

Learning Notes

Mentor: Ms. Bian

Student: ZUO H.L.

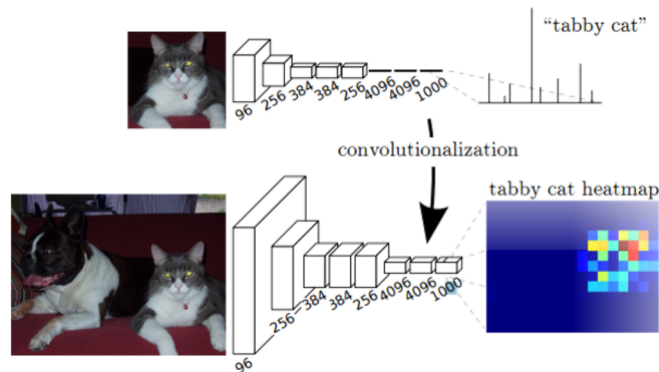
FCN

Fully Convolution Network for Semantic Segmentation

<https://blog.csdn.net/u012759136/article/details/52434826>

1. Main Technics

- Convolutional



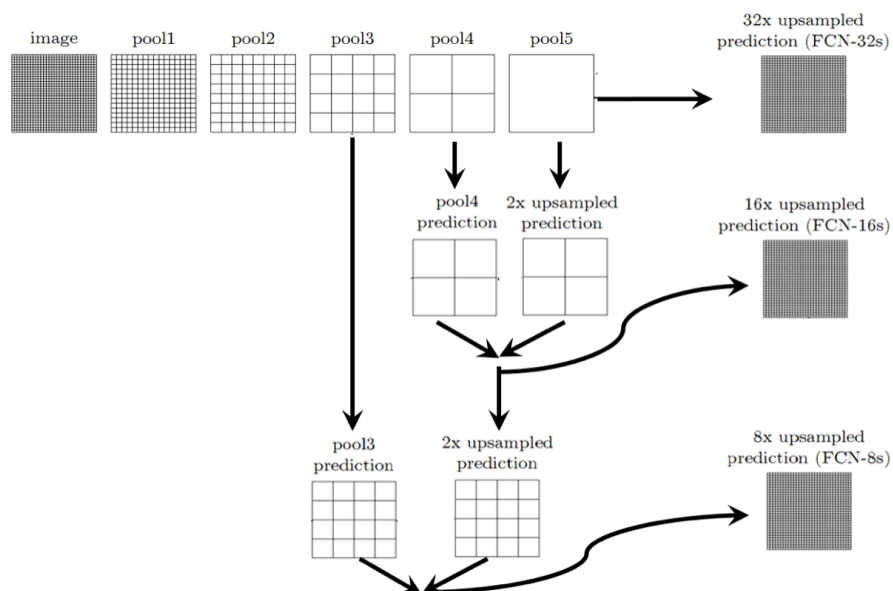
■

- Replace fully connected network using conv so as to capture the spatial information;

- Upsample

- Deconvolution (conv_transpose): 为了还原被 pooling 缩小的图;

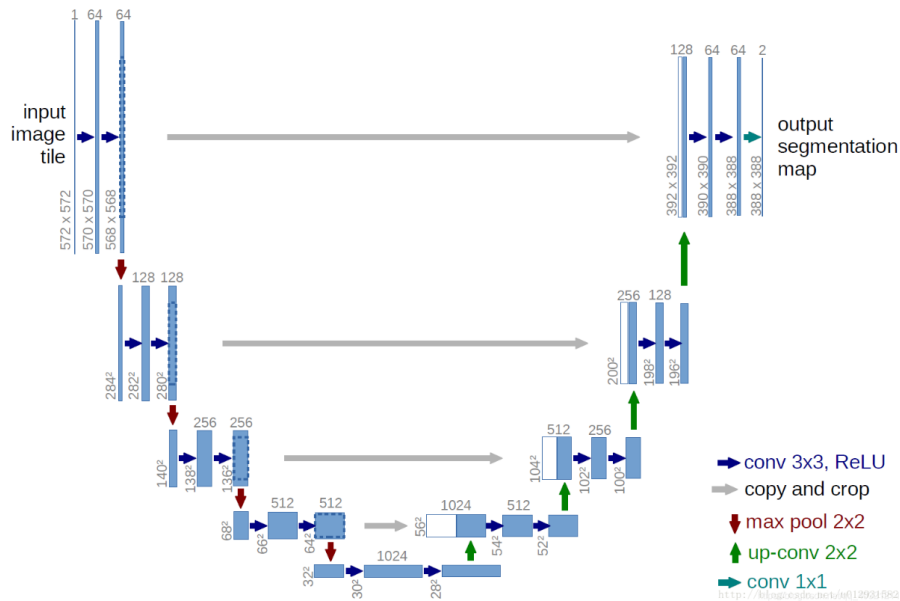
- Skip Layer



■

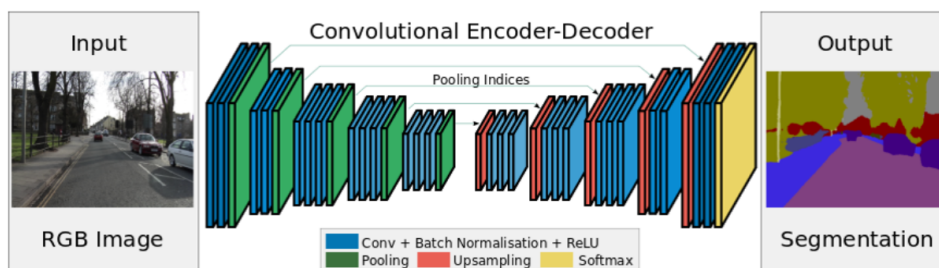
- 通过将不同池化层的输出上采样，得到优化后的输出;

Unet

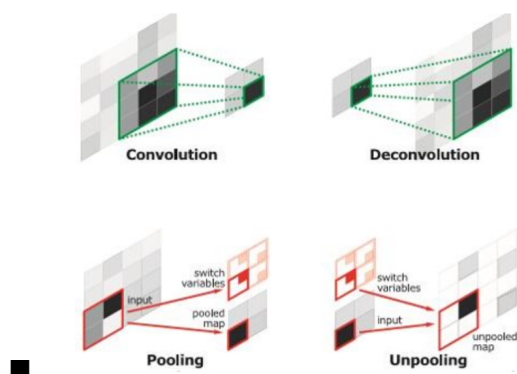


- 采用 FCN 的思想，四个下采样提取目标特征，再通过四个上采样，最后逐个对其像素点进行分割，那么这实际上是一个基于编码器（encode）-解码器（decode）思想。
 - 下采样方法（压缩图像的特征）：使用大步长的卷积；池化；使用 Padding 的卷积。
 - 上采样方法：像素插值（双线性插值、邻近插值——信息丢失少，速度适中），反卷积（转置卷积 T.conv——参数少，速度快），反池化（Unpooling），像素融合（通道信息平铺，不丢失信息）
 - 在 Unet 中使用了转置卷积实现上采样，在 Unet 中每一个上采样 Block 里，运用了一个跳连接把前面一部分特征 Concat 到了上采样后的特征图上。（思想类似 ResNet），目的是使上采样后的特征图有更多的浅层语义信息，增强分割精度，在最后一层直接使用 sigmoid 二分类把 mask 分为前景和背景。

SegNet



- It is Auto-Encoder like.
- The structure mainly applies Deconvolution and Unpooling.



Deeplab v1

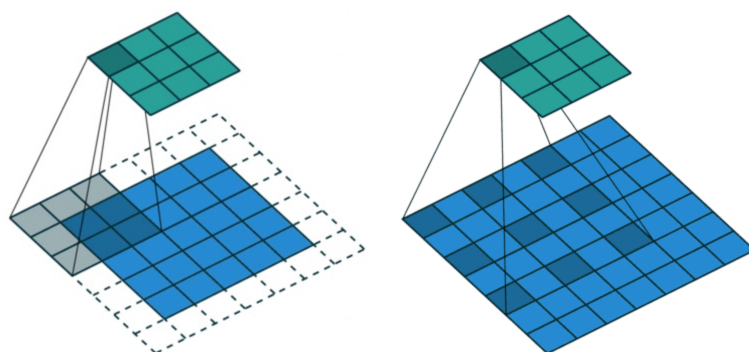
<https://zhuanlan.zhihu.com/p/75333140>

1. Previous models' problems:

- Lose information in pooling;
- Not use of relationship of probability of labels;

2. 空洞卷积, Dilated Convolution: 为了解决 pooling 中信息丢失的问题

(1) 理解



(dilated convolution)

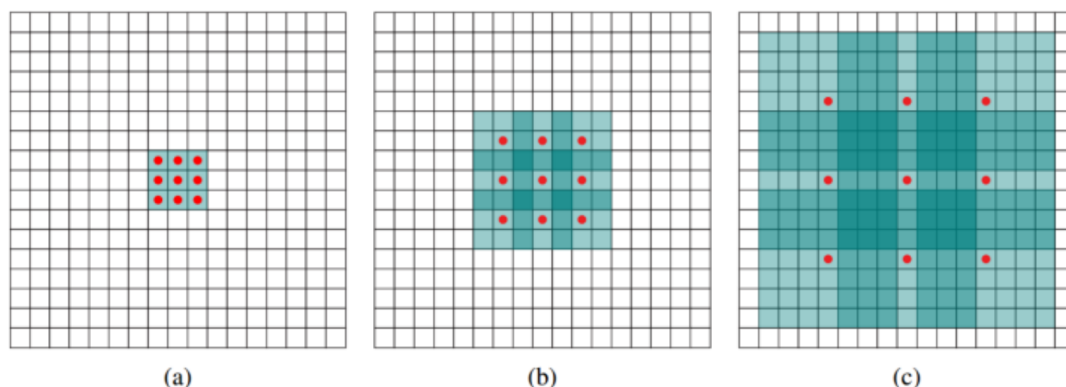
- dilated convolution 多了一个 hyper-parameter 称之为 dilation rate 指的是 kernel 的间隔数量(e.g. 正常的 convolution 是 dilatation rate 1)。

(2) Motivation

- Deep CNN 对于其他任务还有一些致命性的**缺陷**。主要问题有：
 - Up-sampling / pooling layer (e.g. bilinear interpolation) is deterministic. (参数不可学习)
 - 内部数据结构丢失；空间层级化信息丢失。因为 DCNN 是先 pooling 再 upsampling (deconvolution)；
 - 小物体信息无法重建 (假设有四个 pooling layer 则 任何小于 $2^4 = 16$ pixel 的物体信息将理论上无法重建。)

在这样问题的存在下，语义分割问题一直处在瓶颈期无法再明显提高精度，而 dilated convolution 的设计就良好的避免了这些问题。

- 主要的问题就是，想让增加感受野，并且看到更多的信息；



- 从左到右分别是 1-dilated conv, 2-dilated conv, 4-dilated conv;
- 三层依次搭配，就能使感受野达到 15x15;
- 对比传统的 conv 操作，3 层 3x3 的卷积加起来，stride 为 1 的话，只能达到 $(kernel-1)*layer+1=7$ 的感受野，也就是和层数 layer 成线性关系，而 dilated conv 的感受野是指数级的增长。（空洞卷积的主要作用是在增大感受野的同时，不增加参数数量，而且 VGG 中提出的多个小卷积核代替大卷积核的方法，只能使感受野线性增长，而多个空洞卷积串联，可以实现指数增长。）

(3) Drawbacks

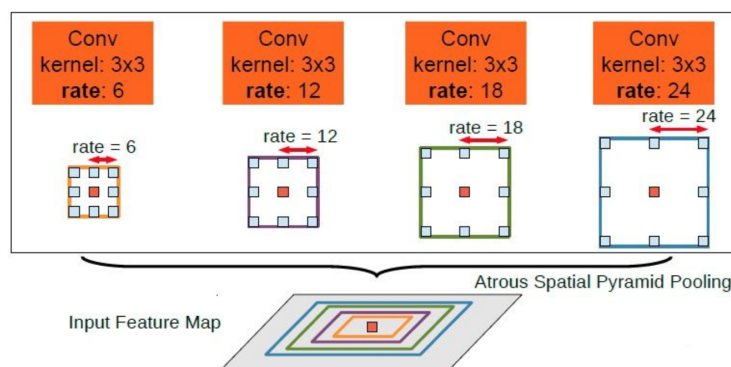
- The gridding effect: 信息的连续性遭到损失，意味着并不是所有的 pixel 都会参与计算；
- Long-ranged information might be not relevant: 在较小物体的表现会下降；

3. CRF, Conditional Random Fields 条件随机场

简单来讲就是每个像素点作为节点，像素与像素间的关系作为边，即构成了一个条件随机场。通过二元势函数描述像素点与像素点之间的关系，鼓励相似像素分配相同的标签，而相差较大的像素分配不同标签，而这个“距离”的定义与颜色值和实际相对距离有关。所以这样 CRF 能够使图片在分割的边界出取得比较好的效果。

Deeplab v2

1. 引入 Atrous Spatial Pyramid Pooling ,ASPP（最重要的改变）



- 可以对不同大小的图像目标有很好的分割效果;
- 多尺度并行;
- 2. 基础层由 VGG16 改为 ResNet;
- 3. 使用不同的学习率;

Deeplab v3

1. 使用了级联模块 ResNet;
2. 改进了 ASPP;
 - a) ASPP 中应用了 BN 层
 - b) 随着采样率的增加, 滤波器中有效的权重减少了(有效权重减少, 难以捕获原距离信息, 这要求合理控制采样率的设置)
 - c) 使用模型最后的特征映射的全局平均池化(为了克服远距离下有效权重减少的问题)