

# **codeExplorer 使用说明书**

最后修订时间: 25.10.10

# 目录

- 1 介绍 ..... 1
- 2 主要功能 ..... 1
  - 2.1 模型仓库 ..... 1
  - 2.2 代码数据集 ..... 2
  - 2.3 任务列表 ..... 4
    - 2.3.1 预处理任务列表 ..... 4
    - 2.3.2 训练任务列表 ..... 6
    - 2.3.3 测试任务列表 ..... 8

# 1 介绍

CodeExplorer 是一个汇集代码扫描工具的软件平台。使用这些代码扫描工具时，需要上传待分析的代码库，然后依次进行代码预处理、模型训练、模型使用和结果可视化等多个步骤（可分离使用）。通过扫描代码，可以发现隐藏在代码中的缺陷（如代码克隆、特征嫉妒等），用于评估和改善软件质量。

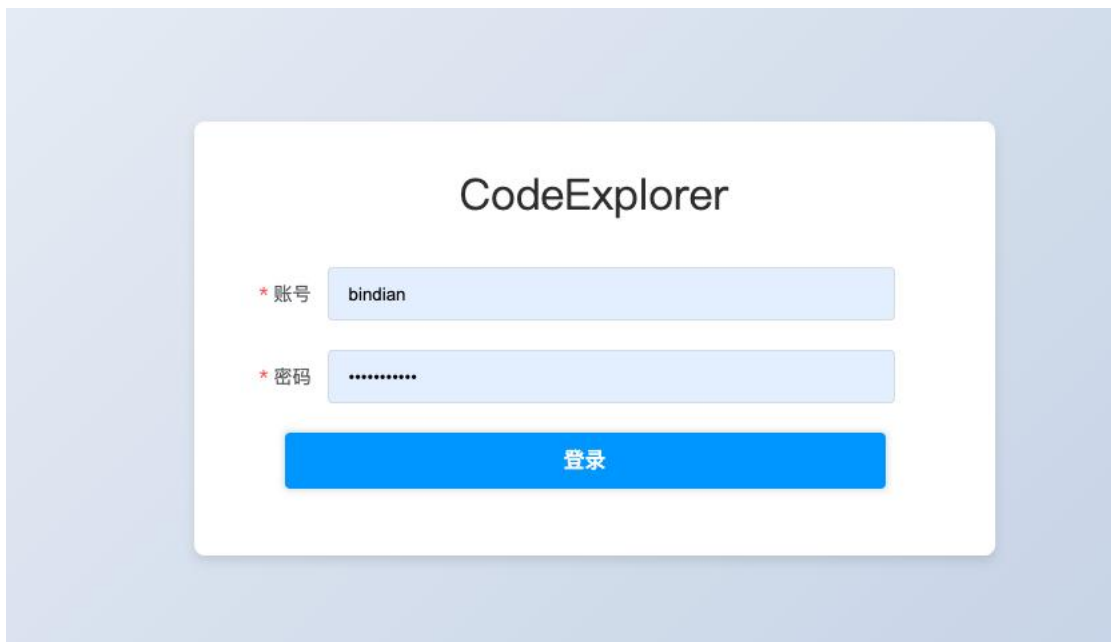
本文旨在帮助读者快速了解网页使用，详细介绍各部分具体内容及功能。接下来，将按照模型仓库、代码数据集、预处理任务列表、训练任务列表、测试任务列表的顺序依次进行介绍。

## 2 主要功能

### 2.1 登录

账号：bindian

密码：bindian2024



### 2.2 模型仓库

模型仓库用于展示当前所有的模型信息，除了表格中展示的基本信息外，点击“查看”按钮可以看到模型训练时设置的参数。点击“删除”按钮可以删除所选模型，**注：此操作不可逆**。模型会在训练任务结束时自动生成，用户不可自行导入，在测试阶段可选表格中任一模型进行代码测试。

CodeExplorer

模型仓库

模型仓库

代码数据集

任务列表

编号	名称	模型	模型类型	描述	创建日期	操作
3	孪生图匹配网络1	CodeGraph4CCDetector	代码克隆检测	一种基于 CFG/PDG 注意力机制的孪生图匹配网络，它可以在 token、语句和图级别捕获关键信息，以充分学习代码语义。	2024-10-28 14:29:37	<a href="#">查看</a> <a href="#">删除</a>
63	测试训练	测试训练	代码克隆检测	模型描述	2024-11-01 10:07:15	<a href="#">查看</a> <a href="#">删除</a>
80	任务名称	输出模型名称	代码克隆检测	模型描述	2024-11-04 13:56:08	<a href="#">查看</a> <a href="#">删除</a>
125	任务名称1	输出模型名称	代码克隆检测	模型描述	2024-12-18 23:04:22	<a href="#">查看</a> <a href="#">删除</a>
126	任务名称	输出模型名称	代码克隆检测	模型描述	2024-12-19 10:54:16	<a href="#">查看</a> <a href="#">删除</a>
127	任务名称1	输出模型名称	代码克隆检测	模型描述	2024-12-25 15:27:17	<a href="#">查看</a> <a href="#">删除</a>
129	任务名称1	输出模型名称1	代码克隆检测	模型描述	2024-12-26 13:36:20	<a href="#">查看</a> <a href="#">删除</a>
130	任务名称1	输出模型名称	代码克隆检测	模型描述	2025-01-20 22:04:30	<a href="#">查看</a> <a href="#">删除</a>
131	任务名称1	输出模型名称	代码克隆检测	模型描述	2025-01-20 22:05:51	<a href="#">查看</a> <a href="#">删除</a>

共 1000 条123456...10> 前往1页

模型仓库列表

CodeExplorer

模型仓库

模型仓库

代码数据集

任务列表

预处理任务列表

训练任务列表

代码克隆检测

代码异味检测

代码漏洞检测

测试任务列表

代码克隆检测

代码异味检测

代码漏洞检测

编号	名称	模型	模型类型	描述	创建日期	操作
3	孪生图匹配网络1	CodeGraph4CCDetector	代码克隆检测	一种基于 CFG/PDG 注意力机制的孪生图匹配网络，它可以在 token、语句和图级别捕获关键信息，以充分学习代码语义。	2024-10-28 14:29:37	<a href="#">查看</a> <a href="#">删除</a>
63	测试训练	测试训练	代码克隆检测	模型描述	2024-11-01 10:07:15	<a href="#">查看</a> <a href="#">删除</a>
80	任务名称	输出模型名称	代码克隆检测	模型描述	2024-11-04 13:56:08	<a href="#">查看</a> <a href="#">删除</a>
125	任务名称1	输出模型名称	代码克隆检测	模型描述	2024-12-18 23:04:22	<a href="#">查看</a> <a href="#">删除</a>
126	任务名称	输出模型名称	代码克隆检测	模型描述	2024-12-19 10:54:16	<a href="#">查看</a> <a href="#">删除</a>
127	任务名称1	输出模型名称	代码克隆检测	模型描述	2024-12-25 15:27:17	<a href="#">查看</a> <a href="#">删除</a>
129	任务名称1	输出模型名称1	代码克隆检测	模型描述	2024-12-26 13:36:20	<a href="#">查看</a> <a href="#">删除</a>
130	任务名称1	输出模型名称	代码克隆检测	模型描述	2025-01-20 22:04:30	<a href="#">查看</a> <a href="#">删除</a>
131	任务名称1	输出模型名称	代码克隆检测	模型描述	2025-01-20 22:05:51	<a href="#">查看</a> <a href="#">删除</a>

共 1000 条123

模型训练参数

任务名称任务名称1

输出模型名称输出模型名称

任务类型代码克隆检测

数据集googlejam4\_src

模型描述模型描述

参数设置

epochs100

batch\_size64

学习率0.0001

学习衰减因子0.5

分类数2

学习率衰减步长50

关闭

模型仓库“查看”按钮

2.3 代码数据集

代码数据集用于展示当前所有的数据集信息，包括已预处理数据集和未预处理数据集。同样的，用户除了可以看到基本信息外，点击“查看”按钮可以看到此数据集中的具体内容。点击“新增原始数据集”可将本地数据上传至网站，用于检测与评估。

CodeExplorer

首页 / 已预处理数据集

模型仓库

代码数据集

任务列表

代码数据集信息表

+ 新增原始数据集

编号	数据集名称	描述	状态	创建日期	操作
7	第二数据集		已处理	2011-11-23 13:02:47	<a href="#">查看</a> <a href="#">删除</a>
8	googlejam4_src		已处理	2011-11-23 13:02:47	<a href="#">查看</a> <a href="#">删除</a>
44	googlejam4_src	数据集描述	已处理	2024-12-19 13:45:21	<a href="#">查看</a> <a href="#">删除</a>
45	googlejam4_src	数据集描述	已处理	2024-12-19 16:11:44	<a href="#">查看</a> <a href="#">删除</a>
54	googlejam4_src	数据集描述	已处理	2025-01-06 15:34:30	<a href="#">查看</a> <a href="#">删除</a>
71	test	数据集描述	已处理	2025-01-12 23:12:26	<a href="#">查看</a> <a href="#">删除</a>
83	googlejam4		未处理	2025-01-13 11:16:55	<a href="#">查看</a> <a href="#">删除</a>
84	googlejam4		未处理	2025-01-13 11:18:00	<a href="#">查看</a> <a href="#">删除</a>
85	test		未处理	2025-01-13 11:18:07	<a href="#">查看</a> <a href="#">删除</a>

共 1000 条 < 1 2 3 4 5 6 ... 10 > 前往 1 页

代码数据集按钮

CodeExplorer

首页 / 已预处理数据集

模型仓库

代码数据集

任务列表

代码数据集信息表

+ 新增原始数据集

编号	数据集名称	描述	状态	创建日期	操作
7	第二数据集		已处理	2011-11-23 13:02:47	<a href="#">查看</a> <a href="#">删除</a>
8	googlejam4_src		已处理	2011-11-23 13:02:47	<a href="#">查看</a> <a href="#">删除</a>
44	googlejam4_src	数据集描述	已处理	2024-12-19 13:45:21	<a href="#">查看</a> <a href="#">删除</a>
45	googlejam4_src	数据集描述	已处理	2024-12-19 16:11:44	<a href="#">查看</a> <a href="#">删除</a>
54	googlejam4_src	数据集描述	已处理	2025-01-06 15:34:30	<a href="#">查看</a> <a href="#">删除</a>
71	test	数据集描述	已处理	2025-01-12 23:12:26	<a href="#">查看</a> <a href="#">删除</a>
83	googlejam4		未处理	2025-01-13 11:16:55	<a href="#">查看</a> <a href="#">删除</a>
84	googlejam4		未处理	2025-01-13 11:18:00	<a href="#">查看</a> <a href="#">删除</a>
85	test		未处理	2025-01-13 11:18:07	<a href="#">查看</a> <a href="#">删除</a>

共 1000 条 < 1 2 3 4 5 6 ... 10 > 前往 1 页

数据集内容

文件结构:

谷歌代码:

源代码:

代码数据集“查看”按钮

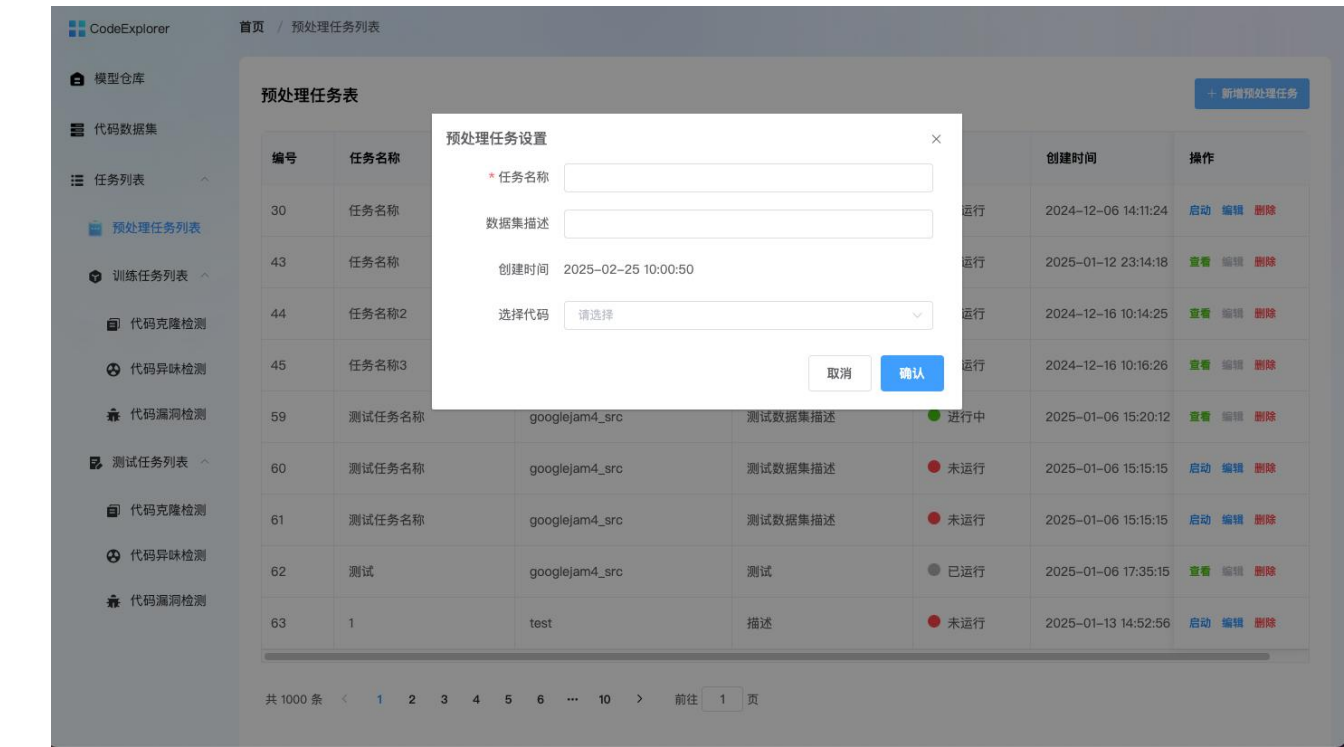
## 2.4 任务列表

任务分为三种：预处理任务、训练任务以及测试任务，三种任务相互独立，但 gpu、模型资源共享。

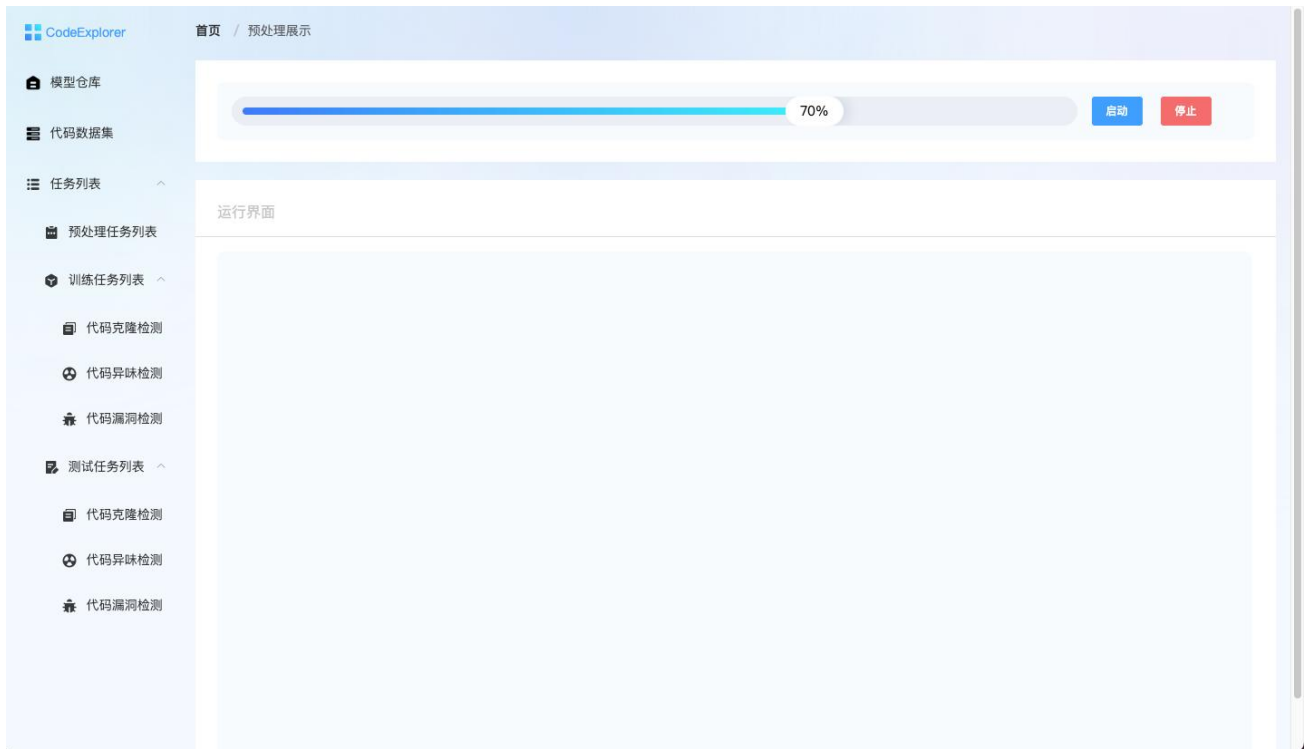
### 2.4.1 预处理任务列表

预处理任务会将未预处理的数据集变为已处理数据集。操作流程为：

1. 进到任务列表-预处理任务列表，点击“新增预处理任务”，填入信息，选择想要预处理的代码。

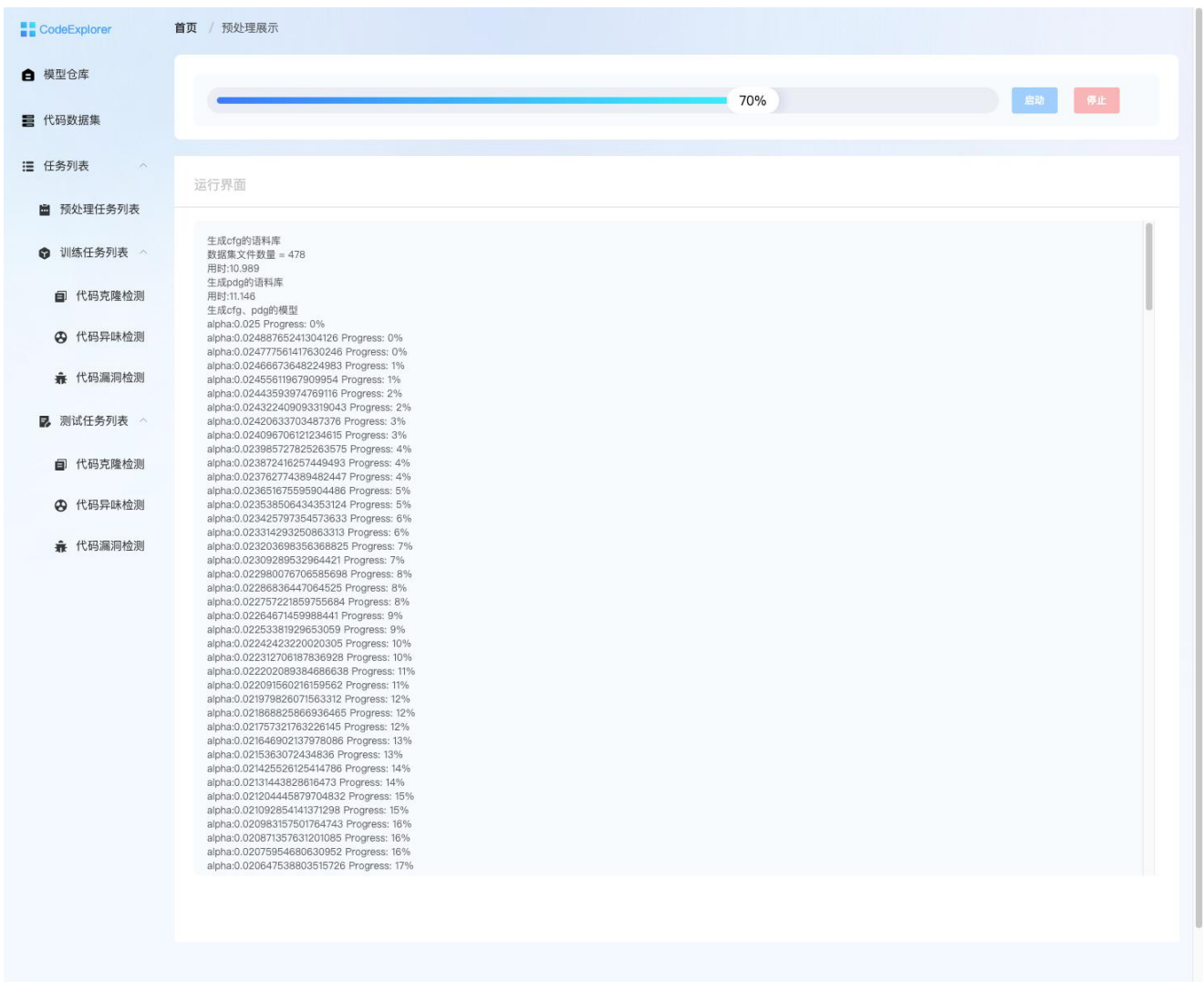


2. 新建完成后可在预处理任务表中看到，任务在未启动前都是可以进行编辑的。
3. 点击“启动”按钮，会跳转到预处理展示界面，点击启动按钮任务，选择空闲的 GPU 后开始执行任务，顶部的进度条会显示总体进行流程，并且用户可随时点击“停止”按钮停止任务。



预处理任务启动界面

4. 任务启动后运行界面会产生控制台信息，方便用户查看执行情况。



预处理任务运行界面

5. 启动成功后用户可随时点击该任务的“查看”按钮查看运行情况。

## 2.4.2 训练任务列表

训练任务有三种，分别是代码克隆检测、代码异味检测以及代码漏洞检测。这里将着重对代码克隆检测流程的描述。

训练任务会对选择的不同数据集训练出不同模型，操作流程为：

1. 进入任务列表-训练任务列表-代码克隆检测，点击“新增训练任务”，输入各项信息，并且可以自定义修改参数设置。注意：此处数据集必须是已处理好的数据集才可在此处展示。

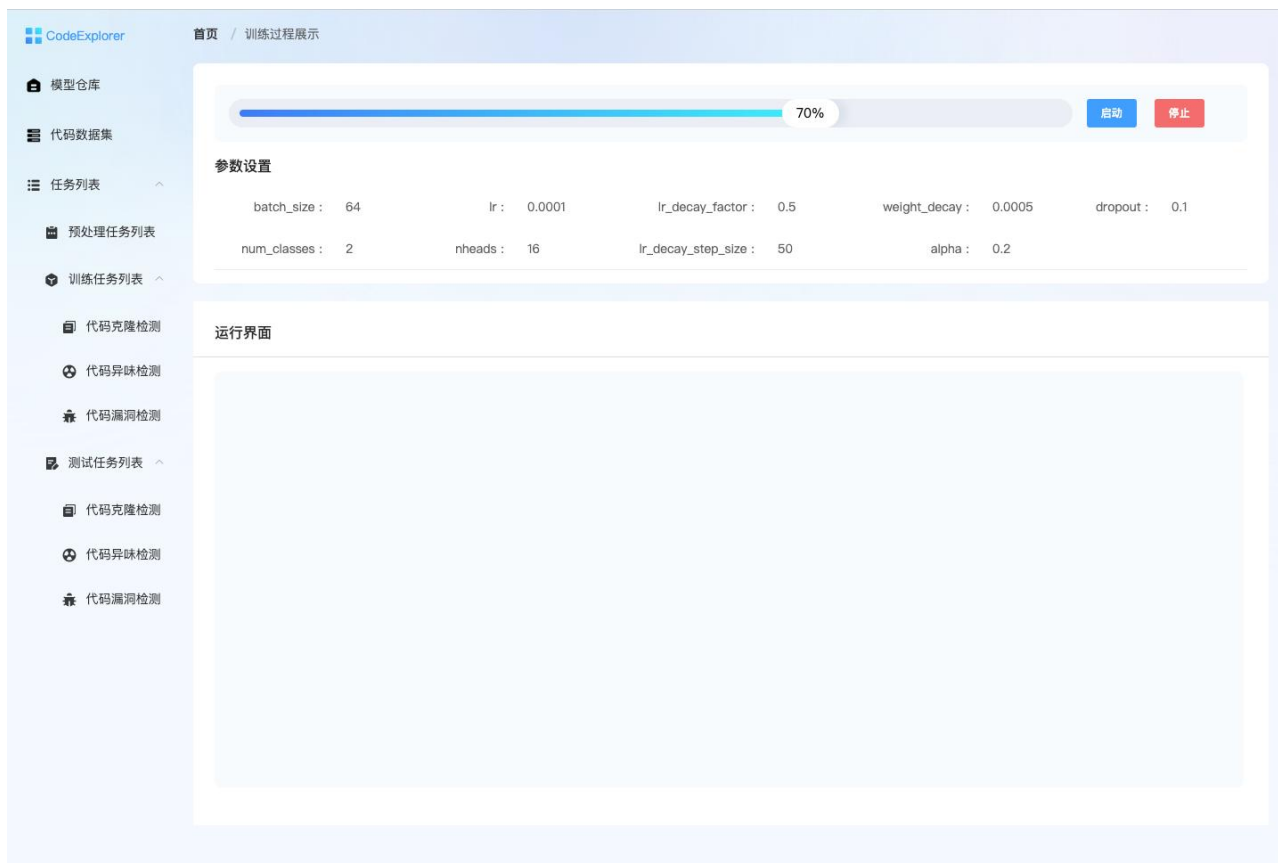


“新增训练任务”按钮界面

2. 新建完成后可在代码克隆检测任务表中看到，任务在未启动前都是可以进行编辑的。

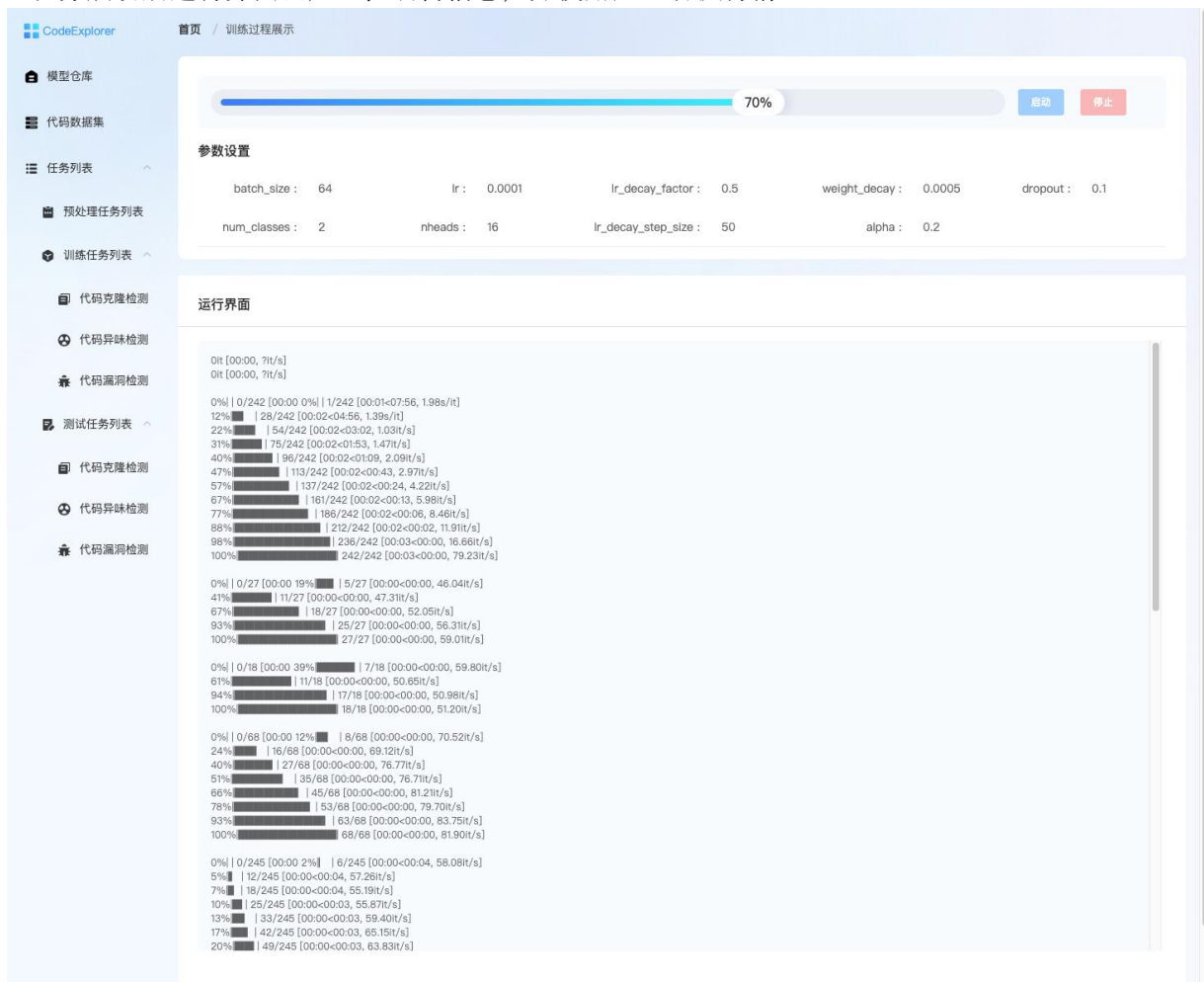
3. 点击“启动”按钮，会跳转到训练过程展示界面，点击启动按钮任务，选择空闲的 GPU 后开始执行任务，顶部的进度条会显示总体进行流程，并且用户可随时点击“停止”按钮停止任务。





训练任务“启动”按钮界面

4. 任务启动后运行界面会产生控制台信息，方便用户查看执行情况。



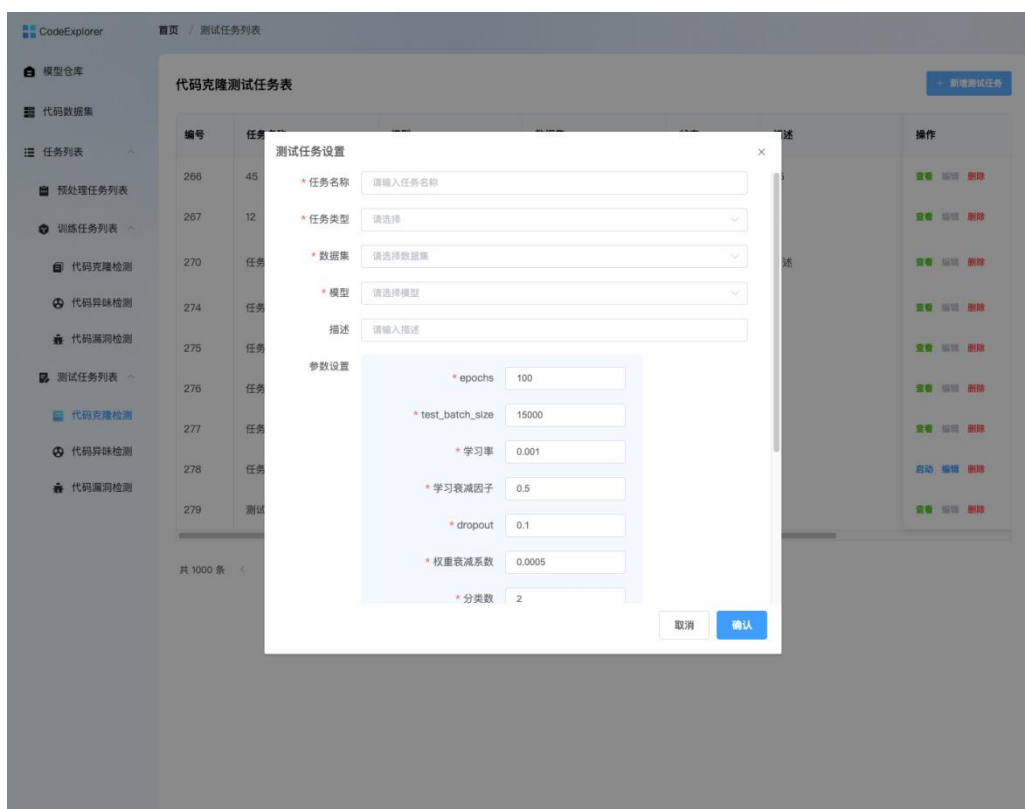
5. 启动成功后用户可随时点击该任务的“查看”按钮查看运行情况。

## 2.4.3 测试任务列表

对应训练任务列表，测试任务同样有三种，这里将着重进行代码克隆检测流程的描述。

测试任务会对生成的模型与用户选择的代码进行检测，分析与评估模型的好坏。操作流程为：

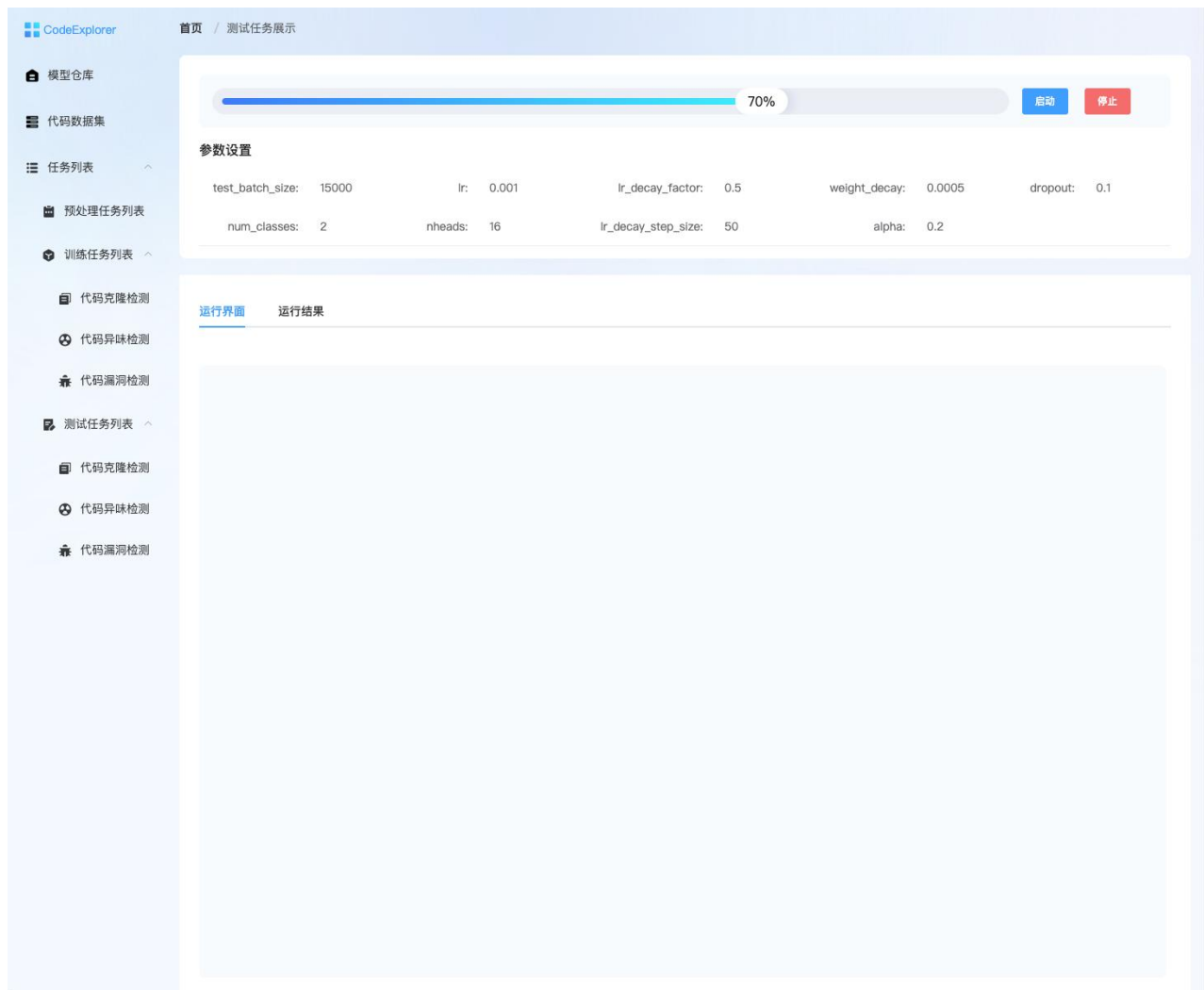
1. 进入任务列表-测试任务列表-代码克隆检测，点击“新增测试任务”，输入各项信息，并且可以自定义修改参数设置。注意：此处数据集必须是已处理好的数据集才可在此处展示。



“新增测试任务”界面

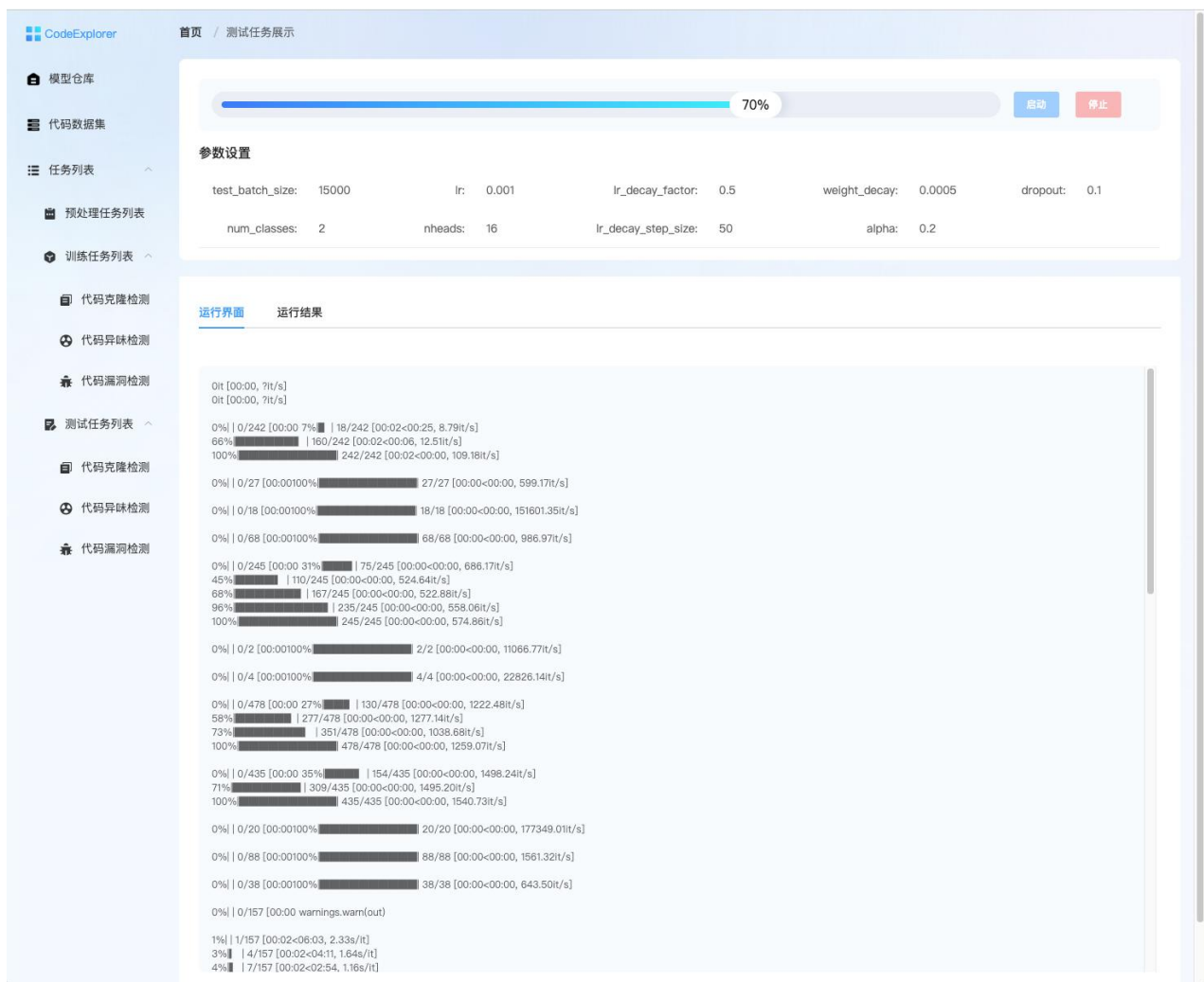
2. 新建完成后可在代码克隆测试任务表中看到，任务在未启动前都是可以进行编辑的。

3. 点击“启动”按钮，会跳转到测试过程展示界面，同样的，点击启动按钮任务，选择空闲的 GPU 后开始执行任务，顶部的进度条会显示总体进行流程，并且用户可随时点击“停止”按钮停止任务。



测试任务“启动”按钮

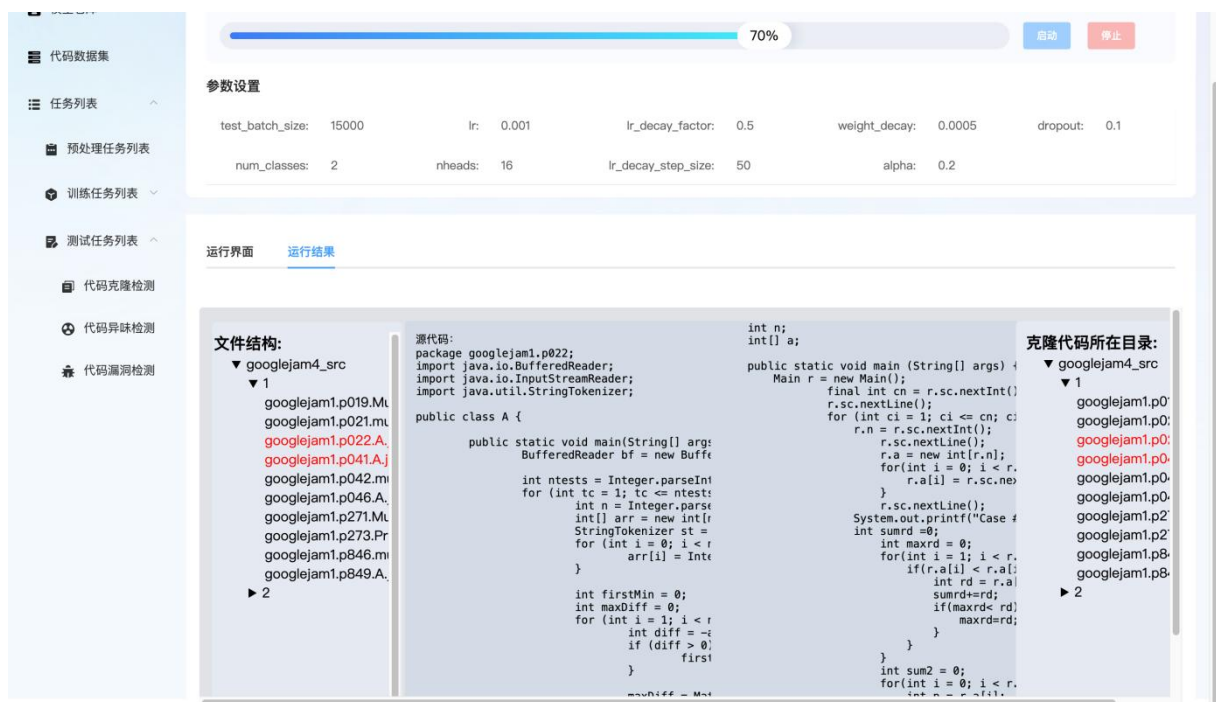
4. 任务启动后运行界面会产生控制台信息，方便用户查看执行情况。



测试任务运行界面

5. 启动成功后用户可随时点击该任务的“查看”按钮查看运行情况。

6. 当程序运行完成后，点击运行结果，可查看检测结果。注：目录中红色字体为疑似克隆文件名，点击出现源代码、疑似克隆代码以及克隆代码所在目录。



测试任务运行结果