

华为云大赛 - ModelArts 平台使用指导

Baseline

文档版本 01

发布日期 2020-03-27



版权所有 © 华为技术有限公司 2019。 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

HUAWEI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默示的声明或保证。

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。

华为技术有限公司

地址: 深圳市龙岗区坂田华为总部办公楼 邮编: 518129

网址: http://www.huawei.com

客户服务邮箱: support@huawei.com

客户服务电话: 4008302118

目 录

1	文档导读	1
2	注册华为云账号	2
3	准备数据	3
4	在 ModelArts 全局配置中添加访问密钥	6
5	创建训练作业	7
6	将生成的模型导入至模型管理	10
7	将模型部署为在线服务	11
8	提交参赛模型	13
9	在 ModelArts Notebook 中进行开发调试	15
10	· 释放资源	17
11	Baseline 代码说明	18

1 文档导读

ModelArts 是面向 AI 开发者的一站式开发平台,您可以使用 ModelArts 完成华为云大赛数据处理、模型训练、推理测试以及结果的提交。

本文档包含使用 ModelArts 完成华为云大赛的全流程指导,通过本文档您可以快速了解 ModelArts 平台的使用,有助于您使用 ModelArts 来完成华为云大赛题目。

您可以根据如下表格快速查找您需要了解的内容。

章节	说明
2 注册华为云账号	使用 ModelArts 前需注册华为云账号并进行实名认证; 如己有华为云账号可跳过此步骤。
3 准备数据	下载竞赛数据集,并将数据集上传至 OBS(华为云对象存储服务)中。
4 在 ModelArts 全局配置中添加访问密钥	使用 ModelArts 前,需要添加访问密 钥。
5 创建训练作业	基于 ModelArts 提供的训练管理功能, 完成模型的训练。
6 将生成的模型导入至模型管理	将训练生成的模型导入至模型管理。
7 将模型部署为在线服务	将模型部署为在线服务,并查看预测结 果是否正确。
8 提交参赛模型	将参赛模型进行参赛发布,后台将进行 自动判分。
9在 ModelArts Notebook 中进行开发调试	在 ModelArts Notebook 中调试训练、预测或推理脚本。
10 释放资源	为避免资源浪费或账号欠费,使用完计 算资源后需及时释放资源。
11 Baseline 代码说明	关于 Baseline 代码的使用方法说明。

2 注册华为云账号

使用华为云服务之前您需要注册华为云帐号,通过华为云帐号,您可以使用所有华为云服务。注册华为云账号具体操作如下。

- 1. 进入华为云首页,点击页面右上角的"注册"。
- 2. 设置手机号、短信验证码、账号名、密码并勾选"我已阅读并同意《华为云用户协议》和《隐私政策声明》",点击"同意协议并注册"。
- 3. 账号注册成功后,登录华为云,点击界面右上角"用户名>实名认证",根据界面 提示信息完成实名认证。

3 准备数据

ModelArts 使用对象存储服务(OBS)存储数据和模型的备份,实现安全、高可靠和低成本的存储需求。因此,在使用 ModelArts 开发模型前,您需要将竞赛数据集上传至您的 OBS 桶中。

本节介绍下载竞赛数据集至本地并解压后,然后通过 OBS Browser 创建桶及文件夹,将竞赛数据集上传至您的 OBS 桶中。具体操作如下。

步骤一: 下载竞赛数据集至本地并解压缩

通过如下链接或竞赛平台赛题说明页面下载数据集 trainval.zip:

https://ma-competitions-bj4.obs.cn-north-4.myhuaweicloud.com:443/SODIC/garbage/datasets/trainval.zip。

将数据集 trainval.zip 下载至本地后解压缩,包含如下文件,可以浏览部分图片以了解数据的情况。

名称	说明
VOC2007	按照PASCAL VOC格式存放的物体检测标注数据,Annotations目录中是xml标签文件,ImageSets/Main/trainval.txt是所有图片的文件名前缀,JPEGImages目录中是图片文件
train_classes.txt	所有的垃圾物体的二级类别名
classify_rule.json	垃圾分类的一级类别和二级类别对应关系字典, key值是一级类别,value是二级类别列表。

步骤二:下载 OBS Browser 并上传数据集

具体操作步骤如下:

- 1. 点此下载 OBS Browser,下载完成后解压缩,点击 obs.exe 运行 OBS Browser。
- 2. 获取访问密钥(AK 和 SK),具体操作可参见 https://support.huaweicloud.com/modelarts_faq/modelarts_05_0004.html。华为云上的 许多产品都需要用到访问密钥,请妥善保存密钥文件。

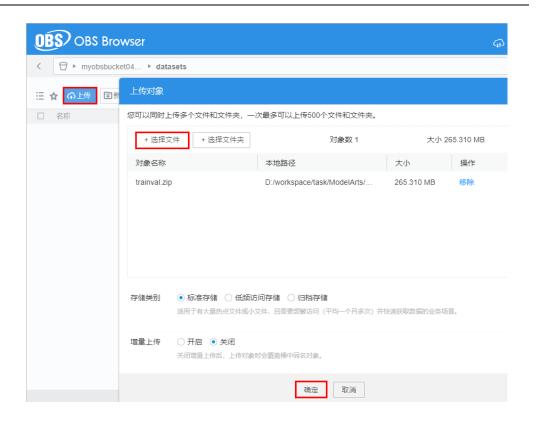
3

3. 在 OBS Browser 添加新账号页面自定义账号名,添加获取的访问密钥,然后点击确定。

账号添加完成后,点击"添加桶",注意:区域要选择"华北-北京四",标准存储、私有、关闭多 AZ,桶名需自定义,OBS 桶名要全局唯一,如提示桶名已存在,则需要您修改为其他名称,比如此处设置桶名为 myobsbucket0408。

添加桶	×
方式	● 创建新桶 ○ 添加外部桶
区域	华北-北京四 ▼ ②
存储类别	● 标准存储○ 低频访问存储○ 归档存储适用于有大量热点文件或小文件,且需要频繁访问(平均一个月多次)并快速获取数据的业务场景。
桶策略	私有公共读公共读写桶的拥有者拥有完全控制权限,其他用户在未经授权的情况下均无访问权限。
\$AZ ②	○ 开启● 关闭多AZ能提高您的数据可用性,同时会采用相对较高的计费标准。价格详情
桶名	myobsbucket0408
	确定取消

点击桶名称进入到桶中,点击"新建文件夹",输入 datasets,确定,点击 datasets 进入到文件夹中,点击"上传"、"选择文件",选择上一步下载的数据集压缩包 trianval.zip,点击"确定"开始上传,您可以在"任务管理"中查看上传进度。



4 在 ModelArts 全局配置中添加访问密钥

使用 ModelArts Notebook、训练作业、模型及服务时可能需要用到对象存储服务,因此 在使用 ModelArts 之前,您需要在 ModelArts 全局配置中添加访问密钥。

在步骤 3 准备数据中已创建过访问密钥,可直接将该密钥添加至 ModelArts。具体操作如下。

添加访问密钥

- 1. 在浏览器中打开 ModelArts 主页,点击"进入控制台",输入账号密码登录。
- 2. 在 ModelArts 左侧导航栏选择"全局配置",点击页面中的"添加访问密钥", 在弹出的对话框中,填写前面步骤获取的访问密钥。
- "访问密钥(AK)"输入密钥文件中的 Access Key Id 字段内容,"私有访问密钥(SK)"输入密钥文件中 Secret Access Key 字段内容。
- 请确保所填写的 AK、SK 为当前账号所获取的。
- 3. 点击"确认",完成访问密钥的添加。

5 创建训练作业

参考本文第 11 节,下载大赛 baseline 模型训练代码 src.zip,解压得到 src 目录,打开 OBS Browser,回到刚创建的桶 myobsbucket0408 目录下,点击"上传"、"选择文件 夹",选择刚解压的 src 目录,点击确定进行上传。

上传模型训练代码至 OBS 完成后,您可以在 ModelArts 训练作业模块创建训练作业进行模型的训练,具体操作如下。

- 1. 点此链接进入 ModelArts 控制台,在左侧导航栏中点击"训练管理"、"训练作业",然后点击页面右侧"训练作业"页签中的"创建",进入创建训练作业页面。
- 2. 在创建训练作业页面填写参数,参数填写说明如下表所示,注意将下表中的桶名 更换成您自己定义的桶名。

参数名	参数值
算法来源	选择"常用框架"
	AI引擎: TensorFlow、TF-1.13.1-python3.6
	代码目录: myobsbucket0408/src/
	启动文件: myobsbucket0408/src/train.py
数据来源	选择"数据存储位置"
	数据存储位置: myobsbucket0408/datasets/,注意,训练过程会将该
	目录下的所有数据都从OBS拷贝到ModelArts中,请不要在该目录
	下放置无关的数据,以免拷贝时间过长
训练输出位置	myobsbucket0408/train_output/,自己新建一个文件夹即可
运行参数	train_url: myobsbucket0408/train_output/,设置训练输出位置后会自
	动生成该参数的值,无需手动设定
	data_url: myobsbucket0408/datasets/,设置数据存储位置后会自动生
	成该参数的值,无需手动设定
	(下面的参数需要手动点击"增加运行参数")
	max_epochs_1:5,第一阶段的训练轮数
	max_epochs_2: 10,整个训练过程的总训练轮数
作业日志路径	myobsbucket0408/logs/,自己新建一个文件夹即可
类型	GPU
规格	GPU: 1*nvidia-v100 16GiB CPU: 8核 64GiB 32GiB,根据自己的需

要选择一个合适的规格即可

所有参数配置完成后,如下图所示:

* 算法来源	算法管理	预置算法	常用框架	自定义	
	选择常用引擎创建议	川练作业。			
	* AI引擎	TensorFlow	▼ TF-1	.13.1-python3.	6 ▼
	*代码目录 ②	/myobsbuc	ket0408/src/		选择
	*启动文件 ?	/myobsbuc	ket0408/src/train.	ру	选择
*数据来源 ⑦	数据集	数据存储位置			
	*数据存储位置	/myobsbuck	et0408/datasets/		选择 ① 立
*训练输出位置 ⑦	/myobsbucket040 请尽量选择空目录来			择	
运行参数 ⑦	train_url		= /myobsbu	cket0408/train	_output
	data_url		= /myobsbu	cket0408/datas	sets/
	max_epochs_1		= 5		Ū
	max_epochs_2		= 10		Ū
	⊕ 増加运行参数				
作业日志路径 ②	/myobsbucket040		选	1,3.1	
	日志默认保存在服务	各,会 不定期清晰	k, 请选择相应路径	用来上传日志。	
* 资源池	公共资源池	专属资源池	ļ.		
* 类型	CPU G	PU			
* 规格	GPU: 1 * nvidia-			-	
	适合场景: GPU计算	尊型,适合标准深	渡学习在GPU上的	代码运行与调测	y .
* 计算节点个数	- 1 +				

3. 参数填完成后,点击"下一步",确认参数无误后,点击"提交"、"查看作业详情",作业的状态会经历"初始化"、"部署中"、"运行中"和"运行成功"四个状态。

训练作业运行完成后,在训练输出位置会自动生成 model 文件夹,且 model 文件夹下包含如下文件或目录。

<		output ▶ V0006 ▶ model
∷≡	★ ○ 上传	下載
	名称	存储类别
	yolo3	-
	classify_rule.json	标准存储
	train_anchors.txt	标准存储
	train_classes.txt	标准存储
	trained_weights_final.h5	标准存储
	config.json	标准存储
	customize_service.py	标准存储

上图中每个文件或目录的含义如下表所示:

文件或目录名	说明
yolo3	从训练代码中复制过来的 yolo3 目录,该目录中包含有 yolo3 模型的定义
classify_rule.json	垃圾分类的标准, key 值是一级类别, value 是二级类别列表。模型预测得到的类别是二级类别, 需要根据该分类标注文件查询得到对应的一级类别名, 最终推理过程输出的结果格式是"一级类别名/二级类别名"
train_anchors.txt	yolo3 模型训练过程中生成的 anchors 参数文件,其中的 anchor 参数在模型推理阶段仍然需要用到
train_classes.txt	训练数据的所有类别名,在推理时用于根据类别 id 查找到类别名
trained_weights_final.h5	训练过程保存的模型参数文件
config.json	ModelArts 平台上"模型管理"和"部署上线"功能必需的推理配置脚本,脚本的内容需要遵循 ModelArts 的相关规范
customize_service.py	ModelArts 平台上"模型管理"和"部署上线"功能必需的推理脚本,脚本的内容需要遵循 ModelArts 的相关规范

6 将生成的模型导入至模型管理

准备好模型及相关的必备文件后,您可以将生成的模型导入至 ModelArts 模型管理。 导入模型具体操作如下。

1. 在 ModelArts 控制台的左侧导航栏点击"模型管理"、"模型",点击右侧页面中的"导入"。

在导入模型页面填写名称,选择元模型来源,可以直接从训练中选择,也可以从 OBS 中选择。如果是从 OBS 中选择,则需要选择到 model 文件夹的上一级目录; 例如,本 baseline model 目录的上一级目录为/myobsbucket0408/train_output。

注意:选择好元模型路径后,"AI 引擎"会自动填充。如未能自动填充,请检查元模型路径是否是 model 目录的上一级目录,或者 model 目录下是否包含模型配置文件 config.json。



2. 点击"立即创建"。需要一点时间来等待模型导入和构建,当模型版本状态为 "正常"后,即导入模型成功,然后进行下一步将模型部署为在线服务的操作。



了 将模型部署为在线服务

导入模型完成后,接下来您可以将模型部署为在线服务。

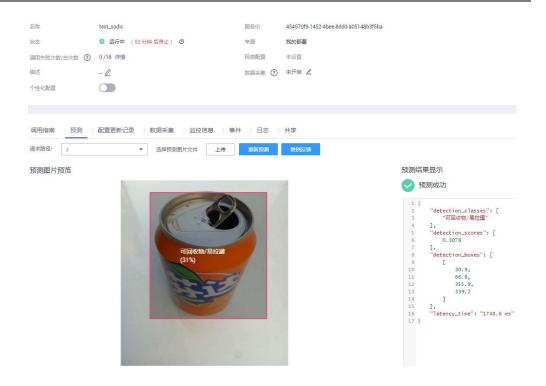
部署为在线服务具体步骤如下。

- 1. 在 ModelArts 左侧导航栏中选择"部署上线 > 在线服务", 然后点击页面中的"部署"。
- 2. 在部署页面填写参数,其中在"模型列表"选择要导入的模型及版本。



3. 点击"下一步",参数确认无误后,点击"提交"。

创建成功后,您可以在在线服务列表中查看部署进度,当状态变为"运行中"后,点击服务名称,进入详情页面,点击"预测",上传图片进行测试,如下图所示。



8 提交参赛模型

模型部署测试成功后,您可以将模型发布到大赛,然后大赛后台会进行自动判分。

步骤一:在 ModelArts 平台上将模型发布到大赛

1. 在 ModelArts 控制台左侧导航栏中选择"模型管理"、"模型",选择页面右侧的模型,点击三角下拉列表展开模型详情,再点击"发布"、"参赛发布",如下图所示。



2. 在参赛模型提交页面填写参数,选择比赛项目和比赛阶段,然后点击"确定", 判分任务的运行结果会以邮件的方式发送到您的华为云账号绑定的邮箱。



13

步骤二:在竞赛平台上查看得分

在大赛平台上点击提交作品,查看作品得分情况。



14

9

在 ModelArts Notebook 中进行开发调试

如果您需要修改模型训练和推理代码,或者想测试模型的效果,可以在 ModelArts Notebook 中进行开发调试。Notebook 中已安装主流的 TensorFlow、MXNet 和 PyTorch 等深度学习框架,以及常用的 Python 模块,是一个开箱即用的在线 AI 开发环境,有不同的 CPU 和 GPU 规格可选。

如果 Notebook 中已安装的 Python 模块不满足您的要求,您可以在其中使用 pip install 安装自己需要的模块,详情请参考《ModelArts 常见 FAQ-V1.0》。

创建 Notebook

- 1. 在 ModelArts 控制台左侧导航栏选择"开发环境 > Notebook", 点击页面中"创建"。
- 2. 在创建 Notebook 页面填写参数,可参考下图填写配置参数。



注意

- "存储配置"建议选择"EVS",这样在 Notebook 实例/home/ma-user/work 目录下的内容能在 notebook 实例关机重启后还能被保存,其他目录下的内容在关机重启后会丢失;
- Notebook 提供 5G 的免费存储空间,超出 5G 的部分在 notebook 关机后也会持续收费;
- 可以分别创建一个 CPU 和一个 GPU 的 Notebook,不必须用到 GPU 的代码调试用 CPU 类型的 notebook 即可,收费更低;
- 3. 点击"下一步"进行规格确认,确认无误后点击"提交"。
 Notebook 创建完成后,可点击 Notebook 作业列表操作栏中"打开",然后进入
 Notebook。

修改模型训练和推理代码

打开 Notebook 后,点击 "Files"页签右上角 "New",选择引擎后进入代码开发界面,然后您可以进行模型训练脚本的编写和运行。

说明: Notebook 支持 TensorFlow、MXNet、PyTorch 等深度学习引擎,您可以基于自身使用习惯来选择。

大赛提供的 baseline 代码已经在 ModelArts Notebook 中编写并调试好模型训练脚本, 您可以直接使用,具体使用方法请查看代码 src/train.py 中开头的注释。

如果您需要自己编写训练脚本,也建议先熟悉大赛 baseline 代码,尤其要注意其中涉及到 OBS 数据交互的代码和 config.json、customize_service.py 两个脚本。

批量图片测试模型

1. 打开已创建好的 Notebook, 点击 upload, 选择已下载好的大赛 baseline 代码 src.zip, 上传,如下图所示,已将 src.zip 上传到了 competition 目录下。



- 2. 点击"New"、"Terminal",将打开一个 linux 终端,复制粘贴执行如下命令: cd /home/ma-user/work/competition/; unzip src.zip 这行命令的意思是进入到上一步的 competition 目录,并解压 src.zip。
- 3. 回到 Notebook 页面,进入到刚解压的 src 目录,找到 test_model.ipynb 文件,点击打开,参考该文件内的说明文档即可以实现对 OBS 上的训练输出模型进行测试,可以进行单张图片测试,也可以计算在测试集上的 mAP 值。AP 的计算方式是参考 VOC2010 之后的方法(https://github.com/rbgirshick/py-faster-rcnn/blob/master/lib/datasets/voc_eval.py),详情请查看 src/cal_mAP.py。

10 释放资源

使用完 ModelArts 相关计算资源后,请务必及时释放资源,否则会造成资源浪费甚至账号欠费。

ModelArts 以下模块会涉及云资源的消耗并会按需或按周期产生费用:自动学习训练和部署,Notebook,训练管理,TensorBoard,在线服务。使用完需要"停止"自动学习部署、Notebook、TensorBoard、在线服务。

停止所有需消耗云资源的作业后,可在 ModelArts 总览页面确认所有计费项都为"0"。



ModelArts 详细计费说明可参见: https://support.huaweicloud.com/price-modelarts/modelarts_07_0002.html。

注意

 公有云的按需使用是采用后付费模式,即先使用再付费,一个小时会出一次话单, 比如: 2019.6.24 04:00~05:00 这段时间使用的费用,在 05:00 时刻采集完成后, 才会出话单,进行扣费。

11 Baseline 代码说明

本 baseline 代码基于 yolo3(https://github.com/qqwweee/keras-yolo3)实现,您可以对比原版代码了解两者之间的差异。

Baseline 代码以压缩包形式提供,解压后得到 src 目录,关键文件说明见下表。

名称	说明
train.py	模型训练启动文件
deploy_scripts	其中包含模型配置文件 config.json 和模型推理脚本 customize_service.py,是华为云 ModelArts 平台上做 模型推理的专属、必备文件
inference.py	基于 customize_service.py 修改得到的模型推理脚本,在 ModelArts 平台内和平台外都可以运行
model_data	存放训练任务相关的类别名文件和 anchors 参数文件
yolo3	存放 yolo3 模型定义脚本
test_model.ipynb	在 ModelArts Notebook 中用于模型测试、计算 mAP的 ipynb 脚本
cal_mAP.py	基于 voc_eval.py 修改得到的 mAP 计算脚本,您可以对比原版代码了解两者之间的差异
其他文件	原版 keras-yolo3 代码自带的文件