

## 融合可变形卷积与注意力机制的糖尿病视网膜多病变分割网络

李春晓<sup>1</sup>,周亚同<sup>1</sup>,单春艳<sup>2</sup>,肖志涛<sup>3</sup>,卜云帆<sup>1</sup>

1.河北工业大学电子信息工程学院,天津 300401; 2.天津医科大学朱宪彝纪念医院肾内科,天津 300134; 3.天津工业大学生命科学学院,天津 300387

**【摘要】**针对糖尿病视网膜病变结构复杂、不同病变尺度差异大等问题,提出一种融合可变形卷积和注意力机制的视网膜多病变分割网络用于糖尿病视网膜多病变自动分割。首先,使用可变形小波编码模块替换原始卷积下采样编码器,以适应病变的不规则形状变化,提取有效特征信息;然后,在瓶颈层引入密集特征感知与聚合模块,通过聚合多个感受野进行多尺度特征的提取,增强深层语义信息;最后,为充分融合解码器输出,提升对边缘信息的识别精度,引入多尺度自适应融合模块对每层解码器输出进行加权,从而获取最准确的分割特征图。在DDR-RLS数据集上进行硬性渗出物、出血点、软性渗出物分割验证,结果发现所提出的网络与原有Unet相比,IoU系数分别提升0.026 2、0.051 8、0.046 5,Dice系数分别提升0.027 1、0.058 1、0.050 4,AUPR值分别提升0.042 3、0.069 1、0.073 4。

**【关键词】**糖尿病视网膜病变;可变形小波编码;密集特征感知与聚合;多尺度自适应融合

**【中图分类号】**R318;TP391.41

**【文献标志码】**A

**【文章编号】**1005-202X(2025)05-0596-10

### A diabetic retinopathy multi-lesion segmentation network integrating deformable convolution and attention mechanism

LI Chunxiao<sup>1</sup>, ZHOU Yatong<sup>1</sup>, SHAN Chunyan<sup>2</sup>, XIAO Zhitao<sup>3</sup>, BU Yunfan<sup>1</sup>

1. School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China; 2. Department of Nephrology, Tianjin Medical University Chu Hsien-I Memorial Hospital, Tianjin 300134, China; 3. School of Life Sciences, Tiangong University, Tianjin 300387, China

**Abstract:** In view of the complex structure of diabetic retinopathy and the large differences in the scales of different lesions, a novel network which integrates deformable convolution and attention mechanism is proposed for automatic diabetic retinopathy multi-lesion segmentation. Specifically, deformable convolution Haar wavelet transform encoder takes place of the original convolutional downsampling encoder to adapt to the irregular shape changes of lesions and extract effective feature information; a dense feature perception and aggregation module is introduced at the bottleneck layer to extract multi-scale features by aggregating multiple receptive fields, thus enhancing deep semantic information; and finally, in order to fully integrate the decoder output and improve the recognition accuracy of edge information, a multi scale adaptive fusion module is used to weight the decoder output of each layer for obtaining the most accurate segmentation feature map. The validation of hard percolation, bleeding point, and soft percolation segmentations on the DDR-RLS dataset reveals that the proposed network shows increases of 0.026 2, 0.051 8 and 0.046 5 in IoU coefficient, 0.027 1, 0.058 1 and 0.050 4 in Dice coefficient, and 0.042 3, 0.069 1 and 0.073 4 in AUPR value, as compared with the original Unet.

**Keywords:** diabetic retinopathy; deformable convolution Haar wavelet transform encoder; dense feature perception and aggregation; multi scale adaptive fusion

**【收稿日期】**2024-12-14

**【基金项目】**京津冀基础研究合作专项(J210008, 21JCZXC00170, H2021202008);天津市医学重点学科(专科)建设项目(TJYXZDXK-032A)

**【作者简介】**李春晓, 硕士, 研究方向:医学图像处理, E-mail: 1694553849@qq.com

**【通信作者】**周亚同, 教授, 研究方向:信息感知与机器学习, E-mail: zyt@hebut.edu.cn

### 前言

糖尿病视网膜病变(Diabetic Retinopathy, DR)是糖尿病患者长期高血压和高血糖导致的微血管疾病,对患者的视力带来严重威胁<sup>[1]</sup>。DR的尽早诊断与及时干预可极大降低患者致盲概率,减少社会医疗资源负担。DR包括软性渗出物、硬性渗出物、出血点、微动脉瘤、黄斑水肿、血管增生等,然而在实际

筛查眼底疾病的过程中,由于病变的面积小、颜色对比度低,识别过程耗费大量的人力。随着计算机辅助医疗的发展,利用图像处理、深度学习等相关技术自动识别相关眼底疾病<sup>[2-3]</sup>,降低不同专家由于先验知识不同带来的检测差异性,有利于对糖尿病患者进行大规模筛查。

早期研究采用传统数字图像处理方法,利用眼底图像的纹理、颜色、对比度、梯度等特征,基于聚类<sup>[4]</sup>、模板匹配<sup>[5]</sup>、阈值<sup>[6]</sup>、边缘检测<sup>[7]</sup>等方法进行眼底图像病变和血管识别,从而判断眼底视网膜病变的严重程度,但是受成像设备和拍摄角度等方面的影响,识别不同的图像需要手动提取不同的特征,不具有普遍适用性。受神经网络快速发展的影响,Sambyal等<sup>[8]</sup>使用亚像素周期混合卷积代替上采样,构建残差U型网络进行眼底病变分割。Huang等<sup>[9]</sup>提出一种结合超像素多特征提取的深度神经网络分割硬性渗出物。为克服眼底病变形状不规则、颜色对比度差异小等困难,有研究通过构建混合注意力关系变换模块<sup>[10]</sup>、尺度变换模块<sup>[11]</sup>、级联多个编码器<sup>[12]</sup>等方式,减轻复杂眼底结构引起的病变检测的模糊性。Wang等<sup>[13]</sup>构建任务感知损失来指导构建视网膜超分辨率图像、视网膜病变分割和视网膜病变分级3个任务,充分利用多任务间的关联性深入挖掘眼底特征。Yao等<sup>[14]</sup>利用硬性渗出物和小胶质细胞的语义相关性,构建双解码器交替反馈特征信息,以增强病变间的相互作用,实现精准分割。Butt等<sup>[15]</sup>利用预训练的卷积神经网络提取特征并传递给多个分类器,Ashwini等<sup>[16]</sup>基于离散小波变换多分辨率分解的方法提取特征,实现对病变严重程度分级,但未实现对病变的精确分割。

上述方法虽在病变识别与分割方面取得很大进

展,但大多针对单一病变进行分割,未考虑DR发病阶段多病变之间的相互影响,因此本研究基于Unet<sup>[17]</sup>,提出一种新的融合可变形卷积与注意力机制的视网膜多病变分割网络(Multi-Lesion Deformable Convolution and Attention Mechanism Unet, MLDA-Unet),根据发病不同阶段多病变产生的关联性,提升对多病变的分割效果。首先,构建可变形小波编码(Deformable Convolution Haar Wavelet Transform Encoder, DCHT)模块提取视网膜图像中的多形变和多尺度的病灶特征,减少小尺度病变在下采样过程中的丢失;其次,在Unet的瓶颈层添加密集特征感知与聚合(Dense Feature Perception and Aggregation module, DFPA)模块,充分扩展图像深层次的语义信息,丰富全局特征;最后,改善Unet网络的编码器输出,使用多尺度自适应融合(Multi Scale Adaptive Fusion, MSFA)模块将不同尺度的编码器输出进行有效融合,丰富预测概率图,以实现多病变准确分割。

## 1 MLDA-Unet模型构建

MLDA-Unet主要包括3个部分:编码部分、上下文信息扩展部分和解码部分(图1)。编码部分由4组级联的DCHT模块组成,使用Haar小波变换将空间信息转换到通道维度,通过 $3 \times 3$ 可变形卷积逐层特征提取,获取深层次语义信息;上下文信息扩展部分插入不同感受野大小的空洞卷积,扩展特征提取范围,丰富全局上下文信息,并结合SE(Squeeze-and-Excitation)注意力<sup>[18]</sup>机制,整合瓶颈层的通道信息;解码部分则使用注意力门控机制<sup>[19]</sup>融合同级解码模块和底层特征信息,MSAF模块对深度解码后的特征图进行自适应融合,充分利用各级解码特征进行视网膜多病变分割。

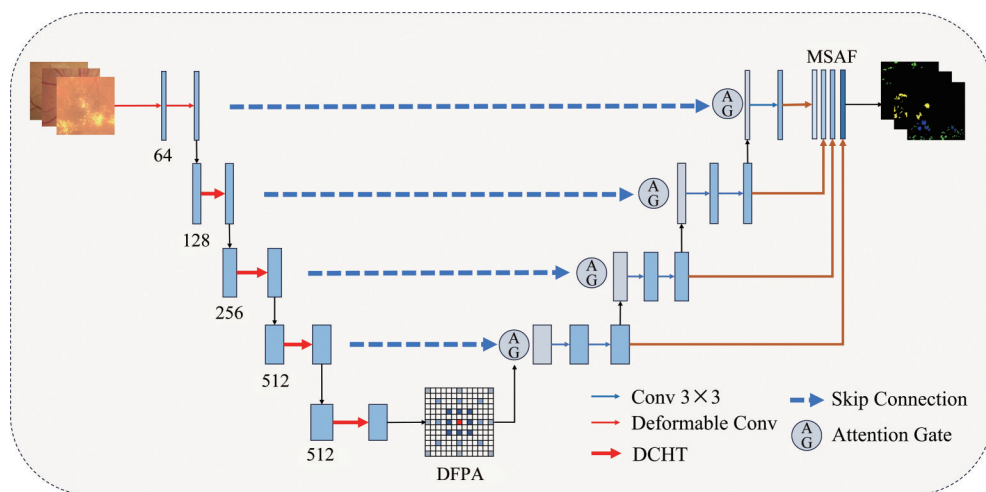


图1 多病变分割网络模型整体结构图

Figure 1 Overall structure of multi lesion segmentation network model

### 1.1 DCHT 模块

由于视网膜病变尺寸小,在Unet网络编码连续下采样池化过程中极易造成小尺度病变丢失,采用Haar小波下采样<sup>[20]</sup>可以在降低分辨率的同时减少语义信息的损失。利用Haar小波处理二维图像数据时,首先,沿水平方向使用一维离散Haar小波变换分解信号,计算相邻像素对的平均值和差值,平均值属于低频分量,包含病变的轮廓信息,而差值属于高频分量,包含病变的细节信息。然后,再沿垂直方向进行二次分解,全面地获取输入图像的低频分量以及水平方向、垂直方向和对角线方向的高频细节信息。在上述信号分解过程,特征图的尺寸压缩到原特征图的1/4,降低了图像的分辨率,拼接后的分量通道数

为原特征图的4倍。最后,通过1×1卷积调整通道数使其与原特征图一致,从而有效地将图像空间维度的信息转换到通道维度,使模型在下采样降低特征图分辨率的同时减少小尺度细节特征的损失。图2展示了利用不同下采样方式获取到的特征图,其中第一张图、第二张图包含的病变为硬渗,第三张图包含的病变为软渗和出血点。通过对比可以发现Haar小波下采样对边缘清晰的明黄色硬渗以及边缘模糊的软渗定位更准确,更有利于保留渗出物的边缘信息,但是无法准确定位大面积暗红色出血,可能是由于Haar小波变换对细小病变高频特征敏感,而对大面积颜色变化的低频特征不敏感。

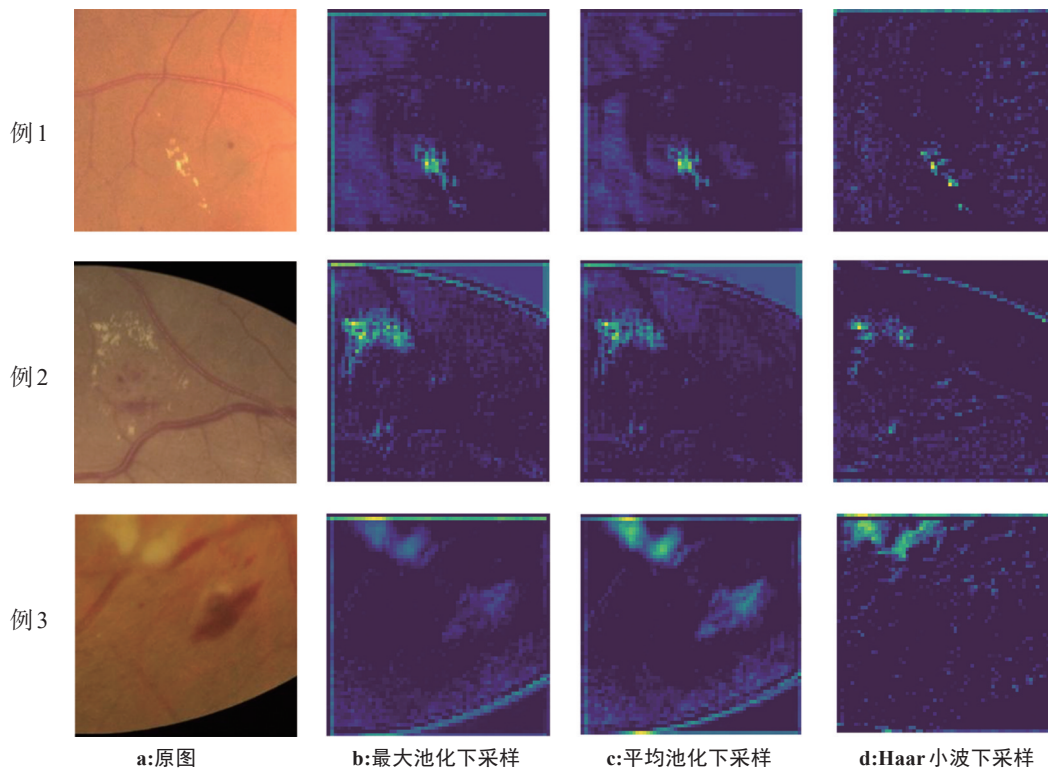


图2 多种下采样方式对比图

Figure 2 Comparison of various downsampling methods

在视网膜眼底图像中不同的位置具有不同尺寸、不同形变大小的病变,标准卷积中卷积核权重固定,难以关注到不同病变之间的形状差异性,而可变形卷积<sup>[21]</sup>在采样时能根据物体的形状进行采样,能更好地适应目标病变多种形状和尺度的变化。因此本研究重构Unet编码器结构,首先使用Haar小波变换对特征图进行下采样,扩大感受野尺寸,然后通过两个可变形卷积模块进行特征提取,并在双层卷积块中间添加残差跳跃连接,防止网络梯度消失,加速网络更快收敛,提出如图3所示的DCHT模块。

### 1.2 DFPA 模块

较大感受野可以获取全局信息,有利于大面积硬性渗出物病变特征的提取,而小感受野更能关注局部信息,适合分割出血点这种小面积病变。整合不同尺度感受野可以获取多病变信息,感知高级语义信息中不同病变之间的关联性,进行多尺度目标的分割。

为避免卷积、池化的方式在增大感受野的同时丧失局部信息,空洞卷积通过在标准卷积核之间填充不同的数量的空洞,感知更大范围特征信息。卷积核大小为 $k$ 的标准卷积,设其空洞率为 $d$ ,则表示在

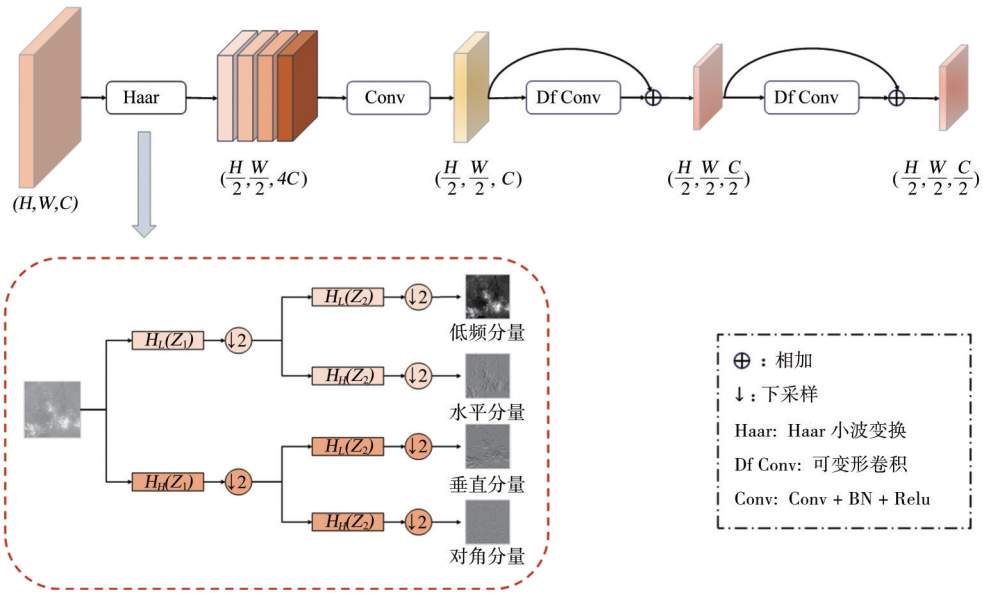


图3 DCHT 模块示意图

Figure 3 Schematic diagram of DCHT module

其中均匀填充  $d - 1$  个空洞, 则其等效空洞卷积的感受野范围  $k'$  为:

$$k' = k + (k - 1) \times (d - 1) \tag{1}$$

在  $3 \times 3$  的卷积核中分别引入 0、2、4 个空洞, 其等效感受野大小分别为 3、7、11。通过调整空洞率和卷积核的组合可在不增加额外计算量的同时获取多感受信息, 以充分获取局部与全局特征, 使卷积核同时关注大面积病变与局部小尺度病变, 然后对不同感受野的特征图进行逐像素相加, 即可获得多尺度空间信息。

在 Unet 的编码器各级特征图中, 瓶颈层的特征

图具有最小的分辨率但具有最大的通道数, 通道数的增多造成部分通道特征冗余, 不同的通道对于分割图的影响权重是不同的, 多尺度特征图丰富了空间信息却忽略了不同通道之间的语义差异性, 为进一步整合通道信息, 增加 SE 注意力模块<sup>[18]</sup>, 该模块通过对每个通道的特征图进行全局池化, 再通过两层全连接层进行自适应通道权重调整, 这样既能关注到多感受野下的空间信息, 又能通过整合通道权重关注分布在通道上的重要语义信息。因此本研组合多感受野模块和 SE 注意力模块, 构建 DFPA 模块以输送给解码器更丰富的上下文信息(图 4)。

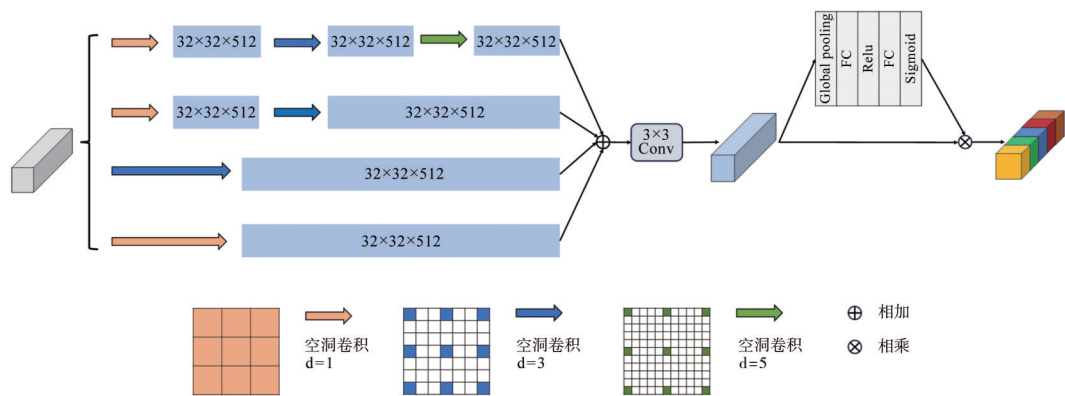


图4 DFPA 模块示意图

Figure 4 Schematic diagram of DFPA module

### 1.3 MSAF 模块

对于 DR 来说, 不同种类的病变具有不同的尺度, 即使相同种类的病变尺度大小也不统一, 为充分捕获病变的多尺度特征, 自适应调整多尺度输出图

对最终预测图的影响权重, 构建 MSAF 模块, 在融合各尺度特征图的同时更好地关注到病变的边缘信息, 提升分割准确率(图 5)。

首先, 将不同尺度的特征图使用  $1 \times 1$  卷积转换通

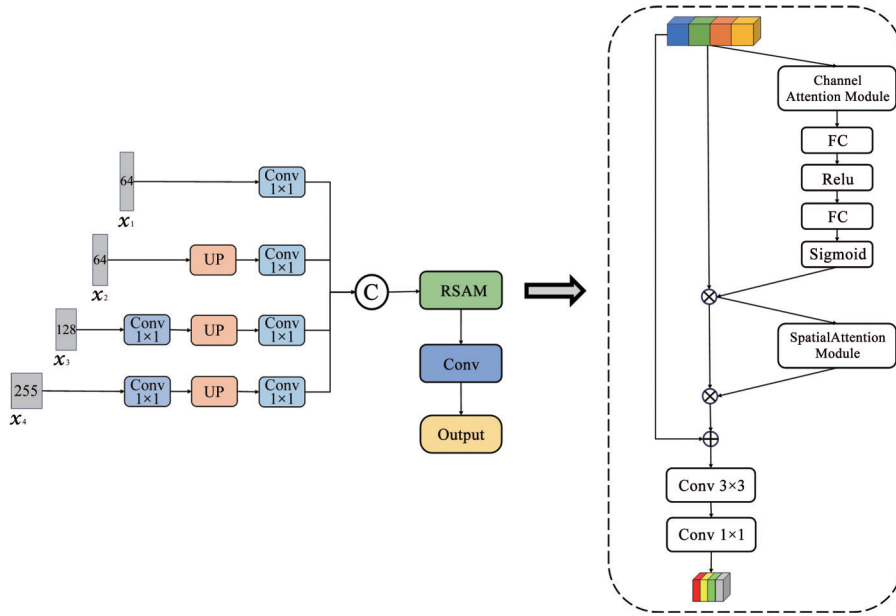


图 5 MSAF 模块示意图

Figure 5 Schematic diagram of MSAF module

道维度,均调整为最后一级解码器输出特征图的通道数 64;然后,使用双线性插值对特征图进行上采样,使其高度  $H$  和宽度  $W$  均为 256,由于最终进行背景及 3 种病变的分割,因此将每级编码器特征图均调整为 4 通道,进行拼接操作后,送入残差尺度注意力模块,同时增强通道和空间的特征表达能力,使得网络对于小尺度病变边缘特征捕获更全面。残差尺度注意力通过在空间注意力机制和通道注意力机制之间添加全连接层及残差连接,动态调整特征图的尺度,其工作流程如下:

经上采样拼接后的特征图  $X$  首先经过通道注意力模块,然后经全连接层尺度压缩后的系数与原始特征图相乘得到  $X_1$ :

$$M_c(X) = \sigma \left( W_1 \delta \left( W_2 \left( \text{Maxpool}(X) \right) \right) + W_1 \delta \left( W_2 \left( \text{Avgpool}(X) \right) \right) \right) \quad (2)$$

$$X_1 = X \otimes \sigma \left( W_1 \delta \left( W_2 \left( M_c(X) \right) \right) \right) \quad (3)$$

其中,  $W_1 \in R^{c \times c}$ 、 $W_2 \in R^{c \times \frac{c}{7}}$  分别为两个全连接层,  $\delta$  代表 ReLU 激活函数,  $\sigma$  为 Sigmoid 激活函数, Maxpool 为空间全局最大池化, Avgpool 为空间全局平均池化,  $\otimes$  表示逐元素乘法。再经空间注意力机制筛选空间维度感兴趣的信息  $X_2$ :

$$M_s(X_1) = \sigma \left( \text{Conv} \left( \text{cat} \left( \text{Maxpool}(X_1), \text{Avgpool}(X_1) \right) \right) \right) \quad (4)$$

$$X_2 = X_1 \otimes M_s(X_1) \quad (5)$$

其中, Conv 代表  $7 \times 7$  卷积。经过上述操作后,再将其与原始特征图  $X$  相加整合,获得动态多尺度调整后的特征图  $X_3$ :

$$X_3 = \text{Conv}_1 \left( \text{Conv}_2 \left( X + X_2 \right) \right) \quad (6)$$

其中,  $\text{Conv}_1$  代表  $1 \times 1$  卷积,  $\text{Conv}_2$  代表  $3 \times 3$  卷积。

## 2 视网膜多病变分割实验设置

### 2.1 数据集预处理

本研究选用公开的 DDR 数据集<sup>[22]</sup>以及医院的私有 RLS 数据集构成混合数据集 DDR-RLS 进行模型训练,已通过医院伦理委员会批准(伦理批件号: ZXYJNYYKMEC2024-4)。DDR 数据集包含 757 张彩色眼底照片,图像分辨率从  $1\ 380 \times 1\ 382$  到  $2\ 736 \times 1\ 824$ ;私有数据集 RLS 使用 Kowa Nonmyd 7 眼底照相机采集,共包含 92 张含有硬性渗出物、软性渗出物、出血点 3 种病变(出血点、软渗、硬渗出现在 DR 早期,临床关注的较高,及时发现可以尽早治疗,而微动脉瘤等小目标标注成本高,新生血管多出现于晚期且数据稀缺,未来将通过搜集补充更多数据集进行进一步研究)的像素级标注的眼底视网膜图像,图像分辨率为  $2\ 992 \times 2\ 000$ 。由于 RLS 数据集规模较小,采用限制对比度自适应直方图均衡化、色彩增强、对比度增强、水平翻转、垂直翻转、旋转  $45^\circ$ 、旋转  $90^\circ$  图像增强方式将训练集进行 7 倍扩充,扩充后的 DDR-RLS 数据集具体划分情况见表 1。

DDR 数据集以及 RLS 数据集的分辨率不统一、眼底视网膜图像周围包含大量黑色边框,且病变像素仅占整张眼底图像的 2%,为充分挖掘眼底图像特征,防止训练过拟合,对数据集进行以下预处理操作:(1)裁除冗余黑边。对图像进行高斯去噪、形态学腐蚀再扩展,然后应用 Otsu 阈值算法分离背景以

表1 DDR-RLS数据集划分情况  
Table 1 Partition of DDR-RLS dataset

数据集	训练集	验证集	测试集
DDR	383	225	149
RLS	47	18	27
RLS扩充	329	126	27
DDR-RLS	712	351	176

及前景信息,仅保留眼底图像及标签中的眼球区域。(2)调整图像大小。将裁剪后的眼底图像的高度进行填充,使高宽比为1,然后调整图像的分辨率为 $1024 \times 1024$ 。(3)切分图像块。训练集和验证集以128为裁剪步长,测试集以256为裁剪步长,将眼底图像及标签裁剪为 $256 \times 256$ 大小的图像块,最后保留有病变的训练集和验证集图像块21785个,测试集图像块911个(图6)。

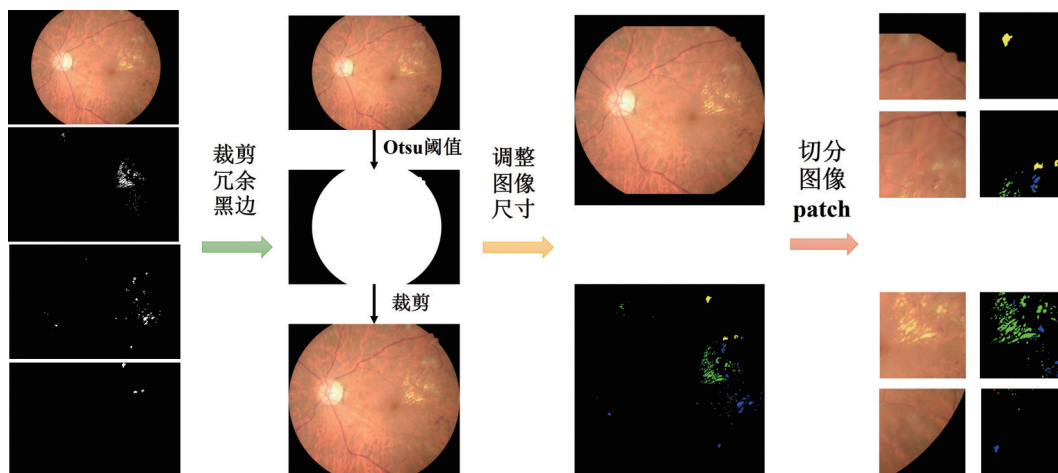


图6 眼底图像及标签预处理流程图

Figure 6 Flowchart of fundus image and label preprocessing

## 2.2 损失函数

本研究使用Dice Loss和加权交叉熵损失函数作为混合损失函数,以克服多病变分割过程中背景与病变像素点分布不均衡的问题,损失函数的定义公式如下:

$$\text{Loss} = \alpha \text{Loss}_{\text{CE}} + (1 - \alpha) \text{Loss}_{\text{Dice}} \quad (7)$$

其中, $\alpha$ 的取值为0.5, $\text{Loss}_{\text{CE}}$ 代表加权交叉熵损失函数, $\text{Loss}_{\text{Dice}}$ 代表Dice损失函数。

在多病变分割任务中,背景像素占比高于病变像素,模型预测结果会更容易预测背景,而加权交叉熵损失函数可以对不同病变设置不同的权重因子,以此来解决对分割结果影响较大的样本,从而提升分割任务的准确性,其定义如式(8)所示:

$$\text{Loss}_{\text{CE}} = -\sum \omega_i y_i \log(\hat{y}_i) \quad (8)$$

其中, $\omega_i$ 代表第*i*类对应的权重, $y_i$ 表示真实的标签, $\hat{y}_i$ 表示模型预测的概率。眼底视网膜图像背景像素占比较大,病变占比较小,且单个出血点像素占比较小,软性渗出物边缘不清晰,容易被误分为硬性渗出物,因此设置背景、硬性渗出物、出血点、软性渗出物的比重分别为0.1、0.2、0.4、0.3。

Dice Loss通过计算模型预测的多病变结果与真实多病变标注之间的相似度来评估模型的性能,其定义公式如下:

$$\text{Loss}_{\text{Dice}} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (9)$$

其中, $|X \cap Y|$ 表示*X*和*Y*之间交集元素的个数, $|X|$ 和 $|Y|$ 分别表示*X*、*Y*中元素的个数。

## 2.3 评价指标

为全面评估所提出的模型对于3种病变的分割能力,采用医学图像处理领域常用的评估指标,包括精确度(Precision)、召回率(Recall)、Dice系数、交并比(Intersection over Union, IoU)、PR曲线下的面积(Area Under the Precision-Recall Curve, AUPR)。公式定义如下:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (10)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (11)$$

$$\text{Dice} = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}} \quad (12)$$

$$\text{IoU} = \frac{A \cap B}{A \cup B} \quad (13)$$

其中,TP代表真实类别为病变,预测结果也为病变;FP代表真实类别为背景,预测结果为病变;FN代表真实结果为病变,预测结果为背景;TN代表真实类别为背景,预测结果为背景。*A*代表模型预测区域,*B*代表真实标注区域。

### 2.4 实验设置

本研究使用 Pytorch 1.8.1 框架、cuda11.1 版本的 GeForce RTX 2080Ti 显卡进行训练,使用 SGD 优化器进行更新参数,动量设为 0.9,学习率设为 0.01,训练轮数设为 100。

## 3 结果与分析

### 3.1 实验结果

基线 Unet 网络与 MLDA-Unet 的多病变分割实验结果见表 2,通过对比发现改进网络对于硬渗、软渗、出血点的分割效果均相对基础网络有不同程度的提升,证明本研究提出的改进网络能有效提升视网膜多病变分割准确率。在硬渗分割中,所有指标

显著提升,其中,Dice 系数提升 0.027 1,AUPR 提升 0.042 3,验证了改进模型能精准定位硬性渗出物;针对易漏检的出血点,召回率大幅提升 0.079 8,IoU 提升 0.051 8,表明改进网络对小目标特征的捕获能力增强;对于边界模糊的软性渗出物,精确度的下降说明假阳性预测增加,而召回率的显著提升表明模型优先考虑降低假阴性风险,即通过提高病变检出敏感性来避免关键病理特征的遗漏,以免假阴性低导致的漏诊延误 DR 患者的关键治疗窗口期,而且 Dice 系数与 AUPR 分别提升 0.050 4、0.073 4,表明改进模型能有效分割软性渗出物。综上所述,MLDA-Unet 模型通过多尺度特征融合和注意力机制优化,在保持高精度的同时显著降低病变漏检率,在小目标和复杂形状的多病变分割任务中取得优异分割效果。

表 2 Unet 网络与改进网络结果对比  
Table 2 Comparison of Unet and the improved network

病变	Unet					MLDA-Unet				
	精确度	召回率	IoU	Dice 系数	AUPR	精确度	召回率	IoU	Dice 系数	AUPR
EX	0.633 4	0.482 2	0.377 0	0.547 6	0.562 0	0.660 8	0.508 5	0.403 2	0.574 7	0.604 3
HE	0.552 3	0.412 9	0.309 4	0.472 6	0.437 5	0.575 0	0.492 7	0.361 2	0.530 7	0.506 6
SE	0.507 8	0.493 7	0.333 9	0.500 7	0.485 4	0.496 8	0.618 9	0.380 4	0.551 1	0.558 8

SE:软性渗出物;EX:硬性渗出物;HE:出血点

### 3.2 消融实验

为验证改进的 DCHT、DFPA、MSAF 对于 3 种病变分割的有效性,本研究进行了消融实验,实验结果见表 3~5,其中,baseline 代表基线模型 Unet。

通过分析表 3 中不同模块 IoU 系数的对比结果可知,3 种模块对于硬性渗出物、出血点、软性渗出物的分割效果均有不同程度提升。其中,DCHT 对于出血点识别效果较为显著,该模块通过减少下采样过程中特征损失,有效提升对小目标的识别效果,DHCT 与其他两个模块融合扩大了感受野、提升对边缘信息的识别精度,因此有利于对较小面积渗出物的识别。本研究提出的 MLDA-Unet 对于边缘清晰且像素分布面积较小的硬性渗出物的分割效果最佳,相比于基线模型对于出血点的分割提升效果最高,证明改进网络对于小目标识别的有效性。

由表 4 各模块 Dice 系数消融对比结果可知,各改进模块及其两两融合模块均相对基线模型的分割效果有所提升,融合 3 个改进模块的网络模型在分割 3 种病变上均获得最高的 Dice 系数,说明改进网络识别结果与真实结果之间有更高的相似度。不同模块对不同病变分割的影响效果不同,DCHT 模块结合 DFPA 模块对于出血点的分割效果低于单一模块,但是对于渗出物的识别均高于基线模型及单一模块,

表 3 IoU 消融结果对比

Table 3 Comparison of IoU in ablation study

baseline	+DCHT	+DFPA	+MSAF	EX	HE	SE
√	×	×	×	0.377 0	0.309 4	0.333 9
√	√	×	×	0.387 5	0.350 6	0.346 7
√	×	√	×	0.381 2	0.347 7	0.342 9
√	×	×	√	0.391 4	0.346 2	0.349 0
√	√	√	×	0.393 9	0.342 4	0.379 3
√	√	×	√	0.401 2	0.354 7	0.359 9
√	×	√	√	0.394 3	0.354 3	0.374 6
√	√	√	√	0.403 2	0.361 2	0.380 4

表 4 Dice 系数消融结果对比

Table 4 Comparison of Dice coefficient in ablation study

baseline	+DCHT	+DFPA	+MSAF	EX	HE	SE
√	×	×	×	0.547 6	0.472 6	0.500 7
√	√	×	×	0.561 6	0.519 2	0.514 8
√	×	√	×	0.556 7	0.516 0	0.510 7
√	×	×	√	0.562 6	0.506 6	0.495 1
√	√	√	×	0.565 1	0.510 1	0.549 6
√	√	×	√	0.572 7	0.523 6	0.529 3
√	×	√	√	0.565 5	0.523 2	0.545 0
√	√	√	√	0.574 7	0.530 7	0.551 1

可能是出血点大多为小像素点,说明两个模块的融

合对小目标的分割提升能力有限;尽管仅添加MSAF模块对出血点的分割仅提升0.034 0,但是DCHT与MSAF模块相结合在仅两个模块融合的网络中获得最优分割效果。

由表5改进网络中不同模块的AUPR值对比可以发现改进网络相比基线模型在识别病变与背景不平衡的眼底图像中仍获得较高的稳定性。图7直观地展示了依次添加DCHT、DFPA、MSAF模块后,分别计算3种病变的分割精确度和召回率,绘制PR曲线,其中红色曲线代表3个模块均添加的改进网络模型的PR曲线,在硬性渗出物、软性渗出物、出血点这3种病变的PR曲线中,红色曲线相比其他曲线均更靠近右上方,AUPR值均获得不同程度提升,其中软

性渗出物提升最高,也可以观察到改进网络针对软性渗出物的PR曲线与基线模型之间相差较大。

表5 改进网络AUPR消融结果对比

Table 5 Comparison of AUPR in ablation study

baseline	+DCHT	+DFPA	+MSAF	EX	HE	SE
√	×	×	×	0.562 0	0.437 5	0.485 4
√	√	×	×	0.580 2	0.499 0	0.530 1
√	×	√	×	0.573 0	0.497 6	0.486 9
√	×	×	√	0.564 9	0.485 3	0.496 6
√	√	√	×	0.600 0	0.482 9	0.553 7
√	√	×	√	0.602 0	0.497 8	0.529 0
√	×	√	√	0.581 9	0.506 9	0.552 3
√	√	√	√	0.604 3	0.506 6	0.558 8

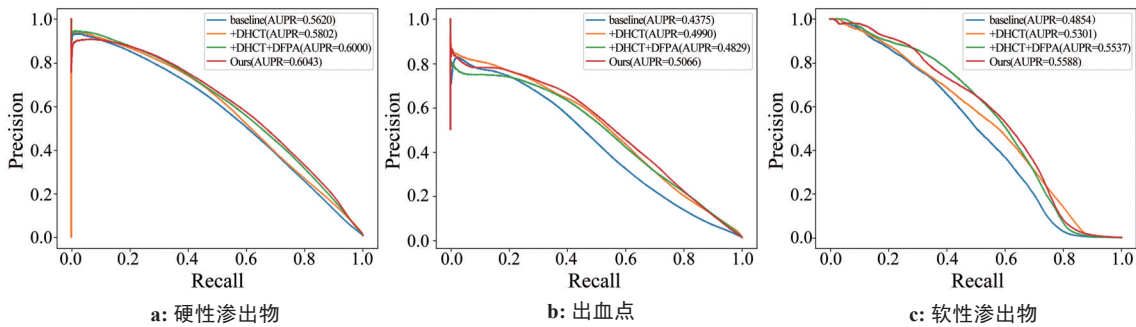


图7 改进模块PR曲线对比

Figure 7 Comparison of PR curves for improved modules

### 3.3 对比实验

通过与常用语义分割模型在同一实验条件下进行实验对比,验证MLDA-Unet的分割性能。PSPNet通过在多个不同尺度上对图像进行池化,构建金字塔池化捕捉图像的上下文信息以提升分割性能<sup>[23]</sup>; Deeplabv3+利用空洞空间金字塔池化将卷积神经网络提取的浅层特征进行扩展,进一步提取深层特征,提升对图像中不同尺度物体的识别能力<sup>[24]</sup>; Unet++改进原始Unet的跳跃连接,类似DenseNet,添加密集跳跃连接和下采样,抓取不同层次的特征,并添加深度监督促进梯度反向传播<sup>[25]</sup>; Attention Unet为促进Unet编码器输出的低级特征与解码器高级语义特征的有效融合,在跳跃连接处增加注意力门控机制,抑制低级特征中冗余特征,加强局部特征的关注度<sup>[19]</sup>; Trans Unet使用Transformer模型替代卷积作为编码器,使用自注意力机制对全局信息进行建模,改善传统Unet在建模长期依赖性方面的局限性<sup>[26]</sup>; Swin Unet使用构建的Swin Transformer Block代替编码器、瓶颈层、解码器构建U型网络, Swin Transformer Block基于原始Transformer结构,通过平移窗口构造自注意力机制,建模图像不同区域之间的关系,并通

过patch扩展层,无需卷积和插值操作既可完成上采样和特征分辨率的恢复<sup>[27]</sup>。

对比实验结果见表6,可以发现除硬性渗出物的IoU和Dice系数低于Deeplabv3+网络,改进网络模型均获得最优分割效果。绘制不同网络模型分割3种病变的PR曲线见图8,可以看出本文方法在3种病变的分割上均更靠近右上角,分割精确度与召回率的综合性能更稳定。PSPNet在所对比网络中分割性能较差,可能是其多个池化模块影响了对于小尺度目标的特征获取; Swin Unet仅在分割硬性渗出物时表现优异,而对于分割小面积的出血点和边缘不清晰的软性渗出物效果不佳; Trans Unet其全局上下文建模能力有助于眼底多病变的分割, IoU分别相对Unet网络提升0.019 1、0.021 3、0.028 0,可见通过改进编码器的全局建模能力也有助于眼底视网膜多病变的分割,但引入Transformer模型会增加网络的复杂度,延长网络训练时间; Unet++与Attention Unet通过不同的方式改进Unet网络的跳跃连接,但相对基础网络分割效果提升不明显,因此寻找更适合的跳跃连接方式去有效融合编码器提取的低级特征与解码器输出的高级特征非常重要; Deeplabv3+在硬性渗

表6 不同网络模型对比实验结果

Table 6 Comparison of different network models

方法	IoU			Dice系数			AUPR		
	EX	HE	SE	EX	HE	SE	EX	HE	SE
Unet	0.377 0	0.309 4	0.333 9	0.547 6	0.472 6	0.500 7	0.562 0	0.437 5	0.485 4
Unet++	0.374 1	0.329 6	0.324 1	0.544 5	0.495 7	0.489 6	0.538 2	0.474 7	0.461 4
PSPNet	0.335 7	0.325 0	0.357 4	0.502 7	0.490 6	0.526 6	0.493 6	0.446 6	0.478 1
DeepLabv3+	0.408 2	0.352 4	0.363 9	0.579 7	0.521 2	0.533 7	0.590 3	0.486 1	0.542 5
Attention Unet	0.372 9	0.329 7	0.283 5	0.543 2	0.495 9	0.441 7	0.542 0	0.456 0	0.444 5
Swin Unet	0.397 3	0.317 6	0.285 2	0.568 6	0.482 0	0.443 8	0.569 6	0.443 8	0.432 5
Trans Unet	0.396 1	0.330 7	0.361 9	0.567 4	0.497 0	0.531 5	0.592 7	0.473 2	0.505 2
本文方法	0.403 2	0.361 2	0.380 4	0.574 7	0.530 7	0.551 1	0.604 3	0.506 6	0.558 8

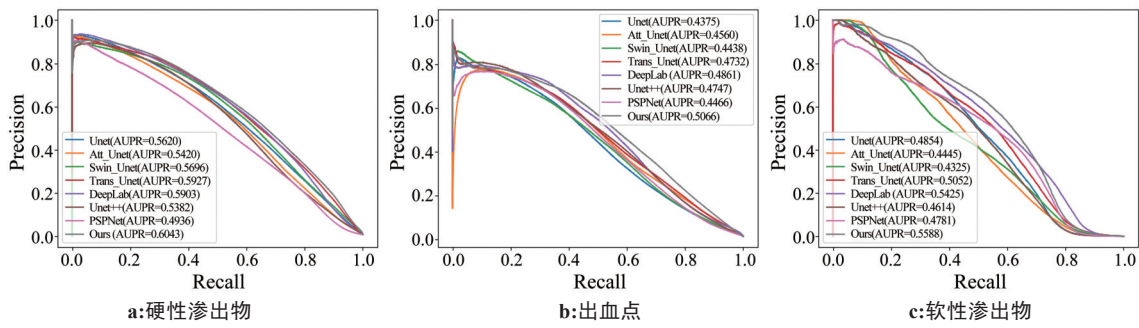


图8 不同网络模型PR曲线对比

Figure 8 Comparison of PR curves of different network models

出物的识别中获得最优性能,但是对于出血点和软性渗出物这两种小病变的识别性能仍有待提升。

### 3.4 分割结果可视化

将3.3节中7种对比病变分割模型与MLDA-Unet的分割结果进行可视化,选取部分视网膜图像的分割结果见图9,对比第一和二张图像的分割结果,可以看出MLDA-Unet相比Unet能减少对于硬性

渗出物、软性渗出物的误判,并有效提升病变边缘分割的准确度;在第三张图像针对大面积出血点的分割中,PSPNet、Trans Unet、Swin Unet识别效果不佳,MLDA-Unet识别出的出血点的轮廓与真实标签最相似;在第四张图片中可以观察到改进后的网络模型相比其它模型对于硬性渗出物的轮廓识别更加清晰准确,并且能正确分割部分小面积渗出物。

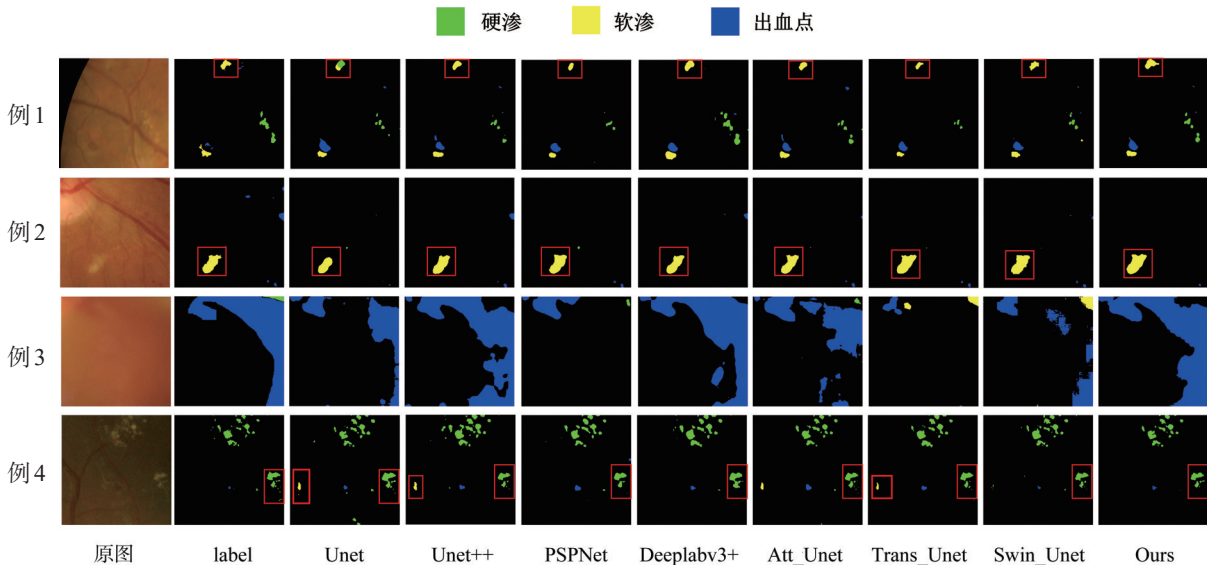


图9 可视化分割结果

Figure 9 Visualization of segmentation results

## 4 结论

本研究提出的 MLDA-Unet 通过结合可变形卷积、空洞卷积和注意力机制分别对 Unet 的编码器、瓶颈层、解码器进行改进。可变形卷积针对目标病变的多种形状进行权重调整,以适应病变结构的复杂形状变化,提取最有效的分割特征;组合多种扩张率空洞卷积可以扩展感受野的尺寸,增强网络对多尺度形变的感知能力;融合空间、通道、尺度多重注意力机制帮助网络关注特征图局部重要区域,提升对局部边缘信息和小目标的分割精度。通过在混合数据集 DDR-RLS 上进行验证,MLDA-Unet 能有效提升对于硬性渗出物、软性渗出物、出血点这 3 种 DR 的分割效果,然而仍存在对硬性渗出物、软性渗出物相似度较大的病变的误识别,并且对于小目标的识别能力仍有待提升。未来研究将从以下 3 个方面进行改进:(1)提升分割病变边缘信息的准确性,同时降低图像噪点对分割效果的影响,使用条件随机场细化图像后处理步骤,充分考虑像素之间的相似性及边缘信息;(2)结合视网膜病变分级等类似图像处理任务,借助多任务之间的关联性辅助进行多病变分割;(3)采用对抗训练和强化学习的方式充分发掘视网膜病变图像的数据分布规律,解决医学注释数据集规模较小的问题,减轻对注释数据集的依赖,并有效适应多设备采集图像,提升应用价值。

## 【参考文献】

[1] 朱芷嫣,陈嘉琪,陈胜凤,等.糖尿病视网膜病变自我管理的知行现状及影响因素的研究进展[J].中国健康教育,2023,39(11):1027-1032.  
Zhu ZY, Chen JQ, Chen SF, et al. Research progress on influence factors related to diabetic retinopathy self-management[J]. Chinese Journal of Health Education, 2023, 39(11): 1027-1032.

[2] Zago GT, Andreão RV, Dorizzi B, et al. Diabetic retinopathy detection using red lesion localization and convolutional neural networks[J]. Comput Biol Med, 2020, 116: 103537.

[3] 刘辉,朱正为,张徐,等.一种结合多尺度特征融合和混合域注意力机制的眼底疾病诊断方法[J].中国医学物理学杂志,2023,40(12):1477-1485.  
Liu H, Zhu ZW, Zhang X, et al. A diagnostic method incorporating multi-scale feature fusion and hybrid domain attention mechanism for fundus diseases[J]. Chinese Journal of Medical Physics, 2023, 40(12): 1477-1485.

[4] Li HY, He HZ, Wen YG. Dynamic particle swarm optimization and K-means clustering algorithm for image segmentation[J]. Optik, 2015, 126(24): 4817-4822.

[5] Kovács G, Hajdu A. A self-calibrating approach for the segmentation of retinal vessels by template matching and contour reconstruction [J]. Med Image Anal, 2016, 29: 24-46.

[6] Aziz T, Ilesanmi AE, Charoenlarnopparut C. Efficient and accurate hemorrhages detection in retinal fundus images using smart window features[J]. Appl Sci, 2021, 11(14): 6391.

[7] Ooi AZ, Embong Z, Abd Hamid AI, et al. Interactive blood vessel segmentation from retinal fundus image based on canny edge detector[J]. Sensors (Basel), 2021, 21(19): 6380.

[8] Sambyal N, Saini P, Syal R, et al. Modified U-Net architecture for semantic segmentation of diabetic retinopathy images[J]. Biocybern Biomed Eng, 2020, 40(3): 1094-1109.

[9] Huang CX, Zong YS, Ding YM, et al. A new deep learning approach for the retinal hard exudates detection based on superpixel multi-feature extraction and patch-based CNN[J]. Neurocomputing, 2021, 452: 521-533.

[10] Huang SQ, Li JN, Xiao YZ, et al. RTNet: relation transformer network for diabetic retinopathy multi-lesion segmentation[J]. IEEE Trans Med Imaging, 2022, 41(6): 1596-1607.

[11] Wu HS, Wang W, Zhong JF, et al. SCS-Net: a scale and context sensitive network for retinal vessel segmentation[J]. Med Image Anal, 2021, 70: 102025.

[12] Geetha Pavani P, Biswal B, Gandhi TK. Simultaneous multiclass retinal lesion segmentation using fully automated RILBP-YNet in diabetic retinopathy[J]. Biomed Signal Process Control, 2023, 86, Part B: 105205.

[13] Wang XF, Xu M, Zhang JC, et al. Joint learning of multi-level tasks for diabetic retinopathy grading on low-resolution fundus images [J]. IEEE J Biomed Health Inform, 2022, 26(5): 2216-2227.

[14] Yao CP, Wang M, Zhu WF, et al. Joint segmentation of multi-class hyper-reflective foci in retinal optical coherence tomography images[J]. IEEE Trans Biomed Eng, 2022, 69(4): 1349-1358.

[15] Butt MM, Iskandar DN, Abdelhamid SE, et al. Diabetic retinopathy detection from fundus images of the eye using hybrid deep learning features[J]. Diagnostics (Basel), 2022, 12(7): 1607.

[16] Ashwini K, Dash R. Grading diabetic retinopathy using multiresolution based CNN[J]. Biomed Signal Process Control, 2023, 86, Part B: 105210.

[17] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.

[18] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Trans Pattern Anal Mach Intell, 2020, 42(8): 2011-2023.

[19] Oktay O, Schlemper J, Folgoc LL, et al. Attention U-Net: learning where to look for the pancreas[EB/OL]. (2018-05-20) [2024-05-12]. <https://arxiv.org/abs/1804.03999>.

[20] Xu GP, Liao WT, Zhang X, et al. Haar wavelet downsampling: a simple but effective downsampling module for semantic segmentation[J]. Pattern Recognit, 2023, 143: 109819.

[21] Dai JF, Qi HZ, Xiong YW, et al. Deformable convolutional networks [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 764-773.

[22] Li T, Gao YQ, Wang K, et al. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening[J]. Inf Sci, 2019, 501: 511-522.

[23] Zhao HS, Shi JP, Qi XJ, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 6230-6239.

[24] Chen LC, Zhu YK, Papandreou G, et al. Encoder-Decoder with atrous separable convolution for semantic image segmentation[C]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 833-851.

[25] Zhou ZW, Rahman Siddiquee MM, Tajbakhsh N, et al. UNet++: a nested U-Net architecture for medical image segmentation[C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer International Publishing, 2018: 3-11.

[26] Chen JN, Lu YY, Yu QH, et al. TransUNet: transformers make strong encoders for medical image segmentation[EB/OL]. (2021-02-08)[2024-06-08]. <https://arxiv.org/abs/2102.04306>.

[27] 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.

(编辑:谭斯允)