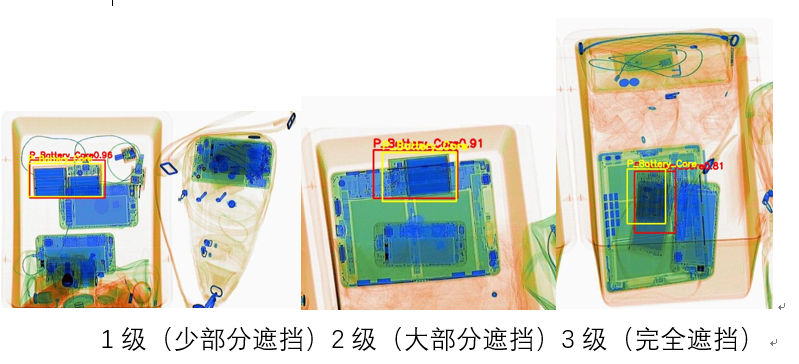
**目标检测遮挡问题**

**1 问题背景**

日常出行中人们往往会选择坐地铁、火车、高铁、飞机等交通工具。这此过程中为了保证乘客的安全，需要对乘客携带的行李使用安检机进行检查。安检机的任务是在众多的物品中识别出所有禁止携带的危险物品，这就是目标检测——识别图像中目标的位置。但是在现实应用中由于场景的不同，目标检测还是有很大的优化空间。例如：在安检的应用中，我们不得不解决的问题就是乘客往往携带的行李包含大量的物品，而且种类繁多，这会对目标造成不同程度的遮挡，如果不能合理解决，将会造成大量的漏检、错捡，会带来很大的不便。

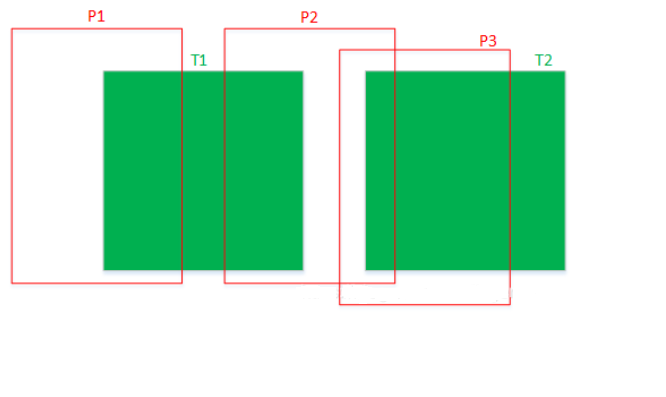
**2 问题分析**

本次实验针对于乘客行李中十分常见的危险品——充电宝，进行目标检测并解决一定程度的带核充电宝和不带核充电宝的遮挡问题。在所提供的数据集中，发现我们的数据集大致可以分为三类（图2.1）：1级（少部分遮挡）、2级（大部分遮挡）、3级（完全遮挡）。

图2.1 目标遮挡等级

遮挡问题可以分为两种类型：

1. 待检测的目标被干扰物体遮挡（Occlusion）：一般的解决办法是使用更大量的数据进行训练模型使其具有更强的特征从而优化效果；
2. 待检测的目标之间相互遮挡（Crowded）：此类情况可以具象为图2.2，假设图中的预测框P1不存在，也即T1应的proposals只有P2，如果T2对应的max score的proposal近似为P3，那么当P2和P3的IoU(交并比)大于阈值（一般在0.5-0.7之间）时，在后处理NMS阶段就会把P2过滤掉，从而导致更严重的情况，T1被漏召回了。

图2.2 目标遮挡

针对此类不精准和漏召回的情况，比较直接的想法是，能不能让P2对应的预测框不受T2的干扰。方法一：可以在后处理阶段采用Soft NMS，就是用稍低一点的分数来代替原有的分数，而不是直接置零，由于Soft NMS可以很方便地引入到目标检测算法中，不需要重新训练原有的模型，因此这是该算法的一大优点。方法二：可以在增加focal loss，以及调整目标检测模型中loss函数的权重，权衡其解决三种程度遮挡的效果。

**3 模型介绍**

针对于本次实验，是检测安检机其传送带上的行李，对于实时的要求较高，所以我们选择one-stage网络中的Single Shot MultiBox Detector（SSD）。SSD是one-stage网络中较为代表性算法，在满足速度要求的同时，保证了一定的精度。SSD是利用单个深度神经网络对图像中目标进行预测的端对端模型。该网络将dounding box的输出空间离散化为一组默认框（default box）,这些默认框在每个特征图位置有不同的宽高比和尺寸。在预测期间，网络针对每个默认框中的每个存在目标类别生成分数，并且对默认框进行一些调整以更好的匹配对象形状。并且SSD的创新点之一是多尺度特征图，即使用不同尺度的默认框在不同的输出层，网络可以组合来自具有不同分辨率的多个特征图的预测，以适应处理各种尺寸的对象。

SSD的优点：

1. 与之前的网络相比，不需要生成将近2000个左右的候选框（region proposal）；
2. 完全消除了proposal生成和后续的像素或重采样阶段，并将所有计算封装在单网络中；
3. 使用多尺度滤波器，预测边界框中的对象类别和偏移，并且将这些滤波器应用于网络后期的多个特征图，以便执行多尺度检测。

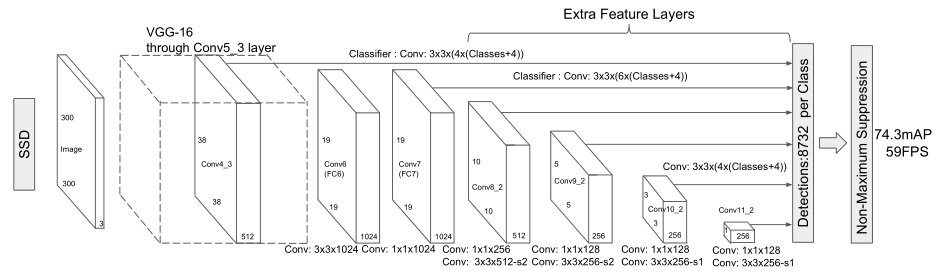
**3.1 网络结构**

SSD使用了一个前馈卷积神经网络（VGG16），针对于方框中的目标类别实例，产生一个固定大小的边界框的集合和分数，增加了卷积特征层，允许多比例特征映射检测，每个特征映射单元都链接到一组默认的边界框。算法步骤如下：

1. 输入一幅图片，将其送入训练好的分类网络（VGG16）获得不同大小的特征映射；
2. 抽取Conv4\_3、Conv7、Conv8\_2、Conv9\_2、Conv10\_2、Conv11\_2层的feature map,然后分别在这些feature map层上面的每个点构造6个不同尺度大小的bounding box,然后分别检测分类，生成多个bbox;
3. 将不同的feature map获得的bbox结合起来，经过NMS方法来抑制一部分重叠或不正确的bbox，生成最终的bbox集合。

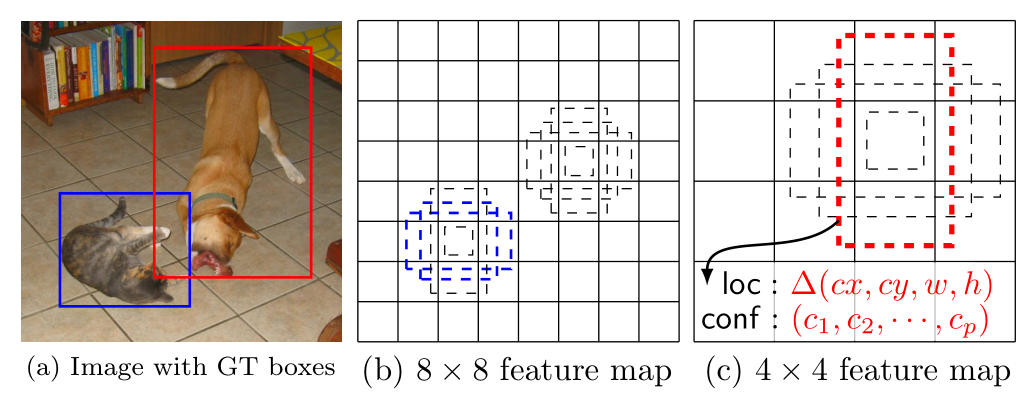
**3.1.1 多尺度特征映射**

SSD的backbone是基于VGG16，但对其进行了一定的修改，从而达到多尺度输出,将全连接层fc6和fc7转化为卷积层conv6（空洞卷积）和conv7，同时将pool5变为3\*3-S1，去掉所有fc8层和dropout层（如图3.1）。SSD算法中使到了conv4\_3,conv\_7，conv8\_2,conv7\_2,conv8\_2,conv9\_2,conv10\_2,conv11\_2这些大小不同的feature maps，其目的是为了能够准确的检测到不同尺度的物体，因为在低层的feature map,感受野比较小，高层的感受野比较大，在不同的feature map进行卷积，可以达到多尺度的目的。

图3.1 SSD的VGG16网络

**3.1.2 默认框**

如图3.2中，在特征图的每个位置预测K个bbox，对于每一个bbox，预测C个类别得分，以及相对于default box的4个偏移量值，这样总共需要 **(C+4)\*K**个预测器，则在m\*n的feature map上面将会产生 **(C+4)\*K\*m\*n**个预测值。SSD最终会有8732个box送入NMS，最后总输出结果。

图3.2SSD framework

默认框的生成规则：

1. 以feature map上每个点的中点为中心（offset=0.5），生成一系列同心的Defalut box（然后中心点的坐标会乘以step，相当于从feature map位置映射回原图位置）；
2. 使用m(SSD300中m=6)个不同大小的feature map 来做预测，最底层的 feature map 的 scale 值为 Smin=0.2，最高层的为Smax=0.95，其他层通过下面的公式计算得到；
3. 使用不同的宽高比（ratio）：[1, 2, 3, 1/2, 1/3]，通过下面的公式计算 default box 的宽度w和高度h；
4. 而对于ratio=0的情况，指定的scale如下所示，即总共有 6 中不同的 default box。

**3.2 训练**

**3.2.1 匹配策略**

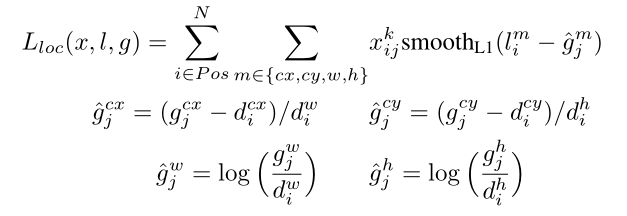
在训练过程中，首先要确定训练图片中的ground truth（真实目标）与哪个先验框来进行匹配，与之匹配的先验框所对应的边界框将负责预测它。SSD的先验框与ground truth的匹配原则主要有两点。第一点：对于图片中每个ground truth，找到与其IOU最大的先验框，该先验框与其匹配，这样，可以保证每个ground truth一定与某个先验框匹配；第二点：对于剩余的未匹配先验框，若某个ground truth的IOU大于某个阈值（一般是0.5），那么该先验框也与这个ground truth进行匹配。

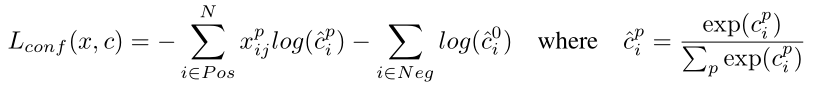
**3.2.2 损失函数**

loss函数分为两部分：计算相应的default box与目标类别的confidence loss以及相应的位置回归：

其中N是match到Ground Truth的default box数量；而alpha参数用于调整confidence loss和location loss之间的比例，默认alpha=1。

位置回归则是采用 **Smooth L1 loss**，loss函数为:

confidence loss是典型的**softmax loss**：

**4 代码及模型训练**

**4.1 代码介绍**

本次实验代码来源于github，仅仅支持VOC0712和COCO数据集的训练。通过更改config.py配置代码，新增SIXray.py本次实验数据集的读取代码。来支持自定义数据集的训练。

模型训练的基本配置预训练基本网络权重采用的vgg-16，学习率设置成0.0001，模型输入300\*300。因为没有GPU，CPU训练速度我们只训练了10000次迭代。数据集的分配我将6000张图片大概安装2比8的比例分成训练集合测试集。随机4000张图片作为训练数据集和2000张图片作为测试数据集。测试时我们会加载已训练好的模型，对2000张图片中的带电芯电池和不带电芯电池的位置进行预测并计算出AP，最后再计算两种类型的mAP。

**4.2 实验效果**

在训练出模型之后，我们对模型进行了测试评估，得到的带电芯电池和不带电芯电池的AP以及两类的mAP值如下图4.1，此图中的AP和mAP值是没有确认

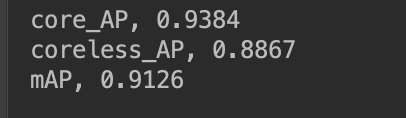
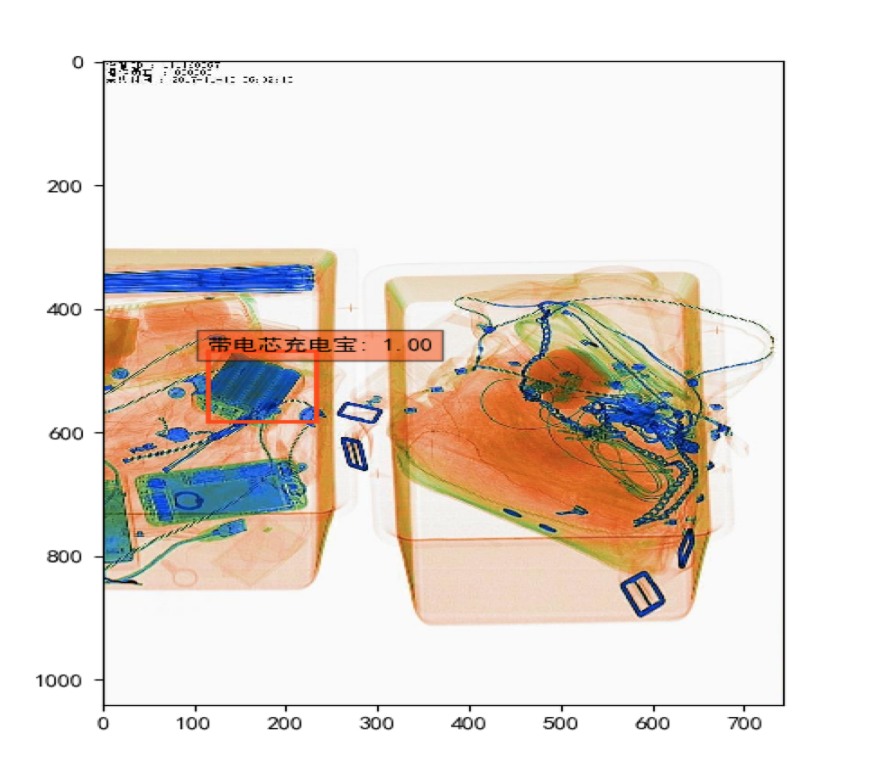


图4.1模型的AP和mAP

遮挡问题的测试值。故我们单独取了几张少部分遮挡、大部分遮挡、完全遮挡的图片进行测试，其中完全不遮挡图片的结果如图4.2，少部分遮挡测试结果如图4.3，大部分遮挡测试结果如图4.4，目标物体相互遮挡测试结果如图4.5。

图4.2无遮挡测试结果

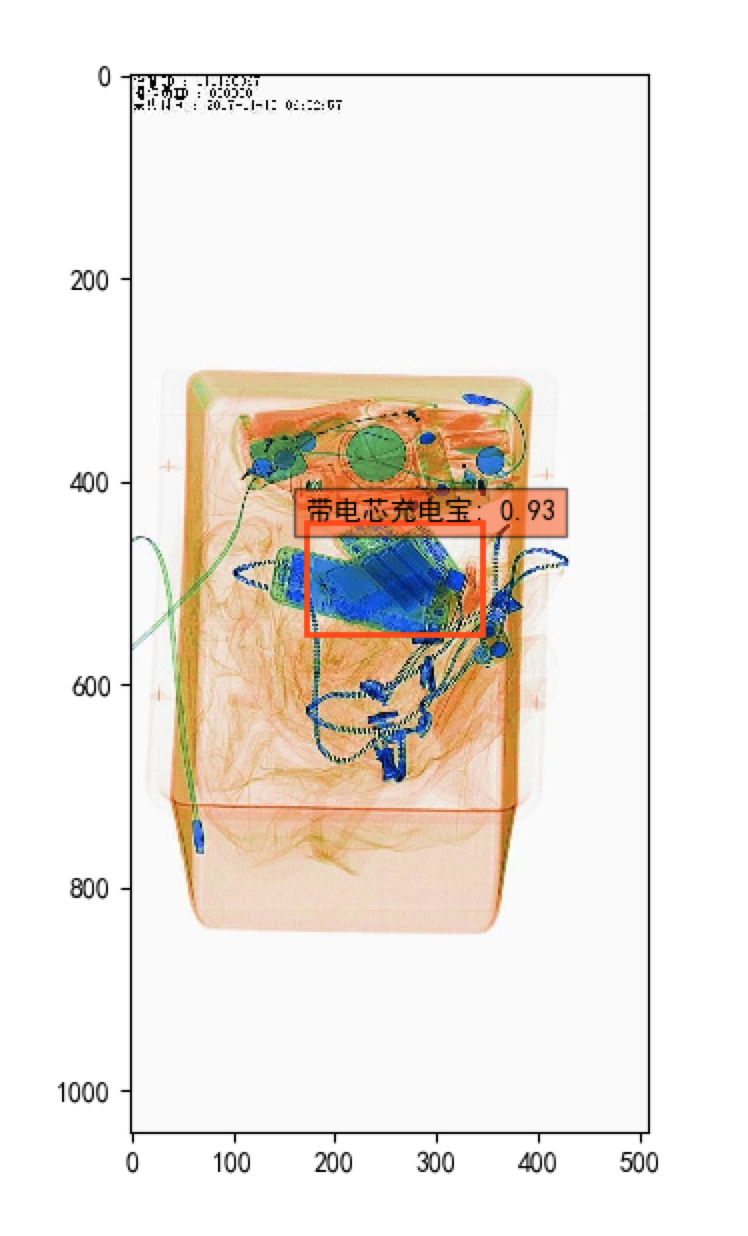


图4.3少部分遮挡测试结果

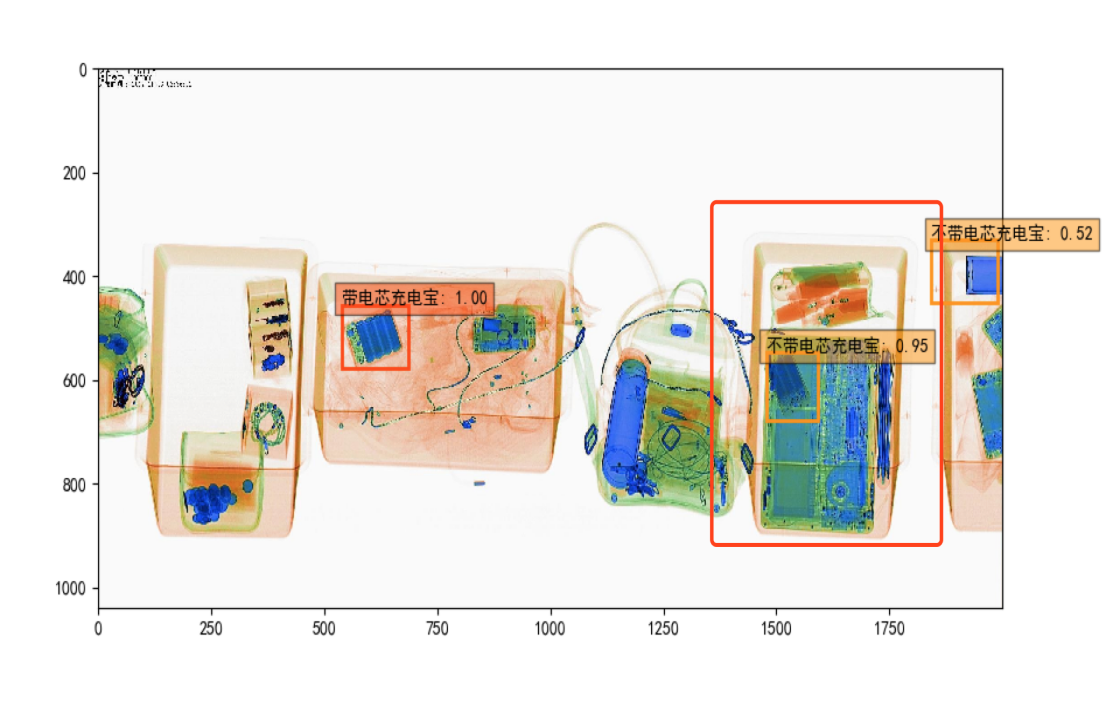


图4.4大部分遮挡测试结果

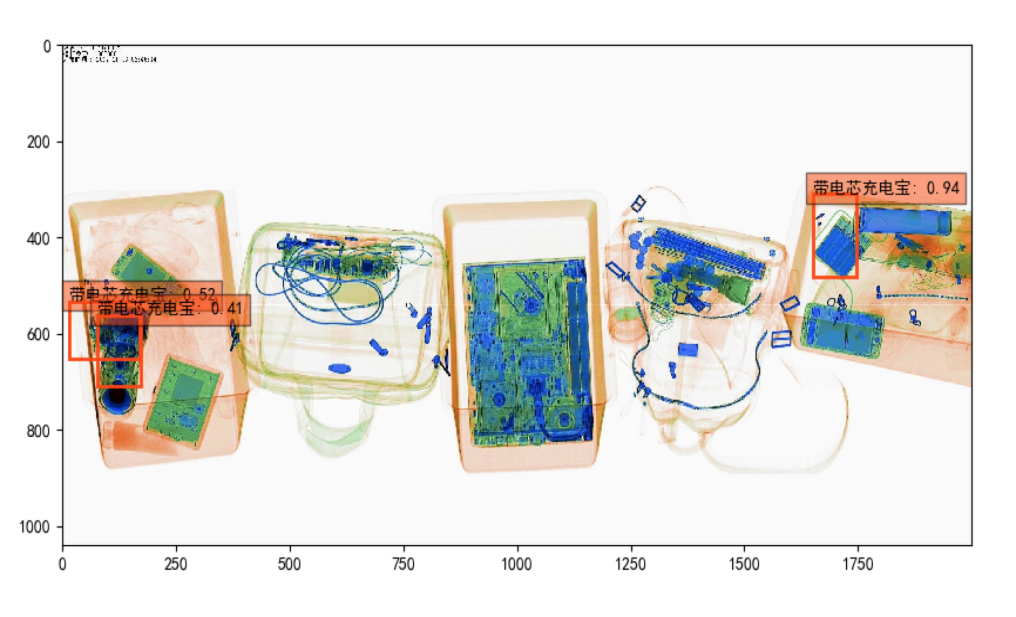


图4.5目标相互遮挡测试结果

从上面结果图中可以看出除了电池在图片的边缘和电池堆叠在一起预测的置信度比较低意外其他情况的预测置信度都达到了90%以上。但是对于大部分遮挡和完全遮挡以及目标物体相互遮挡的预测的效果还不是很理想，时好时坏不稳定。

**5 总结**

从mAP值可以看出模型训练效果还是很好的。我们把模型的输入值改成了512\*512。从提出SSD模型的论文看，输入大小为512\*512的效果明显比输入大小为300\*300的效果有明显的提高。虽然模型的mAP值看起来还不错，但是针对遮挡问题的解决我们也尝试过更改，但是最终没能成功。我们会继续深入研究模型，调参以更好的解决遮挡问题。其次就是我们训练的迭代次数太少了，只有30000次，原来代码对VOC0712和COCO数据集的训练迭代次数为120000和400000。通过这次的学习对SSD模型有了一个很深的认识，接下来会继续学习其模型的实现细节，以支持更多数据集的训练。