         |
| <b>38</b><br>1701 |                      | 【语义分割】城市建筑及道路分割                                    | ☑ 任务ID:1                                  |                                                                |                                         | 吕 训练 茴      | 删除        |
|                   |                      |                                                    | 任务创建成功,若无数据集请先在"黄                         | 牧据总览"创建, 上传训练数据 训练 后, 同                                        | 可以在此处查看最新版本                             |             |           |

## 根据需求选择各项训练配置后,添加训练数据集,点击【开始训练】

|                                                                                   | 开发    | く 〉 训练任务                                                 |                                                                      |                                                                                         |      |         | 0 0 |
|-----------------------------------------------------------------------------------|-------|----------------------------------------------------------|----------------------------------------------------------------------|-----------------------------------------------------------------------------------------|------|---------|-----|
| Eosy                                                                              | 零代码开发 | ↓ 训练方式 A                                                 | utoDL模式 高级                                                           | <u>總版专事</u><br>调參模式 ⑦                                                                   |      |         |     |
| は<br>し<br>数<br>数<br>は<br>は<br>は<br>は<br>は<br>は<br>は<br>は<br>は<br>は<br>は<br>は<br>は |       | ■ 連結配置<br>号出失型<br>部署方式<br>送祥算法<br>高级训练配置<br>叫好字書用四 (***) | <ul> <li>(仅导出源文件</li> <li>服务器SDK</li> <li>默认</li> <li>OFF</li> </ul> | <ul> <li>【可能版专事】</li> <li>● 労出還文件与真体SDK ②</li> <li>③ 適用小型设备SDK ○ 专项适配硬件SDK ③</li> </ul> |      |         |     |
| BEN<br>STB                                                                        |       | 训练完成后问③<br>步发布为模型<br>添加徵据<br>添加训练数据                      | <ul> <li>是 ○ 否</li> <li>+ 请选择</li> <li>数据集</li> </ul>                | 版本                                                                                      | 标签数量 | 操作      |     |
|                                                                                   |       |                                                          | 工件分割                                                                 | VI                                                                                      | 1    | 查看详情 移除 |     |
|                                                                                   |       | 自定义验证集 ⑦<br>自定义测试集 ⑦<br>训练环境                             | OFF<br>OFF<br>7#¥¥M                                                  |                                                                                         |      |         |     |
|                                                                                   |       |                                                          | 名称                                                                   | 規格                                                                                      |      |         |     |
|                                                                                   |       |                                                          | UPU UPU                                                              | interiory Corection ID-62570 CPU & L40GHz                                               |      |         |     |
|                                                                                   |       | 开始训练                                                     |                                                                      |                                                                                         |      |         |     |

在任务总览页任务列表下,可以看到处于训练状态的训练任务,将鼠标放置感叹号图标处,即可查看训练进度

## Step6:模型校验

训练完成后,可在任务列表下,点击【校验】进入单图校验页面

| 选择模型  | xsy-语义分割-test1 | $\sim$ | 部署方式 | 公有云API | $\sim$ | 选择版本 | V3 | ~ |
|-------|----------------|--------|------|--------|--------|------|----|---|
| 启动模型校 | 验服务            |        |      |        |        |      |    |   |
|       |                |        |      |        |        |      |    |   |
|       |                |        |      |        |        |      |    |   |

点击【添加图片】上传本地图片即可完成模型校验

Step7:发布为模型

确认模型效果满意后,可在任务列表下,点击【发布为模型】

输入模型名称、版本描述后点击【确认】正式发布为模型

# 发布至模型仓库

| *模型名称 | 城市道路及建筑分割 |
|-------|-----------|
| 模型类型  | 实例分割      |
| 模型版本  | V1        |
| 版本描述  |           |
|       | 0 / 100   |
|       |           |
|       | 取 消 确 定   |

Step8:导出模型文件或部署SDK

根据您在训练时选择的部署方式,可在模型总览版本列表中,点击【导出模型文件】或【部署】 点击部署进入服务发布界面,选择模型部署设备的芯片型号,点击【发布】,跳转至服务列表页 在服务列表中选择需要导出的模型SDK,点击【导出SDK】,选择存储位置后即可完成模型SDK的导出

|            |            | 部署   | く 〉 服务总览   |             |                 |                         |             |                  |        | 0 0 |
|------------|------------|------|------------|-------------|-----------------|-------------------------|-------------|------------------|--------|-----|
| 1          | EasyCL     | 服务列表 | 飞桨EasyDL推出 | 習能边缘控制台,助力; | 本地部署场景用户高效管理端侧  | 预测设备与服务,极速              | 完成本地数据与模型串I | 、提高模型落地效率。了      | ¥详情 ×  |     |
|            | 总范         |      | 服务器 通用     | 小型设备 专项适配领  | 更件 ⑦            |                         |             |                  | 输入模型名称 | Q   |
|            |            |      | 模型名称       | 发布版本 🔷      | 应用平台            | 发布状态                    | 发布方式        | 发布时间 关           | 操作     |     |
|            | <>>        |      | 工件检测       | 5-V2        | 通用X86 CPU-Linux | <ul> <li>已发布</li> </ul> | 本地部署        | 2021-11-22 20:51 | 导出SDK  |     |
|            | Ŵ          |      | 猫狗分类       | 1-V1        | 英伟达GPU-Windows  | •已发布                    | 本地部署        | 2021-09-27 17:34 | 导出SDK  |     |
|            | 模型         |      |            |             |                 |                         |             |                  |        |     |
|            | 857        |      |            |             |                 |                         |             |                  |        |     |
|            | <b>订</b> 间 |      | •          |             |                 |                         |             |                  |        |     |
|            |            |      |            |             |                 |                         |             |                  |        |     |
|            |            |      |            |             |                 |                         |             |                  |        |     |
|            |            |      |            |             |                 |                         |             |                  |        |     |
|            |            |      |            |             |                 |                         |             |                  |        |     |
|            |            |      |            |             |                 |                         |             |                  |        |     |
|            |            |      |            |             |                 |                         |             |                  |        |     |
|            |            |      |            |             |                 |                         |             |                  |        |     |
|            |            |      |            |             |                 |                         |             |                  |        |     |
| <b>米</b> 女 | <b>汝</b> 打 | 居管理  |            |             |                 |                         |             |                  |        |     |
| 娄          | <b></b>    | 导入   |            |             |                 |                         |             |                  |        |     |

#### の 数据准备

设计分类

对于图像分类任务,需确认分类如何设计,每个分类为你希望识别出的一种结果,如您需要识别动物,则可以以"dog"、"cat"等分别作为一个分 类。

注意:每张图片都应属于一个分类,一个模型最多支持1000个分类,标签名由数字、中英文、中/下划线组成,长度上限256字符。

基于设计好的分类准备图片,有如下要求:

1. 每个分类需要准备20张以上图片,如果需要较好的效果,建议每个分类准备不少于100张图片

2. 训练图片和实际场景要识别的图片拍摄环境一致,举例:如果实际要识别的图片是摄像头俯拍的,训练时也需要使用俯拍角度的图片

3. 每个标签的图片需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强

设计标签

对于物体检测、实例分割任务,需要确认标签如何设计,每种需要识别的目标为一个标签,一张图片中可以有多种标签出现。

注意:单个数据集的标签上限为1000种,标签名由数字、中英文、中/下划线组成,长度上限256字符。

#### 基于设计好的物体检测准备图片,有如下要求:

- 训练图片和实际场景要识别的图片拍摄环境一致,举例:如果实际要识别的图片是摄像头俯拍的,训练图片就不能用网上下载的目标正面图 片
- 2. 每个标签的图片需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强
- 3. 每个模型训练图片量不得低于4张,每个标签建议标注50个框以上
- 心 数据导入

在数据总览页下找到创建完成的数据集点击【导入】

| 猫狗识别 🖸 |       |     |      |          |       | 回 删除 |
|--------|-------|-----|------|----------|-------|------|
| 版本     | 数据集ID | 数据量 | 标注类型 | 标注状态     | 操作    |      |
| V1 🖂   | 7     | 0   | 图像分类 | 0% (0/0) | 导入 导出 |      |

#### 导入要求

#### 图片要求:

- 支持图片类型包括jpg、png、bmp、jpeg,图片大小限制在14M以内
- 图片长宽比在3:1以内,其中最长边小于4096px,最短边大于30px

### 路径要求:

- 无标注信息:导入请确保全部图片保存知同一层文件目录下
- 有标注信息:导入请确保将全部图片与对应标注信息保存至同一层文件目录。该目录下子文件目录及非相关内容(例如压缩包)不导入

导入方式

数据导入无标注信息图片以及有标注信息图片的导入

注:数据导入后依然存储在您设备本地导入路径下,飞桨EasyDL桌面版不会调整您的数据存储路径,因此如更改本地存储路径下的图片将会导致数据集发生变动,如有正使用当前数据集训练的任务,将会导致任务失败.

#### 无标注信息图片导入

进入数据导入界面,在数据标注状态中选择无标注信息

选择训练数据存储的文件夹,选择完成后数据集即导入完成

如需导入多个文件目录的数据,可多次导入

| I | 创建信息                 |         |               |           |                   |
|---|----------------------|---------|---------------|-----------|-------------------|
|   | 数据集ID                | 7       |               | 版本号       | V1                |
|   | 备注                   |         |               |           |                   |
| I | 标注信息                 |         |               |           |                   |
|   | 标注类型                 | 图像分     | 类             | 标注模板      | 单图单标签             |
|   | 数据总量                 | 0       |               | 已标注       | 0                 |
|   | 标签个数                 | 0       |               | 目标数       | 0                 |
| 1 | <b>导入数据</b><br>数据标注状 | 态       | ● 无标注信息       | 🦳 有标注信息   |                   |
|   | 导入方式                 |         | <u>↑</u> 选择目录 | ]         |                   |
|   | 🔒 提示: 1              | 1.导入后请避 | 免改动本地该数据,     | 以免影响数据标注、 | 模型训练功能正常使用        |
|   | 2                    | 2.每次导入仅 | 支持选择唯一目录,     | 如您想快速体验一站 | 式功能,可联网下载已标训练数据样例 |
|   | 1                    | 图像分类训练  | 数据集(coco格式)   |           |                   |
|   | 确认并返                 | 反回      |               |           |                   |

## 有标注信息图片导入

有标注信息导入支持以文件夹命名分类、VOC格式、COCO格式以及平台自定义格式四种

## 以文件夹命名分类导入

数据标注状态选择有标注信息,并选择标注格式为以文件夹命名分类

| 注:以文件夹命名分类仅支持图像分类任务 |           |                |          |           |  |  |
|---------------------|-----------|----------------|----------|-----------|--|--|
| 数据标注状态              | ○ 无标注信息   | ● 有标注信息        |          |           |  |  |
| 导入方式                | ↑ 选择目录    |                |          |           |  |  |
| 标注格式                | 🔵 以文件夹命名分 | )类 ⑦ 🔹 🔿 文oc ⑦ | 🔾 coco 🕐 | ○ 平台自定义 ⑦ |  |  |

以文件夹命名分类方式导入,导入路径下的每一个子文件夹将代表一个分类,子文件夹的名称将代表分类名,子文件夹下的图片将被视为当前分 类下的数据



VOC格式导入

| - 守八奴惦 |  | 导入数据 |
|--------|--|------|
|--------|--|------|

| 数据标注状态 | ○ 无标注信息       | ▶ 有标注信息  |          |           |
|--------|---------------|----------|----------|-----------|
| 导入方式   | <u>个</u> 选择目录 |          |          |           |
| 标注格式   | ○ 以文件夹命名分类    | 🤊  voc 🕐 | 🔿 coco 🕐 | 🔵 平台自定义 ⑦ |

以VOC格式导入,导入路径下应包含JPEGImages以及Annotations两个子文件夹,JPEGImages下存储图片数据,Annotations下存储xml格式的标 注文件,且图片与标注信息一一对应



#### COCO格式导入

| 数据标注状态选择有标注信息,并选择标注格式为COCO格式 |           |                |          |           |  |  |  |
|------------------------------|-----------|----------------|----------|-----------|--|--|--|
| 数据标注状态                       | ○ 无标注信息   | ◯ 有标注信息        |          |           |  |  |  |
| 导入方式                         | ⚠ 选择目录    |                |          |           |  |  |  |
| 标注格式                         | ○ 以文件夹命名分 | )类 ⑦ 🛛 🔿 voc ⑦ | 🔿 coco 🕐 | 🔵 平台自定义 ⑦ |  |  |  |

以COCO格式导入,导入路径下应包含JPEGImages以及Annotations两个子文件夹,JPEGImages下存储图片数据,Annotations下存储Json格式的 一个标注文件,所有图片的标注信息均存储在一个Json文件中

| 1.导入路径内需包括JPEGImages,Annotations两个文件夹,分别包括图片源文件<br>及唯一一个命名为coco_info.json的标注文件下载标注格式样例<br>2.标注文件中标签由数字、中英文、中/下划线组成,长度上限256字符<br>3.图片源文件大小限制在14M内,长宽比在3:1以内,其中最长边需要小于4096pX, | (jpg/png/bmp/jpeg)<br>,最短边需要大于30px |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|
|                                                                                                                                                                             | 1                                  |
| ●<br>与入路径<br>●<br>●<br>●<br>●<br>●<br>●<br>●<br>●<br>●<br>●<br>●<br>●<br>●                                                                                                  | 3                                  |
| Annotations                                                                                                                                                                 | NOR                                |

平台自定义格式导入

数据标注状态选择有标注信息,并选择标注格式为平台自定义格式

■ 导入数据

| 数据标注状态 | ○ 无标注信息       | ○ 有标注信息    |        |           |  |
|--------|---------------|------------|--------|-----------|--|
| 导入方式   | <u>↑</u> 选择目录 |            |        |           |  |
| 标注格式   | ○ 以文件夹命名分     | 类⑦ 🔷 voc ⑦ | 🔾 coco | ◯ 平台自定义 ⑦ |  |

以平台自定义格式导入,导入路径不包含子文件夹,图片数据及标注文件均直接存储在导入路径下,标注信息以Json格式与图片一一对应



导入路径选择完成后,点击【确认并返回】即完成数据导入

## 查看数据集

数据标注完成后可在数据总览页查看数据情况

### 心 查看数据集

数据总览页展示数据集名称、数据集ID、数据量、标注类型、标注状态,鼠标放置在省略号处可查看数据集详细信息

| 1前 删除     |
|-----------|
| LTT MONAY |
|           |
|           |
|           |
|           |
| 回 删除      |
|           |
|           |
|           |

点击【查看与标注】可查看数据集详情

支持查看数据集中数据标注情况,可手动删除数据集中的图片

支持一键删除失效图片

| <b>全部 (69)</b> 有标注信息 (69) | 无标注信息 (0) |    |    |    | 导入图片标注图片        |
|---------------------------|-----------|----|----|----|-----------------|
| 筛选 🗸                      |           |    |    |    | 全选 💿 删除 💿 删除失效图 |
| 标签栏                       | 添加标签      |    |    |    |                 |
| 请输入标签名称                   | Q         |    |    | U  |                 |
| 标签名                       | 标注框数J↑    | 螺丝 | 螺丝 | 螺丝 | 螺丝              |
| 螺丝                        | 232       |    |    |    |                 |
| 螺母                        | 257       |    |    |    |                 |
|                           |           | 螺丝 | 螺丝 | 螺母 | 螺丝              |
|                           |           |    |    |    |                 |
|                           |           | 螺丝 | 螺丝 | 螺丝 | 螺丝              |

点击图片右下角查看大图,可查看放大图片

## 筛选 🗸

| 标签栏      | 添加标签   |    |       |        |    |
|----------|--------|----|-------|--------|----|
| 请输入标签名称  | Q      |    |       |        | A  |
| 标签名      | 标注框数♪♪ | 螺丝 | FK 58 | 螺丝     | 螺丝 |
| 螺丝       | 232    |    | 查看大師  | 图<br>图 |    |
| <b>I</b> |        |    |       |        |    |

放大图片后可查看图片本地地址,点击可打开当前文件存储路径



### 心 数据集导出

支持将数据集导出到指定路径下,点击【导出】将导出数据源文件以及当前数据集已有的标注文件,支持将标注数据导出为voc、coco或平台自 定义格式
创建导出任务
 ×

 导出数据集名称 物体检测数据集V1

 导出数据 导出全部数据,包含源文件及已有的标注文件

 标注格式
 VOC ②
 COCO ②
 平台自定义 ③

 导出路径
 □选择目录
 上次导出路径: /Users/zhangjingsheng/Deskto...

## 创建数据集

## 数据模块主要用于训练数据的导入、标注、导出以及管理

| 000             | 数据                       | く > 我的 | 内数据总览 |      |              | ⑦ ②          |
|-----------------|--------------------------|--------|-------|------|--------------|--------------|
| EasyDL          | 数据总览<br>数据标注             | 创建数据集  |       |      |              | 输入数据集名称或ID Q |
| 总览              | 急党                       | 物体检测2  |       |      |              | 靣 删除         |
| 0               |                          | 数据集ID  | 数据量   | 标注类型 | 标注状态         | 操作           |
| 数据<br><b>《》</b> |                          | 10 💮   | 0     | 物体检测 | 0% (0/0)     | 导入 导出        |
| 开发              | 开文<br>便型<br>正示者<br>CT/FR | 实例分割2  |       |      |              | 意 豊富         |
| 模型              |                          | 数据集ID  | 数据量   | 标注类型 | 标注状态         | 操作           |
| <b>部署</b>       |                          | 9 💬    | 0     | 实例分割 | 0% (0/0)     | 导入 导出        |
| 订阅              |                          | 图像分类2  |       |      |              | 回 删除         |
|                 |                          | 数据集ID  | 数据量   | 标注类型 | 标注状态         | 操作           |
|                 |                          | 8 😔    | 96    | 图像分类 | 100% (96/96) | 查看与标注 导入 导出  |
|                 |                          | 图像分类 🖸 | Z     |      |              | 回 删除         |
|                 |                          | 数据集ID  | 数据量   | 标注类型 | 标注状态         | 操作           |
|                 |                          | 4 💬    | 0     | 图像分类 | 0% (0/0)     | 导入 导出        |

### 发起训练任务前,需要提前准备训练数据,在数据总览页点击创建数据集

| 数据   | く > 我的数据总览 |
|------|------------|
| 数据总览 |            |
| 数据标注 | 创建数据集      |
|      |            |

输入数据集名称,选择数据集标注类型后点击完成即可完成数据集创建,飞桨EasyDL桌面版提供图像分类、物体检测、实例分割三种标注类型可 供选择

| 数据           | く > 我的数据总览                                                                          |  |
|--------------|-------------------------------------------------------------------------------------|--|
| 数据总览<br>数据标注 | * 数据集名称 限制50个汉字内 数据类型 图片                                                            |  |
|              | 标注类型     图像分类     物体检测     实例分割       ・     ・     ・     ・       ・     ・     ・     ・ |  |
|              | 完成<br>EN                                                                            |  |

### 数据集创建完成后可在数据总览页查看创建完成的数据集

| 数据总览<br>数据标注 | 创建数据集   |     |      |          |       | 输入数据集名称或ID Q |
|--------------|---------|-----|------|----------|-------|--------------|
|              | 图像分类数据集 |     |      |          |       | 直 删除         |
|              | 数据集ID   | 数据量 | 标注类型 | 标注状态     | 操作    |              |
|              | 11 💬    | 0   | 图像分类 | 0% (0/0) | 导入一导出 |              |

## 数据标注

飞桨EasyDL桌面版提供了图像分类、物体检测、实例分割的数据标注能力,可在本地高效完成数据标注,数据标注启动前,需先创建数据集中的 标签

### 心 设定标签

## 在数据总览页找到导入完成的数据集,并点击【查看与标注】

| 图像分类标注演示 [2] |     |      |           |             |  |
|--------------|-----|------|-----------|-------------|--|
| 数据集ID        | 数据量 | 标注类型 | 标注状态      | 操作          |  |
| 12           | 96  | 图像分类 | 0% (0/96) | 查看与标注 导入 导出 |  |

### 在查看与标注页面点击【添加标签】创建本次数据中的数据标签



## 筛选 🗸

| 标签栏         | 添加标签 |
|-------------|------|
| 请输入标签名称     | Q    |
| 根据图片内容,选择标签 |      |

在左侧列表栏可看到已添加的标签,并支持搜索标签名称、按照数据量多少对标签进行排序

#### 筛选 🗸

| 标签栏         | 添加标签  |
|-------------|-------|
| 请输入标签名称     | Q     |
| 根据图片内容,选择标签 |       |
| 标签名         | 数据量11 |
| cat         | 0     |
| dog         | 0     |

### の 数据标注

## 图像分类

图像分类场景提供常规标注与批量标注两种标注方式

### 常规标注

标签添加完成后,在【查看与标注】界面,点击图片右下角标注icon进入标注界面



### 筛选 🗸



或在左侧导航栏点击数据标注,选择需要标注的数据集进入到标注界面

| 数据   | く > 数据标注     |
|------|--------------|
| 数据总览 |              |
| 数据标注 |              |
|      |              |
|      |              |
|      |              |
|      |              |
|      |              |
|      |              |
|      | ( 请选择数据集进行标注 |
|      | 请选择数据集へ      |
|      | 图像分类标注演示     |
|      | 图像分类数据集      |
|      |              |

进入标注界面后,可在上方切换查看全部数据、无标注信息数据或有标注信息数据



左侧工具栏提供图片放大、缩小、移动或删除功能



根据图片内容在左侧正确选择图片标签并点击【保存当前标注】完成数据标注



### 批量标注

Ð

Ī

除常规标注外,飞桨EasyDL桌面版提供了批量标注工具,可提高数据标注效率

在数据总览页找到已导入数据的数据集,点击【查看与标注】

| 猫狗识别 [  | 乙 数据集组ID:6  |          |      |           |            | 直 删除 |
|---------|-------------|----------|------|-----------|------------|------|
| 版本      | 数据集ID       | 数据量      | 标注类型 | 标注状态      | 操作         |      |
| V1 💮    | 6           | 96       | 图像分类 | 0% (0/96) | 查看与标注 导入 等 | 导出   |
| 点击【批    | 量标注】进入批量    | 量标注界面    |      |           |            |      |
| 全部 (96) | 有标注信息(0) 无标 | 注信息 (96) |      |           |            |      |

| •        |       |          |            |           | 2 全选 回 删除  |
|----------|-------|----------|------------|-----------|------------|
| 标签栏      | 添加标签  |          |            | (Some     |            |
| 请输入标签名称  | Q     | -        | 100        | 1.34      |            |
| 根据图片内容,递 | 1择标签  | 无标签      | 无标签        | 无标签       | 无标签        |
| 标签名      | 11量訊燈 | NO IN LL | a transis. | 2 D FORME | 2 G 127 AL |
| dog      | 0     |          |            |           |            |
| cat      | 0     |          |            |           | 1998       |
|          |       |          | A 410 200  |           |            |
|          |       | 无标签      | 无标签        | 无标签       | 无标签        |
|          |       |          | *          |           |            |
|          |       | 无标签      | 无标签        | 无标签       | 无标签        |
|          |       |          |            |           |            |

在批量标注界面支持多选或全选本页图片,图片勾选完成后选择相应标签即完成勾选图片的标注

| 全部 (96) 有标注信息 (0) 无标注信息 (96) |                       | 导入图片 关闭批量标注          | 标签栏     | 添加标签  |
|------------------------------|-----------------------|----------------------|---------|-------|
| 筛选 🗸                         | 已选择8个 取消选择            | 择 📗 全选 💿 删除 💿 删除失效图片 | 请输入标签名称 | Q     |
|                              |                       |                      | 根据图片内容, | 选择标签  |
| and a second                 | A State of the second |                      | 标签名     | 数据量11 |
|                              |                       |                      | dog     | 0     |
| 无标签 无标签                      | 无标签 无标签               | 无标签                  | cat     | 0     |
|                              |                       |                      |         |       |
| 无标签 无标签                      | 无标签 无标签               | 无标签                  |         |       |
|                              |                       |                      |         |       |
|                              | 7019132 7019132       | AP MATT              |         |       |

## 物体检测

物体检测数据标注方式为通过矩形框框出图片中需要检测的目标,并对标注框选择标签名,以下为物体检测数据标注流程

在【查看与标注】界面,点击图片右下角标注icon进入标注界面

| <mark>部 (69)</mark> 有标注信息 (0) | 无标注信息 (69) |             |          |          | 导入图片标注        | 图片   |
|-------------------------------|------------|-------------|----------|----------|---------------|------|
| 选 🗸                           |            |             |          |          | 全选 💿 删除 🗴 删除失 | €效图片 |
| 标签栏                           | 添加标签       |             |          |          |               |      |
| 请输入标签名称                       | Q          | 311 S       | (5)      | Califo a | 100           |      |
| 标签名                           | 标注框数J℃     | 无标签 🖸 🄀     | 无标签      | 无标签      | 无标签           |      |
| 螺母                            | 0          | 点击【编辑icon】即 | 可进<br>t" |          |               |      |
| 螟丝                            | 0          |             |          | ×        |               |      |
|                               |            | 无标签         | 无标签      | 无标签      | 无标签           |      |
|                               |            |             | 000      | A.       |               |      |
|                               |            | 无标签         | 无标签      | 无标签      | 无标签           |      |

## 或在左侧导航栏点击数据标注,选择需要标注的数据集进入到标注界面



在标注界面将需要检测的目标框选出,并为目标选择对应的标签

|   |   |          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 保存当前标注(S)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|---|---|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|   |   |          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|   |   | 6        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|   | 0 | Pollin . | Y                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| < | ( | 04       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | >                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|   |   | 请选择标签    | ×                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|   |   | 螺母螺丝     | 1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|   |   |          | はのでのです。     はのでのです。     はのでのです。     はのでのです。     はのでのです。     はのでのです。     はのでのです。     はのでのです。     はのでのです。     はのでのです。     はのでのです。     はののです。     はののです。 | ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・              ・               ・              ・              ・              ・              ・              ・              ・              ・              ・ |

标注过程中应尽可能贴合目标物体,避免未完全框选目标物体或矩形框远大于目标物体,可通过拉动标注框边界调整标注框大小





单张图片标注完成后点击【保存当前标注】即完成单张图片标注



## 实例分割

为实例分割场景数据标注过程负责,飞桨EasyDL-桌面版提供了多种数据标注工具,可根据实际使用需求选择使用

#### 自动识别轮廓标注工具

标注界面工具栏第一个工具为自动识别轮廓工具,开启自动识别工具后,鼠标左键点击目标物体后可自动识别目标物体轮廓

く > 标注



对于误识别区域,可通过右键点击来进行修正

く > 标注 ? 0 全部(69) 无标注信息(69) 有标注信息(0) 自动识别轮廓工具示例 添加标签 标签栏 使用说明: 1.鼠标左击自动识别目标轮廓,鼠标右击可减去"误识别区域"; 2.对于复杂轮廓可通 请输入标签名称 过左右键交替点击来修正识别结果;3. 确定轮廓, 添加标签 Ø 确认轮廓 (Enter) 阈值调整 ⑦ 🗕 -0 螺母 快捷键 1 0 螺丝 快捷键 2 1  $\Diamond$  $\Diamond$ 5 Ů 27 杨注示例 🞯 🞯 🗐 🗐 🐨 🐻 🐨 🏂 🎝 加文件 >

轮廓标注正确后点击【确认轮廓】并为目标添加标签



### 多边形标注

标注界面工具栏第二个工具为多边形标注工具,多边形标注工具是通过多点首尾相连的方式将目标物体圈出的方式完成标注,适用于目标物体结构相对负责的情况

标注完成后可拖动任意一点调整标注结果,标注结果符合要求后为当前标注物选择标签



## 圆形标注工具

标注界面工具栏第三个工具为圆形标注工具,圆形标注工具是通过拉取一个圆将目标物体的方式完成标注,适用于目标物体形状是圆形或近似圆形的情况

标注完成后可拖动调整圆形位置或拖动圆形边缘调整圆形大小,标注结果符合要求后为当前标注物选择标签



## 直线标注工具

标注界面工具栏第四个工具为直线标注工具,直线工具是通过选择两点连成一条直线的方式完成标注,适用于目标物体形状是直线的情况

|            | 全部(69) | 无标注信息(69) | 有标注信息(0) |       |                                               |   |
|------------|--------|-----------|----------|-------|-----------------------------------------------|---|
|            |        |           |          |       | 保存当前标注(S                                      | ) |
| 1          |        |           |          |       |                                               |   |
| $\bigcirc$ |        |           |          |       |                                               |   |
| 0          |        |           |          | 题目示和  |                                               |   |
| S          |        | 相關和這      |          |       |                                               |   |
|            |        |           |          |       |                                               |   |
| $\Diamond$ |        |           |          | 请洗择标签 | $\times$                                      |   |
| 5          |        |           |          | 國田    |                                               | > |
|            |        | 0         |          | 螺丝    | 2                                             |   |
|            |        |           | C        | 删除标注框 |                                               |   |
| $\oplus$   |        |           | S        |       |                                               |   |
| Θ          |        |           | A stream | A     |                                               |   |
| 前          |        |           |          | AREA  | <b>林田田田田田田田</b> 田田田田田田田田田田田田田田田田田田田田田田田田田田田田田 |   |
| ⊌7         |        |           |          |       |                                               |   |
|            | 〈标注示例  |           | 8 0 0    |       |                                               | > |

## 画笔标注工具

标注界面工具栏第五个工具为画笔标注工具,画笔标注工具是通过画笔覆盖一定区域并进行标注的方式,可在标注界面调整画笔大小



1308

如标注区域大于目标物体,可通过橡皮擦将多余区域擦除



### 其他辅助工具

除标注方式外,飞桨EasyDL桌面版仍提供了其他辅助工具,如:操作撤回、操作恢复、移动图片、放大图片、缩小图片、删除以及图片最大化



## 语义分割

为语义分割场景数据标注过程负责,飞桨EasyDL-桌面版提供了多种数据标注工具,可根据实际使用需求选择使用。 自动识别轮廓标注工具

标注界面工具栏第一个工具为自动识别轮廓工具,开启自动识别工具后,鼠标左键点击目标物体后可自动识别目标物体轮廓,对于误识别区域,可 通过右键点击来进行修正,轮廓标注正确后点击【确认轮廓】并为目标添加标签 我的数据总览 > 【图片】语义分割test/V3/查看与标注 > 标注



### 多边形标注

标注界面工具栏第二个工具为多边形标注工具,多边形标注工具是通过多点首尾相连的方式将目标物体圈出的方式完成标注,适用于目标物体结构相对负责的情况

标注完成后可拖动任意一点调整标注结果,标注结果符合要求后为当前标注物选择标签

#### 圆形标注工具

标注界面工具栏第三个工具为圆形标注工具,圆形标注工具是通过拉取一个圆将目标物体的方式完成标注,适用于目标物体形状是圆形或近似圆形的情况

标注完成后可拖动调整圆形位置或拖动圆形边缘调整圆形大小,标注结果符合要求后为当前标注物选择标签

### 直线标注工具

标注界面工具栏第四个工具为直线标注工具,直线工具是通过选择两点连成一条直线的方式完成标注,适用于目标物体形状是直线的情况

### 画笔标注工具

标注界面工具栏第五个工具为画笔标注工具,画笔标注工具是通过画笔覆盖一定区域并进行标注的方式,可在标注界面调整画笔大小

如标注区域大于目标物体,可通过橡皮擦将多余区域擦除

### 其他辅助工具

除标注方式外,飞桨EasyDL桌面版仍提供了其他辅助工具,如:操作撤回、操作恢复、移动图片、放大图片、缩小图片、删除以及图片最大化



# 开发与训练

零代码开发

### の AutoDL模式开发

为零AI开发基础的用户提供的建模方式,内置基于百度文心大模型的成熟预训练模型,可针对用户数据进行算法自动优化,助用户使用少量数据 也能获得具备出色效果与性能的模型。

### 训练配置

AutoDL模式下支持对模型文件导出类型、模型SDK部署方式、训练使用算法类型等内容进行设置。

| 训练配置             |          | 高级版专享                 |                   |
|------------------|----------|-----------------------|-------------------|
| 导出类型             |          | ● 仅导出源文件              | K ⊘               |
| 选择算法 ⑦           | D        | ▶ 超高性能 ⑦ 🔹 高性能 ⑦ 🔷 ह  | 5精度 ⑦ 查看算法性能及适配硬件 |
| 高级训练配置           | 50       |                       |                   |
|                  |          | 以下高级配置选项一般情况推荐不做更改,如实 | 际任务场景有需求,请根据实际调整  |
|                  |          | 输入图片分辨率 ⑦ Auto        | ~                 |
|                  |          | epoch ⑦               | <b>殳</b> 置        |
| 训练完成后际<br>步发布为模型 | 司<br>型 ⑦ | 是 〇 否                 |                   |
|                  |          | *模型名称 害虫识别            |                   |
|                  |          | 模型类型 图像分类             |                   |
|                  |          | 模型版本 V2               |                   |
|                  |          | 版本描述1                 |                   |
|                  |          |                       | 0 / 100           |

## 导出类型

- 导出模型源文件:训练完成后支持将模型源文件导出,模型源文件可通过Paddle-Inference转化至实际应用场景中使用
- **导出模型源文件与离线SDK:**训练完成的模型可发布为在服务器、小型设备、专项适配硬件上直接部署的SDK,覆盖主流芯片与操作系统, 可直接用于业务集成,省去繁琐转化过程

部署方式

- 可发布为在服务器、小型设备、专项适配硬件上直接部署的SDK
- 模型SDK适配设备类型详见 AotuDL模式算法适配硬件 选择算法
- 高精度: 相同训练数据情况下,训练出的模型准确率更高,训练及预测耗时更长,模型体积更大
- 高性能: 相同训练数据情况下,较高精度算法,训练及预测耗时更短,模型准确率平均比高精度算法低3%~5%
- 超高性能: 相对高性能算法,模型体积更小,CPU环境预测速度提升近60%,GPU环境预测速度提升近10%

模型精度与训练数据量的大小、质量成正相关 训练速度与训练数据量的大小以及训练设备的算力情况相关

#### 高级训练配置

高级训练配置开启状态下支持手动设置图像分辨率及迭代轮次 (epoch)

- 图像分辨率:如检测目标在图片中占比较小,可适当调高图片分辨率以得到更高的精度,如检测目标在图片中占比较大,可适当调低图片分 辨率以得到更高的训练效率
- 迭代轮次 (epoch) : 训练集完整参与训练的次数,如模型精度较低,可适当调高迭轮次,使模型训练更完整

系统会根据本地训练资源以及所选算法类型自动采用最优组合,通常情况下推荐不做更改,如实际任务场景有需要,则可根据实际需求调整

#### 训练完成后同步发布为模型

任务训练完成后可通过评估、校验验证任务效果,任务效果满足实际使用要求后发布为模型完成模型部署流程,如当前任务已经过多轮迭代且任 务效果较有保证可勾选训练完成后同步发布为模型,并输入发布为的模型名称以及版本描述,训练成功后将会自动发布为模型

任务与模型一一对应,如当前任务已有版本发布为模型,则当前任务下的其他版本发布时仅支持发布在当前模型下

### 数据配置

### 添加训练数据

- 先选择数据集,再按分类选择数据集里的图片,可从多个数据集选择图片
- 训练时间与数据量大小有关,数据量越大,训练耗时越长

Tips :

- 图像分类任务类型,如只有1个分类需要识别,或者实际业务场景所要识别的图片内容不可控,可以在训练前勾选"增加识别结果为[其他]
   的默认分类"。勾选后,模型会将与训练集无关的图片识别为"其他"
- 如果同一个分类的数据分散在不同的数据集里,可以在训练时同时从这些数据集里选择分类,模型训练时会合并分类名称相同的图片

### 添加自定义验证集

AI模型在训练时,每训练一批数据会进行模型效果检验,以某一张验证图片作为验证数据,通过验证结果反馈去调节训练。可以简单地把AI模型 训练理解为学生学习,训练集则为每天的上课内容,验证集即为每周的课后作业,质量更高的每周课后作业能够更好的指导学生学习并找寻自己 的不足,从而提高成绩。同理AI模型训练的验证集也是这个功效。

注:学生的课后作业应该与上课内容对应,这样才能巩固知识。因此,验证集的标签也应与训练集完全一致。

### 添加自定义测试集

如果学生的期末考试是平时的练习题,那么学生可能通过记忆去解题,而不是通过学习的方法去做题,所以期末考试的试题应与平时作业不能一样,才能检验学生的学习成果。那么同理,AI模型的效果测试不能使用训练数据进行测试,应使用训练数据集外的数据测试,这样才能真实的反

映模型效果。

注:期末考试的内容属于学期的内容,但不一定需要完全包括所学内容。同理,测试集的标签是训练集的全集或者子集即可。

### 配置数据增强策略

深度学习模型的成功很大程度上要归功于大量的标注数据集。通常来说,通过增加数据的数量和多样性往往能提升模型的效果。当在实践中无法 收集到数目庞大的高质量数据时,可以通过配置数据增强策略,对数据本身进行一定程度的扰动从而产生"新"数据。模型会通过学习大量 的"新"数据,提高泛化能力。

你可以在「默认配置」、「手动配置」2种方式中进行选择,完成数据增强策略的配置。

### 默认配置

如果你不需要特别配置数据增强策略,就可以选择默认配置。后台会根据你选择的算法,自动配置必要的数据增强策略。

### 手动配置

飞桨EasyDL提供了大量的数据增强算子供开发者手动配置。你可以通过下方的算子功能说明或训练页面的效果展示,来了解不同算子的功能:

| 算子名          | 功能                         |
|--------------|----------------------------|
| ShearX       | 剪切图像的水平边                   |
| ShearY       | 剪切图像的垂直边                   |
| TranslateX   | 按指定距离(像素点个数)水平移动图像         |
| TranslateY   | 按指定距离(像素点个数)垂直移动图像         |
| Rotate       | 按指定角度旋转图像                  |
| AutoContrast | 自动优化图像对比度                  |
| Contrast     | 调整图像对比度                    |
| Invert       | 将图像转换为反色图像                 |
| Equalize     | 将图像转换为灰色值均匀分布的图像           |
| Solarize     | 为图像中指定阈值之上的所有像素值取反         |
| Posterize    | 减少每个颜色通道的bits至指定位数         |
| Color        | 调整图像颜色平衡                   |
| Brightness   | 调整图像亮度                     |
| Sharpness    | 调整图像清晰度                    |
| Cutout       | 通过随机遮挡增加模型鲁棒性,可设定遮挡区域的长宽比例 |

效果展示

剪切图像的水平边,能更好地识别发生了水平方向形变的图像



×

飞桨EasyDL支持以Intel CPU,或可通过NVIDIA旗下不同型号显卡加速训练。

如果您的计算机有NVIDIA® GPU,且需要使用GPU环境进行训练,请确保满足以下条件:

- Windows 7/8/10/11: 需安装 CUDA 11.2 与 cuDNN v8.2.1
- Ubuntu 16.04/18.4/20.4: 需 CUDA 11.2 与 cuDNN v8.1.1
- CentOS 7:需 CUDA 11.2 与 cuDNN v8.1.1

### CUDA、cuDNN安装指南

您可参考NVIDIA官方文档了解CUDA和CUDNN的安装流程和配置方法,详见:

- CUDA 安装指南: https://docs.nvidia.com/cuda/
- cuDNN安装指南:https://docs.nvidia.com/deeplearning/cudnn/install-guide/index.html

### 心 预置模型调参模式开发

为有一定AI开发基础的开发者提供预置模型调参建模方式,涵盖ResNet、YOLO、PicoDet、MaskRCNN等近30种网络类型,适配大部分场景,开发者只需选择合适的预训练模型以及网络,根据自身经验进行调整,以获得更适合特定场景的模型。

### 训练配置

预置模型调参模式下支持对模型文件导出类型、训练使用算法与网络类型、网络参数等内容进行设置。



### 导出类型

- 导出模型源文件:训练完成后支持将模型源文件导出,模型源文件可通过Paddle-Inference转化至实际应用场景中使用
- 导出模型源文件与离线SDK:训练完成后可直接将模型发布为SDK包,可直接用于业务集成,省去繁琐转化过程

## 预置算法及网络选择

预置模型调参模式中,现已含盖ResNet50\_vd、YOLOv3\_MobileNetV1、SSD\_MobileNetV1、Mask\_RCNN\_R50\_vd\_FPN等图像分类、物体检测、 实例分割3类场景下14类网络,后续还将进一步提升预置网络数量。

## 图像分类

| 算法               | 网络                     | 适用场景                      | 预训练模型    |
|------------------|------------------------|---------------------------|----------|
| ResNet           | ResNet50_vd            | 整体效果稳定、预测时间较短             | ImageNet |
| SENet            | SE_ResNeXt50_32x4d     | 整体效果稳定、预测时间较短             | ImageNet |
| ResNext          | ResNeXt101_32x16d_wsl  | 适用于数据量较大且准确率要求高时,预测时间将会更长 | ImageNet |
| EfficientNot     | EfficientNetB0_small   | 适用于要求预测时间或模型体积尽量小的场景      | ImageNet |
| Encientiver      | EfficientNetB4         | 适用于数据量较大且准确率要求高时,预测时间将会更长 | ImageNet |
| b de la lle bied | MobileNetV2            | 适用于要求预测时间或模型体积尽量小的场景      | ImageNet |
| MobileNet        | MobileNetV3_large_x1_0 | 适用于要求预测时间或模型体积尽量小的场景      | ImageNet |

### 物体检测

| 算法          | 网络                   | 适用场景                | 预训练模型 |
|-------------|----------------------|---------------------|-------|
| CascadeRCNN | Cascade_RCNN_R50_FPN | 适用于模型精度要求更高的场景      | сосо  |
| FasterRCNN  | Faster_RCNN_R50_FPN  | 适用于模型精度要求更高的场景      | COCO  |
|             | YOLOv3_MobileNetV1   | 适用于预测性能要求更高的场景      | COCO  |
| YOLOv3      | YOLOv3_DarkNet53     | 适用于预测性能和模型精度相对折中的场景 | COCO  |
|             | YOLOv3_R50_vd_DCN    | 适用于模型精度要求更高的场景      | сосо  |

### 实例分割

| 算法       | 网络                   | 适用场景          | 预训练模型 |
|----------|----------------------|---------------|-------|
| MaskRCNN | Mask_RCNN_R50_vd_FPN | 模型效果稳定,预测时间适中 | COCO  |

### 高级训练配置 (模型参数)

高级训练配置中提供了预处理参数、训练参数、优化器参数可供选择调整,系统将根据训练任务及数据情况自动设置高级参数,也支持勾选部分 参数进行手动设置。

## 图像分类

### 预处理参数

| 参数                  | 说明                                                                    |
|---------------------|-----------------------------------------------------------------------|
| InputSize           | 输入图像分辨率,值越大训练时显存占用越大。                                                 |
| BatchSize           | 一次训练所选取的样本数,值越大训练时显存占用越大。                                             |
| LabelSmoonthi<br>ng | 标签平滑,是机器学习领域的一种正则化方法,通常用于分类问题,目的是防止模型在训练时过于自信地预测标签,改善泛化<br>能力差的问题。    |
| MixUpAlpha          | 图像混合系数,MixUp是一种数据增强方式,通过将随机的两张样本按比例混合,在训练过程中不会出现非信息像素,从而能够<br>提高训练效率。 |

## 训练参数

| 参数                | 说明                           |
|-------------------|------------------------------|
| Epochs            | 训练集完整参与训练的次数,迭代轮数越多训练耗时越长。   |
| EvalInterval      | 训练中评估的次数,评估间隔越小评估越频繁,训练耗时越长。 |
| ModelSaveInterval | 训练中模型保存间隔,间隔越小训练过程中保存的模型越多。  |
| LogInterva        | 训练日志输出间隔,间隔越小训练日志内容越多。       |

| 参数                   | 说明                                                                                                         |
|----------------------|------------------------------------------------------------------------------------------------------------|
| BaseLearn<br>ingRate | 单BatchSize对应的学习率,会根据BatchSize和卡数线性增大。学习率是在梯度下降的过程中更新权重时的超参数。学习率过高会导<br>致模型难以收敛;过低会可能导致模型收敛速度过慢,建议根据经验设定合理值 |
| WeightDec<br>ay      | 分位L1和L2正则。L1正则化系数,用于权重矩阵稀疏;L2正则化系数,用于防止模型对训练数据过拟合。但系数过大,可能导致欠<br>拟合。                                       |
| WeightDec<br>ayCoeff | 分位L1和L2正则。L1或L2正则:L1正则化系数,用于权重矩阵稀疏;L2正则化系数,用于防止模型对训练数据过拟合。但系数过<br>大,可能导致欠拟合。                               |
| Optimizer            | 优化器是在梯度下降中影响参数更新方式的选项,有Momentum,RMSProp,Adam等。                                                             |
| Momentu<br>m         | 参数更新的计算公式中的动量因子                                                                                            |
| LRDecayTy<br>pe      | 学习率衰减策略,随着训练迭代数的增加,学习率慢慢减小,以帮助模型收敛,常见的策略有阶梯式下降(PiecewiseDecay),余<br>弦下降(CosineDecay)等。                     |
| UseWarm<br>Up        | 是否使用学习率热身策略。                                                                                               |
| WarmUpSt<br>eps      | 学习率热身迭代数,通过该迭代次数后,学习率慢慢增加到指定值。                                                                             |

## 物体检测

### 网络参数

| 参数             | 说明                            |
|----------------|-------------------------------|
| КеерТорК       | 每个图像可以保留的总边界框数。               |
| ScoreThreshold | 过滤掉低置信度分数的边界框的阈值。             |
| NmsThreshold   | 对检测框进行非极大值抑制(NMS)时使用的IOU阈值。   |
| Anchors        | 基于锚框的检测框架里对预置锚框(anchor)大小的设置。 |

## 预处理参数

| 参数        | 说明                        |
|-----------|---------------------------|
| InputSize | 输入图像分辨率,值越大训练时显存占用越大。     |
| BatchSize | 一次训练所选取的样本数,值越大训练时显存占用越大。 |

## 训练参数

| 参数                | 说明                           |
|-------------------|------------------------------|
| Epochs            | 训练集完整参与训练的次数,迭代轮数越多训练耗时越长。   |
| EvalInterval      | 训练中评估的次数,评估间隔越小评估越频繁,训练耗时越长。 |
| ModelSaveInterval | 训练中模型保存间隔,间隔越小训练过程中保存的模型越多。  |
| LogInterva        | 训练日志输出间隔,间隔越小训练日志内容越多。       |

## 优化器参数

| 参数                   | 说明                                                                                                         |  |
|----------------------|------------------------------------------------------------------------------------------------------------|--|
| BaseLearn<br>ingRate | 单BatchSize对应的学习率,会根据BatchSize和卡数线性增大。学习率是在梯度下降的过程中更新权重时的超参数。学习率过高会导<br>致模型难以收敛;过低会可能导致模型收敛速度过慢,建议根据经验设定合理值 |  |
| WeightDec<br>ay      | 分位L1和L2正则。L1正则化系数,用于权重矩阵稀疏;L2正则化系数,用于防止模型对训练数据过拟合。但系数过大,可能导致欠<br>拟合。                                       |  |
| WeightDec<br>ayCoeff | 分位L1和L2正则。L1或L2正则:L1正则化系数,用于权重矩阵稀疏;L2正则化系数,用于防止模型对训练数据过拟合。但系数过<br>大,可能导致欠拟合。                               |  |
| Optimizer            | 优化器是在梯度下降中影响参数更新方式的选项,有Momentum,RMSProp,Adam等。                                                             |  |
| Momentu<br>m         | 参数更新的计算公式中的动量因子                                                                                            |  |
| LRDecayTy<br>pe      | 学习率衰减策略,随着训练迭代数的增加,学习率慢慢减小,以帮助模型收敛,常见的策略有阶梯式下降(PiecewiseDecay),余<br>弦下降(CosineDecay)等。                     |  |
| UseWarm<br>Up        | 是否使用学习率热身策略。                                                                                               |  |
| WarmUpSt<br>eps      | 学习率热身迭代数,通过该迭代次数后,学习率慢慢增加到指定值。                                                                             |  |

## 实例分割

## 网络参数

| 参数             | 说明                            |
|----------------|-------------------------------|
| КеерТорК       | 每个图像可以保留的总边界框数。               |
| ScoreThreshold | 过滤掉低置信度分数的边界框的阈值。             |
| NmsThreshold   | 对检测框进行非极大值抑制(NMS)时使用的IOU阈值。   |
| Anchors        | 基于锚框的检测框架里对预置锚框(anchor)大小的设置。 |

## 预处理参数

| 参数        | 说明                        |  |
|-----------|---------------------------|--|
| InputSize | 输入图像分辨率,值越大训练时显存占用越大。     |  |
| BatchSize | 一次训练所选取的样本数,值越大训练时显存占用越大。 |  |

## 训练参数

| 参数                | 说明                           |
|-------------------|------------------------------|
| Epochs            | 训练集完整参与训练的次数,迭代轮数越多训练耗时越长。   |
| EvalInterval      | 训练中评估的次数,评估间隔越小评估越频繁,训练耗时越长。 |
| ModelSaveInterval | 训练中模型保存间隔,间隔越小训练过程中保存的模型越多。  |
| LogInterva        | 训练日志输出间隔,间隔越小训练日志内容越多。       |

## 优化器参数

| 参致                   | 况明                                                                                                         |
|----------------------|------------------------------------------------------------------------------------------------------------|
| BaseLearn<br>ingRate | 单BatchSize对应的学习率,会根据BatchSize和卡数线性增大。学习率是在梯度下降的过程中更新权重时的超参数。学习率过高会导<br>致模型难以收敛;过低会可能导致模型收敛速度过慢,建议根据经验设定合理值 |
| WeightDec<br>ay      | 分位L1和L2正则。L1正则化系数,用于权重矩阵稀疏;L2正则化系数,用于防止模型对训练数据过拟合。但系数过大,可能导致欠<br>拟合。                                       |
| WeightDec<br>ayCoeff | 分位L1和L2正则。L1或L2正则:L1正则化系数,用于权重矩阵稀疏;L2正则化系数,用于防止模型对训练数据过拟合。但系数过<br>大,可能导致欠拟合。                               |
| Optimizer            | 优化器是在梯度下降中影响参数更新方式的选项,有Momentum,RMSProp,Adam等。                                                             |
| Momentu<br>m         | 参数更新的计算公式中的动量因子                                                                                            |
| LRDecayTy<br>pe      | 学习率衰减策略,随着训练迭代数的增加,学习率慢慢减小,以帮助模型收敛,常见的策略有阶梯式下降(PiecewiseDecay),余<br>弦下降(CosineDecay)等。                     |
| UseWarm<br>Up        | 是否使用学习率热身策略。                                                                                               |
| WarmUpSt<br>eps      | 学习率热身迭代数,通过该迭代次数后,学习率慢慢增加到指定值。                                                                             |
|                      |                                                                                                            |

#### 训练完成后同步发布为模型

任务训练完成后可通过评估、校验验证任务效果,任务效果满足实际使用要求后发布为模型完成模型部署流程,如当前任务已经过多轮迭代且任务效果较有保证可勾选训练完成后同步发布为模型,并输入发布为的模型名称以及版本描述,训练成功后将会自动发布为模型

任务与模型一一对应,如当前任务已有版本发布为模型,则当前任务下的其他版本发布时仅支持发布在当前模型下

### 数据配置

### 添加训练数据

- 先选择数据集,再按分类选择数据集里的图片,可从多个数据集选择图片
- 训练时间与数据量大小有关,数据量越大,训练耗时越长 Tips:
- 图像分类任务类型,如只有1个分类需要识别,或者实际业务场景所要识别的图片内容不可控,可以在训练前勾选"增加识别结果为[其他]的默认分类"。勾选后,模型会将与训练集无关的图片识别为"其他"
- 如果同一个分类的数据分散在不同的数据集里,可以在训练时同时从这些数据集里选择分类,模型训练时会合并分类名称相同的图片 添加自定义验证集

AI模型在训练时,每训练一批数据会进行模型效果检验,以某一张验证图片作为验证数据,通过验证结果反馈去调节训练。可以简单地把AI模型 训练理解为学生学习,训练集则为每天的上课内容,验证集即为每周的课后作业,质量更高的每周课后作业能够更好的指导学生学习并找寻自己 的不足,从而提高成绩。同理AI模型训练的验证集也是这个功效。

注:学生的课后作业应该与上课内容对应,这样才能巩固知识。因此,验证集的标签也应与训练集完全一致。

#### 添加自定义测试集

如果学生的期末考试是平时的练习题,那么学生可能通过记忆去解题,而不是通过学习的方法去做题,所以期末考试的试题应与平时作业不能一样,才能检验学生的学习成果。那么同理,AI模型的效果测试不能使用训练数据进行测试,应使用训练数据集外的数据测试,这样才能真实的反映模型效果。

注:期末考试的内容属于学期的内容,但不一定需要完全包括所学内容。同理,测试集的标签是训练集的全集或者子集即可。

#### 配置数据增强策略

深度学习模型的成功很大程度上要归功于大量的标注数据集。通常来说,通过增加数据的数量和多样性往往能提升模型的效果。当在实践中无法 收集到数目庞大的高质量数据时,可以通过配置数据增强策略,对数据本身进行一定程度的扰动从而产生"新"数据。模型会通过学习大量

的"新"数据,提高泛化能力。

你可以在「默认配置」、「手动配置」2种方式中进行选择,完成数据增强策略的配置。

## 默认配置

如果你不需要特别配置数据增强策略,就可以选择默认配置。后台会根据你选择的算法,自动配置必要的数据增强策略。

## 手动配置

飞桨EasyDL提供了大量的数据增强算子供开发者手动配置。你可以通过下方的算子功能说明或训练页面的效果展示,来了解不同算子的功能:

| 算子名          | 功能                         |  |
|--------------|----------------------------|--|
| ShearX       | 剪切图像的水平边                   |  |
| ShearY       | 剪切图像的垂直边                   |  |
| TranslateX   | 按指定距离(像素点个数)水平移动图像         |  |
| TranslateY   | 按指定距离(像素点个数)垂直移动图像         |  |
| Rotate       | 按指定角度旋转图像                  |  |
| AutoContrast | 自动优化图像对比度                  |  |
| Contrast     | 周整图像对比度                    |  |
| Invert       | 将图像转换为反色图像                 |  |
| Equalize     | 将图像转换为灰色值均匀分布的图像           |  |
| Solarize     | 为图像中指定阈值之上的所有像素值取反         |  |
| Posterize    | 减少每个颜色通道的bits至指定位数         |  |
| Color        | 调整图像颜色平衡                   |  |
| Brightness   | 调整图像亮度                     |  |
| Sharpness    | 调整图像清晰度                    |  |
| Cutout       | 通过随机遮挡增加模型鲁棒性,可设定遮挡区域的长宽比例 |  |

## 效果展示

×

## 剪切图像的水平边,能更好地识别发生了水平方向形变的图像



关闭

### 训练环境配置

飞桨EasyDL支持以Intel CPU,或可通过NVIDIA旗下不同型号显卡加速训练。

如果您的计算机有NVIDIA® GPU,且需要使用GPU环境进行训练,请确保满足以下条件:

- Windows 7/8/10/11:需安装 CUDA 11.2 与 cuDNN v8.2.1
- Ubuntu 16.04/18.4/20.4: 需 CUDA 11.2 与 cuDNN v8.1.1
- CentOS 7:需 CUDA 11.2 与 cuDNN v8.1.1

## CUDA、cuDNN安装指南

您可参考NVIDIA官方文档了解CUDA和CUDNN的安装流程和配置方法,详见:

- CUDA 安装指南:https://docs.nvidia.com/cuda/
- cuDNN安装指南:https://docs.nvidia.com/deeplearning/cudnn/install-guide/index.html

## 心 效果评估

训练完成后,可通过模型评估报告或模型校验了解模型效果。

- 模型评估报告:训练完成后,可以在【任务总览】列表中看到模型效果,以及详细的模型评估报告。
- 模型校验:训练完成后,可以在【任务总览】操作中点击「校验」,实时校验模型效果。

## 模型评估报告

图像分类

## 整体评估

在这个部分可以看到模型训练整体的情况说明,包括基本结论、准确率、F1-score等。这部分模型效果的指标是基于训练数据集,随机抽出部分数据不参与训练,仅参与模型效果评估计算得来。所以**当数据量较少时(如图片数量低于100个),参与评估的数据可能不超过30个,这样得出的模型评估报告效果仅供参考,无法完全准确体现模型效果。** 

查看模型评估结果时,需要思考在当前业务场景,更关注精确率与召回率哪个指标。是更希望减少误识别,还是更希望减少漏识别。前者更需要 关注精确率的指标,后者更需要关注召回率的指标。同时F1-score可以有效关注精确率和召回率的平衡情况,对于希望准确率与召回率兼具的场 景,F1-score越接近1效果越好。评估指标具体的说明如下。

- F1-score: 对某类别而言为精确率和召回率的调和平均数,评估报告中指各类别F1-score的平均数
- 准确率: 基于随机测试集进行计算,为正确分类的样本数与总样本数之比

注意:若想要更充分了解模型效果情况,建议发布模型为API后,通过调用接口批量测试,获取更准确的模型效果。

| ▲ 整体评估 |                 |        |       |                      |        |
|--------|-----------------|--------|-------|----------------------|--------|
| С      | 准确率 ⑦<br>100.0% |        | 0     | F1-score ⑦<br>100.0% |        |
| Top2   |                 | 100.0% | 精确率 ② |                      | 100.0% |
| Тор3   |                 | 100.0% | 召回率 ② |                      | 100.0% |
| Тор4   |                 | 100.0% |       |                      |        |
| Top5   |                 | 100.0% |       |                      |        |
|        |                 |        |       |                      |        |

• top1-top5准确率 对于每一个评估的图片文件,模型会给根据置信度高低,依次给出top1-top5的识别结果,其中top1置信度最高,top5的置信 度最低。那么top1的准确率值是指对于评估标准为"top1结果识别为正确时,判定为正确"给出准确率。top2准确率值是指对于评估标准为"top1 或者top2只要有一个命中正确的结果,即判定为正确"给出的准确率。……以此类推。

## 详细评估

这个部分支持查看模型识别错误的图片示例,以及使用混淆矩阵定位易混淆的分类。

## 识别错误图片示例

通过分标签查看模型识别错误的图片,寻找其中的共性,进而有针对性的扩充训练数据。

### 详细评估 按分类查看错误示例

不同分类的F1-score及对应的识别错误的图片(不包含训练时可能勾选的"其他"类识别错误的图片)



例如,你训练了一个将小番茄和樱桃分类的模型。在查看小番茄分类的错误示例时,发现错误示例中有好几张图片都是带着绿色根茎的小番茄 (与樱桃比较相似)。这种情况下,就需要在小番茄分类的训练集中,多增加一些带绿色根茎的图片,让模型有足够的数据能够学习到带根茎的 小番茄和樱桃的区别。

这个例子中,我们找到的是识别错误的图片中,目标特征上的共性。除此之外,还可以观察识别错误的图片在以下维度是否有共性,比如:图片的拍摄设备、拍摄角度,图片的亮度、背景等等。



### 物体检测

### 整体评估

在这个部分可以看到模型训练整体的情况说明,包括基本结论、mAP、精确率、召回率。这部分模型效果的指标是基于训练数据集,随机抽出部 分数据不参与训练,仅参与模型效果评估计算得来。所以**当数据量较少时(如图片数量低于100个),参与评估的数据可能不超过30个,这样得** 出的模型评估报告效果仅供参考,无法完全准确体现模型效果。

注意:若想要更充分了解模型效果情况,建议发布模型为API后,通过调用接口批量测试,获取更准确的模型效果。

整体评估

cat\_dog V2效果优异 ,建议针对识别错误的图片示例继续优化模型效果。由于目前训练集数据量较少,该结论仅供参考,建议扩充训练集得到更准确的评估效果。 如何优化 效果?



查看模型评估结果时,需要思考在当前业务场景,更关注精确率与召回率哪个指标。是更希望减少误识别,还是更希望减少漏识别。前者更需要 关注精确率的指标,后者更需要关注召回率的指标。同时F1-score可以有效关注精确率和召回率的平衡情况,对于希望准确率与召回率兼具的场 景,F1-score越接近1效果越好。评估指标说明如下

- F1-score: 对某类别而言为精确率和召回率的调和平均数,评估报告中指各类别F1-score的平均数
- mAP: mAP(mean average precision)是物体检测(Object Detection)算法中衡量算法效果的指标。对于物体检测任务,每一类object都可以计算 出其精确率(Precision)和召回率(Recall),在不同阈值下多次计算/试验,每个类都可以得到一条P-R曲线,曲线下的面积就是average
- 精确率:正确预测的物体数与预测物体总数之比。评估报告中具体指经比较F1-score最高的阈值下的结果

• 召回率: 正确预测的物体数与真实物体数之比。评估报告中具体指经比较F1-score最高的阈值下的结果

### 详细评估

在这个部分可以看到不同阈值下的F1-score、模型识别错误的图片示例,以及使用混淆矩阵定位易混淆的标签。

详细评估



不同标签的mAP及对应的识别错误的图片



## 识别错误图片示例

通过分标签查看模型识别错误的图片,直寻找其中的共性,进而有针对性的扩充训练数据;或发现是标注错误,从而直接点击修改标注来将标注 修正

如下图所示,可以通过勾选「误识别」、「漏识别」来分别查看两种错误识别的情况:

| 错 | 吴详情                   |                                       | × |
|---|-----------------------|---------------------------------------|---|
|   | 原标注结果 ⑦               | 模型识别结果                                |   |
| < | AND TO ACC. OLDO AD   |                                       | > |
|   | 正确识别 • 误识别 • 漏识别 🔽 修改 | 标注 如何解读错误示例? 🛛 🔽 正确识别 🔽 误识别 ⑦ 🔽 漏识别 🤇 | Ð |

误识别:红框内没有目标物体(准备训练数据时没有标注),但模型识别到了目标物体观察误识别的目标有什么共性:例如,一个检测电动车的模型,把很多自行车误识别成了电动车(因为电动车和自行车外观上比较相似)。这时,就需要在训练集中为自行车特别建立一个标签,并且在所有训练集图片中,将自行车标注出来。

可以把模型想象成一个在认识世界的孩童,当你告诉他电动车和自行车分别是什么样时,他就能认出来;当你没有告诉他的时候,他就有可能把 自行车认成电动车。

漏识别: 橙框内应该有目标物体(准备训练数据时标注了),但模型没能识别出目标物体观察漏识别的目标有什么共性:例如,一个检测会议室参会人数的模型,会漏识别图片中出现的白色人种。这大概率是因为训练集中缺少白色人种的标注数据造成的。因此,需要在训练集中添加包含白色人种的图片,并将白色人种标注出来。

黄色人种和白色人种在外貌的差别上是比较明显的,由于几乎所有的训练数据都标注的是黄色人种,所以模型很可能认不出白色人种。需要增加 白色人种的标注数据,让模型学习到黄色人种和白色人种都属于「参会人员」这个标签。

以上例子中,我们找到的是识别错误的图片中,目标特征上的共性。除此之外,还可以观察识别错误的图片在以下维度是否有共性,比如: 图片的拍摄设备、拍摄角度,图片的亮度、背景等等。

## 实例分割

### 整体评估

在这个部分可以看到模型训练整体的情况说明,包括基本结论、mAP、精确率、召回率。这部分模型效果的指标是基于训练数据集,随机抽出部 分数据不参与训练,仅参与模型效果评估计算得来。所以**当数据量较少时(如图片数量低于100个),参与评估的数据可能不超过30个,这样得** 出的模型评估报告效果仅供参考,无法完全准确体现模型效果。

注意:若想要更充分了解模型效果情况,建议发布模型为API后,通过调用接口批量测试,获取更准确的模型效果。

#### 整体评估

100.0%

fenge01 V14效果优异,建议针对识别错误的图片示例继续优化模型效果。由于目前训练集数据量较少,该结论仅供参考,建议扩充训练集得到更准确的评估效果。 如何优 化效果?

mAP ⑦ 精确率 ⑦



查看模型评估结果时,需要思考在当前业务场景,更关注精确率与召回率哪个指标。是更希望减少误识别,还是更希望减少漏识别。前者更需要 关注精确率的指标,后者更需要关注召回率的指标。同时F1-score可以有效关注精确率和召回率的平衡情况,对于希望准确率与召回率兼具的场 景,F1-score越接近1效果越好。评估指标说明如下

• F1-score: 对某类别而言为精确率和召回率的调和平均数,评估报告中指各类别F1-score的平均数

75.0%

- mAP: mAP(mean average precision)是物体检测(Object Detection)算法中衡量算法效果的指标。对于物体检测任务,每一类object都可以计算 出其精确率(Precision)和召回率(Recall),在不同阈值下多次计算/试验,每个类都可以得到一条P-R曲线,曲线下的面积就是average
- 精确率:正确预测的物体数与预测物体总数之比。评估报告中具体指经比较F1-score最高的阈值下的结果
- 石回率: 正确预测的物体数与真实物体数之比。评估报告中具体指经比较F1-score最高的阈值下的结果详细评估

在这个部分可以看到不同阈值下的F1-score,以及模型识别错误的图片示例。



```
不同阈值下F1-score表现
```



不同标签的mAP及对应的识别错误的图片



#### 识别错误图片示例

通过分标签查看模型识别错误的图片,寻找其中的共性,进而有针对性的扩充训练数据。

如下图所示,可以通过勾选「误识别」、「漏识别」来分别查看两种错误识别的情况:



 误识别:红色遮盖内没有目标物体(准备训练数据时没有标注),但模型识别到了目标物体观察误识别的目标有什么共性:例如,一个检测 电动车的模型,把很多自行车误识别成了电动车(因为电动车和自行车外观上比较相似)。这时,就需要在训练集中为自行车特别建立一个标 签,并且在所有训练集图片中,将自行车标注出来。

可以把模型想象成一个在认识世界的孩童,当你告诉他电动车和自行车分别是什么样时,他就能认出来;当你没有告诉他的时候,他就有可能把 自行车认成电动车。

•漏识别:橙色遮盖内应该有目标物体(准备训练数据时标注了),但模型没能识别出目标物体观察漏识别的目标有什么共性:例如,一个检测会议室参会人数的模型,会漏识别图片中出现的白色人种。这大概率是因为训练集中缺少白色人种的标注数据造成的。因此,需要在训练集中添加包含白色人种的图片,并将白色人种标注出来。

黄色人种和白色人种在外貌的差别上是比较明显的,由于几乎所有的训练数据都标注的是黄色人种,所以模型很可能认不出白色人种。需要增加 白色人种的标注数据,让模型学习到黄色人种和白色人种都属于「参会人员」这个标签。

以上例子中,我们找到的是识别错误的图片中,目标特征上的共性。除此之外,还可以观察识别错误的图片在以下维度是否有共性,比如: 图片的拍摄设备、拍摄角度,图片的亮度、背景等等。

### 模型校验

#### 图像分类

| •••                                       | 开发                   | く > 校验模型                       |          | 0 0                   |
|-------------------------------------------|----------------------|--------------------------------|----------|-----------------------|
| EasyDL                                    | 零代码开发<br>Notebook 开发 | 选择任务 <b>害虫识别 &gt; </b> 选择版本 V1 | ∨ 训练算法   | MobileNet-MobileNetV2 |
| 10 W                                      | Hotobook 710         | 当前模型准确率 100.00% 评估报告           | 调整阈值 •〇一 | 0.03                  |
| C                                         |                      | 识别结果:                          | 预测分类     | 置信度 > 3.00% ∨         |
| 数据<br><b>开发</b><br>供型<br>服务<br>① (1)<br>① |                      |                                | 菜粉罐      | 99.99%                |
|                                           |                      |                                |          |                       |

当前模型mAP平均精度 89.13% 评估报告

| ME |
|----|

| 只别结果 如何优化效果? |              |
|--------------|--------------|
| 周整阈值         | —— 当前阈值: 0.3 |
| 预测标签         | 置信度>30%      |
| 1. person    | 98.41%       |
| 2. positive  | 94.63%       |
|              |              |
|              |              |
|              |              |
|              |              |
|              |              |
|              |              |
|              |              |

## 实例分割

当前模型mAP平均精度 100.00% 评估报告



| 识别结果 如何优化效果? |         |     |  |  |  |  |  |  |
|--------------|---------|-----|--|--|--|--|--|--|
| 调整阈值         |         | 0.3 |  |  |  |  |  |  |
| 预测标签         | 置信度>30% |     |  |  |  |  |  |  |
| 1. head      | 79.28%  |     |  |  |  |  |  |  |
|              |         |     |  |  |  |  |  |  |
|              |         |     |  |  |  |  |  |  |
|              |         |     |  |  |  |  |  |  |
|              |         |     |  |  |  |  |  |  |
|              |         |     |  |  |  |  |  |  |
|              |         |     |  |  |  |  |  |  |
|              |         |     |  |  |  |  |  |  |
|              |         |     |  |  |  |  |  |  |

# Notebook开发

## の Notebook简介

集成了包括PaddleX、PaddleDetection、PaddleSeg、PaddleClas等端到端开发套件的轻量级IDE,用户可在该模块内进行代码编辑、调试等开发 工作,快速高效的完成各类任务的实现,可对预置模型调参中的模型进行代码级优化。

| ••  | •                                                      |           |                |            |              |         | 飞桨 EasyDL Notebook            |          |       |         |       |     |
|-----|--------------------------------------------------------|-----------|----------------|------------|--------------|---------|-------------------------------|----------|-------|---------|-------|-----|
| В   | ML CodeLab 文件                                          | 编辑        | 宣看             | 运行 内核      | 标签页          | 设置 帮助   |                               |          |       |         |       |     |
|     | · •                                                    | Ŧ         | 0              | 🗐 main,    | cls.ipynb ×  | ≥ 包管理 × |                               |          |       |         | +   : | \$  |
| •   | 按文件名过滤                                                 |           |                | 精选推荐       | 已安装          | 未安装     |                               |          |       |         |       | ٥   |
| ۵   | =/                                                     |           |                | 请输入者       | 称            | ٩       |                               |          |       |         | 0     |     |
| œ   | 名称                                                     | ^   ·     | 修改时间           | 名称         |              |         | 描述                            | 最新版本     | 当前版本  | 操作      |       |     |
| 0   | 🖬 data                                                 |           | 6 天前           | paddlep    | addle        |         | 百度开源的深度学习框架                   | 2.2.2    | 2.2.2 | 更换版本 卸载 |       |     |
| A   | output                                                 |           | 3 小时前          |            |              |         |                               |          |       |         |       |     |
|     | PaddleClas                                             |           | 5 天前           | paddlep    | addle-gpu    |         | 百度开源的深度学习框架gpu版               | 2.2.0rc0 | -     | 安装 卸载   |       |     |
|     | predictor                                              |           | 6 天前           | paddleł    | ub           |         | 飞桨预训练模型应用管理工具                 | 2.2.0    | -     | 安装 卸载   |       |     |
| O   | PretrainedModel                                        |           | 3 4/45/80      |            |              |         |                               |          |       |         |       |     |
| 0   | Y devkit.yaml                                          |           | 3 小时前          | paddlex    |              |         | 飞桨全流程开发工具                     | 2.1.0    | -     | 安装 卸载   |       |     |
| 8   | <ul> <li>Infer.log</li> <li>Imain_cls.ipynb</li> </ul> |           | 3 小时前<br>3 小时前 | paddled    | icr          |         | 飞桨OCR工具库                      | 2.4.0.3  | -     | 安装 卸载   |       |     |
|     |                                                        |           |                | visuald    |              |         | 飞桨可视化分析工具                     | 2.2.3    | 2.2.3 | 更换版本 卸载 |       |     |
|     |                                                        |           |                | paddles    | eg           |         | 飞桨端到端图像分割开发套件                 | 2.4.0    | -     | 安装 卸载   |       |     |
|     |                                                        |           |                | paddled    | las          |         | 飞桨图像识别套件                      | 2.2.1    | -     | 安装 卸载   |       |     |
|     |                                                        |           |                | paddled    | let          |         | 基于飞桨PaddlePaddle的踢到端目标检测套件    | 2.3.0    | -     | 安装 卸载   |       |     |
|     |                                                        |           |                | paddles    | peech        |         | 基于飞桨 PaddlePaddle 的语音方向的开源模型库 | 0.1.2    | -     | 安装 卸载   |       |     |
| 0 8 | 1                                                      | 77%(6.2 G | B/8.0 GB)      | 硬盘: 93%(21 | 7.2 GB/233.5 | GB)     |                               |          |       |         |       | 包管理 |
|     |                                                        |           |                |            |              |         |                               |          |       |         |       |     |

预置开发套件

| 名称                 | 说明                                                           |
|--------------------|--------------------------------------------------------------|
| PaddlePaddle       | 百度开源的深度学习框架                                                  |
| PaddlePaddle (GPU) | 百度开源的深度学习框架GPU版                                              |
| PaddleX            | 飞桨全流程开发工具                                                    |
| PaddleClas         | 基于飞桨PaddlePaddle的端到端图像识别开发套件                                 |
| PaddleDetection    | 基于飞桨PaddlePaddle的端到端目标检测开发套件                                 |
| PaddleSeg          | 基于飞桨PaddlePaddle的端到端图像分割开发套件                                 |
| PaddleOCR          | 飞桨OCR工具库                                                     |
| PaddleSpeech       | 基于飞桨 PaddlePaddle语音方向的开源模型库                                  |
| PaddleHub          | 飞桨预训练模型应用管理工具                                                |
| PaddleSlim         | 专注于深度学习模型压缩的工具库,提供剪裁、量化、蒸馏、和模型结构搜索等模型压缩策略,帮助用户快速实现模型的<br>小型化 |
| Visualdl           | 飞桨可视化分析工具                                                    |

## の Notebook建模示例

本文以物体检测任务类型为例,从启动Notebook到引入数据、训练模型,再到保存模型的全流程。

### 启动Notebook

## step1:在左侧导航栏中选择开发->Notebook开发

| •••                      | 开发                   | く 〉 训练任务                                                                          |                                                                                                                                                                   |                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | ? 0 |  |
|--------------------------|----------------------|-----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|--|
| Caryth<br>35<br>总览<br>设据 | 零代码开发<br>Notebook 开发 | Notebook 开发介绍<br>Notebook 是一种本地交互式开发环境<br>后动 Notebook<br>训练方式                     | 1,省去愈准备开发环境的过程,内嵌1                                                                                                                                                | 百度飞桨框架与强大编辑器,助力高效                       | 完成 AI 开发。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |     |  |
| ★★ 开发 模型 服务 订阅           | 6                    | レンジョン A Carl Andrews Andrews Andrews Andrews Andrews Andrews 法行務機 医学校構成 日本部でのおりてあ | ・          ・         ・          ・         たまたまます。           ・         たまたまます。           ・         たまたまます。           ・         たまたまます。           ・         たまたまます。 | (1) (1) (1) (1) (1) (1) (1) (1) (1) (1) | レビス (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997)<br>した (1997 |     |  |
|                          |                      |                                                                                   |                                                                                                                                                                   |                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |     |  |

## step2:选择开发语言、AI框架、资源规格与工作目录后启动Notebook

| • • • | 启动                                                                                      | 力 Notebook                                           |
|-------|-----------------------------------------------------------------------------------------|------------------------------------------------------|
| 开发语言  | Python 3.7                                                                              |                                                      |
| AI 框架 | PaddlePaddle 2.2.2                                                                      |                                                      |
| 资源规格  | 名称                                                                                      | 规格                                                   |
|       | O CPU                                                                                   | Intel(R) Core(TM) i5–8257U CPU @ 1.40G<br>Hz         |
| 工作目录  | <ul> <li>选择目录</li> <li>当前路径: /Users/heyun02</li> <li>workspace/notebook/runt</li> </ul> | 2/Library/Application Support/paddle-desktop,<br>ime |
|       |                                                                                         | 确认并启动 取消                                             |

### 训练物体检测模型

#### 下载 PaddleDetection 套件

打开进入 Notebook,点击进入终端,本文以 PaddleDetection 代码库 release/2.3 分支为例,输入如下命令克隆PaddleDetection代码库并切换 至release/2.3分支。整个过程需要数十秒,请耐心等待。

```
###### gitee 国内下载比较快
git clone https://gitee.com/paddlepaddle/PaddleDetection.git -b release/2.3
###### github
###### git clone https://github.com/PaddlePaddle/PaddleDetection.git -b release/2.3
```

### 安装环境

```
在终端环境中,安装该版本的 PaddleDetection 代码包依赖的 paddlepaddle-gpu,执行如下命令:
```

python -m pip install paddlepaddle-gpu==2.1.3.post101 -f https://www.paddlepaddle.org.cn/whl/linux/mkl/avx/stable.html

安装完成后,使用 python 或 python3 进入python解释器,输入 import paddle ,再输入 paddle.utils.run\_check()如果出现 PaddlePaddle is installedsuccessfully!,说明成功安装。**准备训练数据** 

训练数据是模型生产的重要条件,优质的数据集可以很大程度上的提升模型训练效果,准备数据可以参考链接。本文所用的安全帽检测数据集可 前往此链接进行下载:下载链接。

### step1:导入用户数据

目前在 Notebook 中不能直接访问您在 飞桨EasyDL 中创建的数据集,需要通过在终端输入数据所在路径。

### step2:数据转换

PaddleDetection 训练所需要的数据格式与 飞桨EasyDL 默认的数据格式有所不同,所以需要利用脚本将导入的数据转为 PaddleDetection 支持的数据格式,并进行3:7切分。

PaddleDetection 默认支持的标注格式为 COCO格式,转换脚本如下:

import os import cv2 import json import glob import codecs import random from pycocotools.coco import COCO def parse\_bml\_json(json\_file): """ 解析标注文件 :return: """ annos = json.loads(codecs.open(json\_file).read()))

labels = annos['labels'] bboxes = [] for label in labels: x1 = label["x1"]y1 = label["y1"] x2 = label["x2"]y2 = label["y2"] id = label["name"] bboxes.append([x1, y1, x2, y2, id]) return bboxes def bbox\_transform(box): .... x1, y1, x2, y2 转为 x1, y1, width, height :return .... box = list(map(lambda x: float(x), box)) box[2] = box[2] - box[0]box[3] = box[3] - box[1]return box def parse\_label\_list(src\_data\_dir, save\_dir): .... 遍历标注文件,获取label\_list :return: .... label\_list = [] anno\_files = glob.glob(src\_data\_dir + "\*.json") for anno\_f in anno\_files: annos = json.loads(codecs.open(anno\_f).read()) for lb in annos["labels"]: label\_list.append(lb["name"]) label\_list = list(set(label\_list)) with codecs.open(os.path.join(save\_dir, "label\_list.txt"), 'w', encoding="utf-8") as f: for id, label in enumerate(label\_list): f.writelines("%s:%s\n" % (id, label)) return len(label\_list), label\_list def bml2coco(src\_dir, coco\_json\_file): ..... 将标注格式转为COCO标注格式 :return: ..... coco\_images = [] coco\_annotations = [] image\_id = 0 anno\_id = 0 image\_list = glob.glob(src\_dir + "\*.[jJPpBb][PpNnMm]\*") for image\_file in image\_list: anno\_f = image\_file.split(".")[0] + ".json" if not os.path.isfile(anno\_f): continue bboxes = parse\_bml\_json(anno\_f) im = cv2.imread(image\_file) h, w, \_ = im.shape image\_i = {"file\_name": os.path.basename(image\_file), "id": image\_id, "width": w, "height": h} coco\_images.append(image\_i)

+ - /1

tor Id, bbox in enumerate(bboxes): # bbox : [x1, y1, x2, y2, label\_name] anno\_i = {"image\_id": image\_id, "bbox": bbox\_transform(bbox[:4]), 'category\_id': label\_list.index(bbox[4]), 'id': anno\_id, 'area': 1.1, 'iscrowd': 0, "segmentation": None} anno\_id += 1 coco\_annotations.append(anno\_i) image\_id += 1 coco\_categories = [{"id": id, "name": label\_name} for id, label\_name in enumerate(label\_list)] coco\_dict = {"info": "info", "licenses": "BMLCloud", "images": coco\_images, "annotations": coco\_annotations, "categories": coco\_categories} with open(coco\_json\_file, 'w', encoding="utf-8") as fin: json.dump(coco\_dict, fin, ensure\_ascii=False) def split\_det\_origin\_dataset( origin\_file\_path, train\_file\_path, eval\_file\_path, ratio=0.7): 按比例切分物体检测原始数据集 :return: .... coco = COCO(origin\_file\_path) img\_ids = coco.getImgIds() items\_num = len(img\_ids) train\_indexes, eval\_indexes = random\_split\_indexes(items\_num, ratio) train\_items = [img\_ids[i] for i in train\_indexes] eval\_items = [img\_ids[i] for i in eval\_indexes] dump\_det\_dataset(coco, train\_items, train\_file\_path) dump\_det\_dataset(coco, eval\_items, eval\_file\_path) return items\_num, len(train\_items), len(eval\_items) def random\_split\_indexes(items\_num, ratio=0.7): 按比例分割整个list的index :return:分割后的两个index子列表 .... offset = round(items\_num \* ratio) full\_indexes = list(range(items\_num)) random.shuffle(full\_indexes) sub\_indexes\_1 = full\_indexes[:offset] sub\_indexes\_2 = full\_indexes[offset:] return sub indexes 1, sub indexes 2 def dump\_det\_dataset(coco, img\_id\_list, save\_file\_path): ..... 物体检测数据集保存 :return: .... imgs = coco.loadImgs(img\_id\_list) img\_anno\_ids = coco.getAnnlds(imglds=img\_id\_list, iscrowd=0)

instances = coco.loadAnns(img\_anno\_ids)

cat\_ids = coco.getCatIds()

1329

categories = coco.ioadCats(cat\_ids) common\_dict = { "info": coco.dataset["info"], "licenses": coco.dataset["licenses"], "categories": categories } img\_dict = { "image\_nums": len(imgs), "images": imgs, "annotations": instances 1 img\_dict.update(common\_dict) json\_file = open(save\_file\_path, 'w', encoding='UTF-8') json.dump(img\_dict, json\_file) class\_nums, label\_list = parse\_label\_list("/home/work/data/\${dataset\_id}/", "/home/work/PretrainedModel/") bml2coco("/home/work/data/\${dataset\_id}/", "/home/work/PretrainedModel/org\_data\_list.json") split\_det\_origin\_dataset("/home/work/PretrainedModel/org\_data\_list.json", "/home/work/PretrainedModel/train\_data\_list.json", "/home/work/PretrainedModel/trainedModel/trainedModel/trainedModel/trainedModel/trainedModel/trainedModel/trainedModel/trainedModel/trai

将上述脚本存放为 convert.py 代码脚本,并将脚本最后两行的 "/home/work/data/\${dataset\_id}/" 均替换为所指定数据集路劲,在终端中运行 即可。 运行代码。

python covert.py

注意:如果报错 No module named 'pycocotools',需要通过如下命令安装相关依赖包,再运行 covert.py 代码。

pip install pycocotools

运行 covert.py 代码成功之后将在 PretrainedModel/ 文件夹下生成对应的数据文件,包括 label\_list.txt、train\_data\_list.json、eval\_data\_list.json、org\_data\_list.json。

### 训练模型

开发者准备好训练数据和安装环境之后即可开始训练物体检测模型。

"/home/work/PretrainedModel/eval\_data\_list.json")

step1:在终端中打开 PaddleDetection 目录

cd /PaddleDetection

### step2:修改yaml配置文件

在PaddleDetection 2.0后续版本,采用了模块解耦设计,用户可以组合配置模块实现检测器,并可自由修改覆盖各模块配置,本文以 configs/yolov3\_darknet53\_270e\_coco.yml 为例:

yolov3\_darknet53\_270e\_coco.yml 主配置入口文件 coco\_detection.yml 主要说明了训练数据和验证数据的路径 runtime.yml 主要说明了公共的运行参数,比如说是否使用GPU、每多少个epoch存储checkpoint等 optimizer\_270e.yml 主要说明了学习率和优化器的配置。 yolov3\_darknet53.yml 主要说明模型、和主干网络的情况。 yolov3\_reader.yml 主要说明数据读取器配置,如batch size,并发加载子进程数等,同时包含读取后预处理操作,如resize、数据增强等等

需要修改/覆盖的参数均可写在主配置入口文件中,主要修改点为训练、验证数据集路径、运行epoch数、学习率等,修改后的主配置文件如下 (注释行即为需要修改的点):

```
_BASE_: [
 '../datasets/coco_detection.yml',
 '../runtime.yml',
 '_base_/optimizer_270e.yml',
 '_base_/yolov3_darknet53.yml',
 '_base_/yolov3_reader.yml',
]
snapshot_epoch: 5
weights: output/yolov3_darknet53_270e_coco/model_final
###### 预训练权重地址
pretrain_weights: https://paddledet.bj.bcebos.com/models/yolov3_darknet53_270e_coco.pdparams
###### coco_detection.yml
num_classes: 2 #实际类别数
TrainDataset:
!COCODataSet
  image_dir: data/${dataset_id}/ #图片地址
  anno_path: PretrainedModel/train_data_list.json # 标注文件
  dataset_dir: /home/work/ # 数据集根目录
  data_fields: ['image', 'gt_bbox', 'gt_class', 'is_crowd']
EvalDataset:
 !COCODataSet
  image_dir: data/${dataset_id}/ #图片地址
  anno_path: PretrainedModel/eval_data_list.json #标注文件
  dataset_dir: /home/work/ # 数据集根目录
###### optimizer_270e.yml
epoch: 50 # 迭代轮数
LearningRate:
base_lr: 0.0001 # 学习率
 schedulers:
- !PiecewiseDecay
  gamma: 0.1
  milestones:
  - 30
  - 45
- !LinearWarmup
  start_factor: 0.
  steps: 400
```

## step3 : 训练模型

在终端中执行以下命令,开始模型训练。

```
cd /PaddleDetection/
python tools/train.py -c configs/yolov3/yolov3_darknet53_270e_coco.yml --eval
```

注意:如果报错 No module named 'lap' 和 No module named 'motmetrics' ,则需要通过如下命令安装相关依赖包,再运行 coversion.py 代 码。 (如果缺失其他模块,也可用类似命令下载安装)

pip install lap motmetrics

### step4:模型评估

在终端中执行以下命令,开始模型评估。

python tools/eval.py -c configs/yolov3/yolov3\_darknet53\_270e\_coco.yml \ -o weights=output/yolov3\_darknet53\_270e\_coco/model\_final

```
运行完成输出如下结果:
```

| Average Precisio | on (AP) @[ 100=0.50:0.95 | area= all                 | maxDets=100 ] = 0.279              |
|------------------|--------------------------|---------------------------|------------------------------------|
| Average Precisio | on (AP) @[ IOU=0.50      | area= all                 | maxDets=100 ] = 0.534              |
| Average Precisio | on (AP) @[ IOU=0.75      | area= all                 | maxDets=100 ] = 0.251              |
| Average Precisio | on (AP) @[ IoU=0.50:0.95 | area= small               | maxDets=100 ] = 0.298              |
| Average Precisio | on (AP) @[ IoU=0.50:0.95 | area=medium               | maxDets=100 ] = -1.000             |
| Average Precisio | on (AP) @[ IOU=0.50:0.95 | area= large               | maxDets=100 ] = -1.000             |
| Average Recall   | (AR) @[ IOU=0.50:0.95    | area= all                 | maxDets= 1 ] = 0.086               |
| Average Recall   | (AR) @[ IOU=0.50:0.95    | area= all                 | maxDets= 10 ] = 0.331              |
| Average Recall   | (AR) @[ IOU=0.50:0.95    | area= all                 | maxDets=100 ] = 0.351              |
| Average Recall   | (AR) @[ IOU=0.50:0.95    | area= small               | maxDets=100 ] = 0.351              |
| Average Recall   | (AR) @[ IOU=0.50:0.95    | area=medium               | maxDets=100 ] = -1.000             |
| Average Recall   | (AR) @[ IOU=0.50:0.95    | area= large               | maxDets=100 ] = -1.000             |
| [01/30 02:28:27] | ppdet.engine INFO: Total | <pre>sample number:</pre> | 150, averge FPS: 25.36864530446337 |

step5:模型预测

在终端中执行以下命令,开始模型预测(注意修改图片路径)。

```
python tools/infer.py -c configs/yolov3/yolov3_darknet53_270e_coco.yml \
--infer_img=/home/work/data/${task_id}/xxx.jpeg \
--output_dir=infer_output/ \
--draw_threshold=0.5 \
-o weights=output/yolov3_darknet53_270e_coco/model_final
```

### step6:导出模型

在终端中执行以下命令,将最佳模型转为可以用于发布的 inference 模型

```
python tools/export_model.py -c configs/yolov3/yolov3_darknet53_270e_coco.yml \
--output_dir=/home/work/PretrainedModel/ \
-o weights=output/yolov3_darknet53_270e_coco/model_final
```

在终端中执行以下命令,将导出模型移至 /PretrainedModel/ 目录。

mv /PretrainedModel/yolov3\_darknet53\_270e\_coco/\* /home/work/PretrainedModel/

#### 心 Http服务运行端注意事项

模型部署启动服务时需对以下文件进行修改:

#### 1. demo\_servering.py, 启动服务设备为cpu/gpu



2. demo\_client.py,发送请求设备任意,只有保证ip和端口号与起服务设备一致即可
<sup>②</sup> 零代码开发转Notebook开发

# 飞桨EasyDL支持将通过零代码开发-预置模型调参模式开发的模型转为对应模型文件在Notebook中打开进行优化。

#### step1:启动Notebook

通过预置模型调整模式完成训练后,在任务总览对应任务中点击【打开Notebook】。



# step2:调整权重参数与网络结构

结合实际需求,在Yaml文件中对网络参数及结构进行调整。



#### step3:运行模型文件

完成网络参数及结构调整后,运行对应main\_xxx.ipynb文件,即可完成环境配置、数据集准备、构建模型、训练模型、评估模型、模型部署等流程。

| ••                         |                 | main_cls.ipynb (2) - 飞菜 EasyDL Notebook                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |               |
|----------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| BML CodeLab 文件             | 编辑 查看           | 运行 内核 标签页 设置 帮助                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |               |
|                            | E 0             | 😝 main_cls.lpynb × 🗉 devkityami × +                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | : q           |
|                            | 1               | ▷ ② ⑦ □ 1: + Code ~ ○ 定位到当前进行Cell Python 3 (ipykernel) ● 3 (ipykernel)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 6             |
| 按文件名过意                     |                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |               |
| ● ■/<br>● 名称               | ↑ 修改时(          | 3.构建模型                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |               |
| 🖬 data                     | 6 天1            | devkit yami 中融合了您在前端设置的参数, 您也可以在 devkit yami 中修改参数以实现模型深度调优                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |               |
| output                     | 3 小时            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |               |
| PaddleClas                 | 5 天1            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |               |
| predictor                  | 6 天             | 4.训练模型                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |               |
| PretrainedModel            | 3 시네)           | 您可以通过下面命令发起训练,这里 -c 用于关联配置文件                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |               |
| Y: devkit.yaml             | 3 小时            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |               |
| infer.log                  | 3 小时            | [2] 1 ipython ./PaddleClas/tools/train.py \                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |               |
| • 🔛 main_cls.ipynb         | 3 小时            | B 2 -c./devkit.yaml                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |               |
|                            |                 | /Applications/で夏EasyOL.app/Contents/Resources/resources/notebook/lib/lib/python3.7/site-packages/paddle/nn/layer/pooling.py:1<br>187: DeprecationWarning: invalid escape sequence<br>"""<br>/Applications/飞襲EasyOL.app/Contents/Resources/resources/notebook/lib/lib/python3.7/site-packages/paddle/incubate/nn/functiona<br>Ufused transformer.py:111: DeprecationWarning: invalid escape sequence \<br>运行时任 752.405秒 結束時间: 2022-05-18 17:11:45                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 1             |
|                            |                 | <b>5. 评估模型</b><br>/output 文件夹内存放有训练过程中保存的模型和日志文件,忽可以加载 best_model 用于模型评估。 套件采用 Top! 作为评估指标,结果存放在 evailing<br>文件                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |               |
|                            |                 | [3] 1 !python ./PaddleClas/tools/eval.py \<br>2 -c ./devkit.yaml \<br>3 -o Global.pretrained_model=./output/MobileNetV2_x2_0/best_model                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |               |
|                            |                 | [2822/45/18 17:11:40] ppc1s NP0:         std : [0.229, 0.224, 0.225]           [2822/45/18 17:11:40] ppc1 NP0:         toat:         interpretation           [2822/45/18 17:11:40] ppc1 NP0:         interpretation         interpretation           [2822/45/18 17:11:40] ppc1 NP0:         interpretation         interpretation           [2822/45/18 17:11:40] ppc1 NP0:         sampler:         interpretation           [2822/45/18 17:11:40] ppc1 NP0:         batch_size:         28           [2822/45/18 17:11:40] ppc1 NP0:         force         drop_last:           [2822/45/18 17:11:40] ppc1 NP0:         note:         istributedBatchSampler           [2822/45/18 17:11:40] ppc1 NP0:         note:         istributedBatchSampler           [2822/45/18 17:11:40] ppc1 NP0:         samfler:         istributedBatchSampler           [2822/45/18 17:11:40] ppc1 NP0:         rate:         rate: | 0             |
| 0 5. 1 @ Python 3 (ipykern | nel)  空闲 CPU: 7 | %内存:76%(6.1 GB/8.0 GB) 硬盘:93%(217.2 GB/23.5 GB) 模式:命令 😵 行1,列1 m                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | nain_cls.ipyn |

# step4:生成模型版本

模型训练成功后,点击左侧菜单栏【生产模型版本】,将训练成功的模型发布至模型中心。



| ■ 牛成模型版本 団                     | [-] main_cls.jpynb × ≣ devkit.yaml ×                                                                                                                       | +   :                       |
|--------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|
| 拉圾分类 V1<br>2022-05-18 17:16:58 | ▶ 保存模型 ×                                                                                                                                                   | Python: 3 (ipykernel) 🔹 🕢   |
|                                | (株存領型単小店, 〒特徴型領型Ξ領型列及具, 支持等面積型源文件 ○ 模型属性 (2) 选择模型文件 部署則2                                                                                                   | <b>「需要反向传播,因此需要将模型进行导成部</b> |
| 9                              | 说明:模型文件应保存在PretrainedModel文件夹中<br>当前项目文件 已选文件                                                                                                              |                             |
| 9                              | devkit_cis/inference.pdipa ×<br>devkit_cis/inference.pdipa ×<br>devkit_cis/inference.pdipa ×<br>devkit_cis/inference.pdipa ×<br>devkit_cis/inference.pdm × |                             |
|                                | devkit_cis/label_list.txt ✓ devkit_cis/label_list.txt X iet                                                                                                |                             |
| •                              | : Pat                                                                                                                                                      | se<br>II II                 |
|                                | , 0.43<br>上一步 生成 8.224                                                                                                                                     | , 8.425]                    |
|                                | [2022/45/18 17:11:53] ppc1s TMF0: num_workers : 4<br>[3073/45/18 17:11:53] ppc1s TMF0: use charact memory · Teue<br>运行时长 5.196秒 越来时间: 2022-05-18 17:11:57  |                             |
|                                |                                                                                                                                                            |                             |

# 模型中心

模型列表

模型列表提供模型管理能力,可查看所有已发布为模型的任务

| 模型列表 |       |      |       |      | 全部   | 模型类型 🗸              | 请输入模型名称  |
|------|-------|------|-------|------|------|---------------------|----------|
| 模型部署 | 模型名称  | 模型ID | 模型来源  | 模型类型 | 版本数量 | 创建时间 🌲              | 操作       |
|      | 模型测试1 | 3    | 零代码开发 | 图像分类 | 1    | 2021-11-24 04:59:   | 07 查看 删除 |
|      | IEC测试 | 1    | 零代码开发 | 图像分类 | 1    | 2021-11-16 11:03:24 | 4 查看 删除  |
|      |       |      |       |      |      |                     |          |

# ලා 查看模型

# 可通过下拉框筛选需要查看的任务类型

|      |      |      |      | [    | 物体检测           | ~    | 请输入核 |
|------|------|------|------|------|----------------|------|------|
| 模型名称 | 模型ID | 模型来源 | 模型类型 | 版本数量 | 全部模型类型<br>图像分类 | 时间 🌲 | 操    |
|      |      |      |      |      | 物体检测           |      |      |
|      |      |      |      |      | 实例分割           |      |      |

点击【查看】可查看模型详情,在模型详情页可查看当前模型下的全部版本、模型版本对应的任务、训练方式、版本描述以及导入时间

# 点击对应任务可跳转至当前模型对应的任务版本

| 模型名称: 楨 | 輕型列表演示 | 模型ID: 4  |            |                     |               |  |  |
|---------|--------|----------|------------|---------------------|---------------|--|--|
| 模型类型:图  | 图像分类   | 模型来源:零代码 | 模型来源:零代码开发 |                     |               |  |  |
| 版本      | 对应任务   | 训练方式     | 描述         | 导入时间 💂              | 操作            |  |  |
| V2      | 7–V2   | 高级调参模式   | 这里是版本描述    | 2021-11-24 02:44:56 | 导出模型文件 ⑦ 部署 ☺ |  |  |
| V1      | 6-V1   | AutoDL模式 |            | 2021-11-22 01:07:31 | 导出模型文件 💮      |  |  |

如需将模型投入实际应用可选择导出模型源文件或将模型部署为离线SDK

仅有训练时导出类型选择为【导出源文件与离线SDK】的任务支持部署为离线SDK

| 版本 | 对应任务 | 训练方式     | 描述      | 导入时间 🌷              | 操作            |
|----|------|----------|---------|---------------------|---------------|
| V2 | 7–V2 | 高级调参模式   | 这里是版本描述 | 2021-11-24 02:44:56 | 导出模型文件 ⑦ 部署 ⊙ |
| V1 | 6-V1 | AutoDL模式 |         | 2021-11-22 01:07:31 | 导出模型文件 ⊙      |

# 模型部署

服务发布界面可选择将模型发布为离线SDK,发布包含以下流程

# 心 选择模型及版本

选择需要发布的模型及版本,选择完成后当前版本对应的训练方式及训练算法将自动展示

| 选择模型 | 模型列表演示             | $\sim$ |
|------|--------------------|--------|
| 模型版本 | V2                 | $\sim$ |
| 训练方式 | 零代码开发-高级调参模式       |        |
| 训练算法 | ResNet-ResNet18_vd |        |

## 心 选择部署环境

部署环境分为服务器、通用小型设备、专项适配硬件,不同类别下对应不同的操作系统及芯片,选择模型在实际应用中部署的系统及芯片点击发 布即可完成模型发布

模型发布分为本地发布与云端发布两种方式,由于模型发布过程对发布环境的操作系统等环境有依赖,因此部分操作系统及芯片类型不支持 在本地发布,飞桨EasyDL为您提供云端环境完成模型发布过程,发布过程请确保本地设备网络链接

|   | 选择部署环境  | 服务器                       |     |
|---|---------|---------------------------|-----|
|   | 选择系统和芯片 | ○ 本地发布                    |     |
|   |         | - 🛆 Linux                 |     |
| ſ |         | 通用X86 CPU 英伟达GPU 华为 Atlas | 300 |
|   |         | 比特大陆SC计算卡                 |     |
|   |         | 选择系统和芯片不能为空               |     |

# 服务列表

服务列表主要用于管理已发布模型SDK,可查看全部已发布模型SDK。

模型SDK可在智能边缘控制台中可视化上完成部署流程,使用流程可参考智能边缘控制台使用文档。

| •••             | 服务   | く 〉 服务总览   |             |                 |                         |            |                                       |                        | 0 0 |
|-----------------|------|------------|-------------|-----------------|-------------------------|------------|---------------------------------------|------------------------|-----|
| EasyDL<br>235   | 服务列表 | 飞桨EasyDL推出 | 3智能边缘控制台,助力 | 力本地部署场景用户高效管理的  | *侧预测设备与服务, <sup>;</sup> | 极速完成本地数据与核 | <sup>夏</sup> 型串联,提高模型落地效 <sup>3</sup> | <mark>客。</mark> 了解详情 × |     |
| 总览              |      | 服务器通用      | 小型设备 专项道    | 适配硬件 ⑦          |                         |            |                                       | 输入模型名称                 | Q   |
| C<br>数据         |      | 模型名称       | 发布版本 📏      | 应用平台            | 发布状态                    | 发布方式       | 发布时间 🛟                                | 操作                     |     |
| <b>&lt;&gt;</b> |      | 猫狗分类       | 1–V1        | 英伟达GPU-Windows  | •已发布                    | 本地部署       | 2021-09-27 17:34                      | 导出SDK                  |     |
|                 |      | 害虫识别       | 5-V2        | 通用X86 CPU-Linux | •已发布                    | 本地部署       | 2021-11-22 20:51                      | 导出SDK                  |     |
| 模型              |      | 猫狗分类       | 10-V1       | 通用X86 CPU-Linux | •已发布                    | 本地部署       | 2021-12-07 20:33                      | 导出SDK                  |     |
| 服务              | •    | 猫狗识别       | 11-V1       | 通用X86 CPU-Linux | •已发布                    | 本地部署       | 2021-12-23 16:12                      | 导出SDK                  |     |
|                 |      |            |             |                 |                         |            |                                       |                        |     |

EasyEdge Intelligent Edge Console (以下简称IEC) 是EasyEdge推出的边缘设备管理的本地化方案。可以运行于多种架构、多系统、多类型的终端之上。通过IEC,用户可以方便地在本地进行

- EasyDL/BML/EasyEdge的SDK的 离线 / 在线激活,服务管理
- 接入本地和远程摄像头,网页中实时预览
- 自动监控和记录相关事件
- 硬件信息的可视化查看 操作系统支持: Linux、Windows

系统CPU架构支持:x86\_64、arm32、arm64

支持添加为服务的SDK版本包括:

- 通用CPU版 (基础版/加速版)
- 通用ARM版
- NVIDIA-GPU版 (基础版/加速版)
- 心 快速开始

根据本地设备类型下载智能边缘控制台

Linux系统-amd64:intel、AMD的64位CPU

Linux系统-arm:树莓派等32位arm CPU

Linux系统-arm64:rk3399、飞腾等 aarch64, 64位的arm CPU

Windows系统-amd64:intel、AMD的64位CPU

根据您设备的操作系统和硬件架构,将二进制和 etc 文件夹拷贝到任一目录,运行即可。(可以通过-cfg参数指定配置文件的路径,默认为./etc/easyedge-iec.yml)

如果是Windows系统,双击运行easyedge-iec.exe 即可如果是Linux系统,终端输入./easyedge-iec 即可运行

#### ./easyedge-iec-linux-amd64

Loading cfg from ./etc/easyedge-iec.yml 2021-09-14T14:49:20 INF0 [EasyEdgeIEC] Start stream service ... 2021-09-14T14:49:20 INF0 [\*EdgeStream] "2021-09-14 06:49:20,443 INFO [EasyEdge] 60439872 " 2021-09-14T14:49:20 WARN [\*EdgeStream] "2021-09-14 06:49:20,443 WARNING [EasyEdge] 139803591698176 EdgeStream is now serving at 127.0.0.1:24402" 2021-09-14T14:49:21 INF0 [EasyEdgeIEC] HLS server disable.... 2021-09-14T14:49:21 INF0 [EasyEdgeIEC] HTP-FLV listen On 0.0.0.0:8103 2021-09-14T14:49:21 WARN [EasyEdgeIEC] Baidu EasyEdge Intelligent Edge Console release 1.0.0, build 20210914 2021-09-14T14:49:21 INF0 [EasyEdgeIEC] RTMP Listen On 0.0.0.0:1935 2021-09-14T14:49:21 INF0 [EasyEdgeIEC] Webservice is now serving at 0.0.0.0:8702

启动之后,打开浏览器,访问 http://{设备ip}:8702/easyedge/iec 即可:



默认用户名密码为 admin / easyedge

| EasyEdge 智能边                           | 缘控制 | 台                                      |                                                                           |                                    |                     |
|----------------------------------------|-----|----------------------------------------|---------------------------------------------------------------------------|------------------------------------|---------------------|
| 总览                                     |     |                                        |                                                                           |                                    |                     |
| ⑦ 本地模型服务 添加模型服务                        |     | ┃基本信息                                  |                                                                           |                                    |                     |
| 模型服务管理                                 |     | 主机名 szth-aip-m12-tia<br>联网状态 <b>在线</b> | 操作系统         linux         MAC地           指纹文件         下载指纹文件         所属用 | 也址 6c:92:bf:46:82:47 IP地址<br>引户 —— | ≟ 10.232.191.23, 19 |
| ② 我的摄像头<br>添加摄像头                       |     | ┃ 资源利用情况                               |                                                                           |                                    |                     |
| 摄像头管理                                  |     | E 0%                                   | E7 09/                                                                    |                                    |                     |
| 本地事件<br>本地事件记录                         |     | 5.976<br>CPU利用率                        | 37.870<br>运行内存                                                            | 55.676<br>储存空间                     |                     |
| <ul> <li>③ 设置</li> <li>系统设置</li> </ul> |     | CPU型号 Intel(R) X<br>CPU核数 <b>48核</b>   | 运行内存 125.64GB<br>已用内存 72.60GB                                             | 总空间 16.39TB<br>已用空间 9.15TB         |                     |
| 运行日志                                   |     | 本地项目管理                                 |                                                                           |                                    |                     |
|                                        |     | 横型服务管理                                 | 序列号管理 >                                                                   | 摄像头管理                              | 本地事件记录              |
|                                        |     | <b>3 0</b><br>数量 运行中                   | <b>0 0</b><br>数量 有效期内                                                     | <b>2 2</b><br>数量 运行中               | 16742<br>⊯≣         |
|                                        |     |                                        |                                                                           |                                    |                     |

心 功能使用说明

# ① 添加模型服务

首先,点击导航栏的「本地模型服务」-「添加模型服务」。在页面中定义服务名称后,将已经下载好的Linux/Windows版本的SDK与IEC关联。

| EasyEdge 智能边缘                          | 控制 | 台      |         |                                         | 操作指南 | a admin ~ |
|----------------------------------------|----|--------|---------|-----------------------------------------|------|-----------|
| 总览                                     |    | 添加模型服务 |         |                                         |      |           |
| ② 本地模型服务                               |    | 新建模型服务 | ş       |                                         |      |           |
|                                        |    | *服务名称  |         |                                         |      |           |
| 模型服务管理                                 |    |        | 土 选择SDK |                                         |      |           |
| <ul> <li>              致的摄像头</li></ul> |    |        |         |                                         |      |           |
| 摄像头管理                                  |    |        |         |                                         |      |           |
| 🖾 本地事件                                 |    | 服务配置   |         |                                         |      |           |
| 本地事件记录                                 |    |        |         |                                         |      |           |
| ◎ 设置                                   |    |        | 参数配置    |                                         |      |           |
| 系统设置                                   |    |        |         |                                         |      |           |
| 运行日志                                   |    |        | 添加并激活   | l i i i i i i i i i i i i i i i i i i i |      |           |
|                                        |    |        |         |                                         |      |           |
|                                        |    |        |         |                                         |      |           |
|                                        |    |        |         |                                         |      |           |
|                                        |    |        |         |                                         |      |           |
|                                        |    |        |         |                                         |      |           |
| 版本号: V1.0.0                            |    |        |         |                                         |      |           |

配置服务,在服务端口不冲突占用的情况下,使用默认即可,点击【添加并激活】

激活完成后即可在「模型服务管理」列表中启动服务,使用后续的操作栏功能。

# 体验本地demo

点击「本地demo体验」即可在立即上传图片进行预测

【物体检测】97741 e7.5检测高性能–Auto分辨率V1



# 接入摄像头

使用接入摄像头功能首先需要添加摄像头,请参考第②步,完成后按照第③步操作

| EasyEdge 智能                             | 边缘控制 | 间台         |                         |                   |      |      | 操作指南                | a) admin ~ |
|-----------------------------------------|------|------------|-------------------------|-------------------|------|------|---------------------|------------|
| 总宽                                      |      | 模型服务管理     |                         |                   |      |      |                     |            |
| ⑦ 本地模型服务 添加模型服务                         |      | 新増服务 查看序列号 | <del>.</del>            |                   |      |      |                     |            |
|                                         |      | 服务名称       | 模型名称                    | 应用平台              | 服务状态 | 服务开关 | 操作                  |            |
| ◎ 我的摄像头                                 |      | 晓宇加速版      | e7.5检测高性能–Auto分<br>耕率V1 | Linux-通用X86 CPU加速 | 运行中  |      | 查看详情 体验本地demo 接入摄像头 | 删除         |
| 添加摄像头                                   |      | zyf_胡乱测试模型 | 测试20210902V1            | Linux-通用X86 CPU   | 运行中  |      |                     |            |
| 摄像头管理<br>二 本地事件<br>② 设置<br>系统设置<br>运行日志 |      |            |                         |                   |      |      |                     |            |
|                                         |      |            |                         |                   |      |      |                     |            |
| 版本号: V1.0.0                             |      |            |                         |                   |      |      |                     |            |

# ② 添加摄像头

导航栏点击「我的摄像头」-「添加摄像头」,定义摄像头名称、备注后即可添加摄像头。支持本地摄像头和网络摄像头。 摄像头添加成功后即 可设置摄像头的运行时间和频率

| EasyEdge 智能边缘控制                                                | 制台                                                                                                                                                                                                   | 操作指南 | a admin ~ |
|----------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|-----------|
| 总宽                                                             | 汤加播像头                                                                                                                                                                                                |      |           |
| <ul> <li>③ 本地模型服务 ^</li> <li>添加模型服务</li> <li>模型服务管理</li> </ul> | * <b>据像头名称</b> 描述入图像头名称<br>备注 描述入图像头指注                                                                                                                                                               |      |           |
| 受 我的摄像头                                                        | 摄像头类型 ◆ 本地摄音头 ○ 网络摄像头<br>清洁洋质像头 · ·                                                                                                                                                                  |      |           |
| 摄像头管理<br>二本地事件 ^<br>本地事件记录<br>② 设置 ^<br>系统设置                   | * 通行时间     00.00000 至 00.0000     + 新提→現     抽紙類率 9項 1 0 9 1 1 4     通紙類率 9項 1 0 9 1 1 4     温馨提示: 1. 通行时间和抽熱频率的设置曲服务主接册像头后,保存地理结果的运行方式,推进结果可在事件记录中重看。     2. 北处的设置不影响描像头实时预定服务指理,崩像头实时预定回在描像头管理组织直着 |      |           |
| <b>运行日志</b>                                                    |                                                                                                                                                                                                      |      |           |
| 版本号: V1.0.0                                                    |                                                                                                                                                                                                      |      |           |

# ③ 摄像头接入模型服务预测

点击「本地模型服务」-「模型服务管理」中,所需接入预测的服务的「接入摄像头」

| EasyEdge 智能边缘控制台 增长 |  |            |                         |                   |      |      | 操作指南              | ī 😑 admin 🗸 |
|---------------------|--|------------|-------------------------|-------------------|------|------|-------------------|-------------|
| 总宽                  |  | 模型服务管理     |                         |                   |      |      |                   |             |
| ③本地模型服务 ^           |  | 新增服务 查看序列号 |                         |                   |      |      |                   |             |
| 添加模型服务              |  |            |                         |                   |      |      |                   |             |
| 模型服务管理              |  | 服务名称       | 模型名称                    | 应用平台              | 服务状态 | 服务开关 | 操作                |             |
| ◎ 我的摄像头 へ           |  | 晓宇加速版      | e7.5检测高性能–Auto分<br>辨率V1 | Linux-通用X86 CPU加速 | 运行中  |      | 查看详情体验本地demo 接入摄像 | 头 <b>删除</b> |
| 添加摄像头               |  | zyf_胡乱测试模型 | 测试20210902V1            | Linux-通用X86 CPU   | 运行中  |      |                   |             |
| 摄像头管理               |  |            |                         |                   |      |      |                   |             |
| 🖹 本地事件 🔷 ^          |  |            |                         |                   |      |      |                   |             |

在弹出的弹窗中选择第②步中添加的摄像头,此时点击确认即可在「摄像头管理」中的实时预览功能中查看摄像头预测结果,识别结果默认不保存。如需保存识别结果,可设置对应的「本地事件触发条件」,根据标签和置信度,将识别结果保存至本地事件记录当中。设置多个标签条件时,IEC会以"或"的逻辑来将所有满足条件的识别结果保存

| 接入摄像头       |                                        |      |
|-------------|----------------------------------------|------|
| 某个ip-camera | ◇ 添加                                   |      |
| 摄像头名称       | 本地时间触发条件 52%                           | 操作   |
| 某个ip-camera | 标签 tomato → 置信度 - 52 % ~ 100% × + 新増一项 | 取消接入 |
|             | 确定取消                                   |      |

## ④ 本地事件

点击导航栏「本地事件记录」,可通过服务名称、摄像头名称、事件记录的时间、标签及置信度来筛选识别结果查看,多个标签及置信度同样也 是"或"的逻辑记录。如有想要删除的事件数据可选择后删除,全选为本页全选。



# 配置项

配置文件etc/easyedge-iec.yml中有关于IEC的各项配置说明,一般无需修改,请确保理解配置项含义之后,再做修改。

##### IEC系统配置 com: #硬件利用率刷新时间间隔: 过低的刷新间隔可能会消耗CPU资源。 hardwareUsageRefreshSecond: 60 #事件监测触发扫描周期 eventTriggerIntervalSecond: 10 # IEC保存SDK等资源的路径:可填写 "default", 也可以直接填写绝对路径 appDataFolder: default # 是否开启DEBUG模式:开启之后,将会打印大量日志,便于追踪系统问题 debug: yes # 是否需要用户名/密码登陆,强烈建议打开!(默认用户名/密码为 admin/easyedge) loginEnabled: yes #登录有效时间,单位秒 sessionMaxAge: 3600 logging: # 是否把日志打印到控制台 toStd: no # 是否把日志输出到文件。关闭后,将无法在页面中正确显示系统日志 toFile: yes loggingFile: ./log/easyedge-iec.log webservice: #WEB服务的监听端口 listenPort: 8702 listenHost: 0.0.0.0 sdk: # GPU SDK所使用的cuda版本: 9 / 10 / 10.2 / 11.0 / 11.1。请安装完cuda之后,这设置正确的版本号。 cudaVersion: 10.2 ##### ------- 以下高级配置一般无需修改 ---------##### !!!注意!!! 请确保理解配置项含义后再做修改 ##### 数据库相关配置 db: sqliteDbFile: ./etc/easyedge-iec.db eventDbFile: ./etc/easyedge-event.db ##### 推流相关配置 livego: listenHost: 0.0.0.0 hlsPort: 7002 apiPort: 8090 flvPort: 8101 rtmpPort: 1935 server: - appName: cameraPreview hls: false api: false flv: true ##### 视频流相关配置 edgestream: listenHost: 127.0.0.1 listenPort: 24402 #摄像头预览:识别结果绘制延迟消失 renderExtendFrames: 10 # 预测队列大小: fps=30时, 延迟约为2秒 inferenceQueueSize: 60 videoEncodeQueueSize: 20 videoEncodeBitRate: 400000

离线SDK部署说明 专项适配硬件离线部署 心 EdgeBoard (FZ) 专用SDK集成文档

# 简介

本文档介绍 EasyEdge/EasyDL在EdgeBoard®边缘计算盒/Lite计算卡上的专用软件的使用流程。

EdgeBoard系列硬件可直接应用于AI项目研发与部署,具有高性能、易携带、通用性强、开发简单等四大优点。

详细硬件参数请在AI市场浏览。

EdgeBoard产品使用手册:https://ai.baidu.com/ai-doc/HWCE/Yk3b86gvp

## 软核版本

| SDK版本  | 对应软核 |
|--------|------|
| 0.5.2+ | 1.4  |
| 0.5.7+ | 1.5+ |

SDK升级需配合EdgeBoard硬件软核升级,建议升级软核为SDK对应版本,否则可能出现结果错误或者其他异常。

可以通过dmesg | grep "DRIVER Version"命令获取EdgeBoard当前的软核版本

## **Release Notes**

SDK对应的软核说明: baidu\_easyedge\_linux\_cpp\_aarch64\_EdgeBoardFZ1.X\_gcc7.5\_v1.Y.Z\_20210813.tar.gz

#### 软核以及SDK更新情况如下表所示:

| 时间        | 版本 (1.Y.X) | 说明   | EdgeBoardFZ1.5对应的软核                        | EdgeBoardFZ1.4对应的软核                        |
|-----------|------------|------|--------------------------------------------|--------------------------------------------|
| 2021.8.23 | 1.0.0      | 第一版! | https://ai.baidu.com/ai-doc/HWCE/Ikqgcqt5x | https://ai.baidu.com/ai-doc/HWCE/Lkqiwlziw |

注意:升级完成相应的软核之后需要重启机器生效。

# 快速开始

开发者从EasyEdge/EasyDL下载的软件部署包中,包含了简单易用的SDK和Demo。只需简单的几个步骤,即可快速部署运行EdgeBoard计算盒。

# 部署包中包含多版本SDK:

- baidu easyedge\_linux\_cpp\_aarch64\_EdgeBoardFZ1.5\*:适用于EdgeBoard 1.5+软核
- baidu easyedge\_linux\_cpp\_aarch64\_EdgeBoardFZ1.4\*:适用于EdgeBoard 1.4+软核

SDK文件结构

| b        | aidu_easyedge_linux_cpp_aarch64_EdgeBoardFZ1.5_*                   |
|----------|--------------------------------------------------------------------|
| $\vdash$ | ReadMe.txt                                                         |
| $\vdash$ | —— bin                                                             |
|          | easyedge_image_inference                                           |
|          | easyedge_serving                                                   |
|          | easyedge_video_inference                                           |
| $\vdash$ | include                                                            |
|          | Leasyedge                                                          |
| $\vdash$ | —— lib                                                             |
|          | libeasyedge.so -> libeasyedge.so.1                                 |
|          | libeasyedge.so.1 -> libeasyedge.so.1.3.1                           |
|          | libeasyedge.so.1.3.1                                               |
|          | libeasyedge_static.a                                               |
|          | libeasyedge_videoio.so -> libeasyedge_videoio.so.1                 |
|          | libeasyedge_videoio.so.1 -> libeasyedge_videoio.so.1.3.1           |
|          | libeasyedge_videoio.so.1.3.1                                       |
|          | libeasyedge_videoio_static.a                                       |
|          | libpaddle_full_api_shared.so -> libpaddle_full_api_shared.so.1.8.0 |
|          | libpaddle_full_api_shared.so.1.8.0                                 |
|          | libverify.so -> libverify.so.1                                     |
|          | libverify.so.1 -> libverify.so.1.0.0                               |
|          | Libverify.so.1.0.0                                                 |
| $\vdash$ | now_sre.log                                                        |
| $\vdash$ | SIC                                                                |
|          | CMakeLists.txt                                                     |
|          | ├─── cmake                                                         |
|          | ├─── common                                                        |
|          | demo_image_inference                                               |
|          | Hemo_serving                                                       |
|          | Letter demo_video_inference                                        |
|          | thirdparty                                                         |
|          | opencv                                                             |

1.1.0+的SDK自带OpenCV, src编译的时候会引用thirdparty/opencv路径下的头文件和库文件。

# Demo使用流程

用户在AI市场购买计算盒之后,请参考以下步骤进行集成和试用。

#### 1. 将计算盒连接电源

指示灯亮起,等待约1分钟。

- 参考EdgeBoard使用文档配置网口或串口连接。登录EdgeBoard计算盒。
- 加载驱动 (开机加载一次即可)。

insmod /home/root/workspace/driver/{zu9|zu5|zu3}/fpgadrv.ko

根据购买的版本,选择合适的驱动。若未加载驱动,可能报错:

Failed to to fpga device: -1

• 设置系统时间 (系统时间必须正确)

date --set "2019-5-18 20:48:00"

## 2. (可选) 启动HTTP服务

部署包中附带了HTTP服务功能,开发者可以进入SDK根目录,运行easyedge\_serving程序启动HTTP服务。

```
###### ./easyedge_serving {RES目录} "" {绑定的host,默认0.0.0.0} {绑定的端口,默认24401}
cd ${SDK_ROOT}
export LD_LIBRARY_PATH=./lib
./demo/easyedge_serving ../../../RES ""
```

Baidu 百度智能云文档

日志显示

2019-07-18 13:27:05,941 INFO [EasyEdge] [http\_server.cpp:136] 547974369280 Serving at 0.0.0.0:24401

则启动成功。此时可直接在浏览器中输入http://{EdgeBoard计算盒ip地址}:24401/,在h5中测试模型效果。



同时,可以调用HTTP接口来访问盒子。具体参考下文接口说明。

EdgeBoard HTTP Server 目前使用的是单线程处理请求。

# 3. 编译运行Demo

编译:

```
cd src
mkdir build && cd build
cmake .. && make
```

运行

./easyedge\_image\_inference {RES资源文件夹路径} {测试图片路径}

便可看到识别结果。

## 使用说明

# 使用流程

激活成功之后,有效期内可离线使用。

- 1. 配置PaddleFluidConfig
- 2. 新建Predictor :global\_controller()->CreateEdgePredictor(config);

3. 初始化 predictor->init()

4. 传入图片开始识别predictor->infer(img, ...);

目前EdgeBoard暂不支持并行多模型计算。

## 接口说明

预测图片

/\*\*

- \* @brief 同步预测接口
- \* inference synchronous
- \* Supported by most chip and engine
- \* @param image: must be BGR , HWC format (opencv default)
- \* @param result
- \* @param threshold
- \* @return
- \*/

```
virtual int infer(
```

cv::Mat &image, std::vector<EdgeResultData> &result, float threshold = 0.1

) = **0**;

# 识别结果说明

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index; // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // object detection field
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
};
```

# 关于矩形坐标

- x1 \* 图片宽度 = 检测框的左上角的横坐标
- y1 \* 图片高度 = 检测框的左上角的纵坐标
- x2 \* 图片宽度 = 检测框的右下角的横坐标
- y2 \* 图片高度 = 检测框的右下角的纵坐标

可以参考demo文件中使用opencv绘制矩形的逻辑。

# HTTP 私有服务请求说明

#### http 请求参数

URL中的get参数:

| 参数        | 说明    |     | 默认值 |
|-----------|-------|-----|-----|
| threshold | 阈值过滤, | 0~1 | 0.1 |

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

Python

import requests

```
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
    'http://127.0.0.1:24401/',
    params={'threshold': 0.1},
    data=img).json()
```

Cpp label=C#

FileStream fs = new FileStream("./img.jpg", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close();

WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close(); response.Close();

Cpp label=C++ 需要安装curl

Baidu 百度智能云文档

```
#include <sys/stat.h>
 #include <curl/curl.h>
 #include <iostream>
 #include <string>
 #define S_ISREG(m) (((m) & 0170000) == (0100000))
 #define S_ISDIR(m) (((m) & 0170000) == (0040000))
 size_t write_callback(void *ptr, size_t size, size_t num, void *data) {
    std::string *str = dynamic_cast<std::string *>((std::string *)data);
    str->append((char *)ptr, size*num);
    return size*num;
 }
 int main(int argc, char *argv[]) {
    const char *post_data_filename = "./img.jpg";
    FILE *fp = NULL;
    std::string response;
    struct stat stbuf = { 0, };
    fp = fopen(post_data_filename, "rb");
    if (!fp) {
      fprintf(stderr, "Error: failed to open file "%s"
 ", post_data_filename);
      return -1;
    }
    if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
      fprintf(stderr, "Error: unknown file size "%s"
 ", post_data_filename);
      return -1;
    }
    CURL *curl;
    CURLcode res;
    curl_global_init(CURL_GLOBAL_ALL);
    curl = curl_easy_init();
    if (curl != NULL) {
      curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
      curl_easy_setopt(curl, CURLOPT_POST, 1L);
      curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
      curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
  curl_easy_setopt(curl, CURLOPT_WRITEFUNCTION, write_callback);
      curl_easy_setopt(curl, CURLOPT_WRITEDATA, &response);
      res = curl_easy_perform(curl);
      if (res != CURLE_OK) {
         fprintf(stderr, "curl_easy_perform() failed: %s
 ", curl_easy_strerror(res));
      }
      std::cout << response << std::endl; // response即为返回的json数据
      curl_easy_cleanup(curl);
    }
    curl_global_cleanup();
    fclose(fp);
    return 0;
 }
```

# Java请求示例

#### http 返回数据

| 字段         | 类型说明   | 其他                                 |
|------------|--------|------------------------------------|
| error_code | Number | 0为成功,非0参考message获得具体错误信息           |
| results    | Array  | 内容为具体的识别结果。其中字段的具体含义请参考预测图像-返回格式一节 |
| cost_ms    | Number | 预测耗时ms,不含网络交互时间                    |

| {                                                                                                                              |
|--------------------------------------------------------------------------------------------------------------------------------|
| "cost_ms": 52,                                                                                                                 |
| "error_code": 0,                                                                                                               |
| "results": [                                                                                                                   |
| {                                                                                                                              |
| "confidence": 0.94482421875,                                                                                                   |
| "index": 1,                                                                                                                    |
| "label": "IronMan",                                                                                                            |
| "x1": 0.059185408055782318,                                                                                                    |
| "x2": 0.18795496225357056,                                                                                                     |
| "y1": 0.14762254059314728,                                                                                                     |
| "y2": 0.52510076761245728                                                                                                      |
| }.                                                                                                                             |
| {                                                                                                                              |
| "confidence": 0.94091796875,                                                                                                   |
| "index": 1.                                                                                                                    |
| "label": "IronMan".                                                                                                            |
| "x1": 0.79151463508605957                                                                                                      |
| "x2": 0.92310667037963867.                                                                                                     |
| "v1": 0.045728668570518494.                                                                                                    |
| "v2": 0.42920106649398804                                                                                                      |
| 3                                                                                                                              |
| ,                                                                                                                              |
| 1                                                                                                                              |
| J                                                                                                                              |
| "x1": 0.79151463508605957,<br>"x2": 0.92310667037963867,<br>"y1": 0.045728668570518494,<br>"y2": 0.42920106649398804<br>}<br>] |

# 预测视频

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类 VideoDecoding,此类提供了获取视频帧数据的便利函数。通过 VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

● 接口

classVideoDecoding :

/\*\*

- \* @brief 获取输入源的下一帧
- \* @param frame\_tensor
- \* @return

\*/

virtual int next(FrameTensor &frame\_tensor) = 0;

/\*\*

- \* @brief 显示当前frame\_tensor中的视频帧
- \* @param frame\_tensor
- \* @return
- \*/

virtual int display(const FrameTensor &frame\_tensor) = 0;

/\*\*

\* @brief 将当前frame\_tensor中的视频帧写为本地视频文件

- \* @param frame\_tensor
- \* @return

virtual int save(FrameTensor &frame\_tensor) = 0;

```
/**
```

\*/

```
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
* @brief 获取视频的width属性
* @return
*/
virtual int get_width() = 0;
/**
* @brief 获取视频的height属性
```

```
* @return
*/
```

```
virtual int get_height() = 0;
```

struct VideoConfig

```
/**
* @brief 视频源、抽帧策略、存储策略的设置选项
*/
struct VideoConfig {
  SourceType source_type; // 输入源类型
  std::string source_value;   // 输入源地址,如视频文件路径、摄像头index、网络流地址
                      // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
 int skip_frames{0};
  int retrieve_all{false};
                     // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
is_needed置为false
                      // 在采取抽帧之前设置视频的fps
  int input_fps{0};
  Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
 bool enable_display{false};
  std::string window_name{"EasyEdge"};
  bool display_all{false};
                      // 是否显示所有frame,若为false,仅显示根据skip_frames抽取的frame
 bool enable_save{false};
  std::string save_path;
                        // frame存储为视频文件的路径
 bool save_all{false};
                       // 是否存储所有frame, 若为false, 仅存储根据skip_frames抽取的frame
  std::map::string, std::string> conf;
};
```

source\_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。source\_value:若source\_type为视频文件,该值为 指向视频文件的完整路径;若source\_type为摄像头,该值为摄像头的index,如对于/dev/videoO的摄像头,则index为0;若source\_type为网络 视频流,则为该视频流的完整地址。skip\_frames:设置跳帧,每隔skip\_frames帧抽取一帧,并把该抽取帧的is\_needed置为true,标记为 is\_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。retrieve\_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被 抽取返回,以作为显示或存储用。 input\_fps:用于抽帧前设置fps。 resolution:设置摄像头采样的分辨率,其值请参考easyedge\_video.h中的 定义,注意该分辨率调整仅对输入源为摄像头时有效。 conf:高级选项。部分配置会通过该map来设置。

#### 注意:

1. 如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2. 使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video\_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo\_video\_inference。

# 错误说明

SDK所有主动报出的错误,均覆盖在EdgeStatus枚举中。同时SDK会有详细的错误日志,开发者可以打开Debug日志查看额外说明:

```
EdgeLogConfig log_config;
log_config.enable_debug = true;
global_controller()->set_log_config(log_config);
```

## FAQ

#### 1. 如何处理一些 undefined reference?

如: undefined reference to `curl\_easy\_setopt@CURL\_OPENSSL\_3'

可以通过安装libcurl3 libcurl-openssl1.0-dev来解决。如果开发者想不想使用低版本的openssl(如Ubuntu 18.04),可以link静态 库easyedge\_static.a,自己指定需要的Library的版本。

示例:修改CMakeList.txt

find\_package(CURL REQUIRED) target\_link\_libraries(easyedge\_demo \${0penCV\_LIBS} easyedge\_static pthread \${CURL\_LIBRARIES} paddle-mobile)

2. error while loading shared libraries: libeasyedge.so.0.4.0: cannot open shared object file: No such file or directory

类似错误包括libpaddle-mobile.so找不到。

直接运行SDK自带的二进制可能会有这个问题,设置LD\_LIBRARY\_PATH为SDK部署包中的lib目录即可。开发者自行使用CMake编译的二进制可以 有效管理.so的依赖。

#### 3. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

 $headers = curl\_slist\_append(headers, "Expect:");$ 

### 4. 预测过程中报内存不足"Killed"

此问题仅出现在ZU5,因为FZ5A带vcu,给他预留的内存过大导致,如果用不到VCU可以把这部分改小。修改/run/media/mmcblk1p1/uEnv.txt:

ethaddr=00:0a:35:00:00:09 uenvcmd=fatload mmc 1 0x3000000 image.ub && bootm 0x3000000

bootargs=earlycon console=ttyPS0,115200 clk\_ignore\_unused cpuidle.off=1 root=/dev/mmcblk1p2 rw rootwait cma=128M

注意中间空行要保留。

如果购买的计算盒较早,驱动文件较旧,而SDK比较新(或SDK比较旧,但是计算盒较新),可能出现结果异常,如结果均为空或者 nan。请参考"软核版本"小节更新软核和驱动版本。

# 6. 编译过程报错file format not recognized

libeasyedge.so: file format not recognized; treating as linker script

下载的SDK zip包需要放到板子内部后,再解压、编译。

7. 提示 driver\_version(1.4.0) not match paddle\_lite\_version(1.5.1)

需更新驱动,否则可能导致结果异常。参考"软核版本"小节。

# 服务器离线部署

の Linux集成文档-C

## 简介

本文档介绍Linux CPP SDK的使用方法。

- 网络类型支持:图像分类,物体检测
- 硬件支持:
  - CPU 基础版: x86\_64
  - NVIDIA GPU: x86\_64 PC
- 操作系统支持: Linux

根据开发者的选择,实际下载的版本可能是以下版本之一:

- EasyDL图像
  - x86 CPU 基础版
  - Nvidia GPU 基础版

#### **Release Notes**

| 时间        | 版本    | 说明   |
|-----------|-------|------|
| 2021.8.23 | 1.0.0 | 第一版! |

# 快速开始

SDK在以下环境中测试通过

- x86\_64, Ubuntu 16.04, gcc 5.4
- x86\_64, Ubuntu 18.04, gcc 7.4
- Tesla P4, Ubuntu 16.04, cuda 9.0, cudnn 7.5
- x86\_64, Ubuntu 16.04, gcc 5.4, XTCL r1.0

依赖包括

- cmake 3+
- gcc 5.4 (需包含 GLIBCXX\_3.4.22) ,gcc / glibc版本请以实际SDK ldd的结果为准
- opencv3.4.11 (可选)
- cuda9.0\_cudnn7 (使用NVIDIA-GPU时必须)
- XTCL 1.0.0.187 (使用昆仑服务器时必须)
- 1. 安装依赖

以下步骤均可选,请开发者根据实际运行环境选择安装。

#### (可选) 安装cuda&cudnn

#### 在NVIDIA GPU上运行必须

对于GPU基础版,若开发者需求不同的依赖版本,请在PaddlePaddle官网下载对应版本的libpaddle\_fluid.so或参考其文档进行编译,覆盖lib文件 夹下的相关库文件。

#### (可选) 安装TensorRT

下载包中提供了对应 cuda9.0、cuda10.0、cuda10.2、cuda11.0和cuda11.1 五个版本的 SDK, cuda9.0 和 cuda10.0 的 SDK 默认依赖的 TensorRT 版本为 TensorRT7.0.0.11, cuda10.2 和 cuda11.0 的 SDK 默认依赖的 TensorRT 版本为 TensorRT7.1.3.4, cuda11.1 的 SDK 默认依赖的 TensorRT 版本为 TensorRT7.2.3.4,请在这里下载对应 cuda 版本的 TensorRT,并把其中的lib文件拷贝到系统lib目录,或其他目录并设置环 境变量。

#### (可选) 安装XTCL

请安装与1.0.0.187版本兼容的XTCL。必要时,请将运行库路径添加到环境变量。

#### 2. 测试Demo

模型资源文件默认已经打包在开发者下载的SDK包中。Demo工程直接编译即可运行。

请先将tar包整体拷贝到具体运行的设备中,再解压缩编译; 在Intel CPU上运行CPU加速版,如果thirdparty里包含openvino文件夹的,必须在 编译或运行demo程序前执行以下命令: source \${cpp\_kit位置路径}/thirdparty/openvino/bin/setupvars.sh

部分SDK中已经包含预先编译的二进制,如 bin/easyedge\_demo, bin/easyedge\_serving,配置LD\_LIBRARY\_PATH后,可直接运行: LD\_LIBRARY\_PATH=../lib ./bin/easyedge\_serving

#### 编译运行:

cd src mkdir build && cd build cmake .. && make ./easyedge\_image\_inference {模型RES文件夹} {测试图片路径} ###### 如果是NNIE引擎,使用sudo运行 sudo./easyedge\_image\_inference {模型RES文件夹} {测试图片路径}

如果希望SDK自动编译安装所需要的OpenCV库,修改cmake的optionEDGE\_BUILD\_OPENCV为ON即可。 SDK会自动从网络下载opencv源码,并 编译需要的module、链接。注意,此功能必须需联网。

cmake -DEDGE\_BUILD\_OPENCV=ON

若需自定义library search path或者gcc路径,修改CMakeList.txt即可。

demo运行效果:



> ./easyedge\_image\_inference ../../../RES 2.jpeg

2019-02-13 16:46:12,659 INFO [EasyEdge] [easyedge.cpp:34] 140606189016192 Baidu EasyEdge Linux Development Kit 0.2.1(20190213) 2019-02-13 16:46:14,083 INFO [EasyEdge] [paddlev2\_edge\_predictor.cpp:60] 140606189016192 Allocate graph success.

2019-02-13 16:46:14,326 DEBUG [EasyEdge] [paddlev2\_edge\_predictor.cpp:143] 140606189016192 Inference costs 168 ms 1, 1:txt\_frame, p:0.994905 loc: 0.168161, 0.153654, 0.920856, 0.779621

```
Done
```

编译demo完成之后,会同时生成一个http服务 运行

###### ./easyedge\_serving {res\_dir} {serial\_key} {host, default 0.0.0.0} {port, default 24401} ./easyedge\_serving ../../../RES "1111-1111-1111-1111" 0.0.0.0 24401

后,日志中会显示

HTTP is now serving at 0.0.0.0:24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,



同时,可以调用HTTP接口来访问服务,具体参考下文接口说明。

# 使用说明

使用该方式,将运行库嵌入到开发者的程序当中。

使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

// step 1: 配置模型资源目录
EdgePredictorConfig config;

```
config.model_dir = {模型文件目录};
  // step 2: 创建并初始化Predictor;在这里选择合适的引擎
 auto predictor = global_controller()->CreateEdgePredictor(config);
 // step 3-1: 预测图像
 auto img = cv::imread({图片路径});
 std::vector<EdgeResultData> results;
 predictor->infer(img, results);
// step 3-2: 预测视频
std::vector<EdgeResultData> results;
FrameTensor frame_tensor;
 VideoConfig video_config;
 video_config.source_type = static_cast<SourceType>(video_type); // source_type 定义参考头文件 easyedge_video.h
 video_config.source_value = video_src;
 /*
 ... more video_configs, 根据需要配置video_config的各选项
 */
 auto video_decoding = CreateVideoDecoding(video_config);
 while (video_decoding->next(frame_tensor) == EDGE_OK) {
    results.clear();
    if (frame_tensor.is_needed) {
      predictor->infer(frame_tensor.frame, results);
      render(frame_tensor.frame, results, predictor->model_info().kind);
    //video_decoding->display(frame_tensor); // 显示当前frame,需在video_config中开启配置
    //video_decoding->save(frame_tensor); // 存储当前frame到视频,需在video_config中开启配置
```

输入图片不限制大小

**SDK参数配置** SDK的参数通过EdgePredictorConfig::set\_config和global\_controller()->set\_config配置。set\_config的所有key 在easyedge\_xxxx\_config.h中。其中

- PREDICTOR前缀的key是不同模型相关的配置,通过EdgePredictorConfig::set\_config设置
- CONTROLLER前缀的key是整个SDK的全局配置,通过global\_controller()->set\_config设置

以序列号为例,KEY的说明如下:

```
/**
* @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号
* 值类型:string
* 默认值:空
*/
static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";
```

```
使用方法如下:
```

```
EdgePredictorConfig config;
config.model_dir = ...;
config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, "1DB7-1111-1111-D27D");
```

具体支持的运行参数配置列表可以参考开发工具包中的头文件的详细说明。

相关配置均可以通过环境变量的方法来设置,对应的key名称加上前缀EDGE\_即为环境变量的key。如序列号配置的环境变量key为EDGE\_PREDICTOR\_KEY\_SERIAL\_NUM,如指定CPU线程数的环境变量key为EDGE\_PREDICTOR\_KEY\_CPU\_THREADS\_NUM。注意:通过代码设置的配置会覆盖通过环境变量设置的值。

# Baidu 百度智能云文档

# 初始化

● 接口

auto predictor = global\_controller()->CreateEdgePredictor(config); predictor->init();

若返回非0,请查看输出日志排查错误原因。

# 预测图像

● 接口

/\*\* \* @brief \* 通用接口 \* @param image: must be BGR , HWC format (opencv default) \* @param result \* @return \*/ virtual int infer( cv::Mat& image, std::vector<EdgeResultData>& result ) = 0; /\*\* \* @brief \* 批量图片推理接口 \* @param image: must be BGR , HWC format (opencv default) \* @param result \* @return \*/ virtual int infer( std::vector<cv::Mat>& image, std::vector<std::vector<EdgeResultData>>& result

) = 0;

图片的格式务必为opencv默认的BGR, HWC格式。

# • 返回格式

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index: // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测、图像分割时才有意义
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割的模型,该字段才有意义
    // 请注意: 图像分割时,以下两个字段会比较大,使用完成之后请及时释放EdgeResultData
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding,游程编码的mask
    // 目标追踪模型,该字段才有意义
    int trackid; // 轨迹id
    int frame; // 处于视频中的第几帧
    EdgeTrackStat track_stat; // 跟踪状态
```

# 关于矩形坐标

};

x1 \* 图片宽度 = 检测框的左上角的横坐标

y1 \* 图片高度 = 检测框的左上角的纵坐标

x2 \* 图片宽度 = 检测框的右下角的横坐标

y2 \* 图片高度 = 检测框的右下角的纵坐标

## 关于图像分割mask

```
cv::Mat mask为图像掩码的二维数组
{
{0,0,0,0,0,0,0,0,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,0,0,0,0,0,0,0},
}
其中1代表为目标区域,0代表非目标区域
```

## 关于图像分割mask\_rle

该字段返回了mask的游程编码,解析方式可参考 http demo

以上字段可以参考demo文件中使用opencv绘制的逻辑进行解析

#### 日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

```
EdgeLogConfig log_config;
log_config.enable_debug = true;
global_controller()->set_log_config(log_config);
```

#### http服务

#### 1. 开启http服务 http服务的启动可以参考demo\_serving.cpp文件。

/\*\*

- \* @brief 开启一个简单的demo http服务。
- \* 该方法会block直到收到sigint/sigterm。
- \* http服务里,图片的解码运行在cpu之上,可能会降低推理速度。
- \* @tparam ConfigT
- \* @param config
- \* @param host
- \* @param port
- \* @param service\_id service\_id user parameter, uri '/get/service\_id' will respond this value with 'text/plain'
- \* @param instance\_num 实例数量,根据内存/显存/时延要求调整
- \* @return
- \*/

template<typename ConfigT>

#### 2. 请求http服务

开发者可以打开浏览器,http://{设备ip}:24401,选择图片或视频来进行测试。

# http 请求方式一:无额外编码

• 图片测试:不使用图片base64格式

URL中的get参数:

| 参数 说明 |           | 说明    |     | 默认值              |   |  |
|-------|-----------|-------|-----|------------------|---|--|
|       | threshold | 阈值过滤, | 0~1 | 如不提供,则会使用模型的推荐阈值 | 淔 |  |

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

# Baidu 百度智能云文档

# Python请求示例

import requests

```
with open('./1.jpg', 'rb') as f:
    img_data = f.read()
    result = requests.post(
        'http://127.0.0.1:24401/',
        params={'threshold': 0.1},
        data=img_data).json()
```

# Java请求示例

http 请求方法二:图片使用base64格式

HTTP方法:POST Header如下:

| 参数           | 值                |  |
|--------------|------------------|--|
| Content-Type | application/json |  |

# Body请求填写:

• 分类网络: body 中请求示例

```
{
"image": "<base64数据>"
"top_num": 5
}
```

# body中参数详情

| 参数          | 是否必<br>选 | 类型     | 可选值<br>范围 | 说明                                                                                         |
|-------------|----------|--------|-----------|--------------------------------------------------------------------------------------------|
| image       | 是        | string | -         | 图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px,<br>支持jpg/png/bmp格式 <b>注意去掉头部</b> |
| top_nu<br>m | 否        | number | -         | 返回分类数量,不填该参数,则默认返回全部分类结果                                                                   |

• 检测和分割网络: Body请求示例:

```
{
"image": "<base64数据>"
}
```

body中参数详情:

| 参数            | 是否必<br>选 | 类型     | 可选值<br>范围 | 说明                                                                                         |
|---------------|----------|--------|-----------|--------------------------------------------------------------------------------------------|
| image         | 是        | string | -         | 图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px,<br>支持jpg/png/bmp格式 <b>注意去掉头部</b> |
| thresho<br>Id | 否        | number | -         | 默认为推荐阈值,也可自行根据需要进行设置                                                                       |

# http 返回数据

| 字段         | 类型说明   | 其他                                 |
|------------|--------|------------------------------------|
| error_code | Number | 0为成功,非0参考message获得具体错误信息           |
| results    | Array  | 内容为具体的识别结果。其中字段的具体含义请参考预测图像·返回格式一节 |
| cost_ms    | Number | 预测耗时ms,不含网络交互时间                    |

返回示例

| {                            |  |  |  |
|------------------------------|--|--|--|
| "cost_ms": 52,               |  |  |  |
| "error_code": 0,             |  |  |  |
| "results": [                 |  |  |  |
| {                            |  |  |  |
| "confidence": 0.94482421875, |  |  |  |
| "index": 1,                  |  |  |  |
| "label": "IronMan",          |  |  |  |
| "x1": 0.059185408055782318,  |  |  |  |
| "x2": 0.18795496225357056,   |  |  |  |
| "y1": 0.14762254059314728,   |  |  |  |
| "y2": 0.52510076761245728    |  |  |  |
| },                           |  |  |  |
| {                            |  |  |  |
| "confidence": 0.94091796875, |  |  |  |
| "index": 1,                  |  |  |  |
| "label": "IronMan",          |  |  |  |
| "x1": 0.79151463508605957,   |  |  |  |
| "x2": 0.92310667037963867,   |  |  |  |
| "y1": 0.045728668570518494,  |  |  |  |
| "y2": 0.42920106649398804    |  |  |  |
| }                            |  |  |  |
| ]                            |  |  |  |
| }                            |  |  |  |

# 其他配置

# 1. 日志名称、HTTP 网页标题设置

通过global\_controller的set\_config方法设置:

global\_controller()->set\_config(easyedge::params::CONTROLLER\_KEY\_LOG\_BRAND, "MY\_BRAND");



# 2. CPU线程数设置

CPU线程数可通过 EdgePredictorConfig::set\_config 配置

EdgePredictorConfig config; config.set\_config(easyedge::params::PREDICTOR\_KEY\_CPU\_THREADS\_NUM, 4);

# 3. 批量预测设置

EdgePredictorConfig config; config.set\_config(params::PREDICTOR\_KEY\_VINO\_MAX\_BATCH\_SIZE, 4); //CPU加速版特殊设置

int batch\_size = 2; // 使用前修改batch\_size再编译、执行 while (get\_next\_batch(imgs, img\_files, batch\_size, start\_index)) {

}

CPU加速版设置PREDICTOR\_KEY\_VINO\_MAX\_BATCH\_SIZE含义:此值用来控制批量图片预测可以支持的最大图片数,实际预测的时候单次预测图 片数不可大于此值。实际预测图片数为std::min(PREDICTOR\_KEY\_VINO\_MAX\_BATCH\_SIZE, batch\_size)。

# FAQ

# 1. 如何处理一些 undefined reference?

如: undefined reference to `curl\_easy\_setopt@CURL\_OPENSSL\_3'

- 方案1:通过安装libcurl3 libcurl-openssl1.0-dev来解决。
- 方案2:如果开发者想不想使用低版本的openssl (如Ubuntu 18.04),可以link静态库easyedge\_static.a,自己指定需要的Library的版本:

示例:修改CMakeList.txt

find\_package(CURL REQUIRED) target\_link\_libraries(easyedge\_demo \${OpenCV\_LIBS} easyedge\_static pthread \${CURL\_LIBRARIES} verify\_static \$(其他需要的库))

其中,其他需要的库视具体sdk中包含的库而定。

## 2. NVIDIA GPU预测时,报错显存不足

如以下错误字样:

paddle.fluid.core.EnforceNotMet: Enforce failed. Expected allocating <= available, but received allocating:20998686233 > available:19587333888.

Insufficient GPU memory to allocation. at [/paddle/paddle/fluid/platform/gpu\_info.cc:170]

请根据显存大小和模型配置。调整合适的初始 fraction\_of\_gpu\_memory。参数的含义参考这里。

## 3. 如何将我的模型运行为一个http服务?

目前cpp sdk暂未集成http运行方式; 0.4.7版本之后,可以通过start\_http\_server方法开启http服务。

#### 4. 运行NNIE引擎报permission denied

日志显示:

open sys: Permission denied open err : Permission denied open err : Permission denied

请使用sudo在root下运行。

#### 5. 运行SDK报错 Authorization failed

情况一:日志显示 Http perform failed: null respond

在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP\_PROXY 的环境变量通过代理处理自己的网络请求。如

```
export HTTP_PROXY="http://192.168.1.100:8888"
./easyedge_demo ...
```

情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx)

此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除~/.baidu/easyedge 目录,再重新激活。

## 6. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl\_slist\_append(headers, "Expect:");

#### 7. 运行二进制时,提示 libverify.so cannot open shared object file

可能cmake没有正确设置rpath,可以设置LD\_LIBRARY\_PATH为sdk的lib文件夹后,再运行:

LD\_LIBRARY\_PATH=\$LD\_LIBRARY\_PATH:../lib ./easyedge\_demo

8. 运行二进制时提示 libopencv\_videoio.so.4.5: cannot open shared object file: No such file or directory

同上面8的问题类似,没有正确设置动态库的查找路径,可通过设置LD\_LIBRARY\_PATH为sdk的thirdparty/opencv/lib文件夹解决

export LD\_LIBRARY\_PATH=\${LD\_LIBRARY\_PATH}:../thirdparty/opencv/lib (tips: 上面冒号后面接的thirdparty/opencv/lib路径以实际项目中路径为准,比如也可能是../../thirdparty/opencv/lib)

#### 9. 编译时报错: file format not recognized

可能是因为在复制SDK时文件信息丢失。请将整个压缩包复制到目标设备中,再解压缩、编译

#### 10. 进行视频解码时,报错符号未找到、格式不支持、解析出的图片为空、无法设置抽帧

请确保安装OpenCV时,添加了-DWITH\_FFMPEG=ON选项(或者GStream选项),并且检查OpenCV的安装日志中,关于Video I/O段落的说明是否为YES。

| <br>Video I/O:      |                     |
|---------------------|---------------------|
| <br>DC1394:         | YES (ver 2.2.4)     |
| <br>FFMPEG:         | YES                 |
| <br>avcodec:        | YES (ver 56.60.100) |
| <br>avformat:       | YES (ver 56.40.101) |
| <br>avutil:         | YES (ver 54.31.100) |
| <br>swscale:        | YES (ver 3.1.101)   |
| <br>avresample:     | NO                  |
| <br>libv4l/libv4l2: | NO                  |
| <br>v4l/v4l2:       | linux/videodev2.h   |
|                     |                     |

如果为NO,请搜索相关解决方案,一般为依赖没有安装,以apt为例:

apt-get install yasm libjpeg-dev libjasper-dev libavcodec-dev libavformat-dev libswscale-dev libdc1394-22-dev libgstreamer0.10-dev libgstreamer0.10-dev libgstreamer-plugins-base0.10-dev libv4l-dev python-numpy libtbb-dev libqt4-dev libgtk2.0-dev libfaac-dev libmp3lame-dev libopencore-ammb-dev libopencore-ammb-dev libtheora-dev libtheora-dev libtvidcore-dev x264 v4l-utils ffmpeg

## 心 Linux集成文档-Python

简介

本文档介绍Linux Python SDK 的使用方法。

- 网络类型支持:图像分类,物体检测
- 硬件支持:
  - Linux x86\_64 CPU
  - Linux x86\_64 Nvidia GPU
- 语言支持: Python 3.5, 3.6, 3.7

#### **Release Notes**

| 时间        | 版本    | 说明   |
|-----------|-------|------|
| 2021.8.23 | 1.0.0 | 第一版! |

#### 快速开始

• 根据引擎的不同,SDK 依赖了不同的底层引擎。根据所需自行安装。

## 安装 paddlepaddle

• 使用x86\_64 CPU预测时必须安装:

pip3 install -U paddlepaddle

若 CPU 为特殊型号,如赛扬处理器(一般用于深度定制的硬件中),请关注 CPU 是否支持 avx 指令集。如果不支持,请在paddle官网安装 noavx 版本

• 使用NVIDIA GPU 预测时必须安装:

pip3 install -U paddlepaddle-gpu

如果环境非 cuda9 cudnn7,请参考paddle文档安装合适的 paddle 版本。 不被 paddle 支持的 cuda 和 cudnn 版本,EasyEdge 暂不支持

# 2. 安装 easyedge python wheel 包

pip3 install -U BaiduAI\_EasyEdge\_SDK-{版本号}-cp36-cp36m-linux\_x86\_64.whl

具体名称以 SDK 包中的 whl 为准。

## 3. 测试 Demo

### 图片预测

输入对应的模型文件夹(默认为RES)和测试图片路径,运行:

python3 demo.py {model\_dir} {image\_name.jpg}

测试效果:



图片加载失败

4. 测试Demo HTTP 服务 输入对应的模型文件夹(默认为RES)、序列号、设备ip和指定端口号,运行:

python3 demo\_serving.py {model\_dir} "" {host, default 0.0.0.0} {port, default 24401}

后,会显示:

Running on http://0.0.0.0:24401/

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片或者视频来进行测试。也可以参考`demo\_serving.py`里 http\_client\_test()函数请求http服务进行推理。



3

图片加载失败

# 使用说明

```
使用流程 demo.py
```

import BaiduAI.EasyEdge as edge

```
pred = edge.Program()
pred.init(model_dir={RES文件夹路径}, device=edge.Device.CPU, engine=edge.Engine.PADDLE_FLUID)
pred.infer_image({numpy.ndarray的图片})
pred.close()
```

demo\_serving.py

```
import BaiduAl.EasyEdge as edge
from BaiduAl.EasyEdge.serving import Serving
```

server = Serving(model\_dir={RES文件夹路径})
####### 请参考同级目录下demo.py里:
####### pred.init(model\_dir=xx, device=xx, engine=xx, device\_id=xx)
####### 对以下参数device\device\_id和engine进行修改
server.run(host=host, port=port, device=edge.Device.CPU, engine=edge.Engine.PADDLE\_FLUID)

# 初始化

● 接口

```
def init(self,
    model_dir,
    device=Device.LOCAL,
    engine=Engine.PADDLE_FLUID,
    config_file='conf.json',
    preprocess_file='preprocess_args.json',
    model_file='model',
    params_file='params',
    graph_file='graph.ncsmodel',
    label_file='label_list.txt',
    device_id=0
    ):
 ....
Args:
   device: Device.CPU
   engine: Engine.PADDLE_FLUID
   model_dir: str
    model dir
   preprocess_file: str
   model_file: str
   params_file: str
   graph_file: str
   label_file: str
   device_id: int
Raises:
   RuntimeError, IOError
Returns:
   bool: True if success
 ....
```

使用 NVIDIA GPU 预测时,必须满足:

- 机器已安装 cuda, cudnn
- 已正确安装对应 cuda 版本的 paddle 版本
- 通过设置环境变量FLAGS\_fraction\_of\_gpu\_memory\_to\_use设置合理的初始内存使用比例

使用 CPU 预测时,可以通过在 init 中设置 thread\_num 使用多线程预测。如:

 $pred.init (model\_dir=\_model\_dir,\ device=edge.Device.CPU,\ engine=edge.Engine.PADDLE\_FLUID,\ thread\_num=1)$ 

# 预测图像

• 接口

| def infer_image(self, img,                               |  |  |  |
|----------------------------------------------------------|--|--|--|
| threshold=0.3,                                           |  |  |  |
| channel_order='HWC',                                     |  |  |  |
| color_format= <mark>'BGR'</mark> ,                       |  |  |  |
| data_type='numpy'):                                      |  |  |  |
|                                                          |  |  |  |
|                                                          |  |  |  |
| Args:                                                    |  |  |  |
| img: np.ndarray or bytes                                 |  |  |  |
| threshold: float                                         |  |  |  |
| only return result with confidence larger than threshold |  |  |  |
| channel_order: string                                    |  |  |  |
| channel order HWC or CHW                                 |  |  |  |
| color_format: string                                     |  |  |  |
| color format order RGB or BGR                            |  |  |  |
| data_type: string                                        |  |  |  |
| image data type                                          |  |  |  |
|                                                          |  |  |  |
| Returns:                                                 |  |  |  |
| list                                                     |  |  |  |
|                                                          |  |  |  |
|                                                          |  |  |  |

• 返回格式: [dict1, dict2, ...]

| 字段         | 类型                   | 取值        | 说明                       |
|------------|----------------------|-----------|--------------------------|
| confidence | float                | 0~1       | 分类或检测的置信度                |
| label      | string               |           | 分类或检测的类别                 |
| index      | number               |           | 分类或检测的类别                 |
| x1, y1     | float                | 0~1       | 物体检测,矩形的左上角坐标 (相对长宽的比例值) |
| x2, y2     | float                | 0~1       | 物体检测,矩形的右下角坐标(相对长宽的比例值)  |
| mask       | string/numpy.ndarray | 图像分割的mask |                          |

# 关于矩形坐标

- x1 \* 图片宽度 = 检测框的左上角的横坐标
- y1 \* 图片高度 = 检测框的左上角的纵坐标
- x2 \* 图片宽度 = 检测框的右下角的横坐标
- y2 \* 图片高度 = 检测框的右下角的纵坐标

可以参考 demo 文件中使用 opencv 绘制矩形的逻辑。

# 结果示例

• i) 图像分类

```
{
    "index": 736,
    "label": "table",
    "confidence": 0.9
}
```

● ii) 物体检测

```
"y2": 0.91211,
"label": "cat",
"confidence": 1.0,
"x2": 0.91504,
"index": 8,
"y1": 0.12671,
"x1": 0.21289
}
```

# FAQ

# 1. 运行时报错 "非法指令" 或 " illegal instruction"

可能是 CPU 缺少 avx 指令集支持,请在paddle官网 下载 noavx 版本覆盖安装

#### 2. NVIDIA GPU预测时,报错显存不:

如以下错误字样:

```
paddle.fluid.core.EnforceNotMet: Enforce failed. Expected allocating <= available, but received allocating:20998686233 > available:19587333888.
Insufficient GPU memory to allocation. at [/paddle/paddle/fluid/platform/gpu_info.cc:170]
```

请在运行 Python 前设置环境变量,通过export FLAGS\_fraction\_of\_gpu\_memory\_to\_use=0.3来限制SDK初始使用的显存量,0.3表示初始使用 30%的显存。如果设置的初始显存较小,SDK 会自动尝试 allocate 更多的显存。

# 3. 我想使用多线程预测,怎么做?

如果需要多线程预测,可以每个线程启动一个Progam实例,进行预测。 demo.py文件中有相关示例代码。

注意: 对于CPU预测,SDK内部是可以使用多线程,最大化硬件利用率。参考init的thread\_num参数。

# 心 纯离线SDK简介

本文档主要说明定制化模型发布后获得的服务器端SDK如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单
- 进入EasyDL社区交流,与其他开发者进行互动

SDK说明

图像分类服务器端SDK支持Linux、Windows两种操作系统。以下为具体的系统、硬件环境支持:

| 操作系统    | 系统支持                 | 硬件环境要求                                                                                                                                                                                                                |
|---------|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Linux   |                      | Intel CPU: x86_64<br>NVIDIA GPU: x86_64                                                                                                                                                                               |
| Windows | 64位 Windows7 及以<br>上 | NVIDIA GPU: x86_64<br>环境依赖:<br>.NET Framework 4.5<br>Visual C++ Redistributable Packages for Visual Studio<br>2013<br>Visual C++ Redistributable Packages for Visual Studio<br>2015<br>GPU依赖:<br>CUDA 9.x + cuDNN 7.x |

#### 激活&使用步骤

1. 本地运行SDK,并完成首次联网激活

通过左侧导航栏查看不同操作系统SDK的开发文档

- 2. 正式使用
- SDK常见问题

通过左侧导航栏查看不同操作系统SDK的FAQ

#### 以下是通用FAQ,如您的问题仍未解决,请在百度智能云控制台内提交工单反馈

- 1、SDK如何激活?有效期是多少?
- 在免费试用期范围内,SDK有效期飞桨EasyDL-桌面版软件的试用期保持一致,通过飞桨EasyDL-高级版发布的SDK,可永久使用。

如有其他异常请在百度智能云控制台内提交工单反馈

<sup>の</sup>Windows集成文档

#### 简介

本文档介绍Windows GPU SDK的使用方法。

- 网络类型支持:图像分类,物体检测
- 硬件支持:
  - NVIDIA GPU
- 操作系统支持
  - 64位 Windows 7 及以上
- 环境依赖 (必须安装以下版本)
  - .NET Framework 4.5
  - Visual C++ Redistributable Packages for Visual Studio 2013
  - Visual C++ Redistributable Packages for Visual Studio 2015
- GPU 版依赖(必须安装以下版本) \* CUDA 9.0.x + cuDNN 7.6.x 或者 CUDA 10.0.x + cuDNN 7.6.x
- 协议
  - HTTP

Release Notes | 时间 | 版本 | 说明 | | ------- | ------ | ------ | | 2021.8.23 | 1.0.0 | 第一版! |

## 快速开始

### 1. 安装依赖

安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642

# Visual C++ Redistributable Packages for Visual Studio 2013

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

# Visual C++ Redistributable Packages for Visual Studio 2015

https://www.microsoft.com/zh-cn/download/details.aspx?id=48145

如果使用GPU版SDK,请安装CUDA + cuDNN

https://developer.nvidia.com/cuda https://developer.nvidia.com/cudnn

#### 注意事项

1. 安装目录不能包含中文

2. Windows Server 请自行开启,选择"我的电脑"——"属性"——"管理"——"添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启 即可。

## 2. 运行离线SDK

解压下载好的SDK,SDK默认使用cuda9版本,如果需要cuda10请运行EasyEdge CUDA10.0.bat切换到cuda10版本,



图片加载失败

点击"启动服务",等待数秒即可启动成功,本地服务默认运行在~~

http://127.0.0.1:24401/

其他任何语言只需通过HTTP调用即可。

```
如启动失败,可参考如下步骤排查:
```



# Demo示例(以图像服务为例)

服务运行成功,此时可直接在浏览器中输入http://127.0.0.1:24401,在h5中测试模型效果。



使用说明
# Baidu 百度智能云文档

# Python 使用示例代码如下

import requests

```
with open('./1.jpg', 'rb') as f:
img = f.read()
```

# C# 使用示例代码如下

FileStream fs = new FileStream("./img.jpg", FileMode.Open); BinaryReader br = new BinaryReader(fs); byte[] img = br.ReadBytes((int)fs.Length); br.Close(); fs.Close(); string url = "http://127.0.0.1:8402?threshold=0.1"; HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url); request.Method = "POST"; Stream stream = request.GetRequestStream(); stream.Write(img, 0, img.Length); stream.Close(); WebResponse response = request.GetResponse(); StreamReader sr = new StreamReader(response.GetResponseStream()); Console.WriteLine(sr.ReadToEnd()); sr.Close();

response.Close();

C++ 使用示例代码如下,需要安装curl

```
###### include <sys/stat.h>
###### include <curl/curl.h>
###### define S_ISREG(m) (((m) \& 0170000) == (0100000))
###### define S_ISDIR(m) (((m) \& 0170000) == (0040000))
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./img.jpg";
  FILE *fp = NULL;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  }
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\" \n", post_data_filename);
     return -1;
  }
  CURL *curl;
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
     curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
     curl_easy_setopt(curl, CURLOPT_POST, 1L);
     curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
     curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
     res = curl_easy_perform(curl);
     if (res != CURLE_OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     curl_easy_cleanup(curl);
  curl_global_cleanup();
  fclose(fp);
  return 0;
```

# 请求参数

| 字段        | 类型    | 取值  | 说明    |
|-----------|-------|-----|-------|
| threshold | float | 0~1 | 置信度阈值 |

HTTP POST Body直接发送图片二进制

**返回参数**|字段|类型|取值|说明||--------|-----|-----|-----||confidence|float|0~1|分类或检测的置信度||label|string||分 类或检测的类别||index|number||分类或检测的类别||x1,y1|float|0~1|物体检测,矩形的左上角坐标(相对长宽的比例值)||x2,y2|float |0~1|物体检测,矩形的右下角坐标(相对长宽的比例值)|

# 关于矩形坐标

- x1 \* 图片宽度 = 检测框的左上角的横坐标
- y1 \* 图片高度 = 检测框的左上角的纵坐标
- x2\*图片宽度 = 检测框的右下角的横坐标
- y2\*图片高度 = 检测框的右下角的纵坐标

FAQ

#### 1. 服务启动失败,怎么处理?

请确保相关依赖都安装正确,版本必须如下: .NET Framework 4.5 Visual C++ Redistributable Packages for Visual Studio 2013 \* Visual C++ Redistributable Packages for Visual Studio 2015

GPU依赖,版本必须如下: \* CUDA 9.0.x + cuDNN 7.6.x 或者 CUDA 10.0.x + cuDNN 7.6.x

GPU加速版(EasyEdge-win-x86-nvidia-gpu-tensorrt)依赖,版本必须如下: CUDA 9.0.x + cuDNN 7.6.x 或者 CUDA 10.0.x + cuDNN 7.6.x TensorRT 7.x 必须和CUDA版本对应

GPU加速版(EasyEdge-win-x86-nvidia-gpu-paddletrt)依赖,版本必须如下: CUDA 11.0.x + cuDNN 8.0.x TensorRT 7.1.3.4 必须和CUDA版本对应

#### 2. 服务调用时返回为空,怎么处理?

调用输入的图片必须是RGB格式,请确认是否有alpha通道。

#### 3. 多个模型怎么同时使用?

SDK设置运行不同的端口,点击运行即可。

# 4. JAVA、C#等其他语言怎么调用SDK?

参考 https://ai.baidu.com/forum/topic/show/943765

# 5. 启动失败,缺失DLL?

打开EasyEdge.log,查看日志错误,根据提示处理 缺失DLL,请使用 https://www.dependencywalker.com/ 查看相应模块依赖DLL缺失哪些,请 自行下载安装

### 6. 启动失败,报错NotDecrypted?

Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

#### 7. 其他问题

如果无法解决,可到论坛发帖: https://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

# 通用小型设备离线部署

# ு Linux集成文档-C++

简介

本文档介绍EasyEdge/EasyDL的Linux CPP SDK的使用方法。

- 网络类型支持: 图像分类 物体检测
- 硬件支持:
  - CPU: aarch64 armv7hf
- 操作系统支持:
  - Linux (Ubuntu, Centos, Debian等)
  - 海思HiLinux
  - 树莓派Raspbian/Debian
  - 瑞芯微Firefly

Release Notes | 时间 | 版本 | 说明 | | --- | ---- | | 2021.8.23 | 1.0.0 | 第一版 ! |

快速开始

- 安装依赖
- 依赖包括

#### Baidu 百度智能云文档

- cmake 3+
- gcc 5.4 以上(需包含 GLIBCXX\_3.4.22), gcc / glibc版本请以实际SDK ldd的结果为准
- opencv3.4.5 (可选)

#### 依赖说明:树莓派

树莓派Raspberry默认为armv7hf系统,使用SDK包中名称中包含 armv7hf\_ARM\_的tar包。如果是aarch64系统,使用SDK包中名称中包含 aarch64\_ARM\_的tar包。

在安装前可通过以下命令查看是32位还是64位 :

getconf LONG\_BIT 32

测试Demo

模型资源文件默认已经打包在开发者下载的SDK包中。

Demo工程直接编译即可运行。

请先将tar包整体拷贝到具体运行的设备中,再解压缩编译;

对于硬件使用为: -Intel Movidius MyRIAD2 / MyRIAD X on Linux x86\_64 / armv7hf / aarch64, 在编译或运行demo程序前执行以下命令: source \${cpp\_kit位置路径}/thirdparty/openvino/bin/setupvars.sh

部分SDK中已经包含预先编译的二进制, bin/easyedge\_demo, bin/easyedge\_serving, 配置LD\_LIBRARY\_PATH后, 可直接运行: LD\_LIBRARY\_PATH=../lib ./bin/easyedge\_serving

编译运行:

cd src mkdir build && cd build cmake .. && make ./easyedge\_image\_inference {模型RES文件夹} {测试图片路径}

如果希望SDK自动编译安装所需要的OpenCV库,修改cmake的optionEDGE\_BUILD\_OPENCV为ON即可。 SDK会自动从网络下载opencv源码,并 编译需要的module、链接。注意,此功能必须需联网。

cmake -DEDGE\_BUILD\_OPENCV=ON .. && make -j16

若需自定义library search path或者gcc路径,修改CMakeList.txt即可。

对于硬件使用为Intel Movidius MyRIAD2 / MyRIAD X 的,如果宿主机找不到神经计算棒Intel® Neural Compute Stick,需要执行以下命令添加 USB Rules:

cp \${cpp\_kit位置路径}/thirdparty/openvino/deployment\_tools/inference\_engine/external/97-myriad-usbboot.rules /etc/udev/rules.d/ sudo udevadm control --reload-rules sudo udevadm trigger sudo ldconfig

demo运行效果:



```
>./easyedge_image_inference ../../../RES 2.jpeg
2019-02-13 16:46:12,659 INFO [EasyEdge] [easyedge.cpp:34] 140606189016192 Baidu EasyEdge Linux Development Kit
0.2.1(20190213)
2019-02-13 16:46:14,083 INFO [EasyEdge] [paddlev2_edge_predictor.cpp:60] 140606189016192 Allocate graph success.
2019-02-13 16:46:14,326 DEBUG [EasyEdge] [paddlev2_edge_predictor.cpp:143] 140606189016192 Inference costs 168 ms
1, 1:txt_frame, p:0.994905 loc: 0.168161, 0.153654, 0.920856, 0.779621
Done
```

测试Demo HTTP 服务

编译demo完成之后,会同时生成一个http服务运行

###### ./easyedge\_serving {res\_dir} {serial\_key} {host, default 0.0.0.0} {port, default 24401} ./easyedge\_serving ../../../RES "1111-1111-1111" 0.0.00 24401

# 后,日志中会显示

HTTP is now serving at 0.0.0.0:24401

字样,此时,开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。



同时,可以调用HTTP接口来访问服务,具体参考下文接口说明。

# 使用说明

使用该方式,将运行库嵌入到开发者的程序当中。

# 使用流程

请优先参考Demo的使用流程。遇到错误,请优先参考文件中的注释解释,以及日志说明。

// step 1: 配置运行参数

EdgePredictorConfig config config.model\_dir = {模型文件目录}; // step 2: 创建并初始化Predictor;这这里选择合适的引擎 auto predictor = global\_controller()->CreateEdgePredictor(config); // step 3-1: 预测图像 auto img = cv::imread({图片路径}); std::vector<EdgeResultData> results; predictor->infer(img, results); // step 3-2: 预测视频 std::vector<EdgeResultData> results; FrameTensor frame\_tensor; VideoConfig video\_config; video\_config.source\_type = static\_cast<SourceType>(video\_type); // source\_type 定义参考头文件 easyedge\_video.h video\_config.source\_value = video\_src; /\* ... more video\_configs, 根据需要配置video\_config的各选项 \*/ auto video\_decoding = CreateVideoDecoding(video\_config); while (video\_decoding->next(frame\_tensor) == EDGE\_OK) { results.clear(); if (frame\_tensor.is\_needed) { predictor->infer(frame\_tensor.frame, results); render(frame\_tensor.frame, results, predictor->model\_info().kind); //video\_decoding->display(frame\_tensor); // 显示当前frame,需在video\_config中开启配置 //video\_decoding->save(frame\_tensor); // 存储当前frame到视频,需在video\_config中开启配置

对于口罩检测模型,将 EdgePredictorConfig config修改为PaddleMultiStageConfig config即可。

口罩检测模型请注意输入图片中人脸大小建议保持在 88到9696像素之间,可根据场景远近程度缩放图片后再传入SDK。

**SDK参数配置** SDK的参数通过EdgePredictorConfig::set\_config和global\_controller()->set\_config配置。set\_config的所有key 在easyedge\_xxxx\_config.h中。其中

- PREDICTOR前缀的key是不同模型相关的配置,通过EdgePredictorConfig::set\_config设置
- CONTROLLER前缀的key是整个SDK的全局配置,通过global\_controller()->set\_config设置

以序列号为例,KEY的说明如下:

```
/**
* @brief 序列号设置;序列号不设置留空时,SDK将会自动尝试使用本地已经激活成功的有效期内的序列号
* 值类型:string
* 默认值:空
*/
static constexpr auto PREDICTOR_KEY_SERIAL_NUM = "PREDICTOR_KEY_SERIAL_NUM";
```

```
使用方法如下:
```

```
EdgePredictorConfig config;
config.model_dir = ...;
config.set_config(params::PREDICTOR_KEY_SERIAL_NUM, "1DB7-1111-1111-D27D");
```

具体支持的运行参数可以参考开发工具包中的头文件的详细说明。

#### 初始化

接口

```
auto predictor = global_controller()->CreateEdgePredictor(config);
predictor->init();
```

若返回非0,请查看输出日志排查错误原因。

#### 预测图像

```
• 接口
```

图片的格式务必为opencv默认的BGR, HWC格式。

# • 返回格式

EdgeResultData中可以获取对应的分类信息、位置信息。

```
struct EdgeResultData {
    int index; // 分类结果的index
    std::string label; // 分类结果的label
    float prob; // 置信度
    // 物体检测活图像分割时才有
    float x1, y1, x2, y2; // (x1, y1): 左上角, (x2, y2): 右下角; 均为0~1的长宽比例值。
    // 图像分割时才有
    cv::Mat mask; // 0, 1 的mask
    std::string mask_rle; // Run Length Encoding, 游程编码的mask
};
```

## 关于矩形坐标

```
x1 * 图片宽度 = 检测框的左上角的横坐标
```

- y1 \* 图片高度 = 检测框的左上角的纵坐标
- x2 \* 图片宽度 = 检测框的右下角的横坐标
- y2 \* 图片高度 = 检测框的右下角的纵坐标

## 关于图像分割mask

```
cv::Mat mask为图像掩码的二维数组
{
{0,0,0,0,0,0,0,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,1,1,1,0,0,0,0},
{0,0,0,0,0,0,0,0,0,0},
}
其中1代表为目标区域,0代表非目标区域
```

#### 关于图像分割mask\_rle

该字段返回了mask的游程编码,解析方式可参考 http demo

# Baidu 百度智能云文档

# 预测视频

SDK 提供了支持摄像头读取、视频文件和网络视频流的解析工具类VideoDecoding,此类提供了获取视频帧数据的便利函数。通过VideoConfig结构体可以控制视频/摄像头的解析策略、抽帧策略、分辨率调整、结果视频存储等功能。对于抽取到的视频帧可以直接作为SDK infer 接口的参数进行预测。

● 接口

classVideoDecoding :

```
/**
* @brief 获取输入源的下一帧
* @param frame_tensor
* @return
*/
virtual int next(FrameTensor &frame_tensor) = 0;
/**
* @brief 显示当前frame_tensor中的视频帧
* @param frame_tensor
* @return
*/
virtual int display(const FrameTensor &frame_tensor) = 0;
/**
* @brief 将当前frame_tensor中的视频帧写为本地视频文件
* @param frame_tensor
* @return
*/
virtual int save(FrameTensor &frame_tensor) = 0;
/**
* @brief 获取视频的fps属性
* @return
*/
virtual int get_fps() = 0;
/**
 * @brief 获取视频的width属性
 * @return
*/
virtual int get_width() = 0;
/**
* @brief 获取视频的height属性
* @return
*/
```

virtual int get\_height() = 0;

struct VideoConfig

```
/**
  * @brief 视频源、抽帧策略、存储策略的设置选项
 */
 struct VideoConfig {
   SourceType source_type;
                          // 输入源类型
   std::string source_value;   // 输入源地址,如视频文件路径、摄像头index、网络流地址
   int skip_frames{0};
                       // 设置跳帧,每隔skip_frames帧抽取一帧,并把该抽取帧的is_needed置为true
   int retrieve_all{false};    // 是否抽取所有frame以便于作为显示和存储,对于不满足skip_frames策略的frame,把所抽取帧的
 is_needed置为false
   int input_fps{0};
                       // 在采取抽帧之前设置视频的fps
   Resolution resolution{Resolution::kAuto}; // 采样分辨率, 只对camera有效
   bool enable_display{false};
   std::string window_name{"EasyEdge"};
   bool display_all{false};
                       // 是否显示所有frame,若为false,仅显示根据skip_frames抽取的frame
   bool enable_save{false};
   std::string save_path;
                         // frame存储为视频文件的路径
   bool save_all{false};
                        // 是否存储所有frame, 若为false, 仅存储根据skip_frames抽取的frame
   std::map<std::string, std::string> conf;
 };
• source_type:输入源类型,支持视频文件、摄像头、网络视频流三种,值分别为1、2、3。
```

- source\_value: 若source\_type为视频文件,该值为指向视频文件的完整路径;若source\_type为摄像头,该值为摄像头的index,如对 于/dev/video0的摄像头,则index为0;若source\_type为网络视频流,则为该视频流的完整地址。
- skip\_frames:设置跳帧,每隔skip\_frames帧抽取一帧,并把该抽取帧的is\_needed置为true,标记为is\_needed的帧是用来做预测的帧。反之,直接跳过该帧,不经过预测。
- retrieve\_all:若置该项为true,则无论是否设置跳帧,所有的帧都会被抽取返回,以作为显示或存储用。
- input\_fps:用于抽帧前设置fps。
- resolution:设置摄像头采样的分辨率,其值请参考easyedge\_video.h中的定义,注意该分辨率调整仅对输入源为摄像头时有效。
- conf:高级选项。部分配置会通过该map来设置。

#### 注意

1.如果使用VideoConfig的display功能,需要自行编译带有GTK选项的opencv,默认打包的opencv不包含此项。

2.使用摄像头抽帧时,如果通过resolution设置了分辨率调整,但是不起作用,请添加如下选项:

video\_config.conf["backend"] = "2";

3.部分设备上的CSI摄像头尚未兼容,如遇到问题,可以通过工单、QQ交流群或微信交流群反馈。

具体接口调用流程,可以参考SDK中的demo\_video\_inference。

## 日志配置

设置 EdgeLogConfig 的相关参数。具体含义参考文件中的注释说明。

EdgeLogConfig log\_config; log\_config.enable\_debug = true; global\_controller()->set\_log\_config(log\_config);

#### http服务

1. 开启http服务 http服务的启动参考demo\_serving.cpp文件。

/\*\*

- \* @brief 开启一个简单的demo http服务。
- \* 该方法会block直到收到sigint/sigterm。
- \* http服务里,图片的解码运行在cpu之上,可能会降低推理速度。
- \* @tparam ConfigT
- \* @param config
- \* @param host
- \* @param port
- \* @param service\_id service\_id user parameter, uri '/get/service\_id' will respond this value with 'text/plain'
- \* @param instance\_num 实例数量,根据内存/显存/时延要求调整
- \* @return \*/ template<typename ConfigT>
- int start\_http\_server( const ConfigT &config, const std::string &host, int port,
  - const std::string &service\_id,
  - int instance\_num = 1);

# 2. 请求http服务

开发者可以打开浏览器,http://{设备ip}:24401,选择图片来进行测试。

#### http 请求方式一:不使用图片base64格式 URL中的get参数:

| 参数        | 说明    |     | 默认值   |             |  |  |  |
|-----------|-------|-----|-------|-------------|--|--|--|
| threshold | 阈值过滤, | 0~1 | 如不提供, | 则会使用模型的推荐阈值 |  |  |  |

HTTP POST Body即为图片的二进制内容(无需base64, 无需json)

Python请求示例

```
import requests
with open('./1.jpg', 'rb') as f:
    img = f.read()
    result = requests.post(
    'http://127.0.0.1:24401/',
    params=['threshold': 0.1},
    data=img).json()
```

## Java请求示例

http 请求方法二:使用图片base64格式 HTTP方法:POST Header如下:

| 参数           | 值                |
|--------------|------------------|
| Content-Type | application/json |

# Body请求填写:

• 分类网络: body 中请求示例

```
{
"image": "<base64数据>"
"top_num": 5
}
```

body中参数详情

| 参数          | 是否必<br>选 | 类型     | 可选值<br>范围 | 说明                                                                                         |
|-------------|----------|--------|-----------|--------------------------------------------------------------------------------------------|
| image       | 是        | string | -         | 图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px,<br>支持jpg/png/bmp格式 <b>注意去掉头部</b> |
| top_nu<br>m | 否        | number | -         | 返回分类数量,不填该参数,则默认返回全部分类结果                                                                   |

• 检测和分割网络: Body请求示例:

```
{
"image": "<base64数据>"
}
```

body中参数详情:

| 参数            | 是否必<br>选 | 类型     | 可选值<br>范围 | 说明                                                                                         |
|---------------|----------|--------|-----------|--------------------------------------------------------------------------------------------|
| image         | 是        | string | -         | 图像数据,base64编码,要求base64图片编码后大小不超过4M,最短边至少15px,最长边最大4096px,<br>支持jpg/png/bmp格式 <b>注意去掉头部</b> |
| thresho<br>Id | 否        | number | -         | 默认为推荐阈值,也可自行根据需要进行设置                                                                       |

# http 返回数据

| 字段         | 类型说明   | 其他                                 |
|------------|--------|------------------------------------|
| error_code | Number | 0为成功,非0参考message获得具体错误信息           |
| results    | Array  | 内容为具体的识别结果。其中字段的具体含义请参考预测图像-返回格式一节 |
| cost_ms    | Number | 预测耗时ms,不含网络交互时间                    |

返回示例

```
{
 "cost_ms": 52,
 "error_code": 0,
 "results": [
  {
      "confidence": 0.94482421875,
      "index": 1,
      "label": "IronMan",
      "x1": 0.059185408055782318,
      "x2": 0.18795496225357056,
      "y1": 0.14762254059314728,
      "y2": 0.52510076761245728
   },
   {
      "confidence": 0.94091796875,
      "index": 1,
      "label": "IronMan",
      "x1": 0.79151463508605957,
      "x2": 0.92310667037963867,
      "y1": 0.045728668570518494,
      "y2": 0.42920106649398804
```

# 其他配置

日志名称、HTTP 网页标题设置

global\_controller()->set\_config(easyedge::params::KEY\_LOG\_BRAND, "MY\_BRAND");

## 效果如下:



# FAQ

#### 1. 如何处理一些 undefined reference / error while loading shared libraries?

如:./easyedge\_demo: error while loading shared libraries: libeasyedge.so.1: cannot open shared object file: No such file or directory 这是因为二进制运行时Id无法找到依赖的库。如果是正确cmake && make 的程序,会自动处理好链接,一般不会出现此类问题。

遇到该问题时,请找到具体的库的位置,设置LD\_LIBRARY\_PATH。

示例一:libverify.so.1: cannot open shared object file: No such file or directory 链接找不到libveirfy.so文件,一般可通过 export LD\_LIBRARY\_PATH=\${LD\_LIBRARY\_PATH}:../../lib 解决(实际冒号后面添加的路径以libverify.so文件所在的路径为准)

示例二:libopencv\_videoio.so.4.5: cannot open shared object file: No such file or directory 链接找不到ibopencv\_videoio.so文件,一般可通 过 export LD\_LIBRARY\_PATH=\${LD\_LIBRARY\_PATH}:../../thirdparty/opencv/lib 解决(实际冒号后面添加的路径以libopencv\_videoio.so所在路 径为准)

#### 2. 如何将我的模型运行为一个http服务?

<del>目前cpp sdk暂未集成http运行方式;</del> 0.4.7版本之后,可以通过start\_http\_server方法开启http服务。

#### 3. 运行NNIE引擎报permission denied

#### 日志显示:

open sys: Permission denied open err : Permission denied open err : Permission denied

请使用sudo在root下运行。

#### 4. 运行SDK报错 Authorization failed

1. 情况一:日志显示 Http perform failed: null respond 在新的硬件上首次运行,必须联网激活。

SDK 能够接受 HTTP\_PROXY 的环境变量通过代理处理自己的网络请求。如

```
export HTTP_PROXY="http://192.168.1.100:8888"
./easyedge_demo ...
```

- 2. 情况二:日志显示failed to get/check device id(xxx)或者Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不 局限于)以下可能的情况:
- MAC地址变化
- 磁盘变更
- BIOS重刷

```
以及系统相关信息。
```

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除~/.baidu/easyedge 目录,再重新激活。

#### 5. 使用libcurl请求http服务时,速度明显变慢

这是因为libcurl请求continue导致server等待数据的问题,添加空的header即可

headers = curl\_slist\_append(headers, "Expect:");

#### 6. 运行NNIE引擎报错 std::bad\_alloc

检查开发板可用内存,一些比较大的网络占用内存较多,推荐内存500M以上

7. 运行二进制时,提示 libverify.so cannot open shared object file

可能cmake没有正确设置rpath,可以设置LD\_LIBRARY\_PATH为sdk的lib文件夹后,再运行:

LD\_LIBRARY\_PATH=\$LD\_LIBRARY\_PATH:../lib ./easyedge\_demo

#### 8. 编译时报错: file format not recognized

可能是因为在复制SDK时文件信息丢失。请将整个压缩包复制到目标设备中,再解压缩、编译

#### <sup>の</sup> 纯离线SDK简介

本文档主要说明定制化模型发布后获得的SDK如何使用,如有疑问可以通过以下方式联系我们:

- 在百度智能云控制台内提交工单
- 前往官方论坛交流,与其他开发者进行互动

# SDK说明

SDK支持iOS、Android、Linux、Windows四种操作系统。以下为具体的系统、硬件环境支持:

| 操作系统         | 系统支持                 | 硬件环境要求                                                                                  |
|--------------|----------------------|-----------------------------------------------------------------------------------------|
| Linux<br>C++ |                      | CPU: AArch64 ARMv7I                                                                     |
|              |                      | Intel CPU x86_64                                                                        |
| Windows      | 64位 Windows7 及以<br>└ | 小児祝秋 .<br>.NET Framework 4.5<br>Vigual C. L. Badistributable Baskages for Vigual Studio |
|              | ±                    | 2013                                                                                    |
|              |                      | Visual C++ Redistributable Packages for Visual Studio 2015                              |

#### 激活&使用SDK

SDK的激活与使用分以下步骤:

- 1. 本地运行SDK,并完成首次联网激活
- 2. 正式使用

#### の Windows集成文档

简介

本文档介绍Windows CPU SDK的使用方法。

- 网络类型支持:图像分类,物体检测
- 硬件支持:
  - Intel CPU \* x86\_64
- 操作系统支持

- 64位 Windows 7 及以上
- 环境依赖 (必须安装以下版本)
  - .NET Framework 4.5
  - Visual C++ Redistributable Packages for Visual Studio 2013
  - Visual C++ Redistributable Packages for Visual Studio 2015
- 协议
  - HTTP

Release Notes | 时间 | 版本 | 说明 | | ------- | ------ | ------ | | 2021.8.23 | 1.0.0 | 第一版 ! |

#### 快速开始

1. 安装依赖

# 安装.NET Framework4.5

https://www.microsoft.com/zh-CN/download/details.aspx?id=42642

#### Visual C++ Redistributable Packages for Visual Studio 2013

https://www.microsoft.com/zh-cn/download/details.aspx?id=40784

#### Visual C++ Redistributable Packages for Visual Studio 2015

https://www.microsoft.com/zh-cn/download/details.aspx?id=48145

# 注意事项

- 1. 安装目录不能包含中文
- 2. Windows Server 请自行开启,选择"我的电脑"——"属性"——"管理"——"添加角色和功能"——勾选"桌面体验",点击安装,安装之后重启 即可。

# 2. 运行离线SDK

解压下载好的SDK,打开EasyEdge.exe。

图片加载失败

点击"启动服务",等待数秒即可启动成功,本地服务默认运行在

http://127.0.0.1:24401/

其他任何语言只需通过HTTP调用即可。

| data          | 2020/4/15 23:51       | 文作 | <b>牛</b> 夹 |   |
|---------------|-----------------------|----|------------|---|
| AI 百度EasyEdge |                       | -  |            | × |
|               |                       |    |            |   |
| Model         | id : Easy             |    |            |   |
| Serial Nur    | and the second second |    | ]          |   |
| Host          | 127.0.0.1             |    | ]          |   |
| Port          | 24401                 |    |            |   |
| 服务状态          | 服务未启动                 |    |            |   |
|               | 是否开机自动启               | 动口 |            |   |
|               | 启动服务                  |    |            |   |
|               |                       |    |            |   |
|               |                       |    |            |   |

# 如启动失败,可参考如下步骤排查:

| •                 |      |
|-------------------|------|
| 🗘 百度智能云           | 故障排查 |
|                   |      |
| 零门槛AI开发           |      |
| 平台EasyDL          |      |
|                   | 2.   |
| WINDOWS离线SDK(桌面版友 |      |
| 布)●回题排查           |      |
|                   |      |

# Demo示例(以图像服务为例)

服务运行成功,此时可直接在浏览器中输入http://127.0.0.1:24401,在h5中测试模型效果。



使用说明

图像服务调用说明

Python 使用示例代码如下

import requests

with open('./1.jpg', 'rb') as f:

# C# 使用示例代码如下

```
FileStream fs = new FileStream("./img.jpg", FileMode.Open);
BinaryReader br = new BinaryReader(fs);
byte[] img = br.ReadBytes((int)fs.Length);
br.Close();
fs.Close();
string url = "http://127.0.0.1:8402?threshold=0.1";
HttpWebRequest request = (HttpWebRequest)HttpWebRequest.Create(url);
request.Method = "POST";
Stream stream = request.GetRequestStream();
stream.Write(img, 0, img.Length);
stream.Close();
WebResponse response = request.GetResponse();
StreamReader sr = new StreamReader(response.GetResponseStream());
```

Console.WriteLine(sr.ReadToEnd()); sr.Close(); response.Close();

C++ 使用示例代码如下,需要安装curl

```
###### include <sys/stat.h>
###### include <curl/curl.h>
###### include <iostream>
###### include <string>
###### define S_ISREG(m) (((m) \& 0170000) == (0100000))
###### define S_ISDIR(m) (((m) \& 0170000) == (0040000))
size_t write_callback(void *ptr, size_t size, size_t num, void *data) {
  std::string *str = dynamic_cast<std::string *>((std::string *)data);
  str->append((char *)ptr, size*num);
  return size*num:
int main(int argc, char *argv[]) {
  const char *post_data_filename = "./img.jpg";
  FILE *fp = NULL;
  std::string response;
  struct stat stbuf = { 0, };
  fp = fopen(post_data_filename, "rb");
  if (!fp) {
     fprintf(stderr, "Error: failed to open file \"%s\"\n", post_data_filename);
     return -1;
  if (fstat(fileno(fp), &stbuf) || !S_ISREG(stbuf.st_mode)) {
     fprintf(stderr, "Error: unknown file size \"%s\" \n", post_data_filename);
     return -1;
  CURL *curl
  CURLcode res;
  curl_global_init(CURL_GLOBAL_ALL);
  curl = curl_easy_init();
  if (curl != NULL) {
    curl_easy_setopt(curl, CURLOPT_URL, "http://127.0.0.1:24401?threshold=0.1");
    curl_easy_setopt(curl, CURLOPT_POST, 1L);
    curl_easy_setopt(curl, CURLOPT_POSTFIELDSIZE_LARGE,(curl_off_t)stbuf.st_size);
    curl_easy_setopt(curl, CURLOPT_READDATA, (void *)fp);
 curl_easy_setopt(curl, CURLOPT_WRITEFUNCTION, write_callback);
    curl_easy_setopt(curl, CURLOPT_WRITEDATA, &response);
    res = curl_easy_perform(curl);
     if (res != CURLE OK) {
       fprintf(stderr, "curl_easy_perform() failed: %s\n", curl_easy_strerror(res));
     std::cout << response << std::endl; // response即为返回的json数据
     curl_easy_cleanup(curl);
  curl_global_cleanup();
  fclose(fp);
  return 0;
```

结果 获取的结果存储在response字符串中。

### 请求参数

| 字段        | 类型    | 取值  | 说明    |
|-----------|-------|-----|-------|
| threshold | float | 0~1 | 置信度阈值 |

HTTP POST Body直接发送图片二进制。

**返回参数**|字段|类型|取值|说明||--------|----|-----|-----|| confidence | float | 0~1 | 分类或检测的置信度 || label | string || 分 类或检测的类别 || index | number || 分类或检测的类别 || x1, y1 | float | 0~1 | 物体检测,矩形的左上角坐标 (相对长宽的比例值) || x2, y2 | float | 0~1 | 物体检测,矩形的右下角坐标 (相对长宽的比例值) |

### 关于矩形坐标

y1 \* 图片高度 = 检测框的左上角的纵坐标

x2 \* 图片宽度 = 检测框的右下角的横坐标

y2 \* 图片高度 = 检测框的右下角的纵坐标

# FAQ

## 1. 服务启动失败,怎么处理?

请确保相关依赖都安装正确,版本必须如下: .NET Framework 4.5 Visual C++ Redistributable Packages for Visual Studio 2013 \* Visual C++ Redistributable Packages for Visual Studio 2015

## 2. 服务调用时返回为空,怎么处理?

调用输入的图片必须是RGB格式,请确认是否有alpha通道。

## 3. 多个模型怎么同时使用?

SDK设置运行不同的端口,点击运行即可。

## 4. JAVA、C#等其他语言怎么调用SDK?

参考 https://ai.baidu.com/forum/topic/show/943765

## 5. 启动失败,报错NotDecrypted?

Windows下使用,当前用户名不能为中文,否则无法正确加载模型。

#### 6. 启动失败,报错 SerialNum无效

日志显示failed to get/check device id(xxx)或者 Device fingerprint mismatch(xxx) 此类情况一般是设备指纹发生了变更,包括(但不局限于)以下可能的情况:

- mac 地址变化
- 磁盘变更
- bios重刷

以及系统相关信息。

遇到这类情况,请确保硬件无变更,如果想更换序列号,请先删除 C:\Users\\${用户名}\.baidu\easyedge 目录,再重新激活。

## 7. 其他问题

如果无法解决,可到论坛发帖: https://ai.baidu.com/forum/topic/list/199 描述使用遇到的问题,我们将及时回复您的问题。

# AutoDL模式算法适配硬件

### ☑ 图像分类

#### 服务器

| 物计类和 | Linux     |        |           |       |       |               |               |                 |                  |              | Windows |
|------|-----------|--------|-----------|-------|-------|---------------|---------------|-----------------|------------------|--------------|---------|
| 身法失望 | 通用x86 CPU | 英伟达GPU | 比特大陆SC计算卡 | 飞腾CPU | 海光DCU | 飞腾+华为Atlas300 | x86+寒武纪MLU270 | 飞腾+百度昆仑XPU-K200 | x86+百度昆仑XPU-K200 | x86+openvino | 英伟达GPU  |
| 超高性能 |           |        |           |       |       |               |               | <b>Ø</b>        | 0                |              | ۷       |
| 高性能  | ۷         |        | ×         |       | ×     | ×             |               |                 | ×                | 0            | 0       |
| 高精度  |           |        |           |       |       |               |               |                 | ۷                |              | ۷       |

# 通用小型设备

| 復計業刊 |       |           |                   |          |            |                  | Windows              |               |           |           |                 |
|------|-------|-----------|-------------------|----------|------------|------------------|----------------------|---------------|-----------|-----------|-----------------|
| 异本关型 | 通用ARM | 通用ARM GPU | 通用ARM+华为Atlas 310 | 华为海思NNIE | 比特大陆SE5计算盒 | 瑞芯微NPU RK3399Pro | 瑞芯微NPU RV1109/RV1126 | 瑞芯微NPU RK3588 | 飞腾+比特大陆SC | 通用x86 CPU | 英特尔Movidius VPU |
| 超高性能 |       |           | <                 | ×        | <b>Ø</b>   |                  | <b>Ø</b>             | 0             | ۷         | •         | <b>Ø</b>        |
| 高性能  |       |           |                   |          | ×          |                  | Ø                    | <b>v</b>      | ×         |           | ×               |
| 高精度  |       |           |                   | ×        |            |                  |                      |               |           |           |                 |

# 专项硬件设备

| 管计卷用 | Linux                   |  |  |  |
|------|-------------------------|--|--|--|
| 异広天空 | Jetson(TX2/Nano/Xavier) |  |  |  |
| 高性能  |                         |  |  |  |

# 心 物体检测

# 服务器

| 物计学到 |           | Linux  |           |       |       |              |               |                 |                 |              |        |
|------|-----------|--------|-----------|-------|-------|--------------|---------------|-----------------|-----------------|--------------|--------|
| 并位关业 | 通用x86 CPU | 英伟达GPU | 比特大陆SC计算卡 | 飞腾CPU | 海光DCU | 飞腾+寒武纪MLU270 | x86+寒武纪MLU270 | 飞腾+百度昆仑XPU-K200 | 飞腾+百度昆仑XPU-K200 | x86+openvino | 英伟达GPU |
| 超高性能 |           |        | •         |       |       |              |               | 0               |                 |              |        |
| 高性能  |           |        | ×         |       | ×     |              |               | 0               | ×               |              |        |
| 高精度  |           |        | ×         | ×     | ×     | ×            | ×             | ×               | ×               | ×            |        |

# 通用小型设备

| 信计采用 |       |           |                   |                 | Linux      | x                |                      |               |           |           | Windows         |
|------|-------|-----------|-------------------|-----------------|------------|------------------|----------------------|---------------|-----------|-----------|-----------------|
| 昇法奕型 | 通用ARM | 通用ARM GPU | 通用ARM+华为Atlas 310 | 通用ARM+寒武纪MLU220 | 比特大陆SE5计算盒 | 瑞芯微NPU RK3399Pro | 瑞芯微NPU RV1109/RV1126 | 瑞芯微NPU RK3588 | 飞腾+比特大陆SC | 通用x86 CPU | 英特尔Movidius VPU |
| 高性能  |       | •         |                   | <               |            | •                |                      | <b>Ø</b>      | 0         | •         | •               |
| 高精度  |       |           |                   | ×               | ×          |                  | 0                    | 0             | ×         |           | ×               |

# 专项适配硬件

| 第二十十五日 | Linux                   |
|--------|-------------------------|
| 昇法失望   | Jetson(TX2/Nano/Xavier) |
| 高性能    |                         |

# 心 实例分割

# 服务器

| 算法类型 | 通用x86 CPU | 英伟达GPU | 通用x86+ GPU |
|------|-----------|--------|------------|
| 默认   |           |        |            |

# 通用小型设备

| 管计类刑 | Windows |
|------|---------|
| 异広天空 | 通用x86   |
| 默认   |         |

# 心 语义分割

# 服务器

| 算法类型 | Linux     | Linux  | Windows    |
|------|-----------|--------|------------|
|      | 通用x86 CPU | 英伟达GPU | 通用x86+ GPU |
| 高性能  |           |        |            |
| 高精度  |           |        |            |

# 通用小型设备

| 管注米刑 | Windows |
|------|---------|
| 并位大王 | 通用x86   |
| 高精度  | 0       |
| 高性能  | 0       |

# 高级调参模式算法适配硬件

# 心 图像分类

| AT-1.41.00   | Linux   |            |         |           |               |               |                 |              |           |                   |                      | Windows       |           |            |
|--------------|---------|------------|---------|-----------|---------------|---------------|-----------------|--------------|-----------|-------------------|----------------------|---------------|-----------|------------|
| 并近失望         | 通用x86芯片 | x86+英伟达GPU | 通用ARM芯片 | 通用ARM GPU | x86+寒武纪MLU270 | 通用ARM+华为昇腾310 | 飞腾+百度昆仑XPU-K200 | 飞腾+寒武纪MLU270 | 比特大陆SC计算卡 | 瑞芯微NPU RK3399 Pro | 瑞芯微NPU RV1109/RV1126 | 現芯積NPU RK3588 | 英伟达Jetson | x86+英伟达GPU |
| MobileNet    | 8       | 0          | 0       | 0         | 0             | 0             | 0               | 0            | 0         | 0                 | 0                    | 0             | 0         | 0          |
| ResNet       | 8       | 8          | 8       | 8         | 0             | 0             | 8               | 0            | 0         | 8                 | 0                    | 8             | 0         | 0          |
| SENet        | 8       | 0          | 0       | 0         | 0             | 0             | 0               | 0            | 0         | 0                 | 0                    | 0             | 0         | 0          |
| EfficientNet | 8       | 8          | ۲       | 8         | ×             | ×             | ×               | ×            | 0         | 8                 | 0                    | 8             | ×         | 0          |
| PPLCNet      | 8       | 8          | 8       | 0         | ×             | ×             | ×               | ×            | ×         | 8                 | ×                    | ۲             | 0         | ø          |
| ResNeXt      | 8       | 0          | 0       | 0         | ×             | 0             | 0               | ×            | 0         | ×                 | ×                    | 8             | ×         | 0          |

# 心 物体检测

| M 11 (0.01  |         | Linux      |         |           |               |               |                 |              |           |                   |                      |               |           | Windows    |
|-------------|---------|------------|---------|-----------|---------------|---------------|-----------------|--------------|-----------|-------------------|----------------------|---------------|-----------|------------|
| 并活失定        | 通用x86芯片 | xB5+英伟达GPU | 通用ARM芯片 | 通用ARM GPU | x86+寒武纪MLU270 | 通用ARM+华为昇腾310 | 飞腾+百度昆仑XPU-K200 | 飞腾+寒武纪MLU270 | 比特大陆SC计算卡 | 瑞芯微NPU RK3399 Pro | 鴉芯鴉NPU RV1109/RV1126 | 期芯積NPU RK3588 | 英伟达Jetson | x86+英伟达GPU |
| CascadeRCNN | 8       | 0          | ۷       | ×         | ×             | ×             | ×               | ×            | ×         | ×                 | ×                    | ×             | ×         | ×          |
| FasterRCNN  | 0       | 0          | 0       | ×         | ×             | ×             | ×               | ×            | ×         | ×                 | ×                    | ×             | ×         | 0          |
| YOLOV3      | 8       | 0          | ۷       | 8         | 0             | Ø             | 0               | 0            | ×         | Ø                 | 0                    | 0             | •         | 0          |
| Picodet     | 8       | 0          | ۷       | 8         | ×             | ۲             | ۲               | ×            | ٥         | ×                 | ×                    | 0             | 0         | ۲          |

|   | 御計新期     | Linu    | IX         | Windows |            |  |
|---|----------|---------|------------|---------|------------|--|
|   | 昇広失空     | 通用x86芯片 | x86+英伟达GPU | 通用ARM芯片 | x86+英伟达GPU |  |
| ſ | MaskRCNN |         |            |         |            |  |

# 心 语义分割

| Adv 54, 32, 104 | Li      | nux        | Wir     | ndows      |
|-----------------|---------|------------|---------|------------|
| 具法类型            | 通用x86芯片 | x86+英伟达GPU | 通用ARM芯片 | x86+英伟达GPU |
| DeepLabV3       |         | 2          |         |            |
| BiSeNetV2       |         | 2          |         |            |
| DeepLabV3p      |         | 2          |         |            |
| OCRNet          |         |            |         |            |
| Fast-SCNN       |         |            |         |            |
| PP-LiteSeg      |         | ۷          |         |            |
| STDCSeg         | 8       | <b>Ø</b>   |         |            |

# 常见问题

#### 心 训练相关问题

心如果使用GPU环境进行训练?

如果您的计算机有NVIDIA® GPU,且需要使用GPU环境进行训练,请确保满足以下条件:

- Windows 10/11:需安装 CUDA 12.0 与 cuDNN v8.9.0
- Ubuntu 18.4/20.4:需 CUDA 12.0 与 cuDNN v8.9.0

您可参考NVIDIA官方文档了解CUDA和CUDNN的安装流程和配置方法,详见:

- CUDA 安装指南: https://docs.nvidia.com/cuda/
- cuDNN安装指南:https://docs.nvidia.com/deeplearning/cudnn/install-guide/index.html

#### 心训练失败怎么办?

若训练失败,训练任务列表中会展示失败原因及解决办法,如有其他问题请提交工单联系我们。

## 心 模型效果相关问题

#### 心如何提升模型效果?

一个模型很难一次性就训练到最佳的效果,可能需要结合模型评估报告和校验结果不断扩充数据和调优。为此我们设计了模型迭代功能,即当模型训练完毕后,会生成一个最新的版本号,首次V1、之后V2……以此类推。可以通过调整训练数据和算法,多次训练,获得更好的模型效果。

若想要提升模型效果,可以尝试以下两种方法:

#### 一、检查并优化训练数据

#### 图像分类

- 1. 检查是否存在训练数据过少的情况,建议每个类别的图片量不少于100个,如果低于这个量级建议扩充。
- 检查不同类别的数据量是否均衡,建议不同分类的数据量级相同,并尽量接近,如果有的类别数据量很高,有的类别数据量较低,会影响模型整体的识别效果。
- 3. 通过模型效果评估报告中的错误识别示例,有针对性地扩充训练数据。
- 检查测试模型的数据与训练数据的采集来源是否一致,如果设备不一致、或者采集的环境不一致,那么很可能会存在模型效果不错但实际测 试效果较差的情况。针对这种情况建议重新调整训练集,将训练数据与实际业务场景数据尽可能一致。

## 物体检测

- 1. 检查是否存在训练数据过少的情况,建议每个标签标注50个框以上,如果低于这个量级建议扩充。
- 检查不同标签的标注框数量是否均衡,建议不同标签的标注框数数据量级相同,并尽量接近,如果有的标签框数很多,有的标签框数很少, 会影响模型整体的识别效果。
- 3. 通过模型效果评估报告中的错误识别示例,有针对性地扩充训练数据。
- 4. 检查测试模型的数据与训练数据的采集来源是否一致,如果设备不一致、或者采集的环境不一致,那么很可能会存在模型效果不错但实际测

试效果较差的情况。针对这种情况建议重新调整训练集,将训练数据与实际业务场景数据尽可能一致。

## 实例分割

- 1. 检查是否存在训练数据过少的情况,建议每个标签标注50个目标以上,如果低于这个量级建议扩充。
- 2. 检查不同标签的标注目标数是否均衡,建议不同标签的标注目标数数据量级相同,并尽量接近,如果有的标签标注的很多,有的标签标注的 很少,会影响模型整体的识别效果。
- 3. 通过模型效果评估报告中的错误识别示例,有针对性地扩充训练数据。
- 检查测试模型的数据与训练数据的采集来源是否一致,如果设备不一致、或者采集的环境不一致,那么很可能会存在模型效果不错但实际测 试效果较差的情况。针对这种情况建议重新调整训练集,将训练数据与实际业务场景数据尽可能一致。

# 二、尝试不同的训练配置

可前往训练配置页面尝试不同的配置组合,因不同数据集在不同的算法上可能表现不一致,所以建议您多尝试不同的算法选型后综合挑选精度最 高的模型使用,你可以选择如下的配置项:

- 增量训练
- 自定义验证集
- 数据增强策略
- 在高级训练配置中增加输入图片分辨率

# 分享我的模型

# 分享EasyDL定制化模型

本文将会引导您将定制化模型售卖至AI市场中,具体的介绍与操作流程,您可以详细查看如下流程。

- 1. 售卖市场介绍
- 2. 模型售卖的准备工作
- 3. 模型售卖的具体流程

心 售卖市场介绍

您在EasyDL定制化训练出的模型可以在百度AI市场(aim.baidu.com)进行售卖。AI市场作为国内首家专注于服务AI产业链的商业平台,集结AI服务商所提供的AI软硬件、解决方案、数据服务、模型算法等产品与服务,为需方提供一站式AI采购平台,帮助供需双方在市场内建立精准的对接与交易通道。

如果您是企业,您可以获取模型销售收入,价格由您自主决定;如果您是个人,您也可通过发布0元模型商品获得社区积分奖励,商品购买一次 将获得1分,50积分即可兑换大礼(社区积分获取兑换详情)。

您将模型售卖至百度AI市场后,对您模型感兴趣的买家可以在市场中一键下单购买,并对模型进行再训练或是部署为服务集成在自己的应用中。 模型在AI市场上架、购买以及买家再训练,均不影响您对原始模型的操作,您可以放心售卖。

#### ⊙ 模型售卖的准备工作

您的模型在AI市场出售需要具备以下条件

- 满足AI市场服务商入驻标准:
- 1. 符合国家相关法律、法规规定
- 2. 完成百度云账号个人或企业实名认证
- 3. 接受《百度AI市场入驻协议》及相关协议及管理规范
- 定制化模型服务满足售卖要求:
- 1. 模型分类属于图像检测、图像分类或情感倾向分析才可予以售卖。并且图像分类、情感倾向分析类只可售卖高精度算法模型。
- 2. 模型最新训练时间在2020.09.25后,且距今不超过1年(模型商品超过训练时间1年,在AI市场将自动下架)。

3. 模型完全由您自主训练,基于已购模型或预置模型再次训练的模型不可售卖。

心 模型售卖的具体流程

您确认符合AI市场服务商入驻标准,且有符合售卖条件的模型,可按如下步骤完成售卖。期间有任何问题,可发邮件至Almarket@baidu.com咨询,我们会尽快与您联系,解答您的疑问,帮您顺利完成。 售卖具体流程如下:

◎ 1、入驻AI市场并成功开店

售卖模型需要在AI市场开店成为服务商,请您按照如下流程操作:

1. 选择身份:根据您的情况,选择以企业或个人身份入驻市场。注:一经选择后不可修改

2. 实名认证:请以模型训练的账号完成认证。企业认证需提交营业执照照片,个人认证需准备身份证照片,请提前准备。

3. 开通店铺:您需提交店铺logo、店铺简介、联系方式等基本信息。通常审核需要1-2个工作日,完成审核后可进入下一步。

心2、选择模型出售

您开店成功后,有两种方案选择模型予以出售。您可以根据您的需要,选择任一一种方案予以出售。

(1) 在EasyDL的AI市场-售卖模型中,选择可售模型列表中的模型,点击"出售模型"后将会跳转至AI市场的商品列表中,商品列表中将会出现一个 已经创建好的草稿,该草稿部分信息已自动填充完成,您需要点击编辑对其余信息予以补充。

| 模型中心         | 售卖模型                                 | 店铺详情                                                                       |
|--------------|--------------------------------------|----------------------------------------------------------------------------|
| 我的模型         | 可售模型 (1) 在售模型 (0)                    | 售卖模型收益是什么?                                                                 |
| 创建模型         |                                      | 您可将已完成训练的模型上架到AI市场(aim.baidu.com)<br>作为呢名商集责你的模型                           |
| 训练模型         | 候型选择: [wtb]英文和對書(ID: 34660) ♥ 「皆候型标准 | 如果您是企业,您可以获取模型销售收入,价格由您自主决<br>空                                            |
| 校验模型         | 部省方式: 公有云API V                       | 如果您是个人,您可通过发布0元模型商品获得社区积分奖<br>助。在早购买一次收益得1分。50和公明可兑换+21                    |
| 发布模型         | 训练版本: V5 🗸                           | 100, 阿田田秋天 八行大行力, 50%万百岁, 75% 71<br>社区积分获取兑换详情><br>多逗演金费暖采 有机会作为优质案例辞得专屋推广 |
| EasyData数据服务 | 出售模型 商品管理                            | * \$5687494493 D 10 10 44 17 20 0000 45 12 30 10 3 16 16 1                 |
| 教掘島皆         |                                      | 售卖模型前需要做哪些准备?                                                              |
| 7/84=24      |                                      | 告头很坚需要任AI印场开活成入版另例, 词总技照如下派性操作:<br>操作:                                     |
| 123300.2     |                                      | 1、这样身份:根据您的情况,这样以正业或个人身份入社中<br>场。注:一经选择后不可修改                               |
| 云服务数据回流      |                                      | 2、实名认证:请以模型训练的账号完成认证。企业认证需<br>提交营业执照照片,个人认证需准备身份证照片,请提前准                   |
| AI市场         |                                      | 备。<br>3、开通店铺:您需提交店铺logo、店铺简介、联系方式等                                         |
| 我的已购模型       |                                      | 基本信息。                                                                      |
| 售卖模型         |                                      |                                                                            |

(2)在AI市场服务商后台的商品列表中选择"商品发布",在商品分类中选择模型算法以及对应的模型分类,右侧出现从"从现有模型发布"按钮, 点击后选中您需要发布的模型。

| 返回商品列表 |                                                                       | 发布商品                   |
|--------|-----------------------------------------------------------------------|------------------------|
| ♥ 基本信息 |                                                                       |                        |
| 商品名称:  | 名称不可重复使用,要求与服务紧密相关,建议以公司简称作为前缀名,2                                     | 2-32个字                 |
| 商品分类:  | 模型算法 / 图像分类 🗸 从现有模型发布                                                 |                        |
|        | 快速发布                                                                  |                        |
| AI能力:  | 您的账号在EasyDL经典版中有如下模型可售卖,一键勾选,快速发布。<br>年内均可在AI市场售卖,到期将会自动下架,重训后即可重新发布。 | 。模型训练完成1               |
| 应用行业:  | 模型名称 部署方式 训练版本                                                        | 训练完成时间                 |
| 商品摘要:  | [wtb]美女和野<br>④ 鲁-公有云API ~ V5 ~<br>V5                                  | 2020-09-27<br>20:45:14 |
| 商品推荐语: | 发布                                                                    | 可添加3个<br>取消            |
|        |                                                                       |                        |

注意:如果您有多个模型但无可售模型,可能原因是您的模型最新训练时间在2020.9.25之前或是模型训练时间超过1年,建议您重新训练模型再予以售卖。

模型商品售卖需要填写四个部分信息:

 基本信息:部分信息已为您默认填充,您还需完成应用行业、商品摘要、商品推荐语、商品图片、产品亮点、产品说明、商品协议等内容予以 填写,图片内容越丰富越直观表达模型应用场景,越容易获得买家的青睐。其中商品协议为您与用户之间的协议,AI市场只提供协议模板,您 在此基础上修改完成上传PDF,买家需要同意后才可下单购买。

| R       | 发布商品                                                                                                                                     | 商品评博:     | 商品详情是用户了解商品; | 并产生购买意思的重要一环,请仔细请写 查看优秀题例                                                |                   |
|---------|------------------------------------------------------------------------------------------------------------------------------------------|-----------|--------------|--------------------------------------------------------------------------|-------------------|
| 基本信息    |                                                                                                                                          |           | 产品完成         | Aa 🖻                                                                     | ĸ"                |
| 商品名称    | Auto-放送一件-V1sik                                                                                                                          |           | 产品说明<br>应用场景 | 目前该模型支持使用EasyOL经典版平台进行交付,购买该模<br>示在你的EasyDL平台的"我的已购模型"中。在EasyDL经典规       | 型后, 将展<br>仮平台, 你可 |
| 商品分类:   | 日本日本 → JARA和学校名 →                                                                                                                        |           | 使用方式         | 、以一键式地将该模型部署成服务。你也可以收集、标注并上作<br>务场景中的数据,对该模型进行二次调优再训练。                   | 9自己真实业            |
|         | RALLEYや行う170.32月、第3日6日1917月。<br>日代产品 → 物理技術<br>第25頁 → 情報分析                                                                               |           | 其他           | EasyDL经典版平台使用方式可参看文档。<br>错损为https://ai.baidu.com/ai-doc/EASYDL/ok38n2w9h |                   |
| A(86,7) | は2月6 ↑ 第12月19 - 4-84,85(1)(F)(A(2)) (10) 9-1(2) 17 (10) 14(1) (10) 14(1) (10) 14(1) (10) (10) 14(1) (10) (10) (10) (10) (10) (10) (10) ( |           |              | 个 商品详情中, 『                                                               | 使用方式』的内容为默认       |
| 后用行业:   | (13.2) + (法法律承担的应用行会 (服务可选3个)                                                                                                            |           | 2即预见>        | 无需修改                                                                     |                   |
|         | 突然监控 ×                                                                                                                                   |           |              |                                                                          |                   |
| 商品编算:   | Int                                                                                                                                      | 相关文档(选项): | 请上传在销售和使用过程。 | 中用户经常用到的文档,包括但不限于:《使用说明》《规格书》(绘)                                         | 89.23 8           |
| 商品推荐语:  | 99 × 最多3个,用空物。enter运分 可适加3个                                                                                                              |           | + 2          | 2件大小: 不超过10Mb<br>2件格式: ppt, doc. pdf等                                    |                   |
|         | 北方容用于首其高品地二級法。该用10字以内的定语描述高品化的 倉積等例                                                                                                      |           |              |                                                                          |                   |
| 商品图片:   | 道上作清晰美能。主体突出的高品型片,避免型片有过多的文字。 <b>下包高品型片态向与残极</b> ,简称也更快制作出符合要求的型片。                                                                       | 商品协议:     |              | 皇着协议范本                                                                   |                   |
|         | 回川です: 720%-60%の正/<br>個月大小: 建立大手60%回小手10MD<br>回用電話: joep, seq<br>回用電話: joep, seq                                                          |           | 商品协议         |                                                                          |                   |
|         |                                                                                                                                          |           |              |                                                                          |                   |

模型信息:模型信息包括模型使用方式、数据标签信息、模型demo素材等信息,您需要重点关注数据标签信息与模型demo素材信息的填写。
 -数据标签:需要您准确说明每个标签的名称与含义,这将帮助买家理解模型用途并基于此进行模型再训练。

-模型demo素材:为了促进模型售卖,AI市场为当前上架的模型免费提供Demo试用服务。因此您需要为Demo提供5张示例图片与文本,方便

|           | ♥ 模型信息        |                                                                      |
|-----------|---------------|----------------------------------------------------------------------|
|           | 模型来源平台:       | EasyOL经典版                                                            |
|           | 模型使用方式:       | ◎ 部署为服务 ? 部署方式:设备编SDK ◎ 再调整 ? 训练平台: EasyDL投典版 □ 获取模型文件               |
|           | 模型数据集:        | ☑ 使用自有数据集训练 69                                                       |
|           | 模型标签数量:       | 2↑                                                                   |
|           | 增删数据标签:       | ● 支持 ○ 不支持                                                           |
|           | 数据标签详情:       | 请准确说明每个数据标签名称以及含义,该内容将会在AI市场公开,帮助浏览者寻找合适的模型商品。您可下载模板,填写对应标签的说明并上传。   |
| 头冢试用模型效果。 |               | 标签 标签说明                                                              |
|           |               | kiwi 0/30                                                            |
|           |               | tomato 0/30                                                          |
|           |               | + 上代称签信品 下载模板                                                        |
|           | ?<br>模型Demo素材 | 请您提供具有代表性约 <mark>5%</mark> 示例图片,帮助购买者快速了解模型功能与效果                     |
|           |               | 提示: 1、Demo示例会公开展现,请您不要提供敏感保密、或有版权风险的图片 2、图片大小10MB以内,支持jpg、jpeg、png格式 |

• **交付信息**:AI市场已为您默认填写好规格信息。对于企业服务商,您的规格信息包括:10天免费试用版与正式版,试用版价格为0,正式版您 可以自主定价;对于个人服务商,您只有一个正式版模型规格,价格为0,但模型一旦售卖将会获得社区积分,可兑换精美礼品。

| * Hate: * Hate:   107.20 gdt                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                           |                                              |         |                |           |  |  |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|----------------------------------------------|---------|----------------|-----------|--|--|
| 10天免費试。     正式版模型     10天免费试。     正式版模型       10天免費试。     10天免費试…     正式版模型       10天免費试。     10天免費试…     正式版模型       10天免費试。     10天免費试…     正式版模型       10天免費试。     10天免費试…     正式服型       10天免費试     10天免費试…     正式服型       10天免費试     10天免費试…     正式服型       10天免費试     10天免費试…     10天免費试…     10天免費试…       10天の算     10天免費试…     10天免費试…     10天免費试…       10日     10天免費试     10天免費试…     10天免費试…       10日     10天免費试     10天免費试…     10天免費试…       10日     10天免費试     10天免費试     10天免費       10日     10日     10日     10日                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | • 规格信息:                                                                   |                                              | • 规格信息: |                |           |  |  |
| 規磁部:     10天免費試用販暖型     規総部:     正式販売型       崩壊型文付目品は0天内有效、仅可支持模型再測体意看效果、不可部者为除务、到期后期<br>此文及。     兄弟の示:     日上之鸟       24000     兄が力記:     母辺分       24000     兄が力記:     日辺 つく)       文付知::     日辺 つく)     日辺 つく)       方     日辺 つく)     日辺 つく)       白田     日辺 つく)     日辺 つく)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 10天免费试                                                                    | 正式版模型                                        | 10天免费试  | 正式版模型          |           |  |  |
| B機型文付目起10天内有效、仅可支持模型再调线查看效果、不可部署为函务、到期后确<br>性实效、         公司方式:         住之马           44/50         24/50         24/50         24/50           大日方式:         成型文付         24/50         24/50           大日方式:         成型文付         24/50         24/50           大日方式:         成型大印         24/50         24/50           大日月         24/50         42/50         24/50           大日         24/                                                                                                                                                                                           | 规格名称:                                                                     | 10天免费试用版模型                                   | 規格名称:   | 正式版模型          |           |  |  |
| 자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련       자료하다     관련                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 规格说明:                                                                     | 自模型交付日起10天内有效,仅可支持模型再训练查看效果、不可部署为服务,到期后模型失效。 | 交易方式:   | 线上交易           |           |  |  |
| kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo       kip     Autoo     Autoo     Autoo <th></th> <th></th> <th>交付类型:</th> <th>模型交付</th> <th></th> <th></th>                                                                                                                                                                                                                                                                 |                                                                           |                                              | 交付类型:   | 模型交付           |           |  |  |
| A4500     定的方式:     使定分       交易方式:     加上交易     高品单位:     个       交付页型:     加型交付     一     一       交付时间:     回     元个     一       交付时间:     同时交付     石気閉:     1       定的方式:     修定分     石気閉:     1       成時方式:     大     日     1       定時方式:     修定分     日     日       自然中位:     ・     ・     1       自然中位:     ・     ・     ・       自然中位:     ・     ・ <th></th> <th></th> <th>交付时间:</th> <th>即时交付</th> <th></th> <th></th>                                                                                                                                                                                                                                                                                                                                                                                           |                                                                           |                                              | 交付时间:   | 即时交付           |           |  |  |
| 전망하기         월교오여         프로마스         프로마스 <th =<="" th=""><th></th><th>44/50</th><th>定价方式:</th><th>枚举定价</th><th></th><th></th></th> | <th></th> <th>44/50</th> <th>定价方式:</th> <th>枚举定价</th> <th></th> <th></th> |                                              | 44/50   | 定价方式:          | 枚举定价      |  |  |
| 交付类型:     模型交付     金根价格:     0     元/       交付时间:     即时交付     有效用:     1年       定价方式:     校型定价     人     人       商品单位:     个     人       金根价格:     免費     ・                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 交易方式:                                                                     | 线上交易                                         | 商品单位:   | $\uparrow$     |           |  |  |
| 한 방당 위         해당 위         적 있 했 · 1 年           定於 方式:         校 単定 위         · · · · · · · · · · · · · · · · · · ·                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 交付类型:                                                                     | 模型交付                                         | 含税价格:   | 0              | 元/个       |  |  |
| 定价方式:         校単定价              役宣展示原介,商品发布后的价格展示效果           商品单位:         个              役宣展示原介,商品发布后的价格展示效果            含税价格:         免费               10 王                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 交付时间:                                                                     | 即时交付                                         | 有效期:    | 1年             |           |  |  |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 定价方式:                                                                     | 校举定价                                         |         |                | *~**      |  |  |
| 会税价格: 免费<br>本応照: 10平                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 商品单位:                                                                     | $\uparrow$                                   | 设重展示    | <b>示原价,商品发</b> | 巾后的价格展示效果 |  |  |
| 本位明・ 10王                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 含税价格:                                                                     | 免费                                           |         |                |           |  |  |
| 712007                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 有效期:                                                                      | 10天                                          |         |                |           |  |  |

售后信息:售后信息包括发票与服务保障。对于企业服务商,您需要为买家提供增值税普票或专票;对于个人服务商,由于商品为零元,您不需要提供发票。服务保障说明已默认为您填写好,无需修改。

| <ul> <li>原服务:</li> <li>● 仅支持开具增值税专用发票(可抵扣)</li> <li>● 仅支持开具增值税普通发票(不可抵扣)</li> <li>● 前两项都支持</li> <li>● 原保期 请输入质保期</li> <li>年 &lt;</li> <li>● 无质保</li> <li>● 终身质保</li> <li>● 终身质保</li> <li>● 就明:</li> <li>8.</li> <li>1. 模型类商品不支持退款,请购买前充分了解模型效果。</li> <li>2. 10天免费试用版在购买成功时生效,生效期内可再训练或将模型部署为服务。</li> <li>3. 正式版模型购买后可在EasyOL经典版平台训练与部署。模型有效期为一年,一年内您可以对购买原始模型进行再训练(但不保证EasyOL新功能的可用性)。有效期结束后,您已部署的服务不会强制下线,您可持续使用。</li> <li>4. 模型购买后不会自动升级或更新,如需使用升级版模型,请重新购买.</li> </ul> | 售后服务信         | 息                                                                                                                                            |                                                                                                          |                       |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|----------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|-----------------------|
| <ul> <li>B保闢: ● 质保期 请输入质保期 年 ● 元质保 ● 各身质保</li> <li>B说明:</li> <li>B说明:</li> <li>1、模型类商品不支持退款,请购买前充分了解模型效果。</li> <li>2、10天免费试用版在购买成功时生效,生效期内可再训练或将模型部署为服务。</li> <li>3、正式版模型购买后可在EasyOL经典版平台训练与部署。模型有效明为一年,一年内您可以对购买原始模型进行再训练(但不保证EasyDL新功能的可用性)。有效期结束后,您已部署的服务不会强制下线,您可持续使用。</li> <li>4、模型购买后不会自动升级或更新,如需使用升级版模型,请重新购买.</li> </ul>                                                                                                                                          | 雲服务:          | ① 仅支持开具增值税专用发票(可抵扣)                                                                                                                          | ○ 仅支持开具增值税普通发票(不可抵扣)                                                                                     | ○ 前两项都支持              |
| <ul> <li>a 说明:</li> <li>1、模型类商品不支持退款,请购买前充分了解模型效果。</li> <li>2、10天免费试用版在购买成功时生效,生效期内可再训练或将模型部署为服务。</li> <li>3、正式版模型购买后可在EasyDL经典版平台训练与部署。模型有效期为一年,一年内您可以对购买原始模型进行再训练(但不保证EasyDL新功能的可用性)。有效期结束后,您已部署的服务不会强制下线,您可持续使用。</li> <li>4、模型购买后不会自动升级或更新,如需使用升级版模型,请重新购买.</li> </ul>                                                                                                                                                                                                 | 务保障:          | ○ 质保期 请输入质保期                                                                                                                                 | 年 🗸 ● 无质保  ○ 终身质保                                                                                        |                       |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | !障说明:<br>(选填) | <ol> <li>模型类商品不支持退款,请购买前</li> <li>10天免费试用版在购买成功时生效</li> <li>正式版模型购买后可在EasyDL经复型进行再训练(但不保证EasyDL新功持续使用。</li> <li>4、模型购买后不会自动升级或更新,</li> </ol> | 充分了解模型效果。<br>、生效期内可再训练或将模型部署为服务。<br>电版平台训练与部署。模型有效期为一年,一年内<br>能的可用性)。有效期结束后,您已部署的服务不<br>如需使用升级版模型,请重新购买. | 您可以对购买原始模<br>会强制下线,您可 |

商品信息填写完整后即可提交审核。提交后,将会由专人进行质量审核,约一个工作日给您结果反馈,

#### Ω 4、售卖并获得收入

您的模型服务审核通过并成功上架后,AI市场的买家将可在市场中购买您的模型服务。成功付款后,收入将由AI市场代收,您可在月底发起结算 申请,AI市场将销售收入打到您的账户中。

# 分享EasyDL定制化API

本文将会引导您将定制化API服务售卖至AI市场中,从而快速获得品牌曝光,并获得丰厚收入。接下来,您将了解如下内容:

- 1. 售卖市场介绍
- 2. API服务售卖的准备工作
- 3. API服务售卖的具体流程

#### ✤ 售卖市场介绍

您在EasyDL定制化训练出的模型可以在百度Al市场 (aim.baidu.com) 进行售卖。Al市场作为国内首家专注于服务Al产业链的商业平台,集结Al服务 商所提供的Al软硬件、解决方案、数据服务、模型算法等产品与服务,为需方提供一站式Al采购平台,帮助供需双方在市场内建立精准的对接与 交易通道。

您将训练完成的API售卖至AI市场中,不仅会获得大量浏览用户关注,增强您企业的品牌曝光。同时,您所训练的服务还将发挥巨大的商业价值, 对您服务感兴趣的买家可以在市场中一键下单购买,并集成到业务中使用,您也将获得相应的收入。 您的服务在AI市场出售定制化API服务需要具备以下条件

- 满足AI市场服务商入驻标准:
- 1. 符合国家相关法律、法规规定,拥有正规的公司资质
- 2. 完成百度云账号企业实名认证
- 3. 接受《百度AI市场入驻协议》及相关协议及管理规范
- 定制化API服务满足售卖要求:
- 1. 定制化识别接口至少被成功调用1次及以上
- 2. 图像分类/声音分类模型准确率>80%, 且总训练数据量>50
- 3. 物体检测模型mAP>60%,且总训练数据量>20
- 心 API服务售卖的具体流程

您确认您的企业/单位符合AI市场服务商入驻标准,且有符合售卖条件的API服务,可按如下步骤完成售卖。期间有任何问题,可发邮件至 Almarket@baidu.com咨询,我们会尽快与您联系,解答您的疑问,帮您顺利完成。售卖具体流程如下:

心1、入驻AI市场并成功开店

售卖模型需要在AI市场开店成为服务商,请您按照如下流程操作:

- 1. 实名认证:请以EasyDL训练者的账号完成企业实名认证。企业认证需提交营业执照照片,请提前予以准备。
- 2. 开通店铺:您需提交店铺logo、店铺简介、联系方式等基本信息。通常审核需要1-2个工作日,完成审核后可进入下一步。
- の2、选择API出售

您开店成功后,有两种方案选择API予以出售。您可以根据您的需要,选择任一一种方案予以出售。

(1) 在EasyDL控制台-公有云服务-售卖服务中,选择可售API列表中的API,点击"发布商品"后将会跳转至AI市场商品发布页面,部分API信息将会 予以填充。

| ✔ 返回EasyDL总览                                                                              | 产品服务 / EasyDL定制训练                                       | §平台 - 售卖已发布服务                                                                       |                                  |                       |   |                   |          |             |            |
|-------------------------------------------------------------------------------------------|---------------------------------------------------------|-------------------------------------------------------------------------------------|----------------------------------|-----------------------|---|-------------------|----------|-------------|------------|
| 经典版                                                                                       | 售卖已发布服务                                                 |                                                                                     |                                  |                       |   |                   |          |             |            |
| <ul> <li>公有云服务</li> <li>。 应用列表</li> <li>• 权限管理</li> <li>• 用量统计</li> <li>• 监控报表</li> </ul> | 已发布服务售卖条件<br>售卖位置: 棵型发<br>售卖条件: • 定制/<br>• 图像夕<br>• 物体科 | 及售卖流程说明 收起教程<br>布成功后,将已上线的定制<br>比识别接口成功调用次数≥1;<br>}发/声音分类模型准确率>8<br>逾测模型mAP>60%,总训练 | 化识别接口发布至<br>0%,总训练数据量<br>数据量>20; | 百度AI市场 进行售卖。<br>1>50; |   |                   |          |             |            |
| <ul> <li>技术文档</li> <li>售卖服务</li> <li>私有服务器部署服务</li> <li>通用设备端服务</li> </ul>                | <ul> <li>・文本5<br/>售卖流程:</li> <li>入社AI市场</li> </ul>      | → 創建幣                                                                               | 3%,文本数据量超〕<br>#<br>X             | 过2万条。<br>发布商品         | → | 售卖获得收入            |          |             |            |
| 专项硬件适配服务                                                                                  | 序号<br>1                                                 | 接口名称<br>邦邦得劲检测                                                                      | 应用类型                             | 模型ID<br>8160          |   | 模型名称<br>吴拓邦的物体检测2 | 模型类型物体检测 | 售卖状态<br>未发布 | 操作<br>发布商品 |

(2)在AI市场服务商后台的商品列表中选择"商品发布",在商品分类中选择软件服务-API接口,右侧出现从"从现有API发布"按钮,点击后选中您 需要发布的API。

#### の3、发布API商品

商品发布需要填写四个部分信息:

- **商品基本信息**:注意:由于EasyDL服务的限制,买家使用的最大并发限制为2QPS,您可在商品信息中提醒买家,API服务暂无法支持高并发的 需求。
- API生产信息:API生产信息主要包括请求参数、请求示例、返回示例以及错误码。由于本部分信息对于买家调用API非常重要,请谨慎填写。具体内容您可按照训练模型的类别,参考EasyDL官方提供的定制化声音识别API文档进行填写。
- 交付信息:选择线上交易,创建规格。在线售卖API已默认添加0元/20次(买家只能购买一次)的体验版规格,方便买家进行简单测试,您可以 根据需要自行删除或添加套餐包规格,例如配置套餐包:配额100次调用,价格10元。
- **售后信息**:售后信息包括发票与服务保障。对于企业服务商,您需要为买家提供增值税普票或专票,服务保障请您按照实际情况填写。

商品信息填写完整后即可提交审核。提交后,将会由专人进行质量审核,约一个工作日给您结果反馈,审核期间,AI市场将会测试调用一次,来确定您的API是否符合售卖标准。

# の4、售卖并获得收入

您的API服务审核通过并成功上架后,AI市场的买家将可在市场中购买您的服务。成功付款后,收入将由AI市场代收,您可在月底发起结算申请,AI市场将销售收入打到您的账户中。

# രം 2022年12月

| 序号 | 块        | 功能描述                                                                   |
|----|----------|------------------------------------------------------------------------|
| 1  | 模型训<br>练 | EasyDL图像·物体检测模型训练新增「半监督学习」功能,支持添加已标注数据和未标注数据共同参与训练,从而获得一个泛<br>化效果更好的模型 |

# 心 2022年09月

| 序号 | 功能模<br>块 | 功能描述                                                                                                                                                                               |
|----|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | 模型评<br>估 | EasyDL图像-图像分类、物体检测模型评估功能新增「模型调优建议」功能,建议将从「基于整个模型」和「基于单个标签」两<br>个角度,给出数据中各因素(对比度、标注框面积等)对模型准确率、漏检率、误检率三个方面造成的影响程度,并提出修<br>改策略建议。                                                    |
| 2  | 模型训<br>练 | EasyDL图像-实例分割新增支持「高性能」算法,性能相较高精度算法提升一倍。                                                                                                                                            |
| 3  | 模型训<br>练 | EasyDL文本-文本创作后端升级文心大模型,从原来的仅提供10B模型调整为高精度(10B)、高性能(1.5B)两类算<br>法,10B模型在效果方面具有2%领先优势,1.5B模型在性能方面具有较大优势,预计1.5B模型训练资源减少到原有的<br>1/8(10B模型训练需要8张v100卡加载,1.5B模型只需要1张v100),预测资源预估约减少为1/3。 |
| 4  | 模型服<br>务 | EasyDL结构化数据-表格数据预测新增支持「本地服务器/通用小型设备x86 CPU部署」,满足用户对离线部署的场景需求。                                                                                                                      |
| 5  | 产品优<br>化 | EasyDL我的模型页面支持模型搜索、模型置顶功能,进一步优化用户模型管理使用体验。                                                                                                                                         |

# 心 2022年07月

| 序号 | 功能模<br>块 | 功能描述                                                                                                                       |
|----|----------|----------------------------------------------------------------------------------------------------------------------------|
| 1  | 模型服<br>务 | EasyDL图像-图像分类、物体检测新增「辨影软硬一体方案」 ,针对图像分类、物体检测操作台在训练模型时支持选择辨影作<br>为专项适配硬件的选项之一,通过发挥辨影在工业质检场景灵活部署、自带显示屏等业务优势,提高用户本地部署的易用<br>性。 |
| 2  | 模型训<br>练 | EasyDL图像-图像分类、物体检测新增支持「超高性能」算法,性能提升1.2倍。                                                                                   |
| 3  | 模型服<br>务 | EasyDL图像-公有云服务支持接口输入为图片URL地址,优化用户服务调用使用体验。                                                                                 |
| 4  | 模型服<br>务 | EasyDL文本-文本创作支持在线H5体验模型推理,优化用户在模型体验阶段的使用体验。                                                                                |
| 5  | 产品优<br>化 | 创建模型数量上调至30,进一步满足用户创建多个模型的开发诉求。                                                                                            |

心 2022年06月

| 序号 | 功能模块 | 〕能描述                                                    |
|----|------|---------------------------------------------------------|
| 1  | 场景应用 | asyDL新增支持「场景范例」,包含工业制造、智慧城市、电商等多行业场景算法,辅助用户快速上手体验AI开发流程 |

| 序号 | 功能模<br>块 | 功能描述                                                                     |
|----|----------|--------------------------------------------------------------------------|
| 1  | 模型种<br>类 | EasyDL新增跨模态技术方向上线「图文匹配」任务类型,满足用户在图文匹配相关场景的应用诉求,例如,知识文档、图谱<br>构建等。        |
| 2  | 场景应<br>用 | EasyDL图像-物体检测新增人脸检测、行人检测、车辆识别3类场景算法能力,进一步满足用户在细分场景对高精度模型的诉<br>求。         |
| 3  | 模型训<br>练 | EasyDL图像·物体检测高级训练配置新增支持「数据不平衡优化」策略,解决用户在数据收集阶段因为样本分布数量不均导致<br>模型精度下降的问题。 |

രു 2022年04月

| 序号 | 功能模块 | 功能描述                                                   |
|----|------|--------------------------------------------------------|
| 1  | 模型种类 | EasyDL图像-图像分割新增模「语义分割」模型类型,充分解决用户在识别外观形状有细分差异场景下的识别难题。 |

# രം 2022年03月

|   | 序号 | 功能模<br>块 | 功能描述                                                     |
|---|----|----------|----------------------------------------------------------|
|   | 1  | 模型种      | EasyDL文本新增「评论观点抽取 」任务类型,多维度收集用户的文本评论观点,常应用于电商购物、政府信箱等业务场 |
| 1 | 类  | 累。       |                                                          |

# രു 2022年01月

| 序号 | 功能模<br>块 | 功能描述                                                                 |
|----|----------|----------------------------------------------------------------------|
| 1  | 模型服<br>务 | EasyDL图像-图像分类&物体检测支持批量预测,支持用户异步获取预测结果的诉求,同时帮助用户进一步利用资源空闲期降<br>低预测成本。 |

# രം 2021年10月

| 序号 | 功能模<br>块 | 功能描述                                                                               |
|----|----------|------------------------------------------------------------------------------------|
| 1  | 模型训<br>练 | EasyDL图像-图像分类分类支持免训练迭代模式,当用户的训练数据存在不断更新的情况,可开启该模式,快速添加新数据至数据库直接迭代模型,不必重新训练,降低训练成本。 |
| 2  | 模型训<br>练 | EasyDL图像-物体检测支持自定义四边形标注和训练,充分满足用户在复杂标注场景下的物体检测标注诉求,提高模型效果精度。                       |

രു 2021年09月

| 序号 | 功能模块 | 功能描述                |                  |         |           |
|----|------|---------------------|------------------|---------|-----------|
| 1  | 模型部署 | EasyDL图像-上线智能边缘控制台, | 充分满足用户在本地管理预测设备、 | 实现服务包统─ | -更新迭代的诉求。 |

# രം 2021年08月

| 序号 | 功能模块 | 功能描述                                        |
|----|------|---------------------------------------------|
| 1  | 模型种类 | EasyDL图像·物体检测支持小目标检测,解决当目标物极小时识别率不佳的复杂场景问题。 |
| 2  | 模型训练 | EasyDL图像-支持增量训练任务                           |
| 3  | 数据服务 | EasyDL图像-图像分割支持自动识别标注                       |

രം 2021年07月

| 序号 | 功能模块 | 功能描述             |
|----|------|------------------|
| 1  | 模型训练 | EasyDL图像支持自动超参搜索 |

രം 2021年06月

| 序号 | 功能模块 | 功能描述                 |
|----|------|----------------------|
| 1  | 模型种类 | EasyDL视频-目标跟踪支持多标签模型 |
| 2  | 数据服务 | EasyDL视频-目标跟踪支持在线标注  |
| 3  | 数据服务 | EasyDL文本-实体抽取支持智能标注  |
| 4  | 数据服务 | EasyDL图像上线噪声样本挖掘     |

രം 2021年05月

| 序号 | 功能模块 | 功能描述                      |
|----|------|---------------------------|
| 1  | 模型种类 | EasyDL文本技术方向新增多语种文本分类模型   |
| 2  | 模型种类 | EasyDL的本地服务器API已支持线上购买    |
| 3  | 模型训练 | EasyDL图像分类高级训练配置支持数据不平衡优化 |
| 4  | 模型训练 | EasyDL图像支持精度提升配置包         |
| 5  | 模型训练 | EasyDL图像支持自定义验证集&自定义测试集   |

രം 2021年03月

| 序号 | 功能模块 | 功能描述           |
|----|------|----------------|
| 1  | 模型种类 | EasyDL视频上线目标跟踪 |

2 模型种类 EaysDL OCR全新上线

# 2021年01月

| 序号 | 功能模块 | 功能描述                                     |
|----|------|------------------------------------------|
| 1  | 模型种类 | EasyDL结构化数据上线时序预测模型                      |
| 2  | 数据服务 | 图像数据导入新增支持COCO格式、导出支持VOC、COCO格式          |
| 3  | 数据服务 | 在大图标注模式下,提供无损压缩的快速浏览模式                   |
| 4  | 数据服务 | 图像分类批量标注                                 |
| 5  | 数据服务 | 图像分割支持导入已标数据                             |
| 6  | 模型训练 | 模型训练支持配置epoch、输入图片分辨率等高级参数               |
| 7  | 模型评估 | 图像分类模型的混淆矩阵分析,支持查看热力图                    |
| 8  | 模型部署 | 支持端云协同服务                                 |
| 9  | 模型部署 | 设备端SDK新增支持Android平台高通骁龙GPU、Linux平台瑞芯微NPU |
| 10 | 模型部署 | 软硬一体方案新增均衡算法,提供在精度和性能上更加平衡的算法选择          |

രം 2020年12月

| 序号 | 功能模块 | 功能描述                                     |
|----|------|------------------------------------------|
| 1  | 数据服务 | 校验模型页面查看检测框,支持按置信度排序,在物体检测框数量较多时查看结果体验更佳 |
| 2  | 模型管理 | 支持模型名称修改                                 |

രം 2020年11月

EasyDL平台全新升级,包含图像、文本、语音、OCR、视频、结构化数据6大方向,及零售行业版,覆盖更多应用场景

| 序号 | 功能模块 | 功能描述                                   |
|----|------|----------------------------------------|
| 1  | 模型种类 | 上线短文本相似度任务                             |
| 2  | 数据服务 | 标签格式支持中文                               |
| 3  | 数据服务 | 标注框支持自定义颜色 ,优化图像分割/物体检测标注模板下用户的标注框浏览体验 |

| 序号 | 功能模块 | 功能描述                                         |
|----|------|----------------------------------------------|
| 1  | 模型种类 | 支持更丰富的技术方向/任务类型:文本实体抽取、语音识别、结构化数据分析,开放目标跟踪邀测 |
| 2  | 模型训练 | 接入AI市场,支持用户交易模型,并基于购买的模型再训练                  |
| 3  | 模型部署 | 图像分类模型,支持适配比特大陆的设备端SDK                       |
| 4  | 模型训练 | 物体检测模型,新增「超高精度」、「均衡」两种算法                     |
| 5  | 模型性能 | 图像分类、检测、分割模型,支持在训练页面查看算法的适配硬件及性能,方便选择算法      |
| 6  | 模型效果 | 文本分类单标签模型,后端框架接入文心,支持高精度与高性能两个算法             |
| 7  | 模型部署 | 文本单标签、多标签、情感倾向分析模型,支持私有API部署                 |
| 8  | 模型部署 | 支持在线购买软硬一体方案专用SDK、按产品线鉴权设备端SDK授权             |

# രു 2020年6月

| 序号 | 功能模块 | 功能描述                                                    |
|----|------|---------------------------------------------------------|
| 1  | 模型训练 | 图像分类、物体检测模型训练时,支持配置数据增强算子                               |
| 2  | 模型部署 | 物体检测模型支持Atlas系列硬件,包括:设备端华为Atlas 200开发板、服务器端Atlas 300加速卡 |

# രം 2020年5月

| 序号 | 功能模块 | 功能描述               |
|----|------|--------------------|
| 1  | 模型部署 | JetsonNano软硬一体方案上线 |

# 心 2020年4月

| 序号 | 功能模块 | 功能描述               |
|----|------|--------------------|
| 1  | 数据服务 | 图像分类支持在线标注         |
| 2  | 模型部署 | 图像分类支持量化加速,提高端部署性能 |

# രം 2020年3月

| 序号 | 功能模块 | 功能描述                               |
|----|------|------------------------------------|
| 1  | 模型部署 | EdgeBoard(VMX)软硬一体方案上线             |
| 2  | 模型部署 | 新增声音分类服务器端SDK                      |
| 3  | 模型部署 | 图像分类设备端基础版SDK,支持Linux系统Atlas200开发板 |

# രു 2019年12月

| 序号 | 功能模块 | 功能描述                              |
|----|------|-----------------------------------|
| 1  | 模型效果 | 物体检测设备端SDK部署高精度算法精度进一步提升,平均精度提升5% |
| 2  | 数据服务 | nlp-序列标注功能上线                      |
| 3  | 模型效果 | nlp方向支持"文心2.0"预训练模型               |

# രം 2019年8月

| 序号 | 功能模块 | 功能描述              |
|----|------|-------------------|
| 1  | 模型种类 | 图像分类本地部署训练新增高精度算法 |
| 2  | 模型效果 | 物体检测模型效果优化        |
| 3  | 模型种类 | EasyDL新增图像分割模型    |

രം 2019年7月

| 序号 | 功能模块 | 功能描述                      |
|----|------|---------------------------|
| 1  | 模型效果 | 物体检测高性能模型平均准确率提升20%       |
| 2  | 模型性能 | 物体检测高性能模型后端时延降低90%,约500ms |

# രം 2019年6月

| 序号 | 功能模块 | 功能描述         |
|----|------|--------------|
| 1  | 数据服务 | EasyDL智能标注功能 |

## の 2019年5月

| 序号 | 功能模块 | 功能描述           |
|----|------|----------------|
| 1  | 模型种类 | EasyDL定制视频分类上线 |

# രം 2019年4月

EasyDL零售行业版上线

| 序号 | 功能模块 | 功能描述                           |
|----|------|--------------------------------|
| 1  | 数据服务 | 物体检测支持多人同时标注数据集                |
| 2  | 模型部署 | 物体检测离线SDK新增支持windows及linux操作系统 |

# 常见问题

# 常见问题

```
₯ 价格常见问题
```

# 模型训练如何计费?

• EasyDL提供各个技术方向均提供免费训练的模式,如需使用更高级付费的资源,请开通付费后使用

图像分类、物体检测、图像分割API如何收费?调用量不够怎么办?

• 每个公有云API有累计10000点的免费调用额度,如需付费使用,请在控制台进行线上购买

# 心 数据相关问题

# 需要上传多少张图片才能训练出效果较好的模型?

• 每个分类至少需要准备20张以上。如果想要较好的效果,建议每个分类准备不少于100张图片。

# 上传图片的总量有限制吗?

• 每个账号下所有数据集的图片总数不能超过200万张。

智能标注功能目前已对图像分类、物体检测、图像分割模型开放,了解功能详情

以下为智能标注相关常见问题

#### "一键标注"和"立即训练"要如何选择?

- 当系统推荐"立即训练",且系统预标注的框确实已非常精准时,可以不用标注剩余数据,直接开始模型训练。此时,仅用当前已标注图片训练的模型,与标注所有数据后训练的模型相比,效果几乎等同
- 如果系统预标注的框还有些不精准,可以启动一键标注,人工确认系统标注的标注框后,再开始训练

# 选择了"立即训练"之后是否还可以"一键标注"?

- 选择"立即训练"之后,系统默认为您结束此次智能标注
- 再次启动智能标注后,您可以通过以下方式进行一键标注:
  - 根据系统提示,进入一键标注
  - 查看系统对"未标注[优先]"图片的预标注,点击"满意预标注结果"后,进入一键标注

# 智能标注结束后,又往数据集上传了新图片,是否可以直接"一键标注"新图片?

- 如果您创建了新的标签、或新上传的图片场景和之前的图片场景差异较大,建议不要使用一键标注,而是从头开始智能标注(即再次筛选关键 图片)
- 如果不是以上情况,再次启动智能标注后,可以通过以下方式进行一键标注:
  - 根据系统提示,进入一键标注
  - 查看系统对"未标注[优先]"图片的预标注,点击"满意预标注结果"后,进入一键标注

# 智能标注中可以增删标签吗?

- 暂不支持。为了保证系统智能标注的效果,建议在启动功能前就创建好所有需要识别的标签
- 如果确实需要增删标签,可以先结束智能标注

# 智能标注中可以增删图片吗?

- 暂不支持。为了保证系统智能标注的效果,建议在启动功能前上传需要标注的所有图片,并删除不相关的图片
- 如果确实需要增删图片,可以先结束智能标注

# 智能标注中可以修改已标注图片的标注框吗?

- 可以。但为了保证智能标注的效果,建议不要大量改动
- 如果确实需要修改大量标注,建议先结束智能标注

# 为什么我已经人工标注了很多图片,但系统预标注依然不准?

- 系统预标注的结果会受以下因素影响:
  - 智能标注期间,对"已标注"图片的标签进行大量改动
  - 曾结束智能标注,并对标签、图片进行增删
- 如果您没有进行以上操作,系统标注结果依然不理想,请在百度智能云控制台内提交工单反馈

# 多个数据集是否可以同时启动智能标注?

• 目前每个账号同一时间仅支持对一个数据集启动智能标注

# 共享中的数据集是否可以启动智能标注?

• 暂不支持。智能标注中的数据集也暂不支持共享,如有疑问请在百度智能云控制台内提交工单反馈

# 智能标注失败了怎么办?

- 可以先尝试稍后重新启动
- 若再次遇到问题,请在百度智能云控制台内提交工单反馈

# 图像分割模型如何正确标注?

- 所有图片中出现的目标物体都需要被标出 (标注可以重叠)
- 标注应包含整个物体,且尽可能不要包含多余的背景
- 如果图片中存在很多相同标签的目标物体,可以使用右侧的锁定按钮。锁定标签后,只需要在左侧标注目标物体即可,不用再重复选择标签

心 训练相关问题

# 数据处理失败或者状态异常怎么办?

 如是是图像分类模型上传处理失败,请先检查已上传的分类命名是否正确,是否存在中文命名、或者增加了空格;然后检查下数据图片量是否 超过上限(每个账户下200万张);再检查图片中是否有损坏。如果自查没有发现问题,请在百度智能云控制台内提交工单反馈

#### 模型训练失败怎么办?

• 如果遇到模型训练失败的情况,请在百度智能云控制台内提交工单反馈

#### 已经上线的模型还可以继续优化吗?

 已经上线的模型依然可以持续优化,操作上还是按照标准流程在训练模型中-选择要优化的模型和数据完成训练,然后在模型列表中更新线上 服务,完成模型的优化

#### 心 模型效果相关问题

#### 如何通过「完整评估结果」里的错误示例优化模型?

- 错误示例中, 左侧是正确的结果, 右侧是模型的识别结果
- 观察模型识别有误的图片有哪些共同点,并有针对性地补充训练数据。比如:当图片比较亮的时候模型都能识别正确,但比较暗的时候模型就 识别错了。这时就需要补充比较暗的图片作为训练数据

#### 我的数据有限,如何优化效果?

• 先申请发布模型,并备注说明希望通过云服务数据管理功能,将实际调用云服务识别的图片加入训练集,不断迭代模型

#### 实际调用服务时模型效果变差?

- 训练图片和实际场景要识别的图片拍摄环境应一致,举例:如果实际要识别的图片是摄像头俯拍的,那训练图片就不能用网上下载的目标正面 图片
- 每个标签的图片需要覆盖实际场景里面的可能性,如拍照角度、光线明暗的变化,训练集覆盖的场景越多,模型的泛化能力越强
- 如果使用的是云服务,可以开通云服务数据管理功能,将实际调用云服务识别的图片加入训练集,不断迭代模型
- \*\*如果训练数据已经达到以上要求,且单个分类/标签的图片量超过200张以上,效果仍然不佳,请在百度智能云控制台内提交工单反馈

#### ∞ 模型上线相关问题

#### 希望加急上线怎么处理?

• 请在百度智能云控制台内提交工单反馈

#### 每个账号可以上线几个模型?是否可以删除已上线的模型?

• 每个账号最多申请发布十个模型,已上线模型无法删除

#### 申请发布模型审核不通过都是什么原因?

 可能原因有,1、经过电话沟通当前模型存在一些问题或者不再使用,如训练数据异常、数据量不够、不想再继续使用等原因,沟通达成一致 拒绝。2、电话未接通且模型效果较差,会直接拒绝。如果需要申诉,请在百度智能云控制台内提交工单反馈

∞ 模型部署相关问题

心平台的部署方式支不支持我的硬件?

# Baidu 百度智能云文档

| 部署类型         | 支持的硬件示例                                       |  |  |  |  |  |
|--------------|-----------------------------------------------|--|--|--|--|--|
| 通用ARM        | 绝大多数安卓、苹果手机;瑞芯微RK32、RK32、RK35系列、树莓派等开发板       |  |  |  |  |  |
| 英特尔神经计算棒     | NCS 1代、NCS 2代                                 |  |  |  |  |  |
| 海思NNIE       | Hi3559AV100/Hi3559CV100等                      |  |  |  |  |  |
| 华为昇腾Atlas开发板 | Atlas200计算盒、Atlas300 计算卡                      |  |  |  |  |  |
| 比特大陆SE计算盒    | Bitmain SE5                                   |  |  |  |  |  |
| 通用x86CPU     | 绝大多数英特尔和AMD CPU                               |  |  |  |  |  |
| 通用x86CPU加速版  | 英特尔志强、酷睿、凌动系列CPU                              |  |  |  |  |  |
| 高通骁龙         | 骁龙660以后芯片的手机                                  |  |  |  |  |  |
| 华为NPU        | mate10,mate10pro,P20,mate20,荣耀v20等            |  |  |  |  |  |
| 华为达芬奇NPU     | mate30,p40,nova6,荣耀v30等                       |  |  |  |  |  |
| 英伟达GPU       | 消费级显卡GeForce系列、RTX系列、TITAN,专业显卡Quadro、Tesla系列 |  |  |  |  |  |
| 英伟达Jetson    | TX2、Nano、Xavier、Xavier NX                     |  |  |  |  |  |

心 快速链接

# 公有云部署

公有云服务API调用常见错误码

# 本地服务器部署

Windows-GPU部署常见问题 Linux-C++部署常见问题 Linux-Python部署常见问题 Linux-Atlas部署常见问题

## 通用小型设备部署

Android部署常见问题 iOS部署常见问题 Windows-CPU部署常见问题 Linux-C++部署常见问题 Linux-Python部署常见问题

# 软硬一体方案部署

EdgeBoard(FZ)部署常见问题 EdgeBoard(VMX)部署常见问题 Jetson部署常见问题

# 智能边缘控制台

# 智能边缘控制台-单节点版

心 EasyEdge 智能边缘控制台-单节点版 IEC

EasyEdge Intelligent Edge Console (以下简称IEC) 是EasyEdge推出的边缘设备管理的本地化方案。可以运行于多种架构、多系统、多类型的终端之上。通过IEC,用户可以方便地在本地进行

- EasyDL/BML/EasyEdge的SDK的 离线 / 在线激活,服务管理
- 接入本地和远程摄像头,网页中实时预览
- 自动监控和记录相关事件
- 硬件信息的可视化查看

支持的系统+CPU架构包括:

• Windows x86\_64 (Windows 7 ~ Windows 10, 暂不支持Windows 11)

• Linux x86\_64 / arm32 / arm64

支持各类常见的AI加速芯片,包括:

- NVIDIA GPU / Jetson 系列
- Baidu EdgeBoard FZ系列
- 比特大陆 Bitmain SC / SE 系列
- 华为 Atlas 系列
- 寒武纪 MLU 系列
- 其他EasyDL/EasyEdge/BML支持的AI芯片
  - 完整列表可参考这里

```
രം Release Note
```

注意:2.0.0之后,默认以系统服务形式安装iec,无法兼容1.x版本的iec

| 版本号     | 发布时间       | 更新说明                                      |
|---------|------------|-------------------------------------------|
| 2.2.2.1 | 2023-08-30 | 系统稳定性优化                                   |
| 2.2.2.0 | 2023-03-16 | 支持多媒体服务器硬件编解码                             |
| 2.2.1   | 2022-11-15 | 修复告警规则删除问题                                |
| 2.2.0   | 2022-10-27 | 新增onvif/gb28181支持;完善端云通信逻辑                |
| 2.0.0   | 2022-03-22 | 支持连接中心节点IECC;支持以系统服务安装                    |
| 1.0.2   | 2021-12-22 | 更新视频预览推流库;新增若干AI芯片支持;支持多种芯片温度、功耗展示;多项性能优化 |
| 1.0.0   | 2021-09-16 | IEC 第一版!                                  |

っ 快速开始

从这里选择您需要的操作系统和CPU架构下载:

- Windows amd64 : intel、AMD的64位x86\_84 CPU
- Linux amd64 : intel、AMD的64位x86\_84 CPU
- Linux arm : 树莓派等32位的ARM CPU
- Linux arm64: RK3399、飞腾等64位的ARM CPU

或者从纯离线服务管理页可下载智能边缘控制台

| 体检测模型        | < <p>     《三 纯离线服务     《     书》     《     书     《     书     》     书     《     书     》     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书     书</p> |  |  |  |  |
|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|
| 创建模型         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |  |  |  |  |
| 训练模型         | 纯离线服务说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |  |  |  |  |
| 校验模型         | 发布纯高线服务,将训练完成的模型部署在本地、高线调用模型,可以选择将模型部署在本地的服务器、小型设备、软硬一体方案专项适配硬件上,<br>通过APL SDK讲一步集成,显示适应不同业务质量;您可下载EasyEdge器能计线控制合,便建管理专个本地模型服务。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |  |  |  |  |
| 发布模型         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |  |  |  |  |
| EasyData数据服务 | 友 他 新 微 分 省 能 工 投 制 音 ① ● 下 軟 替 能 22 課 及 制 音 ①                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |  |  |  |
| 数据总览         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |  |  |  |  |
| 标签组管理        | <b>服务器</b> 通用小型设备 专项适配硬件                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |  |  |  |  |
| 在线标注         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |  |  |  |  |
| 智能标注         | JUIN AFI                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |  |  |  |  |
| 云服务数据回流      | 此处发布、下载的SDK为未授权SDK,需要前往控制台获取序列号激活后才能正式使用。SDK内附有对应版本的Demo及开发文档,开发者可参考源代码完成开发,                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |  |  |  |  |
|              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |  |  |  |  |

您也可以通过先安装多节点版本IECC,通过中心节点来自动连接安装边缘节点。

| 0 EasyEdge-IEC-v2.0.0-linux-amd64-standard > tree . |
|-----------------------------------------------------|
|                                                     |
| easyedge-iec                                        |
| easyedge-iec-setup.sh                               |
| ├─── etc                                            |
| easyedge-iec.yml                                    |
| └─── readme.txt                                     |

以系统服务形式安装(推荐) 以root用户运行bash ./easyedge-iec-setup.sh install 即可

```
[setup]: sudo could not be found
[setup]: Start to install IEC...
[setup]: + bash -c "./easyedge-iec --com.role=edge --service=install"
[setup]: + bash -c "cp /code/EasyEdge-IEC-v2.2.2-linux-amd64-standard/easyedge-iec-setup.sh /usr/sbin/easyedge-iec-setup.sh"
[setup]: Install IEC success!
[setup]: + bash -c "/usr/sbin/easyedge-iec --com.role=edge --service=start"
[setup]: Start to check IEC status...
[setup]: + bash -c "curl -s 127.0.0.1:8702 >/dev/null"
[setup]: + bash -c "curl -s 127.0.0.1:8702 >/dev/null"
[setup]: + bash -c "/usr/sbin/easyedge-iec --com.role=edge --service=status | grep running > /dev/null 2>&1"
[setup]: + bash -c "/usr/sbin/easyedge-iec --com.role=edge --service=status | grep running > /dev/null 2>&1"
[setup]: Hot - C "/usr/sbin/easyedge-iec --com.role=edge --service=status | grep running > /dev/null 2>&1"
[setup]: Hot - C "/usr/sbin/easyedge-iec --com.role=edge --service=status | grep running > /dev/null 2>&1"
[setup]: Hot - C "/usr/sbin/easyedge-iec --com.role=edge --service=status | grep running > /dev/null 2>&1"
[setup]: Hot - C "/usr/sbin/easyedge-iec --com.role=edge --service=status | grep running > /dev/null 2>&1"
[setup]: Hot - C "/usr/sbin/easyedge-iec/easyedge-iec.yml
[easyedge-iec]: default configure file: /etc/easyedge-iec/easyedge-iec.yml
[easyedge-iec]: service usage: service easyedge-iec { start | stop }
[easyedge-iec]: Done!
```

• 日志: /var/log/easyedge-iec/easyedge-iec.log

```
• 系统配置: /etc/easyedge-iec/easyedge-iec.yml
```

• 服务启动/停止: service easyedge-iec { start | stop } (不同操作系统内可能不同,具体命令参考安装日志)

自定义安装 (不推荐) 自定义安装方法仅限于 安装脚本无法识别的情况。

- 拷贝 ./EasyEdge-IEC-v2.0.0-linux-amd64-standard/ 整个目录至自定义文件夹,如/opt/EasyEdge-IEC
- 进入到 /opt/EasyEdge-IEC
- 通过 nohup 等方法运行 ./easyedge-iec-linux-{您的系统架构} amd64: intel、AMD的64位x86\_84 CPU arm: 树莓派等32位的ARM CPU \* arm64: RK3399、飞腾等64位的ARM CPU
- 日志: ./log/easyedge-iec.log
- 系统配置: ./easyedge-iec.yml

の Windows 安装

解压缩之后,安装目录如下所示:

打开命令行 (非powershell) 运行 easyedge-iec-setup.bat install。

如果遇到hang住的情况,可修改命令行配置启动之后,打开浏览器,访问http://{设备ip}:8702/easyedge/iec即可:

D X

| Admini                        | istrator: C:\\                                                                                                 | Windows\sy              | stem32\cmc                         | l.exe                                                |                  |                                                                                       | <b>A</b> |  |
|-------------------------------|----------------------------------------------------------------------------------------------------------------|-------------------------|------------------------------------|------------------------------------------------------|------------------|---------------------------------------------------------------------------------------|----------|--|
| .crosoff<br>) 2018<br>\Users\ | t Windows<br>Microsof<br>\eval>_<br>Windows                                                                    | Versia<br>Corpor        | on 10.0.17<br>Nation。保<br>32\cmd.e | 763.2145]<br>窗所有权利。<br>xe <sup>*</sup> 属性            | -<br>-<br>-<br>- | 还原(R)<br>移动(M)<br>大小(S)<br>最小化(N)<br>最大化(X)<br>关闭(C)<br>编辑(E)<br>默认值(D)<br>属性(P)<br>X | 右键       |  |
| 选项                            | 字体                                                                                                             | 布局                      | 颜色                                 |                                                      |                  |                                                                                       |          |  |
|                               | 示<br>小(S)<br>中(M)<br>大(L)<br>撮 法連<br>編<br>環<br>一<br>て<br>に<br>一<br>間<br>た<br>に<br>一<br>に<br>の<br>一<br>で<br>(M) | 模式(Q)<br>(I)<br>键快捷方    | ← 〕                                | 節令记录<br>缓)中区大小(B):<br>缓)中区数量(N):<br>□ 丢弃旧的副本(D) 取消勾选 | 50               |                                                                                       |          |  |
| 文<br>文<br>〇                   | 本选择<br>启用对选<br>扩展的文                                                                                            | 定内容自<br>本选择键            | / 用TF复制<br>动换行(E)<br>(X)           | //HIXEDUXJERE(C)                                     |                  |                                                                                       |          |  |
| 一当                            | 前代码页                                                                                                           |                         |                                    |                                                      |                  |                                                                                       |          |  |
| 65                            | 001 (UTF                                                                                                       | 8)                      |                                    |                                                      |                  |                                                                                       |          |  |
| □ 個<br>了<br>了解                | 5月旧版招<br>了解有关  <br>更多有关                                                                                        | 制台(U)(<br>日版控制台<br>新控制台 | (需要重新)<br>台模式的详<br>功能的信息           | 自动,影响所有控制台)<br>細信息<br>息                              |                  |                                                                                       |          |  |
|                               |                                                                                                                |                         |                                    |                                                      | R                | STR.                                                                                  |          |  |

启动之后,打开浏览器,访问 http://{设备ip}:8702 即可:

| <b>О</b> павжа                                                                |                        |  |  |
|-------------------------------------------------------------------------------|------------------------|--|--|
| EasyEdge 智能边缘控制台–单节点版                                                         |                        |  |  |
| EnryEdga制造快速器署A模型。可洗化完成模型服务demo体验、摄像头实对推建、识别结果事件记录、做速完成本地数据<br>与概型年联、快速体验模型效果 | 本地管理员登录                |  |  |
|                                                                               |                        |  |  |
|                                                                               | 初始图得: easyedge         |  |  |
|                                                                               | <b>王王</b><br>進作加速 停改进行 |  |  |
| <b>相关产品</b> : EasyCL電门版A开发干台 EML生功能A开发干台 EasyEcope属与功能A服务干台                   | <b>版本号</b> : V222      |  |  |
|                                                                               |                        |  |  |

默认用户名密码为 admin / easyedge
| EasyEdge 智能边缘控制                                       | 则台–单节点版                                  |                                           |                                                         |                                    | 操作指南 🔒 admin 🗸 |
|-------------------------------------------------------|------------------------------------------|-------------------------------------------|---------------------------------------------------------|------------------------------------|----------------|
| 总览                                                    |                                          |                                           |                                                         |                                    |                |
| ○ 本地模型服务<br>添加模型服务<br>模型服务管理<br>量 其的摄像头<br>添加摄像头      | 基本信息<br>主代名<br>取同化念 ● 在线<br>  使作逝行情况概定   |                                           |                                                         |                                    |                |
| 損傷失策理<br>(3) 本地事件<br>本地事件に录<br>(3) 设置<br>系統设置<br>進行日志 | 15.7%<br>止内府<br>运行所行 15008<br>已用所行 24308 | 15.1%<br>47至4<br>急空間 21578<br>日間空間 3.2078 | 18.2%<br>CPUR月年<br>CPUR月年<br>CPUR月年<br>CPUR月年<br>AND(R) |                                    |                |
|                                                       | ★地須目管理 ② 規型服务管理 >> ○ 0 0 21回 送け中         |                                           | ())))))))))))))))))))))))))))))))))))                   | (3) 本地事件记录 ><br>0<br><sup>民国</sup> |                |
| 版本号: V2.2.2                                           |                                          |                                           |                                                         |                                    |                |

## ゆ 功能使用说明

#### の①添加模型服务

首先,点击导航栏的「本地模型服务」-「添加模型服务」。在页面中定义服务名称后,将已经下载好的Linux/Windows版本的SDK与IEC关联。关 联完毕后可按两种激活方式,激活使用SDK。

| EasyEdge 智能边缘控制           | 明台-单节点版          |                                  |                 |    |  |  |  |  |  |
|---------------------------|------------------|----------------------------------|-----------------|----|--|--|--|--|--|
| 总览                        | 添加模型服务           |                                  |                 |    |  |  |  |  |  |
| ② 本地模型服务<br>添加模型服务        | ■ 新建模型服<br>*服务名称 | 新建築加加市                           |                 |    |  |  |  |  |  |
| ● 我的图像头<br>烫如摄像头          |                  | 土 选择SDK                          | m6 x86.zip      |    |  |  |  |  |  |
| 抽錄失管理<br>□ 本地事件<br>本地事件记录 |                  |                                  |                 |    |  |  |  |  |  |
| ② 设置<br>系统设置<br>运行目志      |                  |                                  |                 |    |  |  |  |  |  |
|                           |                  | 变量名称                             | 安量值             | ## |  |  |  |  |  |
|                           |                  | 十 添加环境变量<br>最多支持100组环境变<br>添加并激活 | 量,单个输入框不超过256个字 |    |  |  |  |  |  |
| 版本号: V2.2.2               |                  |                                  |                 |    |  |  |  |  |  |

部分SDK需要提前安装系统依赖,如TRT等,具体请参考EasyDL/BML/EasyEdge SDK使用文件中的环境依赖安装说明

#### 联网激活

1. 在关联SDK完成后,需要在百度智能云控制台对应部署方式管理页中新增测试序列号或购买正式序列号。(图中以服务器版SDK为例)

| ✓ 返回EasyDL总览                 | 产品服务 / EasyDL图像。 | (务 / EasyDL图像 - 服务图线舆线服务 |      |        |       |     |      |            |           |  |
|------------------------------|------------------|--------------------------|------|--------|-------|-----|------|------------|-----------|--|
| EasyDL图像                     | 服务器纯离线服务         | 路纯离线服务管理                 |      |        |       |     |      |            |           |  |
| 公有云部署 🗸 🗸                    | API SDI          | PI SDK                   |      |        |       |     |      |            |           |  |
| EasyEdge本地部署 <               |                  |                          |      |        |       |     |      |            |           |  |
| • 服务器结离线服务                   | 使用说明: (展开)       | 使用说明: (展开查看部署介绍及操作步骤等)   |      |        |       |     |      |            |           |  |
| <ul> <li>设备储结高线服务</li> </ul> | + 购买正式授权         | 新培测试序列号 ?                |      |        |       |     |      |            | 开发文档 工单支持 |  |
| 。 专项硬件纯离线                    | 设备名              | 技术方向                     | 模型ID | 可激活模型数 | 序列号类型 | 序列号 | 激活时间 | 到期日期       | 操作        |  |
|                              | 自定义设备 🖉          | 图像                       |      | 3      | 试用    |     | 未敢活  | 2021-12-31 | 如何激活      |  |
|                              | 自定义设备 🖉          | 图像                       |      | 3      | 试用    |     | 未激活  | 2021-12-31 | 如何激活      |  |
|                              | 自定义设备 之          | 图像                       |      | 3      | 试用    |     | 未激活  | 2021-12-31 | 如何激活      |  |

- 2. 再在IEC中填入所申请的序列号
- 3. 配置服务,在服务端口不冲突占用的情况下,使用默认即可
- 4. 添加并激活

#### 离线激活

1. 在IEC总览页面下载「指纹文件」

| EasyEdge 智能边缘                               | 控制台                                                                                                         | 操作指南 | admin ~ |
|---------------------------------------------|-------------------------------------------------------------------------------------------------------------|------|---------|
| 总览                                          |                                                                                                             |      |         |
| ⑦ 本地模型服务 ^<br>添加模型服务<br>模型服务管理              | ▲<br>主乱名                                                                                                    |      |         |
| ◎ 我的摄像头 ^<br>添加摄像头                          | <sup>^</sup> Ⅰ 黄源和用物说                                                                                       |      |         |
| 摄像头管理<br>E 本地事件 ^                           | <ul><li>▲ 6.4%</li><li>▲ 73.8%</li></ul>                                                                    |      |         |
| 本地事件记录                                      | CPU時間兼 並行內容                                                                                                 |      |         |
| <ul> <li>○ 设置</li> <li>&gt; 系统设置</li> </ul> | <ul> <li>CPU道号 intel(R) X 运行內存 251.3168 总型尚 13.8778</li> <li>CPU地数 28株 已用內存 11.5668 已用空尚 10.2378</li> </ul> |      |         |
| 运行日志                                        |                                                                                                             |      |         |

2. 在百度智能云的控制台中找到SDK对应的管理列表,图中以服务器SDK为例。申请序列号后,点击对应序列号尾部的「离线激活」操作,按指 引激活

| < 返回EasyDL总览                 | 产品服务 / EasyDL图像 - | 服务器纯离线服务    |      |        |       |            |            |             |
|------------------------------|-------------------|-------------|------|--------|-------|------------|------------|-------------|
| EasyDL图像                     | 服务器纯离线服务          | 管理          |      |        |       |            |            |             |
| 公有云部署 ~                      | API SD            |             |      |        |       |            |            |             |
| EasyEdge本地部署 へ               |                   |             |      |        |       |            |            |             |
| <ul> <li>服务器线高线服务</li> </ul> | 使用说明:(展开道         | E看部署介绍及操作步骤 | 等)   |        |       |            |            | ◇ 履开        |
| • 设备端纯高线服务                   | + 购买正式授权          | 新培测试序列号 [   | ?    |        |       |            |            | 开发文档 工单支持   |
| 。 专项硬件纯离线                    | 设备名               | 技术方向        | 模型ID | 可激活模型数 | 序列号类型 | 序列号 激活时间   | 到期日期       | 操作          |
|                              | 自定义设备 🖉           | 122-18      |      | 1      | 试用    | 2021-03-09 | 永久         | 申请延期        |
|                              | 自定义设备 🖉           | 图像          |      | 10     | 试用    | 2021-09-16 | 2021-10-16 | 申请延期        |
|                              | 自定义设备 🖉           | 田像          |      | 10     | 试用    | 未激活        | 激活后30天内有效  | 联网激活 离线激活   |
|                              | 自定义设备 🖉           | 122188      |      | 10     | 试用    | 未激活        | 激活后30天内有效  | 联网激活 南线激活   |
|                              | 自定义设备 🖉           | 图像          |      |        | 试用    | 朱激活        | 激活后31天内有效  | 联网激活   离线激活 |
|                              | 自定义设备 🧷           | 图像          |      |        | 试用    | 未激活        | 激活后31天内有效  | 联网激活丨离线激活   |
|                              |                   |             |      |        |       |            | 每百是元       | 10 × 1 1    |
|                              |                   |             |      |        |       |            | 400 EU     |             |
|                              |                   |             |      |        |       |            |            |             |
|                              |                   |             |      |        |       |            |            |             |

3. 在IEC的添加模型服务页面,上传下载好的授权文件,完成激活



激活完成后即可在「模型服务管理」列表中启动服务,使用后续的操作栏功能。

序列号激活错误码

| 错误码  | 文案                                                        | 描述                            |
|------|-----------------------------------------------------------|-------------------------------|
| 4001 | parameters missing. 参数缺失                                  |                               |
| 4002 | parameters invalid. 参数不合法                                 |                               |
| 4003 | model invalid. 模型信息不合法                                    |                               |
| 4004 | no more activation left. 该序列号和该设备的激活次数超上限                 |                               |
| 4005 | the serial key is out of date. 该序列号过期                     |                               |
| 4006 | the serial key has been activated. 该序列号已被其他设备激活           | 该序列号已被其他设备激活,不能重复激活。          |
| 4007 | account invalid. 序列号不能用于其他账号的模型                           | 序列号不能用于其他账号的模型,只能用于绑定账号的模型。   |
| 4008 | serial key invalid. 序列号不合法                                | 序列号不存在或找不到                    |
| 4009 | bundle id invalid. 包名不合法                                  |                               |
| 4010 | product invalid. 产品不合法                                    | 如easydl的SDK使用BML的序列号来激活,会报该错误 |
| 4011 | platform invalid. 平台不合法                                   |                               |
| 4012 | activate too frequent. 激活太频繁                              | 激活太频繁,请稍后再进行激活。               |
| 4013 | device type and license type not match. 硬件类型和序列号类型不匹<br>配 | 如使用加速版序列号激活基础版SDK会报该错误        |
| 4014 | exceed max activate device num. 超过最大授权数量                  |                               |
| 4015 | technology invalid. 技术类型不合法                               |                               |
| 4016 | exceed max activate entity num. 超过最大模型数量                  |                               |
| 4017 | device invalid. 设备不合法                                     |                               |
| 4018 | model invalid. 模型不合法                                      |                               |

## 体验本地demo





## 接入摄像头

使用接入摄像头功能首先需要添加摄像头并创建告警规则,请参考第②步和第③步,完成后按照第④步操作

注:服务启动后也可参考「模型发布」模块的技术文档进行开发使用,本文档主要介绍IEC使用功能

| EasyEdge 智能边缘控制   | 制台-单节点版   |                 |               |      |      |      |    |  |
|-------------------|-----------|-----------------|---------------|------|------|------|----|--|
| 总览                | 模型服务管理    |                 |               |      |      |      |    |  |
| ○本地模型服务<br>添加模型服务 | 新埔服务 查看序列 | <del>9</del>    |               |      |      |      |    |  |
| 模型服务管理            | 服务名称      | 模型名称            | 应用平台          | 服务来源 | 服务状态 | 服务开关 | 操作 |  |
| 登 我的摄像头           |           | p2-检测-高性能-lowV1 | Linux—通用x86芯片 |      |      |      |    |  |
| 添加摄像头             |           |                 |               |      |      |      |    |  |
| 摄像头管理             |           |                 |               |      |      |      |    |  |
| 囟 本地事件            |           |                 |               |      |      |      |    |  |
| 本地事件记录            |           |                 |               |      |      |      |    |  |
| ◎ 设置              |           |                 |               |      |      |      |    |  |
| 系统设置              |           |                 |               |      |      |      |    |  |
| 运行日志              |           |                 |               |      |      |      |    |  |
|                   |           |                 |               |      |      |      |    |  |
|                   |           |                 |               |      |      |      |    |  |
|                   |           |                 |               |      |      |      |    |  |
|                   |           |                 |               |      |      |      |    |  |
|                   |           |                 |               |      |      |      |    |  |
|                   |           |                 |               |      |      |      |    |  |
| 版本号: V2.2.2       |           |                 |               |      |      |      |    |  |

₯②添加摄像头

导航栏点击「我的摄像头」-「添加摄像头」,定义摄像头名称、备注后即可添加摄像头。支持本地摄像头、网络摄像头、ONVIF协议摄像头和 GB28181协议摄像头。 摄像头添加成功后即可设置摄像头的运行时间和频率

| EasyEdge 智能边缘控制    | 制台-单节点版                                                                                              | 操作指南 💿 ad | min ~ |
|--------------------|------------------------------------------------------------------------------------------------------|-----------|-------|
| 总览                 | <b>添加强度</b> 失                                                                                        |           |       |
| ② 本地模型服务<br>添加模型服务 |                                                                                                      |           |       |
| 模型服务管理             |                                                                                                      |           |       |
| 登 我的摄像头<br>添加摄像头   | ◎◎決夫型 ◆物理療法 ③ 開始提倡法 CWNF GE23181                                                                     |           |       |
| 摄像头管理              |                                                                                                      |           |       |
| 二 本地事件<br>本地事件记录   |                                                                                                      |           |       |
| ◎ 设置<br>系统设置       | + #art-st<br>settiste tage to C 19 - Hat t M                                                         |           |       |
| 运行日志               | 温馨很示:1.这作时间和贴结频率的包置高层场先进始继续来后。保存推进结果的进行方式,推进结果可在事件记录中宣着。<br>2.成此的设置不多种描意头实对预见服务推理,描意头实对预定可在描意头提唱媒认意着 |           |       |
|                    | 21                                                                                                   |           |       |
| 版本号: V2.2.2        |                                                                                                      |           |       |

の ③创建告警规则

导航栏点击「本地事件」-「本地事件记录」,切换至「告警规则管理」tab。点进创建告警规则

| EasyEdge 智能边缘控制         | 则台—单节点版     |        |         |         |      |           | 操作指南 | admin 🔻 | v |
|-------------------------|-------------|--------|---------|---------|------|-----------|------|---------|---|
| 总览                      | 本地事件记录      |        |         |         |      |           |      |         |   |
| ☞ 本地模型服务                | 事件记录 告誓规则管理 |        |         |         |      |           |      |         |   |
| 添加模型服务<br>模型服务管理        | 创建新餐校则      |        |         |         |      |           |      |         |   |
| 登 我的摄像头                 | 音豐名称        | 关联模型服务 | 关联团像头数量 | 视频事件总数量 | 雪豐描述 | esentri o | 1811 |         |   |
| 滚加摄像头<br>               |             |        |         |         |      |           |      |         |   |
| <b>加藤大田</b> 在<br>〇 本地事件 |             |        |         |         |      |           |      |         |   |
| 本地事件记录                  |             |        |         |         |      |           |      |         |   |
| ◎ 设置                    |             |        |         |         | 暂无数据 |           |      |         |   |
| 系统设置                    |             |        |         |         |      |           |      |         |   |
| JE11 D 45               |             |        |         |         |      |           |      |         |   |
|                         |             |        |         |         |      |           |      |         |   |
|                         |             |        |         |         |      |           |      |         |   |
|                         |             |        |         |         |      |           |      |         |   |
|                         |             |        |         |         |      |           |      |         |   |
| 版本号: V2.2.2             |             |        |         |         |      |           |      |         |   |

选择要关联的模型服务,并配置产生告警不同标签需要满足的阈值条件

| EaseEdge 智能边缘控制台-单节点版 图作 |              |                          |  |  |  |  |  |  |  |  |
|--------------------------|--------------|--------------------------|--|--|--|--|--|--|--|--|
| 总览                       | 事件记录 > 新建告答# | <b>投約</b>                |  |  |  |  |  |  |  |  |
| ◎ 本地模型服务                 | 基本信息         |                          |  |  |  |  |  |  |  |  |
| 添加模型服务                   |              |                          |  |  |  |  |  |  |  |  |
| 模型服务管理                   |              |                          |  |  |  |  |  |  |  |  |
| ☺ 找的摄像头                  |              |                          |  |  |  |  |  |  |  |  |
| 添加摄像头                    |              |                          |  |  |  |  |  |  |  |  |
| 摄像头管理                    |              |                          |  |  |  |  |  |  |  |  |
| 歯 本地事件                   |              |                          |  |  |  |  |  |  |  |  |
| 本地事件记录                   |              |                          |  |  |  |  |  |  |  |  |
| ◎ 设置                     | 1 配置信息       |                          |  |  |  |  |  |  |  |  |
| 系統设置                     |              | 事件告置条件 条件关系 <sup>①</sup> |  |  |  |  |  |  |  |  |
| 运行日志                     |              |                          |  |  |  |  |  |  |  |  |
|                          |              |                          |  |  |  |  |  |  |  |  |
|                          |              |                          |  |  |  |  |  |  |  |  |
|                          |              |                          |  |  |  |  |  |  |  |  |
|                          |              |                          |  |  |  |  |  |  |  |  |
|                          |              |                          |  |  |  |  |  |  |  |  |
|                          |              |                          |  |  |  |  |  |  |  |  |
| 版本号: V2.2.2              | 完成           | <b>取</b> 3角              |  |  |  |  |  |  |  |  |

⑦④④摄像头接入模型服务预测

点击「本地模型服务」-「模型服务管理」中,所需接入预测的服务的「接入摄像头」

| EasyEdge 智能边缘控制台 操作指病 💿 |  |            |                        |                   |      |      |                     |    |  |  |
|-------------------------|--|------------|------------------------|-------------------|------|------|---------------------|----|--|--|
| 总宽                      |  | 模型服务管理     |                        |                   |      |      |                     |    |  |  |
| ③ 本地模型服务 /              |  | 新鐵服务 查看库列号 |                        |                   |      |      |                     |    |  |  |
| 添加模型服务                  |  |            |                        |                   |      |      |                     |    |  |  |
| 模型服务管理                  |  | 服务名称       | 模型名称                   | 应用平台              | 服务状态 | 服务开关 | 操作                  |    |  |  |
| ◎ 我的摄像头 /               |  | 晓宇加速版      | e7.5检测高性能Auto分<br>辨率V1 | Linux-通用X86 CPU加速 | 运行中  |      | 查看详情 体验本地demo 提入摄像头 | 删除 |  |  |
| 添加摄像头                   |  | zyf_胡乱测试模型 | 测试20210902V1           | Linux-通用X86 CPU   | 运行中  |      |                     |    |  |  |
| 摄像头管理                   |  |            |                        |                   |      |      |                     |    |  |  |
| □ 本地事件 /                |  |            |                        |                   |      |      |                     |    |  |  |
| *************           |  |            |                        |                   |      |      |                     |    |  |  |

在弹出的弹窗中选择第②步中添加的摄像头,并选择第③步中创建的告警规则,此时点击确认即可在「摄像头管理」中的实时预览功能中查看摄 像头预测结果

| 接入摄像头  |      |    |    | ×   |
|--------|------|----|----|-----|
| fruit; |      | 添加 |    |     |
| 摄像头名称  | 告警规则 |    |    | 操作  |
| fruit  | 水果告警 |    |    |     |
|        |      |    |    |     |
|        |      |    | 取消 | 确 定 |

心 ⑤本地事件

点击导航栏「本地事件记录」,可通过服务名称、摄像头名称、事件记录的时间、标签及置信度来筛选识别结果查看,多个标签及置信度同样也 是"或"的逻辑记录。如有想要删除的事件数据可选择后删除,全选为本页全选。



₯⑥连接到智能边缘控制台-多节点版(IECC)

与中心节点连接之后,边缘节点主程序版本会自动随控制中心版本升级。(>2.0.0)

• Step 1 在IECC中添加边缘节点,选择「边缘节点已安装IEC」,并记录IP地址与Token

| EasyEdge 智能边缘控制              | 1台-多节点版                                                                            |
|------------------------------|------------------------------------------------------------------------------------|
| 总览                           | 边缘节点 > 添加边缘节点                                                                      |
| ⑦ 模型服务中心<br>模型服务管理<br>下发模型服务 | 添加边缘节点<br>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>                                      |
|                              | 激活方式 ○ 边缘节点未安装IEC                                                                  |
|                              | 建每期小人共适的进行。T-L                                                                     |
| 摄像头设备                        | 用装約中心  2元  7地址、IOK的世物,住起体及世文表的)  CG-光幻及重火固 近1)加占。第FF元の自身性  7)近1)な法,代理超过印成功加占起体  元。 |
|                              | 日日1232.191.23.8602                                                                 |
| 事件记录                         |                                                                                    |
|                              | Token b00258b3-6c04-4696-b4dc-e8dd83de941d                                         |
| 系统设置                         | <<br>上一步 校验激活结果                                                                    |
|                              |                                                                                    |

• Step 2 在IEC的系统设置中打开多节点模式,并填入刚才记录的IP地址与Token,点击建立连接

| EasyEdge 智能边缘控制              | 制中心                                                      |
|------------------------------|----------------------------------------------------------|
| 总览                           | 系统设置                                                     |
| ① 本地模型服务<br>添加模型服务<br>模型服务管理 | *事件保存时限 30 ☆ 天<br>*文件保存路径 /opt/baidu/easyedge/iec/events |
|                              | 系统时间显示 ③                                                 |
| 添加摄像头                        | 多节点模式 💿 👥                                                |
| 摄像头管理                        |                                                          |
|                              | 中心节点连接状态 • 未连接(IP地址和Token,讲在"多节点版"的"添加边缘节点"第二步中查看)       |
| 本地事件记录                       |                                                          |
|                              |                                                          |
|                              |                                                          |
| 运行日志                         |                                                          |

• 连接完成后即可在中心节点IECC去监控/管理/应用在边缘节点上的IEC

#### ⑦⑦多媒体服务器使用硬件编解码

点击导航栏「设置」-「系统设置」,会自动检测当前硬件是否支持开启硬件编解码,如支持则可开启

| EasyEdge 智能边缘控制   | 则台单节点版    |                             |     |        |  |  | 操作指南 |  |
|-------------------|-----------|-----------------------------|-----|--------|--|--|------|--|
|                   | 系统设置      |                             |     |        |  |  |      |  |
| ①本地模型服务<br>添加模型服务 |           |                             |     |        |  |  |      |  |
| 模型服务管理            |           |                             |     |        |  |  |      |  |
|                   |           |                             |     |        |  |  |      |  |
| 添加摄像头             |           | + 新端AK/SK ①                 |     |        |  |  |      |  |
| 摄像头管理             |           | AK.                         | cr. | 58.4h  |  |  |      |  |
|                   |           |                             |     | 19811- |  |  |      |  |
| 本地事件记录            |           | 56592b34-ada1-4bc1-b5f2-8c8 | Ø   |        |  |  |      |  |
|                   |           |                             |     |        |  |  |      |  |
| 系统设置              |           |                             |     |        |  |  |      |  |
|                   |           |                             |     |        |  |  |      |  |
|                   |           |                             |     |        |  |  |      |  |
|                   |           |                             |     |        |  |  |      |  |
|                   |           |                             |     |        |  |  |      |  |
|                   | 硬件编解码 🎱 🔅 |                             |     |        |  |  |      |  |
|                   |           |                             |     |        |  |  |      |  |
|                   |           |                             |     |        |  |  |      |  |

心 配置项\*

配置文件easyedge-iec.yml中有关于IEC的各项配置说明,一般无需修改,请确保理解配置项含义之后,再做修改。

#### IEC系统配置 #### ------ 高级配置一般无需修改 -------#### !!!注意!!! 请确保理解配置项含义后再做修改 version: 3 com: # hub: 作为中心节点模式启动。 edge: 作为子节点启动 # role: edge #硬件利用率刷新时间间隔: 过低的刷新间隔可能会消耗CPU资源。 hardwareUsageRefreshSecond: 60 #事件监测触发扫描周期 eventTriggerIntervalSecond: 10 # IEC保存SDK等资源的路径:可填写 "default",也可以直接填写绝对路径 appDataFolder: /var/lib/easyedge-iec/appdata # 是否开启DEBUG模式:开启之后,将会打印大量日志,便于追踪系统问题 debug: no # 是否需要用户名/密码登陆,强烈建议打开!(默认用户名/密码为 admin/easyedge) loginEnabled: yes #登录有效时间,单位秒 sessionMaxAge: 3600 logging: # 是否把日志打印到控制台 toStd: no # 是否把日志输出到文件。关闭后,将无法在页面中正确显示系统日志 toFile: yes # loggingFile: /var/log/easyedge-iec/easyedge-iec.log loggingFolder: /var/log/easyedge-iec/ # 0:info; -1:debug; -2:verbose 。设置为<-1时, SDK HTTP服务会输出DEBUG日志 level: -1 webservice: # WEB服务的监听端口 listenPort: 8702 listenHost: 0.0.0.0 sdk: # GPU SDK所使用的cuda版本: auto(自动检测) / 9 / 10 / 10.2 / 11.0 / 11.1。如果auto检测不正确,可以强制指定版本号。更换版本只 对新添加的模型有效 cudaVersion: auto # AI服务启动时, 额外配置的 LD\_LIBRARY\_PATH(linux) 或者 PATH(windows) libPath: ./ # AI服务启动时,额外配置的其他环境变量。 ENVs: EDGE\_CONTROLLER\_KEY\_LOG\_BRAND: EasyEdge #### EDGE\_CONTROLLER\_KEY\_XXX: XXXX commu: # 普通消息等待respond的超时时间 respondWaitTimeoutSecond: 2 #### 数据库相关配置 db: sqliteDbFile: /var/lib/easyedge-iec/easyedge-iec.db hubDbFile: /var/lib/easyedge-iec/easyedge-iec.hub.db eventDbFile: /var/lib/easyedge-iec/easyedge-event.db fileServerDbFile: /var/lib/easyedge-iec/easyedge-fileserver.hub.db nodeMonitorDbFile: /var/lib/easyedge-iec/easyedge-nodemonitor.hub.db #### 推流相关配置 mediaserver: flvPort: 8715

#### 视频流相关配置

rtmpPort: 8716

```
cugeou cam.
# FATAL 1, ERROR 2, WARNING 3, INFO 4, DEBUG 5, VERBOSE_LEVEL1 6, VERBOSE_LEVEL2 7, VERBOSE_LEVEL3 8
logLevel: 5
listenHost: 0.0.0.0
listenPort: 8710
#摄像头预览:识别结果绘制延迟消失
renderExtendFrames: 10
# 预测队列大小: 如果设置为60,当摄像头fps=30时,视频延迟约为2秒。 降低inferenceQueueSize可以降低预览延迟,但是根据硬件的算
力情况,可能导致模型推理速度跟不上,没有识别结果,不建议设置太低
inferenceQueueSize: 60
videoEncodeBitRate: 400000
#视频采样 & 视频实时预览分辨率设置
# 0: auto, 1: 1080p, 2: 720p, 3: 480p, 4: 360p, 5: 240p
resolution: 0
# 内置多媒体服务配置
# port设为0表示关闭
mediaServerFlvPort: 8713
mediaServerRtmpPort: 8714
mediaServerRtspPort: 0
mediaServerRtpPort: 8716
#### 信令服务器相关配置
sipserver:
listenHost: 0.0.0.0
listenPort: 8708
# 当前域
region: 370700008
 # 当前服务id
deviceId: 370700008200800001
# 用户id固定头部
uid: 3707000081118
# 设备id固定头部
did: 3707000081318
loadbalance
HTTPPort: 8780
TCPPortMin: 30000
TCPPortMax: 31000
UDPPortMin: 30000
UDPPortMax: 31000
```

```
രം FAQ
```

启动服务后,进程中出现两个easyedge-iec进程这是正常现象,IEC通过守护进程的方式来完成更新等操作。

启动服务时,显示端口被占用port already been used 通过修改 easyedge-iecc.yml文件的配置后,再重新启动服务。

安装服务时,报错permission denied 请以管理员身份运行安装程序。

**中心节点重启后,边缘节点IEC 一直离线**中心节点短时间的离线,边缘节点会自动重连。如果中心节点已经恢复在线,边缘节点长时间未自动连接上,可通过边缘节点iec的方法来重新连接(右上角 admin - 重启系统)

IEC 是否有Android / iOS 版本 我们将会在近期发布对Android操作系统的支持

**添加SDK时,报错 SDK不支持该硬件。 SDK not supported by this device** 一般是因为使用的SDK跟硬件不匹配,如 GPU的SDK,硬件没有GPU 卡。 对于Jetson,也可能是Jetpack版本不支持,可以通过查看 本机Jetpack版本和SDK支持的Jetpack版本列表(cpp文件中的文件名来查看)来 匹配。

## 智能边缘控制台-多节点版

心 EasyEdge 智能边缘控制台——多节点版

心 整体介绍

智能边缘控制台 - 多节点版(EasyEdge Intelligent EdgeConsole Center 以下简称IECC),是EasyEdge推出的边缘资源管理、服务应用与管理一站 式本地化方案。

通过IECC,用户可以方便地在中心节点管理子节点:

- 边缘硬件资源的管理与监控
- EasyDL/BML/EasyEdge的SDK的 离线 / 在线激活,服务管理
- 视频流解析,接入本地和远程摄像头,网页中实时预览
- 自动监控和记录相关视频流推理事件

#### 支持的系统+CPU架构包括:

- Windows x86\_64 (Windows 7 ~ Windows 10, 暂不支持Windows 11)
- Linux x86\_64 / arm32 / arm64

#### 支持各类常见的AI加速芯片,包括:

- NVIDIA GPU / Jetson 系列
- Baidu EdgeBoard FZ系列
- 比特大陆 Bitmain SC / SE 系列
- 华为 Atlas 系列
- 寒武纪 MLU 系列
- 其他EasyDL/EasyEdge/BML支持的AI芯片

#### の 连接说明

以下为中心节点(控制中心),边缘节点/子节点,摄像头的连接示意:



其中:

- 控制中心需要有固定IP,而边缘节点可以处于多级子网之下,只需IEC能够主动访问到控制中心节点即可
- 模型服务均运行于各边缘节点之上
- 摄像头均与边缘节点相连

രം Release Note

| 版本号     | 发布时间       | 更新说明                                         |
|---------|------------|----------------------------------------------|
| 2.2.2.1 | 2023-08-30 | 系统稳定性优化                                      |
| 2.2.2.0 | 2023-03-16 | 支持边缘节点多媒体服务器硬件编解码                            |
| 2.2.1   | 2022-11-15 | 修复告警规则删除问题                                   |
| 2.2.0   | 2022-10-27 | 边缘节点新增Android支持;新增onvif/gb28181支持;优化端云通信通道安全 |
| 2.0.0   | 2022-03-25 | 多节点版上线!                                      |
| 1.0.2   | 2021-12-22 | 更新视频预览推流库;新增若干AI芯片支持;支持多种芯片温度、功耗展示;多项性能优化    |
| 1.0.0   | 2021-09-16 | 智能边缘控制台 - 单节点版 IEC 第一版!                      |

の安装

从这里选择您需要的操作系统和CPU架构下载:

- Windows amd64 : intel、AMD的64位x86\_84 CPU
- Linux amd64: intel、AMD的64位x86\_84 CPU
- Linux arm:树莓派等32位的ARM CPU
- Linux arm64: RK3399、飞腾等64位的ARM CPU

或者从纯离线服务管理页可下载智能边缘控制台

| <i>┥</i> ゚゚゚゚゚゚゚゚ヹ゚゚゚¥EasyDL | 产品介绍 | 操作平台 应用案例 使用文档                                                                                                                     |
|------------------------------|------|------------------------------------------------------------------------------------------------------------------------------------|
| 物体检测模型                       |      | 纯高线服务                                                                                                                              |
| 创建模型                         |      |                                                                                                                                    |
| 训练模型                         |      | 纯离线服务说明                                                                                                                            |
| 校验模型                         |      | 发布纯高线服务,将训练完成的模型部署在本地,高线调用模型。可以选择将模型部署在本地的服务器、小型设备、软硬一体方案专项适配硬件上,<br>通过API STK进一批集成 司话该问太同业各场票:你可了就EasuFranswerth线控制台 使建磅理多个本地进用服务 |
| 发布模型                         |      |                                                                                                                                    |
| ☑ EasyData数据服务               |      | 及他加速方 督航五控制台 ⊻ 下軟督能划续控制台 ①                                                                                                         |
| 数据总览                         |      |                                                                                                                                    |
| 标签组管理                        |      | 服务器 通用小型设备 专项适配硬件                                                                                                                  |
| 在线标注                         |      | SDK API                                                                                                                            |
| 智能标注                         |      |                                                                                                                                    |
| 云服务数据回流                      |      | 此处发布、下载的SDK为未授权SDK,需要前往控制 <del>台获取序列号</del> 激活后才能正式使用。SDK内附有对应版本的Demo及开发文档,开发者可参考源代码完成开发。                                          |





Linux 系统

### 通过系统服务形式安装 (推荐)

以管理员运行bash easyedge-iecc-setup.sh install 即可。

[setup]: sudo could not be found [setup]: Start to install IECC... [setup]: + bash -c "./easyedge-iecc --com.role=hub --service=install" [setup]: Install IECC success! [setup]: + bash -c "/usr/sbin/easyedge-iecc --com.role=hub --service=start" [setup]: + bash -c "curl -s 127.0.0.1:8602 >/dev/null" [setup]: + bash -c "/usr/sbin/easyedge-iecc --com.role=hub --service=status | grep running > /dev/null 2>&1" [setup]: HeCC status: OK! [easyedge-iecc]: default configure file: /etc/easyedge-iecc/easyedge-iecc.yml [easyedge-iecc]: default log dir: /var/log/easyedge-iecc/ [easyedge-iecc]: service usage: service easyedge-iecc { start | stop } [setup]: Done!

出现IECC status: OK!字样,表示安装成功。

- 日志: /var/log/easyedge-iecc/easyedge-iecc.log
- 系统配置: /etc/easyedge-iecc/easyedge-iecc.yml
- 服务启动/停止: service easyedge-iecc { start | stop } (不同操作系统内可能不同,具体命令参考安装日志)
- 配置服务自启动:可根据不同操作系统参考这里进行对应配置

可通过 bash easyedge-iecc-setup.sh uninstall 来卸载,以及bash easyedge-iecc-setup.sh upgrade来升级为当前安装包的版本

#### 自定义安装 (不推荐)

自定义安装仅限于 安装脚本无法识别您的操作系统的情况。

- 拷贝 ./EasyEdge-IECC-v{版本号}-{平台}-{架构}/ 整个目录至自定义文件夹,如/opt/EasyEdge-IECC
- 进入到 /opt/EasyEdge-IECC
- 通过 nohup 等方法运行 ./easyedge-iecc --com.role=hub amd64: intel、AMD的64位x86\_84 CPU arm:树莓派等32位的ARM CPU \* arm64: RK3399、飞腾等64位的ARM CPU
- 日志: ./log/easyedge-iecc.log
- 系统配置: ./etc/easyedge-iec.yml

Windows 系统 打开命令行(非powershell)运行 easyedge-iecc-setup.bat install。

| Administrator: C:\Windows\system32\cmd.exe                                                            |   |                                                 |   |    | - | × |
|-------------------------------------------------------------------------------------------------------|---|-------------------------------------------------|---|----|---|---|
| icrosoft Windows [Version 10.0.17763.2145]<br>c) 2018 Microsoft Corporation。保留所有权利。<br>:\Users\eval>_ | - | 还原(R)<br>移动(M)<br>大小(S)<br>最小化(N)<br>最大化(X)     |   | 右键 |   | ^ |
|                                                                                                       | × | <b>关闭(C)</b><br>编辑(E)<br>默认值(D)<br><u>属性(P)</u> | > |    |   |   |

| 光标大小                | 命令记录            |
|---------------------|-----------------|
| ● 小(S)              | 缓冲区大小(B): 50 🔶  |
| 〇中(M)               | (經)山区数母(N)· 4 ▲ |
| ○大(L)               | x0+Exxx((*).    |
|                     | □ 丢弃旧的副本(D)     |
| 编辑选项                |                 |
| □快速编辑模式(Q) 🗲        | 取消勾选            |
| ☑ 插入模式(I)           |                 |
| ☑ 启用 Ctrl 键快捷方式(    | K)              |
| ✓ 粘贴时筛选剪贴板内容        | ₽(F)            |
| □将 Ctrl+Shift+C/V 用 | 作复制/粘贴的快捷键(C)   |
| 文本选择                |                 |
| ☑ 启用对选定内容自动接        | ¥亍(E)           |
| ✓扩展的文本选择键(X)        |                 |
| 当前代码页               |                 |
| 65001 (UTF-8)       |                 |
| 〕 傅田旧版坊割会(□)(李重     | 再新户动 影响所去场制会)   |
|                     |                 |
| 了解有关旧版控制台槽          | EXHIBIT OF HIGH |
| 了解有关旧版控制台模          |                 |

验证安装:启动之后,打开浏览器,访问 http://{设备ip}:8602 即可:

| EasyEdge 智能边缘控制台-多节点版<br><sup>能能达姆技好5 节点系的空机整整体,正定动表示。我用示法的保证场下来加速。<br/>可在4-0节点#5个项层图与使建设描意以出现在。 用意意的人能能反而以正动所的计上的状态</sup> |                       |
|----------------------------------------------------------------------------------------------------------------------------|-----------------------|
|                                                                                                                            | 本地管理员登录               |
|                                                                                                                            |                       |
|                                                                                                                            |                       |
|                                                                                                                            | #₹                    |
|                                                                                                                            | 操作描明 I 修改图码 原本号: V222 |
|                                                                                                                            |                       |
|                                                                                                                            |                       |
| 相关"是:EasyCL来门楼从中没干台 BML生动能从开发干台 EasyEdge能与这部从医务干台                                                                          |                       |
|                                                                                                                            |                       |

**更新服务:**停止服务,下载最新的安装包,替换二进制可执行文件,启动服务。

注:

- 1. 中心节点更新到新版之后,已连接的边缘节点会自动跟随中心节点,自我升级到同样的版本。
- 2. 报错: Text file busy. 一般是因为服务没有停止。

#### の 使用流程

② ①配置对外服务IP地址

- 中心节点所在机器可能存在多网卡,由于边缘节点需要连接中心节点,因此需要配置对外服务IP地址,确定唯一的对外IP
- 导航栏点击「设置」-「系统设置」,设置对外服务IP地址,候选列表为扫描到的IP地址

| EasyEdge 智能边缘控制                                                                                                                                            | 刘台-多节点版                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 操作指南 | admin ~ |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|---------|
| 总览                                                                                                                                                         | 系统设置                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |      |         |
| <ul> <li>○ (供加服务中心)</li> <li>(成加服务管理)</li> <li>下記(供加服务)</li> <li>(以注意清中心)</li> <li>(以注意清中心)</li> <li>(以注意市高)</li> <li>(損償失役者)</li> <li>(二) 第件中心</li> </ul> | Phyle語ameteix 31000 C     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S     S |      |         |
| 事件记录                                                                                                                                                       | 接口服权 <b>+ 新和AK/SX O</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |      |         |
| 系統设置                                                                                                                                                       | AK SK 操作                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |      |         |
| 述(*)日 <b>志</b>                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |      |         |
| 版本号: V2.2.2                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |      |         |

№ ②注册并激活边缘节点

• 导航栏点击「边缘资源中心」-「边缘节点」,点击页面中的「添加边缘节点」按钮

| EasyEdge智能边缘控制台 | -多节点版        |                         |                     |                 |      |        |       |                     |    |
|-----------------|--------------|-------------------------|---------------------|-----------------|------|--------|-------|---------------------|----|
| 总宽              | 边缘节点         |                         |                     |                 |      |        |       |                     |    |
|                 | 已添加的边缘节点     |                         |                     |                 |      |        |       | 漆加边缘节点              |    |
| 下发模型服务          | 边缘节点名称       | 节点状态                    | 节点标识                | CPU             | AI芯片 | 模型服务数量 | 摄像头数量 | 服务添加时间              | 操作 |
| ⊜ 边缘资源中心        | x86-test     | <ul> <li>在线</li> </ul>  |                     | Intel(R) Core(T |      |        |       |                     |    |
| 边缘节点<br>摄像头设备   | 10.138.33.20 | <ul> <li> 高线</li> </ul> | dry_persona<br>Lcpu | Intel(R) Xeon(R |      |        |       | 2022-03-23 19:30:00 |    |

- 注册边缘节点,填写基本信息
- 激活边缘节点,根据边缘节点上是否安装智能边缘控制台-单节点版(IEC)分两种激活方式
  - 边缘节点未安装IEC:复制提供的命令,在边缘节点的终端中输入执行(命令会自动在当前目录,下载单节点版IEC并注册到控制中心)。终端命令执行完成后,在下方校验激活结果,如结果通过即可完成边缘节点的激活

| EasyEdge 智能边缘控制              | 则台-多节点版                                                                                                                                                                                                                                                                                                                                                                           |  |
|------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|                              | 边缘节点 > 深加边缘节点                                                                                                                                                                                                                                                                                                                                                                     |  |
| ◎ 模型服务中心<br>模型服务管理<br>下发模型服务 | 1601218 Т.A<br>С) авсата — (2) явсата                                                                                                                                                                                                                                                                                                                                             |  |
|                              | 加造方式 🕑 <b>边缘节点来安积EC / 边缘节点已安积EC</b>                                                                                                                                                                                                                                                                                                                                               |  |
| 提集 DM<br>摄像失设备               |                                                                                                                                                                                                                                                                                                                                                                                   |  |
|                              | Widewid         Lexx         Addroid           • 978.0 fb;                                                                                                                                                                                                                                                                                                                        |  |
|                              | curl - valit, http://10.127.3199.BMC2/asuryedgy.htearwyr/v/i/downicaal/asuryedgy.htearwyr/v/<br>coetag an A& bann sethigan Imaal, from jaco 10.127.31988002 / aswyddy.htearwyr/vi<br>eybybdichull.21910002/2010000/2010.clavA/GUV/10210.agudeAA/GE20TDHINGE3M/Gan/via<br>sARGebrilwo230CE84/vg20vFormus standard      saftig B+<br>bann / sur/sani/aswydga-soc-setug ot unrastall |  |
|                              |                                                                                                                                                                                                                                                                                                                                                                                   |  |
| 版本号: V2.2.2                  |                                                                                                                                                                                                                                                                                                                                                                                   |  |

• 边缘节点已安装IEC:记录页面中提供的IP地址和Token

| EasyEdge 智能边缘控制                                                      | 川台-多节点版                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |  |
|----------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| 总览                                                                   | ioiste a la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de la constante de |  |
| ② 根型服务中心<br>模型服务管理<br>下发模型服务                                         | 活動2018 节点<br>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |  |
| ◎ 边缘波源中心                                                             | 激活方式 ○ 边缘节点未变极ec ④ 边缘节点已变极ec                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |  |
| 边缘节点 摄像失设备                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |  |
| <ul> <li>二 事件中心</li> <li>事件记录</li> <li>○ 设置</li> <li>系统设置</li> </ul> | P1681_10.12.73.69.8602<br>Towin wyhododullucineamig2058Hs/204mihnani0220ILC.0xXA0UxY/10103 wyae4A0je20TM/NDESMj0al<br>mVIaWOOLMAD/NYNEMANowligDLT052DU/01FLMACOYNLICZVINkg3NCMT0.KVW/25520I-3MK8/EMINA/E<br>X05594Vg22Virhanus                                                                                                                                                                                                                                                                                                                           |  |
| <b>运行日</b> 卷                                                         | 2-19 Bit materia                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |  |
| 版本号: V2.2.2                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |  |

• 在边缘节点的IEC-系统设置中,打开多节点模式开关,将刚才记录的IP地址和Token填入其中,建立连接

| 总览              | 系统设置                                                 |
|-----------------|------------------------------------------------------|
| ⑦ 本地模型服务 添加模型服务 |                                                      |
| 模型服务管理          | *文件保存路径  /opt/baidu/easyedge/iec/events              |
|                 | 系统时间显示 💿 📃                                           |
| 添加摄像头           | 多节点模式 ③                                              |
| 摄像头管理           |                                                      |
|                 | 中心や点连接状态 • 未连接 (IP地址和Lloken,请在"多节点版"的"添加边缘节点"第二步中宣看) |
| 本地事件记录          | 输入IP地址 10.232.191.23:8602                            |
| ③ 设置            | Token 💿                                              |
|                 | 建立连接                                                 |
| 运行日志            |                                                      |
|                 |                                                      |

• 成功激活后可在边缘节点页面中看到一行状态为在线的记录

| EasyEdge智能边缘控制台    | }多节点版    |                                      |      |                 |      |        |       |        | 操作指南 | admin ~ |
|--------------------|----------|--------------------------------------|------|-----------------|------|--------|-------|--------|------|---------|
| 总览                 | 边缘节点     |                                      |      |                 |      |        |       |        |      |         |
| ② 模型服务中心<br>模型服务管理 | 已添加的边缘节点 |                                      |      |                 |      |        |       | 添加边缘节点 |      |         |
| 下发模型服务             | 边缘节点名称   | 节点状态                                 | 节点标识 | CPU             | AI芯片 | 模型服务数量 | 摄像头数量 | 服务添加时间 | 操作   |         |
| 目 边缘资源中心           |          | <ul> <li>         在线     </li> </ul> |      | Intel(R) Xeon(R |      |        |       |        |      |         |

∞③上传并下发模型服务

• 导航栏点击「模型服务中心」-「模型服务管理」,已添加的模型服务页面中点击「添加模型服务」

| EasyEdge智能边缘控制     | 台多节点版     |              |       |         |                     | 操作指南     | admin 🗸 |
|--------------------|-----------|--------------|-------|---------|---------------------|----------|---------|
|                    | 模型服务管理    |              |       |         |                     |          |         |
| ⑦ 模型服务中心<br>模型服务管理 | 已下发的模型服务  | 已添加的模型服务     |       |         |                     | 活动模型服务   |         |
| ト友模型服务             | 服务名称      | 模型名称         | 应用平台  | 已部署节点数量 | 添加时间                | 操作       |         |
| E 10850840         | cpu 加速版模型 | e7.5检测离性能Aut | Linux | 1       | 2022/03/23 19:24:49 | 下发至边缘 删除 |         |

• 上传来自于EasyDL/BML的SDK,目前仅支持Windows/Linux的SDK

| EasyEdge智能边缘控制 | EasyEdge智能边缘控制台–多节点版 |                 |    |  |  |  |  |  |  |
|----------------|----------------------|-----------------|----|--|--|--|--|--|--|
| 总览             | 模型服务管理               | 模型服务管理 > 新建模型服务 |    |  |  |  |  |  |  |
| ☞ 模型服务中心       | 添加模型服纸               | 务               |    |  |  |  |  |  |  |
| 模型服务管理         | *服务名称                |                 |    |  |  |  |  |  |  |
| 下发模型服务         |                      |                 |    |  |  |  |  |  |  |
| 😂 边缘资源中心       | * SDK 🕐              | 土 选择SDK         |    |  |  |  |  |  |  |
| 边缘节点           |                      | 添加              | 取消 |  |  |  |  |  |  |
| 摄像头设备          |                      |                 |    |  |  |  |  |  |  |

- 添加成功后可在已添加的模型服务页面查看添加的模型服务SDK
- 在模型服务SDK上传成功以及边缘节点也添加激活过后,即可将模型服务下发至边缘。点击导航栏-下发模型服务,选择已添加的模型服务, 选择下发的目标节点(支持多节点批量下发)进行模型服务下发

| EasyEdge智能边缘控制     | 台-多节点版   |                            |         |    |               |        | 操作指南 | a) admin 🗸 |
|--------------------|----------|----------------------------|---------|----|---------------|--------|------|------------|
|                    | 1 服务下发配置 |                            |         |    |               |        |      |            |
| ⑦ 模型服务中心<br>模型服务管理 |          | cpu 加速版模型                  |         |    |               |        |      |            |
| 下发模型服务             |          | cpu 加速版模型<br>CPU Linux 普通版 |         | EX | <b>悲节点(0)</b> |        |      |            |
| □ 边缘资源中心<br>边缘劳占   |          | 节点名称                       | 连接状态 标识 |    | 节点名称 连        | 接状态 标识 |      |            |
| 摄像头设备              |          |                            |         |    |               |        |      |            |
|                    |          |                            |         |    |               |        |      |            |
| 事件记录               |          |                            |         |    |               |        |      |            |
|                    |          |                            |         |    |               |        |      |            |
| 系统设置               |          |                            |         |    |               |        |      |            |

确定下发配置后,填入模型服务在边缘节点联网激活运行的序列号(支持批量导入)即可完成模型服务下发,序列号可在智能云控制台获取。
 离线激活的过程可参考IECC中的具体指引

| EasyEdge智能边缘控制               | 台-多节点版                 |           | 操作指南 🔒 admin 🗸 |
|------------------------------|------------------------|-----------|----------------|
| 总宽                           | <b>横型服务管理</b> ≥ 下发模型服务 |           |                |
| ♀ 模型服务中心<br>模型服务管理<br>下发模型服务 | 下发模型服务<br>② 857382     |           |                |
| 目 边缘资源中心                     | 激活方式 • 联网激活 · 高线激活     | 批量导入序列号 🕲 |                |
| 边缘节点<br>摄像头设备                | 边缘节点1: docker-x88 词    |           |                |
| 回 事件中心                       |                        |           |                |

• 完成上述流程后即可在模型服务管理-已下发的模型服务列表中查看记录,并进行下一步应用功能体验

注:完成此步骤后即可在边缘节点进行二次集成已下发的模型服务,具体的集成方式可在文档-某图像任务类型-模型发布中查找对应的 SDK开发文档进行集成开发

| EasyEdge智能边缘控制台  | 台多节点版       |           |          |      |                         |      |      | 操作指南                    | (a) admin $\vee$ |
|------------------|-------------|-----------|----------|------|-------------------------|------|------|-------------------------|------------------|
|                  | 模型服务管理      |           |          |      |                         |      |      |                         |                  |
|                  | 已下发的模型服务    | 已添加的模型服务  |          |      |                         |      |      | <b>下发模型服务</b> 请输入模型服务名称 |                  |
| 下发模型服务           | 服务名称        | 模型名称      | 服务所在边缘节点 | 应用平台 | 服务状态                    | 添加时间 | 服务开关 | 操作                      |                  |
| 已 边缘资源中心<br>边缘节点 | CPU Linux 普 | e7.5检测高性能 |          |      | <ul> <li>运行中</li> </ul> |      |      |                         |                  |

下发时可以通过高级配置设置服务运行的host和port。若不设置,默认host为0.0.0.0,port为系统随机分配的可用端口

心 ④配置摄像头

Step ④ - ⑦ 描述的是如何使用IECC可视化进行视频流式推理与应用,对此有需求的用户建议详细查看后续步骤内容。如仅需对下发的模型服务进行二次集成的用户无需进行后续操作,参考SDK对应的开发文档进行集成即可

• 首先需要确定边缘节点已经接入物理摄像头,可通过USB插口接入,也可通过RTSP/RTMP流式协议接入。在摄像头设备页面点击添加摄像头 按钮,填写对应的信息添加摄像头。支持设置摄像头的运行时间以及摄像头的抽帧频率

| EasyEdge智能边缘控制   | 制台多节点版       |                                                                                                                                              | 操作指南 | a) admin 🗸 |
|------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------|------|------------|
|                  | 摄像头设备 > 添加摄作 | <b>晚</b> 头                                                                                                                                   |      |            |
|                  | 添加摄像头        |                                                                                                                                              |      |            |
| 模型服务管理<br>下发模型服务 |              |                                                                                                                                              |      |            |
|                  |              |                                                                                                                                              |      |            |
| 边缘节点             |              |                                                                                                                                              |      |            |
|                  |              |                                                                                                                                              |      |            |
| 事件记录             | 摄像头类型        | □ 本地描像头 ● 网络描像头                                                                                                                              |      |            |
| ◎ 设置<br>系统设置     |              |                                                                                                                                              |      |            |
|                  |              |                                                                                                                                              |      |            |
|                  |              | + 新聞→現<br>熱熱類率 毎週 1 0 0 0 → 抽取 1 0 0<br>温馨親示・1.注行封病和結核類率的设置指服务技具服像失活,保存強理結果的运行方式,推環結果可在專件记录中設看。<br>2. 此处的设置不影响晶像失实时预定服务相谋,晶像失实时预定可在晶像头管理模块置看 |      |            |
|                  |              | 完成 取消                                                                                                                                        |      |            |

• 添加完成后可在摄像头设备页面查看记录

| EasyEdge智能边缘控制     | 台多节点版   |       |                         |            |       | 操作指南         | a admin ~ |
|--------------------|---------|-------|-------------------------|------------|-------|--------------|-----------|
| 总览                 | 摄像头设备   |       |                         |            |       |              |           |
| ○ 模型服务中心<br>模型服务管理 | 已添加的摄像头 |       |                         |            | 添加摄像头 | 请输入摄像头名称     |           |
| 下发模型服务             | 摄像头名称   | 摄像头标识 | 连接状态                    | 所在节点名称     | 摄像头开关 | 操作           |           |
| 目 边缘资源中心           | lp cam  |       | <ul> <li>运行中</li> </ul> | docker-x86 |       | 实时预览 查看详情 本地 | 事件记录 删除   |

• 点击预览可查看摄像头预览画面



の ⑤配置告警规则

• 导航栏点击「事件中心」-「事件记录」,切换至「告警规则管理」tab。点进创建告警规则



• 选择要关联的模型服务,并配置产生告警不同标签需要满足的阈值条件

| EasyEdge 智能边缘控制                        | 制台-多节点版                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 操作指南 | a) admin 🗸 |
|----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|------------|
| 总览                                     | ●件记录 > 新建告题规则                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |      |            |
| (#型服务中心<br>模型服务管理<br>下发模型服务            | 基本信息<br>*告贤成刻名曰: ★果告算<br>中学文、哲学取了别性组织、不能以下别性开头 2-50个字符                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |      |            |
| 目 边缘资源中心<br>边缘节点                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |      |            |
| 据像失设备<br>ご 事件中心<br>事件20                | ●关环样监察与: 次集成制<br>结选结构在用点合管规则的保证部务                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |      |            |
| <ul> <li>② 设置</li> <li>系统设置</li> </ul> | NR(6.8                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |      |            |
| 递行日志                                   | ● F R ② E 手 F A 雪 整 A 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 1 3 2 |      |            |
| 版本号: V2.2.2                            | EA EA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |      |            |

∞ ⑥模型服务接入视频流预测

• 点击「本地模型服务」-「模型服务管理」中,所需接入预测的服务的「接入摄像头」

| EasyEdge 智能边缘控制                              | 则台多节点版   |              |          |      |      |        |      |                 |       |  |
|----------------------------------------------|----------|--------------|----------|------|------|--------|------|-----------------|-------|--|
| 总览                                           | 模型服务管理   |              |          |      |      |        |      |                 |       |  |
| ① 模型服务中心<br>模型服务管理                           | 已下发的模型服务 | 已添加的模型服务     |          |      |      | 下发模型服务 |      |                 |       |  |
| 下发模型服务                                       | 服务名称     | 模型名称         | 服务所在边缘节点 | 应用平台 | 服务状态 | 下发时间   | 服务开关 | 操作              |       |  |
| <ul> <li>回 边缘资源中心</li> <li>(初線等点)</li> </ul> | 水果识别     | p2-检测-高性能-lo |          |      |      |        |      | 宣看详情 体验demo 接入摄 | 像头 删除 |  |
| 摄像头设备                                        |          |              |          |      |      |        |      |                 |       |  |
| 曲 事件中心                                       |          |              |          |      |      |        |      |                 |       |  |
| 事件记录                                         |          |              |          |      |      |        |      |                 |       |  |
| ◎ 设置                                         |          |              |          |      |      |        |      |                 |       |  |
| 系统设置                                         |          |              |          |      |      |        |      |                 |       |  |
| 运行日志                                         |          |              |          |      |      |        |      |                 |       |  |
|                                              |          |              |          |      |      |        |      |                 |       |  |
|                                              |          |              |          |      |      |        |      |                 |       |  |
|                                              |          |              |          |      |      |        |      |                 |       |  |
|                                              |          |              |          |      |      |        |      |                 |       |  |
|                                              |          |              |          |      |      |        |      |                 |       |  |
| 版本号: V2.2.2                                  |          |              |          |      |      |        |      |                 |       |  |

• 在弹出的弹窗中选择第④步中添加的摄像头,并选择第⑤步中创建的告警规则,此时点击确认即可在「摄像头管理」中的实时预览功能中查看

|         | 接入摄像头  |      |    |     |     |
|---------|--------|------|----|-----|-----|
| 摄像头预测结果 | fruit; | ~    | 添加 |     |     |
|         | 摄像头名称  | 告警规则 |    |     | 操作  |
|         | fruit  | 水果告警 |    |     |     |
|         |        |      |    |     |     |
|         |        |      |    | 取 消 | 确 定 |

## の⑦视频事件告警

• 可在事件中心-事件记录中查看满足时间告警条件的图片记录

| EasyEdge智能边缘控制     | 台-多节点版                                                   | 操作指南 | a) admin 🗸 |
|--------------------|----------------------------------------------------------|------|------------|
| 总览                 | 本地事件记录                                                   |      |            |
| ⑦ 根型服务中心<br>模型服务管理 | 法释迟续节点 deckerx86 ◇ 法择服务 编选择服务 ◇ 法择服务夫 请击联盟继兵 ◇ 时间范围 开始 - |      |            |
| 下发模型服务             | - 新弦涛逸 ~                                                 |      |            |
| ◎ 边缘资源中心           | 事件概刻 事件自动向新 ② 💽                                          | 🗌 全选 | 自 制除       |
| 边缘节点<br>摄像头设备      |                                                          |      |            |
| (回 事件中心)           |                                                          |      |            |
| 事件记录               |                                                          |      |            |

心 高级配置说明

在系统设置 - 高级,可以修改控制中心的高级系统配置

#### IECC 控制中心系统配置 #### ------- 高级配置一般无需修改 ------#### !!!注意!!! 请确保理解配置项含义后再做修改 version: 3 com: # hub: 作为中心节点模式启动。 edge: 作为子节点启动 role: hub #硬件利用率刷新时间间隔: 过低的刷新间隔可能会消耗CPU资源。 hardwareUsageRefreshSecond: 60 # IECC保存SDK等资源的路径:可填写 "default", 也可以直接填写绝对路径 appDataFolder: /var/lib/easyedge-iecc/appdata # 是否开启DEBUG模式:开启之后,将会打印大量日志,便于追踪系统问题 debug: no # 是否需要用户名/密码登陆,强烈建议打开!(默认用户名/密码为 admin/easyedge) loginEnabled: yes #登录有效时间,单位秒 sessionMaxAge: 3600 logging: # 是否把日志打印到控制台 toStd: no # 是否把日志输出到文件。关闭后,将无法在页面中正确显示系统日志 toFile: yes # loggingFile: /var/log/easyedge-iecc/easyedge-iecc.log loggingFolder: /var/log/easyedge-iecc/ # 0:info; -1:debug; -2:verbose level: -1 webservice: #WEB服务的监听端口 listenPort: 8602 listenHost: 0.0.0.0 commu: mgServer: host: 0.0.0.0 port: 8632 HTTPPort: 8620 maxPayload: 8388608 pingIntervalSecond: 30 # 普通消息等待respond的超时时间 respondWaitTimeoutSecond: 2 nodeRefreshIntervalSecond: 30 #### 数据库相关配置 db: sqliteDbFile: /var/lib/easyedge-iecc/easyedge-iecc.db hubDbFile: /var/lib/easyedge-iecc/easyedge-iecc.hub.db eventDbFile: /var/lib/easyedge-iecc/easyedge-event.db hubEventDbFile: /var/lib/easyedge-iecc/easyedge-event.hub.db fileServerDbFile: /var/lib/easyedge-iecc/easyedge-fileserver.hub.db nodeMonitorDbFile: /var/lib/easyedge-iecc/easyedge-nodemonitor.hub.db #### 推流相关配置 mediaserver: flyPort: 8613 rtmpPort: 8614 #### 文件服务器相关配置 fileserver:

fileserver: root: /var/lib/easyedge-iecc/fs

## രം FAQ

启动服务后,进程中出现两个easyedge-iecc进程这是正常现象,IECC通过守护进程的方式来完成更新等操作。

启动服务时,显示端口被占用port already been used 通过修改 easyedge-iecc.yml文件的配置后,再重新启动服务。

安装服务时,报错permission denied 请以管理员身份运行安装程序。

**添加SDK时,报错 SDK不支持该硬件。 SDK not supported by this device** 一般是因为使用的SDK跟硬件不匹配,如 GPU的SDK,硬件没有GPU 卡。对于Jetson,也可能是Jetpack版本不支持,可以通过查看 本机Jetpack版本和SDK支持的Jetpack版本列表(cpp文件中的文件名来查看)来 匹配。

EasyEdge 智能边缘控制台-单节点版 IEC API

# 心概述

欢迎使用EasyEdge 智能边缘控制台-单节点版 IEC。

您可以使用本文档介绍的API对单节点版 IEC服务进行操作。

#### 心 接口概览

单节点版 IEC API 提供下列接口类型:

| 接口类型    | 描述                     |
|---------|------------------------|
| AI服务接口  | 接口包括查询、启动、停止等          |
| 摄像头设备接口 | 接口包括创建、查询、更新、删除、启动、停止等 |
| 物联网设备接口 | 接口包括创建、查询、更新、删除等       |
| 其他接口    | 包括生成鉴权用的Access Token等  |

#### 心 通用说明

API调用遵循HTTP协议。数据交换格式为JSON,所有request/response body内容均采用UTF-8编码。

#### の API认证机制

所有API的安全认证一律采用Access Key与请求签名机制。 Access Key由Access Key ID和Secret Access Key组成,均为字符串。 对于每个HTTP请求,需先调用生成Access Token的接口生成认证字符串。提交认证字符串放在query里。服务端根据生成算法验证认证字符串的正确性。

当服务端接收到用户的请求后,系统将使用相同的SK和同样的认证机制生成认证字符串,并与用户请求中包含的认证字符串进行比对。如果认证 字符串相同,系统认为用户拥有指定的操作权限,并执行相关操作;如果认证字符串不同,系统将忽略该操作并返回错误码。

#### 心 通信协议

支持HTTP调用方式。

心请求结构说明

数据交换格式为JSON,所有request/response body内容均采用UTF-8编码。

请求参数包括如下4种:

| 参数类型        | 说明                                                    |
|-------------|-------------------------------------------------------|
| URI         | 通常用于指明操作类型,如:POST<br>/iec/iapi/v{version}/{type}/{op} |
| Query参数     | URL中携带的请求参数,通常用来传入认证字符串                               |
| HEADER      | 通过HTTP头域传入                                            |
| RequestBody | 通过JSON格式组织的请求数据体                                      |

#### <sup>の</sup>响应结构说明

响应值分为两种形式:

| 响应内容             | 说明                |
|------------------|-------------------|
| HTTP STATUS CODE | 如200,400,403,404等 |
| ResponseBody     | JSON格式组织的响应数据体    |

| 参数      | 类型     | 参数位置  | 描述           | 是否必须 |
|---------|--------|-------|--------------|------|
| version | String | URI参数 | API版本号,当前值为1 | 必须   |

#### ⊙ 日期与时间规范

日期与时间的表示有多种方式。为统一起见,除非是约定俗成或者有相应规范的,凡需要日期时间表示的地方一律采用UTC时间,遵循ISO 8601,并做以下约束:

- 表示日期一律采用YYYY-MM-DD方式,例如2014-06-01表示2014年6月1日。
- 表示时间一律采用hh:mm:ss方式,并在最后加一个大写字母Z表示UTC时间。例如23:00:10Z表示UTC时间23点0分10秒。
- 凡涉及日期和时间合并表示时,在两者中间加大写字母T,例如2014-06-01T23:00:10Z表示UTC时间2014年6月1日23点0分10秒。

```
₯规范化字符串
```

通常一个字符串中可以包含任何Unicode字符。在编程中这种灵活性会带来不少困扰。因此引入"规范字符串"的概念。一个规范字符串只包含百分 号编码字符以及URI(Uniform Resource Identifier)非保留字符(Unreserved Characters)。 RFC 3986规定URI非保留字符包括以下字符:字母 (A-Z, a-z)、数字(0-9)、连字号(-)、点号(.)、下划线(\_)、波浪线(~)。

将任意一个字符串转换为规范字符串的方式是:

- 将字符串转换成UTF-8编码的字节流。
- 保留所有URI非保留字符原样不变。
- 对其余字节做一次RFC 3986中规定的百分号编码(Percent-Encoding),即一个%后面跟着两个表示该字节值的十六进制字母。字母一律采用 大写形式。

示例: 原字符串: this is an example for 测试, 对应的规范字符串: this%20is%20an%20example%20for%20%E6%B5%8B%E8%AF%95。

```
心 服务域名
```

| 服务端点Endpoint                           | 协议   |
|----------------------------------------|------|
| {单节点版 IEC所在机器IP}:{单节点版 IEC监听端口,默认8702} | HTTP |

- の错误码
- 心 错误码格式

当用户访问API出现错误时,会返回给用户相应的错误码和错误信息,便于定位问题,并做出适当的处理。请求发生错误时通过Response Body返回详细错误信息,遵循如下格式:

| 参数名    | 类型     | 说明          |
|--------|--------|-------------|
| status | int    | 表示具体错误类型。   |
| msg    | String | 有关该错误的详细说明。 |

例如:

```
{
"status": 170000,
"message": "发生错误. Error"
}
```

### 心 公共请求参数

```
心 公共请求Query参数
```

当用户访问API时,需要通过query参数传入access\_token,如 http://127.0.0.1:8702/iec/iapi/v1/camera/new?access\_token={access\_token}

```
心获取生成Access Token用的AK、SK
```

| EasyEdge 智能边缘控制                        | 间台-单节点版       |                                                 |    |    | 操作指南 | a admin 🗸 |
|----------------------------------------|---------------|-------------------------------------------------|----|----|------|-----------|
| 总览                                     | 系统设置          |                                                 |    |    |      |           |
| ✿ 本地模型服务<br>添加模型服务<br>模型服务管理           |               | 30 🗘 天<br>/home/work/.baidu/easyedge/iec/events |    |    |      |           |
| ② 我的摄像头<br>添加摄像头                       |               | ●●●<br>+ 新増AK/SK ◎                              |    |    |      |           |
| <b>摄像头管理</b><br>巨 本地事件                 |               | AK                                              | SK | 操作 |      |           |
| 本地事件记录                                 | $\rightarrow$ | ae25d6c1-254f-4351-a1c1-87f61                   |    |    |      |           |
| <ul> <li>② 设置</li> <li>系统设置</li> </ul> |               |                                                 |    |    |      |           |
| 运行日志                                   |               |                                                 |    |    |      |           |
| 版本号: V2.2.2                            | 多节点模式 ③       |                                                 |    |    |      |           |

- ං 获取Access Token
- っ 基本信息

Path: /auth/v1/token

Method : GET

接口描述:

## 心 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

#### Body

| 名称           | 类型     | 是否必须 | 默认值 | 备注                     | 其他信息 |
|--------------|--------|------|-----|------------------------|------|
| grantType    | string | 必须   |     | 固定传入client_credentials |      |
| clientId     | string | 必须   |     | 设置-系统设置-AK             |      |
| clientSecret | string | 必须   |     | 设置-系统设置-SK             |      |

## の 返回数据

| 名称              | 类型     | 是否必须 | 默认值 | 备注       | 其他信息 |
|-----------------|--------|------|-----|----------|------|
| msg             | string | 非必须  |     |          |      |
| status          | number | 必须   |     |          |      |
| data            | object | 非必须  |     |          |      |
| ├── expiresIn   | number | 必须   |     | 过期时间戳,秒级 |      |
| ├── accessToken | string | 必须   |     |          |      |

## の AI服务相关接口

## の 本地服务列表

の 基本信息

Path: /iec/iapi/v1/aiservice/list

Method : GET

#### 接口描述:

心 请求参数

# 心 返回数据

| 名称                      | 类型           | 是否必须 | 默认值 | 备注                          | 其他信息            |
|-------------------------|--------------|------|-----|-----------------------------|-----------------|
| data                    | object<br>[] | 非必须  |     |                             | item 类型: object |
| ├── id                  | number       | 必须   |     |                             |                 |
| ├── uuid                | string       | 必须   |     |                             |                 |
| ├── name                | string       | 必须   |     |                             |                 |
| ├── source              | string       | 必须   |     |                             |                 |
| ├── activationType      | string       | 必须   |     | ONLINE 在线激活 / OFFLINE 纯离线激活 |                 |
| ├── activationSerialNum | string       | 必须   |     |                             |                 |
| ├── host                | string       | 必须   |     |                             |                 |
| ├── port                | number       | 必须   |     |                             |                 |
| ├── env                 | object       | 必须   |     |                             |                 |
| ├──<br>LD_LIBRARY_PATH  | string       | 非必须  |     |                             |                 |
| H AAA1                  | string       | 非必须  |     |                             |                 |
| ├── sssss               | string       | 非必须  |     |                             |                 |
| ├── isServiceOn         | boolean      | 必须   |     |                             |                 |
| ├── serviceType         | string       | 必须   |     | process 进程 / container 容器   |                 |
| ├── serviceStatus       | string       | 必须   |     |                             |                 |
| ├── serviceTip          | string       | 必须   |     |                             |                 |
| ├── createAt            | string       | 必须   |     |                             |                 |
| msg                     | string       | 非必须  |     |                             |                 |
| status                  | number       | 非必须  |     |                             |                 |

### ∞ 启动服务

## の 基本信息

Path: /iec/iapi/v1/aiservice/start

Method : POST

### 接口描述:

## 心 请求参数

#### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

### Body

| 名称        | 类型     | 是否必须 | 默认值 | 备注                | 其他信息 |
|-----------|--------|------|-----|-------------------|------|
| serviceld | number | 必须   |     | 服务id可以从"服务列表"接口获取 |      |

# 心 返回数据

| 名称     | 类型     | 是否必须 | 默认值 | 备注 | 其他信息 |
|--------|--------|------|-----|----|------|
| msg    | string | 非必须  |     |    |      |
| status | number | 非必须  |     |    |      |

心 基本信息

Path: /iec/iapi/v1/aiservice/stop

Method : POST

# 接口描述:

心 请求参数

## Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

## Body

| 名称        | 类型     | 是否必须 | 默认值 | 备注                | 其他信息 |
|-----------|--------|------|-----|-------------------|------|
| serviceld | number | 必须   |     | 服务id可以从"服务列表"接口获取 |      |

## 心 返回数据

| 名称     | 类型     | 是否必须 | 默认值 | 备注 | 其他信息 |
|--------|--------|------|-----|----|------|
| msg    | string | 非必须  |     |    |      |
| status | number | 非必须  |     |    |      |

## 心 更新服务

## 心 基本信息

Path : /iec/iapi/v1/aiservice/update

Method : POST

**接口描述**: 更新服务后服务会自动重启

## 心 请求参数

## Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

Body

| 名称   | 类型       | 是否必须 | 默认值 | 备注                            | 其他信息           |
|------|----------|------|-----|-------------------------------|----------------|
| id   | number   | 非必须  |     |                               |                |
| name | string   | 非必须  |     |                               |                |
| host | string   | 非必须  |     |                               |                |
| port | number   | 非必须  |     |                               |                |
| env  | array [] | 非必须  |     | 环境变量,格式为 [ ["key", "value"] ] | item 类型: array |

心 返回数据

| 名称     | 类型     | 是否必须 | 默认值 | 备注 | 其他信息 |
|--------|--------|------|-----|----|------|
| msg    | string | 非必须  |     |    |      |
| status | number | 非必须  |     |    |      |

∞ 摄像头设备相关接口

心 校验摄像头

Path: /iec/iapi/v1/camera/check-validity

## Method : POST

# 接口描述:

心 请求参数

## Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

#### Body

| 名称         | 类型     | 是否必须 | 默认值 | 备注                           | 其他信息            |
|------------|--------|------|-----|------------------------------|-----------------|
| cameraAddr | string | 必须   |     | 摄像头地址                        |                 |
| cameraType | string | 必须   |     | 摄像头类型,IP表示网络摄像头,LOCAL表示本地摄像头 | 枚举:<br>IP,LOCAL |

# 心 返回数据

| 名称     | 类型      | 是否必须 | 默认值 | 备注                 | 其他信息 |
|--------|---------|------|-----|--------------------|------|
| status | number  | 必须   |     | 状态码,0表示成功          |      |
| msg    | string  | 必须   |     | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须   |     | 摄像头是否合法            |      |

## の 新増摄像头

## 心 基本信息

Path: /iec/iapi/v1/camera/new

Method : POST

## 接口描述:

### の 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

Body

| 名称                | 类型       | 是否必<br>须 | 默认值 | 备注                                                              | 其他信息                              |
|-------------------|----------|----------|-----|-----------------------------------------------------------------|-----------------------------------|
| name              | string   | 必须       |     | 摄像头名称                                                           |                                   |
| camAddr           | string   | 非必须      |     | 摄像头地址,IP或LOCAL时必填                                               |                                   |
| camType           | string   | 必须       |     | 摄像头类型,IP表示网络摄像头,LOCAL表示本地摄像头,ONVIF表示ONVIF摄<br>像头,GB28181表示国标摄像头 | 枚举:<br>LOCAL,IP,ONVIF,GB28<br>181 |
| onvifConfig       | object   | 非必须      |     | ONVIF时必填                                                        |                                   |
| ├— ip             | string   | 必须       |     | IP                                                              |                                   |
| ├── port          | number   | 必须       |     | 端口                                                              |                                   |
| ├<br>username     | string   | 必须       |     | 用户名                                                             |                                   |
| ├<br>password     | string   | 必须       |     | 密码                                                              |                                   |
| gb28181Con<br>fig | object   | 非必须      |     | GB28181时必填                                                      |                                   |
| ├<br>sipName      | string   | 必须       |     | 信令服务器用户名                                                        |                                   |
| ├<br>sipDeviceId  | string   | 必须       |     | 信令服务器设备ID                                                       |                                   |
| ├<br>sipPassword  | string   | 必须       |     | 信令服务器密码                                                         |                                   |
| remark            | string   | 必须       |     | 摄像头备注                                                           |                                   |
| timeRange         | array [] | 必须       |     | 摄像头运行区间                                                         | item 类型: array                    |
| $\vdash$          |          | 非必须      |     |                                                                 |                                   |
| $\vdash$          |          | 非必须      |     | 秒                                                               |                                   |
| frameExtract      | object   | 必须       |     | 抽帧配置                                                            |                                   |
| ├<br>everySecond  | number   | 必须       |     | 每多少秒                                                            |                                   |
| ├── frames        | number   | 必须       |     | 抽多少帧                                                            |                                   |

## 心 返回数据

| 名称       | 类型     | 是否必须 | 默认值 | 备注                 | 其他信息 |
|----------|--------|------|-----|--------------------|------|
| status   | number | 必须   |     | 状态码,0表示成功          |      |
| msg      | string | 必须   |     | 错误信息,status不为0时有意义 |      |
| data     | object | 必须   |     |                    |      |
| ├ id     | number | 必须   |     | 当前节点唯一ID           |      |
| ├── uuid | string | 必须   |     | 全局唯一ID             |      |

## 心 摄像头列表

```
∞ 基本信息
```

Path: /iec/iapi/v1/camera/list

Method : GET

### 接口描述:

の 请求参数

| 参数名称     | 是否必须 | 示例 | 备注   |
|----------|------|----|------|
| pageNo   | 否    |    | 页数   |
| pageSize | 否    |    | 每页数量 |

### ∞ 返回数据

| 名称               | 类型           | 是否必须 | 默认值 | 备注                 | 其他信息                       |
|------------------|--------------|------|-----|--------------------|----------------------------|
| status           | number       | 必须   |     | 状态码,0表示成功          |                            |
| msg              | string       | 必须   |     | 错误信息,status不为0时有意义 |                            |
| data             | object<br>[] | 必须   |     |                    | item 类型: object            |
| ├── id           | number       | 必须   |     | 摄像头ID              |                            |
| ├── uuid         | string       | 必须   |     | 摄像头全局唯一ID          |                            |
| ├── name         | string       | 必须   |     | 摄像头名称              |                            |
| ├── remark       | string       | 必须   |     | 摄像头备注              |                            |
| ├── source       | string       | 必须   |     | 摄像头来源              | 枚举: SELF,IECC              |
| ├── cameraType   | string       | 必须   |     | 摄像头类型              | 枚举: LCOAL,IP,ONVIF,GB28181 |
| ├── cameraStatus | string       | 必须   |     | 摄像头状态              | 枚举: RUNNING,ERROR,STOPPED  |

### 心 摄像头详情

### 心 基本信息

Path: /iec/iapi/v1/camera/get

Method : GET

### 接口描述:

### 心 请求参数

Query

| 参数名称 | 是否必须 | 示例 | 备注    |
|------|------|----|-------|
| id   | 是    |    | 摄像头ID |

心 返回数据

| 名称                   | 类型       | 是否必须 | 默认值 | 备注                 | 其他信息                        |
|----------------------|----------|------|-----|--------------------|-----------------------------|
| status               | number   | 必须   |     | 状态码,0表示成功          |                             |
| msg                  | string   | 必须   |     | 错误信息,status不为0时有意义 |                             |
| data                 | object   | 必须   |     |                    |                             |
| ├── id               | number   | 必须   |     | 摄像头ID              |                             |
| ├── uuid             | string   | 必须   |     | 摄像头全局唯一ID          |                             |
| ├── name             | string   | 必须   |     | 摄像头名称              |                             |
| ├── remark           | string   | 必须   |     | 摄像头备注              |                             |
| ├── source           | string   | 必须   |     | 摄像头来源              | 枚举: SELF,IECC               |
| ├── cameraType       | string   | 必须   |     | 摄像头类型              | 枚举: LCOAL,IP,ONVIF,GB28181  |
| ├── cameraAddr       | string   | 必须   |     | 摄像头地址,IP或LOCAL时有意义 |                             |
|                      | object   | 非必须  |     |                    |                             |
| ├── ip               | string   | 必须   |     | IP                 |                             |
| ├── port             | number   | 必须   |     | 端口                 |                             |
| ├── username         | string   | 必须   |     | 用户名                |                             |
| ├── password         | string   | 必须   |     | 密码                 |                             |
| ├── gb28181Config    | object   | 非必须  |     |                    |                             |
| ├── sipName          | string   | 必须   |     | 信令服务器用户名           |                             |
| ├── sipDeviceId      | string   | 必须   |     | 信令服务器设备ID          |                             |
| ├── sipPassword      | string   | 必须   |     | 信令服务器密码            |                             |
| frameExtractInterval | number   | 必须   |     | 每多少秒               |                             |
| ├── frameExtractNum  | number   | 必须   |     | 抽多少帧               |                             |
| ├── onlineTime       | array [] | 必须   |     | 摄像头运行区间            | item 类型: array              |
| <u> </u>             |          | 非必须  |     |                    |                             |
| <u> </u>             |          | 非必须  |     | 秒                  |                             |
| ├── cameraStatus     | string   | 必须   |     | 摄像头状态              | 枚举: RUNNING, ERROR, STOPPED |

## ₯更新摄像头

```
の 基本信息
```

Path: /iec/iapi/v1/camera/update

Method : POST

#### 接口描述:

心 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

Body

| 名称                 | 类型       | 是否必<br>须 | 默认值 | 备注                                                              | 其他信息                              |
|--------------------|----------|----------|-----|-----------------------------------------------------------------|-----------------------------------|
| id                 | number   | 必须       |     | 摄像头ID                                                           |                                   |
| name               | string   | 必须       |     | 摄像头名称                                                           |                                   |
| camAddr            | string   | 非必须      |     | 摄像头地址,IP或LOCAL时必填                                               |                                   |
| camType            | string   | 必须       |     | 摄像头类型,IP表示网络摄像头,LOCAL表示本地摄像头,ONVIF表示ONVIF摄<br>像头,GB28181表示国标摄像头 | 枚举:<br>LOCAL,IP,ONVIF,GB28<br>181 |
| onvifConfig        | object   | 非必须      |     | ONVIF时必填                                                        |                                   |
| ├── ip             | string   | 必须       |     | IP                                                              |                                   |
| ├── port           | string   | 必须       |     | 端口                                                              |                                   |
| ⊢—<br>username     | string   | 必须       |     | 用户名                                                             |                                   |
| ├<br>password      | string   | 必须       |     | 密码                                                              |                                   |
| gb28181Con<br>fig  | object   | 非必须      |     | GB28181时必填                                                      |                                   |
| ├<br>sipName       | string   | 必须       |     | 信令服务器用户名                                                        |                                   |
| └──<br>sipDeviceId | string   | 必须       |     | 信令服务器设备ID                                                       |                                   |
| ├<br>sipPassword   | string   | 必须       |     | 信令服务器密码                                                         |                                   |
| remark             | string   | 必须       |     | 摄像头备注                                                           |                                   |
| timeRange          | array [] | 必须       |     | 摄像头运行区间                                                         | item 类型: array                    |
| $\vdash$           |          | 非必须      |     |                                                                 |                                   |
| $\vdash$           |          | 非必须      |     | 秒                                                               |                                   |
| frameExtract       | object   | 必须       |     | 抽帧配置                                                            |                                   |
| ├<br>everySecond   | number   | 必须       |     | 每多少秒                                                            |                                   |
| ├── frames         | number   | 必须       |     | 抽多少帧                                                            |                                   |

### 心 返回数据

| 名称     | 类型      | 是否必须 | 默认值 | 备注                 | 其他信息 |
|--------|---------|------|-----|--------------------|------|
| status | number  | 必须   |     | 状态码,0表示成功          |      |
| msg    | string  | 必须   |     | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须   |     | 修改是否成功             |      |

```
心 启动摄像头
```

```
の 基本信息
```

Path: /iec/iapi/v1/camera/start

Method : POST

接口描述:

心 请求参数

Headers

参数名称 参数值 是否必须 示例 Content-Type application/json 是

Body

| 名称 | 类型     | 是否必须 | 默认值 | 备注    | 其他信息 |
|----|--------|------|-----|-------|------|
| id | number | 必须   |     | 摄像头ID |      |

#### 心 返回数据

| 名称     | 类型      | 是否必须 | 默认值 | 备注                 | 其他信息 |
|--------|---------|------|-----|--------------------|------|
| status | number  | 必须   |     | 状态码,0表示成功          |      |
| msg    | string  | 必须   |     | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须   |     | 开启摄像头是否成功          |      |

备注

## ∞ 停止摄像头

```
の 基本信息
```

Path: /iec/iapi/v1/camera/stop

Method : POST

#### 接口描述:

## 心 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

Body

| 名称 | 类型     | 是否必须 | 默认值 | 备注    | 其他信息 |
|----|--------|------|-----|-------|------|
| id | number | 必须   |     | 摄像头ID |      |

## 心 返回数据

| 名称     | 类型      | 是否必须 | 默认值 | 备注                 | 其他信息 |
|--------|---------|------|-----|--------------------|------|
| status | number  | 必须   |     | 状态码,0表示成功          |      |
| msg    | string  | 必须   |     | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须   |     | 关闭摄像头是否成功          |      |

#### 心 删除摄像头

心 基本信息

Path: /iec/iapi/v1/camera/delete

Method : POST

## 接口描述:

心 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

 名称
 类型
 是否必须
 默认值
 备注
 其他信息

 id
 number
 必须
 摄像头ID

### 心 返回数据

| 名称     | 类型      | 是否必须 | 默认值 | 备注                 | 其他信息 |
|--------|---------|------|-----|--------------------|------|
| status | number  | 必须   |     | 状态码,0表示成功          |      |
| msg    | string  | 必须   |     | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须   |     | 删除摄像头是否成功          |      |

## 心物联网设备相关接口

### 心 新増设备

```
心 基本信息
```

Path: /iec/iapi/v1/plc/new

Method : POST

## 接口描述:

#### 心 请求参数

#### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

#### Body

| 名称              | 类型     | 是否必<br>须 | 默认值 | 备注        | 其他信息             |
|-----------------|--------|----------|-----|-----------|------------------|
| name            | string | 必须       |     | 设备名称      |                  |
| deviceType      | string | 必须       |     | 设备类型      | 枚举: modbus,opcua |
| modbus          | object | 非必须      |     | modbus时必填 |                  |
| ├<br>protocol   | string | 必须       |     | 协议类型      | 枚举: tcp,rtu      |
| ├── rtu         | object | 非必须      |     | rtu时必填    |                  |
| ├── port        | string | 必须       |     | 端口        |                  |
| ⊢–-<br>baudrate | number | 必须       |     | 波特率       |                  |
| ├<br>databit    | number | 必须       |     | 数据位       | 枚举: 5,6,7,8      |
| ├<br>stopbit    | number | 必须       |     | 停止位       | 枚举: 1,2          |
| ├<br>parity     | string | 必须       |     | 校验位       | 枚举: N,E,O        |
| ├── tcp         | object | 非必须      |     | tcp时必填    |                  |
| ⊢—<br>address   | string | 必须       |     | 连接地址      |                  |
| ├── port        | number | 必须       |     | 端口        |                  |
| ⊢––<br>slaveld  | number | 必须       |     | 从站号       |                  |
| ├<br>interval   | number | 必须       |     | 采样间隔      |                  |

| opcua                   | object       | 非必须 | opcua时必填                                    |                                                                       |
|-------------------------|--------------|-----|---------------------------------------------|-----------------------------------------------------------------------|
| ├<br>endpoint           | string       | 必须  | 通道地址                                        |                                                                       |
| ├<br>securityPolic<br>y | string       | 必须  | 安全策略                                        | 枚举:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss |
| ├──<br>securityMod<br>e | string       | 必须  | 安全模式                                        | 枚举: None,Sign,SignAndEncrypt                                          |
| ├<br>certificate        | string       | 非必须 | 数字证书                                        |                                                                       |
| ├<br>privateKey         | string       | 非必须 | 密钥证书                                        |                                                                       |
| ├<br>username           | string       | 非必须 | 用户名                                         |                                                                       |
| ├<br>password           | string       | 非必须 | 密码                                          |                                                                       |
| ├<br>timeout            | number       | 必须  | 连接超时时间                                      |                                                                       |
| ├── interval            | number       | 必须  | 采样间隔                                        |                                                                       |
| remark                  | string       | 必须  | 设备备注                                        |                                                                       |
| attributes              | object<br>[] | 必须  | 属性                                          | item 类型: object                                                       |
| ├── id                  | string       | 必须  | 标识符                                         |                                                                       |
| ├── name                | string       | 必须  | 属性名                                         |                                                                       |
| ├── type                | string       | 必须  | 类型                                          | 枚举: bool,int16,int32,int64,float32,float64,string                     |
| ├<br>defaultValue       | string       | 非必须 | 默认值                                         |                                                                       |
| ├── unit                | string       | 必须  | 单位                                          |                                                                       |
| ├<br>required           | boolea<br>n  | 必须  | 是否必填                                        |                                                                       |
| properties              | object<br>[] | 必须  | 测点                                          | item 类型: object                                                       |
| ├── id                  | string       | 必须  | 标识符                                         |                                                                       |
| ├── name                | string       | 必须  | 测点名                                         |                                                                       |
| ├── type                | string       | 必须  | 类型                                          | 枚举: bool,int16,int32,int64,float32,float64,string                     |
| ├── mode                | string       | 必须  | 读写类型                                        | 枚举: ro,rw                                                             |
| ├── unit                | string       | 必须  | 单位                                          |                                                                       |
| ├<br>modbus             | object       | 非必须 | modbus时必填                                   |                                                                       |
| ├── func                | number       | 必须  | 寄存器类型:1.线圈寄存器 2.离散输入寄存器 3.<br>保持寄存器 4.输入寄存器 | 枚举: 1,2,3,4                                                           |
| ⊢–<br>address           | string       | 必须  | 寄存器地址                                       |                                                                       |
| ⊢—<br>quantity          | number       | 必须  | 寄存器数量                                       |                                                                       |
| ├── opcua               | object       | 非必须 | opcua时必填                                    |                                                                       |
|                         | string       | 必须  | 节点ID                                        |                                                                       |

## 心 返回数据

| 名称       | 类型     | 是否必须 | 默认值 | 备注                 | 其他信息 |
|----------|--------|------|-----|--------------------|------|
| status   | number | 必须   |     | 状态码,0表示成功          |      |
| msg      | string | 必须   |     | 错误信息,status不为0时有意义 |      |
| data     | object | 必须   |     |                    |      |
| ├── id   | number | 必须   |     | 当前节点唯一ID           |      |
| ├── uuid | string | 必须   |     | 全局唯一ID             |      |

### 心 设备列表

#### の 基本信息

Path: /iec/iapi/v1/plc/list

Method : GET

## 接口描述:

## 心 请求参数

#### Query

| 参数名称     | 是否必须 | 示例 | 备注   |
|----------|------|----|------|
| pageNo   | 否    |    | 页数   |
| pageSize | 否    |    | 每页数量 |

## 心 返回数据

| 名称             | 类型           | 是否必须 | 默认值 | 备注                 | 其他信息                      |
|----------------|--------------|------|-----|--------------------|---------------------------|
| status         | number       | 必须   |     | 状态码,0表示成功          |                           |
| msg            | string       | 必须   |     | 错误信息,status不为0时有意义 |                           |
| data           | object<br>[] | 必须   |     |                    | item 类型: object           |
| ├── id         | number       | 必须   |     | 设备ID               |                           |
| ├── uuid       | string       | 必须   |     | 设备全局唯一ID           |                           |
| ├── name       | string       | 必须   |     | 设备名称               |                           |
| ├── remark     | string       | 必须   |     | 设备备注               |                           |
| ├── deviceType | string       | 必须   |     | 设备类型               |                           |
| ├── source     | string       | 必须   |     | 设备来源               | 枚举: SELF,IECC             |
| ├── status     | string       | 必须   |     | 设备状态               | 枚举: RUNNING,ERROR,STOPPED |

### 心设备详情

```
心 基本信息
```

Path: /iec/iapi/v1/plc/get

Method : GET

## 接口描述:

心 请求参数

Query

| 参数名称 | 是否必须 | 示例 | 备注   |
|------|------|----|------|
| id   | 是    |    | 设备ID |

## 心 返回数据

| 名称                   | 类型           | 是否必<br>须 | 默认值 | 备注                 | 其他信息             |
|----------------------|--------------|----------|-----|--------------------|------------------|
| status               | number       | 必须       |     | 状态码,0表示成功          |                  |
| msg                  | string       | 必须       |     | 错误信息,status不为0时有意义 |                  |
| data                 | object       | 必须       |     |                    |                  |
| ├── id               | string       | 必须       |     | 设备ID               |                  |
| ├── uuid             | string       | 必须       |     | 设备全局唯一ID           |                  |
| ├── name             | string       | 必须       |     | 设备名称               |                  |
| ├── deviceType       | string       | 必须       |     | 设备类型               | 枚举: modbus,opcua |
| ├── modbus           | object       | 非必须      |     | modbus时必填          |                  |
| ├── protocol         | string       | 必须       |     | 协议类型               | 枚举: tcp,rtu      |
| ├── rtu              | object       | 非必须      |     | rtu时必填             |                  |
| ├── port             | string       | 必须       |     | 端口                 |                  |
| ⊢–-<br>baudrate      | number       | 必须       |     | 波特率                |                  |
| ├── databit          | number       | 必须       |     | 数据位                | 枚举: 5,6,7,8      |
| ├── stopbit          | number       | 必须       |     | 停止位                | 枚举: 1,2          |
| ├── parity           | string       | 必须       |     | 校验位                | 枚举: N,E,O        |
| ├── tcp              | object       | 非必须      |     | tcp时必填             |                  |
| <br>address          | string       | 必须       |     | 连接地址               |                  |
| ├── port             | number       | 必须       |     | 端口                 |                  |
|                      | number       | 必须       |     | 从站号                |                  |
| ⊢—<br>interval       | number       | 必须       |     | 采样间隔               |                  |
| ├── opcua            | object       | 非必须      |     | opcua时必填           |                  |
| - endpoint           | string       | 必须       |     | 通道地址               |                  |
| ⊢–<br>securityPolicy | string       | 必须       |     | 安全策略               |                  |
| ⊢–<br>securityMode   | string       | 必须       |     | 安全模式               |                  |
| ⊢—<br>certificate    | string       | 非必须      |     | 数字证书               |                  |
| ├<br>privateKey      | string       | 非必须      |     | 密钥证书               |                  |
| ⊢––<br>username      | string       | 非必须      |     | 用户名                |                  |
| ⊢—<br>password       | string       | 非必须      |     | 密码                 |                  |
| timeout              | number       | 必须       |     | 连接超时时间             |                  |
| interval             | number       | 必须       |     | 采样间隔               |                  |
| - remark             | string       | 必须       |     | 设备备注               |                  |
| ├── source           | string       | 必须       |     | 设备来源               | 枚举: SELF,IECC    |
| ├── attributes       | object<br>[] | 必须       |     | 属性                 | item 类型: object  |
| ├── id               | string       | 必须       |     | 标识符                |                  |

|    |                 | -            | · - · · · |                                             |                                                          |
|----|-----------------|--------------|-----------|---------------------------------------------|----------------------------------------------------------|
|    | ├── name        | string       | 必须        | 属性名                                         |                                                          |
|    | ├── type        | string       | 必须        | 类型                                          | 枚举:<br>bool,int16,int32,int64,float32,float64,s<br>tring |
| de | ├<br>faultValue | string       | 非必须       | 默认值                                         |                                                          |
|    | ├── unit        | string       | 必须        | 单位                                          |                                                          |
|    | ├── required    | boolea<br>n  | 必须        | 是否必填                                        |                                                          |
| ŀ  | — properties    | object<br>[] | 必须        | 测点                                          | item 类型: object                                          |
|    | ├── id          | string       | 必须        | 标识符                                         |                                                          |
|    | ├── name        | string       | 必须        | 测点名                                         |                                                          |
|    | ├── type        | string       | 必须        | 类型                                          | 枚举:<br>bool,int16,int32,int64,float32,float64,s<br>tring |
|    | ├── mode        | string       | 必须        | 读写类型                                        | 枚举: ro,rw                                                |
|    | ├── unit        | string       | 必须        | 单位                                          |                                                          |
|    |                 | object       | 非必须       | modbus时必填                                   |                                                          |
|    | ├── func        | number       | 必须        | 寄存器类型:1.线圈寄存器 2.离散输入寄存器 3.保持寄存<br>器 4.输入寄存器 | 枚举: 1,2,3,4                                              |
| ad | ├<br>dress      | string       | 必须        | 寄存器地址                                       |                                                          |
| qu | ├<br>antity     | number       | 必须        | 寄存器数量                                       |                                                          |
|    | ├── opcua       | object       | 非必须       | opcua时必填                                    |                                                          |
|    | ├── nodeid      | string       | 必须        | 节点ID                                        |                                                          |
| F  | — status        | string       | 必须        | 设备状态                                        | 枚举: RUNNING,ERROR,STOPPED                                |
| F  | — url           | string       | 非必须       | 测点数据获取websocket连接地址                         |                                                          |
| ⊦  | — realData      | object       | 非必须       | 当前测点数据                                      |                                                          |

## 心 更新设备

# ∞ 基本信息

Path: /iec/iapi/v1/plc/update

Method : POST

### 接口描述:

## 心 请求参数

#### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

## Body

| 名称   | 类型     | 是否必<br>须 | 默认值 | 备注   | 其他信息 |
|------|--------|----------|-----|------|------|
| id   | number | 必须       |     | 设备ID |      |
| name | string | 必须       |     | 设备名称 |      |

| deviceType     | string                                                                                                                               | 必须                                                                                                          | 设备类型                                                                                                                                                                | 枚举: modbus,opcua                                                                                      |
|----------------|--------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| modbus         | object                                                                                                                               | 非必须                                                                                                         | modbus时必填                                                                                                                                                           |                                                                                                       |
| ├<br>protocol  | string                                                                                                                               | 必须                                                                                                          | 协议类型                                                                                                                                                                | 枚举: tcp,rtu                                                                                           |
| ├── rtu        | object                                                                                                                               | 非必须                                                                                                         | rtu时必填                                                                                                                                                              |                                                                                                       |
| ├── port       | string                                                                                                                               | 必须                                                                                                          | 端口                                                                                                                                                                  |                                                                                                       |
| ⊢–<br>baudrate | number                                                                                                                               | 必须                                                                                                          | 波特率                                                                                                                                                                 |                                                                                                       |
| ├<br>databit   | number                                                                                                                               | 必须                                                                                                          | 数据位                                                                                                                                                                 | 枚举: 5,6,7,8                                                                                           |
| ⊢—<br>stopbit  | number                                                                                                                               | 必须                                                                                                          | 停止位                                                                                                                                                                 | 枚举: 1,2                                                                                               |
| ├<br>parity    | string                                                                                                                               | 必须                                                                                                          | 校验位                                                                                                                                                                 | 枚举: N,E,O                                                                                             |
| ├── tcp        | object                                                                                                                               | 非必须                                                                                                         | tcp时必填                                                                                                                                                              |                                                                                                       |
| ⊢––<br>address | string                                                                                                                               | 必须                                                                                                          | 连接地址                                                                                                                                                                |                                                                                                       |
| ├── port       | number                                                                                                                               | 必须                                                                                                          | 端口                                                                                                                                                                  |                                                                                                       |
| ⊢—<br>slaveld  | number                                                                                                                               | 必须                                                                                                          | 从站号                                                                                                                                                                 |                                                                                                       |
| interval       | number                                                                                                                               | 必须                                                                                                          | 采样间隔                                                                                                                                                                |                                                                                                       |
| opcua          | object                                                                                                                               | 非必须                                                                                                         | opcua时必填                                                                                                                                                            |                                                                                                       |
| ⊢              | string                                                                                                                               | 必须                                                                                                          | 通道地址                                                                                                                                                                |                                                                                                       |
| endpoint       |                                                                                                                                      |                                                                                                             |                                                                                                                                                                     |                                                                                                       |
| endpoint<br>   | string                                                                                                                               | 必须                                                                                                          | 安全策略                                                                                                                                                                | 枚举:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss                                 |
| endpoint<br>   | string                                                                                                                               | 必须                                                                                                          | 安全策略<br>安全模式                                                                                                                                                        | 枚举:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss<br>枚举: None,Sign,SignAndEncrypt |
| endpoint<br>   | string<br>string<br>string                                                                                                           | 必须<br>必须<br>非必须                                                                                             | 安全策略<br>安全模式<br>数字证书                                                                                                                                                | 枚举:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss<br>枚举: None,Sign,SignAndEncrypt |
| endpoint<br>   | string<br>string<br>string<br>string                                                                                                 | 必须<br>必须<br>非必须<br>非必须                                                                                      | 安全策略<br>安全模式<br>数字证书<br>密钥证书                                                                                                                                        | 枚举:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss<br>枚举: None,Sign,SignAndEncrypt |
| endpoint<br>   | string<br>string<br>string<br>string<br>string                                                                                       | 必须<br>必须<br>非必须<br>非必须<br>非必须                                                                               | 安全策略<br>安全模式<br>数字证书<br>密钥证书<br>用户名                                                                                                                                 | 枚举:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss<br>枚举: None,Sign,SignAndEncrypt |
| endpoint<br>   | string<br>string<br>string<br>string<br>string<br>string                                                                             | 必须<br>必须<br>非必须<br>非必须<br>非必须<br>非必须                                                                        | 安全策略         安全模式         数字证书         密钥证书         用户名         密码                                                                                                  | 枚挙:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss<br>枚挙: None,Sign,SignAndEncrypt |
| endpoint<br>   | string<br>string<br>string<br>string<br>string<br>string<br>number                                                                   | 必须<br>必须<br>非必须<br>非必须<br>非必须<br>非必须<br>非必须                                                                 | 安全策略         安全模式         数字证书         密钥证书         用户名         密码         连接超时时间                                                                                   | 枚挙:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss<br>枚举: None,Sign,SignAndEncrypt |
| endpoint<br>   | string<br>string<br>string<br>string<br>string<br>string<br>number<br>number                                                         | 必须<br>必须<br>非必须<br>非必须<br>非必须<br>非必须<br>小<br>が须                                                             | 安全策略         安全模式         数字证书         密钥证书         用户名         密码         连接超时时间         采样间隔                                                                      | 枚挙:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss<br>枚挙: None,Sign,SignAndEncrypt |
| endpoint<br>   | string<br>string<br>string<br>string<br>string<br>string<br>number<br>string                                                         | 必须<br>必须<br>非必须<br>非必须<br>非必须<br>非必须<br>な须<br>必须<br>必须                                                      | 安全策略         安全模式         数字证书         密钥证书         席日户名         梁码         连接超时时间         采样间隔         设备备注                                                        | 枚挙:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss<br>枚挙: None,Sign,SignAndEncrypt |
| endpoint<br>   | string<br>string<br>string<br>string<br>string<br>string<br>string<br>number<br>number<br>string<br>object                           | 必须<br>必须<br>非必须<br>非必须<br>非必须<br>非必须<br>小须<br>必须<br>必须                                                      | 安全策略         安全模式         数字证书         密钥证书         席日户名         梁          第         建接超时时间         采样间隔         设备备注         属性                                   | 枚挙:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss<br>枚挙: None,Sign,SignAndEncrypt |
| endpoint<br>   | string<br>string<br>string<br>string<br>string<br>string<br>string<br>number<br>number<br>string<br>object<br>[]                     | 必须<br>必须<br>非必须<br>非必须<br>非必须<br>非必须<br>小须<br>必须<br>必须<br>必须                                                | 安全策略         安全模式         数字证书         密钥证书         用户名         密码         连接超时时间         采样间隔         设备备注         属性         标识符                                  | 枚举:<br>None,Basic256Sha256,Aes128Sha256Rsa0aep,Ae<br>s256Sha256RsaPss                                 |
| endpoint<br>   | string<br>string<br>string<br>string<br>string<br>string<br>string<br>number<br>string<br>object<br>[]<br>string<br>string           | 必须<br>必须<br>非必须<br>非必须<br>非必须<br>非必须<br>非必须<br>必须<br>必须<br>必须<br>必须<br>必须                                   | 安全策略         安全模式         数字证书         数字证书         密钥证书         用户名         家码         建提超时时间         采样间隔         设备备注         属性         标识符         属性名         | 枚挙:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss                                 |
| endpoint<br>   | string<br>string<br>string<br>string<br>string<br>string<br>string<br>number<br>string<br>object<br>[]<br>string<br>string<br>string | 必须<br>必须<br>非必须<br>非必须<br>非必须<br>非必须<br>な须<br>な须<br>必须<br>必须<br>必须<br>必须<br>必须<br>必须<br>必须<br>必须<br>必须<br>必 | 安全策略         安全模式         数字证书         数字证书         密钥证书         席         교         建接超时时间         采样间隔         设备备注         属性         标识符         属性名         类型 | 枚挙:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss                                 |

| defaultValue  | string       | 非必须 | 默认值                                      |                                                   |
|---------------|--------------|-----|------------------------------------------|---------------------------------------------------|
| ├── unit      | string       | 必须  | 单位                                       |                                                   |
| ├<br>required | boolea<br>n  | 必须  | 是否必填                                     |                                                   |
| properties    | object<br>[] | 必须  | 测点                                       | item 类型: object                                   |
| ├── id        | string       | 必须  | 标识符                                      |                                                   |
| ├── name      | string       | 必须  | 测点名                                      |                                                   |
| ├── type      | string       | 必须  | 类型                                       | 枚举: bool,int16,int32,int64,float32,float64,string |
| ├── mode      | string       | 必须  | 读写类型                                     | 枚举: ro,rw                                         |
| ├── unit      | string       | 必须  | 单位                                       |                                                   |
| ├──<br>modbus | object       | 非必须 | modbus时必填                                |                                                   |
| ├── func      | number       | 必须  | 寄存器类型:1.线圈寄存器2.离散输入寄存器3.<br>保持寄存器4.输入寄存器 | 枚举: 1,2,3,4                                       |
| ⊢—<br>address | string       | 必须  | 寄存器地址                                    |                                                   |
| ├<br>quantity | number       | 必须  | 寄存器数量                                    |                                                   |
| ├── opcua     | object       | 非必须 | opcua时必填                                 |                                                   |
| ⊢—<br>nodeid  | string       | 必须  | 节点ID                                     |                                                   |

### 心 返回数据

| 名称     | 类型      | 是否必须 | 默认值 | 备注                 | 其他信息 |
|--------|---------|------|-----|--------------------|------|
| status | number  | 必须   |     | 状态码,0表示成功          |      |
| msg    | string  | 必须   |     | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须   |     | 修改是否成功             |      |

## の 删除设备

```
∞ 基本信息
```

Path: /iec/iapi/v1/plc/delete

Method : POST

接口描述:

心 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

Body

| 名称 | 类型     | 是否必须 | 默认值 | 备注   | 其他信息 |
|----|--------|------|-----|------|------|
| id | number | 必须   |     | 设备ID |      |

心 返回数据
| 名称     | 类型      | 是否必须 默认值 | 备注                 | 其他信息 |
|--------|---------|----------|--------------------|------|
| status | number  | 必须       | 状态码,0表示成功          |      |
| msg    | string  | 必须       | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须       | 删除设备是否成功           |      |

## ∞ 写测点

```
の 基本信息
```

Path: /iec/iapi/v1/plc/property/write

Method : POST

#### 接口描述:

## 心 请求参数

#### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

#### Body

| 名称            | 类型     | 是否必须 默认值 | ī 备注  | 其他信息 |
|---------------|--------|----------|-------|------|
| id            | number | 必须       | 摄像头ID |      |
| propertyld    | string | 必须       | 测点标识符 |      |
| propertyValue | any    | 必须       | 测点值   |      |

#### の 返回数据

| 名称     | 类型      | 是否必须 | 默认值 | 备注                 | 其他信息 |
|--------|---------|------|-----|--------------------|------|
| status | number  | 必须   |     | 状态码,0表示成功          |      |
| msg    | string  | 必须   |     | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须   |     | 修改测点值是否成功          |      |

#### 心功能更新记录

| 时间         | 版本    | 说明  |
|------------|-------|-----|
| 2023-05-25 | 1.0.0 | 第一版 |

EasyEdge 智能边缘控制台-多节点版 IECC API

## の 概述

欢迎使用EasyEdge 智能边缘控制台-多节点版 IEC。 您可以使用本文档介绍的API对多节点版 IEC服务进行操作。

### ∞ 接口概览

多节点版 IEC API 提供下列接口类型:

| 接口类型    | 描述                     |
|---------|------------------------|
| AI服务接口  | 接口包括查询、启动、停止等          |
| 摄像头设备接口 | 接口包括创建、查询、更新、删除、启动、停止等 |
| 物联网设备接口 | 接口包括创建、查询、更新、删除等       |
| 其他接口    | 包括生成鉴权用的Access Token等  |

#### の 通用说明

API调用遵循HTTP协议。数据交换格式为JSON,所有request/response body内容均采用UTF-8编码。

の API认证机制

所有API的安全认证一律采用Access Key与请求签名机制。 Access Key由Access Key ID和Secret Access Key组成,均为字符串。 对于每个HTTP请 求,需先调用生成Access Token的接口生成认证字符串。提交认证字符串放在query里。服务端根据生成算法验证认证字符串的正确性。 当服务 端接收到用户的请求后,系统将使用相同的SK和同样的认证机制生成认证字符串,并与用户请求中包含的认证字符串进行比对。如果认证字符串 相同,系统认为用户拥有指定的操作权限,并执行相关操作;如果认证字符串不同,系统将忽略该操作并返回错误码。

心 通信协议

支持HTTP调用方式。

の请求结构说明

数据交换格式为JSON,所有request/response body内容均采用UTF-8编码。 请求参数包括如下4种:

| 参数类型        | 说明                                                    |
|-------------|-------------------------------------------------------|
| URI         | 通常用于指明操作类型,如:POST<br>/iec/iapi/v{version}/{type}/{op} |
| Query参数     | URL中携带的请求参数,通常用来传入认证字符串                               |
| HEADER      | 通过HTTP头域传入                                            |
| RequestBody | 通过JSON格式组织的请求数据体                                      |

#### 心 响应结构说明

响应值分为两种形式: | 响应内容 | 说明 | |------ | HTTP STATUS CODE | 如200,400,403,404等 | | ResponseBody | JSON格式组织的响应数据体 |

の API版本号

| 参数      | 类型     | 参数位置  | 描述           | 是否必须 |
|---------|--------|-------|--------------|------|
| version | String | URI参数 | API版本号,当前值为1 | 必须   |

⊙ 日期与时间规范

日期与时间的表示有多种方式。为统一起见,除非是约定俗成或者有相应规范的,凡需要日期时间表示的地方一律采用UTC时间,遵循ISO 8601,并做以下约束:

- 表示日期一律采用YYYY-MM-DD方式,例如2014-06-01表示2014年6月1日。
- 表示时间一律采用hh:mm:ss方式,并在最后加一个大写字母Z表示UTC时间。例如23:00:10Z表示UTC时间23点0分10秒。
- 凡涉及日期和时间合并表示时,在两者中间加大写字母T,例如2014-06-01T23:00:10Z表示UTC时间2014年6月1日23点0分10秒。

•

#### <sup>の</sup> 规范化字符串

通常一个字符串中可以包含任何Unicode字符。在编程中这种灵活性会带来不少困扰。因此引入"规范字符串"的概念。一个规范字符串只包含百分 号编码字符以及URI (Uniform Resource Identifier) 非保留字符 (Unreserved Characters)。 RFC 3986规定URI非保留字符包括以下字符:字母 (A-Z, a-z)、数字(0-9)、连字号(-)、点号(.)、下划线(\_)、波浪线(~)。将任意一个字符串转换为规范字符串的方式是:

- 将字符串转换成UTF-8编码的字节流。
- 保留所有URI非保留字符原样不变。
- 对其余字节做一次RFC 3986中规定的百分号编码(Percent-Encoding),即一个%后面跟着两个表示该字节值的十六进制字母。字母一律采用 大写形式。示例:原字符串:this is an example for 测试,对应的规范字符
  - $\texttt{\texttt{\texttt{+}:this\%20is\%20an\%20example\%20for\%20\%E6\%B5\%8B\%E8\%AF\%95}_{\circ}}$

#### 心 服务域名

| 服务端点Endpoint                           | 协议   |
|----------------------------------------|------|
| {多节点版 IEC所在机器IP}:{多节点版 IEC监听端口,默认8702} | HTTP |

心 错误码格式

当用户访问API出现错误时,会返回给用户相应的错误码和错误信息,便于定位问题,并做出适当的处理。请求发生错误时通过Response Body返回详细错误信息,遵循如下格式:

| 参数名    | 类型     | 说明          |
|--------|--------|-------------|
| status | int    | 表示具体错误类型。   |
| msg    | String | 有关该错误的详细说明。 |
|        |        |             |

例如:

| { |                          |
|---|--------------------------|
|   | "status": 170000,        |
|   | "message": "发生错误. Error" |
| } |                          |

### 心 公共请求参数

#### <sup>の</sup>公共请求Query参数

当用户访问API时,需要通过query参数传入access\_token,如 http://127.0.0.1:8702/iec/iapi/v1/camera/new?access\_token={access\_token}

```
の 获取生成Access Token用的AK、SK
```

| EasyEdge 智能边缘控制              | 則台−多节点版  |                                      |              |             | 操作指南 | a admin 🗸 |
|------------------------------|----------|--------------------------------------|--------------|-------------|------|-----------|
| 总览                           | 系统设置     |                                      |              |             |      |           |
| ⑦ 模型服务中心<br>模型服务管理<br>下发模型服务 |          | 通送計 ✓ □<br>30 ☆ 天                    |              |             |      |           |
| □ 边缘资源中心<br>边缘节点<br>摄像头设备    |          | monerwink.baruureasyeugenet.crevenis |              |             |      |           |
| 苗 事件中心<br>事件记录<br>◎ 设置       |          | 检查并更新<br>+ 新墙AK/SK <sup>③</sup>      |              |             |      |           |
| 系统设置 🔶                       | <b>→</b> | AK<br>e26a85f8-bd03-4bfb-9be9-d2aa   | SK<br>****** | 操作<br>复制 删除 |      |           |
| 版本号:V2.2.2                   |          |                                      |              |             |      |           |

- の 获取Access Token
- の 基本信息

Path: /auth/v1/token

Method : GET

接口描述:

### の 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

| 名称           | 类型     | 是否必须 | 默认值 | 备注                     | 其他信息 |
|--------------|--------|------|-----|------------------------|------|
| grantType    | string | 必须   |     | 固定传入client_credentials |      |
| clientId     | string | 必须   |     | 设置-系统设置-AK             |      |
| clientSecret | string | 必须   |     | 设置-系统设置-SK             |      |

| 名称            | 类型     | 是否必须 | 默认值 | 备注       | 其他信息 |
|---------------|--------|------|-----|----------|------|
| msg           | string | 非必须  |     |          |      |
| status        | number | 必须   |     |          |      |
| data          | object | 非必须  |     |          |      |
| ├── expiresIn | number | 必须   |     | 过期时间戳,秒级 |      |
| - accessToken | string | 必须   |     |          |      |

### ⁰ 节点相关接口

```
⊙ 节点列表
```

```
の 基本信息
```

Path: /iec/iapi/v1/node/list

Method : POST

## 接口描述:

节点的新增涉及到激活流程,需要在IEC控制中心页面操作。

### 心 请求参数

### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

### Body

| 名称       | 类型     | 是否必须 | 默认值 | 备注        | 其他信息 |
|----------|--------|------|-----|-----------|------|
| filters  | object | 非必须  |     |           |      |
| ├── name | string | 非必须  |     | 搜索包含子串的节点 |      |

の 返回数据

| 名称             | 类型           | 是否必须 | 默认值 | 备注            | 其他信息            |
|----------------|--------------|------|-----|---------------|-----------------|
| data           | object<br>[] | 非必须  |     |               | item 类型: object |
| ├── id         | number       | 必须   |     |               |                 |
| ├── uuid       | string       | 必须   |     |               |                 |
| ├── name       | string       | 非必须  |     |               |                 |
| ├── tag        | object       | 非必须  |     |               |                 |
| ├── remark     | string       | 非必须  |     |               |                 |
| ├ hostname     | string       | 非必须  |     |               |                 |
| ├ platform     | string       | 非必须  |     | 操作系统          |                 |
| ├── MACAddr    | string       | 非必须  |     | 节点的mac地址,逗号分割 |                 |
| IPAddr         | string       | 非必须  |     | 节点的ip地址,逗号分割  |                 |
| - CPUArch      | string       | 非必须  |     | CPU架构         |                 |
| ├── CPUModel   | string       | 非必须  |     | CPU型号         |                 |
| Here ASICModel | string       | 非必须  |     | Al芯片的型号,逗号分割  |                 |
| isActivated    | boolean      | 非必须  |     | 是否已经激活        |                 |
| ├── createAt   | string       | 非必须  |     |               |                 |
| msg            | string       | 非必须  |     |               |                 |
| status         | number       | 非必须  |     |               |                 |

## 心已添加的模型列表

```
の 基本信息
```

Path: /iec/iapi/v1/aimodel/list

Method : GET

# 接口描述:

心 请求参数

#### っ 返回数据

| 名称                    | 类型           | 是否必须 | 默认值 | 备注                           | 其他信息            |
|-----------------------|--------------|------|-----|------------------------------|-----------------|
| data                  | object<br>[] | 非必须  |     |                              | item 类型: object |
| ├── id                | number       | 必须   |     |                              |                 |
| ├── name              | string       | 必须   |     |                              |                 |
| ├ modelProduct        | string       | 必须   |     |                              |                 |
| ├── modelName         | string       | 必须   |     |                              |                 |
| ├── modelType         | number       | 必须   |     | 1-分类, 2-检测 , 14-语义分割, 6-实例分割 |                 |
| ├── modelSoc          | string       | 必须   |     |                              |                 |
| ├── modelThresholdRec | number       | 必须   |     |                              |                 |
| ├── platform          | string       | 必须   |     |                              |                 |
| ├── form              | string       | 必须   |     |                              |                 |
| ├── isEdgeKit         | boolean      | 必须   |     |                              |                 |
| ├── createAt          | string       | 必须   |     |                              |                 |
| msg                   | string       | 非必须  |     |                              |                 |
| status                | number       | 非必须  |     |                              |                 |

# ⁰下发模型为服务

### の 基本信息

Path: /iec/iapi/v1/aimodel/deploy

## Method : POST

## 接口描述:

# 心 请求参数

## Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

## Body

| 名称                   | 类型           | 是否必须 | 默认值 | 备注                                   | 其他信息            |
|----------------------|--------------|------|-----|--------------------------------------|-----------------|
| id                   | number       | 必须   |     | 模型id,可以从aimode/list接口获得              |                 |
| targetNodes          | object<br>[] | 必须   |     |                                      | item 类型: object |
| ├── id               | number       | 必须   |     | 节点id,可以从 <mark>node/list</mark> 接口获得 |                 |
| ├── activationType   | string       | 必须   |     | ONLINE / OFFLINE                     |                 |
| ├ serialNum          | string       | 非必须  |     | ONLINE 激活时填入序列号                      |                 |
| ├<br>licenseFileCont | string       | 非必须  |     | OFFLINE 激活时填入离线激活的license内容          |                 |
| ├── host             | string       | 非必须  |     | 默认为0.0.0.0                           |                 |
| ├── port             | number       | 非必须  |     | 默认为0,也即系统自动选择                        |                 |
| ├── env              | array []     | 非必须  |     |                                      | item 类型: array  |

### 心 返回数据

| 名称       | 类型           | 是否必须 | 默认值 | 备注 | 其他信息            |
|----------|--------------|------|-----|----|-----------------|
| data     | object<br>[] | 非必须  |     |    | item 类型: object |
| ├── id   | number       | 非必须  |     |    |                 |
| ├── uuid | string       | 非必须  |     |    |                 |
| msg      | string       | 非必须  |     |    |                 |
| status   | number       | 非必须  |     |    |                 |

## 心已下发的服务详情

## 心 基本信息

Path: /iec/iapi/v1/aiservice/detail

Method : POST

## 接口描述:

# 心 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

| 名称 | 类型     | 是否必须 | 默认值 | 备注 | 其他信息 |
|----|--------|------|-----|----|------|
| id | number | 非必须  |     |    |      |

| 名称                 | 类型          | 是否必<br>须 | 默认值 | 备注                     | 其他信息                                                              |
|--------------------|-------------|----------|-----|------------------------|-------------------------------------------------------------------|
| data               | object      | 非必须      |     |                        |                                                                   |
| ├── id             | number      | 必须       |     |                        |                                                                   |
| ├── name           | string      | 必须       |     |                        |                                                                   |
| ├── description    | string      | 必须       |     |                        |                                                                   |
| ├── uuid           | string      | 必须       |     |                        |                                                                   |
| ├ platform         | string      | 必须       |     |                        |                                                                   |
| ├── modelType      | number      | 必须       |     |                        |                                                                   |
| ├── isEdgeKit      | boolea<br>n | 非必须      |     |                        |                                                                   |
| ├── soc            | string      | 非必须      |     | 支持的AI芯片                |                                                                   |
| ├── modelName      | string      | 非必须      |     |                        |                                                                   |
| ├── nodeld         | number      | 必须       |     |                        |                                                                   |
| ├── nodeName       | string      | 非必须      |     |                        |                                                                   |
| ├── CPUArch        | string      | 非必须      |     |                        |                                                                   |
| ├ host             | string      | 必须       |     |                        |                                                                   |
| ├── port           | number      | 必须       |     |                        |                                                                   |
| ├<br>serviceStatus | string      | 必须       |     |                        | 枚举:<br>ERROR,RUNNING,DEPLOYING,DEPLOY_FAILED,OVERDUE,PAUSED,ERROR |
| ├── serviceTip     | string      | 必须       |     |                        |                                                                   |
| ├── isServiceOn    | boolea<br>n | 非必须      |     |                        |                                                                   |
| ├── serviceType    | string      | 必须       |     | process /<br>container |                                                                   |
| ├── createAt       | string      | 非必须      |     |                        |                                                                   |
| ├── updateAt       | string      | 非必须      |     |                        |                                                                   |
| msg                | string      | 非必须      |     |                        |                                                                   |
| status             | number      | 非必须      |     |                        |                                                                   |

### 心 启动服务

## 心 基本信息

Path: /iec/iapi/v1/aiservice/start

## Method : POST

# 接口描述:

### 心 请求参数

### Headers

参数名称参数值是否必须示例备注Content-Typeapplication/json是

| 名称 | 类型     | 是否必须 | 默认值 | 备注   | 其他信息 |
|----|--------|------|-----|------|------|
| id | number | 必须   |     | 模型id |      |

名称 类型 是否必须默认值 备注 其他信息

心已下发的服务列表

```
っ 基本信息
```

Path: /iec/iapi/v1/aiservice/list

Method : POST

#### 接口描述:

接口字段与 "已下发的服务详情" 一致

#### 心 请求参数

#### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

#### 心 返回数据

| 名称                | 类型           | 是否必须 | 默认值 | 备注 | 其他信息            |
|-------------------|--------------|------|-----|----|-----------------|
| data              | object<br>[] | 非必须  |     |    | item 类型: object |
| ├ id              | number       | 必须   |     |    |                 |
| ├── name          | string       | 必须   |     |    |                 |
| ├── description   | string       | 必须   |     |    |                 |
| ├── uuid          | string       | 必须   |     |    |                 |
| ├── platform      | string       | 必须   |     |    |                 |
| ├── modelType     | number       | 必须   |     |    |                 |
| ├── isEdgeKit     | boolean      | 必须   |     |    |                 |
| ├── soc           | string       | 必须   |     |    |                 |
| ├── modelName     | string       | 必须   |     |    |                 |
| ├── nodeld        | number       | 必须   |     |    |                 |
| ├── nodeName      | string       | 必须   |     |    |                 |
| CPUArch           | string       | 必须   |     |    |                 |
| ├── host          | string       | 必须   |     |    |                 |
| ├── port          | number       | 必须   |     |    |                 |
| ├── serviceStatus | string       | 必须   |     |    |                 |
| ├── serviceTip    | string       | 必须   |     |    |                 |
| ├── isServiceOn   | boolean      | 必须   |     |    |                 |
| ├── serviceType   | string       | 必须   |     |    |                 |
| ├── createAt      | string       | 必须   |     |    |                 |
| ├── updateAt      | string       | 必须   |     |    |                 |
| msg               | string       | 非必须  |     |    |                 |
| status            | number       | 非必须  |     |    |                 |

心 基本信息

Path: /iec/iapi/v1/aiservice/stop

Method : POST

# 接口描述:

心 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

Body

| 名称 | 类型     | 是否必须 | 默认值 | 备注 | 其他信息 |
|----|--------|------|-----|----|------|
| id | number | 必须   |     |    |      |

## 心 返回数据

| 名称 | 类型 | 是否必须 | 默认值 | 备注 | 其他信息 |
|----|----|------|-----|----|------|
|    |    |      |     |    |      |

## つ 删除服务

## 心 基本信息

Path: /iec/iapi/v1/aiservice/delete

Method : POST

# 接口描述:

心 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

Body

| 名称 | 类型     | 是否必须 | 默认值 | 备注 | 其他信息 |
|----|--------|------|-----|----|------|
| id | number | 必须   |     |    |      |

の 返回数据

名称 类型 是否必须默认值 备注 其他信息

の 更新服务

```
心 基本信息
```

Path: /iec/iapi/v1/aiservice/update

Method : POST

接口描述:

更新服务后,会自动触发服务重启

の 请求参数

Headers

 参数名称
 参数值
 是否必须 示例
 备注

 Content-Type
 application/json
 是

Body

| 名称   | 类型       | 是否必须 | 默认值 | 备注 | 其他信息           |
|------|----------|------|-----|----|----------------|
| id   | number   | 必须   |     |    |                |
| host | string   | 非必须  |     |    |                |
| port | number   | 非必须  |     |    |                |
| env  | array [] | 非必须  |     |    | item 类型: array |

### ∞ 返回数据

#### 名称 类型 是否必须默认值 备注 其他信息

#### 心 服务请求校验

っ 基本信息

Path: /iec/iapi/v1/aiservice/demo

Method : POST

接口描述: 单次请求下发的服务。 返回体 data 字段的内容即为原始服务的返回内容。可参考:https://ai.baidu.com/aidoc/EASYDL/lk3qy99te#预测图像

#### の 请求参数

#### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

#### Body

| 名称    | 类型     | 是否必须 | 默认值 | 备注          | 其他信息 |
|-------|--------|------|-----|-------------|------|
| id    | number | 必须   |     |             |      |
| image | string | 必须   |     | 图像的base64编码 |      |

#### 心 返回数据

| 名称             | 类型           | 是否必须 | 默认值 | 备注 | 其他信息            |
|----------------|--------------|------|-----|----|-----------------|
| data           | object       | 非必须  |     |    |                 |
| ├── error_code | number       | 非必须  |     |    |                 |
| ├── cost_ms    | number       | 非必须  |     |    |                 |
| ├── results    | object<br>[] | 非必须  |     |    | item 类型: object |
| ├── confidence | number       | 必须   |     |    |                 |
| index          | number       | 必须   |     |    |                 |
| Iabel          | string       | 必须   |     |    |                 |
| ├ x1           | number       | 必须   |     |    |                 |
| ├─_ x2         | number       | 必须   |     |    |                 |
| ├── y1         | number       | 必须   |     |    |                 |
| — y2           | number       | 必须   |     |    |                 |
| msg            | string       | 非必须  |     |    |                 |
| status         | number       | 非必须  |     |    |                 |

心 摄像头设备相关接口

心 校验摄像头

の 基本信息

Path: /iec/iapi/v1/camera/check-validity

Method : POST

# 接口描述:

# 心 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

Body

| 名称         | 类型     | 是否必须 | 默认值 | 备注                           | 其他信息            |
|------------|--------|------|-----|------------------------------|-----------------|
| nodeld     | number | 必须   |     | 节点ID                         |                 |
| cameraAddr | string | 必须   |     | 摄像头地址                        |                 |
| cameraType | string | 必须   |     | 摄像头类型,IP表示网络摄像头,LOCAL表示本地摄像头 | 枚举:<br>IP,LOCAL |

心 返回数据

| 名称     | 类型      | 是否必须 | 默认值 | 备注                 | 其他信息 |
|--------|---------|------|-----|--------------------|------|
| status | number  | 必须   |     | 状态码,0表示成功          |      |
| msg    | string  | 必须   |     | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须   |     | 摄像头是否合法            |      |

心 新增摄像头

心 基本信息

Path: /iec/iapi/v1/camera/new

Method : POST

接口描述:

の 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

| 名称                 | 类型       | 是否必<br>须 | 默认值 | 备注                                                              | 其他信息                              |
|--------------------|----------|----------|-----|-----------------------------------------------------------------|-----------------------------------|
| nodeld             | number   | 必须       |     | 节点ID                                                            |                                   |
| name               | string   | 必须       |     | 摄像头名称                                                           |                                   |
| camAddr            | string   | 非必须      |     | 摄像头地址,IP或LOCAL时必填                                               |                                   |
| camType            | string   | 必须       |     | 摄像头类型,IP表示网络摄像头,LOCAL表示本地摄像头,ONVIF表示ONVIF摄<br>像头,GB28181表示国标摄像头 | 枚举:<br>LOCAL,IP,ONVIF,GB28<br>181 |
| onvifConfig        | object   | 非必须      |     | ONVIF时必填                                                        |                                   |
| ├── ip             | string   | 必须       |     | IP                                                              |                                   |
| ├── port           | number   | 必须       |     | 端口                                                              |                                   |
| ├<br>username      | string   | 必须       |     | 用户名                                                             |                                   |
| ├<br>password      | string   | 必须       |     | 密码                                                              |                                   |
| gb28181Con<br>fig  | object   | 非必须      |     | GB28181时必填                                                      |                                   |
| ├<br>sipName       | string   | 必须       |     | 信令服务器用户名                                                        |                                   |
| ├<br>sipDeviceId   | string   | 必须       |     | 信令服务器设备ID                                                       |                                   |
| └──<br>sipPassword | string   | 必须       |     | 信令服务器密码                                                         |                                   |
| remark             | string   | 必须       |     | 摄像头备注                                                           |                                   |
| tag                | array [] | 必须       |     | 标签                                                              | item 类型: array                    |
| <u> </u>           |          | 非必须      |     |                                                                 |                                   |
| <u> </u>           |          | 非必须      |     |                                                                 |                                   |
| timeRange          | array [] | 必须       |     | 摄像头运行区间                                                         | item 类型: array                    |
| <u> </u>           |          | 非必须      |     |                                                                 |                                   |
| <u> </u>           |          | 非必须      |     | 秒                                                               |                                   |
| frameExtract       | object   | 必须       |     | 抽帧配置                                                            |                                   |
| ├<br>everySecond   | number   | 必须       |     | 每多少秒                                                            |                                   |
| ├── frames         | number   | 必须       |     | 抽多少帧                                                            |                                   |

| 名称       | 类型     | 是否必须 | 默认值 | 备注                 | 其他信息 |
|----------|--------|------|-----|--------------------|------|
| status   | number | 必须   |     | 状态码,0表示成功          |      |
| msg      | string | 必须   |     | 错误信息,status不为0时有意义 |      |
| data     | object | 必须   |     |                    |      |
| ├── id   | number | 必须   |     | 当前节点唯一ID           |      |
| ├── uuid | string | 必须   |     | 全局唯一ID             |      |

# 心 摄像头列表

### 心 基本信息

Path: /iec/iapi/v1/camera/list

## 接口描述:

心 请求参数

### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

#### Body

| 名称       | 类型     | 是否必须 | 默认值 | 备注       | 其他信息 |
|----------|--------|------|-----|----------|------|
| pageNo   | number | 非必须  |     | 页数       |      |
| pageSize | number | 非必须  |     | 每页数量     |      |
| nodeld   | number | 非必须  |     | 节点ID     |      |
| filters  | object | 非必须  |     | 筛选项      |      |
| ├── name | string | 非必须  |     | 按摄像头名称筛选 |      |

## 心 返回数据

| 名称               | 类型           | 是否必须 | 默认值 | 备注                 | 其他信息                        |
|------------------|--------------|------|-----|--------------------|-----------------------------|
| status           | number       | 必须   |     | 状态码,0表示成功          |                             |
| msg              | string       | 必须   |     | 错误信息,status不为0时有意义 |                             |
| data             | object<br>[] | 必须   |     |                    | item 类型: object             |
| ├── id           | number       | 必须   |     | 摄像头ID              |                             |
| ├── uuid         | string       | 必须   |     | 摄像头全局唯一ID          |                             |
| ├── nodeld       | number       | 必须   |     | 节点ID               |                             |
| ├── name         | string       | 必须   |     | 摄像头名称              |                             |
| ├── remark       | string       | 必须   |     | 摄像头备注              |                             |
| ├── cameraType   | string       | 必须   |     | 摄像头类型              | 枚举: LCOAL,IP,ONVIF,GB28181  |
| ├── cameraStatus | string       | 必须   |     | 摄像头状态              | 枚举: RUNNING, ERROR, STOPPED |

## 心 摄像头详情

```
っ 基本信息
```

Path: /iec/iapi/v1/camera/get

Method : GET

接口描述:

の 请求参数

### Query

| 参数名称 | 是否必须 | 示例 | 备注    |
|------|------|----|-------|
| id   | 是    |    | 摄像头ID |

の 返回数据

| 名称                       | 类型       | 是否必须 | 默认值 | 备注                 | 其他信息                        |
|--------------------------|----------|------|-----|--------------------|-----------------------------|
| status                   | number   | 必须   |     | 状态码,0表示成功          |                             |
| msg                      | string   | 必须   |     | 错误信息,status不为0时有意义 |                             |
| data                     | object   | 必须   |     |                    |                             |
| ├── id                   | number   | 必须   |     | 摄像头ID              |                             |
| ├── uuid                 | string   | 必须   |     | 摄像头全局唯一ID          |                             |
| ├── nodeld               | number   | 必须   |     | 节点ID               |                             |
| ├── name                 | string   | 必须   |     | 摄像头名称              |                             |
| remark                   | string   | 必须   |     | 摄像头备注              |                             |
| ├── cameraType           | string   | 必须   |     | 摄像头类型              | 枚举: LCOAL,IP,ONVIF,GB28181  |
| ├── cameraAddr           | string   | 必须   |     | 摄像头地址,IP或LOCAL时有意义 |                             |
| ├── onvifConfig          | object   | 非必须  |     |                    |                             |
| ├── ip                   | string   | 必须   |     | IP                 |                             |
| ├ port                   | number   | 必须   |     | 端口                 |                             |
| ├── username             | string   | 必须   |     | 用户名                |                             |
| ├── password             | string   | 必须   |     | 密码                 |                             |
| ├── gb28181Config        | object   | 非必须  |     |                    |                             |
| ├── sipName              | string   | 必须   |     | 信令服务器用户名           |                             |
| ├── sipDeviceId          | string   | 必须   |     | 信令服务器设备ID          |                             |
| ├── sipPassword          | string   | 必须   |     | 信令服务器密码            |                             |
| ├── frameExtractInterval | number   | 必须   |     | 每多少秒               |                             |
| ├── frameExtractNum      | number   | 必须   |     | 抽多少帧               |                             |
| ├ onlineTime             | array [] | 必须   |     | 摄像头运行区间            | item 类型: array              |
| <u> </u>                 |          | 非必须  |     |                    |                             |
| <u> </u>                 |          | 非必须  |     | 秒                  |                             |
| ├── cameraStatus         | string   | 必须   |     | 摄像头状态              | 枚举: RUNNING, ERROR, STOPPED |

# ℗更新摄像头

```
心 基本信息
```

Path : /iec/iapi/v1/camera/update

Method : POST

### 接口描述:

心 请求参数

### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

| 名称                | 类型       | 是否必<br>须 | 默认值 | 备注                                                              | 其他信息                              |
|-------------------|----------|----------|-----|-----------------------------------------------------------------|-----------------------------------|
| id                | number   | 必须       |     | 摄像头ID                                                           |                                   |
| name              | string   | 必须       |     | 摄像头名称                                                           |                                   |
| camAddr           | string   | 非必须      |     | 摄像头地址,IP或LOCAL时必填                                               |                                   |
| camType           | string   | 必须       |     | 摄像头类型,IP表示网络摄像头,LOCAL表示本地摄像头,ONVIF表示ONVIF摄<br>像头,GB28181表示国标摄像头 | 枚举:<br>LOCAL,IP,ONVIF,GB28<br>181 |
| onvifConfig       | object   | 非必须      |     | ONVIF时必填                                                        |                                   |
| ├── ip            | string   | 必须       |     | IP                                                              |                                   |
| ├── port          | string   | 必须       |     | 端口                                                              |                                   |
| └──<br>username   | string   | 必须       |     | 用户名                                                             |                                   |
| ├<br>password     | string   | 必须       |     | 密码                                                              |                                   |
| gb28181Con<br>fig | object   | 非必须      |     | GB28181时必填                                                      |                                   |
| ├<br>sipName      | string   | 必须       |     | 信令服务器用户名                                                        |                                   |
| ├<br>sipDeviceId  | string   | 必须       |     | 信令服务器设备ID                                                       |                                   |
| ├<br>sipPassword  | string   | 必须       |     | 信令服务器密码                                                         |                                   |
| remark            | string   | 必须       |     | 摄像头备注                                                           |                                   |
| tag               | array [] | 必须       |     | 标签                                                              | item 类型: array                    |
| <u> </u>          |          | 非必须      |     |                                                                 |                                   |
| <u> </u>          |          | 非必须      |     |                                                                 |                                   |
| timeRange         | array [] | 必须       |     | 摄像头运行区间                                                         | item 类型: array                    |
| <u> </u>          |          | 非必须      |     |                                                                 |                                   |
| <u> </u>          |          | 非必须      |     | 秒                                                               |                                   |
| frameExtract      | object   | 必须       |     | 抽岐配置                                                            |                                   |
| ├<br>everySecond  | number   | 必须       |     | 每多少秒                                                            |                                   |
| ├── frames        | number   | 必须       |     | 抽多少帧                                                            |                                   |

| 名称     | 类型      | 是否必须 | 默认值 | 备注                  | 其他信息 |
|--------|---------|------|-----|---------------------|------|
| status | number  | 必须   |     | 状态码,0表示成功           |      |
| msg    | string  | 必须   |     | 错误信息, status不为0时有意义 |      |
| data   | boolean | 必须   |     | 修改是否成功              |      |

# 心 启动摄像头

の 基本信息

Path: /iec/iapi/v1/camera/start

Method : POST

接口描述:

# 心 请求参数

#### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

#### Body

| 名称 | 类型     | 是否必须 | 默认值 | 备注    | 其他信息 |
|----|--------|------|-----|-------|------|
| id | number | 必须   |     | 摄像头ID |      |

### の 返回数据

| 名称     | 类型      | 是否必须 | 默认值 | 备注                 | 其他信息 |
|--------|---------|------|-----|--------------------|------|
| status | number  | 必须   |     | 状态码,0表示成功          |      |
| msg    | string  | 必须   |     | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须   |     | 开启摄像头是否成功          |      |

### ◎ 停止摄像头

## 心 基本信息

Path: /iec/iapi/v1/camera/stop

Method : POST

### 接口描述:

の 请求参数

#### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

### Body

| 名称 | 类型     | 是否必须 | 默认值 | 备注    | 其他信息 |
|----|--------|------|-----|-------|------|
| id | number | 必须   |     | 摄像头ID |      |

### ∞ 返回数据

| 名称     | 类型      | 是否必须 🖁 | 默认值 | 备注                 | 其他信息 |
|--------|---------|--------|-----|--------------------|------|
| status | number  | 必须     |     | 状态码,0表示成功          |      |
| msg    | string  | 必须     |     | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须     |     | 关闭摄像头是否成功          |      |

```
ℬ 删除摄像头
```

```
の 基本信息
```

Path: /iec/iapi/v1/camera/delete

Method : POST

接口描述:

の 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 |
|--------------|------------------|------|----|
| Content-Type | application/json | 是    |    |

Body

| 名称 | 类型     | 是否必须 | 默认值 | 备注    | 其他信息 |
|----|--------|------|-----|-------|------|
| id | number | 必须   |     | 摄像头ID |      |

#### の 返回数据

| 名称     | 类型      | 是否必须 | 默认值 | 备注                 | 其他信息 |
|--------|---------|------|-----|--------------------|------|
| status | number  | 必须   |     | 状态码,0表示成功          |      |
| msg    | string  | 必须   |     | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须   |     | 删除摄像头是否成功          |      |

备注

## 心物联网设备相关接口

心 新増设备

### 心 基本信息

Path: /iec/iapi/v1/plc/new

Method : POST

## 接口描述:

### の 请求参数

#### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

| 名称             | 类型     | 是否必<br>须 | 默认值 | 备注        | 其他信息             |
|----------------|--------|----------|-----|-----------|------------------|
| nodeld         | number | 必须       |     | 节点ID      |                  |
| name           | string | 必须       |     | 设备名称      |                  |
| deviceType     | string | 必须       |     | 设备类型      | 枚举: modbus,opcua |
| modbus         | object | 非必须      |     | modbus时必填 |                  |
| ├<br>protocol  | string | 必须       |     | 协议类型      | 枚举: tcp,rtu      |
| ├── rtu        | object | 非必须      |     | rtu时必填    |                  |
| ├── port       | string | 必须       |     | 発口        |                  |
| ⊢—<br>baudrate | number | 必须       |     | 波特率       |                  |
| ├<br>databit   | number | 必须       |     | 数据位       | 枚举: 5,6,7,8      |
| ├<br>stopbit   | number | 必须       |     | 停止位       | 枚举: 1,2          |
| <br>parity     | string | 必须       |     | 校验位       | 枚举: N,E,O        |
| ├── tcp        | object | 非必须      |     | tcp时必填    |                  |
| ⊢–<br>address  | string | 必须       |     | 连接地址      |                  |

| ├── port                  | number       | 必须  | 端口        |                                                                       |
|---------------------------|--------------|-----|-----------|-----------------------------------------------------------------------|
| ├<br>slaveld              | number       | 必须  | 从站号       |                                                                       |
| interval                  | number       | 必须  | 采样间隔      |                                                                       |
| opcua                     | object       | 非必须 | opcua时必填  |                                                                       |
| ├<br>endpoint             | string       | 必须  | 通道地址      |                                                                       |
| ├──<br>securityPolic<br>y | string       | 必须  | 安全策略      | 枚举:<br>None,Basic256Sha256,Aes128Sha256RsaOaep,Ae<br>s256Sha256RsaPss |
| ├──<br>securityMod<br>e   | string       | 必须  | 安全模式      | 枚举: None,Sign,SignAndEncrypt                                          |
| ├<br>certificate          | string       | 非必须 | 数字证书      |                                                                       |
| ├<br>privateKey           | string       | 非必须 | 密钥证书      |                                                                       |
| ├<br>username             | string       | 非必须 | 用户名       |                                                                       |
| ├<br>password             | string       | 非必须 | 密码        |                                                                       |
| ├<br>timeout              | number       | 必须  | 连接超时时间    |                                                                       |
| interval                  | number       | 必须  | 采样间隔      |                                                                       |
| remark                    | string       | 必须  | 设备备注      |                                                                       |
| tag                       | array []     | 必须  | 标签        | item 类型: array                                                        |
| ⊢                         |              | 非必须 |           |                                                                       |
| ⊢                         |              | 非必须 |           |                                                                       |
| attributes                | object<br>[] | 必须  | 属性        | item 类型: object                                                       |
| ├── id                    | string       | 必须  | 标识符       |                                                                       |
| ├── name                  | string       | 必须  | 属性名       |                                                                       |
| ├── type                  | string       | 必须  | 类型        | 枚举: bool,int16,int32,int64,float32,float64,string                     |
| ├<br>defaultValue         | string       | 非必须 | 默认值       |                                                                       |
| ├── unit                  | string       | 必须  | 单位        |                                                                       |
| ├<br>required             | boolea<br>n  | 必须  | 是否必填      |                                                                       |
| properties                | object<br>[] | 必须  | 测点        | item 类型: object                                                       |
| ├── id                    | string       | 必须  | 标识符       |                                                                       |
| ├── name                  | string       | 必须  | 测点名       |                                                                       |
| ├── type                  | string       | 必须  | 类型        | 枚举: bool,int16,int32,int64,float32,float64,string                     |
| ├── mode                  | string       | 必须  | 读写类型      | 枚举: ro,rw                                                             |
| ├── unit                  | string       | 必须  | 单位        |                                                                       |
| ├<br>modbus               | object       | 非必须 | modbus时必填 |                                                                       |

| ├── func      | number | 必须  | 寄存器类型:1.线圈寄存器2.离散输入寄存器3.<br>保持寄存器4.输入寄存器 | 枚举: 1,2,3,4 |
|---------------|--------|-----|------------------------------------------|-------------|
| ⊢—<br>address | string | 必须  | 寄存器地址                                    |             |
| ├<br>quantity | number | 必须  | 寄存器数量                                    |             |
| ├── opcua     | object | 非必须 | opcua时必填                                 |             |
| ⊢–<br>nodeid  | string | 必须  | 节点ID                                     |             |

| 名称       | 类型     | 是否必须 | 默认值 | 备注                 | 其他信息 |
|----------|--------|------|-----|--------------------|------|
| status   | number | 必须   |     | 状态码,0表示成功          |      |
| msg      | string | 必须   |     | 错误信息,status不为0时有意义 |      |
| data     | object | 必须   |     |                    |      |
| ├ id     | number | 必须   |     | 当前节点唯一ID           |      |
| ├── uuid | string | 必须   |     | 全局唯一ID             |      |

## 心 设备列表

### の 基本信息

Path: /iec/iapi/v1/plc/list

#### Method : POST

## 接口描述:

## 心 请求参数

### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

#### Body

| 名称       | 类型     | 是否必须 | 默认值 | 备注    | 其他信息 |
|----------|--------|------|-----|-------|------|
| pageNo   | number | 非必须  |     | 页数    |      |
| pageSize | number | 非必须  |     | 每页数量  |      |
| nodeld   | number | 非必须  |     | 节点ID  |      |
| filters  | object | 非必须  |     | 筛选项   |      |
| ├── name | string | 非必须  |     | 按名称筛选 |      |

## 心 返回数据

| 名称             | 类型           | 是否必须 | 默认值 | 备注                 | 其他信息                      |
|----------------|--------------|------|-----|--------------------|---------------------------|
| status         | number       | 必须   |     | 状态码,0表示成功          |                           |
| msg            | string       | 必须   |     | 错误信息,status不为0时有意义 |                           |
| data           | object<br>[] | 必须   |     |                    | item 类型: object           |
| ├── id         | number       | 必须   |     | 设备ID               |                           |
| ├── uuid       | string       | 必须   |     | 设备全局唯一ID           |                           |
| ├── nodeld     | number       | 必须   |     | 节点ID               |                           |
| ├── name       | string       | 必须   |     | 设备名称               |                           |
| ├── remark     | string       | 必须   |     | 设备备注               |                           |
| ├── deviceType | string       | 必须   |     | 设备类型               |                           |
| ├── status     | string       | 必须   |     | 设备状态               | 枚举: RUNNING,ERROR,STOPPED |

## 心 设备详情

#### 心 基本信息

Path: /iec/iapi/v1/plc/get

# Method : GET

## 接口描述:

## 心 请求参数

#### Query

| 参数名称 | 是否必须 | 示例 | 备注   |
|------|------|----|------|
| id   | 是    |    | 设备ID |

#### 心 返回数据

| 名称             | 类型     | 是否必<br>须 | 默认值 | 备注                 | 其他信息             |
|----------------|--------|----------|-----|--------------------|------------------|
| status         | number | 必须       |     | 状态码,0表示成功          |                  |
| msg            | string | 必须       |     | 错误信息,status不为0时有意义 |                  |
| data           | object | 必须       |     |                    |                  |
| ├── id         | number | 必须       |     | 设备ID               |                  |
| ├── uuid       | string | 必须       |     | 设备全局唯一ID           |                  |
| ├── nodeld     | number | 必须       |     | 节点ID               |                  |
| ├── name       | string | 必须       |     | 设备名称               |                  |
| ├── deviceType | string | 必须       |     | 设备类型               | 枚举: modbus,opcua |
| ├── modbus     | object | 非必须      |     | modbus时必填          |                  |
| ├── protocol   | string | 必须       |     | 协议类型               | 枚举: tcp,rtu      |
| ├── rtu        | object | 非必须      |     | rtu时必填             |                  |
| ├── port       | string | 必须       |     | 端口                 |                  |
| ├<br>baudrate  | number | 必须       |     | 波特率                |                  |
| ├── databit    | number | 必须       |     | 数据位                | 枚举: 5,6,7,8      |
| ├── stopbit    | number | 必须       |     | 停止位                | 枚举: 1,2          |
| ├── parity     | string | 必须       |     | 校验位                | 枚举: N,E,O        |
| ├── tcp        | object | 非必须      |     | tcp时必填             |                  |
|                |        |          |     |                    |                  |

| address             | string       | 必须   | 连接地址                             |                                                          |
|---------------------|--------------|------|----------------------------------|----------------------------------------------------------|
| ├── port            | number       | 必须   | 端口                               |                                                          |
|                     | number       | 必须   | 从站号                              |                                                          |
| <u> </u>            | number       | /送彡而 | 采样间隔                             |                                                          |
| interval            |              |      |                                  |                                                          |
| - opcua             | object       | 非必须  | opcua时必填                         |                                                          |
| endpoint            | string       | 必须   | 通道地址                             |                                                          |
| securityPolicy      | string       | 必须   | 安全策略                             |                                                          |
| ⊢––<br>securityMode | string       | 必须   | 安全模式                             |                                                          |
| ├<br>certificate    | string       | 非必须  | 数字证书                             |                                                          |
| ⊢—<br>privateKey    | string       | 非必须  | 密钥证书                             |                                                          |
| ⊢—<br>username      | string       | 非必须  | 用户名                              |                                                          |
| ⊢—<br>password      | string       | 非必须  | 密码                               |                                                          |
| ├── timeout         | number       | 必须   | 连接超时时间                           |                                                          |
| ├── interval        | number       | 必须   | 采样间隔                             |                                                          |
| ├── remark          | string       | 必须   | 设备备注                             |                                                          |
| ├── tag             | array []     | 必须   | 标签                               | item 类型: array                                           |
| $\vdash$            |              | 非必须  |                                  |                                                          |
| <u> </u>            |              | 非必须  |                                  |                                                          |
| ├── attributes      | object<br>[] | 必须   | 属性                               | item 类型: object                                          |
| ├── id              | string       | 必须   | 标识符                              |                                                          |
| ├── name            | string       | 必须   | 属性名                              |                                                          |
| ├── type            | string       | 必须   | 类型                               | 枚举:<br>bool,int16,int32,int64,float32,float64,s<br>tring |
| ├<br>defaultValue   | string       | 非必须  | 默认值                              |                                                          |
| ├── unit            | string       | 必须   | 单位                               |                                                          |
| ├── required        | boolea<br>n  | 必须   | 是否必填                             |                                                          |
| ├── properties      | object<br>[] | 必须   | 测点                               | item 类型: object                                          |
| ├── id              | string       | 必须   | 标识符                              |                                                          |
| ├── name            | string       | 必须   | 测点名                              |                                                          |
| ├── type            | string       | 必须   | 类型                               | 枚举:<br>bool,int16,int32,int64,float32,float64,s<br>tring |
| ├── mode            | string       | 必须   | 读写类型                             | 枚举: ro,rw                                                |
| ├── unit            | string       | 必须   | 单位                               |                                                          |
| modbus              | object       | 非必须  | modbus时必填                        |                                                          |
|                     |              |      | 客方哭米刑・1 绊開客方哭 0 室勘絵 λ 客方哭 3 促挂客方 |                                                          |

| ├── func      | number | 必须  | リゴロボス主・シンジョリココロシンの私間ハリココロ シットリッリココ<br>器 4.输入寄存器 | 枚举: 1,2,3,4               |
|---------------|--------|-----|-------------------------------------------------|---------------------------|
| ⊢—<br>address | string | 必须  | 寄存器地址                                           |                           |
| ├<br>quantity | number | 必须  | 寄存器数量                                           |                           |
| ├── opcua     | object | 非必须 | opcua时必填                                        |                           |
| ├── nodeid    | string | 必须  | 节点ID                                            |                           |
| ├── status    | string | 必须  | 设备状态                                            | 枚举: RUNNING,ERROR,STOPPED |
| ├── url       | string | 非必须 | 测点数据获取websocket连接地址                             |                           |
| ├── realData  | object | 非必须 | 当前测点数据                                          |                           |

## ∞ 更新设备

```
の 基本信息
```

Path: /iec/iapi/v1/plc/update

Method : POST

接口描述:

# 心 请求参数

# Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

| 名称             | 类型     | 是否必<br>须 | 默认值 | 备注        | 其他信息             |
|----------------|--------|----------|-----|-----------|------------------|
| id             | number | 必须       |     | 设备ID      |                  |
| name           | string | 必须       |     | 设备名称      |                  |
| deviceType     | string | 必须       |     | 设备类型      | 枚举: modbus,opcua |
| modbus         | object | 非必须      |     | modbus时必填 |                  |
| ├<br>protocol  | string | 必须       |     | 协议类型      | 枚举: tcp,rtu      |
| ├── rtu        | object | 非必须      |     | rtu时必填    |                  |
| ├── port       | string | 必须       |     | 端口        |                  |
| ├<br>baudrate  | number | 必须       |     | 波特率       |                  |
| <br>databit    | number | 必须       |     | 数据位       | 枚举: 5,6,7,8      |
| ⊢––<br>stopbit | number | 必须       |     | 停止位       | 枚举: 1,2          |
| ⊢—<br>parity   | string | 必须       |     | 校验位       | 枚举: N,E,O        |
| ├── tcp        | object | 非必须      |     | tcp时必填    |                  |
| ⊢—<br>address  | string | 必须       |     | 连接地址      |                  |
| ├── port       | number | 必须       |     | 端口        |                  |
| ⊢–<br>slaveld  | number | 必须       |     | 从站号       |                  |

| interval          | number   | 必须     | 采样间隔                                        |                                                   |
|-------------------|----------|--------|---------------------------------------------|---------------------------------------------------|
| opcua             | object   | 非必须    | opcua时必填                                    |                                                   |
| <b>—</b>          | otring   | 以初     | 通送种中                                        |                                                   |
| endpoint          | sung     | 化沙贝    | 地但他近                                        |                                                   |
| ⊢<br>             |          | NIT    | r A tran                                    | 枚举:                                               |
| v                 | string   | 化沙贝    | 女至東略                                        | None,Basic256Sha256,Aes128Sha256RsaUaep,Ae        |
|                   |          |        |                                             |                                                   |
| securityMod       | string   | 非必须    | 安全模式                                        | 枚举: None,Sign,SignAndEncrypt                      |
| е                 |          |        |                                             |                                                   |
| ⊢–<br>certificate | string   | 非必须    | 数字证书                                        |                                                   |
|                   |          |        |                                             |                                                   |
| privateKey        | string   | 非必须    | 密钥证书                                        |                                                   |
| <u> </u>          | string   | 非必须    | 田户夕                                         |                                                   |
| username          | String   | 1F207X |                                             |                                                   |
|                   | string   | 必须     | 密码                                          |                                                   |
|                   |          |        |                                             |                                                   |
| timeout           | number   | 必须     | 连接超时时间                                      |                                                   |
| interval          | number   | 必须     | 采样间隔                                        |                                                   |
| remark            | string   | 必须     | 设备备注                                        |                                                   |
| tag               | array [] | 必须     | 标签                                          | item 类型: array                                    |
| <u> </u>          |          | 非必须    |                                             |                                                   |
| <u> </u>          |          | 非必须    |                                             |                                                   |
| attributes        | object   | 必须     | 属性                                          | item 类型: object                                   |
| ├── id            | string   | 必须     | 标识符                                         |                                                   |
| - name            | string   | 必须     | 属性名                                         |                                                   |
| ├── type          | string   | 必须     | 类型                                          | 枚举: bool,int16,int32,int64,float32,float64,string |
| <u> </u>          |          |        | 题                                           |                                                   |
| defaultValue      | string   | F 化分贝  | 為认1                                         |                                                   |
| ├── unit          | string   | 必须     | 单位                                          |                                                   |
|                   | boolea   | 必须     | 是否必填                                        |                                                   |
| required          | n        |        |                                             |                                                   |
| properties        | []       | 必须     | 测点                                          | item 类型: object                                   |
| ├── id            | string   | 必须     | 标识符                                         |                                                   |
| ├── name          | string   | 必须     | 测点名                                         |                                                   |
| ├── type          | string   | 必须     | 类型                                          | 枚举: bool,int16,int32,int64,float32,float64,string |
| ├── mode          | string   | 必须     | 读写类型                                        | 枚举: ro,rw                                         |
| ├── unit          | string   | 必须     | 单位                                          |                                                   |
| ├<br>modbus       | object   | 非必须    | modbus时必填                                   |                                                   |
| ├── func          | number   | 必须     | 寄存器类型:1.线圈寄存器 2.离散输入寄存器 3.<br>保持寄存器 4.输入寄存器 | 枚举: 1,2,3,4                                       |
| <u> </u>          | etring   | 心义《而   | 寄左哭抽扯                                       |                                                   |

| address        | oung   |     |          |  |
|----------------|--------|-----|----------|--|
| ⊣—<br>quantity | number | 必须  | 寄存器数量    |  |
| ├── opcua      | object | 非必须 | opcua时必填 |  |
| ⊢—<br>nodeid   | string | 必须  | 节点ID     |  |

| 名称     | 类型      | 是否必须 | 默认值 | 备注                 | 其他信息 |
|--------|---------|------|-----|--------------------|------|
| status | number  | 必须   |     | 状态码,0表示成功          |      |
| msg    | string  | 必须   |     | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须   |     | 修改是否成功             |      |

### の 删除设备

## 心 基本信息

Path: /iec/iapi/v1/plc/delete

Method : POST

## 接口描述:

#### 心 请求参数

### Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

Body

| 名称 | 类型     | 是否必须 | 默认值 | 备注   | 其他信息 |
|----|--------|------|-----|------|------|
| id | number | 必须   |     | 设备ID |      |

### 心 返回数据

| 名称     | 类型      | 是否必须 默认值 | 备注                 | 其他信息 |
|--------|---------|----------|--------------------|------|
| status | number  | 必须       | 状态码,0表示成功          |      |
| msg    | string  | 必须       | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须       | 删除设备是否成功           |      |

₯ 写测点

### の 基本信息

Path: /iec/iapi/v1/plc/property/write

Method : POST

## 接口描述:

## 心 请求参数

Headers

| 参数名称         | 参数值              | 是否必须 | 示例 | 备注 |
|--------------|------------------|------|----|----|
| Content-Type | application/json | 是    |    |    |

| 名称            | 类型     | 是否必须 | 默认值 | 备注    | 其他信息 |
|---------------|--------|------|-----|-------|------|
| id            | number | 必须   |     | 设备ID  |      |
| propertyld    | string | 必须   |     | 测点标识符 |      |
| propertyValue | any    | 必须   |     | 测点值   |      |

| 名称     | 类型      | 是否必须 | 默认值 | 备注                 | 其他信息 |
|--------|---------|------|-----|--------------------|------|
| status | number  | 必须   |     | 状态码,0表示成功          |      |
| msg    | string  | 必须   |     | 错误信息,status不为0时有意义 |      |
| data   | boolean | 必须   |     | 修改测点值是否成功          |      |

## 心 功能更新记录

| 时间         | 版本    | 说明  |
|------------|-------|-----|
| 2023-05-25 | 1.0.0 | 第一版 |