# 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。

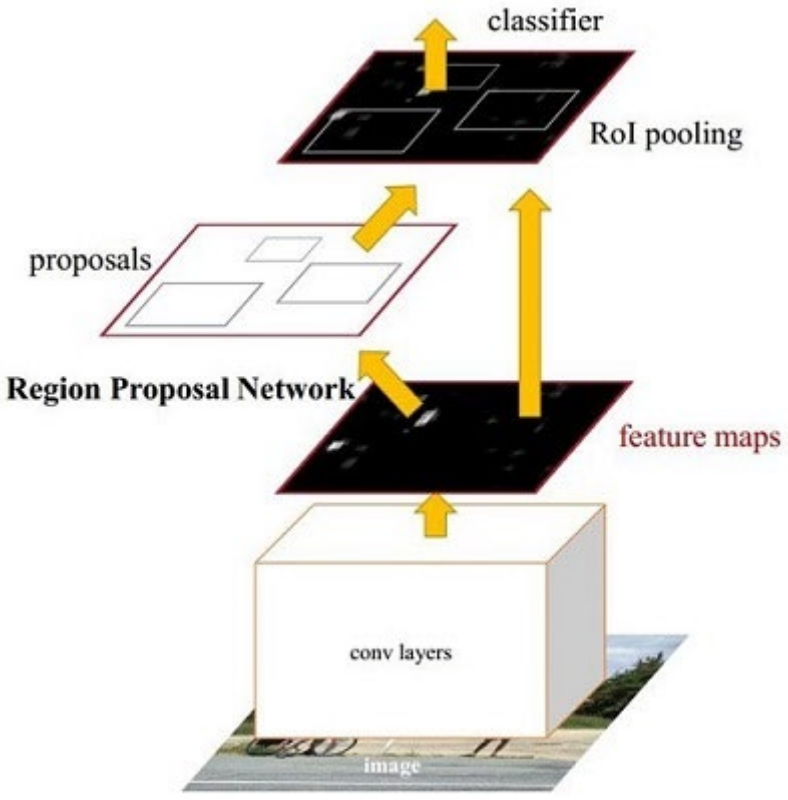
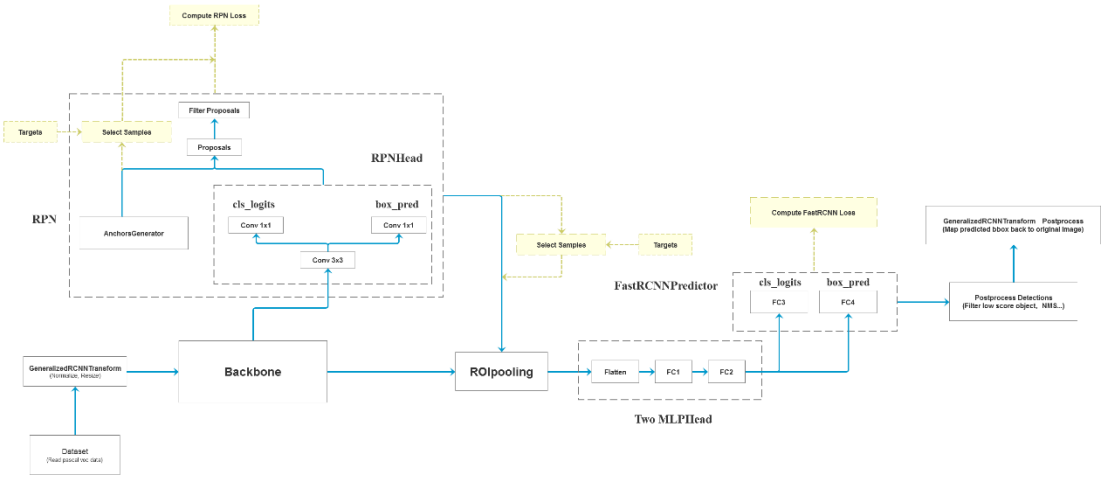


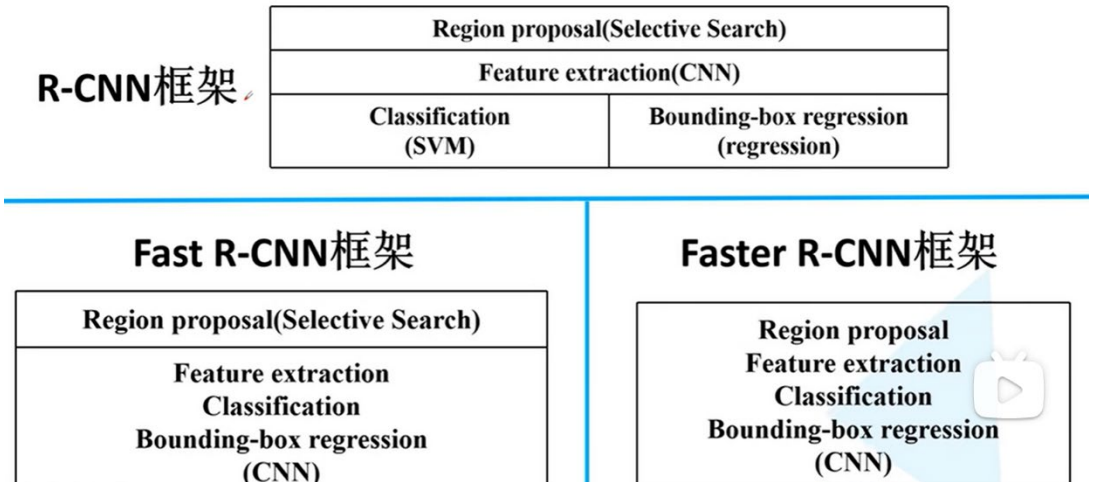

图 1 Faster RCNN 基本结构（来自原论文）

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。

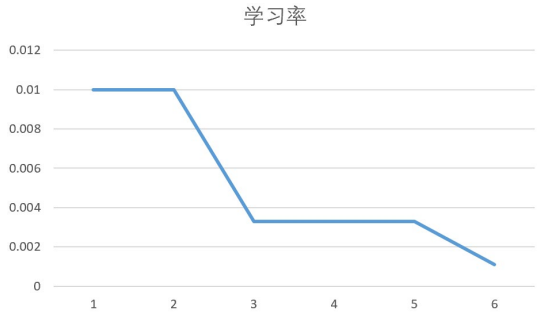



图 4: 学习率

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。

```
COCO results:
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.504
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.798
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.561
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.215
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.396
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.555
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.437
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.617
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.624
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.347
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.519
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.671
```

图 5：使用 COCOAPI 生成的 AP 和 AR 评价结果

```
mAP(IoU=0.5) for each category:
 aeroplane      : 0.8818464402026553
 bicycle        : 0.862657135292305
 bird           : 0.8240104462411914
 boat           : 0.6666045063266888
 bottle         : 0.7135640586561437
 bus            : 0.8737405761766814
 car            : 0.8657412332844033
 cat            : 0.8961703766325009
 chair          : 0.6329175639999185
 cow            : 0.8166582404321598
 diningtable    : 0.6415756305857848
 dog            : 0.8753116112909806
 horse          : 0.8540245909817583
 motorbike      : 0.8686266689578463
 person         : 0.8990112540724376
 pottedplant    : 0.5818968707552233
 sheep          : 0.8471754189376935
 sofa           : 0.7183953029981758
 train          : 0.858950046785303
 tvmonitor      : 0.7889209718352599
```

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。

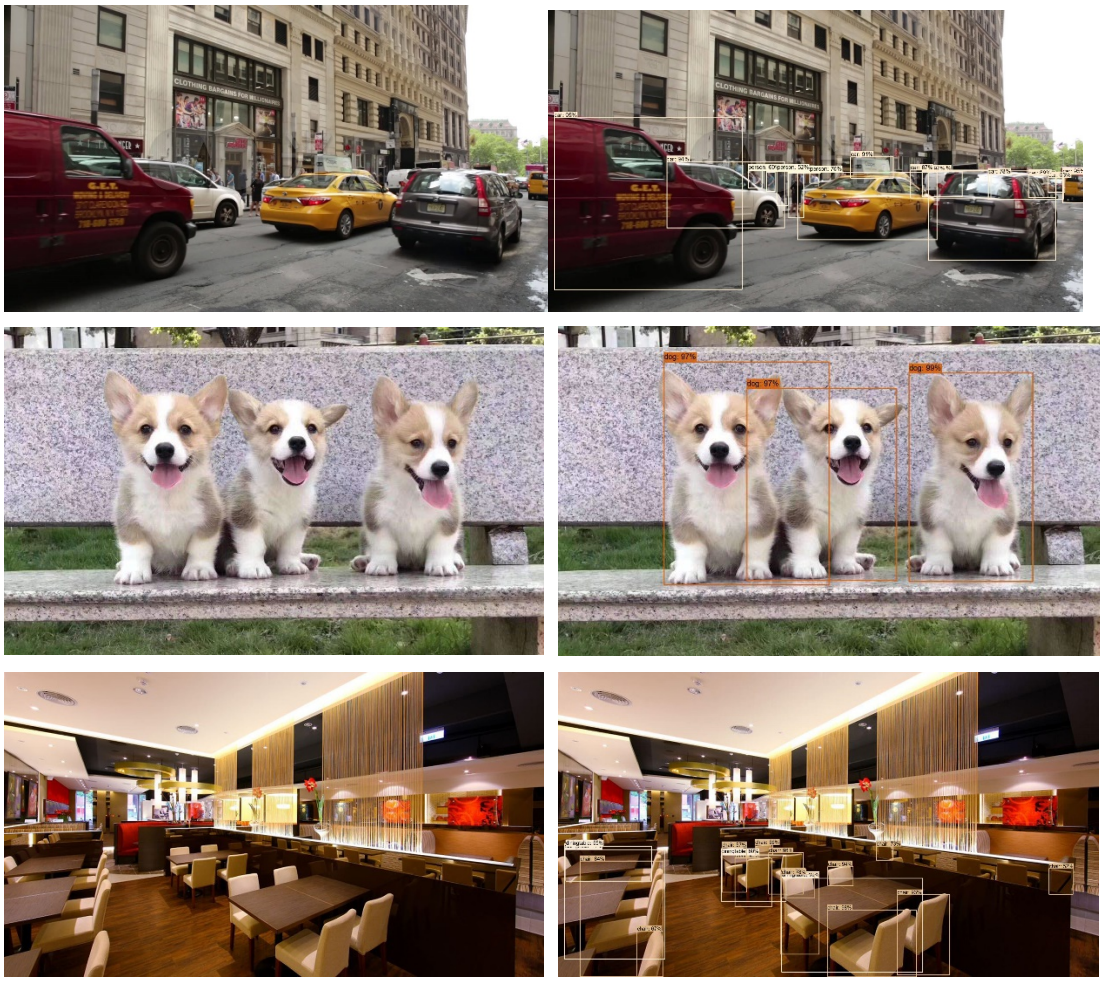


图 7：检测结果