

FCN

一、面临的问题

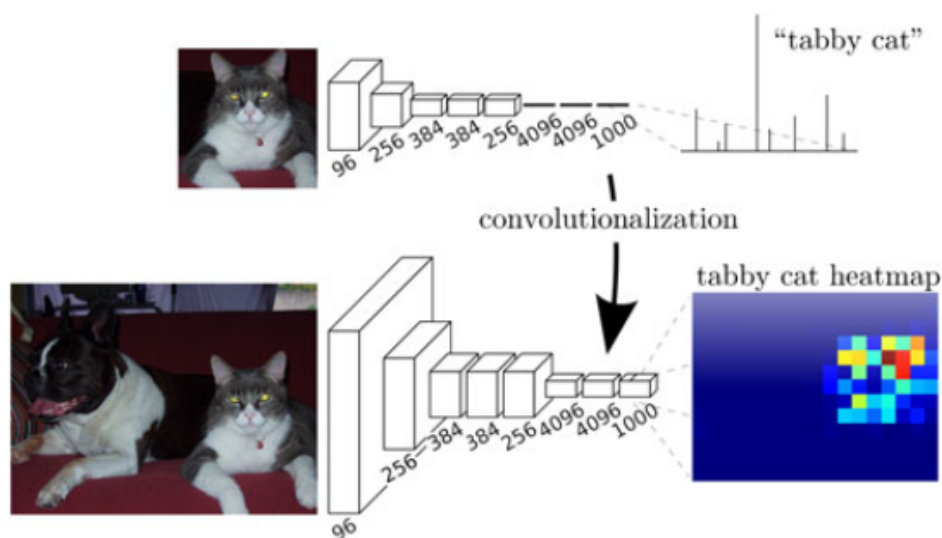
1. 要使用卷积神经网络来处理图像分割

每个像素点都有21种分类的可能性，要对图像做像素级别的预测，即按照每个像素的可能性给出预测值

2. 更深的网络带来更多的信息丢失

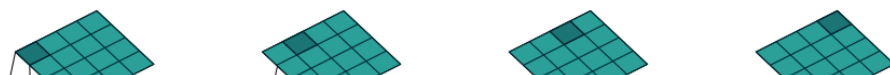
二、解决方案

1. 将全连接层替换为卷积层



2. 对图片进行上采样

反卷积



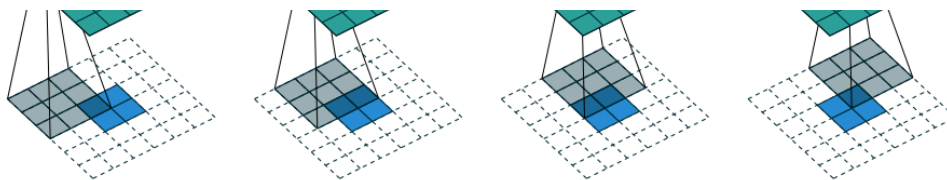


Figure 4.1: The transpose of convolving a 3×3 kernel over a 4×4 input using unit strides (i.e., $i = 4$, $k = 3$, $s = 1$ and $p = 0$). It is equivalent to convolving a 3×3 kernel over a 2×2 input padded with a 2×2 border of zeros using unit strides (i.e., $i' = 2$, $k' = k$, $s' = 1$ and $p' = 2$).

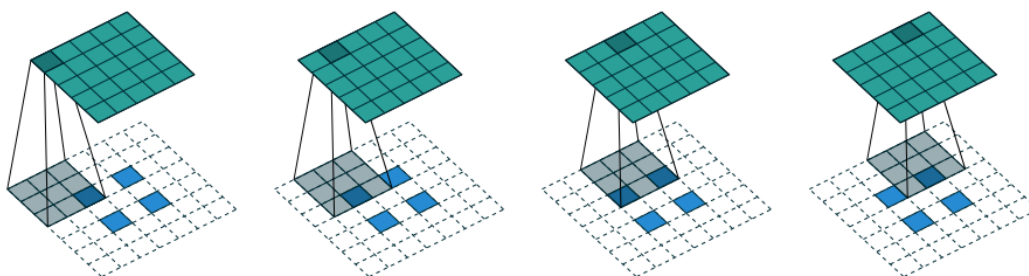
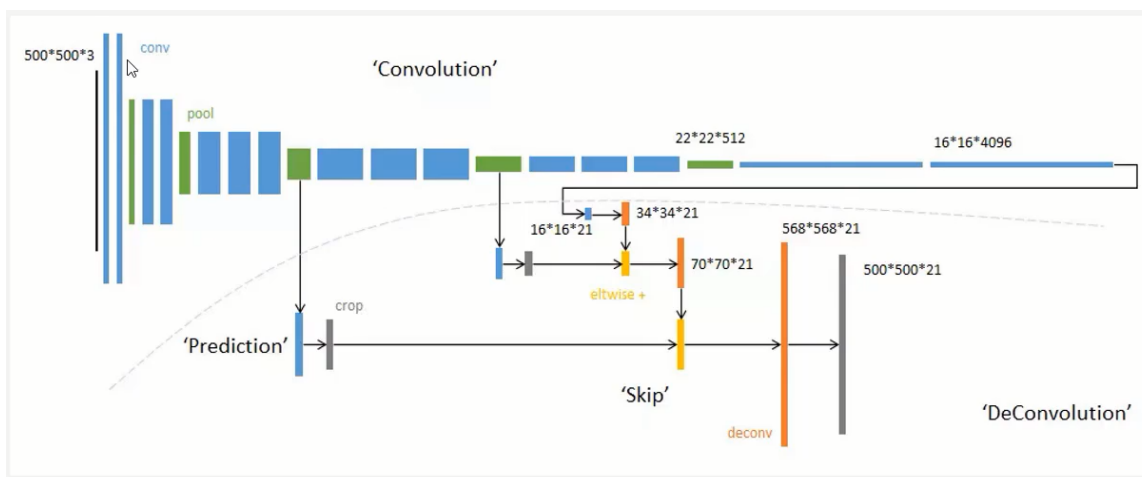


Figure 4.5: The transpose of convolving a 3×3 kernel over a 5×5 input using 2×2 strides (i.e., $i = 5$, $k = 3$, $s = 2$ and $p = 0$). It is equivalent to convolving a 3×3 kernel over a 2×2 input (with 1 zero inserted between inputs) padded with a 2×2 border of zeros using unit strides (i.e., $i' = 2$, $\tilde{i}' = 3$, $k' = k$, $s' = 1$ and $p' = 2$).

3. 使用skip结构跳级连接

在还原图像的时候能够得到更多原图所拥有的信息



三、效果

TABLE 3
Comparison of FCN Variations

pixel acc.	mean acc.	mean IU	f.w. IU
------------	-----------	---------	---------

FCN-32s	90.5	76.5	63.6	83.5
FCN-16s	91.0	78.1	65.0	84.3
FCN-8s at-once	91.1	78.5	65.4	84.4
FCN-8s staged	91.2	77.6	65.5	84.5
FCN-32s fixed	82.9	64.6	46.6	72.3
FCN-pool5	87.4	60.5	50.0	78.5
FCN-pool4	78.7	31.7	22.4	67.0
FCN-pool3	70.9	13.7	9.2	57.6

Learning is end-to-end with batch size one and high momentum, with the exception of the fixed variant that fixes all features. Note that FCN-32s is FCN-VGG16, renamed to highlight stride, and the FCN-poolX are truncated nets with the same strides as FCN-32/16/8s. Scores are evaluated on a subset of PASCAL VOC 2011 segval.⁵

Deeplab V1

一、面临的问题

深度卷积神经网络的位置精度和分辨率不够

-> 两方面：减少位置信息的丢失，提高细节感知能力

二、解决方案

1. 减少位置信息的丢失方面，用空洞卷积取代池化

池化的作用：

- 缩小特征层的尺寸，减少对计算资源的需求
- 快速扩大感受野，利用更多的上下文信息进行分析

池化的劣势：

- 卷积操作具有平移不变性，对位置信息不敏感，这对分类网络是好事，但是对分割网络是坏事，因为丢失了位置信息；池化操作加剧了位置信息的模糊

空洞卷积取代池化：

- 在去掉 pooling 层后，感受野会随之下降，解决方法是引入**空洞卷积**（Atrous Convolution），这样可以在**不增加计算量**的情况下扩大感受野
- 相比于先pooling后再卷积，使用空洞卷积可以让**特征更密**（因为pooling减少了分辨率，这使得空洞卷积和普通卷积的位置不再是一一对应的了）

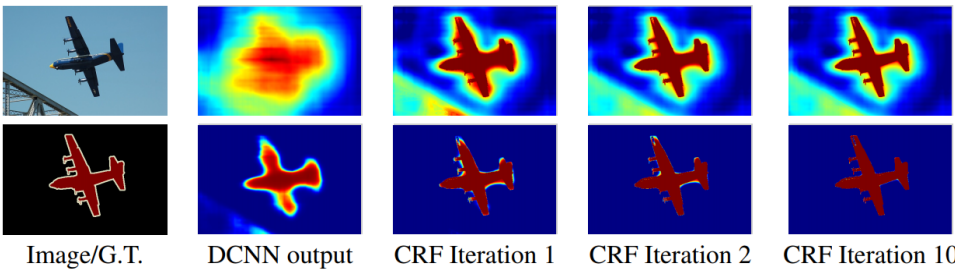
2. 提高细节感知方面，引入CRF

以对象为中心的决策需要对空间变换保持不变性，这限制了模型的空间精度

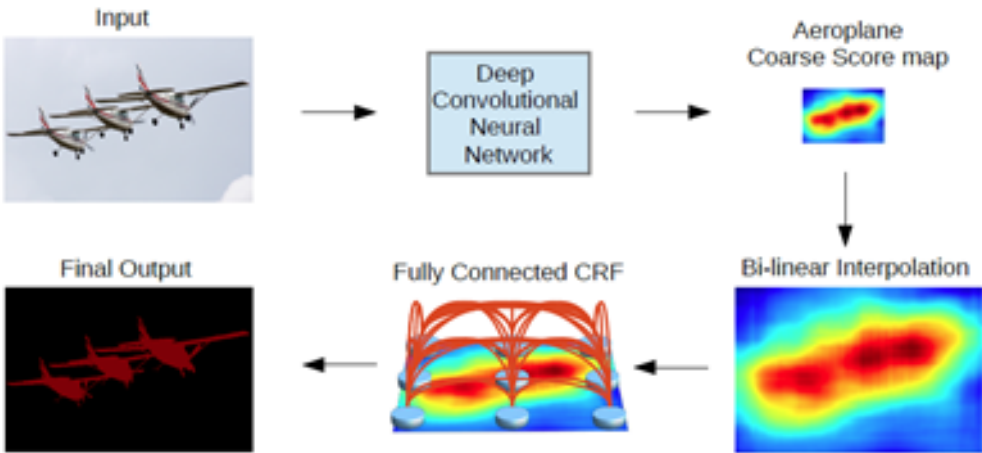
通过采用完全连接的条件随机场 (CRF) 提高模型捕获细节的能力

在进行分割的时候考虑空间上临近的像素并进行评估，依评估结果，按概率选择整张图片整体上的最优的分割方案，达到去除“孤岛”的目的

多次使用CRF进行迭代可以进一步提高分割图像的精度



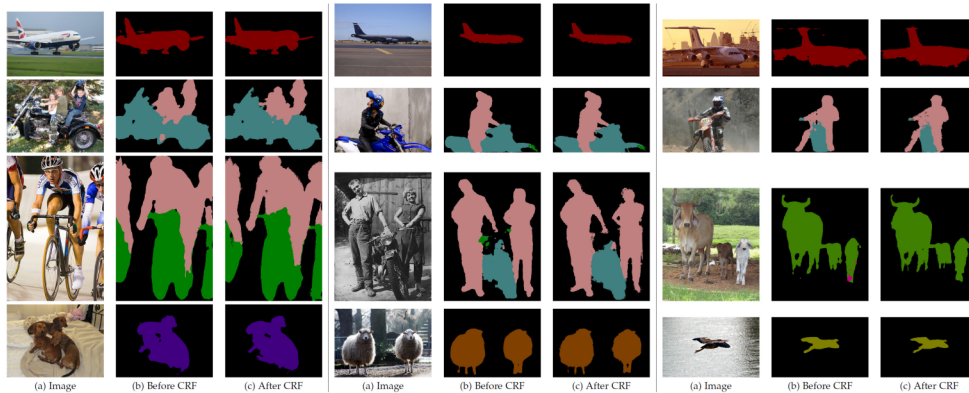
三、整体结构



四、效果

Method	mean IOU (%)
MSRA-CFM	61.8
FCN-8s	62.2
TTI-Zoomout-16	64.4
DeepLab-CRF	66.4
DeepLab-MSc-CRF	67.1
DeepLab-CRF-7x7	70.3
DeepLab-CRF-LargeFOV	70.2

crf的效果:



- 速度: 凭借“atrous”算法, 我们的密集 DCNN 以 8 fps 运行, 而全连接 CRF 的平均场推断需要 0.5 秒,
- 准确性: 我们在 PASCAL 语义分割挑战中获得了最先进的结果, 优于 Mostajabi 等人的次优方法。 (2014) 7.2% 的边际和
- 简单性: 我们的系统由两个相当完善的模块级联组成, DCNNs 和 CRFs。

DeepLab V2

一、面临的问题:

多尺度物体 (the existence of objects at multiple scales)

空洞卷积存在的劣势:

- 局部信息丢失: 由于空洞卷积的计算方式类似于棋盘格式, 某一层得到的卷积结果, 来自上一层的独立的集合, **没有相互依赖**, 卷积结果之间没有相关性
- 远距离获取的信息没有相关性: 由于空洞卷积稀疏的采样输入信号, 使得远距离卷积得到的信息之间没有相关性

DCNN 分数图可以预测物体的存在和粗略位置, 但不能真正描绘它们的边界

二、解决方案

ASPP (Atrous Spatial Pyramid Pooling)

结合 ASPP 的代码来看:

- ASPP 层先对输入做一系列不同 padding 和 dilation 的空洞卷积, 这些卷积通过调整这两个参数以

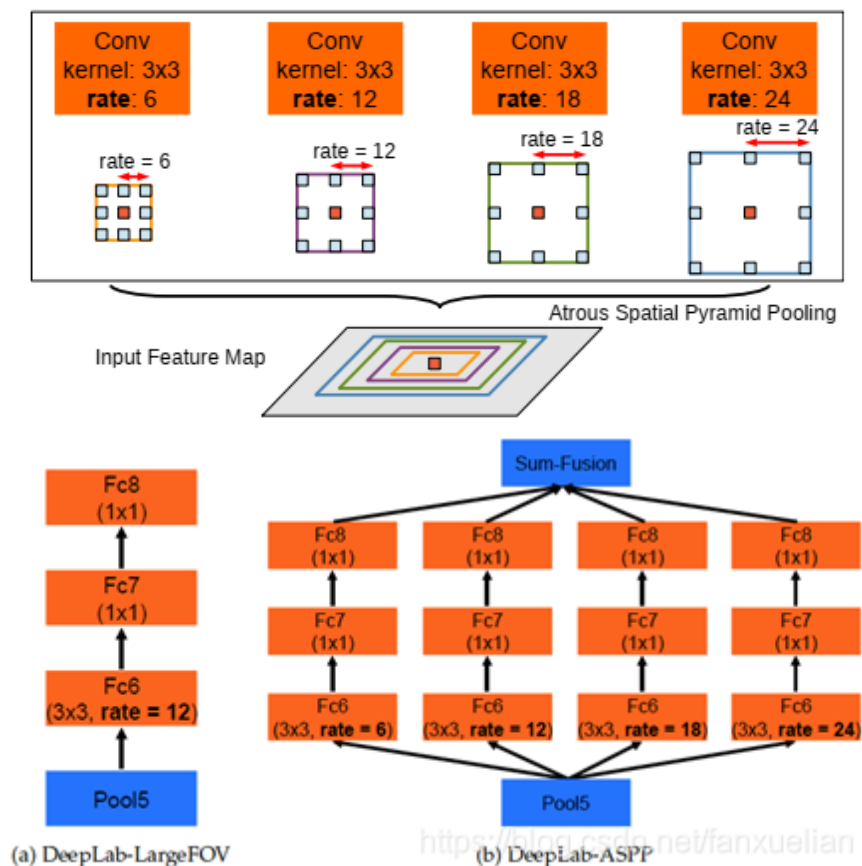
满足输出通道数和尺寸保持一致

- 让这些空洞卷积的输出通过 concat 相加，得到 ASPP 层的输出

ASPP 层解决了空洞卷积信息丢失的劣势，也同时增大了感受野，从不同的尺度上提取特征

在卷积之前以多个速率对给定特征层进行重采样，即使用具有互补有效视野的多个过滤器探测原始图像，从而在多个尺度上捕获对象以及有用的图像上下文。

ASPP 不是重新采样特征，而是使用**具有不同采样率的多个并行空洞卷积层**有效地实现了这种映射



三、效果

Method	mIOU
<i>pre-release version of dataset</i>	
Adelaide Context [40]	66.4
FCN-8s [14]	65.3
DeepLab-CRF-LargeFOV-StrongWeak [58]	64.8
DeepLab-CRF-LargeFOV [38]	63.1
CRF-RNN [59]	62.5
DPN [62]	59.1
Segnet basic [100]	57.0

Segnet extended [100]	56.1
<i>official version</i>	
Adelaide_Context [40]	71.6
Dilation10 [76]	67.1
DPN [62]	66.8
Pixel-level Encoding [101]	64.3
DeepLab-CRF (ResNet-101)	70.4

DeepLab V3

一、面临问题

使用空洞卷积会带来：

- 随着rate的增大，一次空洞卷积覆盖到的有效像素（特征层本身的像素，相应的补零像素为非有效像素）会逐渐减小到1，也就是说太大的 padding 和 dilation 带来了非中心采样点位置过于偏僻，要么太远，要么延申到 padding 补0区域，这些像素采样是没有意义的
- 这就与我们扩展感受野，获取更大范围的特征的初衷相背离了

二、解决方案

1. 改进 ASPP 拓展网络的宽度

措施：

- 使用1x1的卷积：即当rate增大以后3x3卷积的退化形式，替代3x3卷积，减少参数个数
- 增加image pooling：即全局池化，来补充全局特征。具体做法是对每一个通道的像素取平均，之后再上采样到原来的分辨率
- 在空洞卷积之后使用 Batch Normalization

改进结果：

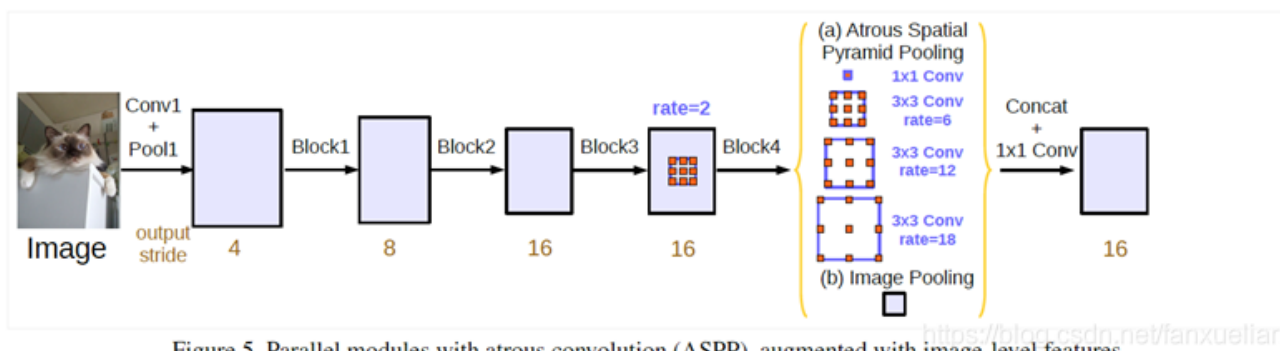


Figure 3. Parallel modules with atrous convolution (ASCT), augmented with image-level features.

2. 使用空洞卷积加深网络

在级联结构的空洞卷积中，加入ResNet的striding结构，从而更容易地在更深的块中捕获远程信息

3. 舍弃CRF

三、效果

Method	Coarse	mIOU
DeepLabv2-CRF [11]		70.4
Deep Layer Cascade [52]		71.1
ML-CRNN [21]		71.2
Adelaide_context [55]		71.6
FRRN [70]		71.8
LRR-4x [25]	✓	71.8
RefineNet [54]		73.6
FoveaNet [51]		74.1
Ladder DenseNet [46]		74.3
PEARL [42]		75.4
Global-Local-Refinement [93]		77.3
SAC_multiple [94]		78.1
SegModel [75]	✓	79.2
TuSimple_Coarse [84]	✓	80.1
Netwarp [24]	✓	80.5
ResNet-38 [86]	✓	80.6
PSPNet [95]	✓	81.2
DeepLabv3	✓	81.3