

文章编号 1004-924X(2026)14-2247-12

融合几何先验与双域特征提纯的 结直肠息肉分割网络

李 梁¹, 毕 利^{1*}, 焦小刚¹, 毕逢辰², 王俊杰^{3,4}

(1. 宁夏大学 信息工程学院, 宁夏 银川 750021;

2. 宁夏医科大学 基础医学院, 宁夏 银川 750021;

3. 中国科学院 合肥物质科学研究院, 安徽 合肥 230031;

4. 中国科学技术大学, 安徽 合肥 230026)

摘要:针对结直肠息肉分割中微小病灶捕捉困难、强光反射伪影干扰及边缘定位不准等瓶颈,提出一种融合几何先验与双域特征提纯的结直肠息肉分割网络 GPDP-Net。首先,利用优化边缘模块捕获显式几何先验,并结合改进型坐标注意力机制增强位置感知;其次,设计物理-频域并行增强瓶颈层,通过物理信息神经网络的演化约束抑制非相干光学噪声,同时利用快速傅里叶变换分支在频谱空间补偿高频信号损失;最后,在级联解码器中引入亚像素还原算子与迭代反向注意力机制以实现病灶细节的保真还原。在四个基准数据集上的评估结果显示,GPDP-Net在各项指标上均展现出较强的竞争力。在包含大量极小尺度目标的ETIS数据集上,其mIoU达到0.701,较次优模型提升1.4%,且边界误差显著降低。实验证明,物理一致性约束与频谱增强的协同机制可有效提升复杂环境下多尺度病灶的分割精度。

关键词:结直肠息肉;物理信息神经网络;快速傅里叶变换;边缘感知;反向注意力机制

中图分类号:TP391 文献标识码:A

doi:10.37188/OPE.20263414.2247 CSTR:32169.14.OPE.20263414.2247

Colorectal polyp segmentation network fusing geometric prior and dual-domain feature purification

LI Liang¹, Billy^{1*}, JIAO Xiaogang¹, BI Fengchen², WANG Junjie^{3,4}

(1. School of Information Engineering, Ningxia University, Yinchuan 750021, China;

2. School of Basic Medical Sciences, Ningxia Medical University, Yinchuan 750021, China;

3. Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China;

4. University of Science and Technology of China, Hefei 230026, China)

* Corresponding author, E-mail: billy68@nxu.edu.cn

Abstract: To address the challenges of tiny lesion segmentation, specular reflection interference, and inaccurate boundary localization during colorectal polyp segmentation, this paper proposes a Geometric Prior and Dual-Domain Purification Network (GPDP-Net). First, an Optimized Edge Module (OEM) is designed to extract explicit geometric priors, and an Improved Coordinate Attention Mechanism is intro-

收稿日期:2026-04-21;修订日期:2026-06-30.

基金项目:宁夏自然科学基金重点项目(No. 2023AAC02011);宁夏重点研发项目(引才专项)(No. 2023BSB03015)

duced to enhance spatial position awareness. Second, a Parallel Dual-Domain Enhancement Hub (PD-DEH) is designed, where the evolutionary constraint based on the Physics-Informed Neural Network (PINN) is employed to suppress incoherent optical noise, while the Fast Fourier Transform Branch is utilized to recover high-frequency information in the frequency domain. Finally, a Sub-pixel Restoration and Iterative Reverse-Attention Decoder (SR-IRAD) is incorporated into the cascaded decoder to achieve high-fidelity restoration of lesion details. Experimental results on four benchmark datasets demonstrate that GPDP-Net achieves strong performance on various evaluation metrics. Specifically, on the ETIS-Larib-PolypDB dataset containing numerous tiny lesions, the proposed network achieves an mIoU of 0.701, which is 1.4 percentage points higher than the second-best method. These results demonstrate that GPDP-Net can effectively improve the segmentation accuracy of colorectal polyps with complex structures and enhance the robustness and generalization capability of the model in challenging clinical scenarios.

Key words: colorectal polyp; physics-informed neural networks; fast Fourier transform; edge awareness; reverse attention mechanism

1 引言

结直肠癌 (Colorectal Cancer, CRC) 是全球范围内发病率和死亡率最高的恶性肿瘤之一。临床研究表明,早期发现并切除胃肠道息肉能够显著阻断其向恶性肿瘤的演变,从而大幅降低 CRC 的死亡率^[1]。结肠镜检查是目前息肉筛查的金标准。然而,在实际临床诊断中,由于息肉的形态、大小多变,且常与周围正常黏膜组织高度相似,加之医生的疲劳和经验差异,息肉的漏诊率仍然居高不下^[2-3]。因此,开发准确、鲁棒的计算机辅助诊断 (Computer-Aided Diagnosis, CAD) 系统进行自动化的息肉分割,对于辅助临床诊断和手术导航具有重要的现实意义。

为了解决上述临床困境,基于深度学习的计算机辅助诊断 (CAD) 系统逐渐成为胃肠道息肉筛查的重要研究方向。Ronneberger 等^[4]提出了一种经典的 U-Net 架构,通过对称的编码器-解码器结构与跳跃连接 (Skip Connections) 实现了医学图像的端到端像素级分割,奠定了该领域的基础。随后,为了克服网络加深带来的梯度消失以及多尺度特征提取不足的问题,Jha 等^[5]在 U-Net 的基础上结合残差块 (Residual Blocks)、空洞空间金字塔池化 (ASPP) 与挤压激励 (SE) 模块提出了 ResUNet++ 模型,在各类息肉数据集上表现出更强的特征表达能力。然而,针对息肉与正常黏膜对比度低且边界模糊的特殊难点,Fan

等^[6]提出了一种并行反向注意力网络 (PraNet),该方法利用反向注意力机制模拟临床医生先定位、后寻边的观察策略,通过逐步挖掘和细化局部边缘细节,在处理高难度息肉边界时展现出了卓越的分割精度。尽管 PraNet 等注意力机制模型取得了显著进展,但针对复杂形态的息肉,尤其是微小息肉和边缘模糊息肉的分割,仍存在巨大挑战。为了防止微小息肉在深度卷积下采样中丢失特征,Wei 等^[7]提出了浅层注意力网络 (Shallow Attention Network, SANet),通过颜色解耦与浅层特征强化,显著提升了小息肉的检出率。同时,针对传统特征融合容易引入冗余背景噪声的问题,Zhao 等^[8]设计了多尺度相减网络 (MSNet),摒弃了传统的拼接操作,利用相邻层特征的相减运算精准提取了息肉边缘的差异化信息。

尽管上述基于卷积神经网络的方法通过设计精细化算子不断优化分割效果,但受限于卷积核的局部感受野,此类方法在捕获长距离上下文依赖方面仍存在固有的局限性^[8-10]。这导致在面对跨尺度差异剧烈或形状极不规则的息肉病灶时,模型容易出现语义不一致以及误诊漏诊现象。为了突破这一瓶颈,研究者开始探索具备强大全局建模能力的视觉 Transformer (ViT) 架构^[11]。Wang 等^[12]提出了一种金字塔视觉 Transformer (PVT) 架构,通过分层结构与空间缩减注意力机制,有效平衡了计算效率与全局感知的深

度。在此基础上,Dong等^[13]针对息肉分割任务提出了Polyp-PVT模型,利用Transformer的自注意力机制在全图范围内建立像素间的长程关联,显著增强了模型对多尺度病灶的表征鲁棒性。随后,为了进一步兼顾全局语义信息与局部细节特征,Zhang等^[14]设计了TransFuse架构,通过并行融合Transformer与CNN双分支,有效弥补了单一架构在细节恢复与边缘细化上的短板。

结肠镜检查依赖于点光源照射,而肠道内壁充满黏液且具有半透明属性,这导致成像过程中普遍存在强烈的镜面反射与光线漫反射干扰^[15-16]。传统的CNN和Transformer模型纯粹依赖数据驱动,难以从底层逻辑上区分病灶边缘与光学伪影,导致在反光强烈的区域极易出现误诊。在经典图像处理理论中,各向异性扩散等物理方程常被用于处理此类复杂的非相干噪声并保护结构边界,但这一物理先验在现有的深度学习分割框架中尚未得到充分利用^[17-19]。其次,针对数据集中的极小病灶,现有架构存在严重的信号稀释问题。无论是CNN的池化操作还是Transformer的空间缩减,本质上都是低通滤波过程。对于微小息肉,其空间特征在深层网络中

被极度压缩,而传统的双线性插值上采样又因其平滑效应进一步模糊了本就微弱的病灶边界,导致模型对小目标的检出率处于较低水平。针对上述痛点,本文提出了融合几何先验与双域特征提纯的结直肠息肉分割网络(Geometric Prior and Dual-Domain Purification Network, GPDP-Net)。

2 模型介绍

针对胃肠道息肉分割中微小病灶捕捉困难、光学反射伪影干扰及模糊边界定位不准等瓶颈,本文构建了一种融合几何先验与双域特征提纯的结直肠息肉分割网络(GPDP-Net),如图1所示。该模型以分层视觉Transformer为表征核心,通过集成优化边缘模块(OEM)显式捕获精细几何先验;在瓶颈层创新性地引入由物理信息神经网络(PINN)与频域增强(FFT)构成的并行分支,通过物理一致性约束与频谱增益实现特征的结构化提纯与去噪;在解码阶段,利用亚像素还原算子配合自适应门控机制实现病灶细节的高保真恢复,显著提升模型在复杂临床场景下的分割精度与泛化鲁棒性。

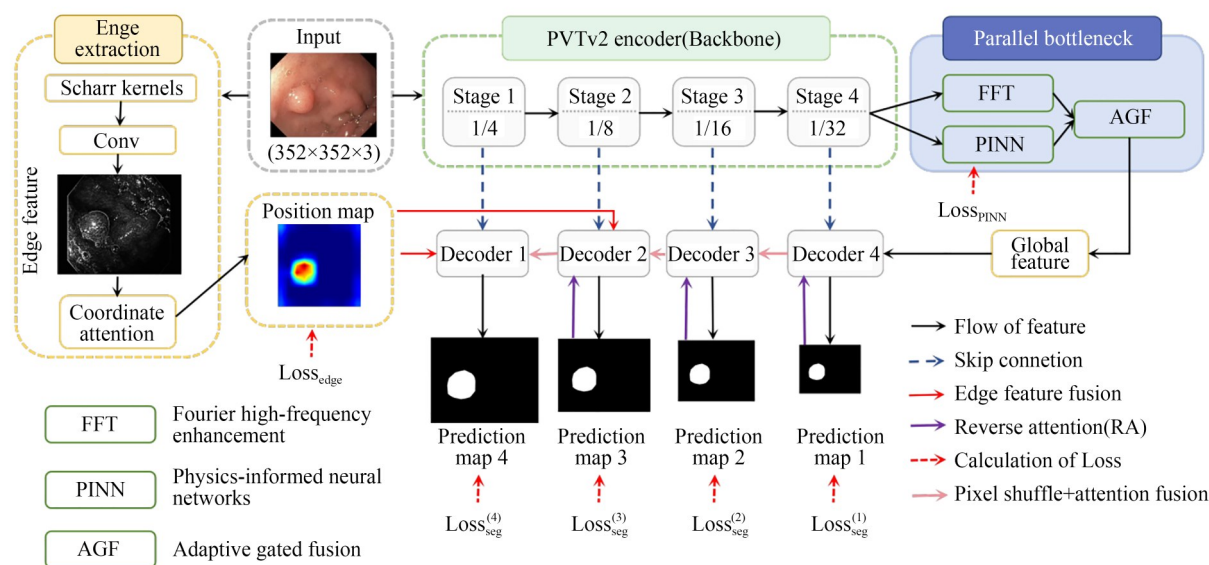


图1 GPDP-Net模型结构图

Fig. 1 GPDP-Net model structure diagram

2.1 优化边缘模块 (Optimized Edge Module, OEM)

本文采用分层视觉 Transformer(PVTv2)^[20]作为主干网络,旨在利用其强大的长程建模能力应对息肉形态的多样性与尺度剧变。然而,受限

于 Transformer 架构固有的低通滤波特性以及特征下采样过程中的空间信息损耗,模型易出现边缘坍塌或微小病灶漏诊。为此,本文设计了优化边缘感知模块,旨在深层语义流中重新注入物理层面的几何先验,其整体结构如图 2 所示。

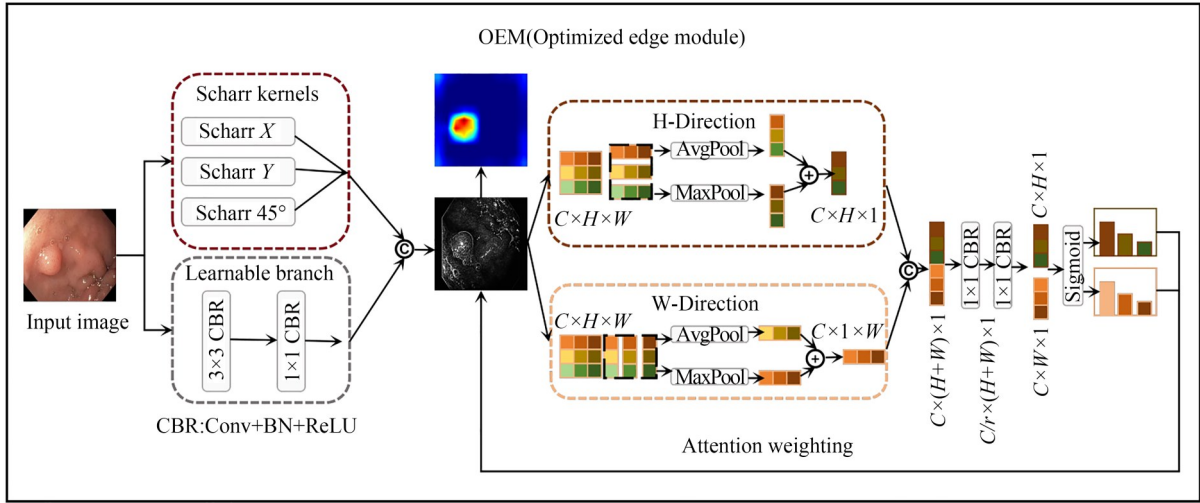


图 2 OEM 模块结构图

Fig. 2 OEM module structure diagram

与常规的隐式特征学习不同, OEM 利用多向 Scharr 梯度算子^[21]显式捕捉原始输入中的局部像素突变, 为模型提供高频对比度引导, 有效弥补了深层特征对模糊边界感知的匮乏。本研究采用的自定义 Scharr 卷积核如图 3 所示。

-4.5	-15	4.5	-4.5	0	4.5	0	4.5	15
0	0	0	-15	0	15	-4.5	0	4.5
-4.5	15	4.5	-4.5	0	4.5	-15	-4.5	0

图 3 自定义 Scharr 卷积核参数示意: (左 1) 水平方向敏感算子, (左 2) 垂直方向敏感算子, (左 3) 45° 方向敏感算子。

Fig. 3 Illustration of the customized Scharr convolution kernels: (Left 1) horizontal-sensitive operator, (Left 2) vertical-sensitive operator, and (Left 3) 45° direction-sensitive operator.

考虑到肠道反光与气泡等引起的伪边缘干扰, 本文引入增强型坐标注意力机制对边缘特征进行精炼。在具体实现上, H 与 W 方向执行并行

双路池化聚合, 以保留高频边缘并抑制噪声:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) + \max_{0 \leq i < W} x_c(h, i), \quad (1)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) + \max_{0 \leq j < H} x_c(j, w), \quad (2)$$

其中: x_c 为通道 c 的输入特征。所得空间向量经拼接及 1×1 卷积交互后, 重新拆分为水平权重 g^h 与垂直权重 g^w 。最终通过逐像素加权生成定位精准的显著性映射图:

$$y_c(i, j) = x_c(i, j) \times \sigma(g_c^h(i)) \times \sigma(g_c^w(j)). \quad (3)$$

这种策略在维持低计算开销的同时, 增强了模型捕获长程空间依赖的能力, 从而有效地从视觉噪声中剥离出息肉结构信息, 确保了注入主干网络的边缘先验具备极高的语义保真度。

2.2 物理-频域并行的双域瓶颈增强模块 (Parallel Dual-Domain Enhancement Hub, PDDEH)

针对结肠镜成像中普遍存在的镜面反射干扰以及微小病灶在深层特征中响应微弱的瓶颈, 本文在编码器与解码器之间构建了一个物理-频域并行的双域瓶颈增强模块, 如图 4 所示。

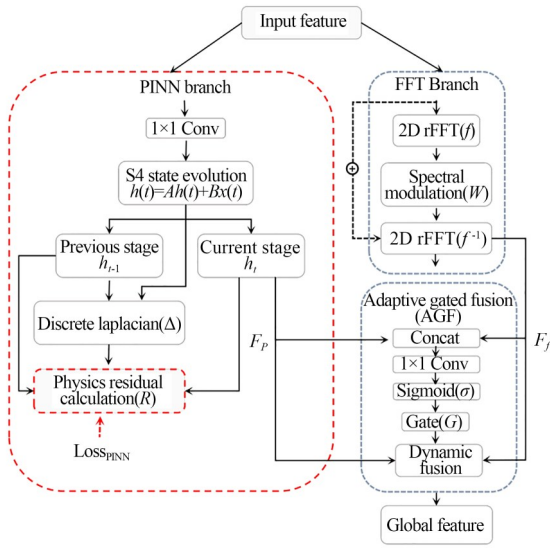


图4 PDDEH模块结构图

Fig. 4 PDDEH Module Structure Diagram

2.2.1 基于物理一致性约束的特征演化分支

本文将瓶颈层的特征变换建模为受控的动力学演化过程,引入物理信息神经网络(PINN)对特征状态转移执行显式正则化^[22]。首先,采用结构化状态空间模型(S4)^[23]作为演化算子。给定输入特征 X ,其状态转移遵循线性常微分方程:

$$\dot{h}(t) = Ah(t) + Bx(t). \quad (4)$$

该方程经离散化处理生成相邻时刻特征状态 h_{t-1} 与 h_t 。其中, A 为状态转移矩阵,负责刻画特征内部演化逻辑与长程空间依赖; B 为输入映射矩阵,定义了外部观测特征对系统的激励强度, h_{t-1} 与 h_t 分别代表特征演化序列中前后时刻的潜在状态,用于表征特征在伪时间维度上的演化增量。

其次,为确保特征演化符合生物组织的物理分布特征,本文引入离散拉普拉斯算子 Δ ^[24] 作为物理评判准则。不同于一阶梯度算子仅关注边缘强度,拉普拉斯算子作为二阶微分算子,其本质是对特征空间曲率的度量。在结肠镜成像环境下,生物组织的物理分布遵循扩散规律,即特征的自发演化倾向于最小化局部的曲率跳变。通过引入离散拉普拉斯算子,本文能够显式捕获并抑制不符合组织物理属性的孤立高频冲击,从而引导状态空间模型学习到具备物理相干性的深层表征。该算子通过预定义的固定卷积核 K_L

对特征空间的空间曲率进行建模:

$$K_L = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \quad (5)$$

根据扩散理论,受控的去噪过程应满足偏微分方程: $\frac{\partial h}{\partial t} = \nu \Delta h$ 。据此,本文构造物理残差 R 衡量神经演化轨迹与物理扩散规律之间的偏差:

$$R = (h_t - h_{t-1}) - \nu(K_L * h_{t-1}), \quad (6)$$

其中: ν 为扩散系数。通过在训练阶段最小化物理损失 $L_{\text{pinn}} = \text{MSE}(R)$,强制 S4 引擎在特征传递过程中自发平滑不符合扩散规律的高频光学伪影,从而提取具备物理鲁棒性的结构特征 F_p 。

2.2.2 频域细节增强分支

为了与物理分支的平滑逻辑互补,频域分支旨在频谱空间内补偿下采样造成的信息损耗。该分支利用二维快速傅里叶变换(2DrFFT)^[25-26]将空间特征映射至频率域,实现全局感受野覆盖。在频谱空间中,本文设计了频谱调制机制,引入可学习的复数权重矩阵 W 对高频分量执行显式增益:

$$\hat{X}_{\text{freq}} = W \odot F(X), \quad (7)$$

其中: F 为傅里叶变换算子。通过频谱加权,模型能够自适应地放大代表病灶边界的高频特征响应。随后,利用逆傅里叶变换 F^{-1} 将信号恢复至空间域,并通过残差连接确保原始语义的稳定性,输出频域增强特征 F_f 。

2.2.3 自适应门控融合机制(Adaptive Gated Fusion, AGF)

由于不同区域对去噪平滑与细节增强的需求存在空间异质性,本文设计了自适应门控融合模块(AGF)执行特征集成。模块首先将 F_p 与 F_f 进行通道拼接,通过 1×1 卷积与 Sigmoid 激活函数生成空间门控权重 G :

$$G = \sigma(\text{Conv}_{1 \times 1}([F_p, F_f])). \quad (8)$$

生成的权重图 G 用于动态调节双分支特征的贡献比例。最终的全局特征 F_{out} 计算如下:

$$F_{\text{out}} = G \odot F_f + (1 - G) \odot F_p. \quad (9)$$

该机制确保模型在抑制反射干扰的同时,以高度的语义保真度保留病灶的细节纹理,实现了物理稳定性与频域辨识度的正交协同。

2.3 亚像素还原与迭代反向注意力解码器 (Sub-pixel Restoration and Iterative Reverse-Attention Decoder, SR-IRAD)

本文设计的亚像素还原与迭代反向注意力解码器(SR-IRAD)通过亚像素空间还原算子与迭代反向注意力机制的融合,实现特征分辨率的保真恢复与病灶边界的精确细化。在层级特征的空间维度扩张阶段,该模块采用亚像素卷积替代传统的双线性插值,以克服线性加权算子导致的信号稀释问题。双线性插值在信号处理层面表现为低通滤波器,会不可避免地平滑深层特征中的高频响应,导致微小病灶的显著性降低。相比之下,亚像素卷积通过对通道维度的特征进行周期性重排,将深度信息直接映射至空间维度,从而在不引入冗余像素的前提下实现分辨率的扩张。给定输入特征 $F_{in} \in R^{C \times H \times W}$,其空间还原过程可表述为:

$$F_{up} = PS(W_p * F_{in} + b_p), \quad (10)$$

其中: W_p 为执行通道扩充的卷积核, PS 为特征重排算子。经空间还原后的特征与经过注意力融合处理后的浅层跳跃连接特征进行交汇,融合后的特征 F_{fused} 有效结合了深层语义的定位信息与浅层编码器的空间细节。为了进一步精细化病灶边缘并抑制非相干背景噪声,SR-IRAD在级联阶段引入了反向注意力机制。该机制利用高层解码阶段生成的预测掩模作为空间引导,通过挖掘主体区域之外的残余特征,实现对边界区域的迭代优化。具体而言,设上一阶段生成的分割预测图为 P_i ,其经过反向操作生成反向注意力权重 A_{rev} ,用以突出病灶边缘及其周围的候选背景:

$$A_{rev} = 1 - \sigma(P_i), \quad (11)$$

其中: $\sigma(\cdot)$ 表示 Sigmoid 函数。随后,该权重与当前层级的融合特征 F_{low} 执行逐像素相乘,强制模型关注被主体区域掩盖或忽略的精细几何结构

$$F_{refined} = Conv_{3 \times 3}(F_{low} \otimes A_{rev}), \quad (12)$$

其中: $\otimes$ 表示逐像素相乘。这种迭代式擦除与细化的逻辑,使得解码器能够自发地对由肠道反光引发的假阳性区域进行修正。此外,由优化边缘提取模块(OEM)生成的位置映射图作为显式的几何先验,在解码器的浅层阶段执行注入。亚像素还原算子确保了这种几何先验在层级传递过

程中的保真度,而反向注意力机制则确保了模型在高分辨率重构时能够精准聚焦于息肉的解剖学边界。

2.4 综合损失函数

为了协同优化前述的优化,本文定义了一个多任务学习范式下的综合损失函数 L_{total} 。该函数通过在空间、物理及几何三个维度施加显式约束,引导模型实现像素精度与物理稳健性的深度对齐。总体目标函数表述为:

$$L_{total} = \sum_{i=1}^K \lambda_i L_{seg}^{(i)} + \alpha L_{pinn} + \beta L_{edge}, \quad (13)$$

其中: $K=4$ 代表级联解码器中的深度监督层级, λ_i, α, β 分别为各任务的超参数权重。

2.4.1 多尺度结构化分割目标(L_{seg})

针对解码器 SR-IRAD 产生的层级化输出 $\{P_i\}_{i=1}^K$,本文采用结构化损失 L_{seg} 进行逐层监督。考虑到息肉病灶在内窥镜影像中极高的像素不平衡性,每一级 L_{seg} 由加权二值交叉熵(Weighted BCE)与加权交并比损失(Weighted IoU)融合而成:

$$L_{seg}(P, G) = L_{wbce}(P, G) + L_{wioU}(P, G), \quad (14)$$

其中: P 代表模型在不同层级输出的预测掩模, G 代表与之相对应的真实标签(Ground Truth)。该表述通过引入硬样本挖掘机制,强制网络在分辨率恢复过程中对微小病灶保持高频响应,同时抑制 L_{wioU} 项以维持分割拓扑结构的完整性。

2.4.2 物理演化一致性正则化(L_{pinn})

为确保瓶颈层中状态空间演化引擎(S4)的行为符合预设的扩散规律,本文构造了物理信息正则化项 L_{pinn} 。该项定义为特征演化轨迹与拉普拉斯空间响应之间的残差:

$$L_{pinn} = E // (h_t - h_{t-1}) - \nu(K_L * h_{t-1}) //^2. \quad (15)$$

此项本质上是在特征流中引入了一个物理定义的低通约束,通过惩罚不符合扩散方程的非相干能量跳变(如光学伪影),强制模型提取具备物理鲁棒性的深层语义特征。

2.4.3 边缘几何保真度约束(L_{edge})

为提升优化边缘模块(OEM)对边界定位的精确度,本文引入了基于 Dice 度量的几何保真度损失 L_{edge} :

$$L_{edge} = 1 - \frac{2 \sum (P_e \cdot G_e) + \epsilon}{\sum P_e + \sum G_e + \epsilon}, \quad (16)$$

其中: P_e 为 OEM 输出的显著性映射图, G_e 为由真

实掩模提取的形态学边缘梯度。该约束确保了边缘流能为解码阶段的反向注意力机制提供高置信度的几何锚点,从而实现对模糊边界的精准对齐。

3 实验和结果

3.1 数据集和评估指标

本文在四个公认的结直肠息肉分割基准数据集上验证模型性能:Kvasir-SEG^[27],CVC-ClinicDB^[28],CVC-ColonDB^[29],以及 ETIS-Larib-PolypDB^[30]。遵循 PraNet 的实验协议,本文将 Kvasir-SEG 与 CVC-ClinicDB 的部分数据合并作为训练集(共 1 450 张),其余数据用于测试模型的泛化能力。特别地,ETIS 数据集因包含大量微小病灶且对比度极低,被作为衡量模型对细微特征捕获能力的核心基准。

在评估体系方面,本文采用了四个维度的互补指标以全面量化性能:平均交并比(mIoU)与 Dice 系数(Dice)用于衡量预测掩模与真实标签在像素层面的重叠精度;结构度量(S_a)^[31]通过结合区域感知与物体感知信息,用于评估模型对息肉全局拓扑结构的捕获能力。此外,考虑到本文提出的优化边缘模块(OEM)与物理一致性约束旨在精细化目标轮廓,本文引入了 95% 豪斯多夫距离(HD95)^[32]作为核心边界评价指标。

3.2 实验设置

本实验基于 PyTorch1.12.1 深度学习框架实现,并在配备 NVIDIA A100 GPU 的高性能计算平台上进行训练与评估,利用该芯片的 Tensor Core 算力通过自动混合精度(AMP)技术与 GradScaler 梯度缩放策略来优化训练效率并确保数值稳定性。在网络优化方面,本文采用 AdamW 优化器,初始学习率设定为 3×10^{-4} ,权重衰减为 1×10^{-2} ,并引入带热身机制的余弦退火调度器^[33],其中前 10 个 Epoch 执行线性热身以将学习率从 1×10^{-6} 提升至预设值,随后在总计 200 个 Epoch 的训练周期内遵循余弦曲线衰减。所有输入图像均统一调整为 352×352 pixels,并应用随机翻转、旋转、缩放及颜色抖动等在线数据增强手段以提升模型的泛化性能。模型参数通过多准则联合损失函数进行迭代更新,该函数由权重为 1.0 的主结构损失(结合加权 BCE 与加

权 IoU)、权重为 0.5 的边缘形态 Dice 损失、权重为 0.3 的外部梯度一致性损失以及权重为 0.1 的内部物理演化残差损失加权组合而成。

3.3 与主流算法的对比试验

为了验证所提模型的有效性,本文将其与 10 种代表性的息肉分割算法进行对比,涵盖了经典的 CNN 架构(如 Unet^[4],Unet++^[34],SFA^[35])、基于注意力的网络(如 PraNet^[6],SANet^[7],MSNet^[36])以及最前沿的视觉 Transformer 模型(如 PVT-Cascade^[13],TransFuse^[14],SSFormer^[37])。表 1~表 4 展示了各模型在四个公开基准数据集上的定量评估结果,最佳性能以粗体显示,次优性能以下划线标示。

表 1 定量比较结果(Kvasir-SEG 数据集)

Tab. 1 Quantitative comparison results(Kvasir-SEG dataset)

DataSet	Methods	mIoU	Dice	S_a	HD95
Kvasir-SEG (seen)	Unet	0.771	0.857	0.856	69.228
	Unet++	0.779	0.862	0.867	68.338
	ACSNet	0.811	0.878	0.893	31.587
	SFA	0.656	0.772	0.707	81.482
	PraNet	0.844	0.901	0.908	35.301
	PVT-Cascade	0.863	<u>0.914</u>	0.928	28.621
	TransFuse	<u>0.864</u>	0.915	0.931	25.159
	SSFormer	0.856	0.907	0.922	27.055
	MSNet	0.855	0.907	0.923	28.281
	SANet	0.821	0.886	0.905	34.172
	Ours	0.865	0.915	<u>0.929</u>	<u>26.651</u>

表 2 定量比较结果(ClinicDB 数据集)

Tab. 2 Quantitative comparison results(ClinicDB dataset)

DataSet	Methods	mIoU	Dice	S_a	HD95
ClinicDB (seen)	Unet	0.774	0.865	0.871	51.383
	Unet++	0.784	0.869	0.874	58.017
	ACSNet	0.778	0.862	0.889	38.387
	SFA	0.607	0.709	0.683	89.033
	PraNet	0.861	0.915	0.891	39.707
	PVT-Cascade	<u>0.869</u>	0.919	0.934	11.682
	TransFuse	0.863	0.912	0.916	14.416
	SSFormer	0.845	0.899	<u>0.939</u>	<u>11.945</u>
	MSNet	0.862	<u>0.916</u>	0.941	15.998
	SANet	0.862	0.914	0.938	12.747
	Ours	0.883	0.927	0.937	13.211

表 3 定量比较结果(ColonDB数据集)

Tab. 3 Quantitative comparison results(ColonDB dataset)

DataSet	Methods	mIoU	Dice	S_a	HD95
ColonDB (unSeen)	Unet	0.501	0.608	0.715	98.254
	Unet++	0.522	0.638	0.759	101.354
	ACSNet	0.665	0.743	0.818	58.703
	SFA	0.426	0.541	0.571	138.34
	PraNet	0.667	0.751	0.819	45.375
	PVT-Cascade	0.689	<u>0.766</u>	0.845	33.905
	TransFuse	<u>0.691</u>	0.765	<u>0.846</u>	41.335
	SSFormer	0.656	0.751	0.834	52.369
	MSNet	0.655	0.735	0.821	52.896
	SANet	0.654	0.728	0.819	33.799
	Ours	0.695	0.771	0.847	<u>33.874</u>

表 4 定量比较结果(ETIS数据集)

Tab. 4 Quantitative comparison results(ETIS dataset)

DataSet	Methods	mIoU	Dice	S_a	HD95
ETIS (unSeen)	Unet	0.367	0.461	0.662	145.681
	Unet++	0.379	0.467	0.672	125.369
	ACSNet	0.584	0.644	0.792	79.485
	SFA	0.321	0.404	0.515	184.942
	PraNet	0.612	0.684	0.812	59.736
	PVT-Cascade	<u>0.687</u>	<u>0.761</u>	<u>0.839</u>	<u>47.359</u>
	TransFuse	0.657	0.733	0.842	51.339
	SSFormer	0.637	0.711	0.824	53.741
	MSNet	0.601	0.669	0.806	77.774
	SANet	0.623	0.692	0.831	58.106
	Ours	0.701	0.778	0.851	41.811

结果可知,在参与训练的数据集(Kvasir-SEG 与 CVC-ClinicDB)中,GPDP-Net 性能稳健,在 CVC-ClinicDB 上达到 0.883 的 mIoU 与 0.927 的 Dice 系数,与 PVT-Cascade 等先进模型相比具有竞争力。在从未参与训练的测试集(CVC-ColonDB 与 ETIS-LaribPolypDB)中,模型展现出了一定的泛化优势。其中,CVC-ColonDB 上的 mIoU(0.695)与 S_a (0.847)指标均略优于对比算法,反映出物理与几何先验在未知环境下的良好适应性。针对包含大量微小病灶的 ETIS 数据集,模型 mIoU 提升至 0.701(约 1.4%增益),其性能提升主要得益于 PDDEH 模块对微弱信号的频域补偿与物理去噪;同时,HD95 指标的下降也证实了 OEM 模块在优化边界对齐精度、

缓解微小病灶边缘模糊方面的积极作用。

3.4 消融实验

为验证各组件贡献,本文在 Kvasir-SEG 与 ETIS 数据集上进行消融实验,结果如表 5 所示。引入优化边缘模块(OEM)显著改善了边界精度,使 ETIS 的 HD95 由 135.21 骤降至 85.12,证实显式几何先验能有效纠正骨干网络下采样导致的边缘坍塌。叠加物理-频域并行的双域瓶颈增强模块(PDDEH)后性能大幅爬升,ETIS 的 mIoU 由 0.525 跃升至 0.668,增益达 14.3%,这主要得益于物理分支(PINN)对背景伪影的平滑以及频域分支(FFT)对微弱信号的频谱补偿。最后,完整架构引入 SR-IRAD 解码器后性能达到最优,ETIS 的 mIoU 升至 0.701,HD95 进一步压低至 41.81。这验证了亚像素还原算子在提升分辨率时对特征强度的保留作用,以及迭代反向注意力对边界残差的精炼效果。综上,消融实验证明 GPDP-Net 各模块共同提升了模型在复杂背景下对多尺度病灶的分割稳健性。

表 5 消融实验结果

Tab. 5 Ablation experiment

Methods	DataSets	mIoU	Dice	S_a	HD95
Baseline	Kvasir-SEG	0.785	0.862	0.868	65.42
	ETIS	0.412	0.521	0.684	135.21
+OEM	Kvasir-SEG	0.812	0.884	0.892	42.16
	ETIS	0.525	0.635	0.742	85.12
+PDDEH	Kvasir-SEG	0.846	0.902	0.915	32.55
	ETIS	0.668	0.752	0.835	48.33
Ours	Kvasir-SEG	0.865	0.915	0.929	26.65
	ETIS	0.701	0.778	0.851	41.81

3.5 可视化分析

为了直观验证所提模型 GPDP-Net 的优越性,图 5 展示了本文方法与 10 种代表性算法在四个基准数据集上的定性对比结果。

在针对极小病灶的检测中(如 ETIS-LaribPolypDB 数据集所示第 5,第 6 行),由于病灶尺寸极小且与周围组织对比度极低,传统卷积架构如 UNet,UNet++ 以及部分注意力模型出现了严重的漏诊现象。相比之下,本文方法能够精准锁定微小目标并完整提取其几何轮廓,这有力印证了叠加物理-频域并行的双域瓶颈增强模块

(PDDEH)中频域分支对微弱信号的频谱补偿能力,以及SR-IRAD解码器通过亚像素还原对空间细节的保真作用。

在处理形状复杂且伴有强烈反光的场景时(如CVC-ClinicDB与Kvasir-SEG数据集),对比模型易受光学伪影误导,导致分割边缘破碎或产生空洞。而本文方法生成的掩模边缘平滑且与真实标签高度贴合,视觉上证实了优化边缘模块(OEM)捕获的显式几何先验以及物理演化分支

(PINN)通过物理一致性约束抑制非相干反光噪声的卓越能力。

此外,在纹理复杂的CVC-ColonDB场景下,本文方法表现出更强的辨识力,能有效过滤肠道褶皱引起的伪边缘,确保模型始终聚焦于真实病灶区域。综上所述,GPDP-Net不仅在常规尺度病灶上具备更精细的边界刻画能力,在应对极具挑战性的微小病灶定位与光学伪影抑制方面,也表现出显著优于当前主流算法的鲁棒性。

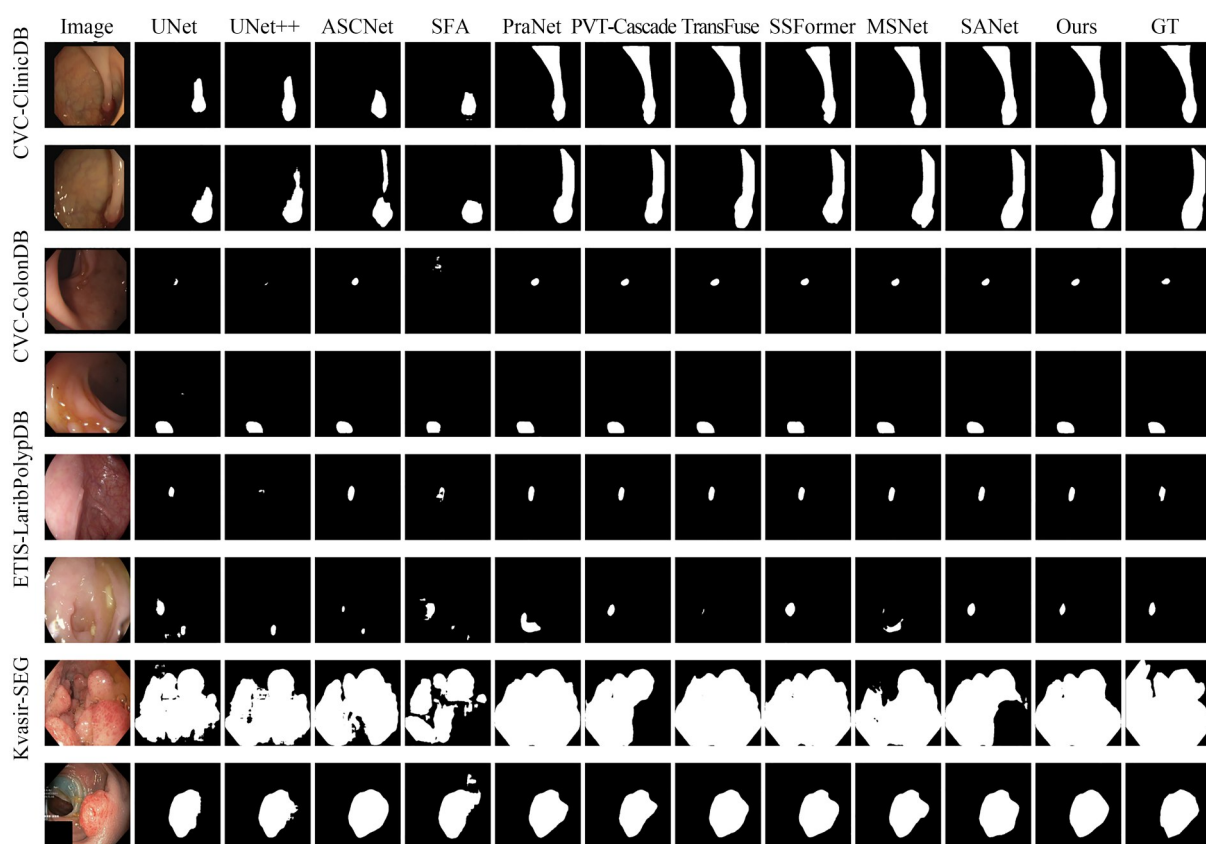


图5 本文方法与主流算法在不同数据集上的定性对比结果

Fig. 5 Qualitative comparison results of the proposed method and other state-of-the-art algorithms on different datasets

4 结 论

本文提出的融合几何先验与双域特征提纯的结直肠息肉分割网络(GPDP-Net),通过多维度的特征挖掘与物理规律引导,有效解决了复杂内镜环境下微小病灶识别困难、光学伪影干扰严重以及边界特征坍塌等关键挑战。实验数据表明,GPDP-Net在四个公开基准数据集上的性

能表现稳健,特别是在包含大量极小尺度目标的ETIS数据集上,模型在保持高检测率的同时,显著降低了边界对齐误差。

本研究的核心论点在于,深度学习模型的性能提升不仅依赖于网络架构的加深或注意力机制的堆叠,更取决于能否将成像过程中的物理先验与底层特征提取逻辑有机结合。实验结果证实,通过状态空间模型模拟特征演化并引入扩散

方程作为物理约束,能够显著抑制肠道反光产生的非相干噪声;而并行的频域增强分支则从频谱空间补偿了下采样过程造成的细节损失。这种双域提纯机制赋予了模型更强的特征表征能力,使其在面对未曾参与训练的新图像域时展现出优异的泛化性能。与当前主流的卷积神经网络及视觉 Transformer 架构相比,本文方法在处理低对比度样本和模糊边缘方面具有明显优势,其结构度量指标与边界距离指标的同步提升说明了模型在捕捉息肉全局拓扑结构与局部精细轮廓上的均衡性。此外,本文提出的亚像素还原策略有效避免了传统插值算法对微弱信号的稀释效应,为极小病灶的保真还原提供了可靠的技术途径。

综上所述,GPDP-Net 为结直肠息肉的自动

化分割提供了兼具物理可解释性与语义精确性的新架构。尽管本工作在多个基准测试中取得了竞争力较强的结果,但在极度复杂且伴有大幅度相机运动的视频流处理上,模型的计算效率仍存在优化空间。未来的研究将重点探索物理演化算子的轻量化实现,并尝试将该多维协同优化框架推广至多类别病种筛查等更广泛的医学影像诊断任务中。

作者贡献声明:

李梁:论文方法的提出,论文构思和撰写;

毕利:论文审核与编辑写作;

焦小刚:论文审核与编辑写作;

毕逢辰:提供资源与验证;

王俊杰:提供资源与数据分析。

参考文献:

- [1] ZAUBER A G, WINAWER S J, O'BRIEN M J, *et al.* Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths [J]. *The New England Journal of Medicine*, 2012, 366 (8): 687-696.
- [2] AHN S B, HAN D S, BAE J H, *et al.* The miss rate for colorectal adenoma determined by quality-adjusted, back-to-back colonoscopies [J]. *Gut and Liver*, 2012, 6(1): 64-70.
- [3] FAN D P, JI G P, ZHOU T, *et al.* PraNet: parallel reverse attention network for polyp segmentation [C]. *Medical Image Computing and Computer Assisted Intervention-MICCAI 2020*. Cham: Springer, 2020: 263-273.
- [4] 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.
- [5] JHA D, SMEDSRUD P H, RIEGLER M A, *et al.* ResUNet++: an advanced architecture for medical image segmentation [C]. 2019 *IEEE International Symposium on Multimedia (ISM)*. December 9-11, 2019, San Diego, CA, USA. IEEE, 2020: 225-2255.
- [6] DONG B, WANG W, FAN D P, *et al.* Polyp-PVT: Polyp segmentation with pyramid vision transformer [J]. *Computerized Medical Imaging and Graphics*, 2023, 102: 102143.
- [7] WEI J, HU Y W, ZHANG R M, *et al.* Shallow attention network for polyp segmentation [C]. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2021*. Cham: Springer, 2021: 699-708.
- [8] CHEN J N, LU Y Y, YU Q H, *et al.* TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation [EB/OL]. 2021; *arXiv*: 2102.04306. <https://arxiv.org/abs/2102.04306>
- [9] CAO H, WANG Y Y, CHEN J, *et al.* Swin-unet: Unet-like Pure Transformer for Medical Image Segmentation [M] *Computer Vision - ECCV 2022 Workshops*. Cham: Springer Nature Switzerland, 2023: 205-218.
- [10] HATAMIZADEH A, TANG Y C, NATH V, *et al.* UNETR: transformers for 3D medical image segmentation [C]. 2022 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. January 3-8, 2022, Waikoloa, HI, USA. IEEE, 2022: 1748-1758.
- [11] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.* An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale [EB/OL]. 2020; *arXiv*: 2010.11929. <https://arxiv.org/abs/2010.11929>
- [12] WANG W H, XIE E Z, LI X, *et al.* Pyramid vision transformer: a versatile backbone for dense

- prediction without convolutions [C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021, Montreal, QC, Canada. IEEE, 2022: 548-558.
- [13] DONG X, YU J, GE W, *et al.* Polyp-PVT: Polyp segmentation with pyramid vision transformer [J]. *arXiv preprint arXiv:2108.06935*, 2021.
- [14] 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.
- [15] AKBARI M, MOHAMMADI M, FATHY M, *et al.* Specular reflection detection and inpainting in colonoscopy images [C]. 2015 *IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, 2015: 454-459.
- [16] BERNAL J, TAJKBAKSH N, SÁNCHEZ F J, *et al.* Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge [J]. *IEEE Transactions on Medical Imaging*, 2017, 36(6): 1231-1249.
- [17] PERONA P, MALIK J. Scale-space and edge detection using anisotropic diffusion [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12(7): 629-639.
- [18] RAISSI M, PERDIKARIS P, KARNIADAKIS G E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations [J]. *Journal of Computational Physics*, 2019, 378: 686-707.
- [19] STEHLE T. Removal of specular reflections in endoscopic images [C]. *Proceedings of the SPIE 6144, Medical Imaging 2006: Image Processing*. San Diego: SPIE, 2006: 61443L.
- [20] 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.
- [21] SCHARR H. *Optimal Operators in Digital Image Processing* [D]. Heidelberg: University of Heidelberg, 2000.
- [22] KARNIADAKIS G E, KEVREKIDIS I G, LU L, *et al.* Physics-informed machine learning [J]. *Nature Reviews Physics*, 2021, 3(6): 422-440.
- [23] GU A, GOEL K, RÉ C. Efficiently Modeling Long Sequences with Structured State Spaces [EB/OL]. 2021: *arXiv: 2111.00396*. <https://arxiv.org/abs/2111.00396>
- [24] GONZALEZ R C, WOODS R E. *Digital Image Processing* [M]. 4th ed., Global ed. New York: Pearson, 2018.
- [25] COOLEY J W, TUKEY J W. An algorithm for the machine calculation of complex Fourier series [J]. *Mathematics of Computation*, 1965, 19(90): 297-301.
- [26] XU W, XU R, WANG C, *et al.* PSTNet: Enhanced polyp segmentation with multi-scale alignment and frequency domain integration [J]. *IEEE Journal of Biomedical and Health Informatics*, 2024, 28(10): 6042-6053.
- [27] JHA D, SMEDSRUD P H, RIEGLER M A, *et al.* Kvasir-SEG: A Segmented Polyp Dataset [C]. *MultiMedia Modeling*. Cham: Springer, 2020: 451-462.
- [28] BERNAL J, SÁNCHEZ F J, VILARINO G, *et al.* WM-DOCA: A new polyp log-on-manifold description for polyp detection in colonoscopy [J]. *Computerized Medical Imaging and Graphics*, 2015, 40: 105-117.
- [29] BERNAL J, VILARINO G, SÁNCHEZ F J, *et al.* Polyp segmentation method in colonoscopy images based on seeds propagation and deformable models [J]. *Computerized Medical Imaging and Graphics*, 2012, 36(5): 392-403.
- [30] SILVA J S, HISTACE A, ROMAIN O, *et al.* Toward automated polyp detection in colonoscopy videos [C]. *Proceedings of the 27th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2014: 432-437.
- [31] CHENG M M, FAN D P. Structure-measure: a new way to evaluate foreground maps [J]. *International Journal of Computer Vision*, 2021, 129(9): 2622-2638.
- [32] TAHA A A, HANBURY A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool [J]. *BMC Medical Imaging*, 2015, 15(1): 29.
- [33] LOSHCHILOV I, HUTTER F. SGDR: stochastic gradient descent with warm restarts [C]. *International Conference on Learning Representations*. Toulon, France, 2017
- [34] ZHOU Z W, RAHMAN SIDDIQUEE M M,

- TAJBAKHS 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, 2018: 3-11.
- [35] FANG Y Q, CHEN C, YUAN Y X, *et al.* Selective feature aggregation network with area-boundary constraints for polyp segmentation [C]. *Medical Image Computing and Computer Assisted Intervention-MICCAI 2019*. Cham: Springer, 2019: 302-310.
- [36] ZHAO X Q, ZHANG L H, LU H C. Automatic Polyp segmentation via multi-scale subtraction network [C]. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2021*. Cham: Springer, 2021: 120-130.
- [37] SHI W T, XU J, GAO P. SSformer: A Lightweight Transformer for Semantic Segmentation [EB/OL]. 2022; *arXiv*: 2208.02034. <https://arxiv.org/abs/2208.02034>

作者简介:



李 梁(2002—),男,四川内江人,硕士,2024年于乐山师范学院获得学士学位,2024年于宁夏大学攻读硕士研究生,主要研究方向为计算机视觉及其自动化。E-mail: Lee_shcolar7@163.com

通讯作者:



毕 利(1968—),女,宁夏银川人,硕士,教授,硕士生导师,1989年于宁夏大学获得学士学位,2002年于华东理工大学获得硕士学位,主要研究方向为信息系统分析与集成、数据挖掘及智能工厂调度。E-mail: billy68@nxu.edu.cn