MPUNet: 基于多尺度特征增强和池化注意力的皮肤病分割网络

吴一帆[1)]，张帆[1,2)*]，王桦[1,3)]，张明丽[1,4)]，张彩明[1,2,5)]

[1)](山东工商学院计算机科学与技术学院 烟台 264003)
[2)](山东省未来智能金融工程实验室(山东工商学院) 烟台 264003)
[3)](鲁东大学计算机与人工智能学院 烟台 264025)
[4)](McGill Centre for Integrative Neuroscience at the Montreal Neurological Institute (McGill University) Montreal H3A 2B4 Canada)
[5)](山东大学软件学院 济南 250101)
(zhangfan@sdtbu.edu.cn)

摘 要：皮肤病的早期诊断高度依赖于皮肤镜图像中病变区域的准确分割。然而病变区域通常具有边界模糊、形状不规则等特点，同时图像还受到毛发、阴影等干扰因素的影响，难以有效地提取关键特征信息，这些因素对皮肤病的分割结果影响很大。为了解决上述问题，提出一种皮肤病分割网络 MPUNet。该网络融合多尺度特征增强模块（MS-Module）和池化注意力模块（PL-Module）。MS-Module 通过引入区分性的空间先验信息，在高和宽 2 个方向上进行通道自适应校准与融合，有效地增强多语义信息的表达能力和对病变边缘区域的特征提取能力；PL-Module 采用池化注意力机制强化全局上下文特征的提取，通过激活相关通道抑制毛发、血管等干扰因素对分割性能的影响，同时缓解训练过程中的梯度消失现象。在 ISIC2018、ISIC2017 和 PH2 公开皮肤病分割数据集上进行实验的结果表明，MPUNet 在 mIoU 指标平均提升 1.57 个百分点。

关键词：多尺度特征增强；池化注意力；皮肤病分割
中图法分类号：TP391.41 **DOI：**10.3724/SP.J.1089.2025-00222

MPUNet: Multi-Scale Feature Enhancement and Pooling Attention Network for Skin Disease Segmentation

Wu Yifan[1)], Zhang Fan[1,2)*], Wang Hua[1,3)], Zhang Mingli[1,4)], and Zhang Caiming[1,2,5)]

[1)](*College of Computer Science and Technology, Shandong Technology and Business University, Yantai* 264003)
[2)](*Shandong Future Intelligent Finance Engineering Laboratory (Shandong Technology and Business University), Yantai* 264003)
[3)](*School of Computer and Artificial Intelligence, Ludong University, Yantai* 264025)
[4)](*McGill Centre for Integrative Neuroscience at the Montreal Neurological Institute (McGill University), Montreal* H3A 2B4 *Canada*)
[5)](*School of Software, Shandong University, Jinan* 250101)

Abstract: Early diagnosis of skin diseases relies heavily on accurate segmentation of lesions in dermoscopic images. However, lesions often have blurred boundaries and irregular shapes. Furthermore, images are often affected by interference from factors such as hair and shadows, making it difficult to effectively extract key features. These factors significantly impact the segmentation results. To address these issues, a skin disease segmentation network (MPUNet) is proposed. This network integrates a multi-scale feature enhancement module (MS-Module) and a pooling attention module (PL-Module). The MS-Module incorporates discriminative spatial

收稿日期：2025-06-28；修回日期：2025-9-17。**基金项目：**国家自然科学基金（62272281）；泰山学者专项基金（tsqn202306274）；山东省高校青年创新科技项目（2023KJ212）；国家自然科学基金与浙江省信息化与工业化融合重点项目（U22A2033）。**吴一帆**（2002—），男，硕士研究生，CCF 学生会员，主要研究方向为图像处理；**张帆**（1986—），男，博士，教授，硕士生导师，CCF 会员，论文通信作者，主要研究方向为计算机图形学、计算机视觉、人工智能；**王桦**（1989—），女，博士，副教授，硕士生导师，主要研究方向为时间序列分析、图像处理；**张明丽**（1984—），女，博士，主要研究方向为计算机视觉、生物医学成像和自然图像处理；**张彩明**（1955—），男，博士，教授，博士生导师，CCF 高级会员，主要研究方向为 CAD&CG、图像处理、计算金融。

prior information and adaptively aligns and fuses channels in both height and width, effectively enhancing the representation of multi-semantic information and the ability to extract features from lesion edges. The PL-Module employs a pooling attention mechanism to enhance the extraction of global contextual features. By activating relevant channels, it suppresses the impact of interference from factors such as hair and blood vessels on segmentation performance, while also alleviating the vanishing gradient phenomenon during training. Experimental results on the ISIC2018, ISIC2017, and PH2 public skin disease segmentation datasets demonstrate that MPUNet achieves an average improvement of 1.57 percentage points in the mean intersection over union (mIoU) metric.

Key words: multi-scale feature enhancement; pooled attention; skin disease segmentation

近年来，皮肤病患者的数量不断上升，而皮肤科医生在辅助诊断过程中仍主要依赖手动描绘病变区域，该方法不仅耗费大量的人力资源，还极为依赖专业经验，且标注过程烦琐、耗时[1]。随着深度学习技术的快速发展，其在医学图像领域表现出强大的特征提取和表达能力，已被广泛地应用于皮肤病的诊断中。深度学习模型能够高效地从皮肤镜图像中提取关键特征，更快速、更准确地识别皮肤病变区域，显著地提升诊断的效率与准确率[2]。尽管当前皮肤病变分割在全监督分割方法中已经取得了显著进展[3-10]，但在实际应用中仍面临诸多挑战：（1）皮肤病图像区域含有毛发、血管等干扰因素，容易对模型的特征提取和判断造成误导；（2）患病皮肤区域与正常皮肤区域之间对比度较低、边界模糊，难以进行准确的分割；（3）不同图像的患病区域的大小、颜色等差异较大。这些因素阻碍了皮肤病分割的发展。近年来，研究者提出许多基于卷积神经网络的医学图像分割方法。DeepLabV3+方法[11]使用金字塔结构并结合残差连接与空洞卷积提取多尺度特征。U-Net[12]作为医学图像分割领域的经典基础网络，其衍生结构医学图像分割任务中均取得了较好的分割性能，如 UNet++[13]、ResUNet[14]，DenseNet[15]和 TransUNet[16]。当前，Transformer 架构凭借超强捕捉远程依赖关系的能力，逐渐在医学图像分割中展现出巨大潜力。杨舟等[17]使用轴向注意力在图像的横轴和竖轴方向上的像素层次聚合信息，同时使用跨通道注意力增强全局建模能力，最终通过融合注意力得到特征信息；周效栋等[18]和张帆等[19-20]动态地调整特征通道和空间位置的权重，使网络能够聚焦于对当前任务更为重要的图像区域，增强特征的表达能力；Wu 等[21]提出 FAT-Net，其充分地融合 CNN 和 Transformer，使用双分支结构分别获取局部和全局的特征信息，有效地提高了分割精度；Tao 等[22]采用双分支自适应特征融合策略，有效地增强特征表达能力，提升模型性能；Zhang 等[23]提出 CF-DAN，通过交叉融合双注意力机制结合局部细节和全局上下文信息，增强了特征的表达能力；He 等[24]提出 H2Former，该模型将 CNN、Transformer 和多尺度特征注意力融合到模块中，显著地增强了模型的特征表达能力。目前，基于 CNN 和 Transformer 混合架构的网络在皮肤病分割任务中取得了显著进展，但是在面对模糊的病变边缘，以及有血管和毛发的干扰因素等情况时，仍存在分割精度不足的问题。

为了解决上述问题，本文提出一种皮肤病分割网络：基于多尺度特征增强和池化注意力网络（multi-scale feature enhancement and pooling attention network，MPUNet），引入多尺度特征增强模块（multi-scale feature enhancement module，MS-Module）和池化注意力模块（pooled attention module，PL-Module）。在训练过程中，利用 PL-Module 对皮肤病变区域进行初步分割，通过池化机制和通道自选择策略有效地减少毛发等干扰因素的影响，从而提升整体分割效果，实现对病变区域的初步定位与干扰抑制。进一步，MS-Module 通过在高和宽 2 个方向上进行多尺度建模提升边缘特征表达能力，实现更精细的分割。MS-Module 和 PL-Module 协同工作，显著地提升了模型的鲁棒性与分割性能。

1 相关工作

1.1 全监督医学图像分割

近年来，随着深度学习技术的快速发展，基于深度学习的图像处理方法受到越来越多的关注。FCN[25]作为深度学习最早应用于图像分割任务的方法之一，首次实现了端到端的像素级预测，并通过跳跃连接融合多尺度特征以提升分割效果；受到 FCN 的启发，U-Net 中引入编码器-解码器结构，并通过跳跃连接将编码器与解码器相连，恢复下采

样过程中丢失的空间信息，凭借其优异的性能，所提出的 U 型结构被广泛地应用于后续的医学图像分割模型中。在 U-Net 的基础上，众多研究者对其跳跃连接、编码器与解码器结构等进行改进，如 UNet++[13]、UNet 3+[26]和 TransUNet[16]等。TransUNet 将 Transformer 中的多头自注意力引入编码器结构，有效地增强了特征表示能力；PraNet[27]强化深层特征的提取能力，显著地提升了分割性能；KiU-Net[28]借鉴 UNet 中的跳跃连接机制，设计了更加轻量化的网络结构；DoubleU-Net[29]通过串联 2 个 U-Net 结构，构建了双阶段分割框架；章东平等[30]提出一种边界线索深度融合的息肉图像分割网络，通过多尺度特征融合和边界增强模块显著地提升息肉区域的识别精度。随着 ViT（vision Transformer）[31]在计算机视觉领域取得成功，Transformer 架构逐渐被引入医学图像分割领域。TransUNet 首次将 Transformer 融入医学图像分割中，取得了优异的效果；TransFuse[32]、MCTrans[33]和 MEDT[34]等模型进一步探索 ViT 在医学分割任务中的应用，推动了该领域的快速发展。目前，医学分割方法已经在全监督领域取得了很好的分割效果。

1.2　Transformer 用于医学分割

早期，Transformer 只应用于自然语言处理任务，直到 ViT 将图像划分为不重叠的图像块，并将其序列化后作为输入，成功地将 Transformer 架构引入图像分类任务。Transformer 在计算机视觉领域的出色性能进一步促进了图像分割相关研究的发展。得益于其强大的全局建模能力，已有许多基于 Transformer 的图像分割方法。Valanarasu 等[35]提出一种融合门控轴向注意力机制和 Transformer 架构的医学图像分割方法，在多个医学分割任务中表现优异；TransUNet 将经典的 UNet 与 Tranformer 相结合，在多器官和心脏分割任务中取得了优异的分割效果；Cao 等[36]提出 Swin-UNet，用 SwinTransformer 模块替代传统 U-Net 中的编码器和解码器结构，显著地提高了分割精度；Xu 等[37]提出 DC-Net，使用 Transformer 模块有效地建模图像的上下文信息；Liu 等[38]提出 CSWin-UNet，将 CSWin 的跨形状窗口自注意力机制嵌入 U-Net 结构，并创新性地引入新的解码器设计，显著地提升了分割性能。研究人员还发现，基于 CNN 和 Transformer 的混合结构在皮肤病分割等任务中的表现优于单一的 Transformer 架构。因此本文提出的 MPUNet 同样采用 CNN 和 Transformer 的混合结构，并以 U-Net 作为基础骨干网络。

1.3　CNN 用于医学分割

近年来，CNN 在医学图像分割领域得到了广泛应用，如 FCN[25]、U-Net[12]、UNet++[13]和 CE-Net[39]。随着深度学习技术的不断发展，基于神经网络的分割方法也在持续改进。Yuan 等[5]设计一种损失函数，以缓解传统交叉熵损失在前景和背景像素识别中的模糊问题；Xie 等[40]提出相互引导的深度卷积网络，显著地提升了皮肤病变区域的分割精度。尽管基于 CNN 的方法在提取局部特征方面表现优异，但是其在建模图像的全局上下文信息方面存在一定的局限性，限制了对复杂结构的感知能力。为了弥补这一不足，研究者逐渐将注意力机制引入医学图像分割模型中。Dai 等[41]提出一种多尺度残差结构的编码器和解码器网络，用于提取皮肤病图像中的多层次特征；Hu 等[42]提出注意力协同网络，通过通道与空间注意力的协同融合提升模型的分割性能；Wu 等[43]提出自适应双重注意模块，可以整合来自不同方向的上下文信息，增强全局特征建模能力；Ruan 等[44]提出 MALUNet，其融合多种注意力机制，提升了对关键区域的感知能力；Sun 等[45]提出多尺度上下文注意力网络，通过融合不同尺度特征信息和上下文信息，有效地缓解了图像分辨率较低带来的分割性能下降问题。

2　本文网络

在临床实践中，医生判断黑色素瘤等皮肤病变时常常依赖于病变区域的边界形态，因此在皮肤病分割任务中对边缘区域特征的精确提取至关重要。针对这一需求，在 MPUNet 中提出 MS-Module 和 PL-Module，以提升模型对病变边缘的细致建模能力。如图 1 所示，MPUNet 在处理边界不规则或病变区域与背景色泽差异较小的图像时，依然能够保持良好的分割性能。其中，绿色的线为真实掩模值；红色的线为预测掩模值。

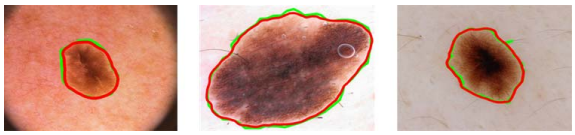


图 1　MPUNet 的分割结果

2.1　概　述

MPUNet 采用经典的 U 形编码器-解码器结构，编码器部分使用封装好的卷积块，而解码器则融合了传统卷积和注意力机制。MPUNet 的结构如图 2 所示，左侧为编码器，右侧为解码器。

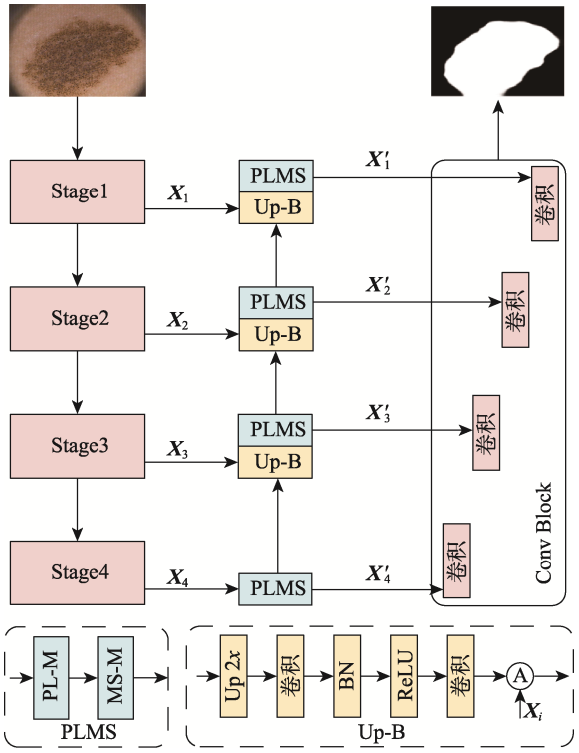


图 2 MPUNet 总体架构

编码器的 4 个层级 Stage1~Stage4 分别生成 4 个输出 $\boldsymbol{X}_1 \sim \boldsymbol{X}_4$，并将其传入解码器。$\boldsymbol{X}_3$、$\boldsymbol{X}_2$ 和 $\boldsymbol{X}_1$ 依次进入 Up-B，并与上一层 PLMS 模块的输出进行融合，最终依次输出结果 $\boldsymbol{X}_3'$、$\boldsymbol{X}_2'$、$\boldsymbol{X}_1'$；每个阶段的输出通过 ConvBlock 进行卷积处理，最终产生预测结果；$\boldsymbol{X}_4$ 进入解码器的 PLMS 模块和 ConvBlock 模块，并进行最终的预测。其中，PLMS 模块由 PL-Module 和 MS-Module 组成，分别负责提取全局上下文信息和局部细节特征。MS-Module 采用传统卷积提取图像的局部特征，而 PL-Module 结合注意力机制与传统卷积提取全局特征；两者协同作用于特定通道与空间位置特征的提取与融合，增强 PLMS 模块对医学图像中细粒度结构的建模能力。Up-B 通过残差连接与多尺度特征融合机制有效地减缓信息流失，并提升解码阶段的特征重建能力，优化了模型的整体分割性能。上采样过程中采用 3×3 卷积核，以有效地融合特征并保持较低的计算复杂度，加速模型最终结果的生成。A 表示将每个阶段的输出 $\boldsymbol{X}_3$、$\boldsymbol{X}_2$ 和 $\boldsymbol{X}_1$ 与 Up-B 的输出进一步结合，提升解码器的重建能力和特征丰富度。

为了深入地理解 MPUNet 的设计思路，有必要探讨 CNN 与 Transformer 在特征提取上的优势与不足。Transformer 擅长全局建模，但对局部细节建模能力相对较弱；而 CNN 在局部细节提取方面表现出色，但缺乏对全局上下文的建模能力。这一局限在皮肤病变图像中尤为明显，尤其是在处理血管和毛发等细节时，常导致分割结果不精确。为了解决这些问题，MS-Module 通过将输入特征划分为大小一致的子特征块，以提升局部建模能力，并为解码器提供更丰富的深层特征；PL-Module 的全局平均池化和全局最大池化操作通过提取上下文信息增强全局感知能力，进一步提升分割准确性。

2.2 MS-Module

在医学图像分割领域中，图像特征的提取至关重要，有效地强化提取的局部信息和全局信息对于提高分割准确性和可靠性具有重要影响。通过增强提取到的特征信息，能够更好地捕捉图像中的关键细节，提升分割模型的性能。局部信息在医学图像处理中的重要性体现在识别病变区域的边缘信息及提供细致的接触细节，然而单一方向上的局部特征提取存在一定的局限性，无法充分地捕捉复杂的空间信息。为了解决这一问题，本文提出 MS-Module 以增强模型提取局部特征的能力，其结构如图 3 所示。该模块在高度和宽度 2 个维度上分别进行局部特征提取，克服单一方向导致的信息提取不足的问题，它还能够解决因通道数减少而引发的特征信息丢失问题，以显著地提升特征的捕捉能力。首先对输入特征 $\boldsymbol{X}$ 进行全局最大池化操作为 $\boldsymbol{X}'$，增强提取全局特征的能力；然后对 $\boldsymbol{X}$ 在高度和宽度维度上分别进行全局平均池化，生成 2 个单向 1D 序列结果，分别表示高度和宽度方向的特征；再将这 2 个 1D 序列分解成 N 个大小相同的独立子特征，本文中设置 N=4。全局最大池化和分解公式分别为

$$\boldsymbol{X}_{\max} = \max\left(\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \boldsymbol{X}_{i3}, \boldsymbol{X}_{i4}\right) \tag{1}$$

$$\boldsymbol{X}_{\mathrm{H}}^{i} = \boldsymbol{X}_{\mathrm{H}}\left[:, (i-1)\times\frac{C}{N} : i\times\frac{C}{N}, :\right] \tag{2}$$

$$\boldsymbol{X}_{\mathrm{W}}^{i} = \boldsymbol{X}_{\mathrm{W}}\left[:, (i-1)\times\frac{C}{N} : i\times\frac{C}{N}, :\right] \tag{3}$$

对高度和宽度的分解操作是相互独立的过程，其中，2 个维度仅针对各自的方向进行均匀划分，分别被分解为若干个相等的子维度。这种独立分解的方式不仅简化特征提取的过程，还可增强模型在局部特征提取方面的能力。通过提取分解后相同大小的子特征，能够有效地获取多语义空间特征，提高模型对局部信息有效捕捉的能力。为了进一步增强模型对多尺度特征信息的捕捉能力，每个子特征都会经过不同内核大小（分别为 3，5，7 和 9）的多尺度特征信息提取操作，这种设计可以帮助模型有

效地提取出每个子特征内部的多样化语义空间结构，解决感受野不足的问题。此外，通过对高度和宽度维度的有效融合，模型能够学习到 2 个维度之间的关系，隐式地建模它们之间的依赖性。多尺度特征信息提取的公式分别为

$$\bar{\boldsymbol{X}}_{\mathrm{H}}^{i}=\mathrm{DWConv1D}_{n_i}^{\frac{C}{N}\to\frac{C}{N}}\left(\boldsymbol{X}_{\mathrm{H}}^{i}\right) \tag{4}$$

$$\bar{\boldsymbol{X}}_{\mathrm{W}}^{i}=\mathrm{DWConv1D}_{n_i}^{\frac{C}{N}\to\frac{C}{N}}\left(\boldsymbol{X}_{\mathrm{W}}^{i}\right) \tag{5}$$

对于高度和宽度维度处理后的结果，通过拼接 Concat 操作将 N 个子特征合并。为了进一步提升模型训练的稳定性和分割效果，对拼接后的特征进行批归一化（batch normalization，BN）处理，确保每个高度和宽度维度的特征合并后的表现一致，该操作有助于加速模型收敛并提高泛化能力。操作公式为

$$\boldsymbol{A}_{\mathrm{H}}=\mathrm{BN}_{\mathrm{H}}^{N}\left(\mathrm{Concat}\left(\bar{\boldsymbol{X}}_{\mathrm{H}}^{1},\bar{\boldsymbol{X}}_{\mathrm{H}}^{2},\cdots,\bar{\boldsymbol{X}}_{\mathrm{H}}^{N}\right)\right) \tag{6}$$

$$\boldsymbol{A}_{\mathrm{W}}=\mathrm{BN}_{\mathrm{W}}^{N}\left(\mathrm{Concat}\left(\bar{\boldsymbol{X}}_{\mathrm{W}}^{1},\bar{\boldsymbol{X}}_{\mathrm{W}}^{2},\cdots,\bar{\boldsymbol{X}}_{\mathrm{W}}^{N}\right)\right) \tag{7}$$

为了确保后续操作的顺利进行，需要对拼接后的高度和宽度维度分别进行额外的维度补全；然后将高度和宽度维度的特征进行相加，并通过一个卷积核大小为 3、步长为 1 的卷积层进行处理；再对卷积操作后的结果进行全局平均池化操作得到输出结果 $\boldsymbol{Y}$。最后将 $\boldsymbol{X}'$ 和 $\boldsymbol{Y}$ 进行通道维度拼接，并使用一个二维卷积层对这 2 个特征进行融合操作。该操作的目的是避免多尺度信息融合时的模型训练过拟合问题。采用 Sigmoid 激活函数对结果进行自适应调整，优化权重并得到最终的局部特征提取结果。通过逐元素相乘的方式将局部注意力特征矩阵与原始特征矩阵相乘，得到最终的局部特征矩阵，公式为

$$\boldsymbol{X}_{\mathrm{s}}=\mathrm{DWconv}\left(\boldsymbol{A}_{\mathrm{H}}+\boldsymbol{A}_{\mathrm{W}}\right) \tag{8}$$

$$\boldsymbol{X}=\mathrm{Sigmoid}\left(\mathrm{Conv}\left(\mathrm{Concat}\left(\boldsymbol{X}_{\max},\boldsymbol{X}_{\mathrm{s}}\right)\right)\right) \tag{9}$$

该过程确保模型在提取局部特征时能够充分地利用多尺度信息的优势，同时保持训练的稳定性和准确性，进一步提高了模型的分割性能。

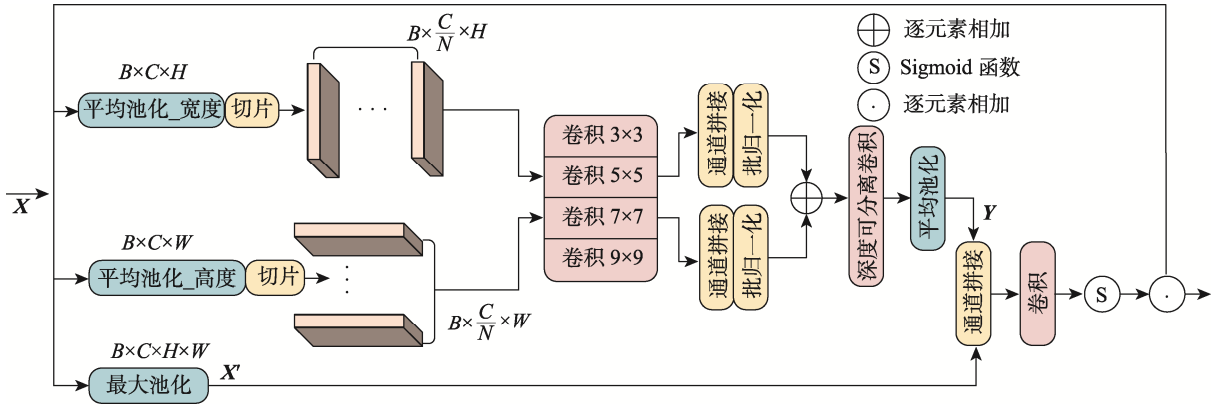


图 3　MS-Module 的结构

2.3　PL-Module

将 PL-Module 与 MS-Module 结合使用，显著地提高了分割准确性。PL-Module 的结构如图 4 所示，通过自适应选择通道，结合卷积和池化注意力机制提取全局上下文特征。该模块的主要功能是对皮肤病变区域进行初步分割，去除大部分背景噪声，能够从整体角度有效地处理毛发遮挡和血管突出等干扰因素，减少这些额外因素对分割结果的负面影响。现有方法通常依赖全局或局部注意力直接建模上下文关系，而 PL-Module 采用异步聚合策略，通过通道分离实现特征处理的明确分工，部分通道专注于局部特征细化；另一部分则通过异步不同形式的池化在皮肤病图像中高效地捕获病灶特征，再使用门控机制融合局部细节与全局上下文，有效地解

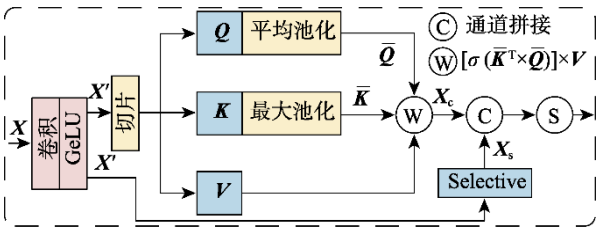


图 4　PL-Module 的结构

决传统注意力局部细节易丢失的问题，实现更均衡的上下文融合。$\boldsymbol{X}\in H^{i}\times W^{i}\times C^{i}$ 表示 PL-Module 的输入特征，其中，$\left(H^{i},W^{i},C^{i}\right)$ 分别表示高度、宽度和通道数。首先将输入特征 $\boldsymbol{X}$ 通过卷积操作并经过激活函数处理，有效地将特征信息进行融合输出为 $\boldsymbol{X}'$，$\boldsymbol{X}'$ 再经过 Selective 操作得到特征 $\boldsymbol{X}_{\mathrm{s}}$，

最后将 $\boldsymbol{X}_{\mathrm{s}}$ 与 $\boldsymbol{X}_{\mathrm{c}}$ 进行融合作为模块结果的输出。此过程中，Selective 操作通过自适应地选择通道，既有效地避免梯度爆炸问题，又能最大程度地提取全局特征信息。同时，对 $\boldsymbol{X}'$ 分解为 $\boldsymbol{Q}$ 、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 这 3 部分，并对每部分都执行各自的操作，这 3 部分分别表示注意力计算的查询矩阵、键矩阵和值矩阵。对 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 分别进行全局平均池化和全局最大池化操作，都采用步长为 2 的 3×3 卷积。其中，$\boldsymbol{Q}$ 的全局平均池化操作用于增强模块的鲁棒性，而 $\boldsymbol{K}$ 的全局最大池化操作则用于增强模块的特征提取能力。池化操作公式为

$$\bar{\boldsymbol{Q}} = P_{\mathrm{avg}}\left(\mathrm{GeLU}\left(\mathrm{Conv2D}\left(\boldsymbol{X}\right)\right)\right) \tag{10}$$

$$\bar{\boldsymbol{K}} = P_{\mathrm{max}}\left(\mathrm{GeLU}\left(\mathrm{Conv2D}\left(\boldsymbol{X}\right)\right)\right) \tag{11}$$

对 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 的池化操作有助于提取特征信息，显著地提高网络最终的分割准确性。

将 $\bar{\boldsymbol{Q}}$ 和 $\bar{\boldsymbol{K}}^{\mathrm{T}}$ 相乘，并通过 Softmax 层对结果进行归一化，避免出现梯度爆炸或消失现象；然后将归一化后的结果与 $\boldsymbol{V}$ 相乘，得到最终的全局注意力为

$$\boldsymbol{X}_{\mathrm{c}} = \boldsymbol{V} \times \left[\mathrm{Softmax}\left(\bar{\boldsymbol{K}}^{\mathrm{T}} \times \bar{\boldsymbol{Q}}\right)\right] \tag{12}$$

将最终的 $\boldsymbol{X}_{\mathrm{c}}$ 与通过自选择通道分支的结果 $\boldsymbol{X}_{\mathrm{s}}$ 进行通道拼接操作，不仅能够保留计算出的重要特征信息，还能最大限度地保留原始的全局特征。将全局注意力特征矩阵通过 Sigmoid 函数进行归一化处理，生成最终的全局注意力特征映射。最后采用逐元素相乘的方式将全局注意力矩阵与初始输入的特征矩阵相乘，得到最终的全局特征矩阵。

2.4 Up-B

近年来，因为多阶段、多尺度的特征融合已经成为医学图像分割的关键技术，所以 U 型医学分割网络常常使用跳跃连接模块融合编码器和解码器不同阶段的输出。本文提出的 Up-B 能够有效地融合来自不同空间尺度的特征信息。在模型的每个输入阶段，首先通过 Up 采样操作将特征维度翻倍；然后使用深度可分离卷积 DWConv 操作有效地融合多层次的输入特征；再通过 BN 和 ReLU 激活函数增强模型的非线性表达能力，进一步避免重要特征的丢失；最后进行一维卷积。Up-B 表示为

$$\mathrm{Up\text{-}B} = \mathrm{Conv}\left(\mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{Conv}\left(\mathrm{Up}\left(\boldsymbol{X}\right)\right)\right)\right)\right) \tag{13}$$

3 实验及结果分析

3.1 数据集

本文选择 3 个公开数据集进行训练和测试，以验证 MPUNet 在皮肤病变分割任务中的有效性，包括 ISIC2018 皮肤病变分割数据集（简称为 ISIC2018 数据集）[46]、ISIC2017 数据集[47]和 PH2 数据集[48]。各个数据集的训练集、验证集和测试集划分情况如表 1 所示。

表 1 实验数据集划分 幅

名称	皮肤病图像	数据集划分		
		训练集	验证集	测试集
ISIC2018	2594	1815	259	520
ISIC2017	2000	1400	200	400
PH2	200	140	20	40

3.2 实验细节

本文所有实验都是在一块 NVIDIA GeForce RTX3090 GPU 上进行，显存容量为 24 GB。为了增强模型的泛化能力，对所有的训练图像进行随机旋转操作用于数据增强。输入图像大小统一调整为 256×256 像素，其中，PH2 数据集的批处理大小为 4，ISIC2018 数据集的批处理大小为 24。在模型训练过程中，采用二进制交叉熵（binary cross entropy，BCE）损失和加权交并比（weighted intersection over union，WIoU）损失的线性组合作为总损失函数。AdamW 作为优化器，初始学习率为 0.000 1；训练过程为 400 个 epoch。加权二元交叉熵（weighted binary cross entropy，WBCE）损失 L_{WBCE} 预测掩模结果与真实掩模结果之间的误差，公式为

$$L_{\mathrm{WBCE}} = \sum_{j}^{i}\left(w \cdot \mathrm{BCE}\left(P, M\right)\right) \Big/ \sum_{j}^{i} w \tag{14}$$

其中，$w = 1+5\cdot\left|P_{\mathrm{avg}}\left(M\right) - M\right|$ 表示加权系数；P 预测模型未经过 Sigmoid 的结果；M 表示数据中的真实掩模值。WIoU 损失 L_{WIoU} 预测掩模结果与真实掩模结果之间的重叠程度，公式为

$$L_{\mathrm{WIoU}} = 1 - \frac{\sum_{j}^{i}\left(P' \cdot M \cdot w\right)+1}{\sum_{j}^{i}\left(P' + M \cdot w\right) - \sum_{j}^{i}\left(P' \cdot M \cdot w\right)+1} \tag{15}$$

其中，i 和 j 分别表示像素位置的行、列索引；P' 表示预测模型经过 Sigmoid 的结果。这 2 种损失函数的线性组合作为 MPUNet 的总损失函数，公式为

$$L_{\text{total}} = \text{mean}\left(L_{\text{WBCE}} + L_{\text{WIoU}}\right) \quad (16)$$

3.3 评价指标

本文采用在计算机视觉和医学图像分割领域中广泛应用的 6 项评价指标对 MPUNet 的分割性能进行全面评估，分别为 Dice 系数 Dice、平均 IoU（mean IoU，mIoU）、灵敏度（Sensitivity，S_{en}）、特异度（Specificity，S_{pec}）、准确性（Accuracy，A_{cc}）和精确率（Precision，P_{re}）。这 6 个指标[49-52]能够从多个维度客观地反映模型的预测表现。其中，Dice 衡量预测区域与真实区域之间的重叠程度，尤其适用于病灶区域较小或形状不规则的情况；mIoU 衡量预测与真实掩模之间的平均 IoU，评估模型整体分割准确性；S_{en} 评估模型检测真实病变区域的能力；S_{pec} 表示模型在识别正常区域方面的性能；A_{cc} 表示模型在所有像素中预测正确的比例；P_{re} 衡量模型预测为病变区域中真正为病变的比例。6 项指标的计算公式分别为

$$\text{Dice} = \frac{2\times \text{TP}}{2\times \text{TP} + \text{FP} + \text{FN}},$$

$$\text{mIoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}},$$

$$S_{\text{en}} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$S_{\text{pec}} = \frac{\text{TN}}{\text{TN} + \text{FP}},$$

$$A_{\text{cc}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}},$$

$$P_{\text{re}} = \frac{\text{TP}}{\text{TP} + \text{FP}}。$$

其中，TP（true positive）表示将真实的病变区域正确预测为病变区域；FP（false postive）表示将正常区域误识别为病变区域；TN（true negative）表示将真实的正常区域正确识别为正常；FN（false negative）表示将病变区域错误地识别为正常区域。

3.4 对比实验

在 ISIC2018、ISIC2017 和 PH2 数据集上，将 MPUNet 与 8 种主流的医学图像分割方法进行实验，结果分别如表 2~表 4 所示。对比方法包括 U-Net[12]、TransUNet[16]、MALUNet[44]、CSAP-UNet[53]、MSDANet[54]、CSWin-UNet[38]、EGE-UNet[55] 和 UCM-NetV2[56]。其中，MALUNet、EGE-UNet、CSAP-UNet 和 UCM-NetV2 是专门为皮肤病图像分割任务设计的网络，其余方法则是通用的医学分割模型。

表 2　在 PH2 数据集上的 9 种方法性能对比　%

方法	Dice	mIoU	S_{en}	S_{pec}	A_{cc}	P_{re}
U-Net[12]	80.14	69.65	83.01	95.96	87.74	86.05
TransUNet[16]	92.48	86.53	90.78	96.80	93.81	95.31
MALUNet[44]	90.19	82.13	85.90	96.75	92.25	94.93
EGE-UNet[55]	91.69	84.65	89.66	95.81	93.26	93.81
CSAP-UNet[53]	92.66	86.62	92.52	93.79	95.27	93.76
MSDANet[54]	92.58	89.23	93.58	95.10	94.48	92.98
CSWin-UNet[38]	93.09	87.47	92.86	95.97	94.62	94.26
UCM-NetV2[56]	91.23	83.88	89.29	95.43	92.89	93.26
MPUNet	**95.18**	**91.00**	**94.37**	**97.33**	**96.03**	**96.38**

注：粗体表示最优值。

表 3　在 ISIC2017 数据集上 9 种方法性能对比　%

方法	Dice	mIoU	S_{en}	S_{pec}	A_{cc}	P_{re}
U-Net[12]	75.11	64.25	63.12	94.23	91.12	82.43
TransUNet[16]	83.24	77.56	87.56	95.56	94.41	87.11
MALUNet[44]	86.81	78.42	85.17	95.62	93.27	91.35
EGE-UNet[55]	87.99	78.56	86.59	97.75	95.73	89.44
CSAP-UNet[53]	88.57	79.65	90.10	84.63	95.64	93.24
MSDANet[54]	87.23	85.30	89.89	95.52	96.68	92.65
CSWin-UNet[38]	90.12	83.65	88.56	95.68	95.74	93.26
UCM-NetV2[56]	90.38	82.45	87.49	98.55	96.42	93.47
MPUNet	**91.42**	**85.33**	**90.16**	**98.98**	**97.12**	**94.86**

注：粗体表示最优值。

表 4　在 ISIC2018 数据集上 9 种方法性能对比　%

方法	Dice	mIoU	S_{en}	S_{pec}	A_{cc}	P_{re}
U-Net[12]	69.16	58.13	67.07	97.39	89.79	87.22
TransUNet[16]	87.61	80.45	89.06	96.78	95.29	83.36
MALUNet[44]	87.90	78.42	88.63	96.62	94.98	97.18
EGE-UNet[55]	88.53	79.42	89.16	96.82	95.24	**97.91**
CSAP-UNet[53]	86.02	78.13	**92.70**	96.55	95.26	94.57
MSDANet[54]	87.38	**85.26**	90.91	95.57	94.65	83.46
CSWin-UNet[38]	87.00	79.59	91.22	96.80	95.42	85.92
UCM-NetV2[56]	87.97	78.52	88.21	96.89	95.14	87.73
MPUNet	**90.10**	83.47	90.05	**98.21**	**96.45**	92.64

注：粗体表示最优值。

从表 2 可以看出，MPUNet 取得了最优的分割性能，其中，在 Dice、mIoU、S_{en}、S_{pec}、A_{cc} 和 P_{re} 指标上，比对比方法中最好的指标分别提高了 2.2%、2.0%、0.8%、0.5%、0.8%和 1.1%。从表 3 可以看出，MPUNet 也取得了最优的分割性能；从表 4 可以看出，MPUNet 在 Dice、S_{pec}、A_{cc} 和 P_{re} 指标上都取得了最优结果，在 mIoU 和 S_{en} 指标上也接近当前的最佳方法。在 ISIC2018 数据集中，由于

数据集的图像分辨率差异较大，因此削弱了局部细节增强模块的效果，且官方划分的训练集中存在较多小目标样本，也影响了 MPUNet 的训练平衡性。在 5 类代表性图像上，7 种分割方法的分割效果如图 5 所示。第 1 行表示该数据集中较为普通的图像，其中，病变区域和正常皮肤区域存在明显的边界和色差，且病变区域是较为完整的形状；第 2 行表示小病变区域的图像；第 3 行表示病变形状复杂但边界清晰，无明显色差的图像；第 4 行表示病变区域与正常皮肤颜色相近，对比度较低的图像；第 5 行表示含有毛发和血管等干扰因素的图像。可以看出，对于第 1 行较为普通的图像，所有方法的分割效果都比较好；在 5 类图像中，MPUNet 均可以有效地提取出图像中的特征，清晰地分割出图像的病变区域。

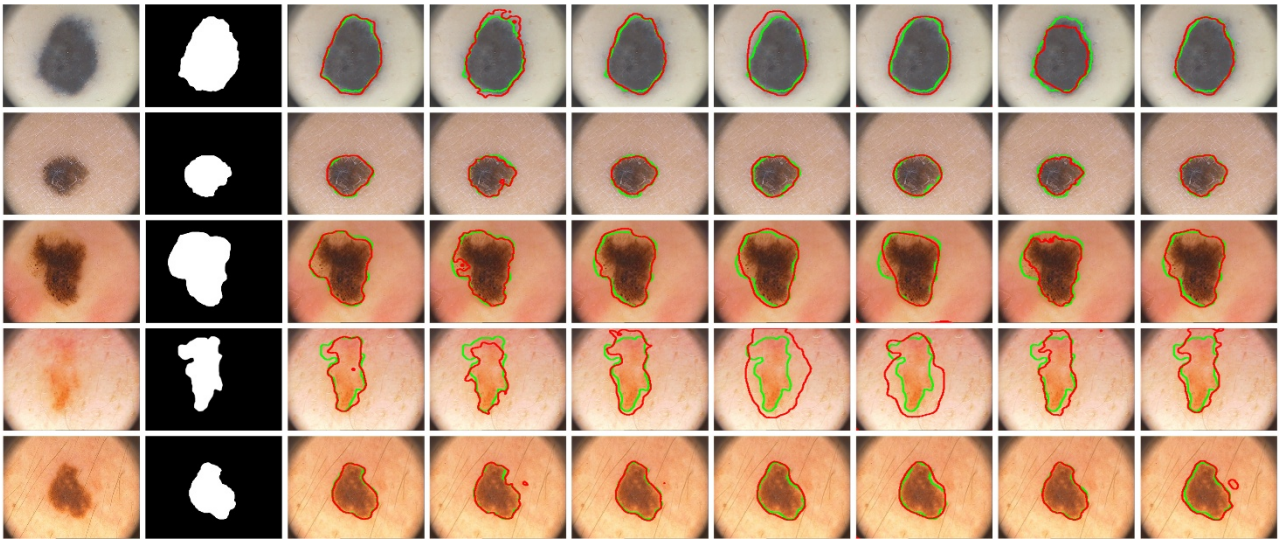


图 5 在 5 类代表性图像上 7 种方法的分割效果视觉对比

3.5 消融实验

通过消融实验，验证本文提出的 MS-Module 和 PL-Module 对 MPUNet 的性能贡献。选择不包含 MS-Module 和 PL-Module 的 G-CASCADE 模型[57]为基线模型，将 MS-Module 和 PL-Module 进行实验，得到单独使用这 2 个模块的模型性能，以及同时使用 2 个模块得到的分割性能。

在 ISIC2018 和 PH2 数据集上进行消融实验，设置批量大小为 4，并确保所有实验中的参数和实验环境保持一致，结果如表 5 所示。可以看出，加入 PL-Module 后，与基线网络相比，MPUNet 的 mIoU 和 Dice 分别提升 0.6%和 0.3%；同时加入 PL-Module 和 MS-Module 后，与基线网络相比，MPUNet 的 mIoU 和 Dice 指标分别提升 1.1%和 0.5%，证明该网络可以有效地提高分割性能。在 ISIC2018 数据集上，加入 PL-Module 后，L_{total} 对比如图 6 所示。可以看出，L_{total} 也有所下降，表明 MPUNet 优于基线模型的结果。为了评估模型效率与性能的平衡性，对不同模块的参数量和浮点运算量对模型的影响进行实验，结果如表 5 所示。在 ISIC2018 数据集上，MS-Module 中多尺度特征提取不同卷积核参数时的性能对比如表 6 所示。可以看出，当卷积核大小分别为(3,5,7,9)时性能最好。

表 5 在 2 个数据集上的消融实验结果对比

基线网络	MS-Module	PL-Module	PH2		ISIC2018		参数量	浮点运算量
			Dice/%	mIoU/%	Dice/%	mIoU/%		
√			94.63	90.07	89.74	82.92	26.63M	5.57G
√	√		94.73	90.23	89.80	83.11	30.49M	6.26G
√		√	94.97	90.63	89.97	83.14	27.02M	5.33G
√	√	√	**95.18**	**91.01**	**90.10**	**83.47**	30.88M	6.02G

注：粗体表示最优值。

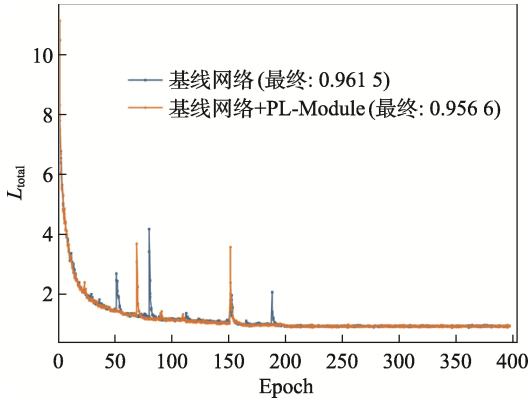


图 6　在 ISIC2018 数据集上 L_{total} 对比

表 6　ISIC2018 上 MS-Module 中不同卷积核参数性能对比　　%

参数	Dice	mIoU	S_{en}	S_{pec}
(3,3,3,3)	89.45	82.75	88.50	96.98
(5,5,5,5)	89.84	82.89	88.65	97.25
(7,7,7,7)	90.02	83.11	89.94	97.35
(3,3,5,5)	89.85	83.03	89.84	97.65
(3,5,7,9)	**90.10**	**83.47**	**90.05**	**98.21**

注：粗体表示最优值。

4　结　语

为了有效地分割皮肤病图像中的病变区域，本文提出的MPUNet，包括MS-Module和PL-Module，通过融合全局与局部特征，可有效地提取病变区域的纹理细节。MPUNet将特征图在高度和宽度方向上划分为若干个相同维度的子特征块，并进行融合处理，增强多尺度特征提取能力，同时提升了对病变边缘信息的捕捉能力。PL-Module负责全局上下文特征的建模，与注重局部细节的MS-Module相互协作，能够更精准地提取图像中的空间结构与通道语义特征，显著地提升分割性能。实验结果表明，MPUNet在皮肤病变分割任务中表现优异，尤其是在处理形状不规则及含有毛发等噪声干扰的图像时，依然能够保持良好的分割效果。未来，将聚焦于编码器结构的优化，探索更高效的特征提取策略，并在保障模型性能的前提下降低网络的空间和时间复杂度。

参考文献（References）：

[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[J]. Computer Methods and Programs in Biomedicine, 2019, 178: 201-218

[2] Hasan M K, Ahamad M A, Yap C H, *et al*. A survey, review, and future trends of skin lesion segmentation and classification[J]. Computers in Biology and Medicine, 2023, 155: Article No.106624

[3] Peruch F, Bogo F, Bonazza M, *et al*. Simpler, faster, more accurate melanocytic lesion segmentation through meds[J]. IEEE Transactions on Biomedical Engineering, 2014, 61(2): 557-565

[4] Jafari M H, Karimi N, Nasr-Esfahani E, *et al*. Skin lesion segmentation in clinical images using deep learning[C] //Proceedings of the 23rd International Conference on Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 337-342

[5] Yuan Y D, Chao M, Lo Y C. Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance[J]. IEEE Transactions on Medical Imaging, 2017, 36(9): 1876-1886

[6] Li H, He X Z, Yu Z, *et al*. Skin lesion segmentation via dense connected deconvolutional network[C] //Proceedings of the 24th International Conference on Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 671-675

[7] Ren L L, Heidari A A, Cai Z N, *et al*. Gaussian kernel probability-driven slime mould algorithm with new movement mechanism for multi-level image segmentation[J]. Measurement, 2022, 192: Article No.110884

[8] Xia J F, Cai Z N, Heidari A A, *et al*. Enhanced moth-flame optimizer with quasi-reflection and refraction learning with application to image segmentation and medical diagnosis[J]. Current Bioinformatics, 2023, 18(2): 109-142

[9] Liu L, Zhao D, Yu F H, *et al*. Performance optimization of differential evolution with slime mould algorithm for multilevel breast cancer image segmentation[J]. Computers in Biology and Medicine, 2021, 138: Article No.104910

[10] Hussien A G, Heidari A A, Ye X J, *et al*. Boosting whale optimization with evolution strategy and Gaussian random walks: an image segmentation method[J]. Engineering with Computers, 2023, 39(3): 1935-1979

[11] 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

[12] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C] //Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Heidelberg: Springer, 2015: 234-241

[13] Zhou Z W, Siddiquee M M R, Tajbakhsh N, *et al*. UNet++: redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE Transactions on Medical Imaging, 2020, 39(6): 1856-1867

[14] Zhang Z X, Liu Q J, Wang Y H. Road extraction by deep residual U-Net[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 749-753

[15] Huang G, Liu Z, van der Maaten L, *et al*. Densely connected convolutional networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 2261-2269

[16] Chen J N, Lu Y Y, Yu Q H, *et al*. TransUNet: transformers make strong encoders for medical image segmentation[OL]. [2025-06-28]. https://arxiv.org/abs/2102.04306

[17] Yang Zhou, Zhang Fan, Zhang Mingli, *et al*. Image de-raining network based on hybrid attention and gate inverted residual module[J/OL]. Journal of Computer-Aided Design & Computer Graphics. [2025-06-28]. https://link.cnki.net/urlid/11.2925.TP.20241029.1418.039(in Chinese)
(杨舟, 张帆, 张明丽, 等. 基于混合注意力与门控倒残差模块的图像去雨网络[J/OL]. 计算机辅助设计与图形学学报. [2025-06-28]. https://link.cnki.net/urlid/11.2925.TP.20241029.1418.039)

[18] Zhou Xiaodong, Zhang Fan, Li Jinjiang, *et al*. Object detection network based on feature enhancement and polynomial interpolation[J/OL]. Journal of Computer-Aided Design & Computer Graphics. [2025-06-28]. https://link.cnki.net/urlid/11.2925.TP.20250206.1619.027(in Chinese)
(周效栋, 张帆, 李晋江, 等. 基于特征增强与多项式插值的目标检测网络[J/OL]. 计算机辅助设计与图形学学报. [2025-06-28]. https://link.cnki.net/urlid/11.2925.TP.20250206.1619.027)

[19] Zhang Fan, Wang Hua, Fan Hui, *et al*. Rational polynomial image magnification based on edge and distance constraints[J]. Scientia Sinica Informationis, 2021, 51(8): 1270-1286(in Chinese)
(张帆, 王桦, 范辉, 等. 基于边缘和距离约束的有理多项式图像放大[J]. 中国科学: 信息科学, 2021, 51(8): 1270-1286)

[20] Zhang Fan, Zhou Hao, Wang Hua, *et al*. Rational polynomial image magnification constrained by feature[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(7): 1047-1057(in Chinese)
(张帆, 周浩, 王桦, 等. 特征约束的有理多项式图像放大[J]. 计算机辅助设计与图形学学报, 2022, 34(7): 1047-1057)

[21] Wu H S, Chen S H, Chen G L, *et al*. FAT-Net: feature adaptive transformers for automated skin lesion segmentation[J]. Medical Image Analysis, 2022, 76: Article No.102327

[22] Tao H L, Li J J, Hua Z, *et al*. DUDB: deep unfolding-based dual-branch feature fusion network for pan-sharpening remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: Article No.5400417

[23] Zhang F, Chen G G, Wang H, *et al*. CF-DAN: facial-expression recognition based on cross-fusion dual-attention network[J]. Computational Visual Media, 2024, 10(3): 593-608

[24] He A L, Wang K, Li T, *et al*. H2Former: an efficient hierarchical hybrid transformer for medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2023, 42(9): 2763- 2775

[25] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 3431-3440

[26] Huang H M, Lin L F, Tong R F, *et al*. UNet 3+: a full-scale connected UNet for medical image segmentation[C] //Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Los Alamitos: IEEE Computer Society Press, 2020: 1055-1059

[27] Fan D P, Ji G P, Zhou T, *et al*. PraNet: parallel reverse attention network for polyp segmentation[C] //Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2020: 263-273

[28] Valanarasu J M J, Sindagi V A, Hacihaliloglu I, *et al*. KiU-Net: towards accurate segmentation of biomedical images using over-complete representations[C] //Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2020: 363-373

[29] Jha D, Riegler M A, Johansen D, *et al*. DoubleU-Net: a deep convolutional neural network for medical image segmentation[C] //Proceedings of the 33rd International Symposium on Computer-Based Medical Systems. Los Alamitos: IEEE Computer Society Press, 2020: 558-564

[30] Zhang Dongping, Li Zheng, Xie Yaguang, *et al*. Boundary cue deep fusion polyp image segmentation network[J]. Journal of Image and Graphics, 2025, 30(5): 1479-1496(in Chinese)
(章东平, 李铮, 谢亚光, 等. 边界线索深度融合息肉图像分割网络[J]. 中国图象图形学报, 2025, 30(5): 1479-1496)

[31] Dosovitskiy A, Beyer L, Kolesnikov A, *et al*. An image is worth 16x16 words: transformers for image recognition at scale[C] //Proceedings of the 9th International Conference on Learning Representations. Washington DC: ICLR, 2021: 1-22

[32] Zhang Y D, Liu H Y, Hu Q. TransFuse: fusing transformers and CNNs for medical image segmentation[C] //Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2021: 14-24

[33] Ji Y F, Zhang R M, Wang H J, *et al*. Multi-compound transformer for accurate biomedical image segmentation[C] //Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2021: 326-336

[34] 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

[35] Valanarasu J M J, Oza P, Hacihaliloglu I, *et al*. Medical transformer: gated axial-attention for medical image segmentation[C] //Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2021: 36-46

[36] Cao H, Wang Y Y, Chen J, *et al*. Swin-UNet: UNet-like pure transformer for medical image segmentation[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2022: 205-218

[37] Xu R T, Wang C W, Xu S B, *et al*. DC-Net: dual context network for 2D medical image segmentation[C] //Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2021: 503-513

[38] Liu X, Gao P, Yu T, *et al*. CSWin-UNet: transformer UNet with cross-shaped windows for medical image segmentation[J]. Information Fusion, 2025, 113: Article No.102634

[39] 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

[40] Xie Y T, Zhang J P, Xia Y, *et al*. A mutual bootstrapping model

for automated skin lesion segmentation and classification[J]. IEEE Transactions on Medical Imaging, 2020, 39(7): 2482-2493

[41] Dai D W, Dong C X, Xu S H, *et al*. Ms RED: a novel multi-scale residual encoding and decoding network for skin lesion segmentation[J]. Medical Image Analysis, 2022, 75: Article No.102293

[42] Hu K, Lu J, Lee D, *et al*. AS-Net: Attention Synergy Network for skin lesion segmentation[J]. Expert Systems with Applications, 2022, 201: Article No.117112

[43] Wu H S, Pan J Q, Li Z Y, *et al*. Automated skin lesion segmentation via an adaptive dual attention module[J]. IEEE Transactions on Medical Imaging, 2021, 40(1): 357-370

[44] Ruan J C, Xiang S C, Xie M Y, *et al*. MALUNet: a multi-attention and light-weight UNet for skin lesion segmentation[C] //Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. Los Alamitos: IEEE Computer Society Press, 2022: 1150-1156

[45] Sun Y H, Dai D W, Zhang Q N, *et al*. MSCA-Net: multi-scale contextual attention network for skin lesion segmentation[J]. Pattern Recognition, 2023, 139: Article No.109524

[46] 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)[OL]. [2025-06-28]. https://arxiv.org/abs/1902.03368

[47] Berseth M. ISIC 2017-skin lesion analysis towards melanoma detection[OL]. [2025-06-28]. https://arxiv.org/abs/1703.00523

[48] Mendonça T, Ferreira P M, Marques J S, *et al*. PH^2-A dermoscopic image database for research and benchmarking[C] //Proceedings of the 35th annual International Conference of the IEEE Engineering in Medicine and Biology Society. Los Alamitos: IEEE Computer Society Press, 2013: 5437-5440

[49] Khan T M, Arsalan M, Robles-Kelly A, *et al*. MKIS-Net: a light-weight multi-kernel network for medical image segmentation[C] //Proceedings of 2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA). Piscataway: IEEE, 2022: 1-8

[50] Naqvi S S, Langah Z A, Khan H A, *et al*. Glan: GAN assisted lightweight attention network for biomedical imaging based diagnostics[J]. Cognitive Computation, 2023, 15(3): 932-942

[51] Javed S, Khan T M, Qayyum A, *et al*. Advancing medical image segmentation with mini-net: a lightweight solution tailored for efficient segmentation of medical images[OL]. [2025-06-28]. https://arxiv.org/abs/2405.17520

[52] Naveed A, Naqvi S S, Khan T M, *et al*. PCA: progressive class-wise attention for skin lesions diagnosis[J]. Engineering Applications of Artificial Intelligence, 2024, 127: 107417

[53] Fan X D, Zhou J, Jiang X L, *et al*. CSAP-UNet: convolution and self-attention paralleling network for medical image segmentation with edge enhancement[J]. Computers in Biology and Medicine, 2024, 172: Article No.108265

[54] Zhang J Q, Luan Z, Ni L N, *et al*. MSDANet: a multi-scale dilation attention network for medical image segmentation[J]. Biomedical Signal Processing and Control, 2024, 90: Article No.105889

[55] Ruan J C, Xie M Y, Gao J S, *et al*. EGE-UNet: an efficient group enhanced UNet for skin lesion segmentation[C] //Proceedings of the 26th International Conference on Medical Image Computing and Computer Assisted Intervention. Heidelberg: Springer, 2023: 481-490

[56] Yuan C Y, Zhao D F, Agaian S S. UCM-NetV2: an efficient and accurate deep learning model for skin lesion segmentation[J]. Journal of Economy and Technology, 2025, 3: 251-263

[57] Rahman M M, Marculescu R. G-CASCADE: efficient cascaded graph convolutional decoding for 2D medical image segmentation[C] //Proceedings of the Winter Conference on Applications of Computer Vision. Los Alamitos: IEEE Computer Society Press, 2024: 7713-7722