|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | | |  |
|  | | |
|  | | |
| 方圆ok1 | | | |
|  | 创新应用大赛ModelArts平台使用指导  **“华为云-东吴杯”数字化转型创新应用大赛Baseline** | | 附件1-16K |
|  | |
| **文档版本** | **01** |
| **发布日期** | **2021-06-15** |
| **华为技术有限公司** | |

|  |
| --- |
| 版权所有 © 华为技术有限公司2020。 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  附件3-版权声明页图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](mailto:support@huawei.com) |
| 客户服务电话： | 4008302118 |

目录

[1 文档导读 4](#_Toc34041409)

[2 获取竞赛数据集 5](#_Toc34041410)

[2.1 在OBS创建桶及文件夹 5](#_Toc34041411)

[2.2 在ModelArts全局配置中添加访问密钥 6](#_Toc34041412)

[2.3 创建Notebook并通过代码拷贝竞赛数据集 7](#_Toc34041413)

[3 在Notebook编写训练脚本 10](#_Toc34041414)

[4 将生成的模型导入至模型管理 13](#_Toc34041415)

[5 将模型部署为在线服务 15](#_Toc34041416)

[6 提交模型判分 17](#_Toc34041417)

[7 释放资源 19](#_Toc34041418)

# 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-东吴杯”2021数字化转型创新应用大赛赛题数据处理、训练建模、推理测试以及结果的提交。

本文档包含使用ModelArts完成“华为云-东吴杯”2021数字化转型创新应用大赛赛题的全流程指导，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来完成比赛。

在ModelArts上实操本文档前，务必确保已经领取并使用华为云现金券，避免造成ModelArts账户欠费。领取指引：<https://competition.huaweicloud.com/information/1000041490/support?track=107>

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 |
| --- | --- |
| [获取竞赛数据集](#ZH-CN_TOPIC_0169740162) | 获取赛题数据集。 |
| 通过训练作业的方式复现baseline | 通过ModelArts训练作业，进行模型的训练。 |
| [将生成的模型导入至模型管理](#ZH-CN_TOPIC_0169740167) | 将训练生成的模型导入至模型管理。 |
| [将模型部署为在线服务](#ZH-CN_TOPIC_0169740168) | 将模型部署为在线服务，查看预测结果。 |
| [提交模型判分](#ZH-CN_TOPIC_0169740169) | 提交测试好的模型进行评分。 |
| [释放资源](#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。 |
| [附录](#附录) | 本baseline中的训练脚本、推理代码，训练数据。 |

# 获取竞赛数据集

在本次大赛中，您可能会使用华为云提供的两种服务：AI开发平台ModelArts和对象存储服务OBS。

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。

ModelArts面向开发者提供一站式AI开发平台。提供基于GPU和CPU的Notebook或训练作业等模型开发能力；提供模型部署、模型发布能能力。

竞赛数据集已上传至华为云对象存储服务（OBS）桶中，您首先需要在OBS创建一个属于您的空桶；然后将竞赛数据集和baseline代码拷贝至您创建的OBS桶中。

竞赛数据集下载链接：<https://marketplace.huaweicloud.com/markets/aihub/datasets/detail/?content_id=cb811923-2d26-4c7e-9be0-32a2d109549d> 完成2.1节相关操作后，可将数据集下载到data目录下。

## 在OBS创建桶及文件夹

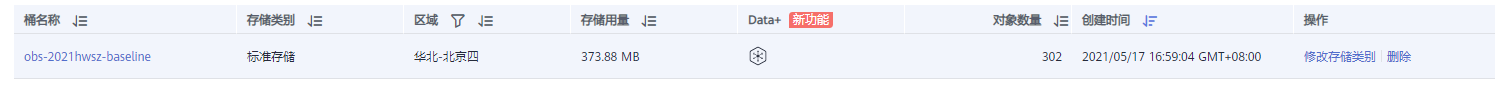
首先在OBS创建桶及文件夹，用于存放竞赛数据集、代码文件及模型训练输出的模型文件，具体操作如下。

1. 登录OBS管理控制台<https://storage.huaweicloud.com/obs/?&region=cn-north-1&locale=zh-cn#/obs/buckets>。
2. 单击页面右上方“创建桶”，在弹出的页面中选择区域、自定义桶名称。本baseline创建的桶名称为obs-2021hwsz-baseline，选手可自由更改。

注意

* “区域”请选择“**华北-北京四**”。
* OBS桶名称全局唯一，不能与已有的任何桶名称重复。
* 注意“多AZ”选择“关闭”。

1. 单击“立即创建”。



1. 在OBS桶列表单击桶名称，然后单击左侧导航栏中“对象”，单击页面中“新建文件夹”，自定义文件夹名称后，单击“确定”。

* 新建一个文件夹存放竞赛数据集(data)；本baseline存放竞赛数据集路径为obs-2021hwsz-baseline/data/。
* 新建一个文件夹存放竞赛baseline代码(code)；本baseline存放竞赛数据集路径为obs-2021hwsz-baseline/code/。
* 新建一个文件夹存放训练完成的模型(ckpt)；本baseline存放竞赛数据集路径为obs-2021hwsz-baseline/ckpt/。
* 新建一个文件夹存放用于发布的模型和代码 (submission)；本baseline存放竞赛数据集路径为obs-2021hwsz-baseline/submission/。



1. data目录下存放训练数据集。注意，这里使用zip压缩格式。不建议选手解压，避免造成训练时造成与OBS过多的交互访问，超出OBS免费请求限制。



1. code目录下存放baseline代码，如下图所示。



1. ckpt用于保存训练好的模型文件，具体使用方式参见第三章，此处为空目录即可。
2. submission用于保存发布模型的相关文件，具体使用方式参见第四章。在该目录下创建model文件夹，后续将model文件夹作为发布模型的目录。此处为空目录即可。



## 在ModelArts全局配置委托授权

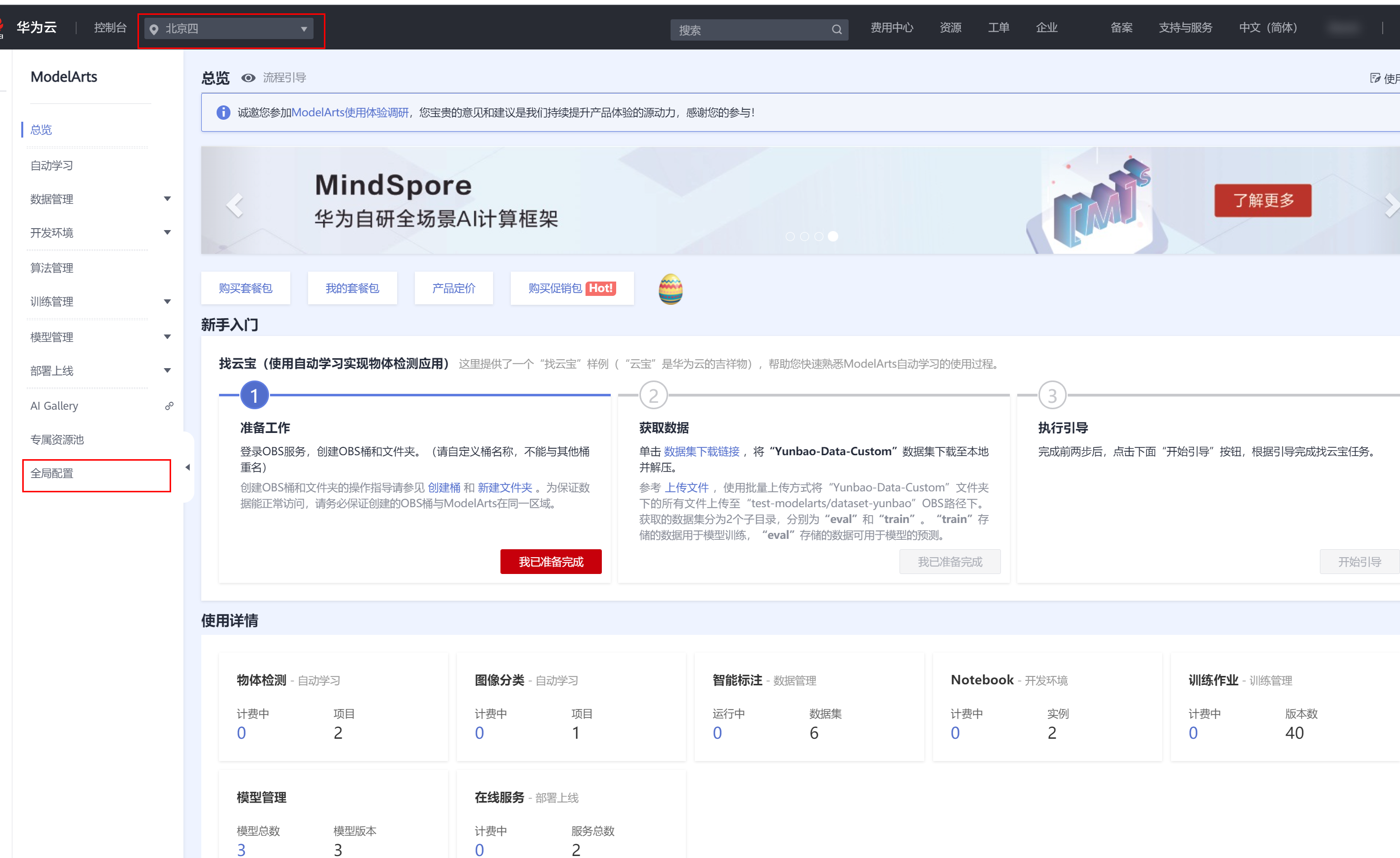
2.1节主要介绍了如何使用OBS桶以及我们桶中存储的文件。接下来本节将介绍如何配置授权，使得ModelArts可以访问OBS。

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，您需要使用委托授权（建议）或者通过访问秘钥授权。

接下来将主要介绍如何使用委托，使得OBS、SWR、IEF等依赖服务ModelArts依赖委托给ModelArts。

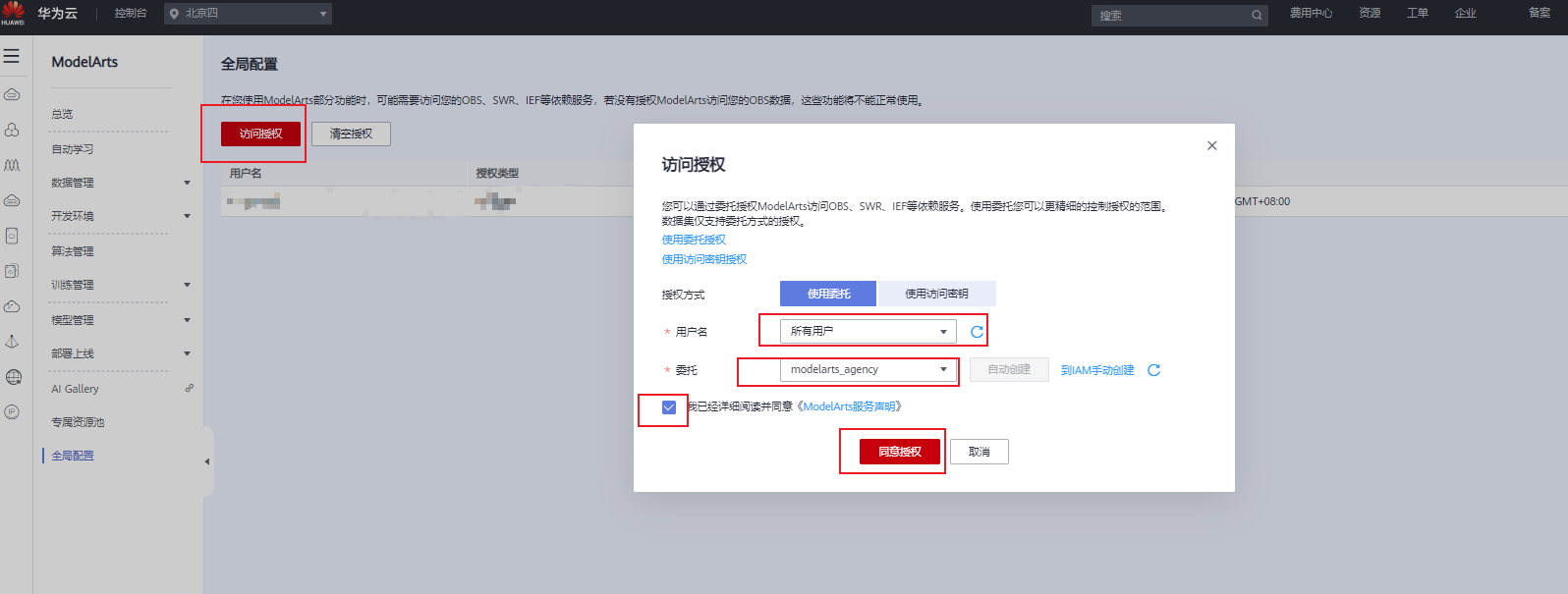
授权配置

使用华为云账号登录ModelArts管理控制台，注意切换为北京四，在左侧导航栏点击“全局配置”，进入“全局配置”页面。

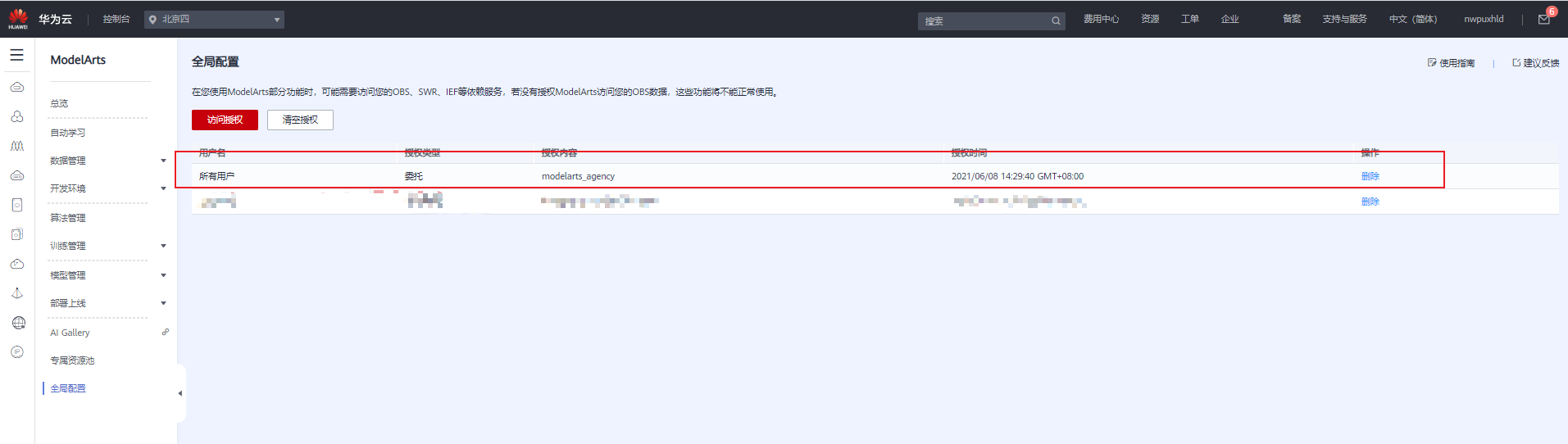


点击访问授权，在弹出的“访问授权”窗口中，“授权方式”选择“使用委托”，选择需要授权的“用户名”及其对应的“委托”。

完成“用户名”和“委托”配置后，勾选“我已经详细阅读并同意《ModelArts服务声明》”，然后单击“同意授权”。



完成配置后，在全局配置列表，可查看到用户的委托配置信息，如下图所示。



至此，我们已经完成了ModelArts中配置委托授权。

说明

* OBS套餐内包括100GB存储、24万次GET请求、6万次PUT请求。通过使用压缩包的形式，可以减少我们与OBS的交互次数。请选手注意余量，避免造成扣费。
* 委托授权生效可能需要1分钟左右的时间。
* 使用ModelArts前注意确认region为北京四

# 通过训练作业的方式复现baseline

在将上述OBS相关数据和代码准备完成之后，您可以在ModelArts中训练代码。

ModelArts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；baseline是基于PyTorch3.6实现。接下来我们将分别从训练脚本介绍、训练作业提交介绍等两个方面对复现baseline展开介绍。

Baseline代码参见本文档附录。

## 3.1 编写模型训练脚本

本baseline已经提供可训练的模型脚本baseline\_released.py。接下来将主要介绍主函数代码（baseline\_released.py）参数和各个文件的作业。选手熟悉后可以自行优化。

1. 主函数代码参数介绍
   1. data\_url:字符串类型，必须，obs上训练数据存储的路径。3.2节会介绍。
   2. train\_url:字符串类型，必须，obs上存储训练好模型的路径。3.2节会介绍。
   3. init\_method:非必须，提交训练作业格式需要。
   4. num\_gpus:int类型，默认值为1。
   5. train-dir:字符串类型，必须，训练作业中训练数据的存储路径。baseline\_released.py会将obs上data\_url下载到train-dir的父目录。如下图所示。moxing为modelarts自研框架。mox.file.copy\_parallel()可以实现将obs中文件下载到本地路径中。



* 1. validate-dir:字符串类型，非必须。若为None，则默认不执行validation，否则针对指定目录执行验证环节。
  2. ckpt-dir:字符串类型，必须。将每个epoch生成的模型文件保存至该目录下。
  3. num-classes:数字类型，必须。缺陷种类个数。
  4. batch-size: 数字类型，必须。
  5. num-epochs: 数字类型，必须。

1. 各个文件的作用：
   1. baseline\_released.py: 用于复现baseline的主入口函数。
   2. utils: 一些通用的函数封装
   3. customize\_service.py: 发布模型所需的主入口函数。参见第四章。
   4. config.json: 发布模型所需要的配置文件，相关字段参考：<https://support.ulanqab.huawei.com/engineers-modelarts/modelarts_23_0092.html#modelarts_23_0092__table1683418482455>。
   5. pretrained: 存放预训练模型。**请将**[**https://download.pytorch.org/models/fasterrcnn\_resnet50\_fpn\_coco-258fb6c6.pth**](https://download.pytorch.org/models/fasterrcnn_resnet50_fpn_coco-258fb6c6.pth)**下载的模型文件放置在该目录下。**
   6. requirements.txt: baseline代码依赖的一些库名称和版本。
2. 若是本地开发修改后的代码，可直接通过浏览器上传到OBS桶中。如下图所示。

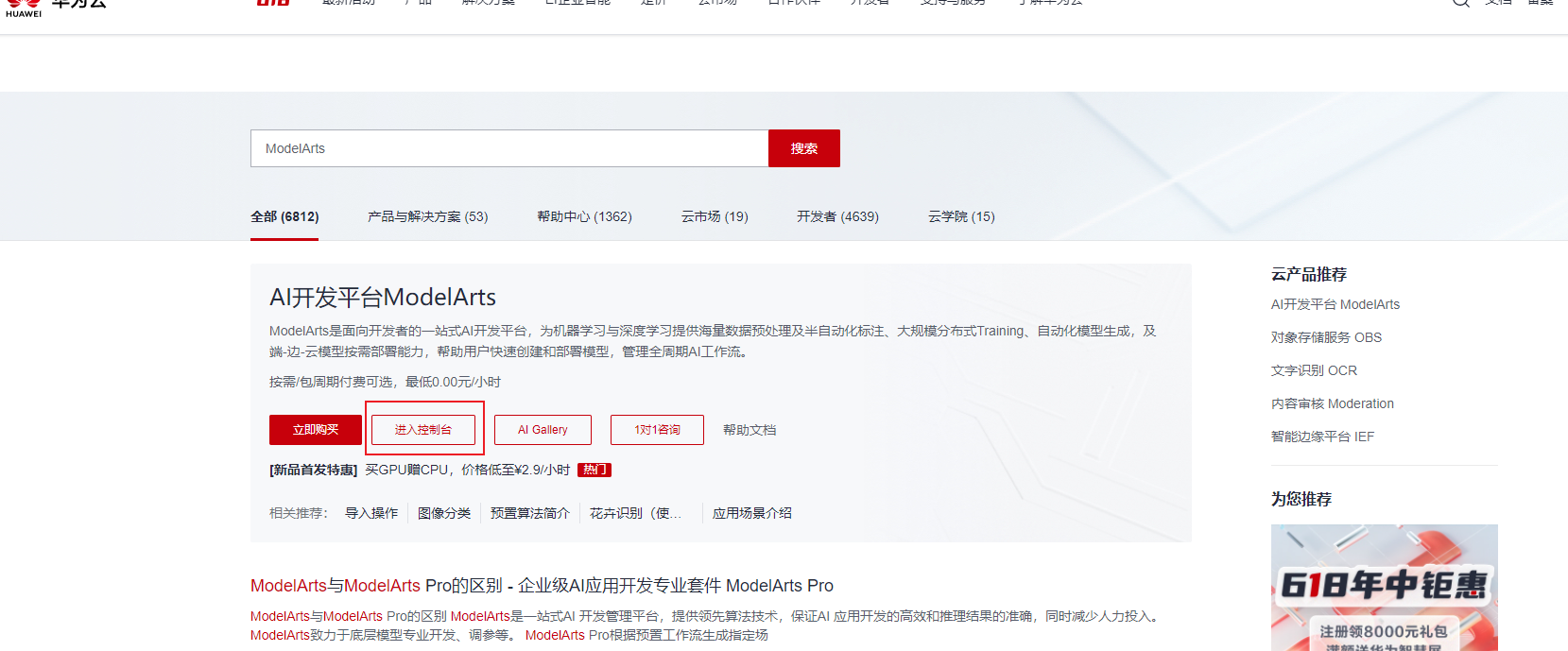




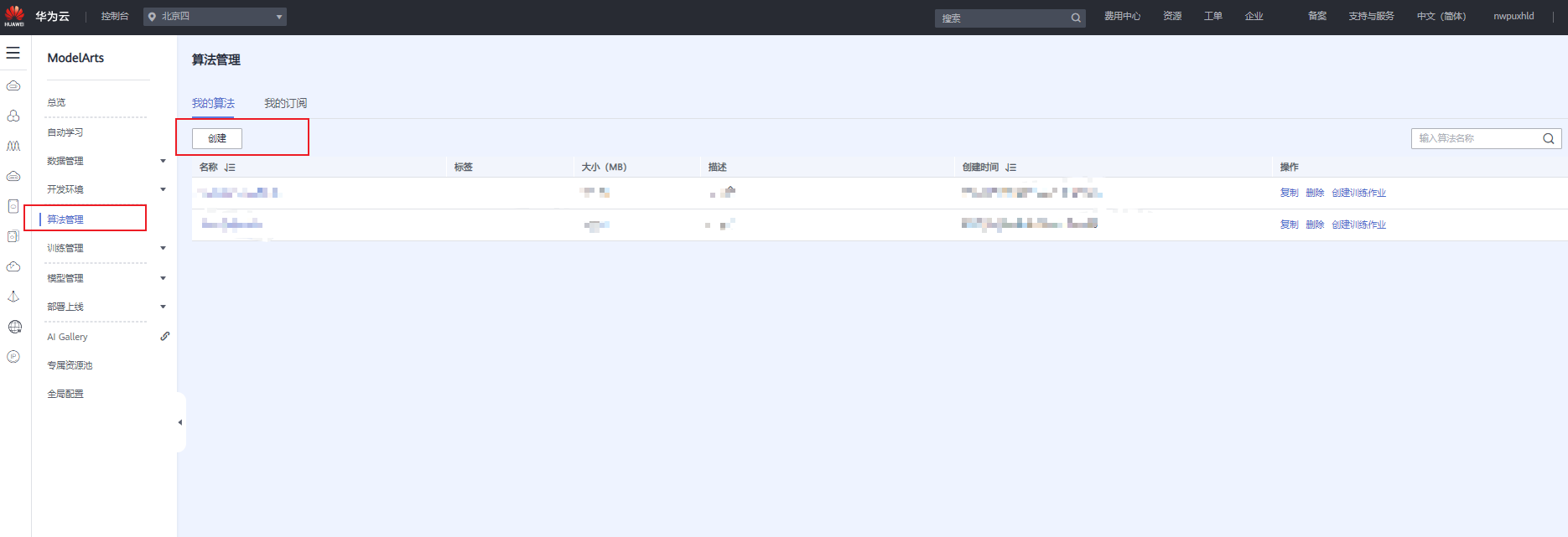
## 3.2 提交训练作业

接下来通过介绍训练作业的方式完成baseline模型的训练。ModelArts提供训练作业和Notebook等方式来训练模型。Notebook由于免费云硬盘额度只有5GB，本次竞赛数据大小为4.87GB。故为了不产生额外费用，本次通过训练作业的方式完成baseline模型训练。若选手有兴趣，可以自行探索notebook方式开发调试训练。

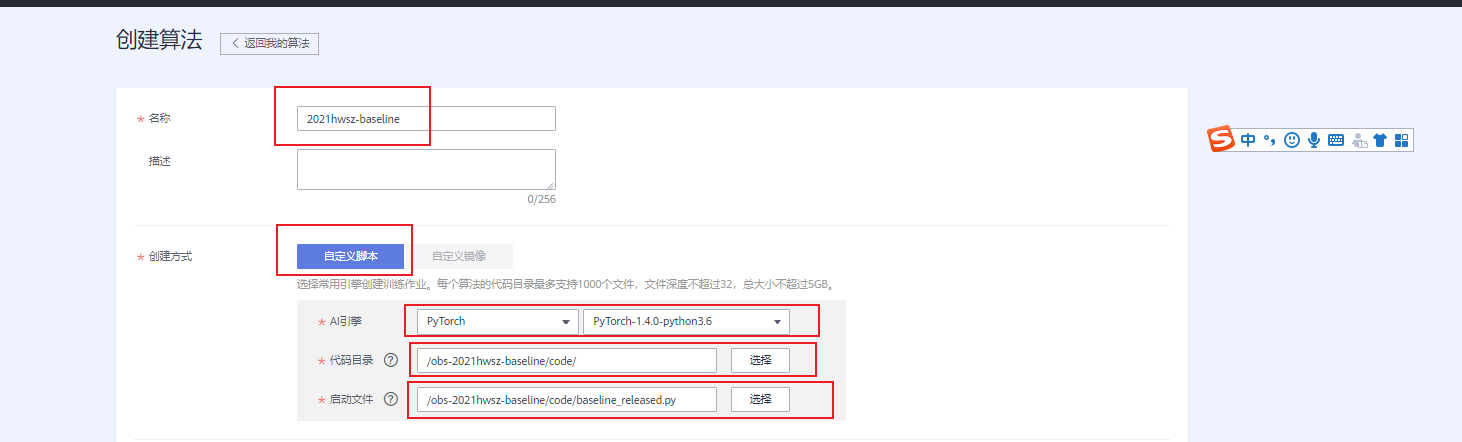
1. 进入ModelArts页面，在华为云搜索MoedlArts，选择进入控制台。



1. 创新新的算法。如下图所示，算法管理->创建。



1. 填写算法名称，AI引擎选择PyTorch-1.4.0-python3.6,代码目录和启动文件的配置如下图所示。



1. 填写启动超参，如下图所示。

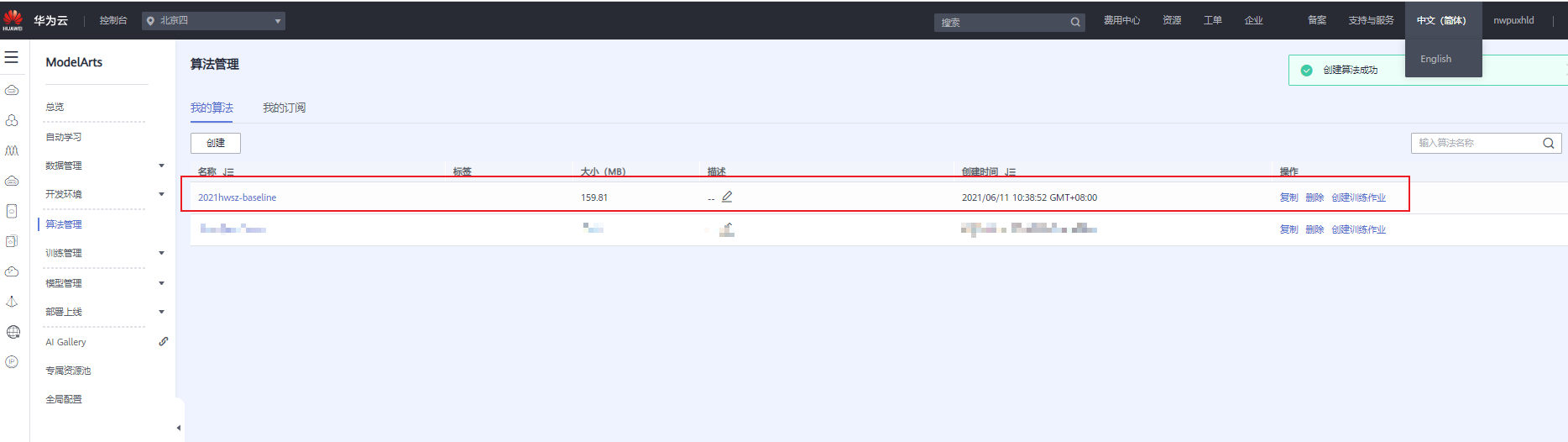


|  |  |
| --- | --- |
| train-dir | /home/work/user-job-dir/data/training |
| num-classes | 10 |
| batch-size | 8 |
| ckpt-dir | /home/work/user-job-dir/code/checkpoints |
| num-epochs | 20 |

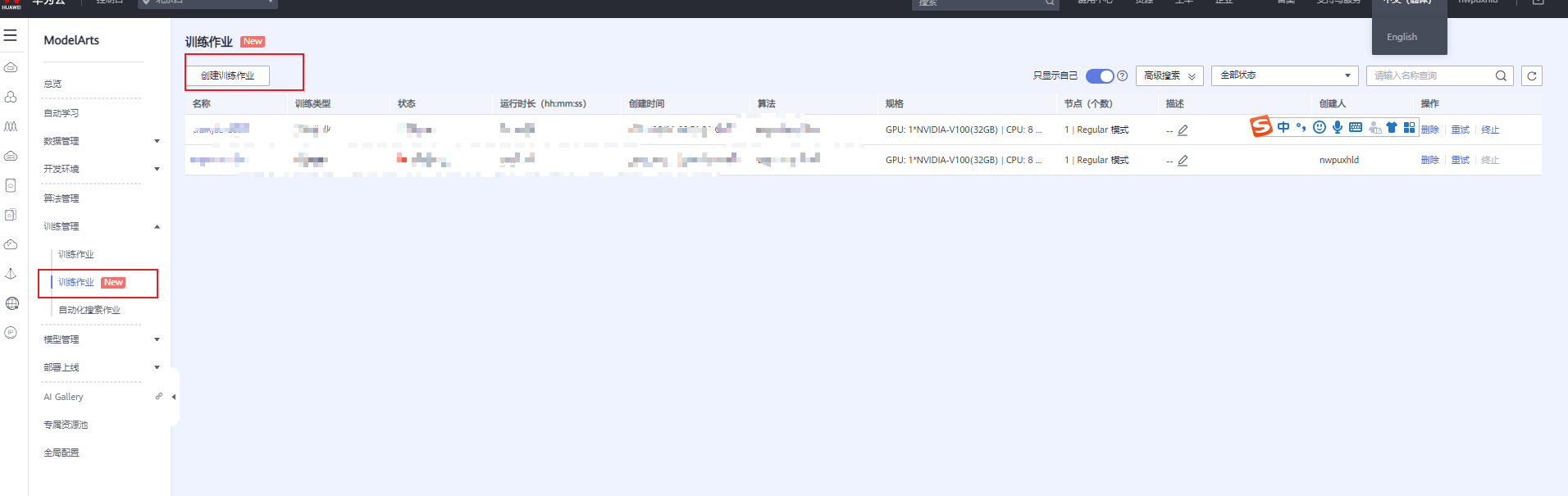
1. 添加一个输出配置项，起名为last\_path。用于存储最后一个pth文件作为评分的模型。如下图所示。



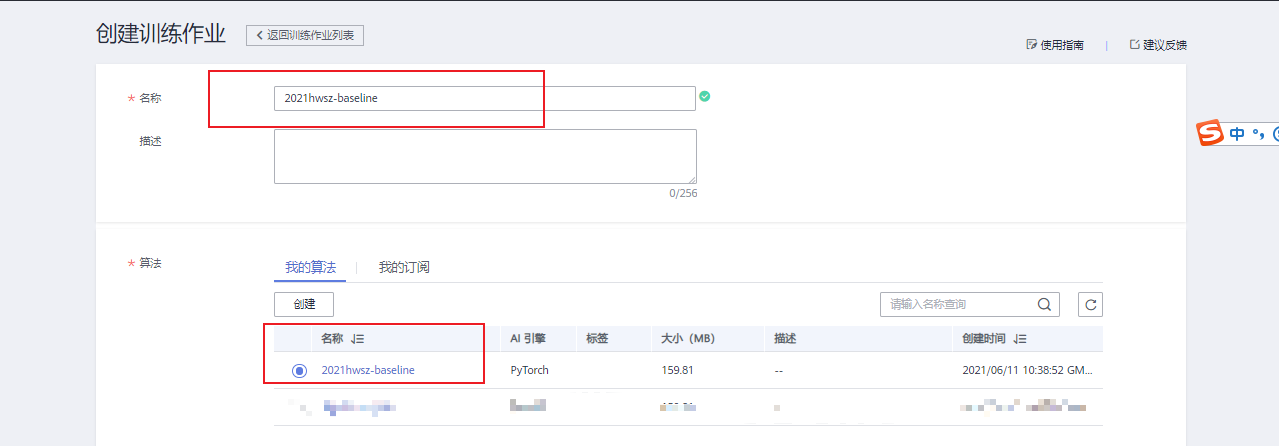
1. 点击提交，创建算法。



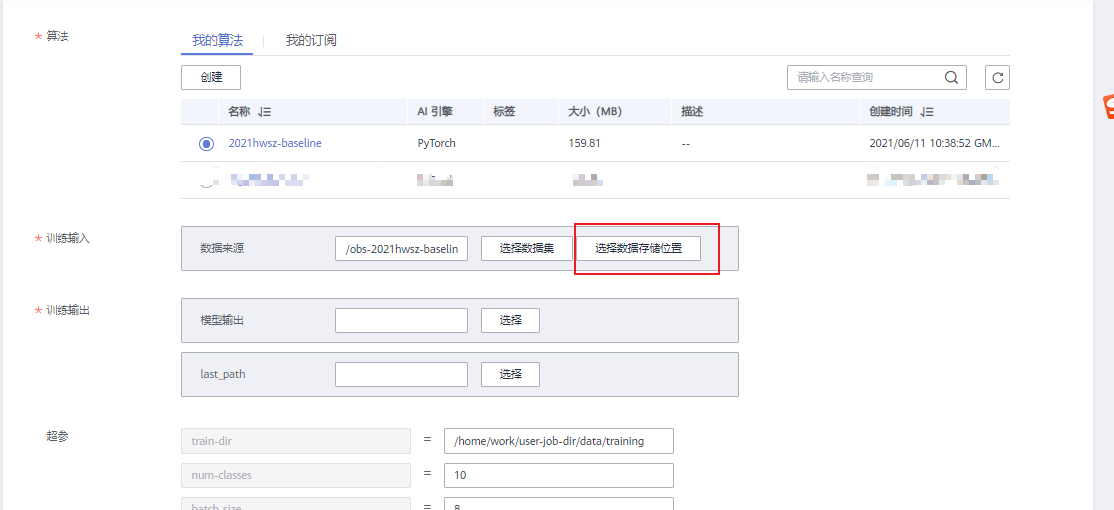
1. 创建新训练作业，如下图所示。训练管理->训练作业（New）->创建训练作业。如果存在训练作业/训练作业（New）请进入训练作业（New），如果只存在训练作业，请进入训练作业。



1. 填写训练作业名称，选择刚创建的算法。如下图所示。



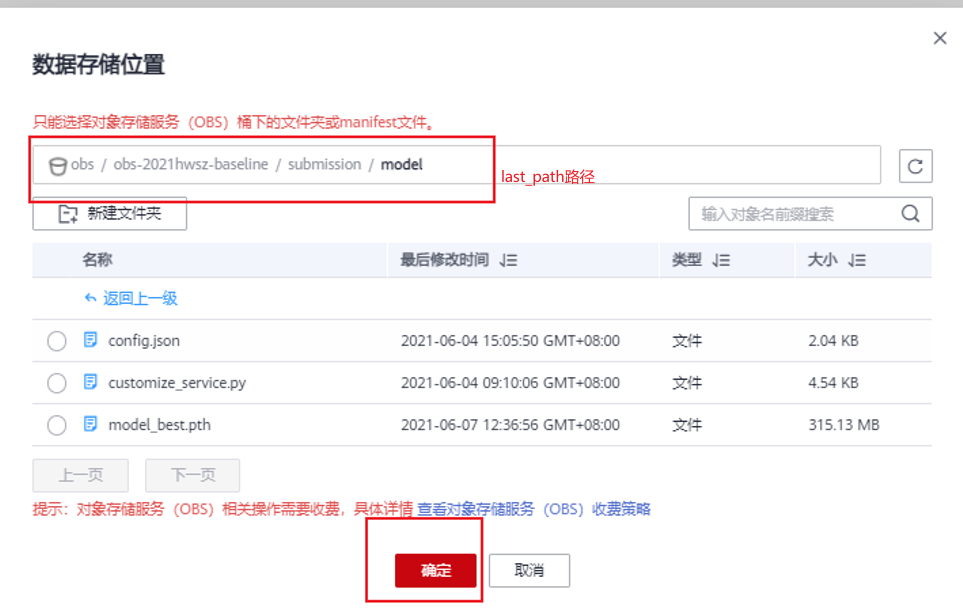
1. 点击选择数据存储位置，选择对应的路径，点击确定。



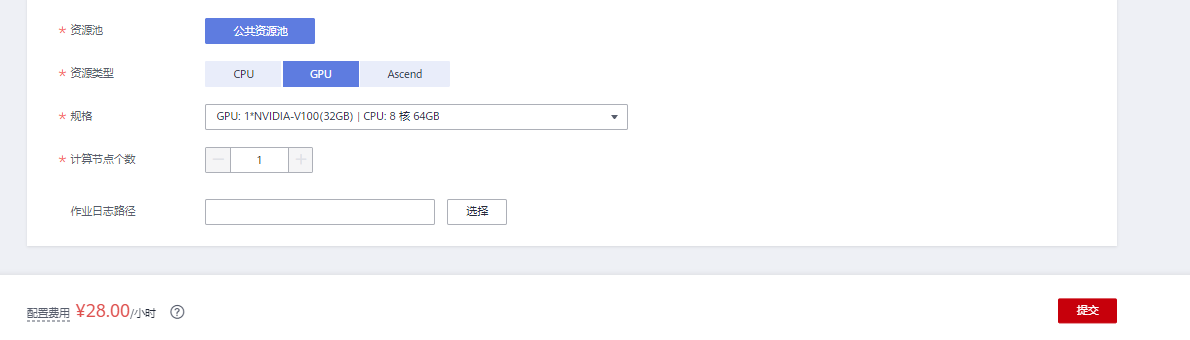


1. 针对训练输出，选择模型输出路径和last\_path路径，如下图所示。



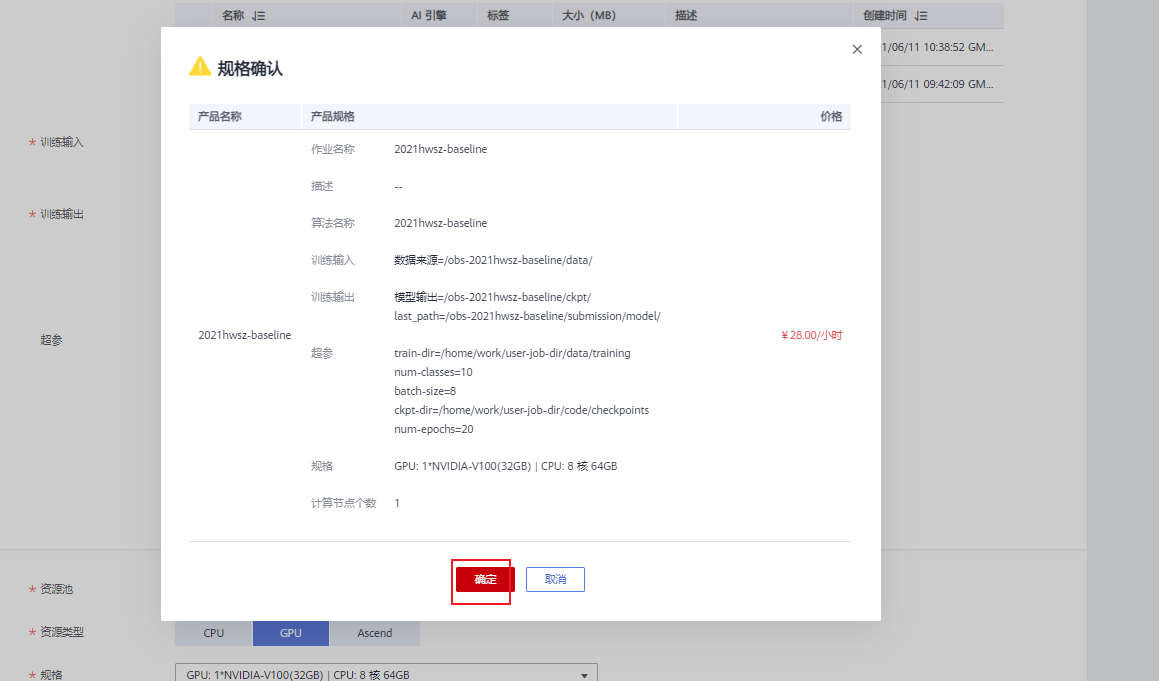


1. 选择计算规格，如下图所示。公共资源池->GPU->1\*NVIDIA-V100或公共资源池->CPU->CPU: 2 核 8GB

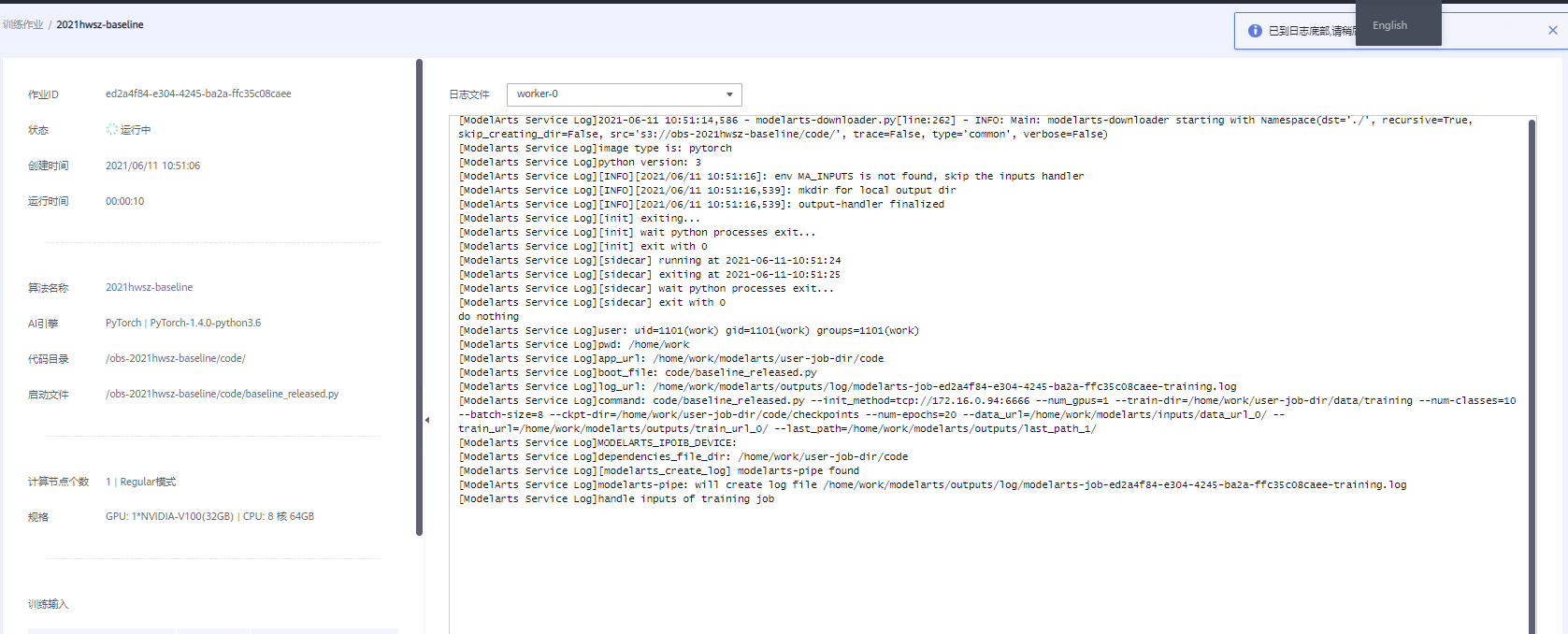




1. 点击提交，确认无误后，开始训练。



1. 通过日志可以实时查看代码输出日志。如下图所示。



1. 通过资源占用情况，可是实时观察显存、内存、CPU、GPU的利用率。如下图所示。



1. 训练完成，会将产生的模型pth文件实时传输到obs桶中（obs://obs-2021hwsz-baseline/ckpt）。同时会自动将最后一个pth文件上传到obs://obs-2021hwsz-baseline/submission/model中，并且重命名为model\_best.pth，以便后续第四章导入模型。**注意，整个训练作业处于“已完成”的状态后，才会触发上传任务。**

说明

* MoXing是华为云ModelArts团队自研的分布式训练加速框架，并提供了一套文件对象API，可以用来读写ModelArts Notebook容器本地文件及OBS文件。
* 从OBS下载文件可能会产生**流量费用**，请选手注意。北京四的收费标准为0.5元/GB。Modelarts中训练作业或者notebook与OBS的传输不收费，属于内网传输。公网传输和跨区传输会产生流量费用。
* OBS上传时限制文件最大上传大小为5GB，若有大于5GB的数据请参考SDK分段上传或拆分为多个文件分批次上传。

# 将生成的模型导入至模型管理

模型训练完成且编写好模型推理代码后，您可以将生成的模型导入至模型管理。

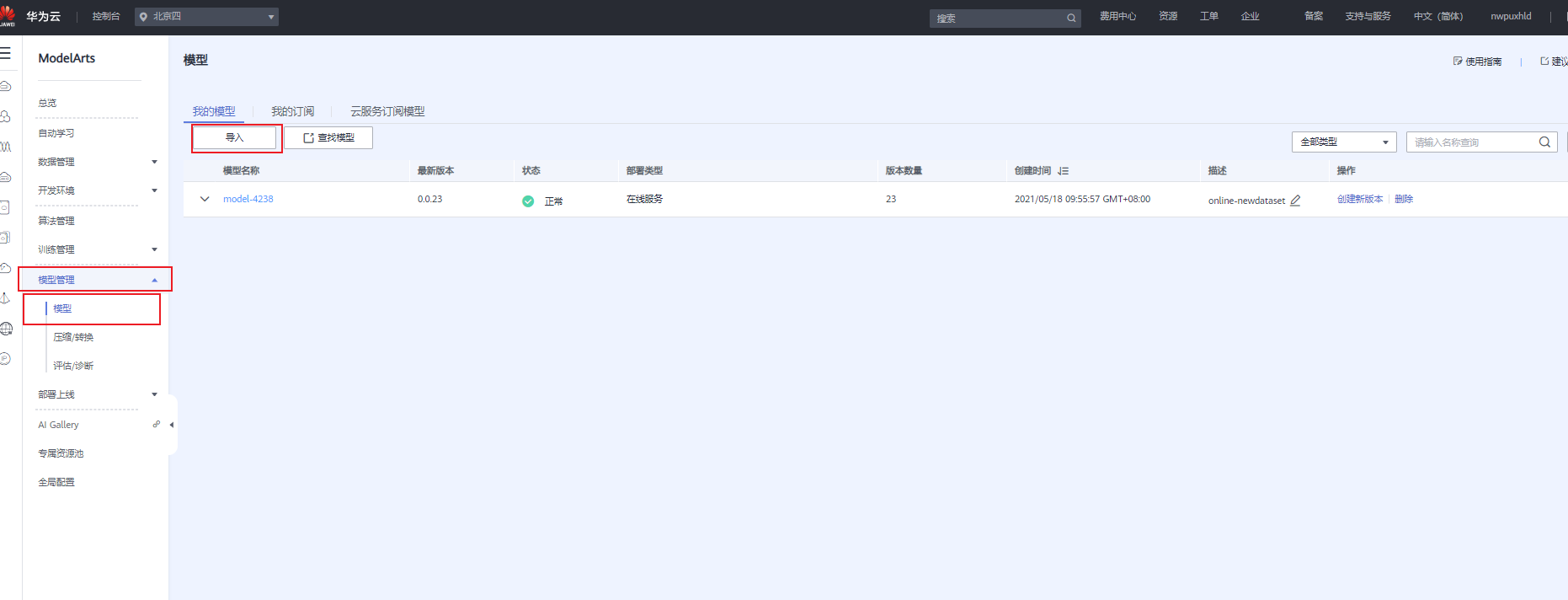
导入模型前，需要将[附件](#附录)中本baseline模型配置文件、推理代码，上传至submission/model文件夹下。

* config.json：用于配置模型运行的在线环境、依赖包等相关信息
* customize\_service.py: 推理服务运行的主入口函数
* model\_best.pth：待加载的模型，通过训练作业自动提交至该目录。也可以手动将本地的模型文件命名为model\_best.pth上传至该目录。

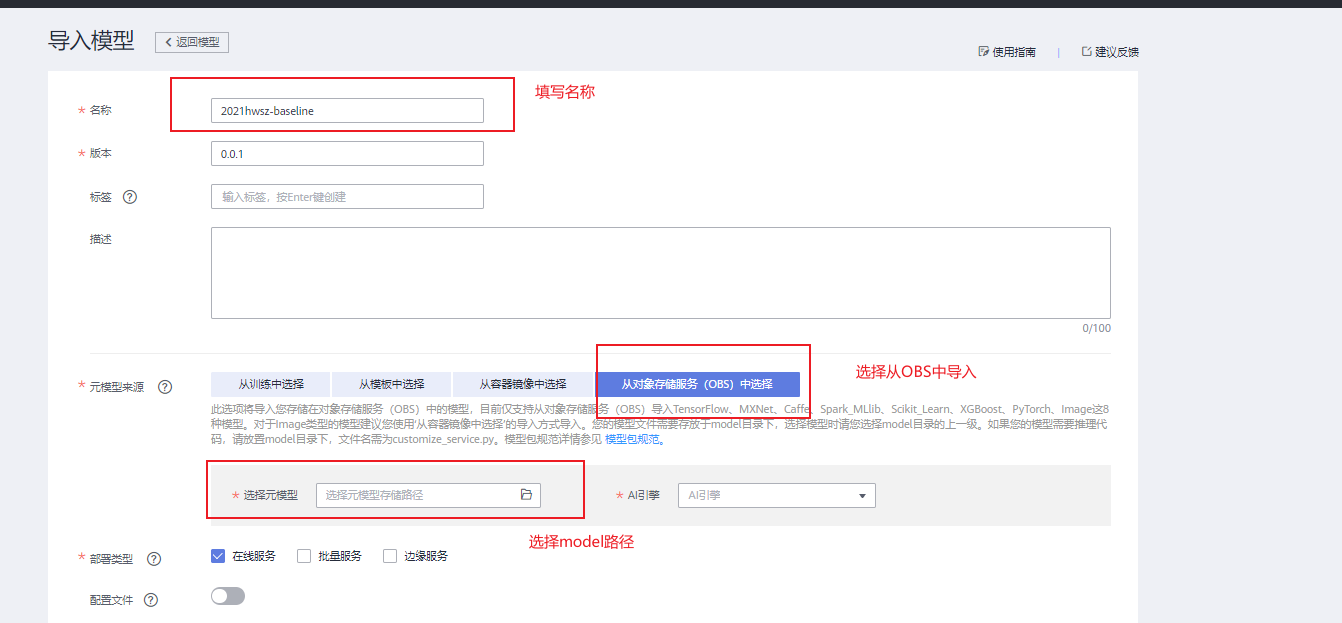


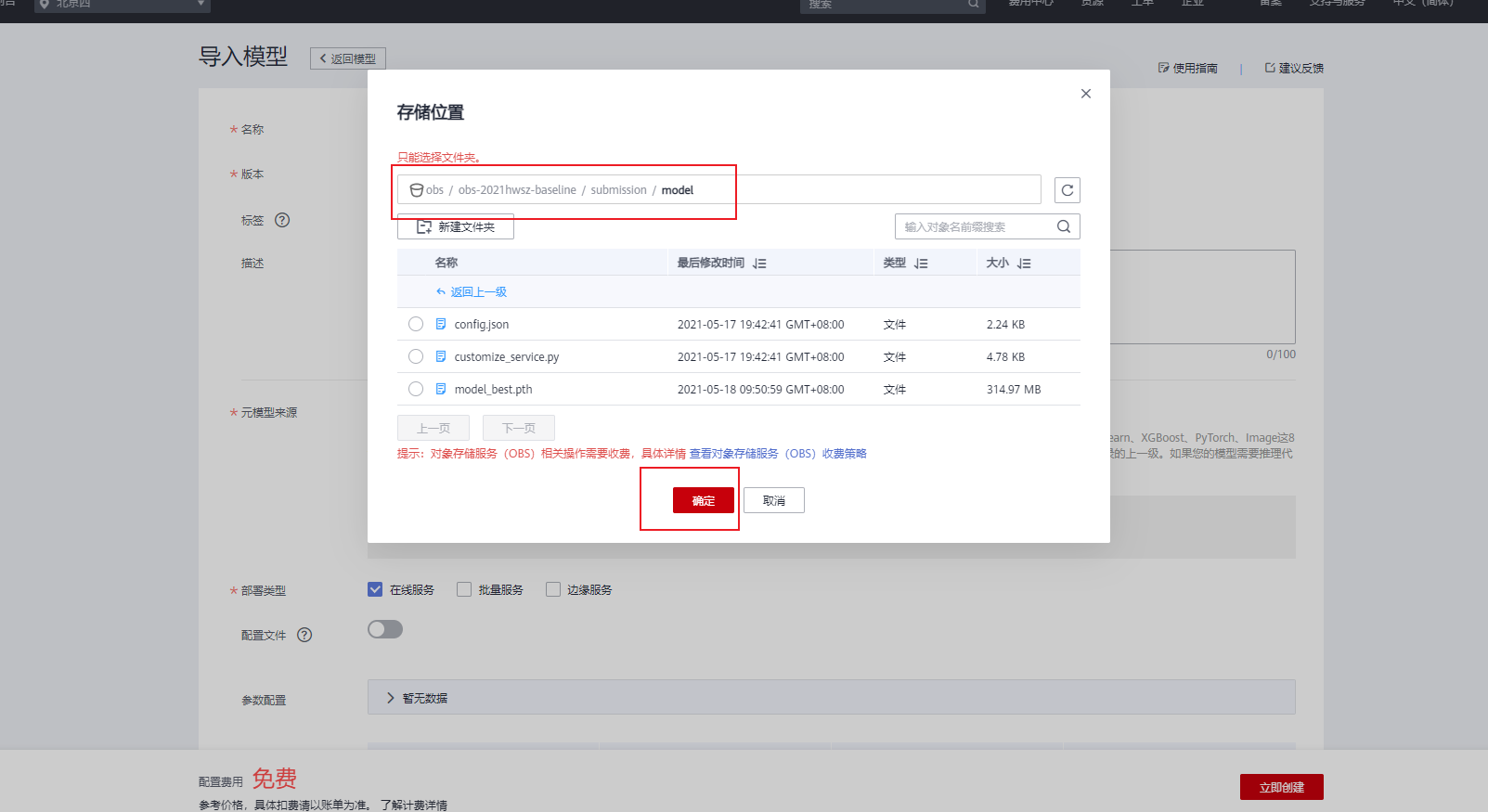
上述三个文件准备齐全后，导入模型具体操作如下。

1. 在ModelArts左侧导航栏选择“模型管理”下的“模型”页面，单击页面中“导入”。

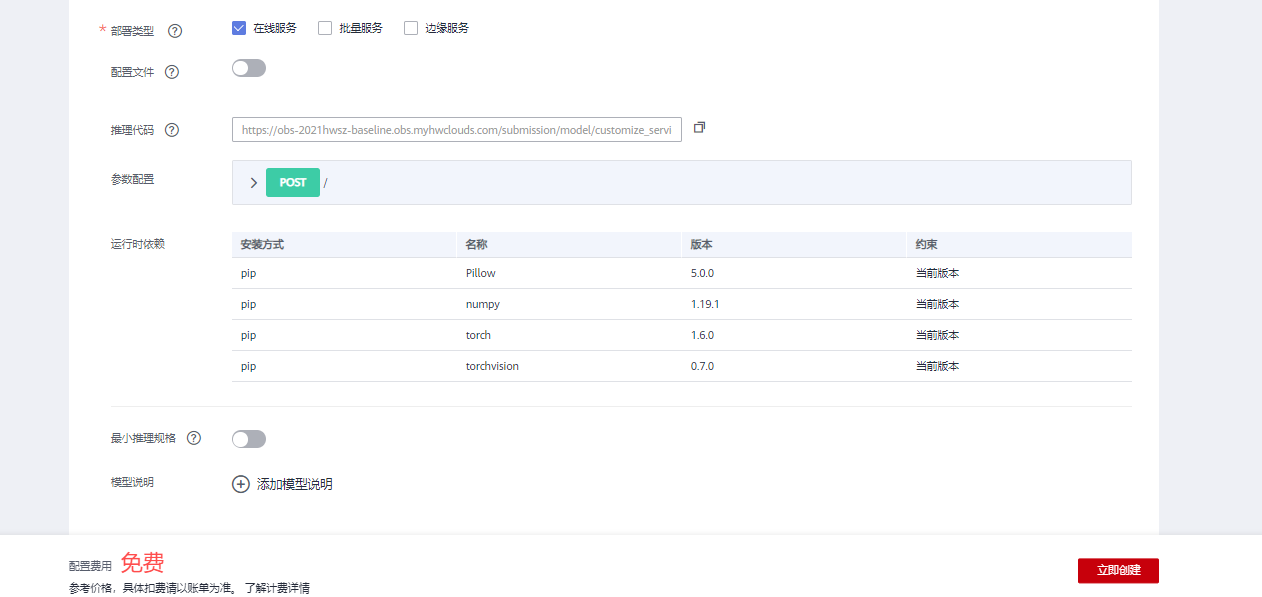


1. 在导入模型页面填写名称，选择元模型来源；其中元模型来源如从OBS中选择，请选择model文件夹；



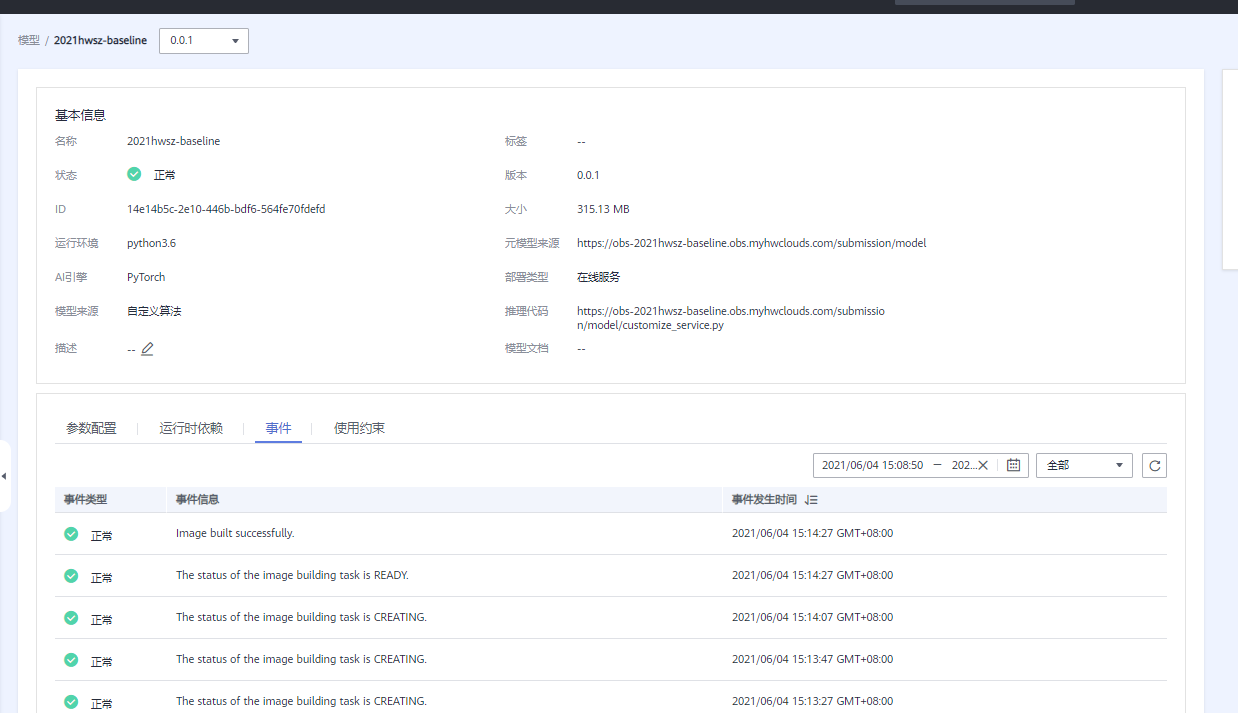


1. 若选择正确则平台会自动解析目录下的config文件完成下图的配置，同时正确识别到推理代码。



1. 单击“立即创建”。

单击模型名称进入模型详情页面，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](#ZH-CN_TOPIC_0169740168)的操作。

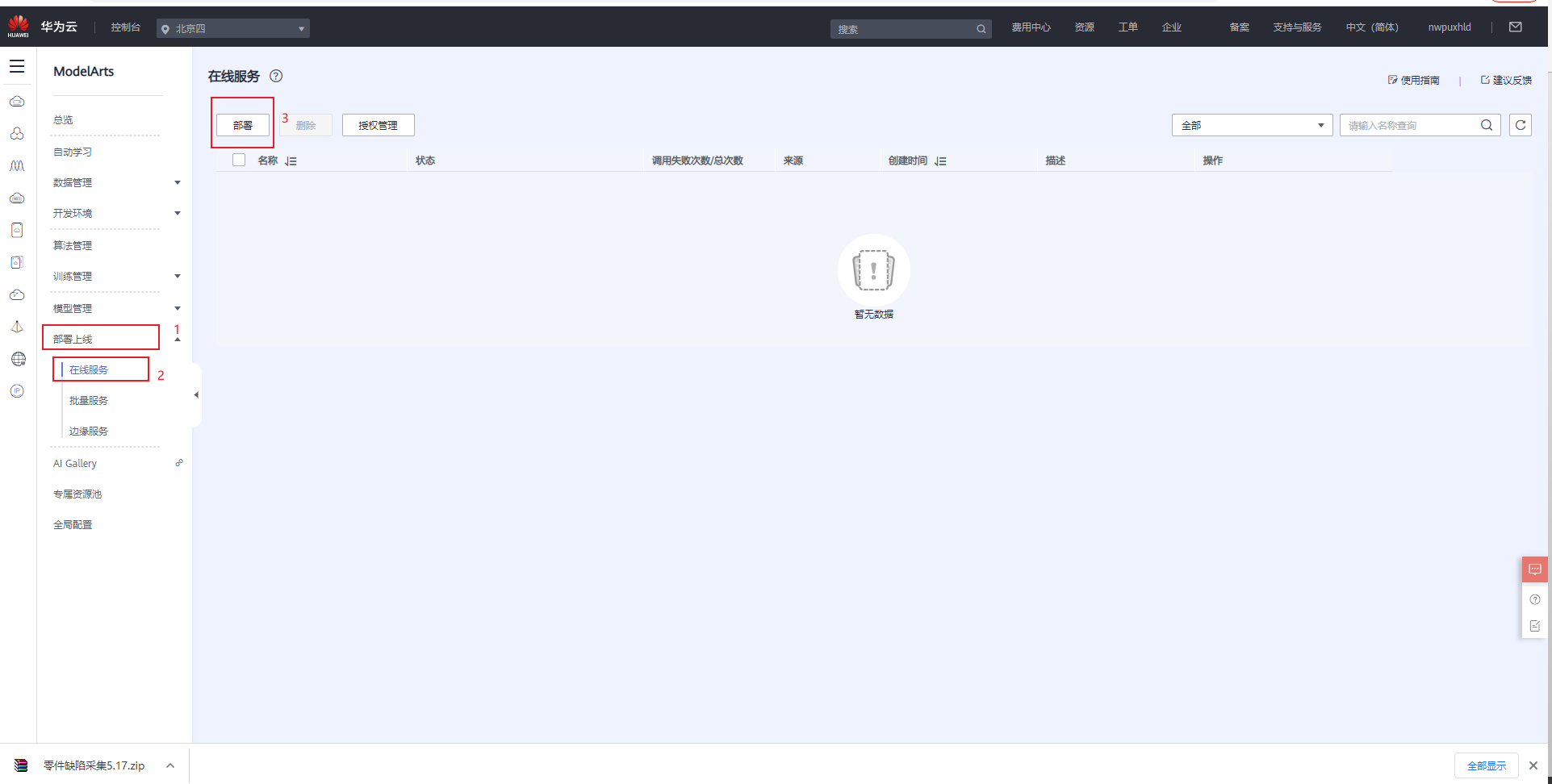


# 将模型部署为在线服务

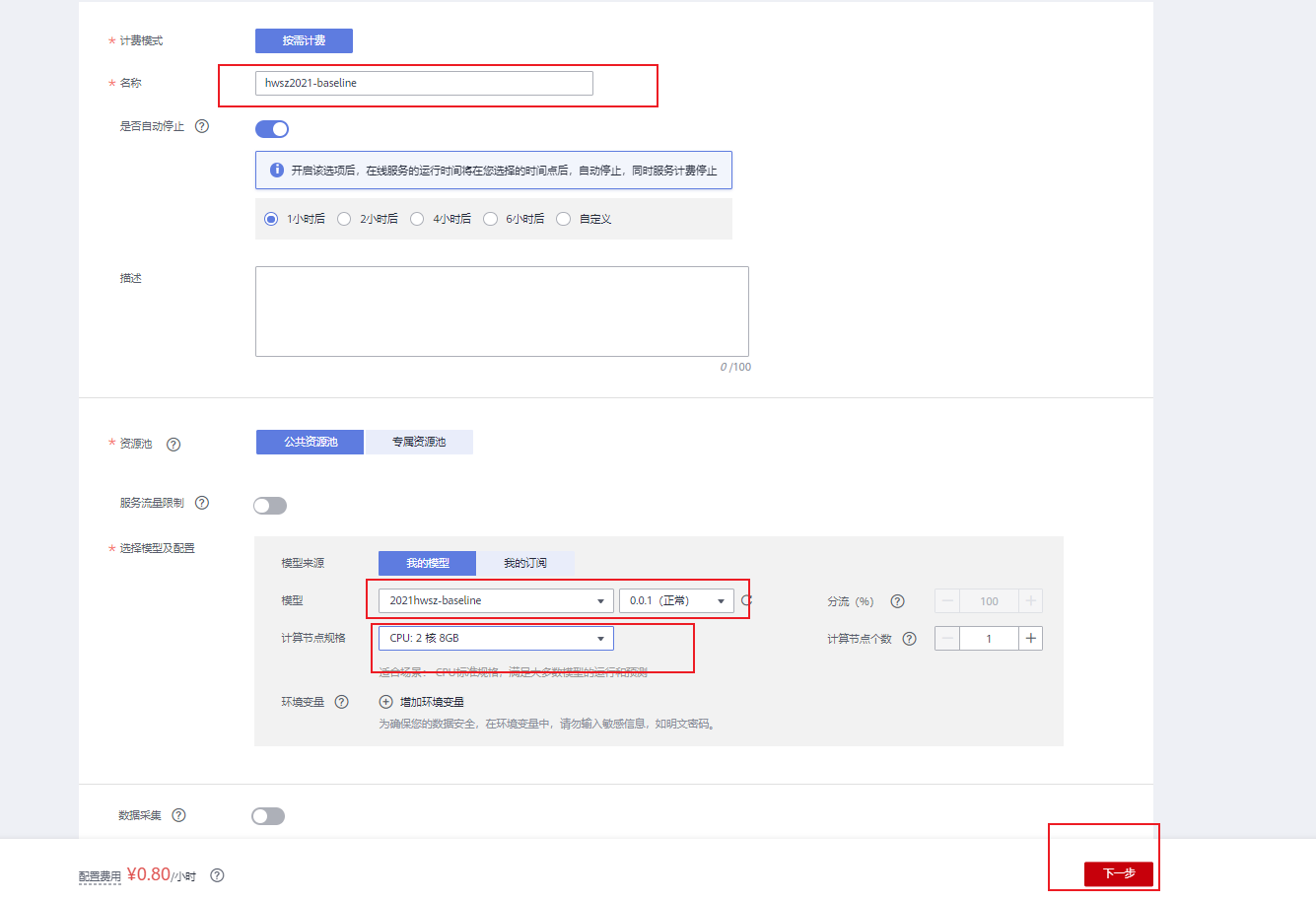
导入模型完成后，接下来您可以将模型部署为在线服务。这里主要是为了验证模型是否正常工作。

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”，然后单击页面中的“部署”。



1. 在部署页面填写参数，选中刚刚导入模型管理中的模型，计算节点可以选择CPU2核8G。



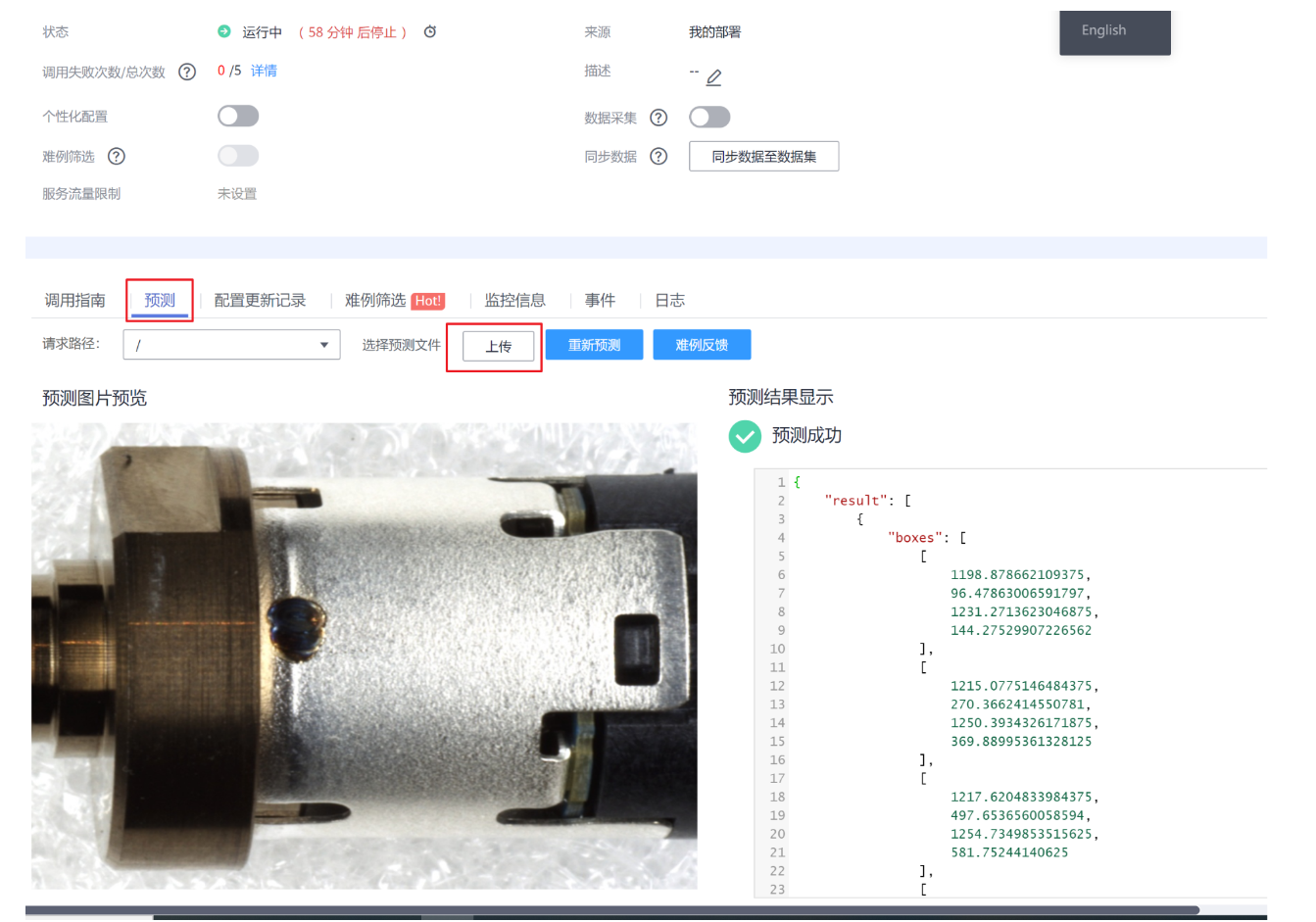
1. 单击“下一步”，参数确认无误后，单击“提交”。

创建成功后，返回在线服务列表，您可以在在线服务列表中查看部署进度；当状态由“部署中”变为“运行中”后，单击服务名称，进入详情页面，如下图所示。

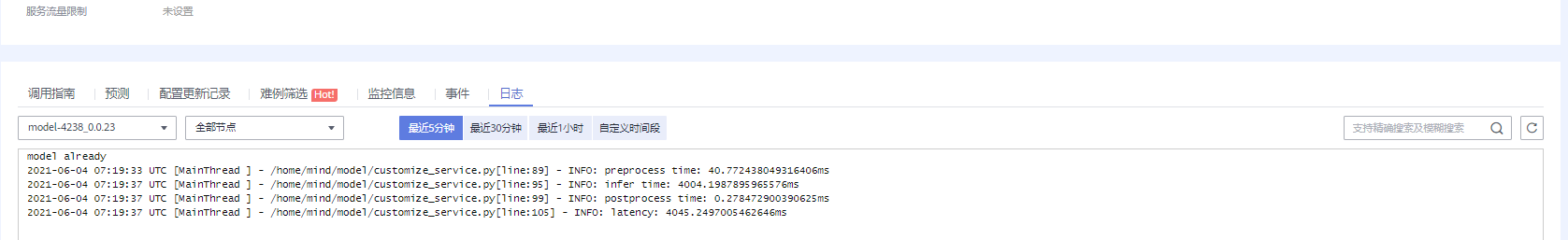




1. 在“预测”页签上传图片进行测试，如下图所示。图片可从训练集中任意选取一张即可。



1. 如果出现报错，选手可以在日志中定位问题，如下图所示



1. 当在线推理测试无误后，选手可以将模型进行参赛发布提交。

说明

* 由于不建议使用模型堆积的策略，建议单张图片的耗时应该控制在5s以内。

# 提交模型判分

模型部署测试后，您可以将模型提交判分，具体操作如下。

**注意：由于每支队伍每天有提交次数限制，为避免浪费提交次数，请在模型提交前务必要部署测试模型，确认模型无误后再提交判分。**

**步骤一：在odelArts上提交模型**

1. 在ModelArts左侧导航栏中选择“模型管理”，单击模型名称前方箭头；然后单击页面右侧操作栏中的“发布”，单击“参赛发布”。



1. 在弹出的“参赛模型提交”对话框中，选择比赛项目、比赛阶段，确认模型名称正确后，然后单击“确定”。



1. 确定后，即成功提交模型判分。在如下界面中可点击“现在加入”，也可以点击“以后再说”或直接点击右上角关掉该对话框。

**步骤二：在竞赛平台上查看分数**

在大赛平台上“提交作品”页签中查看判分分数。其中“提交作品”页面需报名比赛后才会显示。

# 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实例，训练作业，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”。



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。

# 附录

如下RSC\_Baseline.zip文件包含了本指导中涉及的代码文件，文件说明见下表。



| 文件 | 说明 |
| --- | --- |
| baseline\_released.py | 用于复现baseline的主入口函数。 |
| customize\_service.py | 发布模型所需的主入口函数。参见第四章。 |
| config.json | 发布模型所需要的配置文件，相关字段参考：https://support.ulanqab.huawei.com/engineers-modelarts/modelarts\_23\_0092.html#modelarts\_23\_0092\_\_table1683418482455。 |
| requirements.txt | baseline代码依赖的一些库名称和版本。 |
| utils/ | 一些需要用到的函数封装 |
| pretrained/ | 存放预训练模型。 |