

如何在 ModelArt 运行深度学习案例

一、准备数据集和源代码

本次所选案例是“基于 ResNet50 实现毒蘑菇识别实战”：

学习视频：

https://www.bilibili.com/video/BV1ny4y1r7HA?from=search&seid=8373008762929386214&spm_id_from=333.337.0.0

源码：

<https://github.com/mindspore-ai/mindspore-21-days-tutorials/tree/main/chapter3>

案例学习 PPT：

<https://pan.baidu.com/s/1rhWQsbOFO-yV3bCih2rwg>

提取码：eqbp

观看上面资料，自行下载源代码和数据集。

数据集如下所示：

> 此电脑 > DATA (D:) > 360download > mushrooms >			
名称	修改日期	类型	
eval	2021/11/1 16:18	文件夹	
test	2021/11/1 16:18	文件夹	
train	2021/11/1 16:18	文件夹	

> 此电脑 > DATA (D:) > 360download > mushrooms > train			
名称	修改日期	类型	
Agaricus	2021/11/1 16:18	文件夹	
Amanita	2021/11/1 16:18	文件夹	
Boletus	2021/11/1 16:18	文件夹	
Cortinarius	2021/11/1 16:18	文件夹	
Entoloma	2021/11/1 16:18	文件夹	
Hygrocybe	2021/11/1 16:18	文件夹	
Lactarius	2021/11/1 16:18	文件夹	
Russula	2021/11/1 16:18	文件夹	
Suillus	2021/11/1 16:18	文件夹	

源代码如下所示：

« mindspore-21-days-tutorials-main > chapter3 > resnet_ascend >					🔍 搜
名称	修改日期	类型	大小		
.idea	2021/12/2 16:56	文件夹			
ckpt_files	2021/11/1 16:00	文件夹			
src	2021/12/2 16:20	文件夹			
moxing_test.py	2021/11/13 12:55	Python File	1 KB		
README.md	2021/5/12 10:56	MD 文件	3 KB		
resnet50_eval.py	2021/5/12 10:56	Python File	4 KB		
resnet50_predict.py	2021/11/12 21:07	Python File	6 KB		
resnet50_train.py	2021/11/12 0:04	Python File	7 KB		

二、使用 ModelArt，将数据集传输到 OBS（对象存储服务）中

<https://www.huaweicloud.com/>

登录自己的华为云账号，请先确认自己有代金券，再进行下面的操作

在搜索栏搜索 OBS，点击进入 OBS 控制台。



然后点击左侧菜单栏 桶列表，在右边界面点击 新建桶。



请填写下面这些信息，最后点击创建桶。

复制桶配置

选择源桶

该项可选。选择后可复制源桶的以下配置信息：区域 / 数据冗余策略 / 存储类别 / 桶策略 / 默认加密 / 归档数据直读 / 企业项目 / 标签。

区域

华北-北京一

不同区域的云服务产品之间内网互不相通；请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。如何选择区域

桶名称

请输入桶名称

不能和本用户已有桶重名

不能和其他用户已有的桶重名

创建成功后不支持修改

默认存储类别

标准存储

适合高性能、高可靠、高可用，频繁访问场景

多AZ存储

单AZ存储

图片处理

低频访问存储

适合高可靠、低成本，较少访问场景

多AZ存储

单AZ存储

图片处理

归档存储

适合长期存储，基本不访问场景

单AZ存储

费用参考

创建桶时选择的存储类别会作为上传对象的默认存储类别。了解存储类别差异

桶策略

私有

公共读

公共读写

复制桶策略

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。

默认加密

开启默认加密

免费

建议开启默认加密，透明管理全免费，核心数据更安全。

归档数据直读

开启

关闭

关闭归档直读，归档存储类别的数据要先恢复才能访问。归档存储数据恢复和访问会收取相应的费用。价格详情

标签

如果您需要使用同一标签标识多种云资源，即所有服务均可在标签输入框下拉选择同一标签，建议在TMS中创建预定义标签。查看预定义标签

标签键

标签值

您还可以添加10个标签。

创建阶段

使用阶段

OBS桶：创建免费

按需/资源包计费

OBS计费说明

立即创建

创建完之后，将数据集和源代码上传至 OBS 的桶中，在 OBS 界面点击创建的桶。

桶名称	存储类别	区域	数据冗余存储策略	存储容量	Data+	对象数量	创建时间	操作
soul-resnet-new	标准存储	华北-北京四	多AZ存储	10.41 GB		187	2021/11/11 12:50:11 GMT+08:00	修改存储类别 删除
soul-resnet	标准存储	华北-北京四	多AZ存储	7.68 GB		398,633	2021/11/01 16:03:24 GMT+08:00	修改存储类别 删除
mindspore-retail-goods	标准存储	华北-北京四	多AZ存储	0 byte		0	2021/10/25 15:38:01 GMT+08:00	修改存储类别 删除

然后点击左侧菜单栏 对象，再点击上传对象，将数据集、源代码上传

< mindspore-retail-goods

概览

对象

用量统计

访问权限控制

基础配置

域名管理

跨区域复制

数据回源

数据处理

桶清单

Data+

对象

已删除对象

碎片

对象是数据存储的基本单位，在OBS中文件和文件夹都是对象。您可以上传任何类型（文本、图片、视频等）的文件，并在桶中对这些文件进行管理。了解更多

若需要将对象移动到桶内其他位置，推荐下载使用OBS Browser+图形化管理工具。

上传对象

新建文件夹

恢复

更多

名称

存储类别

大小

加密状态

操作

暂无数据

可以直接拖拽本地文件至对应处上传，上传文件有数量大小限制，如何无限制上传，可以借助其他工具，可以考虑 obsutil，那如何使用 obsutil 作为一个小问题，大家自己摸索。

上传对象

超过5GB如何上传?

上传操作将产生请求费用，上传成功后将产生存储费用。

存储类别

标准

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。

对象默认与桶的存储类别相同，也可以根据适用场景修改。[了解更多](#)

上传对象

注意：桶内如有同名文件/文件夹，将被新上传的文件/文件夹覆盖。



拖拽本地文件或文件夹至此处，或[添加文件](#)
(单次最多支持100个文件同时上传，总大小不超过5GB)

加密

当前桶已开启默认加密，在上传对象至当前桶时，默认情况下将对其进行加密。

KMS密钥

KMS-p059

上传

取消

上传好了之后，就可以在 OBS 中看到，我展示我其他桶的内容，你们的桶中应该只有你们上传的数据集和源代码。

若需要桶为移动设备端内其他位置，请[下载使用OBS Browser+图形化管理工具](#)。

上传对象

新建文件夹

依赖

更多

<input type="checkbox"/>	名称	存储类别	大小	加密状态	校验状态	最后修改时间	操作
<input type="checkbox"/>	1108_retail_goods_dataset	--	--	--	--	--	分享 复制
<input type="checkbox"/>	dataset	--	--	--	--	--	分享 复制
<input type="checkbox"/>	human36m_data	--	--	--	--	--	分享 复制
<input type="checkbox"/>	log	--	--	--	--	--	分享 复制
<input type="checkbox"/>	mushrooms	--	--	--	--	--	分享 复制
<input type="checkbox"/>	output	--	--	--	--	--	分享 复制
<input type="checkbox"/>	resnet	--	--	--	--	--	分享 复制

分享 复制 使用过程

接下来我们搜索华为云的 ModelArt，进入管理控制台。

ModelArts

概览

功能模块

强化学习

天筹求解器

盘古大模型

自定义推荐系统

入门&案例

资源&学习

活动&定价

生态合作

AI开发平台ModelArts

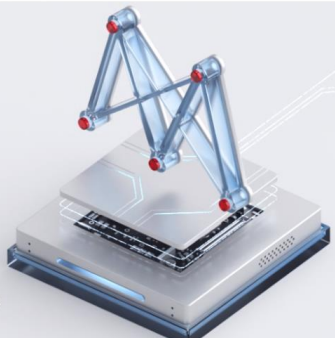
ModelArts 是面向开发者的一站式 AI 开发平台，为机器学习与深度学习提供海量数据预处理及交互式智能标注、大规模分布式训练、自动化模型生成，及端-边-云模型按需部署能力，帮助用户快速创建和部署模型，管理全周期 AI 工作流。

按需/包周期付费可选，最低¥0.00/小时

管理控制台

购买套餐包

帮助文档



点击左侧算法管理，创建一个算法

首页

自动学习

数据管理

开发环境

算法管理

训练作业

模型管理

部署上线

AI Gallery

专家资源站

全局配置

我的算法

我的订阅

创建

输入

名称	标签	大小 (MB)	描述	创建时间	操作
algorithm-a795		0.49	resnet	2021/12/02 16:36:51 GMT+08:00	复制 删除 创建训练作业
algorithm-d9a8		0.02	resnet-1117_train	2021/11/17 14:05:27 GMT+08:00	复制 删除 创建训练作业
algorithm-d9ce		0.04	...	2021/11/13 12:58:18 GMT+08:00	复制 删除 创建训练作业
algorithm-1684		0.43	resnet50_predict	2021/11/12 17:36:15 GMT+08:00	复制 删除 创建训练作业
algorithm-5933		0.42	...	2021/11/11 17:48:30 GMT+08:00	复制 删除 创建训练作业
algorithm-fc27		0.42	...	2021/11/10 23:17:47 GMT+08:00	复制 删除 创建训练作业
algorithm-d426		0.03	15class测试训练	2021/11/10 13:43:53 GMT+08:00	复制 删除 创建训练作业
algorithm-38ff		0.03	15类增量训练	2021/11/02 19:09:41 GMT+08:00	复制 删除 创建训练作业
algorithm-30ed		0.03	15类增量训练	2021/11/02 18:15:06 GMT+08:00	复制 删除 创建训练作业
algorithm-b78c		0.03	预训练网络-有带训练	2021/11/02 15:52:29 GMT+08:00	复制 删除 创建训练作业

10 总条数: 12 2 刷新 1

进行一些简单配置，如下所示：

名称

identity-mushroom

ID

6449d34d-59d4-42fb-9599-a4adb2ef50e4

描述

识别蘑菇

★ 创建方式

自定义脚本

自定义镜像

选择常用引擎创建训练作业。每个算法的代码目录最多支持1000个文件，文件深度不超过32，总大小不超过5GB。

★ AI引擎

Ascend-Powered-Engi...

mindspore_1.3.0-cann_5.0.2-py...

☐ 显示旧版引擎

[引擎升级说明](#)

★ 代码目录

/soul-resnet/resnet_ascend/

选择

?

★ 启动文件

/soul-resnet/resnet_ascend/resnet50_train.py

选择

?

输入数据配置

为您的算法定义处理“输入数据”的参数，在您的算法代码中需要解析该参数获取到训练的数据集

映射名称

数据来源1

代码路径参数

data_url

×

添加约束

☐ 是

☒ 否

添加输入数据配置

输出数据配置

为您的算法定义处理“输出数据”的参数，在您的算法代码中需要解析该参数获取到训练的输出路径

映射名称

输出数据1

代码路径参数

train_url

×

添加输出数据配置

超参

删除

清空

?

超参

添加输出数据配置

删除

清空

?

	名称	类型	默认值	必需	描述	操作
<input type="checkbox"/>	data_url	String		约束 是		删除
<input type="checkbox"/>	train_url	String		约束 是		删除

增加超参

支持的策略

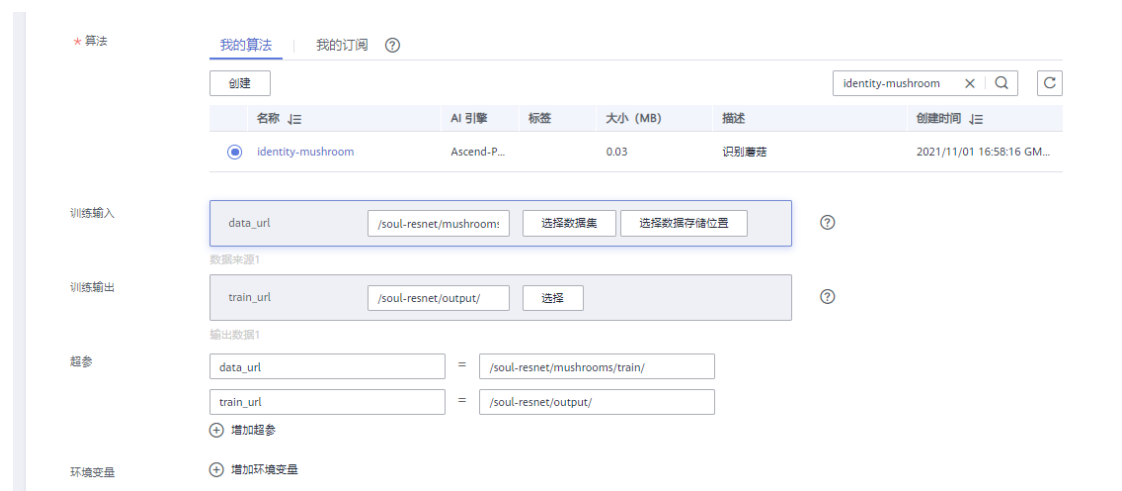
☐ 自动搜索(S)

添加训练约束

☐ 是

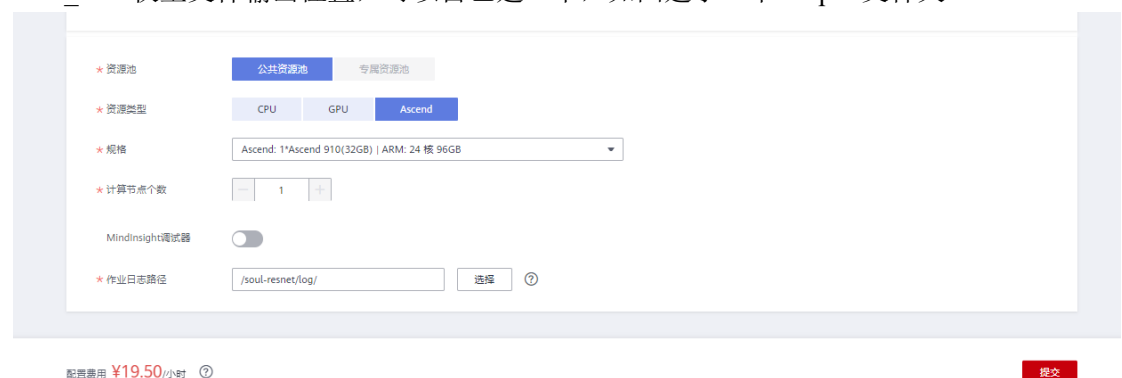
☒ 否

算法创建完成，根据算法创建一个训练作业：



data_url: 训练集位置，即 mushrooms 中 train 文件夹

train_url: 权重文件输出位置，可以自己建一个，如图建了一个 output 文件夹



资源选择 Ascend，因为代码是使用的 Ascend，注意选择 1 * Ascend910 (19.5r/h)

日志路径自己建一个文件夹，选择即可

然后提交，等待代码运行即可。

注：对华为云有使用问题，请现在华为云内搜索关键词看有无教程；若没找到，请在华为云上提交工单，会有工程师 24 小时回复，也可以在群里大家讨论讨论。

三、修改代码以使用 obs 桶中的数据
在创建任务时的补充说明

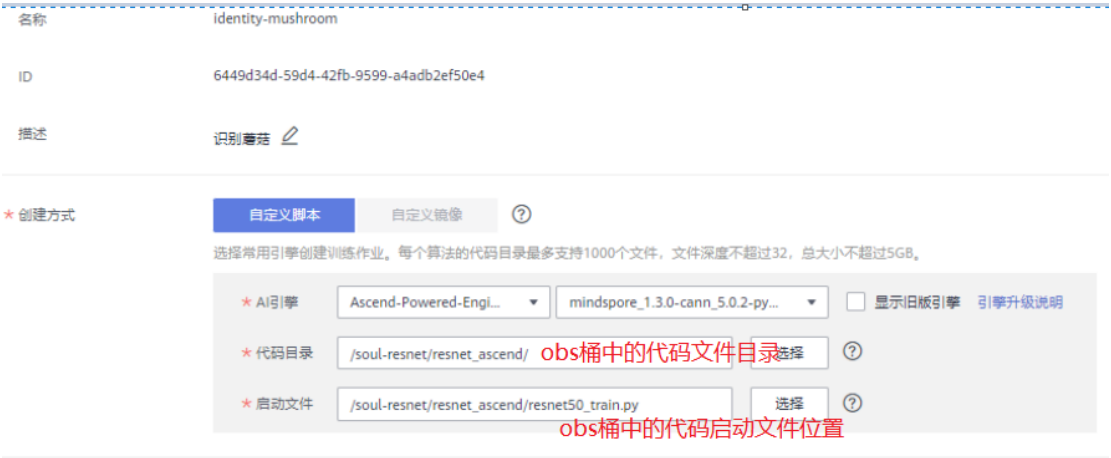


图 1 在创建任务时的补充说明

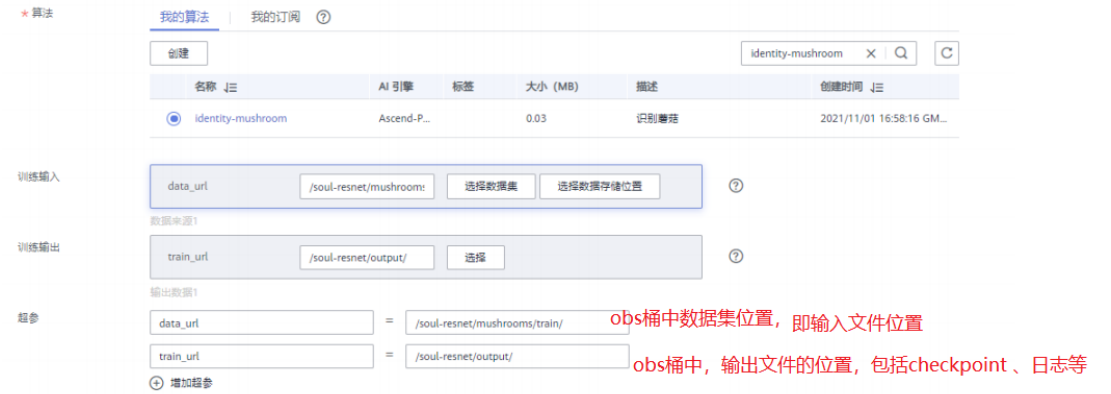


图 2 在配置输入输出文件位置时的补充说明

以上两步都曾用到 **obs** 桶进行配置。
但要在 **python** 代码中实际读取文件时，上图配置的 **data_url** 将会报文件找不到错误。
这是因为 **modelart** 实际运行机器与 **obs** 桶在不同的服务器上。**Modelart** 执行任务时，只会把 **python** 文件读到 **modelart** 服务器上，其它文件没有同步拷贝过去。

故在做出图 2 的配置时，还需修改 **obs** 桶中的源文件。在本例子中，需要额外添加多行代码

```
import moxing as mox //导入库
def resnet50_train(args_opt):
    local_data_path = '/cache/data' //将 obs 数据拷贝到的 modelart 服务器的临时位置
    localckpt_path = '/cache/ckpt_file' //将模型的输出文件（checkpoint）保存到的 modelart 服务器的临时位置
    .
    .
    .
    mox.file.copy_parallel(src_url=args_opt.data_url, dst_url=local_data_path)//将 obs 数据
    拷贝到的 modelart 服务器的 '/cache/data'
    /soul-resnet/mushrooms/train/
//省略中间的训练过程代码
```

.
.

```
mox.file.copy_parallel(src_url=local_ckpt_path, dst_url=args_opt.train_url)
```

//训练结束后，将 modelart 服务器的 '/cache/ckpt_file' 下文件 拷贝回 obs 桶中 `/soul-resnet/output/` 位置。注意：

必须将文件保存回 obs 桶中，否则训练结束 modelart 服务器将自动回收训练时产生的文件，即找不回来了。

obs 桶中数据集位置，所有通过文件读取 和 文件保存方式获得的信息(非 python 文件) 都只能通过给定 mox 库同步拷贝方式。

如果觉得麻烦。推荐 四、使用 notebook 方式使用 modelart 中方式。

四、使用 Notebook 方式使用 modelart

见以下链接。其中也有使用 obs 桶中的数据。

[使用 Mindstudio 调用 Modelarts 进行模型训练 import moxing as mox-CSDN 博客](#)