西北大学学报(自然科学版)
2026 年 10 月,第 56 卷第 5 期,Oct. ,2026,Vol. 56,No. 5
Journal of Northwest University (Natural Science Edition)

·深度学习视觉处理与分析·

基于双编码结构的病理性近视病灶分割

赵婷婷,张元元

西北大学 数学学院,陕西 西安 710127

摘要 病理性近视病灶的精准自动化分割在眼科临床具有关键意义,可减轻眼科医生分析眼底图像负担、提速诊断并为治疗提供可靠依据。但病灶形状大小各异,部分边界模糊且与周围组织相似,这使确定其位置和大小任务复杂,限制了深度学习模型的分割性能。为此,该文提出一种基于双编码器的深度网络,用于处理病灶复杂形态特征。该网络使用 ResNet-34 编码器提取图像局部特征,同时,引入 PVTv2 编码器获取全局特征,全面把握病灶及周围组织信息。借助软注意力机制融合双编码器特征,并依据病灶特征自动分配权重,聚焦关键区域,提升不规则及边界模糊病灶的分割精度。该文设计了空间通道增强注意力模块,使模型敏锐捕捉病灶与周围组织的细微特征差异,精准确定病灶边界和范围。在 PALM19 数据集上的实验结果显示,所提方法的 Dice 系数、交并比 IoU 分别为 84.63%、75.71%,为病理性近视病灶精准分割提供了有效解决方案。

关键词 病理性近视;双编码器;软注意力;空间通道增强注意力

中图分类号:TP391.4 **DOI**:10.16152/j.cnki.xdxbzr.2026-05-012

Pathological myopia lesion segmentation using dual encoding structure

Zhao Tingting, Zhang Yuanyuan

School of Mathematics, Northwest University, Xi'an 710127, China

Abstract In ophthalmic clinical practice, the precise automated segmentation of pathological myopia lesions holds great significance. It can relieve ophthalmologists of the burden of analyzing fundus images, accelerate diagnosis, and offer a reliable basis for treatment. However, lesions vary significantly in shape and size, with some boundaries being blurred and similar to the surrounding tissues. This makes determining their location and size a complex task and restricts the segmentation performance of deep learning models. Therefore, this article puts forward a deep network based on dual encoders for handling the complex morphological features of lesions. This network uses ResNet-34 encoder to extract local features of images. Meanwhile, it introduces a PVTv2 encoder to obtain global features, thus comprehensively grasping the information of lesions and surrounding tissues. By leveraging the soft attention mechanism to fuse the features of the dual encoders and automatically assign weights according to lesion characteristics, focusing on key regions, the segmentation accuracy of irregular and boundary-blurred lesions is enhanced. Simultaneously, this article designs a spatial-channel

收稿日期:2026-05-20 修回日期:2026-06-05

基金项目:国家自然科学基金(12471472)

第一作者:赵婷婷,女,副教授,硕士生导师,从事生物数学研究,ztt0701@nwu.edu.cn。

enhanced attention module, enabling the model to sensitively capture the subtle feature differences between lesions and surrounding tissues, and accurately determine the boundaries and ranges of lesions. Finally, the numerical experimental results on the PALM19 dataset demonstrated that the Dice coefficient and intersection of union(IOU) of the proposed method were 84.63% and 75.71%, providing an effective solution for the accurate segmentation of pathological myopia lesions.

Keywords pathological myopia; dual encoding structure; soft attention; spatial-channel enhanced attention

病理性近视也叫退行性近视,属于高度近视的一种。其主要特征是眼球轴向长度异常伸长,这会使视网膜变薄,并在眼底引发多种可能损害视力的病变。病理性近视是导致视力损害的重要原因,可引发多种视网膜病变[1]。在中低度近视患者中,病理性近视的发生率为 1% ~19%,而在高度近视人群中,这一概率急剧上升到 50% ~70%,这表明高度近视患者面临的风险显著增加[2]。病理性近视往往会导致患者视力受损,给社会带来一定的疾病负担,已逐渐成为全球公共卫生领域需要重点关注的问题之一。流行病学调查显示,病理性近视相关眼底病变是我国不可逆性致盲眼病的主要原因之一[3]。因此,早期诊断对于防止视力进一步恶化至关重要,而精确的自动化分割技术能够减轻眼科医生分析眼底图像的工作负担,提高诊断速度,为治疗提供可靠依据。

病理性近视中,病灶形状大小各异,部分边界模糊,与周围组织相似,这极大地增加了准确确定病灶位置与大小的难度。早期研究对病灶分割的关注较少,多聚焦于形状规则的伴视盘萎缩病灶[4]。病理性近视病灶分割是眼底成像技术中用于精确分离和标记病理性近视眼部病变的特定算法。眼底成像技术,尤其是光学相干断层扫描和眼底摄影,在此过程中至关重要。眼底摄影能够提供高分辨率视网膜图像,有助于医生细致观察与近视相关的视网膜变化。基于眼底图像的视网膜疾病自动分割是治疗视力丧失的重要临床决策辅助工具,也为患者提供了高效、精准的医疗服务,对改善治疗效果意义重大[5]。深度学习作为先进图像处理技术,在病灶分割领域广泛应用。医学图像分割中,基于卷积神经网络的方法较为流行,U-Net 模型[6]凭借其经典的编码器 - 解码器架构成为众多任务的首选。在此基础上,研究者开发了多种改进模型:Attention U-Net[7]引入软注意力机制优化跳跃连接,减少冗余信息,提升灵敏度和预测精度;TransUNet[8]采用 Transformer 模块[9]补充卷积层编码器输出,丰富全局上下文信息,增强理解能力;Transfuse 模型[10]将 U 形结构改造成双分支并行结构,通过并行融合卷积和 Transformer 路径为 Transformer 方法提供空间诱导偏差信息,提升分割性能。这些改进模型既继承了 U-Net 的优点,又弥补了其局限性,为医学图像分割带来了新突破。

基于 U-Net 和 Attention U-Net 等深度学习方法的医学图像分割技术在病理性近视眼底病变区域的精确分割上取得了显著进展。然而,现有数据集的特性给模型性能带来了新的挑战。如图 1 所示,病变区域形状多变,不遵循固定模式,与背景高度相似,在颜色、纹理和形态上差异细微,且病灶位置各异,区分难度大。因此,针对形状多变、边界模糊且与周围组织视觉特征高度相似的微小病灶,分割精度仍有提升空间。根据病灶形状及发展阶段,病理性近视病灶可分为早期伴视盘萎缩、弥漫性萎缩、斑块状萎缩、黄斑萎缩和视网膜脱离 5 类,分别对应图 1(a) ~(e)。图 1 上半部分为眼底彩照,下半部分为对应的病灶图像。针对数据集特点,本文提出一种新型双分支并行结构模型。该模型在含卷积神经网络(CNN)和 Transformer 分支的双编码器结构基础上,加入软注意力机制及空间通道增强注意力模块(spatial-channel enhanced attention, SCEA),有效提升了分割精度和处理效率。具体而言,借助注意力机制和多尺度特征提取,网络能精准捕捉形状不规则的病变特征,解决了传统方法因无法有效利用多尺度特征而导致细微病灶分割性能不佳的问题。鉴于眼底图像中大部分区域为背景,模型期望聚焦于病灶区域,忽略非关键的背景区域,从而集中资源处理病变区域,避免在非关键区域浪费计算资源。

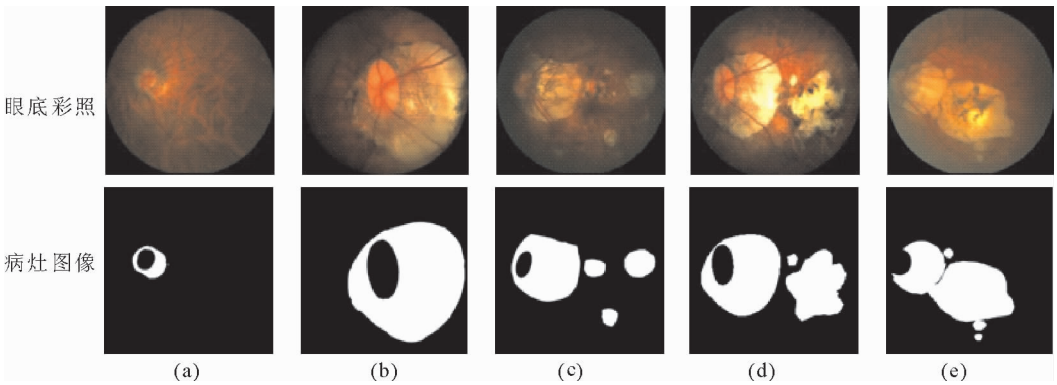


图 1　各类病理性近视眼底彩照与病灶图像

Fig. 1　Fundus color photographs and lesion masks for various categories of pathological myopia

1　研究方法

1.1　PM-Net 整体结构

本文提出的用于病理性近视病灶分割的网络结构 PM-Net 采用双编码器 – 解码器架构,其框架如图 2 所示。首先,利用一个预训练的标准 ResNet-34 编码器[11]提取图像局部特征,该编码器由残差块和最大池化层组成。ResNet-34 中的恒等映射可加快网络收敛速度并有效缓解梯度消失问题。然而,卷积神经网络编码器因结构限制,难以有效捕获全局上下文信息,这可能导致病灶区域分割出现误差[12]。因此,本文引入一个额外分支,使用预训练的 Pyramid Vision Transformer (PVTv2)作为编码器,主要目的是降低计算复杂性并改善分割结果[13-14]。输入的 224 × 224 像素的图像经 PVTv2 编码器处理,该编码器通过重叠分块插入和连续分层 Transformer 块捕获分层图像特征,将输入图像转化为丰富特征表示,从而更全面地捕捉复杂病变区域。借助软注意力机制,可有效融合 PVTv2 的全局特征与 ResNet-34 的局部特征,提升分割精度。双编码器架构整合了 CNN 编码器和 Transformer 编码器,所得特征通道数量是单个编码器输出的 2 倍。但单纯卷积层难以有效捕捉不同特征通道间的关键相关性,会限制分割精度进一步提升。为此,引入空间通道增强注意力模块,挖掘特征通道间依赖性及关键特征相关性。最终,通过双线性插值技术将处理后的特征图上采样至原始图像尺寸,完成精确分割任务。在跳跃连接融合模块(skip-connection fusion module, SCFM),利用分组卷积与点卷积高效实现特征融合及通道间充分信息交互。

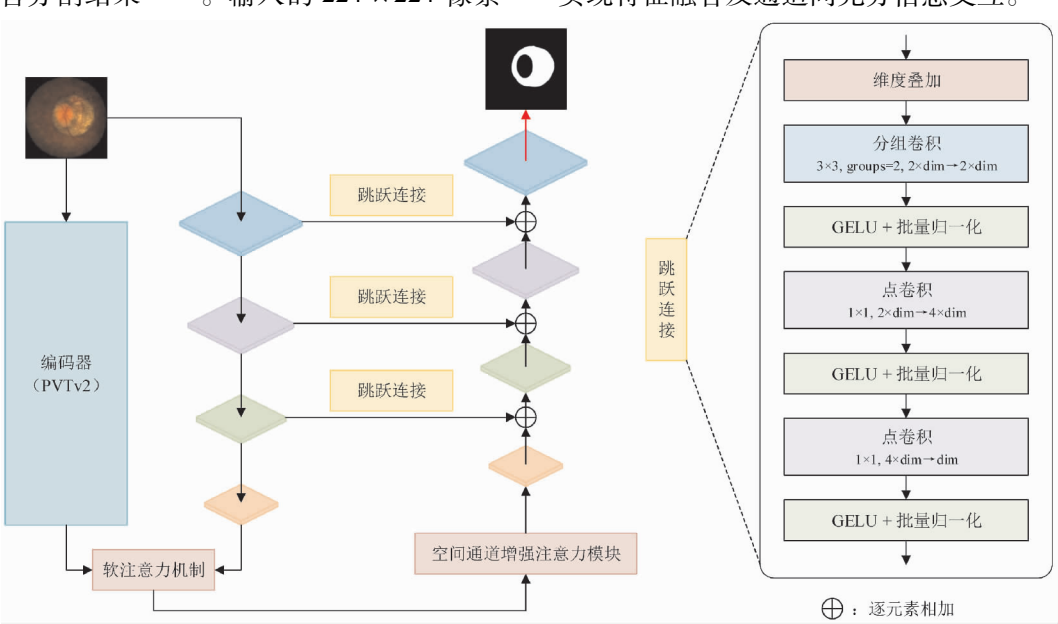


图 2　PM-Net 示意图

Fig. 2　PM-Net schematic

1.2 软注意力机制

PVTv2 借助自注意力机制处理输入图像,使模型能够更加关注全局上下文以及细节处理,PVTv2 所得到的特征图通常蕴含着丰富的空间信息和上下文关系。与之相比,ResNet-34 在捕捉局部特征方面更具优势,尤其擅长表达边缘和纹理信息。而软注意力机制进一步增强了模型的灵活性,它通过学习为输入特征的不同区域分配权重,让模型在预测时能够聚焦于关键特征,同时抑制无关的背景信息。具体来说,它将 2 个编码器的输出进行逐元素相加,并经过 ReLU 激活,以此强化重要特征的表达。接着,通过 1 ×1 卷积和 Sigmoid 函数,计算出注意力系数,该系数再与 PVTv2 编码器的输出特征图逐元素相乘,最终生成加权特征图。本文中的软注意力机制结构如图 3 所示,计算公式如式(1)所示。

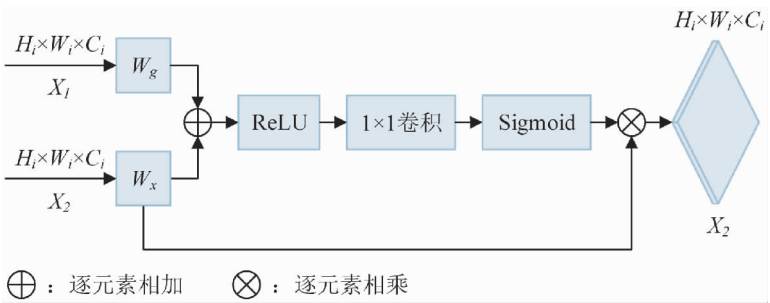


图 3 软注意力机制

Fig. 3 Soft attention mechanism

$$\psi = \mathrm{Sigmoid}(\mathrm{BN}(\mathrm{Conv}_{2d}(\mathrm{ReLU}(W_g \oplus W_x))))$$
$$\boldsymbol{A}_{\mathrm{out}} = W_x \cdot \psi \quad (1)$$

式中:ψ 表示注意力系数;BN 表示批量归一化;Conv_{2d} 表示二维卷积操作;W_g、W_x 为特征变换模块;$\boldsymbol{A}_{\mathrm{out}}$ 为最终输出的加权特征图。

这种动态权重调整策略使模型能够根据引导特征的信息,灵活地判断目标特征的重要性。在病理性近视的病灶分割中,软注意力机制确保关键的局部细节得到适当强调,从而提升模型的表现力。

1.3 空间通道增强注意力模块(SCEA)

在病理性病灶分割任务中,模型的特征提取能力对其性能起着重要的作用。本文提出的双分支结构能够有效捕捉图像的局部特征和多尺度特征。为了在保持高效性的同时增强模型对噪声的鲁棒性,本文引入了空间通道增强注意力模块,其结构如图 4 所示。该模块融合了通道注意力和空间注意力机制,目的在于提升特征提取的效果与鲁棒性。

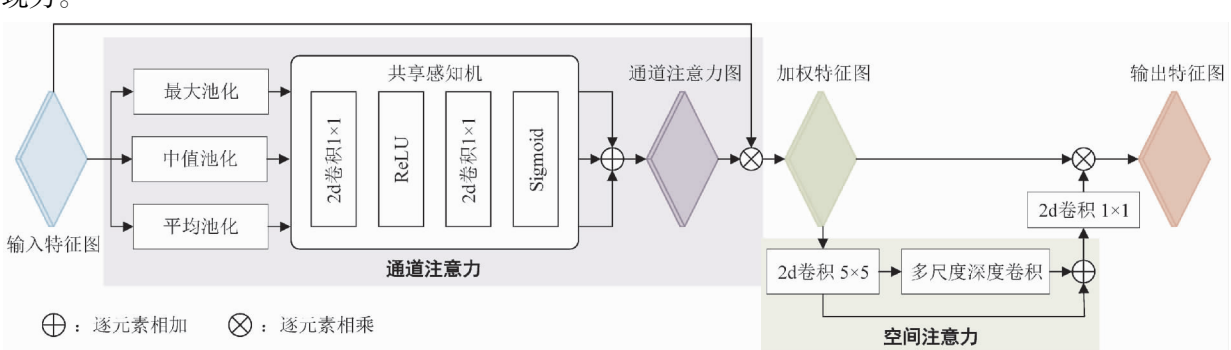


图 4 空间通道增强注意力模块

Fig. 4 Spatial-channel enhancement attention module

在通道注意力的构建上,传统的全局平均池化和全局最大池化在处理噪声时存在一定的局限性。因此,本文融合中值池化、全局最大池化和全局平均池化,构建鲁棒性更优的通道注意力机制。中值池化在有效去除噪声干扰的同时,还能精准地保留关键的特征信息。

在空间注意力子模块的设计上,本文考虑多尺度深度卷积,充分获取特征图的多尺度信息。首先利用 5 ×5 卷积提取输入特征的基础特征,然后经由若干多种尺寸的深度卷积从特征中捕获不同尺度的上下文信息。对多尺度特征逐元素相加,输入 1 ×1 卷积层,生成空间注意力特征图。

基于加权特征和注意力图,进行逐元素相乘运算,完成最终特征的输出。SECA 模块主要包含通道注意力、空间注意力两大核心组件。

1.3.1　通道注意力

通道注意力模块对输入特征图的全局统计信息进行聚合,学习得到通道注意力权重,实现对特征通道的加权,其步骤如下。

1)通过全局平均池化、全局中值池化和全局最大池化,输入特征图得到 3 份尺寸一样(均为 $\mathbf{R}^{C\times1\times1}$,$C$ 表示通道数)的池化结果。

2)将 3 组池化结果输入至包含 2 个 1×1 卷积和 1 个 ReLU 激活函数的共享多层感知器(MLP)。首层卷积依据降维比率 r 实现特征维度从 C 缩减到 C/r;第 2 层卷积实现特征维度的复原,还原为 C。随后经过 Sigmoid 激活,把输出映射至[0,1]区间,生成 3 个注意力图。

3)基于 3 个注意力图,采用逐元素相加生成最终通道注意力图。

4)原始输入特征图与通道注意力图进行逐元素相乘,输出加权后的特征。

$$\begin{aligned}\boldsymbol{F}_C = &\sigma(\mathrm{MLP}(\mathrm{AvgPool}(\boldsymbol{F}))) + \\ &\sigma(\mathrm{MLP}(\mathrm{maxPool}(\boldsymbol{F}))) + \\ &\sigma(\mathrm{MLP}(\mathrm{MedianPool}(\boldsymbol{F})))\end{aligned}$$

$$\boldsymbol{F}' = \boldsymbol{F}_C \otimes \boldsymbol{F} \tag{2}$$

式中:σ 表示 Sigmoid 函数;$\otimes$ 表示逐元素乘法。

1.3.2　空间注意力

空间注意力模块利用多尺度深度卷积捕获特征图的空间关联信息,以此生成对应的空间注意力图,其步骤如下。

1)通过 5×5 深度卷积实现基础特征的提取,该卷积运算过程保证输出特征图与输入尺寸保持改变。

2)为捕获不同感受野对应的多尺度特征信息,将初始卷积输出的特征图分流至多组不同尺寸的深度卷积分支(1×11、1×7 等)。

3)将各深度卷积分支的输出特征执行逐元素求和运算,实现多尺度特征融合。

4)利用 1×1 卷积层对融合特征做映射变换,输出空间注意力权重图。计算公式如下:

$$\boldsymbol{F}_S = \sum_{i=1}^{n} D_i(\boldsymbol{F}') \tag{3}$$

式中:D_i 表示各分支深度卷积运算;n 表示深度卷积分支数量。

基于空间注意力图与通道加权特征图,做逐元素乘法,获取增强后的输出特征。整体模块运算如下:

$$\boldsymbol{F}'' = \mathrm{Conv}_{1\times1}(\boldsymbol{F}_S) \otimes \boldsymbol{F}' \tag{4}$$

1.4　跳跃连接融合模块(SCFM)

传统的跳跃连接通常采用加和或卷积运算来进行特征融合,这种方式较为生硬直接,并且会增加编码器和解码器的负担[15]。而本文则采用了分组卷积和点卷积来实现高效的特征融合以及通道间的信息交换。具体来说,将卷积运算分为 2 组,对编码器特征进行提取,其中,组卷积的核大小为 3×3,步长为 1,填充为 1。为了实现更充分的特征融合,在组卷积之后加入了 2 个反向瓶颈逐点卷积。跳跃连接会将融合前的特征自适应地分配给组卷积,随后逐点 1×1 卷积,在组卷积之后进一步融合特征,实现复杂特征的有效整合。此外,跳跃连接中的每个卷积操作后面都配备了一个 GELU 激活层和 BatchNorm 层,有助于确保特征的非线性变换以及稳定性。

1.5　损失函数

本文使用 Dice 损失和二元交叉熵损失的平均值作为损失函数。Dice 损失函数在处理类别不平衡问题时表现出色,其源于用于衡量 2 个集合相似性的 Dice 系数。Dice 损失函数的计算公式如下:

$$L_{\mathrm{Dice}} = 1 - 2 \times \frac{\sum_i P_i G_i}{\sum_i P_i + \sum_i G_i} \tag{5}$$

式中:i 表示图像中每个像素中的位置索引;P_i 和 G_i 分别代表预测概率图和真实标签图中第 i 个像素的值。

此外,由于病理性近视病灶分割属于像素级二分类问题,本文还引入了二元交叉熵损失函数,与 Dice 损失函数共同构成复合损失函数,旨在减少过拟合现象,同时确保模型能够获得良好的收敛速度。其公式如下:

$$\begin{aligned}L_{\mathrm{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}&[G_i\log(P_i)] + \\ &(1-G_i)\log(1-P_i)\end{aligned} \tag{6}$$

本文损失函数计算公式为

$$L = \mathrm{mean}(L_{\mathrm{Dice}} + L_{\mathrm{BCE}}) \tag{7}$$

2 数值实验

2.1 数据集及预处理

随着屈光不正程度加深,高度近视可能演变为病理性近视,对患者视力造成永久性伤害。因此,病理性近视病灶分割和持续监测至关重要。本文实验所用数据集来自 2019 年国际医学成像大会的 PALM19 挑战赛[16],由中山大学中山眼科中心等机构联合推出,专注于病理性近视诊断和病变区域分割。PALM19 数据集包含约 50% 病理性近视患者和正常受试者的 1 200 张注释完备的视网膜眼底图像,均匀分配到训练数据集、验证数据集和测试数据集,每个集合 400 张图像。数据集涵盖 1 444 ×1 444 和 2 124 ×2 056 这 2 种分辨率图像,按 80% 和 20% 比例随机划分为训练集和测试集[17]。为丰富训练数据多样性并预防模型过拟合,本文采用随机旋转、翻转、缩放和平移等技术进行数据增强。

由于眼底病变形态多样且特征敏感度差异显著,PALM19 数据集中又存在并发多种病灶,若直接用于训练,必然导致类内方差增大,从而削弱网络的诊断准确性。为解决这一问题,本文引入 Gamma 校正算法对原始眼底图像进行深度增强,该方法在保持病灶、血管等微小结构不变的前提下,通过调节图像亮度实现对不同光照条件的鲁棒性。考虑到本实验训练集中样本数量有限,而深度模型对数据量有较高需求,本文对每张原始图像分别应用 γ 值为 0.6、0.8、1.2、1.4 的 4 种典型 Gamma 变换,生成光照梯度各异的 4 组增强样本(见图 5),以成倍扩充训练集规模。原始图像经过这 4 种 Gamma 值变换后,既保留了眼底原始特征,又能模拟从暗光到过曝的多种照明场景,从而有效强化模型对病变大小、形状及灰度对比度等敏感特征的学习能力,显著提升了网络在跨光照域下的泛化能力和诊断性能[18]。

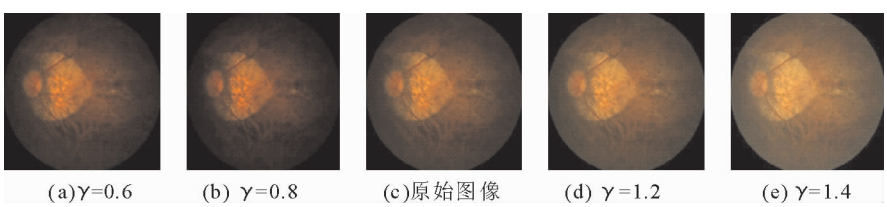


图 5 不同 Gamma 变换值处理后的输入图像

Fig. 5 Input images processed with different Gamma transform values

2.2 实验环境与参数设置

本文实验采用 Adam 优化器进行训练,先以学习率预热策略将初始学习率从 0.001 平滑提升至 0.01,以抑制早期训练的不稳定性。预热结束后再引入指数衰减调度,使学习率稳健下降,进一步加快收敛并减少振荡。训练共进行 200 轮,批量大小为 8,输入图像尺寸统一调整为 224 ×224 像素。所有实验均在 Linux Ubuntu 18.04.4LTS 操作系统、配备 12 GiB 显存的 NVIDIA GeForceRTX 2080 Ti 上完成,编程语言为 Python 3.11.0、框架为 PyTorch 2.2.1。整个过程耗时 8 h 4 min 5 s,且通过实时监控,峰值显存占用约 4.75 GiB。

2.3 实验结果与分析

2.3.1 评价指标

本文使用几种被广泛采纳的性能评价指标,对所提方法的性能进行了全面评估,包括准确性(ACC,式中简记 P_{ACC})、交并比(IoU,式中简记 P_{IoU})、Dice 系数(Dice,式中简记 P_{Dice})、召回率(Recall,式中简记 P_{Recall})以及 F1 分数(F1 score,式中简记 F_1)。计算公式分别为:

$$
\begin{aligned}
P_{\mathrm{ACC}} &= \frac{P_{\mathrm{TP}} + P_{\mathrm{FN}}}{P_{\mathrm{TP}} + P_{\mathrm{FP}} + P_{\mathrm{TN}} + P_{\mathrm{FN}}} \\
P_{\mathrm{Dice}} &= \frac{2P_{\mathrm{TP}}}{2P_{\mathrm{TP}} + P_{\mathrm{FP}} + P_{\mathrm{FN}}} \\
P_{\mathrm{IoU}} &= \frac{P_{\mathrm{TP}}}{P_{\mathrm{TP}} + P_{\mathrm{FP}} + P_{\mathrm{FN}}} \\
P_{\mathrm{Recall}} &= \frac{P_{\mathrm{TP}}}{P_{\mathrm{TP}} + P_{\mathrm{FN}}} \\
P_{\mathrm{Presion}} &= \frac{P_{\mathrm{TP}}}{P_{\mathrm{TP}} + P_{\mathrm{FP}}} \\
F_1 &= \frac{P_{\mathrm{Presion}} \times P_{\mathrm{Recall}}}{P_{\mathrm{Presion}} + P_{\mathrm{Recall}}}
\end{aligned}
\tag{8}
$$

式中:P_{TP}(真阳性)和 P_{TN}(真阴性)分别表示正确分割的病理性近视病灶像素和背景像素的数量;P_{FP}(假阳性)表示被错误标记为病理性近视病灶像素的背景像素的数量;而 P_{FN}(假阴性)表示被错误预测为背景像素的病理性近视病灶像素

的数量,指标值越大,表明分割效果越好。

2.3.2　消融实验

为了验证网络中各组件的有效性,本文进行了消融实验,对比不同方法在配备不同模块时的性能表现,旨在评估 PVTv2 分支、软注意力机制、空间通道增强注意力模块及跳跃连接模块的有效性。以 ResUNet 作为基线模型进行对比,实验结果详见表 1。

表 1　不同模块在 PALM19 数据集上的消融实验结果

Tab. 1　Ablation experiments of different modules on the PALM19 dataset　　单位:%

模型	ACC	Dice	IOU	Recall	F1 分数
基线模型(ResUNet)	95.88	77.83	66.98	78.23	77.83
基线模型 + PVTv2	96.12	80.75	70.48	82.27	80.75
基线模型 + PVTv2 + 软注意力机制	96.54	83.05	73.49	84.67	83.05
基线模型 + PVTv2 + 软注意力机制 + SCEA	96.95	83.83	74.67	81.74	83.83
基线模型 + PVTv2 + 软注意力机制 + SCEA + SCFM	**97.01**	**84.63**	**75.71**	**84.79**	**84.63**

注:加黑数值表示最优结果。

针对 PVTv2 分支的有效性,通过对比“基线模型”与“基线模型 + PVTv2”的各项性能指标,发现引入 PVTv2 解码器分支后,准确率提升了 0.24 百分点,Dice 系数提高了 2.92 百分点,交并比提升了 3.5 百分点,表明 PVTv2 解码器分支能显著增强网络的分割性能。PVTv2 分支具备强大的多尺度特征提取能力,能够捕捉不同层次、不同粒度的图像特征,为网络提供更丰富、更具判别性的特征信息,有助于更精准地识别不同大小和复杂程度的目标,从而提升整体的分割效果。

对于软注意力机制的有效性,本文对比了“基线模型 + PVTv2”与“基线模型 + PVTv2 + 软注意力机制”,引入软注意力机制后,ACC 提升了 0.42 百分点,Dice 系数提高了 2.3 百分点,IOU 提升了 3.01 百分点。软注意力机制使网络能够聚焦于图像中对分割任务更为关键的区域和特征,根据不同区域和特征对于分割结果的重要性分配不同的权重,使网络在处理分割任务时更具针对性和精准性,进而有效提升了分割性能。

对于空间通道增强注意力模块的有效性,本文对比了“基线模型 + PVTv2 + 软注意力机制”和“基线模型 + PVTv2 + 软注意力机制 + SCEA”,添加 SCEA 后,ACC 提升了 0.41 百分点,Dice 系数提高了 0.78 百分点,IOU 提升了 1.18 百分点。空间通道增强注意力模块能够同时从空间和通道 2 个维度对特征进行增强和调整,使网络能够更好地捕捉到特征之间的内在联系以及不同通道所蕴含的重要信息,进一步提升网络对目标的分割能力。

最后,针对跳跃连接模块的有效性,本文对比了“基线模型 + PVTv2 + 软注意力机制 + SCEA”和“基线模型 + PVTv2 + 软注意力机制 + SCEA + SCFM”,添加跳跃连接模块后,ACC 提升了 0.06 百分点,Dice 系数提高了 0.8 百分点,IOU 提升了 1.04 百分点。跳跃连接模块通过一系列卷积操作、激活函数和批量归一化层,对输入特征进行复杂的非线性变换和通道间的信息交互,从而增强特征表达能力。

2.3.3　对比实验

为了验证本文提出的 PM-Net 的优越性,将其与 U-Net、ResUNet、Attention U-Net、TransUNet、Transfuse、EMCAD[19]、ERDUnet[20] 等现有的多种先进方法的分割性能进行了对比。实验结果如表 2 所示,可以看出,本文提出的 PM-Net 模型相较于次优方法 TransUNet,在准确率上提升了 0.51 百分点,Dice 系数提高了 1.97 百分点,交并比提升了 2.94 百分点。

表 2　不同方法在 PALM19 数据集上的分割性能

Tab. 2　Segmentation performance of different methods on the PALM19 dataset　　单位:%

方法	ACC	Dice	IOU	Recall	F1 分数
U-Net	96.07	76.22	65.08	73.86	76.22
ResUNet	95.88	77.83	66.98	78.23	77.83

续表 2 单位:%

方法	ACC	Dice	IOU	Recall	F1 分数
Attention U-Net	95.47	76.53	65.43	80.92	76.53
TransUNet	96.50	82.66	72.77	82.53	82.66
Transfuse	96.71	82.55	72.75	84.54	82.55
EMCAD	96.14	79.59	68.95	81.98	79.59
ERDUnet	94.51	69.16	56.58	77.79	69.16
PM-Net	**97.01**	**84.63**	**75.71**	**84.79**	**84.63**

注:加黑数值表示最优结果。

从上述量化结果来看,PM-Net 在分割结果与真实标签的相似度、分割目标边缘的分割效果以及整体分割准确性等方面均表现优异。结合上一节的消融实验,充分证明了本文所设计的各个模块的有效性。

2.3.4 分割结果可视化

图 6 展示了不同方法对各类病灶进行分割操作后的可视化结果。观察图中第 5 行的可视化结果,可以发现 U-Net 在分割时主要关注与对比度相关的特征信息,这导致其容易将与病灶颜色差异不明显的正常区域误判为病灶区域。Attention U-Net 也存在类似问题,但相较于 U-Net,其分割误差相对较小。进一步分析图 6 中第 2 行的可视化结果,可以看出 PM-Net 在针对较小目标的分割方面表现出色,能有效避免因目标过小而导致的漏诊情况。此外,从第 3 行和第 4 行的结果来看,本文提出的针对病理性近视病灶的 PM-Net 分割方法在准确性方面明显优于其他分割方法,能更精准地界定和划分病理性近视病灶。

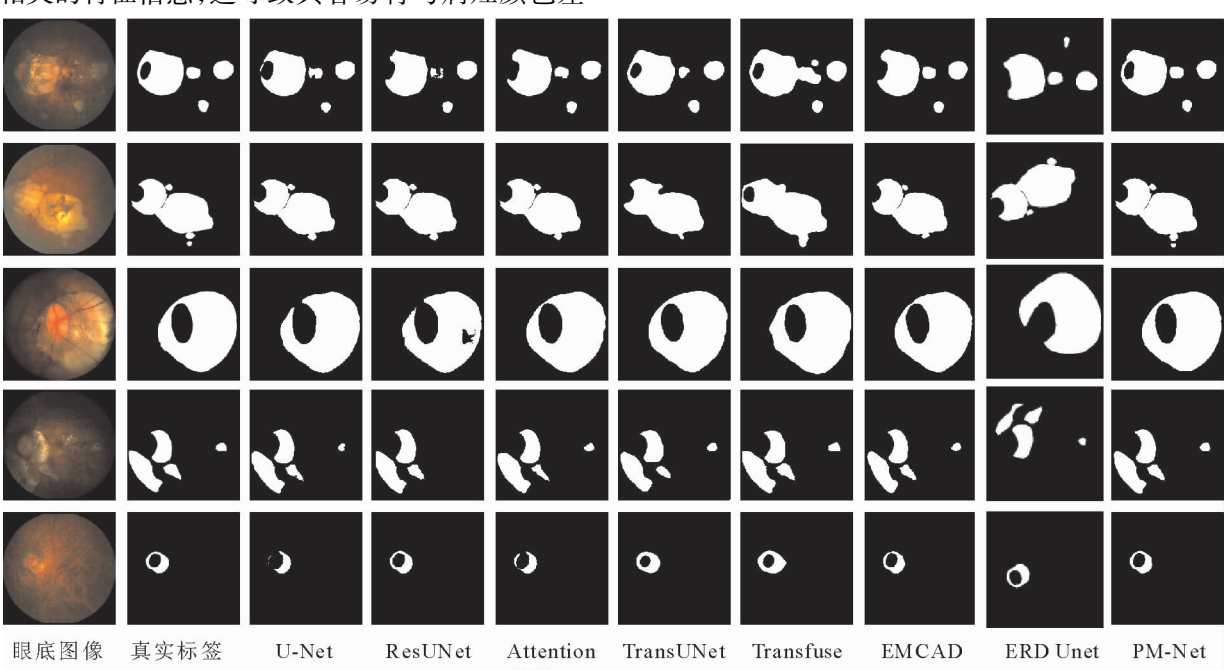


图 6 不同方法的分割结果可视化

Fig. 6 Visualization of segmentation results by different methods

2.3.5 PM-Net 模型的泛化性测试

为了验证 PM-Net 模型的有效性,将模型与 U-Net、ResUNet、Attention U-Net 和 Transfuse 在血管分割数据集 FIVES 上进行了对比实验,测试其泛化能力,结果如表 3 所示。

在 FIVES 血管分割数据集上的对比实验结果表明,PM-Net 的泛化能力尤为突出。具体来看:PM-Net 在分割准确率上达到 97.25%,高于 Attention U-Net(97.10%)和 U-Net(96.99%);在 Dice 系数上也以 77.05% 超越了所有对比模型,

其中,次优模型 ResUNet 为 76.22%;PM-Net 的交并比达到 65.12%,较次优模型 ResUNet 提升了约 3 百分点;同时,其召回率提升至 70.26%,相较于 Attention U-Net(68.55%)显著增强。以上各项指标的一致性领先,充分说明 PM-Net 在陌生的 FIVES 数据域中,依然能够稳健捕捉血管细节,验证了 PM-Net 跨数据集的较强泛化能力。

从图 7 的可视化结果可以看出:U-Net 能较完整地提取主干血管轮廓,却在末梢分支处频繁断裂并遗漏大量细小血管;Attention U-Net 在中粗血管分割上更为平滑,但对极细小分支仍有缺失且部分背景出现虚线条;ResUNet 在保持中等粗细血管连续性方面优于前两者,细节恢复有所提升,但依然存在局部断裂和过分割现象;Transfuse 在细血管感知上最为薄弱,导致整体分支稀疏;PM-Net 则兼顾了主干血管的连贯性与末梢细分支的高覆盖,噪声抑制能力最强,展现出较好的细节还原与连通性,充分验证了其跨域泛化能力。

表 3　不同模型在 FIVES 数据集上的测试结果

Tab. 3　Test results of different models on the FIVES dataset

单位:%

方法	ACC	Dice	IoU	Recall
U-Net	96.99	74.07	61.30	65.34
Attention U-Net	97.10	75.99	61.90	68.55
ResUNet	96.89	76.22	62.04	65.56
Transfuse	91.93	61.25	44.49	62.51
PM-Net	**97.25**	**77.05**	**65.12**	**70.26**

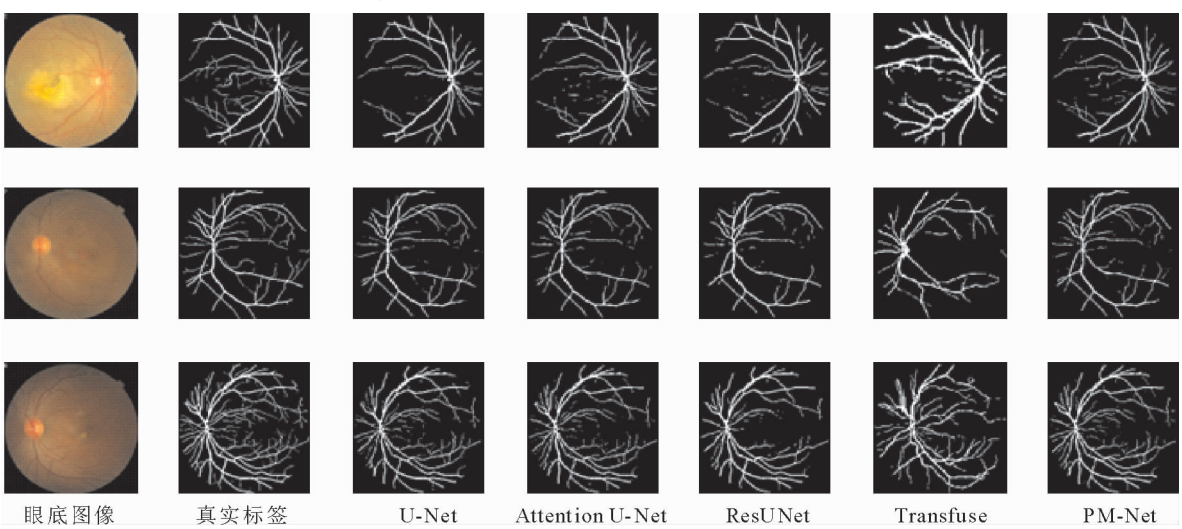


图 7　不同方法在 FIVES 数据集上的分割结果可视化

Fig. 7　Visualization of segmentation results of different models on the FIVES dataset

2.4　局限性分析

在实际临床应用中,PM-Net 仍存在一些不可忽视的局限与挑战。首先,对于体积非常小、边缘细微的病灶,网络往往难以完整捕捉其全部轮廓,预测掩码的边界通常较真实标注明显偏内,导致小病灶的敏感性下降;其次,当出现大范围、形态复杂的病变区时,由于模型在高维特征融合过程中倾向于平滑处理,分割结果的边缘常表现为模糊或欠锐利,难以与专家标注的精细边界完全对齐。尽管如此,PM-Net 能够在大多数中等规模病灶上实现较为准确的提取,其整体分割形状与真实标签高度一致,这说明模型在常见病灶场景下具备良好的稳健性。但针对“小而精”与“广而散”两类极端情况,后续仍需引入更高分辨率特征、多尺度边缘增强或专门的边界感知损失,以进一步提升对极小病灶和超大病变区的分割精度。

3　结语

本文提出了一种应用于病理性近视病灶分割的 PM-Net 网络结构,该结构基于双编码器 - 解码器架构。在特征提取阶段利用 ResNet-34 编码器提取图像的局部特征,但由于卷积神经网络编码器自身结构的局限性,在捕获全局上下文信息方面存在不足,这可能会在病灶区域分割过程中引入误差。因此,本文引入了 PVTv2 编码器分支,以获取更全面的全局特征。同时,借助软注意力机制,将双编码器提取的特征有效结合,从而显著提升分割精度。

在双编码器架构的基础上,进一步设计了空间通道增强注意力模块,旨在深度挖掘特征通道之间的内在依赖性及关键特征的相关性,为提升

分割精度提供有力支持。在跳跃连接部分,采用分组卷积与逐点卷积处理输入特征,确保不同层次的特征信息有效整合。最终,通过双线性插值技术对经过一系列处理后的特征图进行上采样,使其恢复至原始图像尺寸,完成病理性近视病灶的精确分割任务。

参考文献

[1] 王春燕, 艾育德. 病理性近视病因学研究的最新进展[J]. 内蒙古医学杂志, 2004, 36(1): 41-44.
Wang C Y, Ai Y D. Latestprogress of study on etiology of pathological myopia[J]. Inner Mongolia Medical Journal, 2004, 36(1): 41-44.

[2] 张望, 刘康, 高建华, 等. 国际近视研究院关于病理性近视的报告[J]. 中华实验眼科杂志, 2023, 41(4): 366-387.

[3] Yorston D. Retinal diseases and VISION 2020[J]. Community Eye Health, 2003,16(46):19-20.

[4] 李永安. 基于深度学习的病理性近视识别及病灶分割技术研究与实现[D]. 桂林: 桂林理工大学, 2023.

[5] 茹月, 师燕芸. 超广角眼底成像在病理性近视周边视网膜形态改变中的诊疗价值[J]. 中国药物与临床, 2021, 21(19): 3277-3280.

[6] 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, 2015: 234-241.

[7] Oktay O, Schlemperer J, Folgoc L L, et al. Attention U-Net: Learning where to look for the pancreas[J]. Medical Image Analysis, 2020, 58: 101563.

[8] Chen J N, Lu Y Y, Yu Q H, et al. TransUNet: Transformers make strong encoders for medical image segmentation [EB/OL]. (2021-09-08) [2025-09-20]. https://arxiv.org/abs/2102.04306.

[9] 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: Curran Associates Inc., 2017: 5998-6008.

[10] Zhang Y D, Liu H Y, Hu Q. TransFuse: Fusing transformers and CNNs for medical image segmentation[C]// Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer, 2021: 14-24.

[11] He K M, Zhang X Y, Ren S Q, et al. Deep residual learningforimage recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.

[12] Wu H S, Chen S H, Chen G L, et al. FAT-Net: Feature adaptive transformers for automated skin lesion segmentation[J]. Medical Image Analysis, 2022, 76: 102327.

[13] Wang W H, Xie E Z, Li X, et al. PVT v2: Improved baselines with pyramid vision transformer[J]. Computational Visual Media, 2022, 8(3): 415-424.

[14] Jha D, Tomar N K, Biswas K, et al. CT liver segmentation via PVT-based encoding and refined decoding [C]// 2024 IEEE International Symposium on Biomedical Imaging (ISBI). Athens: IEEE, 2024: 1-5.

[15] Tang F H, Ding J R, Quan Q, et al. CMUNEXT: An efficient medical image segmentation network based on large kernel and skip fusion[C]//2024 IEEE International Symposium on Biomedical Imaging (ISBI). Athens: IEEE, 2024: 1-5.

[16] Fu H Z, Li F, Orlando J I, et al. PALM: Pathologic myopia challenge[DB/OL]. IEEE Dataport, 2019. [2025-09-20]. https://doi.org/10.21227/55pk-8z03.

[17] Rauf N, Gilani S O, Waris A. Automatic detection of pathologic myopia using machine learning[J]. Scientific Reports, 2021, 11: 16570.

[18] 符灵利. 基于深度学习的眼底图像血管分割与病理性近视检测研究[D]. 武汉:华中师范大学, 2023.

[19] Rahman M M, Munir M, Marculescu R. EMCAD: Efficient multi-scale convolutional attention decoding for medical image segmentation[C]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle:IEEE, 2024: 11769-11779.

[20] Li H, Zhai D H, Xia Y Q. ERDUnet: An efficient residual double-coding Unet for medical image segmentation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(4): 2083-2096.

(编 辑 李 静)