

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2025)10-3392-19

论文引用格式: Liu W, Zhong M, Liu G W, Wang H N and Ning Q. 2025. Detail-enhanced medical image segmentation focusing on global and intermediate features. Journal of Image and Graphics, 30(10):3392-3410(刘威, 钟森, 刘光伟, 王浩男, 宁倩. 2025. 聚焦全局与中间层的细节增强医学图像分割. 中国图象图形学报, 30(10):3392-3410)[DOI:10.11834/jig.240463]

聚焦全局与中间层的细节增强医学图像分割

刘威^{1,2,3*}, 钟森^{1,2}, 刘光伟⁴, 王浩男⁵, 宁倩^{1,2}

1. 辽宁工程技术大学理学院, 阜新 123000; 2. 辽宁工程技术大学数学与系统科学研究所, 阜新 123000;
3. 辽宁工程技术大学智能工程与数学研究院, 阜新 123000; 4. 辽宁工程技术大学矿业学院, 阜新 123000;
5. 约翰霍普金斯大学, 马里兰州 21218, 美国

摘要: 目的 随着人工智能的发展,深度学习技术在医学图像分割中得到广泛应用。但现有方法往往采用自上而下或自下而上的方式进行特征融合,易忽略或丢失中间层特征信息。此外,现有方法对病灶区域分割边界不够精细。针对上述问题,本文提出一种聚焦全局与中间层特征的细节增强医学图像分割网络(detail-enhanced medical image segmentation network focusing on global and intermediate features, DEMS-GIF)。**方法** 首先通过进一步关注中间层信息,并利用Transformer提取不同区域之间的长距离依赖关系的能力,设计了一种基于Transformer的桥接特征融合模块(Transformer-based bridge feature fusion module, TBBFF),以提升模型的特征提取能力。其次,通过引入反向注意力机制,并结合腐蚀和膨胀操作,提出一种反向注意下的扩缩区域增强上采样策略(expanded and scaled region enhanced upsampling strategy under reverse attention, ESRU),使得模型能够更好地捕捉边界和细节信息。DEMS-GIF模型通过结合TBBFF模块和ESRU策略,进一步提高了分割的准确性。**结果** 在CVC-ClinicDB、DDTI(digital database thyroid image)和Kvasir-SEG 3个数据集上进行对比实验和模块消融实验,评估提出的DEMS-GIF模型,并在CVC-ClinicDB数据集上进行参数消融实验,以了解DEMS-GIF中每个模块和结构内部的有效性。实验结果表明,DEMS-GIF模型的mIoU值分别达到94.74%、84.56%和88.46%,Dice值分别达到94.82%、82.95%和87.44%。与原UNet型通道变换网络相比,mIoU值分别提升3.73%、3.4%和5.24%,Dice值分别提升4.84%、5.45%和6.82%。**结论** 本文提出的DEMS-GIF网络模型较其他先进的分割方法的分割效果更优,表明了其在医学图像分割中的优越性。

关键词: 医学图像分割;特征融合;上采样;反向注意力;腐蚀;膨胀

Detail-enhanced medical image segmentation focusing on global and intermediate features

Liu Wei^{1,2,3*}, Zhong Miao^{1,2}, Liu Guangwei⁴, Wang Haonan⁵, Ning Qian^{1,2}

1. College of Science, Liaoning Technical University, Fuxin 123000, China; 2. Institute of Mathematics and Systems Science, Liaoning Technical University, Fuxin 123000, China; 3. Institute of Intelligent Engineering and Mathematics, Liaoning Technical University, Fuxin 123000, China; 4. College of Mines, Liaoning Technical University, Fuxin 123000, China; 5. Johns Hopkins University, Maryland 21218, USA

收稿日期:2024-08-19;修回日期:2025-01-21;预印本日期:2025-01-28

*通信作者:刘威 liuwei@lntu.edu.cn

基金项目:国家自然科学基金项目(52374123,51974144);辽宁省“兴辽英才计划”项目(XLYC2211085);辽宁省教育厅基本科研项目(LJ212410147013)

Supported by: National Natural Science Foundation of China(52374123,51974144); Liaoning Revitalization Talents Program(XLYC2211085); Basic Scientific Research Project of the Liaoning Provincial Department of Education(LJ212410147013)

Abstract: Objective Medical image segmentation is a crucial and challenging task in the field of not only medical imaging but also modern medicine. Segmentation techniques can be used to precisely locate and measure tumors, blood vessels, and other structures, which facilitate the early diagnosis of the disease and the evaluation of treatment outcomes. In addition, accurate and reliable medical image segmentation can be utilized to monitor disease progression, assist physicians in formulating long-term treatment plans, lay a solid foundation for clinical diagnosis and pathological research, and provide valuable data support. Feature fusion in medical image segmentation can comprehensively capture detailed and global information by combining multi-level, multi-scale, and multimodal features, which improves segmentation accuracy and robustness. This method not only enhances the capability of automated medical image processing and reduces dependence on large amounts of annotated data but also increases the accuracy of clinical decisions. This approach helps doctors make highly reliable judgments in diagnosis and treatment, which promote the development of medical automation. In medical image segmentation, upsampling strategies can effectively restore high-resolution features and detailed information, which enhances the capability of the segmentation model to recognize small structures and boundaries. By adopting appropriate upsampling methods, the spatial information of the original image can be preserved and recovered, which improve the precision of segmentation. This enhancement not only aids in the precise localization of lesion areas and the extraction of biological features but also holds significant importance for clinical diagnosis and treatment decision making. The advancement of artificial intelligence has facilitated the wide application of deep learning techniques. However, these deep learning methods often employ top-down or bottom-up approaches for feature fusion, which can result in the neglect or loss of intermediate layer feature information. Moreover, existing methods still encounter issues with imprecise segmentation boundaries of lesion areas, which leads to the omission of critical information when dealing with fine structures and complex background information. To address these concerns, this study proposes a detail-enhanced medical image segmentation network focusing on global and intermediate features (DEMS-GIF). **Method** First, to address the shortcomings of existing feature fusion methods in capturing complex structures and integrating intermediate features, this study proposes a Transformer-based bridge feature fusion (TBBFF) module. Compared with other feature fusion modules, the TBBFF module focuses more on intermediate feature information and leverages the capability of the Transformer to capture long-range dependencies between different regions. These features enable the network to better understand the overall structure and contextual information of the image, which further enhances the segmentation performance and robustness of the model. Second, to address the issues of excessive smoothness in generated images and the lack of precise boundary segmentation in the areas of lesion using existing methods, this study proposes an expanded and scale-region-enhanced upsampling strategy under reverse attention (ESRU strategy). By incorporating a reverse attention mechanism and combining erosion and dilation operations, the model can better capture boundary and detail information in the target regions of medical images, which enhances the integrity and continuity of the segmented regions. This enhancement, in turn, improves the accuracy and stability of segmentation. In conclusion, the DEMS-GIF model, which integrating the TBBFF module and the ESRU strategy, effectively extracts image details and global information. Thus, it ensures the integrity of the segmented regions and further enhances segmentation accuracy. **Result** We evaluated the superiority of the DEMS-GIF model by comparing it with recent and classic methods through applying them on three different datasets: CVC-ClinicDB, digital database thyroid image(DDTI), and Kvasir-SEG. The experimental results demonstrate that, on the CVC-ClinicDB dataset, the DEMS-GIF model achieved mIoU and Dice scores of 94.74% and 94.82%, respectively. Compared with recently proposed segmentation methods, the mIoU achieved by DEMS-GIF outperformed those of MSUNet, MBSNet, and SCSONet by 3.98%, 4.42%, and 20.13%, respectively. With an increase in training iterations, the train loss of DEMS-GIF decreased significantly and was notably lower than those of ResUNet++ and SCSONet. On the DDTI dataset, the DEMS-GIF model achieved a precision of 85.52% and an mIoU of 84.56%. Compared with traditional network models, DEMS-GIF showed the most significant differences with ResNet++, with precision and mIoU higher by 13.97% and 10.96%, respectively. In addition, the train loss of DEMS-GIF was noticeably lower than those of other models, with the most significant difference observed with MFSNet. On the Kvasir-SEG dataset, the DEMS-GIF model achieved mIoU and Dice scores of 87.44%. Its mIoU was 12.69% higher than that of ResUNet++, and its Dice was 6.82% higher than that of UNet, which demonstrates substantial superiority. Compared with those of other models such as MBSNet, DTA-UNet, and SCSONet, the mIoU of

DEMS-GIF was higher by 6.77%, 3.84%, and 25.02%, respectively, which suggests the best performance. During training, the train loss of DEMS-GIF was significantly lower than those of other network models, except that of MFSNet. We also conducted ablation experiments to understand the effectiveness of each module and structure within DEMS-GIF. The ablation experiments were divided into module ablation and parameter ablation. The module ablation experiments were conducted on the CVC-ClinicDB, DDTI, and Kvasir-SEG datasets, and the parameter ablation experiments were performed on the CVC-ClinicDB dataset. Using Res2Net as the baseline network, the module ablation experiments thoroughly discussed the impact of the TBBFF module and ESRU strategy on model performance. The results demonstrated that the combination of the TBBFF module and ESRU strategy allows the network to focus more on lesion areas, which results in more accurate segmentation. Parameter ablation involved experiments on the threshold and adaptive parameters used in reweighting with erosion and dilation operations. The findings illustrated the contribution of each parameter in the DEMS-GIF model and identified the optimal parameter values. As a result, the model settings are optimized to further enhance the overall performance and reliability of the model. **Conclusion** In this study, we propose DEMS-GIF, which integrates the TBBFF module and the ESRU strategy. The network effectively leverages intermediate layer feature information to integrate features across different scales and levels. It also enhances the focus on boundary and detail information of lesion areas, which realizes efficient extraction and refined processing of image features. This process results in more complete and accurate segmentation regions. Experimental results show that the proposed DEMS-GIF network model outperforms other advanced segmentation methods, which demonstrates its superiority in medical image segmentation.

Key words: medical image segmentation; feature fusion; upsampling; reverse attention; erosion; dilation

0 引言

医学图像分割在现代医疗中扮演着至关重要的角色,它能够对计算机断层扫描(computed tomography, CT)、超声(ultrasound, US)以及磁共振成像(magnetic resonance imaging, MRI)等多种类型的医学图像进行分割,将其中的不同组织、器官和病灶区域自动区分出来,帮助医生更准确地进行疾病的诊断和治疗计划的制定。此外,医学图像分割在手术规划和病变进展监测等方面也提供了重要支持,是现代医疗不可或缺的技术。

然而,医学图像分割面临着诸多挑战。由于不同患者的器官病变在形状、大小、纹理及颜色特征上具有显著差异,因此对分割算法的尺度适应性与鲁棒性提出很高的要求。此外,在许多医学图像中,不同组织或病变的边界非常模糊,难以清晰分辨,如何准确识别和分割这些模糊边界也是需要进一步解决的问题。尽管传统的分割方法如阈值分割、区域生长和边缘检测在一定程度上解决了这些问题,但仍存在显著不足。传统方法通常依赖于手动设置参数和初始条件,导致分割结果受人为主观因素影响较大。因此,开发计算机辅助自动分割方法对提升分割效率、准确性和质量,减少人力资源的消耗是至关

重要的。

现阶段,医学图像分割技术的实现主要依赖于图像处理和计算机视觉方法,为解决其面临的包括形状、大小、纹理和颜色的差异以及边界难以区分等问题,基于深度学习尤其是基于卷积神经网络(convolutional neural network, CNN)的医学图像分割已经成为医疗诊断领域的研究热点问题(Litjens等, 2017),该领域已经有许多器官组织开始应用,如视网膜血管、结肠息肉和脑肿瘤等(Luo等, 2019; Wu等, 2022; El-Assiouti等, 2023)。全卷积神经网络(fully convolutional network, FCN)是深度学习在医学图像领域应用最早的技术,但由于其上下采样结构不一致会导致信息丢失, Ronneberger等人(2015)提出最经典的编码器—解码器结构UNet,进而基于UNet及其变体的结构,如ResUNet(residual UNet)(Alom等, 2018)、DTA-UNet(UNet based on dynamic convolution decomposition and triplet attention)(Li等, 2024b)等通过引入残差块、嵌套跳跃连接等技术,增强了网络的特征提取能力和分割精度。其次,如CA-Net(comprehensive attention-based network)(Gu等, 2021)、MBSNet(multi-branch segmentation network)(Jin等, 2023)等基于注意力机制的结构设计了多种注意力模块,使网络更加关注前景区域和显著特征,提高了模型的自适应能力和特征学习效果。

此外,一些模型如 SF-Net(segmentation-fusion multi-task network)(Liu 等,2022b)通过图像融合作为额外的正则化来学习特征,UC-DDGAN(U-Net3+ and cross-modal attention block dual-discriminator generative adversal network)(王丽芳 等,2022)通过交替提取跨模态特征与深层特征以得到复合特征图,从而实现更充分的多模态特征融合,更好地解决多模态图像分割问题。可见,越来越多的研究者对分割模型进行不同角度的创新和优化,这不仅提升了医学图像分割的准确性和鲁棒性,还为临床诊断和治疗提供了强有力的支持。

但是,现有模型往往容易忽略中间层的特征信息,在进行特征融合时普遍依赖自上而下或自下而上的特征传递方式,导致模型难以充分整合不同尺度和层次的特征。同时,在对医学图像中的高精度边界和细微结构进行分割时,采用的上采样策略通常容易遗漏细节信息,并且对全局信息和局部信息处理不足,从而导致分割结果不够全面和精确,影响诊断的准确性。针对上述问题,本文提出一个聚焦全局与中间层特征的细节增强医学图像分割网络(detail-enhanced medical image segmentation network focusing on global and intermediate features, DEMS-GIF)。

本文主要贡献包括:1)研究设计了一种基于Transformer的桥接特征融合模块(Transformer-based bridge feature fusion module, TBBFF),该模块更加关注中间层特征信息,并利用Transformer进一步整合图像的整体结构和上下文信息,提升模型的特征提取能力;2)提出一种反向注意下的扩缩区域增强上采样策略(expanded and scaled region enhanced upsampling strategy under reverse attention, ESRU),通过引入反向注意力机制,并结合腐蚀和膨胀操作,以期改善对病灶区域分割边界不精细的问题,提高网络对细节分割的准确性;3)针对医学图像分割问题,提出一个聚焦全局与中间层特征的细节增强医学图像分割网络 DEMS-GIF,对图像特征实现了高效提取和精细化处理,显著提升了分割精度,具有较强的泛化能力和鲁棒性;4)所提方法在 CVC-ClinicDB、DDTI(digital database thyroid image)和 Kvasir-SEG 3 个数据集上进行了大量实验,以验证其有效性。同时,利用大量的消融实验证明了本文方法中参数选取的合理性。

1 相关工作

1.1 医学图像分割

医学图像分割是医学成像领域中一项重要且具有挑战性的任务,在现代医学中具有重要意义。通过分割技术,可以准确地定位和测量肿瘤、血管等结构,有助于疾病的早期诊断和治疗效果评估。此外,精确、可靠的医学图像分割可用于监测疾病的进展,辅助医生制定长期治疗计划,为临床诊断和病理学研究奠定可靠的基础,提供宝贵的数据支持。总之,通过不断提升医学图像分割的准确性和可靠性,可以显著提高医疗服务的质量和效率,推动医疗自动化的发展。

随着深度学习技术的迅猛发展,尤其是卷积神经网络(CNN)的应用,医学图像分割的准确性和效率得到了显著提升。尽管基于阈值、区域生长和边缘检测的传统方法在某些特定场景下仍然具有应用价值,但深度学习方法如 UNet、UNet++(Zhou 等,2018)和 ResUNet++(Jha 等,2019)等已经成为当前研究的主流。由于现有的 UNet 在编码器部分对每个下采样层的设计相等,并且简单堆叠卷积,无法充分提取来自不同深度的特征信息,Xu 等人(2023a)设计了一种更深层且更紧凑的分割注意 U 形网络 DCSAU-Net(deeper and more compact split-attention UNet),通过主要特征保留和紧凑分割注意块两种框架高效利用低级和高级语义信息来解决这一问题。Hu 等人(2023)为解决超声图像中的斑点噪声、低对比度和模糊边界等问题,设计了一种交互式双流网络 DSC-Net(detail screeching and compensation network)。该模型使用 Transformer 获取多尺度详细特征,并使用 CNN 提取较少噪声的主体特征,通过细节筛选模块引导多尺度详细特征过滤噪声,并将过滤后的详细特征应用于细节补偿模块,从而为 CNN 流补充丰富的细节。卢玲和漆为民(2023)通过将 Transformer、注意力门控机制及 UNet 相结合构成编码—解码结构,提出一种 CNN 与 Transformer 混合分割模型 TransAGUNet。其中的编码器使用 Transformer 和 CNN 混合架构,提取局部及全局特征。解码器使用 CNN 架构,在跳跃连接部分融入注意力门,将下采样特征图对应的注意力图与下一层上采样后获得的特征图进行拼接,从而融合低层与

高层特征,实现更精细的分割。张家宝和肖志勇(2023)提出一种卷积神经网络和自注意力相结合的双分支网络 LG UNet (local UNet and global Transformer network)。基于 UNet 设计了 Local UNet 分支,利用卷积神经网络的优势,学习分割目标的局部信息。然后在 Global Transformer 分支中,利用 Transformer 全局依赖关系的学习能力来优化分割细节。最后在编码过程中通过交叉融合模块将 Local 分支和 Global 分支的特征图进行融合,将两者优势互补。Chen 等人(2024)为了在确保分割性能的同时减少参数数量和浮点运算,提出一种轻量级网络 SCSONet (spatial-channel synergistic optimization net)。通过引入具有全维度注意力的 ConvStem 模块,使得模型能够有效学习卷积核 4 个维度的互补注意力权重,从而增强对不规则形状病变区域的识别能力,降低参数和计算负担,同时空间通道融合块通过空间和通道的特征融合减少特征冗余,进一步降低参数数量并提升分割结果。Sun 等人(2024)通过整合 Transformer 和双注意力模块到传统的 U 形架构中,提出名为 DA-TransUNet 的新型深度医学图像分割框架。针对医学图像的高细节需求,其优化了双注意力模块的间歇通道,并在每个跳跃连接中应用 DA (dual attention) 模块,有效过滤无关信息。该集成显著增强了模型提取特征的能力,从而改善了医学图像分割的性能。Guo 等人(2024)提出一种简单而有效的混合网络 UCTNet (uncertainty-guided CNN-Transformer hybrid network) 用于医学图像分割。该网络通过不确定性估计让 Transformer 只专注于 CNN 预测中不可靠的区域,以建立全局依赖性,减少功能重叠,从而避免冗余,提高全局特征提取的效率。此外,由于需要处理的区域较少,UCTNet 能够更好地收敛,学习更稳健的特征表示。

显然,这些深度学习的方法在医学图像分割中发挥了重要的作用,但是,深度学习技术在处理图像质量、特征差异以及细节和边缘的准确性上仍具有较大的提升空间,需要进一步改进其在临床应用中的稳定性和可靠性。

1.2 特征融合

医学图像分割中的特征融合通过结合多层次、多尺度和多模态的特征,能够全面捕捉图像的细节和全局信息,从而提高分割的准确性和鲁棒性。这不仅能够增强自动化医疗图像处理的能力,减少对

大量标注数据的依赖,还能够提高临床决策的准确性,帮助医生在诊断和治疗中做出更可靠的判断,推动医疗自动化的发展。

当今,特征融合技术迅速发展。最初的 UNet 是引入了跳跃连接,将编码器和解码器对应层的特征直接相连,显著提高了医学图像分割的效果。Azad 等人(2022)提出 TransNorm 深度分割框架,通过在 UNet 的编码器和跳跃连接中整合 Transformer 模块,利用自注意力机制捕捉全局信息和长距离依赖关系。同时引入空间归一化机制,自适应地重新校准跳跃连接路径上的特征,从而在扩展路径和收缩路径之间更好地融合特征。Zhang 等人(2023b)提出的新型网络架构 TranSEFusionNet 中,空间特征融合模块通过带有跳跃连接的 SEF (squeeze-and-excitation fusion) 模块改善了将深层特征传递给浅层特征的准确性。边缘特征融合模块位于每个解码器层的输出部分,通过细化网络的语义信息,提高了对边缘模糊特征的识别能力。Song 等人(2023)提出 DPCTN (dual path context-aware transformer network) 网络架构,先通过 Transformer 块的通道交叉融合实现高层局部特征的多尺度信息融合,减少冗余信息的影响,再利用双主干特征融合模块有效结合局部和全局高层特征信息,利用解码器逐层细化和丰富边界与区域特征,实现对边界和区域的有效监督。Xu 等人(2023b)提出一种四轴融合框架 FAFuse (four-axis fusion framework),其核心组件是四轴融合模块,该模块通过结合四轴注意力(高度、宽度、主对角线和副对角线轴向注意力)、多尺度卷积、深度可分离卷积及 Hadamard 积的残差结构,有效地融合了全局和局部信息。Liu 等人(2023)为解决现有方法在融合局部特征和全局上下文信息方面的局限性,提出 LGI Net (local-global information interaction network) 医学图像分割架构。其通过改进内部计算,实现网络内局部感知能力和全局上下文信息之间的充分交互。此外,网络还结合了 ECA (efficient channel attention) 模块,有效捕捉通道间的相互作用,并提高层间信息交换能力。Li 等人(2024a)提出一种双语义聚合 Transformer 网络 D-SAT (dual semantic aggregation transformer with dual attention),通过设计双语义特征聚合模块将 CNN 的局部特征细节与 Transformer 的长距离建模能力有效结合,利用条带空间注意机制在不同空间维度上构建

像素级特征关系,并在编码器和解码器之间的跳跃连接中加入特征分布门控注意模块,过滤低级语义信息中的噪声,实现低级和高级语义特征的融合。周新文等人(2024)提出的TRNet(tooth segmentation network with fine-grained receptive fields)模型在融合过程中引入了软性注意力机制,通过对不同尺度、不同层级的特征动态分配权重,使得模型能够根据任务需求自适应地调整特征,避免了单一尺度或层级特征在融合过程中的过度主导。Pan等人(2025)提出一个双重注意力模型DAMAF(dual attention network with multi-level adaptive complementary fusion),通过引入高效注意力和转置注意力机制,实现对医学图像中丰富的空间和通道信息的捕捉。其还利用多层次融合注意块扩展了特征的互补性,逐步增强高层次和低层次特征之间的相关性,同时多层次跳跃注意块强化了相邻层次信息的交互,从而提高特征的表达能力。这些技术的进步为临床应用提供了更加精准和可靠的工具,推动了医疗自动化的发展。

在医学图像分割中,尽管这些特征融合技术取得了重大进步,但仍然存在一些不足之处。现有的方法通常依赖于逐层向上或向下的特征传递,从而导致忽略或丢失中间层特征,影响分割结果的完整性和准确性。同时,这些方法对于不同尺度和层次的特征还是难以做到充分整合,对图像的细节和全局信息捕捉不足,影响分割的精度和鲁棒性。因此,如何有效利用中间层信息,捕捉不同层次的复杂结构,对推动医学图像分割技术在临床应用中进一步发展具有重要意义。

1.3 上采样策略

在医学图像分割中,上采样策略能够有效地恢复高分辨率特征和细节信息,从而提升分割模型对微小结构和边界的识别能力。采用适当的上采样技术,能够保留和恢复原始图像的空间信息,进而提高分割的精度和准确性。这不仅有助于精确定位病变区域和提取生物特征,还对临床诊断和治疗决策具有重要意义。

Liu等人(2022a)提出的LLRNet(local-long range hybrid features network)模型采用级联上采样操作对提取的长距离特征和局部特征进行结合,这些混合特征在上采样过程中被用来定位目标组织的位置,从而确保分割结果的精确度。Zhang等人(2023a)设计了一种新型的神经网络模型SUNet++,

其上采样过程是利用转置卷积操作实现的,通过反转标准卷积操作的输入和输出,有效地增加分辨率,从而帮助模型在解码过程中恢复细节信息,在处理小尺度目标时更为有效。Ouyang等人(2023)提出的LEA U-Net(local feature enhancement and attention U-Net)模型通过转置卷积来恢复图像分辨率,并结合跳跃连接将编码器提取的特征传递到解码器,从而实现上采样操作。同时,注意力机制在上采样过程中增强了与血管分割相关的特征,从而提高了分割的准确性和细节保留程度。Dang等人(2023)提出的LVSegNet(left ventricle segmentation network)模型中,其上采样利用了密集上采样卷积模块来替代常规的上采样方法,通过对经过编码器和中心部分处理后的特征图进行卷积操作,并使用像素重排技术,最大限度地保留了特征图中的重要信息,从而提高图像分割的准确性。Sun等人(2023)提出的新型医学图像分割网络模型DSGA-Net(deeply separable gated transformer and attention strategy network)中,设计了混合三支注意力模块,从而增加上采样过程中的特征数量,同时减少在恢复特征图到原始图像大小时的特征信息损失。Xiong等人(2024)提出的SEA-NET(spiral squeeze-and-excitation and attention net)模型中,上采样通过注意力模块来完成,通过调整特征图的空间信息,将深层语义信息融合到浅层特征图中,避免深层语义信息在连续上采样过程中散失,从而更有效地保留并增强小目标特征,提高分割的精度和效果。Zhu等人(2024)设计了一种双分支交叉注意融合网络(dual-branch cross-attention fusion network, DCFNet),其上采样是通过特征交叉融合(feature cross fusion, FCF)模块和通道注意块(channel attention block, CAB)进行的。FCF模块用于聚合多尺度的特征,以增强解码过程中的特征表达。CAB模块重点强调上采样特征与FCF模块生成特征之间的通道特征,进一步增强解码阶段的细节表示能力。二者共同作用,使得DCFNet能够实现高效且精确的上采样过程,提升分割结果的准确性。Khan等人(2025)提出WaveSeg-UNet轻量级模型,其上采样是通过逆离散小波变换(discrete wavelet transform, DWT)来恢复原始图像尺寸的。在解码器阶段,逆DWT被应用于特征映射,以重建分辨率,并保留重要的细节信息,从而保持上采样过程中的信息完整性。

但是,上采样策略仍面临一些挑战。由于这些方法通常对医学图像的背景和细节处理不足,在需要高精度边界和细微结构分割时容易遗漏重要信息,产生模糊和伪影,以至于分割结果缺乏全面性和精确性。因此,如何进一步优化上采样策略,确保对复杂医学图像的有效分割和诊断支持,成为当前研究的一个关键方向。

2 DEMS-GIF 模型

2.1 整体架构设计

图 1 是本文 DEMS-GIF 的体系结构图,由 Res2Net(Gao 等, 2021)作为主干网络。Res2Net 通过引入更细粒度的分层卷积操作,将每个残差块划分为多个较小的特征组进行特征处理,从而提升特征表示的精细度和多样性。首先,输入

一系列卷积层和池化层处理,在不同层次产生了不同尺度的特征图,而这些特征图在后续编码阶段将分别用于捕获图像的低、中、高 3 级特征,通过层层堆叠的卷积块不断增强特征的丰富度和表达能力。为了降低模型后续计算的复杂度,同时聚合各阶段的有效特征,使用卷积、归一化和激活函数将后三阶段的编码特征通道数统一至 32。其次,利用基于 Transformer 的桥接特征融合模块,将经过通道降维处理的特征进行融合,充分利用中间层的特征信息,形成全局特征图,从而实现图像细节和全局信息的有效提取。在解码阶段,模型连续使用 4 次反向注意下的扩缩区域增强上采样策略来处理不同层次的特征图,进一步捕捉其中的边界和细节信息,确保分割区域的完整性与准确性。最后,使用 1×1 卷积将特征图降为单通道,并使用双线性插值进行上采样操作,输出最终的分割结果。

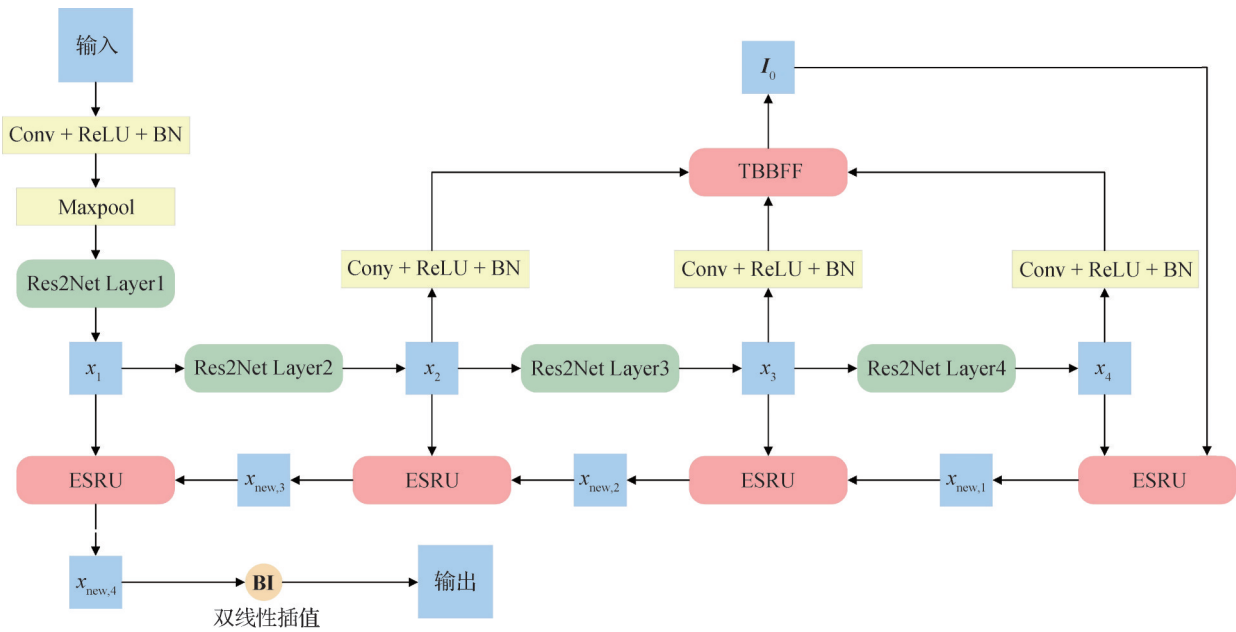


图 1 DEMS-GIF 架构
Fig. 1 The architecture of DEMS-GIF

2.2 基于 Transformer 的桥接特征融合模块

特征融合方法在医学图像分割中具有重要的作用和意义,它能够利用不同特征之间的互补信息提高分割的准确性和模型的鲁棒性。然而,传统的特征融合方法存在特征丢失和信息冗余的问题,在处理多尺度特征和复杂结构时难以有效整合不同尺度和层次的特征,导致模型在识别和分割复杂结构时能力有限。现阶段的一些特征融合方法虽然通过引

入注意力机制等技术提高特征的提取和融合效果,但通常采用的是自上而下或自下而上的方式进行特征融合,而忽略了中间层的特征。中间层特征在捕捉图像细节和全局结构方面起到了桥梁的作用,它既包含了比高层次特征更多的细节信息,又拥有丰富的上下文语义信息,正是这些先天优势,中间层特征对复杂结构或微小病灶的准确识别和分割具有重要意义。因此,针对现有特征融合方法在捕捉复杂

结构和整合中间层特征方面的不足,本文设计一种基于Transformer的桥接特征融合模块(TBBFF),相较于其他特征融合模块,TBBFF模块更加关注中间层特征信息,同时利用Transformer提取不同区域之间的长距离依赖关系的能力,使得网络能够更好地理解图像的整体结构和上下文信息,进一步提升模型的分割性能和鲁棒性。TBBFF模块结构如图2所示。

为后续能够利用Transformer的全局建模能力提取更加丰富和有用的特征,本文先将每个输入的特征图 $I_i \in H_i \times W_i$, $i = 1, 2, 3$ 分别进行展平(flatten)和转置(permute)操作,以适应Transformer编码器的处理格式,具体为

$$I_i = f_{\text{permute}} \left(f_{\text{flatten}} \left(I_i, (B, C, H_i \times W_i) \right), (2, 0, 1) \right) \quad (1)$$

式中, B 是批量大小, C 是通道数, H 和 W 分别是高度

和宽度。将不同尺度的特征图展平并转置后,沿着第一个维度(空间维度)进行连接,形成特征图 I_{cat} ,从而简化特征处理的流程,使模型可以更系统地学习和融合不同尺度的特征。

由于Transformer编码器可以捕捉到特征图中的全局依赖关系,因此该模型将特征图 I_{cat} 输入其中,进而利用Transformer的全局注意力机制生成更加丰富的特征表示,从而关注到整个输入特征图中的所有层次的关系,包括中间层信息。具体为

$$I_i = f_{\text{TransformerEncoder}}(I_{\text{cat}}) \quad (2)$$

此时,特征图 I_i 的尺寸为 $(H_1 \times W_1 + H_2 \times W_2 + H_3 \times W_3, B, C)$ 。接着,对特征图进行还原,这不仅能够保留特征之间的空间关系,而且也能够方便后续的特征融合处理。

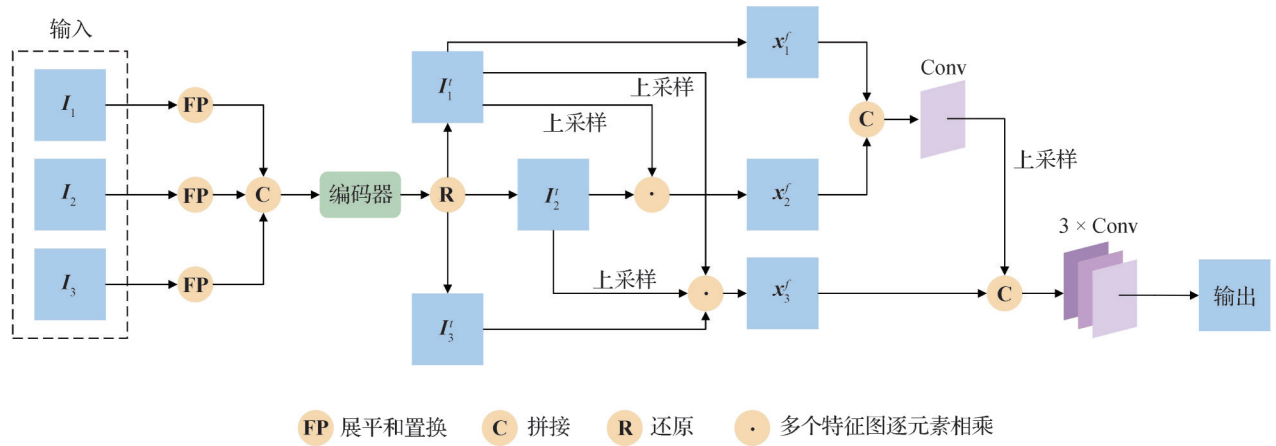


图2 TBBFF模块结构

Fig. 2 The structure of the TBBFF module

通过上采样和逐元素相乘的方式将不同尺度的特征图融合在一起,使得模型能够在每个空间位置上结合不同层次的信息,生成包含丰富细节和全局语义的特征表示。这种融合策略能够注重中间层特征的作用,将其有效地结合到最终的特征表示中,确保了多层次信息的有效利用,增强模型对复杂结构的识别能力,从而提高医学图像分割的准确性。具体为

$$\mathbf{x}_i' = \left(\odot_{i=1}^3 f_{\text{Upsample}}(I_i', s_i) \right) \quad (3)$$

式中, $\odot$ 表示多个特征图逐元素相乘, I_i' 表示经过Transformer编码后得到的特征图, $i = 1, 2, 3$, s_i 表示第 i 个特征图的上采样比例,且满足 $s_i = 2^{3-i}$,

$f_{\text{Upsample}}(I_i', s_i)$ 表示将特征图 I_i' 以比例 s_i 进行上采样操作。

为进一步融合多层次的特征信息,增强模型的特征表达能力,首先将上采样后的低层次特征图 $\mathbf{x}_1'$ 和中间层特征图 $\mathbf{x}_2'$ 进行拼接,并将拼接后的特征图 $\mathbf{x}_2'$ 通过卷积层进一步融合,提取更高层次的特征。然后,将高层次特征图 $\mathbf{x}_3'$ 和上采样后的特征图 $\mathbf{x}_2'$ 进行拼接,同样将拼接后的特征图 $\mathbf{x}_3'$ 通过卷积层再次进行融合。这种方式充分利用中间层特征在整个过程中起到的桥梁作用,使模型能够有效地结合细节信息和全局信息,从而确保了不同层次特征之间的信息传递和综合利用,提高分割的准确性。具体为

$$\begin{cases} X_{\text{fused}} = f_{\text{Conv}}\left(C\left(x_3^f, f_{\text{Upsample}}\left(X_f, 2\right)\right)\right) \\ X_f = f_{\text{Conv}}\left(C\left(x_2^f, f_{\text{Upsample}}\left(x_1^f, 2\right)\right)\right) \end{cases} \quad (4)$$

式中, $C(A, B)$ 表示将特征图 A 和 B 在通道维度上进行拼接, 即对特征图进行卷积操作。

最后, 再对经过拼接和融合得到的特征图 X_3 进行两个卷积操作, 使得模型能够在前面多尺度特征融合的基础上, 进一步提取和处理高级特征, 从而生成更加精细的最终特征图 I_0 , 为后续分割任务提供更具辨识度和分辨力的特征, 提升模型的整体性能。

2.3 反向注意下的扩缩区域增强上采样策略

在医学图像分割任务中, 上采样方法扮演着重要的角色, 优秀的上采样策略不仅能提升图像的质量, 还能显著提高模型的整体性能。但传统的重点区域增强方法通常仅关注前景区域而忽视背景和细节, 分割结果不够精确, 从而使得上采样后的图像缺乏清晰的细节与边界表示。由于医学图像往往包含细微的结构和复杂的背景信息, 容易导致遗漏关键的病变信息。此外, 传统方法生成的图像往往具有过度平滑的效果, 导致分割结果的边界模糊, 不利于精细结构的分割, 并可能引入伪影或不自然的纹理, 影响诊断的准确性。因此, 针对生成图像的过度平滑问题及现有方法对病灶区域分割边界不精细的问题, 本文设计了一种反向注意下的扩缩区域增强上采样策略(expanded and scaled region enhanced upsampling strategy under reverse attention, ESRU), 通过引入反向注意力机制, 并结合腐蚀和膨胀操作, 使得模型能够更好地捕捉医学图像中目标区域的边界和细节信息, 增强分割区域的完整性和连续性, 从而提高分割的精度和稳定性。ESRU 策略结构如图3所示。

首先需要说明的是, 在第1次上采样时, 模型将

融合之后的特征图 $I_0 \in H \times W$ 利用双线性插值方法将其高度和宽度缩小到原来的 $1/4$, 使其与 Transformer 骨干进行特征提取的最小特征图大小相等, 方便后续对于最小特征图的重新赋权。同时, 融合之后的特征图包含不同尺度特征信息, 更容易从上下文层面把控赋权的准确性。具体为

$$I(u, v) = \sum_{i=0}^1 \sum_{j=0}^1 I_0(x_i, y_j) (1 - |u - x_i|) (1 - |v - y_j|) \quad (5)$$

式中, $I_0(x, y)$ 表示输入特征图 I_0 在坐标 (x, y) 的像素值, $x = 4u, y = 4v, x_i$ 和 y_j 表示最近的4个整数坐标, $(1 - |u - x_i|)$ 和 $(1 - |v - y_j|)$ 是插值权重, 基于新坐标 (u, v) 和原始坐标的距离。在后续的上采样过程中, 则是将经过上一次上采样策略的特征图作为输入, 利用双线性插值方法将其高度和宽度扩大为原来的2倍。

其次, 引入反向注意力机制。其能够通过关注背景区域和那些被认为不重要或被忽略的特征, 改善对细节和边缘的恢复效果, 抑制错误分类, 提升模型的特征表达和分类能力。因而, 本文使用反向注意力操作, 得到一个反向的权重掩码 x , 在后续特征融合的过程中选择性强调那些本该被遗弃的有效特征, 具体为

$$x = 1 - \sigma(I) \quad (6)$$

式中, I 表示经过双线性插值方法处理得到的特征图, σ 表示 sigmoid 激活函数。

为有效结合特征和掩码信息, 该模型通过将单通道的掩码扩展为多通道并与原始特征图(Transformer 骨干提取的大小通道数与扩展后的多通道特征图相同)逐元素相乘, 从而为特征处理提供更加丰富的上下文信息。再通过卷积操作, 进一步提取特

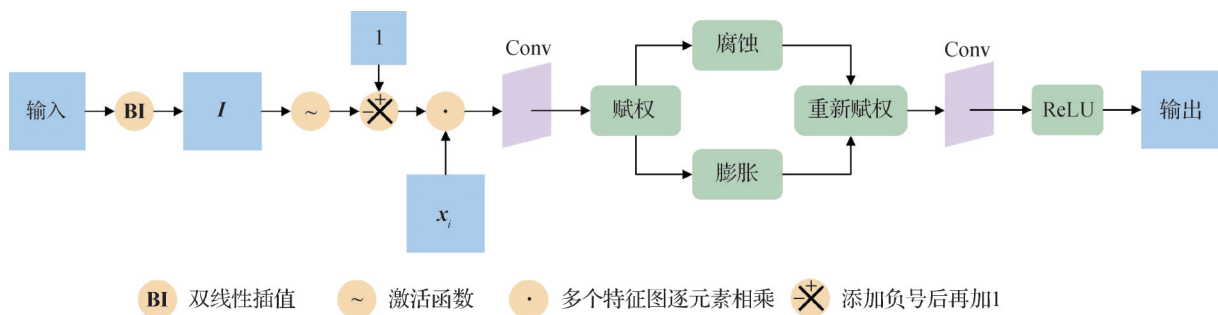


图3 ESRU 策略的结构

Fig. 3 The structure of the ESRU strategy

征图中的局部信息并进行特征变换,为后续提供丰富的特征表示。然后,引入腐蚀和膨胀操作。在图像分割中,腐蚀通过去除不准确的边缘,保留确定的病变区域,从而使得分离出来的病变区域更加准确,而膨胀通过扩展目标区域的边界,进一步扩大病变区域,使网络模型能够重新关注可能的病变区域。通过结合腐蚀和膨胀这两种形态学操作,不仅能够有效处理图像中的噪声和细节,还能使网络在后续的处理过程中更好地识别和分析病变区域的潜在扩展范围,进一步增强分割区域的完整性和连续性。在这一过程中,首先将通过卷积操作得到的特征图 I_c 进行阈值赋权,具体为

$$x_0 = \begin{cases} 1 & I_c \geq 0.6 \\ 0 & I_c < 0.6 \end{cases} \quad (7)$$

然后,结合腐蚀和膨胀操作,对 x 与高层特征图逐元素相乘得到的特征图 G 重新赋权。具体为

$$x_{\text{new}} = \alpha \times G \times x_{\text{ero}} + \beta \times G \times x_{\text{dil}} + \gamma \times G \quad (8)$$

式中, α, β, γ 表示可学习的权重参数, x_{ero} 和 x_{dil} 分别表示特征图 x_0 经过腐蚀和膨胀作用生成的特征图。

3 实验分析

3.1 数据集

3.1.1 CVC-ClinicDB 数据集

CVC-ClinicDB 数据集是广泛用于结肠镜图像分析和竞赛中的标准数据集之一,包含 612 幅从结肠镜检查视频中提取的静态图像,分辨率为 384×288 像素。每幅图像都提供了专家手动标注的息肉分割掩膜,主要用于息肉检测和分割算法的训练和测试。

3.1.2 DDTI 数据集

DDTI 数据集是一个专门用于甲状腺超声图像分析的医学图像数据集,包含 637 幅带有像素级标签的甲状腺超声图像。这些图像的价值在于诊断和病例的多样性,包括甲状腺炎、甲状腺肿、结节和癌症。DDTI 数据集为甲状腺超声图像分析提供了宝贵的资源,对于甲状腺癌的诊断具有重大意义。

3.1.3 Kvasir-SEG 数据集

Kvasir-SEG 数据集是从 Kvasir 数据集中提取出来的一个针对结肠息肉像素级分割的内窥镜数据

集,包含 1 000 幅胃肠息肉图像及其对应的分割掩码,标注了息肉的确切位置和形状。该数据集专门用于开发和评估内镜图像中的息肉分割算法,旨在推动结肠息肉的分割、检测、定位和分类方法的研究和创新。

3.2 评价指标

为了评估模型的分割性能,本文使用召回率(recall)、精度(precision)、病变区域的交并比(intersection over union for lesion region, IoU_{ls})、背景区域的交并比(intersection over union for background region, IoU_{bg})、平均交并比(mean intersection over union, mIoU)以及 Dice 系数作为病变分割的评价指标。其中,recall 衡量的是实际为正样本的像素中被正确预测为正样本的比例,即模型的覆盖能力。precision 衡量模型预测为正样本的像素中实际为正样本的比例,即预测的准确性。 IoU_{ls} 表示病变区域的交并比,衡量分割结果与病变真实标注的重叠情况。 IoU_{bg} 表示背景区域的交并比,衡量分割结果与背景真实标注的重叠情况。 mIoU 是所有类别(包括背景和病变区域)的 IoU 的平均值,提供了模型整体分割性能的衡量。Dice 系数则综合了精度和召回率,作为分割结果与真实标注的相似性度量。

这些评价指标与 4 个值相关,即真阳性(true positive, TP)、真阴性(true negative, TN)、假阳性(false positive, FP)和假阴性(false negative, FN)。

3.3 损失函数

本文研究选用了 BCE Dice Loss。BCE Dice Loss 是一种结合了二元交叉熵(binary cross-entropy, BCE)和 Dice 系数的损失函数,常用于图像分割任务中。这种损失函数能够兼顾 BCE 和 Dice Loss 的优点,平衡像素级别的精度和整体分割区域的一致性,从而在图像分割任务中取得更好的性能。

1)二元交叉熵损失。BCE 损失是经典的分类损失函数,适用于每个像素的二分类任务。它通过计算预测值与真实标签之间的交叉熵,衡量每个像素位置的预测误差。具体为

$$\text{BCE}(p, t) = -[t \log(p) + (1 - t) \log(1 - p)] \quad (9)$$

式中, p 是预测概率, t 是真实标签(0 或 1)。

2)Dice 系数损失。在图像分割中, Dice 系数用于衡量预测分割区域与真实分割区域的重叠。Dice 系数越高,表示预测越准确。具体为

$$Dice(p, t) = \frac{2 \sum_i (p_i t_i) + \epsilon}{\sum_i p_i + \sum_i t_i + \epsilon} \quad (10)$$

式中, p_i 和 t_i 分别是第 i 个像素的预测值和真实值, ϵ 是一个平滑项, 避免分母为 0。

通过将 BCE 和 Dice 损失相结合, 既考虑了像素级的精度, 又考虑了整体的分割质量, 因而在处理图像分割任务时能够取得更好的效果。具体为

$$Loss(p, t) = BCE(p, t) + (1 - Dice(p, t)) \quad (11)$$

3.4 实验细节

所有实验均基于 Ubuntu20.04 操作系统平台, 并使用 Python 3.8 和 PyTorch2.0 进行运算。该系统配备了 Nvidia RTX 3080Ti GPU, 内存容量为 12 GB, 以及 Intel(R) Xeon(R) Silver 4214R CPU, 频率为 2.40 GHz。对于 CVC-ClinicDB、DDTI 以及 Kvasir-SEG 3 个数据集, 输入图像尺寸均设置为 256×256 像素, 并通过平均值和标准偏差进行归一化。本文研究将数据集均以 9:1 的比例随机分为训练集和测试集, 具体来说, CVC-ClinicDB 数据集分为 551 幅图像用于训练, 61 幅用于测试, DDTI 数据集分为 573 幅图像用于训练, 64 幅用于测试, 而 Kvasir-SEG 数据集分为 900 幅图像用于训练, 100 幅用于测试。模型在训练时采用了交叉验证, 其中训练集用来训练与更新参数, 测试集只用来评估模型好坏而不参与模型参数更新, 且训练集与测试集无交集。

为了增强模型的泛化能力, 从而提高模型的性能, 实验对 CVC-ClinicDB、DDTI 以及 Kvasir-SEG 训练数据集使用了数据增强技术, 每幅图像有 30% 的概率会被旋转、水平翻转、垂直翻转和随机遮挡。其中, 旋转角度要求在 0° 到 35° 之间, 水平翻转是指进行左右镜像翻转, 垂直翻转是指进行上下镜像翻转, 随机遮挡中要求在图像上随机遮挡最多 10 个区域, 每个区域的最大高度和宽度为 32 像素。此外, 模型使用 Adam 作为网络优化器, 并将初始学习率设置为 0.0001, 批量大小为 8, 训练轮数为 150 次。损失函数采用 BCE Dice Loss。

由于每个参与评价的模型训练 4 次, 每次都会得到相应的评价指标, 因而本文对得到的评价指标求取了均值与标准差。其中, 均值体现了模型的平均分割能力, 而标准差体现了模型的鲁棒性, 且标准差越小表示模型的鲁棒性越好。

3.5 对比实验结果

为了评价 DEMS-GIF 的优越性, 本文在 3 个不同的数据集 CVC-ClinicDB、DDTI 和 Kvasir-SEG 上将其与最新或经典的方法进行了比较。这 8 种方法是 UNet (Ronneberger 等, 2015)、UNet++ (Zhou 等, 2018)、ResUNet++ (Jha 等, 2019)、MSUNet (multi-scale UNet) (Su 等, 2021)、MFSNet (multi-focus segmentation network) (Basak 等, 2022)、MBSNet (multi-branch segmentation network) (Jin 等, 2023)、DTA-UNet (UNet based on dynamic convolution decomposition and triplet attention) (Li 等, 2024b) 和 SCSNet (spatial-channel synergistic optimization net) (Chen 等, 2024)。结果表明, DEMS-GIF 在 3 个数据集上均取得了较好的分割效果, 具有更好的学习能力和预测能力。

3.5.1 CVC-ClinicDB 数据集上的对比实验结果

模型在 CVC-ClinicDB 数据集上进行实验的结果如表 1 所示。对于医学图像分割, 网络模型在 recall、precision、IoU_{ls}、IoU_{bg}、mIoU 和 Dice 上度量的性能通常受到更多关注。从表 1 可以看出, DEMS-GIF 的 recall 为 95.48%, precision 为 94.52%, IoU_{ls} 为 90.42%, IoU_{bg} 为 99.07%, mIoU 为 94.74%, Dice 为 94.82%。与传统网络相比, DEMS-GIF 具有显著改进, 尤其在 IoU_{ls} 方面, 比 UNet 高出 6.65%, 比 UNet++ 高出 6.48%, 比 ResUNet++ 高出 14.88%。与最近提出的几种分割方法进行比较, DEMS-GIF 的 mIoU 比 MSUNet 高出 3.98%, 比 MBSNet 高出 4.42%, 比 SCSNet 高出 20.13%。

图 4 是 DEMS-GIF 与其他网络在 CVC-ClinicDB 数据集上 mIoU 值的对比, 可以看出, 随着训练次数的增加, DEMS-GIF 更具优越性, 在第 70 个 epoch 后, 其 mIoU 值明显高于其他 8 个网络模型。

图 5 是 DEMS-GIF 与其他网络在 CVC-ClinicDB 数据集上 Train Loss 的对比, 由图可见, 随着训练次数的增加, DEMS-GIF 的 Train Loss 越来越低, 且明显低于 ResUNet++ 和 SCSNet。

3.5.2 DDTI 数据集上的对比实验结果

在 DDTI 数据集上的实验结果如表 2 所示。从表 2 可以看出, DEMS-GIF 的 recall 为 84.97%, precision 为 85.52%, IoU_{ls} 为 72.45%, IoU_{bg} 为 96.68%, mIoU 为 84.56%, Dice 为 82.95%。与传统网络模型相比, DEMS-GIF 与 ResUNet++ 的差距最为明显, 其

表 1 CVC-ClinicDB 数据集上的对比实验结果

Table 1 The comparative experimental results on the CVC-ClinicDB dataset

模型	recall	precision	IoU _{ls}	IoU _{bg}	mIoU	Dice
UNet	90.60±0.55	90.86±0.88	83.77±1.20	98.24±0.22	91.01±0.71	89.98±0.79
UNet++	91.05±0.71	90.66±0.70	83.94±1.00	98.34±0.09	91.14±0.50	89.96±0.83
ResUNet++	84.60±0.70	86.68±1.12	75.54±0.51	97.47±0.25	86.50±0.35	83.52±0.32
MSUNet	89.85±0.55	91.22±0.99	83.07±0.56	98.44±0.13	90.76±0.35	89.63±0.35
MFSNet	93.64±0.64	94.27±0.86	88.88±0.88	99.02±0.03	93.95±0.45	93.58±0.84
MBSNet	89.63±0.38	90.24±0.41	82.56±0.52	98.08±0.12	90.32±0.26	89.29±0.38
DTA-UNet	91.69±0.79	90.18±1.00	84.09±1.25	98.26±0.32	91.17±0.78	90.24±0.88
SCSONet	70.87±3.63	70.80±2.04	54.57±2.15	94.67±0.27	74.61±1.19	66.54±2.17
DEMS-GIF(本文)	95.48±0.27	94.52±0.55	90.42±0.28	99.07±0.03	94.74±0.15	94.82±0.17

注:加粗字体表示各列最优结果。

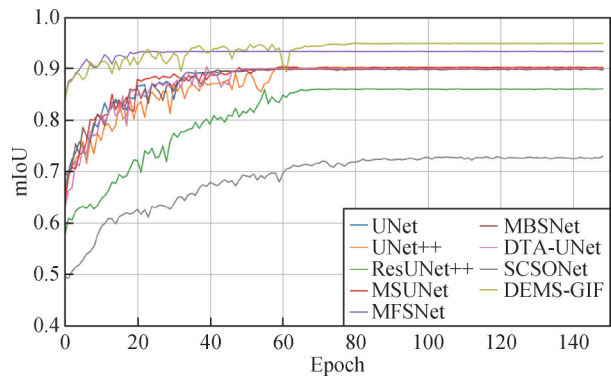


图 4 CVC-ClinicDB 数据集上不同网络的 mIoU 对比
Fig. 4 The mIoU comparison of different networks on the CVC-ClinicDB dataset

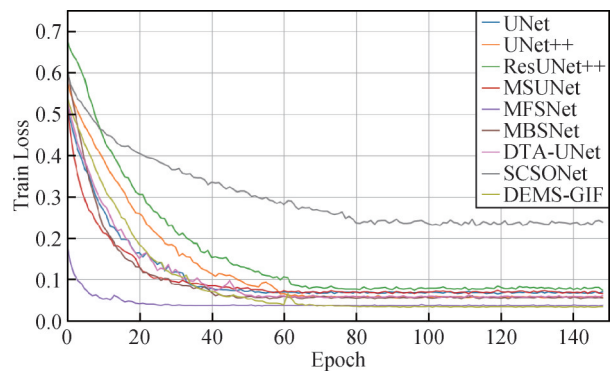


图 5 CVC-ClinicDB 数据集上不同网络的 Train Loss 对比
Fig. 5 The Train Loss comparison of different networks on the CVC-ClinicDB dataset

中 precision、IoU_{ls} 和 mIoU 分别高出 13.97%、19.15%、10.89%。与最近提出的几种分割方法相比,DEMS-GIF 的所有指标也都表现出最好的性能,其 mIoU 分

别比 MBSNet、DTA-UNet 以及 SCSONet 高出 4.23%、2.68%、14.88%。

图 6 是 DEMS-GIF 与其他网络在 DDTI 数据集上 mIoU 值的对比,可以观察到,DEMS-GIF 在第 60 个 epoch 后,其 mIoU 值逐渐趋于稳定,介于 0.8 至 0.9 之间,且高于其他 8 个网络模型。

图 7 是 DEMS-GIF 与其他网络在 DDTI 数据集上 Train Loss 的对比,随着训练次数的增加,DEMS-GIF 的 Train Loss 在第 40~60 个 epoch 之间时已明显低于除 MFSNet 外的其他 7 个网络模型,在第 70 个 epoch 之后,其 Train Loss 明显低于 MFSNet,此时 DEMS-GIF 的 Train Loss 低于所有对比模型。

3.5.3 Kvasir-SEG 数据集上的对比实验结果

在 Kvasir-SEG 数据集上的实验结果如表 3 所示。从表 3 可以看出,DEMS-GIF 的 recall、precision、IoU_{ls}、IoU_{bg}、mIoU、Dice 分别为 98.68%、91.36%、81.40%、95.53%、88.46%、87.44%。其中 Dice 比 UNet 高出 6.82%, recall 比 UNet++ 高出 16.17%, mIoU 比 ResUNet++ 高出 12.69%,与传统分割方法相比,DEMS-GIF 显然具有极大的优越性。与其他模型如 MBSNet、DTA-UNet 以及 SCSONet 相比,DEMS-GIF 的 mIoU 分别高出 6.77%、3.84%、25.02%,达到了最优。

图 8 是 DEMS-GIF 与其他网络在 Kvasir-SEG 数据集上 mIoU 值的对比,由图可见,DEMS-GIF 的 mIoU 值随着训练次数的增加越来越高,在第 40 个

表 2 DDTI数据集上的对比实验结果

Table 2 The comparative experimental results on the DDTI dataset

模型	recall	precision	IoU _{ls}	IoU _{bg}	mIoU	Dice
UNet	80.07±1.05	82.48±1.37	66.32±0.30	96.01±0.07	81.16±0.14	77.50±0.49
UNet++	80.33±0.89	82.43±1.41	66.39±0.25	95.96±0.04	81.18±0.13	77.73±0.15
ResUNet++	74.77±0.69	71.55±3.45	53.30±3.33	94.04±0.55	73.67±1.94	66.84±2.85
MSUNet	81.68±2.08	82.36±1.63	66.84±1.42	95.93±0.18	81.38±0.79	77.94±1.01
MFSNet	83.68±0.79	83.60±0.31	70.04±0.67	96.46±0.13	83.25±0.35	80.82±0.72
MBSNet	78.12±1.29	82.38±0.98	64.73±1.37	95.92±0.20	80.33±0.75	76.53±1.40
DTA-UNet	80.64±0.90	83.15±1.39	67.58±0.27	96.18±0.06	81.88±0.16	78.58±0.34
SCSONet	72.92±2.07	64.36±0.93	46.62±1.36	92.74±0.41	69.68±0.85	61.24±1.35
DEMS-GIF(本文)	84.97±1.39	85.52±1.13	72.45±0.70	96.68±0.09	84.56±0.38	82.95±0.61

注:加粗字体表示各列最优结果。

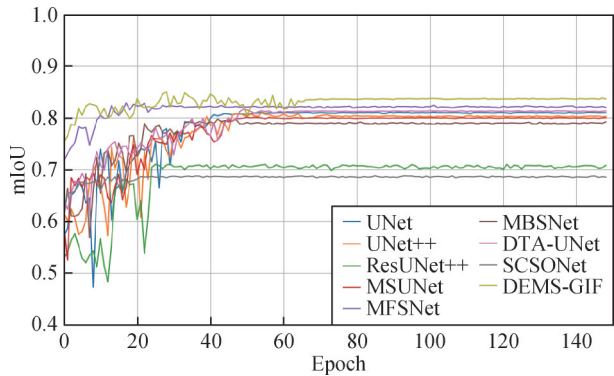


图 6 DDTI数据集上不同网络的mIoU对比
Fig. 6 The mIoU comparison of different networks on the DDTI dataset

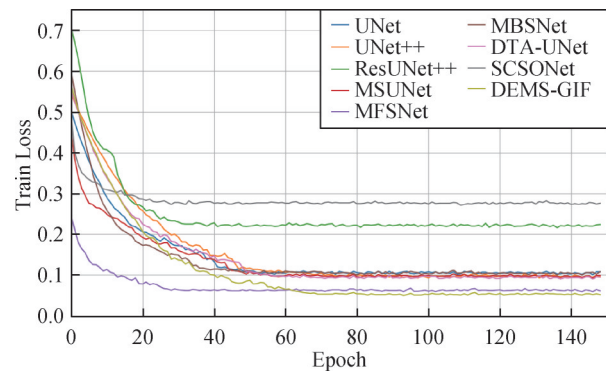


图 7 DDTI数据集上不同网络的Train Loss对比
Fig. 7 The Train Loss comparison of different networks on the DDTI dataset

epoch后,其mIoU值在所有对比模型中最高。

图 9 是 DEMS-GIF 与其他网络在 Kvasir-SEG 数据集上 Train Loss 的对比,DEMS-GIF 的 Train Loss 随

着训练次数的增加越来越低,在第 20 个 epoch 之后,其 Train Loss 已明显低于除 MFSNet 外的其他 7 个网络模型。

3. 5. 4 可视化结果分析

为更加清晰地直观比较不同网络模型对病变区域的分割预测程度,实验对 DEMS-GIF 与其他 8 个网络模型进行了可视化,如图 10 所示。通过比较可以发现,在 CVC-ClinicDB 数据集中,DEMS-GIF 与 MFSNet 的分割结果更接近于真实值,而其他方法均存在一定局限性。在 DDTI 数据集中,显然 DEMS-GIF 能够更为全面地识别病变区域,取得良好的分割效果。在 Kvasir-SEG 上,可以明显看出,SCSONet 效果较差,与 DEMS-GIF 的分割结果相差较大。

综合来看,在不同数据集上,DEMS-GIF 比其他网络能够更完整地分割病变区域,进一步证明了 DEMS-GIF 模型在医学图像分割任务中的显著改进。

3. 6 消融实验结果

在本小节中,本文进行了消融实验研究,以更好地了解所提出的 DEMS-GIF 中每个模块和结构内部的有效性。这项消融研究主要包括两个方面:模块消融和参数消融。其中,模块消融是在 3 个数据集(CVC-ClinicDB、DDTI 和 Kvasir-SEG)上进行的,而参数消融是在 CVC-ClininDB 数据集上进行的。

3. 6. 1 模块消融

为了充分讨论基于 Transformer 的桥接特征融合

表 3 Kvasir-SEG 数据集上的对比实验结果
Table 3 The comparative experimental results on the Kvasir-SEG dataset

模型	recall	precision	IoU _{ls}	IoU _{bg}	mIoU	Dice
UNet	83.49±0.41	86.97±1.11	72.98±1.09	93.45±0.31	83.22±0.70	80.62±0.87
UNet++	82.51±0.80	86.56±0.99	71.93±1.27	93.01±0.31	82.47±0.77	79.68±1.20
ResUNet++	71.65±1.27	82.68±2.38	60.26±1.52	91.28±0.42	75.77±0.97	68.75±1.28
MSUNet	83.08±0.57	86.86±0.20	72.23±0.72	92.94±0.26	82.58±0.48	79.58±0.71
MFSNet	88.32±1.14	90.16±0.69	80.36±0.79	95.32±0.27	87.84±0.51	86.81±0.75
MBSNet	82.21±0.64	85.20±0.75	70.60±1.35	92.79±0.22	81.69±0.76	78.42±1.20
DTA-UNet	87.52±0.71	86.29±0.28	75.50±0.63	93.75±0.29	84.62±0.45	82.71±0.65
SCSNet	63.55±2.05	58.69±0.77	40.60±0.89	86.27±0.15	63.44±0.38	52.36±1.41
DEMS-GIF(本文)	98.68±0.35	91.36±1.00	81.40±0.25	95.53±0.12	88.46±0.16	87.44±0.18

注:加粗字体表示各列最优结果。

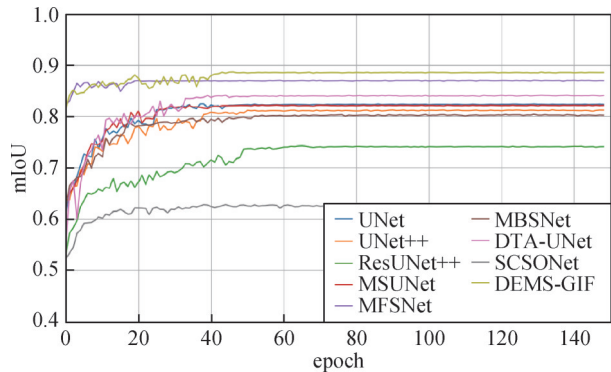


图 8 Kvasir-SEG 数据集上不同网络的 mIoU 对比
Fig. 8 The mIoU comparison of different networks on the Kvasir-SEG dataset

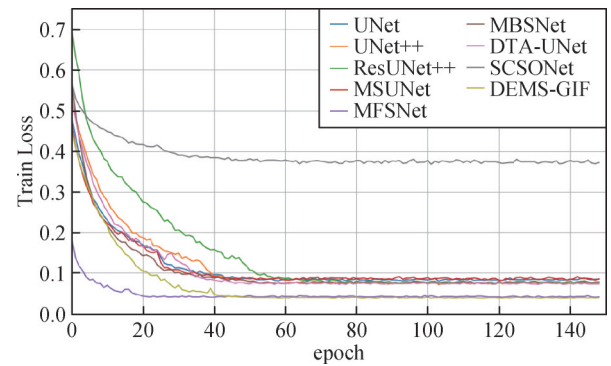


图 9 Kvasir-SEG 数据集上不同网络的 Train Loss 对比
Fig. 9 The Train Loss comparison of different networks on the Kvasir-SEG dataset

模块和反向注意下的扩缩区域增强上采样策略对模型性能的影响,模型使用 Res2Net 作为基线网络,在 CVC-ClinicDB、DDTI 和 Kvasir-SEG 这 3 个数据集上

分别进行了消融研究。实验结果如表 4 所示。方法 A 表示模型采用的是 Res2Net 网络,其上采样方式为原始 UNet 的方式。方法 B 表示的是所设计的 DEMS-GIF 模型的最终结构采用了普通的上采样策略,但加入了基于 Transformer 的桥接特征融合模块。方法 C 表示的是所设计的 DEMS-GIF 模型的最终结构采用了普通融合方式,但加入了反向注意下的扩缩区域增强上采样策略。

在各个数据集上分别与方法 A、B、C 相比,该模型在多项分割指标上均取得了显著提升。在 CVC-ClinicDB 数据集中,DEMS-GIF 与方法 A 相比,precision 和 mIoU 分别提高 4.21% 和 3.48%。与方法 B 相比,precision 和 mIoU 分别提高 2.46% 和 1.32%。与方法 C 相比,precision 和 mIoU 分别提高 1.04% 和 0.45%。方法 B 与方法 A 相比,precision 和 mIoU 分别提高 1.75% 和 2.16%。方法 C 和方法 A 相比,precision 和 mIoU 分别提高 3.17% 和 3.03%。在 DDTI 数据集中,DEMS-GIF 与方法 A 相比,precision 和 mIoU 分别提高 4.60% 和 3.15%。与方法 B 相比,precision 和 mIoU 分别提高 0.45% 和 0.89%。与方法 C 相比,precision 和 mIoU 分别提高 1.82% 和 0.50%。方法 B 与方法 A 相比,precision 和 mIoU 分别提高 4.15% 和 2.26%。方法 C 和方法 A 相比,precision 和 mIoU 分别提高 2.78% 和 2.65%。在 Kvasir-SEG 数据集中,DEMS-GIF 与方法 A 相比,precision 和 mIoU 分别提高 3.82% 和 3.28%。与方法 B 相比,pre-

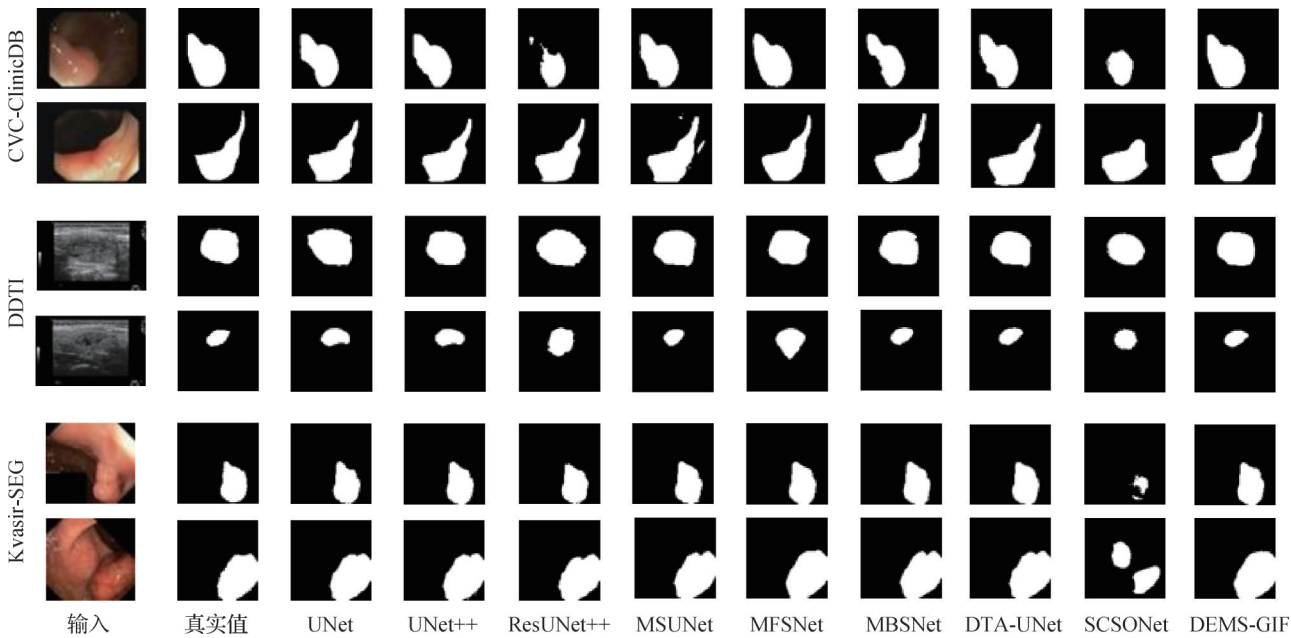


图 10 不同数据集上 DEMS-GIF 与其他模型的可视化结果比较

Fig. 10 Visual comparison of DEMS-GIF and other models on different datasets

cision和mIoU分别提高1.90%和0.48%。与方法C相比,mIoU提高0.30%。方法B与方法A相比,precision和mIoU分别提高1.92%和2.80%。方法C和方法A相比,precision和mIoU分别提高3.83%和2.98%。

模块的消融实验结果表明,基于Transformer的桥接特征融合模块和反向注意下的扩缩区域增强上采样策略的结合,有利于网络更加关注病变区域,从

而进行更为准确的分割。因此,本文研究采取了同时加入TBBFF模块和反向注意下的扩缩区域增强上采样策略的方法,设计了DEMS-GIF模型。

3.6.2 参数消融

为进一步理解DEMS-GIF模型中每个参数的贡献,优化模型设置,本文对模型中涉及到的参数进行了消融研究,从而进一步提升模型的整体性能和可

表 4 DEMS-GIF 结构的消融研究
Table 4 Ablation study of the DEMS-GIF architecture

数据集	模型	recall/%	precision/%	IoU _s /%	IoU _{bg} /%	mIoU/%	Dice/%
CVC-ClinicDB	A	91.48±0.61	90.31±1.37	84.02±0.60	98.51±0.10	91.26±0.35	90.42±0.47
	B	95.24±0.70	92.06±0.47	88.01±0.29	98.84±0.05	93.42±0.16	93.25±0.33
	C	95.70±0.26	93.48±0.51	89.62±0.32	98.97±0.06	94.29±0.17	94.33±0.22
	DEMS-GIF	95.48±0.27	94.52±0.55	90.42±0.28	99.07±0.03	94.74±0.15	94.82±0.17
DDTI	A	82.02±0.71	80.92±1.43	66.68±0.55	96.16±0.2	81.41±0.37	78.53±0.50
	B	83.64±0.96	85.07±1.21	70.76±1.15	96.58±0.16	83.67±0.64	81.54±1.12
	C	85.08±1.68	83.70±1.18	71.58±0.41	96.54±0.13	84.06±0.21	82.20±0.25
	DEMS-GIF	84.97±1.39	85.52±1.13	72.45±0.70	96.68±0.09	84.56±0.38	82.95±0.61
Kvasir-SEG	A	85.05±1.21	87.54±0.83	75.73±0.73	94.62±0.38	85.18±0.55	83.72±0.67
	B	89.33±0.91	89.46±1.29	80.47±0.47	95.50±0.26	87.98±0.32	86.92±0.46
	C	87.84±0.50	91.37±0.49	80.80±0.44	95.52±0.18	88.16±0.29	86.84±0.53
	DEMS-GIF	98.68±0.35	91.36±1.00	81.40±0.25	95.53±0.12	88.46±0.16	87.44±0.18

注:加粗字体表示不同数据集下的各列最优结果。

靠性。其中,TBBFF 模块中的参数主要涉及多头注意力机制的头数和 Transformer 编码器的层数,ESRU 策略中的参数主要涉及利用腐蚀和膨胀操作进行重新赋权时的阈值 I_c 。此外,本文还选取了 ESRU 策略中参数 α,β,γ 的不同值进行实验,进一步分析不同参数的设置对模型性能的影响。

表 5 是对 TBBFF 模块中的多头注意力机制的头数 nhead 和 Transformer 编码器的层数 num_encoder 进行消融研究的结果。根据实验结果发现模型在 nhead = 8, num_encoder = 4 时实验结果最好,且 recall 值为 95.48%, IoU_{is} 值为 90.42%, mIoU 值为 94.74%, Dice 值为 94.82%。

表 5 对 TBBFF 模块中的参数进行消融研究的结果

Table 5 Results of parameter ablation study on the TBBFF module

nhead	num_encoder	recall/%	precision/%	IoU _{is} /%	IoU _{bg} /%	mIoU/%	Dice/%
4	3	95.13±0.55	94.08±0.57	89.82±0.76	99.03±0.08	94.43±0.42	94.29±0.71
8	3	95.17±0.46	94.57±0.45	90.21±0.50	99.10±0.03	94.66±0.26	94.64±0.38
16	3	95.32±0.28	94.18±0.86	90.01±0.71	99.05±0.07	94.53±0.39	94.54±0.48
8	2	94.60±0.76	95.12±0.40	90.14±0.50	99.10±0.05	94.62±0.27	94.49±0.56
8	4	95.48±0.27	94.52±0.55	90.42±0.28	99.07±0.03	94.74±0.15	94.82±0.17

注:加粗字体表示各列最优结果。

表 6 是对 ESRU 策略中利用腐蚀和膨胀操作进行重新赋权时的阈值 I_c 进行消融研究。根据实验结果,发现在阈值为 0.60 时实验结果最好,且 mIoU 值为 94.74%,Dice 达到 94.82%。

表 7 是通过选取 ESRU 策略中的参数 α,β,γ 的

不同值进行实验的结果。通过对比可以发现,不同参数设置会对模型性能带来显著影响。

图 11 展示了不同参数下 DEMS-GIF 模型的 mIoU 和 Dice 对比,显然可以更加直观地看出,参数的不同对模型性能的影响也不同,且当参数为 $\alpha =$

表 6 对利用腐蚀和膨胀操作进行重新赋权的阈值进行消融研究的结果

Table 6 Results of the ablation study on threshold reassignment using erosion and dilation operations

I_c	recall/%	precision/%	IoU _{is} /%	IoU _{bg} /%	mIoU/%	Dice/%
0.50	95.22±0.64	94.03±0.34	89.78±0.48	99.06±0.06	94.41±0.24	94.20±0.53
0.55	95.12±0.20	94.12±0.69	89.75±0.57	99.04±0.05	94.40±0.30	94.32±0.49
0.60	95.48±0.27	94.52±0.55	90.42±0.28	99.07±0.03	94.74±0.15	94.82±0.17
0.65	95.59±0.22	94.08±0.63	90.16±0.45	90.16±0.45	94.61±0.25	94.64±0.27
0.70	95.60±0.35	94.26±0.56	90.26±0.48	99.03±0.05	94.65±0.25	94.74±0.30

注:加粗字体表示各列最优结果。

表 7 不同参数下 DEMS-GIF 模型的实验结果

Table 7 Experimental results of the DEMS-GIF model under different parameters

α	β	γ	recall/%	precision/%	IoU _{is} /%	IoU _{bg} /%	mIoU/%	Dice/%
0.05	0.05	0.9	95.45±0.33	94.22±0.64	90.18±0.54	99.06±0.04	94.62±0.29	94.56±0.49
0.10	0.10	0.8	95.48±0.27	94.52±0.55	90.42±0.28	99.07±0.03	94.74±0.15	94.82±0.17
0.15	0.15	0.7	95.05±0.40	94.58±0.57	90.08±0.39	99.06±0.03	94.58±0.21	94.58±0.26
0.20	0.20	0.6	95.06±0.39	94.39±0.37	89.90±0.15	99.06±0.02	94.48±0.08	94.48±0.11

注:加粗字体表示各列最优结果。

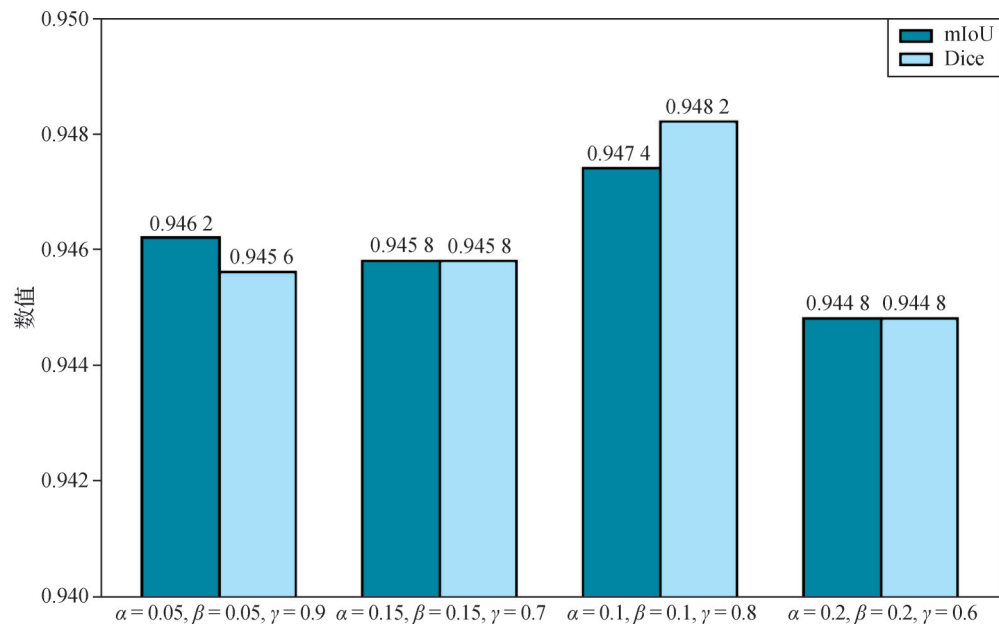


图 11 不同参数下 DEMS-GIF 模型的 mIoU 和 Dice 对比

Fig. 11 Comparison of mIoU and Dice for the DEMS-GIF model under different parameters

0.1, $\beta = 0.1$, $\gamma = 0.8$ 时, DEMS-GIF 模型更具优越性。

4 结 论

本文提出一种聚焦全局与中间层特征的细节增强医学图像分割网络 DEMS-GIF。该网络模型通过结合基于 Transformer 的桥接特征融合模块和反向注意下的扩缩区域增强上采样策略,在充分利用中间层特征信息来整合不同尺度和层次的特征的同时,也进一步关注病灶区域的边界和细节信息,实现了图像特征的高效提取和精细化处理,使得分割区域更具有完整性和准确性。本文在 3 个医学图像分割数据集上对 DEMS-GIF 进行了评估。通过大量实验证明,DEMS-GIF 在 IoU_{ls} 、 IoU_{bg} 、mIoU 以及 Dice 等多个指标上取得了比其他网络模型更高的分数,其中 mIoU 分别为 94.74%、84.56% 和 88.46%,Dice 分别为 94.82%、82.95% 和 87.44%。可见,该模型具有强大的预测和学习能力,在分割任务中展现出良好的泛化能力和鲁棒性,对医学图像分析具有重要的实际应用价值。

在未来的工作中,可以进一步调整模型中的相关参数,优化模型结构,从而将其应用于更多的医学数据集,提高网络模型的泛化性能。

参考文献 (References)

Alom M Z, Hasan M, Yakopcic C, Taha T M and Asari V K. 2018. Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation [EB/OL]. [2024-08-09]. <https://arxiv.org/pdf/1802.06955.pdf>

Azad R, Al-Antary M T, Heidari M and Merhof D. 2022. TransNorm: transformer provides a strong spatial normalization mechanism for a deep segmentation model. IEEE Access, 10: 108205-108215 [DOI: 10.1109/ACCESS.2022.3211501]

Basak H, Kundu R and Sarkar R. 2022. MFSNet: a multi focus segmentation network for skin lesion segmentation. Pattern Recognition, 128: #108673 [DOI: 10.1016/j.patcog.2022.108673]

Chen H Y, Li Z X, Huang X Y, Peng Z W, Deng Y C, Tang L and Yin L. 2024. SCSONet: spatial-channel synergistic optimization net for skin lesion segmentation. Frontiers in Physics, 12: #1388364 [DOI: 10.3389/fphy.2024.1388364]

Dang H, Li M, Tao X X, Zhang G and Qi X Q. 2023. LVSegNet: a novel deep learning-based framework for left ventricle automatic segmentation using magnetic resonance imaging. Computer Communications, 208: 124-135 [DOI: 10.1016/j.comcom.2023.05.011]

El-Assiouti O S, Hamed G, El-Saadawy H, Ebied H M and Khattab D. 2023. RegionInpaint, Cutoff and RegionMix: introducing novel augmentation techniques for enhancing the generalization of brain tumor identification. IEEE Access, 11: 83232-83250 [DOI: 10.1109/access.2023.3301873]

Gao S H, Cheng M M, Zhao K, Zhang X Y, Yang M H and Torr P. 2021. Res2Net: a new multi-scale backbone architecture. IEEE

- Transactions on Pattern Analysis and Machine Intelligence, 43(2): 652-662 [DOI: 10.1109/TPAMI.2019.2938758]
- Gu R, Wang G T, Song T, Huang R, Aertsen M, Deprest J, Ourselin S, Vercauteren T and Zhang S T. 2021. CA-Net: comprehensive attention convolutional neural networks for explainable medical image segmentation. IEEE Transactions on Medical Imaging, 40(2): 699-711 [DOI: 10.1109/TMI.2020.3035253]
- Guo X Y, Lin X, Yang X, Yu L, Cheng K T and Yan Z Q. 2024. UCT-Net: uncertainty-guided CNN-transformer hybrid networks for medical image segmentation. Pattern Recognition, 152: #110491 [DOI: 10.1016/j.patcog.2024.110491]
- Hu K, Zhu Y D, Zhou T X, Zhang Y, Cao C H, Xiao F and Gao X P. 2023. DSC-Net: a novel interactive two-stream network by combining transformer and CNN for ultrasound image segmentation. IEEE Transactions on Instrumentation and Measurement, 72: #5030012 [DOI: 10.1109/TIM.2023.3322993]
- Jha D, Smedsrud P H, Riegler M A, Johansen D, De Lange T, Halvorsen P and Johansen H D. 2019. ResUNet++: an advanced architecture for medical image segmentation//Proceedings of 2019 IEEE International Symposium on Multimedia (ISM). San Diego, USA: IEEE: 225-2255 [DOI: 10.1109/ism46123.2019.00049]
- Jin S Z, Yu S, Peng J, Wang H Y and Zhao Y. 2023. A novel medical image segmentation approach by using multi-branch segmentation network based on local and global information synchronous learning. Scientific Reports, 13(1): #6762 [DOI: 10.1038/s41598-023-33357-y]
- Khan H U, Raza B, Khan M A I and Faheem M. 2025. WaveSeg-UNet model for overlapped nuclei segmentation from multi-organ histopathology images. CAAI Transactions on Intelligence Technology, 10(1): 253-267 [DOI: 10.1049/cit2.12351]
- Li H Y, Ding J Y, Shi X, Zhang Q, Yu P F and Li H S. 2024a. D-SAT: dual semantic aggregation transformer with dual attention for medical image segmentation. Physics in Medicine and Biology, 69(1): #015013 [DOI: 10.1088/1361-6560/acf2e5]
- Li Y, Yan B B, Hou J X, Bai B Y, Huang X Y, Xu C F and Fang L M. 2024b. UNet based on dynamic convolution decomposition and triplet attention. Scientific Reports, 14(1): #271 [DOI: 10.1038/s41598-023-50989-2]
- Litjens G, Kooi T, Bejnordi B E, Setio A A A, Ciompi F, Ghafoorian M, van der Laak J A W M, van Ginneken B and Sánchez C I. 2017. A survey on deep learning in medical image analysis. Medical Image Analysis, 42: 60-88 [DOI: 10.1016/j.media.2017.07.005]
- Liu L J, Li Y, Wu Y L, Ren L L and Wang G L. 2023. LGI Net: enhancing local-global information interaction for medical image segmentation. Computers in Biology and Medicine, 167: #107627 [DOI: 10.1016/j.combiomed.2023.107627]
- Liu L L, Wang Y, Chang J, Zhang P, Liang G B and Zhang H. 2022a. LLRHNet: multiple lesions segmentation using local-long range features. Frontiers in Neuroinformatics, 16: #859973 [DOI: 10.3389/fninf.2022.859973]
- Liu Y, Mu F H, Shi Y and Chen X. 2022b. SF-Net: a multi-task model for brain tumor segmentation in multimodal MRI via image fusion. IEEE Signal Processing Letters, 29: 1799-1803 [DOI: 10.1109/lsp.2022.3198594]
- Lu L and Qi W M. 2023. Spine CT image segmentation based on transformer. Journal of Image and Graphics, 28(11): 3618-3628 (卢玲, 漆为民. 2023. 基于Transformer的脊椎CT图像分割. 中国图象图形学报, 28(11): 3618-3628) [DOI: 10.11834/jig.221084]
- Luo Z M, Zhang Y, Zhou L, Zhang B G, Luo J N and Wu H B. 2019. Micro-vessel image segmentation based on the AD-UNet model. IEEE Access, 7: 143402-143411 [DOI: 10.1109/ACCESS.2019.2945556]
- Ouyang J H, Liu S G, Peng H, Garg H and Thanh D N H. 2023. LEA U-Net: a U-Net-based deep learning framework with local feature enhancement and attention for retinal vessel segmentation. Complex and Intelligent Systems, 9(6): 6753-6766 [DOI: 10.1007/s40747-023-01095-3]
- Pan Y Q, Chen Q H and Fang X. 2025. DAMAF: dual attention network with multi-level adaptive complementary fusion for medical image segmentation. The Visual Computer, 41(4): 2409-2424 [DOI: 10.1007/s00371-024-03543-8]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Song P F, Yang Z, Li J J and Fan H. 2023. DPCTN: dual path context-aware transformer network for medical image segmentation. Engineering Applications of Artificial Intelligence, 124: #106634 [DOI: 10.1016/j.engappai.2023.106634]
- Su R, Zhang D Y, Liu J H and Cheng C D. 2021. MSU-Net: multi-scale U-Net for 2D medical image segmentation. Frontiers in Genetics, 12: #639930 [DOI: 10.3389/fgene.2021.639930]
- Sun G Q, Pan Y Z, Kong W K, Xu Z C, Ma J H, Racharak T, Nguyen L M and Xin J Y. 2024. DA-TransUNet: integrating spatial and channel dual attention with transformer U-Net for medical image segmentation. Frontiers in Bioengineering and Biotechnology, 12: #1398237 [DOI: 10.3389/fbioe.2024.1398237]
- Sun J D, Zhao J Q, Wu X S, Tang C S, Wang S H and Zhang Y D. 2023. DSGA-Net: deeply separable gated transformer and attention strategy for medical image segmentation network. Journal of King Saud University-Computer and Information Sciences, 35(5): #101553 [DOI: 10.1016/j.jksuci.2023.04.006]
- Wang L F, Mi J, Qin P L, Lin S Z, Gao Y and Liu Y. 2022. Medical image fusion using improved U-Net3+ and cross-modal attention blocks. Journal of Image and Graphics, 27(12): 3622-3636 (王丽芳, 米嘉, 秦品乐, 蔺素珍, 高媛, 刘阳. 2022. 改进U-Net3+与跨模

- 态注意力块的医学图像融合. 中国图象图形学报, 27(12): 3622-3636 [DOI: 10.11834/jig.211066]
- Wu C, Long C, Li S J, Yang J J, Jiang F G and Zhou R. 2022. MSRA-former: multiscale spatial reverse attention network for polyp segmentation. *Computers in Biology and Medicine*, 151: #106274 [DOI: 10.1016/j.combiomed.2022.106274]
- Xiong L L, Yi C, Xiong Q L and Jiang S F. 2024. SEA-NET: medical image segmentation network based on spiral squeeze-and-excitation and attention modules. *BMC Medical Imaging*, 24(1): #17 [DOI: 10.1186/s12880-024-01194-8]
- Xu Q, Ma Z C, He N and Duan W T. 2023a. DCSAU-Net: a deeper and more compact split-attention U-Net for medical image segmentation. *Computers in Biology and Medicine*, 154: #106626 [DOI: 10.1016/j.combiomed.2023.106626]
- Xu S K, Xiao D H, Yuan B H, Liu Y, Wang X Y, Li N, Shi L, Chen J L, Zhang J X, Wang Y H, Cao J F, Shao Y Q and Jiang M J. 2023b. FAFuse: a four-axis fusion framework of CNN and transformer for medical image segmentation. *Computers in Biology and Medicine*, 166: #107567 [DOI: 10.1016/j.combiomed.2023.107567]
- Zhang J B and Xiao Z Y. 2023. Gland and colonoscopy segmentation method combining self-attention and convolutional neural network. *Laser and Optoelectronics Progress*, 60(2): #0217002 (张家宝, 肖志勇. 2023. 结合自注意力与卷积神经网络的腺体及息肉分割方法. 激光与光电子学进展, 60(2): #0217002) [DOI: 10.3788/LOP212696]
- Zhang L, Zhang K J and Pan H W. 2023a. SUNet++: a deep network with channel attention for small-scale object segmentation on 3D medical images. *Tsinghua Science and Technology*, 28(4): 628-638 [DOI: 10.26599/TST.2022.9010023]
- Zhang Y Y, Liu L, Han Z Y, Meng F Y, Zhang Y L and Zhao Y W. 2023b. TranSEFusionNet: deep fusion network for colorectal polyp segmentation. *Biomedical Signal Processing and Control*, 86: #105133 [DOI: 10.1016/j.bspc.2023.105133]
- Zhou X W, Zhu Y, Ge J Y, Pan Q J, Wei R and Gu M. 2024. Dental model segmentation network with fine-grained receptive fields and multiscale fusion. *Journal of Image and Graphics*, 29(12): 3786-3799 (周新文, 朱洋, 葛峻沂, 潘钱家, 魏然, 顾敏. 2024. 具有细粒度感受野的多尺度融合口腔模型分割. 中国图象图形学报, 29(12): 3786-3799) [DOI: 10.11834/jig.230769]
- Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2018. UNet++: a nested U-Net architecture for medical image segmentation//*Proceedings of the 4th International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Granada, Spain: Springer: 3-11 [DOI: 10.1007/978-3-030-00889-5_1]
- Zhu C Z, Zhang R M, Xiao Y L, Zou B J, Chai X, Yang Z Z, Hu R and Duan X C. 2024. DCFNet: an effective dual-branch cross-attention fusion network for medical image segmentation. *Computer Modeling in Engineering and Sciences*, 140(1): 1103-1128 [DOI: 10.32604/cmescs.2024.048453]

作者简介

刘威,男,副教授,主要研究方向为深度神经网络和机器学习。E-mail: liuwei@lntu.edu.cn

钟淼,女,硕士研究生,主要研究方向为深度学习和机器学习。E-mail: 472221405@stu.lntu.edu.cn

刘光伟,男,教授,博士生导师,主要研究方向为露天矿开采设计理论和矿业系统工程。E-mail: liuguangwei@lntu.edu.cn

王浩男,男,硕士研究生,主要研究方向为人机交互和机器学习和自然语言处理的交叉。E-mail: hwang298@jh.edu

宁倩,女,硕士研究生,主要研究方向为机器学习和特征选择。E-mail: 472221387@stu.lntu.edu.cn