VOC数据集上训练并测试Faster RCNN

李德民 21210980045

2022年5月10日

**1 VOC数据集简介**

PASCAL VOC数据集初始发布于2005年，一般可用于目标检测、目标分类、目标分割等任务中，该数据集有20个分类标记，主要为人、动物、交通工具、室内设备等。本次实验采用的数据集是VOC2012，取其中80%的图片作为训练集，20%作为测试集。

**2 Faster RCNN算法简介**

在R-CNN和Fast RCNN的基础上，Ross B. Girshick在2016年提出了新的Faster RCNN。从结构层面看，Faster RCNN将特征抽取(feature extraction)，proposal提取，bounding box regression(rect refine)，classification整合在一个网络中，使得综合性能有较大提高，在检测速度方面的优势表现得尤其明显。

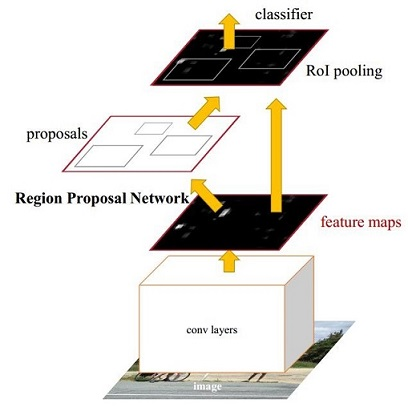


图1 Faster RCNN基本结构（来自原论文）

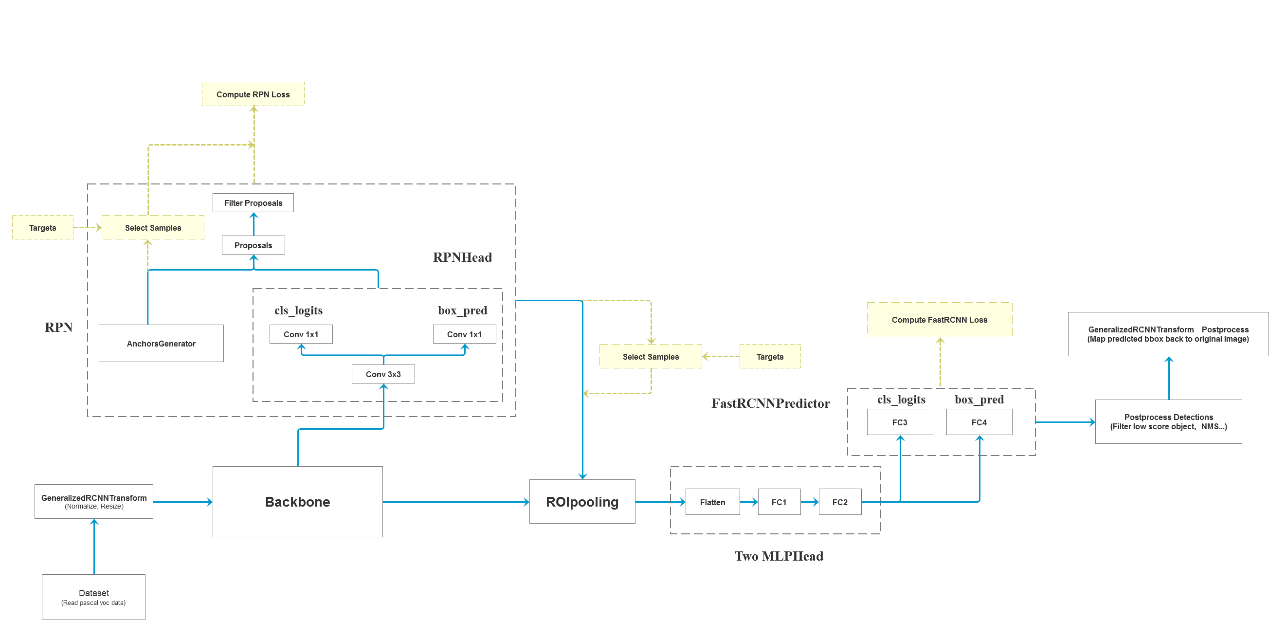
如图1，Faster RCNN其实可以分为4个主要内容：

1、Conv layers。作为一种CNN网络目标检测方法，Faster RCNN首先使用一组基础的conv+relu+pooling层提取image的feature maps。该feature maps被共享用于后续RPN层和全连接层。

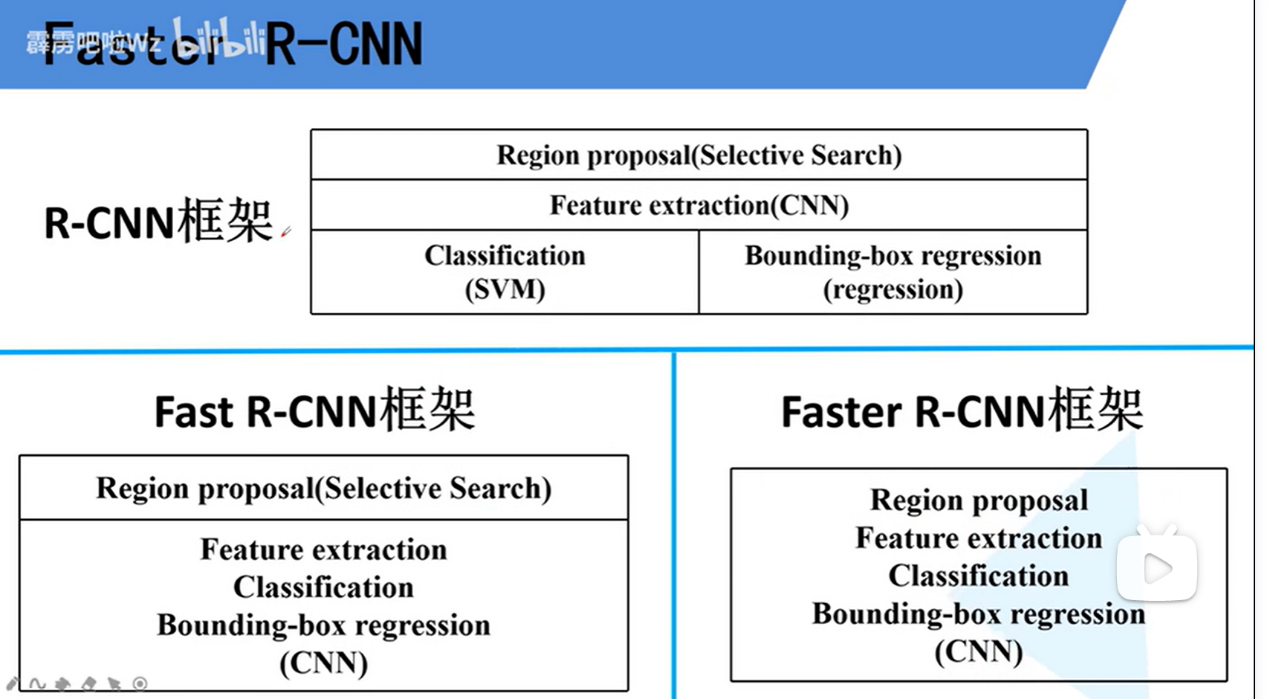
2、Region Proposal Networks。RPN网络用于生成region proposals。该层通过softmax判断anchors属于positive或者negative，再利用bounding box regression修正anchors获得精确的proposals。

3、Roi Pooling。该层收集输入的feature maps和proposals，综合这些信息后提取proposal feature maps，送入后续全连接层判定目标类别。

4、 Classification。利用proposal feature maps计算proposal的类别，同时再次bounding box regression获得检测框最终的精确位置。



图二：Faster RCNN流程结构详解



图三：Faster RCNN与前代模型结构对比

**3 实验设置**

本次实验采用的Faster RCNN网络共含有77784240个参数，它的网络结构如上所示。

采用SGD作为优化器，总计运行了6个 epoch，batch size为8，初始学习率设为 0.01 并逐渐衰减至 0.001089，如图4。6代训练共花费101小时，训练得到的最优模型为 “.\save\_weights” 文件夹下的 “resNetFpn-model-5.pth” 文件，结果已经较为稳定，且与well-trained模型的分类效果很接近。

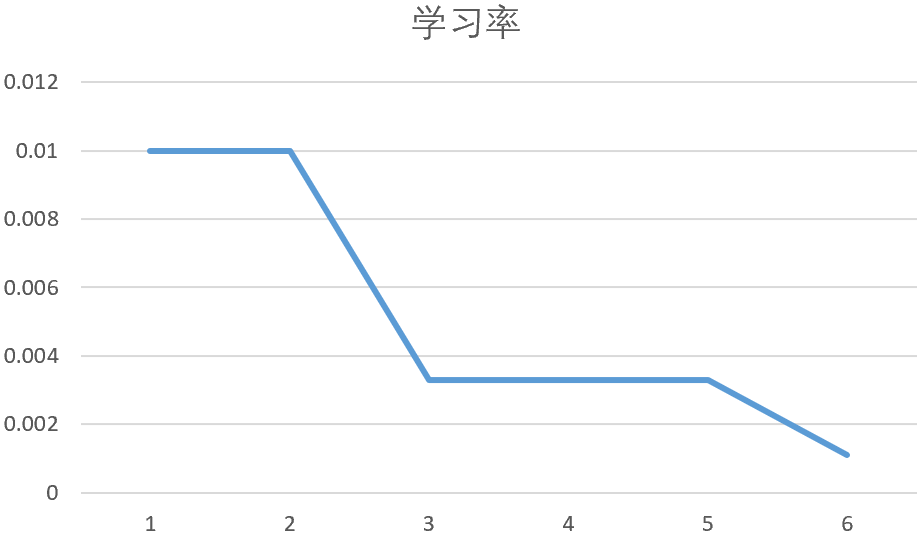


图4: 学习率

在fast rcnn中，损失分为两个部分，分类损失和边界框回归损失，其中分类损失使用的是softmax多分类交叉熵损失，边界框回归损失使用的是smooth L1损失。

**4 评价指标**

目标检测算法常用的性能评价指标包括：检测速度、交并比、精确率、召回率、平均精确率、平均精确率均值等。其中，检测速度 (FPS)，表示算法模型每秒钟所能检测到的图片数量。交并比 (IOU, Intersection over Union)，表示为算法模型产生的预测框和原始标注框的交集与并集的比值，它描述了两个区域的重合程度，其值越高代表算法模型的定位越准确。精确率 (Precision)，表示为分类正确的正样本个数与分类后判别为正样本个数的比值。召回率 (Recall)，表示为分类正确的正样本数与真正的正样本数的比值，衡量的是一个分类器能把所有的正样本都找出来的能力。在通常情况下，精确率越高，则召回率越低。平均精确率 (AP, Average Precision)，表示为 Precision-Recall 曲线下的面积，其值越大，表示分类器对某个类别的检测效果越好。平均精确率均值 (mAP)，表示为所有类别的平均精确率的均值。AR是IoU在[0.5,1.0]上所有recall的平均，是recall-IoU曲线所围成的面积的两倍。IOU=[0.5:0.95] 这个参数的含义是直接把mAP当成AP，然后再把IOU值大于0.5的AP(mAP)，以0.05的增量递增到0.95，也就是把(0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95)IOU值的AP(mAP)的平均值当成最终需要的AP(at IoU=.50:.05:.95)。maxDets=[1,10,100]，该指标的意思是分别保留测试集的每张图上置信度排名第1、前10、前100的预测框，将预测框和真实框比对来计算AP、AR等值。图5和图6展示了训练后模型的主要指标。

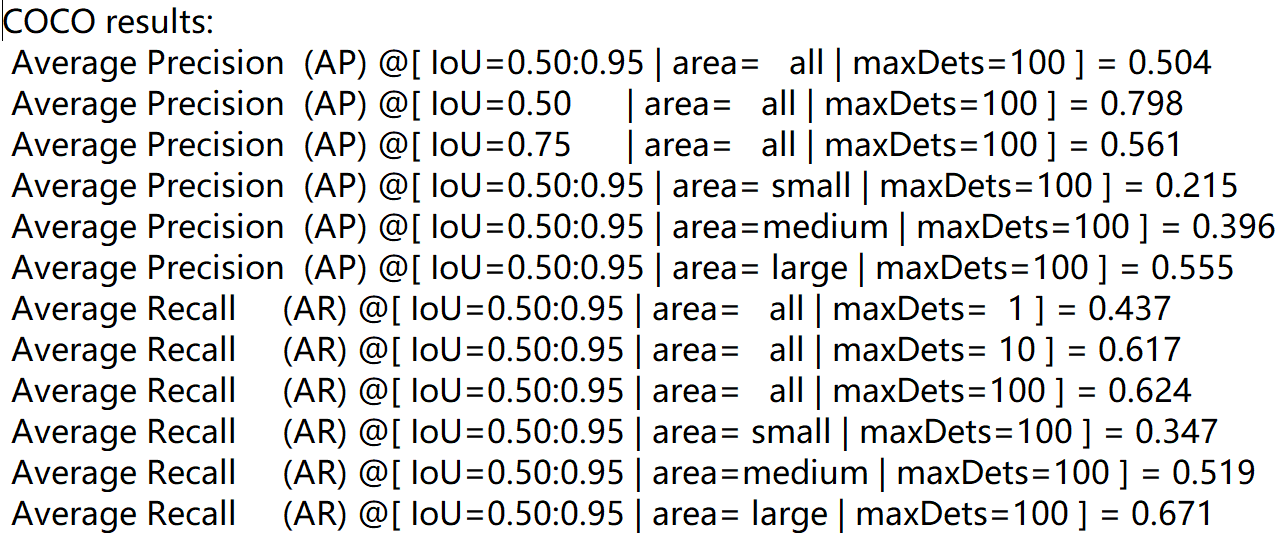


图5：使用COCOAPI生成的AP和AR评价结果

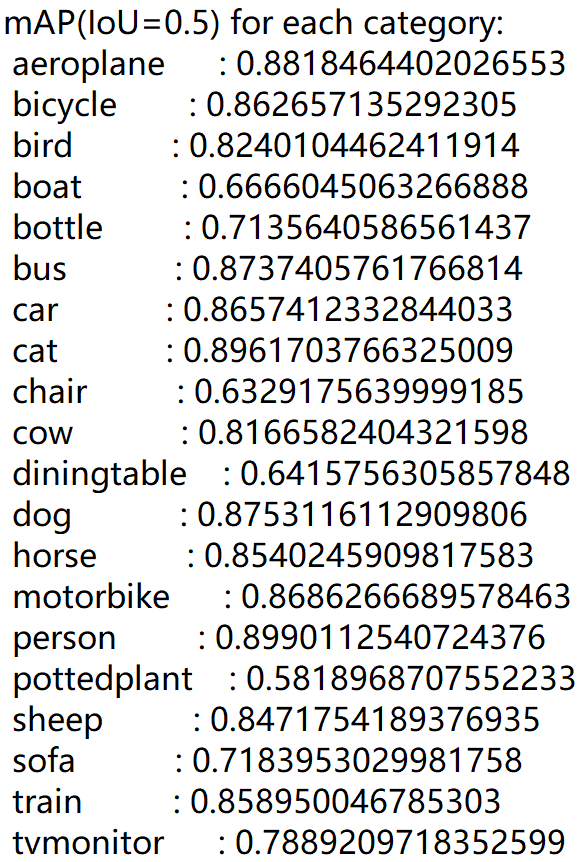
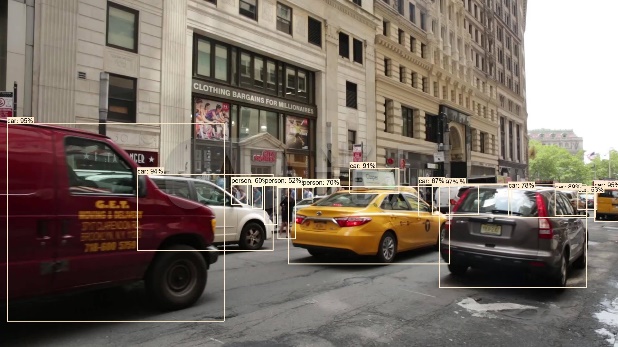


图6：在Iou阈值为0.5下的各类别平均精确率均值

**5 检测结果可视化**

接下来导入训练得到的“resNetFpn-model-5.pth”文件，使用“predict.py”文件来可视化三张不在VOC数据集内，但是包含VOC 中类别物体的图像的检测结果。选用的图片包含人、车、狗、桌子、椅子共计五个类别的物体，检测效果如图7所示。整体而言，模型对各类物体识别效果较好，但并不能准确地识别远处较小的车辆及行人。



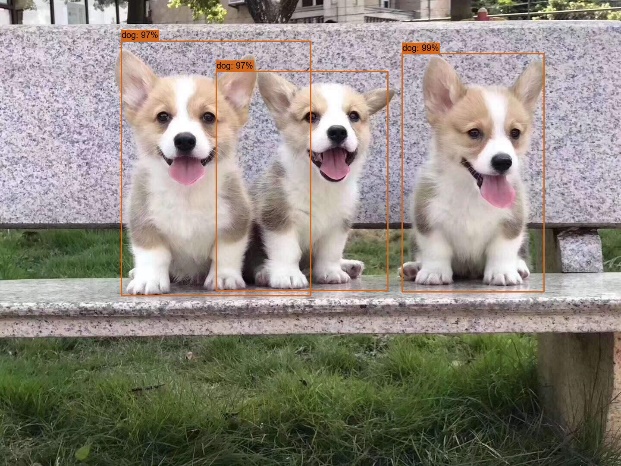
 

图7：检测结果