

中图法分类号: TP18; TP391 文献标识码: A 文章编号: 1006-8961(2026)07-2587-14

论文引用格式: Huang G H, Liu X, Shi Y Y, Ji Y and Wang S H. 2026. EMPD: edge-guided multiperception decoder for medical image segmentation. Journal of Image and Graphics, 31(7):2587-2600(黄冠珩, 刘翔, 石蕴玉, 纪雨, 王硕鸿. 2026. 面向医学图像分割的边缘引导多重感知解码器. 中国图象图形学报, 31(7):2587-2600)[DOI:10.11834/jig.250251]

面向医学图像分割的边缘引导 多重感知解码器

黄冠珩¹, 刘翔^{1*}, 石蕴玉¹, 纪雨¹, 王硕鸿²

1. 上海工程技术大学电子电气工程学院, 上海 201620; 2. 哈佛大学分子与细胞生物学系, 马萨诸塞州 02138, 美国

摘要: **目的** 医学图像分割在医学影像辅助诊断中具有重要的应用价值, 然而传统的神经网络设计很难满足实际需求。为了解决图像边界分割不准确、特征融合有效性不足的问题, 提出一种 EMPD(edge-guided multiperception decoder) 新型解码器, 通过构建有效的边缘信息引导多重感知特征的解码器, 实现医学图像的准确分割。**方法** 在保持 Pyramid Vision Transformer 全局建模能力的同时, 引入边缘信息感知模块有效提升模型对空间结构的感知能力。该模块在边缘区域表现出更高的响应性, 能够更精确地捕捉结构细节, 从而缓解因多次下采样所导致的边缘信息模糊问题; 通过设计双重尺度级联门控模块及三维注意力融合模块, 能够使得网络以更加全面的视角充分利用编码器的信息, 不仅能进行信息筛选, 还能将编码器与解码器信息有效融合, 进而实现了医学图像的准确分割。**结果** 方法在涵盖 3 类医学图像分割任务的 7 个公开数据集上进行了系统性评估, 在 DICE 指标方面表现出稳定的性能提升。具体而言, 在 ClinicDB(colonoscopy videos challengeclinic database) 数据集上, DICE 分数相较于现有主流方法提升了 1.58%; 在 5 个 Polyp 数据集及 ISIC2017(international skin imaging collaboration) 皮肤病变数据集上, 平均提升幅度达 0.77%。此外, 在 Synapse 多器官分割任务中, 本文方法在 DICE 和 mIoU(mean intersection over union) 指标上分别较同一编码器的 PVT-Cascade 提升了 2.16% 和 3.81%, 进一步验证了其在复杂解剖结构建模方面的有效性与泛化能力。**结论** 本文提出的解码器与主流的方法相比具有更优异的性能, 为医学图像分割研究提供了新的工具, 能够实现有效的医疗辅助诊断。

关键词: 边缘感知; 双重尺度级联门控(DSCG); 三维注意力; 特征融合; 医学图像分割

Edge-guided multiperception decoder for medical image segmentation

Huang Guanhu¹, Liu Xiang^{1*}, Shi Yunyu¹, Ji Yu¹, Wang Shuohong²

1. School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2. Department of Molecular and Cellular Biology, Harvard University, Massachusetts 02138, USA

Abstract: Objective Medical image segmentation refers to the process of accurately separating the regions of interest in medical images from the background to extract key information, such as organs and lesion areas, and provide support for subsequent diagnosis, analysis, evaluation, and treatment. Benefiting from the widespread application of deep learning, especially convolutional neural networks and Transformer frameworks, it has evolved from using traditional manual methods

收稿日期: 2025-06-17; 修回日期: 2025-11-09; 预印本日期: 2025-11-16

* 通信作者: 刘翔 xliu@sues.edu.cn

基金项目: 上海市自然科学基金项目 (19ZR1421500)

Supported by: Shanghai Municipal Natural Science Foundation (19ZR1421500)

to automated and universal intelligent segmentation. Many AI-assisted diagnosis and treatment systems are used in practice. Compared with simple pixel computing, existing network structures can recognize various biological information well by extracting and classifying image features, and automated segmentation has achieved substantial progress. However, challenging issues, such as how to solve the difficulty of edge acquisition, how to fully utilize the obtained features, and how to determine which features among multiple ones play a more important role in segmentation judgment, remain. With the deepening of the network layer, the loss of high-frequency detail information leads to blurred and diffused image edges. Reasonable utilization and accurate fusion of multilevel and multiscale feature information directly affect the accuracy of image segmentation. To address these issues, we introduce three modules and propose a novel edge-guided multiperception decoder for medical image segmentation (EMPD). The three modules are edge information perception module, dual-scale cascade gating, and three-dimensional(3D) attention fusion module. **Method** Using EMPD as a decoder after some high-performance encoders can effectively improve the model's segmentation accuracy. The input image first enters the three-dimensional(3D) attention fusion module, which utilizes the encoder input features to interactively fuse positional and channel features. By rearranging the attention mechanism at the pixel location, features from the position, channel, and pixel dimensions are interactively combined to generate a three-dimensional(3D) attention map. This attention map is then applied to the input to extract features at multiple scales for feature enrichment. These features are interactively combined with encoder information from the previous layer through two-scale cascade gating. In this part, we simultaneously extract and interactively merge information from different scales and layers to highlight primary information and suppress secondary information. An edge-aware module processes the original image and performs corresponding downsampling and edge extraction operations at different scales. During the progressive upsampling process, multiple operators extract edge information from different perspectives. This combination of multiple edge information, coupled with the introduction of learnable variables, makes our operator more informative, more learnable, and less expensive than others. The final learned result is inputted into the next layer of three-dimensional(3D) attention fusion module, and the multilayer predicted image is combined with the MUTATION loss to obtain the final segmentation result. **Result** This study conducted experimental evaluations on seven datasets representing three medical image segmentation tasks. The datasets are ClinicDB, ColonDB, CVC-300, Kvasir, and EITS datasets for single-class polyp segmentation; the ISIC2017 dataset for single-class skin lesion segmentation; and the Synapse dataset for multiorgan segmentation. The proposed method achieved a 1.58% improvement in segmentation performance on the ClinicDB dataset compared with state-of-the-art methods. On the five Polyp and ISIC2017 datasets, the Dice score improved by an average of 0.77% compared with the best-performing network. On the Synapse dataset, the Dice score and mean intersection over union improved by 2.16% and 3.81%, respectively, compared with PVT Cascade. In the segmentation of the left and right kidneys, the Dice score improved by 2.92% and 2.50%, respectively. In the large stomach region, the proposed method achieved a 3.28% improvement compared with PVT Cascade. Ablation experiments conducted on this architecture demonstrated that on multiple datasets, the Dice coefficient improved by more than 3 percentage points compared with a simple encoder cascade architecture. Introducing edge information and combining it with an effective fusion strategy substantially improved the problem of poor edge fitting and enhanced the overall segmentation accuracy, proving the effectiveness of the proposed module. However, having abundant edge information is not always good. The number and distribution of edge-aware modules need to be controlled. If the image depth is too large, a small number of learnable parameters may be insufficient to compensate for the conflict between semantic and edge information, potentially leading to a decrease in accuracy. **Conclusion** EMPD, combined with a high-performance encoder, achieved optimal performance on the five datasets, and its results approached the current best results on the two remaining datasets. Its performance surpassed that of mainstream image segmentation methods, and it demonstrated good generalization properties. This work provides new tools and model frameworks for medical image segmentation research, thus helping achieve effective medical-assisted diagnosis.

Key words: edge perception; dual-scale cascaded gating(DSCG); three-dimensional attention; feature fusion; medical image segmentation

0 引言

医学图像分割作为计算机辅助诊断系统中的核心任务之一,在现代医疗中扮演着至关重要的角色。随着医学成像技术的飞速发展,临床中产生了大量复杂的图像数据(Cai等,2020),这些图像为疾病检测、术前规划和治疗评估提供了丰富的信息。然而,这些图像通常具有组织结构复杂、边界模糊以及对比度低等特点(Wang等,2022b),极大地增加了医生人工判读的难度与主观性(Salahuddin等,2022)。传统的手工分割准确率基本可以保证,但对操作者的专业性技能要求较高,往往需要多年的专业培养,通常情况下,一份CT(computed tomography)包含上百幅图像切片,依靠人工查找对医疗人力资源消耗大(李一宸等,2025)。在许多医疗场景下,分割结果直接影响到疾病的诊断准确性与治疗策略的制定(Rahman和Marculescu,2023a)。随着计算机和自动分割技术的发展,衍生出诸多医学图像分割模型,缓解了上述情况的发生,然而依然存在如下的问题。首先,医学图像中的组织边缘不明显,存在过渡模糊的情况(Lee等,2020),在获取图像特征的过程中,随着网络的深度、感受野的增大,高频信息会随网络进一步地压缩,从而导致边界定位误差加大(Marmanis等,2018),在上采样的过程中无法还原出原始的精细边界。其次,传统的跳跃连接在编码器浅层特征与解码器深层特征融合时,存在特征感知弱和语义融合不充分的情况(Wang等,2022a),并且,网络的多尺度特征提取多基于上采样信息而忽视了编码器多尺度纹理信息,不能对编码器信息充分利用。此外,受限于医学图像的固有限制,难以获取充足且高效的标注数据,数据量的不足也加剧了模型准确分割的困难程度(Hesamian等,2019)。

综上所述,医学图像分割仍存在以下难点:1)医学图像固有的边缘特性及下采样的层层压缩使得边界信息难以捕捉;2)不同分割场景下,病灶及器官尺度变化大,结构信息复杂,不同尺寸目标的感知能力弱,有效的图像特征获取困难;3)不同层次获取的特征语义融合程度低,特征信息难以充分利用,导致分割精度不足。

本文的主要贡献如下:1)提出了一种面向医学图像分割的边缘引导多重感知解码器(edge-guided

multiperception decoder, EMPD),将边缘卷积信息表示与上采样信息相结合,减少了边界弥漫情况的产生,提升了分割精度;2)提出一种双重尺度级联门控机制,通过引入多尺度分支增强目标感知能力,并将不同层级的编码器特征经上采样级联融合,结合门控机制实现动态加权,有效提升了模型对细粒度结构与多尺度目标的识别能力;3)设计了一种三维注意力融合模块,结合通道交互与空间交互分支提取全局语义依赖与空间位置信息,对信息分支获取像素粒度级别的加权混合特征,提升了关键区域的响应能力与整体融合表达的判别性。

1 相关工作

1.1 编码器—解码器框架

早期的解码器通常采用简单的卷积层和下采样层的堆叠结构(Badrinarayanan等,2017),这类设计能有效拓展感受野并提取深层特征,然而,重复的堆叠在获取更加广阔感受野的同时也会导致深层网络训练困难的情况(Simonyan和Zisserman,2014),残差神经网络(residual network, ResNet)(He等,2016)引入了残差连接,通过学习输入与输出之间的残差,梯度能够更顺畅地反向传播,使得更深层的网络结构成为可能。Huang等人(2017)通过密集连接实现了高效的特征复用与梯度传导,为构建深层高效解码器提供了理论支撑。为了弥补卷积特征中局部依赖强的问题,部分现代解码器设计开始引入注意力机制,如squeeze and excitation(Hu等,2018)、convolutional block attention module(Woo等,2018)、efficient channel attention(Wang等,2020),通过动态加权的方式实现对通道或空间位置的自适应调整,从而提高网络的表示能力与信息捕捉能力。与此同时,受到自然语言处理中的Transformer架构的启发,视觉方面,Vision Transformer(Dosovitskiy等,2021)与Swin Transformer(Liu等,2021)逐步引入图像分割领域,通过自注意力机制(self-attention)进行全局上下文建模,弥补了卷积神经网络(convolutional neural network, CNN)模型在捕捉图像全局结构关系上的不足(Vaswani等,2017),尤其在长程依赖建模和复杂场景的理解上展现出巨大的优势。为了解决纯卷积网络和纯Transformer网络各自的局限,越来越多的研究开始提出混合模型,结合CNN和Trans-

former 的优势,充分利用卷积网络的局部特征提取能力和 Transformer 的全局建模能力(Zhang 等, 2021)。这类混合网络在编码器中通常先通过卷积层提取局部特征,再通过 Transformer 进行全局信息的捕捉。视觉 Transformer-U 形网络(vision Transformer-based UNet, ViT-UNet)(Zhou 等, 2024)将传统的 U-Net 结构与 Vision Transformer 融合,在网络的编码器部分使用 Transformer 提取全局特征,在解码器部分使用卷积层进行局部信息恢复。Jha 等人(2024)使用基于金字塔视觉 Transformer 的改进基线模型(improved baselines with pyramid vision Transformer, PVTv2)(Wang 等, 2022c)作为编码器,通过层次化解码结构结合残差与全局注意力机制,有效提升医学图像分割中的全局上下文建模与边界的定位能力。引入 Transformer 的 UNet 网络架构(Transformers make strong encoders for medical image segmentation, TransUNet)(Chen 等, 2021)在有限的医学图像数据上通过预训练—微调策略,克服小数据集性能有限的问题。虽然这些方法能提升模型的训练效率和泛化能力,但高效提取跨层级网络的多尺度特征仍是模型适应性和分割准确性的关键(章东平等, 2025)。在本文中,旨在提出一种新的面向医学图像分割的边缘引导多重感知解码器来进一步细化注意力权重并增强融合特征信息。

1.2 边缘信息

边缘信息可以帮助模型精确地捕捉到这些边界,避免因下采样和特征丢失导致的边缘模糊。早期的边缘信息获取通过边缘检测算子进行分割(Boykov 和 Jolly, 2001),传统的 Sobel 算子通过一阶微分运算能够快速定位医学影像中的结构边缘,Canny 算子(Canny, 1986)通过高斯滤波等操作进一步增强了算子的稳定性。U-Net(Ronneberger 等, 2015)通过其编码器—解码器结构中的跳跃连接有效地保留了边缘信息,从而提高了边界的准确性。Alshawh 等人(2023)结合 U-Net,将边缘检测算子注入网络,拓展了特征空间,孙国栋等人(2022)通过噪声抑制与结构增强的预处理方式提升了图像质量,为后续的临床分析提供了更可靠的输入基础。近年来,研究者在图像分割模型中引入了边缘感知机制,这种机制通过专门设计的边缘引导网络、边缘增强模块等进一步加强边缘区域的特征表达。杨可欣等人(2025)根据特征内容本身为每个像素动态生成卷

积核以增强细小纹理,边缘引导与层级聚合网络(edge-guided and hierarchical aggregation network, EHANet)(Tang 等, 2025)对层间不同尺度的特征进行传统卷积,并使用 softmax 函数预测边界图以增强网络分割效果,边缘引导卷积如中心差分卷积(Yu 等, 2020),不同于传统卷积,通过计算中心差分等方法增强对局部细节特征的捕捉,帮助网络聚焦于图像中的重要边缘区域,提升了图像的分割精度。多层边缘注意力网络的医学网络(multilayer edge attention network for medical image segmentation, MEA-Net)(Liu 等, 2022)通过多层边缘注意力机制强化编码阶段的边界特征提取。模型无关边界优化分割方法(model-agnostic boundary refinement for segmentation, SegFix)(Yuan 等, 2020)是一种模型无关的后处理方法,通过学习边界像素的方向信息并进行偏移校正,有效提升了分割结果边界的连贯性与精度。息肉多尺度边缘引导分割网络(multi-scale edge-guided attention network for weak boundary polyp segmentation, MEGANet)(Bui 等, 2024)设计了一种边缘注意力模块,使用拉普拉斯算子与下采样特征相结合,解决了图像的弱边界问题。这些方法虽然在一定程度上提升了边界表达能力,但大多存在边缘响应算子固定或学习能力弱、忽视多层次语义的问题。本文设计了一种基于差分卷积的边缘信息感知模块,通过多方向的差分卷积联合提取边缘响应特征,同时作用于编码器的多个阶段,实现端到端训练下的边界增强与目标精细刻画,有效提升了弱边界区域的分割精度与模型的结构感知能力。

2 方 法

2.1 整体框架

以 Transformer 为框架的模型在获取全局特征时无法有效地关注边界信息,模型针对图像进行准确分割的过程中会存在分割中断或图像弥漫的情况,无法对图像特征信息充分融合。为了解决这个问题,本文提出了一种新的面向医学图像分割的边缘引导多重感知解码器。

图 1 展示了 EMPD 结合 PVTv2 编码器的体系框架。如图 1 所示,位于图像左侧的编码器包含 4 个视觉金字塔模块,从输入的医学图像中捕获不同尺度

的特征组,对于编码器输入,使用三维注意力融合模块丰富编码器特征,实现特征加权,上采样信息与下一层编码器信息的融合在实现级联特征融合的同时,分别送入3个双重尺度门控机制,融合全局与局部信息,选择性增强特征,抑制无效信息。在此之后对应尺度的显式边缘特征与输出图像特征进行拼接,获取多重图像特征及卷积所丢失的边缘信息,再次经过三维注意力融合模块进行有效的信息融合,将每一层的输出汇总,结合每幅分割图以得到最终的分割结果。

2.2 编码器主干

在医学图像任务中,编码器作为特征提取的基石,传统框架难以兼顾局部细节感知与全局语义的建模,尤其在器官间结构复杂,不同组织分布分散的医学图像中表现受限。本文采用PVTv2作为编码器,通过金字塔结构逐步提取多尺度特征,兼具局部空间建模能力与长距离依赖建模能力。PVTv2包含4个阶段,每一阶段均会产生不同尺度的特征图,本文充分利用这种金字塔式的多尺度特征,按照不同层级将不同尺度的特征作为EMPD的输入信息。

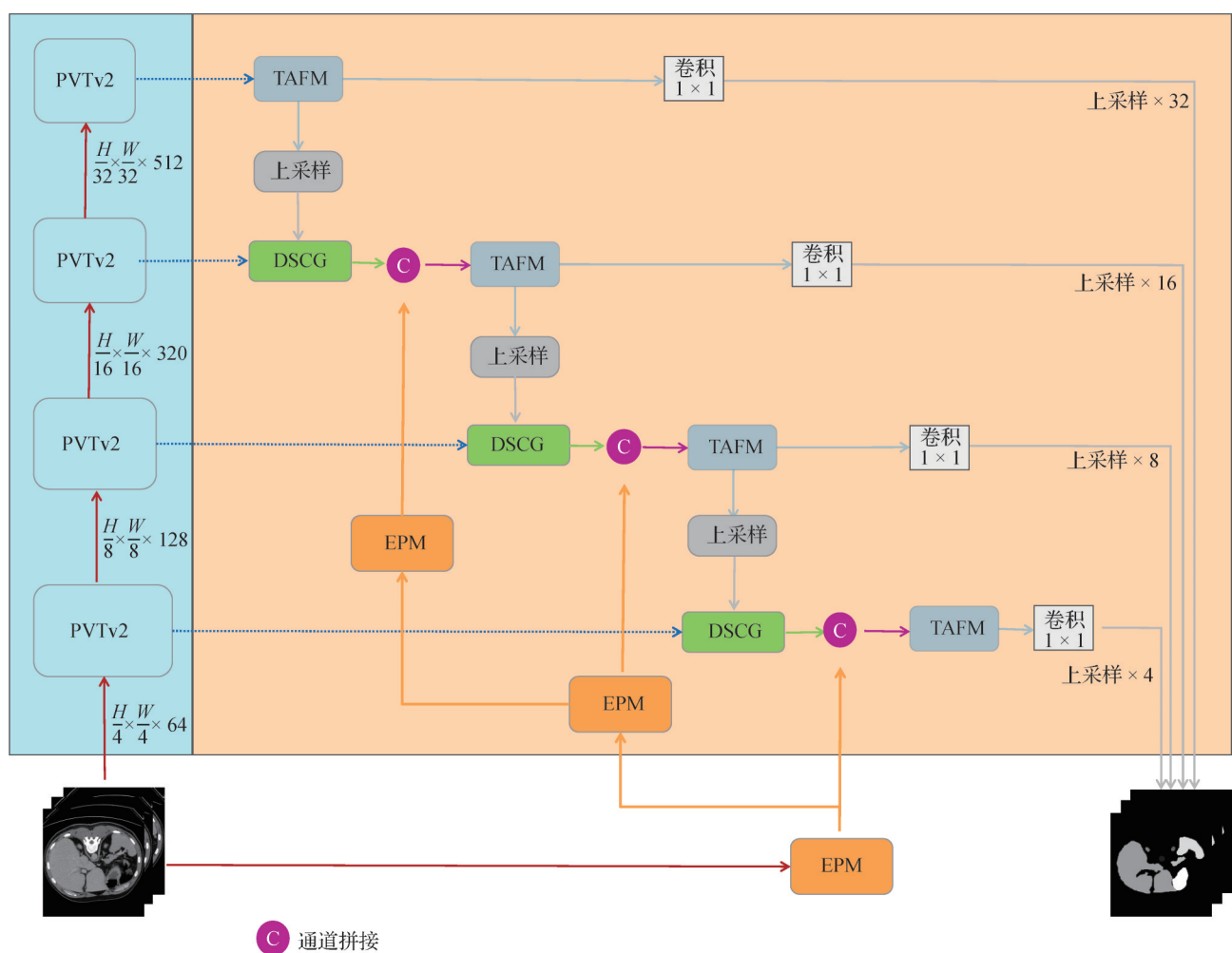


图1 整体框架图

Fig. 1 Overall framework diagram

2.3 边缘信息感知模块(EPM)

Prewitt算子采用两个固定的卷积核分别模拟水平和垂直两方向的梯度,进而提取边缘。Sobel算子在其工作的基础上,对中间行赋予了更高的权重,有效地减少了来自上下行的噪声影响,起到了竖向平滑的效果,Canny算子在此基础上加入高斯降噪等

步骤使得效果进一步提升,但仍然存在自定义阈值影响大、弱边缘易丢失的情况。差分卷积在边缘信息感知模块(edge information perception modules, EPMs)中是通过多种差分卷积结合普通卷积来实现的。以往的工作(Su等,2021)验证了相较于普通卷积特征提取和泛化能力均有提升,近来一些工作

(Chen 等, 2024)将其运用到单幅图像去雾方向。边缘信息感知模块结构如图2所示。

不同的计算方式产生了多种差分卷积,以中心差分卷积(central difference convolution, CDC)为例,通过使卷积核加权为0,当区域变化不大时,计算结果会趋近于0,从而起到对边缘更加敏感的效果。此外还存在角度差分卷积(angular difference convolution, ADC),垂直差分卷积(vertical difference convolution, VDC),水平差分卷积(horizontal difference convolution, HDC)等计算方式。不同的计算方式产生了多种的差分卷积。本文引入边缘感知模块以融合多种差分卷积信息,在丰富边缘信息的同时加入普通卷积,引入了可学习参数,在训练中自动优化,

减少了自定义参数的影响。由于编码器各层之间存在尺度变换不一致的问题,传统的固定下采样策略难以实现有效对齐与信息融合。本文设计了分步卷积的分步变形模块,通过保持每次下采样步长为2,实现不同尺度特征的对齐与融合,有效整合多层卷积信息,提升特征表达能力。

$$F_{\text{EMP}}(x) = F_3 \left\langle \sum_{i=1}^5 F_3(x) \times ki_3 \right\rangle_n \quad (1)$$

式中, $F_{\text{EMP}}(\cdot)$ 代表 EMP 模块的输出, $F_3 \langle \cdot \rangle_n$ 代表 n 次步长为2、卷积核大小为3的下采样卷积, ki_3 在 i 从1到5的过程中分别代表了5种卷积的卷积参数。 $F_3(\cdot)$ 为融合后的单个卷积核大小为3的卷积操作。

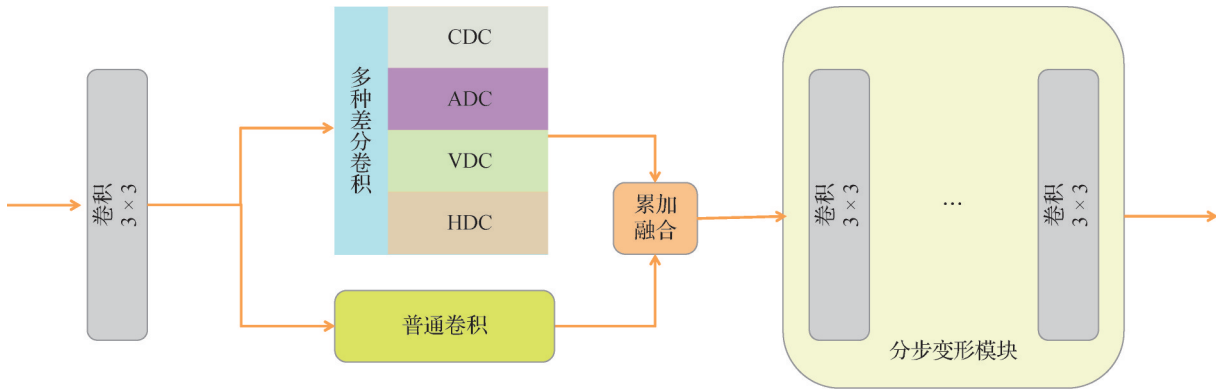


图2 边缘信息感知模块结构图

Fig. 2 Structural diagram of edge information perception module

2.4 双重尺度级联门控(DSCG)

编码器在不同尺度下生成4个特征组,信息融合以获取更加丰富的语义信息可以帮助图像分割精度提升,这也是以UNet为代表的U型网络被广泛作为图像分割基石的原因之一。但是简单的上采样相加会引入冗余的特征信息并导致语义信息不匹配的情况。为了解决这个问题,设计了双重尺度级联门控(dual scale cascaded gate control module, DSCG)处理不同尺度信息融合的情况。图3显示了模型的具体细节。来自不同尺度编码器的特征作为输入分别进入两个支路,每个支路先经过卷积和归一化操作以统一维度与尺度。其次,在模块内部的双分支结构中,显式地提取了具有不同感受野的上下文信息,将两条路径提取的特征在通道维度进行拼接,最终生成门控权重图。该权重图用于动态调整初始输入特征的通道响应,实现对有效信息的强调与无关信息的抑制,进而实现具有多尺度感受野、融合上下文

语义并具备选择性的门控机制。级联门控 DSCG 具体表示为

$$F_u = F_3(BN(F_1(U))) + U \quad (2)$$

$$F_l = F_3(BN(F_1(I))) + F_5(BN(F_1(I))) + I \quad (3)$$

$$Att = \sigma(BN(F_1(RE([F_u, F_l]))) \quad (4)$$

$$DSCG(U, I) = I \times Att + U \quad (5)$$

式中, $[\cdot]$ 代表在通道维度进行拼接, $F_x(\cdot)$ 代表采样 x 大小的卷积核进行卷积, $BN(\cdot)$ 和 $RE(\cdot)$ 分别代表批量规范化和 ReLU(rectified linear unit)激活层, Att 代表注意力图, U 为高层特征, I 为编码器对应层特征, σ 为 sigmoid 函数。

2.5 三维注意力融合模块(TAFM)

仅具备丰富的特征信息并不足以充分发挥其判别能力。针对医学图像中结构复杂、边界模糊等挑战,仍需进一步挖掘特征之间的深层联系与协同作用,以实现更精准的区域识别与边界定位。为此,设

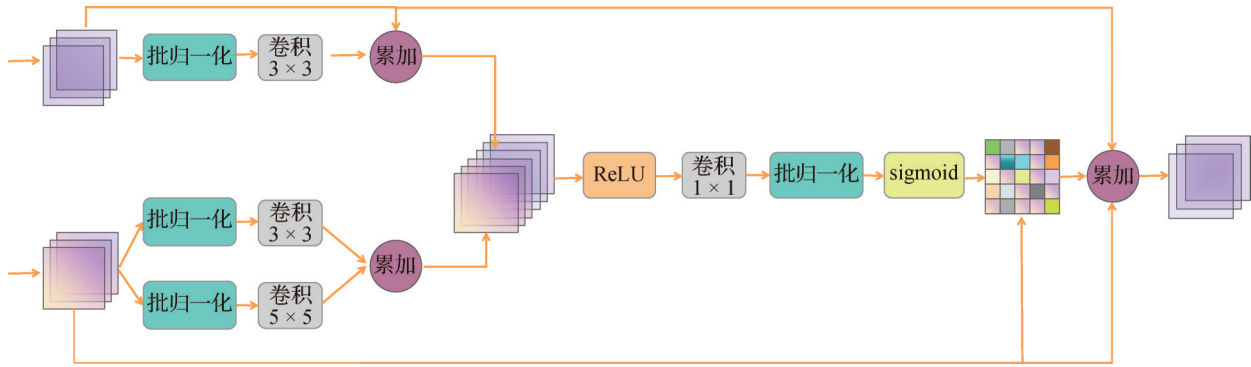


图3 双重尺度级联门控模块结构图

Fig. 3 Structural diagram of dual scale cascaded gate control module

计了一种三维注意力融合模块(three-dimensional attention fusion module, TAFM),从通道、空间位置与像素粒度3个维度动态建模特征依赖关系,增强模型的表达能力与细粒度感知能力。TAFM由通道信息模块(channel information block, CIB)、空间位置信息模块(spatial information block, SIB)及多尺度倒置瓶颈模块(multiscale inverted bottleneck module, MIM)组成。设计与空间和通道双重注意力的Transformer UNet分割网络(dual attention with Transformer U-Net, DA-TransUNet)(Sun 等, 2024)类似, TAFM框架如图4所示。

从图4可以看出,本文采用并行结构对输入进

行信息提取,信息分支通过特征升维与拼接整合信息,利用维度重排与卷积获取像素粒度级别的加权混合特征,扩大感受野并提取融合特征信息注意力图,最后通过多尺度倒置瓶颈模块进一步整合,实现全局与局部信息的有效融合,提升特征表达的精细程度与模型在复杂场景下的稳定表达。

$$Att_{inf}(X) = \sigma \left(F_7 \left(\Phi_{rear} \left[r_{\uparrow 2}(CIB + SIB), r_{\uparrow 2}(X) \right] \right) \right) \quad (6)$$

$$TAFM(X) = MIM(X + Att_{inf}(X) \times X) \quad (7)$$

式中, $[\cdot]$ 代表进行维度拼接, $r_{\uparrow 2}$ 代表在第2维度新增一个维度, Φ_{rear} 代表分组重排操作, $F_7(\cdot)$ 代表

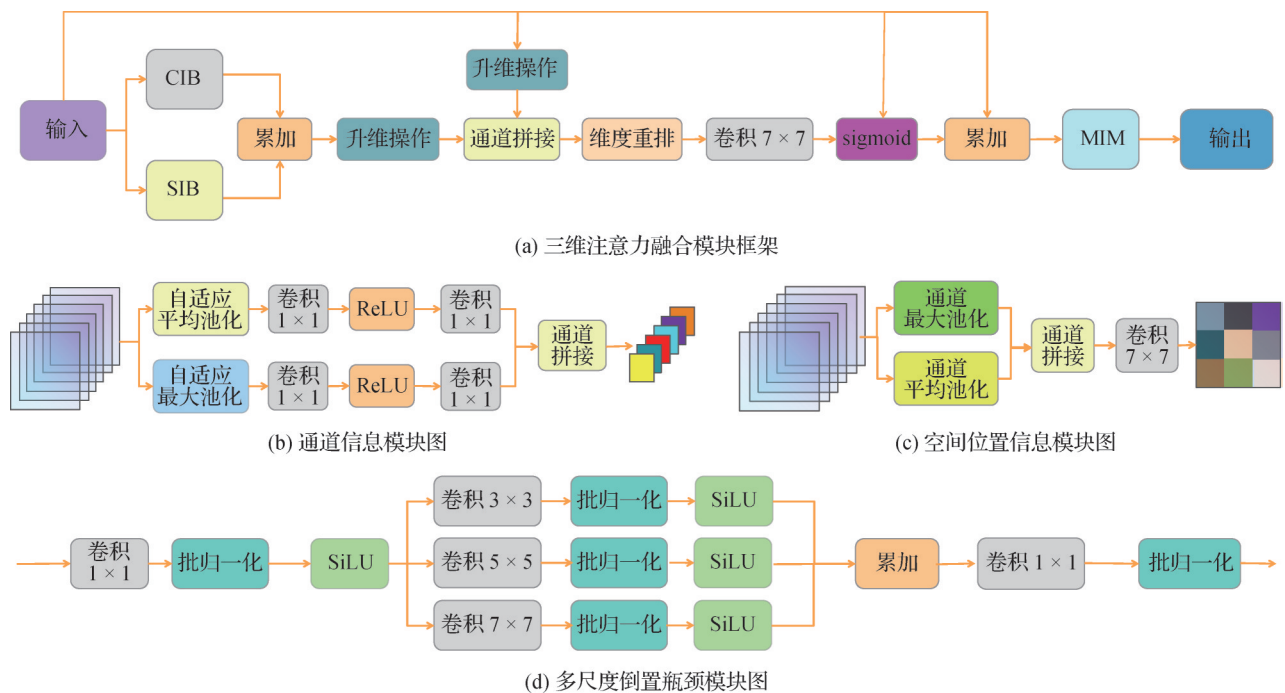


图4 三维注意力融合模块框架图

Fig. 4 Three-dimensional attention fusion module framework diagram ((a) three-dimensional attention fusion module framework; (b) channel information block; (c) spatial information block; (d) multiscale inverted bottleneck module)

7×7 卷积, Att_{inf} 代表信息注意力图, $MIM(\cdot)$ 为多尺度倒置瓶颈模块。

2.5.1 通道信息模块(CIB)

通道信息以结构化方式对各通道特征进行全局建模,从而感知不同通道间的依赖关系,并对关键通道进行响应增强,以提升特征表示的选择性与判别能力。CIB 计算为

$$f(X) = F_1(RE(F_1(X))) \quad (8)$$

$$CIB(X) = f(AAP(X)) + f(AMP(X)) \quad (9)$$

式中, $F_1(\cdot)$ 代表通道映射卷积, 两个 $F_1(\cdot)$ 卷积参考瓶颈模块思想, 先将通道变成原始的 $1/16$ 后升维恢复, 以减少计算量, $AAP(\cdot)$ 和 $AMP(\cdot)$ 代表在空间特征图中自适应最大池化和在空间特征图中自适应平均池化。

2.5.2 空间位置信息模块(SIB)

空间位置信息模块通过生成响应增强图, 对特征图各位置之间的空间依赖关系进行建模, 从而显式编码空间位置信息, 强化目标区域的表达能力, 并提升模型对结构分布的整体感知。SIB 计算为

$$SIB(X) = F_1[CM(X), CA(X)] \quad (10)$$

式中, $CM(X)$ 是对每个空间位置上的全部通道取平均值, $CA(X)$ 是对每个空间位置上的全部通道取最大值。

2.5.3 多尺度倒置瓶颈模块(MIM)

MIM 采用残差连接以保持三维特征结构的连续性, 同时提升模型在梯度传播时的稳定性。类似于移动神经网络架构(inverted residuals and linear bottlenecks, MobileNetV2)(Sandler 等, 2018), 采用倒置瓶颈结构先升维再降维, 在增强信息流动能力的同时, 使用多尺度卷积提升特征表达的多样性, 增强模型的表达能力。MIM 计算为

$$f_i(X) = SL(BN(F_{2i+1}(X))) \quad (11)$$

$$MIM(X) = BN\left(F_1\left(\sum_{i=1}^3 f_i(f_0(X))\right)\right) + X \quad (12)$$

式中, $F_x(\cdot)$ 代表采样 x 大小的卷积核进行卷积, $SL(\cdot)$ 代表 SiLU (sigmoid-weighted linear unit) 激活层。

2.6 损失

传统损失如 Cross Entropy Loss (Krizhevsky 等, 2017), Dice Loss (Milletari 等, 2016) 在多尺度分割中

难以协调各输出分支的优化, 容易导致训练冲突与泛化不足, MUTATION (multi-stage feature-mixing loss aggregation) 损失 (Rahman 和 Marculescu, 2023b) 组合多个阶段输出生成丰富且互补的监督信号, 通过组合预测的方式将模型各阶段输出的预测图进行子集组合, 四阶段输出除空集外共有 15 种组合方式, 对每一种组合结果单独计算损失, 并将所有组合损失进行聚合, 构建隐式集成机制, 从而实现稳定协同训练与性能提升。

$$I_s = \sum_{i \in s} p_i, \quad i \in \{1, 2, 3, 4\} \quad (13)$$

$$L_{total} = \sum_{s \in G} (\omega_{ce} \times L_{CE}(I_s, y) + \omega_{dl} \times L_{DI}(I_s, y)) \quad (14)$$

式中, G 为除空集以外的全部子集序列, p_i 为第 i 层预测图, I_s 为 s 对应子集预测图之和, $L_{CE}(\cdot)$ 为交叉熵损失函数, $L_{DI}(\cdot)$ 为 Dice 损失函数, ω_{ce} 和 ω_{dl} 分别为交叉熵与 Dice 的权重系数。

3 实验结果与分析

3.1 数据集及评价指标

3.1.1 Polyp 数据集

ClinicDB (colonoscopy videos challenge clinic database) 数据集包含 612 幅图像, 其中, 550 幅作为训练集。Kvasir 数据集为 Kvasir-SEG (Kvasir segmentation dataset) 息肉数据集中的 1 000 幅息肉图像, 其中, 900 幅作为网络的训练集。遵循 CASCADE (cascaded attention decoder) 中数据集格式, 将两个数据集中的训练集图像组合作为模型训练集, 同时将剩余的 62 幅 ClinicDB 图像与 100 幅 Kvasir 图像作为各自数据集的测试集。模型同时应用于 CVC-300 (colonoscopy videos challenge)、ColonDB (colonoscopy videos challenge colon database) 以及 ETIS (ETIS larib polyp database) 数据集, 分别包含 60 幅、380 幅以及 196 幅图像。

3.1.2 Synapse 数据集

Synapse 数据集包含 30 例增强腹部 CT 扫描, 每例由 512×512 像素的切片组成。按照 6:4 的比例随机划分, 遵循 TransUNet 数据集格式, 其中 18 例用于训练、12 例用于验证。选取左肾、右肾、主动脉、脾脏、胆囊、肝脏、胃和胰腺共 8 个器官, 用于多器官分割任务。

3.1.3 ISIC2017 数据集

ISIC2017 (international skin imaging collaboration)数据集由国际皮肤影像挑战赛提供,其中共计 2 000 幅图像作为训练集,150 幅图像构成验证集,600 幅图像构成测试集。

3.2 评估指标

在 Polyp 及 ISIC2017 数据集中采用 DICE 得分作为评估标准,在 Synapse 数据集中添加平均交并比 (mean intersection over union, mIoU) 指标共同评估。

3.3 实验设置

本实验基于 Pytorch 框架实现,模型在单个具有 24 GB 显存的 NVIDIA-GeForce-RTX-4090 显卡上进行训练。PVTv2 网络在 ImageNet 数据集进行预训练并作为本模型的编码器部分,训练采用 AdamW 优化器,学习率与权重衰减为 0.000 1,在 Synapse 数据集训练 300 个训练周期,批次大小为 6,其余数据集训练 200 个训练周期,批次大小为 12。本文在 Synapse 数据集上采用 90° 倍数旋转加随机反转后再进行 $[-20^{\circ}, 20^{\circ}]$ 随机小角度旋转,其余数据集未进行数据增强。Polyp 数据集输入转化为 352×352 像素图像大小,使用 $[0.75, 1, 1.25]$ 3 个尺度进行训练。

3.4 数据集实验结果分析

本文在多个医学图像分割数据集上对所提出的 EMPD 解码器进行了全面性能评估,涵盖消化道息肉与皮肤图像的二分类任务 (Polyp 和 ISIC2017) 以及多器官分割的多分类任务 (Synapse)。3 种数据集代表了临床应用中常见的两类任务场景:结构边界不清晰的小目标检测与多类复杂器官分割。

表 1 展示了与当前基于 CNN、Transformer 以及二者融合结构的主流图像分割网络在 Polyp 和

ISIC2017 数据集上的对比结果。可以看出,EMPD 解码器在 4 个数据集上均取得了最高的 DICE 分数,在另外两个数据集上仅次于最优表现。以 ClinicDB 为例,相较于同样采用 PVT 编码器的 PVT-Cascade, DICE 分数提升了 1.58%,在 6 个数据集上平均提升 0.77%。这表明 EMPD 解码器具有良好的泛化性能,能够适应不同器官下的二分类分割任务并稳定提升分割精度。

在多类别的 Synapse 数据集上,进一步评估了模型在复杂器官分割任务中的表现。表 2 给出了各方法的详细对比结果。医学分割大模型如 MedSAM (segment anything in medical images) (Ma 等, 2024) 在输入时需要额外信息指定输入位置,相较于其他模型可以在确定的小范围内进行分割避免误分类的情况,因此在不规则且经常间断性出现的胰腺分割存在优势,与 PVT-Cascade 相比,本文提出的模型在 DICE 和 mIoU 上分别提升了 2.16% 和 3.81%。就具体器官而言,在小器官如左肾与右肾的分割上 DICE 分数分别提升 2.92% 和 2.50%,在胃部较大器官区域提升 3.28%。这主要得益于解码器中引入的多尺度特征整合与边缘感知机制,增强了对小目标与模糊边界结构的识别能力,提升了整体分割精度。

各方法在 Polyp 与 ISIC2017、Synapse 数据集的可视化结果如图 5 和图 6 所示。

3.5 EMPD 不同组成部分的有效性

本文在 3 类医学图像分割任务的 7 个公开数据集上设计并开展了多组消融实验,旨在系统评估所提出模型各个关键模块对整体性能的影响与泛化能力。首先以采用 ImageNet 预训练的 PVTv2 编码器为基础模型,结合级联解码结构构建初始模型框架作

表 1 息肉分割数据集及 ISIC2017 皮肤病变分割数据集结果
Table 1 Results of polyp segmentation dataset and ISIC2017 skin lesion segmentation dataset

方法	ClinicDB	ColonDB	Kvasir	EITS	CVC-300	ISIC2017	均值
U-Net	71.37	53.31	81.03	40.86	71.38	82.60	66.76
AttnUNet	72.53	54.48	81.46	40.49	74.24	82.19	67.56
TransUNet	86.88	64.90	84.99	54.86	83.62	82.66	76.32
TransCascade	91.66	78.02	90.64	72.13	87.59	84.28	84.05
PVT-Cascade	<u>92.87</u>	<u>81.24</u>	<u>91.96</u>	76.79	<u>89.25</u>	82.63	<u>85.79</u>
PVT-EMPD	94.45	81.98	92.83	<u>75.96</u>	89.94	<u>84.19</u>	86.56

注:加粗、下划线字体分别表示各列最优、次优结果。

表 2 Synapse数据集多器官分割结果
Table 2 Multi-organ segmentation results on the Synapse dataset

方法	DICE	mIoU	主动脉	胆囊	左肾	右肾	肝脏	胰腺	脾脏	胃
MedSAM	59.41	52.16	71.80	32.96	71.24	68.57	44.00	73.36	44.00	69.34
U-Net	77.72	68.29	86.84	62.94	81.50	75.27	94.13	59.11	86.58	75.39
AttnUNet	78.27	68.98	88.20	61.67	81.76	78.38	93.07	62.69	86.44	73.97
TransUNet	76.17	65.59	87.69	63.69	75.84	72.37	92.73	54.01	85.84	77.21
PVT-Cascade	81.98	71.28	84.15	<u>69.43</u>	85.07	82.13	94.89	68.68	90.23	<u>81.24</u>
TransCascade	<u>82.74</u>	<u>73.83</u>	87.73	69.04	<u>86.70</u>	<u>82.71</u>	<u>95.08</u>	69.52	<u>91.15</u>	79.99
PVT-EMPD	84.14	75.09	<u>88.18</u>	70.17	87.99	84.63	95.59	<u>70.84</u>	91.22	84.52

注：加粗、下划线字体分别表示各列最优、次优结果。

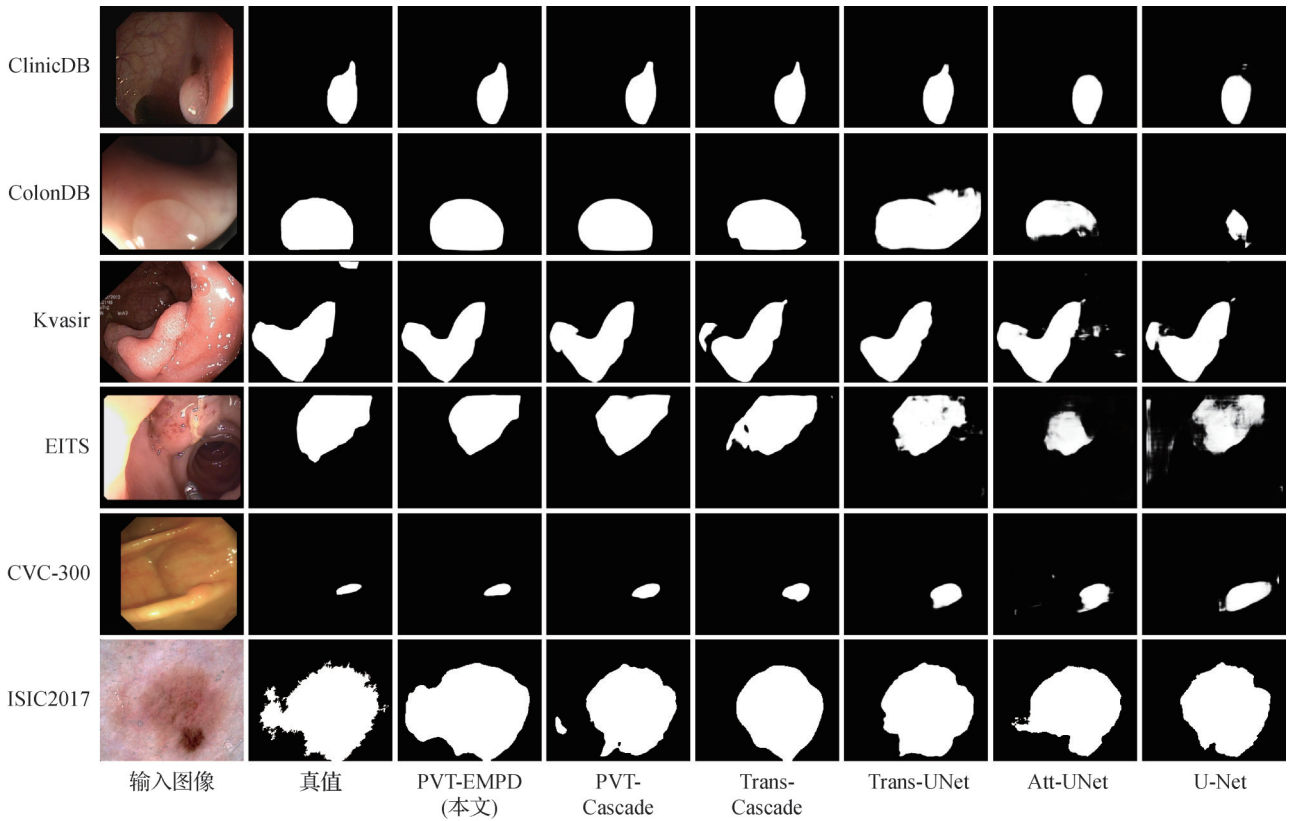


图 5 Polyp 与 ISIC2017 数据集可视化结果
Fig. 5 Visualization results on the Polyp and ISIC2017 datasets

为对比基线。在此基础上,逐步引入本文提出的三大核心模块:三维注意力融合模块、边缘增强模块以及双重尺度注意力模块,并分别评估其独立和联合使用时对分割性能的贡献。实验结果如表 3 所示,三维注意力融合模块显著提升了模型对关键语义区域的响应能力。该模块通过在通道、空间及像素层面上建模多维注意力,有效引导网络聚焦于器官边

界及关键结构区域,从而提升了分割精度。进一步地,加入边缘增强模块后,模型对器官边界的识别能力得到了进一步加强。该模块融合低级特征中的边缘线索,有助于缓解高层特征中的语义模糊问题,从而在提高边界清晰度的同时,保持了全局结构的一致性。双重尺度注意力模块通过在不同尺度上对输入特征进行细粒度建模与筛选,使得网络能够在保

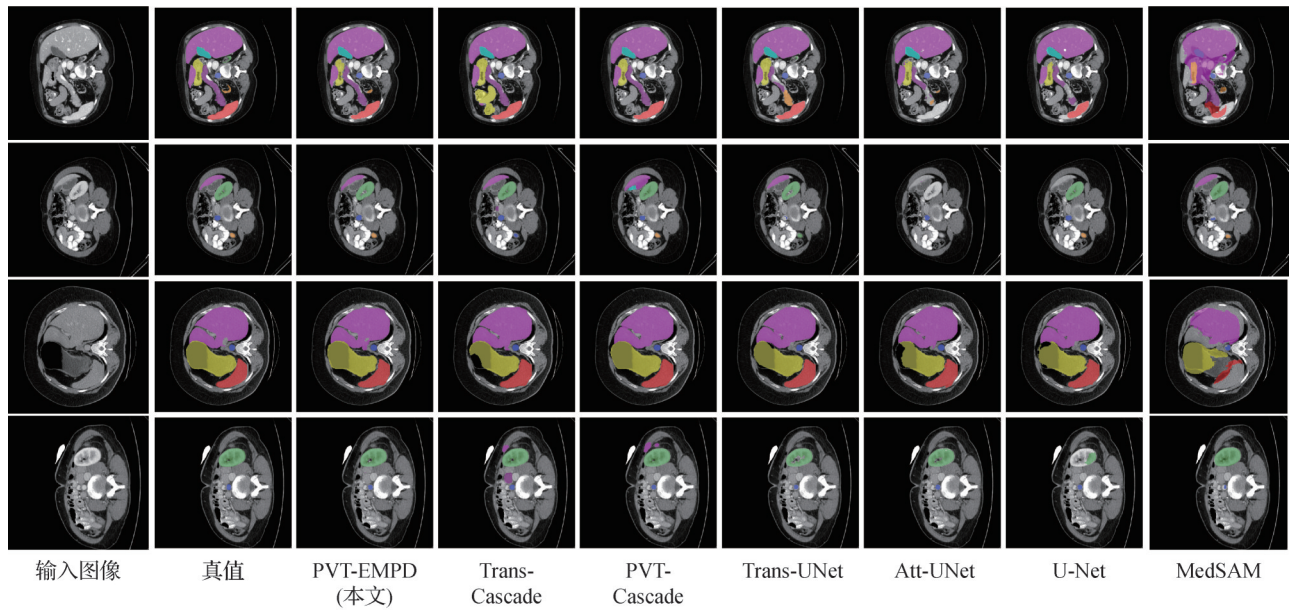


图 6 Synapse数据集可视化结果
Fig. 6 Visualization results on the Synapse dataset

表 3 以PVTv2为骨架的组件消融实验
Table 3 Component ablation experiments using PVTv2 as the skeleton

TAFM	EPM	DSCG	DICE/%						
			Synapse	ClinicDB	ColonDB	Kvasir	EITS	CVC-300	ISIC2017
×	×	×	81.10	92.19	78.34	91.09	72.21	88.77	82.66
√	×	×	81.94	92.86	78.37	91.78	72.55	89.15	82.92
√	×	√	82.16	93.60	79.24	91.90	72.92	89.85	83.15
√	√	×	83.22	93.63	81.09	90.98	74.74	89.77	83.62
√	√	√	84.14	94.45	81.98	92.83	75.96	89.94	84.19

注：加粗字体表示各列最优结果。“×”为未使用，“√”为使用。

留丰富上下文信息的同时,抑制冗余或干扰区域的信息流入,为后续的特征融合与判别提供了更加稳健的表达支持。最终,当三大模块协同工作时,模型在评价指标上均取得了最优表现,验证了各模块之间的互补性与协同增强效果,充分说明了本文方法在提升医学图像分割性能方面的有效性与实用性。

3.6 EPM块位置影响

边缘信息感知模块已经证明其在解码器上的有效性。不同位置结合EPM模块的实验结果如表4所示。在图像的中上层,融合边缘信息带来了明显的收益,但在最深层,由于特征图的高度抽象,主要承载语义信息,此时,少量的参数增加不足以弥补过多边缘特征引入导致的信息冲突,可能会出现准确率略有下降的情况。尽管如此,其性能仍然优于仅在

表 4 不同位置结合EPM模块的实验结果
Table 4 Experimental results combining EPM modules at different locations

位置	参数量/M	DICE/%
[1]	27.19	82.87
[1, 2]	27.99	84.14
[1, 2, 3]	33.68	83.26

注：加粗字体表示最优结果。

浅层单一阶段引入边缘信息的场景。

4 结 论

为提升医学图像分割精度,本文提出一种基于

边缘信息引导的多重感知解码器,该解码器能有效地将边缘信息与多尺度多阶段的图像信息相结合,利用双重门控机制在筛选信息的同时结合三维注意力机制,使得图像能够更加精确地学习到边缘信息以突出重点减少边界弥漫,提升图像分割的准确程度,这对医学图像分割来说十分重要。实验结果相对于其他主流的方法在5个数据集的DICE分数上均取得最好的效果并在其他两个数据集上与最好结果近似,显示了EMPD所具有的优异性能和泛化能力。

在以后的工作中将探索EMPD模型在多模态医学图像分割任务中的表现,以进一步验证其在不同成像手段下的适应性和鲁棒性。此外,还将使用本文方法与医学大模型进行结合,引入语义先验信息以辅助分割模型对复杂结构的理解,从而提升模型的可解释性和诊断价值。

参考文献(References)

- Alshawi R, Hoque M T, Ferdaus M M, Abdelguerfi M, Niles K, Prathak K, et al. 2023. Dual attention U-Net with feature infusion: pushing the boundaries of multiclass defect segmentation [EB/OL]. [2025-09-14]. <https://arxiv.org/pdf/2312.14053.pdf>
- Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]
- Boykov Y Y and Jolly M P. 2001. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images//Proceedings of the 8th IEEE International Conference on Computer Vision. Vancouver, Canada: IEEE: 105-112 [DOI: 10.1109/ICCV.2001.937505]
- Bui N T, Hoang D H, Nguyen Q T, Tran M T and Le N. 2024. MEGANet: multi-scale edge-guided attention network for weak boundary polyp segmentation//Proceedings of 2024 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, HI, USA: IEEE: 4219-4228 [DOI: 10.1109/WACV57701.2024.00780]
- Cai L, Gao J Y and Zhao D. 2020. A review of the application of deep learning in medical image classification and segmentation. *Annals of Translational Medicine*, 8(11): #713 [DOI: 10.21037/atm.2020.02.44]
- Canny J. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6): 679-698 [DOI: 10.1109/TPAMI.1986.4767851]
- Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. 2021. TransUNet: transformers make strong encoders for medical image segmentation [EB/OL]. [2026-01-21]. <https://arxiv.org/pdf/2102.04306.pdf>
- Chen Z X, He Z W and Lu Z M. 2024. DEA-Net: single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE Transactions on Image Processing*, 33: 1002-1015 [DOI: 10.1109/TIP.2024.3354108]
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, et al. 2021. An image is worth 16×16 words: transformers for image recognition at scale//Proceedings of 2021 International Conference on Learning Representations. Vienna, Austria: [s.n.]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hesamian M H, Jia W J, He X J and Kennedy P. 2019. Deep learning techniques for medical image segmentation: achievements and challenges. *Journal of Digital Imaging*, 32(4): 582-596 [DOI: 10.1007/s10278-019-00227-x]
- Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]
- Huang G, Liu Z, Van Der Maaten L and Weinberger K Q. 2017. Densely connected convolutional networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2261-2269 [DOI: 10.1109/CVPR.2017.243]
- Jha D, Tomar N K, Biswas K, Durak G, Medetalibeyoglu A, Antalek M, et al. 2024. CT liver segmentation via PVT-based encoding and refined decoding//Proceedings of 2024 IEEE International Symposium on Biomedical Imaging. Athens, Greece: IEEE: 1-5 [DOI: 10.1109/ISBI56570.2024.10635659]
- Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84-90 [DOI: 10.1145/3065386]
- Lee H J, Kim J U, Lee S, Kim H G and Ro Y M. 2020. Structure boundary preserving segmentation for medical image with ambiguous boundary//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 4816-4825 [DOI: 10.1109/CVPR42600.2020.00487]
- Li Y C, Chen D L, Guo D H and Sun Y. 2025. Integrating spatiotemporal features and temporal constraints for dual-modal breast tumor diagnosis. *Journal of Image and Graphics*, 30(1): 268-281 (李一宸, 陈大力, 郭丁豪, 孙羽. 2025. 融合时空特征与时间约束的双模态乳腺肿瘤诊断. 中国图象图形学报, 30(1): 268-281) [DOI: 10.11834/jig.240217]
- Liu H L, Feng Y, Xu H, Liang S F, Liang H Z, Li S K, et al. 2022. MEA-Net: multilayer edge attention network for medical image seg-

- mentation. *Scientific Reports*, 12 (1) : #7868 [DOI: 10.1038/s41598-022-11852-y]
- Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, et al. 2021. Swin transformer: hierarchical vision transformer using shifted windows// *Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: IEEE: 9992-10002 [DOI: 10.1109/ICCV48922.2021.00986]
- Ma J, He Y T, Li F F, Han L, You C Y and Wang B. 2024. Segment anything in medical images. *Nature Communications*, 15(1): #654 [DOI: 10.1038/s41467-024-44824-z]
- Marmanis D, Schindler K, Wegner J D, Galliani S, Datcu M and Stilla U. 2018. Classification with an edge: improving semantic image segmentation with boundary detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 135: 158-172 [DOI: 10.1016/j.isprsjprs.2017.11.009]
- Milletari F, Navab N and Ahmadi S A. 2016. V-Net: fully convolutional neural networks for volumetric medical image segmentation//*Proceedings of the 4th International Conference on 3D Vision*. Stanford, USA: IEEE: 565-571 [DOI: 10.1109/3DV.2016.79]
- Rahman M M and Marculescu R. 2023a. Medical image segmentation via cascaded attention decoding//*Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, USA: IEEE: 6211-6220 [DOI: 10.1109/WACV56688.2023.00616]
- Rahman M M and Marculescu R. 2023b. Multi-scale hierarchical vision transformer with cascaded attention decoding for medical image segmentation. *Proceedings of Machine Learning Research*, 227: 1526-1544
- 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*. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Salahuddin Z, Woodruff H C, Chatterjee A and Lambin P. 2022. Transparency of deep neural networks for medical image analysis: a review of interpretability methods. *Computers in Biology and Medicine*, 140: #105111 [DOI: 10.1016/j.combiomed.2021.105111]
- Sandler M, Howard A, Zhu M L, Zhmoginov A and Chen L C. 2018. MobileNetV2: inverted residuals and linear bottlenecks//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 4510-4520 [DOI: 10.1109/CVPR.2018.00474]
- Simonyan K and Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition//*Proceedings of the 3rd International Conference on Learning Representations*. San Diego, USA: [s.n.]
- Su Z, Liu W Z, Yu Z T, Hu D W, Liao Q, Tian Q, et al. 2021. Pixel difference networks for efficient edge detection//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, Canada: IEEE: 5097-5107 [DOI: 10.1109/ICCV48922.2021.00507]
- Sun G D, Shi Y Y and Liu X. 2022. Feature extraction algorithm based on carotid artery ultrasound vessels. *Laser and Optoelectronics Progress*, 59(10): #1017002 (孙国栋, 石蕴玉, 刘翔. 2022. 基于颈动脉超声血管的特征提取算法. *激光与光电子学进展*, 59(10): #1017002) [DOI: 10.3788/LOP202259.1017002]
- Sun G Q, Pan Y Z, Kong W K, Xu Z C, Ma J H, Racharak T, et al. 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]
- Tang Y, Zhao D, Pertsau D, Gourinovitch A and Kupryanava D. 2025. Edge-guided and hierarchical aggregation network for robust medical image segmentation. *Biomedical Signal Processing and Control*, 101: #107202 [DOI: 10.1016/j.bspc.2024.107202]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. 2017. Attention is all you need//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, USA: Curran Associates Inc.: 6000-6010 [DOI: 10.5555/3295222.3295349]
- Wang H N, Cao P, Wang J Q and Zaiane O R. 2022a. UCTransNet: rethinking the skip connections in U-Net from a channel-wise perspective with transformer//*Proceedings of the 36th AAAI Conference on Artificial Intelligence*. [s.n.]: AAAI Press: 2441-2449 [DOI: 10.1609/aaai.v36i3.20144]
- Wang Q L, Wu B G, Zhu P F, Li P H, Zuo W M and Hu Q H. 2020. ECA-Net: efficient channel attention for deep convolutional neural networks//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 11531-11539 [DOI: 10.1109/CVPR42600.2020.01155]
- Wang R S, Lei T, Cui R X, Zhang B T, Meng H Y and Nandi A K. 2022b. Medical image segmentation using deep learning: a survey. *IET Image Processing*, 16 (5) : 1243-1267 [DOI: 10.1049/ipr2.12419]
- Wang W H, Xie E Z, Li X, Fan D P, Song K T, Liang D, et al. 2022c. PVT v2: improved baselines with pyramid vision transformer. *Computational Visual Media*, 8(3): 415-424 [DOI: 10.1007/s41095-022-0274-8]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Yang K X, Liu L, Fu X D, Liu L J and Peng W. 2025. Scale feature representation learning network for retinal vessels image segmentation. *Journal of Image and Graphics*, 30(3): 855-069 (杨可欣, 刘骊, 付晓东, 刘利军, 彭玮. 2025. 视网膜血管图像分割的尺度特征表示学习网络. *中国图象图形学报*, 30(3): 855-869) [DOI: 10.11834/jig.240120]

Yu Z T, Zhao C X, Wang Z Z, Qin Y X, Su Z, Li X B, et al. 2020. Searching central difference convolutional networks for face anti-spoofing//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 5294-5304 [DOI: 10.1109/CVPR42600.2020.00534]

Yuan Y H, Xie J Y, Chen X L and Wang J D. 2020. SegFix: model-agnostic boundary refinement for segmentation//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 489-506 [DOI: 10.1007/978-3-030-58610-2_29]

Zhang D P, Li Z, Xie Y G, Wang D Y, Tang S L, Bu Y Z, et al. 2025. Boundary cue deep fusion polyp image segmentation network. Journal of Image and Graphics, 30(5): 1479-1496 (章东平, 李铮, 谢亚光, 王都洋, 汤斯亮, 卜玉真, 等. 2025. 边界线索深度融合息肉图像分割网络. 中国图象图形学报, 30(5): 1479-1496) [DOI: 10.11834/jig.240383]

Zhang Y D, Liu H Y and Hu Q. 2021. TransFuse: fusing transformers and CNNs for medical image segmentation//Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention. Strasbourg, France: Springer: 14-24 [DOI: 10.1007/978-3-030-87193-2_2]

Zhou N, Xu M M, Shen B Q, Hou K, Liu S W, Sheng H, et al. 2024. ViT-UNet: a vision transformer based UNet model for coastal wetland classification based on high spatial resolution imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17: 19575-19587 [DOI: 10.1109/JSTARS.2024.3487250]

作者简介

黄冠琿,男,硕士研究生,主要研究方向为医学影像分析和计算机视觉。E-mail:beihuan1meng@gmail.com

刘翔,通信作者,男,教授,主要研究方向为医学影像分析、计算机视觉与人工生命。E-mail:xliu@sues.edu.cn

石蕴玉,女,讲师,主要研究方向为医学影像分析。E-mail:yunyushi@sues.edu.cn

纪雨,男,讲师,主要研究方向为垂直领域大模型及其优化。E-mail:yuji0201@outlook.com

王硕鸿,女,副研究员,主要研究方向为计算机视觉、医学生物学图像处理 and 机器学习。E-mail:wangsh@fas.harvard.edu