目标检测

吸取论文精华

R-CNN

R-CNN算法是深度神经网络在目标检测领域最早的实践，其主要过程分为四步：

- 利用Selective search算法在图像中生成1000-2000个候选区域
- 归一化为统一尺寸后送入卷积神经网络提取特征
- 神经网路输出的特征送入每一类对应的SVM二分类器分类
- 使用回归器精细修正候选框位置。

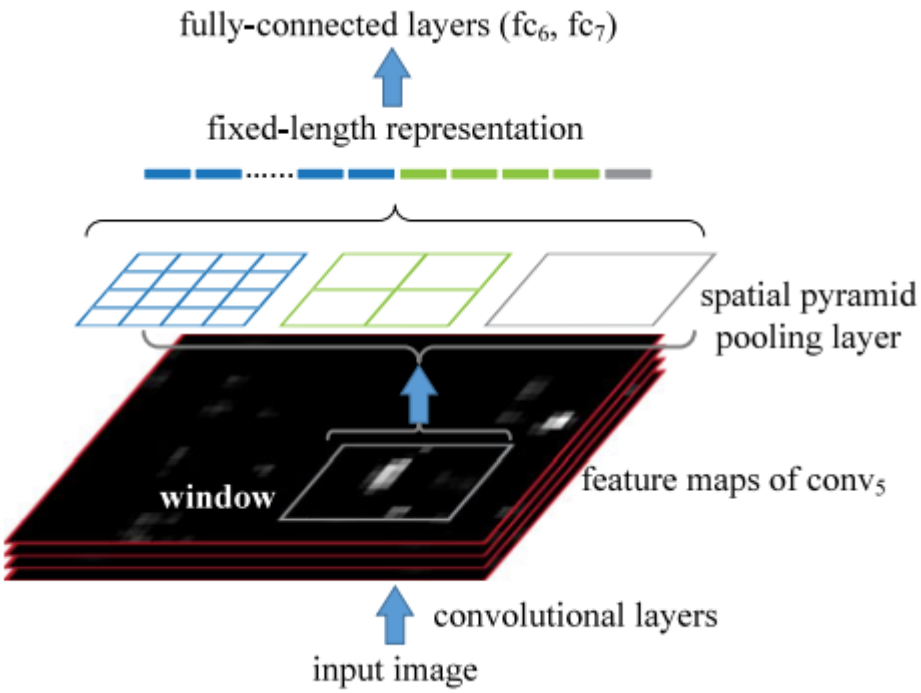
R-CNN的每个候选框都需要归一化后送到神经网路提取特征，计算量大。

SPP-net

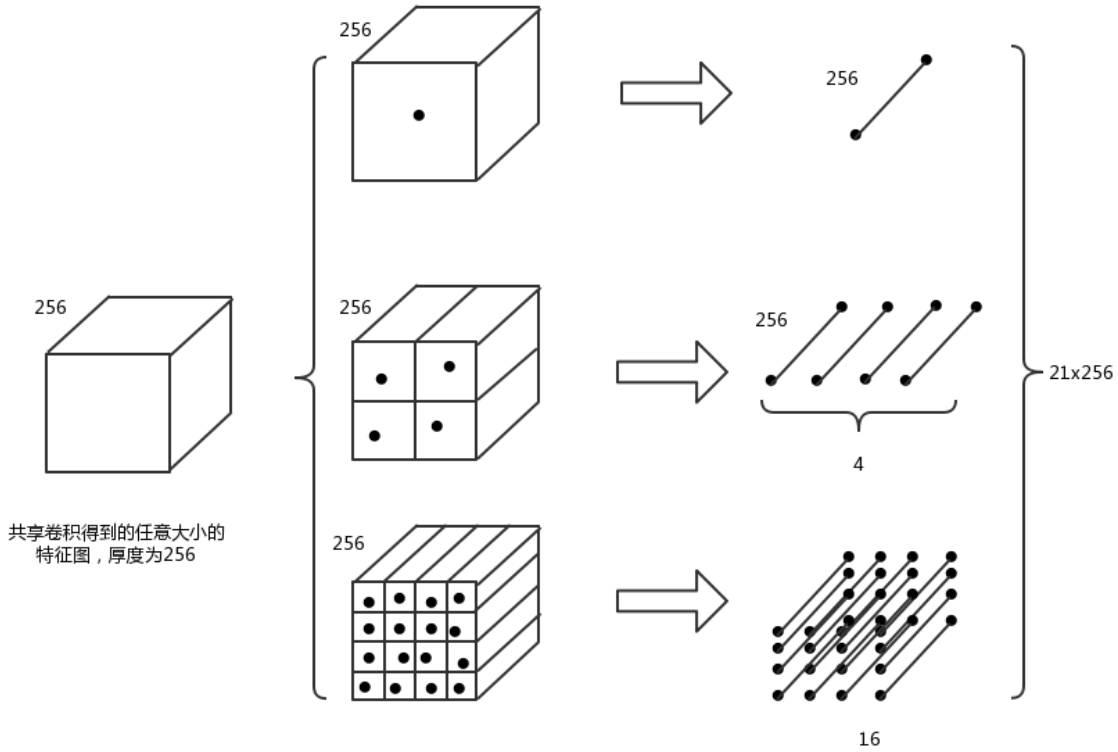
论文：Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition

SPPnet论文总结

本文在R-CNN的基础上思考卷积神经网络是否需要对输入fixed-size。卷积层、池化层对输入大小不敏感，输入不同尺寸对应输出不同大小；全连接层对输入尺寸有限制。所以归一化的过程可以放在全连接层前。而本文提出的归一化工具就是**Spatial Pyramid Pooling（SPP）**，即空间金字塔池化。SPP不仅在目标检测领域实用，而且可以广泛应用在计算机视觉各领域，作为fixed-size的一种解决方案。



基础网络最后卷积层输出的特征图（或者特征图的一部分）送入SPP；池化得到1*1，2*2，4*4的特征矩阵；拉平得到（1+4+16）的矩阵。下图展示了一个h*w*256的特征图转换为21*256的矩阵的SPP过程。SPP的巧妙之处在于巧用池化层将不确定大小的输入转变为相同大小的输出，优于暴力拉伸缩放。



Fast R-CNN

论文：Fast R-CNN

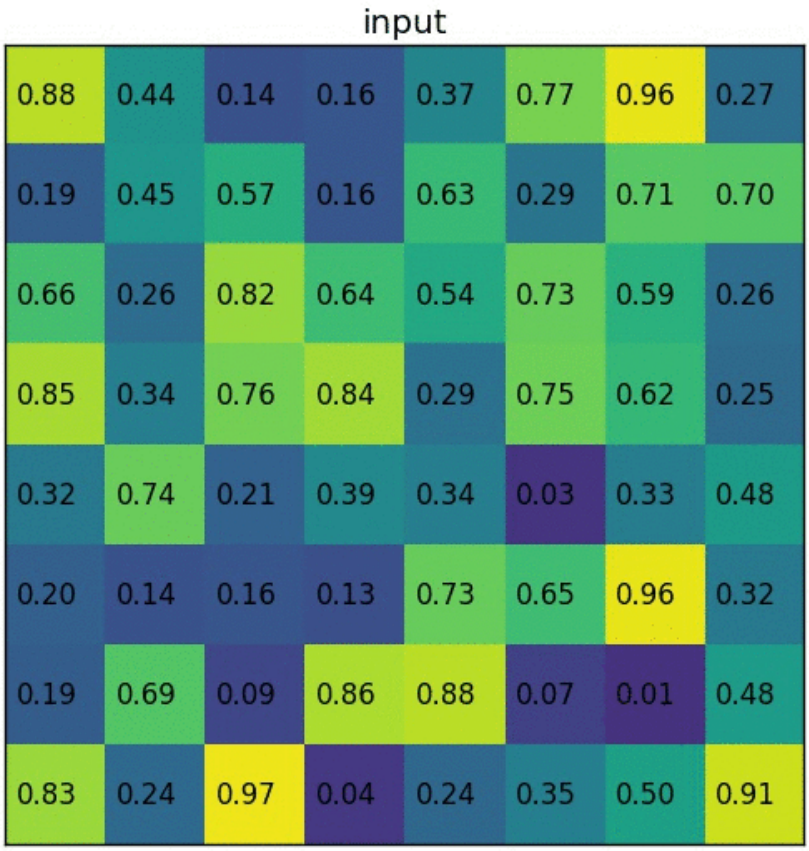
Region of interest pooling explained

Fast R-CNN在R-CNN和SPP-net的基础上做了两点改进：

- 简化空间金字塔池化，提出**ROI pooling**（Region of interest pooling）
- 用卷积神经网络取代类别分类和位置回归，提出了多任务损失函数（即包括类别误差和位置误差）

ROI是表示预选框位置的五维矩阵（index, x1, y1, x2, y2），即图片索引和左上角右下角坐标。ROI pooling过程分为以下四步，下图展示不同尺寸的输入通过ROI pooling得到相同大小的特征图，：

1. 根据ROIs和特征图，得到候选框的信息
2. 将候选框分割为预设尺寸的小块
3. 寻找每个小块的最大值
4. 得到输出矩阵

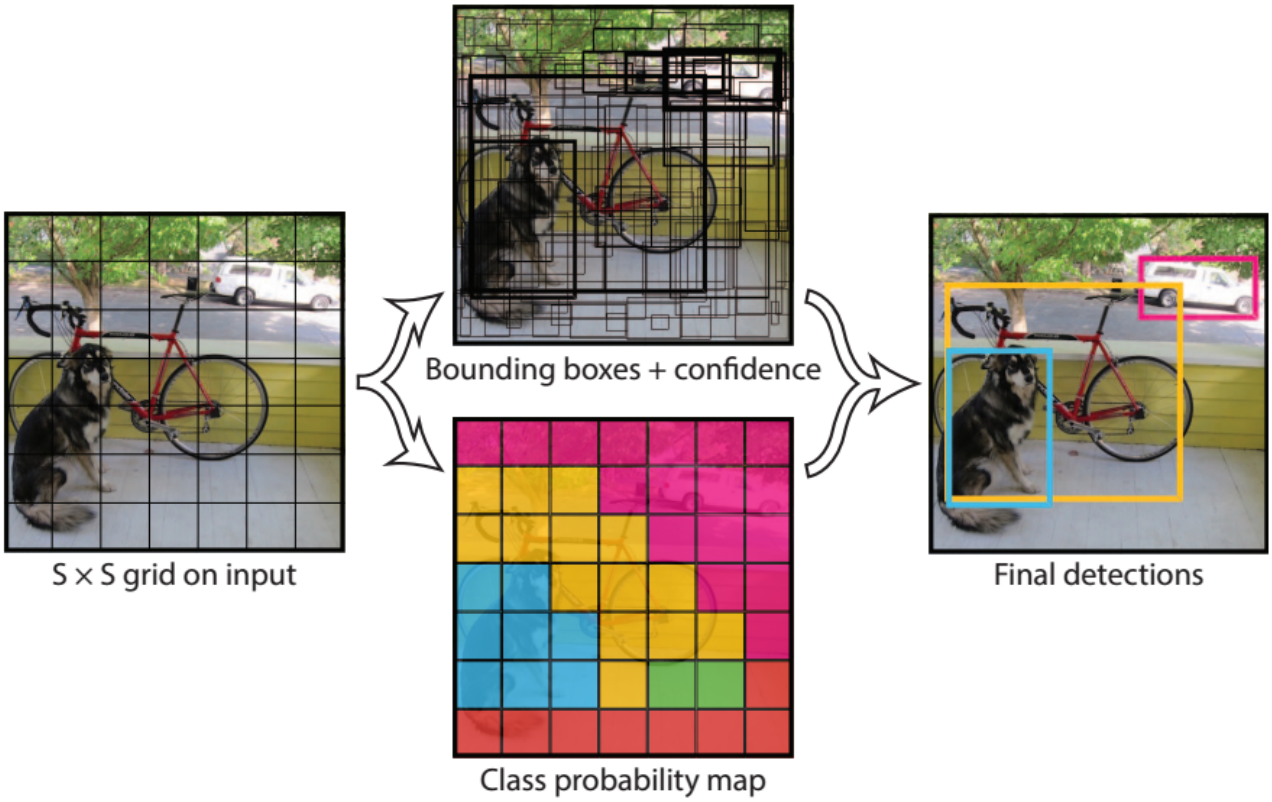


YOLO

论文：You only look once: Unified, real-time object detection

以YOLO为代表的one-stage目标检测算法认为，单独的卷积神经网络就可以完成分类和定位两个任务。YOLO利用卷积层提取信息，全连接层输出类别与位置。检测过程将输入的图片划分为S*S个cell，每个cell负责预测中心点落在该cell的物体的类别和位置。一个cell预测B个bounding box，B个bbox共享类别概率，输出一组5*B+class_num维结果。以B=2，class_num=5为例，每个cell输出为：

$$[x_1, y_1, w_1, h_1, c_1, x_2, y_2, w_2, h_2, c_2, class_1 \dots . class_5]$$



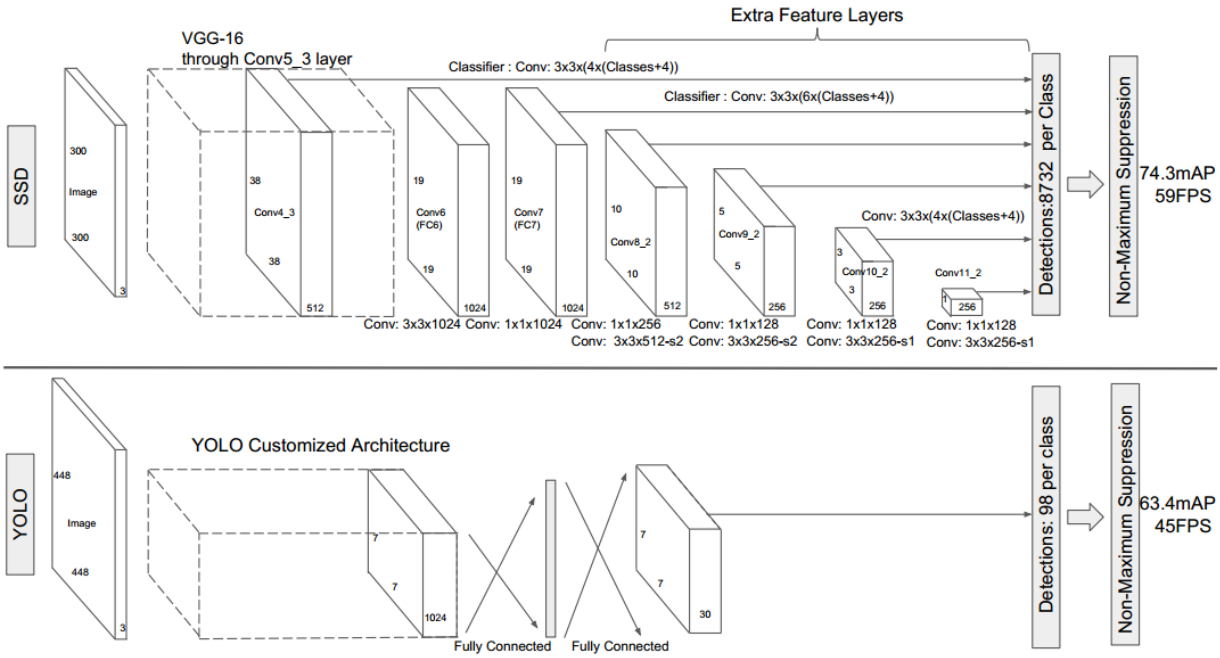
YOLO另一个创新是损失函数。传统的目标检测算法中，类别是分类问题，定位是回归问题；而YOLO将两者统一为回归问题，最近预测每个类别的概率而不是二分类。

SSD

论文：SSD: Single shot multibox detector

SSD与YOLO都采用单个神经网络实现分类定位。相对于YOLO，SSD作了一下改进：

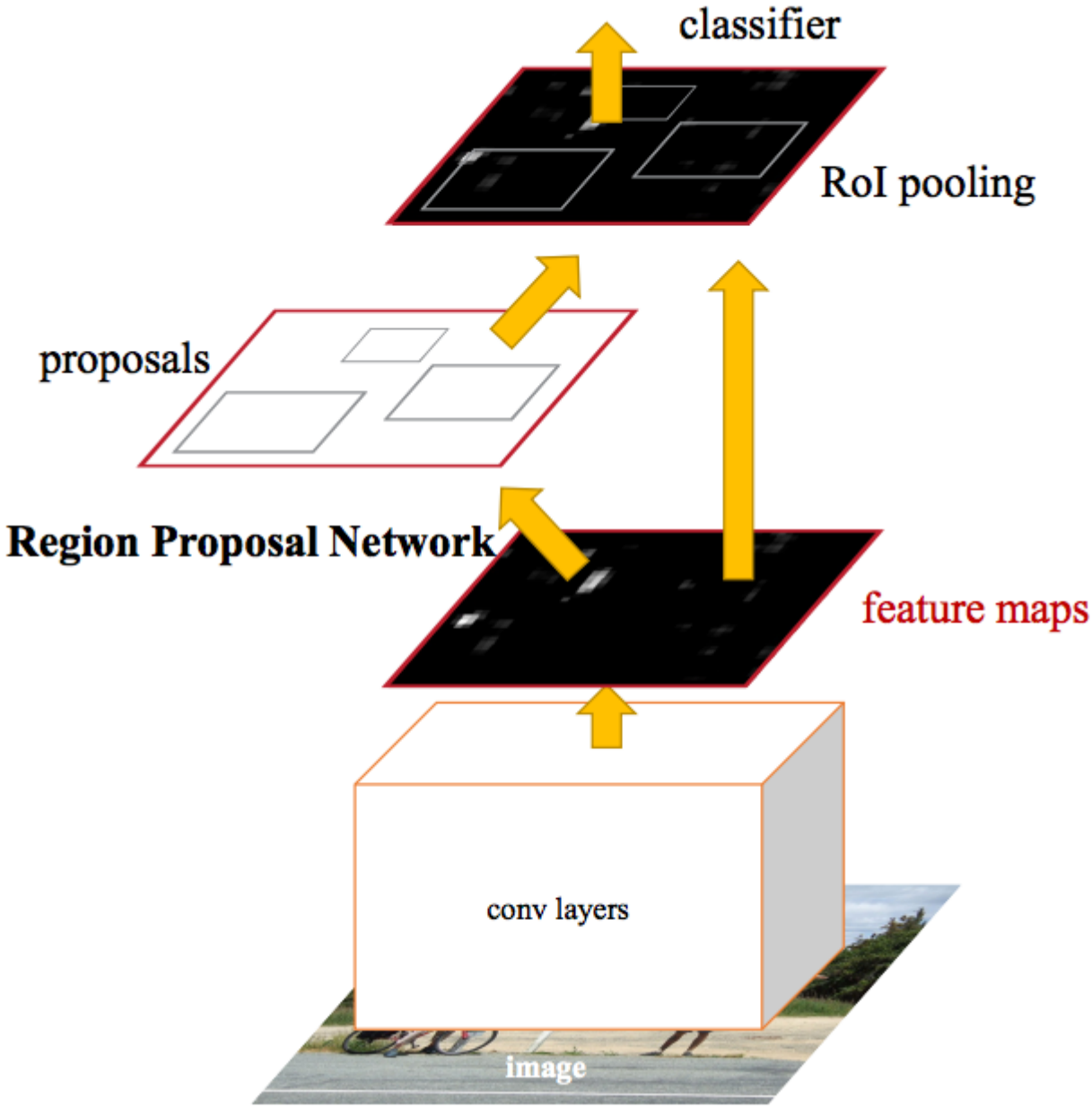
1. 借鉴ResNet的思路，将浅层的特征图连接到最后一层，显著提高了模型的准确度，特别是对小目标的识别能力；
2. 以Faster R-CNN的Anchor Box代替YOLO的Bounding Box；
3. 网络中部分使用DeepLab提出的空洞卷积。



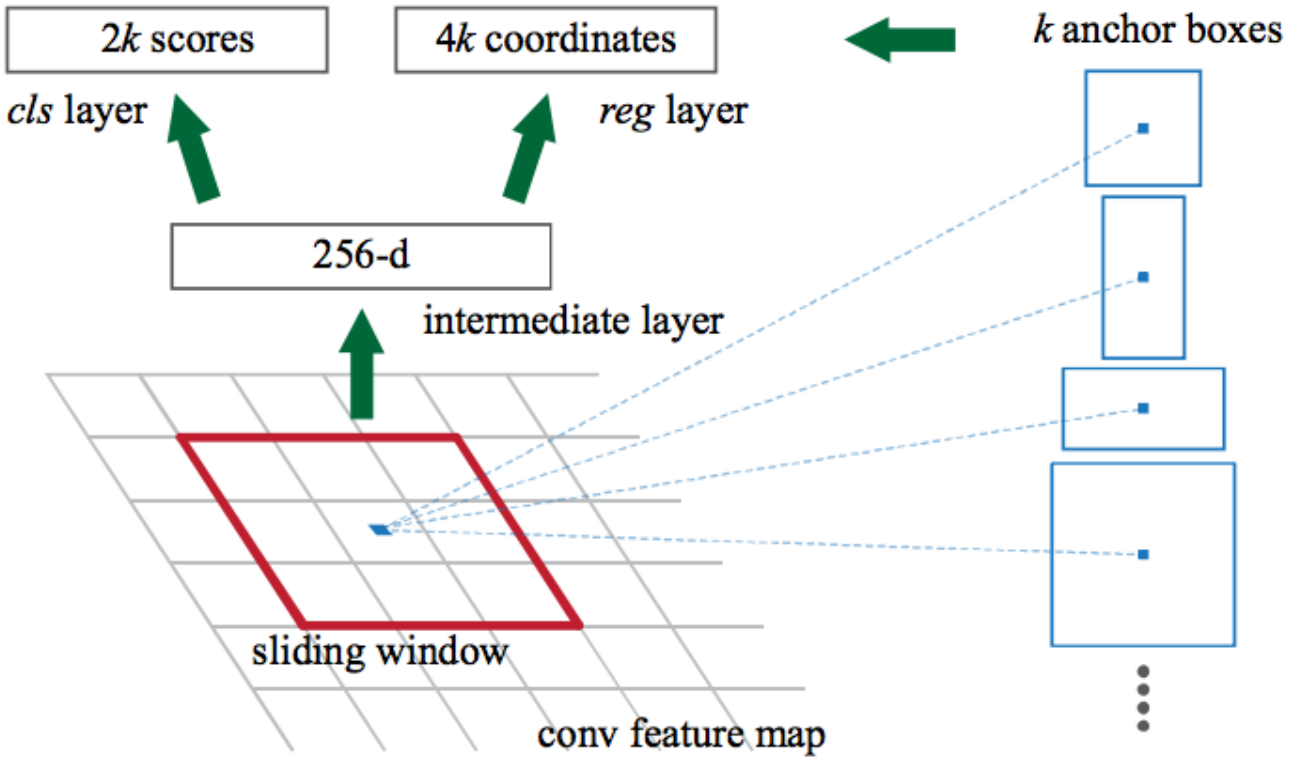
Faster R-CNN

论文：Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

从R-CNN升级到Fast R-CNN，整个算法还剩下一个瓶颈：生成候选框。Faster R-CNN放弃了传统候选框生成算法selective Search，设计了**RPN**（Region Proposal Network）算法，如图。



神经网络输出的特征图，送入RPN，就这么简单的得到候选框。RPN的输入与ROI pooling的输入相同，而且这两层均为单层卷积层（实际上还有1*1的卷积层），整个算法变成全卷积结构。故对于一张图片，只需要运行一次神经网络，节省大量计算开销。RPN生成候选框的过程如下：对于特征图的每个点，生成k个anchor boxes（一般设置3种scale和3种aspect rations，共9个anchor boxes）。每个anchor box预测6个参数，2个为存在物体和不存在物体的概率，另外4个是坐标。如果送入RPN的特征图尺寸为W*H，则预测W*H*k个anchor boxes，可以通过nms等方法过滤后送入ROI pooling。



YOLO v2v3

论文：YOLO9000: better, faster, stronger

R-FCN

FPN