

融合密集空洞注意力金字塔和多尺度的视网膜病变分割

王志鲁¹, 池越¹, 周亚同¹, 单春艳², 肖志涛³, 王劭奇¹

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/天津市内分泌研究所/国家卫健委激素与发育重点实验室/天津市代谢性疾病重点实验室, 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(DR)分割任务中病变区域多尺度特征难以学习、边界模糊等问题,提出一种改进的U型多病变分割模型DDAPNet。首先,对DR图像进行Patch处理,使模型更好地捕捉病变的局部特征;其次在主干特征提取后引入重新设计的密集空洞注意力金字塔(DDAP)模块,扩大感受野,解决病变边界模糊问题;同时采用金字塔切分注意力进行特征增强,然后将二者进行特征融合;最后在跳跃连接中嵌入改进的残差注意力模块,降低浅层冗余信息的干扰。在数据集和医院真实数据集上进行联合验证,实验结果表明,相较于基础模型,DDAPNet模型对微动脉瘤、出血点、软渗出DDR物和硬渗出物的分割在Dice系数上分别提高了4.31%、2.52%、3.39%、4.29%,在mIoU上分别提高了1.80%、2.24%、4.28%、1.98%。该模型对病灶边缘的分割更为连续和平滑,有效提升了软渗出物等视网膜病变的分割性能。

【关键词】糖尿病视网膜病变;密集空洞注意力金字塔;多尺度特征;残差模块

【中图分类号】R318;TP391.41

【文献标志码】A

【文章编号】1005-202X(2024)08-1000-10

Diabetic retinopathy segmentation using dense dilated attention pyramid and multi-scale features

WANG Zhilu¹, CHI Yue¹, ZHOU Yatong¹, SHAN Chunyan², XIAO Zhitao³, WANG Shaoqi¹

1. School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China; 2. Tianjin Medical University Chu Hsien-I Memorial Hospital/Tianjin Institute of Endocrinology/Key Laboratory of Hormones and Development of the National Health Commission/Tianjin Key Laboratory of Metabolic Diseases, Tianjin 300134, China; 3. School of Life Sciences, Tiangong University, Tianjin 300387, China

Abstract: An improved U-shaped multi-lesion segmentation model, namely dense dilated attention pyramid UNet (DDAPNet), is proposed to overcome the difficulty in learning multi-scale features and address the issue of blurry boundaries in diabetic retinopathy (DR) segmentation task. DR images are treated with Patch processing to enhance the model's ability to capture local lesion features. After backbone feature extraction, a redesigned dense dilated attention pyramid module is introduced to expand the receptive field and address the issue of blurry lesion boundaries; and simultaneously, pyramid split attention module is used for feature enhancement; and then, the features output by the two modules are fused. Additionally, an improved residual attention module is embedded within skip connections to reduce interference from shallow redundant information. The joint validation on DDR dataset and real dataset from a specific hospital shows that compared with the original model, DDAPNet model improves the Dice similarity coefficient for segmentations of microaneurysms, hemorrhages, soft exudates and hard exudates by 4.31%, 2.52%, 3.39% and 4.29%, respectively, and increases mean intersection over union by 1.80%, 2.24%, 4.28% and 1.98%, respectively. The proposed model makes the segmentation of lesion edges smoother and more continuous, notably enhancing the segmentation performance for conditions like soft exudates in retinal lesions.

Keywords: diabetic retinopathy; dense dilated attention pyramid; multi-scale feature; residual module

【收稿日期】2024-01-10

【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】王志鲁, 硕士研究生, 研究方向:医学图像处理、机器学习, E-mail: 2516153604@qq.com

【通信作者】池越, 副教授, 研究方向:信息感知与机器学习, E-mail: chiyueliuxin@126.com

前言

糖尿病视网膜病变(Diabetic Retinopathy, DR)作为糖尿病最常见且严重的并发症之一^[1],严重影响着全球范围内数以亿计的患者^[2]。DR是一种因长期高血糖而引起的视网膜病变,其病理过程主要包括微动脉瘤(Microaneurysms, MA)的形成、出血点

(Hemorrhages, HE) 和硬性渗出物 (Hard Exudates, EX) 的产生以及晚期病变中大量的软性渗出物 (Soft Exudates, SE)。通过对视网膜图像中的病变区域分割, 可以准确定位和量化病变范围, 帮助医生进行病情评估和治疗决策。然而, 目前临床对 DR 的诊断主要依赖医生对患者眼底图像的手动观察和判断, 这种方式既费时费力又容易受到医生主观因素的影响。因此, 迫切需要一种自动化的 DR 分割方法。

针对 DR 的分割方法主要分为两大类: 传统分割方法和深度学习方法。传统分割方法主要基于形态学处理、边缘检测、阈值等操作来实现对眼底图像中病变的检测分割。Nur 等^[3]通过动态设置阈值来区分渗出物和背景, 实现对渗出物的分割。高玮玮等^[4]先是利用数学形态学结合阈值法提取出视盘, 然后在此基础上利用形态学重建进行视网膜病变的分割。Kumar 等^[5]采用数学形态学方法对眼底图像进行预处理, 实现了对血管和病灶的分割。Saman 等^[6]通过 Canny 边缘检测提取出病灶候选区域, 然后通过分类器进行微动脉瘤的分割。然而这些传统方法依赖于从图像的颜色、空间和纹理中获取手工特征, 存在一定的局限性, 鲁棒性不高。

近年来, 深度学习技术因其出色的性能不断被应用于图像分割领域。为解决病变区域无法在高水平网络上响应的问题, Guo 等^[7]提出了多尺度特征融合以及多通道 bin 损失。Wang 等^[8]提出一种多级任务联合学习的方法, 解决了视网膜病变分割任务中病变分辨率低的问题。Xiao 等^[9]通过将 HEDNet 网络合并到条件生成对抗网络 cGAN 中, 提出一种用于 DR 病变像素级分割的端到端模型。李贺军^[10]通过引入注意力机制, 优化损失函数, 提出一种改进 U 型网络的视网膜渗出物分割模型。杨知桥^[11]通过改进上下文提取编码模块, 提取出更丰富的高级别特征, 并加入混合注意力机制, 突出病变特征, 实现了对渗出物和出血点的分割。虽然上述方法在 DR 分割方面取得了不错的效果, 但是大多数深度学习方法都只针对单病变分割, 单病变分割难以确定 DR 水平, 不利于医生对病情进行评估, 同时由于 DR 的复杂性和特殊性, 视网膜病变分割任务中还存在多尺度特征提取困难、病灶边界模糊等问题。

针对上述分割问题, 本文提出一种基于密集空洞注意力金字塔和多尺度特征增强的多病变分割模型 (Dense Dilated Attention Pyramid UNet, DDAPNet), 并作出以下贡献: (1) 构建了一种新的密集空洞注意力金字塔 (Dense Dilated Attention Pyramid, DDAP) 模块, 用于扩大感受野, 解决病变区域边界模糊问题; (2) 在 UNet 网络的瓶颈层引入金字

塔切分注意力 (Pyramid Split Attention, PSA), 进一步增强更细粒度的多尺度空间信息的提取; (3) 在跳跃连接之间加入改进的残差注意力模块, 提高网络的表达能力, 加强深层特征和浅层特征的融合。

1 本文方法

1.1 整体网络结构

本文提出的 DDAPNet 模型结构如图 1 所示。该模型整体采用编码解码结构, 主要由编码器模块、DDAP 模块、PSA 模块以及跳跃连接组成。首先, 对 DR 图像进行 Patch 处理, 将处理后的图像送入编码器模块中进行特征提取, 得到深层特征。其中, 编码器模块由 5 个特征编码块组成, 每个块由两个 3×3 卷积层, 一个 ReLU 激活函数层和一个 BN 层组成。接着将深层特征送入到 DDAP 模块中, 用于扩大感受野, DDAP 模块应用在编码器的末端, 由 5 个不同空洞率的 DDAP_Block 密集连接组成。同时引入 PSA 模块进行特征增强, 然后将二者特征融合后送入上采样进行特征恢复。最后, 在跳跃连接之间加入残差注意力 (Residual SE Attention, Rse) 模块, 使 PSA 模块输出的深层特征与同层次的浅层特征进行更好的拼接。

1.2 DDAP 模块

在卷积神经网络中, 网络的大量池化和下采样操作会降低图像的分辨率, 导致模型对病变区域的边界和细节感知不足, 而且无法有效涵盖大范围的图像区域。密集空洞空间金字塔池化 (Dense Atrous Spatial Pyramid Pooling, DenseASPP) 具有较大的感受野, 有助于提高对病变边界的识别和分割^[12]。通过采用不同扩张率的空洞卷积来实现不同感受野大小的特征提取, 并且使用密集连接的方式将不同扩张率的空洞卷积层叠加起来^[13], 使得输出具有更大的感受野和更密集的特征金字塔。对于扩张率为 d 、核大小为 K 的空洞卷积层, 其感受野大小为:

$$R_{K,d} = (d - 1) \times (K - 1) + K \quad (1)$$

若将两个感受野大小分别为 R_1 和 R_2 的卷积层进行叠加, 则新的感受野为:

$$R = R_1 + R_2 - 1 \quad (2)$$

在视网膜病变分割任务中, 虽然通过 DenseASPP 可以获得足够大的感受野, 但是对于视网膜病灶这种微小目标还存在关注度不够, 分割边缘不光滑的缺陷。随着扩张率的增大, 参与计算的有效像素点之间的距离也随之增加, 导致它们之间的关联性减小^[14]。因此对 DenseASPP 进行重新设计, 构建了一种新的 DDAP 模块, 其模型结构如图 2 所示。DDAP 采用扩张率分别为 1、2、3、4、5 的 5 种

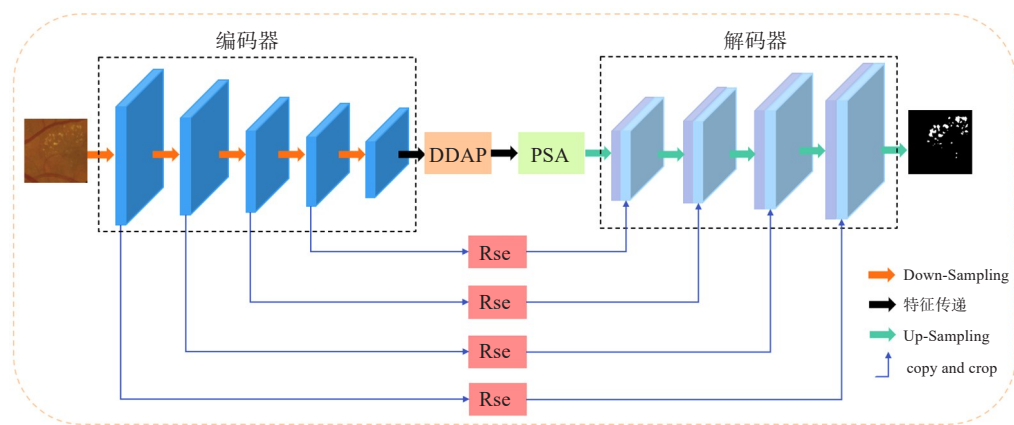


图 1 DDAPNet 模型
Figure 1 DDAPNet model

DDAP_Block 构成特征金字塔,并采用密集连接的网络结构,后一层经过 DDAP_Block 输出的特征图与前面所有层经过 DDAP_Block 输出的特征图合并,最终

将所有层级输出融合。本文设计的 DDAP 模块创新之处主要有两点,第一点设置模型的扩张率为 1、2、3、4、5,第二点重新设计了 DDAP_Block。

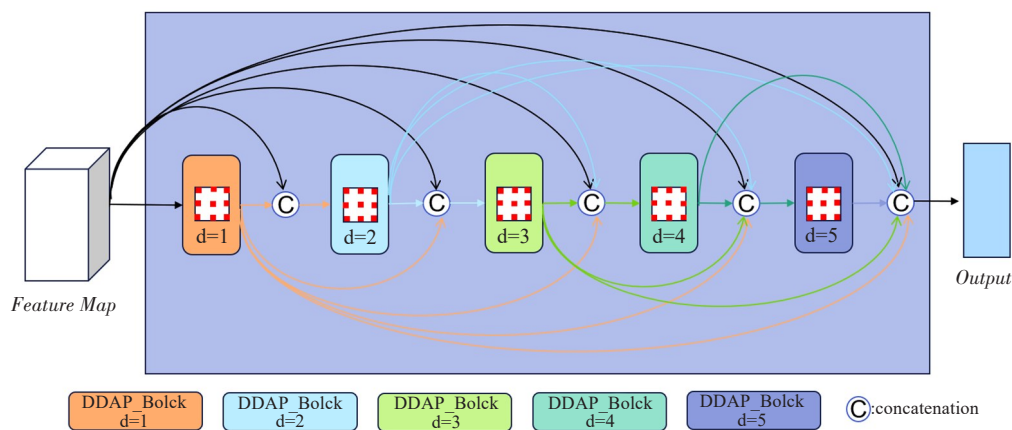


图 2 DDAP 模块
Figure 2 DDAP module

考虑到视网膜病变目标比较小,过大的感受野可能会使图像中的一些重要信息被忽略。由于输入图像的尺寸为 256×256 ,经过编码网络进行 5 次下采样后,图像尺寸变为 16×16 ,为获得多尺度感受野的同时不丢失局部信息,因此将扩张率设置为 1、2、3、4、5。

空洞卷积虽然可以通过增加卷积核的采样间隔来扩大感受野^[15],但是无法根据不同尺度下的特征分布和关系动态调整,可能无法充分捕获有效特征。因此,采用 Dilated Attention 替换空洞卷积。Dilated Attention 是 DilateFormer 中一种简单高效的注意力^[16],它能够模拟小范围内的局部和稀疏的图像块交互,使模型在处理低级信息时,充分利用局部性和稀疏性,而在处理高级信息时,又能模拟远距离依

赖,能够通过自适应地学习不同像素之间的关联性,在每个像素级别上进行关联性建模,更好地整合和利用特征信息。

DDAP_Block 主要是由 Dilated Attention 组成。首先,通过 Dilated Attention 机制对输入特征图进行处理,该过程包括对特征图进行多尺度卷积和特征关联计算,以提取特定尺度和关联性的特征信息。接着,利用 Dropout 机制进行正则化,以防止模型在训练中出现过拟合的情况,提高模型的泛化能力。最终,经过 Dilated Attention 处理后的特征作为模块的输出返回。

1.3 PSA 模块

视网膜病变区域常呈现多尺度和多形态特征,

为进一步增强多尺度特征的提取能力^[17], 捕获更细微的多尺度空间信息, 在DDAP模块输出后引入PSA模块^[18], 然后把DDAP模块输出的特征图和PSA模块输出的特征图进行特征融合, 得到增强后的深层次特征图。

PSA 主要包含两个模块, 即 SPC (Split and Concat)模块和 SEWeight 模块, 其模型结构如图 3 所示。首先, 通过 SPC 模块将输入的视网膜病灶特征

图拆分成 S 块, 获得通道方向的多尺度特征图。接着, 利用 SEWeight 模块对不同尺度的特征图进行注意力提取, 得到每个尺度特征图的通道级注意力向量。然后, 利用 Softmax 对通道级注意力向量进行重新校准, 得到多尺度重新校准权重。最后, 将重新校准后的权重和相应的特征图进行逐元素相乘操作, 得到含有丰富多尺度特征信息的特征图。

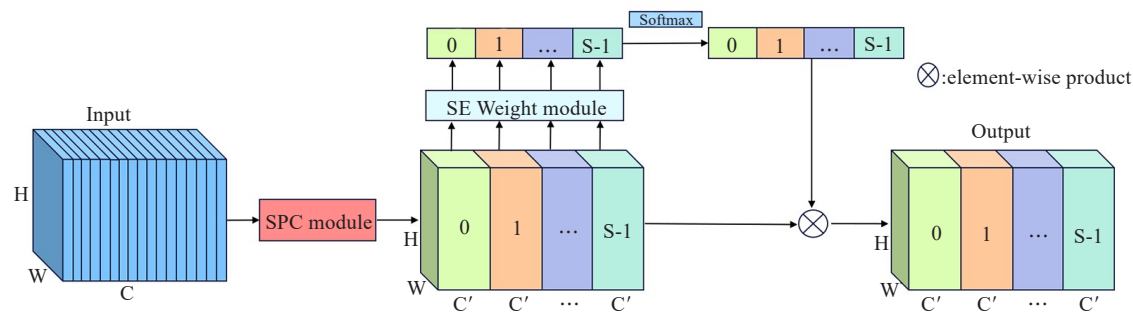


图3 PSA 模块
Figure 3 PSA module

PSA 的关键在于用于多尺度特征提取的 SPC 模块, 其模型结构如图 4 所示。首先, 输入特征图 X 被分割成 S 组 $[X_0, X_1, X_2, \dots, X_{S-1}]$, 其次在多个尺度上并行处理输入张量, 然后利用分组卷积的方法提取出每个信道级特征图上的空间信息, 计算公式如下:

$$F_i = \text{Conv}(k_i \times k_i, G_i)(X_i), i = 0, 1, 2, \dots, S - 1 \quad (3)$$

其中, k_i 为第 i 个内核大小, G_i 为第 i 个组大小, $F_i \in R^{C' \times H \times W}$ 表示不同尺度的特征图。

最后将这些不同尺度的特征图拼接起来, 其计算公式如下:

$$F = \text{Cat}([F_0, F_1, \dots, F_{S-1}]) \quad (4)$$

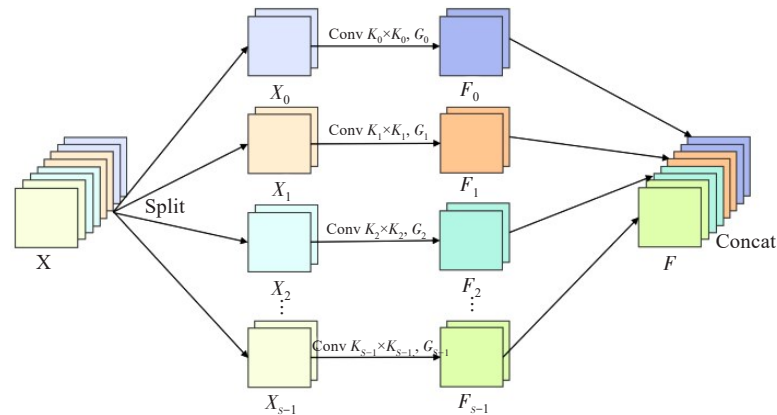


图4 SPC 模块
Figure 4 SPC module

1.4 Rse 模块

通常情况下, 下采样获得的浅层特征中会包含一些冗余信息, 为降低冗余信息的干扰, 本研究对 UNet 网络的跳跃连接进行了改进, 构建了一个 Rse

模块。该模块由残差结构^[19]和 SE (Squeeze-and-Excitation) 注意力机制构成, 这一改进旨在强化网络在特征传递过程中的表征能力, 通过残差结构实现更有效的信息传递, 同时结合 SE 注意力机制有选择

地增强了重要特征表示。

SE注意力是一种轻量级的通道注意力^[20],其模型结构如图5所示。它能对浅层特征图进行通道特征加强,用于计算深浅层特征融合部分特征通道之间的相关性,其过程主要分为挤压(Squeeze)操作和激励(Excitation)操作。Squeeze操作顺着特征维度进行空间特征压缩,将包含全局信息的 $H \times W \times C$ 的特征图经过全局平均池化,压缩为 $1 \times 1 \times C$ 的特征向量,

让每个通道用一个具有全局感受野的数值表示。即输入特征 $U \in R^{H \times W \times C}$ 被压缩为 $X \in R^{1 \times 1 \times C}$,具体公式如下所示:

$$X_c = \frac{1}{W \times H} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j)$$

(5)

其中, U_c 是输入特征 $U \in R^{H \times W \times C}$ 的第 C 个通道的特征图, X_c 是统计向量中的第 C 个元素。

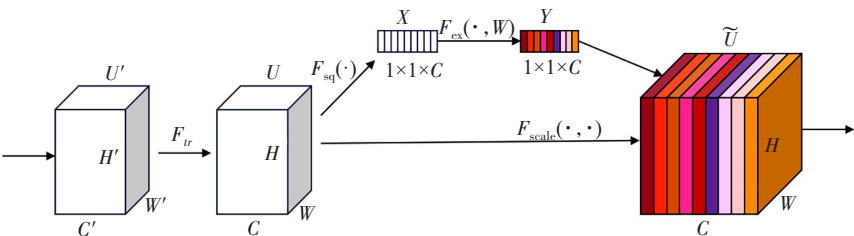


图5 SE注意力
Figure 5 SE attention

Excitation操作连接两个全连接层,对压缩后的特征向量进行通道特征学习,为每个通道赋予不同的权重 Y ,具体公式如下所示:

$$Y = \sigma(\omega_2(\delta(\omega_1(X))))$$

(6)

其中, ω_1 和 ω_2 是使用大小为 1×1 卷积核的卷积操作实现的全连接层, δ 表示ReLU激活函数, σ 表示sigmoid激活函数。

最后将得到的注意力权重 Y 通过相乘操作加权到先前的特征上,得到尺寸为 $H \times W \times C$ 的特征图 \tilde{U} ,具体公式如下所示:

$$\tilde{U} = \text{Fscale}(U_c, Y_c) = Y_c \times U_c$$

(7)

其中,Fscale表示每个特征通道的特征映射与对应的通道权重相乘的运算, $\tilde{U} \in R^{H \times W \times C}$ 。

Rse模块是由残差结构融合SE注意力机制组成,分为主路径和跳跃路径。主路径包括两个分支,第一个分支通过 3×3 的普通卷积处理特征图,第二个分支通过扩张率为3的 3×3 空洞卷积处理特征图,然后将两个分支的输出特征图拼接融合为一个特征图。融合后的特征图被传入SE注意力机制,用于对特征通道进行建模和加权,通过学习得到通道间的重要性权重。跳跃路径由 1×1 的卷积操作和标准化(BatchNorm)组成,通过跳跃路径获得的低级特征,通常是编码(encoder)阶段的特征图。最后,将主路径经过SE注意力机制加权后的高级特征和跳跃路径得到的低级特征进行加和操作后,再经过ReLU激活函数得到最终输出。Rse模块的结构示意图如图6所示。

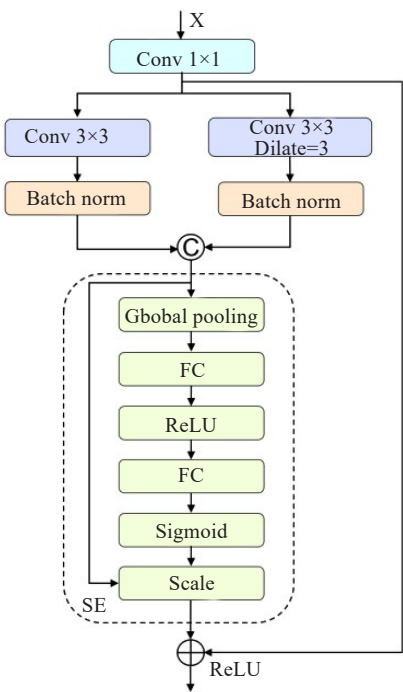


图6 Rse模块
Figure 6 Rse module

2 实验设置

2.1 数据集

本文采用的联合数据集包括来自天津医科大学朱宪彝纪念医院的真实数据集和DDR数据集。其中,医院数据集中带有像素级标注的眼底图像共92张,图像分辨率为 $2\,992 \times 2\,000$,且数据已经做了脱敏处理。DDR数据集是我国最大的DR公开数据集^[21],

共有 757 张眼底图像, 图像分辨率从 1 080×1 920 到 3 456×5 184。为训练更强大和泛化性更好的模型, 本研究将医院数据集和 DDR 公开数据集进行了融合, 即将两个数据集中的训练集、验证集和测试集分别合并成新的训练集、验证集和测试集。由于医院数据集规模比较小, 为平衡两个数据集之间的规模差异, 在数据融合前采用水平翻转、垂直翻转、随机旋转、色彩增强、尺度变换、对比度增强、锐化等方法, 将医院数据集扩充了 8 倍, 数据集划分如表 1 所示。融合后的数据集分辨率大小不一, 在实验时统一分辨率为 256×256。本研究已通过医院伦理委员会批准(伦理批件号: ZXYJNYYKMEC2024-4)。

表 1 数据集划分
Table 1 Division of datasets

数据集	训练集	验证集	测试集	总数
DDR 数据集	383	149	225	757
医院数据集扩充后	376	144	216	736
联合数据集	759	293	441	1 493

2.2 环境设置

实验基于深度学习框架 PyTorch 实现, 使用 PyCharm 集成开发环境, Python 版本为 3.8。在 NVIDIA RTX 3090(24G 显存)的服务器上进行训练调试, Cuda 版本为 11.3。模型训练过程中采用 Adam 优化器进行参数调优, 初始学习率为 0.01, batch size 为 16, 迭代次数为 200, 权值衰减因子设置为 1e-4, 冲量系数为 0.9。

2.3 Patch 处理

由于视网膜图像分辨率较高而病灶目标较小, 并且将整张图像直接输入模型还会增加训练成本。因此本实验对输入图像作 Patch 处理, 具体为: 采用随机裁剪法对训练集进行随机裁剪, 采用滑动窗口法对验证集和测试集进行切片处理, 步长为 128, 处理后的子图大小统一为 256×256。对于处理后不含有病灶的子图, 实验采取剔除的方式, 以确保每张图片至少含有一种病灶。

2.4 损失函数

在模型训练过程中, 由于病灶点占整张图像的比例小, 导致正负样本不平衡, 使得模型倾向于预测较多的像素点, 从而影响分割的准确性。为解决此问题, 本文将交叉熵损失和 Dice 损失进行线性加权组合, 混合损失函数的定义公式如下:

Loss = ρLoss_{Seg} + (1 - ρ)Loss_{Dice} (8)

其中, ρ 的取值为 0.5, Loss_{Seg} 代表交叉熵损失函数, Loss_{Dice} 代表 Dice 损失函数。

交叉熵损失函数是通过计算模型预测结果与真实标签之间的差异来评估模型性能的一种损失函数, 在图像分割任务中, 交叉熵损失函数可以帮助模型更好地学习到像素级别的分类信息, 其定义公式如下:

Loss_{Seg} = y_i × log(p_i) + (1 - y_i) × log(1 - p_i) (9)

其中, y_i 表示真实标签, 取值为 0 或 1; p_i 表示样本属于正类的概率。

Dice 损失函数通过计算模型预测结果与真实标签之间的相似度来评估模型的性能, 可以帮助模型更好地学习到目标形状和边缘信息。其定义公式如下:

Loss_{Dice} = 1 - $\frac{2|X \cap Y|}{|X| + |Y|}$ (10)

其中, |X ∩ 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。

3 实验结果与分析

3.1 评价指标

为了更全面地评估本文算法在 DR 分割上的效果, 采用了定性和定量的分析方法。在定量分析中, 本文使用准确率(Accuracy)、精确度(Precision)、Dice 相似系数(Dice Similarity Coefficient, DSC)和平均交并比(mIoU)4 个指标来评估算法的分割结果, 其计算公式如下:

Accuracy = $\frac{TP + TN}{TP + FN + FP + TN}$ (11)

Precision = $\frac{TP}{TP + FP}$ (12)

DSC = $\frac{2TP}{FP + 2TP + FN}$ (13)

mIoU = $\frac{1}{k} \sum_{i=1}^k \frac{TP}{TP + FP + FN}$ (14)

其中, TP 表示实际为正样本且正确预测为正样本的样本数; FP 表示实际为负样本且错误预测为正样本的样本数; TN 表示实际为负样本且正确预测为负样本的样本数; FN 表示实际为正样本且错误预测为负样本的样本数。

3.2 实验结果

DDAPNet 模型在联合数据集上对 MA、HE、EX 和 SE4 种病灶分割的定量评价指标如表 2 所示。

通过分析表 2 中的各项指标可知, DDAPNet 模型对 4 种病灶的分割都有不错的表现, 其中对于 SE 的分割效果最好, EX 和 HE 的分割效果次之, MA 的分割效果稍差。SE 在大多数评价指标上表现最好,

表 2 DDAPNet 模型分割效果
Table 2 Segmentation performance of DDAPNet model

病灶类型	准确率	精确度	DSC	mIoU
EX	0.980 3	0.762 4	0.672 5	0.789 6
MA	0.991 5	0.692 2	0.579 8	0.668 7
HE	0.973 1	0.745 9	0.633 8	0.742 2
SE	0.980 6	0.870 6	0.674 0	0.760 9

这说明 DDAPNet 模型中添加的多尺度特征提取和增强模块有助于准确地捕获和分割 SE 的边界和细节特征。EX 在 DSC 和 mIoU 上表现较好,这说明添加的 PSA 和残差注意力机制,在特征融合和边缘捕获方面带来了优势。HE 在精确度上表现较高,这说明模型能够较好地地区分 HE,在 DSC 和 mIoU 上(0.633 8 和 0.742 2)表现稍逊,这可能是因为 HE 相对于其他病灶边缘模糊。MA 在准确率方面表现最好(0.991 5),DSC 和精确度(0.579 8 和 0.692 2)较低,可能是因为 MA 的大小和分布导致难以被模型准确地识别和分割。

3.3 消融实验

用多对变体结构去论证 DDAPNet 模型中的 DDAP 模块、Rse 模块、PSA 模块对病变分割结果的改进效果,通过消融实验来验证各模块的有效性,各个变体模型对 DR 分割的结果如表 3~表 6 所示。基础网络 UNet 用“Baseline”表示,“-DDAP”表示在 UNet 网络中添加 DDAP 模块,“-Rse”表示在 UNet 网络中添加 Rse 模块,“-PSA”表示在 UNet 网络中添加 PSA 模块,“-DDAP-Rse”表示在 UNet 网络中添加 DDAP 模块和 Rse 模块,“-DDAP-PSA”表示在 UNet 网络中添加 DDAP 模块和 PSA 模块,“-Rse-PSA”表示在 UNet 网络中添加 Rse 模块和 PSA 模块,“DDAPNet”表示在 UNet 网络中添加 DDAP、PSA 和 Rse 模块,即本文所提出的方法。

表 3 各个变体分割模型的准确率
Table 3 Accuracy of each segmentation model

模型	MA	HE	EX	SE
Baseline	0.991 0	0.969 2	0.978 1	0.978 6
-DDAP	0.991 7	0.971 0	0.978 9	0.979 5
-Rse	0.991 8	0.969 1	0.978 8	0.979 2
-PSA	0.991 5	0.971 2	0.979 0	0.980 2
-DDAP-Rse	0.991 9	0.971 1	0.979 8	0.980 2
-DDAP-PSA	0.992 0	0.972 3	0.979 7	0.979 8
-Rse-PSA	0.991 9	0.971 8	0.979 7	0.980 0
DDAPNet	0.991 5	0.973 1	0.980 3	0.980 6

表 4 各个变体分割模型的精确度
Table 4 Precision of each segmentation model

模型	MA	HE	EX	SE
Baseline	0.621 9	0.681 2	0.706 7	0.842 6
-DDAP	0.653 6	0.704 0	0.725 6	0.853 2
-Rse	0.650 3	0.700 4	0.727 3	0.850 5
-PSA	0.665 4	0.709 2	0.725 1	0.858 0
-DDAP-Rse	0.668 3	0.726 7	0.749 6	0.866 2
-DDAP-PSA	0.670 3	0.737 3	0.741 6	0.865 4
-Rse-PSA	0.665 0	0.722 1	0.749 3	0.865 2
DDAPNet	0.692 2	0.745 9	0.762 4	0.870 6

表 5 各个变体分割模型的 DSC
Table 5 Dice similarity coefficient of each segmentation model

模型	MA	HE	EX	SE
Baseline	0.536 7	0.608 6	0.638 6	0.631 1
-DDAP	0.543 1	0.610 4	0.647 4	0.646 3
-Rse	0.541 7	0.616 1	0.648 8	0.653 6
-PSA	0.548 0	0.618 2	0.647 7	0.649 8
-DDAP-Rse	0.556 3	0.622 5	0.655 3	0.658 7
-DDAP-PSA	0.562 6	0.625 8	0.657 8	0.658 1
-Rse-PSA	0.554 4	0.623 6	0.650 5	0.654 5
DDAPNet	0.579 8	0.633 8	0.672 5	0.674 0

表 6 各个变体分割模型的平均交并比
Table 6 Mean intersection over union of each segmentation model

模型	MA	HE	EX	SE
Baseline	0.650 7	0.719 8	0.746 8	0.741 1
-DDAP	0.653 1	0.732 9	0.750 8	0.745 4
-Rse	0.651 2	0.729 5	0.749 1	0.746 2
-PSA	0.652 2	0.734 0	0.751 4	0.750 6
-DDAP-Rse	0.657 4	0.724 8	0.755 3	0.756 1
-DDAP-PSA	0.660 7	0.736 3	0.756 4	0.751 5
-Rse-PSA	0.660 1	0.735 0	0.754 6	0.754 4
DDAPNet	0.668 7	0.742 2	0.789 6	0.760 9

由表 3 可知,对于 MA、HE、EX 和 SE 这 4 种病灶的分割,各个变体模型相较于基础模型 Baseline 都有不同的提升,这说明各个变体模型的有效性。对于 HE、EX 和 SE 的分割,本文模型在准确率方面获得了最佳表现,而对于 MA 的分割,DDAPNet 模型没有表现出最优的结果,这可能是 MA 目标比较小,很容易

被误判为背景,因此在准确率方面表现不佳。

由表4可知,各个变体模型相较于Baseline在精确度上均有不同程度的提升,且本文模型在这4种病灶上都表现出了最优结果,充分证明了DDAP模块、PSA模块以及Rse模块的有效性。在MA分割方面,DDAPNet模型相对于Baseline和仅添加单一模块的模型,改进效果更为显著,精确度提升较多。在HE和EX分割方面,添加单一模块(-DDAP、-Rse、-PSA)或其组合(-DDAP-Rse、-DDAP-PSA、-Rse-PSA)均有一定程度的提升,但效果相对较轻微。

由表5可知,各个变体模型对MA、HE、EX和SE 4种病灶的分割相较于Baseline在DSC上都有提升,这表明各个改进模型的有效性和可行性。其中本文模型在EX和SE分割上效果最佳(0.672 5和0.674 0),对于MA的分割效果提升显著,提高了4.31%。

由表6可知,在所有病灶类型上,DDAPNet模型的mIoU均高于其他模型,显示出更好的分割效果。单一模块的引入(DDAP、Rse、PSA)有助于提高mIoU,但提升相对较轻微。同时引入两个模块(-DDAP-Rse、-DDAP-PSA、-Rse-PSA)或所有模块(DDAPNet)的组合有助于进一步提高mIoU,表现出更好的分割效果。其中DDAP模块对于HE和EX的分割效果提升最多,这表明DDAP模块对于这两种病灶的特征提取更为有效。

3.4 对比实验

为了进一步验证本文提出的DDAPNet模型的分割效果,将其与UNet^[22]、UNet++^[23]、DeepLab v3+^[24]和SegNet^[25]模型进行对比实验,在联合数据集上分别对4种主要病灶进行了分割实验,分割结果如表7~表10所示。

由表7可知,在MA分割方面,DDAPNet模型在精确度、DSC和mIoU上取得了最佳表现,这表明它在MA分割任务中更有效地捕捉到了病变区域的边界和细节信息。MA通常呈小且散乱的特点,并且与周围背景的对比度较低,很容易被误判为背景,因此在准确率方面没有达到最佳表现。

由表8可知,在HE分割方面,DDAPNet模型在准确率、DSC和mIoU 3个指标中表现最佳,这表明本文模型能够增强网络对零散出血点病灶的感知能力,在捕捉出血点病灶的整体形状和边界方面具有较好的性能。尽管DDAPNet在其他方面表现优异,但在精确度方面未能达到最佳水平,精确度为0.745 9。HE病灶可能包含许多微小的、模糊的特征,这使得一些像素容易被误分类为正类别,降低了模型的精确度。UNet++在精确度方面达到了0.747 4,比DDAPNet稍高。这可能是因为UNet++在其架构中集成了更多

表7 不同模型对微动脉瘤分割性能对比
Table 7 Comparison of different models for microaneurysms segmentation

模型	准确率	精确度	DSC	mIoU
UNet	0.991 0	0.621 9	0.536 7	0.650 7
UNet++	0.989 1	0.643 1	0.568 1	0.661 6
DeepLab v3+	0.992 9	0.648 5	0.566 4	0.640 3
SegNet	0.990 5	0.637 0	0.572 0	0.655 2
DDAPNet	0.991 5	0.692 2	0.579 8	0.668 7

表8 不同模型对出血点分割性能对比
Table 8 Comparison of different models for hemorrhage segmentation

模型	准确率	精确度	DSC	mIoU
UNet	0.969 2	0.681 2	0.608 6	0.719 8
UNet++	0.970 5	0.747 4	0.624 7	0.730 4
DeepLab v3+	0.971 1	0.734 3	0.630 2	0.729 1
SegNet	0.970 0	0.732 7	0.623 1	0.728 6
DDAPNet	0.973 1	0.745 9	0.633 8	0.742 2

表9 不同模型对硬渗出物分割性能对比
Table 9 Comparison of different models for hard exudate segmentation

模型	准确率	精确度	DSC	mIoU
UNet	0.978 1	0.706 7	0.638 6	0.746 8
UNet++	0.978 3	0.737 7	0.646 9	0.755 1
DeepLab v3+	0.977 5	0.740 9	0.637 0	0.751 8
SegNet	0.976 4	0.711 0	0.632 4	0.740 3
DDAPNet	0.980 3	0.762 4	0.672 5	0.789 6

表10 不同模型对软渗出物分割性能对比
Table 10 Comparison of different models for soft exudate segmentation

模型	准确率	精确度	DSC	mIoU
UNet	0.978 6	0.842 6	0.631 1	0.741 1
UNet++	0.980 4	0.750 2	0.641 0	0.736 1
DeepLab v3+	0.978 4	0.766 7	0.635 6	0.745 6
SegNet	0.979 5	0.743 1	0.626 1	0.729 9
DDAPNet	0.980 6	0.870 6	0.674 0	0.760 9

的跳跃连接和特征融合机制,有助于更好地保留细微特征和边界信息,提高了一些难以分类的像素的分类准确性。

由表9可知,在EX分割方面,本文模型的分割结果优于其它4种模型,这表明引入DDAP模块可以提取到更为丰富的多尺度特征,能够突出病灶的边缘信息,提高DR的分割精度。与其他模型相比,DDAPNet在精确度方面表现出色,这意味着它能够更准确地识别出EX,减少了误分割。

由表10可知,在SE分割方面,本文模型在4种评价指标上都取得了最优值,DDAPNet通过增强对多尺度特征的提取和对边界细节的捕捉,对SE的分割效果有显著改进。这种改进使得模型能更准确地识别和表达SE病灶,从而提高了精确度、DSC和mIoU等关键指标,

表现出较高的分割准确性和全局表达能力。

为了更直观的评价本文模型的分割性能,将DDAPNet与其它4种模型的分割结果进行可视化对比,如图7所示。图7a~g依次为原始视网膜图像、Ground-Truth、UNet、UNet++、DeepLab v3+、SegNet以及DDAPNet分割结果图。通过对比可以看到DDAPNet模型的分割效果要优于其他模型,对细微且复杂的视网膜病灶图像分割效果较好,能够有效分割出病变区域,得到的分割图更接近于Ground-Truth,并且生成的边缘更为平滑,同时也没有产生多余的分割部分。

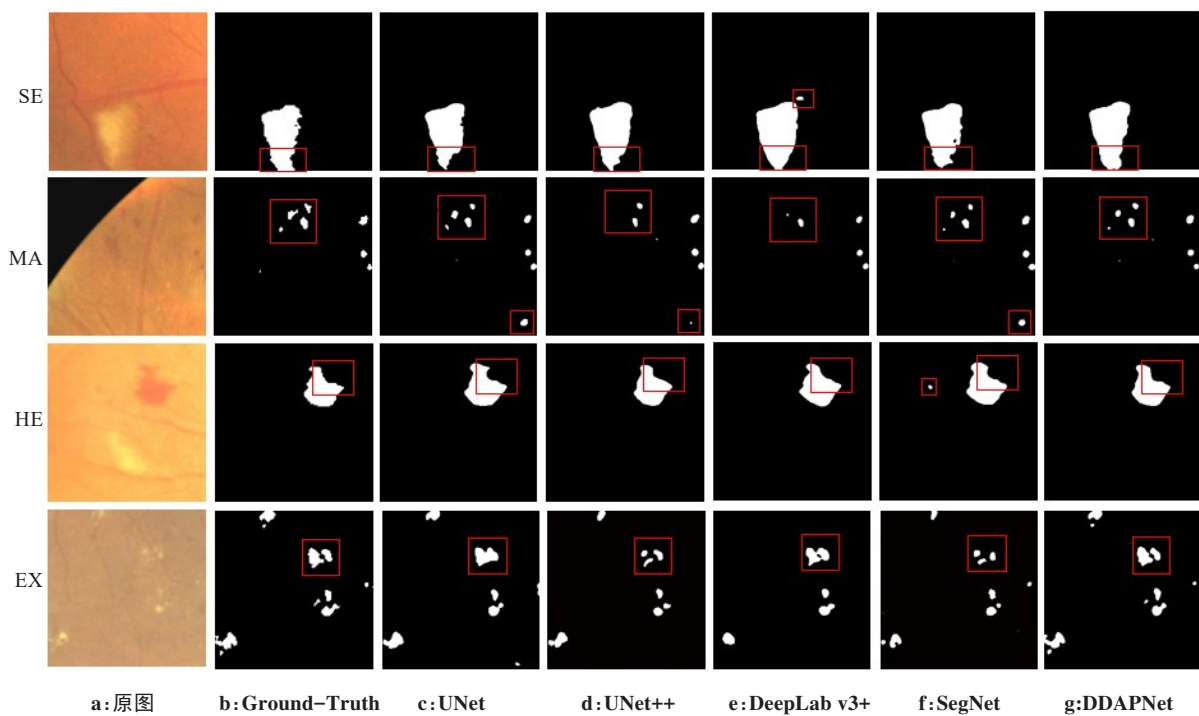


图7 4种病灶分割结果
Figure 7 Segmentation results of 4 types of lesions

3.5 Grad-CAM可视化分析

为了更加直观地评价所提出的DDAP模块的有效性,证明该模块可以扩大感受野、增强对病灶关键特征的提取。在实验过程中使用梯度加权类激活映射(Gradient-weighted Class Activation Mapping, Grad-CAM)方法进行可视化解释^[26]。通过生成的热力图,可以直观地观察模型在关注哪些区域,并评估其对关键特征的提取能力,其效果如图8所示。通过对比基础网络UNet和添加DDAP模块后的UNet可知,添加了DDAP模块的模型生成的热力图中红色区域更广泛、更集中,特别是在视网膜病变区域周围。这表明该模型对病变区域的关注程度更高,能够更

准确地感知边缘和细微特征。相比之下,未使用DDAP模块的UNet生成的热力图中的高亮区域相对较小,而且比较分散。因此,进一步说明了DDAP模块的有效性,有助于改善分割结果的准确性和鲁棒性。

4 结 论

本文提出的视网膜病变分割模型DDAPNet实现了对MA、HE、EX和SE 4种病灶的联合分割,通过结合DDAP模块、PSA模块以及Rse模块解决了DR分割任务中存在的多尺度特征提取困难、边界模糊等问题,并且采用加权交叉熵损失和Dice损失的联合

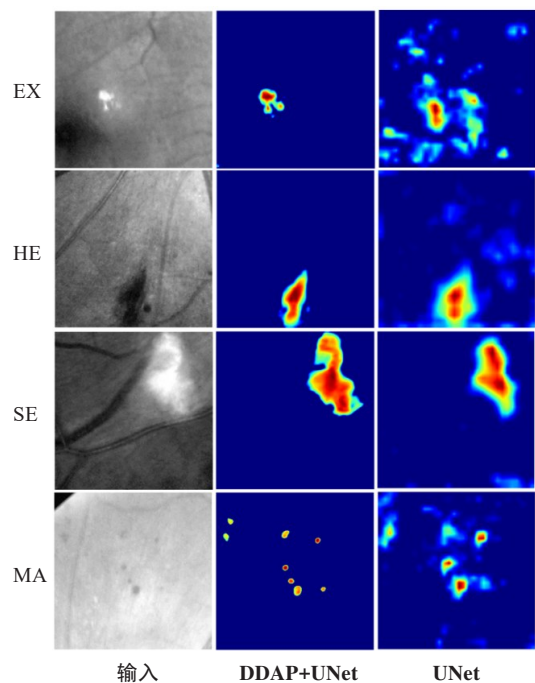


图8 4种病灶的可视化结果图
Figure 8 Visualization results of 4 types of lesions

损失函数来优化DDAPNet,通过在DDR公开数据集和医院真实数据集上联合验证,本文模型较已有模型均有不错的提升。未来的研究方向可以在进一步优化网络结构和注意力机制的基础上,探索更多先进的图像处理和分割方法,以实现更准确和可靠的病变分割效果。

【参考文献】

[1] 张凤俊,李晶明,刘秋平. 糖尿病视网膜病变发病机制及潜在治疗研究进展[J]. 眼科新进展, 2020, 40(7): 677-685.
Zhang FJ, Li JM, Liu QP. Pathogenesis and potential treatment of diabetic retinopathy[J]. Recent Advances in Ophthalmology, 2020, 40(7): 677-685.

[2] Walter T, Klein JC, Massin P, et al. A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina[J]. IEEE Trans Med Imaging, 2002, 21(10): 1236-1243.

[3] Nur N, Tjandrasa H. Exudate segmentation in retinal images of diabetic retinopathy using saliency method based on region[J]. J Phys Conf Ser, 2018, 1108(1): 012110.

[4] 高玮玮,沈建新,王玉亮. 基于数学形态学的快速糖尿病视网膜病变自动检测方法[J]. 光谱学与光谱分析, 2012, 32(3): 760-764.
Gao WW, Shen JX, Wang YL. Method of fast and automated detection of diabetic retinopathy based on mathematical morphology [J]. Spectroscopy and Spectral Analysis, 2012, 32(3): 760-764.

[5] Kumar S, Adarsh A, Kumar B, et al. An automated early diabetic retinopathy detection through improved blood vessel and optic disc segmentation[J]. Opt Laser Technol, 2020, 121: 105815.

[6] Saman G, Gohar N, Noor S, et al. Automatic detection and severity classification of diabetic retinopathy[J]. Multimed Tools Appl, 2020, 79(43): 31803-31817.

[7] Guo S, Li T, Kang H, et al. L-seg: an end-to-end unified framework

for multi-lesion segmentation of fundus images[J]. Neurocomputing, 2019, 349: 52-63.

[8] 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.

[9] Xiao QQ, Zou JX, Yang MQ, et al. Improving lesion segmentation for diabetic retinopathy using adversarial learning [C]//International Conference on Image Analysis and Recognition. Cham: Springer International Publishing, 2019: 333-344.

[10] 李贺军. 基于深度学习的糖尿病视网膜病变分析与研究[D]. 桂林: 桂林理工大学, 2022.
Li HJ. Analysis and research of diabetic retinopathy based on deep learning[D]. Guilin: Guilin University of Technology, 2022.

[11] 杨知桥,张莹,王新杰,等. 改进U型网络在视网膜病变检测中的应用研究[J]. 计算机科学与探索, 2022, 16(8): 1877-1884.
Yang ZQ, Zhang Y, Wang XJ, et al. Application research of improved U-shaped network in detection of retinopathy[J]. Journal of Frontiers of Computer Science & Technology, 2022, 16(8): 1877-1884.

[12] Yang MK, Yu K, Zhang C, et al. DenseASPP for semantic segmentation in street scenes[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 3684-3692.

[13] Yin W, Zhou D, Fan T, et al. Image segmentation of skin lesions based on dense atrous spatial pyramid pooling and attention mechanism[J]. Journal of Biomedical Engineering, 2022, 39(6): 1108-1116.

[14] Fei C, Luo J. DtUNet: A transformer-based UNet combined with denseaspp module for pancreas segmentation [C]//2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, 2022: 1-7.

[15] Chen LC, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Trans Pattern Anal Mach Intell, 2018, 40(4): 834-848.

[16] Jiao JY, Tang YM, Lin KY, et al. DilateFormer: multi-scale dilated transformer for visual recognition[J]. IEEE Trans Multimedia, 2023, 25: 8906-8919.

[17] Pang SC, Du AA, Yu ZM, et al. Correlation matters: multi-scale fine-grained contextual information extraction for hepatic tumor segmentation [C]//Advances in Knowledge Discovery and Data Mining. Cham: Springer International Publishing, 2020: 462-474.

[18] Zhang H, Zu KK, Lu J, et al. EPSANet: an efficient pyramid squeeze attention block on convolutional neural network[C]//Computer Vision-ACCV 2022. Cham: Springer Nature Switzerland, 2023: 541-557.

[19] 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.

[20] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.

[21] 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.

[22] 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.

[23] 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.

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

[25] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(12): 2481-2495.

[26] Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 618-626.

(编辑:薛泽玲)