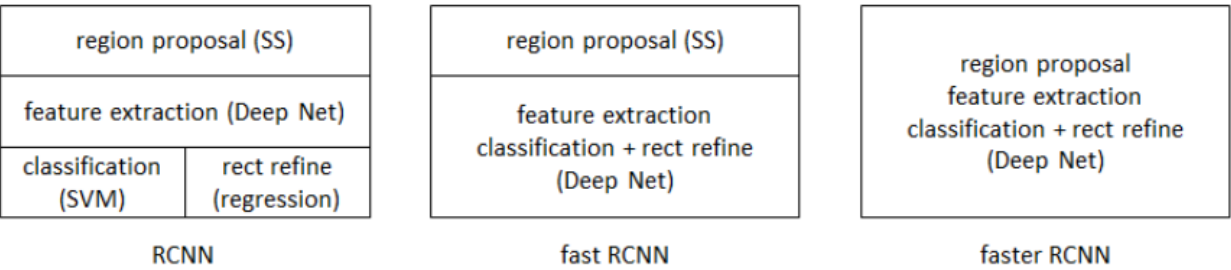


Faster RCNN算法

思想

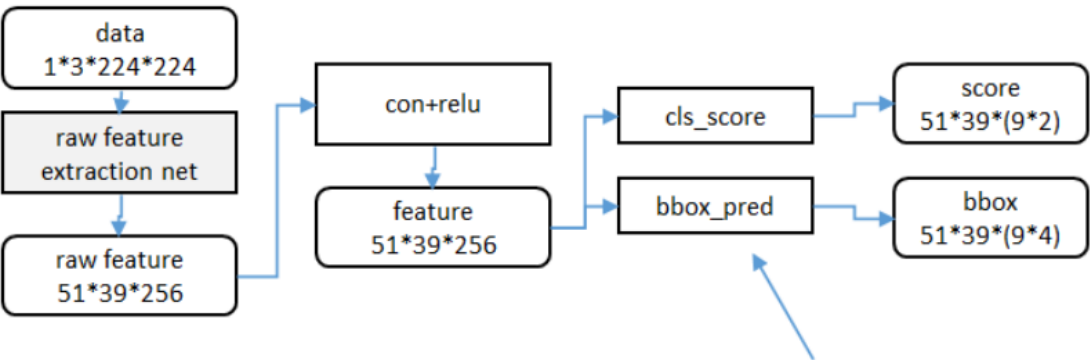
- 将目标检测的四个步骤（候选区域生成、特征提取、分类、位置精修）被统一到一个深度网络框架中。



- Faster RCNN可以简单地看做“区域生成网络+Fast RCNN”的系统，用区域生成网络代替Fast RCNN中的Selective Search方法。

区域生成网络：结构

- 基本思想：在提取好的特征图上，对所有可能得候选框进行判别。由于后续还有位置精修步骤，所以候选框实际比较稀疏。



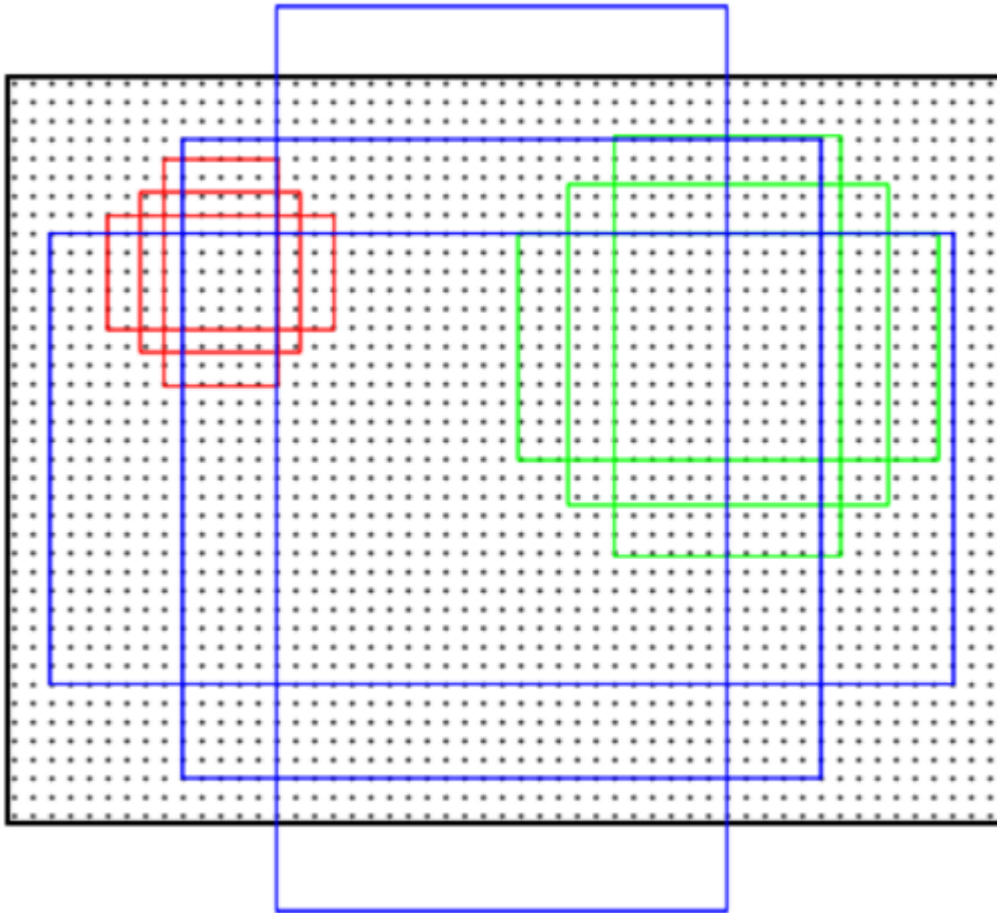
Consider 9 ‘anchors’ on each of the 51*39 positions

特征提取

- 原始特征提取（灰色框）包含若干层conv+relu，直接套用ImageNet上常见的分类网络即可。
- 额外添加一个conv+relu层，输出**51*39*256**维特征。

候选区域 (anchor)

- 特征可以看做一个尺度51*39的256通道图像，对于该图像的每一个位置，考虑9个可能的候选窗口：三种面积{ 128^2, 256^2, 512^2 } X 三种比例 { 1:1, 1:2, 2:1 }。这些候选窗口称为anchors。



- 在整个Faster R-CNN算法中，有三种尺度。
 - 原图尺度：原始输入的大小。
 - 归一化尺度：输入特征提取网络的大小。
 - 网络输入尺度：输入特征检测网络的大小。

窗口分类和位置精修

- 分类层 (cls_score) 输出每一个位置上，9个anchor属于前景和背景的概率。
- 窗口回归层 (bbox_pred) 输出每一个位置上，9个anchor对应窗口应该平移缩放参数。
- 对于每一个位置来说，分类层从256维特征中输出属于前景和背景的概率。
- 窗口回归层从256维特征中输出4个平移缩放参数。
- 就局部来说，这两层全连接网络。就全局来说，由于网络在所有位置的参数相同，所以实际尺寸为1x1的卷积网络实现。

区域生成网络：训练

样本

- 考察训练集中的每张图像：
 - 对每个标定的真值候选区域，与其重叠比例最大的anchor记为前景样本。
 - 对（1）剩余的anchor，如果其与某个标定重叠比例大于0.7，记为前景样本。如果其与任意一个标定的重叠比例都小于0.3，记为背景样本。

- 对 (1) (2) 剩余的anchor，弃去不用。
- 跨越图像边界的anchor弃去不用。

代价函数

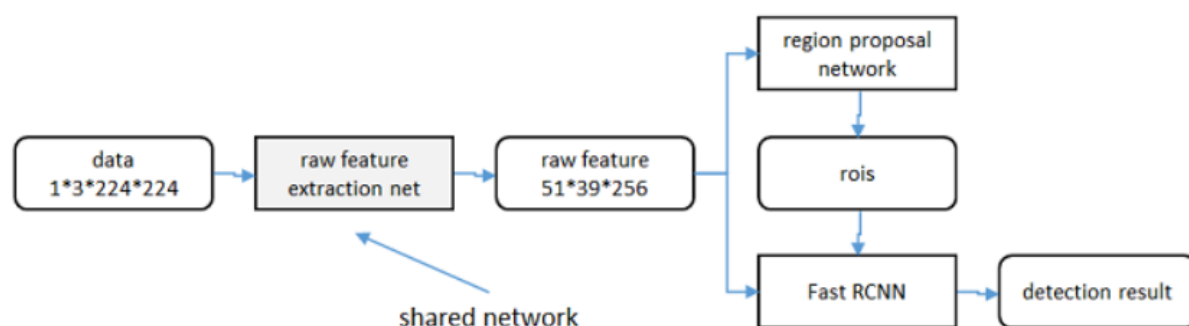
- 同时最小化两种代价：
 - 分类误差
 - 前景样本的窗口位置偏差

超参数

- 原始特征提取网络使用ImageNet的分类样本初始化，其余新增层随机初始化。
- 每个mini-batch包含从一张图像中提取的256个anchor，前景背景样本1:1。
- 前60K迭代，学习率为0.001，后20K迭代，学习率为0.0001。
- momentum设置为0.9，weight decay设置为0.0005。

共享特征

- 区域生成网络（RPN）和Fast RCNN都需要一个原始特征提取网络（灰色）。这个网络使用ImageNet的分类库得到初始参数 W_0 。



轮流训练

- 从 W_0 开始，训练RPN。用RPN提取训练集上的候选区域。
- 从 W_0 开始，用候选区域训练Fast RCNN，参数记为 W_1 。
- 从 W_1 开始，训练RPN...
- 具体操作时，仅执行两次迭代，并在训练时冻结了部分层。

近似联合训练

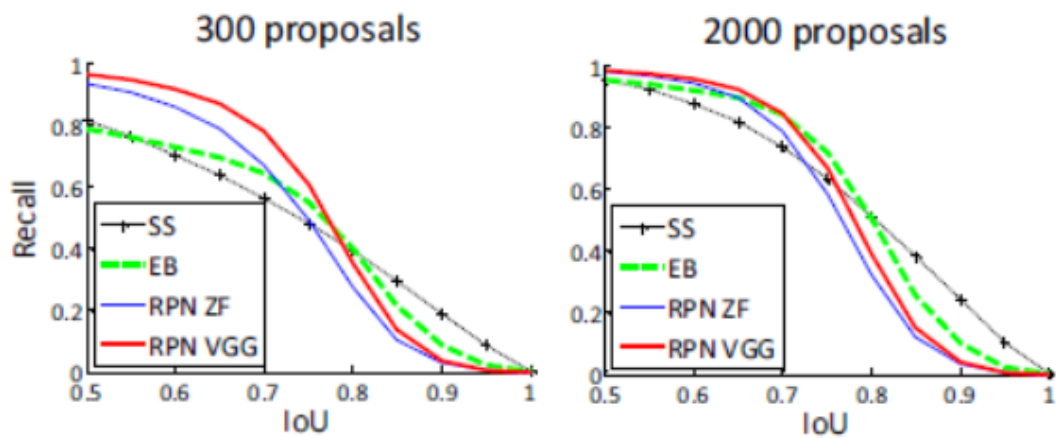
- 直接在上图结构上训练。在backward计算梯度时，把提取的ROI区域当作固定值看待。在backward更新参数时，来自RPN和来自Fast RCNN的增量合并输入原始特征提取层。

联合训练

- 直接在上图结构训练。但在backward计算梯度时，要考虑ROI区域的变化影响。

实验

- 与Selective Search方法相比，当每张图生成的候选区域从2000减少到300时，本文RPN方法（红蓝）的召回率下降不大，说明RPN方法的目的性更明确。



- 使用更大的COCO库训练，直接在PASCAL VOC上测试，准确率提升6%。说明Faster RCNN迁移性良好，没有过拟合。

training data	2007 test
VOC07	69.9
VOC07+12	73.2
VOC07++12	-
COCO (no VOC)	76.1
COCO+VOC07+12	78.8
COCO+VOC07++12	-