



# 深度学习计算服务平台 HyperDL

## 用户使用手册

### (V4.5)

文档编号: ZKHY-H4-D-YF-001-V1.3

编写日期: 2023.07.20



## 目 录

1.	引言 .....	3
2.	术语、缩略词 .....	3
3.	用户登录 .....	5
3.1.	登录平台 .....	5
4.	首页 .....	6
4.1.	文件管理 .....	7
4.2.	镜像管理 .....	8
5.	样本库 .....	11
5.1.	原始数据 .....	11
5.2.	标注项目 .....	16
5.3.	样本管理 .....	25
6.	算法库 .....	33
6.1.	预置算法 .....	33
6.2.	用户算法 .....	37
6.3.	代码管理 .....	38
7.	训练平台 .....	40
7.1.	Notebook .....	40
7.2.	预置算法训练 .....	44
7.3.	用户算法训练 .....	45
7.4.	超参调优任务 .....	46
8.	模型库 .....	47
8.1.	模型管理 .....	47
8.2.	预测任务 .....	49
8.3.	评估任务 .....	51
8.4.	评估引擎 .....	53
8.5.	转换任务 .....	54
8.6.	转换工具 .....	56
9.	推理平台 .....	57
9.1.	云端服务 .....	57
9.2.	运行环境 .....	58
9.3.	边缘服务 .....	58
9.4.	边缘设备 .....	59
10.	流程化教程 .....	60



## 1. 引言

HyperDL-V4.5 是中科弘云科技有限公司推出的训推一体化人工智能计算服务平台，提供涵盖数据标注、模型开发、模型训练、模型部署、推理服务的一站式 AI 模型开发及推理服务解决方案，以“可视化”操作和“自动化”的流程管理让用户“零代码”即可快速上线 AI 应用。

本文档针对平台使用提供模块解析、流程化使用教程，使用户能根据教程快速上手使用平台。

查看流程化使用教程请直接跳转至 [10. 流程化教程](#)；

注意事项：

- 平台推荐使用浏览器 chrome（100 版本 以上）、Edge（110.0 版本 以上）、safari（15.1 版本以上）；推荐分辨率为 1280\*720 或者更高；
- 在平台进行内部页面跳转时浏览器可能会进行拦截，需要进行授权后才能进行跳转使用；

## 2. 术语、缩略词

针对于平台上的专业术语和缩略词进行说明：

名称	说明
K8s	Kubernetes（通常简称为 K8s）是一个开源的容器编排平台，用于自动化部署、扩展和管理容器化应用程序。Kubernetes 的目标是简化容器化应用程序的部署、管理和维护，使开发者能够更加专注于应用程序的逻辑而不是基础架构的细节。官方网址是： <a href="https://kubernetes.io/">https://kubernetes.io/</a>
Docker	Docker 是一个开源的容器化平台，允许开发者将应用程序、服务以及它们的依赖项打包到一个独立的、可移植的容器中。这些容器包含了一切运行所需的代码、运行时、系统工具、系统库等，使得应用程序能够在任何环境中快速部署、运行和迁移。Docker 的容器化技术解决了应用程序在不同环境中部署和运行的一致性问题，简化了开发、测试和生产环境的管理。官方网址是： <a href="https://www.docker.com/">https://www.docker.com/</a>
镜像/容器	镜像是一个预先配置好的、可执行的软件包，它包含了运行应用程序所需的所有文件、代码、库和依赖项。镜像可以看作是一个容器的静态模板，可以用来创建容器实例。Docker 是容器化技术的一种实现，它使用镜像作为容器的基础。通过 Docker 镜像，开发者可以快速部署、测试和运行应用程序，而无需担心底层系统

	和依赖项的差异。镜像是 Docker 容器化技术的核心概念，它使得容器在不同环境中具有相同的行为，为应用程序的开发和部署提供了便利和一致性。
集群	服务器集群是由多台独立的服务器组成的网络系统，它们协同工作以提供更高的性能、可靠性和可扩展性。
角色	角色是平台内置功能，通过创建不同的角色分配不同的菜单权限；
组织	组织是平台内置功能，通过创建不同的组织分配不同的资源、创建自己组织下的用户；
用户	用户是平台内置在组织下功能，用户下用来新建用户分配权限；
Notebook Jupyter	“Notebook” 是一个交互式的计算环境，其中可以结合代码、文本、图像和其他富媒体元素来创建和共享文档。Notebook 的最大特点是可以在单个界面中实现编程和文档撰写，使得数据分析、可视化和实验变得更加方便和易于理解。最常见的 Notebook 应用是 Jupyter Notebook，它最初是 IPython 项目的一部分，现在已经成为一种广泛使用的开源工具。
原始数据	原始数据是未进行标注的数据。
样本数据	样本数据是进行标注后的数据，由原始数据和标注结果组成。
样本库	样本库是平台功能模块；包含原始数据、标注项目、样本管理三个功能。分别提供对数据的上传、标注、清洗以及纳管等功能
算法库	算法库是平台功能模块；包含预置算法、用户算法、代码管理三个功能。对算法和代码进行管理、集成；
模型库	模型库是平台功能模块；包含模型管理、预测任务、评估任务、评估引擎、转换任务、转换工具六个功能；对算法训练完成的模型进行预测、评估、转换以及发布在线服务等操作；
模型转换	模型转换是把模型从一种表示形式转换为另一种表示形式的过程。平台目前内置了 air 转 om、onnx 转 trt 等工具；
资源类型	资源类型指的是加速卡的芯片类型，平台目前支持 GPU、NPU、MLU、DCU、CPU 等多种加速卡芯片类型
计算框架	计算框架是一种用于进行大规模数据处理和计算的软件工具或库。平台目前支持 TensorFlow、PyTorch、MindSpore、PaddlePalle、OpenMPI、SKLearn、XGBoost；
超参调优	超参调优用于选择最优的超参数组合，以提高模型的性能和泛化能力。超参是模型训练过程中需要人为指定的参数，例如学习率、批大小、迭代次数、正则化参数等，它们对模型的性能和训练过程有重要影响。平台目前内置了 GPU、NPU 的图像分类算法；
云端服务	云端服务是一种通过互联网提供计算资源、存储空间、应用程序和其他服务的模式。针对平台内置算法平台提供了 Triton、



	MIndSpore 以及自定义服务提供使用；
边缘服务	边缘服务是一种将计算、存储和网络服务移动到物理设备接近数据源或用户的边缘位置的计算模式。针对平台内置算法平台提供了 Triton、MindSpore 以及自定义服务提供使用；

### 3. 用户登录

使用平台前，需要开通用户账号，向系统管理员申请用户账号及密码。

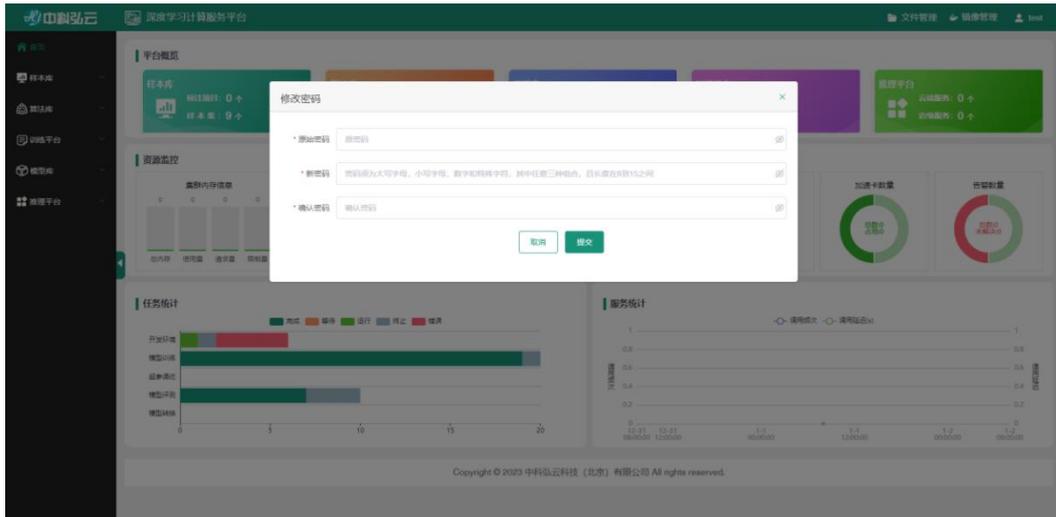
#### 3.1. 登录平台

平台登录地址：<https://platformip:port>（platform 是平台服务的 IP 地址，port 是服务端口，具体地址请联系系统管理员）。

➤ 登录界面如下图所示：



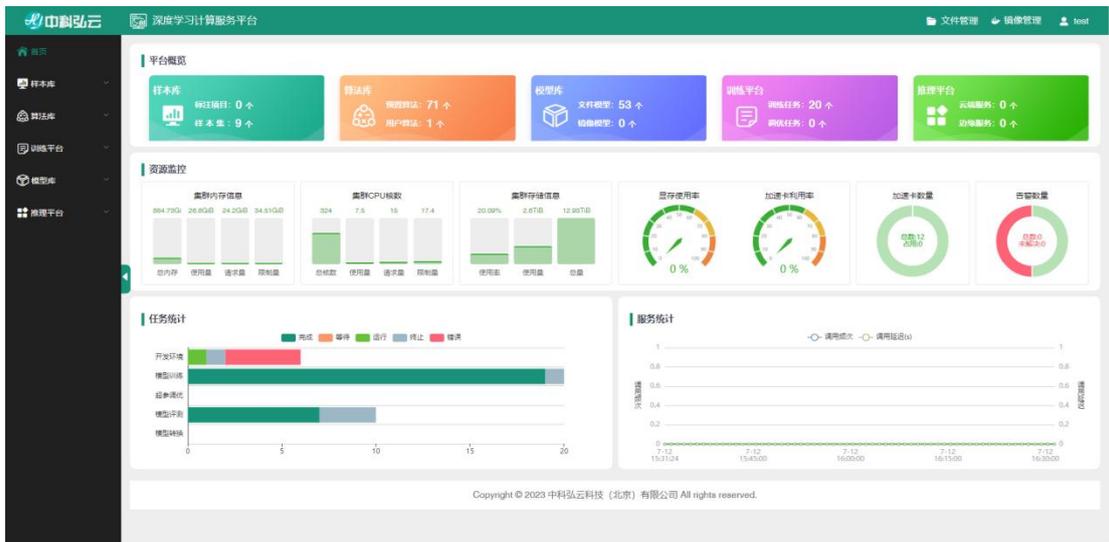
➤ 在异构资源管理平台创建完成用户后，登录深度学习服务平台。首次登录会提示修改密码，按照提示进行修改，重新登录即可。温馨提示：没有特殊说明，新建用户的缺省密码为 Hyper123!，设置密码的长度为 8 位字符以上，由大小写字母、数字或特殊符号三种组成以上。



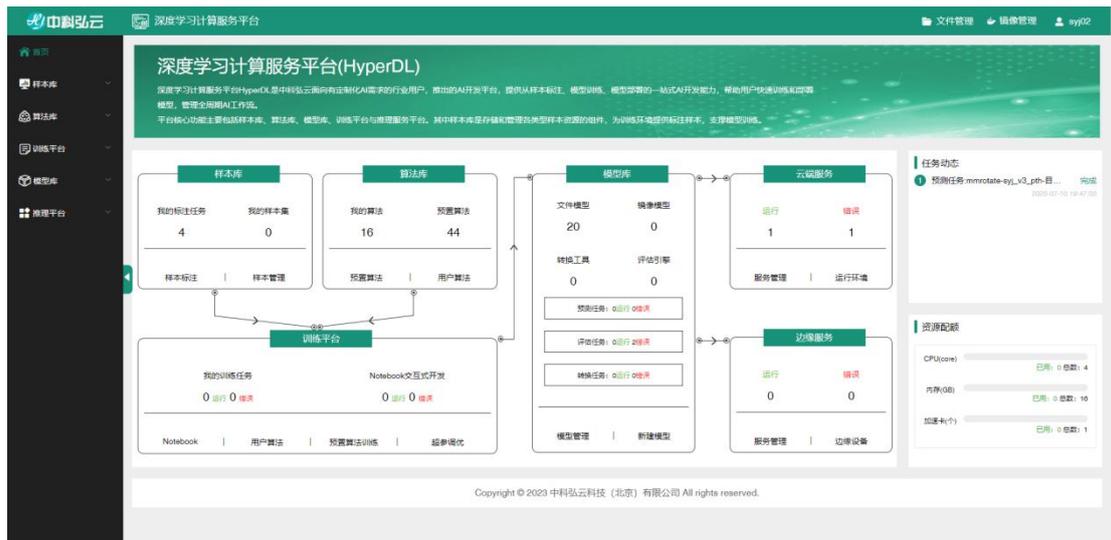
#### 4. 首页

成功登录到平台后能看到平台的首页；深度学习服务平台分为管理员和普通用户角色，不同角色用户所呈现首页是不同。如下图。

- 管理员首页：管理员首页展现平台数字资产统计信息，算力资源使用、以及计算任务统计信息。



- 普通用户首页：展现平台业务流程，以及用户创建或可以使用数字资产或任务信息。

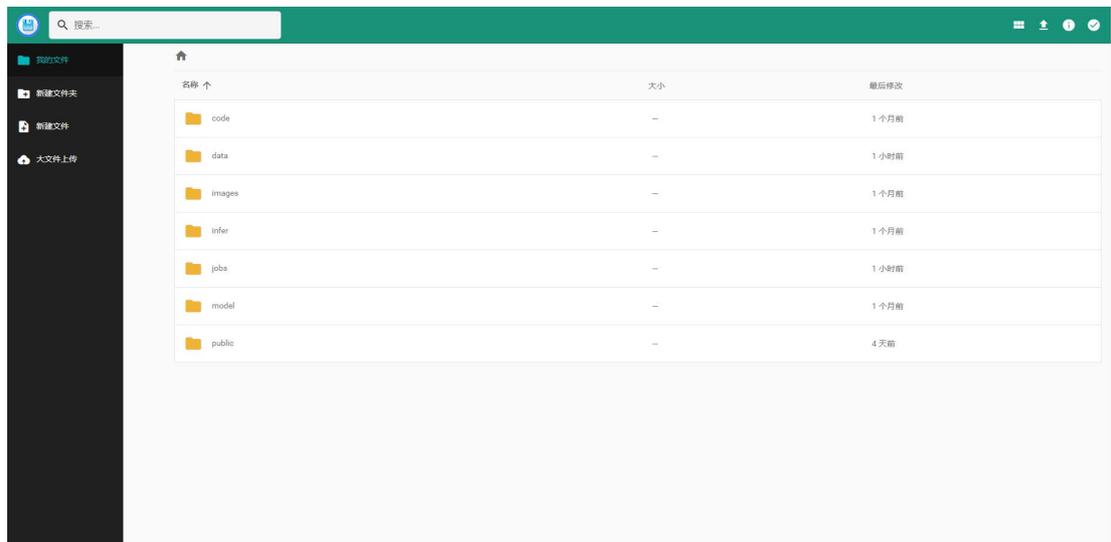


区别在于管理员对平台进行整体把控，能看到是平台整体的使用情况以及所有用户的任务。而普通用户只是能看到自己用户的使用情况以及自己创建的任务；

#### 4.1. 文件管理

文件管理提供新建文件、文件上传、文件下载、文件浏览、文件编辑等功能，大文件上传支持断点续传。对于大文件(如大于 500MB)，推荐使用大文件上传进行传输。

点击文件管理会跳转到个人工作目录下，平台为用户分配个人工作目录，提供使用；每个用户只能访问自己的工作目录，无权访问其他用户目录。



用户工作目录结构如下：

**code:** 用于存放用户的代码等相关文件；

**data:** 用于存放用户的样本数据；

- images:** 用于存放用户的镜像文件；
- infer:** 用于存放用户预测评测相关的数据；
- jobs:** 用于存放用户的作业任务（数据、训练、预测、评估、转换、服务）；
- model:** 用于存放用户的训练发布模型；
- public:** 关联平台提供的公开样例数据，只读权限，无权创建、编辑、和删除文件；

**提示：**以上平台内置目录请勿删除。

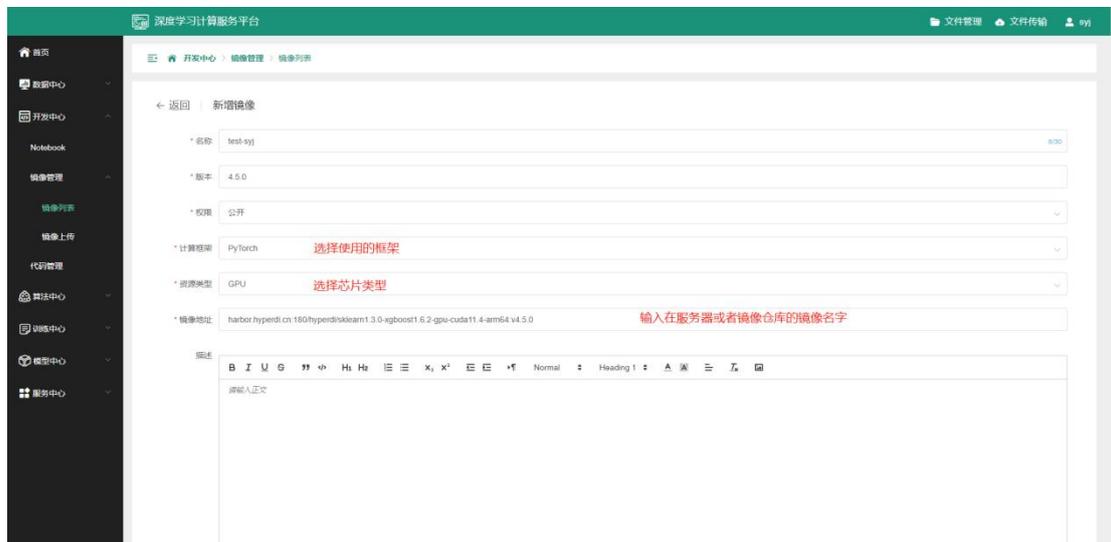
## 4.2. 镜像管理

镜像管理对平台内使用的镜像进行纳管，同时也支持用户上传自定义的镜像到平台上进行使用。

### 4.2.1. 新增镜像

新增镜像：在平台添加服务器上已有的镜像；后台 load 完成的镜像或是上传到 Harbor 仓库的镜像；

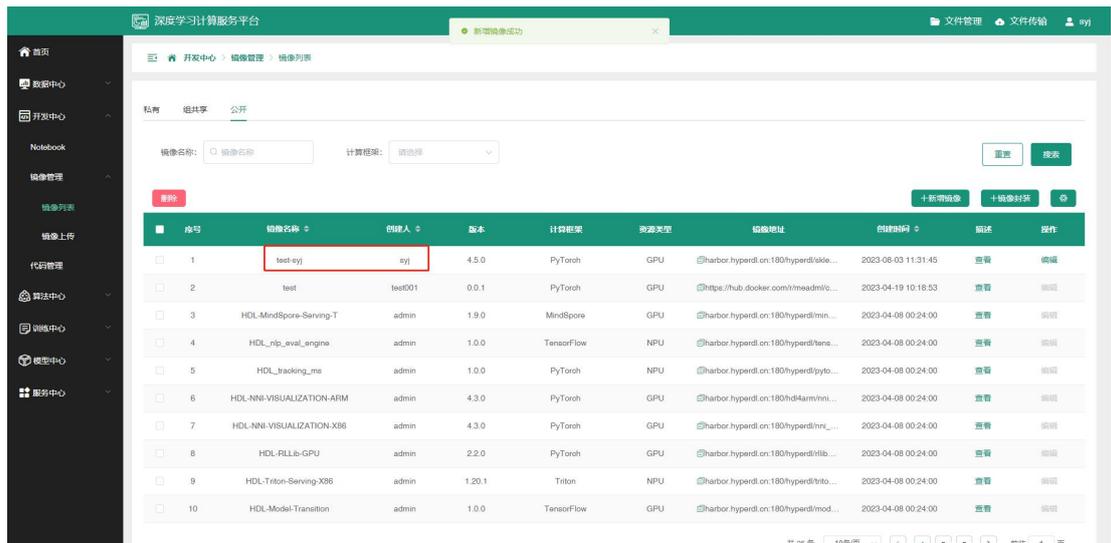
- 点击镜像管理→镜像列表→+新增镜像，进行镜像添加；输入名称、版本号、权限、选择框架、芯片类型、镜像地址进行添加；



**提示：**资源类型为镜像支持运行的芯片类型。如镜像需要运行在 GPU 卡上，则需选择 GPU，如果不需要加速卡，则选择 CPU。

镜像地址栏中填写镜像在 Harbor 镜像仓库中地址，平台缺省 Harbor 仓库地址为 harbor.hyperdl.cn:180。

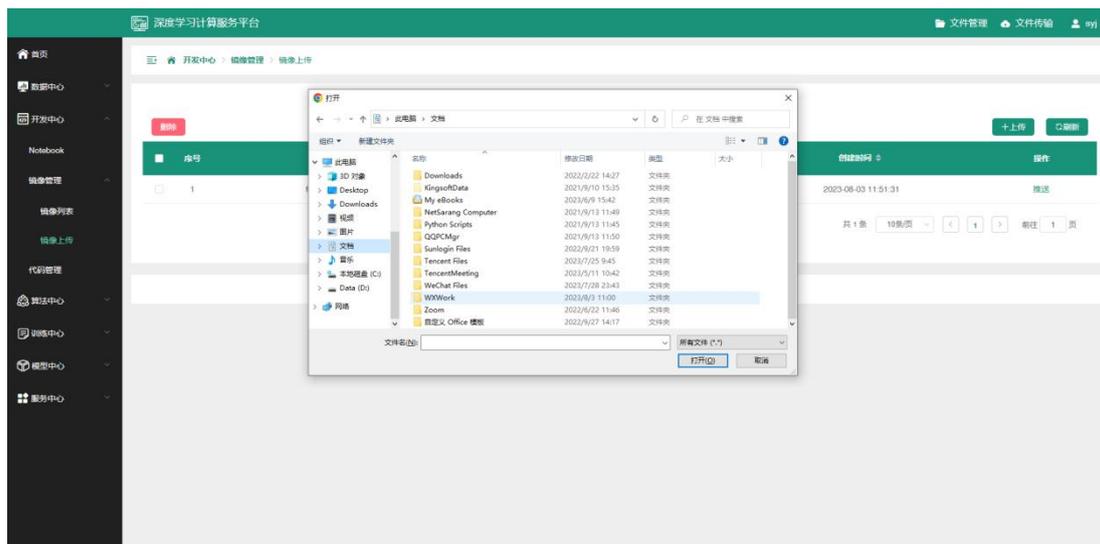
- 添加完成跳转到页面进行查看；



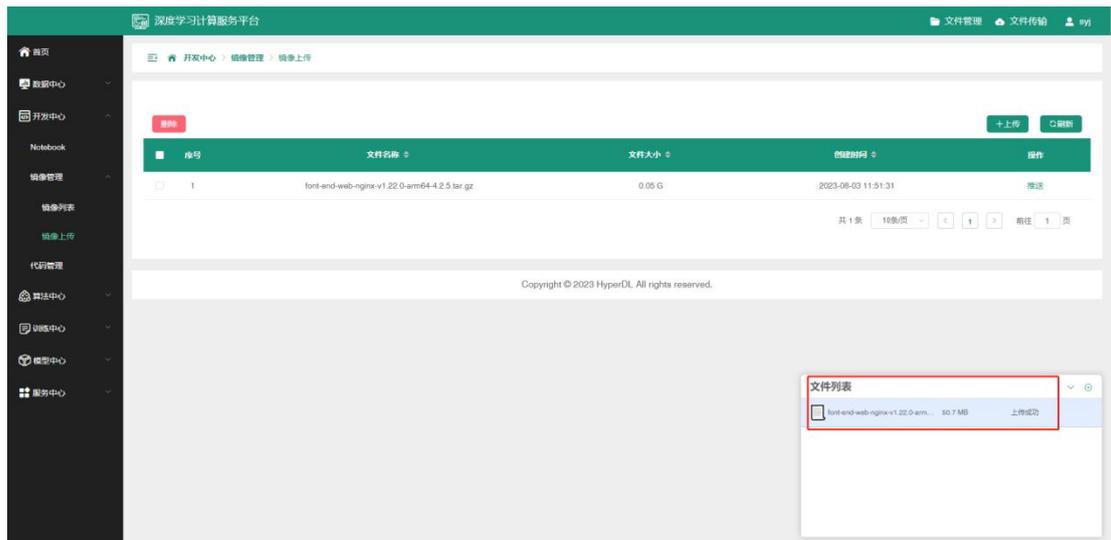
## 4.2.2. 镜像上传

镜像上传：上传打包成 tar、img 的镜像并推送到 HarBor 仓库；

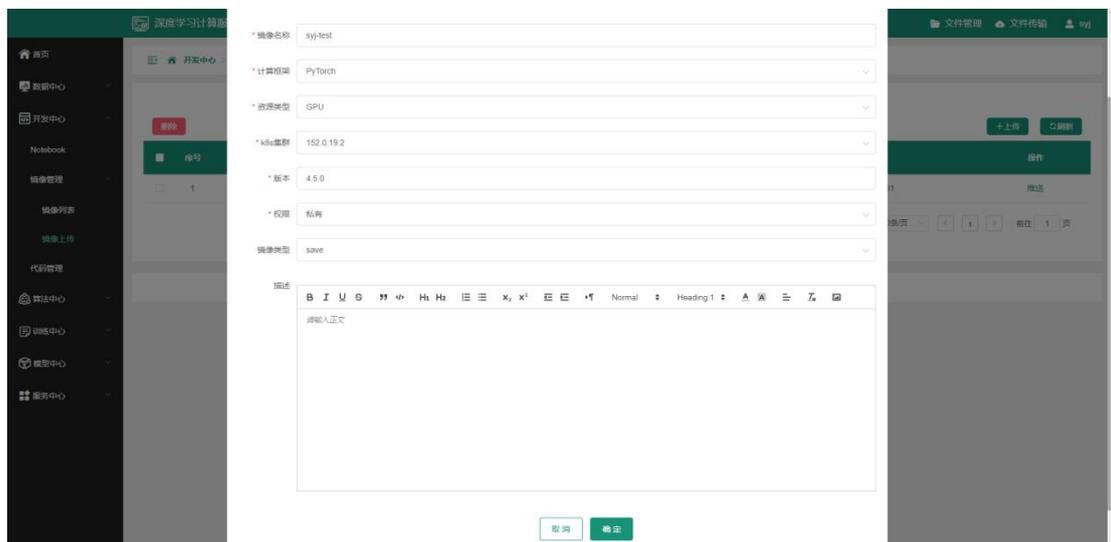
- 点击镜像上传-->+上传选择需要上传的镜像进行上传；



- 上传成功会在右下角文件列表进行展示；

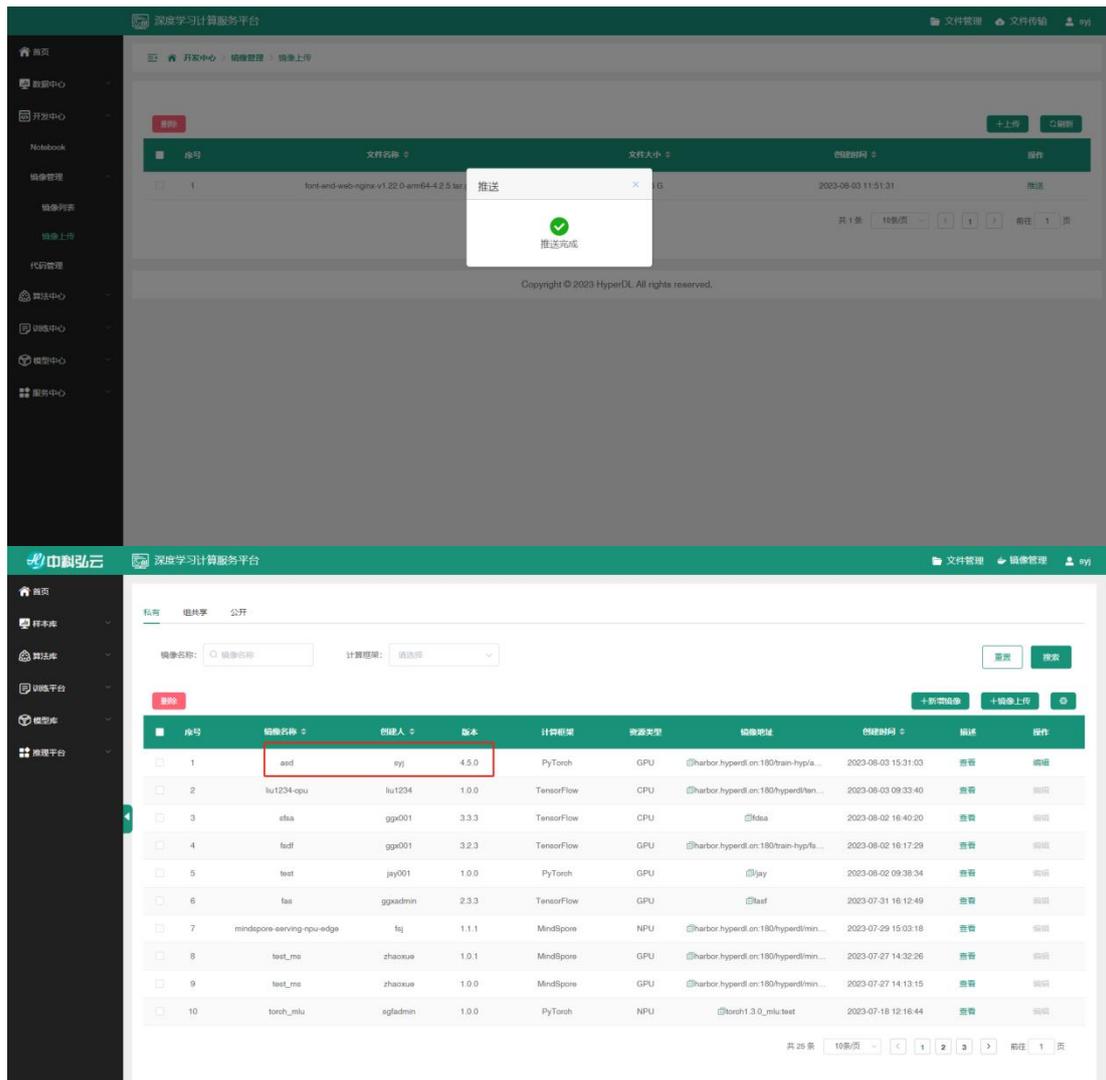


- 在操作栏点击推送，填写信息后进行推送；



镜像类型分为 save 类型和 export 类型，取决于镜像保存的方式。当镜像是以 docker save 方式保存时选择 save，以 docker export 方式保存时选择 export。缺省情况下以 save 方式。

- 等待推送完成在镜像列表进行查看



## 5. 样本库

样本库提供原始数据、样本标注、样本管理功能，有原始数据管理、标注项目、样本管理三个模块，分别提供对数据上传、清洗，样本标注，样本增强、以及样本纳管等功能；

**原始数据：**将原始数据上传至平台，清洗处理，提供给后续标注使用；

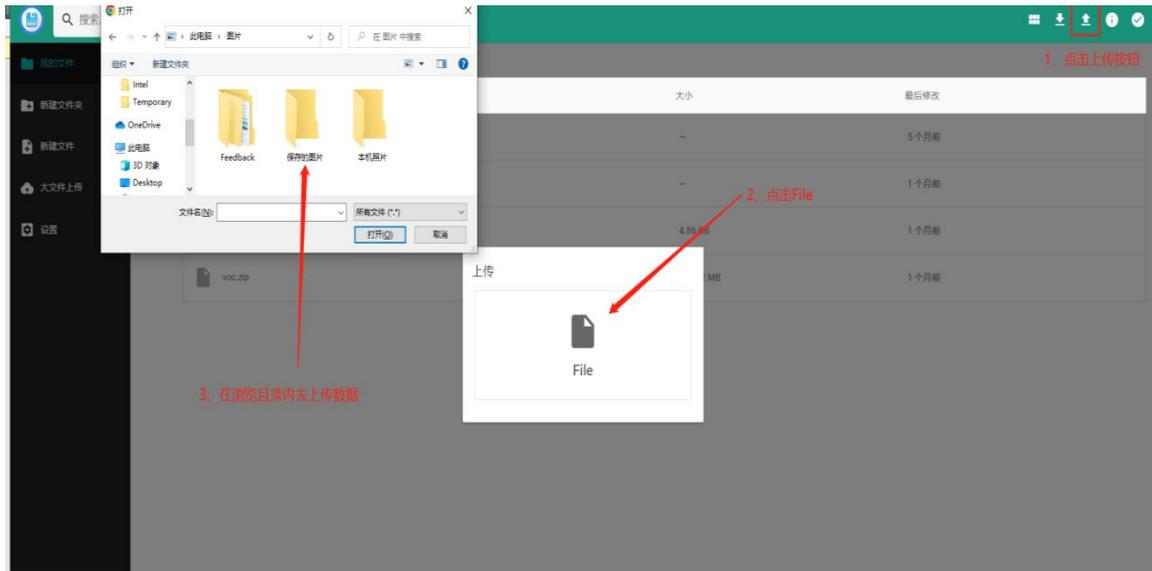
**标注项目：**对原始数据进行标注；

**样本管理：**对标注完成或外部导入数据进行纳管，最终提供给算法训练使用；

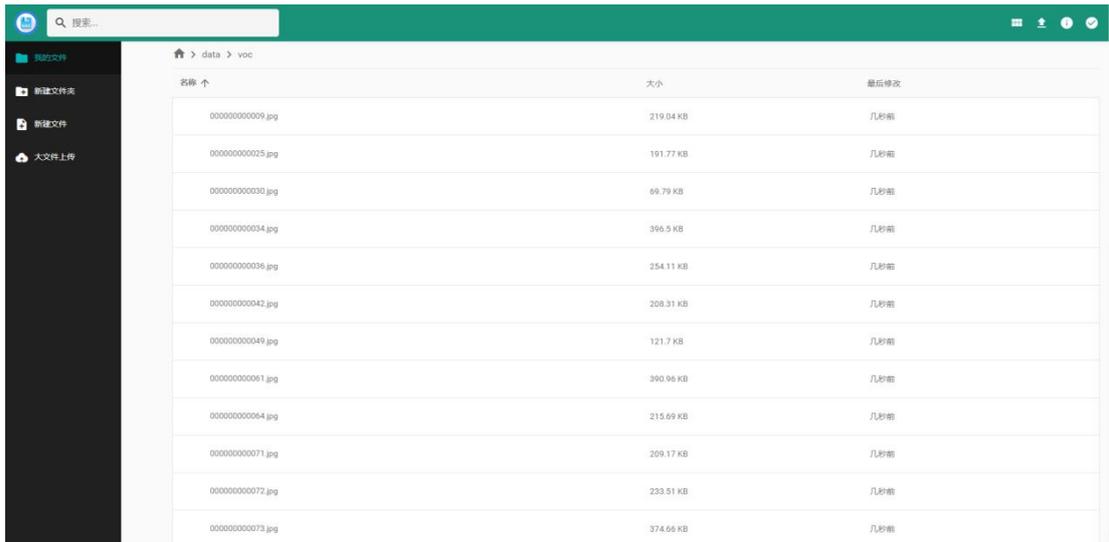
### 5.1. 原始数据

原始数据是对用户的真实数据进行上传并纳管，方便后续在平台上进行标注操作；在进行标注之前需要对原始数据进行上传；

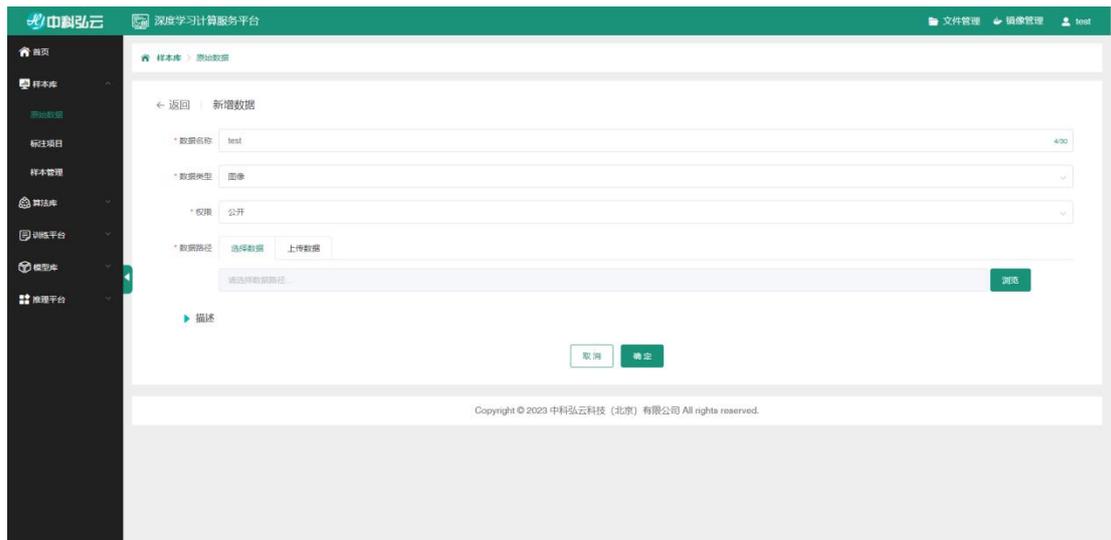
- 点击文件管理-->data 目录下上传自己的数据



- 上传完成可查看上传的数据;

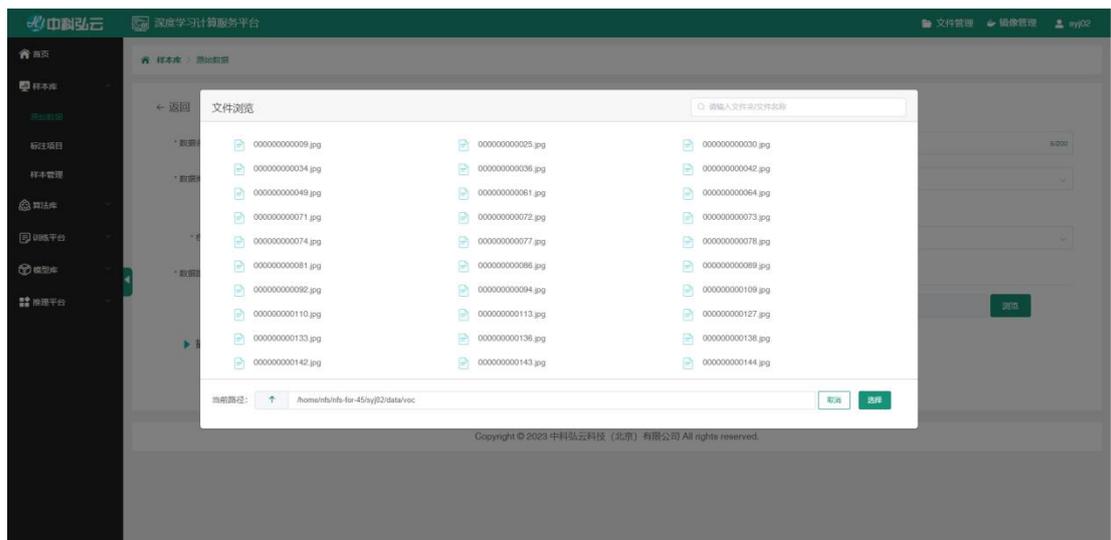


- 点击样本库-->原始数据--> +数据，进行原始数据的上传;

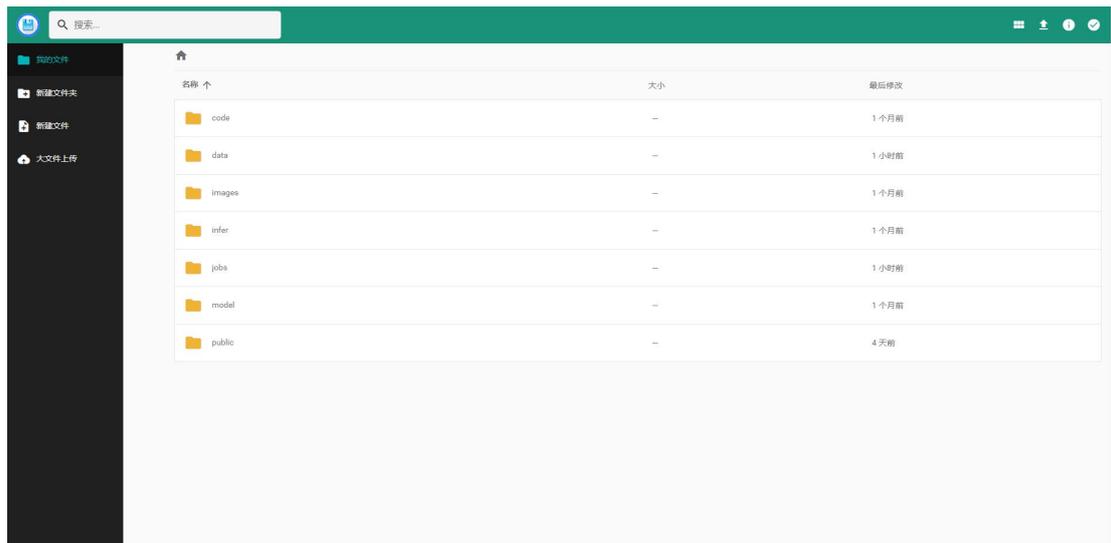


在数据路径中有两个选择：

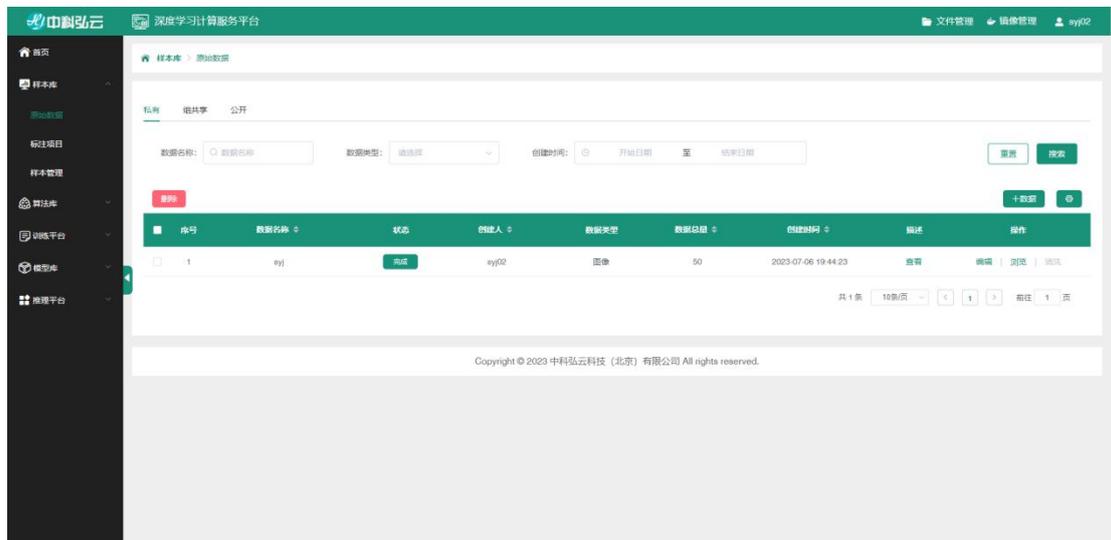
- 1、选择数据从个人目录下导入数据（可以浏览通过文件管理上传的数据进行导入）；



- 2、上传数据是会跳转到文件管理页面，然后可以从本地上传数据到个人目录下（相当于是跳转到最开始的文件管理进行数据上传使用）；



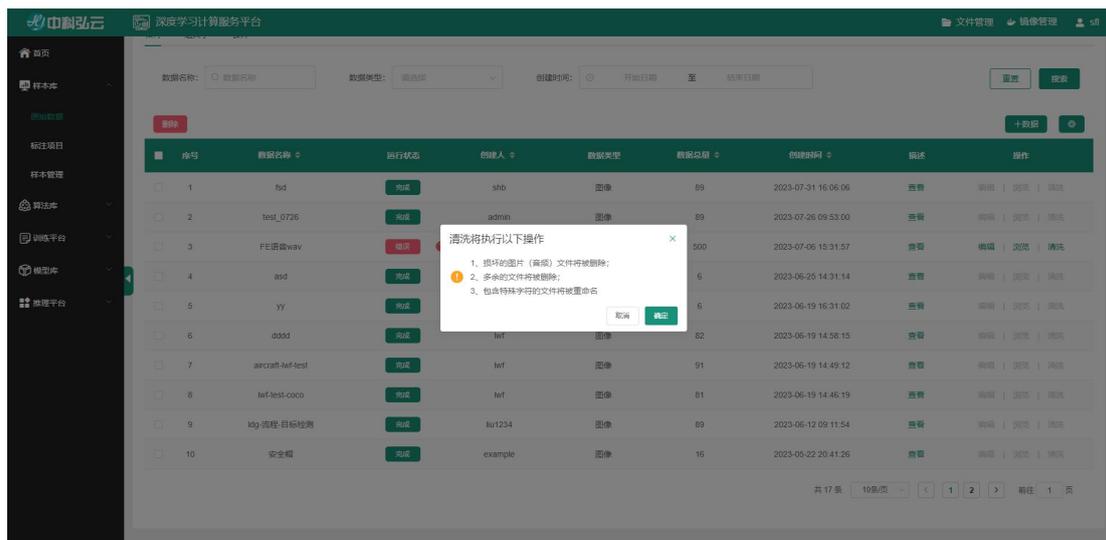
➤ 上传完成后的页面



通过操作栏可以对完成的任务进行操作。

➤ 点击编辑可以修改任务名称，权限：



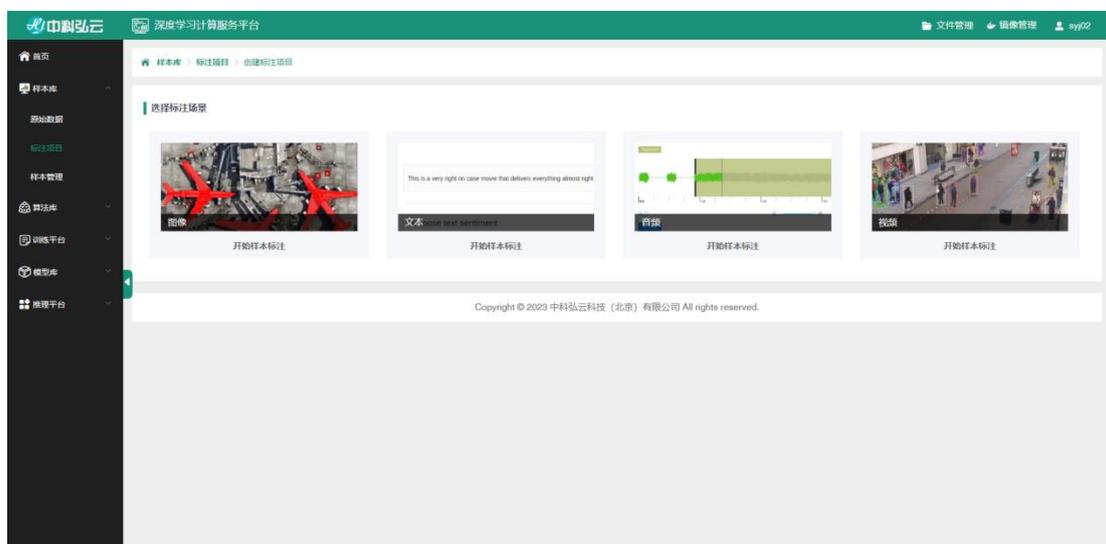


清洗操作说明：1) 损坏的文件是指平台的工具无法打开或读取；2) 多余文件是与数据类型不相符的文件，例如是图像数据集，合规的格式为 png, jpg, jpeg 格式文件，其他文件为非合规；3) 文件名中包含特殊字符，如空格，\*，! 等，重新命名。

## 5.2. 标注项目

添加完成原始数据后就可以对添加的原始数据进行标注；

- 点击标注项目--> +项目标注选择要标注的场景进行 workflow 选择；



平台支持图像、文本、音频、视频四种数据类型的标注。图像支持的数据格式类型包括 png, jpg, jpeg, 文本支持的文件格式类型为 txt, csv, 音频文件格式为 mp3, wav, 视频文件格式为 mp4。文件后缀不区分大小写。

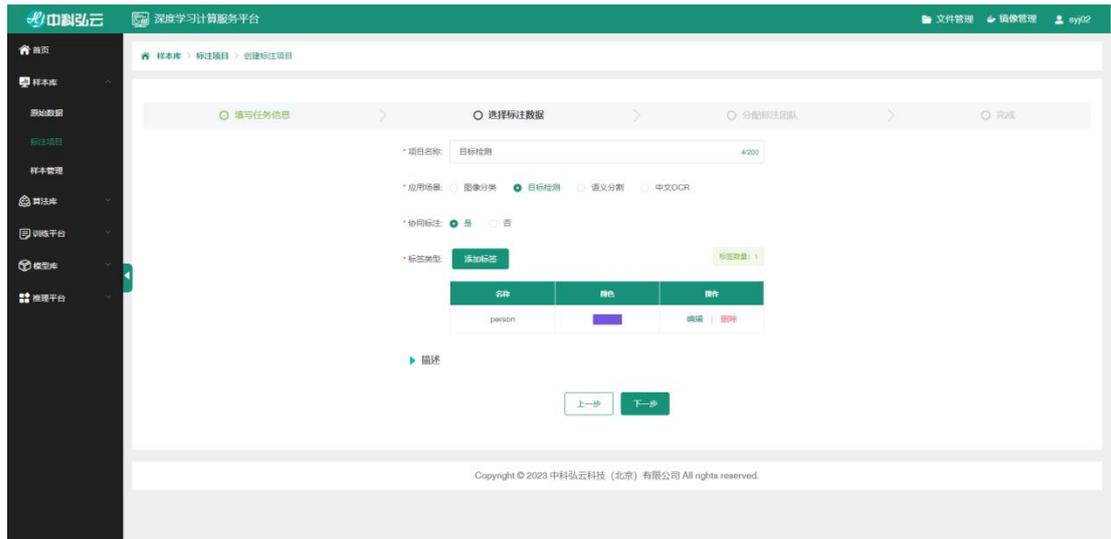
标注任务以目标检测场景为示例进行操作；

**应用场景：** 目标检测；

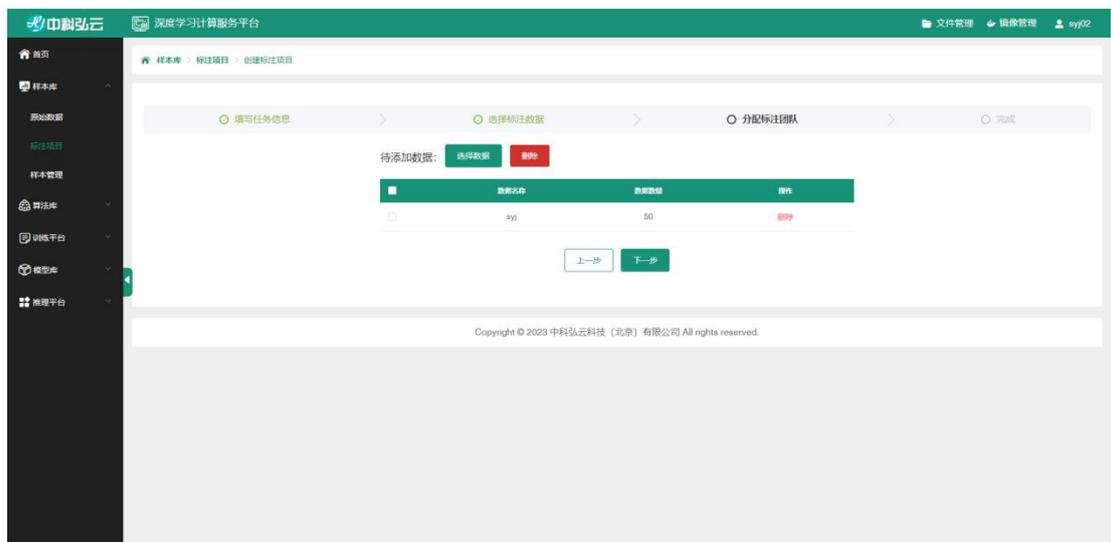
**协同标注：** 是否选择多人一起标注这份数据；选择是就是多人，否只能自己用户进行标注；

**添加标签：** 标签支持手动一个个新增；也支持外部导入进行批量导入；（导入规则文本文件内一行一个类别）

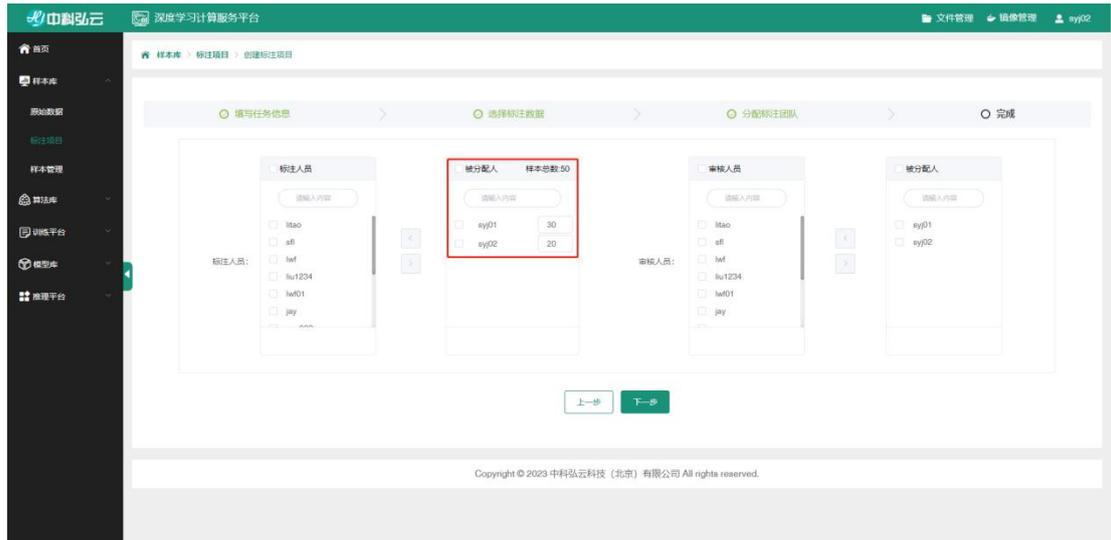
- 填写项目名称、选择标注的场景、是否协同标注、添加标签进行下一步；



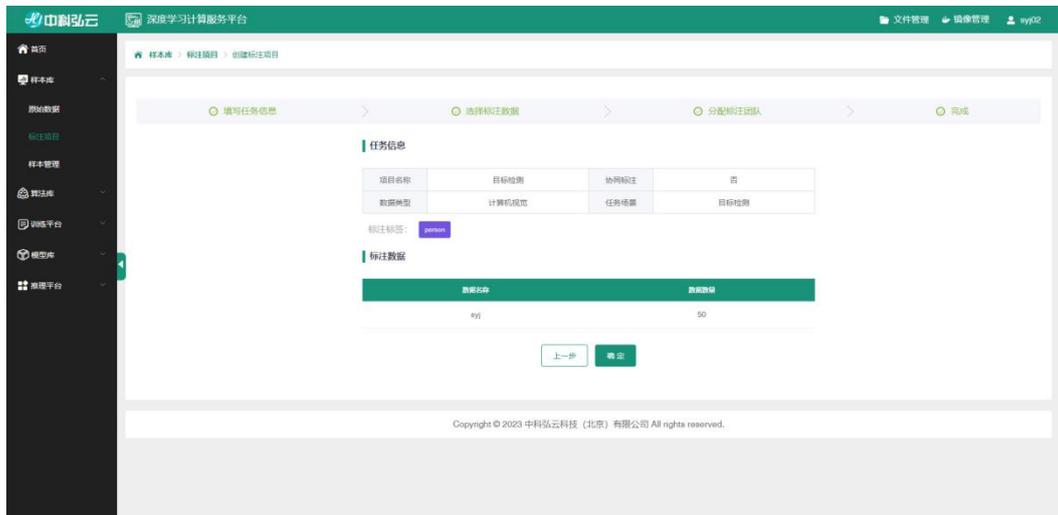
- 选择刚添加的数据集进行下一步操作；



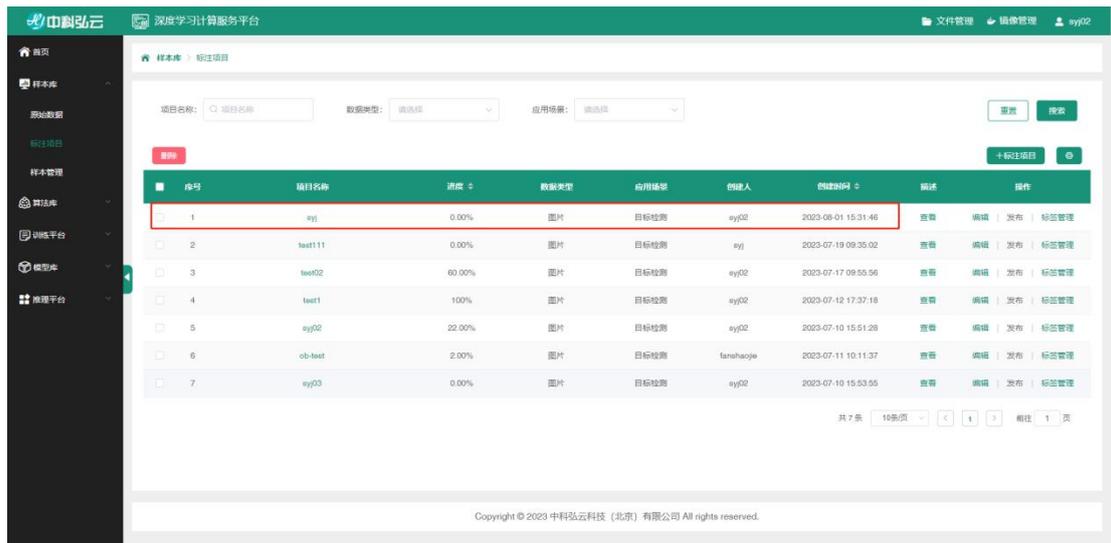
- 点击下一步进行标注团队分配，选择标注人员分配标注数量，选择审核人员进行审核；



- 最后进行确认没有问题点击确定创建完成标注项目；

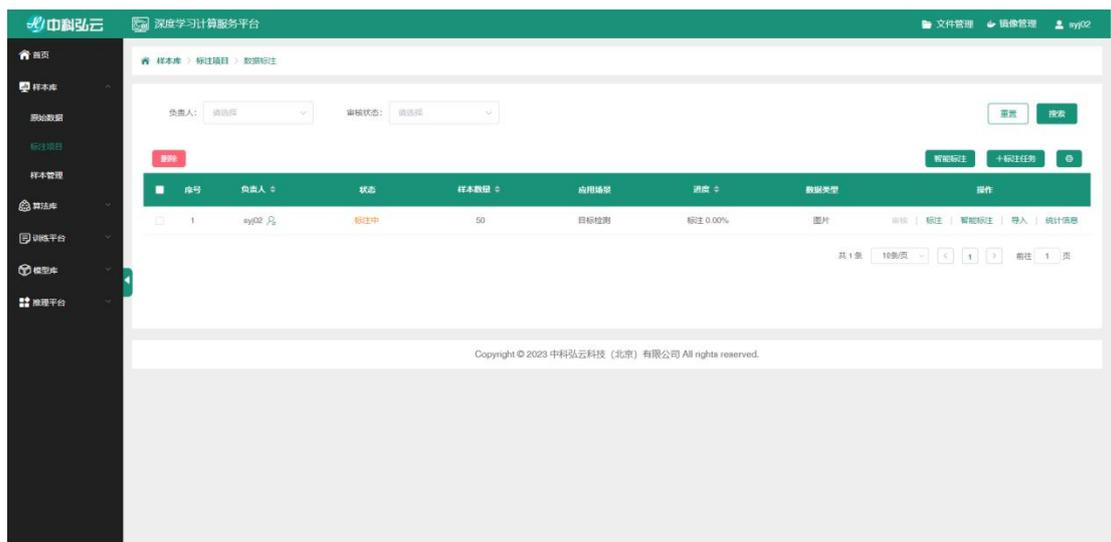


- 成功以后跳转到标注项目查看创建好的项目；

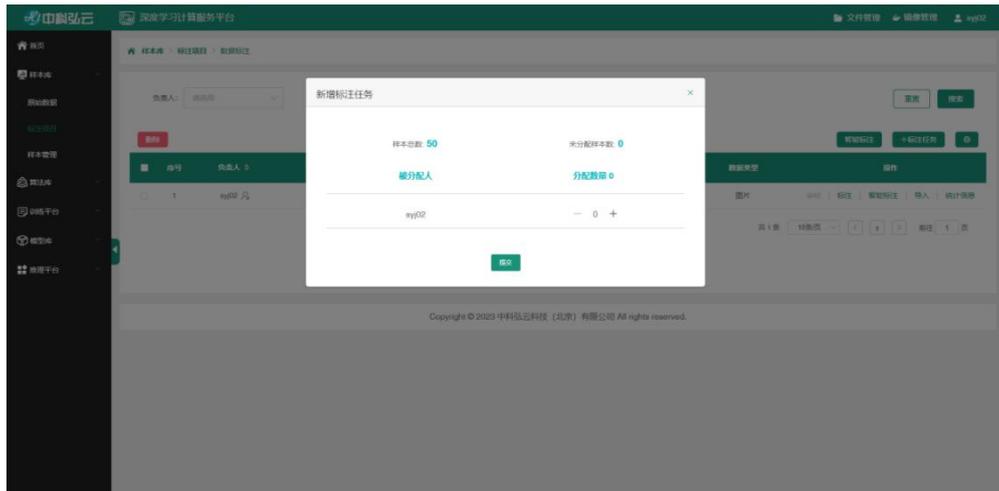


### 5.2.1. 数据标注

➤ 点击项目名称，进入到数据标注操作页面；

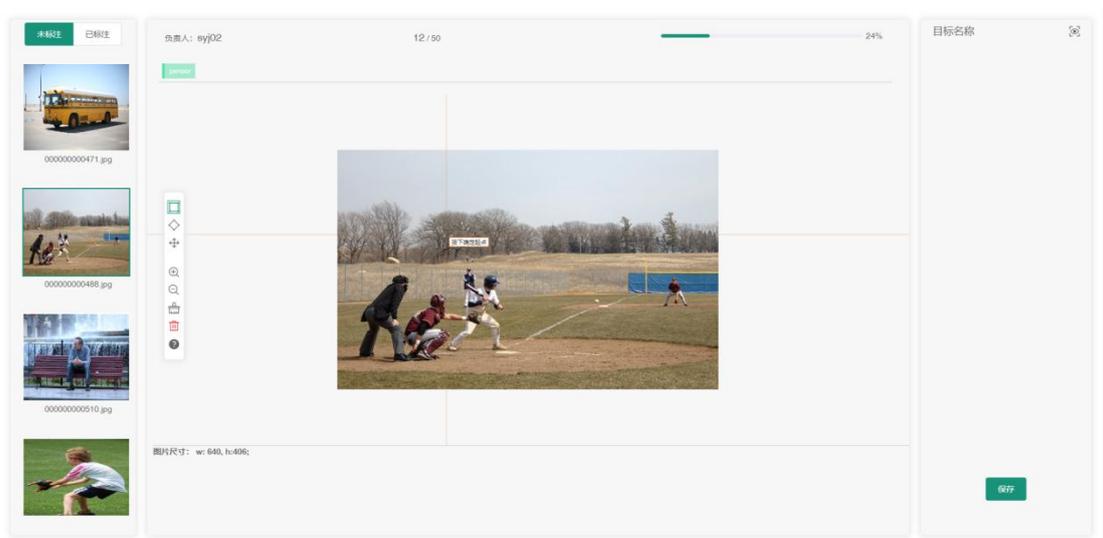


➤ 点击+标注任务，可以对协同标注人员分配任务；



- 分配完成以后点击标注-->未标注里选择要标注的图像，即可跳转到图像的标注页面可以开始对图像进行标注；

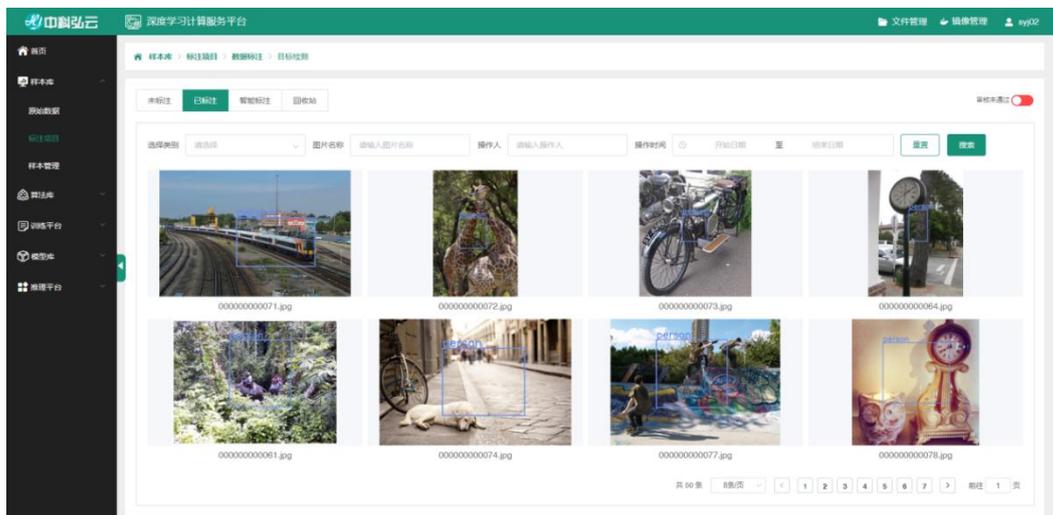
提示： 如果点击标注未跳转到新的页签，是由于浏览器可能会进行拦截，需要进行授权后才能进行跳转使用；



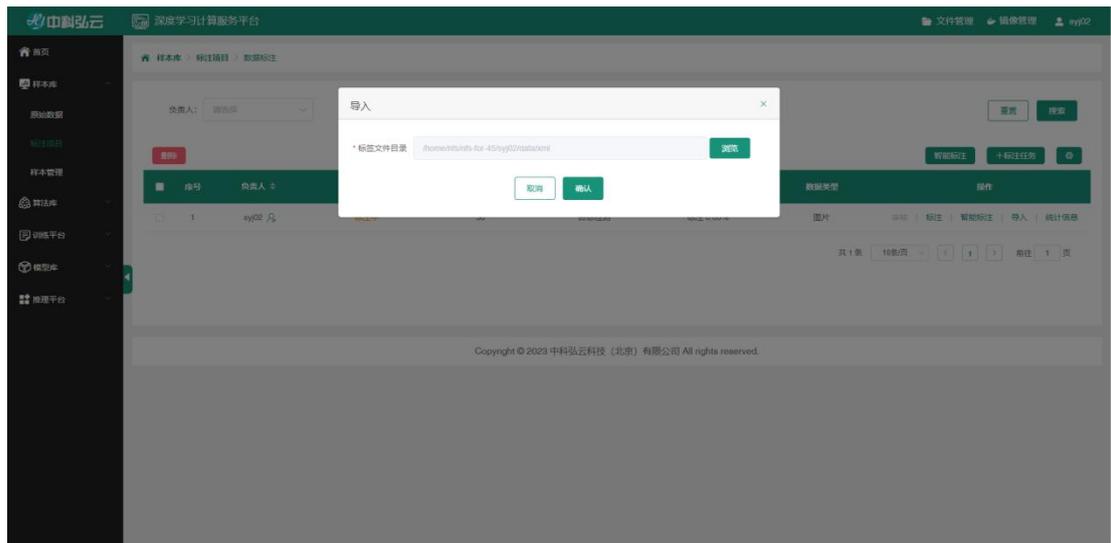
- 点击上方的标签-->然后开始进行标注-->标注完成点击保存完成标注；



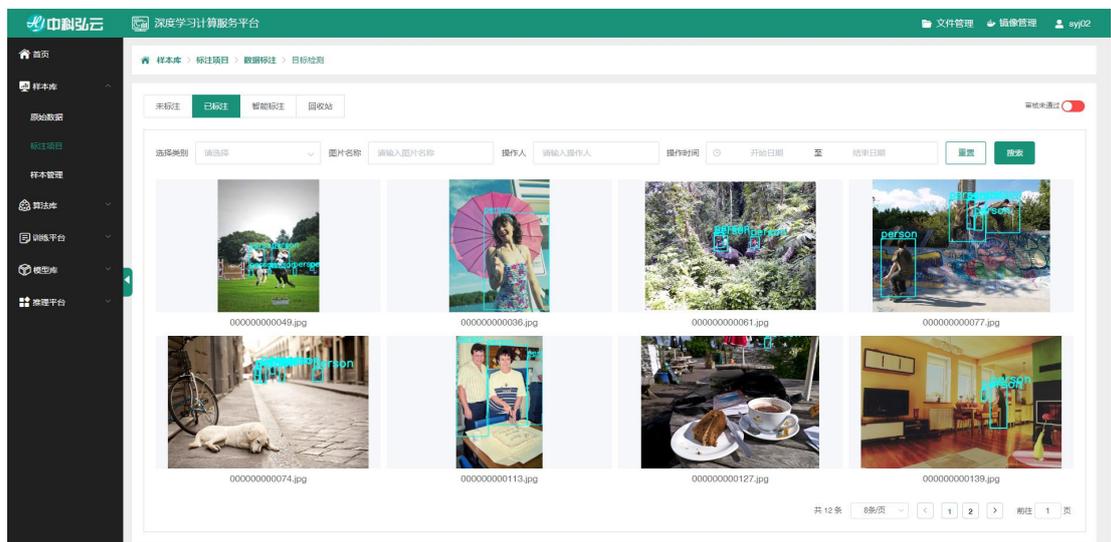
➤ 标注完成可以看到未标注里是没有图像的，在已标注里查看标注完成的图像；



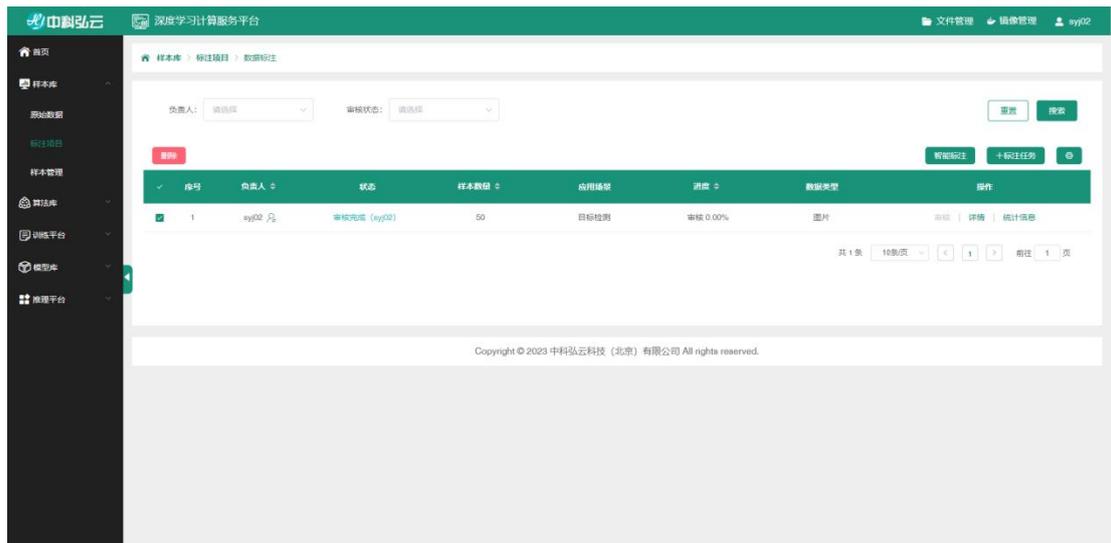
➤ 同时平台也支持外部导入数据（比如一共 50 张图像，用户有 15 张是已经标注完成的就可以导入到平台，就可以看到已标注完成 15 张还剩余 35 张未标注），详细操作为点击导入选择已经标注完成的部分标注文件，可以载入；



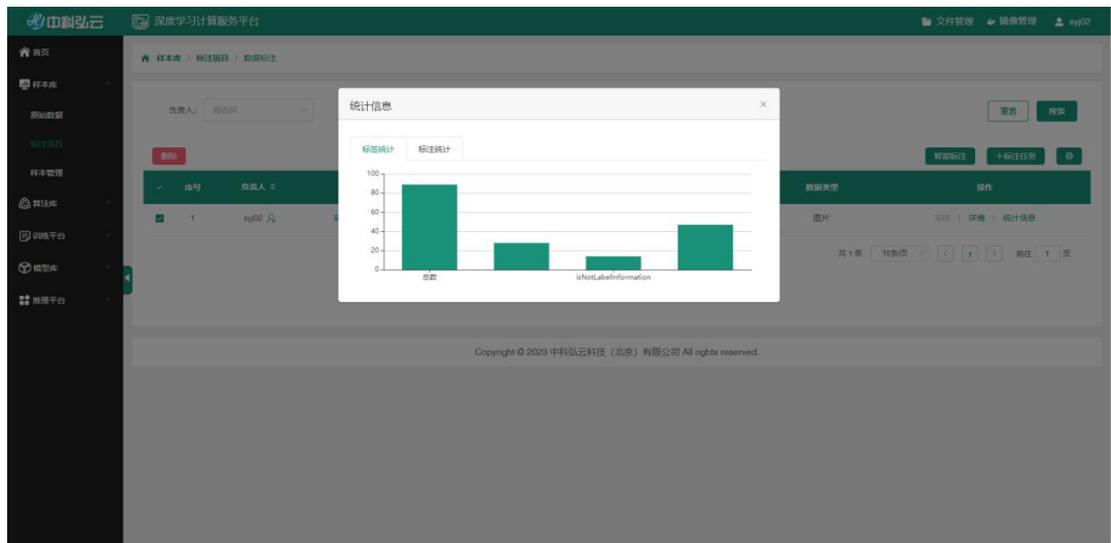
➤ 导入完成点击标注→在已标注里可以看到我们导入的标注框；



➤ 标注完成后既可以进行审核，审核完成这份数据就标注完成了，后续可以对标注完成的数据进行纳管和训练；



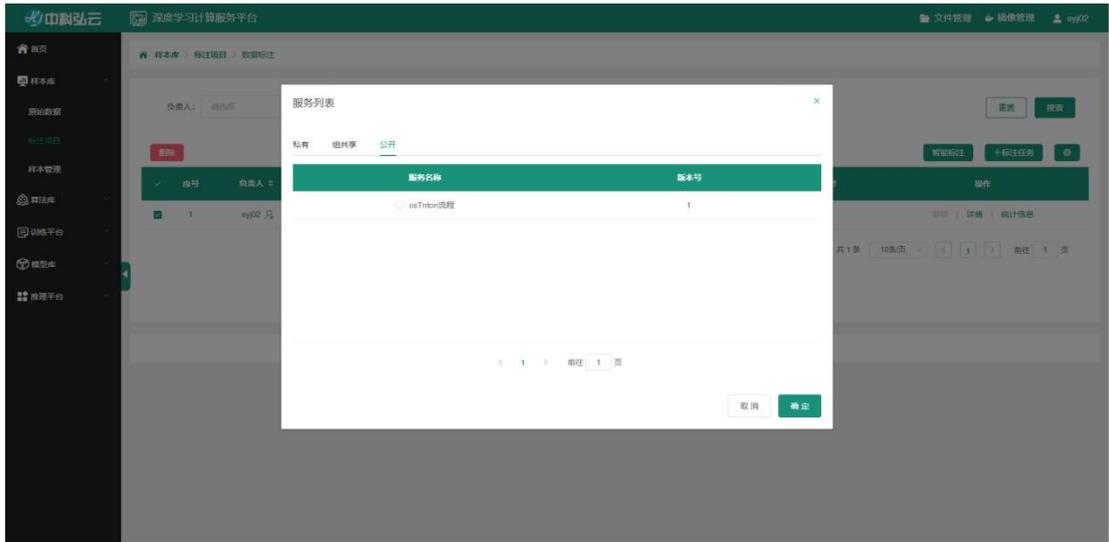
- 点击统计信息能看到标注完成的标签统计和标注统计详情；



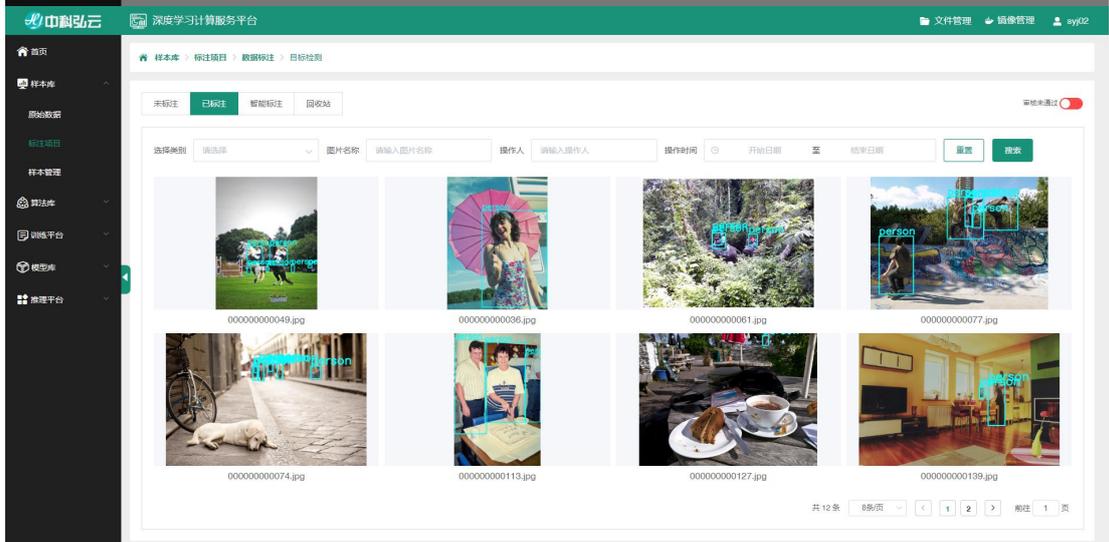
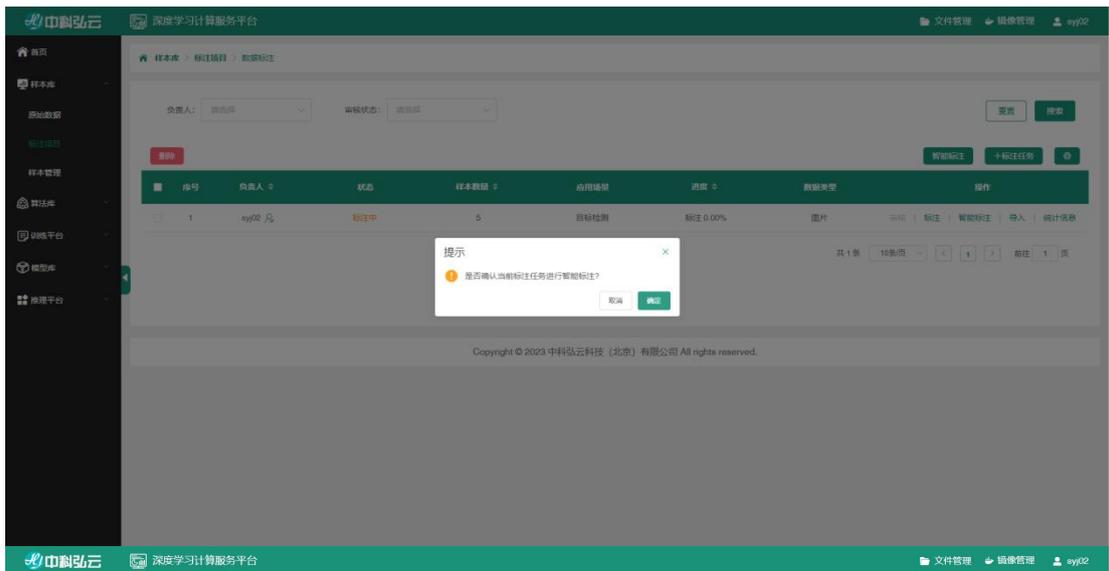
## 5.2.2. 智能标注

平台针对目标检测场景内置了智能标注功能，对于平台内置算法训练的模型发布服务后，可以进行智能标注，从而形成闭环；

- 点击项目标注-->智能标注，选择已经发布起来的服务（服务发布详情见 [8. 推理平台](#)）；



- 选择好要调用的服务后，点击操作栏里的智能标注，等待智能标注完成后，在项目已完成里可以查看智能标注后的图像标签，并进行修改；

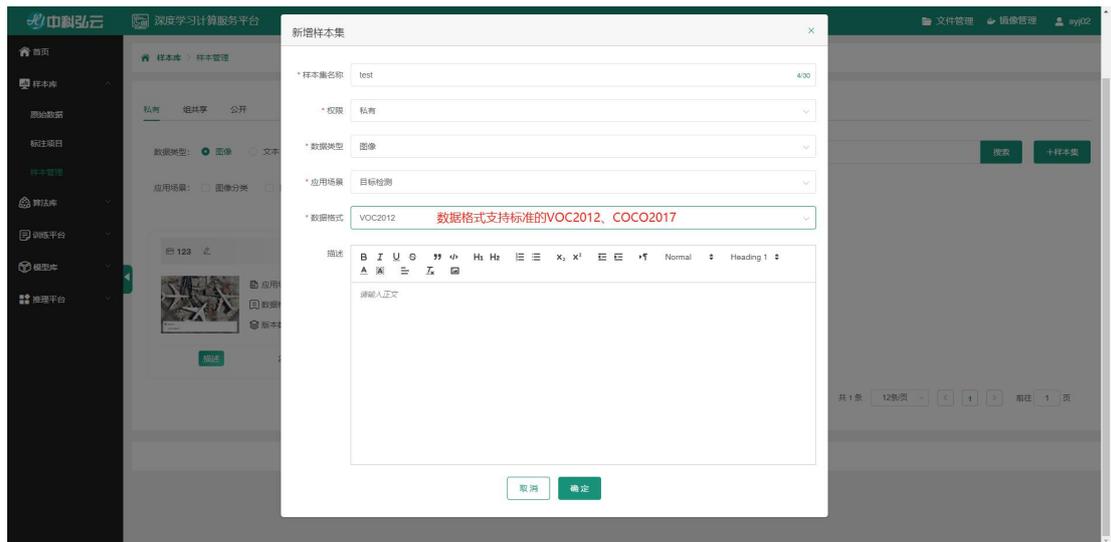


### 5.3. 样本管理

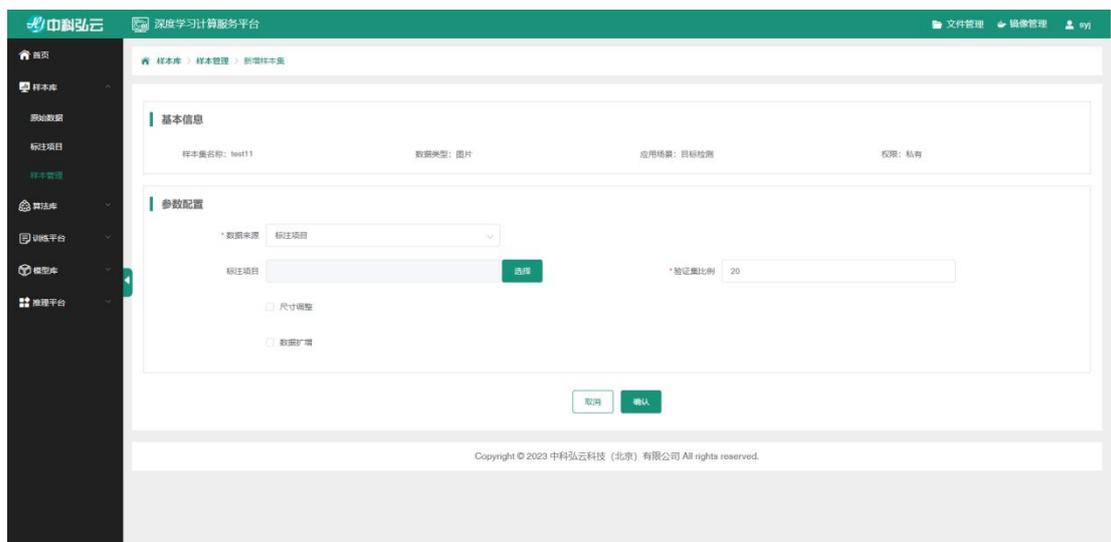
样本管理对平台标注的样本数据管理，同时支持将外部标注的样本数据导入平台，进行管理。

#### 5.3.1. 标注数据导入

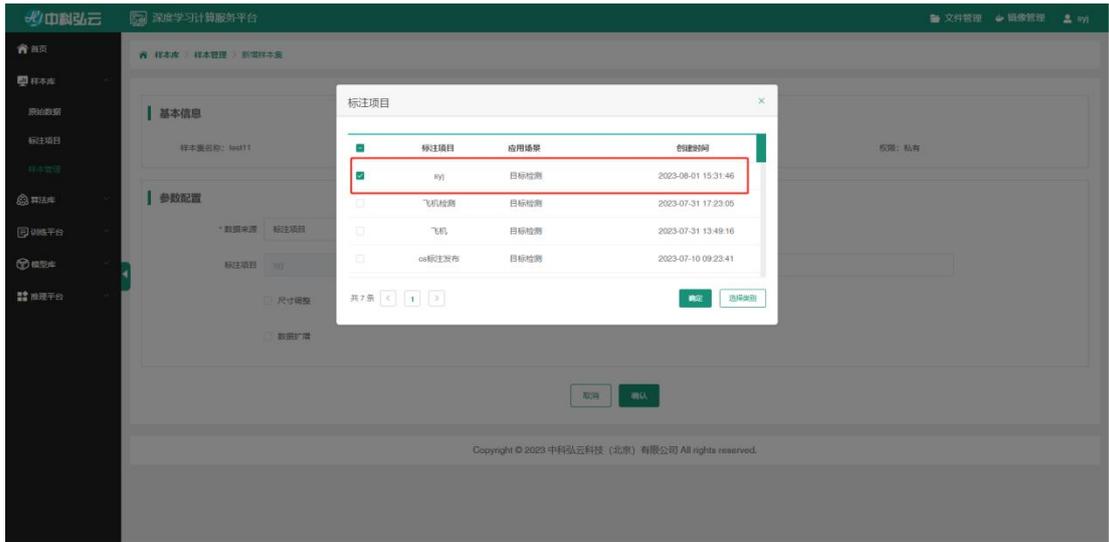
- 点击样本管理-->+样本集进行样本库的添加；



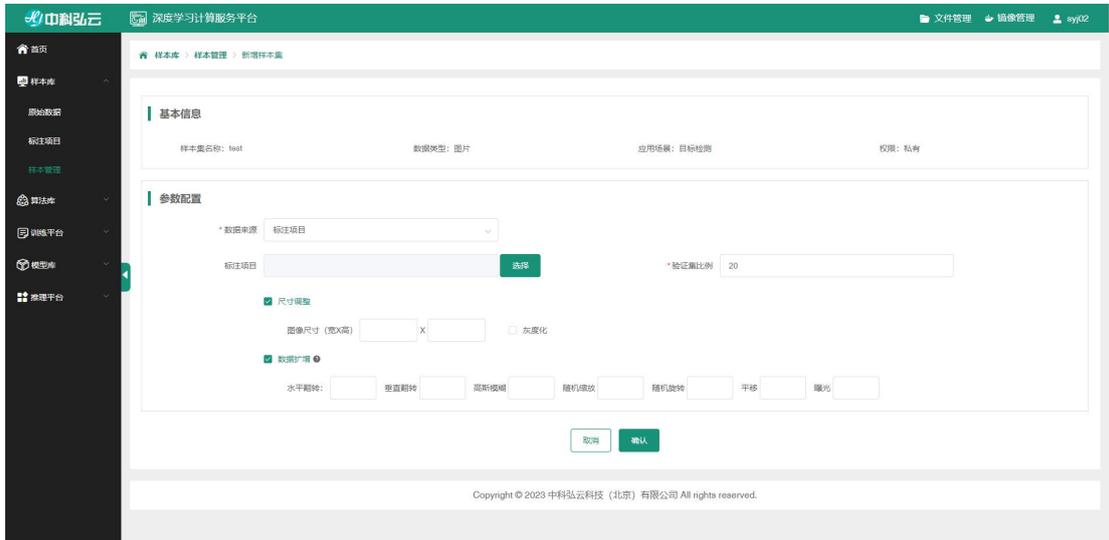
- 添加完成后点击+版本进行数据的导入；



- 点击选择导入标注完成的数据；（同时还可以筛选导入的类别。比如有多个类，如人、猫、狗，我们选择类别只选择人，导入的数据就只会导入类别为人的数据）

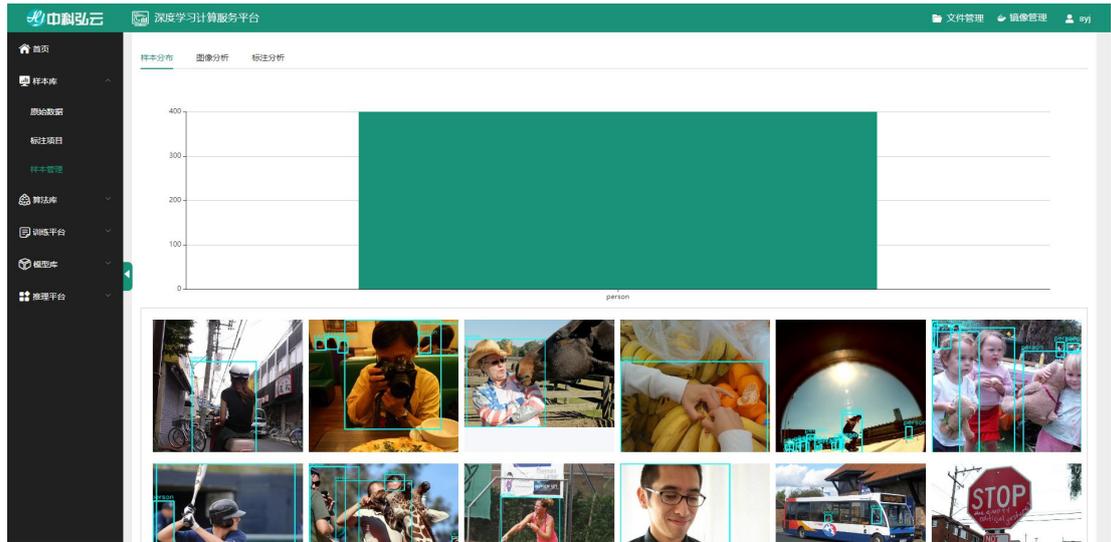


➤ 导入过程中还有数据扩增功能提供使用；



数据扩增支持水平旋转、垂直旋转、高斯模糊、随机缩放、随机旋转、平移、和曝光六种方式，框中填写要扩增样本集的比例，为正整数。

➤ 点击导入完成的数据可查看导入详情；



➤ 图像分析

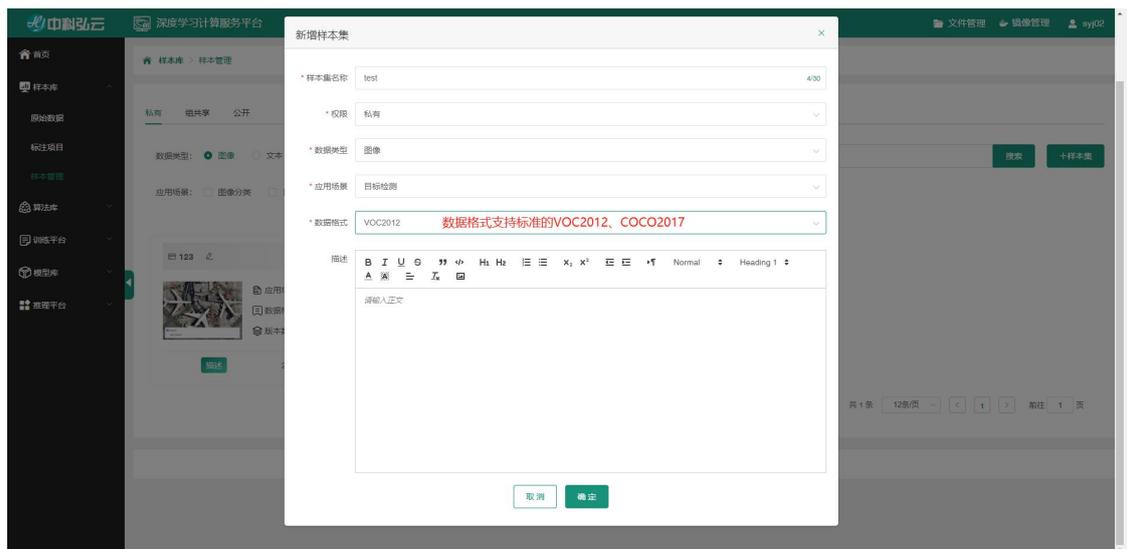


➤ 标注分析

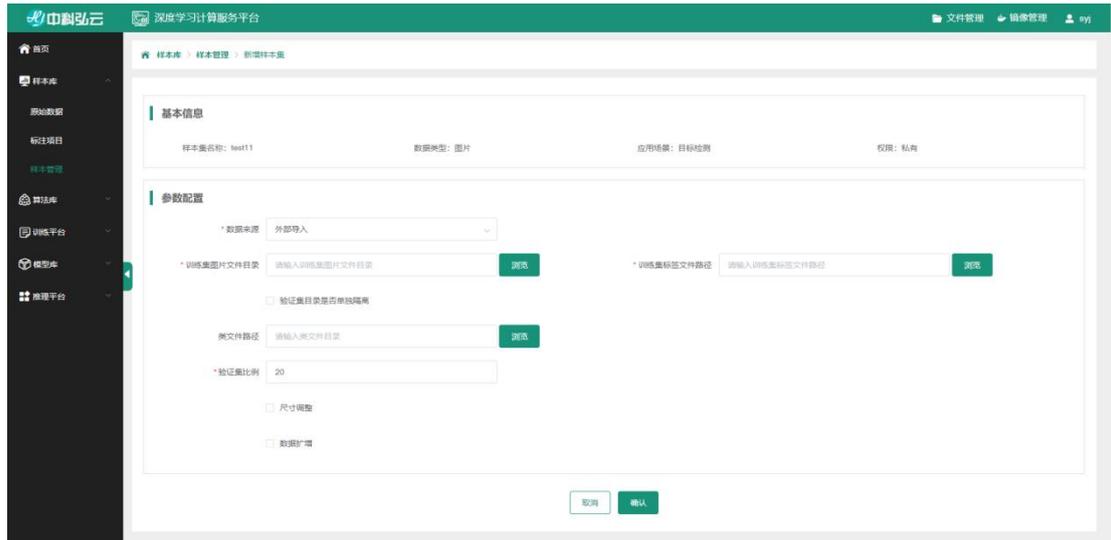


### 5.3.2. 外部数据导入

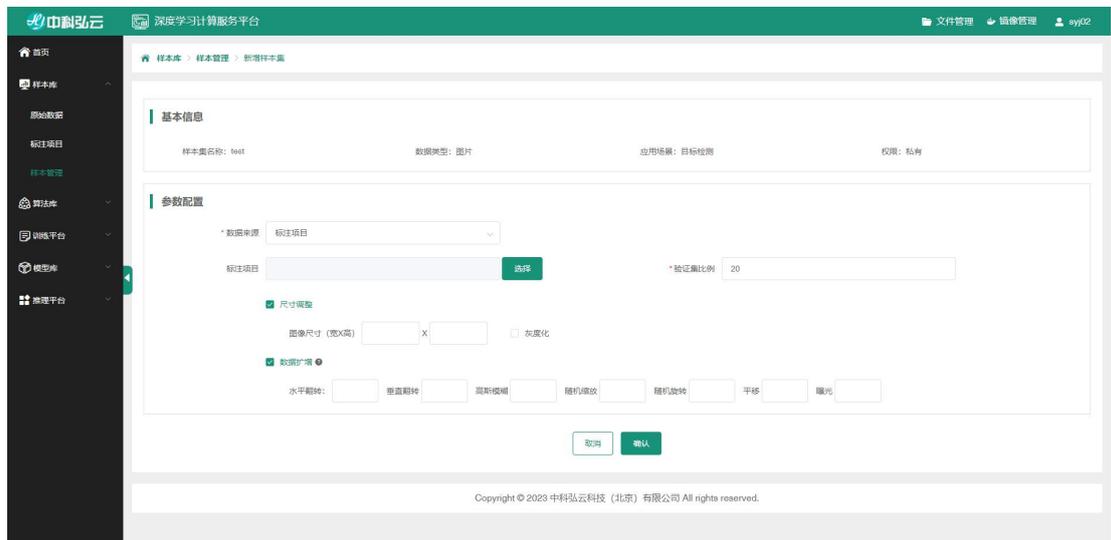
➢ 点击样本管理-->+样本集进行样本库的添加;



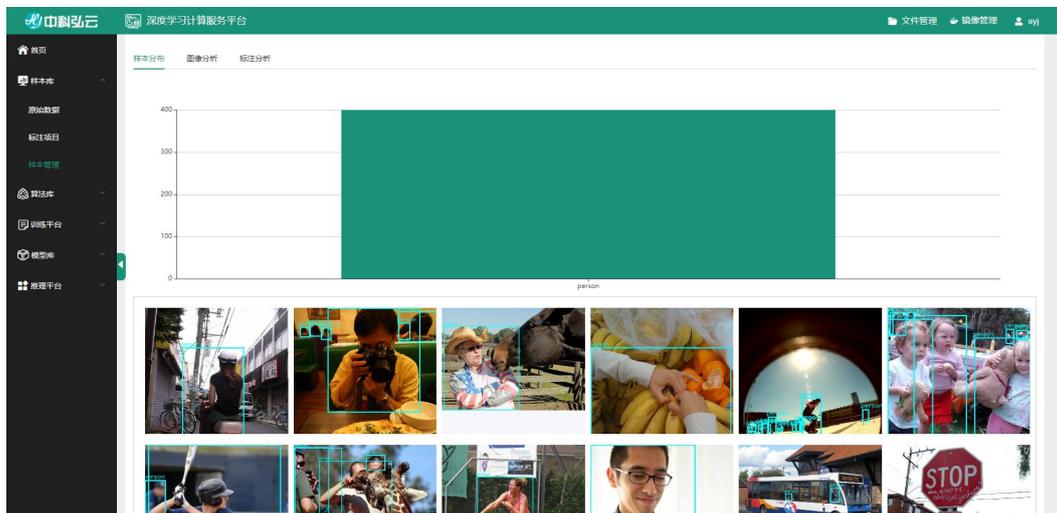
➢ 添加完成后点击+版本进行数据的导入。详情见 [4.3.2.1. 目标检测 VOC 外部导入](#)、[4.3.2.2. 目标检测 COCO 外部导入](#);



➤ 导入过程中还有数据扩增功能提供使用；



➤ 点击导入完成的数据可查看导入详情；



➤ 图像分析



➤ 标注分析



5.3.2.1. 目标检测 VOC 外部导入

➤ VOC 数据格式外部导入



➤ 目录结构

```
└─ VOC
  └─ Annotations
    ├── 2010_000152.xml
    ├── 2010_000159.xml
    └─ 2010_000162.xml
  └─ JPEGImages
    ├── 2010_000152.jpg
    ├── 2010_000159.jpg
    └─ 2010_000162.jpg
  └─ label.txt
```

5.3.2.2. 目标检测 COCO 外部导入

➤ COCO 数据格式外部导入



文件浏览

test	train	val
test.json	train.json	val.json

目标检测支持coco大json数据格式

当前路径:



➤ 目录结构

```
├─ VOC
│  ├─ colormap.txt
│  ├─ label.txt
│  ├─ train
│  │  ├─ images
│  │  │  ├─ 2010_000162.jpg
│  │  │  ├─ new-2010_000159_20220617-081244-0xdb2c.jpg
│  │  │  ├─ new-2010_000159.jpg
│  │  │  └─ new-2010_000162.jpg
│  │  └─ png_labels
│  │     ├─ 2010_000162.png
│  │     ├─ new-2010_000159_20220617-081244-0xdb2c.png
│  │     ├─ new-2010_000159.png
│  │     └─ new-2010_000162.png
│  └─ val
│     ├─ images
│     │  └─ 2010_000159.jpg
│     └─ png_labels
│        └─ 2010_000159.png
```

## 6. 算法库

算法库管理平台算法，包括预置算法、用户算法、代码管理三个模块，对算法和代码进行管理、集成；

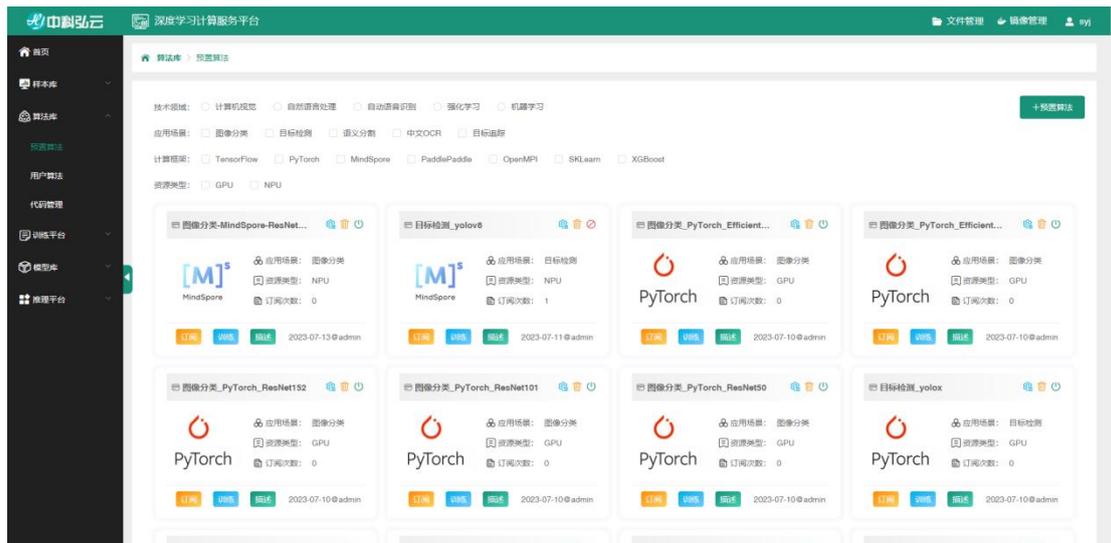
**预置算法：**平台内置的多种框架不同加速卡的 70+ 个算法；

**用户算法：**当平台内置的算法不满足用户需求时，用户可以按照平台规则去修改集成算法到平台上去使用；

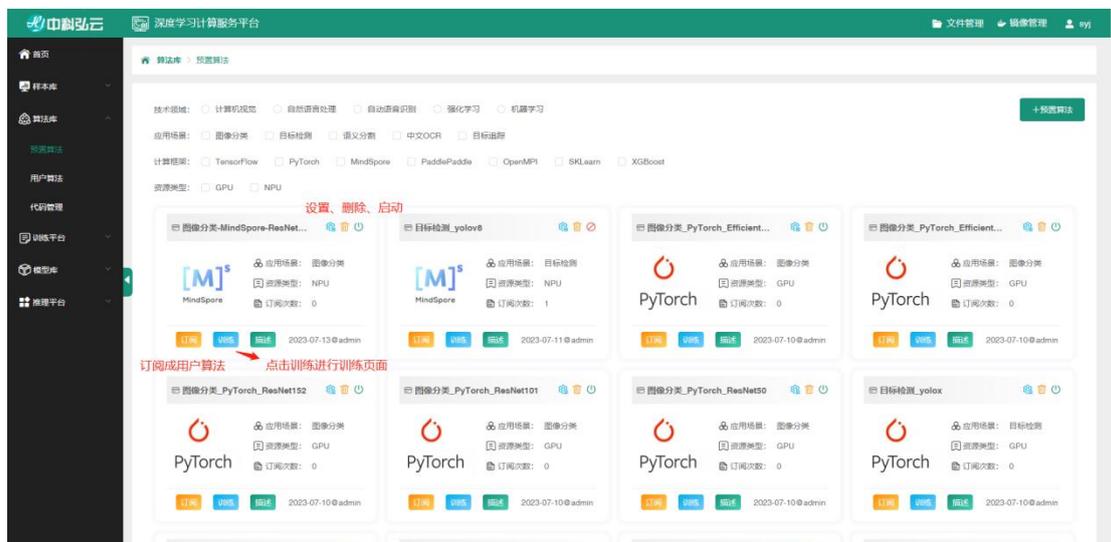
**代码管理：**当用户环境有代码仓库时，可以连接代码仓库，对代码进行拉取推送等功能；

### 6.1. 预置算法

➤ 点击算法库→预置算法可查看到平台内所有预置算法；



- 通过点击算法页上的按钮可对预置算法进行操作；



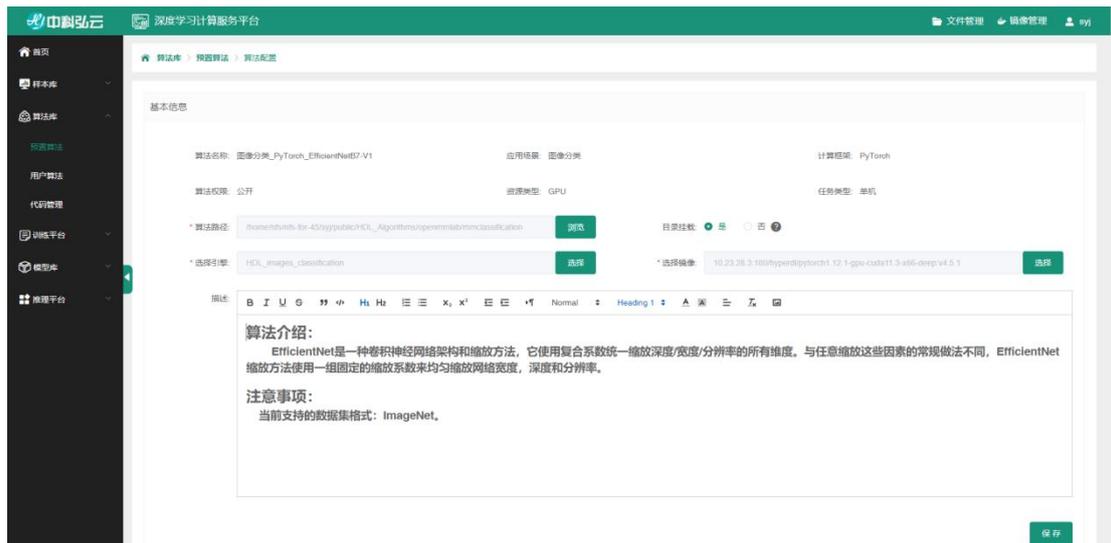
说明：拥有系统管理员角色的用户可以对预置算法进行操作，普通用户没有操作按钮。

## 6.1.1. 算法设置

### 6.1.1.1. 基本信息

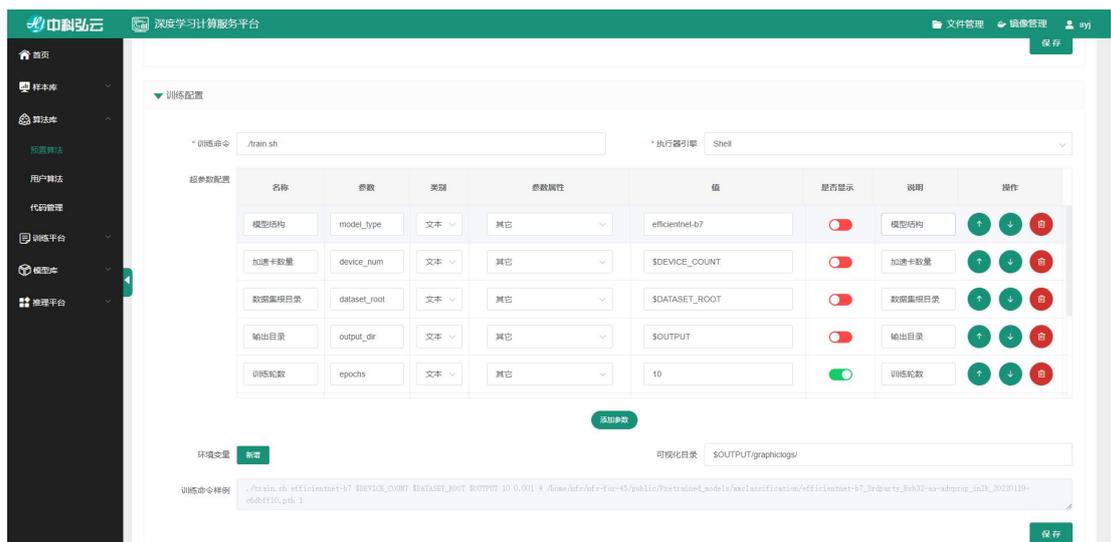
通过点击设置按钮可进入算法的配置页面，页面内提供了算法的基本信息、训练配置以及测试配置；

- 基本信息是配置了算法的运行镜像、代码目录、评估所执行的评估引擎（详情见 [7.4 评估引擎](#)）以及算法的描述；



### 6.1.1.2. 训练配置

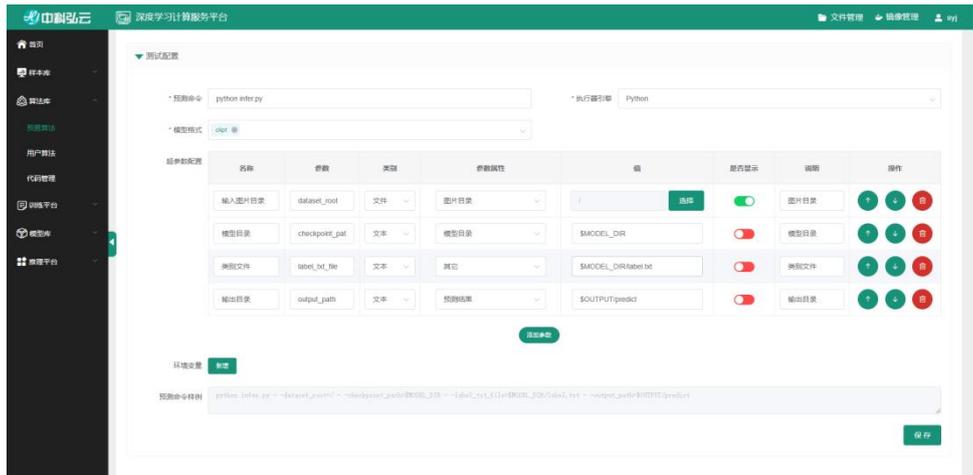
- 训练配置内可以查看并修改配置好的训练命令、参数；



说明：超参数配置列表中“名称”项为页面显示参数名，“参数”为训练脚本接收的参数名，“是否显示”为红色时，任务配置页面不显示该参数，为绿色时显示该参数。

### 6.1.1.3. 测试配置

- 测试配置内可以查看并修改配置好的预测命令、参数；

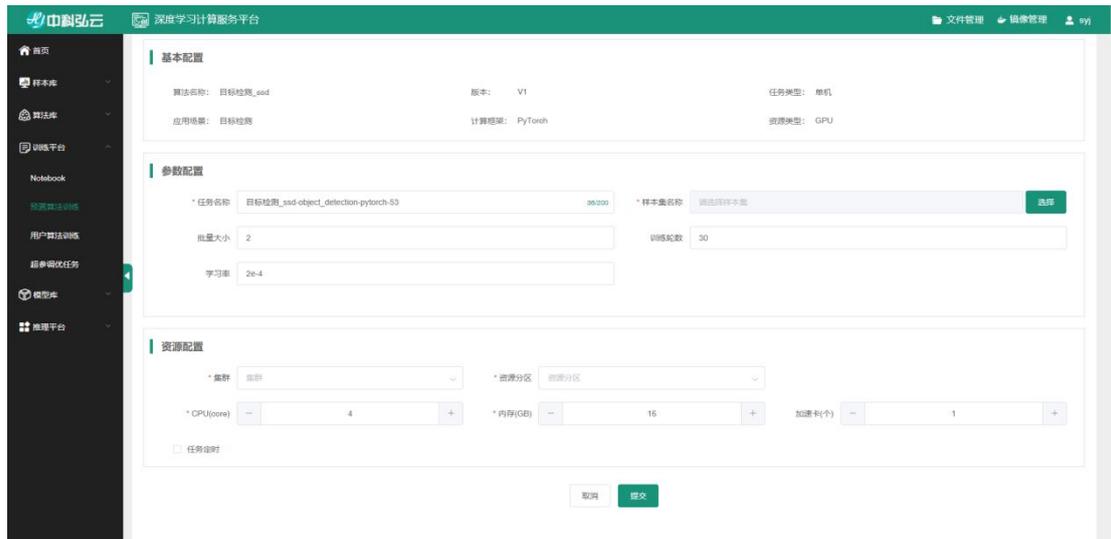


### 6.1.2. 订阅算法

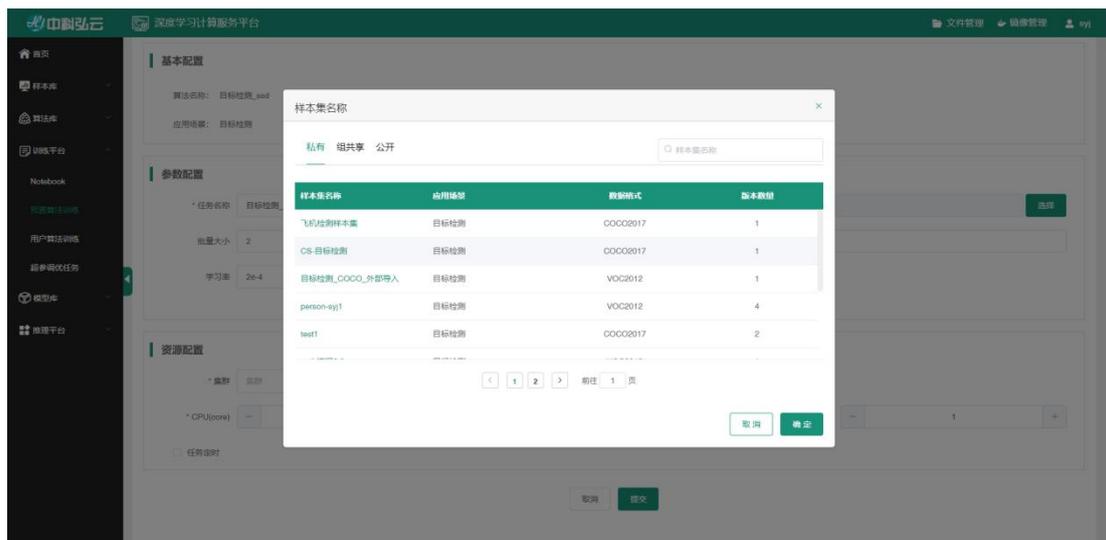
点击订阅会把仓库算法订阅成用户算法，提供给用户进行代码修改、调试参数使用；

### 6.1.3. 算法训练

➤ 点击训练会跳转到训练界面；



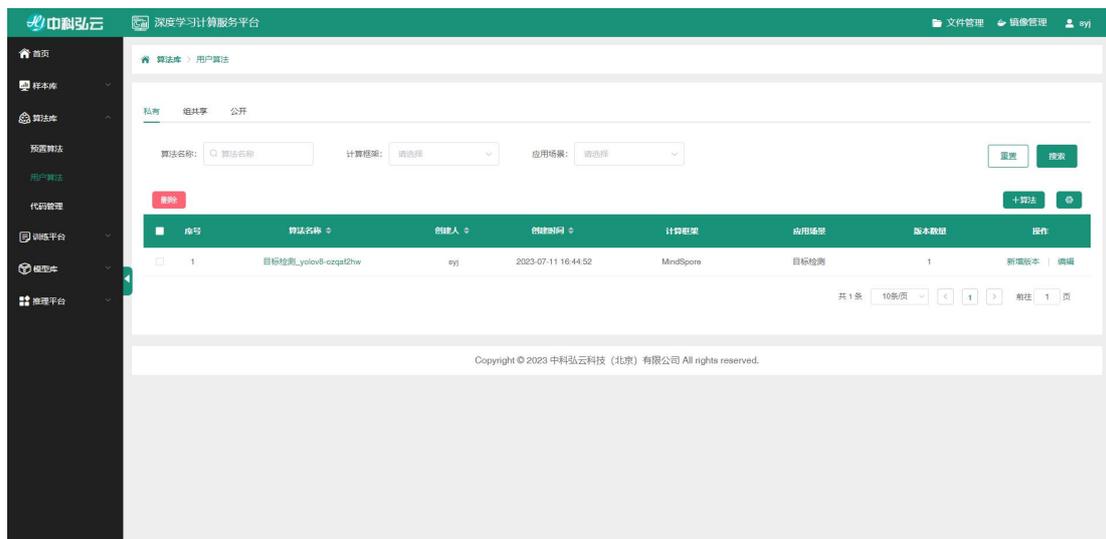
➤ 点击数据集就可以选择我们刚才发布完成到样本库的数据进行训练；



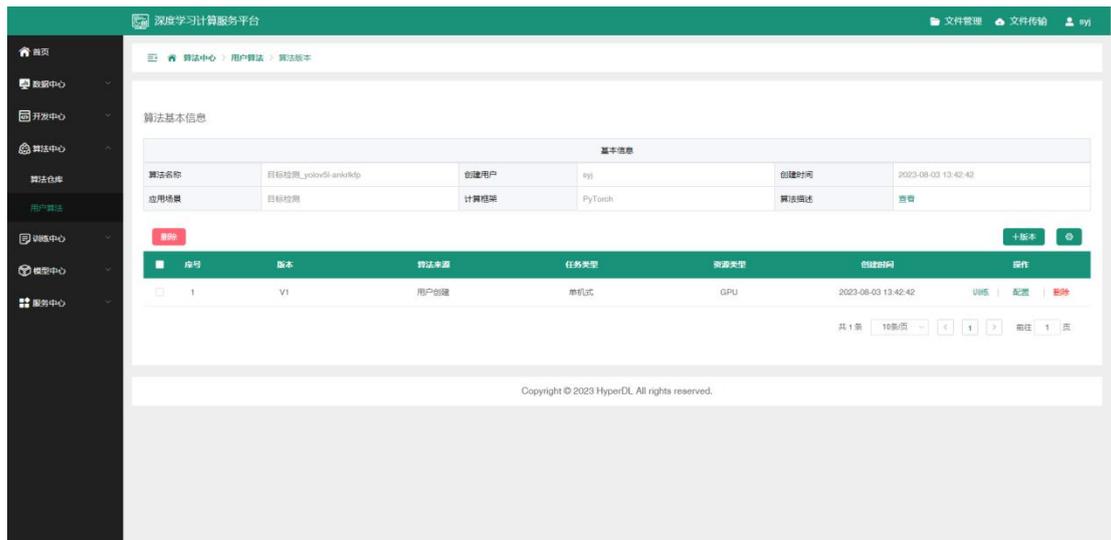
## 6.2. 用户算法

用户可以按照平台规则去修改集成算法到平台上去使用，或者订阅内置算法成为用户算法去修改使用；

- 点击算法库-->用户算法可查看用户算法列表



- 点击算法名称可查看代码版本、点击+版本可添加新的代码版本配置用户自己的算法；详情配置见 [6.2. 预置算法训练配置](#)；

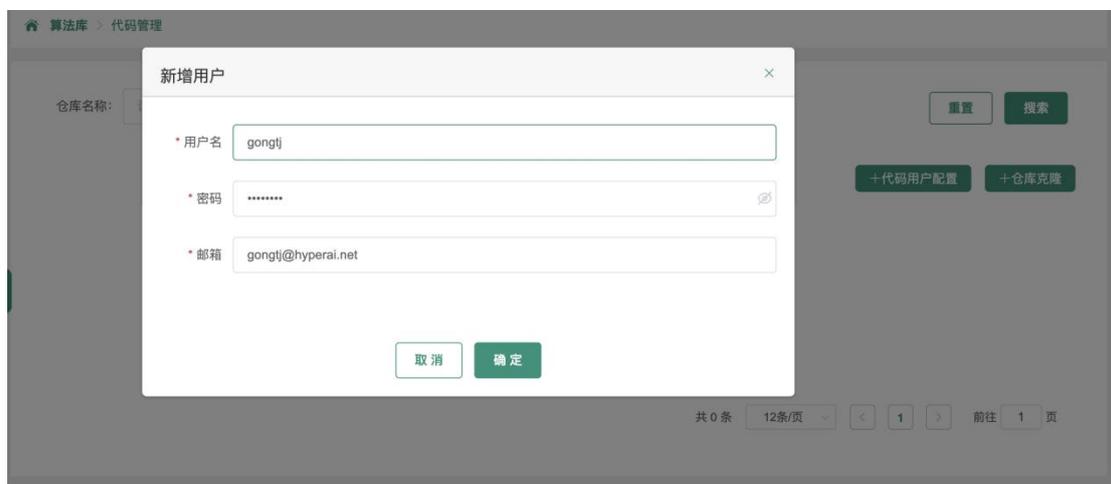


### 6.3. 代码管理

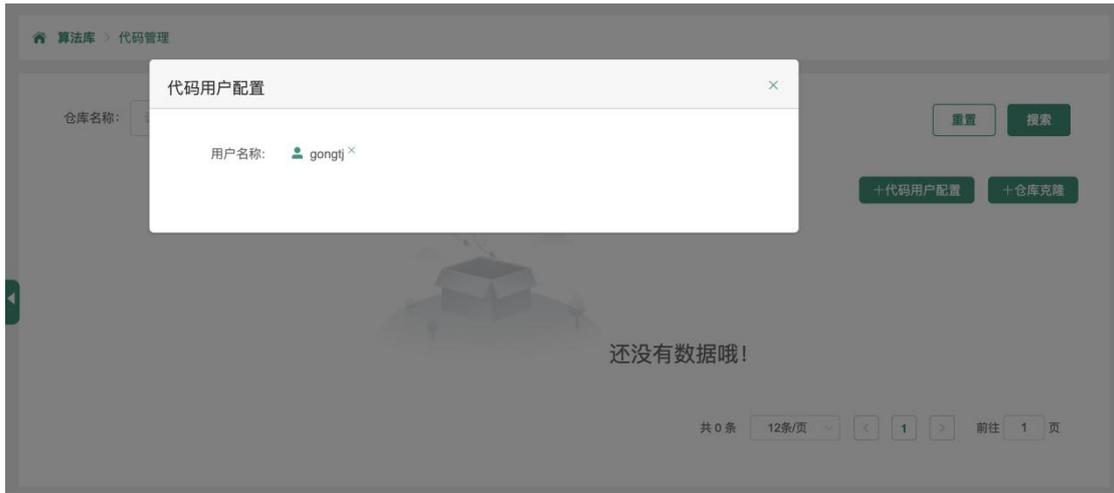
连接代码仓库后可实现拉取、上传功能；



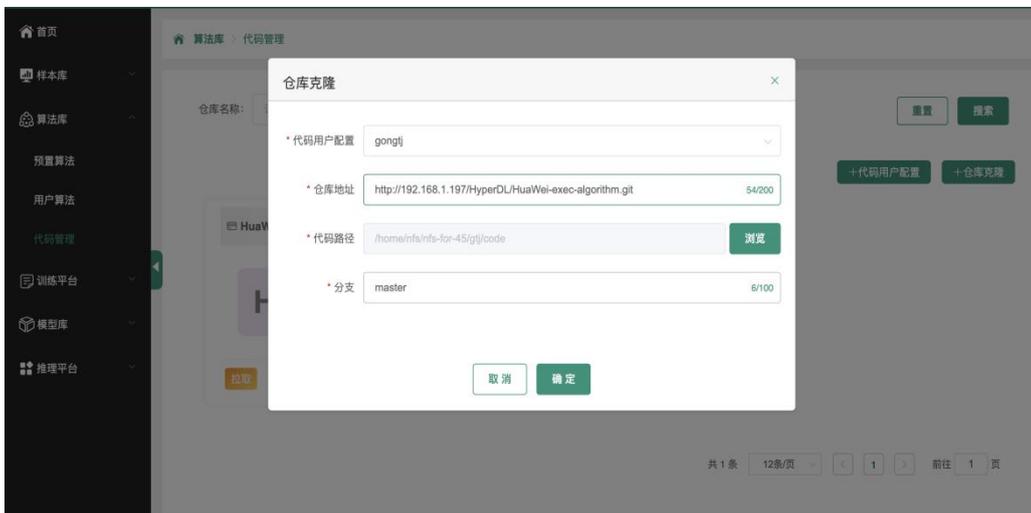
➤ 点击+代码用户配置 按钮，新增代码用户；配置用户名、密码、及邮箱参数。



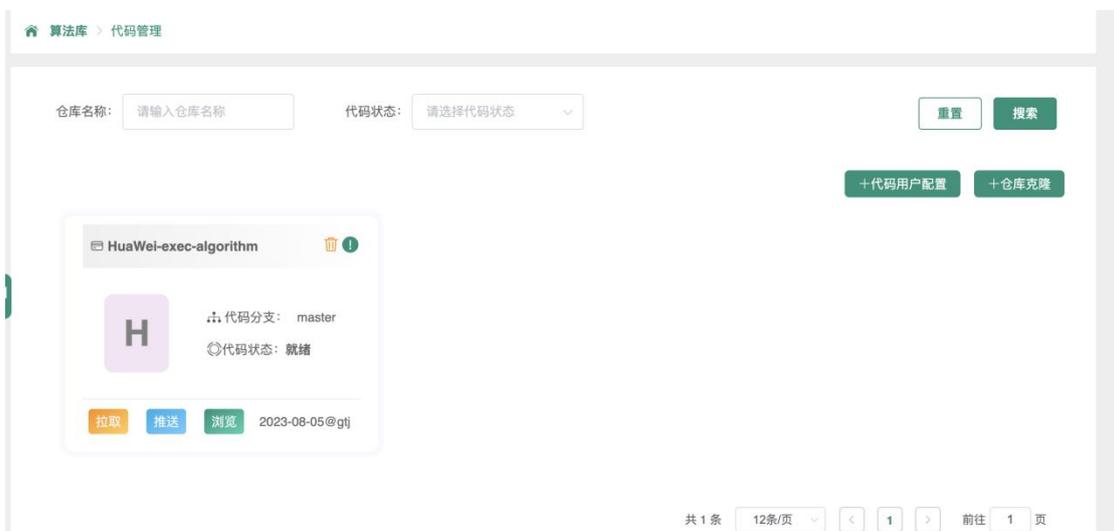
说明：用户、密码和邮箱为公司私有 gitlab 系统的用户账号，密码和邮箱。



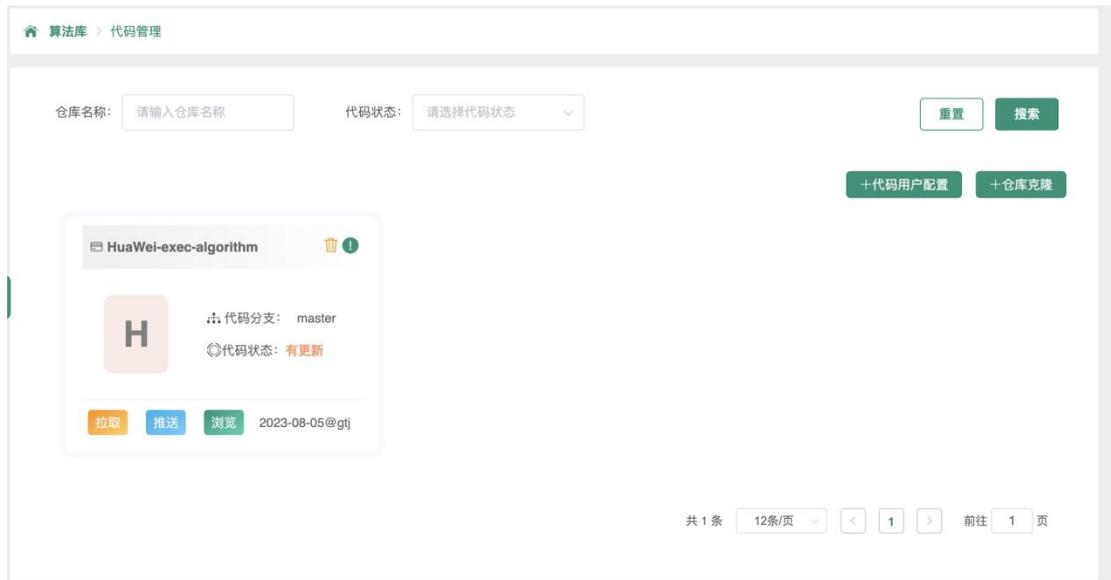
➤ 点击+仓库克隆，克隆代码到用户 code 目录下，



配置仓库地址、代码路径、分支等参数，确定后，平台自动拉取项目代码。



- 对代码进行拉取、推送、和浏览； 当代码有更新时，代码状态为“有更新”，提示要推送到 git 仓库。



## 7. 训练平台

训练平台提供预置算法和用户算法的模型训练，支持单机训练、分布式训练、以及超参调优方式训练，包括 Notebook、预置算法训练、用户算法训练、超参调优任务四个模块；可在平台进行在线开发训练、预置算法训练、用户算法训练以及超参调优训练；

**Notebook：** 提供开发环境在平台上进行在线开发；

**预置算法训练：** 纳管预置算法训练任务；

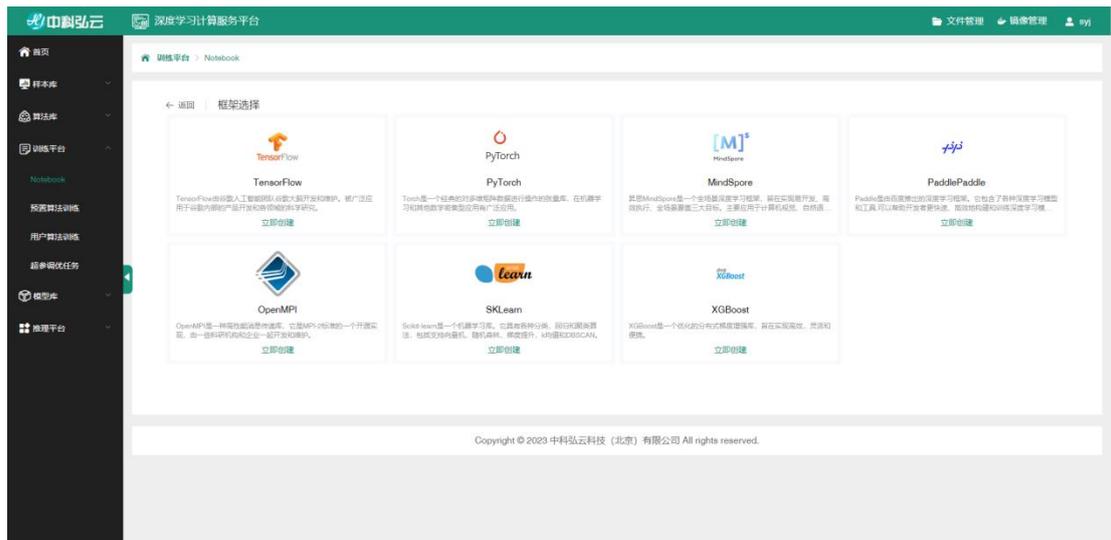
**用户算法训练：** 纳管用户算法训练任务；

**超参调优任务：** 纳管超参调优训练任务；

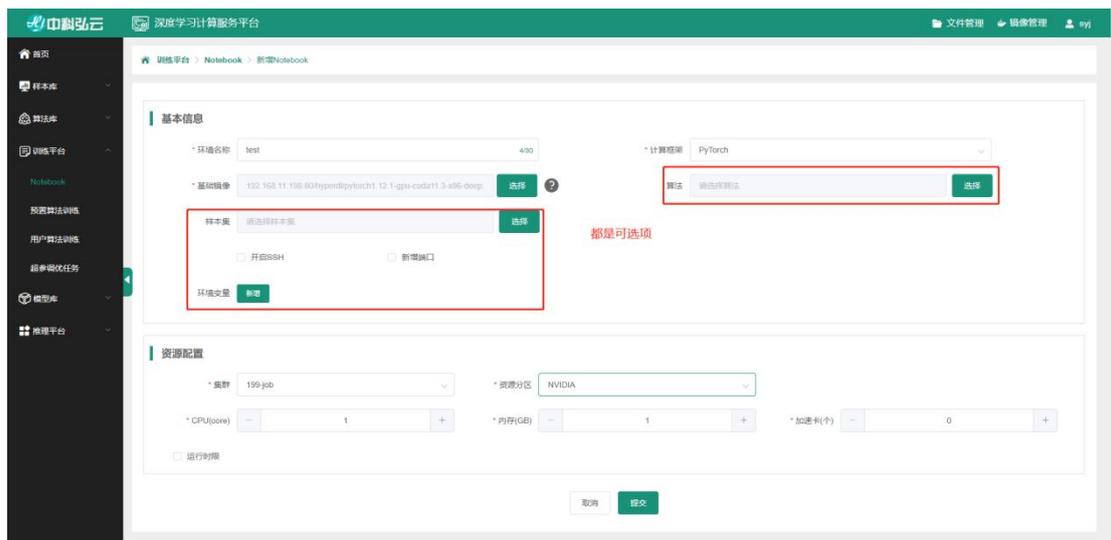
### 7.1. Notebook

Notebook 为算法开发人员提供容器环境，提供了远程开发功能，通过开启 SSH 连接，用户本地 SSH 可以远程连接到 HyperDL 的 Notebook 开发环境中，调试和运行代码；

- 点击训练平台-->Notebook-->+Notebook 可以进行框架的选择

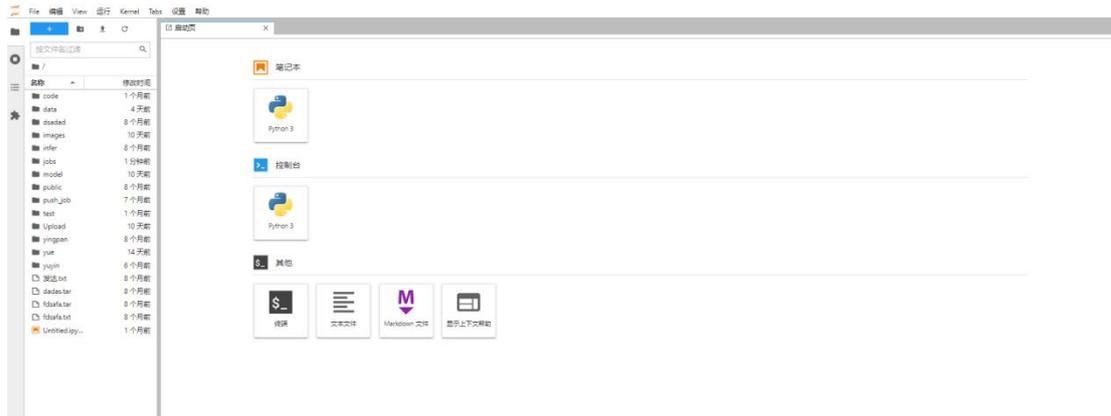


- 选择好框架进入到新增 Notebook，填写相关字段、配置镜像、算法和样本集以及 ssh、端口等都是可选项；

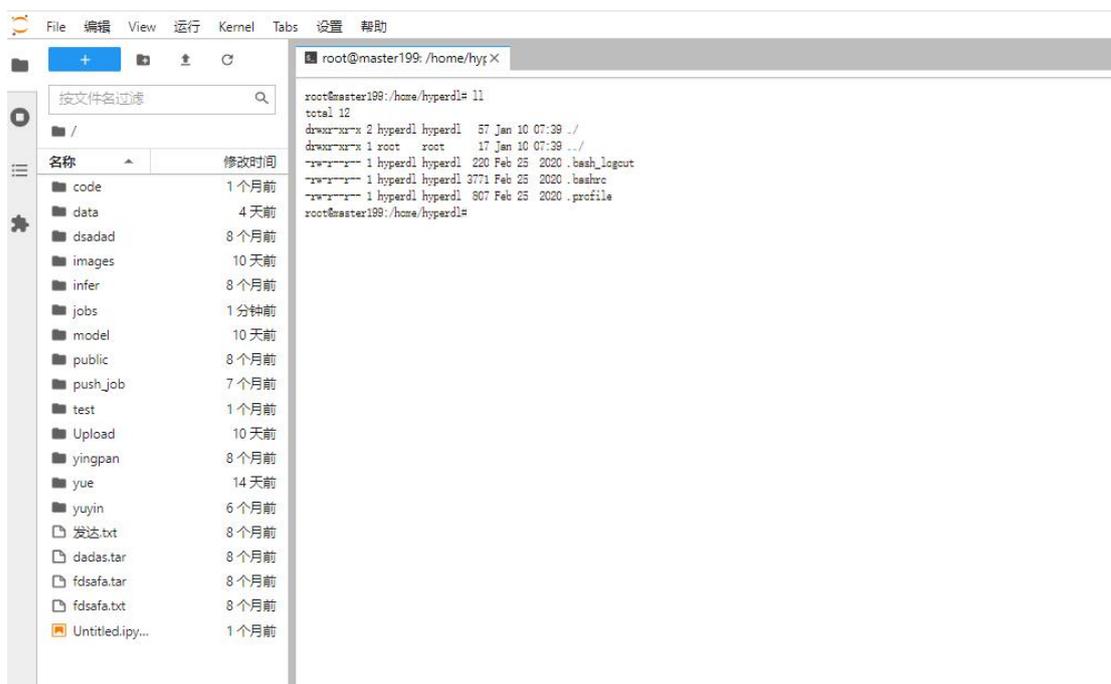


运行时限设置 Notebook 运行的时长（单位为分钟），超过后自动停止。

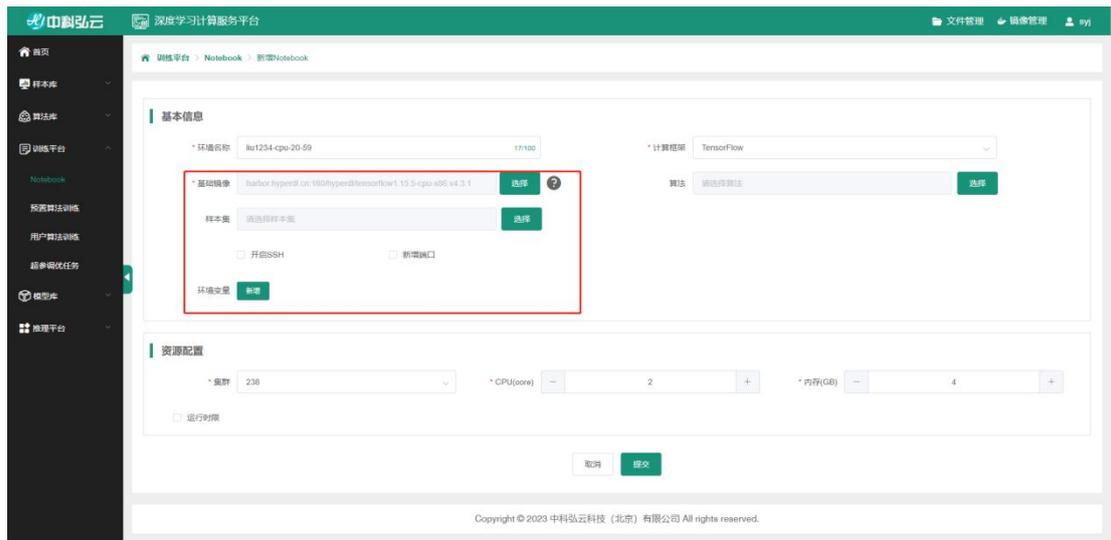
- 启动完成点击 Jupyter 进入操作页面；



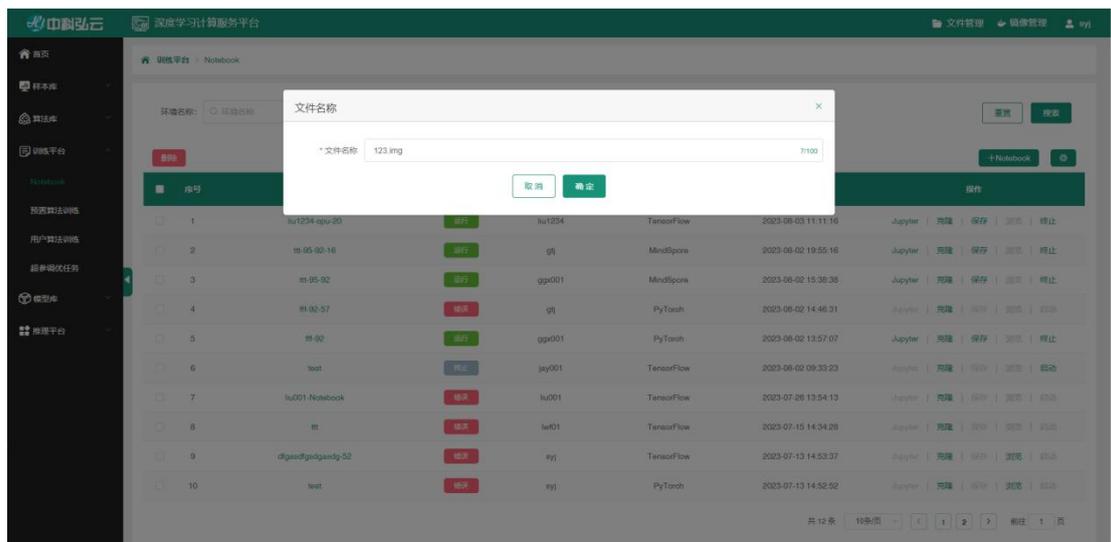
➤ 点击终端进入开发页面进行开发；



➤ 点击克隆复制一个新的任务出来使用；

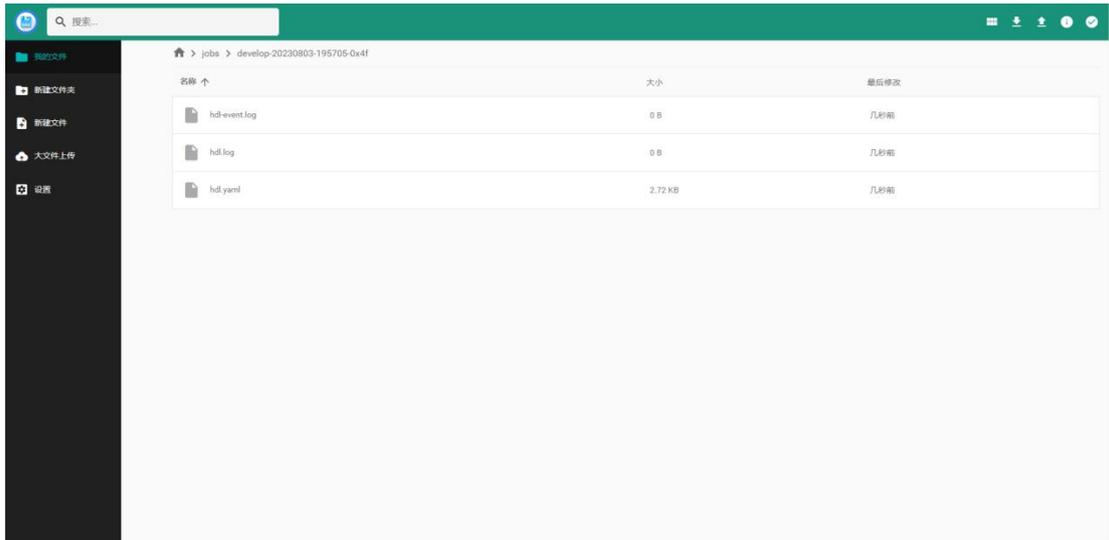


- 点击保存，输入想要保存的镜像名称，对开发的 Notebook 进行保存；保存完成会推送到 Harbor 仓库以及镜像列表中。

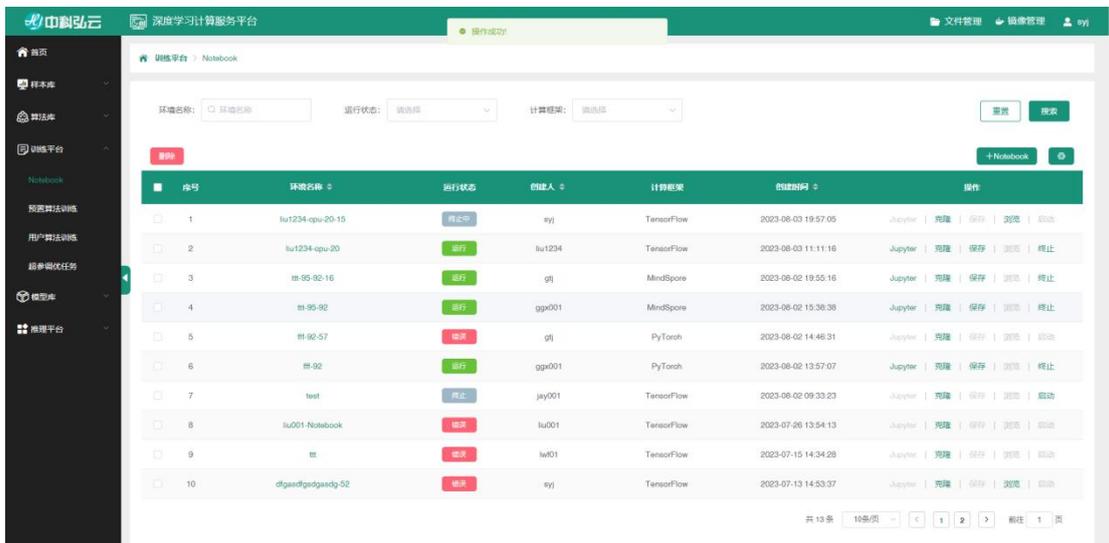


说明：镜像保存适用于对容器中安装新的软件包和工具，保存为新镜像。

- 点击浏览查看配置文件、打印日志等；

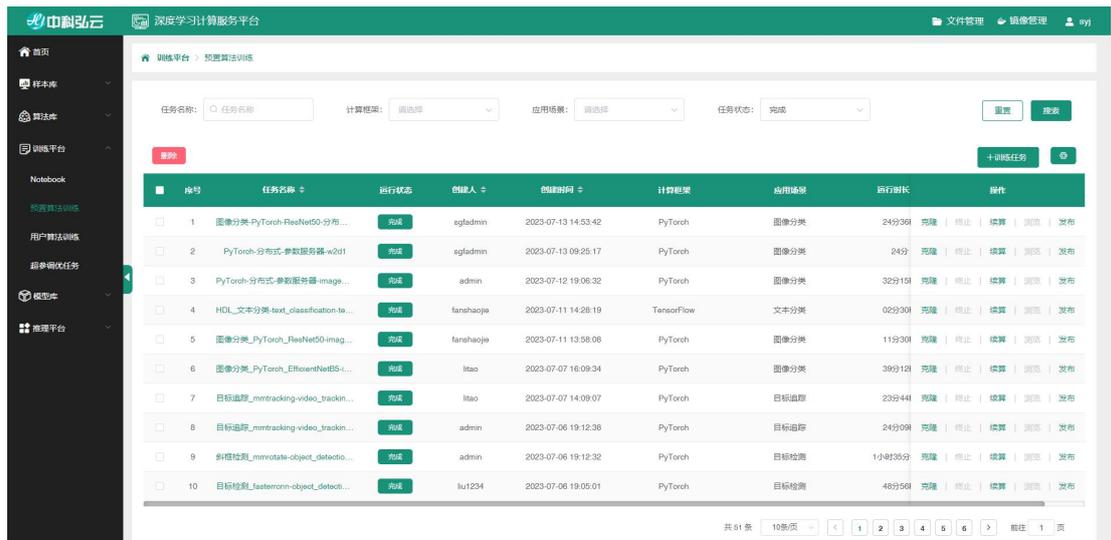


➤ 点击终止会终止正在运行的任务



## 7.2. 预置算法训练

➤ 点击训练平台-->预置算法训练可以查看到预置算法训练的任务；



通过操作栏按钮可以对任务进行操作；

**克隆：**复制一个训练任务参数；

**终止：**终止正在运行的任务；

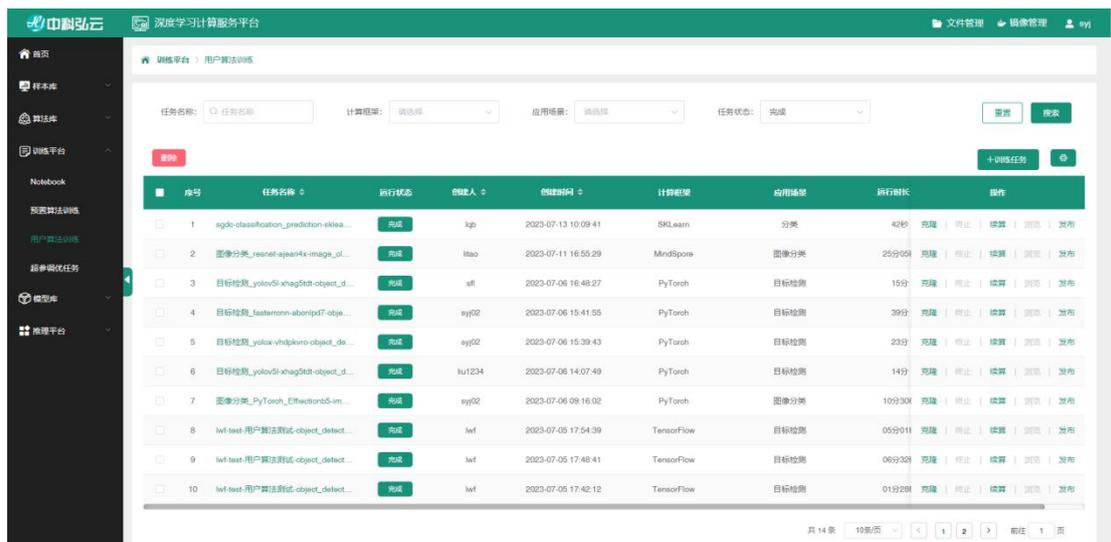
**续算：**在任务训练中断后可进行继续训练；

**浏览：**查看训练过程中产生的日志、模型等文件；

**发布：**把训练完成的模型发布到模型仓库进行纳管；

### 7.3. 用户算法训练

➤ 点击训练平台-->用户算法训练可以查看到用户算法训练的任务；



## 7.4. 超参调优任务

- 点击训练平台-->超参调优任务可以查看到超参调优算法训练的任务：

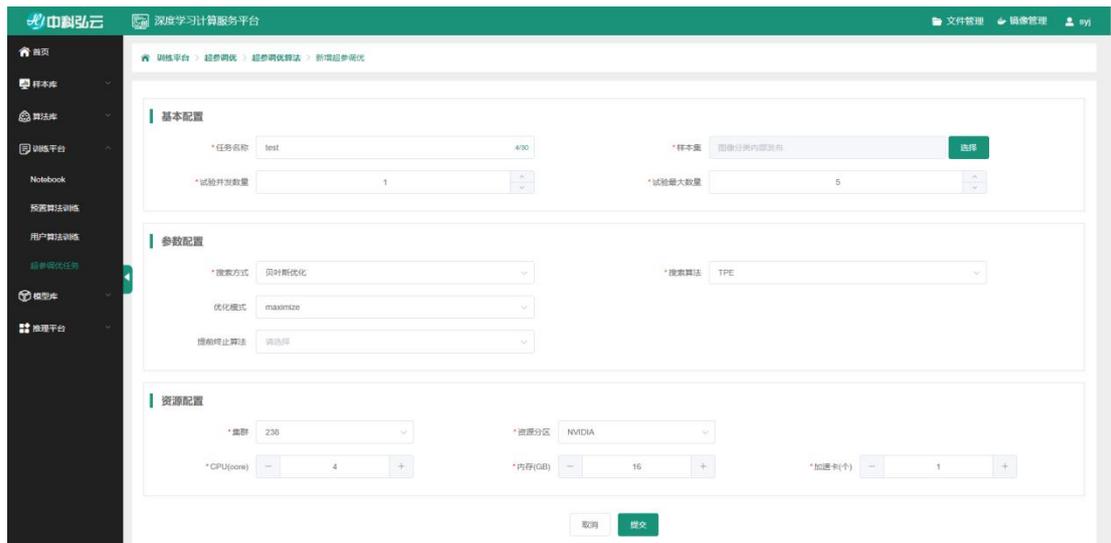
序号	任务名称	运行状态	创建人	创建时间	应用框架	计算框架	运行耗时	操作
1	on主流程-超参	运行	liao	2023-07-13 10:21:19	图像分类	PyTorch	6小时23分	克隆   终止   调参   可视化   发布
2	on主流程-超参	完成	liao	2023-07-11 16:56:52	图像分类	PyTorch	5小时21分	克隆   终止   调参   可视化   发布
3	on目标检测	停止	liao	2023-07-10 17:49:30	目标检测	MindSpore	0分	克隆   终止   调参   可视化   发布
4	ldg目标检测	完成	ku1234	2023-07-06 19:10:34	目标检测	PyTorch	18分	克隆   终止   调参   可视化   发布
5	on1.5	完成	liao	2023-07-06 17:31:04	图像分类	PyTorch	0分	克隆   终止   调参   可视化   发布
6	测试	完成	ali	2023-07-06 16:48:32	目标检测	PyTorch	23分	克隆   终止   调参   可视化   发布
7	测试	完成	hw01	2023-07-06 10:27:26	目标检测	PyTorch	14分0	克隆   终止   调参   可视化   发布
8	测试	停止	hw01	2023-07-06 10:22:36	目标检测	PyTorch	0分	克隆   终止   调参   可视化   发布
9	test	停止	gga001	2023-07-05 17:48:33	图像分类	PyTorch	0分	克隆   终止   调参   可视化   发布
10	ldg-图像分类	完成	ku1234	2023-07-04 18:04:05	图像分类	PyTorch	3小时20分	克隆   终止   调参   可视化   发布

- 点击+新建创建新的超参调优任务：

The interface displays a grid of algorithm cards for selection. Each card includes the algorithm name, application framework, source type, and creation details.

算法名称	应用框架	资源类型	创建时间
图像分类PyTorch-NPU-ARM	PyTorch	NPU	2023-06-26@admin
图像分类PyTorch-GPU-X86	PyTorch	GPU	2023-06-26@admin
目标检测_Yolov5-ARM	MindSpore	NPU	2023-06-26@admin
目标检测_Yolov5-X86	PyTorch	GPU	2023-06-26@admin
PyTorch_自定义	PyTorch	GPU	2023-06-26@admin
TensorFlow_自定义	TensorFlow	GPU	2023-06-26@admin
MindSpore_自定义	MindSpore	NPU	2023-06-26@admin

- 选择想要使用的算法点击训练，跳转到训练参数配置填写任务名称、样本集，配置参数配置、资源进行提交即可训练；



## 8. 模型库

模型库对平台中模型进行管理，包括模型管理、预测任务、评估任务、评估引擎、转换任务、转换工具六个模块；对算法训练完成的模型进行预测、评估、转换以及发布在线服务等操作；

**模型管理：**对训练的算法模型或者外部模型进行导入纳管使用；

**预测任务：**纳管预测任务；

**评估任务：**纳管评估任务；

**评估引擎：**提供内置转换引擎；

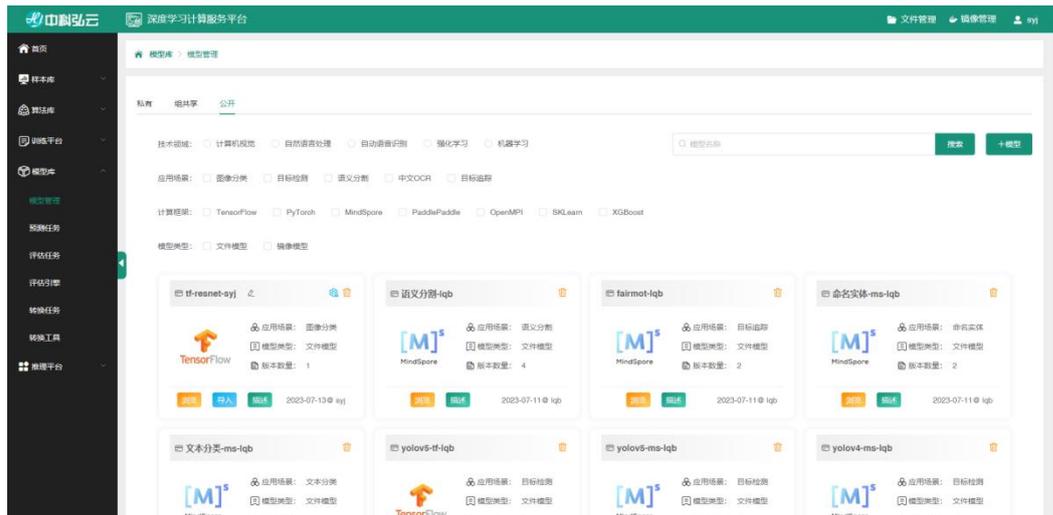
**转换任务：**纳管转换任务；

**转换工具：**提供内置转换工具；

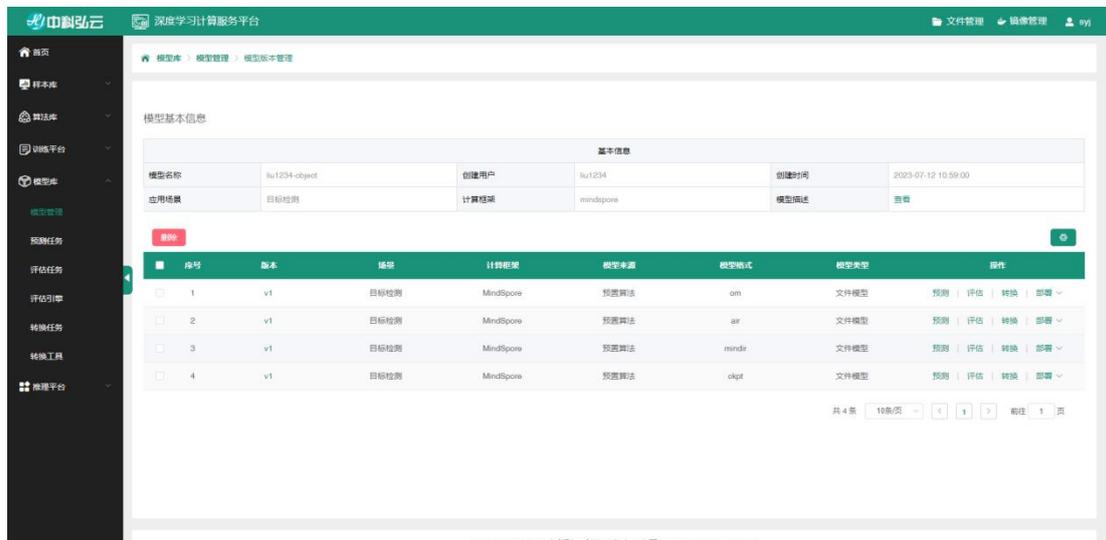
### 8.1. 模型管理

针对于训练平台训练的模型，模型库对发布的模型进行统一的纳管使用；

- 点击模型库-->模型管理可以查看到发布成功的模型仓库；



- 选择想要操作的模型藏库，查看模型版本信息；



通过操作栏按钮可以对模型进行操作；

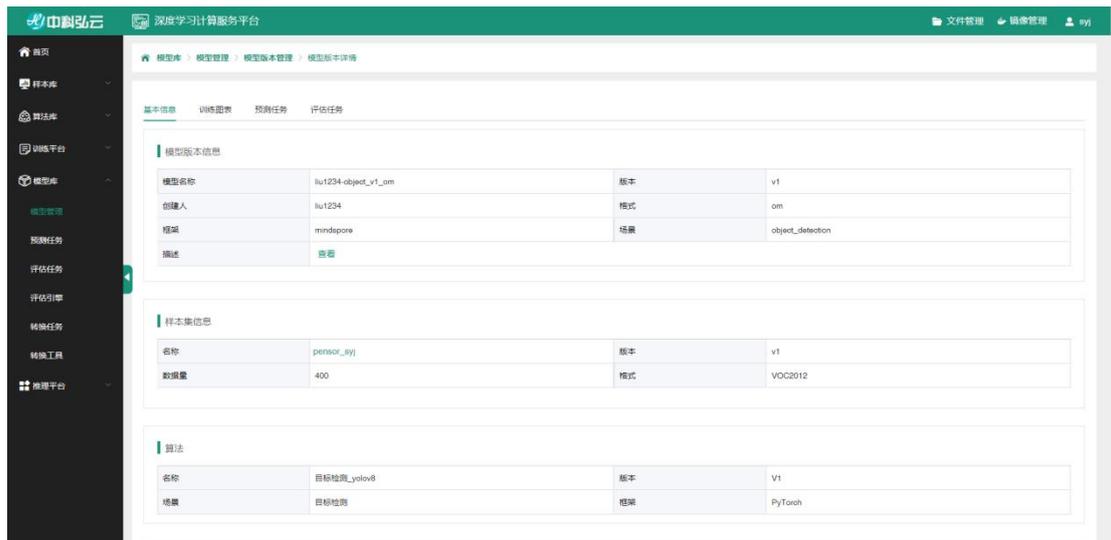
**预测：**对模型进行预测；

**评估：**对模型评估，输出评估指标；

**转换：**对模型进行格式转换；

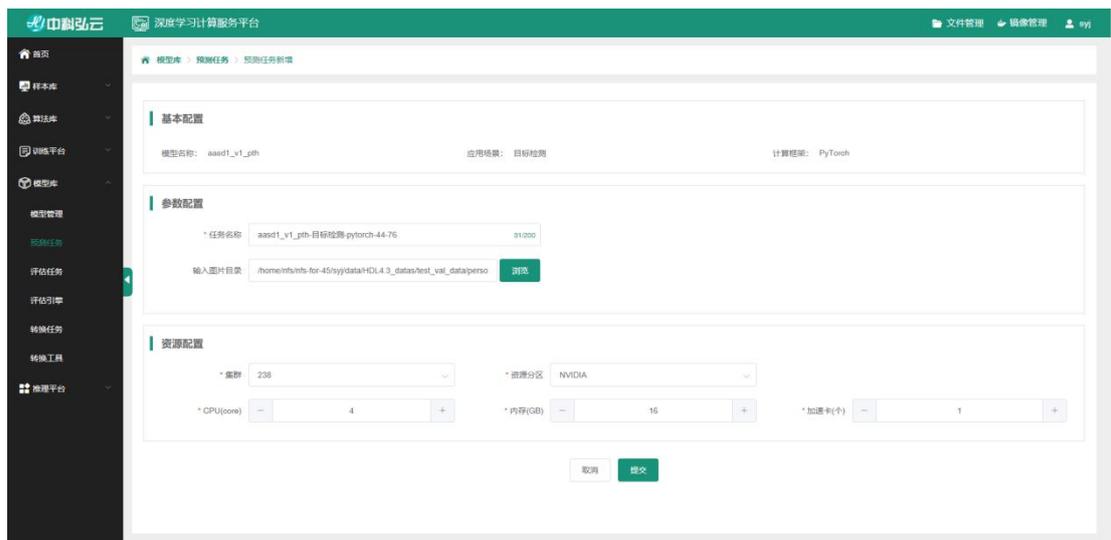
**部署：**将模型进行部署，发布在线推理服务；

- 点击模型版本号可以查看训练信息、样本信息、训练图表等等信息；

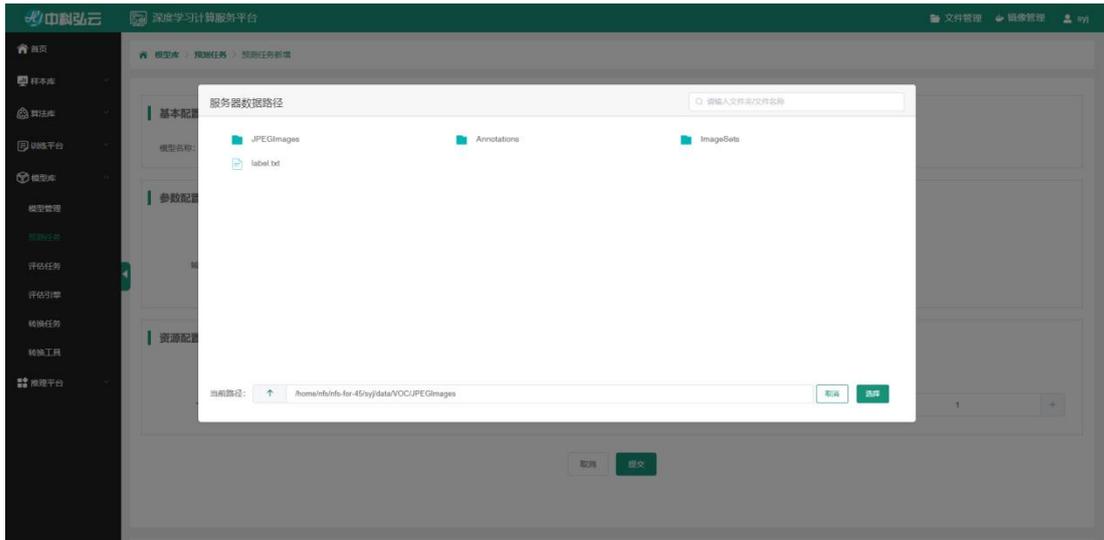


## 8.2. 预测任务

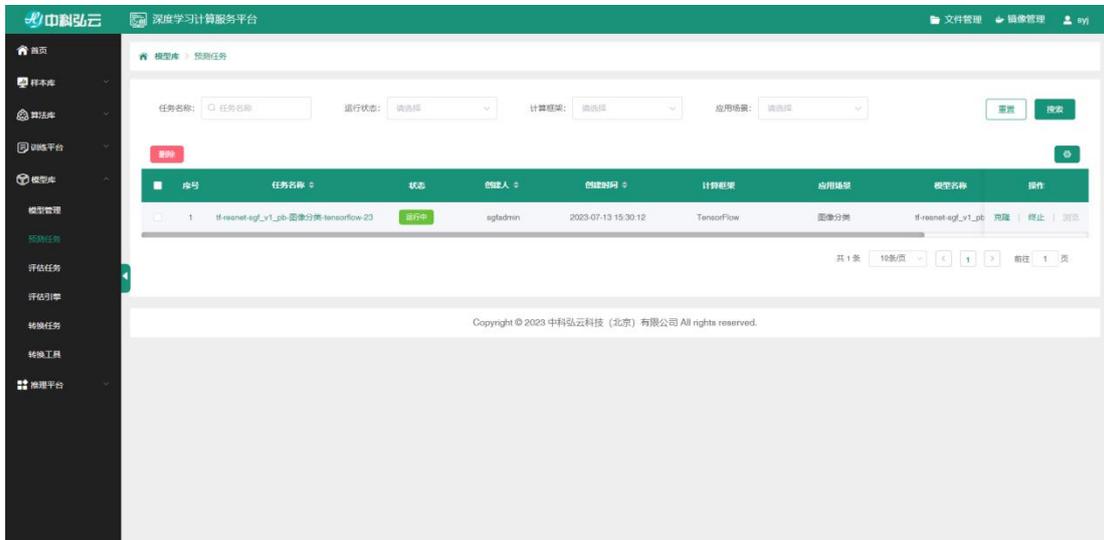
- 从模型库选择想要预测的模型点击预测即可跳转到预测任务进行选择；



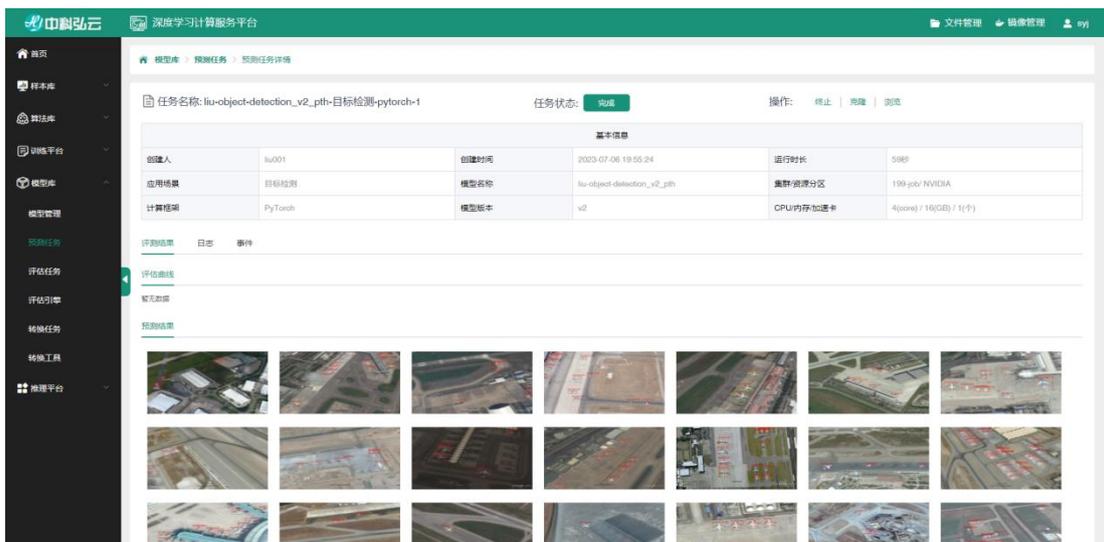
- 点击浏览选择想要进行预测数据进行预测；



➤ 点击模型库-->预测任务可以查看到模型跑的预测任务；

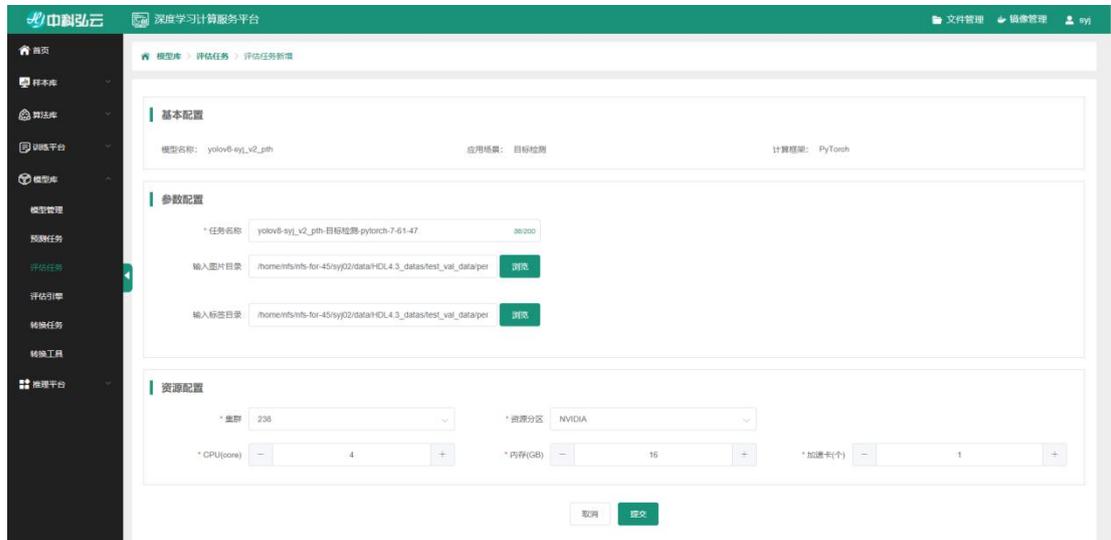


➤ 点击预测完成的任务可以在评测结果中看到预测完成的影像；

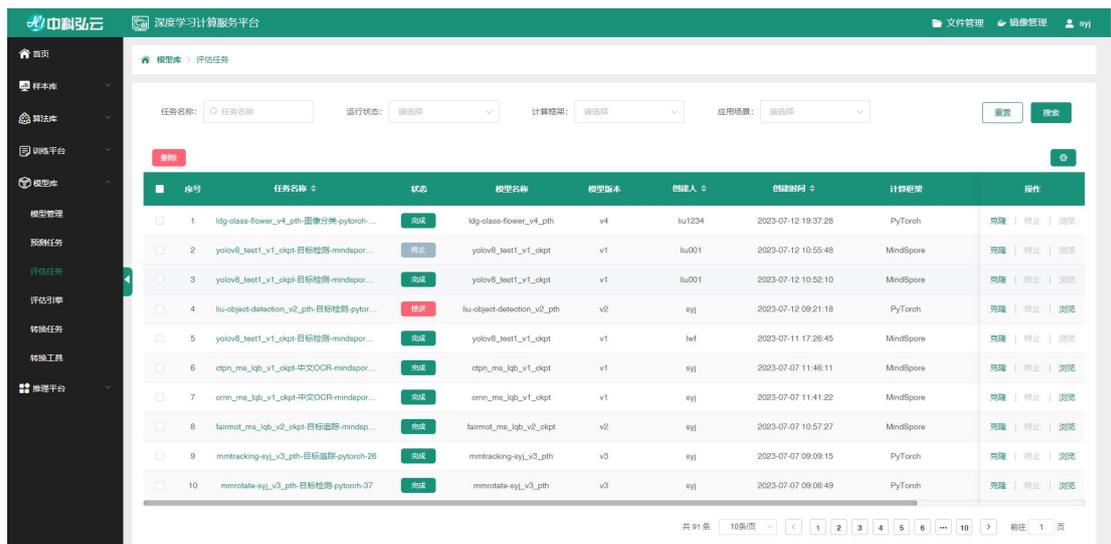


### 8.3. 评估任务

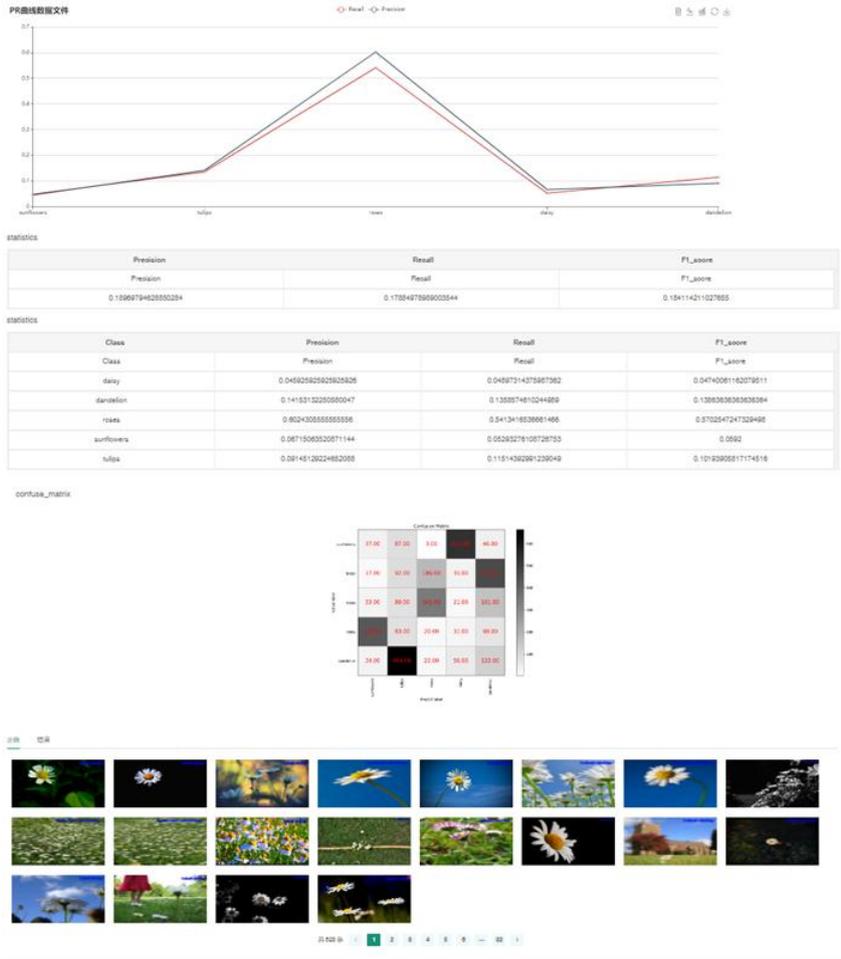
- 从模型库选择想要评估的模型，点击评估即可跳转到评估任务进行选择；目标检测模型的评估，点击浏览分别传入图像和真实标签后进行评估；



- 点击模型库-->评估任务可以查看到模型跑的评估任务；

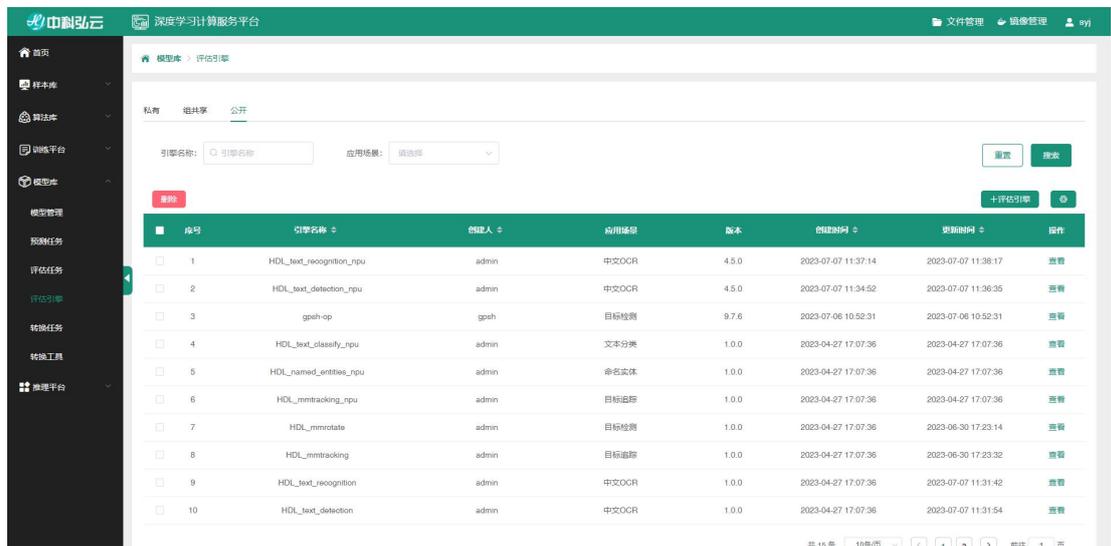


- 点击评估完成的任务可以在评测结果中看到数据的评估指标；（下图为图像分类评估的指标和结果）。



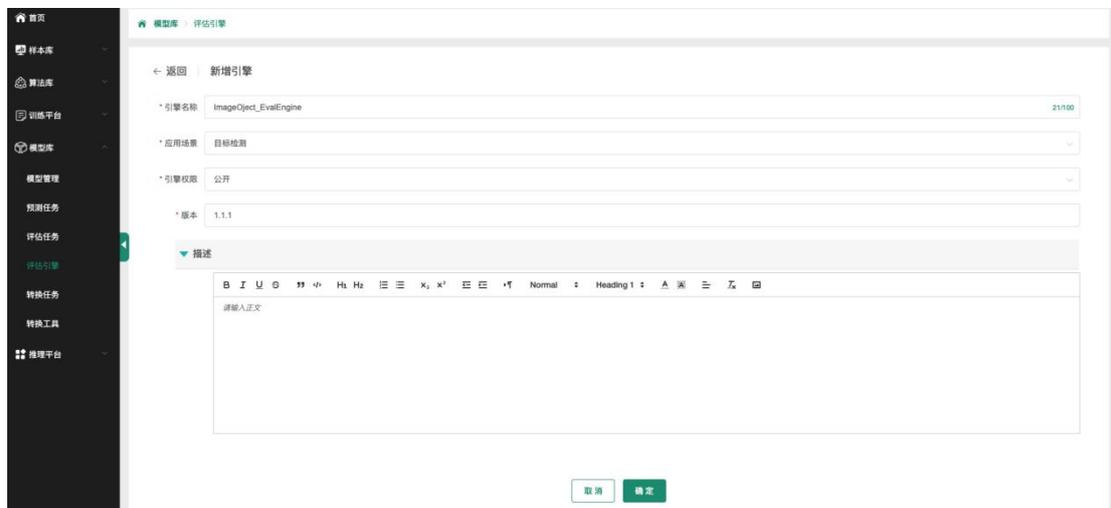
## 8.4. 评估引擎

- 点击模型库-->评估引擎-->公开可以查看到平台内置的评估引擎；



评估可视化的图表是由评估引擎生成的，平台上不同场景对应使用相关的评估引擎；同时平台也支持用户自定义评估引擎；

- 自定义评估引擎，点击+评估引擎，配置评估引擎属性参数，点击确定。



序号	引擎名称	创建人	应用场景	版本	创建时间	更新时间	操作
1	ImageObject_EvalEngine	gjl	目标检测	1.1.1	2023-08-05 13:57:44	2023-08-05 13:57:44	配置
2	test	gpx001	目标检测	2.2.2	2023-08-04 19:05:44	2023-08-04 19:06:23	查看
3	fas	gpx001	图像分类	2.3.2	2023-08-02 19:52:14	2023-08-02 19:52:15	查看
4	hdi-mi-xgboost-regression	lqb01	回归	4.5.1	2023-08-01 15:54:45	2023-08-01 16:38:16	查看
5	hdi-mi-xgboost-classify	lqb01	分类	4.5.1	2023-08-01 14:27:46	2023-08-01 14:41:09	查看
6	hdi-mi-classify	syj	分类	4.5.0	2023-07-26 14:25:41	2023-07-26 18:12:09	查看
7	hdi-mi-regression	syj	回归	4.5.0	2023-07-26 13:46:52	2023-07-26 16:32:04	查看
8	HDL_text_recognition_npu	admin	中文OCR	4.5.0	2023-07-07 11:37:14	2023-07-07 11:38:17	查看
9	HDL_text_detection_npu	admin	中文OCR	4.5.0	2023-07-07 11:34:52	2023-07-07 11:36:35	查看
10	HDL_text_classify_npu	admin	文本分类	1.0.0	2023-04-27 17:07:36	2023-04-27 17:07:36	查看

➤ 点击评估引擎列表页，点击配置，进行评估引擎配置。

基本信息

引擎名称: ImageObject\_EvalEngine    应用场景: 目标检测    版本: 1.1.1    描述: 查看

权限: 公开

配置信息

\* 执行镜像: 请选择执行镜像 [选择]    \* 评估脚本: /home/hf/hf-for-45/HyperAI-Evaluate-Engine-Dist/ImageObject\_EvalEngine/522 [选择]

\* 评估命令: [输入框]    \* 执行器引擎: Python

参数	名称	参数名	参数类别	参数属性	参数值	是否显示	操作
暂无参数							

添加参数

环境变量	变量	值	操作
暂无参数			

执行镜像为评估程序运行的镜像，评估脚本为评估程序。评估命令为运行评估命令。执行器引擎为 Python 或 shell。参数为评估程序的参数配置。

## 8.5. 转换任务

➤ 点击模型库-->转换任务可以查看到模型进行的转换任务；



中科弘云 深度学习计算服务平台 文件管理 镜像管理 99%

模型库 > 转换任务

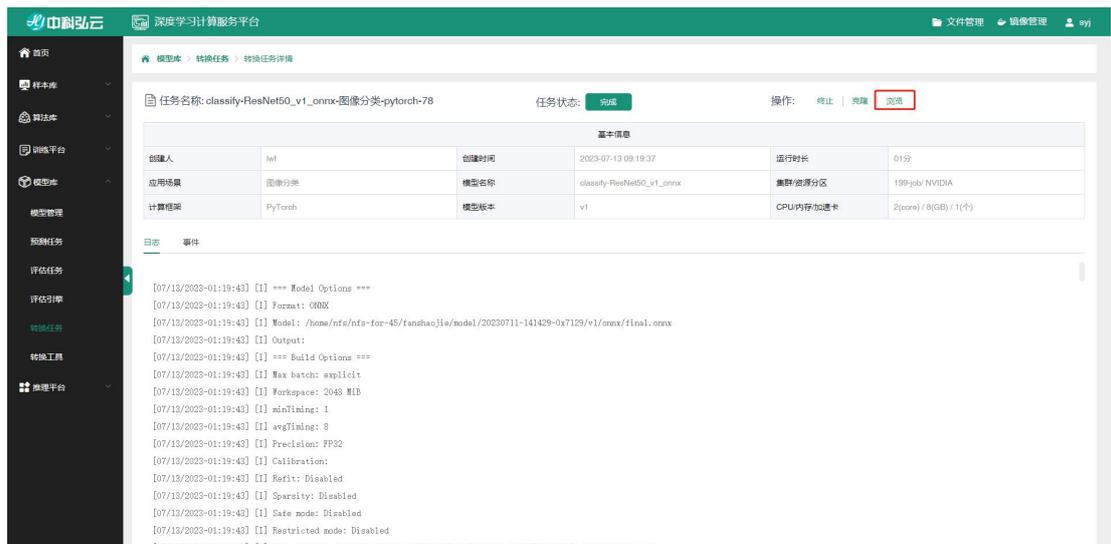
任务名称:  运行状态:  计算框架:  应用场景:

序号	任务名称	状态	模型名称	模型版本	创建人	创建时间	计算框架	操作
1	classify-FeeNet50_v1_onnx-图像分类-pyt...	完成	classify-FeeNet50_v1_onnx	v1	lwf	2023-07-13 09:19:37	PyTorch	克隆   停止   删除
2	classify-FeeNet50_v1_onnx-图像分类-pyt...	完成	classify-FeeNet50_v1_onnx	v1	lu001	2023-07-12 10:20:42	PyTorch	克隆   停止   删除
3	FE目标检测_v1_ckpt-目标检测-mindspore...	错误	FE目标检测_v1_ckpt	v1	lwf	2023-07-11 17:36:00	MindSpore	克隆   停止   删除
4	lu-object-detection_v2_onnx-目标检测-pyt...	错误	lu-object-detection_v2_on...	v2	lu001	2023-07-06 19:59:43	PyTorch	克隆   停止   删除
5	FE目标检测_v1_ckpt-目标检测-mindspore...	停止	FE目标检测_v1_ckpt	v1	slf	2023-07-06 17:35:26	MindSpore	克隆   停止   删除
6	ldg-autodi-class_v1_pth-图像分类-pytorch...	停止	ldg-autodi-class_v1_pth	v1	lu1234	2023-07-06 14:15:46	PyTorch	克隆   停止   删除
7	ldg-autodi-class_v1_pth-图像分类-pytorch...	停止	ldg-autodi-class_v1_pth	v1	lu1234	2023-07-06 14:15:45	PyTorch	克隆   停止   删除

共 7 条

Copyright © 2023 中科弘云科技（北京）有限公司 All rights reserved.

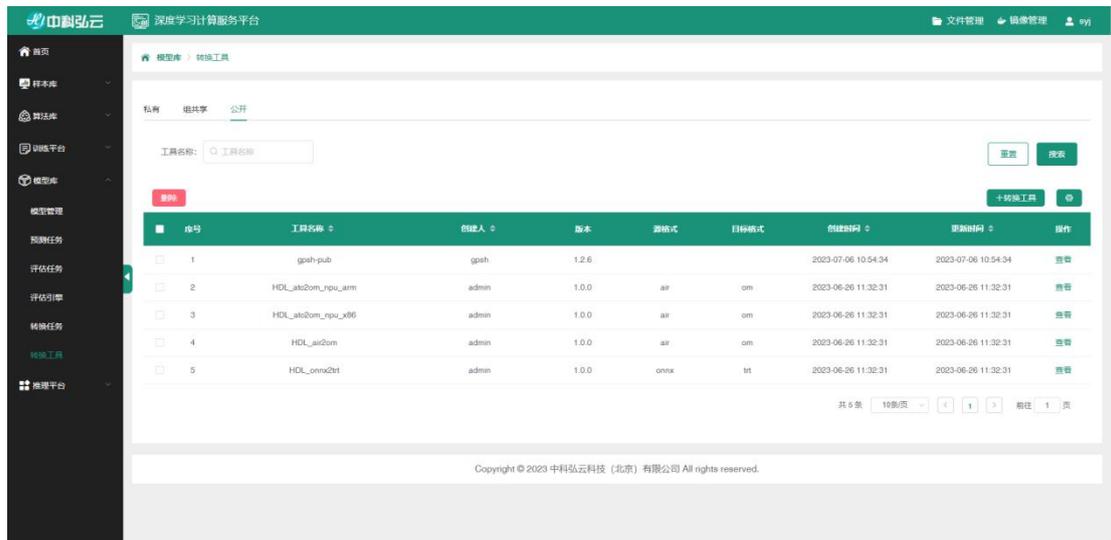
- 点击转换完成的任务可以查看转换的日志，点击浏览功能查看转换完成的模型；



转换成功在对应的模型仓库内也能看到完成的模型；

## 8.6. 转换工具

- 点击模型库-->转换工具可以查看平台内置的转化工具；



转换任务都是由转换工具进行转换的；

## 9. 推理平台

推理平台提供将模型发布成推理服务，并监控推理服务。推理平台分为云端服务、运行环境、边缘服务、边缘设备四个模块；

云端服务：纳管发布在云端的在线服务；

运行环境：纳管云端服务的运行环境；

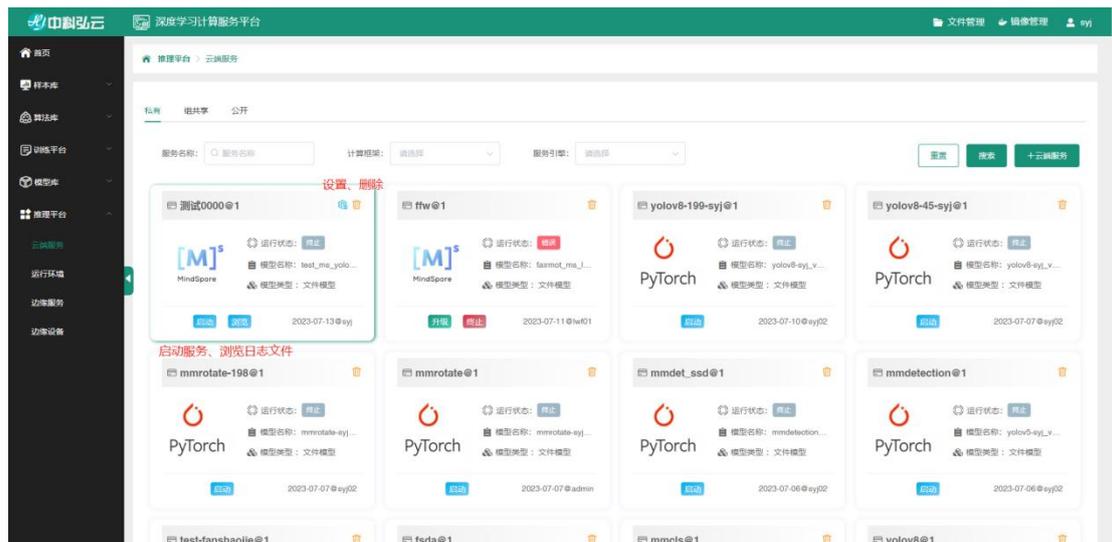
边缘服务：纳管发布在边缘设备的在线服务（以边缘服务维度展示）；

边缘设备：纳管发布在边缘设备的在线服务（以边缘设备维度展示）；

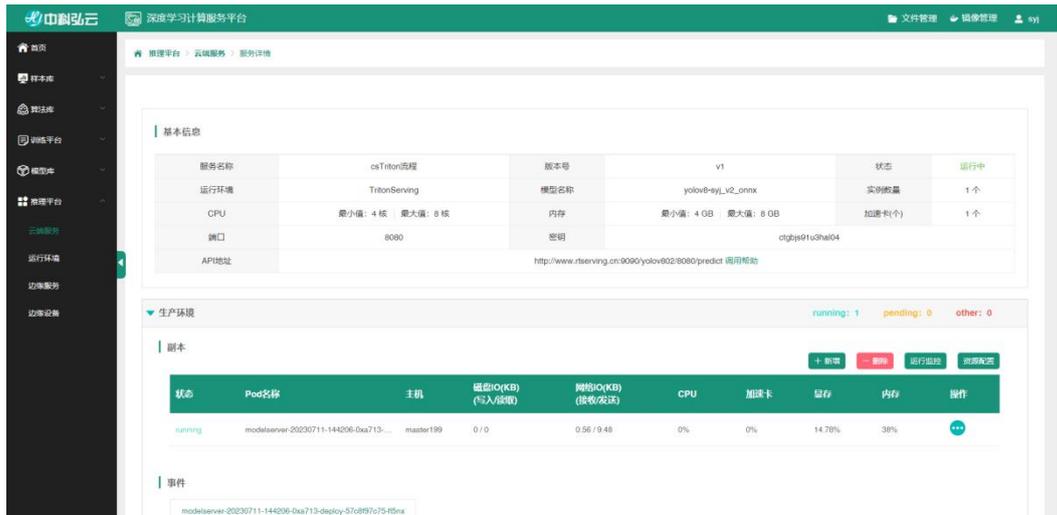
### 9.1. 云端服务

平台支持多种场景、多种框架不同算法的服务，使用详情见《中科弘云-深度学习计算服务平台-HyperDL-V4.5-服务使用说明.docx》；

- 点击推理平台-->云端服务可以查看到发布成功的服务；

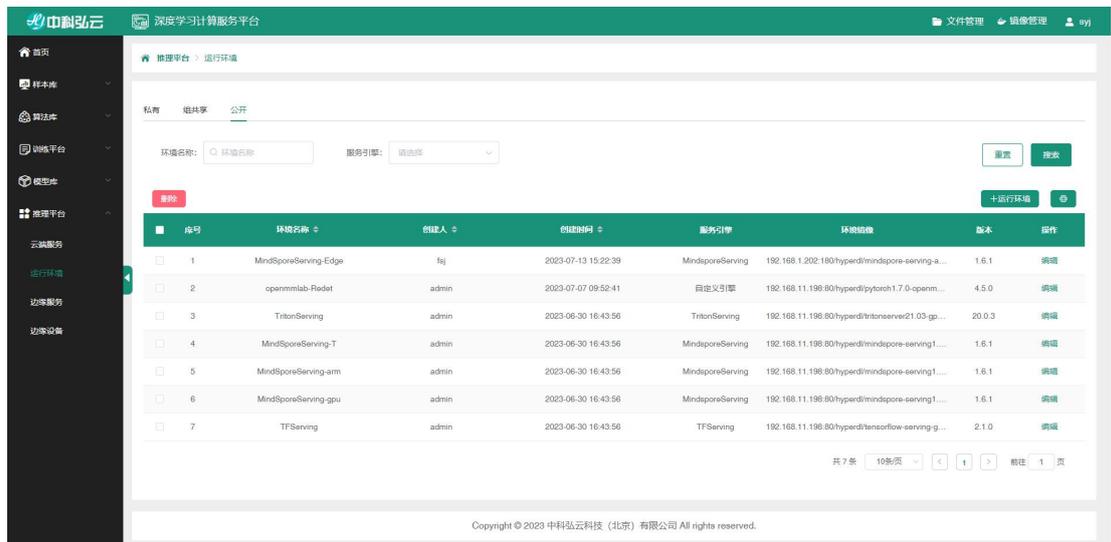


- 点击服务页面进入服务详细信息；



## 9.2. 运行环境

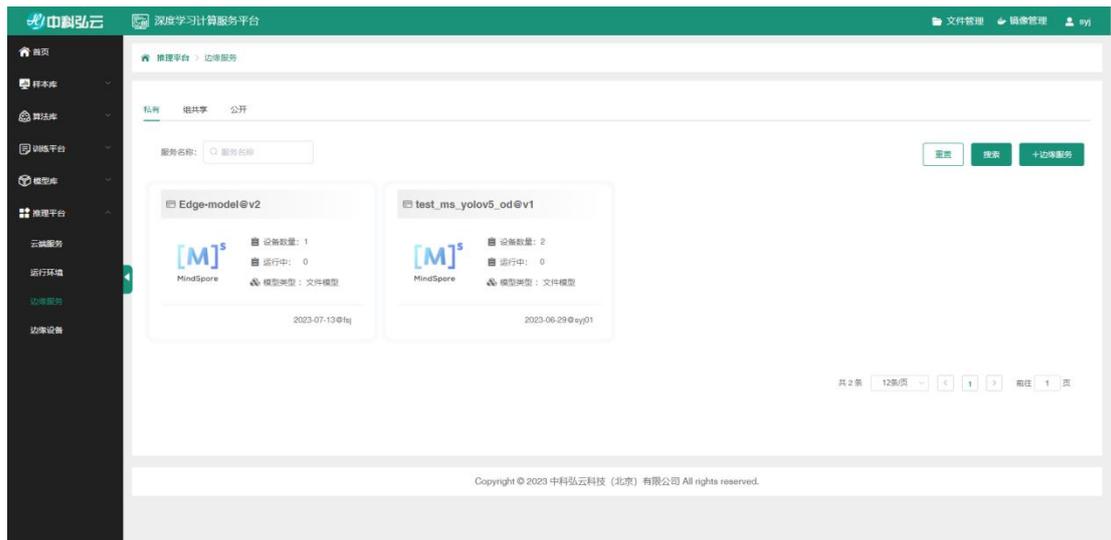
- 点击推理平台-->运行环境可以平台内置的服务环境；



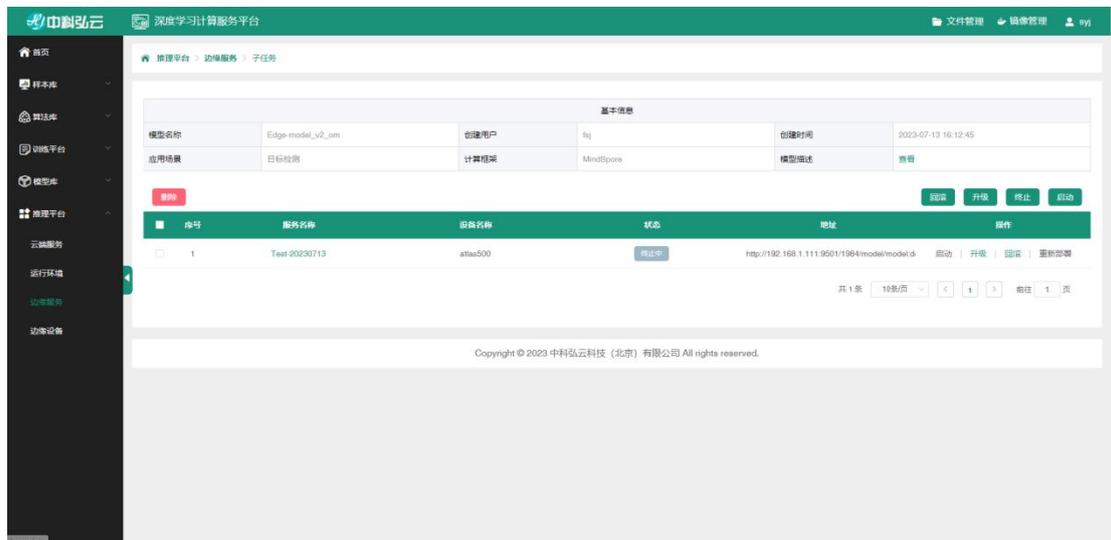
平台内置了 MindSpore-GPU、NPU 服务、TensorFlow 服务、TriTOn 服务；同时也支持用户添加自己的环境；

## 9.3. 边缘服务

- 点击推理平台-->边缘服务可以查看运行的边缘服务；

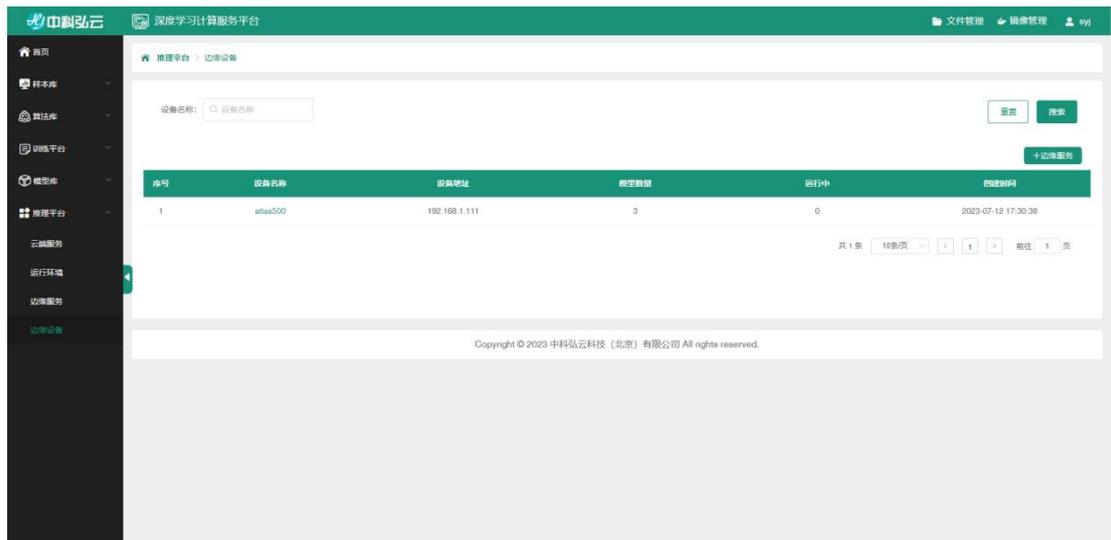


➤ 点击服务页面查看服务运行详情；

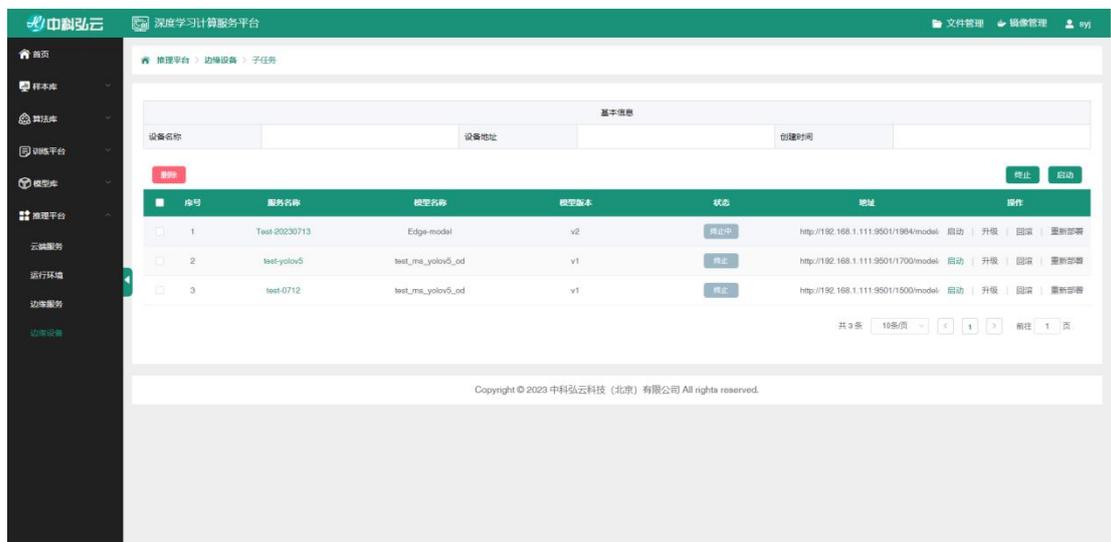


## 9.4. 边缘设备

➤ 点击推理平台-->边缘设备可以查看已有的边缘设备；



➤ 点击边缘设备查看跑在上面的服务：



## 10. 流程化教程

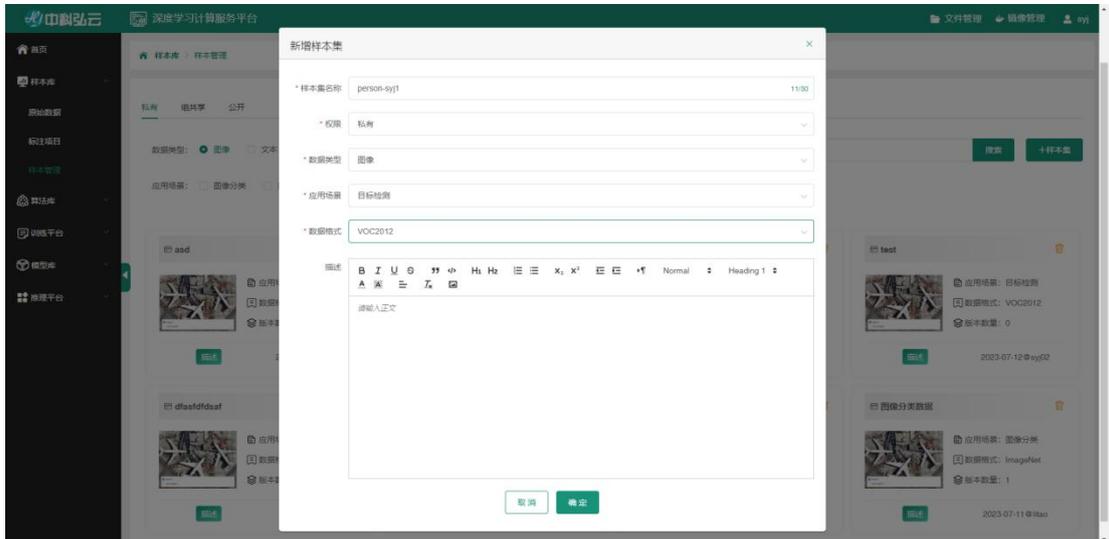
以下教程以管理员用户、目标检测场景、外部导入数据为示例进行使用演示：  
流程：

- 1、新建一个目标检测样本仓库；
- 2、外部导入一份目标检测数据集；
- 3、使用预置算法训练导入的数据集；
- 4、查看算法训练详细情况；
- 5、发布模型到模型仓库；

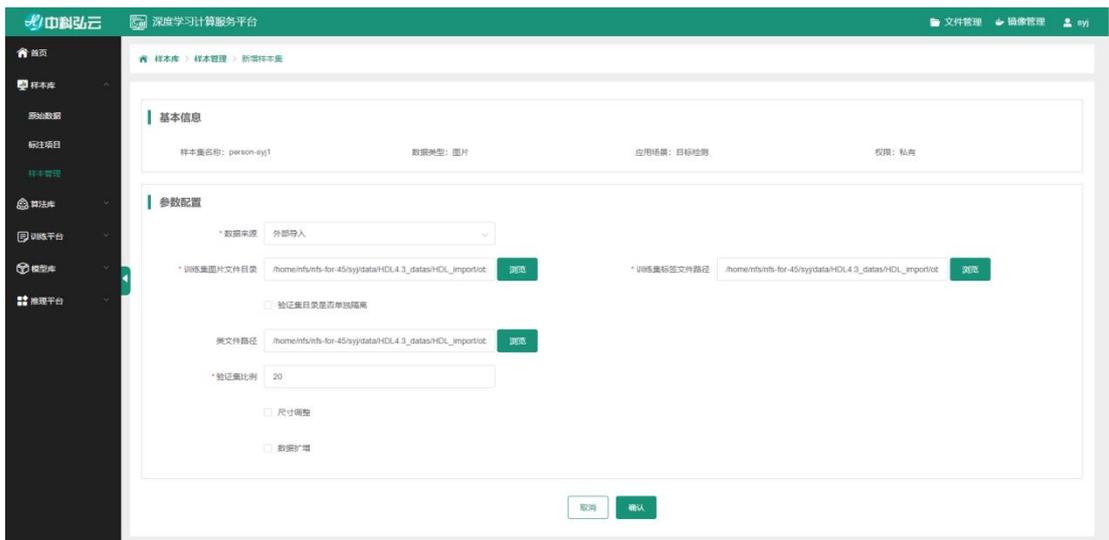


- 6、对模型进行预测、评估；
- 7、发布模型服务并在线进行调用；

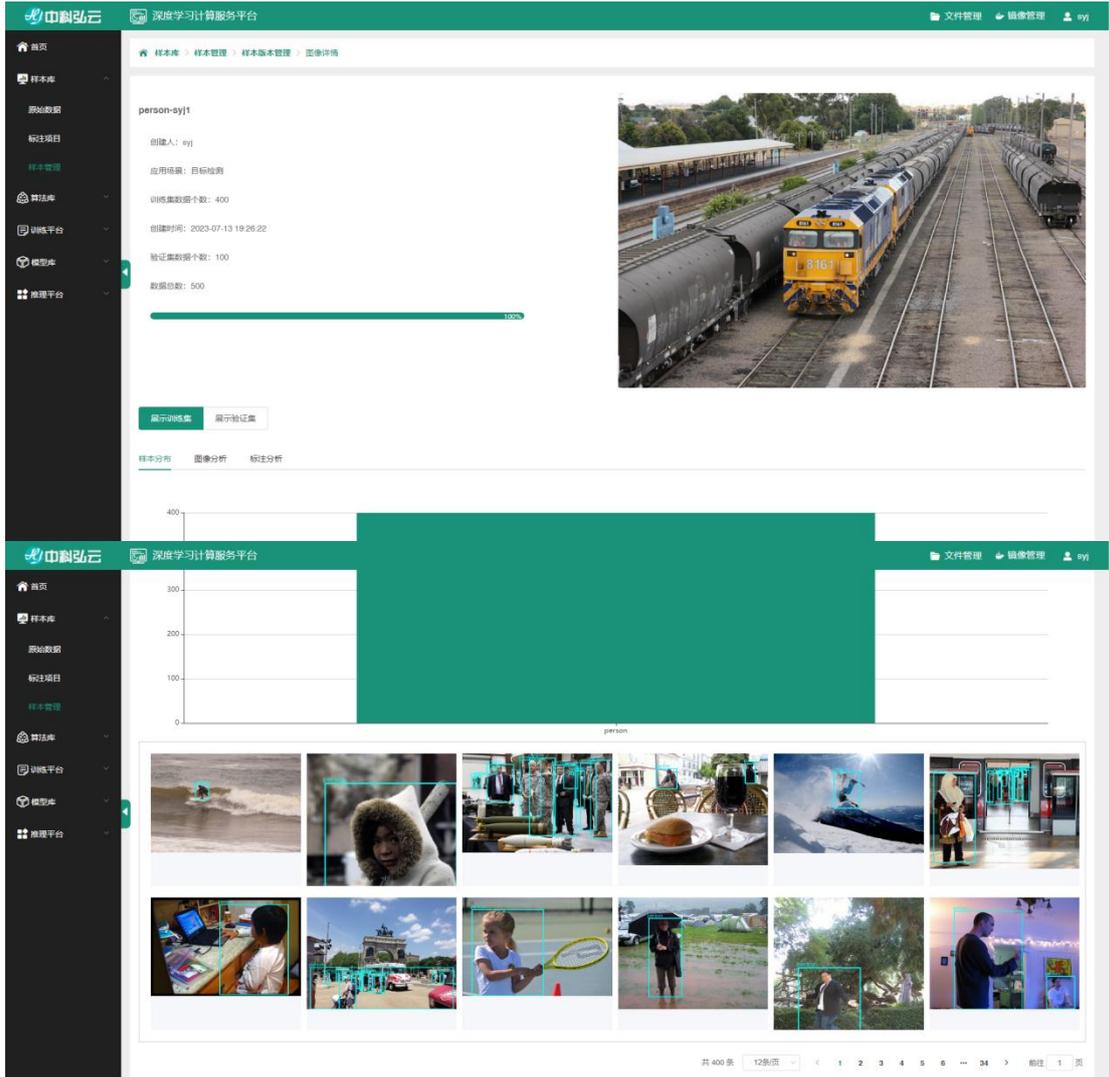
➤ 首先在样本管理新建一个目标检测样本仓库；



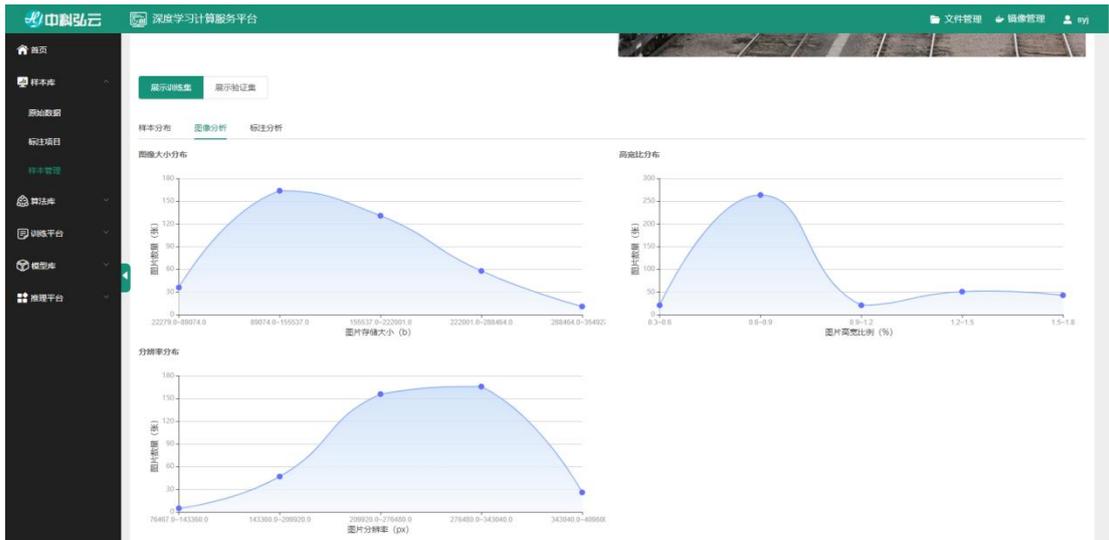
➤ 外部导入一个目标检测数据集，导入参考 [4.3.2. 外部数据导入](#)；



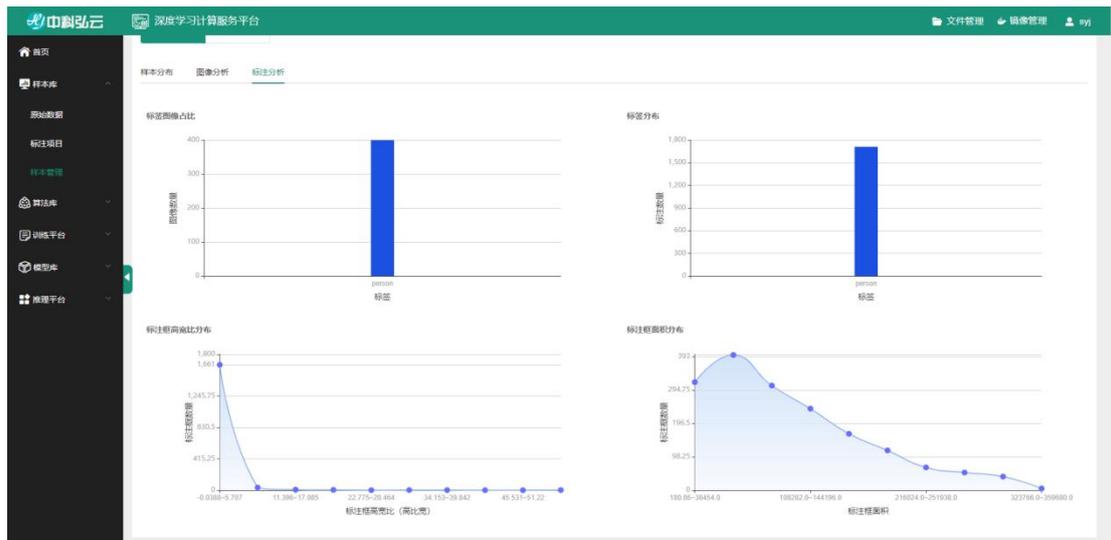
➤ 导入完成点击版本号可查看详情；



➤ 查看图像分析；



➤ 查看标注分析；



➤ 在预置算法-->计算机视觉-->目标检测中选择一个算法点击训练；

➤ 选择刚导入的数据集、修改参数、配置资源进行提交；

➤ 待算法运行起来点击进去可查看日志、参数、事件、图表等信息；

训练完成后，可查看日志来判断算法训练情况：

资源	显卡	使用率	使用率	温度
1	nvidia	加速卡	显存	温度 42.70

➤ 训练完成后可查看图表、日志来判断算法训练情况；

创建人	创建时间	运行时长
hu1234	2023-07-06 19:05:01	48分06秒

资源分区	CPU内存加速卡	计算结果
199-job/ NVIDIA	4(core) / 16(GB) / 1(个)	PyTorch

应用场景	算法名称	样本集名称
目标检测	@ 目标检测_fasterronn	@ lsg_目标检测_飞机v1

训练完成后的图表显示了各项指标的收敛情况：

- RPN位置损失: 0.074
- RPN分类损失: 0.06
- 物体损失: 0.494
- 类别损失: 0.239
- 总损失: 0.866
- 平均精确度: 0.758

创建人	创建时间	运行时长
hu1234	2023-07-06 19:05:01	48分06秒

资源分区	CPU内存加速卡	计算结果
199-job/ NVIDIA	4(core) / 16(GB) / 1(个)	PyTorch

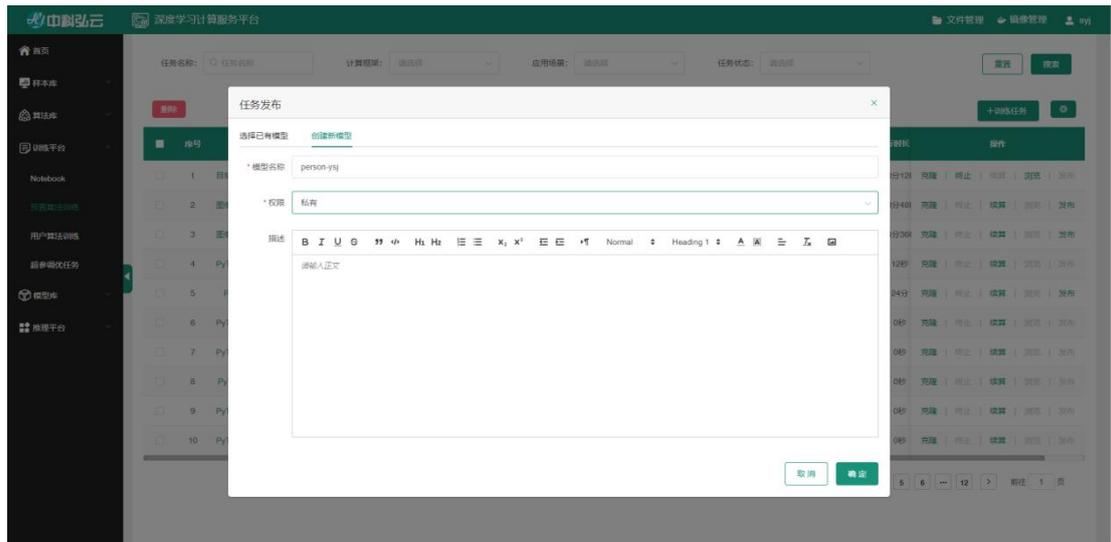
应用场景	算法名称	样本集名称
目标检测	@ 目标检测_fasterronn	@ lsg_目标检测_飞机v1

训练完成后的日志显示了详细的性能指标：

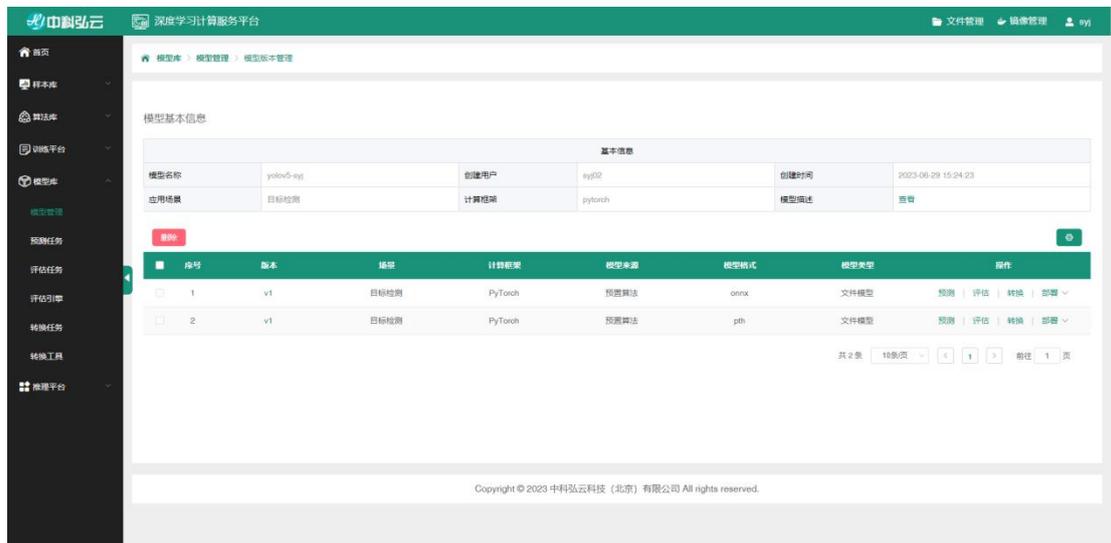
```

Average Precision (AP) @[ IoU=0.50:0.95 | area= all | maxDets=100 ] = 0.38875
Average Precision (AP) @[ IoU=0.50 | area= all | maxDets=1000 ] = 0.76173
Average Precision (AP) @[ IoU=0.75 | area= all | maxDets=1000 ] = 0.29938
Average Precision (AP) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.07543
Average Precision (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.41177
Average Precision (AP) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.53064
Average Recall (AR) @[ IoU=0.50:0.95 | area= all | maxDets=100 ] = 0.45637
Average Recall (AR) @[ IoU=0.50:0.95 | area= all | maxDets=200 ] = 0.45637
Average Recall (AR) @[ IoU=0.50:0.95 | area= all | maxDets=1000 ] = 0.45637
Average Recall (AR) @[ IoU=0.50:0.95 | area= small | maxDets=1000 ] = 0.19740
Average Recall (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=1000 ] = 0.50090
Average Recall (AR) @[ IoU=0.50:0.95 | area= large | maxDets=1000 ] = 0.58354
    
```

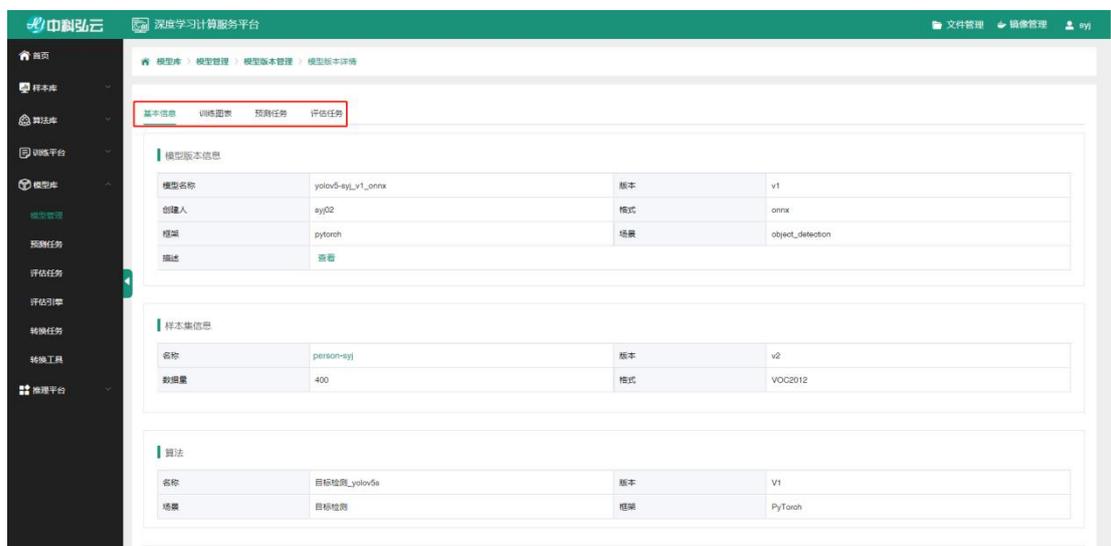
- 模型训练完成点击发布，把模型发布到模型仓库进行使用；



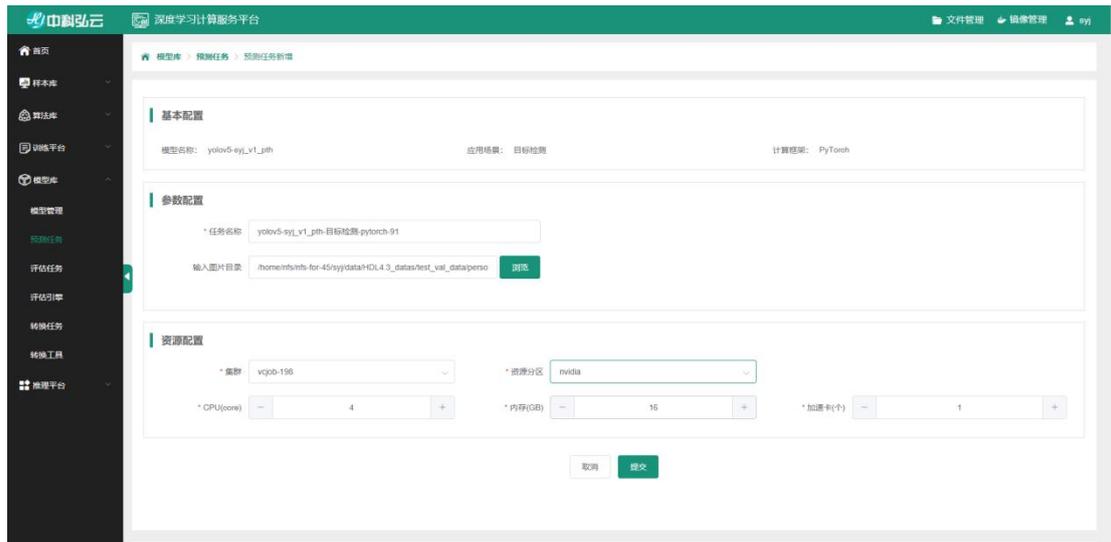
- 点击模型页查看导入的模型；



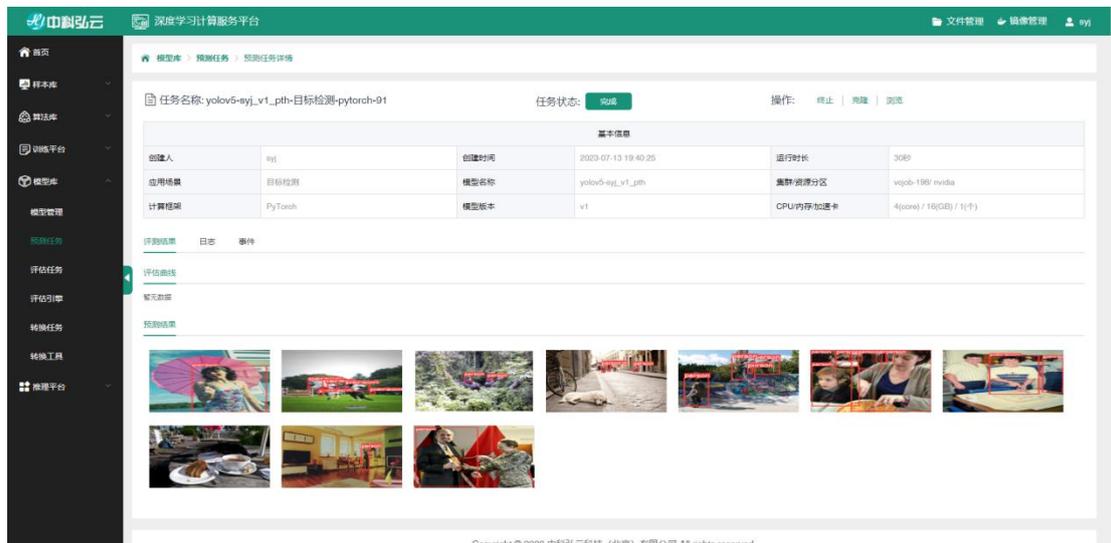
- 点击模型版本号可以查看训练信息、样本信息、训练图表等等信息；



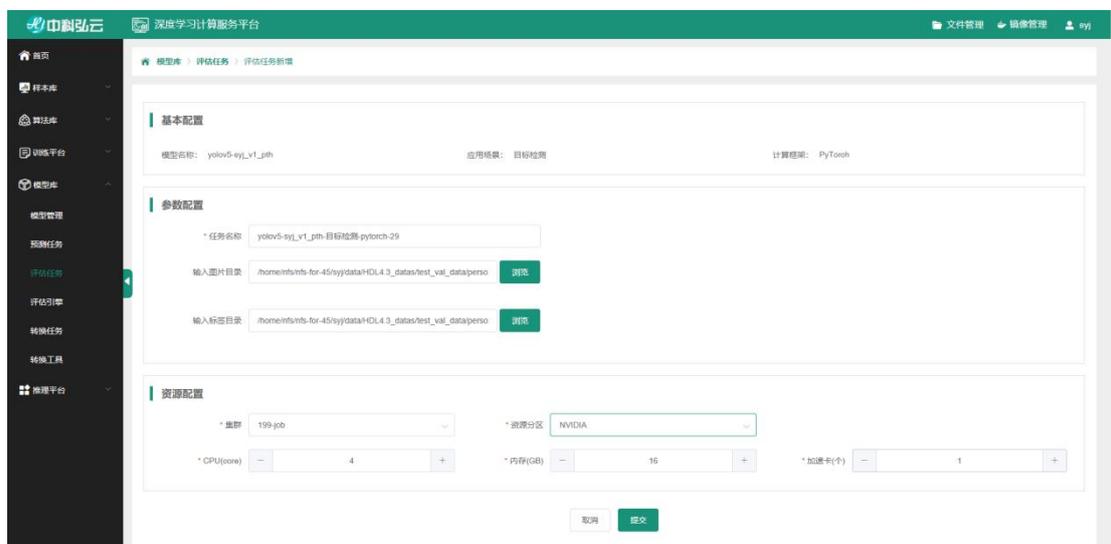
➤ 点击预测选择要测试的数据集进行预测；



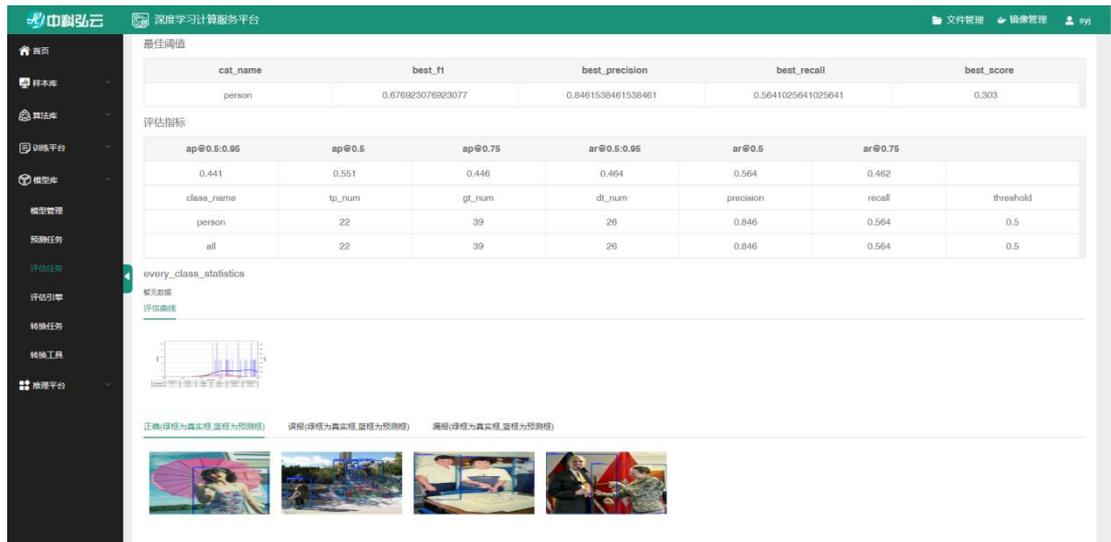
➤ 预测完成点击任务查看详情；



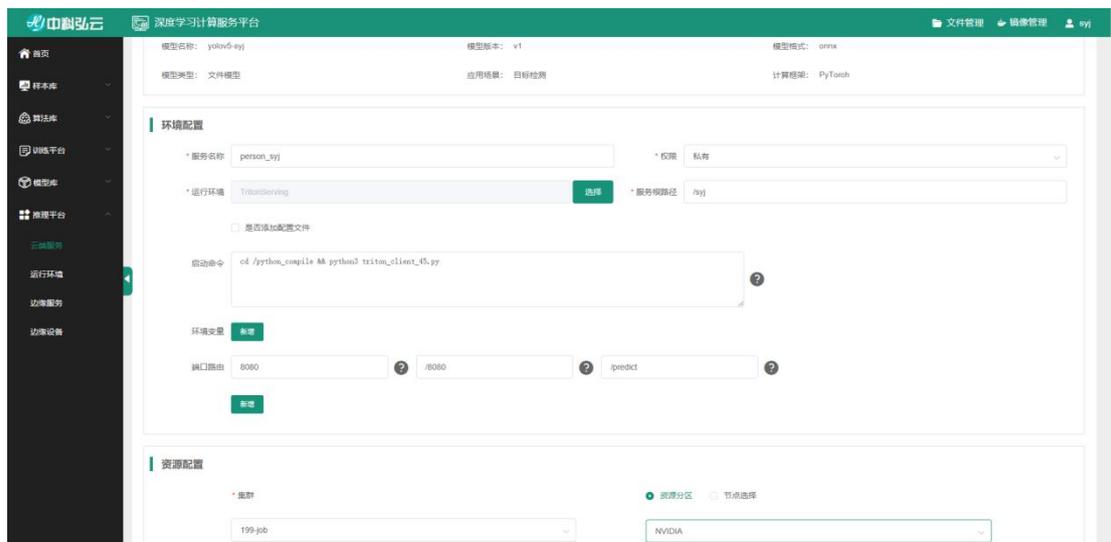
➤ 返回模型仓库，点击评估，选择需要评估的数据和真实标签文件进行评估；



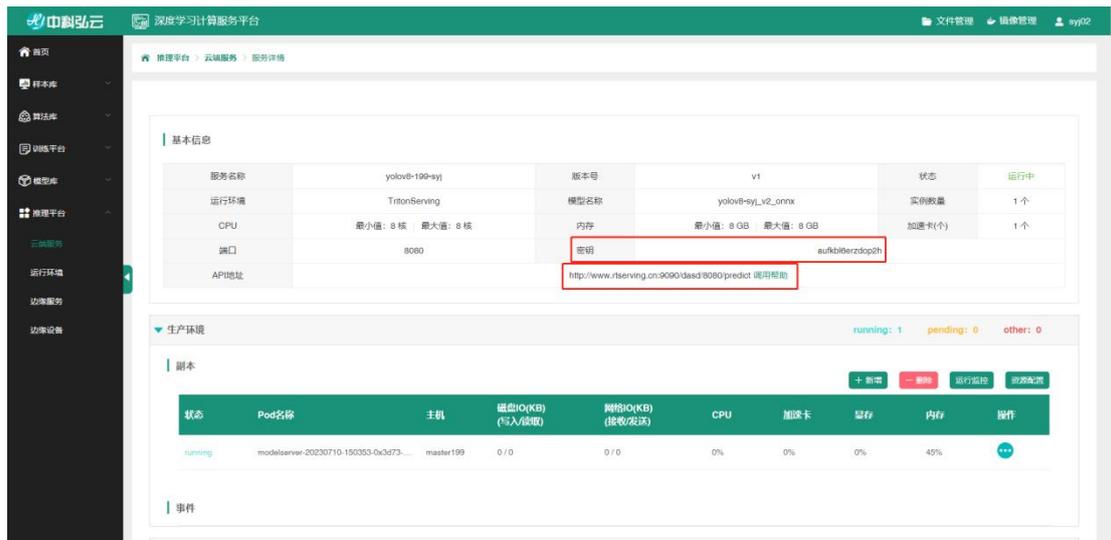
➤ 评估完成点击任务查看详情；



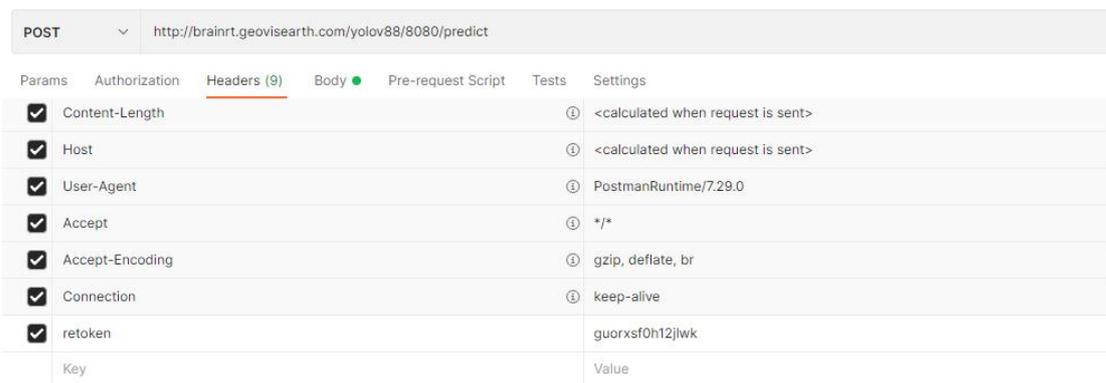
➤ 返回模型仓库，点击部署选择云端服务，填写参数选择运行环境，进行发布；  
（详细发布调用见《中科弘云-深度学习计算服务平台-HyperDL-V4.5-服务使用说明.docx》文档）；



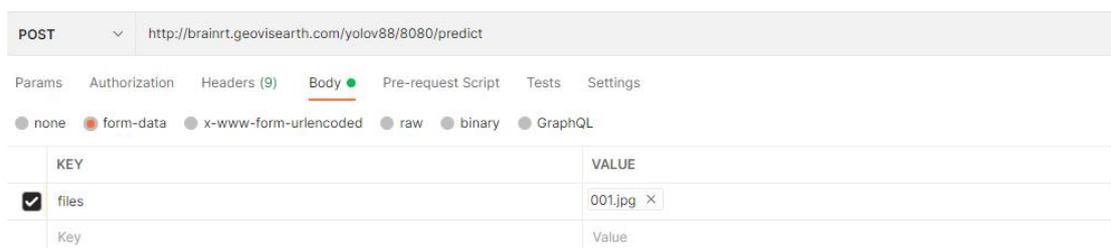
➤ 启动后点击查看服务详情；



- 使用 Postman 调用测试服务能否使用；
- 打开软件填写参数----→headers 下填写 retoken----→guorxsf0h12jlwk（密钥）；



- 选择 Body 下的 form-data 下填写 files 选择图像进行上传；



- 选择 post 请求进行调用，返回识别结果；



http://brainrt.geovisearth.com/yolov88/8080/predict

Save

POST http://brainrt.geovisearth.com/yolov88/8080/predict Send

Params Authorization Headers (9) Body Pre-request Script Tests Settings

none form-data x-www-form-urlencoded raw binary GraphQL

KEY	VALUE	DESCRIPTION	***	Bulk Edit
<input checked="" type="checkbox"/>	files	001.jpg		
Key	Value	Description		

Body Cookies Headers (6) Test Results

Status: 200 OK Time: 1506 ms Size: 318 B Save Response

Pretty Raw Preview Visualize JSON

```
1  {
2    "code": 200,
3    "result": {
4      "instances": [
5        {
6          "result": "[['aircraft', 0.79, [588.36, 634.4, 478.74, 678.77]]]"
7        }
8      ]
9    },
10   "run_time": 1.12449849949646
11 }
```