

泛化增强与动态感知的结直肠息肉分割网络

王 森^{1,2}, 史彩娟^{1,2}, 蔡 澳^{1,2}, 王 睿^{1,2}, 于鑫阳^{1,2}, 程旭东^{1,2}, 陈伟彬³

(1. 华北理工大学人工智能学院, 唐山 063210; 2. 河北省工业智能感知重点实验室, 唐山 063210; 3. 华北理工大学附属医院, 唐山 063000)

摘 要: 随着计算机辅助医疗诊断系统和医学图像分割技术的快速发展, 结直肠镜检查性能得到了极大的提升, 可有效帮助临床医生对息肉病变作出快速准确的判断并制定治疗方案。然而, 在临床实践中, 息肉分割面临众多挑战, 如不同患者的息肉所处肠道环境不同, 息肉大小不同、形状各异等。为了应对这些挑战, 提升结直肠息肉分割模型的泛化能力和学习能力, 提出了一种泛化增强与动态感知网络 (Generalization enhancement and dynamic perception network, GEDPNet)。GEDPNet 使用金字塔视觉 Transformer (PVT_v2) 作为主干, 重点设计了泛化增强 (Generalization enhancement, GE) 模块、动态感知 (Dynamic perception, DP) 模块和级联聚合 (Cascade aggregation, CA) 模块。首先, GE 模块创新性地从提取息肉域不变特征的角度来提升模型的泛化性, 从而有效缓解不同患者的息肉所处肠道环境不同导致的分割性能不佳问题; 同时, 该模块还通过提取丰富的层内多尺度信息来应对息肉尺寸多样化的挑战。其次, DP 模块能够有效地动态感知全局信息和局部信息, 捕获息肉的语义位置信息和边界纹理等细节信息。最后, CA 模块将不同层级的含有不同语义信息的多尺度特征有效聚合, 保证息肉信息的完整性, 进一步提升了分割性能。所提 GEDPNet 模型在 5 个息肉数据集 Kvasir-SEG、CVC-ClinicDB、CVC-T、CVC-ColonDB 和 ETIS 上进行了测试, mDice 分别达到 0.930、0.946、0.911、0.825 和 0.806; mIoU 分别达到 0.883、0.902、0.848、0.747 和 0.733; MAE 分别达到 0.019、0.005、0.005、0.025 和 0.013。此外, 所提方法与 20 种经典及先进的息肉图像分割方法进行了性能比较, 比经典息肉分割方法 PraNet 的 mIoU 分别提高了 4.3%、5.3%、5.1%、10.7% 和 16.6%。结果表明, 本文所提的 GEDPNet 对肠道环境差异大、尺寸不一及形状各异的息肉具有较好的感知能力, 可有效提升模型的息肉分割精度和泛化能力。

关键词: 图像分割; 息肉分割; 泛化增强; 动态感知; 多尺度

中图分类号: TP391

文献标志码: A

Generalization Enhancement and Dynamic Perception of Colorectal Polyp Segmentation Network

WANG Sen^{1,2}, SHI Caijuan^{1,2}, CAI Ao^{1,2}, WANG Rui^{1,2}, YU Xinyang^{1,2}, CHENG Xudong^{1,2}, CHEN Weibin³

(1. School of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China; 2. Hebei Key Laboratory of Industrial Intelligent Perception, Tangshan 063210, China; 3. The Affiliated Hospital of North China University of

基金项目: 唐山市人才项目 (A202110011); 北京市现代信息科学与网络技术重点实验室开放课题 (XDXX2301); 华北理工大学杰出青年基金 (JQ201715)。

收稿日期: 2024-08-14; **修订日期:** 2025-01-09

Science and Technology, Tangshan 063000, China)

Abstract: With the rapid development of computer-aided medical diagnosis systems and medical image segmentation technologies, the performance of colorectal endoscopy has been significantly improved, effectively helping clinicians make quick and accurate judgments on polyp lesions and formulate appropriate treatment plans. However, in clinical practice polyp segmentation faces numerous challenges, such as different intestinal environments in different patients, and varying sizes and shapes of polyps. To address these challenges and enhance the generalization and learning abilities, the generalization enhancement and dynamic perception network (GEDPNet) is proposed. GEDPNet utilizes the pyramid vision Transformer (PVT_v2) as its backbone and focuses on the design of three key modules: the generalization enhancement (GE) module, the dynamic perception (DP) module, and the cascade aggregation (CA) module. Firstly, the GE module innovatively improves the model's generalization by extracting the polyp domain-invariant features, effectively alleviating the problem of poor segmentation caused by different intestinal environments of polyps in different patients. Meanwhile, the GE module also addresses the challenge of diverse polyp sizes by extracting rich multi-scale information intra each layer. Secondly, the DP module is able to dynamically perceive the global and local information, and then to effectively capture the position information as well as the boundaries or textures of polyps. Finally, the CA module can fully aggregate multi-scale features at different levels to obtain rich semantic information, ensuring the integrity of polyp information and further enhancing segmentation performance. To verify the effectiveness of the proposed GEDPNet, extensive experiments are conducted on five polyp datasets: Kvasir-SEG, CVC-ClinicDB, CVC-T, CVC-ColonDB, and ETIS. On these five polyp datasets, the mDice of the proposed GEDPNet achieves 0.930, 0.946, 0.911, 0.825, and 0.806, respectively; mIoU of that achieves 0.883, 0.902, 0.848, 0.747, and 0.733, respectively; and MAE of that achieves 0.019, 0.005, 0.005, 0.025, and 0.013, respectively. Furthermore, the proposed GEDPNet has been compared with 20 classical and advanced polyp image segmentation methods and it outperforms nearly all of them. Notably, mIoU of GEDPNet has been improved by 4.3%, 5.3%, 5.1%, 10.7%, and 16.6% respectively on these five polyp datasets compared to that of classical polyp segmentation method PraNet. These results indicate that the proposed GEDPNet exhibits superior dynamic perception capabilities for polyps with significant variations in intestinal environments, sizes, and shapes, so it can effectively enhance the polyp segmentation accuracy and model's generalization.

Key words: image segmentation; polyp segmentation; generalization enhancement (GE); dynamic perception (DP); multi-scale

引 言

结直肠癌(Colorectal cancer, CRC)是全球第三大常见癌症,在癌症致死率中排名第二。结直肠癌通常从息肉病变发展而来,通过结肠镜检查及早发现息肉病变,能够有效降低结直肠癌的发病率,从而挽救更多的生命。传统手工结肠镜检查高度依赖于医生的经验水平,且费时费力^[1]。近年,计算机辅助医疗诊断及医学图像分割技术快速发展,结肠镜检查效果得到了极大提升,使得息肉分割的准确度不断提高,从而帮助临床医生对息肉病变作出快速准确的判断,并制定进一步的治疗方案。

近年,卷积神经网络(Convolutional neural network, CNN)被广泛应用于医学图像分割,如 Ronneberger 等^[2]提出的 U-Net 将医学图像分割推向了新的高度。此后,各种基于 U-Net 网络的变体算法不

断被提出,例如 U-Net++^[3]、ResUNet++^[4]等。这些算法为息肉分割提供了编解码模型设计思路, PraNet^[5]、UACANet^[6]、CaraNet^[7]、HRNetED^[8]、MBANet^[9]等息肉分割模型相继出现,使得息肉分割性能不断提升。虽然 CNN 能有效提取息肉的局部信息,但对于全局信息的提取还非常有限。

随着 Transformer 在计算机视觉领域的发展与应用,视觉 Transformer (Vision Transformer, ViT)^[10] 因其出色的全局信息获取能力而被广泛应用于分割任务。为了获取更为丰富的多尺度信息, Wang 等^[11]在 ViT 的基础上提出了金字塔视觉 Transformer (Pyramid vision Transformer, PVT_v2) 主干网络。随后, PVT_v2 被广泛用于结直肠息肉分割,多种基于 PVT_v2 的模型,如 Polyp-PVT^[12]、SSFormer^[13]、PVT-Cascade^[14]和 PVT-PMFFD^[15]等被提出。尽管现有基于 PVT_v2 的息肉分割方法取得了不错的性能,但由于息肉所处环境差异大,尺寸和形状各异等问题,分割性能还比较有限。

针对上述问题, Yue 等^[16]提出了 BRNet, 主要设计边界生成模块和边界引导细化模块来提升模型的学习和泛化能力。Cai 等^[17]提出了 CCFNet, 有效地利用跨层次和多尺度的上下文信息, 设计高级特征级联模块、跨层集成模块和全局上下文增强模块来改善模型的学习和泛化能力。上述方法虽然一定程度上提升了模型的泛化能力, 但针对于不同患者肠道内部环境差异大, 以及图像采集中光照、方向和位置对息肉成像影响等问题, 没有从提取息肉本质特征(即域不变特征^[18])的角度来有效提升模型的泛化性。另外, Li 等^[19]提出了 CIFG-Net, 通过设计跨层信息处理模块来增强 PVT_v2 提取的多层多尺度特征, 增强各层特征间的关联性, 融合得到更为细致的特征, 从而提高息肉分割准确率。吴港等^[20]提出了 FTDC-Net, 通过扩大模型的感受野使模型捕捉到结肠镜图像中更大范围内的多尺度信息。Xiao 等^[21]提出了 CTNet, 通过设计的对比 Transformer 主干获得长距离依赖关系的多尺度特征, 设计自多尺度交互模块有效获取多尺度信息, 并设计信息收集模块以获得具有高级语义信息的高分辨率特征图, 从而使模型能够在不同尺度上很好地分割息肉。这 3 种方法虽然利用不同层级特征获得了丰富的多尺度信息, 在一定程度上提升了对不同尺度息肉的分割性能, 但也仍然存在一些不足: (1) 忽略了同一层级特征的层内多尺度信息; (2) 在解码器逐层恢复分辨率的过程中, 特征融合时并不能动态地感知全局信息和局部信息, 且忽略了息肉目标本身较大的形状差异。

因此, 针对息肉环境差异大和尺寸形状各异等问题, 本文提出了一种息肉分割模型, 即泛化增强与动态感知网络 (Generalization enhancement and dynamic perception network, GEDPNet), 通过提取息肉的域不变特征来提升模型的泛化性, 同时提取丰富的层内多尺度信息, 以及动态感知全局信息和局部信息, 来进一步提升模型分割息肉的精度。GEDPNet 以 PVT_v2 为主干, 着重设计了泛化增强 (Generalization enhancement, GE) 模块、动态感知 (Dynamic perception, DP) 模块和级联聚合 (Cascade aggregation, CA) 模块。GE 模块通过设计的域不变特征提取子模块有效提取息肉域不变特征, 增强模型的泛化能力, 同时能够有效关注层内多尺度信息; DP 模块利用全局感知子模块和局部与空间感知子模块来动态感知精准的全局信息和局部信息, 避免层级差异; CA 模块通过有效聚合不同层级的多尺度特征, 充分捕获息肉的语义位置和细节纹理等信息。所提模型 GEDPNet 在多个息肉数据集上进行了丰富的实验, 实验结果验证了所提模型优于其他经典以及最新的模型, 取得了良好的息肉分割性能。

总体来说, 本文的主要贡献总结如下:

(1) 提出一种新的息肉分割方法 GEDPNet, 创新地提取息肉域不变特征来提升模型的泛化性; 同时通过获取丰富的层间和层内多尺度信息, 动态感知全局和局部信息, 进一步提升息肉分割精度。

(2) 设计 GE 模块, 通过在其内部设计域不变特征提取子模块来提取息肉域不变特征, 增强模型的泛化性, 同时有效获取层内多尺度信息;

(3) 设计 DP 模块来动态感知全局和局部信息, 设计 CA 模块获取丰富的层间多尺度信息, 进而捕获息肉的语义和细节信息;

(4)在多个息肉数据集上进行实验,实验结果表明GEDPNet在息肉分割任务上具有出色的学习能力和泛化能力。

1 本文方法

面对息肉在不同患者肠道内所处环境差异大、尺寸多变且形状各异等困难,为进一步提升息肉分割的性能,本文提出GEDPNet,其结构如图1所示。

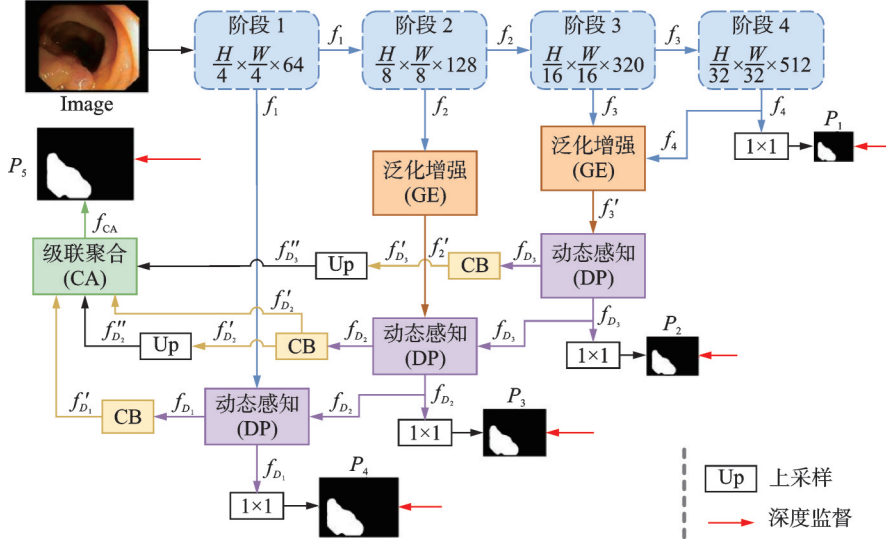


图1 泛化增强与动态感知网络结构图

Fig.1 Structural diagram of generalization enhancement and dynamic perception network (GEDPNet)

GEDPNet模型主要由PVT_v2主干和解码器组成,其中着重设计了解码器中的GE、DP和CA模块。GEDPNet模型具体工作过程如下:输入一幅息肉图像 $I \in \mathbb{R}^{H \times W \times C}$,使用带有预训练权重的主干PVT_v2进行特征提取,获得4个不同层级的多尺度原始特征 $f_i \in (H \times W)/2^{i+1} \times C_i (i=1, 2, 3, 4)$,其中 H 和 W 分别为输入图像的高度和宽度, C_i 为第 i 层的通道数。将原始特征 f_2 与 f_3 分别输入到两个GE模块,分别实现层内多尺度特征增强,并且提取该层的息肉域不变特征,得到多尺度泛化增强特征 $f_i' (i=2, 3)$ 。接着,原始特征 f_1, f_4 和多尺度泛化增强特征 f_2', f_3' 分别输入到3个DP模块,动态感知精准的全局和局部信息并将其进行有效融合,避免由层级语义差异导致的信息丢失,得到感知融合特征 $f_{D_i} (i=1, 2, 3)$ 。然后,利用 3×3 卷积和批归一化(Convolution and batch normalization, CB)对感知融合特征 $f_{D_i} (i=1, 2, 3)$ 进行通道对齐和特征规范,得到规范对齐特征 $f_{D_i}' (i=1, 2, 3)$,其中 f_{D_2}' 和 f_{D_3}' 再分别经过上采样进行空间尺度对齐,得到空间尺度对齐特征 f_{D_2}'' 和 f_{D_3}'' 。最后,将特征 $f_{D_1}', f_{D_2}', f_{D_3}', f_{D_2}'', f_{D_3}''$ 输入到CA模块进行特征聚合,充分聚合不同层级的含有不同语义信息的多尺度特征,充分捕获息肉语义及细粒度信息,得到级联聚合特征 f_{CA} 。

综上,本文重点设计的GE、DP和CA模块分别从3个角度对特征进行增强,在三者共同协作下能够有效提升息肉分割性能。首先,GE模块能够提取息肉域不变特征和层内多尺度特征,为DP模块提供包含丰富上下文信息的多尺度泛化增强特征;然后,DP模块利用全局感知子模块和局部与空间感知子模块来动态感知GE模块提供的特征中的全局信息,以及包含边界和纹理细节的局部信息;最后,CA模块有效融合不同层级DP模块输出的具有不同的空间分辨率和语义信息的多尺度特征,从而充分捕

获息肉的语义位置和纹理细节,提升模型的息肉分割性能。

1.1 泛化增强模块

由于每个患者肠道环境差异大,且采集息肉图像时受光线、角度等因素影响,导致息肉分割困难;另外,结直肠息肉大小不同、形状各异,尽管采用PVT_v2主干获得了4层多尺度特征,但是每一层的层内特征仅有一个尺度,不足以应对尺寸多变的息肉目标。因此,本文设计GE模块,提取域不变特征和丰富的层内多尺度信息,有效提升模型的泛化能力和分割精度。GE模块具体结构如图2所示。

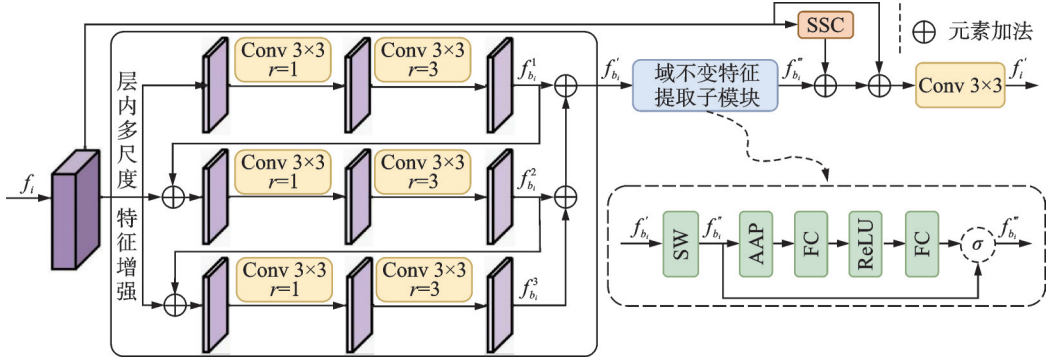


图2 泛化增强模块结构图

Fig.2 Structural diagram of generalization enhancement (GE) module

具体工作过程如下:

首先,进行层内多尺度特征增强,采用3个并行分支来处理原始特征 $f_i (i=2,3)$,每个分支经过不同空洞率的卷积,且3个分支间相加连接,从而使每一层特征获得更为丰富的层内多尺度信息。具体来说,每个分支先分别经过1个 3×3 的常规卷积,再经过1个 3×3 的空洞卷积(空洞率分别为1、3和5),得到第 i 层中层内多尺度特征增强的第 j 个分支的输出 $f_{b_i}^j (j=1,2,3)$,所有卷积均进行补零操作,维持特征图尺寸。最后,将3个分支结果逐像素相加,获得层内多尺度增强特征 $f_{b_i}' (i=2,3)$ 。 $f_{b_i}^j (j=1,2,3)$ 和 $f_{b_i}' (i=2,3)$ 的获取可以分别表示为

$$f_{b_i}^j = \begin{cases} \text{Conv}_{3 \times 3, r=1}(\text{Conv}_{3 \times 3}(f_i)) & j=1 \\ \text{Conv}_{3 \times 3, r=3}(\text{Conv}_{3 \times 3}(f_i \oplus f_{b_i}^1)) & j=2 \\ \text{Conv}_{3 \times 3, r=5}(\text{Conv}_{3 \times 3}(f_i \oplus f_{b_i}^2)) & j=3 \end{cases} \quad (1)$$

$$f_{b_i}' = \sum_{j=1}^3 f_{b_i}^j \quad (2)$$

式中: $\text{Conv}_{3 \times 3}(\bullet)$ 表示 3×3 卷积, r 表示扩张率。

其次,为了提升模型对不同患者肠道环境中的息肉分割的泛化能力,同时去除层内多尺度增强特征中包含的冗余和噪声,设计了域不变特征提取(Domain invariant feature extraction, DIFE)子模块。DIFE采用可切换白化(Switchable whitening, SW)方法和压缩与激励的操作,去除特征中的冗余和相关性,从而提取息肉域不变特征。SW内部包含5种归一化和白化方法,分别是批归一化、实例归一化、层归一化、批白化和实例白化,通过动态调整这些归一化和白化方法的权重来适应不同的任务。在面对不同域时,SW更多地选择实例白化和实例归一化来减少源域和目标域之间的差异,从而有效地应对不同患者肠道内部的息肉外观差异。实例白化针对单个样本进行操作,计算每个样本的均值和协方差矩

阵,并利用这些统计量对特征进行白化处理,旨在去除单个样本内部特征之间的相关性^[18]。

DIFE子模块的具体工作过程如下:层内多尺度增强特征 f'_{b_i} 经过可切换白化,获得去除相关性的白化特征 $f''_{b_i}(i=2,3)$,再经过一个压缩与激励的通道权重分配过程,以保留更重要的信息,从而获得域不变特征 $f'''_{b_i}(i=2,3)$,该过程可以表示为

$$f''_{b_i} = \text{SW}(f'_{b_i}) \quad (3)$$

$$f'''_{b_i} = f''_{b_i} \odot \sigma(\text{FC}(\text{R}(\text{FC}(\text{AAP}(f''_{b_i})))) \quad (4)$$

式中: $\text{SW}(\cdot)$ 表示可切换白化,“ $\odot$ ”表示元素乘法, $\text{AAP}(\cdot)$ 表示自适应平均池化, $\text{FC}(\cdot)$ 表示全连接层, $\text{R}(\cdot)$ 表示ReLU激活函数, $\sigma(\cdot)$ 表示Sigmoid函数。

最后,采用特征空间和通道重构注意力 $\text{SSConv}^{[22]}$ 对原始特征 $f_i(i \in [2,3])$ 进行增强和噪声抑制,将其结果与域不变特征 f'''_{b_i} 相加,同时采用原始特征做残差连接补充整体信息,避免信息丢失,最终经过一个 3×3 卷积输出多尺度泛化增强特征 $f'_i(i \in [2,3])$,该过程可以表示为

$$f'_i = \text{Conv}_{3 \times 3}(f'''_{b_i} + \text{SSConv}(f_i) + f_i) \quad (5)$$

式中: $\text{SSConv}(\cdot)$ 表示特征空间和通道重构注意力操作。

1.2 动态感知模块

在解码器逐层恢复分辨率的过程中,需要融合原始特征 f_1, f_4 和多尺度泛化增强特征 f'_2, f'_3 来保证输入图像信息的完整。但是,不同层级间存在信息差异,直接融合容易造成信息丢失。因此,本文设计了DP模块,其中设计了全局感知(Global perception, GP)子模块和局部与空间感知(Local and spatial perception, LSP)子模块来动态感知精准的全局信息和局部信息。设计基于Transformer的GP子模块和基于CNN的LSP子模块来分别获取长期依赖关系和边界及纹理细节信息,同时捕获小息肉目标。DP整体结构如图3所示。

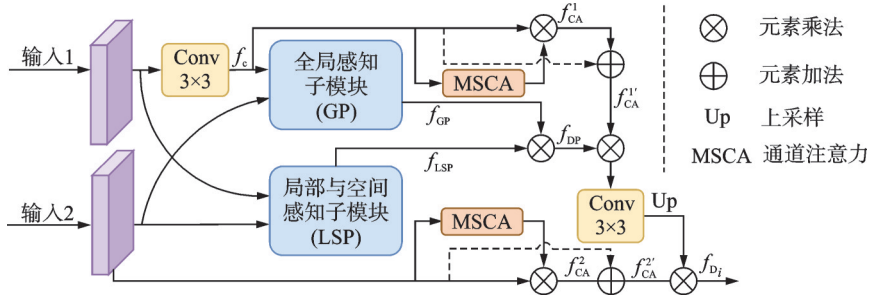


图3 动态感知模块结构图

Fig.3 Structural diagram of dynamic perception (DP) module

具体工作过程如下:DP模块有两个输入。首先,输入1经过 3×3 卷积对齐通道,得到通道对齐特征 f_c ,将其与输入2同时输入到GP子模块,得到全局感知特征 f_{GP} ;输入1和输入2同时直接输入到LSP子模块,得到局部与空间感知特征 f_{LSP} ;最后将 f_{GP} 与 f_{LSP} 相乘得到特征 f_{DP} 。其次, f_c 经过通道注意力模块 $\text{MSCA}^{[23]}$,被赋予新的通道权重后得到特征 f_{CA}^1 ,用 f_c 做残差连接与 f_{CA}^1 相加,得到特征 f_{CA}^1' 。同时,输入2也经过MSCA模块,被赋予新的通道权重后得到特征 f_{CA}^2 ,再用输入2做残差连接与 f_{CA}^2 相加,得到特征 f_{CA}^2' 。最后,特征 f_{DP} 与 f_{CA}^1' 相乘后经过卷积和上采样,再与 f_{CA}^2' 相乘,最终得到DP模块输出的感知融合特征 $f_{D_i}(i=1,2,3)$ 。

1.2.1 全局感知子模块

为了使两个特征图在融合时能更加关注全局信息,获得长距离依赖关系,设计了GP子模块,其整体结构如图4所示。GP利用交叉注意力在两个不同输入之间建立双向的关联和信息传递,并能动态地调整特征图之间的信息传递权重,从而使融合后的特征更加全面和精准。

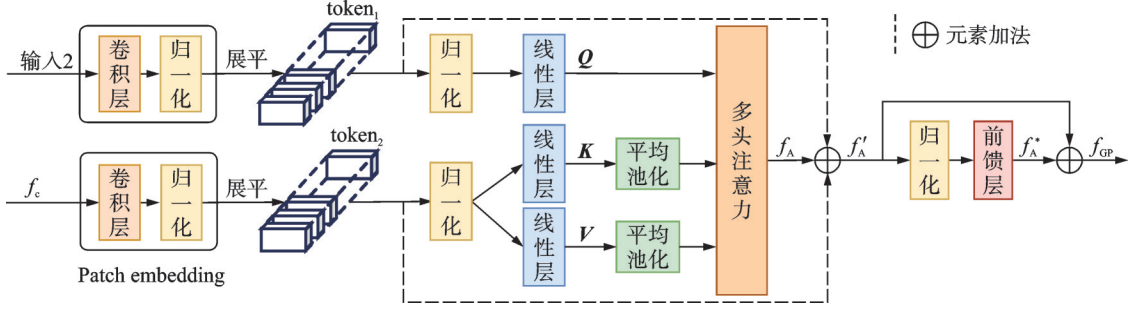


图4 全局感知子模块结构图

Fig.4 Structural diagram of global perception (GP) sub-module

具体工作过程如下：

首先,使用卷积层和归一化对两个输入分别分块后再展平,得到具有序列化的 token_1 和 token_2 。然后,对 token_1 和 token_2 分别进行归一化,并输入到线性层进行线性映射,计算出查询矩阵 Q (Query),键矩阵 K (Key)和值矩阵 V (Value)。不同于传统自注意力机制中 Q 、 K 和 V 都来自相同的输入,本文中的 Q 由输入2计算得到, K 和 V 由通对齐特征 f_c 计算得到。

为了减少计算复杂度,对 K 和 V 分别进行平均池化^[11],然后与 Q 一起输入到多头注意力中,使输入2中的每个位置关注通道对齐特征 f_c 的所有位置,有效融合,从而得到注意力特征 f_A 。 Q 和 K 用于注意力权重矩阵计算,并与 V 相乘,则多头注意力的计算过程可表示为

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

式中 d_k 为键向量的维度,用于缩放点积。

接着,注意力特征 f_A 与 token_1 和 token_2 相加之后得到 f'_A ,将其归一化后输入到前馈层得到前馈特征 f''_A ,该求取过程可表示为

$$f''_A = \text{FC}_2(\text{DWC}(\text{G}(\text{FC}_1(f'_A)))) \quad (7)$$

式中: $\text{DWC}(\bullet)$ 表示深度卷积, $\text{FC}_1(\bullet)$ 和 $\text{FC}_2(\bullet)$ 表示线性全连接映射, $\text{G}(\bullet)$ 表示GELU激活函数。

最后,将注意力特征 f'_A 和前馈特征 f''_A 相加,得到GP子模块输出,即全局感知特征 f_{GP} ,此过程可表示为

$$f_{GP} = f'_A + f''_A \quad (8)$$

1.2.2 局部与空间感知子模块

为了更好地捕捉到息肉目标的几何形状、局部细节信息以及小息肉目标,提高特征表示的灵活性,本文利用可变形卷积设计了LSP子模块,其结构如图5所示。

具体工作过程如下：

首先,输入1经过 3×3 卷积增强局部建模能力,同时提取偏移量Offsets;然后,输入1和输入2同时利用共享的偏移量Offsets进行可变形卷积,相比于常规卷积能够更准确地拟合息肉形状,图6给出了常

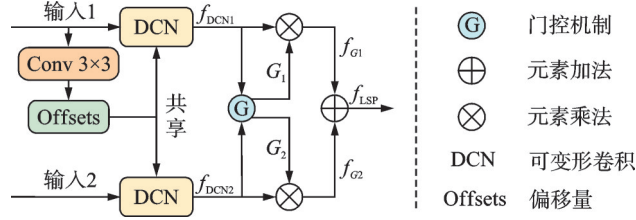


图5 局部与空间感知子模块结构图

Fig.5 Structural diagram of local and spatial perception(LSP)sub-module

规卷积与可变形卷积对息肉拟合的对比图,可变形卷积控制步长同时将两个特征图的尺寸对齐,分别得到空间感知特征 f_{DCN1} 和 f_{DCN2} ,偏移量Offsets的共享能让两个不同层特征关注的点更加相似;之后,将 f_{DCN1} 和 f_{DCN2} 经过门控机制,通过平均池化、 1×1 卷积以及Sigmoid函数赋予两个分支不同的权重 G_1 和 G_2 ;最后,用所得权重对 f_{DCN1} 和 f_{DCN2} 进行加权求和取平均,最终得到LSP子模块的输出,即局部与空间感知特征 f_{LSP} 。上述过程可表示为

$$\text{Offsets} = \text{Conv}_{3 \times 3}(\text{输入1}) \quad (9)$$

$$f_{DCN2} = \text{DCN}(\text{offsets}, \text{输入2}) \quad (10)$$

$$G_{1,2} = \sigma(\text{Conv}_{1 \times 1}(\text{AAP}(f_{DCN1,2}))) \quad (11)$$

$$G_{1,2} = \sigma(\text{Conv}_{1 \times 1}(\text{AAP}(f_{DCN1,2}))) \quad (12)$$

$$f_{LSP} = \frac{G_1 \cdot f_{DCN1} + G_2 \cdot f_{DCN2}}{2} \quad (13)$$

式中: $\text{Conv}_{1 \times 1}$ 表示 1×1 卷积,DCN($\bullet$)表示可变形卷积。

1.3 级联聚合模块

为了有效融合不同层级的具有不同的空间分辨率和语义信息的多尺度特征,本文设计了CA模块,充分捕获息肉的语义位置和纹理细节等信息,保证信息的完整性来进一步提升模型分割性能。CA整体流程如图7所示。

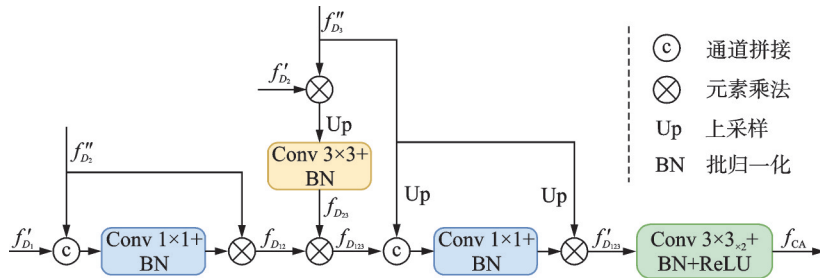
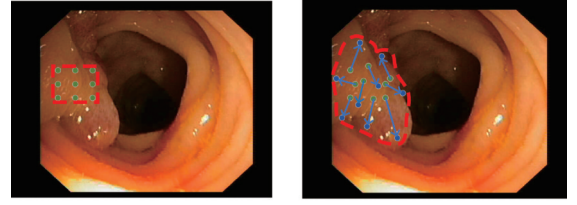


图7 级联聚合模块结构图

Fig.7 Structural diagram of cascade aggregation (CA) module



(a) Ordinary Convolution (b) Deformable Convolution

图6 常规卷积与可变形卷积对息肉拟合的对比图

Fig.6 Comparison of polyp fitting between ordinary and deformable convolutions

具体工作过程如下:

CA 模块接收经过通道对齐且规范后的特征 f'_{D_1} 、 f'_{D_2} 和 f'_{D_3} 以及再经过空间对齐的特征 f''_{D_2} 和 f''_{D_3} 。首先 f'_{D_2} 与 f'_{D_1} 进行拼接、卷积、相乘操作得到低层融合特征 $f_{D_{12}}$, 因为低层特征富含的细节信息较多, 但同时存在更多的噪声, 这些操作能尽可能地捕获 f'_{D_1} 与 f'_{D_2} 在空间和通道上的有效细节特征并去除不相关的噪声; 其次, 特征 f'_{D_2} 与 f''_{D_3} 相乘, 并经过 3×3 卷积去除噪声得到高层融合特征 $f_{D_{23}}$, 最大限度地保留了高层特征中的语义信息, 再与低层融合特征 $f_{D_{12}}$ 相乘得到初步聚合特征 $f_{D_{123}}$; 最后用噪声最少的高层特征 f''_{D_3} 与 $f_{D_{123}}$ 进行拼接、卷积和相乘操作得到高层指导特征 $f'_{D_{123}}$, 用高层特征指导去除初步聚合特征的冗余信息, 实现全局和局部信息的交互, 在不增加干扰噪声的同时, 全面获得息肉整体语义信息和边缘纹理等细节信息; 最后用两个 3×3 的卷积和 BN 以及 ReLU 激活函数, 进一步过滤聚合后的特征, 减少冗余并规范特征数值, 得到最后的输出 f_{CA} 。上述过程可表示为

$$f_{D_{12}} = f'_{D_2} \odot \text{Conv}_{1 \times 1}([\![f'_{D_2}, f'_{D_1}]\!]) \quad (14)$$

$$f_{D_{23}} = f_{D_{12}} \odot \text{Conv}_{3 \times 3}(\text{Up}(f'_{D_2} \odot f''_{D_3})) \quad (15)$$

$$f'_{D_{123}} = f''_{D_3} \odot \text{Conv}_{3 \times 3}([\![f_{D_{123}}, \text{Up}(f''_{D_3})]\!]) \quad (16)$$

$$f_{CA} = \text{ReLU}(\text{BN}(\text{Conv}_{3 \times 3 \times 2}(f'_{D_{123}}))) \quad (17)$$

式中: $\text{Up}(\cdot)$ 表示上采样, $[\![\cdot]\!]$ 表示按通道维度拼接, $\odot$ 表示元素乘法。

综上可知, 级联聚合模块能够有效聚合富含语义位置信息及细节信息的不同层级的多尺度特征。

1.4 损失函数

本文采用息肉分割领域常用的损失函数组合, 包括基于全局约束的加权交并比损失 L_{IOU}^w 和基于局部(像素级)约束的加权二元交叉熵损失 L_{BCE}^w 。对 PVT_v2 最高层预测输出 P_1 、DP 模块的 3 个预测输出 $P_i (i \in [2, 3, 4])$ 以及 CA 模块输出的最终预测图 P_5 进行深度监督, 每个预测图都上采样到与真值图 (Ground truth, GT) 相同的大小。因此, 本文所提模型 GEDPNet 的总损失函数 L_t 可表示为

$$L_t = \sum_{i=1}^5 L(P_i, \text{GT}) \quad (18)$$

式中: GT 表示真值图, L 表示混合损失函数, 可定义为

$$L = L_{\text{IOU}}^w + L_{\text{BCE}}^w \quad (19)$$

2 实验与结果分析

本节对所提 GEDPNet 模型在 5 个数据集上进行了实验验证。首先, 对所用数据集、评估指标和实验设置等进行简单介绍, 然后着重对实验结果和消融实验等进行分析。

2.1 数据集及评估指标

为了评估所提模型 GEDPNet 的学习能力和泛化能力, 本文在 Kvasir-SEG^[24]、CVC-ClinicDB^[25]、CVC-ColonDB^[26]、ETIS^[27] 和 CVC-T^[28] 五个公开的结肠镜息肉分割数据集上对所提模型性能进行实验验证和分析, 这些数据集提供了多样化且具有代表性的结直肠息肉图像样本。本文遵循与 PraNet^[5] 相同的数据集设置, 表 1 给出了 5 个数据集的描述信息和相应的训练、测试划分情况。

实验采用 3 个广泛应用的评价指标来评估模型的分割性能, 包括: 平均 Dice 系数 (mean Dice coefficient, mDice)、平均交并比 (mean Intersection over union, mIoU) 和平均绝对误差 (Mean average error, MAE)。mDice 用来衡量预测分割与真实分割之间的相似度, 其值范围在 0 到 1 之间; mIoU (也称为 Jaccard 指数) 计算预测区域与真实区域的交集与并集之比, 同样取值于 0 到 1 之间; MAE 评估预测分割图与真

值图之间的像素级精度,它计算预测值与真实值之间的平均绝对差异,MAE值越小,表示分割越精确。

表 1 数据集细节
Table 1 Details of datasets

数据集	图片尺寸/(像素 $\times$ 像素)	样本总数/张	训练样本/张	测试样本/张
Kvasir-SEG	332 $\times$ 487~1 920 $\times$ 1 072	1 000	900	100
CVC-ClinicDB	384 $\times$ 288	612	550	62
CVC-T	574 $\times$ 500	60	—	60
CVC-ColonDB	574 $\times$ 500	380	—	380
ETIS	1 225 $\times$ 966	196	—	196

注:“—”表示该数据集不用于训练。

2.2 实验设置

本文实验在 NVIDIA RTX 3090 GPU (24 GB)上进行,CUDA 版本为 11.8。所提模型基于 PyTorch 框架开发,并使用在 ImageNet 上预训练的 PVT_v2 作为整个网络的主干。训练过程中采用 Adam 优化算法,学习率最初设置为 5×10^{-5} ,分别在 30、50、80 个迭代时以 0.5 倍下降,批大小设置为 8,共迭代 150 次。与 PraNet^[5]的设置一致,实验中所有输入图像尺寸被设置为 352 像素 $\times$ 352 像素,采用多尺度训练策略 {0.75, 1, 1.25},无其他数据增强手段。

2.3 定量分析

医学图像分割任务不仅关注模型的学习能力,同时也关注模型的泛化能力,从而更好地应用于临床实践。因此,为了更加全面地验证所提模型的有效性,与文献[20]方法类似,对 GEDPNet 及相关模型的学习能力和泛化能力分别进行了验证分析。比较的相关模型算法共 20 个,包括 U-Net^[2]、PraNet^[5]、SANet^[29]、UACANet-S^[6]、MSNet^[30]、BDG-Net^[31]、CaraNet^[7]、ECTransNet^[32]、HarDNet-CPS^[33]、CFANet^[34]、RA-DENet^[35]、PPNet^[36]、ConvMLPSeg^[37]、CCFNet^[17]、NPD-Net^[38]、IC-PolypSeg-EF7^[39]、MEGANet^[40]、BRNet^[16]、RTA-Former-L^[41]和 CIFG-Net^[19]。

2.3.1 学习能力分析

为了验证所提模型 GEDPNet 的学习能力,本实验在 Kvasir-SEG 和 CVC-ClinicDB 数据集上进行。数据集被划分为训练集和测试集,其中测试集属于模型学习能力测试样本。实验中所提方法 GEDPNet 与其他先进方法的比较结果如表 2 所示,其中比较方法的指标数据均来自原文。

由表 2 可知:(1)GEDPNet 在 3 个指标上均获得了最优性能,如在 Kvasir-SEG 上,GEDPNet 的 mDice、mIoU 和 MAE 分别达到了 0.930、0.883 和 0.019;在 CVC-ClinicDB 上,其 mDice、mIoU 和 MAE 分别达到了 0.946、0.902 和 0.005。(2)所提 GEDPNet 与其他方法相比,性能基本有较大提升,如在 Kvasir-SEG 上与性能次优方法 CIFG-Net 相比,mIoU 提升了 0.7%;在 CVC-ClinicDB 上与 BRNet 相比,mIoU 提升了 2.4%,但是二者的 MAE 相同。(3)在这两个数据集上,GEDPNet 的 mDice 比经典医学图像分割方法 U-Net 的 mDice 均高出 10% 以上,比经典息肉图像分割方法 PraNet 的 mIoU 均高出 4% 以上。(4)从 GEDPNet 的各指标标准差可以看出,如两个数据集上 MAE 的标准差分别为 0.021 和 0.01,表明所提模型具有相对稳定的分割精度。

实验结果表明,本文所提模型 GEDPNet 相比于其他方法在学习能力上有着明显优势,这主要归因于 GEDPNet 中本文设计的 3 个模块:GE 模块赋予了模型更多的多尺度信息,适应尺寸多变的息肉;DP 模块能够动态感知精准的全局信息和局部信息,从而有效获取息肉的整体定位和细节信息;CA 模块能

表2 不同模型学习能力比较

Table 2 Comparison of learning ability for different models

模型	年份	Kvasir-SEG			CVC-ClinicDB		
		mDice ↑	mIoU ↑	MAE ↓	mDice ↑	mIoU ↑	MAE ↓
U-Net ^[2]	2015	0.818	0.746	0.055	0.823	0.755	0.019
PraNet ^[5]	2020	0.898	0.840	0.030	0.899	0.849	0.009
SANet ^[29]	2021	0.904	0.847	—	0.916	0.859	—
UACANet-S ^[6]	2021	0.905	0.852	0.026	0.916	0.870	0.008
MSNet ^[30]	2021	0.907	0.862	0.028	0.921	0.879	0.008
BDG-Net ^[31]	2022	0.915	0.865	0.021	0.916	0.864	—
CaraNet ^[7]	2022	0.918	0.865	0.023	0.936	0.887	0.007
ECTransNet ^[32]	2023	0.901	0.847	—	0.923	0.878	—
HarDNet-CPS ^[33]	2023	0.911	0.856	0.025	0.917	0.887	0.008
CFANet ^[34]	2023	0.915	0.861	0.023	0.933	0.883	0.007
RA-DENet ^[35]	2023	0.918	0.870	0.021	0.921	0.865	0.011
PPNet ^[36]	2023	0.920	0.874	0.024	0.921	0.878	0.008
ConvMLPSeg ^[37]	2023	0.920	0.869	—	0.924	0.870	—
CCFNet ^[17]	2024	0.910	0.855	0.024	0.920	0.870	0.007
NPD-Net ^[38]	2024	0.910	0.855	0.026	0.928	0.878	0.008
IC-PolypSeg-EF7 ^[39]	2024	0.910	0.859	0.026	0.938	0.890	0.007
MEGANet ^[40]	2024	0.913	0.863	0.025	0.938	0.894	0.006
BRNet ^[16]	2024	0.918	0.871	0.026	0.921	0.878	0.005
RTA-Former-L ^[41]	2024	0.923	0.875	—	0.938	0.888	—
CIFG-Net ^[19]	2024	0.925	0.876	0.021	0.938	0.891	0.006
GEDPNet(ours)	2024	0.930(0.033)	0.883(0.021)	0.019(0.018)	0.946(0.030)	0.902(0.024)	0.005(0.010)

注：“↑”表示数据越高越好，“↓”表示数据越低越好；加粗数据为最优值；“—”表示无法获取的数据，()中为标准差。

够有效聚合富含语义及细节信息的不同层级的多尺度特征。3个模块相辅相成,使得模型 GEDPNet 具有出色的学习能力。

2.3.2 泛化能力分析

为了验证所提模型的泛化能力,本实验将在 Kvasir-SEG 和 CVC-ClinicDB 训练集上训练得到的模型 GEDPNet,在 CVC-T、CVC-ColonDB 和 ETIS 数据集上进行测试。另外,实验中所提模型 GEDPNet 还与其他先进方法进行性能比较,实验结果如表 3 所示,公平起见表格中比较方法的指标数据均来自原文。由表 3 可知:(1)GEDPNet 在 3 个指标上都获得了最优性能,如在 CVC-ColonDB 上,GEDPNet 的 mDice、mIoU 和 MAE 分别达到了 0.825、0.747 和 0.025;如在 ETIS 上,GEDPNet 的 mDice、mIoU 和 MAE 分别达到了 0.806、0.733 和 0.013;在 CVC-T 上,GEDPNet 的 mDice、mIoU 和 MAE 分别达到了 0.911、0.483 和 0.005。(2)所提 GEDPNet 与其他方法相比,性能基本均有较大提升,如在 CVC-ColonDB 上与性能次优方法 RTA-Former-L 相比,mIoU 提升了 1.3%;在 ETIS 上与性能次优方法 CIFG-Net 相比,mIoU 提升了 0.7%,但是 mDice 和 MAE 均没有提升;在 CVC-T 上与性能次优方法 ECTransNet 相比,mDice 提升了 0.4%。(3)所提模型和经典医学图像分割模型相比,在 3 个数据集上 GEDPNet 的 mDice 比 U-Net 分别提升了至少 20%、30% 和 40%,比经典息肉图像分割方法 PraNet 分别高出了

3.9%、17.8% 和 11.6%。(4)从 GEDPNet 的各指标标准差可以看出,如 3 个数据集上 MAE 的标准差分别为 0.030、0.031 和 0.016,表明了所提模型具有相对稳定的分割泛化能力。

表 3 不同模型泛化能力比较

Table 3 Comparison of generalization ability for different models

模型	年份	CVC-ColonDB			ETIS			CVC-T		
		mDice ↑	mIoU ↑	MAE ↓	mDice ↑	mIoU ↑	MAE ↓	mDice ↑	mIoU ↑	MAE ↓
U-Net ^[2]	2015	0.512	0.444	0.061	0.398	0.335	0.036	0.710	0.627	0.022
PraNet ^[5]	2020	0.709	0.640	0.045	0.628	0.567	0.031	0.871	0.797	0.010
SANet ^[29]	2021	0.753	0.670	—	0.750	0.654	—	0.888	0.815	—
UACANet-S ^[6]	2021	0.783	0.704	0.034	0.694	0.615	0.023	0.902	0.837	0.006
MSNet ^[30]	2021	0.755	0.678	0.041	0.719	0.664	0.020	0.869	0.807	0.010
BDG-Net ^[31]	2022	0.804	0.725	—	0.756	0.679	—	0.899	0.831	—
CaraNet ^[7]	2022	0.773	0.689	0.042	0.747	0.672	0.017	0.903	0.838	0.007
ECTransNet ^[32]	2023	0.766	0.687	—	0.722	0.655	—	0.907	0.840	—
HarDNet-CPS ^[33]	2023	0.729	0.658	0.037	0.690	0.619	0.014	0.891	0.826	0.008
CFANet ^[34]	2023	0.743	0.665	0.039	0.732	0.655	0.014	0.893	0.827	0.008
RA-DENet ^[35]	2023	0.768	0.686	0.032	0.760	0.688	0.019	0.891	0.821	0.007
PPNet ^[36]	2023	0.791	0.726	0.028	0.784	0.716	0.013	0.899	0.839	0.006
ConvMLPSeg ^[37]	2023	0.793	0.717	0.033	0.723	0.676	0.014	0.893	0.822	0.007
CCFNet ^[17]	2024	0.768	0.683	0.035	0.747	0.669	0.014	0.892	0.827	0.007
NPD-Net ^[38]	2024	0.812	0.729	0.029	0.782	0.700	0.017	0.876	0.803	0.011
IC-PolypSeg-EF7 ^[39]	2024	0.799	0.728	0.030	0.758	0.670	0.015	0.907	0.846	0.009
MEGANet ^[40]	2024	0.793	0.714	0.040	0.739	0.665	0.037	0.899	0.834	0.007
BRNet ^[16]	2024	0.795	0.723	0.031	0.760	0.696	0.005	0.898	0.836	0.007
RTA-Former-L ^[41]	2024	0.818	0.734	—	0.795	0.714	—	0.891	0.815	—
CIFG-Net ^[19]	2024	0.815	0.733	0.030	0.806	0.726	0.013	0.901	0.834	0.007
GEDPNet(ours)	2024	0.825 (0.045)	0.747 (0.040)	0.025 (0.030)	0.806 (0.040)	0.733 (0.032)	0.013 (0.031)	0.911 (0.021)	0.848 (0.016)	0.005 (0.016)

上述结果表明,针对不同患者肠道环境差异较大,以及光照、角度等对息肉图像采集的影响,所提 GEDPNet 表现出了比其他模型更好的泛化能力,这主要归因于所设计的 GE 模块能够有效去除特征之间的相关性,提取息肉域不变特征。尽管如此,GEDPNet 在部分数据集上的泛化能力仍有待进一步提升。

2.3.3 可视化分析

为了更直观地展示本文方法 GEDPNet 在不同数据集上的整体性能,在 Kvasir-SEG、CVC-ClinicDB、CVC-T、CVC-ColonDB 和 ETIS 这 5 个数据集上对所提模型 GEDPNet 的学习能力和泛化能力进行了可视化分析,并与其他 10 个先进方法进行性能比较。实验中采用 mDice、mIoU 两个指标,结果如图 8 所示。从图 8 中可以观察到,所提 GEDPNet 具有最优的学习和泛化能力,且性能更稳定。此实验再次验证了本文所设计的 GE、DP 和 CA 模块的有效性。

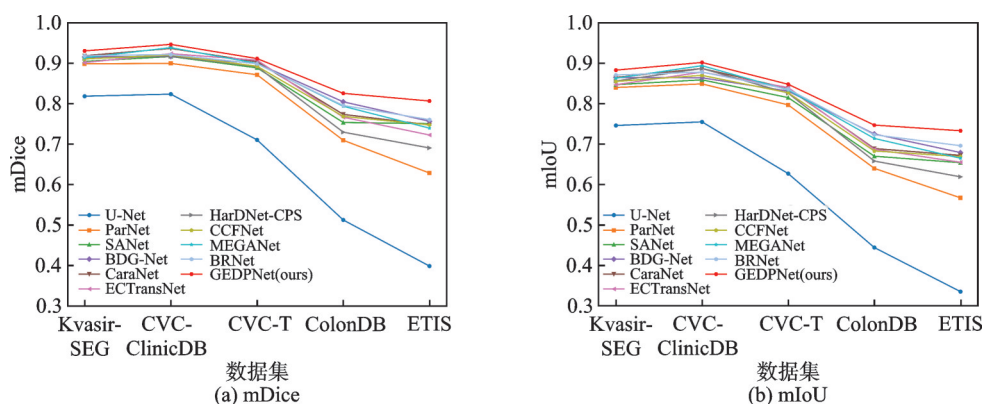


图8 不同指标可视化分析

Fig.8 Visual analysis of different indices

2.4 定性分析

为了更直观地展示息肉分割效果,本节对所提方法 GEDPNet 进行了定性分析,通过实验给出了尺寸不同、形状各异、环境差异大等不同形态的息肉分割结果视觉图,同时与其他 6 种方法 U-Net、SAnet、PraNet、CFANet、ECTransNet 和 MEGANet 进行了比较,结果如图 9 所示,其中从上到下每行图片分别来自 CVC-ClinicDB、CVC-T、Kvasir-SEG、ColonDB 和 ETIS 数据集。

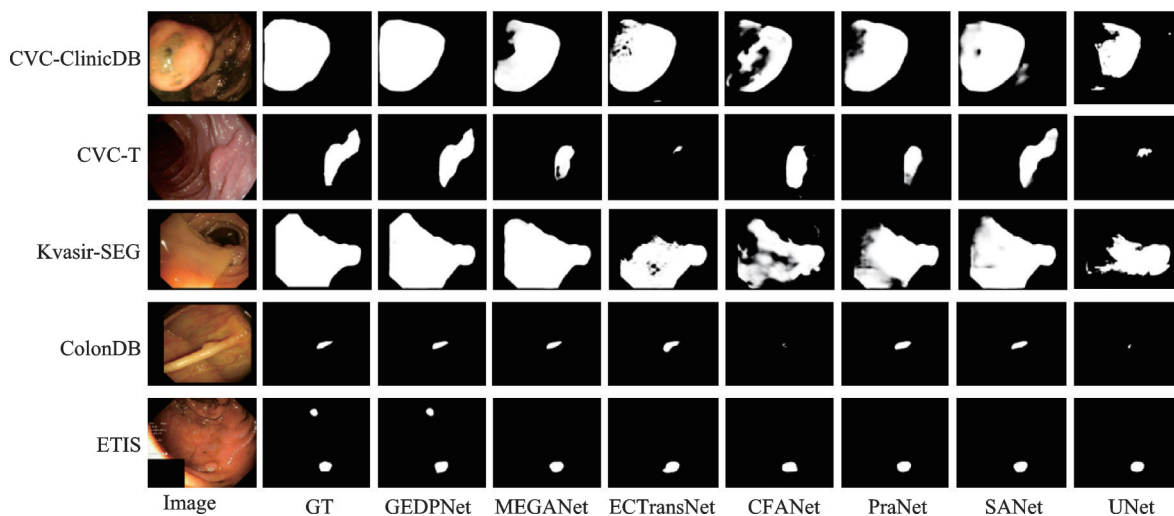


图9 视觉比较结果

Fig.9 Visual comparison results

从图 9 可以看出:(1)针对不同形态息肉,与其他方法相比,本文所提模型在 5 个数据集上均获得了最优分割效果;(2)针对环境差异大的息肉(第 1 行和第 2 行),GEDPNet 能够克服颜色、深度、褶皱和光线等环境变化,获得出色的分割结果,而 ECTransNet 等则不能很好地分割出完整息肉;(3)对于不同尺寸的息肉(第 3 行和第 4 行),GEDPNet 均能很好地分割出息肉,而 CFANet 等方法对小尺寸息肉分割性能有限;(4)针对多个息肉(第 5 行),GEDPNet 能同时有效分割多个息肉目标,而其他方法则没有分割出全部息肉。该实验从视觉角度直观展示了所提方法良好的息肉分割性能,尤其是对于所处环境差异

大的息肉,这主要归因于模型中设计的GE、DP和CA模块。

虽然所提模型GEDPNet在绝大多数场景下表现出了较好的分割性能,但在部分特定场景下,也会产生错误的分割结果,图10给出了部分示例。从图10可以看出:(1)面对小区域病灶时,GEDPNet未能精准识别目标区域,而误将背景分割为目标(第1列和第2列);(2)当息肉特征不明显时,GEDPNet误将突出的背景区域作为息肉进行了分割(第3列和第4列);(3)面对密集的小区域病灶时,GEDPNet无法精准定位密集且连续待分割的区域(第5列)。因此,未来工作中将进一步优化所提模型,提升对特征不明显的息肉目标、以及小目标尤其是密集小目标的分割性能。

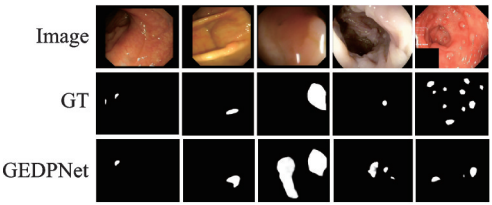


图10 错误分割示例

Fig.10 Examples of incorrect segmentation

2.5 消融分析

2.5.1 模块消融实验

本实验对所提模型GEDPNet中设计的GE、DP及CA模块在CVC-ClinicDB和ETIS数据集上分别进行消融实验,两个数据集上的结果分别体现了各个模块对模型学习能力和泛化能力的影响。实验中基线以PVT_v2作为主干,解码器仅采用上采样和拼接操作的跳跃连接,构成U型网络。实验采用控制变量法,分别添加GE、DP、CA模块和其任意组合。实验中采用mDice、mIoU两个指标,结果如表4所示。

表4 模块消融实验结果

Table 4 Module ablation experimental results

模型	GE	DP	CA	CVC-ClinicDB		ETIS	
				mDice ↑	mIoU ↑	mDice ↑	mIoU ↑
Baseline	—	—	—	0.910	0.861	0.765	0.684
#1	✓			0.935	0.887	0.793	0.715
#2		✓		0.935	0.891	0.779	0.699
#3			✓	0.923	0.872	0.784	0.709
#4	✓	✓		0.942	0.896	0.796	0.720
#5	✓		✓	0.941	0.893	0.800	0.724
#6		✓	✓	0.941	0.895	0.798	0.720
GEDPNet	✓	✓	✓	0.946	0.902	0.806	0.733

注:加粗数据为最优值。

从表4可以看出:

(1)从CVC-ClinicDB数据集上可以看出,GE、DP和CA单个模块均能使模型的学习能力得到提升。与基线模型相比,加入GE模块后,模型的mDice和mIoU均提升了2.5%,这是因为GE模块能应对尺寸多变的息肉目标,提取域丰富的层内多尺度信息,有效提升模型的分割精度;加入DP模块后,模型的mDice和mIoU分别提升2.5%和3.0%,这是因为DP模块能够获取全局和局部精准信息来提升模型分割性能,避免了直接融合而造成信息丢失;加入CA模块后,模型的mDice和mIoU分别提升了1.3%和1.1%,这是因为CA模块能够有效聚合多层特征,利用息肉的语义位置和纹理细节等信息来保证信息的完整性从而提升模型分割性能。

(2)从ETIS数据集上可以看出,GE、DP和CA单个模块均能使模型的泛化能力得到提升。与基线模型相比,加入GE模块后,模型的mDice和mIoU分别提升了2.8%和3.1%,主要是因为GE模块能够去除特征相关性,提取息肉域不变特征,以适应不同环境下的息肉分割,使模型的泛化能力大幅提升;加入DP模块后,模型的mDice和mIoU分别提升1.4%和1.5%,因为其可以动态感知未训练图像的全

局和局部特征;加入CA模块后,模型的mDice和mIoU分别提升1.9%和2.5%,其能够充分聚合原始图像的高维语义信息与息肉的纹理信息,从而有效识别分割模型未见过的息肉区域。

(3)在CVC-ClinicDB数据集和ETIS数据集上,GE、DP和CA两两模块组合比单独使用一个模块时更能促进模型学习能力和泛化能力的提升。如同时采用GE和DP模块的模型,比基线模型的mIoU分别提升了3.5%和3.6%;同时采用GE和CA模块,模型的mIoU指标分别提升了3.2%和4.0%;同时采用DP和CA模块,模型的mIoU指标分别提升了3.4%和3.6%。

(4)当同时使用GE、DP和CA这3个模块时,也即本文所提模型GEDPNet,在CVC-ClinicDB和ETIS数据集上,比基线方法的mDice分别提升3.6%和4.1%,mIoU分别提升4.1%和4.9%。这表明GE模块的泛化能力、DP模块的动态信息感知能力、CA模块的语义与纹理信息聚合能力同时发挥作用,有效提升了所提模型的整体学习能力和泛化性能。

2.5.2 消融可视化分析

为了清晰展示所设计模块对分割性能的影响,本实验对GE、DP和CA这3个模块的消融结果进行了可视化分析。实验结果如图11所示,其中绿色代表正确的息肉分割部分,黄色表示未从背景中分割出的息肉部分,红色表示把背景误分割成息肉的部分。

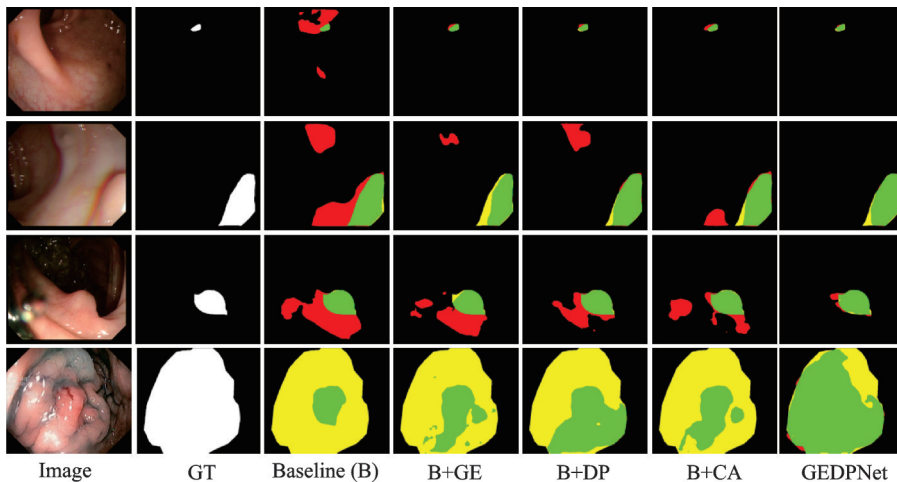


图11 消融可视化结果

Fig.11 Ablation visualization results

从图11可以看出:(1)基线(第3列)方法分割结果中出现大量错误,如有的丢失了大部分息肉,有的将一个目标分割出多个目标;(2)基线中加入GE模块后(第4列),病灶区域的分割面积有所提升,误判的背景区域也有所减少,这主要归因于GE模块可以学习到息肉的域不变特征,去除层内多尺度增强特征中包含的冗余和噪声,有效区别目标区域和背景;(3)基线中加入DP模块后(第5列),有效缓解了小区域息肉分割不充分、背景区域错误分割等问题,同时大幅提升了真实目标分割精度,这是因为DP模块具有良好的动态感知局部和全局信息的能力;(4)基线中加入CA模块后(第6列),能够有效融合不同层级的多尺度特征,充分捕获息肉的语义位置和纹理细节;(5)同时采用GE、DP和CA模块(第7列),即本文所提GEDPNet,能够获得最佳分割效果。本实验的可视化结果再一次验证了本文设计的GE、DP和CA模块的有效性。

2.5.3 可切换白化消融分析

为了验证GE模块中SW对模型泛化性能的影响,本节在CVC-ClinicDB和ETIS数据集上分别对SW进行了消融实验分析,同时将所提方法GEDPNet和基线方法进行了比较,结果如表5所示。

表 5 可切换白化消融实验结果
Table 5 SW ablation experimental results

模型	方法	CVC-ClinicDB		ETIS	
		mDice $\uparrow$	mIoU $\uparrow$	mDice $\uparrow$	mIoU $\uparrow$
#1	Baseline(B)	0.910	0.861	0.765	0.684
#2	B+GE (w/o SW)	0.932	0.886	0.784	0.708
#3	B+GE	0.935	0.887	0.793	0.715
#4	GEDPNet (w/o SW)	0.943	0.897	0.800	0.724
#5	GEDPNet	0.946	0.902	0.806	0.733

注:加粗数据为最优值。

从表5可以看出:(1)在 CVC-ClinicDB数据集上,#2和#3分别表示基线模型加入不包括SW的GE模块和包含SW的GE模块,#3比#2的mDice提升了0.3%;#4为GEDPNet去掉SW后的模型,#5为包含SW的GEDPNet,#5比#4的mDice提升了0.3%。(2)在 ETIS数据集上,#2和#3分别表示基线模型加入不包括SW的GE模块和包含SW的GE模块,#3比#2的mDice提升了0.9%;#4为GEDPNet去掉SW后的模型,#5为包含SW的GEDPNet,#5比#4的mDice提升了0.6%。实验结果表明,引入SW的GE模块能够有效提升模型的性能,特别是泛化能力,主要是因为GE模块利用SW能够有效去除特征间的相关性,提取息肉域不变特征。

2.6 迁移性分析

为了验证本文设计的GE、DP和CA模块对不同主干的迁移性,本实验对各个模块进行了迁移性分析。实验中对基于Transformer的主干PVT_v2和Swintransformer_base以及基于CNN的主干Res2Net_50和Convnext_tiny这4个不同主干在5个数据集(Kvasir-SEG、CVC-ClinicDB、CVC-T、CVC-ColonDB和ETIS)上分别进行了实验。实验采用mDice和mIoU作为指标,结果如图12~15所示,其中“主干+”表示在不同主干上添加了本文设计的模块。

从图12~15可以看出:(1)对于Res2Net主干,引入本文设计模块后mDice和mIoU均提升了30%以上。(2)对于Convnext和Swintransformer主干,引入设计模块后性能也有较大程度的提升。例如在CVC-ColonDB和ETIS上,Convnext的mIoU分别提升了5.4%和6.7%;Swintransformer的mIoU分别提升了4.5%和6.8%。(3)对于PVT_v2主干,引入设计模块后在两个数据集上均获得了比其他主干更优的息肉分割性能,因此,本文所提模型GEDPNet中采用PVT_v2作为主干。上述实验结果表明,所

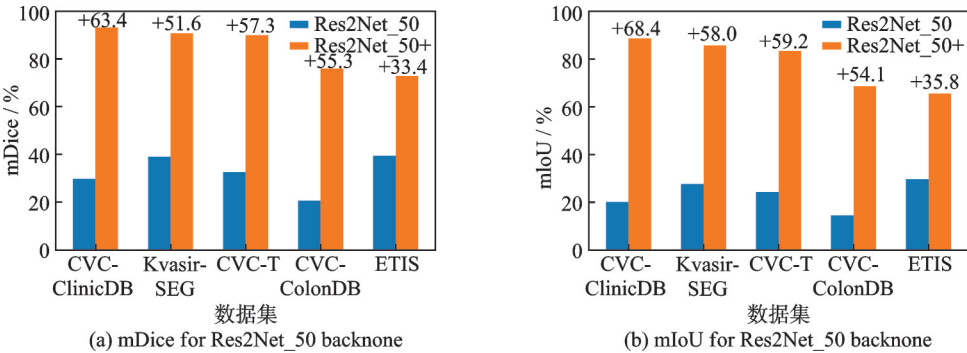


图12 迁移性分析(Res2Net_50)

Fig.12 Transferability analysis(Res2Net_50)

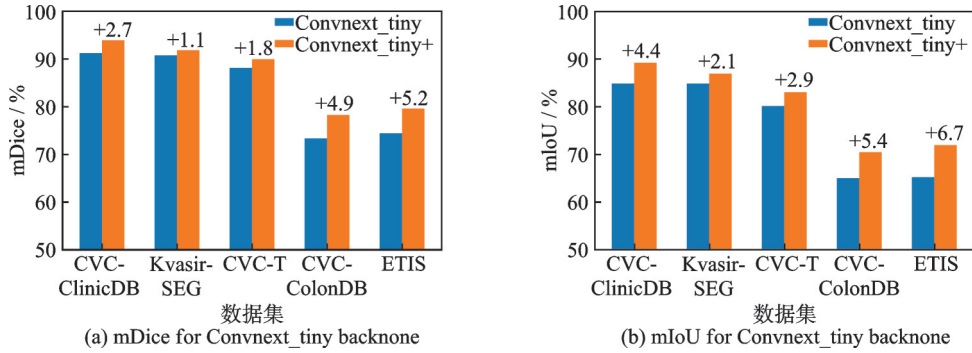


图 13 迁移性分析(Convnext_tiny)

Fig.13 Transferability analysis(Convnext_tiny)

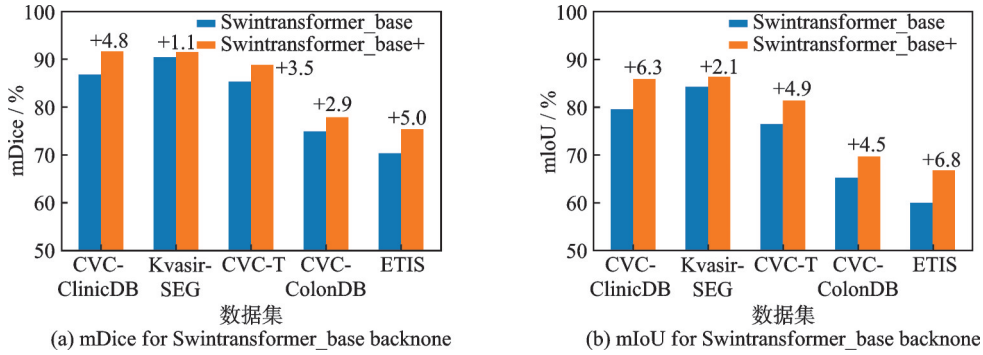


图 14 迁移性分析(Swintransformer_base)

Fig.14 Transferability analysis(Swintransformer_base)

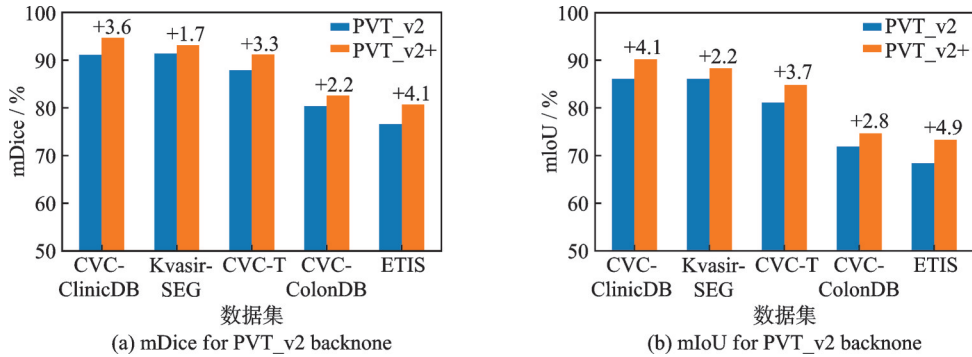


图 15 迁移性分析(PVT_v2)

Fig.15 Transferability analysis(PVT_v2)

设计的GE、DP和CA模块具有很好的迁移性和普适性,不管针对基于CNN的主干还是基于Transformer的主干,均能有效提升模型的息肉分割精度。

3 结束语

本文提出了一种泛化增强与动态感知的息肉分割网络GEDPNet来提升息肉分割性能。首先,设计GE模块,创新性地从提取息肉域不变特征的角度提升模型的泛化性,同时获取层内多尺度信息;然

后设计DP模块,动态感知精准的全局信息和局部信息,获取长期依赖关系和局部细节信息;最后设计CA模块,聚合不同层级且含有不同语义信息的多尺度特征,充分捕获息肉的语义位置和细节纹理等信息,提升分割性能。该模型在不同数据集上进行了实验验证,同时与其他相关方法进行了比较。结果表明,所提模型能有效提升息肉分割性能。特别地,所提模型GEDPNet具有良好的泛化能力,在存在域差异的复杂结直肠息肉分割场景下依然能够有效分割息肉。总之,GEDPNet不仅能够提升息肉分割的准确性,还为解决结直肠息肉分割中的多尺度信息处理和泛化能力增强提供了新思路,能够进一步促进结直肠息肉临床辅助诊断效果。

尽管所提模型GEDPNet能够提升复杂场景下结直肠息肉分割精度,但模型的参数量和计算复杂度仍然存在较大局限。因此,未来工作中,将针对这一问题进行模型轻量化设计与研究,有效提高模型分割效率,以满足结直肠息肉分割临床实际需求。此外,也将对模型的泛化性和鲁棒性进一步深入研究,以满足其他医学图像分割任务和医学场景的需要。

参考文献:

- [1] 考文涛,李明,马金刚.卷积神经网络在结直肠息肉辅助诊断中的应用综述[J].计算机科学与探索,2024,18(3):627-645.
KAO Wentao, LI Ming, MA Jingang. Review of application of convolutional neural network in auxiliary diagnosis of colorectal polyps[J]. Journal of Frontiers of Computer Science and Technology, 2024, 18(3): 627-645.
- [2] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Munich, Germany: Springer International Publishing, 2015: 234-241.
- [3] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKHSH N, et al. UNet++: A nested U-Net architecture for medical image segmentation[C]//Proceedings of Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Granada, Spain: Springer International Publishing, 2018: 3-11.
- [4] JHA D, SMEDSRUD P H, RIEGLER M A, et al. ResUNet++: An advanced architecture for medical image segmentation [C]//Proceedings of 2019 IEEE International Symposium on Multimedia (ISM). [S.l.]: IEEE, 2019.
- [5] FAN D P, JI G P, ZHOU T, et al. PraNet: Parallel reverse attention network for polyp segmentation[C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer International Publishing, 2020: 263-273.
- [6] KIM T, LEE H, KIM D. UACANet: Uncertainty augmented context attention for polyp segmentation[C]//Proceedings of the 29th ACM International Conference on Multimedia. [S.l.]: ACM, 2021: 2167-2175.
- [7] LOU A, GUAN S, KO H, et al. CaraNet: Context axial reverse attention network for segmentation of small medical objects [C]//Proceedings of Medical Imaging 2022: Image Processing. [S.l.]: SPIE, 2022, 12032: 81-92.
- [8] 林佳俐,李永强,徐希舟,等.边缘概率分布引导的结直肠息肉高分辨率分割网络[J].中国图象图形学报,2023,28(12): 3897-3910.
LIN Jiali, LI Yongqiang, XU Xizhou, et al. Edge-distribution-guided high-resolution network for colorectal polyp segmentation [J]. Journal of Image and Graphics, 2023, 28(12): 3897-3910.
- [9] 杨瑞君,陈丽叶,程燕.基于多尺度边缘感知和增强的息肉图像分割[J].计算机工程与应用,2025,61(1): 272-281.
YANG Ruijun, CHEN Liye, CHENG Yan. Polyp image segmentation based on multi-scale edge perception and enhancement [J]. Computer Engineering and Applications, 2025, 61(1): 272-281.
- [10] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[EB/OL]. (2020-10-22). <https://doi.org/10.48550/arXiv.2010.11929>.
- [11] WANG W, XIE E, LI X, et al. PVT v2: Improved baselines with pyramid vision transformer[J]. Computational Visual Media, 2022, 8(3): 415-424.
- [12] DONG B, WANG W, FAN D P, et al. Polyp-PVT: Polyp segmentation with pyramid vision transformers[EB/OL]. (2021-08-16). <https://doi.org/10.26599/AIR.2023.9150015>.

- [13] WANG J, HUANG Q, TANG F, et al. Stepwise feature fusion: Local guides global[C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022: 110-120.
- [14] RAHMAN M M, MARCULESCU R. Medical image segmentation via cascaded attention decoding[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. [S.l.]: IEEE, 2023: 6222-6231.
- [15] 庞飞翔, 丁德锐, 罗康. 一种并行注意力的金字塔视觉 Transformer 的结肠息肉分割网络[J/OL]. 小型微型计算机系统, 1-9 [2024-05-11]. <http://kns.cnki.net/kcms/detail/21.1106.TP.20240510.1838.006.html>.
PANG Feixiang, DING Derui, LUO Kang. Parallel attention pyramid visual Transformer for colon polyp segmentation network[J/OL]. Journal of Chinese Computer Systems, 1-9[2024-05-11]. <http://kns.cnki.net/kcms/detail/21.1106.TP.20240510.1838.006.html>.
- [16] YUE G, LI Y, JIANG W, et al. Boundary refinement network for colorectal polyp segmentation in colonoscopy images[J]. IEEE Signal Processing Letters, 2024, 31: 954-958.
- [17] CAI D, ZHAN K, TAN Y, et al. Cross-level context fusion network for polyp segmentation in colonoscopy images[J]. IEEE Access, 2024, 12: 35366-35377.
- [18] PAN X, ZHAN X, SHI J, et al. Switchable whitening for deep representation learning[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2019: 1863-1871.
- [19] LI W, HUANG Z, LI F, et al. CIFG-Net: Cross-level information fusion and guidance network for polyp segmentation[J]. Computers in Biology and Medicine, 2024, 169: 107931.
- [20] 吴港, 全海燕. 一种基于特征融合的息肉分割双解码模型[J]. 数据采集与处理, 2024, 39(4): 954-966.
WU Gang, QUAN Haiyan. A double-decoding model for polyp segmentation based on feature fusion[J]. Journal of Data Acquisition and Processing, 2024, 39(4): 954-966.
- [21] XIAO B, HU J, LI W, et al. CTNet: Contrastive transformer network for polyp segmentation[J]. IEEE Transactions on Cybernetics, 2024, 54(9): 5040-5053.
- [22] LI J, WEN Y, HE L. SCConv: Spatial and channel reconstruction convolution for feature redundancy[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2023: 6153-6162.
- [23] DAI Y, GIESEKE F, OEHMCKE S, et al. Attentional feature fusion[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. [S.l.]: IEEE, 2021: 3560-3569.
- [24] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Kvasir-SEG: A segmented polyp dataset[C]//Proceedings of the 26th International Conference on MultiMedia Modeling. Daejeon, South Korea: Springer International Publishing, 2020: 451-462.
- [25] BERNAL J, SÁNCHEZ F J, FERNÁNDEZ-ESPARRACH G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians[J]. Computerized Medical Imaging and Graphics, 2015, 43: 99-111.
- [26] TAJBAKHSH N, GURUDU S R, LIANG J. Automated polyp detection in colonoscopy videos using shape and context information[J]. IEEE Transactions on Medical Imaging, 2015, 35(2): 630-644.
- [27] SILVA J, HISTACE A, ROMAIN O, et al. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer[J]. International Journal of Computer Assisted Radiology and Surgery, 2014, 9: 283-293.
- [28] VÁZQUEZ D, BERNAL J, SÁNCHEZ F J, et al. A benchmark for endoluminal scene segmentation of colonoscopy images[J]. Journal of Healthcare Engineering, 2017, 2017(1): 4037190.
- [29] WEI J, HU Y, ZHANG R, et al. Shallow attention network for polyp segmentation[C]//Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI 2021. Strasbourg, France: Springer International Publishing, 2021: 699-708.
- [30] WANG S, CONG Y, ZHU H, et al. Multi-scale context-guided deep network for automated lesion segmentation with endoscopy images of gastrointestinal tract[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 25(2): 514-525.
- [31] QIU Z, WANG Z, ZHANG M, et al. BDG-Net: Boundary distribution guided network for accurate polyp segmentation[C]//Proceedings of Medical Imaging 2022: Image Processing. [S.l.]: SPIE, 2022, 12032: 792-799.
- [32] LIU W, LI Z, LI C, et al. ECTransNet: An automatic polyp segmentation network based on multi-scale edge complementary[J]. Journal of Digital Imaging, 2023, 36(6): 2427-2440.

- [33] YU T, WU Q. HarDNet-CPS: Colorectal polyp segmentation based on harmonic densely united network[J]. Biomedical Signal Processing and Control, 2023, 85: 104953.
- [34] ZHOU T, ZHOU Y, HE K, et al. Cross-level feature aggregation network for polyp segmentation[J]. Pattern Recognition, 2023, 140: 109555.
- [35] WANG K, LIU L, FU X, et al. RA-DENet: Reverse attention and distractions elimination network for polyp segmentation[J]. Computers in Biology and Medicine, 2023, 155: 106704.
- [36] HU K, CHEN W, SUN Y Z, et al. PPNet: Pyramid pooling based network for polyp segmentation[J]. Computers in Biology and Medicine, 2023, 160: 107028.
- [37] JIN Y, HU Y, JIANG Z, et al. Polyp segmentation with convolutional MLP[J]. The Visual Computer, 2023, 39(10): 4819-4837.
- [38] YU Z, ZHAO L, LIAO T, et al. A novel non-pretrained deep supervision network for polyp segmentation[J]. Pattern Recognition, 2024, 154: 110554.
- [39] CHEN Z, WANG K, LIU Y. Efficient polyp segmentation via integrity learning[C]//Proceedings of 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S.l.]: IEEE, 2024: 1826-1830.
- [40] BUI N T, HOANG D H, NGUYEN Q T, et al. MEGANet: Multi-scale edge-guided attention network for weak boundary polyp segmentation[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. [S.l.]: IEEE, 2024: 7985-7994.
- [41] LI Z, YI M, UNERI A, et al. RTA-Former: Reverse transformer attention for polyp segmentation[C]//Proceedings of the 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). [S.l.]: IEEE, 2024.

作者简介:



王森(1999-),男,硕士研究生,研究方向:医学图像分割和深度学习,E-mail:wang-sanmu1029@163.com。



史彩娟(1977-),通信作者,女,博士,教授,研究方向:计算机视觉、图像处理和深度学习等,E-mail:scj-blue@163.com。



蔡澳(2000-),男,硕士研究生,研究方向:医学图像分割和深度学习。



王睿(1998-),男,硕士研究生,研究方向:伪装目标检测和弱监督学习。



于鑫阳(1999-),女,硕士研究生,研究方向:医学图像分割和深度学习。



程旭东(1999-),男,硕士研究生,研究方向:目标检测和域自适应学习。



陈伟彬(1982-),男,副主任医师,硕士生导师,研究方向:X线、CT及MRI等医学影像诊断。

(编辑:王静)