

# 多尺度自适应上下文融合的皮肤病变图像分割网络<sup>①</sup>



练纯青<sup>1</sup>, 李维乾<sup>1</sup>, 陈金广<sup>1</sup>, 徐子竞<sup>2</sup>, 马丽丽<sup>1</sup>

<sup>1</sup>(西安工程大学 计算机科学学院, 西安 710600)

<sup>2</sup>(黑龙江大学 计算机与大数据学院, 哈尔滨 150080)

通信作者: 李维乾, E-mail: wqli@xpu.edu.cn

**摘要:** 针对皮肤图像中存在目标分割区域尺度变化大、病变区域误分割及边界模糊等问题, 本文提出了一种名为 MSANet 的皮肤病变分割方法. 该方法以 pyramid vision Transformer v2 (PVT v2) 网络模型作为骨架网络, 结合 Transformer 和卷积神经网络 (CNN) 的优势, 通过对多层融合解码策略进行改进, 提升了皮肤病变分割的准确性. 首先, 解码部分利用分割门控注意力模块 (SGA) 获取多尺度全局和局部空间特征, 增强模型对上下文信息的捕获能力. 随后采用多尺度上下文注意力模块 (MCA) 抽取并整合通道和位置信息, 提升网络对病灶区域的精准定位. 在 ISIC2017 和 ISIC2018 数据集上的实验结果表明, MSANet 的主要指标 *Dice* 系数分别达到了 90.12% 和 90.91%, *mIoU* 分别为 85.82% 和 84.27%, 其分割性能优于现有方法.

**关键词:** 皮肤病变图像分割; 多层融合; 多尺度特征; 注意力机制; Transformer

引用格式: 练纯青, 李维乾, 陈金广, 徐子竞, 马丽丽. 多尺度自适应上下文融合的皮肤病变图像分割网络. 计算机系统应用, 2025, 34(8): 188-196. <http://www.c-s-a.org.cn/1003-3254/9915.html>

## Multi-scale Adaptive Context Fusion Network for Skin Lesion Image Segmentation

LIAN Chun-Qing<sup>1</sup>, LI Wei-Qian<sup>1</sup>, CHEN Jin-Guang<sup>1</sup>, XU Zi-Jing<sup>2</sup>, MA Li-Li<sup>1</sup>

<sup>1</sup>(School of Computer Science, Xi'an Polytechnic University, Xi'an 710600, China)

<sup>2</sup>(School of Computer and Big Data, Heilongjiang University, Harbin 150080, China)

**Abstract:** To address challenges such as large-scale variations in target segmentation regions, mis-segmentation of lesion areas, and blurred boundaries in skin images, this study proposes a novel method for skin lesion segmentation, named MSANet. This approach utilizes the pyramid vision Transformer v2 (PVT v2) as the backbone network, integrating the strengths of both Transformer and convolutional neural networks (CNNs). By improving the multi-layer fusion decoding strategy, the proposed method significantly enhances the accuracy of skin lesion segmentation. The decoding process incorporates a split gated attention block (SGA) to capture multi-scale global and local spatial features, thus enhancing the model's ability to capture contextual information. The multi-scale contextual attention (MCA) module is employed to extract and integrate both channel and positional information, improving the network's precision in lesion localization. Experimental results on the ISIC2017 and ISIC2018 datasets demonstrate that MSANet achieves *Dice* scores of 90.12% and 90.91%, and *mIoU* scores of 85.82% and 84.27%, respectively, outperforming existing methods in segmentation performance.

**Key words:** skin lesion image segmentation; multi-layer fusion; multi-scale feature; attention mechanism; Transformer

① 基金项目: 陕西省自然科学基金基础研究计划 (2023-JC-YB-826)

收稿时间: 2024-12-30; 修改时间: 2025-01-20, 2025-02-11; 采用时间: 2025-02-18; csa 在线出版时间: 2025-06-27

CNKI 网络首发时间: 2025-06-30

## 1 引言

恶性黑色素瘤是最具侵袭性和致命性的皮肤病变之一,尽管其在皮肤癌中的发病率较低,但致死率相对较高<sup>[1]</sup>.在皮肤癌的治疗过程中,精准地从皮肤镜获得的图像中分割出病变区域显得尤为关键.手动标注方法因过程复杂、耗时长、误差较大等问题,在临床应用中难以实现大规模标注,因此迫切需要能够自动进行且可以精确分割图像的技术,以满足临床工作对图像处理的需求.随着深度学习技术的迅猛发展,基于深度卷积神经网络(CNN)的图像分割方法被提出,并成功应用于皮肤病变分割任务<sup>[2,3]</sup>.如U-Net<sup>[4]</sup>通过在其编码器和解码器之间引入跳跃连接,有效融合了深层语义信息与浅层细节信息,展示出卓越的分割性能;而基于U-Net变体的UNet++<sup>[5]</sup>、ResUNet++<sup>[6]</sup>及U2Net<sup>[7]</sup>等模型,凭借其复杂的编码器-解码器结构,在医学图像分割领域同样展现出了引人注目的性能.为克服CNN中感受野的局限性,使得模型能够捕获像素间的长距离依赖关系,Chen等人<sup>[8]</sup>通过引入扩张卷积以有效扩大卷积核的感受野,从而在不增加计算量的情况下获取更丰富的上下文信息.Xie等人<sup>[9]</sup>采用分组卷积,成功减少了计算开销并加快了网络的训练速度.Howard等人<sup>[10]</sup>引入了深度可分离卷积,大幅降低了模型参数量和计算量.同时,Woo等人<sup>[11]</sup>提出的CBAM模块通过融合空间注意力和通道注意力,有效地增强了特征图的表达能力,进一步提升了卷积神经网络的性能.尽管这些方法扩大了特征感受野、提升了模型训练速度、增强了特征图表达能力,但其在捕获远程上下文信息方面仍存在不足.

近年,众多基于Transformer的方法相继出现,在各类视觉任务中取得重大突破.Transformer最初是为自然语言处理(NLP)中的序列到序列预测任务而设计,它借助自注意力机制学习输入标记之间的相互关系,在捕获远距离依赖和从宏观层面理解语义差异方面表现出色.为了将这种能力迁移至视觉领域,Dosovitskiy等人<sup>[12]</sup>于2021年提出了视觉变换器(ViT),该方法将图像划分成固定大小的块,并将其转换为向量序列,再通过多头自注意力(multi-head self-attention, MHSA)与多层感知机(multilayer perceptron, MLP)技术,有效捕捉并学习图像中长距离依赖关系.同年,Liu等人<sup>[13]</sup>提出层级化视觉模型Swin Transformer,通过窗口注意力机制提升模型对局部区域信息的捕捉能力.相比之下,

Wang等人<sup>[14]</sup>在2022年提出的金字塔视觉Transformer(pyramid vision Transformer, PVT)展现出更显著的优势,其通过金字塔结构和空间缩减注意力机制,有效融合了全局与局部信息且能够降低计算复杂度,使PVT在目标检测和图像分割等任务中具有较大优势.尽管该模型在处理全局特征方面表现出色,但在像素间的局部上下文信息方面仍存在限制,这可能导致对模糊边界的预测不够精确.在全面捕获全局到局部细节方面仍有提升空间,且在解码策略方面仍面临挑战.

受上述文献的启发,以PVT v2网络模型作为骨架网络,通过改变编码器结构,构建用于恶性黑色素瘤分割的多尺度自适应上下文融合网络(MSANet).总的来说,本文的主要贡献如下.

(1) 多尺度自适应上下文融合网络(MSANet):提出了MSANet,这是一款专为医学图像分割任务,特别是皮肤病变图像分割而设计的分割模型.该模型通过多尺度处理和自适应上下文融合技术,优化了医学图像分割模型.

(2) 关键创新模块:设计了两个模块,即SGA和MCA.这些模块分别作用于多尺度上下文信息的有效融合和病灶区域的精准定位,显著提升了医学图像分割任务的精确度.

(3) 广泛的实验验证:在ISIC2017、ISIC2018两个公开的皮肤数据集上进行了大量实验.实验结果表明,MSANet在分割性能和泛化能力方面达到了当前先进水平,优于以往的方法.

## 2 相关工作

### 2.1 Vision Transformer

在计算机视觉领域,vision Transformer (ViT) 凭借其革命性的多头自注意力(MHSA)机制,已经成为具有里程碑意义的模型.相较于传统卷积层,MHSA层的引入突破了局部感受野的限制,显著提升了模型在处理复杂视觉场景时的表现力.Dosovitskiy等人<sup>[12]</sup>通过将图像分割成固定大小的块,并依次通过多个Transformer编码块处理,首次为图像识别领域引入了ViT,实现了前所未有的端到端图像识别效果,为多种视觉任务设定了新的性能基准.此外,基于ViT框架的后续研究成功集成了Transformer的全局处理能力与卷积神经网络的局部细节捕捉能力,由此产生了CvT<sup>[15]</sup>、CeiT<sup>[16]</sup>等混合模型架构.同时,通过金字塔结构的引入,

研究者在从卷积网络过渡到 Transformer 的过程中,实现了对多尺度信息的更优利用,进一步提升了模型在目标检测、语义分割等多种下游任务中的性能,其中 Swin Transformer<sup>[13]</sup>、PVT<sup>[14]</sup>和 SegFormer<sup>[17]</sup>等模型尤为突出。在医学图像分割领域,ViT 也显示出巨大潜力,如 TransFuse<sup>[18]</sup>模型凭借其强大的编码能力,在精确识别医学图像中病变区域方面取得了显著成果,为医疗诊断提供了强有力的支持。

然而,尽管 ViT 在多个领域展现了其卓越性能,但在恶性黑色素瘤图像分割中,数据集规模小的问题限制了模型的训练效果。为克服这一难题,我们采用了在大型数据集(例如 ImageNet)上预训练的 Transformer 编码器,以借助丰富的预训练特征来弥补本地数据集的不足。此外,设计了一种多尺度级联混合语义解码器,以优化模型在捕捉像素间局部上下文信息方面的不足,从而提升网络处理多尺度目标和边界细节的能力,显著增强了分割精确度。

## 2.2 皮肤病灶分割

恶性黑色素瘤图像分割作为医学图像分割的一个关键子集,由于深度学习特别是卷积神经网络(CNN)的应用而取得了显著进展。全卷积网络(FCN)和 U-Net<sup>[4]</sup>的开发与实施显著推动了这一领域的进步,为像素级准确分割皮肤病灶提供了坚实的基础。SkinNet<sup>[3]</sup>是基于 U-Net 的变体,其主要在编码器部分引入扩张卷积以增加感受野并提高网络性能。除此之外,还有诸如 MSAU-Net<sup>[19]</sup>等 U-Net 变体网络,通过在网络的瓶颈处插入注意力机制来模拟分层表示,以非线性方式聚合多级表示,从而选择代表性特征。更进一步,为了设计轻量级的分割模型,DSNet<sup>[20]</sup>和 Separable-Unet<sup>[21]</sup>在模型中使用了深度可分离卷积,以减少传统卷积的参数数量。ASCU-Net<sup>[22]</sup>则创新性地融合了三重注意力机制,使得网络能够更专注于与目标更为紧密相关的区域,进而提升皮肤病变分割的精确度。

2023年,Rahman 等人<sup>[23]</sup>提出的一种创新性的多尺度分层变换器网络(MERIT),通过引入级联注意力解码机制,并在不同窗口大小中计算自注意力,从而提升了分割的精确性。同时,注意力机制在医学图像分割领域的应用也取得了长足进展,通过与 CNN 和基于 Transformer 的架构相结合,使得医学图像的细节和边缘信息得以更精准地提取和呈现。例如,Du 等人<sup>[24]</sup>于 2023 年提出 AViT,通过引入轻量级适配器模块和提

示生成器,缓解了 ViT 在数据需求方面的问题,显著提高了皮肤病灶分割的性能。与这些方法不同,我们提出了一种多尺度自适应上下文融合网络,能够更有效地聚合多尺度上下文信息、精准定位目标区域,并精细处理边缘细节,使其在恶性黑色素瘤病灶分割任务中表现更为出色。

## 3 方法

本节首先介绍用于恶性黑色素瘤图像分割的整体网络 MSANet,然后详细介绍所提出的两个模块:分割门控注意力模块(SGA)、多尺度上下文注意力模块(MCA)。

### 3.1 整体结构

为增强网络对上下文特征信息的有效表示,针对皮肤病灶的特点,本文以 PVT v2 网络模型为骨架网络,通过添加分割门控注意力模块(SGA)和多尺度上下文注意力模块(MCA)两个模块,对解码器部分进行改进,提出了一种用于恶性黑色素瘤分割的多尺度自适应上下文融合网络(MSANet),如图 1 所示。

编码器部分采用在 ImageNet 数据集上预训练的 PVT v2,通过逐层提取皮肤病灶图像的空间和语义信息,随后编码器输出的特征图经过解码器部分进一步处理,经过 SGA 提取多尺度特征,然后通过高效的特征上采样(FUC)模块对特征图进行上采样,使得特征图在不同尺度上能更好地进行融合和处理。特征图也将送入 MCA 模块,进一步整合不同尺度的通道和位置信息,增强对病灶区域的精准定位。最后引入具有可学习权重的多尺度特征预测模块(MSP),以有效整合阶段性预测结果,形成最终预测图,整体流程见图 1。

### 3.2 分割门控注意力模块

在恶性黑色素瘤图像的自动分割任务中,准确捕捉病变区域的边界与细节是一个核心挑战。传统的分割模型往往依赖于全局信息来确定大致轮廓,或通过局部信息捕捉细节,缺乏有效的机制来融合这两者,导致分割结果出现精度不足或边界模糊的情况。而在医学图像中的恶性黑色素瘤分割任务中,病灶的形状和大小差异较大,这使得现有的注意力机制往往无法在不同尺度和形态的病灶区域间做出精细地区分,从而影响分割精度。为克服这一难题,本文设计了一种分割门控注意力模块(SGA),通过结合全局和局部特征信息,充分挖掘不同尺度下的上下文信息,确保模型实现对病变区域更准确、更全面地分割。

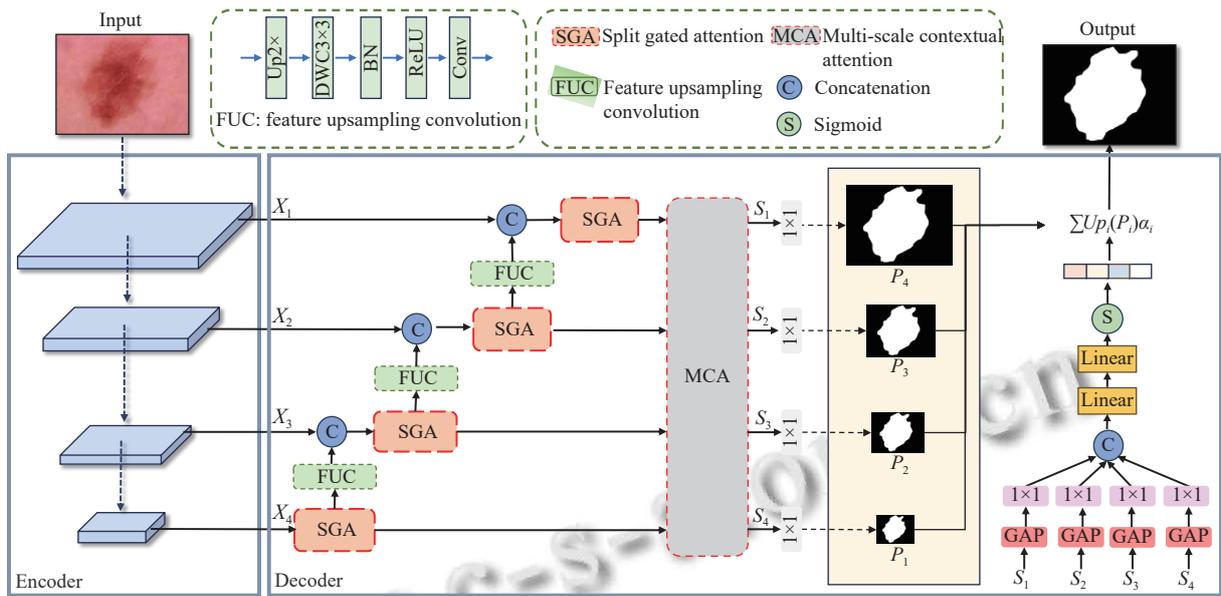


图1 模型整体结构图

受 Ruan 等人<sup>[25]</sup>提出的分组扩张卷积的启发,我们设计了 SGA 模块,通过结合全局和局部特征信息,确保分割结果的精细度和完整性,如图 2 所示.全局信息为模型提供了宏观层面的结构认知,这有助于识别皮肤病灶区域的整体轮廓及其与周围环境的联系,从而使模型能够更精确地确定病灶位置;同时,局部信息的捕捉则关注于微观层面的细节,这对于获取病变区域的边缘特征至关重要.与传统的全局或局部信息单独处理的模型不同,SGA 模块由两个子单元构成:分割卷积单元 split Conv unit (SCU) 和门控注意力单元 gated attention unit (GAU),实现了多尺度信息的高效融合,显著提升了分割精度.

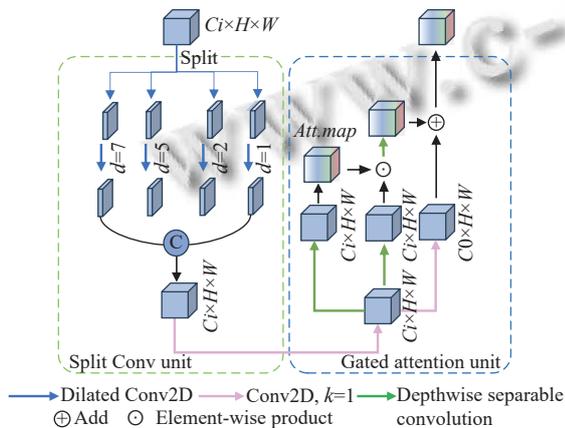


图2 分割门控注意力模块结构图

从编码器输出特征后,SCU 将特征图  $X$  作为输入沿通道维度分割为 4 部分,通过扩张率为 1 和 2 的卷

积得到局部特征信息,通过扩张率为 5 和 7 的深度可分离卷积获得全局特征.然后,在通道维度上进行级联操作以恢复特征图的大小并通过卷积操作得到输出特征图  $X'$ ,使得全局和局部信息进行交互.可以表示为:

$$x_1, x_2, x_3, x_4 = \text{Chunk}_4(X) \quad (1)$$

$$x'_1, x'_2, x'_3, x'_4 = W_1(x_1), W_2(x_2), W_5(x_3), W_7(x_4) \quad (2)$$

$$X' = W(\text{Concat}(x'_1, x'_2, x'_3, x'_4)) \quad (3)$$

其中,  $\text{Concat}$  表示连接操作,  $W$  表示普通卷积操作.

为弱化无关信息,增强 SCU 模块对多尺度信息的利用率,引入门控注意力单元 (GAU).对于 GAU 来说,通过深度可分离卷积生成与输入特征形状相同的注意力图,以抑制 SCU 传输的特征信息中的不重要区域,从而使模型更加关注重要信息.最后,应用残差连接操作来获得输出.

$$\text{Att} = \delta(DW(X')) \quad (4)$$

$$O_1 = DW(DW(X') \odot \text{Att}) + W(X') \quad (5)$$

其中,  $\delta$  是 Sigmoid 函数,  $DW$  表示深度可分离卷积,  $\odot$  表示逐元素乘法.

输出特征经过比例因子为 2 的上采样  $Up(i)$ 、 $3 \times 3$  的深度可分离卷积  $DW(\cdot)$ 、批归一化  $BN(\cdot)$ 、 $ReLU(\cdot)$  激活函数和  $1 \times 1$  卷积  $Conv(\cdot)$  完成对当前特征的上采样操作,以将当前维度与下一层跳跃连接相匹配完成特征融合.上述过程可以用式 (6) 表示:

$$FUC(x) = Conv(ReLU(BN(DW(U_p(O_1)))))) \quad (6)$$

### 3.3 多尺度上下文注意力模块

在医学图像分割中,尤其是对恶性黑色素瘤这种形态复杂、边界不清晰的病变区域,精确定位是一个重大挑战.传统的注意力机制虽然能够提高模型对显著特征的关注,但它们大多专注于单尺度特征的处理,无法充分利用多尺度的上下文信息.现有的注意力,如SE(squeeze-and-excitation)和CBAM(convolutional block attention module),已被广泛研究和应用以提高网络性能,但它们通常侧重于通道信息的提升,对空间位置和方向信息的处理较为薄弱.这在处理恶性黑色素瘤图像分割任务时尤为突出,因为病变区域的尺度差异和多样性常导致单一尺度的注意力机制无法有效捕获目标区域的全部特征,从而影响分割精度.

为解决这一问题,并获取更丰富的空间和语义信息,增强注意力机制对目标区域的定位精度,受Coordinate attention<sup>[26]</sup>的启发,本文引入多尺度上下文注意力模块(MCA),并将其用于医学图像分割领域,如图3所示.与传统方法不同,MCA模块不仅能捕获跨通道的信息,还可以同时捕获方向和位置的敏感信息,这有助于模型更加准确地定位和识别感兴趣的病灶区域,然后将所得的特征图互补地应用于输入特征图以增强模型对病灶区域的表示,从而提升分割的精度.

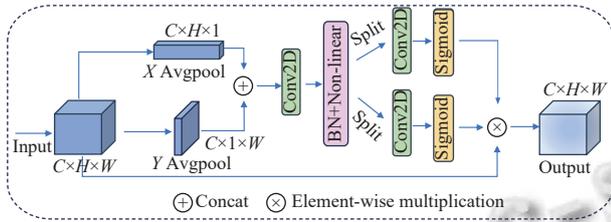


图3 多尺度上下文注意力模块结构图

MCA将通道注意力分解为两个并行的1D特征编码过程,分别沿水平和垂直方向进行1D平均池化操作,生成一对方向感知图.从而能够在一个方向上捕获长距离依赖性,同时在另一个方向上保留精确位置信息,位置信息分别生成映射 $F^h$ 和 $F^w$ ,如下所示:

$$F^h = \frac{1}{W} \sum_{0 \leq i \leq w} x_c(h, i) \quad (7)$$

$$F^w = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w) \quad (8)$$

接着将从两个方向编码得到的特征图进行拼接,对注意力图进行卷积和Sigmoid函数处理,分别得到

注意力图 $g^h$ 和 $g^w$ ,可以表示为:

$$g^h = \delta(Con\delta(Concat(F^h, F^w))) \quad (9)$$

$$g^w = \delta(Con\delta(Concat(F^h, F^w))) \quad (10)$$

其中,Concat(·)表示连接操作, $\delta$ 代表Sigmoid激活函数,Conv(·)表示卷积运算.

最后,将注意力权重应用于原始的输入特征图,得到输出特征图 $F'$ 以增强或减弱特定区域的特征表示,可以表示为:

$$F'(i, j) = F_c(i, j) \times g^h \times g^w \quad (11)$$

之后,在获得解码器不同阶段的输出特征 $S_i$ 后,经过引入的多尺度预测模块(MSP),在MSP中,多阶段自适应联合损失函数允许整体监督信息并行反馈到网络的不同特征阶段.具体来说,输出特征 $S_i$ 采用4个并行的 $1 \times 1$ 卷积和上采样操作获取不同阶段对应的输出二进制掩码 $P_i$ .同时将输出特征图 $S_1, S_2, S_3, S_4$ 经过全局平均池化进行空间信息压缩.然后使用4个并行的 $1 \times 1$ 卷积来匹配池化后特征的维度,将匹配后的特征图先进行通道加法融合.之后通过两个连续的全连接层进一步编码,利用Sigmoid函数得到权重系数.最后采用自适应加权加法得到预测输出 $\sum U_{p_i}(P_i)\alpha_i$ ,其中, $\alpha_i$ 表示可学习的权重系数,可以根据输入特征调整每个尺度的损失重要性,使得网络能够拟合原始数据的尺度分布,实现更好的分割性能.

### 3.4 损失函数

损失函数能够通过缩小预测值和真实值之间的差距来优化模型,在皮肤病灶图像分割任务中,我们利用二元交叉熵损失BCE和加权IoU损失对每个阶段的损失进行监督.损失函数公式如式(12)所示,其中, $P_i$ 表示在多尺度密集并行解码模块获得输出特征图 $S_i$ 后经过4个并行的 $1 \times 1$ 卷积和上采样操作获取到的解码器第*i*阶段的预测掩码.权重参数 $\alpha$ 和 $\beta$ 分别控制二元交叉熵和加权IoU的函数损失贡献,在后续实验中,我们通过调整损失函数的权重 $\alpha$ 和 $\beta$ 来分析其对模型性能的影响,具体结果将在第4节中详细介绍. $G$ 表示真实值ground truth.

$$L_i = \alpha L_{IoU}(P_i, G) + \beta L_{BCE}(P_i, G) \quad (12)$$

因此解码器输出金字塔中所有阶段预测质量的总损失可以表示为:

$$L_{overall} = \sum L_i, i \in \{1, 2, 3, 4\} \quad (13)$$

## 4 实验

本节介绍实验中采用的数据集、实施细节以及评估指标,在 ISIC2017 和 ISIC2018 数据集上与 SOTA (state-of-the-art) 方法进行比较分析.最后,通过消融实验验证本文提出的每个模块的有效性.

### 4.1 数据集

我们在公开数据集 ISIC2017<sup>[27]</sup>和 ISIC2018<sup>[28]</sup>上进行了大量实验,为了统一模型的输入尺寸,所有图像的分辨率均调整为 256×256,以此对提出的模型进行训练和评估.这两个数据集均由国际皮肤成像协作组织 (ISIC) 提供,以 7:3 的比例将数据集随机划分为实验的训练集和测试集.表 1 显示了两个数据集的信息.

表 1 数据集样本分布详情

Dataset	Total sample	Training sample	Test sample
ISIC2017	2150	1500	650
ISIC2018	2694	1886	808

### 4.2 评估指标

实验使用准确度 ( $Acc$ )、敏感度 ( $Sen$ )、特异性 ( $Spe$ )、Dice 相似系数 ( $Dice$ ) 和并集平均交集 ( $mIoU$ ) 作为研究网络性能的评价依据.其中准确度是指准确分类的像素占图像中所有像素的比率;敏感度表示检测病变皮肤像素数量与真实病变皮肤像素数量比率的关系;特异性表示检测正常皮肤像素数量与真实正常皮肤像素数量比率的关系.定义如下:

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

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

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

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (17)$$

$$mIoU = \frac{TP}{TP + FP + FN} \quad (18)$$

其中,  $TP$ 、 $TN$ 、 $FP$  和  $FN$  分别代表真阳性、真阴性、假阳性和假阴性.  $TP$  和  $TN$  分别表示像素正确分类为皮肤病变和背景,而  $FP$  和  $FN$  分别表示像素错误分类为皮肤病变和背景.

### 4.3 与当今流行技术比较

为了充分证实本文方法的有效性,将本文实验结果与其他应用于医学图像分割的模型进行比较.表 2

和表 3 展示了多种医学图像分割模型在 ISIC2017 和 ISIC2018 数据集上的实验结果.

表 2 ISIC2017 数据集的对比实验结果 (%)

Model	$mIoU$	$Dice$	$Acc$	$Sen$	$Spe$
U-Net <sup>[4]</sup>	76.98	86.99	95.65	86.82	97.43
UTNetV2 <sup>[29]</sup>	77.35	87.23	95.84	84.85	98.05
TransFuse <sup>[18]</sup>	79.21	88.40	96.17	87.14	97.98
MobileNetV3 <sup>[30]</sup>	77.69	87.44	—	—	—
UNeXt-S <sup>[31]</sup>	78.26	87.80	95.95	87.04	97.74
MALUNet <sup>[32]</sup>	78.78	88.13	96.18	84.78	98.47
PraNet <sup>[33]</sup>	77.93	86.63	93.26	83.92	97.25
CaraNet <sup>[34]</sup>	77.75	86.67	95.93	86.32	96.97
EGE-UNet <sup>[25]</sup>	79.81	88.77	—	—	—
TransUNet <sup>[35]</sup>	76.37	85.86	94.23	80.41	95.65
Swin-UNet <sup>[36]</sup>	80.89	81.99	94.76	88.06	96.05
BCDU-Net <sup>[37]</sup>	79.20	78.11	91.63	76.46	97.09
GA-Net <sup>[38]</sup>	83.12	90.52	94.07	<b>88.07</b>	98.21
MSANet (Ours)	<b>85.82</b>	<b>90.12</b>	<b>96.47</b>	87.96	<b>98.59</b>

表 3 ISIC2018 数据集的对比实验结果 (%)

Model	$mIoU$	$Dice$	$Acc$	$Sen$	$Spe$
U-Net <sup>[4]</sup>	77.86	87.55	94.05	85.86	96.69
UTNetV2 <sup>[29]</sup>	78.97	88.25	94.32	87.60	96.48
TransFuse <sup>[18]</sup>	80.63	89.27	94.66	<b>91.28</b>	95.74
MobileNetV3 <sup>[30]</sup>	78.55	87.98	—	—	—
UNeXt-S <sup>[31]</sup>	79.09	88.33	94.39	87.15	96.72
MALUNet <sup>[32]</sup>	80.25	89.04	94.62	89.74	96.19
PraNet <sup>[33]</sup>	78.70	87.50	93.74	86.67	95.79
CaraNet <sup>[34]</sup>	78.20	87.00	96.23	87.43	97.67
EGE-UNet <sup>[25]</sup>	80.94	89.46	—	—	—
TransUNet <sup>[35]</sup>	80.60	88.03	95.57	83.14	96.73
Swin-UNet <sup>[36]</sup>	82.79	88.98	96.83	90.10	97.16
BCDU-Net <sup>[37]</sup>	81.10	85.10	93.70	78.50	<b>98.20</b>
MSANet (Ours)	<b>84.27</b>	<b>90.91</b>	<b>96.34</b>	90.97	97.81

从表 2 和表 3 中得出结论,与其他新型的改进模型相比,MSANet 模型性能具有显著提升,在 ISIC2017 和 ISIC2018 数据集上分别达到了 90.12% 和 90.91% 的  $Dice$  得分.对于 ISIC2018 数据集,本文模型在  $Dice$  系数上比 SOTA 方法提高了 1.64%,在  $mIoU$  上提高了 2.13%,其他指标也有类似的改进.除此之外,与目前医学领域中先进的 Transformer 架构模型相比,本文模型也有明显的提升.在 ISIC2017 数据集上,GA-Net 的  $Sen$  值达到了 88.07%,比 MSANet 高出 0.11%,这是因为 GA-Net 采用了独特的数据增强技术与最新的卷积自适应融合网络架构,从而有效提高了分割性能.然而,MSANet 的其他指标均高于 GA-Net.从整体上来说,MSANet 通过多层次特征的有效整合、目标区域

的精确捕捉以及边界细节的精准处理, 能够从丰富的皮损信息中提取关键特征, 从而大幅提升恶性黑色素瘤病变分割的准确性.

图4展示了不同方法对恶性黑色素瘤图像的分割结果, 本文方法能够生成更接近真实标签的预测掩码, 特别是在处理边缘模糊的病灶时表现得更加精细准确. 实验结果表明, MSANet 不仅能够精确定位病灶区域, 还能清晰分割皮肤病变的边缘轮廓, 进一步验证了其有效分割皮肤病变区域的优越性能.

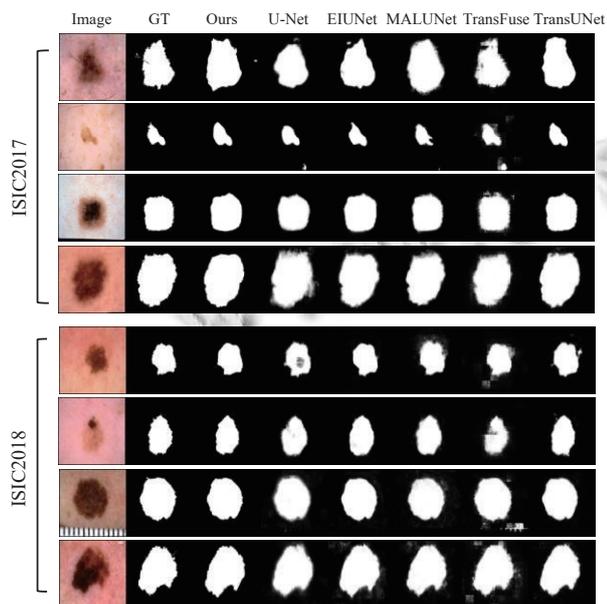


图4 不同网络分割结果可视化图

同时, 我们进行了不同损失函数权重设置的对比实验, 具体配置如表4所示. 表4展示了不同权重配置下的模型表现, 通过调整二元交叉熵损失 BCE 权重 $\alpha$ 和加权 IoU 损失权重 $\beta$ , 探索它们对皮肤病变分割性能的影响.

表4 损失函数权重设置对比 (%)

Dataset	Weight setting	Dice	mIoU	Acc	Sen	Spe
ISIC2017	$\alpha=1, \beta=1$	<b>90.12</b>	<b>85.82</b>	<b>96.47</b>	<b>87.96</b>	<b>98.59</b>
	$\alpha=1.2, \beta=1$	90.30	85.70	96.50	88.00	98.55
	$\alpha=1, \beta=1.2$	90.12	85.82	96.47	87.96	98.59
	$\alpha=0.8, \beta=1$	89.50	84.80	96.10	87.30	98.30
ISIC2018	$\alpha=1, \beta=1$	<b>90.91</b>	<b>84.27</b>	<b>96.34</b>	<b>90.97</b>	97.81
	$\alpha=1.2, \beta=1$	90.10	84.13	96.30	90.20	<b>97.85</b>
	$\alpha=1, \beta=1.2$	90.70	84.00	96.20	90.90	97.60
	$\alpha=0.8, \beta=1$	90.20	83.80	96.10	90.50	97.40

从表4中的实验结果可以看出, 当加权 IoU 损失权重 $\alpha$ 和二元交叉熵损失权重 $\beta$ 均设置为1时, 模型的

表现最佳. 这表明在此配置下, 加权 IoU 损失和二元交叉熵损失对模型的贡献保持了平衡, 二者共同优化模型的分割精度和定位能力.

#### 4.4 消融实验

我们在 ISIC2017 和 ISIC2018 数据集上进行了进一步的消融实验, 以证明每个模块对皮肤病变精准分割的必要性和有效性. 通过 *Dice*、*mIoU*、*Acc*、*Sen* 和 *Spe* 作为评估指标, 探讨各模块对整体模型分割性能的影响, 所有实验均使用相同的分层 PVT v2 编码器模块作为基础网络. 表5展示了每个模块在消融实验中的具体表现, 以验证其有效性.

表5 消融实验结果 (%)

Dataset	Metric	Dice	mIoU	Acc	Sen	Spe
ISIC2017	w/o SGA	88.15	84.32	89.50	87.80	91.30
	w/o MCA	87.64	83.93	88.20	86.20	90.50
	MSANet (Ours)	<b>90.12</b>	<b>85.82</b>	<b>96.47</b>	<b>87.96</b>	<b>98.59</b>
ISIC2018	w/o SGA	89.57	83.02	90.10	88.30	93.00
	w/o MCA	89.13	82.19	89.00	87.50	92.20
	MSANet (Ours)	<b>90.91</b>	<b>84.27</b>	<b>96.34</b>	<b>90.97</b>	<b>97.81</b>

我们进行了一项实验来评估“缺失 SGA (w/o SGA)”和“缺失 MCA (w/o MCA)”的 MSANet 网络效果. 表5显示, 缺少任何一个模块均会导致模型性能降低. 缺少 SGA 模块时, ISIC2017 数据集上的 *Dice* 系数从 90.12% 降至 88.15%, 在 ISIC2018 数据集上, *Dice* 系数从 90.91% 降至 89.57%, 这表明 SGA 模块通过提取全局和局部的多尺度信息, 确保了模型在不同尺度下准确捕捉和表达特征, 是提升整体分割性能的重要因素. 缺少 MCA 模块时, ISIC2017 和 ISIC2018 数据集上的 *Dice* 系数分别降至 87.64% 和 89.13%, 表明 MCA 模块能够精确定位病灶区域, 通过结合通道和位置信息, 显著提升模型对目标区域的识别能力. 消融实验结果清晰展示了每个模块对皮肤病灶分割模型的重要性, 进一步验证了它们对整体模型性能的贡献.

更进一步, 我们对消融实验结果进行了可视化分析, 如图5所示. 与基准模型相比, MSANet 表现更优. 缺少 SGA 模块时 (图5(c)), 皮肤分割图像边缘过于平滑, 病灶的边界区域未能有效识别, 这凸显了 SGA 模块在提高病灶全局和局部分割精细度方面的重要性; 缺少 MCA 模块时 (图5(d)), 模型易受背景噪声干扰, 导致模型对病变区域做出错误预测, 这表明 MCA 模块显著增强了目标区域的精准定位能力, 极大提升了模型对病灶区域定位的准确性.

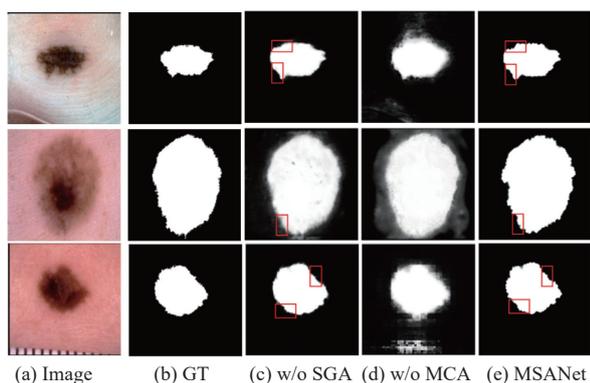


图5 消融实验结果的可视化图

## 5 结论与展望

本文提出了一种多尺度自适应上下文融合网络MSANet模型。该模型包含SGA和MCA两个模块,其中,SGA模块能够有效提取局部和全局的多尺度特征,增强模型对不同尺度信息的捕捉能力;MCA模块通过结合通道和位置信息,显著提升了模型对病灶区域的精准定位能力。通过在ISIC2017和ISIC2018两个数据集上的对比实验和消融实验,证明MSANet模型能够有效应对病灶区域尺度变化大、误分割和边界模糊等挑战,显著提升皮肤病变区域的分割效果。尽管MSANet在多项实验中表现出色,但其计算复杂度相对较高,可能影响实际应用的处理效率。未来工作将致力于优化模型结构,提高计算效率,探索更轻量化的注意力机制,以增强模型对小尺度和复杂形状病变区域的处理能力。此外,MSANet目前主要针对皮肤病变的图像分割任务,未来将进一步探索在其他医学影像分割任务中的应用潜力。

### 参考文献

- 1 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]
- 2 Li W, Raj ANJ, Tjahjedi T, *et al.* Digital hair removal by deep learning for skin lesion segmentation. Pattern Recognition, 2021, 117: 107994. [doi: 10.1016/j.patcog.2021.107994]
- 3 Vesal S, Ravikumar N, Maier A. SkinNet: A deep learning framework for skin lesion segmentation. Proceedings of the 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC). Sydney: IEEE, 2018. 1–3.
- 4 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.
- 5 Zhou ZW, Siddiquee MMR, Tajbakhsh N, *et al.* UNet++: A nested U-Net architecture for medical image segmentation. Proceedings of the 4th International Workshop, DLMIA 2018, and the 8th International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Granada: Springer, 2018. 3–11.
- 6 Jha D, Smedsrud PH, Riegler MA, *et al.* ResUNet++: An advanced architecture for medical image segmentation. Proceedings of the 2019 IEEE International Symposium on Multimedia (ISM). San Diego: IEEE, 2019. 225–230.
- 7 Peng SR, Guo CH, Wu X, *et al.* U2Net: A general framework with spatial-spectral-integrated double u-net for image fusion. Proceedings of the 31st ACM International Conference on Multimedia. Ottawa: ACM, 2023. 3219–3227.
- 8 Chen LC, Papandreou G, Schroff F, *et al.* Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587, 2017.
- 9 Xie SN, Girshick R, Dollár P, *et al.* Aggregated residual transformations for deep neural networks. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 5987–5995.
- 10 Howard AG, Zhu ML, Chen B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.
- 11 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19.
- 12 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations. OpenReview.net, 2021.
- 13 Liu Z, Lin YT, Cao Y, *et al.* Swin Transformer: Hierarchical vision Transformer using shifted windows. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 9992–10002.
- 14 Wang WH, Xie EZ, Li X, *et al.* PVT v2: Improved baselines with pyramid vision Transformer. Computational Visual Media, 2022, 8(3): 415–424.
- 15 Wu HP, Xiao B, Codella N, *et al.* CvT: Introducing convolutions to vision Transformers. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 22–31.

- 16 Yuan K, Guo SP, Liu ZW, *et al.* Incorporating convolution designs into visual Transformers. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 559–568.
- 17 Xie EZ, Wang WH, Yu ZD, *et al.* SegFormer: Simple and efficient design for semantic segmentation with Transformers. Proceedings of the 35th International Conference on Neural Information Processing Systems. Curran Associates Inc., 2021. 12077–12090.
- 18 Zhang YD, Liu HY, Hu Q. TransFuse: Fusing Transformers and CNNs for medical image segmentation. Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Strasbourg: Springer, 2021. 14–24.
- 19 Alahmadi MD. Multiscale attention U-Net for skin lesion segmentation. IEEE Access, 2022, 10: 59145–59154. [doi: [10.1109/ACCESS.2022.3179390](https://doi.org/10.1109/ACCESS.2022.3179390)]
- 20 Hasan MK, Dahal L, Samarakoon PN, *et al.* DSNet: Automatic dermoscopic skin lesion segmentation. Computers in Biology and Medicine, 2020, 120: 103738. [doi: [10.1016/j.combiomed.2020.103738](https://doi.org/10.1016/j.combiomed.2020.103738)]
- 21 Tang P, Liang QK, Yan XT, *et al.* Efficient skin lesion segmentation using Separable-Unet with stochastic weight averaging. Computer Methods and Programs in Biomedicine, 2019, 178: 289–301. [doi: [10.1016/j.cmpb.2019.07.005](https://doi.org/10.1016/j.cmpb.2019.07.005)]
- 22 Tong XZ, Wei JY, Sun B, *et al.* ASCU-Net: Attention gate, spatial and channel attention U-Net for skin lesion segmentation. Diagnostics, 2021, 11(3): 501. [doi: [10.3390/diagnostics11030501](https://doi.org/10.3390/diagnostics11030501)]
- 23 Rahman MM, Marculescu R. Multi-scale hierarchical vision Transformer with cascaded attention decoding for medical image segmentation. Medical Imaging with Deep Learning. Nashville: PMLR, 2023. 1526–1544.
- 24 Du SY, Bayasi N, Hamarneh G, *et al.* AViT: Adapting vision Transformers for small skin lesion segmentation datasets. Proceedings of the 26th International Conference on Medical Image Computing and Computer Assisted Intervention. Vancouver: Springer, 2023. 25–36.
- 25 Ruan JC, Xie MY, Gao JS, *et al.* EGE-UNet: An efficient group enhanced U-Net for skin lesion segmentation. Proceedings of the 26th International Conference on Medical Image Computing and Computer Assisted Intervention. Vancouver: Springer, 2023. 481–490.
- 26 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13708–13717.
- 27 Berseth M. ISIC 2017—skin lesion analysis towards melanoma detection. arXiv:1703.00523, 2017.
- 28 Codella N, Rotemberg V, Tschandl P, *et al.* Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). arXiv:1902.03368, 2019.
- 29 Gao YH, Zhou M, Liu D, *et al.* A data-scalable Transformer for medical image segmentation: Architecture, model efficiency, and benchmark. arXiv:2203.00131, 2022.
- 30 Howard A, Sandler M, Chen B, *et al.* Searching for MobileNetV3. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 1314–1324.
- 31 Valanarasu JMJ, Patel VM. UNeXt: MLP-based rapid medical image segmentation network. Proceedings of the 25th International Conference on Medical Image Computing and Computer Assisted Intervention. Singapore: Springer, 2022. 23–33.
- 32 Ruan JC, Xiang SC, Xie MY, *et al.* MALUNet: A multi-attention and light-weight U-Net for skin lesion segmentation. Proceedings of the 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Las Vegas: IEEE, 2022. 1150–1156.
- 33 Fan DP, Ji GP, Zhou T, *et al.* PraNet: Parallel reverse attention network for polyp segmentation. Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention. Lima: Springer, 2020. 263–273.
- 34 Lou AG, Guan SY, Ko H, *et al.* CaraNet: Context axial reverse attention network for segmentation of small medical objects. Proceedings of the 2022 SPIE 12032, Medical Imaging 2022: Image Processing. San Diego: SPIE, 2022: 120320D.
- 35 Chen JN, Lu YY, Yu QH, *et al.* TransUNet: Transformers make strong encoders for medical image segmentation. arXiv:2102.04306, 2021.
- 36 Cao H, Wang YY, Chen J, *et al.* Swin-Unet: Unet-like pure Transformer for medical image segmentation. Proceedings of the 2023 European Conference on Computer Vision. Tel Aviv: Springer, 2023. 205–218.
- 37 Azad R, Asadi-Aghbolaghi M, Fathy M, *et al.* Bi-directional ConvLSTM U-Net with densely connected convolutions. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul: IEEE, 2019. 406–415.
- 38 Zhou LS, Liang LM, Sheng XQ. GA-Net: Ghost convolution adaptive fusion skin lesion segmentation network. Computers in Biology and Medicine, 2023, 164: 107273. [doi: [10.1016/j.combiomed.2023.107273](https://doi.org/10.1016/j.combiomed.2023.107273)]

(校对责编:王欣欣)