

面向皮肤癌的双分支多尺度分割网络研究<sup>\*</sup>李 寅<sup>1</sup> 韩基泰<sup>2</sup> 张凯文<sup>1</sup> 于 敏<sup>1</sup> 刘小庆<sup>1</sup>

(1. 南京信息工程大学自动化学院 南京 210044; 2. 无锡学院自动化学院 无锡 214105)

**摘要:** 皮肤镜图像分割因病变形态复杂、边界模糊及噪声干扰, 面临诸多挑战。提出深度学习模型 MelanoFusionNet, 采用 ResNet50 与改进的 Mamba VSS 构成双分支编码器, 结合多尺度注意力融合解码器 (MSAFD), 整合局部与全局信息。解码器中的扩展内核分组门 (EKGG)、分层尺度感知注意力模块 (HSAM) 提升特征融合与边界精细度, 深度可分离上采样块 (DURB) 提高重建效率。模型以 Dice 与边界损失加权训练。在 ISIC2018 数据集上的实验显示, 其多项指标优于主流方法, 能为皮肤癌计算机辅助诊断提供可靠支持。

**关键词:** 皮肤病变分割; MelanoFusionNet; 双编码器模型; 多尺度注意力融合; 状态空间模型

**中图分类号:** TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Research on a dual-branch multi-scale segmentation  
network for skin cancer diagnosisLi Yin<sup>1</sup> Han Jitai<sup>2</sup> Zhang Kaiwen<sup>1</sup> Yu Min<sup>1</sup> Liu Xiaqing<sup>1</sup>

(1. School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, China;

2. School of Automation, Wuxi University, Wuxi 214105, China)

**Abstract:** Dermoscopic image segmentation faces significant challenges due to complex lesion morphology, blurred boundaries and noise interference. This study proposes a deep learning model named MelanoFusionNet, which employs a dual-branch encoder composed of ResNet50 and an improved Mamba VSS module. A multi-scale attention fusion decoder (MSAFD) is introduced to integrate both local and global information. Within the decoder, the extended kernel grouped gate (EKGG) and the hierarchical scale-aware attention module (HSAM) enhance feature fusion and boundary precision, while the depthwise upsample refinement block (DURB) improves reconstruction efficiency. The model is trained with a weighted combination of Dice and boundary losses. Experiments on the ISIC2018 dataset demonstrate that MelanoFusionNet outperforms mainstream methods across multiple metrics and provides reliable support for computer-aided skin cancer diagnosis.

**Keywords:** skin lesion segmentation; MelanoFusionNet; dual-branch encoder; multi-scale attention fusion; state space model

## 0 引 言

黑色素瘤作为最具致命性的皮肤癌类型, 其早期实现精准诊断对于患者的生存预后具有至关重要的意义。皮肤镜图像分割技术致力于精确地勾勒出病变区域的边界, 是计算机辅助诊断系统中不可或缺且极为关键的一个步骤<sup>[1]</sup>。然而, 由于黑色素瘤在大小、形状、颜色和纹理上存在显著差异, 且其边界常常模糊、与周围健康组织的对比度低, 该任务仍然充满挑战。近年来, 深度学习的最新进展显著提升了医学图像分割的性能<sup>[2]</sup>。在这一领域中, 以 U-

Net<sup>[3]</sup> 及其变体为代表的卷积神经网络 (convolutional neural network, CNN)<sup>[4-7]</sup> 能有效提取局部层次化特征, 但受归纳偏置限制, 难以捕获长程依赖关系。视觉 Transformer (vision transformer, ViT)<sup>[8]</sup> 虽具备全局上下文建模能力, 却因计算复杂度高而难以在资源受限环境中应用。

为兼顾全局建模与局部细节表达, 状态空间模型 Mamba 以线性复杂度实现长序列依赖建模, 展现出新的应用潜力<sup>[9]</sup>。近年来, 以 Mamba 为代表的状态空间模型 (state space model, SSM) 凭借其线性计算复杂度与长序

列全局建模的平衡优势<sup>[10]</sup>,在计算机视觉领域受到广泛关注,并迅速被引入医学图像分析。在皮肤病变分割任务中,现有研究主要沿两个方向探索 Mamba 的潜力:1)将其与 CNN 进行混合架构设计,以融合局部细节与全局上下文;2)在小样本学习等特定范式中对其进行增强,以应对标注数据稀缺的挑战。在混合架构方面,研究主要致力于结合 Mamba 的全局建模能力与 CNN 的局部特征提取优势。Zhang 等<sup>[11]</sup>提出的 HCMUNet 采用双分支编码器,由 CNN 分支捕获局部细节,并行 Mamba 分支建立长程依赖,并通过特征融合模块整合信息,在医学超声图像分割中取得了优于纯 CNN 或 Transformer 模型的性能,验证了该策略的有效性。梁宵等<sup>[12]</sup>提出的空间通道双自编码器,通过分离空间维度的结构特征与通道维度的语义特征进行无监督异常检测,进一步证明了双分支结构在分维度特征学习中的优势,这种设计可有效避免单一分支对局部细节或全局上下文的遗漏。为应对数据稀缺问题,Guo 等<sup>[13]</sup>将 Mamba 与对抗性训练、双向扫描机制结合,增强了模型在少量标注样本下的泛化能力,为临床标注成本高的困境提供了解决方案。此外,成荣等<sup>[14]</sup>通过轻量化设计证明了 Mamba 在资源受限场景下的应用潜力,而基于 Mamba 的 3D 目标检测与时间序列预测方法,也为处理皮肤三维图像或时序随访数据提供了重要启发。

然而,单纯依赖全局特征容易忽略医学图像中关键的边界与纹理细节<sup>[15-16]</sup>。为有效弥补现有研究存在的这一不足,本研究提出了一种配备双分支编码器的网络架构 MelanoFusionNet。该模型的核心在于其独特的双

分支编码器设计,其中 ResNet50 分支<sup>[17]</sup>负责提取丰富的局部特征,Mamba 分支则用于捕获全面的全局上下文表示,二者协同工作,确保模型既能聚焦细节,又具备上下文感知能力。在解码器端,引入分层尺度感知注意力模块(hierarchical scale-aware attention module, HSAM)与扩展内核分组门(extended kernel grouped gate, EKGG),这两个模块有效细化并融合多尺度特征,同时抑制无关的背景噪声,最终显著提升边界分割的准确性。除此之外,本研究的另一个关键创新在于提出了特征融合策略。采用了一种分层融合过程,EKGG 作为一个智能的特征融合枢纽,其中来自编码器跳跃连接和解码器上采样路径的特征,通过使用大核分组卷积的门控注意力机制进行动态整合。这使得融合过程能够在更大的感受野内自适应地进行,确保 CNN 分支的高分辨率空间细节和 Mamba 分支的全局语义上下文在整个解码过程中得到最优利用。

## 1 方 法

### 1.1 整体结构

本研究提出的 MelanoFusionNet 是一款专为皮肤镜图像分割设计的深度学习模型,旨在解决病变形态复杂、边界模糊及背景噪声等问题。模型整体基于 U-Net 框架,通过双分支编码器与多尺度注意力融合解码器(multi-scale attention fusion decoder, MSAFD)的协同设计,实现了对局部细节与全局上下文的高效建模,并提升了边界分割的精度与鲁棒性。模型的整体结构如图 1 所示。

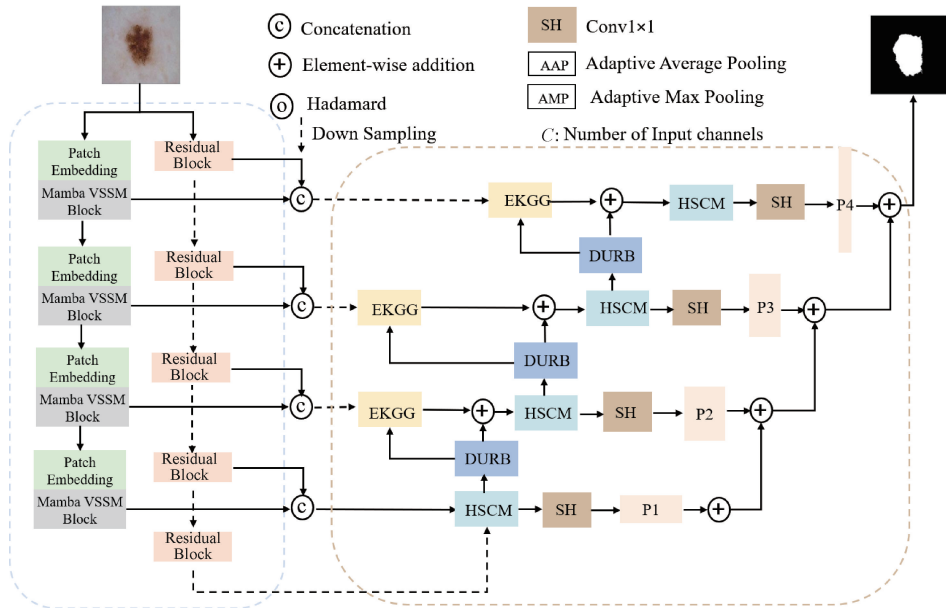


图 1 MelanoFusionNet 总体架构

Fig. 1 MelanoFusionNet overall architecture

在编码器部分,MelanoFusionNet 采用双分支并行架构。其中,ResNet50 分支专注于提取病灶区域的细粒度边

缘轮廓、纹理模式和局部对比度变化局部特征;其残差连接机制能够有效缓解深层网络训练中的梯度消失问题,并

保证关键细节特征在网络中得以传递与保留。另一分支为改进型 Mamba 视觉状态空间模型 (vision state space block, VSS) 分支,通过集成空洞空间金字塔池化 (atrous spatial pyramid pooling, ASPP) 与二维选择性扫描 (selective scan 2D, SS2D) 机制,在保持线性计算复杂度的同时实现长程依赖建模,能够捕获病灶与周围组织的全局语义与结构信息。两条分支优势互补,使模型在获取精细局部细节的同时兼顾全局上下文理解。

在解码器部分, MelanoFusionNet 设计了 MSAFD 来逐级重建高分辨率分割结果。其核心模块包括 EKGG 利用大核分组卷积的门控机制动态融合来自编码器的跳跃连接与解码器上采样特征,从而突出关键区域并抑制背景噪声; HSAM 结合通道注意力、空间注意力与多尺度卷积块,实现多尺度上下文的充分利用并细化病灶边界; 深度可分离上采样优化块 (depthwise upsample refinement block, DURB) 在分辨率恢复过程中保持细节结构并降低计算成本; 以及分段头 (segmentation head, SH) 将各级解码特征映射为分割预测并通过多级输出融合进一步提升边界精细度和整体分割质量。

整体而言, MelanoFusionNet 通过双分支特征提取与多尺度注意力解码的协同设计,在捕获病灶细节信息和保持全局结构一致性之间实现了良好的平衡,从而有效提高了皮肤病变分割任务的准确性与鲁棒性,为临床辅助诊断提供了可靠支持。

## 1.2 编码器部分

### 1) Resnet50 分支

在 MelanoFusionNet 的编码器中, ResNet50 分支主要用于提取皮肤病变的局部细粒度特征,包括病灶边缘、纹理模式及局部对比度等信息。得益于其深层残差结构,该分支能够在保证较深网络深度的同时有效缓解梯度消失问题,并保留关键细节特征。通过多层卷积和残差单元的逐级提取, ResNet50 分支为解码器提供了高分辨率的局部信息支撑,有助于实现病灶区域的精确分割与边界刻画。

### 2) Mamba VSS

VSS 是 Mamba 架构中用于视觉任务的核心模块,其原理为基于 SSM 的离散化理论,通过 SS2D 将二维图像沿多个方向展开为一维序列,并利用选择性机制动态调整参数,以线性时间复杂度高效捕捉长程依赖关系,其原理如图 2 所示。该模块采用双分支并行架构,结合深度卷积与层归一化,通过特征融合与残差连接保留局部细节、增强模型稳定性,最终输出结构化特征以支持复杂视觉任务。传统 VSS Block 在捕捉局部细节与全局上下文关系的平衡上存在不足,尤其是对长距离依赖关系的建模效率较低,其模块如图 3(a) 所示。

为此,本研究在 VSS 模块中引入 ASPP 结构,皮肤病变在尺度、纹理复杂度及边界清晰度方面具有高度不确定性,单一尺度卷积难以同时捕获局部细节与长距离上下文

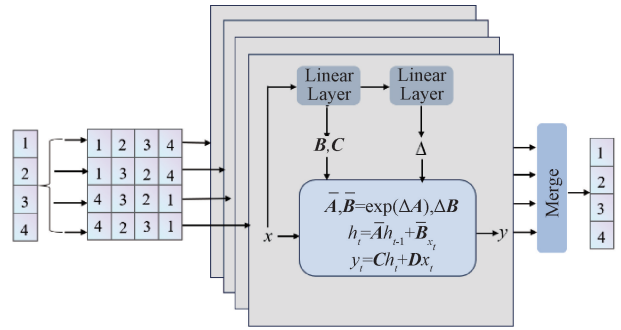


图 2 SS2D 计算流程

Fig. 2 SS2D computational flowchart

依赖。ASPP 通过不同膨胀率的空洞卷积并行提取多尺度语义特征,能够在不显著增加计算成本的前提下扩大有效感受野,使模块具备对大范围病灶区域轮廓与细粒度结构同时建模的能力。与金字塔池化模块 (pyramid pooling module, PPM)<sup>[18]</sup>、Inception<sup>[19]</sup> 以及感受野模块 (receptive field block, RFB)<sup>[20]</sup> 等多尺度模块相比, ASPP 在密集预测任务中表现更为稳定,且结构更容易与 Mamba 的序列扫描机制结合。

此外,由于 ASPP 含多分支空洞卷积,若全部采用标准卷积会引入较大的参数量与计算量,这与 Mamba 模块强调“线性复杂度、高效建模”的设计理念相矛盾。为此,本文将 ASPP 中的普通卷积统一替换为深度可分离卷积,既保持了多扩张率卷积的多尺度感受野优势,又将计算量显著降低,实现对全局语义与局部空间结构的高效融合。上述设计使改进后的 Mamba VSS 模块在理论上更符合医学图像的多尺度特性需求,同时在实际推理阶段具有更高的效率与可部署性。

改进后的模块如图 3(b) 所示,输入数据被分流至两个并行分支。第 1 个分支中,输入先经线性层处理,再引入 ASPP,通过不同扩张率的空洞卷积增强多尺度特征提取能力。随后应用 SiLu 激活函数。第 2 个分支则先通过线性层转换,再经深度可分离卷积提取局部特征,并同样使用 SiLu 激活函数。之后输入被送入 SS2D 模块进行全局特征建模。两个分支的输出通过层归一化 LayerNorm 进行标准化处理。随后将第 1 分支的输出与第 2 分支特征逐元素相乘完成融合,最终通过线性层混合特征,并与原始输入的残差连接相加,形成改进型 Mamba VSS 模块的最终输出。为降低模块的计算复杂度和参数量,将 ASPP 中的普通卷积替换为深度可分离卷积。从而在基本保持其多尺度感受野与模型性能的前提下,大幅提升运行效率,使模型更适合计算资源受限的部署环境,其结构如图 4 所示。

## 1.3 解码器部分

本研究提出的解码器采用 MSAFD 架构,用于充分利用编码器各阶段提取的多尺度特征,实现对复杂病灶边缘

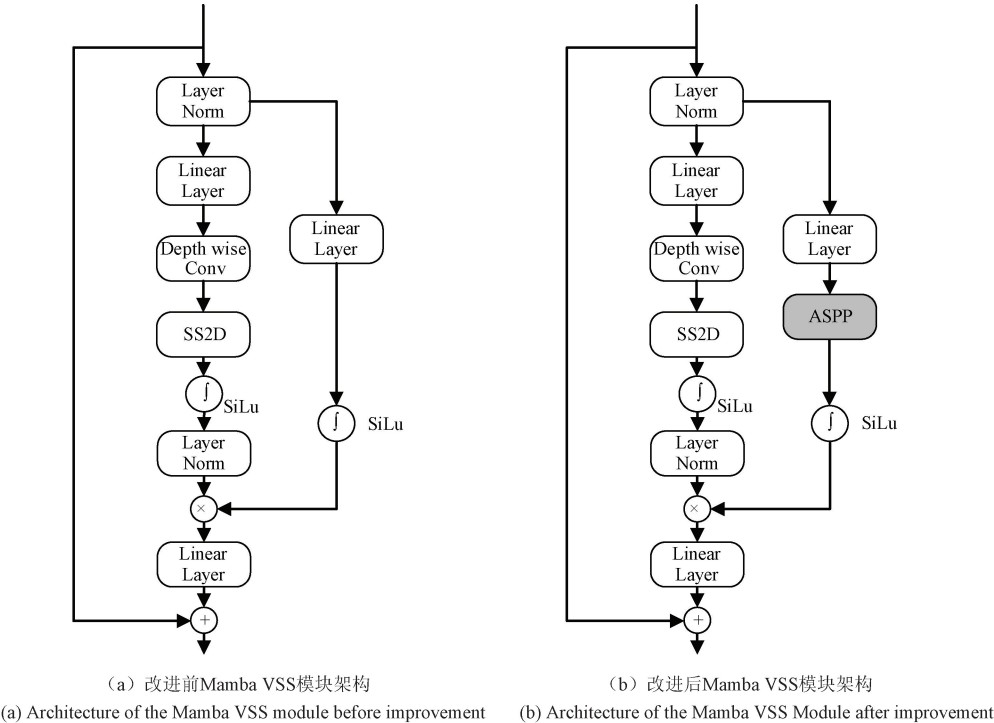


图 3 VSS 架构图与改进后的 VSSB 架构

Fig. 3 VSS Architecture diagram and the improved VSSB architecture

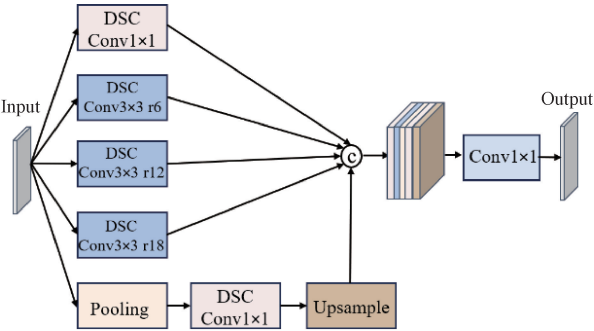


图 4 ASPP 模块架构

Fig. 4 ASPP module architecture

与低对比度区域的精细分割。解码器由 4 个关键模块构成:HSAM、EKGG、DURB 以及 SH。在整体工作流程上,解码器首先通过 HSAM 模块对编码器不同层次的特征进行提炼与增强;再借助 EKGG 实现与跳跃连接特征的高效融合;随后,DURB 模块逐步上采样并优化特征图,使其与对应分辨率的特征相匹配;最后,SH 将各阶段解码后的特征映射为分割预测,并通过多级输出的逐像素融合获得最终分割结果。该解码器结构在保持计算效率的同时,显著提升了模型在细粒度病灶边界识别及多尺度目标分割中的表现,为医学图像分割提供了更鲁棒的特征恢复与重建能力。

1) EKGG

EKGG 旨在通过引入更高效的注意力门控机制来增

强跳跃连接与解码器特征的融合。其核心思想是利用高层特征  $g$  作为门控信号,选择性地突出低层特征  $x$  中与目标相关的区域,同时抑制背景干扰。与 Attention U-Net<sup>[21]</sup>采用的  $1 \times 1$  卷积不同,EKGG 在门控函数  $q_{att}(\cdot)$ 中分别对  $g$  和  $x$  施加  $3 \times 3$  分组卷积  $GC_g(\cdot)$ 和  $GC_x(\cdot)$ ,以较低的计算成本捕获更广的空间上下文信息。两者的卷积结果经批量归一化  $BN(\cdot)$ <sup>[22]</sup>处理后相加,并通过 ReLU 激活函数  $R(\cdot)$ 进行非线性变换,随后再施加  $1 \times 1$  卷积  $C(\cdot)$ 与  $BN(\cdot)$ 得到单通道特征图,最终通过 Sigmoid 函数  $\sigma(\cdot)$ 输出注意力系数。其流程图如图 5 所示,该注意力系数通过逐元素乘法对输入特征  $x$  进行加权,实现公式化表达:

$$q_{att}(g, x) = \sigma(BN(C(R(BN(GC_g(g)) + BN(GC_x(x)))))$$

(1)

$$EKGG(g, x) = q_{att}(g, x) \odot x$$

(2)

其中,  $\odot$  表示 Hadamard 乘积。通过上述门控机制,高层语义信息得以引导低层特征流动,强化了模型对关键区域的响应并抑制冗余特征。使解码器能够在保持较低计算量的前提下捕获更丰富的空间上下文,有助于提升医学图像分割中病灶区域的边缘识别与细节刻画。

2) HSAM

HSAM 旨在对来自编码器不同阶段的特征图进行细粒度增强与多尺度融合,以更好地提取复杂病灶区域的边界与上下文信息。该模块由通道注意力块(channel attention block, CAB)、空间注意力块(spatial attention

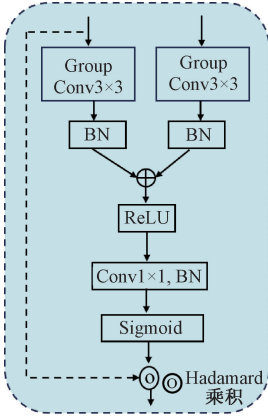


图 5 EKG 架构

Fig. 5 EKG architecture

block, SAB)<sup>[23]</sup> 和多尺度卷积块 (multi-scale convolution module, MCM) 组成, 三者协同作用以实现特征的通道选择、空间聚焦及多尺度上下文建模, 流程如图 6(a) 所示。其公式如下:

$$\text{HSAM}(\mathbf{x}) = \text{MCM}(\text{SAB}(\text{CAB}(\mathbf{x}))) \quad (3)$$

为进一步提高多尺度信息表达, HSAM 引入了基于 MobileNetV2 反向残差结构<sup>[24]</sup> 改进的 MCM。在 MCM 中, 首先使用逐点卷积  $\text{PWC}_1(\cdot)$  对输入特征  $\mathbf{x}$  进行通道扩展, 扩展因子设置为 2, 经批量归一化  $\text{BN}(\cdot)$  和  $\text{ReLU}_6(\cdot)$  后, 送入多尺度深度卷积单元  $\text{MSDC}(\cdot)$ , 以提取跨尺度上下文信息。考虑到深度卷积对通道依赖的忽略, 模型进一步引入通道混洗操作<sup>[25]</sup> 以促进跨组特征交互; 最后, 通过另一逐点卷积  $\text{PWC}_2(\cdot)$  和  $\text{BN}(\cdot)$  将通道数恢复到原始维度, 并实现通道间依赖建模。MCM 的完整流程如图 6(b) 所示, 其数学表达如式(4)所示。

$$\text{MCM}(\mathbf{x}) = \text{BN}(\text{PWC}_2(\text{CS}(\text{MSDC}(\text{R}_6(\text{BN}(\text{PWC}_1(\mathbf{x})))))) \quad (4)$$

其中,  $\text{MSDC}(\cdot)$  由不同卷积核大小  $k_s$  的并行深度可分离卷积组成, 其流程如图 6(c) 所示。

$$\text{MSDC}(\mathbf{x}) = \sum_{k_s \in \text{KS}} \text{DWCB}_{k_s}(\mathbf{x}) \quad (5)$$

并通过残差连接实现逐级正则化:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \text{MSDC}(\mathbf{x}_t) \quad (6)$$

SAB 流程如图 6(d) 所示。模型通过沿通道维度的最大池化  $\text{Ch}_{\max}(\cdot)$  与平均池化  $\text{Ch}_{\text{avg}}(\cdot)$  提取局部响应特征, 并经  $7 \times 7$  卷积融合局部上下文后生成空间注意力权重, 再由 Sigmoid 函数  $\sigma(\cdot)$  输出, 最终通过 Hadamard 乘积对输入特征  $\mathbf{x}$  进行加权:

$$\text{SAB}(\mathbf{x}) = \sigma(\text{LKC}([\text{Ch}_{\max}(\mathbf{x}), \text{Ch}_{\text{avg}}(\mathbf{x})])) \otimes \mathbf{x} \quad (7)$$

CAB 旨在为不同通道分配差异化的重要性权重, 从而突出关键特征并抑制冗余信息, 其流程如图 6(e) 所示。CAB 首先在空间维度上分别施加自适应最大池化  $P_m(\cdot)$  和自适应平均池化  $P_a(\cdot)$ , 以提取每 4 个通道的全局统计特征<sup>[26]</sup>。随后, 对于池化得到的特征图, 依次通过逐点卷

积  $C_1(\cdot)$  和 ReLU 激活函数  $R(\cdot)$  将通道数按比例  $r=1/16$  进行压缩, 以降低计算复杂度并提炼关键信息。接着, 再利用另一个逐点卷积  $C_2(\cdot)$  将通道数恢复至原始维度。两条池化分支的输出特征相加后, 经 Sigmoid 函数  $\sigma(\cdot)$  映射得到通道注意力权重, 最后通过 Hadamard 乘积将其与输入特征  $\mathbf{x}$  相融合, 实现对重要通道的自适应增强, 其公式如式(8)所示。

$$\text{CAB}(\mathbf{x}) = \sigma(C_2(R(C_1(P_m(\mathbf{x})))) + C_2(R(C_1(P_a(\mathbf{x})))) \otimes \mathbf{x} \quad (8)$$

CAB 则通过自适应加权的方式强化对关键通道的响应。通过上述模块的分层协作, HSAM 能够有效增强模型对多尺度上下文的感知能力并突出目标病灶区域, 为解码器提供更具判别力的特征表示, 从而提高医学图像分割中对模糊边界与低对比度目标的识别性能。

### 3) DURB

DURB 旨在在解码阶段实现高效的逐级上采样与特征增强, 从而使当前层特征与来自编码器的跳跃连接特征在分辨率和语义上保持一致, 其架构如图 7 所示。

DURB 首先利用上采样算子  $\text{Up}(\cdot)$  将输入特征  $\mathbf{x}_i$  的空间分辨率按比例因子  $s=2$  进行扩展, 以恢复更高分辨率的特征图。随后, 使用  $3 \times 3$  深度可分离卷积  $\text{DWC}(\cdot)$  对上采样后的特征进行细化, 通过捕获局部空间上下文来补偿上采样带来的信息损失; 接着, 将其依次送入批量归一化  $\text{BN}(\cdot)$  与 ReLU 激活函数  $R(\cdot)$  以提升特征稳定性和非线性表达能力; 最后通过  $1 \times 1$  卷积  $C_{1 \times 1}(\cdot)$  调整通道数, 使其与下一阶段解码器或跳跃连接特征保持一致。其计算过程可表示为:

$$\text{DURB}(\mathbf{x}) = C_{1 \times 1}(\text{ReLU}(\text{BN}(\text{DWC}(\text{Up}(\mathbf{x})))) \quad (9)$$

与传统上采样方式相比, DURB 通过在上采样后引入轻量化的深度可分离卷积, 有效地在保持计算效率的同时增强了特征细节与局部结构信息, 缓解了直接上采样导致的模糊与边缘退化问题。在整个解码器结构中, DURB 逐级对不同分辨率特征进行空间尺寸恢复并进行优化, 为后续与跳跃连接特征的融合奠定基础, 有助于提升模型在医学图像分割任务中对病灶细粒度边界和结构的重建能力。

### 4) SH

SH 作为解码器的最终输出模块, 旨在将经多级特征增强与上采样后的解码特征图映射到具体的分割结果。SH 主要负责将不同阶段的精炼特征转化为与目标任务对应的类别概率图, 实现像素级预测。在本模型中, 对于来自解码器第  $i$  阶段的特征图  $\mathbf{F}_i$ , 其通道数为  $ch_i$ , SH 通过一个  $1 \times 1$  卷积  $\text{Conv}_{1 \times 1}(\cdot)$  将高维特征压缩或变换为目标类别数  $C$  (对于二值分割任务则为单通道)。该过程的数学形式可表示为:

$$\mathbf{S}_i = \text{SH}(\mathbf{F}_i) = \text{Conv}_{1 \times 1}(\mathbf{F}_i), \mathbf{S}_i \in \mathbb{R}^{C \times H_i \times W_i} \quad (10)$$

其中,  $H_i$  和  $W_i$  分别为第  $i$  阶段特征图的高度和宽

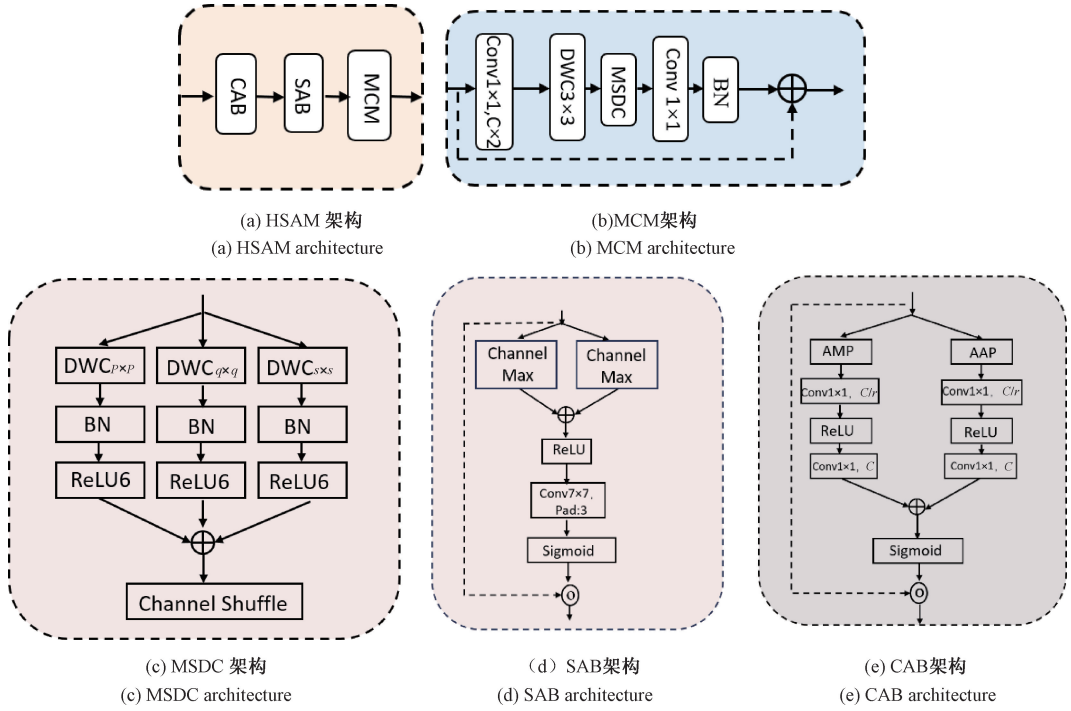


图 6 HSAM 架构以及 HSAM 中各个子模块架构

Fig. 6 HSAM architecture and the architecture of each sub-module in HSAM

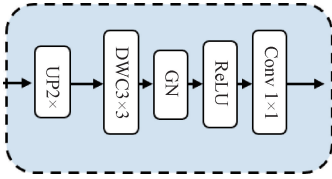


图 7 DURB 架构

Fig. 7 DURB architecture

度。在整体推理过程中,解码器的 4 个阶段均通过 SH 模块独立生成分割预测图 $\{S_1, S_2, S_3, S_4\}$ ,这些预测结果在空间维度上进行逐像素融合,从而得到最终的高分辨率分割输出:

$$S_{\text{final}} = \sum_{i=1}^4 S_i \quad (11)$$

通过这种多阶段输出与融合机制,SH 不仅能够利用深层语义特征提升整体分割精度,还能保留浅层特征提供的边缘与细节信息,实现对复杂病灶边界及细粒度结构的更好刻画。同时,1×1 卷积的使用保证了模块的轻量化特性,避免引入过多参数开销,适合在多尺度解码结构中高效部署。

#### 1.4 损失函数

本工作采用 Dice 损失与边界损失 BoundaryLoss 的组合,以协同优化分割区域的整体一致性和边界精确度。Dice 损失函数定义如下:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N i = 1^N p_i g_i}{\sum_{i=1}^N p_i^2 + \sum_{i=1}^N g_i^2} \quad (12)$$

其中,  $p_i \in [0, 1]$  和  $g_i \in [0, 1]$  分别表示像素  $i$  的预测概率和真实标签。该损失通过计算两个集合的交叠度,有效缓解了前景与背景间的类别不平衡问题,但其区域优化的特性对边界误分类不够敏感。

为弥补 Dice 损失在边界精细度上的不足,本研究引入边界损失,其公式为:

$$\mathcal{L}_{\text{Boundary}} = \sum_{i \in \partial G} \partial G \phi_G(i) \cdot p_i + \sum_{i \in \partial P} \phi_P(i) \cdot (1 - p_i) \quad (13)$$

其中,  $\partial G$  和  $\partial P$  分别表示真实轮廓和预测轮廓,  $\phi_G(i)$  是像素  $i$  到真实边界的距离符号映射。该函数直接惩罚预测边界与真实边界在空间上的偏离。最终的组合损失函数为二者的加权和:  $\mathcal{L}_{\text{Total}} = \alpha \mathcal{L}_{\text{Dice}} + \beta \mathcal{L}_{\text{Boundary}}$ 。这种设计使模型在 Dice 损失的驱动下保证区域完整性的同时,受边界损失的引导专注于学习轮廓的几何细节,从而在整体分割精度和边界贴合度上达到更优的平衡。

## 2 实验设计与结果分析

### 2.1 数据集介绍

本研究选用 ISIC2018 数据集作为模型训练与评估的主要数据来源。该数据集由国际皮肤影像协作组织 (international skin imaging collaboration, ISIC) 公开发布,是皮肤病变分析领域最具权威性和广泛应用的标准基准之一。数据集共包含 2 594 张高分辨率皮肤镜图像,每幅图像均由经验丰富的皮肤科专家进行精准标注,并附有对应的病变区域分割掩码,能够为算法提供可靠的监督信

号。图像来源于多个临床医疗中心,涵盖了不同人群、成像设备和病变类型,因此数据集在图像分布和病灶表现上具有较高的多样性与代表性。

## 2.2 实验条件设置

本节实验在 Windows 11 操作系统下进行,硬件环境采用 NVIDIA GeForce RTX 3090 GPU,显存为 24 GB,以确保在高分辨率皮肤镜图像分割任务中的训练效率与计算稳定性。实验基于 PyTorch 2.1.0 深度学习框架实现,使用 Python3.10 作为编程语言,并在 PyCharm 集成开发环境中完成代码编写与调试。CUDA 版本为 11.8,用于加速 GPU 上的并行计算。

在训练策略方面,批处理大小(batch size)设为 8,初始学习率设置为 0.000 2,并采用学习率衰减策略:每完成 20 个训练轮次(epoch),学习率按原值的 1/10 进行更新,以促进模型在后期收敛。总 epoch 为 300。所有卷积层权重使用 Kaiming 初始化,以改善深层网络的梯度传播并加快收敛速度。这些配置在保证训练稳定性和高效性的同时,有助于模型在复杂皮肤镜图像分割任务中获得更佳性能。

## 2.3 双分支特征可视化与协同机制分析

为直观验证所提出双分支编码器在特征提取层面的功能分工与互补性,本研究基于 Grad-CAM++ 可解释性方法,对 ResNet50 分支、改进 MambaVSS 分支以及二者融合后的编码器输出特征进行了系统可视化分析,其结果如图 8 所示。该分析旨在从空间关注模式的角度揭示不同编码策略在局部细节建模与全局语义感知方面的差异,为双分支结构设计的合理性提供直观证据。

图 8(b)为 ResNet50 单分支编码器的特征响应热力图。可以观察到,该分支的高响应区域主要集中于病灶边缘、内部纹理突变区域以及毛发、血管等高频结构附近。这一现象表明,基于卷积神经网络的 ResNet50 受其局部感受野与层级卷积结构影响,更倾向于捕获局部空间变化显著的区域,对边界轮廓与细粒度纹理信息具有较强敏感性。然而,其关注范围相对局限,对病灶整体结构的覆盖不够完整,在病灶内部低对比度区域的激活强度较弱,容易导致全局语义理解不足。

图 8(c)为改进后 MambaVSS 单分支编码器的特征响应结果。与 ResNet50 分支相比,该分支的热力图呈现出更大范围、更均匀的激活分布,高响应区域不仅覆盖病灶主体区域,还延伸至与病变语义相关的周围组织。这得益于 MambaVSS 中引入的选择性扫描状态空间建模机制以及 ASPP 模块对多尺度上下文的增强,使模型能够在保持线性计算复杂度的同时有效捕获长程依赖关系与全局结构信息。然而,由于其建模重心偏向全局语义,该分支在病灶边缘处的激活相对平滑,对精细轮廓和局部纹理变化的刻画能力相对有限。

图 8(d)为 ResNet50 与改进 MambaVSS 组成的双分

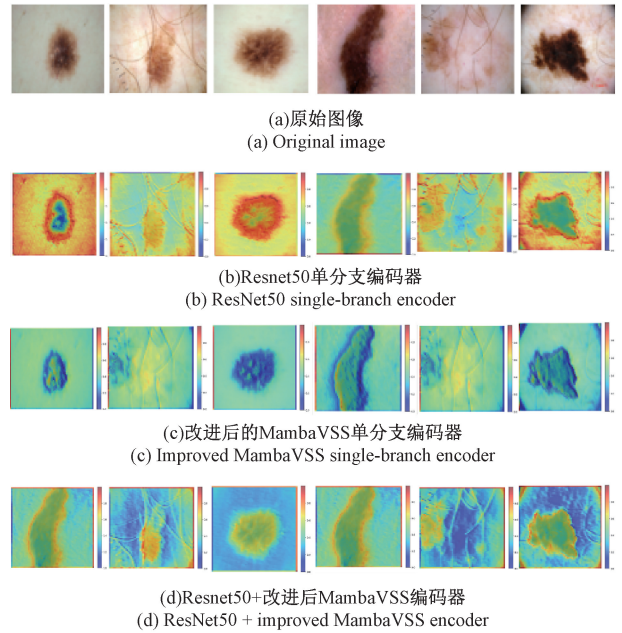


图 8 不同编码器组成的热力图组

Fig. 8 Heatmap sets of different encoder combinations

支编码器融合特征的可视化结果。可以看出,融合后的热力图同时具备清晰的病灶边缘响应与完整的病灶区域覆盖:一方面,病灶轮廓及细节结构被显著强化;另一方面,病灶内部区域及其整体形态获得了连续且稳定的高响应。这表明解码阶段在 EKG 与 HSAM 等模块的作用下,成功整合了 CNN 分支提供的局部高频细节信息与 Mamba 分支提供的全局上下文语义,实现了特征层面的有效互补。

综合图 8(b)~(d)的对比分析可以得出结论:ResNet50 分支在局部细节与边界建模方面具有优势,改进 MambaVSS 分支在全局结构与多尺度上下文感知方面表现突出,而双分支融合能够同时兼顾二者的优势。该可视化结果从特征响应层面直观验证了 MelanoFusionNet 双分支编码器设计的合理性,也为其在复杂皮肤病变分割任务中取得优异性能提供了有力的解释支撑。

## 2.4 消融实验

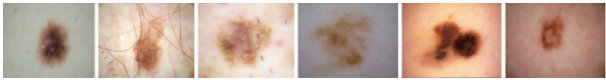
为验证所提出模型中各关键模块的有效性,本研究在 ISIC2018 数据集上设计了系列消融实验。实验从基础 U-Net 模型出发,逐步引入改进组件,包括多尺度注意力融合解码器 MSAFD、组合损失函数、双分支编码器(ResNet50+改进 MambaVSS),并评估不同组合下模型在特异性 Sp、灵敏度 Se、平均交并比 mIoU、Dice 系数及边界距离 HD95 指标上的表现。其结果如表 1 所示,可视化结果如图 9 所示。

结果显示,引入 MSAFD 解码器后即 Model\_2,模型在 Se 和 Dice 上明显提升,同时 HD95 显著下降,表明解码器的多尺度卷积与注意力机制增强了对病灶特征的捕捉能

表 1 在 ISIC2018 数据集上的消融实验量化评估结果

Table 1 Quantitative evaluation results of the ablation study on the ISIC2018 dataset

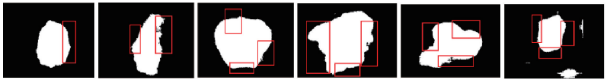
| 模型名称    | 模块组合      |       |      | 评估指标         |              |              |       |               |
|---------|-----------|-------|------|--------------|--------------|--------------|-------|---------------|
|         | VSS Block | MSAFD | 损失函数 | Sp           | Se           | mIoU         | Dice  | HD95/pixel    |
| Model_1 | ×         | ×     | ×    | 0.939        | 0.898        | 0.863        | 0.831 | 29.903        |
| Model_2 | ×         | ✓     | ×    | 0.934        | 0.944        | 0.854        | 0.841 | 22.403        |
| Model_3 | ×         | ✓     | ✓    | 0.826        | 0.952        | 0.848        | 0.830 | 27.538        |
| Model_4 | ✓         | ×     | ×    | <b>0.960</b> | 0.882        | 0.876        | 0.843 | 22.407        |
| Model_5 | ✓         | ✓     | ×    | 0.944        | 0.901        | 0.864        | 0.836 | 22.178        |
| Model_6 | ✓         | ×     | ✓    | 0.957        | 0.938        | 0.866        | 0.832 | 21.853        |
| Model_7 | ✓         | ✓     | ✓    | <b>0.950</b> | <b>0.949</b> | <b>0.867</b> | 0.836 | <b>19.455</b> |



(a) 原始图片  
(a) Original image



(b) 原始标签  
(b) Original label



(c) Model\_1



(d) Model\_2



(e) Model\_3



(f) Model\_4



(g) Model\_5



(h) Model\_6



(i) Model\_7

图 9 在 ISIC2018 数据集上的消融实验可视化分割结果  
Fig. 9 Visualized segmentation results of the ablation study on the ISIC2018 dataset

数即 Model\_3 进一步优化了区域一致性和边界贴合度。将编码器替换为双分支结构 (Model\_4 和 Model\_5) 后, 模型在 Sp、mIoU 和 Dice 上继续提升, 其中 MambaVSS 分支通过 ASPP 和全局上下文建模增强了对复杂病灶结构的感知, 而 ResNet50 分支保留了关键局部细节, 使得分割边界更平滑且与真实标签更接近。改进后的 MambaVSS 与损失函数组合 (Model\_6) 在大范围病灶区域上表现较好, 但在边缘细节、小结构、低对比区域上, 分割结果不够精细。最终整合全部改进模块 (Model\_7) 后, 模型在 ISIC2018 数据集上达到 Sp0.950、Se0.949、mIoU0.867、Dice0.851, 并将 HD95 降低至 19.45 像素, 较基础模型显著改善。

消融实验结果证明, MelanoFusionNet 中 MSAFD 解码器、EKGG 和 HSAM 模块对分割边界的精细化和细粒度病灶识别起到关键作用; 双分支编码器有助于实现局部与全局特征的有效融合; 而组合损失函数则在提升整体分割精度的同时优化了边界定位。这些组件的协同作用共同提升了模型在复杂皮肤病变分割任务中的鲁棒性和临床应用价值。

2.5 对比试验

为全面评估所提出 MelanoFusionNet 的分割性能, 本研究在 ISIC2018 数据集上与多种主流皮肤病变分割模型进行了对比实验。对比模型包括 UNet、UNet++<sup>[27]</sup>、VM-UNet<sup>[28]</sup>、SwinDeepLab<sup>[29]</sup>、Simpleunet、Scalefusionnet<sup>[30]</sup> 与 Ac\_Mambaseg<sup>[31]</sup>。这些模型分别代表了基于卷积网络的传统方法、引入注意力机制的改进方案及结合 Transformer 或状态空间模型的先进结构, 能够全面反映不同特征提取策略在皮肤病变分割任务中的性能水平。各对比模型在 ISIC2018 数据集上的定量评估结果如表 2 所示。

在相同数据集划分、预处理流程与训练策略下, 对比实验分别计算了各模型在 Sp、Se、mIoU、Dice 以及 95% Hausdorff 距离 HD95 等指标上的表现。从表 2 可以看出, MelanoFusionNet 在多项指标上均取得最优或次优成绩, 尤其在边界相关指标上表现突出; 其在 mIoU 和 Dice

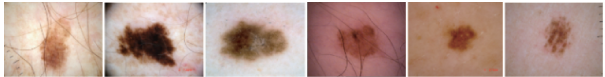
表 2  不同网络在 ISIC2018 数据集上的实验量化评估结果

Table 2  Quantitative evaluation results of experiments with different networks on the ISIC2018 dataset

| 模型             | Sp           | Se           | mIoU         | Dice         | HD95/pixel    |
|----------------|--------------|--------------|--------------|--------------|---------------|
| UNet           | 0.939        | 0.898        | 0.863        | 0.831        | 29.903        |
| UNet++         | 0.946        | 0.927        | 0.877        | 0.860        | 22.627        |
| VMUNet         | 0.964        | 0.920        | <b>0.882</b> | 0.855        | 20.651        |
| SwinDeepLab    | 0.923        | 0.922        | 0.855        | 0.832        | 30.641        |
| Simpleunet     | 0.897        | 0.901        | 0.819        | 0.786        | 30.804        |
| ScaleFusionNet | 0.955        | 0.883        | 0.870        | 0.837        | 22.417        |
| AC-MambaSeg    | <b>0.966</b> | 0.882        | 0.876        | 0.843        | 22.407        |
| 本模型            | 0.958        | <b>0.949</b> | 0.879        | <b>0.871</b> | <b>19.455</b> |

系数上分别达到 0.879 和 0.871,并将 HD95 降低至 19.45 pixel,显著优于其他对比模型,说明所提出的双分支编码器与多尺度注意力解码器在边界刻画和区域一致性方面具有更强的表达能力。

为进一步直观比较不同模型的分割效果,本文选取具有代表性的测试样本,对各对比方法的分割结果进行了可视化分析,其结果如图 10 所示。可以观察到,传统 CNN 模型在病灶边界模糊或低对比度区域容易出现边界不完整或过分割现象;引入 Transformer 或 Mamba 结构的模型在整体轮廓建模方面有所改善,但在细粒度边缘处仍存在一定偏差。相比之下,MelanoFusionNet 在图 10 中呈现出更加平滑、连贯且贴合真实标签的分割轮廓,在复杂形态、弱对比病灶区域中有效减少了漏检与伪分割问题。



(a) 原始图像  
(a) Original image



(b) 原始标签  
(b) Original label



(c) UNet



(d) UNet++



(e) SwinDeepLab



(f) Simpleunet



(g) ScaleFusionNet



(h) AC-MambaSeg



(i) MelanoFusionNet

图 10  不同网络的可视化分割结果

Fig. 10  Visualized segmentation results of different networks

综合表 2 的定量评估结果与图 10 的定性对比分析可以得出 MelanoFusionNet 在分割精度、边界刻画能力以及模型鲁棒性方面均优于现有对比方法,验证了所提出双分支多尺度分割网络在皮肤病变分割任务中的有效性和临床应用潜力。

3  结    论

本研究针对皮肤镜图像分割中存在的病变形态多样、边界模糊及噪声干扰等挑战,提出了一种创新性的深度学习模型 MelanoFusionNet。该模型通过构建双分支编码器与 MSAFD 的协同架构,有效解决了传统方法在处理复杂医学图像时的多尺度变异和边界模糊问题。

本模型的核心创新——双分支编码器结构被证实有效。ResNet50 分支确保了局部细节特征的保留,而改进的集成 ASPP 的 MambaVSS 分支显著增强了对病变全局上下文和长距离依赖关系的建模能力,两者的协同工作为高精度分割奠定了坚实基础。

解码器中引入的 MSAFD,特别是 EKGG 与 HSAM,在细化特征和精确边界定位方面发挥了关键作用,可视化结果直观显示了其在生成光滑、准确病灶轮廓上的优势。

Dice 损失与边界损失的组合有效协同优化了分割区域的整体一致性和边界精确度,在 Dice 和 HD95 定量指标

和定性结果上均取得了良好平衡。

在 ISIC2018 数据集上的实验结果表明, MelanoFusionNet 在 Sp、Se、mIoU 和 Dice 系数上分别达到了 0.958、0.949、0.879 和 0.871, 同时 HD95 降低至 19.45 pixel, 显著优于现有基于 CNN、Transformer 及其他混合方法的性能。

综上所述, MelanoFusionNet 通过其独特的双分支编码器与多尺度注意力融合解码器设计, 在皮肤镜图像分割任务中取得了卓越的性能, 展现了在医学图像处理领域的广泛应用前景和重要临床价值。

## 参考文献

- [1] Hasan M K, Dahal L, Samarakoon P N, et al. DSNet: Automatic dermoscopic skin lesion segmentation[J]. Computers in Biology and Medicine, 2020, 120: 103738.
- [2] Bi L, Kim J, Ahn E, et al. Step-wise integration of deep class-specific learning for dermoscopic image segmentation[J]. Pattern Recognition, 2019, 85: 78-89.
- [3] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-assisted Intervention, 2015: 234-241.
- [4] Zhong L W, Li T S, Cui M, et al. DSU-Net: Dual-stage U-Net based on CNN and Transformer for skin lesion segmentation[J]. Biomedical Signal Processing and Control, 2025, 100: 107090.
- [5] Jebin B M, Shyla S I, Bel K N S, et al. Brain tumor detection and classification using U-Net and CNN with brain texture pattern analysis[J]. Biomedical Signal Processing and Control, 2025, 110: 108156.
- [6] Ferdinandus F X, Santoso J, Setiawan E I, et al. 3D visualization for lung COVID-19 infection based on U-Net CNN segmentation[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-11.
- [7] 周宝康, 曹爽, 高洪涌, 等. 基于双编码器的鼻中隔医学图像分割模型[J]. 电子测量技术, 2025, 48(15): 150-158.
- Zhou Baokang, Cao Shuang, Gao Hongyong, et al. A medical image segmentation model for nasal septum based on dual encoder[J]. Electronic Measurement Technology, 2025, 48(15): 150-158.
- [8] Naas M, Mzoughi H, Njeh I, et al. An explainable AI for breast cancer classification using vision Transformer(ViT)[J]. Biomedical Signal Processing and Control, 2025, 108: 108011.
- [9] Gu A, Dao T. Mamba: Linear-time sequence modeling with selective state spaces[PP/OL]. V2. (2024-05-31) [2025-11-14]. <https://doi.org/10.48550/arXiv.2312.00752>.
- [10] Fang Z Y, Liu J J, Shi K X, et al. When Mamba meets xLSTM: An efficient and precise method with the xLSTM-VMUNet model for skin lesion segmentation[PP/OL]. V1. (2024-11-14) [2025-11-14]. <https://doi.org/10.48550/arXiv.2411.09363>.
- [11] Zhang C, Zhou X Y, Cui Y J, et al. HCMU-Net: A hybrid CNN and Mamba network for medical ultrasound image segmentation[J]. Biomedical Signal Processing and Control, 2026, 112: 108683.
- [12] 梁宵, 陈莹. 基于知识蒸馏的空间通道双自编码器无监督异常检测[J]. 电子测量与仪器学报, 2025, 39(3): 146-158.
- Liang Xiao, Chen Ying. Unsupervised anomaly detection based on spatial-channel dual autoencoder with knowledge distillation[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(3): 146-158.
- [13] Guo B J, Huang W H, Wang X Y. ABE-Mamba: Few-shot medical image segmentation via adversarial bidirectional enhanced Mamba[J]. Expert Systems with Applications, 2026, 298: 129897.
- [14] 成荣, 朱文忠, 王文. UltraLight CrackNet: 基于 VMamba 的轻量化裂缝分割网络[J]. 电子测量技术, 2025, 48(22): 224-234.
- Cheng Rong, Zhu Wenzhong, Wang Wen. UltraLight CrackNet: A lightweight crack segmentation network based on VMamba [J]. Electronic Measurement Technology, 2025, 48(22): 224-234.
- [15] Qiao J B, Liao J C, Li W, et al. Hi-Mamba: Hierarchical Mamba for efficient image super-resolution[PP/OL]. V1. (2024-10-14) [2025-11-14]. <https://doi.org/10.48550/arXiv.2410.10140>.
- [16] 周杰伟, 李智, 张绍荣, 等. 融合多级注意力与多尺度信息的铁轨缺陷分割网络[J]. 电子测量与仪器学报, 2025, 39(7): 140-150.
- Zhou Weijie, Li Zhi, Zhang Shaorong, et al. A rail defect segmentation network integrating multi-level attention and multi-scale information[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(7): 140-150.
- [17] 许孔晨, 汤怀锋, 杨海倩, 等. 基于改进 ResNet-50 的智能流式细胞分析方法[J]. 电子测量技术, 2025, 48(18): 41-52.
- Xu Kongchen, Tang Huaifeng, Yang Haiqian, et al. An intelligent flow cytometry analysis method based on improved ResNet-50[J]. Electronic Measurement Technology, 2025, 48(18): 41-52.

- [18] 张绍文, 史卫亚, 张世强, 等. 基于加权感受野和跨层融合的遥感小目标检测[J]. 电子测量技术, 2023, 46(18): 129-138.  
Zhang Shaowen, Shi Weiya, Zhang Shiqiang, et al. Remote sensing small target detection based on weighted receptive field and cross-layer fusion [J]. Electronic Measurement Technology, 2023, 46(18): 129-138.
- [19] 陈佳琳, 尚志武, 张雷. 基于 IDRSN-BiLSTM 的铣削加工表面粗糙度预测方法[J]. 仪器仪表学报, 2024, 45(4): 27-36.  
Chen Jialin, Shang Zhiwu, Zhang Lei. A surface roughness prediction method for milling based on IDRSN-BiLSTM [J]. Chinese Journal of Scientific Instrument, 2024, 45(4): 27-36.
- [20] Wang J, Li X Q, Zhou L H, et al. Adaptive Receptive Field Enhancement Network based on attention mechanism for detecting the small target in the aerial image[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-18.
- [21] Jia X F, Wang W J, Zhang M, et al. Atten-Nonlocal UNet: Attention and non-local Unet for medical image segmentation[J]. Computers in Biology and Medicine, 2025, 191: 110129.
- [22] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//32nd International Conference on Machine Learning(ICML), 2015: 6-11.
- [23] Chen L, Zhang H W, Xiao J, et al. SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5659-5667.
- [24] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [25] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6848-6856.
- [26] Cao M H, Zhao J C, Zhang Y, et al. An improved Lidar point cloud target detection algorithm for networked vehicles[C]//21st International Conference on Optical Communications and Networks(ICOCN), 2023.
- [27] Zhou Z W, Rahman Siddiquee M M, Tajbakhsh N, et al. UNet++: A nested U-Net architecture for medical image segmentation [C]//International Workshop on Deep Learning in Medical Image Analysis, 2018: 3-11.
- [28] Ruan J C, Li J C, Xiang S C. VM-UNet: Vision Mamba UNet for medical image segmentation[PP/OL]. V2. (2024-11-08) [2025-11-14]. <https://doi.org/10.48550/arXiv.2402.02491>.
- [29] Duan G H, Cao Y K, Cui X X, et al. SwinDeepLab: Swin Transformer-based DeepLab v3+ for gland segmentation [C]//16th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics(CISP-BMEI), 2023: 1-6.
- [30] Qamar S, Furqan Qadri S, Alroobaea R, et al. ScaleFusionNet: Transformer-guided multi-scale feature fusion for skin lesion segmentation [J]. Scientific Reports, 2025, 15(1): 34393.
- [31] Nguyen V T, Pham V T, Tran T T. AC-MambaSeg: An adaptive convolution and Mamba-based architecture for enhanced skin lesion segmentation [C]//International Conference on Green Technology and Sustainable Development, 2024: 13-26.

## 作者简介

**李寅**, 硕士研究生, 主要研究方向为深度学习、人工智能。

E-mail: 15190306507@163.com

**韩基泰**(通信作者), 博士, 副教授, 主要研究方向为电池、深度学习、人工智能。

E-mail: 860085@cw Xu.edu.cn

**张凯文**, 硕士研究生, 主要研究方向为深度学习、机器人机械臂。

**于敏**, 硕士研究生, 主要研究方向为目标检测与计算机视觉。

**刘小庆**, 硕士研究生, 主要研究方向为目标检测与计算机视觉。