

目标检测框架Faster-RCNN

目标检测框架Faster-RCNN

- ▶ 框架中用到的一些技术介绍
- ▶ Faster-RCNN整体框架
- ▶ 训练时采用的一些策略与参数设置

RCNN系列检测算法介绍

- ▶ RCNN、fast-rcnn、faster-rcnn
- ▶ 此系列算法采用策略：
 - 1、提取建议区域，R代表region proposal。
 - 2、利用CNN对建议区域进行分类
- ▶ 提取建议区域方法的发展：
 - 滑动窗口、select search/edge box、rpn
- ▶ 其它深度学习检测策略，利用CNN强大的表述能力直接对目标位置进行回归，例如YOLO。

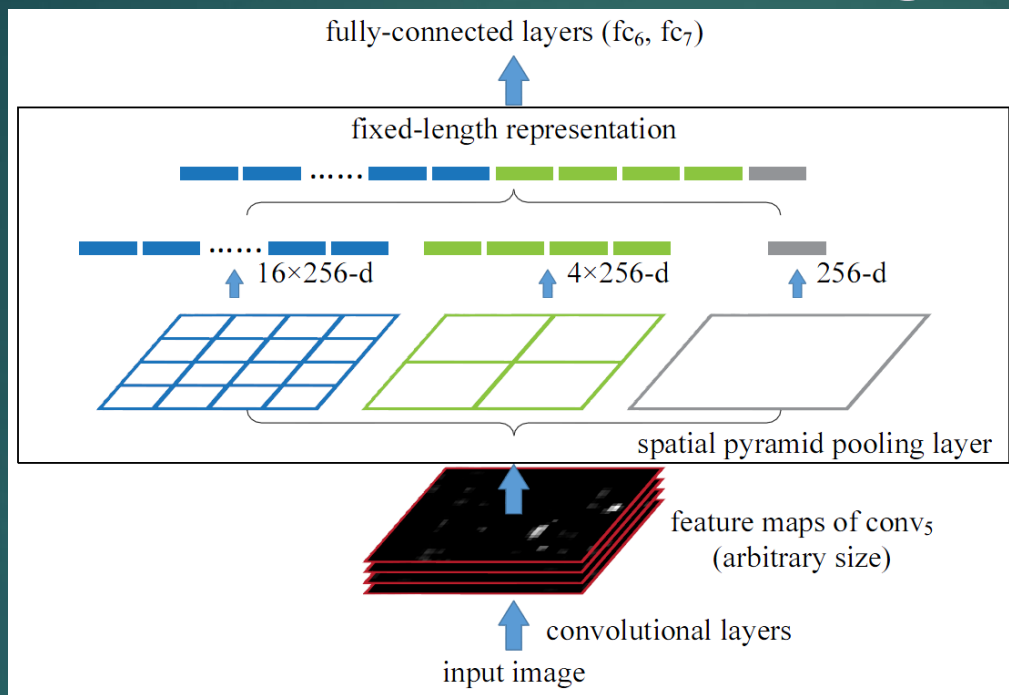
SPP Net

1. 结合空间金字塔方法实现CNNs的多尺度输入



一般CNNs后接full-connect layer或者classifier，它们都需要固定的输入尺寸。因此不得不对输入数据进行crop或warp，这些预处理会造成数据的丢失或几何学上的失真。SPP Net 的第一个贡献是将空间金字塔的思想加入到CNNs中，实现了数据的多尺度输入。

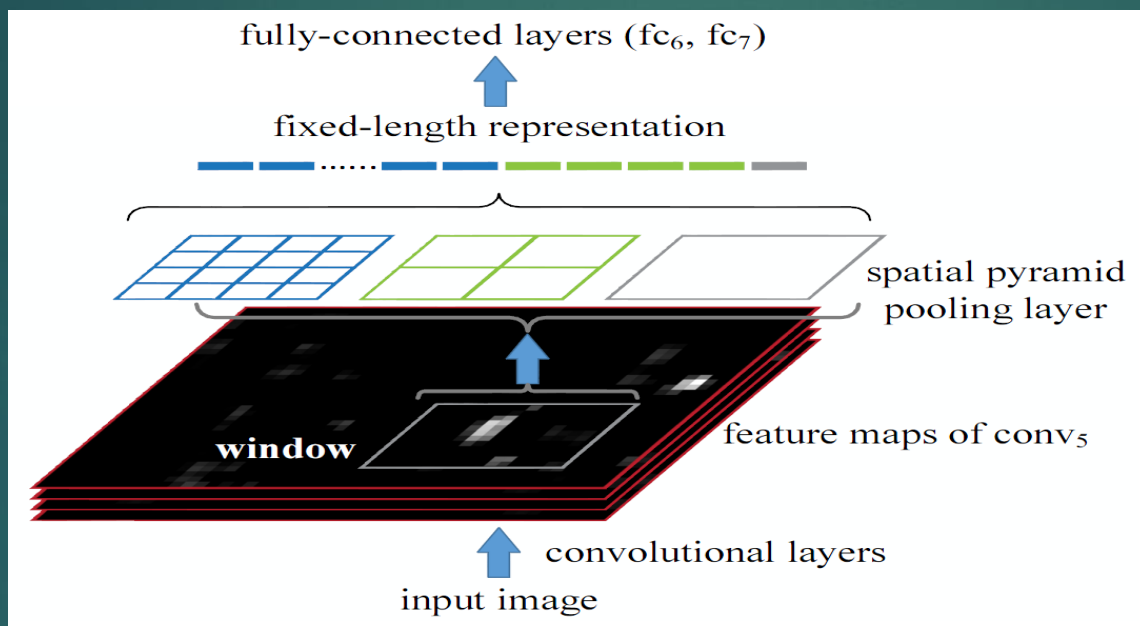
Spatial Pyramid Pooling Layer (SPP)



如图，在卷基层和全连接层之间加入SPP layer。此时网络的输入可以是任意尺寸，在SPP layer中每一个pooling的filter会根据输入调整大小，而SPP的输出尺寸始终是固定的。

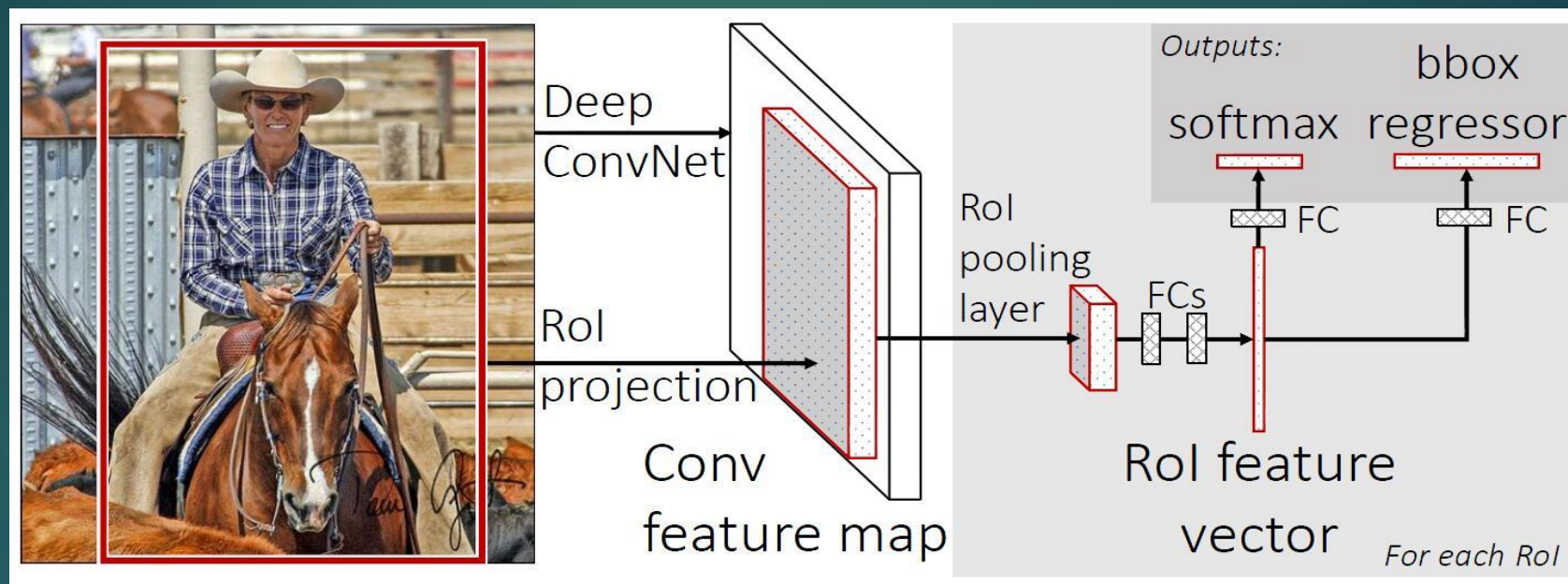
2. 只对原图提取一次卷积特征

在R-CNN中，每个proposed region先rescale成统一大小，然后分别作为CNNs的输入，这样是很低效的。



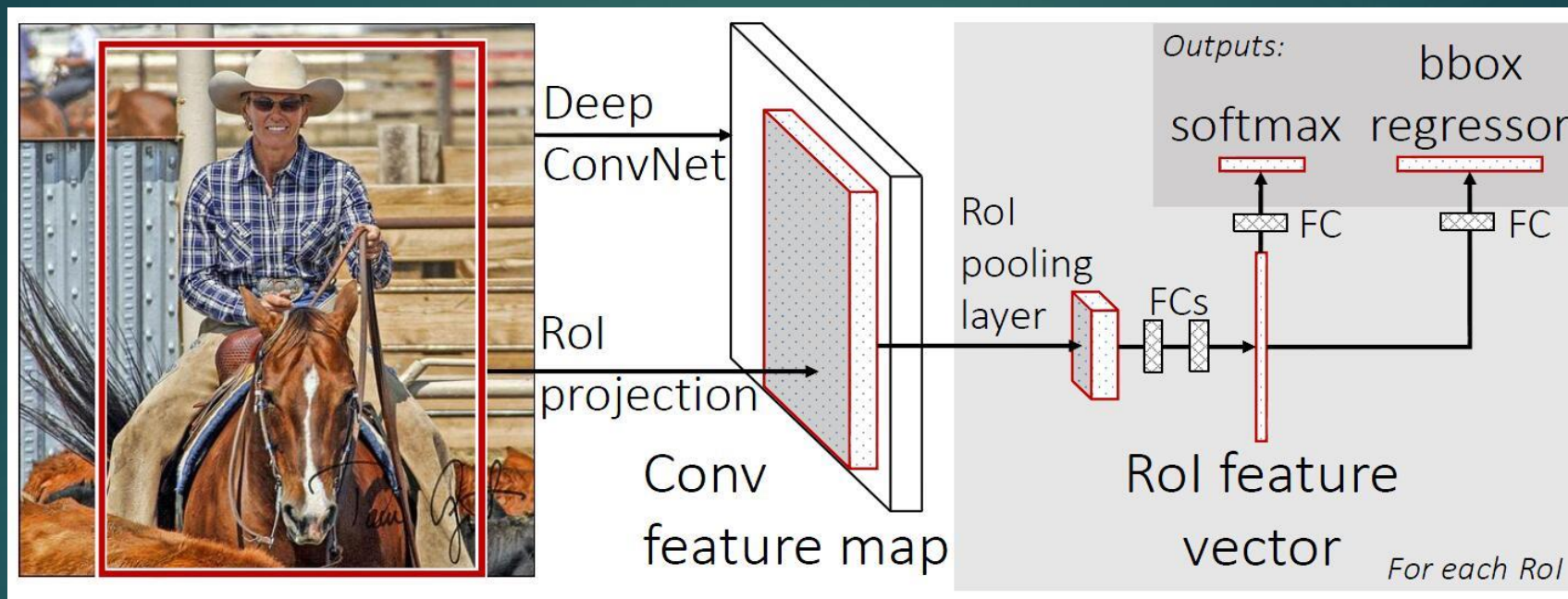
在SPP Net中，只对原图进行一次卷积得到整张图的feature map，然后找到每个proposed region在feature map上的映射patch，将此patch作为每个proposed region的卷积特征输入到SPP layer和之后的层。节省了大量的计算时间，比R-CNN有一百倍左右的加速。

Fast R-CNN整体结构



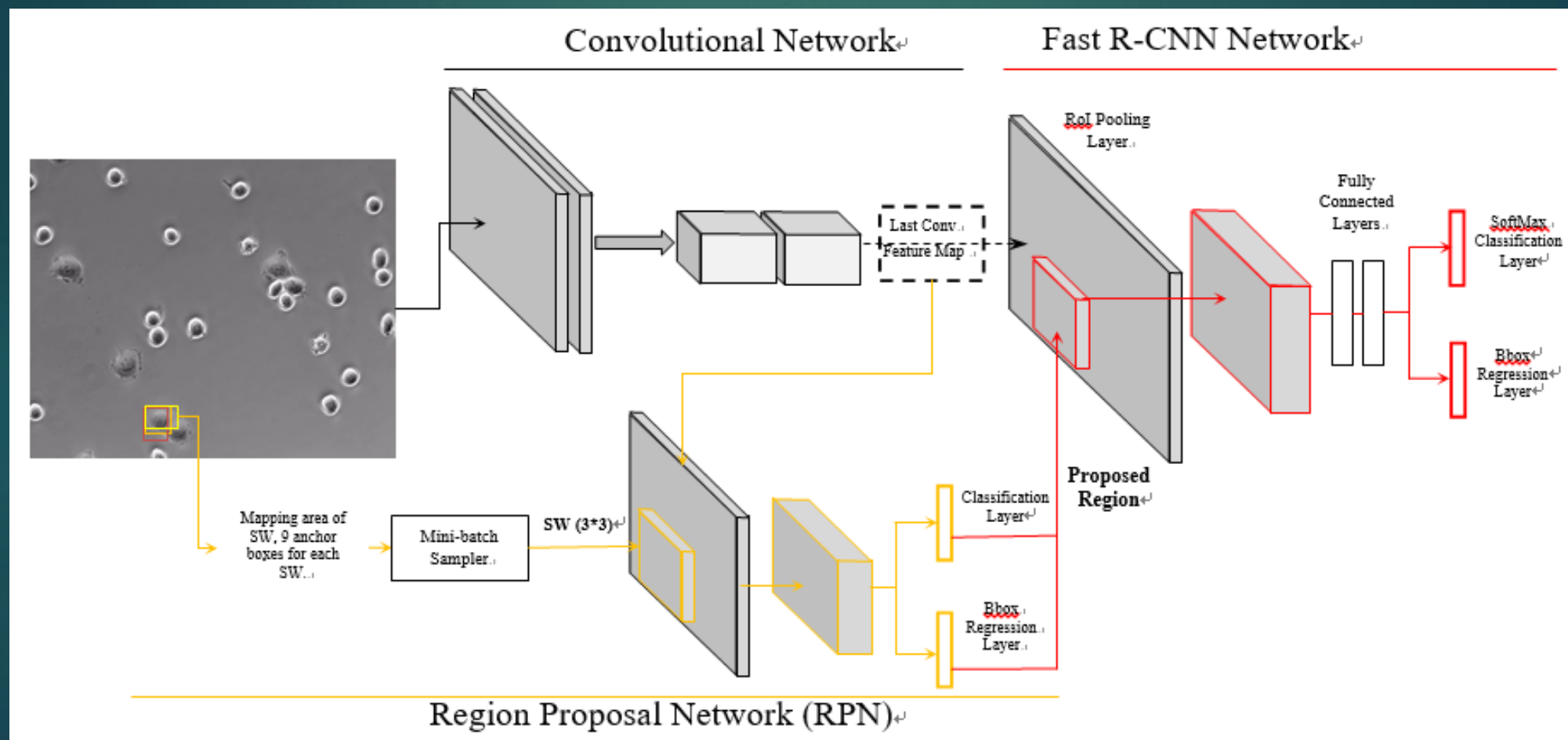
如图，Fast R-CNN的网络有**两个输出层**，一个softmax，一个bbox regressor（相对的R-CNN,SPP Net中分类和回归是两个部分，这里集成在了同一个网络中）。而且加入了一个RoI pooling layer（类似于一个尺度的SPP layer）。注意：Fast R-CNN提取建议区域的方法依然是select search。

Rol pooling layer



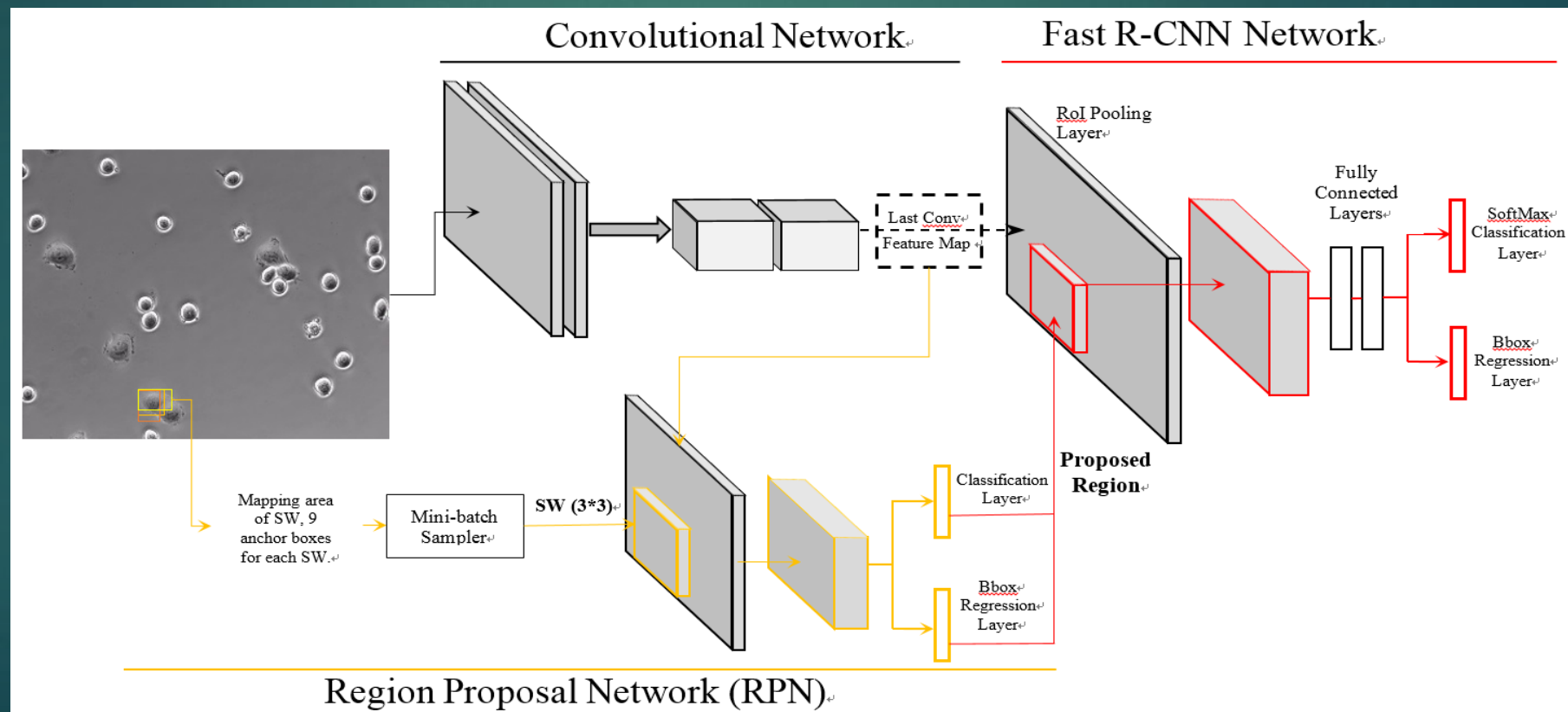
- 这是SPP pooling的一个简化版本，可以看做是只有一个尺度 filter的‘金字塔’。输入是N个整幅图的feature map和一组R个Rol（proposed region）。每个特征映射都是 $H \times W \times C$ ，每个Rol是一个元组 (n, r, c, h, w) ， n 是特征映射的索引， r, c, h, w 分别是Rol的左上角坐标和高与宽。输出是max-pooling过得特征映射 $H' \times W' \times C$ ，如上图红色框线。

Faster-RCNN整体框架

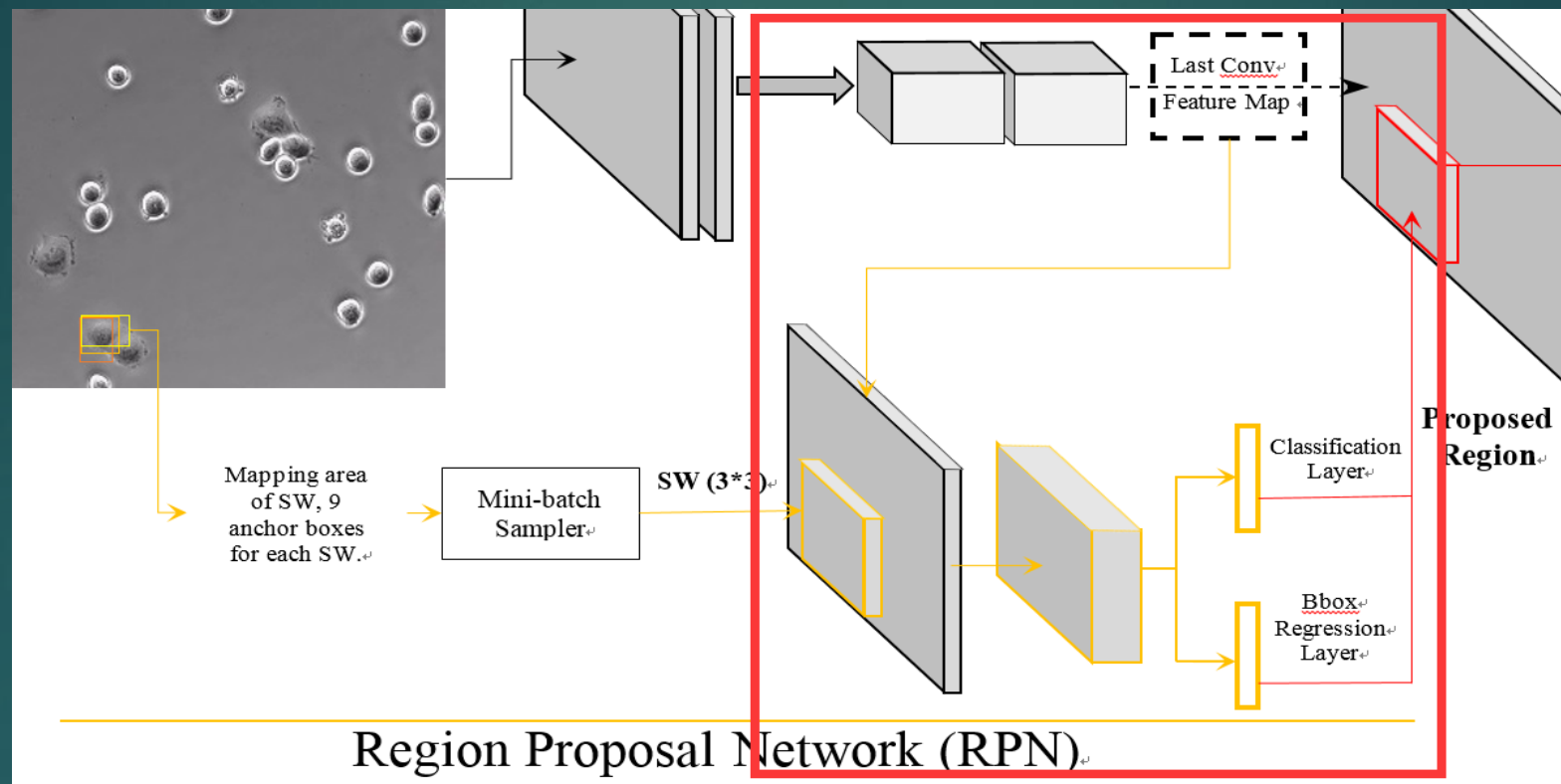


Faster R-CNN

- ▶ Faster R-CNN的主要贡献是设计了提取建议区域的网络Region Proposal Network (RPN)。代替了费时的select search，使检测速度大为提高。下图为Faster R-CNN的结构图，黄色部分为RPN，可以看出除了RPN，其它部分继承了FR-CNN的结构。

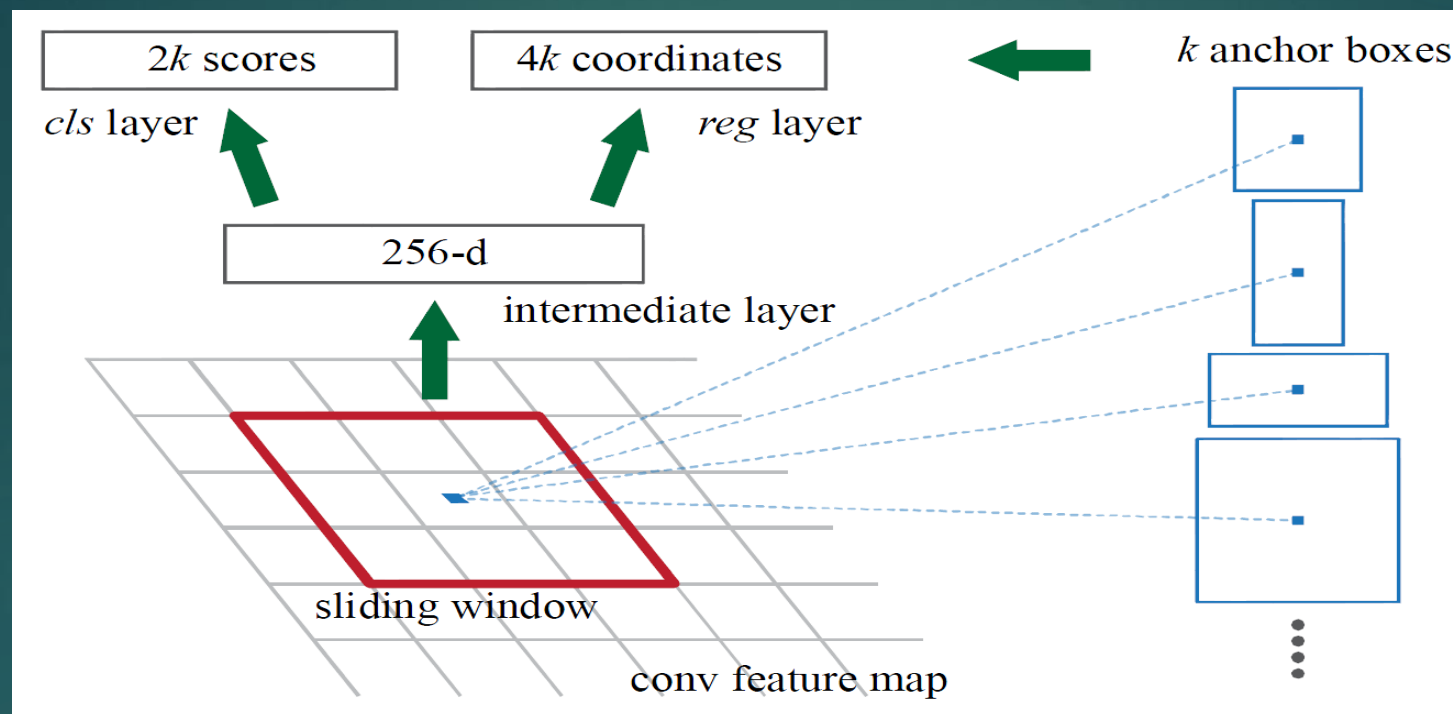


RPN整体结构



RPN的网络结构类似于FR-CNN，连接与最后卷基层输出的feature map，有一个RoI层，两个输出层，一个输出滑窗为建议区域的概率，另一个输出bbox回归的offset。其训练方式也类似于FR-CNN。注意：RPN与FR-CNN共用卷积层。

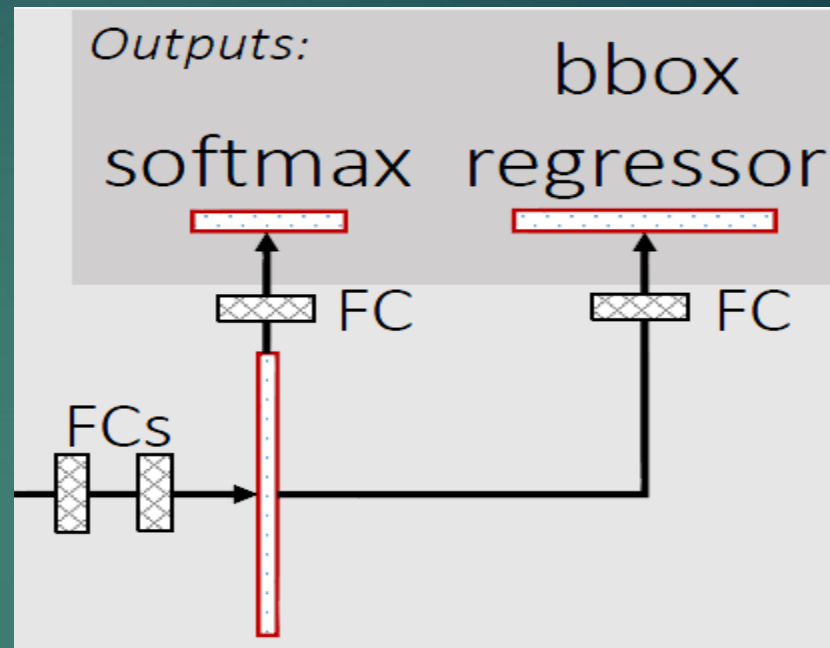
RPN



RPN通过一个滑动窗口（图中红色框）连接在最后一个卷积层输出的feature map上，然后通过全连接层调整到256-d的向量，作为输出层的输入。同时每个滑动窗对应 k 个anchor boxes，在论文中使用3个尺寸和3个比例的 $3*3=9$ 个anchor。每个anchor对应原图上一个感受野，通过这种方法提高scale-invariant。

Multi-task loss

FR-CNN的有两个网络输出层，将原来与网络分开的bbox regression的操作整合在了网络中。并设计了一个同时优化两个输出层的loss函数。



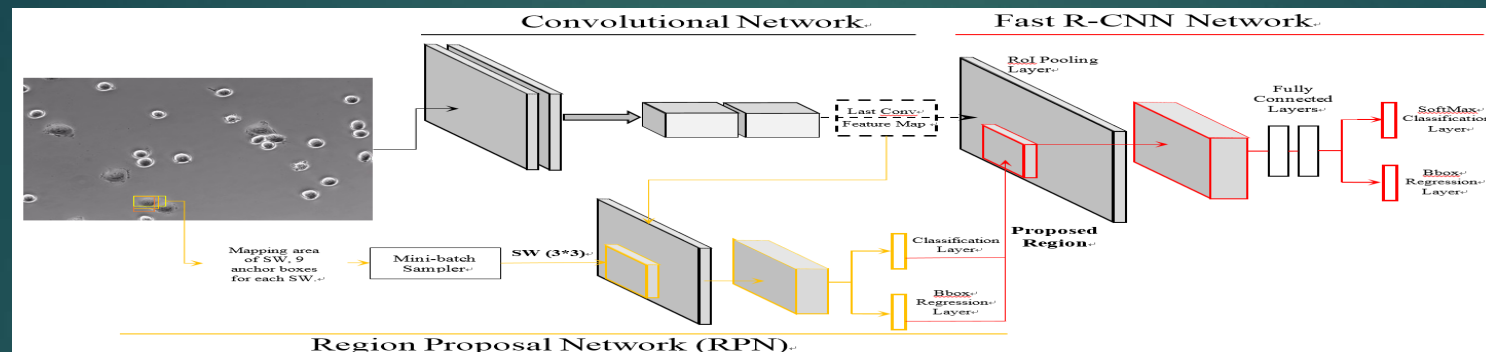
$$L(p, k^*, t, t^*) = L_{cls}(p, k^*) + \lambda[k^* \geq 1]L_{loc}(t, t^*)$$

其中，第一部分是softmax层的loss函数， p 是每个建议区域为目标概率， k^* 是一个表示类别的标签。第二部分是bbox regression层的loss函数， $t^* = (t_x^*, t_y^*, t_w^*, t_h^*)$ 是bbox的标签offset， $t = (t_x, t_y, t_w, t_h)$ 是bbox预测的offset。

Rol-centric sampling与Image-centric sampling

- ▶ Rol-centric sampling: 从所有图片的所有Rol中随机均匀取样，这样每个SGD的mini-batch中包含了不同图像中的样本（SPP Net采用）。SPP Net的反向传播没有到SPP pooling之前的层，因为反向传播需要计算每一个Rol感受野的卷基层，通常会覆盖整幅图像，又慢又耗内存。FR-CNN想要解决这个限制。
- ▶ Image-centric sampling: mini-batch采用分层采样，先对图像采样，再对Rol采样。将采样的Rol限定在个别图像内，这样同一图像的Rol共享计算和内存。通过这种策略，实现了端到端的反向传播，可以fine-tuning整个网络。

RPN与FR-CNN共享卷基层



为了使共用的卷积层在训练RPN和FR-CNN时都会收敛，论文里设计了一个四步训练的策略：

- (1)：对RPN进行end-to-end的训练，这里网络使用ImageNet pre-trained model进行初始化。
- (2)：使用第一步RPN生成的建议区域训练FR-CNN，这里也使用ImageNet pre-trained model进行初始化。
- (3)：使用上一步FR-CNN的参数初始化RPN，固定卷基层，只fine-tune RPN独有的层。（在此步已共享卷积层）
- (4)：固定卷基层，只fine-tune FR-CNN独有的层。

训练时采用的一些策略与参数设置

► 1、使用ImageNet分类模型初始化。

使模型具有一定的提取图像通用特征的能力，是目前在训练数据较少和检测任务中的常用做法。

► 2、solver配置文件中的参数

初始学习率0.001，学习率衰减系数0.1，SGD方法，权重衰减项0.0005。在RPN和Fast-RCNN中都是这样。

► 3、训练样本选择方法与其参数设置

训练时采用的一些策略与参数设置

► 3、训练样本选择方法与其参数设置

RPN中参数的设置

按照Image-centric sampling, 每个batch中的图片数。ims_per_batch 1或2

每个batch中的正负样本数。 batch_size 256

每个batch中正样本占得比率。 fg_fraction 0.5

与Ground Truth的IOU大于阈值0.75的ROI作为正样本。 fg_thresh=0.75

与Ground Truth的IOU在阈值0到0.35之间的ROI作为负样本。
bg_thresh_hi=0.35、bg_thresh_lo=0

feature map到原图映射的变化率为16, 这个数值主要是由池化降维算出, 与其步长选择有关。 feat_stride=16

训练时采用的一些策略与参数设置

► 3、训练样本选择方法与其参数设置

Fast-RCNN中参数的设置

ims_per_batch 1或2

batch_size 128

每个batch中正样本占得比率。 fg_fraction 0.25

与GT的IOU大于阈值0.6的ROI作为正样本。 fg_thresh=0.6

与GT的IOU在阈值0.1到0.5之间的ROI作为负样本。 bg_thresh_hi=0.5、
bg_thresh_lo=0.1