



SDA U-Mamba: 基于频域动态特征融合与双极路由注意力的医学图像分割

彭晨阳, 何立风, 王梦溪, 杜晓刚, 王营博, 路艳, 雷涛

引用本文:

彭晨阳, 何立风, 王梦溪, 等. SDA U-Mamba: 基于频域动态特征融合与双极路由注意力的医学图像分割[J]. *智能系统学报*, 2026, 21(1): 284-294.

PENG Chenyang, HE Lifeng, WANG Mengxi, et al. SDA U-Mamba: spectral-domain dynamic fusion and bipolar routing attention for medical image segmentation[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(1): 284-294.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202508032>

您可能感兴趣的其他文章

差异特征融合的无监督SAR图像变化检测

Unsupervised SAR image change detection based on difference feature fusion

智能系统学报. 2021, 16(3): 595-604 <https://dx.doi.org/10.11992/tis.202103011>

利用混合高斯和拓扑结构的人体“鬼影”抑制算法

Human “ghost” suppression algorithm using Gaussian mixture model and topology

智能系统学报. 2021, 16(2): 294-302 <https://dx.doi.org/10.11992/tis.201912030>

结合模糊特征检测的鲁棒核相关滤波跟踪法

Robust KCF tracking algorithm combined with fuzzy feature detection

智能系统学报. 2021, 16(2): 323-329 <https://dx.doi.org/10.11992/tis.201912010>

基于语义分割的简洁线条肖像画生成方法

Concise line portrait generation method based on semantic segmentation

智能系统学报. 2021, 16(1): 134-141 <https://dx.doi.org/10.11992/tis.202101003>

多特征融合的异视角目标关联算法

Target association from different perspectives based on multi-feature fusion

智能系统学报. 2020, 15(5): 847-855 <https://dx.doi.org/10.11992/tis.202006037>

基于卷积神经网络的遥感图像分类研究

Classification of remote-sensing image based on convolutional neural network

智能系统学报. 2018, 13(4): 550-556 <https://dx.doi.org/10.11992/tis.201706078>

DOI: 10.11992/tis.202508032

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20251219.1353.003>

SDA U-Mamba: 基于频域动态特征融合与双极路由注意力的医学图像分割

彭晨阳^{1,2}, 何立风¹, 王梦溪^{1,2}, 杜晓刚^{1,2}, 王莹博^{1,2}, 路艳³, 雷涛^{1,2}

(1. 陕西科技大学电子信息与人工智能学院, 陕西西安 710021; 2. 陕西科技大学陕西省人工智能联合实验室, 陕西西安 710021; 3. 陕西科技大学轻工科学与工程学院, 陕西西安 710021)

摘要:以 Mamba 为代表的状态空间模型 (state space models, SSM) 凭借其出色的长程依赖建模能力与较低的线性计算复杂度, 在医学图像分割领域展现出广阔的应用前景。但该方法对图像逐像素展平的处理方式会破坏图像空间结构, 导致局部细节提取不足, 且缺乏对器官与病灶的聚焦机制, 在复杂背景下易引入冗余背景信息。为解决上述问题, 本文提出了一种融合谱域动态特征与注意力机制的医学图像分割网络——频域动态注意力 U 型 Mamba (spectral dynamic attention U-Mamba, SDA U-Mamba)。该网络采用分层 U 型结构设计, 分别从空间连续性建模和区域聚焦能力两方面对 Mamba 进行优化。在网络浅层设计 Mamba 空频注意力模块, 通过融合空域卷积、频域变换和金字塔自注意力结构以提升模型的局部信息感知能力与多尺度上下文建模效果; 在网络深层引入双极路由注意力模块, 通过动态路由选择与稀疏激活机制增强模型对医学图像器官或病灶的表征。实验结果表明, SDA U-Mamba 在 BUSI、CVC-ClinicDB 与 CHAOS-Liver 这 3 个公开医学数据集上的分割性能显著优于当前主流方法, 平均交并比 (intersection over union, IoU) 提升 2.61%。本文所提算法可用于临床医学图像分割。

关键词: Mamba; 医学图像分割; 状态空间模型; 谱动态特征融合; 傅里叶变换; 注意力; U-Net 架构; 多尺度特征建模
中图分类号: TP391.4 文献标志码: A 文章编号: 1673-4785(2026)01-0284-11

中文引用格式: 彭晨阳, 何立风, 王梦溪, 等. SDA U-Mamba: 基于频域动态特征融合与双极路由注意力的医学图像分割 [J]. 智能系统学报, 2026, 21(1): 284-294.

英文引用格式: PENG Chenyang, HE Lifeng, WANG Mengxi, et al. SDA U-Mamba: spectral-domain dynamic fusion and bipolar routing attention for medical image segmentation [J]. CAAI transactions on intelligent systems, 2026, 21(1): 284-294.

SDA U-Mamba: spectral-domain dynamic fusion and bipolar routing attention for medical image segmentation

PENG Chenyang^{1,2}, HE Lifeng¹, WANG Mengxi^{1,2}, DU Xiaogang^{1,2},
WANG Yingbo^{1,2}, LU Yan³, LEI Tao^{1,2}

(1. School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China; 2. Shaanxi Provincial Joint Laboratory of Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China; 3. School of Light Industry Science and Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China)

Abstract: State space models (SSM) represented by Mamba have demonstrated broad application prospects in medical image segmentation owing to their excellent long-range dependency modeling and low linear computational complexity. However, the pixel-wise flattening operation used in these methods destroys the spatial structure of images, leading to insufficient extraction of local details. Additionally, they lack a mechanism to focus on organs and lesions, making them prone to introducing redundant background information in medical images with complex contexts. To address these issues, this study proposes spectral dynamic attention U-Mamba (SDA U-Mamba), a medical image segmentation network that integrates spectral-domain dynamic features with attention mechanisms. The network adopts a hierarchical U-shaped architecture, optimizing Mamba in two aspects: spatial continuity modeling and regional focusing capability. Specifically, a Mamba Spatial-Frequency Attention module is introduced into the network's shallow layers. This module enhances the model's local information perception and multi-scale context modeling performance by fusing spatial convolution, frequency-domain transformation, and a pyramid self-attention structure. Furthermore, a Bipolar Routing Attention module is introduced in the deep layers, strengthening the network's representation of organs and lesions in medical images through dynamic routing selection and sparse activation mechanisms. Experimental results show that SDA U-Mamba achieved significantly better segmentation performance than current state-of-the-art methods across three public medical datasets (BUSI, CVC-ClinicDB, and CHAOS-Liver), with an average improvement in intersection over union (IoU) of 2.61%. The results indicate that the proposed algorithm can be applied to clinical medical image segmentation.

Keywords: Mamba; medical image segmentation; state space model; spectral-dynamic feature fusion; Fourier transform; attention mechanism; U-Net architecture; multi-scale feature modeling.

收稿日期: 2025-08-28. 网络出版日期: 2025-12-22.

基金项目: 国家自然科学基金项目 (62271296; 62201334); 陕西省创新能力支撑计划项目 (2025RS-CXTD-012); 陕西省重点研发计划项目 (2024GX-YBXM-121); 西安市中青年科技创新领军人才项目 (25ZQRC00019); 陕西省教育厅科学研究计划项目 (23JJP014; 23JJP022).

通信作者: 雷涛. E-mail: leitao@sust.edu.cn.

随着视觉表征学习的不断发展, 卷积神经网络 (convolutional neural network, CNN)^[1-4] 和视觉 Transformer^[5-7] 已成为计算机视觉领域的两类主流框架。CNN 通过局部感受野的层级堆叠有效

提取空间特征,但其固定尺度的卷积核在建模全局上下文关系方面存在固有局限。相比之下,ViT利用自注意力机制实现跨空间范围的全局特征交互,显著提升了长距离依赖建模能力,然而,其注意力矩阵的计算复杂度呈平方增长,导致模型在处理高分辨率图像时面临较高的计算负担。这种互补性的架构特点促使研究者探索能够兼顾全局建模能力与计算效率的新型网络范式。

近年来,以 Mamba 为代表的状态空间模型(state space models, SSM) 凭借其在长程依赖建模方面的优势以及线性计算复杂度的特性,受到了研究人员的广泛关注。但现有 Mamba 系列分割方法仍存在明显局限:大多数网络虽通过调整结构适配图像空间特性,却多局限于空域单维度优化,既未针对医学图像的关键边缘纹理开展频域信息建模,也缺乏对器官、病灶的显式聚焦机制,难以有效规避背景冗余信息干扰。为此,本文提出一种融合谱域动态特征与注意力机制的医学图像分割网络—频域动态注意力 U 型 Mamba(spectral dynamic attention U-Mamba, SDA U-Mamba), 该网络能够增强模型对二维空间结构的保持能力及对病灶或器官的聚焦能力。具体而言,其通过空频协同建模弥补现有 Mamba 方法频域信息缺失的短板,借助双极路由注意力机制实现对病灶与器官的主动聚焦,显著区别于现有 Mamba 方法的全局无差别建模与单一空域优化思路。在乳腺肿瘤、息肉和腹部器官 3 个分割任务中的实验结果表明,所提方法在多个性能指标上均优于现有主流方法。本文的主要贡献如下:

1) 针对 Mamba 在医学图像分割中存在的空间结构断裂、频域信息缺失的问题,提出了用于 Mamba 的空频注意力模块。该模块以视觉状态空间模块(visual state-space block, VSS-Block)为核心,通过引入谱动态特征融合模块(spectral dynamic feature fusion, SDFF)增强模型对图像中边缘纹理和频域信息的建模能力,同时嵌入自注意力金字塔池化模块(self-attention pyramid pooling, SAPP)缓解 U 型网络上下采样过程中的语义不匹配问题,进而提升模型的多尺度上下文建模能力。

2) 针对 Mamba 在医学图像分割中缺乏对器官或病灶聚焦建模、易引入冗余背景信息的问题,本文在 U 型网络深层引入双极路由注意力模块(bipolar routing attention, BRA)。该模块通过构建动态稀疏注意力路径,实现特征之间的选择性激活与抑制,从而有效增强模型对图像中器官或病灶的语义响应能力,削弱无关背景对分割结果的干扰。

3) 基于频域动态特征融合与双极路由注意力建模,本文提出了面向医学图像分割的 SDA U-Mamba 网络,该网络不仅能够捕捉医学图像的局部细节信息,而且能够有效建模医学图像的长距离上下文信息,实现对医学图像的精准分割。实验结果显示,该方法在 BUSI、CVC-ClinicDB 和 CHAOS-Liver 数据集上平均交并比(intersection over union, IoU)分别高达 65.73%、86.46% 和 96.67%,较次优方法平均提升了 2.61%。

1 状态空间模型及医学图像分割相关工作

1.1 状态空间模型

SSM 作为一种描述系统动态行为的数学框架,长期以来由于其较高的计算复杂度与内存开销,难以在实践中广泛应用。为解决这一瓶颈问题,S4^[8]创新性地对系统矩阵进行正态低秩参数化,并利用伍德伯里恒等式重构卷积计算过程,成功将其计算复杂度降至线性水平,显著降低了模型的训练成本。在此基础上,S6^[9]提出了输入依赖的动态参数替代静态结构约束,同时结合硬件友好的并行扫描算法,将理论上的线性复杂度有效转化为 GPU 上的实际高效计算,使其训练速度较 S4 提升 3 倍以上。Mamba^[9]则进一步基于 S6,提出了选择性状态空间机制,通过动态稀疏化策略降低冗余状态传递,同时设计了内存优化方案,缓解了中间状态存储带来的瓶颈,显著提升了长序列建模的效率,这一系列改进为 Mamba 在视觉任务中的广泛应用奠定了坚实基础。

随着 Mamba 相关技术的快速发展,研究者开始将其引入视觉建模领域,并针对不同视觉任务展开优化与适配。VMamba^[10]作为首个适配二维视觉任务的 Mamba 架构,通过引入双向扫描机制与跨窗口状态共享策略,有效缓解了 Mamba 在方向敏感性强及局部与全局特征耦合方面的不足。针对图像分类任务,Vision Mamba^[11]结合双向状态建模与位置编码,提出了具备空间感知能力的 Vim 模块,显著提升了模型对空间结构的理解能力。针对语义分割任务,Pan-Mamba^[12]构建了多尺度特征金字塔,并创新性地采用分层状态共享机制,实现了从局部细节到全局语义信息的有效协同与优化。针对移动端部署,EfficientVMamba^[13]提出了可分离状态空间核与动态通道剪枝策略,在显著降低计算开销和内存占用的同时,有效提升了模型的精度和部署效率。针对医学图像分割任务,Mamba-SEA^[14]结合状态空间模型与边缘增

强机制,通过引入专门的边缘感知模块,强化了模型对解剖结构边界的捕捉能力,同时利用自适应状态调整策略,提升了模型对复杂医学图像细节信息的表征能力。

综上所述,状态空间模型经历了从理论突破到高效应用的关键演进。S4 与 S6 的创新显著提升了计算效率和训练速度,为 Mamba 风格架构的广泛应用奠定了基础。Mamba 通过选择性状态空间机制和内存优化,进一步推动了其在长序列建模中的应用发展。未来,随着深度学习技术与计算机硬件的不断进步,状态空间模型有望在复杂视觉任务中发挥更重要的作用。

1.2 医学图像分割网络

随着深度学习技术的快速发展,医学图像分割领域已涌现出大量的基于深度学习的高效方法。当前主流的医学图像分割方法大致可以分为基于 CNN 的医学图像分割方法、基于 Transformer 的医学图像分割方法和基于 Mamba 的医学图像分割方法 3 类。

基于 CNN 的医学图像分割方法通过局部感受野的卷积运算分层提取图像特征,其典型的编解码器架构利用下采样学习高层语义特征,再通过上采样恢复分辨率,最终结合跳跃连接实现病灶与器官的精准分割。其中,U-Net^[15]作为该范式的开创性模型,首次将这种编解码器结构与跳跃连接机制应用于医学图像分割任务,显著提升了分割精度,并奠定了后续研究的基础。针对 CNN 架构中 U-Net 网络的改进主要从二维医学图像分割和三维医学图像分割两个维度展开。在二维医学图像分割领域,研究聚焦于优化二维卷积网络架构,U-Net++^[16]通过引入嵌套式密集跳跃连接并结合深度监督,有效缩小了编码器与解码器特征之间的语义差距;U-Net3+^[17]构建了全尺度特征金字塔,利用密集特征融合策略增强多分辨率特征间的互补性;SelfReg-UNet^[18]则进一步深挖 UNet 的固有局限,针对编码器与解码器间监督信号不对称导致的语义丢失以及深层特征冗余问题,设计了语义一致性正则化与内部特征蒸馏双机制,以即插即用的方式平衡监督信号并减少冗余特征。而在三维医学图像分割领域,研究致力于设计高效的三维卷积网络以处理体数据,V-Net^[19]率先采用 3D 卷积实现端到端的体数据分割,设计了包含残差连接的编解码器结构,有效增强了模型在三维空间的特征提取能力;在此基础上,DenseVNet^[20]引入了 DenseNet 风格的密集连接结构,强化了特征重用与梯度流动,进

一步提升了模型在三维分割任务中的训练稳定性与性能表现,特别适用于具有复杂结构的多器官分割场景;MedNeXt^[21]则以纯 3D 卷积架构为核心,通过迭代增大卷积核尺寸的策略避免医疗小数据下的性能饱和,结合带残差的编解码块保持跨尺度语义丰富性,在多模态 3D 分割任务中刷新榜单,展现了现代纯卷积架构的潜力。

基于 Transformer 的医学图像分割方法依托注意力机制实现全局特征建模,显著增强了模型对长程依赖关系的捕获能力。现有研究按注意力实现形式主要分为标准自注意力与改进型注意力两类。在标准自注意力中,TransUNet^[22]提出混合架构,将 CNN 提取的局部特征图转化为序列输入预训练好的 ViT 编码器,再与 U-Net 解码器集成以协同局部与全局特征;UNETR^[23]则设计了纯 Transformer 架构,直接采用 ViT 处理图像块序列,并利用跳跃连接将多尺度编码器特征传递至卷积解码器。然而,这类方法在高分辨率图像下面临计算开销大、局部特征表达不足的问题。为此,改进型注意力应运而生。Swin-Unet^[24]构建了基于移位窗口的分层 Swin Transformer 架构,在非重叠局部窗口计算自注意力并实现跨窗口连接;DC-SAU-Net^[25]引入了可变形卷积与双通道空间注意力块,前者通过自适应采样提升形变鲁棒性,后者利用通道注意力与空间注意力加权关键特征;SAT^[26]则通过引入多模态解剖学知识注入与文本提示交互机制,实现了多类别医学目标的通用分割能力,显著降低模型对任务专用设计的依赖。

基于 Mamba 的医学图像分割方法利用结构化状态空间方程对序列数据进行建模,能够以亚线性的计算复杂度高效捕获全局上下文信息,有效解决了传统方法在处理大尺寸医学图像时面临的长距离依赖建模与计算效率之间的平衡难题。根据模型中状态空间模块的构成方式,现有方法主要可分为纯 SSM 架构与混合 SSM 架构两大类。在纯 SSM 架构研究中,VM-UNet^[27]率先构建了基于视觉状态空间模块的 U 形网络,并通过门控选择模块在跳跃连接中动态调制跨层特征流的空间与通道权重;SegMamba^[28]则提出了层级化的 Mamba-in-Mamba(MiM)架构,将 3D 医学体数据划分为非重叠的块序列,在外层利用状态空间模型建模全局块间依赖关系,内部通过双向状态空间扫描机制实现对局部切片内空间上下文信息建模,并进一步设计跨层级特征融合模块实现全局语义与局部细节的协同优化。在混合架构研究中,Swin U-Mamba^[29]结合了滑动窗口机制与状态

空间模型, 将二维医学图像划分为非重叠的局部窗口, 在每个窗口内部利用状态空间层高效建模局部上下文信息, 并通过层级化的窗口移位设计实现跨窗口的信息交互和多尺度特征融合; CM-UNet^[30] 则采用分阶段特征提取, 浅层保留 CNN 对图像的局部特征提取优势, 深层引入状态空间模块建模图像的长程依赖, 兼顾解剖细节刻画与全局上下文理解。

2 基于 Mamba 的医学图像分割方法

2.1 网络架构

为了解决 Mamba 模型在医学图像分割中存在的频域特征提取不足、器官或病灶区域建模不足及空间连续性断裂等问题, 本文提出了 SDA U-Mamba 模型。该模型通过在 VMamba 的视觉状

态空间模块中引入空频融合模块和注意力机制, 有效提升了模型对图像中器官或病灶的感知能力与多尺度特征融合能力, 其整体网络架构如图 1(a) 所示。从图 1(a) 中可以看出, 该框架采用经典的 U 型结构, 利用编解码特征的跳跃连接实现了低层细节特征与高层语义信息的有效融合。但与传统 U 型架构不同的是, 本模型设计了分层 U 型架构, 其中前 3 层引入了 Mamba 空频注意力模块, 而后两层则采用了双极路由注意力模块。具体而言, Mamba 空频注意力模块用于增强全局语义建模并有效捕捉图像的频域特征, 而双极路由注意力模块则专注于器官或病灶的自适应聚焦, 这种设计使模型在浅层网络侧重全局上下文建模, 深层网络专注器官或病灶语义信息增强, 从而整体提升了模型在医学图像分割中的精度。

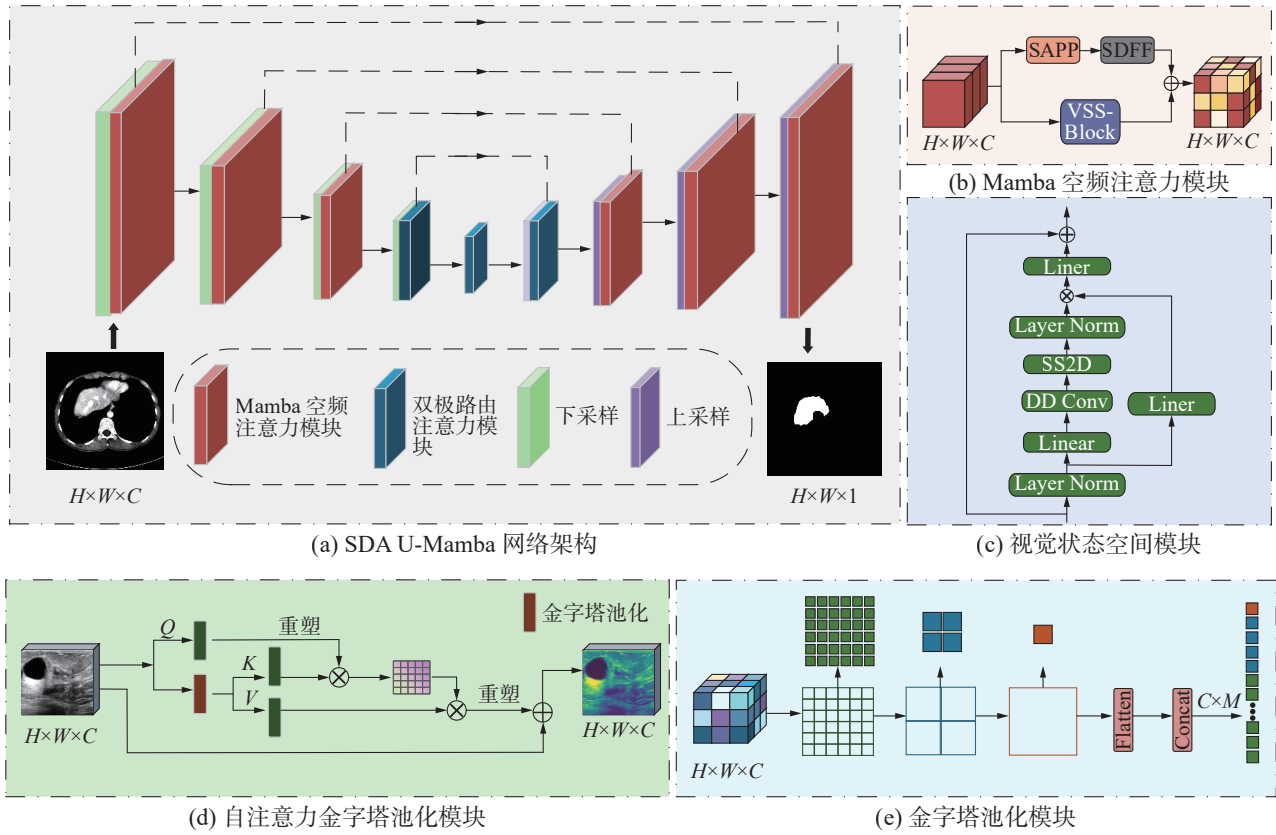


图 1 网络整体架构及各组合模块结构

Fig. 1 Overall architecture of the network and structures of each composite module

2.2 Mamba 空频注意力模块

为解决传统 Mamba 模型在医学图像分割中存在的空间结构连续性断裂、频域信息缺失等问题, 本文提出了一种 Mamba 空频注意力模块, 如图 1(b) 所示, 该模块由视觉状态空间模块、谱动态特征融合模块和自注意力金字塔池化模块组成, 构建了并行协同的全局-局部特征学习框架, 并实现了跨空间频率维度的信息融合与优化。

其中, VSS-Block 作为长程建模分支, 摒弃了传统 Mamba 在处理图像时采用的逐像素扫描策略, 它创新性地设计了双向扫描机制, 在状态转移过程中保持了图像固有的二维空间结构连续性, 从而能够高效捕获医学图像全局上下文中的长距离依赖关系, 显著强化了模型的全局语义表达能力。然而, VSS-Block 在医学图像边缘纹理和局部细节的建模上仍然存在不足。为此, 本文

设计了谱动态特征融合模块,该模块融合傅里叶频域分析与可变形卷积操作,增强了模型对医学图像中频域信息与复杂边界区域的响应能力。此外,针对U型架构中下采样导致的高层语义信息丢失及上采样过程中的语义非对齐问题,本文设计了自注意力金字塔池化模块,该模块结合空间金字塔池化的多尺度上下文捕获能力与空间注意力机制的区域感知特性,实现上下文特征的精细校准,显著增强了模型对医学图像多尺度上下文信息的建模能力和复杂语义区域的表征能力。

综合来看,VSS-Block实现了对医学图像的全局特征建模能力,谱动态特征融合模块强化了对医学图像局部信息的表征,而自注意力金字塔池化模块则有效提升了模型对医学图像多尺度语义的动态适配能力。三者协同构建的Mamba空频注意力模块实现了更精细、更全面的特征表达,在提升医学图像分割精度方面展现出显著优势。

2.2.1 视觉状态空间模块

视觉状态空间模块作为VMamba架构的核心模块,其结构如图1(c)所示,旨在解决二维视觉数据的长程依赖建模难题,其核心功能单元SS2D通过四向正交展开策略将二维特征图分解为水

平、垂直及对角方向的伪序列流,分别在每个方向实施选择性状态空间建模。这一过程不仅充分利用了图像的空间信息,还通过多方向的展开和融合,增强了模型对图像中复杂结构的捕获能力。具体实现如下:

$$\mathbf{Z}_d = \text{expand}(\mathbf{Z}, d)$$

$$\overline{\mathbf{Z}}_v = \text{S6}(\mathbf{Z}_d)$$

$$\overline{\mathbf{Z}} = \text{merge}(\overline{\mathbf{Z}}_1, \overline{\mathbf{Z}}_2, \overline{\mathbf{Z}}_3, \overline{\mathbf{Z}}_4)$$

式中: \mathbf{Z} 为输入的特征图, $\overline{\mathbf{Z}}$ 为输出特征图, d 为4个不同的扫描方向, $\text{expand}(\cdot)$ 和 $\text{merge}(\cdot)$ 分别为扫描扩展操作和扫描合并操作。

2.2.2 谱动态特征融合模块

医学图像中的器官或病灶往往具有不规则形状和复杂纹理,这对模型的细节建模能力提出了更高要求。然而,传统卷积运算卷积核尺寸固定、感受野受限,即使通过堆叠多层卷积扩大感受野范围,依然难以充分捕捉局部细节和模糊边缘信息。为此,本文提出谱动态特征融合模块作为对VSS-Block的有效补充,其结构如图2(a)所示。该模块既增强了模型对细粒度纹理的感知能力,又提升了模型对频率信息的解析能力。

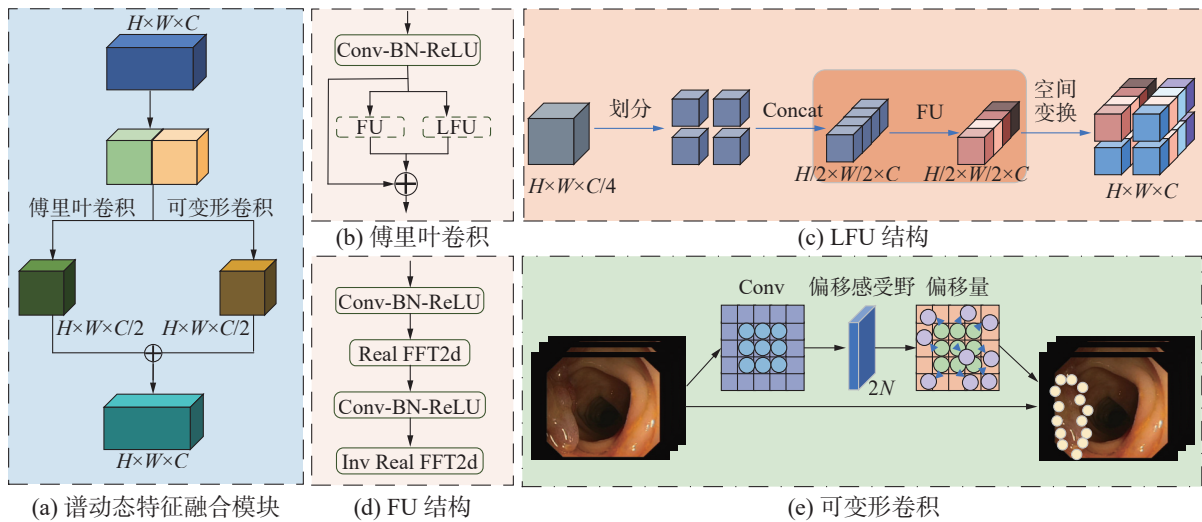


图 2 谱动态特征融合模块及各组成部分结构

Fig. 2 Spectral dynamic feature fusion module and structures of each component

基于傅里叶变换的谱分析理论^[31]、频域分量与全局特征的高度耦合为频域卷积提供了坚实的数学基础,即单一频谱分量的调整即可触发全局特征的重构,使频域操作能够高效建模全局上下文,其结构如图2(b)所示。与之互补的可变形卷积通过引入可学习的偏移参数,动态调整卷积核采样点的空间分布,突破了传统卷积核几何结构的刚性限制,其结构如图2(e)所示。该模块通过融合傅里叶变换的频域建模能力与可变形卷积的

空间自适应性,构建了具备空频协同感知能力的混合卷积机制,通过频域分支解析纹理的频率分布规律以强化细粒度特征表达,同时利用可变形卷积动态调整采样位置以适配不规则病灶形态,既解决了传统卷积感受野与细节感知失衡的问题,又弥补了单一空频建模对医学图像复杂场景适配不足的短板。具体过程如下:

首先,给定输入特征图 $\mathbf{X} \in \mathbf{R}^{H \times W \times C}$,其中 H 、 W 、 C 分别表示图像的高度、宽度与通道数。模块

将输入划分为空域分支与频域分支进行并行建模。在频域分支中, 首先通过 1×1 卷积、批归一化及激活函数对输入特征进行通道压缩与映射, 得到初步特征表示。

$$X_c = \text{ReLu}(\text{BN}(\text{Conv}_{1 \times 1}(X)))$$

随后, 对空间特征施加二维傅里叶变换得到复数频谱, 并通过可学习的 1×1 卷积层联合调制频谱的实部与虚部分量, 再经逆傅里叶变换重构空间特征, 实现对全频段信息的自适应增强建模。

$$F_{fu} = \text{FU}(X_c)$$

同时, 将特征分块后送入局部傅里叶变换对低频信息进行处理, 得到低频增强特征, 因为在医学图像中, 低频信息决定器官或病灶的整体形态和位置, 而高频分量则包含细节和噪声, 通过增强低频特征, 可以提升模型对整体结构的把握能力。

$$F_{flu} = \text{LFU}(X_c)$$

频域分支的最终输出通过逐元素相加融合压缩特征、FU 输出与 LFU 输出, 并通过 1×1 卷积投影得到:

$$F_{fre} = \text{Conv}_{1 \times 1}(X_c + F_{fu} + F_{flu})$$

在空域分支中, 为增强模型对语义边界的建模能力, 模块引入可变形感知机制。首先, 利用 3×3 卷积对输入图像 X 生成空间偏移场 $\Delta p_n \in \mathbf{R}^{H \times W \times C}$ 。

$$\Delta p_n = \text{Conv}_{3 \times 3}(X)$$

随后, 结合该偏移信息调整标准卷积核的采样位置, 实现对局部结构的动态感知。输出特征表示为

$$F_{\text{def}(p_0)} = \sum_{p_0 \in R} w_n \cdot X(p_0 + p_n + \Delta p_n(p_0))$$

式中: p_0 为当前卷积中心位置, R 为标准感受野, w_n 为卷积权重。最终, 模块将频域增强特征与可变形感知特征进行加权融合, 输出结果为

$$F_{\text{out}} = \alpha F_{\text{fre}} + (1 - \alpha) \cdot F_{\text{def}}$$

式中融合系数 $\alpha \in [0, 1]$, 固定设置为 0.5, 反映频域与空域特征在融合过程中的相对贡献。

2.2.3 自注意力金字塔池化模块

尽管 U 型结构在医学图像分割中已广泛应用, 其编解码对称设计有助于整合多尺度特征, 但在下采样和上采样过程中, 模型常因固定感受野、尺度缩放及插值操作导致语义信息损失和上下文表达不一致, 表现为对图像中器官或病灶的结构重建不完整、边界模糊等问题。特别是对于形态复杂或尺寸变化显著的器官或病灶区域, 模型缺乏对多尺度上下文的动态适配能力, 难以兼顾局部细节与全局结构表达。

为缓解上述问题, 本文提出自注意力金字塔

池化模块, 如图 1(d) 所示。区别于传统金字塔池化仅侧重多尺度特征聚合或单一自注意力仅聚焦局部区域的局限, 该模块创新性地多尺度上下文感知与空间注意力机制进行协同优化设计, 通过在不同池化尺度上动态分配注意力权重, 使模型能根据医学图像中器官与病灶的语义优先级, 自适应强化关键尺度的特征表达, 同时抑制无关尺度的冗余信息。该模块被嵌入至编码器和解码器的主干路径中, 通过融合多尺度上下文感知与空间注意力机制, 增强模型在上下采样过程中的语义适配能力与区域表达能力。

具体而言, SAPP 首先采用不同尺度的自适应平均池化 (如 1×1 、 3×3 、 6×6 和 8×8 网格划分) 提取多粒度上下文信息, 并通过双线性插值还原至原始特征图尺寸后进行拼接, 构建出一个多尺度上下文特征金字塔, 其结构如图 1(e) 所示。随后引入空间注意力机制, 对各尺度特征进行逐位置的显式加权, 增强模型对图像中器官或病灶的响应能力, 并抑制背景干扰, 从而实现上下文信息的选择性聚合与结构引导。

该模块不仅在局部到全局层面维持了空间结构的连续性, 也有效缓解了因固定感受野和尺度缩放带来的上下文语义偏移问题, 从而提升了模型对结构变化剧烈或语义复杂区域的适应能力。其核心计算过程如下式所示:

$$\begin{aligned} Y_k &= \text{Pool2d}(X, (k, \max(1, \lfloor \text{ar} \times k \rfloor))) \\ Z &= \text{Concat}(\text{Flatten}(Y_1), \text{Flatten}(Y_2), \\ &\quad \text{Flatten}(Y_3), \text{Flatten}(Y_4)) \end{aligned}$$

式中: ar 为宽高比 (H/W); Pool2d 为自适应平均池化操作; Flatten 为将特征图展平为一维向量, 并在最后一个维度拼接起来, 得到最终的输出特征图 Z 。

2.3 双极路由注意力模块

在医学图像分割任务中, 器官或病灶通常呈现边界模糊、形态多变等特征, 使得模型在深层特征表达过程中容易出现语义扩散现象, 即对目标区域的响应被背景信息稀释, 导致模型对医学图像器官或病灶的聚焦能力减弱。尤其对于 Mamba 这类以长程依赖建模为主的状态空间模型, 尽管其具备优秀的全局上下文建模能力, 但由于缺乏显式的位置引导与区域选择机制, 在全局上下文建模中易引入大量冗余背景信息, 从而削弱模型对图像中器官或病灶的判别能力。

为解决上述问题, 本文在 U-Net 架构深层引入了双极路由注意力模块^[32], 该模块的核心双层动态路由框架借鉴自 BiFormer, 但在其基础上针对医学图像分割需求进行了优化。其结构如图 3 所示, 以实现图像中器官或病灶的聚焦建模。

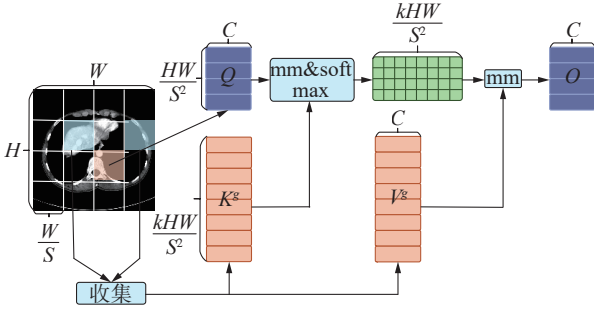


图 3 BRA 模块结构

Fig. 3 Structure of the BRA module

该模块采用双层动态路由策略，首先在区域级别筛选语义相关性最强的位置，再在细粒度层面进行局部特征聚合与语义强化，构建由粗至细的多级注意力引导机制。与此同时，BRA 引入了结构感知增强模块，融合局部上下文信息，进一步提升模型对器官或病灶边缘与微小结构的响应能力，并有效抑制背景区域的冗余干扰。此外，BRA 模块通过引入令牌压缩与稀疏激活策略，大幅降低了计算复杂度和内存占用，保证在资源受限的硬件环境下仍具备高效的推理能力和良好的实时性能。双极路由注意力模块算法的伪代码如下：

输入: $x \in \mathbf{R}^{B \times H \times W \times C}$

输出: $y \in \mathbf{R}^{B \times C \times H \times W}$

- 1) $x \leftarrow F.pad(x, pad_s, pad_b)$;
- 2) $x \leftarrow Reshape(x) \in (N, P^2, h, w, C)$;
- 3) $(q, kv) \leftarrow QKVLinear(x)$;
- 4) $q_{pix} \leftarrow Reshape(q) \in (N, P^2, hw, C)$;
- 5) $kv_{pix} \leftarrow Downsample(kv)$;
- 6) $q_{win} \leftarrow Mean_{h,w}(q), k_{win} \leftarrow Mean_{h,w}(kv [: C_{qk}])$;
- 7) $lepe \leftarrow DWConv(kv [C_{qk} :])$;
- 8) $(r_w, r_i dx) \leftarrow TopkRouting(q_{win}, k_{win})$;
- 9) $kv_{sel} \leftarrow KVgather(r_i dx, r_w, kv_{pix})$;
- 10) $(k_{pix}, v_{pix}) \leftarrow Split(kv_{sel})$;
- 11) $attn \leftarrow Softmax(q_{pix} \cdot k_{pix} / \sqrt{d})$;
- 12) $out \leftarrow attn \cdot v_{pix}$;
- 13) $out \leftarrow Rearrange(out) + lepe$;
- 14) $y \leftarrow W_o(out)$;
- 15) **if** $auto_{pad} = True$ **then** $y \leftarrow Crop(y)$;

算法 1 中用伪代码对算法进行了总结，下面将给出具体计算过程：

1) 区域划分和线性投影：将 2D 输入特征图 $X \in \mathbf{R}^{H \times W \times C}$ 划分为 $S \times S$ 个不重叠的区域，获得每个区域的特征维数，随后将得到的特征图 $X^r \in \mathbf{R}^{S^2 \times HW/S^2 \times C}$ 通过线性投影导出查询 Q 、键 K 、值 $V \in \mathbf{R}^{S^2 \times HW/S^2 \times C}$ 。

$$Q = X^r W^q, K = X^r W^k, V = X^r W^v$$

式中 $W^q, W^k, W^v \in \mathbf{R}^{C \times C}$ 为投影矩阵。

2) 区域间路由：首先分别计算每个区域的 Q 和 K 的平均值，产生区域级查询 $Q^r, K^r \in \mathbf{R}^{S^2 \times C}$ ，接着通过应用 Q^r 和 K^r 之间的矩阵乘法导出区域到区域邻接矩阵 $A^r \in \mathbf{R}^{S^2 \times S^2}$ ，最后使用 $topkIndex()$ 运算符得到路由索引矩阵 $I^r \in \mathbf{R}^{S^2 \times k}$ ，仅保留每个查询区域的 top-k 最相关区域：

$$A^r = Q^r (K^r)^T$$

$$I^r = topkIndex(A^r)$$

3) 令牌对令牌注意：由于路由区域可能在空间上分散在整个特征图上，因此需要收集路由区域中的键和值张量，然后细粒度的令牌对令牌注意力被应用于这些键值张量。

$$K^g = gather(K, I^r), V^g = gather(V, I^r)$$

$$Attention(Q, K^g, V^g) =$$

$$softmax((Q(K^g)^T) / \sqrt{C}) V^g + LCE(V)$$

式中: $K^g, V^g \in \mathbf{R}^{S^2 \times k \times HW/S^2 \times C}$ 是收集的键和值张量，函数 $LCE(\cdot)$ 使用深度卷积进行参数化。

3 实验结果与分析

3.1 数据集

为验证所提出方法在多种医学图像分割任务中的通用性与鲁棒性，本文在乳腺超声图像数据集 (breast ultrasound images, BUSI)、结肠镜图像数据集 (colorectal cancer-clinic dataset, CVC-ClinicDB) 以及肝脏 CT 数据集 (combined healthy abdominal organ segmentation, CHAOS-Liver) 3 个公开数据集上进行了系统评估。BUSI 数据集包含 780 张乳腺超声图像，涵盖正常、良性及恶性 3 类样本，图像质量受噪声、阴影等伪影影响严重，主要用于乳腺肿块分割任务，测试模型在低信噪比条件下的病灶定位与边界感知能力；CVC-ClinicDB 是一个用于结肠镜图像中息肉分割的公开数据集，共包含 612 张高分辨率图像，具有丰富的色彩变化与形态复杂性，是评估模型在自然腔镜环境中进行精细分割性能的重要基准；CHAOS-Liver 数据集来源于 MICCAI 2019 CHAOS 挑战，包含 T1 和 T2 加权的磁共振 (magnetic resonance, MR) 及计算机断层扫描 (computed tomography, CT) 扫描影像，图像分辨率高、对比度差异显著，适用于验证模型在大尺度器官结构下的区域建模与边界保持能力。

3.2 实验设置

本研究采用 PyTorch 框架，在 NVIDIA RTX 3090 GPU 平台上进行实验。为保障实验一致性，对 3 类医学图像数据集进行了统一预处理，将所有图像尺寸调整为 224 像素 \times 224 像素。实验设置方面采用 Adam 优化器 (初始学习率为 0.000 1)，

批处理大小设为 2, 选用二元交叉熵 (binary cross-entropy, BCE) 损失与戴斯相似系数 (Dice similarity coefficient, DICE) 损失的加权组合作为损失函数, 以兼顾整体分割精度和小目标区域的识别性能。为公平比较模型性能, 所有实验均不使用预训练权重或后处理技术。在数据增强方面, 仅采用随机旋转和翻转两种基础增强方式, 在保证数据多样性的同时避免过强的先验干扰。每个数据集均训练 400 个 epoch 以确保充分收敛。通过在解剖结构和病灶类型各异的医学图像数据集上进行验证, 本实验方案能够全面评估所提方法在不同分割任务中的适用性与鲁棒性。

3.3 实验结果

与其他方法对比以此来评估本文的模型, 基于 CNN 和 Transformer 的模型: U-Net^[15]、Swin-Uet^[24] 和 CiT^[33]; 基于 Kan 的模型: U-Kan^[34]; 基于 MLP 的模型: Rolling-Unet^[35]; 基于 Mamba 的模型: VM-

Unet^[27]、VM-UnetV2^[36]、Swin U-Mamba^[29] 和 Ultra-Light VM-Unet^[37]。并采取 IoU、Dice、HD95 和 HD 作为评估指标。

表 1~3 分别给出了 SDA U-Mamba 在 BUSI、CVC-ClinicDB 和 CHAOS-Liver 数据集上的实验结果。图 4 为不同模型各个数据集上分割结果。结果表明, SDA U-Mamba 在 3 个数据集上的 IoU 和 Dice 指标均优于现有方法, 说明模型在预测区域与真实分割区域之间具有更高的重叠度和语义一致性, 能够有效应对医学图像中的噪声干扰及组织结构不均等挑战。尤其是 Dice 系数对小体积目标更为敏感, 其提升进一步验证了 SDA U-Mamba 在小目标分割中的优势, 既保持了大目标的分割性能, 又兼顾了边界细节和细粒度区域的准确表达。此外, 3 个数据集涵盖了不同成像模态与器官结构, 结果显示 SDA U-Mamba 在多样化数据分布下表现稳定, 体现了良好的鲁棒性和泛化能力。

表 1 BUSI 数据集实验结果

Table 1 Experimental results on the BUSI dataset

Methods	Backbone	IoU ↑	Dice ↑	HD↓	HD95↓	参数量/MB
U-Net ^[15]	CNN	58.39	72.56	51.54	16.43	34.53
Swin-Unet ^[24]	Transformer	58.92	74.72	33.73 ↓3.37	5.05	27.15
CiT ^[33]	CNN+Transformer	58.89	72.92	41.53	17.26	20.55
U-Kan ^[34]	Kan	60.80	75.54	37.58	4.97	6.35
Rolling-Unet ^[35]	MLP	61.87	76.34	40.14	3.58 ↓0.98	28.32
VM-Unet ^[27]	Mamba	61.26	75.37	56.26	13.77	34.62
VM-Unet V2 ^[36]	Mamba	56.62	72.28	38.16	4.56	17.91
Swin U-Mamba ^[29]	Mamba	60.67	74.70	48.50	18.78	55.06
UltraLight VM-UNet ^[37]	Mamba	52.51	67.66	55.11	27.56	0.04
SDA U-Mamba	Mamba	65.73 ↑4.47	78.72 ↑2.38	37.10	5.20	7.56

注: 红色代表最优结果, 蓝色代表次优结果, 箭头后数字为较次优方法提升指标。

表 2 CVC-ClinicDB 数据集实验结果

Table 2 Experimental results on the CVC-ClinicDB dataset

Methods	Backbone	IoU ↑	Dice↑	HD↓	HD95↓	参数量/MB
U-Net ^[15]	CNN	84.96	90.2	33.51	2.98	34.53
Swin-Unet ^[24]	Transformer	60.51	76.87	39.77	5.95	27.15
CiT ^[33]	CNN+Transformer	70.45	82.6	39.56	17.86	20.55
U-Kan ^[34]	Kan	78.08	87.66	35.08	2.64 ↓0.28	6.35
Rolling-Unet ^[35]	MLP	86.17	91.70	18.82 ↓2.55	2.92	28.32
VM-Unet ^[27]	Mamba	70.22	81.84	38.28	12.4	34.62
VM-Unet V2 ^[36]	Mamba	63.63	77.57	32.35	5.05	17.91
Swin U-Mamba ^[29]	Mamba	82.12	89.53	34.76	7.25	55.06
UltraLight VM-UNet ^[37]	Mamba	48.83	62.98	50.44	26.01	0.04
SDA U-Mamba	Mamba	86.46 ↑0.29	92.52 ↑0.82	21.37	3.32	7.56

注: 红色代表最优结果, 蓝色代表次优结果, 箭头后数字为较次优方法提升指标。

表 3 CHAOS-Liver 数据集实验结果
Table 3 Experimental results on the CHAOS-Liver dataset

Methods	Backbone	IoU ↑	Dice ↑	HD ↓	HD95 ↓	参数量/MB
U-Net ^[15]	CNN	96.35	97.65	7.01	0.050	34.53
Swin-UNet ^[24]	Transformer	92.73	95.94	13.35	0.723	27.15
CiT ^[33]	CNN+Transfomer	94.74	96.83	7.89	0.111	20.55
U-Kan ^[34]	Kan	96.48	97.72	9.92	0.060	6.35
Rolling-UNet ^[35]	MLP	96.01	97.48	8.74	0.530	28.32
VM-UNet ^[27]	Mamba	96.11	98.01	7.36	0.009	34.62
VM-UNet V2 ^[36]	Mamba	95.90	97.70	7.77	0.056	17.91
Swin U-Mamba ^[29]	Mamba	95.79	97.32	6.57 ↓0.79	0.036 ↓0.007	55.06
UltraLight VM-UNet ^[37]	Mamba	93.99	96.60	8.09	0.170	0.04
SDA U-Mamba	Mamba	96.67 ↑0.19	98.28 ↑0.27	7.75	0.043	7.56

注：红色代表最优结果，蓝色代表次优结果，箭头后数字为较次优方法提升指标。

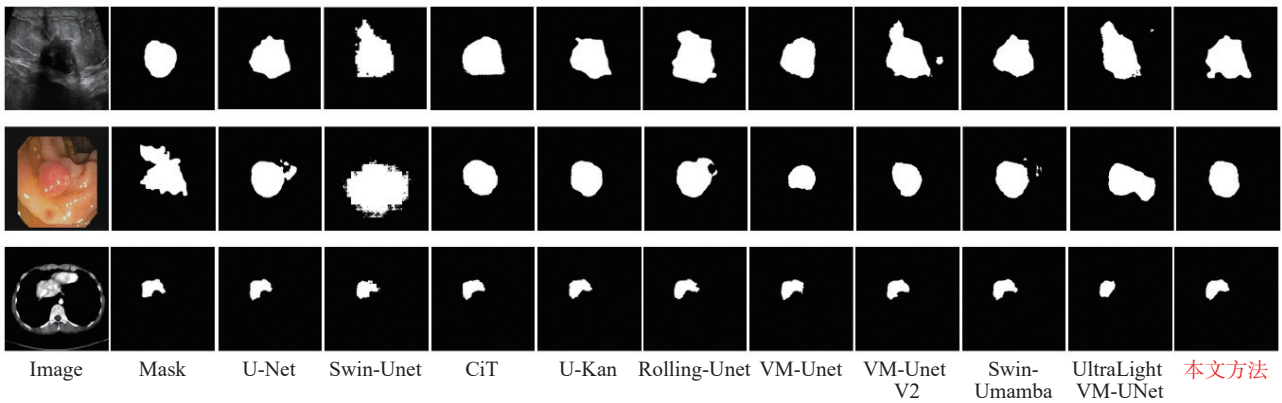


图 4 不同模型各个数据集上分割结果

Fig. 4 Segmentation results of different models on various datasets

需要指出的是，尽管在多个指标上表现出色，SDA U-Mamba 在 3 个数据集上的 HD95 指标未达到最优，表明模型在边界定位方面仍存在一定的改进空间。造成这一现象的可能原因包括：一是数据样本规模有限，限制了模型对边界变化的充分学习；二是空频注意力模块虽能捕捉边缘纹理的空频特征，但在特征优化优先级上更侧重全局上下文建模，对边界细节的强化程度不够；三是双极路由注意力模块主要聚焦器官与病灶的整体区域聚焦，未针对边界像素设计专门的特征引导机制，这些因素共同导致 HD95 指标未能达到最优。未来工作将重点致力于提升模型对边界细节的表达能力，以进一步增强其在复杂临床场景中的泛化表现。

3.4 消融实验

为了研究各个模块对模型性能的影响，本文在 BUSI 数据集上进行了消融实验。如表 4 所示，本文给出了 SAPP 与 SDFP 对模型性能的影响。结果表明，SDFP 在提升效果上最为显著，SAPP

次之，而两者联合使用时性能最佳，说明两模块在空间与频域特征提取上具有良好的互补性，能有效增强 Mamba 在医学图像分割中的表现。表 5 进一步分析了不同尺度的金字塔池化窗口对分割效果的影响，当窗口大小设置为 (1, 3, 6, 8) 时，模型性能最优，说明合理的多尺度配置有助于更全面地捕捉图像上下文特征。最后，表 6 给出了 BRA 模块中路由区域数 k 对注意力机制的影响，结果发现 $k=4$ 时模型表现最佳，表明在保持局部空间相关性的同时，限制路由区域的数量有助于减少冗余信息干扰，提升模型对图像中器官或病灶的聚焦能力与整体分割性能。

表 4 各模块消融实验
Table 4 Ablation experiments for each module

SAPP	SDFP	BRA	IoU	Dice	HD	HD95
		√	62.11	75.90	40.42	12.48
√		√	65.04	77.63	38.25	9.66
	√	√	65.18	78.19	41.65	12.47
√	√		65.37	78.14	38.21	5.09
√	√	√	65.73	78.72	38.10	7.20

表 5 金字塔池化模块不同尺度消融实验

Table 5 SPP ablation experiments at different scales

SPP尺度	IoU	Dice	HD	HD95
(1,2,3,6)	64.26	77.40	49.43	15.37
(1,2,8,12)	64.29	77.56	46.22	13.04
(1,3,6,8)	65.73	78.72	38.10	7.20

表 6 BRA 不同 top-k 消融实验

Table 6 Ablation experiments of BRA with different top-k values

top-k	IoU	Dice	HD	HD95
3	64.67	77.96	38.73	10.54
4	65.73	78.72	38.10	7.20
5	63.25	76.54	42.32	14.17
6	63.01	76.67	44.85	12.48

4 结束语

本研究提出了一种融合频域动态特征与注意力机制的医学图像分割网络 SDA U-Mamba, 旨在应对传统 Mamba 模型在医学图像中存在的空间结构连续性受损与器官或病灶聚焦能力不足方面的挑战。具体而言, SDA U-Mamba 在网络浅层设计 Mamba 空频注意力模块, 通过空域卷积、频域变换与金字塔自注意力机制的联合建模, 有效提升模型对边缘细节与多尺度上下文的感知能力, 强化空间信息的表达连续性; 在网络深层引入双极路由注意力模块, 采用稀疏动态路由策略显式增强模型对器官或病灶的响应能力, 同时维持 Mamba 在建模长程依赖中的计算优势。此外, SDA U-Mamba 具有良好的模块化特性, 关键模块可即插即用地集成至多种主干架构中, 适应不同任务需求。未来, 本研究将进一步挖掘 Mamba 架构在医学图像分割中的潜力, 持续优化其结构设计与建模机制, 以提升其在复杂临床场景中的泛化能力与实用价值。

参考文献:

- [1] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017(6): 1229–1251.
ZHOU Feiyan, JIN Linpeng, DONG Jun. Review of convolutional neural network[J]. Chinese journal of computers, 2017(6): 1229–1251.
- [2] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. Pattern recognition letters, 2020, 133: 157–165.
- [3] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2261–2269.
- [4] TAN Mingxing, LE Q V. Efficientnet: rethinking model scaling for convolutional neural networks[C]//International Conference on Machine Learning. Long Beach: PMLR, 2019.
- [5] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[C]//International Conference on Learning Representations. Virtual: ICLR, 2021.
- [6] LEI Tao, SUN Rui, DU Xiaogang, et al. SGU-Net: shape-guided ultralight network for abdominal image segmentation[J]. IEEE journal of biomedical and health informatics, 2023, 27(3): 1431–1442.
- [7] 周丽娟, 毛嘉宁. 视觉 Transformer 识别任务研究综述[J]. 中国图象图形学报, 2023, 28(10): 2969–3003.
ZHOU Lijuan, MAO Jianing. Vision Transformer-based recognition tasks: a critical review[J]. Journal of image and graphics, 2023, 28(10): 2969–3003.
- [8] GU A, GOEL K, RÉ C. Efficiently modeling long sequences with structured state spaces[C]//International Conference on Learning Representations. Virtual: ICLR, 2022.
- [9] GU A, DAO T. Mamba: linear-time sequence modeling with selective state spaces[EB/OL]. (2023–12–01)[2025–04–20]. <https://arxiv.org/abs/2312.00752>.
- [10] JIAO Jianbin, LIU Yue, LIU Yunfan, et al. VMamba: visual state space model[C]//Advances in Neural Information Processing Systems. Vancouver: Neural Information Processing Systems Foundation, Inc., 2024.
- [11] ZHU Lianghui, LIAO Bencheng, ZHANG Qian, et al. Vision mamba: efficient visual representation learning with bidirectional state space model[C]//Proceedings of the 41st International Conference on Machine Learning. Baltimore: PMLR, 2024.
- [12] HE Xuanhua, CAO Ke, ZHANG Jie, et al. Pan-Mamba: effective pan-sharpening with state space model[J]. Information fusion, 2025, 115: 102779.
- [13] PEI Xiaohuan, HUANG Tao, XU Chang. Efficientmamba: atrous selective scan for light weight visual mamba[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver: AAAI, 2025.
- [14] CHENG Zihan, GUO Jintao, ZHANG Jian, et al. Mamba-sea: a mamba-based framework with global-to-local sequence augmentation for generalizable medical image segmentation[J]. IEEE transactions on medical imaging, 2025, 44(9): 3741–3755.
- [15] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer Assisted Intervention. Munich: MICCAI Society, 2015.
- [16] ZHOU Zongwei, RAHMAN SIDDIQUEE M M, TAJBAKHS N, et al. UNet++: a nested UNet architecture for medical image segmentation[C]//Proceedings of the 4th International Workshop on Deep Learning in Medical Image Analysis. Granada: DLMIA, 2018.
- [17] HUANG Huimin, LIN Lanfen, TONG Ruofeng, et al. UNet 3+: a full-scale connected UNet for medical image

- segmentation[C]//2020 IEEE International Conference on Acoustics, Speech and Signal Processing. Barcelona: IEEE, 2020.
- [18] ZHU Wenhui, CHEN Xiwen, QIU Peijie, et al. SelfReg-UNet: self-regularized UNet for medical image segmentation[C]//International Conference on Medical Image Computing and Computer Assisted Intervention. Morocco: MICCAI Society, 2024.
- [19] MILLETARI F, NAVAB N, AHMADI S A. V-Net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3D Vision. Stanford: IEEE, 2016.
- [20] GIBSON E, GIGANTI F, HU Yipeng, et al. Automatic multi-organ segmentation on abdominal CT with dense V-networks[J]. *IEEE transactions on medical imaging*, 2018, 37(8): 1822–1834.
- [21] ROY S, KOEHLER G, ULRICH C, et al. MedNeXt: transformer-driven scaling of ConvNets for medical image segmentation[C]//International Conference on Medical Image Computing and Computer Assisted Intervention. Vancouver: MICCAI Society, 2023: 405–415.
- [22] ZAGORUYKO S, KOMODAKIS N. Learning to compare image patches via convolutional neural networks[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015.
- [23] HATAMIZADEH A, TANG Yucheng, NATH V, et al. UNETR: transformers for 3D medical image segmentation[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2022.
- [24] CAO Hu, WANG Yueyue, CHEN J, et al. Swin-UNet: UNet-like pure transformer for medical image segmentation[C]//European Conference on Computer Vision. Tel Aviv: ECVA, 2023.
- [25] XU Qing, MA Zhicheng, NA He, et al. DCSAU-Net: a deeper and more compact split-attention U-Net for medical image segmentation[J]. *Computers in biology and medicine*, 2023, 154: 106626.
- [26] LIU Xueyu, SHI Guangze, WANG Rui, et al. Segment any tissue: one-shot reference guided training-free automatic point prompting for medical image segmentation[J]. *Medical image analysis*, 2025, 102: 103550.
- [27] RUAN Jiacheng, LI Jincheng, XIANG Suncheng. VM-UNet: vision mamba UNet for medical image segmentation[EB/OL]. (2024-02-04)[2025-04-20]. <https://arxiv.org/abs/2402.02491>.
- [28] XING Zhaohu, YE Tian, YANG Yijun, et al. SegMamba: long-range sequential modeling mamba for 3D medical image segmentation[C]//International Conference on Medical Image Computing and Computer Assisted Intervention. Marrakesh: MICCAI Society, 2024.
- [29] LIU Jiarun, YANG Hao, ZHOU Hongyu, et al. Swin-UMamba: mamba-based UNet with ImageNet-based pre-training[C]//International Conference on Medical Image Computing and Computer Assisted Intervention. MICCAI Society, 2024.
- [30] LIU Mushui, DAN Jun, LU Ziqian, et al. CM-UNet: Hybrid CNN-Mamba UNet for remote sensing image semantic segmentation[EB/OL]. (2024-02-04)[2025-02-20]. <https://arxiv.org/abs/2405.10530>.
- [31] CHI Lu, JIANG Borui, MU Yadong. Fast fourier convolution[J]. *Advances in neural information processing systems*, 2020, 33: 4479–4488.
- [32] ZHU Lei, WANG Xinjiang, KE Zhanghan, et al. BiFormer: vision transformer with bilevel routing attention[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023.
- [33] LEI Tao, SUN Rui, WANG Xuan, et al. CiT-net: convolutional neural networks hand in hand with vision transformers for medical image segmentation[C]//Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. Maca: International Joint Conferences on Artificial Intelligence Organization, 2023.
- [34] LI Chenxin, LIU Xinyu, LI Wuyang, et al. U-KAN makes strong backbone for medical image segmentation and generation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Philadelphia: AAAI, 2025.
- [35] LIU Yutong, ZHU Haijiang, LIU Mengting, et al. Rolling-UNet: revitalizing MLP's ability to efficiently extract long-distance dependencies for medical image segmentation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver: AAAI, 2024.
- [36] ZHANG Mingya, YU Yue, JIN Sun, et al. VM-UNET-V2: rethinking vision mamba UNet for medical image segmentation[C]//International Symposium on Bioinformatics Research and Applications. Kunming: CCF, 2024.
- [37] WU Renkai, LIU Yinghao, NING Guochen, et al. UltraLight VM-UNet: parallel vision mamba significantly reduces parameters for skin lesion segmentation[J]. *Patterns*, 2025, 6(11): 101298.

作者简介:



彭晨阳, 硕士研究生, 主要研究方向为计算机视觉、机器学习。E-mail: 2046591497@qq.com。



何立风, 教授, 博士, 主要研究方向为数字图像处理、机器学习。主持国家自然科学基金 2 项, 发表学术论文 60 余篇, 申请国家发明专利 10 余项。E-mail: helifeng@ist.aichipu.ac.jp。



雷涛, 教授, 博士, 陕西科技大学电子信息与人工智能学院副院长, 主要研究方向为人工智能、计算机视觉。主持国家自然科学基金 5 项, 发表学术论文 100 余篇, 其中 12 篇论文入选 ESI 高被引论文。E-mail: leitao@sust.edu.cn。