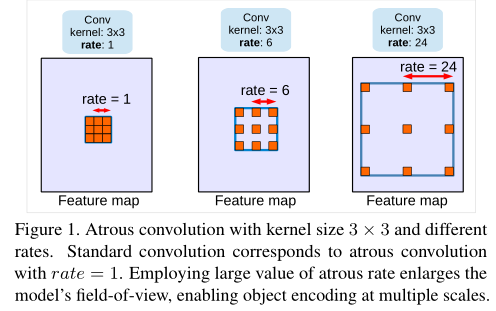
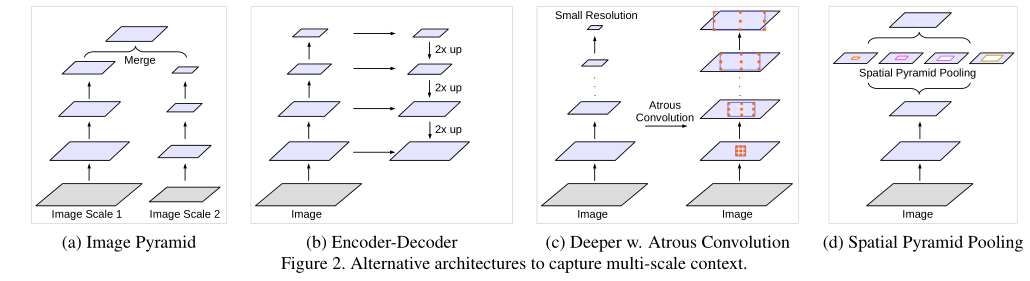
## Rethinking Atrous Convolution for Semantic Image Segmentation

### DNN进行语义分割的两大挑战：

* 1. 由DCNN中的下采样（如pooling,卷积步长等）因为丢失了很多空间信息,造成的特征图分辨率降低，这种局部的不变性会对密集的预测任务造成干扰。
* 本文提出的解决方法：带孔卷积（Atrous Convolution），它主要是通过移除网络最后几层的下采样操作以及对应的滤波器的上采样操作，来提取更紧凑的特征，相当于在不同的滤波器权重之间插入holes，其决定了DCNNs计算的特征的分辨率，而不增加新的额外学习参数。
* 

2. 物体存在的多尺度问题

解决方法：



第一种办法是对不同的尺寸的图像进行并行的处理，再融合在一起。

第二办法是使用编码和解码的对称结构，利用来自编码器部分的多尺度特征并从解码器部分恢复空间分辨率。

第三种办法是在原始的网络上级联一些模块，额外的模块级联在原始网络的顶部以捕获长距离信息。

第四种则是空间金字塔池化（spatial pyramid pooling），通过提供不同的滤波器产生的特征图，或者是采用不同尺寸的池化操作，从而以多种尺度捕捉物体。

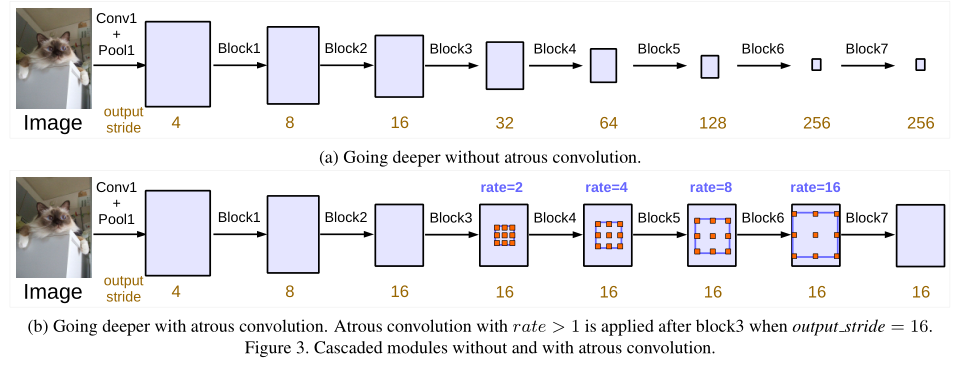
### Atrous Convolution for Dense Feature Extraction：

Atrous convolution能够明确控制FCN中特征响应的密集程度。用输出步长来表示输入图像空间分辨率与最终输出分辨率的比值。对于图像分类任务的DCNN，最终特征响应（在完全连接的层或全局池之前）比输入图像维度小32倍，因此输出stride = 32。如果想要将DCNN中的计算特征响应的空间密度加倍（即，输出步长= 16），将卷积层被rate = 2 Atrous Convolution的替换。 这使我们能够提取更密集的特征响应，而无需学习任何额外的参数。

### Going Deeper with Atrous Convolution

以串行方式设计 atrous convolution 模块 

* 复制ResNet的最后一个block，如图的 block4，并将复制后的blocks以串行方式级联；



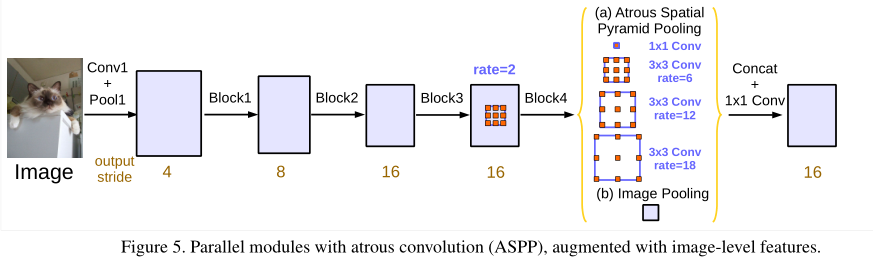
* 各block中进行3次3x3卷积， 除了最后一个block， 其它block的最后一个convolution的步长都为2。
* 这种网络模型设计的动机，引入的stride能更容易的捕获较深的blockes中的大范围信息。例如，整体图像feature可以融合到最后一个小分辨率的feature map中，如上图（a）。
* 不过，这种连续的步长式设计，对于语义分割是不利的，会破坏图像的细节信息。因此，这里采用由期望output\_stride值来确定atrous convolution的rete进行模型设计，如上图（b）。
* 采用串行的ResNet，级联block为block5、block6、block7，均为block4的复制，如果没有 atrous convolution， 其output\_stride=256。

### 多网格方法(Multi-grid Method)

* 在block4-block7内的三个卷积使用不同的rate，Multi-grid = （r 1，r 2，r 3），
* 卷积最终的 atrous rate 等于单元 rate与对应的 rate 的乘积. 例如，当 output\_stride=16， Multi\_Grid=(1,2,4)时, block4 中三个卷积rate 分别为：rates=2∗(1,2,4) = (2,4,8).

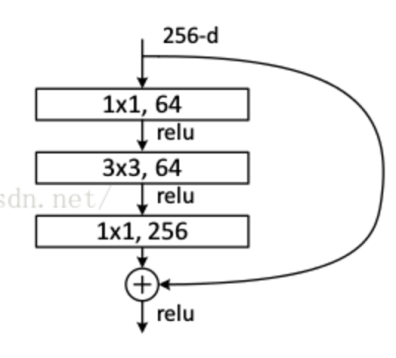
### Atrous Spatial Pyramid Pooling

* ASPP, 采用四个并行的不同atrous rate的atrou convolution对特征图进行处理，灵感来自spatial pyramid pooling。
* 不同atrous rates的ASPP有效捕捉多尺度信息。但是，随着采样率变大，有效滤波器权重（即有效特征区域，而不是补零区域的权重）的数量变得更小。当3×3的滤波器采用不同的采样率，应用于一个65×65的特征图时， 在rate值接近特征图大小的极端情况下，3×3滤波器不是捕获整个图像上下文，而是退化为简单的1×1滤波器。
* 为了克服这个问题并将全局上下文信息结合到模型中，
* 采用全局平均池化(global average pooling)对模型的 feature map 进行处理，将得到的图像级特征输入到256个 1×1 convolution 中，并加上 batch normalization，然后将特征进行双线性上采样(bilinearly upsample)到特定的空间维度.



* Conv1:7x7,64

Pool1:3x3,stride =2

* Block1:3次残差模块,64
* Block2:4次残差模块,128
* Block3:6次残差模块,256
* Block4:3次残差模块,512
* 

改善的ASPP包括（a），当输出步长= 16时，包括一个 1×1 卷积 和三个3×3 卷积， rate=（6,12,18），以及 （b）图像级特征，如图所示。当输出步幅= 8时，rate加倍。 然后将所有分支的结果特征串联起来，并在最终的1×1卷积之前通过另一个1×1卷积（也有256个滤波器和批量归一化），从而产生最终的预测

### 实验结果：

