中图法分类号： 文献标识码： 文章编号：1006-8961(XXXX)XX-0001-16

论文引用格式：Hu Keli, Huang Jie, Fan Kangxin, Wang Chen, Zhu Hancan, Zhao Liping, Ren Haowei, Hu Jianhao. A polyp detection network with multi-level and multi-scale feature fusion enhancement based on Mamba architecture[J/OL]. Journal of Image and Graphics, XXXX:1-16. DOI: 10.11834/jig.250492.(胡珂立, 黄杰, 范康鑫, 王臣, 祝汉灿, 赵利平, 任浩炜, 胡剑浩. Mamba架构下多级多尺度特征融合增强的息肉检测网络[J/OL]. 中国图象图形学报,XXXX:1-16. DOI: 10.11834/jig.250492.)[DOI:10.11834/jig.250492]

Mamba架构下多级多尺度特征融合增强的息肉检测网络

胡珂立[1,2,3],黄杰[1],范康鑫[1],王臣[1*],祝汉灿[2],赵利平[1],任浩炜[2],胡剑浩[2]

1. 绍兴文理学院智能工程学院,浙江 绍兴 312000; 2. 绍兴文理学院附属医院,浙江 绍兴 312000; 3. 绍兴北大信息技术科创中心,浙江 绍兴 312000

摘 要：目的 针对结肠镜息肉图像边界模糊、尺度变化范围大的问题,现有分割模型常因计算冗余、特征表征能力不足,难以兼顾分割精度与处理效率。为此,本文提出一种基于Mamba架构的多级多尺度特征融合增强网络(multi-level and multi-scale feature fusion enhanced network, MMF-MambaNet),通过轻量化网络架构与高效上下文感知机制,提升模型对息肉边界细节的捕捉能力。**方法** MMF-MambaNet以Mamba变体网络VSSM(visual state space model, VSSM)作为骨干网络,通过嵌入局部序列样式转换增强模块对特征均值与方差进行随机重参数化处理,有效增强骨干特征多样性,网络核心包含三大关键模块:一是引入跨尺度上下文自感知注意力(cross-scale aware self-attention,CASA),通过深层特征引导,使各层级特征在保留细节的同时融入高层语义上下文;二是构建自适应细节增强模块(adaptive detail enhancing module, ADEM),以某一尺度为锚点,融合其他尺度的结构信息,构建语义完整的息肉细节表示;三是构建自适应全局-局部融合门控(adaptive global-local integration gate, AGLI),以轻量结构实现细粒度语义引导,助力网络聚焦语义可靠、边界清晰息肉区域,提升分割精度。此外,模型设计反向特征融合路径,将编码器输出的深层全局语义特征与浅层细节特征有机结合,最终生成包含清晰边界线索的精细化分割掩膜。**结果** 在CVC-300、CVC-ClinicDB、Kvasir、CVC-ColonDB及ETIS-LaribPolypDB五大公开数据集上开展实验,将MMF-MambaNet与PraNet、VM-UNET-V2和Polyp-Mamba等8种医学图像分割领域的先进算法进行对比。在五个数据集上,与VM-UNET相比,mDice评估指标分别提升1.1%、5.0%、4.2%、9.1%和2.6%。此外,CVC-ColonDB数据集上的实验结果表明,与VM-UNET-V2相比,在mDice和mIoU上分别提升4.6%、6.2%。**结论** 本文提出的MMF-MambaNet克服了在处理息肉边界模糊、尺度多变问题时的息肉区域提取精度与效率失衡局限,为结肠镜息肉的精准分割提供了高效可行的模型支撑。

关键词： Mamba;多尺度;多级;息肉检测;深度学习

A polyp detection network with multi-level and multi-scale feature fusion enhancement based on Mamba architecture

Hu Keli[1,2], Huang Jie[1], Fan Kangxin[1], Wang Chen[1*], Zhu Hancan[2], Zhao Liping[1], Ren Haowei[2], Hu Jianhao[2]

1. School of Intelligent Engineering, Shaoxing University, Shaoxing 312000, China; 2. Affiliated Hospital of Shaoxing University, Shaox-

收稿日期:2025-10-05;**修回日期**:2026-03-18

* **通信作者**:王臣 wangchen3024@163.com

基金项目:浙江省自然科学基金重点项目(LZ24F020006);教育部人文社会科学研究青年基金项目(25YJCZH082);国家自然科学基金资助项目(62271321, 61603258)

Supported by: Zhejiang Provincial Natural Science Foundation of China (LZ24F020006); Humanities and Social Sciences YouthFoundation of Ministry of Education of China (25YJCZH082); National Natural Science Foundation of China (62271321, 61603258)

ing 312000, China

Abstract: Objective Accurate segmentation of colorectal polyps in colonoscopy images plays a crucial role in the early diagnosis and prevention of colorectal cancer. However, colonoscopy images often exhibit blurred boundaries, irregular shapes, and large scale variations, which pose significant challenges for automated segmentation. Conventional convolutional neural network (CNN)-based methods are limited in modeling long-range dependencies, often leading to inaccurate boundary localization and missed detection of small lesions. Transformer-based models improve global context modeling but suffer from quadratic computational complexity and high memory consumption when processing high-resolution medical images. Recently, the Mamba architecture based on selective state space models has shown strong potential for efficient sequence modeling with linear complexity. Nevertheless, directly applying Mamba-based models to polyp segmentation remains challenging due to insufficient local feature extraction and limited multi-scale feature interaction. To address these issues, this paper proposes a Multi-level and Multi-scale Feature Fusion Enhanced Network based on the Mamba architecture (MMF-MambaNet) to improve segmentation accuracy while maintaining computational efficiency. **Methods** MMF-MambaNet adopts an encoder-decoder architecture and employs the Visual State Space Model (VSSM), a variant of Mamba, as the backbone to efficiently capture long-range contextual dependencies. To enhance feature diversity and robustness, a Local Sequence-wise Style Transformation Augmentation (LSA) module is integrated into the backbone. This module performs stochastic re-parameterization by perturbing the mean and variance of feature distributions, thereby improving model generalization under varying imaging conditions. To strengthen multi-scale representation and boundary perception, three specialized modules are incorporated. First, a Cross-scale Aware Self-Attention (CASA) module is introduced to facilitate hierarchical feature interaction. CASA uses deep semantic features to guide shallow feature refinement and integrates channel and spatial attention mechanisms to effectively fuse multi-level features while preserving fine boundary details. Second, an Adaptive Detail Enhancing Module (ADEM) is designed to improve structural representation across scales. Using an anchor-scale fusion strategy, this module adaptively integrates information from different scales to construct more complete detail representations of polyps, thereby enhancing boundary structures and improving segmentation of small lesions. Third, an Adaptive Global-Local Integration (AGLI) gating module is introduced in the decoder stage. This lightweight mechanism adaptively integrates global semantic features with local detailed features from skip connections, enabling the model to emphasize reliable regions while suppressing background interference. Furthermore, MMF-MambaNet introduces a reverse feature fusion pathway to combine deep semantic features with shallow spatial information through progressive upsampling and feature aggregation. A pyramid pooling module (PPM) is also employed to capture multi-scale global contextual information, improving semantic consistency in the segmentation results. The model is optimized using a hybrid Binary Cross-Entropy and Dice loss, which balances pixel-level classification accuracy and region-level overlap while alleviating class imbalance in medical image segmentation. **Results** Extensive experiments were conducted on five publicly available colorectal polyp segmentation datasets: CVC-300, CVC-ClinicDB, Kvasir-SEG, CVC-ColonDB, and ETIS-LaribPolypDB. The proposed MMF-MambaNet MMF-MambaNet was compared against eight state-of-the-art segmentation methods, including PraNet, VM-UNET-V2, and Polyp-Mamba. Experimental results demonstrate that MMF-MambaNet achieves competitive and stable performance across multiple datasets. Compared with the VM-UNet baseline, MMF-MambaNet improves the mean Dice coefficient (mDice) by 1.1%, 5.0%, 4.2%, 9.1%, and 2.6% on the CVC-300, CVC-ClinicDB, Kvasir-SEG, CVC-ColonDB, and ETIS-LaribPolypDB datasets, respectively. On the challenging CVC-ColonDB dataset, the proposed model further surpasses VM-UNetV2 with improvements of 4.6% in mDice and 6.2% in mean Intersection over Union (mIoU). Visual comparisons show that MMF-MambaNet can accurately delineate polyp boundaries even in cases with low contrast, irregular shapes, or small lesion sizes. In addition to segmentation accuracy, the model demonstrates favorable computational efficiency. With 18.84M parameters and 5.74 GFLOPs, MMF-MambaNet maintains a relatively lightweight architecture while achieving an inference speed of 57.58 frames per second. Ablation studies further confirm the effectiveness of the proposed modules, showing that CASA, ADEM, and AGLI each contribute to improving segmentation accuracy and boundary preservation. **Conclusion** This study presents MMF-MambaNet, a Mamba-based network that integrates multi-level and multi-scale feature fusion mechanisms for colorectal

polyp segmentation. By combining efficient sequence modeling with adaptive feature interaction strategies, the proposed approach effectively addresses the challenges of blurred boundaries and scale variations in colonoscopy images. Experimental results demonstrate that MMF-MambaNet achieves strong segmentation performance while maintaining lightweight computational complexity, making it a promising solution for computer-aided colonoscopy diagnosis and assisting early detection of colorectal cancer.

Key words: Mamba; Multi-scale; Multi-level; Polyp Detection; Deep Learning

引言

结肠癌是常见的恶性肿瘤之一，其早期诊断和治疗对患者的生存率和生活质量具有重要意义(FY等，2019)。然而，在结肠镜检查及手术领域，图像反馈的病灶区自动定位问题仍然是一个亟待解决的难题。目前，结肠镜检查主要依赖医生的经验和肉眼观察，存在以下问题：结肠镜操作过程中，病灶区域定位不准确，容易导致漏诊和误诊(YuT等，2022)；人工定位病灶区域耗时较长，增加患者痛苦和手术风险；结肠镜图像反馈存在一定局限性，无法精确勾勒病灶区域边界(De等，2023)。

早期探索阶段，Jerebko等采用Canny边缘检测器来分割计算机断层扫描(CT)结肠造影中的息肉候选图像(Jerebko等，2003)。Yao等提出了一种基于知识引导的强度调整和模糊C均值聚类相结合的方法(Yao等，2004)，用于CT结肠成像中自动分割结肠息肉，引入了更多的图像处理技巧，以提高分割的准确性。Gross等提出了第一个针对结肠镜窄带图像的息肉自动分割算法(Gross等，2009)。Hwang等提出了一种无监督方法(Hwang等，2010)，用于在无线胶囊内窥镜视频中检测息肉。

深度学习在计算机视觉领域的突破性进展为结肠息肉分割带来了新的机遇(吴琪琪等，2026)。研究者们开始将深度学习模型应用于息肉分割任务。Ronneberger等提出了一种用于生物医学图像分割的UNet网络(Ronneberger等，2015)，该网络在ISBI细胞追踪挑战赛中取得了较好的成绩，为后续的息肉分割研究提供了新的框架。Tajbakhsh等提取图像边缘图(Tajbakhsh等，2016)，依托边缘信息区分息肉与非息肉结构，剔除非息肉边缘后，结合形状特征实现息肉区域的定位。Fang等提出了一个具有息肉面积和边界约束的选择性特征聚合网络(Fang等，2019)，进一步提高了分割的精度。Fan等提出了一种平行反向注意网络(Fan等，2020)，用于结肠镜图像中的息肉准确分割，该方法强调了上下文信息的利用。Zhang等提出了一种基于Transformer的并行分支架构(Zhang等，2021)，将Transformer和CNN相结合，为息肉分割提供了新的视角。Ling等提出了一种高斯概率引导的语义融合方法(Ling等，2023)，该方法通过逐步融合息肉位置的概率信息与解码器的输出，进一步提升了分割性能。Albert等提出了Mamba模型(Albert等，2023)，通过结合选择性状态空间SSM (selective state space, SSM)和Transformer架构，使其在自然语言处理、时间序列分析等多个领域具有广泛的应用潜力。

目前，基于CNN(convolutional neural networks, CNN)和Transformer的医学图像分割面临着许多挑战。比如CNN在长距离建模能力上存在不足，而Transformer则受到其二次计算复杂度的制约。Mamba模型因其在语言处理、基因组学和音频分析等各个领域的应用而脱颖而出。Mamba模型采用线性时间序列建模架构，结合了选择性状态空间，可在不同模式中提供较好性能。Mamba的主要优势之一是能够应对传统Transformer在处理长序列时的计算复杂度挑战。通过将选择机制集成到其状态空间模型中，Mamba可以根据序列中每个token的相关性有效地决定是传播还是丢弃信息。该选择性方法可加快推理速度，吞吐率比标准Transformer高出五倍。现有模型在结肠镜息肉分割任务中仍存痛点，CNN存在长距离建模能力不足的缺陷，易造成微小息肉漏检、病灶边缘定位模糊，难以满足临床精度需求；Transformer受二次计算复杂度制约，在处理高分辨率结肠镜图像时，计算效率低下且显存消耗过大，不利于常规硬件部署；Mamba虽具备线性复杂度与高效长序列建模优势，但直接迁移至该任务仍面临挑战，包括局部特征提取薄弱、显存占用较高等。结合息肉分割任务特性和Mamba架构优势进行针对性网络结构设计，是实现息肉分割性能优化的关键。

本文贡献概括如下:1)提出了一种基于Mamba架构的多级多尺度特征融合增强的息肉检测网络MMF-MambaNet,该网络引入经区域特征强化的VSSM主干,同时结合跨尺度上下文自感知注意力(CASA),以及自适应全局-局部融合门控(AGLI),实现边界细节与语义信息的协同优化。2)设计了反向特征融合路径,通过上采样聚合编码器深层全局特征,并融合浅层细节信息,有效解决息肉边界模糊与尺度变化问题,提升分割掩膜的精细化程度。3)在CVC-300、CVC-ClinicDB、Kvasir、CVC-ColonDB及ETIS-LaribPolypDB多个数据集上的实验表明,MMF-MambaNet在平均Dice(mDice)和平均交并比(mIoU)上优于现有先进方法,且模型更轻量,验证了其在结肠息肉分割任务中的有效性。

1 相关工作

Ruan等提出的VM-UNet(Ruan等,2024),将Mamba结构融入UNet的模型,VM-UNet引入了视觉态空间(VSS)块作为基础块以捕捉广泛的上下文信息,并构建了一个非对称的编码器-解码器结构,在ISIC17、ISIC18和Synapse数据集上超越UNet++/UNetv2。Wu等提出了一种用于处理深度特征的并行视觉Mamba策略(Wu等,2024),在保持总体处理通道数不变的同时,以最低的计算负载实现了出色的性能。Wang等在U-Net架构中引入了Visual-Mamba块(Wang等,2024),以改善医学图像分析中的远距离依赖建模,这提供了一种新的方法来处理长序列数据,并在医学图像分割领域中取得了优越的性能。

由于Mamba模块的提出,进而激发了一系列基于Mamba的混合模型的研究(刘建明等,2026)。Ma等提出了医学图像分割框架U-Mamba(Ma等,2024),其采用CNN与SSM混合块,结合自配置机制,成功用于3D腹部器官、内窥镜图像等分割任务,去得了较好结果。Hatamizadeh等提出了MambaVision(Hatamizadeh等,2024),是一种新颖的混合Mamba-Transformer架构,提高了原始Mamba架构的准确性和图像吞吐量,提高了模型捕获全局上下文和长距离空间依赖的能力。WangZ等提出了一种弱监督学习框架Weak-Mamba-Unet(Wang等,2024),该框架利用了卷积神经网络(CNN)、视觉Transformer(ViT)和VisualMamba(VMamba)架构,用于医学图像分割,其在MRI心脏分割数据集的Dice系数达到0.9171,准确率达到0.9963。Gong等提出nnMamba(Gong等,2025),其通过融合CNN局部特征提取与SSM全局上下文建模优势,用于分割、分类及检测任务,在BraTS2023等6个数据集上表现优异。

在结肠镜息肉分割领域,众多基于Mamba的衍生模型不断涌现。Xu等人提出的Polyp-Mamba通过融合多频特征感知技术与门控选择机制,完成了跨尺度的语义特征精准提取(Xu等,2024),在五个息肉分割数据集上取得了领先性能。Xie等提出的ProMamba将提示学习引入视觉Mamba框架之中(Xie等,2024),有效提升了模型的泛化能力,提高了分割精度,为模型性能的优化提供了新的思路。VM-Unet(Ruan等,2024)和VM-UNet-v2(Zhang等,2024)模型将Mamba框架强大的序列表示能力,有机地融入经典的Unet编码解码架构中,进而展现出较好的分割性能。Wang等人提出的S^3-Mamba通过融合增强视觉状态空间块、张量基跨特征多尺度注意力与正则化学习策略(Wang等,2025),实现了对小尺寸病变的敏感性提升与精准特征提取,在CVC-ClinicDB息肉数据集上表现优异。Jiao等人提出的TM-UNet通过引入令牌记忆(Token-Memory)模块与Mamba增强型序列建模机制(Jiao等,2025),有效解决了长序列特征建模效率低、关键病灶特征记忆不持久等难题,在包括息肉分割在内的多个医学图像分割数据集上展现出高效且精准的分割性能。Dutta等提出SAM-Mamba(Dutta等,2025),结合SAM与Mamba,引入含MSD和Mamba块的Mamba-Prior模块,提升息肉分割的全局上下文捕捉与零样本泛化能力。Li等提出含OD表示的边界感知模型(Li等,2025),结合方向导数与不确定边界图,提升息肉分割边界精度。

2 方 法

MMF-MambaNet整体采用编码器解码器结构,结合多尺度特征融合与注意力机制实现结肠息肉分割。基于VMamba编码器主体架构,Cheng等将局部序列样式转换增强模块LSA(local sequence-wise style transformation augmentation, LSA)引入并替换

VMamba中的SS2D模块，构建了VSSDG模块（Cheng等，2025）。LSA对特征均值与方差进行随机重参数化处理，有效增强了网络骨干特征多样性。MMF-MambaNet首先将输入图像划分为补丁并嵌入特征，再经多层VSSDG模块逐级提取多尺度特征。对编码器输出的多尺度特征，先通过跨尺度上下文自感知注意力（CASA）模块，利用深层特征上下文优化浅层特征；再经Translayer降维后，通过自适应细节增强模块（ADEM）进行多尺度特征融合，生成不同分辨率的融合特征图。解码器部分由多层解码模块组成，结合金字塔池化模块（PPM）增强全局上下文，通过上采样逐步恢复特征图分辨率，并利用自适应全局-局部融合门控（AGLI）强化跳跃连接特征融合，最终通过分类器输出分割结果。

总体上，MMF-MambaNet基于Mamba架构的高效序列建模能力，通过联合局部扰动特征增强、多尺度注意力机制，以及多级多尺度特征提取与融合优势，实现对息肉区域精准定位。

2.1 编码器

MMF-MambaNet编码器以VMamba架构为基础，通过堆叠VSSDG块构建下采样层级结构。每个编码层对输入特征的处理包含两个关键维度：借助状态空间模型单元捕捉长距离依赖关系与全局上下文，同时依赖卷积操作提取局部细节特征并实现层间传播。VSSDG内的LSA通过随机生成掩码对特征局部区域进行扰动混合，在保留整体特征分布的同时引入局部随机性，增强了特征的泛化能力与模型的抗干扰性，这一设计使得不同尺度特征在信息交互与融合过程中能学习到更稳健的跨尺度关联模式。

与传统UNet结构不同，该编码器在特征提取过程中，先通过VSSDG生成从高分辨率浅层到低分辨率深层的多尺度特征f_1、f_2、f_3、f_4（其中f_1富含细节信息，f_4蕴含全局语义），随后在编码阶段末端引入CASA模块实施自顶向下的特征优化：以最深层特征f_4为起点，依次指导f_3、f_2的优化，最终作用于f_1。通过跨尺度特征注入与注意力机制的结合，浅层特征在保留局部边缘信息的同时，持续吸收深层语义上下文，而LSA模块带来的局部扰动进一步促进了不同尺度特征表征的一致性与判别性，为后续解码过程提供了经过多维度优化的特征基础，MMF-MambaNet整体网络结构如图1所示。

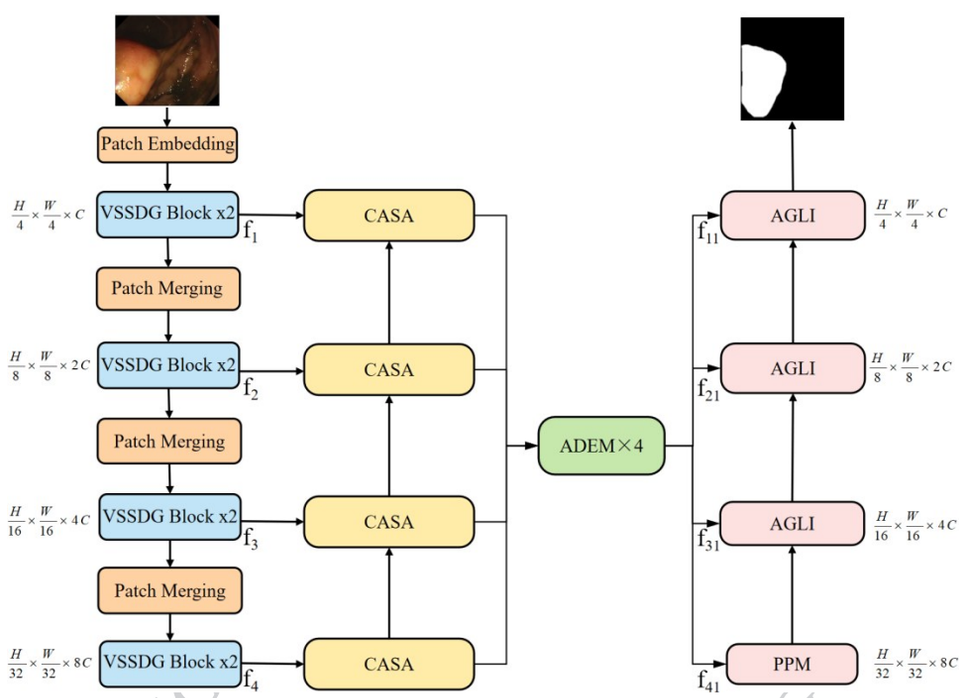

图1 MMF-MambaNet模型结构

2.1.1 跨尺度自注意力模块（CASA）

在医学图像分割中，网络不同层级特征存在天然互补性：浅层特征富含位置与边界信息但语义表达较弱，深层特征具备丰富全局语义却空间分辨率较低，难以精准定位边缘。如何实现多尺度特征的有效融合，是编码器设计的核心难题。跨尺度感知

自注意力(CASA)模块如图2所示,CASA模块结合CBAM(Woo等,2018)通道注意力与空间注意力的串行计算框架轻量化设计理念,针对多尺度特征融合痛点,扩展了多尺度逐层反馈融合的模块,采用独立的1x1卷积替代共享MLP机制,旨在通过强化不同池化特征,结合深层语义特征逐层引导浅层特征学习,使各层级特征在保留细节的同时融入高层语义上下文,从而提升分割结果的语义一致性与边缘精度。

与Polyp-Mamba中SAS模块直接拼接多尺度特征的处理方式不同,CASA采用逐层反馈策略,每层引入更深层特征作为辅助信息,经通道与空间分辨率对齐后融合,避免了语义混杂与冗余,实现结构化的递进式跨尺度语义增强。SAS模块直接拼接不同层级特征,未对语义层级差异做筛选处理,易因浅层细节与深层语义的直接耦合引发信息冗余。同时,CASA集成通道与空间注意力机制,通道注意力动态调节语义通道重要性,空间注意力聚焦息肉边缘等关键区域,即使最深层也能通过自注意力精炼特征。SAS模块则主要依赖VSS块内嵌的Mamba隐式序列注意力,未针对息肉分割任务设计专用的空间通道注意力分支,对微小息肉和病灶边界的关注度不足。CASA先融合当前特征与深层上下文,再依次施加通道与空间注意力,通过双重加权强化特征关联性与任务相关性。CASA模块的输出为经过跨尺度融合和双重注意力加权的特征。

$$\boldsymbol{M}_c = \sigma_{sig}\left(\boldsymbol{W}_c(x + c)\right) \tag{1}$$

$$\boldsymbol{M}_s = \sigma_{sig}\left(\boldsymbol{W}_s\left((x + c)\odot\boldsymbol{M}_c\right)\right) \tag{2}$$

$$y = (x + c)\odot\boldsymbol{M}_c\odot\boldsymbol{M}_s \tag{3}$$

式中W_c为通道注意力向量,W_s为空间注意力向量,$\sigma_{sig}(\cdot)$为sigmoid函数,$\odot$为逐元素乘积,c是更深层的上下文特征(High Level Feature),x是当前尺度的特征(Low Level Feature),M_c是通道权重图,M_s是空间权重图。

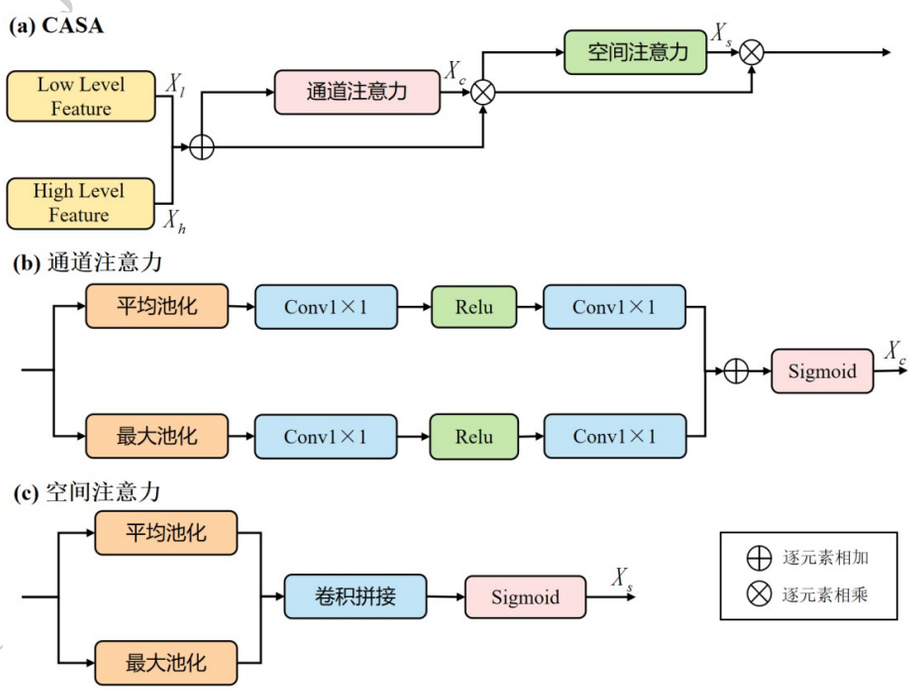

图2 CASA模块结构

2.1.2 局部序列样式转换增强(LSA)

LSA(局部序列风格变换增强模块)针对Mamba中选择性扫描机制的特性,将Mamba模块中的VSS块改进为VSSDG块(Cheng等,2025),通过以LSA模块替代原有的SS2D操作,显式抑制Mamba输入相关矩阵中的领域特定特征。该模块旨在通过序列级局部风格增强,进一步提升模型的泛化能力。LSA模块如图3所示,主要包含扫描扩展(Scan Expanding)、选择性状态空间模型(SSMs)模块、风格增强操作(Style Augmentation)、序列混合(Sequence-wise Mixup)以及扫描合并(Scan Merging)五部分组件,对每个序列从四个方向实施序列级风格增强。

具体操作流程如下：从Mamba中间层获取的特征图，初始维度为$R^{B\times C\times W\times H}$（B=批次大小、C=通道数、W=宽度、H=高度），通过扫描扩展将特征图重塑为1D序列$f_h(x)\in R^{B\times D\times L}$（D=序列特征维度，由通道数C转换而来；L=序列长度，由W×H计算得到），使特征适配Mamba的序列处理模式。将$f_h(x)$输入选择性SSM模块，利用Mamba的输入依赖矩阵和选择性扫描机制，捕捉序列内的长程语义依赖，输出经过依赖建模的序列特征$f_h(x)$（维度保持$R^{B\times D\times L}$）。然后对每个样本x_i的序列特征$f_h(x_i)$，按通道计算均值$\mu(f_h(x_i))$和标准差$\sigma(f_h(x_i))$。

$$\mu\left(f_h(x_i)\right)=\frac{1}{L}\Sigma_{l=1}^{L}f_h(x_i)^l \tag{4}$$

$$\sigma\left(f_h(x_i)\right)=\sqrt{\frac{l}{L}\sum_{l=1}^{L}[f_h(x_i)^l-\mu\left(f_h(x_i)\right)]^2+\varepsilon} \tag{5}$$

式中l为序列内元素的位置索引（取值范围1~L），$f_h(x_i)^l$表示样本x_i序列特征中第l个元素；ϵ为平滑系数（默认取值10^{-6}）。基于批次内所有样本的统计量估计特征均值和标准差的不确定性$\sum_{\mu}^{2}(f_h(x))$和$\sum_{\sigma}^{2}(f_h(x))$，量化领域偏移导致的统计量波动。

$$\sum\nolimits_{\mu}^{2}(f_h(x))=\frac{1}{B}\sum_{i=1}^{B}\left(\mu\left(f_h(x_i)\right)-\mathbb{E}\left[\mu\left(f_h(x_i)\right)\right]\right)^2 \tag{6}$$

$$\sum\nolimits_{\sigma}^{2}(f_h(x))=\frac{1}{B}\sum_{i=1}^{B}\left(\sigma\left(f_h(x_i)\right)-\mathbb{E}\left[\sigma\left(f_h(x_i)\right)\right]\right)^2 \tag{7}$$

式中B为批次大小。基于高斯分布$\epsilon_\mu\sim(0,1)$、$\epsilon_\sigma\sim(0,1)$生成带随机扰动的新统计量$\beta(f_h(x_i))$、$\gamma(f_h(x_i))$，通过在原始均值μ和标准差σ上叠加高斯分布扰动，来模拟不同领域的风格差异。

$$\beta\left(f_h(x_i)\right)=\mu\left(f_h(x_i)\right)+\int_{\mu}\sum\nolimits_{\mu}(f_h(x)),\varepsilon_\mu\sim\mathbb{N}(0,1) \tag{8}$$

$$\gamma\left(f_h(x_i)\right)=\sigma\left(f_h(x_i)\right)+\varepsilon_\sigma\sum\nolimits_{\sigma}(f_h(x)),\varepsilon_\sigma\sim\mathbb{N}(0,1) \tag{9}$$

用新统计量β和γ替换原始特征的统计量，得到风格增强后的序列特征$f_h^{aug}(x_i)$

$$f_h^{aug}(x_i)=\beta\left(f_h(x_i)\right)+\left(\frac{f_h(x_i)-\mu\left(f_h(x_i)\right)}{\sigma\left(f_h(x_i)\right)}\right)\cdot\gamma\left(f_h(x_i)\right) \tag{10}$$

完成特征风格变换后，对每个样本x_i生成随机掩码$\mathcal{M}_i$。

$$\boldsymbol{M}_{i,j}^{mask}=\mathbb{I}\left(0\leqslant\frac{j-j_{start}}{L}<P\right),j\in\{1,2,\cdots,L\} \tag{11}$$

式中j为序列位置，j_{start}为随机起始位置，P为掩码比例，默认0.75。

将掩码扩展至$R^{D\times L}$（与特征维度匹配），通过元素级乘法混合$f_h^{aug}(x_i)$和$f_h(x_i)$，得到最终局部增强特征$f_h^{mixed}(x_i)$。

$$f_h^{mixed}(x_i)=f_h^{aug}(x_i)\odot\boldsymbol{M}_i^{mask}+f_h(x_i)\odot(1-\boldsymbol{M}_i^{mask}) \tag{12}$$

式中$\odot$表示逐元素乘积，用于将掩码与特征图按像素位置相乘。最后通过扫描合并将$f_h^{mixed}(x_i)$从1D序列$R^{B\times D\times L}$重塑回2D空间特征图$R^{B\times C\times W\times H}$。

为保持标记依赖关系，采用连续序列扰动替代随机像素选择，确保特征结构完整性和语义关系的稳定性。该模块通过对局部随机连续子序列的特征统计量进行多元高斯分布建模，生成具有多样性的特征表示，从而增强模型对不同域偏移的鲁棒性。

2.1.3 自适应细节增强模块（ADEM）

ADEM如图4所示，其核心思想是以某一尺度为锚点，融合其他尺度的结构信息，构建语义完整的细节表示。区别于传统特征拼接或固定加权方式，ADEM采用锚点驱动和多源注入相融合的策略，对于多尺度特征列表f_1-f_4，指定某一尺度为锚定尺度，其余尺度特征经分辨率对齐（上/下采样）、卷积处理后生成细节掩码，通过掩码与锚定特征的动态交互实现跨尺度融合。

具体而言，ADEM首先通过平行3×3卷积预处理各尺度特征。如图4(b)空间自适应细节调制子模块（spatial adaptive detail modulation，SADM），对非锚定特征，调整其分辨率至锚定尺度，经卷积和Sigmoid激活生成细节掩码$\mathcal{M}_i$，刻画该尺度在锚定尺度下的位置显著性。锚定特征X_a经自身卷积$\mathcal{W}_a$后，与所有非锚定尺度的掩码逐元素相乘，最终输出融合特征$\mathcal{Y}$，其数学表达式为：

$$\boldsymbol{Y}_a=\boldsymbol{W}_a(\boldsymbol{X}_a) \tag{13}$$

$$\boldsymbol{M}_i=\sigma_{sig}\left(\boldsymbol{W}_i\left(X_i^{(a)}\right)\right),i\neq a \tag{14}$$

$$\boldsymbol{Y}=\boldsymbol{Y}_a\prod_{i\neq a}\boldsymbol{M}_i \tag{15}$$

式中$\mathcal{Y}_a$为锚定尺度特征经专属卷积核精炼后的基础特征，a为锚定尺度索引，$X_i^{(a)}$为对齐至锚定尺度的

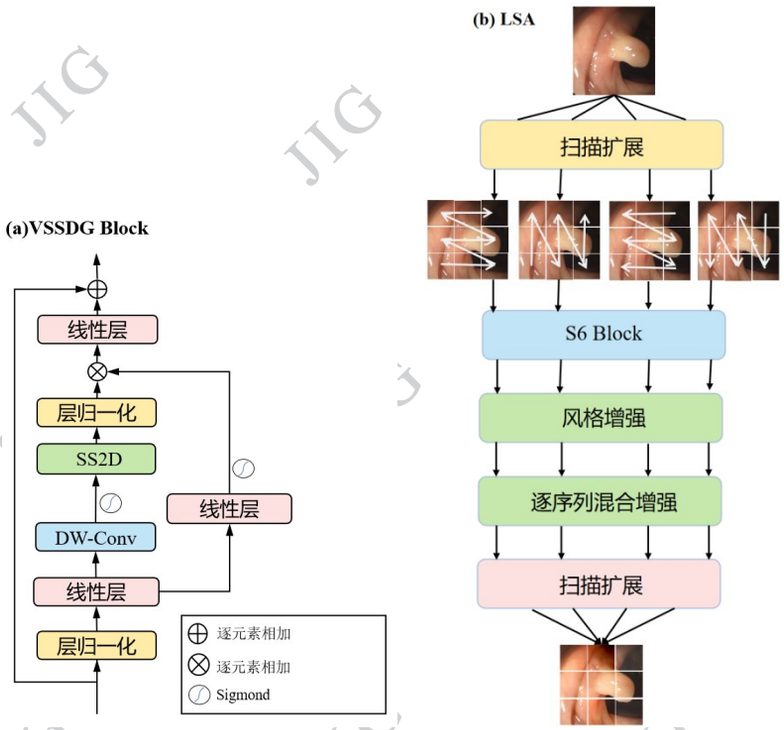

图3　VSSDG Block结构图

特征。

ADEM具备双向信息流动能力:浅层高分辨率锚定时,深层特征提供语义引导,强化语义一致区域;深层锚定时,浅层特征注入边缘线索,弥补定位不足。该模块通过学习动态细节掩码,避免固定结构的局限性,实现按需增强,突出目标边界与细小结构,抑制无关冗余。实验表明,ADEM显著提升模型对复杂结构和遮挡场景的边缘保留与细粒度恢复能力,增强分割结果的边界精度与结构一致性。

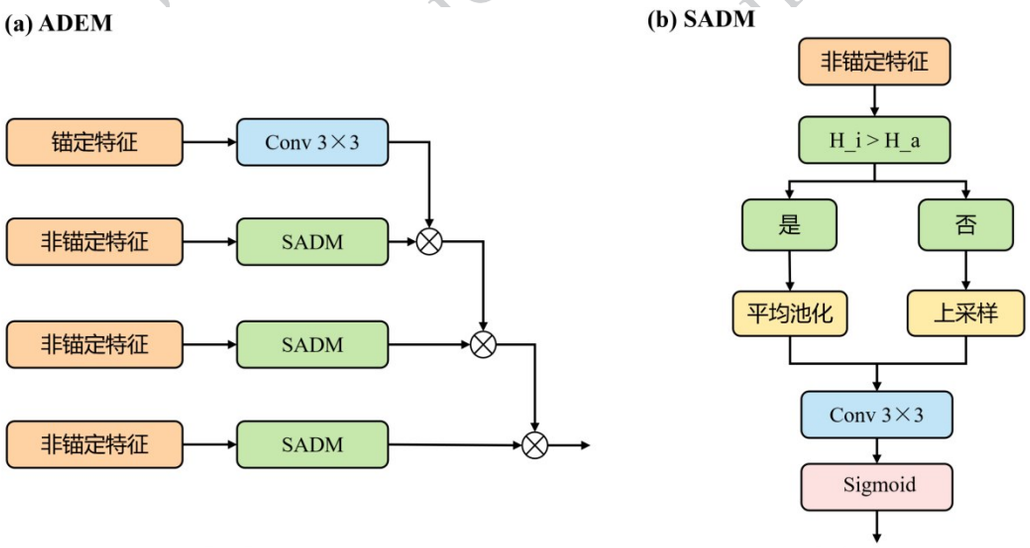

图4　ADEM模块结构图

2.2　解码器

该模型的解码器部分采用层级化设计,旨在将编码器输出的多尺度特征逐步融合并恢复空间分辨率,最终生成精确的分割结果。

首先,解码器接收经CASA模块优化及Trans-layer降维后的多尺度特征(f_1、f_2、f_3、f_4),通过ADEM

模块以锚定融合策略整合跨尺度信息，生成四个不同分辨率的融合特征（f_{11}、f_{21}、f_{31}、f_{41}）。其中，最深层融合特征f_{41}会先经过金字塔池化模块（PPM），通过多尺度池化操作捕获全局上下文，增强语义一致性。

随后，解码器采用四级DecoderBlock进行逐步上采样：以池化PPM（pyramid pooling module，PPM）处理后的f_{41}为起点，依次与f_{31}、f_{21}、f_{11}通过AGLI门控机制融合浅层细节特征，每级模块通过转置卷积实现2倍分辨率提升，并结合卷积层与注意力机制精炼特征。最终通过final_conv输出分割结果，同时引入深监督机制，对中间特征进行监督并加权融合，进一步提升分割精度。

整个解码器通过多尺度融合、全局上下文建模及注意力机制的结合，在恢复空间细节的同时保持语义一致性，有效处理复杂医学图像中目标边界模糊、尺度多变等问题。

2.2.2　金字塔池化模块（PPM）

解码器中的金字塔池化模块（PPM）旨在通过多尺度上下文建模增强特征的全局语义感知能力（Zhao等，2017）。该模块包含多个并行分支，每个分支对应不同尺度的自适应平均池化操作（如bins=（1，2，3，6）），将输入特征图分别池化到不同分辨率，以捕获从全局到局部的多层次上下文信息。每个分支经1×1卷积压缩通道维度、批量归一化和ReLU激活后，通过双线性插值上采样至输入特征尺寸，与原始特征图拼接形成多尺度特征融合表示。最后，通过1×1卷积层将通道数恢复为输入维度，实现对不同尺度目标结构的语义一致性建模。这一设计有效弥补了深层特征因空间分辨率降低导致的边缘定位不足问题，通过整合全局语义线索与局部细节特征，显著提升模型对复杂场景中目标区域的分割精度与结构完整性。

2.2.3　自适应全局-局部融合模块（AGLI）

在解码阶段，如何有效整合编码器的细节特征与解码层的语义特征是关键挑战。为此引入AGLI，如图5所示，其设计借鉴AttentionU-Net的门控机制并优化结构（Oktay等，2018），旨在通过注意力机制实现跳跃连接中语义与细节的动态融合。

AGLI接收编码器的局部细节特征X_{local}与解码层的全局语义特征X_{global}，首先通过1×1卷积和批量归一化（BN）对两支输入进行维度统一与降维，得到投影后的特征$\hat{X}_g$和$\hat{X}_l$。将二者相加并经ReLU激活后，再通过1×1卷积和Sigmoid函数生成单通道注意力权重图Ψ（取值范围0-1），用于刻画局部特征各位置的重要性。最终融合输出，计算公式表示为：

$$\hat{\boldsymbol{X}}_{\mathrm{g}}=\mathrm{BN}\left(Conv_{1\times1}^{(\mathrm{g})}\left(\boldsymbol{X}_{\mathrm{global}}\right)\right) \tag{16}$$

$$\hat{\boldsymbol{X}}_{l}=\mathrm{BN}\left(Conv_{1\times1}^{(1)}\left(\boldsymbol{X}_{\mathrm{local}}\right)\right) \tag{17}$$

$$h=\mathrm{ReLu}\left(\hat{\boldsymbol{X}}_{\mathrm{g}}+\hat{\boldsymbol{X}}_{1}\right) \tag{18}$$

$$\boldsymbol{\Psi}=\sigma_{\mathrm{sig}}\left(Conv_{1\times1}^{(\psi)}(h)\right),0\leqslant\boldsymbol{\Psi}\leqslant1 \tag{19}$$

$$\boldsymbol{Y}=\boldsymbol{X}_{\mathrm{global}}+\boldsymbol{\Psi}\odot\boldsymbol{X}_{\mathrm{local}} \tag{20}$$

式中⊙表示逐元素相乘，通过权重图选择性保留与全局语义一致的局部细节，抑制冲突区域的冗余信息。

AGLI门控机制可自适应调节局部与全局特征融合比例。语义一致时增强细节保留，反之抑制干扰，提升目标边缘重建准确性与语义结构一致性；其以轻量结构实现细粒度语义引导，弥合特征融合的语义-细节断层，助力解码器聚焦语义可靠、边界清晰区域，提升分割目标精度与边缘感知能力。

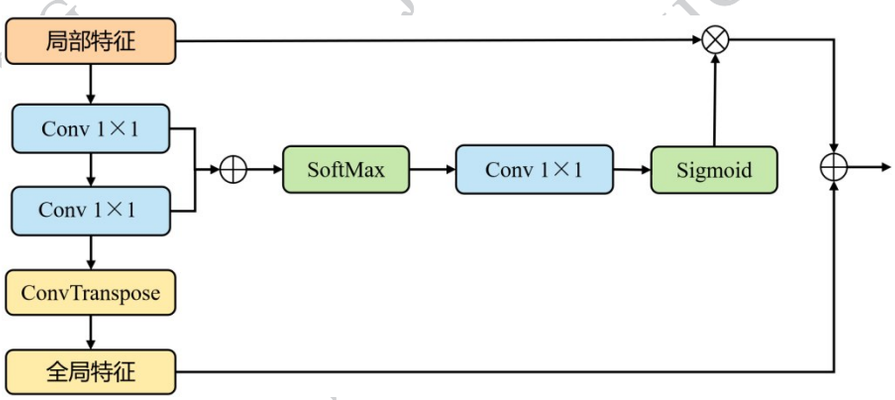

图5　AGLI模块结构图

2.3 损失函数

本文采用二元交叉熵-骰子组合损失(BceDiceLoss)对息肉掩码进行监督,该损失函数融合了二元交叉熵BCE(binary cross entropy, BCE)损失和骰子(Dice)损失的优势,能够同时优化像素级概率分布和区域级空间重叠度。在息肉分割任务中,该损失可有效处理背景与病灶间的类别不平衡问题,并提升小病灶区域的分割精度。总损失定义为:

$$L_{local} = W_{\mathrm{b}} \cdot L_{\mathrm{Bce}}(P,G) + W_{\mathrm{d}} \cdot L_{\mathrm{Dice}}(P,G) \quad (21)$$

式中L_{Bce}和L_{Dice}分别表示二元交叉熵损失和骰子损失。Bce、DiceLoss通过权重参数$\mathcal{W}_b$和$\mathcal{W}_d$调节两种损失的贡献比例(实验中均设置为1)。其中L_{Bce}通过衡量每个像素的预测概率与真实标签的差异,引导模型学习像素级类别分布;L_{Dice}基于预测区域与真实区域的重叠度计算,更关注小病灶区域的分割完整性。P是模型输出的息肉掩码预测结果,G为对应的真实标签掩码。

3 实 验

3.1 实验数据集

本文采用该领域研究国际公认的五个公开的结直肠息肉分割数据集进行实验分析。数据集基本信息和数据划分方式如表1所示。其中CVC-ClinicDB数据集中的550张图像和Kvasir-SEG数据集中的900张图像作为训练样本,其余图像作为测试样本。CVC-ClinicDB数据集由23名患者的31段序列中提取得到,共含612张分辨率为384×288的息肉图像。本研究选取其中550张用于训练,其余作为测试集。Kvasir-SEG数据集包含1000张息肉图像,分辨率范围从334×487至1920×1070,其中900张用于训练,剩余用于测试。此外,CVC-ColonDB与CVC-300分别包含380张和300张息肉图像,分辨率均为574×500。ETIS-LaribPolypDB数据集由44个不同息肉的34个视频序列中提取的196张图像组成,分辨率统一为1966×966。该数据集中息肉形态复杂,部分目标尺寸较小,且边界与周围组织高度相似,因而更具挑战性。

3.2 实验环境及参数设置

本文方法基于PyTorch实现,所有实验均基于Linux操作系统(Ubuntu 20.04 LTS)完成,CUDA版本为12.4,PyTorch版本为2.1.0,Mamba-SSM版本为1.2.0。使用NVIDIA GeForce RTX 4090 GPU(24GB显存)训练模型。训练过程采用AdamW优化器,初始学习率设置为2e-4,采用1e-4权重衰减以防止过拟合。实验中,图像尺寸统一标准化为352×352像素,数据增强策略包括随机水平翻转、随机垂直翻转、随机裁剪及随机亮度/对比度调整,批大小设置为16,整体训练过程持续300轮。

表1 数据集信息

Table 1 Dataset information

Datasets	图像数量	训练集	测试集	验证集
CVC-300	300	0	300	300
CVC_ClinicDB	612	550	62	62
Kvasir-SEG	1000	900	100	100
CVC-ColonDB	380	0	380	380
ETIS-LaribPolypDB	196	0	196	196

3.3 评价指标

本文采用平均IoU(mIoU)和平均Dice(mDice)作为评价指标。Dice系数常用于衡量预测结果与真实标注在像素级的重叠程度,其值通过交集像素的两倍与两者总像素数之比计算得到。与此不同,IoU更关注分割区域整体的一致性,通过交叠区域占并集的比例来评价模型性能。

3.4 不同模型的结果比较

本文LP-MambaNet模型与8种先进方法进行比较,包括MSNet(Wu等,2020)、PraNet(Fan等,2020)、Polyp-PVT(Dong等,2023)、Polyp-Mamba(Xu等,2024)、VM-Unet(Ruan等,2024)、VM-UNetV2(Zhang等,2024)、SAM-Mamba(Dutta等,2025)、SSformer+OD(Li等,2025)。其中Polyp-Mamba、VM-Unet、VM-UNetV2以及SAM-Mamba是基于mamba的算法模型。MSNet通过设计减法单元、金字塔多尺度聚合及LossNet监督解决特征融合冗余问题;PraNet通过并行部分解码器(PPD)聚合高层特征并生成全局地图、反向注意力(RA)模块挖掘边界线索,以准确分割结肠镜图像息肉;Polyp-PVT采用Transformer编码器,并引入级联融合模块(CFM)、伪装识别模块(CIM)和相似度聚合模块(SAM),优化跨层特征融合、抑制特征噪音、提升表征能力;Polyp-Mamba基于Mamba(状态空间模型),通过尺度感知语义模块(SAS)捕捉跨尺度依赖性、全局语义注入

模块(GSI)缩小全局与局部语义差距,并借SAS实现跨尺度特征融合;VM-Unet首次提出纯基于状态空间模型(SSM)的U形架构,以视觉状态空间(VSS)块为基础模块用于医学图像分割;VM-UNetV2与VM-Unet相比跳跃连接变成了SDI,增强低高层特征融合;SAM-Mamba在SAM编码器中引入Mamba-Prior模块以提升分割精度与零样本泛化能力;SSformer+OD在SSformer模型基础上引入了可学习的导向导数(OD),通过捕捉像素与边界在方向和距离上的关系并生成不确定边界权重图来增强边界区域特征。MSNet、PraNet、VM-Unet及VM-UNetV2的实验结果均基于本文所设定的实验参数与训练流程计算得出，Polyp-PVT、Polyp-Mamba、SAM-Mamba和SSformer+OD模型的实验结果则直接引自相应文献中的分析数据,采用的训练和测试数据与本文一致。

如表2所示,MMF-MambaNet以95.1%的mDice和91.1%的mIoU位列第一,优于所有对比模型。相较于传统CNN模型,其mDice较MSNet、PraNet分别提升3.0%、5.2%;相较于Transformer类模型Polyp-PVT,mDice、mIoU分别提升1.4%、2.2%。在基于Mamba的模型中,MMF-MambaNet在两个指标上均取得优势,mDice较Polyp-Mamba、VM-UNetV2、SAM-Mamba分别提升1.0%、0.8%、0.9%;mIoU较Polyp-Mamba、VM-UNetV2、SAM-Mamba分别提升1.5%、1.8%、2.4%,且较2024年提出的Mamba模型VM-Unet提升更为显著。这表明MMF-MambaNet的多级多尺度特征融合与上下文感知机制,有效提升了息肉分割精度,尤其在边界模糊场景下表现更优。

表2 CVC_ClinicDB数据集上不同算法的结果

Table 2 Results of different algorithms on the CVC_ClinicDB dataset

/%

表3 Kvasir数据集上不同算法的结果

Table 3 Results of different algorithms on the Kvasir dataset

/%

表4 CVC-ColonDB数据集上不同算法的结果

Table 4 Results of different algorithms on the CVC-ColonDB dataset

/%

表5显示了MMF-MambaNet在ETIS-LaribPolypDB数据集上Dice和IoU虽略低于VM-UNetV2,但远超早期方法(如MSNet、PraNet),在mDice指标上较MSNet提升幅度超10%,较PraNet提

表3　显示了MMF-MambaNet在Kvasir数据集中的出色表现,两项指标优于所有比较模型。

Dataset	Networks	mDice	mIoU
CVC_ClinicDB	MSNet	92.1	87.9
	PraNet	89.9	84.9
	PolypPVT	93.7	88.9
	Polyp-Mamba	94.1	89.6
	VM-Unet	90.1	81.9
	VM-UNetV2	94.3	89.3
	SAM-Mamba	94.2	88.7
	SSformer+OD	92.0	87.0
	MMF-MambaNet	95.1	91.1

表4　显示了MMF-MambaNet在CVC-ColonDB数据集中的表现同PolypPVT、Polyp-Mamba和SSformer+OD较接近,较其余方法均获得了较大优势。PolypPVT所搭载的伪装识别模块,可针对性地抑制背景干扰,这一技术优势使其性能表现略优于MMF-MambaNet。

Dataset	Networks	mDice	mIoU
Kvasir	MSNet	90.7	86.2
	PraNet	89.8	84.0
	PolypPVT	91.7	86.4
	Polyp-Mamba	91.9	86.7
	VM-Unet	89.1	80.8
	VM-UNetV2	91.3	84.2
	SAM-Mamba	92.4	87.3
	SSformer+OD	91.6	86.4
	MMF-MambaNet	93.3	87.4

Dataset	Networks	mDice	mIoU
CVC-ColonDB	MSNet	75.5	67.8
	PraNet	70.9	64.0
	PolypPVT	80.8	72.7
	Polyp-Mamba	79.1	71.3
	VM-Unet	71.3	55.3
	VM-UnetV2	75.8	61.0
	SSformer+OD	79.2	71.4
	MMF-MambaNet	80.4	67.2

升接近20%,且较其他基于Mamba的模型均有较大提升,证明其在应对小目标区域和模糊边界时的竞争力。

表5 ETIS-LaribPolypDB数据集上不同算法的结果

Table 5 Results of different algorithms on the ETIS-LaribPolypDB Dataset

/%

Dataset	Networks	mDice	mIoU
ETIS-LaribPolypDB	MSNet	71.9	66.4
	PraNet	62.8	56.7
	PolypPVT	78.7	70.6
	Polyp-Mamba	75.6	66.8
	VM-Unet	79.8	66.4
	VM-UNetV2	83.9	72.3
	SSformer+OD	76.7	69.0
	MMF-MambaNet	82.4	70.0

表6显示了MMF-MambaNet在CVC-300数据集上的可靠表现,其Dice系数达89.7%,对息肉区域的重叠分割精度与目标定位准确性均处于较好水平,具备一定的性能优势。PolypPVT、Polyp-Mamba的表现稍优于MMF-MambaNet,基于对医学图像语义信息和细节特征融合的重视,VM-UnetV2表现出一定优势。

表6 CVC-300数据集上不同算法的结果

Table 6 Results of different algorithms on the CVC-300 dataset

/%

Dataset	Networks	mDice	mIoU
CVC-300	MSNet	86.9	80.7
	PraNet	87.1	79.7
	PolypPVT	90.1	83.3
	Polyp-Mamba	90.6	84.0
	VM-Unet	88.6	79.6
	VM-UnetV2	92.2	84.1
	SSformer+OD	88.3	81.4
	MMF-MambaNet	89.7	81.3

综合CVC-ClinicDB、Kvasir、CVC-ColonDB、ETIS-LaribPolypDB、CVC-300五大主流数据集的对比结果,尽管PolypPVT、Polyp-Mamba和VM-UnetV2在部分数据集中的表现稍优于MMF-MambaNet,但MMF-MambaNet总体上展现出了在综合性能上的优势。在ETIS-LaribPolypDB这一高挑战数据集(小息肉、复杂背景)中,MMF-MambaNet性能虽略低于VM-UNetV2,但远超MSNet、PraNet等模型,仍具强竞争力;在CVC-ColonDB数据集中,MMF-MambaNet相较VM-UNetV2取得了较大优势;在CVC-ClinicDB数据集上,其Dice系数和IoU均超越所有对比模型,包括Polyp-Mamba、SAM-Mamba等最新架构,实现性能最优;在Kvasir、CVC-ColonDB、CVC-300数据集上,其核心指标同样优于MSNet、PraNet、VM-Unet等多数主流方法。

图6呈现了MSNet、PraNet、VM-Unet、VM-UNetV2与本文MMF-MambaNet模型在预测阶段的分割效果对比。从结果可见,在面临息肉成像尺寸、色彩或形态差异大等挑战时,本文模型均能较好完成对息肉区域的识别与分割任务。即便遇到息肉颜色与周边组织颜色相近、边界模糊的情况,或是息肉与周围组织存在显著色彩反差的场景,该模型也可依托其学习到的特征信息,实现息肉与背景组织的有效区分,展现出比较稳定的分割性能。

3.5 消融实验

为评估MMF-MambaNet各核心模块的作用,在5个息肉分割数据集上开展消融实验(表7),分别移除LSA、ADEM、AGLI后,模型Dice与IoU均有不同程度下降:其中移除ADEM模块降幅最显著(Dice平均降3.8%、IoU平均降5.7%),其次是LSA(Dice平均降1.6%、IoU平均降2.6%)与AGLI(Dice平均降3.4%、IoU平均降4.9%)。这表明三个模块均对模型性能有正向贡献,ADEM在多尺度特征融合中作用关键,LSA助力局部细节聚焦,AGLI强化高低层特征协调,三者协同保障了MMF-MambaNet的分割性能。

3.5 轻量化与效率比较

除评估模型分割精度外,本文还对不同模型的计算复杂度进行了分析,以验证模型在工程实用性上的表现。实验采用模型参数量(Params)、计算量(FLOPs)及推理速度(FPS)作为核心评估指标,结果如表8所示(输入图像分辨率为256x256),所有测试均在NVIDIA GeForce RTX 4090上完成。其中,

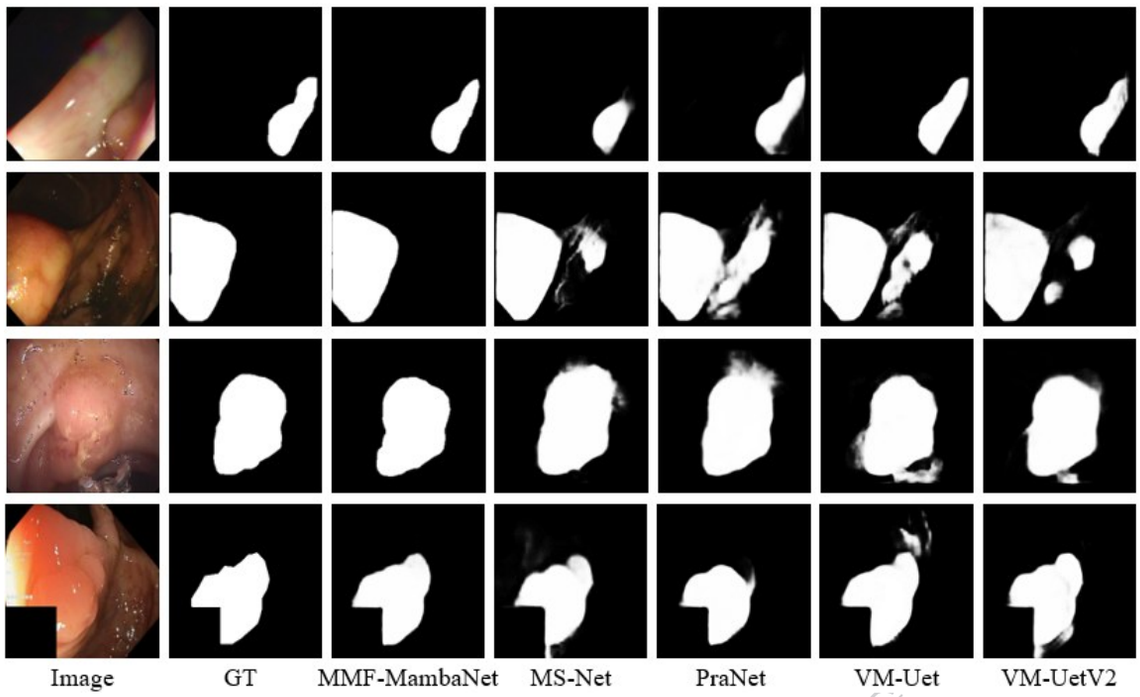

图6　分割结果可视化

MMF-MambaNet的综合计算性能表现最优：其参数量（18.84M）与计算量（5.74G）保持轻量化水平，仅略高于VM-UNetV2，而推理速度（57.58FPS）显著优于MSNet、PraNet、VM-Unet及VM-UNetV2等对比模型，充分体现出在保证模型轻量化的同时，MMF-MambaNet具备更高效的推理能力。

总体上，通过LSA、CASA、ADEM及AGLI等模块的协同设计，MMF-MambaNet实现了多级多尺度特征的结构化融合与增强，展现出优于主流模型的分割精度，尤其在边界细节捕捉和小息肉检出场景中表现较突出；模型保持了轻量化特性，推理速度明显优于VM-Unet、PraNet等模型，具备良好的临床部署潜力。面对息肉形态复杂、边界与周围组织高度相似等极具挑的战场景，需进一步优化模型适配能力。

4　结　论

本文构建了一种基于Mamba架构的结直肠息肉分割模型MMF-MambaNet，通过优化特征提取与跨尺度融合机制，提升对息肉多样形态及模糊边界的分割精度。该模型以视觉状态空间模块为基础，结合局部序列风格变换增强、跨尺度上下文自感知注意力及自适应细节增强等模块，实现多尺度特征的精准捕捉与动态融合，有效应对息肉图像中尺度多变、边界模糊等挑战。尽管MMF-MambaNet展现出较强的分割性能，但仍存在一定局限，模型在图像上的复杂息肉分割精度尚有提升空间，且多模块融合导致推理速度略低于轻量型网络。未来将通过精简特征交互路径提升推理效率，结合半监督学习等策略扩展对低质量、少标注数据的适应能力。

表7　消融实验结果表

Table 7　Results of ablation experiments

/%

Datasets	Networks	mDice	mIoU
CVC_ClinicDB	w/o LSA	91.1	83.6
	w/o ADEM	90.8	83.2
	w/o AGLI	91.3	83.9
	MMF-MambaNet	95.1	90.6
Kvasir	w/o LSA	92.0	85.2
	w/o ADEM	91.3	84.0
	w/o AGLI	91.8	84.9
	MMF-MambaNet	93.3	87.4
CVC-ColonDB	w/o LSA	79.1	65.4
	w/o ADEM	74.9	59.8
	w/o AGLI	77.0	62.7
	MMF-MambaNet	80.4	67.2
ETIS-LaribPolypDB	w/o LSA	81.9	69.3
	w/o ADEM	77.1	62.7
	w/o AGLI	75.4	60.5
	MMF-MambaNet	82.4	70.0
CVC-300	w/o LSA	89.0	80.2
	w/o ADEM	87.7	78.1
	w/o AGLI	88.1	79.9
	MMF-MambaNet	89.7	81.3

表8　计算复杂度、GPU内存占用与推理时间对比

Table 8　Comparison of computational complexity, GPU memory usage, and inference time

Model	Params(M)	FLOPs(G)	帧/S ↑
MSNet	27.69	16.93	30.43
PraNet	30.52	13.08	25.78
VM-Unet	35.62	7.56	20.61
VM-UNetV2	17.91	4.40	32.58
MMF-MambaNet	18.84	5.74	57.58

参考文献

Bernal J, Sánche F J, Fernández-Esparrach G, Gil D, Rodríguez C and Vilariño F. 2015. WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. Computerized Medical Imaging and Graphics, 43: 99-111. [DOI: 10.1016/j.compmedimag.2015.02.007]

Cheng Z, Guo J, Zhang J, Qi L, Zhou L, Shi Y and Gao Y. 2025. Mamba-Sea: A Mamba-based Framework with Global-to-Local Sequence Augmentation for Generalizable Medical Image Segmentation. IEEE Transactions on Medical Imaging, 44(9): 3741-3755. [DOI:10.1109/TMI.2025.3564765]

De Cristofaro E, Lolli E, Migliozzi S, Sincovih S, Marafini I, Zorzi F, et al. 2023. Frequency and Predictors of Dysplasia in Pseudopolyp-like Colorectal Lesions in Patients with Long-Standing Inflammatory Bowel Disease. Cancers, 15(13): 3361. [DOI: 10.3390/cancers15133361]

Dong B, Wang W, Fan D P, Li J, Fu H and Shao L. 2021. Polyp-pvt: Polyp segmentation with pyramid vision transformers [EB/OL]. https://arxiv.org/pdf/2108.06932

Dutta T K, Majhi S, Nayak D R and Jha D. 2025.SAM-Mamba: Mamba Guided SAM Architecture for Generalized Zero-Shot Polyp Segmentation. IEEE/CVF Winter Conference on Applications of Computer

Vision (WACV). Waikoloa, USA: IEEE: 4655-4664. [DOI: 10.1109/WACV58584.2025.00472]

Fang Y, Chen C, Yuan Y, Tong K Y, et al. 2019. Selective feature aggregation network with area-boundary constraints for polyp segmentation//Proceedings of the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention-MICCAI 2019. Shenzhen, China: Springer: 302-310. [DOI: 10.1007/978-3-030-32245-8_33]

Fan D P, Ji G P, Zhou T, et al. 2020. PraNet: Parallel reverse attention network for polyp segmentation//Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention-MICCAI 2020. Cham: Springer: 263-273. [DOI: 10.1007/978-3-030-59719-1_26] YangF, ChenL and WangZ J. 2019. MicroRNA-32 inhibits the proliferation, migration and invasion of human colon cancer cell lines by targeting E2F transcription factor 5. European review for medical and pharmacological sciences, 23 (10. [DOI:10.26355/eurrev_201910_19094]

Gong H, Kang L, Wang Y, Wang Y, Wan X, Wu X and Li H. 2025. nnMamba: 3D biomedical image segmentation, classification and landmark detection with state space model//Proceedings of the 22nd International Symposium on Biomedical Imaging (ISBI). IEEE: 302-310. [DOI:10.1109/ISBI56570.2025.00008]

Gross S, Kennel M, Stehle T, Wulff J, Tischendorf J, Trautwein C and Aach T. 2009. Polyp segmentation in NBI colonoscopy//Proceedings of Bildverarbeitung für die Medizin: Algorithmen—Systeme—Anwendungen. Heidelberg, Germany: Springer: 252-256. [DOI:10.1007/978-3-540-92990-1_61]

Gu A and Dao T. 2023. Mamba: linear-time sequence modeling with selective state spaces [EB/OL]. https://arxiv.org/pdf/2312.00752.pdf

Hatamizadeh A and Kautz J. 2024. MambaVision: a hybrid Mamba-Transformer vision backbone [EB/OL]. https://arxiv.org/pdf/2407.08083.pdf

Hwang S and Celebi M E. 2010. Polyp detection in wireless capsule endoscopy videos based on image segmentation and geometric features//Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Dallas, USA: IEEE: 678-681. [DOI:10.1109/ICASSP.2010.5495581]

Jha D, Riegler M A, Johansen D, Halvorsen P and Johansen H D. 2020. DoubleU-Net: a deep convolutional neural network for medical image segmentation//Proceedings of the 33rd IEEE International Symposium on Computer-Based Medical Systems (CBMS). Rochester, USA: IEEE: 558-564. [DOI: 10.1109/CBMS49503.2020.00111]

Jiao Y, Xu Q, Luo Y, He X, Chen Z and Duan W. 2025. TM-UNet: Token-Memory Enhanced Sequential Modeling for Efficient Medical Image Segmentation [EB/OL]. https://arxiv.org/pdf/2511.12270.pdf

Jerebko A K, Teerlink S, Franaszek M and Summers R M. 2003. Polyp segmentation method for CT colonography computer-aided detection//Proceedings of Medical imaging: Physiology and Function: Methods, Systems, and Applications. Bellingham, USA: SPIE: 359-369. [DOI:10.1117/12.463584]

Li Y, Cheng M, Zheng X, Ji R and Chen J. 2025. Oriented-derivative Representation for Boundary-aware Polyp Segmentation. IEEE Transactions on Multimedia, 27: 7608-7618. [DOI:10.1109/TMM.2025.3445678]

Ling T, Wu C, Yu H, Cai T, Wang D, Zhou Y, et al. 2023. Probabilistic Modeling Ensemble Vision Transformer Improves Complex Polyp Segmentation//Proceedings of the 26th International Conference on Medical Image Computing and Computer-Assisted Intervention -MICCAI. Cham: Springer: 572-581. [DOI: 10.1007/978-3-031-43994-5_56]

Liu J M, Cao S H and Zhang Z P. 2026. Medical image segmentation with vision Mamba and adaptive multiscale loss fusion [J]. Journal of Image and Graphics, 31 (1): 0335-0348. (刘建明, 曹圣浩, 张志鹏. 2026. 融合视觉 Mamba 与自适应多尺度损失的医学图像分割. 中国图象图形学报, 31(1): 0335-0348) [DOI:10.11834/jig.250224]

Ma J, Li F and Wang B. 2024. U-mamba: Enhancing long-range dependency for biomedical image segmentation [EB/OL]. https://arxiv.org/pdf/2401.04722.pdf

Oktay O, Schlemper J, Folgoc L L, Lee M, Heinrich M, Misawa K, et al. 2018. Attention u-net: Learning where to look for the pancreas [EB/OL]. https://arxiv.org/pdf/1804.03999.pdf

Ronneberger O, Fischer P and Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and computer-assisted intervention-MICCAI. Munich, Germany: Springer: 234-241. [DOI:10.1007/978-3-319-24574-4_28].

Ruan J and Xiang S. 2024. VM-UNet: vision mamba UNet for medical image segmentation [EB/OL]. https://arxiv.org/pdf/2402.02491.pdf

Tajbakhsh N, Gurudu S R and Liang J M. 2016. Automated polyp detection in colonoscopy videos using shape and context information. IEEE Transactions on Medical Imaging, 35(2): 630-644. [DOI: 10.1109/TMI.2015.2487997]

Wang G, Li Y, Chen W, Ding M, Cheah, W P, Qu R, et al. 2025. S^3-Mamba: Small-Size-Sensitive Mamba for Lesion Segmentation//Proceedings of the 39th AAAI Conference on Artificial Intelligence. Washington, DC, USA: AAAI: 7655-7664. [DOI: 10.1609/aaai.v39i7.26789]

Wang Z, Cheng J Q, Zhang Y, Cui G and Li L. 2024. Mamba-UNet: UNet-like pure visual Mamba for medical image segmentation [EB/OL]. https://arxiv.org/pdf/2402.05079.pdf

Wang Z and Ma C. 2024. Weak-Mamba-UNet: Visual Mamba makes

CNN and ViT work better for scribble-based medical image segmentation [EB/OL]. https://arxiv.org/pdf/2402.10887.pdf

Wu Q Q, Deng X, Shao H J and Wang F. 2026, ViBound-Net: a hybrid attention network for colorectal polyp segmentation integrating visual and boundary-focused mechanisms. Journal of Image and Graphics, 31(2): 0001-0018. (吴琪琪,邓星,邵海见,王飞. 2026. 视觉与边界融合的混合注意力结直肠息肉分割. 中国图象图形学报,31(2): 0001-0018).[DOI:10.11834/jig.250488]

Wu R, Liu Y, Liang P and Chang Q. 2024. Ultralight vm-unet: Parallel vision mamba significantly reduces parameters for skin lesion segmentation [EB/OL]. https://arxiv.org/pdf/2403.20035.pdf

Wu Y, Shen X, Bu F and Tian J. 2020. Ultrasound image segmentation method for thyroid nodules using ASPP fusion features. IEEE Access, 8: 172457-172466. [DOI: 10.1109/ACCESS. 2020. 3023558]

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 (ECCV): 3-19. [DOI: 10.1007/978-3-030-01234-2_1]

Xie J, Liao R, Zhang Z, Yi S, Zhu Y and Luo G. 2024. ProMamba: Prompt-Mamba for polyp segmentation [EB/OL]. https://arxiv.org/pdf/2403.13660.pdf

Xu Z, Tang F, Chen Z, Zhou Z, Wu W, Yang Y, et al. 2024. PolypMamba: Polyp Segmentation with Visual Mamba//Proceedings of the 27th International Conference on Medical Image Computing and Computer-Assisted Intervention -MICCAI. Marrakesh, Morocco: Springer: 510-521. [DOI:10.1007/978-3-031-52437-7_49]

Yao J, Miller M, Franaszek M and Summers R M. 2004. Colonic polyp segmentation in CT colonography-based on fuzzy clustering and deformable models. IEEE Transactions on Medical Imaging, 23 (11): 1344-1352. [DOI:10.1109/TMI.2004.836385]

Yu T, Lin N, Zhang X, Pan Y, Hu H, Zheng W, et al. 2022. An end-to-end tracking method for polyp detectors in colonoscopy videos. Artificial Intelligence in Medicine, 131: 102363. [DOI:10.1016/j.artmed.2022.102363]

Zhang M, Yu Y, Jin S, Gu L, Ling T and Tao X. 2024. VM-UNET-V2: rethinking vision Mamba UNet for medical image segmentation// Proceedings of International Symposium on Bioinformatics Research and Applications. Singapore: Springer: 335-346. [DOI: 10.1007/978-981-99-6505-5_29]

Zhang Y, Liu H 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 - MICCAI Strasbourg, France: Springer: 14-24. [DOI:10.1007/978-3-030-87196-3_2]

Zhao H, Shi J, Qi X, Wang X and Jia J. 2017. Pyramid scene parsing network//Proceedings of the IEEE Conference on computer vision and pattern recognition (CVPR). Honolulu, USA: IEEE: 2881-2890. [DOI:10.1109/CVPR.2017.309]