# 工大标志

# GPU计算实验报告

## 实验三 基于昇腾CANN的目标检测应用（AscendCL接口）

学院：

姓名：

学号：

### 一、实验预习（10分）

1. 注册华为云账号：https://www.huaweicloud.com/
2. 课程内容预习：

https://education.huaweicloud.com/courses/course-v1:HuaweiX+CBUCNXA024+Self-paced/about

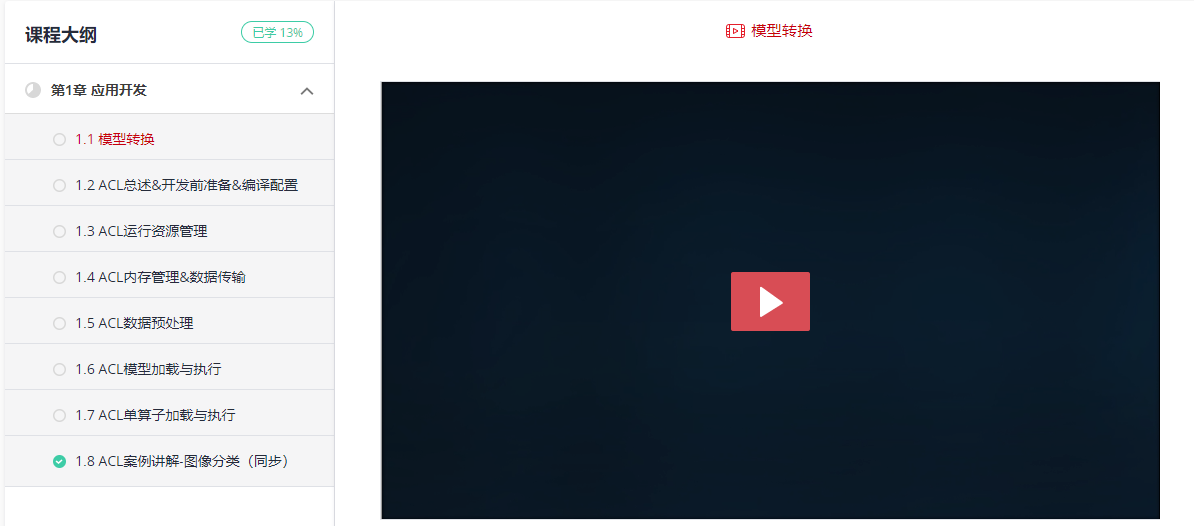


图1 昇腾学院发布的ASCENDCL课程

https://www.bilibili.com/video/BV16K4y1976L?spm\_id\_from=333.999.0.0

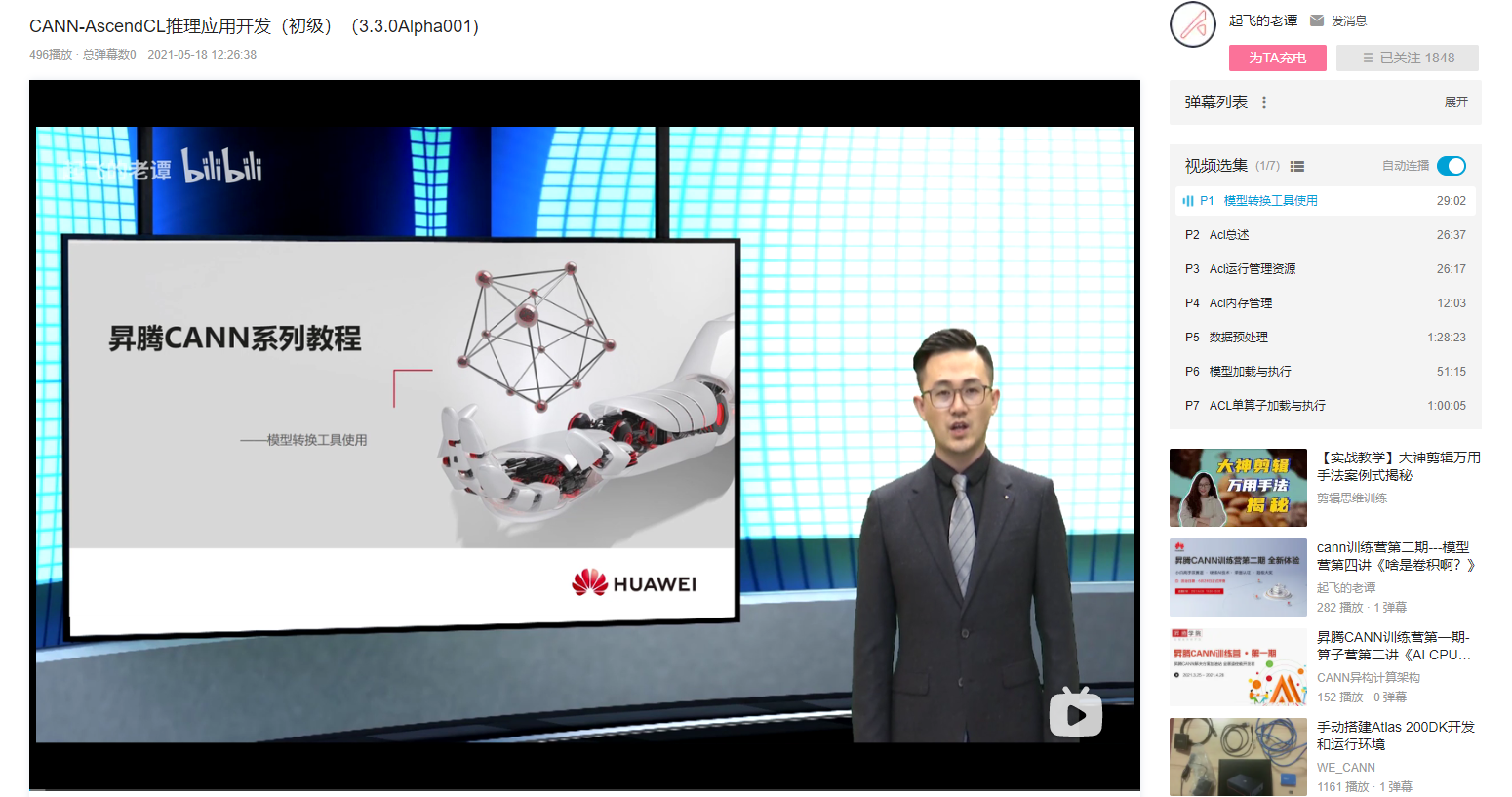


图2 华为谭老师bilibili发布的视频

1. 问题（10分）：

（回答请使用红色文字）

1. 在昇腾310上执行推理，模型格式必须是om，需要使用ATC工具进行模型转换，目前支持哪些深度学习框架的模型转换？

A、Caffe B、TensorFlow C、MindSpore D、PyTorch

答：AB

1. 模型转换时设置静态AIPP参数，模型生成后，AIPP参数被保存在om模型中，其功能主要有什么？

A、图像裁切对齐 B、色域转换 C、归一化 D、图片解码与缩放

答：BC

1. 下面那些选项是AscendCL能够完成的？
2. 运行资源管理（Device管理、Context管理、Stream管理）
3. 内存管理
4. 数据预处理DVPP
5. 模型加载与执行
6. 算子加载与执行

答：ABCDE

1. DVPP用于数据预处理，能够对图像解码、缩放，输出YUV420sp格式的图片，解码对输出图片的宽高有对齐要求，宽要求128对齐，高16对齐，缩放也对宽高有对齐要求，具体为：

答：

宽：（16）对齐

高：（2）对齐

1. ASCENDCL中推理所需的输入输出数据，是通过一种特定的数据结构来组织的，称为Dataset，所有的输入组成了一个Dataset，所有的输出组成了一个Dataset，但对于很多模型来说，输入不止一个，那么所有的输入集合叫Dataset，其中的每一个输入叫什么？

答：input\_

## 二、实验目标

1. 了解华为昇腾全栈开发工具MindStudio及其离线模型转换功能；
2. 了解如何使用ASCENDCL开发基于华为昇腾处理器的神经网络推理应用。

## 三、实验内容

完成链接中的实验：

https://lab.huaweicloud.com/testdetail.html?testId=458&ticket=ST-947652-UdMQSjr2EsqynL7E0Qu4Dkqz-sso



图3 基于昇腾AI处理器的目标检测应用（ASCENDCL）

参考文档：

https://support.huaweicloud.com/aclcppdevg-cann504alpha1infer/aclcppdevg\_01\_0001.html



图4 AscendCL官方文档

## 四、实验过程（80分）

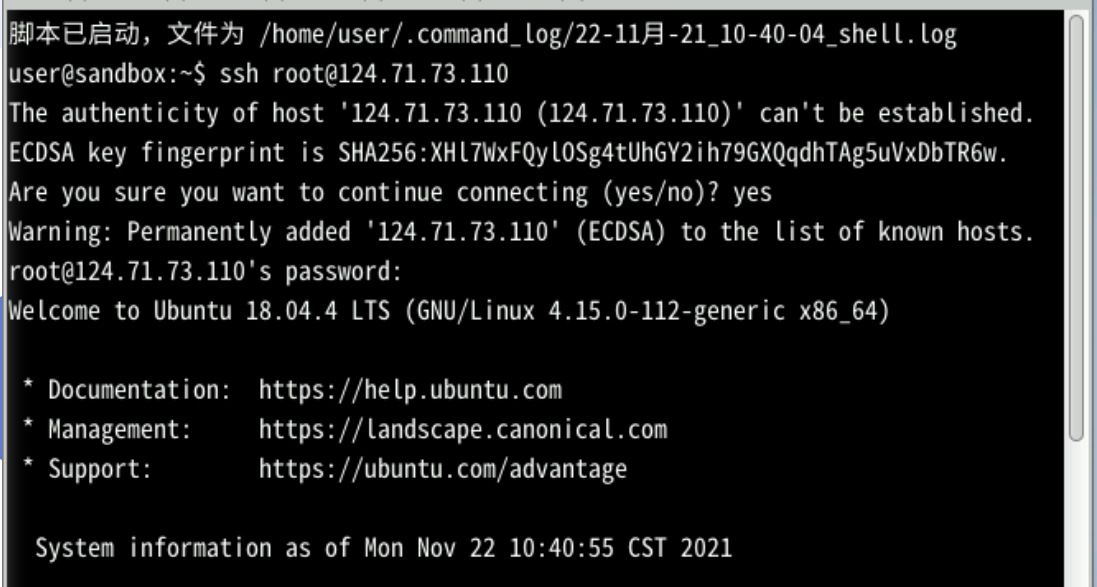
1. 环境准备（10分）

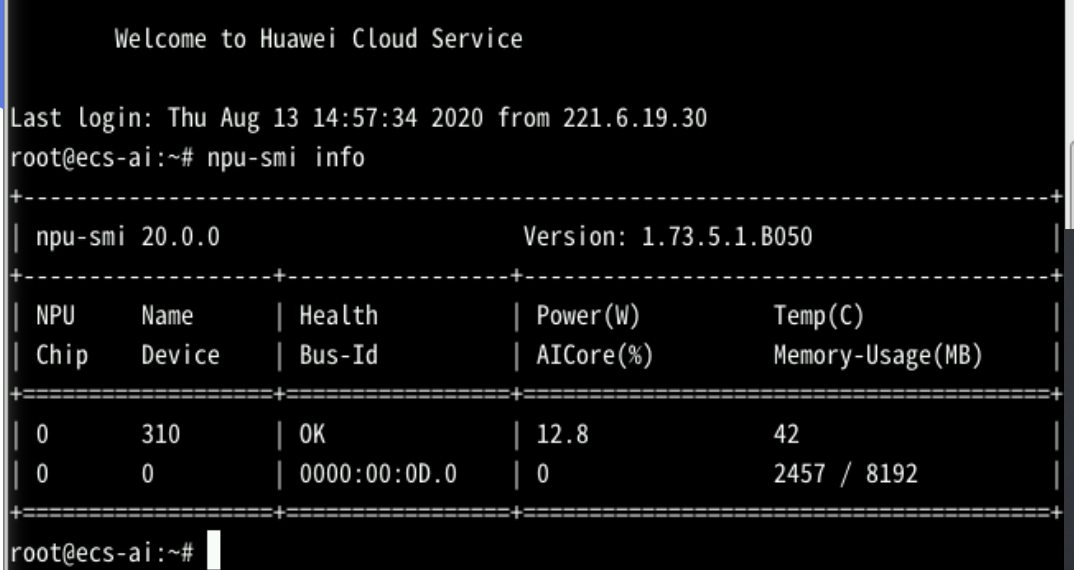
完成弹性云服务器ECS创建，获取公网IP，如XX.XX.XX.XX，然后通过终端工具登录云服务器（密码为云服务器创建时自己设置的密码）：

ssh root@XX.XX.XX.XX

并且输入下列命令获取云服务器上昇腾芯片的信息并截图：

npu-smi info





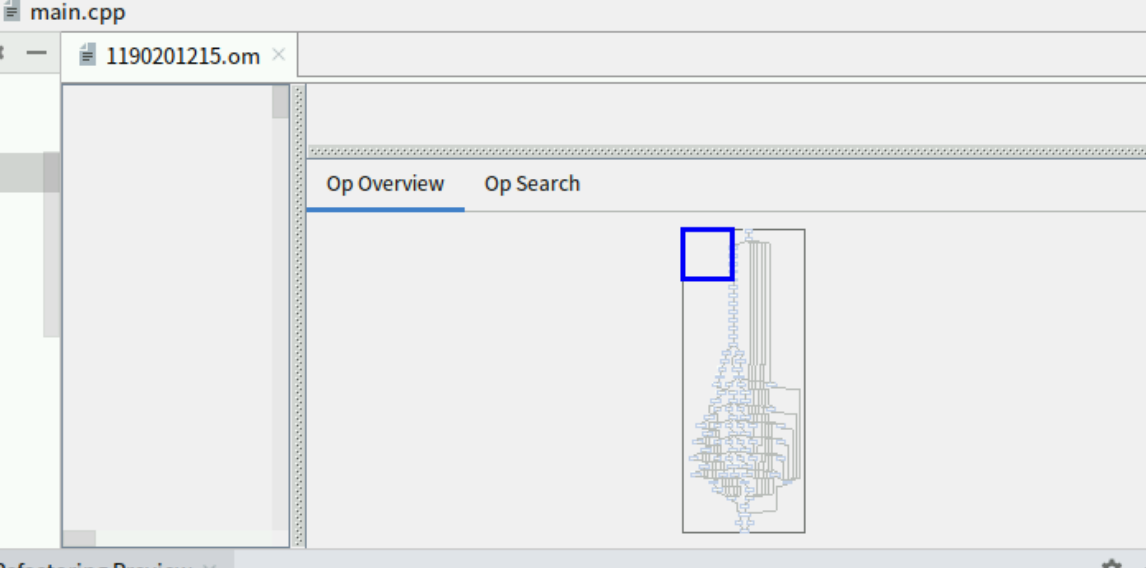
2. 模型转换（30分）

对于已经转换成功的.om模型文件，可以在MindStudio界面呈现其网络拓扑结构，并可以查看模型所使用的算子：

https://support.huaweicloud.com/usermanual-mindstudio302/atlasms\_02\_0060.html

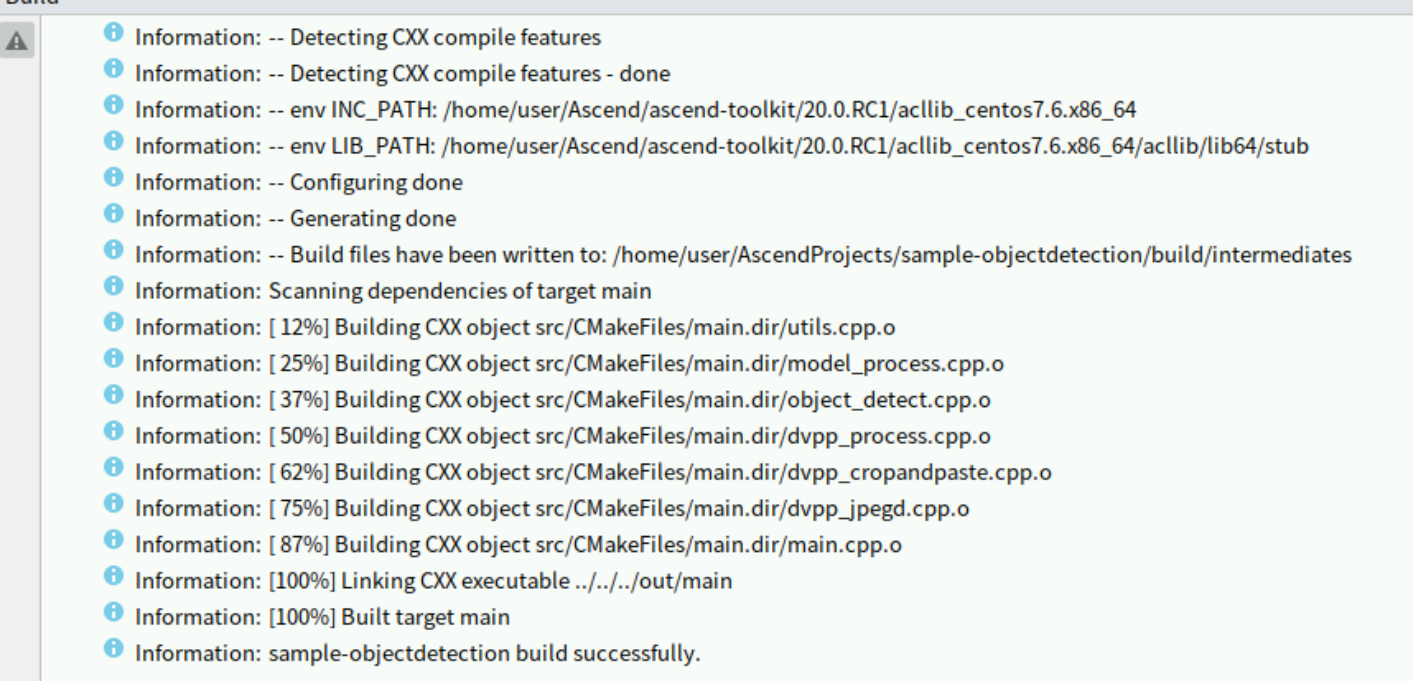
依次单击MindStudio菜单栏“Ascend > Model Visualizer”，选择转换成功的.om模型，将om模型可视化界面截图。



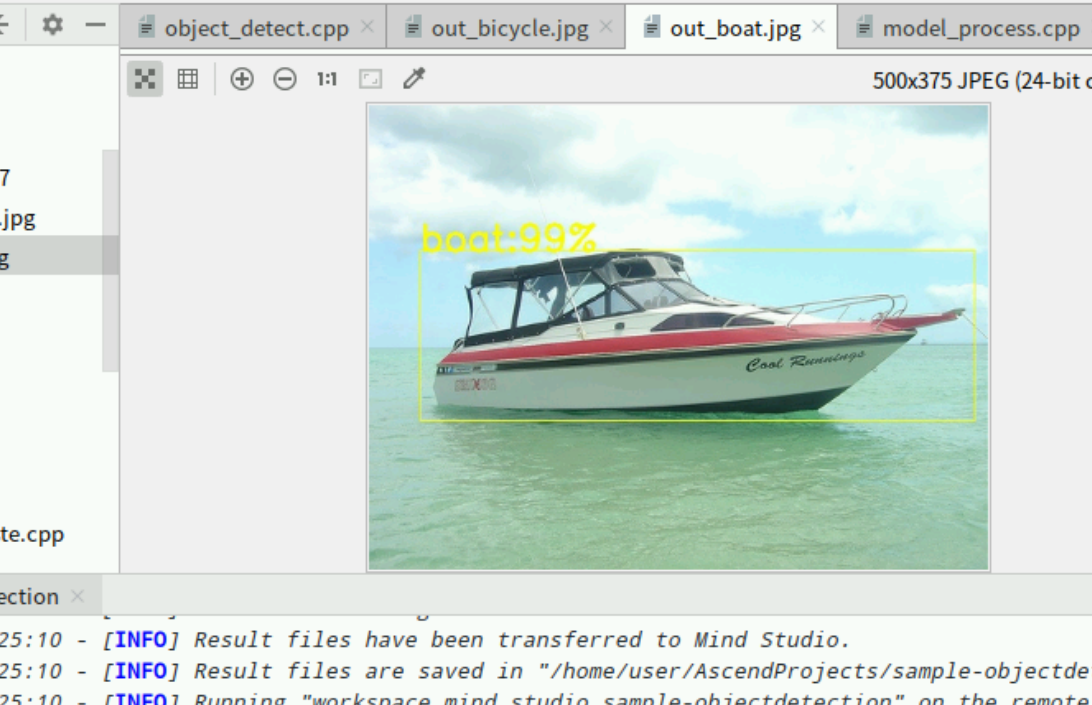


3. 编译运行（30分）

按实验手册将代码补全，并编译运行，MindStudio会将工程发送到云服务器，完成推理后再将结果回传到沙箱环境，请将两个测试用例的推理结果截图。

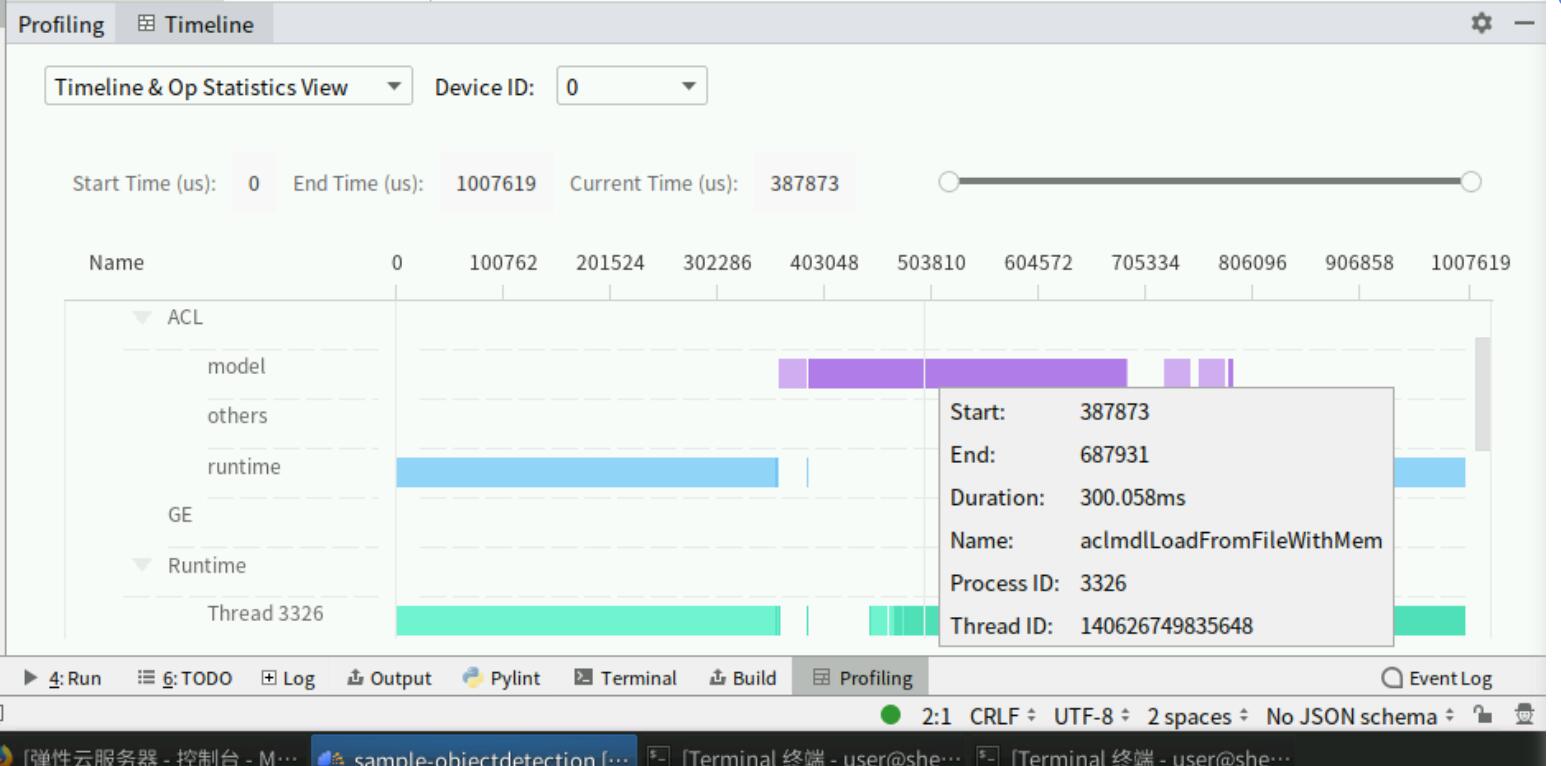




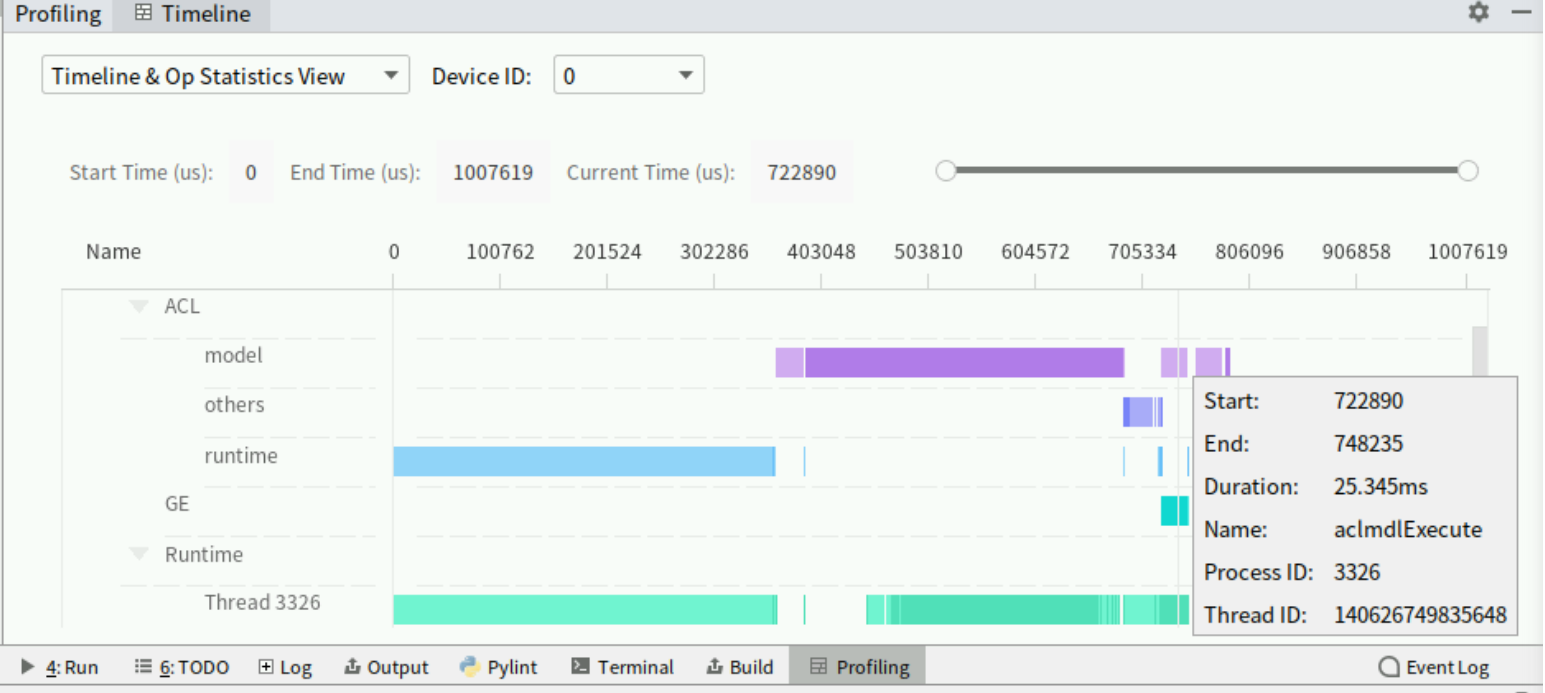


4. 模型性能（10分）

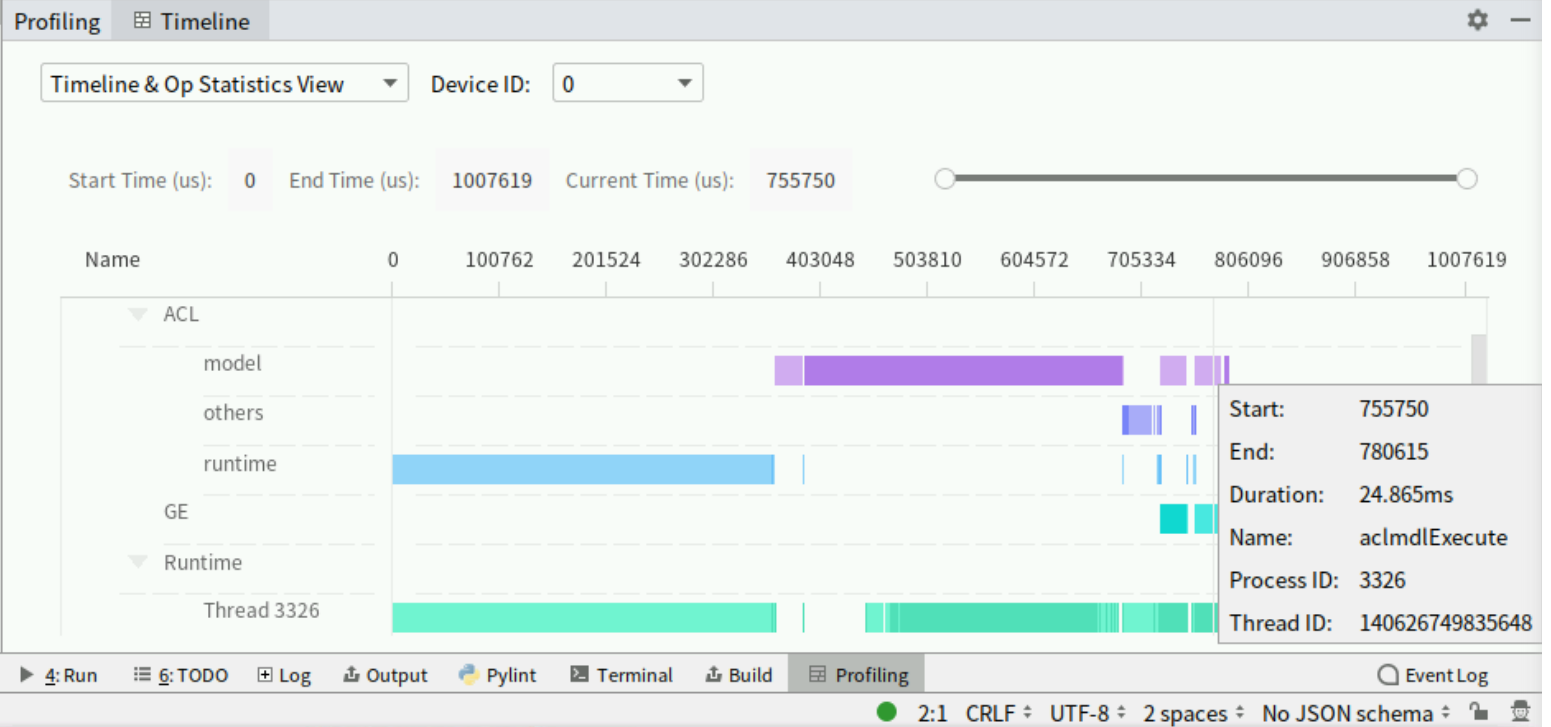
运行Profiling 查看推理性能，截图并指出模型从文件加载进内存耗时以及两张图片推理耗时。



模型从文件加载进内存耗时



第一张图片推理耗时



第二张图片推理耗时

模型从文件加载进内存耗时：300ms

第一张图像推理耗时：25.3ms

第二张图像推理耗时：24.8ms

## 五、代码分析（10分）

（由于实验环境有时间限制，可以将代码下载到本地，参考华为实验手册与课程PPT分析）

源码工程：

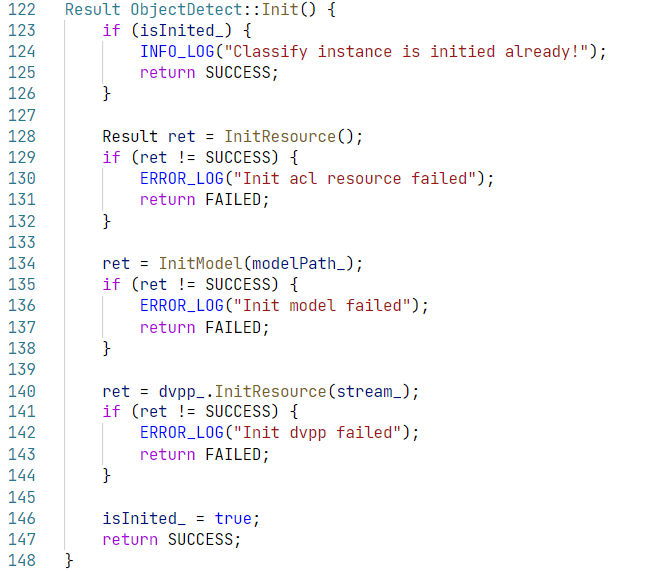
https://sandbox-experiment-resource-north-4.obs.cn-north-4.myhuaweicloud.com/shengteng-image-target/sample-objectdetection.zip

简单分析：

https://kqybjxvfo3.feishu.cn/docs/doccnr1WWpezXYWeeUykUKUpi8g#F8Ufjc

分析main.cpp可知，代码主要有五部分：初始化、数据预处理、执行推理、数据后处理、资源销毁。

1. 初始化（包括ASCENDCL资源初始化、模型初始化、dvpp资源初始化）（2分）



请查看InitResource函数的实现代码，ASCENDCL资源初始化时进行了哪些操作？

答：初始化acl init ;

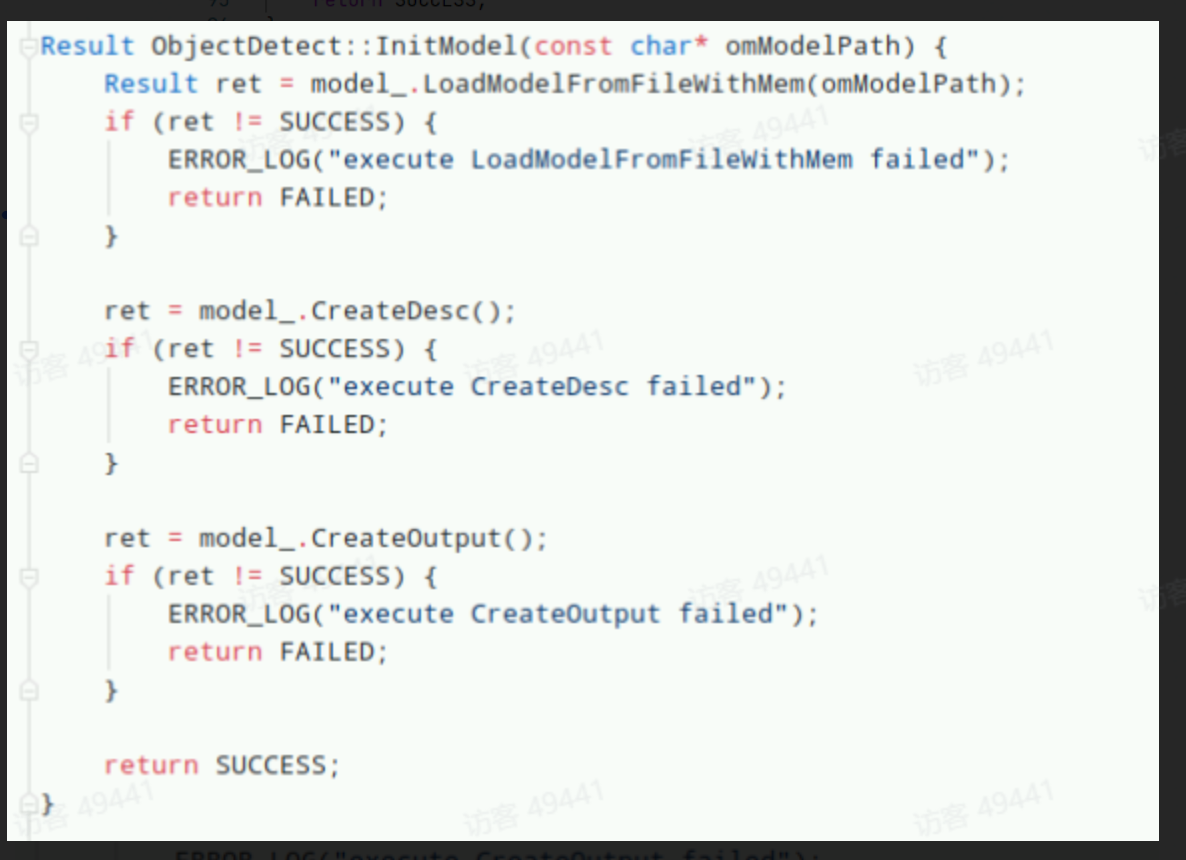
打开设备open device

创建create context(set current)

创建create stream

请查看InitModel函数的实现代码，模型初始化时进行了哪些操作？

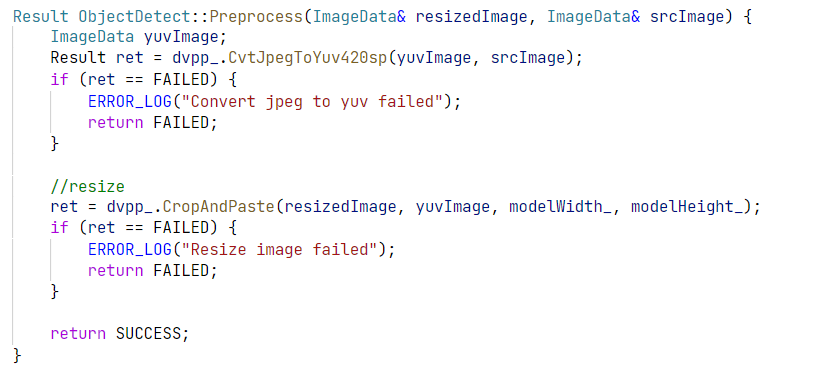
答：



模型加载、创建模型描述信息、创建模型输出

createdesc createoutput 将数据存储内存

1. 数据预处理（2分）

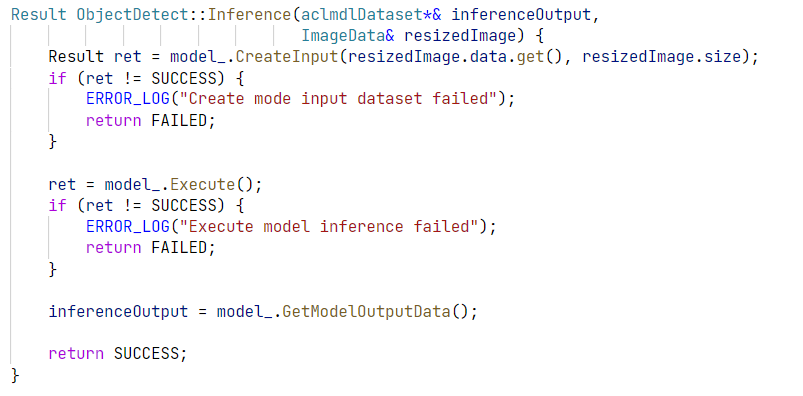


这一部分主要有两个封装好的函数，其作用分别是：

CvtJpegToYuv420sp：将JPEG格式的图像转化为可处理的Yuv420sp图像

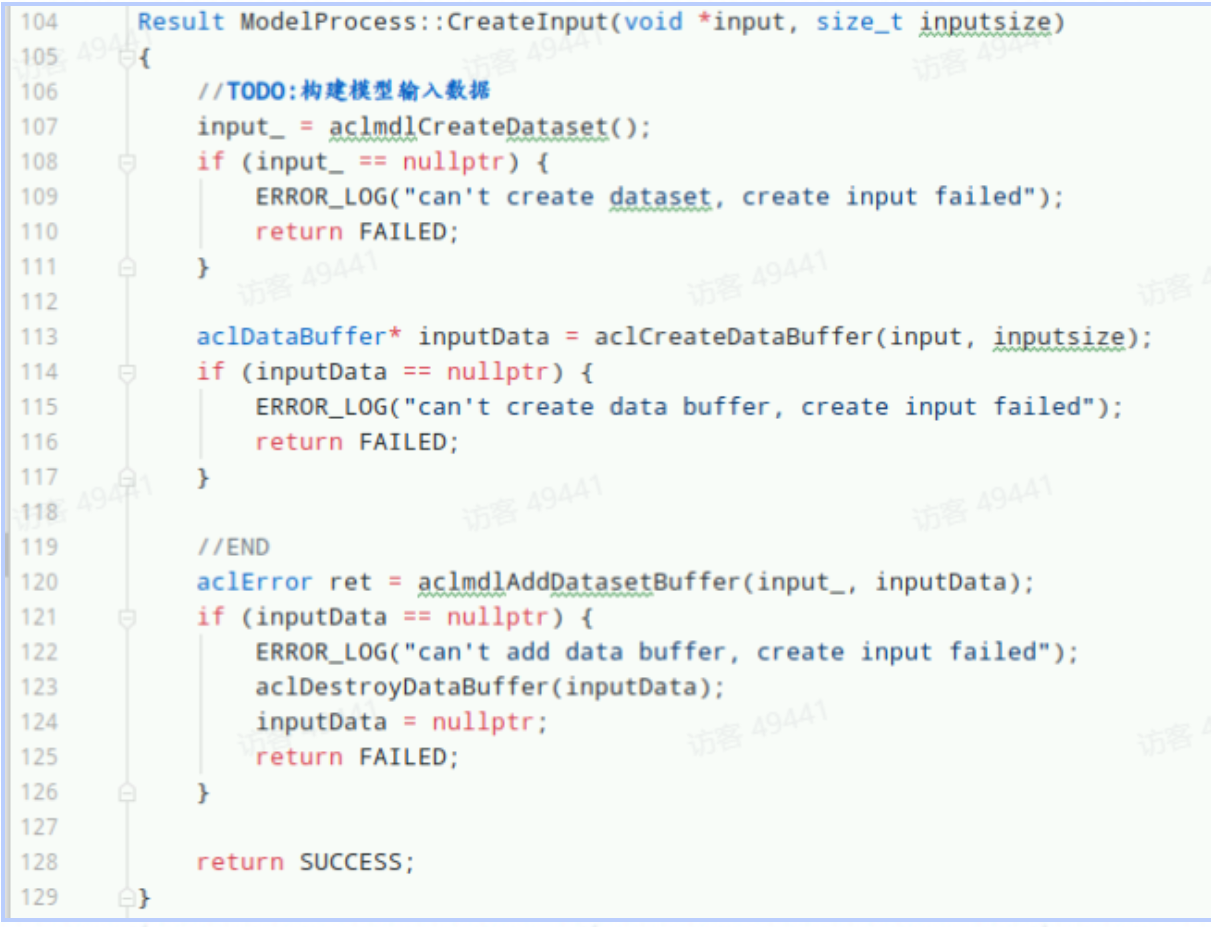
CropAndPaste：使用DVPP对图片缩放

1. 执行推理（2分）



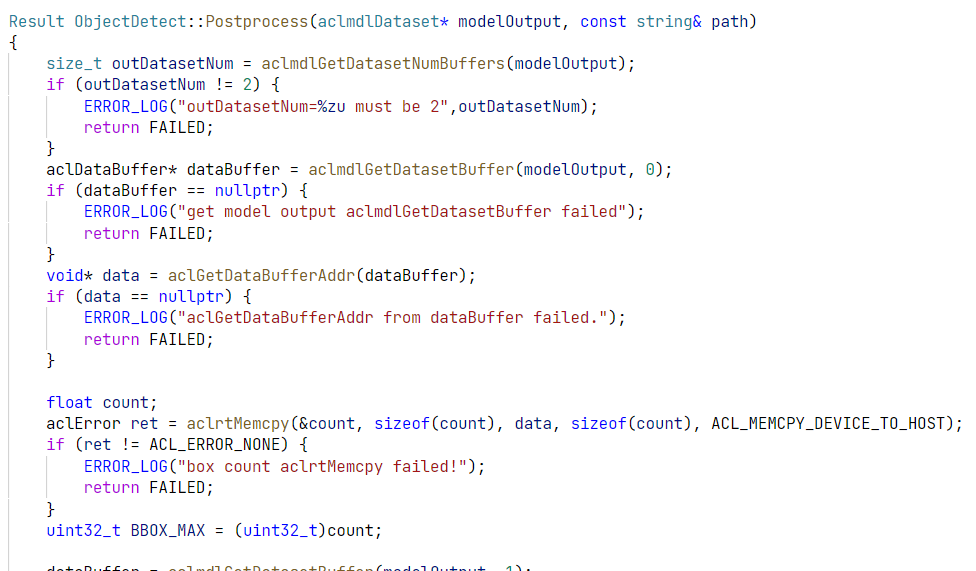
请查看CreateInput函数的实现代码，模型输入数据是怎样构造的？

答：



从aclmdCreateDataset();中读取输入数据。并开辟一段内存空间。ASCENDCL中推理所需的输入输出数据，是通过一种特定的数据结构来组织的，称为Dataset，所有的输入组成了一个Dataset，所有的输出组成了一个Dataset，但对于很多模型来说，输入不止一个，那么所有的输入集合叫Dataset。

1. 数据后处理（2分）

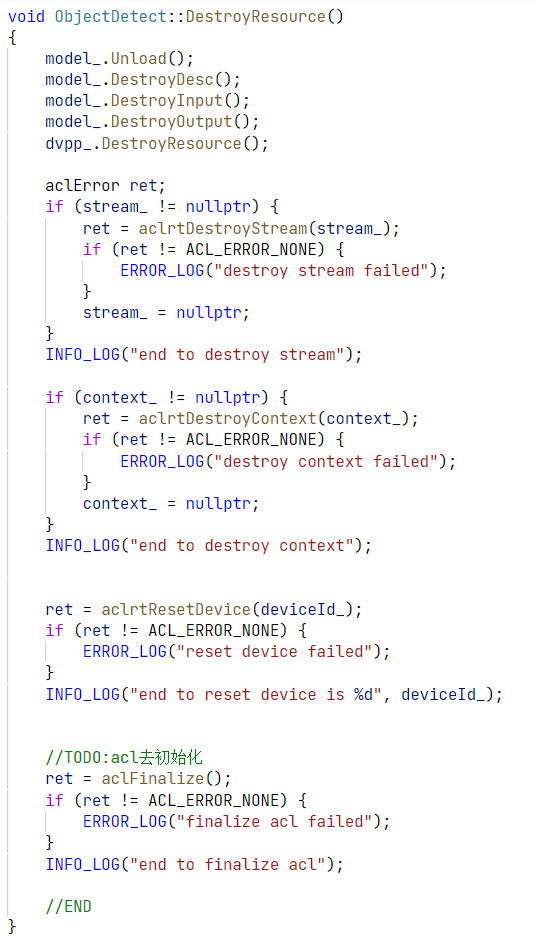


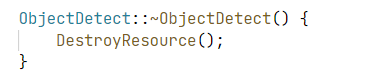


上图为后处理的代码，它将模型输出的每一个检测结果（包括边界框点的位置、检测标签、置信度）在输入图像进行标注，简单分析一个检测结果占用多少字节？

答：不会…

1. 资源销毁（2分）





~ObjectDetect()为析构函数，在该对象生命周期结束时自动调用，释放内存。其会调用DestroyResource()函数，该函数进行了哪些操作？

答：进行dvpp销毁，释放与图像相关联的内存，卸载模型销毁数据

## 六、实验感想

1.实验内容挺多

2.一步步跟着助教并照着实验指导书还是有点不熟练

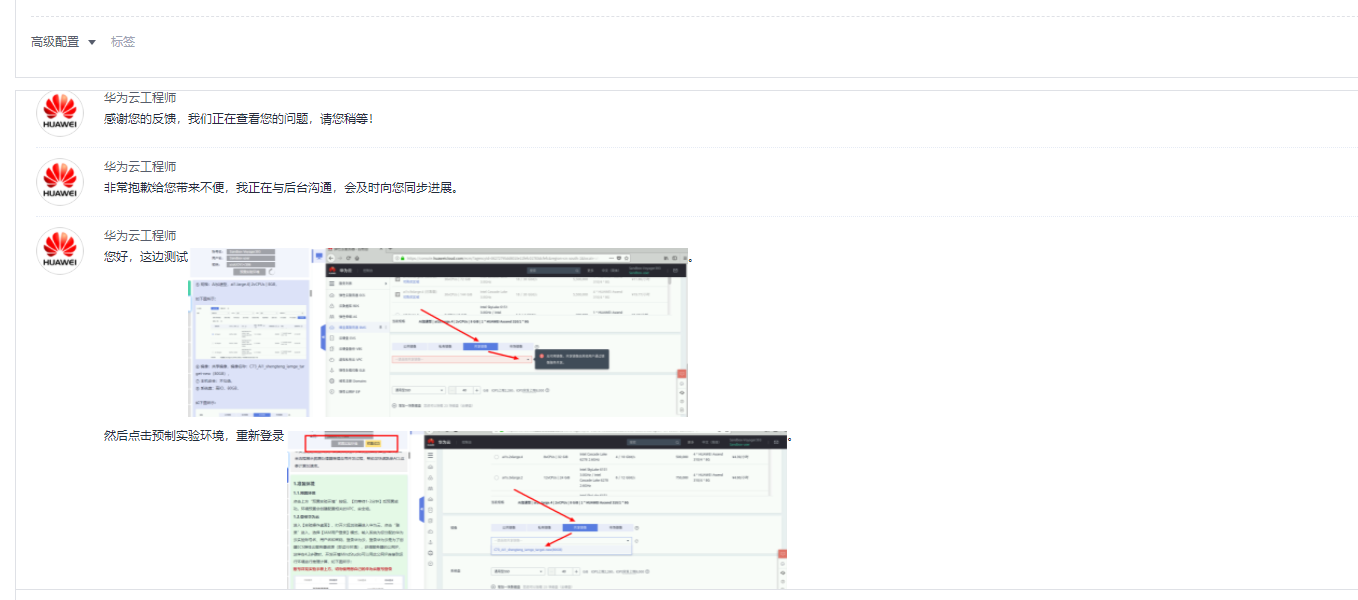
3．课后重做一遍

## 附录：FAQ

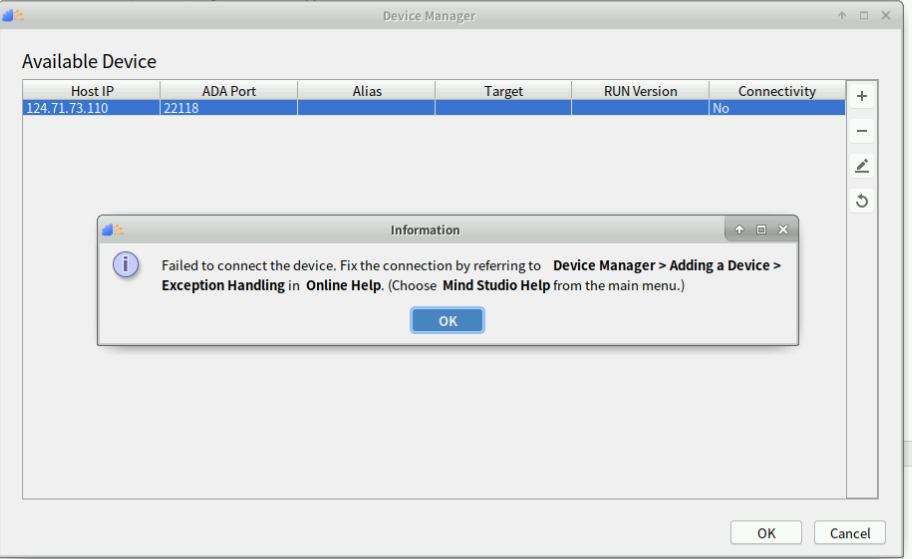
1. 找不到共享镜像？



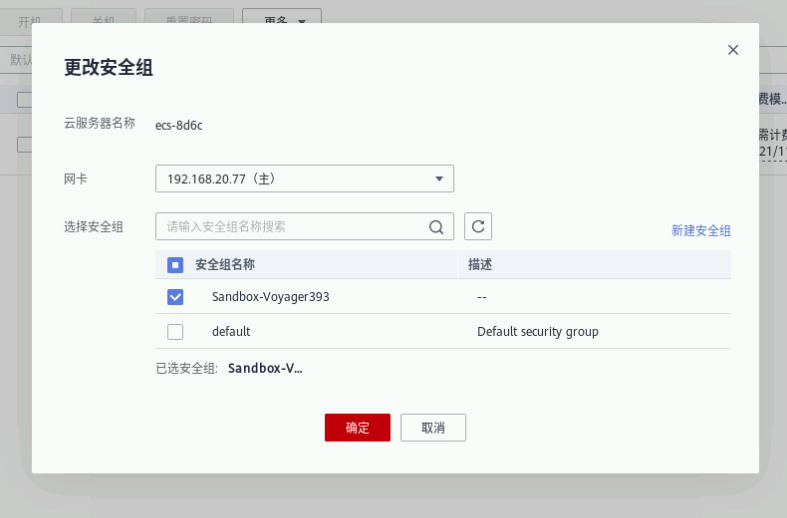
预置实验环境再重新登陆：

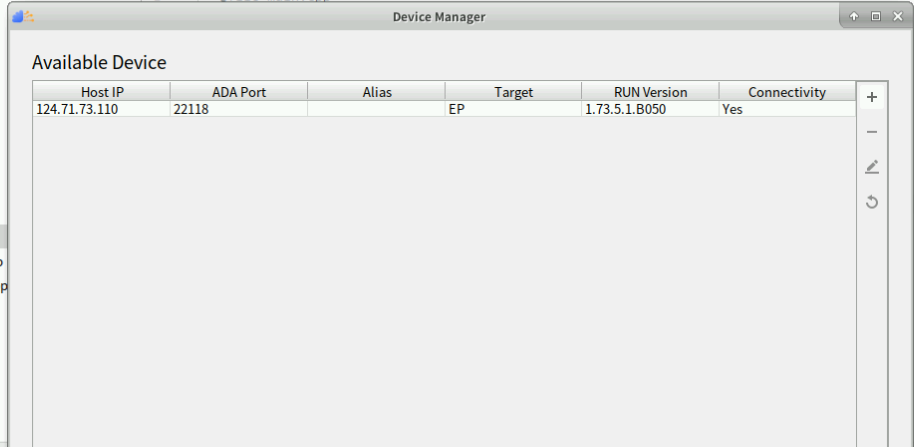


1. MindStudio无法连接云服务器？



更改安全组策略即可：





1. MindStudio运行项目，提示ERROR，并且错误信息中IP地址和端口与自己在Device Manager设置的不一致？

原因：MindStudio可能存在bug，由于某些原因，.idea/workspace.xml中hostip没有改变

解决一：将Device Manager中配置好的删掉，再重新配置一遍

解决二：按下图所示，选择AI Core，并Apply，再取消AI Core，再Apply

