

基于边缘关键点和边缘注意力的黑色素瘤图像分割方法

王娜¹, 贾伟^{1,2}, 赵雪芬¹, 高宏娟^{1,2}

1. 宁夏大学信息工程学院, 宁夏 银川 750021; 2. 宁夏“东数西算”人工智能与信息安全重点实验室, 宁夏 银川 750021

【摘要】黑色素瘤图像的高精度分割对早期诊断和提高患者的生存率至关重要。然而,由于黑色素瘤的边缘区域模糊,呈现出不规则的形状,使得现有分割方法难以获取边缘特征信息,影响了黑色素瘤图像分割的准确率。为解决该问题,提出了一种基于边缘关键点和边缘注意力的黑色素瘤图像分割方法。首先,在编码器中设计了点渲染的边缘关键点选择模块和组合卷积变压器块,通过边缘关键点选择,引导获取边缘的局部特征和全局特征。然后,在编码器中设计边缘细化模块用于细化深层网络的边缘特征,最后,在跳连接中设计了多尺度边缘注意力模块,能够更好地捕获图像多尺度的边缘形状特征。将所提方法在ISIC 2018和PH²两个数据集上进行实验,实验结果表明,与现有分割方法相比,所提方法具有较好的分割性能。

【关键词】黑色素瘤;图像分割;多尺度;边缘注意力;边缘关键点选择

【中图分类号】R318;TP391

【文献标志码】A

【文章编号】1005-202X(2024)10-1225-12

Melanoma image segmentation method based on edge key points and edge attention

WANG Na¹, JIA Wei^{1,2}, ZHAO Xuefen¹, GAO Hongjuan^{1,2}

1. School of Information Engineering, Ningxia University, Yinchuan 750021, China; 2. Ningxia Key Laboratory of Artificial Intelligence and Information Security for Channeling Computing Resources from the East to the West, Yinchuan 750021, China

Abstract: High-precision segmentation of melanoma images is crucial for early diagnosis and improving patient survival. However, the blurring of the edge region of melanoma, which presents irregular shapes, makes it difficult for existing segmentation methods to obtain edge feature information, affecting the accuracy of melanoma image segmentation. To solve this problem, a melanoma image segmentation method based on edge key points and edge attention is proposed. An edge key point selection module for point rendering and a combined convolution transformer block are designed in the encoder to guide the acquisition of local and global features of the edge by selecting edge key points. Then, the edge refinement module is designed in the encoder to refine the edge features of the deep network, and finally, the multi-scale edge attention module is designed in the skip connection, which enables the capture of the edge shape features at multiple scales. The tests on two datasets (ISIC 2018 and PH²) demonstrate that the proposed method has better segmentation performance than the existing segmentation methods.

Keywords: melanoma; image segmentation; multi-scale; edge attention; edge key point selection

前言

黑色素瘤是一种罕见但危险的皮肤癌,其高度转移性导致了高致死率^[1]。虽然黑色素瘤仅占皮肤癌病例的1%,但却造成了大多数皮肤癌死亡病例。

近30年来,黑色素瘤发病人数每年稳步增加^[2]。由于早期黑色素瘤症状不明显,容易被误认为良性痣,因此延迟诊断会显著降低治疗率。早期诊断对于黑色素瘤的治疗和预后至关重要。如果及早发现,并在发病初期进行相应治疗,那么患者的治愈率是非常高的,可达到99%。然而,延迟诊断会导致治愈率急剧下降至14%^[3]。黑色素瘤病灶区域具有模糊的边缘、不均匀的纹理、多变的大小和形状以及其他影响因素:如毛发、手术标记等,给诊断带来了很大的挑战^[4]。即使对于专业皮肤科医生来说,人工检查皮肤镜图像也是一件极易出错且耗时的工作,并且存在着很大的主观性。随着计算机视觉技术的飞速发

【收稿日期】2024-04-15

【基金项目】国家自然科学基金(62062057,12062021);宁夏自然科学基金(2022AAC03005)

【作者简介】王娜,硕士研究生,主要研究方向:医学图像处理与分析,
E-mail: 2970722507@qq.com

【通信作者】贾伟,博士,副教授,主要研究方向:医学图像处理与分析,
E-mail: jiawnx@163.com

展,图像分割技术被越来越多地应用于黑色素瘤图像的分析,可靠的自动分割方法可以很好地帮助医生进行诊断^[5]。传统的黑色素瘤图像分割方法有边缘检测^[6]、阈值估计^[7]、区域生长聚类^[8]等,这些方法主要依靠人工标注,不仅实现难度大、效率低,而且分割效果也不理想。近年来,研究者们针对医学图像分割问题,提出了许多基于U-Net框架的网络,例如AttU-Net^[9]、ResUNet^[10]、UNet3+^[11]等。越来越多的学者致力于提高黑色素瘤图像分割的准确性,Sarker等^[12]提出了一种黑色素瘤图像分割模型,将跳跃连接和金字塔池化相结合,并引入负对数似然和端点误差的损失函数来呈现边缘形状。Hanselmann等^[13]提出了一种基于多注意机制的CANet,该网络综合考虑空间、信道和尺度三个维度的信息,并结合上下文信息和门控机制对边缘进行优化。Pang等^[14]提出了一种聚合交互策略,能够有效利用多级特征防止因黑色素瘤图像分辨率的差异而造成的干扰。由于黑色素瘤图像边缘不清、形态多样,给预测带来了很大的困难。虽然已有的基于U-Net的分割方法在分割中获得了显著的效果,但是由于这些分割方法使用卷积神经网络,只能捕获局部特征,难以获得全局特征信息,导致分割后的图像仍存在形态不清晰、边缘不清晰等问题。Transformer也被广泛应用于医学图像的各类具体任务中,与具有局部感受野的卷积神经网络相比,Transformer能够捕捉全局信息^[15]。对于皮肤黑色素瘤分割任务,Wang等^[16]在其研究中提出了一种全新的边缘敏感Transformer结构,将边缘敏感门控模块引入到Transformer中,使其能够充分利用边缘关联的先验信息来获取更多的局部信息。Wu等^[17]提出了一种特征适应性的Transformer网络,该网络融合了Transformer与CNN的双重分支结构,能够较好地实现图像的特征提取,但对色彩变化较大、对比度较低的黑色素瘤图像的边缘分割效果不理想。

将U型网络和Transformer相结合是近几年的一个研究热点^[18]。Swin-Unet^[19]是一个以Transformer作为主干的U形网络,通过利用Transformer来弥补U-Net自身对远程相关性的捕捉不足缺陷。与传统方法相比,Swin-Unet方法可以更好地提高黑色素瘤图像的分割精度。但是,由于黑色素瘤的边缘区域模糊,呈现出不规则的形状,Swin-Unet难以准确分割出黑色素瘤的边缘区域,导致分割效果不佳。为了解决这个问题,在Swin-Unet的基础上,提出一种基于边缘关键点和边缘注意力的黑色素瘤图像分割方法(Melanoma Image Segmentation Method Based on Edge Key Points and Edge Attention, MSBKA)。本文

主要贡献如下:(1)提出了一种黑色素瘤图像分割方法MSBKA,该方法在编码器部分设计了点渲染的边界关键点选择模块(Edge Key Points Selection Module For Point Rendering, EKSM)、组合卷积变压器块(Combination Convolutional Transformer Block, CCTB)和边缘细化模块(Edge Refinement Module, ERM),通过EKSM选择的边缘关键点逐层引导CCTB对边缘局部特征和全局特征的提取,ERM负责在深层网络中细化边缘特征。(2)为突出边缘特征,抑制边缘周围的噪声干扰,在跳连接中设计了多尺度边缘特征注意力模块(Multi-scale Edge Feature Attention Module, MEAM),在多个尺度上聚合边缘特征信息,并将其传输至解码器,以区分各种边缘形状的特征。

1 黑色素瘤图像的分割方法

图1是本文所提方法MSBKA的总体结构图,设计了一种对称的编码器和解码器架构,包含4个主要组成部分:编码器、瓶颈器、解码器和跳跃连接。在编码器部分:把输入的图像分成4×4大小的相互不重叠的图像补丁,每个补丁的特征维数为48。补丁通过线性嵌入层,将各个补丁的特征维度嵌入到EKSM,EKSM负责选择正确的边缘关键点,从而准确定位黑色素瘤的边缘位置和基本形状。CCTB负责从全局和局部两个角度进行特征学习,得到分层的特征表示。补丁合并层包括下采样和增加维度。在最后一组CCTB之后加入了一个ERM,目的是补充边缘特征细节,使得分割边缘精细化。瓶颈器是由两个Swin Transformer块组成的,通过压缩编码器所得的高维特征,降低了特征维数,作用是将编码器的高级特征与解码器的上采样特征进行融合。解码器由3个分层的解码块构成,每个解码块由一个补丁扩展层和两个Swin Transformer块组成,目的是将瓶颈器的特征逐步上采样,直至恢复到与原始输入图像相同的分辨率。此外,把MEAM加入到跳跃连接中,能够使分割网络很好地适应黑色素瘤病变区域形状的变化。

1.1 编码器的改进

1.1.1 组合卷积变压器块 对于黑色素瘤的分割来说,捕获全局和局部信息是至关重要的。全局信息可以帮助模型理解病变的整体结构,确定其与背景的关系^[20],有助于更精准地定位病变区域。而局部信息则可以捕捉病变区域的边缘特征,进一步提升模型的性能。为了获得更充分的全局信息空间细节并且建立长距离像素依赖性,设计了CCTB,其结构如图2所示。首先,使用深度可分离卷积来替代

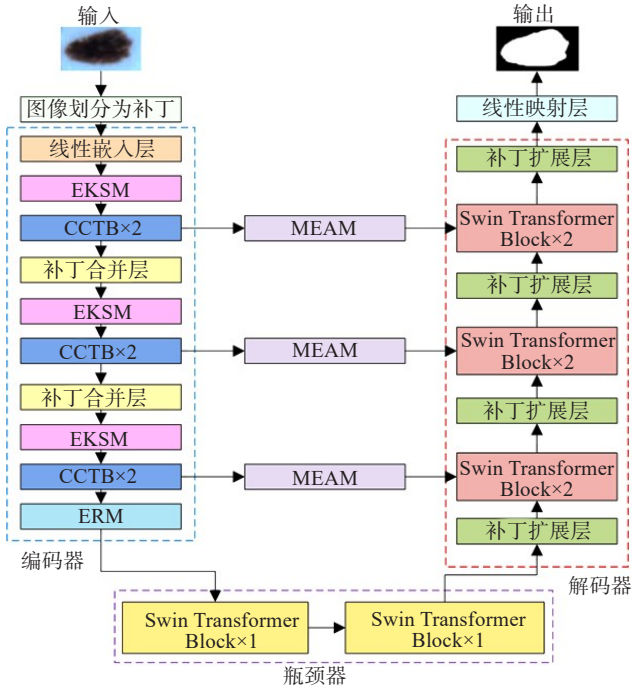


图1 MSBKA的总体结构图

Figure 1 Overall structure diagram of MSBKA

Swin-Unet 网络中的线性投影^[21]。具体来说,在一个二维重塑的令牌上执行3个深度可分离卷积操作。与线性投影相比,由深度可分离卷积构成的投影层不仅能够更好地捕捉局部空间语义信息,而且无需位置编码。接着,将这3个卷积操作后得到的查询向量 Q 、关键向量 K 和数值向量 V 传递给代理注意力机制进行计算^[22],得到特征 F_A ,这个特征 F_A 随后与原图相加。相加后的结果通过一个归一化层(Layer Normalization, LN)和一个残差多层感知器(Residual MLP, ResMLP)进行处理,得到的结果再与原图相加,最终得到的输出不仅保留了原图的信息,还融入了通过深度可分离卷积和代理注意力机制提取的局部空间语义信息。CCTB的计算公式如下:

$$Q = \text{DepthwiseConv}_1(X) \quad (1)$$

$$K = \text{DepthwiseConv}_2(X) \quad (2)$$

$$V = \text{DepthwiseConv}_3(X) \quad (3)$$

$$F_A = \text{AAT}(Q, K, V) \quad (4)$$

$$Y = \text{ResMLP}(\text{LN}(F_A + X)) + F_A \quad (5)$$

其中, X 表示输入特征图, Y 表示输出特征图, DepthwiseConv_1 、 DepthwiseConv_2 、 DepthwiseConv_3 分别为卷积核大小为1、2、3的深度可分离卷积,AAT为代理注意力。

在 Swin-Unet 移动窗口的基础上采用了代理注意力,它由两个 Softmax 注意力组成,运算上等效为一种广义的线性注意力,将两个高性能注意力 Softmax 注意力和线性注意力 Matmul 有机结合起来,

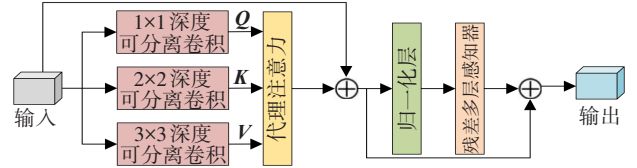


图2 CCTB的结构图

Figure 2 Structure diagram of CCTB

充分发挥两者的优势,代理注意力结构如图3所示。代理注意力在传统的注意力三元组(Q, K, V)中引入了一组额外的代理向量 A ,定义了一种新的四元注意力机制(Q, A, K, V)。首先把 A 当作查询向量, Q 与 K 进行 Matmul 和 Softmax 计算,得到的结果再与 V 进行 Matmul 计算,得到代理特征 V_Q 。然后,以 A 作为关键向量, K 和 Q 进行第二次 Matmul 和 Softmax 计算,得到代理特征 V_K 。最后,将代理特征 V_Q 当作数值向量, V 和 V_K 再进行一次 Matmul 计算,获得最终输出特征信息 F_A 。这样,就不需要计算 Q 和 K 之间相似度,而是通过代理向量来完成每个 Query-Key 之间的信息互换。

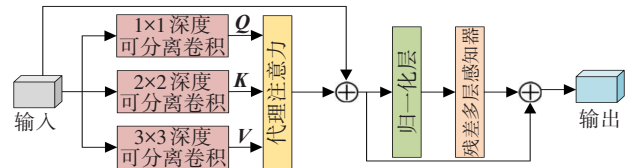


图3 代理注意力结构图

Figure 3 Structure diagram of agent attention

1.1.2 点渲染的边界关键点选择模块 黑色素瘤病变区域与周边皮肤组织相近,目标病变区域与背景区域的分界线模糊不清、对比度较低,使得分割任务变得更加困难,导致图像分割结果与标记之间在边缘部分存在一定偏差。在模型的特征提取过程中,对边缘点的识别十分重要,仅靠边缘空间注意力无法感知模糊病变区域中复杂的边缘。因此,本文提出了EKSMM,结构如图4所示。该模块先使用一个自适应边缘增强模块(Adaptive Edge Enhancement Module, AEEM)增强输入图像的高频信息部分,再经过一个PointRend模块正确选择边缘点^[23]。

AEEM用于增强黑色素瘤病变区域的边缘特征,定位模糊边缘的位置。首先输入的特征图经过一个 1×1 卷积层和一个批归一化(Batch Normalization, BN)层, 1×1 卷积的作用是降维,卷积操作都附加BN层以进行正则化。利用特征金字塔模块(Feature Pyramid Module, FPM)来提取不同分辨率的特征^[24]。

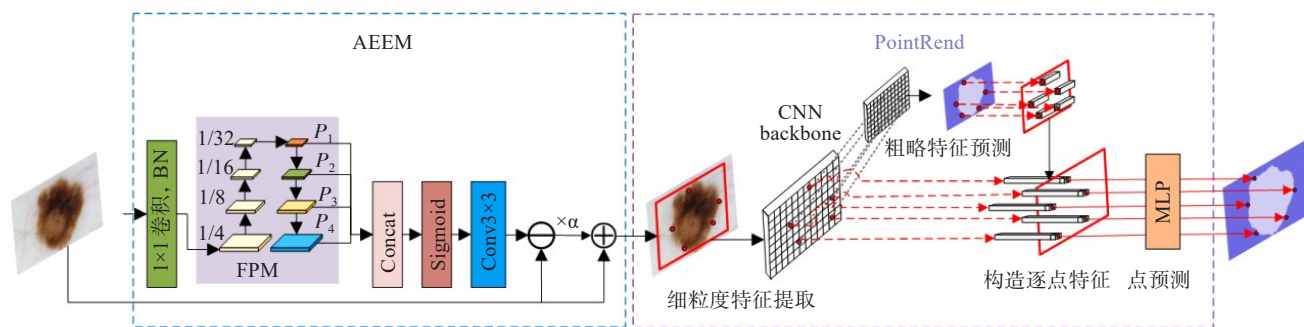


图4 EKSM的结构图

Figure 4 Structure diagram of EKSM

在FPM中,每一层都会生成相应尺度的特征图,这些特征图的大小分别为原图尺寸的1/4、1/8、1/16和1/32。将FPM输出的不同尺度特征图 P_1 、 P_2 、 P_3 、 P_4 进行连接操作。通过Sigmoid函数,激活模糊边缘上的候选点图。再通过一个卷积核大小为 3×3 的可训练卷积滤波器平滑输入图像,在模型训练时优化滤波器参数。随后,平滑后的图像被输入图像减去,以获得图像中的高频信息。接着,通过一个乘法因子 α 控制增强。最后,将增强后的图像和输入图像叠加。模型训练时,自适应表现在更新卷积滤波器的内核参数和对局部的高频信息进行处理。AEEM的计算公式如下:

$$P_i = \text{FPM}(\text{Conv}_1(\text{BN}(F_{\text{in}}))) \quad (6)$$

$$F_{\text{out}} = F_{\text{in}} + (F_{\text{in}} - \text{Conv}_3(\delta(C(P_1, P_2, P_3, P_4)) \cdot \alpha)) \quad (7)$$

其中, F_{out} 表示输出特征图, F_{in} 表示输入特征图, δ 表示Sigmoid函数,FPM表示特征金字塔模块, $P_i(i=1, 2, 3, 4)$ 表示经过FPM得到的特征图, Conv_1 表示 1×1 的卷积层, Conv_3 表示可训练卷积滤波器, $\alpha \in [0, 1]$ 。

PointRend是点渲染的边缘关键点选择模块的核心方法,该方法是一个迭代上采样的过程。在持续进行上采样的过程中,选取 N 个与周围点不同的点,这些点位于黑色素瘤病变区域的边缘。对于这 N 个点,首先进行细粒度的特征提取,随后利用CNN backbone进行更深入的特征学习^[25]。基于学习到的特征,进行初步的特征预测,并计算每个点相应的表征向量。这个表征向量组合了细粒度特征和粗略特征预测得到的特征,在选定的点上构造逐点特征。在细粒度特征方面,针对每个采样点进行了特征向量提取。然而,对于细粒度特征的特征映射,可能仅涵盖较低级别的信息,因此需要进行粗略的特征预测进行补充,并提供更多的全局背景信息。最后,使用MLP对这个表征向量进行预测,将这个点的归属进行分类,判断这个点是否是黑色素瘤边缘点,从而提升边缘分割的效果。

1.1.3 边缘细化模块 黑色素瘤的形状往往是不规则的,其周边的组织也是不规则的,其形状可分为表浅蔓延型、雀斑型、结节型等,呈斑状。考虑到黑色素瘤图像边缘形状的不规则性,在编码部分不断下采样过程中,其边缘特征细节难以捕捉,这可能导致分割错误的概率增加。本文设计了ERM,确保MSBKA网络对黑色素瘤的分割结果达到满意的效果。设计该模块的原理在于利用亚像素卷积和Sigmoid函数在特征图中引入更多的非线性因素,其结构如图5所示。首先,输入的特征图通过 1×1 的卷积操作以降低其通道维度。随后,这些特征图会分别进入两条并行支路进行处理。在第一条支路上,特征图会先经过ReLU激活函数的处理,然后通过亚像素卷积进行上采样,以恢复特征图的尺寸并进行边缘特征细化。处理完成后,得到的结果会与原始输入特征图进行相加操作得到特征图 Y 。在第二条支路中,特征图经过两次Sigmoid激活函数的处理,然后与经过 1×1 次卷积处理后的特征图进行点乘运算,得到特征图 S 。最后,将最后两条支路的输出进行拼接操作,得到最终的输出结果。ERM的计算公式如下:

$$Y = F_{\text{in}} + Y\text{Conv}(R(\text{Conv}_1(F_{\text{in}}))) \quad (8)$$

$$S = \text{Conv}_1(F_{\text{in}}) \odot \text{Tanh}(\delta(\text{Conv}_1(F_{\text{in}}))) \quad (9)$$

$$F_{\text{out}} = \text{Concat}(Y, S) \quad (10)$$

其中, F_{out} 表示输出特征图, F_{in} 表示输入特征图, Y 表示第一条支路得到的特征图, S 表示第二条支路得到的特征图, Conv_1 是 1×1 的卷积层, $Y\text{Conv}$ 则是亚像素卷积层, R 表示ReLU激活函数, δ 代表Sigmoid激活函数,Concat则表示拼接操作。

1.2 多尺度边缘特征注意力

黑色素瘤病灶区域形状复杂,呈斑状,可分为浅表扩散型、雀斑型、结节型等。单一尺度的设计方法往往不能准确分割黑色素瘤图像病灶区域。同时,直接将编码器的特征传递到解码器会导致糟糕的分割结果。为优化黑色素瘤病灶图像分割准确度,本

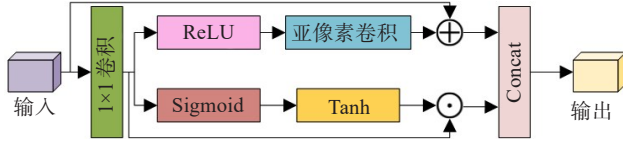


图5 ERM结构图

Figure 5 Structure diagram of ERM

文提出在下采样过程中将MEAM加入到跳跃连接中,在不同尺度上将编码阶段的通道、像素和空间特征聚合在一起再传递到解码器中,能够很好地适应黑色素瘤病变区域的尺度和形状的变化,从而实现黑色素瘤图像的分割。

MEAM如图6所示,对输入特征图进行了3个并行操作,分别是1×1卷积、膨胀率为2的3×3空洞卷积和膨胀率为4的5×5空洞卷积。卷积核大小为1×1的卷积可以压缩通道数,使用3×3的卷积核可以提取较小范围的特征信息,关注局部的细节信息,膨胀率为4的5×5空洞卷积将感受野逐渐扩大,可关注更大范围的特征信息,使用3个不同尺度的卷积核,可以获得不同尺度的特征信息。在1×1卷积支路上采用优化通道注意力(Optimize Channel Attention, OCA),膨胀率为4的5×5空洞卷积支路上采用残差像素注意力(Residual Pixel Attention, RPA),膨胀率为2的3×3空洞卷积支路上采用边缘空间注意力(Edge Space Attention, ESA),将这3个注意力得到不同尺度的特征图进行拼接,通过1×1卷积将通道数与输入保持一致,随即得到输出和原图大小一致的特征图。

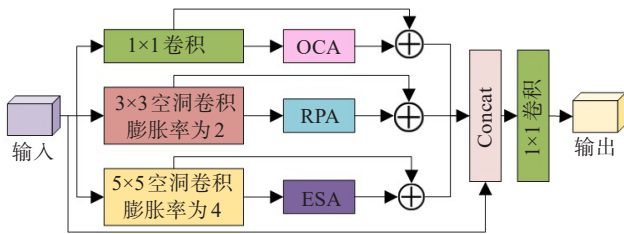


图6 MEAM结构图

Figure 6 Structure diagram of MEAM

1.2.1 优化通道注意力 OCA的结构如图7所示,最前端添加了一个3×3的卷积层、一个BN层和一个ReLU层,统一了输出的通道数量,从而优化每个维度的特征图,特别是低维特征。利用最大池化操作后对整个特征块进行覆盖,用以提取黑色素瘤图像的纹理信息,平均池化操作可以很好地保留背景信息,也减少了过拟合现象。然后输入到共享的多层感知器以实现特征转换和信息重组,得到特征图 F_M

和 F_A 。再将特征信息 F_M 和 F_A 相加,经过一个Sigmoid函数进行激活。OCA的计算公式如下:

$$F_M = C_{CBR}(\text{MP}(\text{SMLP}(F_{in}))) \quad (11)$$

$$F_A = C_{CBR}(\text{AP}(\text{SMLP}(F_{in}))) \quad (12)$$

$$F_{out} = \delta(F_M + F_A) \quad (13)$$

其中, F_{out} 表示输出特征图, F_{in} 表示输入特征图, δ 表示Sigmoid激活函数, C_{CBR} 表示一个3×3的卷积层和一个BN层和一个ReLU层,MP表示最大池化,AP表示平均池化,SMLP表示共享感知器层。

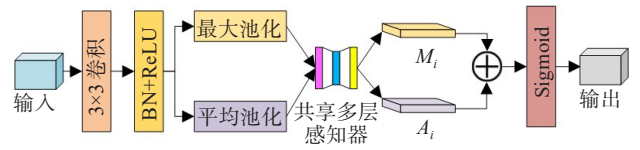


图7 OCA的结构图

Figure 7 Structure diagram of OCA

1.2.2 残差像素注意力 在黑色素瘤分割过程中,前景和背景之间边缘处的像素需要更多地关注周围的像素,以更好地提取这部分特征信息。因此本文提出了RPA,其结构如图8所示。在像素注意力之前加入一个残差块,学习残差信息有助于提高颜色映射性能。残差块由两个3×3卷积层、1个ReLU激活函数和1个BN层组成。在卷积层中,第一层旨在增加特征数量映射,第二层卷积连接前面的卷积层。ReLU激活函数和BN层生成所有特征映射。在残差块的末尾,采用像素注意力(Pixel Attention, PA)^[26],它由2个1×1卷积层、一个ReLU层、1个Sigmoid函数构成。经过PA的特征图与残差块得到的特征图相乘,再与输入特征图相加得到最终的像素特征图。RPA的计算公式如下:

$$F_{out} = F_{in} + (C_{CBRC}(F_{in})) \otimes (\text{Conv}_1(R(\text{Conv}_1(F_{in})))) \quad (14)$$

其中, F_{out} 表示输出特征图, F_{in} 表示输入特征图, δ 表示Sigmoid函数, C_{CBRC} 表示一个3×3的卷积层、一个BN层、一个ReLU层和一个3×3的卷积层, R 表示ReLU激活函数, Conv_1 表示1×1的卷积层。

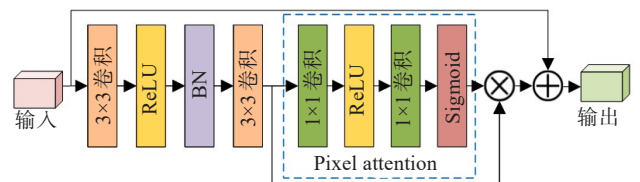


图8 RPA的结构图

Figure 8 Structure diagram of RPA

1.2.3 边缘空间注意力 图9是所提出的ESA结构。边缘信息可以被视为一种空间特征。在此基础上,边缘空间注意力采用了Sobel边缘检测方法。Sobel算子用于检测边缘,能够提取水平和垂直方向上的边缘^[27],随后将这两个正交方向的边缘进行组合。其中, S_h 代表水平方向的边缘,而 S_v 则代表垂直方向的边缘。从上层的输出得到输入特征 $F_i(i=1,2,3)$,将输入特征分别与水平边缘和垂直边缘相乘得到 F_{ih} 和 F_{iv} ,再对 F_{ih} 和 F_{iv} 沿着水平和垂直方向进行卷积计算,最后将两个方向的边缘信息通过正交计算合并。得到的边缘特征通过全局最大池化和全局平均池化选择边缘特征中的最大值和平均值。将池化后的结

果连接起来后,进行卷积运算重新获得空间权重,并使用Softmax函数来激活特征。最终,激活的特征与输入特征相乘得到输出结果。ESA的计算公式如下:

$$F_s = \sqrt{F_{iv}^2 + F_{ih}^2} \quad (15)$$

$F_{out} = F_{in} \otimes \delta(\text{Conv}_1(C(\text{GMP}(F_s), \text{GAP}(F_s))))$ (16)
其中, F_{out} 表示输出特征图, F_{in} 表示输入特征图, F_s 表示经过Sobel算子的边缘特征图, δ 表示Sigmoid函数, Conv_1 表示 1×1 的卷积层, C 表示Concat拼接操作, GMP 表示全局最大池化, GAP 表示全局平均池化。

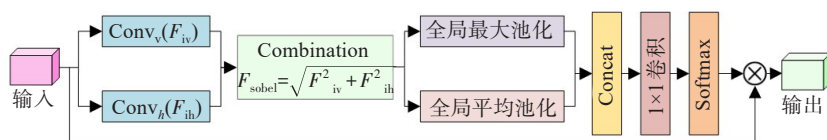


图9 ESA的结构图

Figure 9 Structure diagram of ESA

1.3 损失函数

黑色素瘤分割可看作像素级二分类问题,病变区域之外都是背景^[28],类别不平衡是黑色素瘤图像分割任务中的一个主要挑战。在网络训练中,一般采用二分类交叉熵作为损失函数进行监督。然而,仅仅依靠二分类交叉熵损失函数并不能充分关注边缘,难以精确分割病变边缘。所以,本文采用了一种基于边缘的形状感知损失函数,在边缘难以分割的情况下,通过增加基于形状的系数来降低损失。为了进一步优化模型,提出了一个联合损失函数,它由二分类交叉熵损失和基于边缘的形状感知损失函数组成。同时,在模型训练过程中分别为两个损失函数分配动态权重,实现针对性监督,同时监督黑色素瘤病变区域的整体特征和边缘特征的提取。联合损失函数公式如下:

$$L = \lambda L_{\text{BCE}} + (1 - \lambda) L_{\text{SA}} \quad (17)$$

其中, L_{BCE} 是二分类交叉熵损失函数, L_{SA} 是基于边缘的形状感知损失函数。 λ 是取值范围为 $[0, 1]$ 的权重参数, λ 初始值设定为1,模型专注于病变区域的空间特性。随着训练的深入, L_{BCE} 权重逐渐减小直至趋近于0, L_{SA} 权重逐渐增大到1,在这个过程中,模型越来越关注边缘细节特征。

2 实验结果与分析

2.1 数据集介绍

实验所用数据集是国际皮肤成像协会(the International Skin Imaging Collaboratio, ISIC)2018年在国际皮肤镜图像分割竞赛中提供的ISIC 2018分割数据集和来自西班牙佩德罗医院的PH²数据集。ISIC 2018数据集包含519张黑色素瘤图像及其真实标签,图像尺寸从 556×679 像素到 4499×6748 像素不等。在实验中对其进行二维切片并且随机分为训练集、验证集和测试集,364张图像用于训练,52张图像用于验证,103张图像用于测试。PH²数据集包含40张黑色素瘤图像,所有图像的尺寸均为 577×769 像素。其中,20张图像被用于训练,5张图像被用于验证,剩下15张图像用于测试。为了丰富数据的多样性,对这些图像应用了图像转换技术,包括水平和垂直翻转、随机旋转以及随机裁剪等。最终,将图像的分辨率调整至 224×224 像素进行模型训练。

2.2 实验设置

所有实验都是linux操作系统上进行,运行内存大小为32 GB,GPU为NVIDIA GeForce RTX 3090的硬件平台上进行的,开发环境为基于深度学习框架Pytorch3.8。使用ALSEM优化器进行训练,初始学习率设置为 $10e-4$,重量衰减 $10e-8$,批次大小为16,迭代100次。

2.3 评价指标

为了更好地对模型分割效果作出评价,在实验中使用准准确度(Accuracy, ACC)、杰卡尔德系数

(Jaccard Index, JI)、骰子系数(Dice Coefficient, DC)、敏感度(Sensitivity, Sen)和特异度(Specificity, Spe)5个性能指标。ACC代表模型预测正确的样本占整体样本数量的比例,而JI和DC则用来衡量预测样本与真实样本的重叠程度。另外,Sen表示模型成功预测出的阳性样本占有所有阳性样本的比例,Spe则代表模型成功预测出的阴性样本占有所有阴性样本的比例。5个评价指标定义如下所示:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

(18)

$$JI = \frac{TP}{FP + TP + FN}$$

(19)

$$DC = \frac{2 \times TP}{FP + 2 \times TP + FN}$$

(20)

$$Sen = \frac{TP}{TP + FN}$$

(21)

$$Spe = \frac{TN}{TN + FP}$$

(22)

其中,TP(真阳性)表示正确分割的黑色素瘤病变像素的数量,TN(真阴性)表示正确分割的背景像素的数量,FP(假阳性)表示把背景像素错误预测为黑色素瘤病变像素的数量,FN(假阴性)表示把黑色素瘤病变像素错误预测为背景像素的数量。

2.4 实验结果分析

2.4.1 权重参数λ的选择实验 在ISIC 2018和PH²两个数据集上的权重参数λ选择实验结果如表1所示。通过λ来平衡L_{BCE}和L_{SA}两个损失函数,L_{BCE}是二分类交叉熵损失函数,L_{SA}是基于边缘的形状感知损失函数。从表中数据,可以得知当λ=0.4时分割效果最好,即当MSBKA对病变区域的边缘细节特征更加关注时,损失最小。

表1 权重参数λ的选择实验
Table 1 Selection experiment of weight parameter λ

λ	ACC		JI		DC		Sen		Spe	
	ISIC 2018	PH ²	ISIC 2018	PH ²	ISIC 2018	PH ²	ISIC 2018	PH ²	ISIC 2018	PH ²
1.0	0.958	0.956	0.818	0.819	0.855	0.856	0.897	0.896	0.968	0.967
0.9	0.959	0.956	0.819	0.819	0.858	0.859	0.898	0.897	0.969	0.969
0.8	0.959	0.957	0.821	0.820	0.859	0.861	0.899	0.898	0.970	0.969
0.7	0.960	0.959	0.823	0.820	0.861	0.863	0.902	0.901	0.971	0.970
0.6	0.961	0.959	0.825	0.822	0.863	0.864	0.903	0.902	0.972	0.971
0.5	0.961	0.960	0.826	0.822	0.865	0.865	0.904	0.902	0.973	0.972
0.4	0.963	0.961	0.827	0.823	0.867	0.866	0.905	0.903	0.975	0.973
0.3	0.962	0.960	0.819	0.822	0.860	0.865	0.903	0.902	0.974	0.971
0.2	0.959	0.959	0.817	0.821	0.859	0.861	0.901	0.902	0.973	0.970
0.1	0.957	0.958	0.815	0.819	0.856	0.860	0.899	0.898	0.972	0.968

2.4.2 消融实验 为了验证CCTB、EKSM、ERM和MEAM的有效性,在ISIC 2018和PH²两个数据集上进行了大量的实验,在ISIC 2018数据集上的消融实验结果如表2所示。Swin-Unet+CCTB表示在Swin-Unet解码器部分分别添加3个CCTB×2,CCTB可以获得更充分的全局信息空间细节并且建立长距离像素依赖性,了解病变的整体结构并确定其与背景的关系,从而更准确地定位病变区域。Swin-Unet+EKSM表示在编码器与解码器之间添加1个EKSM,EKSM能够正确选择边缘关键点,引导编码器更准确地定位黑色素瘤的位置。Swin-Unet+ERM表示在编码器第三层加入1个ERM,ERM可以很好地捕获病变区域的边缘特征,以达到细化黑色素瘤边缘的目的。

的。Swin-Unet+MEAM表示在跳连接中添加3个MEAM,MEAM可以提取多尺度特征信息,在不同尺度上将编码阶段的通道、像素和空间特征聚合在一起再传递到解码器中,能够很好地适应黑色素瘤病变区域形状的变化。Swin-Unet+CCTB+EKSM表示添加3个CCTB×2和1个EKSM。Swin-Unet+CCTB+ERM表示添加3个CCTB×2和1个ERM。Swin-Unet+CCTB+MEAM表示添加3个CCTB×2和3个MEAM。Swin-Unet+EKSM+ERM表示添加3个EKSM、1个ERM。Swin-Unet+EKSM+MEAM表示添加3个EKSM,跳跃连接中嵌入3个MEAM。Swin-Unet+ERM+MEAM表示添加1个ERM,跳跃连接中嵌入3个MEAM。Swin-Unet+CCTB+EKSM+

表2 在 ISIC 2018数据集上的消融实验
Table 2 Ablation study on ISIC 2018 dataset

方法	ACC	JI	DC	Sen	Spe
Swin-Unet	0.952	0.815	0.857	0.897	0.961
Swin-Unet+CCTB	0.957	0.819	0.859	0.899	0.965
Swin-Unet+EKSM	0.955	0.819	0.858	0.898	0.962
Swin-Unet+ERM	0.954	0.820	0.859	0.899	0.95
Swin-Unet+MEAM	0.954	0.820	0.860	0.899	0.963
Swin-Unet+CCTB+EKSM	0.958	0.822	0.861	0.900	0.968
Swin-Unet+CCTB+ERM	0.959	0.823	0.862	0.901	0.968
Swin-Unet+CCTB+MEAM	0.958	0.822	0.862	0.902	0.970
Swin-Unet+EKSM+ERM	0.958	0.823	0.861	0.901	0.970
Swin-Unet+EKSM+MEAM	0.957	0.822	0.862	0.902	0.968
Swin-Unet+ERM+MEAM	0.956	0.822	0.861	0.900	0.968
Swin-Unet+EKSM+ERM+MEAM	0.960	0.824	0.863	0.903	0.972
Swin-Unet+CCTB+EKSM+ERM	0.961	0.826	0.864	0.903	0.971
Swin-Unet+CCTB+EKSM+MEAM	0.961	0.825	0.863	0.904	0.972
Swin-Unet+CCTB+ERM+MEAM	0.962	0.825	0.863	0.903	0.974
Swin-Unet+CCTB+EKSM+ERM+MEAM	0.963	0.827	0.867	0.905	0.975

ERM 表示增加 3 个 CCTB×2、3 个 EKSM 和 1 个 ERM。Swin-Unet+CCTB+EKSM+MEAM表示增加 3 个 CCTB×2、3 个 EKSM, 跳跃连接中嵌入 3 个 MEAM。Swin-Unet+CCTB+ERM+MEAM 表示增加 3 个 CCTB×2、1 个 ERM, 跳跃连接中嵌入 3 个 MEAM。Swin-Unet+EKSM+ERM+MEAM表示添加 3 个 EKSM、1 个 ERM、3 个 MEAM。Swin-Unet+CCTB+EKSM+ERM+MEAM表示添加 3 个 CCTB×2、3 个 EKSM、1 个 ERM,跳跃连接中嵌入 3 个 MEAM。对添加 CCTB、EKSM、ERM 和 MEAM之前和之后的结果进行对比。增加 CCTB、EKSM、ERM 和 MEAM 后方法具有更好的输出结果,与 Swin-Unet 相比,在 ISIC 2018 数据集上 ACC、JI、DC、Sen 和 Spe 5 个性能指标分别提高了 1.1%、1.2%、1%、0.8% 和 1.4%,这表明 CCTB、EKSM、ERM 和 MEAM 的组合提高了分割性能。

在 PH²数据集上进行的消融实验结果如表 3 所示,消融实验所采取的模块组合与 ISIC 2018 数据集上的设置保持一致。通过在基线网络 Swin-Unet 上 CCTB、EKSM、ERM 和 MEAM 模块以及模块的组合,分割性能得到了提升。与 Swin-Unet 相比,在 PH²

数据集上, Swin-Unet+CCTB+EKSM+ERM+MEAM 带来了 0.7% (ACC)、0.9% (JI)、1.1% (DC)、0.9% (Sen)和 0.8%(Spe)的性能提升。在 PH²数据集上的消融实验结果表明了 CCTB、EKSM、ERM 和 MEAM 模块的组合对于提高方法性能起到了积极的作用。

2.4.3 现有方法与 MSBKA 方法的对比 在 ISIC 2018 和 PH² 两个数据集上,将提出的 MSBKA 与 AttU-Net^[9]、UNet3+^[11]、CANet^[13]、SSFormer^[29]、TransUNet^[30]、Swin-Unet^[19]进行了分割性能比较。其中,AttU-Net、UNet3+为常用的医学图像分割网络,CANet 是专门针对于黑色素瘤图像的分割网络,SSFormer 为有代表性 Transformer 的分割网络,TransUNet、Swin-Unet 为卷积神经网络和 Transformer 相结合的典型分割网络。

表 4 列出了 MSBKA 方法在 ISIC 2018 数据集上与其他 6 种先进方法的比较结果。从表中数据可以看出,提出的 MSBKA 方法的 ACC、JI、DC、Sen 和 Spe 指标均是最高的,分别达到了 96.3%、82.7%、86.7%、90.5%和 97.5%,比常用医学分割网络 AttU-Net 提高了 2.5%、8.3%、4.4%、2% 和 1.5%,比基线网络 Swin-Unet 提高了 1.1%、1.2%、1%、0.8% 和 1.4%。原因在于

表 3 在 PH²数据集上的消融实验
Table 3 Ablation study on PH² dataset

方法	ACC	JI	DC	Sen	Spe
Swin-Unet	0.954	0.814	0.855	0.894	0.965
Swin-Unet+CCTB	0.956	0.817	0.858	0.895	0.968
Swin-Unet+EKSM	0.954	0.818	0.857	0.896	0.968
Swin-Unet+ERM	0.954	0.818	0.856	0.895	0.967
Swin-Unet+MEAM	0.953	0.818	0.857	0.896	0.967
Swin-Unet+CCTB+EKSM	0.957	0.819	0.859	0.897	0.968
Swin-Unet+CCTB+ERM	0.958	0.818	0.860	0.899	0.968
Swin-Unet+CCTB+MEAM	0.957	0.820	0.862	0.898	0.969
Swin-Unet+EKSM+ERM	0.958	0.819	0.862	0.897	0.969
Swin-Unet+EKSM+MEAM	0.959	0.818	0.859	0.899	0.968
Swin-Unet+ERM+MEAM	0.957	0.820	0.860	0.898	0.969
Swin-Unet+EKSM+ERM+MEAM	0.961	0.820	0.862	0.902	0.970
Swin-Unet+CCTB+EKSM+ERM	0.960	0.821	0.863	0.901	0.970
Swin-Unet+CCTB+EKSM+MEAM	0.959	0.822	0.864	0.902	0.972
Swin-Unet+CCTB+ERM+MEAM	0.960	0.822	0.865	0.900	0.971
Swin-Unet+CCTB+EKSM+ERM+MEAM	0.961	0.823	0.866	0.903	0.973

MSBKA 方法的优势,其 CCTB 模块能够更全面地捕捉全局信息空间的细微之处,并建立长距离的像素依赖关系,从而更深入地理解病变的整体构造及其与背景之间的关联。EKSM 则精准地挑选边缘关键点,引导编码器精确捕捉黑色素瘤的边缘。同时,ERM 表现出色,能够精细地描绘病变区域的边缘特征,进一步锐化黑色素瘤的边缘轮廓。此外,MEAM 通过提取多尺度特征信息,能够在不同尺度上有效整合编码阶段的通道、像素和空间特征,并将这些特征传递给解码器。这使得 MEAM 具有出色的灵活性,能够适应黑色素瘤病变区域形状的变化。与其他 6 种方法相比,本章提出的 MSBKA 方法在分割性能上表现出了显著的优势,获得了更高的分割性能得分,从而验证了其有效性。

为了更直观地比较不同网络的分割性能,图 10 给出了 MSBKA 与 AttU-Net、UNet3+、CANet、SSFormer、TransUNet、Swin-Unet 6 种模型在 ISIC 2018 数据集上的分割指标得分柱状图。可以很明显地看出本文提出的 MSBKA 方法在 5 个性能指标上的得分都是最高的。整体来看,AttU-Net、UNet3+、CANet、SSFormer、TransUNet、Swin-Unet、MSBKA 的分割性能得分逐渐提高,分割效果越来越好,但是在 SSFormer 有所下降,SSFormer 局部获取信息能力不足,会导致分割结果出现漏分割、错分割的现象。

表 4 在 ISIC 2018 数据集上的分割性能比较
Table 4 Comparison of segmentation performance on ISIC 2018 dataset

方法	ACC	JI	DC	Sen	Spe
AttU-Net ^[9]	0.938	0.744	0.823	0.885	0.960
UNet3+ ^[11]	0.933	0.745	0.836	0.891	0.961
CANet ^[13]	0.942	0.765	0.840	0.894	0.965
SSFormer ^[29]	0.931	0.764	0.832	0.882	0.945
TransUNet ^[30]	0.946	0.798	0.843	0.890	0.957
Swin-Unet ^[19]	0.952	0.815	0.857	0.897	0.961
MSBKA	0.963	0.827	0.867	0.905	0.975

AttU-Net、UNet3+、CANet 3 种 UNet 的变体模型分割性能低于 TransUNet、Swin-Unet、MSBKA,后 3 个模型都将 CNN 与 Transformer 相结合,利用 CNN 进行特征提取,利用 Transformer 进行全局关系建模或上下文理解。

图 11 给出了 ISIC 2018 数据集的一些可视化分割效果图。图中展示了不同网络在面对形状不规则的病变区域的分割结果,第 3、4、5 列分别是 AttU-Net、UNet3+、CANet 的分割结果图,从这几列可以观察到,这 3 种 UNet 网络的变体都存在过度分割的问题,分割结果与标签对比来看,原本不属于病变区域的

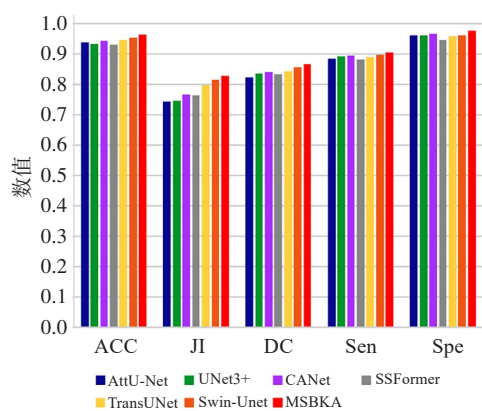


图 10 不同方法在 ISIC 2018 数据集的分割性能比较
Figure 10 Comparison of segmentation performance of different methods on ISIC 2018 dataset

像素也被划分到病变区域。第6列是SSFormer的分割结果,可以看出它的分割效果并不是很好,如第2行第6列病变区域缺失。第7、8两列是TransUNet、Swin-Unet的分割结果,分割效果较好,病变区域较为准确,但边缘不够明显。最后一列是本文方法MSBKA的分割效果图,可以看到它在边缘处的分割表现优于其他网络,尤其是图中红框标注边缘处的表现,有很精细的边缘。

表5列出了MSBKA方法在PH²数据集上与其他6种先进方法的比较结果。从表中数据可以看出,提出的MSBKA方法的ACC、JI、DC、Sen和Spe的指标上均是最高,分别达到了96.1%、82.3%、86.6%、

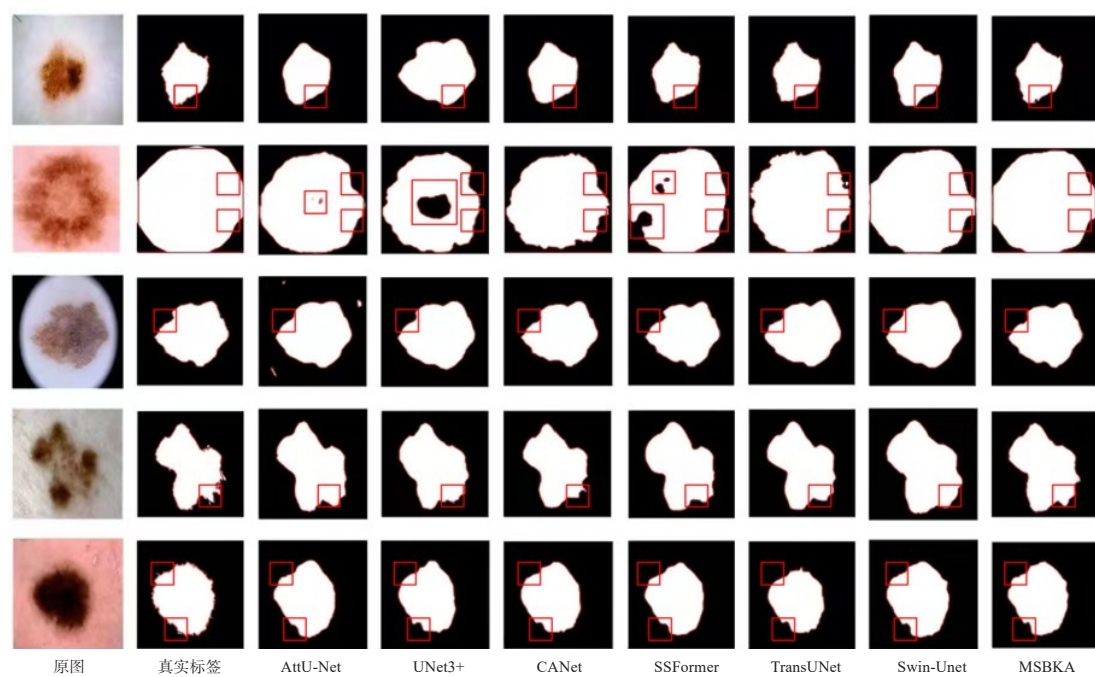


图 11 不同方法在 ISIC 2018 数据集的可视化结果
Figure 11 Visualization results of different methods on ISIC 2018 dataset

90.3%和97.3%,比常用医学分割网络 AttU-Net 提高了3%、4.2%、4.1%、1.6%和1%,比基线网络 Swin-Unet 提高了1%、0.9%、1.1%、0.9%和0.8%。这是因为MSBKA方法通过CCTB从全局和局部两个角度对边缘特征进行提取,设计了EKSM正确选择边缘关键点,准确定位边缘位置,提出ERM使得边缘分割效果更加精细化,再经过MEAM的多尺度边缘信息提取,使得边缘细节更加丰富。

为了更直观地比较不同网络在图像分割性能上的优劣,图12展示了MSBKA与6种模型在PH²数据集上的分割指标得分柱状图。从图中可以清晰地看

表 5 在 PH²数据集上的分割性能比较

Table 5 Comparison of segmentation performance on PH² dataset

方法	ACC	JI	DC	Sen	Spe
AttU-Net ^[9]	0.931	0.781	0.825	0.887	0.963
UNet3+ ^[11]	0.939	0.783	0.828	0.889	0.964
CANet ^[13]	0.945	0.792	0.843	0.891	0.967
SSFormer ^[29]	0.937	0.778	0.835	0.890	0.957
TransUNet ^[30]	0.948	0.799	0.846	0.889	0.962
Swin-Unet ^[19]	0.951	0.814	0.855	0.894	0.965
MSBKA	0.961	0.823	0.866	0.903	0.973

出,本文提出的MSBKA方法在5个性能指标上的得分均为最高。总体趋势上,这7种模型的分割性能得分逐步提高,表明分割效果在持续优化。然而,SSFormer的得分出现了下降的情况,这可能是由于SSFormer在局部信息获取方面的能力不足,进而可能导致分割结果中出现漏分割、错分割的现象。在对比各模型时,可以发现AttU-Net、UNet3+、CANet这3种基于UNet的变体模型在分割性能上略逊于TransUNet、Swin-Unet和MSBKA。尽管这些模型在某些方面表现出色,但它们分割性能仍不及本文提出的MSBKA方法。

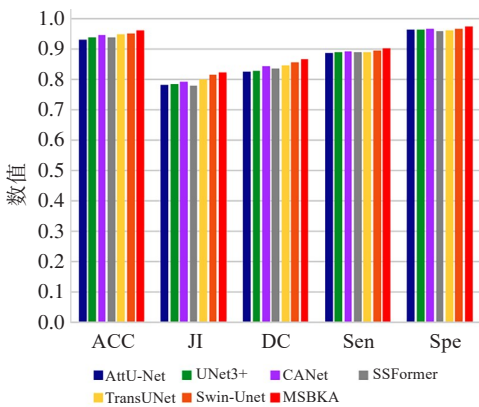


图12 不同方法在PH²数据集的分割性能比较
Figure 12 Comparison of segmentation performance of different methods on PH² dataset

图13给出了PH²数据集的一些可视化分割效果图。尽管所有模型在黑色素瘤病变区域上实现大体的分割,但在准确性方面却存在显著差异。AttU-Net、UNet3+、CANet模型对病变区域的漏分割和错分割现象较为显著。SSFormer模型在处理病变边缘不明显的图像时,分割效果也较差,易导致漏分割。相对而言,TransUNet和Swin-Unet的分割结果有所改善,但在边缘轮廓细节上仍显得不够清晰和准确。本文提出的MSBKA模型在分割黑色素瘤病变区域时,很少出现漏分割和错分割的情况,同时能够更好地保留病变区域轮廓边缘的细节信息。通过与AttU-Net、UNet3+、CANet、SSFormer、TransUNet、Swin-Unet等6种方法比较,尤其是在图中红框标出的位置,可以发现MSBKA模型在边缘处能取得较好的分割效果,说明模型识别病变边缘方面的能力得到了显著提升。

3 结 语

本文提出了一种U型网络MSBKA,解决了黑色素瘤图像中大小形状不规则、边缘模糊的问题。MEAM利用并行的卷积和3种不同的注意力机制,在不降低特征图像分辨率的情况下捕获多尺度上下文信息。设计了CCTB,通过卷积神经网络和Transformer的组合,从局部和全局两个层面获取黑色素瘤图像的边缘特征信息,提升网络对不同形状的病变区域的表示能力。此外,设计了EKSM和ERM,

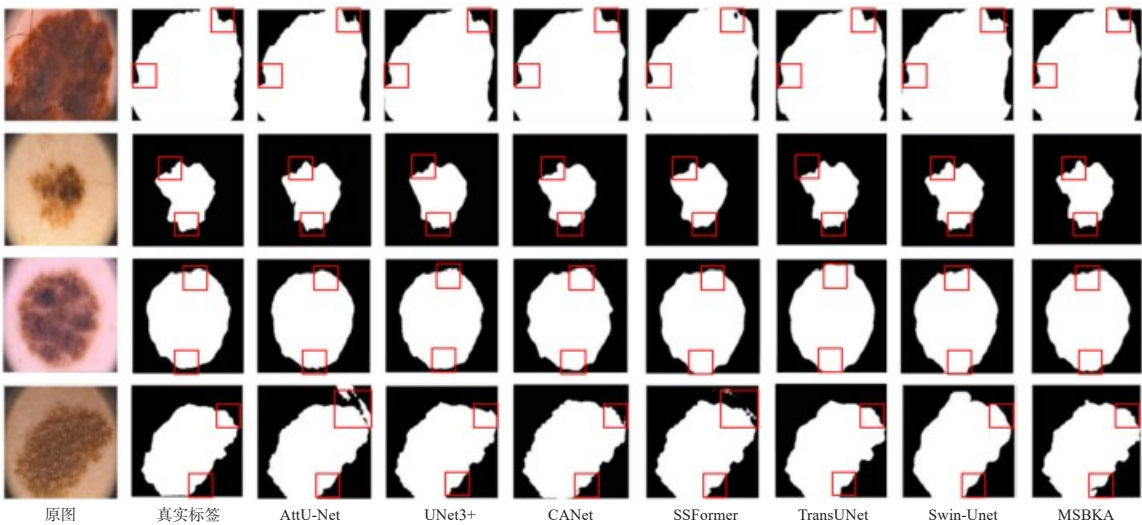


图13 不同方法在PH²数据集的可视化结果
Figure 13 Visualization results of different methods on PH² dataset

提高了网络处理不规则大小形状和模糊边缘的能力,使得分割图的边缘更加精细。在ISIC 2018和PH²两个数据集上的实验和比较研究证明了MSBKA

方法的有效性。在今后的研究中,将进一步研究如何减少干扰因素来提高黑色素瘤的分割效果。

【参考文献】

- [1] Capdehourat G, Corez A, Bazzano A, et al. Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions[J]. *Pattern Recognit Lett*, 2011, 32(16): 2187-2196.
- [2] Nadratowski A, Shoots-Reinhard B, Shafer A, et al. Evidence-based communication to increase melanoma knowledge and skin checks[J]. *JID Innov*, 2024, 4(2): 100253.
- [3] Esteve A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks[J]. *Nature*, 2017, 542(7639): 115-118.
- [4] Haenssle HA, Fink C, Schneiderbauer R, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists[J]. *Ann Oncol*, 2018, 29(8): 1836-1842.
- [5] Yu LQ, Chen H, Dou Q, et al. Automated melanoma recognition in dermoscopy images via very deep residual networks[J]. *IEEE Trans Med Imaging*, 2017, 36(4): 994-1004.
- [6] Chakkaravarthy AP, Chandrasekar A. An automatic segmentation of skin lesion from dermoscopy images using watershed segmentation [C]//2018 International Conference on Recent Trends in Electrical, Control and Communication (RTECC). Piscataway, NJ, USA: IEEE, 2018: 15-18.
- [7] Abbas AA, Guo X, Tan WH, et al. Combined spline and B-spline for an improved automatic skin lesion segmentation in dermoscopic images using optimal color channel[J]. *J Med Syst*, 2014, 38(8): 80.
- [8] Garg S, Jindal B. Skin lesion segmentation using k-mean and optimized fire fly algorithm[J]. *Multimed Tools Appl*, 2021, 80(5): 7397-7410.
- [9] Wang SH, Li L, Zhuang XH. AttU-Net: attention U-net for brain tumor segmentation[C]//Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Cham: Springer International Publishing, 2022: 302-311.
- [10] Xiao X, Lian S, Luo ZM, et al. Weighted res-UNet for high-quality retina vessel segmentation[C]//2018 9th International Conference on Information Technology in Medicine and Education (ITME). Piscataway, NJ, USA: IEEE, 2018: 327-331.
- [11] Huang HM, Lin LF, Tong RF, et al. UNet 3+: a full-scale connected UNet for medical image segmentation[C]//ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ, USA: IEEE, 2020: 1055-1059.
- [12] Sarker MMK, Rashwan HA, Akram F, et al. SLSDeep: skin lesion segmentation based on dilated residual and pyramid pooling networks [C]//Medical Image Computing and Computer Assisted Intervention -MICCAI 2018. Cham: Springer International Publishing, 2018: 21-29.
- [13] Hanselmann M, Strauss T, Dormann K, et al. CANet: an unsupervised intrusion detection system for high dimensional CAN bus data[J]. *IEEE Access*, 2020, 8: 58194-58205.
- [14] Pang YW, Zhao XQ, Zhang LH, et al. Multi-scale interactive network for salient object detection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 9410-9419.
- [15] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2017: 6000-6010.
- [16] Wang JC, Wei L, Wang LS, et al. Boundary-aware transformers for skin lesion segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer International Publishing, 2021: 206-216.
- [17] Wu HS, Chen SH, Chen GL, et al. FAT-net: feature adaptive transformers for automated skin lesion segmentation[J]. *Med Image Anal*, 2022, 76: 102327.
- [18] 傅励瑶, 尹梦晓, 杨锋. 基于Transformer的U型医学图像分割网络综述[J]. *计算机应用*, 2023, 43(5): 1584-1595.
Fu LY, Yin MX, Yang F. Transformer based U-shaped medical image segmentation network: a survey[J]. *Journal of Computer Applications*, 2023, 43(5): 1584-1595.
- [19] Cao H, Wang YY, Chen J, et al. Swin-Unet: Unet-like pure transformer for medical image segmentation[C]//Computer Vision-ECCV 2022 Workshops. Cham: Springer Nature Switzerland, 2023: 205-218.
- [20] 魏坤, 沈记全, 赵艳梅. MAUNet: 用于皮肤病变分割的轻量级模型[J]. *中国医学物理学杂志*, 2023, 40(5): 555-561.
Wei K, Shen JQ, Zhao YM. MAUNet: a lightweight model for skin lesion segmentation[J]. *Chinese Journal of Medical Physics*, 2023, 40(5): 555-561.
- [21] Zhang R, Zhu F, Liu JY, et al. Depth-wise separable convolutions and multi-level pooling for an efficient spatial CNN-based steganalysis[J]. *IEEE Trans Inf Forensics Secur*, 2020, 15: 1138-1150.
- [22] Han DC, Ye TZ, Han YZ, et al. Agent attention: on the integration of SoftMax and linear attention[EB/OL]. (2023-12-14)[2024-04-02]. <https://arxiv.org/abs/2312.08874>.
- [23] Kirillov A, Wu YX, He KM, et al. PointRend: image segmentation as rendering[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 9796-9805.
- [24] Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 936-944.
- [25] Wang CY, Liao HY, Wu YH, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2020: 1571-1580.
- [26] Zhao HY, Kong XT, He JW, et al. Efficient image super-resolution using pixel attention[C]//Computer Vision-ECCV 2020 Workshops. Cham: Springer International Publishing, 2020: 56-72.
- [27] 马伟, 龚超凡, 徐士彪, 等. 自顶向下引导式特征融合的物体轮廓检测网络[J]. *计算机辅助设计与图形学学报*, 2021, 33(6): 855-863.
Ma W, Gong CF, Xu SB, et al. Top-down guided fusion network for object contour detection[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2021, 33(6): 855-863.
- [28] 陈菁菁, 李小霞, 吕念祖. 结合通道权重更新与密集残差金字塔空间注意力的皮肤病变分割方法[J]. *中国医学物理学杂志*, 2023, 40(1): 39-46.
Chen JJ, Li XX, Lü NZ. Skin lesion segmentation method combining channel weight update and dense residual pyramid spatial attention[J]. *Chinese Journal of Medical Physics*, 2023, 40(1): 39-46.
- [29] Shi WT, Xu J, Gao P. SSformer: a lightweight transformer for semantic segmentation [C]//2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP). Piscataway, NJ, USA: IEEE, 2022: 1-5.
- [30] Chen JN, Lu YY, Yu QH, et al. TransUNet: transformers make strong encoders for medical image segmentation[EB/OL]. (2021-02-08)[2024-04-02]. <https://arxiv.org/abs/2102.04306>.

(编辑:薛泽玲)