ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 40,No. 4,Jul. 2025,pp. 887－900
DOI:10. 16337/j. 1004-9037. 2025. 04. 004

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

面向多模态心脏影像的多分支协同分割模型

肖　瑞，邵　伟

（南京航空航天大学人工智能学院，南京 211106）

摘　要：精确的心脏结构分割对于心脏血管疾病辅助诊断和术前的准确评估有着重要的意义。不同模态的影像之间在空间分布和语义表达上存在显著差异，但现有方法多采用单分支网络结构，难以充分融合多模态信息，在多模态任务上缺乏泛化能力。针对这一问题，提出一种融合状态空间模型Mamba与卷积模型的多分支协同分割网络MCNet（Multi-modal collaborative network）。该网络主要由3个模块构成：基于Mamba与卷积神经网络的双分支特征提取器、动态特征融合模块以及Mamba解码器。特征提取器的双分支分别侧重于提取全局语义与局部细节特征，动态特征融合模块根据图像动态调整多种融合路径的权重，从而实现不同分支的动态特征整合。本文提出的方法在心脏的MRI数据集ACDC与超声数据集CAMUS上进行了充分实验。实验结果表明，本文方法通过基于混合专家（Mixture of experts，MoE）机制的动态特征融合模块，动态调整Mamba全局特征和CNN局部特征的融合权重，在边界清晰的ACDC数据集中，平均Dice和交并比IoU分别达到0.845和0.779，在边界模糊的CAMUS数据集中的平均Dice和IoU分别达到0.883和0.796，均优于目前主流方法。同时，消融实验进一步验证了每个模块的有效性。MCNet通过MoE机制实时调整全局和局部特征的融合权重，在保证全局感知的同时提升了结构细节完整性，为多模态心脏影像分割提供了高效而鲁棒的解决方案。

关键词：医学影像分割；多模态医疗影像；心脏结构分割；Mamba；动态特征融合；多分支协同分割
中图分类号：TP18;R445　　**文献标志码**：A

Multi-branch Collaborative Segmentation Model for Multi-modal Cardiac Imaging

XIAO Rui，SHAO Wei

（College of Artificial Intelligence，Nanjing University of Aeronautics & Astronautics，Nanjing 211106，China）

Abstract：Precise structural segmentation of the heart is important for the adjunctive diagnosis of cardiovascular disease and accurate preoperative evaluation. There are significant differences between images of different modalities in terms of spatial distribution and semantic expression，but existing methods mostly use single-branch network structures，which are unable to fully integrate multi-modal information and lack generalization capabilities in multi-modal tasks. To address this problem，this paper proposes a multi-branch collaborative segmentation network，i.e. multi-modal collaborative network（MCNet），which fuses the state space model Mamba with the convolutional model. The network is mainly composed of three modules：A dual-branch feature extractor based on Mamba and convolutional neural networks，a dynamic feature fusion module，and a Mamba decoder. The dual branches of the feature extractor focus on extracting global semantic and local detail features，respectively，and the dynamic feature fusion module

收稿日期：2025-06-22；**修订日期**：2025-07-20

dynamically adjusts the weights of multiple fusion paths according to the image, thus realizing dynamic feature integration in different branches. The proposed method is fully experimented on the MRI dataset ACDC of the heart and the ultrasound dataset CAMUS. Experimental results show that the proposed method, through a dynamic feature fusion module based on the mixture of experts (MoE) mechanism, dynamically adjusts the fusion weights of Mamba global features and CNN local features. In the ACDC dataset with clear boundaries, the average Dice and intersection over union (IoU) values reach 0.845 and 0.779, respectively. In the CAMUS dataset with blurred boundaries, the average Dice and IoU values reach 0.883 and 0.796, respectively, both of which outperform current mainstream methods. Additionally, ablation experiments further validate the effectiveness of each module. MCNet uses the MoE mechanism to dynamically adjust the fusion weights between global and local features in real time, enhancing structural detail integrity while maintaining global perception, thereby providing an efficient and robust solution for multi-modal cardiac image segmentation.

Key words: medical imaging segmentation; multi-modal medical imaging; segmentation of the heart structure; Mamba; dynamic feature fusion; multi-branch collaborative segmentation

引　　言

心脏在维持全身血液循环和保障供血方面起着关键作用。心脏血管疾病一直是全球范围内的重要公共健康问题[1],它不仅严重威胁个人的生命和健康,而且给医疗保健系统和社会经济发展带来沉重负担。准确的诊断是实施有效治疗的基础,因此心脏疾病的早期筛查和准确评估已逐渐成为临床医学和基础研究的重点方向。影像学[2]在心脏结构和功能评估中的应用,为心脏疾病的诊断和病理机制的确定提供了重要支持,有助于早期发现心肌损伤、瓣膜异常及血流动力学障碍等关键病变,显著提高诊断效率和治疗时效。

医学影像为准确识别和全面评估心脏疾病提供了重要的数据基础。磁共振成像(Magnetic resonance imaging, MRI)[3],作为一种无创且高分辨率的成像方法,在心脏结构成像方面具有独特的优势。通过结合静态和动态成像序列,心脏磁共振成像不仅能清晰呈现心肌结构、心腔形态和血管走向,还能动态评估收缩和舒张功能、心肌灌注状态、组织纤维化和瘢痕形成及其他病理过程。T1加权序列有助于分析心肌组织成分和脂肪浸润,而T2加权序列对水肿和炎症等病理变化更为敏感;晚期钆增强(Late gadolinium enhancement, LGE)技术可有效识别心肌梗死和纤维化区域。这些多序列成像技术的共同应用,使得心脏核磁共振成像在冠心病、心肌病、心肌炎等多种心脏疾病的诊断、病情评估和疗效监测中发挥着越来越重要的作用,并为实现个体化治疗提供了坚实的成像基础。

超声心动图(Echocardiography)作为临床应用最为广泛的心脏影像技术[4],具有无创、实时、方便和低成本等特点,被广泛用于评估心腔大小、心肌厚度、瓣膜结构和功能以及心脏收缩和舒张功能。传统的二维超声可观察心脏解剖结构和瓣膜运动,而M型超声适用于定量测量心腔尺寸和心壁厚度,彩色多普勒成像可用于分析血流方向和速度,确定是否存在瓣膜反流或狭窄等异常。虽然超声成像在图像分辨率方面不如磁共振成像,但其便携性和动态实时性的特点使其在临床实践中仍然不可替代。然而,通过人工方式对磁共振图像和超声进行诊断既主观又耗时。

针对人工的不足,现有的研究提出了很多基于深度学习的超声影像和MRI影像的分割方法,辅助医生诊断。卷积神经网络(Convolutional neural network, CNN)作为最早广泛应用于医学图像分析的深度学习模型,能够有效提取局部空间特征,并在图像分割任务中展现出良好的性能。代表性结构如

U-Net[5]及其变体,在多种心脏影像分割任务中被广泛采用,能够实现精确的边界定位和结构识别。然而,CNN在建模长距离依赖关系方面存在一定局限,难以全面捕捉全局上下文信息。

为了克服卷积神经网络在全局信息感知方面的不足,近年来推出了基于注意力机制的Transformer架构。Transformer最初在自然语言处理领域取得了重大突破,而后也逐渐被应用到医学图像分割任务中。有别于卷积神经网络只能关注邻近区域,Transformer通过注意力机制能够捕捉图像中两个相隔很远区域的依赖关系,从而提高复杂结构识别的准确性。典型方法如TransUNet[6]和Swin-UNet[7]将Transformer模块与CNN进行融合,在保持细节感知能力的同时增强了对全局的感知能力,显著提升了分割效果。虽然Transformer具有出色的建模能力,但其自注意力机制在图像上具有二次方的计算复杂度,导致单张MRI图像推理时间增加2～3倍。为此,研究者近期提出了一种新型神经网络架构Mamba[8],其设计灵感来源于状态空间模型(State space model, SSM),它具备更高效的序列建模能力。相较于Transformer,Mamba被认为是一种兼具效率和表现力的替代方案,它通过线性递归操作捕捉长程依赖关系,大幅降低了计算资源需求,并且通过SSM减少了模型参数,降低了内存占用。在医学图像分割任务中,Mamba结构表现出良好的性能和可扩展性,为处理心脏磁共振成像和超声图像等医学数据提供了新的解决思路。但是,现有的Mamba-based模型,如VM-UNet[9]仍然存在局部细节丢失的问题,在对心肌这类对边界细节要求高的分割任务中,比传统CNN方法表现更差。

因此,如何兼顾全局特征与局部细节特征,仍是当前研究的挑战之一。本文提出了一种基于混合专家(Mixture of experts, MoE)机制的融合框架MCNet(Multi-modal collaborative network),通过专家选择机制自适应地融合Mamba和CNN分支的输出贡献,使模型能够根据输入图像的特征差异,动态调整多种融合路径的权重,得到最优的融合特征。这种结构不仅提高了模型的表达能力和泛化性能,而且为高质量分割超声和磁共振心脏成像提供了更准确、更高效的解决方案。

本文的主要贡献如下:

(1)提出一种融合Mamba与CNN的双分支架构,并引入MoE机制根据输入图像内容实时调整特征融合权重,从而实现动态特征融合。该方法充分结合了Mamba优越的长程依赖建模能力与CNN对局部空间细节的高效提取能力,通过MoE机制根据输入图像特征自适应分配多融合路径的权重输出,从而提升模型在不同成像模态下的泛化性能。

(2)在多心脏MRI与超声影像数据集上进行了充分的实验验证,包括与现有主流分割方法的对比实验以及对模型关键模块的消融实验。实验结果表明,本文所提出的方法在MRI模态下较跨模态迁移模型(如Swin-Unet)Dice系数提升约3.1%,在准确率、泛化性和模型稳定性方面均优于现有方法。

1　相关工作

1.1　卷积神经网络

CNN在图像处理中具有强大的特征提取能力,是医学图像分析中应用最广泛的深度学习架构之一。CNN通过局部感受野、权重共享和池化操作,有效捕捉图像中的空间结构信息,可以自动学习多层次、多尺度的图像特征,能自动学习多层次、多尺度的图像特征,大大提高了医学图像的分类、检测和分割性能。

经典的U-Net结构由Ronneberger等[5]于2015年提出,其编码器-解码器对称设计及跳跃连接机制有效融合了低层空间细节与高层语义信息,极大推动了医学图像分割的发展。随后,许多基于U-Net的变体被提出,如Attention U-Net[10]引入注意力机制强化重要特征的表达,3D U-Net[11]扩展至三维医学影像数据处理,进一步提升了分割的精度和适用范围。

除此之外,深度残差网络(ResNet)[12]、密集连接网络(DenseNet)[13]等也被广泛应用于医学图像领

域,它通过更深入的特征学习和梯度转移优化,增强了模型的性能和训练稳定性。此外,卷积神经网络在超声波图像和磁共振心脏图像的分割中均表现出较高的准确性和鲁棒性,成为临床辅助诊断的重要工具。

1.2 Mamba神经网络

随着大规模序列建模任务对效率和表现力的要求越来越高,状态空间模型(SSM)在视觉领域的应用引起了广泛关注。Mamba是一种基于选择性SSM(Selective SSM)的神经网络架构,由Gu等[8]于2023年提出,其核心思想是利用结构化状态空间模型实现长序列依赖关系的线性时间建模,与传统的变换器相比,Mamba在保持建模深度的同时,大大降低了计算复杂度,提高了推理效率。

Mamba的主要优势在于它以卷积方式实现了递归更新,在GPU上实现了并行计算,并突破了以往基于循环或自注意机制的长距离建模的瓶颈。在一些自然语言和视觉任务中,Mamba的表现优于Transformer和RNN。最近,一些研究也尝试将Mamba应用于医学图像分析任务,如图像分类和器官分割,但其描绘局部细节的能力仍不如CNN,尤其是在边界清晰度和纹理保存方面,因此,将Mamba与具有较强局部感知能力的模型融合成为提高其在医学图像任务中性能的一个重要方向。

1.3 混合专家融合机制

MoE机制是一类引入稀疏性和模块选择策略的神经网络结构,最早由Jacobs等[14]提出,后经Google的GShard[15]、Switch Transformer[16]等模型发展成为大规模模型训练的重要架构。MoE的基本思想是在每次前向传播过程中,仅激活一小部分“专家”子网络,并通过门控机制根据输入特征决定专家的权重组合,从而在保证模型表达能力的同时显著减少计算成本。

近年来,MoE机制在自然语言处理和计算机视觉任务中得到了广泛应用,并逐渐扩展到医学图像分析领域。在图像分割任务中,MoE可以有效整合不同类型的特征提取分支(如CNN、Transformer、SSM等),实现结构层面的优势互补,同时,通过动态选择机制,它可以针对不同的解剖结构或图像模式自适应地调整信息融合方法,从而提高模型的适应性和鲁棒性。在多模态图像(如核磁共振和超声)或具有复杂异质组织结构的心脏图像的分割中,MoE提供了灵活高效的融合框架。

2 本文方法

如图1所示,本文所提出的MCNet模型主要由3部分组成,分别是Mamba和卷积神经网络两分支特征提取器、MoE动态选择器和Mamba解码器。两分支特征提取器负责提取心脏影像的特征,两个特征分别专注于全局信息和局部信息,然后通过MoE对特征融合模块进行动态选择,得到各融合模块的权重比例。通过这个比例计算得到融合特征,然后由Mamba解码器得到该图像的掩膜。通过动态选择特征融合的方式,最终实现多模态心脏图像的分割。

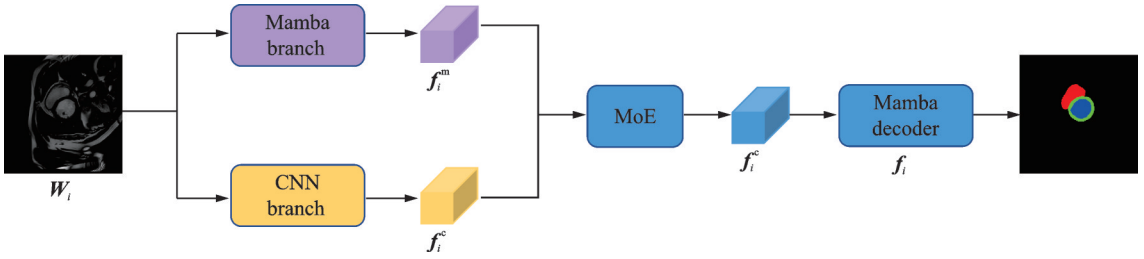


图1 MCNet的网络结构

Fig.1 Network structure of MCNet

2.1　基本定义

本文定义一个包含N个心脏多模态影像$\boldsymbol{X}=\{\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_N\}$的多类别分割任务,其中$\boldsymbol{Y}_i$为$\boldsymbol{W}_i$对应的类别标签。对于$\boldsymbol{W}_i$,本文分别利用Mamba和卷积神经网络双分支架构提取图像的特征,记为$\boldsymbol{f}_i^{\mathrm{m}}=\{\boldsymbol{f}_i^{\mathrm{m},1},\boldsymbol{f}_i^{\mathrm{m},2},\cdots,\boldsymbol{f}_i^{\mathrm{m},K}\}$和$\boldsymbol{f}_i^{\mathrm{c}}=\{\boldsymbol{f}_i^{\mathrm{c},1},\boldsymbol{f}_i^{\mathrm{c},2},\cdots,\boldsymbol{f}_i^{\mathrm{c},K}\}$。随后,再利用MoE机制对特征进行动态选择。最后利用Mamba重建出掩膜M_i。

2.2　两分支特征提取器

在MCNet中,特征提取器由两个并行的分支构成:Mamba分支和卷积神经网络分支。该设计旨在充分利用这两个分支在全局和局部感知方面的互补优势:

(1)Mamba分支:该分支采用Mamba结构作为核心模块,利用状态空间模型框架捕捉长距离依赖关系,有效获取图像中的全局上下文信息。它的结构类似于Transformer中的自我注意力机制,但在计算效率和序列建模方面展现出更优的性能。通过多层堆叠,Mamba分支可以构建对心脏解剖结构具有更强辨别力的全局特征。

(2)CNN分支:该分支基于传统卷积神经网络架构,由多层卷积层和池化层组成,重点提取图像的局部细节特征。由于心脏MRI图像和超声影像中的许多边缘信息和细节结构对分割精度至关重要,卷积分支的局部建模能力能够有效补充Mamba分支在空间细节方面的不足。

双分支分别输出两种不同类型的特征表示,即

$$\boldsymbol{f}_i^{\mathrm{m}}=F_{\mathrm{m}}(\boldsymbol{X}_i) \tag{1}$$

$$\boldsymbol{f}_i^{\mathrm{c}}=F_{\mathrm{c}}(\boldsymbol{X}_i) \tag{2}$$

式中:$\boldsymbol{f}_i^{\mathrm{m}}$和$\boldsymbol{f}_i^{\mathrm{c}}$分别表示两分支特征提取器提取的特征;$F_{\mathrm{m}}$和$F_{\mathrm{c}}$分别表示Mamba特征提取器和卷积特征提取器;$\boldsymbol{X}_i$表示心脏影像或超声影像。

通过该双分支结构,MCNet能够同时关注图像的全局结构与局部细节,为后续分割提供更丰富、更准确的特征支撑。

2.3　MoE动态特征融合模块

为了实现特征层面的动态融合,MCNet引入了一个基于MoE的动态选择模块,根据不同模态心脏影像的内容,对来自Mamba分支和卷积分支的特征进行多模块的自适应加权融合。

2.3.1　路由网络

如图2(a)所示,针对来自双分支的特征$\boldsymbol{f}_i^{\mathrm{m}}$和$\boldsymbol{f}_i^{\mathrm{c}}$,本文采用路由网络(Gating network)[17]将两个特征分别输入到一个模块中。该模块中,经过线性层和均方根归一化层(Rooted mean square layer normalization,RMSNorm)初步压缩和标准化特征,然后经过高斯误差线性单元(Gaussian error linear unit,GELU)引入非线性表达,最后取平均值稳定训练过程。处理完的双分支特征经过拼接,和1个线性层改变形状得到4个特征融合模块的权重$\boldsymbol{L}=\{l_1,l_2,l_3,l_4\}$,可表示为

$$\boldsymbol{L}=\boldsymbol{W}*\mathrm{mean}(\mathrm{GELU}(\mathrm{RMSNorm}(\boldsymbol{W}_1\boldsymbol{f}_i^{\mathrm{m}})))+\boldsymbol{W}*\mathrm{mean}(\mathrm{GELU}(\mathrm{RMSNorm}(\boldsymbol{W}_2\boldsymbol{f}_i^{\mathrm{c}}))) \tag{3}$$

式中:$\boldsymbol{W}$、$\boldsymbol{W}_1$和$\boldsymbol{W}_2$分别为3个线性层的可学习参数,维度分别为$C\times 4$、$C\times C$和$C\times C$,其中C为特征的通道数;mean、GELU和RMSNorm分别表示取平均、高斯误差线性单元和均方根归一化层。

2.3.2　双时相特征聚合模块

如图2(b)所示,Mamba和卷积神经网络分别提取出全局信息和局部细节信息。针对这种感受野差距大的特征,本文采用了双时相特征聚合模块(Bitemporal feature aggregation module,BFAM)[18]。

BFAM首先通过4个不同膨胀率的卷积,聚合低级纹理信息和全局信息,使得特征更加的丰富。对

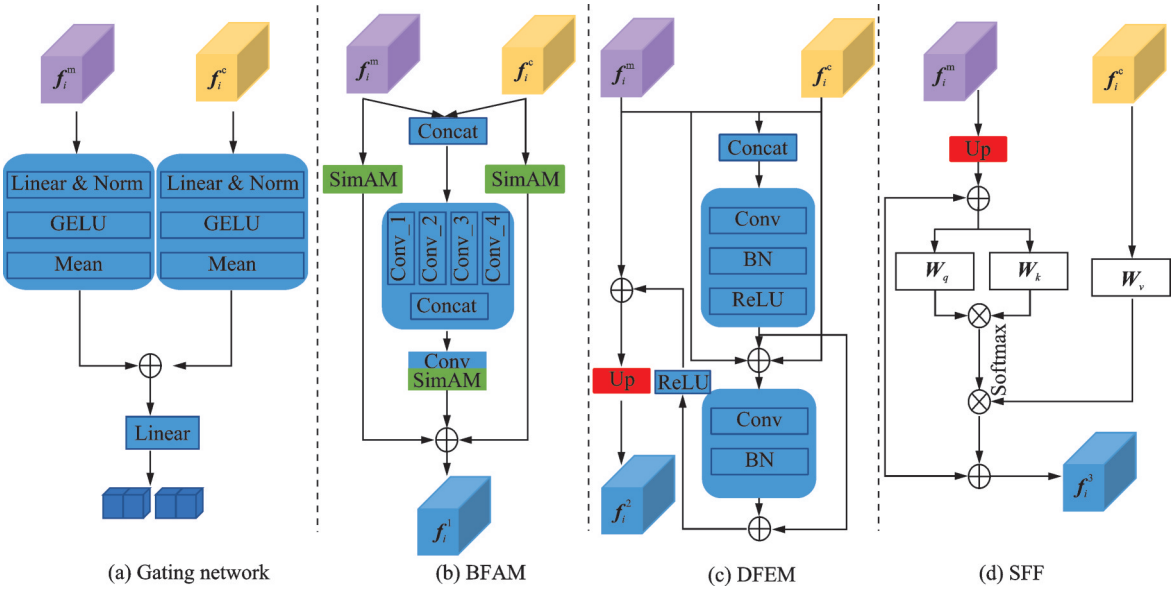


图2 MoE模块的路由模块和特征融合模块

Fig.2 The routing module and feature fusion module of the MoE module

于输入的双分支特征，BFAM将两个特征进行拼接，然后使用不同膨胀率的卷积提取多尺度特征，然后通过卷积进行降维，再使用SimAM(Simple attention module)注意力机制[19]强化特征，即

$$f_{3\times3}^{d}=\mathrm{Conv}_{3\times3}^{d}(\mathrm{Concat}(f_i^{\mathrm{m}},f_i^{\mathrm{c}})) \quad d=\{1,2,3,4\} \tag{4}$$

$$f_{\mathrm{cat}}=\mathrm{SimAM}(\mathrm{Conv}_{1\times1}(\mathrm{Concat}(f_{3\times3}^{1},f_{3\times3}^{2},f_{3\times3}^{3},f_{3\times3}^{4}))) \tag{5}$$

式中：d表示膨胀率，Conv表示卷积层，Concat表示通道连接层，SimAM表示SimAM注意力层。通过计算邻近特征间的差异，有助于提升特征的区分度，提升心脏结构边界分割精度。

然后使用SimAM注意力机制提取双分支特征的更精确的局部细节，再利用刚刚提取的f_{cat}分别相乘，得到$f_i^{\mathrm{m}'}$、$f_i^{\mathrm{c}'}$，即

$$f_i^{\mathrm{m}'}=\mathrm{SimAM}(f_i^{\mathrm{m}})\times f_{\mathrm{cat}} \tag{6}$$

$$f_i^{\mathrm{c}'}=\mathrm{SimAM}(f_i^{\mathrm{c}})\times f_{\mathrm{cat}} \tag{7}$$

最后，将得到的所有特征利用SimAM注意力层和卷积层聚合在一起，即

$$f_i^{1}=\mathrm{SimAM}(\mathrm{Conv}_{3\times3}(f_{\mathrm{cat}}+f_i^{\mathrm{m}'}+f_i^{\mathrm{c}'})) \tag{8}$$

通过使用BFAM来整合不同感受野的特征，整合的目的是通过合并两个分支的特征来增强变化区域内的空间信息。

2.3.3 深度特征提取模块

如图2(c)所示，为了进一步增强特征融合后的表征能力，本文设计了深度特征提取模块(Deep feature extraction module, DFEM)，用于在保留原始特征信息的同时提升高层语义特征的表达效果。DFEM主要对融合后的特征进行进一步处理，提取更加深层次的结构信息，为最终的分割解码提供支持。

首先将两分支特征在通道维度上进行拼接，并通过逐元素求和操作实现初步融合。在融合后的特征上引入1×1卷积以压缩通道维度，从而在减小计算复杂度的同时完成特征的非线性变换与重组。为避免压缩过程造成信息损失，DFEM引入了残差拼接策略，将原始输入特征与压缩后的特征相连接，以保留低层细节信息并增强特征的完整性。

随后,DFEM利用3×3卷积进一步提取融合特征中的深层语义信息。同时,考虑到Mamba中包含的全局信息对分割的准确性非常重要,DFEM还对该模块的原始输出特征和提取的深层语义特征进行了逐元素求和,加强了两分支特征之间的协同表达。最后,通过SimAM注意力机制对融合后的特征进行加权和求和,从空间和通道两个维度突出显著区域,进一步增强对关键特征的聚焦能力和不同心脏结构的判别能力。

$$\boldsymbol{f}_i^{\mathrm{cat}}=\mathrm{Conv}_{1\times1}(\mathrm{Concat}(\boldsymbol{f}_i^{\mathrm{m}},\boldsymbol{f}_i^{\mathrm{c}})+(\boldsymbol{f}_i^{\mathrm{m}}+\boldsymbol{f}_i^{\mathrm{c}})) \tag{9}$$

$$\boldsymbol{f}_i^{\mathrm{r}}=\mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}_{3\times3}(\boldsymbol{f}_i^{\mathrm{cat}}))+\boldsymbol{f}_i^{\mathrm{cat}})+\boldsymbol{f}_i^{\mathrm{m}} \tag{10}$$

$$\boldsymbol{f}_i^{2}=\mathrm{Conv}_{3\times3}(\mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}_{3\times3}(\boldsymbol{f}_i^{\mathrm{r}})))) \tag{11}$$

式中:Conv、ReLU和BN分别表示卷积层、ReLU激活函数和批归一化层。

通过上述步骤,DFEM实现了多源特征的深度整合,为后续分割提供了更加丰富和高质量的特征支撑。

2.3.4　空间特征融合模块

为了有效融合不同尺度的空间信息,增强低层特征的语义表达能力,本文引入了空间特征融合(Spatial feature fusion,SFF)模块[20]。该模块旨在利用高层特征图中丰富的语义信息对低层特征图进行指导,在保留低层空间细节的同时,实现语义增强,从而提高模型在多模态分割任务中的准确性与鲁棒性。

如图2(d)所示,SFF模块以双分支特征作为输入,对特征$\boldsymbol{f}_i^{\mathrm{m}}$进行上采样操作防止两个特征图形状冲突,然后引入注意力机制建模两个特征图的关联性,具体操作如下

$$\boldsymbol{f}_i^{\mathrm{m}'}=\mathrm{Up}(\boldsymbol{f}_i^{\mathrm{m}},\mathrm{size}(\boldsymbol{f}_i^{\mathrm{c}})) \tag{12}$$

$$\boldsymbol{f}_i^{3}=\mathrm{Softmax}((\boldsymbol{W}_q\boldsymbol{f}_i^{\mathrm{m}'}\times\boldsymbol{W}_k\boldsymbol{f}_i^{\mathrm{m}'}))\times(\boldsymbol{W}_v\boldsymbol{f}_i^{\mathrm{c}})+\boldsymbol{f}_i^{\mathrm{m}'} \tag{13}$$

式中:q和k是上采样的Mamba全局信息和CNN的串联,而v是CNN的局部信息。此操作使全局信息能够提供更丰富、更准确的上下文信息,从而指导CNN特征以计算像素的关系并增强其表示语义信息的能力。

第4个Drop模块是选择Mamba的全局信息作为融合后的特征,即

$$\boldsymbol{f}_i^{4}=\boldsymbol{f}_i^{\mathrm{m}} \tag{14}$$

于是,可以得到最终的特征融合结果

$$\boldsymbol{f}_i=\boldsymbol{L}\cdot(\boldsymbol{f}_i^{1},\boldsymbol{f}_i^{2},\boldsymbol{f}_i^{3},\boldsymbol{f}_i^{4}) \tag{15}$$

2.4　损失函数

为监督模型的训练过程,本文选用交叉熵损失(Cross-entropy loss)函数[21]作为主要优化目标函数。该损失函数广泛应用于多分类语义分割任务中,能够有效度量预测概率分布与真实标签分布之间的差异,从而指导模型收敛至更优的判别边界。ACDC数据集验证集的收敛情况如图3所示,可以看到验证集到50个epoch基本收敛,说明本文采用的交叉熵损失函数能有效地指导模型收敛。交叉熵损失函数公式具体为

$$L(\boldsymbol{x},\boldsymbol{y})=-\boldsymbol{y}_i\log_2P(\boldsymbol{x}_i) \tag{16}$$

式中:$\boldsymbol{y}_i$表示真实标签,$P(\boldsymbol{x}_i)$表示预测标签的概率分布。

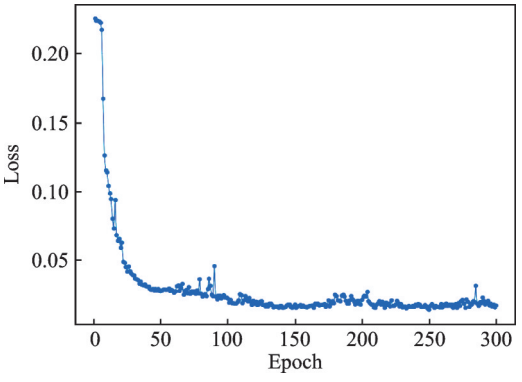


图3　ACDC验证集的损失变化曲线

Fig.3　Loss curve on the ACDC validation set

3 实验分析与结果

3.1 数据集和实验设置

为了验证本文所提方法在心脏图像分割任务中的有效性与通用性，在两个公开心脏医学图像数据集上进行了实验。两个数据集分别为心脏MRI影像数据集ACDC(Automated cardiac diagnosis challenge)[22]和心脏超声影像数据集CAMUS(Cardiac acquisitions for multi-structure ultrasound segmentation)[23]。

ACDC数据集是心脏MRI图像分割领域广泛使用的基准数据集，由MICCAI 2017提供，包含来自5类心脏病病人的短轴心脏MRI序列图像，包括正常(NOR)、心肌肥厚(HCM)、扩张型心肌病(DCM)、心肌梗死后(MINF)及右心室异常(RV)等，每个样本均提供3类完整的结构标注：左心室(LV)、右心室(RV)和心肌(MYO)。

CAMUS数据集是一个公开的心脏超声成像数据集，包含500名患者的四腔室心脏超声图像，涵盖收缩末(ES)和舒张末(ED)关键帧的医生手绘标签。该数据的分割目标为左心室(LV)和左心房(LA)，图像采集过程具有临床多样性高和真实场景干扰的特点。

在数据预处理上，为统一不同模态图像的尺寸和统计分布，本文将所有输入图像统一缩放至256像素×256像素，并在训练时进行归一化处理，以提高模型的鲁棒性和泛化能力。同时，两个数据集分别按原论文推荐的8∶2∶5和8∶1∶1划分训练/验证/测试集。实验在NVIDIA RTX 3090 GPU平台上进行，使用PyTorch框架实现，优化器选择AdamW，初始学习率设为1×10^{-3}，batch size设为16，训练轮次为300。

本文实验主要包括以下3部分：

(1) 多模态心脏分割实验：分别在ACDC(MRI)和CAMUS(超声)数据集上独立训练并验证模型性能，以评估所提方法在不同模态下的适应性与泛化能力。

(2) 对比实验：在ACDC数据集上与主流方法进行对比，包括U-Net、TransUnet、Swin-Unet、VM-UNet[23]和SCUNet++[24]，采用统一训练策略与评价指标，比较不同方法在心脏MRI图像分割中的性能差异。

(3) 消融实验：在两个数据集上分别移除本文使用的4个关键模块——SFF、BFAM、DFEM和Drop，分析各模块对模型性能的具体贡献。

为全面评估模型在磁共振成像和超声两个模态下的心脏影像分割任务中的性能表现，采用3种主流的量化指标，包括Dice系数(Dice similarity coefficient, DSC)、交并比(Intersection over union, IoU)和像素精度(Accuracy, Acc)。这些指标从重叠程度、区域匹配程度和整体预测准确性等3个不同维度综合衡量模型性能，是医学图像分割领域中被广泛接受的评价标准。

3.2 实验结果

3.2.1 多模态心脏分割实验

本实验旨在验证MCNet在差异大的两种影像模态下的鲁棒性和泛化能力。具体来说，本论文的模型在ACDC(心脏核磁共振成像)和CAMUS(心脏超声波)数据集上进行了训练和测试，以评估其在结构边界成像清晰的磁共振成像和结构模糊、伪像复杂的超声图像中的适应性和泛化性能。同时，考虑到两种模态的心脏影像在图像分辨率、对比度和组织结构表达方面的显著差异，该模型必须具备强大的特征提取和融合能力，才能在两个数据集上实现理想的分割性能。

实验结果如表1所示，其中$DSC_{class\,i}$($i=1,2,3$)表示数据集的单类标签的Dice值，比如$DSC_{class\,1}$表示数据集第一类的Dice值，在ACDC数据集中第一类为右心室(RV)，在CAMUS数据集中第一类为左心室(LV)。由表1不难看出，MCNet在磁共振成像和超声两个模态的数据集上均取得了优异的分割性能。在ACDC数据集中，MCNet对左心室(LV)、右心室(RV)和心肌(MYO)这3个解剖结构的所有图

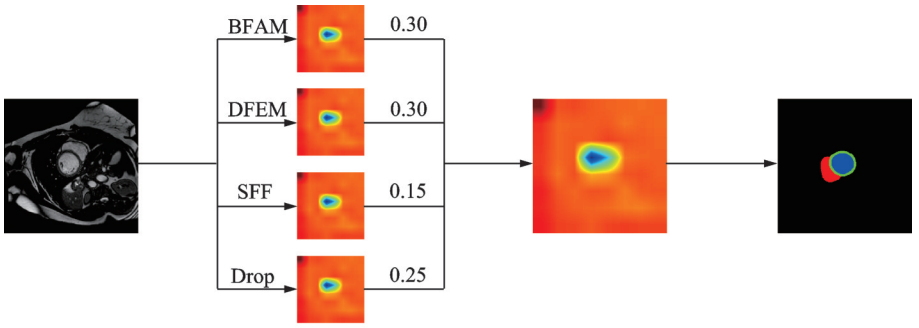
表1　MCNet在ACDC数据集和CAMUS数据集上的分割结果

Table 1　Segmentation results of MCNet on ACDC and CAMUS datasets

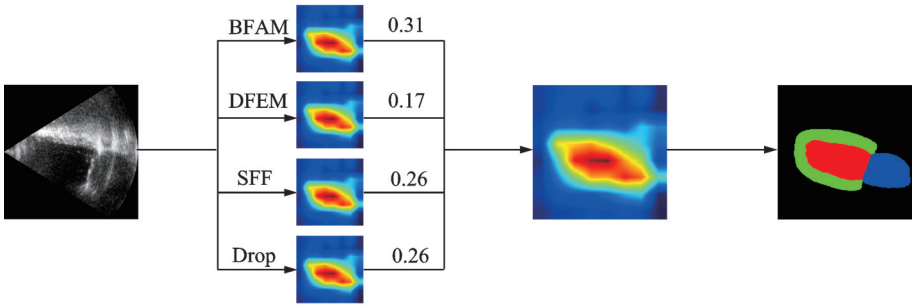
数据集	DSC_{class1}	DSC_{class2}	DSC_{class3}	DSC_{mean}	IoU	ACC
ACDC	0.780 8	0.828 9	0.925 3	0.845 0	0.779 2	0.994 3
CAMUS	0.900 7	0.855 7	0.893 1	0.883 2	0.796 2	0.954 9

像平均Dice系数分别达到0.93、0.78和0.83,展现了模型在成像清晰的模态下,对边界细节和形态结构具有良好的建模能力。在成像模糊、边界不清晰的CAMUS数据集中,MCNet对左心室(LV)和左心房(LA)的Dice分别达到0.90和0.89,由此不难知晓,在噪声干扰较大的超声图像中,模型也能对心脏的不同结构进行定位和分割。

图4展示了MCNet在ACDC(MRI)和CAMUS(超声)两种模态下的分割结果。可以观察到,模型在差异性巨大的两个模态下均能清晰还原心脏结构的边界与形态,表现出良好的分割性能。MCNet通过4个不同的特征融合模块分别对两个分支提取的特征进行融合,并通过路由网络计算出的权重系数(Logits),对4种融合得到的特征的加权整合,最终获得自适应融合特征,并经由解码器重建为分割掩膜。



(a) Weight distribution of the ACDC dataset



(b) Weight distribution of the CAMUS dataset

图4　多分支协同分割网络可视化结果

Fig.4　Visualization results of multi-branch collaborative segmentation network

同时,模型的自适应性能还能从表2中的定量分析结果可见,MCNet在不同模态下对各融合模块的侧重程度存在显著差异,反映出模型在不同影像模态下具有良好的适应性。在表2中,权重分析均在两个模态的测试集下进行,其中DFEM模块在ACDC数据集中所有测试图像的平均权重达到0.31,而在CAMUS数据集中仅为0.17;相反地,SFF模块在ACDC中所有测试图像的平均权重仅为0.15,但在CAMUS中提升至0.26。上述结果表明,MCNet能够根据两种模态的图像的差异动态调整特征的融合策略,从而更有效地提取与分割任务相关的判别性特征,提升在多模态医学影像分割任务中的鲁棒性

表2 不同模态下MCNet各特征融合模块的平均Logits权重分析

Table 2 Analysis of average Logits weights for MCNet feature fusion modules under different modalities

数据集	BFAM	DFEM	SFF	Drop
ACDC	0.303 4	0.307 4	0.146 2	0.243 0
CAMUS	0.314 8	0.169 3	0.262 3	0.253 6

与泛化能力。

综合来看,MCNet依托其Mamba和CNN双分支结构和动态特征融合机制,在应对不同成像模态的结构特征变化方面表现出较强的自适应能力,为多模态医学图像分割提供了稳定有效的解决方案。

3.2.2 对比实验

为了进一步验证所提模型的有效性,本文在心脏磁共振成像数据集ACDC上与几种主流的语义分割方法进行了系统的对比实验。对比模型包括经典的卷积神经网络方法U-Net、基于Transformer结构的TransUNet、Swin-Unet和最近提出的SCUNet++[24],以及最近提出的Mamba结构的VM-UNet。为确保实验的公平性,它们与本文MCNet的对比实验都在相同的训练设置和数据预处理策略下进行训练和评估。

在ACDC数据集上与已有方法的对比结果如表3所示。从表3可以看出,在多个指标上,MCNet均优于其他方法。其中,在对主要的左心室(LV)的分割任务中,MCNet的Dice系数达到0.93,超过简单卷积结构的U-Net约2.3%,分别超过Transformer结构的TransUNet、Swin-Unet和SCUNet++约3.9%、3.7%和2.8%,超过有Mamba状态空间结构的VM-UNet约1.5%;同时,平均Dice和IoU分别达到84.50%和77.92%,均优于其他5种方法,说明在成像清晰的ACDC数据集上,MCNet模型与其他主流方法相比具有更强的分割性能。

表3 在ACDC数据集上与已有分割方法的比较

Table 3 Comparison with existing segmentation methods on the ACDC dataset

方法	DSC_{RV}	DSC_{MYO}	DSC_{LV}	DSC_{mean}	IoU
U-Net	0.735 6	0.787 5	0.901 9	0.808 3	0.734 3
TransUNet	0.729 7	0.755 0	0.886 2	0.790 3	0.710 3
Swin-Unet	0.733 7	0.764 7	0.887 9	0.795 5	0.713 9
SCUNet++	0.741 9	0.776 6	0.897 0	0.805 2	0.723 6
VM-UNet	0.765 9	0.809 6	0.910 2	0.828 6	0.755 8
MCNet (Ours)	0.780 8	0.828 9	0.925 3	0.845 0	0.779 2

在CAMUS数据集上与已有方法的对比结果如表4所示,在多个指标上,MCNet均优于其他方法。其中,在对主要的左心室(LV)的分割任务中,MCNet的Dice系数达到0.90,超过简单卷积结构的U-Net约4.5%,分别超过Transformer结构的TransUNet、Swin-Unet和SCUNet++约2.5%、2.3%和2.9%,超过有Mamba状态空间结构的VM-UNet约1.0%;同时,平均的Dice、IoU分别达到88.32%和79.62%,均优于其他5种方法,说明在边界不清的CAMUS数据集上,MCNet模型与其他主流方法相比具有更强的分割性能。

同时,对比实验的可视化结果如图5、6所示。在两个数据集中,MCNet在差异化明显的心脏结构下,也能表现出稳定的分割性能。从图5的前两张图像可以看出,Swin-Unet和VM-UNet在分割小区域的心脏结构时,预测掩膜出现离群的点,分割的心脏结构存在锯齿;从第3张和第4张图可以看出,U-Net和Swin-Unet在对区域边界的分割时,如果边界不清晰,区域不明显,可能会出现漏检的情况。

表4　在CAMUS数据集上与已有分割方法的比较

Table 4　Comparison with existing segmentation methods on the CAMUS dataset

方法	DSC_{LV}	DSC_{MYO}	DSC_{LA}	DSC_{mean}	IoU
U-Net	0.855 2	0.761 4	0.846 8	0.821 1	0.706 0
TransUNet	0.874 9	0.813 7	0.873 8	0.854 1	0.752 1
Swin-Unet	0.876 9	0.815 6	0.863 8	0.852 1	0.748 1
SCUNet++	0.870 9	0.839 3	0.881 1	0.863 8	0.766 0
VM-UNet	0.879 7	0.840 9	0.882 7	0.867 8	0.770 5
MCNet (Ours)	0.900 7	0.855 7	0.893 1	0.883 2	0.796 2

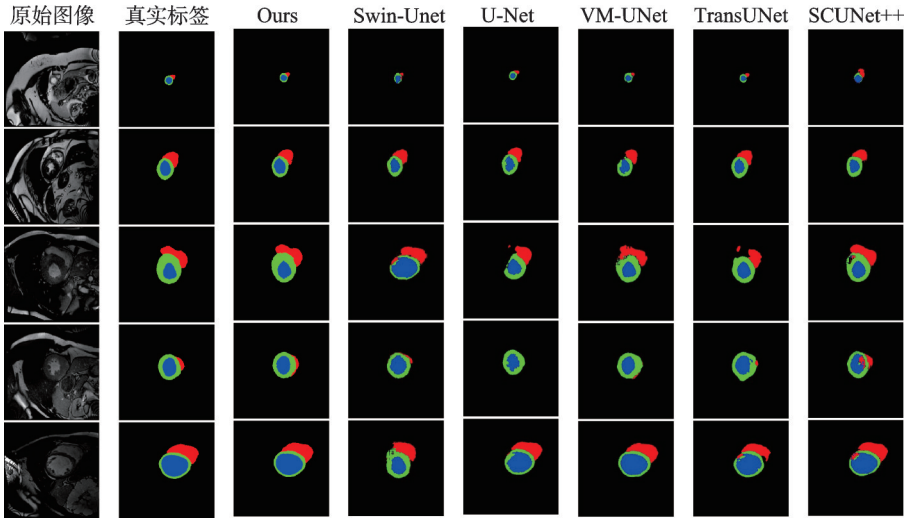


图5　在ACDC数据集上对比实验的可视化结果

Fig.5　Visualization results of comparative experiments on the ACDC dataset

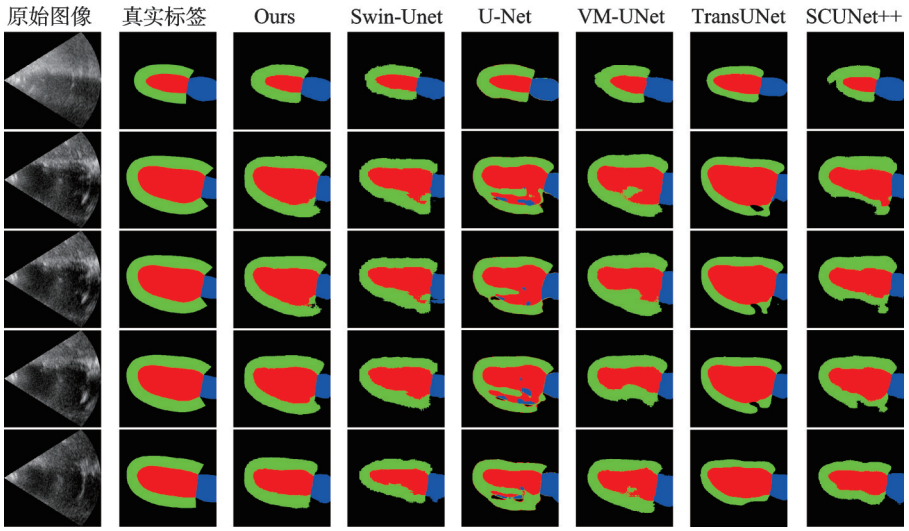


图6　在CAMUS数据集上对比实验的可视化结果

Fig.6　Visualization results of comparative experiments on the CAMUS dataset

综上,MCNet在左心室(蓝色)区域分割精度均优于其他方法。

从图6不难看出,简单卷积结构的U-Net分割效果不佳,容易出现离群的点;其次,基于Transformer结构的TransUNet、Swin-Unet和SCUNet++,在心肌(绿色区域)的处理上,容易出现断层的现象,基于Mamba结构的VM-UNet则容易出现误检的情况,将右心室(红色区域)检测为心肌。而本文的MCNet方法由于引入了Mamba状态空间分支对长距离语义的建模能力,以及卷积分支对局部纹理细节的保留,同时通过动态特征融合模块根据图像内容自动调整融合策略,进一步提高了模型对不同结构的判别能力,使得MCNet在可视化结果中展现出明显优势。

3.2.3 消融实验

为了验证MCNet中动态特征融合模块中各特征融合模块对模型性能的贡献,本文设计了各模块的消融实验,即在分别去除4个特征融合模块(SFF、BFAM、DFEM、Drop)后,对模型进行独立的训练和测试,最后与完整模型进行比较。所有实验在ACDC数据集和CAMUS数据集上进行,用于评价的指标包括Dice系数和IoU。

表5、6展示了消除单个特征融合模块的定量评估结果,其中前4行为双分支编码器,然后消除单个特征融合模块,例如"Mamba+CNN/-SFF"表示使用Mamba和CNN双分支,消除SFF特征融合模块的结果,最后2行为完整模型的实验结果。

表5 在ACDC数据集上的MCNet消融实验结果

Table 5 Ablation experimental results of MCNet on the ACDC dataset

编码器/融合模块	DSC_{RV}	DSC_{MYO}	DSC_{LV}	DSC_{mean}	IoU
Mamba+CNN/-SFF	0.749 8	0.803 4	0.907 6	0.820 3	0.748 4
Mamba+CNN/-BFAM	0.761 1	0.798 6	0.902 2	0.820 6	0.751 4
Mamba+CNN/-DFEM	0.746 5	0.790 8	0.909 8	0.815 7	0.740 4
Mamba+CNN/-Drop	0.753 1	0.802 7	0.907 5	0.821 1	0.749 0
MCNet (Ours)	0.780 8	0.828 9	0.925 3	0.845 0	0.779 2

表6 在CAMUS数据集上的MCNet消融实验结果

Table 6 Ablation experimental results of MCNet on the CAMUS dataset

编码器/融合模块	DSC_{LV}	DSC_{MYO}	DSC_{LA}	DSC_{mean}	IoU
Mamba+CNN/-SFF	0.871 4	0.821 9	0.862 9	0.852 1	0.750 2
Mamba+CNN/-BFAM	0.876 0	0.830 3	0.856 0	0.854 1	0.754 6
Mamba+CNN/-DFEM	0.876 5	0.812 3	0.835 9	0.841 6	0.736 4
Mamba+CNN/-Drop	0.885 0	0.833 2	0.855 7	0.858 0	0.760 3
MCNet (Ours)	0.900 7	0.855 7	0.893 1	0.883 2	0.796 2

从表5、6中结果可以看出,在去除空间特征融合(SFF)模块后,在ACDC数据集中,对边界细节清晰度要求高的右心室(RV)和心肌(MYO)分割任务,较完整模型Dice值分别降低了3.1%和2.5%;在CAMUS数据集中,左心房(LA)和心肌(MYO)分割的Dice值分别降低了3.0%和3.4%,这表明SFF在融合浅层空间细节和深层语义信息方面有积极作用。消除双时相融合模块(BFAM)后,在ACDC数据集中,心肌(MYO)区域的分割精度降低了很多,Dice降低了3.0%;在CAMUS数据集中,左心房(LA)区域的分割精度降低最多,Dice降低了3.7%,这表明BFAM能够有效加强边界信息表达,缓解模糊区域的误分割问题。消除深度特征提取模块(DFEM)后,在ACDC数据集中,心肌(MYO)区域的分割精度下降最多,Dice较完整模型降低了3.8%;在CAMUS数据集中,左心房(LA)区域的分割精度降

低最多,Dice较完整模型降低了5.7%,这表明DFEM模块能有效增强模型的全局语义建模和局部细节表达能力,使得整体分割结果更加连贯,提升对复杂区域的识别能力。消除Drop模块后,在ACDC数据集中,右心室(RV)和心肌(MYO)较完整模型,Dice值分别降低了2.8%和2.6%;在CAMUS数据集中,心肌(MYO)和左心房(LA)较完整模型,Dice值分别降低了2.3%和3.7%,说明Drop模块在不同模态下均对模型性能起到关键作用。

由上可知,完整的MCNet模型通过动态特征融合模块集成了4个模块的优势,实现了对多模态心脏影像中全局语义与局部结构的协同建模,这种动态集成策略不仅提升了整体分割精度,还有效缓解了单一模块在复杂医学图像中可能存在的局限性。

4　结束语

本文针对医学图像分割任务中局部细节表达与全局上下文建模难以兼顾的问题,提出了一种融合Mamba与CNN的混合架构,并引入MoE机制以实现动态专家选择。该方法充分结合了Mamba在长程依赖建模中的高效性与CNN在局部结构提取方面的优势,通过门控机制实现两者的优势互补,提升了模型在心脏MRI与超声影像分割中的准确性与泛化能力。

在两个差异性大的模态数据集上的实验结果表明,本文所提出的模型在多个评价指标上优于现有主流方法,尤其是在边界清晰度、复杂结构处理方面,表现出更强的鲁棒性和适应性。本文的研究工作为深度学习模型在医学图像智能分析中的集成与优化提供了新思路,也为实现精准高效的临床辅助诊断奠定了基础。本文的研究仍然存在局限性,模型目前只在MRI和超声两个模态数据集中进行实验,在其他模态(如低剂量CT影像等)场景下可能存在性能下降的情况,可通过调整Mamba分支的序列建模步长以适应CT的断层特征,同时优化CNN分支的卷积核大小增强密度差异捕捉能力。同时,MoE模块导致模型参数量增加,计算复杂度和内存占用均比基础模型大,后续的研究将进一步探索MoE模型的轻量化设计,考虑采用稀疏激活策略,仅对关键解剖区域激活对应专家模块,降低计算成本,并将其拓展到更多器官和疾病的多模态图像分析场景中,以推动智能医学图像处理技术的实用化和普及化发展。

参考文献:

[1] 刘明波, 何新叶, 杨晓红, 等.《中国心血管健康与疾病报告2023》要点解读[J]. 临床心血管病杂志, 2024, 40(8): 599-616.
LIU Mingbo, HE Xinye, YANG Xiaohong, et al. Interpretation of report on cardiovascular health and diseases in China 2023 [J]. Journal of Clinical Cardiology, 2024, 40(8): 599-616.

[2] 杨印凯,万鹏,石航,等.基于多模态超声对比学习的肝癌诊断方法[J]. 数据采集与处理, 2024, 39(4): 874-885.
YANG Yinkai, WAN Peng, SHI Hang, et al. Liver cancer diagnosis method on multi-modal ultrasound contrast learning[J]. Jouranl of Data Acquisition and Processing, 2024, 39(4): 874-885.

[3] BRAHIM K, QAYYUM A, LALANDE A, et al. A 3D network based shape prior for automatic myocardial disease segmentation in delayed-enhancement MRI[J]. IRBM, 2021, 42(6): 424-434.

[4] ISLAM M R, QARAQE M, SERPEDIN E. CoST-UNet: Convolution and swin Transformer based deep learning architecture for cardiac segmentation[J]. Biomedical Signal Processing and Control, 2024, 96: 106633.

[5] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]// Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Munich, Germany: Springer International Publishing, 2015: 234-241.

[6] CHEN J, LU Y, YU Q, et al. TransUNet: Transformers make strong encoders for medical image segmentation[EB/OL]. (2021-02-08). https://doi.org/10.48550/arXiv.2102.04306.

[7] CAO H, WANG Y, CHEN J, et al. Swin-Unet: Unet-like pure Transformer for medical image segmentation[EB/OL]. (2021-05-12). https://doi.org/10.48550/arXiv.2105.05537.

[8] GU A, DAO T. Mamba: Linear-time sequence modeling with selective state spaces[EB/OL]. (2023-12-01). https://doi.org/

10.48550/arXiv.2312.00752.

[9] RUAN J, LI J, XIANG S. VM-UNet: Vision Mamba UNet for medical image segmentation[EB/OL]. (2024-11-08). https://doi.org/10.48550/arXiv.2402.02491.

[10] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: Learning where to look for the pancreas[EB/OL]. (2018-05-21). https://arxiv.org/pdf/1804.03999.pdf.

[11] ÇIÇEK Ö, ABDULKADIR A, LIENKAMP S S, et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation[C]//Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016. Athens, Greece: Springer International Publishing, 2016: 424-432.

[12] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2016: 770-778.

[13] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2017: 4700-4708.

[14] JACOBS R A, JORDAN M I, NOWLAN S J, et al. Adaptive mixtures of local experts[J]. Neural Computation, 1991, 3(1): 79-87.

[15] LEPIKHIN D, LEE H J, XU Y, et al. GShard: Scaling giant models with conditional computation and automatic sharding [C]//Proceedings of International Conference on Learning Representations. [S.l.]: ICLR, 2021.

[16] FEDUS W, ZOPH B, SHAZEER N. Switch Transformers: Scaling to trillion parameter models with simple and efficient sparsity[J]. Journal of Machine Learning Research, 2022, 23(120): 1-39.

[17] ZHENG H, WEI D, ZHENG Y. MoME: Mixture of multimodal experts for cancer survival prediction[EB/OL]. (2024-06-14). https://doi.org/10.48550/arXiv.2406.09696.

[18] ZHANG Z, BAO L, XIANG S, et al. B2CNet: A progressive change boundary-to-center refinement network for multi-temporal remote sensing images change detection[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 11322-11338.

[19] YANG L, ZHANG R Y, LI L, et al. SimAM: A simple, parameter-free attention module for convolutional neural networks [C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2021: 11863-11874.

[20] MA X, YANG J, HONG T, et al. STNet: Spatial and temporal feature fusion network for change detection in remote sensing images[C]//Proceedings of 2023 IEEE International Conference on Multimedia and Expo (ICME). [S.l.]: IEEE, 2023: 2195-2200.

[21] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup: Beyond empirical risk minimization[EB/OL]. (2017-10-25). https://doi.org/10.48550/arXiv.1710.09412.

[22] LECLERC S, SMISTAD E, PEDROSA J, et al. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography[J]. IEEE Transactions on Medical Imaging, 2019, 38(9): 2198-2210.

[23] BERNARD O, LALANDE A, ZOTTI C, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?[J]. IEEE Transactions on Medical Imaging, 2018, 37(11): 2514-2525.

[24] CHEN Y, ZOU B, GUO Z, et al. SCUNet++: Swin-UNet and CNN bottleneck hybrid architecture with multi-fusion dense skip connection for pulmonary embolism CT image segmentation[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. [S.l.]: IEEE, 2024: 7759-7767.

作者简介：

肖瑞(2002-),男,硕士研究生,研究方向:心脏影像分析,E-mail: 4627xxr@nuaa.edu.cn。

邵伟(1986-),**通信作者**,男,副教授,研究方向机器学习、医学图像处理,E-mail:shaowei20022005@nuaa.eud.cn。

(编辑:王静)