目标检测文献综述

1.RCNN：[Rich feature hierarchies for accurate object detection and semantic segmentation](file:///Users/lianlian/Documents/machinelearning/neuralNetwork/CV/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.pdf)

key point: （1） 使用CNN进行自底向上的定位和分割物体

（2） 标签数据少时，使用监督预训练网络之后再对网络进行fine tune

步骤：（1）生成与目标种类无关的region proposals

（2）CNN网络提取特征

（3）针对每个类的线性SVM二分类器

对每个步骤进行详细说明：

region proposals：使用selective search搜索rp，提取两千个region proposals

提取特征：采用CNN提取目标区域中图像的特征，首先第一步需要将调整图像尺寸使其能够输入到CNN中，采用了直接对图像进行缩放的方式，输入尺寸为227\*227

检测：讲每个候选区域的特征输入到SVM中进行检测，得到该类的得分，之后根据IOU值使用极大值抑制的方法去除掉IOU大于某个阈值的region，为什么不直接用神经网络softmax层做分类而是用SVM？因为SVM和CNN对正负样本的定时方式不同，CNN在训练时对标注比较宽松，SVM则会比较严格，只有当bounding box将物体都包进去了才将它标位物体类别。在论文中好像有提到CNN中IOU阈值设为0.5，SVM阈值设为0.3.

问题：相当于每检测一幅图像需要输入2000个region图像数据用CNN对它做特征提取，大部分时间都花在这一步，之后再对每个region做21个svm二分类，不过是线性运算较快。

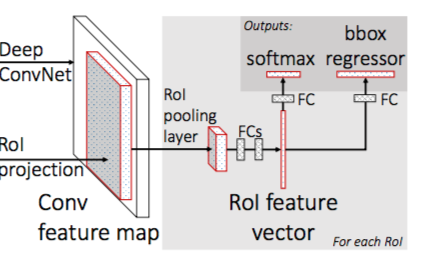
2.fast-rcnn: [Fast R-CNN](file:///Users/lianlian/Documents/machinelearning/neuralNetwork/CV/Girshick_Fast_R-CNN_ICCV_2015_paper.pdf)

借鉴了SPPnet，速度方面进行了改进，同时增加了精度，RCNN速度慢的问题在于没有共享从同一幅图中提取出来的特征，中间进行了多次重复计算，SPPnet最早提出共享计算的方法，和RCNN一样，SPP依然一开始对图像进行selective search，选出两千个候选框，然后用卷积网络对输入图像进行特征提取，得到feature maps，之后根据某种对应方式得到feature maps 上对应的候选框，然后再对候选框进行SVM分类，和RCNN相比，共享了CNN的计算结果。

Fast-RCNN开始同样是将输入图像和选择性搜索得到的ROI输入到CNN中，提取得到feature maps，然后对每个候选框经过ROI pooling layer和全连接层，得到一固定长度的特征向量，然后将每个ROI的特征向量分别输入一个softmax层得到分类结果和一个regressor得到检测区域的回归结果。

和SPPnet最大的不同在于ROI pooling layer：

每一个RoI都有一个四元组（r,c,h,w）表示，其中（r，c）表示左上角，而（h，w）则代表高度和宽度。这一层使用最大池化（max pooling）来将RoI区域转化成固定大小的H\*W的特征图。假设一个RoI的窗口大小为h\*w,则转换成H\*W之后，每一个网格都是一个h/H \* w/W大小的子网，利用最大池化将这个子网中的值映射到H\*W窗口即可。



之后的训练步骤使用了多任务损失函数来计算整体的损失，是一个end-to-end的过程。

3.faster-cnn：[Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks](5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf)

fast-rcnn实现了特征参数的贡献，对proposals的计算成为限制物体检测的主要瓶颈，二faster-rcnn主要贡献在于使用RPN网络生成region proposals ，大大减少了计算proposals的损耗。