VMA-UNet:基于Mamba的多尺度医学图像分割网络

王海,李亚鸽,林愉萱,陆雪松*
(中南民族大学 生物医学工程学院, 武汉 430074)

摘 要 卷积神经网络(CNN)在医学图像分割中取得了显著的进展,但其在捕捉长距离依赖信息方面存在局限性.虽然Transformer模型在处理远程依赖方面表现出色,但自注意力机制导致了较高的计算成本.为了解决这些问题,提出了多尺度医学图像分割网络VMA-UNet(VMamba ASPP U-Net),它融合了VMamba的VSS块结构和空洞空间卷积池化金字塔(ASPP)模块.VMA-UNet利用VSS块的线性复杂度特性,实现高效的全局信息建模,并结合ASPP模块在多个尺度上捕捉医学图像中的关键特征.通过在ACDC、COVID-19 CT和Synapse等数据集上的广泛实验,结果表明:VMA-UNet在分割精度和计算效率上均优于基于CNN和Transformer方法,显示了其在不同任务中的竞争力.VMA-UNet克服了CNN在捕捉远程信息方面的局限性,实现了高效的多尺度建模,展现了其在医学图像分割中的巨大潜力.

关键词 医学图像分割;VMamba技术;ASPP模块;多尺度建模
中图分类号 TP391.4 **文献标志码** A **文章编号** 1672-4321(2026)01-0051-09
doi:10.20056/j.cnki.ZNMDZK.20250824

VMAU-Net: A multi-scale medical image segmentation network based on Mamba

WANG Hai, LI Yage, LIN Yuxuan, LU Xuesong*
(College of Biomedical Engineering, South-Central Minzu University, Wuhan 430074, China)

Abstract Convolutional Neural Networks (CNN) have achieved remarkable progress in medical image segmentation, but it exhibits limitations in capturing long-range dependencies. While Transformer models excel in handling long-range dependencies, the self-attention mechanism incurs high computational costs. To address these issues, VMA-UNet(VMamba ASPP U-Net) is proposed, it is a multi-scale medical image segmentation network that integrates the VSS (Visual State-Space) block structure from VMamba and the Atrous Spatial Pyramid Pooling (ASPP) module. VMA-UNet leverages the linear complexity of the VSS block to enable efficient global information modeling and incorporates the ASPP module to capture critical features of medical images at multiple scales. Extensive experiments conducted on datasets including ACDC, COVID-19 CT, and Synapse demonstrate that VMA-UNet outperforms CNN- and Transformer-based methods in terms of segmentation accuracy and computational efficiency. By overcoming the limitations of CNN in modeling long-range dependencies and enabling efficient multi-scale modeling, VMA-UNet showcases its immense potential in medical image segmentation tasks.

Keywords medical image segmentation; VMamba; ASPP; multi-scale modeling

医学图像分割的目标是自动或半自动地识别和分割出医学图像中的重要结构,为病理研究和临床诊断提供可靠依据,帮助医生做出更为准确的判断.其鲁棒性和准确性对临床诊断和治疗(如计算机辅助诊断、术前评估和图像引导手术)有着至关重要的作用[1-2].

随着深度学习技术的迅速发展,卷积神经网络(Convolutional Neural Networks, CNN)在医学成像领

收稿日期 2024-11-24
* **通信作者** 陆雪松(1975-),男,教授,博士,研究方向:医学图像分析,E-mail:xslu-scuec@hotmail.com
基金项目 湖北省自然科学基金资助项目(2016CFB489);中央高校基本科研业务费专项资助项目(CZZ24014)

域的应用中已占据主导地位.U-Net[3]及其变体,如Attention U-Net[4]、U-Net++[5]、U-Net3+[6]和nnU-Net[7]等,在医学图像分割方面取得了显著成功,其关键在于包含编解码器的U型架构设计.虽然CNN在特征学习上表现出色,但由于卷积算子的固有局部性,它们在捕获远程信息方面的能力受到显著制约.这种局限性可能导致特征提取不充分,进而影响分割结果的准确性.

为此,研究者们从自然语言处理领域取得突破的Transformer[8]模型中受到启发.Vision Transformer(ViT)的引入有效解决了远程信息捕获不足的问题,开启了利用自注意力机制捕捉全局信息的新纪元,为医学图像分割提供了更加精确的全局视角.TransUNet[9]是首个将CNN与ViT结合的模型,通过CNN提取局部特征,利用ViT进行全局信息建模,从而提升了医学图像分割的性能.紧随其后,UNETR[10]将ViT应用于三维数据分割,通过多头自注意力和多层感知机构建主编码器提取全局信息,并跳跃连接到CNN解码器,推动了医学图像分割的发展.nnformer[11]引入局部和全局自注意力机制来学习体素特征,充分发挥了Transformer在医学图像分割的优势.此外Swin-UNet[12]单纯采用Swin Transformer[13]构建U形架构的编码器和解码器,应用于二维医学图像分割.尽管Transformer在处理远程依赖关系方面表现出色,但其自注意力机制的计算成本仍然较高.因此,如何有效降低Transformer的高计算成本一直是研究热点.

最近,状态空间模型(State Space Model, SSM)在计算机视觉领域受到广泛关注,尤其是Mamba结构模型展现出的长距离依赖建模能力和线性复杂度优势.U-Mamba率先将CNN和Mamba相结合,充分利用CNN的局部特征提取能力和SSM的全局信息捕获能力,性能超越了传统的CNN和Transformer架构.SegMamba[14]则提出了基于Mamba三维医学图像分割模型,与基于CNN和Transformer混合架构相比,不仅保持了出色的推理效率,而且在空间维度上面也表现了卓越的远程建模能力.此外,LightM-UNet将U-Net与Mamba集成到轻量化架构中,实验表明其能够有效捕捉远程信息,同时计算量更小,性能优于传统CNN和Transformer方法.VM-UNet作为首个纯基于SSM的医学图像分割模型,在长距离依赖建模和计算效率方面超越了基于CNN和Transformer的方法,进一步凸显了SSM在医学图像分割领域的应用前景.

多尺度建模已经被证明[15-16]能够产生丰富的语义特征.许多方法利用多尺度信息来提升医学图像分割的性能,如DRINet[17]提出了一种残差初始模块组成的扩张反卷积,高效率捕获多尺度信息;CE-Net[18]引入了残差多核池化块,采用各种大小的池化操作来有效地编码多尺度上下文特征;MultiResUNet[19]从三个连续的卷积块获得输出,并将它们连接起来以不同尺度提取空间特征.本文提出以Mamba为核心模块的U型网络,采用空洞空间卷积池化金字塔(Atrous Spatial Pyramid Pooling, ASPP)[20]对高级特征图进行建模,捕捉多种尺度的上下文信息应用于医学图像分割.在三个数据集上进行了详细的实验验证,结果表明该VMAU-Net(VMamba ASPP U-Net)在医学图像分割中具有较强的竞争力.

1 方法

1.1 SSM原理

状态空间模型SSM,通过中间隐式状态$h(t)\in\mathcal{R}^N$将一维输入函数或序列$x(t)\in\mathcal{R}$映射到输出$y(t)\in\mathcal{R}$,上述过程可以表示为线性常微分方程:

$$\begin{aligned} h'(t) &= \boldsymbol{A}h(t)+\boldsymbol{B}x(t)\\ y(t) &= \boldsymbol{C}h(t)\end{aligned}, \tag{1}$$

其中$\boldsymbol{A}\in\mathcal{R}^{N\times N}$表示状态矩阵,$\boldsymbol{B}\in\mathcal{R}^{N\times 1}$和$\boldsymbol{C}\in\mathcal{R}^{N\times 1}$表示线性投影参数.

为了满足深度学习的需求,该连续系统必须经历离散化过程.具体来说,可以引入一个时间尺度参数Δ,并使用固定的离散化规则将$\boldsymbol{A}$和$\boldsymbol{B}$,转换为离散参数$\overline{\boldsymbol{A}}$和$\overline{\boldsymbol{B}}$,通常采用零阶保持器作为离散化规则,其定义如下:

$$\begin{aligned} \bar{\boldsymbol{A}} &= \exp(\Delta\boldsymbol{A})\\ \bar{\boldsymbol{B}} &= (\Delta\boldsymbol{A})^{-1}(\exp(\Delta\boldsymbol{A})-\boldsymbol{I})\cdot\Delta\boldsymbol{B}\end{aligned}, \tag{2}$$

因此,式(1)经过离散化得到式(3)

$$\begin{aligned} h_t &= \bar{\boldsymbol{A}}h_{t-1}+\bar{\boldsymbol{B}}x_t\\ y_t &= \boldsymbol{C}h_t\end{aligned}, \tag{3}$$

为了提高计算效率和可扩展性,通过全局卷积的方式对式(3)进行计算:

$$\begin{aligned} \bar{\boldsymbol{K}} &= \left(\boldsymbol{C}\bar{\boldsymbol{B}},\boldsymbol{C}\overline{\boldsymbol{A}\boldsymbol{B}},\cdots,\bar{\boldsymbol{A}}^{L-1}\bar{\boldsymbol{B}}\right)\\ \mathbf{y} &= \mathbf{x}*\bar{\boldsymbol{K}}\end{aligned}, \tag{4}$$

其中,L表示输入序列x的长度,$\bar{\boldsymbol{K}}\in\mathcal{R}^L$作为SSM的卷积核,并且$*$表示卷积运算.

1.2 总体架构

VMAU-Net的构架如图1所示.主要包括Patch

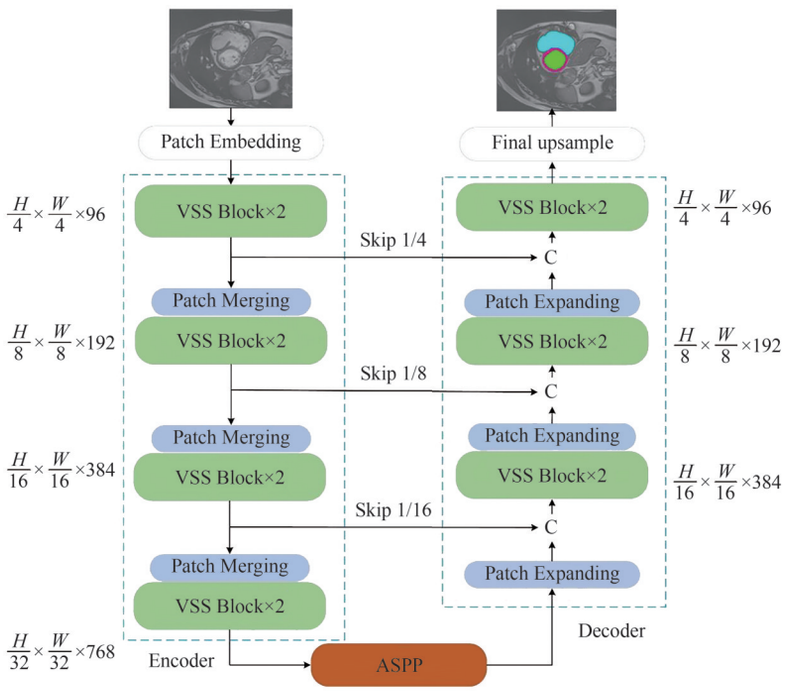

图 1　VMA-UNet 网络结构图

Fig. 1　VMA-UNet network structure

Embedding、编码器、ASPP 模块、解码器、线性投影层及跳跃连接. 首先，Patch Embedding 将输入图像划分为 4×4 的不重叠补丁，并嵌入到 96 维特征空间中. 保留了局部细节，为多尺度特征提取奠定基础. 随后，图像进入编码器进行多层次处理，利用 VSS（Visual State-Space）块和 Patch Merging 层逐步提取特征并下采样. 每次下采样将分辨率减半，通道数加倍. 最终将特征图从 $\frac{H}{4}\times\frac{W}{4}\times96$ 缩减到 $\frac{H}{32}\times\frac{W}{32}\times768$，从而提升了网络捕捉局部和全局信息的能力. 瓶颈层的 ASPP 模块通过不同扩张率的卷积捕捉多尺度上下文信息，增强对复杂结构的感知能力. 解码器采用对称设计，由 VSS 块和 Patch Expanding 层组成，通过上采样逐步恢复空间细节，确保输出与编码器特征大小一致，最终生成分割结果.

其中编码器通过逐步下采样特征图提取多尺度信息，每层由多个 VSS 块和 Patch Merging 组成. VSS 块保持分辨率与维度一致，Patch Merging 则逐层降低分辨率、增加通道数，从而增强不同尺度下的上下文表达能力.

解码器通过逐步恢复特征图分辨率并融合编码器的跳跃连接特征，实现高效特征重建. 每一层使用两个连续的 VSS 块处理特征，确保输入和输出的分辨率与维度一致. 同时，解码器引入 Patch Expanding 层，将特征图分辨率扩大 2 倍，特征维度减半，使特征逐层上采样，最终恢复到输入图像的原始分辨率.

跳跃连接确保了细节信息在恢复过程中的完整保留，通过将编码器中的高分辨率特征直接传递到解码器的对应层，有效保留了关键的空间信息. 这种简洁高效的特征融合方式显著提升了分割精度，同时避免了计算成本的增加.

1.3　VSS

以 SSM 为基础，VSS 模块通过引入二维选择性扫描（2D Selective Scan，SS2D）策略，实现由 Mamba 到 Vision Mamba 的转变. 如图 2 所示，输入张量经过 Layer Normalization 层后，被分成两个分支. 在第一

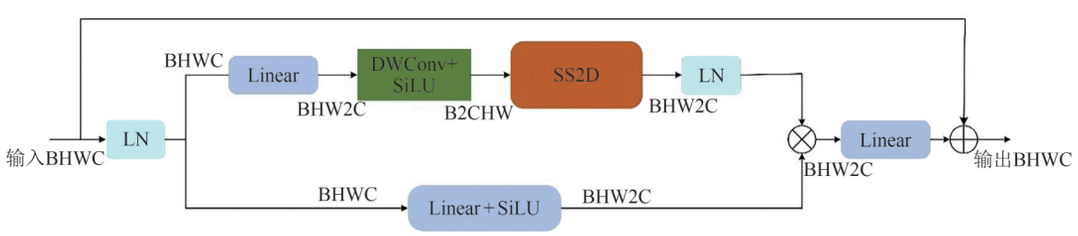

图 2　VSS 块结构

Fig. 2　VSS block structure

个分支中通过一个线性层和一个激活函数SiLU[21]；在第二个分支中，经过线性层，深度可分离卷积和激活函数SiLU处理，输入到SS2D模块中进行特征提取，随后使用Layer Normalization对特征进行规范化，最后，两个分支的输出被执行逐元素合并，进一步使用线性层混合特征，并将此结果进行残差连接形成VSS块的输出.

在SS2D模块的特征提取过程中，首先通过交叉扫描将输入特征沿四个方向展开：按行、按列、翻转后按行、翻转后按列.这些展开后的特征序列被堆叠，形成多方向表示，以捕捉不同方向的信息.接着，S6操作利用S4状态空间模型的动态适应性，自动调整模型的建模能力，从而有效处理大规模输入并保持对长距离依赖的敏感性，提升全局建模的灵活性与精准度.最后，特征进入交叉融合阶段，重新从四个方向获取特征并相加融合，恢复至与输入相同的维度.通过这一多方向展开与融合策略，SS2D模块在二维数据中能够更全面地捕捉方向性与位置性特征，增强图像特征提取和边缘细节表达能力，提升模型的鲁棒性和精度.

1.4 ASPP

如图3所示，ASPP模块通过不同扩张率(r1，r6，r12，r18表示不同的空洞率，空洞率越大，感受野越大)的卷积操作，从多个尺度捕捉图像的上下文信息，增强了对多尺度结构的感知能力.同时，ASPP利用全局平均池化进一步强化了全局上下文信息的捕捉.最后，通过1×1卷积将多尺度特征融合为统一的低维特征表示，从而有效地提升了分割性能.

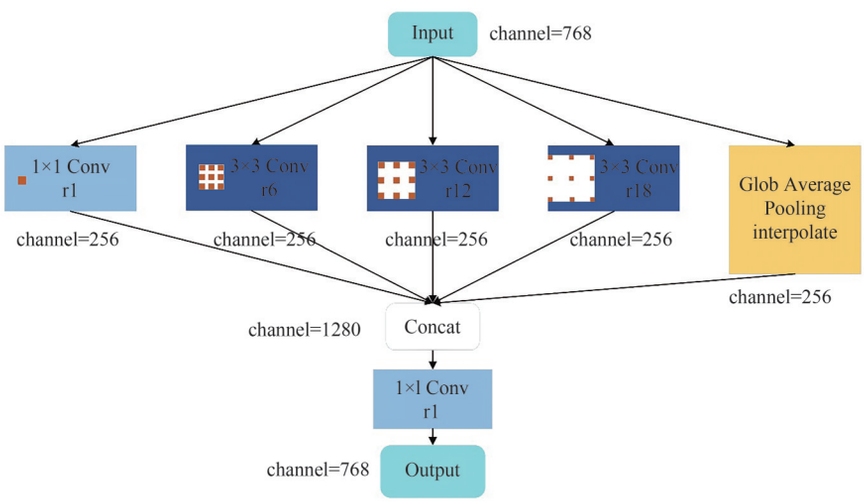

图3 ASPP结构

Fig. 3 ASPP structure

2 实验与结果

2.1 数据集

2.1.1 自动心脏诊断挑战数据集(ACDC)

该数据集包含100名不同患者的MRI扫描图像，每个患者的MRI图像标注了心脏的主要结构，包括左心室、右心室和心肌.每个体数据由28至40层切片组成，切片厚度在5 mm至10 mm之间.数据集按照7:1:2比例划分训练集(共1353个切片)，验证集(共145个切片)以及测试集(共404个切片)，在验证集上选择最佳权重进行最终测试.

2.1.2 COVID-19CT肺和感染分割数据集(COVID19 CT-Seg)

该数据集包含1836张COVID-19切片和1637张非COVID-19切片，所有切片均来自20次COVID-19 CT扫描.每个CT体的平均切片数量为175张，切片大小为512×512或630×630在每次CT扫描中，左肺、右肺和新冠肺炎感染区域均被独立标注.为了更集中地进行新冠肺炎感染区域的分割，本研究将左肺和右肺的标签统一为一个类别，同时保持新冠肺炎感染区域的标签不变.数据集按照7:1:2比例划分为训练集(共2504个切片)，验证集(286个轴向切片)以及测试集(共683个切片)，在验证集上选择最佳权重进行最终测试.

2.1.3 Synapse多器官分割数据集

Synapse是一个公开的多器官分割数据集，包括了30例腹部CT扫描，共有3779张轴向切片对比增强的腹部临床CT图像.每个CT体由85-198个512×512的切片组成.本文采用18个病例(2212个切片)作为

训练数据集，12 个病例（1567 个切片）作为测试数据集．我们使用平均 Hausdorff 距离和平均 Dice 相似系数作为评估指标来评价 CT 图像中 8 个器官的分割性能，包括主动脉（Aorta）、胆囊（Gallbladder）、脾（Spleen）、左肾（Kidney（L））、右肾（Kidney（R））、肝脏（Liver）、胰腺（Pancreas）和胃（Stomach）．

2.2　评价指标

Dice 相似系数（DSC），Hausdorff 距离 95%（HD95）、精确度（Precision）、召回率（Recall）和交并比（IoU）被用于评估分割准确性．HD95 计算的是地面真实值和预测点集之间表面距离的第 95 个百分位数．DSC、Precision、Recall 和 IoU 与以下四个值相关：真阳性（TP）、真阴性（TN）、假阳性（FP）和假阴性（FN）．这些指标的计算公式如下：

$$\text{Dice} = \frac{2\sum_{i=1}^{I} T_i P_i}{\sum_{i=1}^{I} T_i + \sum_{i=1}^{I} P_i}, \tag{5}$$

$$\text{HD} = \max\left\{ \max_{t' \in T'} \min_{p' \in P'} \| t' - p' \|, \max_{p' \in P'} \min_{t' \in T'} \| p' - t' \| \right\}, \tag{6}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{8}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \tag{9}$$

其中 T 和 P 分别代表体素值的真实值和预测值，T' 和 P' 分别代表地面真值和预测表面点集．

2.3　实验细节

所有实验均在 Python 3.10、Pytorch 2.0.1、CUDA 11.8 和 Ubuntu 20.04 环境下进行．训练硬件为 NVIDIA GeForce RTX 2080Ti（11 GB 显存）支持．所有图像被重采样到 224 × 224 尺寸，作为 ACDC、COVID19 CT-Seg 和 Synapse 数据集的输入．为了提高模型的泛化能力，训练过程中应用了随机翻转和旋转等数据增强策略．受 Swin-UMamba[22] 模型启发，在网络的编码器中加载了预训练的 VMamba-Tiny[23] 模型权重．VMA-UNet 采用随机初始化，批量大小为 20，优化器为随机梯度下降，学习率为 0.01，动量为 0.9，权重衰减为 0.0001，并使用了余弦退火调度器（见公式 11）进行学习率调节．损失函数由交叉熵损失和骰子损失的加权平均构成：

$$\text{loss}_{\text{total}} = w_1 \text{loss}_{\text{ce}} + w_2 \text{loss}_{\text{dice}}, \tag{10}$$

其中 $\text{loss}_{\text{total}}$，$\text{loss}_{\text{ce}}$，$\text{loss}_{\text{dice}}$ 分别为总损失，交叉熵损失和骰子损失．根据参数微调的结果，w_1 和 w_2 分别设置为 0.4 和 0.6 是最佳的．

余弦衰减调度器表示如下：

$$\eta_t = \eta_T + \frac{\eta_0 - \eta_T}{2}\left(1 + \cos\left(\frac{\pi t}{T}\right)\right), \tag{11}$$

其中 η_0 表示初始学习率，T 表示更新学习率的最大步数，设置为 20，η_T 表示 T 处的学习率，t 表示时间．

2.4　实验结果对比

2.4.1　ACDC 数据集上的实验结果

表 1 展示了 VMA-UNet 在 ACDC 数据集上的定量实验结果．该模型在 DSC、Recall 和 IoU 等关键指标上均优于其他方法，分别达到 87.71%、89.37% 和 80.84%，较次优模型分别提高了 1.15%、2.54% 和 1.29%．这些显著提升表明 VMA-UNet 在心脏 MRI 分割任务中具备卓越的特征捕捉能力，并在全局信息建模方面展现了强大潜力．然而，模型的 HD95 较高，表明尽管整体分割效果较好，但在处理细微边缘时存在一定不足，特别是在边界模糊或轻微偏差的情况下，局部误差可能导致 HD95 增加．尽管边缘分割精度尚需进一步优化，但 VMA-UNet 在大多数指标上的出色表现充分证明了其强大的竞争力．

表 1　ACDC 数据集分割性能的定量比较，最优结果用粗体表示，次优结果用下划线表示．

Tab. 1　Quantitative comparison of ACDC data set segmentation performance, optimal results are shown in bold, sub-optimal results are shown in underline

Methods	Average					Right ventricle		Myocardium		Left ventricle	
	DSC ↑	HD95 ↓	Precision ↑	Recall ↑	IoU ↑	DSC ↑	HD95 ↓	DSC ↑	HD95 ↓	DSC ↑	HD95 ↓
V-Net[24]	80.87	3.06	89.21	81.38	72.76	81.43	3.66	75.79	3.06	85.38	2.47
U-Net[3]	84.23	2.73	<u>89.71</u>	82.72	75.51	85.72	<u>1.75</u>	77.41	3.72	<u>89.55</u>	2.73
U-Net++[5]	86.01	2.57	87.55	86.41	<u>79.55</u>	**88.61**	1.98	83.58	<u>1.91</u>	85.84	3.81
Attention UNet[4]	85.62	3.46	88.33	85.87	78.67	87.29	2.98	**84.89**	2.78	84.67	4.62
nn-UNet[7]	84.15	2.55	**90.55**	83.75	76.61	86.06	**1.51**	78.60	4.44	87.80	**1.70**
TransUNet[9]	<u>86.56</u>	**2.28**	84.63	83.42	76.77	87.15	2.97	83.34	**1.89**	<u>89.20</u>	<u>1.97</u>
Swin-UNet[12]	85.78	<u>2.47</u>	86.56	<u>86.83</u>	78.54	<u>88.54</u>	2.08	81.82	2.03	86.97	3.31
VMA-UNet	**87.71**	2.80	87.64	**89.37**	**80.84**	88.37	2.70	<u>84.10</u>	2.51	**90.67**	3.34

图 4 展示了 ACDC 数据集中不同方法的分割表现．简单场景（如场景（b））中，各方法均表现良好；而在复杂场景（如场景（a）和（c））中，VMA-UNet 表现尤为出色．例如，场景（a）中，其他方法（如 U-

Net++)在边界区域存在欠分割问题,而VMA-UNet能够精确分割;场景(c)中,VMA-UNet在左心室位置的分割效果显著优于其他方法,其他方法甚至不能分割出左心室轮廓.综合分析表明,VMA-UNet在全局信息捕获与细节建模方面具备显著优势,为医学图像分割树立了新标杆.

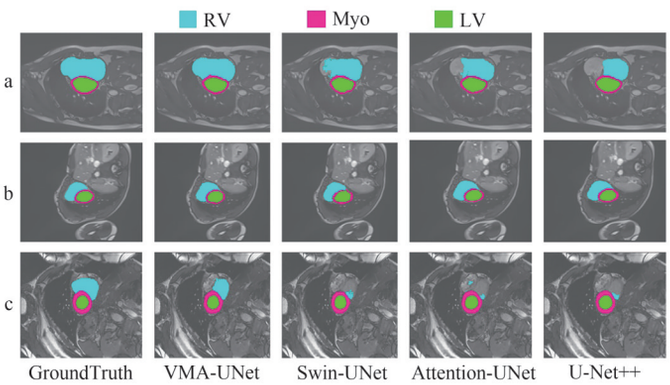

图4 ACDC数据集分割性能的定性比较

Fig. 4 Qualitative comparison of ACDC data set segmentation performance

图5展示了在ACDC数据集上的训练收敛情况.与Swin-UNet相比,VMA-UNet表现出了更为出色的训练损失下降趋势,其损失迅速减少并在训练过程中达到了较低且稳定的值.尽管Attention-UNet和U-Net++损失下降速度较快,但它们最终可学到的特征较为有限.相反,VMA-UNet能够有效适应并捕捉训练数据中的关键信息,从而在医学图像分割任务中展现了更为优异的性能和优势.

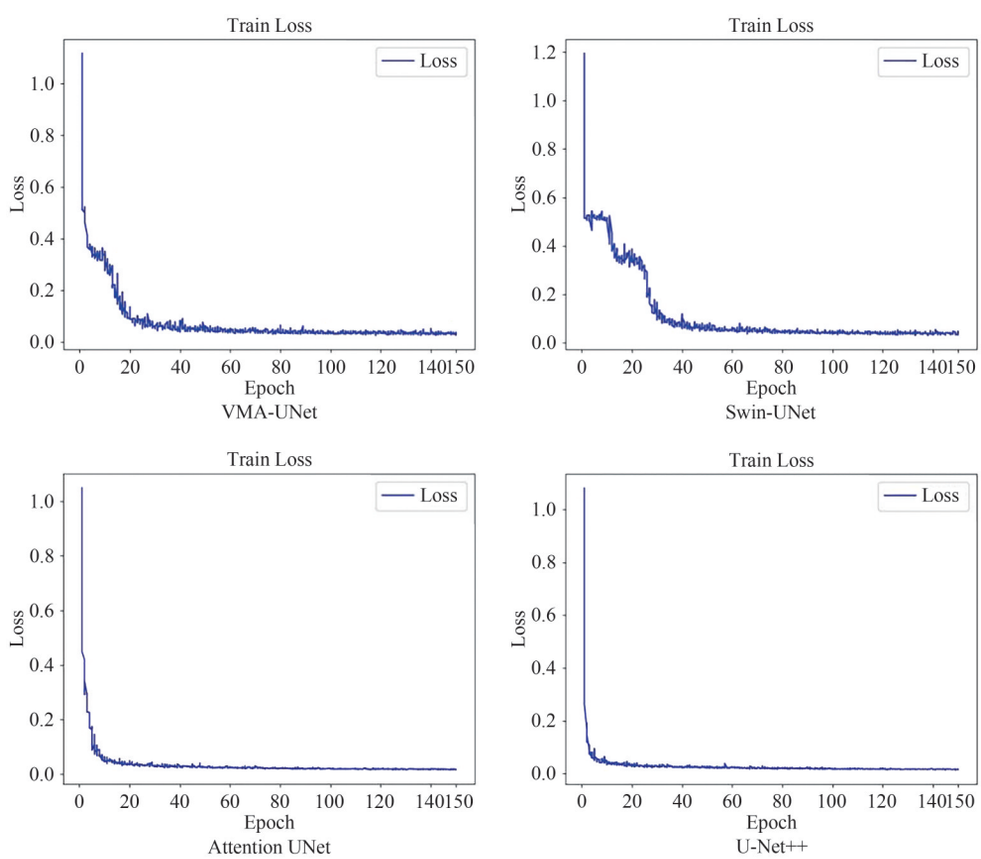

图5 ACDC数据集训练损失收敛

Fig. 5 Training loss convergence diagram of ACDC data set

2.4.2 COVID19 CT-Seg数据集上的实验结果

表2展示了COVID-19 CT-Seg数据集上的定量实验结果.与其他方法相比,VMA-UNet在DSC(77.52%)、HD95(38.03 mm)、Recall(75.98%)和IoU(67.72%)四项指标上均取得了最佳表现.其中,DSC、Recall和IoU分别较次优方法提升了1.03%、0.87%和0.54%;

同时，HD95 减少了 3.39 mm，显示出在捕捉复杂感染区域边界时的显著优势．较高的 IoU 值进一步表明，VMA-UNet 在实际临床应用中具有更高的可靠性和适用性．

表 2　COVID19 CT-Seg 数据集分割性能的定量比较，最优结果用粗体表示，次优结果用下划线表示

Tab. 2　Quantitative comparison of the segmentation performance of the COVID19 CT-Seg dataset. optimal results are shown in bold and sub-optimal results are underlined

Methods	Average					Infection		Lung	
	DSC ↑	HD95 ↓	Precision ↑	Recall ↑	IoU ↑	DSC ↑	HD95 ↓	DSC ↑	HD95 ↓
V-Net[24]	66.18	73.90	74.45	<u>75.11</u>	55.65	52.50	74.13	79.86	73.67
U-Net[3]	72.84	50.69	80.52	75.04	60.67	<u>63.35</u>	**48.65**	82.32	52.72
U-Net++[5]	76.26	45.64	**87.97**	72.77	66.19	61.07	75.70	**91.45**	<u>15.58</u>
Attention UNet[4]	71.15	47.73	<u>84.48</u>	68.11	61.21	51.90	76.32	90.40	19.14
nn-UNet[7]	70.53	56.48	74.08	73.45	57.83	60.76	<u>58.72</u>	82.56	56.53
TransUNet[9]	<u>76.49</u>	<u>41.42</u>	81.98	66.16	57.92	61.91	67.23	<u>91.08</u>	16.00
Swin-UNet[12]	75.35	59.19	82.81	73.10	<u>67.18</u>	60.55	70.72	90.17	47.66
VMA-UNet	**77.52**	**38.03**	82.02	**75.98**	**67.72**	**64.15**	61.07	90.88	**14.99**

2.4.3　Synapse 数据集上的实验结果

表 3 展示了在 Synapse 多器官分割数据集上的定量结果．VMA-UNet 在平均 DSC 及各个器官的 DSC 上均实现了最佳表现，平均 DSC 达 80.81%，相较其他方法显著提升．尤其在 Kidney（L）、Liver、Pancreas 和 Stomach 的分割任务中，VMA-UNet 分别达到了 85.68%、94.70%、63.06% 和 80.95%，取得了优异成绩，充分展示了其在较大器官分割中的高鲁棒性与精确度．

表 3　Synapse 数据集上的定量分割结果，最优结果用粗体表示

Tab. 3　Quantitative segmentation results on the Synapse dataset，the optimal results are shown in bold

Methods	DSC ↑	HD ↓	Aorta	Gallbladder	Kidney（L）	Kidney（R）	Liver	Pancreas	Spleen	Stomach
V-Net[24]	68.81	-	75.34	51.87	77.10	**80.75**	87.84	40.05	80.56	56.98
DARR[25]	69.77	-	74.74	53.77	72.31	73.24	94.08	54.18	89.90	45.96
U-Net[3]	76.85	39.70	89.07	**69.72**	77.77	68.60	93.43	53.93	86.67	75.58
Attention-UNet[4]	77.77	36.02	**89.55**	68.88	77.98	71.11	93.57	58.04	87.30	75.75
R50 U-UNet[9]	74.68	36.87	87.74	63.66	80.60	78.19	93.74	56.90	85.87	75.58
R50 Att-UNet[9]	75.57	36.97	55.92	63.91	79.20	72.71	93.56	49.37	87.19	74.95
R50 ViT[9]	71.29	32.87	73.73	55.13	75.80	72.20	91.51	45.99	81.99	73.95
TransUNet[9]	77.48	31.69	87.23	65.13	81.87	77.02	94.08	55.86	85.08	75.62
Swin-UNet[12]	79.13	**21.55**	85.47	66.53	83.28	79.61	94.29	56.58	**90.66**	76.60
VMA-UNet	**80.81**	25.25	87.48	68.50	**85.68**	79.54	**94.70**	**63..06**	86.55	**80.95**

2.5　消融实验

为了验证 VMA-UNet 中各个模块的有效性，我们在 ACDC 和 COVID-19 CT-Seg 数据集上进行了实验，并给出网络的参数量，采用 DSC 作为性能评估指标．实验结果如表 4 所示．从参数量的角度来看，VMamba 模型的参数量最小，但其 DSC 值低于其他模型．引入 ASPP 模块后，虽然参数量有所增加，但 DSC 显著提升，且参数量仍小于其他模型，表现仅次于只使用 VMamba 的情况．

在未引入 ASPP 模块的情况下，Mamba 模型在 ACDC 和 COVID-19 CT 数据集上的平均 DSC 分别为 86.31% 和 74.43%．而当加入 ASPP 模块后，模型的分割性能显著提升，平均 DSC 分别提高至 87.71% 和 77.52%．这表明，ASPP 模块通过采用不同扩张率的卷积操作，能够有效捕捉多尺度上下文信息，从而

表 4　ACDC 与 COVID-19 CT-Seg 数据集消融实验

Tab. 4　ACDC and COVID-19 CT-Seg data set ablation experiment

DATASET	Methods	Params/（M）	Flops	DSC ↑
				Average
ACDC	TransUNet	105.28	25.38	86.56
	Swin-UNet	27.16	5.92	85.78
	VMamba	**19.12**	**3.51**	86.31
	TransUNet+ASPP	111.97	26.25	85.88
	Swin-UNet+ASPP	33.84	6.24	<u>86.93</u>
	VMA-UNet	25.81	3.83	**87.71**
COVID-19 CT	TransUNet	105.28	25.38	76.49
	Swin-UNet	27.16	5.92	75.35
	VMamba	**19.12**	**3.51**	74.43
	TransUNet+ASPP	111.97	26.25	75.79
	Swin-UNet+ASPP	33.84	6.24	<u>76.81</u>
	VMA-UNet	25.81	3.83	**77.52**

显著增强模型在复杂感染区域的分割能力.

此外,当保留ASPP模块并将VMamba替换为TransUNet或SwinUNet时,结果显示,TransUNet与ASPP的组合在ACDC和COVID-19 CT数据集上的DSC都有所下降,而SwinUNet与ASPP的组合虽然在两个数据集上都实现了DSC提升,但参数量也有所增加,且不如VMamba与ASPP组合的表现.这些结果进一步突出展示了Mamba模型在保持较低参数量的同时,仍具备较强的分割性能,强调了其线性复杂度和可行性.

3 讨论与总结

3.1 讨论

本文提出的VMA-UNet模型通过引入VSS块和ASPP模块,有效克服了传统CNN在捕捉远程信息时的局限性,同时降低了计算成本.实验结果表明,VMA-UNet在ACDC、COVID-19 CT-Seg和Synapse数据集上展现了出色的多尺度建模能力和全局信息捕获性能.首先,VSS块通过线性计算复杂度实现高效的全局建模,相比于基于Transformer的模型(如Swin-UNet[12]),在减少计算成本的同时保持了优异的分割性能.其次,ASPP模块增强了模型的多尺度信息处理能力,能够有效捕捉医学图像中的关键特征,尤其在复杂的医学图像分割任务(如心脏和多器官分割)中显著提升了分割精度.然而,本研究也存在一些局限性.首先,模型性能在很大程度上依赖于预训练权重的初始化,未来可探讨专为医学图像分割任务设计的端到端预训练方法,以进一步提升性能.此外,边缘区域的分割精度仍有提升空间,尤其在ACDC数据集中,HD95指标较高,未来需加强对边缘区域的处理.最后,本研究仅限于单模态医学图像,未来计划扩展至多模态医学图像分割任务,并探索其在3D医学图像分割中的潜力.

3.2 结论

本文提出了基于VMamba和ASPP模块融合的VMA-UNet模型,能够高效捕捉医学图像中的远程依赖信息,并在多尺度上实现全局上下文建模.通过在ACDC、COVID-19 CT-Seg和Synapse数据集上的实验验证,VMA-UNet在分割性能和计算效率上均优于传统的CNN和基于Transformer的方法.未来研究将致力于优化模型在多模态医学图像中的表现,并进一步探索如何在降低计算复杂度的同时提升分割精度的方法.

参 考 文 献

[1] SALPEA N, TZOUVELI P, KOLLIAS D. Medical image segmentation: A review of modern architectures [M]// Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2023.

[2] LI J, CHEN J, TANG Y, et al. Transforming medical imaging with Transformers? A comparative review of key properties, current progresses, and future perspectives[J]. Medical Image Analysis, 2023, 85: 102762.

[3] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[M]// Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015.

[4] OKTAY O, SCHLEMPER J, LE FOLGOC L, et al. Attention U-net: Learning where to look for the pancreas[EB/OL]. 2018: 1804.03999.https://arxiv.org/abs/1804.03999v3

[5] ZHOU Z, RAHMAM SIDDIQUEE M M, TAJBAKHSH N, et al. Unet++: A nested u-net architecture for medical image segmentation[C]//4th International Workshop, DLMIA 2018, Granada: Springer International Publishing, 2018: 3-11.

[6] HUANG H, LIN L, TONG R, et al. UNet 3: A full-scale connected UNet for medical image segmentation[C]//2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona: IEEE, 2020: 1055-1059.

[7] ISENSEE F, JAEGER P F, KOHL S A A, et al. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation [J]. Nature Methods, 2021, 18(2): 203-211.

[8] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach: Curran Associates Inc. 2017:6000-6010.

[9] CHEN J, MEI J, LI X, et al. TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers[J]. Medical Image Analysis, 2024, 97: 103280.

[10] HATAMIZADEH A, TANG Y, NATH V, et al. UNETR: Transformers for 3D medical image segmentation[C]// 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE, 2022: 1748-1758.

[11] ZHOU H Y, GUO J, ZHANG Y, et al. nnFormer: Volumetric medical image segmentation *via* a 3D transformer[J]. IEEE Transactions on Image Processing,

2023，32：4036-4045.

[12] CAO H，WANG Y，CHEN J，et al. Swin-unet：Unet-like pure transformer for medical image segmentation[M]// Lecture Notes in Computer Science. Cham：Springer Nature Switzerland，2023：.

[13] LIU Z，LIN Y，CAO Y，et al. Swin transformer：Hierarchical vision transformer using shifted windows[C]// 2021 IEEE/CVF International Conference on Computer Vision（ICCV）. Montreal：IEEE，2021：9992-10002.

[14] XING Z，YE T，YANG Y，et al. SegMamba：Long-range sequential modeling mamba for 3D medical image segmentation[M]//Lecture Notes in Computer Science. Cham：Springer Nature Switzerland，2024.

[15] SZEGEDY C，LIU W，JIA Y，et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition（CVPR）. Boston：IEEE，2015：1-9.

[16] SZEGEDY C，VANHOUCKE V，IOFFE S，et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition（CVPR）. Las Vegas：IEEE，2016：2818-2826.

[17] CHEN L，BENTLEY P，MORI K，et al. DRINet for medical image segmentation[J]. IEEE Transactions on Medical Imaging，2018，37(11)：2453-2462.

[18] GU Z，CHENG J，FU H，et al. CE-net：Context encoder network for 2D medical image segmentation[J]. IEEE Transactions on Medical Imaging，2019，38(10)：2281-2292.

[19] IBTEHAZ N，RAHMAN M S. MultiResUNet：Rethinking the U-Net architecture for multimodal biomedical image segmentation[J]. Neural Networks，2020，121：74-87.

[20] CHEN L C，PAPANDREOU G，KOKKINOS I，et al. DeepLab：Semantic image segmentation with deep convolutional nets，atrous convolution，and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2018，40(4)：834-848.

[21] ELFWING S，UCHIBE E，DOYA K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning[J]. Neural Networks，2018，107：3-11.

[22] LIU J，YANG H，ZHOU H Y，et al. Swin-UMamba：Mamba-based UNet with ImageNet-based pretraining[M]// Lecture Notes in Computer Science. Cham：Springer Nature Switzerland，2024.

[23] LIU Y，TIAN Y，ZHAO Y，et al. VMamba：Visual state space model[EB/OL]. 2024：2401.10166. https://arxiv.org/abs/2401.10166v3

[24] 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（3DV）. Stanford：IEEE，2016：565-571.

[25] FU S，LU Y，WANG Y，et al. Domain adaptive relational reasoning for 3D multi-organ segmentation[M]//Lecture Notes in Computer Science. Cham：Springer International Publishing，2020.

（责编&校对　雷建云）