

融合多尺度特征和注意力机制的肺炎病灶分割方法

郭乙轩¹, 管江南¹, 武杰¹, 姚旭峰²

1. 上海理工大学健康科学与工程学院, 上海 200093; 2. 上海健康医学院医学影像学院, 上海 201318

【摘要】针对肺炎病灶区域大小形状差异大、区域边界不清晰等问题,提出一种融合多尺度特征和注意力机制的分割模型,该模型由多尺度视觉状态空间块构成编码器与解码器的主要结构,利用级联的不同大小的卷积核捕捉不同病灶的多尺度上下文信息,有效应对肺炎病灶形状大小的显著变异性。同时,在跳跃连接层中引入CBAM模块,通过空间和通道维度的双分支交互机制,增强编码器与解码器特征的语义一致性,减少特征融合时的信息丢失。在多个公开肺部病理图像数据集上进行实验,在Large COVID-19 CT scan slice与CNCB两个肺炎病灶分割数据集上的Dice相似系数分别达到87.7%与85.1%,IOU分别达到79.7%与75.3%,并在LUNA16肺结节数据集上验证模型的泛化能力,实验结果表明,相较于传统U-Net架构与transformer架构该模型分割结果具有显著提升。

【关键词】肺炎病灶; 图像分割; 特征融合; 多尺度; 注意力机制; 医学图像处理

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2026)07-0939-07

Pneumonia lesion segmentation method integrating multi-scale feature fusion and attention mechanisms

GUO Yixuan¹, GUAN Jiangnan¹, WU Jie¹, YAO Xufeng²

1. School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 2. School of Medical Imaging, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China

Abstract: A segmentation model integrating multi-scale features and attention mechanisms is developed to address the challenges of large variations in the size and shape of pneumonia lesions as well as unclear boundaries. The model employs multi-scale visual state space blocks as the main components of both the encoder and decoder. It utilizes cascaded convolution kernels of different sizes to capture multi-scale contextual information from varying-sized lesions, thereby effectively coping with the significant variations in the shape and scale of pneumonia lesions. Meanwhile, a convolutional block attention module is embedded in the skip connection layers. Through a dual-branch interaction mechanism across spatial and channel dimensions, the semantic consistency of encoder-decoder features is improved, and the information loss during feature fusion is alleviated. Experiments are conducted on several public lung pathological image datasets. On the Large COVID-19 CT scan slice dataset and the CNCB pneumonia lesion segmentation dataset, the proposed model achieves Dice similarity coefficients of 87.7% and 85.1%, and IoU of 79.7% and 75.3%, respectively. In addition, the generalization performance of the model is verified on the LUNA16 lung nodule dataset. Experimental results demonstrate that the proposed model exhibits superior segmentation performance over conventional U-Net architectures and Transformer-based models.

Keywords: pneumonia lesion; image segmentation; feature fusion; multi-scale; mixed attention; medical image processing

前言

【收稿日期】2026-03-15

【基金项目】上海市科学技术委员会地方院校能力建设项目
(23010502700)

【作者简介】郭乙轩, 硕士研究生, 研究方向: 医学图像处理与分析,
E-mail: 1670188964@qq.com

【通信作者】武杰, 博士, 讲师, 研究方向: 医学图像处理与分析,
E-mail: wujie3773@sina.com

肺炎作为一种高传染性急性呼吸道疾病,对全球公共卫生和社会经济造成深远影响^[1]。胸部计算机断层扫描(CT)因其能够直观呈现肺部病变细节,成为肺炎早期筛查与病程监控的关键技术。然而,传统人工阅片效率低且主观性强,尤其在医疗资源紧张地区,自动化病灶分割技术对缓解诊断压力、提升治疗决策准确性具有重要意义^[2]。

深度学习技术的发展为医学图像分割带来突破^[3],其中U-Net及其变体凭借U型架构和跳跃连接

设计成为主流方法^[4]。陈梦飞等^[5]提出 RPR-MLP 分割网络,在肺肿瘤 CT 图像分割中取得不错的效果。Alshomrani 等^[6]于 2024 年提出 SAA-UNet,通过空间注意力与门控机制提升分割精度。TDD-UNet 引入 Transformer 结构增强远程依赖建模能力^[7]。然而,传统卷积神经网络(CNN)受限于局部感受野,难以捕捉长程上下文关联;而 Transformer 虽能建模全局依赖,却面临计算复杂度高、对高分辨率医学图像处理效率低下的挑战。

近年来,视觉状态空间(Visual State Space, VSS)模型因其线性复杂度和高效长序列建模能力引发关注^[8-10]。作为 VSS 模型的代表性方法,Mamba 通过选择性状态空间机制动态处理输入信息^[11],在捕获全局依赖的同时保持计算高效性。基于 Mamba 的架构,Vision Mamba 在医学图像分类中展现潜力,其通过 VSS 块实现特征融合与全局上下文建模的平衡,但现有的医学图像分割方法仍存在对多尺度特征利用不充分、病灶内部纹理建模能力不足等问题。

针对上述挑战,本文提出一种融合多尺度特征和注意力机制的肺炎病灶分割方法,通过以下创新提升肺炎病灶分割性能:(1)构建一个基于编码器解码器结构的分割网络,不同于传统 CNN 网络,该模型的编码器基本结构由补丁合并层与 VSS 块构成,解码器的每一层结构由补丁拓展层与 VSS 块构成。(2)提出了一种改进的 VSS 块,在 VSS 块末端集成多尺

度特征融合模块,通过并联的不同大小的卷积核提取不同感受野的纹理特征,有效建模感染区域的复杂结构变化。(3)为增强编码器与解码器特征的语义一致性,在跳跃连接中嵌入卷积注意力(CBAM)模块,通过通道维度的跨层交互与空间维度的区域聚焦,增强不同层级特征的互补性,尤其是病灶边缘等细节区域的定位准确度。(4)在多个公开肺部图像数据集上进行实验。实验结果表明,该方法显著提升分割精度与鲁棒性,为临床应用提供了高效解决方案。

1 方法

1.1 Mamba-UNet 网络结构

Mamba-UNet 结合了 Mamba 的视觉状态空间块与 U-Net 的 U 型结构,旨在克服传统 CNN 网络和 Transformer 在处理长程依赖和计算效率方面的局限性。用多尺度视觉状态空间块(Multi-Scale Vision State Space, MSVSS)代替传统 U-Net 中的特征提取,增强该模型特征提取和图像内部依赖关系的能力,结合注意力机制,使得模型关注编码过程中通道和空间关系的相互作用,这种增强的跳跃连接,在弥合编码器和解码器特征之间的语义差距中发挥重要作用,对图像分割的整体性能和准确性至关重要。Mamba-UNet 网络结构见图 1。

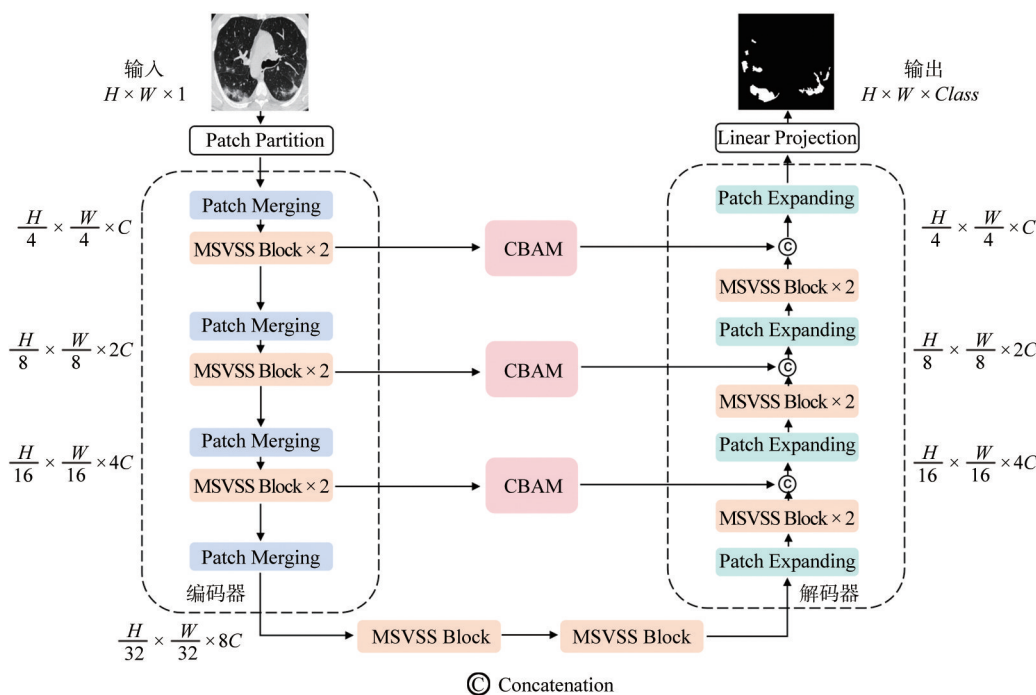


图1 Mamba-UNet 结构
Figure 1 Mamba-UNet architecture

Mamba-UNet 的整体结构如图 1 所示, Mamba-UNet 包含补丁嵌入层、编码器、解码器、投影层以及跳跃连接层, Mamba-UNet 的网络结构从获取 $H \times W \times 1$ 图像数据开始, 该数据通过使用 Vision Mamba 中 Patch Partition 处理, 补丁嵌入层将输入的图像 $x \in \mathbb{R}^{H \times W \times C}$ 划分为 4×4 的非重叠补丁, 其尺寸固定为 $H/4 \times W/4 \times 16$, 记该数据维度为 $H/4 \times W/4 \times C$, 然后数据流经过两个多尺度视觉状态空间块, 该模块用于特征提取, 该操作不改变输入张量的特征维度, 之后经历 3 个阶段, 其中每个阶段都包含 1 个补丁合并层, 以降低输入特征的高度和宽度, 同时增加通道数量, 再经过 2 个多尺度视觉状态空间模块进行特征提取。在经过最开始的 3 对 MSVSS 模块后, 数据还将被送入跳跃连接层的 CBAM 模块, 经过一系列处理, 以匹配多尺度视觉状态空间模块输出的特征维度。

经过一系列处理后, 数据特征维度为 $H/32 \times W/32 \times 8C$, 随后将其送入解码器。在解码器中, 数据经历 3 个阶段, 每个阶段包含补丁扩展操作来增加数据的高度和宽度, 同时减少特征通道数, 随后送入 MSVSS。最后经过补丁扩展和线性投影, 生成 $H \times W \times C$ 的分割图像。

1.2 MSVSS 模块

为了捕捉图像中不同尺度的信息, 本文提出 MSVSS 模块, 通过在 VSS 块末端引入多尺度特征融合模块来解决图像分割中分割不同大小病灶不准确的难点。MSVSS 模块是本文模型的核心结构, 结构如图 2 所示, 经过层归一化后, 输入分为两个部分, 在第一个部分中, 输入经过线性层和激活函数, 在第二个部分中, 输入经过线性层, 深度可分离卷积和激活函数, 随后送入 SS2D 块增强特征提取, 之后特征通过归一化层标准化, 然后将第一部分的输出与之合并送入线性层, 将此结果与残差连接结合在一起^[12], 再经过归一化层送入多尺度特征融合模块, 最后, 将其与残差连接结合构成 MSVSS 块的输出。激活函数默认为 Sigmoid。MSVSS 的定义由式(1)和式(2)给出:

$$\begin{aligned} \hat{f}_i^d &= C_2(\sigma_1(C_1(\text{LN}(f_i^d))) \times S_{2D}(\text{LN}(f_i^d))) + f_i^d(1) \\ f_{i+1}^d &= \text{MS}(\text{LN}(\hat{f}_i^d)) + \hat{f}_i^d \end{aligned} \quad (2)$$

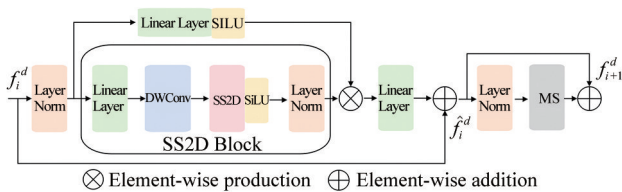


图2 多尺度视觉状态空间块

Figure 2 Multi-scale vision state space block

其中, f_i^d 和 f_{i+1}^d 表示第 i 阶段的输入和输出特征图, LN 表示归一化层, σ_1 表示 Sigmoid 激活函数, MS 表示多尺度特征融合模块, C_1 与 C_2 表示线性投影。

如图 3 所示, 多尺度特征融合模块由一层平行的卷积层, 两个线性层和激活函数构成, 笔者采用 3 个不同大小的深度卷积进行卷积操作^[13]。多尺度特征融合模块的定义由式(3)、式(4)给出:

$$m_i^{\text{out}1} = \sigma_2(C_3(m_i^{\text{in}})) \quad (3)$$

$$m_i^{\text{out}2} = C_4\left(\sum_{\text{size} \in \text{SIZE}} \text{DWConv}_{\text{size}}(m_i^{\text{out}1}) + m_i^{\text{out}1}\right) \quad (4)$$

其中, m_i^{in} 和 $m_i^{\text{out}2}$ 表示第 i 阶段多尺度特征融合中输入和输出的张量, $m_i^{\text{out}1}$ 表示第一个线性层的输出, C_3 和 C_4 表示线性投影, $\text{DWConv}_{\text{size}}$ 表示深度卷积的卷积核大小。SIZE 定义了一组平行的卷积层, 卷积核大小为 $1 \times 1, 3 \times 3, 5 \times 5$ 。 σ_2 表示 GELU 激活函数^[14]。

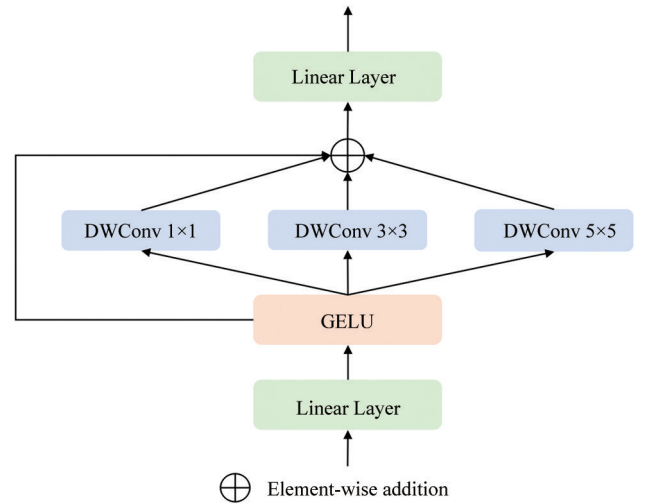


图3 多尺度特征融合模块

Figure 3 Multi-scale feature fusion block

1.3 CBAM 模块

为进一步增强编码器与解码器之间特征融合的有效性, 本研究在跳跃连接层中引入 CBAM 模块^[15]。CBAM 模块通过串联通道注意力块与空间注意力块, 在不同维度上自适应地重新加权特征响应, 从而提升网络对病灶区域的关注度, 抑制无关背景信息, 其结构如图 4 所示。

通道注意力机制用于自适应地调整不同通道特征的重要性, 从而增强特征表达能力, 其结构如图 5 所示。对于输入特征图 $F \in \mathbb{R}^{C \times H \times W}$, 首先通过全局平均池化(GAP)和全局最大池化(GMP)获得两个通道描述符 F_{avg} 与 F_{max} :

$$F_{\text{avg}} = \text{GAP}(F) \quad (5)$$

$$F_{\text{max}} = \text{GMP}(F) \quad (6)$$

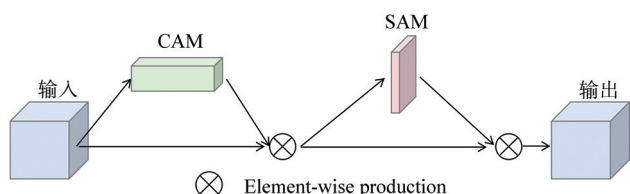


图4 CBAM模块
Figure 4 CBAM block

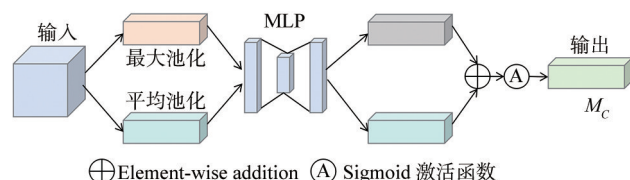


图5 通道注意力块
Figure 5 Channel attention block

然后将二者输入共享的多层感知机(MLP),采用通道压缩以减少参数数量和防止过拟合。经逐元素相加后,通过 Sigmoid 激活函数 σ 生成通道注意力权重图,通道注意力 M_c 定义如式(7)所示:

$$M_c(F) = \sigma(\text{MLP}(F_{\text{avg}}) + \text{MLP}(F_{\text{max}})) \quad (7)$$

最终利用该注意力图对输入特征进行加权:

$$F = M_c(F) \otimes F \quad (8)$$

其中, $\otimes$ 表示逐元素相乘操作。该过程能够突出病灶相关的显著通道,加强纹理、密度变化与边缘特征,抑制无关通道响应,从而提高网络的全局语义一致性与特征判别能力。

空间注意力模块用于在空间维度上确定应当关注的位置,以突出病灶区域并抑制背景噪声,结构如图6所示,对于经过通道注意力优化的特征图 $F \in R^{C \times H \times W}$,空间注意力首先在通道方向上进行平均池化 F_{avg} 和最大池化 F_{max} :

$$F_{\text{avg}} = \text{AvgPool}(F) \quad (9)$$

$$F_{\text{max}} = \text{MaxPool}(F) \quad (10)$$

将二者拼接后,经过 7×7 大小的卷积并通过 Sigmoid 激活函数 σ 得到空间注意力权重图 M_s :

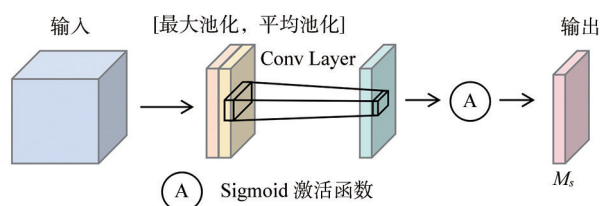


图6 空间注意力模块
Figure 6 Spatial attention block

$$M_s(F) = \sigma(f^{7 \times 7}([F_{\text{avg}}, F_{\text{max}}])) \quad (11)$$

然后通过加权运算获得最终的空间增强特征:

$$F = M_s(F) \otimes F \quad (12)$$

注意力机制可在空间层面上强化对肺炎病灶区域的响应,提升网络对模糊边界、小面积及弥散病灶的识别能力。通过串联通道注意力块与空间注意力块,CBAM 模块在全局与局部尺度上实现了特征的自适应增强,从而提高肺炎病灶分割的精度与鲁棒性。在本文改进网络中,CBAM 模块被嵌入于每一层的跳跃连接中,在与解码器特征拼接前对编码器特征进行注意力重校准,可以显著提高跨层特征的语义一致性与空间定位精度。

2 实验

2.1 数据集介绍

本研究采用两个公开肺部CT病理图像数据集评估模型分割肺炎病灶区域的性能,还通过LUNA16肺结节数据集来评估本文模型的泛化能力。Large COVID-19 CT scan slice dataset数据集来源于意大利、俄罗斯等跨国联合研究项目,原始数据整合自Maftouni 等^[16]整理的Curated COVID-CT数据库与COVID-19 CT病灶分割数据集,通过标准化预处理流程形成统一研究样本库。该数据集包含113名病人的2729张肺部病理切片图像。CNCB肺炎CT数据集由中国胸部CT图像调查联盟构建^[17],所使用的数据集主要包含经专家标注的胸部CT影像,覆盖多家医院与不同感染阶段的患者,影像分辨率高、病灶特征丰富。平台提供的标注包括感染区域轮廓及对应的病灶类型信息,该数据集包含150名肺炎患者的750个CT切片。在模型泛化实验中使用Lung Nodule Analysis 2016(LUNA16)公开数据集^[18],该数据集包含888例胸部低剂量CT扫描影像,并附有由多名放射科专家标注的肺结节信息及肺结节区域掩膜。LUNA16数据具有分辨率高、影像质量一致、标注精度高等特点,常被用于肺部影像分析与自动检测算法研究。本研究使用肺结节掩膜数据,用以评估模型在不同数据分布下的泛化能力。实验数据集的80%划分为训练集,10%划分为验证集,剩余10%为测试集。

2.2 实验设置

本文实验基于并行智算云平台构建训练环境,硬件配置采用NVIDIA GeForce RTX 4090显卡,24 GB GDDR6X 显存,与AMD EPYC 7402 处理器协同计算,配备60GB DDR4 内存及NVMe SSD 高速存储系统。软件环境部署于Ubuntu 22.04 操作系统,依托CUDA 11.8加速库与PyTorch 2.0.1 深度学习框架,在

Python 3.10.11 环境下进行模型训练。在网络训练过程中网络优化采用AdamW算法,设置初始学习率为0.000 1,并应用余弦退火策略动态调整至最小学习率0.000 01,以加权Dice交叉熵损失函数监督训练100个周期,Dice损失与交叉熵损失权重均取0.5,每批次输入16张CT切片。

2.3 评价指标

本文采用如下几个指标评估模型病灶分割任务的性能:交并比(IOU)、Dice相似系数(DSC)及F1分数。各指标基于像素级混淆矩阵计算,定义如下:

$$IOU = \frac{TP}{TP + FN + FP} \tag{13}$$

$$DSC = \frac{2TP}{2TP + FN + FP} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{17}$$

其中,TP为正确预测的病灶像素数,FP为误诊为病灶像素数,FN为漏诊的真实病灶像素数,TN为正确识别的背景像素数。

3 结果与分析

3.1 不同模型对比实验

本文通过对比实验评估模型性能,对比模型包括U-Net^[19]、Res-UNet^[20]、nn-UNet^[21]、Trans-UNet^[22]、Swin-UNet^[23]、VM-UNet^[9]。通过消融实验评估多尺度特征融合模块和CBAM模块对图像分割性能的影响。

在Large COVID-19 CT scan slice dataset数据集

上的实验结果如表1所示,本文模型在DSC和IOU两项核心分割指标上均取得最优表现,分别达到87.7%和79.7%。模型参数量略高于部分对比模型,但分割性能提升显著。与基于卷积结构的分割模型相比,本文模型通过引入MSVSS块与CBAM模块,在不同感受野下融合全局与局部特征信息,增强空间与通道间的信息交互与特征选择能力,有效提升病灶区域边界的识别能力。相比于Trans-UNet,本文模型的DSC与IOU分别提升1.1%和2.4%,提升分割效果的同时也降低模型参数量。与基于Mamba类的模型VM-UNet相比,本文模型的DSC与IOU分别提升0.9%和1.4%。Large COVID-19 CT scan slice dataset数据集的肺炎病灶分割测试结果如图7所示,为了更直观地展示模型在局部区域的分割效果,在原图中使用红色矩形框标示感兴趣区域,并在图像右上角给出对应区域的放大图,以突出显示模型在病灶边界及细节结构处的分割表现。

表1 大型肺炎CT数据集上对比实验的评价指标
Table 1 Evaluation metrics of comparative experiments on the large-scale pneumonia CT dataset

方法	DSC/%	IOU/%	F1/%	参数量/M
U-Net	81.7	71.9	89.1	31.4
Res-UNet	82.3	72.4	88.7	35.6
nn-UNet	85.7	75.8	90.7	41.3
Trans-UNet	86.6	77.3	91.6	105.2
Swin-UNet	86.5	76.5	88.9	62.1
VM-UNet	86.8	78.3	91.6	66.7
本文模型	87.7	79.7	92.2	68.2

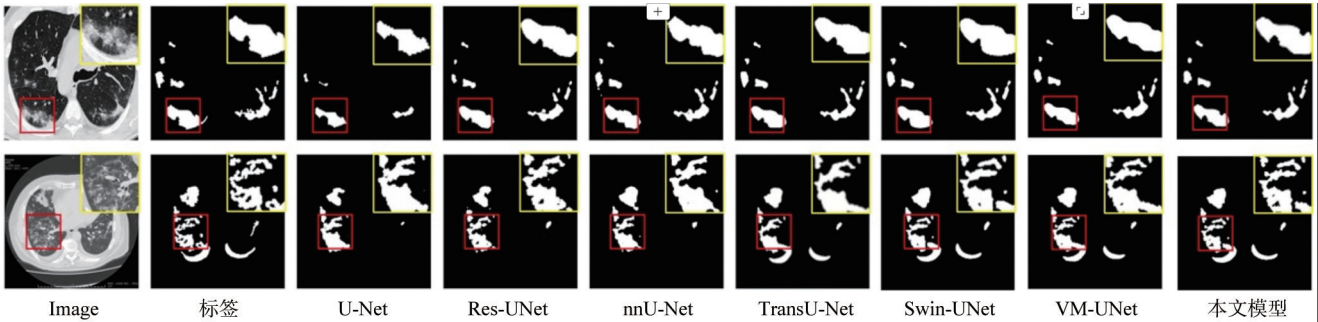


图7 大型肺炎CT数据集上的分割结果

Figure 7 Segmentation results on the large-scale pneumonia CT dataset

第1、2行图像分别代表不同患者的图像

在CNCB肺炎CT数据集上进行的实验结果如表2所示。与对比模型相比,本文模型在各项评价指标上均取得明显提升,对比U-Net,DSC与IOU分别

提高了6.6%和6.8%,相较于Res-UNet,DSC与IOU分别提高5.7%和5.9%,对比nn-UNet,DSC与IOU分别提高1.9%和2.0%,对比transformer类模型

Trans-UNet 与 Swin-UNet, DSC 提高 2.6% 和 3.4%, IOU 提高 2.3% 和 3.4%。相较于 VM-UNet, DSC 与 IOU 分别提高 1.5% 与 1.2%。表明模型在病灶区域的识别与分割精度上表现更优。对比实验结果验证本文方法在不同肺炎病灶数据集下的稳健性,能够在 CNCB 肺炎 CT 数据集上有效捕获肺炎病灶的形态特征,并保持较高的分割一致性。该方法在 CNCB 数据集上的性能进一步证明其在肺炎病灶分割任务中的应用潜力。

CNCB 肺炎 CT 数据集的肺炎病灶分割测试结果如图 8 所示,可以观察到基于传统的 CNN 类模型的分割边缘出现类过平滑现象,缺少细节特征,对边缘像素的错误分类说明这些模型未能有效地提取到肺炎病灶边缘特征,而本文模型克服了边缘难以分割的问题,更能完整清晰地分割病灶边缘。

表 2 CNCB 肺炎 CT 数据集上对比实验的评价指标
Table 2 Evaluation metrics of comparative experiments on the CNCB pneumonia CT dataset

方法	DSC/%	IOU/%	F1/%	参数量/M
U-Net	78.5	68.5	86.7	31.4
Res-UNet	79.4	69.4	86.5	35.6
nn-UNet	83.2	73.3	89.6	41.3
Trans-UNet	82.5	73.0	88.8	105.2
Swin-UNet	81.7	71.9	88.1	62.1
VM-UNet	83.6	74.1	89.3	66.7
本文模型	85.1	75.3	90.8	68.2

本文在 LUNA16 数据集上对肺结节分割性能进行评估。肺结节相较于肺炎病灶区域更小,更考验

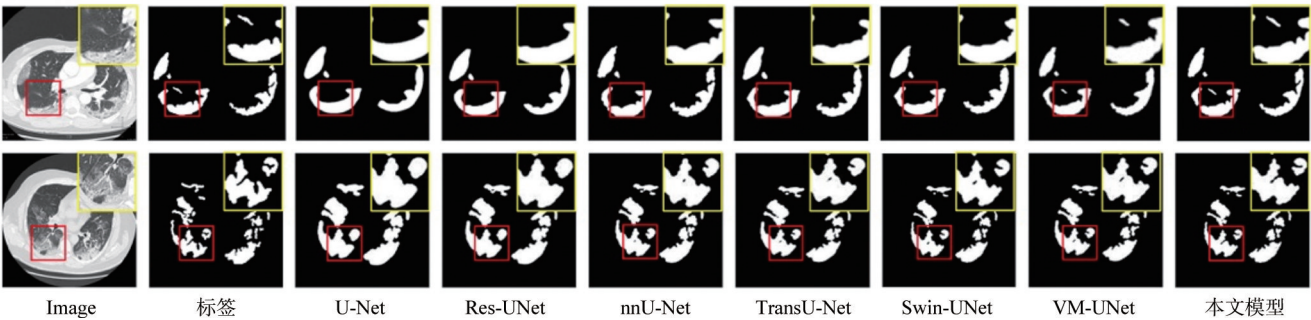


图 8 CNCB 肺炎 CT 数据集上的分割结果
Figure 8 Segmentation results on CNCB pneumonia CT dataset

第 1、2 行图像分别代表不同患者的图像

模型对微小病灶分割的有效性。对比实验结果如表 3 所示,本文方法在 DSC 和 IOU 指标分别达到 84.5% 和 74.3%,相比 CNN 类与 transformer 类模型有不同程度的提升。本文方法在不同数据集下仍能准确分割目标区域,体现出较强的泛化能力与鲁棒性。

3.2 消融实验

为探究 CBAM 模块和多尺度特征融合模块对模型性能的影响,笔者在上文所述的 3 个数据集上设计了消融实验,分别移除了模型中的这两个模块。实验结果如表 4 所示,其中 Case A 不引入多尺度特征融合模块也不引入 CBAM 模块,Case B 仅引入多尺度特征融合模块,Case C 仅引入 CBAM 模块,Case D 同时引入多尺度特征融合模块和 CBAM 模块。

实验结果表明,仅引入多尺度特征融合模块与仅引入 CBAM 模块的 DSC 和 IOU 均有一定程度的下降。这些结果表明,本研究提出的改进策略是有效的。CBAM 模块和多尺度特征融合模块的加持,使得模型更能捕

表 3 LUNA16 肺结节 CT 数据集上对比实验的评价指标
Table 3 Evaluation metrics of comparative experiments on the LUNA16 lung nodule CT dataset

方法	DSC/%	IOU/%	F1/%	参数量/M
U-Net	78.1	67.7	86.3	31.4
Res-UNet	78.9	68.4	86.7	35.6
nn-UNet	82.6	72.5	89.7	41.3
Trans-UNet	81.4	71.3	88.4	105.2
Swin-UNet	82.1	71.8	88.8	62.1
VM-UNet	83.2	72.7	89.6	66.7
本文模型	84.5	74.3	91.0	68.2

捉图像长距离依赖关系,提高对重要区域以及不同大小病灶的注意力,进而提升分割能力。基于在大型肺炎 CT 数据集、CNCB 肺炎 CT 数据集、LUNA16 数据集上的消融实验结果,可以量化 CBAM 模块与多尺度特

表4 3个数据集上消融实验的结果(%)
Table 4 Results of ablation experiments on the 3 datasets (%)

模型	大型肺炎CT数据集			CNCB肺炎CT数据集			LUNA16数据集		
	DSC	IOU	F1	DSC	IOU	F1	DSC	IOU	F1
Case A	85.1	76.8	90.8	82.7	72.1	89.4	82.3	71.7	89.4
Case B	86.5	78.4	91.6	83.8	73.6	90.0	83.5	72.9	90.2
Case C	86.8	78.9	91.8	84.2	74.2	90.4	83.8	73.4	90.5
Case D	87.7	79.7	92.2	85.1	75.3	90.8	84.5	74.3	91.0

征融合模块对模型分割性能的提升作用。实验设置的4组对照,Case A性能较低,其DSC在3数据集分别为85.1%、82.7%、82.3%;Case B和Case C均实现性能提升,且Case C提升更明显,DSC较Case A分别提高1.7%、1.5%、1.5%;Case D的各项指标均最优,DSC较Case A最高提升2.6%,较单一模块组有0.8%~1.2%的增加,充分证明多尺度特征融合模块与CBAM模块的有效性。

4 结论

本文中提出一种融合多尺度特征和注意力机制的肺炎病灶分割方法,并探索基于VSS模型在图像分割领域的应用。Mamba-UNet作为一种新兴的分割网络,能够有效捕捉长距离依赖关系,依靠长序列建模能力,增强对病灶形态变异,不规则边界的适应性。对复杂医学图像的分割起到重要作用。在VSS末端引入多尺度特征融合模块,通过级联深度卷积整合不同感受野特征,解决小病灶漏检问题。在跳跃连接层中引入CBAM模块,更加有效得整合通道和空间的特征以进行分割。通过在多个数据集上的对比实验与消融实验,验证本文改进模型的有效性。本文改进模型结合Mamba-UNet的优势,并针对肺炎病灶分割任务进行优化,在性能上超越传统的UNet及其变体,为肺炎病灶分割提供一种更为精确有效的方法。尽管本文方法在多个数据集上取得较好的实验结果,但仍存在一定局限性。实验数据规模仍相对有限,模型主要在公开数据集上进行验证,尚未在大规模真实临床数据环境中进行系统评估。其次,虽然模型引入了VSS结构,但Mamba扫描机制在医学图像分割任务中的特征建模能力仍有进一步研究空间。未来的研究将进一步扩展多中心临床数据集,并探索三维视觉状态空间模型,以充分利用CT体数据的空间信息。同时,通过优化Mamba扫描机制与序列建模策略,进一步提升模型对复杂医学图像结构特征的表达能力和分割性能。

【参考文献】

[1] Machkovech HM, Hahn AM, Garonzik Wang J, et al. Persistent SARS-CoV-2 infection: significance and implications[J]. Lancet Infect Dis, 2024, 24(7): e453-e462.
[2] Oulefki A, Agaia S, Trongtirakul T, et al. Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images [J]. Pattern Recognit, 2021, 114: 107747.

[3] Hertel R, Benlamri R. Deep learning techniques for COVID-19 diagnosis and prognosis based on radiological imaging [J]. ACM Comput Surv, 2023, 55(12): 260.
[4] Ehab W, Huang LN, Li YM. UNet and variants for medical image segmentation[J]. Int J Netw Dyn Intell, 2024, 3: 100009.
[5] 陈梦飞, 王姚芬, 王海玲, 等. 基于多尺度卷积与并行反向注意力的医学图像分割[J]. 中国医学物理学杂志, 2025, 42(1): 27-36.
Chen MF, Wang RF, Wang HL, et al. Medical image segmentation based on multi-scale convolution and parallel reverse attention [J]. Chinese Journal of Medical Physics, 2025, 42(1): 27-36.
[6] Alshomrani S, Arif M, Al Ghamdi MA. SAA-UNet: spatial attention and attention gate UNet for COVID-19 pneumonia segmentation from computed tomography[J]. Diagnostics, 2023, 13(9): 1658.
[7] Huang XP, Chen JX, Chen MZ, et al. TDD-UNet: transformer with double decoder UNet for COVID-19 lesions segmentation[J]. Comput Biol Med, 2022, 151(Pt A): 106306.
[8] Zhang SL, Liu JY, Qian TY, et al. Prompt-guided dual-path UNet with Mamba for medical image segmentation[EB/OL]. (2025-03-25)[2026-01-10]. <https://arxiv.org/abs/2503.19589>.
[9] Wang ZY, Zheng JQ, Zhang YC, et al. Mamba-UNet: UNet-like pure visual mamba for medical image segmentation[EB/OL]. (2024-03-30)[2026-01-10]. <https://arxiv.org/abs/2402.05079>.
[10] Yang XF, Luo ZY, Wu YL, et al. TMU: transmission-enhanced mamba-UNet for medical image segmentation [C]//Advanced Intelligent Computing Technology and Applications. Singapore: Springer Nature Singapore, 2024: 428-438.
[11] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.
[12] Ramachandran P, Zoph B, Le QV. Searching for activation functions [EB/OL]. (2017-10-17) [2026-01-10]. <https://arxiv.org/abs/1710.05941>.
[13] Chollet F. Xception: deep learning with depthwise separable convolutions [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 1800-1807.
[14] Hendrycks D, Gimpel K. Gaussian error linear units (GELUs) [EB/OL]. (2023-06-06)[2026-01-10]. <https://arxiv.org/abs/1606.08415>.
[15] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module [C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
[16] Maftouni M, Law ACC, Shen B, et al. A robust ensemble-deep learning model for COVID-19 diagnosis based on an integrated CT scan images database [C]//IISE Annual Conference and Expo 2021. Norcross, GA, USA: Institute of Industrial and Systems Engineers, 2021: 632-637.
[17] Zhang K, Liu XH, Shen J, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography [J]. Cell, 2020, 181(6): 1423-1433.e11.
[18] Setio AAA, Traverso A, de Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge [J]. Med Image Anal, 2017, 42: 1-13.
[19] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.
[20] Diakogiannis FI, Waldner F, Caccetta P, et al. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data [J]. ISPRS J Photogramm Remote Sens, 2020, 162: 94-114.
[21] Isensee F, Jaeger PF, Kohl SAA, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation [J]. Nat Methods, 2021, 18(2): 203-211.
[22] Chen JN, Mei JR, Li XH, et al. TransUNet: rethinking the U-Net architecture design for medical image segmentation through the lens of transformers [J]. Med Image Anal, 2024, 97: 103280.
[23] Cao H, Wang YY, Chen J, et al. Swin-UNet: Unet-like pure transformer for medical image segmentation [C]//Computer Vision-ECCV 2022 Workshops. Cham: Springer Nature Switzerland, 2023: 205-218.

(编辑:薛泽玲)