

基于深度学习的图像分割：网络结构设计

青暮 CV技术指南 6天前

收录于话题
#CV技术总结

73个

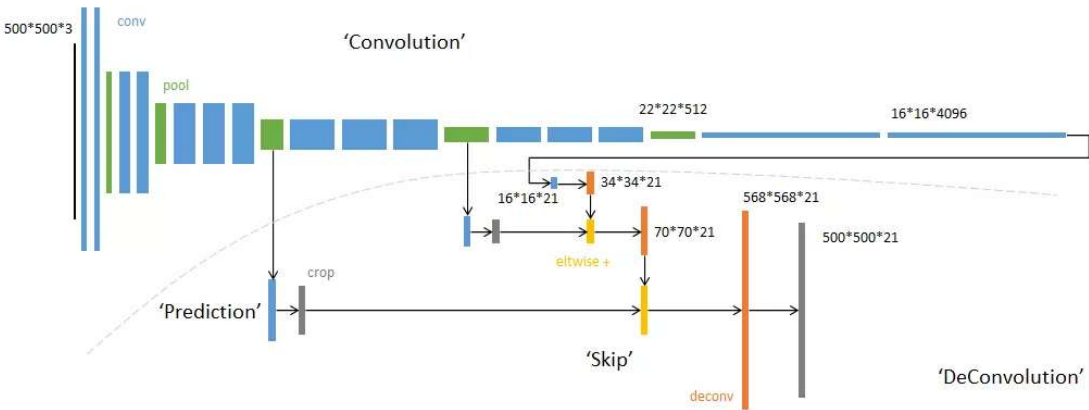
点击上方“CV技术指南”，关注我们

前言 文章总结了利用CNNs进行图像语义分割时，针对网络结构的创新，这些创新点主要包括新神经架构的设计（不同深度、宽度、连接和拓扑结构）和新组件或层的设计。前者是利用已有的组件组装复杂的大型网络，后者是更偏向于设计底层组件。

本文转载自新机器视觉

1. 图像语义分割网络结构创新

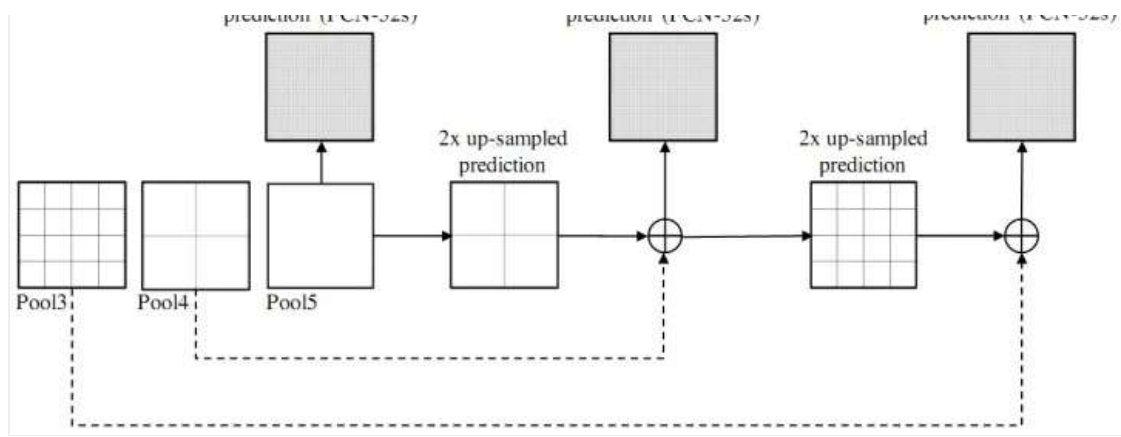
1.1 FCN网络



FCN整体架构简图

单独将FCN网络列出来是因为FCN网络是第一个从全新的角度来解决语义分割问题的网络。此前的基于神经网络的图像语义分割网络是利用以待分类像素点为中心的图像块来预测中心像素的标签，一般用CNN+FC的策略构建网络，显然这种方式无法利用图像的全局上下文信息，而且逐像素推理速度很低；而FCN网络舍弃全连接层FC，全部用卷积层构建网络，通过转置卷积以及不同层特征融合的策略，使得网络输出直接是输入图像的预测mask，效率和精度得到大幅度提升。

32x up-sampled prediction (FCN-32s) 16x up-sampled prediction (FCN-32s) 8x up-sampled prediction (FCN-32s)



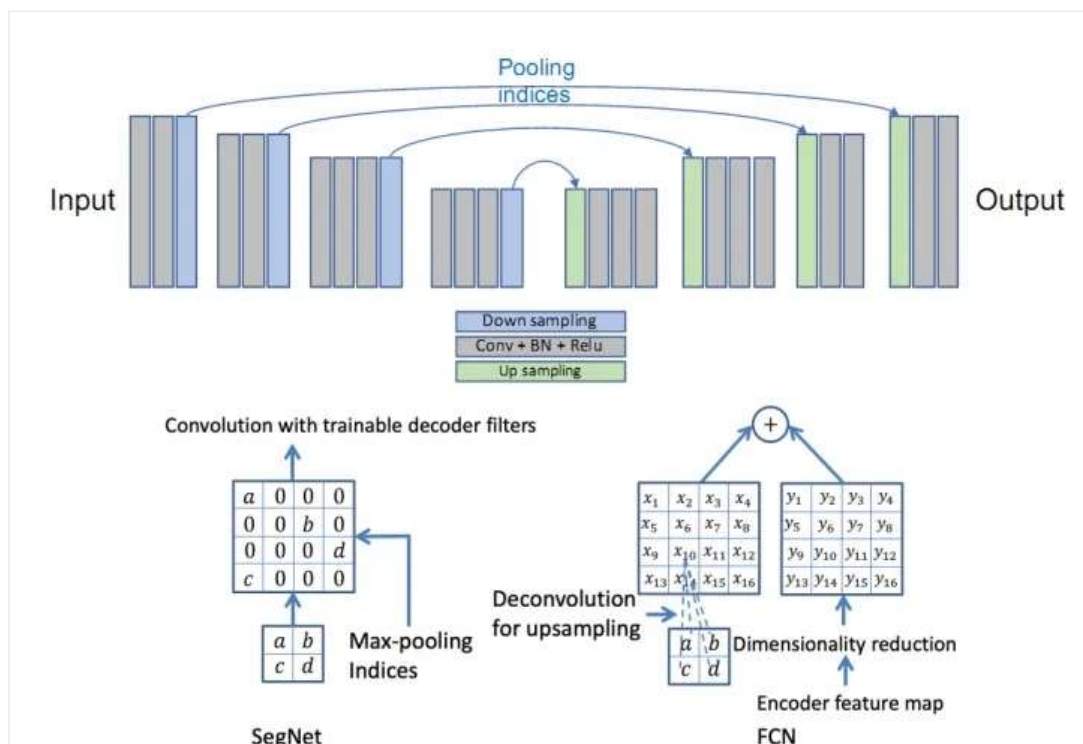
FCN不同层特征融合示意图

创新点：全卷积网络（不含fc层）；转置卷积deconv（反卷积）；不同层特征图跳跃连接（相加）

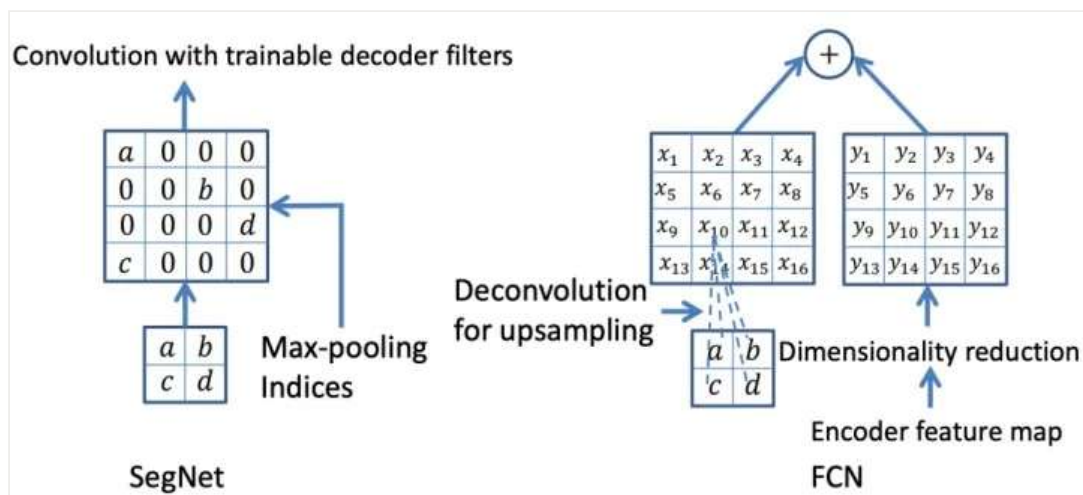
1.2 编解码结构（Encoder-decoder）

- **SegNet**和FCN网络的思路基本一致。编码器部分使用VGG16的前13层卷积，不同点在于Decoder部分Upsampling的方式。FCN通过将特征图deconv得到的结果与编码器对应大小的特征图相加得到上采样结果；而SegNet用Encoder部分maxpool的索引进行Decoder部分的上采样（原文描述：the decoder upsamples the lower resolution input feature maps. Specifically, the decoder uses pooling indices computed in the max-pooling step of the corresponding encoder to perform non-linear upsampling.）。

创新点：Encoder-Decoder结构；Pooling indices。



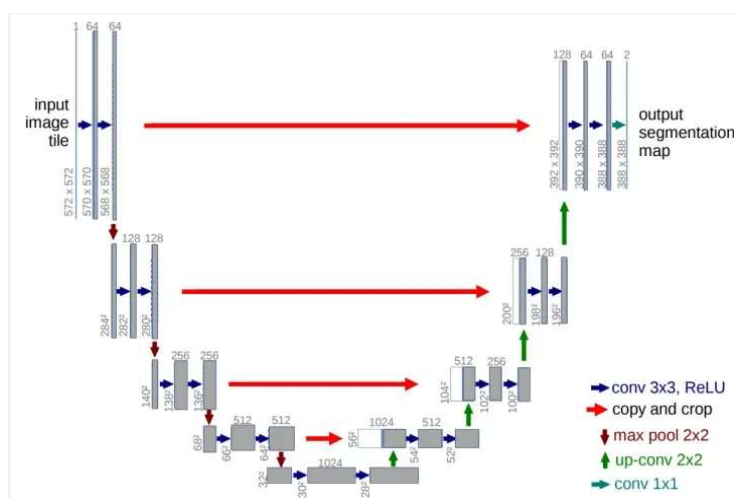
SegNet网络



SegNet与FCN的Upsample方式对比

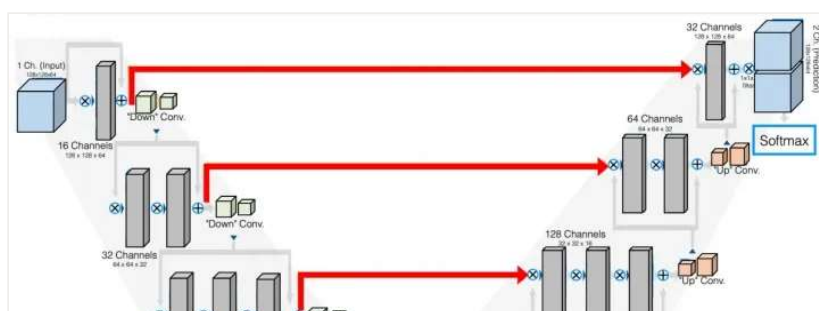
- U-Net网络最初是针对生物医学图像设计的，但由于其出色的性能，现如今UNet及其变体已经广泛应用到CV各个子领域。UNet网络由U通道和短接通道（skip-connection）组成，U通道类似于SegNet的编解码结构，其中编码部分（contracting path）进行特征提取和捕获上下文信息，解码部分（expanding path）用解码特征图来预测像素标签。短接通道提高了模型精度并解决了梯度消失问题，特别要注意的是短接通道特征图与上采样特征图是拼接而不是相加（不同于FCN）。

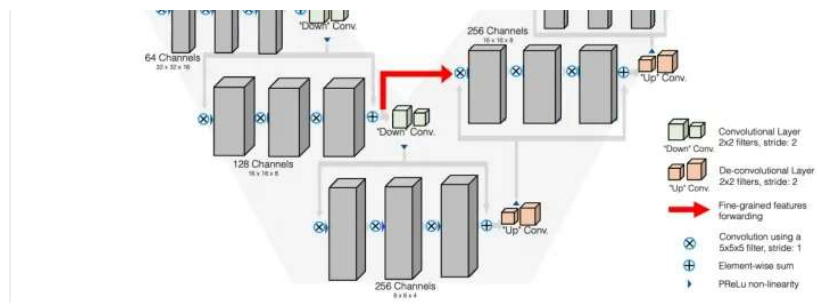
创新点： U型结构；短接通道（skip-connection）



U-Net网络

- V-Net网络结构与U-Net类似，不同在于该架构增加了跳跃连接，并用3D操作物替换了2D操作以处理3D图像（volumetric image）。并且针对广泛使用的细分指标（如Dice）进行优化。

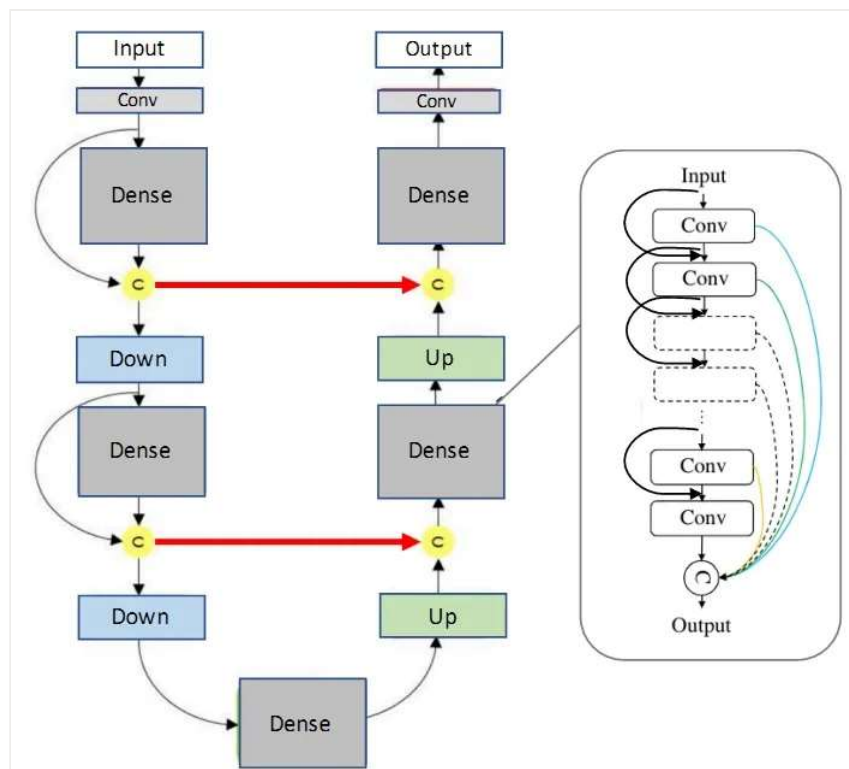




V-Net网络

创新点：相当于U-Net网络的3D版本

- FC-DenseNet（百层提拉米苏网络）**（paper title: The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation）该网络结构是由用密集连接块（Dense Block）和UNet架构组建的。该网络最简单的版本是由向下过渡的两个下采样路径和向上过渡的两个上采样路径组成。且同样包含两个水平跳跃连接，将来自下采样路径的特征图与上采样路径中的相应特征图拼接在一起。上采样路径和下采样路径中的连接模式不完全同：下采样路径中，每个密集块外有一条跳跃拼接通路，从而导致特征图数量的线性增长，而在上采样路径中没有此操作。（多说一句，这个网络的简称可以是Dense Unet，但是有一篇论文叫Fully Dense UNet for 2D Sparse Photoacoustic Tomography Artifact Removal，是一个光声成像去伪影的论文，我看到过好多博客引用这篇论文里面的插图来谈语义分割，根本就不是一码事好么 =_=||，自己能分清即可。）

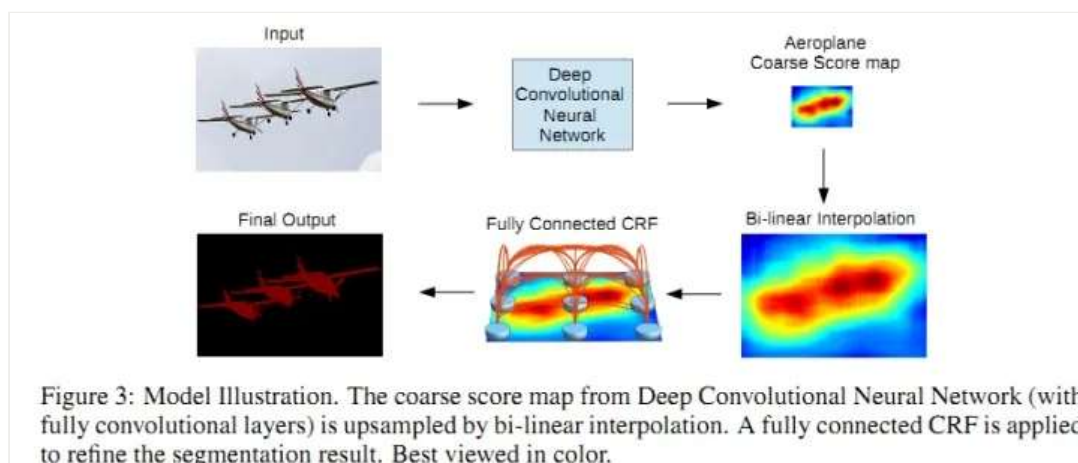


FC-DenseNet（百层提拉米苏网络）

创新点：融合DenseNet与U-Net网络（从信息交流的角度看，密集连接确实要比残差结构更强大）

- Deeplab系列网络是在编解码结构的基础上提出的改进版本，2018年DeeplabV3+网络在VOC2012和Cityscapes数据集上的表现优异，达到SOTA水平。DeepLab系列共有V1、V2、V3和V3+共四篇论文。简要总结一些各篇论文的核心内容：

1) DeepLabV1：融合卷积神经网络和概率图模型：CNN+CRF，提高了分割定位精度；



2) DeepLabV2：ASPP（扩张空间金字塔池化）；CNN+CRF

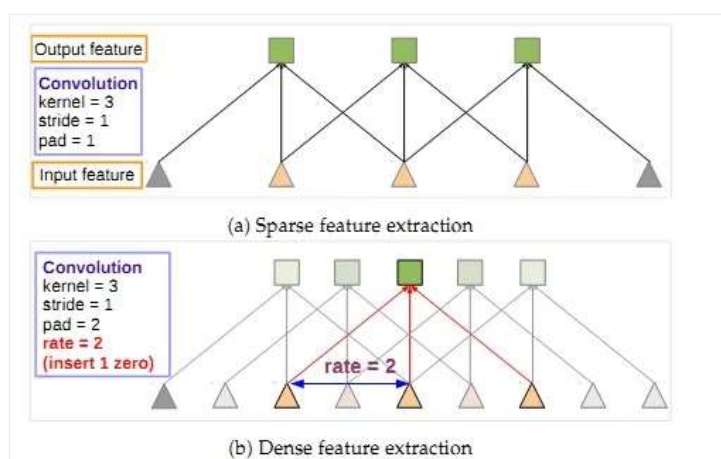


Fig. 2: Illustration of atrous convolution in 1-D. (a) Sparse feature extraction with standard convolution on a low resolution input feature map. (b) Dense feature extraction with atrous convolution with rate $r = 2$, applied on a high resolution input feature map.

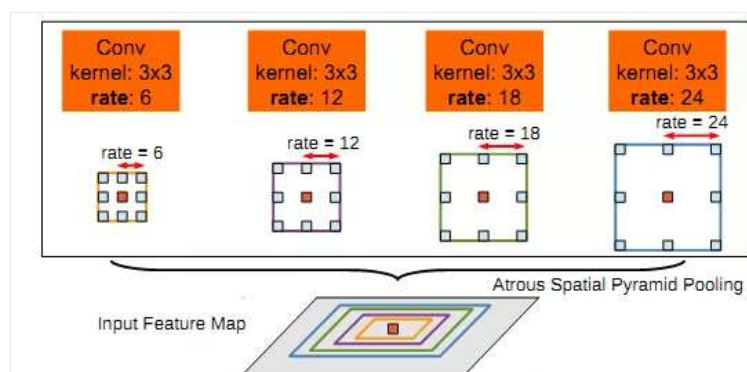
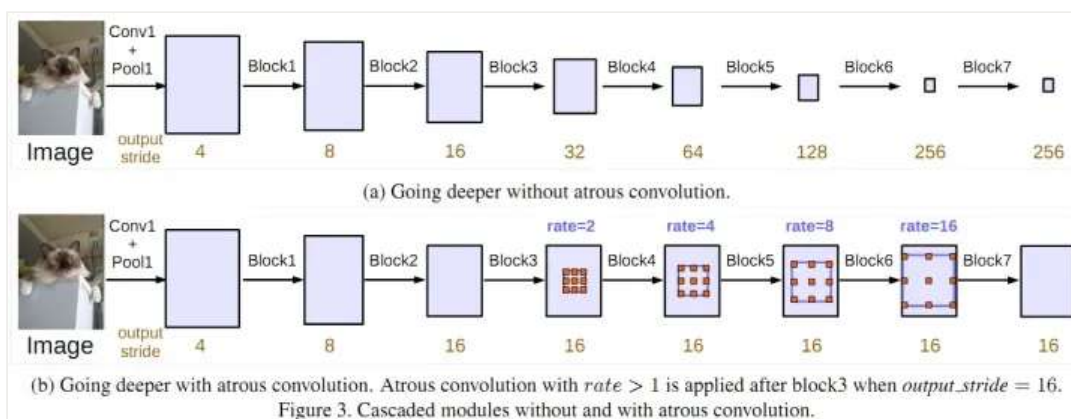
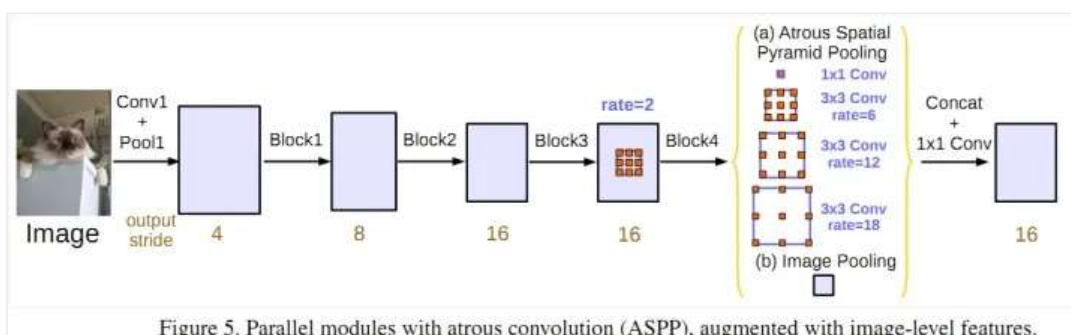


Fig. 4: Atrous Spatial Pyramid Pooling (ASPP). To classify the center pixel (orange), ASPP exploits multi-scale features by employing multiple parallel filters with different rates. The effective Field-Of-Views are shown in different colors.

3) DeepLabV3: 改进ASPP, 多了1*1卷积和全局平均池化 (global avg pool) ; 对比了级联和并联空洞卷积的效果。

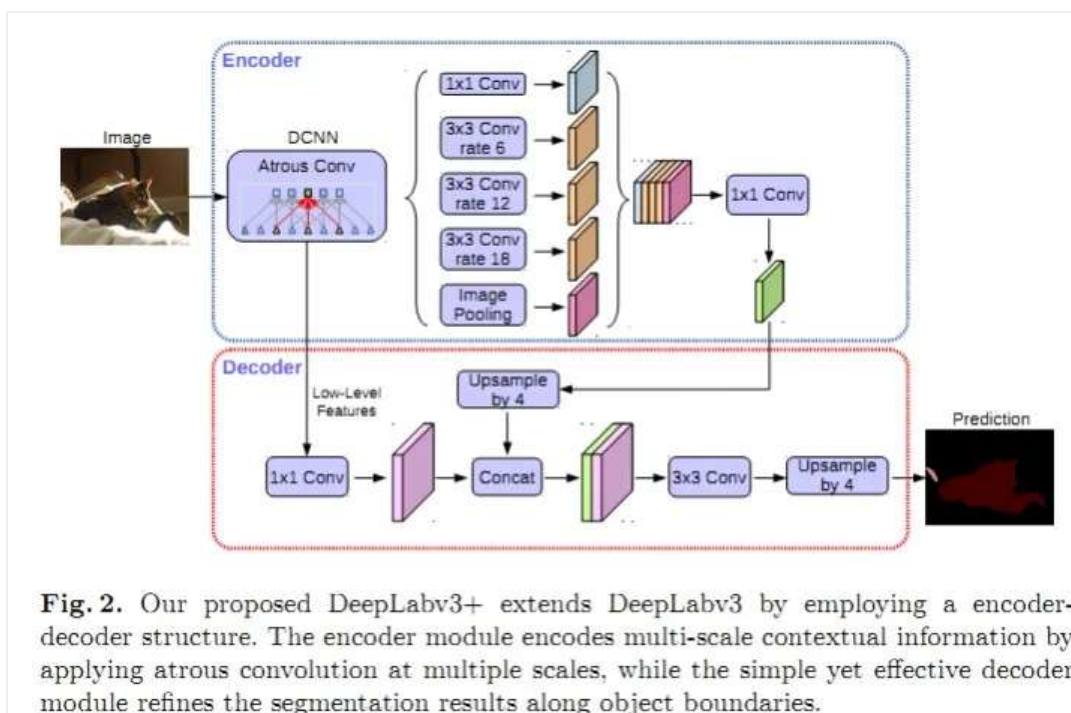


级联空洞卷积



并联空洞卷积 (ASPP)

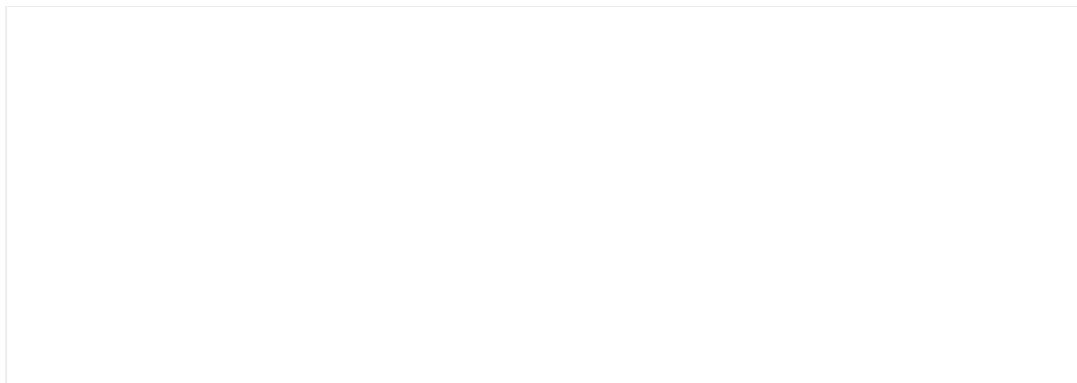
4) DeepLabV3+: 加入编解码架构思想, 添加一个解码器模块来扩展DeepLabv3; 将深度可分离卷积应用于ASPP和解码器模块; 将改进的Xception作为Backbone。



DeepLabV3+

总的来说，**DeepLab系列的核心贡献**: 空洞卷积；ASPP；CNN+CRF（仅V1和V2使用CRF，应该是V3和V3+通过深度网络解决了分割边界模糊的问题，效果要比加了CRF更好）

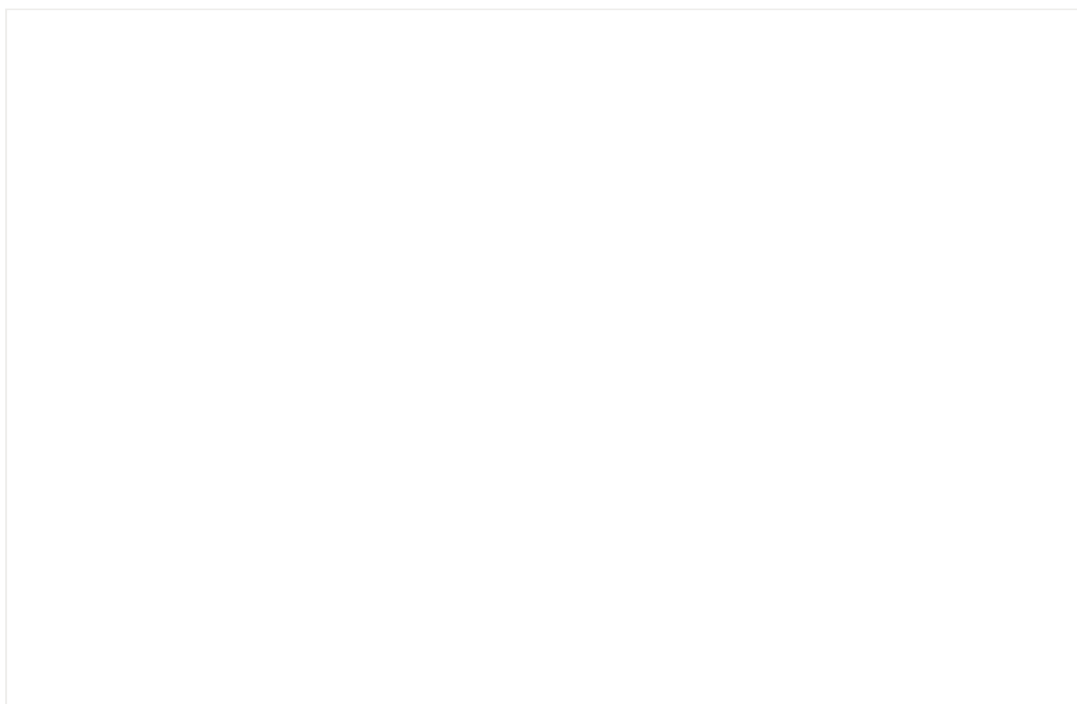
- **PSPNet** (pyramid scene parsing network) 通过对不同区域的上下文信息进行聚合，提升了网络利用全局上下文信息的能力。在PSPNet，金字塔池化生成的不同层次的特征图最终被flatten并concat起来，再送入全连接层以进行分类，消除了CNN要求图像分类输入大小固定的限制。而在PSPNet中，使用的策略是：pooling-conv-upsample，然后拼接得到特征图，然后进行标签预测。



PSPNet网络

创新点：多尺度池化，更好地利用全局图像级别的先验知识来理解复杂场景

- **RefineNet**通过细化中间激活映射并分层地将其连接到结合多尺度激活，同时防止锐度损失。网络由独立的Refine模块组成，每个Refine模块由三个主要模块组成，即：剩余卷积单元(RCU)，多分辨率融合(MRF)和链剩余池(CRP)。整体结构有点类似U-Net，但在跳跃连接处设计了新的组合方式（不是简单的concat）。个人认为，这种结构其实非常适合作为自己网络设计的思路，可以加入许多其他CV问题中使用的CNN module，而且以U-Net为整体框架，效果不会太差。



创新点：Refine模块

1.3 降低计算复杂的网络结构

也有很多工作致力于降低语义分割网络的计算复杂度。一些简化深度网络结构的方法：张量分解；通道/网络剪枝；稀疏化连接。还有一些利用NAS（神经架构搜索）取代人工设计来搜索模块或整个网络的结构，当然AutoDL所需的GPU资源会劝退一大批人。因此，也有一些人使用随机搜索来搜索小的多的ASPP模块，然后基于小模块来搭建整个网络模型。

网络轻量化设计是业内共识，移动端部署不可能每台机器配一张2080ti，另外耗电量、存储等问题也会限制模型的推广应用。不过5G如果能普及的话，数据就可以全部在云端处理，会很有意思。当然，短期内（十年），5G全方位部署不知道是否可行。

1.4 基于注意力机制的网络结构

注意力机制可以定义为：使用后续层/特征图信息来选择和定位输入特征图中最具判断力（或显著性）的部分。简单地可以认为是给特征图加权的一种方式（权值通过网络计算得到），按照权值的作用方式的不同，可以分为通道注意力机制（CA）和空间注意力机制（PA）。FPA（Feature Pyramid Attention，特征金字塔注意力）网络是一种基于注意力机制的语义分割网络，它将注意力机制和空间金字塔相结合，以提取用于像素级标记的精密特征，而没有采用膨胀卷积和人为设计的解码器网络。

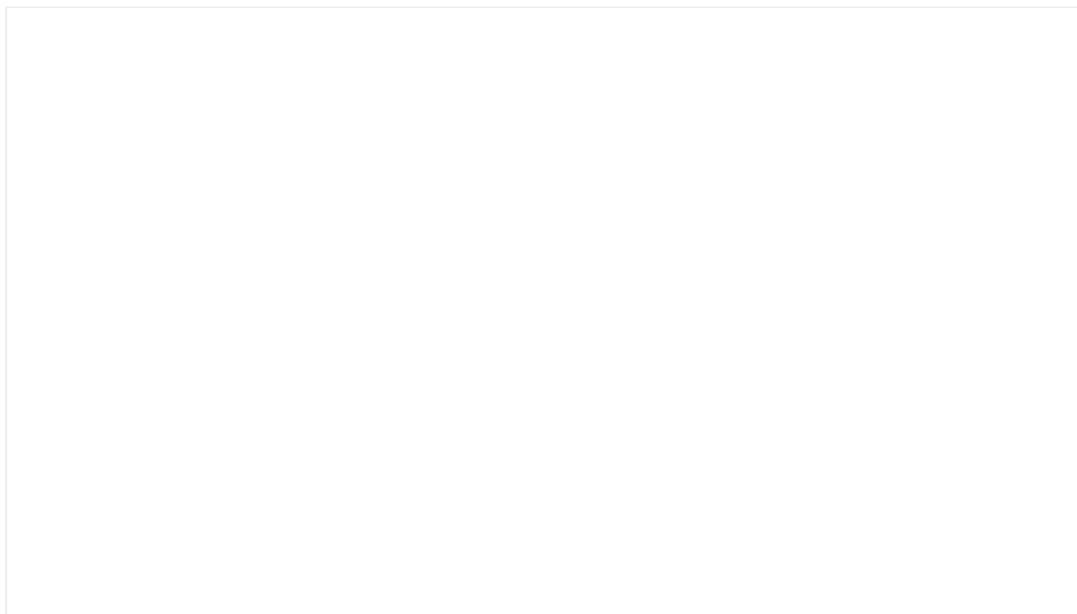
1.5 基于对抗学习的网络结构

Goodfellow等人在2014年提出了一种对抗的方法来学习深度生成模型，生成对抗网络（GANs）中需要同时训练两个模型：捕获数据分布的生成模型G，和估计样本来自训练数据的概率的判别模型D。

- G是一个生成式的网络，它接收一个随机的噪声 z （随机数），通过这个噪声生成图像
- D是一个判别网络，判别一张图片是不是“真实的”。它的输入参数是 x （一张图片），输出 $D(x)$ 代表 x 为真实图片的概率，如果为1，就代表100%是真实的图片，而输出为0，就代表不可能是真实的图片。

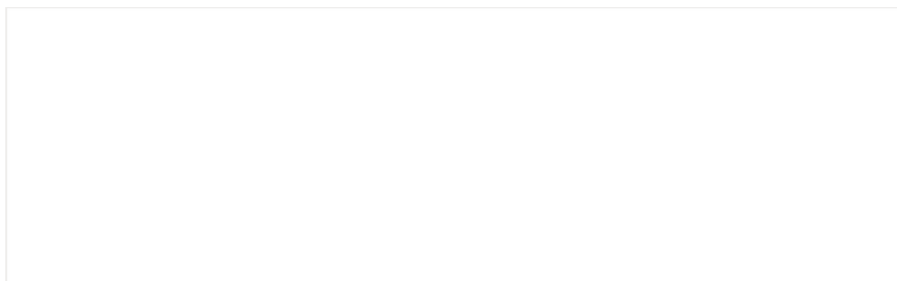
G的训练程序是将D错误的概率最大化。可以证明在任意函数G和D的空间中，**存在唯一的解决方案，使得G重现训练数据分布，而 $D=0.5$** 。训练过程中，生成网络G的目标就是尽量生成真实的图片去欺骗判别网络D。而D的目标就是尽量辨别出G生成的假图像和真实的

图像。这样，G和D构成了一个动态的“博弈过程”，最终的平衡点即纳什均衡点。在G和D由神经网络定义的情况下，整个系统可以用反向传播进行训练。



GANs网络结构示意图

受到GANs启发，Luc等人训练了一个语义分割网络（G）以及一个对抗网络（D），对抗网络区分来自ground truth或语义分割网络（G）的分割图。G和D不断地博弈学习，它们的损失函数定义为：

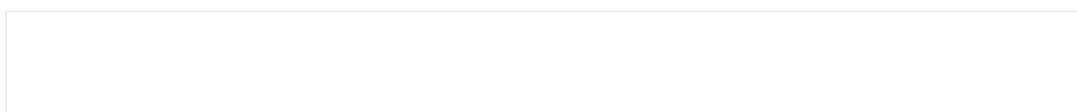


GANs loss function

其中 θ_s 和 θ_a 分别代表分割网络（G）和对抗网络（D）的参数。 l_{mce} 和 l_{bce} 分别是多类和二类交叉熵损失，两部分共同组成网络的损失函数。

回顾一下原始的GAN损失函数：

GANs的损失函数体现的是一种零和博弈的思想，原始的GANs的损失函数如下：



损失的计算位置是在D（判别器）输出处，而D的输出一般是fake/true的判断，所以整体可以认为是采用了二分类交叉熵函数。由GANs的损失函数的形式可知，训练要分为两部分：

首先是maxD部分，因为训练一般是先保持G（生成器）不变训练D的。D的训练目标是正确区分fake/true，如果我们以1/0代表true/fake，则对第一项E因为输入采样自真实数据所以我们期望D(x)趋近于1，也就是第一项更大。同理第二项E输入采样自G生成数据，所以我们期望D(G(z))趋近于0更好，也就是说第二项又是更大。所以是这一部分是期望训练使得整体更大了，也就是maxD的含义了。这部分只更新D的参数。

第二部分保持D不变（不进行参数更新），训练G，这个时候只有第二项E有用了，关键来了，因为我们要迷惑D，所以这时将label设置为1(我们知道是fake，所以才叫迷惑)，希望D(G(z))输出接近于1更好，也就是这一项越小越好，这就是minG。当然判别器哪有这么好糊弄，所以这个时候判别器就会产生比较大的误差，误差会更新G，那么G就会变得更好了，这次没有骗过你，只能下次更努力了（引自<https://www.cnblogs.com/walter-xh/p/10051634.html>）。这时候只更新G的参数。

从另一个角度看GANs，判别器（D）相当于一种特殊的损失函数（由神经网络构成，不同于传统的L1、L2、交叉熵等损失函数）。

另外GANs训练方式特殊，存在梯度消失、模式崩溃等问题（目前好像有办法可以解决），但其设计思想确实是深度学习时代一个伟大的发明。

1.6 小结

基于深度学习的图像语义分割模型大多遵循编码器-解码器体系结构，如U-Net。近几年的研究成果表明，膨胀卷积和特征金字塔池可以改善U-Net风格的网络性能。在第2节中，我们总结一下，如何将这些方法及其变体应用于医学图像分割。

2. 网络结构创新在医学图像分割中的应用

这部分介绍一些网络结构创新在2D/3D医学图像分割中的应用研究成果。

2.1 基于模型压缩的分割方法

为了实现实时处理高分辨率的2D/3D医学图像（例如CT、MRI和组织病理学图像等），研究人员提出了多种压缩模型的方法。weng等人利用NAS技术应用于U-Net网络，得到了在CT，MRI和超声图像上具有更好的器官/肿瘤分割性能的小型网络。Brugger通过利用组归一化（group normalization）和Leaky-ReLU（leaky ReLU function），重新设计了U-Net架构，以使网络对3D医学图像分割的存储效率更高。也有人设计了参数量更少的扩张卷积module。其他一些模型压缩的方法还有权重量化（十六位、八位、二值量化）、蒸馏、剪枝等等。

2.2 编码-解码结构的分割方法

Drozdal提出了一种在将图像送入分割网络之前应用简单的CNN来对原始输入图像进行归一化的方法，提高了单子显微镜图像分割、肝脏CT、前列腺MRI的分割精度。Gu提出了在主干网络利用扩张卷积来保留上下文信息的方法。Vorontsov提出了一种图到图的网络框架，将具有ROI的图像转换为没有ROI的图像（例如存在肿瘤的图像转换为没有肿瘤的健康图像），然后将模型去除的肿瘤添加到新的健康图像中，从而获得对象的详细结构。Zhou等人提出了一种对U-Net网络的跳跃连接重新布线的方法，并在胸部低剂量CT扫描中的结节分割，显微镜图像中的核分割，腹部CT扫描中的肝脏分割以及结肠镜检查视频中的息肉分割任务中测试了性能。Goyal将DeepLabV3应用到皮肤镜彩色图像分割中，以提取皮肤病变区域。

2.3 基于注意力机制的分割方法

Nie提出了一种注意力模型，相比于baseline模型（V-Net和FCN），可以更准确地分割前列腺。SinHa提出了一种基于多层注意力机制的网络，用于MRI图像腹部器官分割。Qin等人提出了一个扩张卷积模块，以保留3D医学图像的更多细节。其他基于注意力机制的医学图像分割论文还有很多。

2.4 基于对抗学习的分割网络

Khosravan提出了从CT扫描中进行胰腺分割的对抗训练网络。Son用生成对抗网络进行视网膜图像分割。Xue使用全卷积网络作为生成对抗框架中的分割网络，实现了从MRI图像分割脑肿瘤。还有其他一些成功应用GANs到医学图像分割问题的论文，不再一一列举。

2.5 基于RNN的分割模型

递归神经网络（RNN）主要用于处理序列数据，长短期记忆网络（LSTM）是RNN的一个改进版本，LSTM通过引入自环（self-loops）使得梯度流能长期保持。在医学图像分析领域，RNN用于对图像序列中的时间依赖性进行建模。Bin等人提出了一种将全卷积神经网络与RNN融合的图像序列分割算法，将时间维度上的信息纳入了分割任务。Gao等人利用CNN和LSTM来对脑MRI切片序列中的时间关系进行建模，以提高4D图像中的分割性能。Li等人先用U-Net获得初始分割概率图，后用LSTM从3D CT图像中进行胰腺分割，改善了分割性能。其他利用RNN进行医学图像分割的论文还有很多，不再一一介绍。

2.6 小结

这部分内容主要是分割算法在医学图像分割中的应用，所以创新点并不多，主要还是对不同格式（CT还是RGB，像素范围，图像分辨率等等）的数据和不同部位数据的特点（噪声、对象形态等等），经典网络需要针对不同数据进行改进，以适应输入数据格式和特征，这样能更好的完成分割任务。虽然说深度学习是个黑盒，但整体上模型的设计还是有

章可循的，什么策略解决什么问题、造成什么问题，可以根据具体分割问题进行取舍，以达到最优的分割性能。

部分参考文献：

- 1 Deep Semantic Segmentation of Natural and Medical Images: A Review
- 2 NAS-Unet: Neural architecture search for medical image segmentation. IEEE Access, 7:44247-44
- 3 Boosting segmentation with weak supervision from image-to-image translation. arXiv preprint
- 4 Multi-scale guided attention for medical image segmentation. arXiv preprint arXiv:1906.02849
- 5 SegAN: Adversarial network with multi-scale L1 loss for medical image segmentation.
- 6 Fully convolutional structured LSTM networks for joint 4D medical image segmentation. In 201
- 7 <https://www.cnblogs.com/walter-xh/p/10051634.html>

-----版权声明-----

仅用于学术分享，若侵权请联系删除

若本文对您有帮助，请点个“分享”和“在看”吧

点击下方名片，关注CV技术指南



CV技术指南

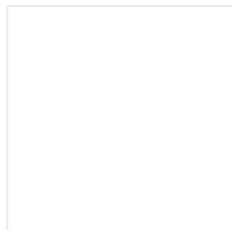
计算机视觉技术总结，计算机视觉论文分享和解读。
99篇原创内容

公众号

注：设置星标才能及时看到公众号更新的文章

在公众号中回复关键字“技术总结”可获取公众号原创技术总结的汇总pdf

扫描下方二维码添加个人微信号，备注“公众号读者”，邀请加入读者交流群。



其它文章

padding在深度学习模型中重要吗？

计算机视觉中的图像标注工具总结

使用 Ray 将 PyTorch 模型加载速度提高 340 倍
轻量化模型系列--GhostNet：廉价操作生成更多特征
深度学习有哪些trick?
技巧：如何将内存占用降低90%
如何看待计算机视觉未来的走向
从CVPR 2021的论文看计算机视觉的现状
CV 面试问题详解宝典--目标检测篇
卷积神经网络的复杂度分析
2021年小目标检测最新研究综述
汇总 | 国内外经典开源数据集
CV技术指南--精华文章汇总分类
TorchShard 简介 | Pytorch Lightning Flash 简介
Pytorch代码调试工具--torchsnooper
资源分享 | 使用 FiftyOne 加快您的论文写作速度
计算机视觉中的小样本学习综述
计算机视觉入门路线 | 知识蒸馏的简要概述
论文创新的常见思路总结 | GPU多卡并行训练总结
各种 Optimizer 梯度下降优化算法回顾和总结
CV方向的高效阅读英文文献方法总结
Siamese network总结 | FLOPS与FLOPs区别
归一化方法总结 | 欠拟合与过拟合技术总结
NMS总结 | 损失函数技术总结
注意力机制技术总结 | 特征金字塔技术总结
池化技术总结 | 数据增强方法总结

阅读 357

分享

收藏

赞 1

在看 1

收藏此内容的人还喜欢

同样都是调参，为什么人家的神经网络比我牛逼 100 倍？

目标检测与深度学习 阅读 385

神经网络中的泛化总结

CV技术指南 阅读 469

基于深度学习的目标检测算法原理

人工智能感知信息处理算法研究院 阅读 333

