

中图法分类号: 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-18

论文引用格式: Xu Yingcheng, Zhang Yunfeng, Liu Peide, Yao Xunxiang, Chen Yujie. Colorectal polyp segmentation integrating frequency-domain enhancement and geometric guidance [J/OL]. Journal of Image and Graphics, XXXX: 1-18. DOI: 10.11834/jig.260341. (徐英城, 张云峰, 刘培德, 姚勋祥, 陈宇杰. 融合频域增强和几何引导的肠息肉医学图像分割[J/OL]. 中国图象图形学报, XXXX: 1-18. DOI: 10.11834/jig.260341.) [DOI:10.11834/jig.260341]

融合频域增强和几何引导的肠息肉医学图像分割

徐英城¹, 张云峰², 刘培德^{1*}, 姚勋祥², 陈宇杰³

1. 山东财经大学管理科学与工程学院, 山东济南 250014; 2. 山东财经大学计算机与人工智能学院, 山东济南 250014; 3. 山东大学数学学院, 山东济南 250100

摘要: 目的 结肠镜图像中的肠息肉区域常受低对比度、镜面反光、黏膜纹理相似和边界形态复杂等因素影响, 现有分割方法在高频细节提取、弱边界恢复和复杂轮廓建模方面仍存在不足。为提高复杂内镜场景下息肉区域的分割精度, 本文提出一种融合频域增强与几何引导的肠息肉医学图像分割网络。方法 该网络在编码阶段设计频域增强 Transformer 块, 在多头注意力和前馈网络中引入快速傅里叶变换, 将频谱信息嵌入全局上下文建模与局部细节提取过程, 以增强对边缘纹理和结构细节的联合表征。在跳跃连接阶段, 构建双频带特征融合模块, 通过径向频域掩膜分离低频轮廓信息和高频边界响应, 减少跨层传递中的边缘细节损失。在解码阶段, 设计双阶段几何引导模块, 并结合曲率感知几何监督损失, 对边界位置、有符号距离场分布和局部曲率变化进行联合优化。结果 在 Kvasir、ClinicDB、ColonDB、EndoScene 和 ETIS 等 5 个公开数据集上进行实验, 并与现有医学图像分割方法进行对比。本文方法在 5 个数据集上的平均交并比 (intersection over union, IoU) 和平均 Dice 相似系数 (Dice similarity coefficient, Dice) 分别为 78.55% 和 85.14%, 较当前主流方法分别提升 0.83 和 1.37 个百分点。结论 本文方法能够在保持区域一致性的同时改善边界恢复质量, 对低对比度和复杂形态息肉具有较好的分割效果和跨数据集泛化能力。

关键词: 结肠镜图像; 息肉分割; 频域增强; 几何引导; 曲率感知

Colorectal polyp segmentation integrating frequency-domain enhancement and geometric guidance

Xu Yingcheng¹, Zhang Yunfeng², Liu Peide^{1*}, Yao Xunxiang², Chen Yujie³

1. School of Management Science and Engineering of Shandong University of Finance and Economics, Jinan, 250014, China; 2. School of Computing and Artificial Intelligence of Shandong University of Finance and Economics, Jinan, 250014, China; 3. School of Mathematics of Shandong University, Jinan, 250100, China

Abstract: Objective Colorectal polyp segmentation is an important task in colonoscopic image analysis because accurate delineation of polyp regions can provide useful support for lesion localization and subsequent computer-aided diagnosis. However, colonoscopic images often contain low-contrast lesions, specular reflections, mucosal texture interference, and

收稿日期: 2026-06-12; 修回日期: 2026-10-08

* 通信作者: 刘培德 liupd1966@126.com

基金项目: 山东省自然科学基金项目 (项目编号: ZR2023QF161); 山东省高等学校青年创新团队 (项目编号: 2022KJ185); 山东省创新联合基金重点项目 (项目编号: ZR2024LZH00); 山东省自然科学基金项目 (项目编号: ZR2025QC717)

Supported by: Project supported by the Natural Science Foundation of Shandong Province, China (Grant No. ZR2023QF161); The Shandong Provincial Youth Innovation Team Development Plan for Higher Education Institutions (Grant No. 2022KJ185); Shandong Province Innovation and Joint Fund Key Project (Grant No. ZR2024LZH00); Project supported by the Natural Science Foundation of Shandong Province, China (Grant No. ZR2025QC717)

irregular boundary shapes. These factors make the separation between polyp regions and surrounding tissues difficult, especially for small, flat, or weak-boundary polyps. Existing medical image segmentation methods have achieved considerable progress with convolutional neural networks, Transformer-based architectures, and boundary-enhancement strategies. Nevertheless, many of them still focus mainly on spatial-domain feature learning. During hierarchical encoding, downsampling, and skip-connection fusion, high-frequency structural cues related to polyp edges and fine textures may be weakened. In addition, common boundary-enhancement methods usually strengthen boundary responses or introduce auxiliary edge supervision, but they are less effective in modeling geometric continuity. To address these issues, this study proposes a colorectal polyp medical image segmentation network that integrates frequency-domain enhancement and geometry-guided boundary modeling. **Method** The proposed network follows an encoder-decoder architecture and introduces frequency-domain and geometric priors into different stages of feature representation and boundary reconstruction. In the encoding stage, a Spectral-enhanced Transformer Block is designed to improve the joint modeling of global structures and local details. Specifically, fast Fourier transform is embedded into the multi-head attention and feed-forward processes, allowing the network to enhance frequency-sensitive responses while preserving the semantic representation learned in the spatial domain. This design helps the encoder capture structural information that is closely related to weak polyp boundaries, subtle texture changes, and local contour variations. In the skip-connection stage, a Dual-band Feature Fusion Module is constructed to reduce the loss of boundary details during cross-layer feature transmission. The module decomposes encoder features into low-frequency and high-frequency components through a radial frequency-domain mask. The low-frequency branch is used to preserve region-level contour consistency, whereas the high-frequency branch emphasizes boundary transitions and fine-grained texture responses. These complementary components are then aligned and fused with decoder features, which enables the decoder to recover both global polyp shape and local boundary details more effectively. In the decoding stage, a Dual-stage Geometry-guided Module is introduced to incorporate signed distance and curvature information into boundary reconstruction. The module first estimates a coarse signed distance field and derives curvature-related cues from the distance representation. Narrow-band weighting and curvature-variation weighting are then used as geometric priors to guide the recovery of ambiguous boundary regions and locally curved contours. Furthermore, a Curvature-aware Geometric Supervision Loss is designed to jointly constrain the final segmentation output, boundary localization, signed distance field regression, and local curvature consistency. Through this joint optimization, the proposed network learns not only region-level overlap but also geometric consistency around complex polyp boundaries. **Result** Experiments were conducted on five public colorectal polyp segmentation datasets, including Kvasir, ClinicDB, ColonDB, EndoScene, and ETIS. The proposed method was compared with 19 representative medical image segmentation methods, including general segmentation networks and polyp-oriented segmentation models. IoU and Dice were used as quantitative evaluation metrics because they can reflect region-level consistency and segmentation overlap. On the Kvasir dataset, the proposed method achieved an IoU of 85.78% and a Dice score of 91.04%. On the ClinicDB dataset, the IoU and Dice score reached 86.18% and 91.21%, respectively. On the ColonDB dataset, the proposed method achieved 70.62% IoU and 78.03% Dice, indicating its effectiveness under more challenging cross-dataset conditions. On the EndoScene dataset, the IoU and Dice score reached 83.37% and 89.65%, respectively. On the ETIS dataset, which contains many low-contrast and ambiguous polyp cases, the proposed method achieved 66.79% IoU and 75.76% Dice. The average IoU and average Dice over the five datasets were 78.55% and 85.14%, respectively, which were 0.83 and 1.37 percentage points higher than existing state-of-the-art methods. Qualitative visualization results also showed that our model produced fewer missing regions and fewer boundary errors in low-contrast, small-target, and complex-contour cases. **Conclusion** The proposed frequency-domain enhanced and geometry-guided segmentation network improves colorectal polyp segmentation by combining spectral detail modeling, dual-band feature transmission, signed distance representation, and curvature-aware supervision. The experimental results demonstrate that the method achieves high segmentation accuracy on five public datasets and shows promising cross-dataset generalization. The method is particularly effective for weak-boundary, low-contrast, and complex-shaped polyps, where both high-frequency detail preservation and geometric boundary modeling are important for accurate segmentation.

Key words: colonoscopic image; polyp segmentation; frequency-domain enhancement; geometric guidance; curvature

awareness

论文引用格式: Xu Yingcheng, Zhang Yunfeng, Liu Peide, Yao Xunxiang, Chen Yujie. Colorectal polyp segmentation integrating frequency-domain enhancement and geometric guidance— SCID [J/OL]. Journal of Image and Graphics. DOI: 10.11834/jig.260341. (徐英城, 张云峰, 刘培德, 姚勋祥, 陈宇杰. 融合频域增强和几何引导的肠息肉医学图像分割— SCID [J/OL]. Journal of Image and Graphics. DOI:10.11834/jig.260341.)

0 引言

结直肠癌是全球范围内发病率和死亡率均较高的恶性肿瘤之一,早筛早诊早治对降低疾病负担具有重要意义(Morgan 等,2023)。结肠镜检查及癌前腺瘤性息肉的及时识别与切除,仍是结直肠癌预防和临床干预的重要手段(Zauber 等,2012)。然而,在实际内镜检查过程中,肠息肉识别与勾画容易受到肠道准备质量、视觉疲劳以及病灶形态复杂等多种因素影响,尤其是小息肉、扁平息肉及低对比度病灶更易与周围黏膜背景混淆,进而造成漏检、误检或轮廓定位偏差(Qayoom 等,2025)。因此,面向结肠镜图像的自动肠息肉分割不仅有助于增强病灶可见性、提高病灶区域定位与边界描绘的准确性,而且能够为后续检测、定量分析与辅助诊疗提供更加可靠的基础支撑,具有重要的理论意义和应用价值(Chen 等,2024)。

近年来,以U形网络(U-shaped network, U-Net)及其变体为代表的编码器-解码器框架,通过跳跃连接融合浅层细节与深层语义特征,已成为医学图像分割的经典基线模型(Ronneberger 等,2015; Liu 和 Mo, 2025)。在此基础上,卷积神经网络(convolutional neural network, CNN)对局部纹理、边缘特征和形态细节具有较强的感知能力,因此在边界清晰的病灶区域分割中得到了广泛应用(彭晏飞和孙伟强, 2026; Tong 等, 2024)。随着Transformer在视觉任务中的成功应用,学者们进一步将其长距离依赖建模能力引入医学图像分割,通过自注意力机制增强全局上下文建模能力,并形成CNN与Transformer的混合架构,以兼顾局部细节提取和全局语义建模(刘善

洁和童孟军,2026)。这类方法在整体分割性能上取得了明显进展,但现有模型大多仍围绕空间域特征展开设计,其改进重点通常集中于多尺度特征融合、卷积增强或混合编码结构,对频域结构信息捕捉与高频细节恢复仍显不足。

对于结肠镜图像而言,息肉边缘、纹理突变、细小凸起及局部轮廓变化等信息往往对应于中高频成分,而这些信息恰恰是提升弱边界、小目标和低对比区域分割精度的关键线索(黄冠珩 等,2026)。现有网络在卷积聚合、跨层交互等过程中,容易造成高频细节衰减,尤其在编码器逐级下采样和编码器-解码器间跳跃连接的特征融合过程中,边缘结构与纹理细节容易受到背景噪声、阴影反光及黏液等干扰的影响而进一步衰减(Zhang 等, 2024; Xue 等, 2024)。最近,一些研究开始尝试从频域角度改善医学图像分割性能,如通过小波变换、Gabor滤波和傅里叶变换增强结构表达,以补充空间域难以稳定捕获的细粒度信息(Yan 等,2025; Chen C 等,2023)。然而,现有肠息肉分割中的频域建模方法多侧重于编码阶段的特征重建或频域参与特征融合,对语义提取与细节恢复等关键路径上的协同频域建模仍相对不足,难以同时兼顾深层特征判别与浅层精细结构进行联合建模。

除中高频信息表征不足外,肠息肉分割的另一项核心挑战在于边界定位。结肠镜图像中普遍存在反光、阴影和局部遮挡等现象,息肉轮廓又具有显著的形态异质性和局部弯曲变化,导致病灶边界常呈现模糊、不连续甚至局部塌陷的特征(贾立新 等, 2023; Yue 等, 2024)。为提升边界定位精度与轮廓恢复质量,已有研究从边界增强、边界分支以及边界注意力等角度开展了探索,在一定程度上改善了模糊边界区域的分割效果(Xue 等,2020; Karimi 和 Salcudean, 2020)。然而,这类方法多通过边界显著性增强或附加边界监督来提升边界区域的可分性,主要侧重于边界区域的定位与响应强化,而对局部曲率变化、轮廓连续性等几何先验的结构化建模仍较为有限。对于形态异质性强、局部边界起伏显著的肠息肉而言,仅依赖边界响应强化或附加边界监督,往往难以稳定刻画边界形状的连续变化。因此,有必要进一步将几何先验引入边界建模过程,使模型

在增强边界可分性的同时具备对边界几何结构的建模能力。

为解决肠息肉图像分割中高频细节易衰减、弱边界区域难恢复以及复杂轮廓难以准确刻画的问题,本文提出一种融合频域增强与几何引导的肠息肉分割网络框架。首先,在编码器中设计频域增强Transformer块(spectral-enhanced transformer block, STB),在多头注意力和前馈网络中引入快速傅里叶变换,以增强网络对全局频域结构和局部细节信息的联合表征能力。其次,在跳跃连接层中构建双频带特征融合模块(dual-band feature fusion module, DFFM),利用径向频域掩膜将编码器特征划分为高频分量和低频分量,并通过不同频带信息的融合以减少跨层传递过程中轮廓纹理和边缘细节的丢失。进一步,在解码阶段构建双阶段几何引导模块(dual-stage geometry-guided module, DGM),通过有符号距离场统一表示息肉边界内外空间关系,并从预测距离场中解析曲率变化信息,以增强模型对局部轮廓起伏、弱边界区域和复杂形态变化的建模能力。最后,结合曲率感知几何监督损失(curvature-aware geometric supervision loss, CGSL),从边界位置约束和几何形状约束两个层面对网络进行优化,从而提高肠息肉分割的整体精度与边界恢复质量。

本文的主要贡献总结如下:1)提出一种融合频域增强与几何引导的肠息肉分割网络框架:通过将频域特征与几何先验统一到特征提取和边界恢复过程中,从而增强模型对低对比度、弱边界和复杂形态息肉的分割能力;2)设计了频域增强Transformer块,在多头注意力和前馈网络中引入快速傅里叶变换,使频谱信息参与全局上下文与局部细节的联合建模;3)提出了双频带特征融合模块利用径向频域掩膜将编码器特征分解为低频轮廓分量和高频边界分量,同时保留区域主体结构信息和边界细节信息;4)设计了双阶段几何引导模块,引入有符号距离场、窄带权重和曲率变化信息,使网络在解码阶段能够利用几何先验引导边界恢复,并在训练过程中对边界位置、距离场分布和局部曲率一致性进行联合优化。在5个公开肠息肉分割任务数据集中与现有主流方法对比,验证了本文分割网络框架对提升复杂场景下肠息肉分割性能的有效性。

1 相关工作

1.1 基于空间域特征的医学图像分割方法

医学图像分割长期以空间域特征学习为主,其中U-Net凭借编码器-解码器结构及跳跃连接机制成为该领域最具代表性的基础框架(Ronneberger等, 2015)。在此基础上,现有研究主要围绕特征融合、注意力机制与上下文信息增强等方面进行改进(Gu等, 2019)。UNet++通过嵌套密集跳跃连接层进一步缩小编码器与解码器之间的语义差距以提升跨层特征融合效果(Zhou等, 2018); Attention UNet引入注意力门控机制,使网络聚焦目标区域特征并抑制无关背景干扰(Oktay等, 2018); I2U-Net采用双路径编码结构,强化了低层结构细节与高层语义信息的联合建模能力(Dai等, 2024)。为弥补卷积网络在长程依赖和全局上下文关系建模方面的不足,双编码器结构、Transformer以及状态空间模型也被逐步引入医学图像分割(Cao等, 2022; Ruan等, 2025)。其中,TransUNet将Transformer嵌入U型框架编码器中,利用全局自注意力补充卷积的局部建模特性(Chen等, 2021); TransFuse通过并行CNN与Transformer分支同时建模低层空间细节和全局依赖关系(Zhang等, 2021); 双编码器U-Net(dual-encoder U-Net, DEU-Net)采用双编码器协同提取局部卷积特征和全局上下文信息(Karimi等, 2023)。近期, Mamba类状态空间模型逐渐被引入医学图像分割任务,并在局部结构表达方面表现出较好的潜力(Wu等, 2025), U-Mamba利用状态空间模型在降低复杂度的同时提升长程依赖建模能力(Ma等, 2024)。

在肠息肉分割任务中,现有研究通常从卷积模块、注意力机制和多尺度特征融合等空间域机制提升分割性能。浅层注意力网络(shallow attention network, SANet)通过注意力机制抑制背景噪声并保留小息肉的高分辨率特征(Wei等, 2021); M2UNet结合MetaFormer与多尺度上采样强化多层解码特征之间的聚合能力(Trinh等, 2023)。总体来看,空间域方法虽已取得较好的分割效果,但在面对边界模糊、纹理相近和高频轮廓细节易丢失等问题时,仍容易出现低频语义占优而高频特征响应不足的现象。

1.2 基于频域增强的医学图像分割方法

为弥补空间域建模对全局结构表示和细微纹理

刻画方面的不足,学者们开始将快速傅里叶变换、小波变换等频域方法引入医学图像分割模型,利用不同频率成分补充语义与纹理信息。全局频域UNet(global frequency domain UNet, GFUNet)在U-Net结构引入傅里叶变换,通过双域编码模块联合利用空间域和频域特征(Li 等, 2023);频域滤波UNet(frequency domain filter UNet, FDFUNet)通过多尺度频域滤波器从频域特征中捕获全局信息,从而缓解通道冗余问题并降低全局建模成本(Chen 等, 2024);金字塔傅里叶可变形网络(pyramid Fourier deformable network, PFD-Net)设计快速傅里叶卷积残差模块,进一步丰富编码器输出的局部和全局特征(Yang and Zhang, 2024)。总体而言,频域建模为医学图像分割提供了区别传统空间卷积的特征提取视角,有助于补充卷积网络难以稳定捕获的全局结构信息和细粒度纹理特征。

在肠息肉分割任务中,由于息肉区域具有低对比度、内镜高光和黏膜纹理复杂等特点,仅依赖空间域卷积往往难以有效提取轮廓与纹理线索。AdaptUNet利用小波变换提取传统空间域处理中易被忽略的特征,以增强对病灶区域轮廓纹理的刻画(Rajasekar 等, 2024);DSHNet利用动态频谱解耦器分离低频和高频分量,根据不同频率信息的特性来提取关键特征(Wang H L 等, 2025);小波增强边界感知网络(wavelet-enhanced boundary aware network, WBANet)设计小波引导特征增强模块融合空间域与频域信息,从而改善息肉结构的清晰度(Wang Y P 等, 2025)。上述研究验证了频域特征建模在肠息肉分割任务中的有效性,但现有方法多在单一阶段引入频域方法,并且在筛选中高频信息方面仍有提升空间。而本文方法与已有频域分割方法的区别在于:并非单纯依赖频域特征增强,而是将频域细节和几何边界建模进行结合,以提升弱边界和复杂轮廓息肉的分割稳定性。

1.3 基于几何边界建模的医学图像分割方法

近年来,除强化主干网络之外,几何边界建模也成为提升医学图像分割精度的重要方向。一类方法通过边缘特征提取、增加边界分支等来改善病灶区域的边界定位能力。边界引导网络(boundary guidance network, BGDNet)设计专门的边界提取模块,以加强解码过程中的边界特征恢复(Xu 等, 2024);边界感知上下文神经网络(boundary-aware

context neural network, BA-Net)在编码器子网络的每个阶段中提出一种金字塔式边缘提取模块用于捕捉多粒度边缘信息,从而辅助病灶边界定位(Wang 等, 2022)。另一类方法从几何建模角度出发,将形状先验信息编码为距离图引入分割模型。几何感知互学习(geometry-aware mutual learning, GAML)从病灶内外区域的几何关系出发,借助几何信息引导的图推理模块深入挖掘复杂形态与模糊边界下的结构几何特征(Wang K 等, 2025);此外,还有研究通过多尺度特征和符号距离图从区域语义表示到边界细节恢复的转换,从而改善整体分割效果(Liu 等, 2022)。

在肠息肉分割任务中,息肉在结肠镜图像中通常表现出形状多变、边界与背景黏膜易混淆、局部凸起或凹陷明显等特点,单纯依赖区域特征难以获得精细的轮廓边界定位结果。多尺度差分网络(multi-scale subtraction network, MSNet)通过多尺度差分特征建模强调不同层级之间的结构差异,从而缓解定位偏移和边缘模糊问题(Zhao 等, 2021);PEFNet在上采样层定义位置嵌入特征块,将位置信息逐步转化为高质量分割掩码,并最大限度地保留边界特征(Nguyen-Mau 等, 2023);多尺度边缘引导注意力网络(multi-scale edge-guided attention network, MEGANet)将拉普拉斯算子融入边缘引导注意力机制中,通过强化边缘区域的关键高频响应,提升弱边界场景下的轮廓建模能力(Bui 等, 2024)。总体来看,肠息肉分割中的几何建模方法已逐步由边界增强拓展至形状约束、距离映射和几何关系建模等层面,但对边界局部曲率变化等关键几何属性的刻画仍显不足。

2 方法

2.1 整体框架

本文提出一种融合空间域与频域特征的肠息肉分割网络,引入有符号距离图和曲率变化作为几何先验信息,其总体结构如图1所示,以解决肠息肉内镜图像中边界模糊、局部凸起复杂以及背景纹理相似导致的误分割问题。该网络的整体框架以编码器-解码器为基础,沿跳跃连接层传递语义信息,包含4个核心结构:(1)由卷积层与频域增强Transformer块组成的编码器,用于联合建模局部纹理细节和全局上下文依赖;(2)以双频带特征融合模块为

关键组成部分的跳跃连接层,用于在解码阶段强化区域轮廓信息和边界细节信息;(3)基于有符号距离图的双阶段几何引导模块,用于在解码过程中逐步引入边界和曲率先验;(4)面向分割结果、边界响应、有符号距离场(signed distance field, SDF)和曲率变化的几何监督损失,用于提高模型对复杂边界形态和细粒度轮廓结构的恢复能力。

给定肠息肉内镜图像 $I \in \mathbb{R}^{H \times W \times D}$, I 首先进入编码器中的卷积分支进行空间域特征提取,保留局部纹理、边缘梯度与细粒度空间结构,再经由残差网络提取分层局部特征,对应分辨率为 $[H/2, H/4, H/8]$, 输入到由 12 个 STB 模块组成的频域特征增强分支,并在注意力机制和前馈神经网络(feed-forward net-

work, FFN)中显式增强频域响应,从而形成兼具局部辨识度与全局一致性的混合特征。随后,编码特征经由跳跃连接层中的 DFFM 划分出低频分量与高频分量,在中高频分支中突出边界转折、局部凸起和细粒度纹理细节,在低频分支中补充区域主体轮廓与结构一致性,进一步融合得到深层解码特征 d_1 和 d_2 。最后在解码阶段,网络将利用生成的粗几何先验通过 DGM 获取粗几何信息并在高分辨率中完成精细几何重建,输出分割图、边界图、有符号距离图及其对应的曲率图。为进一步约束网络在边界区域和曲率变化区域的几何一致性,本文构建曲率感知损失函数,对主

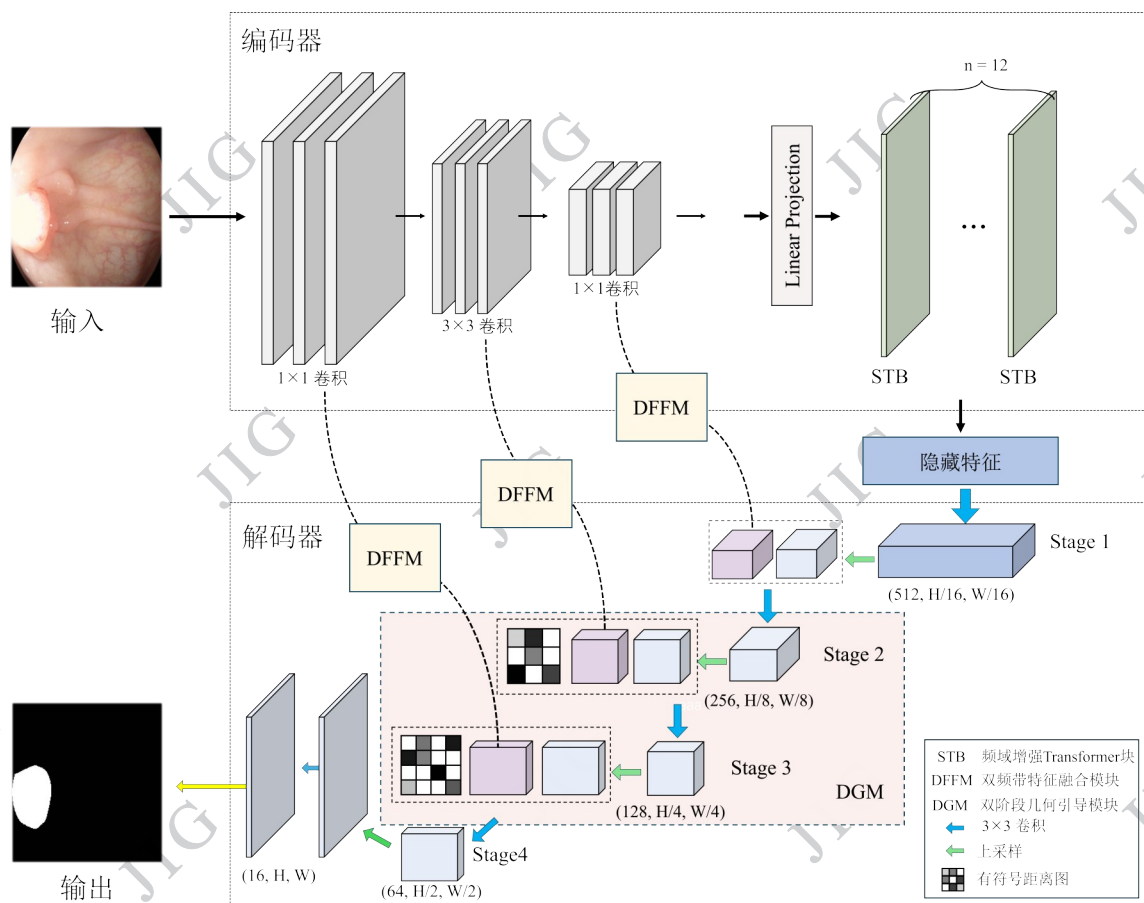


图1 本文方法框架图

Fig. 1 Architecture of the proposed framework

分割、SDF回归、曲率学习和零水平集一致性进行联合优化,在保证区域一致性的同时,更有效地恢复肠息肉复杂边界处的局部形态变化。

2.2 频域增强 Transformer 块

在结肠镜图像中,息肉与邻近黏膜的颜色和纹

理可能相近,弱边界区域的局部差异不足以独立支持准确分割,需要结合目标整体结构与周围组织的上下文信息,而局部边缘过渡和细小轮廓变化又提供了重要的定位线索。针对上述联合表征需求,本文构建频域增强 Transformer 块,如图2所示。该块

在多头注意力和前馈网络中引入快速傅里叶变换 (fast Fourier transform, FFT), 从全局上下文建模过程中引入频谱先验, 并在局部重构过程中融合空间域和频域特征。这种结构使频谱信息以结构化方式嵌入编码过程, 既保留了 Transformer 建立全局依赖的优势, 也避免了频域分支对原始空间域语义的过度覆盖。

在多头注意力子层中, 首先通过线性映射得到查询、键和值矩阵 Q, K, V 。 Q 和 K 主要用于计算不同空间位置之间的相关性, 决定特征聚合的权重分布; V 则承载每个位置的特征内容信息。为了在保持空间注意力权重稳定的前提下, 本文方法仅对 V 施加 FFT, 增强特征中的中高频成分, 进而使得在后续特征聚合中保留更多包括边界突变、纹理变化和细粒度结构等局部结构细节。具体地, 值分支的频域映射可写为

$$V_f = F(V) \quad (1)$$

式中, $F(\cdot)$ 表示 FFT 算子。为自适应筛选不同通道在频域中对应的高频响应, 本文在频域 V 分支中引入轻量化挤压激励 (squeeze - excitation module, SE) 模块进行通道级加权, 通过全局统计信息建模通道间依赖关系, 自适应增强与息肉边界、纹理变化相关的频率通道, 并抑制由反光、噪声造成的无效通道响应, 得到增强后的频域特征。随后, 利用逆快速傅里叶变换 (inverse fast Fourier transform, IFFT) 将其映射回原特征域, 并与空间域分支进行残差融合, 即

$$V'_f = \alpha F^{-1}(\tilde{V}_f) + V \quad (2)$$

式中, $F^{-1}(\cdot)$ 表示 IFFT, α 为频域分支的残差系数, $\tilde{V}_f$ 为经过 SE 模块的空间域值特征。这种在值通路内部引入频谱先验的方式有利于将结构敏感信息直接注入注意力传播过程中, 锐化了微弱的息肉边界, 保留了细微的纹理变化。

在 STB 的前馈神经网络中, 给定输入特征 X , 先通过逐点映射和深度卷积得到 2 个中间分支 X_1 和 X_2 , 再通过门控激活单元对特征进行动态调制, 将学习到的门控权重与特征逐元素相乘, 得到输出特征 X_3 , 即

$$X_3 = GELU(X_1) \odot X_2 \quad (3)$$

式中, $GELU(\cdot)$ 表示高斯误差线性单元 (Gaussian error linear unit, GELU) 激活函数, $\odot$ 表示逐元素乘法。由于息肉边界、细小凸起和局部曲率变化通常

表现为局部邻域内的非均匀响应, 本文在门控前引入深度卷积, 使特征分支能够先捕获邻域内的纹理变化、边缘梯度和局部形态差异, 通过 $GELU(\cdot)$ 激活后的分支响应形成非线性门控权重, 对具有较强判别性的局部结构响应进行保留或放大, 对背景纹理和弱相关响应进行抑制 (黄冠晖 等, 2026)。

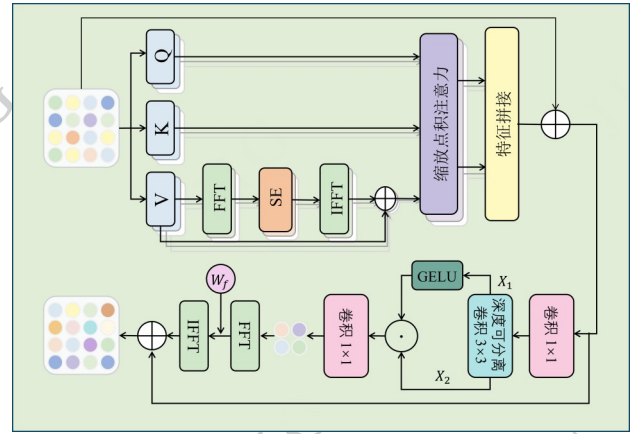


图2 频域增强Transformer块结构

Fig. 2 Structure of the Spectral-enhanced Transformer Block

对于结肠镜图像而言, 息肉边缘、局部凸起、黏膜纹理突变和弱边界过渡通常表现为局部区域内灰度或颜色分布的快速变化, 这类变化在频域中主要反映为中高频成分。在得到空间域特征 X_3 后, 本文对其进行卷积和分块划分, 并对每个块应用 FFT 将其投影到频域中。通过对局部块进行 FFT, 网络可以在局部范围内显式感知不同频率成分, 并借助可学习频率调制矩阵 W_f 对关键频率响应进行加权。随后经 IFFT 与分块重组恢复到空间域后, 得到前馈网络的频域增强特征 X' , 即

$$X' = F^{-1}(F(X_3) \cdot W_f) + X_3 \quad (4)$$

这种设计能够针对局部块的频率成分进行选择性的增强, 增强后的频率响应重新作用于局部纹理和边界结构, 从而在保持空间结构一致性的同时, 强化了与边缘、精细纹理相关的高频信息 (Yang 和 Zhang, 2024)。

2.3 双频带特征融合模块

结肠镜图像中的息肉分割需要分别维持目标区域的完整性和定位小尺度轮廓变化。然而, 反光边缘与背景黏膜纹理也可能产生较强的局部变化, 因此不能将高频响应直接视为息肉边界。为此, 本文在跳跃连接层中构建双频带特征融合模块, 对编码

器特征进行频带分解,并增强分解后的低频和高频分量,利用边缘权重图对高频响应进行选择性地调制,最后与解码器特征对齐融合,从而在跨层传递过程中同时保持轮廓一致性与边界敏感性。

给定第 l 层编码器特征为 $X_l^e \in \mathbb{R}^{H \times W \times D}$, 对应解码器上采样特征为 $X_l^d \in \mathbb{R}^{H \times W \times D}$ 。首先,对 X_l^e 进行 FFT 处理,并利用径向掩膜将其划分为低频分量和高频分量,即

$$F_l^{low} = F^{-1}(M_l \odot F(X_l^e)) \quad (5)$$

$$F_l^{high} = F^{-1}((1 - M_l) \odot F(X_l^e)) \quad (6)$$

式(5)和(6)中, M_l 由径向阈值构造的低频掩膜,由径向阈值来控制低频掩膜在频谱平面中的覆盖范围。低频分量主要表示息肉区域的整体轮廓和结构一致性,高频分量更关注边界突变和局部纹理变化。当阈值较小时,更多频率成分被划入高频分支,高频分支能够获得更多边界和纹理响应,但也更容易混入内镜反光、黏膜纹理和噪声成分;当阈值较大时,低频分支能够保留更多区域主体轮廓和中低频结构信息,但过大的低频覆盖范围可能削弱高频分支对边界突变的表达。随后,本文对低频分量和高频分量进行频域增强。此外,考虑到高频分量中易混入由反光、噪声或背景纹理造成的高频响应,本文在高频分支中引入边缘权重图,对增强后的高频结果进行逐像素乘法调制,即

$$F_l'^{low} = \delta_{low}(F_l^{low}) \quad (7)$$

$$F_l'^{high} = \delta_{high}(F_l^{high}) \odot G_l \quad (8)$$

式(7)和(8)中, G_l 表示与第 l 层编码器特征同尺度的边缘权重图, $\delta_{low}(\cdot)$ 和 $\delta_{high}(\cdot)$ 分别为低频增强算子和高频增强算子。值得强调的是, G_l 是由当前编码器特征 X_l^e 经过卷积和激活函数等操作动态生成的,在训练过程中,生成 G_l 的卷积层和归一化参数通过反向传播进行更新。最后,将增强后的低频特征和高频特征进行重组,并与解码器特征 X_l^d 对齐后沿通道维度拼接,再通过深度可分离卷积完成跨层特征融合,即

$$F_l = D(\text{Cat}(A(F_l'^{low}, F_l'^{high}), X_l^d)) \quad (9)$$

式中 $A(\cdot)$ 表示双线性插值对齐算子, $\text{Cat}(\cdot)$ 表示通道拼接, $D(\cdot)$ 表示深度可分离卷积融合。这种设计使跳跃连接层在向解码器传递信息时,不再只是简单补充空间分辨率,而是从高低频两个分支提升解码阶段对复杂边界和弱对比区域的重建能力。

2.4 双阶段几何引导模块

在低对比度、局部反光或阴影干扰下,息肉轮廓的一部分可能缺少清晰的外观线索,不规则轮廓中的转折与局部弯曲区域也容易在解码过程中出现定位偏差。因此,本文在解码器中构建双阶段几何引导模块,使几何先验信息直接参与高分辨率边界恢复。如图1所示,在 stage2、stage3 阶段上引入不同粒度的几何引导模块对解码特征进行处理并生成先验信息,最终融合语义信息和几何先验实现对息肉边界的细粒度重建。

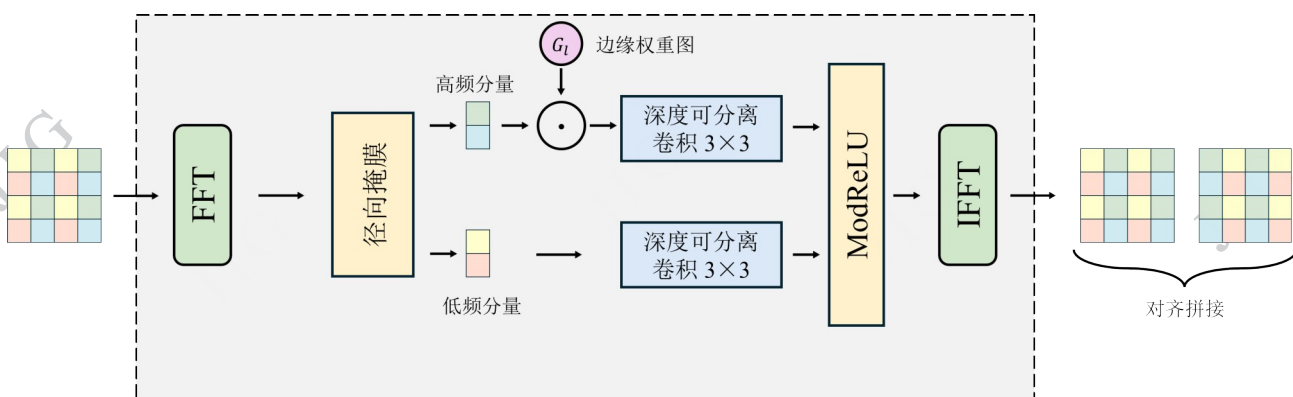


图3 双频带特征融合模块结构

Fig. 3 Structure of Dual-band Feature Fusion Module

给定 stage2 解码阶段的解码特征 $D_2 \in \mathbb{R}^{C \times \frac{H}{8} \times \frac{W}{8}}$, 构建粗粒度几何头,生成边界位置和几何场的初始估计。具体地,粗粒度几何头通过双曲正切映射得

到归一化粗 SDF 图,为降低局部噪声对曲率估计的影响,本文先对 Φ 进行平滑处理,再分别采用水平和垂直 Sobel 差分算子计算梯度,进而由单位法向量散

度得到曲率响应。

$$\Phi = \tanh(f_{sdf}(D)) \quad (10)$$

$$n = \frac{\nabla \Phi}{\sqrt{\Phi_x^2 + \Phi_y^2 + \varepsilon}} \quad (11)$$

$$\kappa = \tanh\left(\frac{S_x(n_x) + S_y(n_y)}{s_\kappa}\right) \quad (12)$$

式中 κ 表示对应曲率, Φ 表示任意尺度上的归一化 SDF 图, $n = (n_x, n_y)$ 表示单位法向量, $\varepsilon = 10^{-6}$ 为数值稳定项, $s_\kappa = 3.0$ 为曲率响应缩放因子。SDF 图表示边界零水平集及其邻域的距离场分布, 而曲率则由距离场计算得到的法向量散度刻画局部弯曲程度。随后, 本文将依据粗 SDF 图与零水平集之间的距离关系构造窄带权重, 通过计算粗曲率幅值构造曲率变化权重, 进一步增强对边界邻域的连续性和对局部弯曲区域的敏感性, 即

$$W_b(x) = \exp\left(-\frac{|\Phi(x)|}{\tau_b}\right) \quad (13)$$

$$W_\kappa(x) = \tanh\left(\frac{|\kappa(x)|}{\tau_\kappa}\right) \quad (14)$$

式中, $W_b(x)$ 表示边界窄带权重, 越接近 SDF 零水平集的像素权重越大; $W_\kappa(x)$ 表示曲率变化权重, 用于突出局部弯曲程度较大的边界区域; τ_b 和 τ_κ 分别表示窄带和曲率归一化参数。在 DGM 中, 给定解码特征 D 和几何张量 C , 将窄带权重和曲率变化权重作为几何先验信息引导后续解码特征的恢复过程, 即

$$C = \text{Cat}(W_b, W_\kappa) \quad (15)$$

$$D' = D \odot (1 + \sigma(C)) + C \quad (16)$$

式中, $\sigma(\cdot)$ 表示激活函数。同样地, 在 stage3 的高分辨率阶段, 通过细粒度几何头联合恢复区域语义、边界响应和几何场, 实现边界位置与局部弯曲形态的一致性重建。

2.5 曲率感知几何监督损失函数

对于边界模糊或轮廓不规则的息肉, 局部边界偏移可能仅涉及少量像素, 区域重叠目标对这类几何误差的约束仍不充分。因此, 本文在主分割监督之外, 引入距离场、曲率和零水平集一致性约束, 分别关注边界邻域的内外距离关系、局部弯曲形态以及边界分支与距离场表示之间的对应关系。引入归一化有符号距离场 Φ , 根据 $\Phi = 0$ 表示对应目标边界、 $\Phi < 0$ 和 $\Phi > 0$ 分别表示边界内外两侧区域这一

几何关系对分割结果和边界形态进行联合约束。主分割分支采用由加权二元交叉熵 (weighted binary cross-entropy, WBCE) 和加权交并比 (weighted intersection over union, WIoU) 组成的结构损失, 即

$$L_{seg} = L_{WBCE}(\hat{Y}, Y) + L_{WIoU}(\hat{Y}, Y) \quad (17)$$

式中, Y 表示真实分割掩码。为增强边界邻域的几何监督, 本文根据真实 SDF 构造窄带权重 $W(x)$ 。该权重在 $|\Phi(x)|$ 接近 0 的边界区域取得较大值, 在远离边界的位置逐渐减小, 因此 SDF 回归损失和曲率监督损失为

$$L_{sdf} = \frac{\sum_{x \in \Omega} W(x) \rho_s(\hat{\Phi}(x) - \Phi(x))}{\sum_{x \in \Omega} W(x) + \varepsilon} \quad (18)$$

$$L_{curv} = \frac{\sum_{x \in \Omega} W(x) \rho_c(\hat{\kappa}(x) - \kappa(x))}{\sum_{x \in \Omega} W(x) + \varepsilon} \quad (19)$$

式中, Ω 表示像素集合, 统一使用 Φ 和 $\hat{\Phi}$ 表示真实 SDF 和预测 SDF, κ 和 $\hat{\kappa}$ 表示真实曲率图和预测曲率图, $\rho_s(\cdot)$ 表示平滑 L1 误差, $\rho_c(\cdot)$ 表示 Charbonnier 误差, ε 为防止分母为 0 的稳定常数项。SDF 损失用于约束边界内外的距离场分布, 曲率监督损失用于约束由距离场导出的局部轮廓变化。

由于 SDF 的零水平集对应目标边界, 本文进一步构造零水平集一致性损失, 使边界预测概率与由 SDF 派生的边界距离场响应保持一致。

$$B_\Phi(x) = \exp\left(-\frac{|\hat{\Phi}(x)|}{\sigma_{zero}}\right) \quad (20)$$

$$L_{zero} = \frac{1}{|\Omega|} \sum_{x \in \Omega} \rho_s(\sigma(\hat{B}(x)) - B_\Phi(x)) \quad (21)$$

式中, $\hat{\Phi}(x)$ 表示预测 SDF 图, $B_\Phi(x)$ 表示 SDF 派生的边界距离场响应, $\sigma(\cdot)$ 表示 Sigmoid 函数, $\rho_s(\cdot)$ 表示平滑 L1 误差。 L_{zero} 使得边界分支在预测 SDF 接近 0 的边界位置时产生较强响应, 并在远离零水平集的位置抑制边界响应, 从而增强边界预测效果与边界几何一致性。因此, 本文将总损失定义为

$$L_{total} = L_{seg} + \lambda_1 L_{bd} + \lambda_2 L_{sdf} + \lambda_3 L_{curv} + \eta_1 L_{zero} + \eta_2 L_{aux} \quad (22)$$

式中, L_{bd} 表示由 WBCE 和加权 Dice 组成的边界监督损失, L_{aux} 表示辅助监督项, λ_1 、 λ_2 、 λ_3 、 η_1 和 η_2 分别表示控制损失项的权重参数。本文在训练中采用课程式几何监督策略, 随着训练过程逐步收缩边界窄带权重的取值范围, 并相应提高几何监督项的权重。

3 实验与分析

3.1 实验数据集

为全面评估本文方法在肠息肉分割任务中的分割精度与跨数据集泛化能力,选取 Kvasir、ClinicDB、ColonDB、EndoScene 和 ETIS 这 5 个公开数据集开展实验,其统计信息如表 1 所示。上述数据集覆盖了多种结肠镜成像场景,既包含高分辨率、近距离观测条件下的息肉图像,也包含受低对比度、光照波动、镜面反光、运动模糊及局部遮挡等因素干扰的复杂样本。

表 1 肠息肉分割的公开数据集统计

Table 1 Statistics of public datasets for colorectal polyp segmentation

数据集	类别	样本量	训练集	测试集
Kvasir	可见	1000	900	100
ClinicDB	可见	612	550	62
ColonDB	不可见	380	-	380
EndoScene	不可见	60	-	60
ETIS	不可见	196	-	196

Kvasir 和 ClinicDB 具有较高标注质量和较丰富的视觉多样性,作为可见数据集用于模型训练与评估;ColonDB 和 ETIS 中小尺度、低对比度且形态差异显著的息肉样本占比较高,作为不可见数据集用于检验模型的跨数据集鲁棒性;EndoScene 则包含一定数量具有模糊和遮挡干扰的挑战性样本,可进一步用于评估模型在分布偏移条件下的适应能力。通过在上述 5 个数据集上的联合测试,可较为系统地验证本文方法在区域分割精度、边界刻画能力及复杂临床场景泛化性能等方面的综合表现。

3.2 实验设置与评价指标

本文所有实验均在 NVIDIA GeForce RTX 4090D 图形处理器(graphics processing unit, GPU)上完成,基于 PyTorch 框架实现。模型采用 Adam 优化器进行训练,初始学习率设置为 1×10^{-4} ,权重衰减系数设置为 1×10^{-8} 。为兼顾训练初期的稳定收敛与后期的持续优化能力,学习率调度策略采用带热重启的余弦退火机制,每 10 个 epoch 进行一次重启,并设置周期扩展系数 $T_{multi} = 2$;在训练后期,将学习

率下限设为 1×10^{-5} ,以避免学习率过度衰减导致性能波动。实验中 batch size 设为 8,总训练轮次设为 300。为增强模型的鲁棒性与泛化能力,对训练样本采用随机旋转、水平翻转和垂直翻转等常规数据增强策略。数据划分选取 ClinicDB 和 Kvasir 中共 1450 幅图像作为训练集,并从上述 5 个数据集中划分 798 幅图像作为测试集(Chen W L 等, 2023);所有输入图像在送入网络前统一调整为 224×224 像素。为保证对比实验的公平性,所有对比方法均在相同的数据划分、输入尺度与优化设置下重新训练。

为定量评价本文方法的分割性能,选取 2 个医学图像分割任务中常用的评价指标 Dice 和 IoU。设预测分割区域为 P ,真实标注区域为 G ,则 Dice 和 IoU 的计算公式分别为

$$Dice = \frac{2|P \cap G|}{|P| + |G|} \quad (23)$$

$$IoU = \frac{|P \cap G|}{|P \cup G|} \quad (24)$$

式中, Dice 主要用于衡量预测结果与真实标注在区域层面的重叠程度,对局部轮廓偏差较为敏感;IoU 对过分割和欠分割具有更高的误差敏感性,能够更真实地反映模型的区域定位精度。在后续实验结果的分析中,采用的平均 Dice 和平均 IoU 为 5 个公开数据集对应指标的算术平均值,不按各数据集样本数量进行加权计算。由于本文方法同时从频域特征增强、边界几何感知等层面提升分割质量,因此采用 Dice 与 IoU 进行评价,能够较为全面地反映模型在整体区域一致性和边界定位方面的性能差异。

3.3 性能评估

3.3.1 与现有主流方法的对比实验

为验证本文方法在肠息肉分割任务中的有效性与先进性,在 EndoScene、ClinicDB、ColonDB、Kvasir 和 ETIS 等 5 个公开数据集上,将本文方法与 19 种具有代表性的医学图像分割方法进行对比,实验结果如表 2 所示。对比方法主要从通用医学图像分割方法和面向肠息肉分割任务的专用方法中进行选取。通用医学图像分割方法涵盖了卷积神经网络、注意力机制以及 CNN-Transformer 混合架构,可以从通用分割模型角度评估本文方法的整体性能。面向肠息肉分割任务的专用方法包含了与本文方法较为接近的空间域特征建模、频域信息辅助建模、边界增强等方法,因而能够较为全面地检验本文方法的分割

表 2 各模型在五个公开数据集上的性能比较

Table 2 Performance comparison of different models on five public datasets

模型	EndoScene		ClinicDB		ColonDB		Kvasir		ETIS	
	IoU (%)	Dice (%)	IoU (%)	Dice (%)	IoU (%)	Dice (%)	IoU (%)	Dice (%)	IoU (%)	Dice (%)
UNet	72.94	81.02	84.49	89.09	58.23	65.81	82.11	87.12	47.03	53.16
U-Mamba	70.44	79.42	81.35	87.37	54.73	63.50	78.75	85.75	42.83	49.21
H-vmunet	70.63	79.40	81.35	86.40	55.85	67.54	77.67	85.29	43.77	47.60
CAUNet	69.67	77.02	77.47	84.16	60.41	68.90	79.54	86.43	46.50	53.69
UNet++	71.78	79.17	81.69	87.62	57.82	65.16	81.70	87.04	42.79	48.65
FUNet	68.47	75.92	81.39	87.79	61.61	69.43	81.53	86.83	43.05	51.12
AttUNet	77.32	84.89	81.74	88.70	61.35	68.88	83.82	88.51	47.15	53.32
CENet	74.90	83.62	80.86	87.38	63.57	71.84	82.37	87.82	53.19	60.75
TransFuse	75.99	84.42	82.90	90.03	64.08	72.72	83.27	88.22	58.69	67.59
SwinUNet	81.64	88.07	85.80	<u>90.53</u>	69.92	77.37	84.16	<u>89.58</u>	59.87	67.70
DEU-Net	<u>82.98</u>	<u>89.41</u>	83.80	88.57	69.28	76.42	83.49	88.37	62.11	69.20
TransUNet	79.67	86.03	85.26	90.45	<u>69.94</u>	77.32	84.82	89.34	61.83	69.05
VMUNet	81.75	87.36	<u>85.38</u>	90.47	70.21	<u>77.73</u>	<u>85.21</u>	89.23	<u>66.05</u>	<u>74.03</u>
SANet	79.51	85.86	83.91	87.61	67.12	75.31	82.73	88.41	63.42	73.05
MSNet	78.72	84.92	84.93	88.12	67.87	75.52	83.26	88.72	64.41	71.93
PEFNet	77.27	85.12	81.35	85.46	63.85	76.31	81.27	87.11	60.23	63.64
M ² UNet	77.41	85.83	83.27	89.08	67.43	76.72	82.41	88.03	61.56	67.66
MEGANet	79.43	85.94	83.94	90.08	69.42	76.23	83.32	89.32	64.57	72.91
WBANet	75.99	86.06	83.86	88.77	66.38	77.19	82.58	88.78	62.10	69.72
本文方法	83.37	89.65	86.18	91.21	70.62	78.03	85.78	91.04	66.79	75.76

注:加粗字体表示每列最优值,下划线表示每列次优值。

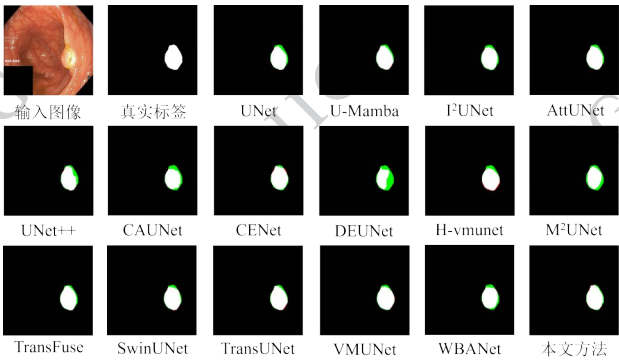
效果。

进一步结合不同数据集的成像特点进行分析,本文方法在 ClinicDB 和 Kvasir 这类图像质量较高、标注较规范、息肉主体较清晰的数据集上,现有方法普遍已经能够取得较高的分割精度,但本文方法仍能在较高基线上继续取得提升,尤其在 Kvasir 数据集上,IoU 和 Dice 分别达到 85.78% 和 91.04%,较最优对比方法分别提升 0.57 和 1.46 个百分点,说明本文方法不仅适用于困难样本,也能够进一步提升常规场景下的特征表达质量和轮廓恢复能力。本文方法在 EndoScene、ETIS 和 ColonDB 这 3 个不可见数据集上均取得最优结果,其中在 ColonDB 上的 IoU 和 Dice 分别达到 70.62% 和 78.03%,说明其对于复杂纹理背景下的不规则息肉结构具有更强的识别能

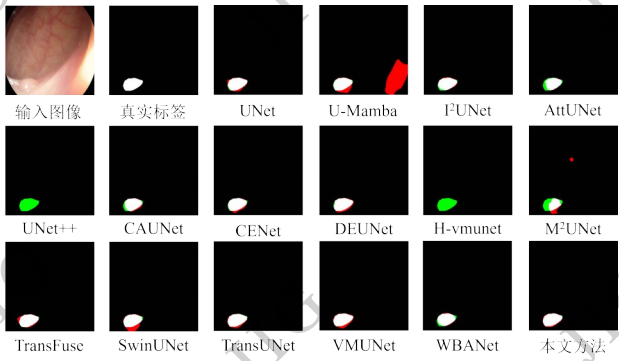
力,也表明频域增强能够有效补偿空间域特征在弱纹理和低对比度区域中的表达不足。特别是在 ETIS 数据集上,其样本具有对比度低、边界模糊等特点,同时还伴随较强的光照波动、镜面反射等因素。

3.3.2 模型复杂度分析

由表 3 可见,本文方法的参数量为 135.87M,高于表中其他对比方法。从浮点运算量看,本文方法的浮点运算量(floating-point operations, FLOPs)为 33.92G,虽然高于 VMUNet、SwinUNet、DEU-Net 和 TransFuse 等轻量或低计算量模型,但低于 UNet、AttUNet 和 SANet 等计算量较高的方法,说明所提模块的参数规模增加并未导致最高的前向计算量,使得计算开销仍处于可接受范围。参数量高的主要原



图(a)



图(b)

图4 Kvasir和ETIS数据集上代表性方法的可视化分割结果
Fig. 4 Visualization results of representative methods on the Kvasir and ETIS datasets

因是网络结构在编码阶段引入频域增强 Transformer 块,并在跳跃连接和解码阶段进一步加入双频带特征融合模块、双阶段几何引导模块以及边界、SDF 和曲率预测分支,使模型能够同时学习区域语义、高频边界和局部几何变化。

然而,本文方法并非以轻量化部署为主要目标,而是面向弱边界、低对比度和复杂形态息肉分割任

务,重点提升边界恢复质量和跨数据集分割稳定性。因此,结合表2结果可知,本文方法在5个公开数据集上的平均IoU和平均Dice均优于对比方法,表明频域增强与几何引导结构能够在增加一定参数规模的情况下换取更好的弱边界和复杂轮廓建模能力。

3.4 消融实验

3.4.1 结构消融

为了系统地评估本文方法中每个模块的贡献,我们在具有不同成像特征和难度水平的多个数据集进行了结构化消融实验,以 UNet 模型作为 Baseline 并采用平均 Dice 和平均 IoU 作为评价指标。消融实验结果如表4所示,4组单一模块实验分别对应了仅引入 STB、DFFM、DGM 和 CGSL 的情形,后续2组双模块组合实验则用于验证频域部分和几何部分的协同作用,为便于对照表4,图5中序号分别对应表4中从上到下的8组消融设置。此外,本文为进一步分析各核心模块内部设计的合理性,继续从 STB 内部结构、DFFM 频带划分参数和 DGM 两阶段边缘图、曲率图变化等3个方面进行补充分析。

由表4可见,4个单一模块在基线模型基础上均带来了稳定的性能提升,表明本文设计的各组成部分均具有独立有效性。其中,单独引入 STB 后,平均 IoU 和平均 Dice 分别提升至 72.73% 和 79.64%,说明编码器中的频域增强有助于提升全局上下文建模能力和深层语义表达;单独引入 DFFM 时,平均 IoU 和平均 Dice 分别达到 73.55% 和 80.39%,说明基于双频带分解的跳跃连接融合能够更充分地保留高低频互补信息。对于几何部分,单独引入 DGM 和 CGSL 时,较基线模型平均 IoU 分别提升了 4.28 和 3.16 个百分点,平均 Dice 分别提升了 4.24 和 3.58 个百分点,表明无论是在解码阶段引入几何引导,还

表3 不同模型的复杂度比较

Table 3 Complexity comparison of different models

模型	参数量/M	浮点运算量/G	模型	参数量/M	浮点运算量/G
UNet	34.53	50.22	VMUNet	44.27	5.77
UNet++	9.16	26.58	I²UNet	29.65	9.36
AttUNet	34.88	51.07	SANet	27.70	44.40
TransFuse	33.35	10.24	DEU-Net	31.75	4.77
SwinUNet	27.17	5.91	MEGANet	44.19	28.84
TransUNet	105.28	25.36	本文方法	135.87	33.92

注:方法均在相同输入尺寸下统计参数量和浮点运算量。

是在训练阶段引入曲率感知损失函数,均有助于改善目标轮廓恢复与边界定位精度。

表 4 本文方法在五个公开数据集上的消融实验结果
Table 4 Ablation study results of the proposed method on five public datasets

模块					平均 IoU(%)	平均 Dice(%)
Baseline	STB	DFFM	DGM	CGSL		
√					69.36	76.05
√	√				72.73	79.64
√		√			73.55	80.39
√			√		73.64	80.29
√				√	72.52	79.63
√	√	√			76.91	83.84
√			√	√	76.60	83.50
√	√	√	√	√	78.55	85.14

注:加粗字体为每列最优值,“√”表示引入对应模块。

进一步地,频域部分的联合改进在平均 IoU 和平均 Dice 分别达到 76. 91% 和 83. 84%,这说明频域部分能够有效提升全局语义建模和细节补偿能力。几何部分的联合改进,在平均 IoU 和平均 Dice 取得了 76. 60% 和 83. 50% 的结果,有助于提高模型对目标轮廓结构的恢复精度和边界定位的稳定性。当 4 个模块共同作用时,完整模型获得了最优结果,平均 IoU 和平均 Dice 分别达到 78. 55% 和 85. 14%,较基线模型分别提升 9. 19 和 9. 09 个百分点,同时也明显优于两组双模块组合结果,表明频域增强与几何引导具有良好的互补性。

根据图 5 给出的不同消融设置下的可视化结果,基线模型在小目标、边界模糊和低对比度场景下容易出现区域缺失、边界粘连及背景误检等问题。在分别引入单一模块后,目标主体区域的分割完整性和边界质量均得到不同程度改善,但局部仍存在误差区域。相较之下,频域组合模型在目标区域恢复和细节保留方面表现更优,几何组合模型在轮廓规整性和边界定位方面更具优势,而完整模型在不同场景下均与真实标签(ground truth, GT)最为接近,误差区域最少,能够更准确地刻画息肉的整体区域与细粒度边界。

为验证 STB 内部结构设计的合理性,本文在保持其他模块不变的情况下,在标准 Transformer 块基础上,依次引入对 V 分支和 FFN 进行 FFT,以及 SE 模块进行内部消融。由表 5 可见,标准 Transformer 块

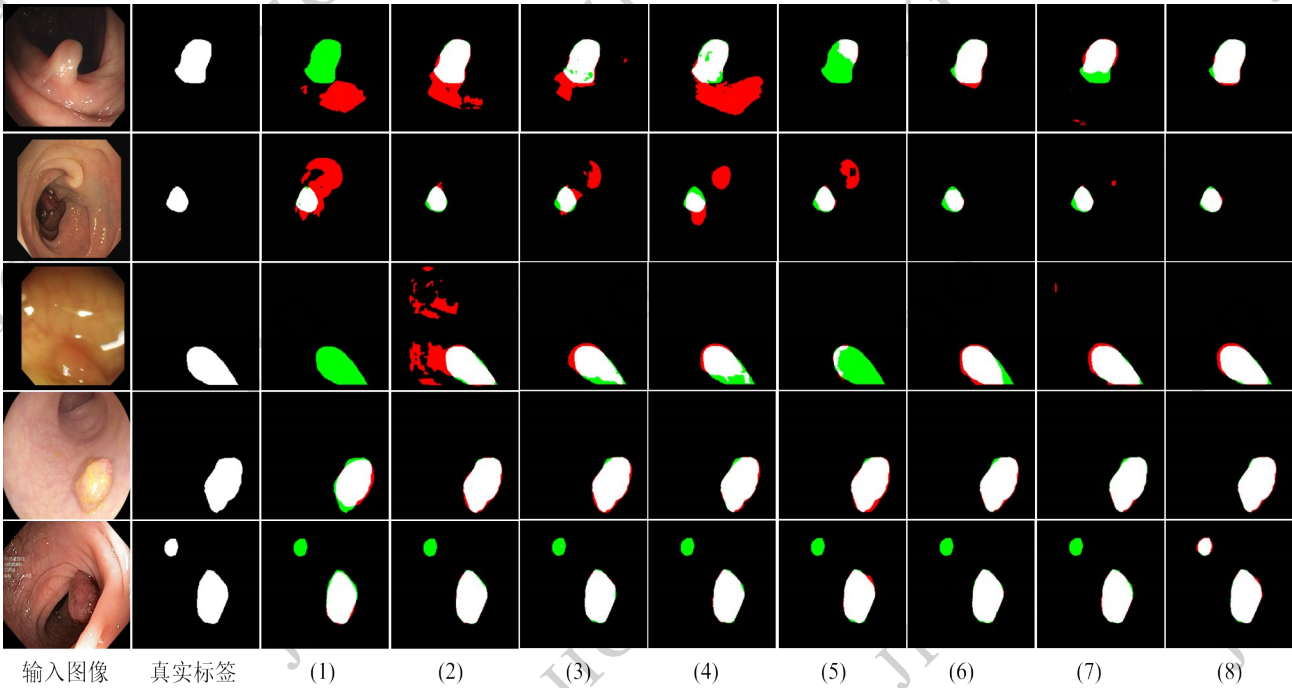
的平均 IoU 和平均 Dice 分别为 76. 99% 和 83. 65%。在仅对 V 分支引入 FFT 后,平均 IoU 和平均 Dice 分别提升至 77. 35% 和 84. 25%,说明在保持 Q, K 空间相关性计算不变的情况下,对 V 分支进行频域增强能够为注意力聚合过程补充边界纹理和局部结构信息。进一步加入频域 SE 后,平均 IoU 和平均 Dice 分别提升至 77. 97% 和 84. 72%,表明通道级频域重标定能够自适应筛选与息肉边界、纹理变化相关的有效频率响应,并抑制由反光和背景纹理引起的无效高频干扰。最后,完整 STB 模块取得最优结果,平均 IoU 和平均 Dice 分别达到 78. 55% 和 85. 14%,说明注意力 V 分支频域增强、频域 SE 通道筛选和前馈网络局部频域增强之间具有互补作用,能够共同提升模型对弱边界和复杂局部轮廓的表征能力。

为系统比较 DGM 分为粗、细两阶段几何引导的贡献,本文从 5 个数据集中选出代表性样本,给出 DGM 模块两阶段边缘图和曲率图的可视化结果。图 6 中每行对应不同数据集中的样本图,从左至右依次为输入图像、GT 边缘图、粗阶段预测边缘图、细阶段预测边缘图、GT 曲率图、粗阶段预测曲率图和细阶段预测曲率图。从图 6 中可以看出,粗阶段几何分支已经能够在较低分辨率下给出目标边界的大致位置,并形成与真实轮廓相一致的初始曲率响应,但其结果仍存在边界偏粗、局部响应扩散和高曲率区域刻画不够集中的现象。经过细阶段几何引导后,预测边缘与真实边缘的重合度进一步提高,弱边

表 5 STB 模块在五个公开数据集上的内部消融实验结果
Table 5 Internal ablation results of STB on five public datasets

模块设置	模块			平均 IoU (%)	平均 Dice (%)
	V+FFT	SE	FFN+FFT		
标准 Transformer 块				76.99	83.65
标准 Transformer 块+仅对 V 分支频域增强	√			77.35	84.25
标准 Transformer 块+仅对 V 分支频域增强+SE	√	√		77.97	84.72
完整 STB 模块	√	√	√	78.55	85.14

注:加粗字体为每列最优值,“√”表示引入对应模块。



(5)Baseline+CGSL、(6)Baseline+STB+DFFM、(7)Baseline+DGM+CGSL、(8)本文方法)

图 5 本文方法消融实验的可视化结果。((1)Baseline、(2)Baseline+STB、(3)Baseline+DFFM、
Fig. 5 Visualization results of ablation experiments of the proposed method. ((1)Baseline、(2)Baseline+STB、(3)Baseline+DFFM、
(4)Baseline+DGM、(5)Baseline+CGSL、(6)Baseline+STB+DFFM、(7)Baseline+DGM+CGSL、(8)Proposed model)

界区域的连续性更好,局部细长结构、边界转折及小尺度凸起位置的轮廓恢复也更加准确。与此同时,细阶段预测曲率图在目标轮廓附近呈现出更集中的响应分布,在局部弯曲显著区域与 GT 曲率图保持了更高一致性,说明细阶段并非对粗阶段结果的简单重复,而是在粗几何先验基础上完成了高分辨率的结构细化与形状校正。结合表 3 中单独引入 DGM 时平均 IoU 和平均 Dice 分别较基线提升 4.28 和 4.24 个百分点,以及几何部分联合引入后平均 IoU 和平均 Dice 分别达到 76.60% 和 83.50% 的结果,可以认为,DGM 的前一阶段负责建立边界邻域和几何场的初始表示,后一阶段则利用窄带权重和曲率先

验进一步强化弱边界和复杂轮廓区域的恢复能力,从而提升模型在低对比度、形态不规则和局部高曲率息肉场景下的分割稳定性。

3.4.2 参数消融

为系统比较 DFFM 中高低频分解的径向阈值对分割性能的影响,本文在保持其他训练参数不变的情况下,将径向阈值分别设置为 0.15、0.20、0.25、0.30、0.35、0.40、0.45、0.50,并统计 5 个公开数据集上的平均 IoU 和平均 Dice 作为结果进行分析。该阈值决定低频掩膜在频谱平面中的覆盖范围,进而影响低频轮廓分量和高频边界分量之间的信息分配。

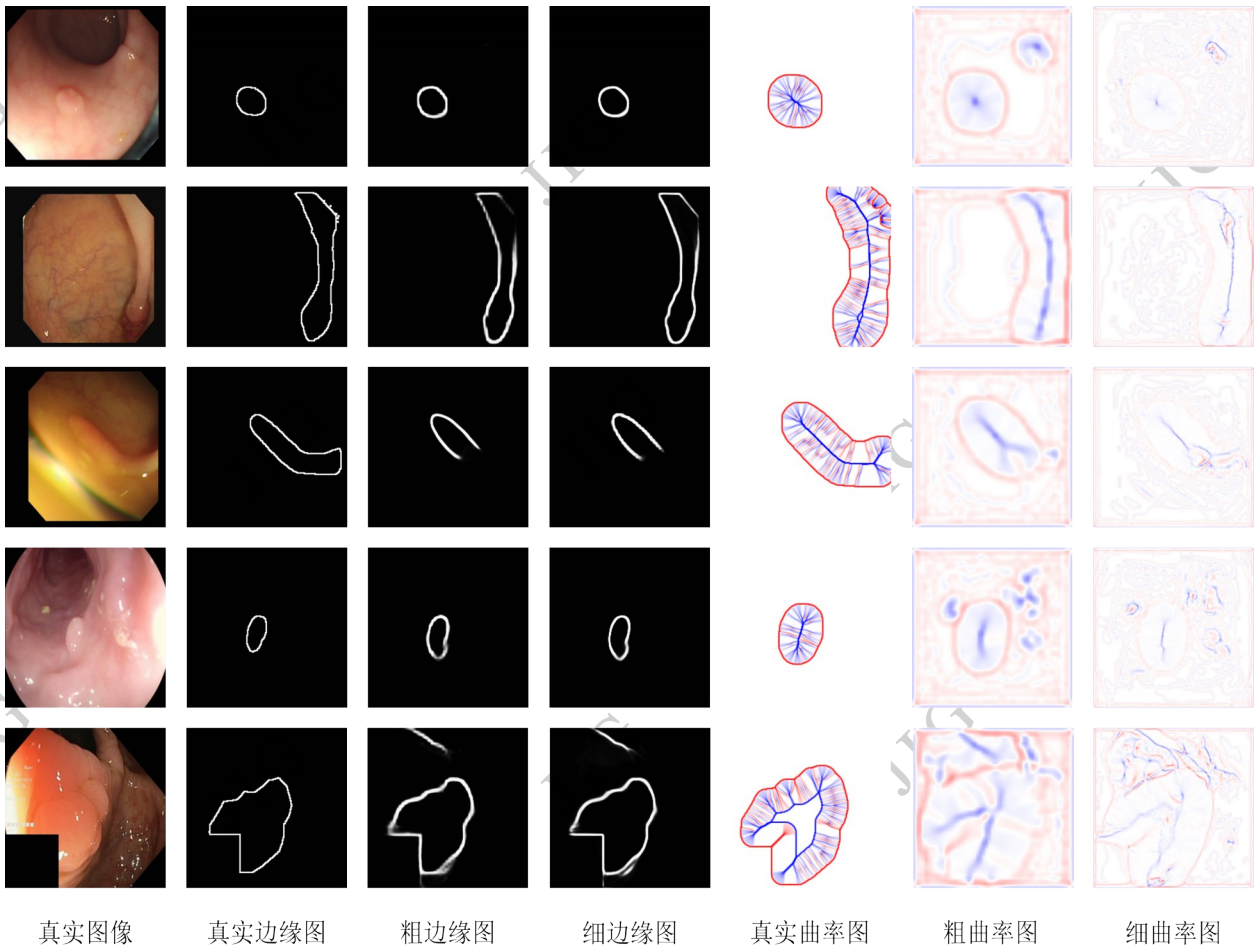


图6 DGM的两阶段可视化结果
Fig. 6 Internal visualization results of DGM.

由表6可见,当径向阈值从0.15增大至0.35时,平均IoU和平均Dice整体呈上升趋势,并在径向阈值为0.35时达到最优,分别为78.55%和85.14%。这说明在本文参数设置下,适当扩大低频覆盖范围有助于保留息肉区域的主体轮廓和结构一致性,同时仍能够通过高频分支保留必要的边界细节。

当阈值进一步增大至0.40、0.45和0.50时,模型性能开始下降,说明过大的低频范围会压缩高频分支的有效表达空间,使部分边界突变、局部纹理和弱边界信息被过度归入低频分量,从而削弱高频分支对精细边界的增强作用。综合平均IoU和平均Dice,本文最终选取0.35作为DFFM的高低频分解阈值。为进一步分析损失函数中几何监督权重对模型性能的影响,本文对SDF回归损失、曲率监督损失和零水平集一致性损失进行联合敏感性分析。SDF回归损失用于约束边界内外的距离场分布,曲率监

表6 DFFM高低频分解径向阈值的参数敏感性分析
Table 6 Parameter sensitivity analysis of the radial threshold for frequency decomposition in DFFM

径向阈值	平均IoU (%)	平均Dice (%)
0.15	76.99	83.63
0.20	76.92	83.46
0.25	77.25	84.32
0.30	77.21	83.87
0.35	78.55	85.14
0.40	77.5	84.33
0.45	76.49	83.12
0.50	76.32	82.93

注:加粗字体为每列最优值。

督损失用于刻画局部轮廓弯曲变化,零水平集一致性损失用于增强预测边界与SDF零水平集之间的对应关系。考虑到上述损失函数之间具有协同作用,

本文未对各权重进行独立网格搜索,而是在几何监督权重组合基础上设置不同缩放系数,以分析几何监督强度变化对分割性能的影响。

表 7 几何监督损失权重敏感性分析
Table 7 Sensitivity analysis of geometric supervision loss weights

权重缩放系数	平均 IoU (%)	平均 Dice (%)
0.5	76.52	83.29
0.75	76.95	83.70
1	77.83	84.65
1.25	78.55	85.14
1.5	77.77	84.37

注:加粗字体为每列最优值。权重缩放系数表示在几何监督权重组合基础上,对 SDF 回归、曲率监督和零水平集一致性损失权重进行同步缩放。

由表 7 可见,当权重缩放系数由 0.5 增大至 1.25 时,模型的平均 IoU 和平均 Dice 整体提升,并在缩放系数为 1.25 时取得最优结果,分别为 78.55% 和 85.14%。这说明适当增强几何监督强度,有助于网络更充分地学习边界邻域的距离场分布、局部曲率变化和零水平集一致性,从而提升弱边界和复杂轮廓区域的恢复能力。当缩放系数继续增大至 1.5 时,平均 IoU 和平均 Dice 有所下降,说明过强的几何监督可能对主分割目标产生一定干扰,使模型过度关注边界局部结构而削弱区域分割稳定性。综合平均 IoU 和平均 Dice 结果,本文最终将几何监督损失权重缩放系数设置为 1.25。

4 结 论

为解决肠息肉内镜图像中高频细节易衰减、弱边界区域难恢复以及复杂轮廓难以准确刻画等问题,本文提出一种融合频域增强与几何引导的肠息肉分割网络。该方法在编码器中通过频域增强 Transformer 块强化全局结构与局部纹理信息的联合表征,在跳跃连接层利用双频带特征融合模块协同建模高频边界细节与低频轮廓信息,并在解码阶段结合双阶段几何引导模块与曲率感知损失函数,对边界位置、距离场分布及局部曲率变化进行联合约束,从而提升了模型对弱对比、小目标及复杂形态息

肉的分割能力。实验结果表明,本文方法在 5 个公开数据集的 IoU 与 Dice 指标均取得最优结果,平均 IoU 和平均 Dice 分别达到 78.55% 和 85.14%,在肠息肉分割任务中表现出较好的分割精度和跨数据集泛化能力。虽然本文方法的计算开销仍然处于可接受的范围,但频域增强和几何引导模块增加了模型的参数规模,可能会提高实际部署成本。后续研究将从轻量化编码器、频域模块压缩、几何分支参数共享和知识蒸馏等方面进一步优化模型结构,在保持弱边界和复杂轮廓分割性能的同时提高推理效率。

参考文献(References)

Bui N T, Hoang D H, Nguyen Q T, Tran M T and Le N. 2024. MEGANet: multi-scale edge-guided attention network for weak boundary polyp segmentation//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 7985-7994 [DOI: 10.1109/WACV57701.2024.00780]

Gao H, Wang Y Y, Chen J, Jiang D S, Zhang X P, Tian Q, et al. 2022. Swin-Unet: Unet-like pure transformer for medical image segmentation//Proceedings of the European Conference on Computer Vision Workshops. Cham: Springer Nature Switzerland: 205-218 [DOI: 10.1007/978-3-031-25066-8_9]

Chen C, Zhou K N, Qi S Y, Lu T and Xiao R X. 2023. A learnable Gabor convolution kernel for vessel segmentation. Computers in Biology and Medicine, 158: 106892 [DOI: 10.1016/j.compbiomed.2023.106892]

Chen J N, Lu Y Y, Yu Q H, Luo X D, Adeli E, Wang Y, et al. 2021. TransUNet: transformers make strong encoders for medical image segmentation[EB/OL]. [2026-05-31]. <https://arxiv.org/abs/2102.04306>

Chen W L, Zhang R, Zhang Y F, Bao F X, Lv H X, Li L H, et al. 2023. Pact-Net: parallel CNNs and Transformers for medical image segmentation. Computer Methods and Programs in Biomedicine, 242: 107782 [DOI: 10.1016/j.cmpb.2023.107782]

Chen Y F, Zhang X Q, Peng L F, He Y D, Sun F and Sun H J. 2024. Medical image segmentation network based on multi-scale frequency domain filter. Neural Networks, 175: 106280 [DOI: 10.1016/j.neunet.2024.106280]

Dai D W, Dong C X, Yan Q S, Sun Y H, Zhang C Y, Li Z F, et al. 2024. I2U-Net: a dual-path U-Net with rich information interaction for medical image segmentation. Medical Image Analysis, 97: 103241 [DOI: 10.1016/j.media.2024.103241]

Gu Z W, Cheng J, Fu H Z, Zhou K, Hao H Y, Zhao Y T, et al. 2019. CE-Net: context encoder network for 2D medical image segmentation. IEEE Transactions on Medical Imaging, 38(10): 2281-2292

[DOI: 10.1109/TMI.2019.2903562]

Huang G H, Liu X, Shi Y Y, Ji Y and Wang S H. 2026. EMPD: An edge-guided multi-perception decoder for medical image segmentation. *Journal of Image and Graphics*, 31(7): 2587-2600 (黄冠辉, 刘翔, 石蕴玉, 纪雨, 王硕鸿. 2026. 面向医学图像分割的边缘引导多重感知解码器. *中国图象图形学报*, 31(7): 2587-2600) [DOI: 10.11834/jig.250251]

Jia L X, Hu Y B, Jin Y, Xue Z Z, Jiang Z W and Zheng Q F. 2023. Colorectal polyp segmentation neural network combined with multiple attention mechanisms. *Journal of Computer-Aided Design & Computer Graphics*, 35(3): 463-473 (贾立新, 胡奕标, 金燕, 薛智中, 姜智伟, 郑秋富. 2023. 融合多种注意力机制的结肠息肉分割神经网络. *计算机辅助设计与图形学学报*, 35(3): 463-473) [DOI: 10.3724/SP.J.1089.2023.19366]

Karimi A, Faez K and Nazari S. 2023. DEU-Net: dual-encoder U-Net for automated skin lesion segmentation. *IEEE Access*, 11: 134804-134821 [DOI: 10.1109/ACCESS.2023.3337528]

Karimi D and Salcudean S E. 2020. Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions on Medical Imaging*, 39(2): 499-513 [DOI: 10.1109/TMI.2019.2930068]

Li P H, Zhou R, He J, Zhao S F and Tian Y. 2023. A global-frequency-domain network for medical image segmentation. *Computers in Biology and Medicine*, 164: 107290 [DOI: 10.1016/j.combiomed.2023.107290]

Liu J Q and Mo J Q. 2025. CAU-Net: U-shaped medical image segmentation network based on compound attention//*Proceedings of the 2024 International Conference on Computer Vision and Image Processing*. Hangzhou, China: SPIE, 13521: 1352107 [DOI: 10.1117/12.3057922]

Liu S J and Tong M J. 2026. Swin-EdgeNet: medical image segmentation with multi-scale edge enhancement based on Transformer. *Journal of Image and Graphics*, 1-14 (刘善洁, 童孟军. 2026. Swin-EdgeNet: 基于Transformer的多尺度边缘增强医学图像分割模型. *中国图象图形学报*, 1-14) [DOI: 10.11834/jig.250592]

Liu X W, Yang L, Chen J G, Yu S Y and Li K Q. 2022. Region-to-boundary deep learning model with multi-scale feature fusion for medical image segmentation. *Biomedical Signal Processing and Control*, 71: 103165 [DOI: 10.1016/j.bspc.2021.103165]

Ma J, Li F F and Wang B. 2024. U-Mamba: enhancing long-range dependency for biomedical image segmentation [EB/OL]. [2026-05-31]. <https://arxiv.org/abs/2401.04722>

Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag C J, Laversanne M, et al. 2023. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN: Gut, 72(2): 338-344 [DOI: 10.1136/gutjnl-2022-327736]

Nguyen-Mau T H, Trinh Q H, Bui N T, Thi P T V, Nguyen M V, Cao X C, et al. 2023. PEFNet: positional embedding feature for polyp

segmentation//*Proceedings of the 29th International Conference on Multimedia Modeling*. Cham: Springer Nature Switzerland: 240-251 [DOI: 10.1007/978-3-031-27818-1_20]

Oktaç O, Schlemper J, Folgoc L L, Lee M, Heinrich M P, Misawa K, et al. 2018. Attention U-Net: learning where to look for the pancreas[EB/OL]. [2026-05-31]. <https://arxiv.org/abs/1804.03999>

Peng Y F and Sun W Q. 2026. Dual-path encoding and adaptive receptive-field-driven medical image segmentation. *Journal of Image and Graphics*, 31(1): 320-334 (彭晏飞, 孙伟强. 2026. 双路径编码与自适应感受野驱动的医学图像分割. *中国图象图形学报*, 31(1): 320-334) [DOI: 10.11834/jig.250128]

Qayoom A, Xie J and Ali H. 2025. Polyp segmentation in medical imaging: challenges, approaches and future directions. *Artificial Intelligence Review*, 58(6): 169 [DOI: 10.1007/s10462-025-11173-2]

Rajasekar D, Theja G, Prusty M R and Chinara S. 2024. Efficient colorectal polyp segmentation using wavelet transformation and AdaptUNet: a hybrid U-Net. *Heliyon*, 10(13): e33655 [DOI: 10.1016/j.heliyon.2024.e33655]

Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//*Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*. Cham: Springer International Publishing: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]

Ruan J C, Li J C and Xiang S C. 2025. VM-UNet: vision Mamba UNet for medical image segmentation. *ACM Transactions on Multimedia Computing, Communications, and Applications*, [DOI: 10.1145/3767748]

Tong Y, Chen Z Q, Zhou Z J, Hu Y, Li X and Qiao X B. 2024. An edge-enhanced network for polyp segmentation. *Bioengineering*, 11(10): 959 [DOI: 10.3390/bioengineering11100959]

Trinh Q H, Bui N T, Nguyen-Mau T H, Nguyen M V, Phan H M, Tran M T, et al. 2023. M2UNet: MetaFormer multi-scale upsampling network for polyp segmentation//*Proceedings of the 31st European Signal Processing Conference*. Helsinki, Finland: IEEE: 1115-1119 [DOI: 10.23919/EUSIPCO58844.2023.10290110]

Wang H L, Wang K N, Hua J, Tang Y, Chen Y, Zhou G Q, et al. 2025. Dynamic spectrum-driven hierarchical learning network for polyp segmentation. *Medical Image Analysis*, 101: 103449 [DOI: 10.1016/j.media.2024.103449]

Wang K, Lu Y T, Liu J and Zhang X H. 2025. GAML: geometry-aware mutual learning for polyp segmentation. *Biomedical Signal Processing and Control*, 108: 107965 [DOI: 10.1016/j.bspc.2025.107965]

Wang R, Chen S, Ji C, Fan J and Li Y. 2022. Boundary-aware context neural network for medical image segmentation. *Medical Image Analysis*, 78: 102395 [DOI: 10.1016/j.media.2022.102395]

Wang Y P, Tian Q, Chu J H and Lu W. 2025. A wavelet-enhanced boundary aware network with dynamic fusion for polyp segmenta-

- tion. *Neurocomputing*, 640: 130259 [DOI: 10.1016/j.neucom.2025.130259]
- Wei J, Hu Y W, Zhang R M, Li Z, Zhou S K and Cui S G. 2021. Shallow attention network for polyp segmentation//*Medical Image Computing and Computer-Assisted Intervention - MICCAI 2021*, Cham: Springer International Publishing: 699-708 [DOI: 10.1007/978-3-030-87193-2_66]
- Wu R K, Liu Y H, Liang P C and Chang Q. 2025. H-vmunet: high-order vision Mamba UNet for medical image segmentation. *Neurocomputing*, 624: 129447 [DOI: 10.1016/j.neucom.2025.129447]
- Xu R, Xu C, Li Z, Zheng T, Yu W and Yang C. 2024. Boundary guidance network for medical image segmentation. *Scientific Reports*, 14(1): 17345 [DOI: 10.1038/s41598-024-67554-0]
- Xue H, Luo Y G, Liu M and Li L. 2024. A lighter hybrid feature fusion framework for polyp segmentation. *Scientific Reports*, 14(1): 23179 [DOI: 10.1038/s41598-024-72763-8]
- Xue Y, Tang H, Qiao Z, Gong G Z, Yin Y, Qian Z, et al. 2020. Shape-aware organ segmentation by predicting signed distance maps//*Proceedings of the AAAI Conference on Artificial Intelligence*. New York, USA: AAAI Press: 12565-12572 [DOI: 10.1609/aaai.v34i07.6946]
- Yan S, Yang B Q, Chen A H, Zhao X M and Zhang S Q. 2025. Multi-scale convolutional attention frequency-enhanced transformer network for medical image segmentation. *Information Fusion*, 119: 103019 [DOI: 10.1016/j.inffus.2025.103019]
- Yang C R and Zhang Z H. 2024. PFD-Net: pyramid Fourier deformable network for medical image segmentation. *Computers in Biology and Medicine*, 172: 108302 [DOI: 10.1016/j.combiomed.2024.108302]
- Yue G H, Zhuo G B, Yan W Q, Zhou T W, Tang C, Yang P, et al. 2024. Boundary uncertainty aware network for automated polyp segmentation. *Neural Networks*, 170: 390-404 [DOI: 10.1016/j.neunet.2023.11.050]
- Zauber A G, Winawer S J, O'Brien M J, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey B F, et al. 2012. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *New England Journal of Medicine*, 366(8): 687-696 [DOI: 10.1056/NEJMoa1100370]
- Zhang Y D, Liu H Y and Hu Q. 2021. TransFuse: fusing Transformers and CNNs for medical image segmentation//*Medical Image Computing and Computer-Assisted Intervention - MICCAI 2021*. Cham: Springer International Publishing: 14-24 [DOI: 10.1007/978-3-030-87193-2_2]
- Zhang Y H, Yang G, Gong C J, Zhang J H, Wang S and Wang Y T. 2024. Polyp segmentation with interference filtering and dynamic uncertainty mining. *Physics in Medicine & Biology*, 69(7): 075016 [DOI: 10.1088/1361-6560/ad2b94]
- Zhao X Q, Zhang L H and Lu H C. 2021. Automatic polyp segmentation via multi-scale subtraction network//*Medical Image Computing and Computer-Assisted Intervention - MICCAI 2021*. Cham: Springer International Publishing: 120-130 [DOI: 10.1007/978-3-030-87193-2_12]
- Zhou Z W, Rahman Siddiquee M M, Tajbakhsh N and Liang J M. 2018. UNet++: a nested U-Net architecture for medical image segmentation//*Deep Learning in Medical Image Analysis*. Cham: Springer International Publishing: 3-11 [DOI: 10.1007/978-3-030-00889-5_1]

作者简介

徐英城,男,博士研究生,主要研究方向为医学图像分割。E-mail: 241106010@stu.sdufe.edu.cn

刘培德,通信作者,男,教授,主要研究方向为计算机视觉、垂直领域大模型及其优化。E-mail: liupd1966@126.com

张云峰,男,教授,主要研究方向为计算机视觉、多模态通用大模型优化。E-mail: yfzhang@sdufe.edu.cn

姚勋祥,男,讲师,主要研究方向为计算机视觉、超分辨率重建。E-mail: Xunxiang.Yao@sdufe.edu.cn

陈宇杰,女,硕士研究生,主要研究方向为计算机视觉、医学图像分割。E-mail: 15513138063@163.com