基于交叉视觉状态空间与多分支交互注意力的医学图像分割

薛 伟[1],陈创慧[1],杜明洋[2],钟 平[3],郑 啸[1]
(1. 安徽工业大学计算机科学与技术学院,安徽马鞍山 243032;2. 国防科技大学电子对抗学院,安徽合肥 230037;
3. 国防科技大学电子科学学院,湖南长沙 410073)

摘 要: 医学图像分割是智慧医疗领域的关键技术,旨在精准识别并分割影像中的器官或病变区域,为临床诊断与治疗决策提供可靠的量化依据. 近年来,基于卷积神经网络(Convolutional Neural Network,CNN)的医学图像分割方法因其优异的局部特征提取能力得到广泛应用. 然而,受限于卷积操作固有的局部感受野,CNN在建模长距离空间依赖和全局上下文信息方面仍存在不足. 尽管基于Transformer的方法通过自注意力机制实现了对全局特征的建模,但计算复杂度随序列长度的平方增长,制约了其实际应用效率. 针对上述问题,本文提出一种新的医学图像分割网络,该网络包含交叉视觉状态空间(Cross-Vision State Space,C-VSS)和多分支交互注意力(Multi-Branch Interactive Attention,MBIA)两个核心模块. C-VSS模块融合卷积操作的局部感知优势与状态空间的长序列建模能力,通过双分支协作策略,在保持线性计算复杂度的同时,实现对局部和全局特征的有效提取与融合. MBIA模块则通过多分支架构增强多尺度上下文信息的表征能力,并在编码器与解码器之间建立双向信息交互通道,实现跨层特征的动态融合,从而提升模型对复杂结构的感知能力. 为验证所提方法的有效性,在CVC-ColonDB、ISIC2017、ISIC2018和COVID-19这4个公开医学图像分割数据集上开展试验. 结果表明:与次优方法相比,本文方法在交并比(Intersection over Union,IoU)指标上分别提升了约0.94、0.83、1.04和2.28个百分点,在Dice相似系数(Dice Similarity Coefficient,DSC)指标上分别提升了约0.63、0.50、1.56和1.51个百分点. 此外,平均数(Average,Avg)指标在4个数据集上分别达到91.51%、91.74%、91.30%和88.78%,均优于所有对比方法,展现出最优性能,充分验证了所提方法在分割性能上的优越性. 进一步开展消融实验以验证核心模块的作用,实验表明:单独移除C-VSS模块后,IoU指标分别下降3.62、2.15、1.69和2.13个百分点,DSC指标分别下降2.25、1.29、1.02和1.40个百分点;单独移除MBIA模块后,IoU指标分别下降10.11、0.50、1.08和1.97个百分点,DSC指标分别下降6.54、0.30、0.65和1.30个百分点. 实验结果充分证明C-VSS与MBIA模块的有效性,且MBIA模块对性能提升的贡献更为显著,二者协同作用可进一步优化模型性能.

关键词: 医学图像分割;交叉视觉状态空间(C-VSS);多分支交互注意力(MBIA);动态特征融合;卷积神经网络(CNN);Transformer

基金项目: 国家自然科学基金(No.62441207);马鞍山市科技创新攻坚计划项目(No.2024RGZN001)

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112(2025)09-3331-14

电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250642

A Medical Image Segmentation Network Based on Cross-Visual State Space and Multi-Branch Interactive Attention

XUE Wei[1], CHEN Chuang-hui[1], DU Ming-yang[2], ZHONG Ping[3], ZHENG Xiao[1]
(1. *School of Computer Science and Technology, Anhui University of Technology, Maanshan, Anhui* 243032, *China*;
2. *College of Electronic Engineering, National University of Defense Technology, Hefei, Anhui* 230037, *China*;
3. *College of Electronic Science and Technology, National University of Defense Technology, Changsha, Hunan* 410073, *China*)

Abstract: Medical image segmentation is a key technology in the field of smart healthcare, aiming to accurately identify and segment organs or pathological regions within images, thereby providing reliable quantitative evidence for clinical diagnosis and treatment decision-making. In recent years, medical image segmentation methods based on convolutional neural network (CNN) have been widely adopted due to their excellent capability in extracting local features. However, due to the

收稿日期:2025-07-23;录用日期:2025-09-22;责任编辑:覃怀银

inherent local receptive field of convolution operations, CNN still suffers from limitations in modeling long-range spatial dependencies and global contextual information. Although Transformer-based methods achieve global feature modeling through the self-attention mechanism, their computational complexity grows quadratically with sequence length, limiting their efficiency in practical applications. To mitigate the aforementioned issues, this paper proposes a new medical image segmentation network, which mainly consist of two core modules: cross-vision state space (C-VSS) and multi-branch interactive attention (MBIA). The C-VSS module integrates the local perception advantage of convolutional operation with the long-sequence modeling capability of state space model. Through a dual-branch collaborative strategy, it achieves effective extraction and fusion of local and global features while maintaining linear computational complexity. The MBIA module enhances the representation of multi-scale contextual information through a multi-branch architecture and establishes bidirectional information interaction pathways between the encoder and the decoder to enable dynamic fusion of cross-level features, thereby improving the model's ability to perceive complex structures. Experimental results on four public medical image segmentation datasets, including CVC-ColonDB, ISIC2017, ISIC2018, and COVID-19, demonstrate that our method outperforms the second-best approach by approximately 0.94, 0.83, 1.04, and 2.28 percentage points in intersection over union (IoU) and 0.63, 0.50, 1.56, and 1.51 percentage points in dice similarity coefficient (DSC), respectively. In addition, the proposed method achieves average (Avg) scores of 91.51%, 91.74%, 91.30%, and 88.78% on the four datasets, respectively, all of which are higher than those of the comparative methods, demonstrating its superior segmentation performance. Furthermore, ablation studies show that removing the C-VSS module alone leads to a decrease of 3.62, 2.15, 1.69, and 2.13 percentage points in IoU, and 2.25, 1.29, 1.02, and 1.40 percentage points in DSC, respectively. Removing the MBIA module alone results in a decline of 10.11, 0.50, 1.08, and 1.97 percentage points in IoU, and 6.54, 0.30, 0.65, and 1.30 percentage points in DSC, respectively. The experimental results fully verify the effectiveness of the C-VSS and MBIA modules, indicate that the MBIA module contributes more significantly to performance improvement, and reveal a notable synergy between the two.

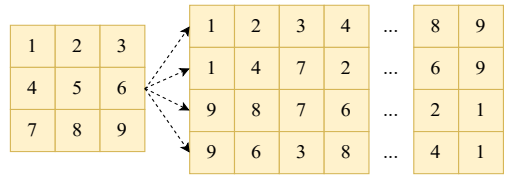
Key words: medical image segmentation; cross-visual state space module; multi-branch interactive attention module; dynamic feature fusion; convolutional neural network; Transformer

Foundation Item(s): National Natural Science Foundation of China (No.62441207); Science and Technology Innovation Tackle Plan Project of Maanshan (No.2024RGZN001)

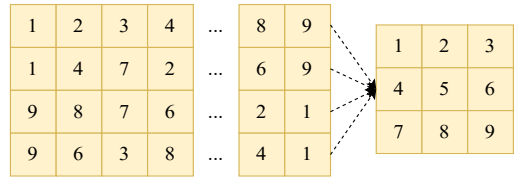
1 引言

医学图像分割在智慧医疗领域具有重要的研究价值,其目标是将器官、病灶、血管等感兴趣区域从计算机断层扫描(Computed Tomography,CT)、磁共振成像(Magnetic Resonance Imaging,MRI)、超声等医学图像中分离出来[1]. 早期的分割主要依赖于传统的图像处理技术,如基于阈值分割方法、基于边缘检测分割方法等[2],这些方法对医学图像的质量和噪声敏感,泛化能力较弱[3]. 随着深度学习的发展,基于卷积神经网络(Convolutional Neural Network,CNN)和Transformer[4]的医学图像分割方法实现了从输入图像到分割结果的直接映射,提升了分割的精度与效率. 其中,具有代表性的基于CNN的方法包括迭代边界优化小样本分割网络(Iterative Boundary Refinement Few-Shot Segmentation Network,IBR-FSS-Net)[5]、结构重参数化与多尺度深度监督网络(Structural Reparameterization and Multi-scale Deep Supervision Network,SR&MDS-Net)[6]和MADGNet(Modality-Agnostic Domain Generalizable medical image segmentation by multi-frequency in multi-scale attention)[7]. 这些方法通过引入不同的注意力机制,细化特征的提取过程,从而增强模型对关键区域的感知能力. 但由于卷积操作具有局部感受野的固有特性,CNN在建模图像的长程依赖关系上存在局限性[8],难以有效捕捉医学图像中相距较远的结构间的语义关联. 基于Transformer的方法通过引入位置编码和自注意力机制,实现对图像全局上下文信息的直接建模,有效捕捉了图像中不同结构间的关系,代表性工作包括UCTransNet(rethinking the skip connections in U-Net from a Channel - wise perspective with Transformer)[9]、多尺度卷积调制网络(Multi-Scale Convolution Modulation Network,MSCMNet)[10]和多分支多尺度注意力网络(Multi-branch and multi-scale Attention Network,M2ANet)[11]等. 为进一步结合CNN与Transformer的优势,文献[12]提出一种将U-Net(用于生物医学图像分割的卷积网络)[13]与Swin-Unet(用于医学图像分割的类Unet纯Transformer)结合的双分支网络,实现了局部和全局信息的有效融合. 文献[14]将Swin Transformer与U-Net编码器结合,提出一种基于双通道自注意力的分割方法. 上述方法取得了良好性能,但普遍受限于Transformer的二次计算复杂度,在处理高分辨率医学图像时,易受硬件和内存资源的限制.

状态空间模型(State Space Model,SSM)尤其是结构化状态空间模型(Structured state space model,S4)[15]的引入提供了一种新的解决方案.SSM利用状态空间方程,在保持线性计算复杂度的同时,有效地对长序列依赖关系进行建模[16].在此基础上,文献[17]提出一种具有选择性的SSM模型:Mamba,该模型通过参数化网络中的参数,使其能够依据不同序列内容动态调整其状态转移过程,进一步提升了SSM模型在处理长序列任务中的性能与效率.随后,文献[18]提出Vision Mamba,该模型集成双向的SSM模块,实现了更高效的全局特征建模,并在效率和性能方面超越了传统的Transformer架构.文献[19]引入2D选择扫描(2D Selective Scan,SS2D)机制,通过从4个方向的扫描对图像中的全局特征进行建模.SS2D作为Mamba模型向二维视觉任务扩展的关键技术之一,提升了SSM在视觉表达方面的能力,使其能够直接应用于图像处理任务中.此外,文献[20]提出SMM-UNet模型,通过有效整合局部和全局特征,在降低模型计算复杂度的同时,提升了其对形态多变的分割任务的适用性.尽管SS2D在图像中表现出良好的全局建模能力,但其对图像中的所有区域(包括感兴趣区域和背景区域)均采用统一的关注机制,缺乏对关键局部结构的针对性建模能力.如图1所示,SS2D将输入沿着不同的扫描路径展开为序列,对每条序列独立应用SSM进行特征变换.然而,这种方法为所有特征赋予相同的权重,未区分语义重要性不同的区域,导致边界或微小病灶区域在全局状态更新过程中被平滑,从而影响模型在复杂医学图像分割任务中的性能.



(*a*) 扫描扩展



(*b*) 扫描合并

图1　SS2D结构[19]

针对这一问题,本文在SS2D的基础上进一步拓展SSM在医学图像分割任务中的适用性,提出一种交叉视觉状态空间(Cross-Vision State Space,C-VSS)模块,以提升模型对关键区域的建模精度和语义表达能力.此外,U-Net[13]作为最具有代表性的医学图像分割框架之一,其采用对称的编码-解码结构,并利用跳跃连接以补充下采样丢失的细节信息,这对于医学图像分割至关重要.目前,包括ACC-Unet(A Completely Convolutional Unet model for the 2020s)[21]、SMESwin Unet[22]、Mamba-UNet(UNet-like pure visual Mamba for medical image segmentation)[23]在内诸多模型证明了U-Net框架在医学图像分割任务上的有效性.受此启发,本文以C-VSS为基础单元构建U型结构,提出了基于C-VSS与多分支交互注意力(Multi-Branch Interactive Attention,MBIA)的分割网络.本文的主要贡献如下:

(1)C-VSS模块,用于提取并融合图像的局部和全局特征.C-VSS模块结合卷积运算与SSM的优势,通过引入双分支协作策略,实现对卷积操作提取的局部特征与SSM提取的全局依赖关系的多层次融合.

(2)提出MBIA模块,以增强不同尺度信息间的交互能力.MBIA模块通过在编码器和解码器之间建立双向信息传递通道,实现跨层级的动态特征融合,提升多尺度特征的表示能力.

(3)基于C-VSS模块与MBIA模块,提出基于C-VSS与MBIA的网络(Cross-Visual state space and Multi-Branch interactive Network,CVMBNet)用于医学图像分割.实验结果表明:与现有主流方法相比,CVMBNet在多个数据集上均表现出更优的性能,实现了分割精度的提升.

2　相关工作

2.1　基于CNN与Transformer的分割

CNN通过局部感受野与滑动窗口机制,有效捕获医学图像中的边缘、纹理等局部特征,为病灶分割和器官识别等任务提供可靠的特征表示.然而,这种基于局部感受野的特征提取方式难以建模医学图像中长距离依赖关系[24].为了缓解这一局限性,文献[25]设计了一种多尺度减法聚合模块,通过计算不同尺度特征间的差异信息并融合到解码器,增强了模型对像素级和结构级特征的辨别能力.文献[26]提出一种基于深度卷积的多尺度解码器,通过结合多尺度信息与注意力机制,细化了模型的特征解码过程.上述方法均在卷积运算的基础上,引入多尺度处理或注意力机制等手段,以提高模型对特征的提取和融合能力.与之不同的是,文献[27]提出一种层次化的Swin Transformer模型,通过滑动窗口机制以及跨窗口交互策略,在降低模型计算复杂度的同时,实现了对长距离依赖关系的建模.在此基础上,文献[28]将Swin Transformer块作为网络的基础单元,构建了一个完全基于Transformer的U-Net网

络. CASCADE(CASCADE attention decoding)[29]在Transformer的基础上,融合卷积注意力模块设计了一种层级化级联注意力解码器,用于增强模型提取上下文语义信息的能力. 此外,部分研究通过重构自注意力机制的计算形式,在提升模型性能的同时提高了计算效率. 例如,MultiTrans[30]调整了标准自注意力机制中矩阵乘法的运算顺序,并通过引入头共享机制,降低了自注意力模块的计算复杂度. 文献[31]提出一种轻量级的极化多尺度特征自注意力网络,通过将每个注意力点的关键向量的计算简化为全局关键向量的计算,将标准自注意力的计算复杂度从二次复杂度降低到一次复杂度.

上述方法均是在CNN与Transformer架构上引入多尺度信息和注意力机制来提高模型对多层次上下文信息和关键区域特征的感知能力,在一定程度上取得了良好的性能. 然而,基于CNN的方法在全局特征提取上具有局限性,基于Transformer的方法其计算复杂度与输入序列二次方成正比. 尽管已有研究提出重构标准自注意力机制来降低模型计算复杂度,但此类方法仍在性能与效率的平衡上面临挑战,为此,本文提出一种基于SSM架构的医学图像分割网络,在保持线性计算复杂度的同时,提高模型的性能.

2.2 基于SSM的分割

SSM利用方程和矩阵对输入序列进行系统化建模与分析,从而增强模型在处理长距离依赖关系方面的能力. 作为主流的SSM模型,Mamba[17]结合选择性机制和硬件感知技术,在保持计算效率的同时,有效地捕捉序列中的长距离依赖关系. SS2D[19]将Mamba的适用范围从一维扩展到二维空间,提高了Mamba的视觉表达能力. 在随后的研究中,文献[32]结合SSM与CNN,提出了用于医学图像分割的Mamba模型:U-mamba. 文献[33]在SS2D基础上提出视觉状态空间(Visual State Space, VSS)模块来构建VM-UNet,这是医学图像分割中首个完全基于SSM的U-Net模型. VM-UNet V2(rethinking Vision Mamba UNet for medical image segmentation)[34]改进了VM-UNet,通过引入新的跳跃连接实现高级和低级语义信息的融合. Swin-UMamba(mamba-based unet with imagenet-based pretraining)[35]在VSS块的基础上,进一步引入上采样块来恢复图像的分辨率. 文献[36]将SS2D引入到高阶视觉领域,提出了H-vmunet(High-order vision mamba UNet)用于医学图像分割任务.

上述方法基于U-Net架构,通过引入SS2D改进其编码器、解码器及跳跃连接,在提升分割性能的同时,推动了SSM在医学图像分割领域的应用. 然而,SS2D存在一定的局限性,其在建模全局依赖关系时,削弱了模型整合边缘纹理等局部特征的能力,从而影响分割边界的准确性. 针对上述问题,本文提出C-VSS模块以构建U-Net的编码器和解码器. 与现有方法不同,C-VSS模块采用双分支并行结构,两条分支对不同的通道特征使用交叉SS2D(Cross-SS2D, C-SS2D)模块和卷积运算以提取全局和局部特征,并通过交叉相乘机制实现动态融合,增强了特征的表达能力. 此外,为了进一步强化多尺度特征的双向信息交互能力,本文提出MBIA模块. MBIA模块将编码器输出的多尺度特征沿通道划分为4个子块,并引入Mamba机制对各子块进行跨分支特征交互,从而增强不同尺度信息间的上下文关联.

3 本文方法

3.1 总体结构

如图2(a)所示,本文提出的CVMBNet网络采用经典的编码器-解码器架构,主要由输入层、编码器、解码器、MBIA模块和输出层构成. 输入层采用7×7卷积核进行特征映射,用于将输入图像转换为模型所需的中间特征维度;相应地,输出层使用7×7卷积核作为分割头,以实现特征图到分割结果的转换. 编码器与解码器均采用本文所提的C-VSS模块作为基础构建单元,其中编码器通过下采样操作逐层降低空间分辨率并扩展特征维度,解码器则通过上采样操作逐步恢复图像分辨率和特征维度. 特别地,在输入层和输出层、对应层级的编码器与解码器之间,本文设计了MBIA模块作为跳跃连接,该模块能够有效促进不同尺度信息之间的双向信息交互,增强网络对多尺度特征的融合能力. 下面将详细解释本文所提的模块.

3.2 C-VSS模块

基于SS2D,本文提出C-VSS模块,用于提取图像中的局部细节信息和全局上下文信息. 进一步分析可知,单一分支结构的提取方式在兼顾局部细节与全局上下文时,存在表征能力有限的问题;相比之下,双分支结构不仅能够提高模型性能,还可以降低计算复杂度,基于此,本文采用双分支协作策略,以促进特征的提取和融合. 具体而言,输入特征在通道维度上被均匀划分为两个子分支,每个分支独立处理其对应的通道特征. 在此基础上,本文对两路特征进行逐元素交叉相乘,实现非线性互补融合;随后将经变换后的双路特征沿通道维度拼接,以得到完整语义信息的输出特征. 下面将详细介绍C-VSS模块的构成.

3.2.1 SS2D

SS2D是SSM的一种二维扩展形式. 如图1所示,SS2D包含扫描扩展与扫描合并两个阶段. 在扫描扩展阶段,输入的特征图首先沿着4个不同的方向(左上到右下、左下到右上、右下到左上、右上到左下)进行展

开,转换为一维向量. 每个一维向量独立送入SSM中进行特征提取,这种方式使得图像中的每一个元素能够整合不同的位置信息,从而捕捉更加丰富的空间依赖关系. 在扫描合并阶段,从4个方向得到的序列按照原来方向进行合并相加,这一过程将输出信息重新映射回二维空间,并恢复到输入图像的大小. 通过上述扫描扩展与合并机制,SS2D能够有效建模图像在高度和宽度两个维度上的依赖关系,从而更全面地捕捉图像中的全局结构.

3.2.2　C-SS2D

如图2(*c*)所示,C-SS2D主要由SS2D、层归一化(LayerNorm,LN)和多层感知机(MultiLayer Perceptrons,MLP)组成的. 对输入图像$x' \in \mathbb{R}^{Bs\times Cs\times H\times W}$(其中,Bs为批大小,Cs为通道数,$H$和$W$分别为高度和宽度),首先通过LayerNorm层进行归一化以缓解内部协变量偏移,提升训练的稳定性. 其次,特征图经过SS2D模块以建模空间长程依赖关系. 该模块的输出与原始输入特征通过残差连接进行初步融合,得到包含全局信息的增强表示. 再次,融合后的特征经过LayerNorm层,并通过MLP层完成非线性特征变换. 最后,将MLP层的输出与前一阶段的输出经过二次残差连接以缓解深层网络中的梯度消失问题. 上述过程表示如下:

$$\begin{aligned} o &= \mathrm{SS2D}\left(\mathrm{LN}(x')\right) + x' \\ o_{\mathrm{C\text{-}SS2D}} &= \mathrm{MLP}\left(\mathrm{LN}(o)\right) + o \end{aligned} \tag{1}$$

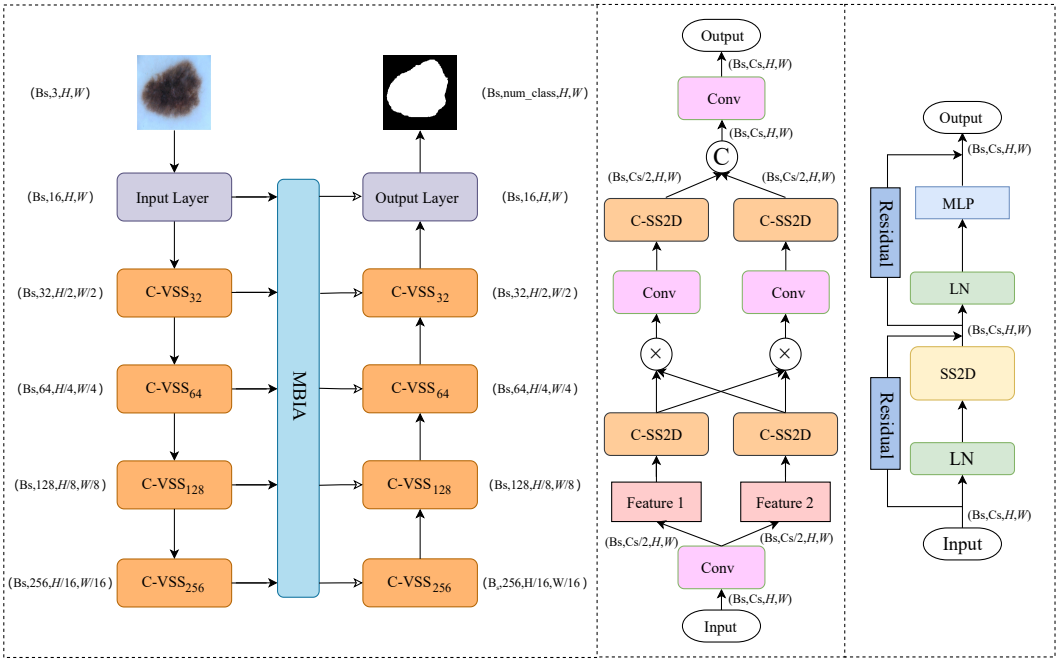
其中,o表示中间结果;$o_{\mathrm{C\text{-}SS2D}}$表示C-SS2D模块的输出.

3.2.3　C-VSS模块

如图2(*b*)所示,C-VSS模块主要由卷积块和C-SS2D模块组成. 其中,卷积运算是为了提取图像中的局部细节信息,而C-SS2D模块是为了提取图像中的全局上下文信息. 具体地,输入图像$x \in \mathbb{R}^{Bs\times Cs\times H\times W}$首先通过卷积层以进行初步特征提取,随后沿着通道维度将其划分为两个子特征图$f_1 \in \mathbb{R}^{Bs\times \frac{Cs}{2}\times H\times W}$和$f_2 \in \mathbb{R}^{Bs\times \frac{Cs}{2}\times H\times W}$,并分别对$f_1$和$f_2$使用C-SS2D模块以提取特征$m_{ij}/n_{ij}$. 其次,对于上述输出特征,采用交叉相乘的方式进一步融合空间与通道维度的信息,以增强特征表达能力. 再次,分别对上述输出结果使用卷积操作和C-SS2D来进一步提取图像中的局部特征和全局上下文特征,得到$\bar{f}_{ij}$. 最后,将两路处理后的特征沿着通道维度进行拼接和卷积操作,以融合完整的语义信息. 上述过程表示如下:

$$\begin{aligned} f_{ij} &\leftarrow \mathrm{Conv}(x), \text{其中} i,j \in [1,2] \text{且} i \neq j \\ m_{ij} &= n_{ij} = o_{\mathrm{C\text{-}SS2D}}(f_{ij}) \\ \bar{f}_{ij} &= o_{\mathrm{C\text{-}SS2D}}\left(\mathrm{Conv}(m_{ij} * n_{ji})\right) \\ o_{\mathrm{C\text{-}VSS}} &= \mathrm{Conv}\left(\mathrm{Cat}(\bar{f}_{ij})\right) \end{aligned} \tag{2}$$

其中,i、j表示子特征图的下标;Conv表示卷积操作;Cat表示拼接操作;$o_{\mathrm{C\text{-}VSS}}$表示C-VSS模块的输出结果;$m_{ij} = n_{ij}\bar{f}_{ij}$表示中间结果.



(*a*) CVMBNet总体网络架构　　(*b*) C-VSS模块架构　　(*c*) C-SS2D模块架构

图2　网络架构

3.3 MBIA模块

如图3(*a*)所示,本文设计MBIA模块作为网络的跳跃连接,其核心组成Inter-kernel如图3(*b*)所示. 对5个不同尺度的输入$\mathrm{En}_k \in \mathbb{R}^{\mathrm{Bs}\times \mathrm{Cs}_k \times H\times W}$, $k\in[1,5]$, Inter-kernel首先将输入沿通道维度进行拼接并归一化得到$\mathrm{En}_{\mathrm{all}} \in \mathbb{R}^{\mathrm{Bs}\times \sum_{k=1}^{5}\mathrm{Cs}_k \times H\times W}$. 紧接着,将$\mathrm{En}_{\mathrm{all}}$分为4个子块,每个子块包含来自不同尺度的通道特征以便于特征交互. 其次,采用Mamba模块对4个子块进行全局特征交互,在此基础上,引入乘加操作(Multiply Add Operation,MAO),即式(3)来优化梯度,从而增强深层网络训练的稳定性. 再次,完成特征交互后,将4个通道特征图进行融合,并通过LayerNorm层进行归一化处理. 最后,将融合后的特征图通过MLP层进行特征变换,再通过BN(BatchNorm)和ReLu激活函数,进一步优化特征表达. 上述过程表示如下:

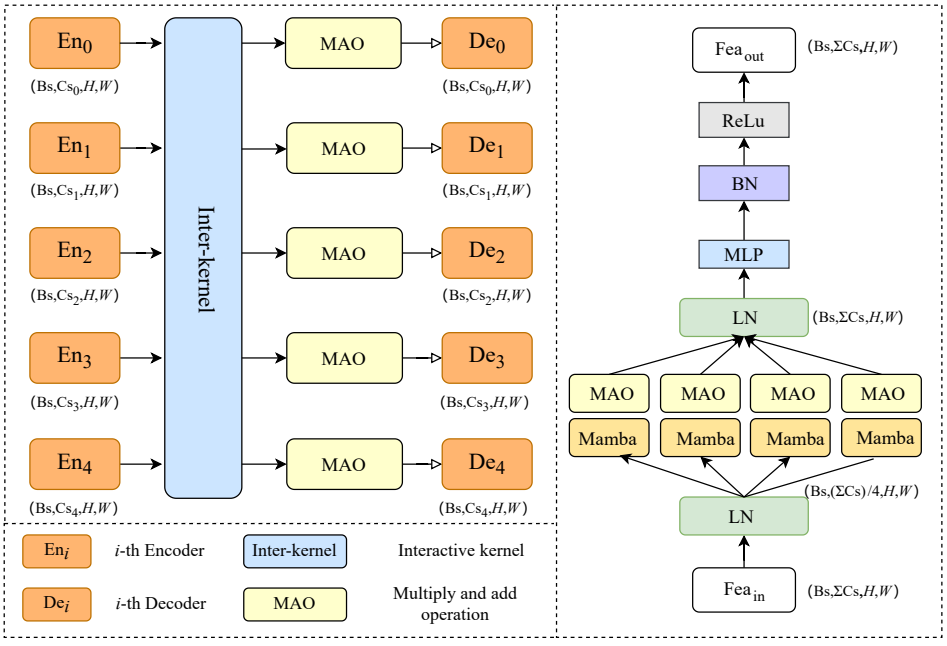
$$l_s \leftarrow \mathrm{LN}(\mathrm{En}_{\mathrm{all}}), s\in[1,4]$$

$$\begin{aligned} \bar{l}_s &= \mathrm{MAO}\left(\mathrm{Mamba}(l_s)\right) \\ o_{\mathrm{Inter-kernel}} &= \mathrm{ReLu}\left(\mathrm{BN}\left(\mathrm{MLP}\left(\mathrm{LN}\left(\mathrm{Cat}(\bar{l}_s)\right)\right)\right)\right) \end{aligned} \tag{3}$$

其中,l_s表示特征图$\mathrm{En}_{\mathrm{all}}$的分割结果;$\bar{l}_s$表示中间结果;MAO表示乘加操作;$\mathrm{MAO}\left(\mathrm{Mamba}(l_s)\right)=l_s*(1+\mathrm{Mamba}(l_s))$;$o_{\mathrm{Inter-kernel}}$表示Inter-kernel的输出. 为了与解码器的结构相对应,本文将$o_{\mathrm{Inter-kernel}}$沿通道维度划分为5个分支,并对5个分支进行MAO操作,即式(4)以增强特征的复用,MBIA模块的数学表达式如下:

$$\begin{aligned} \mathrm{En}_{\mathrm{all}} &= \mathrm{Cat}(\mathrm{En}_k) \\ \overline{\mathrm{De}}_k &\leftarrow o_{\mathrm{Inter-kernel}}(\mathrm{En}_{\mathrm{all}}) \\ \mathrm{De}_k &= \mathrm{MAO}(\overline{\mathrm{De}}_k), k\in[1,5] \end{aligned} \tag{4}$$

其中,$\overline{\mathrm{De}}_k$表示将$o_{\mathrm{Inter-kernel}}$沿通道划分的结果;$\mathrm{De}_k$表示MBIA的5个输出结果;$\mathrm{MAO}(\overline{\mathrm{De}}_k)=\mathrm{En}_k*(1+\overline{\mathrm{De}}_k)=\mathrm{En}_k*\left(1+o_{\mathrm{Inter-kernel}}(\mathrm{En}_{\mathrm{all}})\right)$.



(*a*) MBIA (*b*) Inter-kernel

图3 MBIA结构及核心设计部分Inter-kernel

3.4 损失函数

本文采用二元交叉熵(Binary Cross Entropy,BCE)损失和Dice损失的混合损失来优化模型. 其中,BCE损失用于提升像素级别的分类准确性,而Dice损失主要用于增强预测区域与真实标注之间的空间重叠度. 为了准确地表达该损失函数,以N表示图像总像素数量,y_i表示第i个像素的真实标签,p_i表示其预测结果,BCE损失表示为

$$L_{\mathrm{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i\ln(p_i)+(1-y_i)\ln(1-p_i)\right] \tag{5}$$

Dice损失表示为

$$L_{\mathrm{Dice}} = 1-\frac{2\sum_{i=1}^{N}p_i y_i+\epsilon}{\sum_{i=1}^{N}p_i+\sum_{i=1}^{N}y_i+\epsilon} \tag{6}$$

其中,ϵ表示平滑项. 综上所述,本文所使用的总损失函数为$L_{\mathrm{Total}}=L_{\mathrm{BCE}}+L_{\mathrm{Dice}}$.

4　实验设置

4.1　数据集处理

CVC-ColonDB[①]是结直肠息肉分割数据集,其训练集包含304张图像,验证集和测试集各包含38张图像,所有图像均包含有医学专家手工标注的像素级息肉分割掩模. 在本实验中,沿用上述数据集的划分方法,将CVC-ColonDB数据集分为训练集、验证集和测试集.

ISIC2017[②]包含2 000张皮肤镜图像,有3种不同类型的皮肤疾病. 在本实验中,对该数据集进行划分,其中1 250张作为训练集,150张作为验证集,剩余600张作为测试集,以评估模型的泛化性能.

ISIC2018[③]包含2 594张图像,有7种不同类型的皮肤病变. 所有图像均配有高质量的分割标签. 在本实验中,对该数据集进行划分,分别选取1 750张、250张和594张图像作为实验的训练集、验证集和测试集,以用于模型的训练、超参数调优以及性能评估.

COVID-19[④]是用于肺结节分割任务的肺部CT图像数据集,该数据集包含有2 534张带有相应分割标签的CT图像,将该数据集依据1 750∶250∶534的比例划分为训练集、验证集和测试集,用于模型训练、参数调优及性能评估.

为提升模型的泛化性能,本文在数据预处理阶段对所有的数据集均采用随机增强操作. 具体而言,对输入数据进行标准化处理,并且依次采用随机水平和垂直翻转、随机缩放、随机旋转操作,以扩充样本的多样性. 此外,为保持输入数据空间维度的一致性,所有样本在输入前均采用双线性插值算法缩放至256×256大小.

4.2　评价指标

本文使用交并比(Intersection over Union, IoU)、Dice相似系数(Dice Similarity Coefficient, DSC)、准确率(Accuracy, Acc)、特异度(Specificity, Spe)、敏感度(Sensitivity, Sen)和平均数(Average, Avg)作为分割性能的评价指标,以避免单一指标的局限性. 其中,$IoU=TP/(TP+FP+FN)$用于衡量预测结果与真实结果之间的重叠程度;$DSC=2TP/(2TP+FP+FN)$用于衡量预测值与真实值之间的相似程度;$Acc=(TN+TP)/(TN+TP+FP+FN)$用于衡量模型正确预测样本占总样本数的比例;$Spe=TN/(TN+FP)$反映模型正确识别负类样本的能力;$Sen=TP/(TP+FN)$用于衡量模型正确识别正类样本的能力;$Avg=(IoU+DSC+Acc+Spe+Sen)/5$用于量化模型输出指标的集中趋势. 在上述指标中,TP表示真正例,FP表示假正例,FN表示假反例,TN表示真反例.

4.3　对比算法与参数设置

本文选取多种具有代表性的模型进行对比实验,具体如下:基于CNN架构的U-Net[13]、ACC-UNet[21]和MADGNet[7];基于Transformer架构的UCTransNet[9]、Swin-Unet[28];基于SSM架构的VM-UNet[33]、VM-UNet V2[34]和H-vmunet[36]. 在模型优化方面,本文采用AdamW优化器进行参数更新以提升训练过程的稳定性和收敛速度. 具体训练设置如下:初始学习率设为1×10^{-3},并结合余弦退火学习率调度策略进行动态调整;所有模型均训练250个轮次,以保证各个模型都能充分收敛.

5　实验结果

本文在保持相同实验设置的前提下,对每组实验独立重复3次,取指标的平均值和标准差作为最终评估结果,在固定随机种子的条件下,能够保持复现性.

5.1　定量结果

为了定量评估所提方法的分割性能,表1~表4分别列出了所有对比方法在各项指标上的实验结果. 其中,粗体表示在对应指标上的最优性能,下划线表示次优结果. 特别说明的是,本文所提到的提升,均是与次优结果进行对比. 从实验结果可以看出,本文所提方法在4个数据集上的多个评价指标中均取得了最优表现,充分验证了本文方法在医学图像分割任务上的优越性能.

在CVC-ColonDB数据集上,CVMBNet在前5个评价指标上均取得最优结果,相较于次优的方法,性能分别提升约0.94(IoU)、0.63(DSC)、0.06(Acc)、0.02(Spe)和0.80(Sen)个百分点. 这表明:CVMBNet在结肠息肉分割数据集上优势明显,尤其是在IoU和DSC值上提升明显. 在ISIC2017数据集上,CVMBNet在IoU、DSC和Spe这3项指标中取得最优结果,相较于次优的方法,分别提升了约0.83、0.50和0.11个百分点. 然而,在该数据集的Acc和Sen指标上,VM-UNet V2和ACC-UNet的表现更优. 这一结果表明:CVMBNet在ISIC2017数据集上具有良好的形态学建模能力,能够有效识别负类样本,但在准确性和敏感度上还有进一步优化的空间. 在ISIC2018数据集上,CVMBNet在前4项评价指标上均取得了最优的性能,分别较次优结果提升了1.04、1.56、0.37和0.13个百分点,但是其在Sen评价指标上没有取得最优. 同样地,在COVID-19数据集上,CVMBNet

① https://www.kaggle.com/datasets/pavancshekar/cvc-colondb/data.

② https://challenge.isic-archive.com/data/#2017.

③ https://challenge.isic-archive.com/data/#2018.

④ https://www.kaggle.com/datasets/piyushsamant11/ pidata-new-names/data?status=pending&suggestionBundleId=974&selectedOnly=true.

表1 CVC-ColonDB数据集结果 单位:%

指标 对比	U-Net 2015	ACC-UNet 2023	MADGNet 2024	UCTransNet 2022	Swin-Unet 2022	VM-UNet 2024	VM-UNetv2 2024	H-vmunet 2025	CVMBNet Ours
IoU	72.97±1.87	73.98±0.14	78.25±1.81	77.78±0.36	69.06±0.50	70.92±7.63	78.99±0.24	79.98±3.99	**80.92±0.65**
DSC	81.41±1.41	82.33±0.36	86.21±1.46	85.91±0.31	79.92±0.34	82.75±5.23	88.26±0.15	88.82±2.46	**89.45±0.39**
Acc	98.60±0.12	98.31±0.01	98.61±0.21	98.83±0.00	98.40±0.09	98.03±0.57	98.69±0.00	98.79±0.24	**98.85±0.04**
Spe	98.88±0.01	98.32±0.00	98.86±0.28	98.92±0.01	99.03±0.13	99.05±0.22	99.57±0.07	99.41±0.04	**99.43±0.04**
Sen	83.14±2.09	87.58±0.46	88.08±1.31	86.87±0.39	81.42±0.90	81.51±6.19	84.42±1.21	88.00±3.83	**88.88±0.14**
Avg	87.00±1.10	88.10±0.19	90.00±1.01	89.66±0.21	85.57±0.39	86.45±3.97	89.99±0.33	91.00±2.11	**91.51±0.25**

表2 ISIC2017数据集结果 单位:%

指标 对比	U-Net 2015	ACC-UNet 2023	MADGNet 2024	UCTransNet 2022	Swin-Unet 2022	VM-UNet 2024	VM-UNetv2 2024	H-vmunet 2025	CVMBNet Ours
IoU	78.46±3.33	80.61±2.66	80.79±2.62	80.91±1.93	79.71±3.28	77.79±7.41	78.15±5.67	83.01±0.51	**83.84±0.02**
DSC	86.30±2.63	87.91±2.02	87.96±1.97	88.09±1.32	87.43±2.22	87.31±4.70	87.62±3.57	90.71±0.30	**91.21±0.01**
Acc	95.97±0.19	96.29±0.16	96.40±0.17	96.34±0.07	96.26±0.29	97.09±0.02	**97.18±0.32**	96.43±0.11	96.61±0.00
Spe	97.06±0.04	96.67±0.25	97.37±0.09	97.01±0.14	96.95±0.51	98.20±0.19	98.17±0.26	98.55±0.05	**98.66±0.04**
Sen	89.78±1.49	**91.27±0.45**	89.84±1.32	90.67±0.55	90.73±0.44	88.89±2.46	85.64±4.11	87.85±0.36	88.36±0.18
Avg	89.51±1.54	90.55±1.11	90.47±1.23	90.60±0.80	90.22±1.35	89.86±2.96	89.35±2.79	91.31±0.27	**91.74±0.05**

表3 ISIC2018数据集结果 单位:%

指标 对比	U-Net 2015	ACC-UNet 2023	MADGNet 2024	UCTransNet 2022	Swin-Unet 2022	VM-UNet 2024	VM-UNetv2 2024	H-vmunet 2025	CVMBNet Ours
IoU	79.88±0.87	81.05±1.00	81.58±1.21	80.90±1.09	81.08±1.20	75.93±5.42	74.52±6.48	80.09±2.55	**82.62±0.28**
DSC	87.58±0.46	88.10±0.56	88.77±0.69	88.11±0.75	88.45±0.68	86.21±3.50	85.24±4.26	88.92±1.57	**90.48±0.17**
Acc	94.53±0.92	94.81±1.19	94.91±1.13	94.64±1.19	95.06±1.10	95.55±0.06	95.32±0.24	94.69±1.32	**95.92±0.08**
Spe	95.86±0.63	95.22±1.83	94.52±1.50	94.92±1.03	94.80±1.50	97.40±0.06	97.44±0.21	96.57±1.64	**97.57±0.12**
Sen	89.57±0.99	89.86±0.59	**92.03±0.03**	90.67±1.00	91.86±0.84	85.98±2.90	84.31±3.63	88.46±0.43	89.93±0.06
Avg	89.48±0.77	89.81±1.03	90.36±0.91	89.85±1.01	90.25±1.06	88.21±2.39	87.37±2.96	89.75±1.50	**91.30±0.14**

表4 COVID-19数据集结果 单位:%

指标 对比	U-Net 2015	ACC-UNet 2023	MADGNet 2024	UCTransNet 2022	Swin-Unet 2022	VM-UNet 2024	VM-UNetv2 2024	H-vmunet 2025	CVMBNet Ours
IoU	69.34±0.24	71.30±0.24	69.99±0.25	69.25±0.50	62.74±0.20	62.09±8.35	72.41±0.07	72.06±1.01	**74.69±1.41**
DSC	80.13±0.11	81.97±0.11	80.89±0.25	80.02±0.65	75.24±0.17	76.28±6.37	83.99±0.04	83.76±0.68	**85.50±0.93**
Acc	99.36±0.03	99.41±0.04	99.38±0.01	98.33±1.00	99.20±0.00	99.29±0.17	99.51±0.00	99.50±0.02	**99.56±0.02**
Spe	99.43±0.03	99.55±0.03	99.45±0.01	98.93±0.45	99.28±0.01	99.69±0.06	99.77±0.00	99.76±0.01	**99.79±0.00**
Sen	86.28±0.72	85.97±0.11	86.71±0.18	**86.98±0.40**	84.41±0.49	73.84±7.48	82.76±0.02	83.03±0.59	84.34±1.36
Avg	86.91±0.23	87.64±0.11	87.28±0.14	86.70±0.60	84.17±0.17	82.24±4.49	87.69±0.03	87.62±0.46	**88.78±0.74**

在前4项评价指标中均表现最佳,提升幅度分别为2.28、1.51、0.05和0.02个百分点. 上述实验结果表明:CVMBNet在两个不同医学图像数据集上均展现出良好的泛化能力和稳定性,但是在病灶的敏感度上仍然有提升的空间. 进一步地,实验结果表明:本文提出的方法在4个数据集上的Avg指标均优于其他方法,分别达到91.51%、91.74%、91.30%和88.78%,充分验证了该方法在不同数据集上的有效性和稳定性.

值得注意的是,在4个数据集的实验结果中,次优的结果主要集中在VM-UNet、VM-UNet V2和H-vmunet中,这些模型均是基于SSM架构构建的,这从侧面验证了SSM在医学图像分割任务上的可行性,以及本文使用SSM进行全局特征提取策略的合理性. 此外,如图4所示,CVMBNet的计算复杂度为4.69 G,参数量为2.99 M,相较于其他方法,CVMBNet在保持较好的分割性能的同时,有着更高的计算效率,更适合于资源受限的临床场景.

5.2 定性结果

图5~图8展示了所提方法及对比方法在4个数据集上的定性结果,其中:(*a*)表示原始图像,(*b*)表示分

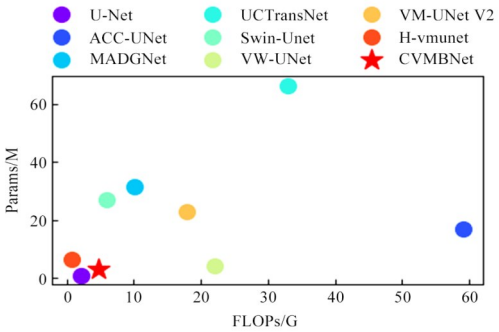
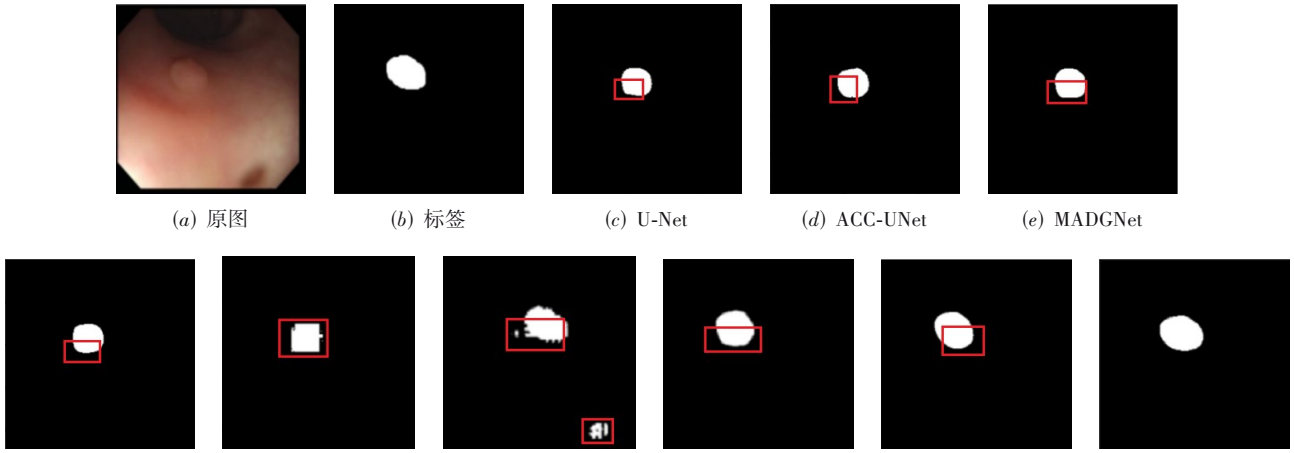


图 4　模型间 Params 和 FLOPs 对比

割图像的真实标签,(*k*)表示本文所提方法的分割结果,(*c*)~(*j*)表示其他对比方法的分割结果. 定性结果表明:本文提出方法能够准确区分目标组织与周围的区域并生成边界清晰、结构完整的分割结果,有效避免了现有方法在边缘模糊或过度分割上存在的问题. 具体而言,如图 5(*a*)所示,在 CVC-ColonDB 数据集上,由于息肉区域与邻近非息肉区域在结构和纹理特征上的相似性,现有方法出现了将非息肉区域误判为息肉区域,导致分割结果出现息肉区域尺寸不一致[如图 5(*c*)~图 5(*f*)]、轮廓模糊[如图 5(*g*)]或者假阳性[如图 5(*h*)]等现象. 相比之下,本文提出的模型能够同时关注息肉区域与非息肉区域在局部细节以及整体结构上的差异,从而有效降低因区域相似带来的误分割的风险. 如图 6(*a*)和图 7(*a*)所示,在 ISIC2017 和 ISIC2018 数据集上,由于皮肤病区域边界通常表现出不规则、模糊的特性,现有方法在边缘分割的精度不高,出现了边界过于平滑[如图 6(*d*)和图 7(*j*)]或者锯齿状[如图 6(*g*)和图 7(*h*)]的情况. 与之相比,本文提出的方法通过 C-VSS 模块实现对边缘信息的自适应特征选择,并通过 MBIA 模块对多分支上下文信息进行交互与融合,提升了模型对复杂边缘特征的表达能力. 同样地,如图 8(*a*)所示,在 COVID-19 数据集上,由于肺部病灶区域边界高度不规则、正常区域与肺部病灶区域的对比度较低,现有方法出现了病灶区域粘连、将邻近血管误判为病灶的情况,如图 8(*c*)~图 8(*h*). 本文所提方法通过 C-VSS 模块的局部和全局信息的提取以及 MBIA 模块对跨层级特征的双向信息传递,可以在保持清晰的分割边界的同时,降低肺部病灶区域的误判率.



(*a*) 原图　(*b*) 标签　(*c*) U-Net　(*d*) ACC-UNet　(*e*) MADGNet

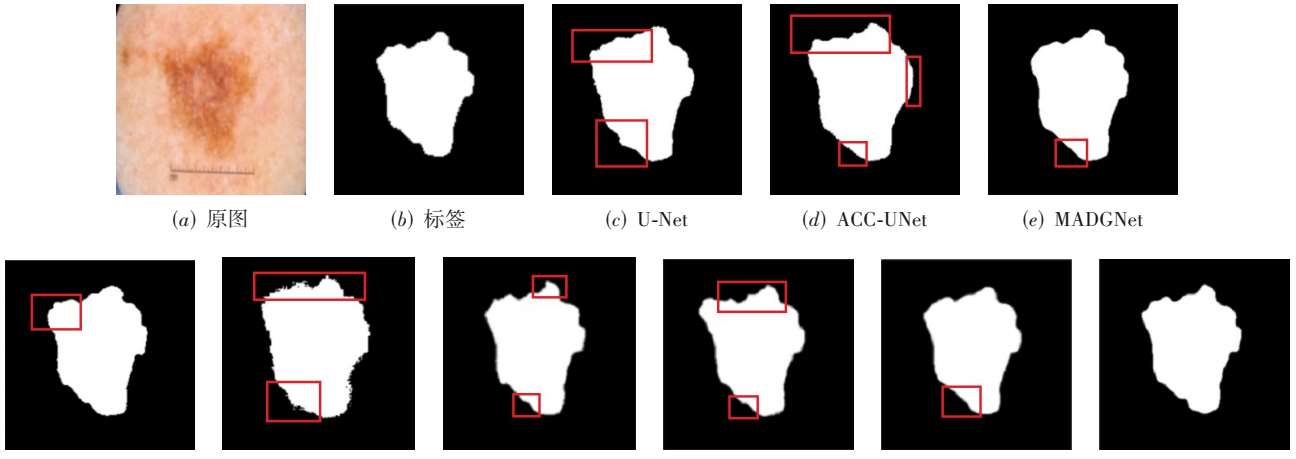
(*f*) UCTransNet　(*g*) Swin-Unet　(*h*) VM-UNet　(*i*) VM-UNet V2　(*j*) H-vmunet　(*k*) CVMBNet(Ours)

图 5　CVC-ColonDB 数据集的定性结果

上述定量与定性结果表明:本文提出的方法在 4 个公开数据集上优于当前的主流方法,取得良好的分割性能. 具体而言,C-VSS 模块通过双分支协作策略,实现了局部细节特征与全局上下文信息的有效融合,进而提升了模型对模糊边界的识别能力;同时,MBIA 模块引入多分支注意力机制,通过跨尺度特征信息的传递与融合,增强了模型对复杂病灶形态的建模能力.

5.3　消融实验

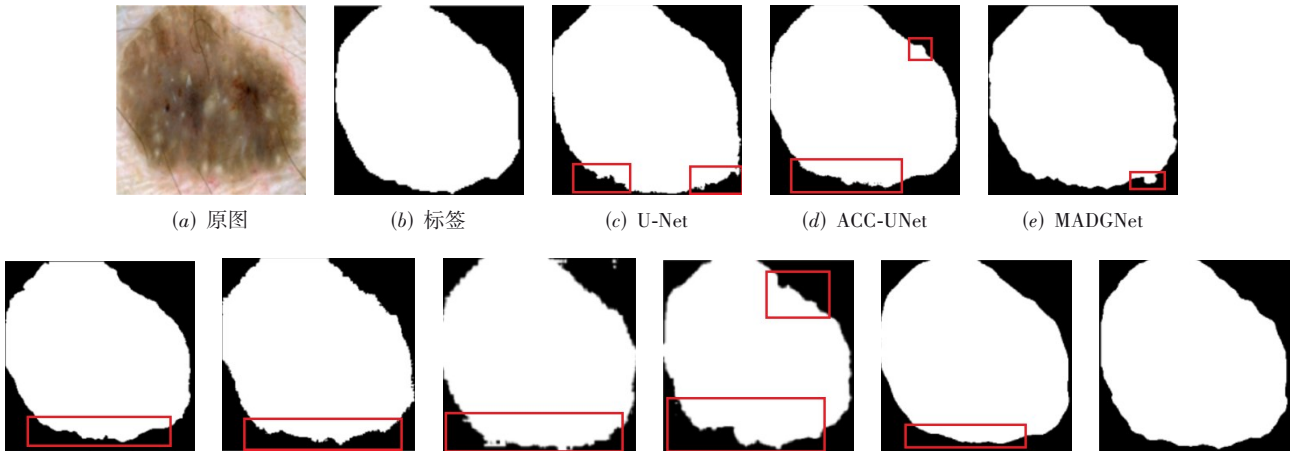
为了进一步验证所提方法的有效性,本文通过逐步引入不同模块设计了一系列的消融实验,其结果如图 9 所示. 其中,蓝色柱状图表示 U-Net(with Conv),即没有使用 C-VSS 和 MBIA 连接模块的消融结果;橙色柱状图表示 U-Net(with MBIA),即只使用 MBIA 连接模块的消融结果;黄色柱状图对应 U-Net(with C-VSS),即只使用 C-VSS 模块的消融结果;绿色柱状图则表示完整的 CVMBNet,即同时使用 C-VSS 和 MBIA 两个模块的实验结果. 如图 9 所示,本文提出的 C-VSS 模块和 MBIA 模块在消融实验中均表现出明显的性能提升效果,具体表现如下:C-VSS 模块的引入增强了模型在局部细节特征和全局上下文特征的提取能力,从而提升了模型整体分割性能. 而 MBIA 模块的引入进一步增强了多尺



(*a*) 原图 (*b*) 标签 (*c*) U-Net (*d*) ACC-UNet (*e*) MADGNet

(*f*) UCTransNet (*g*) Swin-Unet (*h*) VM-UNet (*i*) VM-UNet V2 (*j*) H-vmunet (*k*) CVMBNet(Ours)

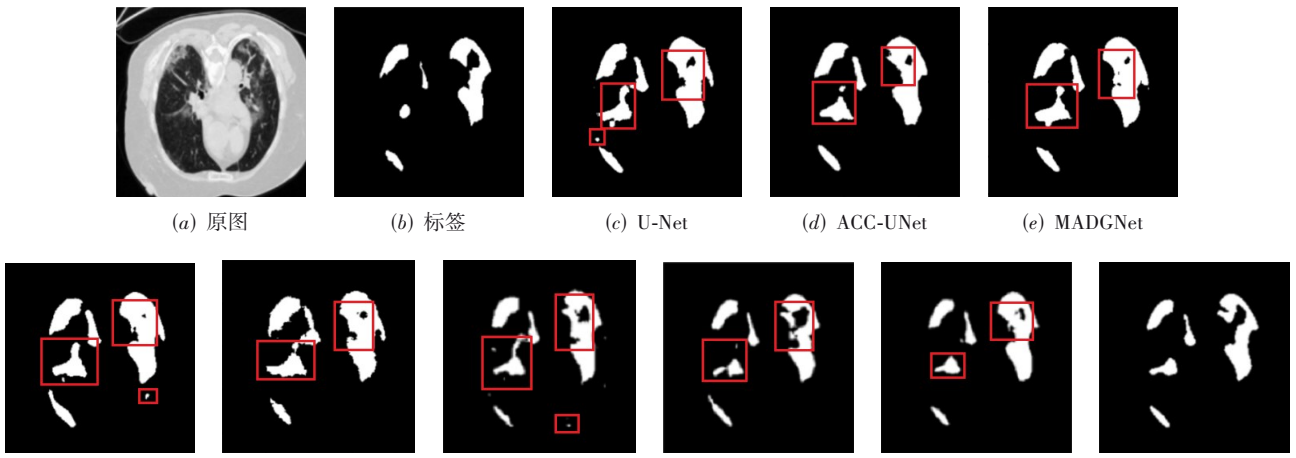
图6 ISIC2017数据集的定性结果



(*a*) 原图 (*b*) 标签 (*c*) U-Net (*d*) ACC-UNet (*e*) MADGNet

(*f*) UCTransNet (*g*) Swin-Unet (*h*) VM-UNet (*i*) VM-UNet V2 (*j*) H-vmunet (*k*) CVMBNet(Ours)

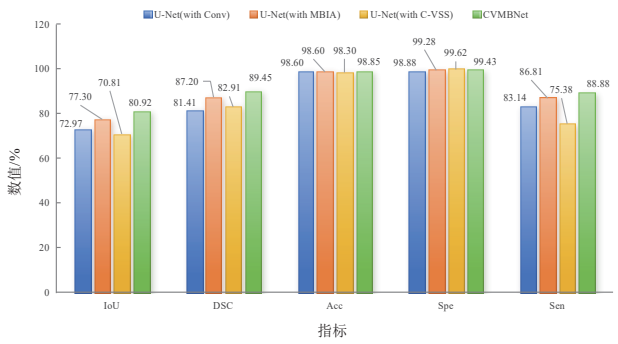
图7 ISIC2018数据集的定性结果



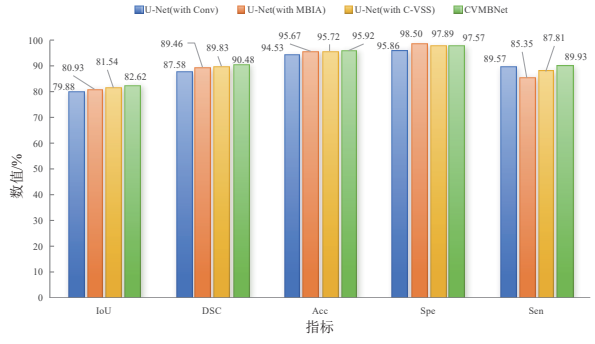
(*a*) 原图 (*b*) 标签 (*c*) U-Net (*d*) ACC-UNet (*e*) MADGNet

(*f*) UCTransNet (*g*) Swin-Unet (*h*) VM-UNet (*i*) VM-UNet V2 (*j*) H-vmunet (*k*) CVMBNet(Ours)

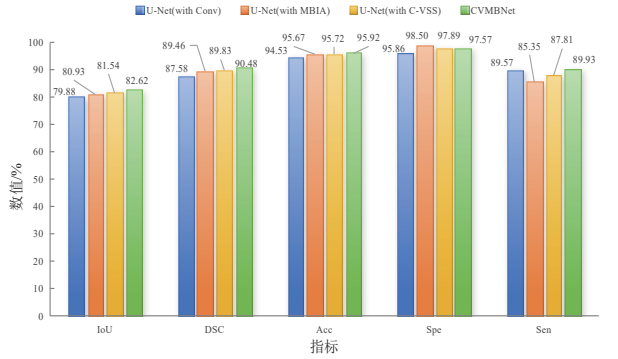
图8 COVID-19数据集的定性结果



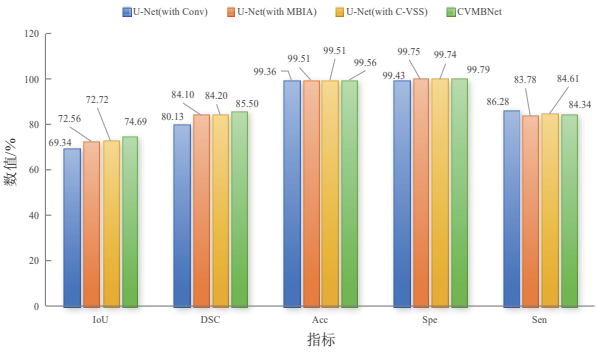
(*a*) CVC-ColonDB数据集的消融结果



(*b*) ISIC2017数据集的消融结果



(*c*) ISIC2018数据集的消融结果



(*d*) COVID-19数据集的消融结果

图9　消融实验结果

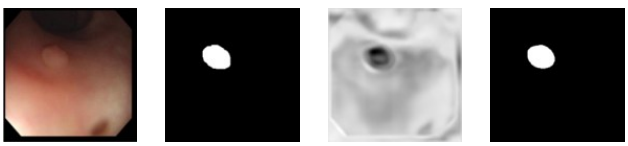
度特征间的交互和融合,无论是在U-Net(with MBIA)还是CVMBNet,均表现出更强的特征表示能力.具体而言,在CVC-ColonDB数据集上,相较于U-Net(with Conv)的消融结果,U-Net(with MBIA)在5个评价指标上分别提升了4.33、5.79、0.00、0.40和3.67个百分点,而相较于U-Net(with C-VSS),CVMBNet模型则分别提升了10.11、6.54、0.55、−0.19和13.50个百分点.与此同时,相较于U-Net(with Conv),U-Net(with C-VSS)在DSC和Spe指标上分别提升了1.50和0.74个百分点.这一结果表明:在该数据集上,相较于C-VSS模块,MBIA模块在该数据集的作用更明显.值得注意的是,CVMBNet在Spe指标上出现0.19个百分点的下降,其潜在的原因是结肠镜图像中息肉与非息肉区域纹理非常相似,导致模型对负类样本的识别能力受到一定影响.在ISIC2017数据集上,与U-Net(with Conv)和U-Net(with C-VSS)的消融结果相比,U-Net(with MBIA)和CVMBNet模型分别在前4个指标上有不同程度上的提升,其中U-Net(with MBIA)提升了3.23、3.62、0.19、1.59个百分点,而CVMBNet则分别提升了0.50、0.30、0.11、0.10个百分点.此外,相较于U-Net(with Conv),U-Net(with C-VSS)在前4个指标上分别提升了4.88、4.61、0.53和1.50个百分点.上述结果表明:本文所提MBIA模块在一定程度上提高了模型对边界识别的能力.同时,U-Net(with C-VSS)相较于U-Net(with Conv)的提升进一步验证了C-VSS模块在局部细节与全局上下文信息融合方面的有效性,同时也说明所提方法在处理病灶边界模糊、结构复杂等挑战性场景中具备良好的适应性和稳定性.在ISIC2018数据集上,相较于U-Net(with Conv)的消融结果,U-Net(with MBIA)在IoU、DSC、Acc和Spe指标上分别提升了1.05、1.88、1.14和2.64个百分点,而与U-Net(with C-VSS)相比,CVMBNet在IoU、DSC、Acc和Sen指标上分别提升了1.08、0.65、0.20和2.12个百分点.与此同时,相较于U-Net(with Conv),U-Net(with C-VSS)在IoU、DSC、Acc和Spe指标上分别提升了1.66、2.25、1.19和2.03个百分点.这一结果表明:在该数据集上,C-VSS模块与MBIA模块展现出一定的互补性,C-VSS模块通过局部和全局特征的协同提取,降低了模型在分割过程中的假阳性;而MBIA模块则通过多分支特征的交互与融合,增强了模型对复杂病灶区域的特征的表达能力,从而共同推动整体性能的提升.在COVID-19数据集上,从消融结果来看,相较于U-Net(with Conv),U-Net(with MBIA)在IoU和DSC指标上分别提升了3.22和3.97个百分点,而相较于U-Net(with C-VSS)的消融结果,CVMBNet则分别提升了1.97和1.30个百分点;此外,相较于

U-Net(with Conv),U-Net(with C-VSS)在IoU和DSC指标上分别提升了3.38和4.07个百分点.上述结果表明:本文所提模块能够有效增强模型对病灶区域的定位精度以及对复杂边界的识别能力.进一步分析显示,相较于U-Net(with Conv)与U-Net(with C-VSS),U-Net(with MBIA)与CVMBNet在Acc和Spe指标上分别提升了0.15、0.32和0.05、0.05个百分点;而U-Net(with C-VSS)相比U-Net(with Conv)在Acc和Spe上也分别提升了0.15和0.31个百分点.这说明:所提方法不仅在病灶边界分割方面具有优势,还在整体分类准确性与特异性方面表现出良好的性能提升效果,进一步验证了模型在低对比度、高复杂度医学图像分割任务中的鲁棒性与实用性.此外,为了进一步验证C-SS2D模块的性能,本文将其核心组件C-SS2D结构替换为标准的SS2D结构进行实验.然而,在训练过程中发现模型无法正常反向传播,导致训练无法继续进行.这一现象表明:C-SS2D结构在保证网络训练的稳定性和特征表达的完整性起到至关重要的作用.

综上所述,本文提出的C-VSS和MBIA模块在多个医学图像分割任务中均表现出良好的性能.实验结果表明:C-VSS模块通过局部与全局特征的协同提取,有效降低了假阳性率;MBIA模块通过增强多尺度特征间的交互与融合,提升了模型的整体表达能力.上述结果充分验证了所提方法在不同模态医学图像分割中的有效性与鲁棒性.

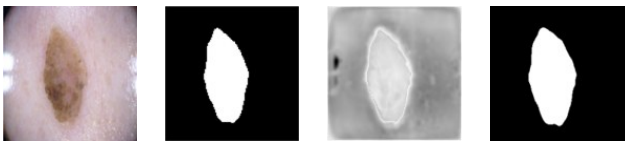
5.4 可解释性实验

为了更直观地评估模型的性能,本文通过纹理图可视化模型内部的注意力分布区域,结果如图10~图13所示,其中:(*a*)表示原图,(*b*)表示原图的分割标签,(*c*)表示纹理图,(*d*)表示本文分割结果.可以看出,尽管不同模态的图像在形态上存在差异,本文模型通过局部和全局特征的协同提取与融合,能够有效地恢复图像中的纹理信息.进一步分析(*c*)部分的纹理图可以看出,本文模型对病灶区域的纹理重建比其他区域更为清晰,这表明本模型不仅能够有效恢复图像的整体纹理信息,还特别强调对病灶区域的关注.这种特性有助于提高病灶区域的识别精度和边界分割质量,从而进一步验证了本文所提方法的有效性.



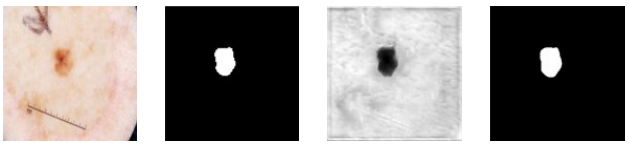
(*a*) 原图 (*b*) 分割标签 (*c*) 纹理图 (*d*) CVMBNet

图10 CVC-ColonDB数据集纹理图



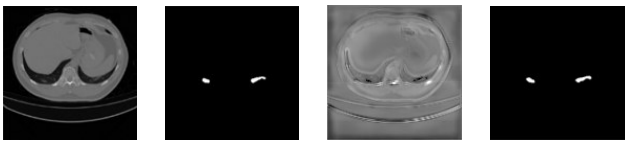
(*a*) 原图 (*b*) 分割标签 (*c*) 纹理图 (*d*) CVMBNet

图11 ISIC2017数据集纹理图



(*a*) 原图 (*b*) 分割标签 (*c*) 纹理图 (*d*) CVMBNet

图12 ISIC2018数据集纹理图



(*a*) 原图 (*b*) 分割标签 (*c*) 纹理图 (*d*) CVMBNet

图13 COVID-19数据集纹理图

6 结论

医学图像分割任务因其结构复杂、边界模糊及噪声干扰等特性,对算法的精度与效率提出了较高的要求.尽管当前基于CNN与Transformer等架构的方法在局部特征提取与全局上下文建模方面取得了一定进展,但仍面临计算复杂度高、边缘分割精度不足等问题.为应对上述挑战,本文提出一种融合C-VSS与MBIA的医学图像分割方法.其中,C-VSS模块通过双分支协作策略实现局部和全局信息的提取与融合,提升了SSM在医学图像分割中的适用性;MBIA模块则通过构建编码器与解码器之间的双向信息交互机制,实现多层级特征的动态融合,增强了模型对多尺度信息的表达能力.为验证所提方法的有效性,在4个公开医学图像数据集上的实验结果表明:所提方法在分割精度与计算效率方面均优于当前主流方法,展现出良好的性能优势.未来工作将聚焦于该模型在三维医学图像分割任务中的扩展与优化.

参考文献

[1] XIAO H G, LI L, LIU Q Y, et al. Transformers in medical image segmentation: A review[J]. Biomedical Signal Processing and Control, 2023, 84: 104791.

[2] 吴玉超, 林岚, 王婧璇, 等. 基于卷积神经网络的语义分割在医学图像中的应用[J]. 生物医学工程学杂志, 2020, 37(3): 533-540.
WU Y C, LIN L, WANG J X, et al. Application of seman-

tic segmentation based on convolutional neural network in medical images[J]. Journal of Biomedical Engineering, 2020, 37(3): 533-540. (in Chinese)

[3] MEIBURGER K M, ACHARYA U R, MOLINARI F. Automated localization and segmentation techniques for B-mode ultrasound images: A review[J]. Computers in Biology and Medicine, 2018, 92: 210-235.

[4] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. (2023-08-02) [2025-07-21]. https://arxiv.org/abs/1706.03762.

[5] 贾熹滨, 郭雄, 王珞, 等. 一种迭代边界优化的医学图像小样本分割网络[J]. 自动化学报, 2024, 50(10): 1988-2001.
JIA X B, GUO X, WANG L, et al. A small sample segmentation network for medical images based on iterative boundary optimization[J]. Acta Automatica Sinica, 2024, 50(10): 1988-2001. (in Chinese)

[6] 刘金平, 吴娟娟, 张荣, 等. 基于结构重参数化与多尺度深度监督的COVID-19胸部CT图像自动分割[J]. 电子学报, 2023, 51(5): 1163-1171.
LIU J P, WU J J, ZHANG R, et al. Toward automated segmentation of COVID-19 chest CT images based on structural reparameterization and multi-scale deep supervision[J]. Acta Electronica Sinica, 2023, 51(5): 1163-1171. (in Chinese)

[7] NAM J H, SYAZWANY N S, KIM S J, et al. Modality-agnostic domain generalizable medical image segmentation by multi-frequency in multi-scale attention[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 11480-11491.

[8] RAYED M E, SAJIBUL ISLAM S M, NIHA S I, et al. Deep learning for medical image segmentation: State-of-the-art advancements and challenges[J]. Informatics in Medicine Unlocked, 2024, 47: 101504.

[9] WANG H N, CAO P, WANG J Q, et al. UCTransNet: Rethinking the skip connections in U-Net from a channel-wise perspective with transformer[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(3): 2441-2449.

[10] 周新民, 熊智谋, 史长发, 等. 基于多尺度卷积调制的医学图像分割[J]. 电子学报, 2024, 52(9): 3159-3171.
ZHOU X M, XIONG Z M, SHI C F, et al. Medical image segmentation based on multi-scale convolution modulation[J]. Acta Electronica Sinica, 2024, 52(9): 3159-3171. (in Chinese)

[11] XUE W, CHEN C H, QI X, et al. M2ANet: Multi-branch and multi-scale attention network for medical image segmentation[J]. Chinese Physics B, 2025, 34(8): 080703.

[12] 雷涛, 张峻铭, 杜晓刚, 等. 基于混洗特征编码与门控解码的医学图像分割网络[J]. 电子学报, 2024, 52(12): 4142-4152.
LEI T, ZHANG J M, DU X G, et al. Medical image segmentation network based on shuffled feature encoding and gated decoding[J]. Acta Electronica Sinica, 2024, 52(12): 4142-4152. (in Chinese)

[13] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer, 2015: 234-241.

[14] 宋艳涛, 路云里 . SwinT-Unet: 基于双通道自注意力机制的超声图像分割方法[J]. 电子学报, 2024, 52(11): 3835-3846.
SONG Y T, LU Y L. SwinT-unet: Ultrasound image segmentation based on two-channel self-attention mechanism[J]. Acta Electronica Sinica, 2024, 52(11): 3835-3846. (in Chinese)

[15] GU A, GOEL K, RÉ C. Efficiently modeling long sequences with structured state spaces[EB/OL]. (2022-08-05)[2025-07-21]. https://arXiv.org/abs/2111.00396.

[16] ZHU Q F, FANG Y, CAI Y Z, et al. Rethinking scanning strategies with vision mamba in semantic segmentation of remote sensing imagery: An experimental study[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 18223-18234.

[17] GU A, DAO T. Mamba: Linear-time sequence modeling with selective state spaces[EB/OL]. (2023-12-01) [2025-07-21]. https://arXiv.org/abs/2312.00752.

[18] ZHU L H, LIAO B C, ZHANG Q, et al. Vision mamba: Efficient visual representation learning with bidirectional state space model[EB/OL]. (2024-11-14) [2025-08-30]. https://arXiv.org/abs/2401.09417.

[19] LIU Y, TIAN Y J, ZHAO Y Z, et al. VMamba: Visual state space model[EB/OL]. (2024-12-29) [2025-07-21]. https://arXiv.org/abs/2401.10166.

[20] LI G J, HUANG Q H, WANG W, et al. Selective and multi-scale fusion Mamba for medical image segmentation[J]. Expert Systems with Applications, 2025, 261: 125518.

[21] IBTEHAZ N, KIHARA D. ACC-UNet: A completely convolutional UNet model forthe2020s[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2023. Cham: Springer, 2023: 692-702.

[22] WANG Z H, MIN X K, SHI F Y, et al. SMESwin Unet: Merging CNN and transformer for medical image segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2022. Cham: Springer, 2022: 517-526.

[23] WANG Z Y, ZHENG J Q, ZHANG Y C, et al. Mamba-UNet: UNet-like pure visual mamba for medical image segmentation[EB/OL]. (2024-03-30)[2025-08-30]. https://arXiv.org/abs/2402.05079.

[24] YAO W J, BAI J J, LIAO W, et al. From CNN to transformer: A review of medical image segmentation models[J]. Journal

of Imaging Informatics in Medicine, 2024, 37(4): 1529-1547.
[25] ZHAO X Q, JIA H P, PANG Y W, et al. M^{2}SNet: Multi-scale in multi-scale subtraction network for medical image segmentation[EB/OL]. (2025-09-30)[2025-07-21]. https://arXiv.org/abs/2303.10894.
[26] RAHMAN M M, MUNIR M, MARCULESCU R. EMCAD: Efficient multi-scale convolutional attention decoding for medical image segmentation[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 11769-11779.
[27] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 9992-10002.
[28] CAO H, WANG Y Y, CHEN J, et al. Swin-unet: Unet-like pure transformer for medical image segmentation[C]// Computer Vision - ECCV 2022 Workshops. Cham: Springer, 2023: 205-218.
[29] RAHMAN M M, MARCULESCU R. Medical image segmentation via cascaded attention decoding[C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2023: 6211-6220.
[30] ZHANG Y H, BALESTRA G, ZHANG K, et al. MultiTrans: Multi-branch transformer network for medical image segmentation[J]. Computer Methods and Programs in Biomedicine, 2024, 254: 108280.
[31] ZHONG J H, TIAN W H, XIE Y L, et al. PMFSNet: Polarized multi-scale feature self-attention network for lightweight medical image segmentation[J]. Computer Methods and Programs in Biomedicine, 2025, 261: 108611.
[32] MA J, LI F F, WANG B. U-mamba: Enhancing long-range dependency for biomedical image segmentation[EB/OL]. (2024-01-09)[2025-08-30]. https://arXiv.org/abs/2401.04722.
[33] RUAN J C, LI J C, XIANG S C. VM-UNet: Vision mamba UNet for medical image segmentation[EB/OL]. (2024-11-08)[2025-07-21]. https://arXiv.org/abs/2402.02491.
[34] ZHANG M Y, YU Y, JIN S, et al. VM-UNET-V2: Rethinking vision mamba UNet for medical image segmentation[C]//Bioinformatics Research and Applications. Singapore: Springer, 2024: 335-346.
[35] LIU J R, YANG H, ZHOU H Y, et al. Swin-UMamba: Mamba-based UNet with ImageNet-based pretraining[C]// Medical Image Computing and Computer Assisted Intervention-MICCAI 2024. New York: ACM, 2024: 615-625.
[36] WU R K, LIU Y H, LIANG P C, et al. H-vmunet: High-order Vision Mamba UNet for medical image segmentation[J]. Neurocomputing, 2025, 624: 129447.

作者简介

薛 伟 男,1986年11月出生于江苏省南通市. 现为安徽工业大学计算机科学与技术学院副院长、副教授、博士生导师. 主要研究方向为机器学习、计算机视觉、数据挖掘. 中国电子学会会员编号:E190188441M.
E-mail: xuewei@ahut.edu.cn

陈创慧 女,2000年9月出生于广东省茂名市. 现为安徽工业大学计算机科学与技术学院硕士研究生. 主要研究方向为医学图像分割.
E-mail: chenchuanghui16@foxmail.com

杜明洋 男,1994年7月出生于安徽省蚌埠市. 现为国防科技大学电子对抗学院讲师. 主要研究方向为雷达智能感知与对抗. 中国电子学会会员编号:E190087642M.
E-mail: dumingyang17@nudt.edu.cn

钟 平 男,1979年6月出生于四川省内江市. 现为国防科技大学电子科学学院研究员、博士生导师. 主要研究方向为计算机视觉、机器学习、模式识别.
E-mail: zhongping@nudt.edu.cn

郑 啸 男,1975年11月出生于福建省莆田市. 现为安徽工业大学副校长、教授、博士生导师. 主要研究方向为工业互联网、群智感知网络、数据隐私保护.
E-mail: xzheng@ahut.edu.cn