

集成全尺度融合和循环注意力的医学图像分割网络

单昕昕, 李凯, 文颖

引用本文

单昕昕, 李凯, 文颖. [集成全尺度融合和循环注意力的医学图像分割网络](#)[J]. 计算机科学, 2024, 51(5): 100-107.

SHAN Xinxin, LI Kai, WEN Ying. [Medical Image Segmentation Network Integrating Full-scale Feature Fusion and RNN with Attention](#) [J]. Computer Science, 2024, 51(5): 100-107.

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)

[面向前提选择的新型图约简表示与图神经网络模型](#)

New Graph Reduction Representation and Graph Neural Network Model for Premise Selection

计算机科学, 2024, 51(5): 193-199. <https://doi.org/10.11896/jsjcx.230300193>

[基于特征注意力提纯的显著性目标检测模型](#)

Salient Object Detection Based on Feature Attention Purification

计算机科学, 2024, 51(5): 125-133. <https://doi.org/10.11896/jsjcx.230300018>

[基于渐进式多尺度Transformer的图像去雾算法](#)

Multi Scale Progressive Transformer for Image Dehazing

计算机科学, 2024, 51(5): 117-124. <https://doi.org/10.11896/jsjcx.230300049>

[基于跨模态信息过滤的视觉问答网络](#)

Cross-modal Information Filtering-based Networks for Visual Question Answering

计算机科学, 2024, 51(5): 85-91. <https://doi.org/10.11896/jsjcx.230300202>

[基于语义扩充和HDGCN的虚假新闻联合检测技术](#)

Unified Fake News Detection Based on Semantic Expansion and HDGCN

计算机科学, 2024, 51(4): 299-306. <https://doi.org/10.11896/jsjcx.230700170>

集成全尺度融合和循环注意力的医学图像分割网络

单昕昕 李凯 文颖

华东师范大学通信与电子工程学院上海市多维度信息处理重点实验室 上海 200241

(51184506005@stu.ecnu.edu.cn)

摘要 深度学习中的编解码网络在图像特征提取和分层特征融合方面具有卓越的性能,常被用于医学图像分割。但是,目前主流的编解码网络分割方法仍面临编码和解码阶段单一网络挖掘的图像特征信息不足,以及仅使用简单的跳跃连接而无法充分利用全尺度特征包含的粗粒度信息和细粒度信息等问题。为了解决上述问题,提出了一种集成全尺度融合和循环注意力的医学图像分割网络。首先,在 U-Net 编码器中加入了结合多层感知机(MLP)的卷积 MLP 模块来提取图像的全局特征信息,用于扩大编码器的特征感受野。其次,通过全尺度特征融合模块使得各尺度跳跃连接特征进行粗粒度信息和细粒度信息的有效融合,减小各尺度跳跃连接特征间的语义差异,突出图像的关键特征信息。最后,解码器通过提出的结合循环神经网络(RNN)和注意力机制的循环注意力解码模块(RADU)来逐级精细化图像特征信息,加强特征提取的同时避免信息冗余,并得到高精度分割结果。在 4 个数据集上将所提方法与主流较优的方法进行比较,所提方法在像素精度和骰子相似系数两个指标上的图像分割精度均有提高。因此,所提出的用于医学图像分割的编解码网络利用全尺度特征融合模块和循环注意力解码模块,能够获得较优异的高精度分割结果,并且模型具有良好的噪声鲁棒性和抗干扰能力。

关键词: 医学图像分割;编解码网络;多层感知机;全尺度特征融合;注意力机制;循环神经网络

中图分类号 TP391.7

Medical Image Segmentation Network Integrating Full-scale Feature Fusion and RNN with Attention

SHAN Xinxin, LI Kai and WEN Ying

Shanghai Key Laboratory of Multidimensional Information Processing, School of Communication and Electronic Engineering, East China Normal University, Shanghai 200241, China

Abstract The encoder-decoder network in deep learning has excellent performance in image feature extraction and hierarchical feature fusion, and is often used in medical image segmentation. However, the current mainstream encoding and decoding network segmentation methods still face two problems: 1) in encoding and decoding stages, image feature information mined by a single network may be insufficient; 2) encoder-decoder networks using simple skip connections cannot fully exploit the contextual information of full-scale features. Therefore, aiming at the shortcomings of the existing methods, an encoder-decoder network integrating full-scale feature fusion and RNN with attention for medical image segmentation is proposed. At first, the convolutional multi-layer perceptron(MLP) module combined with MLP is introduced in U-Net encoder to further expand the feature receptive field of the encoder. Secondly, by the full-scale feature fusion module, the skip connection features of each scale are effectively fused with coarse-grained information and fine-grained information. This operation reduces the semantic difference between the skip-connection features of each scale and highlights the key feature information of the image. Finally, the decoder refines the image feature information level by level through the proposed recurrent attention decoding module(RADU) combining recurrent neural network(RNN) and attention mechanism, which strengthens feature extraction while avoiding information redundancy, and obtains the final segmentation results. The proposed method is compared with the mainstream algorithms on BrainWeb, MR-brainS, HVSMR and Choledoch datasets, the image segmentation precision is improved in pixel accuracy and dice similarity coefficient. Therefore, experimental results show that by introducing the full-scale feature fusion module and the proposed RADU, the proposed method can achieve excellent segmentation results in image segmentation applications and has good noise robustness and

到稿日期:2023-04-16 返修日期:2023-08-16

基金项目:国家自然科学基金(62273150);上海市自然科学基金(22ZR1421000);上海市优秀学术/技术带头人计划项目(21XD1430600);上海市科学技术委员会资助项目(22DZ2229004)

This work was supported by the National Natural Science Foundation of China(62273150), Shanghai Natural Science Foundation(22ZR1421000), Shanghai Outstanding Academic Leaders Plan(21XD1430600) and Science and Technology Commission of Shanghai Municipality(22DZ2229004).

通信作者:文颖(ywen@cs.ecnu.edu.cn)

anti-interference ability.

Keywords Medical image segmentation, Encoder-Decoder network, Multi-layer perceptron, Full-scale feature fusion, Attention mechanism, Recurrent neural network

1 引言

医学图像分割任务是对图像中的各类解剖结构或者病理区域进行像素级的分类,在神经科学和疾病诊断中发挥着关键作用。在医疗影像学诸多技术中,核磁共振成像(Magnetic Resonance Imaging, MRI)具有图像较清晰、多模态、多方位和多平面的成像优势,使得人体中的组织解剖关系和相邻关系能清楚地呈现,进而充分提高医生的诊断效率。但是,依靠专家手动进行图像注释的传统方式既繁琐又耗时,因此,探索全自动医学图像分割的方法引起了人们的兴趣。现有的医学图像分割手段可以大致分为两类:基于传统机器学习的分割方法和基于深度神经网络的分割方法。基于传统机器学习的分割方法有 K-means 聚类算法^[1]和模糊 C-means 等^[2-3],这些方法在某些完整和高对比度图像(如核磁共振脑切片图像)分割上取得了较好的成果。但是,此类方法迭代仍需要高额的处理时间,并且 MR 图像中常出现的强度不均匀和图像噪声也影响算法的鲁棒性和泛化能力。

近年来,基于深度学习的分割方法在医学图像分割任务上的应用取得了显著的进展。深度学习分割方法的成功得益于深度神经网络强大的图像特征提取能力,且通过对图像特征信息的处理能得到精细化的分割结果。深度神经网络模型包括卷积神经网络(Convolutional Neural Networks, CNN)、多层感知机(Multilayer Perceptron, MLP)以及 Transformer^[4]等。现今卷积神经网络已经在医学图像分割领域取得了一定成就,例如全卷积网络(FCN)^[5]、U-Net^[6]和 SegNet^[7]等。其中,U-Net 是一个具有代表性的医学图像分割网络。U-Net 在编码阶段利用分级 CNN 提取图像特征,解码阶段则通过反卷积和跳跃连接实现编解码器的特征交互融合,从而得到较好的图像分割预测图。基于 U-Net,研究者们提出了很多改进网络,例如,结合注意力机制的 Attention U-Net^[8]、增加特征精细化处理的 DFM^[9]、融合时序监督^[10]和基于特征融合^[11]的分割模型,这些方法均使得图像分割的精度得到进一步提升。然而,基于卷积神经网络的方法中的卷积运算的内在局部性,导致它缺乏对全局特征相关性建模的能力。Transformer 和多层感知机能够利用自注意力机制和全连接网络来捕获图像的全局信息,能提取到卷积神经网络较难获取的图像特征。因此,在医学图像分割任务中,Transformer 通常会被嵌入到深度神经网络的编码器或者解码器中,通过扩大特征感受野,来提高分割精度。例如,TransUNet^[12]通过将 U-Net 编码器中最深层次的特征替换为从视觉 Transformer (ViT)^[13]中提取的特征;TCRNet^[14]将 ViT 和卷积神经网络一并作为分割网络的编码器;而 TrSeg^[15]则使用 Transformer 作为解码器模块,解码从 CNN 编码器中提取的特征并得到分割图像。Transformer 的兴起激发了研究者利用深度网络模型对图像分割领域进行探索。最新提出的 MLP-Mixer^[16]表明,多层感知机的图像特征提取能力不

亚于 Transformer。多层感知机结合卷积是一个新的突破点,例如卷积 MLP^[17]构建了卷积神经网络和 MLP 相结合的图像特征提取器,应用于图像分割任务时取得了不错的效果。UNetX^[18]集成了卷积和多层感知机的医学图像分割网络,取得了优于同类方法的分割预测结果,且模型的参数量明显降低。

然而,目前大多基于深度学习的医学图像分割方法仍存在不足之处。第一,网络输入图像的尺寸要求有限制。大多数基于 Transformer 的方法^[19-20]都需要数据集有大小规整方正的训练与测试图像,但是广泛收集的真实医学图像大多是尺寸任意的,而图像大小的调整将导致图像变形失真并影响分割结果。第二,编码和解码阶段中单一网络挖掘的图像特征信息可能不足^[21]。多层感知机网络提取的是全局特征信息,而 CNN 的优势是聚焦局部的信息。考虑到由 CNN 编码器提取的多尺度特征映射可以被视作序列数据,因此可以使用循环神经网络(Recurrent Neural Network, RNN)来挖掘特征映射的时间信息和上下文信息。故合理地结合多种网络可以获得更全面的特征信息,从而提升网络性能。第三,U-Net 的简单跳跃连接结构难以将各尺度跳跃连接特征中的粗粒度信息和细粒度信息进行有效融合,故各尺度跳跃连接特征间存在语义差异,从而限制了网络的分割性能^[22]。最近提出的 CDRN^[23]通过 RNN 将粗粒度特征信息进行存储传递,一定程度上弥补了尺度间的特征交互,但仍缺乏单一尺度特征对细粒度特征信息的融合。因此,需要探索一种能有效融合全尺度特征信息的方法。

针对上述问题,本文提出了集成全尺度融合和循环注意力的医学图像分割网络,具体贡献如下:

1)在编码器中嵌入多层感知机加强了网络对图像全局特征信息提取的能力;并且通过本文提出的全尺度特征融合模块对编码器提取的各尺度特征图进行粗粒度信息和细粒度信息的有效融合,使得跳连接特征之间的语义差异减小,突出了图像的关键特征信息。

2)提出的循环注意力解码模块(RADU)结合循环卷积神经网络和注意力机制,一方面使得解码器挖掘时间以及上下文特征信息的能力进一步提升,另一方面通过注意力机制来自适应地控制各自通道权重,加强了有效信息的利用。

3)在 BrainWeb^[24], MRBriens^[25], HVS MR2016^[26]和 Choleloch^[27] 4 个数据集上将所提方法与主流较优算法进行了对比实验。结果表明,所提出的方法在医学图像分割应用中显著提升了分割精度指标,在 Choleloch 数据集上的定量指标像素精度(PA)提升达 2.96%。

2 本文方法

对于医学图像分割任务,本文所提方法的网络结构主要分为 3 个部分:卷积 MLP 与 CNN 结合的特征编码器、全尺度特征融合模块和基于循环注意力解码模块(RADU)的特征

解码器,其结构图如图1所示。图像分割网络分为3个处理阶段。第一阶段,利用CNN骨干网络和卷积MLP网络作为特征编码器,将其分别用于提取输入图像局部的层次特征信息和全局的特征信息。第二阶段,将编码器提取的各层次特征利用提出的全尺度特征融合模块进行多尺度的特征交互

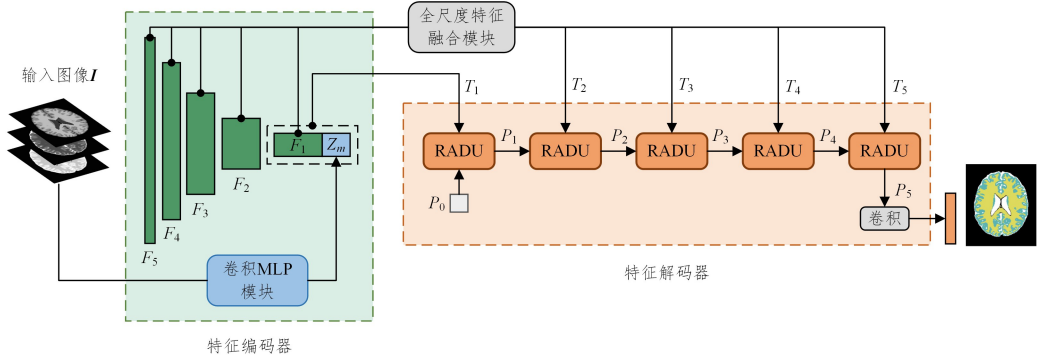


图1 图像分割网络结构

Fig. 1 Structure of the image segmentation network

2.1 卷积MLP与CNN结合的特征编码器

由于MLP具有提取特征同时保留图像全局特征信息的优势,CNN则在提取和恢复局部空间特征信息方面有十分出色的效果,因此将两者结合起来以获得更全面的特征信息是一个很好的解决方案。

给定多模态医学图像 I ,利用U-Net作为骨干网络提取5层不同尺度的特征映射,得到了具有不同分辨率的分层特征集合 $\{F_i\}_{i=1}^5 \in \mathbb{R}^{c_i \times h_i \times w_i}$,其中, c_i , h_i 和 w_i 分别表示特征图 F_i 的通道数、高度和宽度。随着网络层数的增加,其特征图 F_i 空间大小随网络层数减半,通道数量逐层加倍。

考虑到卷积CNN网络聚焦于图像局部信息的特征提取,本论文采用卷积MLP^[17]来补充图像特征的全局信息。卷积MLP模块如图2所示。输入图像 I 经过2次卷积组得到初始特征映射,该卷积组由 3×3 深度可分离卷积和池化层下采样组成,MLP模块和卷积下采样构成的卷积MLP阶段共有3个。将初始特征映射输入所提的卷积MLP模块,得到输出的特征映射。最终,将该输出的特征映射与U-Net编码器最后层得到的级联,作为特征编码器输出的特征映射。

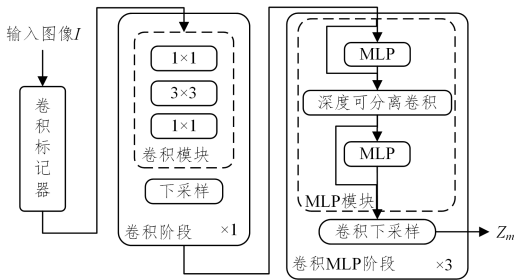


图2 卷积MLP模块的结构

Fig. 2 Structure of conv MLP

2.2 全尺度特征融合模块

图像分割任务中,跳跃连接通过融合粗粒度特征和细粒度特征来减小各尺度特征差异,有利于提升网络的分割精度^[22]。普通U-Net模型的直接跳跃连接由于缺少各尺度特征间的交互,无法进行粗粒度信息和细粒度信息的有效融合。

融合,生成包含全尺度特征的各级用于跳跃连接的特征映射。第三阶段,设计了一个包含循环卷积网络和注意力机制的RADU模块,用于细化不同尺度的特征映射,最后通过卷积模块将特征向量的通道数映射到分割的类数,从而得到最终的分割图像。

受UNet3+^[28]中特征交互方式的启发,本文提出了一个新的全尺度特征融合模块。由于 F_1 为最深层特征,与卷积MLP模块输出的 Z_m 级联得到 T_1 ,故模块从 T_2 开始进行特征融合处理。全尺度特征融合模块使跳跃连接的各级特征进行全尺度范围的交互,从而得到融合了各级特征后的跳连接特征 $\{T_i\}_{i=2}^5 \in \mathbb{R}^{c_i \times h_i \times w_i}$ 。

以 T_2 支路为例来阐述模块中各支路的工作流程,其结构如图3所示。其中, \hat{F}_1 即为特征 F_1 ,down和up分别表示将特征图下采样和上采样到 (h_2, w_2) (特征 F_2 的高度和宽度),null表示不做采样操作。 C 表示卷积操作,这里统一为 3×3 卷积,输出通道数为64。 χ 操作首先将输入的5层特征按通道级联叠加为单一特征映射(通道数为 $64 \times 5 = 320$),该特征映射再经过输出通道数为320的 3×3 卷积、批标准化操作和ReLU激活函数,得到融合特征映射 $\hat{F}_2 \in \mathbb{R}^{320 \times h_2 \times w_2}$ 。最后,为了将 \hat{F}_2 的特征通道数调整为 c_2 (特征 F_2 的通道数), \hat{F}_2 经过 ψ_2 操作,再与编码器输出特征 F_2 相加,得到输出特征 T_2 。 ψ_2 中包含一个输出通道数为 c_2 的卷积、批标准化操作和ReLU激活函数。

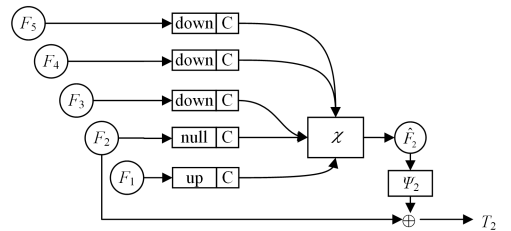


图3 全尺度特征融合模块中 T_2 支路的流程结构图

Fig. 3 Flow chart of T_2 branch in full-scale feature fusion module

模块各支路的融合特征 \hat{F}_i (如图3中的 \hat{F}_2)包含来自特征编码器的较大尺度特征图 $\{F_k\}_{k=i+1}^5$ 、相同尺度的特征图 F_i 和来自该模块生成的较小尺度的融合特征 $\{\hat{F}_k\}_{k=1}^{i-1}$,从而完整地捕获细粒度到粗粒度的语义信息。全尺度融合的跳连接特征 T_i 可以通过式(1)计算得到:

$$\mathbf{T}_i = \mathbf{F}_i \oplus \psi_i(\hat{\mathbf{F}}_i), i=2,3,4,5 \quad (1)$$

其中:

$$\hat{\mathbf{F}}_i = \begin{cases} \mathbf{F}_1, & i=1 \\ \chi(C(\text{down}(\mathbf{F}_k))_{k=i+1}^5, C(\mathbf{F}_i), C(\text{up}(\hat{\mathbf{F}}_k))_{k=1}^{i-1}), & i=2,3,4,5 \end{cases}$$

其中, i 表示不同尺度的层数, \mathbf{F}_i 表示通过特征编码器得到的特征图, $C(\cdot)$ 表示深度可分离卷积运算, $\text{down}(\cdot)$ 和 $\text{up}(\cdot)$ 分别表示下采样和上采样的操作, $\chi(\cdot)$ 表示通道叠加的操作, $\psi_i(\cdot)$ 表示第 i 层通道调整的操作(将通道的数量调整为 c_i), \oplus 表示逐像素相加运算。最后, 依据式(1), 全尺度融合的跳连接特征 \mathbf{T}_i 由各尺寸融合特征图 $\hat{\mathbf{F}}_i$ 经过卷积维度调整和残差结构调整后得到。

2.3 循环注意力解码模块

鉴于 RNN 具有挖掘特征编码中的时间序列和上下文信息的能力, 受 TCRNet 方法^[14]中编码器结构的启发, 我们构建了一个新的结合卷积 RNN、光流场^[29]和注意力机制的循环注意力解码模块(Recurrent And Attention Decoding Unit, RADU)。特征解码器通过 RADU 逐级精细化图像特征信息, 加强特征提取的同时避免信息冗余, 从而得到高精度分割结果。从特征编码器输出的 \mathbf{T}_1 和全尺度特征残差融合模块输出的 $\{\mathbf{T}_i\}_{i=2}^5$ 开始, 利用 RADU 得到逐级精细的特征图 $\{P_i\}_{i=1}^5$ 。循环注意力解码模块的结构如图 3 所示。

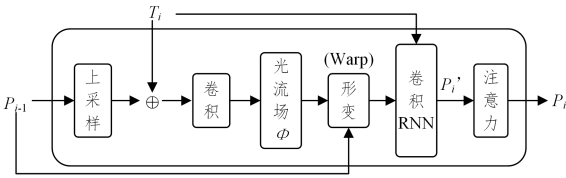


图 4 循环注意力解码模块的结构

Fig. 4 Structure of recurrent and attention decoding unit(RADU)

RADU 首先将序列特征图 $P_{i-1} \in \mathbb{R}^{c_{i-1} \times h_{i-1} \times w_{i-1}}$ 上采样至 \mathbf{T}_i 的尺寸, 再将两者级联起来。为了提高 P_{i-1} 的高分辨率特征, 通过对级联的特征映射执行卷积运算来预测特征流场 $\Phi \in \mathbb{R}^{2 \times h_i \times w_i}$ 。对于特征图图像域 Ω 中的每个像素点 p , 利用特征流场来指导 P_{i-1} 的形变。形变公式如式(2)所示:

$$\text{Warp}(P_{i-1}) = P_{i-1}(p_x + \Phi(p)_x, p_y + \Phi(p)_y) \quad (2)$$

其中, $\forall p \in \Omega$, 下标的 x 和 y 是像素点的坐标。

此外, RADU 中的卷积 RNN 操作如式(3)所示:

$$P_i' = \sigma(C(W_P P_{i-1}) + C(W_T T_i)) \quad (3)$$

其中, $\sigma(\cdot)$ 表示 ReLU 激活函数, $C(\cdot)$ 表示卷积运算, W_P 和 W_T 分别表示隐藏状态 P_{i-1} 和跳连接输入 T_i 的权重矩阵。

所提出的方法采用注意力机制来重新分配通道信息和空间信息的权重。考虑到全尺度特征融合模块引入了额外的多尺度特征信息, 而简单地将高层特征和浅层特征融合的方式通常会带来信息的冗余和混乱, 为避免冗余信息的干扰, 在解码器单元中加入了卷积注意力模块(Convolutional Block Attention Module, CBAM)^[30]来增加对有效信息的利用效率。RADU 中的注意力机制的计算过程如式(4)所示:

$$P_i = M_c(M_c(P_i') \odot P_i') \odot (M_s(P_i') \odot P_i') \quad (4)$$

注意力机制依次推断出一维的通道注意力图 $M_c \in \mathbb{R}^{c_i \times 1 \times 1}$ 和二维空间注意力图 $M_s \in \mathbb{R}^{1 \times h_i \times w_i}$ 。公式中的 \odot 表示逐元素乘法。在计算过程中, 通道注意力值沿着空间维度传播, 而空间注意力沿着通道维度传播, P_i 是单个该模块的最终输出。

总的来看, P_i 可以逐层通过输入的 P_{i-1} 和 T_i 来生成, 如式(5)所示:

$$P_i = \text{RADU}(P_{i-1}, T_i; \phi), P_i = \text{RADU}(P_{i-1}, T_i; \phi) \quad (5)$$

其中, $i = \{2, 3, 4, 5\}$, $P_i \in \mathbb{R}^{c_i \times h_i \times w_i}$, P_0 是初始化为 0 的张量, ϕ 表示在 RADU 中的所有参数。

最后, 利用 3×3 卷积将特征图 P_5 的通道数映射到分割的类别数, 得到最终的分割结果。

2.4 模型损失函数

影响分割模型总体性能的因素不仅有网络结构设计, 损失函数也发挥着关键作用。本文讨论的是医学图像的多类分割问题, 单一的交叉熵损失函数不能很好地解决在分割任务中普遍存在的类别分布不平衡问题。针对这个问题, 本文采用了 GDL 损失函数(Generalized Dice Loss)^[31]联合多分类的交叉熵损失函数来训练网络, 从而提高网络的分割精度。

GDL 损失函数可以自适应地将类别进行加权, 同时, 交叉熵损失函数能加快网络的收敛速度。GDL 损失函数(L_{gdl})是多分类的损失函数, 式中每个类别都会有对应的自适应性权重。GDL 损失函数的计算式为:

$$L_{\text{gdl}} = 1 - 2 \frac{\sum_{m=1}^M (\omega_m \sum_{n=1}^N (g_{mn} p_{mn}))}{\sum_{m=1}^M (\omega_m \sum_{n=1}^N (g_{mn} + p_{mn}))} \quad (6)$$

其中, g_{mn} 表示类别 m 在第 n 个位置像素的真值, 而 p_{mn} 表示相应的预测值。 ω_m 表示在 M 个类别中第 m 个类别的权重, 定义为:

$$\omega_m = 1 / \sum_{n=1}^N g_{mn} \quad (7)$$

多分类的交叉熵损失函数(L_{ce})为:

$$L_{\text{ce}} = - \frac{1}{N} \sum_{n=1}^N \sum_{m=1}^M (g_{mn} \log p_{mn}) \quad (8)$$

最后, 网络训练的总损失函数公式为:

$$L = L_{\text{gdl}} + \lambda L_{\text{ce}} \quad (9)$$

其中, λ 为超参数, 用来控制 L_{gdl} 和 L_{ce} 之间的比重, 本文中设置为 1.2。

3 实验及分析

本章定量评估所提出的方法在医学图像分割任务的性能。实验数据集包括两个脑图像数据集、一个心脏数据集以及一个胆总管数据集。首先分别介绍这 4 个数据集和实现细节, 接着进行消融研究以测试不同模块组合以及不同数据集之间的性能差异, 而后将本文方法与其他主流方法进行对比实验, 最后通过噪声实验来验证模型的鲁棒性。

3.1 数据集

本文的实验数据集包括 4 个公开数据集, 分别为: BrainWeb 数据集^[24]、MICCAI 2013 MRBrainS Challenge 数据集^[25]、HVSMMR 2016 Challenge 心脏数据集^[26]以及 Choledoch 数据集^[27]。将这 4 个数据集全部按照 6:4 的比例划分为训练集

和测试集。BrainWeb 数据集是一个基于大脑解剖结构的仿真数据集,其中包含 T1、T2 和 PD 3 种模态的 MR 图像。本文中使用了 399 张图像,其中 239 张图像用于训练,160 张图像用于测试。MRBrainS 数据集包含了实际环境中大脑的 MR 图像,其中包含 T1、T1IR 和 FLAIR (Fluid Attenuated Inversion Recovery) 3 种模态的 MR 图像,选择 104 张图像用于训练,70 张用于测试。HVS MR 数据集旨在分割心血管 MR 图像中的血池和心肌。本文选择了 10 个 MRI 集及其真实标注进行网络训练和评估,其中 1868 个切片用于训练,1473 个切片用于测试。Choledoch 数据集包含胆总管的切片图像,分为癌变区域和正常区域,整个数据集包含 514 张图像,仍按照上述比例来划分训练集和测试集。数据集的选择是依据一些已发表的论文^[14,23,32]中的选取方式,实验结果为 5 次实验的平均值。

3.2 实验环境与模型配置

本文中的实验运行框架为 Pytorch 1.6.0 版本,实验运行设备为 Intel © Xeon © Gold 6230 CPU @ 2.10 GHz (总的 CPU 核数为 40)。在训练过程中,模型采用 Adam 优化算法促使损失函数趋向最小来更新网络参数,初始的学习率设置为 0.0006,权重衰减值为 0.0005^[23]。同时,训练中的每批数据量大小设置为 1,总的迭代次数是 30000。实验中所采用的定量评价指标是像素精度 (Pixel Accuracy, PA)^[33] 和骰子相似系数 (Dice Similarity Coefficient, DSC)^[34]。

3.3 消融实验

为了验证所提出方法中各个模块的有效性,在上述 4 个实验数据集上分别测试了删除该模块后模型的分割效果。具体来说,分别删除卷积 MLP 模块、全尺度特征残差融合模块、RADU 模块中的注意力结构 (CBAM) 以及总损失函数中的 GDL 损失函数,并将所得结果与原模型进行比较。实验比较结果如表 1 所列。综合来看,提出的分割模型在 4 个数据集上均取得了较优的结果,证实了所提方法中各模块的有效性。例如,在 Choledoch 数据集上,所提出的全尺度特征融合模块使得模型精度指标明显提升,PA 提升了 3.06%,DSC 提升了 2.68%。在 MRBrainS 数据集上,去除卷积 MLP 模块的模型相比原模型指标反而有提升,造成该现象的原因可能是

提升指标需要的是关注图像的细节信息,但卷积 MLP 模块使得网络过分关注结构上的信息。

表 1 数据集上的模块消融实验

Table 1 Ablation experimental results on datasets

模型	(%)							
	BrainWeb		MRBrainS		HVS MR		Choledoch	
	PA	DSC	PA	DSC	PA	DSC	PA	DSC
无卷积 MLP	99.69	99.38	97.41	91.43	93.53	87.86	78.27	75.00
无全尺度特征融合	99.70	99.40	97.35	91.19	93.05	87.15	76.28	72.50
无 CBAM	99.72	99.43	97.28	90.96	94.00	87.82	77.80	74.91
无 GDL 损失函数	99.72	99.42	97.27	90.96	93.63	87.82	75.94	73.20
原模型	99.74	99.46	97.35	91.22	94.21	88.39	79.34	75.18

3.4 对比实验

我们采用了 12 种主流图像分割网络作为比较方法来评估本文方法的分割性能,包括基于传统卷积神经网络方法的分割网络 FCN^[5]、SegNet^[7]、U-Net^[6]、Attention U-Net^[8]、U-Net3+^[28]、CRDN^[23]、DFM^[9]、SFNet^[29]、集成相异性准则与熵率超像素的图像分割模型^[32],以及基于 CNN 和 Transformer 结合的主流分割网络 TransUNet^[11]、MT-UNet^[20] 和 TCRNet^[14]。为了保证比较的公平性,将 UNet 作为所有需要骨干网络的模型的骨干;CDRN 中使用的是普通卷积 RNN;TransUNet,MT-UNet 和 TCRNet 中输入图像的分辨率是 256×256。

4 个数据集的实验结果如表 2 所列。表中上半部分是主流的 CNN 模型的实验结果,下半部分是结合 Transformer 结构的模型的实验结果。从表中可以看出,大部分基于卷积神经网络的模型仍然具有较好的分割效果,比如 CDRN 在数据集上的分割精度超过了 TransUNet;HVS MR 数据集中的训练图像数据由于尺寸大小不统一且不方正,不适用几个需要指定输入图像尺寸的模型 (如 DFM,TransUNet 和 MT-UNet),若强行将图像形变,分割精度会明显下降。从 4 个数据集的实验结果来看,本文方法均取得了优异的结果,其相较于目前最优的 TCRNet 在 4 个数据集上的 DSC 值分别提高了 0.06%、0.36%、1.73% 和 2.38%;尤其是在 Choledoch 数据集上,本文方法的 PA 值和 DSC 值比 TCRNet 模型高出 2.96% 和 2.38%。

表 2 在 BrainWeb,MRBrainS,HVS MR 和 Choledoch 上的模型对比实验结果

Table 2 Experiment result comparison on BrainWeb,MRBrainS,HVS MR and Choledoch datasets

方法	来源	BrainWeb		MRBrainS		HVS MR		Choledoch	
		PA	DSC	PA	DSC	PA	DSC	PA	DSC
FCN	CVPR 2015	95.79	91.76	95.64	86.18	91.79	82.95	66.35	61.84
U-Net	MICCAI 2015	99.45	98.92	97.05	90.21	92.79	85.93	65.57	62.10
SegNet	TPAMI 2017	97.15	94.55	95.06	84.48	90.27	80.99	65.66	61.46
Attn-UNet	MIDL 2018	99.71	99.40	97.09	90.33	93.72	87.89	68.20	64.98
U-Net3+	ICASSP 2020	99.05	98.12	96.73	89.19	92.87	85.66	71.09	67.36
CRDN	AAAI 2020	99.67	99.34	97.18	90.68	93.88	88.00	72.50	66.83
SFNet	ECCV 2020	99.48	99.02	97.02	90.07	92.98	85.93	69.02	64.20
DFM	MICCAI 2020	99.33	98.62	96.70	89.17	—	—	64.53	59.58
Gu 等	中国图像图形学报 2022	98.93	97.71	92.46	84.74	—	—	—	—
TransUNet	CVPR 2021	94.40	88.97	92.54	76.72	—	—	69.57	61.11
MT-UNet	ICASSP 2022	99.05	98.13	96.83	89.60	—	—	70.63	67.57
TCRNet	ICASSP 2022	99.70	99.40	97.25	90.86	93.49	86.66	76.38	72.80
Ours	—	99.74	99.46	97.35	91.22	94.21	88.39	79.34	75.18

然后,为了验证所提模型结构的有效性,使用 thop 工具来计算模型参数量和 FLOPs。表 3 以 Choledoch 数据集上的实验结果为例,展示了与本文方法相关的比较方法的模型参数量和 FLOPs。本文方法与表中其他方法相比,FLOPs 和模型参数量相对较高,但有效提高了分割精度,并且对计算资源的要求不高。

表 3 Choledoch 数据集上不同模型的参数量和 FLOPs

Table 3 Parameters and FLOPs of different models on Choledoch dataset

方法	DSC/% \uparrow	FLOPs \downarrow	参数量 \downarrow
U-Net 3+	67.36	198.057×10^9	26.975×10^6
CRDN	66.83	1.206×10^9	1.206×10^6
SFNet	64.20	5.117×10^9	1.254×10^6
DFM	59.58	66.219×10^9	34.536×10^6
Ours	75.18	234.570×10^9	55.823×10^6

3.5 噪声干扰实验

由于医学图像实际获取过程中会受外部环境以及器械自身因素的影响,例如噪声和偏移场的干扰,图像容易出现伪影、灰度不均匀等情况。为了验证模型对噪声和偏移场干扰的鲁棒性,在 BrainWeb 数据集上将本文模型与 FCN,U-Net, SegNet 和 CRDN 这 4 个分割精度差异较明显的模型进行对比。

BrainWeb 数据集提供了 5 类不同等级的高斯噪声图像,噪声等级分别为 1%,3%,5%,7%和 9%。如图 5 所示,在低噪声等级时,各模型之间受影响程度较小,而当噪声逐渐增大时,本文方法仍具有明显优势。此外,BrainWeb 数据集提供了受偏移场干扰的数据,图 6 说明了 4 种方法的分割结果随着 INU 的增加而下降,但本文提出的方法仍然保持着较高的指标。

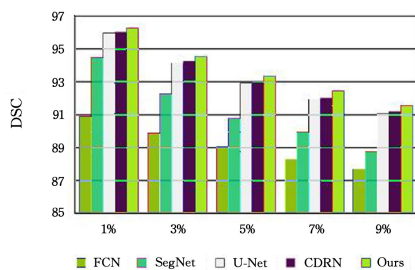


图 5 BrainWeb 图像在受到高斯噪声干扰下的实验结果

Fig. 5 Image segmentation results with Gaussian noise interference

on BrainWeb dataset

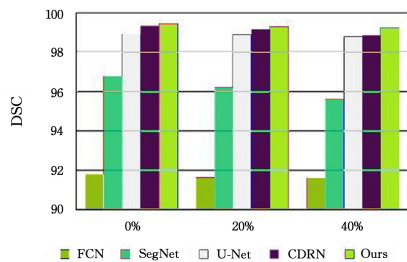


图 6 BrainWeb 图像在受到偏移场干扰下的实验结果

Fig. 6 Image segmentation results with Offset field interference

on BrainWeb dataset

由此可得出结论:该模型具有对图像噪声干扰和图像灰度不均匀现象的鲁棒性。

3.6 分割结果可视化

为了进一步显示本文方法的分割细节,对 6 种典型分割方法和提出方法的图像分割结果进行了可视化。图 7 显示了各方法在 BrainWeb 和 MRBrainS 上的可视化结果。

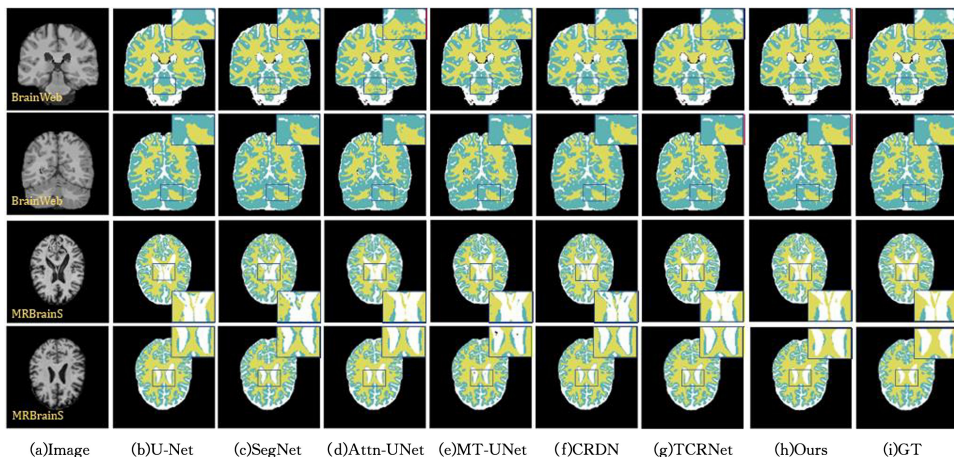


图 7 各种方法在 BrainWeb 和 MRBrainS 图像上分割的可视化实验结果(电子版为彩图)

Fig. 7 Visualization of image segmentation results of various methods on BrainWeb and MRBrainS datasets

图中第一列展示的是两个数据集的测试图像,其余各列分别为 U-Net, SegNet, Attn-U-Net, MT-U-Net, CRDN, TCR-Net 和本文方法的分割结果及分割真值,黑色区域为背景,黄色区域为白质,绿色区域为灰质,白色区域为脑脊液。可以看出,U-Net 和 SegNet 的分割结果较差,其余方法间差异不大。但从与其他方法细节的比较看来,本文方法在可视化结果中与真值的差异是较小的,表明其分割精度较高,证实了该方法的有效性。

结束语 针对目前主流医学图像分割方法存在的图像特征信息挖掘不充分和缺少多尺度特征间融合信息等局限性,本文提出了一种集成全尺度融合和循环注意力的图像分割网络。首先,在编码过程中加入卷积 MLP 模块提取的全局特征信息,进一步扩大编码器的特征感受野。其次,通过全尺度特征融合模块使得各尺度跳连接特征进行粗粒度信息和细粒度信息的有效融合,减小了各层特征间的语义差异。最后,解码器通过提出的结合 RNN 和注意力机制的 RADU 来逐级

精细化图像特征信息,加强特征提取的同时避免信息冗余,并得到最终的分割结果。

实验结果表明,与目前主流的分割网络相比较,本文提出的分割网络在 BrainWeb, MRBrains, HVSMR 和 Choledoch 这 4 个数据集上都取得了明显的分割精度提升。其中,在 Choledoch 数据集上,所提方法相较于目前最优网络的 PA 指标和 DSC 指标精度提升分别高达 3% 和 2.3%,验证了本文模型的有效性和出色性能。

但本文算法也存在局限性,仍然有很大提升空间,具体来说:1)全尺度特征融合模块和循环解码单元模块结构仍然较为复杂,存在计算量较大的问题,后续工作可以致力于降低计算复杂度和网络剪枝;2)本文的任务聚焦于 2D 医学图像分割,目前 3D 医学图像处理已成为热门问题,可以进一步研究如何将网络迁移到 3D 图像分割任务中。

参 考 文 献

- [1] MACQUEEN J. Some methods for classification and analysis of multivariate observations[C]// Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. 1967:281-297.
- [2] DUNN J C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters [J]. Journal of Cybernetics, 1973, 3(3): 32-57.
- [3] GONG M, LIANG Y, SHI J, et al. Fuzzy c-means clustering with local information and kernel metric for image segmentation [J]. IEEE Transactions on Image Processing, 2012, 22(2): 573-584.
- [4] ASHISH V, NOAM S, NIKI P, et al. Attention is all you need [C]// Proceedings of 31st Conference on Neural Information Processing Systems. 2017:6000-6010.
- [5] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [6] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]// Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. 2015:234-241.
- [7] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(12): 2481-2495.
- [8] OKTAY O, SCHLEMPER J, FOLGOC L, et al. Attention U-Net: Learning where to look for the pancreas[C]// Conference on Medical Imaging with Deep Learning. 2018:1-10.
- [9] CHENG F, CHEN C, WANG Y, et al. Learning directional feature maps for cardiac MRI segmentation[C]// Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2020:108-117.
- [10] YE Z, WU M. Choroidal Neovascularization Segmentation Combining Temporal Supervision and Attention Mechanism [J]. Computer Science, 2021, 48(8): 118-124.
- [11] BAI X, MA Y, WANG W. Segmentation Method of Edge-guided Breast Ultrasound Images Based on Feature Fusion [J]. Computer Science, 2023, 50(3): 199-207.
- [12] CHEN J, LU Y, YU Q, et al. TransUNet: Transformers make strong encoders for medical image segmentation [C]// Proceedings of the IEEE International Conference on Computer Vision. 2021:1-13.
- [13] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [C]// International Conference on Learning Representations. 2021:1-22.
- [14] SHAN X, MA T, GU A, et al. TCRNet: Make Transformer, CNN and RNN complement each other [C]// Proceedings of International Conference on Acoustics, Speech and Signal Processing. 2022:1441-1445.
- [15] JIN Y, HAN D, KO H. TrSeg: Transformer for semantic segmentation [J]. Pattern Recognition Letters, 2021, 148: 29-35.
- [16] TOLSTIKHIN I O, HOULSBY N, KOLESNIKOV A, et al. Mlp-mixer: An all-mlp architecture for vision [C]// Proceedings of Neural Information Processing Systems. 2021:1-16.
- [17] LI J, HASSANI A, WALTON S, et al. ConvMLP: Hierarchical convolutional MLPs for vision [J]. arXiv:2109.04454, 2021.
- [18] VALANARASU J M J, PATEL M V. UNeXt: MLP-based rapid medical image segmentation network [C]// Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. 2022:23-33.
- [19] VALANARASU J M J, OZA P, HACIHALILOGLU I, et al. Medical Transformer: Gated axial-attention for medical image segmentation [C]// Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. 2021:36-46.
- [20] WANG H, XIE S, LIN L, et al. Mixed Transformer U-Net for medical image segmentation [C]// Proceedings of International Conference on Acoustics, Speech and Signal Processing. 2022: 2390-2394.
- [21] ZHOU P, GONG S, ZHONG S, et al. Image Semantic Segmentation Based on Deep Feature Fusion [J]. Computer Science, 2020, 47(2): 126-134.
- [22] WANG H, CAO P, WANG J, et al. UCTransNet: Rethinking the skip connections in u-net from a channel-wise perspective with Transformer [C]// Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence. Vancouver. 2022: 7966-7978.
- [23] WEN Y, XIE K, HE L. Segmenting medical MRI via recurrent decoding cell [C]// Proceedings of The Thirty-Forth AAAI Conference on Artificial Intelligence. 2020:12452-12459.
- [24] COCOSCO C A, KOLLOKIAN V, KWAN K S, et al. Brainweb: Online interface to a 3D MRI simulated brain database [J]. NeuroImage, 1997, 5(4): part 2/4, S425.
- [25] MENDRIK A M, VINCKEN K L, KUIJFF H J, et al. MRBrainS challenge: online evaluation framework for brain image segmentation in 3T MRI scans [J]. Computational Intelligence and

- Neuroscience, 2015(1):1-16.
- [26] PACE D F, DALCA A V, GEVA T, et al. Interactive whole-heart segmentation in congenital heart disease[C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015:80-88.
- [27] ZHANG Q, LI Q, YU G, et al. A Multidimensional Choledoch Database and Benchmarks for Cholangiocarcinoma Diagnosis [J]. IEEE Access, 2019, 7:149414-149421.
- [28] HUANG H, LIN L, TONG R, et al. UNet 3+: A full-scale connected UNet for medical image segmentation[C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2020:1055-1059.
- [29] LI X, YOU A, ZHU Z, et al. Semantic flow for fast and accurate scene parsing [C] // Proceedings of European Conference on Computer Vision, 2020:775-793.
- [30] WOO S, PARK J, LEE J, et al. CBAM: Convolutional Block Attention Module[C]//Proceedings of the European Conference on Computer Vision, 2018:3-19.
- [31] SUDRE C H, LI W Q, VERCAUTEREN T, et al. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations[C]//Proceedings of the 3rd MICCAI International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2017:240-248.
- [32] GU A, SHAN X, WEN Y. An image segmentation model with integrated dissimilarity criterion and entropy rate super-pixel [J]. Journal of Image and Graphics, 2022, 27(11):3267-3279.
- [33] AL-DMOUR H, AL-ANI A. A clustering fusion technique for MR brain tissue segmentation [J]. Neurocomputing, 2018, 275:546559.
- [34] HASTIE T, TIBSHIRANI R. Discriminant adaptive nearest neighbor classification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(6):607-616.



SHAN Xinxin, born in 1996, Ph.D. Her main research interests include computer vision and image processing.



WEN Ying, born in 1975, professor, Ph.D supervisor, is a member of CCF (No. F2169M). Her main research interests include computer vision, image processing and artificial intelligence.

(责任编辑:柯颖)