

基于多尺度特征融合和双重注意力机制的肝脏CT图像分割

黄扬林, 胡凯, 郭建强, 彭诚

引用本文

黄扬林, 胡凯, 郭建强, 彭诚. 基于多尺度特征融合和双重注意力机制的肝脏CT图像分割[J]. 计算机科学, 2022, 49(11A): 210800162-9.

HUANG Yang-lin, HU Kai, GUO Jian-qiang, Peng Cheng Key Laboratory of Intelligent Computing & Information Processing of Ministry of Education, Xiangtan University, Xiangtan, Hunan 411105, China. [Liver CT Images Segmentation Based on Multi-scale Feature Fusion and Dual Attention Mechanism](#) [J]. Computer Science, 2022, 49(11A): 210800162-9.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于子空间特征相互学习的MRI与PET/SPECT图像融合](#)

MRI and PET/SPECT Image Fusion Based on Subspace Feature Mutual Learning
计算机科学, 2022, 49(11A): 211000171-6. <https://doi.org/10.11896/jsjcx.211000171>

[结合注意力机制的多尺度特征融合图像去雾算法](#)

Multi-scale Feature Fusion Image Dehazing Algorithm Combined with Attention Mechanism
计算机科学, 2022, 49(5): 50-57. <https://doi.org/10.11896/jsjcx.210400093>

[基于DenseNet和混合域注意力的COVID-19低剂量CT图像质量评价](#)

Image Quality Assessment for Low-dose-CT Images of COVID-19 Based on DenseNet and Mixed Domain Attention
计算机科学, 2021, 48(11A): 620-624. <https://doi.org/10.11896/jsjcx.201200252>

[使用模糊聚类的胶囊网络在图像分类上的研究](#)

Study on Image Classification of Capsule Network Using Fuzzy Clustering
计算机科学, 2019, 46(12): 279-285. <https://doi.org/10.11896/jsjcx.190200315>

[面向RP的工业CT切片数据格式转换软件开发](#)

计算机科学, 2008, 35(1): 263-265.

基于多尺度特征融合和双重注意力机制的肝脏 CT 图像分割

黄扬林 胡凯 郭建强 彭诚

湘潭大学智能计算与信息处理教育部重点实验室 湖南湘潭 411105

湘潭大学计算机学院·网络空间安全学院 湖南湘潭 411105

(yanglinhuang@yeah.net)

摘要 肝脏疾病是医学上最常见的疾病之一,对其进行精确的分割是辅助肝脏疾病诊断及手术规划的必要步骤。然而,由于肝脏 CT 图像的复杂性,肝脏分割仍然是一个极具挑战性的问题。以往的研究大多简单地使用拼接或求和操作来融合不同语义,导致无法充分利用其互补性。针对这一问题,提出了一种基于多尺度特征融合和双重注意力机制的网络模型 MD-AUNet。首先利用分层多尺度注意力下采样模块中分层级的双重注意力机制有效地融合不同尺度特征信息,提取富含空间信息的特征表示。然后通过全局注意力上采样模块获取高层特征的全局上下文用于对低层特征信息加权,从而选择更为精确的空间信息。同时在网络训练时采用深层监督策略,以学习不同解码层的层次表示。最后提出了一种简洁有效的后处理方法,用于进一步细化 MD-AUNet 粗分割结果。在医院采集的肝脏数据集(经专家手动标注)上的实验结果表明,所提算法在主观视觉感受和客观评价指标方面均优于其他现有肝脏分割算法,其平均像素精度、平均交并比和 Dice 相似系数分别为 97.6%,95.4%和 95.5%。

关键词: CT 图像;肝脏分割;多尺度特征融合;双重注意力机制;MD-AUNet

中图法分类号 TP391

Liver CT Images Segmentation Based on Multi-scale Feature Fusion and Dual Attention Mechanism

HUANG Yang-lin, HU Kai, GUO Jian-qiang and Peng Cheng

Key Laboratory of Intelligent Computing & Information Processing of Ministry of Education, Xiangtan University, Xiangtan, Hunan 411105, China

School of Computer Science & School of Cyberspace Science, Xiangtan University, Xiangtan, Hunan 411105, China

Abstract Liver disease is one of the most common diseases in medicine, and accurate segmentation of liver disease is a necessary step to assist liver disease diagnosis and surgical planning. However, liver segmentation is still a challenging task due to the complexity of liver CT images. With the deepening of research, people begin to consider combining high-level semantics with low-level semantics to further enhance the segmentation effect. However, most of previous studies simply use splicing or summation operation to fuse different semantics, resulting in failure to make full use of its complementarity. To solve the above problems, a network(MD-AUNet) based on multi-scale feature fusion and dual attention mechanism is proposed in this paper. Firstly, the hierarchical dual attention mechanism in the hierarchical multi-scale attention down-sampling module(HAM) is used to effectively fuse feature information of different scales and extract feature representations rich in spatial information. Then, the global context of high-level features is obtained through the global attention up-sampling module(GAM) for weighting the low-level feature information, so as to select more accurate spatial information. At the same time, deep supervision strategy is used in network training to learn the hierarchical representation of different decoding layers. Moreover, a concise and effective post-processing method is proposed to refine the coarse segmentation result of MD-AUNet. Experimental results on the liver datasets collected by the hospital(manually annotated by experts) demonstrate that the proposed algorithm is superior to other existing liver segmentation algorithms in subjective visual perception and objective evaluation indicators, and its mean pixel accuracy, mean IoU and Dice are 97.6%, 95.4%, and 95.5% respectively.

Keywords CT image, Liver segmentation, Multi-scale feature fusion, Dual attention mechanism, MD-AUNet

1 引言

肝脏是人体内体积最大的实质性脏器和最重要的消化器

官,也是原发性或继发性肿瘤发展的共同部位。从医学影像中将肝脏组织准确地分割出来,是计算机辅助肝脏疾病诊断与手术规划中一个基础且至关重要的步骤^[1]。计算机断层

基金项目:国家自然科学基金(62272404);湖南省大学生创新创业训练计划项目(S202010530031)

This work was supported by the National Natural Science Foundation of China(62272404) and Innovation and Entrepreneurship Training Program for Hunan University Students(S202010530031).

通信作者:胡凯(kaihu@xtu.edu.cn)

成像(Computed Tomography, CT)由于具有扫描时间快、图像清晰等优点,被广泛应用于肿瘤的诊断。目前,临床多采用人工手动分割的方式从CT图像中获取肝脏区域。这一过程需要由具有相关专业知识和大量实践经验的医生对CT图像进行逐像素的语义标注,极其消耗时间和精力,且易受到医生经验差别以及主观因素的影响^[2]。因此,研究在CT图像中的肝脏自动分割方法具有重要意义。

肝脏分割方法一般可分为传统分割方法和机器学习方法两大类。传统分割方法主要包括阈值法^[3]、区域增长法^[4]、水平集法^[5]以及分水岭法^[6]等。这类方法易受到CT图像灰度信息以及肝脏区域形状等因素的影响,导致对噪声点与图像对比度非常敏感。因此,这类方法并不能很好地胜任肝脏分割这项工作。

此外,大量基于机器学习的分割方法也被提出。例如, Vorontsov等^[7]提出了基于支持向量机(Support Vector Machines, SVM)的图像分割算法,通过提取不同尺度的方向梯度直方图特征来训练SVM分类器,以实现对肝脏的初步分割。Achanta等^[8]提出了基于无监督的聚类算法,通过K-means方法对像素相似度聚类,以完成肝脏肿瘤分割。Li等^[9]通过继承算法AdaBoost来学习肝脏肿瘤区域的不同强度分布,以实现肝脏CT图像中肿瘤边界的定位。上述这些经典的机器学习方法大多需要事先指定图像的特征,并且依据指定的特征来训练模型,导致在非均质、低对比度的肝脏CT图像上分割表现不佳。

近年来,基于卷积神经网络(Convolutional Neural Networks, CNNs)的深度学习因具有高效的特征提取能力和良好的特征表达能力,在医学影像分割领域取得了巨大的成功。在肝脏分割方向上,众多研究者受全卷积神经网络^[10](Fully Convolutional Network, FCN)启发,提出了一系列改进的肝脏分割模型。例如, Bi等^[11]利用残差网络^[12](Residual Neural Network, ResNet)的优点,采用全卷积神经网络结合级联的多尺度融合方式,来实现对LiTS2017肝脏肿瘤数据集的分割。Sun等^[13]设计了一种多通道全卷积神经网络,从CT图像中分割肝脏。结果表明,该网络能有效地融合不同阶段的高级特征,提高网络模型的精度。特别是, Ronneberger等^[14]在FCN基础上修改并扩张,提出了U-Net模型。该网络因其卓越的性能成为了众多医学分割模型的基准网络,并有众多研究者基于该方法提出了改进的语义分割模型。如Zhou等^[15]提出的U-Net++架构、Huang等^[16]提出的U-Net3+架构等,均有效提高了肝脏分割的准确性。

上述方法虽然在肝脏CT图像分割任务中取得了不错的效果,但仍存在一些不足。首先,现有方法中的特征融合大多是将不同层次特征简单地通过拼接或求和的方式进行融合^[14-16],并未充分利用高层特征和低层特征的互补性。同时,这种简单的融合方式还会受到高层相似度的影响,使得低层相似度在区分细粒度差异方面失效^[17-18]。其次,现有方法在解码网络上大多直接使用低层特征帮助高层特征恢复图像细节,然而这造成了有用信息的丢失与无用信息的冗余,从而使网络模型的性能下降。

针对上述问题,本文提出了一种基于多尺度特征融合和双重注意力机制的网络模型MD-AUNet。网络整体上是一个U型架构,在编码器子网络和解码器子网络中分别融入了

分层多尺度注意力下采样模块和全局注意力上采样模块。其中,分层多尺度注意力下采样模块通过分层级的双重注意力机制,分别对空间和通道维度的语义相互依赖性进行建模,使不同尺度特征利用注意力掩模协同优化,从而提取更精确的特征表示。全局注意力上采样模块则利用高层特征的全局上下文对低层特征信息进行加权,引导低层特征选择空间信息,从而减少有用信息的丢失与无用信息的冗余。通过上述结构设计,本文网络可以有效地提取与融合多尺度特征,并通过深层监督学习进一步从聚合特征图中学习层次表示,从而改善最终的语义分割效果。

本文的主要贡献概括如下:

(1)提出了一种端到端的基于多尺度特征融合和双重注意力机制的网络模型MD-AUNet,用于解决肝脏CT图像分割问题。

(2)提出了两种多尺度特征融合方法,用于网络的编码与解码。在编码过程中,本文通过分层多尺度注意力下采样模块中分层级的双重注意力机制,使不同尺度特征利用注意力掩模协同优化,提取更精准的特征表示。在解码过程中,则通过全局注意力上采样模块,利用高层特征的全局上下文对低层特征信息进行加权,引导低层特征选择空间信息,从而减少有用信息的丢失与无关信息的涉及。

(3)引入了一种简洁有效的后处理方法来细化所提方法的分割结果,其目的是修正细碎的错分区域和减轻斑点噪声的影响。实验结果表明,本文提出的后处理方法提升了对易混淆像素点的分类效果,优化了最终的分割结果。

2 相关工作

2.1 基于卷积神经网络的肝脏分割

近年来,基于卷积神经网络的分割方法被广泛应用于医学图像领域。2014年, Long等^[10]提出了全卷积神经网络,创新性地从图像级别的分类进一步延伸到像素级别的分类,为语义分割任务提供了新的思路。研究者们也开始考虑将其应用于肝脏分割方向。Ben-Cohen等^[19]使用3个邻近切片作为输入,以此来训练全卷积神经网络,实现对肝脏的分割。但该方法分割的图像边缘粗糙、结果不够精细。Sun等^[13]则在全卷积神经网络的基础上,采用三通道并行的方式同时提取3个相邻切片特征并将其进行融合,进一步提高了肝脏分割结果的准确性。但是由于肝脏分割任务存在样本数据少、分割精度要求高的特点,上述基于全卷积神经网络的方式并未取得令人满意的效果。为了改善上述问题, Ronneberger等^[14]在全卷积神经网络的基础上,提出了在较少数据样本下就能达到高分割精度的U-Net网络模型,该模型目前在医学图像分割中得到了广泛应用。众多研究者也在U-Net网络模型基础上,展开了对肝脏分割任务的研究。Christ等^[20]使用级联的U-Net进行肝脏粗分割后,利用3D条件随机场进一步细化分割结果,以获得最终的肝脏分割区域。为了进一步提升网络的性能, Schnurr等^[21]在卷积层结构上进行改进,用稀疏的卷积层代替U-Net中原有的卷积层,并在肝脏肿瘤挑战赛(Liver Tumor Segmentation challenge, LiTS)上获得了0.32%的Dice系数提升。Zhou等^[15]则考虑到了U-Net中纯跳跃连接带来的语义上不相似的特征融合问题,引入了嵌套和密集的连接来减少编码器和解码器之间的语义

差距,并在肝脏分割结果上获得了较大提升。Huang 等^[16]进一步设计了编码器与解码器之间的相互连接以及解码器之间的内部连接,来从全尺度捕获细粒度细节和粗粒度语义,以此获得更为精确的肝脏分割结果。

2.2 注意力机制

近年来,在计算机视觉领域,基于空间域和通道域的注意力机制得到了广泛的应用。Oktay 等^[22]提出了一种空间注意力机制,通过在特征拼接前加入一个集成注意力门,来重新调整输出特征,以获取更好地特征表示。Hu 等^[23]提出了一种通道注意力机制,通过对特征通道间的相关性进行建模,把重要的特征进行强化来提升准确率。为了更好地达到增强特征的效果,研究者们进一步探索了如何将空间注意力与通道注意力机制进行有效地融合。Roy 等^[24]受文献^[23]的启发,提出了一种并行的空间和通道的压缩与激励模块 scSE,使特征映射分别在空间和通道上进行重新校准。通过在全脑分割和器官分割任务上的评估结果,验证了该方法的有效性。Woo 等^[25]通过级联和并行的方式整合空间注意力和通道注意力,用于进一步提升特征提取能力。Fu 等^[26]通过分别对空间和通道维度的语义相互依赖性进行建模,并将其输出求和,来进一步改善特征表示,此外,为了更有效地融合多尺度信息,以获取更好的特征表示,Li 等^[27]将注意力机制和空间金字塔结构相结合,来提取更精准密集的特征,通过在实验数据集上的结果表明,该方法能够更有效地学习特征表示;Sinha 等^[28]通过多尺度注意力机制,来捕获更加广泛和丰富的上下文依赖关系,从而获得更精准的结果;Tao 等^[29]则通过分层级的注意力机制结合多尺度预测,来提升语义分割的结果。

3 本文方法

图 1 给出了本文方法的概述。首先对训练过程中的数据集进行预处理操作,其次引入一种高效的神经网络 MD-AU-Net 来分割肝脏区域。在此基础上,使用 Focal Loss 作为本文网络的损失函数,以减小数据不平衡的影响。最后通过一种简洁有效的后处理方法,进一步细化 MD-AU-Net 的分割结果。

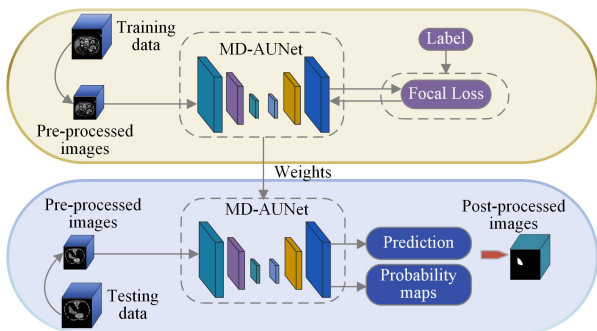


图 1 所提方法的概述

Fig. 1 Overview of the proposed method

3.1 预处理

预处理是卷积神经网络训练的关键步骤,它可以便于网络的训练。首先,为了排除无关器官的影响并增加肝脏器官的图像对比度,本文采用窗口技术对原始肝脏 CT 图像进行预处理,在 Hounsfield 的单位值 $[0, 110]$ 范围内开窗。其次,为了加快模型的收敛速度与防止过拟合现象,本文对上述

得到的图像进行了归一化操作。

3.2 网络模型

图 2 给出了本文网络模型的结构,它整体为一个 U 型编码-解码架构,并利用跳跃连接来传递编码层和解码层之间的信息。其中,编码器每一层均由残差块构成,用于提取特征信息。另一方面,解码器对特征进行连续上采样,逐步恢复图像的分辨率。

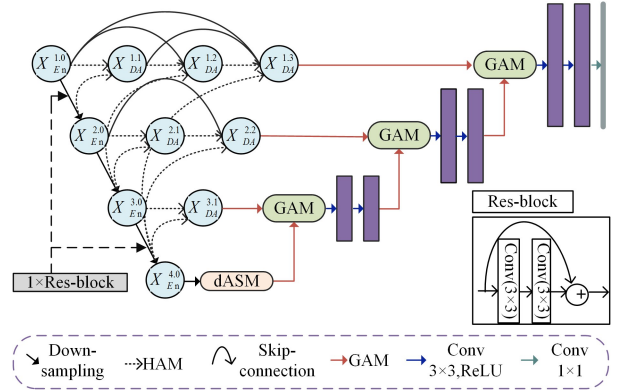


图 2 网络模型 MD-AU-Net 的整体结构

Fig. 2 Overall architecture of MD-AU-Net model

本文网络模型在编码器子网络和解码器子网络中分别引入了分层多尺度注意力下采样模块和全局注意力上采样模块,用于融合多尺度特征信息,以得到更精确的特征表示。在网络中间层,则融合了基于 DenseASPP 改进的中间层模块,用于在保持图像分辨率的前提下获取更大的感受野。

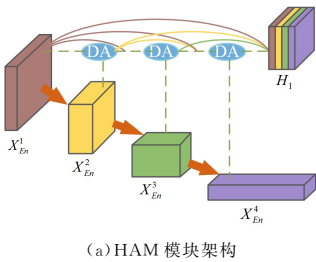
为了优化训练过程,本文使用深层监督策略^[30]来训练网络,以从聚合特征图中学习层次表示。具体来说,本文网络在不同层次的解码特征上进行监督学习。首先通过 1×1 卷积操作将 4 个上采样输出特征的通道数减少为 2,以实现特征图维度的变换以及跨通道的交互和信息整合,然后使用 Softmax 函数对其进行像素级的分类,最后将解码器最后一层的输出作为网络最终的语义分割结果。

接下来将分别对本网络模型中的分层多尺度注意力下采样模块、基于 DenseASPP 改进的中间层模块和全局注意力上采样模块进行详细介绍。

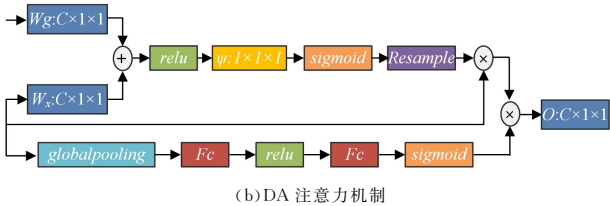
3.2.1 分层多尺度注意力下采样模块

卷积神经网络中,不同尺度的特征图展示着不同的信息。低层语义特征图捕捉丰富的空间信息,能够突出器官的边界;高层语义特征图则体现了器官所在的位置信息。但是,在逐步的下采样和上采样过程中,这些微妙的信号可能会逐渐稀释。因此,高效地融合高、低层特征,是改善分割模型的关键。基于此,本文提出了一种分层级的双重注意力机制,利用空间和通道注意力掩模来指导和优化特征的提取,使提取的多尺度融合特征不仅关注到细节信息,同时关注到语义信息,帮助网络更精确地分割。下面将对该模块进行详细介绍。

分层多尺度注意力下采样模块(HAM)的结构如图 3(a)所示,该图说明了如何构造 H_1 特征。其中, $X_{En}^{i,c}$ 表示来自编码器第 i 个阶段的特征,特征的通道数和尺寸分别为 c 和 $h \times w$ 。将第 i 个阶段特征和第 $i+1$ 个阶段特征作为双重注意力机制(DA),如图 3(b)所示的输入。接下来将分别从空间注意力机制和通道注意力机制上对该模块进行进一步的说明。



(a) HAM 模块架构



(b) DA 注意力机制

图 3 分层多尺度注意力下采样模块

Fig. 3 Hierarchical multi-scale attention down-sampling module

在空间注意力机制上,首先分别对第 i 个阶段特征图和第 $i+1$ 个阶段特征图做 1×1 卷积操作,以获得相同的通道数与尺寸大小:

$$X_{En}^{i,1} = GN(f^{1 \times 1}(X_{En}^i)) \quad (1)$$

$$X_{En}^{i+1,1} = GN(f^{1 \times 1}(X_{En}^{i+1})) \quad (2)$$

其中, $f^{1 \times 1}$ 表示标准卷积操作,卷积核大小为 1×1 ,通道数为 c ; GN 为组标准化操作。再对两个层次特征对应元素相加后进行 ReLU 激活操作:

$$U_{En}^{i,i+1} = \text{ReLU}(X_{En}^{i,1} \oplus X_{En}^{i+1,1}) \quad (3)$$

其中, ReLU 为线性整流激活函数; \oplus 表示对应位置元素相加。经过此阶段,得到了融合第 i 层特征与第 $i+1$ 层特征的新融合特征 $U_{En}^{i,i+1}$ 。之后通过 1×1 卷积操作并使用 Sigmoid 函数激活,将特征值归一化到 0 和 1 之间,得到注意力掩模 $M_{i,i+1}$:

$$M_{i,i+1} = \frac{1}{1 + \exp(-f^{1 \times 1}(U_{En}^{i,i+1}))} \quad (4)$$

其中, $f^{1 \times 1}$ 表示标准卷积操作,卷积核大小为 1×1 ,通道数为 1;进一步地,将此阶段获得的注意力掩模重采样至原来 X_{En}^i 的大小;将得到的权重图与 X_{En}^i 相乘,即可使目标区域获得更多的关注:

$$U_i^s = \text{Resample}(M_{i,i+1}) \otimes X_{En}^i \quad (5)$$

其中, Resample 表示重采样操作; \otimes 表示对应位置元素相乘; U_i^s 为获取了空间注意力的特征。

在通道注意力机制上,受文献[23]的启发,融入 SE Block 来获取通道注意力。其具体过程如下:首先采用 Global Average pooling 作为 Squeeze 操作,将二维的特征通道变成一个实数,使得该实数在某种程度上具有全局感受野。接着通过两个全连接层去建模通道间的相关性,并输出与输入特征同样数目的权重。其中,在两个全连接层中间引入 ReLU 函数进行激活,使其具有更多的非线性,从而更好地拟合通道间的复杂相关性。最后通过 Sigmoid 激活函数获得 0 到 1 之间的权重 S 。

将上述归一化后的权重 S 加权到 U_i^s 每个通道的特征上:

$$X_{En}^{concat} = U_i^s \cdot S_c \quad (6)$$

其中, S_c 表示通道的权重值; X_{En}^{concat} 为经过了双重注意力机制

后得到的特征。最后对该特征图进行非线性映射操作,实现信息的整合,增加更多的非线性因素与提高最终的泛化能力:

$$O_i = \text{Relu}(GN(f^{1 \times 1}(X_{En}^{concat}))) \quad (7)$$

其中, $f^{1 \times 1}$ 表示标准卷积操作,卷积核大小为 1×1 ,通道数为 c ; O_i 为双重注意力机制最终的输出特征。

3.2.2 基于 DenseASPP 改进的中间层模块

肝脏分割任务中,由于输入的 CT 图像具有较高的分辨率,用于此任务的网络应具有较大的感受野。卷积神经网络中往往会采用池化操作来获取更大感受野的特征,但这会降低特征图的分辨率,导致细节信息的丢失。为了保持在分辨率不变的情况下获取更大的感受野,本文通过在网络中间层添加一系列扩张卷积组合级联来使后面的神经元获取更大的感受野,并采用密集连接的方式整合多尺度特征信息。下面将对该模块进行详细介绍。

基于 DenseASPP 改进的中间层模块 (dASM) 的结构如图 4 所示。首先通过采用 4 个膨胀率 d 分别为 3, 6, 12 和 18, 核大小 k 为 3, 通道数 c 为 64 的卷积层构成特征金字塔。其等效接受域计算式为:

$$R = (d-1) \times (k-1) + k \quad (8)$$

进一步地,卷积层叠加的接受域计算式为:

$$R = R_{c1} + R_{c2} - 1 \quad (9)$$

其中, R_{c1} 和 R_{c2} 分别为第一个卷积层接受域和第二个卷积层接受域。

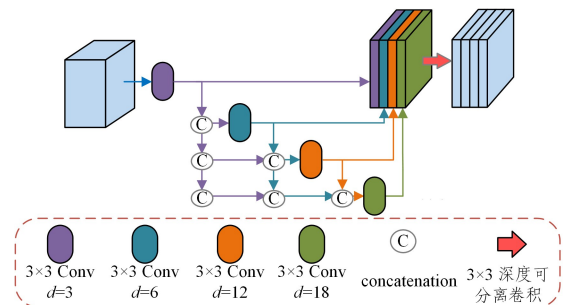


图 4 基于 DenseASPP 改进的中间层模块

Fig. 4 Improved central layer module based on DenseASPP

通过以上扩张卷积组合获得的大感受野,使其能够编码更高层次的语义,并且通过密集的特征连接,使得中间层特征图上的神经元能够从多个尺度对语义信息进行编码。

然后将上述扩张卷积组合与一个 $256 \times 3 \times 3$ 的深度可分离卷积结合,保证在参数数量和运算成本较低的情况下实现信息的整合。

3.2.3 全局注意力上采样模块

现有 U 型架构网络中,大部分解码结构都采用跳跃连接策略来直接融合多尺度信息,以帮助逐步恢复图像信息。这类方法虽然取得了一定的效果,但也带来了计算消耗增大以及信息冗余等问题。受文献[31]的启发,本文根据高层特征富含语义信息,能引导低层特征进行学习的特点,通过采用高层特征全局池化提供的全局信息作为引导选择低层特征。下面将对该模块进行详细介绍。

全局注意力上采样模块 (GAM) 的结构如图 5 所示。首先,通过对高层特征采用全局池化操作来获取全局信息,执行 1×1 卷积、组归一化和 ReLU 非线性激活操作,以减少 CNN

特征映射的通道数与降低计算消耗。然后,与低层特征逐元素相乘,用于将注意力放到目标区域上,缓解信息冗余问题,从而获得更精确的特征表示。最后,将高层特征与低层特征在通道维度进行拼接。

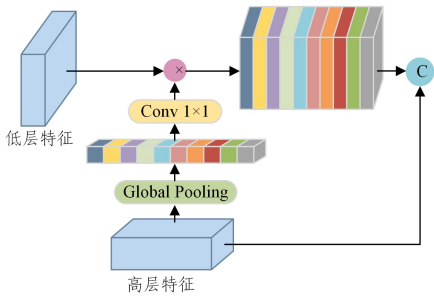


图5 全局注意力上采样模块

Fig. 5 Global attention up-sampling module

3.3 损失函数

本文网络在4个不同层次的解码特征上进行监督学习。具体来说,本文采用区域插值的方式对语义标签图下采样,将其尺寸分别调整为 256×256 , 128×128 , 64×64 和 32×32 ,用来进行深层监督训练。对4个上采样层的输出特征分别使用 1×1 卷积和Softmax函数计算每个像素的分类概率,再将二分类Focal Loss^[32]作为损失函数添加到上述4个语义层次中的每个层次。总体损失函数描述为:

$$L_{total} = \frac{1}{n} \sum_{i=1}^n -\alpha_i (1 - P_i)^{\gamma} \log P_i \quad (10)$$

其中:

$$P_i = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \quad (11)$$

通过上述损失函数,可以对网络模型进行端到端的训练。同时利用跳跃连接将分层多尺度注意力下采样模块的信息传递到解码层,使得在反向传播阶段,辅助损失产生的梯度可以更容易传递到对应的编码层,从而有效避免网络的梯度消失问题,提升整体的参数更新能力。

3.4 图像后处理

3.4.1 全连接条件随机场

本文网络在解码器部分采用上采样操作,使特征图恢复至原图尺寸。然而,这个过程造成了特征信息丢失,从而导致分类目标边界模糊问题。为了提高结果的精度,本文采用全连接条件随机场作为图像后处理方法^[33-34]。该方法通过结合原始肝脏CT图像中像素之间的关系,来对卷积神经网络所得到的像素分类结果进行处理,优化图像中粗糙和不确定性的标记,修正细碎的错分区域,以获取更细致的分割边界。

具体来说,本文方法将网络模型的输出概率分布图作为全连接条件随机场的一元势能,而原始肝脏CT图像则提供二元势能中的位置和颜色信息。其中,全连接条件随机场的能量函数如下:

$$E(x) = \sum_{i=1}^N \psi_U(x_i) + \sum_E \psi_p(x_i, x_j) \quad (12)$$

能量函数中的第一项为一元势能函数,用于衡量当像素点*i*的观测值为 y_i 时,该像素点属于类别标签 x_i 的概率。能量函数中的第二项 $\psi_p(x_i, x_j)$ 为二元势能函数,用于衡量两事件同时发生的概率 $P(x_i, x_j)$,其描述了像素之间的关系,

并将颜色和相对距离最近的像素归为一类,计算式如下:

$$\psi_p(x_i, x_j) = U(x_i, x_j) \sum_{m=1}^M \omega^m K_G^m(f_i, f_j) \quad (13)$$

其中, $U(x_i, x_j)$ 为标签兼容项,它约束了像素间传导的条件,只有相同标签条件下,能量才可以相互传导; ω^m 为权值参数,用于平衡函数; $K_G^m(f_i, f_j)$ 为特征函数,表达式为:

$$K_G^m(f_i, f_j) = \exp\left(-\frac{1}{2}(f_i, f_j)^T \Lambda^{(m)}(f_i - f_j)\right) \quad (14)$$

其中, f_i 和 f_j 代表像素*i*与像素*j*的特征向量。

通过上述方法对肝脏CT图像进行分割,当相似区域的像素点被判别为不同类时,能量函数值会变得较大;当存在差异的区域被判别为同一类时,也会产生较大的能量函数值。通过多次迭代,使能量函数值最小化来获得最终结果,这样有效利用了整个图像的信息来细化分割边缘,从而使分割精度得到提高。

3.4.2 孔洞填充

由于医学图像会不可避免地产生斑点噪声,导致分割的肝脏区域会伴有少量孔洞出现。为了进一步优化最终分割结果,本文采用基于漫水填充法^[35]的孔洞填充算法,来填补小孔洞噪声。具体过程如算法1所示。

算法1 孔洞填充算法

输入:原始二值图像 **I**, 阈值 **T**

输出:最终分割结果 **FPre**

1. $I' \leftarrow$ 创建像素初始值与 **I** 相同的图像
2. ROI Area \leftarrow 确定图像 I' 的 ROI 区域
3. Sp \leftarrow 初始化图像 I' 的种子点
4. WHILE Sp 未被遍历完
5. FOR pixel \in 4 邻域 DO
6. IF $| \text{Sp 像素值} - \text{pixel 像素值} | < T$
7. New pixel \leftarrow 将 pixel 添加进 ROI 区域
8. END IF
9. END FOR
10. Sp \leftarrow 将 New pixel 设置为新的种子点
11. END WHILE
12. FPre \leftarrow 原始图像 **I** | 种子填充后图像 I' 取反
13. RETURN FPre

4 实验结果与分析

4.1 实验数据集与评估指标

4.1.1 数据集

本文使用由专家标注的临床肝脏CT图像分割数据集,共包含420张成对的肝脏CT图像和标签。在实验评估中,使用400个实例进行训练(其中80张作为验证集),使用20个实例进行测试。

4.1.2 评估指标

本文使用平均像素精度、平均交并比和Dice相似系数这3种常用的语义分割评价指标对算法性能进行验证,具体表达式如下。

平均像素精度(Mean Pixel Accuracy),如式(15)所示:

$$MPA = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{TP+FP} \quad (15)$$

平均交并比(Mean IoU),如式(16)所示:

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{k+1+FN+FP+TP} \quad (16)$$

Dice相似系数公式,如式(17)所示:

$$Dice(P, T) = \frac{2TP}{FP + 2TP + FN} \quad (17)$$

其中, TP 表示被判定为正样本, 事实上也是正样本; FP 表示被判定为正样本, 但事实上是负样本; FN 表示被判定为负样本, 但事实上是正样本。

4.2 网络参数设置

本文在开源的深度学习框架 Tensorflow 和 Keras 上实现并训练所提出的网络模型。首先对数据集进行窗宽(window width)和窗位(window level)调整, 以获得肝脏区域的最佳显示效果。然后对所有肝脏 CT 图像通过区域插值的方式将图片大小调整为 256×256 作为网络输入, 并且采用同样的方式将语义标签图的尺寸调整为 $256 \times 256, 128 \times 128, 64 \times 64$ 和 32×32 , 用于进行深层监督训练。最后将 4 个解码层的输出和标签间的二分类 Focal Loss 求和并取平均值, 作为网络最终的损失函数。本文网络在 NVIDIA GeForce GTX 1050Ti GPU 上使用自适应矩估计(Adam)^[36] 优化算法进行训练, 训练时批处理大小设置为 2, 初始学习率设置为 0.001。

4.3 实验结果分析

本文通过评估在数据集上不同实验方案的结果来验证所提方法的有效性。首先, 为了研究 MD-AUNet 中两个注意力模块的性能, 本文通过两个注意力模块消融实验结果来评估其影响。其次, 为了验证改进中间层模块的高效性, 将其与现有方法进行了对比。此外, 为了分析其他可能影响本文网络模型性能的因素, 本文对编码器、特征融合层次这两个因素进行了评估。更进一步, 为了证明本文网络 MD-AUNet 的高性能, 将其与最先进方法进行了比较。最后, 为了评估后处理的有效性和普适性, 本文比较了有无后处理以及后处理方法融入其他网络两种情况下的性能。

4.3.1 两种注意力模块的性能分析

本节讨论本文提出的分层多尺度注意力下采样模块(HAM)和全局注意力上采样模块(GAM)在网络模型中的作用。本文通过替换网络模型中两个特征融合模块来验证其对最终分割性能的影响。实验结果如表 1 所列, 本文首先评估了网络在去除两个特征融合模块情况下的性能, 具体在保持网络其他因素不变的情况下, 采用拼接操作来代替所有特征融合模块, 在表 1 中用 Concat 表示此网络。其分割的平均像素精度、平均交并比和 Dice 相似系数分别为 92.1%, 84.8% 和 83.3%, 比本文网络 MD-AUNet 分别减少了 5.1%, 10.0% 和 11.6%。这说明本文提出的两个注意力模块显著提升了最终的语义分割准确度。

表 1 两种注意力模块对网络性能的影响

Table 1 Influence of two attention modules on network

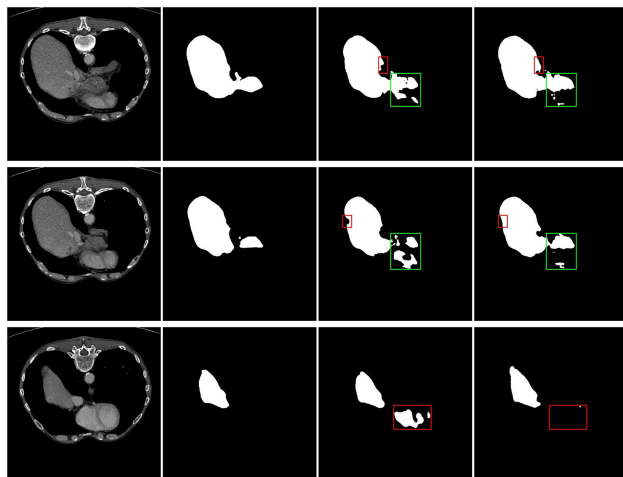
Method	HAM	GAM	MPA/%	MIOU/%	Dice/%
Concat	×	×	92.1	84.8	83.3
MD-AUNet	×	√	94.3	85.0	83.4
MD-AUNet	√	×	95.1	86.4	85.4
MD-AUNet	√	√	97.2	94.8	94.9

进一步地, 本文评估了网络在分别去除分层多尺度注意力下采样模块和全局注意力上采样模块情况下的性能。

(1) 分层多尺度注意力下采样模块

如表 1 中的第 2 行所示, 去除分层多尺度注意力下采样模块后, 网络的平均像素精度, 平均交并比和 Dice 相似系数

分别降低了 2.9%, 9.8% 和 11.5%。图 6 进一步给出了使用或去除分层多尺度注意力下采样模块的分割结果。通过可视化对比结果可以看出, 使用分层多尺度注意力下采样模块的网络能够准确编码位置信息, 并且在灰度变化的位置产生更高的响应, 获取到更精准的细节轮廓信息。而对于灰度值相近而类别不同的区域, 本模块能利用注意力协同优化, 融合不同尺度互补信息的优势, 更有效地适应复杂的分割情况。



(a) 原始图像 (b) 语义标签 (c) 未使用 HAM (d) 使用 HAM

图 6 分层多尺度注意力下采样模块的对比结果

Fig. 6 Comparison of hierarchical multi-scale attention down-sampling modules

为了进一步验证本模块中双重注意力机制对特征融合效果的影响, 本文在其他因素不变的情况下, 分别用 Attention Gate(空间注意力机制)、SE Block(通道注意力机制)、scSE(并行的双重注意力机制)、CBAM(串行的双重注意力机制)替换本模块中的双重注意力机制来进行结果的对比。如表 2 所列, 本文的双重注意力机制在分割性能上具有较大的优势, 这得益于融合不同尺度特征时的独特设计。其中空间注意力模块将高级特征的语义信息引入到低级特征, 以此弥补了两者间的语义差距, 有效地融合了特征。通道注意力模块则对特征通道间的相关性进行建模, 将重要的特征进行强化, 以此提升整体性能。

表 2 注意力机制的性能对比

Table 2 Performance comparison of attention mechanisms

注意力机制	Params/MB	MPA/%	MIOU/%	Dice/%
Attention Gate ^[22]	1.96	95.9	88.6	88.2
SE Block ^[23]	2.60	93.1	85.5	84.5
scSE ^[24]	2.61	93.4	81.6	80.1
CBAM ^[25]	2.64	93.4	87.2	86.6
Our Method	2.38	97.2	94.8	94.9

(2) 全局注意力上采样模块

如表 1 中的第 3 行所列, 去除全局注意力上采样模块后, 网络的平均像素精度, 平均交并比和 Dice 相似系数分别降低了 2.1%, 8.4% 和 9.5%。图 7 进一步给出了使用或去除全局注意力上采样模块的分割结果。通过可视化对比结果可以看出, 使用全局注意力上采样模块的网络在分割的肝脏宏观形状和局部细节上均得到了提升。这表明本模块可以有效利用

高层特征的全局上下文来引导低层特征进行空间信息选择,从而减少有用信息的丢失与无用信息的冗余。

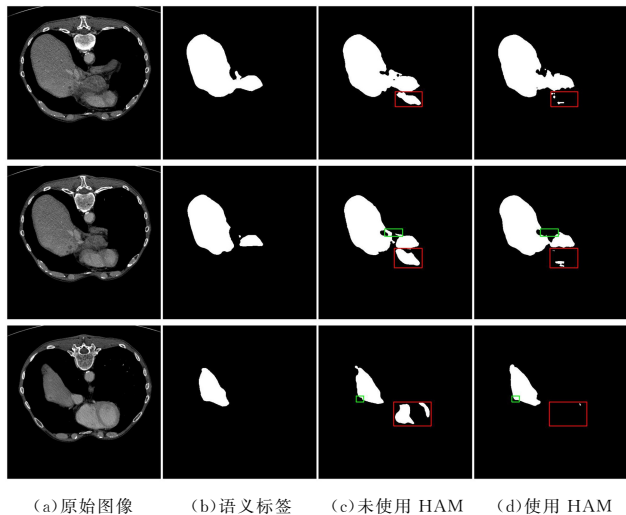


图7 全局注意力上采样模块的对比结果

Fig. 7 Comparison of global attention up-sampling modules

4.3.2 中间层模块的性能分析

本节讨论本文提出的基于 DenseASPP 改进的中间层模块(dASM)在网络模型中的作用。本文分别通过两个 $256 \times 3 \times 3$ 的卷积层、两个 $256 \times 3 \times 3$ 的深度可分离卷积层以及 DenseASPP 模块替换中间层模块来进行结果的对比。

如表 3 所列,替换中间层模块后,评估指标均出现不同程度的下降。并且在计算消耗上,本文方法仅比直接采用深度可分离卷积作为中间层方法的参数略多。这表明本文方法能够在不引入大量参数的前提下,使网络的感受野更大,更有效地整合多尺度特征信息。

表3 中间层模块对网络性能的影响

Table 3 Performance comparison of central layer modules

中间层	Params/MB	MPA/%	MIoU/%	Dice/%
普通卷积	2.73	93.4	92.2	92.2
深度可分离卷积	1.97	94.3	93.3	93.4
DenseASPP 模块	6.65	94.2	85.6	84.7
Our Method	2.38	97.2	94.8	94.9

4.3.3 其他影响因素分析

本节进一步分析了可能影响网络模型性能的其他因素,分别讨论了编码器、特征融合层数对最终分割性能的影响。

(1) 编码器

本文为了验证在不同编码器情况下网络模型的性能,分别将 ResNet(深度扩展)、InceptionNet V1(宽度扩展)、EfficientNet B0(深度、宽度、分辨率扩展)中部分模块融入到本网络中作为编码器来进行结果的对比。如表 4 所列,使用 ResNet 中的 Res Block 作为编码器能够在较低的计算消耗下获取较好的性能。

表4 编码器对网络性能的影响

Table 4 Impact of encoder on network performance

Encoder	Params/MB	MPA/%	MIoU/%	Dice/%
ResNet ^[12]	2.38	97.2	94.8	94.9
InceptionNet V1 ^[37]	1.62	88.3	87.1	86.3
EfficientNet B0 ^[38]	4.09	92.4	89.9	89.8

具体来说,相比 InceptionNet V1 中的 Inception Module,其参数量有所增加,却带来了显著的准确度提升。而相比 EfficientNet B0 中的 MBConv Block,其则在准确度和参数量上均有较大优势。(2)去除某些特征融合层

为了验证多层次特征融合的效果,本文通过去除网络中不同层次融合层来评估网络性能。如表 5 所列,去除任何层次的特征融合层均会降低网络的最终分割性能。这表明本文网络融合的各个层次特征,能够有效利用特征之间的互补信息,学习到更具鲁棒性的特征表示。

表5 特征层次对网络性能的影响

Table 5 Impact of feature layer on network performance

特征层次	MPA/%	MIoU/%	Dice/%
不融入 Res-1 层	93.7	91.9	91.8
不融入 Res-2 层	95.3	86.5	85.6
不融入 Res-3 层	94.1	91.8	91.8
不融入 Res-4 层	95.1	83.8	82.6
MD-AUNet	97.2	94.8	94.9

4.3.4 与其他算法的对比结果

本文将所提算法与现有算法在专家标注的肝脏 CT 图像数据集上进行对比实验。为了比较的公平性,本文使所有算法的编码与解码层次保持相同。如表 6 所列,本文算法在 3 种评估指标上均优于现有方法。具体来说,本文网络 MD-AUNet 在测试集上的平均像素精度、平均交并比和 Dice 相似系数分别为 97.2%,94.8%和 94.9%;本文方法的平均像素精度、平均交并比和 Dice 相似系数相比 U-Net 算法分别提高了 5.8%,4.6%和 4.9%;相比 UNet++算法,分别提高了 3.2%,2.9%和 3.1%;相比 UNet3++算法,分别提高了 1.9%,2.0%和 2.2%;相比 Attention U-Net 算法,分别提高了 3.1%,1.8%和 1.9%。上述算法均采用带有跳跃连接的编码-解码架构,本文方法优于同层次其他网络归功于多尺度特征融合时采用的结构设计与注意力机制,可以有效地融合高、低层特征,获取更精准的特征表示。

表6 与其他算法的性能比较

Table 6 Performance comparison with other algorithms

Method	MPA/%	MIoU/%	Dice/%
U-Net ^[14]	91.4	90.2	90.0
UNet++ ^[15]	94.0	91.9	91.8
UNet3++ ^[16]	95.3	92.8	92.7
Attention U-Net ^[22]	94.1	93.0	93.0
Ours	97.2	94.8	94.9

此外,为了更直观地反映肝脏区域的分割效果,图 8 给出了不同算法在专家标注的肝脏 CT 图像数据集上的可视化对比结果。由图 8 可以看出,本文算法获得了更为精确的分割结果。具体来说,其他算法在分割结果上均出现了不同程度的肝脏区域丢失现象,这源自算法本身无法有效地捕获感受野以外的肝脏区域位置信息。本文算法则充分利用了高层特征与低层特征的互补性,使上述问题得到了缓解。此外,上述算法均不同程度地出现了混淆像素点被分类错误的问题。而本文通过在特征融合模块中引入注意力机制,来帮助网络着重关注有效的信息区域,以此来获得优于其他算法的表现。

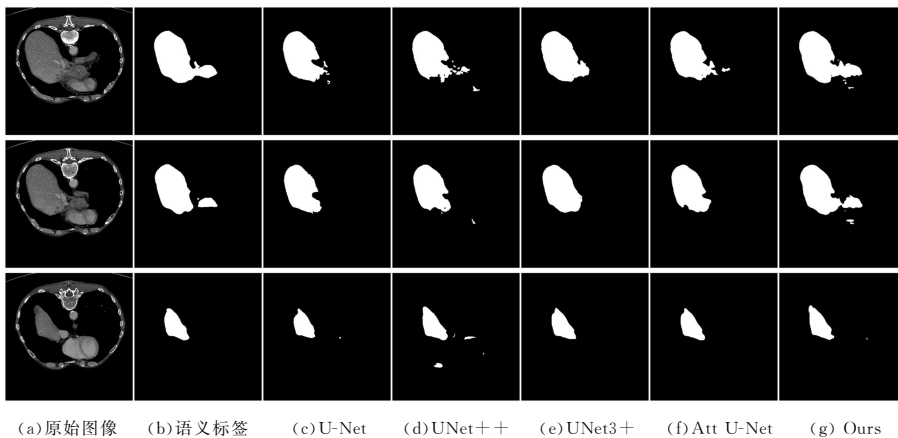


图8 可视化分割结果对比

Fig. 8 Comparison of visual segmentation results

4.3.5 图像后处理的性能分析

本节讨论本文所采用的图像后处理方法在图像分割中的作用。根据肝脏CT图像粗分割结果,观察到本文网络容易出现混淆像素点被集中分类错误的问题,从而导致性能的降低。如表7中的第5行所示,经过后处理步骤,本文方法的平均像素精度、平均交并比和Dice相似系数分别提高了0.4%,0.6%和0.6%。为了更直观地感受后处理方法的有效性,在图9中进一步展示了使用或去除后处理的分割结果。结果表明,在采用本文图像后处理方法后,有效地去除了肝脏分割中的假阳性区域以及填充了真阳性区域,使最终的分割结果得到了进一步的细化。

表7 图像后处理对分割结果的影响

Table 7 Performance comparison of post-processing methods

(单位:%)

Post-processing	MPA		MIoU		Dice	
	✓	✗	✓	✗	✓	✗
U-Net ^[14]	92.4	91.2	91.5	90.2	91.6	90.0
UNet++ ^[15]	94.1	94.0	92.5	91.9	92.3	91.8
UNet3+ ^[16]	95.5	95.3	93.2	92.8	93.0	92.7
Attention U-Net ^[22]	94.2	94.1	93.1	93.0	93.1	93.0
Ours	97.6	97.2	95.4	94.8	95.5	94.9

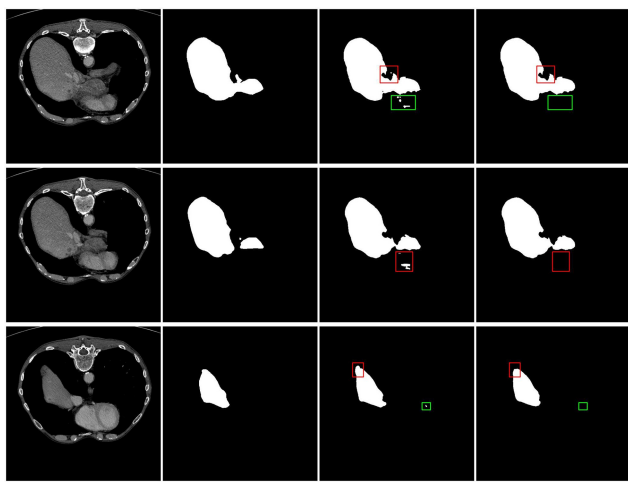


图9 图像后处理结果对比

Fig. 9 Results comparison of image post-processing

为了进一步探索本文图像后处理方法的普适性,通过在UNet,UNet++,UNet3+以及 Attention U-Net 算法得到的初步分割结果基础上,采用本文图像后处理方法,对粗分割

结果进行进一步细化。从表7中可以看出,这些算法在采用本文图像后处理方法后,分割结果均得到了进一步的提升,这表明该方法具有较好的普适性。

结束语 本文提出了一种高效的肝脏分割方法,其中包含两种创新性的特征融合模块:分层多尺度注意力下采样模块和全局注意力上采样模块。通过将它们集成到一个U型编码-解码网络中,有效地融合多尺度特征。其中,分层多尺度注意力下采样模块通过分层级的双重注意力机制充分利用不同尺度特征的互补信息,来提取富含空间信息的特征表示。全局注意力上采样模块通过高层特征的全局上下文对低层特征进行引导,选择更为精确的空间信息。其次,在网络中间层引入了基于DenseASPP改进的中间层模块,使网络获取更大的感受野。此外,本文证实了采用残差块作为编码器能有效的提取特征信息,以及融合多个层次特征可以更有效地提升网络最终的分割精度。更进一步地,本文针对肝脏粗分割结果,提出了一种简洁有效的后处理方法,以此来细化肝脏分割结果。实验结果表明,本文方法在所采用的肝脏数据集上的分割性能优于其他现有方法。

本文方法依然有进一步提升的空间,如不能对小目标肝脏区域进行精确分割等。在未来研究中,将考虑轻量级技术在网络中的应用以提升整体的效率,以及结合多任务训练策略进一步提升分割精度。

参考文献

- [1] HE L, WU Q. Automatic liver segmentation based on three-dimensional convolutional neural network[J]. Chinese Journal of Medical Physics, 2018, 35(6): 581-589.
- [2] WANG L D, WANG J, HU C J, et al. Liver Segmentation by Using an Optimal Framework for CT Images[J]. Chinese Journal of Computers, 2016, 39(7): 1477-1489.
- [3] CASTRO P, HUERGA C, LA GLARÍA, et al. Target volume segmentation of PET images by an iterative method based on threshold value[J]. Rev Esp Med Nucl Imagen Mol, 2014, 33(6): 331-339.
- [4] RUSKÓ L, BEKES G, FIDRICH M. Automatic segmentation of the liver from multi- and single-phase contrast-enhanced CT images[J]. Medical Image Analysis, 2009, 13(6): 871-882.
- [5] WU Y, HE C. Indirectly regularized variational level set model for image segmentation[J]. Neurocomputing, 2016, 171: 194-208.
- [6] LIM S J, JEONG Y Y, LEE C W, et al. Automatic segmentation

- of liver in CT images using the watershed algorithm base on morphological filtering[C]// Proceedings of SPIE. 2004;1658-1666.
- [7] VORONTSOV E,ABI-JAOUDEH N,KADOURY S. Metastatic liver tumor segmentation using texture-based omni-directional deformable surface models[C]// International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging. Springer,2014;74-83.
- [8] RACHANTA,SHAJI A. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods [J]. IEEE Press, 2012 (11): 2274.
- [9] LI Y,HARA S,SHIMURA K. A machine learning approach for locating boundaries of liver tumors in ct images[C]//18th International Conference on Pattern Recognition(ICPR'06). IEEE, 2006,1;400-403.
- [10] SHELHAMER E, LONG J, DARRELL T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2017, 39 (4): 640-651.
- [11] BI L,KIM J,KUMAR A, et al. Automatic Liver Lesion Detection using Cascaded Deep Residual Networks[J]. arXiv: abs/1704.02703,2017.
- [12] HE K,ZHANG X,REN S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition(CVPR). 2016;770-778.
- [13] SUN C, GUO S, ZHANG H, et al. Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs[J]. Artificial Intelligence in Medicine, 2017, 83: 58-66.
- [14] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. arXiv: abs/1505.04597,2015.
- [15] ZHOU Z, SIDDIQUEE M M R, TAJBAKHSH N, et al. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation [J]. IEEE Transactions on Medical Imaging, 2020, 39(6): 1856-1867.
- [16] HUANG H L. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation[C]// 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2020). 2020;1055-1059.
- [17] YU W, YANG K Y, YAO H X, et al. Exploiting the complementary strengths of multi-layer CNN features for image retrieval[J]. Neurocomputing, 2017, 237; 235-241.
- [18] WANG X L, LIU F, JIAO L C. Multiscale Bayesian Image Segmentation Fusing Context Information[J]. Chinese Journal of Computers, 2005, 28(3): 386-391.
- [19] BEN-COHEN A, DIAMANI T, KLANGE D, et al. Fully Convolutional Network for Liver Segmentation and Lesions Detection [C]// International Workshop on Deep Learning in Medical Image Analysis. 2016;77-85.
- [20] CHRIST P F. Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields[C]// Medical Image Computing and Computer-Assisted Intervention(MICCAI 2016). 2016;415-423.
- [21] SCHNURR A K, SCHAD L R, ZÖLLNER F G. Sparsely Connected Convolutional Layers in CNNs for Liver Segmentation in CT [C] // Bildverarbeitung für die Medizin 2019. Springer Vieweg, Wiesbaden, 2019; 80-85.
- [22] OKTAY O. Attention U-Net: Learning Where to Look for the Pancreas[J]. arXiv abs/1804.03999,2018.
- [23] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-Excitation Networks[C]// IEEE Transactions on Pattern Analysis and Machine Intelligence. 2020;2011-2023.
- [24] ROY A G, NAVAB N. Concurrent Spatial and Channel 'Squeeze & Excitation' [C]// Fully Convolutional Networks. Medical Image Computing and Computer Assisted Intervention(MICCAI 2018). 2018;421-429.
- [25] WOO S, PARK J. CBAM: Convolutional Block Attention Module[C]// European Conference on Computer Vision. 2018;3-19.
- [26] FU J, LIU J, TIAN H, et al. Dual Attention Network for Scene Segmentation[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). 2019;3141-3149.
- [27] LI H, XIONG P. Pyramid Attention Network for Semantic Segmentation[J]. arXiv, abs/1805.10180, 2018.
- [28] SINHA A, DOLZ J. Multi-Scale Self-Guided Attention for Medical Image Segmentation[C]// IEEE Journal of Biomedical and Health Informatics. 2021;121-130.
- [29] TAO A, SAPRA K, CATANZARO B. Hierarchical Multi-Scale Attention for Semantic Segmentation [J]. arXiv: abs/2005.10821,2020.
- [30] DUAN L J, SUN Q C, QIAO Y H, et al. Attention-Aware and Semantic-Aware Network for RGB-D Indoor Semantic Segmentation[J]. Chinese Journal of Computers, 2021, 44(2): 275-291.
- [31] XIONG L H. Pyramid Attention Network for Semantic Segmentation[J]. arXiv, abs/1805.10180, 2018.
- [32] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Trans PatternAnalMach Intell, 2020, 42(2): 318-327.
- [33] SHU H H, XIA H X, GA L H, et al. Water Body Extraction of High Resolution Remote Sensing Image based on Improved U-Net Network[J]. Journal of Geo-information Science, 2020, 22(10): 2010-2022.
- [34] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv:1412.7062,2014;97-105.
- [35] CHEN J X, JIA X M. Real-time Color Object Recognition Method Based on Flood Fill Algorithm[J]. Chinese Journal of Computer Simulation, 2012, 29(3): 4-9.
- [36] DIEDERIK P K, JIMMY B, YOSHUA B, et al. A Method for Stochastic Optimization[C]// ICLR(Poster). 2015.
- [37] SZEGEDY C. Going deeper with convolutions[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015;1-9.
- [38] TAN M X, LE Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks [J]. arXiv: abs/1905.11946, 2019.



HUANG Yang-lin, born in 2000, undergraduate. His main research interests include deep learning and medical image processing.



HU Kai, born in 1984, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include machine learning, pattern recognition, bioinformatics, and medical image processing.