**机器学习——任务一（遮挡问题）**

组 长： 董善为 ZF1906204

小组成员： 刘一萱 ZF1906209

刘 云 ZF1906274

应昌君 ZF1906299

1. 问题描述

任务：训练一个模型用于目标检测任务：输入一张3通道安检x光照片，识别图片中给定类别的物体，并输出物体类别以及位置信息。其中需要识别的类别有带芯充电宝和不带芯充电宝两种。两类物体在各照片中可能会出现不同程度的遮挡。

难点：图片上物品的遮挡问题处理

1. 解决方案
2. 遮挡的解决方案

通过观察图片，我们小组一致认为遮挡不构成问题，可以直接进行训练。

1. 使用模型简介
   1. 目标检测简介

目标检测原理是定位图像中目标的存在，并在该目标周围绘制一个边界框。目前有两种目标检测的模型，分别为Two-stage detectors和One-stage detectors，Two-stage detectors常见的模型架构有RCNN、Fast-RCNN、Faster-RCNN，One-stage detectors常见的模型架构有YOLO、SSD等，其中本次使用的是yolov3模型。

* 1. Yolov3模型原理简介

Yolov3由一个用于提取图片特征的卷积神经网络层Darknet、一个多尺度特征生成器层和一个多分类网络层构成，在目标检测这一任务上实现了one-stage：输入原始图像，经过整个网络，直接输出目标的类别和位置信息。

该模型相比大多数目标检测模型来说，显著的优点是训练较为容易。因为其前置的特征提取网络基于Darknet53，是一种类似于Resnet53但训练速度快很多的网络结构。相比于ssd中的vgg16网络，Darknet网络不需要预训练权重，而这次任务由于数据集的特殊性，无法加载在常规数据集上的预训练权重。所以Yolov3更适合这次任务。

* 1. 训练过程主要步骤如下

1. 先验框匹配

确定图片中的ground truth（真实目标）与哪个先验框来进行匹配，用与之匹配的先验框所对应的边界框将负责预测它。匹配原则有以下两点：首先，对于图片中每个ground truth，找到与其IOU最大的先验框进行匹配（IOU计算公式），若一个先验框没有与任何ground truth进行匹配，那么该先验框只能与背景匹配；其次，对于剩余的未匹配的先验框，若与某个ground truth的IOU大于某个阈值（一般为0.5），那么该先验框也与这个ground truth进行匹配，若多个ground truth与某个先验框IOU大于阈值，那么先验框只与最大的先验框进行匹配。将与ground truth匹配到的先验框成为正样本。

1. 损失函数

损失函数主要用于计算分类损失和位置损失，此处给出的为位置误差与置信度误差的加权和：

log() —

其中：，

其中，N是先验框的数量，i为先验框序号，j为真实框序号，p为类别序号；∊{1, 0} 是一个指示参数，取1时表示第i个先验框与第j个ground truth匹配，且该ground truth的类别为p；l为先验框对应边界框的位置预测值，g为ground truth的位置参数，其中，位置误差采用；表示第i个先验框对应类别p的预测概率；权重系数α通过交叉验证设置为1。

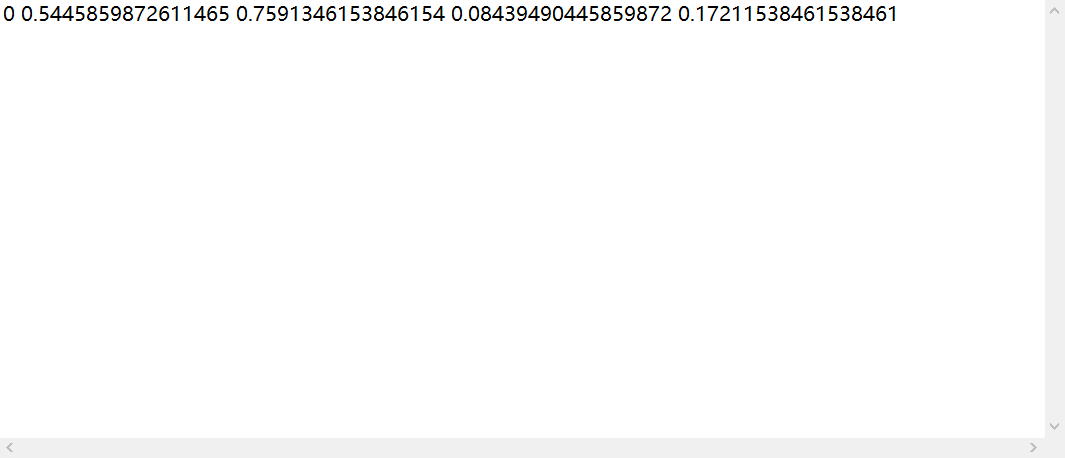
1. 预测过程

对于每个预测框，首先根据类别置信度确定其类别与置信度值，并过滤掉属于背景的预测框。然后根据置信度阈值过滤掉阈值较低的预测框。对于留下的预测框进行解码，根据先验框得到其真实的位置参数，解码后一般还需要做clip，防止预测框位置超出图片。解码之后，一般需要根据置信度进行降序排列，然后仅保留top-k个预测框。最后就是进行NMS算法，通过设置IOU阈值过滤掉那些重叠度较大的预测框。最后剩余的预测框就是检测结果。

1. 样本的预处理

（1）图片先在保持比例的基础上pad至正方形，然后统一resize到416\*416。

（2）将所给Annotations文件夹中的标签转换成如下格式以适应项目：



以空格分割的五个数值分别代表类别、中心点横坐标、中心点纵坐标、框的宽度、框的高度，其中后四项通过除以图片的宽度和高度值缩放到[0，1]区间内。

1. 代码实现过程

已上传至github.com/semirol/yolov3\_pytorch

1. 实现结果

代码运行环境：centos

python环境：3.6

依赖：

numpy

torch>=1.0

torchvision

matplotlib

tensorflow

tensorboard

terminaltables

pillow

tqdm

cuda版本：9.0

显卡：NVIDIA Tesla V100

Batchsize：16

学习率：0.01

训练数据量：5700

Epoch：100

Map：0.91

