ModelArts

数据标注

文档版本 01

发布日期 2022-03-24





版权所有 © 华为技术有限公司 2022。 保留一切权利。

非经本公司书面许可,任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部,并不得以任何形式传播。

商标声明



HUAWE 和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标,由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束,本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定,华为公司对本文档内容不做任何明示或暗示的声明或保证。

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,本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

目录

| 1 数据标注简介 | |
|----------------|----|
| 2 人工标注 | 3 |
| 2.1 创建标注作业 | 3 |
| 2.2 图片标注 | |
| 2.2.1 图像分类 | |
| 2.2.2 物体检测 | 14 |
| 2.2.3 图像分割 | 19 |
| 2.3 文本标注 | 25 |
| 2.3.1 文本分类 | 25 |
| 2.3.2 命名实体 | 28 |
| 2.3.3 文本三元组 | 31 |
| 2.4 音频标注 | 35 |
| 2.4.1 声音分类 | 35 |
| 2.4.2 语音内容 | 38 |
| 2.4.3 语音分割 | 40 |
| 2.5 视频标注 | 42 |
| 3 智能标注 | 46 |
| 3.1 创建智能标注作业 | 46 |
| 3.2 难例确认 | 49 |
| 4 团队标注 | 52 |
| 4.1 团队标注简介 | |
| 4.2 创建和管理团队 | 52 |
| 4.2.1 管理团队 | 52 |
| 4.2.2 管理成员 | 54 |
| 4.3 创建团队标注任务 | 55 |
| 4.4 登录数据标注平台 | 59 |
| 4.5 启动团队标注任务 | 61 |
| 4.6 审核团队标注任务结果 | 62 |
| 4.7 验收团队标注任务结果 | 64 |

◆ 数据标注简介

模型训练过程中需要大量已标注的数据,因此在模型训练之前需要进行数据标注作业。ModelArts为用户提供了标注数据的能力:

- **人工标注**:用户创建单人标注作业,对数据进行手工标注。
- **智能标注**:在标注一定量的数据情况下,用户可以通过启动智能标注任务对数据 进行自动标注,提高标注的效率。
- **团队标注**:对于大批量的数据,用户可以通过创建团队标注作业,进行多人协同标注。

人工标注

对于不同类型的数据,用户可以选择不同的标注类型。当前ModelArts支持如下类型的标注作业:

- 图片
 - 图像分类:识别一张图片中是否包含某种物体。
 - 物体检测:识别出图片中每个物体的位置及类别。
 - 图像分割:根据图片中的物体划分出不同区域。
- 音频
 - 声音分类:对声音进行分类。
 - 语音内容:对语音内容进行标注。
 - 语音分割:对语音进行分段标注。
- 文本
 - 文本分类:对文本的内容按照标签进行分类处理。
 - 命名实体:针对文本中的实体片段进行标注,如"时间"、"地点"等。
 - 文本三元组:针对文本中的实体片段和实体之间的关系进行标注。
- 视频

视频标注:识别出视频中每个物体的位置及分类。目前仅支持mp4格式。

智能标注

除了人工标注外,ModelArts还提供了智能标注功能,快速完成数据标注,为您节省70%以上的标注时间。智能标注是指基于当前标注阶段的标签及图片学习训练,选中系统中已有的模型进行智能标注,快速完成剩余图片的标注操作。

目前只有"图像分类"和"物体检测"类型的数据集支持智能标注功能。

团队标注

数据标注任务中,一般由一个人完成,但是针对数据集较大时,需要多人协助完成。 ModelArts提供了团队标注功能,可以由多人组成一个标注团队,针对同一个数据集进 行标注管理。

团队标注功能当前仅支持"图像分类"、"物体检测"、"文本分类"、"命名实体"、"文本三元组"、"语音分割"类型的数据集。

不同类型数据集支持的功能列表

其中,不同类型的数据集,支持不同的功能,详细信息请参见表1-1。

表 1-1 不同类型数据集支持的功能

| 数据集 类型 | 标注类型 | 人工标注 | 智能标注 | 团队标注 |
|-----------|-------|------|------|------|
| 图片 | 图像分类 | 支持 | 支持 | 支持 |
| | 物体检测 | 支持 | 支持 | 支持 |
| | 图像分割 | 支持 | - | - |
| 音频 | 声音分类 | 支持 | - | - |
| | 语音内容 | 支持 | - | - |
| | 语音分割 | 支持 | - | 支持 |
| 文本 | 文本分类 | 支持 | - | 支持 |
| | 命名实体 | 支持 | - | 支持 |
| | 文本三元组 | 支持 | - | 支持 |
| 视频 | 视频标注 | 支持 | - | - |
| 自由格式 | - | - | - | - |
| 表格 | - | - | - | - |

2 人工标注

2.1 创建标注作业

由于模型训练过程需要大量有标签的数据,因此在模型训练之前需对没有标签的数据添加标签。您可以通过创建单人标注作业或团队标注作业对数据进行手工标注,或对任务启动智能标注添加标签,快速完成对图片的标注操作,也可以对已标注图片修改或删除标签进行重新标注。

标注作业支持的数据类型

对于不同类型的数据集,用户可以选择不同的标注任务,当前ModelArts支持如下类型的标注任务。

- 图片
 - 图像分类:识别一张图片中是否包含某种物体。
 - 物体检测:识别出图片中每个物体的位置及类别。
 - 图像分割:根据图片中的物体划分出不同区域。
- 音频
 - 声音分类:对声音进行分类。
 - 语音内容:对语音内容进行标注。
 - 语音分割:对语音进行分段标注。
- 文本
 - 文本分类:对文本的内容按照标签进行分类处理。
 - 命名实体:针对文本中的实体片段进行标注,如"时间"、"地点"等。
 - 文本三元组:针对文本中的实体片段和实体之间的关系进行标注。
- 视频

视频标注:识别出视频中每个物体的位置及分类。目前仅支持mp4格式。

前提条件

在进行数据标注前,需要创建相应类型的数据集。具体步骤参考创建数据集。

操作步骤

- 1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理> 数据标注",进入 "数据标注"管理页面。
- 2. 单击"创建标注作业",进入"创建标注作业"页面,根据需求创建不同类型的标注作业。
 - a. 填写标注作业基本信息,标注作业的"名称"和"描述"。

图 2-1 标注作业基本信息



b. 根据您的需求,选择"标注场景"和"标注类型"。

图 2-2 选择标注场景和标注类型



- c. 针对不同类型的标注作业,需填写参数不同,请参考如下类型标注作业对应 的参数介绍。
 - 图片(图像分类、物体检测、图像分割)
 - 音频(声音分类、语音内容、语音分割)
 - 文本(文本分类、命名实体、文本三元组)
 - 视频
- d. 参数填写无误后,单击页面右下角"创建"。

标注作业创建完成后,系统自动跳转至数据标注管理页面,针对创建好的标 注作业,您可以执行智能标注、发布、修改和删除等操作。

图片(图像分类、物体检测、图像分割)

图 2-3 图像分类和物体检测类型的参数

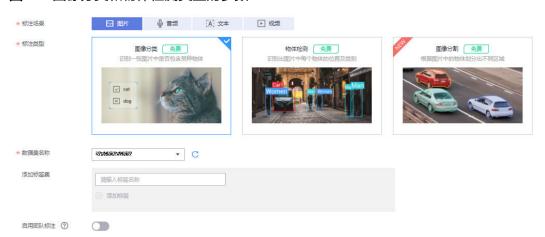


表 2-1 图片类型标注作业的详细参数

| 参数名称 | 说明 |
|--------|--|
| 数据集名称 | 选择支持当前标注类型的数据集。 |
| 添加标签集 | • 设置标签名称 :在标签名称文本框中,输入标签名称。长度为1~ 1024字符。 |
| | ● 添加标签: 单击"添加标签"可增加多个标签。 |
| | • 设置标签颜色 : "物体检测"和"图像分割"类型标注作业需设置 此参数。在每个标签右侧的标签颜色区域下,可在色板中选择颜 色,或者直接输入十六进制颜色码进行设置。 |
| | • 设置标签属性:针对"物体检测"类型标注作业,在设置完标签颜色后,可在右侧单击加号,增加对应的标签属性。标签属性用于区分同一标签物体的不同属性。例如,黄色小猫、黑色小猫。标签为cat,颜色为不同的标签属性。 |
| 启用团队标注 | 选择是否启用团队标注。图像分割暂不支持团队标注,当选择图像分 割类型时,界面不显示此参数。 |
| | 启用团队标注功能,需填写对应的团队标注任务"类型",同时选择对应的"标注团队"及参与标注的"团队成员"。参数详细介绍请参见创建团队标注任务。 |
| | 在启用"团队标注"前,需确保您已经在"标注团队"管理页面,添加相应的团队以及成员。如果没有标注团队,可直接从界面链接跳转至"标注团队"页面,添加您的团队并为其添加成员。详细指导请参见添加团队。 |
| | 启用团队标注功能的数据集,在创建完成后,可以在"标注类型"中 看到"团队标注"的标识。 |

音频(声音分类、语音内容、语音分割)

图 2-4 声音分类、语音内容、语音分割类型的参数



表 2-2 音频类型标注作业的详细参数

| 夜 2-2 自 <u>妙</u> 天空你注诉亚的许细多数 | | |
|-------------------------------------|--|--|
| 参数名称 | 说明 | |
| 数据集名称 | 选择支持当前标注类型的数据集。 | |
| 添加标签集(声音分类) | "声音分类"类型的标注作业可以添加标签集。 • 设置标签名称:在标签名称文本框中,长度为1~1024字符。 • 添加标签:单击"添加标签"可增加多个标签。 | |
| 标签管理(语音分割) | ● 单标签 单标签 单标签适用于一段音频标注只有一种类别的音频,通常标注一个标签。 - 设置标签名称:在"标签名"列输入标签名称。长度为1~1024字符。 - 设置标签颜色:在"标签颜色"列设置标签颜色。可在色板中选择颜色,或者直接输入十六进制颜色码进行设置。 ● 多标签 多标签 多标签 多标签适用于多维度标注,例如在一段音频标注噪音与人说话的声音两种类别,其中说话的声音还可以标注为不同人的声音。单击"新建标签类别"可添加多个标签类别,一个标签类别可以包含多个标签。"标签类别"和"标签名"只能是中文、字母、数字、英文句号、下划线或中划线组成的合法字符串。长度为1~256字符。 - 设置标签类别:在"标签类别"输入标签类别的名称。 - 设置标签名称:在"标签名"输入标签名称。 - 添加标签:单击"添加标签"可增加多个标签。 | |
| 启用语音内容标 注(语音分割) | 仅"语音分割"类型数据集支持设置,默认关闭。如果启用此功能,支持针对语音内容进行标注。 | |

| 参数名称 | 说明 |
|------------------|---|
| 启用团队标注 (语音分割) | 仅"语音分割"类型支持团队标注,因此选择创建语音分割类型时,支持设置是否启用团队标注。 |
| | 启用团队标注功能,需填写对应的团队标注任务"类型",同时 选择对应的"标注团队"及参与标注的"团队成员"。参数详细 介绍请参见 <mark>创建团队标注任务</mark> 。 |
| | 在启用"团队标注"前,需确保您已经在"标注团队"管理页面,添加相应的团队以及成员。如果没有标注团队,可直接从界面链接跳转至"标注团队"页面,添加您的团队并为其添加成员。详细指导请参见 <mark>添加团队</mark> 。 |
| | 启用团队标注功能的数据集,在创建完成后,可以在"标注类型"中看到"团队标注"的标识。 |

文本(文本分类、命名实体、文本三元组)

图 2-5 文本分类、命名实体、文本三元组类型的参数



表 2-3 文本类型标注作业的详细参数

| 参数名称 | 说明 |
|-------|-----------------|
| 数据集名称 | 选择支持当前标注类型的数据集。 |

说明 参数名称 添加标签集(文 ● 设置标签名称: 在标签名称文本框中,输入标签名称。长度为 本分类、命名实 1~1024字符。 体) • 添加标签: 单击"添加标签"可增加多个标签。 ● **设置标签颜色**:在每个标签右侧的标签颜色区域下,可在色板 中选择颜色,或者直接输入十六进制颜色码进行设置。 #3399ff 针对"文本三元组"类型的数据集,需要设置实体标签和关系标 添加标签集(文 本三元组) 签。 • 实体标签:需设置标签名以及标签颜色。可在颜色区域右侧单 击加号增加多个标签。 • 关系标签: 关系标签为两个实体之间的关系。需设置起始实体 和终止实体,您需要先添加至少2个实体标签后,再添加关系 标签。 实体标签 请输入标签名称 **▼** + 关系标签 请输入标签名称 起始实体 ▼ 终止实体 启用团队标注 选择是否启用团队标注。 启用团队标注功能,需填写对应的团队标注任务"类型",同时 选择对应的"标注团队"及参与标注的"团队成员"。参数详细 介绍请参见创建团队标注任务。 在启用"团队标注"前,需确保您已经在"标注团队"管理页 面,添加相应的团队以及成员。如果没有标注团队,可直接从界 面链接跳转至"标注团队"页面,添加您的团队并为其添加成 员。详细指导请参见添加团队。 启用团队标注功能的数据集,在创建完成后,可以在"标注类 型"中看到"团队标注"的标识。

视频

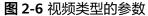




表 2-4 视频类型标注作业的详细参数

| 参数名称 | 说明 |
|-------|---|
| 数据集名称 | 选择支持当前标注类型的数据集。 |
| 添加标签集 | • 设置标签名称 :在标签名称文本框中,输入标签名称。长度为 1~1024字符。 |
| | • 添加标签: 单击"添加标签"可增加多个标签。 |
| | • 设置标签颜色 :在每个标签右侧的标签颜色区域下,可在色板中选择颜色,或者直接输入十六进制颜色码进行设置。 |

2.2 图片标注

2.2.1 图像分类

由于模型训练过程需要大量有标签的图片数据,因此在模型训练之前需对没有标签的 图片添加标签。您可以通过手工标注或智能一键标注的方式添加标签,快速完成对图 片的标注操作,也可以对已标注图片修改或删除标签进行重新标注。

针对图像分类场景,开始标注前,您需要了解:

- 图片标注支持多标签,即一张图片可添加多个标签。
- 标签名是由中文、大小写字母、数字、中划线或下划线组成,且不超过32位的字符串。

开始标注

1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理> 数据标注",进入 "数据标注"管理页面。

- 2. 在标注作业列表中,基于"标注类型"选择需要进行标注的标注作业,单击标注 作业名称进入标注作业标注详情页。
- 3. 在标注作业标注详情中,展示此标注作业下全部数据。

同步新数据

ModelArts会自动将数据集中新增的数据同步至标注作业,包含数据及当前标注作业支持的标注信息。

为了快速获取数据集中最新数据,可在标注作业详情页的"全部"、"未标注"或 "已标注"页签中,单击"同步新数据",快速将数据集中的数据添加到标注作业 中。

筛选数据

在标注作业详情页面,默认展示作业中全部数据,您可以在"全部"、"未标注"或页签下,在筛选条件区域,单击"展开",添加筛选条件,快速过滤出您想要查看的数据。

支持的筛选条件如下所示,您可以设置一个或多个选项进行筛选。

- 难例集:难例或非难例。
- 标签:您可以选择全部标签,或者基于您指定的标签,选中其中一个或多个。
- 样本创建时间:1个月内、1天内或自定义,如果选择自定义,可以在时间框中指定明确时间范围。
- 文件名或目录:根据文件名称或者文件存储目录筛选。
- 标注人:选择执行标注操作的帐号名称。
- 样本属性:表示自动分组生成的属性。只有启用了自动分组任务后才可使用此筛选条件。
- 数据属性: 暂不支持。

图 2-7 筛选条件



标注图片(手工标注)

在标注作业详情页中,展示了此数据集中"全部"、"未标注"和"已标注"的图片,默认显示"全部"的图片列表。单击图片,即可进行图片的预览,对于已标注图片,预览页面下方会显示该图片的标签信息。

1. 在"未标注"页签,勾选需进行标注的图片。

- 手工点选:在图片列表中,单击勾选图片左上角的选择框,进入选择模式,表示图片已勾选。可勾选同类别的多个图片,一起添加标签。
- 批量选中:如果图片列表的当前页,所有图片属于一种类型,可以在图片列表的右上角单击"选择当前页",则当前页面所有的图片将选中。
- 2. 为选中图片添加标签。
 - a. 在右侧的"添加标签"区域中,单击"标签名"右侧的文本框中设置标签。 单击"标签名"右侧的文本框,然后从下拉列表中选择已有的标签。如果已 有标签无法满足要求时,直接在文本框中添加新标签。
 - b. 单击"确定"。此时,选中的图片将被自动移动至"已标注"页签,且在 "未标注"和"全部"页签中,标签的信息也将随着标注步骤进行更新,如 增加的标签名称、各标签对应的图片数量。

图 2-8 添加标签



查看已标注图片

在标注任务详情页,单击"已标注"页签,您可以查看已完成标注的图片列表。图片缩略图下方默认呈现其对应的标签,您也可以勾选图片,在右侧的"选中文件标签"中了解当前图片的标签信息。

修改标注

当数据完成标注后,您还可以进入已标注页签,对已标注的数据进行修改。

• 基于图片修改

在标注作业详情页面,单击"已标注"页签,然后在图片列表中选中待修改的图片(选择一个或多个)。在右侧标签信息区域中对图片信息进行修改。

修改标签:在"选中文件标签"区域中,单击操作列的编辑图标,然后在文本框中输入正确的标签名,然后单击确定图标完成修改。

删除标签:在"选中文件标签"区域中,单击操作列的删除图标删除该标签。此操作仅删除选中图片中的标签。

图 2-9 编辑标签

选中文件标签



• 基于标签修改

在标注作业详情页,单击"标签管理"页签,显示全部标签列表。

- 修改标签:单击操作列的修改按钮,然后在弹出的对话框中输入修改后的标签名,然后单击"确定"完成修改。修改后,之前添加了此标签的图片,都将被标注为新的标签名称。
- 删除标签:单击操作列的删除按钮,之前添加了此标签的图片,都将删除此标签。

图 2-10 全部标签的信息



添加数据

除了同步数据集中的新数据外,您还可以在标注作业中,直接添加图片,用于数据标注。添加的数据将先导入至标注任务关联的数据集中,然后标注任务会自动同步数据集中最新的数据。

1. 在标注作业详情页面,单击"全部"、"已标注"或"未标注"页签,然后单击 左上角"添加数据",选择添加数据。

图 2-11 添加数据



2. 在弹出的导入对话框中,选择数据来源和导入方式,选择导入的数据路径和数据标注状态。

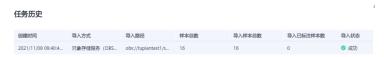
图 2-12 添加图片

▲ 注意: 您导入的数据将先导入至标注任务关联的数据集中,然后标注任务会自动同步数据集最新的数据。 ★ 数据来源 OBS 本地上传 ***** 导入方式 提前将需要导入的数据集文件放入您有权限访问的对象存储服务 (OBS) 目录中。 标注文件格式说明 * 导入路径 /a-test-obs/dxf/import/labels/tmp/ 未标注 日标注 取消 导入 ▲ 注意: 您导入的数据将先导入至标注任务关联的数据集中,然后标注任务会自动同步数据集最新的数据。 OBS 本地上传 ★ 上传数据存储路径 /a-test-obs/dxf/import/labels/tmp/ ① 文件上传 上传数据 未标注 巴标注

3. 在导入对话框中,单击"确定",完成添加数据的操作。 您添加的图片将自动呈现在"全部"的图片列表中,也可单击"添加数据>查看历史记录",进入"任务历史"界面,可查看相应的导入历史。

确定 取消

图 2-13 查看历史数据



删除图片

通过数据删除操作,可将需要丢弃的图片数据快速删除。

在"全部"、"未标注"或"已标注"页面中,依次选中需要删除的图片,或者选择"选择当前页"选中该页面所有图片,然后单击左上角"删除图片"。在弹出的对话框中,根据实际情况选择是否勾选"同时删除源文件",确认信息无误后,单击"确定"完成图片删除操作。

其中,被选中的图片,其左上角将显示为勾选状态。如果当前页面无选中图片时, "删除图片"按钮为灰色,无法执行删除操作。

□ 说明

如果勾选了"同时删除源文件",删除图片操作将删除对应OBS目录下存储的图片,此操作可能 会影响已使用此源文件的其他数据集或数据集版本,有可能导致展示异常或训练/推理异常。删 除后,数据将无法恢复,请谨慎操作。

2.2.2 物体检测

由于模型训练过程需要大量有标签的图片数据,因此在模型训练之前需对没有标签的 图片添加标签。您可以通过手工标注或智能一键标注的方式添加标签,快速完成对图 片的标注操作,也可以对已标注图片修改或删除标签进行重新标注。

针对物体检测场景,开始标注前,您需要了解:

- 图片中所有目标物体都要标注。
- 目标物体清晰无遮挡的,必须画框。
- 画框仅包含整个物体。框内包含整个物体的全部,画框边缘不可与待标注的物体的边缘轮廓相交,在此基础之上确保边缘和待标注物体间不要留着空隙,避免背景对模型训练造成干扰。

开始标注

- 1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理> 数据标注",进入 "数据标注"管理页面。
- 2. 在标注作业列表中,基于"标注类型"选择需要进行标注的标注作业,单击标注 作业名称进入标注作业标注详情页。
- 3. 在标注作业标注详情中,展示此标注作业下全部数据。

同步新数据

ModelArts会自动将数据集中新增的数据同步至标注作业,包含数据及当前标注作业支持的标注信息。

为了快速获取数据集中最新数据,可在标注作业详情页的"全部"、"未标注"或 "已标注"页签中,单击"同步新数据",快速将数据集中的数据添加到标注作业 中。

筛选数据

在标注作业详情页面,默认展示作业中全部数据,您可以在"全部"、"未标注"或页签下,在筛选条件区域,单击"展开",添加筛选条件,快速过滤出您想要查看的数据。

支持的筛选条件如下所示,您可以设置一个或多个选项进行筛选。

- 难例集:难例或非难例。
- 标签:您可以选择全部标签,或者基于您指定的标签,选中其中一个或多个。
- 样本创建时间:1个月内、1天内或自定义,如果选择自定义,可以在时间框中指定明确时间范围。
- 文件名或目录:根据文件名称或者文件存储目录筛选。
- 标注人:选择执行标注操作的帐号名称。
- 样本属性:表示自动分组生成的属性。只有启用了自动分组任务后才可使用此筛选条件。

数据属性: 暂不支持。

图 2-14 筛选条件



标注图片(手工标注)

标注作业详情页中,展示了此数据集中"全部""未标注"和"已标注"的图片,默认显示"全部"的图片列表。

- 1. 在"未标注"页签图片列表中,单击图片,自动跳转到标注页面。在标注页面, 常用按钮的使用可参见表2-6。
- 2. 在页面上方工具栏选择合适的标注图形,系统默认的标注图形为矩形。本示例使用矩形工具进行标注。

□ 说明

页面左侧可以选择多种形状对图片进行标注。标注第一张图片时,一旦选择其中一种,其 他图片默认使用此形状进行标注,用户可以根据自己需求再进行切换。

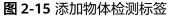
表 2-5 支持的标注框

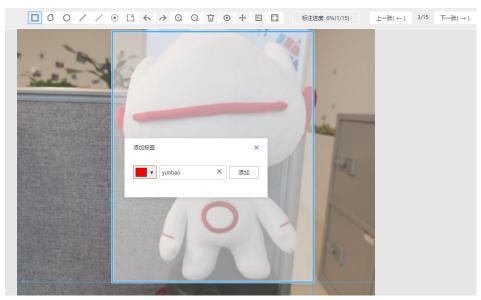
| 图标 | 使用说明 |
|-------|---|
| | 矩形。鼠标单击标注对象左上角边缘位置,界面将出现矩形 框,移动鼠标使得矩形框覆盖标注对象,然后单击完成标 注。 |
| O | 多边形。在标注对象所在范围内,鼠标左键单击完成一个点的标注,沿着物体的形状边缘,通过鼠标指定多个点,最终单击到第一个点的位置,由所有的点组成一个多边形形状。 使得需标注的对象在此标注框内。 |
| 0 | 圆形。在标注对象中,选择物体的中心点位置,单击鼠标确 定圆心,然后移动鼠标,使得圆形框覆盖标注对象,然后再 单击鼠标完成标注。 |
| / | 直线。在标注对象中,选择物体的起始点,单击鼠标确定直 线的起始点,然后使得直线覆盖标注对象,然后再单击鼠标 完成标注。 |
| are e | 虚线。在标注对象中,选择物体的起始点,单击鼠标确定虚 线的起始点,然后使得虚线覆盖标注对象,然后再单击鼠标 完成标注。 |

| 图标 | 使用说明 |
|----|--------------------------|
| • | 点。单击图片中的物体所在位置,即可完成点的标注。 |

3. 在弹出的添加标签文本框中,直接输入新的标签名,在文本框前面选中标签颜色,然后单击"添加"。如果已存在标签,从下拉列表中选择已有的标签,单击"添加"。

逐步标注图片中所有物体所在位置,一张图片可添加多个标签。完成一张图片标注后,可单击图片下方图片列表,快速选中其他未标注的图片,然后在标注页面中执行标注操作。





4. 单击页面上方"返回数据标注预览"查看标注信息,在弹框中单击"确定"保存 当前标注并离开标注页面。

选中的图片被自动移动至"已标注"页签,且在"未标注"和"全部"页签中,标签的信息也将随着标注步骤进行更新,如增加的标签名称、标签对应的图片数量。

表 2-6 标注界面的常用按钮

| 按钮图标 | 功能说明 |
|---------------|----------------|
| ~ | 撤销上一个操作。 |
| \rightarrow | 重做上一个操作。 |
| Q | 放大图片。 |
| Q | 缩小图片。 |
| 亡 | 删除当前图片中的所有标注框。 |

| 按钮图标 | 功能说明 |
|----------|---|
| © | 显示或隐藏标注框。只有在已标注图片中可使用此操作。 |
| 4 | 拖动,可将标注好的框拖动至其他位置,也可以选择框的边缘, 更改框的大小。 |
| Ð | 复位,与上方拖动为同组操作,当执行了拖动后,可以单击复位 按钮快速将标注框恢复为拖动前的形状和位置。 |
| | 全屏显示标注的图片。 |

查看已标注图片

在标注作业详情页,单击"已标注"页签,您可以查看已完成标注的图片列表。单击图片,可在右侧的"当前文件标签"中了解当前图片的标签信息。

修改标注

当数据完成标注后,您还可以进入已标注页签,对已标注的数据进行修改。

• 基于图片修改

在标注作业详情页面,单击"标注"页签,选择"已标注"页签,然后在图片列表中选中待修改的图片,单击图片跳转到标注页面,在右侧标签信息区域中对图片信息进行修改。

- 修改标签: "标注"区域中,单击编辑图标,在文本框中输入正确的标签名,然后单击确定图标完成修改。也可以单击标签,在图片标注区域,调整标注框的位置和大小,完成调整后,单击其他标签即可保存修改。
- 删除标签:在"标注"区域中,单击删除图标即可删除此图片中的标签。标签删除后,单击页面左上角的"返回数据标注预览"离开标注页面,在弹出对话框中保存标注。图标的标签全部删除后,该图片会重新回到"未标注"页签。

图 2-16 编辑物体检测标签



• 基于标签修改

在标注作业详情页中,单击"标签管理"页签,在标签管理详情页,显示全部标 签的信息。

- 修改标签:单击操作列的"修改",然后在弹出的对话框中输入修改后的标签和颜色,然后单击"确定"完成修改。修改后,之前添加了此标签的图片,都将被标注为新的标签名称。
- 删除标签:单击操作列的"删除",或选中需要删除的标签名称,单击上方的"删除"可删除此标签。

图 2-17 物体检测的全部标签





添加数据

除了同步数据集中的新数据外,您还可以在标注作业中,直接添加图片,用于数据标注。添加的数据将先导入至标注任务关联的数据集中,然后标注任务会自动同步数据 集中最新的数据。

1. 在标注作业详情页面,单击"全部"、"已标注"或"未标注"页签,然后单击 左上角"添加数据",选择添加数据。

图 2-18 添加数据



2. 在弹出的导入对话框中,选择数据来源和导入方式,选择导入的数据路径和数据标注状态。

图 2-19 添加图片

导入

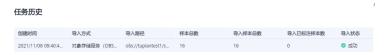




3. 在导入对话框中,单击"确定",完成添加数据的操作。

您添加的图片将自动呈现在"全部"的图片列表中,也可单击"添加数据>查看历史记录",进入"任务历史"界面,可查看相应的导入历史。

图 2-20 查看历史数据



删除图片

通过数据删除操作, 可将需要丢弃的图片数据快速删除。

在"全部"、"未标注"或"已标注"页面中,依次选中需要删除的图片,或者选择"选择当前页"选中该页面所有图片,然后单击左上角"删除图片"。在弹出的对话框中,根据实际情况选择是否勾选"同时删除源文件",确认信息无误后,单击"确定"完成图片删除操作。

其中,被选中的图片,其左上角将显示为勾选状态。如果当前页面无选中图片时, "删除图片"按钮为灰色,无法执行删除操作。

□ 说明

如果勾选了"同时删除源文件",删除图片操作将删除对应OBS目录下存储的图片,此操作可能 会影响已使用此源文件的其他数据集或数据集版本,有可能导致展示异常或训练/推理异常。删 除后,数据将无法恢复,请谨慎操作。

2.2.3 图像分割

由于模型训练过程需要大量有标签的图片数据,因此在模型训练之前需对没有标签的 图片添加标签。您可以通过在ModelArts控制台进行标注,也可以对已标注图片修改或 删除标签进行重新标注。

针对图像分割场景,开始标注前,您需要了解:

- 图片中需要提取轮廓的物体都要标注。
- 支持使用多边形标注和极点标注。
 - 多边形标注,根据目标物体的轮廓绘制多边形。
 - 极点标注,在目标物体轮廓的最上、最左、最下、最右的位置分别标注四个极点,极点要在物体的轮廓上。系统将根据标注的极点推理出物体的轮廓。对于背景比较复杂的图片,极点标注效果不佳,推荐使用多边形标注。

门 说明

极点标注功能仅支持北京一与北京四区域。

● 多边形标注时,标注框或极点,必须在图片范围内,超出图片将导致后续作业异常。

开始标注

- 1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理> 数据标注",进入 "数据标注"管理页面。
- 2. 在标注作业列表中,基于"标注类型"选择需要进行标注的标注作业,单击标注 作业名称进入标注作业标注详情页。
- 3. 在标注作业标注详情中,展示此标注作业下全部数据。

同步新数据

ModelArts会自动将数据集中新增的数据同步至标注作业,包含数据及当前标注作业支持的标注信息。

为了快速获取数据集中最新数据,可在标注作业详情页的"全部"、"未标注"或 "已标注"页签中,单击"同步新数据",快速将数据集中的数据添加到标注作业中。

筛选数据

在标注作业详情页面,默认展示作业中全部数据,您可以在"全部"、"未标注"或页签下,在筛选条件区域,单击"展开",添加筛选条件,快速过滤出您想要查看的数据。

支持的筛选条件如下所示,您可以设置一个或多个选项进行筛选。

- 难例集:难例或非难例。
- 标签: 您可以选择全部标签,或者基于您指定的标签,选中其中一个或多个。
- 样本创建时间:1个月内、1天内或自定义,如果选择自定义,可以在时间框中指定明确时间范围。
- 文件名或目录:根据文件名称或者文件存储目录筛选。
- 标注人:选择执行标注操作的帐号名称。
- 样本属性:表示自动分组生成的属性。只有启用了自动分组任务后才可使用此筛 选条件。
- 数据属性: 暂不支持。

图 2-21 筛选条件



标注图片(手工标注)

标注作业详情页中,展示了此标注作业中"未标注"和"已标注"的图片,默认显示"全部"的图片列表。

- 1. 在"未标注"页签图片列表中,单击图片,自动跳转到标注页面。在标注页面, 常用按钮的使用可参见表2-8。
- 2. 选择标注方式。

在标注页面,上方工具栏提供了常用的<mark>标注方式及常用按钮</mark>,系统默认的标注方式为多边形标注。选择多边形标注或极点标注。

□ 说明

标注第一张图片时,一旦选择其中一种,其他所有图片都需要使用此方式进行标注。

图 2-22 工具栏



表 2-7 标注方式

| 图标 | 使用说明 |
|-------|---|
| Q | 多边形。在标注对象所在范围内,鼠标左键单击完成一个点的标注,沿着物体的形状边缘,通过鼠标指定多个点,最终单击到第一个点的位置,由所有的点组成一个多边形形状。 使得需标注的对象在此标注框内。 |
| • • • | 极点标注。在目标物体轮廓的最上、最左、最下、最右的位 置分别标注四个极点,极点要在物体的轮廓上。系统将根据 标注的极点推理出物体的轮廓。 |

表 2-8 工具栏常用按钮

| 按钮图标 | 功能说明 |
|---------------|---------------------------|
| < | 撤销上一个操作。 |
| \rightarrow | 重做上一个操作。 |
| Q | 放大图片。 |
| Q | 缩小图片。 |
| Ū | 删除当前图片中的所有标注框。 |
| Ø | 显示或隐藏标注框。只有在已标注图片中可使用此操作。 |

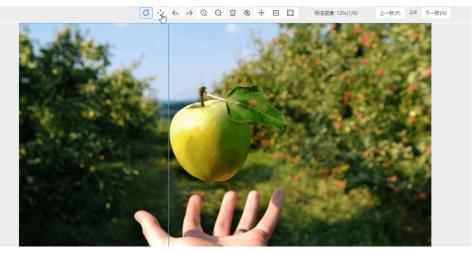
| 按钮图标 | 功能说明 |
|------|---|
| 4 | 拖动,可将标注好的框拖动至其他位置,也可以选择框的边 缘,更改框的大小。 |
| Ð | 复位,与上方拖动为同组操作,当执行了拖动后,可以单击 复位按钮快速将标注框恢复为拖动前的形状和位置。 |
| | 全屏显示标注的图片。 |

3. 标注物体。

以极点标注为例。识别图片中的物体,单击左键分别定位物体的最上、最左、最 下、最右的位置点。确定位置后,将弹出对话框,填入标签名称,单击确定添加 物体的标签。确定后系统将自动推理出物体的轮廓。

完成一张图片标注后,可单击图片下方图片列表,快速选中其他未标注的图片, 然后在标注页面中执行标注操作。





4. 单击页面上方"返回数据标注预览"查看标注信息,在弹框中单击"确定"保存 当前标注并离开标注页面。

选中的图片被自动移动至"已标注"页签,且在"未标注"和"全部"页签中, 标签的信息也将随着标注步骤进行更新,如增加的标签名称、标签对应的图片数 量。

查看已标注图片

在标注作业详情页中,单击"已标注"页签,您可以查看已完成标注的图片列表。单 击图片,可在右侧的"当前文件标签"中了解当前图片的标签信息。

修改标注信息

当数据完成标注后,您还可以进入已标注页签,对已标注的数据进行修改。

在数据标注详情页面,单击"已标注"页签,然后在图片列表中选中待修改的图片, 单击图片跳转到标注页面,在右侧标签信息区域中单击此图片已添加的标注信息。

- 修改标签: "标注"区域中,单击编辑图标,在弹出框中输入正确的标签名或标签颜色,然后单击"确定"完成修改。也可以单击标签,在图片标注区域,调整标注框的位置和大小,完成调整后,单击其他标签即可保存修改。
- 修改图片标注信息:在图片展示区,显示物体边缘,可单击蓝色圆点,将标注框 调整至物体边缘。
- 删除标签:在"标注"区域中,单击删除图标即可删除此图片中的标签。图片的标签全部删除后,该图片会重新回到"未标注"页签。

标注标注信息修改后,单击页面左上角的"返回数据标注预览"离开标注页面,在弹出对话框中单击"确定"保存修改。

图 2-24 编辑标注信息



添加数据

除了同步数据集中的新数据外,您还可以在标注作业中,直接添加图片,用于数据标注。添加的数据将先导入至标注任务关联的数据集中,然后标注任务会自动同步数据 集中最新的数据。

1. 在标注作业详情页面,单击"全部"、"已标注"或"未标注"页签,然后单击 左上角"添加数据",选择添加数据。

图 2-25 添加数据



在弹出的导入对话框中,选择数据来源和导入方式,选择导入的数据路径和数据标注状态。

图 2-26 添加图片

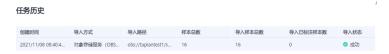
导入



3. 在导入对话框中,单击"确定",完成添加数据的操作。

您添加的图片将自动呈现在"全部"的图片列表中,也可单击"添加数据>查看历史记录",进入"任务历史"界面,可查看相应的导入历史。

图 2-27 查看历史数据



删除图片

通过数据删除操作,可将需要丢弃的图片数据快速删除。

在"全部"、"未标注"或"已标注"页面中,依次选中需要删除的图片,或者选择"选择当前页"选中该页面所有图片,然后单击左上角"删除图片"。在弹出的对话框中,根据实际情况选择是否勾选"同时删除源文件",确认信息无误后,单击"确定"完成图片删除操作。

其中,被选中的图片,其左上角将显示为勾选状态。如果当前页面无选中图片时, "删除图片"按钮为灰色,无法执行删除操作。

□ 说明

如果勾选了"同时删除源文件",删除图片操作将删除对应OBS目录下存储的图片,此操作可能会影响已使用此源文件的其他数据集或数据集版本,有可能导致展示异常或训练/推理异常。删除后,数据将无法恢复,请谨慎操作。

2.3 文本标注

2.3.1 文本分类

由于模型训练过程需要大量有标签的数据,因此在模型训练之前需对没有标签的文本添加标签。您也可以对已标注文本进行修改、删除和重新标注。

针对文本分类场景,是对文本的内容按照标签进行分类处理,开始标注前,您需要了解:

- 文本标注支持多标签,即一个标注对象可添加多个标签。
- 标签名是由中文、大小写字母、数字、中划线或下划线组成,且不超过32位的字符串。

开始标注

- 1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理> 数据标注",进入 "数据标注"管理页面。
- 2. 在标注作业列表中,基于"标注类型"选择需要进行标注的标注作业,单击标注 作业名称进入标注作业标注详情页。
- 3. 在标注作业标注详情中,展示此标注作业下全部数据。

同步新数据

ModelArts会自动将数据集中新增的数据同步至标注作业,包含数据及当前标注作业支持的标注信息。

为了快速获取数据集中最新数据,可在标注作业详情页的"未标注"页签中,单击"同步新数据",快速将数据集中的数据添加到标注作业中。

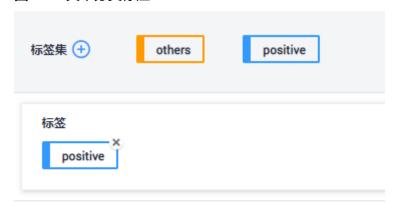
标注文本

标注作业详情页中,展示了此标注作业中"未标注"和"已标注"的文本,默认显示"未标注"的文本列表。

1. 在"未标注"页签文本列表中,页面左侧罗列"标注对象列表"。在列表中单击 需标注的文本对象,选择右侧"标签集"中的标签进行标注。一个标注对象可添 加多个标签。

以此类推,不断选中标注对象,并为其添加标签。

图 2-28 文本分类标注



标注对象文本内容:

Chinese test_1 for classification.

2. 当所有的标注对象都已完成标注,单击页面下方"保存当前页",完成"未标注"列表的文本标注。

添加标签

• 在"未标注"页签添加:单击页面中标签集右侧的加号,然后在弹出的"新增标签"页中,添加标签名称,选择标签颜色,单击"确定"完成标签的新增。

图 2-29 添加标签(1)



• 在"已标注"页签添加:在右侧单击页面中全部标签加号,然后在弹出的"新增标签"页中,添加标签名称,选择标签颜色,单击"确定"完成标签的新增。

图 2-30 添加标签(2)



图 2-31 新增标签

新增标签



查看已标注文本

在标注作业详情页,单击"已标注"页签,您可以查看已完成标注的文本列表。您也可以在右侧的"全部标签"中了解当前标注作业支持的所有标签信息。

修改标注

当数据完成标注后,您还可以进入已标注页签,对已标注的数据进行修改。

• 基于文本修改

在标注作业详情页,单击"已标注"页签,然后在文本列表中选中待修改的文本。

在文本列表中,单击文本,当文本背景变为蓝色时,表示已选择。当文本有多个标签时,可以单击文本标签上方的 x 删除单个标签。

• 基于标签修改

在标注作业详情页,单击"已标注"页签,在图片列表右侧,显示全部标签的信息。

- 批量修改:在"全部标签"区域中,单击操作列的编辑图标,然后在文本框中修改标签名称,选择标签颜色,单击"确定"完成修改。
- 批量删除:在"全部标签"区域中,单击操作列的删除图标,在弹出对话框中,可选择"仅删除标签"或"删除标签及仅包含此标签的标注对象",然后单击"确定"。

添加文件

除了同步新数据外,您还可以在标注详情页面中,直接添加数据,用于数据标注。

- 1. 在标注作业详情页面,单击"未标注"页签,然后单击左上角"添加数据"。
- 2. 在弹出的导入对话框中,选择数据来源、导入方式、导入路径等参数,导入数据。单击确定。

导入数据的详细操作介绍请参见导入操作。

图 2-32 导入数据



删除文件

通过数据删除操作,可将需要丢弃的文件数据快速删除。

- 在"未标注"页面中,单击选中需要删除的文本对象,然后单击左上角"删除",即可完成文本的删除操作。
- 在"已标注"页面中,选中待删除的文本对象,然后单击"删除",删除单个文本。或者选择"选择当前页"选中该页面所有文本,然后单击左上角"删除",即可完成当前页所有文本的删除操作。

其中,被选中的文本,其背景将显示为蓝色。

2.3.2 命名实体

命名实体场景,是针对文本中的实体片段进行标注,如"时间"、"地点"等。开始标注前,您需要了解:

● 实体命名标签名是由中文、大小写字母、数字、中划线或下划线组成,且不超过 32位的字符串。

开始标注

- 1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理> 数据标注",进入 "数据标注"管理页面。
- 2. 在标注作业列表中,基于"标注类型"选择需要进行标注的标注作业,单击标注 作业名称进入标注作业标注详情页。
- 3. 在标注作业标注详情中,展示此标注作业下全部数据。

同步新数据

ModelArts会自动将数据集中新增的数据同步至标注作业,包含数据及当前标注作业支持的标注信息。

为了快速获取数据集中最新数据,可在标注作业详情页的"未标注"页签中,单击"同步新数据",快速将数据集中的数据添加到标注作业中。

标注文本

标注作业详情页中,展示了此标注作业中"未标注"和"已标注"的文本,默认显示 "未标注"的文本列表。

1. 在"未标注"页签文本列表中,页面左侧罗列"标注对象列表"。在列表中单击需标注的文本对象,在右侧标签集下显示的文本内容中选中需要标注的部分,然后选择右侧"标签集"中的标签进行标注。一个标注对象可添加多个标签。以此类推,不断选中标注对象,并为其添加标签。

图 2-33 命名实体标注



2. 单击页面下方"保存当前页"完成文本标注。

添加标签

• 在"未标注"页签添加:单击页面中标签集右侧的加号,然后在弹出的"新增标签"页中,添加标签名称,选择标签颜色,单击"确定"完成标签的新增。

图 2-34 添加命名实体标签(1)



• 在"已标注"页签添加:在右侧单击页面中全部标签加号,然后在弹出的"新增标签"页中,添加标签名称,选择标签颜色,单击"确定"完成标签的新增。

图 2-35 添加命名实体标签(2)

全部标签 🕀

| 数量 操作 | 対号 | 3 | 2 立 | 地点 | 5 | 2 立 | 时间 | 5 | 2 立 | 人名 | 5 | 2 立 |

图 2-36 新增命名实体标签

新增标签



查看已标注文本

在数据集详情页,单击"已标注"页签,您可以查看已完成标注的文本列表。您也可以在右侧的"全部标签"中了解当前数据集支持的所有标签信息。

修改标注

当数据完成标注后,您还可以进入"已标注"页签,对已标注的数据进行修改。

在数据集详情页,单击"已标注"页签,在右侧标签信息区域中对文本信息进行修 改。

• 基于文本修改

在数据集详情页,单击"已标注"页签,然后在文本列表中选中待修改的文本。 手工点选删除:在文本列表中,单击文本,当文本背景变为蓝色时,表示已选择。在页面右侧,单击文本标签上方的、删除单个标签。

• 基于标签修改

在数据集详情页,单击"已标注"页签,在图片列表右侧,显示全部标签的信息。

- 批量修改:在"全部标签"区域中,单击操作列的编辑按钮,然后在文本框中添加标签名称,选择标签颜色,单击"确定"完成修改。
- 批量删除:在"全部标签"区域中,单击操作列的删除按钮,在弹出对话框中,可选择"仅删除标签"或"删除标签及仅包含此标签的标注对象",然后单击"确定"。

添加文件

除了同步新数据外,您还可以在标注详情页面中,直接添加数据,用于数据标注。

- 1. 在标注作业详情页面,单击"未标注"页签,然后单击左上角"添加数据"。
- 2. 在弹出的导入对话框中,选择数据来源、导入方式、导入路径等参数,导入数据。单击确定。

导入数据的详细操作介绍请参见导入操作。

图 2-37 导入数据



删除文件

通过数据删除操作,可将需要丢弃的文件数据快速删除。

- 在"未标注"页面中,单击选中需要删除的文本对象,然后单击左上角"删除",即可完成文本的删除操作。
- 在"已标注"页面中,选中待删除的文本对象,然后单击"删除",删除单个文本。或者选择"选择当前页"选中该页面所有文本,然后单击左上角"删除",即可完成当前页所有文本的删除操作。

其中,被选中的文本,其背景将显示为蓝色。

2.3.3 文本三元组

三元组标注适用于标注出语句当中形如(主语/Subject,谓词/Predicate,宾语/Object)结构化知识的场景,标注时不但可以标注出语句当中的实体,还可以标注出实体之间的关系,其在依存句法分析、信息抽取等自然语言处理任务中经常用到。

文本三元组类型的数据标注,需要关注两种标签,"实体标签"和"关系标签"。 "关系标签"需设置对应的"起始实体"和"终止实体"。

- 支持设置多个"实体标签"和"关系标签"。一个文本数据中,也可以标注多个 "实体标签"和"关系标签"
- 创建数据集时定义的"实体标签",不支持删除。

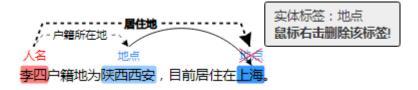
注意事项

在开始标注之前,需确保标注作业对应的"实体标签"和"关系标签"已定义好。 "关系标签"需设置对应的"起始实体"和"终止实体"。"关系标签"只能添加至 其设置好的"起始实体"和"终止实体"之间。 例如,如<mark>图2-38</mark>所示,当两个文本都被标注为"地点",那么针对这两个实体,无法添加本示例中的任意一个关系标签。当无法添加某个关系标签时,界面将显示一个红色的叉号,如图2-39所示。

图 2-38 实体标签和关系标签的示例



图 2-39 无法添加关系标签



开始标注

- 1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理> 数据标注",进入 "数据标注"管理页面。
- 2. 在标注作业列表中,基于"标注类型"选择需要进行标注的标注作业,单击标注 作业名称进入标注作业标注详情页。
- 3. 在标注作业标注详情中,展示此标注作业下全部数据。

同步新数据

ModelArts会自动将数据集中新增的数据同步至标注作业,包含数据及当前标注作业支持的标注信息。

为了快速获取数据集中最新数据,可在标注作业详情页的"未标注"页签中,单击"同步新数据",快速将数据集中的数据添加到标注作业中。

标注文本

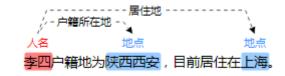
标注作业详情页中,展示了此标注作业中"未标注"和"已标注"的文本,默认显示 "未标注"的文本列表。 1. 在"未标注"页签文本列表中,页面左侧罗列"标注对象列表"。在列表中单击需标注的文本对象,选中相应文本内容,在页面呈现的实体类型列表中选择实体名称,完成实体标注。

图 2-40 实体标注



2. 在完成多个实体标注后,鼠标左键依次单击起始实体和终止实体,在呈现的关系 类型列表中选择一个对应的关系类型,完成关系标注。

图 2-41 关系标注



3. 当所有的标注对象都已完成标注,单击页面下方"保存当前页"完成"未标注" 列表的文本标注。

□ 说明

"文本三元组"类型的数据集,不支持在标注页面修改标签,需要进入"标签管理"页面,修改 "实体标签"和"关系标签"。

修改标注

当数据完成标注后,您还可以进入已标注页签,对已标注的数据进行修改。

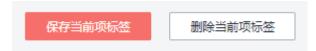
在标注作业详情页,单击"已标注"页签,在左侧文本列表中选中一行文本,右侧区域显示具体的标注信息。将鼠标移动至对应的实体标签或关系类型,单击鼠标右键,可删除此标注。单击鼠标左键,依次单击连接起始实体和终止实体,可增加关系类型,增加关系标注。

图 2-42 在文本中修改标签



您也可以在单击页面下方的"删除当前项标签"按钮,删除选中文本对象中的所有标签。

图 2-43 删除当前项标签



添加文件

除了同步新数据外,您还可以在标注详情页面中,直接添加数据,用于数据标注。

- 1. 在标注作业详情页面,单击"未标注"页签,然后单击左上角"添加数据"。
- 2. 在弹出的导入对话框中,选择数据来源、导入方式、导入路径等参数,导入数据。单击确定。

导入数据的详细操作介绍请参见导入操作。

图 2-44 导入数据



删除文件

通过数据删除操作,可将需要丢弃的文件数据快速删除。

- 在"未标注"页面中,单击选中需要删除的文本,然后单击左上角"删除",即可完成文本的删除操作。
- 在"已标注"页面中,选中待删除的文本,然后单击"删除",删除单个文本。 或者勾选"选择当前页"选中该页面所有文本,然后单击左上角"删除",即可 完成当前页所有文本的删除操作。

其中,被选中的文本,其背景将显示为蓝色。如果当前页面无选中文本时,"删除"按钮为灰色,无法执行删除操作。

2.4 音频标注

2.4.1 声音分类

由于模型训练过程需要大量有标签的音频数据,因此在模型训练之前需对没有标签的音频添加标签。通过ModelArts您可对音频进行一键式批量添加标签,快速完成对音频的标注操作,也可以对已标注音频修改或删除标签进行重新标注。

开始标注

- 1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理> 数据标注",进入 "数据标注"管理页面。
- 2. 在标注作业列表中,基于"标注类型"选择需要进行标注的标注作业,单击标注 作业名称进入标注作业标注详情页。
- 3. 在标注作业标注详情中,展示此标注作业下全部数据。

同步新数据

ModelArts会自动将数据集中新增的数据同步至标注作业,包含数据及当前标注作业支持的标注信息。

为了快速获取数据集中最新数据,可在标注作业详情页的"未标注"和"已标注"页签中,单击"同步新数据",快速将数据集中的数据添加到标注作业中。

标注音频

标注作业详情页中,展示了此标注作业中"未标注"和"已标注"的音频,默认显示 "未标注"的音频列表。单击音频左侧 ,即可进行音频的试听。

- 1. 在"未标注"页签,勾选需进行标注的音频。
 - 手工点选:在音频列表中,单击音频,当右上角出现蓝色勾选框时,表示已 勾选。可勾选同类别的多个音频,一起添加标签。
 - 批量选中:如果音频列表的当前页,所有音频属于一种类型,可以在列表的右上角单击"选择当前页",则当前页面所有的音频将选中。
- 2. 添加标签。
 - a. 在右侧的"标签"区域中,单击"标签"下侧的文本框中设置标签。

方式一(已存在标签): 单击"标签"下方的文本框,在快捷键下拉列表中选择快捷键,然后在标签文本输入框中选择已有的标签名称,然后单击"确定"。

方式二(新增标签):在"标签"下方的文本框中,在快捷键下拉列表中选择快捷键,然后在标签文本输入框中输入新的标签名称,然后单击"确定"。

b. 选中的音频将被自动移动至"已标注"页签,且在"未标注"页签中,标签的信息也将随着标注步骤进行更新,如增加的标签名称、各标签对应的音频数量。

□ 说明

快捷键的使用说明: 为标签指定快捷键后,当您选择一段音频后,在键盘中按一下快捷键,即可为此音频增加为此快捷键对应的标签。例如"aa"标签对应的快捷键是"1",在数据标注过程中,选中1个或多个文件,按"1",界面将提示是否需要将此文件标注为"aa"标签,单击确认即可完成标注。

快捷键对应的是标签,1个标签对应1个快捷键。不同的标签,不能指定为同一个快捷键。 快捷键的使用,可以大大提升标注效率。

图 2-45 添加音频标签



查看已标注音频

在标注作业详情页,单击"已标注"页签,您可以查看已完成标注的音频列表。单击音频,可在右侧的"选中文件标签"中了解当前音频的标签信息。

修改标注

当数据完成标注后,您还可以进入"已标注"页签,对已标注的数据进行修改。

• 基于音频修改

在标注作业详情页面,单击"已标注"页签,然后在音频列表中选中待修改的音频(选择一个或多个)。在右侧标签信息区域中对标签进行修改。

- 修改标签:在"选中文件标签"区域中,单击操作列的编辑图标,然后在文本框中输入正确的标签名,然后单击确定图标完成修改。
- 删除标签:在"选中文件标签"区域中,单击操作列的删除图标删除该标签。

• 基于标签修改

在标注作业详情页面,单击"已标注"页签,在音频列表右侧,显示全部标签的信息。

图 2-46 全部标签信息

全部标签 1



修改标签:单击操作列的编辑图标,然后在弹出的对话框中输入修改后的标签名,然后单击"确定"完成修改。修改后,之前添加了此标签的音频,都将被标注为新的标签名称。

- 删除标签:单击操作列的删除图标,在弹出的对话框中,根据提示框选择需要删除的对象,然后单击"确定"完成删除。

添加音频

除了同步新数据外,您还可以在标注详情页面中,直接添加数据,用于数据标注。

- 1. 在标注作业详情页面,单击"未标注"页签,然后单击左上角"添加数据"。
- 2. 在弹出的导入对话框中,选择数据来源、导入方式、导入路径等参数,导入数据。单击确定。

导入数据的详细操作介绍请参见导入操作。

图 2-47 导入数据



删除音频

通过数据删除操作,可将需要丢弃的音频数据快速删除。

在"未标注"或"已标注"页面中,选中需要删除的音频,或者选择"选择当前页"选中该页面所有音频,然后单击左上角"删除音频",在弹出的对话框中,根据实际情况选择是否勾选"同时删除源文件",确认信息无误后,单击"确定"完成音频删除操作。

其中,被选中的音频,其右上角将显示为勾选状态。如果当前页面无选中音频时, "删除音频"按钮为灰色,无法执行删除操作。

□ 说明

如果勾选了"同时删除源文件",删除音频操作是将删除对应OBS目录下存储的音频。此操作可能会影响已使用此源文件的其他数据集或数据集版本,有可能导致展示异常或训练/推理异常。删除后,数据将无法恢复,请谨慎操作。

2.4.2 语音内容

由于模型训练过程需要大量有标签的音频数据,因此在模型训练之前需对没有标签的 音频添加标签。通过ModelArts您可对音频进行一键式批量添加标签,快速完成对音频 的标注操作,也可以对已标注音频修改或删除标签进行重新标注。

开始标注

- 1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理> 数据标注",进入 "数据标注"管理页面。
- 2. 在标注作业列表中,基于"标注类型"选择需要进行标注的标注作业,单击标注 作业名称进入标注作业标注详情页。
- 3. 在标注作业标注详情中,展示此标注作业下全部数据。

同步新数据

ModelArts会自动将数据集中新增的数据同步至标注作业,包含数据及当前标注作业支持的标注信息。

为了快速获取数据集中最新数据,可在标注作业详情页的"未标注"页签中,单击"同步新数据",快速将数据集中的数据添加到标注作业中。

标注音频

标注作业详情页中,展示了此数据集中"未标注"和"已标注"的音频,默认显示 "未标注"的音频列表。

- 1. 在"未标注"页签左侧音频列表中,单击目标音频文件,在右侧的区域中出现音频,单击音频下方。 即可进行音频播放。
- 2. 根据播放内容,在下方"语音内容"文本框中填写音频内容。
- 3. 输入内容后单击下方的"确认标注"按钮完成标注。音频将被自动移动至"已标注"页签。

图 2-48 语音内容音频标注



查看已标注音频

在标注作业详情页,单击"已标注"页签,您可以查看已完成标注的音频列表。单击音频,可在右侧的"语音内容"文本框中了解当前音频的内容信息。

修改标注

当数据完成标注后,您还可以进入"已标注"页签,对已标注的数据进行修改。

在标注作业详情页,单击"已标注"页签,然后在音频列表中选中待修改的音频。在 右侧标签信息区域中修改"语音内容"文本框中的内容,单击下方的"确认标注"按 钮完成修改。

添加音频

除了同步新数据外,您还可以在标注详情页面中,直接添加数据,用于数据标注。

- 1. 在标注作业详情页面,单击"未标注"页签,然后单击左上角"添加数据"。
- 2. 在弹出的导入对话框中,选择数据来源、导入方式、导入路径等参数,导入数据。单击确定。

导入数据的详细操作介绍请参见导入操作。





删除音频

通过数据删除操作,可将需要丢弃的音频数据快速删除。

在"未标注"或"已标注"页面中,选中需要删除的音频,然后单击左上角"删除音频",在弹出的对话框中,根据实际情况选择是否勾选"同时删除源文件",确认信息无误后,单击"确定"完成音频删除操作。

□ 说明

如果勾选了"同时删除源文件",删除音频操作是将删除对应OBS目录下存储的音频。此操作可能会影响已使用此源文件的其他数据集或数据集版本,有可能导致展示异常或训练/推理异常。删除后,数据将无法恢复,请谨慎操作。

2.4.3 语音分割

由于模型训练过程需要大量有标签的音频数据,因此在模型训练之前需对没有标签的音频添加标签。通过ModelArts您可对音频添加标签,快速完成对音频的标注操作,也可以对已标注音频修改或删除标签进行重新标注。

开始标注

- 1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理> 数据标注",进入 "数据标注"管理页面。
- 2. 在标注作业列表中,基于"标注类型"选择需要进行标注的标注作业,单击标注 作业名称进入标注作业标注详情页。
- 3. 在标注作业标注详情中,展示此标注作业下全部数据。

同步数据源

ModelArts会自动从数据集输入位置同步数据至数据集详情页,包含数据及标注信息。 为了快速获取OBS桶中最新数据,可在数据集详情页的"未标注"页签中,单击"同步数据源",快速将通过OBS上传的数据添加到数据集中。

标注音频

标注作业详情页中,展示了此标注作业中"未标注"和"已标注"的音频,默认显示 "未标注"的音频列表。

- 1. 在"未标注"页签左侧音频列表中,单击目标音频文件,在右侧的区域中出现音频,单击音频下方 pp可进行音频播放。
- 2. 根据播放内容,选取合适的音频段,在下方"语音内容"文本框中填写音频标签 和内容。





3. 输入内容后单击下方的"确认标注"按钮完成标注。音频将被自动移动至"已标注"页签。

查看已标注音频

在标注作业详情页,单击"已标注"页签,您可以查看已完成标注的音频列表。单击音频,可在右侧的"语音内容"文本框中了解当前音频的内容信息。

修改标注

当数据完成标注后,您还可以进入"已标注"页签,对已标注的数据进行修改。

- 修改标签:在数据集详情页,单击"已标注"页签,然后在音频列表中选中待修改的音频。在右侧标签信息区域中修改"语音内容"中的"标签"和"内容",单击下方的"确认标注"按钮完成修改。
- 删除标签:单击目标编号操作列的 □ ,删除该段音频的标注。您也可以单击标注 音频文件上方的叉号删除标注,然后单击"确认标注"。

添加音频

除了同步新数据外,您还可以在标注详情页面中,直接添加数据,用于数据标注。

- 1. 在标注作业详情页面,单击"未标注"页签,然后单击左上角"添加数据"。
- 2. 在弹出的导入对话框中,选择数据来源、导入方式、导入路径等参数,导入数据。单击确定。

导入数据的详细操作介绍请参见导入操作。





删除音频

通过数据删除操作,可将需要丢弃的音频数据快速删除。

在"未标注"或"已标注"页面中,选中需要删除的音频,然后单击左上角"删除音频",在弹出的对话框中,根据实际情况选择是否勾选"同时删除源文件",确认信息无误后,单击"确定"完成音频删除操作。

□ 说明

如果勾选了"同时删除源文件",删除音频操作是将删除对应OBS目录下存储的音频。此操作可能会影响已使用此源文件的其他数据集或数据集版本,有可能导致展示异常或训练/推理异常。删除后,数据将无法恢复,请谨慎操作。

2.5 视频标注

由于模型训练过程需要大量有标签的视频数据,因此在模型训练之前需对没有标签的视频添加标签。通过ModelArts您可对视频添加标签,快速完成对视频的标注操作,也可以对已标注视频修改或删除标签进行重新标注。

□ 说明

视频标注仅针对视频帧进行标注。

开始标注

- 1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理> 数据标注",进入 "数据标注"管理页面。
- 2. 在标注作业列表中,基于"标注类型"选择需要进行标注的标注作业,单击标注 作业名称进入标注作业标注详情页。
- 3. 在标注作业标注详情中,展示此标注作业下全部数据。

同步数据源

ModelArts会自动从数据集输入位置同步数据至数据集详情页,包含数据及标注信息。

为了快速获取OBS桶中最新数据,可在数据集详情页的"全部"或"未标注"页签中,单击"同步数据源",快速将通过OBS上传的数据添加到数据集中。

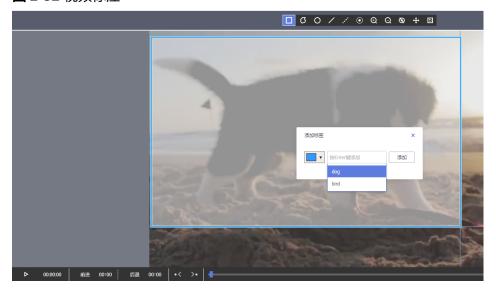
视频标注

标注作业详情页中,展示了此数据集中"未标注"和"已标注"的视频。

- 1. 在"未标注"页签左侧视频列表中,单击目标视频文件,打开标注页面。
- 2. 在标注页面中,播放视频,当视频播放至待标注时间时,单击进度条左侧的暂停 按钮,将视频暂停至某一帧对应的画面。
- 3. 在上方区域选择标注框,默认为矩形框。使用鼠标在视频画面中框出目标,然后在弹出的添加标签文本框中,直接输入新的标签名,在文本框前面选中标签颜色,单击"添加"完成1个物体的标注。如果已存在标签,从下拉列表中选择已有的标签,然后单击"添加"完成标注。逐步此画面中所有物体所在位置,一帧对应的画面可添加多个标签。

支持的标注框与"物体检测"类型一致,详细描述请参见物体检测章节的**表2标** 注界面的常用按钮。

图 2-52 视频标注



4. 上一帧对应的画面标注完成后,在进度条处单击播放按钮继续播放,在需要标注 处暂停,然后重复执行步骤**3**完成整个视频的标注。

单击界面右上角的"标签类别",在"标签类别"的详情页将呈现当前视频带标注的时间点。

图 2-53 当前文件标签信息

当前文件标签



5. 单击页面左上角"返回数据标注预览",页面将自动返回标注作业详情页面,同时,标注好的视频将呈现在"已标注"页签下。

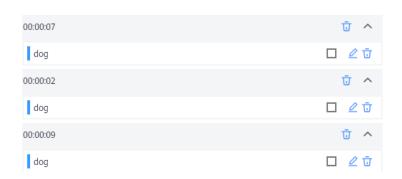
修改标注

当数据完成标注后,您还可以进入"已标注"页签,修改标注数据。

- 在"已标注"页签下,双击目标视频文件,在标注页面的右上方选择"标签类别"进入"当前文件标签"详情页面,可单击时间点右侧小三角展开详情,您可以修改或删除标签。
- 修改标签:在"当前文件标签"详情页,单击标签右侧的编辑按钮,标签名称可进行修改。
- 删除标签:在"当前文件标签"详情页,单击标签右侧的删除按钮,将直接删除 此标签。如果单击画面时间右侧的删除按钮,将删除此画面下的所有标签。

图 2-54 修改标注

当前文件标签



添加视频

除了同步新数据外,您还可以在标注详情页面中,直接添加数据,用于数据标注。

- 1. 在标注作业详情页面,单击"未标注"页签,然后单击左上角"添加数据"。
- 在弹出的导入对话框中,选择数据来源、导入方式、导入路径等参数,导入数据。单击确定。
 - 导入数据的详细操作介绍请参见OBS导入数据。

图 2-55 导入数据



删除视频

通过数据删除操作,可将需要丢弃的视频数据快速删除。

在"全部"、"未标注"或"已标注"页面中,依次选中需要删除的视频,或者选择"选择当前页"选中该页面所有视频,然后单击左上角"删除视频"。在弹出的对话

框中,根据实际情况选择是否勾选"同时删除源文件",确认信息无误后,单击"确定"完成视频删除操作。

其中,被选中的视频,其左上角将显示为勾选状态。如果当前页面无选中视频时, "删除视频"按钮为灰色,无法执行删除操作。

山 说明

如果勾选了"同时删除源文件",删除视频操作将删除对应OBS目录下存储的视频,此操作可能会影响已使用此源文件的其他数据集或数据集版本,有可能导致展示异常或训练/推理异常。删除后,数据将无法恢复,请谨慎操作。

3智能标注

3.1 创建智能标注作业

除了人工标注外,ModelArts还提供了智能标注功能,快速完成数据标注,为您节省70%以上的标注时间。智能标注是指基于当前标注阶段的标签及图片学习训练,选中系统中已有的模型进行智能标注,快速完成剩余图片的标注操作。

背景信息

- 目前只有"图像分类"和"物体检测"类型的标注作业支持智能标注功能。
- 启动智能标注时,需标注作业存在至少1种标签,且每种标签已标注的图片不少于5张。
- 启动智能标注时,必须存在未标注图片。
- 启动智能标注前,保证当前系统中不存在正在进行中的智能标注任务。
- 检查用于标注的图片数据,确保您的图片数据中,不存在RGBA四通道图片。如果存在四通道图片,智能标注任务将运行失败,因此,请从数据集中删除四通道图片后,再启动智能标注。

启动智能标注作业

- 1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理 > 数据标注",进入"数据标注"管理页面。
- 2. 在标注作业列表中,选择"物体检测"或"图像分类"类型的标注作业,单击操作列的"智能标注"启动智能标注作业。
- 3. 在弹出的"启动智能标注"对话框中,选择智能标注类型,可选"主动学习"或者"预标注",详见表3-1和表3-2。

表 3-1 主动学习

| 参数 | 说明 |
|---------|---|
| 智能标注 类型 | "主动学习"。"主动学习"表示系统将自动使用半监督学习、难例筛选等多种手段进行智能标注,降低人工标注量,帮助用户找到难例。 |

| 参数 | 说明 |
|------|--|
| 算法类型 | 针对"图像分类"类型的数据集,您需要选择以下参数。 "快速型":仅使用已标注的样本进行训练。 "精准型":会额外使用未标注的样本做半监督训练,使得模型精度更高。 |

表 3-2 预标注

| 参数 | 说明 |
|-------------|--|
| 智能标注 类型 | "预标注"。"预标注"表示选择用户模型管理里面的模型,选择模型时需要注意模型类型和数据集的标注类型相匹配。预标注结束后,如果标注结果符合平台定义的标准标注格式,系统将进行难例筛选,该步骤不影响预标注结果。 |
| 选择模型 及版本 | • "我的模型"。您可以根据实际需求选择您的模型。您需要在目标模型的左侧单击下拉三角标,选择合适的版本。您的模型导入参见导入模型。 |
| | "我的订阅"。您可以根据实际需求选择AI Gallery中已订阅的模型。您需要在目标模型的左侧单击下拉三角标,选择合适的版本。查找模型参见我的订阅模型。 |
| 计算节点 规格 | 在下拉框中,您可以选择目前ModelArts支持的节点规格选项。 |
| 计算节点 个数 | 默认为1。您可以根据您的实际情况选择,最大为5。 |

□ 说明

- 针对"物体检测"类型的标注作业,选择"主动学习"时,只支持识别和标注矩形框。
- 当系统中智能标注作业过多时,可能会出现排队的情况,导致作业一直处于"标注中"的状态。请您耐心等待,系统会按照顺序完成标注作业。

图 3-1 启动智能标注(图像分类)

启动智能标注

智能标注类型

② 主动学习 ① 预标注
系统将自动使用半监督学习,难例筛选等多种手段进行智能标注,降低人工标注量,帮助用户找到难例。

③ 快速型 ② 精准型

限时免费

创建免费,使用阶段按训练的时长收费。 了解计费详情

提交

图 3-2 启动智能标注(物体检测)



图 3-3 启动智能标注(预标注)



- 4. 完成参数设置后,单击"提交",即可启动智能标注。
- 5. 在标注作业列表中,单击标注作业名称进入"标注作业详情"页。
- 6. 在"数据集概览页标注作业详情页",选择"标注"页签,单击"待确认"页签,即可查看智能标注进度。

您也可以在该页签,"启动智能标注"或者查看"智能标注历史"

图 3-4 标注进度



- 7. 智能标注完成后,"待确认"页面将呈现所有标注后的图片列表。
 - 图像分类标注作业

在"待确认"页面查看标签是否准确,勾选标注准确的图片,然后单击"确认"完成智能标注结果的确认。确认完成后的图片将被归类至"已标注"页面下。

针对标为"难例"的图片,您可以根据实际情况判断,手工修正标签。详细操作及示例请参见•针对"图像分类"数据集。

- 物体检测标注作业

在"待确认"页面,单击图片查看标注详情,查看标签及目标框是否准确, 针对标注准确的图片单击"确认标注"完成智能标注结果的确认。确认完成 后的图片将被归类至"已标注"页面下。

针对标为"难例"的图片,您可以根据实际情况判断,手工修正标签或目标框。详细操作及示例请参见•针对"物体检测"数据集。

3.2 难例确认

在数据量很大的标注任务中,标注初期由于已标注图片不足,智能标注的结果无法直接用于训练。若对所有的未标注数据——进行调整确认仍然需要较大的人力和时间成本。为了更快地完成标注任务,在对未标注数据进行智能标注的任务中,ModelArts嵌入了自动难例发现功能。该功能会对剩余未标注图片的标注优先级给出建议。因为标注优先级高的图片的智能标注结果未达到预期,所以称之为难例。

ModelArts平台提供的自动难例发现功能,在智能标注以及数据采集筛选过程中,将自动标注出难例,建议对难例数据进一步确认标注,然后将其加入训练数据集中,使用此数据集训练模型,可得到精度更高的模型。首先,针对智能标注和采集筛选任务,难例的发现操作是系统自动执行的,无需人工介入,仅需针对标注后的数据进行确认和修改即可,提升数据管理和标注效率。其次,您可以基于难例的情况,补充类似数据,提升数据集的丰富性,进一步提升模型训练的精度。

在数据集管理中,对难例的管理有如下场景。

- 智能标注后,确认难例
- 将数据集中的数据标注为难例

□ 说明

目前只有"图像分类"和"物体检测"类型的数据集支持难例发现功能。

智能标注后,确认难例

"智能标注"任务执行过程中,ModelArts将自动识别难例,并完成标注。当智能标注结束后,难例标注结果将呈现在"待确认"页签,建议您对难例数据进行人工修正,然后确认标注。

- 1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理>数据标注",进入 "数据标注"管理页面。
- 2. 在标注作业列表中,选择"物体检测"或"图像分类"类型的标注作业,单击标注作业名称进入"标注作业详情"。
- 3. 在"标注作业详情页",选择"标注>待确认"页签,查看并确认难例。

□ 说明

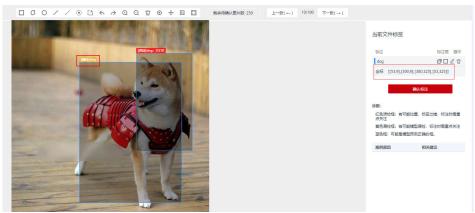
只有当智能标注任务完成后,待确认页签才会显示标注数据。否则,此页签内容为空。智 能标注操作请参见**创建智能标注作业**。

- 针对"物体检测"标注作业

在"待确认"页签中,单击图片展开标注详情,查看图片数据的标注情况,如标签是否准确、目标框位置添加是否准确。如果智能标注结果不准确,建议手工调整标签或目标框,然后单击"确认标注"。完成确认后,重新标注的数据将呈现在"已标注"页签下。

如<mark>图3-5</mark>所示的难例,dog标签的目标框位置不准确,使用标注框重新标注,如图中的"漏检"目标框,然后需要将原先标注错误的目标框删除,即"误检"标签框。手工调整后,单击"确认标注"完成难例确认。

图 3-5 物体检测的难例确认



- 针对"图像分类"标注作业

在"待确认"页签中,查看标注难例的图片,其添加的标签是否准确。勾选标注不准确的图片,删除错误标签,然后在右侧"标签名"处添加准确标签。单击"确认",勾选的图片及其标注情况,将呈现在"已标注"页签下。

如<mark>图3-6</mark>所示,选中的图片为标注错误图片,在右侧删除错误标签,然后在标签名处添加"狗"的标签,然后单击"确认",完成难例确认。

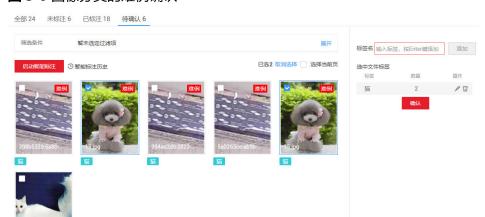


图 3-6 图像分类的难例确认

将数据集中的数据标注为难例

针对标注作业中,已标注或未标注数据,也可以将图片数据标注为难例。标注为难例的数据,对后续模型训练中,通过内置规则提升模型精度。

- 1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理>数据标注",进入 "数据标注"管理页面。
- 2. 在标注作业列表中,选择"物体检测"或"图像分类"类型的标注作业,单击数据集名称进入"标注作业详情页"。
- 3. 在标注作业详情页,选择"标注"页签",单击"已标注"、"未标注"或"全部"页签,勾选需标注为难例的图片,然后单击"难例批处理 > 确认为难例"。 完成标注后,图片预览时,其右上角将显示为"难例"。

图 3-7 确认为难例



4 团队标注

4.1 团队标注简介

数据标注任务中,一般由一个人完成,但是针对数据集较大时,需要多人协助完成。 ModelArts提供了团队标注功能,可以由多人组成一个标注团队,针对同一个数据集进 行标注管理。

□□说明

团队标注功能当前仅支持"图像分类"、"物体检测"、"文本分类"、"命名实体"、"文本三元组"、"语音分割"类型的数据集。

数据标注任务中,一般由一个人完成,但是针对数据集较大时,需要多人协助完成。 ModelArts提供了团队标注功能,可以由多人组成一个标注团队,针对同一个数据集进 行标注管理。

团队标注功能当前仅支持"图像分类"、"物体检测"、"文本分类"、"命名实体"、"文本三元组"、"语音分割"类型的数据集。

针对启用团队标注功能的数据标注任务,支持创建团队标注任务,将标注任务指派给不同的团队,由多人完成标注任务。同时,在成员进行数据标注过程中,支持发起验收、继续验收以及查看验收报告等功能。

团队标注功能是以团队为单位进行管理,数据集启用团队标注功能时,必须指定一个团队。一个团队可以添加多个成员。

- 一个帐号最多可添加10个团队。
- 如果数据集需要启用团队标注功能,当前帐号至少拥有一个团队。如果没有,请执行添加团队操作添加。

4.2 创建和管理团队

4.2.1 管理团队

团队标注功能是以团队为单位进行管理,数据集启用团队标注功能时,必须指定一个团队。一个团队可以添加多个成员。

背景说明

- 一个帐号最多可添加10个团队。
- 如果数据集需要启用团队标注功能,当前帐号至少拥有一个团队。如果没有,请 执行**添加团队**操作添加。

添加团队

- 1. 在ModelArts管理控制台左侧导航栏中,选择"数据管理>标注团队",进入"标注团队"管理页面。
- 2. 在"标注团队"管理页面,单击"添加团队"。
- 3. 在弹出的"添加团队"对话框中,填写团队"名称"和"描述",然后单击"确定"。完成标注团队的添加。

图 4-1 添加团队

添加团队



团队添加完成后,"标注团队"管理页面呈现新添加的团队,在页面右侧区域,可以查看团队详情。新添加的团队,其成员列表为空,请参考<mark>添加成员</mark>操作,为您的团队添加成员。

删除团队

当已有的团队不再使用,您可以执行删除操作。

在"标注团队"管理页面中,选中需删除的团队,然后单击"删除"。在弹出的对话框中,确认信息无误后,单击"确定"完成团队删除。

图 4-2 删除团队



4.2.2 管理成员

新添加的团队,其成员列表为空。您需要根据实际情况添加即将参与标注任务的成员信息。

一个团队最多支持添加100个成员,当超过100时,建议分为多个团队进行管理。

添加成员

- 1. 在ModelArts管理控制台左侧导航栏中,选择"数据管理>标注团队",进入"标注团队"管理页面。
- 在"标注团队"管理页面,从左侧团队列表中选择一个团队,单击团队,其右侧 区域将呈现"团队详情"。
- 3. 在"团队详情"区域,单击"添加成员"。
- 4. 在弹出的"添加成员"对话框中,按照分类,可选择"当前账号"或"其他账号"。选择"当前账号",需要填写成员的"IAM用户名"、"角色"、"邮箱"、"描述"。选择"其他账号",需要填写成员的"账号ID"、"IAM用户ID"、"角色"、"描述",单击左下角"添加账号ID"。填写完成之后,然后单击"确定"。

邮箱作为团队管理中的唯一标识,不同成员不能使用同一个邮箱。您填写的邮箱 地址将被记录并保存在ModelArts中,仅用于ModelArts团队标注功能,当成员删 除后,其填写的邮箱信息也将被一并删除。

其中,"角色"支持"Labeler"、"Reviewer"和"Team Manager", "Team Manager"只能设置为一个人。



图 4-3 当前账号添加成员

图 4-4 其他账号添加成员



成员添加完成后,团队详情区域中将呈现此成员的信息。

修改成员信息

团队中的成员,当其信息发生变化时,可以编辑其基本情况。

- 1. 在"团队详情"区域,选择需修改的成员。
- 2. 在成员所在行的"操作"列,单击"修改"。在弹出的对话框中,修改其"描述"或"角色"。

成员的"邮箱"无法修改,如果需要修改邮箱地址,建议先删除此成员,然后再基于新的邮箱地址添加新成员。

"角色"支持"Labeler"、"Reviewer"和"Team Manager","Team Manager"只能设置为一个人。

删除成员

• 删除单个成员

在"团队详情"区域,选择需要删除的成员, 单击"操作"列的"删除"。在弹出的对话框中,确认信息无误后,单击"确定"完成删除操作。

• 批量删除

在"团队详情"区域,勾选需删除的成员,然后单击"删除"。在弹出的对话框中,确认信息无误后,单击"确定"完成多个成员的删除操作。

图 4-5 批量删除



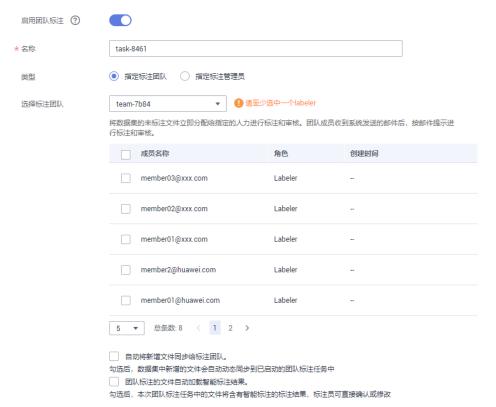
4.3 创建团队标注任务

如果您在创建标注作业时,即启用团队标注,且指派了某一团队负责标注,系统将默 认基于此团队创建一个标注任务。您可以在创建数据标注任务后,在"标注任务进 展"页面查看此任务。 您还可以重新创建一个团队标注任务,指派给同一团队的不同成员,或者指派给其他 标注团队。

团队标注作业的创建方式

从控制台的"数据管理>数据标注"页面进入,创建标注作业时,打开"启用团队标注"开关,同时指定一个标注团队,或者指定标注管理员。

图 4-6 创建团队标注作业



从控制台的"数据管理>数据集"页面进入,对于同一个数据集,可以创建多个团队标注任务。

图 4-7 在数据集列表中创建团队标注任务



图 4-8 在数据集概览页中,选择标注任务页签,点击"新建标注任务",创建团 队标注任务



图 4-9 在数据集详情页创建团队标注任务



🗀 说明

- 当标注人员第一次收到标注邮件时,会告知标注的地址(labelConsole)以及默认初始密码。
- 只有当创建团队标注任务时,标注人员才会收到邮件。创建标注团队及添加标注团队的成员并不会发送邮件。此外,当所有样本都是已标注状态时,创建团队标注任务也不会收到邮件。
- 标注任务创建完成后,会将所有未标注状态的样本分配给标注人员。分配采用随机均分的策略,不支持重复分配。

创建团队标注任务

同一个数据集,支持创建多个团队标注作业,指派给同一团队的不同成员,或者指派给其他标注团队。

1. 登录ModelArts管理控制台,在左侧菜单栏中选择"数据管理 >数据集",打开数 据集列表。

- 2. 在数据集列表中,选择支持团队标注的数据集,单击数据集名称进入数据集概览 页。
- 3. 在数据集概览页中,单击"标注任务"页签,可查看此数据集已有的标注任务。 单击左上角的"新建标注任务"开始创建新任务。

图 4-10 标注任务



或者也可以从"数据管理>数据标注"页面进入,单击"创建标注作业"进入创建标注作业页面。

- 4. 在弹出的"创建标注作业"页面中,填写相关参数,然后单击"确定",完成任务创建。
 - "名称":设置此任务的名称。
 - "标注场景":选择标注作业的任务类型。
 - "标签集":展示当前数据集已有的标签及标签属性。
 - **"启用团队标注"**:选择打开,并配置如下团队标注相关参数。
 - "类型":设置任务类型,支持"指定标注团队"或"指定标注管理员"。
 - "选择标注团队":任务类型设置为"指定标注团队",需在此参数中 指定一个团队,同时勾选此团队中某几个成员负责标注。下拉框中将罗 列当前帐号下创建的标注团队及其成员。
 - "选择标注接口人":任务类型设置为"指定标注管理员",需在所有团队的"Team Manager"中选择一人作为管理员。
 - "自动将新增图片同步给标注团队":根据需要选择是否将任务中新增的数据自动同步给标注人员。
 - "团队标注的图片自动加载智能标注结果":根据需要选择是否将任务中智能标注待确认的结果自动同步给标注人员。

□ 说明

团队标注加载智能标注结果的处理步骤:

- 如果类型选择"指定标注团队",需要先创建团队标注任务,然后执行智能标注任务。
- 如果类型选择"指定标注管理员",需要登录labelconsole先执行"分配任务",然后执行智能标注任务。

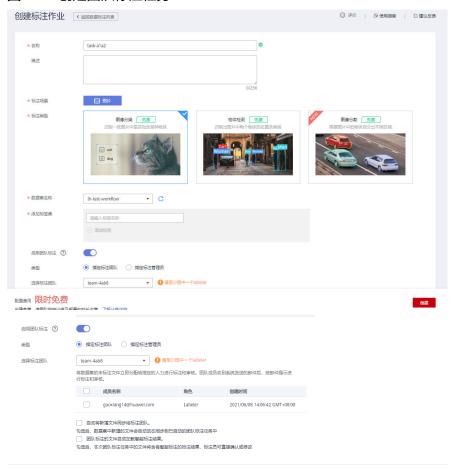


图 4-11 创建团队标注任务

任务创建完成后,您可以在"标注任务"页签下看到新建的任务。

4.4 登录数据标注平台

在ModelArts中,一般用户使用数据标注功能,直接是在"数据管理"模块操作,此模块包含数据管理所有能力,数据集管理、数据标注、数据导入导出、智能标注、团队标注和管理等。

而数据标注平台,仅在团队标注场景下使用,此平台仅包含基础的数据预览、数据标注能力,面向团队标注任务中的Labeler和Reviewer使用。

数据标注平台的访问地址可以查看如下表格获取,另外,如果团队成员绑定了邮箱,可以收到任务通知邮件,成员也可以通过邮件中给出的地址访问LabelConsole平台。 登录的用户名密码也在任务邮件中提供。针对此平台,无需注册华为云帐号,直接使用任务邮件中提供的用户名密码即可。

表 4-1 数据标注平台访问地址

| 局点 | LabelConsole访问地址 |
|------------|---|
| 华北-北 京四 | https://cn-north-4-modelarts-label.huaweicloud.com/label-console/ |

| 局点 | LabelConsole访问地址 |
|------------|---|
| 华北-北 京一 | https://cn-north-1-modelarts-label.huaweicloud.com/label-console/ |
| 华东-上 海一 | https://cn-east-3-modelarts-label.huaweicloud.com/label-console/ |
| 华南-广 州 | https://cn-south-1-modelarts-label.huaweicloud.com/label-console/ |
| 中国-香港 | https://ap-southeast-1-modelarts-label.huaweicloud.com/label-console/ |
| 亚太-新 加坡 | https://ap-southeast-3-modelarts-label.huaweicloud.com/label-console/ |
| 亚太-曼 谷 | https://ap-southeast-2-modelarts-label.huaweicloud.com/label-console/ |

山 说明

首次登录数据标注平台时,请务必修改密码。如果未修改,将导致下一次无法正常登录。 针对数据标注平台老用户(同一个邮箱地址,之前已接收过团队标注任务,再次接收任务时,称 为老用户),团队标注邮件中将不提供用户名和密码,仅提供数据标注平台地址。

登录后,仅显示当前用户(此邮箱用户)相关的团队标注任务及其相关数据。

图 4-12 新用户接收的任务邮件



图 4-13 数据标注平台登录页面

IAM用户登录



图 4-14 标注页面



4.5 启动团队标注任务

在标注任务创建后,被分配任务的团队成员将收到一封通知邮件,标题为"您有新的标注任务待查收"。

在邮件详情中,单击标注任务链接进入数据标注平台,使用您的"邮箱地址"及"初始密码"登录标注平台系统。首次登录时,需根据提示修改密码。

登录数据标注平台后,可查看到分配的标注任务,单击任务名称,可进入标注页面。 不同类型的数据集,标注方式不同,详细请参见:

- 图像分类
- 物体检测
- 文本分类
- 命名实体

• 文本三元组

在标注平台中,每个成员可查看"未标注"、"待修正"、"已驳回"、"待审核"、"审核通过"、"验收通过"的图片信息。请及时关注管理员驳回以及待修正的图片。

当团队标注任务中,分配了Reviewer角色,则需要对标注结果进行审核,审核完成后,再提交给管理员验收。

图 4-15 成员标注平台



4.6 审核团队标注任务结果

团队标注成员完成后,团队审核者可以在数据标注平台上对标注结果进行审核。

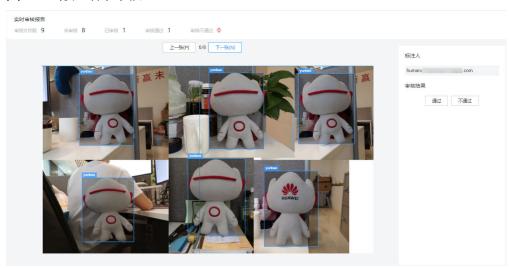
1. 登录数据标注平台,在任务列表中单击"审核",发起审核。

图 4-16 发起审核



2. 在审核页面中,审核人员可以查看"未审核"、"已审核"、"审核通过"、 "审核不通过"的样本。

图 4-17 标注结果审核



3. 审核人员可以在审核页面的右侧选择"审核结果"("通过"或"不通过")。

当选择审核结果为"通过"时,需设置"验收评分"(分"A"、"B"、 "C"、"D"四个选项,"A"表示最高分),如<mark>图4-18</mark>所示。当选择审核结果 为"不通过"时,可以在文本框中写明驳回原因,如<mark>图4-19</mark>所示。

图 4-18 审核通过



图 4-19 审核不通过

标注人 human/ com 审核结果 请输入审核意见 0/256

4.7 验收团队标注任务结果

任务验收(管理员)

• 发起验收

当团队的成员已完成数据标注,标注作业的创建者可发起验收,对标注结果进行抽验。只有当标注成员存在标注完成的数据时,才可以发起验收,否则发起验收按钮为灰色。

- a. 录ModelArts管理控制台,在左侧菜单栏中选择"数据管理 >数据标注",打 开数据标注管理页。
- b. 在数据标注作业列表,选择团队标注作业,点击作业名称进入"标注作业详情页",单击右上角"验收",发起验收。
- c. 在弹出的对话框中,设置"抽样策略",可设置为"按百分比",也可以设置为"按数量"。设置好参数值后,单击"确定"启动验收。
 - "按百分比":按待验收图片总数的一定比例进行抽样验收。
 - "按数量":按一定数量进行抽样验收。

图 4-20 发起验收

发起验收



d. 验收启动后,界面将展示实时验收报告,您可以在右侧选择"验收结果" ("通过"或"不通过")。

当选择验收结果为"通过"时,需设置"验收评分"(分"A"、"B"、 "C"、"D"四个选项,"A"表示最高分),如<mark>图4-22</mark>所示。当选择验收 结果为"不通过"时,可以在文本框中写明驳回原因,如<mark>图4-23</mark>所示。

图 4-21 查看实时验收报告



图 4-22 设置验收结果为"通过"

验收结果



图 4-23 设置验收结果为"不通过"

验收结果

| not good | | | |
|----------|----|--------|-------|
| | | | 8/256 |
| | 通过 | 确认为不通过 | 跳过 |

• 继续验收

针对未完成验收的任务,可以继续验收。针对未发起过验收流程的任务,不支持"继续验收",按钮为灰色。

在"标注任务进展"页签中,针对需继续验收的任务,单击"继续验收"。系统直接进入"实时验收报告"页面,您可以继续验收未验收的图片,设置其"验收结果"。

• 完成验收

在完成验收窗口,您可以查看本标注作业的验收情况,如抽样文件数等,同时设置如下参数,然后进行验收。只有完成验收,标注信息才会同步到标注作业的已标注页面中。

一旦标注数据完成验收,团队成员无法再修改标注信息,只有数据集创建者可修 改。

表 4-2 完成验收的参数设置

| 参数 | 说明 | | | | |
|----------|---|--|--|--|--|
| 对已标注数据修改 | ● 不覆盖:针对同一个数据,不使用当前团队标注的结果 覆盖已有数据。 | | | | |
| | ● 覆盖:针对同一个数据,使用当前团队标注的结果覆盖已有数据。覆盖后无法恢复,请谨慎操作。 | | | | |

| 参数 | 说明 |
|------|--|
| 通过范围 | 全部: 当前团队标注完成的所有数据。包含验收通过、 未验收和验收不通过的。即本数据集的所有抽样文件 数。 |
| | 全部不通过:当前团队标注完成的所有数据不通过验收,即将所有标注数据驳回给标注人员。全部数据指验收通过、未验收和验收不通过的所有数据,即本数据集的所有抽样文件数。 |
| | 验收通过和未验收的数据:针对抽样文件中验收通过和 未验收的数据,通过验收。验收不通过的数据将驳回给 标注人员。 |
| | 验收通过的数据:针对抽样文件中验收通过的数据,通过验收。未验收和验收不通过的数据将驳回给标注人员。 |

图 4-24 完成验收

完成验收

验收通过后,标注结果才会同步到数据集已标注



查看验收报告

针对进行中或已完成的标注任务,都可以查看其验收报告。在"标注任务进展"页签中,单击"验收报告",即可在弹出的"验收报告"对话框中查看详情。

图 4-25 查看验收报告

验收报告

| 进行中验收统 | 计信息 | | | | |
|--------|-----|-------|---|-------|---|
| 验收通过率 | -% | 抽样文件数 | - | 未验收 | |
| 已验收 | | 验收通过 | - | 验收不通过 | |
| 已完成验收统 | 计信息 | | | | |
| 验收通过率 | -% | 抽样文件数 | | 未验收 | - |
| 已验收 | | 验收通过 | - | 验收不通过 | |
| | | | | | |
| | | | | | |

删除标注任务

验收结束后,针对不再使用的标注任务,您可以在"标注任务进展"页签下,单击任 务所在行的删除。任务删除后,未验收的标注详情将丢失,请谨慎操作。但是数据集 中的原始数据以及完成验收的标注数据仍然存储在对应的OBS桶中。