# 机器学习大作业报告

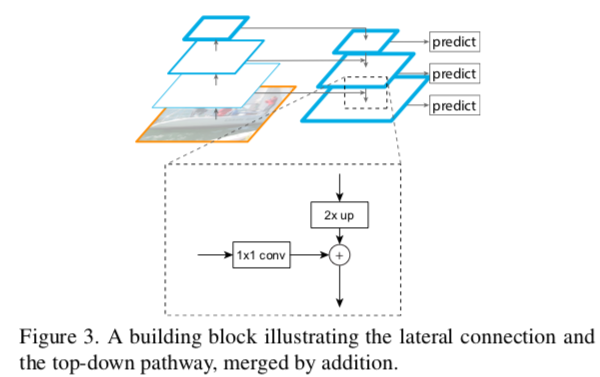
### 1问题描述：

数据集中的安检图片都存在不同等级的遮挡问题，遮挡会严重影响检测器识别危险品的准确率。安检机返回x光图像为RGB彩色图像。训练集中危险品包括带电芯充电宝和不带电芯充电宝两个类别。训练集中共有6000张图片，每张图片都拥有一个危险品所在的位置标注文件。根据训练集训练出能够识别并标注危险品的模型。

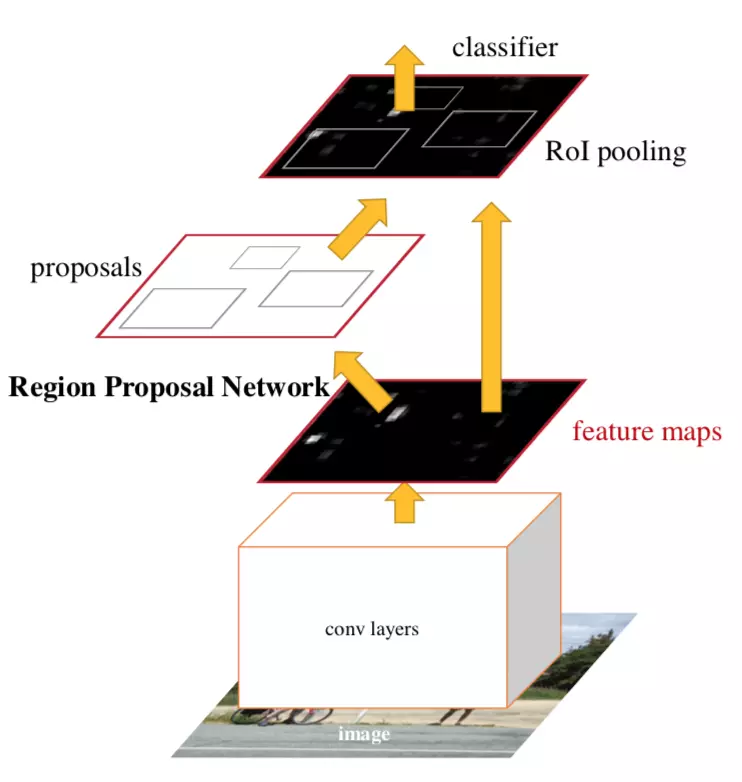
### 2模型介绍：

解决此问题使用的网络是基于resnet50的Faster-RCNN网络。并在特征提取中整合FPN来高效提取图片中各维度的特征。使用reset50为主干是因为图片中有遮挡，所以想要使用深层网络结构来提高特征提取的准确性，而对于普通的深层网络，准确率可能由于层数的加深而降低，所以使用了reset50网络来保证一定的准确率。

浅层的网络比较注重细节，所以我们可以用浅层的网络来提取坐标，深层的网络更注重语义的提取，所以我们可以用深层的网络来准确目标，所以我们使用FPN构建特征金字塔，同时获取浅层和深层的网络信息。我们的特征金字塔可以在速度和准确率之间进行权衡，可以通过它获得更加鲁棒的语义信息。其网络结构如下图所示：



Faster-RCNN在结构上将特征提取，proposal提取，边框回归和分类整合到一个网络中，以达到综合性能较高，检测速度较快的效果。其网络结构如下图所示：



如图1，Faster-RCNN可以分为四个主要内容：

1. 卷积层（conv layers），用于提取图片的特征，输入为整张图片，输出为提取出的特征，即feature maps。
2. RPN网络（Regin Proposal Network），用于推荐候选区域，输入为图片（feature maps），输出为多个候选区域。
3. RoI pooling，将不同大小的输入转换为固定长度的输出。
4. 分类和回归，这一层的输出是最终目的，输出候选区域所属的类和候选区域在图像中的精确位置。

FPN是一种利用常规CNN模型来高效提取图片中各维度特征的方法。在计算机视觉中，多维度的目标检测一直以来都是通过将缩小或扩大后的不同维度图片作为输入来生成出反映不同维度信息的特征组合。这种办法虽然能够有效地表达出图片的各种维度特征，但却对硬件计算能力及内存大小有较高的要求。

而FPN通过利用常规CNN模型内部自底向上各层对图片不同维度的特征表达结构，提出了一种在单一图片视图下对其生成多维度特征表达的方法。它可以有效地增强常规CNN模型，从而生成出表达能力更强的feature maps以提供下一阶段任务来使用。

在本次作业中，它是一种加强主干网络CNN特征表达的方法。

### 3实现：

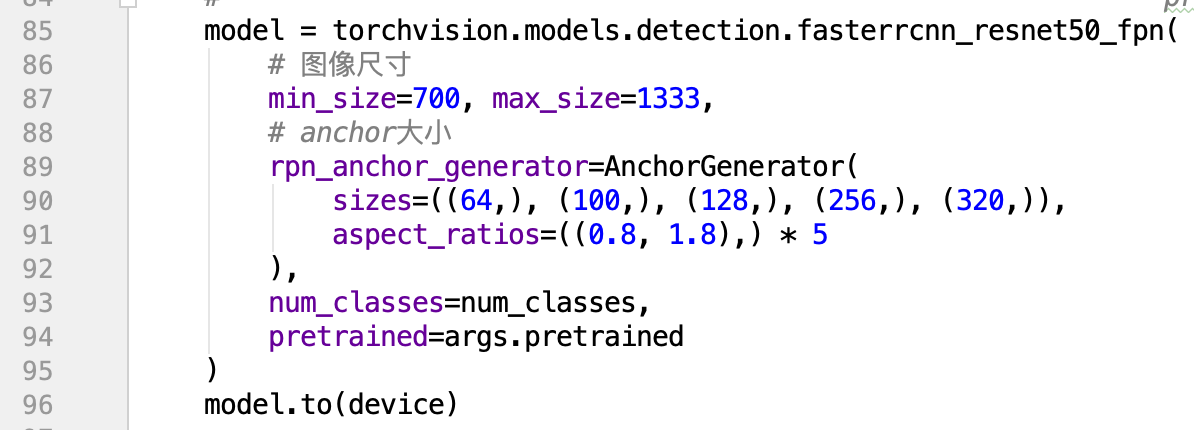
本次作业使用pytorch实现：

1. 图像预处理。

将训练集转为voc2012格式。

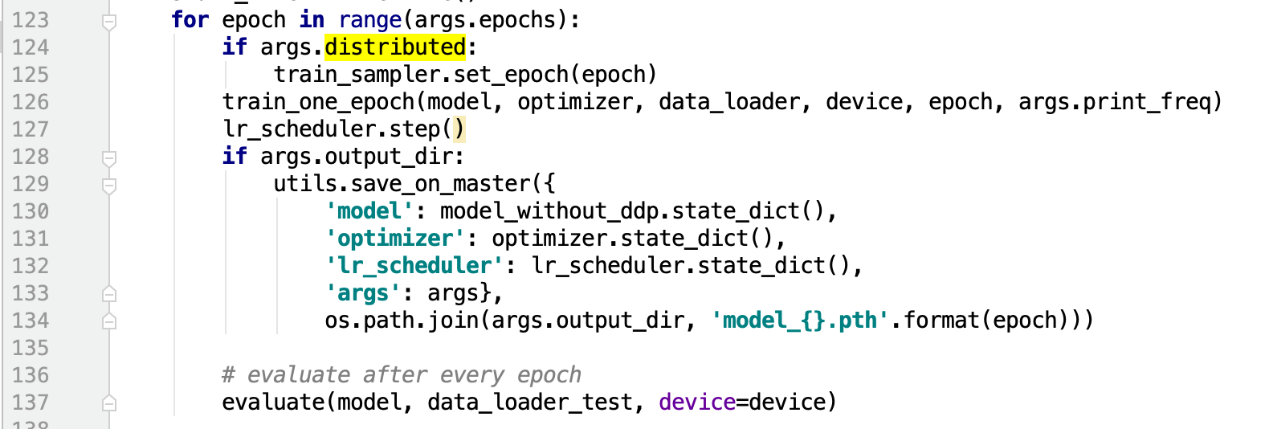
1. 模型训练。

训练集中共有6000张图片，按5500:500来从训练集中划分出验证集。训练模型调用方法：torchvision.models.detection.fasterrcnn\_resnet50\_fpn。代码如下图所示：



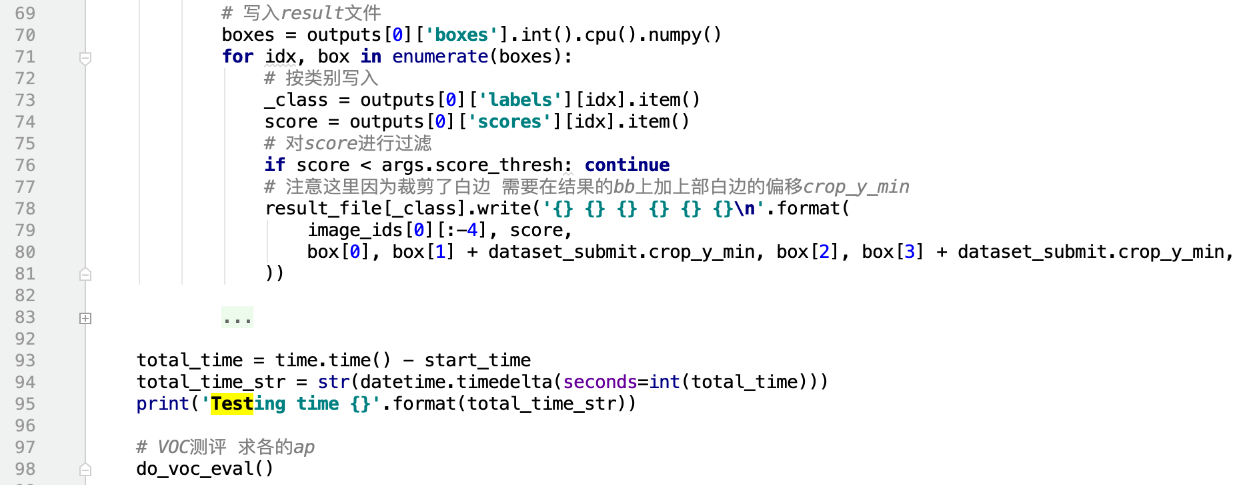
1. 模型保存。

代码如下图所示：

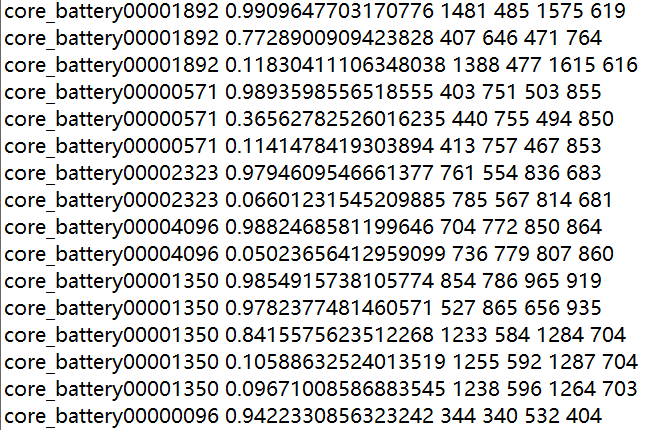


1. 模型测试与结果存储。

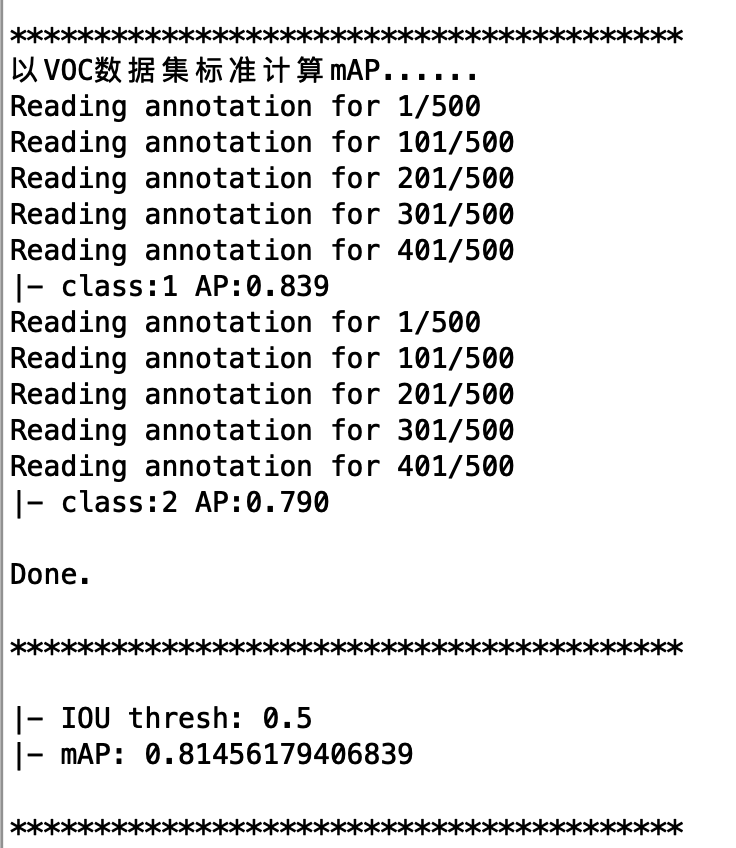
对训练的模型使用验证集进行测试。测试代码如下图所示：



测试结果如下图所示：



计算mAP，结果如下图所示：



### 4.小组分工

董晓云：数据集转换

张淑慧：模型搭建，模型训练

张 伟：资料收集、报告文档

张紫娴：模型测试、答辩

杨晓晨：资料收集、PPT制作、模型调参