

EasyData 文档



【版权声明】

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】



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

目录	2
EasyData平台简介	3
什么是EasyData	3
数据管理	5
数据查看	5
图片数据导入	6
文本数据导入	17
音频数据导入	23
视频数据导入	27
数据导出	31
表格数据导入	33
数据质检	34
数据采集	36
数据采集功能简介	36
接入摄像头采集图片	36
采集摄像头管理	36
原始数据查看	44
云服务数据回流	46
配置数据授权	46
原始数据查看	46
数据清洗	47
清洗任务管理	47
文本数据清洗	52
数据标注	56
在线标注	56
选择数据集版本	56
图像智能标注	62
文本智能标注	67
寻求标注支持	72
多人标注	73
标签组管理	77
数据增强	79
增强任务管理	79
联系我们	82
产品升级记录	83
数据增强	84

EasyData平台简介

什么是EasyData

🔗 产品介绍

EasyData是百度大脑推出的智能数据服务平台，支持面向各行各业有AI开发的企业用户及开发者提供一站式数据服务工具。主要围绕AI开发过程中所需要数据采集、数据清洗、数据标注等业务需求提供完整的数据服务。目前EasyData已经支持图片、文本、音频、视频四类基础数据的处理,也初步支持机器学习数据的存储。

同时EasyData已与EasyDL、BML数据管理模块打通，可以将EasyData处理的数据应用于EasyDL、BML模型训练。

[了解EasyDL](#)

[了解BML](#)

🔗 功能介绍

🔗 数据采集

EasyData提供两种数据采集方案

1. 从摄像头采集图片数据：

- **采集方案：**提供本地采集软件，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EasyData管理

- **方案优势：**

操作便捷：直接对接采集摄像头硬件，自动将数据从本地传至云端

采集效率高：减少数据中转环节，采集、抽帧、上传一站式解决

2. 通过云服务数据回流采集数据

- **采集方案：**AI模型训练后可能需要持续迭代和优化模型效果，可以通过调用EasyDL、BML云服务接口并开通该采集服务，将实际业务数据及识别结果可视化查看，针对性的选择高质量数据。

- **方案优势：**

数据匹配度高：直接对接预测服务云服务接口，训练数据与实际业务匹配性更强

难例挖掘：支持通过置信度等多种维度筛选图片，挖掘难例，更有针对性地补充AI模型训练数据

🔗 数据标注

- **预置丰富的标注模板**

- 图像: 图像分类、物体检测、图像分割
- 文本: 文本分类、短文本相似度、情感倾向分析、文本实体抽取
- 音频: 声音分类
- 视频: 视频分类

- **支持智能标注**

智能标注提供人机交互协作标注，最高降低90%的标注成本。目前智能标注已支持物体检测、图像分割、文本分类等数据类型。

- **支持多人标注**

多人标注通过团队协作完成标注任务，提高标注效率。目前多人标注已支持图像、文本、语音、视频等数据方向。

- 提供数据标注服务

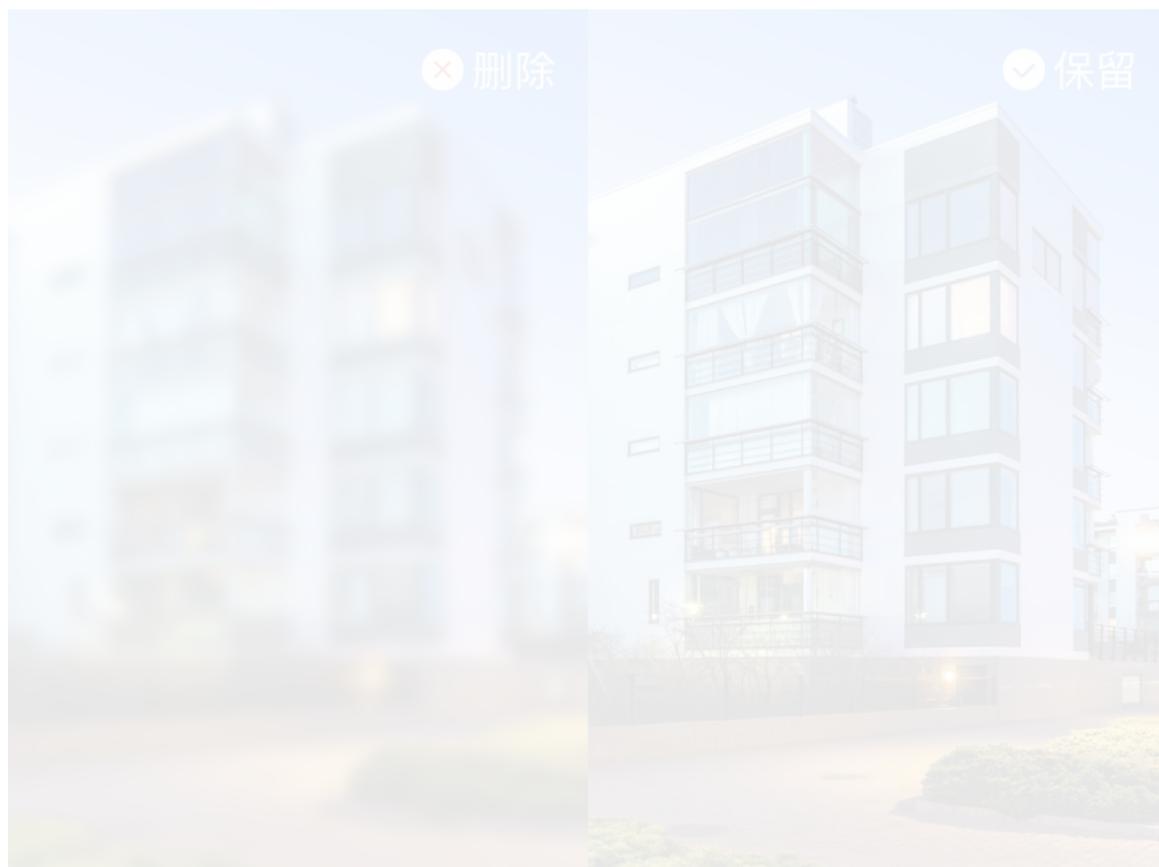
百度众测及数据服务商标注支持EasyData已全面对接百度众测及AI市场中的优质数据服务商，可以通过EasyData平台面向百度众测提交详细的标注需求。



🔄 数据清洗

支持四类数据清洗策略：

- 图片去模糊：过滤清晰度较低的图片，保证数据质量
- 图片去重：过滤大量重复的图片，提高关键图片处理效率
- 图片批量裁剪：批量裁剪图片中的无关元素，提升数据质量
- 图片旋转：校正采集图片的角度，方便进行下一步的处理
- 过滤无人脸图片：通过调用百度人脸检测服务过滤出图片中没有人脸的图片
- 过滤无人体图片：通过调用百度人体检测或人像分割服务过滤出图片中没有人体的图片



数据管理

数据查看

目录

1. [查看无标注信息数据](#)
2. [查看有标注信息数据](#)

[查看无标注信息数据](#)

数据筛选

无标注信息支持根据数据来源、导入日期、标注日期进行筛选及查看

我的数据总览 > 【图片】图片分类/V1/查看

无标注信息 有标注信息

fish-杜勇V1版本的图片列表

筛选

查看有标注信息数据

数据筛选

有标注信息支持根据标注状态、数据来源、导入日期、标注日期进行筛选及查看

- 标注状态：可以选择待确认及已确认，其中待确认的数据有两个来源，来源1为智能标注机器标注结束但尚未完成人工确认的数据；来源2为云服务调用采集而来的数据。对于待确认的数据可以通过查看大图一键确认。
- 数据来源：可以选择本地上传、摄像头采集、云服务调用。
- 导入日期：指筛选相应日期导入完成的数据
- 标注日期：指筛选相应日期标注完成的数据

无标注信息 有标注信息

aaaV2版本的图片列表

筛选

查看大图与确认标注信息

点击查看数据可以通过点击确认标注，将待确认数据更改为已标注数据。

图片数据导入

目录

1. [导入未标注数据](#)
2. [导入已标注数据](#)

🔗 导入未标注数据

为选定的图像数据集导入未标注数据（即原始图片）的方式有两种。一种为从本地选择图片文件；一种为从线上【EasyData-数据服务】中其余同为图片类型数据集中导入。

🔗 从本地上传图片导入

选择未标注数据-本地-上传图片，点击上传图片后选择本地图片文件进行导入。

数据格式要求

- 1、图片类型为jpg/png/bmp/jpeg，单次上传限制100个文件。
- 2、图片大小限制在14M内，长宽比在3：1以内，其中最长边需要小于4096px，最短边需要大于30px。
- 3、单个数据集大小限制为10万图片，超出后可被忽略。

特别注意：

- 对同一数据集存在多个内容完全一致的图片，将会做去重处理。
- 为保证模型训练效果，所上传的图片应与实际业务场景的图片（光线、角度、采集设备）尽可能一致

Baidu 大脑 | EasyData

我的数据总览 > 螺丝螺母测试/V1/导入

创建信息

数据集名称	螺丝螺母测试	版本号	V1
标注类型	物体检测	标注模板	矩形框标注

标注信息

数据总量	0	已标注	0
标签个数	0	目标数	0
待确认	0	大小	0M

数据清洗

暂未做过数据清洗任务

导入数据

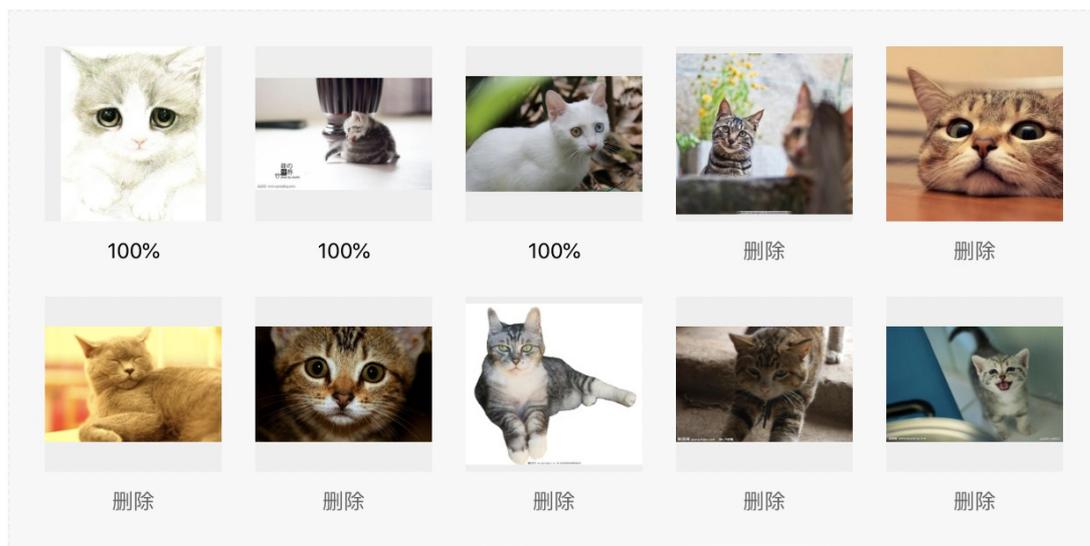
数据标注状态 无标注信息 有标注信息

导入方式 本地数据集 上传图片

上传图片 上传图片

确认并返回

上传图片



开始上传

继续添加

从本地上传压缩包导入

选择**未标注数据-本地-上传压缩包**，点击上传压缩包后选择本地图片压缩包完成数据的批量导入。

数据格式要求

- 1、压缩包仅支持zip格式，大小限制5GB以内
- 2、压缩包内图片格式要求为：图片类型为jpg/png/bmp/jpeg，图片大小限制在14M内，长宽比在3：1以内，其中最长边需要小于4096px，最短边需要大于30px。
- 3、单个数据集大小限制为10万图片，超出后可被忽略。

特别注意：

- 对同一数据集存在多个内容完全一致的图片，将会做去重处理。
- 为保证模型训练效果，所上传的图片应与实际业务场景的图片（光线、角度、采集设备）尽可能一致

Baidu 百度智能云 | EasyData

我的数据总览 > 螺丝螺母测试/V1/导入

创建信息

数据集名称	螺丝螺母测试	版本号	V1
标注类型	物体检测	标注模板	矩形框标注

标注信息

数据总量	0	已标注	0
标签个数	0	目标数	0
待确认	0	大小	0M

数据清洗

暂未做过数据清洗任务

导入数据

数据标注状态 无标注信息 有标注信息

导入方式 本地数据集 上传压缩包

上传压缩包 上传压缩包

确认并返回

导入线上已有数据

选择未标注数据-已有数据集-导入已有数据集，支持将已经上传在其他EasyData同为图片类的数据集进行导入。在视觉任务中往往会有存在更换模型类型的可能性，比如某个业务场景问题既可以用图像分类解决又可以用物体检测解决，通过这个功能可以有效地快速切换数据类型，分别测试效果。

由于这里选择上传的是未标注数据，那么在导入线上已有数据的情况下，可以选择导入同为图片类型数据集的全部数据但去掉标注；或者选择未被标注的部分。

Baidu 大脑 | EasyData

我的数据总览 > 螺丝螺母测试/V1/导入

创建信息

数据集名称	螺丝螺母测试	版本号	V1
标注类型	物体检测	标注模板	矩形框标注

标注信息

数据总量	0	已标注	0
标签个数	0	目标数	0
待确认	0	大小	0M

数据清洗

暂未做过数据清洗任务

导入数据

数据标注状态 无标注信息 有标注信息

导入方式

数据集选择

导入该数据集 全部数据(不带标注) 未标注数据

[确认并返回](#)

🔗 导入摄像头采集数据

选择导入摄像头采集数据，跳转至摄像头原始数据页面，选择设备及时间后，点击【保存至数据集】[了解如何从摄像头采集数据](#)

我的数据总览 > 螺丝螺母测试/V1/导入

创建信息

数据集名称	螺丝螺母测试	版本号	V1
标注类型	物体检测	标注模板	矩形框标注

标注信息

数据总量	0	已标注	0
标签个数	0	目标数	0
待确认	0	大小	0M

数据清洗

暂未做过数据清洗任务

导入数据

数据标注状态 无标注信息 有标注信息

导入方式

[从摄像头采集原始数据中导入](#)

EasyData数据服务 <三> 原始数据查看

摄像头原始数据查看

在这里查看摄像头采集的全部数据，按设备、时间筛选数据，并添加到数据集中，以进行下一步的数据清洗、标注等

选择设备 视频抽帧接入: 好设备啊-6... gym_抽帧测... gym_抽帧测... rms-test zdzl测试设备 rsm-rk [展开](#)

图片接入: gym_usb_vi... gym_video londonbridge [清空](#)

选择时间 2020/3/29 - 2020/4/28 [📅](#)

[查看筛选结果](#)

筛选结果

[保存至数据集](#) [批量删除](#)



导入云服务数据回流数据

选择云服务调用数据，可以导入云服务调用的原始数据。

说明：【云服务数据回流数据】是指通过在实际业务系统中集成云服务接口（目前支持EasyDL云服务接口、BML云服务接口），同时开通服务后，可以将接口的图片上传在EasyData平台中来。**注意：**当导入数据标注类型选择了无标注信息，则仅导入原始数据

Baidu 大脑 | EasyData

☰ 我的数据总览 > 螺丝螺母测试/V1/导入

创建信息 ▾

数据集名称	螺丝螺母测试	版本号	V1
标注类型	物体检测	标注模板	矩形框标注

标注信息 ▾

数据总量	0	已标注	0
标签个数	0	目标数	0
待确认	0	大小	0M

数据清洗

暂未做过数据清洗任务

导入数据

数据标注状态 无标注信息 有标注信息

导入方式 从云服务调用获取数据 ▾

注：这里仅导入云服务调用接口的原始图片数据，若希望将接口返回的识别结果作为标签信息一起导入，请切换至【有标注信息】进行导入

从云服务调用获取数据

☞ 导入已标注数据

☞ 导入图像分类已标注数据

从本地导入

图像分类数据从本地导入已标注好的数据，支持整体以一个.zip格式的压缩包上传。**单次上传压缩包整体大小不超过5G**
注意，为了使平台能正确解析标注图片，请按照以下方式整理标注文件再打包上传

☰ 数据管理与标注 > 车厘子和樱桃识别/导入

创建信息

数据ID	65	数据名称	车厘子和樱桃识别
数据类型	图片	创建时间	2019.10.19 00:03:40
标注类型	图像分类	标注模板	单图单标签

标注信息

数据总量	127	已标注	90(进度70.87%)
标签个数	17		

导入数据集

数据标注状态 未标注数据 已标注数据

导入方式 本地 ▾ 上传压缩包 ▾

上传压缩包 上传压缩包

确认并返回

1、图片及对应的标注文件一一对应，是文件名需要完全相同，否则会导入失败。

- 图片格式要求

1. 图片类型为jpg/png/ bmp/jpeg 。
2. 图片大小限制在14M内，长宽比在3：1以内，其中最长边需要小于4096px，最短边需要大于30px。
3. 单个数据集大小限制为10万图片，超出后可被忽略。

- 特别注意：

- 对同一数据集存在多个内容完全一致的图片，将会做去重处理。
- 为保证模型训练效果，所上传的图片应与实际业务场景的图片（光线、角度、采集设备）尽可能一致

Output	File Name	Size
	tt0084058.jpg	24 KB
	tt0084058.json	52 字节
	tt0084867.jpg	45 KB
	tt0084867.json	73 字节
	tt0085121.jpg	35 KB
	tt0085121.json	72 字节
	tt0085154.jpg	29 KB
	tt0085154.json	71 字节
	tt0085159.jpg	32 KB
	tt0085159.json	32 字节
	tt0085208.jpg	36 KB
	tt0085208.json	32 字节

2、json文件内容格式示例如下：

单图单标签

```
{
  "labels":
  [
    {"name": "Drama"}
  ]
}
```

单图多标签

```
{
  "labels":
  [
    {"name": "Drama"},
    {"name": "Romance"}
  ]
}
```

从已有数据集导入已标注数据

从已有数据集导入已标注数据，支持选择数据集及导入的具体标签进行导入。

数据管理与标注 > 车厘子和樱桃识别/导入

标注信息

数据总量	127	已标注	90(进度70.87%)
标签个数	17		

导入数据集

数据标注状态

未标注数据

已标注数据

导入方式

已有数据集

导入线上已有数据集

数据集选择

【单图单标签】图像分类-单图单标签

选择标签

全选

potato

sdf

sdfasdfs

🔗 导入物体检测已标注数据

从本地导入已标注数据

物体检测数据从本地导入已标注好的数据，支持整体以一个.zip格式的压缩包上传。单次上传压缩包整体大小不超过5G
注意，为了使平台能正确解析标注图片，请按照以下方式整理标注文件再打包上传

三> 数据管理与标注 > 车厘子和樱桃识别/导入

I 创建信息

数据ID	65	数据名称	车厘子和樱桃识别
数据类型	图片	创建时间	2019.10.19 00:03:40
标注类型	图像分类	标注模板	单图单标签

I 标注信息

数据总量	127	已标注	90(进度70.87%)
标签个数	17		

I 导入数据集

数据标注状态	<input type="radio"/> 未标注数据 <input checked="" type="radio"/> 已标注数据
导入方式	<input type="text" value="本地"/> <input type="text" value="上传压缩包"/>
上传压缩包	<input type="button" value="上传压缩包"/>
	<input type="button" value="确认并返回"/>

1、图片及对应的标注文件一一对应，是文件名需要完全相同，否则会导入失败。

• 图片格式要求

1. 图片类型为jpg/png/ bmp/jpeg 。
2. 图片大小限制在14M内，长宽比在3：1以内，其中最长边需要小于4096px，最短边需要大于30px。
3. 单个数据集大小限制为10万图片，超出后可被忽略。

• 特别注意：

- 对同一数据集存在多个内容完全一致的图片，将会做去重处理。
- 为保证模型训练效果，所上传的图片应与实际业务场景的图片（光线、角度、采集设备）尽可能一致。
- 单张图片中的标注框须限制在500个之内。如有特殊需求，请通过[提交工单](#)咨询。

Output	
 tt0084058.jpg	24 KB
 tt0084058.json	52 字节
 tt0084867.jpg	45 KB
 tt0084867.json	73 字节
 tt0085121.jpg	35 KB
 tt0085121.json	72 字节
 tt0085154.jpg	29 KB
 tt0085154.json	71 字节
 tt0085159.jpg	32 KB
 tt0085159.json	32 字节
 tt0085208.jpg	36 KB
 tt0085208.json	72 字节

2、json文件内容格式示例如下：

矩形框标注

```
{
  "labels": [
    {
      "name": "car",
      "x1": 278.7673,
      "y1": 38.76027,
      "x2": 328.51887,
      "y2": 101.77642
    },
    {
      "name": "car",
      "x1": 333.0601,
      "y1": 56.9776,
      "x2": 378.78652,
      "y2": 126.21715
    }
  ]
}
```

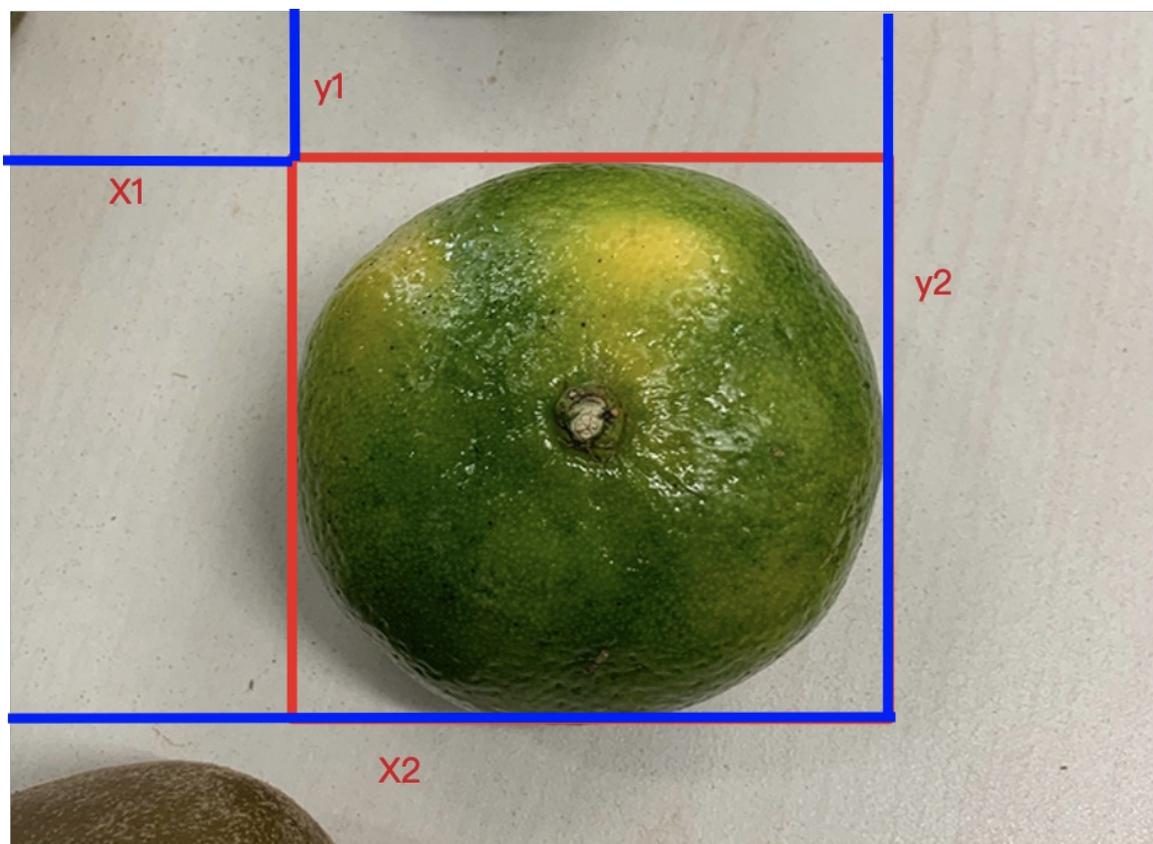
其中x1、y1、x2、y2含义如下

x1：目标定位位置的长方形左上顶点的水平坐标

y1：目标定位位置的长方形左上顶点的垂直坐标

x2：目标定位位置的长方形右下顶点的水平坐标

y2：目标定位位置的长方形右下顶点的垂直坐标



从已有数据集导入

从已有数据集导入已标注数据，支持选择数据集及导入的具体标签进行导入。

I 导入数据集

数据标注状态

未标注数据

已标注数据

导入方式

已有数据集

导入线上已有数据集

数据集选择

【矩形框标注】晓宇_小番茄_奇异果

选择标签

全选

kiwi

tomato

确认并返回

文本数据导入

🔗 导入文本分类无标注数据

进入到新创建的文本分类数据集中，如果您手中的数据是未标注数据，可以在导入数据中，数据标注状态选择“无标注信息”。

☰ 我的数据总览 > myj-文本分类数据集/V1/导入

创建信息 ▾

数据集名称	myj-文本分类数据集	版本号	V1
标注类型	文本分类	标注模板	单文本单标签

标注信息 ▾

数据总量	0	已标注	0
标签个数	0	目标数	0
待确认	0	大小	0M

数据清洗

暂未做过数据清洗任务

导入数据

数据标注状态 无标注信息 有标注信息

导入方式

平台提供两种方式导入数据，一种是从「本地上传」，一种是从「线上已有数据集」

在「本地上传」数据过程中需要注意以下几点：

- 可支持单个txt文本文件上传或将多个文本文件以压缩包的方式统一上传
- 上传过程中存在文本内容完全一样的样本，将会做去重处理
- 文本文件类型为txt，单次上传限制100个文本文件
- 单个文本大小限制在4M以内，文本文件大小限制长度最大4096个UTF-8字符
- 单个数据集大小限制为10万文本文件，超出后会被忽略

「线上已有数据集」导入时，平台将筛选出同为「文本分类」的数据集以供挑选。

单个文本的样例，请见下图：



🔗 导入短文本匹配无标注数据

进入到新创建的文本分类数据集中，如果您手中的数据是未标注数据，可以在导入数据中，数据标注状态选择“无标注信息”。

三> 我的数据总览 > myj-文本匹配数据集/V1/导入

创建信息

数据集名称	myj-文本匹配数据集	版本号	V1
标注类型	短文本匹配	标注模板	短文本匹配

标注信息

数据总量	0	已标注	0
标签个数	0	目标数	0
待确认	0	大小	0M

数据清洗

暂未做过数据清洗任务

导入数据

数据标注状态 无标注信息 有标注信息

导入方式 本地数据集 上传文本

上传文本 上传文本

确认并返回

平台提供两种方式导入数据，一种是从「本地上传」，一种是从「线上已有数据集」

从「本地上传」数据过程中需要注意以下几点：

以文本文件上传时：

1. 支持上传文件类型为txt，编码仅支持UTF-8，单次上传限制100个文本文件，单个数据集大小限制为10万条数据，超出后会被平台忽略
2. 短文本匹配的数据格式要求为“文本内容1\t文本内容2 \n”，一行表示一组数据，每组数据字符数建议不超过512个，可上传多个文本文件

关于文本数据的样例，如下图：

1	喜欢打篮球的男生喜欢什么样的女生	爱打篮球的男生喜欢什么样的女生
2	我手机丢了,我想换个手机	我想买个新手机,求推荐
3	大家觉得她好看吗	大家觉得跑男好看吗?
4	求秋色之空漫画全集	求秋色之空全集漫画
5	晚上睡觉带着耳机听音乐有什么害处吗?	孕妇可以戴耳机听音乐吗?
6	学日语软件手机上的	手机学日语的软件
7	打印机和电脑怎样连接,该如何设置	如何把带无线的电脑连接到打印机上
8	侠盗飞车罪恶都市怎样改车	侠盗飞车罪恶都市怎么改车
9	什么花一年四季都开	什么花一年四季都是开的
10	看图猜一电影名	看图猜电影!
11	这上面写的是什	胃上面是什么
12	建议您重新注册,辛苦您了。	可以的,您注销成功后,可以重新注册的,辛苦您了。
13	小草有什么的特点,可以象征什么?	小草有什么特点可以象征什么
14	校验失败了,	您好,您还是访客状态呢
15	尼玛什么意思	尼玛啊是什么意思?
16	自找苦吃的地方是哪儿?	自找苦吃的地方是哪儿
17	尾号4位多少	尾号是多少后4位
18	谢文东能在哪里看	谢文东在哪里看
19	新概念英语第二册练习册41课答案	新概念英语第二册练习册21练习答案
20	过年送礼送什么好	过年前什么时候送礼?
21	壮丁是什么生肖	欲钱买壮丁指的是什么生肖?
22	这是什么蔬菜啊	这是什么蔬菜呀?
23	最近有什么好电视剧看?	现在看电影都是有哪些号网站啊

以压缩包方式上传时：

1. 压缩包格式为.zip格式，单个压缩包限制5G以内
2. 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\n”，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512个
3. 单个文本文件大小限制在4M内，文本文件大小限制长度最大4096 UTF-8字符

「线上已有数据集」导入时，平台将筛选出同为「短文本匹配」的数据集以供挑选。

🔗 导入文本分类有标注数据

如果您手中的数据为已经标注好的数据，可以选择数据标注状态为“有标注信息”。平台提供两种方式导入数据，一种是从本地上传，一种是从「线上已有数据集」导入

☰ 我的数据总览 > myj-文本分类数据集/V1/导入

I 创建信息 ▾

数据集名称	myj-文本分类数据集	版本号	V1
标注类型	文本分类	标注模板	单文本单标签

I 标注信息 ▾

数据总量	0	已标注	0
标签个数	0	目标数	0
待确认	0	大小	0M

I 数据清洗

暂未做过数据清洗任务

I 导入数据

数据标注状态 无标注信息 有标注信息

导入方式

已标注数据，需要以压缩包的形式进行上传，同时需要参考如下要求：

- 上传已标注文件要求格式为zip格式压缩包，同时压缩包大小在5GB以内
- 压缩包内需要包括单个文本文件（txt）及同名的json格式标注文件
- 文本文件大小限制长度最大4096个UTF-8字符。单个数据集大小限制为10万文本文件量，超出后将被忽略

压缩包的样例如下：

名称
224237.json
224237.txt
224238.json
224238.txt
224239.json
224239.txt
224240.json
224240.txt
224241.json
224241.txt
224242.json
224242.txt
224243.json
224243.txt
224244.json
224244.txt

导入成功后，便可以在已标注数据空间中，查看数据。

🔗 导入短文本匹配有标注数据

如果您手中的数据为已经标注好的数据，可以选择数据标注状态为“有标注信息”。平台提供两种方式导入数据，一种是从本地上传，一种是从「线上已有数据集」导入

在本地上传时，仅支持压缩包的方式上传，您可以根据上传文件的弹窗提示来准备上传的文件，具体详细要求如下：

1. 压缩包格式为.zip格式，单个压缩包限制5G以内
2. 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，每个文本可以有多个短文本组数据，每组数据字符数建议不超过512个
3. 单个文本文件大小限制在4M内，文本文件大小限制长度最大4096 UTF-8字符

导入成功后，便可以在已标注数据空间中，查看数据。

🔗 导入文本实体抽取无标注数据

选择数据类型为“文本”，标注类型选择“文本实体抽取”，标注模板中使用默认选项“文本实体抽取”。

如果您手中的数据是未标注数据，可以选择数据标注状态为“无标注信息”。平台暂不支持上传有标注信息的文本实体抽取数据。

本地上传，包括三种上传方式，分别为「上传TXT文本」、「上传压缩包」、「上传Excel」；

通过文本上传时，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4096个字符（约2000个字），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

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4096个字符（约2000个字），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4096个字符（约2000个字），超出将被截断

文件格式支持xlsx格式，单次可上传100个文件

🔗 文本实体抽取去重策略

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。

音频数据导入

目录

1. [导入未标注数据](#)
2. [导入已标注数据](#)

🔗 导入未标注数据

为选定的图像分类数据集导入未标注数据（即原始音频）的方式有两种。一种为从本地选择音频文件；一种为从线上【EasyData-数据服务概览】中其余同为音频类型数据集中导入。

从本地上传音频导入

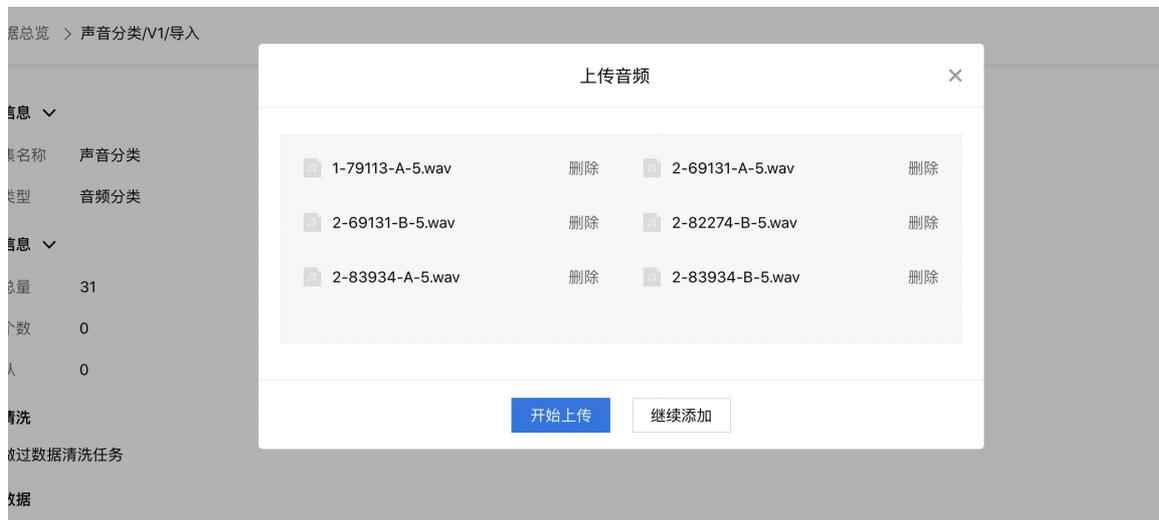
选择未标注数据-本地-上传音频，点击上传音频后选择本地音频文件进行导入。

数据格式要求

- 1、音频类型wav/mp3/m4a，单次上传限制100个文件
- 2、单个音频文件大小限制在4M内，长度限制10s以内
- 3、您的账户下数据集数量限制为10万个音频，如果需要提升数据额度，可在平台提交工单

特别注意：

- 1、对同一数据集存在多个内容完全一致的音频，将会做去重处理。。
- 2、为保证模型训练效果，所上传的音频应与实际业务场景的音频（采集设备等）尽可能一致。



从本地上传压缩包导入

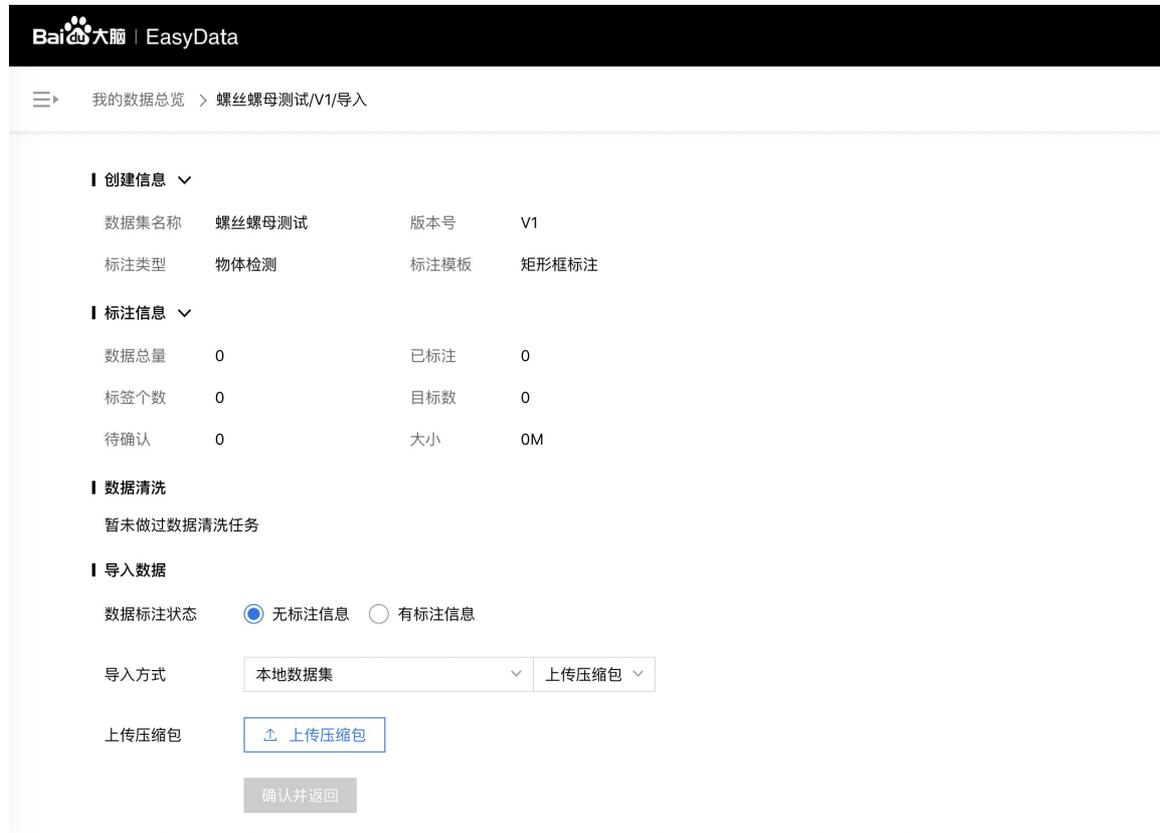
选择**无标注信息-本地-上传压缩包**，点击上传压缩包后选择本地音频压缩包完成数据的批量导入。

数据格式要求

- 1、压缩包仅支持zip格式，大小限制5GB以内
- 2、压缩包内音频类型wav/mp3/m4a，音频大小限制在4M内，长长度限制10s以内
- 3、单个数据集大小限制为10万音频，超出后可被忽略。

特别注意：

- 对同一数据集存在多个内容完全一致的音频，将会做去重处理。
- 为保证模型训练效果，所上传的音频应与实际业务场景的音频（采集设备）尽可能一致



🔗 导入线上已有数据

选择**无标注信息-已有数据集-导入已有数据集**，支持将已经上传在其他Easydata同为音频的数据集进行导入。

Baidu 大脑 | EasyData

☰ 我的数据总览 > 声音分类/V1/导入

创建信息 ▾

数据集名称	声音分类	版本号	V1
标注类型	音频分类	标注模板	短音频单标签

标注信息 ▾

数据总量	31	已标注	0 (进度0.0%)
标签个数	0	目标数	0
待确认	0	大小	13.04M

数据清洗

暂未做过数据清洗任务

导入数据

数据标注状态 无标注信息 有标注信息

导入方式 线上已有数据集 ▾

数据集选择 声音分类 V1 ▾

导入该数据集 全部数据(不带标注) 未标注数据集

确认并返回

🔗 导入已标注数据

🔗 导入音频分类有标注信息数据

从本地导入

音频分类数据从本地导入已标注好的数据，支持整体以一个.zip格式的压缩包上传。**单次上传压缩包整体大小不超过5G**注意，为了使平台能正确解析标注音频，请按照以下方式整理标注文件再打包上传

1. 上传已标注文件要求格式为zip格式压缩包，同时压缩包大小在5GB以内
2. 压缩包内需要包括音频源文件（wav/mp3/m4a）及同名的json格式标注文件，详细请见示例压缩包
3. 单音频文件类型wav/mp3/m4a，单文件大小限制在4M内，长度限制10s以内
4. 您的账户下数据集数量限制为10万个音频，如果需要提升数据额度，可在平台提交工单

● 特别注意：

- 对同一数据集存在多个内容完全一致的音频，将会做去重处理。
- 音频及对应的标注文件一一对应，是文件名需要完全相同，否则会导入失败。
- 为保证模型训练效果，所上传的音频应与实际业务场景的音频（采集设备）尽可能一致

2、json文件内容格式示例如下：

单图单标签

```
{
  "labels":
  [
    {"name": "Drama"}
  ]
}
```

单图多标签

```
{
  "labels":
  [
    {"name": "Drama"},
    {"name": "Romance"}
  ]
}
```

从已有数据集导入已标注数据

从已有数据集导入已标注数据，支持选择数据集及导入的具体标签进行导入。

☰
我的数据总览 > 声音分类/V1/导入

标注信息 ▾

数据总量	31	已标注	0 (进度0.0%)
标签个数	0	目标数	0
待确认	0	大小	13.04M

数据清洗

暂未做过数据清洗任务

导入数据

数据标注状态 无标注信息 有标注信息

导入方式

数据集选择

选择标签

<input type="checkbox"/>	全选
<input type="checkbox"/>	cat
<input type="checkbox"/>	dog
<input type="checkbox"/>	human

确定并导入

视频数据导入

目录

1. [导入未标注数据](#)
2. [导入已标注数据](#)

🔗 导入未标注数据

为选定的视频分类数据集导入未标注数据（即原始视频）的方式有两种。一种为从本地选择视频文件；一种为从线上【EasyData-数据服务概览】中其余同为视频类型数据集中导入。

🔗 从本地上传视频导入

选择**未标注数据-本地-上传视频**，点击上传视频后选择本地视频文件进行导入。

数据格式要求

- 1、视频类型mp4/mov，单次上传限制100个文件
- 2、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10s以内
- 3、您的账户下数据集数量限制为10万个视频，如果需要提升数据额度，可在平台提交工单

特别注意：

- 1、对同一数据集存在多个内容完全一致的视频，将会做去重处理。
- 2、为保证模型训练效果，所上传的视频应与实际业务场景的视频（采集设备等）尽可能一致。

☰ 我的数据总览 > gym_review_video/V1/导入

I 创建信息 ▾

数据集名称	gym_review_video	版本号	V1
标注类型	视频分类	标注模板	短视频单标签

I 标注信息 ▾

数据总量	18	已标注	3 (进度16.7%)
标签个数	2	目标数	0
待确认	0	大小	10.19M

I 数据清洗

暂未做过数据清洗任务

I 导入数据

数据标注状态 无标注信息 有标注信息

导入方式 本地数据集 ▾ 上传视频 ▾

上传视频 📁 上传视频

确认并返回

🔗 从本地上传压缩包导入

选择**无标注信息-本地-上传压缩包**，点击上传压缩包后选择本地视频压缩包完成数据的批量导入。

数据格式要求

- 1、压缩包仅支持zip格式，大小限制5GB以内
- 2、压缩包内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10s以内
- 3、单个数据集大小限制为10万视频，超出后可被忽略。

特别注意：

- 对同一数据集存在多个内容完全一致的视频，将会做去重处理。
- 为保证模型训练效果，所上传的视频应与实际业务场景的视频（采集设备）尽可能一致

我的数据总览 > gym_review_video/V1/导入

创建信息

数据集名称	gym_review_video	版本号	V1
标注类型	视频分类	标注模板	短视频单标签

标注信息

数据总量	18	已标注	3 (进度16.7%)
标签个数	2	目标数	0
待确认	0	大小	10.19M

数据清洗

暂未做过数据清洗任务

导入数据

数据标注状态 无标注信息 有标注信息

导入方式 本地数据集 上传压缩包

上传压缩包 上传压缩包

确认并返回

我的数据总览 > gym_review_video/V1/导入

创建信息

数据集名称	gym_review_video	版本号	V1
标注类型	视频分类	标注模板	短视频单标签

标注信息

数据总量	18	已标注	3 (进度16.7%)
标签个数	2	目标数	0
待确认	0	大小	10.19M

数据清洗

暂未做过数据清洗任务

导入数据

数据标注状态 无标注信息 有标注信息

导入方式 本地数据集 上传压缩包

上传压缩包 上传压缩包

确认并返回

导入线上已有数据

选择无标注信息-已有数据集-导入已有数据集，支持将已经上传在其他Easydata同为视频的数据集进行导入。

我的数据总览 > gym_review_video/V1/导入

标注类型	视频分类	标注模板	短视频单标签
标注信息 ▾			
数据总量	18	已标注	3 (进度16.7%)
标签个数	2	目标数	0
待确认	0	大小	10.19M
数据清洗			
暂未做过数据清洗任务			
导入数据			
数据标注状态	<input checked="" type="radio"/> 无标注信息 <input type="radio"/> 有标注信息		
导入方式	线上已有数据集 ▾		
数据集选择	视频导入-分类 V6 ▾		
导入该数据集	<input checked="" type="radio"/> 全部数据(不带标注) <input type="radio"/> 未标注数据		
确认并返回			

🔗 导入已标注数据

🔗 导入视频分类有标注信息数据

从本地导入

视频分类数据从本地导入已标注好的数据，支持整体以一个.zip格式的压缩包上传。单次上传压缩包整体大小不超过5G
注意，为了使平台能正确解析标注视频，请按照以下方式整理标注文件再打包上传

1. 上传已标注文件要求格式为zip格式压缩包，同时压缩包大小在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，详细请见示例压缩包
3. 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
4. 您的账户下数据集数量限制为10万个视频，如果需要提升数据额度，可在平台提交工单

● 特别注意：

- 对同一数据集存在多个内容完全一致的视频，将会做去重处理。
- 视频及对应的标注文件一一对应，是文件名需要完全相同，否则会导入失败。
- 为保证模型训练效果，所上传的视频应与实际业务场景的视频（采集设备）尽可能一致

2、json文件内容格式示例如下：

```
{
  "labels":
  [
    {"name": "Drama"}
  ]
}
```

单图多标签

```
{
  "labels":
  [
    {"name": "Drama"},
    {"name": "Romance"}
  ]
}
```

从已有数据集导入已标注数据

从已有数据集导入已标注数据，支持选择数据集及导入的具体标签进行导入。

我的数据总览 > gym_review_video/V1/导入

创建信息 ▾

数据集名称	gym_review_video	版本号	V1
标注类型	视频分类	标注模板	短视频单标签

标注信息 ▾

数据总量	18	已标注	3 (进度16.7%)
标签个数	2	目标数	0
待确认	0	大小	10.19M

数据清洗

暂未做过数据清洗任务

导入数据

数据标注状态 无标注信息 有标注信息

导入方式 本地数据集 ▾ 上传压缩包 ▾

上传压缩包 📁 上传压缩包

确认并返回

数据导出

🔗 数据导出入口

数据内容不为空的数据集版本支持数据导出，可以在相应的数据集列表找到数据导入的功能入口。

EasyData数据服务 <三 总览

创建数据集

版本	数据量	最近导入状态	标注类型	标注状态	发布状态	操作
V1	18	已完成	视频分类	16% (3/18)	未发布	查看 导入 导出 标注 清洗 发布 去训练

版本	数据量	最近导入状态	标注类型	标注状态	发布状态	操作
V1	31	已完成	音频分类	0% (0/31)	未发布	查看 导入 导出 标注 清洗 发布 去训练

版本	数据量	最近导入状态	标注类型	标注状态	发布状态	操作

我的数据总览 > 螺丝螺母测试/V1/导出

创建信息

数据集名称	螺丝螺母测试	版本号	V1
标注类型	物体检测	标注模板	矩形框标注

标注信息

数据总量	19	已标注	0 (进度0.0%)
标签个数	2	目标数	0
待确认	0	大小	12.57M

数据清洗

暂未做过数据清洗任务

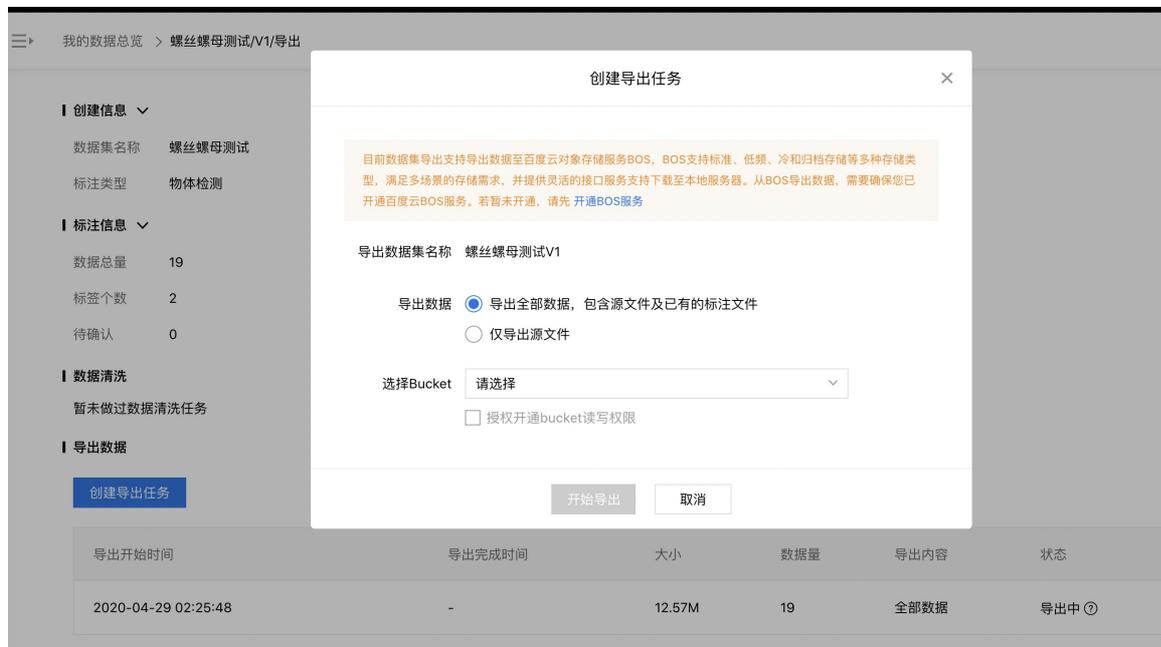
导出数据

创建导出任务

创建数据导出任务

目前数据集导出支持导出数据至百度云对象存储服务BOS，BOS支持标准、低频、冷和归档存储等多种存储类型，满足多场景的存储需求，并提供灵活的接口服务支持下载至本地服务器。从BOS导出数据，需要确保您已开通百度云BOS服务。若暂未开通，请先[开通BOS](#)

根据页面提示，依次选择导出数据内容并选择bucket、勾选授权服务后即可启动数据导出。



数据导出状态

提交数据导出后，可以通过页面数据导出列表查看导出状态。导入完成后可以前往BOS查看导出结果。**注意**：如果遇到导出失败可能由网络不稳定等原因造成，建议稍后重试。若反复失败，请[提交工单](#)。



数据导出结果

- 若选择导出全部数据，则可以在目标bucket看到数据源文件及标注json文件
- 若只选择导出源文件，则可以在目标bucket看到原始数据文件

表格数据导入

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

创建数据集

数据集需要先定义，然后再导入数据。

单击“数据总览”，进入数据集列表页面。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。

单击“数据总览”，进入数据集列表页面。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。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。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。单击“确认并返回”完成导入操作。

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

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) + 导入图片 **质检报告** 批量标注示例

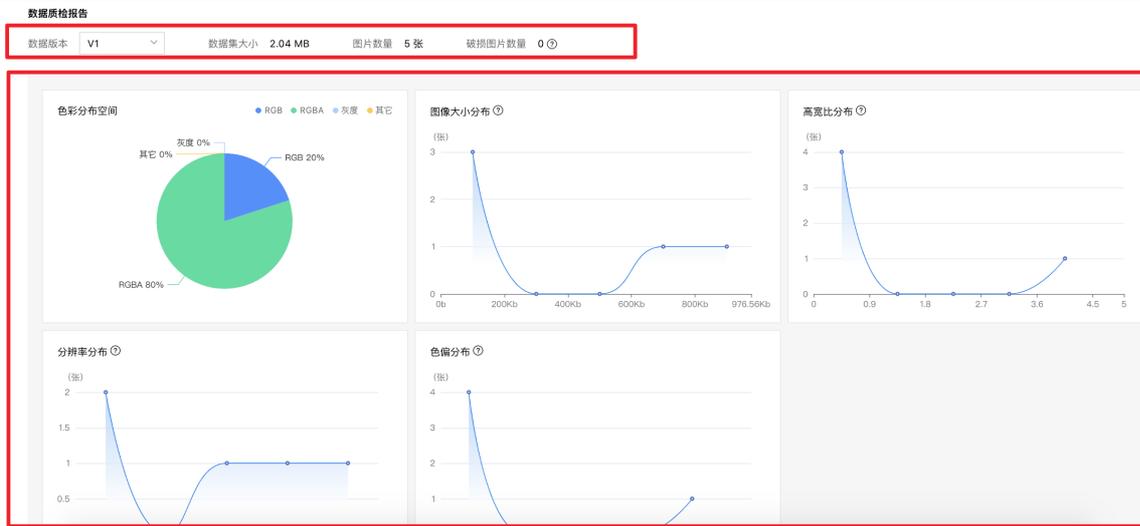
的V1版本的图片列表 筛选

本页全选 删除



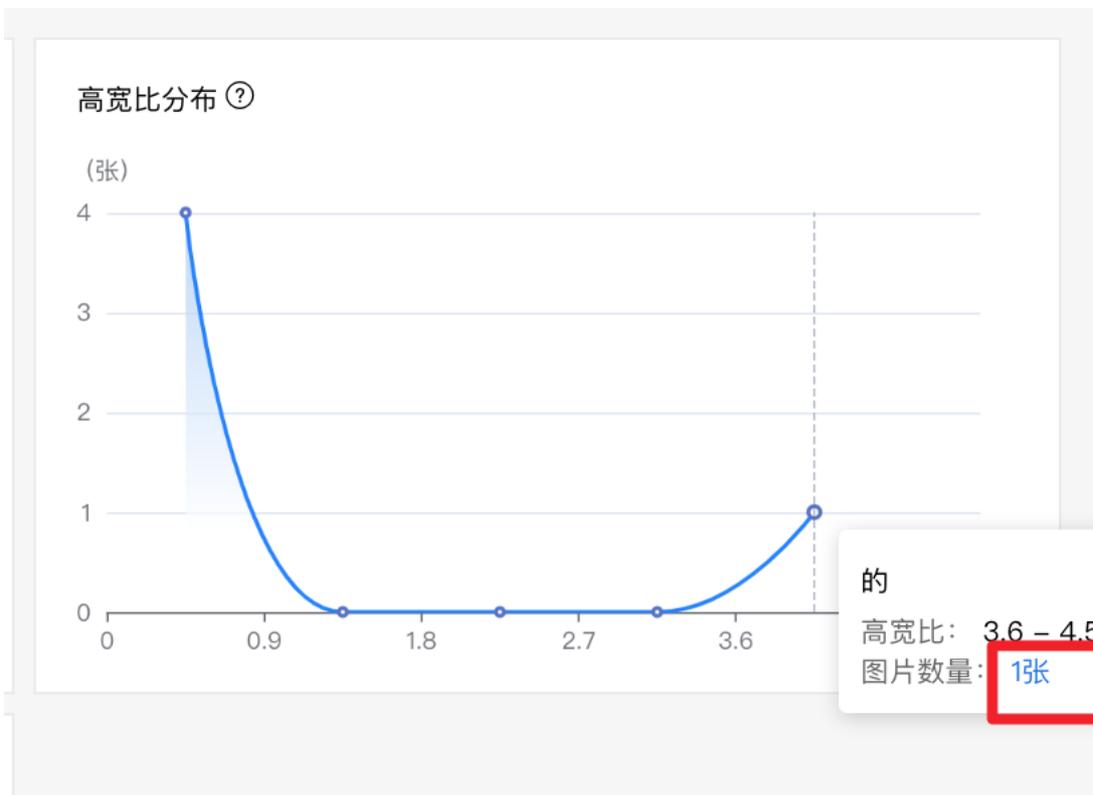
2, 指标查看 本期报告分为整体指标和分布指标两类。整体指标包括数据集存储大小、图片数量、破损图像数三类；分布指标包括色彩分布空间、图像存储大小分布、高宽比分布、分辨率分布、色偏分布五类。

可以通过切换数据集版本查看不同版本下质检报告。



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

数据采集

数据采集功能简介

数据采集是数据处理和服务的必经步骤。实际应用中，数据采集的方式非常多样，往往与最终的数据应用方式密切相关。

为了提高数据采集及后续数据处理和服务的效率，EasyData数据采集目前已支持以下两种图片采集方式，如果您有更多需求，也欢迎[提交工单](#)联系我们：

- 接入摄像头采集图片

提供本地采集软件，支持手动截图、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EasyData管理

- 云服务调用获取数据

AI模型训练后可能需要持续迭代和优化模型效果，通过调用云服务接口（支持EasyDL、BML云服务接口）并在EasyData开通相关服务，可以将实际业务数据及识别结果可视化查看，针对性的选择高质量数据用于模型迭代

接入摄像头采集图片

采集摄像头管理

本地软件安装

简介

软件功能

在本地添加、管理用于数据采集的摄像头。支持两种摄像头数据接入方式：

- 视频抽帧接入

适用采集的数据为视频流、需要进一步抽帧为图片的场景

- 图片接入

适用采集的数据为图片、可以通过FTP协议传输到本地目录的场景

运行环境

网络要求

- 软件的工作需要连接网络至Internet

系统要求

- windows x86 64位架构的PC
 - 操作系统：Windows7 Windows10
- linux x86 64位架构
 - 操作系统：桌面linux系统，例如Ubuntu 16.04，Ubuntu 18.04，CentOS等
- linux arm 64位架构的开发版，例如rk3399
 - 操作系统：rk开发版官方Ubuntu
- 树莓派3B+，树莓派4
 - 操作系统：树莓派Raspbian操作系统
- Macbook，iMac系列

- 操作系统：Mac OS

如何运行软件

第一步：

在[官网](#)下载本地软件，解压对应平台的tar包或zip包（如Capture_Tool_Linux_XXX.tar.gz），放于任意目录

第二步：

进入目录（须进入目录执行./EasyData_Capture_Tool），命令行运行目录中的可执行文件EasyData_Capture_Tool

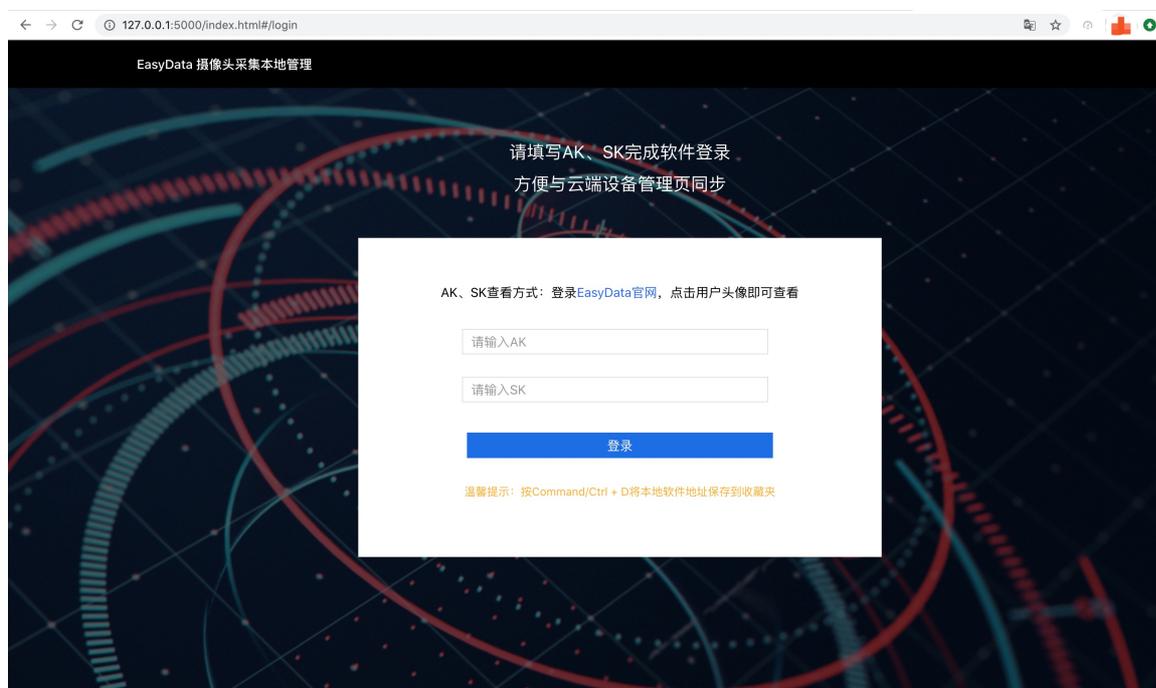
第三步：

在浏览器中输入机器的IP地址加端口号以及首页地址index.html，默认端口号5000打开软件

如果是本机操作，IP地址为127.0.0.1，在浏览器中输入<http://127.0.0.1:5000/index.html>即可

运行成功：

如果能跳转至登录界面说明软件运行成功，否则请确认您的防火墙等设施是否阻止了访问。如果还有异常请[提交工单](#)寻求百度EasyData开发人员支持



如何登录软件

成功运行软件后，需要使用账号唯一的AK、SK登录软件。

AK、SK查看方式：登录[EasyData官网](#)，点击用户头像即可查看：



使用流程

请参考[采集摄像头本地管理](#)文档

软件配置说明

软件包中的easy_data_config.py为软件的配置文件，请在有必要的时候再配置，且确保配置文件符合python语法

- EASY_DATA_ENABLE_VIRTUAL - 如果需要以FTP的方式接入图片数据需要为True，否则可以设置为False，默认为True
- EASY_DATA_LOG_DEBUG - 设置为True会在控制台打印更多信息，便于调试问题但会增大log文件大小。默认为True
- EASY_DATA_SERVER_PORT - 服务器的http端口号，默认为5000，可以修改为自己需要的端口号
- EASY_DATA_FFmpeg_LOCATION - 项目依赖的ffmpeg库地址，默认在软件包中third_part/ffmpeg目录

特别说明

- 首次运行后在软件目录中会产生一些子目录，对这些目录下文件或目录的操作可能会导致软件异常，请避免操作这些目录和文件
- 软件目录中的log目录内是一些log信息，当软件运行异常时可以将目录中当日的log文件提交给百度EasyData的开发者定位问题
- FTP服务方式验证用户名称为匿名用户anonymous，密码为空。同时也可以使用用户easydata，密码easydata。

更新日志

更新日期	版本号	更新内容
2020.04.29	0.1.0	初版上线，支持linux系统，x86以及部分arm
2020.05.20	0.2.0	支持windows系统，支持系统自带摄像头以及外接的USB摄像头
2020.9.3	0.3.1	上线Mac OS，树莓派3，4支持，并新增rtmp视频源支持

摄像头本地管理

简介

登录摄像头管理本地软件，即可在本地添加、管理用于数据采集的摄像头。

目前摄像头的接入支持以下两种方式，您可以根据页面提示或参考文档进行设备的添加和管理。

- 视频抽帧接入

适用采集的数据为视频流、需要进一步抽帧为图片的场景

- 图片接入

适用采集的数据为图片、可以通过FTP协议传输到本地目录的场景

如果使用中遇到任何问题，欢迎[提交工单](#)联系我们

EasyData摄像头采集本地管理 本地管理 云端管理 使用文档 工单支持

摄像头采集本地管理 点击收起

设备采集的数据可通过「视频抽帧接入」、「图片接入」两种方式同步到云端。
每个账号默认可添加5个设备，向云端传输总计10万张图片，原始图片将在云端保留60天，请及时在云端将图片保存到数据集中。如有更多需求，请提交工单或加入官方QQ群 (xxx) 联系我们。

视频抽帧接入流程：

- 1.添加设备
- 2.填写摄像头路径/视频流地址
- 3.设置抽帧策略
- 4.抽帧产生的图片通过网络传输到云端

图片接入流程：

- 1.添加设备
- 2.生成接入的本地目录
- 3.将数据通过FTP协议传输到本地目录
- 4.本地目录里的图片通过网络传输到云端

添加设备

视频抽帧接入 图片接入

设备名称	摄像头路径/视频流地址	备注	本地状态	联网传输状态	抽帧启停控制	操作
摄像头1号	collections/8883624/ai-%2B...	广渠路	正常	正常	ON	查看云端数据
摄像头2号	collections/8883624/ai-%2B...	广渠路	未连接	传输中止	OFF	抽帧设置 编辑 删除 查看云端数据

1 2 3 4 5 6 7 >

视频抽帧接入

添加设备

点击页面上的「添加设备」按钮，选择「视频抽帧接入」

×

添加设备

选择数据接入方式：

视频抽帧接入
 适用采集的数据为视频流、需要进一步抽帧为图片的场景

图片接入
 适用采集的数据为图片、可以通过FTP协议传输到本地目录的场景。

开始

取消

第一步：填写基本信息

- 摄像头路径/视频流地址
- 设备名称
- 备注

视频抽帧接入 ✕

1 基本信息 2 校验设备 3 抽帧设置

请选择 摄像头路径 视频流地址

填写路径

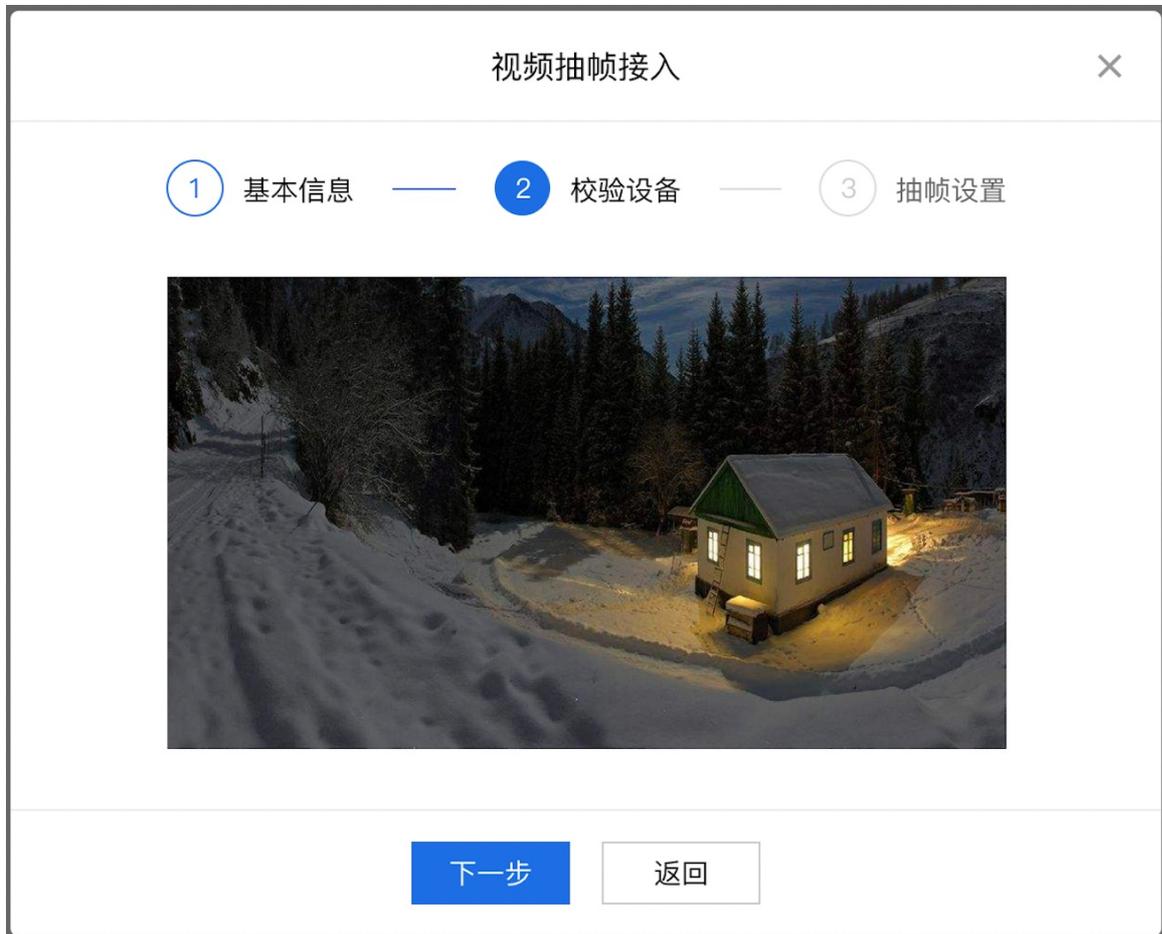
设备名称

备注

下一步 取消

第二步：校验设备

- 摄像头连接正常时，即可看到当前拍摄的图像



抽帧设置

第一步：点击抽帧设置，完成自动抽帧策略的设定

[视频抽帧接入](#)

设备名称	摄像头路径/视频流地址	备注	本地状态	联网传输状态	抽帧启停控制	操作
我的测试	FaceTime HD Camera	我的测试	正常	正常	<input type="checkbox"/> OFF	手动截屏 抽帧设置 编辑 删除 查看云端数据

- 支持以秒、分钟、小时为单位进行抽帧
- 支持添加多个抽帧时间段



第二步：如存在手动截图需求，点击手动截图，完成截图的选取



选取完

毕的截图，可点击查看已有截图进行快速浏览与后续保存等处理

摄像头原始数据查看

在这里查看摄像头采集的全部数据，按设备、时间筛选数据，并添加到数据集中，以进行下一步的数据清洗、标注等

选择设备 视频抽帧接入： 我的测试 [清空](#)

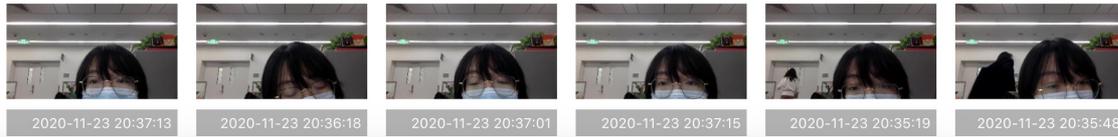
选择时间

[查看筛选结果](#)

筛选结果

[保存至数据集](#)

[批量删除](#)



查看设备状态

在设备列表中，可以查看设备的本地状态、联网传输状态，并进行抽帧的启停控制

- 本地状态：包括摄像头是否正常连接、抽帧软件是否正常运行
- 联网传输状态：抽帧软件是否联网，数据是否正常传输到云端
- 抽帧启停控制：与状态无关，用于控制是否进行抽帧

设备名称	摄像头路径/视频流地址	备注	本地状态	联网传输状态	抽帧启停控制	操作
摄像头1号	collections/8883624/ai-%2B	广渠路	正常	正常	<input checked="" type="checkbox"/>	查看云端数据
摄像头2号	collections/8883624/ai-%2B...	广渠路	未连接	传输中止	<input type="checkbox"/>	抽帧设置 编辑 删除 查看云端数据

设备管理

在设备列表的操作栏，可以进行设备的管理：

- 修改抽帧设置：调整抽帧频率或抽帧时间（注意：抽帧启停控制为ON时，不支持修改抽帧设置，需要先停止抽帧再操作）
- 编辑设备信息：修改设备名、备注、摄像头路径/视频流地址
- 删除设备
- 查看云端数据：在EasyData在线平台查看当前设备采集的数据

图片接入

添加设备

点击页面上的「添加设备」按钮，选择「图片接入」

第一步：填写基本信息：

- 采集单张图片的时间间隔：即图片拍摄的时间间隔
- 设备名称
- 备注

图片接入
✕

采集单张图片
时间间隔 秒 ▼ /图片

设备名称

备注

生成本地目录
取消

第二步：获得本地目录，并将摄像头采集的图片通过[FTP协议](#)传输到目录，FTP的用户名为匿名用户：anonymous

生成本地目录
✕

接入的本地目录已生成，请将数据通过[FTP协议](#)传输到以下目录

XXXXXXXXXXXXXXXXXXXXX 复制

确定

查看设备状态

在设备列表中，可以查看设备的本地状态、联网传输状态

- 本地状态：图片是否以指定的时间间隔传输到本地目录
- 联网传输状态：抽帧软件是否联网，数据是否正常传输到云端

设备名称	本地目录	备注	本地状态	联网传输状态	操作
摄像头1号	xxxxxxx	广渠路	正常	正常	编辑 删除 查看云端数据
摄像头2号 🔗	xxxxxxx 🔗	广渠路 🔗	异常 🕒	正常	编辑 删除 查看云端数据

设备管理

在设备列表的操作栏，可以进行设备的管理：

- 编辑设备信息：修改设备名、备注、采集单张图片时间间隔
- 删除设备
- 查看云端数据：在EasyData在线平台查看当前设备采集的数据

🔗 摄像头云端管理

简介

在EasyData[摄像头管理页面](#)下载本地软件，成功完成[安装](#)并[添加设备](#)后，即可在EasyData在线平台看到本地的设备（同步展示设备名称、备注等信息）及对应的采集数据。

查看设备状态

云端的设备状态，表示本地软件是否联网在线。当状态为离线时，采集的数据无法传输到EasyData平台。

查看采集数据

每个设备的右侧，都有「查看数据」的操作，点击即可查看该设备采集的数据。

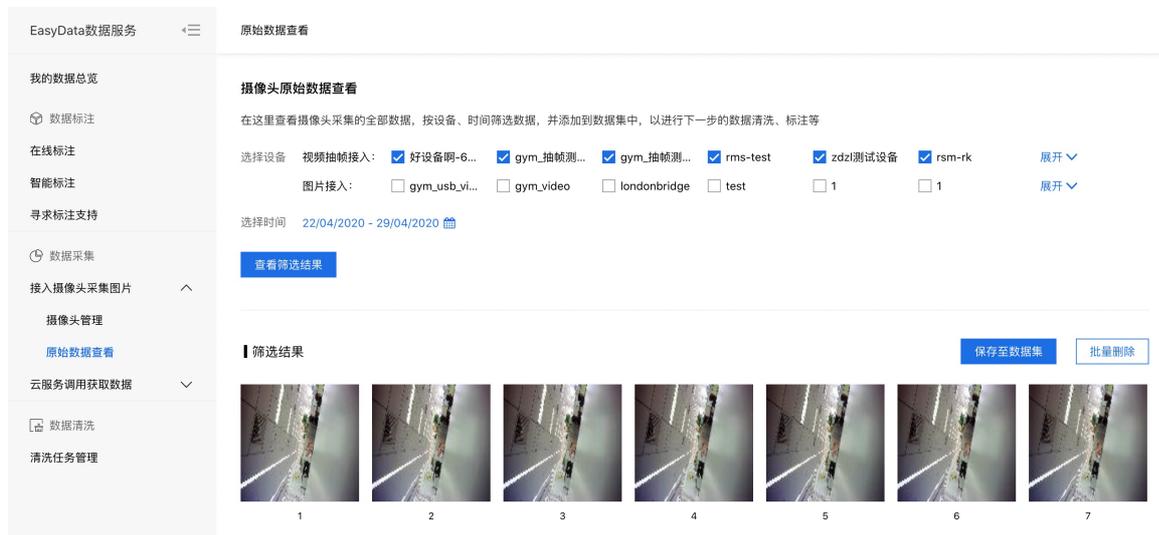
[视频抽帧接入](#) [图片接入](#)

设备ID	设备名称	备注	状态	操作
bb56bb7a-89f1-11ea-8027-38f9d3cca13f	asdasd	asd	离线 ⓘ	查看数据
397aafce-89ee-11ea-9cbd-9cd21e1891fb	test	test	离线 ⓘ	查看数据
dbcdb38a-8916-11ea-88eb-38f9d3cca13f	测试打包	asdsadasd	离线 ⓘ	查看数据
257e2094-890a-11ea-80d4-38f9d3cca13f	测试打包	asd	离线 ⓘ	查看数据
70f4d1ca-887a-11ea-8578-4cedfb96cab1	新-摄像头	猪场测试	在线	查看数据
cea5d8da-885e-11ea-b4ce-9cd21e1891fb	gym_usb_test	usb_for_pm	离线 ⓘ	查看数据

原始数据查看

🔗 简介

摄像头采集[原始数据查看](#)页面，支持查看所有摄像头采集的数据。通过设置筛选条件，也可以同时查看单个/多个摄像头采集的数据。



查看方法

- 选择设备：选择数据来源的摄像头
- 选择时间：选择数据采集的时间段
- 点击「查看筛选结果」

保存数据

原始数据查看页面中的数据均为临时存储的数据，请及时操作将图片保存到数据集中：

- 目前单个账号支持存储的图片上限数为10万张
- 原始图片自采集日起，将在云端保留60天

如有更多需求，请[提交工单](#)联系我们

点击图片查看区右上角的「保存至数据集」，即可将当前查看的图片批量保存：



云服务数据回流

配置数据授权

简介

在EasyData授权管理已发布云服务的调用数据，查找公有云服务模型API识别错误的的数据，保存到数据集进行清洗、标注后，用于下一次模型优化训练。

只需在[EasyDL](#)、[BML](#)训练模型并发布公有云API，即可使用本服务。

EasyDL、BML云服务调用数据管理服务

目前EasyDL、BML已全面支持图像类模型的云服务调用数据管理，详情参考：

[EasyDL图像分类](#)、[EasyDL物体检测](#)、[EasyDL图像分割](#)、[BML视觉任务](#)

添加云服务数据授权

第一步，在使用服务之前，需要选择公有云API、同意数据存储协议，完成授权



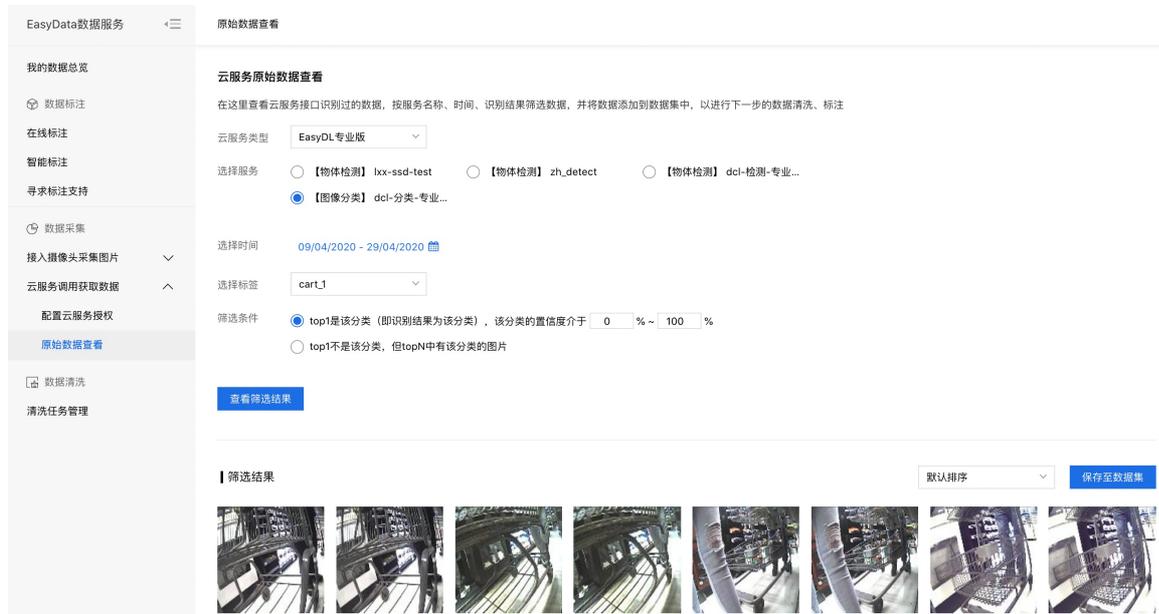
第二步，添加完成后，即可在列表中看到对应的公有云服务，并查看相关数据

云服务名称	云服务类型	ID	模型类型	接口地址	操作
老市民的图像分类	EASYDL经典版	42	图像分类	laoshimindetuxiangfenlei	查看数据 取消授权
钢铁侠	EASYDL经典版	45	图像分类	gangtiexia	查看数据 取消授权
flower_5_shupeng_6f	EASYDL经典版	35	图像分类	flower_5_shupeng_6	查看数据 取消授权
flower_5_shupeng_6f	EASYDL经典版	35	图像分类	flower_5_shupeng_6	查看数据 取消授权
物体检测-1h	EASYDL经典版	41	图像分类	object_1h	查看数据 取消授权

原始数据查看

简介

云服务调用[原始数据查看](#)页面，支持查看所有已授权的云服务调用数据（调用公有云API识别的数据）。通过设置筛选条件，也可以更有针对性地查看API识别错误、识别较差的数据。



查看方法

- 选择服务：选择已授权的公有云服务
- 选择时间：选择调用服务的时间段
- 选择标签并设置筛选条件：不同模型类型有不同的筛选条件，可以多次调整、查看，锁定公有云API识别错误、识别较差的数据范围
- 点击「查看筛选结果」

保存数据

原始数据查看页面中的数据均为临时存储的数据（仅保留30天），请及时操作将图片保存到数据集中。

点击图片查看区右上角的「保存至数据集」，即可将当前筛选出的图片批量保存：



数据清洗

清洗任务管理

在这里对数据集中的图片进行去模糊、去近似、旋转、镜像等多种清洗，以及过滤无人脸图片、过滤无人体图片等高级清洗服务，提升数据质量，方便进行下一步的数据标注等操作。

新建清洗任务

1. 选择数据集版本

选择清洗前后数据集名称及版本时，该数据集版本如果正被用于导入、智能标注、训练、清洗本身、label集图片的删除与更新则不可被选定。

为保证清洗任务顺利进行：当您的清洗前后数据集版本为两个数据集时，两个数据集版本均会为您保留；请确保清洗前数据版本中图片个数少于50000张，如您有大规模数据清洗需求，建议通过拆分数数据集完成。

I 请选择数据集

提交任务后，新添加到清洗前数据集的数据不会被清洗

清洗前	请选择数据集	版本
清洗后	请选择数据集	版本

2. 选择清洗方式

选择清洗方式时，有通用与高级清洗方案两种：如您选择通用清洗方案，同一数据集版本最多可添加三种通用清洗方式；如您选择高级清洗方案，同一数据集版本最多可添加一种高级清洗方式；通用与高级方案不可同时选择。

清洗任务管理 > 新建清洗任务

I 请选择清洗方式

通用清洗方案

最多可添加3种清洗方式，裁剪、旋转、镜像仅支持无标注信息数据

<input type="checkbox"/> 去近似	▼
<input type="checkbox"/> 去模糊	▼
<input type="checkbox"/> 裁剪	▼
<input type="checkbox"/> 旋转	▼
<input type="checkbox"/> 镜像	▼

高级清洗方案

当您的清洗前数据集版本暂无标注类型，选择检测服务则默认清洗后数据集版本类型为检测，选择分割服务则默认清洗后数据集版本标注类型为分割

<input type="radio"/> 过滤无人脸图片	调用百度人脸检测服务完成过滤 服务已开通，免费试用中，当日不限调用总量，QPS 2，调用中APP: 我的人脸应用 ID:19633905	申请配额升级
<input type="radio"/> 过滤无人体图片		

3. 具体清洗方式介绍如下：

● 去近似

根据示例中参考图片，通过拖拽圆点选取对应相似度取值，只会保存相似度高于此值的图片

去近似

两图相似度大于等于此值保留一张: 0.57 0  1

相似度参考



原图



相似度0.75



相似度0.50



相似度0.25



相似度0.08

保存

取消

去模糊

根据示例中参考图片，通过拖拽圆点选取对应清晰度取值，只会保存清晰度高于此值的图片

保留清晰度大于等于此值的图片: 4959 0  10000

清晰度参考



清晰度5



清晰度50



清晰度100



清晰度500



清晰度5000

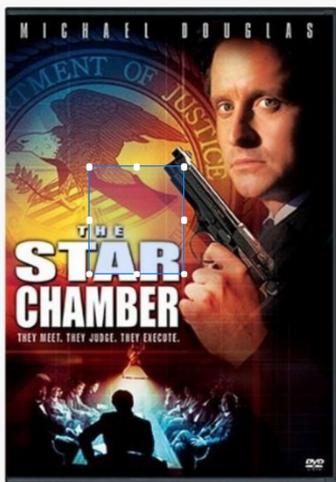
保存

取消

裁剪

在随机图片中通过拖拽裁剪框边缘各点调节框大小，最终针对所有图片只保留框选出区域

下图为gym_wtjcV1中随机抽取的一张图片，裁剪后所有图片均只保留下图框选出的区域，请调节框的大小

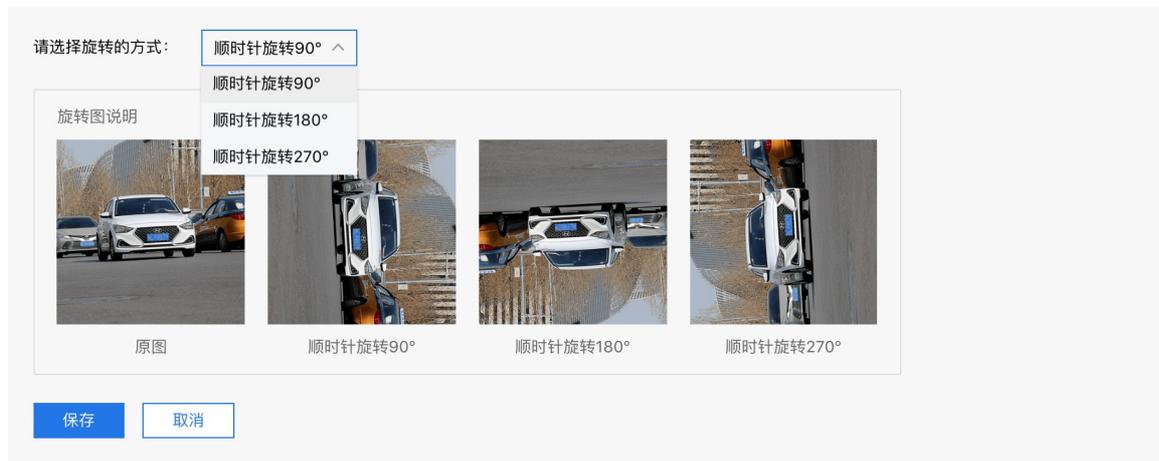


保存

取消

旋转

根据示例中参考图片，通过下拉框选择不同的顺时针旋转角度，最终针对所有图片做旋转操作



- 镜像

水平镜像是指将图像的左右部分以图像垂直中轴线为中心进行镜像对换；垂直镜像是指将图像的上下两部分以图像水平中轴线为中心进行镜像对换；中心镜像是指将图像以图像水平中轴线和垂直中轴线的交点为中心进行镜像对换，相当于将图像先后进行水平镜像、垂直镜像。

根据示例中参考图片，通过下拉框选择不同的镜像方式，最终针对所有图片做对应镜像操作



- 过滤无人脸图片

该服务帮助您从清洗前数据集版本中过滤出无人脸图片，过滤机制为通过调用百度AI开放平台提供的人脸检测服务完成高级清洗。

当您的清洗前数据集版本暂无标注类型时，选择人脸检测服务则默认清洗后数据集版本类型为检测；当您的清洗前数据集版本有标注类型时，需保证该标注类型为物体检测。

使用清洗服务的前置条件需您在百度AI开放平台开通该服务，创建您的APP ID，根据业务需求实现服务调用，具体步骤分为两步：

第一步，<https://ai.baidu.com/tech/face/detect> 去人脸检测服务落地页点击【立即使用】，完成服务开通、创建APPID后，您可以免费试用该服务，人脸检测服务当前的免费机制为当日不限调用总量，QPS限制为2；

第二步，如您希望配额升级，点击【申请配额升级】完成付费购买。



使用清洗服务时，需注意保证您的单图中人脸个数不要超过10张；您可以选择将人脸画框位置信息是否保存至清洗后数据集版本，若选择，则在您的清洗后数据集版本中新建一个为face的标签。



• 过滤无人体图片

该服务帮助您从清洗前数据集版本中过滤出无人体图片，过滤机制为通过调用百度AI开放平台提供的人体检测或人像分割服务完成高级清洗。

建议：当前过滤无人体图片服务的解决场景主要聚焦在视频监控领域，因此，针对实拍场景数据的清洗效果明显优于网图。

当您的清洗前数据集版本暂无标注类型时，选择检测服务则默认清洗后数据集版本标注类型为检测，选择分割服务则默认清洗后数据集版本标注类型为分割；当您的清洗前数据集版本标注模板为图像分割时，默认为您选择图像分割服务，标注模板为非图像分割时，默认为您选择人体检测服务。

使用清洗服务的前置条件需您在百度AI开放平台开通该服务，创建您的APP ID，根据业务需求实现服务调用，具体步骤分为两步：

第一步，去往对应服务落地页点击【立即使用】，完成服务开通、创建APPID后，您可以免费试用该服务，服务当前的免费机制为当日免费调用50000次，QPS限制为2；

第二步，如您希望配额升级，点击【申请配额升级】完成付费购买。

您可以选择将人脸画框位置信息是否保存至清洗后数据集版本，若选择，则在您的清洗后数据集版本中新建一个为body的标签。

🔗 管理清洗任务

1. 任务概览

通过列表页展示任务如下字段：任务序号、清洗方式、清洗前后数据集版本、清洗状态、操作

任务序号	清洗方式	清洗前数据集	清洗后数据集	清洗状态	操作
54	去近似	gym_review_picture-V1	gym_review_picture-V1	清洗完成	查看任务详情
53	去近似	gym_review_picture-V1	gym_review_picture-V1	清洗完成	查看任务详情
52	去近似	gym_review_picture-V2	gym_review_picture-V2	清洗完成	查看任务详情
51	去近似、去模糊	gym_review_picture-V2	gym_review_picture-V3	清洗完成	查看任务详情

2. 查看详情

点击查看任务详情可看到任务开始与完成时间、提交数据量、清洗方式、清洗结果（点击可跳转至清洗后数据集版本详情页）



3. 其他情况

当您选择了清洗服务后，会存在清洗结果为零的情况（例如，过滤无人脸服务完成过滤后发现清洗后数据集版本中有效数据为零），在这种情况下，您将无法查看清洗结果。

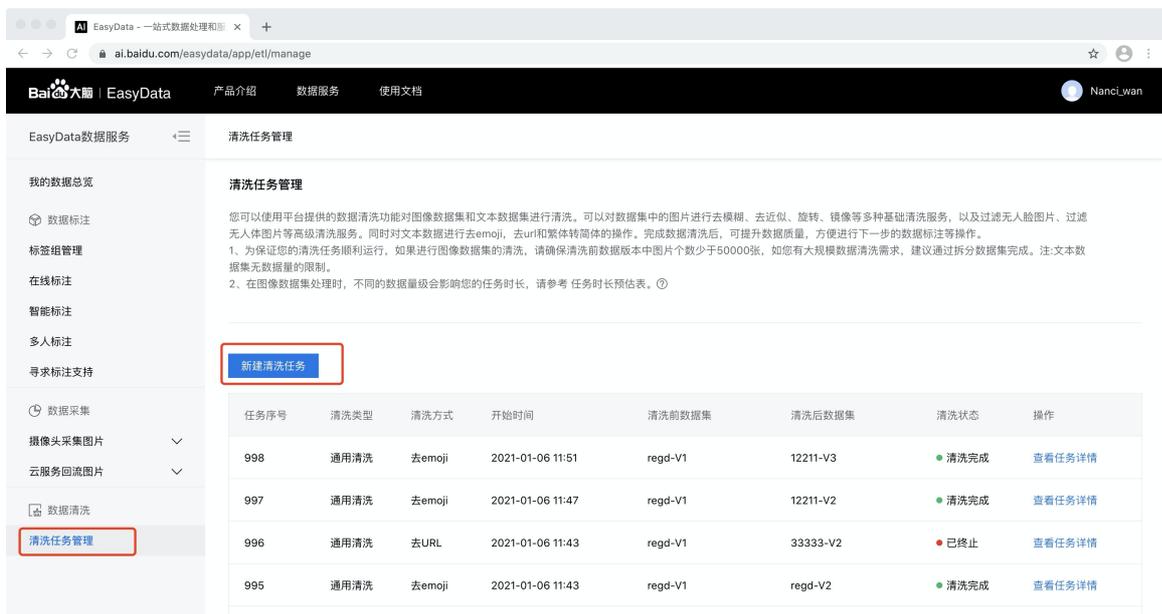
当您选择了高级清洗服务后，如您的服务欠费导致清洗服务终止，这是在任务详情页将显示服务欠费，您可以前往对应服务落地页完成续费复用后在清洗任务列表页选择【继续清洗】。

文本数据清洗

您可以使用平台提供的数据清洗功能对文本数据集进行清洗。当您的文本数据中存在emoji(表情等符号)，大量的url时，您可以使用该数据清洗的功能去除。您也可以试用该功能将文本数据中的繁体字转为简体字。需要注意的是，文本数据清洗功能仅支持未标注数据的清洗。您可以参考以下步骤完成文本数据的清洗。

1. 创建清洗任务

点击左侧导航栏中的【清洗任务管理】，可进入清洗任务管理页面，点击【新建清洗任务】即可跳转到创建页面，清洗方式选择【文本数据清洗】。



2、选择数据集版本

为保证清洗任务顺利进行：当您的清洗前后数据集版本为两个数据集时，两个数据集版本均会为您保留；在清洗前选择数据集及版本号，清洗后数据集的名称您可以选择和清洗前保持一致，新增该数据集版本号，也可以创建新的数据集名称，版本号为V1。在选择数据集版本时您需要注意以下几点：

- 清洗前的数据集需为非空且未标注的数据集；
- 目前本系统不支持清洗前后的数据集版本相同；您可以新建清洗后的数据集版本；
- 清洗后数据集版本标注信息需与清洗前保持一致；
- 清洗后数据集的名称可以和清洗前不同

Baidu 百度智能云 | EasyData 产品介绍 数据服务 使用文档

清洗任务管理 > 新建清洗任务

清洗方式： 图像数据清洗 文本数据清洗

请选择数据集
提交任务后，新添加到清洗前数据集的数据不会被清洗

清洗前	请选择数据集	版本
清洗后	请选择数据集	版本

请选择清洗方式
 通用清洗方案
最多可添加3种清洗方式

- 去除emoji
- 去除URL
- 繁体转简体

提交 返回

3、选择数据集清洗方式

文本数据的清洗方式目前仅支持通用清洗方案，共有三种，根据您清洗文本数据的需要，您可同时选择1到多种清洗方式(最多可添加3种清洗方式)，以下分别对三种清洗方式进行具体介绍。

- 去除emoji：是指去掉清洗前文本中的表情等符号

Baidu 百度智能云 | EasyData 产品介绍 数据服务 使用文档

清洗任务管理 > 新建清洗任务

清洗前 请选择数据集 版本

清洗后 请选择数据集 版本

请选择清洗方式

通用清洗方案

最多可添加3种清洗方式

去除emoji

源文本 价廉物美的好选择，晚上到机住十分合适，服务很热情😄😄，房间没地毯很清爽，海景很好看❤️

清洗后 价廉物美的好选择，晚上到机住十分合适，服务很热情，房间没地毯很清爽，海景很好看

去除URL

繁体转简体

提交 返回

- 去除url：是指去掉文本数据中的网页链接

EasyData - 一站式数据处理和... x +

ai.baidu.com/easydata/app/etl/add

Baidu 百度智能云 | EasyData 产品介绍 数据服务 使用文档 Nanci_wan

清洗任务管理 > 新建清洗任务

清洗后 请选择数据集 版本

请选择清洗方式

通用清洗方案

最多可添加3种清洗方式

去除emoji

去除URL

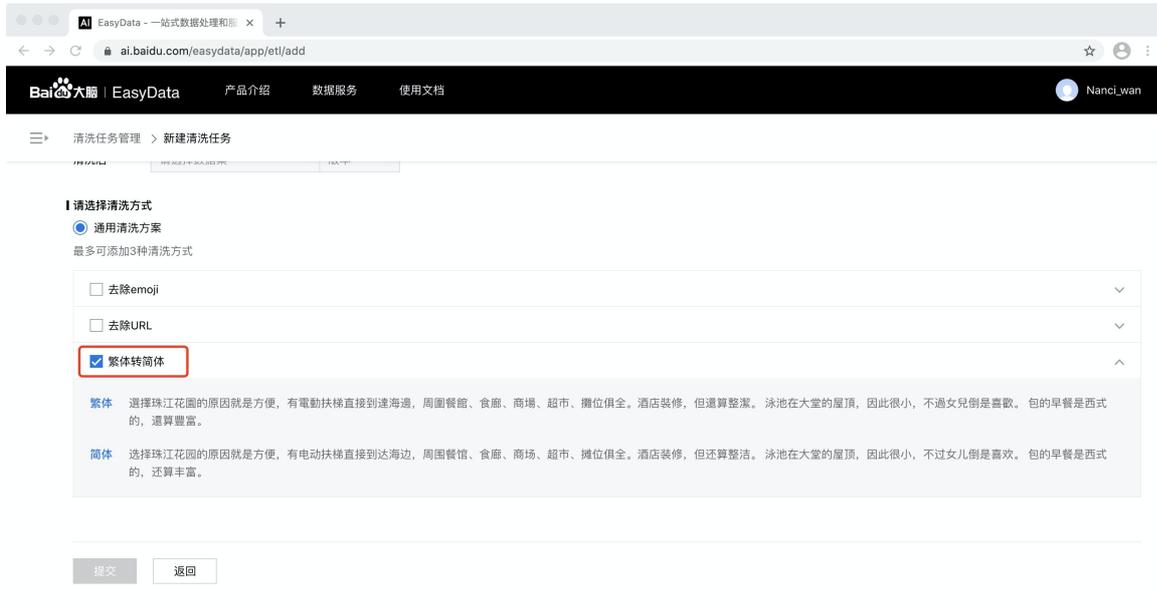
源文本 阿刀田高的《他人同士<https://douban.com/subject/1662858>》《圆梦<https://douban.com/subject/1662861>》《落樱飞雪<https://douban.com/subject/1662860>》《不可思议的声音<https://douban.com/subject/1331463>》《危险的童话<https://douban.com/subject/1662859>》有钱的话可以去看看，我对这套系列的评价正是蔡国庆当年那首freestyle里唱的那样：实在是棒棒棒棒，我无话可说。

清洗后 阿刀田高的《他人同士》《圆梦》《落樱飞雪》《不可思议的声音》《危险的童话》有钱的话可以去看看，我对这套系列的评价正是蔡国庆当年那首freestyle里唱的那样：实在是棒棒棒棒，我无话可说。

繁体转简体

提交 返回

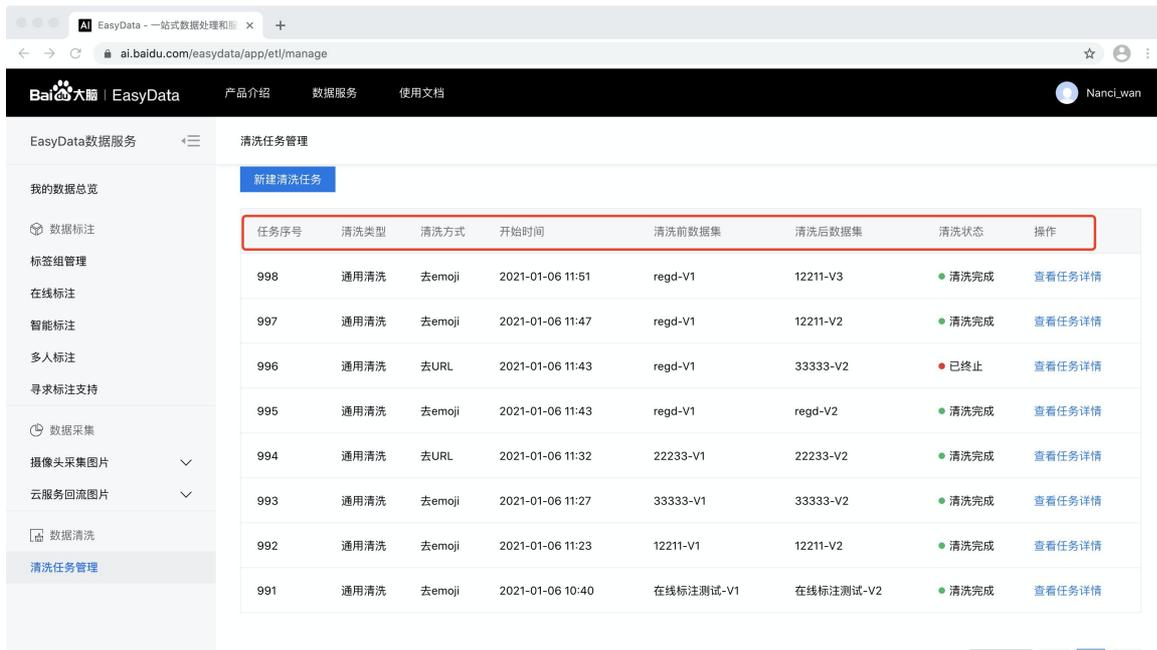
- 繁体转简体：是指讲文本中繁体字转为简体字，通常情况下在使用简体字可以取得更好的模型效果。



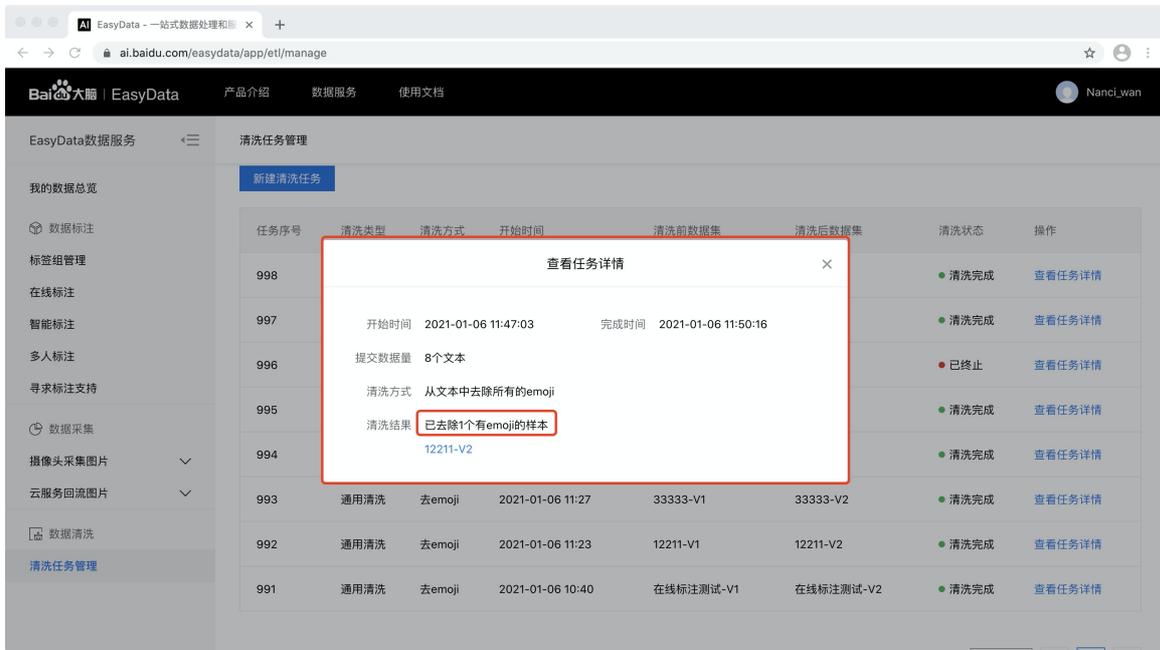
4、清洗任务管理

在完成以上步骤后，点击提交，即可完成清洗任务的创建，并返回清洗任务管理页面。

- **任务概览** 通过列表页展示任务如下字段：任务序号、清洗类型、开始时间、清洗方式、清洗前后数据集版本、清洗状态、操作。如您在清洗状态为正在清洗中时，想结束清洗任务，您可以点击终止任务。



- **查看详情** 点击查看任务详情可看到清洗的结果，如在本文的实例中试用的数据集已去除1个有emoji的样本，说明清洗前文本数据中有一个样本包含emoji，目前已将该样本中emoji去除。点击清洗后数据集即可查看清洗后数据的详细内容。需要注意的是，文本数据清洗后默认了数据去重策略，如您的数据中存在样本重复的情况，将自动去重。



数据标注

在线标注

选择数据集版本

- 快速导入已有数据集

点击在线标注左侧导航，选择已有数据集版本进行标注，支持图片、文本、音频、视频等类型



请选择数据集及版本进行标注

请选择数据集 版本

- 创建数据集

想要重新上传数据集，在我的数据总览页下点击创建数据集并导入

[创建数据集](#)

版本	数据量	最近导入状态	标注类型	标注状态	发布状态	操作
v1	0	导入中	短文本匹配	0% (0/0)	未发布	修改数据服务项 新增版本 全部版本 删除

展示对应标注模板

当选定数据集版本后会根据数据集版本类型、选择的标注分类及模板，展示对应的标注操作框，数据类型、标注类型及模板关联关系如下：

- 图片

图像分类：单图单标签模板、单图多标签模板

物体检测：矩形框标注模板

图像分割：图像分割模板

- 文本

文本分类：单文本单标签模板 短文本匹配：短文本匹配模板 情感倾向分析：情感倾向分析模板 文本实体抽取：文本实体抽取模板

- 音频

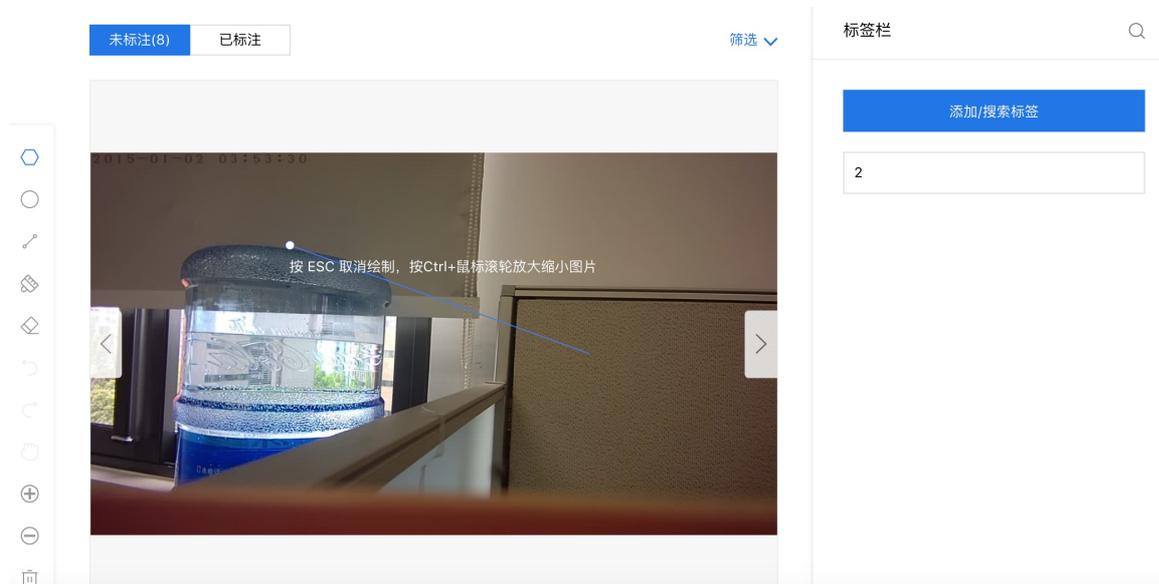
音频分类：短音频单标签模板

- 视频

视频分类：短视频单标签模板

🔗 进行标注操作

在未标tab下，针对以上提供的标注模板进行标注操作，操作流程不同模板基本一致，图像方向以图像分割举例：



第一步，右侧标签栏新建标签，如果图像分类则支持新建多个标签

第二步，中间标注画板按照指导提示进行框选，选定标注框勾选方式（多边形框/圆形框）

第三步，左侧工具栏可对标注框进行画笔大小设置、橡皮擦、图片放大、图片缩小、图片删除、图片添加等具体操作

第四步，完成标注后该图片将在已标tab下展示

文本方向以文本实体抽取举例：

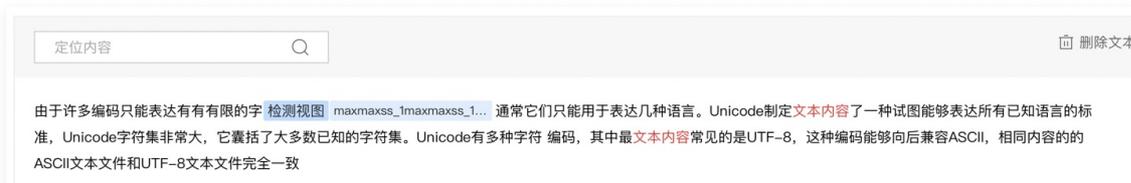
第一步，创建标签 进入到待标注的数据集，您需要在右侧的标签栏中创建标签，点击「添加/搜索标签」后，即可输入标签名称，注意平台支持中英文数字中下划线、长度不超过256字符的标签名，如下图：



第二步，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。您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：



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。

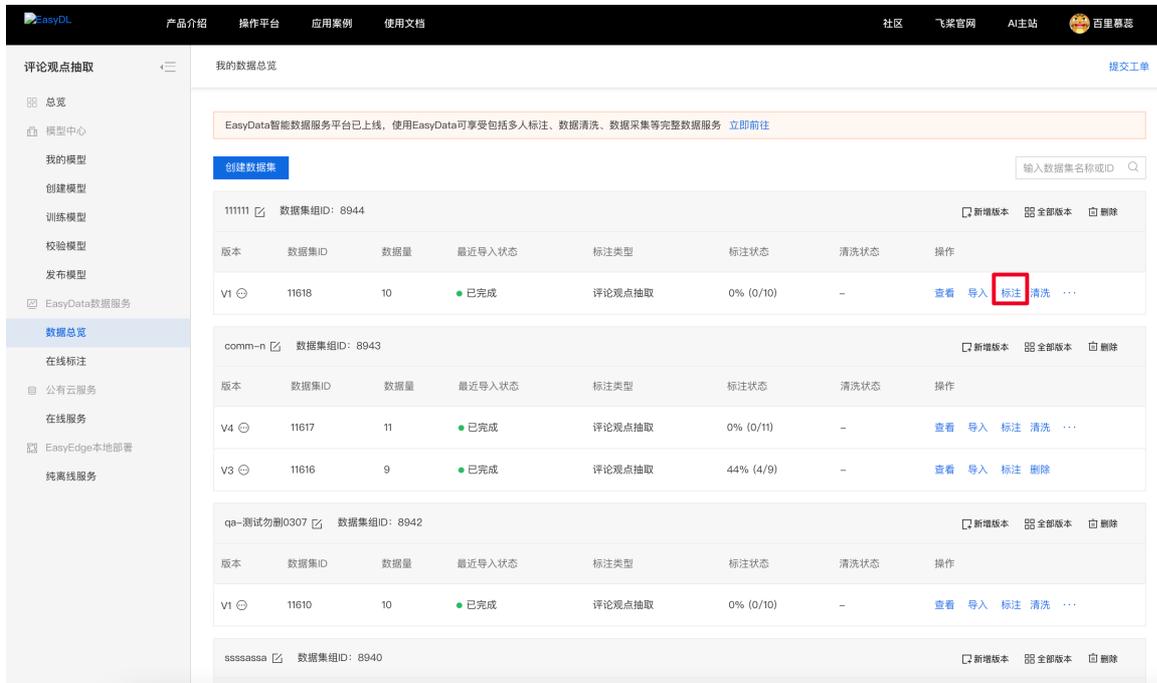


第三步，查看标注信息 您可以在标签中，点击「高亮」图标按钮，来达到让文本区域高亮显示标注信息，以方便您的查看标注情况。图标按钮如下图所示：

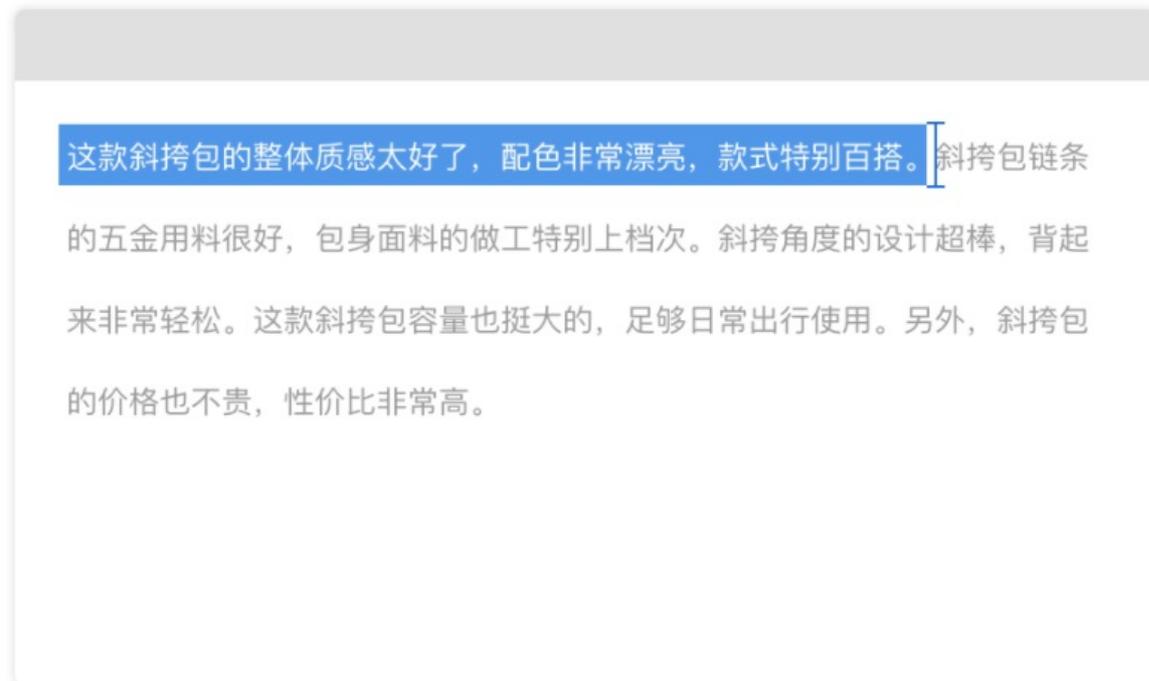


评论观点抽取数据标注：

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。



1. 选取评价片段



2. 选取

评价对象 (必填)

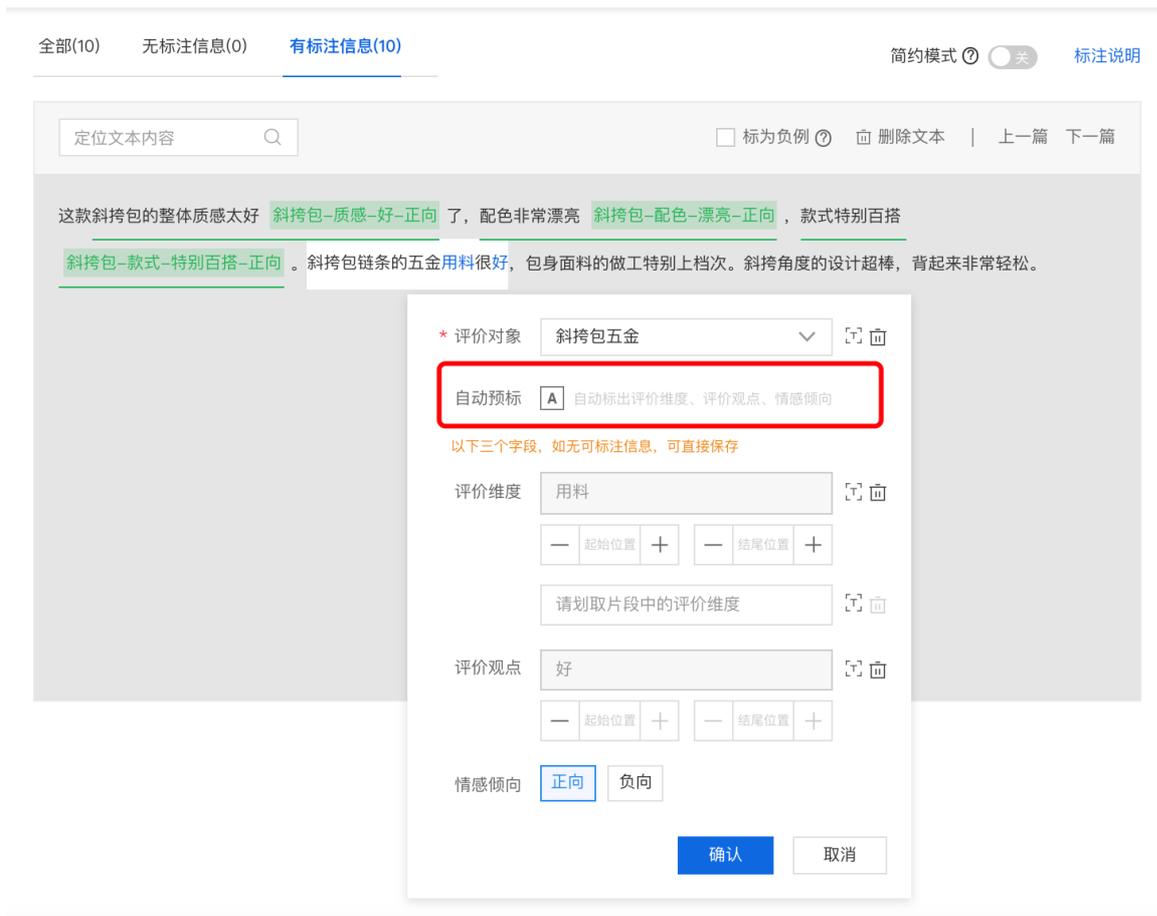
画选评价片段后，会出现弹窗如下，请激活图中按钮划选评价对象或手动输入评价对象。



3.自动预标注

- 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



- 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 关 标注说明



4.完成标注 完成标注后效果如图：



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。

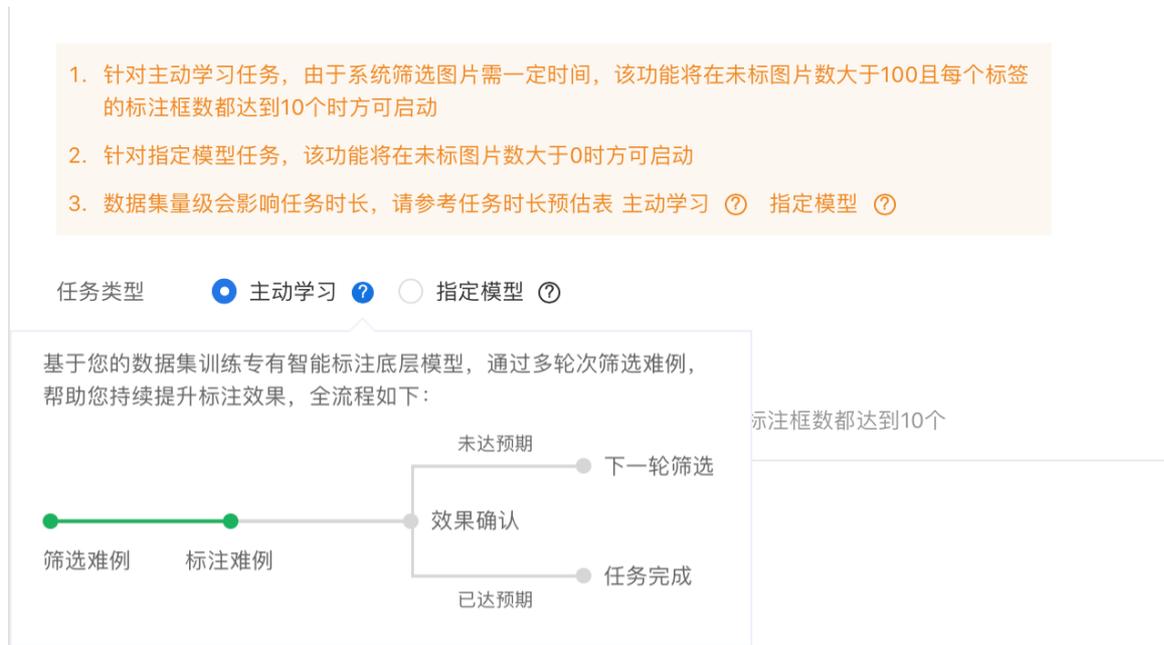
6.简约模式 开启简约模式后, 会隐藏掉详细的标注结果 (评价对象-评价维度-评价观点-情感倾向), 方便您阅读文本内容, 如图为简约模式与普通模式对比：



图像智能标注

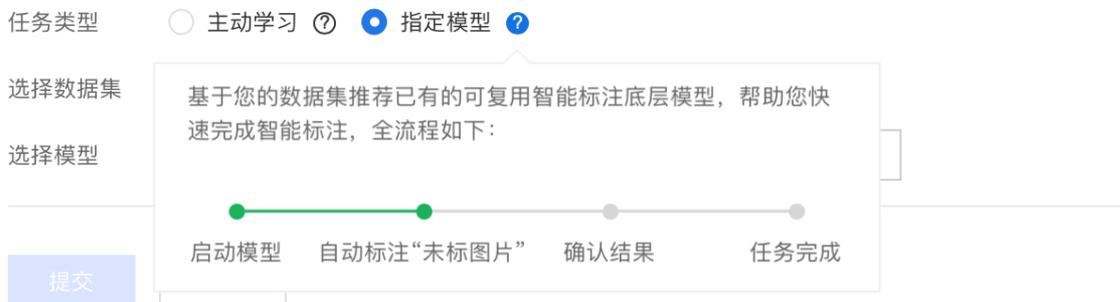
使用智能标注功能可降低数据的标注成本，物体检测、图像分割、文本分类数据集支持您使用智能标注功能。其中，物体检测、图像分割数据提供两种智能标注方式供您选择：主动学习、指定模型。

主动学习即指基于您的数据集训练专有智能标注底层模型，通过多轮次筛选难例，帮助您持续提升标注效果。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。流程如下图：



指定模型即指基于您的数据集推荐已有的可复用智能标注底层模型，帮助您快速完成智能标注，流程如下图：

1. 针对主动学习任务，由于系统筛选图片需一定时间，该功能将在未标图片数大于100且每个标签的标注框数都达到10个时方可启动
2. 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
3. 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](#) [指定模型](#)



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，请先检查一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
- 针对主动学习任务，由于系统筛选图片需一定时间，该功能将在未标图片数大于100且每个标签的标注框数都达到10个时方可启动
- 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

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

选择主动学习任务

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，在此起见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。[了解功能详情](#)
您的智能标注任务预计需要38分钟完成

关闭智能标注



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 ● 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			查看记录
5	3111	tj-智能标注-检测-1110			查看记录
6	2831	py3升级-智能标注-物体检测			查看记录
7	1965	物体检测-多人标注用			查看记录
8	2981	赵粤专属数据集1			查看记录

当前您处于第1轮难例阶段（共4轮）
已为您筛选出本轮难例图片，请确认该轮难例图片

● 任务完成
● 下一轮筛选

[知道了](#)

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) **待确认标注(5)**

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

标注结果

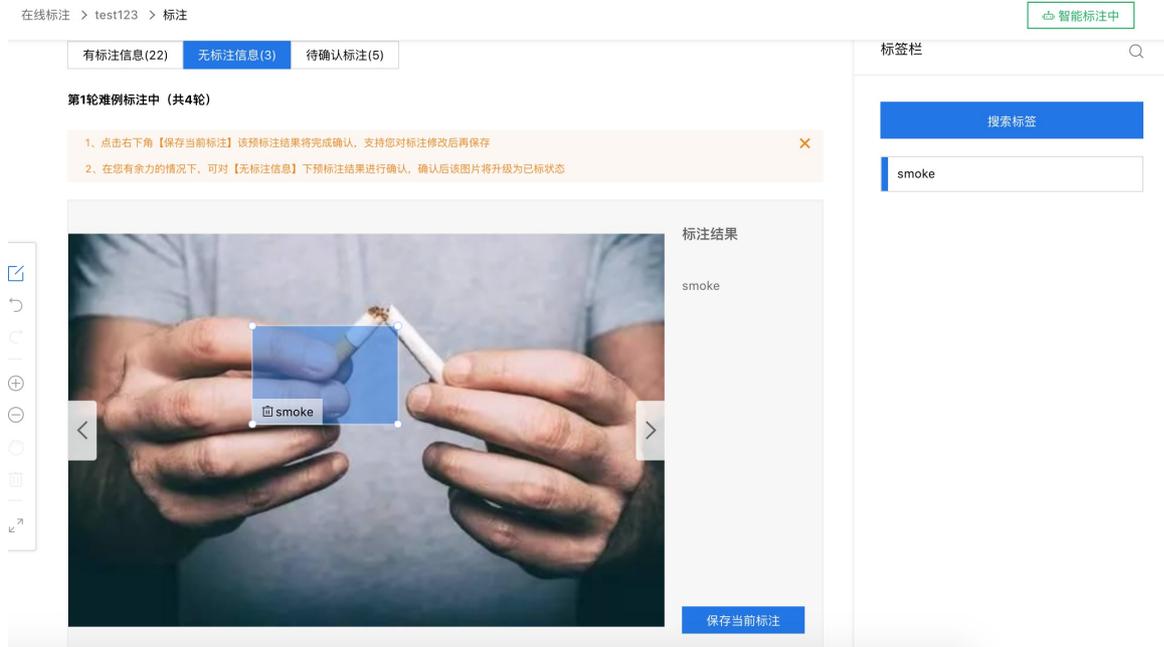
请在右侧选择标签

[保存当前标注](#)

标签栏

smoke

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

当您为难例完成确认后，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



选择指定模型任务

选择模型

支持选择来自EasyDL/BML训练完成的模型，可通过列表页查看该模型名称、ID、版本、创建时间、模型效果辅助您快速定位需复用模型，列表如下图：

选择数据集 dy-智能标注物体检测 V1

您的智能标注任务预计需要 **6-7分钟** 完成

选择模型 来自训练平台 (70个) 来自已有智能标注任务 (10个)

EasyDL BML

输入模型名称或ID

模型名称	模型ID	版本数量	模型来源
> 小麦计数	111955	1	EasyDL
> nose or mouth	110766	1	EasyDL
> 麻将识别-test	110624	1	EasyDL
> hyb_超高精度无验证集	108890	1	EasyDL
> lsf_test	108823	1	EasyDL

支持您选择已成功运行的智能标注任务生成的底层模型，在数据集相似（标签、应用场景）情况下，您可以选择该类模型，完成对新数据集的快速标注，列表如下图：

来自训练平台 (70个) 来自已有智能标注任务 (10个)

输入数据集名称或ID

模型名称	数据集版本ID	数据集组名称	创建时间	标注模板	数据集标签总数	智能标注图片数
<input type="radio"/> 0323-yyw-V2-2021-03-23 16:39:19	167203	0323-yyw	2021-03-23 16:39:19	物体检测	1个	10张
<input type="radio"/> 航拍飞机识别v1-V1-2021-03-23 15:33:26	144446	航拍飞机识别v1	2021-03-23 15:33:26	物体检测	1个	20张
<input type="radio"/> person_smart-V2-2021-03-17 18:14:59	165215	person_smart	2021-03-17 18:14:59	物体检测	1个	33张
<input type="radio"/> dy-智能标注物体检测-V1-2021-02-24 18:55:05	159681	dy-智能标注物体检测	2021-02-24 18:55:05	物体检测	2个	10张
<input type="radio"/> xj_10.5_检测智能标注-V1-2020-12-18 20:15:57	142959	xj_10.5_检测智能标注	2020-12-18 20:15:57	物体检测	2个	40张

查看进度

任务提交后您可点击查看进度，实时了解任务运行详情。

创建智能标注任务

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
1	159681	dy-智能标注物体检测	V1	指定模型	运行中	查看进度 确认结果 中止任务

确认结果

当您的智能标注底层模型复用成功后，可点击确认结果页，对图片所有打上的自动标签进行确认，您可以通过单图确认，也支持您对基本效果满意情况下一键确认。对图片自动打上的标签完成确认后，该张图片将升级为已标图片。

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

🔗 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：

- 在论坛咨询

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。

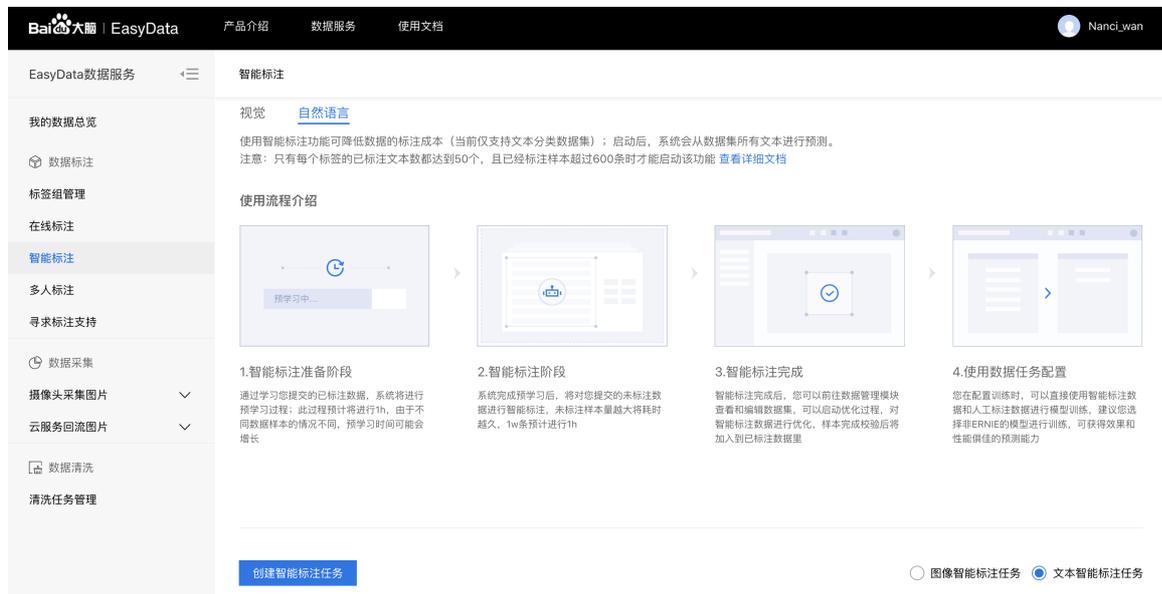
文本智能标注

欢迎您使用「文本智能标注」。您可以通过提供少量人工标注数据和大量无标注数据，通过文本智能标注能力进行自动标注，并将需要人工优先复审的样本筛选出来，辅助您快速完成数据标注工作。您可以获得大规模的智能标注数据，并将数据用于模型的训练。本文将介绍说明如何利用「文本智能标注」获得智能标注数据，以及智能标注与后续如何使用智能标注数据训练模型。

详细使用「文本智能标注」的流程如下：

一、启动智能标注获得智能标注数据 Step1 启动智能标注任务前的准备工作 智能标注任务，是对一个数据集中的未标注数据进行智能标注。请您先在「数据管理/标注」模块，创建数据集，并上传一定量的已标注数据和未标注数据。具体数据量的要求，请见下文「创建文本智能标注任务」。

Step2 创建文本智能标注任务 您可以在「智能标注」功能页面下，选中「自然语言」查看文本智能标注的流程概述和说明。



在页面上，通过点击「创建智能标注任务」来选择对应操作的数据集。选择「文本分类」任务，数据集选择您想智能标注的数据集。



注意，请先检验所选择的数据集是否满足以下条件：

数据集中已标注数据量超过600条 每个标注标签的数据量超过50条 未标注数据的数据量超过600条 选择好数据集后，点击「启动智能标注」，平台将对您提交的数据集进行启动前的校验，并可在下方的任务栏中，查看到新建设的智能标注任务，任务状态处于「数据校验中」。

Step3 智能标注过程 当数据校验无误后，系统将自动启动智能标注，页面将显示当前智能标注所处的流程状态。如页面图中所示，智能标注过程分为三个部分，第四个部分为「前往模型训练」的说明，后文将详细说明。

下面，将对智能标注所涉及的三个阶段进行说明：

阶段一：智能标注预学习阶段

在本阶段，系统将使用您提交的已标注数据进行机器学习。预学习阶段预计耗时约40-60分钟不等，根据您的数据集中已标注数据的量而定。如果您需要中止预学习阶段，您可以点击「关闭智能标注」，则系统将重置已完成的智能学习状态。此时您的数据集中「已标注」和「未标注」的数据不会产生变化。

阶段二：智能标注阶段

完成智能标注预学习阶段，系统将自动进入智能标注阶段。此时系统将对您提交数据集下的「未标注」数据进行智能标注，智能标注过程的耗时跟您提交的未标注数据量正相关，即数据量越大，耗时将越长。您可以根据需要，也可「关闭智能标注」。此时注意，关闭智能标注后，已经完成的智能标注的数据，将被恢复为未标注数据。

Tips：按照我们的估算，预计5w条未标注数据，将消耗2小时进行智能标注。

阶段三：完成智能标注，查看智能标注数据

当完成智能标注后，您可前往您的数据集查看。此时数据集将被分为「未标注」、「已标注（人工）」和「已标注（智能）」三个标签。完成智能标注的数据，将放置在「已标注（智能）」中，您可以对智能标注数据进行人工校验，完成校验的数据，将被转移到「已标注（人工）」，代表此样本确认无误，后文将详细说明智能标注的原理。智能标注前的「已标注」数据将放入「已标注（人工）」中，且数据不发生变化。

您也可以「已标注（智能）」中，查看本次智能标注的准确率，以及智能标注数据的总样本数。如果您对智能标注的准确率不满意，可以点击「优化智能标注效果」，来对全部的智能标注数据进行优化。



智能标注准确率: 81.67% 智能标注样本数: 7479

您可以对智能标注结果进行人工校验，完成校验的数据将加入到“已标注·人工”中，如果您不满意智能标注效果，可点击优化智能标注效果进行优化。如果您满意智能标注效果，可选择智能标注及人工标注数据共同用于模型的训练。 [技术原理说明](#)

全部标签 (2个)	0的文本列表				本页全部确认
标签名称	操作	序号	文本内容摘要	标签	操作
0	删除	1	地板太滑，孩子洗澡出来摔了一跤，幸无大碍。宾馆反馈 2008年6月23日：您...	0	确认 查看
1	删除	2	我想，如果我们想分辨人间的正义，那么，这本书就能告诉你，用谁的眼光来看...	0	确认 查看
		3	首先应该讲携程的服务是非常到位的，为本人提供了不少方便。其次讲讲扬子岛...	0	确认 查看
		4	对这个欲装的系统不感冒，从装费劲 01，bios光盘模式该ide 网查linux改xp有的...	0	确认 查看
		5	装linux，说明书却是教windows使用，没有linux使用说明，也没有重装系统注意...	0	确认 查看

二、使用智能标注数据训练模型 如果您对智能标注的数据质量满意，您可以直接使用「已标注（人工）」和「已标注（智能）」下的数据，对模型进行训练。后文有详细介绍此步骤的原理。

您在配置任务时，在选择数据集的过程中，可以在「选择数据集-可选项」下，勾选「使用智能标注数据」，如果此数据集下有智能标注数据，训练任务时将会使用「已标注（智能）」的数据。

添加数据

选择历史训练任务 [更换版本](#)

选择数据集

+ 请选择

可选项

[全部添加](#)

已选项

[全部删除](#)

公开数据集 我的数据集

使用智能标注数据

三、启动智能标注数据优化流程 如果您对数据质量不满意，可以在「已标注（智能）」页面中，通过点击「优化智能标注」进入优化流程。

全部(8249) 无标注信息 有标注信息·人工 有标注信息·智能

完成校验样本: 0 / 300 智能标注准确率: 81.67%

启动效果提升 退出优化效果

温馨提示: 启动优化智能标注成功, 请您对系统挑选的样本进行校验标注

启动成功 筛选样本 等待您标注已筛样本

知道了

序号	文本内容摘要	标签	操作
1	作为妈妈, 我个人很喜欢这本书, 但是在怀疑这本书是否适合给孩子看? 这本...	0	确认 查看
2	原产哉通过无印良品及一些设计展的设计来展开全书, 继而分析目前日本设计的走...	0	确认 查看
3	没有磨难的爱情不会长久, 哀哀和何适, 他们俩人的问题从来都不是因为第三者...	0	确认 查看
4	上高中的时候就在同学的文章中读到了一句话“学会放弃才意味着成功”, 当时就...	0	确认 查看

进入后, 平台将为您提供至多300条的优先校验样本, 优先校验样本是平台挑选的机器较难确认的样本, 同时也是对提升智能标注准确率最有帮助的样本。此部分样本将按照优先级从高到底排序, 建议您按顺序逐一校验。

在优化流程的页面中, 您将发现已完成校验的数量, 当前的智能标注准确率情况。当您完成100条以上的样本校验后, 「启动效果提升」按钮将置亮, 您可以点击并启动优化智能标注的能力。

全部(8249) 无标注信息 有标注信息·人工 有标注信息·智能

完成校验样本: 116 / 300 智能标注准确率: 81.67%

启动效果提升 退出优化效果

已为您筛选出300条优先提升智能标注效果的样本, 建议您至少完成前100条样本的校验, 过多的样本校验能够更好地提升智能标注的效果。完成校验后, 系统将自动重新评估智能标注的准确率, 并进入下一轮的样本筛选; 您可以随时结束此优化过程。

序号	文本内容摘要	标签	操作
21	9月初去南京, 住的这家酒店, 看网上评价还不错, 可是到了酒店后感觉特不舒服, 客...	1	撤销 查看
22	inter 的SSD硬盘 速度超级快, 现在用笔记本在也不用像以前那样开机后就小心...	1	撤销 查看

此过程将对您完成校验的样本重新学习, 并对剩余智能标注数据进行重新标注。此过程不可关闭, 请您耐心等待。完成优化后, 将为您展示智能标注准确率的提升状况, 并同时为您再次筛选出优先标注样本进行再次的校验和提升。通常情况下, 此过程将持续3次, 即您完成3次的样本校验。您也可以根据智能标注优化的效果, 选择退出优化流程。

Tips: 您也可以在启动优化智能标注前增加「未标注」数据, 在优化智能标注过程中, 也将把新提交的「未标注」数据进行智能标注。

如果您点击「退出优化效果」或您点击进入到了其他页面中, 您在优化流程中的进度将会保存。下次返回后, 将可以继续继续进行。

对样本的校验过程, 将和文本的标注过程类似。当您发现样本的智能标注有误时, 您可以点击选择正确的标签, 点击「确认后, 样本将被放入到「已标注(人工)」中。

如果您对当前页的所有样本都判断为正确, 则可以点击「本页全部确认」将全部样本添加到「已标注(人工)」中。

四、再次启动智能标注 如果您新建了一个文本智能标注任务, 并且完成过一次智能标注。此时, 您可以继续在此数据集下, 增加「未标注」数据, 并对新增的「未标注」数据进行智能标注。

智能标注

智能标注 使用智能标注功能可降低数据的标注成本（当前仅支持文本分类数据集）；启动后，系统会从数据集所有文本进行预测。

展开

创建智能标注任务

图像智能标注任务 文本智能标注任务

序号	数据集版本ID	数据集组名称	版本	智能标注状态	操作
1	147329	在线标注	V1	● 已完成	再次启动 查看结果
2	139811	test5	V1	● 数据校验失败 ?	重新提交数据

Tips：您在再次启动智能标注前，建议先完成至少一次「优化智能标注」的过程。此时，智能标注的数据将比「优化智能标注」前质量更高。

五、文本智能标注原理说明 训练模型过程中，通常需要经历数据集准备（标注）、任务网络配置开发、模型的训练和部署等重要过程。很多时候，模型训练在数据准备阶段遇到数据量不足的问题，使模型开发过程迟迟不能启动。

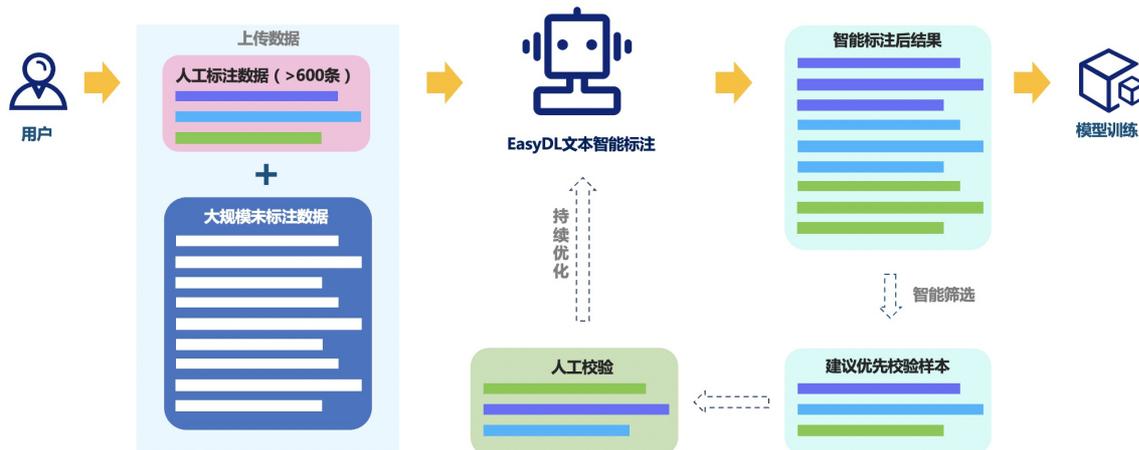
平台推出的文本智能标注功能，目标是通过少量的已标注数据样本，来获得大规模的智能标注数据，通过减少人工逐一校验的工作，使用智能标注数据来训练小型网络模型，以获得效果和性能更优的模型预测服务。使用文本智能标注数据来训练模型的原理，即当前较为流行的「模型蒸馏」。

下面，将为您简要描述智能标注和模型蒸馏的原理。

平台智能标注原理 平台提供的文本智能标注，是使用了当前中文模型效果最好的预训练模型ERNIE2.0。完成对少量人工标注数据学习后，可以对未标注数据进行预测，从而获得智能标注数据。由于预测的数据质量，与您提供的人工标注数据的质量强相关。建议您完成智能标注后，查看数据标注的质量效果，然后根据数据情况来判断是否进入到优化智能标注的流程中。

优化智能标注的过程中，系统根据算法挑选出优先标注样本。您完成对优先校验样本的人工校验后，系统将使用此部分样本重新训练模型，从而获得更精准的智能标注数据。

智能标注数据训练模型 ERNIE完成对人工标注数据学习后，生成的模型称之为「教师模型」，通过预测的大规模无监督语料，把「教师模型」的泛化能力通过模型训练教给「学生模型」。如下图所示：



所以，您可以使用智能标注数据，在训练任务配置中，不使用ERNIE预训练模型。通过选择小型的网络，即可训练出效果逼近教师模型ERNIE Large的模型。且此模型的也保留了小网络的预测性能，可以在您的业务中达到具体业务使用的目标。

六、文本智能标注常见问题 问：智能标注可以支持哪几种类型的数据？

答：目前平台提供的文本智能标注，仅支持「单文本-单标签」的数据类型。

问：我是否能够对智能标注数据进行删除操作？

答：可以，您可以处理所有智能标注数据。通常情况下，您可以根据后续模型训练所需，对某些样本进行删除，以减少模型训

练过程中的干扰。

问：为什么效果没有提升到一定程度，就提示已经完成优化智能标注？

答：因为您的智能标注数据过少，通过人工标注即可更好的完成校验；

问：为什么我使用智能标注数据训练模型效果不佳。不如只使用人工已标注的数据训练的效果？

答：您可以先进入「优化智能标注」过程，平台将为您挑选提升效果最佳的样本，将您的智能标注数据进行优化。同时您也需要查看数据集的数据分布是否与验证集的分布保持一致。

问：我将优化流程中的样本，都删除了，没法启动优化智能标注了，改怎么办？

答：您可以对此数据集增加「未标注」数据，然后返回到「智能标注」任务页面下，找到此数据集对应的智能标注任务，点击「再次启动」。系统将重新启动智能标注过程，并重置「优化流程」

问：完成智能标注后，我是否能在数据集新增带有新标签的标注数据？

答：您可以上传标注数据。但请注意，如果您上传的已标注数据带有新的标签（或您在已经完成智能标注任务的数据集上新增标签），将会影响智能标注质量。如果您需要使用带有新标签的数据，建议您新建数据集，并将全部人工标注数据导入后，再根据智能标注启动条件来启动新的智能标注任务。

问：为什么启动智能标注任务，或启动优化任务时，有时会不成功？

答：目前平台集中公测阶段，因算力限制，仅提供给一种任务的使用。即如果您已经启动了智能标注任务，则无法在另一个智能标注任务中启动优化任务。如果您有业务中的需求，可加入技术交流群，联系工作人员；

问：非中文语种的数据，可以使用智能标注吗？

答：目前平台仅支持中文的智能标注。不支持其他语种的智能标注任务。

问：为什么我使用智能标注数据训练模型，效果无明显提升？

答：不同场景和行业的数据集，使用智能标注数据提升效果不同。建议您针对智能标注数据进行优化流程，在优化流程中将有优先校验样本进行人工校验。完成至少一次优化流程，然后再使用智能标注数据进行模型训练。

寻求标注支持

目前EasyData支持提交数据需求发布至百度众包服务或第三方数据标注服务商两种方式：

🔗 百度众包服务

- 服务介绍

一站式的数据众包服务，可根据特定领域、特定场景的客户需求，提供定制化的数据获取与加工方案，提交以下表单内容后EasyData团队将把您的需求内部转于百度众包服务

- 服务创建

您在提交百度众测服务时，需要对您的数据类型、数据量、标注预算、数据示例、联系方式等以表单形式提交

如果您对众测服务有其他诉求，可以在其他需求中作为备注进行描述

EasyData数据服务 < 寻求标注支持

我的数据总览

- 数据标注
- 在线标注
- 智能标注
- 寻求标注支持
- 数据采集
- 接入摄像头采集图片
- 云服务调用获取数据
- 数据清洗
- 清洗任务管理

提交寻求百度众测服务

数据类型: 图片 文本 音频 视频 其他

数据量: 请选择

大致预算: 请选择

上传文件: 请上传三个示例数据文件

联系方式: 请输入联系方式

其他需求: 请描述下标注类型, 希望标注的效果等 (选填) 0/50

第三方数据标注服务商

目前AI市场已经入驻了十余家优质的数据标注和处理服务商, 提供丰富的数据标注服务, 根据需求进行对接

多人标注

使用多人标注功能可通过团队协作方式, 提高数据标注效率。任务发起后, 系统会向团队成员自动分发标注任务, 成员提交任务后, 管理员可以进行标注结果验收。

管理员创建任务

多人标注页点击创建任务, 管理员可选择需要标注的数据集, 设置当前数据集的标注标签, 选择标注团队, 如果开启标注审核, 则需选择审核团队, 提交任务后系统将启动任务分发。

EasyData数据服务 < 返回 创建多人标注任务

数据集管理

数据标注

标签组管理

在线标注

智能标注

多人标注

寻求标注支持

数据处理

数据清洗

数据增强

数据回流

云服务回流管理

云服务回流数据查看

摄像头采集管理

摄像头采集数据查看

基本信息

* 任务名称 请输入任务名称

* 选择数据集 请选择数据集 [创建数据集](#)

智能标注/多人标注/清洗/共享/导入导出中数据集不可选、且数据集未标数据需大于0

标注任务样例 (支持20M以内 pdf 文件)

* 数据标注状态 未标注

任务分配

* 选择标注团队 请选择 [创建团队](#)

每人标注数量 1个

* 是否审核 是 否

* 任务截止日期 2023/10/6

* 标注完数据集 覆盖已有版本 生成新版本

- 任务名称：自定义
- 选择数据集：下拉可选择由您创建的未被共享、清洗、标注的所有数据集版本

- 标注任务样例：您可上传格式为pdf的标注规范文档，上传后标注员可在标注页面进行查看
- 数据标注状态：包含未标注和已标注，可选择未标注数据进行标注，也可选择已标注数据进行二次校验
- 选择标注团队：下拉可选择您管理的多人标注团队，如需新建团队，可点击创建团队
- 每人标注数量：按照每人任务数*人数小于等于可标注总数的逻辑，系统随机均等分配
- 是否审核：平台支持配置是否审核，开启审核后标注员完成标注后任务会随机分配给审核团队成员进行审核，支持配置审核抽检比例。
- 任务截止日期：任务需在设置截止日期前完成标注，任务过期后无法再进行标注，注意：任务截止日期临近仍未完成标注，建议修改截止日期，避免任务过期后无法再继续标注
- 标注完数据集：您可以选择是覆盖当前已选择版本，或在当前版本号维度上新增一个版本号，后者将会生成一个新的数据集版本

任务ID	任务名称	数据来源	标注数据ID	标注数据名称	标注类型	任务状态	创建时间	操作
606663	ddd	数据集	1657959	头盔检测1 V1	物体检测	标注中 已标注: 0/1	2023-09-26 10:53:10	任务详情 查看进度 终止任务 删除
79	jingtest1014	数据集	115748	test3 V1	物体检测	已终止 (不保留标注)	2020-10-13 17:28:01	任务详情 删除
51	jingtest4	数据集	115750	test3 V2	物体检测	已终止 (标注保留)	2020-09-23 16:56:17	任务详情 验收任务 删除
49	jingtest2	数据集	115748	test3 V1	物体检测	验收已完成	2020-09-23 15:57:38	任务详情 查看数据集 删除
38	jingtest	数据集	115727	test2 V1	图像分割	验收已完成	2020-09-22 14:57:54	任务详情 查看数据集 删除

管理员新建标注团队

在创建任务过程中，如果管理员尚无标注团队，可以点击新建团队完成标注成员添加。创建完毕的标注团队，管理员也可对成员信息进行增删改查。

< 返回 创建团队

1 EasyData迁入百度智能云控制后，原有标注任务、团队成员需重新配置。

团队类型 标注团队 审核团队

* 团队名称

团队描述 0/50

* 团队成员

成员用户名	备注	状态	操作
暂无成员			

+ 添加成员

如需新增子用户，可前往 [IAM控制台](#) 新建子用户

确认 取消

- 团队类型：分为标注团队/审核团队

- 团队名称：自定义
- 团队描述：自定义
- 团队成员：可将当前账号创建的子账户添加为团队成员，注意：1个团队的标注成员上限为20人

EasyData迁移至百度智能云Console后，多人标注任务仅可分发给百度智能云子用户，原通过邮箱添加的团队人员将会失效，您需要通过主账号重新创建和管理标注团队。

如您需要创建百度智能云子用户，可通过团队管理页面入口或者通过超链接

(<https://console.bce.baidu.com/iam/#/account/user/list>)进入IAM界面创建子用户，如子用户仅用于多人标注则无需配置权限，也无需绑定百度账号。

标注成员标注任务

1. 标注成员登陆百度控制台，进入EasyData产品页面多人标注-我接收的任务。



注意：百度智能云控制台无法搜索到EasyData独立产品，可通过ai.baidu.com->导航->开发平台->EasyData数据服务平台，进入EasyData产品页面。



2. 我接收的任务列表，可以点击具体任务启动标注，进入标注详情页完成标注任务，支持标注员多次登录完成标注，注意：标注成员不可以对标注标签进行增删改查
3. 点击页面右上角提交标注任务，注意：标注任务一旦提交后不可再进行二次编辑，提交后标注员在我接收的任务列表，可看到该条任务记录

审核成员审核任务

如果管理员在创建多人标注任务时，开启了审核功能，那么在标注成员提交标注任务后，审核成员将开始审核任务。

1. 审核成员标注成员登陆百度控制台，进入EasyData产品页面多人标注-我接收的任务，在“审核任务”按键下，查看待审核的任务。
2. 在“操作”列中，选择“去审核”，在审核详情页面中，选择“通过”或“不通过”标注信息。
3. 审核完成后，点击右上角的“提交任务”，任务状态变为“审核完成”。

管理员管理任务

管理员可对待标注、标注中、验收已完成的标注任务进行对应操作：

多人标注

任务ID	任务名称	数据集ID	来源数据集	标注类型	任务状态	创建时间	操作
154	我的中国心	2132	tj-新的数据集 V1	物体检测	待标注	2020-09-17 17:12:41	任务详情 查看进度 终止任务 删除
153	任务数量	2180	tj-多人标注-情感的 V1	情感倾向分析	待标注	2020-09-17 14:17:55	任务详情 查看进度 终止任务 删除
152	zzy-23	2150	yyw-audio V2	音频分类	待标注	2020-09-17 13:52:15	任务详情 查看进度 终止任务 删除
151	zzy-12	2267	YYW-DUAL V5	图像分类	待标注	2020-09-17 13:49:29	任务详情 查看进度 终止任务 删除
150	zzy-123	2269	zzy-多人标注 V1	图像分类	待标注	2020-09-17 13:48:04	任务详情 查看进度 终止任务 删除
149	tjddd	2299	tj-回归限制 V2	物体检测	待标注	2020-09-17 13:46:15	任务详情 查看进度 终止任务 删除
148	新的回归任务-删-新	2298	tj-回归限制 V1	物体检测	标注中	2020-09-17 13:38:19	任务详情 查看进度 终止任务 删除
143	测试任务	2299	tj-回归限制 V2	物体检测	验收已完成	2020-09-16 21:29:12	任务详情 查看数据集 删除
142	非供热热	2298	tj-回归限制 V1	物体检测	验收已完成	2020-09-16 18:31:53	任务详情 查看数据集 删除
141	覆盖的数据集	2298	tj-回归限制 V1	物体检测	验收已完成	2020-09-16 18:24:05	任务详情 查看数据集 删除

重要操作如下：

- 查看进度：支持对标注成员的标注进度、标注任务的总进度进行查看，同时，对于单个标注成员，可以提前结束该标注成员的任务，也可点击【任务移交】将成员进行替换，以避免某标注成员在标注过程中因不可抗因素无法继续完成标注任务，移交时可选择是否将上位标注成员的已标内容生效

多人标注 > 查看进度

数据集ID: 2298 名称: 新的回归任务-删-新 团队人数: 2 每人标注数量: 0 任务总量: 25 整体任务完成度: 4% (已标注: 1, 已删除: 0)

序号	邮箱	已标注数量	删除数量	剩余待标量	标注进度	状态	创建时间	操作
348	tia****g06@baidu.com	1	0	11	8%	标注中	2020-09-17 13:40:01	查看任务链接 结束标注 任务移交
347	yan****wei@baidu.com	0	0	13	0%	结束标注(保留修改)	2020-09-17 13:40:01	查看任务链接 任务移交

任务移交

下一位标注成员邮箱:

[确定](#) [取消](#)

每页显示: 20

- 终止任务：管理员可以提前结束整个任务，并选择是否将已标内容生效，结束后标注成员均不可再继续标注

Baidu 百度智能云 | EasyData 产品介绍 数据服务 使用文档

EasyData数据服务 < 多人标注

终止任务

终止后所有成员不可再进入标注页面

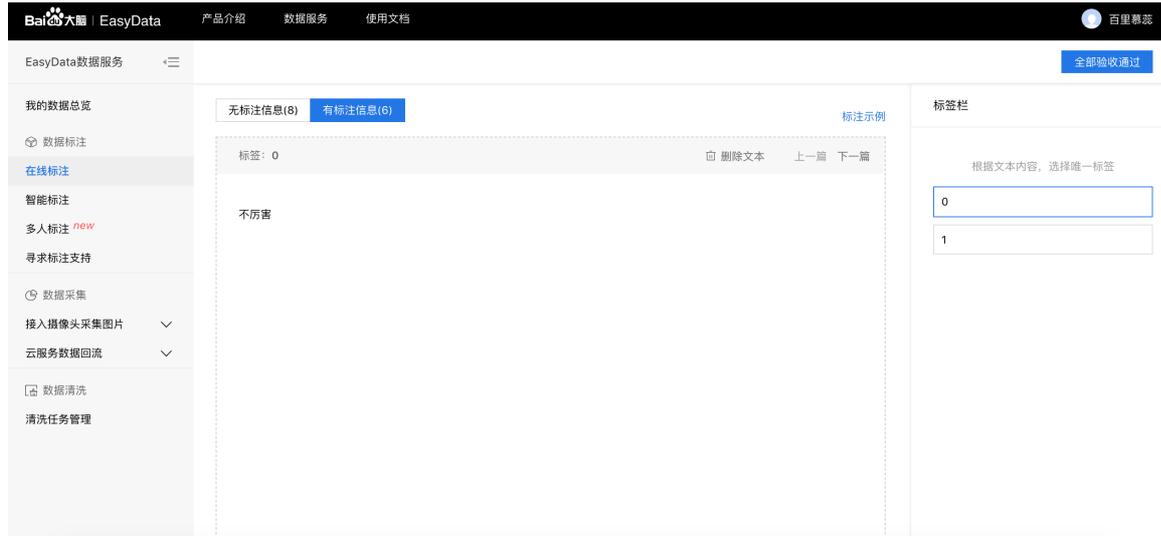
已标注内容是否生效: 是 否

[确定](#) [取消](#)

任务ID	任务名称	数据集ID	来源数据集	标注类型	任务状态	创建时间	操作
154	我的中国心	2132	tj-新的数据集 V1	物体检测	待标注	2020-09-17 17:12:41	任务详情 查看进度 终止任务 删除
153	任务数量	2180	tj-多人标注-情感的 V1	情感倾向分析	待标注	2020-09-17 14:17:55	任务详情 查看进度 终止任务 删除
152	zzy-23	2150	yyw-audio V2	音频分类	待标注	2020-09-17 13:52:15	任务详情 查看进度 终止任务 删除
151	zzy-12	2267	YYW-DUAL V5	图像分类	待标注	2020-09-17 13:49:29	任务详情 查看进度 终止任务 删除
150	zzy-123	2269	zzy-多人标注 V1	图像分类	待标注	2020-09-17 13:48:04	任务详情 查看进度 终止任务 删除
149	tjddd	2299	tj-回归限制 V2	物体检测	待标注	2020-09-17 13:46:15	任务详情 查看进度 终止任务 删除

- 验收任务：默认所有标注员在任务有效期内点击任务提交后，管理员可对任务进行验收；也支持管理员针对特殊情况，点击

【终止任务】后提前将任务结束进行验收。管理员点击【任务验收】后会跳转至该任务标注详情页进行标注内容校验（支持二次修改），完成校验后可点击右上角【全部验收通过】，验收通过后，管理员可在我的数据总览列表页对该数据集版本进行后续清洗、导出等更多数据处理操作



标签组管理

标签组管理主要解决您在标签数较多情况下，批量添加、修改、删除标签的使用诉求。

创建标签组

从左侧导航进入标签组列表页，点击创建标签组弹窗，输入标签名、标签描述（非必选）即可完成创建



管理标签组

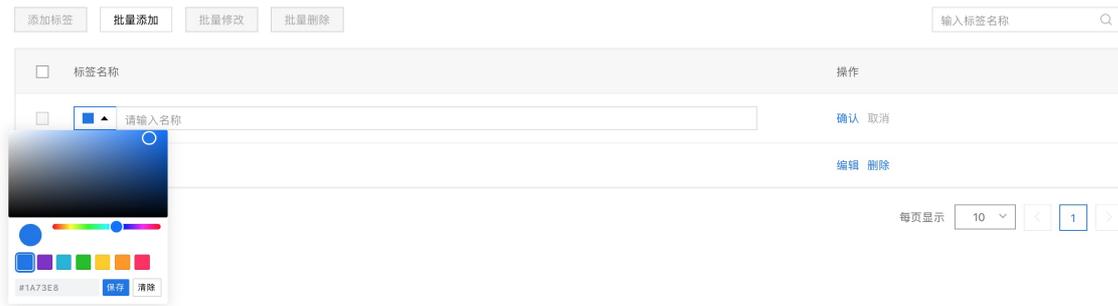
针对某个标签组，平台提供针对单个标签组进行删除、修改标签组名称描述等通用功能。此外，支持用户对某个标签组进行管理。



添加标

签

点击【添加标签】支持手动完成新标签添加，添加过程支持用户对标签颜色自定义添加，标签名称支持中英文数字中下划线共256字符。



批量添加

当您的标签数较多时，可点击【批量添加】，通过导入csv、xls、txt完成标签名批量导入。



批量修改、删除

当勾选左侧多选框时，您可以对所选定标签进行批量修改颜色、批量删除操作。



应用标签组

您可以选定对应的数据集版本进入在线标注页面，点击右侧【添加】即可选择导入对应的标签组为目标数据集版本完成快速添加标签。

注意：一个数据集版本最多支持添加1000个标签。

在线标注 > zq_666 > 标注 zq_666 V1

无标注信息(102)有标注信息(1)



<>

标注结果

请在右侧选择标签

标签栏

添加 ▲

添加标签

添加标签组

根据图片内容，选择唯一标签

的非

非

的

离开的时间发dlkfds

可来得及发离开的时间发多少

发离开的时间发看到老师

立刻带积分

数据增强

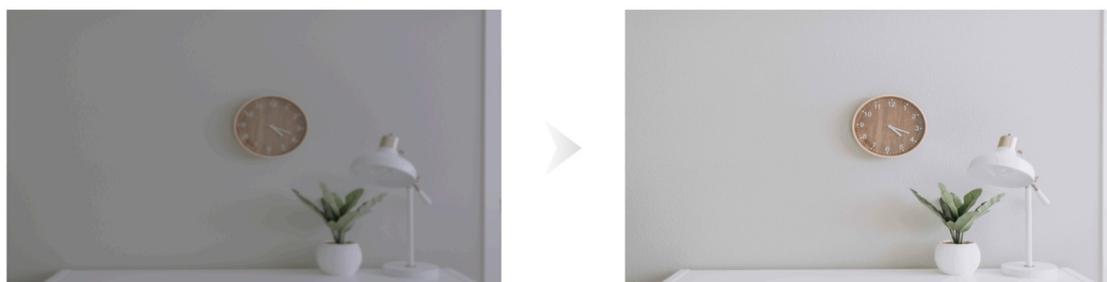
增强任务管理

目录

- [1. 创建数据增强任务](#)
- [2. 查看数据增强任务](#)

增强任务管理

数据增强是对数据进行扩充、提升样本丰富度的有效手段。当在实践中无法收集到数目庞大的高质量数据时，可以通过数据增强策略，对数据本身进行一定程度的扰动和扩充，从而产生“新”数据。在训练时会通过学习大量的“新”数据，提高模型的泛化能力。



创建数据增强任务

- 进入[百度EasyData首页](#)，点击导航栏的数据处理
- 选择“数据处理>增强任务管理”，进入“数据增强任务列表”页面。

任务序号	增强类型	增强方式	算子处理策略	开始时间	增强前数据集	增强后数据集	增强状态	操作
98	物体检测	Solarize_add、ShearX_BBox、Cutout	并行遍历	2022-05-30 11:42:40	tj检测2张图-V2	tj_0530_det_小-V1	● 已完成	查看任务详情
97	物体检测	Posterize、Equalize、Solarize_Only_BBoxes、BBox_Cutout、Equalize_Only_BBoxes、Flip_Only_BBoxes	串行叠加	2022-05-30 11:37:49	tj-大图-已标注-V3	tj_0503_det_局部-V2	● 已完成	查看任务详情
96	物体检测	ShearX_BBox、ShearX_Only_BBoxes、ShearY_BBox、ShearY_Only_BBoxes、TranslateX_BBox、TranslateX_Only_BBoxes、TranslateY_BBox、TranslateY_Only_BBoxes、Rotate_BBox	并行遍历	2022-05-30 11:36:56	tj-大图-已标注-V3	tj_0503_det_局部-V1	● 已完成	查看任务详情

3. 点击“新建增强任务”，进入创建数据增强任务页面，填写配置信息，包括：

增强任务管理 > 新建增强任务

请选择标注类型

- 数据类型 图像类
- 标注类型 图像分类 物体检测 (暂不支持自定义四边形标注模板) 图像分割
- 增强区域 图片全局增强 全局以及标注框局部增强

请选择数据集

温馨提示：增强任务未结束之前，该数据集会暂时锁定，期间无法对其进行导入导出等操作，以及开启数据清洗、智能标注或多人标注任务。

数据输入

选择标签 全选

- aaa
- bbb

数据输出

请选择增强算子

- AutoContrast 效果展示
- Invert 效果展示

- 数据标注类型：目前增强任务仅支持图片类数据，请选择图片类型下的具体标注类型，当前支持类型有图像分类、物体检测（仅支持矩形框标注）、图像分割。
- 选择数据集：选择数据增强的输入输出数据集和对应版本。当数据集为已标注数据时，可选择部分标签。增强任务未结束之前，该数据集会暂时锁定，期间无法对其进行导入导出等操作，以及开启数据清洗、智能标注或多人标注任务。
- 选择增强算子：在选择好标注类型和数据集后，会对应展示该类别下可用的数据增强算子。点击算子名称或“效果展示”，可查看各个算子的图文示例。

请选择增强算子

ShearX 收起效果

剪切图像的水平边，能更好地识别发生了水平方向变形的图像



ShearY 效果展示

- 选择算子处理策略：当前提供两种算子处理策略，分别为串行叠加和并行遍历。选择算子完成后，平台会给出提示：当前账号下已有xx张图片，预计该次任务会产生xx张新图片，个人账号的图片数上限为20w张，请保证有足够的空间存放增强后数据集。

串行叠加：



并行遍历：



4. 完成配置后点击“提交”，完成数据增强任务的创建。

5. 返回“数据增强任务列表”页面

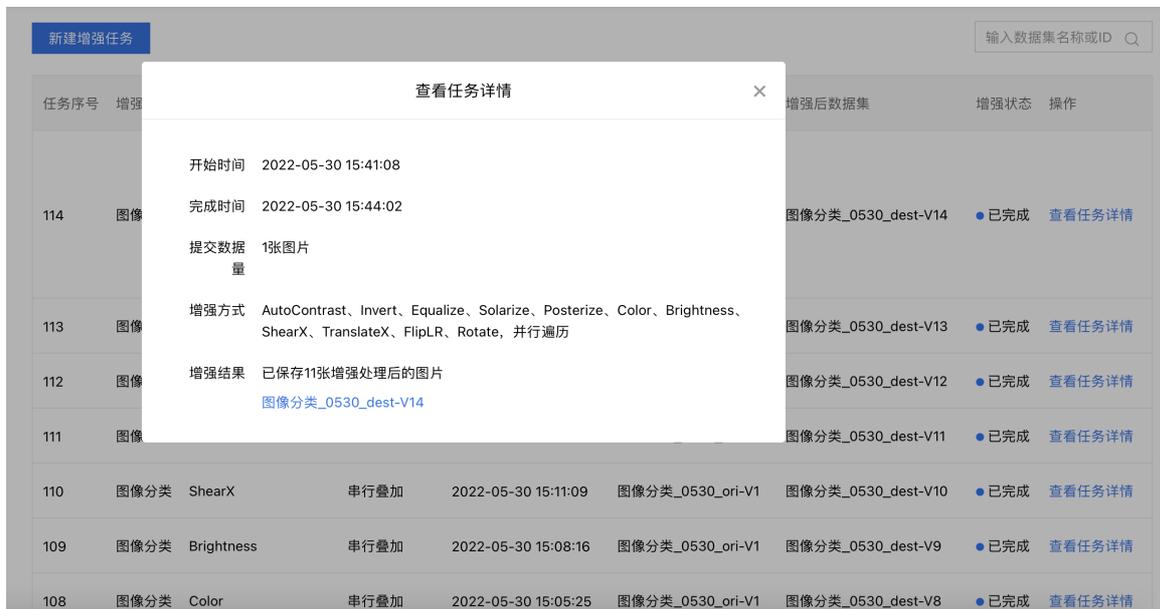
任务序号	增强类型	增强方式	算子处理策略	开始时间	增强前数据集	增强后数据集	增强状态	操作
98	物体检测	Solarize_add、ShearX_BBox、Cutout	并行遍历	2022-05-30 11:42:40	tj检测2张图-V2	tj_0530_det_小-V1	● 已完成	查看任务详情
97	物体检测	Posterize、Equalize、Solarize_Only_BBoxes、BBox_Cutout、Equalize_Only_BBoxes、Flip_Only_BBoxes	串行叠加	2022-05-30 11:37:49	tj-大图-已标注-V3	tj_0503_det_局部-V2	● 已完成	查看任务详情
96	物体检测	ShearX_BBox、ShearX_Only_BBoxes、ShearY_BBox、ShearY_Only_BBoxes、TranslateX_BBox、TranslateX_Only_BBoxes、TranslateY_BBox、TranslateY_Only_BBoxes、Rotate_BBox	并行遍历	2022-05-30 11:36:56	tj-大图-已标注-V3	tj_0503_det_局部-V1	● 已完成	查看任务详情

查看数据增强任务

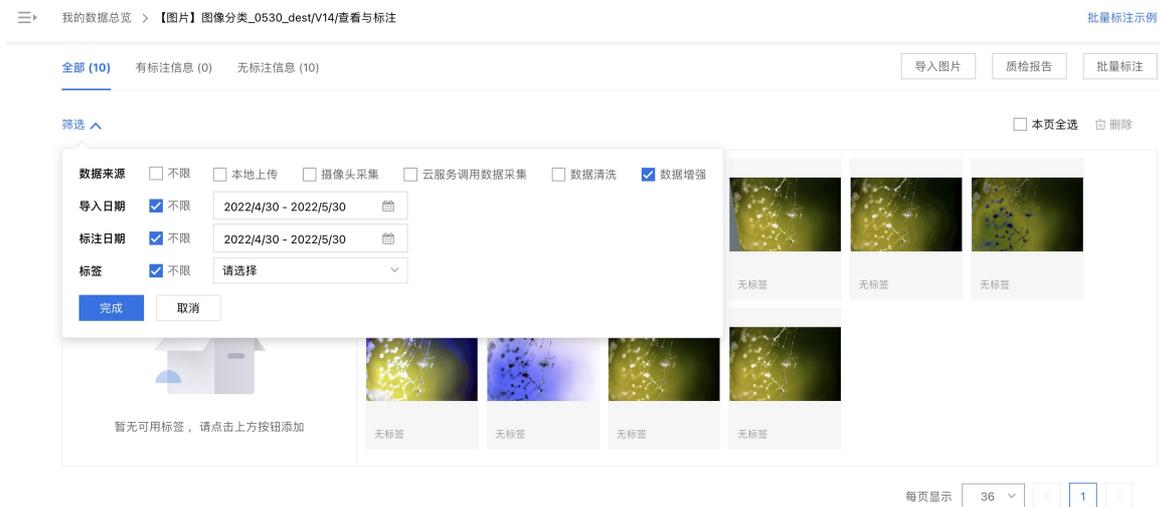
查看数据增强任务的基本信息，包括当前任务的序号、增强类型以及对应的算子列表和算子处理策略、任务的开始时间、增强前数据集、增强后数据集以及当前的任务状态等。

1. 选择“数据处理>增强任务管理”，进入“数据增强任务列表”页面。

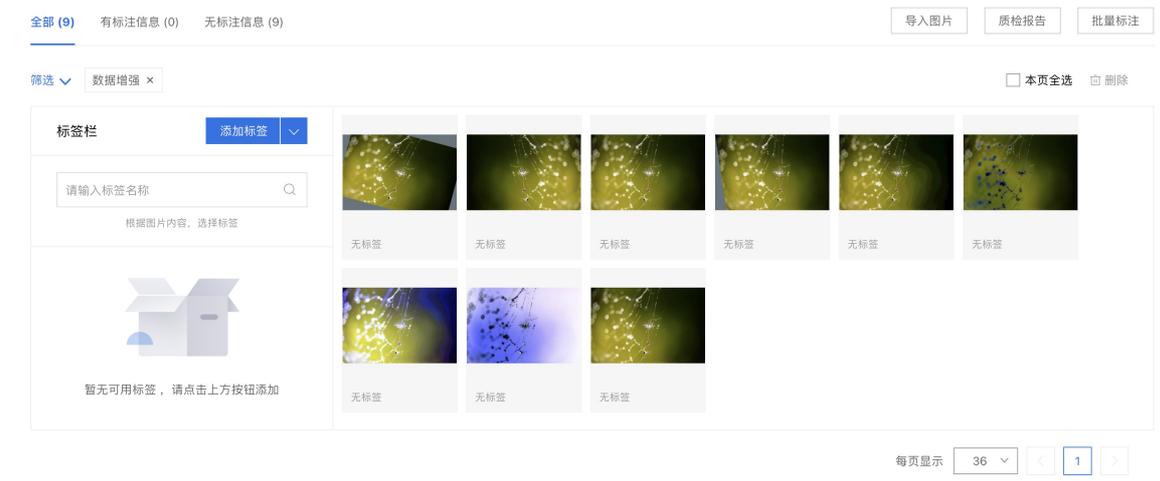
2. 点击“查看详情”，可查看当前任务的详细信息。



3. 点击增强结果中保存后的数据集链接，进入数据集：



4. 进入筛选，选择数据来源“数据增强”并点击完成，即可展示出任务生成的增强后图片，此时可进行人工的查看与审核。



联系我们

您在使用EasyData过程中可以通过以下任何方式联系我们：

1、在论坛咨询

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](#)

2、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](#)

3、添加EasyDL及EasyData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。

产品升级记录

本文为您介绍平台功能及文档更新的最新动态，方便您及时了解版本新增特性及功能变更。

2020年11月更新记录

| 序号 | 模块 | 功能描述 |

- | 1 | 智能标注 | 对图像分割、物体检测智能标注流程进行交互体验升级
- | 2 | 在线标注 | 对数据集标签定义规则新增支持中文标签功能
- | 3 | 在线标注 | 对标注框新增自定义颜色设置功能
- | 4 | 数据导入 | 文本实体抽取模板支持已标注数据支持导入Excel、API数据
- | 5 | 数据导入 | 【EasyDL】短文本相似度模板支持导入API数据
- | 6 | 数据导入 | 对数据导入报错原因进行细分提示优化
- | 7 | 数据导入 | 支持按照图片、视频、音频数据集中的不同标签分别导入数据
- | 8 | 数据回流 | 针对云服务回流的数据支持带识别结果快速保存
- | 9 | 数据采集 | 支持用户通过手动截图方式获取视频截图作为训练数据

2020年9月更新记录

| 序号 | 模块 | 功能描述 |

- | 1 | 智能标注 | 新增智能标注时长预估提示：针对智能标注任务实时提示任务时长
- | 2 | 多人标注 | 新增多人标注功能：针对标注任务支持轻量级团队协作完成
- | 3 | 在线标注 | 新增文本实体抽取标注模板
- | 4 | 数据导入 | 物体检测类型数据下新增支持导入VOC格式数据
- | 5 | 数据管理 | 新增数据集版本删除、数据集搜索功能
- | 6 | 数据采集 | 新增支持RTMP协议、支持Mac操作系统、树莓派开发板
- | 7 | 数据类型 | EasyDL经典版智能数据服务模块新增支持ML表格数据导入
- | 8 | 体验升级 | 打平EasyDL经典版、专业版数据模块与EasyData独立平台的交互体验

2020年7月更新记录

| 序号 | 模块 | 功能描述 |

- | 1 | 数据清洗 | 针对图像数据新增支持按照对是否有人脸、有人体进行图片过滤，支持客户的定制化清洗诉求
- | 2 | 在线标注 | 针对物体检测、图像分割，上线全屏标注功能，方便用户快速定位标注目标
- | 3 | 在线标注 | EasyDL经典版智能数据服务模块、EasyData针对文本数据新增支持情感倾向分析标注模板
- | 4 | 在线标注 | EasyDL专业版智能数据服务模块针对文本数据新增支持序列标注模板

| 5 | 体验升级 | 初步打平EasyDL经典版、专业版数据模块与EasyData独立平台的交互体验

2020年5月更新记录（平台上线）

| 序号 | 模块 | 功能描述 |

| 1 | 数据采集 | 支持用户接入摄像头、图片传输协议等方式进行图像数据本地采集后传输至云端进行管理 & 处理

| 2 | 在线标注 | 支持图像分类、物体检测、图像分割、文本分类、文本相似度、音频分类、视频分类等标注模板

| 3 | 智能标注 | 针对物体检测、图像分割、文本分类支持发起智能标注任务

| 4 | 数据清洗 | 针对图片支持去模糊、去重、旋转、镜像、裁剪等清洗

| 5 | 云服务数据回流 | AI模型训练后可能需要持续迭代和优化模型效果，可以通过调用EasyDL云服务接口并开通该采集服务，将实际业务数据及识别结果可视化查看，针对性的选择高质量数据。

数据增强