CN 51-1346/O4 ISSN 1003-501X (印刷版) ISSN 2094-4019 (网络版)

光电工程
Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

融合Transformer自适应特征选择的结直肠息肉分割

梁礼明，康婷，王成斌，陈康泉，李俞霖

引用本文:

梁礼明，康婷，王成斌，等. 融合Transformer自适应特征选择的结直肠息肉分割[J]. 光电工程，2025， **52**(3): 240279.

Liang L M, Kang T, Wang C B, et al. Colorectal polyp segmentation via Transformer-based adaptive feature selection[J]. *Opto-Electron Eng*, 2025, **52**(3): 240279.

https://doi.org/10.12086/oee.2025.240279

收稿日期: 2024-11-29; 修改日期: 2025-01-24; 录用日期: 2025-02-06

相关论文

轻量型Swin Transformer与多尺度特征融合相结合的人脸表情识别方法

李艳秋，李胜赵，孙光灵，颜普

光电工程 2025, **52**(1): 240234 doi: 10.12086/oee.2025.240234

基于多尺度特征增强的高效Transformer语义分割网络

张艳，马春明，刘树东，孙叶美

光电工程 2024, **51**(12): 240237 doi: 10.12086/oee.2024.240237

结合极化自注意力和Transformer的结直肠息肉分割方法

谢斌，刘阳倩，李俞霖

光电工程 2024, **51**(10): 240179 doi: 10.12086/oee.2024.240179

面向道路场景语义分割的移动窗口变换神经网络设计

杭昊，黄影平，张栩瑞，罗鑫

光电工程 2024, **51**(1): 230304 doi: 10.12086/oee.2024.230304

更多相关论文见光电期刊集群网站

OEE 光电工程 Opto-Electronic Engineering

http://cn.oejournal.org/oee

OE_Journal Website

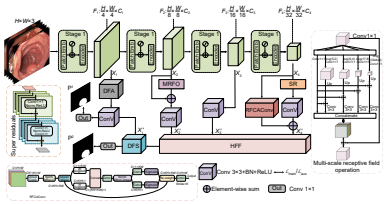
Opto-Electronic Engineering
光 电 工 程

Article

2025 年, 第 52 卷, 第 3 期

DOI: 10.12086/oee.2025.240279 CSTR: 32245.14.oee.2025.240279

融合 Transformer 自适应特征选择的结直肠息肉分割



梁礼明，康　婷*，王成斌，陈康泉，李俞霖

江西理工大学电气工程与自动化学院，江西 赣州 341000

摘要：针对结直肠息肉分割中区域误分割和目标定位精度不足等挑战，本文提出一种融合 Transformer 自适应特征选择的结直肠息肉分割算法。首先通过 Transformer 编码器提取多层次特征表示，涵盖从细粒度到高层语义的多尺度信息；其次设计双重聚焦注意力模块，通过融合多尺度信息、空间注意力和局部细节特征，增强特征表达与辨识能力，显著提升病灶区域定位精度；再次引入分层特征融合模块，采用层次化聚合策略，加强局部与全局特征的融合，强化对复杂区域特征的捕捉，有效减少误分割现象；最后结合动态特征选择模块的自适应筛选与加权机制，优化多分辨率特征表达，去除冗余信息，聚焦关键区域。在 Kvasir、CVC-ClinicDB、CVC-ColonDB 和 ETIS 数据集上进行实验验证，其 Dice 系数分别达到 0.926、0.941、0.814 和 0.797。实验结果表明，本文算法在结直肠息肉分割任务中具有优越性能和应用价值。

关键词：结直肠息肉；Transformer；双重聚焦注意力模块；动态特征选择模块

中图分类号：TP391.4　　**文献标志码：**A

梁礼明，康婷，王成斌，等. 融合 Transformer 自适应特征选择的结直肠息肉分割 [J]. 光电工程，2025，**52**(3): 240279

Liang L M, Kang T, Wang C B, et al. Colorectal polyp segmentation via Transformer-based adaptive feature selection[J]. *Opto-Electron Eng*, 2025, **52**(3): 240279

Colorectal polyp segmentation via Transformer-based adaptive feature selection

Liang Liming, Kang Ting*, Wang Chengbin, Chen Kangquan, Li Yulin

College of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China

Abstract: To address challenges such as regional mis-segmentation and insufficient target localization accuracy in colorectal polyp segmentation, this paper proposes a colorectal polyp segmentation algorithm that integrates adaptive feature selection based on a Transformer. Firstly, the Transformer encoder is employed to extract multi-level feature representations, capturing multi-scale information from fine-grained to high-level semantics. Secondly, a dual-focus attention module is designed to enhance feature representation and recognition capabilities by integrating multi-scale information, spatial attention, and local detail features, significantly improving the localization accuracy of lesion areas. Thirdly, a hierarchical feature fusion module is introduced, which adopts a hierarchical aggregation strategy to strengthen the fusion of local and global features, enhancing the capture of complex regional features and effectively reducing mis-segmentation. Finally, a dynamic feature selection module is

收稿日期：2024−11−29；修回日期：2025−01−24；录用日期：2025−02−06

基金项目：国家自然科学基金资助项目 (51365017，61463018)；江西省自然科学基金资助项目 (20192BAB205084)；江西省教育厅科学技术研究青年项目 (GJJ2200848)

*通信作者：康婷，1833075267@qq.com。

incorporated with adaptive selection and weighting mechanisms to optimize multi-resolution feature representation, eliminate redundant information, and focus on key areas. Experiments conducted on the Kvasir, CVC-ClinicDB, CVC-ColonDB, and ETIS datasets achieved Dice coefficients of 0.926, 0.941, 0.814, and 0.797, respectively. The experimental results demonstrate that the proposed algorithm exhibits superior performance and application value in the task of colorectal polyp segmentation.

Keywords: colorectal polyps; Transformer; dual-focus attention module; dynamic feature selection module

1 引　言

结直肠癌是全球范围内高发的恶性肿瘤之一，严重威胁人类健康，其发病率和死亡率近年来呈现逐渐上升的趋势。这种疾病通常由结直肠息肉恶变引起，而息肉在早期阶段大多表现为良性[1]。如果能在息肉阶段及时发现并进行有效处理，可显著降低癌症发生的风险，提高患者的生存率。因此，结直肠息肉的准确检测和分割对于辅助诊断、手术规划以及治疗监控具有重要意义[2]。然而，由于息肉形态各异、边缘模糊且易与背景组织混淆，如何精确且高效地实现息肉的分割，已成为当前研究中的难点和热点问题。

传统图像分割方法，包括基于区域生长、直方图分析及边缘检测的算法，在一定程度上能够满足简单图像的分割需求[3]。但面对医学图像中复杂的解剖结构和细微特征，这些方法往往因缺乏全局信息建模能力，难以取得令人满意的分割结果。深度学习技术的快速发展为图像分割问题提供强大的工具，尤其是基于卷积神经网络 (CNN) 的模型在医学图像分割领域中取得显著进展。例如，U-Net[4] 通过其编码器-解码器对称结构，能够提取多层次的特征并生成像素级分割结果。然而，U-Net 对复杂场景中多尺度信息的处理能力有限，对目标分割区域连贯性仍存在不足。Diakogiannis 等[5] 基于 ResUNet 设计了一种深度残差结构，通过引入残差模块，增强特征传递能力，并有效缓解深层网络的梯度消失问题，但在小目标分割上的表现仍有不足。Yin 等[6] 提出的 DCRNet 基于上下文关系建模机制，有效融合多层次特征并提升息肉区域的分割精度，但对病灶区域定位仍有待提高。Lou 等[7] 提出的 CaraNet 结合轴向反向注意力机制和上下文信息增强模块，显著增强模型对小目标区域的细节捕捉能力，但复杂网络结构导致其计算效率较低，限制实时应用场景。Huang 等[8] 提出的 MSEG 通过轻量化的 HardNet 编码器和高效解码器，在保证分割精度的同时实现高达 86 f/s 的实时性能，但该模型在应对光照变化时的鲁棒性仍需提升。

近年来，Transformer 在计算机视觉任务中表现出强大的全局建模能力，其在医学图像分割中的应用成为新趋势。例如，Shi 等[9] 提出的 SSFormer-S 结合 Transformer 机制，通过轻量级结构设计，在降低计算复杂度的同时，强化了全局特征捕捉能力，但其在边界细节的处理上仍显不足。Wu 等[10] 设计的 MSRAFormer 结合多尺度反向注意力机制，有效提升了其对复杂形态息肉的分割精度，但模型在处理低分辨率特征时仍存在一定信息丢失。

针对上述挑战，本文提出一种融合 Transformer 自适应特征选择的结直肠息肉分割算法 (TAFS-Net)。具体而言，该算法采用 PVTv2 主干网络提取全局和多尺度特征，利用其线性复杂度注意力机制，通过多层次特征提取捕获从局部细节到全局语义的丰富信息，为复杂场景中的息肉分割奠定坚实基础；在浅层特征处理中，设计双重聚焦注意力模块，通过多尺度信息建模与空间注意力优化，融合浅层局部细节和上下文语义关系，显著提升对边界模糊和复杂背景中息肉区域的定位精度，有效减少误检和漏检现象；在深层特征处理中，引入分层特征融合模块，采用分层聚合策略，充分整合深层全局特征与局部细节信息，优化目标边界的表达能力，使分割结果更加连贯和精准；最后，动态特征选择模块自适应地筛选出关键通道和空间位置，去除冗余特征并通过加权机制聚焦于边缘和重要区域，进一步提升对小尺寸息肉和多目标区域的敏感性与适应性。

2 网络整体架构

2.1 总体结构

为应对结直肠息肉分割中多尺度信息建模不足、细节处理薄弱及噪声干扰等挑战，本文提出一种融合 Transformer 自适应特征选择的结直肠息肉分割算法，其结构如图 1 所示，其中 $\mathcal{L}_{\text{main}}$ 表示主损失 (main

loss)，$\mathcal{L}_{aux}$表示辅助损失 (auxiliary loss)。首先采用金字塔视觉 Transformer (PVTv2)[11] 逐层提取多尺度特征，生成 64×88×88 (X_1)、128×44×44 (X_2)、320×22×22 (X_3) 和 512×11×11 (X_4) 四种分辨率的特征图，涵盖从低级细节到高级语义的多层次信息。其次利用双重聚焦注意力模块对低级特征X_1进行多尺度交互建模与边界强化，增强复杂区域的语义表达能力，同时通过多尺度感受野操作、3×3 卷积、感受野注意力卷积及残差处理对低级高分辨率特征X_2、中层特征X_3和最高级特征X_4进行全局语义信息构建与细节语义增强，优化特征对目标区域辨别能力和边界敏感性，为后续任务提供兼具全局视野与细节精度的高质量特征表征。再次引入分层特征融合模块对X_2'、X_3'和X_4'采用层次化聚合策略，强化特征的重构与边界捕捉能力。最后，动态特征选择模块以自适应机制对不同分辨率的特征进行筛选和加权，进一步优化特征表达并抑制冗余与噪声干扰，从而生成最终的高精度分割结果。

2.2 双重聚焦注意力模块

结直肠息肉由于形态多样、大小不一且易与背景组织相混淆，导致病灶区域的定位精度较低。为应对这一问题，本文受到文献 [12-13] 的启发，设计了一种双重聚焦注意力模块 (dual-focus attention module, DFA)，其结构如图 2 所示，其中 g 表示特征图通道数，σ 表示方差计算。

DFA 模块主要由多尺度聚焦注意力 (multi-scale focused attention branch, MFA) 分支和局部细节增强 (local detail enhancement branch, LDE) 分支组成。其中，MFA 分支通过交互加权捕获长短时记忆的多尺度信息，LDE 分支则利用局部特征与方差加权增强多尺度空间特征。DFA 模块通过融合多尺度信息、空间注意力和局部细节特征，增强特征的表达能力和辨别力，解决病灶区域定位不准问题。若给定输入$X_1 \in \mathbb{R}^{64\times88\times88}$，DFA 模块首先将输入$X_1$特征图并行通过多尺度聚焦注意力分支和局部细节增强分支，然后将两分支输出特征$X_{output1}$和$X_{output2}$与原始输入特征相加，形成最终融合特征X_1'，其表达式为

$$X_1' = X_{output1} + X_{output2} + X_1 . \tag{1}$$

2.2.1 多尺度聚焦注意力分支

首先将输入特征图按通道分组，并分别通过 L1 层和 L2 层操作生成两组多尺度特征图 $K1$ 和 $K2$。随后 $K1$ 和 $K2$ 经过全局池化处理，并应用 Softmax 函数生成相应的权重，量化每组特征对其他特征的相对重要性。最后利用这些计算得到的注意力权重对分组特征进行加权交互，并融合各组特征以生成最终的输出特征图$X_{output1}$。这一过程通过精确聚焦不同尺度和空间位置，优化特征图的表达能力和区分细节的敏感

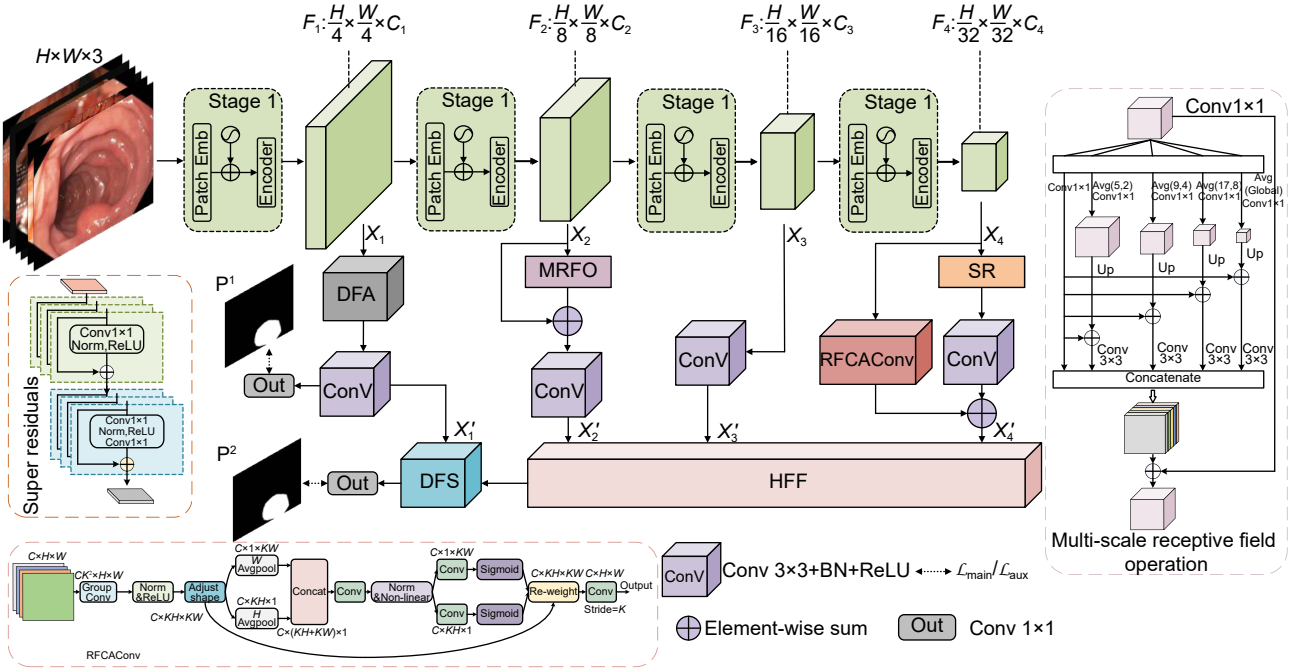


图 1 基于 Transformer 自适应特征选择的结直肠息肉分割算法

Fig. 1 A Transformer-based adaptive feature selection algorithm for colorectal polyp segmentation

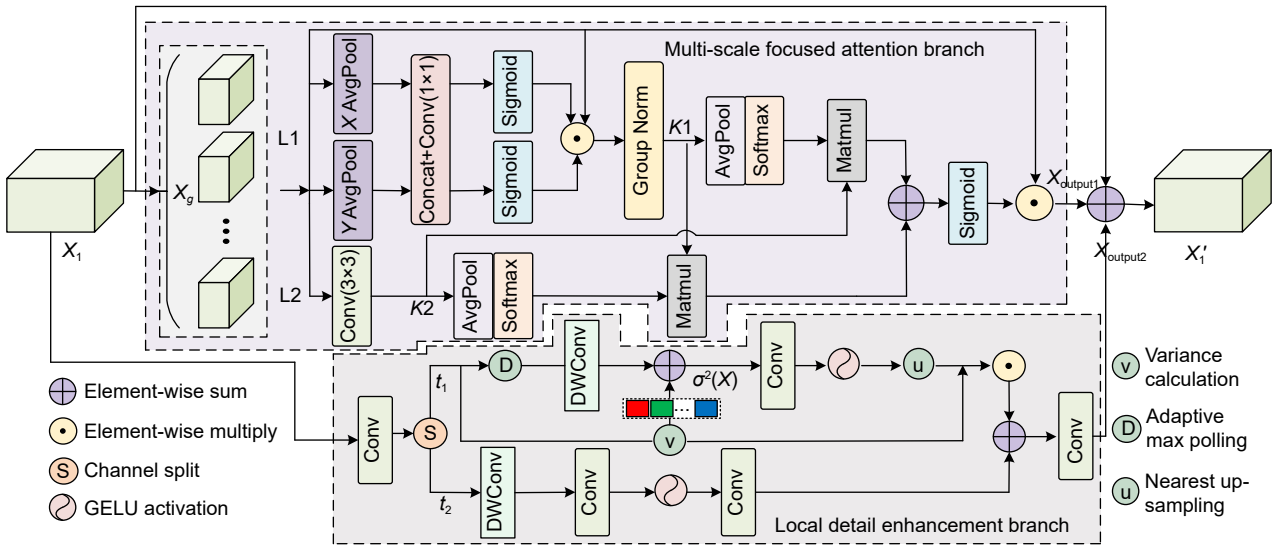


图 2 双重聚焦注意力模块

Fig. 2 Dual-focus attention module

性。其中，L1 层首先对每组特征图进行水平方向和垂直方向的自适应平均池化，提取包含方向性信息的特征图。随后，将池化结果通过 1×1 卷积融合，生成融合特征，并进一步拆分为水平方向和垂直方向的权重映射。接着，这些权重映射经过 Sigmoid 激活后，与原始分组特征图逐元素相乘，生成第一组多尺度特征 $K1$，并通过 Group Normalization 进行标准化处理。L2 层则直接通过 3×3 卷积生成另一组多尺度特征 $K2$，其表达式分别为

$$\begin{cases} W_{11} = \mathrm{Softmax}(\mathrm{GAP}(K1)) \\ W_{21} = \mathrm{Softmax}(\mathrm{GAP}(K2)) \end{cases}, \tag{2}$$

$$X_{\mathrm{output1}} = \mathrm{reshape}(W_{11} \cdot K2 + W_{21} \cdot K1), \tag{3}$$

式中：Softmax(·) 表示归一化；GAP(·) 表示全局平均池化；reshape(·) 表示张量变换操作。

2.2.2 局部细节增强分支

首先通过 1×1 卷积将输入特征图 X_1 分为 t_1 和 t_2 两个分支，其中 t_1 分支用于多尺度特征提取和加权，t_2 分支用于局部双重嵌入。在 t_1 分支中，特征图首先通过自适应池化和深度卷积提取多尺度信息，并在不同尺度上计算方差特征；然后利用可训练参数对这些特征进行加权融合，得到融合后的多尺度特征。t_2 分支则通过具有隐藏 GELU 激活的两个 1×1 卷积进行局部特征增强，并利用深度卷积捕捉细节特征。最后将 t_1 和 t_2 两分支的输出进行融合，并通过 1×1 卷积生成最终输出特征图 X_{output2}，从而在保留多尺度上下文信息的同时，自适应增强局部细节，为后续任务提供更加丰富的空间信息和更强的细节表达能力，其表达式分别为

$$t_{1\mathrm{s}} = \mathrm{Conv}_{3\times3}\left(\mathrm{AdaptiveMaxPool}\left(t_1, \left(\frac{H}{8}, \frac{W}{8}\right)\right)\right), \tag{4}$$

$$t_{1\mathrm{out}} = t_{1\mathrm{v}} \cdot \mathrm{Upsample}(\mathrm{GELU}(\mathrm{Conv}1\times1(t_{1\mathrm{s}} \cdot \alpha + t_{1\mathrm{v}} \cdot \beta)), size = (H, W)), \tag{5}$$

$$t_{2\mathrm{out}} = \mathrm{Conv}\left(\mathrm{GELU}\left(\mathrm{Conv}_{1\times1}(\mathrm{Conv}_{3\times3}(t_2))\right)\right), \tag{6}$$

$$X_{\mathrm{output2}} = \mathrm{Conv}_{1\times1}(t_{1\mathrm{out}} + t_{2\mathrm{out}}), \tag{7}$$

式中：$t_{1\mathrm{v}}$ 是输入特征 $X \in \mathbb{R}^{C\times H\times W}$ 在空间维度上的方差；α 和 β 是可学习参数，分别对 $t_{1\mathrm{s}}$ 和 $t_{1\mathrm{v}}$ 进行加权；Conv(·) 是深度卷积操作；AdaptiveMaxPool(·) 是自适应最大池化；Upsample(·) 是上采样操作；GELU(·) 是非线性激活操作。

2.3 分层特征融合模块

在结直肠息肉分割任务中，有效融合局部特征和全局特征能够增强模型对细小结构和边缘信息的捕捉能力，从而提高在复杂背景下的分割精度。为此，本文引入一种分层特征融合模块 (hierarchical feature fusion, HFF)[14]，该模块通过聚合不同尺度特征图的全局与局部信息，实现多尺度特征的高效融合，从而增强息肉分割效果，其结构如图 3 所示。

HFF 模块将最高层特征 $X_4' \in \mathbb{R}^{32\times44\times44}$ 作为全局特征输入，用于捕捉语义信息；中层特征 $X_3' \in \mathbb{R}^{32\times44\times44}$

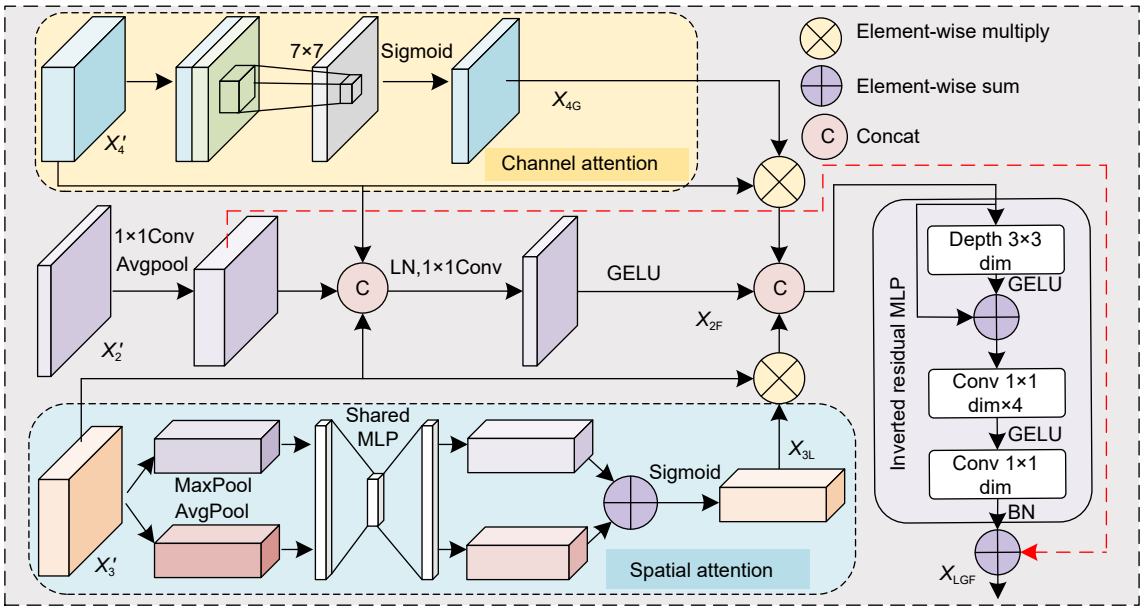


图 3 分层特征融合模块

Fig. 3 Hierarchical feature fusion module

作为局部特征输入，通过空间注意力聚焦于关键区域的细节；低层高分辨率特征$X_2' \in \mathbb{R}^{32\times88\times88}$则作为融合特征，为边缘和细节信息提供支持。具体来说，HFF模块首先对局部特征X_3'应用空间注意力机制，利用最大池化和平均池化生成空间权重图以增强细节特征；对全局特征X_4'使用通道注意力，通过自适应注意力生成的通道权重保留关键语义信息。同时，将融合特征X_2'经过降维和池化后，与局部X_3'和全局特征X_4'拼接，并通过规范化和非线性激活，生成融合后的特征X_{2F}。然后，将空间和通道增强后的特征X_{3L}、X_{4G}与融合特征X_{2F}拼接，并通过 LayerNorm 和残差连接生成输出特征X_{LGF}。通过这种方式，HFF 模块不仅能够保持全局语义信息，还能有效增强边缘和细节信息，从而显著提高分割精度和稳定性。其具体数学描述分别为

$$X_{LGF}^{1} = \text{DropPath}(\text{IRMLP}(\text{LayerNorm}(\text{Concat}(X_{4G}, X_{3L}, X_{2F})))), \tag{8}$$

$$X_{LGF} = \text{AvgPool}(\text{Conv}(X_2')) + X_{LGF}^{1}, \tag{9}$$

式中：DropPath(p) 为随机深度跳跃，p 为丢弃概率；IRMLP(·) 为残差多层感知机操作；LayerNorm(·) 为层归一化；Concat(·) 为特征拼接；AvgPool(·) 为平均池化操作。

2.4 动态特征选择模块

不同分辨率的特征承载不同层次的重要信息，但直接利用这些多尺度特征容易引入冗余信息和噪声，尤其在边界模糊或细节复杂的区域，可能导致误分割现象。因此，如何有效提取多尺度特征中的关键信息，抑制无效或干扰特征，成为提升分割精度的重要挑战。受文献 [15-16] 的启发，设计出一种动态特征选择模块 (dynamic feature selector, DFS)，该模块能从不同分辨率特征中动态筛选出关键的通道和空间位置，去除低质量特征，并利用自适应加权机制增强模型对边缘和重要区域的关注，从而提升模型在复杂场景中的分割表现。DFS 模块由特征选择与特征聚合两阶段组成，其结构如图 4 所示。

2.4.1 特征选择阶段

首先，对高层特征 (语义特征)X_{LGF}和低层特征 (细节特征)X_1'分别进行通道和空间注意力建模。在此过程中，高层特征通过注意力机制强化全局语义表达，确保模型能够捕捉到目标的整体形态和语义信息；低层特征则通过抑制无关区域的噪声，突出边缘和纹理细节，使分割模型在复杂区域的边界表达更加清晰。然后，通过建立高层特征对低层特征的引导关系，对低层特征进行校准，去除其中的噪声信息，并利用低层特征为高层特征补充细节信息。最后，生成经过优化的高层和低层特征H_e和L_e，为后续聚合提供高质量的输入，其具体数学表达式分别为

$$\begin{cases} H_e = \text{FSC}(X_{LGF}) \\ L_e = \text{FSC}(X_1') \end{cases}, \tag{10}$$

$$\text{FSC}(M) = \text{ChannelShuffle}\ (\text{Concat}(\sigma(W_c \cdot \text{GAP}(M_0) + b_c) \cdot M_0, \sigma(W_s \cdot \text{GN}(M_1) + b_s) \cdot M_1)), \tag{11}$$

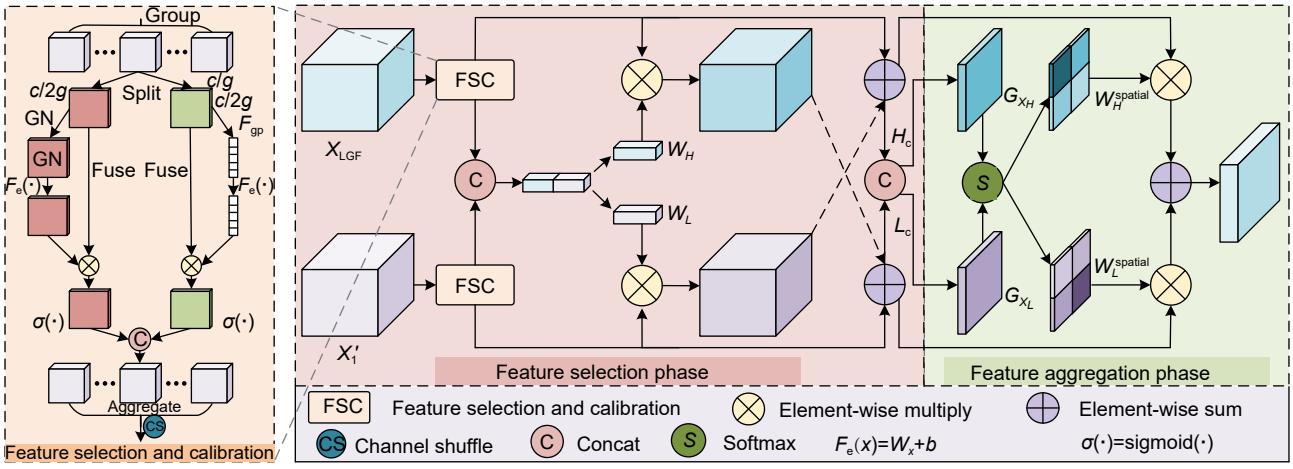


图 4 动态特征选择模块

Fig. 4 Dynamic feature selection module

$$L_c = L_e + \sigma(W_H \cdot \mathrm{ReLU}(W_L \cdot \mathrm{GAP}(\mathrm{Concat}(H_e, L_e)))) \cdot H_e , \tag{12}$$

$$H_c = H_e + \sigma(W_H \cdot \mathrm{ReLU}(W_L \cdot \mathrm{GAP}(\mathrm{Concat}(L_e, H_e)))) \cdot L_e , \tag{13}$$

式中：σ表示 Sigmoid 激活函数；ChannelShuffle(·)表示通道混洗操作；M_0、M_1为通道分组后的两组特征；GAP(·)表示全局平均池化操作；W_c、W_s为通道和空间注意力的可学习参数，用于生成权重；W_L、W_H为两层全连接层的权重，用于生成通道加权向量；b_c、b_s表示通道和空间注意力的偏置。

2.4.2 特征聚合阶段

首先将校准后的高层特征H_c和低层特征L_c拼接在一起，为生成动态权重提供输入。然后，通过 1×1 卷积和 Softmax 机制分别生成高层和低层特征的空间权重，为动态调整高层和低层特征的贡献比例提供依据。最后，根据生成的权重对高层和低层特征进行加权求和，生成最终的聚合特征图，其具体数学表达式分别为

$$W_L^{\mathrm{spatial}} = \mathrm{Softmax}(\mathrm{Conv}_{1\times 1L}(\mathrm{Concat}(H_c, L_c))) , \tag{14}$$

$$W_H^{\mathrm{spatial}} = \mathrm{Softmax}(\mathrm{Conv}_{1\times 1H}(\mathrm{Concat}(H_c, L_c))) , \tag{15}$$

$$F_{\mathrm{output}} = W_L^{\mathrm{spatial}} \cdot L_c + W_H^{\mathrm{spatial}} \cdot H_c , \tag{16}$$

式中：W_L^{spatial}、W_H^{spatial}为低层和高层的动态空间加权权重，由 Softmax 归一化生成。

3 实验结果的讨论与分析

3.1 数据集

为评估网络在结直肠息肉分割任务中的有效性、泛化能力及准确性，本文选用了四个公开数据集开展实验。第一个数据集 CVC-ClinicDB[17] 由医学图像计算与计算机辅助干预协会提供，涵盖多种类型的息肉标注；第二个数据集 Kvasir-SEG[18] 来自挪威奥斯陆大学医院，由专家标注；第三个数据集 CVC-ColonDB[19] 由美国梅奥诊所发布，提供了广泛的标注数据；第四个数据集 ETIS-LaribPolypDB[20] 来源于 MICCAI 息肉检测挑战赛，包含复杂的息肉形态。在实验设计中，本文从 CVC-ClinicDB 数据集中随机抽取 550 张图像，与 Kvasir-SEG 数据集的 900 张图像共同组成训练集，其余图像与 ETIS-LaribPolypDB 和 CVC-ColonDB 的全部图像构成测试集。为统一训练过程，所有训练集图像的分辨率均调整为 352×352。各数据集细节与划分如表 1 所示。

表 1 数据集细节及划分

Table 1 Dataset details and division

Dataset	Image resolution	Train data	Test data	Image type
CVC-ClinicDB	384×288	550	62	Image
Kvasir-SEG	Size variation	900	100	Image and mask
CVC-ColonDB	574×500	0	380	Image
ETIS	1226×996	0	196	Image

3.2 实验环境与参数设计

本实验在 Windows 10 操作系统下进行，使用 PyTorch 3.9 (Facebook Inc., 美国) 和 CUDA 12.1 (Nvidia Inc., 美国) 完成模型构建与训练。硬件平台配置包括 Nvidia GeForce GTX 4060 Ti 显卡和 Intel Core i5-13600KF 处理器。训练过程中，采用结合加权二进制交叉熵和加权交并比 (IoU) 的联合损失函数，优化器选择 Adam，初始学习率设置为 5×10^{-5}，动量系数为 0.9。学习率按 50 轮衰减，衰减率设为 0.1。批次大小为 4，总训练轮数为 100 轮。为提升模型的泛化能力，训练过程中引入了多尺度训练策略，图像缩放比例设置为{0.75,1,1.25}。为保证对比实验的一致性，所有训练与测试均在相同数据集、学习率及优化策略下进行。

3.3 损失函数

在图像分割任务中，传统上使用加权二进制交叉熵损失函数来评估预测结果与真实标签之间的差异。然而，在结直肠息肉分割问题中，目标区域通常较小，这导致传统损失函数在训练过程中容易出现不理想的效果。为克服这一挑战，采用加权 IoU 损失函数能够通过计算预测边界框与真实边界框的交并比来减轻小目标区域分割时的性能问题。因此，本研究结合加权二进制交叉熵损失和加权 IoU 损失的优势，以更全面地评估模型的分割能力。该损失函数的表达式分别为

$$L_{\mathrm{BCE}}=-\frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(1+\lambda\beta_{ij})\sum_{i=1}^{q}\varphi(g_{ij}=l)\log P(p_{ij}=l|\alpha)}{\sum_{i=1}^{H}\sum_{j=1}^{W}\lambda\beta_{ij}}, \tag{17}$$

$$L_{\mathrm{IoU}}=1-\frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(g_{ij}+p_{ij})(1+\lambda\beta_{ij})}{\sum_{i=1}^{H}\sum_{j=1}^{W}(g_{ij}+p_{ij}-g_{ij}\times p_{ij})\times(1+\lambda\beta_{ij})}, \tag{18}$$

$$L=L_{\mathrm{BCE}}+L_{\mathrm{IoU}}, \tag{19}$$

式中：$\varphi(\cdot)$为标记像素类别的指数函数；λ为超参数；$l\in(0,1)$用于区分病变区域和非病变区域；$P(p_{ij}=l|\alpha)$代表预测结果的概率值；β_{ij}为权重值，范围为 (0,1)，该值越大说明像素与周围区域像素值差距大。

3.4 评估指标

为评估结直肠息肉的分割效果，本文采用 6 个常用的指标，包括 Dice 相似性系数、平均交并比 (MIoU)、召回率 (SE)、精确率 (PC)、F2 得分和平均绝对误差 (MAE)。这些指标能够全面衡量模型在分割任务中的性能，其中 Dice 相似性系数、平均交并比 (MIoU) 计算公式分别为

$$Dice=\frac{2|X\cap Y|}{|X|+|Y|}, \tag{20}$$

$$MIoU=\frac{|X\cap Y|}{|X|+|Y|-|X\cap Y|}, \tag{21}$$

$$SE=\frac{TP}{TP+FN}, \tag{22}$$

$$PC=\frac{TP}{TP+FP}, \tag{23}$$

$$F2=\frac{5\times SE\times PC}{4\times PC+SE}, \tag{24}$$

$$MAE=\frac{1}{N}\sum|Y-X|, \tag{25}$$

式中：X为预测输出图像；Y为专家标注的金标签图像；TP为预测结果中正确分割的前景像素数目；FN为预测结果中被错误分类为前景像素数目；FP为预测结果中被错误分类为背景像素数目；N为结直肠息肉图像中的像素点总数。

3.5 算法对比实验

为验证本文算法在结直肠息肉分割中的性能，本文在 Kvasir、CVC-ClinicDB、CVC-ColonDB 和 ETIS 四个公开数据集上，与 U-Net、DCRNet、CaraNet、MSEG、SSFormer-S、MSRAFormer 和 Polyp-PVT[21]7 种分割算法进行对比实验，并将比较结果列于表 2 和表 3，表中加粗部分表示各项指标中的最优表现。

如表 2 和表 3 所示，本文算法在四个公开数据集上的分割性能均展现显著优势，多个指标表现出卓越的全局与局部特征建模能力。由表 2 可知，在 Kvasir 数据集中，本文算法的 Dice (0.926) 和 MIoU (0.878) 均为最优，其中 Dice 反映预测区域与真实标签区域的重叠程度，数值越高表示分割结果与真实区域越吻合。F2(0.919) 达到次优水平，优于大多数其他方法，同时在 SE (0.917) 和 PC (0.955) 上表现卓越，分别展现对目标区域的高检出率和精准度，体现对复杂背景下目标区域的强大捕获能力。在 CVC-ClinicDB 数据集中，本文算法的 Dice (0.941)、MIoU (0.896)、SE (0.957) 和 F2 (0.949) 均为最优，特别是 F2 作为查准率和查全率的综合指标，表明本文算法在分割性能的平衡性上优势显著；此外，PC (0.934) 和 MAE (0.006) 表现突出，说明算法能够在高精度分割的同时有效控制误检和漏检，实现更鲁棒的分割结果。

表 2 Kvasir 和 CVC-ClinicDB 数据集上不同网络分割结果

Table 2 Segmentation results of different networks on Kvasir and CVC-ClinicDB datasets

Dataset	Method	Dice	MIoU	SE	PC	F2	MAE
Kvasir	U-Net[4]	0.818	0.746	0.856	0.857	0.827	0.055
	DCRNet[6]	0.888	0.825	0.902	0.904	0.891	0.035
	CaraNet[7]	0.922	0.872	0.915	0.941	**0.921**	0.019
	MSEG[8]	0.899	0.842	0.900	0.923	0.896	0.028
	SSFormer-S[9]	0.925	0.877	0.914	0.944	0.917	**0.018**
	MSRAFormer[10]	0.923	0.873	0.915	0.952	0.917	0.024
	Polyp-PVT[21]	0.917	0.864	0.913	0.947	0.914	0.023
	Ours	**0.926**	**0.879**	**0.917**	**0.955**	0.919	0.023
CVC-ClinicDB	U-Net[4]	0.823	0.755	0.834	0.839	0.827	0.019
	DCRNet[6]	0.899	0.846	0.913	0.893	0.906	0.010
	CaraNet[7]	0.934	0.890	0.944	0.940	0.939	0.006
	MSEG[8]	0.912	0.866	0.924	0.935	0.918	0.007
	SSFormer-S[9]	0.918	0.875	0.905	**0.939**	0.910	0.007
	MSRAFormer[10]	0.924	0.874	0.945	0.920	0.932	0.008
	Polyp-PVT[21]	0.937	0.889	0.949	0.936	0.945	0.006
	Ours	**0.941**	**0.896**	**0.957**	0.934	**0.949**	**0.006**

表 3 CVC-ConlonDB 和 ETIS 数据集上不同网络分割结果

Table 3 Segmentation results of different networks on CVC-ColonDB and ETIS datasets

Dataset	Method	Dice	MIoU	SE	PC	F2	MAE
CVC-ConlonDB	U-Net[4]	0.512	0.444	0.523	0.621	0.510	0.061
	DCRNet[6]	0.707	0.632	0.776	0.719	0.723	0.052
	CaraNet[7]	0.748	0.683	0.753	0.893	0.746	0.035
	MSEG[8]	0.738	0.669	0.752	0.806	0.739	0.038
	SSFormer-S[9]	0.774	0.698	0.777	0.837	0.766	0.036
	MSRAFormer[10]	0.782	0.707	0.803	**0.874**	0.181	0.028
	Polyp-PVT[21]	0.808	0.727	0.821	0.849	0.809	0.031
	Ours	**0.814**	**0.732**	**0.849**	0.824	**0.825**	**0.028**
ETIS	U-Net[4]	0.398	0.335	0.482	0.439	0.429	0.036
	DCRNet[6]	0.550	0.486	0.746	0.504	0.600	0.095
	CaraNet[7]	0.728	0.661	0.775	0.814	0.750	0.017
	MSEG[8]	0.703	0.632	0.739	0.710	0.720	0.015
	SSFormer-S[9]	0.769	0.698	0.856	0.743	0.800	0.016
	MSRAFormer[10]	0.750	0.679	0.811	0.745	0.777	0.013
	Polyp-PVT[21]	0.787	0.706	0.867	**0.774**	0.820	0.013
	Ours	**0.797**	**0.716**	**0.889**	0.761	**0.834**	**0.018**

由表 3 可知，在 CVC-ColonDB 数据集中，本文算法的 Dice (0.814)、MIoU (0.732) 和 F2 (0.825) 均为最优，其中 SE 达到 0.849，为所有方法中最高，展现对边界模糊区域和小目标的强检测能力。同时，MAE 仅为 0.028，与 MSRAFormer 相当，表明分割误差极小。在 ETIS 数据集中，本文算法在 Dice (0.797)、MIoU (0.716)、SE (0.889) 和 F2 (0.834) 四项指标上均取得最优表现，其中高 SE 值证明算法对复杂背景下息肉区域的显著捕获能力，F2 值则进一步展现对查准率与查全率的良好平衡性。同时，PC

(0.761) 达到次优水平，仅略低于 Polyp-PVT 的 0.774，而 MAE (0.018) 虽略高于最优，但仍处于较低水平。综合来看，本文算法在四个数据集上的综合性能优势显著。与传统的 U-Net 和 DCRNet 相比，本文方法在 Dice 和 MIoU 等关键指标上实现了 30%~50% 的提升；与基于 CNN 的 CaraNet 和 MSEG 相比，本文方法在全局信息建模与边界捕获能力上表现更为优越；而在与其他基于 Transformer 的 Polyp-PVT、SSFormer-S 和 MSRAFormer 的对比中，算法在 Dice、SE 和 F2 等核心指标上同样表现突出，特别是在 CVC-ColonDB 和 ETIS 数据集中展现更强的目标区域检测能力和边界处理能力。综合实验结果表明，本文提出的分割算法能够在复杂背景下更准确地分割出息肉区域，显著提升分割精度与泛化能力。

表 4 展示了本文算法与其他 6 种算法在处理 CVC-ClinicDB 数据集时的参数性能对比。通过参数量、计算复杂度 (GFLOPs) 和训练时间的综合分析可以看出，本文算法表现出卓越的综合优势。在参数量方面，本文方法仅为 26.05 M，相较于 U-Net、DCRNet 和 MSRAFormer 显著减少，充分体现了模型设计的轻量化特点。在计算复杂度上，本文方法的 11.00 GFLOPs 虽略高于 Polyp-PVT 和 SSFormer-S，但较 CaraNet 和 MSRAFormer 显著降低，展现了性能与复杂度之间的高度平衡。训练时间方面，本文算法耗时仅为 183 s，较 U-Net、DCRNet 和 CaraNet 显著缩短，即使与 Polyp-PVT 和 SSFormer-S 相比，也展现出明显的效率优势。综合分析，本文方法能在性能、复杂度与效率之间实现良好的权衡，为结直肠息肉分割任务提供更高效、更实用的解决方案。

表 4　不同网络性能对比 (CVC-ClinicDB)

Table 4　Performance comparison of different networks (CVC-ClinicDB)

Method	Parameters/M	GFLOPs	Train/(round·s^{-1})
U-Net	34.53	65.52	309
DCRNet	28.70	53.00	285
CaraNet	44.54	11.45	256
SSFormer-S	29.31	10.11	220
MSRAFormer	68.96	21.29	199
Polyp-PVT	**25.12**	**5.30**	233
Ours	26.05	11.00	183

图 5 直观地展示本文算法在四个数据集上训练过程中 Dice 系数的变化趋势，清晰体现算法的分割性能。从整体曲线来看，Dice 系数的提升过程平滑且快速，早期训练阶段 (前 20 个 epoch) 即可达到较高水平，展现了良好的收敛性与高效性。其中，Kvasi 和 CVC-ClinicDB 两个数据集表现尤为优异，其 Dice 系数在约第 58 个 epoch 时趋于稳定，分别达到 0.926 和 0.941 的最优水平，表明算法在这些数据集上的分割效果接近完美。相比之下，CVC-ColonDB 和 ETIS 数据集的 Dice 系数起点较低，但经过训练后，CVC-ColonDB 达到了 0.81 左右的高水平，而 ETIS 稳定在 0.76 附近，尽管略低于其他数据集，仍充分展现了算法在复杂场景下的鲁棒性和适应性。总体来看，图中曲线无明显过拟合现象，模型在各数据集上均表现出优越的分割性能，进一步验证了算法的稳定性与广泛适用性。

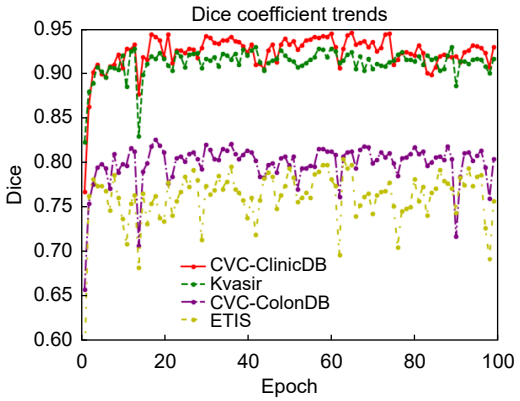


图 5　Dice 系数变化趋势图

Fig. 5　Trend chart of Dice coefficient changes

图 6 和图 7 为上述 8 种算法在 Kvasir、CVC-ClinicDB、CVC-ColonDB 和 ETIS 四个公开数据集上的可视化分割结果，自上而下依次为原始图像 (image)、标注掩码 (mask)、以及 U-Net、DCRNet、Caranet、MSEG、SSFormer-s、MASRFormer、PolypPVT 和本文算法的分割结果。为了更清晰地比较，原始图像中病变区域用蓝色方框标注，同时分割结果采用黄色、红色和绿色三种颜色区分：黄色表示正确分割的病变区域 (真阳性)，红色为错误分割的病变区域 (假阳性)，绿色为未分割出的病变区域 (假阴性)。如图 6 第 2 列和图 7 第 8 列所示，U-Net、DCRNet 和 CaraNet 在处理小尺寸息肉时难以准确捕捉病灶区域的细节特征，常出现漏检现象，尤其在复杂背景或多息肉场景下更为明显。如图 6 第 4 列和图 7 第 3 列所示，MSEG 尽管通过多尺度特征聚合提升了对小息肉的分割能力，但在处理相邻多个小息肉区域时容易导致边界模糊，区域连贯性欠缺；SSFormer-s 虽具有全

局信息建模能力，但在分割边缘不够平滑，容易在小息肉附近引入伪影，干扰实际分割效果。如图 6 第 9 列和图 7 第 9 列所示，MASRFormer 和 PolypPVT 虽然在大病灶分割方面表现较好，但对于小尺寸息肉的敏感性不足，在单幅图像中含有多个小息肉的场景下，常出现漏检或对小病灶区域的分割不完整。与上述算法相比，本文算法通过多项创新设计显著优化了对小尺寸息肉及多小息肉区域的分割能力。首先，基于金字塔视觉 Transformer (PVTv2) 的分层建模能力，本文算法能够有效提取小病灶区域的细粒度特征，为后续处理提供丰富的多尺度特征表征。其次，双重聚焦注意力模块对低级特征进行了细粒度语义建模和边界信息强化，使得小息肉区域的分割更加精准，显著减少漏检现象。然后，分层特征融合模块通过分层聚合策略，增强了对小尺寸病灶间边界的刻画能力，有效提升了在单幅图像中包含多个小息肉场景下的区域连贯性。最后，动态特征选择模块通过自适应加权机制，进一步优化多分辨率特征筛选，提升了对复杂小息肉区域的敏感性，抑制噪声和伪影的同时增强了对小尺寸病灶的鲁棒性。综合分割结果显示，本文算法在小尺寸息肉及多小息肉场景的分割中，较其他算法具备更高的精准性和鲁棒性。无论是分割评价指标还是可视化分割结果，本文算法均优于其他七种对比算法，特别是在小病灶区域的捕捉与细节刻画上，充分展现了其在复杂医学场景下的应用潜力。

目前在临床应用中，白光成像和窄带成像 (NBI) 技术通常是并行使用的，分别提供病灶区域的全局视图和精细结构信息。白光图像虽然具有更广的视野和直观性，但在处理细小病灶、模糊边界和复杂背景时容易出现误检或漏检。相比之下，NBI 图像通过特定波长的光增强微血管和黏膜表面结构的对比度，能够更清晰地显示病灶的边界和纹理细节。然而，NBI 图像的特点也带来了挑战：其背景复杂、光照不均现象显著，且病灶区域的纹理信息丰富但易混淆，这对分割算法的边界处理能力、细节捕捉能力及鲁棒性提出了更高要求。本文算法若应用于 NBI 图像，其模块特性可充分应对上述挑战：基于 PVTv2 的多尺度建模能力能够提取 NBI 图像中从局部细节到全局语义的多层次特征；双重聚焦注意力模块强化细粒度语义建模与边界信息表达，提高对微血管和小病灶区域的敏感性；分层特征融合模块通过聚合全局和局部信息，增强对复杂形态病灶区域的适应性；动态特征选择模块利用自适应加权机制筛选关键特征，抑制伪影和冗余信息，同时优化复杂背景下的分割性能。总体而言，本文算法能够在 NBI 图像中展现更高的分割精度和鲁棒性，不仅对细小病灶区域具有更强的捕捉能力，还能提供平滑连贯的边界分割效果，为 NBI 成像在结直肠息肉检测中的应用提供高质量支持，有望进一步提升诊断的准确性与智能化水平。

3.6 模块消融实验

为深入分析本文算法各模块对整体分割性能的贡献，本文在 CVC-ClinicDB 和 ETIS 数据集上进行了模块消融研究。具体设置如下：G1 代表在分层 Transformer 编码器的基础上，引入分层特征融合模块和动态特征选择模块；G2 代表在分层 Transformer

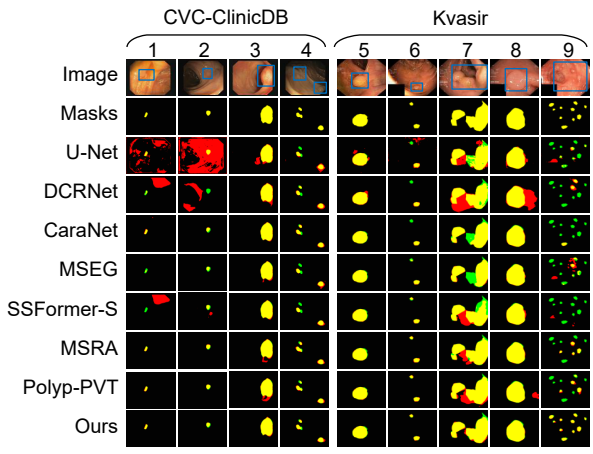


图 6 CVC-ClinicDB 和 Kvasir 数据集上不同网络的可视化分割结果

Fig. 6 Visualization of segmentation results of different networks on CVC-ClinicDB and Kvasir datasets

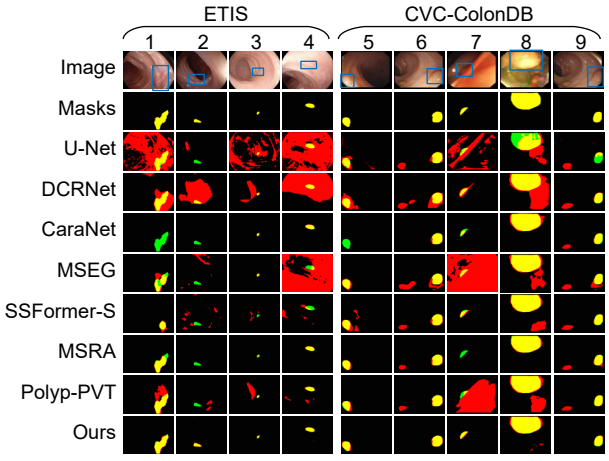


图 7 CVC-ColonDB 和 ETIS 数据集上不同网络的可视化分割结果

Fig. 7 Visualization of sgmentation results of different networks on CVC-ColonDB and ETIS datasets

编码器的基础上，添加双重聚焦注意力模块和动态特征选择模块；G3 代表在分层 Transformer 编码器的基础上，结合双重聚焦注意力模块与分层特征融合模块；G4 为完整模型，即在分层 Transformer 编码器基础上，集成双重聚焦注意力模块、分层特征融合模块和动态特征选择模块，以验证所提方法的整体有效性。消融实验结果见表 5 和表 6，其中，各项指标的最优表现以加粗形式标注。从表中可知，G1 方法在 CVC-ClinicDB 和 ETIS 两个数据集上的 Dice 系数和召回率相比 G4 方法均有所降低，说明双重聚焦注意力模块能够通过捕捉目标区域的细节信息和复杂区域的语义关系，显著增强模型在复杂多变环境中的分割准确性，尤其对边界模糊的息肉区域具有重要作用。G2 方法在 CVC-ClinicDB 和 ETIS 两个数据集上的所有指标均低于 G4 方法，这表明分层特征融合模块能够通过分层聚合策略，整合不同分辨率的特征信息，提升模型对息肉区域形状和大小变化的适应性，并对精确率 (PC) 表现出显著的增强作用。G3 方法在 CVC-ClinicDB 和 ETIS 两个数据集上的平均交并比 (MIoU) 和 F2 系数分别较 G4 方法降低了 0.030、0.028 和 0.018、0.028，这说明动态特征选择模块能够有效筛选不同分辨率的特征信息，优化分割区域的特征表达，提升模型的鲁棒性。特别是在处理不规则形状和复杂背景的息肉时，该模块对 Dice 系数、平均交并比和召回率均有显著的贡献。消融实验验证了各模块对整体算法的贡献。

4 结　论

本文提出一种融合 Transformer 自适应特征选择的结直肠息肉分割算法，旨在应对结直肠息肉分割任务中因复杂背景干扰导致的区域分割误差及目标定位精度不足等问题。算法基于 PVTv2 主干网络，逐层提取息肉图像的多级特征表征，涵盖从局部细节到全局语义的丰富信息。双重聚焦注意力模块通过多尺度交互建模与边界优化，有效提升病灶区域的定位精确性；分层特征融合模块通过分层聚合策略，强化特征重构能力并改善细节捕捉效果；动态特征选择模块利用自适应加权机制，对多分辨率特征进行优化筛选，同时抑制噪声干扰。TAFS-Net 在 Kvasir-SEG、CVC-ClinicDB、ETIS-LaribPolypDB 和 CVC-ColonDB 四个数据集上的召回率 (SE) 分别达到 0.917、0.957、0.849 和 0.889，与现有算法相比，该方法在复杂背景中展现出更优异的分割性能，为结直肠息肉的诊断和治疗提供了有力支持。未来研究可进一步探索轻量化模型设计，以降低计算成本满足实时分割需求，同时通过引入多中心医疗数据集和多模态数据增强模型的泛化能力和鲁棒性；此外，结合半监督学习和自监督学习技术充分利用无标签数据，为小样本医疗分割任务提供新思路；最后，将分割算法与临床辅助诊断系统相结合，实现病灶严重程度预测和息肉分类分析，可进一步提升算法在实际临床应用中的价值，为结直肠疾病的智能诊断和个性化治疗提供全面技术支持[22]。

表 5　各模块在 CVC-ClinicDB 数据集上的消融研究结果

Table 5　Ablation results of each module on the CVC-ClinicDB dataset

Method	DFA	HFF	DFS	Dice	MIoU	SE	PC	F2
G1		√	√	0.929	0.882	0.935	**0.941**	0.931
G2	√		√	0.930	0.884	0.950	0.924	0.938
G3	√	√		0.921	0.866	0.922	0.935	0.921
G4	√	√	√	**0.941**	**0.896**	**0.957**	0.934	**0.949**

表 6　各模块在 ETIS 数据集上的消融研究结果

Table 6　Ablation results of each module on the ETIS dataset

Method	DFA	HFF	DFS	Dice	MIoU	SE	PC	F2
G1		√	√	0.782	0.705	0.856	0.748	0.816
G2	√		√	0.787	0.705	0.875	0.755	0.825
G3	√	√		0.780	0.698	0.839	**0.771**	0.806
G4	√	√	√	**0.797**	**0.716**	**0.889**	0.761	**0.834**

利益冲突：所有作者声明无利益冲突

参考文献

[1] Xie B, Liu Y Q, Li Y L. Colorectal polyp segmentation method combining polarized self-attention and Transformer[J]. *Opto-Electron Eng*, 2024, **51**(10): 240179.
谢斌, 刘阳倩, 李俞霖. 结合极化自注意力和 Transformer 的结直肠息肉分割方法[J]. 光电工程, 2024, **51**(10): 240179.

[2] Lin L, Lv G Z, Wang B, et al. Polyp-LVT: polyp segmentation with lightweight vision transformers[J]. *Knowledge-Based Syst*, 2024, **300**: 112181.

[3] Zhang Y, Ma C M, Liu S D, et al. Multi-scale feature enhanced Transformer network for efficient semantic segmentation[J]. *Opto-Electron Eng*, 2024, **51**(12): 240237.
张艳, 马春明, 刘树东, 等. 基于多尺度特征增强的高效 Transformer 语义分割网络[J]. 光电工程, 2024, **51**(12): 240237.

[4] 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*, 2015: 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

[5] Diakogiannis F I, Waldner F, Caccetta P, et al. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data[J]. *ISPRS J Photogramm Remote Sens*, 2020, **162**: 94−114.

[6] Yin Z J, Liang K M, Ma Z Y, et al. Duplex contextual relation network for polyp segmentation[C]//*Proceedings of 2022 IEEE 19th International Symposium on Biomedical Imaging* (*ISBI*), 2022: 1–5. https://doi.org/10.1109/ISBI52829.2022.9761402.

[7] Lou A G, Guan S Y, Ko H, et al. CaraNet: context axial reverse attention network for segmentation of small medical objects[J]. *Proc SPIE*, 2022, **12032**: 120320D.

[8] Huang C H, Wu H Y, Lin Y L. HarDNet-MSEG: a simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 FPS[Z]. arXiv: 2101.07172, 2021. https://doi.org/10.48550/arXiv.2101.07172.

[9] Shi W T, Xu J, Gao P. SSformer: a lightweight transformer for semantic segmentation[C]//*Proceedings of 2022 IEEE 24th International Workshop on Multimedia Signal Processing* (*MMSP*), 2022: 1–5. https://doi.org/10.1109/MMSP55362.2022.9949177.

[10] Wu C, Long C, Li S J, et al. MSRAformer: multiscale spatial reverse attention network for polyp segmentation[J]. *Comput Biol Med*, 2022, **151**: 106274.

[11] Wang W H, Xie E Z, Li X, et al. PVT v2: improved baselines with pyramid vision transformer[J]. *Comput Visual Media*, 2022, **8**(3): 415−424.

[12] Ouyang D L, He S, Zhang G Z, et al. Efficient multi-scale attention module with cross-spatial learning[C]//*Proceedings of ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), 2023: 1–5. https://doi.org/10.1109/ICASSP49357.2023.10096516.

[13] Zheng M J, Sun L, Dong J X, et al. SMFANet: a lightweight self-modulation feature aggregation network for efficient image super-resolution[C]//*Proceedings of the 18th European Conference on Computer Vision*, 2024: 359–375. https://doi.org/10.1007/978-3-031-72973-7_21.

[14] Huo X Z, Sun G, Tian S W, et al. HiFuse: hierarchical multi-scale feature fusion network for medical image classification[J]. *Biomed Signal Process Control*, 2024, **87**: 105534.

[15] Chen X K, Lin K Y, Wang J B, et al. Bi-directional cross-modality feature propagation with separation-and-aggregation gate for RGB-D semantic segmentation[C]//*Proceedings of the 16th European Conference on Computer Vision*, 2020: 561–577. https://doi.org/10.1007/978-3-030-58621-8_33.

[16] Zhang Q L, Yang Y B. SA-Net: shuffle attention for deep convolutional neural networks[C]//*Proceedings of ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), 2021: 2235–2239. https://doi.org/10.1109/ICASSP39728.2021.9414568.

[17] 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]. *Comput Med Imaging Graphics*, 2015, **43**: 99−111.

[18] 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. https://doi.org/10.1007/978-3-030-37734-2_37.

[19] Tajbakhsh N, Gurudu S R, Liang J M. Automated polyp detection in colonoscopy videos using shape and context information[J]. *IEEE Trans Med Imaging*, 2016, **35**(2): 630−644.

[20] Silva J, Histace A, Romain O, et al. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer[J]. *Int J Comput Assisted Radiol Surg*, 2014, **9**(2): 283−293.

[21] Dong B, Wang W H, Fan D P, et al. Polyp-PVT: polyp segmentation with pyramid vision transformers[Z]. arXiv: 2108.06932, 2024. https://doi.org/10.48550/arXiv.2108.06932.

[22] Li D X, Li D H, Liu Y, et al. Progressive CNN-transformer semantic compensation network for polyp segmentation[J]. *Opt Precis Eng*, 2024, **32**(16): 2523−2536.
李大湘, 李登辉, 刘颖, 等. 渐进式 CNN-Transformer 语义补偿息肉分割网络[J]. 光学 精密工程, 2024, **32**(16): 2523−2536.

作者简介

梁礼明 (1967-)，男，江西吉安人，硕士，教授，硕士生导师，主要研究方向为机器学习、模式识别与图像处理等。

E-mail: 9119890012@jxust.edu.cn

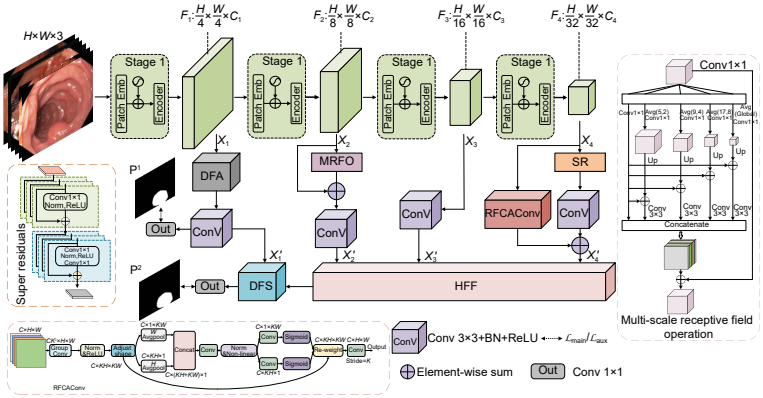
【通信作者】康婷 (2001-)，女，江西赣州人，硕士研究生，主要研究方向为机器学习、模式识别与图像处理等。

E-mail: 1833075267@qq.com

扫描二维码，获取PDF全文

Colorectal polyp segmentation via Transformer-based adaptive feature selection

Liang Liming, Kang Ting*, Wang Chengbin, Chen Kangquan, Li Yulin



A Transformer-based adaptive feature selection algorithm for colorectal polyp segmentation

Overview: Colorectal cancer ranks among the most common and life-threatening diseases worldwide, with colorectal polyps identified as the primary precursors. Accurate detection and segmentation of polyps are essential for preventing cancer progression and improving patient outcomes. However, existing segmentation methods face persistent challenges, including regional mis-segmentation, low localization accuracy, and difficulties in capturing the complex features of polyps. To overcome these limitations, this study presents a novel colorectal polyp segmentation algorithm that integrates Transformer-based adaptive feature selection to improve segmentation accuracy and robustness.

The proposed approach utilizes a Transformer encoder to extract multi-level feature representations, capturing information from fine-grained details to high-level semantics. This enables a comprehensive understanding of the morphology of polyps and their surrounding tissues. To further improve feature representation, a dual-focus attention module is introduced, which integrates multi-scale information, spatial attention, and local detail features. This module enhances lesion localization accuracy and reduces errors arising from the complex structures of polyps.

To address regional mis-segmentation, a hierarchical feature fusion module is developed. By employing a hierarchical aggregation strategy, this module strengthens the integration of local and global features, allowing the model to better capture intricate regional characteristics. Additionally, a dynamic feature selection module is incorporated to optimize multi-resolution feature representations. Through adaptive selection and weighting mechanisms, this module eliminates redundant information and focuses on critical regions, improving segmentation precision.

Extensive evaluations were conducted on four widely used datasets: Kvasir, CVC-ClinicDB, CVC-ColonDB, and ETIS. The algorithm achieved Dice coefficients of 0.926, 0.941, 0.814, and 0.797, respectively, surpassing state-of-the-art segmentation methods. These results highlight the model's robustness, accuracy, and generalization ability across datasets with diverse imaging characteristics and complexities.

In conclusion, this study demonstrates the effectiveness of integrating Transformer-based adaptive feature selection, dual-focus attention, hierarchical feature fusion, and dynamic feature optimization. The proposed algorithm provides a comprehensive and innovative solution to the challenges of colorectal polyp segmentation, offering significant potential for clinical applications in early cancer diagnosis and treatment.

Liang L M, Kang T, Wang C B, et al. Colorectal polyp segmentation via Transformer-based adaptive feature selection[J]. *Opto-Electron Eng*, 2025, **52**(3): 240279; DOI: 10.12086/oee.2025.240279

Foundation item: National Natural Science Foundation of China (51365017, 61463018), the Natural Science Foundation of Jiangxi Province (20192BAB205084), and the Youth Project of Science and Technology Research of the Jiangxi Provincial Department of Education (GJJ2200848)

College of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China
* E-mail: 1833075267@qq.com