关于upsampling的相关知识在cs231n的lecture11

Fully Convolutional Networks for Semantic Segmentation

Abstract

Introduction

Related work

Fully Convolutional Networks

Segmentation Architecture

Results

Conclusion

Abstract

卷积网络是强大的视觉模型，可以产生特征层次结构。本文展示的网络是端到端。像素到像素的，表现优异。本文构建了全卷积网络，该网络采用任意大小的输入，并通过有效的推理和学习产生相应大小的输出。本文定义了一种跳跃结构，该架构将来自深层粗糙层的语义信息与来自浅层细节层的外观信息相结合，以产生准确和详细的分割。网络处理时间也很快

动机：

背景：

贡献：本文定义了一种跳跃结构，该架构将来自深层粗糙层的语义信息与来自浅层细节层的外观信息相结合，以产生准确和详细的分割。网络处理时间也很快

卷积网络推动了识别的进步。卷积网络不仅改善了整体图像分类，而且还在结构化输出的局部任务上取得进展。这些包括边界框对象检测，部分和关键点预测。从粗略推理到细微推理的下一步是对每个像素进行预测。先前的方法使用了用于语义分割的卷积，但是具有缺点。

全卷积网络在语义分割上训练端到端的像素到像素，超过了现有技术。据我们所知，这是首次训练FCN端到端（1）进行像素预测，（2）进行有监督的预训练。

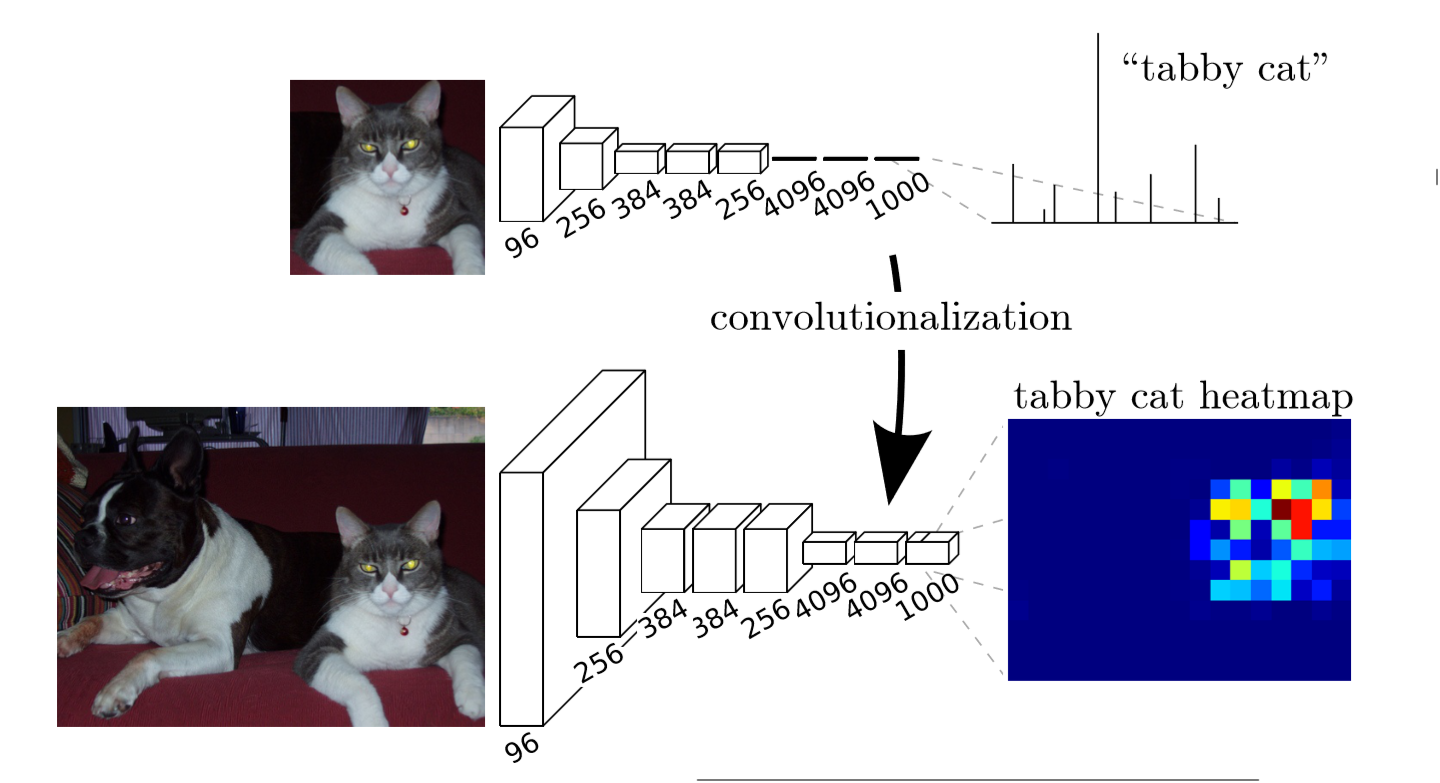
语义分割面临语义和位置之间的内在问题：全局信息解决了是什么，位置信息解决了在哪里

深度特征层次结构将位置和语义信息编码进一个局部到全局的金字塔内。

我们定义了一个跳跃架构（skip architecture）来利用这个特征谱，它结合了深层的，粗糙的语义信息和浅层的，细微的外观信息

为什么CNN需要对输入图像尺寸做要求：

因为CNN最后有全连接层，而全连接层的输入是固定大小的。



3.全卷积网络

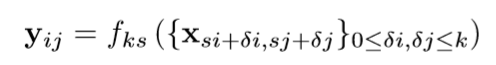
h×w×d,

更高的层中的位置对应的它们通过路径相连的图像中的位置，被称为接受域（receptive fields）.

就是输出featuremap某个节点的响应对应的输入图像的区域就是感受野。

卷积网络建立在平移不变性之上，其基础组成部分（卷积，池化，激活函数）在局部输入区域进行，只依赖于相对空间坐标。

作者提出了一个函数



其中k是核大小（kernel size），s是步长或二次采样参数，fks决定层的类型，是卷积还是池化还是激活函数。

Adapting classifiers for dense prediction

典型的识别网络，包括LeNet，AlexNet，都需要输入固定，并且产生的输出是非空间的。这些网络的全连接层的尺寸固定并且丢弃了空间坐标。全连接层可以被视为另一种卷积层，将全连接层转化为卷积层就可以获得全卷积网络，从而输出heatmap。

3.2 Shift-and-stitch is filter rarefaction

通过将输入的移位版本的输出拼接在一起，可以从粗略输出中获得密集预测。

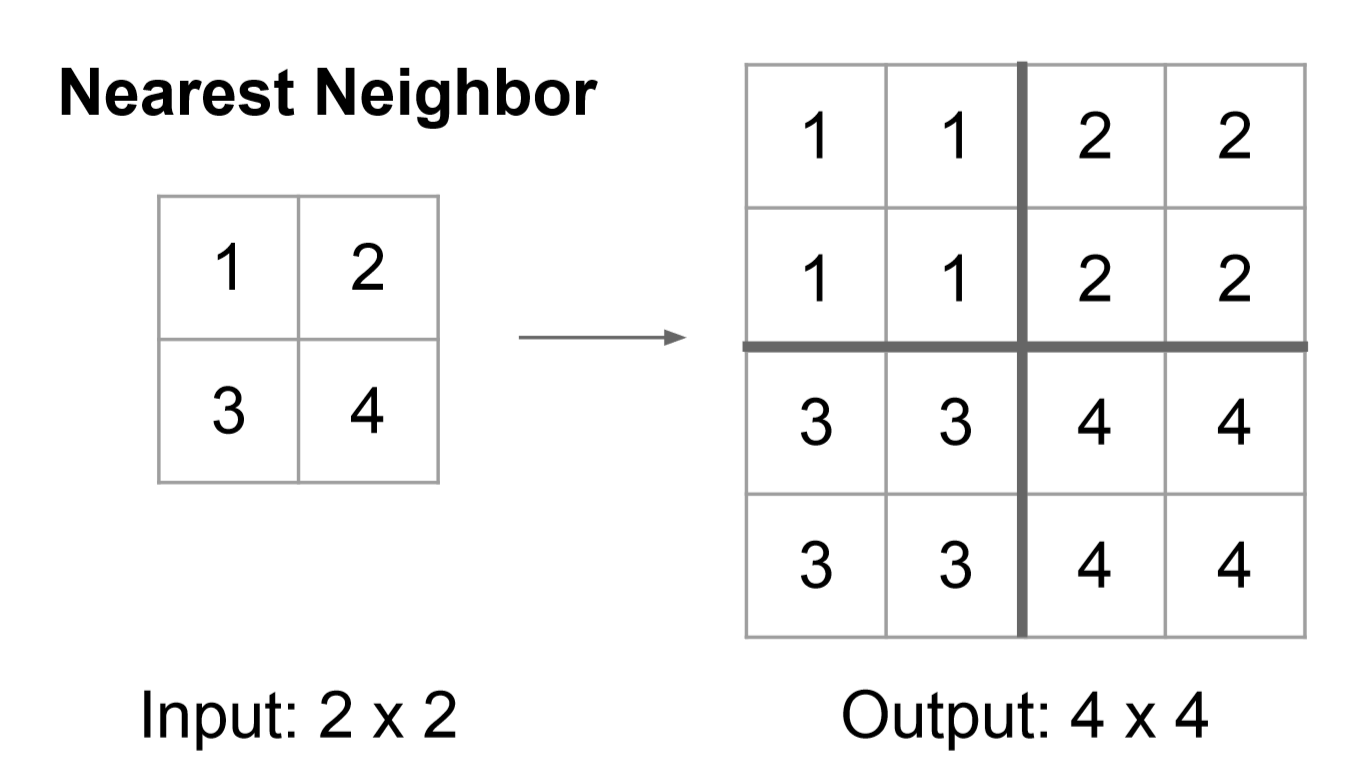
3.3 Upsampling is backwards strided convolution

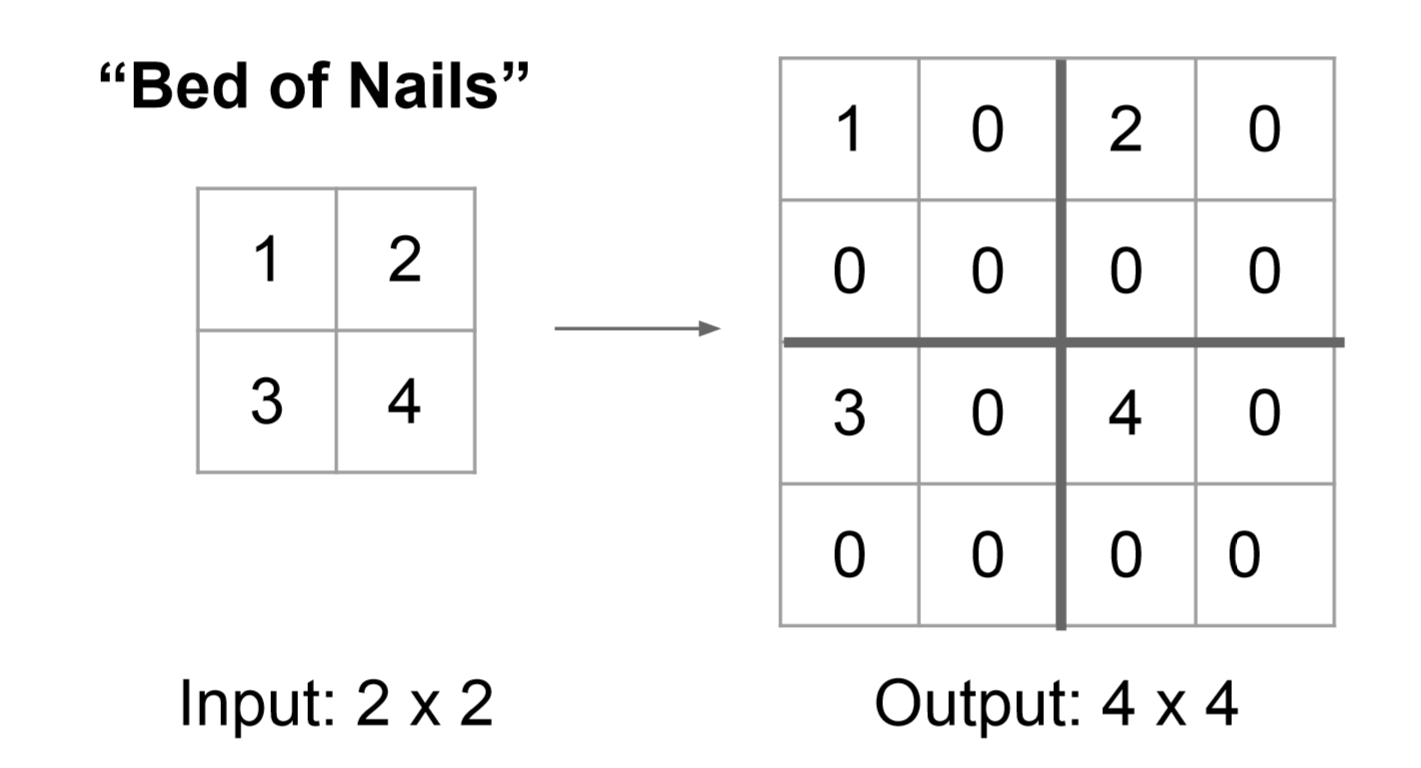
3.4 Patchwise training is loss sampling

Backwards convolution (sometimes called deconvolution)

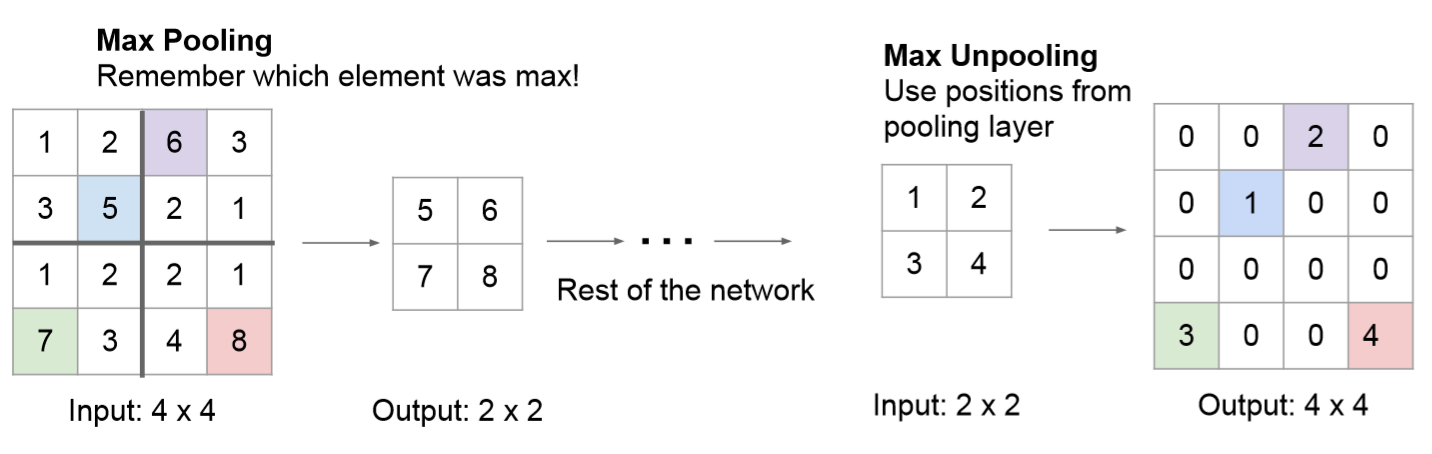
一些upsampling的方法

Unpooling 去池化

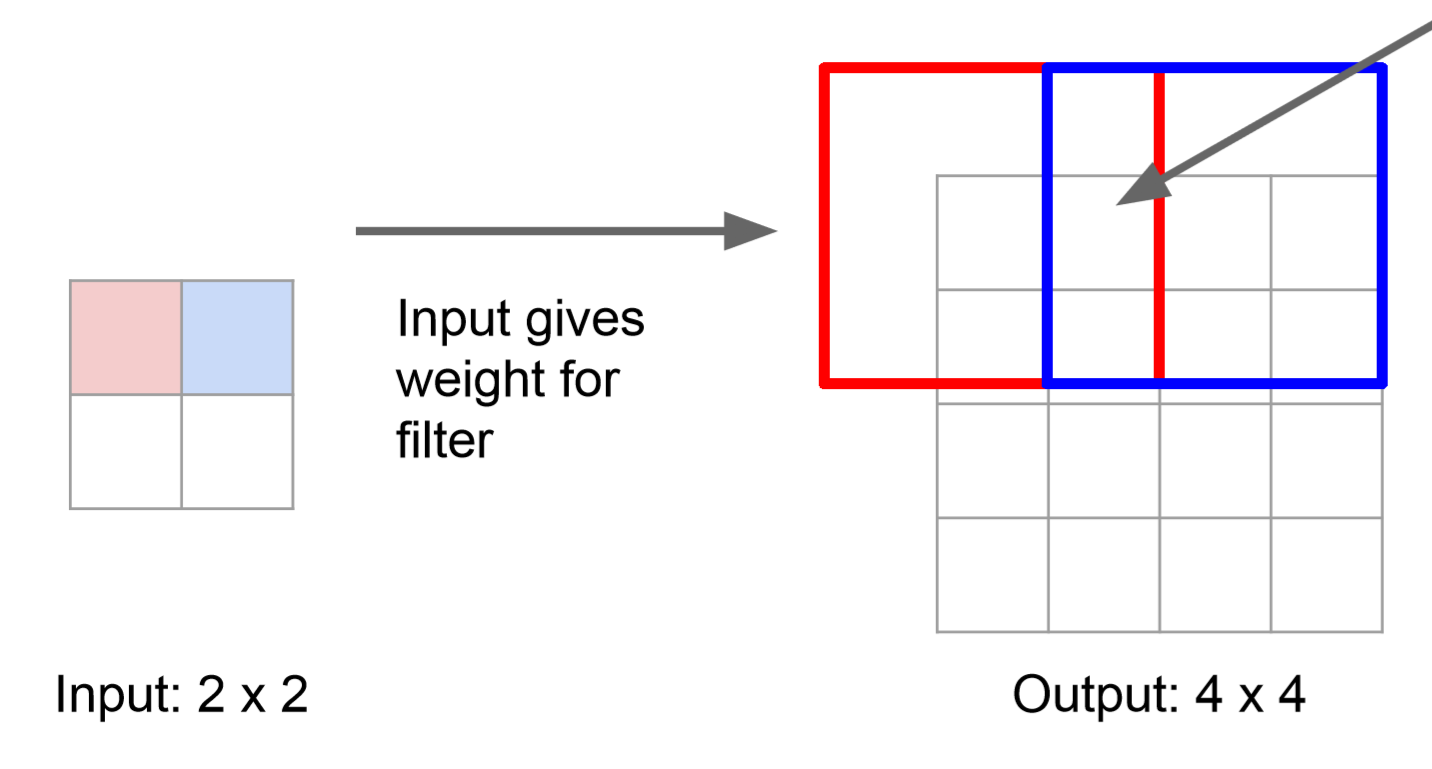




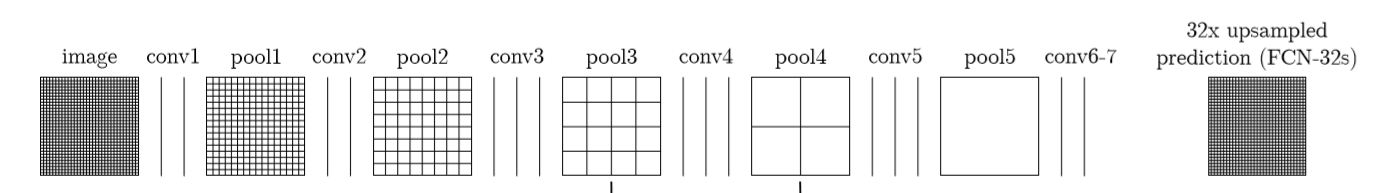
Max unpooling

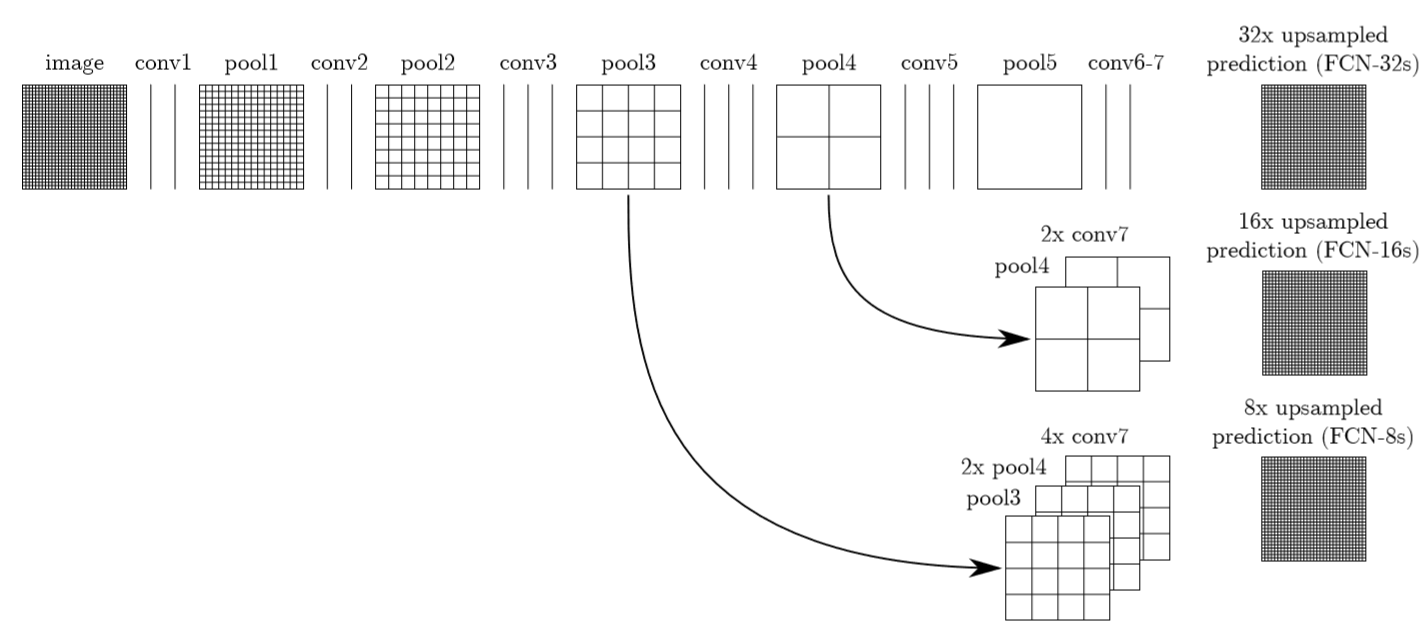


Transpose convolution



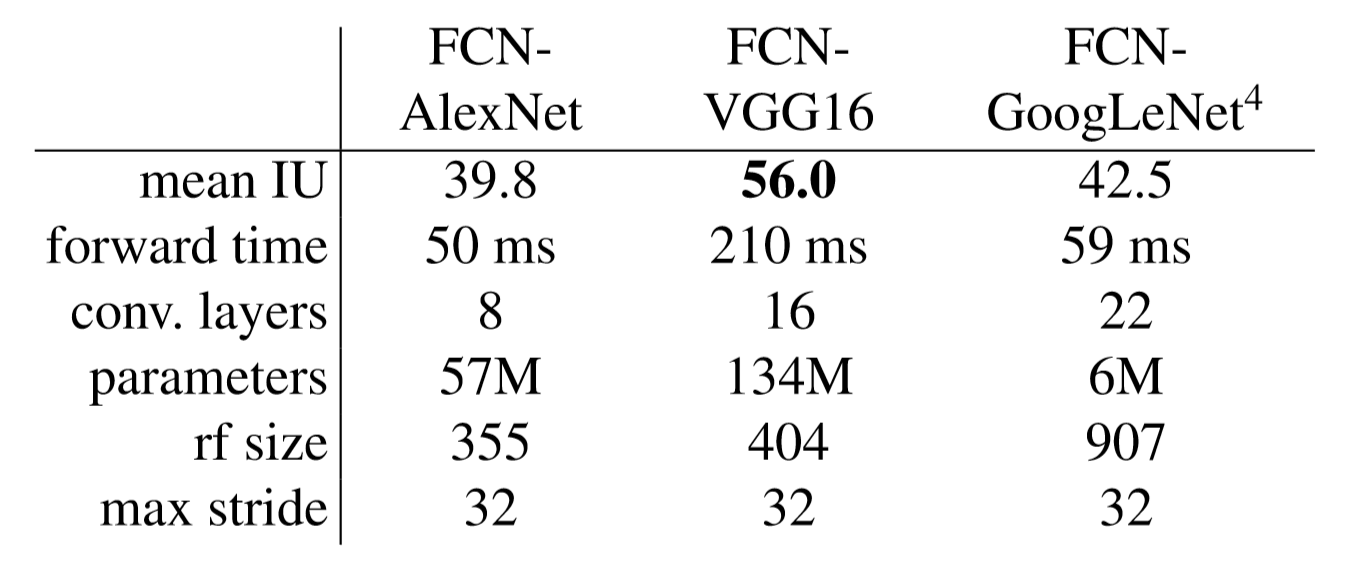
4. Segmentation Architecture

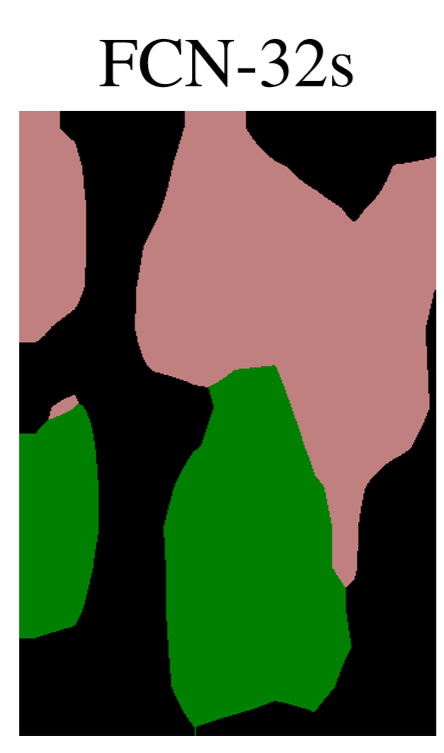


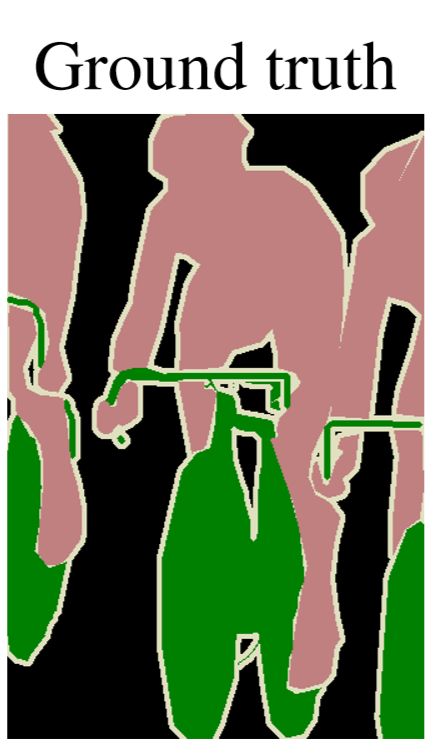


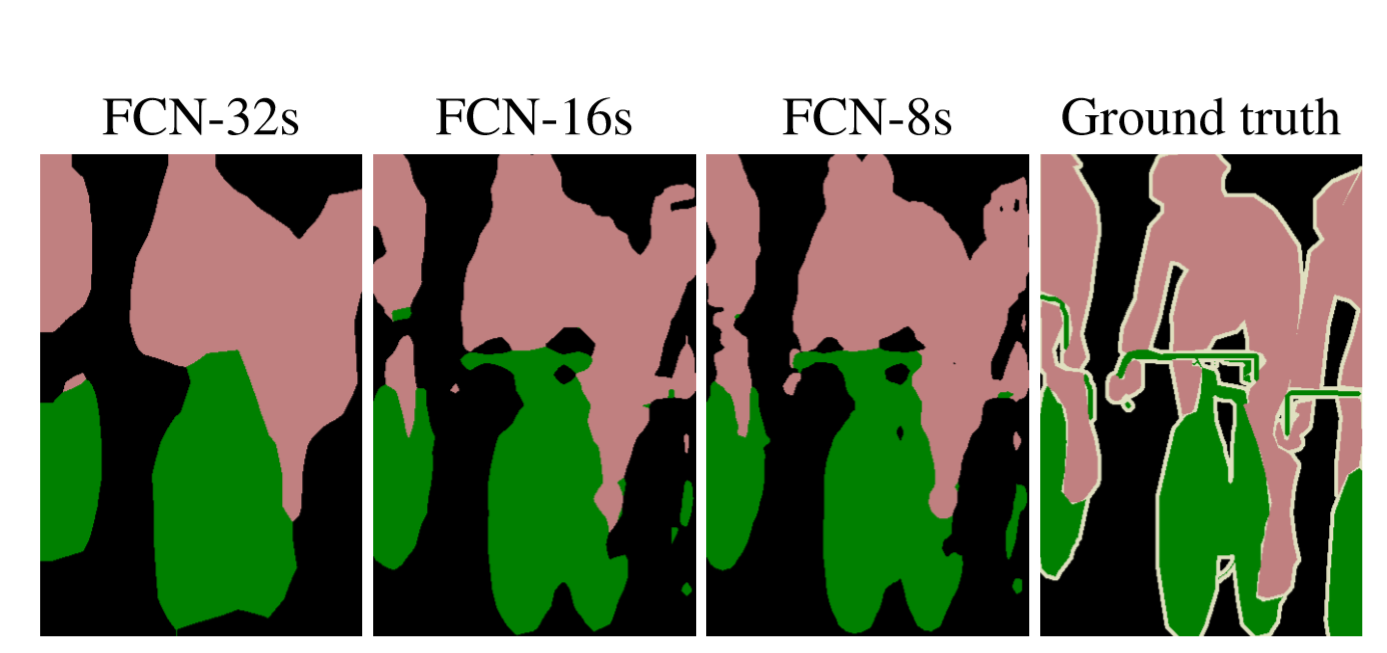
作者对AlexNet，VGG，GoogLeNet进行微调成全卷积网络，在粗糙输出处添加1×1的卷积，，有21个通道来对应PASCAL的种类，之后加上deconvolutional层，把粗糙输出上采样为像素密集输出。

结果表明VGG的效果最好。



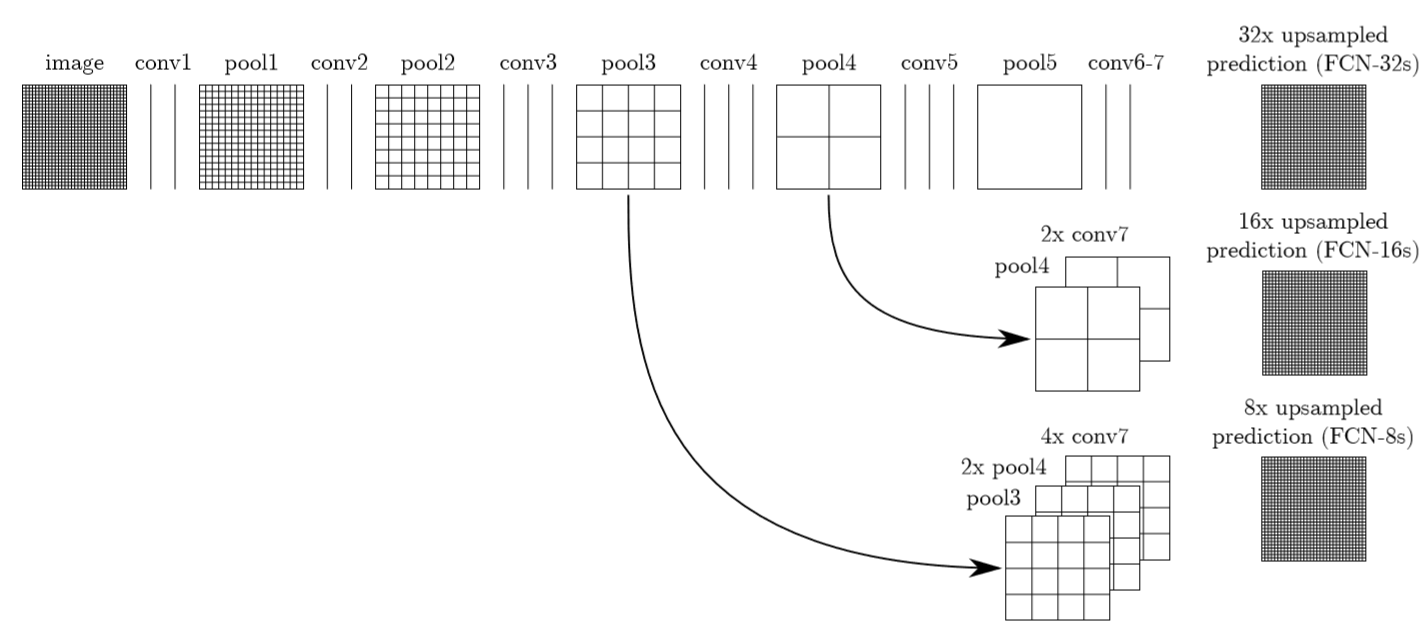






单看FCN-32s的效果过于粗糙，最终预测层的32像素步长的反卷积导致上采样输出中的细节尺度不高。

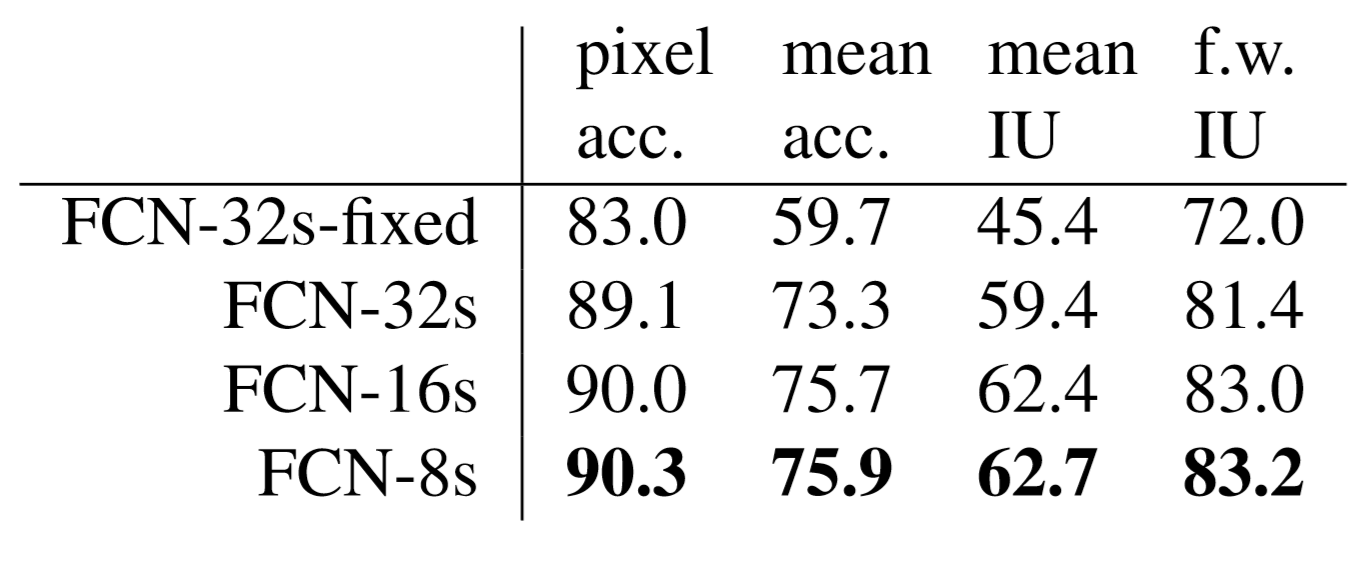
解决方法：添加跳跃（skip）结构



* 对原图进行卷积conv1、pool1后图像缩小为1/2
* 对图像进行第二次卷积conv2、pool2后图像缩小为1/4；
* 对图像进行第三次卷积conv3、pool3后图像缩小为1/8，此时保留pool3的featuremap；
* 对图像进行第四次卷积conv4、pool4后图像缩小为1/16，此时保留pool4的featuremap；
* 对图像进行第五次卷积conv5、pool5后图像缩小为1/32，然后把原来CNN操作过程中的全连接变成卷积操作的conv6、conv7，图像的featuremap的大小依然为原图的1/32
* 对于FCN-32s，网络里面有5个pool，所以conv7的特征图是原始图像1/32，然后再经过32x upsampled prediction 图片变回原输入图大小
* 对于FCN-16s，在pool4（原图大小的1/16）层上添加1×1卷积来产生额外的类预测，，，，在conv7上进行2x upsampling操作，之后将二者输出融合在一起，将融合结果进行16x upsampled prediction。

对于FCN-8s，对conv7进行4x upsample， pool4进行2x upsample，最后把4x conv7，2x pool4与pool3进行fuse，得到求和后的特征图，再进行8x upsampled prediction。

测试结果是FCN-8s略优于FCN-16s，FCN-32s，但只是有了一点点的提升，所以之后就不再进行更浅层的融合了



4.3 Experimental framework

完全卷积网络是一类丰富的模型，其中现代分类网络是一种特殊情况。 认识到这一点，将这些分类网络扩展到分割，并通过多分辨率层组合改进架构，大大提高了现有技术水平，同时简化并加速了学习和推理。