

# 深度学习算法在皮肤病变区域分割中的研究进展<sup>①</sup>



王星皓, 刘 静, 仇大伟

(山东中医药大学 医学信息工程学院, 济南 250355)

通信作者: 刘 静, E-mail: [liuj\\_jn@163.com](mailto:liuj_jn@163.com)

**摘 要:** 随着计算机辅助诊断技术快速发展, 对皮肤病变筛查的效率明显提高, 但皮肤镜图像中众多干扰因素, 致使皮肤病变区域自动分割成为难题. 由于 Skip-Connecting 模型在小目标和模糊边界分割、Atrous-Convolution 模型在大范围和多尺度病变分割的良好表现, 成为皮肤病变图像分割的两大主流方向; Transformer 模型凭借其强大的全局建模能力和对长距离依赖关系的捕捉能力, 未来将与传统深度学习模型进一步耦合, 以优化分割性能. 基于此, 本文系统梳理 Skip-Connecting 模型与 Atrous-Convolution 模型在皮肤病变分割中的应用进展及其衍生网络的改进策略; 重点分析 Transformer 模型在该领域的应用范式及其与传统模型的耦合方法; 并对视觉状态空间模型 (Mamba)、生成对抗网络、扩散模型等新兴架构的探索性研究进行分析. 最后, 针对当前临床研究中存在的局限性及分割效果不佳的问题, 提出相应的解决思路, 并对未来的研究方向进行展望.

**关键词:** 图像分割; Transformer; 皮肤病变; 深度学习; 计算机视觉; 视觉状态空间模型

引用格式: 王星皓, 刘静, 仇大伟. 深度学习算法在皮肤病变区域分割中的研究进展. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/10121.html>

## Research Progress in Deep Learning Algorithms for Skin Lesion Segmentation

WANG Xing-Hao, LIU Jing, QIU Da-Wei

(School of Medical Informational Engineering, Shandong University of Traditional Chinese Medicine, Jinan 250355, China)

**Abstract:** The rapid development of computer-aided diagnosis technology has significantly improved the efficiency of skin lesion screening. However, the automatic segmentation of skin lesion remains challenging due to various interfering factors in dermoscopic images. Models based on Skip-Connecting have demonstrated strong performance in segmenting small targets and fuzzy boundaries, and those utilizing Atrous-Convolution excel in handling large-scale and multi-scale lesion segmentation, establishing these as two predominant approaches in skin lesion image segmentation. Furthermore, the Transformer model, renowned for its powerful global modeling and long-distance dependency capture capabilities, is poised for deeper integration with the traditional deep learning model to further optimize the segmentation performance. Accordingly, this review systematically outlines the application progress of the Skip-Connecting model and the Atrous-Convolution model in skin lesion segmentation and the improvement strategy of their derivative network. It focuses on analyzing the application paradigm of the Transformer model in this field and its coupling method with traditional models. Exploratory research on emerging frameworks, such as the visual state space model (Mamba), the generative confrontation network, and the diffusion model, is also examined. Finally, in light of the limitations in clinical research and the underlying causes of suboptimal segmentation effect, corresponding solutions are proposed, and future research directions are prospected.

**Key words:** image segmentation; Transformer; skin lesion; deep learning; computer vision; visual state space model

① 基金项目: 国家自然科学基金面上项目 (82174528); 山东中医药大学科学研究基金 (KYZK2024M14); 山东中医药大学研究生提质创新课题 (YJSTZCX 2025071, YJSTZCX2025069)

收稿时间: 2025-09-22; 修改时间: 2025-10-11, 2025-11-11; 采用时间: 2025-11-17; csa 在线出版时间: 2026-02-06

深度学习的快速发展推动了自动化解方案在生物医学诊断中的应用,显著提高了诊断的效率与准确性<sup>[1]</sup>。一般而言,疾病诊断的自动化解方案涉及图像分割、特征提取和图像分类这3个关键环节。其中,最为关键的一步是病变区域的自动分割,旨在将病变区域与正常组织清晰地区分开来。引入深度学习,通过其强大的学习能力、高效的特征提取能力和对图像复杂的非线性映射能力,可以从图像中自动提取由低到高的多层次特征。相比传统方法,极大地缩短了图像特征信息的提取时间,并显著提高了分割的精度。

皮肤病被认为是最普遍的健康问题之一,其分为多个种类,包括基底细胞癌、黑色素瘤、上皮内癌以及鳞状细胞癌<sup>[2]</sup>。全球范围内,皮肤癌困扰着许多人,根据世界卫生组织的报告,全球每年记录有200万-300万的非黑色素瘤皮肤癌病例和132000例黑色素瘤皮肤癌病例<sup>[3,4]</sup>。鉴于皮肤镜图像中病灶形态学的复杂多样性、图像边界的低饱和度以及视觉干扰因素的普遍存在,传统的图像分割方法往往难以达到高精度和鲁棒性的要求<sup>[5,6]</sup>。故许多研究者针对皮肤病变区域分割的相关工作进行了归纳综述,旨在分析皮肤镜图像研究工作的发展,加快推进皮肤病变分割领域的研究及临床应用。如张杰等<sup>[7]</sup>总结了卷积神经网络、残差网络和迁移学习方法在皮肤黑色素瘤病灶分割工作中的应用。孙晓楠等<sup>[8]</sup>从卷积神经网络与注意力机制这两种深度学习分割方法对皮肤镜图像分割研究进展工作进行了论述。尽管这些综述对皮肤病变分割进行了阶段性总结,但仍存在明显的覆盖局限:张杰等<sup>[7]</sup>的综述主要聚焦于传统CNN架构在黑色素瘤分割中的应用;孙晓楠等<sup>[8]</sup>则侧重于注意力机制的讨论。同时,这些工作存在3方面不足:(1)模型覆盖范围局限于传统CNN架构,未系统涵盖Transformer、Mamba等新兴范式;(2)缺乏对多模态融合、轻量化设计、生成模型等前沿方向的深入讨论;(3)对Skip-Connecting与Atrous-Convolution这两类主流方法的改进机制缺乏横向对比分析。

针对上述工作仍存在的局限,本文在系统归纳Skip-Connecting模型与Atrous-Convolution模型改进策略的基础上,重点补充了Transformer模型的多样化融合范式,并将视觉状态空间模型(Mamba)、扩散模型等新兴技术纳入综述范围,为研究者提供一个涵盖经典方法与前沿探索的完整技术图谱。

本文整体结构按照经典架构-混合范式-新兴技术-未来展望的技术演进顺序。首先聚焦Skip-Connecting结构,详细分析U-Net、SegNet及其衍生网络中的优化策略,并探讨Atrous-Convolution系列模型在处理大范围病变和尺度变化等挑战方面的成效。之后主要分析Transformer在皮肤图像分割中的应用价值,包含其独立架构的改进措施及CNN与Transformer的融合样式。在第4节进一步补充视觉状态空间模型(Mamba)、生成对抗网络及扩散模型等新兴架构的前沿动态。最后归纳当前研究尚存的局限,并就未来可能的研究方向提出前瞻性看法。本文呈现了从传统CNN到Transformer等新型架构的纵向演进路径,也囊括了不同模型相互的横向比较,期望为该领域的研究者提供系统而完备的方法借鉴。

## 1 Skip-Connecting模型在皮肤镜图像分割中的应用

### 1.1 基于U-Net网络的皮肤镜图像分割

U-Net<sup>[9]</sup>采用对称解码器结构,通过跳级连接与编码器紧密耦合。其解码器执行上采样恢复分辨率时,融合对应编码器层的特征图。该设计使解码器在提升分辨率的同时,复用编码器中的空间细节,有效补偿上采样的信息损失,从而生成更精细准确的分割结果,结构如图1所示。

UNet在处理包含复杂结构和细微特征的医学图像时,分割精度常难以满足要求。为缓解这一问题,Abid等<sup>[10]</sup>提出Double U-Net,通过叠加两阶段U-Net结构并引入ASPP模块以增强多尺度特征提取能力,但该方法因缺乏动态注意力机制,仍易受复杂背景干扰。在此基础上,Rehman等<sup>[11]</sup>构建了Attention Res-UNet,通过嵌入注意力门控机制动态聚焦目标区域并抑制背景噪声,同时引入残差连接以缓解梯度消失的问题,然而其卷积操作在跨尺度特征融合方面仍显不足。Saadati等<sup>[12]</sup>进一步引入自注意力机制,借助Transformer模块建模全局上下文依赖关系,并结合多尺度空洞注意力降低计算开销。为兼顾模型精度与推理效率,Le等<sup>[13]</sup>设计了轻量化架构MAA-UNet,编码器加入抗锯齿层以减少下采样过程中的偏移方差,解码器融入注意力模块以强化关键特征。针对CNN在长距离建模方面的不足及Transformer的计算瓶颈,Ruan等<sup>[14]</sup>提出基于状态空间模型的VM-UNet,其视觉状态空间块能够高

效捕捉全局与局部特征, 并采用非对称编解码器结构优化多分辨率特征重建. 为进一步提升编码器与解码器间的信息传递效率, Amin 等<sup>[15]</sup>提出一种边界感知分割网络, 通过融合 CNN 与 Transformer 的混合架构, 结合边界注意力模块增强跨层特征一致性. 同时在跳跃

连接处嵌入边界注意力机制, 以动态调整特征权重, 此设计显著提升了边缘分割精度. 此外, 网络还引入深度监督机制, 通过在多个解码器层级添加辅助损失函数, 增强中间层特征表达能力, 从而有效缓解了梯度消失问题.

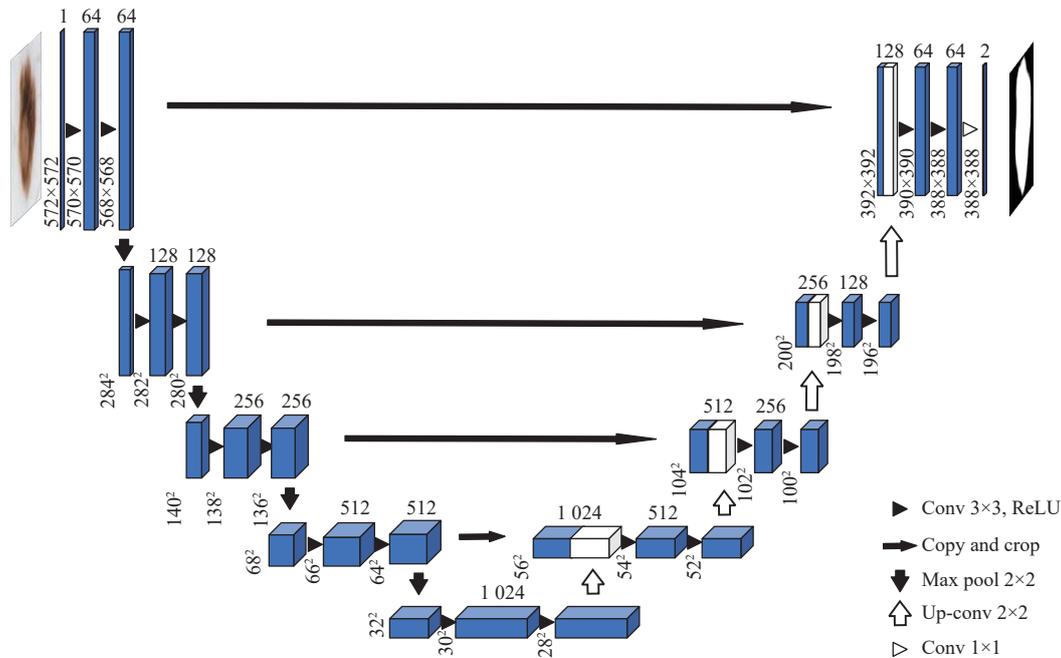


图1 U-Net 网络模型结构图

除了网络结构的优化, 边界精细化策略也成为 U-Net 改进的重要方向. Mirikharaji 等<sup>[16]</sup>系统分析了深度学习在皮肤病变分割中的进展, 指出自然和人工伪影的存在使得皮肤病变分割成为具有挑战性的任务, 强调了显式边界检测分支与分割分支的双任务学习框架能够增强边界表达能力, 通过边界响应图引导分割分支聚焦边界区域, 显著提升边界分割的 F1 分数. 此外, 多任务联合优化策略也被证明能够提升 U-Net 的性能. Akram 等<sup>[17]</sup>提出结合 Mask R-CNN 和 ResNet50 实现皮肤病变的分割与分类联合优化. 同时, Sarwar 等<sup>[18]</sup>将蚁群优化算法引入 Hybrid ResUNet 模型进行超参数优化, 通过联合训练使分割性能得到提升. 实验表明, 联合训练策略能够实现任务间知识迁移, 证明了多任务学习在医学图像分析中的有效性. 这些研究共同推动了 U-Net 架构从单一分割任务向多任务、多模态融合方向的发展.

### 1.2 基于 SegNet 网络的皮肤镜图像分割

SegNet 是一种语义分割 CNN 架构<sup>[19]</sup>. 其结构如

图 2 所示, 编码器通过卷积和池化层提取多尺度特征. 关键创新在于池化时记录池化区域索引, 解码器使用非池化操作, 依据记录的索引进行确定性上采样, 恢复空间结构. 此机制简化解码器设计, 减少参数, 在提升推理速度并降低内存占用的同时保持精度.

在医学图像分割中, SegNet 对图像细微结构的捕捉能力存在一定局限. 为改善这一问题, Kuang 等<sup>[20]</sup>提出 BEA-SegNet, 引入体边解耦模块, 将卷积特征分别解析为区域特征与边缘特征, 从而增强对复杂医学图像的处理能力, 但该方法在精确定位病变区域方面仍显不足, 易受图像噪声干扰. 在此基础上, Taghizadeh 等<sup>[21]</sup>将 YOLOv3 与 SegNet 相结合, 先通过微调后的 F-YOLOv3 检测病变位置, 再将其输入 F-SegNet 完成像素级分割, 有效提升了定位与分割的协同性能. 针对病变形态与尺寸的个体差异, Kumar 等<sup>[22]</sup>提出的 Deep-SegNet 采用更深层次的特征提取机制, 结合模糊认知图<sup>[23]</sup>引入模糊权值以建模图像不确定性, 从而增强模型对形态变化的适应能力. 为改善 SegNet 作为“黑盒”

模型的可解释性问题, Narayanan 等<sup>[24]</sup>设计了 IARS SegNet, 具体的, 其在基线架构中融合了跳跃连接、残差卷积和全局注意力机制. 跳跃连接有助于强化轮廓细节的学习, 残差卷积支持更深层次的网络建模, 注意力机制能够提取更具判别力的特征图, 从而增强模型

在临床场景中的可解释性. 为进一步提升模型综合性能, Ahmed 等<sup>[25]</sup>提出 Twin-SegNet, 通过联合优化前景分割与背景组织分割两个任务, 增强模型对目标与背景的区分能力, 并引入小波卷积块以强化边缘信息的提取效果.

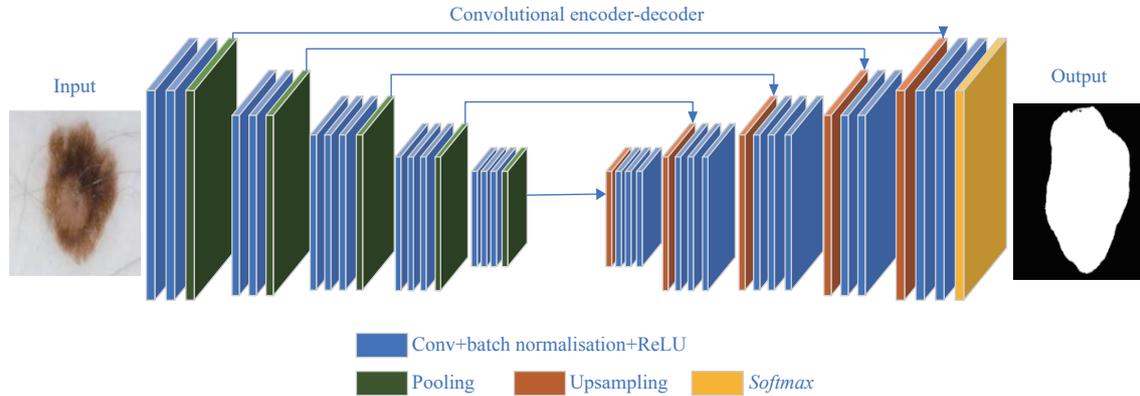


图2 SegNet 网络模型结构图

### 1.3 Skip-Connecting 模型改进方法小结

Skip-Connecting 模型通过引入注意力机制、多尺度模块、抗锯齿层等结构创新, 在细节保留、边界处理与模型轻量化方面取得了显著进展. 为更清晰地总结这些改进方法的共性、优势与局限, 本文在表 1

中归纳了常见结构改进策略的核心思想与典型算法实例; 并在表 2 中进一步对比了基于 Skip-Connecting 模型的不同方法在多个公开数据集上的性能表现, 以供研究者对 Skip-Connecting 模型的进一步优化展开研究.

表 1 结构改进常用方法总结

方法	核心思想	优点	缺点	算法实例
空洞空间金字塔池化	利用不同膨胀率空洞卷积并行提取多尺度特征, 结合全局平均池化获取多尺度上下文信息	多尺度特征提取; 有效扩大感受野; 保持分辨率	膨胀率设置敏感; 计算资源消耗大; 结构复杂度高	DU-Net <sup>[10]</sup>
注意力门机制	为输入特征分配权重, 突出显著特征, 抑制无关信息, 动态调整关注区域以提升特征表达	抑制背景噪声; 自适应动态调整特征权重; 模块化设计, 可移植性强	依赖上下文信息质量; 需要更精细的参数调节, 训练难度高	AttRes-UNet <sup>[11]</sup>
多尺度空洞注意力机制	结合多尺度空洞卷积与注意力机制, 提取多尺度特征并分配权重, 增强显著特征, 抑制无关特征	多尺度特征提取; 抑制无关特征; 对复杂场景的特征表达能力强	计算开销大; 参数优化难度高; 存在信息冗余风险	Dilated-UNet <sup>[12]</sup>
抗锯齿层	在下采样前引入低通滤波, 减少信息丢失和别名效应, 保留更多有效特征	减少信息丢失; 提升模型鲁棒性; 可无缝兼容现有深度学习架构	网络设计难度大; 细节信息损失; 额外计算成本高	MAA-UNet <sup>[13]</sup>
视觉状态空间块	投影视觉信息至隐状态空间利用循环结构建模动态序列, 结合显式状态更新捕获时序依赖与全局上下文	建模时序依赖; 增强全局信息表达; 参数效率高; 适应多模态输入	训练不稳定, 收敛难度大; 不适用于简单静态视觉任务	VM-UNet <sup>[14]</sup>
体边界偶块	分离与融合体特征和边界特征, 联合建模目标区域与边界信息, 提高分割准确性	边界精细度增强; 对复杂目标分割效果高; 可避免边界模糊与丢失	边界依赖性强, 在边界不清晰场景中效果受限; 计算成本高	BEA-SegNet <sup>[20]</sup>
模糊认知图块	构建加权概念节点因果图, 利用模糊逻辑推理表征复杂系统的模糊关系与动态演化	直观性强; 处理模糊信息能力强; 模拟系统动态变化, 适应复杂网络变化	依赖领域知识, 主观性强; 权重确定困难; 缺乏精确性	DeepSegNet <sup>[22]</sup>
背景前景联合优化	建模背景与前景关系, 整体优化, 利用背景增强前景区分性, 提升检测或分割性能	前景准确性高减少误检; 前景背景对比, 可提高边界精度; 适应复杂场景	数据依赖性强, 标签不准确时效果下降; 模型泛化能力受限; 模型训练难度大	Twin-SegNet <sup>[25]</sup>

表2 基于 Skip-Connecting 模型的皮肤镜图像分割方法总结

文献	针对问题	基础模型	改进方式	改进内容	数据集	算法性能
Abid等 <sup>[10]</sup>	细微特征分割的能力不足	U-Net	网络融合	双重U-Net网络叠加; 引入空洞空间金字塔池化	HAM10000	Dice (0.764 5)
Rehman等 <sup>[11]</sup>	关键区域定位问题	U-Net	结构改进	引入注意力门机制	ISIC2018	Dice (0.894 1); Jac (0.811 6)
Saadati等 <sup>[12]</sup>	远程依赖关系建模不足	U-Net	结构改进	融合自注意力Transformer 模块; 引入多尺度空洞注意力机制	ISIC2018	DSC (0.914 7); ACC (0.964 7)
Le等 <sup>[13]</sup>	锯齿效应与下采样偏差	U-Net	结构改进	集成抗锯齿层	ISIC2018	Dice (0.881); Jac (0.809)
Ruan等 <sup>[14]</sup>	上下文信息捕获不足	U-Net	结构改进	加入视觉状态空间模块	ISIC2017 ISIC2018	DSC (0.890 3); ACC (0.962 9) Dice (0.897 1); ACC (0.949 1)
Kuang等 <sup>[20]</sup>	整体与边界特征耦合问题	SegNet	结构改进	加入体边界偶模块	ISIC2018	Dice (0.903)
Taghizadeh等 <sup>[21]</sup>	位置检测与分割的协同问题	SegNet	网络融合	融合YOLO-v3网络	ISIC2018	Dice (0.928 1); ACC (0.951 6)
Kumar等 <sup>[22]</sup>	复杂病变特征处理困难	SegNet	结构改进	提高网络深度; 引入模糊认知图模块	ISIC2018	Dice (0.920); ACC (0.933)
Narayanan等 <sup>[24]</sup>	病变轮廓复杂; 深层模型训练难度大	SegNet	结构改进	加入跳跃连接、残差卷积、全局注意力机制	PH <sup>2</sup>	Dice (0.971 2); Jac (0.923 3)
Ahmed等 <sup>[25]</sup>	背景组织与前景目标的区分困难; 边缘信息提取不足	SegNet	结构改进	使用背景前景联合优化; 引入小波卷积块	ISIC2018	F1-score (0.916 1)

## 2 Atrous-Convolution 模型在皮肤镜图像分割中的应用

皮肤病变自动分割对提升临床诊疗效率与医学研究价值显著, 但其精准分割面临病灶复杂性、图像干扰等挑战. 传统方法(如区域生长<sup>[26]</sup>、强度阈值处理<sup>[27]</sup>)依赖人工特征设计, 抗干扰性差且缺乏统一标准. DeepLab 网络凭借更精细的分割效果与优越的多尺度处理能力成为主流解决方案.

DeepLab 系列的标志性模块 ASPP (atrous spatial pyramid pooling) 并行使用多个膨胀率的空洞卷积捕获多尺度特征. 空洞卷积 (atrous convolution, 又称膨胀卷积) 是通过在标准卷积核中引入空洞 (holes) 扩展感受野, 其核心优势在于无需增加参数量或降低分辨率即可捕获更大范围的上下文信息. 给定输入特征图  $x$  和卷积核  $w$ , 膨胀率为  $r$  的空洞卷积输出  $y[i]$  计算如下:

$$y[i] = \sum_{k=1}^K x[i+r \cdot k] \cdot w[k] \quad (1)$$

其中,  $K$  为卷积核大小,  $r$  控制采样间隔. 这种设计使模型能够同时捕获精细纹理与广域结构, 适用于皮肤病变形态多样、尺度跨度大的场景. 当前改进研究聚焦两大方向: 1) 单网络结构优化: 通过传统模块与卷积结构创新增强特征提取能力; 2) 多网络融合策略: 整合互补性网络架构以协同发挥基础网络与辅助网络优势.

### 2.1 基于单网络结构的 DeepLab 改进方法

为了提高分割准确度, Azad 等<sup>[28]</sup>提出 Attention DeepLabv3+ (Att-DeepLabv3+) 模型, 采用两个阶段的

注意力机制来分割皮肤病变. 此策略通过为每个通道分配权重来捕获一组特征图通道之间的关系. 通道注意力允许网络通过上下文门控机制对信息丰富且有意义的通道进行重点强调, 并利用二级注意力策略来整合不同层次的卷积, 这有助于网络专注于与目标更相关的视野, 分割效果如图 3 所示. 通过对比可以看出, Att-DeepLabv3+ 的分割结果与真实掩码 (ground truth) 更为接近, 其边界刻画更为精细, 包含了更多的局部细节, 同时有效避免了图像内视觉干扰的误判, 证明了注意力机制在提升分割精度方面的有效性. 针对边缘特征提取仍存在不足的问题, Tuyet 等<sup>[29]</sup>提出 CRF 优化的 DeepLab, 引入条件随机场概率图模型, 利用像素间依赖关系及位置语义信息对预测结果进行精细化处理, 显著提升边缘细节的平滑度与准确性.

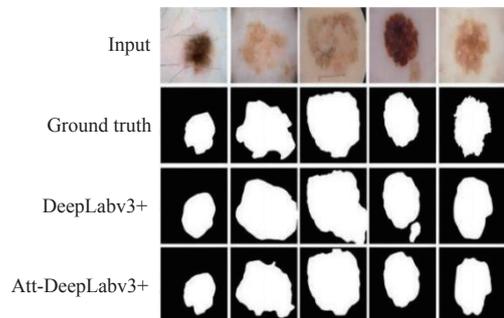


图3 不同网络的分割效果

### 2.2 基于多网络融合的 DeepLab 改进方法

在 YOLO 网络<sup>[30-32]</sup>思路指导下, Bagheri 等<sup>[33]</sup>提出 YOLO-DeepLab 两阶段模型: 第 1 阶段用 CNN 估

计病变位置尺寸,第2阶段通过归一化网络(融合YOLO-v3<sup>[34]</sup>网络与DeepLabv3+双网络输出)标准化输入,提升DeepLabv3+分割精度.针对传统2DIA方法手工特征局限,Rew等<sup>[35]</sup>设计混合分割方案:采用DeepLabv3+(主干为Inception-ResNet-v2<sup>[36]</sup>)进行像素级分割,结合LightGBM<sup>[37]</sup>与形态学处理<sup>[38]</sup>提高分割质量并提取皮肤特征,但是存在样本不均衡问题.Bagheri等<sup>[39]</sup>提出Retina-DeepLab:利用RetinaNet<sup>[40]</sup>的Focal loss解决正负样本失衡问题,先检测病变位置并裁剪标准化图像,再由DeepLabv3+分割.为提升分割不确定性处理能力,Czajkowska等<sup>[41]</sup>串联双DeepLabv3+模型:首模型定位ROI,次模型结合ResNet-50主干与模糊连接性分析模块实现精细分割.针对计算效率问题,Zafar等<sup>[42]</sup>融合轻量化预训练模型MobileNetV2<sup>[43]</sup>与DeepLabv3+,通过迁移学习提取深层特征并优化边界分割,使

模型能够在移动设备上实现实时分割,为皮肤病变筛查的普及化提供了技术支持.

### 2.3 Atrous-Convolution 模型改进方法小结

基于Atrous-Convolution的模型通过扩展感受野和融合多尺度特征,显著增强了对大范围及形态复杂病变的分割性能.当前相关改进可归纳为两类主要路径:一是对单一网络结构进行优化,如引入注意力机制、嵌入Transformer模块或结合条件随机场等方法;二是探索多网络融合策略以集成互补优势.不同方法在提升分割精度、强化边缘细节以及增强模型鲁棒性等方面展现出各自特点.为系统评估其技术特性,表3从改进内容与针对问题等维度对比了多种改进模型的优缺点,表4进一步列出了各模型在PH<sup>2</sup>、ISIC2017两个公共数据集上的量化结果,为客观比较其实际性能提供了依据.

表3 基于Atrous-Convolution的改进方法对比

算法模型	针对问题	改进内容	优点	缺点
Attention-DeepLabv3+ <sup>[28]</sup>	特征选择不当与特征融合信息冗余问题	引入两阶段注意力策略	增强全局与局部信息交互;提升特征表达能力;多尺度特征融合;边缘信息增强	计算复杂度高;对数据依赖性强;参数调试难度大
HSCRF-BN-DeepLab <sup>[29]</sup>	边缘细节处理不足;语义信息与像素位置关系的融合不足	输出预测阶段添加条件随机场	细化边缘;增强小目标表现;消除孤立像素点特征;整体分割,全局一致性强	主干网络和后处理优化目标不一致性;局限于边缘与小目标
YOLO-DeepLab <sup>[33]</sup>	病灶大小与位置的敏感性	融合YOLO-v3对分割前阶段进行归一化处理	提高目标定位精度;减少计算冗余;增强对小目标的分割性能;复杂场景适应性提升	误检目标影响分割性能;不适合无目标(全景分割)场景
HSENet <sup>[35]</sup>	分割结果细化与特征提取相互影响问题	模型分割阶段融合Inception-ResNet-v2网络后处理引入LightGBM	多尺度特征提取增强;特征表达能力提升;改善小目标识别效果;提升整体分割性能	优化效果不确定性;系统复杂度高;泛化能力依赖
RetinaNet <sup>[39]</sup>	正负样本分布不均;多任务协作问题	检测阶段级联RetinaNet网络;引入Focal loss作为损失函数	强化小目标检测能力;缓解类别不平衡问题;提高多任务性能;模型鲁棒性强	优化过程复杂;对标签质量敏感;多任务共享特征时存在任务间冲突
FEDL-CC <sup>[41]</sup>	感兴趣区域检测不足;边界模糊细节丢失	双DeepLabv3+模型进行模糊连接	多视角特征融合,提升复杂场景的识别能力;强化上下文信息;边界分割准确性高	模型推理时间长;额外集成资源消耗大;加剧过拟合风险;训练复杂度高
SMA-DeepLab <sup>[42]</sup>	模型复杂度高	引入MobileNetV2作为预训练模型	降低模型复杂度;模型推理速度快;迁移学习能力强;硬件需求低	特征表达能力受限;全局上下文信息不足;面对高精度分割任务准确度低

表4 不同网络改进模型的性能对比

文献	PH <sup>2</sup> 数据集				ISIC2017数据集		
	Specificity	Accuracy	Jaccard	Dice	Specificity	Accuracy	Jaccard
Azad等 <sup>[28]</sup>	0.9896	0.9657	0.9657	—	0.9901	0.9698	0.9698
Tuyet等 <sup>[29]</sup>	—	—	—	—	—	0.9543	—
Bagheri等 <sup>[33]</sup>	—	—	—	0.8757	0.9608	0.9429	0.7996
Bagheri等 <sup>[39]</sup>	—	—	—	—	—	—	0.8004
Czajkowska等 <sup>[41]</sup>	—	—	0.8985	0.8749	0.9651	0.9418	0.8004
Zafar等 <sup>[42]</sup>	—	0.9591	0.9048	—	—	0.9591	0.9048

### 3 Transformer 在皮肤镜图像分割中的应用

Transformer 凭借其自注意力机制与全局上下文建模能力, 为皮肤镜图像分割提供了新范式<sup>[44]</sup>. 其核心价值在于解决病灶边界模糊、形态不规则及复杂纹理等挑战. Transformer 架构 (如图 4 所示) 的核心在于自注意力机制 (self-attention), 其通过建模序列内元素间的全局依赖关系突破了卷积操作的局部性约束. 在视觉任务中, 输入图像首先被划分为不重叠的 patch, 并线性投影为特征嵌入  $X \in R^{N \times D}$  ( $N$  为 patch 数量,  $D$  为特征维度). 自注意力机制通过可学习矩阵  $W_Q, W_K, W_V \in R^{D \times d_k}$  将  $X$  映射为查询 (query,  $Q$ )、键 (key,  $K$ )、值 (value,  $V$ ) 表示:

$$Q = XW_Q, K = XW_K, V = XW_V \quad (2)$$

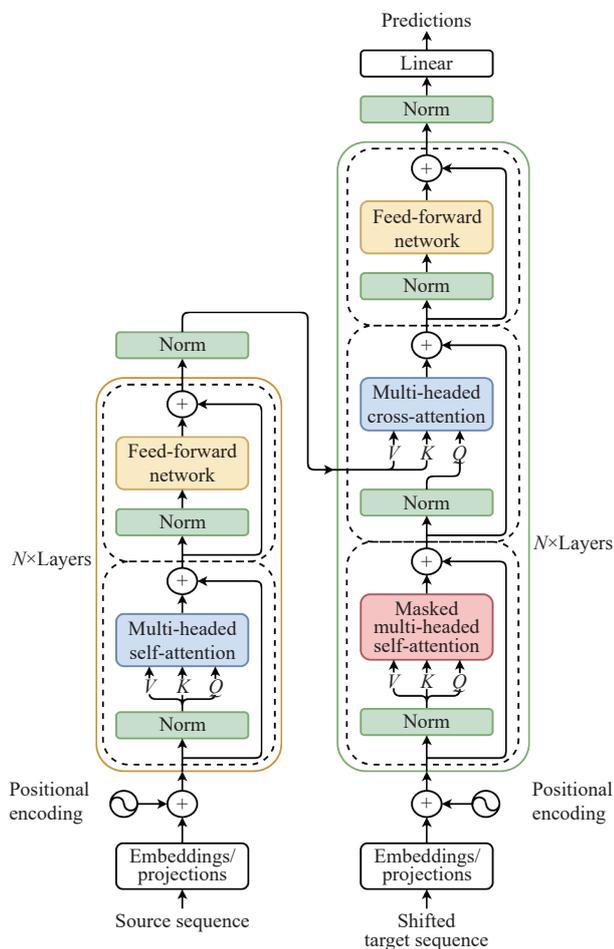


图 4 Transformer 模型架构

注意力权重通过缩放点积计算,  $Softmax$  归一化后与  $V$  加权求和:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

该机制使每个 patch 能够与全图所有 patch 建立直接联系, 有效捕获长程依赖.

Transformer 系列模型在皮肤病变区域分割中的应用主要分为两类: 单一 Transformer 架构 (直接处理图像块序列, 完全替代 CNN) 与 CNN-Transformer 耦合架构 (结合 CNN 局部特征提取与 Transformer 全局建模优势).

#### 3.1 Transformer 架构改进

单一 Transformer 架构在皮肤病变分割中面临局部细节提取不足的缺陷. Wang 等<sup>[45]</sup>提出 BATransformer, 集成边界注意力门生成块级注意力图引导模型聚焦模糊边界, 但在边界极度模糊或强干扰场景下失效. 为此, Xu 等<sup>[46]</sup>设计的 SkinFormer 利用峰度引导的统计计数算子来量化纹理特征, 并通过构建统计纹理增强 Transformer 融合多尺度统计信息, 但其丢失空间位置上下文, 难以建模语义区域的空间依赖. Xin 等<sup>[47]</sup>遂提出 SapFormer (结构如图 5 所示): 采用多混合 Transformer 编码器捕获多尺度位置特征, 基于 Transformer 解码器的自适应特征框架动态生成位置感知参数以优化区域特征分布, 但其结构复杂且高分辨率细节处理能力受限. Yang 等<sup>[48]</sup>进一步开发 MSPMformer: 以 PVTv2 为编码器提取多尺度特征; 多尺度感知模块通过深度卷积聚焦局部特征解决色差问题; 全局自适应融合模块结合减少冗余, 在低维特征层实现全局自适应权重融合.

#### 3.2 CNN-Transformer 耦合进展

由于传统卷积神经网络在捕获全局上下文信息方面的固有局限性, Wu 等<sup>[49]</sup>提出了一种基于经典编码器-解码器架构的新型特征自适应 Transformer 网络 FAT-Net, 在保留经典 CNN 架构的同时集成了一个额外的 Transformer 分支, 以有效地捕获远程依赖关系和全局上下文信息, 通过无缝集成 CNN 和 Transformer 分支, 不仅可以提取丰富的局部特征, 还可以捕获重要的全局上下文信息. 受 CNN-Transformer 耦合方向的启发, Li 等<sup>[50]</sup>提出 SUTrans-NET, 采用 CNN 和 Transformer 并行双编码器结构, 在每一层进行动态交互融合图像信息. 该方法有效结合了 CNN 的局部特征提取能力和 Transformer 的全局依赖建模能力, 但在处理医学临床问题时存在数据量大和计算量大等问题, 基于此, Wu 等<sup>[51]</sup>提出基于递归门卷积的高阶空间交互的 HorUNet 模型, 并在跳跃连接部分加入多维融合机制,

形成 MHorUNet 模型架构. 递归门卷积的高阶交互机制不仅具有 Transformer 成功的关键因素, 而且保留了卷积本身的优异特性. 实验结果表明, 该方法在多个指

标上表现最佳. 这证实了该模型在医学图像分割任务中能够生成更准确的分割边界, 并在保持高精度的同时展现出更强的泛化能力.

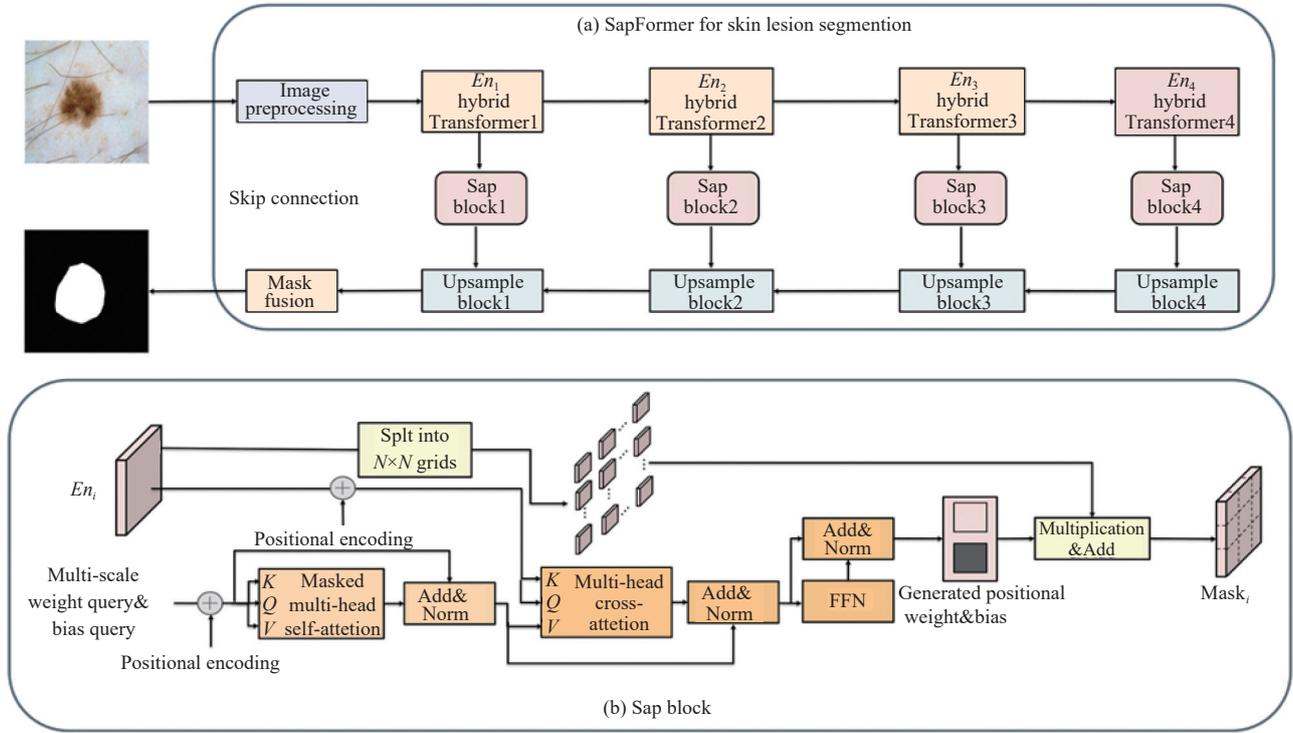


图 5 SapFormer 模型结构图

为突破 CNN-Transformer 混合架构的卷积瓶颈, Azad 等<sup>[52]</sup>提出了 TransDeepLab, 一种用于医学图像分割的新型 DeepLab-Transformer 架构. 具体来说, 此架构利用具有移位窗口的分层 Swin-Transformer 来扩展 DeepLabv3, 并对 Atrous 空间金字塔池化模块进行建模. Swin Transformer 引入窗口化自注意力 (window-based multi-head self-attention,  $W-MSA$ ) 与移位窗口机制 (shifted window MSA,  $SW-MSA$ ).  $W-MSA$  将特征图划分为  $M \times M$  不重叠窗口, 仅在窗口内计算注意力, 复杂度降至  $O(N)$ :

$$W-MSA(z^l) = Attention_{window}(z^l) + z^l \quad (4)$$

$SW-MSA$  通过循环移位窗口实现跨窗口信息交互:

$$SW-MSA(z^{l+1}) = Attention_{shifted-window}(shift(z^{l+1})) + z^{l+1} \quad (5)$$

这种层次化设计在保持线性复杂度的同时实现全局建模, 为医学图像分割提供了高效解决方案. 通过将编码器提取的特征视为序列, 对每个目标像素生成查

询向量 (query), 计算其与全图所有像素键向量 (key) 的点积得到注意力分数 (量化全局依赖关系强度), 经  $Softmax$  归一化为权重后加权聚合值向量 (value), 最终生成融合全局上下文的目标特征表示. 该机制使模型能精准建模病灶与正常组织的复杂空间关系. 但模型难以进行长距离特征建模, 且基于 Transformer 的模型往往不太关注局部信息, 从而导致边界精度较低.

为了解决上述问题, Peng 等<sup>[53]</sup>提出了一种融合 Transformer 和 U-Net 的双分支网络 DBTU-Net, 利用注意力密集的 U-Net 来捕捉局部特征, 并采用视觉 Transformer 对图像的长期依赖关系和局部贡献分数进行建模, 从而实现局部和全局特征的综合提取. 注意力密集 U-Net 包括一个三重融合注意力模块, 该模块可提取高度、宽度和通道维度的特征, 帮助 U-Net 捕获通道和空间位置之间的相互依赖关系. U-Net 模型对卷积运算的依赖阻碍了其捕获对准确医学图像分割至关重要的长距离依赖关系的能力. 此外, U 形结构受到编码器和解码器之间语义差距的阻碍.

基于此, Eskandari 等<sup>[54]</sup>通过设计一种尺度间上下文融合方法, 在编码器的每个阶段使用注意力相关性来自适应地组合每个阶段的上下文, 改进传统跳过连接路径中使用的典型串联过程以减少语义差距, 但其刚性相似度量无法适应病变形变, 且缺乏跨阶段长程空间关系建模. 针对此类问题, Cai 等<sup>[55]</sup>提出了一种基于 Transformer U-Net 的皮肤癌图像精确驱动分割模型, 称为 BiADATU-Net, 它将可变形的注意力 Transformer 和双向注意力块集成到 U-Net 中. 编码器部分利用具有双注意力块的可变形注意力 Transformer, 允

许自适应学习全局和局部特征.

### 3.3 Transformer 模型优化方法小结

Transformer 模型凭借其强大的全局建模能力, 为皮肤病变分割提供了新的解决路径. 本节从单一 Transformer 架构与 CNN-Transformer 耦合架构两个方向, 综述了其在皮肤镜图像分割中的最新进展. 为系统比较 Transformer 类算法的技术特点与性能表现, 本文在表 5 中总结了各优化方法的技术优势与局限; 并在表 6 中汇总了多个数据集上的详细性能指标, 为后续研究提供了实证依据与对比基线.

表 5 用于皮肤病变分割的 Transformer 算法对比

优化方法	针对问题	技术优势	技术局限
BATransformer <sup>[45]</sup>	Transformer难以处理模糊边界	引入边界注意力门, 通过识别模糊边界生成注意力图; 结合全局依赖建模与局部边界先验, 提升边界分割精度	高度依赖边界识别的准确性, 在低对比度、强干扰场景下边界引导易失效
SkinFormer <sup>[46]</sup>	病变区域的统计纹理特征忽略	设计峰度引导统计算子, 量化纹理分布特征; 通过统计纹理融合模块, 显式建模纹理信息	过度依赖统计特征, 丢失像素间空间关系与语义位置差异
SapFormer <sup>[47]</sup>	模型难以适应不同位置的特征分布差异	设计自适应位置感知解码器, 动态生成位置相关参数; 显式建模空间关系, 优化区域特征分布	结构复杂, 计算开销大, 动态参数机制导致训练不稳定; 对极细微局部特征的捕捉效率有限
MSPMformer <sup>[48]</sup>	多尺度特征融合效率低, 易受毛发干扰	多尺度感知模块: 聚焦局部细节; 全局自适应融合: 抑制冗余特征, 增强多尺度融合	计算复杂度高; 卷积的局部性限制超细微结构建模
FAT-Net <sup>[49]</sup>	传统CNN难以建模长距离依赖, 单一Transformer架构易丢失局部细节	双分支架构, 同步提取局部特征与全局上下文	双分支特征仅简单拼接, 缺乏深度交互机制
TransDeepLab <sup>[52]</sup>	卷积架构限制病变多尺度建模与长距依赖捕获能力	分层Swin-Transformer消除卷积瓶颈; 多速率窗口注意+空洞采样高效多尺度全局建模	移位窗口自注意力计算复杂度高, 显存消耗大; 对显微级结构分割精度有
DBTU-Net <sup>[53]</sup>	现有模型对病变局部异质性建模不足, 全局-局部特征融合僵化	高度、宽度、通道维度捕捉空间-通道依赖; 利用ViT输出动态加权关键区域特征	跳跃连接采用原始串联, 多尺度特征对齐能力弱; 注意力模块未适配不同分辨率特征, 小目标分割精度波动大
ISCFUT <sup>[54]</sup>	多尺度特征融合噪声大, 引发病变边界断裂	通过注意力相关性图自适应加权融合相邻尺度特征; 替代串联操作, 显著减少编码器-解码器语义差距	注意力相关性依赖固定余弦相似度, 无法适应非刚性形变; 仅融合相邻尺度, 跨阶段长程依赖建模缺失
BiADATU-Net <sup>[55]</sup>	注意力机制对病变形变适应性差, 空间拓扑关系建模不足	动态调整注意力焦点, 适应几何形变; 联合全局-局部路径建模空间约束; 筛选位置敏感特征	卷积骨架的平移不变性限制位置敏感特征表达; ASPP模块的空洞卷积计算冗余, 多尺度效率低

表 6 Transformer 相关算法性能对比

算法	数据集	评价指标					
		Sensitivity	IoU	Accuracy	Specificity	Dice	Jaccard
BAT <sup>[45]</sup>	ISIC2017	0.893 8	0.810 7	0.965 2	0.980 1	0.895 5	—
	ISIC2018	0.904 6	0.821 8	0.952 2	0.967 6	0.902 2	—
SkinFormer <sup>[46]</sup>	PH <sup>2</sup>	—	87.4	—	—	92.7	—
	ISBI2018	—	87.6	—	—	93.2	87.9
SapFormer <sup>[47]</sup>	ISIC2017	—	0.864	0.943	—	0.926	—
	ISIC2018	—	0.894	0.957	—	0.943	—
MSPMformer <sup>[48]</sup>	ISIC2018	—	—	0.962 3	—	0.929 6	0.876 0
FAT-Net <sup>[49]</sup>	ISIC2017	—	0.765 3	0.932 6	—	0.850 0	—
	ISIC2018	—	0.820 2	0.957 8	—	0.890 3	—

表 6 Transformer 相关算法性能对比 (续)

算法	数据集	评价指标					
		Sensitivity	IoU	Accuracy	Specificity	Dice	Jaccard
TransDeepLab <sup>[52]</sup>	ISIC2017	0.8971	—	0.9705	0.9886	0.9239	—
	ISIC2018	0.8756	—	0.9654	0.9889	0.9122	—
	PH <sup>2</sup>	0.9161	—	0.9657	0.9896	0.9456	—
DBTU-Net <sup>[53]</sup>	ISIC2017	0.9005	—	0.9680	0.9837	0.9168	—
	ISIC2018	0.8939	—	0.9647	0.9818	0.9093	—
	PH <sup>2</sup>	0.9471	—	0.9623	0.9690	0.9401	—
ISCFUT <sup>[54]</sup>	ISIC2017	0.9321	—	0.9698	0.9793	0.9257	—
	ISIC2018	0.9284	—	0.9630	0.9723	0.9136	—
BiADATU-Net <sup>[55]</sup>	PH <sup>2</sup>	0.9496	—	0.9687	—	0.9463	0.8998
	ISIC2017	0.8727	—	0.9618	—	0.8990	0.8270
	ISIC2018	0.9137	—	0.9626	—	0.9008	0.8343

### 4 新兴架构探索

除传统 CNN、Transformer 架构外,生成对抗网络、扩散模型和视觉状态空间模型等新兴方法为皮肤病变分割提供了新思路,特别是在小样本学习、边界精细化、长程依赖建模等方面展现出独特优势。

#### 4.1 生成对抗网络 (GAN) 辅助分割

生成对抗网络通过生成器 (generator) 与判别器 (discriminator) 的对抗训练,能够增强分割网络对病变边界的判别能力,对抗损失函数定义为:

$$L_{GAN} = E_{x,y}[\log D(x,y)] + E_x[\log(1 - D(x,G(x)))] \quad (6)$$

其中,  $x$  为输入图像,  $y$  为真实掩码,  $G$  为生成器,  $D$  为判别器. 该损失函数的第 1 项鼓励判别器对真实样本对做出正确判断, 第 2 项同时训练生成器  $G$  生成逼真的分割图  $G(x)$  以欺骗判别器  $D$ , 并训练判别器  $D$  能够识破这种欺骗。

Innani 等<sup>[56]</sup>提出基于 GAN 的皮肤病变分割框架 Efficient-GAN (EGAN) (模型架构如图 6 所示), 其模型

设计采用轻量化的生成器-判别器架构. 生成器采用基于挤压激励 (squeeze-excitation) 的复合缩放编码器, 配合非对称横向连接的解码器捕获密集特征; 判别器使用 PatchGAN 架构分类  $16 \times 16$  图像块来区分真实掩码与生成掩码. 该框架引入形态学平滑损失函数处理模糊边界. 其轻量化版本 MGAN 基于 MobileNetV2 架构, 仅包含 2.2M 参数, 推理速度 13 f/s, 为移动端部署提供了可行方案. 为了进一步提升分割效果, Zou 等<sup>[57]</sup>提出了一种基于对抗训练的全局和局部语义特征感知的生成对抗网络 GLSFA-GAN, 用于皮肤损伤分割. 具体地, 在生成器中, 设计了多尺度的局部特征融合模块和有效通道注意力模块, 用于获取皮损区域的多尺度局部详细信息. 此外, 在生成器的编码器和解码器之间的瓶颈处使用全局上下文提取模块来捕获更多关于病变的全局语义特征和空间信息. 之后, 使用对抗训练策略使判别器辨别生成的标签和分割预测图, 这有助于生成器产生更准确的分割图。

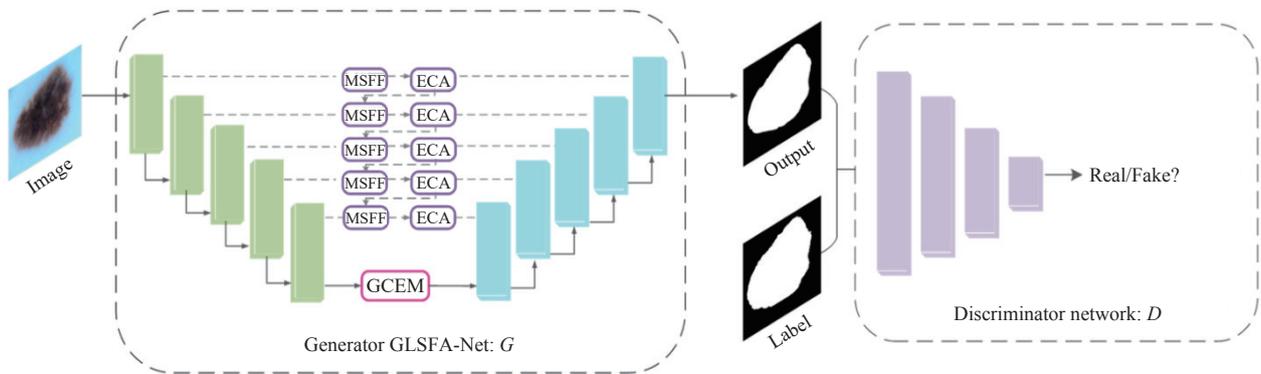


图 6 EGAN 模型结构图

#### 4.2 扩散模型的概率分割范式

扩散模型通过将分割任务建模为一个参数化的马

尔可夫链, 能够生成更精细、更具多样性的分割掩码. 其核心思想是实现从简单噪声分布到复杂目标掩码分

布的渐进式恢复,该过程包含前向扩散与反向去噪两个阶段.前向过程通过固定策略逐步向真实数据添加高斯噪声,最终将其转化为纯噪声:

$$q(x_t|x_{t-1}) = N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \quad (7)$$

其中,  $\beta_t$  为噪声调度参数,它随着时间步长  $t$  从  $\beta_1$  单调增加至  $\beta_T$ ,  $\beta_t$  的序列决定了噪声添加的速率和幅度.在皮肤镜图像分割任务中,由于需要保留病变区域的细微结构和模糊边界,  $\beta_t$  的设定通常采用余弦调度等相对平缓的策略,避免在早期扩散步骤中引入过强噪声导致关键纹理信息的过早丢失.这种平缓的噪声注入方式有助于模型在后续的去噪过程中,更精细地恢复出病变的形态细节.反向去噪过程则旨在学习一个参数化的生成模型,该模型从纯噪开始,通过另一个马尔可夫链逐步去除噪声,最终还原出高质量的分割掩码.该过程通过神经网络预测每一步所添加的噪声或均值,其训练目标通常为最小化预测噪声与真实噪声之间的均方误差.通过这种结构化的渐进式生成,扩散模型在医学图像分割中能够有效建模复杂的条件数据分布,从而在保持结构合理性的同时捕捉到病变区域的精细细节和不确定性.

Wu 等<sup>[58]</sup>提出的 Medical SAM adapter 将 SAM (segment anything model) 与扩散模型结合,该方法通过适配器模块将 SAM 的视觉先验知识迁移至医学图像分割,扩散过程逐步细化初始掩码,在边界模糊的病变区域表现优异.但其迭代去噪过程导致计算开销较大,实时分割应用受限.

为提升计算效率, Saharia 等<sup>[59]</sup>开发了级联扩散分割模型 CDS,采用由粗到精的两阶段架构.该模型首先在低分辨率 ( $64 \times 64$ ) 条件下通过扩散过程快速生成初步分割掩码,随后利用超分辨率扩散模块将其细化至高分辨率 ( $512 \times 512$ ) 结果.这一设计在保持生成质量的同时大幅压缩了推理耗时,为扩散模型在医学图像分割中的实际部署提供了可行思路.在此基础上, Ho 等<sup>[60]</sup>进一步提出无分类器引导扩散模型 CFG-Diff,通过在生成过程中引入临床标签(如良恶性分类)作为条件信号,引导模型生成与特定类别语义一致的分割结果.这一条件生成机制不仅增强了对输出内容的控制能力,也为处理类别分布不均衡的医学数据集提供了新的技术路径.

### 4.3 视觉状态空间模型 (Mamba)

视觉状态空间模型通过状态空间方程建模序列数

据,其核心在于将一维序列的演化过程模拟为一个隐式动态系统,通过隐状态  $h_t$  的连续性传递,能够有效捕获数据中的长程依赖关系.其中, Mamba 架构采用选择性状态空间机制 (selective SSM),通过可学习选择门控制信息流:

$$h_t = \bar{A}h_{t-1} + \bar{B}x_t, y_t = Ch_t \quad (8)$$

其中,  $\bar{A} = \exp(\Delta A)$ ,  $\bar{B} = \Delta B$ ,  $\Delta$  为时间步长参数.这种选择性机制赋予了模型强大的上下文感知能力,使其能够根据图像内容,自适应地将计算资源聚焦于病变关键区域等信息丰富的部分,并有效抑制背景噪声的干扰,从而在保证高效计算的同时,显著提升了对复杂病理结构的建模精度与鲁棒性.

Mamba 模型凭借其线性计算复杂度,在处理高分辨率医学图像时展现出显著的效率优势,尤其在边界精细化任务中,其分割精度超越了传统 Transformer 架构. Ma 等<sup>[61]</sup>提出的 U-Mamba 模型(结构见图 7)引入了双向状态空间扫描机制,通过前向与后向两个方向的语义建模增强了特征表达能力.

为进一步提升多尺度建模能力, Wang 等<sup>[62]</sup>开发了 Mamba-UNet,采用层次化状态空间模块与跨层融合机制.该模型通过在不同分辨率层级部署参数独立的状态空间模块,并借助可学习的状态映射矩阵促进跨尺度信息交互,使 Mamba 在保持线性复杂度的同时实现了与 CNN 和 Transformer 相当的多尺度特征提取能力.在模型基础架构的持续优化中, Dao 等<sup>[63]</sup>提出了 Mamba-2,针对选择性扫描机制进行改进.该模型引入多头选择性扫描架构,使每个注意力头能够学习不同的扫描模式(水平、垂直、对角线或螺旋方向),再通过注意力机制进行特征聚合.这种设计使模型能够从多视角捕获图像特征,显著增强了对复杂病变形态的建模能力.

### 4.4 新兴架构的协同潜力

综上所述,生成对抗网络 (GAN) 通过对抗训练机制有效增强边界判别能力,并为数据增强提供了新途径;扩散模型利用其概率建模特性,能够生成细节丰富的分割掩码,同时有效处理医学图像中的不确定性因素, Mamba 模型以线性计算复杂度实现了高效的全局依赖建模,为处理高分辨率医学图像提供了可行的技术方案.这些创新共同推动着皮肤病变分割技术向更高精度、更强泛化能力和更优计算效率的方向发展.

进一步的研究可关注这些架构的协同融合,从而在保持分割精度的同时,显著提升模型的计算效率与鲁棒

性.如表7所示,这类融合创新有望为皮肤病变分割带来新的技术突破.

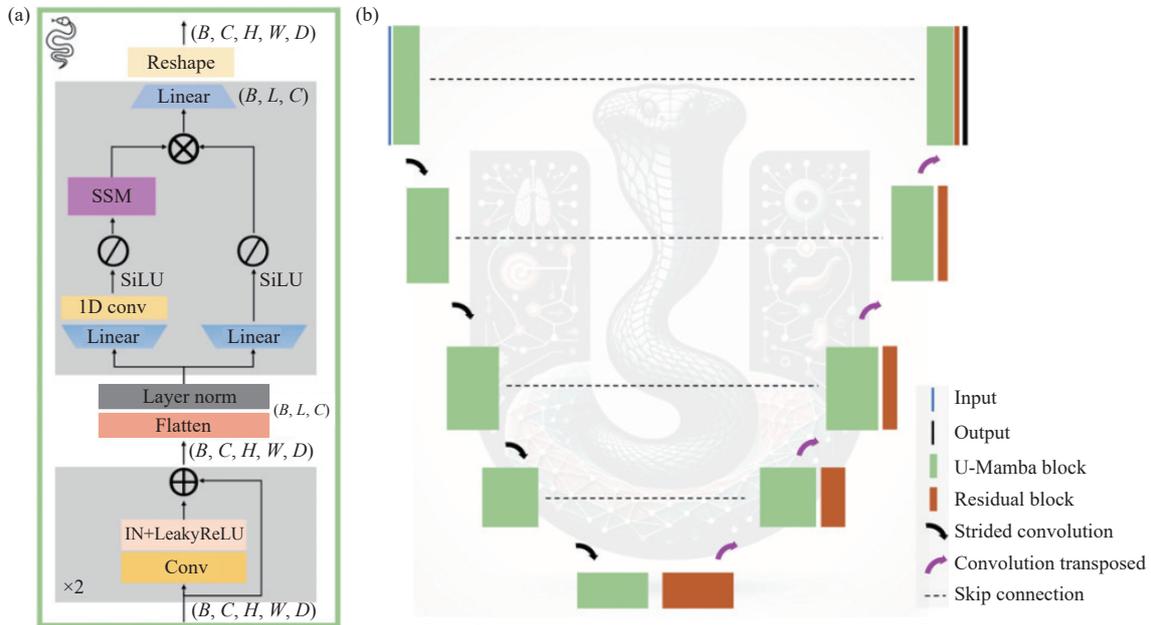


图7 U-Mamba 模型结构图

表7 新兴架构协同融合分析

融合方案	技术路线	技术优势	技术局限
GAN+Mamba	Mamba生成器替代传统CNN生成器;保持PatchGAN判别器	线性复杂度提升生成效率;全局依赖增强生成质量;对抗训练保持边界判别能力	梯度冲突:对抗梯度与状态空间平稳梯度冲突,易致模式崩溃 状态不稳:对抗训练干扰Mamba隐状态的长期一致性
Diffusion+Transformer	扩散模型生成多假设掩码;Transformer全局优化选择最优假设	概率建模提供多样化候选;Transformer全局推理增强鲁棒性	计算开销大:扩散迭代与Transformer二次复杂度串联,显存与时间成本高 特征对齐困难:扩散过程的多时间步特征难以有效作为Transformer条件输入
Mamba+Diffusion	Mamba编码器提取特征;扩散解码器细化掩码	Mamba高效提取全局特征;扩散模型精细化边界;保持计算效率	条件映射低效:Mamba的单一隐状态难以指导复杂的多步扩散去噪过程 训练目标失衡:判别式编码器与生成式解码器的训练目标难以协同
GAN+Diffusion+Mamba	三者串联:Mamba特征提取→扩散初步分割→GAN边界细化	综合三者优势:效率+精度+鲁棒性	训练链脆弱:多阶段误差累积,端到端训练因损失函数博弈而难以收敛 超参调优复杂:多组件组合导致超参数搜索空间爆炸

### 5 总结与展望

通过将深度学习技术应用于皮肤病变区域的图像分割,不仅能提升诊断的效率,还能减少由医生主观因素引发的漏诊和误诊等问题;另外,皮肤镜图像的分割精度和准确性都得到了提升,这在临床治疗中起到了不可替代的作用.尽管如此,相关网络的应用研究仍然面临许多挑战和考验,基于深度学习的皮肤镜图像分割的研究工作还需进一步完善和优化,以解决目前仍

存在的问题,推动医学图像分割技术的持续进步.

(1) 在进行皮肤病变区域的划分时,由于皮肤的生理结构影响,病灶区域的大小、形态和位置都可能发生改变,这可能导致在医学影像中所观察到的目标区域外观发生显著变化,从而对临床分析产生影响.

(2) 由于在皮肤镜图像成像的过程中,光照强度的改变可能导致图像出现显著差异,同时还会受到毛发遮挡和噪声等外部因素的影响,因此深度学习模型在

分割皮肤病变区域的边缘时效果并不理想。

(3) 在进行皮肤病变区域图像分割的过程中,单一化的模型架构改进难以在保持高计算效率的同时,兼顾对高分辨率图像的长距离语义依赖建模以及对病变边界细节的精准划分,导致模型在复杂病理场景下的分割精度与视觉合理性均存在局限。

考虑到当前存在的临床和算法技术问题,本文结合已有的临床经验和算法改进方法,对深度学习在皮肤镜图像分割领域的未来研究方向进行了展望,认为未来的研究应该重点围绕以下几个方面进行改进。

(1) 由于人体皮肤具有独特的形态学结构,这使得从根本上解决分割易受非目标特征区域影响的问题很困难。未来可进一步探索生成对抗网络与扩散模型在数据合成与结构建模中的潜力。在 Skip-Connecting 架构中嵌入生成对抗机制,生成具备高真实性与结构多样性的皮肤病灶样本,可以有效扩充训练集并增强模型对形态变化的适应能力;扩散模型则可用于生成高质量、多形态的病变图像,甚至联合分割任务进行端到端训练,提升模型在复杂场景下的泛化性能。对于 Atrous-Convolution 类模型,可引入具有全局感知能力的视觉状态空间模型,通过其长序列建模优势捕捉远离区域的语义依赖,增强对病灶整体与局部上下文的理解,从而抑制非目标区域的干扰。

(2) 为了解决图像边缘模糊和噪声过大的问题,可以在图像预处理阶段对图像数据进行增强处理,或采用直方图均衡化技术重新分配图像的像素信息。实验证明,上述方法可以提高皮肤病变区的分割精度。此外,生成对抗网络和扩散模型也可用于图像去噪、超分辨率重建和边缘增强,生成更干净、结构更清晰的训练样本。在半监督与自监督学习中,可引入扩散式自监督预训练,通过对图像施加多级噪声并学习去噪过程,使模型掌握更鲁棒的语义特征。同时,基于 Mamba 的编码器因其在长程依赖建模上的优势,可在自监督对比任务中更好地保留空间与边缘一致性,提升细节分割质量。

(3) 为解决当前深度学习改进方式较为单一的问题,可以在架构层面推动更深层次的多元技术融合。在现有 CNN-Transformer 混合架构的基础上,引入视觉 Mamba 模型作为补充或替代方案,借助其线性计算复杂度有效建模长距离依赖,更好地适应高分辨率医学图像的处理需求。生成对抗网络和扩散模型也可被整

合至解码阶段,作为后处理模块或联合优化目标,从而显著提升边界预测的准确性与视觉合理性。

此外,应积极探索基于 Mamba 或扩散模型的跨模态融合机制,实现图像特征与临床文本、患者元数据等多源信息的有效融合。此类融合不仅能增强模型的可解释性,也有助于提升深度学习在真实临床场景中的适用性,为构建下一代皮肤病变分割系统奠定技术基础。

### 参考文献

- 1 Li YX, Shen LL. Skin lesion analysis towards melanoma detection using deep learning network. *Sensors*, 2018, 18(2): 556. [doi: 10.3390/s18020556]
- 2 Parisi R, Shah H, Everdell E, *et al.* First-time office visit for suspicious skin lesion evaluation as a predictor of high-risk melanoma. *Melanoma Research*, 2023, 33(6): 555–556. [doi: 10.1097/CMR.0000000000000930]
- 3 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, 2019, 69(1): 7–34. [doi: 10.3322/caac.21551]
- 4 Siegel RL, Miller KD, Wagle NS, *et al.* Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 2023, 73(1): 17–48. [doi: 10.3322/caac.21763]
- 5 Kirillov A, Mintun E, Ravi N, *et al.* Segment anything. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision*. Paris: IEEE, 2023. 3992–4003.
- 6 Khasanov R, Ramires A, Grinenko V, *et al.* In-plane magnetic penetration depth in Sr<sub>2</sub>RuO<sub>4</sub>: Muon-spin rotation and relaxation study. *Physical Review Letters*, 2023, 131(23): 236001. [doi: 10.1103/PhysRevLett.131.236001]
- 7 张杰, 赵惠军, 李贤威, 等. 基于深度学习检测方法检测皮肤镜图像中黑色素瘤的研究进展与展望. *医疗卫生装备*, 2018, 39(11): 90–95. [doi: 10.7687/j.issn1003-8868.2018.11.090]
- 8 孙晓楠, 陆奎, 陈晨, 等. 融合注意力卷积神经网络在皮肤镜分割中的应用研究进展. *沈阳医学院学报*, 2024, 26(5): 514–523. [doi: 10.16753/j.cnki.1008-2344.2024.05.014]
- 9 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.
- 10 Abid I, Almakdi S, Rahman H, *et al.* A convolutional neural network for skin lesion segmentation using double U-Net architecture. *Intelligent Automation & Soft Computing*, 2022, 33(3): 1407–1421.

- 11 Rehman A, Butt MA, Zaman M. Attention Res-UNet: Attention residual UNet with focal tversky loss for skin lesion segmentation. *International Journal of Decision Support System Technology (IJDSST)*, 2023, 15(1): 1–17.
- 12 Saadati D, Manzari ON, Mirzakuchaki S. Dilated-UNet: A fast and accurate medical image segmentation approach using a dilated Transformer and U-Net architecture. *arXiv:2304.11450*, 2023.
- 13 Le PT, Pham BT, Chang CC, *et al.* Anti-aliasing attention U-Net model for skin lesion segmentation. *Diagnostics*, 2023, 13(8): 1460. [doi: [10.3390/diagnostics13081460](https://doi.org/10.3390/diagnostics13081460)]
- 14 Ruan JC, Li JC, Xiang SC. VM-UNet: Vision mamba UNet for medical image segmentation. *arXiv:2402.02491*, 2024.
- 15 Amin J, Azhar M, Arshad H, *et al.* Skin-lesion segmentation using boundary-aware segmentation network and classification based on a mixture of convolutional and Transformer neural networks. *Frontiers in Medicine*, 2025, 12: 1524146. [doi: [10.3389/fmed.2025.1524146](https://doi.org/10.3389/fmed.2025.1524146)]
- 16 Mirikharaji Z, Abhishek K, Bissoto A, *et al.* A survey on deep learning for skin lesion segmentation. *Medical Image Analysis*, 2023, 88: 102863. [doi: [10.1016/j.media.2023.102863](https://doi.org/10.1016/j.media.2023.102863)]
- 17 Akram A, Rashid J, Jaffar MA, *et al.* Segmentation and classification of skin lesions using hybrid deep learning method in the Internet of Medical Things. *Skin Research and Technology*, 2023, 29(11): e13524. [doi: [10.1111/srt.13524](https://doi.org/10.1111/srt.13524)]
- 18 Sarwar N, Irshad A, Naith QH, *et al.* Skin lesion segmentation using deep learning algorithm with ant colony optimization. *BMC Medical Informatics and Decision Making*, 2024, 24(1): 265. [doi: [10.1186/s12911-024-02686-x](https://doi.org/10.1186/s12911-024-02686-x)]
- 19 Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481–2495. [doi: [10.1109/TPAMI.2016.2644615](https://doi.org/10.1109/TPAMI.2016.2644615)]
- 20 Kuang HL, Liang YX, Liu N, *et al.* BEA-SegNet: Body and edge aware network for medical image segmentation. *Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Houston: IEEE, 2021. 939–944.
- 21 Taghizadeh M, Mohammadi K. The fast and accurate approach to detection and segmentation of melanoma skin cancer using fine-tuned YOLOv3 and SegNet based on deep transfer learning. *arXiv:2210.05167*, 2022.
- 22 Kumar KA, Vanmathi C. Segmentation and detection of skin cancer using fuzzy cognitive map and deep Seg Net. *Soft Computing*, 2024, 28(5): 4575–4592. [doi: [10.1007/s00500-024-09644-9](https://doi.org/10.1007/s00500-024-09644-9)]
- 23 Stylios CD, Georgopoulos VC, Malandraki GA, *et al.* Fuzzy cognitive map architectures for medical decision support systems. *Applied Soft Computing*, 2008, 8(3): 1243–1251. [doi: [10.1016/j.asoc.2007.02.022](https://doi.org/10.1016/j.asoc.2007.02.022)]
- 24 Narayanan VS, Sikha OK, Benitez R. IARS SegNet: Interpretable attention residual skip connection SegNet for melanoma segmentation. *IEEE Access*, 2024, 12: 126122–126134. [doi: [10.1109/ACCESS.2024.3404224](https://doi.org/10.1109/ACCESS.2024.3404224)]
- 25 Ahmed S, Hasan MK. Twin-SegNet: Dynamically coupled complementary segmentation networks for generalized medical image segmentation. *Computer Vision and Image Understanding*, 2024, 240: 103910. [doi: [10.1016/j.cviu.2023.103910](https://doi.org/10.1016/j.cviu.2023.103910)]
- 26 Reddy DA, Roy S, Kumar S, *et al.* A scheme for effective skin disease detection using optimized region growing segmentation and autoencoder based classification. *Procedia Computer Science*, 2023, 218: 274–282. [doi: [10.1016/j.procs.2023.01.009](https://doi.org/10.1016/j.procs.2023.01.009)]
- 27 Jumawi WAH, El-Zaar A. Gumbel (EVI)-based minimum cross-entropy thresholding for the segmentation of images with skewed histograms. *Applied System Innovation*, 2023, 6(5): 87. [doi: [10.3390/asi6050087](https://doi.org/10.3390/asi6050087)]
- 28 Azad R, Asadi-Aghbolaghi M, Fathy M, *et al.* Attention DeepLabv3+: Multi-level context attention mechanism for skin lesion segmentation. In: Bartoli A, Fusiello A, eds. *Computer Vision—ECCV 2020 Workshops*. Cham: Springer, 2020. 251–266.
- 29 Tuyet VTH, Binh NT. Melanoma classification via hybrid saliency and conditional random field with bottleneck to optimize DeepLab. *International Journal of Online and Biomedical Engineering*, 2023, 19(10): 140–155.
- 30 Randar S, Shah V, Kulkarni H, *et al.* YOLOv8-based frameworks for liver and tumor segmentation task on LiTS. *SN Computer Science*, 2024, 5(6): 741. [doi: [10.1007/s42979-024-03097-5](https://doi.org/10.1007/s42979-024-03097-5)]
- 31 Kang M, Ting CM, Ting FF, *et al.* ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation. *Image and Vision Computing*, 2024, 147: 105057. [doi: [10.1016/j.imavis.2024.105057](https://doi.org/10.1016/j.imavis.2024.105057)]
- 32 Wei WY, Li JY, Wu XQ, *et al.* High-through cell micronucleus image detection method combining multi-attention mechanism and YOLOv5. *Biomedical Signal Processing and Control*, 2024, 87: 105496. [doi: [10.1016/j.bspc.2023.105496](https://doi.org/10.1016/j.bspc.2023.105496)]

- 33 Bagheri F, Tarokh MJ, Ziaratban M. Semantic segmentation of lesions from dermoscopic images using YOLO-DeepLab networks. *International Journal of Engineering*, 2021, 34(2): 458–469.
- 34 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.
- 35 Rew J, Kim H, Hwang E. Hybrid segmentation scheme for skin features extraction using dermoscopy images. *Computers, Materials and Continua*, 2021, 69(1): 801–817. [doi: [10.32604/cmc.2021.017892](https://doi.org/10.32604/cmc.2021.017892)]
- 36 Szegedy C, Ioffe S, Vanhoucke V, *et al.* Inception-v4, inception-ResNet and the impact of residual connections on learning. *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. San Francisco: AAAI Press, 2017. 4278–4284.
- 37 Ke GL, Meng Q, Finley T, *et al.* LightGBM: A highly efficient gradient boosting decision tree. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 3149–3157.
- 38 Joseph SIT, Sasikala J, Sujitha Juliet D. Retraction note: A novel vessel detection and classification algorithm using a deep learning neural network model with morphological processing (M-DLNN). *Soft Computing*, 2019, 23(8): 2693–2700. [doi: [10.1007/s00500-024-10198-z](https://doi.org/10.1007/s00500-024-10198-z)]
- 39 Bagheri F, Tarokh MJ, Ziaratban M. Skin lesion segmentation from dermoscopic images by using Mask R-CNN, Retina-DeepLab, and graph-based methods. *Biomedical Signal Processing and Control*, 2021, 67: 102533. [doi: [10.1016/j.bspc.2021.102533](https://doi.org/10.1016/j.bspc.2021.102533)]
- 40 Lin TY, Goyal P, Girshick R, *et al.* Focal loss for dense object detection. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 2999–3007.
- 41 Czajkowska J, Badura P, Korzekwa S, *et al.* Automated segmentation of epidermis in high-frequency ultrasound of pathological skin using a cascade of DeepLab v3+ networks and fuzzy connectedness. *Computerized Medical Imaging and Graphics*, 2022, 95: 102023. [doi: [10.1016/j.compmedimag.2021.102023](https://doi.org/10.1016/j.compmedimag.2021.102023)]
- 42 Zafar M, Amin J, Sharif M, *et al.* DeepLabv3+-based segmentation and best features selection using slime mould algorithm for multi-class skin lesion classification. *Mathematics*, 2023, 11(2): 364. [doi: [10.3390/math11020364](https://doi.org/10.3390/math11020364)]
- 43 Sandler M, Howard A, Zhu ML, *et al.* MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 4510–4520.
- 44 Han K, Wang YH, Chen HT, *et al.* A survey on visual Transformer. arXiv:2012.12556, 2020.
- 45 Wang JC, Wei L, Wang LS, *et al.* Boundary-aware Transformers for skin lesion segmentation. *Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention*. Strasbourg: Springer, 2021. 206–216.
- 46 Xu RT, Wang CW, Zhang JG, *et al.* SkinFormer: Learning statistical texture representation with Transformer for skin lesion segmentation. *IEEE Journal of Biomedical and Health Informatics*, 2024, 28(10): 6008–6018. [doi: [10.1109/JBHI.2024.3417247](https://doi.org/10.1109/JBHI.2024.3417247)]
- 47 Xin C, Liu ZF, Ma YZ, *et al.* Transformer guided self-adaptive network for multi-scale skin lesion image segmentation. *Computers in Biology and Medicine*, 2024, 169: 107846. [doi: [10.1016/j.combiomed.2023.107846](https://doi.org/10.1016/j.combiomed.2023.107846)]
- 48 Yang G, Geng Z, Wang Q. MSPMformer: The fusion of Transformers and multi-scale perception modules skin lesion segmentation algorithm. *IEEE Access*, 2024, 12: 128602–128617. [doi: [10.1109/ACCESS.2024.3446808](https://doi.org/10.1109/ACCESS.2024.3446808)]
- 49 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)]
- 50 Li YQ, Tian TH, Hu J, *et al.* SUTrans-NET: A hybrid Transformer approach to skin lesion segmentation. *PeerJ Computer Science*, 2024, 10: e1935. [doi: [10.7717/peerj-cs.1935](https://doi.org/10.7717/peerj-cs.1935)]
- 51 Wu RK, Liang PC, Huang X, *et al.* MHorUNet: High-order spatial interaction UNet for skin lesion segmentation. *Biomedical Signal Processing and Control*, 2024, 88: 105517. [doi: [10.1016/j.bspc.2023.105517](https://doi.org/10.1016/j.bspc.2023.105517)]
- 52 Azad R, Heidari M, Shariatnia M, *et al.* TransDeepLab: Convolution-free Transformer-based DeepLab v3+ for medical image segmentation. *Proceedings of the 5th International Workshop on Predictive Intelligence in Medicine*. Singapore: Springer, 2022. 91–102.
- 53 Peng WQ, Li JJ, Lai DD, *et al.* DBTU-Net: A dual branch network fusing Transformer and U-Net for skin lesion segmentation. *IEEE Access*, 2025, 13: 101262–101273. [doi: [10.1109/ACCESS.2025.3578295](https://doi.org/10.1109/ACCESS.2025.3578295)]
- 54 Eskandari S, Lumpp J, Sanchez Giraldo L. Skin lesion segmentation improved by Transformer-based networks with inter-scale dependency modeling. *Proceedings of the 14th*

- International Workshop on Machine Learning in Medical Imaging. Vancouver: Springer, 2023. 351–360.
- 55 Cai LL, Hou KK, Zhou S. Intelligent skin lesion segmentation using deformable attention Transformer U-Net with bidirectional attention mechanism in skin cancer images. *Skin Research and Technology*, 2024, 30(8): e13783. [doi: [10.1111/srt.13783](https://doi.org/10.1111/srt.13783)]
- 56 Innani S, Dutande P, Baid U, *et al.* Generative adversarial networks based skin lesion segmentation. *Scientific Reports*, 2023, 13(1): 13467. [doi: [10.1038/s41598-023-39648-8](https://doi.org/10.1038/s41598-023-39648-8)]
- 57 Zou RY, Zhang JH, Wu YF. Skin lesion segmentation through generative adversarial networks with global and local semantic feature awareness. *Electronics*, 2024, 13(19): 3853. [doi: [10.3390/electronics13193853](https://doi.org/10.3390/electronics13193853)]
- 58 Wu JD, Wang ZY, Hong MX, *et al.* Medical SAM adapter: Adapting segment anything model for medical image segmentation. *Medical Image Analysis*, 2025, 102: 103547. [doi: [10.1016/j.media.2025.103547](https://doi.org/10.1016/j.media.2025.103547)]
- 59 Saharia C, Chan W, Chang HW, *et al.* Palette: Image-to-image diffusion models. *Proceedings of the 2022 ACM SIGGRAPH Conference Proceedings*. Vancouver: ACM, 2022. 15.
- 60 Ho J, Salimans T. Classifier-free diffusion guidance. *arXiv:2207.12598*, 2022.
- 61 Ma J, Li FF, Wang B. U-Mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv:2401.04722*, 2024.
- 62 Wang ZY, Zheng JQ, Zhang YC, *et al.* Mamba-UNet: UNet-like pure visual mamba for medical image segmentation. *arXiv:2402.05079*, 2024.
- 63 Dao T, Gu A. Transformers are SSMS: Generalized models and efficient algorithms through structured state space duality. *arXiv:2405.21060*, 2024.

(校对责编: 李慧鑫)