

中图法分类号: TP18; TP391 文献标识码: A 文章编号: 1006-8961(2026)01-0335-14

论文引用格式: Liu J M, Cao S H and Zhang Z P. 2026. Medical image segmentation with vision Mamba and adaptive multiscale loss fusion. 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]

融合视觉 Mamba 与自适应多尺度损失的医学图像分割

刘建明^{1,2}, 曹圣浩^{1*}, 张志鹏²

1. 江西师范大学计算机信息工程学院, 南昌 330022; 2. 江西师范大学数字产业学院, 上饶 334000

摘要: 目的 在医学图像分割领域, 传统基于卷积神经网络(convolutional neural network, CNN)的模型在捕捉长距离依赖信息方面存在固有局限, 而基于视觉 Transformer(vision Transformer, ViT)的模型其自注意力机制的计算复杂度与图像尺寸呈平方关系, 在资源有限的现实环境中难以部署。为了解决这些问题, 提出一种融合视觉 Mamba 和自适应多尺度损失的医学图像分割方法 VMAML-UNet(medical image segmentation with vision Mamba and adaptive multi-scale loss)。方法 VMAML-UNet 采用编码器—解码器架构。在编码阶段, 设计了融合小波卷积的视觉 Mamba 块, 以线性复杂度提取病变区域的精确特征并扩大感受野, 并通过块合并进行下采样。解码阶段同样引入融合小波卷积的视觉 Mamba 块并利用块扩展进行上采样。跳跃连接中, 提出小波卷积注意力聚合模块, 用于提取并融合不同尺度下的图像特征。此外, 设计了柯尔莫哥洛夫—阿诺德网络(Kolmogorov-Arnold network, KAN)调控多尺度加权损失, 动态调控各层级损失权重。**结果** 在 BUSI(breast ultrasound images dataset)、GlaS(gland segmentation in histology images challenge dataset)和 CVC(CVC-ClinicDB dataset)3 个异质性显著的医学图像数据集上的实验结果表明, 与主流的 VM-UNet(vision Mamba UNet)等采用 Mamba 的医学图像分割方法相比取得显著的性能提升。在 BUSI 数据集上, 交并比(intersection over union, IoU)和 F1 分数分别提升 2.72% 和 2.02%; 在 GlaS 数据集上, IoU 和 F1 分数分别提升 3.38% 和 1.89%; 在 CVC 数据集上, IoU 和 F1 分数分别提升 2.51% 和 1.42%。**结论** 提出的 VMAML-UNet 采用基于视觉 Mamba 的线性复杂度的长距离依赖建模与基于 KAN 的动态损失优化机制, 显著减少了计算成本, 同时提升了模型对复杂医学图像的分割精度。该模型在 3 个数据集上的优异表现证明了其在不同医学图像场景下的广泛适用性和高效性。

关键词: 状态空间模型(SSM); 柯尔莫哥洛夫—阿诺德网络(KAN); 小波卷积; 多尺度加权损失; 连续流

Medical image segmentation with vision Mamba and adaptive multiscale loss fusion

Liu Jianming^{1,2}, Cao Shenghao^{1*}, Zhang Zhipeng²

1. School of Computer and Information Engineering, Jiangxi Normal University, Nanchang 330022, China;

2. School of Digital Industry, Jiangxi Normal University, Shangrao 334000, China

Abstract: Objective Medical image segmentation is crucial for identifying anatomical structures and regions of interest in medical images, playing a critical role in diagnosis and treatment planning. Although traditional convolutional neural net-

收稿日期: 2025-05-22; 修回日期: 2025-06-23; 预印本日期: 2025-06-30

* 通信作者: 曹圣浩 202341600093@jxnu.edu.cn

基金项目: 国家自然科学基金项目(62266022); 江西省自然科学基金项目(20242BAB25110)

Supported by: National Natural Science Foundation of China(62266022); Natural Science Foundation of Jiangxi Province, China (20242BAB25110)

work (CNN)-based models have shown notable success, they often struggle to capture long-range dependencies, resulting in suboptimal feature extraction and segmentation performance. This limitation is particularly problematic in medical imaging, where accurate and detailed segmentation is necessary for reliable diagnoses. Transformer-based models that use the self-attention mechanism excel in global context modeling but demonstrate quadratic computational complexity with image size, increasing their computational cost for dense medical image segmentation tasks and hindering efficient real-world applications. Recent studies indicate that state-space models such as Mamba can simulate long-range dependencies with linear complexity. Furthermore, Kolmogorov-Arnold networks (KAN) possess powerful nonlinear modeling capabilities suitable for complex medical image features. However, traditional static weighting strategies ineffectively adapt to the dynamic nature of medical image data. Aiming to address these challenges, VMAML-UNet, a novel medical image segmentation framework combining KAN, is proposed to regulate multiscale weighted losses and visual Mamba for efficient long-range dependency modeling. **Method** The VMAML-UNet method adopts an encoder-decoder architecture, a widely used and effective design in deep learning for image segmentation tasks. In the encoding stage, a novel visual Mamba block (WCVM block) is introduced, incorporating wavelet convolutions to extract precise and localized features from lesion regions with linear computational complexity. The use of wavelet convolutions enables the model to expand its receptive field, which is critical for capturing long-range dependencies within the image. The visual Mamba block enhances feature extraction by improving the representation of critical areas within the image, thereby addressing the issue of insufficient feature capture. Furthermore, the encoding stage incorporates downsampling through block merging, which effectively reduces data dimensionality while retaining important features. In the decoding phase, WCVM blocks are reused, and block expansion is employed to perform upsampling. This approach aids in accurately reconstructing the segmentation mask with high accuracy, ensuring that fine details are preserved throughout the process. The skip connections between the encoder and decoder are designed to transfer critical information from low to high layers of the network. This study introduces a new component: the wavelet convolution attention aggregation (WCAA) module. The WCAA module is designed to fuse and refine features from multiple scales, both spatially and across channels, which allows the model to capture more complex, multidimensional patterns within the image. This module is particularly useful for improving the quality of segmentation in images where the regions of interest are surrounded by similar tissue, making them harder to differentiate. Additionally, a KAN-regulated multiscale weighted loss module is introduced to dynamically capture the nonlinear features and inter-layer dependencies among outputs from different stages of the model. This module addresses the limitations of traditional static weighting strategies, which fail to adapt to the dynamic nature of feature representations extracted at different layers. Specifically, the KAN module applies KAN convolutions to the final three decoder layers to generate multiscale segmentation masks, which are then used to compute hierarchical losses. These losses are then combined with the corresponding encoder outputs to form the multiscale weighted loss. Finally, this loss is integrated with the loss computed from the true labels and predicted masks, enabling effective backpropagation and model training. **Result** Aiming to evaluate the performance of the proposed VMAML-UNet model, experiments on three diverse and heterogeneous medical image datasets were conducted: the BUSI dataset, the GlaS dataset, and the CVC dataset. These datasets were selected because they represent different types of medical images with varying complexity and noise levels. Experimental results show that the VMAML-UNet outperforms other segmentation methods, such as VM-UNet, which also employs VSS blocks for segmentation. Specifically, on the BUSI dataset, the VMAML-UNet model achieved an improvement in intersection over union (IoU) and an improvement in F1 score by 2.72% and 2.02%, respectively, compared to VM-UNet. The BUSI dataset, which contains breast ultrasound images, presents challenges due to the noise and variability in image quality. However, the proposed model showed notable improvements in addressing these issues. On the GlaS dataset, which contains eye fundus images for glaucoma detection, the VMAML-UNet model achieved 3.38% and 1.89% improvements in IoU and F1 score, respectively. Glaucoma is a leading cause of blindness, and accurate segmentation of the optic nerve head is crucial for effective diagnosis. The strong performance of the VMAML-UNet model on this dataset highlights its capability to capture fine details in medical images. Similarly, on the CVC dataset, which comprises colonoscopy images, the model demonstrated improvements of 2.51% in IoU and 1.42% in F1 score. These results further confirm that the proposed VMAML-UNet model substantially improves segmentation performance across different types of medical images. **Conclusion**

Through the integration of wavelet convolution-enhanced visual state-space (VSS) blocks, the proposed VMAML-UNet notably reduces computational costs and effectively addresses the limitations of CNN and Transformer-based models in medical image segmentation. The superior performance of this model across three datasets highlights its broad applicability and efficiency in various medical imaging scenarios, offering valuable insights into the development of highly efficient and robust medical image segmentation methods.

Key words: state space model (SSM); Kolmogorov-Arnold network (KAN); wavelet convolution; multi-scale weighted loss; continuous flow

0 引言

医学图像分割(Ronneberger等,2015;Myronenko,2019)是一项富有挑战性的任务,它通过辅助医生快速诊断病理,显著提高患者护理效率。近年来,深度学习技术在此领域得到广泛应用并取得突破性进展。其中,基于卷积神经网络(convolutional neural network,CNN)和Transformer的模型表现突出,在医学图像分割任务中成果显著。以U-Net(Ronneberger等,2015)为代表的CNN模型因其简洁结构和强扩展性而备受关注。例如,U-Net++(Zhou等,2018)设计了嵌套U型网络,通过深度融合多尺度特征提升分割精度;Att-Unet(Oktay等,2018)在跳跃连接处引入注意力门(attention gate),利用动态权重分配聚焦目标区域并抑制背景信息;U-NeXt(Valanarasu和Patel,2022)与Rolling U-Net(Liu等,2024b)则采用卷积和多层感知机(multi layer perceptron,MLP)结合的U型架构,优化了网络结构与性能。许多后续改进(Çiçek等,2016;Ruan等,2022a,b,2023;Li等,2024b;Zhou等,2024)均基于此经典U形架构。在Transformer模型方面,TransUNet(Chen等,2021)作为开创性工作之一,首次在编码器中引入ViT(vision Transformer)(Dosovitskiy等,2021)提取特征,并在解码器结合CNN,展现了强大的全局信息建模能力。随后,TransFuse(Zhang等,2021)提出ViT与CNN的并行架构,以同时捕捉局部细节和全局上下文。此外,Swin-UNet(Cao等,2023)结合Swin Transformer(Liu等,2021)与U形结构,首次实现了纯Transformer的U形模型。

然而,无论是CNN模型还是Transformer模型都存在固有局限性(Liu等,2023)。CNN模型受限于局部感受野,难以有效捕捉长距离依赖,常导致特征提取不充分,影响分割效果。Transformer模型虽然擅

长全局建模,但其自注意力机制的计算复杂度与图像尺寸呈平方关系(Dosovitskiy等,2021;Vaswani等,2017;Mu等,2024),带来巨大计算负担,尤其在医学图像分割这类密集预测任务中。这些模型的不足促使研究者探索新架构,以期在有效捕捉长距离信息的同时,保持线性计算复杂度。

最近,以Mamba(Gu和Dao,2024)为代表的状态空间模型(state space model,SSM)(Kalman,1960)已成为传统CNN和ViT的强大竞争对手。Mamba不仅能够建立长距离依赖关系,还展现出与输入尺寸呈线性关系的计算复杂度。基于SSM的模型在多个领域得到了深入研究,包括语言理解(Gu和Dao,2024)、通用视觉任务(Zhu等,2024;Liu等,2025a)等。特别是在医学图像分割领域,U-Mamba(Ma等,2024)是近期提出的一种新型的SSM-CNN混合模型,首次将其应用于医学图像分割任务。VM-UNet(vision Mamba U-Net)(Ruan等,2024)是首个纯SSM驱动的医学图像分割模型,其核心架构是通过状态空间模型高效捕获全局上下文信息,为后续研究提供了基础框架。MDSVM-UNet++(multi-directional snake vision Mamba UNet++)(刘建明和唐煜城,2024)则结合多方向蛇形卷积和视觉残差Mamba实现了冠状动脉的分割。虽然上述工作已将SSM应用于医学图像分割任务,但基于SSM的模型在该领域的性能仍有待探索,且上述U-Net变体都存在可解释性不足的问题。柯尔莫哥洛夫—阿诺德网络(Kolmogorov-Arnold network,KAN)(Bodner等,2025)凭借其出色的可解释性,尝试揭开传统网络结构的“黑箱”,展现了“白箱”网络研究的巨大潜力。近期研究(Liu等,2025b)表明,KAN有望成为多层感知机的一种有力替代方案。

为了有效利用Mamba的长距离依赖建模能力和KAN强大的非线性建模能力及其可解释性,本文提出一种融合视觉Mamba与自适应多尺度损失的

医学图像分割模型(medical image segmentation with vision Mamba and adaptive multi-scale loss, VMAML-UNet)。具体而言,VMAML-UNet由以下部分组成:编码器、解码器、小波卷积视觉 Mamba(wavelet convolution vision Mamba, WCVM)模块、小波卷积注意力聚合(wavelet convolution attention aggregation, WCAA)模块和 KAN 调控的多尺度损失(KAN-regulated multi-scale weighted loss, KMWL)模块。其中编码器和解码器都包含 WCVM 模块,用于精准提取病变区域的特征、增大感受野并以线性复杂度建模长距离依赖关系。在跳跃连接部分,采用 WCAA 模块可以有效地提取并融合不同尺度中的图像特征信息,且生成的特征图在空间和通道维度上都得到了丰富和细化。此外,基于 KAN 强大的非线性表征能力, KMWL (KAN-regulated multi-scale weighted loss)损失能精准捕捉各层级输出的非线性特征和层级间的特征依赖关系,并对各层级损失权重进行动态调控。

本文的主要贡献总结如下:1)提出一种新的医学图像分割模型 VMAML-UNet:通过线性复杂度的长距离建模与基于 KAN 动态损失优化机制,显著提升模型对复杂医学图像的分割精度;2)提出一种 KAN 调控的多尺度加权损失 KMWL:在 KAN 强大的非线性表征能力支持下,能精准捕获层级间的特征依赖关系,并对各层级损失权重进行动态调控;3)设计了一种基于小波卷积的注意力聚合模块 WCAA:在跳跃连接部分,采用该模块可以有效整合并提取不同尺度中的图像特征信息,且生成的特征图在空间和通道维度上都得到丰富和细化;4)在多个医学图像分割任务数据集中与当前最前沿的(state-of-the-art, SOTA)方法对比有显著性能提升。

1 相关工作

1.1 状态空间模型

基于状态空间的模型如 HiPPO (high-proorder polynomial projection operators)(Gu 等, 2020)和 LSSL (linear state space layer)(Gu 等, 2021)等最初源自控制理论中的线性状态空间方程,主要用于序列数据建模任务。结构化状态空间序列(structured state space sequence model, S4)模型(Gu 等, 2022),仅需序列长度线性复杂度的计算即可实现长距离依赖建

模,显著提升了效率。近期, Mamba 作为一种新型架构,在自然语言处理任务中展现出与 CNN 及 Transformer 相媲美的性能,其通过引入数据依赖的 SSM 层扩展了 S4 框架,推动了 SSM 模型的进一步发展。受此启发, Vision Mamba (Zhu 等, 2024)和 VMamba (Liu 等, 2024b)引入了交叉扫描模块和良好的分层架构设计,在通用视觉领域实现了基于数据驱动 SSM 层的模型架构,性能显著提升。在医学图像分割领域, U-Mamba (Ma 等, 2024)在 SSM 中嵌入卷积运算,以将卷积层的局部建模能力与 SSM 的全局建模能力相结合。 SegMamba (Xing 等, 2024)作为 SSM 与 CNN 的混合模型,初步展现了 SSM 在医学图像分割中的潜力。然而,纯 SSM 模型在分割任务中的能力尚未得到充分挖掘。 VM-UNet (Ruan 等, 2024)是首个纯 SSM 驱动的医学图像分割模型,其核心架构是通过状态空间模型高效捕获全局上下文信息,为后续研究提供了基础框架。

1.2 柯尔莫哥洛夫—阿诺德网络

柯尔莫哥洛夫—阿诺德(Kolmogorov-Arnold)定理表明,任何连续函数均可表示为有限个一元连续函数的复合,这一理论奠定了构建通用神经网络模型的数学基础。后续研究进一步证明了这一点,即前馈神经网络(feed-forward neural network, FNN)具备通用逼近能力(Hornik 等, 1989),为深度学习的发展提供了理论支持。基于柯尔莫哥洛夫—阿诺德定理,研究者提出 KAN 这种新型神经网络架构。KAN 由一系列级联的柯尔莫哥洛夫—阿诺德层组成,每个层均包含一组可学习的一维激活函数。该网络结构在高维复杂函数的逼近方面展现出卓越的性能,并已在多个应用领域得到验证。KAN 具有较强的理论可解释性和模型可解释性。研究者对 KAN 的优化特性及其收敛性进行了深入分析,证明了其优异的逼近能力及泛化性能。后续研究进一步提出一种深度 KAN 模型,并将其应用于图像分类等任务。此外,研究者还将 KAN 拓展至时间序列预测和控制问题领域。最近研究指出,由于 KAN 的高效性和可解释性, KAN 可以作为多层感知机的有力替代方案(Liu 等, 2025b)。鉴于此,研究者提出一种基于 U-KAN (Li 等, 2024a)的医学图像分割方法,通过可学习激活函数捕获图像的非线性特征。后续又有研究提出将 KAN 和 Mamba 相结合的方法用于医学图像分割,如 KAN-Mamba FusionNet (Agrawal 等, 2025)、

KM-UNet (KAN Mamba UNet) (Zhang, 2025)。尽管 KAN 取得了重要进展, 但与 Mamba 相比, 基于 KAN 架构的方法存在一个缺点, 即它无法对图像中复杂的空间特征进行自注意力机制的学习 (Woo 等, 2018)。目前, KAN 作为一种具有坚实理论基础的新型神经网络模型, 尚未广泛整合至主流计算机视觉框架之中。因此, 本文尝试在该方向开展探索性研究, 旨在设计一种结合 KAN 的新型视觉 Mamba 网络架构, 并在图像分割领域对其有效性进行验证。

1.3 可变卷积

1.3.1 小波卷积

小波变换卷积 (wavelet transform convolution, WTConv) (Finder 等, 2025) 是一种将小波变换与传统卷积操作相结合的创新方法并要求输入与输出通道保持相同, 简称小波卷积, 旨在利用小波变换的多分辨率分析能力增强特征提取效果。小波变换通过不同尺度和位置的基函数对信号进行分解, 能够同时捕捉时域和频域信息 (Wang 等, 2021), 与傅里叶变换不同, 其局部性特点使其在处理非平稳信号 (如图像中的边缘、纹理等) 具有独特的优势。在图像处理领域, 小波卷积广泛应用于图像去噪、增强等任务, 能够有效地去除噪声, 同时保留图像的重要细节和边缘信息。在图像分割任务中, 小波卷积被引入到网络模型中, 替代传统深度可分离卷积 (Finder 等, 2025), 以提升模型对病变区域等关键特征的提取能力, 增大感受野, 从而更全面地捕捉病变区域的特征信息, 提升分割精度。

1.3.2 KAN 卷积

KAN 卷积 (Bodner 等, 2025) 的提出受柯尔莫哥洛夫—阿诺德定理启发以应对传统 CNN 的改进需求。传统 CNN 在计算机视觉任务中表现出色, 但存在一些局限性, 如参数量较大、训练速度较慢等。KAN 卷积通过引入可学习的非线性激活函数, 试图在保持 CNN 高效空间处理能力的同时, 克服这些缺点。传统卷积核由固定的权重参数构成, 其数值可学习但形式是线性的, 而 KAN 卷积通过将卷积核的每个元素替换为可学习的非线性激活函数, 不仅显著增强了模型的非线性表达能力, 还提升了模型对复杂数据模式的适应性。KAN 卷积广泛用于改进经典分割网络如 U-Net (Ronneberger 等, 2015), 通过替换传统卷积层, 模型能同时捕捉到丰富的局部特征和非线性特征。KAN 卷积通过理论创新与跨领

域应用, 成为提升模型非线性能力的重要工具, 尤其在医学图像分割领域展现出广阔前景。

2 本文方法

本文提出的 VMAML-UNet 整体架构如图 1 所示, 给定输入图像 $X \in \mathbb{R}^{H \times W \times 3}$, 其中 H 是图像高度, W 是图像宽度, 3 表示图像通道数。

在编码器部分, 图像块嵌入层首先将输入图像分割为不重叠的 4×4 像素块, 随后将图像的通道维度映射为 C , C 默认为 96。经过映射后, 图像维度被压缩为 $X' \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times C}$, 并通过层归一化 (layer normalization, LN) (Ba 等, 2016) 处理后输入编码器。编码器由 4 个阶段构成, 前 3 个阶段末尾均采用块合并操作对特征进行下采样。各阶段分别部署 2 个 WCVM 模块, 通道数依次为 $C, 2C, 4C, 8C$ 。

在解码器部分, 同样包含 4 个阶段, 后 3 个阶段起始处采用块扩展操作对特征进行上采样。前 3 个阶段部署 2 个 WCVM 块, 第 4 个阶段部署 1 个 WCVM 块, 通道数依次为 $C, 2C, 4C, 8C$ 。解码器末端通过最终投影层进行特征重建, 首先进行 4 倍上采样恢复原始分辨率, 再通过线性投影调整通道数以匹配分割目标。在跳跃连接部分模型采用 WCAA 模块跳跃连接, 能够融合编码器阶段输出的特征, 增强解码器的性能, 该模块的输出与上采样后的解码器输出进行相加操作作为下一层解码器的输入。此外, 类似于深度监督 (Zhou 等, 2018), 后 3 个层级解码器输出的特征图分别通过 KAN 卷积进行掩码预测并计算多尺度分层损失, 随后与各层级解码器的输出一同作为 KMWL 模块的输入进行动态调控损失权重并计算多尺度加权损失。

2.1 融合小波卷积的视觉 Mamba 块

融合小波卷积的视觉 Mamba 块如图 2 所示。给定输入特征图 $W_{in} \in \mathbb{R}^{\tilde{H} \times \tilde{W} \times \tilde{C}}$, 在经过层归一化后, 输出特征图分成两个分支。在第 1 个分支中, 首先使用线性层将输入特征通道数扩展到 $\lambda \times \tilde{C}$, λ 表示预定义的通道调整因子; 随后经过 SiLU (sigmoid linear unit) 激活函数 (Elfwing 等, 2018) 处理后输出特征为 W_1 。在第 2 个分支中, 使用线性层将输入特征通道数扩展到 $\lambda \times \tilde{C}$, 并经卷积核大小 (kernel size) 为 3×3 的小波卷积和激活函数处理后送入 2D 选择性扫

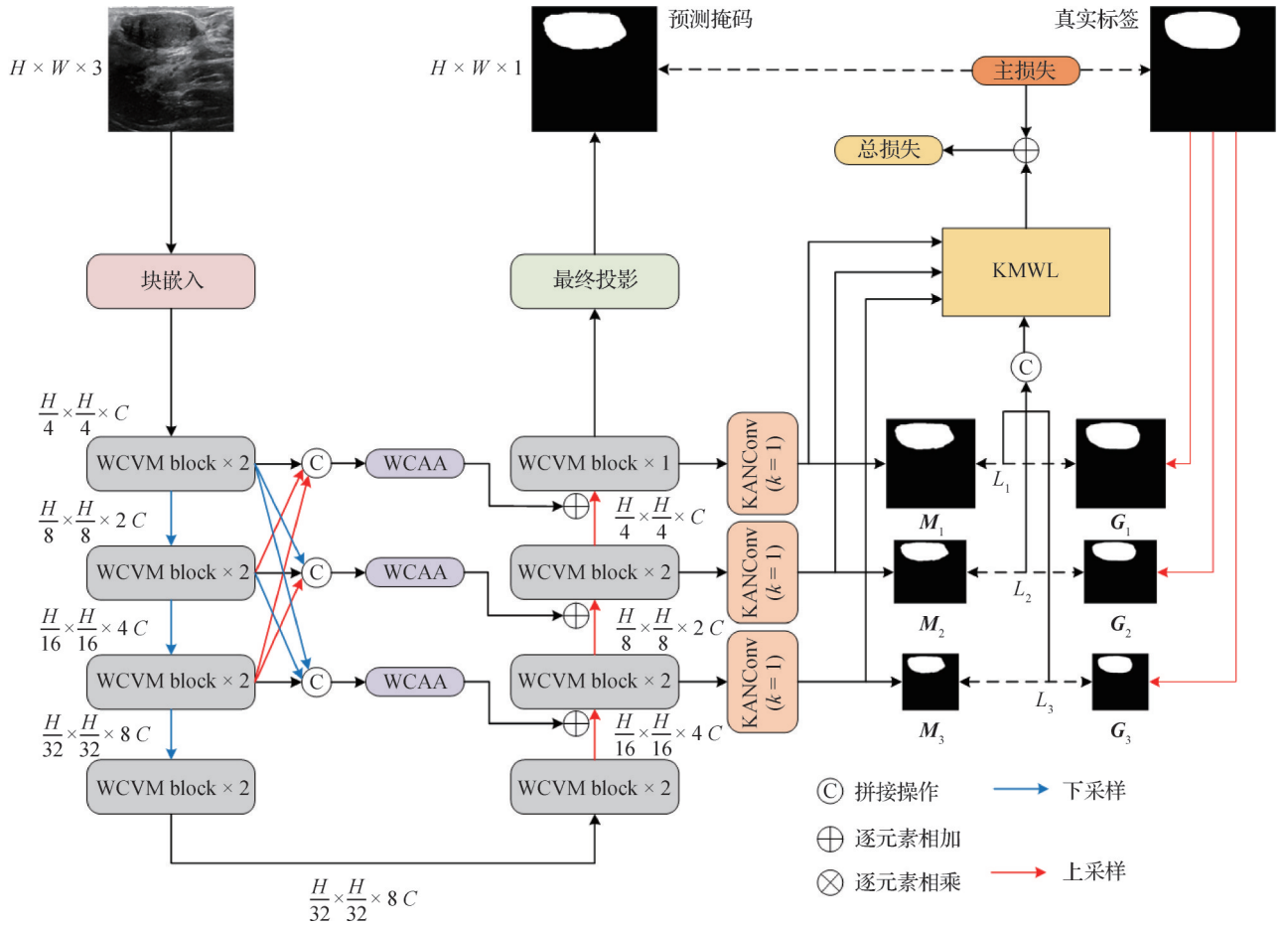


图1 网络框架

Fig. 1 Network framework

描(2D selective scan, SS2D)模块(Liu等, 2025a)以进一步提取特征;随后特征通过层归一化操作进行标准化后输出特征为 W_2 ,并与第1个分支的输出进行Hadamard乘积以聚合来自两个分支的特征;最后,通过一个线性层将聚合特征的通道数投影回 $\tilde{C}$,并将输出结果与输入特征图进行逐元素相加执行特征融合,形成WCVM模块的输出 W_{out} 。上述过程数学表达式为

$$W_1 = \text{SiLU}(\text{Linear}(\text{LN}(W_{in}))) \quad (1)$$

$$W_2 = \text{LN}\left(\text{SS2D}\left(\text{SiLU}\left(\text{WTConv}_{k=3}\left(\text{Linear}(\text{LN}(W_{in}))\right)\right)\right)\right) \quad (2)$$

$$W_{out} = \text{Linear}(W_1 \odot W_2) + W_{in} \quad (3)$$

式中, LN 表示层归一化, Linear 表示线性层, SiLU 表示激活函数, $\text{WTConv}_{k=3}$ 表示 3×3 的小波卷积, SS2D 表示2D选择性扫描模块。 $\odot$ 表示Hadamard乘积。

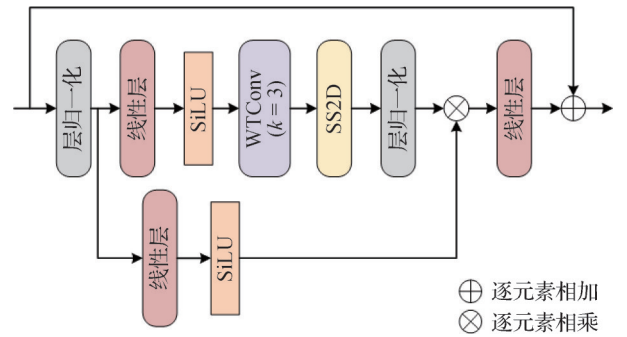


图2 融合小波卷积的视觉Mamba块

Fig. 2 Wavelet convolution vision Mamba block

SS2D模块由3部分组成:扫描扩展(scan expanding)操作、S6(linear-time sequence modeling with selective state spaces)模块和扫描合并(scan merging)操作。如图3所示,扫描扩展操作将输入图像沿4个不同方向,即从左下到右上、从右上到左下、从左上到右下、从右下到左上展开成序列。这些序列由S6模块进行特征提取,确保不同方向的信息

得到充分扫描,从而捕捉多样化的特征。接着,扫描合并操作将4个方向的序列求和并合并,将输出图像恢复到与输入图像相同的大小。其中,S6模块源自 Mamba,在 S4 的基础上引入了选择性机制,通过输入调整 SSM 的参数,使模型能够区分并保留相关信息,同时过滤掉不相关的信息。

2.2 小波卷积注意力聚合模块

图4描述了用于细化图像特征的WCAA模块。给定前3个编码器阶段的输出,通过不同尺度的上采样或下采样后的特征为 F_1^e 、 F_2^e 和 F_3^e ,并将其拼接为 $F_{in} \in \mathbf{R}^{H \times W \times C_1}$ 。组合后的特征 F_{in} 被输入到WCAA模块中进行细化。在WCAA模块中,主要分为多尺度融合(multi-scale fusion)、空间聚合(spatial aggregation)和通道聚合(channel aggregation)3个部分(Liu等,2024a)。 F_{in} 先通过 1×1 卷积进行通道投影得到

特征 F_1 ,并使通道数从 C_1 降到 C_2 ,其中, $C_2 = \frac{C_1}{\alpha}$, α

表示预定义的通道调整因子。在多尺度融合部分,首先经过不同卷积核大小的小波卷积进行特征提取,并将输出结果执行加法运算得到多尺度融合特征 F_2 。上述过程数学表达式为

$$F_{in} = \text{Concat}(F_1^e, F_2^e, F_3^e) \quad (4)$$

$$F_1 = \text{Conv}_{k=1}(F_{in}) \quad (5)$$

$$F_2 = \sum_{k \in KS} \text{WTConv}_k(F_1) \quad (6)$$

式中,Concat表示拼接操作, $\text{Conv}_{k=1}$ 表示 1×1 的卷积,WTConv_k表示小波卷积,其中KS定义了一组卷积核,其值为 $\{3 \times 3, 5 \times 5, 7 \times 7\}$ 。

在空间聚合部分,首先将 F_2 在通道维度分别进行全局平均池化(average pooling, AvgPool)和最大池化(max pooling, MaxPool),得到两个 $H \times W \times 1$ 的特征图,并将这两个特征图进行拼接操作得到

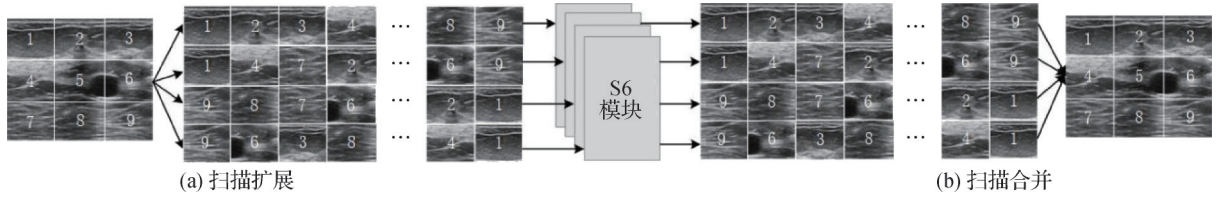


图3 2D选择性扫描模块

Fig. 3 2D selective scan module ((a) scan expanding; (b) scan merging)

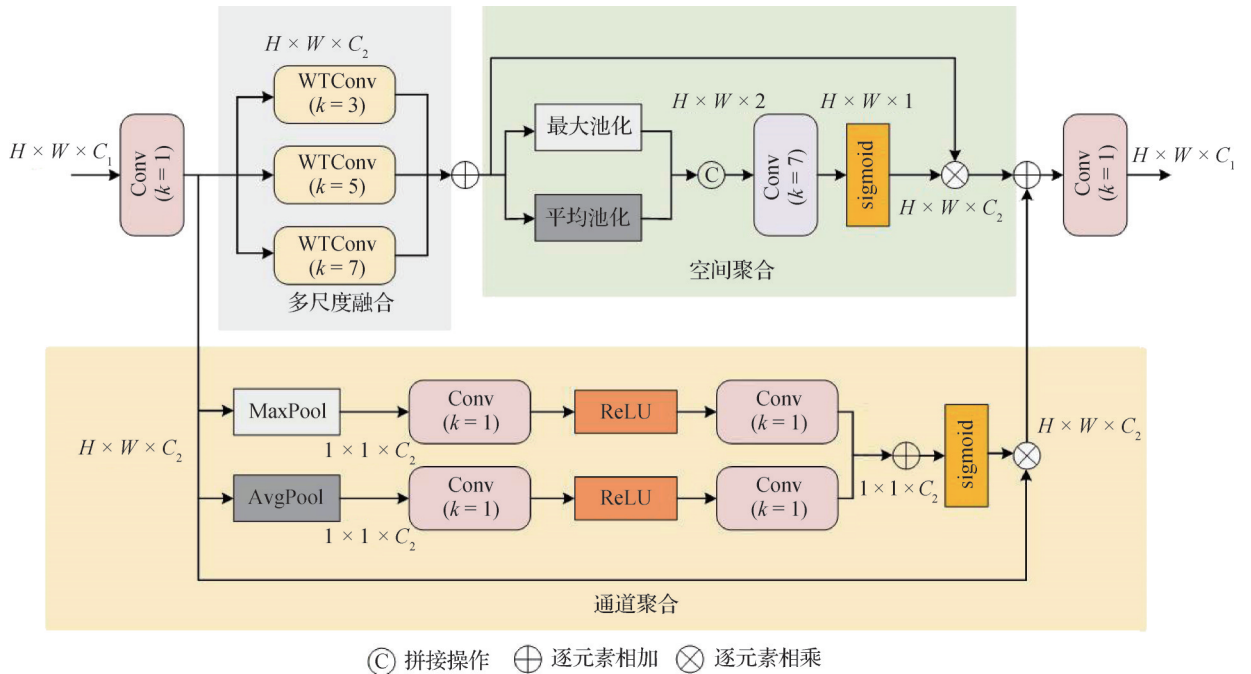


图4 小波卷积注意力聚合模块

Fig. 4 Wavelet convolution attention aggregation

特征 F_3 ; 随后经过 7×7 的卷积进一步提取特征以及 sigmoid 激活函数处理后与 F_2 进行 Hadamard 乘积以生成空间注意力图 F_s 。上述过程数学表达式为

$$F_3 = \text{Concat}(\text{AvgPool}(F_2), \text{MaxPool}(F_2)) \quad (7)$$

$$F_s = \text{sigmoid}(\text{Conv}_{k=7}(F_3)) \odot F_2 \quad (8)$$

式中, sigmoid 表示激活函数, $\text{Conv}_{k=7}$ 表示 7×7 的卷积, Concat 表示拼接操作, AvgPool 表示平均池化, MaxPool 表示最大池化。

在通道聚合部分, 首先对 F_1 在空间维度上分别进行全局最大池化和平均池化; 随后共同经过 1×1 卷积、ReLU (rectified linear unit) 激活函数和 1×1 卷积处理后得到特征 F_4 和 F_5 ; 接着将两个分支的输出执行加法运算, 所得结果经 sigmoid 激活函数处理后与 F_1 执行 Hadamard 乘积以生成通道注意力图 F_c ; 最后将空间注意力图和通道注意力图执行加法运算以融合空间特征和通道特征, 并将输出结果通过 1×1 卷积将通道数投影回 C_1 , 得到该模块的输出 F_{out} 。上述过程数学表达式为

$$F_4 = \text{Conv}_{k=1}(\text{ReLU}(\text{Conv}_{k=1}(\text{MaxPool}(F_1)))) \quad (9)$$

$$F_5 = \text{Conv}_{k=1}(\text{ReLU}(\text{Conv}_{k=1}(\text{AvgPool}(F_1)))) \quad (10)$$

$$F_c = \text{sigmoid}(F_4 + F_5) \odot F_1 \quad (11)$$

$$F_{\text{out}} = \text{Conv}_{k=1}(F_s + F_c) \quad (12)$$

式中, ReLU 、 sigmoid 表示激活函数。WCAA 模块能够增强空间维度和通道维度的特征表达, 从而为后续网络层提供更加丰富的特征表示。通过引入 WCAA 模块并在其中应用多尺度小波卷积, 不仅能增强特征提取效果, 还能使最终得到的特征图具备更强的空间信息和通道信息。

2.3 损失函数

在本文的医学图像分割任务中, 采用了二元交叉熵和 Dice 相似系数 (binary cross-entropy and Dice loss, BCEDice) 混合损失函数作为模型优化的目标函数。该损失函数通过结合二元交叉熵损失 (binary cross entropy loss, BCE) 与 Dice 相似系数损失, 有效平衡了像素级分类精度与区域重叠率的优化目标, 其定义为

$$L_{\text{BCEDice}} = \lambda_1 \times L_{\text{BCE}} + \lambda_2 \times L_{\text{Dice}} \quad (13)$$

$$\begin{cases} L_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [\gamma_i \log(\hat{\gamma}_i) + (1 - \gamma_i) \log(1 - \hat{\gamma}_i)] \\ L_{\text{Dice}} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \end{cases} \quad (14)$$

式中, N 表示样本总数, γ_i 和 $\hat{\gamma}_i$ 分别表示真实标签和预测掩码。集合 $|X|$ 和 $|Y|$ 分别表示真实标注和预测结果。参数 λ_1 和 λ_2 用于控制式中不同损失项的相对重要性, 且默认情况下均设定为 1。

2.4 KAN 调控多尺度加权损失

给定真实标签 G 、分割掩码 P 、后 3 个解码器层级输出的特征图为 $F_1^d \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times C}$ 、 $F_2^d \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times 2C}$ 和 $F_3^d \in \mathbb{R}^{\frac{H}{16} \times \frac{W}{16} \times 4C}$, 各特征图分别经过独立的 KAN 卷积捕捉非线性特征并生成各级预测掩码, 假定为 M_1 、 M_2 和 M_3 。同时, 使用最近邻插值法 (nearest) 将真实标签分别下采样至原始分辨率的 $\frac{1}{4}$ 、 $\frac{1}{8}$ 和 $\frac{1}{16}$, 以生成与各级预测掩码分辨率相同的真实标签, 假定为 G_1 、 G_2 和 G_3 。随后将其分别与对应的预测掩码进行逐像素的二元交叉熵和 Dice 相似系数混合损失计算, 并堆叠成多尺度损失向量 $\Omega \in \mathbb{R}^3$ 。上述过程数学表达式为

$$M_i = \text{KANConv}_{k=1}(F_i^d) \quad (15)$$

$$G_i = \text{Downsample}(G) \quad (16)$$

$$L_i = \text{BCEDice}(G_i, M_i) \quad (17)$$

$$\Omega = \text{Concat}(L_1, L_2, L_3) \quad (18)$$

式中, $i \in \{1, 2, 3\}$, $\text{KANConv}_{k=1}$ 表示 1×1 的 KAN 卷积, Downsample 表示下采样, M_i 表示各层级预测掩码, G_i 表示各层级真实标签, G 表示输入图像的真实标签, BCEDice 表示二元交叉熵 Dice 相似系数混合损失函数, L_i 表示各层级的损失值。

如图 5 所示, 将 F_1^d 、 F_2^d 和 F_3^d 一并输入到 KMWL 模块, 分别经全局最大池化后在通道维上进行特征拼接得到 F_m , 随后进入 KAN 层 (Li 等, 2024a) 以进一步捕捉非线性特征, 并经 softmax 函数归一化处理输出权重系数向量 $\Psi \in \mathbb{R}^3$, 最后与多尺度损失向量进行向量点积运算得到 KMWL 模块的输出, 即多尺度加权损失 L_m 。最终输入图像的真实标签与预测掩码经损失函数计算所得主损失与多尺度加权损失进行线性组合后得到最终损失 L_t 。

上述过程数学表达式为

$$F_m = \text{Concat}(\text{MaxPool}(F_1^d, F_2^d, F_3^d)) \quad (19)$$

$$\Psi = \text{softmax}(\text{KANLayer}(F_m)) \quad (20)$$

$$L_w = \langle \Omega, \Psi \rangle \quad (21)$$

$$L_i = \text{BCEDice}(G, P) + \mu L_w \quad (22)$$

式中, KANLayer 表示 KAN 层, softmax 表示归一化处理, BCEDice 表示二元交叉熵 Dice 相似系数混合损失函数, $\langle \cdot, \cdot \rangle$ 表示向量的点积运算, μ 为超参数, 其值在 0 到 1 之间。该设计通过 KAN 网络动态感知不同层级特征的优化状态, 使网络在训练过程中自适应调整各尺度特征的贡献权重。相比传统静态加权方法, 本机制在 KAN 的强非线性表征能力支持下, 能更精准捕捉层级间的特征依赖关系。

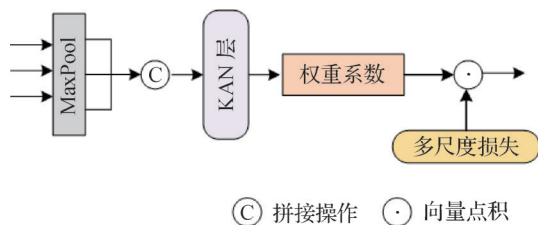


图 5 KAN 调控多尺度加权损失模块

Fig. 5 KAN-regulated multi-scale weighted loss module

3 实验设计和性能评估

3.1 数据集

本研究对所提方法进行了全面评估, 测试基于 3 个具有明显差异且异质性较强的数据集, 每个数据集都呈现出独特的特征、不同的数据规模以及分辨率差异较大的图像。这些数据集通常用于图像分割和生成等任务, 为方法的有效性和适应性提供了稳健的测试基准。

3.1.1 BUSI 乳腺超声数据集

BUSI (breast ultrasound images dataset) 乳腺超声数据集 (Al-Dhabyani 等, 2020) 由超声影像构成, 其中包含正常、良性及恶性乳腺癌病例的图像, 并附带相应的分割标注图。在本研究中, 使用了 647 例超声图像, 涵盖了良性及恶性乳腺肿瘤病例。所有图像均统一调整至 256×256 像素的分辨率。该数据集提供了一套完整的影像资源, 有助于乳腺肿瘤的检测与分类, 并为医学专业人士及研究人员提供了重要的学术洞察。

3.1.2 GlaS 腺体病理数据集

GlaS (gland segmentation in histology images chal-

lenge dataset) 腺体病理数据集 (Valanarasu 等, 2021) 由 31 组影像序列中的 612 帧标准清晰度 (standard definition, SD) 图像构成。每帧图像的分辨率为 384×288 像素, 并采集自 23 名患者。该数据集来源于西班牙巴塞罗那的临床医院。数据集中所有影像序列均采用 Olympus Q160AL 和 Q165L 设备进行采集, 并配备 Extra II 视频处理器进行处理。按照常见的数据处理规范, 本研究选取了 GlaS 数据集中的 165 幅图像, 并统一调整为 512×512 像素的分辨率以进行后续实验。

3.1.3 CVC-ClinicDB 息肉数据集

CVC (CVC-ClinicDB dataset) 息肉数据集 (Bernal 等, 2015), 常简称为 CVC 数据集, 是一种公开可用的资源, 主要用于结肠镜检查视频中的息肉诊断研究。该数据集共包含 612 幅图像, 每幅图像的分辨率均为 384×288 像素, 并精确提取自 31 组不同的结肠镜检查序列。这些图像涵盖了多种类型的息肉实例, 使其在息肉检测算法的开发与评估中尤为重要。为了确保不同数据集在研究中的一致性, 本文将 CVC-ClinicDB 数据集中的所有图像统一调整至 256×256 像素的分辨率。

3.2 实验环境与配置

本文实验都是基于 Python 3.8 和 PyTorch 1.13.0-cuda11.8 实现的, 所有实验都使用了一个 NVIDIA Tesla V100 GPU, 内存为 32 G。对于 BUSI、GlaS 和 CVC 数据集, 训练时, 学习率 (learning rate) 设定为 1×10^{-3} 。采用 Adam 优化器训练模型, 并使用余弦退火 (cosine annealing) 学习率调度策略, 其最小学习率设定为 1×10^{-5} 。数据集按 80% 用于训练、20% 用于验证和测试的比例随机划分。数据增强应用了基本的图像变换, 包括随机旋转和翻转。整个模型训练共进行了 600 轮。

3.3 评估指标

本研究从定性和定量两个维度对分割结果进行评估, 并使用交并比 (intersection over union, IoU) 和 F1 分数 (F1 score, F1) 两个关键指标进行比较。IoU 衡量预测前景区域与真实前景区域的空间重叠程度, 直接量化分割边界的空间匹配精度, 关注局部像素级误差, 定义为两者的交集面积与并集面积的比值。F1 分数是精确率 (precision) 与召回率 (recall) 的调和平均数, 通过调和平均降低精确率与召回率的偏倚, 综合评估模型分割性能。二者结合可全面反

映模型分割性能。IoU 和 F1 分数的值越高说明模型分割效果越好。

3.4 性能评估

本文提出的 VMAML-UNet 模型在 BUSI、GlaS 和 CVC 数据集上与先进模型进行对比。包括 U-Net (Ronneberger 等, 2015)、Att-UNet (Oktay 等, 2018)、U-Net++ (Zhou 等, 2018)、U-NeXt (Valanarasu 和 Patel, 2022)、Swin-UNet (Cao 等, 2023)、Rolling-UNet (Liu 等, 2024b)、U-Mamba (Ma 等, 2024)、U-KAN (Li 等, 2024a)、KAN-Mamba FusionNet (Agrawal 等, 2025)、KM-UNet (Zhang, 2025) 和 VM-UNet (Ruan 等, 2024)。

本文使用 VM-UNet 模型作为基准模型 (baseline), 性能评估采用 IoU 和 F1 分数两项指标, 结果如表 1 所示。可以看出, VMAML-UNet 在所有数据集上均优于基准模型和其他模型。

表 1 各种分割方法对比

Table 1 Comparison of various segmentation methods

%

方法	BUSI		GlaS		CVC	
	IoU	F1	IoU	F1	IoU	F1
U-Net	57.22	71.91	86.66	92.79	83.79	91.06
Att-UNet	55.18	70.22	86.84	92.89	84.52	91.46
U-Net++	57.41	72.11	87.07	92.96	84.61	91.53
U-NeXt	59.06	73.08	84.51	91.55	74.83	85.36
Swin-UNet	52.11	68.51	82.44	90.38	73.49	84.72
Rolling-UNet	61.00	74.67	86.42	92.63	82.87	90.48
U-Mamba	61.81	75.55	87.01	93.02	84.79	91.63
U-KAN	63.38	77.42	87.64	93.37	85.05	91.88
KAN-Mamba FN	62.41	76.41	84.92	91.82	84.93	91.72
KM-UNet	65.42	78.79	87.51	93.27	85.01	91.79
VM-UNet	62.76	77.12	87.33	93.24	87.13	93.12
VMAML-UNet (本文)	65.48	79.14	90.71	95.13	89.64	94.54

注:加粗字体表示各列最优结果。KAN-Mamba FN表示 KAN-Mamba FusionNet。

此外,进一步验证了将本文方法作为网络基准模型时的高效性。如表 2 所示,在多个数据集上统计了模型的参数量 (params) 和浮点运算次数 (floating point operations per second, FLOPs)。实验结果表

明,本文方法不仅在分割精度上超越多数分割算法,在计算效率方面也展现出相当水平。总体而言,在分割精度与计算效率的权衡分析中,本文方法展现出最优的综合性能。

同时,本文对比了超参数 μ 对模型性能的影响,结果如表 3 所示,可以看出,当 $\mu=0.5$ 时达到最优平衡。

表 2 各种分割方法的参数量与浮点运算次数的统计表

Table 2 Comparison of parameters and FLOPs across various segmentation methods

方法	Params/M	FLOPs/G
U-Net	35.53	524.20
Att-UNet	35.90	533.10
U-Net++	36.60	1 109
U-NeXt	1.470	4.58
Swin-UNet	27.17	9.02
Rolling-UNet	1.78	16.82
U-Mamba	86.30	2 087
U-KAN	6.35	14.02
KAN-Mamba FusionNet	13.96	6.67
KM-UNet	17.66	7.35
VM-UNet	27.43	4.11
VMAML-UNet (本文)	35.18	6.60

注:加粗字体表示各列最优结果。

表 3 超参数 μ 对模型性能的影响

Table 3 The influence of hyperparameter μ on model performance

%

μ	BUSI		GlaS		CVC	
	IoU	F1	IoU	F1	IoU	F1
0.40	65.33	78.69	90.52	95.05	89.31	94.23
0.45	65.12	78.63	90.67	95.11	89.28	94.18
0.50	65.48	79.14	90.71	95.13	89.64	94.54
0.55	64.87	78.44	90.56	95.06	89.13	94.12
0.60	64.82	78.41	90.58	95.08	89.24	94.16

注:加粗字体表示各列最优结果。

3.5 消融实验

为全面分析本文方法,以 VM-UNet 为基准模型,在 BUSI、GlaS 和 CVC 数据集上进行消融实验,结果如表 4 所示。可以看出,基准模型中引入 WTConv

层时,在 BUSI、GlaS 和 CVC 数据集上的 IoU 和 F1 分数分别提升 0.88% 和 0.66%、1.19% 和 0.67%、1.28% 和 0.73%。引入 WCAA 模块时,各数据集上的 IoU 和 F1 分数分别提升 1.65% 和 1.23%、1.21% 和 0.68%、1.65% 和 0.94%。引入 KMWL 模块时,各数据集上的 IoU 和 F1 分数分别提升 1.90% 和 1.42%、0.95% 和 0.52%、2.75 和 1.55。同时引入 WTConv 层、WCAA 和 KMWL 模块能够进一步提升性能,各数据集上的 IoU 和 F1 分数分别提升 2.72% 和 2.02%、3.38% 和 1.89%、2.51% 和 1.42%,说明 WTConv、WCAA 和 KMWL 模块具有相互促进作用。需要指出的是,在 CVC 数据集上单独引入 KMWL 取得最好效果,可能是因为息肉目标通常大小、形态各异,边界有时模糊,而 KMWL 这种能够自适应地、非

线性地加权不同尺度、不同层级特征的损失函数,能更好地指导网络学习区分不同大小和难度的息肉区域,尤其在捕捉边界和细节方面优于传统损失函数。

3.6 分割结果可视化

本文进一步针对所有数据集进行全面的定性比较,将其他方法与所提出的 VMAML-UNet 方法分割的结果进行可视化,结果如图 6 所示,第 1 列是原始图像,第 2 列是真实标签,其余各列是不同方法的分割结果。相比之下,本文方法相较于其他方法产生了更少的假阳性结果,证明其在抑制噪声预测方面具有显著优势。同时,本文方法的分割结果在器官边界和形态细节方面也展现出更精细的表征能力,这些结果能够充分验证其实现精细化分割的卓越性能。

表 4 消融实验结果
Table 4 Ablation test results

方法	/%					
	BUSI		GlaS		CVC	
	IoU	F1	IoU	F1	IoU	F1
Baseline	62.76	77.12	87.33	93.24	87.13	93.12
Baseline+WTConv	63.64	77.78	88.52	93.91	88.41	93.85
Baseline+WCAA	64.41	78.35	88.54	93.92	88.78	94.06
Baseline+KMWL	64.66	78.54	88.28	93.76	89.88	94.67
Baseline+WTConv+WCAA+KMWL	65.48	79.14	90.71	95.13	89.64	94.54

注:加粗字体表示各列最优结果。

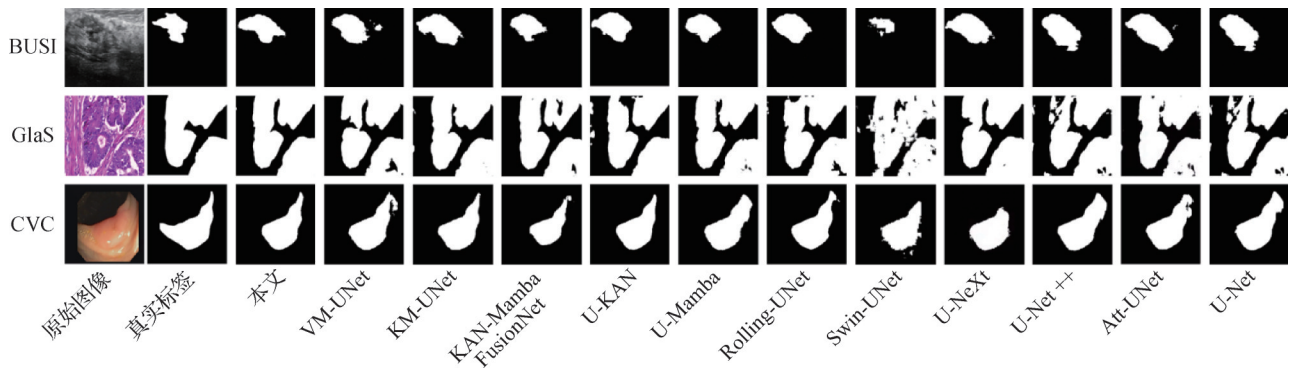


图 6 不同方法的可视化分割结果

Fig. 6 Visualization of segmentation results from different methods

4 结 论

医学图像分割作为精准医疗与疾病诊断的关键技术,对算法精度与效率的要求不断提升。现有基

于 CNN 的医学图像分割方法对长距离依赖关系的建模能力有限,很难充分利用图像中的语义信息,并且基于 Transformer 的方法复杂度太高,而 Mamba 能够以线性复杂度建模长距离依赖关系。同时,KAN 强大的非线性建模能力,有助于更精准地分割复杂

的医学图像结构。针对上述原因,本文提出一种采用KAN调控多尺度加权损失和视觉Mamba的医学图像分割方法。首先,在编码器和解码器阶段均引入了小波卷积的视觉Mamba块,不仅精准提取了病变区域特征、增大了感受野,还以线性复杂度高效建模了长距离依赖关系;其次,在跳跃连接阶段设计了小波卷积注意力聚合模块,能有效整合不同尺度的图像特征信息,使生成的特征图在空间和通道维度上得到了丰富和细化;最后,为突破传统静态加权策略的局限,设计了KAN调控的多尺度加权损失模块,在不同层级的解码器中使用KAN卷积输出多尺度预测掩码并计算多尺度的分层损失,通过动态调控各层级损失权重,精准捕捉了层级间的特征依赖关系。在3个公开的医学数据集上进行了对比实验。实验结果表明,本文方法在交并比和F1分数这两个指标上均获得了明显提升且取得了最好的效果。未来,将进一步优化模型的性能和效率,并继续探究模型在其他任务中的应用,如图像检测、图像配准、图像重建以及化学反应连续流的工艺优化等。

参考文献(References)

- Agrawal A, Agrawal A, Gupta S and Bagade P. 2025. KAN-Mamba FusionNet: redefining medical image segmentation with non-linear modeling [EB/OL]. [2025-05-10].
<https://arxiv.org/pdf/2411.11926.pdf>
- Al-Dhabyani W, Gomaa M, Khaled H and Fahmy A. 2020. Dataset of breast ultrasound images. *Data in Brief*, 28: #104863 [DOI: 10.1016/j.dib.2019.104863]
- Ba J L, Kiros J R and Hinton G E. 2016. Layer normalization [EB/OL]. [2025-05-10]. <https://arxiv.org/pdf/1607.06450.pdf>
- Bernal J, Sánchez 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]
- Bodner A D, Tepsich A S, Spolski J N and Pourteau S. 2025. Convolutional Kolmogorov-Arnold networks [EB/OL]. [2025-05-10].
<https://arxiv.org/pdf/2406.13155.pdf>
- Cao H, Wang Y Y, Chen J, Jiang D S, Zhang X P, Tian Q, et al. 2023. Swin-unet: Unet-like pure transformer for medical image segmentation//*Proceedings of 2023 European Conference on Computer Vision*. Tel Aviv, Israel: Springer: 205-218 [DOI: 10.1007/978-3-031-25066-8_9]
- Chen J N, Lu Y Y, Yu Q H, Luo X D, Adeli E, Wang Y, et al. 2021. TransUNet: transformers make strong encoders for medical image segmentation [EB/OL]. [2025-05-10].
<https://arxiv.org/pdf/2102.04306.pdf>
- Çiçek Ö, Abdulkadir A, Lienkamp S S, Brox T and Ronneberger O. 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation//*Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016*. Athens, Greece: Springer: 424-432 [DOI: 10.1007/978-3-319-46723-8_49]
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, et al. 2021. An image is worth 16 × 16 words: transformers for image recognition at scale [EB/OL]. [2025-05-10].
<https://arxiv.org/pdf/2010.11929.pdf>
- Elfwing S, Uchibe E and Doya K. 2018. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks*, 107: 3-11 [DOI: 10.1016/j.neunet.2017.12.012]
- Finder S E, Amoyal R, Treister E and Freifeld O. 2025. Wavelet convolutions for large receptive fields//*Proceedings of the 18th European Conference on Computer Vision*. Milan, Italy: Springer: 363-380 [DOI: 10.1007/978-3-031-72949-2_21]
- Gu A and Dao T. 2024. Mamba: linear-time sequence modeling with selective state spaces [EB/OL]. [2025-05-10].
<https://arxiv.org/pdf/2312.00752.pdf>
- Gu A, Dao T, Ermon S, Rudra A and Ré C. 2020. HiPPO: recurrent memory with optimal polynomial projections//*Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates Inc.: 1474-1487
- Gu A, Goel K and Ré C. 2022. Efficiently modeling long sequences with structured state spaces [EB/OL]. [2025-05-10].
<https://arxiv.org/pdf/2111.00396.pdf>
- Gu A, Johnson I, Goel K, Saab K, Dao T, Rudra A, et al. 2021. Combining recurrent, convolutional, and continuous-time models with linear state-space layers//*Proceedings of the 35th International Conference on Neural Information Processing Systems*. Virtual Event: Curran Associates Inc.: 572-585
- Hornik K, Stinchcombe M and White H. 1989. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2 (5): 359-366 [DOI: 10.1016/0893-6080(89)90020-8]
- Kalman R E. 1960. A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering*, 82(1): 35-45 [DOI: 10.1115/1.3662552]
- Li C X, Liu X Y, Li W Y, Wang C, Liu H Y, Liu Y F, et al. 2024a. U-KAN makes strong backbone for medical image segmentation and generation [EB/OL]. [2025-05-10].
<https://arxiv.org/pdf/2406.02918.pdf>
- Li H, Zhai D H and Xia Y Q. 2024b. ERDUNet: an efficient residual double-coding UNet for medical image segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4): 2083-

- 2096 [DOI: 10.1109/TCSVT.2023.3300846]
- Liu J M and Tang Y C. 2025. Two-stage coronary artery segmentation via multidirectional snake convolution and vision Mamba. *Journal of Image and Graphics*, 30(10): 3242-3254 (刘建明, 唐煜城. 采用多方向蛇形卷积和视觉残差 Mamba 的两阶段冠状动脉分割方法. *中国图象图形学报* [DOI: 10.11834/jig.240538])
- Liu J X, Chen Y, Ni B B and Yu Z B. 2023. Joint global and dynamic pseudo labeling for semi-supervised point cloud sequence segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(10): 5679-5691 [DOI: 10.1109/TCSVT.2023.3253210]
- Liu M S, Dan J, Lu Z Q, Yu Y L, Li Y M and Li X. 2024a. CM-UNet: hybrid CNN-Mamba UNet for remote sensing image semantic segmentation [EB/OL]. [2025-05-10]. <https://arxiv.org/pdf/2405.10530.pdf>
- Liu Y, Tian Y J, Zhao Y Z, Yu H T, Xie L X, Wang Y W, et al. 2025a. VMamba: visual state space model//*Proceedings of the 38th International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates Inc.: 103031-103063
- Liu Y T, Zhu H J, Liu M T, Yu H Y, Chen Z H and Gao J. 2024b. Rolling-UNet: revitalizing MLP's ability to efficiently extract long-distance dependencies for medical image segmentation//*Proceedings of the 38th AAAI Conference on Artificial Intelligence*. Vancouver, Canada: AAAI: 3819-3827 [DOI: 10.1609/aaai.v38i4.28173]
- Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, et al. 2021. Swin transformer: hierarchical vision transformer using shifted windows//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: IEEE: 9992-10002 [DOI: 10.1109/ICCV48922.2021.00986]
- Liu Z M, Wang Y X, Vaidya S, Ruehle F, Halverson J, Soljačić M, et al. 2025b. KAN: Kolmogorov-Arnold networks [EB/OL]. [2025-05-10]. <https://arxiv.org/pdf/2404.19756.pdf>
- Ma J, Li F F and Wang B. 2024. U-Mamba: enhancing long-range dependency for biomedical image segmentation [EB/OL]. [2025-05-10]. <https://arxiv.org/pdf/2401.04722.pdf>
- Mu P, Wu G Y, Liu J Y, Zhang Y D, Fan X and Liu R S. 2024. Learning to search a lightweight generalized network for medical image fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(7): 5921-5934 [DOI: 10.1109/TCSVT.2023.3342808]
- Myronenko A. 2019. 3D MRI brain tumor segmentation using autoencoder regularization//*4th International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Granada, Spain: Springer: 311-320 [DOI: 10.1007/978-3-030-11726-9_28]
- Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, et al. 2018. Attention U-net: learning where to look for the pancreas [EB/OL]. [2025-05-10]. <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 2015*. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Ruan J C, Li J C and Xiang S C. 2024. VM-UNet: Vision Mamba UNet for medical image segmentation [EB/OL]. [2025-05-10]. <https://arxiv.org/pdf/2402.02491.pdf>
- Ruan J C, Xiang S C, Xie M Y, Liu T and Fu Y Z. 2022a. MALUNet: a multi-attention and light-weight UNet for skin lesion segmentation//*Proceedings of 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Las Vegas, USA: IEEE: 1150-1156 [DOI: 10.1109/BIBM55620.2022.9995040]
- Ruan J C, Xie M Y, Gao J S, Liu T and Fu Y Z. 2023. EGE-UNet: an efficient group enhanced UNet for skin lesion segmentation//*Proceedings of the 26th International Conference on Medical Image Computing and Computer Assisted Intervention*. Vancouver, Canada: Springer: 481-490 [DOI: 10.1007/978-3-031-43901-8_46]
- Ruan J C, Xie M Y, Xiang S C, Liu T and Fu Y Z. 2022b. MEW-UNet: multi-axis representation learning in frequency domain for medical image segmentation [EB/OL]. [2025-05-10]. <https://arxiv.org/pdf/2210.14007.pdf>
- Valanarasu J M J, Oza P, Hacıhaliloğlu I and Patel V M. 2021. Medical transformer: gated axial-attention for medical image segmentation//*Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention-MICCAI 2021*. Strasbourg, France: Springer: 36-46 [DOI: 10.1007/978-3-030-87193-2_4]
- Valanarasu J M J and Patel V M. 2022. UNeXt: MLP-based rapid medical image segmentation network//*Proceedings of the 25th International Conference on Medical Image Computing and Computer Assisted Intervention*. Singapore, Singapore: Springer: 23-33 [DOI: 10.1007/978-3-031-16443-9_3]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. 2017. Attention is all you need//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, USA: Curran Associates Inc.: 6000-6010
- Wang T, Lu C H, Sun Y N, Yang M, Liu C and Ou C S. 2021. Automatic ECG classification using continuous wavelet transform and convolutional neural network. *Entropy*, 23(1): #119 [DOI: 10.3390/e23010119]
- 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)*. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Xing Z, Ye T, Yang Y, Liu G and Zhu L. 2024. SegMamba: long-range sequential modeling Mamba for 3D medical image segmentation//*Proceedings of the 27th International Conference on Medical Image Computing and Computer Assisted Intervention*. Marrakesh,

Morocco: Springer: 578-588 [DOI: 10.1007/978-3-031-72111-3_54]

Zhang Y B. 2025. KM-UNet KAN Mamba UNet for medical image segmentation [EB/OL]. [2025-05-10].
<https://arxiv.org/pdf/2501.02559.pdf>

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

Zhou T, Zhou Y, Li G Y, Chen G and Shen J B. 2024. Uncertainty-aware hierarchical aggregation network for medical image segmentation. IEEE Transactions on Circuits and Systems for Video Technology, 34(8): 7440-7453 [DOI: 10.1109/TCSVT.2024.3370685]

Zhou Z W, Rahman Siddiquee M M, Tajbakhsh N and Liang J M. 2018. UNet++: a nested U-Net architecture for medical image segmentation//Deep Learning in Medical Image Analysis and Multimodal

Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018. Granada, Spain: Springer: 3-11 [DOI: 10.1007/978-3-030-00889-5_1]

Zhu L H, Liao B C, Zhang Q, Wang X L, Liu W Y and Wang X G. 2024. Vision Mamba: efficient visual representation learning with bidirectional state space model [EB/OL]. [2025-05-10].
<https://arxiv.org/pdf/2401.09417.pdf>

作者简介

刘建明,男,副教授,主要研究方向为医学图像处理和小样本学习。E-mail:liujianming@jxnu.edu.cn

曹圣浩,通信作者,男,硕士研究生,主要研究方向为计算机视觉。E-mail: 202341600093@jxnu.edu.cn

张志鹏,女,硕士研究生,主要研究方向为小样本学习和医学图象处理。E-mail:202341600157@jxnu.edu.cn