

基于金字塔注意力与双路径融合的皮肤病变图像分割^①



宋存利¹, 傅景鑫¹, 王 依¹, 张雪松¹, 时维国²

¹(大连交通大学 轨道智能工程学院, 大连 116052)

²(大连交通大学 电气工程学院, 大连 116052)

通信作者: 时维国, E-mail: swgdl@163.com

摘 要: 在皮肤病变图像分割任务中, U-Net 在处理皮肤镜图像时存在多尺度适应性不足、跨层特征融合低效及计算冗余导致边缘信息丢失等问题. 本文提出层次化金字塔注意力网络 HPANet (hierarchical pyramid attention network), 通过金字塔注意力模块和双路径特征融合机制, 实现了多尺度特征捕获和跨层特征传递的双重优化. 其中, 双路径自适应融合模块结合 CNN 与 Transformer 双分支特征, 通过通道注意力与压缩空间注意力增强互补特征的信息交互, 并利用双线性交互与残差连接缓解特征稀释问题. 金字塔注意力模块结合分层多核卷积、深度可分离下采样及分块空间通道注意力机制, 显著提升多尺度病变特征捕获能力. 实验结果表明, 本架构在 ISIC 2017、ISIC 2018 数据集集中的表现均超越主流模型, 证实其在病变边界保留与小病灶检测方面的双重优势.

关键词: 金字塔注意力; 多尺度特征融合; 医学图像分割; 深度学习

引用格式: 宋存利,傅景鑫,王依,张雪松,时维国.基于金字塔注意力与双路径融合的皮肤病变图像分割.计算机系统应用,2026,35(1):178-187. <http://www.c-s-a.org.cn/1003-3254/10050.html>

Skin Lesion Image Segmentation Based on Pyramid Attention and Dual-path Fusion

SONG Cun-Li¹, FU Jing-Xin¹, WANG Yi¹, ZHANG Xue-Song¹, SHI Wei-Guo²

¹(School of Intelligent Rail Transit Engineering, Dalian Jiaotong University, Dalian 116052, China)

²(School of Electrical Engineering, Dalian Jiaotong University, Dalian 116052, China)

Abstract: In skin lesion image segmentation tasks, U-Net suffers from insufficient multi-scale adaptability, inefficient cross-layer feature fusion, and computational redundancy that leads to edge information loss when processing dermoscopy images. This study proposes a hierarchical pyramid attention network (HPANet) that achieves dual optimization of multi-scale feature capture and cross-layer feature transmission through pyramid attention module and dual-path feature fusion mechanism. The dual-path adaptive fusion module combines CNN and Transformer dual-branch features, enhancing information interaction of complementary features through channel attention and compressed spatial attention, while utilizing bilinear interaction and residual connections to alleviate feature dilution problems. The pyramid attention module integrates hierarchical multi-kernel convolution, depthwise separable downsampling, and patch-wise spatial-channel attention mechanisms to significantly improve multi-scale lesion feature capture capability. Experimental results demonstrate that the proposed architecture outperforms mainstream models on both ISIC 2017 and ISIC 2018 datasets, confirming its dual advantages in lesion boundary preservation and small lesion detection.

Key words: pyramid attention; multi-scale feature fusion; medical image segmentation; deep learning

① 基金项目: 国家自然科学基金 (62276042); 辽宁省教育厅项目 (LJKMZ20220828)

收稿时间: 2025-06-10; 修改时间: 2025-06-30, 2025-08-01; 采用时间: 2025-08-15; csa 在线出版时间: 2025-11-26

CNKI 网络首发时间: 2025-11-27

黑色素瘤是一种高致死率的皮肤癌,也是世界上增长速度最快的癌症之一,近年来,该病每年导致近6万人因此丧命^[1].研究表明,癌细胞一旦扩散至全身组织,患者的生存率将骤降至约10%,但若能在早期发现并干预,患者的治愈率可高达90%,因此,精准的早期诊断对提高患者存活率至关重要.

传统的医学图像分割方法多依赖支持向量机(SVM)、随机森林(RF)和马尔可夫随机场(MRF)等经典机器学习技术^[2],或采用阈值分割、区域生长与边缘检测等传统图像处理手段.虽然这些方法在处理结构简单、边界清晰的图像时能取得一定效果,但在面对复杂背景、图像噪声、边缘模糊及病灶形态变化等问题时,往往难以实现精确分割,进而影响诊断的准确性与可靠性.近年来,深度学习,特别是卷积神经网络(CNN),通过自动学习图像特征,提高了分割精度,有效克服了传统方法的局限,并催生了多种经典网络架构.其中,以U-Net^[3]为代表的卷积神经网络分割架构,凭借其结构简单、高效的编码器-解码器设计,在医学图像分割任务中表现出色.其高度的灵活性与可扩展性也为后续大量研究提供了良好的基础,推动了多种架构的优化与创新.例如,2019年,UNet++^[4]在U-Net基础上引入嵌套解码器和密集跳跃连接,强化特征提取,提升多尺度分割精度与鲁棒性.Chen等人^[5-8]提出的DeepLab系列(v1, v2, v3, v3+)通过引入空洞卷积(atrous convolution)和空间金字塔池化(ASPP)模块,提高了模型对不同尺度病变区域的感知能力,并在多个版本迭代中不断优化分割精度与计算效率.Garbaz等人^[9]提出的DMFC-UFormer通过深度多尺度分解卷积(DMFC)和多尺度特征提取(MFFE)模块,有效扩展了感受野并增强了特征多样性,同时利用空间通道分离特征注意力(SCPFA)模块和基于注意力的特征稳定(AFS)模块,在处理小病变区域和复杂背景时表现出色,显著提升了分割精度.Badshah等人^[10]提出了ResBCU-Net,该架构结合了残差设计、批量归一化和双向长短时记忆网络,防止模型在皮肤病变分割中学习重复特征,并通过批量归一化加快训练,减少过拟合,同时利用BConvLSTM缓解长期依赖关系中的梯度消失问题.密集卷积连接层则增强了特征传播,进一步提升了分割精度.Ruan等人^[11]提出EGE-UNet,引入GHPA和GAB模块,通过多尺度特征提取和特征融合,实现了超低参数量和计算量,使其更适合移动医疗应用.

尽管CNN模型在医学图像分割中表现优异,但其主要依赖堆叠卷积层逐步扩大感受野,虽能有效提取局部特征,却难以充分建模图像中的全局信息,限制了性能的进一步提升.相比之下,Transformer^[12]架构通过自注意力机制捕捉不同位置间的依赖关系,在全局特征建模方面展现出显著优势.因此,融合CNN的局部特征提取能力与Transformer的全局建模能力,以实现信息处理的全局与局部协同,正成为当前研究的热点方向.Chen等人^[13]提出的TransUNet结合Transformer的全局上下文建模能力与U-Net的局部特征提取能力,在医学图像分割中取得了显著的性能提升.Tagnamas等人^[14]提出的SCA-InceptionUNeXt通过引入空间通道注意力机制(SCA),增强了医学图像中关键特征的表达,显著提高了分割精度,尤其在处理小病变区域和复杂边界时表现出色.Nam等人^[15]提出的TransGUNet是一种创新的医学图像分割模型,通过结合注意力机制的跨尺度图神经网络和基于熵的特征选择空间注意力,有效减少了编码器与解码器之间的语义差距,并在多种模态和临床场景下展现出卓越的性能和效率.Zhong等人^[16]提出的PMFSNet通过引入极化多尺度特征自注意力模块(PMFS block),在轻量化的网络架构中实现了全局与局部特征处理的平衡,显著提升了医学图像分割的效率与性能.Liu等人^[17]提出的CSWinUNet通过引入交叉窗状自注意力机制,显著增强了特征提取的效率和感受野交互能力,同时在解码器中采用内容感知重装配算子进行上采样,有效保留了目标边缘和细节特征,从而在复杂医学图像分割任务中实现了更高的分割精度和计算效率.

2021年,Zhang等人^[18]提出的TransFuse结合了浅层CNN与Transformer,通过BiFusion模块有效整合两者特征,实现高效协同预测,捕捉底层空间特征与高层语义信息,显著提升了模型在皮肤病变分割中的表现.然而,采用ResNet和DeiT(data-efficient image Transformer)作为骨干网络增加了模型复杂度和训练时间,对计算资源提出了更高的要求.同年,Cao等人^[19]提出了分层Swin Transformer作为编码器,并结合带有跳跃连接的Transformer架构的对称解码器,摒弃了传统的卷积操作,有效获取了皮肤病变图像中的全局和长程语义信息.2022年,Wu等人^[20]提出的FAT-Net通过特征自适应变换器和记忆高效解码器融合局部特征与全局信息,但在特征融合时采用逐元素相加,未充分考虑两

编码器特征差异. 2024年, Huang等人^[21]提出 ADF-Net, 包含多阶段双流混合框架(MDHF)和焦点注意力解码器(FAD), 实现粗粒度和细粒度特征的综合表示, 有效解决皮肤病变分割中的大尺度变化、边界模糊等问题, 显著提升分割精度. 然而, 以上模型都没有涉及对于跨模态特征高效融合、多尺度精细建模以及轻量化计算优化的综合解决方案.

尽管 U-Net 及其改进方法在医学图像分割中取得显著成效, 但仍存在跨层特征交互低效、多尺度语义对齐偏差及边界细节易丢失等问题, 限制了其在复杂场景下的表现. 传统跳跃连接难以建模高阶特征交互^[3], 导致多尺度信息传递效率低下^[18]; 卷积操作依赖单一尺度^[22], 难以兼顾大小病变区域; 而常用的插值或反卷积上采样易造成边界模糊^[23]. 此外, 引入 Transformer 虽能提升性能, 却显著增加计算负担^[24], 影响实际部署效率.

针对上述问题, 本文提出层次化金字塔注意力网络(hierarchical pyramid attention network, HPANet), 一种通过多尺度金字塔卷积提取局部和全局特征, 并结合包含通道-空间注意力机制的双路径特征融合分割架构. 具体来说, 金字塔注意力模块(pyramid attention mo-

dule) 采用分层多核卷积与分块空间通道注意力, 同步提取局部-全局特征, 结合深度可分离下采样压缩计算量, 实现多尺度病变区域的精准捕获; 双路径自适应融合(dual adaptive fusion, DAF)模块通过通道-空间双注意力引导的双路径融合机制, 动态校准编码器-解码器的高阶特征交互, 抑制冗余噪声. 为了验证 HPANet 的有效性与其泛化能力, 在皮肤病变数据集 ISIC 2017^[25]和 ISIC 2018^[26]上进行了实验, 并与现有主流方法进行了比较. 实验结果表明, 本研究提出的网络模型在性能上优于当前最先进的算法.

1 方法

如图 1 所示, 层次化金字塔注意力网络(HPANet)融合全局上下文与局部结构特征, 实现多层次信息互补. 该网络采用双路径架构: Transformer 路径基于轻量级 Vision Transformer (DeiT), 高效提取图像的全局长距离依赖特征, 并通过两个级联上采样模块(UP)逐步恢复空间分辨率; CNN 路径集成 PAM 模块, 增强局部区域的多尺度建模能力, 精准捕捉边缘与纹理细节, 同时利用 Dropout 策略提升泛化性能.

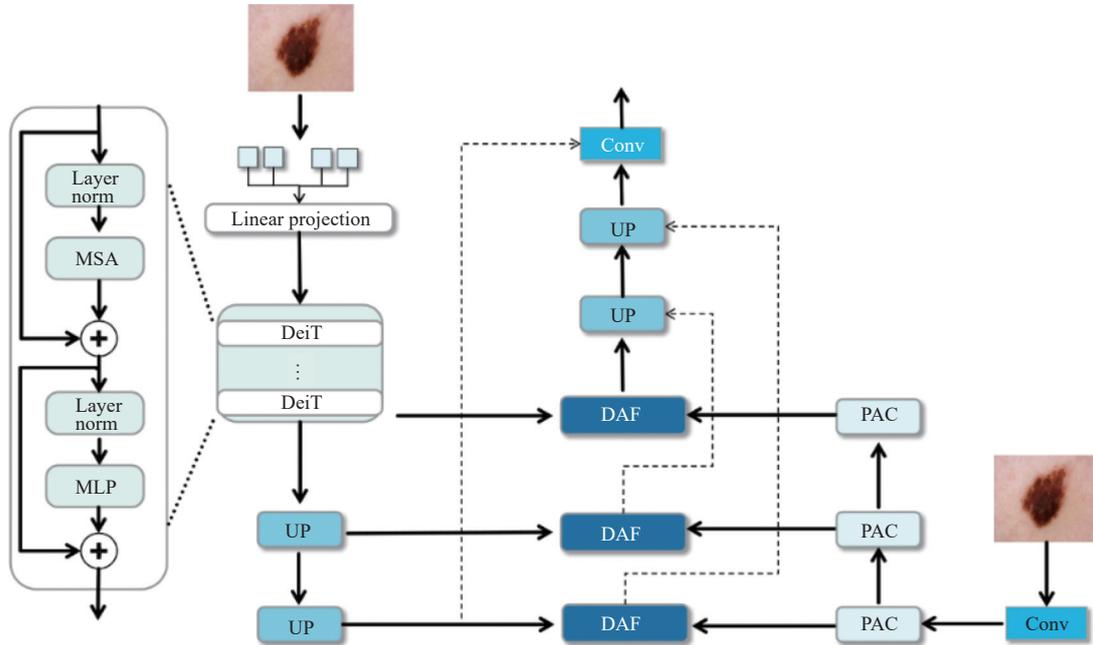


图 1 本文层次化金字塔注意力网络(HPANet)

为协调 Transformer 与 CNN 在特征表达上的差异, HPANet 引入 DAF 模块, 在浅层、中层和高层之间构建多级融合机制, 结合通道显著性建模与空间注意力机制, 增强对关键区域的响应能力. 融合特征经注意

力引导的上采样模块进一步整合, 优化解码阶段的细节恢复.

解码器采用三支路输出结构, 并引入多层次监督策略以提升细粒度特征学习效果: 主支路对融合特征

进行卷积操作并上采样至原图尺寸,两个辅助支路分别监督 Transformer 路径与融合路径输出,保障不同语义层级的有效学习.所有预测结果均通过双线性插值统一上采样,确保输出一致性.

HPANet 有效结合 Transformer 的长距离建模能力与 CNN 的局部结构感知优势,在多尺度融合中实现高效协同,显著提升病变区域的边缘感知与语义建模精度.图 2 展示了 MLP 与 Transformer encoder 的结构细节,便于理解网络机制.

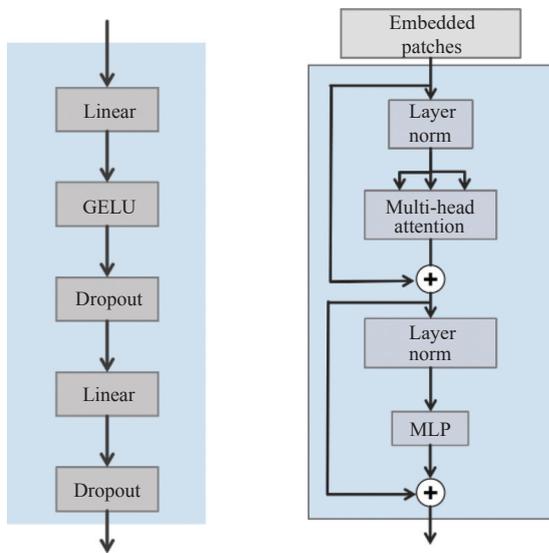


图 2 MLP 和 Transformer 模块

1.1 双路径自适应融合模块

为解决 CNN 与 Transformer 特征融合中异构特征对齐不精确和通道语义关联不充分的问题,本文提出了一种基于自适应通道关联学习的双路径融合策略:双路径自适应融合 (DAF) 模块 (见图 3),该模块通过引入通道注意力机制的自相关建模,显著提升了多模态特征的特征能力和融合效率.具体来说,该模块以 CNN 路径特征 g 和 Transformer 路径特征 x 作为输入,首先针对两个分支的不同特性采用差异化的注意力增强策略:对于 CNN 分支,通过池化操作并行提取通道维度的最大值和均值特征生成通道压缩表示,再经过 7×7 卷积和 *Sigmoid* 激活生成空间注意力图,有效增强对局部细节和边缘信息的响应.对于 Transformer 分支,将输入特征同时投影为 Query (Q)、Key (K) 和 Value (V) 这 3 个注意力分量,首先通过 Query 和 Key 的乘积获得通道间相关性能量矩阵 $E = Q \otimes K^T$,其中 $E = R^{B \times C \times C}$,

表示每一对通道之间的语义关联强度.为了增强注意力对比度,我们采用基于最大值的反差增强策略,即对 E 的每一行元素计算其最大值 $\max(E)$,并与原始能量矩阵作差得到残差图 $\max(E) - E$,该操作可强化对低响应通道的关注,提升对判别性特征的建模能力.随后通过归一化得到注意力权重,对 Value 进行加权融合,最后引入学习缩放因子 γ 并与原始输入特征 x 做残差连接,得到增强后的特征图 C_{att} .

在获得注意力增强特征的基础上,我们进一步设计了双线性交互建模机制,将输入特征 g 和 x 分别通过 1×1 卷积映射为中间表示 W_g 和 W_x ,随后进行逐元素乘法捕获交互信息,并通过 3×3 卷积生成双线性融合特征 bp ,有效建模了跨模态特征间的非线性交互关系.最终,DAF 将经注意力增强的 CNN 特征、Transformer 特征及双线性交互特征进行通道拼接,通过残差融合模块统一整合以保证梯度流动稳定性,同时采用 Dropout 机制以抑制过拟合,从而实现了异构特征间的深度语义耦合与自适应融合.DAF 模块的数学公式如下所示:

$$CP(X) = \text{Concat}(\text{MaxPool}(X), \text{AvgPool}(X)) \quad (1)$$

$$bp = \text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 1}(g) \odot \text{Conv}_{1 \times 1}(x)) \quad (2)$$

$$S_{att} = \text{Sigmoid}(\text{Conv}_{7 \times 7}(CP(g))) \odot g \quad (3)$$

$$C_{att} = \gamma \times \text{Softmax}(\max(E) - E) \otimes V + x \quad (4)$$

$$DAF(g, x) = \text{Residual}(\text{Concat}(S_{att}, C_{att}, bp)) \quad (5)$$

其中, g 和 x 分别代表 CNN 和 Transformer 分支的输入特征图, bp 表示双线性池化后的特征图. \odot 代表逐元素乘法, $\text{Conv}_{k \times k}$ 表示 $k \times k$ 滤波器的卷积操作. $CP(\cdot)$ 代表通道池化操作, S_{att} 是带有空间注意力权重的特征图. $E = Q \otimes K^T$, Q 代表能量矩阵, V 是值的投影矩阵, γ 是可学习参数, C_{att} 是经过通道注意力处理后的特征图. *Softmax* 操作用于计算归一化的注意力分布. *Residual*(\cdot) 代表残差连接操作, *Concat*(\cdot) 代表特征拼接.最终的 $DAF(g, x)$ 是融合后的输出特征图.

1.2 金字塔注意力模块

基于多尺度信息融合的思想,设计了金字塔注意力模块 (pyramid attention module, PAM),如图 4 所示.具体而言, PAM 模块采用不同大小的卷积核 (3×3 、 5×5 、 7×7),在不同尺度上提取特征,并根据分组卷积

1×1 卷积的可训练权重, $Conv_{k \times k}$ 表示具有 $k \times k$ 滤波器的卷积操作, $Concat$ 代表通道维度上的拼接操作. $\sigma(\cdot)$ 是 *Sigmoid* 激活函数, $Pool(\cdot)$ 代表池化操作, \odot 表示逐元素乘法 (用于注意力加权), $ReLU(\cdot)$ 为 *ReLU* 激活函数. 最终, $Dropout(\cdot)$ 用于随机丢弃部分神经元, 以防止过拟合.

1.3 损失函数

本文采用融合加权的交并比损失 (weighted intersection-over-union loss) 和加权的二元交叉熵损失 (weighted binary cross-entropy loss) 构建基础目标函数, 其数学公式如式 (10) 所示:

$$L_{\text{base}} = W_{\text{IoU}} + W_{\text{BCE}} \quad (10)$$

其中, 特别对边缘区域的像素点施加更高的权重系数^[27], 以强化模型对轮廓细节的敏感性. 为进一步优化训练稳定性, 设计了一种层次化监督机制: 除主干网络输出外, 额外对 Transformer 分支的深层特征 (\tilde{F}^2) 和首个融合阶段的中间特征 (F^0, F^2) 进行联合监督. 最终的目标函数通过可调参数实现多层次特征约束, 具体如式 (11) 所示:

$$L_{\text{final}} = \lambda_1 \cdot D(G, \Psi(\tilde{F}^2)) + \lambda_2 \cdot D(G, \Psi(F^2)) + \lambda_3 \cdot D(G, \Psi(F^0)) \quad (11)$$

其中, $\lambda_1, \lambda_2, \lambda_3$ 为动态平衡系数, G 表示标注真值, Ψ 为特征解码器, D 表示基础损失计算模块. 该策略通过梯度传播路径的多样性增强, 显著提升了模型对全局语义与局部细节的协同建模能力.

2 实验与结果分析

2.1 数据集

本研究在 ISIC 2017 和 ISIC 2018 皮肤病变分割数据集上进行实验, 以验证所提出模型的有效性. 这两个数据集均来源于国际皮肤影像合作组织 (international skin imaging collaboration, ISIC). ISIC 2017 数据集包含 2 000 张训练图像、150 张验证图像和 600 张测试图像, 每张图像均配备相应的病变区域分割标签. ISIC 2018 数据集相较于 ISIC 2017 数据集, 规模更大且更加多样化, 包含 2 594 张训练图像和 1 000 张测试图像. 每张图像均配有像素级分割标签, 涵盖了不同类型的皮肤病变 (如黑色素瘤、良性皮损等).

为进一步评估模型在跨器官与不同病灶类型分割

任务中的泛化能力, 本研究补充了在 CVC-ClinicDB 数据集上的实验. 该数据集是结肠镜视频筛查领域的重要基准, 专门用于结肠镜视频图像中的息肉自动分割任务. CVC-ClinicDB 数据集共包含从多个检查视频中提取的 612 张图像, 其中的息肉在大小、形状和外观上具有多样性, 对分割算法的鲁棒性提出了较高要求.

在所有实验中, 输入图像均调整至 256×256 的统一分辨率, 以适配所设计的网络结构, 并保证训练过程的稳定性. 所有数据均经过归一化处理, 将像素值缩放至 $[0, 1]$ 区间, 以加速模型收敛. 为了避免模型过拟合, 在训练过程中采用了随机水平翻转、随机旋转、随机裁剪和亮度调整等数据增强策略, 从而提升模型在皮肤病变和结肠息肉等多种场景下的泛化能力.

2.2 实验环境与参数设置

本文深度学习框架的开发环境是 PyTorch 1.8.1, CUDA 版本为 11.1, 汇编语言为 Python 3.8.12, 本文所有实验均在 Windows 操作系统下完成. 实验中使用的显卡为 NVIDIA GeForce RTX 4080, 显存为 16 GB. 本文使用的优化器是 Adam, 损失函数是二进制交叉熵函数, 初始学习率设置为 1×10^{-4} , 最大 epoch 为 100 轮, 每 10 轮验证一次, Batch 大小设置为 8. 在训练中, 若损失未降低, 则调整学习率. 最终选取验证精度最高的模型作为最终结果.

2.3 消融实验

HPANet 是基于 PAM 和 DAF 构建的医学图像分割网络. 为验证模型中各模块的有效性, 本实验采用模块级消融策略, 通过逐步添加 PAM 和 DAF 模块, 评估它们对分割性能的影响, 统一使用 ISIC 2017 数据集作为测试集. 实验结果汇总于表 1, 分割可视化对比见图 5. 对比指标包括 *Dice* 相似系数 (*Dice*)、*IoU* (intersection over union) 和准确率 (*ACC*) 这 3 项, 计算方式如式 (12)–式 (14):

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (12)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (13)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (14)$$

本实验首先分析了 PAM 模块对多尺度特征建模的影响. 与基线模型^[18]相比, Baseline+PAM 提升了 *Dice* 相似系数 (从 85.02% 提升至 86.20%)、*IoU* (从

76.68% 提升至 78.49%) 和准确率 (从 90.87% 提升至 92.26%)。这一结果表明, PAM 模块通过金字塔卷积增强了不同尺度特征的捕获能力, 从而提高了模型在分割病变区域时的准确性。

其次, 实验也验证了 DAF 模块在特征融合方面的贡献。Baseline+DAF 的 *Dice* 和 *IoU* 分别达到 86.17% 和 77.84%。DAF 模块通过双分支结构, 分别提取低层细节特征和高层语义特征, 并结合通道注意力与空间

注意力机制, 实现跨层特征的自适应加权融合与对齐, 从而有效增强模型的分割性能及边界细节的保留能力。

表 1 在 ISIC 2017 数据集上对 HPANet 进行皮肤病变分割的消融研究 (%)

Method	<i>Dice</i>	<i>IoU</i>	<i>ACC</i>
Baseline	85.02	76.68	90.87
Baseline+PAM	86.20	78.49	92.26
Baseline+DAF	86.17	77.84	91.75
Baseline+PAM+DAF	87.11	79.40	93.34

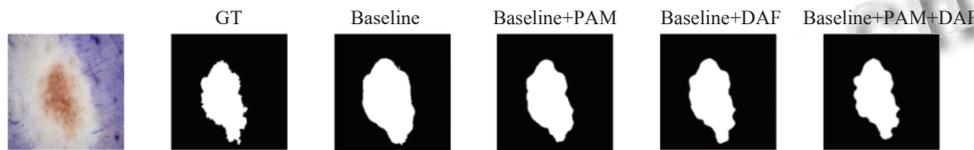


图 5 不同模块在 ISIC 2017 数据集上的可视化结果

为了进一步地分析 PAM 模块与 DAF 模块的协同作用, 本实验在基线模型中同时引入两者 (Baseline+PAM+DAF), 结果表明, *Dice* 提升至 87.11%, *IoU* 提升至 79.40%, 准确率提高至 93.34%。相比单独引入 PAM 或 DAF 模块, 协同作用带来了更显著的提升, 进一步验证了 PAM 模块对多尺度特征的增强作用, 以及 DAF 模块在跨层信息融合中的优势。

此外, 本研究还展示了原模型与 HPANet 的总损失评估收敛曲线如图 6 所示, 从图 6 中可以看到 HPANet 的损失曲线低于原模型, 表明 HPANet 在训练过程中具有更快的收敛速度和更优的优化效果, 这也进一步验证了该模型在皮肤病变分割任务中具备更强的特征表达能力和泛化性能。

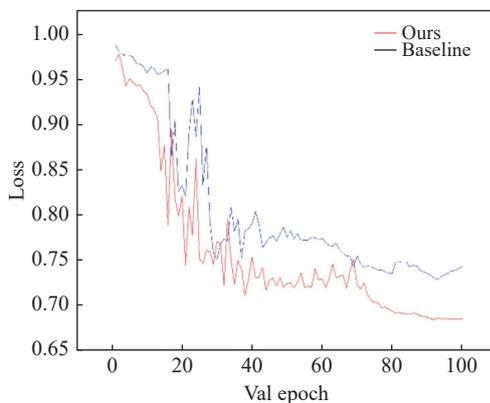


图 6 总损失评估图

综上所述, 消融实验结果表明, PAM 模块和 DAF 模块在皮肤病变分割任务中均能有效提升模型性能, 尤其在多尺度建模与特征融合方面具有显著的互补性,

最终使分割结果更加精准。

2.4 对比实验

为了验证模型的分割性能, 我们在 ISIC 2017 和 ISIC 2018 这两个数据集上进行了对比实验, 将 HPANet 模型与 5 种主流分割方法进行比较: U-Net, UNet++, UCTransNet^[28], DA-TransUNet^[29], HTC-Net^[30]。使用 *Dice*、*IoU*、*SE* (sensitivity)、*SP* (specificity) 和 *ACC* 来评价这些方法的分割性能。

$$SE = \frac{TP}{TP + FN} \quad (15)$$

$$SP = \frac{TN}{TN + FP} \quad (16)$$

在 ISIC 2017 和 ISIC 2018 数据集上的实验结果如表 2、表 3 所示, HPANet 在皮肤病变分割任务中展现出显著优势。后续表格中加粗表示最优结果。

表 2 不同方法在 ISIC 2017 数据集上的分割性能 (%)

Model	Year	<i>Dice</i>	<i>IoU</i>	<i>SE</i>	<i>SP</i>	<i>ACC</i>
U-Net	2015	82.76	75.21	86.93	88.67	89.56
UNet++	2018	84.37	76.83	88.36	89.43	90.73
UCTransNet	2022	85.79	77.91	89.57	89.77	91.28
DA-TransUNet	2024	86.53	78.15	91.46	90.46	92.17
HTC-Net	2024	86.21	79.27	91.67	89.63	92.85
Ours	2025	87.11	79.40	90.22	90.85	93.34

实验结果表明, HPANet 在所有评估指标上均取得了良好的表现。其中, 在 ISIC 2017 数据集上, HPANet 的 *Dice* 和 *IoU* 分别达到 87.11% 和 79.40%; 在 ISIC 2018 数据集上, 其 *Dice* 和 *IoU* 分别为 87.78% 和 80.62%。此外, 与近年来的先进模型 UCTransNet、DA-

TransUNet 等相比, HPANet 在 *Dice*、*IoU* 及准确率等方面均表现优异, 进一步验证了其在分割任务中的精度与鲁棒性。

在图 7 中给出 HPANet 在 ISIC 2017 和 ISIC 2018 可视化分割图, 从图中可以看出, HPANet 在较小目标区域的预测结果仍然是最优的, 能够有效捕捉远程依赖关系, 减少特征信息的丢失。显然, 本文提出的方法在分割病变边缘区域上展现了显著的准确性, 即便是微小的病灶, 也能实现精确的定位和边界分割, 充分展

示了其强大的性能。整体而言, HPANet 在分割精度和细节捕捉方面展现了卓越的能力。

表 3 不同方法在 ISIC 2018 数据集上的分割性能 (%)

Model	Year	<i>Dice</i>	<i>IoU</i>	<i>SE</i>	<i>SP</i>	<i>ACC</i>
U-Net	2015	83.26	77.25	83.01	88.41	88.67
UNet++	2018	84.74	78.41	84.72	89.53	89.36
UCTransNet	2022	85.49	78.86	86.54	90.69	90.68
DA-TransUNet	2024	85.93	79.50	89.39	89.47	92.45
HTC-Net	2024	86.51	80.38	90.16	90.61	92.91
Ours	2025	87.78	80.62	89.78	91.28	93.21

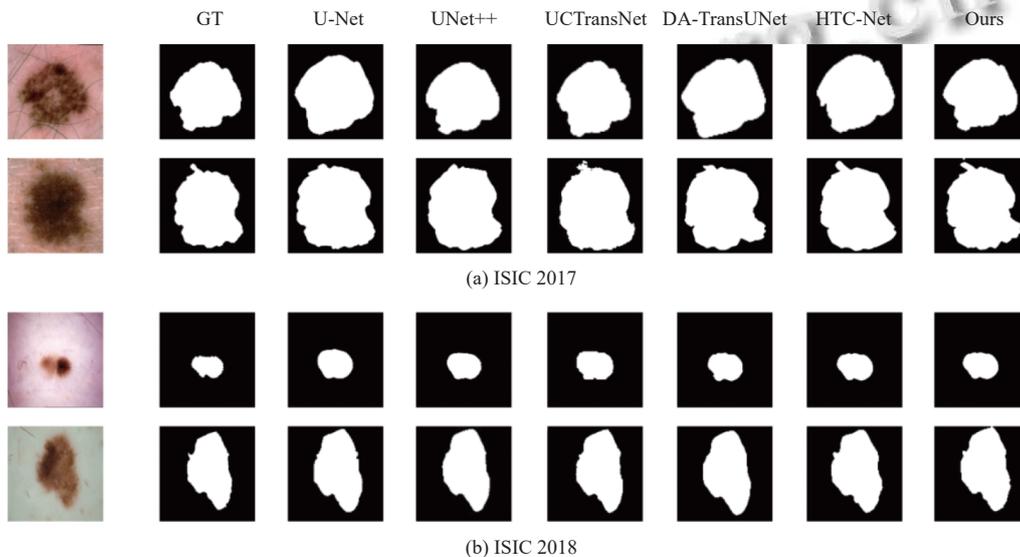


图 7 不同方法在 ISIC 2017、ISIC 2018 数据集上的可视化结果

为验证该模型的泛化性, 我们还在 CVC-ClinicDB 上进行了实验 (见表 4、图 8), 与 U-Net、PraNet^[31]、SANet^[32]、VM-UNET^[33] 等主流模型进行对比。实验结果表明, 本模型在该数据集上取得了 87.68% 的 *Dice* 系数、79.48% 的 *IoU* 和 98.20% 的准确率, 全面优于对比模型, 展现出更强的分割性能和跨数据集的泛化能力。

表 4 不同方法在 CVC-ClinicDB 数据集上的分割性能 (%)

Model	Year	<i>Dice</i>	<i>IoU</i>	<i>ACC</i>
U-Net	2015	83.67	74.03	96.24
PraNet	2020	87.09	78.87	97.75
SANet	2021	86.92	78.79	97.51
VM-UNET	2024	87.33	80.22	97.83
Ours	2025	87.68	79.48	98.20

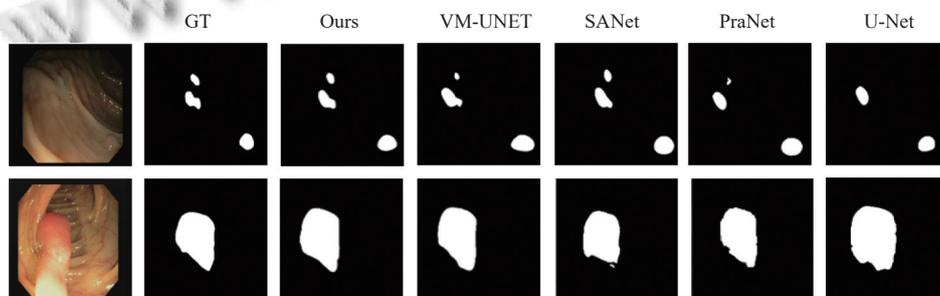


图 8 不同方法在 CVC-ClinicDB 数据集上的可视化结果

3 结论

本研究针对传统 U-Net 结构在跨层特征交互、特

征对齐及多尺度建模方面的不足, 提出了层次化金字塔注意力网络, 并设计了两大关键模块: DAF 模块与 PAM

模块。

DAF 模块通过双路径融合机制,将双线性池化与通道-空间注意力机制有机结合,增强高阶特征的交互表达能力,有效抑制冗余噪声,显著提升边缘区域的识别效果。PAM 模块采用多级金字塔卷积结构以扩展感受野,并引入区域自适应机制实现特征的动态增强,从而提升对小目标的检测能力,尤其在病理切片等细粒度分割任务中表现出色。

总体而言,本文提出的两大核心改进有效弥补了 U-Net 在结构设计上的局限性,显著提升了模型在多尺度、多模态医学图像分割任务中的适应性及泛化能力,展现出良好的临床辅助诊断潜力。

参考文献

- 1 Chatterjee S, Dey D, Munshi S. Integration of morphological preprocessing and fractal based feature extraction with recursive feature elimination for skin lesion types classification. *Computer Methods and Programs in Biomedicine*, 2019, 178: 201–218. [doi: [10.1016/j.cmpb.2019.06.018](https://doi.org/10.1016/j.cmpb.2019.06.018)]
- 2 Seo H, Badiei Khuzani M, Vasudevan V, *et al.* Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications. *Medical Physics*, 2020, 47(5): e148–e167.
- 3 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention*. Munich: Springer, 2015. 234–241.
- 4 Zhou ZW, Siddiquee MR, Tajbakhsh N, *et al.* UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 2019, 39(6): 1856–1867.
- 5 Chen LC, Zhu YK, Papandreou G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018. 801–818.
- 6 Chen LC, Papandreou G, Kokkinos I, *et al.* Semantic image segmentation with deep convolutional nets and fully connected CRFs. *Proceedings of the 3rd International Conference on Learning Representations*. San Diego, 2015.
- 7 Chen LC, Papandreou G, Kokkinos I, *et al.* DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834–848. [doi: [10.1109/TPAMI.2017.2699184](https://doi.org/10.1109/TPAMI.2017.2699184)]
- 8 Chen LC, Papandreou G, Schroff F, *et al.* Rethinking atrous convolution for semantic image segmentation. *arXiv:1706.05587*, 2017.
- 9 Garbaz A, Oukdach Y, Charfi S, *et al.* DMFC-UFormer: Depthwise multi-scale factorized convolution Transformer-based UNet for medical image segmentation. *Biomedical Signal Processing and Control*, 2025, 101: 107200. [doi: [10.1016/j.bspc.2024.107200](https://doi.org/10.1016/j.bspc.2024.107200)]
- 10 Badshah N, Ahmad A. ResBCU-Net: Deep learning approach for segmentation of skin images. *Biomedical Signal Processing and Control*, 2022, 71: 103137. [doi: [10.1016/j.bspc.2021.103137](https://doi.org/10.1016/j.bspc.2021.103137)]
- 11 Ruan JC, Xie MY, Gao JS, *et al.* EGE-UNet: An efficient group enhanced UNet for skin lesion segmentation. *Proceedings of the 26th International Conference on Medical Image Computing and Computer-assisted Intervention*. Vancouver: Springer, 2023. 481–490.
- 12 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 13 Chen JN, Lu YY, Yu QH, *et al.* TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv:2102.04306*, 2021.
- 14 Tagnamas J, Ramadan H, Yahyaouy A, *et al.* SCA-InceptionUNeXt: A lightweight spatial-channel-attention-based network for efficient medical image segmentation. *Knowledge-based Systems*, 2025, 311: 113166. [doi: [10.1016/j.knosys.2025.113166](https://doi.org/10.1016/j.knosys.2025.113166)]
- 15 Nam JH, Syazwany NS, Lee SC. TransGUNet: Transformer meets graph-based skip connection for medical image segmentation. *arXiv:2502.09931*, 2025.
- 16 Zhong JH, Tian WH, Xie YL, *et al.* PMFSNet: Polarized multi-scale feature self-attention network for lightweight medical image segmentation. *Computer Methods and Programs in Biomedicine*, 2025, 261: 108611. [doi: [10.1016/j.cmpb.2025.108611](https://doi.org/10.1016/j.cmpb.2025.108611)]
- 17 Liu X, Gao P, Yu T, *et al.* CSWin-UNet: Transformer UNet with cross-shaped windows for medical image segmentation. *Information Fusion*, 2025, 113: 102634. [doi: [10.1016/j.inffus.2024.102634](https://doi.org/10.1016/j.inffus.2024.102634)]
- 18 Zhang YD, Liu HY, Hu Q. TransFuse: Fusing Transformers and CNNs for medical image segmentation. *Proceedings of the 24th International Conference on Medical Image*

- Computing and Computer Assisted Intervention. Strasbourg: Springer, 2021. 14–24.
- 19 Cao H, Wang YY, Chen J, *et al.* Swin-Unet: Unet-like pure Transformer for medical image segmentation. Proceedings of the 2023 European Conference on Computer Vision. Tel Aviv: Springer, 2023. 205–218.
- 20 Wu HS, Chen SH, Chen GL, *et al.* FAT-Net: Feature adaptive Transformers for automated skin lesion segmentation. Medical Image Analysis, 2022, 76: 102327. [doi: [10.1016/j.media.2021.102327](https://doi.org/10.1016/j.media.2021.102327)]
- 21 Huang ZW, Deng HM, Yin SC, *et al.* ADF-Net: A novel adaptive dual-stream encoding and focal attention decoding network for skin lesion segmentation. Biomedical Signal Processing and Control, 2024, 91: 105895. [doi: [10.1016/j.bspc.2023.105895](https://doi.org/10.1016/j.bspc.2023.105895)]
- 22 Zhao HS, Shi JP, Qi XJ, *et al.* Pyramid scene parsing network. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 2881–2890.
- 23 Isensee F, Jaeger PF, Kohl SAA, *et al.* nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 2021, 18(2): 203–211. [doi: [10.1038/s41592-020-01008-z](https://doi.org/10.1038/s41592-020-01008-z)]
- 24 Huang G, Liu Z, Van Der Maaten L, *et al.* Densely connected convolutional networks. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 4700–4708.
- 25 Codella NCF, Gutman D, Celebi ME, *et al.* Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). Proceedings of the 15th IEEE International Symposium on Biomedical Imaging. Washington: IEEE, 2018. 168–172.
- 26 Codella N, Rotemberg V, Tschandl P, *et al.* Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). arXiv:1902.03368, 2019.
- 27 Qin XB, Zhang ZC, Huang CY, *et al.* BASNet: Boundary-aware salient object detection. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 7479–7489.
- 28 Wang HN, Cao P, Wang JQ, *et al.* UCTransNet: Rethinking the skip connections in U-Net from a channel-wise perspective with Transformer. Proceedings of the 36th AAAI Conference on Artificial Intelligence. AAAI Press, 2022. 2441–2449.
- 29 Sun GQ, Pan YZ, Kong WK, *et al.* DA-TransUNet: Integrating spatial and channel dual attention with Transformer U-Net for medical image segmentation. Frontiers in Bioengineering and Biotechnology, 2024, 12: 1398237. [doi: [10.3389/fbioe.2024.1398237](https://doi.org/10.3389/fbioe.2024.1398237)]
- 30 Tang H, Chen YB, Wang T, *et al.* HTC-Net: A hybrid CNN-Transformer framework for medical image segmentation. Biomedical Signal Processing and Control, 2024, 88: 105605. [doi: [10.1016/j.bspc.2023.105605](https://doi.org/10.1016/j.bspc.2023.105605)]
- 31 Fan DP, Ji GP, Zhou T, *et al.* PraNet: Parallel reverse attention network for polyp segmentation. Proceedings of the 23rd International Conference on Medical Image Computing and Computer-assisted Intervention. Lima: Springer, 2020. 263–273.
- 32 Wei J, Hu YW, Zhang RM, *et al.* Shallow attention network for polyp segmentation. Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Strasbourg: Springer, 2021. 699–708.
- 33 Ruan JC, Li JC, Xiang SC. VM-UNET: Vision mamba UNet for medical image segmentation. arXiv:2402.02491, 2024.

(校对责编: 张重毅)