

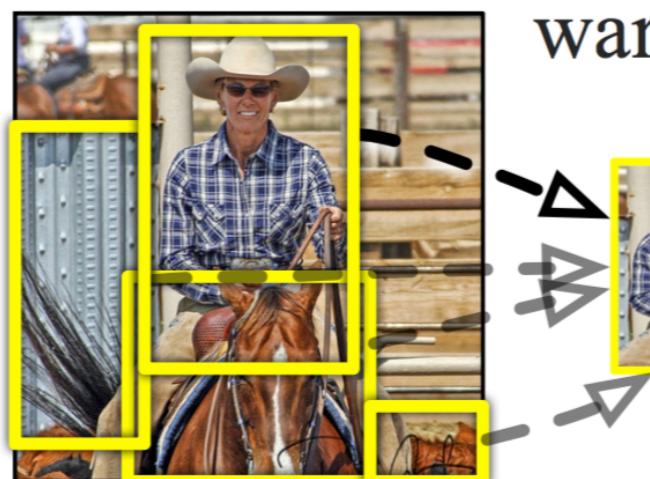
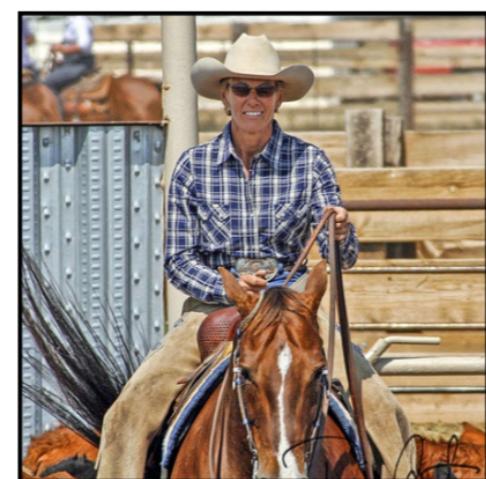
R-CNN  
SPP-net  
Fast R-CNN  
Faster R-CNN

# 图像分类 VS 物体识别



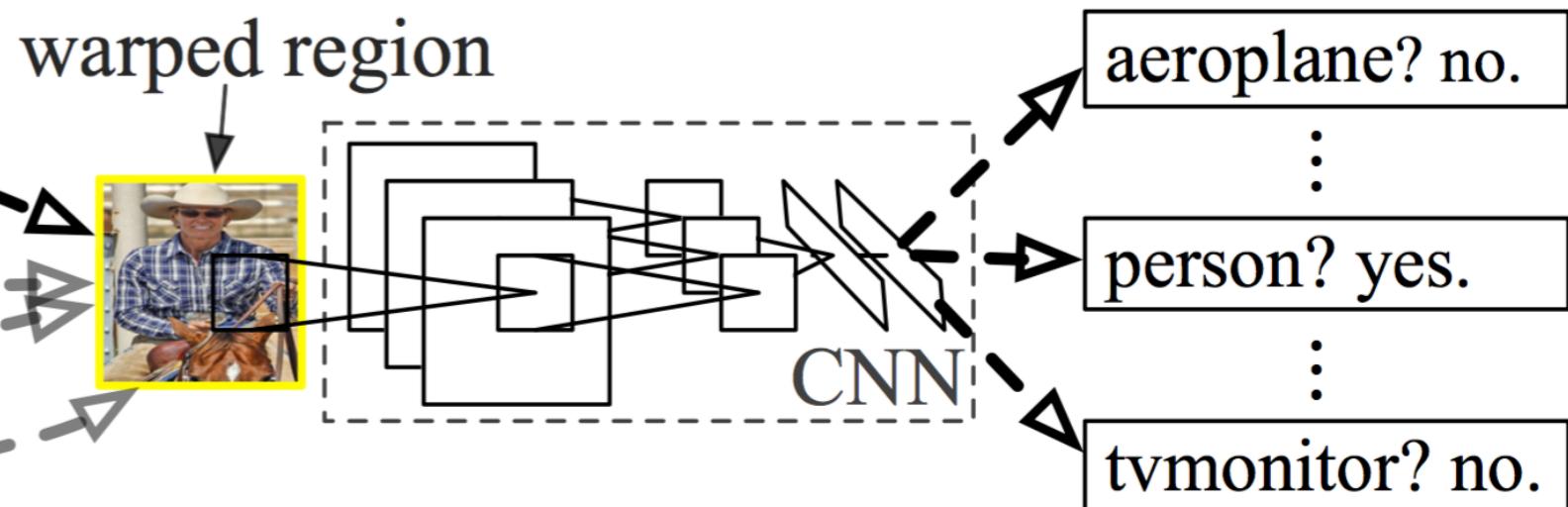
# R-CNN

**R-CNN: *Regions with CNN features***



1. Input  
image

2. Extract region  
proposals (~2k)

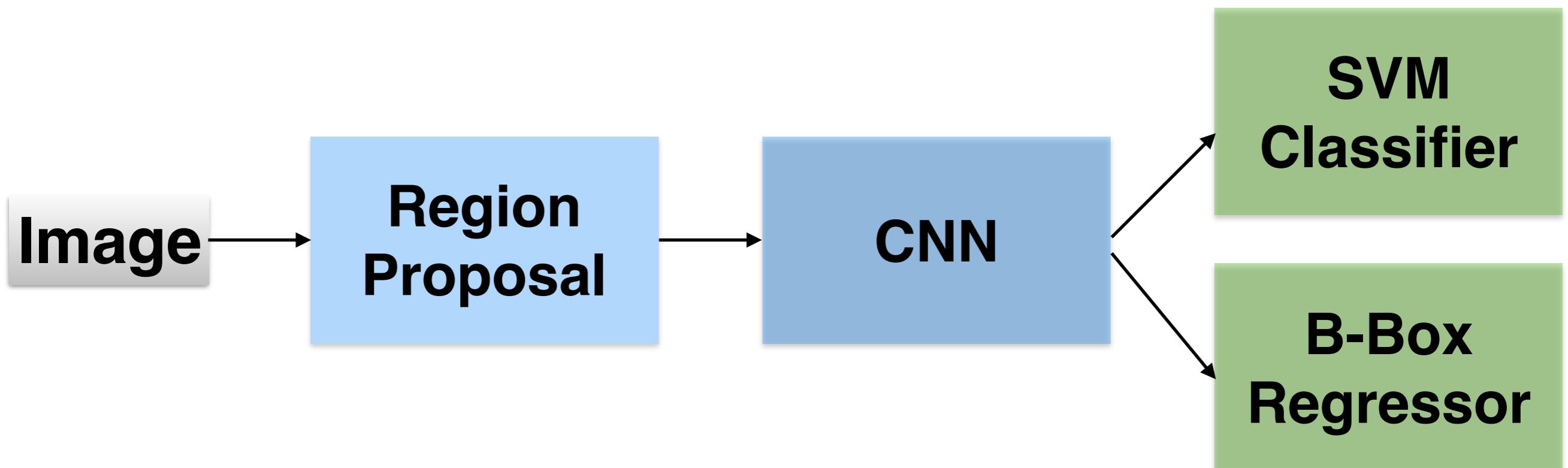


3. Compute  
CNN features

4. Classify  
regions

Selective Search 方法  
输入任意图像，输出  
2000个候选矩形框

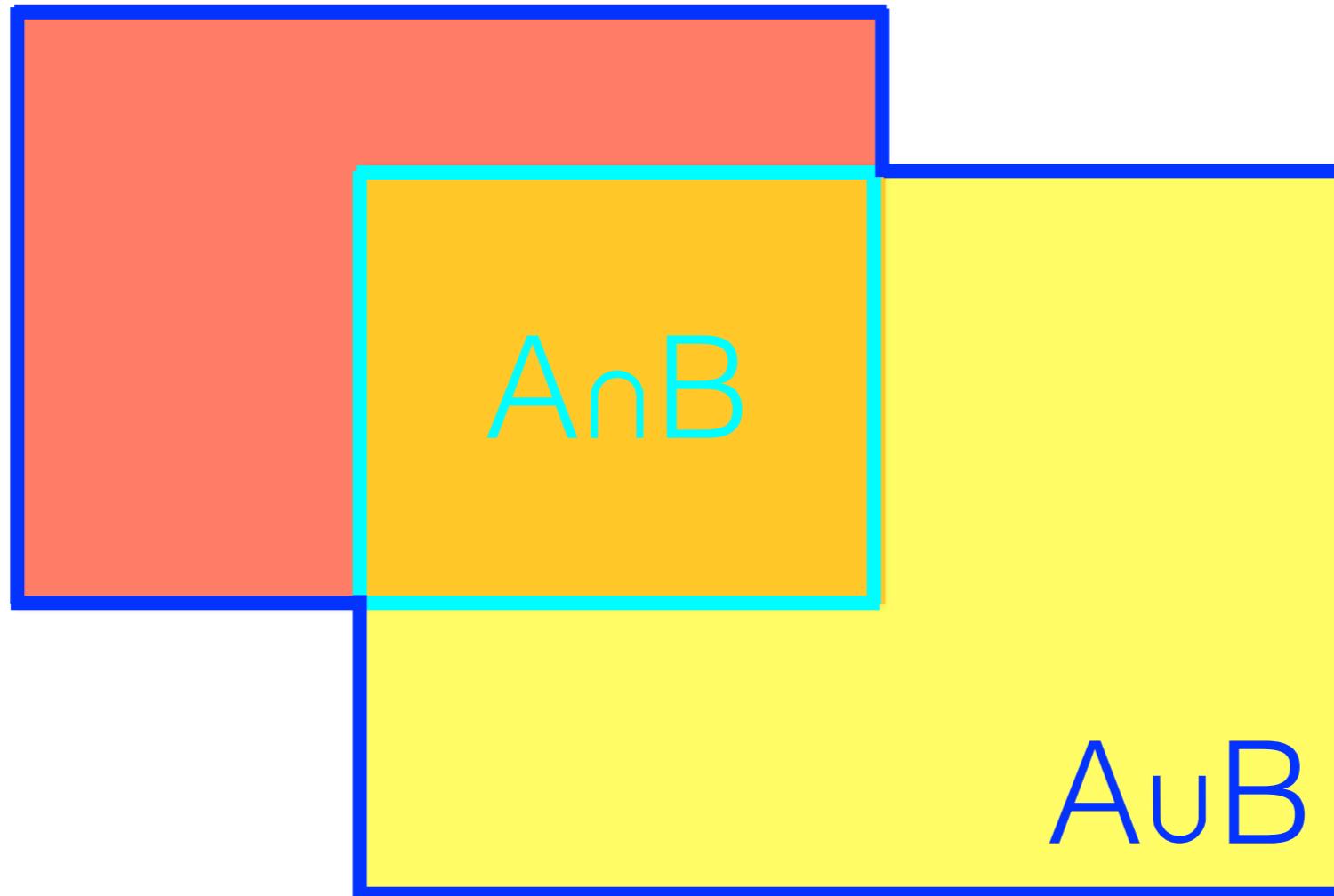
# R-CNN



# SVM分类器

- 一个SVM分类器用于一种特定物体类别的判定，输入为CNN计算得来的特征向量，输出为框内物体输入此类别概率
- 对于单个SVM分类器，正样本定义：输入候选框与至少一个此类别物体正确的矩形框重合度超过一定阈值即为正样本
- 负样本定义参见[1]的附录B

# IOU值 - 两个矩形框的重合度

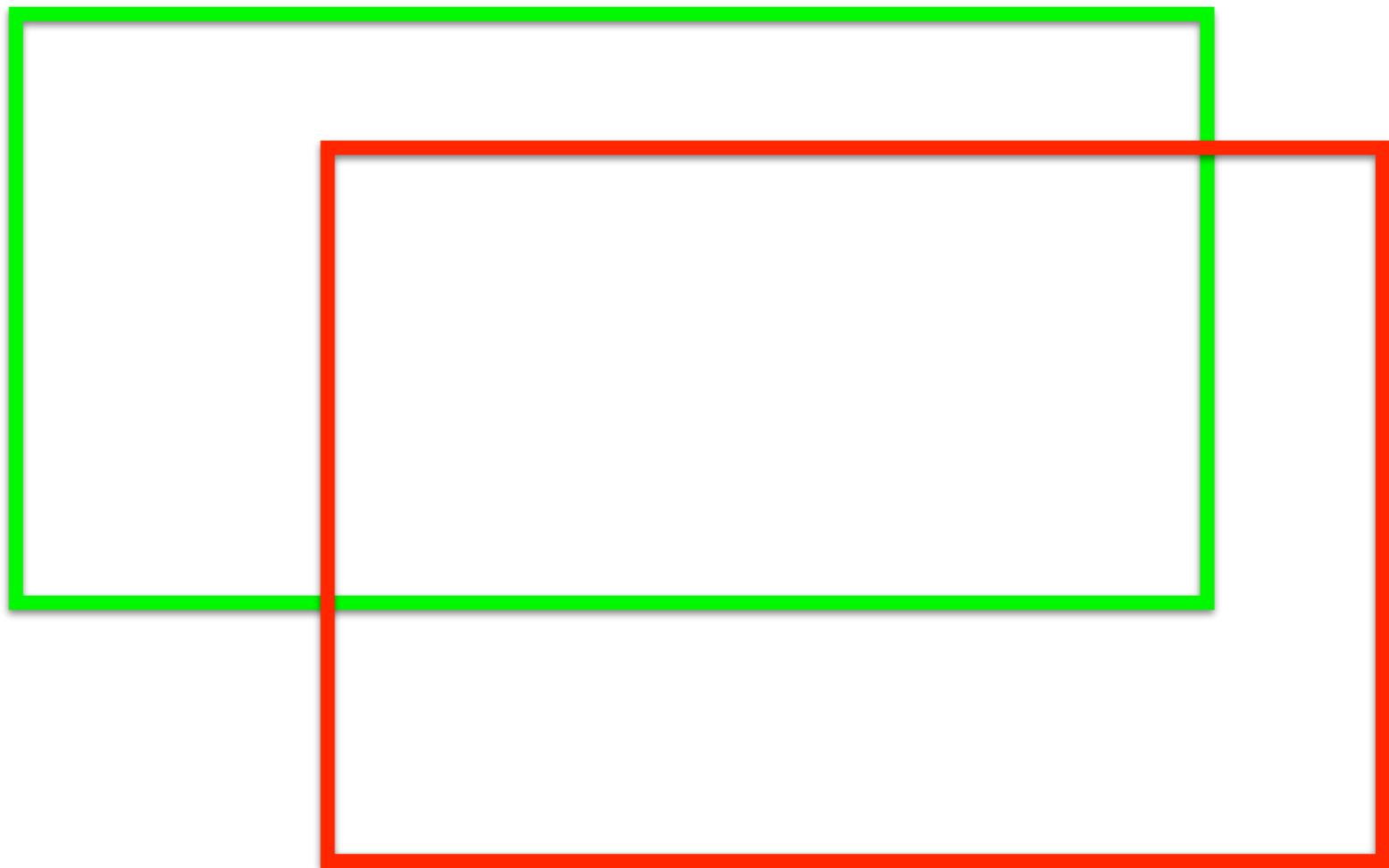


Intersection  
Over  
Union

$$IOU = \frac{A \cap B}{A \cup B}$$

# B-Box Regressor

- Bounding-Box Regressor
- 用于修正当前输入的候选框
- 输出4个值：中心横向平移距离，中心纵向平移距离，横向宽度缩放比例，纵向高度缩放比例



红色 - 候选矩形框  
绿色 - 正确矩形框



红色 - 候选矩形框  
绿色 - 正确矩形框



红色 - 候选矩形框

绿色 - 正确矩形框



红色 - 候选矩形框

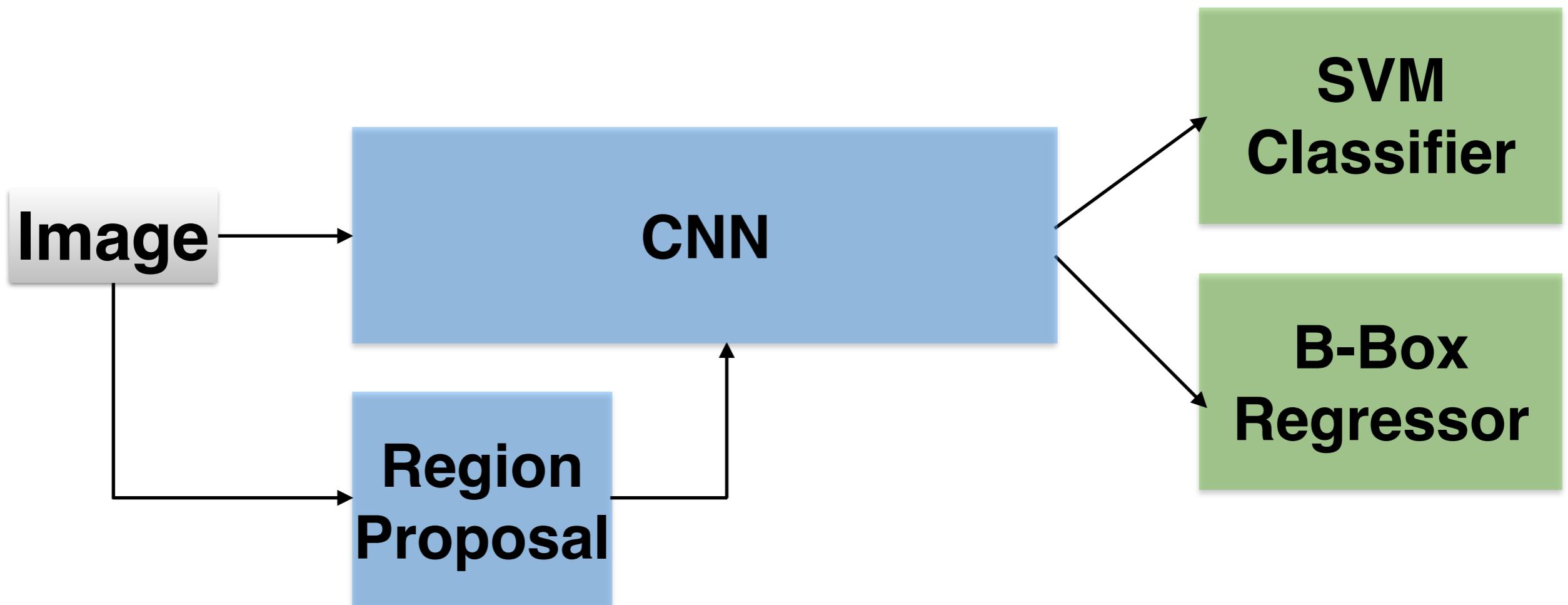
绿色 - 正确矩形框



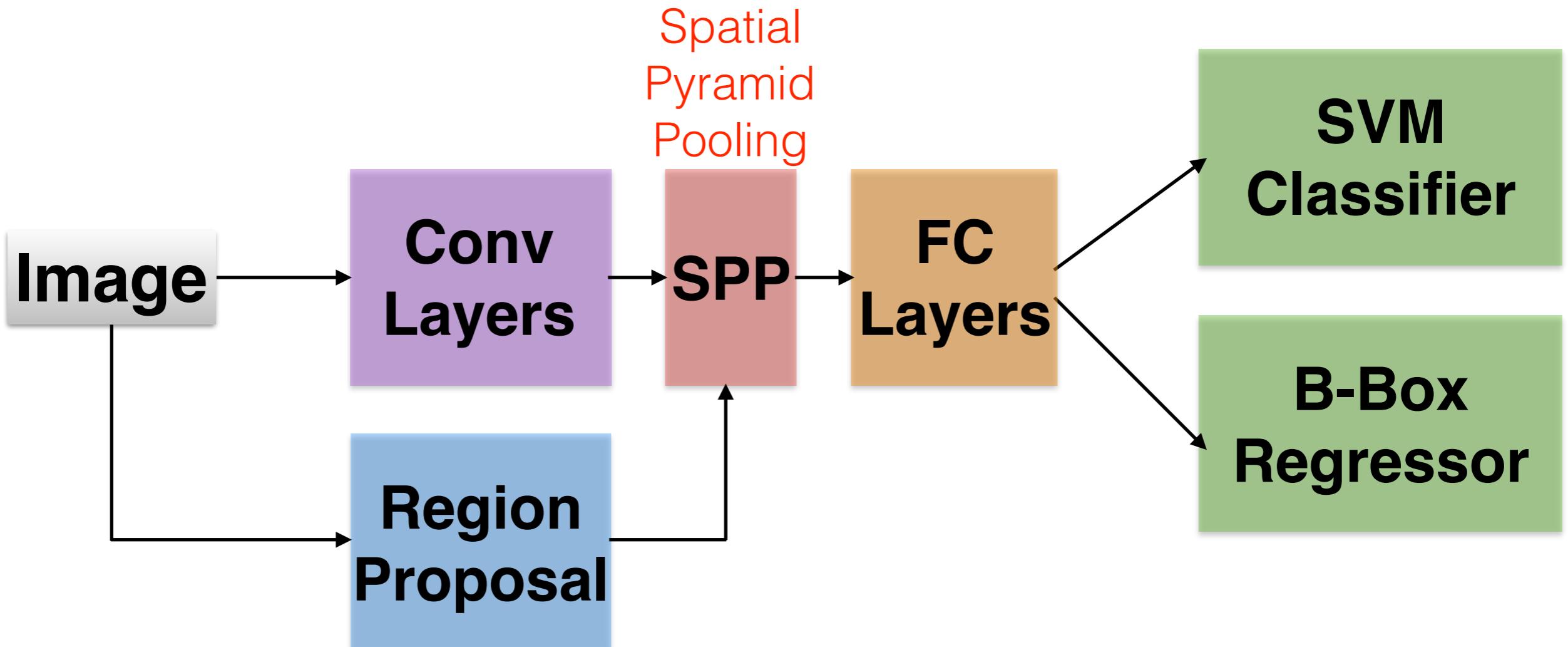
红色 - 候选矩形框

绿色 - 正确矩形框

# SPP-net



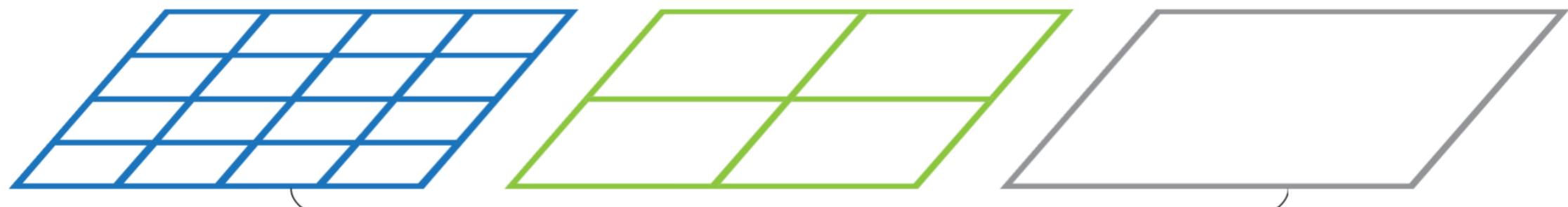
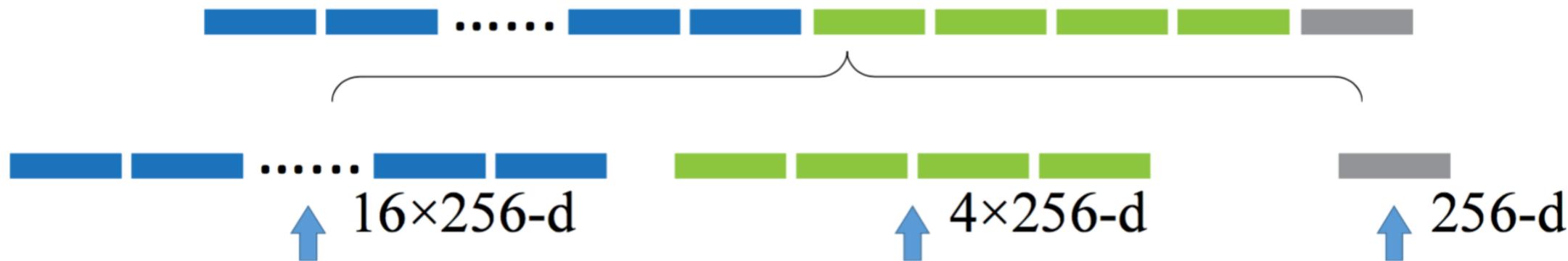
# SPP-net



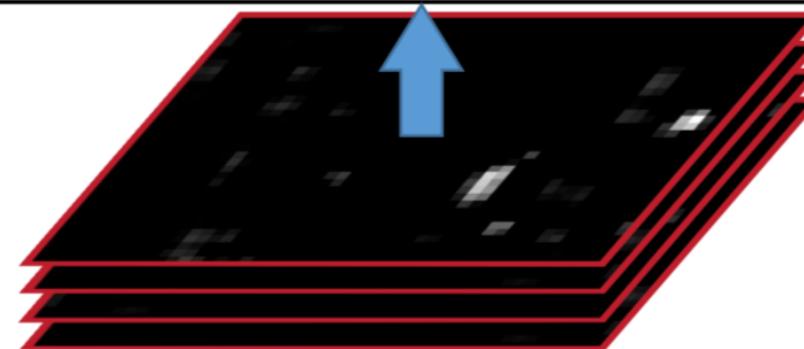
fully-connected layers ( $fc_6$ ,  $fc_7$ )



fixed-length representation



spatial pyramid pooling layer

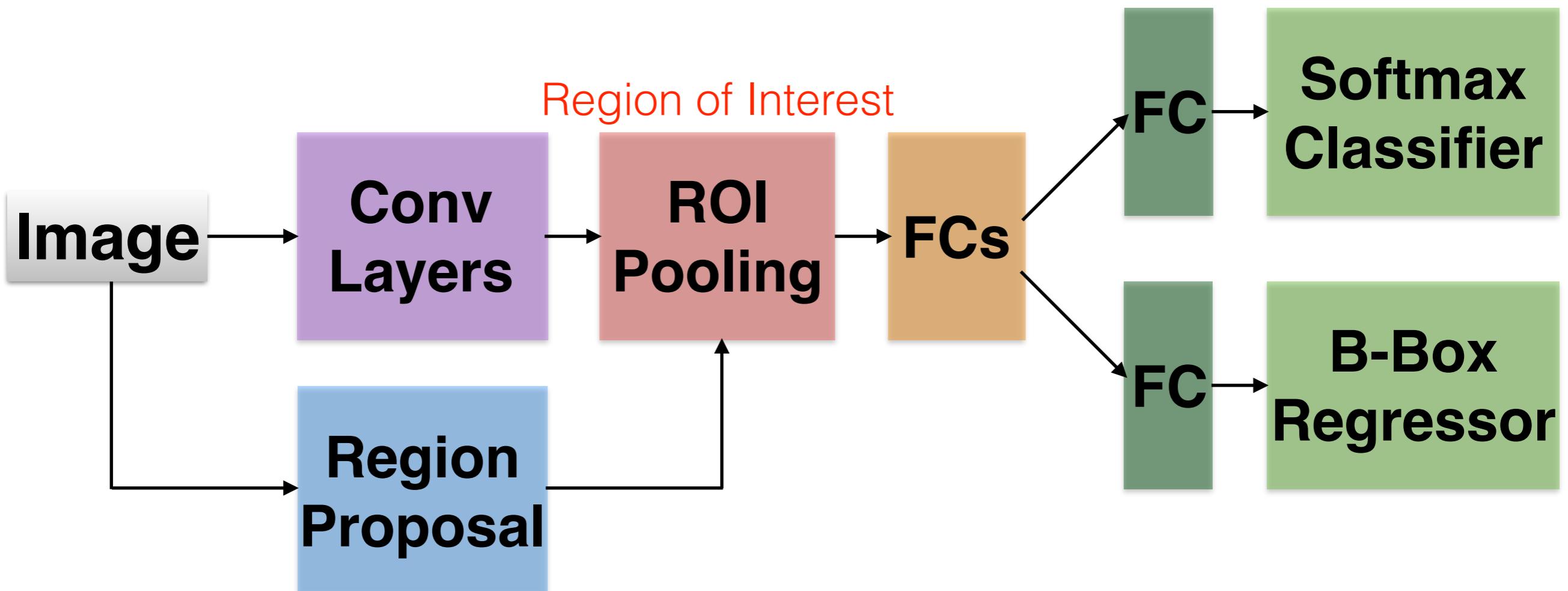


feature maps of  $conv_5$   
(arbitrary size)

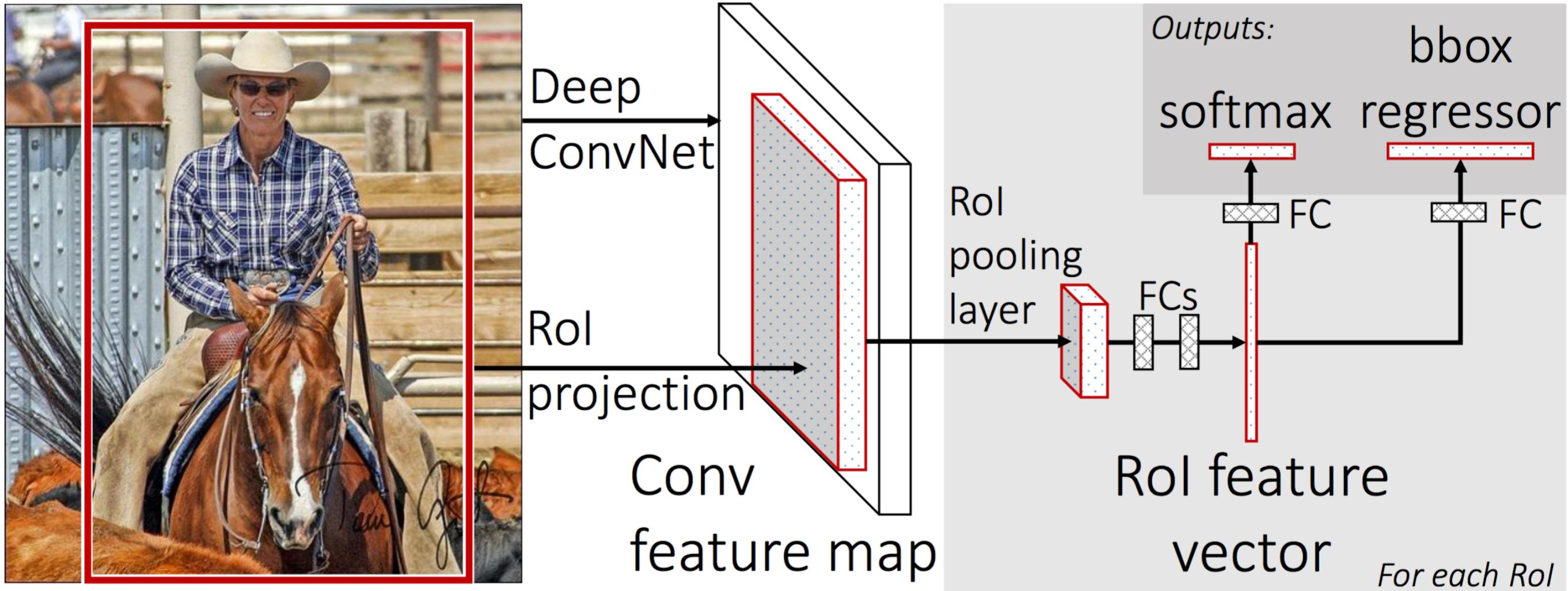


convolutional layers  
input image

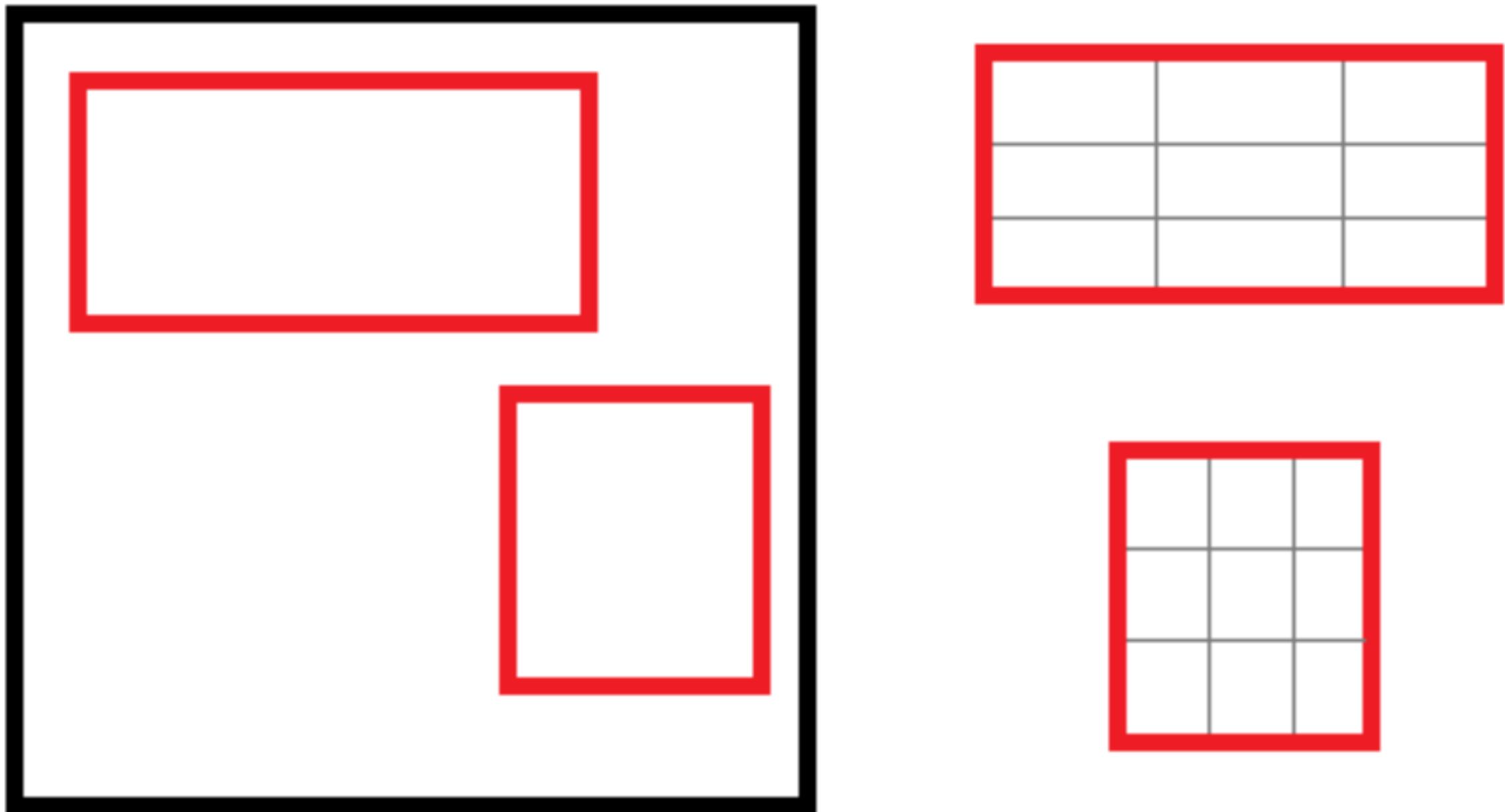
# Fast R-CNN



# Fast R-CNN

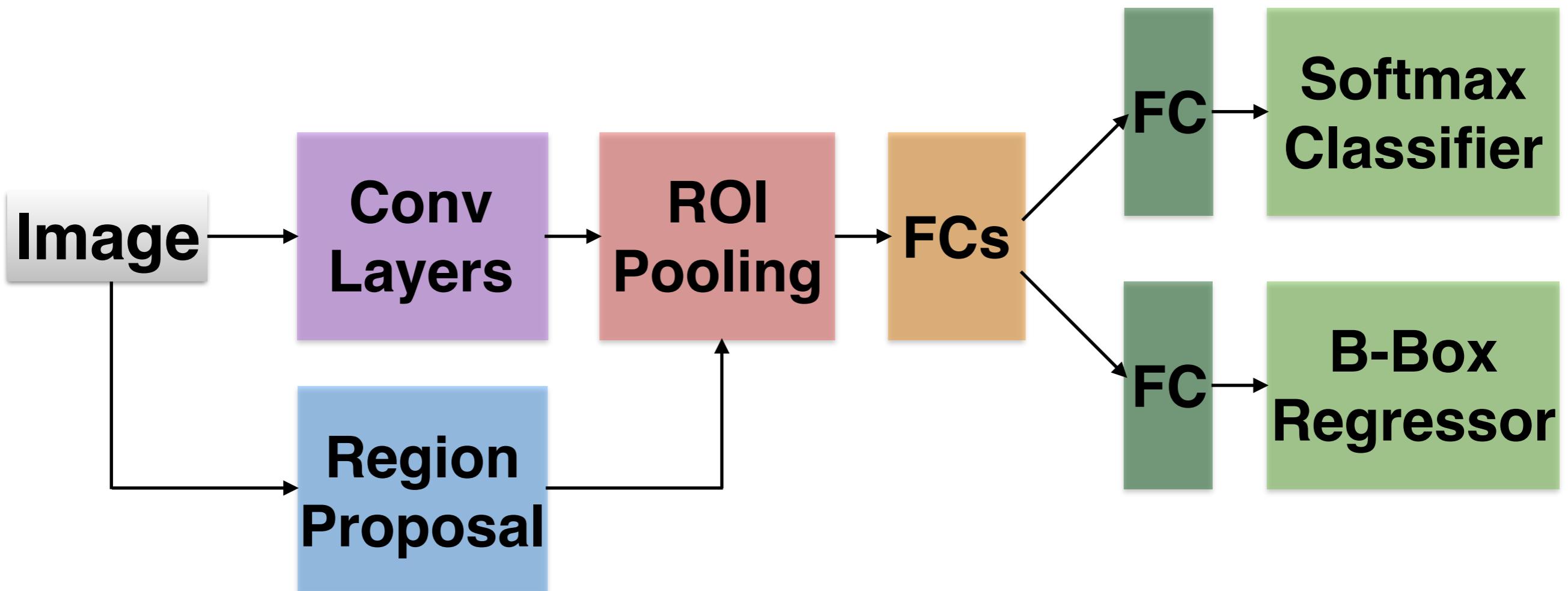


# ROI Pooling

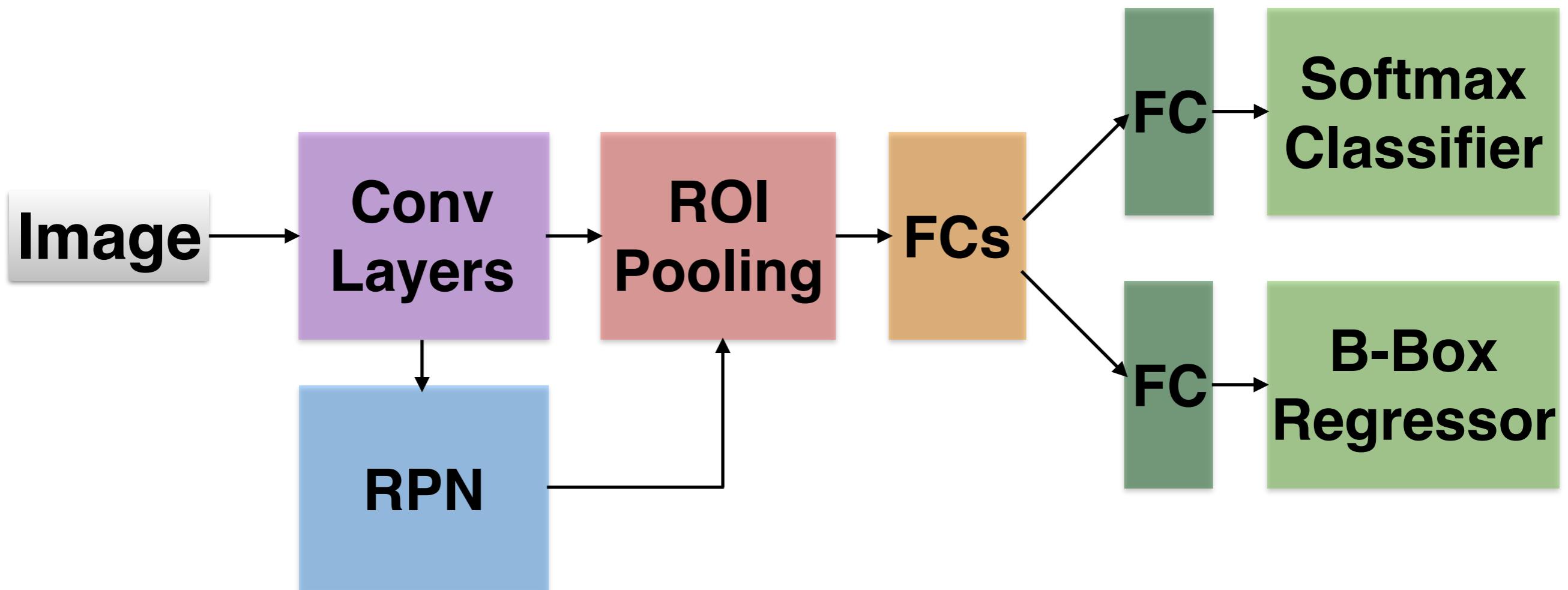


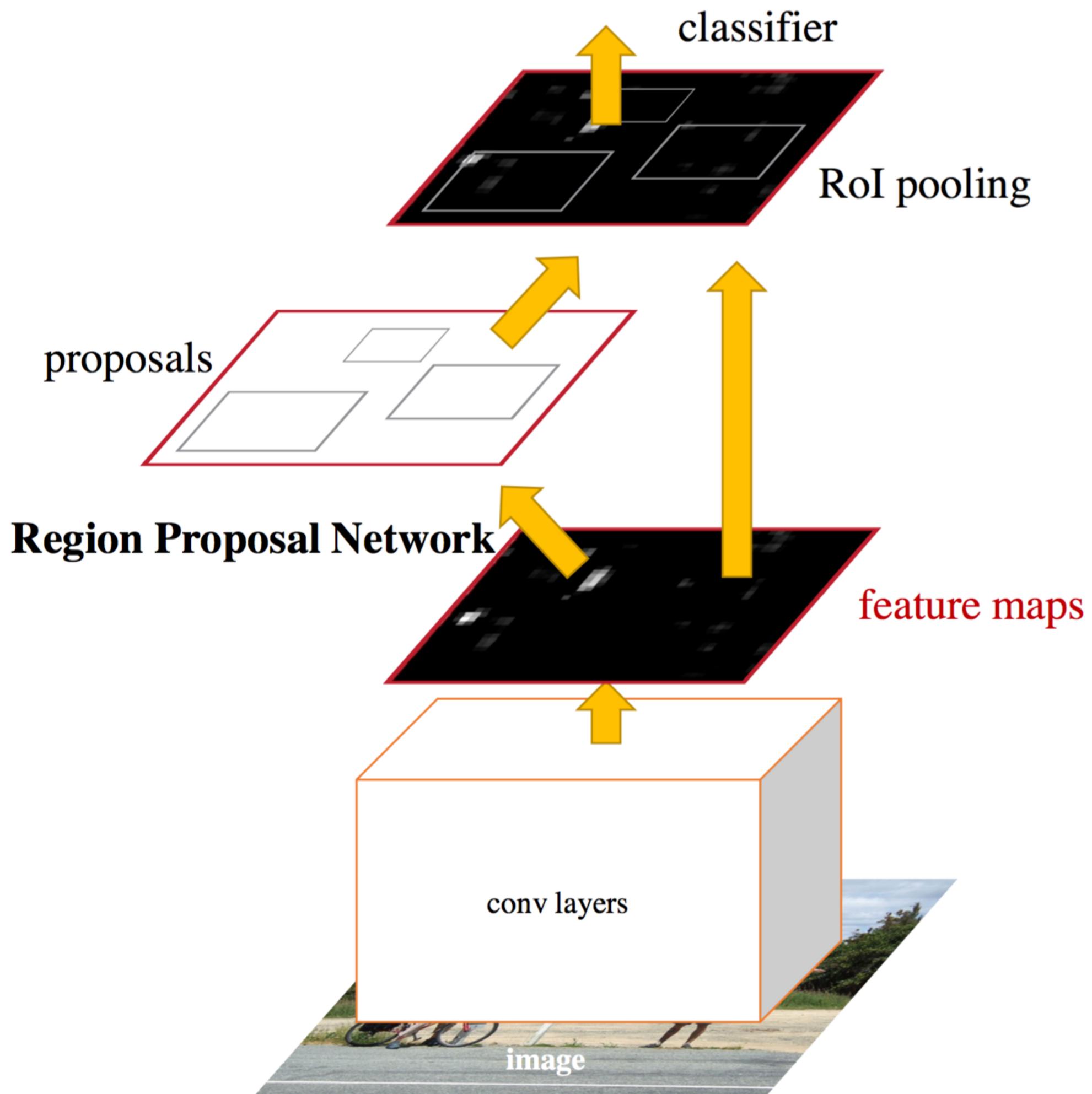
将每个候选区域均匀分成 $M \times N$ 块，对每块进行max pooling。将特征图上大小不一的候选区域转变为大小统一的数据，送入下一层

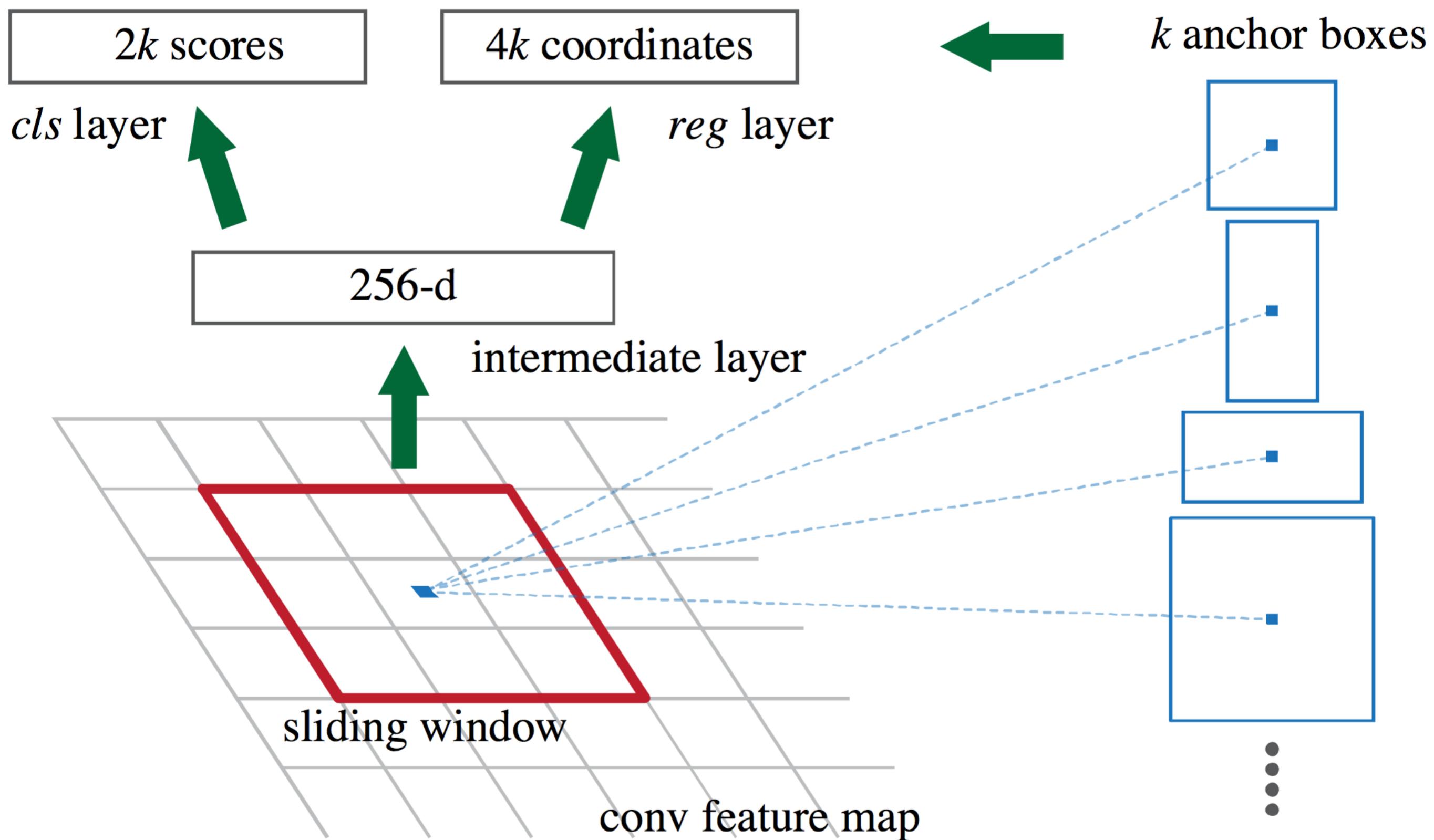
# Fast R-CNN



# Faster R-CNN







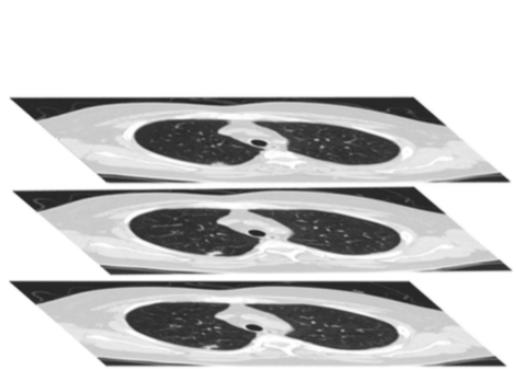
绿色 - 正确矩形框



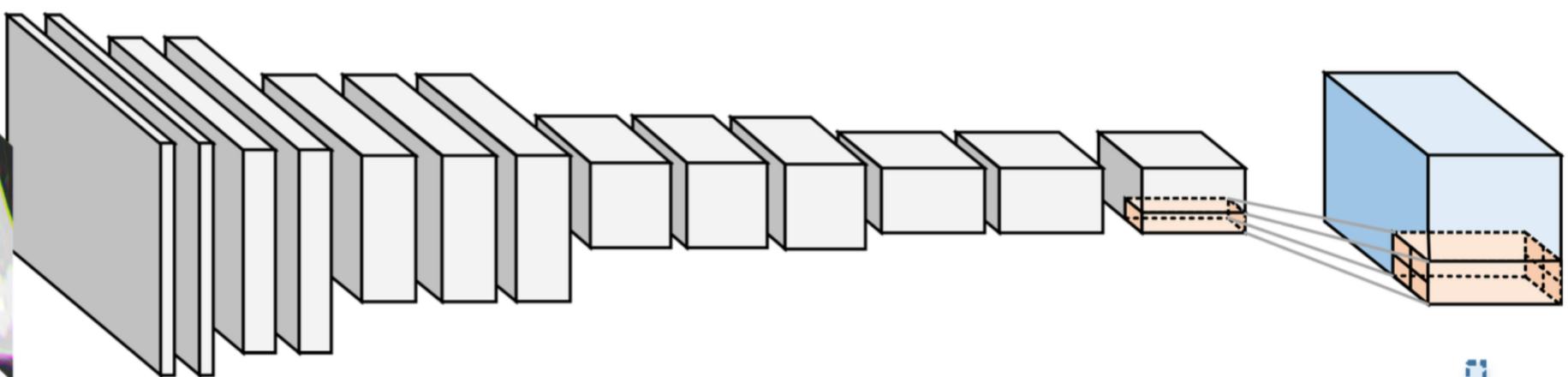
蓝色 - Anchor

## Stage 1. Candidate Detection

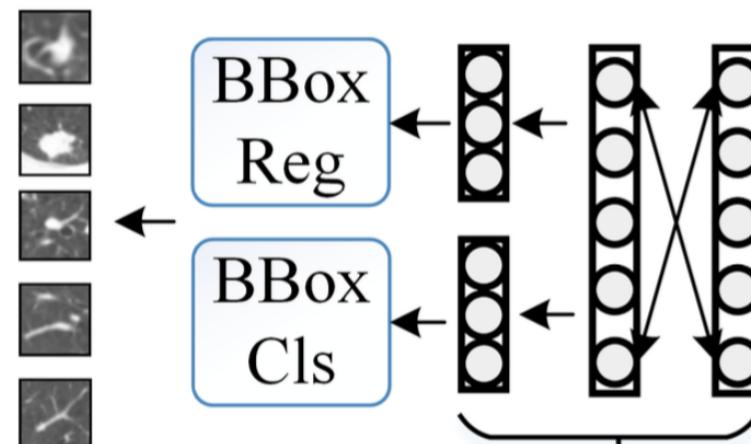
Input Image



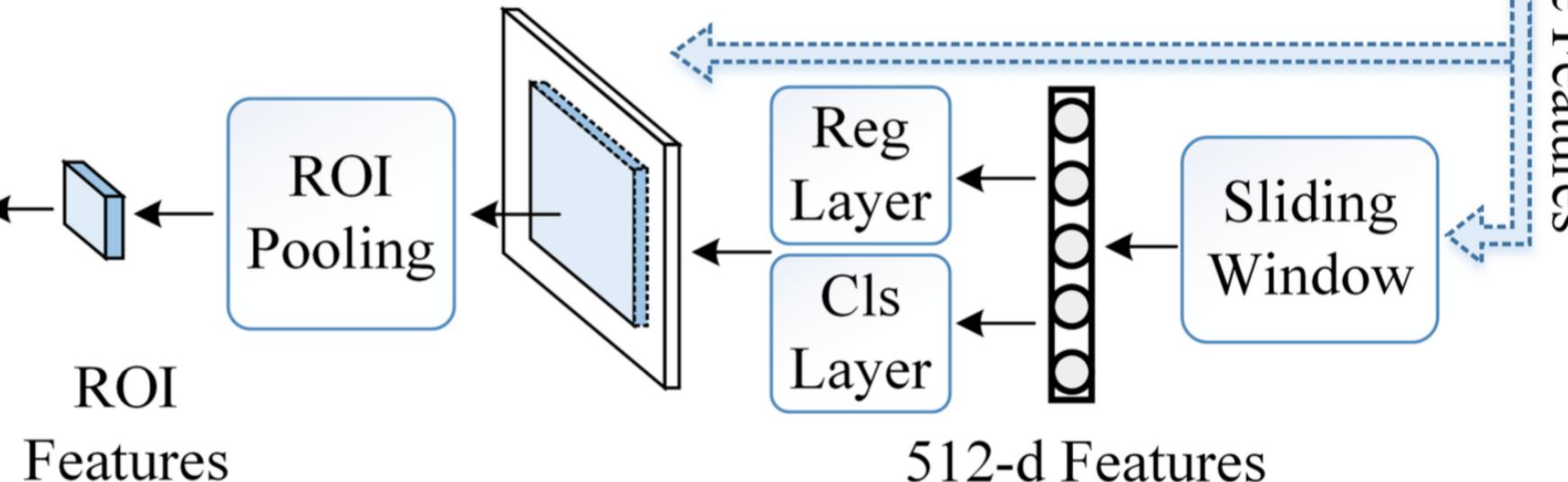
The First 5-group Convs of VGG-16



ROI Classifier



Region Proposal Network



Candidates

FCs

ROI Features

512-d Features

Share Features