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

卷积网络推动了识别的进步。Convnet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：全局信息解决了本地信息在何处解决的问题。

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。

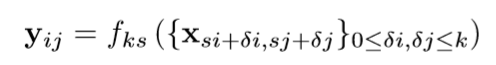
3.全卷积网络

h×w×d,

更高的层中的位置对应于它们与路径相连的图像中的位置，被称为接受域（receptive fields）.

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。

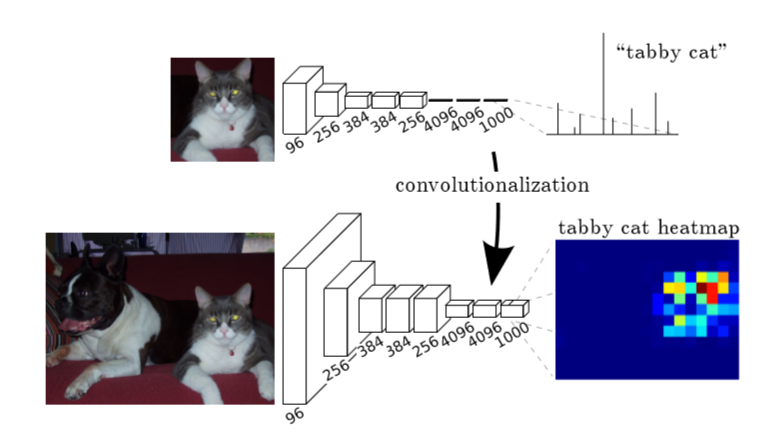
作者提出了一个函数



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。



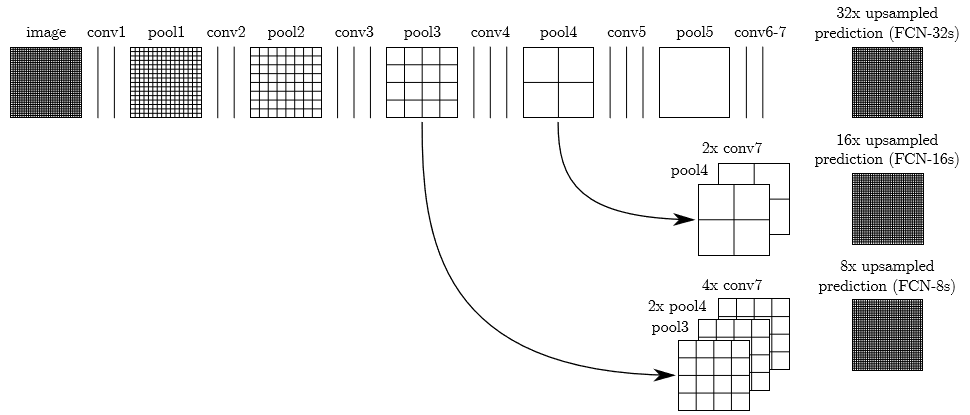
3.2 Shift-and-stitch is filter rarefaction

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。

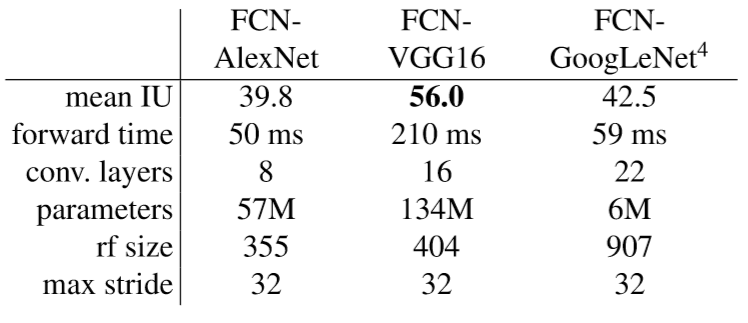
3.3 Upsampling is backwards strided convolution

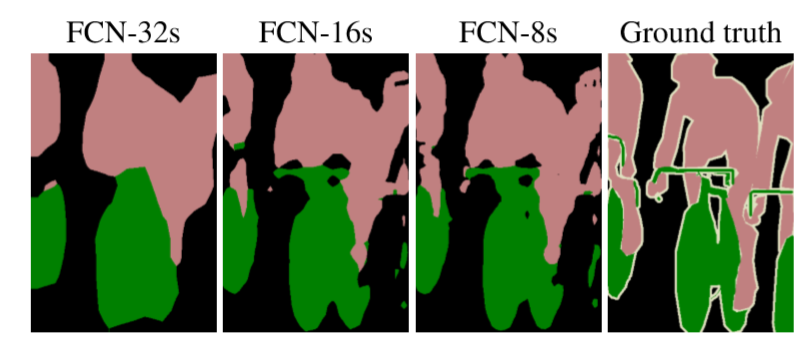
3.4 Patchwise training is loss sampling

4. Segmentation Architecture



作者对AlexNet，VGG，GoogLeNet进行微调成全卷积网络，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。

* 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，在FCN中的卷积是不会改变图像大小（或者只有少量像素的减少，特征图大小基本不会小很多），然后再经过32x upsampled prediction 图片变回32x32。
* 对于FCN-16s，作者在conv7先进行一个2x conv7操作，这里是增加1个卷积层，卷积后特征图的大小为conv7的**2**倍，此时2x conv7与pool4的大小一样，之后对pool4与2x conv7进行一个fuse操作，将fuse结果进行16x upsampled prediction，与FCN-32s一样，也是增加一个卷积层，卷积后的大小为输入图像的**16**(2^4)倍，我们知道pool4的大小是原图的1/16，放大16倍，就是32x32，这样最后图像大小也变为原来的大小，至此完成了一个16s的upsample。
* 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，在FCN中的卷积是不会改变图像大小（或者只有少量像素的减少，特征图大小基本不会小很多），然后再经过32x upsampled prediction 图片变回32x32。
* 对于FCN-16s，作者在conv7先进行一个2x conv7操作，这里是增加1个卷积层，卷积后特征图的大小为conv7的**2**倍，此时2x conv7与pool4的大小一样，之后对pool4与2x conv7进行一个fuse操作，将fuse结果进行16x upsampled prediction，与FCN-32s一样，也是增加一个卷积层，卷积后的大小为输入图像的**16**(2^4)倍，我们知道pool4的大小是原图的1/16，放大16倍，就是32x32，这样最后图像大小也变为原来的大小，至此完成了一个16s的upsample。

在进行FCN-8s的测试后，发现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。