

# 基于内容引导异构双解码器的息肉图像分割

王丽黎<sup>1,2</sup>, 郑益聪<sup>1</sup>, 郭东旭<sup>1</sup>, 王振宁<sup>1</sup>

(1. 西安理工大学 自动化与信息工程学院, 陕西 西安 710048;  
2. 西安市无线光通信与网络研究重点实验室, 陕西 西安 710048)

**摘要:** 针对结直肠图像中息肉尺寸大小不一、边界模糊以及内窥镜图像清晰度受限等问题, 提出了一种基于内容引导异构双解码器的特征融合网络 (HCGFNet)。HCGFNet 中编码器网络采用异构多路径自适应特征融合模块 (HAF), 通过异构多数据流更精准地捕获复杂肠道环境中隐匿的各类小型息肉与周边特征信息。解码器网络采用内容引导特征融合注意力机制 (CGFA), 逐层处理解码阶段特征图中干扰信息, 细化目标边缘分割效果并辅助重建灰度图。最终分别在 KvasirSEG、CVC-ClinicDB、CVC-ColonDB 息肉数据集上进行广泛对比, 结果表明, 所设计的 HCGFNet 相较于目前主流模型, 在各项性能中均有提升。引入 HAF、CGFA 模块后, 各项性能较基准模型提升 2% ~ 5%, 较最先进模型提升 1% ~ 2%。

**关键词:** 医学图像分割; 结肠息肉; 异构多路径; 特征融合; 灰度重建

中图分类号: TP391.41

文献标志码: A

DOI: 10.19358/j.issn.2097-1788.2026.02.005

**中文引用格式:** 王丽黎, 郑益聪, 郭东旭, 等. 基于内容引导异构双解码器的息肉图像分割 [J]. 网络安全与数据治理, 2026, 45(2): 34-43.

**英文引用格式:** Wang Lili, Zheng Yicong, Guo Dongxu, et al. Content-guided heterogeneous dual-decoder network for polyp image segmentation [J]. Cyber Security and Data Governance, 2026, 45(2): 34-43.

## Content-guided heterogeneous dual-decoder network for polyp image segmentation

Wang Lili<sup>1,2</sup>, Zheng Yicong<sup>1</sup>, Guo Dongxu<sup>1</sup>, Wang Zhenning<sup>1</sup>

(1. School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China;  
2. Key Laboratory of Wireless Optical Communication and Network Research, Xi'an 710048, China)

**Abstract:** In response to issues such as varying polyp sizes, blurred boundaries in colorectal images, and limited clarity in endoscopic images, this paper proposes a feature fusion network based on content-guided heterogeneous dual decoders (HCGFNet). In HCGFNet, the encoder network employs the proposed Heterogeneous Multi-path Adaptive Feature Fusion module (HAF), which captures concealed small polyps and surrounding feature information in complex intestinal environments more accurately through heterogeneous multi-data streams. The decoder network utilizes the designed Content-Guided Feature Fusion Attention mechanism (CGFA) to process interfering information in feature maps during the decoding phase layer by layer, refining the segmentation of target edges and assisting in grayscale image reconstruction. Extensive comparative experiments conducted on the Kvasir-SEG, CVC-ClinicDB, and CVC-ColonDB polyp datasets demonstrate that the proposed HCGFNet outperforms current mainstream models across various performance metrics. After incorporating the HAF and CGFA modules, performance improvements range from 2% to 5% compared to the baseline model and from 1% to 2% compared to state-of-the-art models.

**Key words:** medical image segmentation; colorectal polyps; heterogeneous multipath; feature fusion; grayscale reconstruction

## 0 引言

目前, 结直肠癌是全球最常见的恶性肿瘤之一, 其发病率和死亡率也在众多肿瘤中位居前列。肠息肉作为结肠癌前期主要病变因素, 发现并切除可以有效阻断结直肠癌的发生, 但结肠镜检查仍存在漏诊的现

象, 其中以锯齿状病变息肉 (Sessile Serrated Lesions with dysplasia, SSL) 为主。由于 SSL 多数具备形状不规则、边界模糊、尺寸微小等特点, 需要一个自动化系统辅助医生完成息肉检测。

传统医学图像分割术主要包括阈值法<sup>[1]</sup>、区域生

长法<sup>[2]</sup>、边缘检测法<sup>[3]</sup>以及基于聚类<sup>[4]</sup>的分割算法等,但在实际临床诊断中,肠息肉常隐匿于肠道组织中,且表面易被粘液或粪便所覆盖,提取到的浅层特征不足以准确定位息肉。直到近年随着深度学习技术的飞速发展,全卷积网络(Fully Convolutional Networks, FCN)<sup>[5]</sup>可通过像素级别的预测来分割息肉,然而,基于FCN的方法对细节特征恢复以及多尺度特征处理仍存在不足。自2015年,基于编解码器的网络模型因其卓越的性能,逐渐占据图像分割领域的主导地位,如U-Net<sup>[6]</sup>、U-Net++<sup>[7]</sup>和ResUNet++<sup>[8]</sup>等,但在面临形状复杂、纹理模糊的微型SSL息肉时,上述模型的分割性能依旧受限。在此前提下,对模型的提升应聚焦于捕捉息肉周围更多的上下文和语义信息。鉴于此,Wang等人<sup>[9]</sup>设计了多尺度特征融合单元,合并息肉的上下文信息与边界细节,从而使模型能够适应各种尺寸的息肉特征;Li等人<sup>[10]</sup>提出了MSCFF-Net模型,通过多尺度注意力机制抑制了图像背景噪声的干扰;Qin等人<sup>[11]</sup>提出了FFA-Net,通过引入空间注意力机制生成空间重要性图,该指标能够自适应地区分不同像素位置的重要级别,提升模型对重要特征提取能力的同时抑制并去除干扰信息,但并未考虑到解码阶段分配至每一特征层中的冗余信息;为了抑制此干扰信息,Chen等人<sup>[12]</sup>提出的DEA-Net对编码阶段特征进行优化,获得更加稳定的特征掩码。

基于以上研究,为了更好地捕捉SSL肠息肉,本文提出了一种新的肠镜息肉分割方法,该方法主要通

过增强上下采样过程中多尺度特征的捕获能力以及对重要特征信息的还原来提高分割精度。本文贡献如下:

(1) 提出一种基于内容引导异构双解码器结构的多尺度特征融合网络(Hetero-path Content Guidance Feature Fusion Net, HCGFNet),利用自编码器网络和注意力机制辅助模型对目标特征信息的捕捉与恢复,并通过跳跃连接充分融合低级空间信息与高级语义空间。

(2) 提出一种异构多路径自适应特征融合模块(Hetero-path Adaptive Feature Fusion, HAF),利用多路并联的结构与注意力机制来精确提取息肉附近上下文特征信息,从而区分肠息肉与肠内壁。

(3) 提出一种内容引导特征融合注意力机制(Content-Guided Feature Fusion Attention, CGFA),对每一条通道分配独一无二的重要性权重系数,实现对重要特征信息的引导。

在实验中添加各个模块后,均大幅提升了息肉边界的分割精度。在Kvasir-SEG、CVC-ClinicDB、CVC-ColonDB三个数据集中均取得了良好的效果。

## 1 网络模型

### 1.1 网络结构

针对SSL肠息肉复杂多变的形态以及与周围黏膜组织高度相似导致边界轮廓模糊的问题,本文提出HCGFNet应用于结肠息肉分割,整个网络由一组编码器网络和两组解码器网络构成,编码器网络为两组解码器网络所共用,网络结构如图1所示。

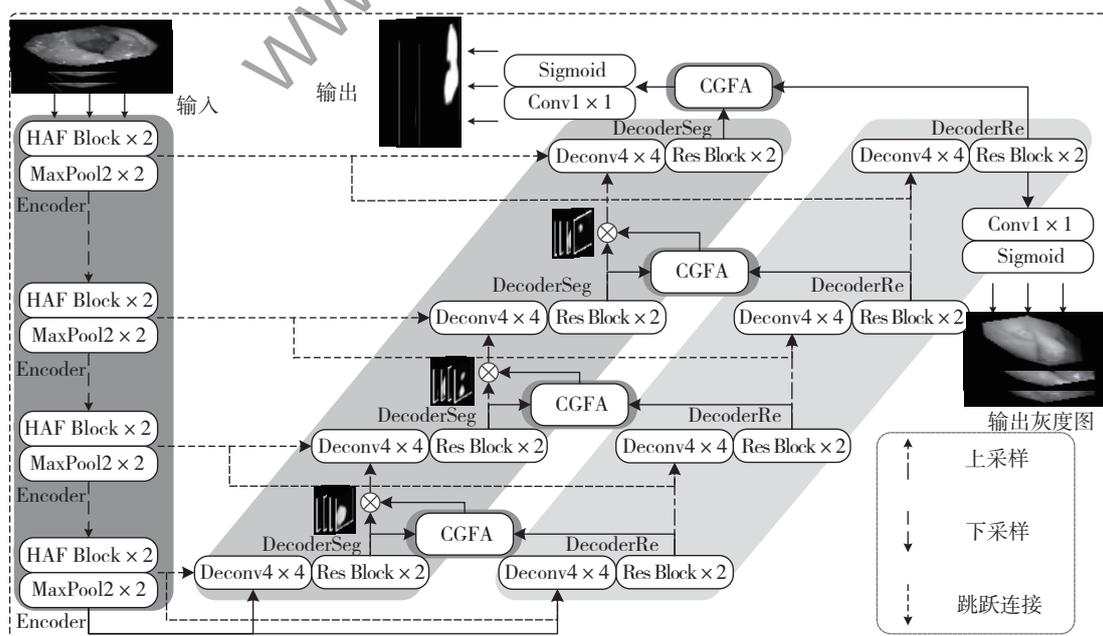


图1 内容引导异构双解码器特征融合网络结构

受多分支解码架构的启发, HCGFNet 设计了分割解码分支 (DecoderSeg)、灰度重建解码分支 (DecoderRe)。其中 DecoderRe 通过重建灰度图辅助 DecoderSeg 实现精准分割。考虑到现有多分支架构在编码阶段提取特征方式单一, 在解码阶段, 现有网络模型对特征图的优化仅停留在原始图像层面, 忽略了采样过程中被编码至各个阶段特征层的干扰信息, 外加内窥镜影像中本身存在雾化、噪声等干扰信息, 使得息肉分割效果在一定程度上受限。因此, HCGFNet 在编码阶段, 通过 HAF 模块捕捉多尺度特征信息的同时自适应调节各个通道的权重系数, 提取更加丰富的细节特征。在解码阶段, 通过 CGFA 模块去除解码层中干扰特征信息, 引导模型关注重要特征, 为每一通道分配权重系数, 辅助分割段, 得到最终的预测特征图。

首先, 编码器网络接收原始输入息肉图像后, 提取特征的同时逐层进行下采样, 将编码信息分别传输给两组解码器网络 DecoderSeg 与 DecoderRe。随后, 将压缩后的特征图在空间维度加倍, 同时利用跳跃连接的方式将编码器与解码器相同维度的特征图进行特征映射<sup>[13]</sup>拼接, 优化梯度流动。再以双支解码器输出的分割掩码与重建灰度图作为输入, 传输到 CGFA 模块中, 得到特征权重系数引导分割解码器作为输入再传输到 DecoderSeg 的下一级解码器, 以此类推, 直至将最后一层 CGFA 模块的输出经过特征激活模块, 生成最后的分割掩码。DecoderRe 则输出最终生成的重建灰度图。

模型中编码器网络由 4 个编码器模块组成, 每个模块先经过两次 HAF 模块实现动态特征筛选, 提取到高精度特征后下采样。两组解码器网络也均由 4 个解码器模块组成, 每个模块由  $4 \times 4$  转置卷积以及两个残差模块组成, 不仅能够还原特征图的空间维度, 复原

在下采样过程中可能丢失的部分特征, 还能通过两个残差模块学习网络在反向传播中所需的重要特征, 提高模型对息肉边缘特征的捕捉能力。

### 1.2 异构多路径自适应特征融合模块 (HAF)

由于肠息肉具有尺度各异、形状不一的特性, 单路径、单尺度卷积核因其感受野受限, 捕获上下文信息能力较弱, 难以同时对息肉边缘特征与整体特征信息的捕获。因此, 本节提出了 HAF 模块, 采用多路径、多尺度卷积核异构并行的连接架构, 有效解决了上述缺陷, 应用不同路径针对性学习息肉不同尺度边缘特征信息, 并将不同卷积分支提取的特征信息在通道维度上进行拼接, 结合 SE 注意力机制二次提取全局特征信息与局部细节特征, 融合多尺度特征信息, 获取丰富语义信息。

具体而言, HAF 采取主路径与旁路径融合的方式提取特征信息, 结合实例归一化和参数修正单元加速网络收敛, 高效精确地捕捉 SSL 息肉特征, 结构图如图 2 所示。主路径采用 5 路异构并行数据流的结构, 通过不同形状卷积核实现对不同层次不同尺度的特征图进行捕捉。令输入特征为  $X$ , 各卷积路径的输出为  $Y_i$  ( $X, Y_i \in \mathbf{R}^{C \times H \times W}$ ), 式中,  $H$  和  $W$  分别表示高和宽,  $C$  表示通道数,  $i$  表示从左至右路径编号,  $i \in \{1, 2, 3, 4, 5\}$ 。

路径 1 保留输入原始分辨率特征信息的同时避免丢失高频信息。路径 2、3 专注处理边缘、纹理等细节特征<sup>[14]</sup>, 通过调整通道间的依赖关系, 实现跨通道特征信息融合, 增强通道间的协同工作。路径 2、3 级联合合并构成路径 4, 重组特征通道间的权重关系并通过两次归一化处理对细节特征重提取, 确保特征分布平稳, 抑制深层网络中梯度异常的问题。路径 5 在输入、输出特征空间维度不变的前提下, 增强了模型对息肉

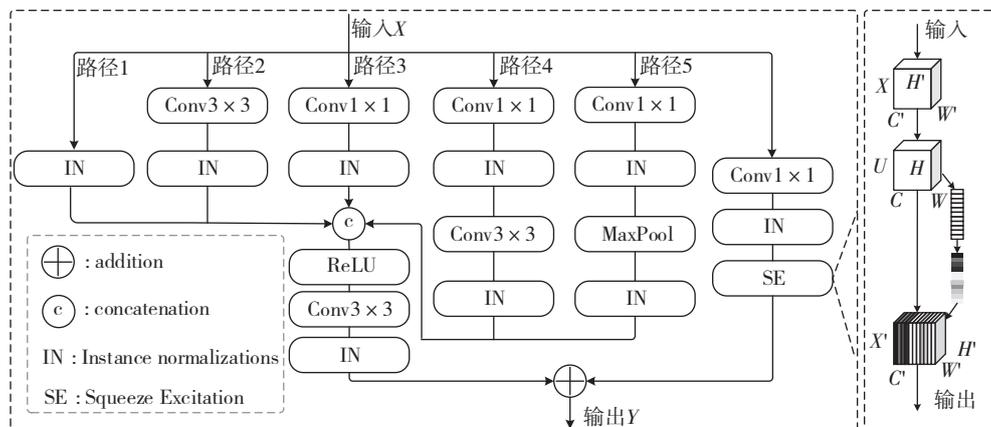


图 2 异构多路径自适应特征融合模块

边缘细节变化的鲁棒性，并通过减少参数量来防止过拟合，从而保证模型的泛化能力且便于与其他路径输出的特征进行通道维度拼接。

对上述五条路径输出特征在通道维度进行融合得  $Y_{\text{fused}} = \sum_1^5 Y_k$ ,  $Y_{\text{fused}}$  经 ReLU 激活函数抑制负相位响应后通过卷积层进行空间调制，最后由归一化层消除域偏移，输出精炼特征  $Y_{\text{refined}}$ 。

$$Y_{\text{refined}} = \text{IN} (C_{3 \times 3} (\varphi (Y_{\text{fused}}))) \quad (1)$$

式中， $\text{IN}(\cdot)$  表示实例归一化运算， $C_{j \times j}(\cdot) \in \mathbf{R}^{j \times j \times C \times C}$  表示卷积核大小为  $j \times j$  的卷积层， $\varphi(\cdot)$  表示 ReLU 激活函数， $\varphi(x) = \max(0, x)$ 。

五条路径分别对应了恒等映射、空间卷积、通道交互、级联卷积以及池化增强的功能。同时，批量引入实例归一化层增加模型稳定性提高模型的泛化能力。

旁路引入残差连接结构，有效缓解伴随网络深度的增加而出现的梯度异常问题，促进深层特征与浅层特征的融合。再经过通道注意力机制 SE (Squeeze-and-Excitation) 模块动态调整特征通道的权重，强化重要特征通道表达的同时抑制冗余信息，更易于捕捉微型息肉边界特征。

$$Y_{\text{SEbf}} = \text{IN} (C_{1 \times 1} (X)) \quad (2)$$

$$Y_{\text{SE}} = F_s (W_2 \cdot \delta (W_1 \cdot \text{GAP} (Y_{\text{SEbf}}))) \odot Y_{\text{SEbf}} \quad (3)$$

式中， $Y_{\text{SEbf}}$  表示输入特征经过  $1 \times 1$  卷积计算与归一化层后的输出， $F_s(\cdot)$  为 Sigmoid 激活函数， $W_1 \in \mathbf{R}^{C/r \times C}$ ， $W_2 \in \mathbf{R}^{C \times C/r}$  均表示全连接层权重， $\text{GAP}(\cdot)$  表示全局平均池化， $\odot$  表示通道缩放。最后，对双路径进行残差相加得到最终融合特征  $Y$ 。

$$Y = Y_{\text{refined}} + Y_{\text{SE}} \quad (4)$$

综上，针对肠道息肉边界形态走势复杂、大小差异较大等问题，该模型结合多路异构、多尺度并行的模型架构，实现不同路径对不同尺度边缘特征信息的针对性提取，从而提高息肉边缘分割效果。

### 1.3 内容引导特征融合注意力机制 (CGFA)

经典通道空间注意力机制，对空间、通道注意力进行单一串、并联交互，受其固有框架限制，在微小目标分割中难以达到临床标准。本文提出了新的通道空间注意力机制 CGFA，通过压缩空间、通道维度的方式，大幅提高模型对通道、空间维度上特征信息的关注度，着重提升模型对息肉边缘细节特征与微小隐匿型息肉位置特征的捕获能力。同时，摒弃了传统通道、空间注意力间单一并联、级联的交互方式，CGFA 在并联交互后，通过 Shuffle 操作在通道维度上进行打乱重排，增强通道与空间注意力权重信息的交互性的同

时使模型聚焦于每一通道的特征信息，使每条通道获取独一无二的空间重要性分布图 (Spatial Importance Map, SIM)，引导模型聚焦于各个通道，使 HCGFNet 在解码阶段恢复高精度特征信息，去除干扰信息，从而提升对 SSL 型息肉的分割精度。内容引导特征融合注意力机制模块如图 3 所示。

CGFA 模块由通道处理分支与空间处理分支两部分构成，其中，输入特征通过通道处理分支，将特征空间维度压缩至  $1 \times 1$ ，但在通道维度上保留高分辨率，对不同特征通道的重要性进行分析处理，聚焦于获取息肉外形边缘等特征信息，最终生成通道权重系数。同时，输入特征通过空间处理分支，将通道维度压缩至 1，保留了原有空间维度上的高分辨率，分析不同区域的重要性及其相互关联，聚焦于获取息肉在复杂环境中的具体位置，最终生成空间权重系数。

具体而言，通道处理分支通过两路  $1 \times 1$  卷积对输入特征  $X \in \mathbf{R}^{C \times H \times W}$  在通道维度上进行压缩，并通过 Reshape 函数对压缩后的三维张量重塑为二维矩阵，其中一路通过计算空间位置相似度矩阵，与另一路重塑二维矩阵点积，得到重加权特征矩阵，依此对通道注意力权重初次校准，最后还原通道数并结合输入特征  $X$  得到通道分支权重系数  $W_c$ 。同样，空间处理分支通过两路  $1 \times 1$  卷积对输入特征  $X$  在通道维度上进行压缩，其中一路通过对空间信息压缩后的张量进行非线性激活，突出重要通道信息，以此权重系数引导空间权重系数二次校准，最终恢复空间结构，结合输入特征  $X$  得到空间分支权重系数  $W_s$ 。

$$W_{c1} = C_{1 \times 1} (\varphi_1 (C_{1 \times 1} (X))) \odot F_s (\varphi_2 (C_{1 \times 1} (X))) \quad (5)$$

$$W_c = F_s [\text{LN} (C_{1 \times 1} (W_{c1}))] \otimes_c X \quad (6)$$

$$F_{\text{sm}} (X) = \sum_{j=1}^{N_p} \frac{e^{x_j}}{\sum_{m=1}^{N_p} e^{x_m}} x_j \quad (7)$$

$$W_{s1} = F_{\text{sm}} (\varphi_1 (F_{\text{gp}} (C_{1 \times 1} (X)))) \odot \varphi_2 (C_{1 \times 1} (X)) \quad (8)$$

$$W_s = F_s [\varphi_3 (W_{s1})] \otimes_s X \quad (9)$$

式中， $C_{K \times K}(\cdot)$  表示卷积核大小为  $K \times K$  的卷积层， $\varphi_1$ 、 $\varphi_2$ 、 $\varphi_3$  分别表示三个张量重塑操作符， $F_{\text{sm}}(\cdot)$  表示 Softmax 函数操作符， $\odot$  表示矩阵点积运算符， $F_s(\cdot)$  表示 Sigmoid 激活函数， $\text{LN}(\cdot)$  表示归一化处理， $\otimes_c$ 、 $\otimes_s$  表示通道乘法运算符与空间乘法运算符， $F_{\text{gp}}(\cdot)$  表示全局操作符。随后将空间权重系数与通道权重系数进行逐元素相加融合得到粗略融合系数  $W_{\text{cos}} \in \mathbf{R}^{C \times H \times W}$ 。

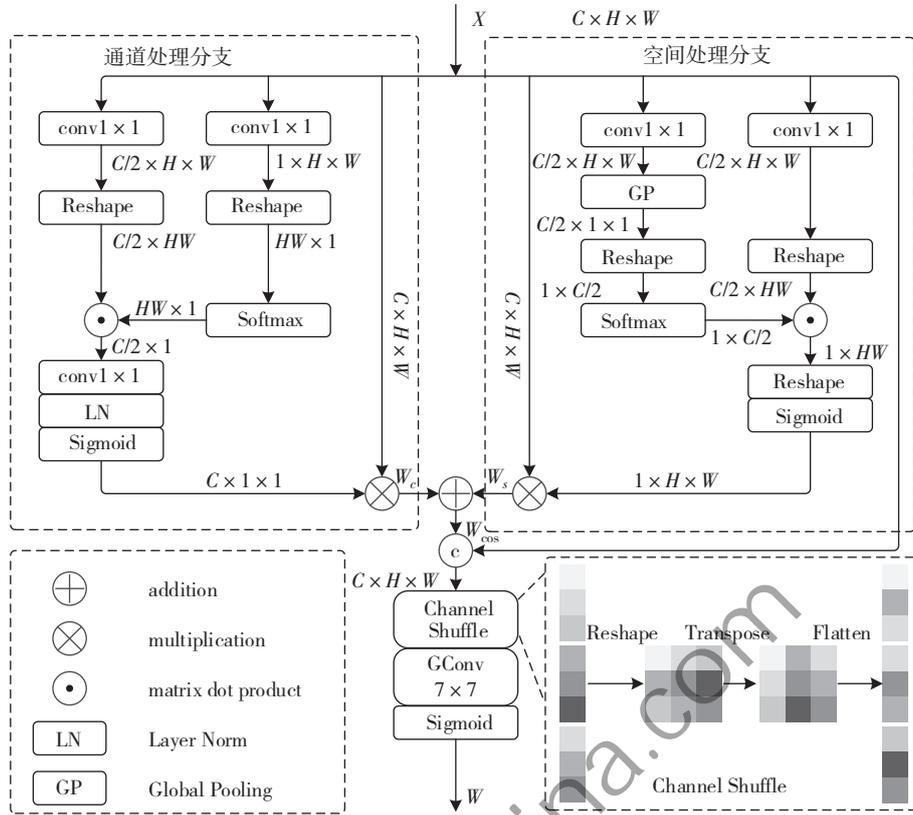


图3 内容引导特征融合注意力机制

$$W_{cos} = W_c + W_s \quad (10)$$

为了获取更精确的 SIM,  $W_{cos}$  结合原始输入特征  $X$ , 将两者通道拼接后随机排列组合, 大幅降低参数量的同时以原始特征图作为参考, 对通道权重系数再调整, 确保每条通道可获得独一无二的 SIM, 输出最终权重系数  $W$ , 引导模型聚焦各个通道的关键区域从而提升整个模型的息肉分割精确度。

$$W = F_s [C_{g7 \times 7} (CS([W_{cos}, X]))] \quad (11)$$

式中  $C_{gk \times k}(\cdot)$  表示内核为  $k$  的分组卷积层,  $CS(\cdot)$  代表通道随机排列组合操作。

综上, 模型通过对空间、通道维度上的压缩以及 Shuffle 操作, 大幅提高了对微小息肉位置、边缘特征的感知与捕获能力, 引导模型保留恢复重要特征信息。

## 2 实验结果与分析

### 2.1 实验环境与配置

本文提出的 HCGFNet 是基于 PyTorch 实现的, 网络训练使用随机梯度下降优化器, 初始学习率设置为 0.01, 动量设置为 0.9, 为了使模型更具鲁棒性, 设置权重衰减为 0.000 1。整个训练过程共 50 个轮次。输入图像分辨率修改为  $256 \times 256$ , 批大小设置为 16, 在

对比、消融实验中, 针对各模型均固定统一随机种子。硬件环境使用 GPU 为 RTX4070TiSuper 的服务器, 显存为 32 GB, 训练框架为 PyTorch1.13.1。

### 2.2 数据集

为了验证 HCGFNet 模型的性能, 采用三个公开肠道息肉数据集 Kvasir-SEG<sup>[15]</sup>、CVC-ClinicDB<sup>[16]</sup>、CVC-ColonDB<sup>[17]</sup>对模型进行全方位评估。

Kvasir-SEG 数据集共有 1 000 张经专家标注与核实的胃肠息肉图像, 包含各类大小以及不同分布的息肉图像信息, 具备较为广泛的分辨率 (从  $720 \times 560$  到  $1 920 \times 1 072$ )。选择 80% 的数据作为训练集, 20% 作为验证集。

CVC-ClinicDB 数据集是从结肠镜检查视频中提取的 612 张静态息肉图像, 每张图像的分辨率为  $384 \times 288$ 。其中息肉图像同样具有大小位置各不相同以及背景干扰丰富的特点。选择 80% 的数据作为训练集, 20% 作为验证集。

CVC-ColonDB 数据集由 15 个不同结肠镜检查影像中提取到 380 张静态帧作为数据图像, 每张图像的分辨率为  $574 \times 500$ 。其中包含扁平息肉、带蒂息肉等微小息肉类型, 能够较好验证模型对隐匿息肉的分割性能。选择 80% 的数据作为训练集, 20% 作为验证集。

## 2.3 评估指标

为了客观地评估与比较不同模型间的性能差别，遵循现有的 SOTA 方法，在分割结果上采用了五种不同的度量，包含精确率 (Precision)、召回率 (Recall)、F1 分数、分割精度 (IoU)、相似性系数 (Dice)。

## 2.4 实验结果分析

### 2.4.1 不同算法对比实验分析

为了验证 HCGFNet 的综合分割性能，将其与 10 种先进方法进行比较，包括 U-Net、DDANet<sup>[18]</sup>、TransUNet<sup>[19]</sup>、UNeXt<sup>[20]</sup>、TransCeption<sup>[21]</sup>、Rolling-Unet<sup>[22]</sup>、U-Mamba<sup>[23]</sup>、U-KAN<sup>[24]</sup>、CMUNeXt<sup>[25]</sup>、CSWin-UNet<sup>[26]</sup>。其中均采用训练集、验证集 8:2 的方式进行实验，最终得到的性能指标如表 1 ~ 表 3 所示，其中 F1 分数、Recall、Precision、IoU 和 Dice 指标数值越大表示息肉分割性能越好，加粗数值表示最优值，下划线数值表示次优值。

表 1 数据集 Kvasir-SEG 中各对比实验

方法	年份	F1	Recall	Precision	IoU	Dice
U-Net	2015	0.821 2	0.836 4	0.836 5	0.716 3	0.821 0
DDANet	2020	0.861 2	0.858 6	0.897 5	0.780 0	0.857 6
TransUNet	2021	0.875 4	0.864 3	0.881 3	0.793 2	0.868 8
UNeXt	2022	0.879 0	0.871 5	0.883 2	0.792 6	0.877 5
TransCeption	2023	0.878 3	0.871 2	0.879 9	0.806 4	0.874 3
Rolling-Unet	2024	<u>0.882 1</u>	0.875 1	0.885 4	0.807 9	0.876 7
U-Mamba	2024	0.881 5	0.876 2	0.884 6	0.811 2	0.879 3
U-KAN	2024	0.878 5	0.870 8	0.880 1	0.802 0	0.878 5
CMUNeXt	2024	0.878 1	0.908 9	<u>0.894 0</u>	0.758 7	0.851 3
CSWin-UNet	2025	0.880 7	<b>0.911 3</b>	0.889 0	<u>0.814 5</u>	<u>0.886 2</u>
本文	2025	<b>0.887 6</b>	<u>0.909 7</u>	<b>0.902 4</b>	<b>0.819 8</b>	<b>0.887 4</b>

表 2 数据集 CVC-ClinicDB 中各对比实验

方法	年份	F1	Recall	Precision	IoU	Dice
U-Net	2015	0.823 0	0.815 6	0.834 3	0.755 3	0.857 8
DDANet	2020	0.876 8	0.881 2	0.894 0	0.843 1	0.891 5
TransUNet	2021	0.891 3	0.894 7	0.901 5	0.858 9	0.901 1
UNeXt	2022	0.892 0	0.885 3	0.894 6	0.812 0	0.892 0
TransCeption	2023	0.910 5	<u>0.912 1</u>	0.915 7	<u>0.864 3</u>	0.903 0
Rolling-Unet	2024	0.891 2	0.894 0	0.887 0	0.854 8	0.902 0
U-Mamba	2024	0.910 3	0.902 8	0.898 3	0.842 3	0.901 4
U-KAN	2024	<u>0.917 6</u>	0.891 5	0.885 6	0.851 2	0.899 5
CMUNeXt	2024	0.883 0	0.875 7	0.880 4	0.835 2	0.885 7
CSWin-UNet	2025	0.916 7	0.905 2	<u>0.919 8</u>	0.862 5	<u>0.903 8</u>
本文	2025	<b>0.919 0</b>	<b>0.918 2</b>	<b>0.921 1</b>	<b>0.877 2</b>	<b>0.916 4</b>

表 3 数据集 CVC-ColonDB 中各对比实验

方法	年份	F1	Recall	Precision	IoU	Dice
U-Net	2015	0.512 3	0.510 4	0.615 7	0.430 0	0.510 3
DDANet	2020	0.745 8	0.735 2	0.762 4	0.645 7	0.721 1
TransUNet	2021	0.578 5	0.565 8	0.595 3	0.573 6	0.668 3
UNeXt	2022	0.672 9	0.680 4	0.710 5	0.617 9	0.692 4
TransCeption	2023	0.735 4	0.765 8	0.750 7	0.638 6	0.725 1
Rolling-Unet	2024	0.763 6	0.799 8	0.785 2	<u>0.721 8</u>	0.743 2
U-Mamba	2024	0.798 9	0.810 5	0.805 7	0.712 9	0.768 5
U-KAN	2024	0.810 5	0.800 3	0.811 0	0.702 7	0.740 1
CMUNeXt	2024	0.785 8	<u>0.835 7</u>	0.801 9	0.680 4	0.755 4
CSWin-UNet	2025	<u>0.815 7</u>	0.835 4	<b>0.830 6</b>	0.718 9	<u>0.789 3</u>
本文	2025	<b>0.828 9</b>	<b>0.855 3</b>	<u>0.825 8</u>	<b>0.732 7</b>	<b>0.798 5</b>

在 Kvasir-SEG 数据集上，本文模型与最新的 CSWin-UNet 相比虽在 Recall 系数上降低 0.16%，但在 Precision 上提升了 1.34%，F1 分数提升了 0.69%，IoU 系数提升 0.53%，Dice 系数提升 0.12%，相比于近年其他 SOTA 模型在各项指标系数上提升可达 1% ~ 2%。在 CVC-ClinicDB 数据集上，本文模型效果达到最优值，相较于近年 SOTA 网络模型，分割性能指标提升 1% ~ 3%。在 CVC-ColonDB 数据集上，本文模型与最新的 CSWin-UNet 相比虽在 Precision 系数上降低 0.48%，但在 F1 分数中提升了 1.32%，Recall 系数上提升 1.99%，IoU 系数提升 1.38%，Dice 系数提升 0.92%，相比于近年其他 SOTA 模型在各项指标系数上提升可达 1% ~ 4%。鉴于三个数据集均包含各类复杂肠道环境下大小不一的肠道息肉，足以验证本文模型同其他模型相比更好地完成了对肠息肉的分割。

### 2.4.2 定性分析

为了更直观地展示 HCGFNet 模型在临床影像观测中的有效性，分别在三类数据集中对本文方法与其余 10 种先进分割模型进行可视化对比，对不同尺寸形状以及边界模糊度的息肉分割实验结果如图 4 ~ 6 所示。其中，图 4 针对性展示了边缘走势不规则以及形状从小至大的息肉图片。图 5 针对性展示了处于复杂肠道背景下不同大小形态的息肉图片。图 6 则针对性展示了边缘平滑所导致息肉与相邻肠道组织色泽高度一致，难以区分的息肉类型图片。分割结果表明，U-Net、DDANet、TransUNet 和 UNeXt 算法模型受固有框架限制均存在不同程度的漏检、错检现象。相比之下，HCGFNet 不仅能准确捕获到微小以及隐匿息肉的位置

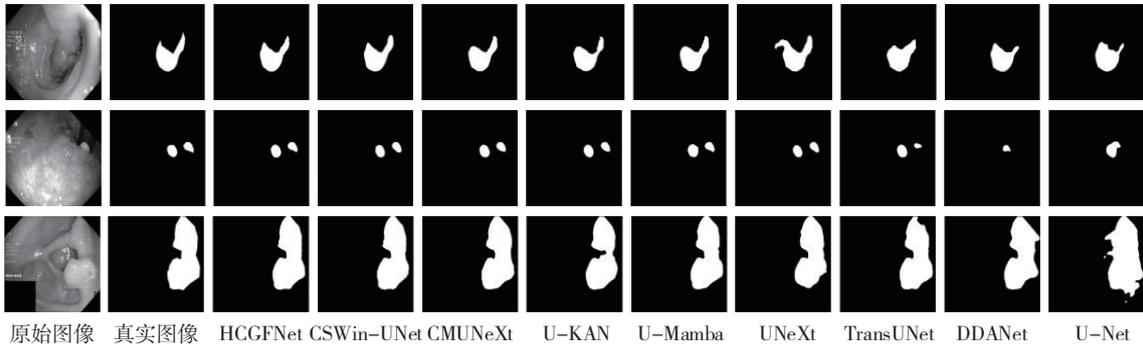


图4 不同算法在 Kvasir-SEG 数据集上可视化

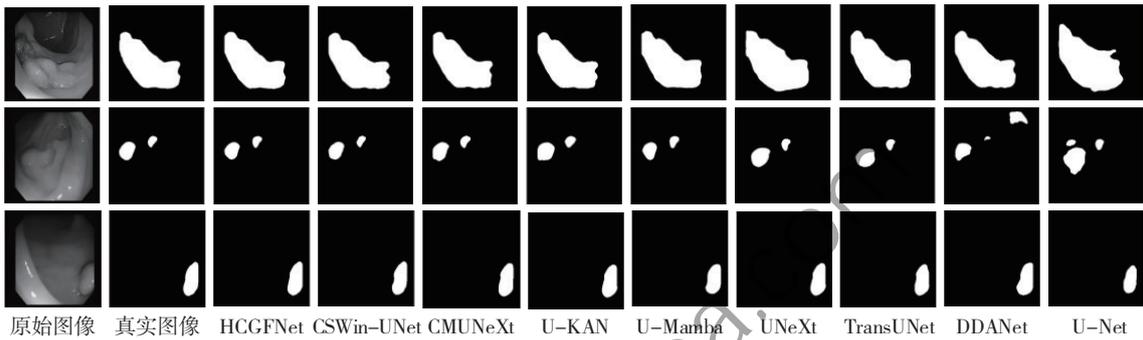


图5 不同算法在 CVC-ClinicDB 数据集上可视化

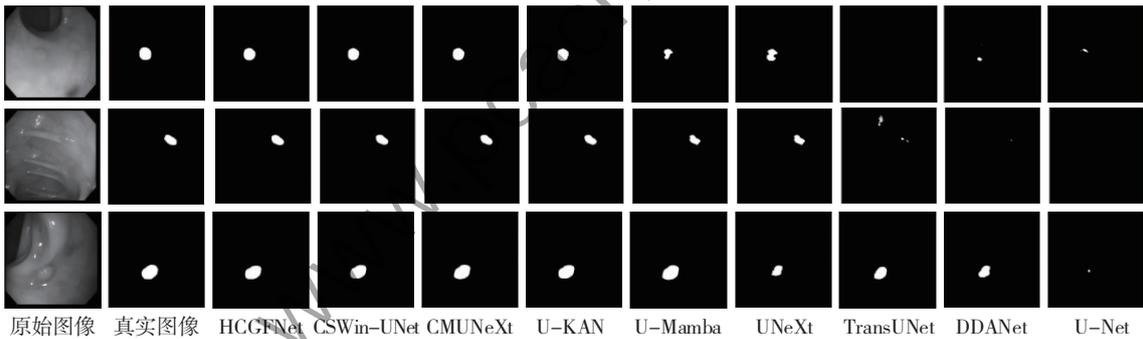


图6 不同算法在 CVC-ColonDB 数据集上可视化

特征信息，还能在复杂肠道环境中精准学习到各类息肉的边缘信息。CSWin-UNet、CMUNeXt、U-KAN 和 U-Mamba 模型在针对边缘走势不规则、棱角分明的 SSL 息肉时，在多处拐角点未能明确捕获、保留边缘细节特征，而本文模型则清晰展现出息肉边缘轮廓，表明模型对边缘细节特征具备更强的学习捕捉能力，进而分割结果也更接近真实分割标签。

综上，模型结合多核异构卷积结构与空间通道注意力机制可更精确地捕捉边缘齿状不规则类型息肉的边缘走向与变化。结合特征引导注意力机制，对微小隐匿型息肉也同样具备优秀的分割能力，能够更精确地捕捉到微小息肉的位置、形状等细节特

征信息。

### 2.4.3 泛化实验结果分析

为进一步验证 HCGFNet 模型的泛化能力，本文选择在 Kvasir-SEG 数据集集中进行训练，在 CVC-ClinicDB、CVC-ColonDB 数据集集中进行验证，并与近年 SOTA 模型进行对比，实验结果如表 4、表 5 所示。在 CVC-ClinicDB 数据集中，HCGFNet 的泛化性能分别在 IoU、Dice、Recall 指标中相较 CSWin-UNet、U-KAN、CMUNeXt 提升 1% ~ 3%，在 Precision 指标中仅比 CSWin-UNet 下降 0.6%。在 CVC-ColonDB 数据集中，HCGFNet 在各项指标上均有 1% ~ 5% 的提升。综上，定量分析结果表明，HCGFNet 在针对不同肠道环境、不同

类型息肉时具备更优越的泛化能力。

表4 CVC-ClinicDB 数据集上泛化实验结果

方法	IoU	Dice	Recall	Precision
U-KAN	0.820 2	0.855 7	0.857 4	0.861 2
CMUNeXt	0.819 6	0.861 0	0.860 3	0.859 4
CSWin-UNet	0.840 7	0.872 1	0.883 5	<b>0.911 3</b>
本文	<b>0.851 5</b>	<b>0.873 9</b>	<b>0.897 3</b>	0.904 4

表5 CVC-ColonDB 数据集上泛化实验结果

方法	IoU	Dice	Recall	Precision
U-KAN	0.702 4	0.735 8	0.779 2	0.808 3
CMUNeXt	0.672 7	0.732 1	0.832 9	0.817 4
CSWin-UNet	0.708 3	0.769 7	0.844 6	0.831 0
本文	<b>0.712 9</b>	<b>0.771 4</b>	<b>0.851 7</b>	<b>0.832 8</b>

#### 2.4.4 消融实验结果分析

HAF 被设计用于提取更加丰富的细节特征，并将这些特征信息融合进分割网络，引导下采样输出更精确的特征掩码。为了验证 HAF 模块的重要性，在实验中去除了 HAF 模块，随机提取 16 个输出通道特征热力图对使用和未使用 HAF 的模型分割过程进行可视化热力特征图比较，如图 7 所示。结合表 6、表 7 数据，模型在去除 HAF 模块后，实线方框中息肉边缘纹理以及细节特征表现能力变弱，息肉成像出现虚影，在

Kvasir-SEG 数据集上的性能下降约 1% ~ 3%，在 CVC-ClinicDB 数据集上的性能下降约 1% ~ 2%。在集成 HAF 模块后，如图 7 虚线方框中的息肉成像所示，可准确地提取到息肉以及息肉边缘纹理特征。

表6 Kvasir-SEG 数据集上消融实验结果

消融实验	HAF	CGFA	IoU	Dice
实验 1		✓	0.797 6	0.861 2
实验 2	✓	✓	0.819 8	0.887 4

表7 CVC-ClinicDB 数据集上消融实验结果

消融实验	HAF	CGFA	IoU	Dice
实验 1		✓	0.863 5	0.902 3
实验 2	✓	✓	0.877 2	0.916 4

CGFA 被设计用于增强解码段重要特征、抑制干扰信息，引导神经网络捕获目标特征。通过在模型中去除 CGFA，以及与 CBAM、SE、CA 等经典通道空间注意力机制相比较，进行性能对比分析。根据表 8、表 9 数据可知，分别用 CBAM、SE、CA 替换 CGFA 时，Kvasir-SEG 数据集中 IoU 系数下降了 1% ~ 3%，Dice 系数下降了 3% ~ 5%，CVC-ClinicDB 数据集中 IoU 系数、Dice 系数均下降 1% ~ 3%，充分证明了 CGFA 模块在引导神经网络抑制冗余信息增强目标特征方面起到了关键作用。

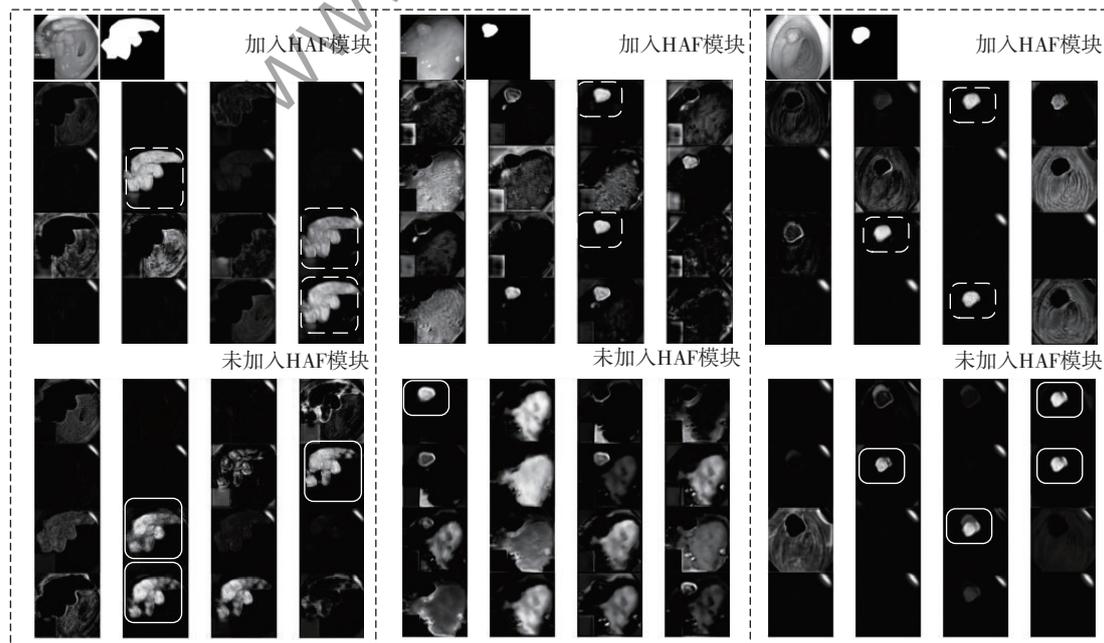


图7 HAF 消融实验可视化热力图

表8 Kvasir-SEG数据集上CGFA消融实验结果

消融实验	HAF	CGFA	CBAM	SE	CA	IoU	Dice
实验1			✓			0.781 3	0.836 8
实验2				✓		0.792 4	0.820 5
实验3					✓	0.780 8	0.848 3
实验4		✓				0.805 9	0.851 6
实验5	✓	✓				0.819 8	0.887 4

表9 CVC-ClinicDB数据集上CGFA消融实验结果

消融实验	HAF	CGFA	CBAM	SE	CA	IoU	Dice
实验1			✓			0.838 6	0.882 4
实验2				✓		0.842 1	0.901 7
实验3					✓	0.839 0	0.891 8
实验4		✓				0.859 9	0.905 1
实验5	✓	✓				0.877 2	0.916 4

此外, HCGFNet 集成异构特征融合模块 (HAF)、内容引导特征融合模块 (CGFA) 后, 在两个数据集上进行广泛测试, 其中各项性能得到显著提升。具体而言, 在 Kvasir-SEG、CVC-ClinicDB 数据集中 IoU 系数、Dice 系数再提升 1% ~ 2%, 通过 HAF、CGFA 的协同工作, HCGFNet 在息肉分割任务中实现了性能跃升。

### 3 结论

本文针对 SSL 息肉特性, 引入了一种新型端到端息肉分割网络 HCGFNet。通过多尺度模块扩大感受视野, 自适应调整通道权重系数, 捕捉复杂环境下息肉边缘特征。特征引导模块应用于网络上采样过程中输出轮廓掩码以及灰度重建图, 有效抑制采样过程中被编入特征层的干扰信息, 还原息肉细节特征。通过这些设计, HCGFNet 在处理 SSL 肠息肉中取得了良好的效果, 提高了模型对微小息肉的感知表达能力。模型虽在分割精度、准确性上取得明显成效, 但就模型参数量、时间性能方面仍有提升空间, 后续可进一步设计合适的轻量化模块, 加入到现有工作成果中, 在保持高精度的同时达到轻量化的标准。

### 参考文献

[1] XIA S R, KRISHNAN S M, TJOA M P, et al. A novel methodology for extracting colon's lumen from colonoscopic images [J]. Journal of Systemics Cybernetics and Informatics, 2003, 1 (2): 7-12.

[2] HWANG S, OH J H, TAVANAPONG W, et al. Polyp detection in colonoscopy video using elliptical shape feature [C]//2007 IEEE International Conference on Image Processing, 2007: 465-468.

[3] TAJBAKHS N, GURUDU S R, LIANG J. Automatic polyp detection using global geometric constraints and local intensity variation patterns [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2014; 17th International Conference. Springer, 2014: 179-187.

[4] MAGHSOUDI O H. Superpixel based segmentation and classification of polyps in wireless capsule endoscopy [C]// 2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), 2017: 1-4.

[5] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 3431-3440.

[6] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [J]. Medical Image Computing and Computer-Assisted Intervention, 2015, 9351: 234-241.

[7] ZHOU Z W, SIDDIQUEE M R, 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, 2018: 3-11.

[8] JHA D, SMEDSRUD P H, RIEGLER M A, et al. ResUNet++: an advanced architecture for medical image segmentation [C]//IEEE International Symposium on Multimedia, 2019: 225-230.

[9] WANG Z N, LIU Z, YU J F, et al. Multi-scale nested UNet with transformer for colorectal polyp segmentation [J]. Journal of Applied Clinical Medical Physics, 2024, 25 (6): 14351.

[10] LI Z H, ZHANG L, YIN S L, et al. MSCFF-Net: multi-scale context feature fusion network for polyp segmentation [J]. Multimedia Systems, 2025, 31 (3). DOI: 10.1007/s00530-025-01776-8.

[11] QIN X, WANG Z L, BAI Y C, et al. FFA-Net: feature fusion attention network for single image dehazing [C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34 (7): 11908-11915.

[12] CHEN Z X, HE Z W, LU Z M. DEA-Net: single image dehazing based on detail-enhanced convolution and content-guided attention [J]. IEEE Transactions on Image Processing, 2024, 33: 1002-1015.

[13] 张文豪, 瞿绍军. 基于双解码器结构的多尺度注意力特征融合网络的视网膜血管分割 [J]. 计算机工程与科学, 2023, 45 (12): 2175-2185.

[14] 郭祥振, 李思潼, 郭祥振, 等. 基于多任务联合注意力的结肠息肉分割网络 [J]. 计算机工程, 2024, 50 (2): 327-336.

[15] 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, 2020: 451-462.

- [16] BERNAL J, SÁNCHEZ F J, FERNANDEZ-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.
- [17] TAJBAKSH N, GURUDU S R, LIANG J. Automated polyp detection in colonoscopy videos using shape and context information [J]. *IEEE Transactions on Medical Imaging*, 2016, 35 (2): 630 – 644.
- [18] TOMAR N K, JHA D, ALI S, et al. DDANet: dual decoder attention network for automatic polyp segmentation [C]//ICPR International Workshops and Challenges, 2021: 307 – 314.
- [19] CHEN J N, LU Y Y, YU Q H, et al. TransUNet: Transformers make strong encoders for medical image segmentation [J]. *Computer Vision and Pattern Recognition*, 2021.
- [20] VALANARASU J M J, PATEL V M. UNeXt: MLP-based rapid medical image segmentation network [C]// International Conference on Medical Image Computing and Computer Assisted Intervention, 2022: 23 – 33.
- [21] AZAD A, JIA Y W, AGHDAM E K, et al. Enhancing medical image segmentation with TransCeption: a multi-scale feature fusion approach [J]. arXiv: 2301.10847, 2023.
- [22] LIU Y T, ZHU H J, LIU M T, et al. Rolling-Unet: revitalizing MLP's ability to efficiently extract long-distance dependencies for medical image segmentation [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38 (4): 3819 – 3827.
- [23] MA J, LI F F, WANG B. U-Mamba: enhancing long-range dependency for biomedical image segmentation [J]. arXiv preprint arXiv: 2401.04722, 2024.
- [24] LI C, LIU X, LI W, et al. U-KAN makes strong backbone for medical image segmentation and generation [J]. arXiv preprint arXiv: 2406.02918, 2024.
- [25] TANG F H, GING J R, WANG L T. CMUNeXt: an efficient medical image segmentation network based on large kernel and skip fusion [C]//2024 IEEE International Conference on Smart Mechatronics, 2024: 89 – 94.
- [26] LIU X, GAO P, YU T, et al. CSWin-UNet: Transformer UNet with cross-shaped windows for medical image segmentation [J]. *Information Fusion*, 2025, 113: 102634.

(收稿日期: 2025 – 12 – 18)

#### 作者简介:

王丽黎 (1968 –), 女, 硕士, 教授, 主要研究方向: 图像处理、人工智能。

郑益聪 (1997 –), 通信作者, 男, 硕士研究生, 主要研究方向: 图像处理。E-mail: 317861447@qq.com。

郭东旭 (2001 –), 男, 硕士研究生, 主要研究方向: 图像处理。

# 版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部

www.pcachina.com