

# 模型推理服务说明文档

版本	修订日期	版本更新说明
1.1.0	20210902	更新Mxnet版本到1.8.0.post0 增加REID和Landmark推理服务
1.0.0	20210707	初版

## 1. “试一试”功能

### 1.1 检测模型

#### 1.1.1 数据准备

检测模型训练结束后，在输出目录下获取以下文件并放到一个目录下，如/models

1) 训练导出的模型文件（文件名必须一致）

- detect\_centernet-0000.params
- detect\_centernet-symbol.json

2) 类别名称文件

- synset.txt

3) 模型输入信息文件

- signature.json

#### 1.1.2 启动推理服务

docker启动示例

```
nvidia-docker run -d --rm -v /models:/models -p 8080:8080 -p 8081:8081  
mxnet_server:1.1.0 bash start_detection.sh
```

`-v /models:/models`

冒号前面的 /models 为数据准确时文件的存放路径，根据实际存放路径修改

冒号后面的 /models 为映射到容器内的路径，不可修改

`-p 8080:8080`

推理服务端口

`-p 8081:8081`

推理服务管理端口，可不暴露

### 1.1.3 推理接口

接口URL

[http://127.0.0.1:8080/detect\\_centernet/predict](http://127.0.0.1:8080/detect_centernet/predict)

请求方式

POST

Content-Type

form-data

请求Body参数

参数名	示例值	参 数 类 型	是 否 必 填	参数 描述
data	/data/aios/server/test_detect/test/000000000307.jpg	File	是	输入 图片 文件 名
conf_threshold	0.15	Text	否	阈值 参数

```
{
  "data": {
    "objectList": [
      {
        "targetType": "person",
        "confidence": "0.287",
        "rect": [
          280.1095886230469,
          358.2354736328125,
          339.7184143066406,
          445.73468017578125
        ]
      },
      {
        "targetType": "dog",
        "confidence": "0.154",
        "rect": [
          311.428955078125,
          221.19863891601562,
          468.3990478515625,
          358.7715148925781
        ]
      }
    ]
  }
}
```

参数名	示例值	参数类型	参数描述
results	-	Object[]	输入图片文件名
data.objectList	-	Object[]	目标列表
data.objectList.targetType	person	String	类别
data.objectList.confidence	0.287	Number	置信度
data.objectList.rect	[280.1, 358.2, 339.7, 445.7]	Number	人脸框

## 1.2 分类模型

### 1.2.1 数据准备

分类模型训练结束后，在输出目录下获取以下文件并放到一个目录下（如/models）

1) 训练导出的模型文件（文件名必须一致）

- classification-0000.params
- classification-symbol.json

2) 类别名称文件

- synset.txt

3) 模型输入信息文件

- signature.json

### 1.2.2 启动推理服务

docker启动示例

```
nvidia-docker run -d --rm -v /models:/models -p 8080:8080 -p 8081:8081  
mxnet_server:1.1.0 bash start_classification.sh
```

`-v /models:/models`

冒号前面的 `/models` 为数据准确时文件的存放路径，根据实际存放路径修改

冒号后面的 `/models` 为映射到容器内的路径，不可修改

`-p 8080:8080`

推理服务端口

`-p 8081:8081`

推理服务管理端口，可不暴露

### 1.2.3 推理接口

接口URL

<http://127.0.0.1:8080/classification/predict>

请求方式

POST

Content-Type

form-data

请求Body参数

参数名	示例值	参数类型	是否必填	参数描述
data	paper_002162.jpg	File	是	输入图片文件名
conf_threshold	0.8	Text	否	阈值参数

成功响应示例

```
{
  "results": [
    {
      "probability": 0.9804968237876892,
      "class": "paper"
    }
  ]
}
```

参数名	示例值	参数类型	参数描述
results	-	Object	结果
results.probability	0.9804968237876892	Number	置信度
results.class	paper	String	类别

1.3 Landmark模型

1.3.1 数据准备

Landmark模型训练结束后，在输出目录下获取以下文件并放到一个目录下（如/models）

1) 训练导出的模型文件（文件名必须一致）

- landmark-0000.params
- landmark-symbol.json

2) 模型输入信息文件

- signature.json

1.3.2 启动推理服务

docker启动示例

```
nvidia-docker run -d --rm -v /models:/models -p 8080:8080 -p 8081:8081
mxnet_server:1.1.0 bash start_landmark.sh
```

-v /models:/models

冒号前面的 `/models` 为数据准确时文件的存放路径，根据实际存放路径修改

冒号后面的 `/models` 为映射到容器内的路径，不可修改

`-p 8080:8080`

推理服务端口

`-p 8081:8081`

推理服务管理端口，可不暴露

### 1.3.3 推理接口

#### 接口URL

<http://127.0.0.1:8080/landmark/predict>

#### 请求方式

POST

#### Content-Type

form-data

#### 请求Body参数

参数名	示例值	参数类型	是否必填	参数描述
data	/data/aio/data/landmark/test/4821642_1.jpg	File	是	-

#### 成功响应示例

```
{
  "keypoints": [
    {
      "x": 104.41558074951172,
      "y": 57.488616943359375
    },
    {
      "x": 103.67813110351562,
      "y": 83.98870849609375
    },
    .....
    {
      "x": 110.82903289794922,
      "y": 162.0066680908203
    }
  ]
}
```

参数名	示例值	参数类型	参数描述
keypoints	-	Text	关键点
keypoints.x	104.41558074951172	Text	关键点x坐标
keypoints.y	57.488616943359375	Text	关键点y坐标

## 1.4 REID模型

### 1.4.1 数据准备

REID模型训练结束后，在输出目录下获取以下文件并放到一个目录下（如/models）

1) 训练导出的模型文件（文件名必须一致）

- reid-0000.params
- reid-symbol.json

2) 模型输入信息文件

- signature.json

### 1.4.2 启动推理服务

docker启动示例

```
nvidia-docker run -d --rm -v /models:/models -p 8080:8080 -p 8081:8081  
mxnet_server:1.1.0 bash start_reid.sh
```

`-v /models:/models`

冒号前面的 /models 为数据准确时文件的存放路径，根据实际存放路径修改

冒号后面的 /models 为映射到容器内的路径，不可修改

`-p 8080:8080`

推理服务端口

`-p 8081:8081`

推理服务管理端口，可不暴露

### 1.4.3 推理接口

接口URL

<http://127.0.0.1:8080/reid/predict>

请求方式

POST

Content-Type

form-data

请求Body参数

参数名	示例值	参数类型	是否必填	参数描述
data	/data/aios/data/reid/train/0/1204_c1s5_038341_03.jpg	File	是	-

成功响应示例

```
{
  "image_feature": [
    0,
    0.3337107002735138,
    2.9779863357543945,
    0.0026022610254585743,
    0.32378092408180237,
    .....
    0.2650204598903656,
    0.03645819053053856,
    3.923445701599121,
    1.1394240856170654
  ]
}
```

参数名	示例值	参数类型	参数描述
image_feature	{}	Text	