软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2021,32(2):519−550 [doi: 10.13328/j.cnki.jos.006104]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于 U-Net 结构改进的医学影像分割技术综述*

殷晓航, 王永才, 李德英

(中国人民大学 信息学院,北京 100872)

通讯作者: 王永才, E-mail: ycw@ruc.edu.cn

摘 要: 深度学习在医学影像分割领域得到广泛应用,其中,2015 年提出的 U-Net 因其分割小目标效果较好、结构具有可扩展性,自提出以来受到广泛关注.近年来,随着医学图像割性能要求的提升,众多学者针对 U-Net 结构也在不断地改进和扩展,比如编解码器的改进、外接特征金字塔等.通过对基于 U-Net 结构改进的医学影像分割技术,从面向性能优化和面向结构改进两个方面进行总结,对相关方法进行了综述、分类和总结,并介绍图像分割中常用的损失函数、评价参数和模块,进而总结了针对不同目标改进 U-Net 结构的思路和方法,为相关研究提供了参考.

关键词: U-Net;医学影像分割;结构改进;深度神经网络;技术综述

中图法分类号: TP391

中文引用格式: 殷晓航,王永才,李德英.基于 U-Net 结构改进的医学影像分割技术综述.软件学报,2021,32(2):519−550. http://www.jos.org.cn/1000-9825/6104.htm

英文引用格式: Yin XH, Wang YC, Li DY. Suvery of medical image segmentation technology based on U-Net structure improvement. Ruan Jian Xue Bao/Journal of Software, 2021,32(2):519−550 (in Chinese). http://www.jos.org.cn/1000-9825/6104.htm

Suvery of Medical Image Segmentation Technology Based on U-Net Structure Improvement

YIN Xiao-Hang, WANG Yong-Cai, LI De-Ying

(School of Information, Renmin University of China, Beijing 100872, China)

Abstract: The application of deep learning in the field of medical image segmentation has attracted great attentions, among which the U-Net proposed in 2015 has been widely concerned because of its good segmentation effect and scalable structure. In recent years, with the improvement of the performance requirements of medical image segmentation, many scholars are improving and expanding the U-Net structure, such as the improvement of encoder-decoder, or the external feature pyramid, and so on. In this study, the medical image segmentation technology based on U-Net structure improvement is summarized from the aspects of performance-oriented optimization and structure-oriented improvement. Related methods are reviewed, classified and summarized. The paper evaluates the parameters and modules, and then summarizes the ideas and methods for improving the U-Net structure for different goals, which provides references for related research.

Key words: U-Net; medical image segmentation; structural improvement; deep neural network; technology survey

随着深度学习技术的快速发展,深度学习在医学影像领域的应用吸引了广泛的研究和关注.其中,如何自动识别和分割医学影像中的病灶是最受关注的问题之一.为解决这一问题,2015 年,Ronneberger 等人在 MICCAI 会议发表 U-Net[1],是深度学习在医学影像分割中的突破性的进展.U-Net 是基于 FCN(fully convolutional network)改进而成,包括编码器、瓶颈(bottleneck)模块、解码器几部分组成,由于其 U 型结构结合上下文信息和训练速度快、使用数据量小,满足医学影像分割的诉求,而在医学影像分割中广泛应用.U-Net 的结构如图 1 所

* 基金项目: 国家自然科学基金(61972404, 61672524, 11671400)
Foundation item: National Natural Science Foundation of China (61972404, 61672524, 11671400)
收稿时间: 2020-05-09; 修改时间: 2020-06-02; 采用时间: 2020-06-20; jos 在线出版时间: 2020-07-27

© 中国科学院软件研究所 http://www.jos.org.cn

示.由于病灶的形状的多样性和不同器官结构的差异性,仅使用 U-Net 结构分割病灶无法满足对于精准度、速度等的需求.

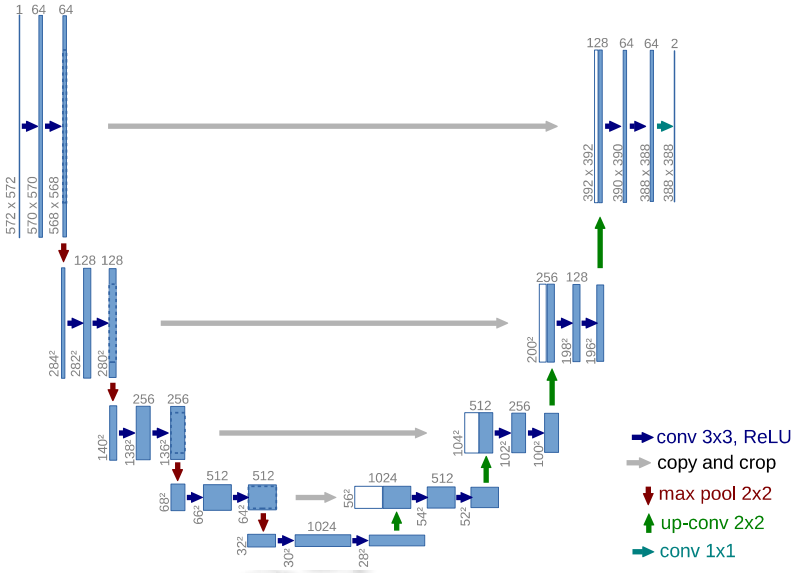


Fig.1 U-Net network structure diagram[1]

图 1 U-Net 网络结构图[1]

U-Net 自发表以来,其编码器-解码器-跳连的网络结构启发了大量基于 U-Net 结构改进的医学影像分割方法.随着深度学习技术的发展,包括注意力机制、稠密模块、特征增强、评价函数改进等基于 U-Net 的基础结构,将这些深度神经网络发展的最新技术引入到医学影像分割应用中,成为被广泛采取的改进方法.这些相关工作或者面向不同的优化目标,或者通过结构改进、添加新模块等手段,提高医学影像分割的准确性、运算效率、适用范围等.由于相关工作众多,而且大多数工作是结合实际问题,不断地加入新的思想,现有文献中对 U-Net 结构改进的相关工作尚缺少较好的综述和总结的工作.本文拟从改进目的和改进手段两个方面对近几年基于 U-Net 结构改进的医学影像分割的工作进行综述.

- 面向性能优化的改进工作主要包括:(1) 将 U-Net 扩展到 3D 图像[2,3];(2) 增强相关特征,抑制无关特征[4−13];(3) 改进计算速度、内存占用[14−22];(4) 改进特征融合方法[19,23−30];(5) 针对小样本训练数据集的改进[31−34];(6) 提高泛化能力的改进[35,36].
- 针对 U-Net 模块结构的改进主要包括:(1) 针对编码器、解码器结构的改进[37−45];(2) 针对损失函数的改进[2,7,32,41,46−49];(3) 对瓶颈(bottleneck)模块结构的改进[9,31,50];(4) 增加数据流路径的改进[49,51];(5) 采用自动结构搜索的改进[52]等方面.

图 2 给出了本文对 U-Net 相关研究工作的分类方法.虽然有的相关工作同时被两个层面包含,但这种分类总结能使得我们更清晰地了解该工作的改进目的和实现目的的改进手段.针对每类改进的具体方法,本文较详细的介绍了方法的主要设计思想、改进效果、所使用的数据集、评价指标等,并最后给出对相关方法的整体的总结和比较.此外,本文还提炼出 U-Net 结构改进中一些常见的基础结构模块,这些基础结构模块对深度学习网络结构的改进具有较为普遍的借鉴意义.

本文第 1 节介绍医学影像分割深度神经网络中的一些常见的损失函数和评价参数.第 2 节、第 3 节从两个方面、11 个子类总结和介绍基于 U-Net 结构改进的医学影像分割的相关研究工作.第 4 节提炼医学影像分割研究中常见的一些特殊结构.第 5 节对文中所提的算法进行总结、对比和展望.

© 中国科学院软件研究所 http://www.jos.org.cn

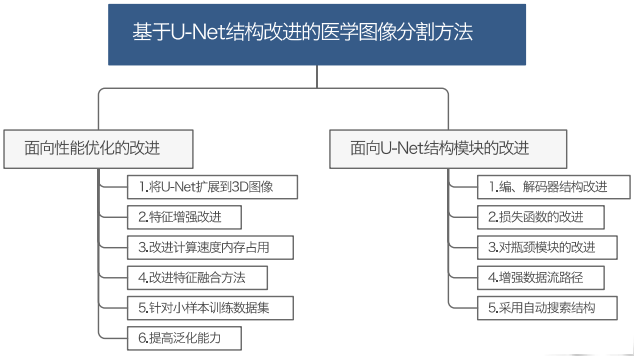


Fig.2 Medical image segmentation method based on U-Net structure improvement

图 2 基于 U-Net 结构改进的医学影像分割方法

1 医学影像分割中神经网络常采用的评价参数和损失函数

损失函数和评价参数是训练网络是必不可少的部分:损失函数表示预测和目标之间的差异,常用交叉熵、Dice loss 等判断训练模型与真值之间的差异;分割评价参数是评价图像分割好坏的重要参数,常用 Dice 等评价网络模型的优劣.本节主要列举几个图像分割神经网络中常用的评价参数和损失函数.

1.1 评价参数

在评价参数之前,先要介绍一下机器学习中的混淆矩阵.混淆矩阵主要是解决二分类问题[53].其中,*TP*=True Positive=真阳性,*FP*=False Positive=假阳性,*FN*=False Negative=假阴性,*TN*=True Negative=真阴性.

1.1.1 精确率(precision)

精确率表示的是预测为正的样本中有多少被预测正确.

$$P=\frac{TP}{TP+FP} \tag{1}$$

1.1.2 召回率(recall)

召回率就是召回目标类别,即表示样本中的正样本有多少被预测正确.

$$R=\frac{TP}{TP+FN} \tag{2}$$

1.1.3 准确率(accuracy)

准确率是评估获得所有成果中目标成果所占的比率.

$$ACC=\frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

1.1.4 综合评价指标(*F*-measure)

F-Measure 是综合精确率和召回率的评估指标,用于反映整体的情况.

$$F=\frac{(\alpha^2+1)PR}{\alpha^2(P+R)} \tag{4}$$

当α=1 时,

$$F1=\frac{2PR}{P+R} \tag{5}$$

1.1.5 IoU(intersection over union)/Jaccard Index

IoU 又称为 Jaccard Index[54],是目标检测常用到的评价参数,通过预测边框和真实边框的比值计算两个样本的相似度或者重叠度.我们分别用 *Vseg*,*Vgt* 表示两个轮廓区域所包含的点集(*Vseg* 为预测,*Vgt* 为真实标签),范围[0,1],则

© 中国科学院软件研究所 http://www.jos.org.cn

$$J=\frac{|V_{seg}\cap V_{gt}|}{|V_{seg}\cup V_{gt}|}=\frac{|V_{seg}\cap V_{gt}|}{|V_{seg}|+|V_{gt}|-|V_{seg}\cap V_{gt}|}=\frac{TP}{FP+TP+FN} \tag{6}$$

1.1.6 mIoU(mean intersection over union)

mIoU 为语义分割的标准度量,在每个类上计算 IoU 之后进行平均.由于其简洁、代表性强,大多数研究人员都采用该标准报告结果.假设共有 k+1 个类(从 L_0 到 L_k,其中包含一个空类或背景),p_{ij} 表示本属于类 i 但被预测为类 j 的像素数量.即 p_{ii} 表示真正的数量,而 p_{ij},p_{ji} 表示假正、假负,则 MIoU 定义为

$$MIoU=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ij}}{\sum_{j=o}^{k}p_{ij}+\sum_{j=0}^{k}p_{ij}-p_{ij}} \tag{7}$$

1.1.7 Dice coefficient

Dice 是一种评估两个轮廓区域相似度的函数,通常用于计算两个样本的相似度或者重叠度,其范围为[0,1].

$$Dice=2\frac{|V_{seg}\cap V_{gt}|}{|V_{seg}|+|V_{gt}|}=\frac{2TP}{FP+2TP+FN} \tag{8}$$

Jaccard Index 和 Dice coefficient 之间的换算公式为

$$J=\frac{D}{2-D},\ D=\frac{2J}{1+J} \tag{9}$$

1.1.8 SSIM

SSIM 是图像质量评价结构相似性指标,是基于样本 x 和 y 之间对于亮度、对比度、结构这 3 个方面进行比较,其范围为[0,1],值越大,两图像之间的差异越小[55].

$$SSIM(x,y)=[l(x,y)]^{\alpha}\cdot[c(x,y)]^{\beta}\cdot[s(x,y)]^{\gamma} \tag{10}$$

其中,图像照明度比较部分为 $l(x,y)=\frac{2\mu_x\mu_y+c_1}{\mu_x^2+\mu_y^2+c_1}$,图像对比度比较部分为 $c(x,y)=\frac{2\sigma_x\sigma_y+c_2}{\sigma_x^2+\sigma_y^2+c_2}$,图像结构比较部分为 $s(x,y)=\frac{\sigma_{xy}+c_3}{\sigma_x\sigma_y+c_3}$.其中,$\mu_x$ 为 x 的均值,μ_y 为 y 的均值;σ_x^2 为 x 的方差,σ_y^2 为 y 的方差,σ_{xy} 为 x 和 y 的协方差;c_1~c_3 为常数,避免除 0.

1.2 损失函数

损失函数的设计常要考虑数据集的特点,比方说,Focal loss 就是用来处理数据集中的难分样本.Dice 系数可以用来处理数据分类不均衡的情况,其中,不均衡很多情况下是由于背景和待分割区域之间的面积对比不均衡导致的.对于二分类,可以只考虑待分割区域,即是本文中的 Dice loss 函数;那么当对于多种类的分割时,同样可以只计算待分割区域的 Dice 系数,这样就可以避免背景占比太大,造成的数据集分类不均的情况.

1.2.1 交叉熵损失函数

设 y'是模型的输出,在 0-1 之间.对于正样本而言,输出越大,意味着损失越小;对于负样本而言,越小,则损失越小.所以,交叉熵的定义为

$$H=-y\log y'-(1-y)\log(1-y')=\begin{cases}-y\log y', & y=1\\ -(1-y)\log(1-y'), & y=0\end{cases} \tag{11}$$

1.2.2 Focal loss

Focal loss[56]是在交叉熵损失函数基础上进行的修改,主要是为了解决 one-stage 目标检测中正负样本比例严重失衡的问题.该损失函数降低了大量简单负样本在训练中所占的权重,也可理解为一种困难样本挖掘.

$$L_{ft}=\begin{cases}-\alpha(1-y')^{\gamma}\log y', & y=1\\ -(1-\alpha)y'^{\gamma}\log(1-y'), & y=0\end{cases} \tag{12}$$

γ>0 减少易分类样本的损失,使得更关注于困难的、错分的样本.平衡因子α用来平衡正负样本.实验证明,γ=2 最优.

© 中国科学院软件研究所 http://www.jos.org.cn

1.2.3 Dice loss

Dice loss[2]在感兴趣的解剖结构仅占据扫描的非常小的区域,从而使学习过程陷入损失函数的局部最小值.所以,要加大前景区域的权重.

$$L_{Dice}=1-Dice \tag{13}$$

1.2.4 Tversky loss

Tversky[57]系数是 Dice 系数和 Jaccard 系数的一种广义系数,V_{seg} 为预测标签,V_{gt} 为真实标签,公式定义为

$$T(V_{seg},V_{gt})=\frac{|V_{seg}\cap V_{gt}|}{|V_{seg}\cap V_{gt}|+\alpha|V_{seg}-V_{gt}|+\beta|V_{gt}-V_{seg}|} \tag{14}$$

$$L_{Tversky}=1-(V_{seg},V_{gt}) \tag{15}$$

当$\alpha=\beta=0.5$ 时,Tversky loss 为 Dice loss;当$\alpha=\beta=1$ 时,Tversky 系数就是 Jaccard 系数.$A-B$ 则意味着是 FP(假阳性),而 $B-A$ 则意味着是 FN(假阴性);α和β分别控制假阴性和假阳性.通过调整α和β,可以控制假阳性和假阴性之间的权衡[58].

2 基于 U-Net 面向性能优化的改进方法

本节和第 3 节将重点介绍基于 U-Net 改进的各类的图像分割方法,本节介绍面向性能优化的改进方法.现有的工作主要在应用范围、特征增强、训练速度优化、训练精度、特征融合、小样本训练集以及泛化能力提升几方面对 U-Net 提出各种改进进行研究,这些工作对网络结构进行了不同的变体,或是针对不同的问题加入了不同的结构.

2.1 将U-Net扩展到3D图像

生物医学影像是不同位置的切片构成的一组三维图,所以传统的 2D 图像处理模型处理 3D 的医学影像时会存在问题:一是效率不高,二是会丢失大量的上下文.针对这一问题,Ozgun Cicek 等人基于 U-Net 提出了 3D U-net[3],其网络结构如图 3 所示.3D U-net 输入输出是三维图像,提升了 U-Net 模型对三维图像的分割准确性.

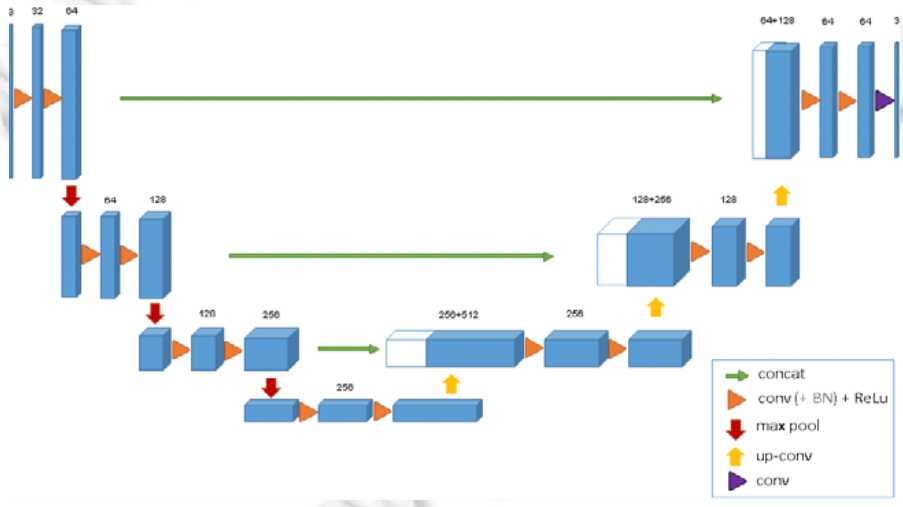


Fig.3 3D U-Net structure diagram[3]

图 3 3D U-Net 结构图[3]

另一个代表性工作是 Fausto Millemari 等人提出的 V-net[2],结构如图 4 所示.在输入 3D 图像按照通道拆分的同时,在每一层加入残差结构,以确保短时间收敛.降采样采用卷积操作替换最大池化操作,有利于在接下来的网络层中减小输入信号的尺寸的同时扩大特征感受野范围,并提出 Dice-based loss 这个新的损失函数.

© 中国科学院软件研究所 http://www.jos.org.cn

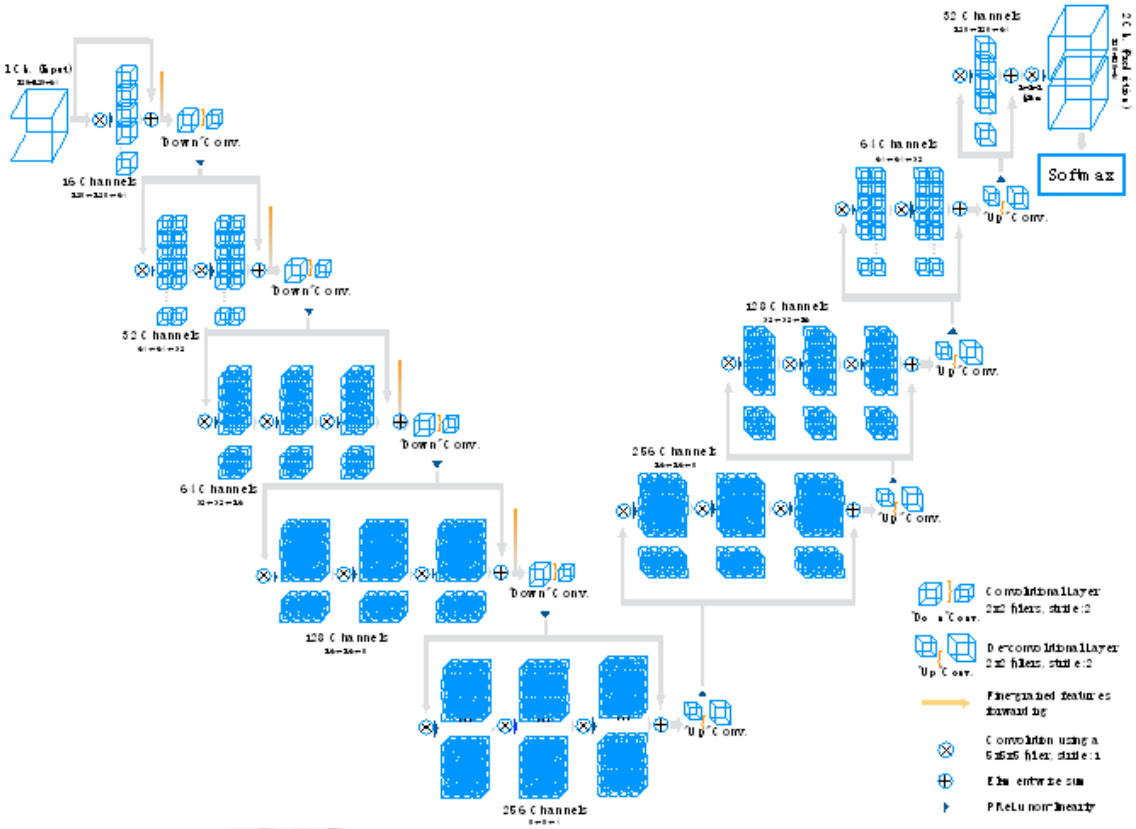


Fig.4 V-Net structure diagram[2]

图 4 V-Net 结构图[2]

2.2 针对增强相关特征,抑制无关特征的改进方

医学影像中,由于病变区的位置较其他无关特征更多,那么在分割过程中,聚焦目标特征、抑制无关特征就极为重要.一般在编解码器和瓶颈处加入 SE 或者 attention 模块,这两个模块都可以从空间和通道两个方面进行激励,以达到增强特征的效果.

2.2.1 加入 SE 模块

SE(squeeze-and-excitation)是在 2018 年 CVPR 上提出通过学习的方式来自动获取到每个特征通道的重要程度,然后依照这个重要程度去提升有用的特征,并抑制对当前任务用处不大的特征[59].Roy 等人[4]引入 3 个 SE 模型扩展结构分,别串联在 U-Net 的编码和解码结构中.

(1) 第 1 种是信道 SE(cSE),通过全局池化提取最能表现特征的通道,再将信息融合到原有的 tensor 中.

(2) 第 2 种是空间 SE(sSE),提取一张特征图划分特征区域,再将特征区域信息融合到原有的 tensor 中.

(3) 第 3 种同时进行空间和信道 SE(scSE),是 cSE 与 sSE 的合并输出.

实验结果表明:空间激励要比通道激励产生更高的对分割更为重要的增益;与标准的网络相比,scSE 虽然增加了一些计算复杂度,但是分割性能更好.

2.2.2 加入 attention 块

Attention 可以解释为将计算资源偏向信号最具信息性的部分的方法.一般在图像分割中,由于病灶较小且形状变化较大,常在 encoder 和 decoder 对应特征拼接之前,或是在 U-Net 的瓶颈处增加 attention 模块来减少假阳性预测.

© 中国科学院软件研究所 http://www.jos.org.cn

(1) 在 encoder 和 decoder 之间加入 attention 模块

Oktay O 等人在 2018 年提出的 Attention U-net[5]在 U-Net 在 encoder 和 decoder 中对应的特征进行拼接之前加入了一个集成注意力门(AGs),重新调整了 encoder 的输出特征.该模块将生成一个门控信号 g,用以消除不相关和嘈杂的歧义在跳过连接中的响应,以突出通过跳过连接传递的显着特征.attention 模块的内部结构如图 5 所示.

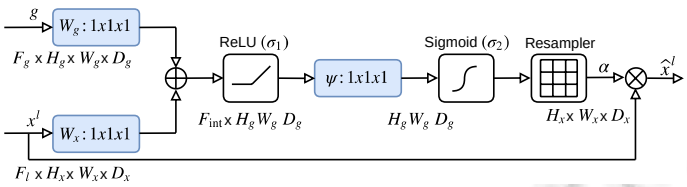


Fig.5 Internal structure of Attention module in Attention U-net[5]

图 5 Attention U-net 中 Attention 模块内部结构[5]

Li 等人在 2019 年提出了敏感连接注意力 U-net(CASU)[6],用于视网膜血管的精细分割.CASU 的网络结构与 Attention U-net 的网络结构相同,但是在 Attention 模块的结构上,CASU 采用不同的设计方式,如图 6 所示.G 是并行编码模块的输出,X 是前一个解码模块的输出.G 和 X 经过 Attention 门的处理后,再同 G 拼接.由于注意门的参数更新不仅取决于解码层传递的梯度,而且还取决于编码器层传递的梯度,其 AGs 最终采用的是可以提高训练过程中细节特征的质量和影响的 Up-link.实验结果表明,该方法能够有效地提高分割模型的注意权重.

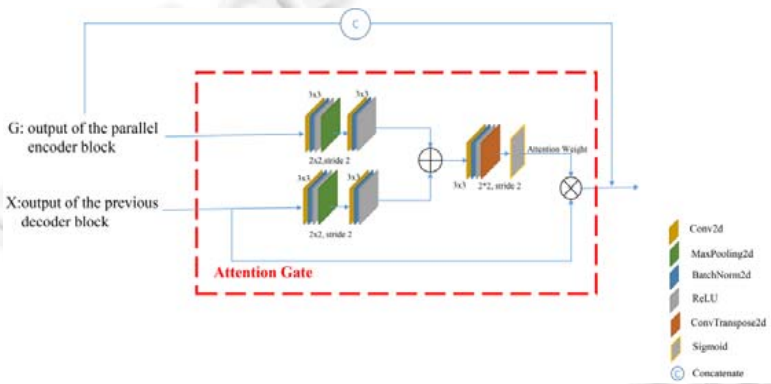


Fig.6 Structure diagram of Attention module in CASU[6]

图 6 CASU 中 Attention 模块的结构图[6]

Ni 等人在 2019 年提出的 RAUNet[7]加入了增强注意力模块(AAM)用于融合多层次特征和捕获上下文信息,来解决白内障手术器械分割中的镜面反射问题.RAUNet 的增强注意力模块(AAM)结构如图 7 所示.

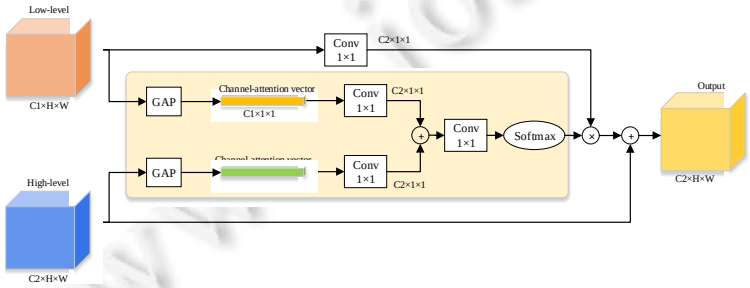


Fig.7 Structure diagram of AAM in RAUNet[7]

图 7 RAUNet 中 AAM 的结构图[7]

AAM 对语意依赖进行建模,以强调目标通道.其主要通过全局平均池化分别提取高层和低层的全局上下文信息和语义特征,并分别压缩成一个 attention 向量后对语意依赖项进行编码,突出关键特征并过滤背景信息.

Zhou 等人提出了轮廓感知信息聚合网络 CIA-Net[8],用于解决细胞核簇和不同器官形状的差异性的问题.CIA-Net 在编解码器之间建立多层次的横向连接,分层地充分利用金字塔特征,通过 encoder 早期层的纹理信息,可以帮助 Nuclei decoder 中分辨率低但具有强语义的层来细化细节,如图 8(a)所示.

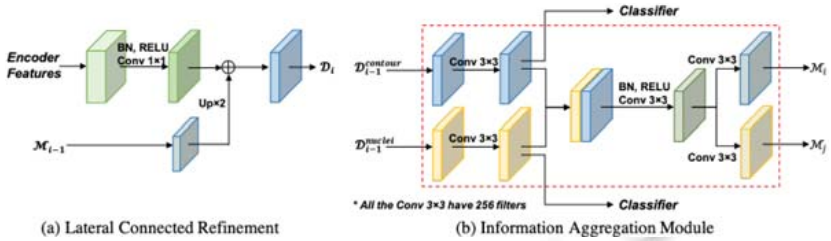


Fig.8 Structure diagram of IAM in CIA-Net[8]

图 8 CIA-Net 中 IAM 结构图[8]

CIA-Net 将 decoder 分为 Nuclei decoder 和 Contour decoder,两者之间加入信息聚合模块(IAM),将核信息和轮廓信息双向融合(如图 8(b)所示).此外,为了防止网络依赖于单一层次的区分特征,在每个阶段引入深度监控机制[60],加强对多层次上下文信息的学习,这也有利于通过缩短反向传播路径来训练更深层次的网络体系结构.

(2) 在 bottleneck 处加入 Attention 模块

bottleneck 是 U 型网络收缩路径和扩张路径中间的部分.Wang 等人提出的巩膜分割模型——ScleraSegNet[9]采用丢弃了全联接层的 VGG16 作为 encoder,瓶颈处增加了 bottleneck 模块用以编码最有区别的语义特征,其信息特征按照空间和信道进行分解,采用 4 个 attention 模块:(1) Channel attention module(CAM),由 SEnet[59]引出,结构图如图 9(b)所示;(2) Spatial attention module(SAM)[61],结构如图 9(c)所示;(3) Parallel channel attention and spatial attention module[61],将 CAM 和 SAM 并联后相加,结构如图 9(d)所示;(4) Sequential channel attention and spatial attention module[62]在平均池化层增加了最大池化层,然后将 CAM 和 SAM 模型顺序串联起来所得到,结构图如图 9(e)所示.经过在不同数据集上验证,作者提出了的这个方法在准确性上和泛化能力上都取得了显著的效果.

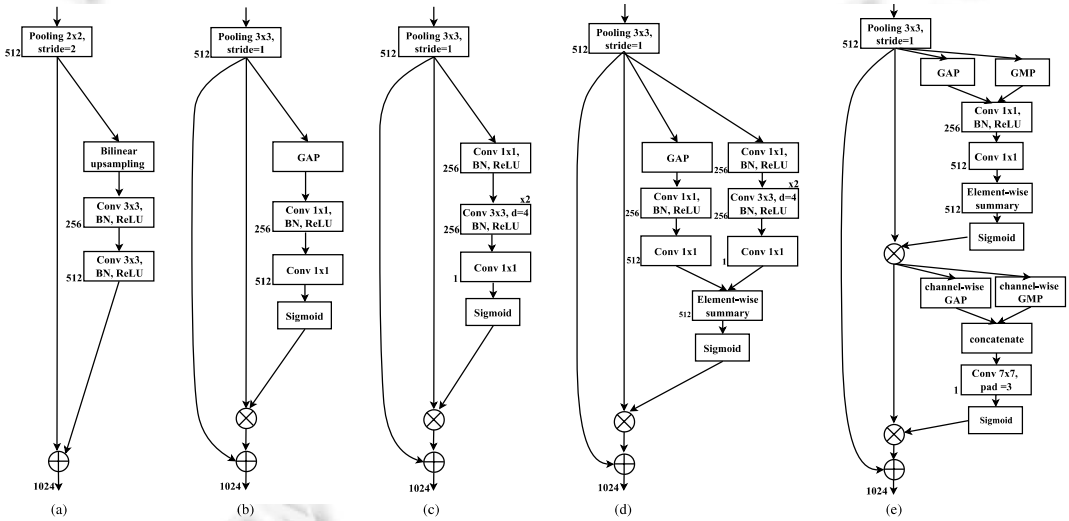


Fig.9 Internal structure diagram of Bottleneck in ScleraSegNet[9]

图 9 ScleraSegNet 的 bottleneck 内部结构图[9]

© 中国科学院软件研究所 http://www.jos.org.cn

(3) 在 decoder 中加入 attention 模块

DA 3D-UNet[10]在 3D Unet 的基础上将上采样替换成 DUpsampling[63],以提高解码器中图像的质量.在解码器的最后两层加入由空间 attention 和通道 attention 组合而成的双注意力模块,将大范围的、多通道的特征集中在关键位置、通道中.

(4) 在跳连中加入 attention 模块

Li 等人提出的 ANU-Net[11]在 Unet++的跳连中加入了 Attention Gate,以提升网络对于形状多样性病灶的分割效果.Attention Gate 的输入分为两个部分:解码器的上采样特征(g)和编码器中相应的特征(f),其上采样特征(g)作为门信号来增强编码器中相应的深度特征,从而增强相关特征、抑制无关特征,结构图如图 10 所示.

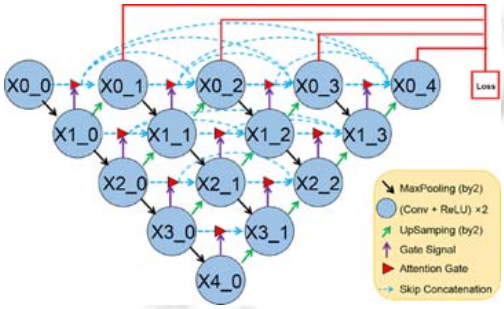


Fig.10 ANU-Net structure diagram[11]

图 10 ANU-Net 的结构图[11]

(5) 在编解码器单元中加入 attention 模块

徐宏伟等人提出的 RDA-Unet[12]采用残差双注意力(RAD)模块作为编解码器单元,RAD 将通道注意力机制和空间注意力机制相结合,保证特征提取的准确性,并加入残差结构以防止梯度消失,更好地定位肾脏的边界.

2.2.3 其他方法

Liu 等人在 U-Net 的基础上加深网络层数至 7 层,并将编码器的池化层直接与解码器对应的层级联,以减少信息的损失,并称此网络为 IU-Net[64].由于分割复杂的肝脏切片容易产生低质量的分割,其采用图割算法[65],在前景和背景选取种子点,建立一个图,利用最大流算法找到加权图的最小割集,最终得到分割较好的图像.他们将 IU-Net 和图割相结合的网络最终命名为 GIU-Net[64].Mu-net[13]将经过下采样后不同尺度的特征图分别输入 U-Net,再将不同尺度的 U-Net 的输出经过上采样到上一层 U-net 的输入,帮助上一层减少对于低频信息的计算,从而更加聚焦于病灶的分割.

2.3 针对内存占用、计算速度的改进方法

由于 2D 卷积容易丢失上下文信息,而 3D 卷积占用 CPU 量过大,为了减少内存,一种方法是扩大 patch 中的 volume,另一种方法是用较小的 batch size 训练.这些方法毕竟是有局限性的,因而以下几个工作改进网络中部分模块,减少内存,提升运算速度.

2.3.1 加入稠密卷积块

Li 等人于 2018 年提出由有效提取切片的 2D Dense Net 和提取肝脏病灶分割中上下文信息的 3D Dense Net 组成 H-DenseUNet[14],即稠密融合 U 型网络.该网络先采用 ResNet 粗略的分割肝脏图像,然后在感兴趣区域(ROI)中,利用 2D Dense Net 和 3D Dense Net[66,67]有效探测切片内和切片间的特征.在 H-DenseUNet 的结构上,MMMDF[15]将 2D/3D DenseNet 替换成多模态 2D-ResUNet 和 3D-ResUNet,以 2D 网络的快速分割结果来指导 3D 模型的学习并实施分割.

另一方面,光声成像(PAT)测量用的声波经过稀疏采样后可以用于图像重建,但会导致较为严重的图像信息缺失[68].Steven Guan 针对这一问题提出一种全密集连接的 FD-UNet[16],用于重建稀疏采样的 2D PAT 图像,其基

© 中国科学院软件研究所 http://www.jos.org.cn

于 U-Net,在编解码器引入 Dense connectivity 密集连接,避免了冗余特征的学习,增强了信息流动,在性能接近的前提下进一步减少了网络参数,降低了计算成本,进行图像重建时可更加快捷(如图 11 所示).

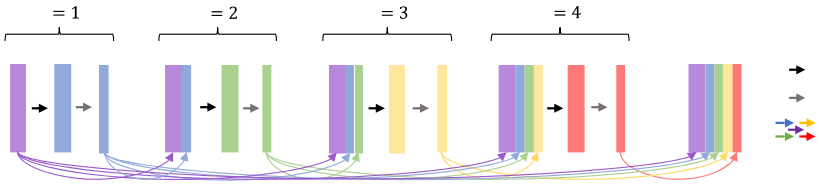


Fig.11 2D dense connection in FD-Net[16]

图 11 FD-Net 中 2D 稠密连接[16]

2.3.2 加入可逆结构

Robin 等人提出了 Partially Reversible U-Net[17],将 U-Net 的编解码器每个单元采用可逆序列[69],同时使用传统的不可逆操作来进行下采样和上采样以及跳跃连接.这种完全可逆的体系结构比传统的 U-Net 节省了大量的内存,因为激活只需要在每个可逆序列的末尾和不可逆的组件上保存.

2.3.3 加入 SE 残差块

Zhu 等人提出的 AnatomyNet[46](如图 12 所示)以端到端的方式联合分割所有 organs-at-risks(OARs),接收一个原始的全容积 CT 图像作为输入,并将所有 OARs 的掩模与图像一起返回.该结构与 U-Net 的不同在于:只在第 1 个编码块中采样了下采样层,使得下一层中的特征映射和梯度比其他网络结构占用更少的 GPU 内存.移除第 2~第 4 个编码器块中的下采样层,采用了 SE 残差块[59]学习有效特征,以提升分割小解剖结构的性能.

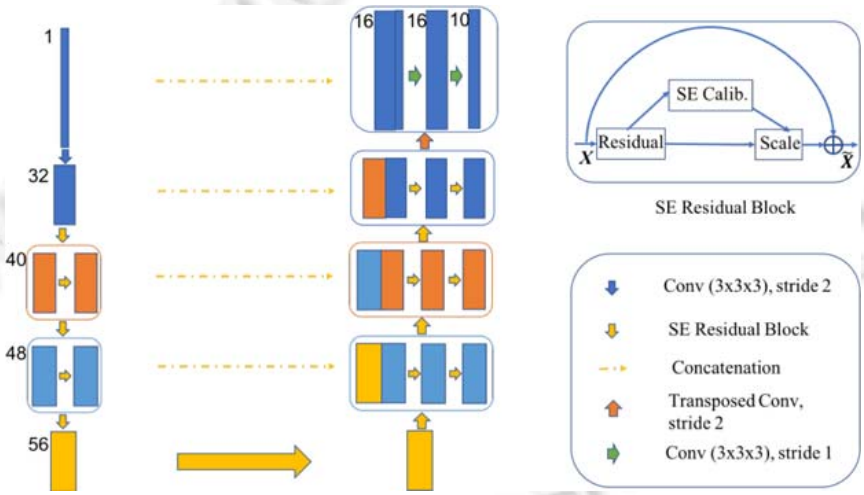


Fig.12 Anatomynet structure diagram. The first layer is down sampling, and the rest is replaced by SE residual block[46]

图 12 AnatomyNet 结构图.第 1 层是下采样,其余由 SE residual block 代替[46]

2.3.4 通道分组

Chen 等人在 2019 年为弥补三维 MRI 脑肿瘤分割模型效率和准确性不可共存的问题,提出一种新的三维扩张多纤维网络(DMFNet)[18].DMFNet 建立在多光纤单元(MF)[70]的基础上,利用有效的群卷积,引入加权的三维扩展卷积运算,获得多尺度的图像分割表示,从而减少参数以提升运算效率.

图 13(a)、图 13(b)采用通道分组思想,将循环通道分成多个组,这样可以减少特征映射和核心之间的连接,从而显著地节省参数.而图 13(c)中的 Multiplexer 主要是用于不同的 fiber 之间交换信息.在图 13(d)中增加了扩张卷积,这种加权求和策略可以从不同视角自动选择有价值的信息.

© 中国科学院软件研究所 http://www.jos.org.cn

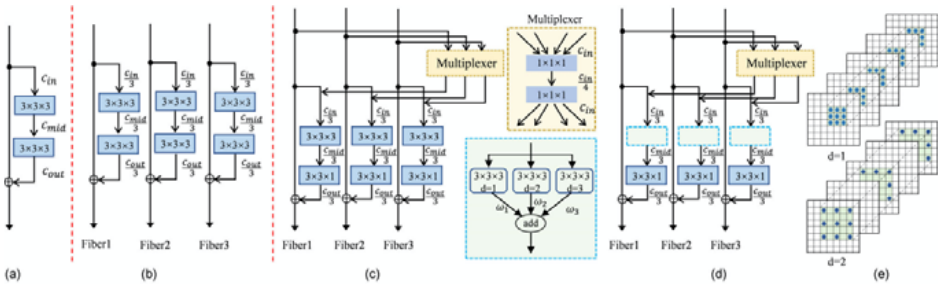
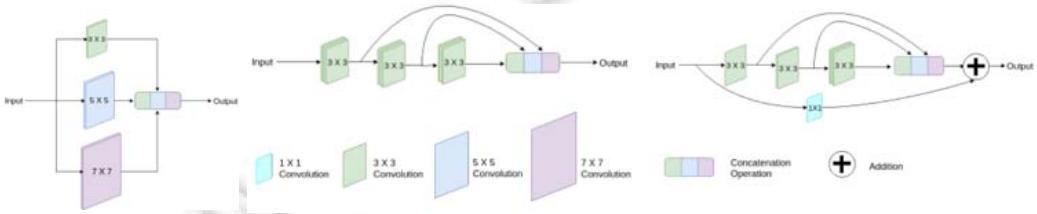


Fig.13 MF unit of dilated multi-fiber net (DMFnet)[18]

图 13 DMFnet 多纤维网络的 MF 单元[18]

2.3.5 加入 Inception 块

Nabil 等人在 MultiResUNet[19]中提出了 MultiRes 模块代替 U-Net 的每一层解决多分辨率分析同时又节省内存、提高速度,其将 U-Net 编解码器的每个单元替换为 MultiRes 模块.受 inception[71]启发,作者先将大小为 3×3,5×5,7×7 的卷积层并联,以实现多层分辨率的分析(如图 14(a)所示),然后采用更小、更轻量级的 3×3 卷积块近似代替 5×5,7×7 的卷积操作以减少存储(如图 14(b)所示),再将 filter 的个数从 1 逐渐增加 3 来减少前一层滤波器数量带来的二次效应[72],并且增加 1×1 卷积层的残差连接以更好地保证空间信息,最后构成 MultiRes 模块.



(a) (b) (c)

Fig.14 MultiRes model of MultiResUnet[19]

图 14 MultiResUnet 中 MultiRes 模块[19]

DENSE-Inception U-net[20]将 inception 和残差模块以及稠密模块相结合,如图 15 所示,其中,采用 Inception-Res 取代标准的卷积来增加网络宽度;采用 Dense-Inception 模块,在增加网络深度的同时,又不会增加参数的数量;上下采样采用 Inception 模块,保证图像分割的准确性.其在肺部图像分割和血管图像分割、脑肿瘤图像分割方面都有很好的性能.

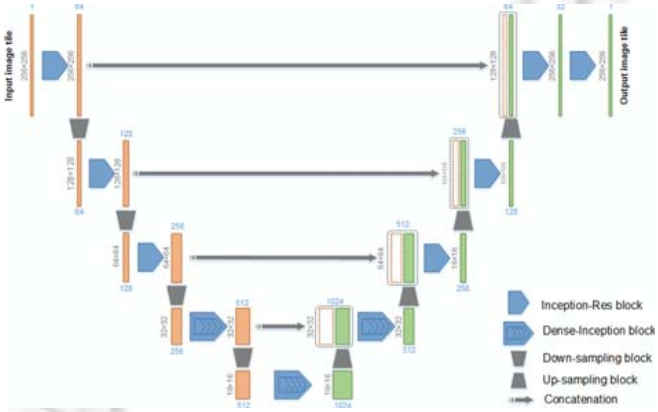


Fig.15 DENSE-INception U-net structure diagram[20]

图 15 DENSE-INception U-net 结构图[20]

2.3.6 其他方法

Li 等人提出了 PBR-Unet[21],主要包括提取像素级概率图的功能提取模块和用于精细分割的双向递归模块,如图 16 所示.用 2D Unet 提取概率图,用于指导精细分割;双向递归模块将上下文信息集成到整个网络中,避免了传播过程中空间信息的丢失,从而节省内存.徐等人提出了基于级联 Vnet-S[22]的单一器官分割法,在 V-net 的基础上减少 V-net 的编解码器的卷积单元,减小卷积核的大小,在跳连中加入 Dropout 缓解过拟合,以减少 3D 卷积带来内存占用问题,提升运算速度.

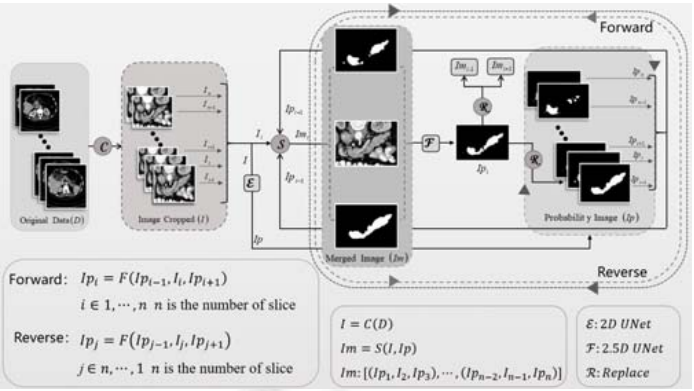


Fig.16 PBR-Unet structure diagram[21]

图 16 PBR-Unet 结构图[21]

2.4 针对特征融合的改进方法

特征融合其实更多意义上讲的是网络的上下文特征的融合、不同模态特征的融合.上下文特征的融合可以从编解码器中加入新的模块 DAC 和 RMP 帮助融合信息,如 CE-Net[73];也可以在跳连阶段增加编解码器信息的流动,如 MultiResUNet[19]、Unet++[23]或者去掉跳连、增加信息聚合的 DFA-Net[24];或者外接特征金字塔从不同分辨率角度保证分割的准确性,如 MFP-Unet[25].对于不同模态的融合,可将编码器分别提取各模态之间的信息再进行融合,如深度级联脑肿瘤分割方法[26]、Dense Multi-path U-Net[27]、IVD-Net[28].

2.4.1 上下文特征的融合

(1) 编解码器加入新的模块

为了获取更高层次的信息,并保留 2D 医学影像分割中的医学信息,Gu 等人在编解码器的基础上加入上下文提取模块,从而构成新的网络 CE-Net[73],整体框架如图 17 所示.

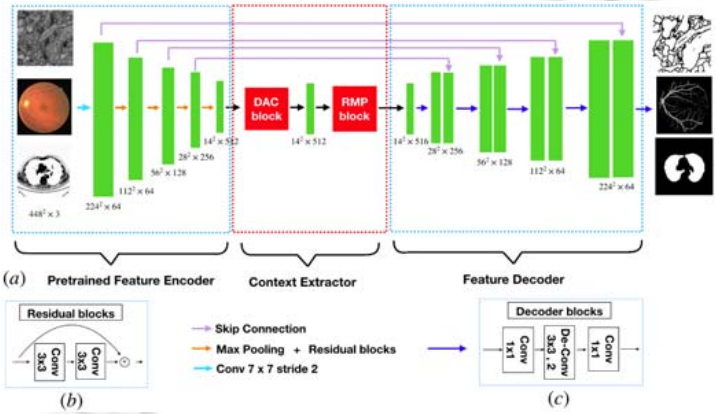


Fig.17 CE-Unet network structure diagram[73]

图 17 CE-Unet 网络结构图[73]

© 中国科学院软件研究所 http://www.jos.org.cn

上下文提取模块主要包括 DAC(dense atrous convolution module)模块和 RMP 模块(residual multi-kernel pooling).作者受 inception-ResNet-V2[74]模块和扩张卷积启发提出了 DAC 模块,以编码高层级的语义特征图.RMP 模块采用残差多核池化方法,主要依靠多核有效视场来检测不同大小的目标,解决医学影像中物体尺寸的巨大变化.

(2) 在跳连处改进

U-Net 的跳连结构主要是融合上下文的语义特征,以更好地分割病灶。但是简单的级联使得高层级和低层级的语义信息融合容易造成重要语意丢失,因而针对这一问题,相关工作提出了多种改进特征融合的方法.Nabil 等人在 MultiResUNet[19]中提出:由一系列带有残差连接的卷积层构成的 Res path 取代 U-Net 的级联,使低级特征经过进一步处理再与高级特征级联,以消除编码器的低级特征和解码器的高级特征融合时造成的语义差异(如图 18 所示).Zhou 等人从另一方面对跳连进行改进,提出带有深度监控的嵌套的密集跳连路径的 Unet++[23],结构如图 19 所示.

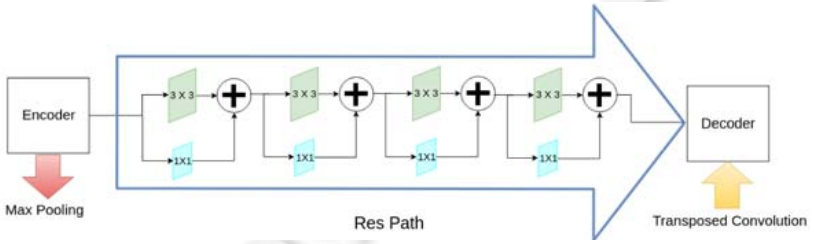


Fig.18 Skip connection of MultiResUnet[19]

图 18 MultiResUnet 的跳连[19]

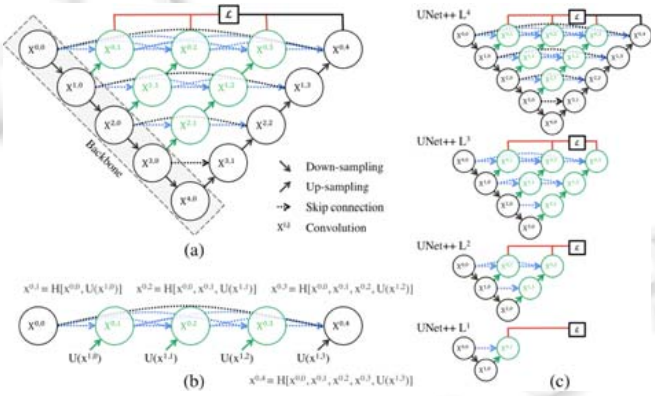


Fig.19 UNet++ network structure diagram[23]

图 19 UNet++的网络结构图[23]

Jin 等人在跳连处加入 attention 机制,提出了三维混合残差注意感知分割网络 RA-UNet[29],用于精确提取肝脏兴趣体积(VOI),并从肝脏 VOI 中分割肿瘤.该网络在 3D U-net 的基础上,除了第 1 层和最后一层外,其他都由残差结构堆叠,以实现增加深度而不会产生梯度爆炸,并在跳连处加入 Wang 等人提出的残差注意力模型[75],分为用于处理原始特征的主干分支、用于增强特征抑制噪声的软掩膜分支.

杨兵等人所提出的深度特征聚合网络 DFA-Net[24](如图 20 所示)直接去掉 U-Net 跳连并称其为基础层,加入中间层、聚合层以及特征聚合模块 FAM,帮助更好地融合上下文信息.Nikhil 等人提出的 U-Det[30]将 U-Net 的跳连替换为 Bi-FPN[76],如图 21 所示.其拥有自顶向下和自底向上的路径的同时,每个节点为不同的输入加入不同权重,以强调不同输入的重要性.而 Li 所提出的 DPSN[77]在跳连处采用特征金子塔,高度提取抽象编码器的特征之后,再通过跳连和解码器进行级联,将高层语义特征和低层语义特征更好的融合.

© 中国科学院软件研究所 http://www.jos.org.cn

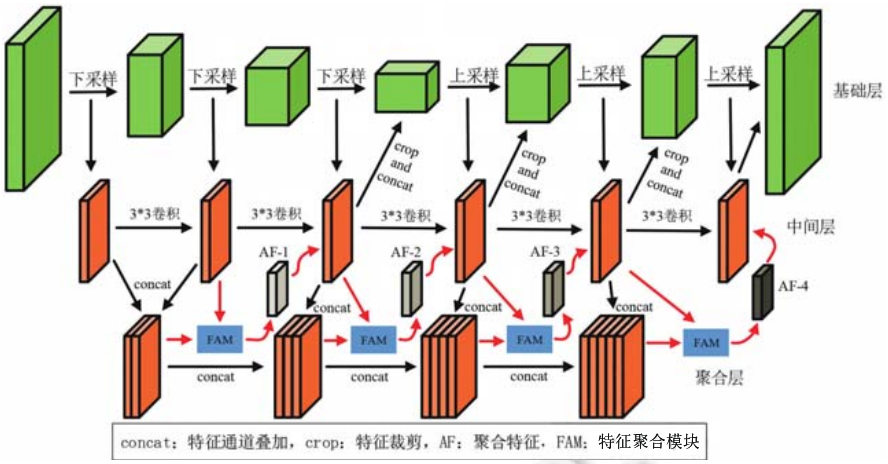


Fig.20 DFA-Net network structure diagram[24]

图 20 DFA-Net 网络结构图[24]

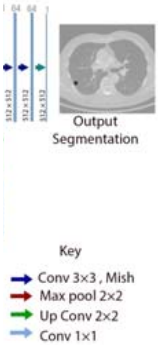


Fig.21 U-Det network structure diagram[30]

图 21 U-Det 网络结构图[30]

(3) 外接特征金字塔

Moradi 等人提出了 MFP-Unet[25],该网络在 U-Net 基础上外接特征金字塔网络 FPN[78],从扩展路径的各个层次提取特征,最后将提取的特征串联,形成 64 通道的最终特征映射,并传给用于特征分类滤波器中,以提升语义对分割的贡献.

2.4.2 不同模态的特征融合

传统的 U-Net 网络对于多模态图像输入,采用先混合处理再输入的方式,这样操作容易丢失不同模态的部分信息.Lachinov 等人[26]为解决这一问题,提出一种深度级联脑肿瘤分割方法,主要是将编码器并行分出几个路径分别学习不同模态的特征表示,在跳连和 bottleneck 处采用像素最大化操作再与解码器级联,网络结构如图 22 所示.

而 Dolz 等人提出的 Dense Multi-path U-Net[27]在编码器多路径的基础上加入 Dense Net 的思想,以解决缺血性中风病灶的位置和形状的高度变异性.此方法首先将输入端图像混合的方式,变成在不同路径中对每个模式进行处理,以更好地利用其独特的信息,如图 23 所示;然后在不同模态之间建立稠密连接,改善数据流,减轻梯度消失;并且扩展了非对称 inception 卷积块,代替最大池化操作,其多扩张率的卷积操作,从不同尺度上提取特征,更好的捕获上下文信息.同年,Jose Dolz 等人在 Dense Multi-path U-Net[27]的基础上提升编码器的多路径稠密性,提出了 IVD-Unet[28],主要对椎间盘(IVD)图像进行分割.

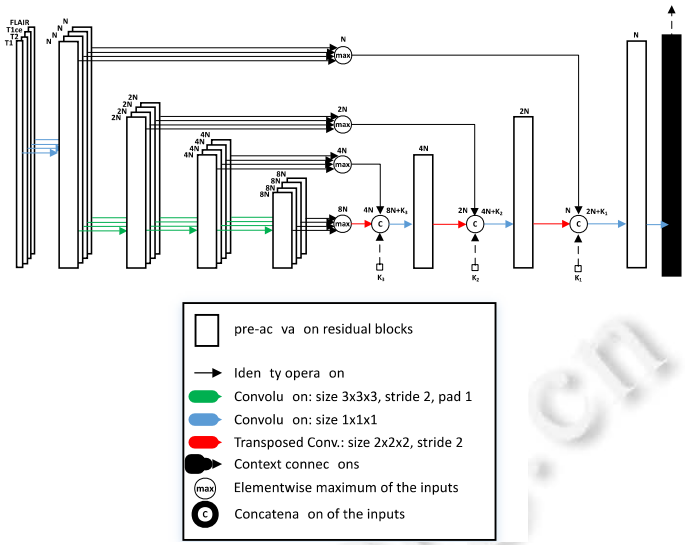


Fig.22 Gliomas segmentation and cascaded U-Net network structure diagram[26]

图 22 Glioma 分割与级联 U-Net 网络结构图[26]

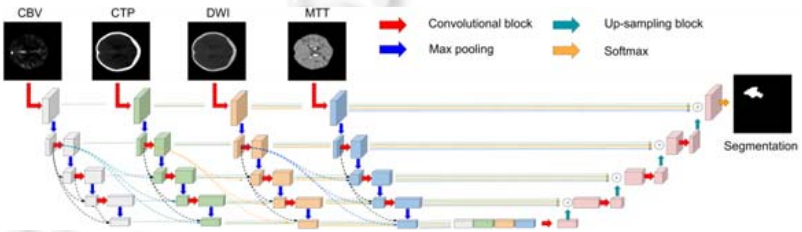


Fig.23 Dense Multi-path U-Net network structure diagram[27]

图 23 Dense Multi-path U-Net 网络结构图[27]

2.5 针对提高泛化能力的改进方法

在临床实践中,医学影像是从不同的供应商处获取,从特定的源域训练的 U-Net 再传输到不同的目标域时,性能会急剧下降.泛化能力是指网络可以混合训练来自不同提供商的图像,具体方法有两种:从内部结构提出适应不同提供商图像的域适配器,如 3D U^2-net[35]采用在编解码器单元内采用 Adapter 找到合适网络训练的卷积,从而适应不同提供商提供的图像;从外部接入 Cycle-GAN 网络,如 Yan 等人提出的 Unet-GAN[36],包括一个用于适应供应商的非配对生成对抗网络(CycleGAN)[79],一个用于对象分割的 LV-Unet.在图 24 中,LV-Unet 是由数据集 S 训练的满足 LV(左心房)分割的分割网络;CycleGAN 是一个为未配对的图像到图像转换而设计的既定架构,包括两个生成器 G_S 和 G_T,代表源域和目标域;两个辨别器 D_S 和 D_T,辨别是原始图像还是转换图像.

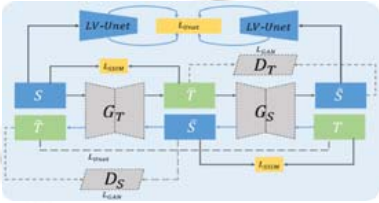


Fig.24 U-netGAN network structure diagram[36]

图 24 U-netGAN 网络结构图[36]

© 中国科学院软件研究所 http://www.jos.org.cn

2.6 针对小样本训练数据集的改进

医学影像由于涉及隐私问题和标注成本高的问题,其数据集数量极少.针对小样本训练数据集的问题,一般从重复的网络结构和数据集标签两个方面进行.

- 重复网络结构可以在 bottleneck 重复使用 SRU 门控单元[31],也可以重复使用整个 U-Net 如 Bridged U-net[32].
- 从数据本身标签入手,可以结合贝叶斯训练给定数据标签再进行网络训练,提升分割准确性;或者将一幅图像仅使用一个单一全局标签,以降低对数据量的要求.

2.6.1 重复网络结构

Wang 等人[31]提出在 U-Net 的瓶颈处加入重复单元结构:双门控递归单元(DRU)或单门控递归单元(SRU),可以在数据集和计算能力有限的情况下进行训练.DRU 在 GRU[80]上进行改进,能有效地细化迭代分割,但浪费内存,因而提出简化成单门控的 SRU 代替 DRU,其精度并未变化.

Chen 等人提出了 Bridged U-net[32],采用 U-Net 桥连接的方法,在多个层次上充分利用不同的特征,加速神经网络的收敛.网络结构如图 25 所示.两个 U-Net 之间的桥连接采用级联,U-Net 的跳连采用加法,可以达到网络的最好表现形式.激活函数采用 ELU 和 ReLU 相结合,解决了单纯使用 ELU 的随着网络不断深入的饱和问题.

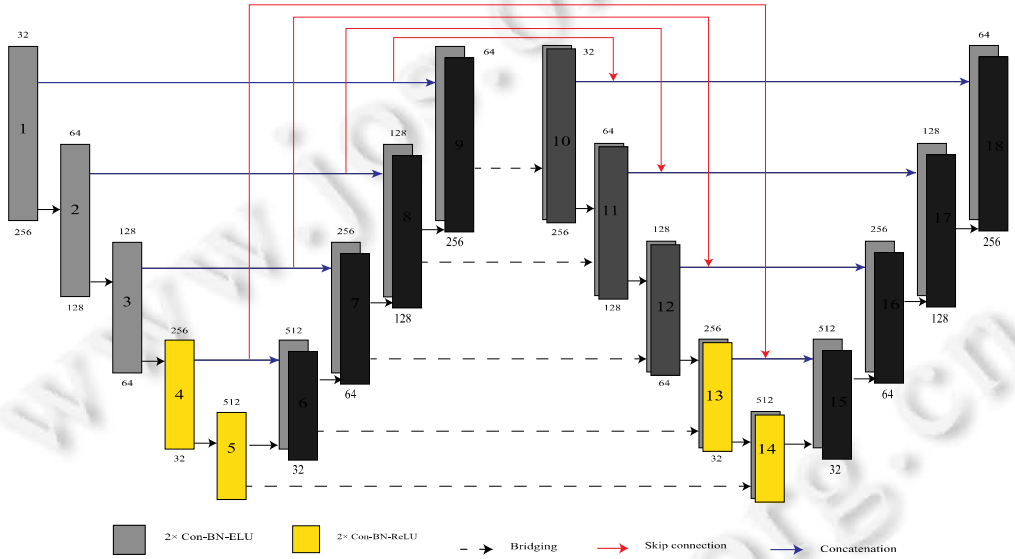


Fig.25 Bridged U-net network structure diagram[32]

图 25 Bridged U-net 的网络结构图[32]

2.6.2 计算数据标签

U2-NET[33]提出一个具有认知不确定性反馈的 BAYESIAN U-NET 模型,用于病理 OCT 扫描中光感受器层的分割.通过贝叶斯对于给定数据和标签进行后验概率计算,再通过 U-Net 进行训练.

Florian Dubost 提出了 GP-Unet[34],用弱标签来检测病灶的卷积神经网络,也就是每幅图像只需要一个单一的全局标签“病变计数”就可以训练,网络结构如图 26 所示.GP-Unet 是一个具有完全卷积结构的回归网络,结合一个全局池层,将 3D 输出聚合成一个指示病变数量的标量.在测试时,GP-Unet 首先运行网络来估计病变的数量,再移除全局池层来计算输入图像大小的定位图.

© 中国科学院软件研究所 http://www.jos.org.cn

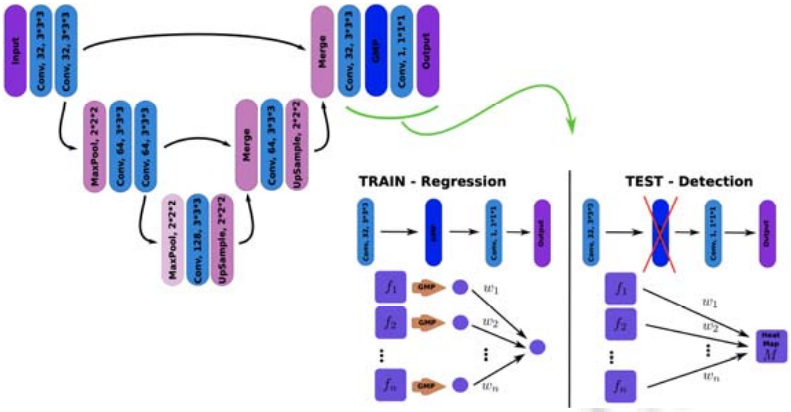


Fig.26 GP-Unet network structure diagram[34]

图 26 GP-Unet 网络结构图[34]

3 基于 U-Net 面向结构模块的改进

本节对于 U-Net 面向结构模块的改进作出总结,现有工作主要在编解码器、bottleneck 和损失函数改进、数据流增强以及自动搜索结构几方面对 U-Net 提出各种改进型研究.这些工作对网络结构进行了不同的变体,或是针对不同的问题加入了不同的结构.

3.1 针对编码器解码器结构的改进方法

编、解码器的改进可分为 3 部分:一是卷积操作的改进,如加入协调引导卷积、长短残差结构;二是编解码器单元的改进,如可以可形变卷积块、循环残差卷积和概率模块;三是上、下采样的改进,如可以采用长短残差结构和最近邻插值的方法.

3.1.1 卷积操作改进

肺叶的鉴别和诊断对疾病的诊断和治疗具有重要意义,少数肺病在肺叶有区域性的病变,准确分割肺叶极为重要,Wang 等人[37]提出一种基于利用协调引导卷积的深度神经网络,从胸部 CT 图像中自动分割肺叶的方法.其首先采用自动肺分割方法提取 CT 图像中的肺面积,然后利用 V-net 对肺叶进行分割.协调卷积部分结构图如图 27 所示.

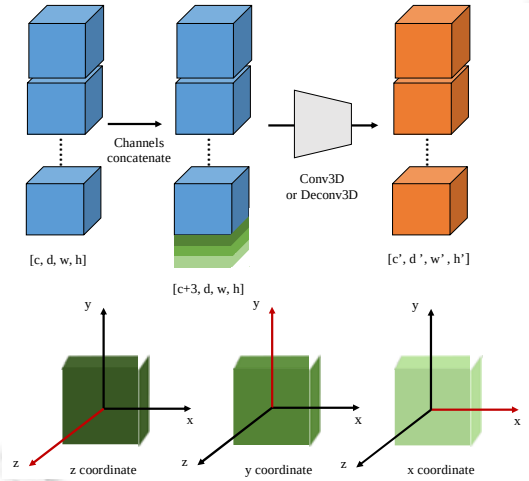


Fig.27 Coordination-guided convolution proposed by Wang, *et al*.[37]

图 27 Wang 等人提出的协调卷积[37]

为了减少不同肺叶的错误分类,文中采用协调引导卷积(CoordConvs)[81]来生成肺叶位置信息的附加特征图.CoordConv 是对经典卷积层的简单扩展,通过添加额外的坐标通道来集成位置信息.

对于模态的急性亚急性脑中风病灶 MRI 图像分割,Albert 等人[38]提出在采样过程中平衡患者和健康的人的 MRI 图像采样,并且在 U-Net 的网络结构中加入长短残差结构代替卷积操作和下采样操作在保证精度的同时减少参数,改进结构如图 28 所示.

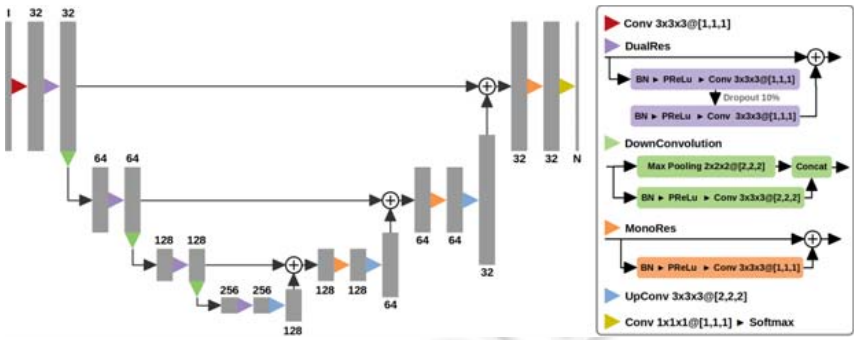


Fig.28 Network structure diagram proposed by Albert, *et al*.[38]

图 28 Albert 等人提出的网络结构图[38]

3.1.2 编、解码器单元改进

Jin 等人提出 DUNet[39],其是在 U-Net 的框架的基础上,用可变形卷积块[82]作为编码器、解码器的每一个单元.可变形卷积块通过学习局部、密集和自适应的感受野来模拟不同形状和尺度的视网膜血管,以达到准确分割.具体是在标准卷积使用的网格采样位置上添加偏移量,而偏移量是从附加卷积层生成的先前特征映射中学习的.因此,变形能够适应不同的尺度、形状、方向等.图 29 给出可变形卷积与普通卷积方法差异的示意图.

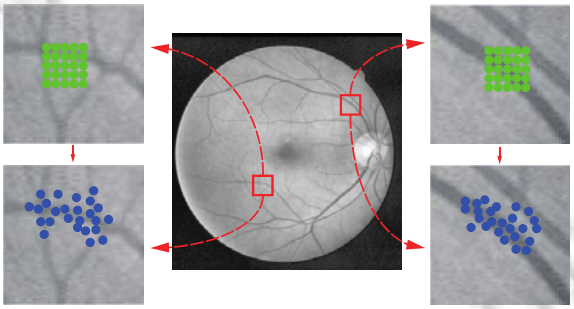


Fig.29 Comparison of deformable convolution and normal convolution in DUNet[39]

图 29 DUNet 中可变形卷积和正常卷积对比[39]

蒋等人提出的 I-Unet[40]在 U-Net 的基础上改进编解码器单元,编码器采用由扩张卷积、inception 和 RCL 层组成的 Conv-Block,解码器采用反卷积、RCL 层组成的 Deconv-Block,通过扩大感受野进行多尺度特征融合.

何承恩等人基于 3D-Unet 提出了 3D-HDC-Unet[41],在编码器的每个单元中加入混合膨胀卷积残差块,以不断变化的膨胀率改变棋盘效应[83]给分割带来的负面影响.

Alom 等人提出了 R2U-Net[42],该方法将残差连接和循环卷积结合起来,用于替换 U-Net 中原来的子模块,其改进结构如图 30 所示,图中环形箭头表示循环连接.图 31 展示了几种不同的子模块内部结构图.该方法保证网络深度的同时,减轻梯度消失的影响,在提取低级特征有显著效果,多应用于视网膜血管分割.

Kohl 等人提出 HPU-net[43],一个结合 U-Net 和条件变分自动编码器(cVAE)的能够考虑多尺度变化的层次概率分割网络,网络结构如图 32 所示.该网络分为采样过程和训练过程,在采样过程中,解码器额外对延迟的空间网格采样.在训练过程中,采用条件概率分布对网络进行训练.

© 中国科学院软件研究所 http://www.jos.org.cn

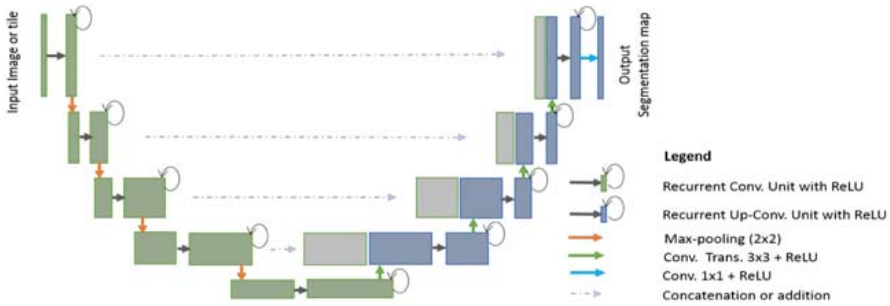
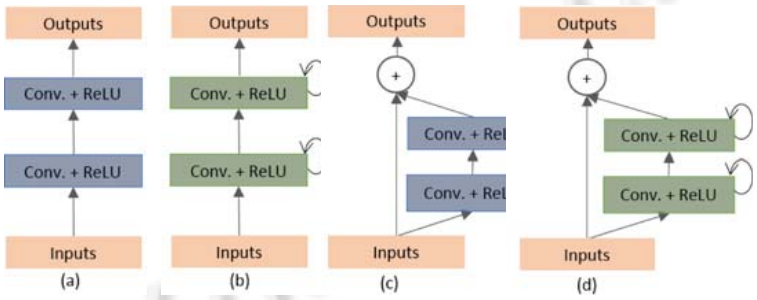


Fig.30 R2U-Net network structure diagram[68]

图 30 R2U-Net 网络结构图[68]



(a) 常规的 U-Net 中使用的方法;(b) 在(a)的基础上使用循环卷积代替原有卷积;
(c) 使用残差连接的方式;(d) 该文章提出的结合(b)和(c)的循环残差卷积模块

Fig.31 Structure diagram of Recurrent convolution in R2U-Net[42]

图 31 R2U-Net 中循环卷积结构图[42]

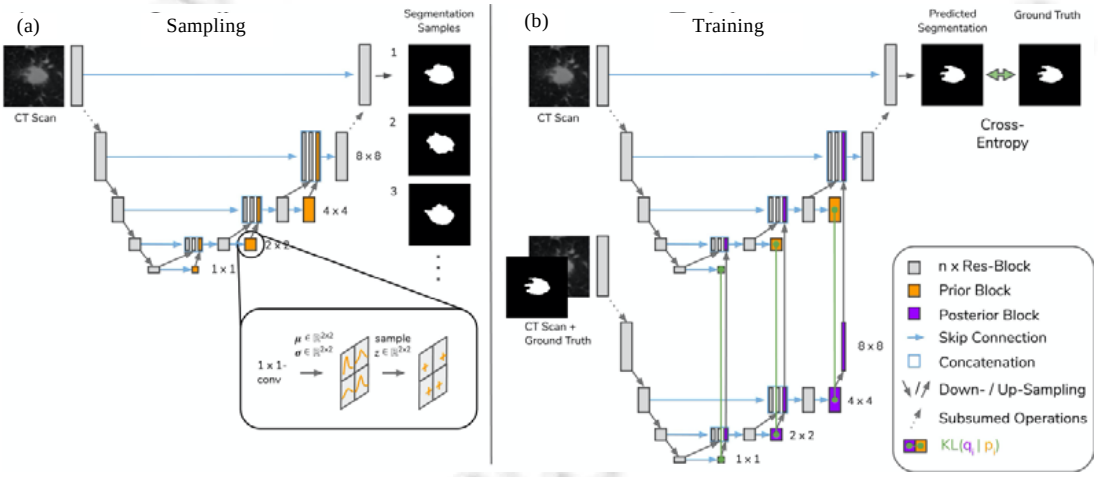


Fig.32 HPU-net network structure diagram[43]

图 32 HPU-net 网络结构图[43]

3.1.3 上、下采样改进

在微创手术中,准确地追踪到手术器械的位置是十分重要的.针对内窥镜图像中的分割和识别外科器械问题,Hasan 等人[44]为缓解转置卷积导致“不均匀重叠”也就是棋盘格形状的伪影问题,提出了 U-NetPlus 网络结构,将 VGG-11 和 VGG-16[84]作为编码器这种预先训练的编码器[85],通过规避与目标数据相关联的优化挑战,加快

了收敛速度[86].

Wang 等人提出了 Non-local U-Nets[45],在 U-Net 的基础上,对于输入输出模块采用残差结构,对于上下采样采用外部嵌套残差结构的全局聚合模块,从而减弱单一卷积操作所带来的信息丢失问题.

3.2 基于损失函数的改进方法

神经网络训练过程中,使用损失函数计算每次迭代的结果与真实值之间的差距,从而指导下一步训练向正确的方向进行.损失函数改进主要解决的是类不平衡的问题,主要是从函数自身和两个损失函数相结合两个方面进行改进.

3.2.1 函数自身改进

Dice loss 函数的一个局限性在于 *FP* 与 *FN* 的检测权重相等,这将导致分割图有较高的准确率和较低的召回率.像在皮肤病变者众数据极不平衡,感兴趣区域极小,*FN* 需要比 *FP* 高很多才能提高召回率.V-net[2]中提出了一个基于 dice coefficient 的损失函数,也就是对分割求偏导数,从而不需要为不同类别分配权重,就可以建立前景、背景平衡.

$$D=\frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2+\sum_i^N g_i^2} \tag{16}$$

$$\frac{\partial D}{\partial p_j}=2\left[\frac{g_j\left(\sum_i^N p_i^2+\sum_i^N g_i^2\right)-2p_j\left(\sum_i^N p_i g_i\right)}{\left(\sum_i^N p_i^2+\sum_i^N g_i^2\right)^2}\right] \tag{17}$$

对于 Dice loss 的预测接近真实情况时效果不佳引起震荡的问题,Chen WL 等人提出一个新的损失函数 Cos-Dice[32]损失函数,来加速学习进程.

$$L_{CosDice}=\cos^Q\left(\frac{\pi}{2}\cdot DSC\right),\ Q>1 \tag{18}$$

3.2.2 两个函数的混合

针对医学数据类不平衡的问题,Abraham 等人基于 Tversky index 提出了一个广义的损失函数 Focal Tversky Loss(FTL)[47].与 Dice loss 函数相比,这个函数在训练较小结构可以更好地权衡准确率与召回率之间的关系(其中,c 为类别,TI_c 为 Tversky index).

$$FTL_c=\sum_c(1-TL_c)^{\frac{1}{\gamma}} \tag{19}$$

而 AnatomyNet[46]和 3D-HDC-Unet[41]采用 Dice 系数和 Focal loss 相结合的方式解决这一问题.

$$L=L_{Dice}+\lambda L_{Focal} \tag{20}$$

实验结果表明,λ=0.5 时效果最好.

RAUNet[7]提出的 Cross Entropy Log Dice(CEL-Dice)结合了交叉熵的稳定性和类不平衡不影响 Dice loss 的特性,因此,它有比 Dice loss 更好的稳定性,比交叉熵更好地解决类不平衡的问题(H 为交叉熵,D 为 Dice loss).

$$L=(1-\alpha)H-\alpha\log(D) \tag{21}$$

Zhong 等人[48]提出了交叉熵和 Dice loss 损失组合新的形式:

$$L=(1-\alpha)H+\alpha D \tag{22}$$

3.3 针对增加数据流路径的改进方法

针对于数据流的改进主要是从两个方面:一是采用 DenseNet 的思想,增加网络中不同模块之间的连接;二是将 U-Net 网络串行使用两次,也就是桥连接,从而达到信息成倍数流通的目的.

U-Net 的各种变体都包含编码器和解码器,但是对于数据流路径数量是有限的,Zhang 等人提出的 MDU-Net[51]将 DenseNet 的思想应用于编、解码器、跳连中,直接融合高层和低层相邻的不同比例尺的特征映射,增强当前层的特征传播.这在很大程度上提高了信息流的编解码能力(如图 33 所示).

© 中国科学院软件研究所 http://www.jos.org.cn

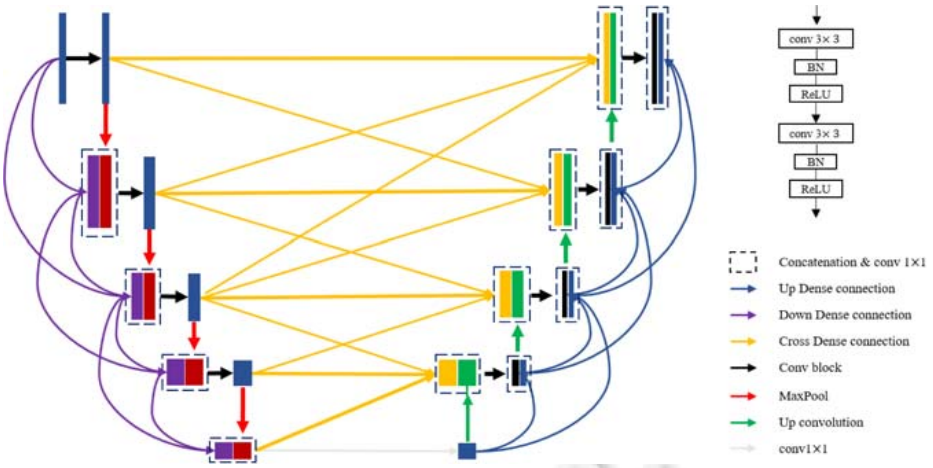


Fig.33 MDU-Net network structure diagram[51]

图 33 MDU-Net 的网络结构图[51]

Zhuang 等人提出的 LadderNet[49],其结构类似于桥连接,与之最大区别在于两点:其一,LadderNet 用加法取代 U-Net 中跳连采用的级联,两个并行 U-Net 对应层也采用加法;其二,LadderNet 采用了一个新的共享权重残差块(如图 34 所示),解决了多编码器解码器分支来的参数增多、训练难度增加的问题.这个共享权重残差块由 3 个部分组成:跳连、递归卷积以及 dropout 正则化.其在同一块中的两个卷积层共享参数可以看作一个递归层,两个卷积层之间加入 dropout 避免过拟合.

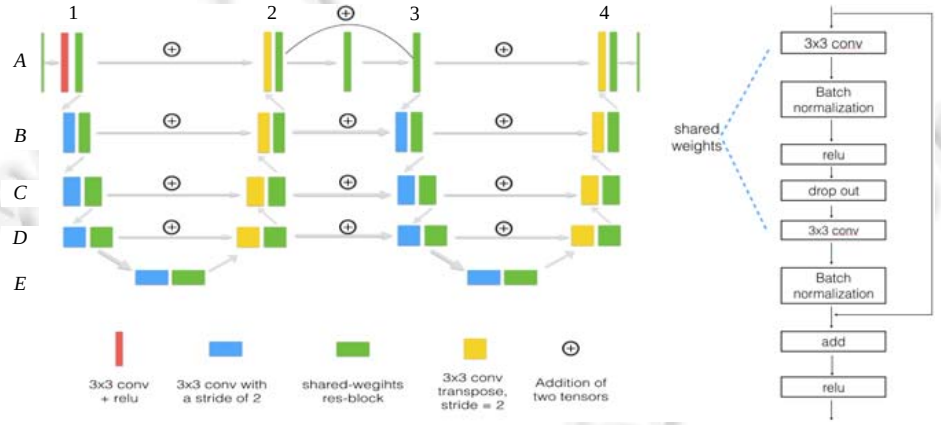


Fig.34 LadderNet network structure diagram[49]

图 34 LadderNet 网络结构图[49]

3.4 基于自动搜索最优网络结构的改进方法

Ken 等人借鉴了网络结构搜索(NAS),提出了 SegNAS3D[52]三维图像分割网络结构搜索,以解决三维图像分割中大量手动调参和网络体系结构优化的问题,如图 35 所示.文中提出了在每一层加入一个新的块结构 Mg-Blk,该块结构是由可学习块 Block[87]、空间 dropout 和可选择残差连接组成.文中最重要的是这个可学习块 block 的学习训练,文中将一个块结构表示成一个有向无环图,如图 36 所示,每一个节点代表一个特征图,每一条边代表一次操作.矩阵的行和列为输入节点和输出节点,矩阵中的数值为扩张卷积的扩张率,通过学习节点数以及扩张率来训练整个网络的准确性.

© 中国科学院软件研究所 http://www.jos.org.cn

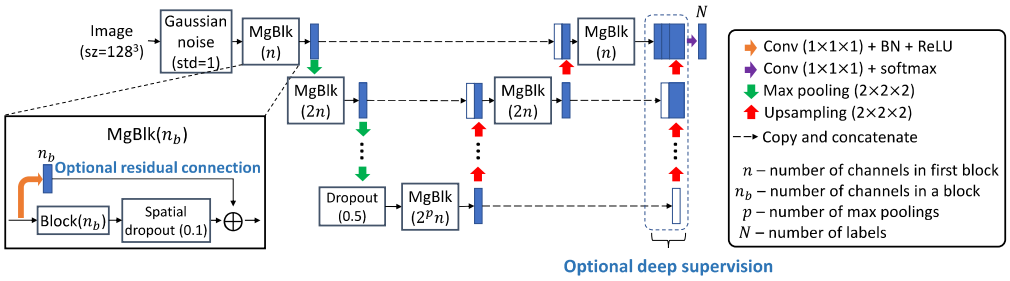


Fig.35 SegNAS3D network structure diagram[52]

图 35 SegNAS3D 网络结构图[52]

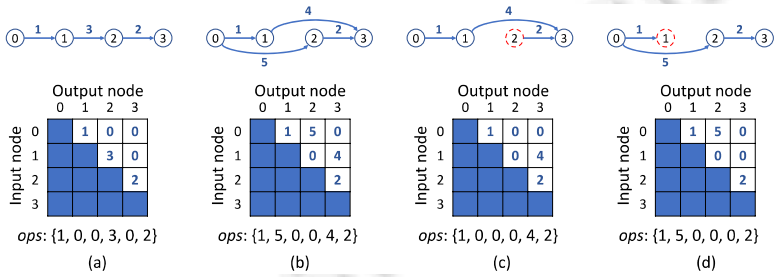


Fig.36 SegNAS3D expansion coefficient selection[52]

图 36 SegNAS3D 扩张系数选择[52]

3.5 基于瓶颈(bottleneck)的改进方法

Bottleneck 是 U 型网络收缩路径和扩张路径中间的部分,其主要接受了所有来自编码器提取的特征信息,并将分割好的的图像通过解码器恢复到原有分辨率,因而其重要性可想而知.一般对于 bottleneck 的改进,多采用 attention 机制,以更好地关注分割细节.Wang 等人提出的巩膜分割模型 ScleraSegNet[9]在 bottleneck 中采用 4 种 attention 机制,将通道注意力和空间注意力相结合,以更好地分割.而 Wang 等人[31]受 GRU 的启发,在 bottleneck 处重复使用 SRU 模型,在保证分割精度的同时,又减轻参数过多带来的影响.

而 Li 等人提出了新的改进方法 BSU-Net[50],其先将 U-Net 的编解码器以及 bottleneck 进行改进,加入了 Inception、Dense 模块和扩张卷积,并称该网络为 Base U-Net.然后将 Base U-Net 按照是否去掉跳连分为 Encoding U-Net 和 Segmentation U-Net,再将两者的 bottleneck 部分连接.大多数的网络金队输入输出有监督,而 BSU-Net 通过将 U-Net 改进,称自动编码器,弥补了 bottleneck 的监督空白,并且可以进一步提取 bottleneck 处的信息,对于图像进行更好的分割.网络结构如图 37 所示.

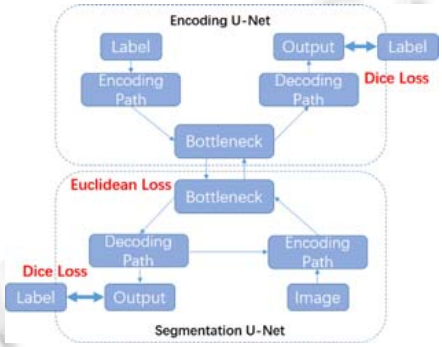


Fig.37 BS U-Net network structure diagram[50]

图 37 BS U-Net 网络结构图[50]

4 U-Net 结构改进中常用结构模块

由于图像分割任务的目标不同,其网络结构也不尽相同.通过对解决问题的分类,我们总结归纳出如下适用于不同问题的网络结构模块,从而帮助大家针对不同问题,快速找到适用的模块以组成网络结构.

4.1 残差结构

一般来说,增加神经网络的宽度和深度可以提高网络的表达性能.但如果简单地增加网络的层数,就会面对梯度消失或是梯度爆炸的问题.何凯明等人为此提出了残差网络[88],将每两层网络中增加一个跳连,以保证增加层数后不会削弱网络的表达性能.残差结构如图 38 所示.在 V-net[2]、MultiResUNet[19]、RDA-Unet[12]等结构中都有采用,可以放置于编码解码器单元,也可在下采样中使用.

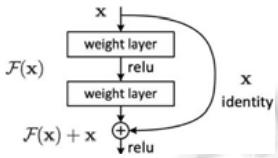


Fig.38 Residual structure[88]

图 38 残差结构[88]

4.2 Attention模块

Attention 的主要思想是强化特征,在医学影像里常用于分割细节较多的血管、肺叶,其内部结构不固定,可根据强调的内容自行设计,例如强调特征图称为 Spatial Attention、强调通道为 Channel Attention.Attention U-net[5]、CASU[6]、RAUNet[7]、CIA-Net[8]都是在解码器中加入 attention 模块,Attention U-net 的 AGs 是加入门控信号以消除噪声,CASU 的 AGs 采用 Up-link 以聚焦特征,IAM 是将核、轮廓两个解码器分支的信息聚合.这三者都是空间信息方面的特征聚焦,而 AAM 主要是通过强调目标通道从而聚焦特征;ScleraSegNet[9]是在瓶颈处加入 CAM 和 SAM 块,从空间和通道两个方面进行特征的聚焦.表 1 给出了不同的 Attention 模块的总结.

Table 1 Network model with attention mechanism

表 1 加入 attention 机制的网络模型

网络简称	改进结构	模块图号	Attention 机制	功能
Attention U-net[5]	解码器	图5	AGs	门控 g 消除噪声
CASU[6]	解码器	图6	AGs	Uplink 聚焦特征
RAUNet[7]	解码器	图7	AAM	强调目标通道
CIA-Net[8]	解码器	图8	IAM	核和轮廓信息聚合
ScleraSegNet[9]	瓶颈处	图9	CAM+SAM	通道+空间attention

4.3 SE模块

通道的压缩和激励,通过找到特征较好的通道,强调这一通道,压缩不相关的通道,以减少参数,强化分割精度.无论 AnatomyNet[46],还是 Roy 等人[4]提出的通道和空间挤压激励,都是运用了这一思想.SKNet[89]在 SENet 的基础上结合了 Merge-and-RunMappings 以及 attention on inception block,用多尺度特征汇总信息,来按照通道指导侧重使用哪个核的表征,从而减少了参数增加路径和动态选择.因而在之后的图像分割中,可以考虑采用 SKNet 所提出的方法代替 SE 结构.

4.4 DenseNet模块

DenseNet 主要是将每一层的网络复制到下一层的全连接结构,增强数据流动的同时减少计算量.医学影像分割中,常将 Dense 的应用于编码器与解码器的全连接、编解码器的单元,从而达到减少参数提升计算精度.H-DenseUNet[14]和 FD-UNet[16]都是在编解码器的单元中引用 Dense 模块,不同在于:前者是将 2D 和 3D Dense 的模块相结合,用 2D 网络指导 3D 网络分割;而后者仅仅在编解码器每个单元中使用 Dense 模块以减少参

© 中国科学院软件研究所 http://www.jos.org.cn

数.Dense Multi-path U-Net[27]和 IVD-Unet[28]的网路结构大部分相同,后者在前者的基础上将编码器的稠密性提升到了一个新的高度,将不同模态之间的信息更好地交互.MDU-Net[51]将稠密的思想应用在编码器的每个单元之间、解码器的每个单元之间,并且跳连也不再单单是编解码器对应层的连接,采用全连接的方式将高级信息和低级信息充分融合利用,表 2 给出了稠密模块的总结.

Table 2 Application of DenseNet thought in network

表 2 应用 DenseNet 思想的网络

网络简称	改进结构	Dense 形式	作用
H-DenseUNet[14]	编解码器	2D+3D Dense 模块	保证信息流
FD-UNet[16]	编解码器	Dense block	减少参数
Dense Multi-path U-Net[27]	编码器	不同模态稠密连接	增强不同模态之间的信息流
IVD-Unet[28]	编码器	不同模态超级稠密连接	增强不同模态的信息流
MDU-Net[51]	编解码器、跳连	全网络稠密连接	增强信息流动

4.5 Inception模块

在医学影像中,病灶所占区域在不同的图像中变化极大,因而卷积核的选择非常重要.然而较小的卷积核适合局部信息,较大的内核更适合捕获全局分布的信息.Inception 将扩张率不同、大小不同的卷积都囊括在内,通过不同尺度提取特征,从而达到精确提取特征的目的,例如 CE-Net[73]中的 DAC 模块就是为了更好地提取特征而设计.但由于 Inception 中包含不同大小的卷积核,大卷积核相比于小卷积核计算量要多很多,因而 MultiResUNet[19]选择用 3×3 代替 5×5,7×7 的卷积核,而 Dense Multi-path U-Net[27]和 IVD-Unet[28]采用非对称 inception 模块,也就是将 $n\times n$ 的卷积拆解成 $n\times 1$,$1\times n$ 的卷积,从而降低计算量.

4.6 CycleGAN模块

生成对抗性网络(GAN)将图像从一个领域转换成另外一领域,例如将斑马转换成马.在医学影像中,由于不同源域的图像放在一起训练将导致极大的问题,通过将图像先经过 CycleGAN 再进行训练,从而达到更好的训练效果,例如 Unet-GAN(如图 39 所示)[36].

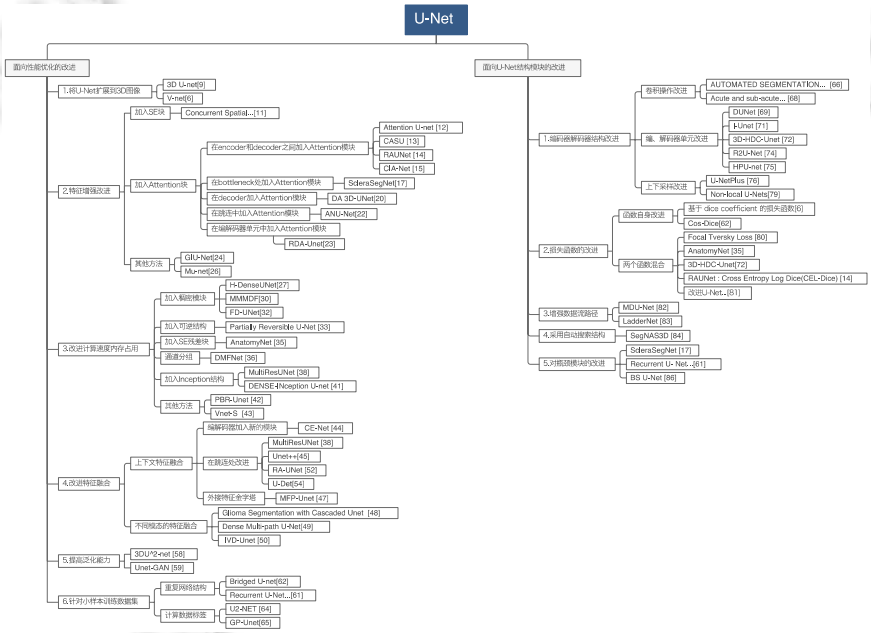


Fig.39 U-net extended structure technology development flow chart

图 39 U-Net 扩展结构技术发展流程图

© 中国科学院软件研究所 http://www.jos.org.cn

5 总结与建议

对于上文介绍的基于 U-Net 扩展结构的技术脉络总结如图 39 所示,按照面向性能优化和面向结构改进总结成表 3 和表 4.主要从网络维度、改进结构、亮点、数据集、评价参数这 5 个方面进行总结.

1) 网络的维度主要是 2D、3D、2D 和 3D 的混合:2D 网络为基本网络;3D 网络因为其具有 2D 网络没有的上下文信息;2D 和 3D 网络的融合以 2D 网络的快速分割结果来指导 3D 模型的学习并实施分割,主要在减少内存、提升计算速度中使用.
2) 对于改进的结构,包括编解码器、bottleneck、跳连、卷积操作、上下采样、损失函数、外接其他结构以及数据标签.编解码器的改进包括编解码器单元的改进、上下采样、卷积操作的改进,在大部分问题中都有应用,而像跳连、数据标签这类改进主要是为解决特征融合、小样本数据集问题.
3) 在数据集方面,相关方法采用了几乎各不相同的数据集,主要包括脑肿瘤(BraTS)、视网膜血管、肝脏、胰脏、腰间盘、乳腺癌、前列腺等.由于各种方法采用的数据集不尽相同,本文列举数据集以提供一个数据集名称索引.由于不同方法评价所采用的数据集不尽相同,所以不同方法难以进行横向比较.
4) 从指标函数方面,可见主要评价指标是 Dice scores,IOU 次之.表中总结了各个方法采取的评价指标以及指标数值,帮助大家在选择网络模型时有一个参考和比较.

Table 3 Based on U-Net, it classifies the performance optimization and compared from the aspects of improved structure and highlights

表 3 基于 U-Net 面向性能优化进行分类,并从改进结构、亮点等方向进行对比

分类	网络简称	时间	维度	改进结构	亮点	数据集	评价参数	数值
基于 3D 图像分割的改进	3D U-Net[3]	2016	3D	编解码器	卷积 2D 变 3D	Enopus kidney embryos	IoU	0.732
	V-net[2]	2016	3D	编解码器	加入残差结构	PROMISE 2012	Avg. Hausdorff distance	5.71± 1.20mm
增强相关特征,抑制无关特征	Concurrent Spatial...[4]	2018	3D	编解码器	scSE	MALC	global Dice scores	0.843± 0.062
	Attention U-net[5]	2018	2D	解码器	AGs	Gastric cancer	Dice scores	0.767± 0.132
	CASU[6]	2019	3D	解码器	AGs	DRIVE	ROC AUC	0.915 7
	RAUNet[7]	2019	3D	解码器	AAM	Cata7	Avg.Dice	0.977 1
	CIA-Net[8]	2019	3D	解码器	信息聚合模块(IAM)	MoNuSeg 2018 MICCAI	*F*1-score	0.845 8
	ScleraSegNet[9]	2019	2D	bottleneck	4 个 Attention 模块替换 bottleneck 块	UBIRIS, MICHE-I and MASD	Precision	0.921 1
	DA 3D-UNet[10]	2019	3D	解码器、上采样	双注意力模块、DUpsampling	LIDC	MIoU	0.894
	RDA-Unet[12]	2019	2D	编解码器单元	残差双注意力模块	从医院共采集 79 位肾囊肿患者的 CT 图像	DSC	0.965
	GIU-Net[64]	2020	2D	解码器	池化层级连,分割后使用图割算法	codalab	DSC	0.9505± 0.0132
	ANU-Net[11]	2020	2D	跳连	Attention gate	LiTS/ CHAO	Dice	0.9815/ 0.9479
	Mu-net[13]	2020	3D	重复 U 型网	多尺度 U-Net	2PM image datasets	PSNR (N1-B1)	26.41
针对内存占用、计算速度的改进方法	FD-Unet[16]	2018	2D	卷积操作	全密集模块	Circles	AVERAGE SSIM	0.82± 0.07
	H-DenseUNet[14]	2018	2/3D	U-Net 放入其他结构	2D+3D Dense Net	MICCAI 2017 LiTS 和 3DIRCADb	Dice scores	0.937± 0.02
	AnatomyNet[46]	2018	3D	编解码器	SE 残差模块	MICCAI head and neck segmentation 2015	Avg. Dice	0.792 5

© 中国科学院软件研究所 http://www.jos.org.cn

Table 3 Based on U-Net, it classifies the performance optimization and compared from the aspects of improved structure and highlights (Continued 1)

表 3 基于 U-Net 面向性能优化进行分类,并从改进结构、亮点等方向进行对比(续 1)

分类	网络简称	时间	维度	改进结构	亮点	数据集	评价参数	数值
针对内存占用、计算速度的改进方法	DMFNet[18]	2019	3D	编解码器	DMF	BraTS-2018	Dice scores	0.850 9
	Partially Reversible U-Net[17]	2019	3D	编解码器	部分可逆结构	BraTS 2018	Dice scores	0.856 2
	PBR-Unet[21]	2019	3D	U-Net 放入其他结构	功能提取模块和双向递归模块	NIH Pancreatic segmentation	Avg. Dice	83.35± 5.02
	MultiResUNet[19]	2019	2D/ 3D	跳连	MultiRes block	BraTS17	5-Fold Cross Validation	78.1936± 0.7868
	Vnet-S[22]	2019	3D	编解码器、跳连	减少卷积单元和卷积核,跳连加入 Dropout	MICCAI 2017 LiTS	Dice	0.96
	DENSE-Inception U-net[20]	2020	2D/ 3D	卷积操作、下采样和上采样操作	卷积采用 Dense-Inception 模块,上下采样使用 inception 模块	MICCAI BraTS 2017	Average Dice score	0.958 2
	MMMDF[15]	2020	2D/ 3D	编解码器	2D ResUnet+ 3D ResUet	421 名 NPC 患者 T1W, T2W,T1C 3 种脑结构 MR 图像	Dice	0.805
针对小样本训练数据集的改进	GP-Unet[34]	2017	3D	数据标签	每幅图像只需要一个单一的全局标签	subset of the Rotterdam Scan Study	Sensitivity	0.62
	Bridged U-net[32]	2018	2D	重复 U 型网	U-net 桥连接	PROMISE12	Mean vDSC	89.56
	U2-NET[33]	2019	2D	数据标签	贝叶斯训练给定数据标签	Vienna Reading	Dice scores	0.8943± 0.04
	Recurrent U-net…[31]	2019	2D	bottleneck	双门控递归单元(DRU)和单门控递归单元(SRU)	DRIVE	mIOU	0.849
针对特征融合	UNet++[23]	2018	2D	跳连	在跳连阶段采用密集块和深度监控	cell nuclei	Jaccard/IOU	92.63
	Glioma Segmentation with Cascaded Unet[26]	2018	3D	编码器	分别学习各模态再组合	BraTS 2018	Avg. Dice	0.908
	Dense Multi-path U-Net[27]	2018	3D	编码器	密集多路径	ISLES	DSC	0.635± 0.186
	IVD-Net[28]	2018	3D	编码器	超密集多路径	LITS	Avg. Dice	0.961
	CE-Net[73]	2019	2D	编解码器	DAC 和 RMP	ORIGA	Accuracy	0.955± 0.003
	MultiResUNet[19]	2019	3D	跳连	Res path	BraTS17	5-Fold Cross Validation	78.1936± 0.7868
	RA-UNet[29]	2019	3D	跳连	Augmented Attention Module	MICCAI 2017 LiTS and the 3DIRCADb	mDice	0.977 1
	MFP-Unet[25]	2019	2D	外接	外接特征金字塔网络	validation	Dice scores	0.97± 0.13
	DFA-Net[24]	2020	3D	跳连	特征聚合模块 FAM	brainweb	Dice (WM GM CSF)	95.84,94.91, 94.68
	U-DET[30]	2020	2D	跳连、激活函数	Bi-FPN,Mish 激活函数	LUNA-16	DSC	0.828 2

© 中国科学院软件研究所 http://www.jos.org.cn

Table 3 Based on U-Net, it classifies the performance optimization and compared from the aspects of improved structure and highlights (Continued 2)

表 3 基于 U-Net 面向性能优化进行分类,并从改进结构、亮点等方向进行对比(续 2)

分类	网络简称	时间	维度	改进结构	亮点	数据集	评价参数	数值
针对提高泛化能力	Unet-GAN[36]	2019	2D	U-Net 放入其他结构单元内	CycleGAN	GE	Dice scores	0.867±0.035
	3DU2-net[35]	2019	3D		Adapter	Base01 Heart	Dice scores	93.26

Table 4 Based on U-Net, it classifies the structural module improvement, and compares the improved structure and highlights

表 4 基于 U-Net 面向结构模块改进进行分类,并从改进结构、亮点等方向进行对比

分类	网络简称	时间	维度	改进结构	亮点	数据集	评价参数	数值
针对编码器解码器的改进	Acute and sub-acute...[38]	2018	3D	卷积操作、下采样和上采样操作	长短残差网络结构	ISLES 2015	Avg. Dice	0.59±0.31
	DUNet[39]	2018	3D	编解码器	可形变卷积	DRIVE	Accuracy	0.969 7
	R2U-Net[42]	2018	2D	编码解码器单元	循环残差卷积	Retina Vessels	mIOU (%)	0.821
	HPU-net[43]	2019	2D	解码器	cVAE	LIDC	IoUrec	0.97±0.00
	Automated segmentation...[37]	2019	3D	卷积操作	协调引导卷积	LUNA16	Dice coefficient	0.947 (0.044)
	U-NetPlus[44]	2019	2D	上采样	Nearest-Neighbor interpolation	MICCAI 2017EndoVis	Dice scores	0.902
	I Unet[40]	2019	2D	编解码器	加入 Incep-tion 和 RCL	ISIC Archive	Dice	0.871
	3D-HDC-Unet[41]	2020	3D	编码器	混合膨胀卷积	BraTS2017	DSC(WT, TC,ET)	0.90,0.80, 0.70
	Non-local U-Nets[45]	2020	2D	上下采样	全局聚合模块(global aggregation block)	3D multimodality isointense infant brain MR image	Dice	0.923 9
基于损失函数改进的方法	V-Net[2]	2016	3D	损失函数	Dice-based loss	PROMISE 2012	Avg. Dice	0.869±0.033
	Focal Tversky loss...[47]	2018	2D	损失函数	Focal Tversky Loss(FTL)	BUS 2017	Dice scores	0.804±0.024
	Bridged U-Net[32]	2018	2D	损失函数	cos-Dice loss Func-tion	PROMISE12	Dice scores	89.96
	RAUNet[7]	2019	3D	损失函数	CEL-Dice	Cata7	mIOU (%)	0.956 2
	AnatomyNet[46]	2019	3D	损失函数	混合损失函数	MICCAI head and neck seg-mentation2015	Avg. Dice	0.792 5
	3D-HDC-Unet[41]	2020	3D	损失函数	混合损失函数 HLF	BraTS2017	DSC(WT, TC,ET)	0.90,0.80, 0.70
	改进 U-Net…[48]	2020	2D	损失函数	二值交叉熵和 Dice loss 函数	LIDC-IDRI	DSC	84.48
针对数据流路径的改进	LadderNet[49]	2018	2D	重复 U 型网、编解码器	桥连接、共享权重残差块	DRIVE	F1-score	0.820 2
	MDU-Net[51]	2018	2D	编解码器、交叉连接	多尺度密集交叉连接	MICCAI2015 GlaS	Dice scores	0.928
自动搜索最优网络结构	SegNAS3D[52]	2019	3D	编解码器	MgBlk 块	MP-RAGE	Dice	0.817±0.003
针对 Bottle neck 的改进	Recurrent U-net...[31]	2019	2D	bottleneck	双门控递归单元(DRU)和单门控递归单元(SRU)	DRIVE	mIOU	0.849
	ScleraSegNet[9]	2019	2D	bottleneck	4 个 attention 模块替换 bottleneck 块	UBIRIS.v2	Precision	0.921 1
	BSU-Net[50]	2020	2D	bottleneck	Encoding U-Net	LiTS 2017	DPC	0.961 0

© 中国科学院软件研究所 http://www.jos.org.cn

随着器官结构差异化、病灶形状多样化,U-Net 已经无法满足所有病灶高精准性的分割.随着 attention 机制、Dense 模块、Inception 模块以及残差结构、图割等模块的发展和完善,近期一些工作在 U-Net 的基础上加入不同的模块,以实现对于不同病灶的精准分割.基于 U-Net 扩展结构多种多样,因而我们进一步总结了针对不同的目的的网络结构改进方法,总结出几个方案供参考,见表 5.

1) 对于分割微小细节,例如视网膜血管、肺叶等,可在残差的基础上使用可变形卷积或者扩张卷积,根据空间、通道激励,在解码器或 bottleneck 中选择加入 attention 模块或者 SE 模块.
2) 针对提升计算速度的问题,可以采用的基本结构包括扩张卷积、残差结构、SE 残差块等;主要改进的特殊结构包括从编解码器单元加入稠密模块或者可逆结构,对于损失函数可选择 Dice-based loss.
3) 对于特征融合问题,可以将网络编码器按照不同模态分别提取特征再进行总体融合,不同模态之间加入稠密卷积,增加信息流减少参数,或者解码器外接特征金子塔辅助特征融合.

Table 5 Suggestions on using structure for different problems

表 5 针对不同问题的使用结构建议

解决问题	基本结构	特殊结构	放置位置	损失函数
分割 微小细节	扩张卷积、可形变卷积、残差结构	attention 机制、SE 块	decoder 或者 bottleneck	–
提升 计算速度	扩张卷积、残差结构、SE 残差块	稠密卷积、 可逆结构	编解码器单元、 上采样、下采样	Dice-based loss or Dice loss
特征融合	残差结构	Inception、金字塔 特征图、稠密卷积	编解码器、 跳连、外接	–

综上所述,本文分类总结和分析了基于 U-Net 结构改进的医学图像分割方法,从改进的性能指标目的和改进的主要结构特点,对主要的相关工作进行了总结和介绍.基于对现有工作的综述,提炼出基于 U-Net 进行结构改进的一些常用改进模块和常用改进方法,可以作为未来这个领域研究工作的参考.对于 U-Net 的未来应用,U-Net 不仅应用于医学影像分割,在其他领域也有较好的应用效果,例如应用于人群识别的 W-net[90],应用于航空图像的 ResUNet-a[91]、TernausNet[85]、FlowS-Unet[92]、ST-Unet[93]等多种改进方法.将 U-Net 扩展到更多的应用领域,并进一步提升 U-Net 的特征提取和识别准确性,提高计算效率,是未来可行的研究方向.

References:

[1] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Proc. of the Int'l Conf. on Medical Image Computing and Computer-assisted Intervention (MICCAI 2015). 2015. 234–241.

[2] Milletari F, Navab N, Ahmadi S, *et al*. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: Proc. of the Int'l Conf. on 3D Vision. 2016. 565–571.

[3] Cicek O, Abdulkadir A, Abdulkadir A, *et al*. 3D U-net: Learning dense volumetric segmentation from sparse annotation. In: Proc. of the Medical Image Computing and Computer-assisted Intervention. 2016. 424–432.

[4] Roy AG, Navab N, Navab N, *et al*. Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks. In: Proc. of the Medical Image Computing and Computer-assisted Intervention. 2018. 421–429.

[5] Oktay O, Schlemper J, Folgoc LL, *et al*. Attention U-net: Learning where to look for the pancreas. arXiv:1804.03999, 2018.

[6] Li R, Li M, Li J, *et al*. Connection sensitive attention U-net for accurate retinal vessel segmentation. arXiv:1903.05558, 2019.

[7] Ni ZL, Bian GB, Zhou XH, *et al*. RAUNet: Residual attention U-net for semantic segmentation of cataract surgical instruments. arXiv:1909.10360, 2019.

[8] Zhou Y, Onder OF, Dou Q, *et al*. CIA-net: Robust nuclei instance segmentation with contour-aware information aggregation. In: Proc. of the Int'l Conf. on Information Processing in Medical Imaging (IPMI 2019). 2019. 682–693.

[9] Wang C, He Y, Liu Y, *et al*. ScleraSegNet: An improved U-net model with attention for accurate sclera segmentation. In: Proc. of the Int'l Conf. on Biometrics. 2019. 1–8.

[10] Wang P, Qiang Y, Yang XT, Hou TX. Double attention 3D-UNet for lung nodule segmentation. Computer Engineering, 2020 (in Chinese with English abstract). https://doi.org/10.19678/j.issn.1000-3428.0057019

© 中国科学院软件研究所 http://www.jos.org.cn

[11] Li C, Tan Y, Chen W, *et al*. ANU-Net: Attention-based nested U-Net to exploit full resolution features for medical image segmentation. Computers and Graphics, 2020,90:11−20.

[12] Xu HW, Yan PX, Wu M, Xu ZY, Sun YB. Automated segmentation of cystic kidney in CT images using residual double attention motivated U-Net model. Application Research of Computers, 2020,37(7):2237−2240 (in Chinese with English abstract).

[13] Lee S, Negishi M, Urakubo H, *et al*. Mu-Net: Multi-scale U-net for two-photon microscopy image denoising and restoration. Neural Networks, 2020,125:92−103.

[14] Li X, Chen H, Qi X, *et al*. H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE Trans. on Medical Imaging, 2018,37(12):2663−2674.

[15] Hong YJ, Meng TB, Li HJ, Liu LZ, Xu SY, Guo SW. Deep segmentation method of tumor boundaries from MR images of patients with nasopharyngeal carcinoma using multi-modality and multi-dimension fusion. Journal of Zhejiang University (Engineering Science), 2020,54(3):566−573 (in Chinese with English abstract).

[16] Guan S, Khan AA, Sikdar S, *et al*. Fully dense UNet for 2D sparse photoacoustic tomography artifact removal. IEEE Journal of Biomedical and Health Informatics, 2020,24(2):568−576.

[17] Brügger R, Baumgartner CF, Konukoglu E. A partially reversible U-net for memory-efficient volumetric image segmentation. In: Shen D, *et al*., eds. Proc. of the Int'l Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI 2019). Springer-Verlag, 2019.

[18] Chen C, Liu X, Ding M, *et al*. 3D dilated multi-fiber network for real-time brain tumor segmentation in MRI. In: Proc. of the Int'l Conf. on MICCAI 2019. 2019. 184−192.

[19] Ibtehaz N, Rahman MS. MultiResUNet: Rethinking the U-net architecture for multimodal biomedical image segmentation. arXiv: 1902.04049v1, 2019.

[20] Zhang Z, Wu C, Coleman S, *et al*. DENSE-INception U-net for medical image segmentation. Computer Methods and Programs in Biomedicine, 2020,192:Article No.105395.

[21] Li J, Lin X, Che H, *et al*. Probability map guided bi-directional recurrent UNet for pancreas segmentation. arXiv:1903.00923, 2019.

[22] Xu BQ, Ling TH. Automatic segmentation algorithm for single organ of CT images based on cascaded Vnet-S network. Journal of Computer Applications, 2019,39(8):2420−2425 (in Chinese with English abstract).

[23] Zhou Z, Siddiquee MR, Tajbakhsh N, *et al*. UNet++: A nested U-net architecture for medical image segmentation. arXiv:1807. 10165, 2018.

[24] Yang B, Liu XF, Zhang J. Medical image segmentation based on deep feature aggregation network. Computer Engineering, 2020 (in Chinese with English abstract). http://kns.cnki.net/kcms/detail/31.1289.TP.20200407.0948.001.html

[25] Moradi S, Oghli MG, Alizadehasl A, *et al*. MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography. Physica Medica, 2019,67:58−69.

[26] Lachinov D, Vasiliev E, Turlapov V, *et al*. Glioma segmentation with cascaded UNet. arXiv:1810.04008, 2018.

[27] Dolz J, Ayed IB, Desrosiers C, *et al*. Dense multi-path U-net for ischemic stroke lesion segmentation in multiple image modalities. arXiv:1810.07003, 2018.

[28] Dolz J, Desrosiers C, Ayed IB, *et al*. IVD-Net: Intervertebral disc localization and segmentation in MRI with a multi-modal UNet. arXiv:1811.08305, 2018.

[29] Jin Q, Meng Z, Sun C, *et al*. RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans. arXiv:1811. 01328, 2018.

[30] Keetha NV, Babu PSA, Annavarapu CSR. U-Det: A modified U-Net architecture with bidirectional feature network for lung nodule segmentation. arXiv:2003.09293, 2020.

[31] Wang W, Yu K, Hugonot J, *et al*. Recurrent U-net for resource-constrained segmentation. arXiv preprint arXiv:1906.04913, 2019.

[32] Chen W, Zhang Y, He J, *et al*. Prostate segmentation using 2D bridged U-net. In: Proc. of the Int'l Joint Conf. on Neural Network. 2019. 1−7.

[33] Orlando JI, Seebock P, Bogunovic H, *et al*. U2-Net: A Bayesian U-net model with epistemic uncertainty feedback for photoreceptor layer segmentation in pathological oct scans. In: Proc. of the Int'l Symp. on Biomedical Imaging. 2019. 1441−1445.

© 中国科学院软件研究所 http://www.jos.org.cn

[34] Dubost F, Bortsova G, Adams HH, *et al*. GP-Unet: Lesion detection from weak labels with a 3D regression network. In: Proc. of the Medical Image Computing and Computer-assisted Intervention. 2017. 214−221.

[35] Huang C, Huang C, Han H, *et al*. 3D U^2-net: A 3D universal U-net for multi-domain medical image segmentation. In: Proc. of the Medical Image Computing and Computer-assisted Intervention. 2019. 291−299.

[36] Yan W, Wang Y, Gu S, *et al*. The domain shift problem of medical image segmentation and vendor-adaptation by Unet-GAN. In: Proc. of the Medical Image Computing and Computer-assisted Intervention. 2019. 623−631.

[37] Wang W, Chen J, Zhao J, *et al*. Automated segmentation of pulmonary lobes using coordination-guided deep neural networks. In: Proc. of the Int'l Symp. on Biomedical Imaging. 2019. 1353−1357.

[38] Clerigues A, Valverde S, Bernal J, *et al*. Acute and sub-acute stroke lesion segmentation from multimodal MRI. arXiv:1810.13304, 2018.

[39] Jin Q, Meng Z, Meng Z, *et al*. DUNet: A deformable network for retinal vessel segmentation. arXiv:1811.01206v1, 2018.

[40] Jiang HD, Ye XN. An improved skin disease image segmentation algorithm based on I-Unet network, Modern Electronics Technique, 2019,42(12):52−56 (in Chinese with English abstract).

[41] He CE, Xu HJ, Wang Z, *et al*. Automatic segmentation algorithm for multimodal magnetic resonance-based brain tumor images. Acta Optica Sinica, 2020,40(6):66−75 (in Chinese with English abstract).

[42] Alom MZ, Hasan M, Yakopcic C, *et al*. Recurrent residual convolutional neural network based on U-net (R2U-net) for medical image segmentation. arXiv:1802.06955, 2018.

[43] Kohl SA, Romeraparedes B, Maierhein KH, *et al*. A hierarchical probabilistic U-net for modeling multi-scale ambiguities. arXiv:1905.13077, 2019.

[44] Hasan SM, Linte CA. U-NetPlus: A modified encoder-decoder U-net architecture for semantic and instance segmentation of surgical instrument. arXiv:1902.08994, 2019.

[45] Wang Z, Zou N, Shen D, *et al*. Non-local U-Nets for biomedical image segmentation. In: Proc. of the AAAI Conf. on Artificial Intelligence. 2020. 6315−6322.

[46] Zhu W, Huang Y, Zeng L, *et al*. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. arXiv:1808.05238v2, 2019.

[47] Abraham N, Khan NM. A novel focal Tversky loss function with improved attention U-net for lesion segmentation. In: Proc. of the Int'l Symp. on Biomedical Imaging. 2019. 683−687.

[48] Zhong SH, Guo XM, Zheng YN. Improved U-Net network for lung nodule segmentation. Computer Engineering and Applications, 2020,56(17):203−209 (in Chinese with English abstract).

[49] Zhuang J. LadderNet: Multi-path networks based on U-net for medical image segmentation. arXiv:1810.07810, 2018.

[50] Song LI, Tso GK, Kaijian HE, *et al*. Bottleneck feature supervised U-net for pixel-wise liver and tumor segmentation. arXiv:1810.10331, 2020.

[51] Zhang J, Jin Y, Xu J, *et al*. MDU-Net: Multi-scale densely connected U-net for biomedical image segmentation. arXiv:1812.00352, 2018.

[52] Wong KC, Moradi M. SegNAS3D: Network architecture search with derivative-free global optimization for 3D image segmentation. arXiv:1909.05962, 2019.

[53] Foody GM. Status of land cover classification accuracy assessment. Remote Sensing of Environment, 2002,80(1):185−201.

[54] Jaccard P. The distribution of the flora in the alpine zone. New Phytologist, 1912,11(2):37−50.

[55] Horé A, Ziou D. Image quality metrics: PSNR vs. SSIM. In: Proc. of the 20th Int'l Conf. on Pattern Recognition (ICPR 2010). Istanbul: IEEE Computer Society, 2010. 2366−2369. [doi: 10.1109/ICPR.2010.579]

[56] Lin TY, Goyal P, Girshick R, *et al*. Focal loss for dense object detection. arXiv:1708.02002, 2017.

[57] Physica A, Boris Podobnik A, Jia Shao C, *et al*. Features of similarity. Psychological Review,1977,84(4):327−352.

[58] Salehi SS, Salehi SS, Erdogmus D, *et al*. Tversky loss function for image segmentation using 3D fully convolutional deep networks. arXiv:1706.05721, 2017.

[59] Hu J, Shen L, Albanie S, *et al*. Squeeze-and-excitation networks. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2020, 42(8):2011−2023. [doi: 10.1109/TPAMI.2019.2913372]

© 中国科学院软件研究所 http://www.jos.org.cn

[60] Dou Q, Yu L, Chen H, *et al*. 3D deeply supervised network for automated segmentation of volumetric medical images. Medical Image Analysis, 2017,41(41):40−54.

[61] Park J, Woo S, Lee J, *et al*. BAM: Bottleneck attention module. arXiv:1807.06514, 2018.

[62] Woo S, Park J, Lee J, *et al*. CBAM: Convolutional block attention module. In: Proc. of the European Conf. on Computer Vision. 2018. 3−19.

[63] Tian Z, He T, Shen C, *et al*. Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation. In: Proc. of the Computer Vision and Pattern Recognition. 2019. 3126−3135.

[64] Liu Z, Song Y, Sheng VS, *et al*. Liver CT sequence segmentation based with improved U-Net and graph cut. Expert Systems with Applications, 2019,126:54−63.

[65] Boykov Y, Veksler O. Graph Cuts in Vision and Graphics: Theories and Applications. Springer-Verlag, 2006. 79−96.

[66] Huang G, Liu Z, Laurens VDM, *et al*. Densely connected convolutional networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2017. 2261−2269.

[67] Tu Z. Auto-context and its application to high-level vision tasks. In: Proc. of the Computer Vision and Pattern Recognition. 2008. 1−8.

[68] Ntziachristos V, Ripoll J, Wang LV, *et al*. Looking and listening to light: The evolution of whole-body photonic imaging. Nature Biotechnology, 2005,23(3):313−320.

[69] Gomez AN, Ren M, Urtasun R, *et al*. The reversible residual network: Backpropagation without storing activations. In: Proc. of the Neural Information Processing Systems. 2017. 2214−2224.

[70] Chen Y, Kalantidis Y, Li J, *et al*. Multi-fiber networks for video recognition. In: Proc. of the European Conf. on Computer Vision. 2018. 364−380.

[71] Szegedy C, Vanhoucke V, Ioffe S, *et al*. Rethinking the inception architecture for computer vision. In: Proc. of the Computer Vision and Pattern Recognition. 2016. 2818−2826.

[72] Szegedy C, Liu W, Jia Y, *et al*. Going deeper with convolutions. In: Proc. of the Computer Vision and Pattern Recognition. 2015. 1−9.

[73] Gu Z, Cheng J, Fu H, *et al*. CE-Net: Context encoder network for 2D medical image segmentation. IEEE Trans. on Medical Imaging, 2019,38(10):2281−2292.

[74] Szegedy C, Ioffe S, Vanhoucke V, *et al*. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: Proc. of the National Conf. on Artificial Intelligence. 2016. 4278−4284.

[75] Wang F, Jiang M, Qian C, *et al*. Residual attention network for image classification. In: Proc. of the Computer Vision and Pattern Recognition. 2017. 6450−6458.

[76] Tan M, Pang R, Le QV, *et al*. EfficientDet: Scalable and efficient object detection. arXiv:1911.09070, 2019.

[77] Li M, Dong S, Gao Z, *et al*. Unified model for interpreting multi-view echocardiographic sequences without temporal information. Applied Soft Computing, 2020,88:Article No.106049.

[78] Lin T, Dollar P, Girshick R, *et al*. Feature pyramid networks for object detection. In: Proc. of the Computer Vision and Pattern Recognition. 2017. 936−944.

[79] Wang Z, Bovik AC, Sheikh HR, *et al*. Image quality assessment: from error visibility to structural similarity. IEEE Trans. on Image Processing, 2004,13(4):600−612.

[80] Cho K, Van Merrienboer B, Bahdanau D, *et al*. On the properties of neural machine translation: Encoder-decoder approaches. In: Proc. of the Empirical Methods in Natural Language Processing. 2014. 103−111.

[81] Liu R, Lehman J, Molino P, *et al*. An intriguing failing of convolutional neural networks and the CoordConv solution. In: Proc. of the Neural Information Processing Systems. 2018. 9605−9616.

[82] Dai J, Qi H, Xiong Y, *et al*. Deformable convolutional networks. In: Proc. of the Int'l Conf. on Computer Vision. 2017. 764−773.

[83] Yu F, Koltun V, Funkhouser T, *et al*. Dilated residual networks. In: Proc. of the Computer Vision and Pattern Recognition. 2017. 636−644.

[84] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2015.

© 中国科学院软件研究所 http://www.jos.org.cn

[85] Iglovikov V, Shvets AA. TernausNet: U-net with VGG11 encoder pre-trained on ImageNet for image segmentation. arXiv:1801.05746, 2018.

[86] He K, Girshick R, Dollar P, *et al*. Rethinking ImageNet pre-training. In: Proc. of the 2019 IEEE/CVF Int'l Conf. on Computer Vision (ICCV). 2019. 4917−4926. [doi: 10.1109/ICCV.2019.00502]

[87] Liu H, Simonyan K, Yang Y, *et al*. DARTS: Differentiable architecture search. In: Proc. of the Int'l Conf. on Learning Representations. 2019.

[88] He K, Zhang X, Ren S, *et al*. Deep residual learning for image recognition. In: Proc. of the Computer Vision and Pattern Recognition. 2016. 770−778.

[89] Li X, Wang W, Hu X, *et al*. Selective kernel networks. In: Proc. of the Computer Vision and Pattern Recognition. 2019. 510−519.

[90] Valloli VK, Mehta K. W-Net: Reinforced U-net for density map estimation. arXiv:1903.11249, 2019.

[91] Diakogiannis FI, Waldner F, Caccetta P, *et al*. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. arXiv:1904.00592, 2019.

[92] Gu L, Xu SQ, Zhu LQ. Detection of building changes in remote sensing images via FlowS-Unet. Acta Automatica Sinica, 2020, 46(6):1291−1300 (in Chinese with English abstract).

[93] Yu B, Yin H, Zhu Z, *et al*. ST-UNet: A spatio-temporal U-network for graph-structured time series modeling. arXiv:1903.05631, 2019.

附中文参考文献:

[10] 王磐,强彦,杨晓棠,侯腾璇.双注意力 3D-UNet 肺结节分割网络.计算机工程,2020. https://doi.org/10.19678/j.issn.1000-3428.0057019

[12] 徐宏伟,闫培新,吴敏,徐振宇,孙玉宝.基于残差双注意力 U-Net 模型的 CT 图像囊肿肾脏自动分割.计算机应用研究,2020,37(7):2237−2240.

[15] 洪炎佳,孟铁豹,黎浩江,刘立志,徐硕瑀,郭圣文.多模态多维信息融合的鼻咽癌 MR 图像肿瘤深度分割方法.浙江大学学报(工学版),2020,54(3):566−573.

[22] 徐宝泉,凌彤辉.基于级联 Vnet-S 网络的 CT 影像单一器官自动分割算法.计算机应用,2019,39(8):2420−2425.

[24] 杨兵,刘晓芳,张纠.基于深度特征聚合网络的医学图像分割.计算机工程,2020. http://kns.cnki.net/kcms/detail/31.1289.TP.20200407.0948.001.html

[40] 蒋宏达,叶西宁.一种改进的 I-Unet 网络的皮肤病图像分割算法.现代电子技术,2019,42(12):52−56.

[41] 何承恩,徐慧君,王忠,等.多模态磁共振脑肿瘤图像自动分割算法研究.光学学报,2020,40(6):66−75.

[48] 钟思华,郭兴明,郑伊能.改进 U-Net 网络的肺结节分割方法.计算机工程与应用,2020,56(17):203−209.

[92] 顾炼,许诗起,竺乐庆.基于 flows-UNet 的遥感图像建筑物变化检测.自动化学报,2020,46(6):1291−1300.

殷晓航(1995－),女,硕士生,主要研究领域为深度学习,医学图像分割.

李德英(1965－),女,博士,教授,博士生导师,CCF 高级会员,主要研究领域为传感器网络,社会网络,图优化理论与算法,算法设计与分析.

王永才(1978－),男,博士,副教授,博士生导师,CCF 专业会员,主要研究领域为物联网,智能感知,网络定位,视觉感知,惯导融合定位.

© 中国科学院软件研究所 http://www.jos.org.cn