

doi:10.3969/j.issn.1003-3106.2025.11.015

引用格式:李明蓉. 基于注意力机制与小波变换的皮肤病变分割方法[J]. 无线电工程, 2025, 55(11): 2274-2282.
[A Skin Lesions Image Segmentation Method Based on Attention Mechanism and Wavelet Transform[J]. Radio Engineering, 2025, 55(11): 2274-2282.]

基于注意力机制与小波变换的皮肤病变分割方法

李明蓉

(贵州大学 大数据与信息工程学院, 贵州 贵阳 550025)

摘要: 皮肤病变发病率上升,使皮肤癌早期筛查变得尤为关键。皮肤病变图像分割任务在通道信息建模、结构适应性及特征融合方面存在不足,易导致病变边界识别不准确和关键上下文信息利用不充分。针对这些问题,提出一种基于注意力机制与小波变换的皮肤病变图像分割方法——AW-SkinNet。该方法通过双分支协同注意力提取空间与通道依赖特征,结合小波变换增强频域表征,并利用轻量注意力引导的子像素上采样提升细节恢复与上下文理解。在 ISIC-2017 和 ISIC-2018 皮肤病变分割数据集上的实验表明,所提方法相较于现有皮肤病变图像分割方法实现了更高的分割精度。

关键词: 皮肤病;图像分割;注意力机制;小波变换

中图分类号:R739.5;TP391.41

文献标志码:A

开放科学(资源服务)标识码(OSID):



文章编号:1003-3106(2025)11-2274-09

A Skin Lesions Image Segmentation Method Based on Attention Mechanism and Wavelet Transform

LI Mingrong

(College of Big Data and information Engineering, Guizhou University, Guiyang 550025, China)

Abstract: The rising incidence of skin lesions has made early screening for skin cancer increasingly critical. However, existing methods for skin lesion image segmentation often suffer from limitations in channel-wise information modeling, structural adaptability, and feature fusion, which can lead to inaccurate boundary delineation and insufficient utilization of crucial contextual information. To address these issues, a skin lesions image segmentation method based on attention mechanism and wavelet transform, termed AW-SkinNet, is proposed. The proposed approach employs a dual-branch collaborative attention module to extract spatial and channel-dependent features, integrates wavelet transform to enhance frequency-domain representations, and incorporates lightweight attention-guided sub-pixel upsampling to improve detail restoration and contextual understanding. Experimental results on the ISIC-2017 and ISIC-2018 skin lesion segmentation datasets demonstrate that the proposed method achieves higher segmentation accuracy compared with existing approaches for skin lesion image segmentation.

Keywords: skin lesion; image segmentation; attention mechanism; wavelet transform

0 引言

近年来,皮肤病变的发病率持续上升^[1],引发了医学界对皮肤癌的高度关注。皮肤癌具有显著的致死风险,但通过早期检测和及时干预,可显著提升患者的生存率。传统的皮肤病变识别主要依赖临床医生的视觉判断与经验积累,此类方式在效率与准确性方面均存在一定局限^[2]。为提高诊断的客观性与自动化水平,计算机辅助诊断(Computer-Aided Diagnosis, CAD)技术应运而生。在 CAD 系统中,对病灶区域的精确分割是至关重要的一步,其质量直

接决定了后续诊断流程的有效性。因此,皮肤病变图像的自动分割问题已成为近年来深度学习研究领域的重点方向,促使多种先进模型与算法不断被提出与优化。

早期针对皮肤病变区域的分割研究主要依托传统图像处理技术,涵盖人工精细标注、半自动化方法以及多种自动化策略,如基于灰度阈值^[3]、边缘检

收稿日期:2025-07-16

基金项目:国家自然科学基金(52168011,62166006)

Foundation Item: National Natural Science Foundation of China(52168011, 62166006)

测^[4]以及图论模型^[5]的实现方案。尽管上述方法在计算上较为高效,适用于特定场景,但普遍存在鲁棒性不足的问题,易受到伪影、遮挡物以及标注误差的影响。其中,阈值法和区域合并策略常面临过度分割现象,而图论方法虽然具备较强的建模能力,却因算法复杂性较高、适应性较差等问题,难以在多样化的临床图像中广泛推广。

随着深度学习技术的迅猛发展,其在挖掘大规模图像数据潜在特征以及自动学习高层语义表示方面展现出显著优势,为克服传统分割方法的性能瓶颈提供了全新思路。在众多模型中,卷积神经网络(Convolutional Neural Network, CNN)^[6]作为关键基础架构,在医学图像分割任务中取得了突破性进展。2017年,Shelhamer等^[7]提出的全卷积网络,首次实现了对图像的像素级预测,为后续研究提供了范式参考。然而,这类早期模型在建模图像全局上下文信息和融合多尺度特征方面能力有限,难以充分适应皮肤病变图像中常见的边界模糊、形态变化复杂等挑战,影响了分割精度的进一步提升。

U-Net^[8]架构及其系列变体因其编码器-解码器的结构优势,在医学图像分割领域迅速占据主导地位。为进一步增强模型性能,研究者提出了多种结构优化方案。例如,UNet++^[9]通过引入深层次嵌套和多级跳跃连接,改善了特征融合效果,但在处理低对比度图像时仍存在性能瓶颈。DoubleU-Net^[10]通过双重U-Net结构增强语义信息提取,但代价是模型参数量和训练时间的大幅提升。CE-Net^[11]结合空洞卷积与多尺度池化模块,在处理边界模糊图像上展现出一定优势,但在应对稀有病变时表现仍不理想。王国凯等^[12]提出的MSBF-Net网络,通过分裂池化、全尺度特征融合和边界融合的方法,有效减少了空间信息损失并增强了上下文信息传递。白雪飞等^[13]针对病变区域边界模糊问题提出的BEMA-UNet通过空间多尺度注意力和边缘增强机制提升了分割精度。

近年来,Transformer^[14]架构的引入为分割精度带来新突破,如TransUNet^[15]利用自注意力机制有效捕捉长距离依赖关系,实现更为细致的区域分割,但其庞大的参数规模导致训练过程中易出现梯度消失或爆炸等问题。陶惜婷等^[16]提出的并行双分支网络架构PDTransCNN,通过ResNet34提取局部上下文特征,并结合Vision Transformer来捕获长距离依赖关系,然而模型计算复杂度和参数量较高。此外,部分研究探索了融合CNN、Transformer和Mamba^[17]等结构的混合模型,并引入深度可分离卷

积^[18]、频域变换^[19]等技术以扩大感受野与提升特征表达能力。然而,这类模型往往伴随着计算资源消耗过大、泛化能力不强或特征连续性不足等局限性。

尽管以CNN为基础的方法在皮肤病变分割领域取得了显著进展,并在实际应用中占据主导地位,但其固有的结构特性仍存在一定局限。具体而言,在视觉感知任务中Transformer架构已被广泛应用以建模图像中长距离的依赖关系,然而现有注意力机制在处理视觉特征时存在通道维度信息建模不足和空间注意力缺乏结构适应性的问题。此外,在语义分割^[20]、图像恢复^[21]等任务中,现有通道注意力机制多集中于特征响应强度的自适应调整,缺乏对图像频域结构的显式建模,这可能导致对高频纹理区域和边界的抑制,从而影响分割边缘的精度。另一方面,常规的上采样操作虽然能够提升分辨率,但在信息传递中缺乏对跳跃连接特征通道重要性的建模,且在多尺度特征融合时容易引入冗余甚至噪声信息,导致解码器无法充分利用关键上下文,影响最终分割质量。

针对上述问题,本文提出了一种AW-SkinNet模型,以解决医学图像分割模型中通道维度信息建模不足和空间注意力缺乏结构适应性,缺乏对图像频域结构的显式建模以及常规的上采样操作缺乏对跳跃连接特征通道重要性的建模等问题,从而提高医学图像分割模型的分割精度。

1 研究方法

1.1 AW-SkinNet 模型框架

本文提出的AW-SkinNet模型建立于编码器与解码器架构之上,网络模型结构如图1所示。

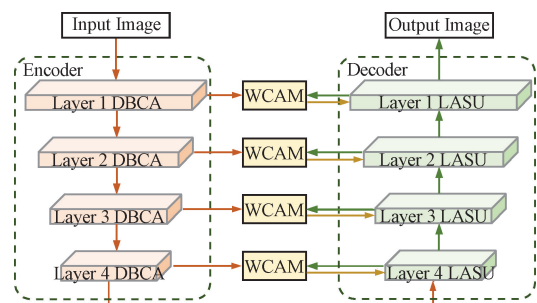


图1 AW-SkinNet 模型结构

Fig. 1 Structure of AW-SkinNet model

首先,针对皮肤病变形态多变且边界模糊的特性,模型通过双分支协同注意力机制(Dual-Branch Collaborative Attention, DBCA)对输入图像进行特征

提取。其中,对比协方差通道注意力(Contrastive-Covariance Channel Attention, CCA)分支通过建模通道间的协方差关系捕获不同通道特征之间的全局依赖关系与语义协同结构,解决通道维度信息建模不足的问题,以更充分地利用上下文信息并提升对病变整体语义的理解。空间感知引导注意力(Spatially-Aware Guided Attention, SGA)分支针对空间注意力缺乏结构适应性的问题,通过建模不同空间位置间的动态交互关系,引导模型聚焦与病变结构更相关的区域,增强对不规则形状病变的空间感知能力。

随后,针对常规方法缺乏对图像频域结构建模、易抑制高频边界细节的问题,模型引入基于小波变换的通道注意力模块(Wavelet-based Channel Attention Module, WCAM)。该模块利用小波变换的多尺度分解能力在频域处理特征,以更好地保留和增强对分割至关重要的高频边界信息,并结合通道注意力机制实现更强的上下文感知与细粒度结构建模能力。

最后,为提升解码器语义信息恢复的准确性和空间结构重建的质量,同时兼顾计算效率,设计了一种轻量而高效的轻量注意力引导的子像素上采样模块(Lightweight Attention-guided Sub-pixel Upsampling Module, LASU)来替代传统解码器中的上采样操作,以提升解码器对细节的恢复能力与整体上下文的理解能力,实现图像重建并获得最终的预测输出图像。

1.2 DBCA

为了更有效地捕捉图像特征中的通道依赖结构与空间上下文信息,本文提出了一种 DBCA。整个模块包括 2 个分支:CCA 分支与 SGA 分支,其结构如图 2 所示。

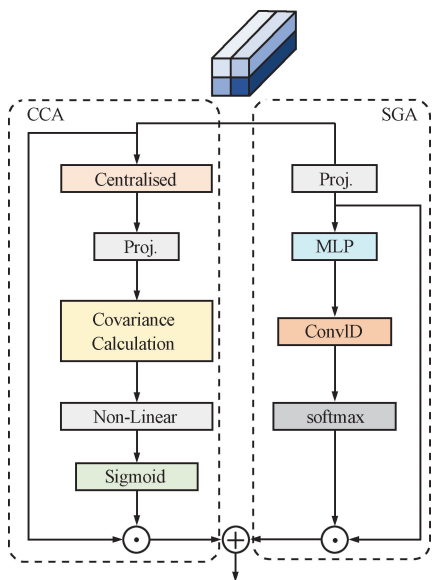


图 2 DBCA
Fig. 2 DBCA

传统通道注意力主要依赖平均池化或一维卷积对每个通道进行单独加权,忽略了通道之间的协同关系。为此,引入协方差矩阵作为统计工具,用以捕捉不同通道之间的变化相关性,从而建立更为鲁棒和表达力更强的通道注意力建模方式。

在空间维度上,为解决传统自注意力对长距离依赖建模能力强但对局部结构表达不足的问题,设计了基于空间感知的卷积注意力模块,利用 *Query* 向量通过函数映射生成一个用于局部感知的卷积核,并作用于 *Key* 和 *Value* 特征,从而实现内容自适应的局部卷积操作,使空间注意力具备更强的动态建模能力。

1.2.1 CCA

设输入特征为 $X_c \in \mathbb{R}^{B \times L \times C}$, 其中 B 为批量大小, L 为展平后的空间维度, C 为通道数。首先,对每个通道进行中心化处理以去除均值偏移:

$$\tilde{X}_c = X_c - \frac{1}{L} \sum_{i=1}^L X_c^{(i)}, \quad (1)$$

式中: \tilde{X}_c 表示中心化处理的特征,通过线性映射将其映射到一个较低维度 r ,以减少协方差计算的复杂度。

$$Z_c = \tilde{X}_c \cdot W_r \in \mathbb{R}^{B \times L \times r}, \quad (2)$$

式中: $W_r \in \mathbb{R}^{C \times r}$ 表示一个线性变换权重矩阵。对每个样本独立地计算协方差矩阵:

$$C_{cov}^{(b)} = \frac{1}{L-1} Z_c^{(b)T} \cdot Z_c^{(b)} \in \mathbb{R}^{r \times r}, \quad (3)$$

式中: $C_{cov}^{(b)}$ 表示第 b 个批次的协方差矩阵,其编码了通道嵌入之间的线性关系。通过非线性变换后将其投影回原通道空间:

$$A_c^{(b)} = W_a^T \cdot ReLU[C_{cov}^{(b)}] \cdot W_a, \quad (4)$$

式中: $W_a \in \mathbb{R}^{r \times C}$ 表示用于将协方差矩阵映射回通道维度的权重矩阵。最后,通过逐通道乘法将统计学习得到的协同注意力作用于原始通道表示:

$$X_c^{att, (b)} = X_c^{(b)} \odot \sigma[A_c^{(b)}], \quad (5)$$

式中: $\sigma(\cdot)$ 表示逐元素 Sigmoid 激活。这种方式不再对每个通道独立加权,而是依据其与其他通道的统计协同程度进行联合建模,从而提升表示的全局一致性和结构表达能力。

1.2.2 SCA

设 X_s 为输入特征(与 X_c 相同), $W_q, W_k, W_v \in \mathbb{R}^{C \times d}$ 为线性投影矩阵, d 为空间注意力的维度, $Q_s, K_s, V_s \in \mathbb{R}^{B \times L \times d}$ 为空间位置上的查询 (*Query*)、键 (*Key*) 和值 (*Value*):

$$Q_s = X_s W_q, \quad (6)$$

$$K_s = X_s W_k, \quad (7)$$

$$V_s = X_s W_v. \quad (8)$$

通过一个轻量级的感知机为每个 *token* 动态生成一组卷积核权重:

$$\mathbf{W}_{\text{dyn}} = \text{MLP}(\mathbf{Q}_s) \in \mathbb{R}^{B \times N \times k}, \quad (9)$$

式中: k 为局部卷积核大小。接下来,对于每一个 *token*,利用其对应的权重对整个 *Key* 和 *Value* 序列执行 1D 卷积操作:

$$\tilde{\mathbf{K}}_s^{(n)} = \text{Conv1D}[\mathbf{K}_s, \mathbf{W}_{\text{dyn}}^{(n)}], \quad (10)$$

$$\tilde{\mathbf{V}}_s^{(n)} = \text{Conv1D}[\mathbf{V}_s, \mathbf{W}_{\text{dyn}}^{(n)}], \quad (11)$$

得到局部结构增强后的 *Key* 与 *Value* 表示。采用标准点积注意力结构计算 *token* 与其上下文之间的匹配程度,单个注意力头的空间注意力计算如下:

$$\alpha_n = \text{softmax}\left[\frac{\mathbf{Q}_s^{(n)} \cdot \tilde{\mathbf{K}}_s^{(n)\text{T}}}{\sqrt{C}}\right]. \quad (12)$$

然后通过加权求和得到注意力输出特征 O_n :

$$O_n = \alpha_n \cdot \tilde{\mathbf{V}}_s^{(n)}. \quad (13)$$

1.2.3 双分支特征融合

双分支分别建模通道统计依赖和空间区域局部上下文信息,并通过并行协同融合,提升整体网络在多尺度视觉信息表达上的性能与泛化能力。通过相加的方式进行特征融合得到输出特征 Y :

$$Y = \mathbf{X}_c^{\text{att},(b)} + O_n. \quad (14)$$

1.3 WCAM

为了解决传统通道注意力模块在建模跨尺度上下文关系时易忽略多频率信息的问题,提出了一种基于小波变换^[22]的通道注意力模块,其结构如图3所示。

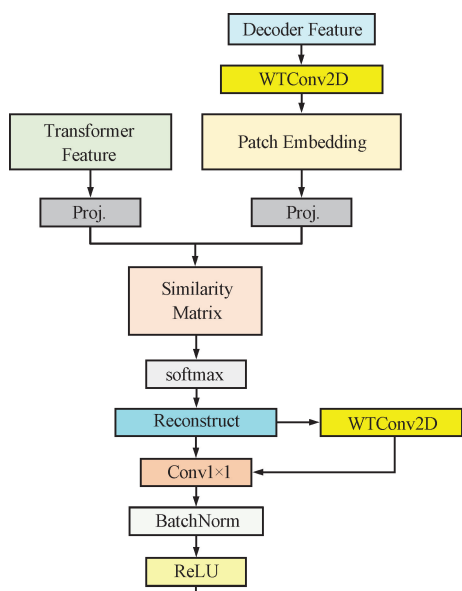


图3 WCAM
Fig.3 WCAM

该模块通过引入小波域的多尺度分解能力,增强特征图在频域下的表征能力,并结合通道注意力机制,实现更强的上下文感知与细粒度结构建模能力。

在该模块中,设解码器分支提取的特征图输入为 $\mathbf{D} \in \mathbb{R}^{B \times C \times H \times W}$,其中, H 、 W 是空间维度。

首先,本文对该特征图进行小波增强操作,使用自定义的 Wavelet Transform 卷积层 (WTConv2D) 对其进行频率域建模,得到增强后的特征 \mathbf{D}_{wt} :

$$\mathbf{D}_{\text{wt}} = \text{WTConv2D}(\mathbf{D}). \quad (15)$$

该 WTConv2D 模块内部首先执行若干层离散小波变换 (Discrete Wavelet Transform, DWT) 分解操作,将输入图像分解为低频子带 LL 与多个高频子带 $\{LH, HL, HH\}$ 。随后,对各个子带施加深度卷积操作,并通过逆小波变换 (Inverse Wavelet Transform, IWT) 将其重建回原始空间维度,增强了高频细节和结构轮廓的表达能力。

接着,将增强后的特征 \mathbf{D}_{wt} 进行块嵌入,通过一个步长为 *patch size* 的卷积将其转换为序列表示:

$$\mathbf{D}_p = \text{PatchEmbed}(\mathbf{D}_{\text{wt}}) \in \mathbb{R}^{B \times N \times C}, \quad (16)$$

式中: $N = \left(\frac{H}{P}\right) \times \left(\frac{W}{P}\right)$, P 为 *patch* 的边长。引入来自

Transformer 分支的中间特征 $\mathbf{T} \in \mathbb{R}^{B \times N \times C_T}$,并对二者分别进行线性投影以生成查询、键和值向量:

$$\mathbf{Q} = \mathbf{D}_p \mathbf{W}_q, \quad (17)$$

$$\mathbf{K} = \mathbf{T} \mathbf{W}_k, \quad (18)$$

$$\mathbf{V} = \mathbf{T} \mathbf{W}_v, \quad (19)$$

式中: \mathbf{W}_q 、 \mathbf{W}_k 、 \mathbf{W}_v 为学习得到的权重矩阵。接下来计算通道注意力的相关性矩阵。将 *Query* 与 *Key* 相乘,得到相似性矩阵:

$$\mathbf{A} = \mathbf{Q} \mathbf{K}^T \in \mathbb{R}^{B \times N \times C}. \quad (20)$$

为提升注意力分布的稳定性与泛化能力,本文在 softmax 前对其施加 Instance Normalization 操作:

$$\tilde{\mathbf{A}} = \text{softmax}[\text{IN}(\mathbf{A})]. \quad (21)$$

随后利用该注意力矩阵与 *Value* 相乘,生成通道增强后的响应特征:

$$\mathbf{F}_{\text{attn}} = \tilde{\mathbf{A}} \mathbf{V} \in \mathbb{R}^{B \times N \times C}. \quad (22)$$

再通过一个线性投影 \mathbf{W}_r 将其变换回原始通道维度,并通过重建恢复空间尺寸,得到初步融合特征:

$$\mathbf{F}_{\text{rec}} = \text{Reconstruct}(\mathbf{F}_{\text{attn}}, \mathbf{W}_r) \in \mathbb{R}^{B \times C \times H \times W}. \quad (23)$$

为了进一步提升该特征的频域表达能力,将 \mathbf{F}_{rec} 再次进行小波重建增强:

$$\mathbf{F}_{\text{wt}} = \text{WTConv2D}(\mathbf{F}_{\text{rec}}). \quad (24)$$

最后,将 \mathbf{F}_{rec} 与 \mathbf{F}_{wt} 沿通道维度拼接,并通过 1×1 卷积进行特征融合:

$$F_{fused} = GeLU\{BN[W_f(F_{rec}; F_{wt})]\}, \quad (25)$$

式中: $W_f \in \mathbb{R}^{2C \times C}$ 表示融合卷积权重。输出的 F_{fused} 作为解码器与 Transformer 间融合后的特征传递至上采样模块, 继续完成分割预测。

1.4 LASU

本文提出了一种轻量而高效的 LASU, 用于替代传统解码器中的上采样单元, 以提升语义信息恢复的准确性和空间结构重建的质量, 其结构如图 4 所示。该模块集成了 PixelShuffle^[23] 子像素重排方式、通道注意力引导机制与深度可分离卷积结构, 实现了解码器在性能与效率之间的有效权衡。

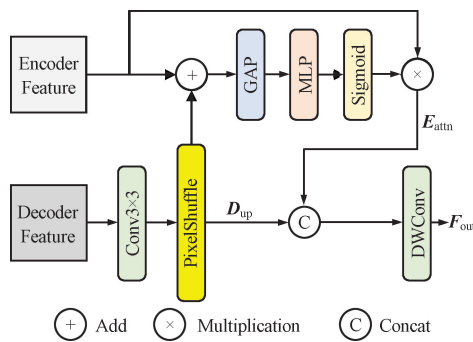


图 4 LASU
Fig. 4 LASU

具体而言, 在上采样阶段, 采用 PixelShuffle 算法进行子像素级别的空间重建, 相较于传统转置卷积, PixelShuffle 可避免棋盘伪影的出现, 并具备更平滑的上采样过程。

LASU 模块的输入包括 2 个分支: 解码器分支输入特征图记作 $D \in \mathbb{R}^{C \times H \times W}$, 编码器分支提供的经过跳跃连接的特征记作 $E \in \mathbb{R}^{C \times H \times W}$ 。首先, 对解码器特征进行上采样, 提升其空间分辨率至与编码器特征相同, 再使用子像素重排函数将通道维映射至更高分辨率的空间维度:

$$D_{up} = \pi[Conv_{3 \times 3}(D)] \in \mathbb{R}^{C \times rH \times rW}, \quad (26)$$

式中: $Conv_{3 \times 3}(\cdot)$ 表示通道扩展卷积操作, $\pi(\cdot)$ 表示 PixelShuffle 操作, r 表示上采样倍数, 本文设定为 2。

在融合来自编码器的跳跃连接特征时, 引入了通道注意力机制, 通过全局平均池化 (Global Average Pooling, GAP) 与多层感知机 (Multilayer Perceptron, MLP) 生成通道权重向量 $\alpha \in \mathbb{R}^C$, 对该特征进行动态加权:

$$\alpha = \sigma\{MLP[GAP(E + D_{up})]\}, \quad (27)$$

$$E_{attn} = \alpha \cdot E, \quad (28)$$

式中: $\sigma(\cdot)$ 表示 Sigmoid 激活函数, 用于将注意力向量归一化至 (0, 1), 保证每个通道的重要性权重

处于合理范围。该加权特征 E_{attn} 与上采样后的解码器特征 D_{up} 进行通道级拼接, 形成融合特征。为进一步提高参数效率, 利用深度可分离卷积替代传统卷积层, 得到最终的特征:

$$F_{out} = DWConv[Concat(E_{attn}, D_{up})]. \quad (29)$$

2 损失函数

为提升二分类医学图像分割任务中对目标区域的识别能力, 并有效缓解目标与背景之间的类别不平衡问题, 采用了加权二元交叉熵损失 (Binary Cross Entropy Loss, BCE)^[24] 与加权 Dice 损失^[25] 的加权组合, 构成最终的复合损失函数 Loss, 其表达式如下:

$$Loss = \lambda_{Dice} \cdot L_{Dice} + \lambda_{BCE} \cdot L_{BCE}, \quad (30)$$

式中: λ_{Dice} 与 λ_{BCE} 分别表示骰子系数 (Dice) 与 BCE 损失的权重系数, L_{Dice} 表示加权 Dice 损失, L_{BCE} 表示加权 BCE。

该损失函数结合了 Dice 损失对区域重叠的敏感性与 BCE 在像素层面上的稳定优化效果, 尤其适用于目标稀疏或类间不平衡严重的医学图像分割场景。

3 实验结果与分析

3.1 实验准备

3.1.1 数据集

ISIC-2017 与 ISIC-2018^[26] 皮肤病变分割数据集均由国际皮肤图像协作组织 (International Skin Imaging Collaboration, ISIC) 构建, 并作为国际医学图像计算与计算机辅助干预会议 (International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI) 旗下皮肤病变分析挑战赛的核心基准数据集, 旨在推动 CAD 技术在皮肤病变自动识别与分割中的应用与发展。这 2 个数据集设计用于皮肤镜图像中黑色素瘤等皮肤病变的检测与分割任务, 为相关研究提供了标准化的图像资源与标注依据。需要注意的是, 尽管该挑战赛也提供了独立的测试集用于最终评估, 但其真实标注结果并未公开。因此, 为了确保实验结果的可靠性、公平性和可重复性, 本文均采用五折交叉验证方法对训练集进行划分与评估。具体做法是将训练集划分为 5 个大小相近的子集, 在每一轮实验中使用其中 4 个子集进行训练, 剩余一个子集用于验证, 并循环进行 5 轮实验。最终的评估结果为 5 次实验结果的平均值, 能够有效减小由于样本划分带来的偶然性误差, 提升模型性能评价的稳健性。

3.1.2 实验环境与参数配置

本文实验在 NVIDIA 4060 Laptop GPU 上进行训练,基于 Python3.8 及 PyTorch 深度学习框架实现。将所有图像的大小调整为 224 pixel×224 pixel,然后输入网络。优化器为 AdamW^[27],根据初步实验结果,模型在 200 个训练轮次内即达到收敛。学习率为 0.001,权重衰减为 0.0005。

3.1.3 评价指标

为了有效评价网络的分割性能,选择 Dice 和 95% Hausdorff^[28] 距离 (HD95) 作为验证和测试的评估指标。

$$Dice = \frac{2TP}{2TP+FP+FN} \times 100\%, \quad (31)$$

式中: TP 、 FP 、 FN 分别表示真阳性、假阳性和假阴性。为了使结果更加稳定和令人信服,进行了 5 次交叉验证,并获得了平均值和标准差。

3.2 对比实验

为了验证 AW-SkinNet 模型在 ISIC-2017 和 ISIC-2018 数据集上的性能,将其与当前多种先进的分割方法进行了全面对比。这些方法涵盖了不同结构和机制的设计,包括传统的 U-Net^[6] 及其多个变体,例如引入嵌套和密集跳跃连接的 UNet++^[7]、采用注意力机制的 Attention U-Net^[29],以及融合多分辨率特征的 MultiResUNet^[30]。此外,还比较了近年来将 Transformer 引入医学图像分割的代表性模型,如 MEDT^[31]、Swin-UNet^[32] 和 UDTransNet^[33],以及基于状态空间模型 Mamba 的医学图像分割架构 VM-UNet V2^[34] 和 UltraLight VM-UNet^[35]。

表 1 展示了在 ISIC-2017 数据集上与基于 U-Net 及其变体的方法相比,AW-SkinNet 模型在平均 Dice 得分上较最优的 MultiResUNet 提升了 3.12%,在平均 HD95 上减少了 2.01%。与结合 Transformer 的方法比,AW-SkinNet 模型在平均 Dice 得分上较最优的 UDTransNet 提升了 2.84%,在平均 HD95 上降低了 1.46%。与基于 Mamba 的方法比,AW-SkinNet 模型在平均 Dice 得分上较最优的 UltraLight VM-UNet 提升了 0.97%,在平均 HD95 上降低了 1.05%。

表 2 展示了在 ISIC-2018 数据集上与基于 U-Net 及其变体的方法相比,AW-SkinNet 模型在平均 Dice 得分上较最优的 MultiResUNet 提升了 2.23%,同时在平均 HD95 上减少了 2.07%。与结合 Transformer 的方法相比,AW-SkinNet 模型在平均 Dice 得分上较最优的 UDTransNet 提升了 1.52%,在平均 HD95 上降低了 0.22%。与基于 Mamba 的方法相

比,AW-SkinNet 模型在平均 Dice 得分上较最优的 UltraLight VM-UNet 提升了 1.78%,在平均 HD95 上降低了 0.93%。

综合来看,这些数据充分证明了 AW-SkinNet 模型在整体分割性能上的显著优势。

表 1 ISIC-2017 数据集上的对比实验

Tab. 1 Comparison experiments on ISIC-2017 dataset

单位:%

方法	Avg Dice ↑	Avg HD95 ↓
U-Net	79.08	15.94
UNet++	82.76	14.68
Attention U-Net	87.83	14.05
MultiResUNet	88.74	12.17
MEDT	84.63	14.59
Swin-UNet	88.46	12.93
UDTransNet	89.02	11.62
VM-UNet V2	90.31	11.47
UltraLight VM-UNet	90.89	11.21
AW-SkinNet	91.86	10.16

表 2 ISIC-2018 数据集上的对比实验

Tab. 2 Comparison experiments on ISIC-2018 dataset

单位:%

方法	Avg Dice ↑	Avg HD95 ↓
U-Net	88.15	14.57
UNet++	88.97	12.31
Attention U-Net	89.05	13.34
MultiResUNet	89.20	12.70
MEDT	87.67	15.28
Swin-UNet	88.80	12.75
UDTransNet	89.91	10.85
VM-UNet V2	89.73	11.03
UltraLight VM-UNet	89.65	11.56
AW-SkinNet	91.43	10.63

3.3 消融实验

为验证 AW-SkinNet 中各个模块的有效性,ISIC-2018 数据集上进行了消融实验,探索 DBCA、WCAM 和 LASU 对模型性能的贡献,如表 3 所示。

表 3 AW-SkinNet 模型的消融实验

Tab. 3 Ablation experiments of AW-SkinNet

单位:%

方法	Avg Dice ↑	Avg HD95 ↓
Baseline	89.91	10.85
w/o WCAM&LASU	89.96	10.93
w/o LASU	90.77	10.76
AW-SkinNet	91.43	10.63

本实验中采用的 Baseline 为 UDTransNet,在去除所有模块的基础模型上,Dice 和 HD95 的平均值分别为 89.91% 和 10.85%。引入 DBCA 模块后,模

型通过建模通道间相关性与空间依赖关系,增强局部与全局特征的表达能力,平均 Dice 提升至 89.96%,验证了其对特征表征的优化作用。进一步加入 WCAM 模块后,模型的平均 Dice 达到 90.77%,平均 HD95 降至 10.76%,说明多层次注意力融合结构在提升语义区分能力方面有显著效果。

最终,当 3 个模块全部集成形成完整的 AW-SkinNet 时,模型取得了最优性能,平均 Dice 提高至 91.43%,平均 HD95 降至 10.63%,全面优于其他组合,表明 DBCA、WCAM 与 LASU 模块在提高模型精度、细化边界表达以及保持结构完整性方面均发挥了关键作用,三者协同构成了高效的分割结构。

3.4 可视化分析

该模型在 ISIC-2017 数据集和 ISIC-2018 数据集上进行了训练和验证。为了进一步说明分割结果和验证所提出的方法的有效性,如图 5 和图 6 所示,AW-SkinNet 模型在多个任务中的定量分割结果与真实标注高度一致。在边界清晰或边界模糊等不同情况下,现有方法常出现分割不足或过度分割的现象。而实验结果显示,AW-SkinNet 在这 2 类场景下均取得了更为鲁棒的分割效果。定量分析表明,该模型实现了在子区域的精确分割,验证了模型在皮肤病变分割任务中的有效性,提升了分割精度。

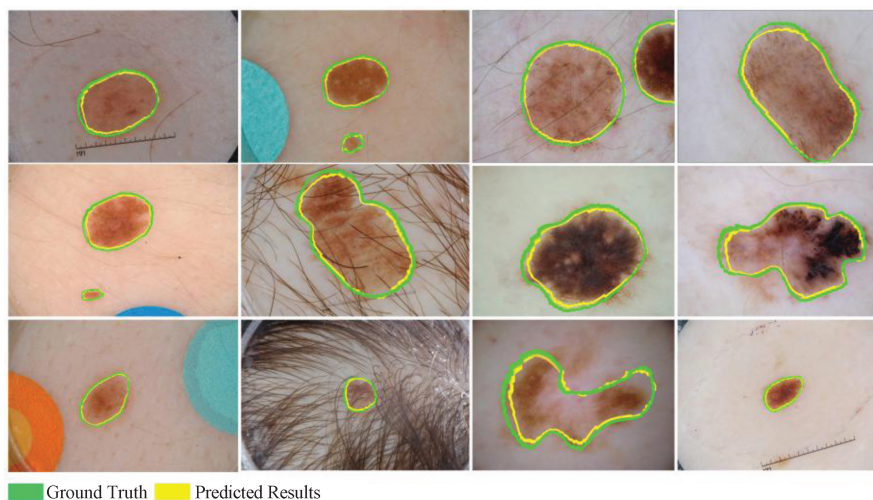


图 5 ISIC-2017 数据集上的可视化分割结果

Fig. 5 Visual segmentation results on ISIC-2017 dataset

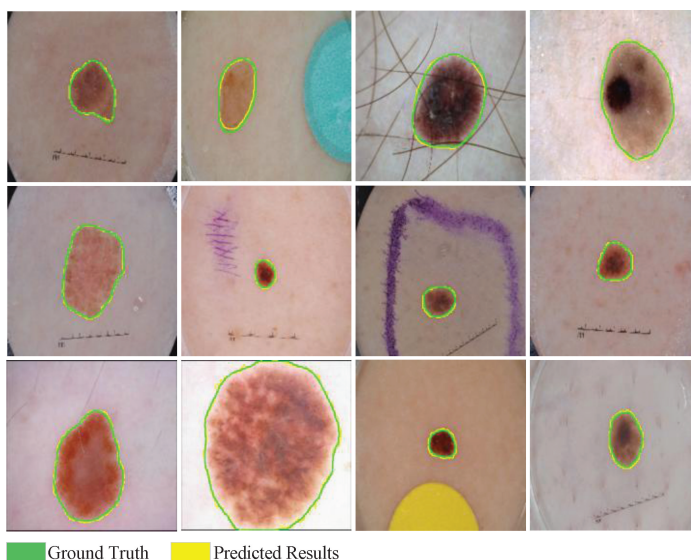


图 6 ISIC-2018 数据集上的可视化分割结果

Fig. 6 Visual segmentation results on ISIC-2018 dataset

4 结束语

本文提出了一种 AW-SkinNet 模型。通过 DB-CA 分别捕捉空间和通道间的依赖关系,融合小波变换以增强频域特征表达,并引入轻量级注意力引导的子像素上采样模块,提高了细节还原和上下文感知能力,从而实现精确的皮肤病变图像分割。在 ISIC-2017 和 ISIC-2018 皮肤病变分割数据集上的可视化分析与 ISIC-2018 数据集上进行的消融实验表明,本文方法优于其他先进的方法,实现了皮肤病灶的精确分割,并生成了高质量的分割图。



参考文献

- [1] YADAV R, BHAT A. A Systematic Literature Survey on Skin Disease Detection and Classification Using Machine Learning and Deep Learning [J]. *Multimedia Tools and Applications*, 2024, 83(32): 78093–78124.
- [2] 孔令军,王茜雯,包云超,等.一种基于深度学习的臂丛神经超声图像分割网络[J].*无线电工程*, 2021, 51(9): 841–847.
- [3] MA X P, SONG H D, JIA X, et al. An Improved V-Net Lung Nodule Segmentation Model Based on Pixel Threshold Separation and Attention Mechanism [J]. *Scientific Reports*, 2024, 14(1): 4743.
- [4] YE Y F, XU K, HUANG Y H, et al. DiffusionEdge: Diffusion Probabilistic Model for Crisp Edge Detection [EB/OL]. (2024-01-04) [2025-05-10]. <https://arxiv.org/abs/2401.02032>.
- [5] KUMAR J S, ARCHANA B, MURALIDHARAN K, et al. Graph Theory: Modelling and Analyzing Complex System [J]. *Metallurgical and Materials Engineering*, 2025, 31(3): 70–77.
- [6] 陈彦桥,张泽勇,陈金勇,等.基于 CNN 和 RFC 的极化 SAR 图像分类[J].*无线电工程*, 2021, 51(2): 87–91.
- [7] SHELHAMER E, LONG J, DARRELL T. Fully Convolutional Networks for Semantic Segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640–651.
- [8] 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*. Munich: Springer International Publishing, 2015: 234–241.
- [9] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHS N, et al. UNet++: A Nested U-Net Architecture for Medical Image Segmentation [C] // *International Workshop on Deep Learning in Medical Image Analysis*. Granada: Springer International Publishing, 2018: 3–11.
- [10] JHA D, RIEGLER M A, JOHANSEN D, et al. DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation [C] // *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*. Rochester: IEEE, 2020: 558–564.
- [11] GU Z W, CHENG J, FU H Z, et al. Ce-Net: Context Encoder Network for 2D Medical Image Segmentation [J]. *IEEE Transactions on Medical Imaging*, 2019, 38(10): 2281–2292.
- [12] 王国凯,张翔,王顺芳.多尺度和边界融合的皮肤病变区域分割网络[J].*计算机科学与探索*, 2024, 18(7): 1826–1837.
- [13] 白雪飞,靳智超,王文剑,等.融合边缘增强与多尺度注意力的皮肤病变分割[J].*计算机科学*, 2023, 50(4): 96–102.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need [EB/OL]. (2017-06-12) [2025-05-11]. <https://arxiv.org/abs/1706.03762>.
- [15] CHEN J N, MEI J R, LI X H, et al. Transunet: Rethinking the U-Net Architecture Design for Medical Image Segmentation Through the Lens of Transformers [J]. *Medical Image Analysis*, 2024, 97: 103280.
- [16] 陶惜婷,叶青.融合 CNN 和 Transformer 的并行双分支皮肤病灶图像分割[J].*计算机应用研究*, 2024, 41(8): 2554–2560.
- [17] MA X P, ZHANG X K, PUN M O. Rs³Mamba: Visual State Space Model for Remote Sensing Image Semantic Segmentation [J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 1–5.
- [18] CHOLLET F. Xception: Deep Learning with Depthwise Separable Convolutions [C] // *2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 1800–1807.
- [19] MAKIHARA Y, SAGAWA R, MUKAIGAWA Y, et al. Gait Recognition Using a View Transformation Model in the Frequency Domain [C] // *European Conference on Computer Vision*. Berlin: Springer, 2006: 151–163.
- [20] ZHU Q F, CAI Y Z, FANG Y, et al. Samba: Semantic Segmentation of Remotely Sensed Images with State Space Model [J]. *Heliyon*, 2024, 10(19): 38495.
- [21] GUO H, LI J M, DAI T, et al. Mambair: A Simple Baseline for Image Restoration with State-Space Model [C] // *European Conference on Computer Vision*. Cham: Springer, 2024: 222–241.
- [22] LIU Y J, FAN K F, ZHOU W J. FPWT: Filter Pruning via Wavelet Transform for CNNs [J]. *Neural Networks*, 2024, 179: 106577.

- [23] IBRAHEM H, SALEM A, KANG H S. Pixel Shuffling Is All You Need: Spatially Aware ConvMixer for Dense Prediction Tasks [J]. *Pattern Recognition*, 2025, 158: 111068.
- [24] HURTIK P, TOMASIELLO S, HULA J, et al. Binary Cross-entropy with Dynamical Clipping [J]. *Neural Computing and Applications*, 2022, 34(14): 12029–12041.
- [25] SUDRE C H, LI W Q, VERCAUTEREN T, et al. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations [C] // *international Workshop on Deep Learning in Medical Image Analysis*. [S.l.]: Springer, 2017: 240–248.
- [26] ZHU W H, TIAN J Y, CHEN M Z, et al. MSS-Unet: A Multi-spatial-shift MLP-based Unet for Skin Lesion Segmentation [J]. *Computers in Biology and Medicine*, 2023, 168: 107719.
- [27] ZHOU P, XIE X Y, LIN Z C, et al. Towards Understanding Convergence and Generalization of Adamw [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(9): 6486–6493.
- [28] TAHA A A, HANBURY A. An Efficient Algorithm for Calculating the Exact Hausdorff Distance [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(11): 2153–2163.
- [29] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: Learning Where to Look for the Pancreas [EB/OL]. (2018-04-11) [2025-05-11]. <https://arxiv.org/abs/1804.03999>.
- [30] IBTEHAZ N, RAHMAN M S. MultiResUNet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation [J]. *Neural Networks*, 2020, 121: 74–87.
- [31] QI Q F, LIN L Y, ZHANG R, et al. MEDT: Using Multimodal Encoding-Decoding Network as in Transformer for Multimodal Sentiment Analysis [J]. *IEEE Access*, 2022, 10: 28750–28759.
- [32] CAO H, WANG Y Y, CHEN J, et al. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation [C] // *European Conference on Computer Vision*. Tel Aviv: Springer, 2022: 205–218.
- [33] WANG H N, CAO P, YANG J Z, et al. Narrowing the Semantic Gaps in U-Net with Learnable Skip Connections: The Case of Medical Image Segmentation [J]. *Neural Networks*, 2024, 178: 106546.
- [34] ZHANG M Y, YU Y, JIN S, et al. VM-UNET-V2: Rethinking Vision Mamba UNet for Medical Image Segmentation [C] // *International Symposium on Bioinformatics Research and Applications*. Kunming: Springer, 2024: 335–346.
- [35] WU R K, LIU Y H, LIANG P C, et al. UltraLight VM-UNET: Parallel Vision Mamba Significantly Reduces Parameters for Skin Lesion Segmentation [EB/OL]. (2024-04-24) [2025-05-11]. <https://arxiv.org/pdf/2403.20035>.

作者简介

李明蓉 女, (2000—), 硕士研究生。主要研究方向: 图像分割、深度学习。