Large Kernel Matters

论文信息

论文地址：Large Kernel Matters ——Improve Semantic Segmentation by Global Convolutional Network

发表日期：8 Mar 2017

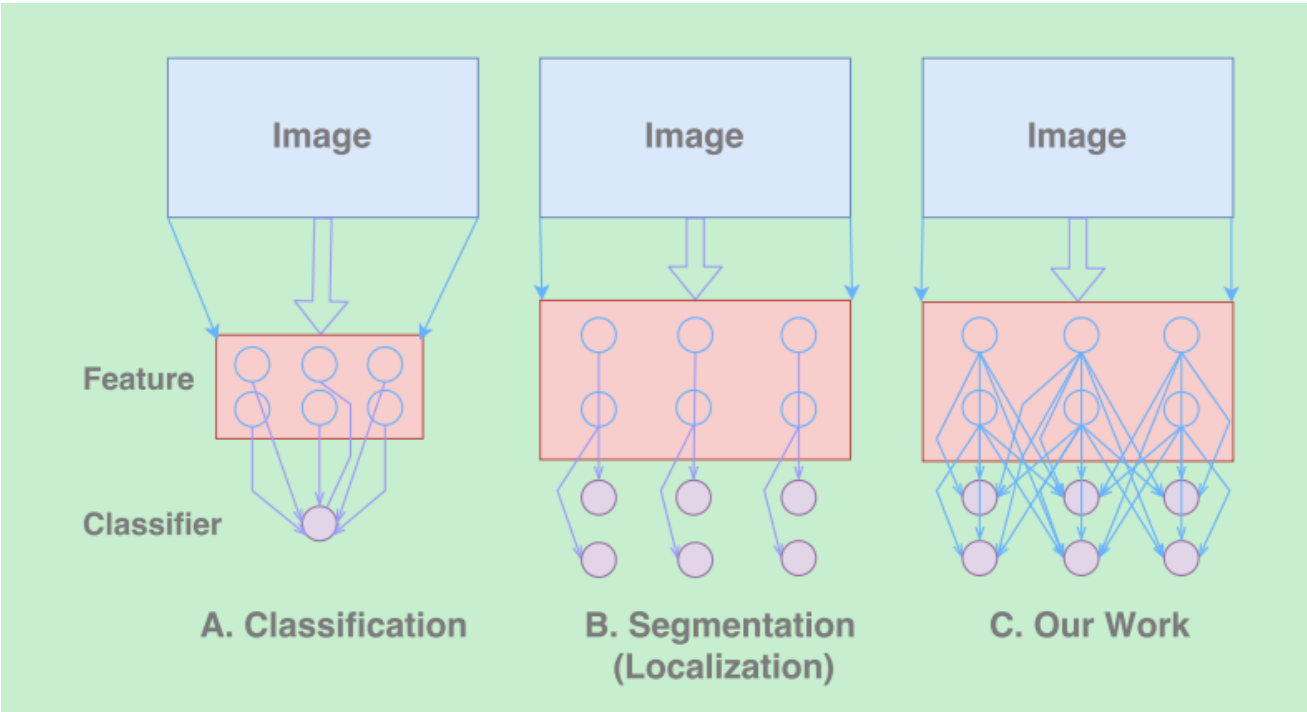
创新点

1. 提出全局卷积网络（Global Convolutional Network，GCN），用以同时提高语义分割中分类和定位的准确度。
2. 提出Boundary Refinement block（BR），用以提高物体边界的定位。

思想

在语义分割中，有两个任务，一个是对像素进行分类，一个是对像素进行定位。这两个任务通常是矛盾的：对于分类任务而言，模型需要对各种变换（比如旋转和平移）具有不变形，但是对于定位任务而言，模型又必须对各种变换保持敏感，因为每个像素都需要在正确的位置上进行分类。

现在很多语义分割方法的主要目标是为了定位，比如下图B所示，这可能会降低分类性能。因此本文提出了GCN来解决以上的矛盾，同时改善分类和定位的性能，如下图C。



基于以上的讨论，本文提出了模型的设计思路：

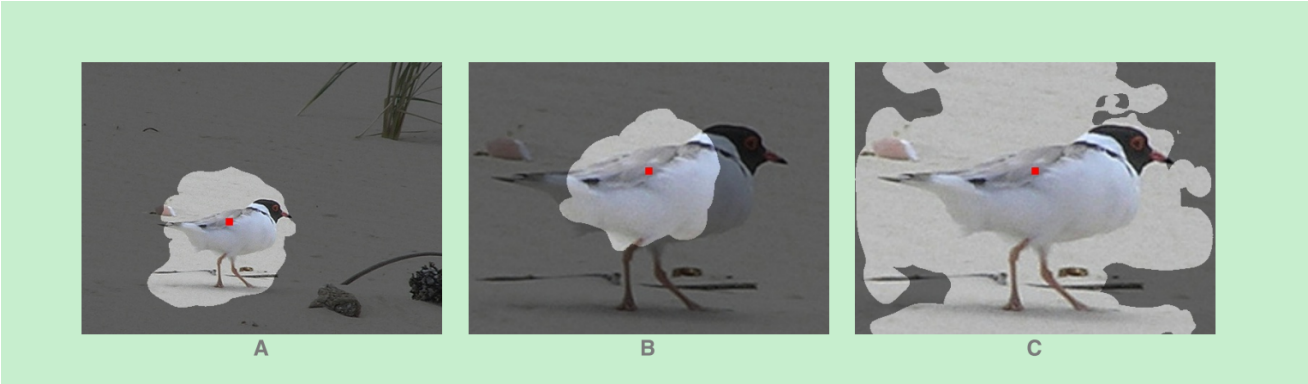
1. 从定位角度出发，模型使用FCN来保持位置信息，不使用全连接层或者global pooling，因为这些操作会丧失定位信息。
2. 从分类角度出发，网络应该使用大卷积核，这样能够使分类器具备更强的分类能力来应对各种变换。

3. 使用Boundary Refinement block来进行边界对齐，该模块使用残差结构。

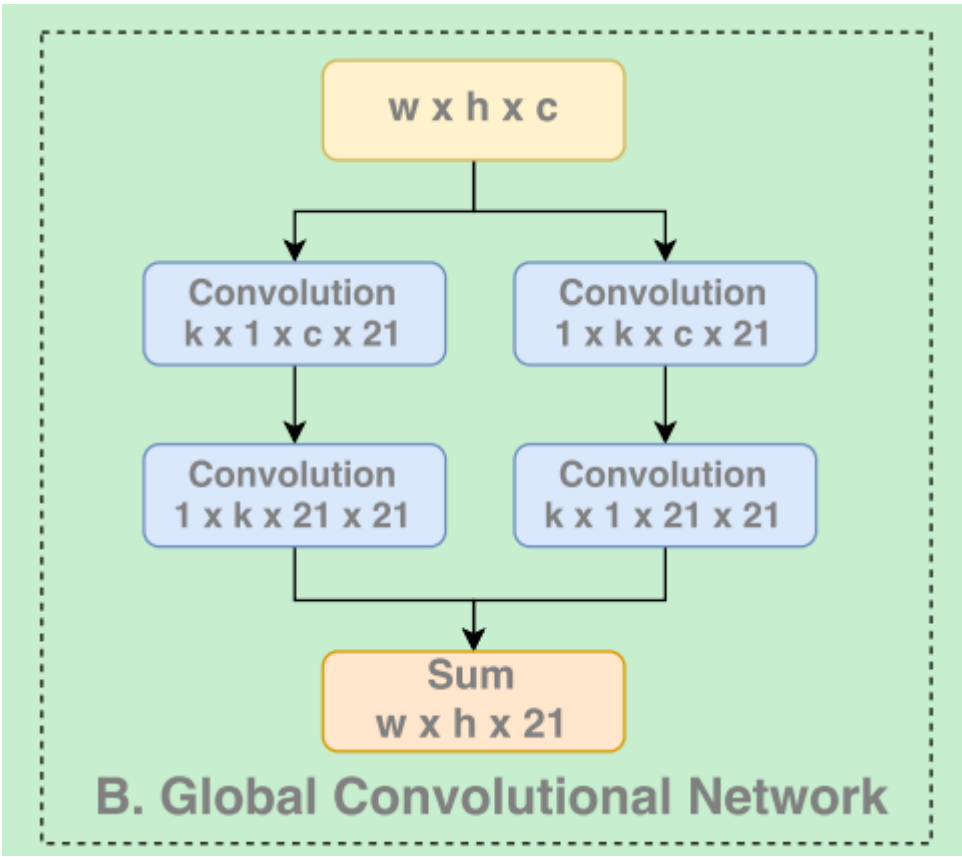
模型

1.GCN

作者发现，目前state-of-the-art的语义分割方法的设计都是为了更好地进行定位，这在某种程度上可能会降低分类的效果。而分类效果变差可能由于感受野造成的。作者举了个例子，如下图所示，当分类器有效感受野（注意这里的有效感受野不同于感受野。有效感受野的概念来自于论文：Object detectors emerge in deep scene cnns ，这篇论文中称，尽管GoogleNet和ResNet等深度网络的感受野通常都大于原图，但网络只能在感受野的一个很小区域获得有效信息，称为有效感受野）足以覆盖整个物体的时候，这时候分类器可能可以正确地分类，但是当图像的尺度变大的时候，这个时候有效感受野就会覆盖不了整个物体，这对分类是非常不利的。

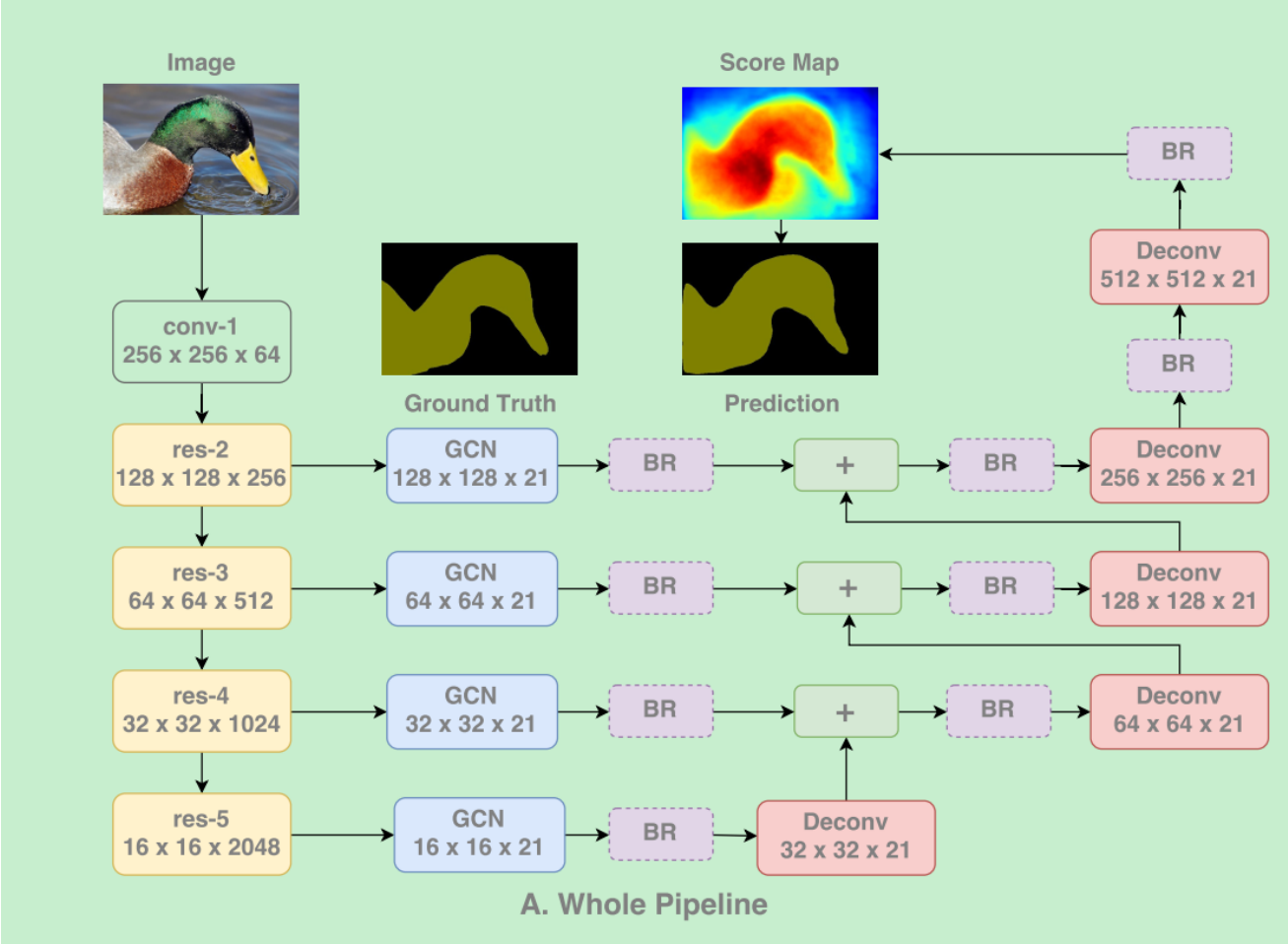


基于此，作者提出，网络设计要使用FCN，不能使用fully-connected layer或者global pooling，否则丧失位置。另外，卷积的kernel size越大越好。但是越大的卷积核，计算量也越大，所以作者提出使用1xk+kx1和kx1+1xk的卷积组合，这相当于连接了feature map的kxk大小的区域。另外值得注意的是，这里的卷积之后并没有使用非线性激活函数。



2.整体框架

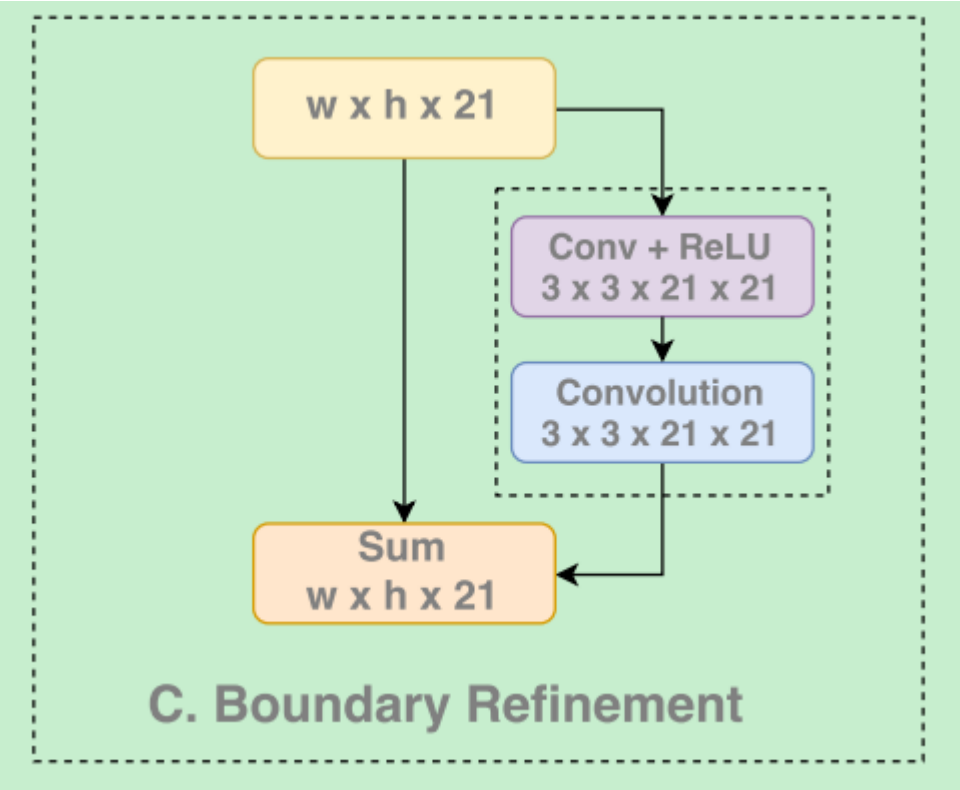
整体框架如下图所示：



从图中可以看到几点：

1. 基础网络使用了ResNet作为特征提取，使用FCN作为语义分割的框架。
2. 使用了ResNet中不同stage的feature map，因此是多尺度架构。
3. GCN模块则用于产生低分辨率的的score map，并上采样与更高分辨率的score map加和产生新的score map。
4. 经过最后的上采样，就输出了预测结果。

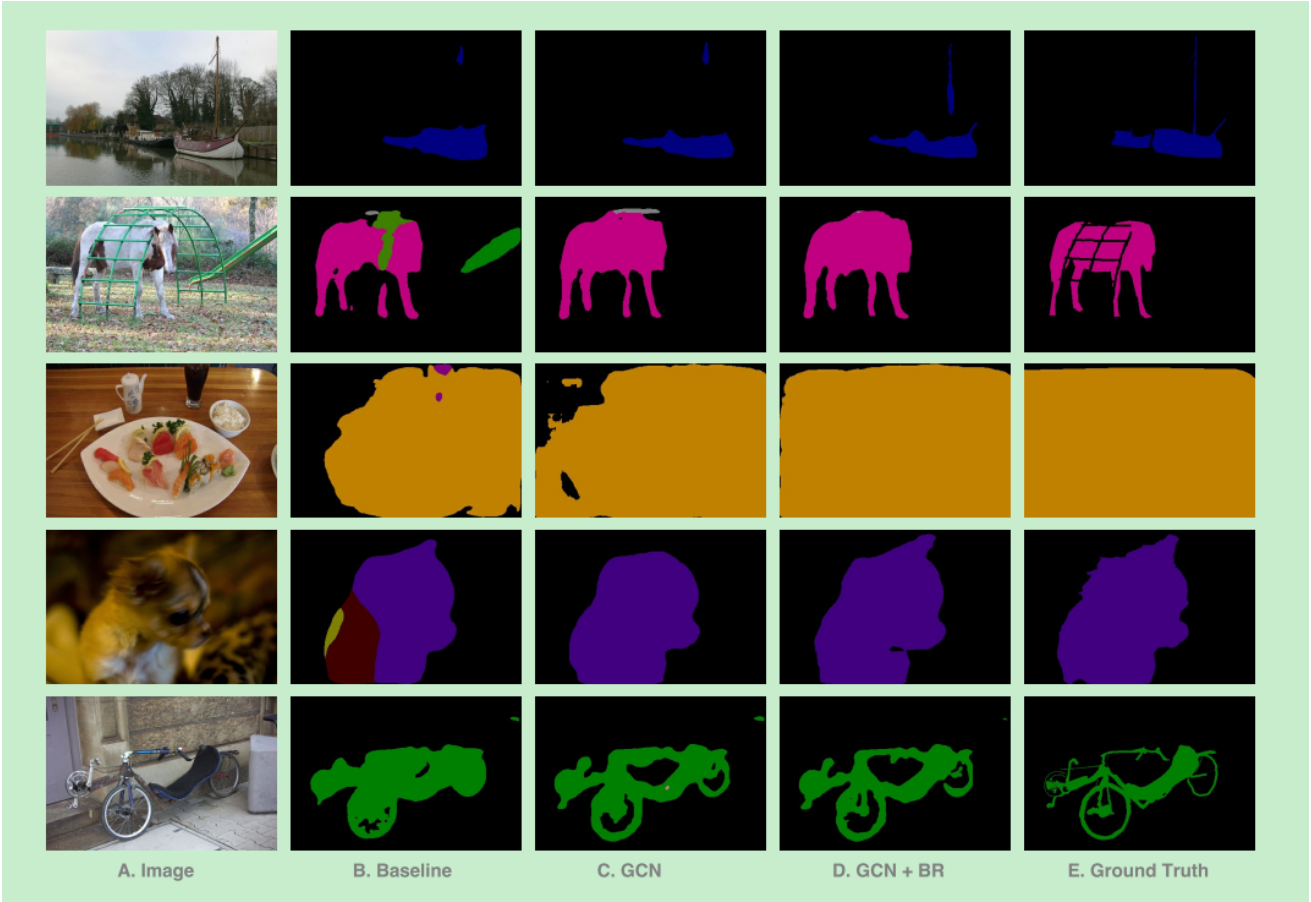
另外，需要注意的是，在整个网络中还加入了很多Boundary Refinement ，这种结构如下图：



可见这是一种残差连接结构。顶部的wxhx21是一种粗粒度的score map，侧边的残差连接可以对boundary进行refine。两者加和就可以达到Boundary Refinement。

实验

作者进行了很多实验，这里只看PASCAL VOC 2012的一些可视化结果：


A. Image
B. Baseline
C. GCN
D. GCN + BR
E. Ground Truth