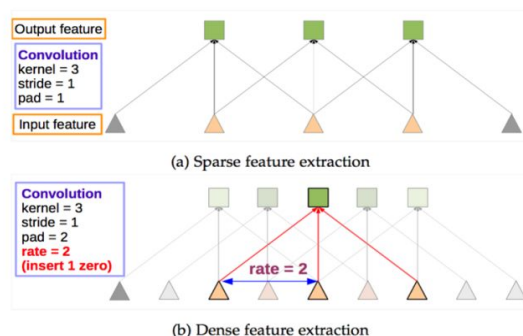


1. CVPR 2018 文章：[DenseASPP for Semantic Segmentation in Street Scenes](#)

作者来自初创无人驾驶公司 DeepMotion Inc. 后三位作者为微软亚洲研究院研究员。

(1) 问题介绍：

- Atrous convolution 被提出以在扩大感受野的同时不增加过多的计算量。



(a) 是正常的卷积，最下面是输入特征，最上面是输出特征，中间是卷积。卷积就是那三条线。

(b) 是带洞卷积，中间三条红色的线就是卷积核大小为 3 时的卷积过程， $rate=n$ ，则应该每间隔 $n-1$ 个点进行一次卷积。Note that standard convolution is a special case in which $rate\ r = 1$ 。

- ASPP 被提出以融合不同尺寸的特征。

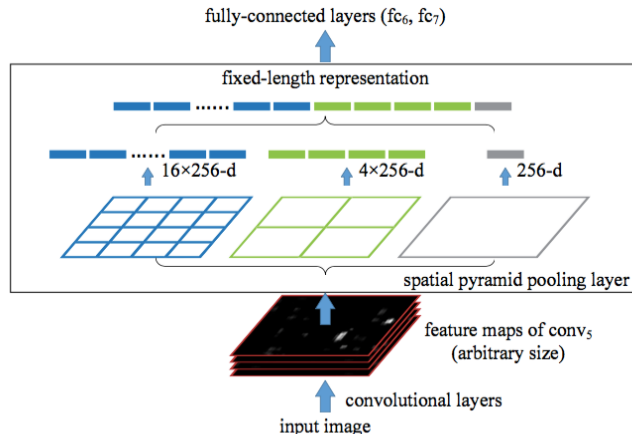


Figure 3: A network structure with a **spatial pyramid pooling layer**. Here 256 is the filter number of the conv₅ layer, and conv₅ is the last convolutional layer.

将任意尺寸的 feature map 分别切分成 16、4、1 份（这里 16、4、1 即为各个层 bin 的数量），再对每一份进行池化操作（一般选择 max pooling），将池化后的结果拼接得到固定长度的特征向量（图中的 256 为 filter 的个数），送入全连接层进行后续操作。

ASPP 能够使得输出的 Feature Map 融合多个不同 d 所带来的多个尺寸的感受野，包含更丰富的信息。

- 问题：ASPP 的 d 不能过大（如 $d>24$ ），否则模型性能会急剧下降，因为每次卷积

参与决策的点距离中心点过远，相关性不是非常大。需要设计一种结构能够在融合多种尺寸感受野特征图的同时还保持较大、合理的感受野。

(2) 网络结构

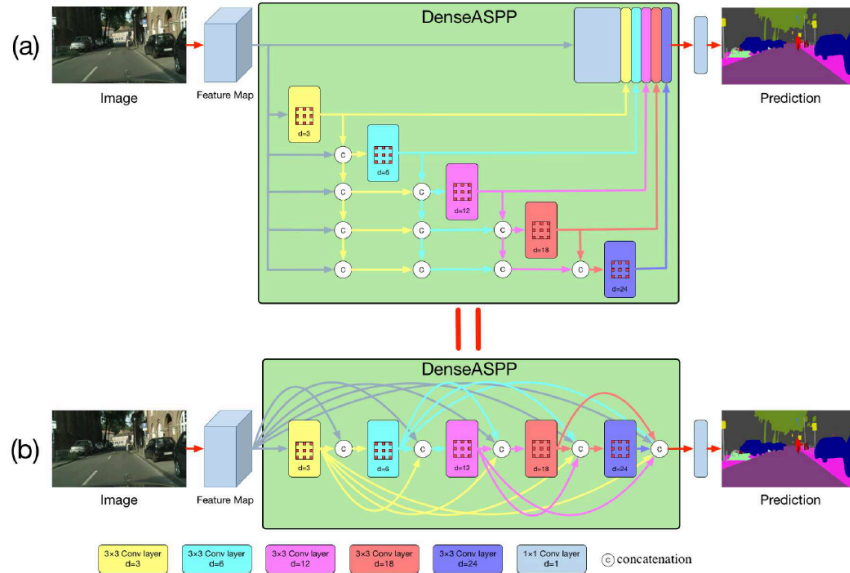


Figure 2. The structure of DenseASPP, (a) illustrate DenseASPP in detail, the output of each dilated convolutional layer is concatenated with input feature map, and then feed into the next dilated layer. Each path of DenseASPP compose a feature representation of correspond scale. (b) illustrate this structure in a more concrete version

先将图片送入去除 Pooling 和全连接层的 DenseNet，得到 Feature Map。然后如图(a)所示，经过不同 d 的卷积层，得到更高层的特征。这些特征之间采用级连方式相接，并在经过卷积层后使用 Pyramid Pooling 来得到相同维度的特征，拼接在一起。送入卷积层后将较多的 Channel 数量压缩到 3，然后经过上采样得到预测结果。

(3) DenseASPP 的优点

- DenseASPP 通过级连的方式，在由经过 $d=3$ 的卷积运算的特征图的基础上，再次进行 $d=6$ 的卷积运算。间接的达到了比 $d=6$ 卷积感受野更大的目的。
- 不同 level 的特征进行级连的结构，保证了低层的结构信息能够和高层的语义信息进行很好的融合。避免两者只有其一导致分割性能下降。

(4) 实验结果

Table 1. Category-wise comparison on the Cityscapes test set.

Methods	mIoU	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle
FCN-8s [16]	65.3	97.4	78.4	89.2	34.9	44.2	47.4	60.1	65	91.4	69.3	93.9	77.1	51.4	92.6	35.3	48.6	46.5	51.6	66.8
DeepLabv2-CRF [2]	70.4	97.9	81.3	90.3	48.8	47.4	49.6	57.9	67.3	91.9	69.4	94.2	79.8	59.8	93.7	56.5	67.5	57.5	57.7	68.8
FRRN [19]	71.8	98.2	83.3	91.6	45.8	51.1	62.2	69.4	72.4	92.6	70	94.9	81.6	62.7	94.6	49.1	67.1	55.3	53.5	69.5
RefineNet [15]	73.6	98.2	83.3	91.3	47.8	50.4	56.1	66.9	71.3	92.3	70.3	94.8	80.9	63.3	94.5	64.6	76.1	64.3	62.2	70
PEARL [11]	75.4	98.4	84.5	92.1	54.1	56.6	60.4	69	74	92.9	70.9	95.2	83.5	65.7	95	61.8	72.2	69.6	64.8	72.8
GCN [18]	76.9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DUC [21]	77.6	98.5	85.5	92.8	58.6	55.5	65	73.5	77.9	93.3	72	95.2	84.8	68.5	95.4	70.9	78.8	68.7	65.9	73.8
PSPNet [24]	78.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ResNet-38 [22]	78.4	98.5	85.7	93.1	55.5	59.1	67.1	74.8	78.7	93.7	72.6	95.5	86.6	69.2	95.7	64.5	78.8	74.1	69	76.7
DenseASPP(Ours)	80.6	98.7	87.1	93.4	60.7	62.7	65.6	74.6	78.5	93.6	72.5	95.4	86.2	71.9	96.0	78.0	90.3	80.7	69.7	76.8

