

## 卷积神经网络在医学图像分割中的研究进展

徐航<sup>1</sup>, 随力<sup>1</sup>, 张靖雯<sup>2</sup>, 赵彦富<sup>1</sup>, 李月如<sup>1</sup>

1. 上海理工大学医疗器械与食品学院, 上海 200093; 2. 上海交通大学医学院附属仁济医院放射科, 上海 200127

**【摘要】**卷积神经网络(CNN)是目前计算机视觉和模式识别中效果最为突出的算法。CNN拥有强大的空间识别能力,可以从图像中提取高阶的空间特征,同时通过共用卷积核的方式大幅减少参数量,从而在提升网络性能的同时保持总参数量在一个合理的、可运算的范畴。部分采用无监督学习的CNN算法可以在没有先验知识的条件下实现一定程度的图像语义分割,大幅减少人工读图的负担。本研究就CNN在医学图像分割中的研究进展和使用CNN时的具体技巧及其效果进行综述。以使用CNN为核心的深度学习工具解决医学图像分割的课题为中心,展示了CNN在有监督学习、半监督学习及无监督学习中的巨大潜力,分析比较了现有方案的优点与不足,探讨了未来CNN在医学图像领域的前进方向。

**【关键词】**卷积神经网络;医学图像;图像分割;深度学习;综述

**【中图分类号】**R318

**【文献标志码】**A

**【文章编号】**1005-202X(2019)11-1302-05

### Progress on convolutional neural network in medical image segmentation

XU Hang<sup>1</sup>, SUI Li<sup>1</sup>, ZHANG Jingwen<sup>2</sup>, ZHAO Yanfu<sup>1</sup>, LI Yueru<sup>1</sup>

1. School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;

2. Department of Radiology, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China

**Abstract:** Convolutional neural network (CNN) is regarded as the state-of-the-art algorithm in computer vision and pattern recognition. CNN which is excellent for spatial recognition can extract hierarchical features from images. The number of parameters is greatly reduced by sharing kernels, thus improving the network performance and keeping the number of total parameters in a reasonable and computable range. To some extent, some CNN algorithms based on unsupervised learning can perform semantic segmentation without prior knowledge, releasing the burden of manual works. Herein the research progresses of CNN in medical image segmentation as well as the techniques and outcomes of CNN are reviewed. Focusing on the topic of using deep learning with CNN as core to realize medical image segmentation, the great potentials of CNN in supervised learning, semi-supervised learning and unsupervised learning are introduced. By analyzing and comparing advantages and shortages of existing methods, the prospects of CNN in medical imaging are discussed.

**Keywords:** convolutional neural network; medical image; image segmentation; deep learning; review

### 前言

近年来,卷积神经网络(Convolutional Neural Network, CNN)已经在计算机视觉和模式识别领域中证实了它的应用价值<sup>[1]</sup>。自从2012年AlexNet的诞生以来,计算机视觉各类任务的准确率有了质的飞跃,有的甚至可以超过接受训练的人类<sup>[2]</sup>。CNN在医学影像领域的应用则从2015年开始有显著的增

长,医学图像计算和计算机辅助干预国际会议(Medical Image Computing and Computer Assisted Intervention, MICCAI)上关于CNN和深度学习的论文数量也逐年增多,可以说深度学习技术,尤其是CNN已经逐渐渗透了整个医学图像分析领域<sup>[3]</sup>。CNN拥有强大的空间识别能力,可以从原始输入中提取出高层特征信息,这些特征就好比构成原始图像的“笔画”,通过对这些特征进行分析和识别就可以轻松地完成对原始图像的分割操作。CNN的另一大特征就是可以通过共用卷积核和上下采样来大幅降低整个网络的参数量。有研究证明神经网络的深度是影响网络性能的关键。然而随着网络层数的增加,参数量的爆炸式增长为实际运算带来了难以承

**【收稿日期】**2019-07-12

**【作者简介】**徐航,研究生在读,研究方向:医学影像技术,E-mail: Buster\_X@live.com

**【通信作者】**随力,博士,教授,研究方向:神经工程,E-mail: lsui@usst.edu.cn

受的负担,使得神经网络训练需要大量的资源和时间。CNN的每一个卷积层只使用一个远小于图像尺寸的卷积核(可能存在多个通道),让这个卷积核在滑过图像的同时更新其参数。这样就既能充分利用图像整体的空间信息,又能有效限制参数量。全卷积网络更是完全放弃了参数量庞大的全连接层,全部采用卷积层,减少参数量的同时在图像分割任务上也有着出色的表现<sup>[4]</sup>。CNN的强大性能和巨大潜力激发了大量研究者们对其的钻研,新的方法和技巧层出不穷。本研究针对CNN在医学图像分割领域的研究进展进行综述。

## 1 CNN在图像分割中的应用

### 1.1 有监督学习与全卷积网络

常规的滑窗式CNN可以通过在指定像素的周围多次提取一小块图像进行运算,逐个像素进行判别,从而完成图像分割任务<sup>[4]</sup>。实际操作上就像是用一个卷积核滑过整幅图像,符合我们对卷积的一般认识,然而这其实是一种效率很低的做法,因为卷积核滑动的步幅绝大多数情况都小于卷积核的尺寸,这就导致了相邻步幅之间重复的内容很多,带来冗余的计算。同时,由于一次仅对相当于卷积核尺寸的一小块图像进行运算,网络很难把握图像整体的空间特征。另外也有研究表示池化层会导致输出的分辨率远低于输入<sup>[5]</sup>。U-Net由Ronneberger等<sup>[6]</sup>于2015年提出,这个全卷积网络最早是被应用在医学图像的分割上的,由于其强大的性能,很快又被其他领域借鉴使用。成功训练一个深度网络需要大量的人为标注的数据集,ImageNet大赛使用的训练数据集中图片的数量往往都以万为单位。医学图像由于设备之间差异和人员的精力时间所限,难以获取规格相近的数据。U-Net近乎是一个对称的网络,由收缩路径和扩张路径组成,其中,收缩路径与典型的CNN相似,用于获取空间信息;扩张路径则由上采样和 $2\times 2$ 的卷积层组成,用于精确定位。两条路径上分别具有数量相同的上采样层和下采样层,相对应的层之间由跳过连接结合,把收缩路径和扩张路径上获取的特征联系在了一起。从训练的角度来看,U-Net可以一次性获取整幅图像的空间信息,同时有效避免滑窗式CNN相邻步幅之间输入内容的大面积重叠。这样的网络结构大幅提升了分割的准确性,同时训练的速度极快且不需要大量的标记数据集。在2014年和2015年ISBI细胞示踪挑战中,该网络于仅有35幅训练图像的数据集“PhC-U373”上取得92%的IOU(Intersection Over Union,用于表示预测和真

实情况之间的相关度,该值越高,相关度越高),于仅有20幅训练图像的数据集“DIC-HeLa”上取得77.5%的IOU<sup>[7]</sup>。V-Net可以理解成U-Net的一个改良版。Milletari等<sup>[1]</sup>在U-Net的基础上引入残差单元,并用Dice损失层替换交叉熵损失函数,有效解决在训练过程中模型陷入局部最小值而停止,不再寻求全局最小值的问题,提升了分割的精度。同时,V-Net可以在3D的磁共振(Magnetic Resonance, MR)图像上进行工作,可以说不论是在精度还是维度上都对U-Net进行了明显的改良。V-Net在PROMISE 2012(一系列前列腺的MR图像)数据集上取得了82.39的高分,与当年的最佳成绩Imorphics的84.36相差甚微<sup>[8]</sup>。同等硬件条件下,完成一幅图像的训练Imorphics需要8 min,V-Net仅需要1 s。

残差网络由He等<sup>[9]</sup>于2015年提出,可以说是深度学习网络结构上最新的重大发明。残差网络允许原始输入信息直接传递到后面的层中,此时网络学习的内容就不再是完整的输出了,而是输出与输入的差,即残差。残差结构有利于保护输入信息的完整性,同时简化学习目标和难度,一经发表就被大量的研究者应用在他们的网络结构中,U-Net中使用的跳过连接就是一种残差结构。Yu等<sup>[10]</sup>使用了一个全卷积残差网络(Fully Convolutional Residual Network, FCRN)进行皮肤黑色素瘤损伤的精确分割,荣获ISBI 2016年该项分割任务的第二名。该研究将38、50、101层的FCRN和未采用残差结构的VGG-16及GoogleNet(22层)进行对比,结果表明FCRN的表现全面优于未采用残差结构的这两个网络,且50层FCRN的表现最为出色。Chen等<sup>[11]</sup>将残差网络推广到3D的MR影像中,实现脑部体积图像的分割。实际上在这项任务中他们使用了两个残差网络:(1)VexResNet用于对 $T_1$ 权重图像进行初步分割;(2)Auto-context VexResNet则对 $T_1$ -IR、 $T_2$ -FLAIR等多模图像和初步分割好的 $T_1$ 权重图像执行自动上下文算法,取得更为精确的分割结果。值得一提的是,VexResNet使用了多重分支结构<sup>[11]</sup>,在不同的分支上采取不同的卷积策略,这样的结构也普遍为残差网络所使用。

### 1.2 无监督学习与自编码器

在CNN的应用过程中无监督学习越来越受到重视,研究者们都希望能创造一种能真正意义上替代具有专业知识的人类完成医学图像分割的工具。无监督学习下,网络不需要医师标注好的分类结果作为训练材料,只需要对输入图像进行多层卷积等操作便可以得到其高层空间特征,从而将感兴趣区域



和图像背景区分开来。自编码器<sup>[12]</sup>是一个经典的无监督学习模型,通过一个编码器和一个解码器来重构输入。最初的自编码器中,编码器和解码器都只有一层。随着深度学习思想的渗透,编码器和解码器的层数逐渐增多,层层堆叠,所以这样的自编码器也被称作栈式自编码器(Stacked Auto-Encoder, SAE)<sup>[13]</sup>或深度自编码器。同时,卷积层和池化层也被应用在自编码器中,以增强对图像输入的处理能力,通过反复的卷积和池化操作,便能在整个自编码器正中间的隐含层中得到一个维度远小于原始输入的特征向量。这个特征向量可以通过配套的解码器来近似还原成原始输入,所以自编码器在图像分割中常常用于提取图像的高阶特征。从MR图像中分割出视觉传导通路具有相当的难度,该组织狭长且结构多变,同时与周围组织的解剖特征区别不明显,要想实现病理意义上的划分(例如分割出胶质瘤)就更具挑战了。Mansoor等<sup>[13]</sup>利用SAE在MR图像中推断出视觉传导通路的特征,实现了左视神经、右视神经及视交叉的全自动分段模型划分,实验中的Dice相关系数高达0.779,为当时的最佳成绩,同时相比于常规方法拥有相同的鲁棒性。研究表示SAE在低对比度区域的效果尤其突出,同时使用SAE提取高阶特征也使得他们的方法可同时应用于健康和病态的组织上。Guo等<sup>[14]</sup>发现以往用于从脑部MR图像中划分成人海马体的方法并不适用于1岁以下的婴儿,并推测这是因为海马体和周围的脑组织之间缺乏具有足够判别力的特征表示,他们提出使用SAE实现无监督学习,同时从 $T_1$ 和 $T_2$ 权重图像中提取互补特征的方法来解决这个问题,并在2周至9个月大的婴儿的脑部图像上证明了该方法的有效性。Moriya等<sup>[15]</sup>将Yang等<sup>[16]</sup>的无监督学习图像聚类法推广到3D的医学图像上。他们首先训练一个包含了3个卷积层、1个池化层和2个全连接层的CNN来提取图像中每一小块的特征表示;然后,再用K均值法进行聚类,将这些特征表示分为K类并给每一类赋予标签;最后给每一小块最中心的几个体素赋予和该一小块所属类相同的标签,即将聚类的结果投影到原图上,完成图像的分割。K的数值可以根据实际需要进行调整,且无需再提取特征,标签的赋予也仅用作区分,不需要专业的医学知识。Moriya等<sup>[15]</sup>将K设置为3,将肺部的CT图像划分为3部分:侵袭性癌变、非侵袭性癌变和正常组织,相比直接应用K均值法和多阈值大津法,该方法取得了更为优秀的结果。Bao等<sup>[17]</sup>使用了一种相当新颖的方式来提升脑部MR图像分割的质量。他们使用了一个多重规格结构的CNN,其每一层有

尺寸不同的多个卷积核,分别进行运算后在最后几层进行汇总。多重规格结构的CNN被用于对图像进行特征提取并赋予每一小块可能的标签;再使用随机游走法强制实现标签的连续性,即相邻的小块应尽可能拥有同样的标签,使得分割的结果更加平滑<sup>[18]</sup>。此法在IBSR数据集上取得了平均0.822的准确度,在LPBA40数据集上取得了平均0.850的准确度,相较于之前的PBL法<sup>[19]</sup>(其上述两项准确度分别为0.760、0.843),不论是在精度还是速度上都有显著提升。

### 1.3 半监督学习与预训练

预训练就是在正式训练前,利用某些初始化方式为网络内参数赋予初值,然后再开始训练。此时训练中参数更新的幅度较小,也有锁定前几层网络参数,仅让训练对最后几层做出调整的方法,这种手法也被称为微调<sup>[20]</sup>。较为早期的网络参数初始化主要还是采取统计学手段,例如数据标准化,但本质上还是一个随机取值的过程。近年来,各领域的研究普遍表明,一些特殊的预训练方法可以明显提升网络的学习效率和精度,其中,有些方法本身就应用了CNN,故值得单独讨论,如上文提到的自编码器就经常应用于CNN的预训练。Kallenberg等<sup>[21]</sup>将他们使用稀疏自编码器进行无监督预训练的网络CSAE(Convolutional Sparse Auto-Encoder)应用在乳腺密度分割上。这种无监督预训练和有监督微调的组合也被称作半监督学习。而Sun等<sup>[22]</sup>在他们关于乳腺癌症状区域的分割筛选上则采用了另外一种半监督学习方式。该研究面临的情况是在3 158个样本的训练集中,仅有100例有标注,其余为无标注数据。他们先通过21个图像数据指标同时对标注数据和无标注数据进行特征提取及分类;然后,通过同类的标注数据为无标注数据赋予标签,从而转化为标注数据,反复迭代这一过程,直至所有数据都得到标注;最后,通过CNN进行划分。该研究赋予标签的准确率在80%左右,最终结果的准确率为82.43%。值得关注的是,在3 158个样本全都正确标记的情况下,同样的卷积网络可以取得85.52%的准确率,可见半监督学习可以在不损失太多正确率的情况下,大幅减少人为标注数据的工作量,或是在只有少量样本被标注的情况下依然能较好地完成分割任务。该研究通过主成分分析的方法实现特征提取,通过CNN进行更高层的特征提取,有望进一步提升赋予标签的准确率。

类似的方法也被应用在3D以及动态医学图像上。Brosch等<sup>[23]</sup>使用含卷积的受限玻尔兹曼机对他

们的3D卷积编码器进行预训练,然后将这个类编码器的网络应用在多发性硬化的损伤分割上,在MICCAI 2008和ISBI 2015两个数据集上取得了顶尖的成绩。Yu等<sup>[24]</sup>将微调应用到了动态医学图像上,在超声心动图里左心室的分割任务中大量使用预训练和微调。开始分割前,使用一批有标记的数据对CNN进行预训练,开始分割任务后,CNN会动态地使用第一帧和剩余帧不断地进行微调。尽管他们的网络结构相当简单,但创新性的动态微调帮助他们在51组超声心动图的分割任务中取得与人为分割相接近的结果,并且明显优于现有的其他手段。另外也有不少工作是基于迁移学习完成的,即使用在类似领域训练好的网络进行参数初始化后再进行微调,可以一定程度解决目标任务训练数据不足的问题。Chen等<sup>[25]</sup>对一个在PASCAL VOC数据集<sup>[26]</sup>上训练好的模型在他们的超声图像集上进行微调,并使用它完成了5个不同视角的超声图像下左心室的分割,结果与人为分割结果相比也毫不逊色。

## 2 挑战与展望

深度CNN作为近几年来计算机视觉的新宠,已逐渐渗透医学影像整个领域,在解剖结构及症状区域的识别、分割以及医学图像的增强、配准及融合等方面都取得了令人眼前一亮的成果。CNN应用于医学图像分割具有得天独厚的优势,可以通过多层卷积提取输入图像的高层空间特征并以此为依据执行分割,在人眼难以识别并区分区域,效果尤其显著。CNN在医学图像分割领域的应用意义重大,一方面提高了分割算法的效率和精度,另一方面解放了人力资源,减轻医师的负担。随着不断研究与发展,相信不久的将来无监督学习将会在各种任务中扮演更为重要的角色,甚至可以在某些任务中真正意义上替代人类,完成没有先验知识条件下的图像分割。然而,CNN的进一步推广也面临着若干挑战:(1)现有的硬件水平普遍难以负担CNN带来的海量运算。本文提及的实验绝大多数都将图像缩小或裁剪后再提供给CNN,分辨率等级仅限于各维度几百像素,这对于精确定位和诊断来说还是远远不够的。理想情况下我们希望有足够强大的硬件设备能一次性读入整张高分辨率的图片,同时在网络结构中不需要下采样降低分辨率以减轻运算负担。(2)大部分任务中CNN的精度比起人为分割仍有一定差距,所以目前几乎所有的医学图像分割CNN都仅处于实验阶段,并未投入实际使用。但随着各种网络结构和训练技巧的推陈出新,CNN精度超过一般从业人员

的势头已经很明显,并有望在未来的几年内达到专家水平。(3)医学图像数据因为设备间差异和技师操作习惯等原因,难以取得大量同规格的标记数据作为训练资料。这个问题在医疗联网逐渐普及的如今有望在短期内得到解决,另外一方面也可以寄希望于无监督学习的发展,直接对现有数据进行特征挖掘,而不依赖于人为的标记。总而言之,CNN的应用使医学图像分割向着自动化迈出了确实的一步。

## 【参考文献】

- [1] MILLETARI F, NAVAB N, AHMADI S A. V-net: fully convolutional neural networks for volumetric medical image segmentation[C]. 2016 Fourth International Conference on 3D Vision, 2016: 565-571.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Commun ACM, 2017, 60(6): 84-90.
- [3] LITJENS G, KOOI T, BEJNORDI B E, et al. A survey on deep learning in medical image analysis[J]. Med Image Anal, 2017, 42: 60-88.
- [4] LONG J, SHEHMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [5] SHEHMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(4): 640-651.
- [6] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]. International Conference on Medical Image Computing and Computer-assisted Intervention, 2015: 234-241.
- [7] MASKA M, ULMAN V, SVOBODA D, et al. A benchmark for comparison of cell tracking algorithms[J]. Bioinformatics, 2014, 30(11): 1609-1617.
- [8] VINCENT G, GUILLARD G, BOWES M. Fully automatic segmentation of the prostate using active appearance models[C]. PROMISE 12: MICCAI 2012 Grand Challenge on Prostate MR Image Segmentation, 2012: 1-8.
- [9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [10] YU L Q, CHEN H, DOU Q, et al. Automated melanoma recognition in dermoscopy images via very deep residual networks[J]. IEEE Trans Med Imaging, 2017, 36(4): 994-1004.
- [11] CHEN H, DOU Q, YU L, et al. VoxResNet: deep voxelwise residual networks for volumetric brain segmentation [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, 4: 39.
- [12] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.
- [13] MANSOOR A, CERROLAZA J J, IDREES R, et al. Deep learning guided partitioned shape model for anterior visual pathway segmentation[J]. IEEE Trans Med Imaging, 2016, 35(8): 1856-1865.
- [14] GUO Y, WU G, COMMANDER L A, et al. Segmenting hippocampus from infant brains by sparse patch matching with deep-learned features [C]. International Conference on Medical Image Computing and

- Computer-Assisted Intervention. Cham: Springer, 2014: 308-315.
- [15] MORIYA T, ROTH H R, NAKAMURA S, et al. Unsupervised segmentation of 3D medical images based on clustering and deep representation learning [C]. Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging. International Society for Optics and Photonics, 2018, 10578: 1057820.
- [16] YANG J, PARIKH D, BATRA D. Joint unsupervised learning of deep representations and image clusters [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 5147-5156.
- [17] BAO S, CHUNG A S. Multi-scale structured CNN with label consistency for brain MR image segmentation[J]. Comput Method Biomec, 2018, 6(1): 113-117.
- [18] GRADY L. Random walks for image segmentation[J]. IEEE Trans Pattern Anal Mach Intell, 2006, 28(11): 1768-1783.
- [19] COUPÉ P, MANJÓN J V, FONO V, et al. Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation[J]. Neuroimage, 2011, 54(2): 940-954.
- [20] RUDER S. Transfer learning-machine learning's NextFrontier[R/OL]. [2017-04-01].
- [21] KALLENBERG M, PETERSEN K, NIELSEN M, et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring[J]. IEEE Trans Med Imaging, 2016, 35(5): 1322-1331.
- [22] SUN W Q, TSENG T, ZHANG J Y, et al. Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data [J]. Comput Med Imaging Graph, 2017, 57: 4-9.
- [23] BROSCHE T, TANG L Y, YOO Y, et al. Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation[J]. IEEE Trans Med Imaging, 2016, 35(5, SI): 1229-1239.
- [24] YU L, GUO Y, WANG Y Y, et al. Segmentation of fetal left ventricle in echocardiographic sequences based on dynamic convolutional neural networks[J]. IEEE Trans Biomed Eng, 2017, 64(8): 1886-1895.
- [25] CHEN H, ZHENG Y, PARK J H, et al. Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016: 487-495.
- [26] CHEN L C, PAPANDREOU G, KOKKINOS I A, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Trans Pattern Anal Mach Intell, 2018, 40(4): 834-848.

(编辑:谭斯允)