

1、Deeplabv1

DeepLab是结合了深度卷积神经网络(DCNNs)和概率图模型(DenseCRFs)的方法。在实验中发现DCNNs做语义分割时精准度不够的问题，根本原因是DCNNs的高级特征的平移不变性(即高层次特征映射，根源在于重复的池化和下采样)。

针对信号下采样或池化降低分辨率，DeepLab是采用的atrous(带孔)算法扩展感受野，获取更多的上下文信息。另外DeepLab采用完全连接的条件随机场(CRF)提高模型捕获细节的能力。

论文的模型基于VGG16，在Titan GPU上运行速度达到了8FPS，全连接CRF平均推断需要0.5s，PASCAL VOC-2012 达到71.6% IOU accuracy。

2. Deeplabv2

deeplabv2是相对于deeplabv1基础上的优化。deeplabv1在三个方向努力解决，但是问题依然存在：特征分辨率的降低、物体存在多尺度，DCNN的平移不变性。因DCNN连续池化和下采样造成分辨率降低，DeepLabv2在最后几个最大池化层中去除下采样，取而代之的是使用空洞卷积，以更高的采样密度计算特征映射。物体存在多尺度的问题，deeplabv1中是用多个MLP结合多尺度特征解决，虽然可以提供系统的性能，但是增加特征计算量和存储空间。论文受到我们受到spatial pyramid pooling(SPP)的启发，提出了一个类似的结构，在给定的输入上以不同采样率的空洞卷积并行采样，相当于以多个比例捕捉图像的上下文，称为ASPP(atrous spatial pyramid pooling)模块。

Deeplabv2 在之前的基础上又给出了三个方面的贡献：

一是使用Atrous Convolution 代替原来上采样的方法，比之前得到更高像素的score map，并且增加了感受野的大小；

二是使用ASPP 代替原来对图像做预处理resize 的方法，使得输入图片可以具有任意尺度，而不影响神经网络中全连接层的输入大小；

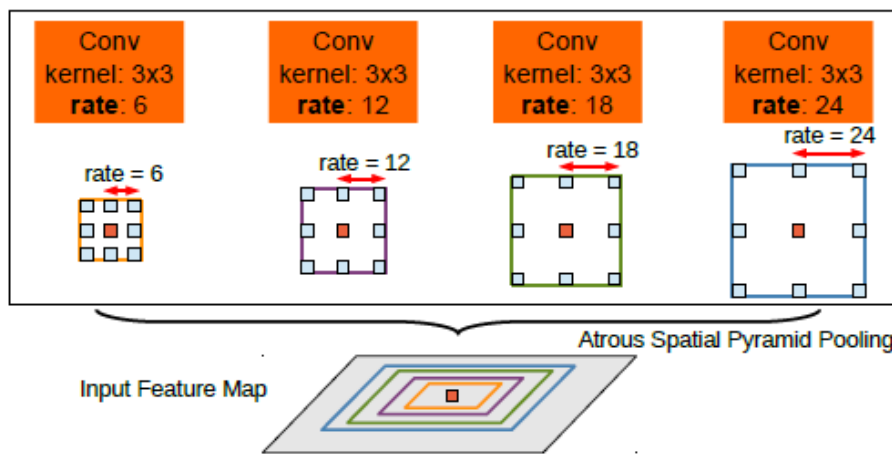


Fig. 4: Atrous Spatial Pyramid Pooling (ASPP). To classify the center pixel (orange), ASPP exploits multi-scale features by employing multiple parallel filters with different rates. The effective Field-Of-Views are shown in different colors.

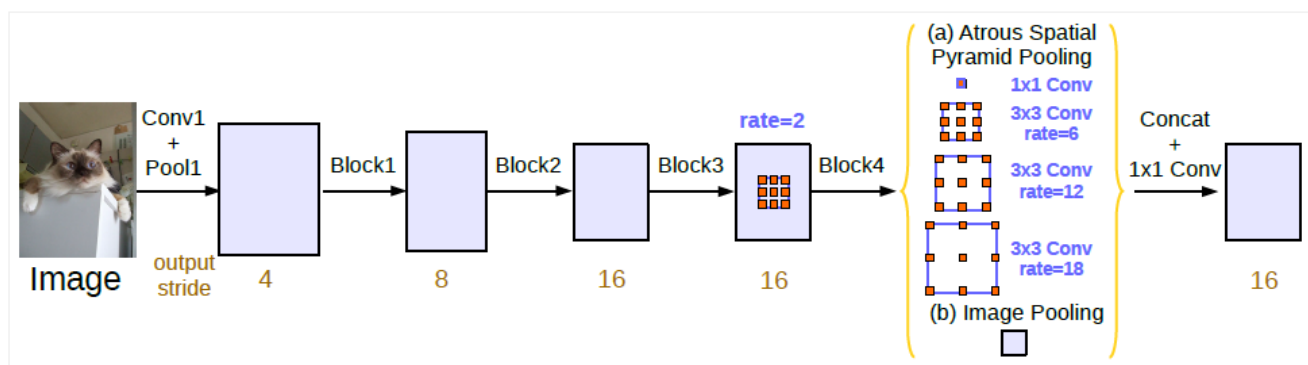


Figure 5. Parallel modules with atrous convolution (ASPP), augmented with image-level features.

三是使用全连接的CRF，利用低层的细节信息对分类的局部特征进行优化。

论文模型基于ResNet，在NVIDIA Titan X GPU上运行速度达到了8FPS，全连接CRF平均推断需要0.5s，在耗时方面和deeplabv1无差异，但在PASCAL VOC-2012 达到79.7 mIOU。

3、deeplabv3

deeplab延续到deeplabv3系列，依然是在空洞卷积做文章，但是探讨不同结构的方向。

deeplabv3论文图2比较多种捕获多尺度信息的方式：a. Image Pyramid：将输入图片放缩成不同比例，分别应用在DCNN上，将预测结果融合得到最终输出。b. Encoder-Decoder：利用Encoder阶段的多尺度特征，运用到Decoder阶段上恢复空间分辨率(代表工作有FCN、SegNet、PSPNet等工作)。c. Deeper w. Atrous Convolution：在原始模型的顶端增加额外的模块，例如DenseCRF，捕捉像素间长距离信息。d. Spatial Pyramid Pooling：空间金字塔池化具有不同采样率和多种视野的卷积核，能够以多尺度捕捉对象。