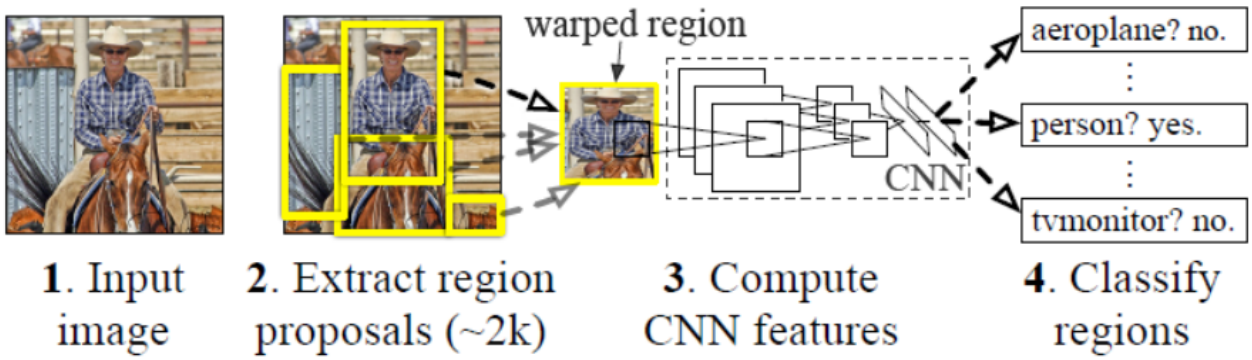
RCNN

- R-CNN采用`AlexNet`。
- R-CNN采用`Selective Search`技术生成Region Proposal。
- R-CNN在`ImageNet`上先进行预训练，然后利用成熟的权重参数在`PASCAL VOC`数据集上进行fine-tune。
- R-CNN用`CNN`提取特征，然后用一系列的`SVM`做类别预测。
- R-CNN的Bbox位置回归基于`DPM`的灵感，自己训练了一个线性回归模型。
- R-CNN的语义分割采用`CPMC`生成Region。

流程

- RCNN算法分为4个步骤：
  - 一张图像生成1K~2K个候选区域。
  - 对每个候选区域，使用深度网络提取特征。
  - 特征送入每一类的SVM分类器，判别是否属于该类。
  - 使用回归器精细修正候选框位置。



利用预训练与微调解决标注数据缺乏问题

- 采用在`ImageNet`上使用`ILSVRC 2012`数据集已经训练好的模型，然后在`PASCAL VOC`数据集上进行`fine-tune`。
- 基层网络：`AlexNet`
- 训练策略采用`SGD`训练，初始学习率为`0.001`，mini-batch大小为`128`。

候选区域生成

- 使用了Selective Search方法从一张图像生成约2000~3000个候选区域。基本思路如下：
  - 使用一种分割手段，将图像分割成小区域。
  - 查看现有小区域，合并(交并比 >= 0.5)可能性最高的两个区域。重复直到整张图像合并成一个区域位置。
  - 输出所有曾经存在过的区域，所谓候选区域。

合并规则

- 优先合并一下四种区域：
  - 颜色（颜色直方图）相近的。

- 纹理（梯度直方图）相近的。
- 合并后总面积小的。
- 合并后，总面积在其BBox中所占比例大的。

多样化与后处理

- 为了可能不遗漏候选区域，上述操作在多个颜色空间中同时进行。在一个颜色空间中，使用上述四条规则的不同组合进行合并。所有颜色空间与所有规则的全部结果，在去除重复后，都作为候选区域输出。

特征提取

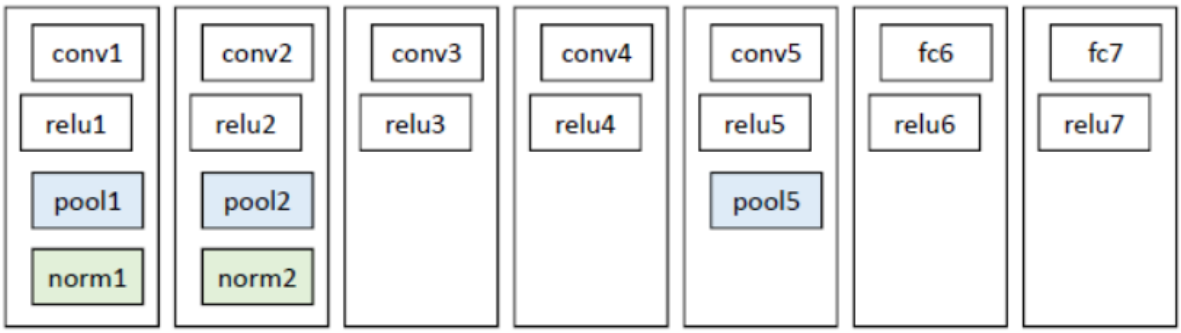
预处理

- 使用深度网络提取特征之前，首先把候选区域归一化成同一尺寸`227 x 227`。

预训练

网络结构

- 基本借鉴Hintona 2012年在Image Net上的分类网络。



此网络提取的特征为`4096`维，之后送入一个`4096->1000`的全连接层进行分类。

训练数据

- 使用ILVCR 2012的全部数据进行训练，输入一张图片，输出1000维的类别编号。

调优训练

网络结构

- 同样使用上述网络，最后一层换成4096->21的全连接网络，学习率0.001，每一个batch包含32个正样本（属于20类）和96个背景。

训练数据

- 使用PASCAL VOC 2007的训练集，输入一张图片，输出21维的类别编号，表示20类+背景。
- 考虑一个候选框和当前图像上所有标定框重叠面积最大的一个。如果重叠比例大于0.5，则认为此候选框为此类标定的类别；否则认为此候选框为背景。

类别判断

分类器

- 对每一类目标，使用一个线性SVM二类分类器进行判别。输入为深度网络输出的4096维特征，输出是否属于此类。
- 由于负样本很多，使用hard negative mining方法。
- 采用`SVM分类器`，而不采用`CNN（softmax）分类器`的原因：
    - SVM 训练和 CNN 训练对正负样本的定义方式是不同的，在 CNN 进行训练时，需要对训练数据进行阈值较低的标注，即 IoU(重叠度)大于 0.5 就可以标注为正样本，因为 CNN 在训练过程中容易过拟合，需要大量的训练数据，如果阈值标注过高，就会导致 CNN 训练样本数很少。而 SVM 训练只需要少量的样本，所以需要 IoU=1才可以标注为正样本。

正样本

- 本类的真值标定框。

负样本

- 考察每一个候选框，如果和本类所有标定框的重叠都小于0.3，认为其为负样本。

位置精修

回归器

- 对每一类目标，使用一个线性回归器进行精修。正则项λ = 10000。
- 输入为深度网络pool5层的4096维特征，输出为xy方向的缩放和平移。

训练样本

- 判定为本类的候选框中，和真值重叠面积大于0.6的候选框。
- R-CNN 将候选区域与 GroundTrue 中的 box 标签相比较，如果 IoU > 0.5（IOU的值通过级联的方式来优化），说明两个对象重叠的位置比较多，于是就可以认为这个候选区域是 Positive,否则就是 Negetive.