深度学习在图像语义分割领域的应用

Abstract

1.Introduction

2.背景与基本理论

1.迁移学习，迁移学习的用处和必要性，使用条件，

2.数据预处理与数据扩充，加速收敛，正则化

3.卷积神经网络

1.经典分类网络，AlexNet，VGG，讲这两个，googlenet顺便提一下

2.优化网络，ResNet，DenseNet

3.基本的分割网络，FCN（最经典），U-Net（写一下医疗相关的，借鉴那片论文），SegNet,讲一下segnet的一些变种

4.分割网络的解码器变种

4.优化方法

1.条件随机场CRF

2.多尺度预测

5.讨论，讲一下评估方法，参考segnet和那篇综述的论文内容

写一下比较结果，包括segnet和综述的内容。本文的数据来自所包含的网络的论文。

6.结论，分割，还有未来展望。

Abstract

近几年人工智能吸引了越来越多人的注意，部分得益于人工智能的重要组成部分深度学习的快速发展，也因此得到很大的发展。目标识别和语义分割一直是计算机视觉领域中传统场景理解的两个最常见问题。由于深度学习的应用，尤其是卷积神经网络在图像分类方面的惊人的效果，越来越多的计算机视觉研究员将目光放在了用深度学习处理语义分割上，目前已经取得了显著的效果。本文总结了深度学习的一些背景和基础理论与优化方法，并详细介绍了近几年神经网络的发展和一些经典而常用的卷积神经网络，最后对目前的几种网络进行评估分析与一点未来研究方向的思考。

Introduction

神经网络这个名词并不新鲜，在上个世纪便已经被提出，但是发展神经网络需要的两个重要条件，海量的数据和高速的运算机器在当时无法得到满足，所以这项技术便被搁置。经过神经网络研究科学家对其进行算法上的优化，以及目前数据量的激增和计算硬件的高速发展，神经网络尤其是卷积神经网络引起了越来越多研究人员的兴趣。

视觉信息一直是人类接收的所有信息中的主要组成部分，人类也因此在视觉分析与处理上进化到很高的水准。人类可以在很短时间内学会识别物体，并且可以在不知道物体是属于哪类的情况下进行分割，比如医学图像方面，可能会存在未知的目标，但是它们仍然可以在图像内被分割，以进一步研究。

在计算机视觉的漫长历史中，图像语义分割一直是一个重要的研究问题。目前许多应用正在兴起，都迫切需要准确而有效的分割机制，如自动驾驶，室内导航，虚拟现实或增强现实系统等等。在之前的研究中，目标检测和语义分割领域已经又人员提出了一些算法，部分算法很有效。但是在深度学习或者说卷积神经网络再一次被应用到目标检测上后，取得了十分惊艳的结果。AlexNet在2012年的ImageNet大型视觉识别比赛（ILSVRC）中取得top5错误率15.3%的惊人成绩，领先第二名超过10个百分点。至此引发了深度学习与神经网络的再一次研究热潮。随着分类卷积网络的不断优化，人们将目光放到了对每个像素进行预测的语义分割领域。研究人员提出了先进而巧妙的网络结构，在计算机视觉语义分割领域取得了十分优异的成绩。

在本文的下一部分，我们将讨论一些深度学习的背景以及详细介绍常用的神经网络

2.基础理论与训练技术

3.卷积神经网络

在近期的深度学习研究中，科研人员提出了很多的卷积网络架构，用以解决各种各样的问题，都取得了一定的成效。其中，某些深度卷积神经网络已经为这个领域做出了极其重要的贡献，它们已经成为了众所周知的标准，得到后来研究者的广泛使用。它们目前被用作许多分割网络架构的基础和编码器模块。由于它们的重要性，我们将在这一部分对其进行详细介绍。

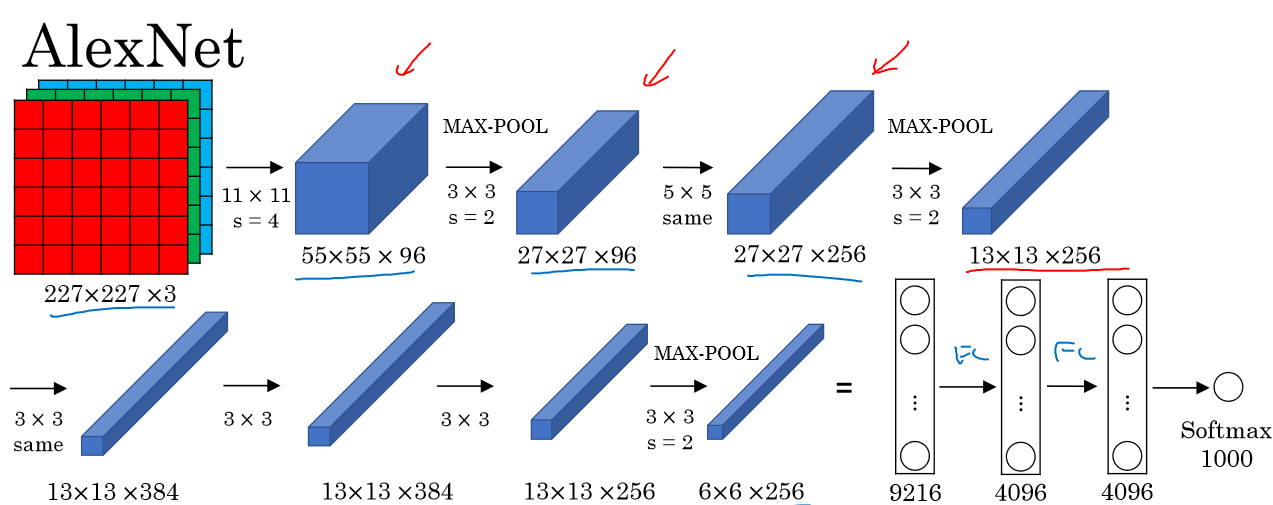
3.1经典分类网络

3.1.1 AlexNet

AlexNet算是将深度学习运用到大规模图像处理的开山之作，在2010年ImageNet上120万高分辨率图像进行1000种分类，比当时最先进的结果表现都要好。AlexNet还在ILSVRC 2012上取得了top5错误率15.3%的好成绩，比第二名高了10个百分点。

AlexNet创造性的首次使用ReLU（Rectified Linear Units）非线性函数代替tanh和sigmoid作为激活函数，基于梯度下降法的训练时间大大减少。ReLU函数形式为f(x) = max(0,x)，即在x<0的范围内函数为0，在x>0的范围内，函数为x。

网络含有8个权重层，其中五个卷积层和3个全连接层，最后一个全连接层连接到1000类的softmax分类器上。网络最大化多项逻辑回归目标，这相当于最大化预测分布下正确标签的对数概率的训练案例的平均值。在每个卷积层和全连接层之后都会应用ReLU非线性函数。网络的结构示意图如下所示



下面讲一下AlexNet使用的数据扩充方法。第一点是第一种形式的数据增强包括生成图像平移（image translations）和水平反射（horizontal reﬂections）。方法是通过在256×256的图像中随机提取224×224大小的块（以及其水平反射），并在这些提取的块上训练神经网络，将训练集扩大了2048倍。测试的时候，通过提取10个224×224的块并对其进行预测，以及对网络的softmax层对10个块的预测求平均值。第二种方式包括改变训练图像的RGB通道的强度。通过ImageNet训练集在RGB像素值上进行PCA（Principal Component Analysis)），对于每一个训练图片，增加其多倍的主部分。该方案近似符合自然图像的重要特性，即目标身份对于照明的强度和颜色的变化是不变的。

3.1.2 VGGNet

VGGNet 由牛津大学的Visual Geometry Group在论文《Very Deep Convolutional Networks for Large-Scale Image Recognition》中提出，网络的命名也是来自于作者所在的小组名的缩写。VGGNet在ILSVRC2013取得了top5高达92.7%的正确率，对日后无数的网络产生了深远的影响，很多的后续网络结构都是基于VGG进行微调。该网络的主要研究贡献是关于卷积网络的深度在大规模图像识别设置中的准确性的影响，这一点为后面关于网络性能的研究做出了尝试。网络创造性地使用了非常小（3×3）卷积filter，此网络深度较大，为16层到19层，并表现出显著的性能提升。 VGG16是指网络中有16个权重层，包括13个卷积层和3个全连接层。

网络创造性地使用了非常小（3×3）卷积filter，根据理论计算，2个3×3卷积层连在一起相当于一个5×5的卷积层，三个连在一起相当于单个7×7层。使用小的卷积核的优点是合并了三个非线性层而不是一个非线性形层，可以优化决策函数，其次可以减少参数数量。至此，使用小的卷积核的方法被后来的研究者广泛使用。网络中还使用1×1大小的卷积，可以通过修正函数增加额外的非线性。

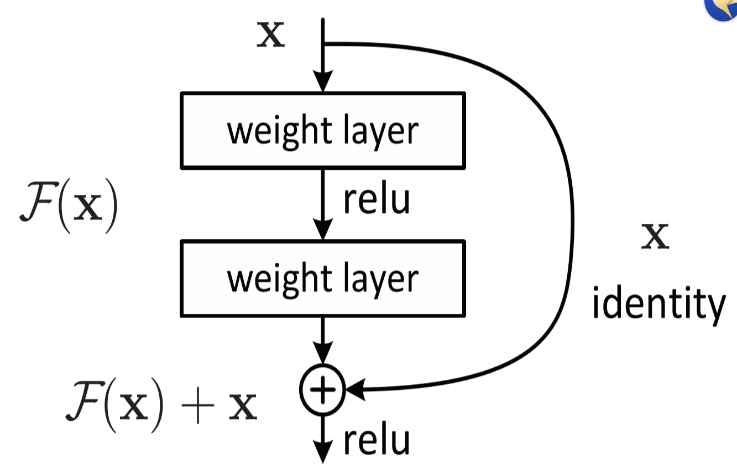
AlexNet和VGGNet作为图像分类的经典网络，对后面网络的发展产生了深远的影响。后续的一些优化训练的网络乃至进行语义分割的网络都有着它们的影子，下面我们介绍一下对神经网络的结构进行精妙的设计从而大大改善了训练的一些网络。

3.2训练优化网络

这一部分的主要内容是介绍对网络的结构进行创造性的修改，从而大大地优化了训练过程和预测结果的网络。

3.2.1ResNet

深度卷积神经网络在图像分类上效果显著。在VGG之后，研究人员发现了卷积网络深度对其性能的影响，，网络可以集成从低级到高级的特征以及分类器，并且网络越深。特征级别越高。随着网络深度的增加，会出现梯度消失和梯度爆炸的问题，这个问题通过归一初始化部分解决。但是网络继续加深会出现退化问题（degradation），深度越深，越难以训练。退化问题是指，理论上当网络深度增加后，训练好的网络至少不会比更浅的网络表现得差，但是事实发现更深的网络会表现得反而不如浅层的网络，这就是退化问题。残差网络ResNet的提出了新的方法，将一个层的输入跳过一个或多个层的连接，到达后面层的输出中。此方法不增加额外参数也不增加计算复杂度。实验表明，深度残差网络很容易优化，并且可以轻松的从网络深度的增加中获得准确度增益。

图2是ResNet中的残差学习块，也是网络最具有创新性的部分，下面具体介绍残差块能够有效的原因。

原来的函数是H(x)，现在将其改为F(x)+x，这两种表达的效果相同，但是优化的难度却并不相同。假设F(x)+x的优化会比H(x)容易，网络深度增加，求解器难以通过多个非线性层接近恒等映射，但利用残差网络则很容易。如果恒等映射最优，只需将权重趋向0即可。如果恒等映射不是最优，只要最优函数本身更接近恒等映射而不是零映射，则网络参考恒等映射来学习要比把该函数当成新的映射来学习要容易。

网络还提出了Bottleneck结构，使用三层卷积（1×1,3×3,和1×1）替换之前的二层卷积（3×3和3×3），1×1卷积用来减少维度，从而减少训练时间。

ResNet的网络结构可以称之为天才的设计，简单而又易于实施，效果又十分惊艳，其最主要的贡献是大大优化了深度很大时的神经网络的训练，仅在论文中作者就提出了50，101，152层的网络，这样的深度如果没有ResNet加持是无法实现训练的。在后续的研究来看，很多网络结构都采用了ResNet的结构进行设计。

3.2.2

和基本理论，训练神经网络的一些技术方法。