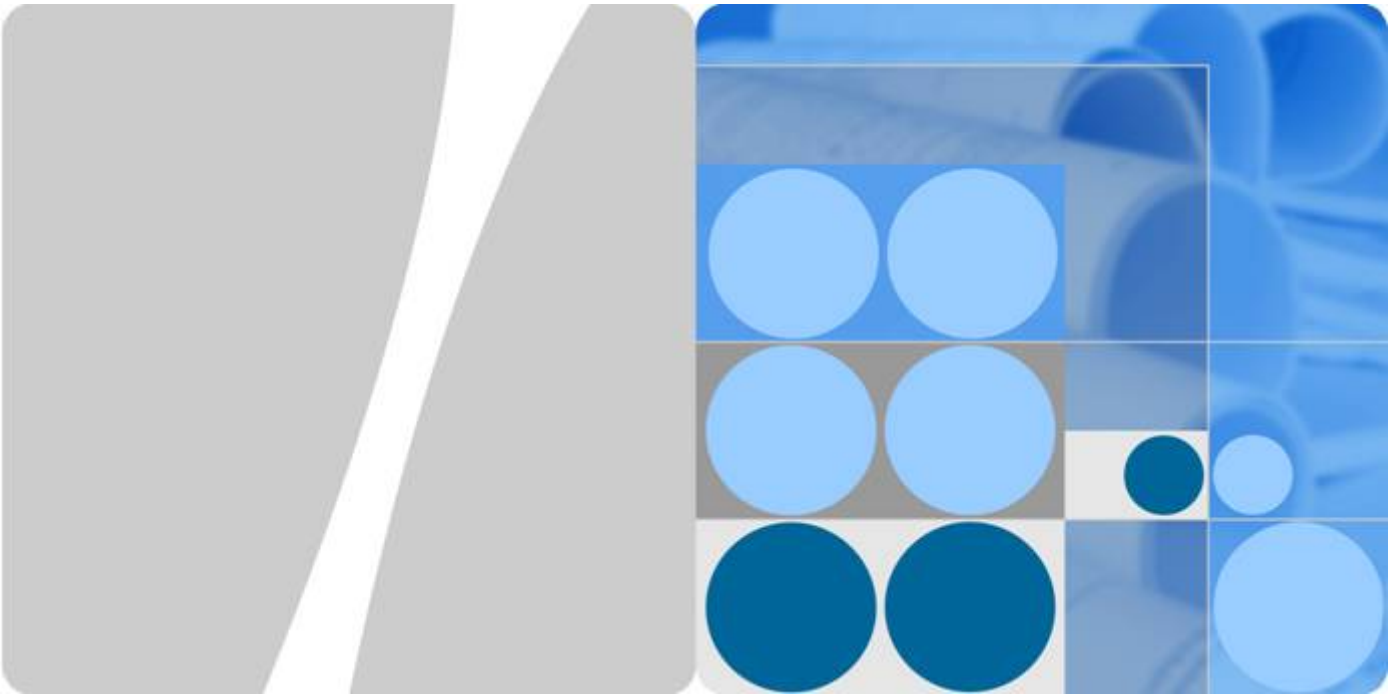


ModelArts

用户指南

发布日期　2018-11-15

目 录

1 文档导读

本文档包含了使用ModelArts前的准备工作、如何使用ModelArts以及各功能模块操作的相关指导，您可以根据表1-1查找您需要的内容。

表 1-1 文档导读

阶段	章节
使用ModelArts前必做的准备工作	● 申请ModelArts ● ModelArts全局配置 ● 上传数据至OBS
了解如何使用ModelArts，并基于入门样例快速熟悉ModelArts三种AI模型开发方式流程和相关操作	● 自动学习：零基础构建AI模型 ● 预置算法：快速生成模型 ● 代码开发：AI开发全流程管理
熟悉ModelArts自动学习相关操作	● 自动学习
熟悉ModelArts管理控制台中数据管理、开发环境、训练作业、模型管理及部署上线相关操作	● 数据管理 ● 开发环境 ● 训练作业 ● 模型管理 ● 部署上线
查看ModelArts市场中共享的数据集和模型	● 市场

2 准备工作

2.1 申请 ModelArts

当前ModelArts处于公测阶段，需要申请公测并通过审核才能使用ModelArts。

您可以登录华为云，进入ModelArts管理控制台，单击“立即申请”，并填写申请信息以申请ModelArts公测。（您也可以直接点击**https://console.huaweicloud.com/modelarts/?region=cn-north-1#/beta**，进入申请ModelArts公测页面）

说明

- 申请完成后请等待系统管理员审核，您可在“我的公测”页面查看审批状态。当审批状态为“审批通过”时，则可以开始使用ModelArts。
- 公测审批当前是人工审批，一般需要等待2~3天，如遇节假日顺延。

2.2 ModelArts 全局配置

由于使用ModelArts时，您使用的Notebook、训练作业、模型和服务需要使用对象存储功能；若没有添加访问密钥，则无法使用对象存储功能。因此在使用ModelArts进行AI模型开发前，您需要完成以下2个步骤：

1. 获取访问密钥
2. 添加访问密钥

获取访问密钥

步骤1 登录华为云，打开“我的凭证”页面（您可直接点击**https://console.huaweicloud.com/iam/#/myCredential**，进入“我的凭证”页面）。

步骤2 单击“管理访问密钥”页签，单击“新增访问密钥”按钮。

步骤3 在弹出的“新增访问密钥”对话框，输入当前用户的登录密码，通过已验证手机或已验证邮箱进行验证，输入对应的验证码，如**图2-1**所示。

图 2-1 新增访问密钥

步骤4 单击“确定”。

步骤5 根据浏览器提示，保存密钥。密钥会直接保存到浏览器默认的下载文件夹中。

步骤6 打开下载下来的“credentials.csv”文件既可获取到访问密钥（AK和SK）。

----结束

添加访问密钥

步骤1 在ModelArts控制台界面，单击左侧导航栏“全局配置”。

步骤2 单击“全局配置”界面中的“添加访问密钥”，在弹出的对话框中，填写获取的访问密钥，如图2-2所示。

图 2-2 添加访问密钥

说明

- “访问密钥”填写Access Key Id，“私有访问密钥”填写Secret Access Key。
- 请确保所填写的AK、SK为当前账号所获取的。

步骤3 单击“确认”，完成访问密钥的添加。

----结束

2.3 上传数据至 OBS

使用ModelArts处理数据时，需要先将数据上传到华为云对象存储服务（OBS）桶中。您可以登录**OBS管理控制台**，单击“创建桶”。OBS桶创建完成后，可以在您创建的OBS桶中创建文件夹，然后再进行数据的上传。

说明

由于当前ModelArts部署在华北-北京一，您在创建OBS桶时请选择华北-北京一。

OBS上传数据的详细操作请参见**《对象存储服务控制台指南》**。

3 如何使用 ModelArts

3.1 ModelArts 使用简介

ModelArts是面向AI开发者的一站式开发平台，通过AI开发全流程管理助您智能、高效地创建AI模型和一键部署到云、边、端。

ModelArts不仅支持自动学习功能，还预置了多种已训练好的模型，同时集成了Jupyter Notebook，提供在线的代码开发环境。通过ModelArts您可以使用以下三种方式进行AI模型的开发：

1. 如果您是业务开发者，无AI开发经验，您可以使用ModelArts的自动学习功能，进行零基础构建AI模型，详细介绍请参见**自动学习：零基础构建AI模型**。
2. 如果您是AI初学者，有一定AI开发经验，您可以上传业务数据至OBS，然后对ModelArts预置的算法进行重训练，从而得到新模型，详细介绍请参见**预置算法：快速生成模型**。
3. 如果您是AI工程师，熟悉代码编写和调测，您可以使用ModelArts提供的在线代码开发环境，编写训练代码进行AI模型的开发，详细介绍请参见**代码开发：AI开发全流程管理**。

3.2 自动学习：零基础构建 AI 模型

ModelArts自动学习功能可以根据标注数据自动设计模型、自动调参和训练、自动压缩和部署模型，不需要代码编写和模型开发经验，即可实现零基础构建AI模型。

当前自动学习支持快速创建图像分类、物体检测和预测分析模型。图像分类是识别图片中是否是某类物体。物体检测是识别图片里每个物体的位置、类别。预测分析是对结构化数据做出分类或者数值预测。

使用流程

使用自动学习零基础构建AI模型流程如**图3-1**所示。

图 3-1 自动学习使用流程

入门样例场景描述

“云宝”是华为云的吉祥物，本样例基于ModelArts平台，详细介绍如何使用自动学习对预置的云宝数据集进行训练，快速构建云宝图像识别应用。

操作步骤主要分为4部分：

1. **数据准备**：通过ModelArts市场中预置的数据集，创建自动学习项目中所需要的云宝数据集。
2. **创建物体检测项目**：使用云宝数据集创建物体检测项目。
3. **自动训练**：创建项目后，发布训练。
4. **部署上线**：完成模型训练后，进行服务的部署及测试。

说明

- 使用ModelArts开始本样例操作前，请完成准备工作中的**申请ModelArts**和**ModelArts全局配置**。

数据准备

步骤1 登录“ModelArts”管理控制台，单击左侧导航栏“市场”。在“数据集”页签，找到自动学习对应的云宝预置数据集“Yunbao-Data-Set”。

步骤2 单击进入该预置数据集Yunbao-Data-Set详情页面，单击“导入到我的数据集”，页面会自动跳转到“数据管理>数据集”页面进行创建。

步骤3 查看创建的云宝数据集（Yunbao-Data-Set）。在“数据集目录”页签获取创建的云宝数据集的桶信息，如图3-2所示。

图 3-2 云宝数据集桶信息

----结束

创建物体检测项目

步骤1 在ModelArts管理控制台，单击左侧导航栏“自动学习”。

步骤2 单击物体检测方框中的“创建项目”，在弹出的对话框中填写项目名称，训练数据选择云宝数据集OBS路径/ yunbao-data-set-c79800ac-abb5-40f4-8398-60e50f503649/Data/（确保目录结构正确，选择到Data层），如图3-3所示。

图 3-3 创建项目

步骤3 单击“确定”，完成物体检测项目创建。

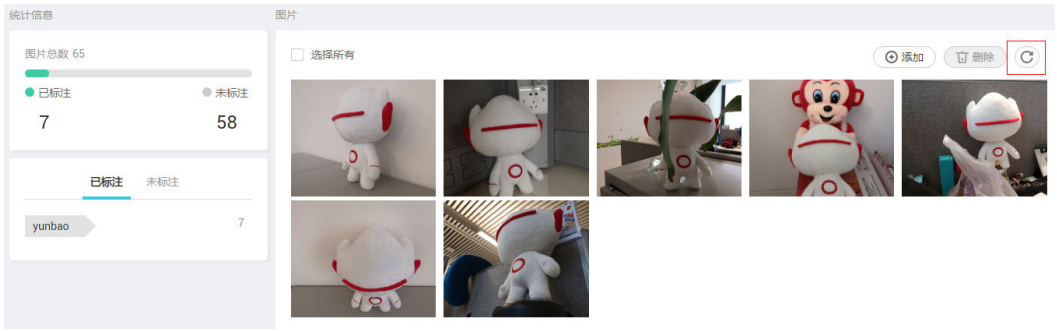
----结束

自动训练

步骤1 在自动学习项目列表中单击项目名称，进入当前自动学习项目“数据标注”页签。

步骤2 在数据标注页面单击  ，执行数据源同步操作，同步完成后如图3-4所示。

图 3-4 数据源同步



说明

- 预置的云宝数据集部分图片已经标注，您可以直接使用已经标注的图片进行训练。您也可以在“未标注”页签，完成所有图片的标注，图片标注操作请参见**图片标注**。

步骤3 单击右侧的“开始训练”，启动模型训练。

----结束

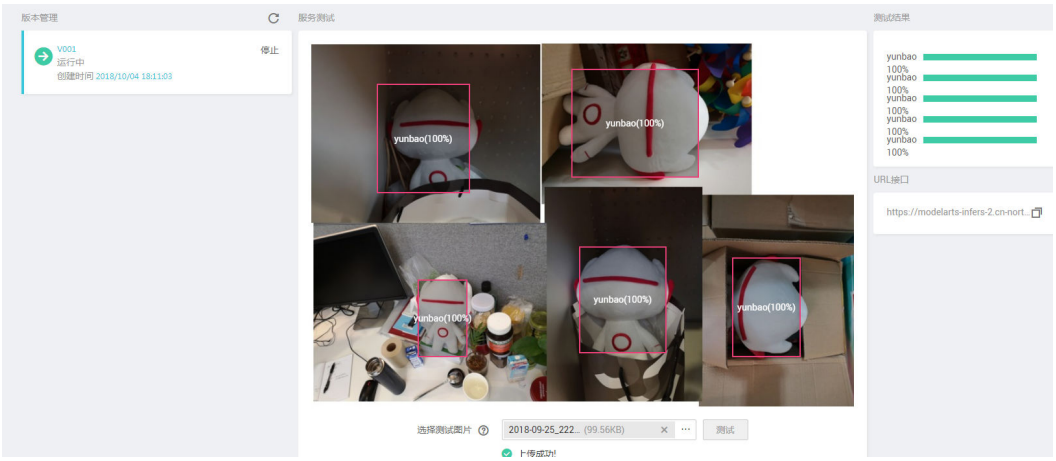
部署上线

步骤1 当“模型训练”页签中训练状态为“已完成”时，单击页签右侧“部署”，完成物体检测模型的部署。

步骤2 部署成功后，可在“部署上线”页签上传图片测试，如图3-5所示。

如模型准确率不满足预期，可在“数据标注”页签中添加图片并进行标注，重新进行模型训练及部署上线。

图 3-5 服务测试



----结束

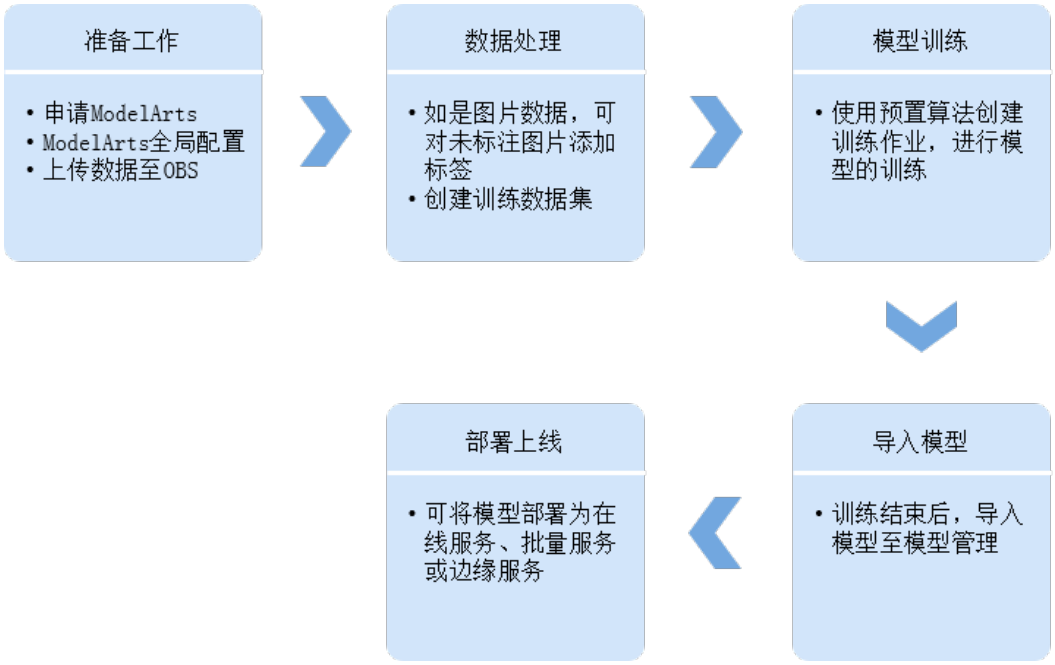
3.3 预置算法：快速生成模型

ModelArts预置了多种已经训练好的模型，您可以上传自己的业务数据对预置模型进行重训练，无需编写代码即可快速生成模型。

使用流程

使用预置算法快速生成模型流程如图3-6所示。

图 3-6 预置算法使用流程



入门样例场景描述

本样例基于ModelArts平台，详细介绍如何使用flowers数据集对预置ResNet_v1_50模型进行重训练，快速构建花卉图像分类应用。

操作步骤主要分为4部分：

1. **数据准备**：下载flowers数据集，上传数据至华为云对象存储服务（OBS）中。
2. **模型训练**：使用flowers训练集对ResNet_v1_50模型重训练。
3. **导入模型**：训练结束后，导入模型至模型管理。
4. **部署上线**：将得到的模型，部署为在线预测服务，并添加图片进行测试。

说明

- 使用ModelArts开始本样例操作前，请完成准备工作中的**申请ModelArts**和**ModelArts全局配置**。

数据准备

步骤1 下载并解压缩数据集压缩包“flower_photos.tgz”，flowers数据集的下载路径为：**https://dls-public-data.obs.cn-north-1.myhwclouds.com/tensorflow/modelarts_flowers.zip**。

步骤2 参见**上传数据至OBS**，将数据集上传至OBS桶中（假设OBS桶路径为：“s3://modelarts-example/datasets/flowers_split”）。

该路径下包含了模型训练需要的所有图像文件， 目录下有5种类别，分别为：daisy, dandelion, roses, sunflowers, tulips，目录结构为：

```
s3://modelarts-example/datasets/flowers_split
    |- 01.jpg
    |- 01.txt
    |- ...
    |- n.jpg
```

```
    |- n.txt
    |- ...
```

----结束

模型训练

步骤1 单击ModelArts左侧导航栏的“训练作业”，在“预置算法”页签找到名称为ResNet_v1_50的模型，单击右侧“创建训练”。

步骤2 在“创建训练作业”界面填写参数，如图3-7所示。

其中“数据的存储位置”(s3://modelarts-example/datasets/flowers_split)，即数据所在的父目录；“运行参数”，增加参数max_epoches=10，max_training_time=100000（防止数据量大时获取数据速度慢，训练时间过长而退出训练）；“训练输出位置”请选择一个路径（建议新建一个文件夹s3://modelarts-example/log）用于保存输出模型和预测文件。

说明

参数max_epoches：1个epoch代表整个数据集，此处表示训练10个epoch，数值可更改，不填写时使用默认值。

图 3-7 使用预置算法创建训练作业

步骤3 参数确认无误后，单击“立即创建”完成训练作业的创建。

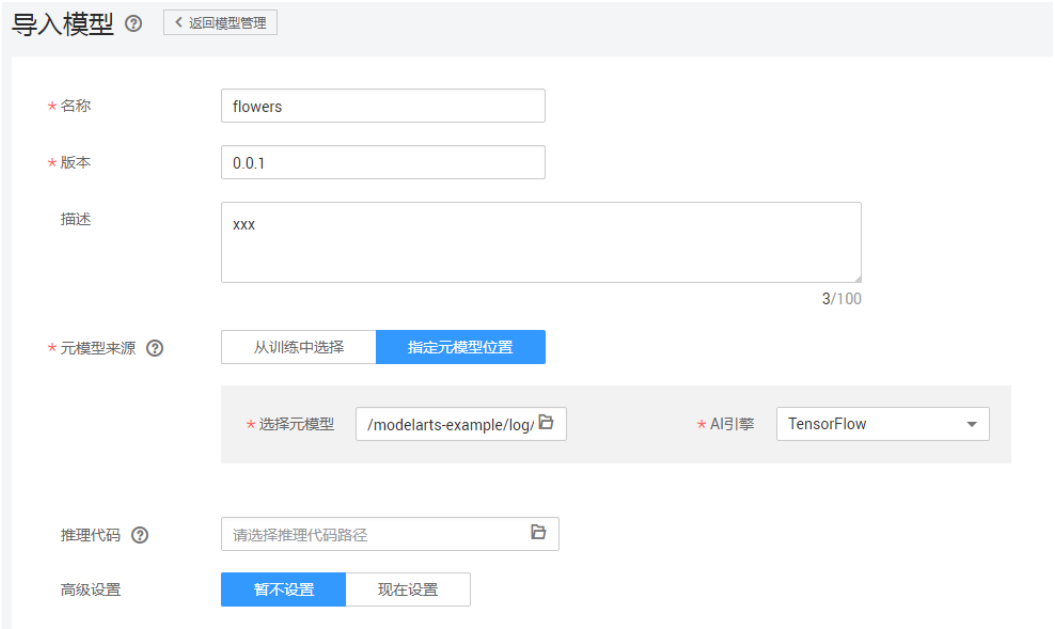
----结束

导入模型

步骤1 在ModelArts“模型管理”界面，单击左上角“导入”，参考图3-8填写参数。

其中“元模型来源”选择“指定元模型位置”，“选择元模型”的路径与训练模型中“训练输出位置”保持一致（s3://modelarts-example/log），“AI引擎”选择“TensorFlow”。

图 3-8 导入模型



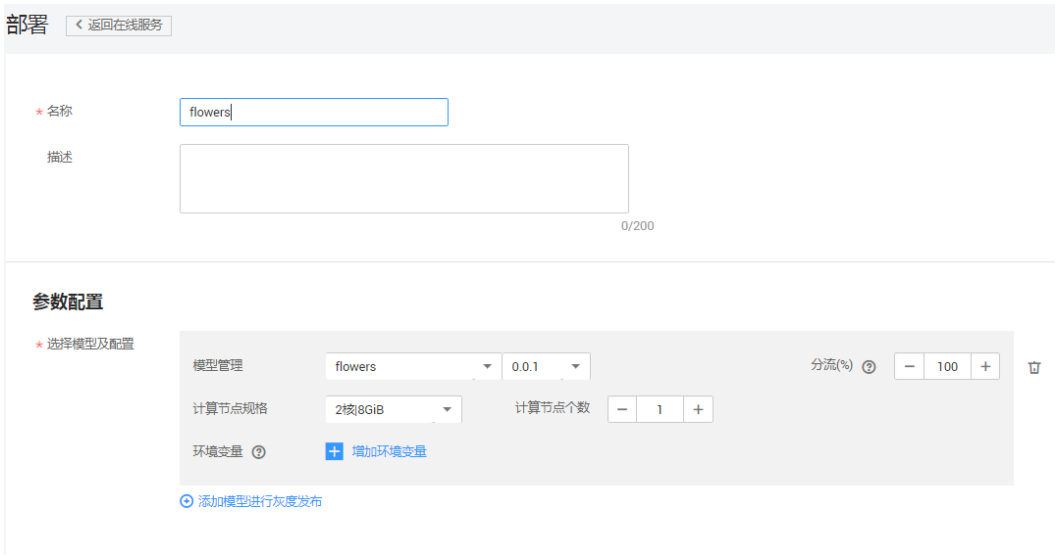
步骤2 单击“立即创建”。当模型状态为“正常”时，则表示导入成功。

----结束

部署上线

步骤1 在ModelArts“部署上线”界面，单击“在线服务”页签上方的“部署”，参考图3-9填写参数，单击“立即创建”，完成模型的部署。

图 3-9 部署为在线服务



步骤2 在“在线服务”界面中，单击服务名称，进入详情页面，可添加图片进行测试。

----结束

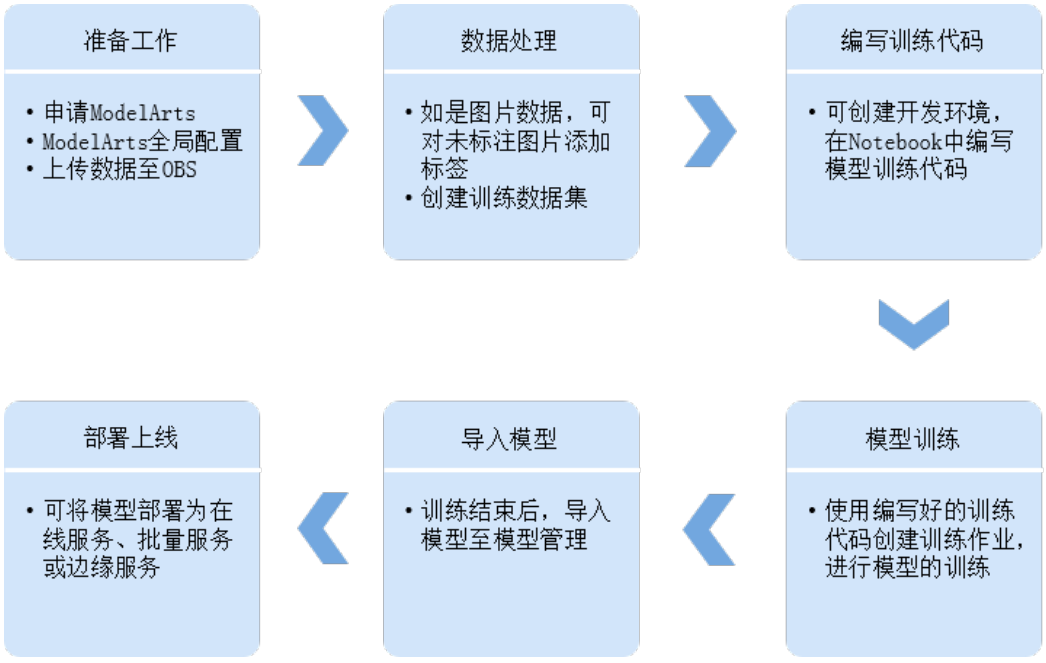
3.4 代码开发：AI 开发全流程管理

ModelArts集成了Jupyter Notebook，您可以通过创建开发环境，自行编写模型训练代码和调测，然后使用编写好的训练代码创建训练作业，进行模型的训练和部署。

使用流程

通过自行编写代码进行AI模型开发流程如图3-10所示。

图 3-10 代码开发使用流程



入门样例场景描述

MNIST数据集为手写数字图像数据集，每张图片大小为28*28像素，每张图片上有一个手写阿拉伯数字，数字分别为从0~9。

本样例基于ModelArts平台，详细介绍如何使用MXNet框架编写模型训练代码，然后进行模型的训练和部署，实现MNIST数据集的手写数字图像识别应用。

操作步骤主要分为5部分：

1. **数据准备**：下载MNIST数据集，上传数据至OBS。
2. **编写训练代码**：创建Notebook，编写模型训练代码。
3. **模型训练**：使用编写好的模型训练代码创建训练作业，进行模型的训练。
4. **导入模型**：训练结束后，导入模型，得到新模型。
5. **部署上线**：将得到的模型，部署为在线预测服务。

说明

- 使用ModelArts开始本样例操作前，请完成准备工作中的**申请ModelArts**和**ModelArts全局配置**。

数据准备

步骤1 下载MNIST数据集。下载路径为：**http://data.mxnet.io/data/mnist/** 。

数据集文件说明如下：

- t10k-images-idx3-ubyte.gz：验证集，共包含10000个样本。
- t10k-labels-idx1-ubyte.gz：验证集标签，共包含10000个样本的类别标签。
- train-images-idx3-ubyte.gz：训练集，共包含60000个样本。
- train-labels-idx1-ubyte.gz：训练集标签，共包含60000个样本的类别标签。

步骤2 解压缩4个压缩文件，参见**上传数据至OBS**，将数据集上传至华为云OBS桶 （假设OBS桶路径为：s3://mxnet-test/data/）。

----结束

编写训练代码

步骤1 单击ModelArts左侧导航栏“开发环境”，进入“Notebook”页签。

步骤2 单击左上角的“创建”。在“创建Notebook”界面填写参数，其中镜像类型选择MXNet-1.2.1-python3.6或MXNet-1.2.1-python2.7，存储位置为s3://mxnet-test/code/。

步骤3 在开发环境列表中，单击所创建开发环境右侧的“打开”，进入Jupyter Notebook文件目录界面，如**图3-11**所示。单击右上角的“New”，选择“Python 3”，进入代码开发界面。

图 3-11 Notebook 开发界面

步骤4 参见如下训练代码，在Cell中填写数据代码。单击运行按钮，运行代码。

```
import mxnet as mx
import argparse
import logging
import os
# load data
def get_mnist_iter(args):
    train_image = args.data_url + 'train-images.idx3-ubyte'
    train_lable = args.data_url + 'train-labels.idx1-ubyte'
    train = mx.io.MNISTIter(image=train_image,
                            label=train_lable,
                            data_shape=(1, 28, 28),
                            batch_size=args.batch_size,
                            shuffle=True,
                            flat=False,
                            silent=False,
                            seed=10)
    return train
# create network
def get_symbol(num_classes=10, **kwargs):
    data = mx.symbol.Variable('data')
    data = mx.sym.Flatten(data=data)
    fc1  = mx.symbol.FullyConnected(data = data, name='fc1',
num_hidden=128)
    act1 = mx.symbol.Activation(data = fc1, name='relu1',
act_type="relu")
    fc2  = mx.symbol.FullyConnected(data = act1, name = 'fc2',
num_hidden = 64)
    act2 = mx.symbol.Activation(data = fc2, name='relu2',
act_type="relu")
    fc3  = mx.symbol.FullyConnected(data = act2, name='fc3',
num_hidden=num_classes)
    mlp  = mx.symbol.SoftmaxOutput(data = fc3, name = 'softmax')
    return mlp
def fit(args):
    # create kvstore
    kv = mx.kvstore.create(args.kv_store)
    # logging
    head = '%(asctime)-15s Node[' + str(kv.rank) + '] %(message)s'
    logging.basicConfig(level=logging.DEBUG, format=head)
    logging.info('start with arguments %s', args)
    # get train data
    train = get_mnist_iter(args)
    # create checkpoint
    checkpoint = mx.callback.do_checkpoint(args.train_url if kv.rank ==
0 else "%s-%d" % (
        args.train_url, kv.rank))
    # create callbacks after end of every batch
```

```
    batch_end_callbacks = [mx.callback.Speedometer(args.batch_size,
args.disp_batches)]
    # get the created network
    network = get_symbol(num_classes=args.num_classes)
    # create context
    devs = mx.cpu() if args.num_gpus == 0 else [mx.gpu(int(i)) for i in
range(args.num_gpus)]
    # create model
    model = mx.mod.Module(context=devs, symbol=network)
    # create an initialization method
    initializer = mx.init.Xavier(rnd_type='gaussian', factor_type="in",
magnitude=2)
    # create params of optimizer
    optimizer_params = {'learning_rate': args.lr, 'wd' : 0.0001}
    # run
    model.fit(train,
              begin_epoch=0,
              num_epoch=args.num_epochs,
              eval_data=None,
              eval_metric=['accuracy'],
              kvstore=kv,
              optimizer='sgd',
              optimizer_params=optimizer_params,
              initializer=initializer,
              arg_params=None,
              aux_params=None,
              batch_end_callback=batch_end_callbacks,
              epoch_end_callback=checkpoint,
              allow_missing=True)
    if args.export_model == 1 and args.train_url is not None and
len(args.train_url):
        import moxing.mxnet as mox
        end_epoch = args.num_epochs
        save_path = args.train_url if kv.rank == 0 else "%s-%d" %
(args.train_url, kv.rank)
        params_path = '%s-%04d.params' % (save_path, end_epoch)
        json_path = ('%s-symbol.json' % save_path)
        logging.info(params_path + 'used to predict')
        pred_params_path = os.path.join(args.train_url, 'model',
'pred_model-0000.params')
        pred_json_path = os.path.join(args.train_url, 'model',
'pred_model-symbol.json')
        mox.file.copy(params_path, pred_params_path)
        mox.file.copy(json_path, pred_json_path)
        for i in range(1, args.num_epochs + 1, 1):
            mox.file.remove('%s-%04d.params' % (save_path, i))
        mox.file.remove(json_path)
if __name__ == '__main__':
    # parse args
    parser = argparse.ArgumentParser(description="train mnist",

formatter_class=argparse.ArgumentDefaultsHelpFormatter)
    parser.add_argument('--num_classes', type=int, default=10,
                        help='the number of classes')
    parser.add_argument('--num_examples', type=int, default=60000,
                        help='the number of training examples')
    parser.add_argument('--data_url', type=str, default='s3://obs-lpf/
data/', help='the training data')
    parser.add_argument('--lr', type=float, default=0.05,
                        help='initial learning rate')
    parser.add_argument('--num_epochs', type=int, default=10,
                        help='max num of epochs')
```

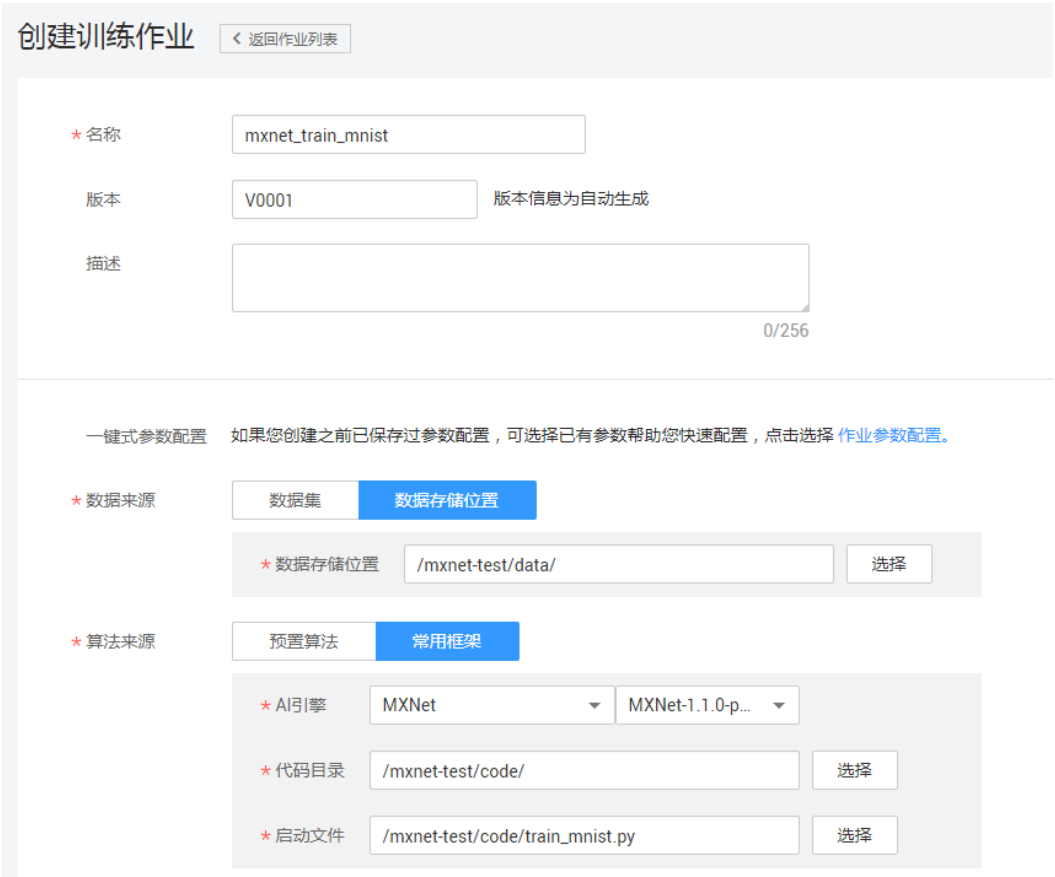
```
    parser.add_argument('--disp_batches', type=int, default=20,
                        help='show progress for every n batches')
    parser.add_argument('--batch_size', type=int, default=128,
                        help='the batch size')
    parser.add_argument('--kv_store', type=str, default='device',
                        help='key-value store type')
    parser.add_argument('--train_url', type=str, default='s3://obs-lpf/
ckpt/mnist',
                        help='the path model saved')
    parser.add_argument('--num_gpus', type=int, default='0',
                        help='number of gpus')
    parser.add_argument('--export_model', type=int, default=1, help='1:
export model for predict job \
                                                                    0:
not export model')
    args, unkown = parser.parse_known_args()
    fit(args)
```

----结束

模型训练

步骤1 在ModelArts管理控制台“训练作业”界面，单击左上角的“创建”，参考图3-12填写训练作业参数。

图 3-12 创建训练作业



其中，运行参数“num_epochs”为训练需要迭代的次数，默认10。“batch_size”为训练的每一步包含的样本数量大小， 默认128。参数“kv_store”，如果是单计算节点设

置为‘local’或‘device’，如果是多计算节点设置为‘dist_sync’或‘dist_sync_device’，如图3-13所示。

图 3-13 运行参数

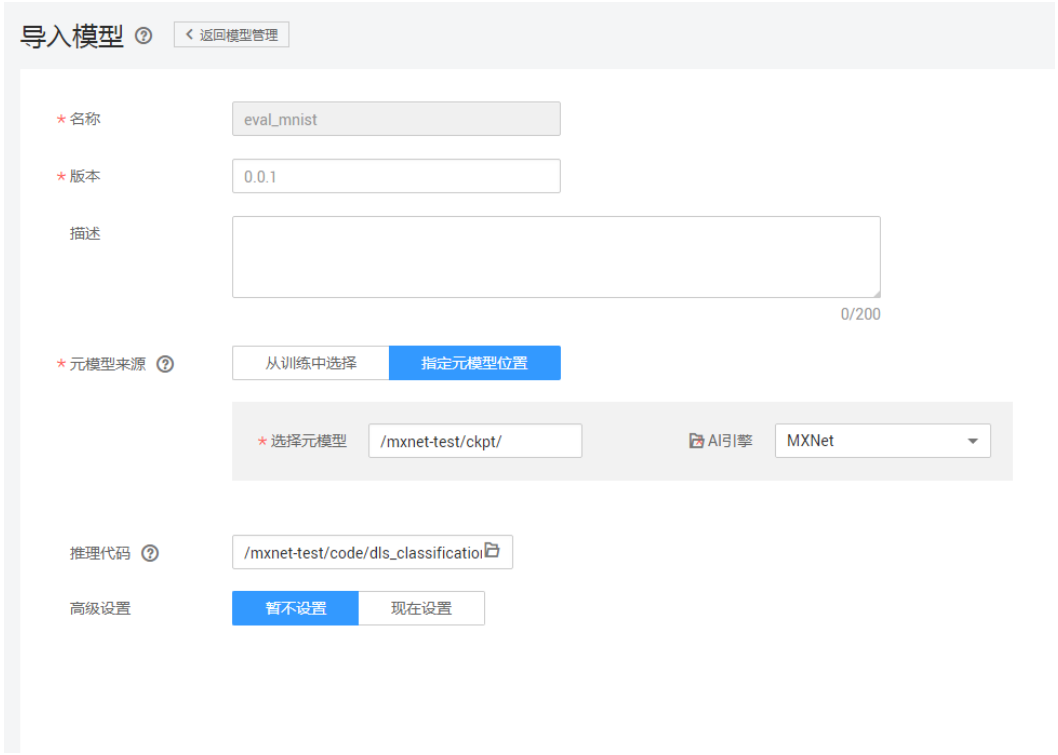
步骤2 单击“立即创建”，完成训练作业的创建。

----结束

导入模型

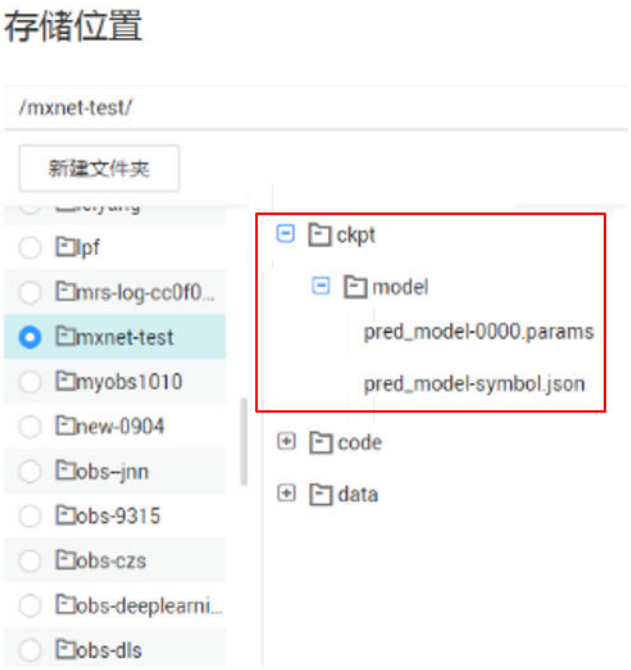
步骤1 在“模型管理”界面，单击左上角“导入”，参考图3-14填写参数。

图 3-14 导入模型



其中，元模型的路径需要设置为.params和.json文件所在文件夹的上一层文件夹，比如图3-15红框的路径：/mxnet-test/ckpt/。

图 3-15 元模型指定路径



“推理代码”参考**dls_classification_service.py**，并将推理代码上传到obs，并将其文件路径设置到模型参数配置中的“推理代码”输入框。

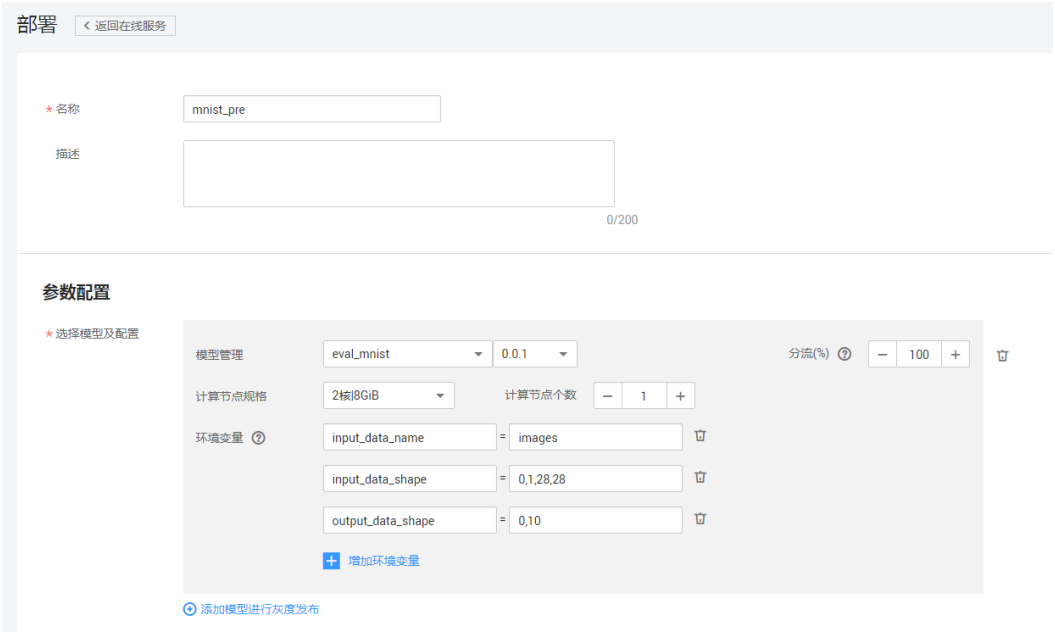
步骤2 单击“立即创建”，完成模型的导入。

----结束

部署上线

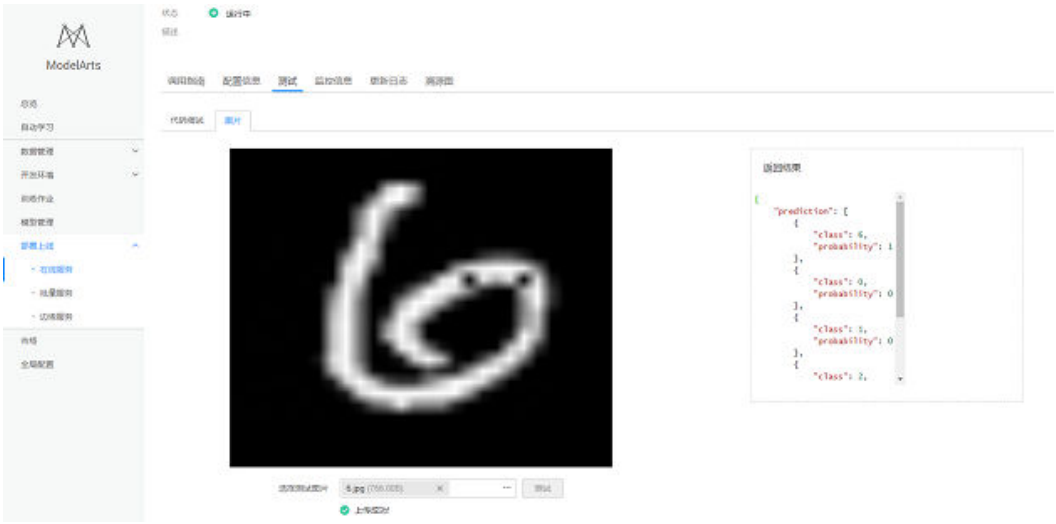
步骤1 在ModelArts“部署上线”界面，单击“在线服务”页签上方的“部署”，参考**图3-16**填写参数，单击“立即创建”，完成模型的部署。其中“input_data_name”为输入数据的名称，“input_data_shape”为输入数据需要的形状，“output_data_shape”为模型输出数据的形状，“0,10”表示输出的类别在0到10之间。

图 3-16 部署为在线服务



步骤2 在“在线服务”界面中，单击服务的名称，进入服务详情页面，可添加图片进行测试，如图3-17所示。其中'class'就是类的名称，比如第一项class为6，就是数字6这一类。类后的参数为概率，即预测为6的可能性为100%，别的数字都为0。

图 3-17 在线服务测试



如果您想使用自己的手写图片，可以参考**make_your_mnist.py**，只需修改图片数据的路径即可，其中原图需要使用黑底白字。

----结束

4 自动学习

4.1 自动学习简介

使用ModelArts自动学习开发AI模型，无需编写代码，您只需上传数据，通过自动学习提供的UI向导完成数据标注、发布训练即可训练出高质量模型，然后部署上线并进行测试。

自动学习界面说明

自动学习界面主要分为2部分，界面上方列举了当前支持的自动学习项目类型，单击“创建项目”可**创建自动学习项目**；界面下方为已创建的自动学习项目列表，您可以在列表右上方根据项目的类型进行过滤显示，或者在文本框中输入名称，单击 进行查询。

自动学习项目列表说明如**表4-1**所示。

表 4-1 自动学习项目列表说明

参数名称	说明
项目名称	项目的名称。 单击项目名称可进入当前项目详情页面。
项目类型	项目的类型。 当前支持图像分类、物体检测和预测分析三种类型的项目。
训练状态	当前项目最新版本的训练状态。 包括未开始、初始化、部署中、提交失败、运行中、删除中、已完成、运行失败、已丢失。
OBS路径	当前项目数据存放的OBS路径。
创建时间	项目创建的时间。
描述	项目的简要描述。

参数名称	说明
操作	对已创建项目的操作。 ● “删除”：删除当前自动学习项目。 ● “停止”：停止当前训练状态为“运行中”的自动学习项目。

4.2 创建自动学习项目

使用ModelArts自动学习前，首先需要创建一个自动学习项目，本节为您介绍如何创建自动学习项目以及OBS数据路径规则要求。

步骤1 单击自动学习界面中“创建项目”，弹出“创建项目”对话框，如图4-1所示。

图 4-1 创建项目

步骤2 在“创建项目”对话框中填写参数，参数说明如表4-2所示。

表 4-2 参数说明

参数名称	说明
名称	项目的名称。 名称只能包含数字、字母、下划线和中划线，长度不能超过20位且不能为空。

参数名称	说明

训练数据	OBS数据路径。
	1. 图像分类、物体检测项目的OBS数据路径需符合以下规则：
	- 如不需要提前上传进行标注的图片，请创建一个空文件夹用于存放工程后期生成的文件，例如：/obs-xxx/data-cat。
	- 如需要提前上传进行标注的图片，请创建一个空文件夹，然后将需要标注的图片存放在该文件夹下，图片的文件格式支持JPG、JPEG、PNG、BMP。
	- 如您将已标注好的图片上传至OBS，请按照如下规范上传：
	`|-- data_url`
	`|--a.jpg`
	`|--a.txt（or a.xml）`
	`|--b.jpg`
	`|--b.txt（or b.xml）`
	`...`
	`|--label_map_dict`
	其中，“data_url”为文件夹名，图像分类图片和标签文档（.txt）需同名，物体检测图片和标签文档（.xml）同名；“label_map_dict”为标签的索引文件，标签索引值为标签名-四位字母或数字组成随机字符串，例如cat-JNac。
	且data_url文件夹下不允许存在以上说明的文件之外的任何其他文件夹及文件。
	2. 预测分析项目的OBS数据路径需符合以下规则：
	- 输入数据的OBS路径应指向数据文件，且数据文件不能直接放在OBS桶的根目录下，例如：/obs-xxx/data/input.csv。
	- 输入数据的格式必须为csv格式，数据文件不包括表头，有效数据行数必须大于150行。

描述	对项目的简要描述。

步骤3 单击“确定”，完成自动学习项目的创建。

----结束

4.3 图像分类

4.3.1 数据标注

图像分类“数据标注”界面主要分为4部分内容，如图4-2所示，4个区域的内容介绍如表4-3所示。

图 4-2 图像分类数据标注界面

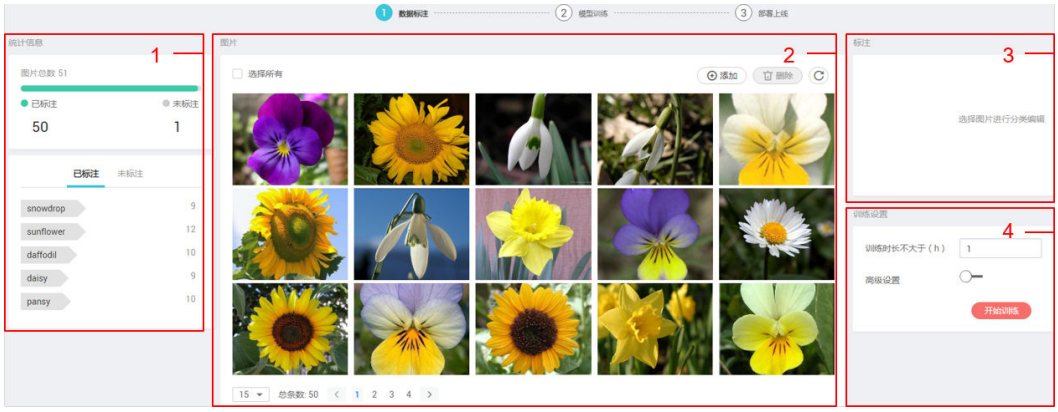


表 4-3 界面内容介绍

区域	说明
1	图片的统计信息，包含图片总数、已标注图片数量及标签名称、未标注图片数量。
2	图片的相关操作，可对图片进行添加、删除、预览操作，并可同步OBS桶中最新上传的图片。同时可以单个或批量选择图片，然后在区域3完成图片标签的添加。 图片的相关详细操作请参见： ● 图片预览 ● 添加图片 ● 删除图片 ● 数据源同步
3	图片标注，详细标注操作请参见图片标注。
4	完成数据标注后，设置训练参数，开始自动训练。

图片预览

单击浮于图片上方的“图片预览”，即可进行图片的预览。

图片标注

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。通过ModelArts您可对图片进行一键式批量添加标签，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，具体操作如下。

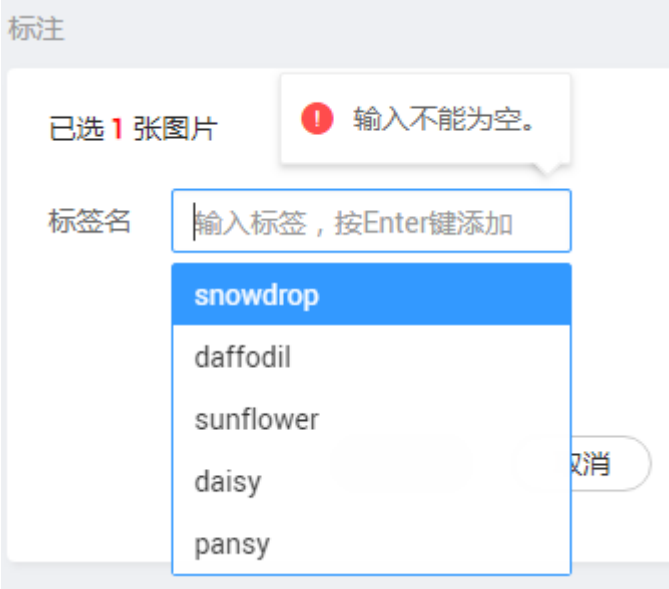
说明

一张图片支持添加多个不同的标签。

- 添加标签
  a. 选择未标注的图片。单击区域1中的“未标注”，然后在区域2单击浮于图片上方的“选择图片”依次选中图片，或勾选上方“选择所有”选中该页面所有图片。

b. 添加标签。选中图片后，在区域3输入标签名称或从弹出的列表中选择已添加的标签，然后按Enter键添加，如图4-3所示。

图 4-3 图像分类图片标注



c. 单击“确定”，完成选中图片的标注操作。

- 删除或修改单个图片标签

a. 单击区域1中的“已标注”，然后在区域2单击浮于图片上方的“选择图片”选中图片。

b. 单击区域3中标签右侧 ×，删除该标签。然后在上方输入新的标签名称，或从弹出的列表中选择已添加的标签，然后按Enter键添加，如图4-4所示。

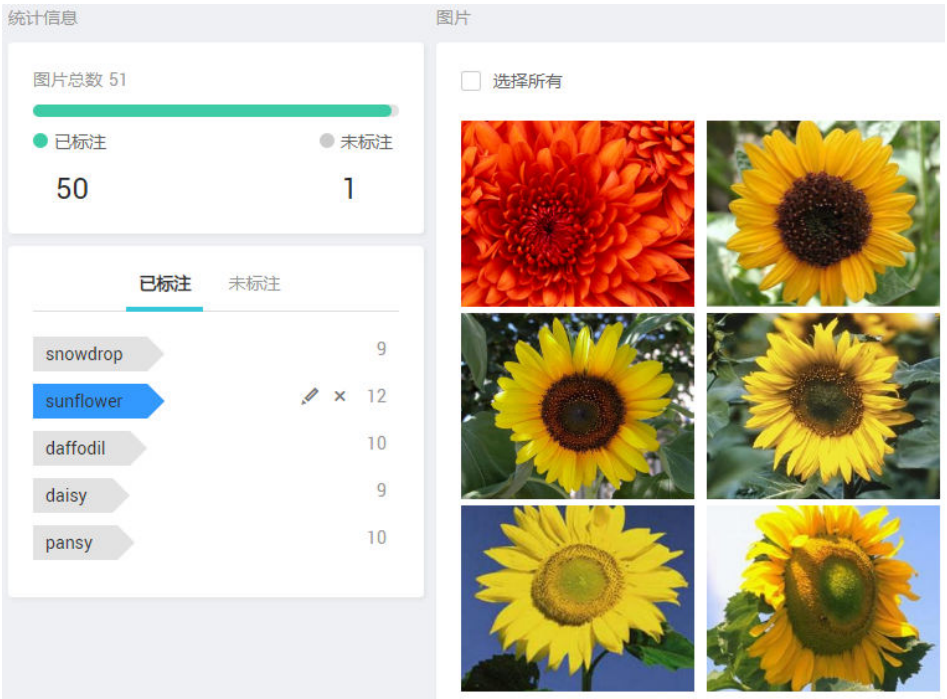
图 4-4 删除并添加新标签

c. 单击“确定”，完成单个图片标签的删除并添加新标签的操作。

- 批量删除或修改图片标签

a. 单击区域1中的“已标注”，然后单击下方需要批量修改或删除的标签名称，如图4-5所示。

图 4-5 批量删除或修改图片标签



b. 单击标签右侧 ，在弹出的对话框中可重命名标签名称。或者单击标签右侧 ，在弹出对话框中，可选择“删除标签”或“删除标签及图片”。

添加图片

通过数据添加操作，可将您本地的图片快速添加到ModelArts，同时自动拷贝到创建项目时选择的OBS桶中。单击区域2中的“添加”，在弹出的对话框中单击“添加文件”并选择要添加的图片，即可完成图片的添加操作。

说明

图片只支持JPG、JPEG、PNG、BMP，且一次上传所有图片的总大小不能超过8MB。

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。单击浮于图片上方的“选择图片”依次选中需要删除的图片，或者勾选上方“选择所有”选中该页面所有图片，然后单击区域2中“删除”，即可完成图片的删除操作。

数据源同步

为了快速获取用户OBS桶中最新图片，可单击区域2右上角的 ，快速将通过OBS上传的图片数据添加到ModelArts。

4.3.2 模型训练

完成图片标注后，可进行模型的训练。本节为您介绍如何进行模型训练的发布、查看训练结果。

发布训练

由于用于训练的图片，至少有2种以上的分类，每种分类的图片数不少于5张。因此在发布训练之前，请确保已标注的图片符合要求，否则“开始训练”会处于灰色状态。

训练设置如图4-6所示，参数设置可参考表4-4所示。

图 4-6 训练设置

表 4-4 训练设置参数说明

名称	说明
训练时长不大于（h）	设置训练时长。 输入值不能小于0.05，如不设置，默认为1。
推理时间不小于（ms）	设置推理时间。 输入值不能小于50，如不设置，默认为300。
推理环境	设置推理环境，当前有2核|8GiB和2核|8GiB |1*P4两种可选。

查看训练结果

训练结束后，可在图像分类“模型训练”界面查看模型训练结果。

图像分类“模型训练”界面主要分为3部分内容，如图4-7所示，3个区域的内容介绍如表4-5所示。

图 4-7 图像分类模型训练界面

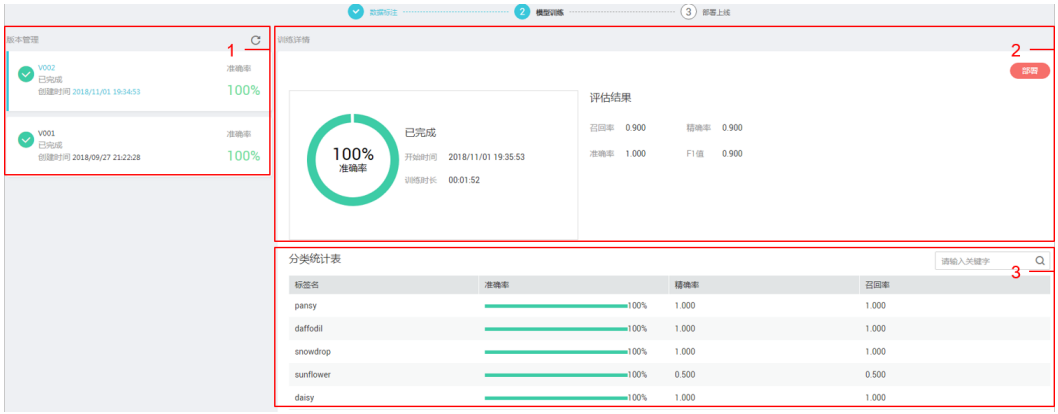


表 4-5 界面内容介绍

区域	说明
1	版本管理。 模型训练的版本，包括版本名称、版本状态、创建时间以及准确率。
2	训练详情。 当前版本的训练详情，包括： ● 训练准确率、训练状态、开始训练时间以及训练时长。 ● 评估结果。
3	分类统计表。 包括不同标签对应的准确率、精确率和召回率。可在左上角文本框中输入标签名，单击 进行查询。

4.3.3 部署上线

模型训练完成后，您可在训练版本管理中选择最优的训练模型部署上线。

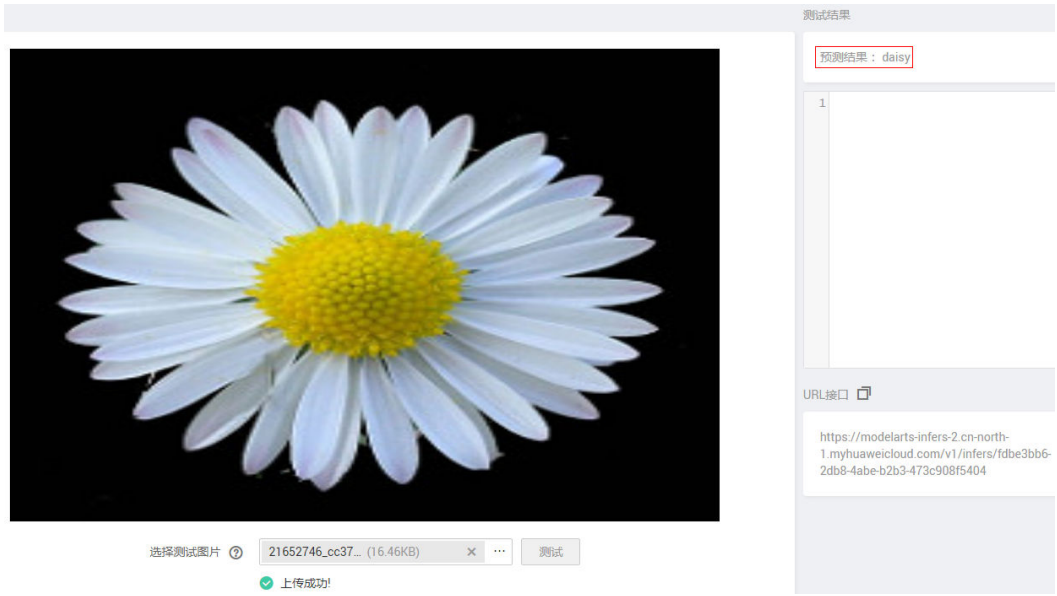
部署模型

完成模型训练后，可选择准确率理想且训练状态为“已完成”的版本部署上线。单击“模型训练”页签右上角“部署”，即可完成模型的部署操作。

服务测试

模型部署完成后，您可添加图片进行测试。在“部署上线”界面，单击图片选择按钮“ ··· ”，然后选择图片。图片上传成功后，单击“测试”即可进行服务的测试，右侧会显示测试结果，如图4-8所示。

图 4-8 服务测试



4.4 物体检测

4.4.1 数据标注

物体检测“数据标注”界面主要分为3部分内容，如图4-9所示，3个区域的内容介绍如表4-6所示。

图 4-9 物体检测数据标注界面

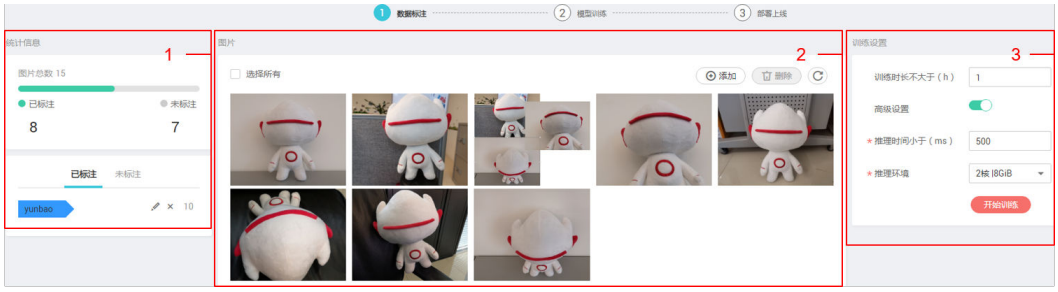


表 4-6 界面内容介绍

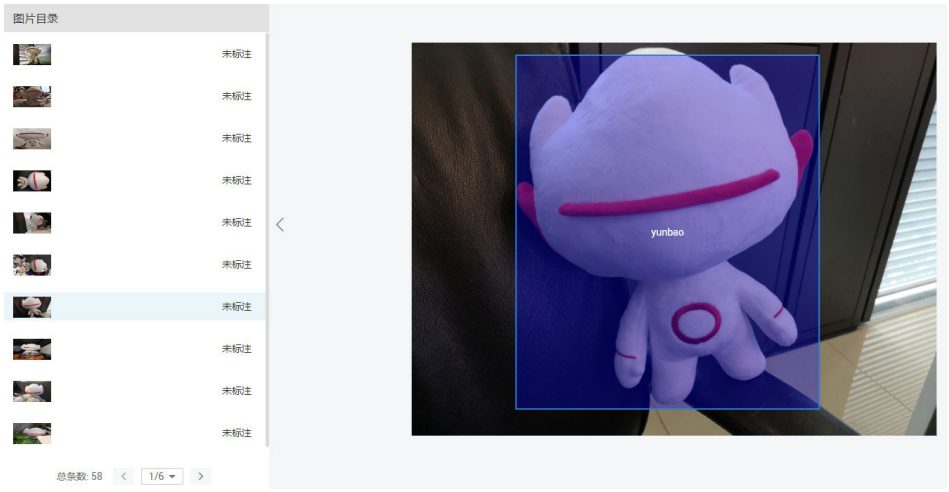
区域	说明
1	图片的统计信息，包含图片总数、已标注图片数量及标签名称、未标注图片数量。

区域	说明
2	图片的相关操作，可对图片进行添加、删除操作，并可同步OBS桶中最新上传的图片。同时可以单个或批量选择图片，进行图片的标注操作。 图片的相关详细操作请参见： ● 图片标注 ● 添加图片 ● 删除图片 ● 数据源同步
3	完成数据标注后，设置训练参数，开始训练。

图片标注

- 物体标注
  a. 单击浮于图片上方的“图片标注”，进入图片标注界面。
  b. 用鼠标框选图片中的物体区域，然后在弹出的对话框中输入标签名称，单击“确定”，如图4-10所示。

  图 4-10 物体检测图片标注

  c. 完成当前图片中所有物体标注后，可选择左侧图片目录中其他未标注图片，重复上一步骤，完成所有图片的标注。
- 删除或修改单个图片物体标注
  a. 单击区域1中的“已标注”。
  b. 单击浮于图片上方的“图片标注”，进入图片标注界面。（单击“已标注”后您可以单击下方标签名称，区域2将仅显示含此标签图片。）
  c. 单击页面右侧的，可删除当前物体标注，并可参见物体标注重新对图片中物体进行标注。单击页面右侧的，可修改当前物体标注名称。
- 批量修改或删除图片物体标注

a. 单击“已标注”，然后单击下方需要批量修改或删除的图片物体标注名称，如图4-11所示。

图 4-11 批量修改或删除图片物体标注

b. 单击标签右侧 ，在弹出的对话框中可重命名标签名称。或者单击标签右侧 ，在弹出对话框中，可选择“删除标签”或“删除标签及图片”。

添加图片

通过数据添加操作，可将用户本地计算机的图片快速添加到ModelArts，同时自动拷贝到创建项目时选择的OBS桶中。单击区域2中的“添加”，在弹出的对话框中单击“添加文件”并选择要添加的图片，即可完成图片的添加操作。

说明

图片只支持JPG、JPEG、PNG、BMP，且一次上传所有图片的总大小不能超过8MB。

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。单击浮于图片上方的“选择图片”依次选中需要删除的图片，或者勾选上方“选择所有”选中该页面所有图片，单击区域2中“删除”，即可完成图片的删除操作。

数据源同步

为了快速获取用户OBS桶中最新图片，可单击区域2右上角的 ，快速将通过OBS上传的图片数据添加到ModelArts。

4.4.2 模型训练

完成数据标注后，可进行模型的训练。本节为您介绍如何进行模型训练的发布、查看训练结果。

发布训练

完成数据标注后，可在“数据标注”页面设置训练参数，训练设置如图4-12所示，参数说明如表4-7所示。

图 4-12 训练设置

表 4-7 训练设置参数说明

名称	说明
训练时长不大于（h）	设置训练时长。 输入值不能小于0.05，如不设置，默认为1。
推理时间不小于（ms）	设置推理时间。 输入值不能小于50，如不设置，默认为500。
推理环境	设置推理环境，当前有2核|8GB和2核|8GB |1*P4两种可选。

查看训练结果

训练结束后，可在物体“模型训练”界面查看模型训练结果。

4.4.3 部署上线

模型训练完成之后，您可在训练版本管理中选择最优训练结果进行模型部署。

部署模型

完成模型训练后，可选择准确率理想且训练状态为“已完成”的版本部署上线。单击“模型训练”页签右上角“部署”，即可完成模型的部署操作。

服务测试

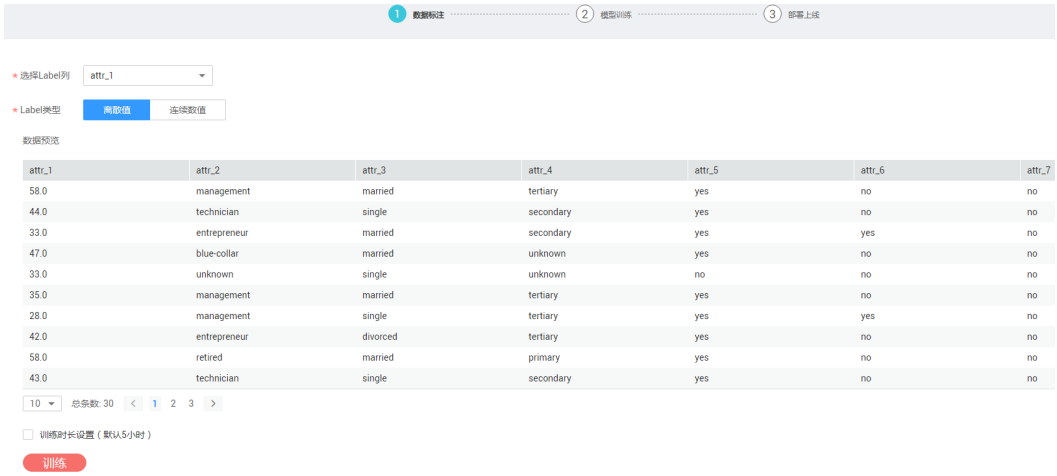
模型部署完成后，您可添加图片进行测试。在“部署上线”界面，单击图片选择按钮“ ··· ”，然后选择图片。图片上传成功后，单击“测试”即可进行服务的测试。

4.5 预测分析

4.5.1 数据标注

在预测分析“数据标注”界面，可预览数据并完成Label列及类型的选择，如图4-13所示。

图 4-13 预测分析数据标注界面



选择 Label 及类型

步骤1 选择数据Label列，在“选择Label列”下拉框中输入需要设置为Label列的名称，单击进行搜索。搜索出来后，单击名称即完成数据Label列的选择。

步骤2 选择Label类型，在“Label类型”选定Label列的数据类型。

----结束

4.5.2 模型训练

完成数据Label列及类型选择后，可进行模型的训练。本节为您介绍如何进行模型训练的发布、查看训练结果。

发布训练

完成数据Label列及类型选择后，单击图4-13左下角“训练”，即可开始模型的训练。

查看训练结果

训练结束后，可在“模型训练”界面查看模型训练结果。预测分析“模型训练”界面主要分为版本管理、训练详情两部分内容，如图4-14所示。

图 4-14 预测分析模型训练界面

表 4-8 界面内容介绍

区域	说明
版本管理	模型训练的版本，包括版本名称、版本状态、创建时间。
训练详情	当前版本的训练详情，包括评估结果、ROC曲线、P-R曲线。其中，单击  ，可将ROC曲线或P-R曲线保存为图片。

4.5.3 部署上线

模型训练完成之后，您可在训练版本管理中选择最优训练结果进行模型部署。

部署模型

完成模型训练后，可选择训练状态为“已完成”的版本部署上线。单击“模型训练”页签右上角“部署”，即可完成服务的部署操作。

服务测试

模型部署完成后，您可输入代码进行测试。在“部署上线”界面的“代码调试”框中输入调试代码，然后单击 “测试”，即可在“返回结果”中输出测试结果，如图4-15所示。

图 4-15 服务测试

代码调试说明如下。

- 输入代码

```
{
"meta": {
"uuid": "10eb0091-887f-4839-9929-cbc884f1e20e"
},
"data": {
"count": 1,
"req_data": [
{
"attr_1": "58",
"attr_2": "management",
"attr_3": "married",
"attr_4": "tertiary",
"attr_5": "yes",
"attr_6": "yes",
"attr_7": "123"
}
]
}
}
```

  其中，attr_1~ attr_7为输入的预测数据。

- 返回结果

```
{
"meta": {
"uuid": "10eb0091-887f-4839-9929-cbc884f1e20e"
},
"result": {
"service_name": "d6909770-f78d-4a9b-ab9d-daad70ccc337",
"service_version": "d657dd99-68f6-438e-9a76-9e8336bf37be",
"count": 1,
"resp_data": [
{
"probabilitycol": 0,
"attr_7": "123",
"attr_6": "yes",
"attr_5": "yes",
"attr_4": "tertiary",
"attr_3": "married",
"attr_2": "management",
"attr_1": 58,
```

```
"predictioncol": "no"
}
]
}
}
```

其中，predictioncol为attr_7的预测结果。

5 数据管理

5.1 数据管理简介

在使用数据进行模型训练之前，您可在“数据管理”页面对数据进行处理、创建数据集并进行数据集版本管理。

- 数据标注

  可创建数据标注作业，可进行图像分类、物体检测两种类型的人工标注。

- 数据集

  可创建数据集、发布多个版本，并可进行版本之间的对比。

5.2 数据标注

在数据管理“数据标注”页面列举了用户所创建的数据标注作业，单击左上方“创建”可创建新的标注作业，创建步骤可参见**创建数据标注作业**。

创建数据标注作业后，单击作业名称可进入数据标注页面，图像分类标注请参见**图像分类数据标注**，物体检测标注请参见**物体检测数据标注**。

创建数据标注作业

步骤1 单击数据标注页面左上方“创建”。

步骤2 在“创建数据标注作业”页面填写参数，单击“创建”。

说明

当前人工标注类型仅支持图像分类和物体检测两种类型的标注作业。

----结束

5.3 数据集

数据管理“数据集”界面主要分为4部分内容，4个区域内容介绍如**表5-1**所示。

图 5-1 数据集界面

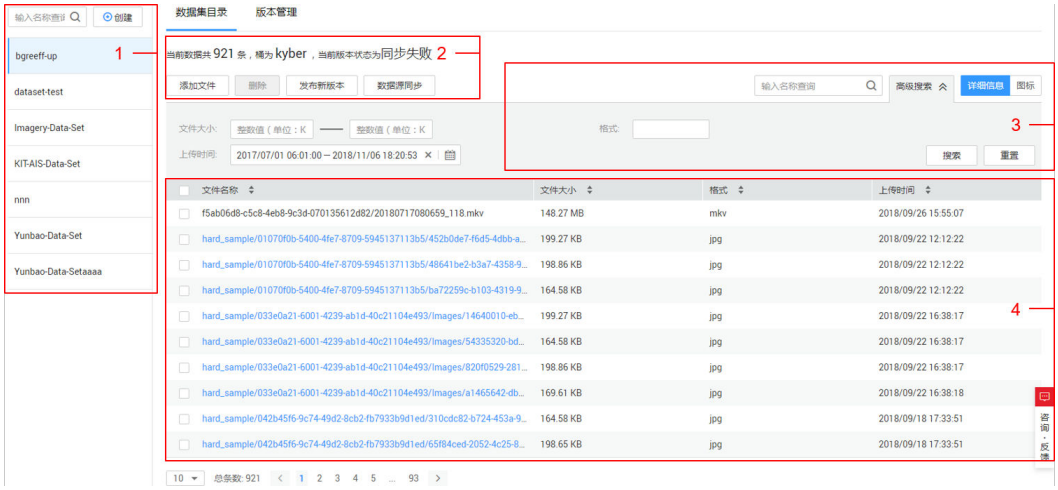


表 5-1 界面内容介绍

区域	说明
1	数据集列表，列举用户所创建的数据集，同时可进行如下操作： ● 查询：输入数据集名称单击查询。 ● 修改：单击“”可修改数据集名称、描述。 ● 删除：单击“”可对数据进行删除。 ● 创建：单击“创建”可创建数据集，详细步骤参见创建数据集。
2	当前版本数据集信息，包括当前数据的数量、桶名称及当前版本状态。 同时可进行如下操作： ● 添加文件：单击“添加文件”，在弹出的对话框中选择文件，单击“确定”，完成图片的添加操作。 ● 删除：选中区域4中的图片，单击“删除”，完成图片的删除操作。 ● 发布新版本：单击“发布新版本”，可在弹出的对话框中填写描述，单击“确定”，完成新版本发布操作。 ● 数据源同步：单击“数据源同步”，可快速将通过OBS上传的图片数据添加到ModelArts。
3	可选择“详细信息”和“图标”两种显示方式。在“详细信息”显示模式下，可单击文件名称对图片文件进行预览。 可输入文件名称进行简单搜索，或者输入文件大小范围、格式、上传时间段进行高级搜索。
4	当前版本数据集文件列表，包含文件名称、文件大小、格式及上传时间。

创建数据集

步骤1 单击图5-1区域1中的“创建”。

步骤2 在弹出的对话框中输入名称、描述，并选择数据集存储路径。

步骤3 单击“确定”，完成数据集的创建。

----结束

版本管理

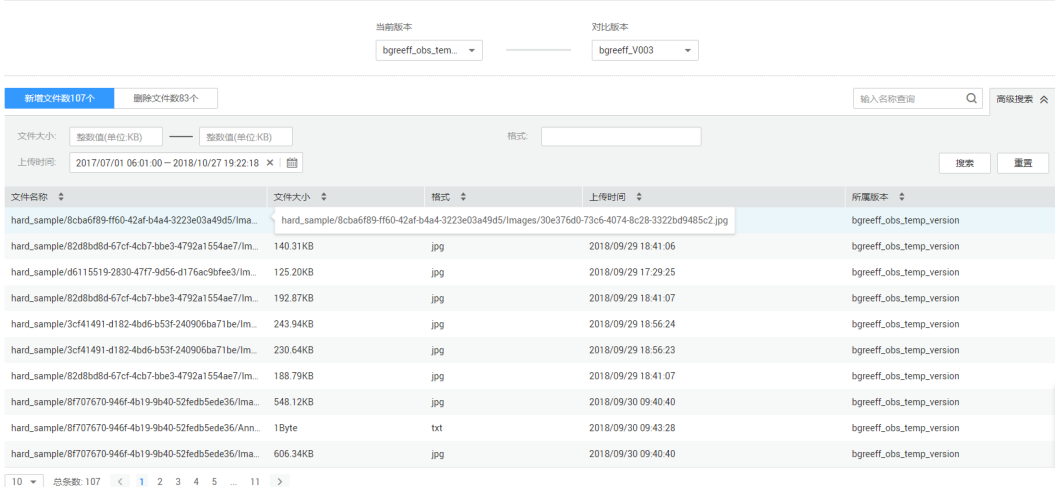
在数据集“版本管理”界面，可查看版本演进过程、版本名称、状态等信息，并且可进行数据集版本的对比，如图5-2所示。

图 5-2 版本管理

数据集版本对比

单击图5-2右侧的“对比”，可进入数据集版本对比界面，如图5-3所示。

图 5-3 数据集版本对比



6 开发环境

开发环境简介

ModelArts集成了基于开源的Jupyter Notebook，可为您提供在线的交互式开发调试工具。您可以通过创建开发环境，自行编写模型训练代码和调测，然后基于该代码进行模型的训练。创建Notebook具体操作请参见创建Notebook。

说明

- 您也可以在本地环境进行代码的开发和调测，然后把已调测好的代码上传至OBS桶中，无需创建开发环境。
- 关于Jupyter Notebook的操作问题，请参见**Jupyter Notebook使用文档**。

开发环境界面说明

在开发环境“Notebook”界面，列举了用户所创建的Notebook。您可以在列表的右上方根据状态进行过滤显示，或者在文本框中输入Notebook名称，单击 进行查询。Notebook列表说明如**表6-1**所示。

表 6-1 Notebook 列表说明

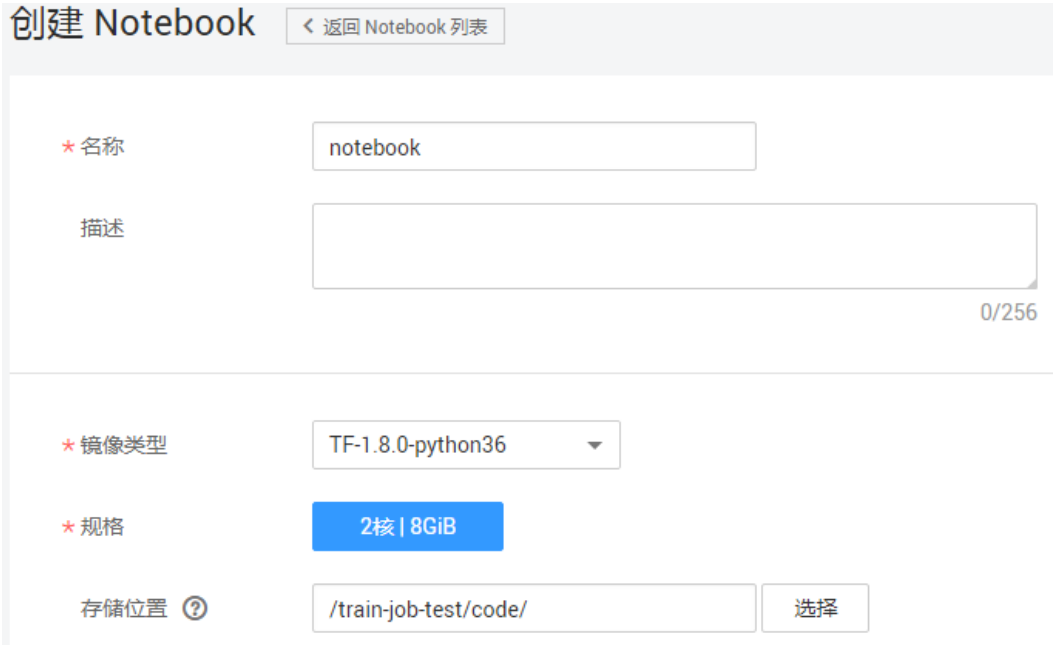
参数名称	说明
名称	用户所创建的Notebook名称。
状态	当前Notebook的运行状态，包括启动中、运行中、停止中、停止、未知。
描述	对Notebook的简要描述，在创建Notebook时填写。
创建时间	用户创建Notebook的时间。
操作	对Notebook的操作，包括： ● “打开”：进入Notebook进行代码开发。 ● “启动”：启动Notebook。 ● “停止”：停止Notebook。 ● “删除”：删除Notebook。

单击Notebook名称前的 ，可以查看该Notebook的详情，包括名称、镜像类型和存储位置信息。

创建 Notebook

步骤1 单击Notebook界面左上方的“创建”，进入创建笔记本页面，如图6-1所示。

图 6-1 创建 Notebook



步骤2 填写Notebook名称和描述，选择镜像类型、规格、存储位置，参数说明如表6-2所示。

表 6-2 参数说明

参数名称	说明
名称	Notebook的名称。只能包含数字、字母、下划线和中划线，长度不能超过20位且不能为空。
描述	对Notebook的简要描述。
镜像类型	当前支持的镜像类型有TF-1.8.0-python36、TF-1.8.0-python27、MXNet-1.2.1-python3.6、MXNet-1.2.1-python2.7、Caffe-1.0.0-python2.7、ML-1.0.0-python27、Spark-2.2.0-python2.7。
规格	当选择的镜像类型为ML-1.0.0-python27时，规格为1核|2GiB。选择其他镜像类型时，规格为2核|8GiB。
存储位置	在Notebook中编写的代码所存储的OBS桶。

步骤3 单击“立即创建”，完成Notebook的创建。

----结束

打开 Notebook

单击Notebook列表右侧的“打开”，进入Jupyter Notebook文件目录界面，如图6-2所示。

图 6-2 Jupyter Notebook 开发界面

您可以通过Jupyter界面控制台上的“ModelArts Examples”选项卡，访问ModelArts提供的所有Jupyter Notebook示例，示例说明如表6-3所示。

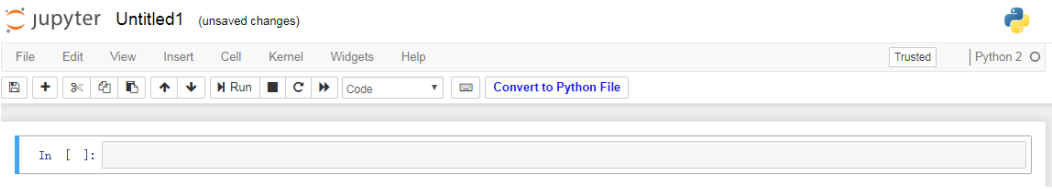
表 6-3 示例说明

示例名称	说明
pyspark	使用基于spark.ml库实现一个"银行理财预期"的二分类，算法选择为逻辑回归，使用的镜像为Spark-2.2.0-python2.7。
sklearn	使用基于sk-learn库实现一个"银行理财预期"的二分类，算法选择为随机森林，使用的镜像为ML-1.0.0-python27。
xgboost	使用xgboost实现一个"银行理财预期"的二分类，算法选择为xgboost分类器，使用的镜像为ML-1.0.0-python27。

Notebook示例使用nbexamplesJupyter扩展，它允许您查看示例笔记本的只读版本或创建它的副本，以便您可以修改和运行它。有关nbexamples扩展的更多信息，请参阅**https://github.com/danielballan/nbexamples**。

单击图6-2“Files”页签右上角“New”，可进入代码开发界面，如图6-3所示。其中，通过“convert to python file”按钮，可直接将您输入的代码保存为py文件到工作目录。

图 6-3 代码开发界面



7 训练作业

7.1 训练作业简介

ModelArts为用户提供了基于CPU+GPU的模型训练环境，您可以使用已完成的模型训练代码创建训练作业。为了方便用户在训练模型时调整参数，ModelArts还提供了可视化功能，您可以将训练作业时得到的Summary文件（TensorBoard日志文件）中的计量指标进行图形化效果展示。

训练作业界面说明

在ModelArts “训练作业”界面，您可进行训练作业的创建、作业参数管理以及TensorBoard相关操作。

在“训练作业”界面，列举了用户所创建的训练作业，如图7-1所示。在该界面您可以**创建训练作业**，也可以在列表右上方根据状态进行过滤显示，或在文本框中输入作业名称，单击 进行查询，训练作业列表说明如表7-1所示。

图 7-1 训练作业界面

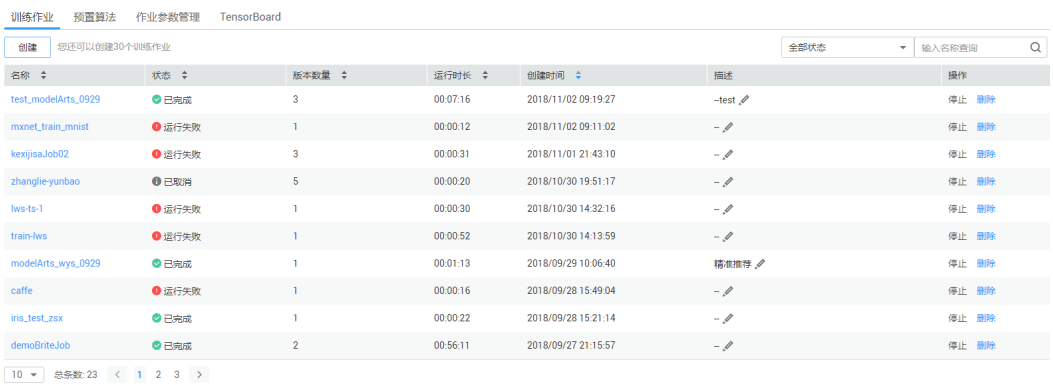


表 7-1 训练作业列表说明

参数名称	说明
名称	训练作业的名称。单击名称，进入该训练作业详情页面，详情参见查看训练作业详情。
状态	当前训练作业运行状态。包括初始化、运行中、已完成、运行失败、删除中、部署中、作业丢失、已取消、提交失败。
版本数量	当前训练作业的版本数量。可单击训练作业名称进入训练作业详情，单击“修改”，启动新版本的训练作业。
运行时长	当前训练作业的运行时长。
创建时间	当前训练作业的创建时间。
描述	当前训练作业的简要描述。
操作	对训练作业的操作。 ● “停止”：停止运行中的训练作业。 ● “删除”：删除当前训练作业。

创建训练作业

步骤1 单击图7-1左上方“创建”，进入创建训练作业界面，如图7-2所示。

图 7-2 创建训练作业

步骤2 在创建训练作业界面填写参数，参数说明如表7-2所示。

表 7-2 参数说明

参数名称	说明
名称	训练作业的名称。
版本	训练作业的版本，版本信息为自动生成。
描述	训练作业的简要描述。
数据来源	训练作业所需要的数据。 ● “数据集”：选择数据集及版本。 ● “数据存储位置”：从OBS桶中选择训练数据。

参数名称	说明
算法来源	训练作业的算法。 ● “预置算法”：ModelArts预置的算法，目前有5种可选，详细介绍请参见预置算法。 ● “常用框架”：选择AI引擎和版本，选择代码目录及启动文件。
运行参数	设置代码中的命令行参数值。请确保参数名称和代码的参数名称一致。
训练输出位置	选择训练结果的存储位置。
计算节点规格	当前有2核|8GiB|1*P100和2核|8GiB两种规格可选。
计算节点个数	计算节点个数，最大为2。

说明

- 如您创建之前已保存过参数配置，可选择已有参数快速配置参数。单击“作业参数配置”，在弹出对话框中选择训练参数。
- 您也可以保存当前参数配置。勾选“保存训练参数”，填写训练参数名称和描述，完成当前参数配置的保存。

步骤3 单击“立即创建”，完成训练作业的创建。

----结束

查看训练作业详情

在训练作业列表中，单击作业名称即可进入训练作业详情页面。以AI引擎为TensorFlow的训练作业为例：“版本管理”页签主要分为5部分，如图7-3所示，5个区域的内容介绍如表7-3所示。

图 7-3 训练作业详情界面

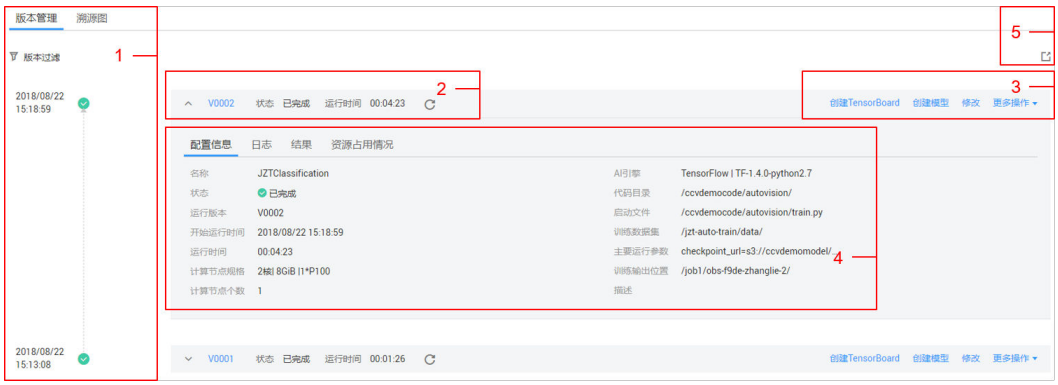


表 7-3 界面内容介绍

区域	说明
1	版本过滤显示，可单击 ，选择显示的版本。
2	当前版本的信息，包括版本名称、状态、运行时间。
3	对当前版本的一些操作，包括： ● “创建TensorBoard”：基于当前训练版本创建TensorBoard，详细参见创建TensorBoard作业。 **说明** TensorBoard目前只支持TensorFlow引擎，因此只有采用AI引擎为TensorFlow的训练作业才可以创建TensorBoard作业。 ● “创建模型”：基于当前训练版本创建模型，详细参见导入模型。 ● “修改”：修改当前训练参数，单击“确定”，启动新版本的训练作业。 ● “更多操作”：可保存当前训练的作业参数、停止当前运行中的训练和删除当前版本训练作业。
4	当前版本的配置信息、日志、结果和资源占用情况。
5	训练版本对比，可查看训练版本之间的对比，包含运行参数、F1值、召回率、精确率、准确率。

在“溯源图”页签可查看数据、训练、模型及服务之间的溯源图。

7.2 预置算法

在“预置算法”页签，列举了所有ModelArts预置的已训练好的模型，您可以在列表的右上方文本框中输入模型名称，单击 进行查询操作，如图7-4所示，预置模型列表说明如表7-4所示。

图 7-4 预置算法

	名称	用途	引擎类型	精度	大小（MB）	创建时间	操作
⌄	retinanet_resnet_v1_50	物体的类别和位置	TensorFlow,TF-1.8.0-python2.7	83.15%(mAP)	255.15	2018/11/07 11:32:44	创建训练
⌄	inception_v3	图像标签	TensorFlow,TF-1.8.0-python2.7	78.00%(top1), 93.90%(top5)	103.78	2018/11/06 21:22:57	创建训练
⌄	darknet_53	图像标签	MXNet,MXNet-1.1.0-python2.7	78.56%(top1), 94.43%(top5)	158.89	2018/11/06 10:57:56	创建训练
⌄	ResNet_v1_50	图像标签	TensorFlow,TF-1.8.0-python2.7	74.2%(top1), 91.7%(top5)	200.84	2018/06/21 13:53:07	创建训练
⌄	Faster_RCNN_ResNet_v1_50	物体的类别和位置	TensorFlow,TF-1.8.0-python2.7	73.6%(mAP)	281.16	2018/06/21 13:53:07	创建训练
⌄	Faster_RCNN_ResNet_v2_50	物体的类别和位置	MXNet,MXNet-1.1.0-python2.7	80.05%(mAP)	182.09	2018/05/03 10:07:04	创建训练
⌄	SegNet_VGG_BN_16	像素级分类标签	MXNet,MXNet-1.1.0-python2.7	89%(pixel acc)	112.41	2018/05/03 10:07:04	创建训练
⌄	ResNet_v2_50	图像标签	MXNet,MXNet-1.1.0-python2.7	75.55%(top1), 92.6%(top5)	97.76	2018/03/28 14:33:00	创建训练

表 7-4 模型列表说明

参数名称	说明
名称	模型的名称，例如：“ResNet_v1_50”。单击名称前的 ，查看模型详情信息，如表7-5所示。

参数名称	说明
用途	模型的应用场景，包括图像标签、物体的类别和位置、像素级分类标签。
引擎类型	模型训练时所使用的深度学习引擎类型。
精度	模型的识别精度。 ● 图像标签模型使用“top1”、“top5”。 ● 物体的类别和位置模型使用“mAP”。 ● 像素级分类标签模型用“pixel acc”。
大小(MB)	模型的大小，单位为MB。
创建时间	模型的创建时间。
操作	对当前模型的操作。 “创建训练”：使用该预置模型创建训练作业，操作步骤请参见使用预置模型创建训练。

单击模型名称前的 ，可以查看该模型的详情，如表7-5所示。

表 7-5 模型详情

参数名称	说明
训练数据集	训练该模型所使用的数据集。
数据格式	使用该模型时要求的数据输入格式。
运行参数	训练该模型时的所使用的参数。
引擎类型	训练该模型时所使用的深度学习引擎。
模型输出	模型的输出结果。
描述	模型的简要描述。

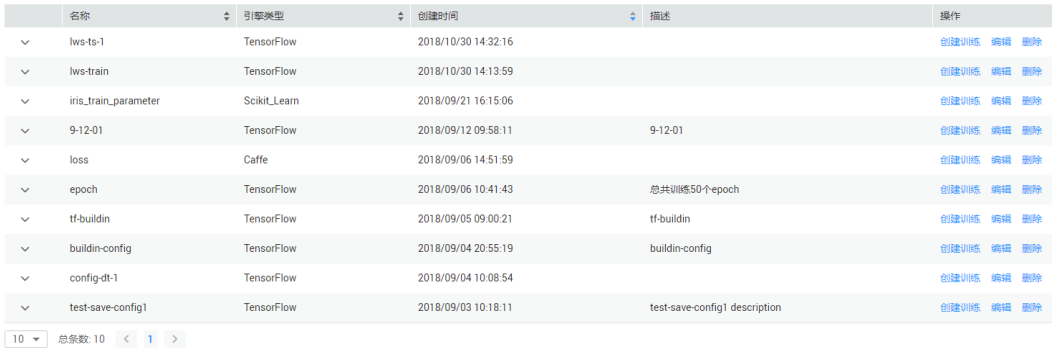
使用预置模型创建训练

单击预置算法列表右侧的“创建训练”，即可进入创建训练作业页面，参数填写可参见创建训练作业。

7.3 作业参数管理

在“作业参数管理”界面，列举了用户所创建的训练参数，如图7-5所示。您可以对训练参数进行编辑、删除操作，还可以基于已创建的训练参数快速创建训练。

图 7-5 作业参数管理



7.4 TensorBoard

对于采用AI引擎为TensorFlow的训练作业，您可以使用模型训练时产生的Summary文件来创建TensorBoard作业。

为了保证训练结果中输出Summary文件，您需要进行如下操作：

1. 对于TensorFlow用户：需要在代码里添加Summary相关代码。
2. 对于MoXing用户：要求MoXing中mox.run中的参数save_summary_steps>0，并且超参summary_verbosity≥1。

说明

- 如果您想显示其他指标，可以在model_fn的返回值类型mox.ModelSpec中log_info中添加张量（仅支持0阶张量，即标量），添加的张量会被写入到Summary文件中。
- 如果您希望在Summary文件中写入更高阶的张量，只需要在model_fn中使用TensorFlow原生的tf.summary的方式添加即可。

在“TensorBoard”页签，列举了用户所创建的TensorBoard作业。您可以在右上方根据状态进行过滤显示，或者在文本框中输入TensorBoard作业名称，单击 进行查询。TensorBoard作业列表说明如表7-6所示。

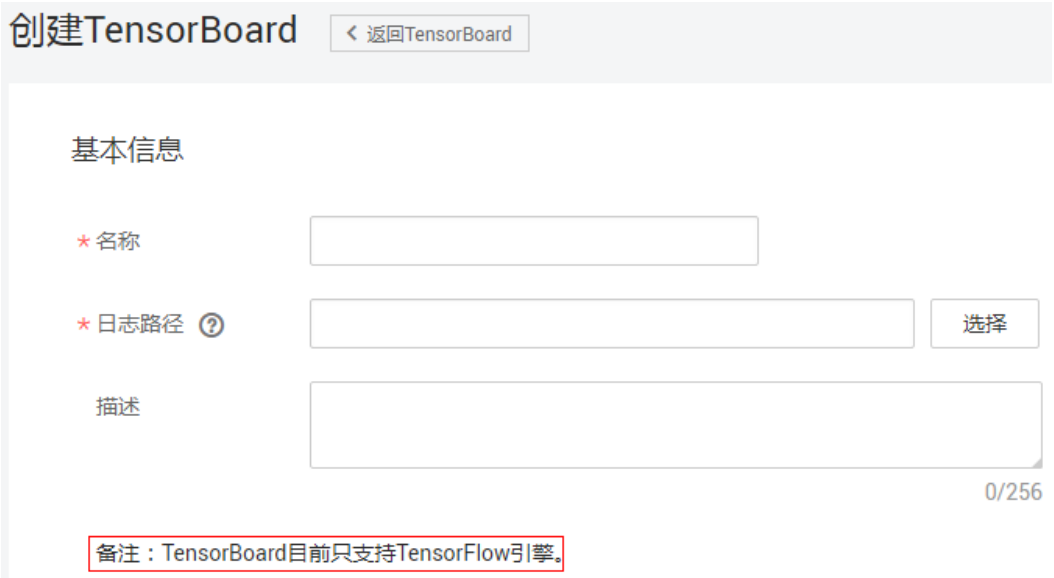
表 7-6 TensorBoard 作业列表说明

参数名称	说明
名称	TensorBoard作业的名称。
状态	当前TensorBoard作业运行状态，包括初始化、排队中、运行中、已取消、作业丢失。
运行时长	TensorBoard作业的运行时长。
创建时间	TensorBoard作业的创建时间。
描述	TensorBoard作业的简要描述。
操作	对TensorBoard作业的操作，包括： ● “停止”：停止运行中或创建中的TensorBoard作业。 ● “删除”：删除当前TensorBoard作业。 ● “运行”：运行当前TensorBoard作业。

创建 TensorBoard 作业

步骤1 单击TensorBoard页签左上方“创建”，进入创建TensorBoard界面，如图7-6所示。

图 7-6 创建 TensorBoard



步骤2 填写TensorBoard作业名称、选择日志路径。

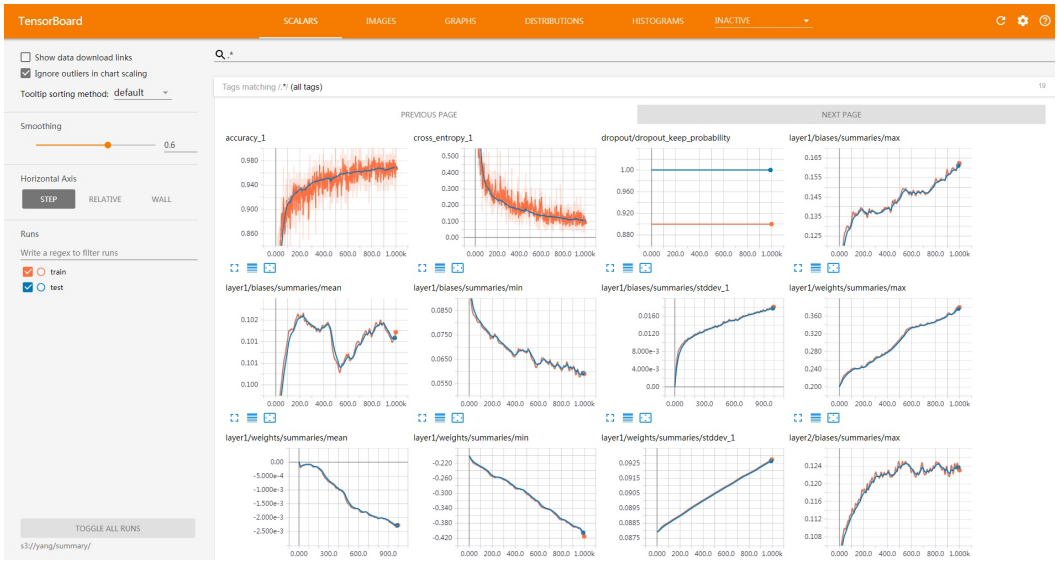
步骤3 单击“立即创建”，完成TensorBoard作业的创建。

----结束

打开 TensorBoard 作业

单击TensorBoard作业名称，即可打开TensorBoard作业显示界面，如图7-7所示。

图 7-7 TensorBoard 界面



8 模型管理

模型管理简介

AI模型的开发和调优往往需要大量的迭代和调试，数据集的变化、训练代码或参数的变化都可能会影响模型的质量，如不能统一管理开发流程元数据，可能会出现无法重现最优模型的现象。

ModelArts模型管理可导入所有训练版本生成的模型，可实现对所有迭代和调试的模型进行统一管理，通过数据、训练和模型之间的版本演进溯源图，还可实现模型的溯源管理。

模型管理界面说明

ModelArts “模型管理”界面主要分为2部分，如图8-1所示，2个区域内容介绍如表8-1所示。

图 8-1 模型管理界面

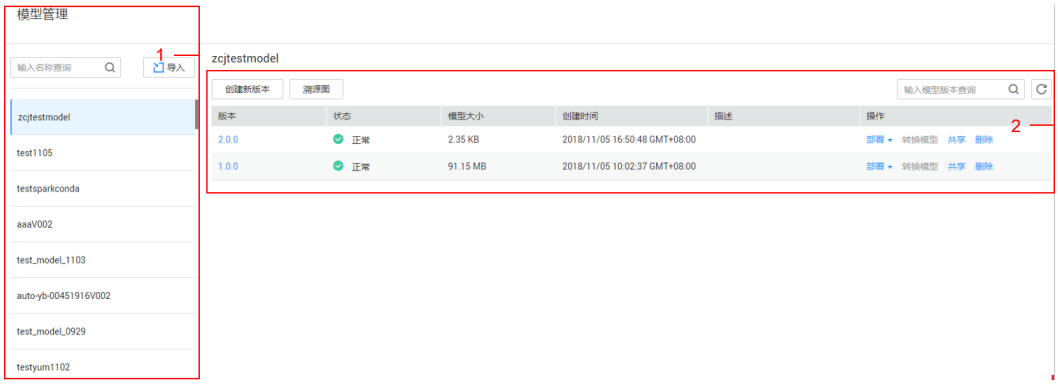


表 8-1 界面内容介绍

区域	说明
1	模型列表，列举了用户所创建的模型，同时可进行如下操作： ● 查询：输入模型名称单击 查询。 ● 删除：选中模型后，单击模型右侧“ ”可删除当前选中的模型。 ● 导入模型：单击“导入”可导入新模型，详细步骤参见**导入模型**。
2	列举了当前模型版本信息，可查看版本详情、溯源图，可创建新版本、部署模型、共享模型以及删除版本。 模型版本相关详细操作请参见： ● **创建新版本** ● **查看版本详情** ● **部署模型** ● **共享模型**

导入模型

步骤1 单击**图8-1**区域1上方的“导入”，进入导入模型页面。

步骤2 在导入模型页面填写相关参数，参数说明如**表8-2**所示。

表 8-2 参数填写说明

参数名称	说明
名称	模型的名称
版本	设置所创建模型的版本。
描述	模型的简要描述。
元模型来源	可从训练中选择，或指定元模型位置。 ● “从训练中选择”：选择训练作业及版本。 ● “指定元模型位置”：选择元模型存储路径和AI引擎，当前支持TensorFlow、MXNet、Caffe、Spark_MLlib、AutoModel、Scikit_Learn、XGBoost几种引擎。
推理代码	自定义模型推理处理逻辑。 通过配置推理代码，可实现对推理请求API的输入预处理或推理结果输出的后处理。当前仅支持Python代码，且当训练作业引擎为MXNet时，必须添加推理代码，代码实现可参见**https://github.com/huawei-clouds/modelarts-mxnet**。

参数名称	说明
高级配置	● 入参 定义模型的输入参数，便于模型的元数据管理。当需要将模型部署为批量服务时必须设置正确的入参，批量服务根据此入参构造推理请求体。 ● 出参 定义模型的输出参数，便于模型的元数据管理。 ● 运行时依赖 – 从ModelArts中训练出的模型且未自定义推理代码，则此项可不填，ModelArts会自动配置所有运行时依赖。 – 如自定义推理代码时引入了第三方库，则需要补充依赖包，通过配置安装方式、安装包名称、版本、约束等。

步骤3 单击“立即创建”，完成模型的导入。

----结束

创建新版本

单击图8-1中区域2左上角的“创建新版本”，进入导入模型界面，参见导入模型填写参数，单击“立即创建”，完成新版本的创建操作。

部署模型

单击图8-1中区域2右上角的“部署”，可快速将当前版本模型部署为在线服务、批量服务、边缘服务，部署参数设置可分别参见在线服务、批量服务、边缘服务。

查看版本详情

单击版本名称即可进入版本详情页面。

共享模型

通过ModelArts共享功能，您可以将您训练好的模型共享给您指定的用户。您也可以在控制台“市场”中查看其它用户的共享给您的模型以及您的共享。

将您的模型共享给其他用户操作如下。

步骤1 单击图8-1区域2右上角的“共享”。

步骤2 在弹出的对话框中，填写发布者名称、选择模型图标和共享的用户ID，如图8-2所示，参数填写如表8-3所示。

图 8-2 共享模型

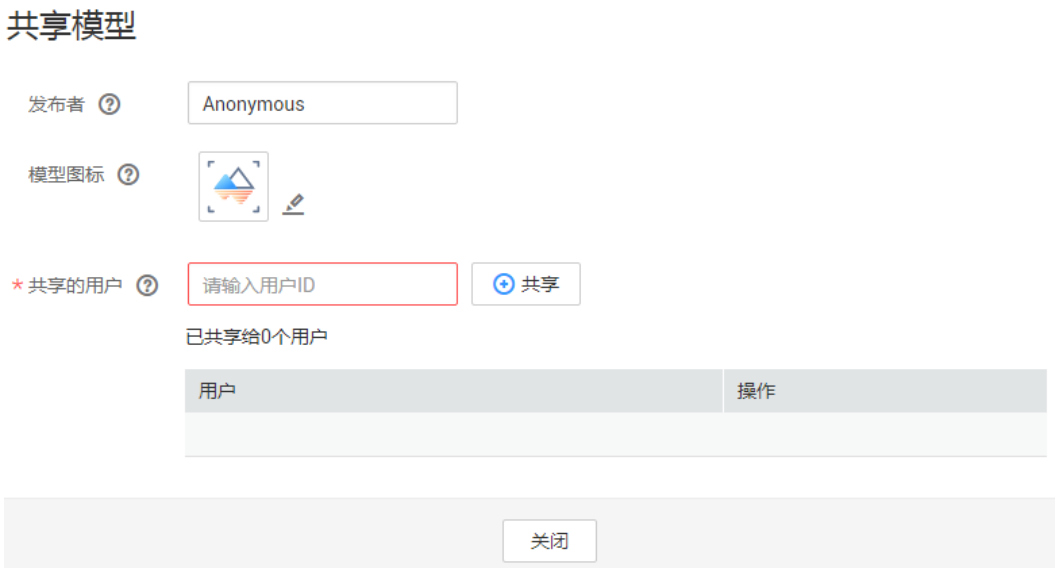


表 8-3 参数说明

名称	说明
发布者	模型市场中显示的发布者名称，共享后将不能修改。
模型图标	模型市场中显示的模型图标。可单击“ ”选择共享图标，且共享后将不能修改。
共享的用户	输入用户ID。 **说明** ● 用户ID可从“我的凭证”中获取。 ● 输入多个用户ID时，用户ID之间请用“，”隔开，且不允许出现特殊字符及空格。

步骤3 单击“关闭”，完成模型的共享操作。

您可以在ModelArts管理控制台“市场”中“**我的共享**”页签查看您的共享。

----结束

9 部署上线

9.1 部署上线简介

在完成训练作业并生成模型后，可在“部署上线”页面对模型进行部署。ModelArts当前支持三种部署类型：在线服务、批量服务和边缘服务。

- **在线服务**

  在线服务对每一个推理请求会同步给出推理的结果。成功部署为在线服务后，可查看服务详情并进行服务测试。

- **批量服务**

  批量服务是对批量数据进行推理的批量推理作业。

- **边缘服务**

  边缘服务是将模型部署到华为云智能边缘平台。

9.2 在线服务

在线服务界面说明

在部署上线“在线服务”界面，列举了用户所创建的在线服务，如**图9-1**所示。您可以在右上方文本框中输入服务名称，单击 进行查询，服务列表说明如**表9-1**所示。

单击左上角“部署”可以部署新的服务，具体操作请参见**部署为在线服务**。

说明

公测期间，在线服务运行30分钟后自动停止。如需继续使用，请重新启动服务。

图 9-1 在线服务

表 9-1 在线服务列表说明

参数名称	说明
服务	在线服务的名称。单击服务的名称，可查看服务详情。
状态	当前在线服务的状态，包括部署中、运行中、停止、异常、告警。 **说明** ● 当状态为“异常”时，请检查参数配置并重新部署。 ● 当状态为“告警”时，需要停止当前服务并重新启动。
推理方式	服务的推理方式，在线服务。
调用失败次数/调用总次数	服务调用的失败次数/总次数。
发布时间	当前在线服务的发布时间。
描述	在线服务的简要描述。
操作	对服务的操作，包括： ● “修改”：可对服务的描述、选择模型及配置进行修改。 ● “停止”：停止当前运行中的服务。 ● “启动”：启动当前已停止的服务。

部署为在线服务

步骤1 单击在线服务界面左上方的“部署”，在部署页面填写参数，如图9-2所示，参数说明如表9-2所示。

图 9-2 部署界面

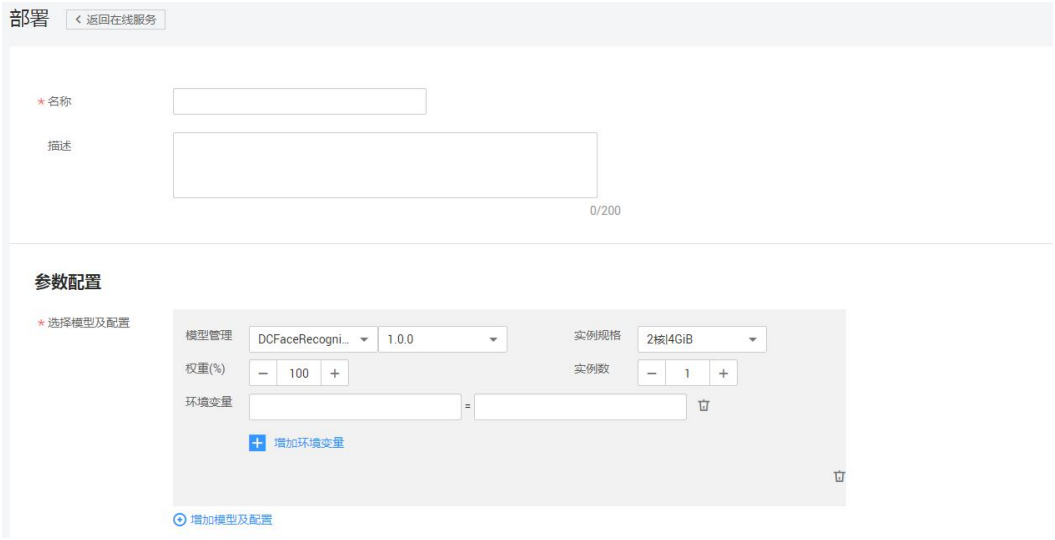


表 9-2 参数说明

参数名称	说明
名称	在线服务的名称。
描述	在线服务的简要描述。
模型管理	选择训练好的模型及版本。
分流	设置当前实例节点的流量占比。 如您仅部署一个版本模型，请设置为100%。如您添加多个版本进行灰度发布，多个版本分流之和设置为100%。
计算节点规格	当前有2核|8GiB、2核|8GiB|1*P4两种可选。
计算节点个数	当前可选1或2个计算节点个数。
环境变量	设置环境变量。
添加模型进行灰度发布	ModelArts提供多版本支持和灵活的流量策略，您可以通过使用灰度发布，实现模型版本的平滑过渡升级。

步骤2 完成参数填写后，单击“立即创建”。

----结束

查看服务详情

单击服务名称，即可进入服务详情页面。在线服务详情页面分为3部分内容，如图9-3所示，3个区域的内容介绍如表9-3所示。

图 9-3 在线服务详情页面

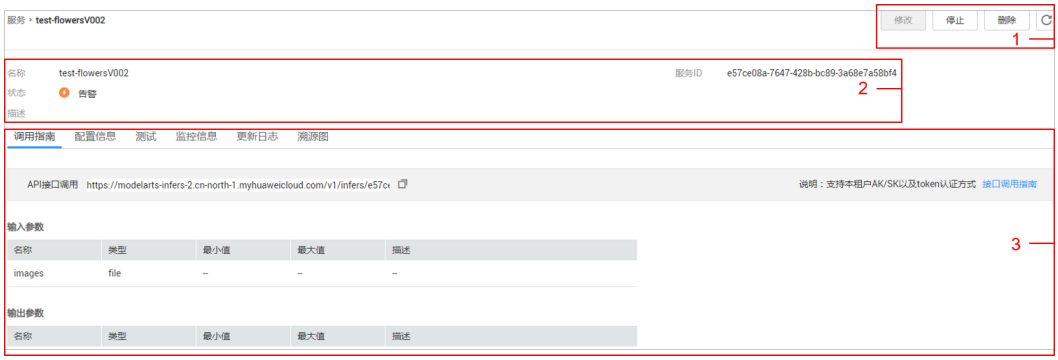


表 9-3 界面内容介绍

区域	说明
1	对服务的操作，包括： ● “停止”：停止当前运行中的服务。 ● “删除”：删除当前服务。 ● “修改”：修改当前服务。
2	服务的相关信息，包括服务名称、服务ID、状态和描述。
3	服务的调用指南、配置信息、测试、监控信息、更新日志及溯源图。

服务测试

在“测试”页签可进行代码调试或添加图片测试，图片测试如图9-4所示，代码调试说明请参见服务测试。

图 9-4 添加图片测试

访问在线服务

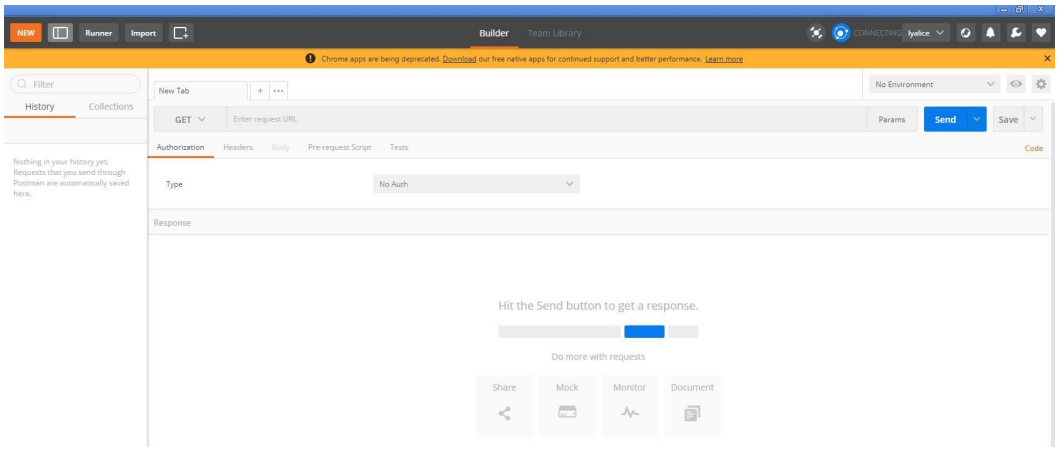
若在线服务的状态处于“运行中”，则表示在线服务已部署成功，您可以使用以下两种方式向在线服务发起预测请求。

方式一：使用图形界面的软件进行预测（以Postman为例）

步骤1 下载Postman软件并安装，或直接在chrome浏览器添加postman扩展程序（也可使用其它支持发送post请求的软件）。

步骤2 打开Postman，如图9-5所示。

图 9-5 Postman 界面



步骤3 在Postman界面填写参数，以图像分类举例说明。

- 选择POST任务，将在线服务的调用地址（通过在线服务详情界面-调用指南页签查看，如图9-6所示）复制到POST后面的方框。Headers栏的Key值填写为“X-Auth-Token”，Value值为您获取到的Token（关于如何获取token，请参考获取请求认证），如图9-7所示。

图 9-6 查看调用 URL

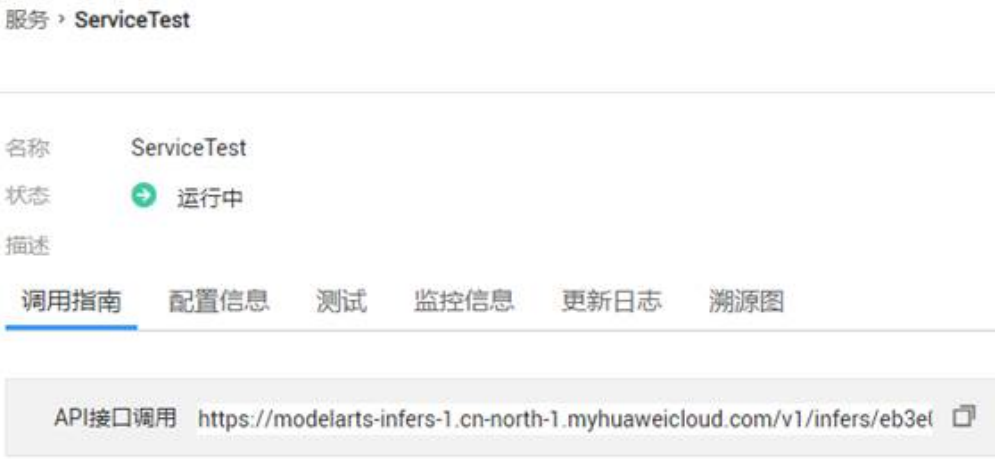
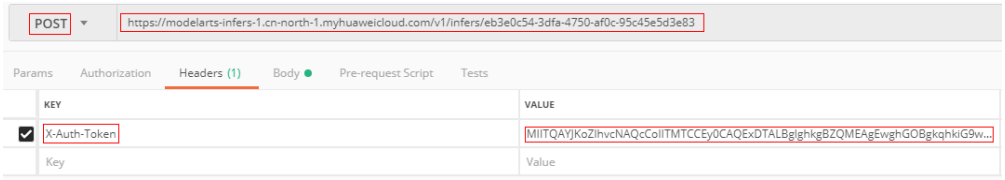


图 9-7 参数填写

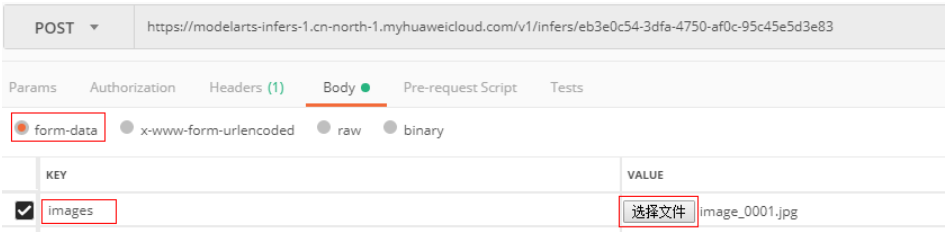


- 在Body栏下，根据模型的输入参数不同，可分为2种类型：文件输入、文本输入。

  a. 文件输入

  选择“form-data”。在“KEY”值填写模型的入参，比如本例中预测图片的参数为“images”。然后在“VALUE”值，选择文件，上传一张待预测图片（当前仅支持单张图片预测），如图9-8所示。

  图 9-8 填写 Body

  b. 文本输入

  选择“raw”，选择JSON(application/json)类型，在下方文本框中填写请求体，请求体样例如下：

  ```
  {
    "meta": {
      "uuid": "10eb0091-887f-4839-9929-cbc884f1e20e"
    },
    "data": {
      "req_data": [
        {
          "sepal_length": 3,
          "sepal_width": 1,
          "petal_length": 2.2,
          "petal_width": 4
        }
      ]
    }
  }
  ```

  其中，meta中可携带uuid，调用时传入一个uuid，返回预测结果时回传此uuid用于跟踪请求，**如无此需要可不填写meta。**data包含了一个req_data的数组，可传入单条或多条请求数据，其中每个数据的参数由模型决定，比如本例中的sepal_length、sepal_width等。

步骤4 参数填写完成，点击“send”发送请求，结果会在Response下的对话框里显示。

- 文件输入形式的预测结果样例如图9-9所示，返回结果的字段值根据不同模型可能有所不同。
- 文本输入形式的预测结果样例如图9-10所示，请求体包含meta及data。如输入请求中包含uuid，则输出结果中回传此uuid。如未输入，则为空。data包含了一个resp_data的数组，返回单条或多条输入数据的预测结果,其中每个结果的参数由模型决定，比如本例中的sepal_length、predictresult等。

图 9-9 文件输入预测结果

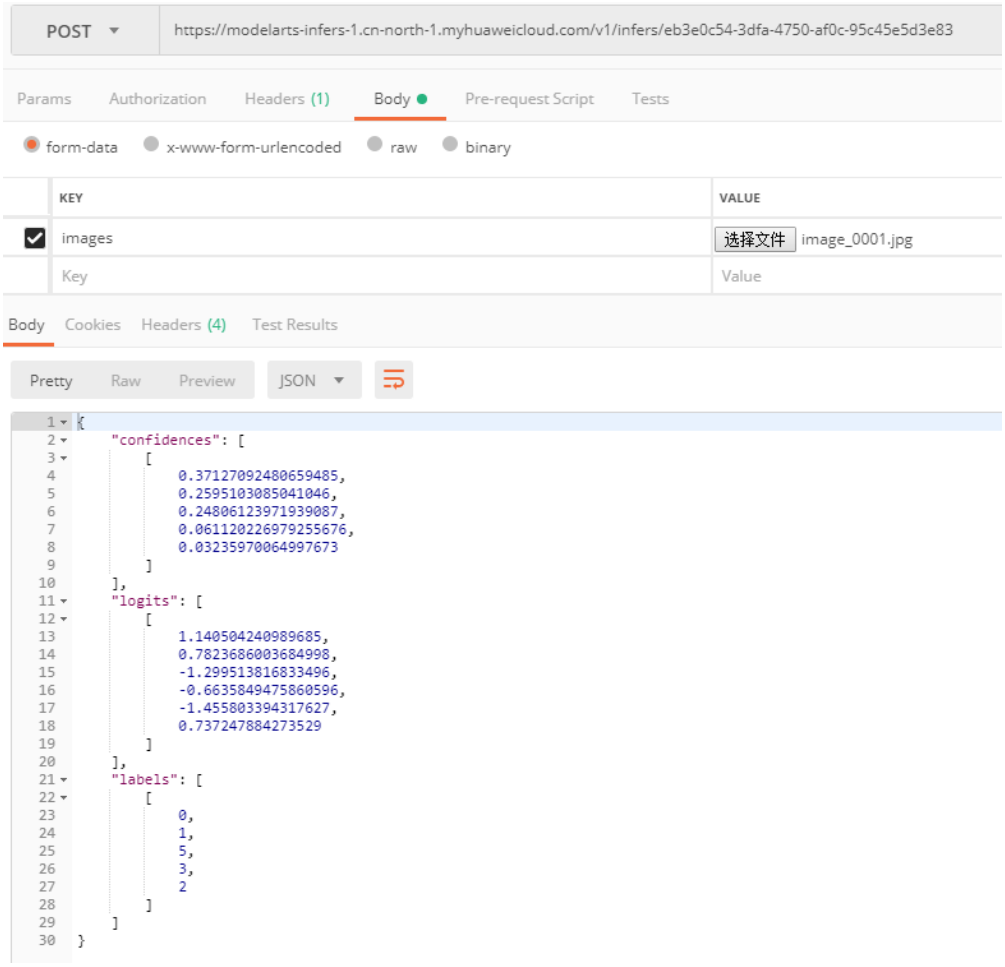
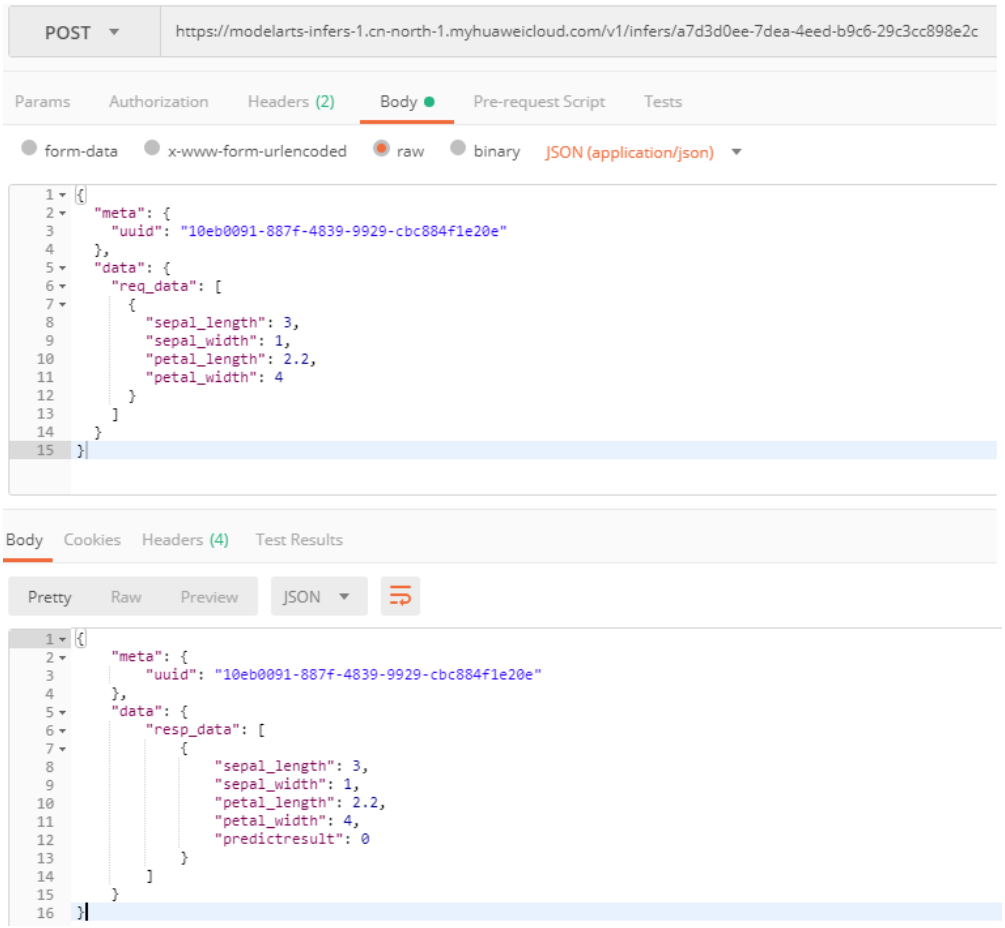


图 9-10 文本输入预测结果



----结束

方式二：使用curl命令发送预测请求

使用curl命令发送预测请求的命令格式也分为文件输入、文本输入两类。

1. 文件输入

   curl -F '**images**=@**图片路径**' -H '**X-Auth-Token**:**Token值**" -X POST **在线服务地址**。

   - “-F”是指上传数据的是文件，本例中参数名为**images**，这个名字可以根据具体情况变化，@后面是图片的存储路径。
   - “-H”是post命令的headers，headers的key值为**X-Auth-Token**，这个名字为固定的， **Token值**是用户获取到的token值（关于如何获取token，请参考**Token认证**）。
   - “POST”后面跟随的是在线服务的调用地址。

   curl命令文件输入样例：

   ```
   curl -F 'images=@/home/data/test.png' -H 'X-Auth-
   Token:MIISkAY***80T9wHQ==' -X POST https://modelarts-infers-1.cn-
   north-1.myhuaweicloud.com/v1/infers/eb3e0c54-3dfa-4750-
   af0c-95c45e5d3e83
   ```

2. 文本输入

   ```
   curl -d '{
   "meta": {
   "uuid": "10eb0091-887f-4839-9929-cbc884f1e20e"
   },
   ```

```
"data": {
"req_data": [
{
"sepal_length": 3,
"sepal_width": 1,
"petal_length": 2.2,
"petal_width": 4
}
]
}
}-H 'X-Auth-Token:MIISkAY***80T9wHQ==' -X POST https://modelarts-infers-1.cn-north-1.myhuaweicloud.com/v1/infers/eb3e0c54-3dfa-4750-af0c-95c45e5d3e83
```

- “-d”是Body体的文本内容。

9.3 批量服务

批量服务界面说明

在部署上线“批量服务”界面，列举了用户所创建的批量服务。您可以在界面右上方文本框中输入服务名称，单击 进行查询，服务列表说明如表9-4所示。

单击左上角“部署”可以部署新的服务，具体操作请参见部署为批量服务。

表 9-4 批量服务列表说明

参数名称	说明
服务	服务的名称。单击服务的名称，可查看服务详情。
状态	当前服务的状态，包括部署中、运行中、运行完成、停止、异常。 **说明** 当状态为“异常”时，请检查参数配置并重新部署。
推理方式	服务的推理方式，批量服务。
发布时间	批量服务的发布时间。
描述	批量服务的简要描述。
操作	对服务的操作，包括： ● “修改”：可对服务的描述、选择模型及版本、数据目录位置、训练参数、计算节点规格及个数进行修改。 ● “启动”：启动当前批量服务。 ● “停止”：停止当前批量服务。

部署为批量服务

步骤1 单击批量服务界面上方的“部署”，在部署页面填写参数，参数说明如表9-5所示。

表 9-5 参数说明

参数名称	说明
名称	批量服务的名称。
描述	批量服务的简要描述。
选择模型及版本	选择训练好的模型及版本。
输入数据目录位置	选择输入数据的目录位置。
环境变量	设置环境变量。
输出数据目录位置	选择批量预测结果的保存位置。
计算节点规格	当前有2核|8GiB、2核|8GiB|1*P4两种可选。
计算节点个数	当前可选1或2个计算节点个数。

步骤2 完成参数填写后，单击“立即创建”，完成部署。

----结束

查看服务详情

单击服务名称，即可进入服务详情页面。批量服务详情页面分为3部分内容，如图9-11所示，3个区域的内容介绍如表9-6所示。

图 9-11 批量服务详情页面

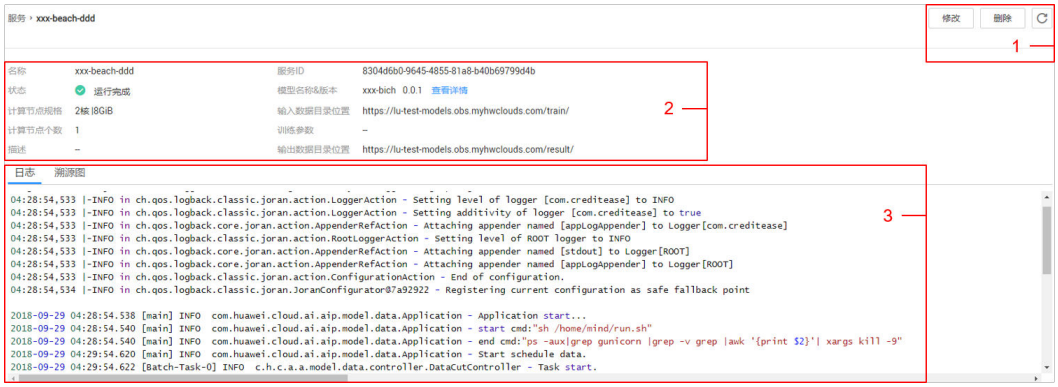


表 9-6 界面内容介绍

区域	说明
1	对服务的操作。 ● “修改”：修改当前批量服务参数。 ● “删除”：删除当前批量服务。
2	服务的相关信息，包括服务名称、服务ID、状态、数据目录位置等。
3	服务的日志和溯源图。

9.4 边缘服务

边缘服务界面说明

在部署上线“边缘服务”界面，列举了用户所创建的边缘服务。您可以在界面右上方文本框中输入服务名称，单击 进行查询，服务列表说明如表9-7所示。

单击左上角“部署”可以部署新的服务，具体操作请参见部署为边缘服务。

表 9-7 边缘服务列表说明

参数名称	说明
服务	边缘服务的名称。单击服务的名称，可查看服务详情。
状态	当前服务的状态，包括部署中、运行中、停止、异常。 **说明** 当状态为“异常”时，请检查参数配置并重新部署。
推理方式	服务的推理方式，边缘服务。
发布时间	边缘服务的发布时间。
描述	边缘服务的简要描述。
操作	对服务的操作，包括： ● “修改”：可对服务的描述、计算节点规格修改，可添加环境变量、边缘节点。 ● “启动”：启动当前边缘服务。

部署为边缘服务

步骤1 单击边缘服务界面上方的“部署”，在部署页面填写参数，参数说明如表9-8所示。

表 9-8 参数说明

参数名称	说明
名称	边缘服务的名称。
描述	边缘服务的简要描述。
模型管理	选择模型及版本。
计算节点规格	当前有2核|8GiB、2核|8GiB|1*P4两种可选。
环境变量	添加环境变量。

步骤2 完成参数填写后，单击选择边缘节点“添加”，在弹出的“添加节点”对话框中选择节点。选择好节点后，单击“确定”。

说明

边缘节点是您自己的边缘计算设备，用于运行边缘应用，处理您的数据，并安全、便捷地和云端应用进行协同。

步骤3 单击“立即创建”，完成边缘服务的部署。

----结束

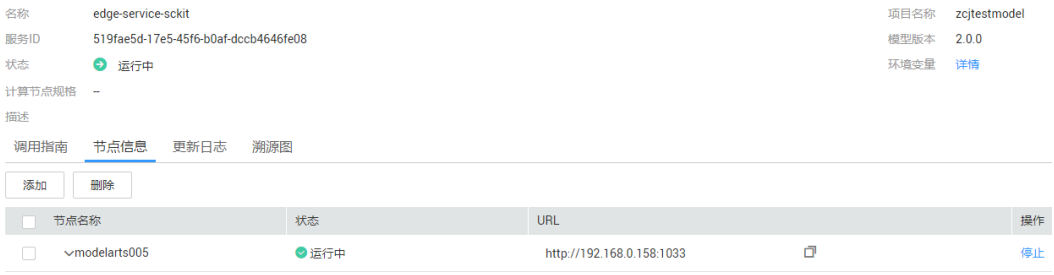
查看服务详情

单击边缘服务名称，可进入当前边缘服务详情页面。

访问边缘服务

当边缘服务和边缘节点的状态都处于“运行中”，如图9-12所示，则表示边缘服务已在边缘节点成功部署。

图 9-12 服务状态



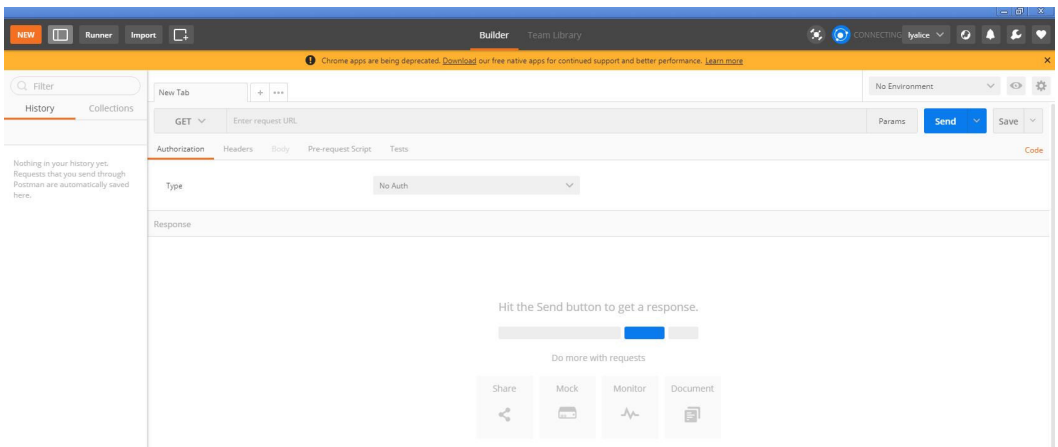
您可以通过以下两种方式，在能够访问到边缘节点的网络环境中，对部署在边缘节点上的边缘服务发起预测请求。

方式一：使用图形界面的软件进行预测（以Postman为例）

步骤1 下载Postman软件并安装，或直接在chrome浏览器添加postman扩展程序（也可使用其它支持发送post请求的软件）。

步骤2 打开Postman，如图9-13所示。

图 9-13 Postman 界面



步骤3 在Postman界面填写参数，以图像分类举例说明。

- 选择POST任务，将某个边缘节点的调用地址（通过边缘服务详情界面-节点信息页签查看，如图9-14所示）复制到POST后面的方框，如图9-15所示。

图 9-14 查看调用 URL

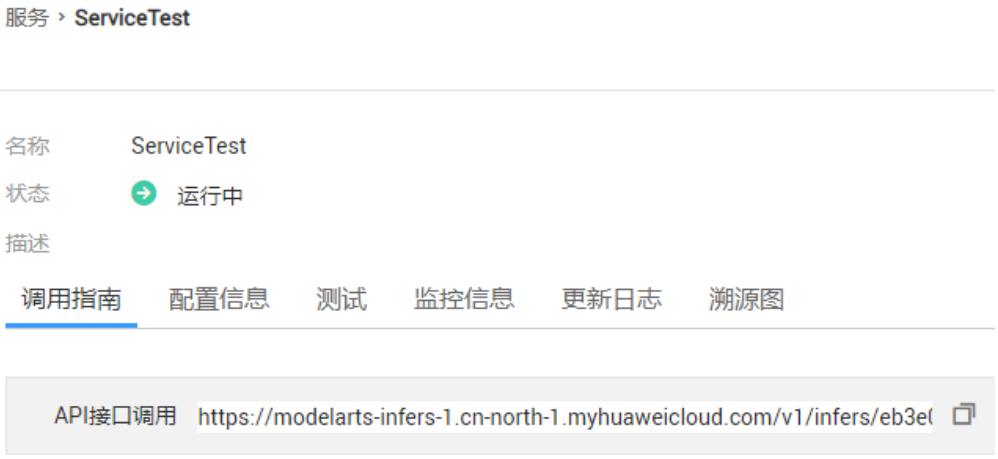
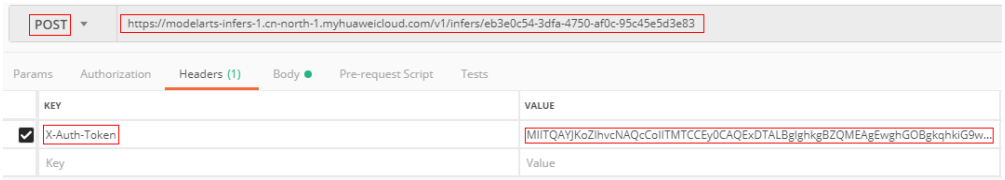


图 9-15 参数填写

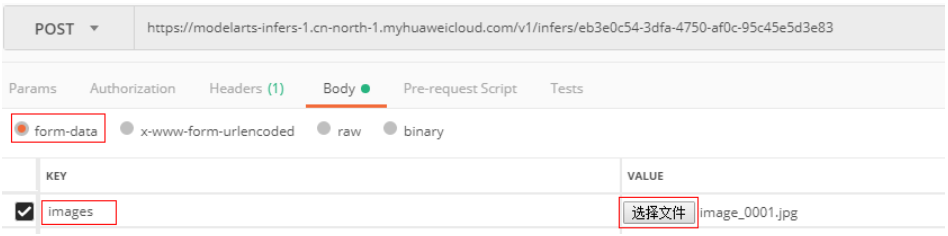


- 在Body栏下，根据模型的输入参数不同，可分为2种类型：文件输入、文本输入。

a. 文件输入

选择“form-data”。在Key值填写模型的入参，比如本例中预测图片的参数为“images”。然后在value值，选择文件，上传一张待预测图片（当前仅支持单张图片预测），如图9-16所示。

图 9-16 填写 Body



b. 文本输入

选择“raw”，选择JSON(application/json)类型，在下方文本框中填写请求体，请求体样例如下。

```
{
"meta": {
"uuid": "10eb0091-887f-4839-9929-cbc884f1e20e"
},
"data": {
"req_data": [
```

```
{
"sepal_length": 3,
"sepal_width": 1,
"petal_length": 2.2,
"petal_width": 4
}
]
}
}
```

其中，meta中可携带uuid，调用时传入一个uuid，返回预测结果时回传此uuid用于跟踪请求，**如无此需要可不填写meta**。data包含了一个req_data的数组，可传入单条或多条请求数据，其中每个数据的参数由模型决定，比如本例中的sepal_length、sepal_width等。

步骤4 参数填写完成，点击“send”发送请求，结果会在Response下的对话框里显示。

- 文件输入形式的预测结果样例如图9-17所示，返回结果的字段值根据不同模型可能有所不同。
- 文本输入形式的预测结果样例如图9-18所示，请求体包含meta及data。如输入请求中包含uuid，则输出结果中回传此uuid。如未输入，则为空。data包含了一个resp_data的数组，返回单条或多条输入数据的预测结果。其中每个结果的参数由模型决定，比如本例中的sepal_length、predictresult等。

图 9-17 文件输入预测结果

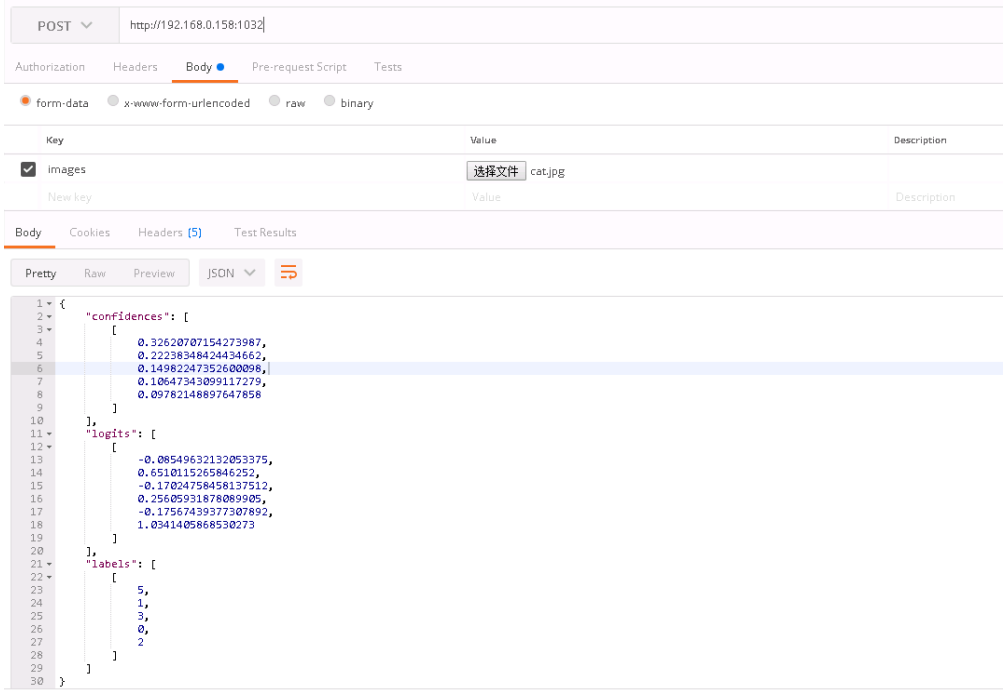
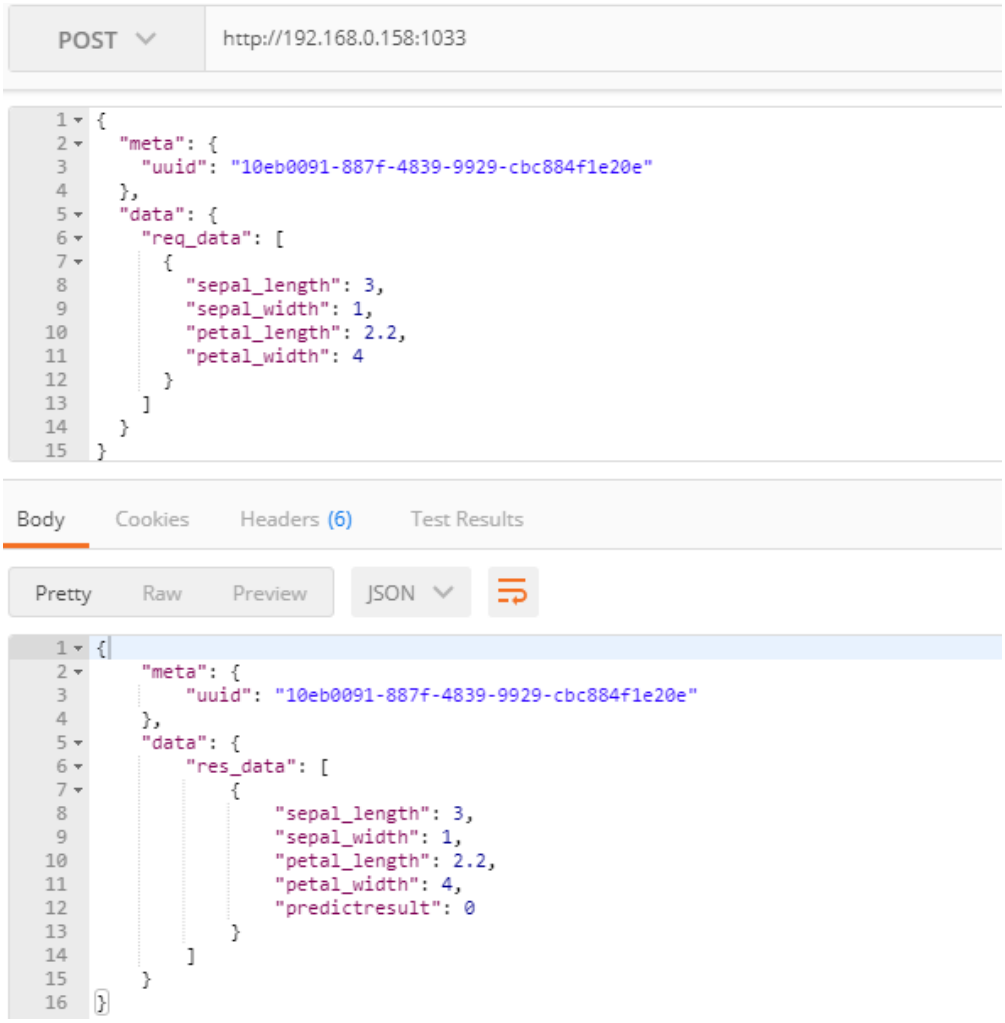


图 9-18 文本输入预测结果



----结束

方式二：使用curl命令发送预测请求

使用curl命令发送预测请求的命令格式也分为文件输入、文本输入两类。

1. 文件输入

   curl -F '**images**=@**图片路径**'-X POST **边缘节点服务地址**。

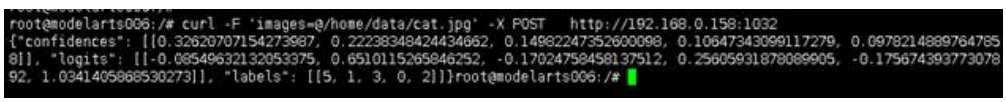
   - “-F”是指上传数据的是文件，本例中参数名为**images**，这个名字可以根据具体情况变化，@后面是图片的存储路径。
   - “POST”后面跟随的是边缘节点的调用地址。

   curl命令文件输入预测样例：

   ```
   curl -F 'images=@/home/data/cat.jpg' -X POST   http://
   192.168.0.158:1032
   ```

   预测结果如图9-19所示。

图 9-19 curl 命令文件输入预测结果



2. 文本输入

```
curl -d '{
"meta": {
"uuid": "10eb0091-887f-4839-9929-cbc884f1e20e"
},
"data": {
"req_data": [
{
"sepal_length": 3,
"sepal_width": 1,
"petal_length": 2.2,
"petal_width": 4
}
]
}
}' -X POST 边缘节点服务地址
```

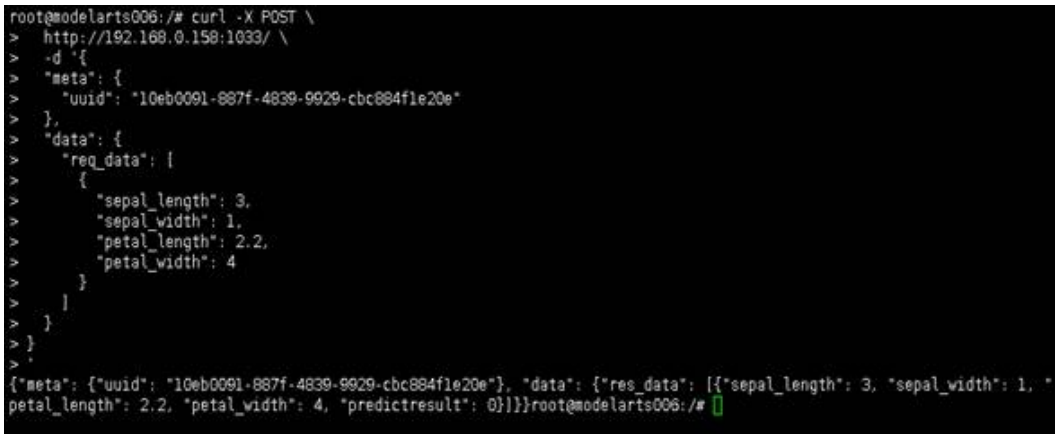
- “-d”是Body体的文本内容，如模型为文本输入，则需要用此参数。

curl命令文本输入预测样例：

```
curl -d '{
"meta": {
"uuid": "10eb0091-887f-4839-9929-cbc884f1e20e"
},
"data": {
"req_data": [
{
"sepal_length": 3,
"sepal_width": 1,
"petal_length": 2.2,
"petal_width": 4
}
]
}
}' -X POST http://192.168.0.158:1033
```

预测结果如图9-20所示。

图 9-20 curl 命令文本输入预测结果



10 市场

ModelArts市场中列举了其他用户共享的模型、ModelArts预置的数据集及您共享给其他用户的模型。

共享模型

在“模型”界面中，列举了其他用户共享给您的所有模型，如图10-1所示。您可按共享类型和模型类型分类显示，或者在搜索框中输入模型名称、模型类型，单击 进行查询。单击模型名称可进入模型详情页面，您可查看当前模型详情，可单击详情页右上角“保存到我的模型”将当前模型导入至“模型管理”，同时您还可以提交您对当前模型的评论和评分。

说明

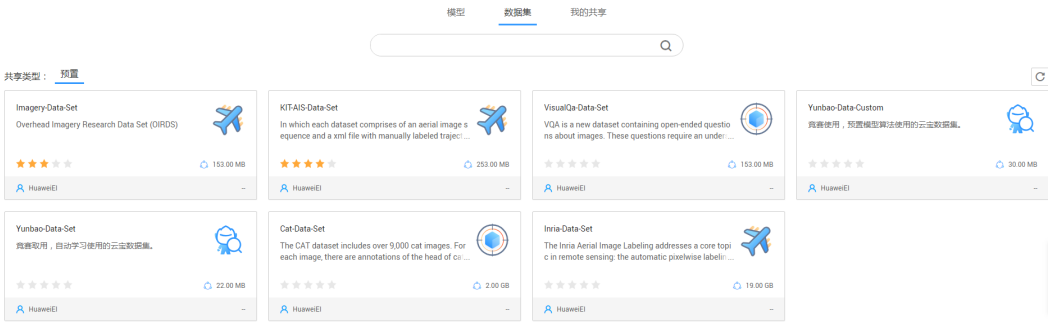
共享类型“公共”中的模型为所有用户可见，“私人”中的模型为仅您或部分用户可见。

图 10-1 模型界面

ModelArts 预置数据集

在“数据集”界面中，列举了ModelArts预置的所有数据集，如图10-2所示。单击数据集名称可进入数据集详情页面，单击详情页右上角“导入到我的数据集”将当前数据集导入至“数据管理->数据集”，同时您还可以提交您对当前数据集的评论和评分。

图 10-2 数据集界面



我的共享

在“我的共享”界面中，列举了您共享给所有用户或您指定用户的模型。您可按共享类型或模型类型进行分类显示，或者在搜索框中输入模型名称、模型类型，单击进行查询。

单击模型名称可进入模型详情页面，可查看当前模型详情。单击详情页面右上角“修改共享”，在弹出的对话框中修改共享模型设置。

说明

- 共享类型“公共”中的模型为所有用户可见，“私人”中的模型为共享给指定用户的模型，仅指定的用户可见。
- 服务公测期间，仅支持共享给指定用户。

11 修订记录

发布日期	修改说明
2018-11-15	第二次正式发布。 ● 修改 修改了**使用flowers数据集对预置ResNet_v1_50模型进行重训练**数据下载路径、部分步骤。
2018-11-08	第一次正式发布。