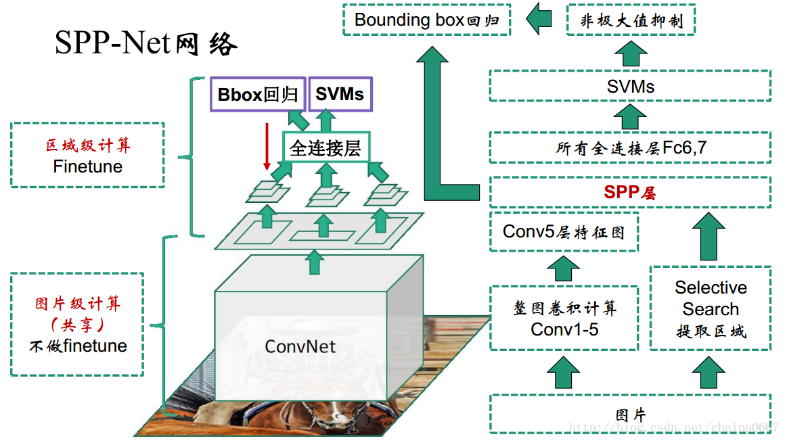
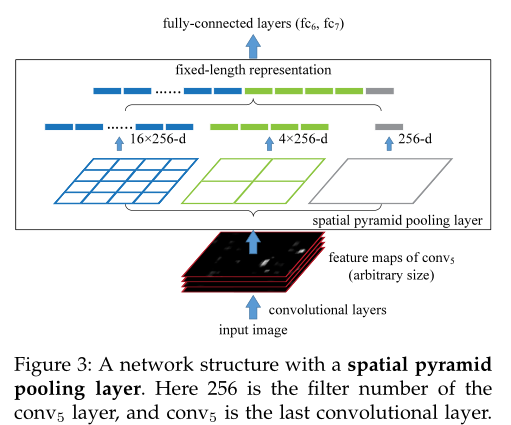
**[Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition]**

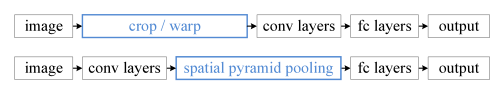
**IEEE 2015 何凯明**

**主要思想：**。SPP是一种池化方式，特点是对任意输入，可得到固定长度输出。SPP-Net是采用SPP对RCNN的改进，避免重复计算，大大提高速度。比较简单，直接见下文论文文细节部分。

**贡献/缺点：**主要贡献有两点：1）**空间金字塔池化**。传统CNN中，卷积层对输入尺寸没有要求，但**全连接层的固定节点数要求卷积层输出固定尺寸的特征图**，这需要在卷积之前首先通过cropping（图片区域丢失）或wraping（目标形变）将输入调至fixed input。在conv层和fc层之间引入SPP，可以让网络的输入是任意大小、任意形状。2）卷积共享。RCNN中，是对2000个regions分别进行卷积、全连接并分类，存在大量**重复计算**导致计算开销大，SPP只需要在原图进行一次卷积运算，然后在conv层后的卷积层上基于regions池化至所需尺寸送入fc，这样的**卷积共享避免了重复计算**，提速了两个数量级。缺点：训练时只能finetune fc层，不能finetune conv层，但这个问题在Fast RCNN中通过多任务损失函数与Roi Pooling提出得以解决。另外，SPP-Net的训练过程依然是一个多阶段的训练，这一点和R-CNN一样，并未改进

**论文细节：**





**空间金字塔池化原理。**左图，对于任意尺寸输入（SPP-Net中是conv5层的feature map，也可以应用在其他网络中），SPP将输入特征平均分为16块，4块和1块，并在每一块（Bin）上做Max pooling，同时特征的厚度保持不变，最后将这些特征串接，便可以得到**fixed length的特征向量**输入到全连接层。文章提出将SPP置于网络的卷积层和全连接层之间替换普通池化，对卷积层的特征向量**分通道**池化，卷积层可以实现在任意尺寸的图像输入下输出固定尺寸的特征向量输出，方便全连接层的连接。SPP的层数和每层的块大小可定制。比如上图，卷积层最后一层是256个filter，SPP设定为(4\*4,2\*2,1)，则对于任意图像尺寸，conv最终输出是256\*(16+4+1)维特征。

**SPP-Net测试过程。**右图，之前RCNN的测试过程是ss生成2000个区域，对每个区域先进行crop/warp至固定尺寸(227\*227)，先后进行conv,fc，然后再SVM，NMS和bounding box回归。SPP-Net从fc层开始，与RCNN是相同的，不同点是，先直接对任意尺寸图像输入进行conv，得到conv5的feature map，此时再ss生成2000区域，经过坐标变换直接在feature map的对应区域进行SPP，得到fixed length 一维向量连接fc。避免了RCNN中大量的重复计算，所以速度上提高了两个数量级。

**SPP-Net训练。**首先拿到在ImageNet预训练的AlexNet模型，用AlexNet计算Conv5层特征，根据ss生成的区域建议，从Conv5上提取到对应的SPP特征，用提取到的特征finetune全连接层（把AlexNet当做分类模型来训练）。AlexNet训练好之后，用fc7层的特征训练SVM分类器，用SPP特征训练bounding box（这里和R-CNN一样）。另外一个重要的点是，SPP使**多尺度训练**成为可能，也是第一次提出并应用多尺度训练。做法是，既然采用SPP后对输入图像的尺寸不再要求了，训练是可以在不同尺度下进行，在每个epoch（或batch）中，设置的输入尺寸不同，好处是：1)提高尺度不变性、减小过拟合；2)在不同尺度下提取特征，收敛更快，准确率提升。

精髓：https://blog.csdn.net/v1\_vivian/article/details/73275259