

# 基于多尺度特征和上下文聚合的结肠息肉图像分割网络<sup>①</sup>



许海英, 徐健皓, 陈平华

(广东工业大学 计算机学院, 广州 510006)

通信作者: 陈平华, E-mail: [phchen@gdut.edu.cn](mailto:phchen@gdut.edu.cn)

**摘要:** 为解决结肠息肉图像语义分割任务中存在的边界不清晰以及分割结果不连贯、不完整甚至丢失的问题, 提出一种基于多尺度特征和上下文聚合的结肠息肉图像分割网络 (colon polyp image segmentation network based on multi-scale features and contextual aggregation, MFCA-Net). 网络选择 PvTv2 作为特征提取主干网络, 设计多尺度特征互补模块 (MFCM) 用以提取丰富的多尺度局部信息, 减少息肉形态变化对分割结果的影响; 设计全局信息增强模块 (GIEM), 构建嵌入位置注意力的大核深度卷积实现对息肉的精确定位, 提升网络辨别复杂背景的能力; 设计高级语义引导的上下文聚合模块 (HSCAM), 以全局特征引导局部特征, 差异性互补和交叉融合浅层细节信息与深层语义信息, 提升分割的连贯性和完整性; 设计边界感知模块 (BPM), 结合传统图像处理方法与深度学习优化边界特征, 实现细粒度分割, 进而获取更清晰的边界. 实验表明, 在 Kvasir、ClinicDB、ColonDB 和 ETIS 等公开的结肠息肉图像数据集上, 所提出的网络均取得相较于当前主流算法更高的 *mDice* 与 *mIoU* 分数, 具有更高的分割准确率和更强的鲁棒性.

**关键词:** 结肠息肉图像分割; 多尺度特征; 上下文聚合; Transformer; 注意力机制

引用格式: 许海英, 徐健皓, 陈平华. 基于多尺度特征和上下文聚合的结肠息肉图像分割网络. 计算机系统应用, 2025, 34(3): 115-123. <http://www.c-s-a.org.cn/1003-3254/9797.html>

## Colon Polyp Image Segmentation Network Based on Multi-scale Features and Contextual Aggregation

XU Hai-Ying, XU Jian-Hao, CHEN Ping-Hua

(School of Computer, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract:** To solve unclear boundaries and incoherent, incomplete, or even lost segmentation results in the semantic segmentation task of colon polyp images, a colon polyp image segmentation network named colon polyp image segmentation network based on multi-scale features and contextual aggregation (MFCA-Net) is proposed. The network selects PvTv2 as the backbone network for feature extraction. The multi-scale feature complement module (MFCM) is designed to extract rich multi-scale local information and reduce the influence of polyp morphology changes on segmentation results. The global information enhancement module (GIEM) is designed. A large-kernel deep convolution embedded with positional attention is constructed to accurately locate polyps and improve the network's ability to distinguish complex backgrounds. The high-level semantic-guided context aggregation module (HSCAM) is designed. It guides local features with global features, complements differences, and cross-fuses shallow details and deep semantic information to improve the coherence and integrity of segmentation. The boundary perception module (BPM) is designed. Boundary features are optimized by combining traditional image processing methods and deep learning methods to

<sup>①</sup> 基金项目: 广东省重点领域研发计划 (2023B1111050010, 2020B0101100001)

收稿时间: 2024-09-08; 修改时间: 2024-09-30; 采用时间: 2024-10-14; csa 在线出版时间: 2025-01-16

CNKI 网络首发时间: 2025-01-17

achieve fine-grained segmentation and obtain clearer boundaries. Experiments show that the proposed network obtains higher *mDice* and *mIoU* scores compared with current mainstream algorithms on the publicly available colon polyp image datasets such as Kvasir, ClinicDB, ColonDB, and ETIS, and has higher segmentation accuracy and robustness.

**Key words:** colon polyp image segmentation; multi-scale feature; contextual aggregation; Transformer; attention mechanism

结直肠癌 (CRC) 是由结肠或直肠黏膜上皮细胞异常增生引起的恶性肿瘤。全球范围内, 结直肠癌发病率和死亡率均位居恶性肿瘤的前列。在我国, 2022 年新增的 CRC 病例高达 51.7 万例, 死亡人数达到 24 万<sup>[1]</sup>。大多数结直肠癌病例源于结肠内长期存在的良性息肉, 这些息肉随着时间推移可能转化为癌细胞。因此, 及时切除早期息肉被视为阻断癌变进程、降低结直肠癌发病风险的最有效措施。临床实践中, 标准化结肠内镜筛查通常用于发现并切除肠道内的异常病变。截至 2019 年, 在中国已有 200 万人受益于标准化内镜筛查, 结直肠癌早期检出率和治愈率分别达到 72.64% 和 83.38%。然而, 结肠镜检查高度依赖医生的主观判断和专业水平, 存在遗漏微小息肉及误诊的风险。因此, 设计一种自动且准确的结肠息肉分割方法, 对提高筛查的效率和准确性具有重要意义。

近年来, 随着深度学习的兴起, 基于卷积神经网络的结肠息肉分割方法应用越来越普及。Long 等<sup>[2]</sup>提出完全卷积神经网络 (FCN), 开创了卷积神经网络在图像分割领域的应用。随后, Brandao 等<sup>[3]</sup>针对结肠息肉分割任务设计了 FCN 变体 FCN-VGG 网络, 引入去除全连接层和分类层的 VGG 网络增强息肉特征提取能力。2015 年, 一种采用 U 形结构的卷积神经网络 U-Net<sup>[4]</sup>从众多神经网络中脱颖而出, 在医学图像分割任务中表现极为出色, 这归功于其编码器-解码器结构能够较好地利用上下文信息来增强特征表达能力。此后, 围绕 U-Net 的改进研究不断涌现, 如 UNet++<sup>[5]</sup>、ResUNet++<sup>[6]</sup>、DoubleUNet<sup>[7]</sup>等。相比其他医学图像, 结肠息肉图像具有形态多样性、低对比度、大小变化范围大等特点, 具有一定的挑战性。为应对这些挑战, Fan 等<sup>[8]</sup>提出反向注意力网络 (parallel reverse attention network, PraNet), 使用反向注意力模块从息肉区域中挖掘边界线索, 建立起区域和边界之间的关系, 同时使用并行部分解码器生成全局映射图引导网络进行分割, 该网络在多个息肉图像数据集上均取得了最优性能, 但由于其生成的全局映射图为高维度低分辨率特征图, 对小息肉感

知能力较差, 分割精度受限。Zhong 等<sup>[9]</sup>从息肉颜色分布差异的角度出发, 提出 SANet, 采用颜色交换操作解耦图像内容和颜色, 迫使网络更多地关注息肉目标的形状和结构<sup>[10]</sup>。Qiu 等<sup>[11]</sup>从边界信息入手, 提出边界分布引导网络 (BDG-Net), 设计边界分布生成模块 (BDGM) 聚合高级特征并生成边界分布图辅助分割。Liu 等<sup>[12]</sup>则从多尺度特征融合角度出发, 提出一个利用多尺度通道空间融合网络的实时自动分割框架 (MCSF-Net), 该框架利用多尺度特征融合模块, 结合空间和通道注意力机制, 有效融合高维度多尺度特征, 并且在形状模块中使用简单的减法操作突出息肉边界, 提高分割精度。

2020 年, 首个应用于计算机视觉领域的 Transformer 网络 vision Transformer<sup>[13]</sup>被提出。Transformer<sup>[14]</sup>的引入为医学图像分割性能提升带来了新的思路。Chen 等<sup>[15]</sup>提出 TransUNet 架构, 首次将 Transformer 嵌入 U-Net 结构, 显著提升了结肠息肉分割的准确性和鲁棒性, 开创了新的医学图像分割方法。Zhang 等<sup>[16]</sup>提出 TransFuse 网络, 创建了一种新的特征融合方式——BiFusion 模块, 有效融合来自 CNN 和 Transformer 的多层次特征, 显著增强了息肉分割性能。Wang 等<sup>[17]</sup>提出 PvT 网络, 利用金字塔结构增强特征提取能力, 针对性优化了对不同尺度息肉的分割效果。

为应对息肉分割的现实挑战, 本文通过研究提高息肉分割结果完整性和连续性的策略, 提出一个基于多尺度特征和上下文聚合的结肠息肉图像分割网络 MFCA-Net (colon polyp segmentation network based on multi-scale features and contextual aggregation)。具体工作如下。

1) 提出了一种结肠息肉图像分割网络 MFCA-Net, 包括多尺度特征互补模块、全局信息增强模块、高级语义引导的上下文聚合模块以及边界感知模块, 通过模块间的协同作用, 提升对结肠息肉图像分割的准确性和鲁棒性。

2) MFCA-Net 从多尺度视角出发, 分层次挖掘并增强浅层特征和深层特征, 同时利用高级语义引导的上下文聚合模块进行特征融合, 生成精确分割图。边缘

感知模块则探索了提取息肉边界特征的新范式,结合传统图像处理方法与深度学习方法优化边界特征,纠正分割中边界预测不准确的问题。

3) 在 Kvasir<sup>[18]</sup>、ClinicDB<sup>[19]</sup>、ColonDB<sup>[20]</sup>和 ETIS<sup>[21]</sup>等 4 个公开的结肠息肉图像数据集上进行了多个对比实验。实验结果表明,在  $mDice$  和  $mIoU$  评价指标上,所提出的 MFCA-Net 超越了包括 U-Net、PraNet、TransFuse 等现有网络。同时,设计了系列消融实验,验证 MFCA-Net 中各模块、方法的有效性。

## 1 网络结构

结肠息肉图像分割主要面临两大挑战:一是因息肉尺寸变化和形态多样性所引起的分割结果不完整、不连续甚至丢失;二是息肉与周围背景对比度不足或噪声干扰导致的边界分割不清晰。为应对上述挑战,本文提出了一种基于多尺度特征和上下文聚合的结肠息

肉图像分割网络 MFCA-Net,如图 1 所示。网络采用 PvTv2 (pyramid vision Transformer v2)<sup>[22]</sup>编码器处理输入图像,从 4 个不同阶段提取特征图,其大小分别对应原图分辨率的 1/4、1/8、1/16 和 1/32,并分别标记为  $F_1$ 、 $F_2$ 、 $F_3$  和  $F_4$ 。其中,  $F_1$ 、 $F_2$ 、 $F_3$  包含了丰富细节信息,将其定义为浅层特征,  $F_4$  则提供了丰富的语义信息,将其定义为深层特征。为减少计算开销,网络省略了大量跳跃连接,改为通过多尺度特征互补模块充分挖掘浅层特征的局部表示。全局信息增强模块利用多尺度深度卷积结合注意力机制提升息肉区域定位能力。随后,通过高级语义引导的上下文聚合模块,将局部特征与全局高维特征图有效融合,生成准确的分割图。此外,边界感知模块应用小波变换<sup>[23]</sup>和边缘检测算子<sup>[24]</sup>,结合反向注意力机制<sup>[25]</sup>和逆残差结构<sup>[26]</sup>,提升对息肉边界细节的捕捉能力,改善网络在处理复杂结构时的性能。

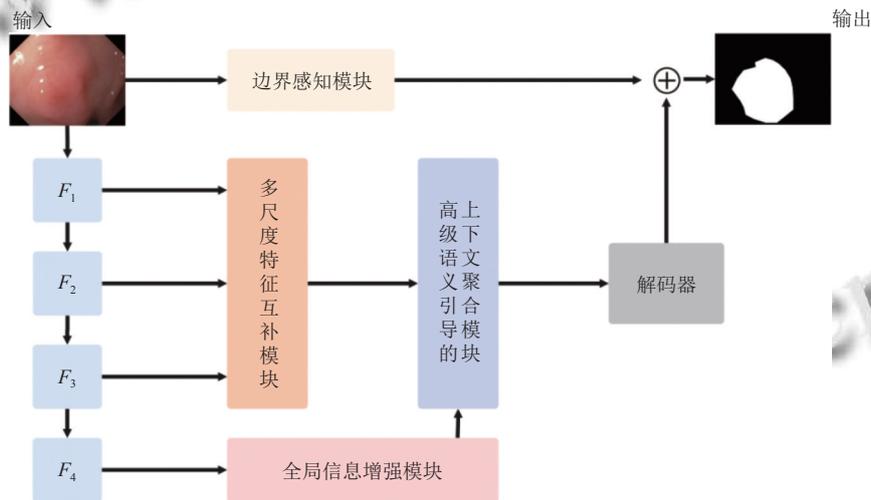


图1 MFCA-Net 框架

### 1.1 多尺度特征互补模块

多尺度特征互补模块包含局部特征增强分支和局部特征注入分支<sup>[27]</sup>,旨在通过特征交互与融合丰富局部上下文信息的有效表示,提升编码器的特征表示能力,如图 2 所示。浅层特征  $F_1$ 、 $F_2$ 、 $F_3$  同时输入到局部特征增强分支,生成增强后的局部特征表示;随后,增强后的特征与初始的浅层特征  $F_1$ 、 $F_2$ 、 $F_3$  融合,得到模块输出特征  $F'$ 。

局部特征增强分支的主要创新在于提出了结合深度可分离卷积的自适应多尺度特征学习策略<sup>[28]</sup>,如图 3 所示。首先,  $F_1$  和  $F_3$  分别经过下采样和上采样调

整至与  $F_2$  一致的空间尺寸,随后通过拼接和  $1 \times 1$  卷积操作得到融合特征后进入 3 个平行分支:第 1 个分支通过最大池化操作和  $3 \times 3$ 、 $5 \times 5$ 、 $7 \times 7$  这 3 个尺度的深度可分离卷积进行特征提取;第 2 个分支通过双线性插值和相同的多尺度深度可分离卷积处理特征;第 3 个分支则将第 1 和第 2 分支处理后的特征进行拼接,以补偿特征交互过程中可能丢失的信息。这种设计策略应用了不同尺度的深度可分离卷积为网络提供了多样的感知能力,使网络能够获得更为全面的特征表示;同时,结合多并行结构,允许网络自适应地选择当前任务最合适的特征尺寸与感受野,有效捕获不

同尺度的特征信息,增强网络对复杂息肉目标的区分能力.

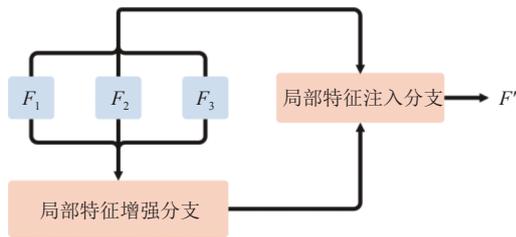


图2 多尺度特征互补模块

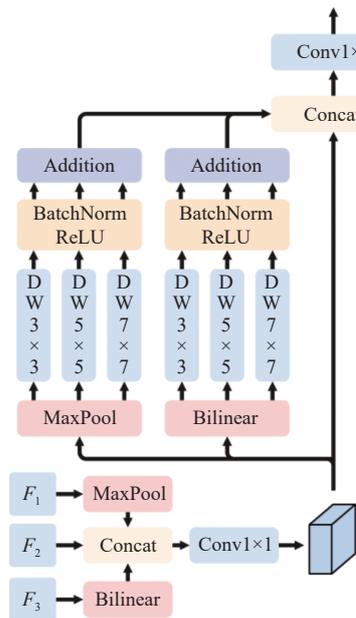


图3 局部特征增强分支

局部特征注入分支的算法流程图如图4所示。 $F_1$ 、 $F_2$ 、 $F_3$ 被对齐至一致的空间尺寸后进行拼接,并通过 $1 \times 1$ 卷积调整通道数;随后,将 $F'$ 通过 $1 \times 1$ 卷积和 Sigmoid 函数激活处理,生成语义权重,并与对齐后的特征进行元素层级的加权融合;最后,通过跳跃连接与 $F'$ 相加,得到模块输出 $F_{local}$ 。通过融合操作使每个像素在特征图中均受到相应语义权重影响,使网络能够更准确地捕捉关键语义信息,从而提高息肉分割的精度和质量。

### 1.2 全局信息增强模块

全局信息增强模块采用结合位置注意力<sup>[29]</sup>的多尺度卷积注意力(ca-MSCA)模块<sup>[30]</sup>,捕获息肉图像特征的远程依赖关系,提高网络对息肉区域的精确定位,提升分割性能。该模块的整体结构类似 ViT,但将其中的自注意力机制替换为 ca-MSCA 模块,如图5所示。ca-MSCA 模块主要包括4个部分,分别为:用于聚合局部

特征信息的 $5 \times 5$ 深度卷积、用于捕获多尺度上下文的多分支深度卷积、用于整合不同尺度特征信息的位置注意力机制,以及用于模拟不同通道之间关系的 $1 \times 1$ 卷积。另外,在模块最后, $1 \times 1$ 卷积的输出将被视为注意力权重,借助跳跃连接对模块的初始输入进行重新加权,得到模块输出特征。

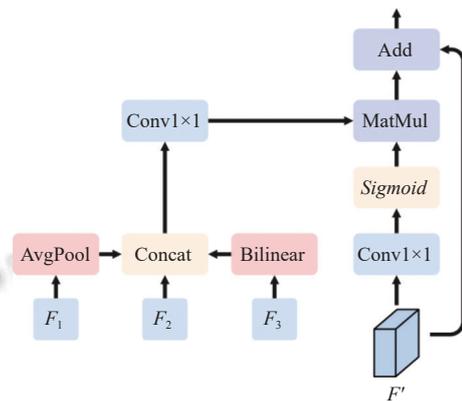


图4 局部特征注入分支

多分支深度卷积的每个分支均采用一对深度条形卷积,例如 $1 \times 7$ 和 $7 \times 1$ 的深度条形卷积,以串联形式代替对应尺度的标准大核卷积。在各个分支中,深度条形卷积的内核大小分别设置为7、11和21。借助条形卷积的轻量化特点,模块能够在减少计算负担的同时保持强大的特征表征能力,对分割场景中存在的条状息肉特征进行有效提取,提高对息肉区域的定位精度。

结肠息肉图像分割任务中,位置信息是捕捉息肉空间结构的关键。模块采用位置注意力机制作为捕获远程依赖关系的主要方法,提升网络辨别复杂背景的能力<sup>[31]</sup>,如图6所示。位置注意力机制使用两个一维全局平均池化操作替代常规的二维全局平均池化操作,分别沿水平和垂直方向进行特征压缩,促使注意力机制捕获远程空间交互和精确位置信息。

### 1.3 高级语义引导的上下文聚合模块

为有效利用全局特征指导局部特征在关键区域进行精确预测,设计高级语义引导的上下文聚合模块。模块通过充分利用上下文信息,为网络提供更全面的图像背景信息以及分割目标间的关系,提高网络对息肉边界、形状和内部结构的感知能力,如图7所示。模块首先通过 $1 \times 1$ 卷积、Sigmoid 激活函数和双线性插值操作对全局特征权重进行重新标定,这一过程量化了全局特征对不同位置像素的重要性;随后,将量化后的全局特征与局部特征进行元素层级融合,同时,引入跳跃

连接将上采样后的初始全局特征与融合结果相加, 弥补融合过程中可能造成的信息损失, 得到输出特征  $F_{out}$ . 计算推导式如式 (1)–式 (4) 所示:

$$F_l = Conv1 \times 1(AvgPool(F_{local})) \quad (1)$$

$$F_g = Bilinear(Sigmoid(Conv1 \times 1(F_{global}))) \quad (2)$$

$$F_{lg} = MatMul(F_l, F_g) \quad (3)$$

$$F_{out} = F_{lg} + Bilinear(Conv1 \times 1(F_{global})) \quad (4)$$

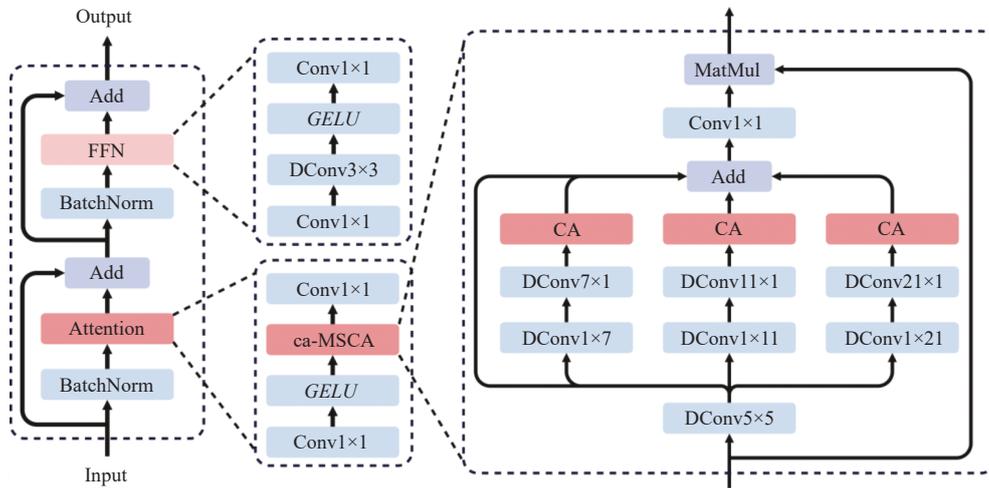


图5 全局信息增强模块

于局部特征  $F_{local}$  所计算的反向注意力权重相乘, 得到输出特征  $F_a$ .

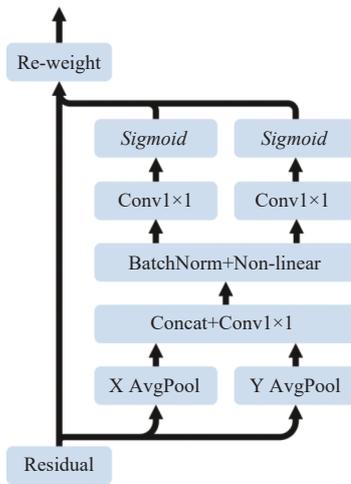


图6 位置注意力机制

### 1.4 边界感知模块

边界感知模块对输入图像应用小波变换和边缘检测算子, 提取边界特征, 结合反向注意力机制和逆残差结构, 提升网络对边界细节的捕捉能力. 具体而言, 如图 8 所示, 模块采用 8 个方向 ( $0^\circ$ 、 $22.5^\circ$ 、 $45^\circ$ 、 $67.5^\circ$ 、 $90^\circ$ 、 $112.5^\circ$ 、 $135^\circ$ 和  $157.5^\circ$ ) 的边缘检测算子捕获不同方向上的边缘梯度信息; 同时, 利用小波变换将输入图像分解成不同频率和方向的子图像提取边缘特征; 接着, 将两者相加, 以充分融合多尺度、多方向、多频率的息肉边界信息; 随后, 将该融合结果与基

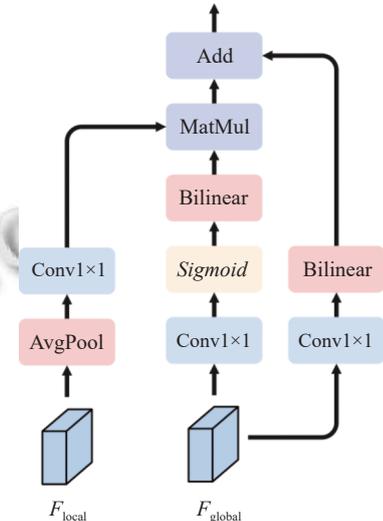


图7 高级语义引导的上下文聚合模块

接下来, 应用结合空间注意力机制的逆残差结构, 进一步增强反向注意力所输出的特征  $F_a$ , 以获得更为精细的边界特征. 具体而言, 首先通过  $1 \times 1$  卷积扩充通道数, 随后应用空间注意力机制计算特征图的空间注意力权重, 这一步骤有助于模型聚焦关键边界信息, 同时抑制无关背景噪声; 最后, 通过  $1 \times 1$  卷积恢复原通道数, 在压缩高维特征以减少计算负担和参数量的同时

使关键特征信息得到保留. 通过上述逆残差结构的处理, 反向注意力特征得到了进一步的细化和增强. 逆残

差结构的设计使其能够有效地融合和强化多层次的特征信息, 尤其在边界细节特征的捕捉上表现出色.

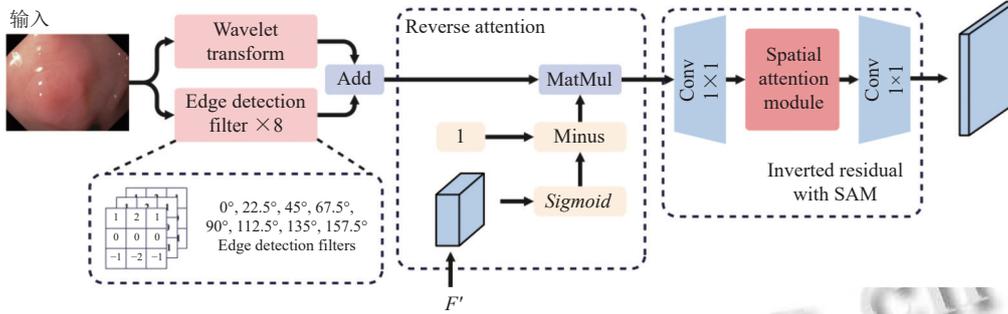


图8 边界感知模块

## 2 实验

实验环境为: CPU AMD EPYC 9754 128-Core, 内存 60 GB, GPU NVIDIA GeForce RTX3090, 显存 24 GB, Python 3.8, Linux Ubuntu 操作系统, PyTorch 2.0.0 深度学习框架.

### 2.1 实验数据集

实验选取了广泛应用于胃肠道疾病研究的4个公开医学图像数据集, 分别是Kvasir、ClinicDB、ColonDB和ETIS. 这些数据集因高质量的图像和精确的分割标签, 在胃肠道疾病的诊断与治疗中扮演着重要角色. 其中, Kvasir数据集包含了1000张不同分辨率的息肉图像及其相应的标签, 内容涵盖多种胃肠道病变形态, 包括但不限于息肉、溃疡和出血等. 对比实验中, 该数据集的900张图像被用于网络训练, 剩余100张则用于网络测试. ClinicDB数据集专注于结直肠癌的检测和研究, 包含612张统一尺寸为388×284像素的结肠镜图像, 其中550张图像用于网络训练, 62张用于网络测试. ColonDB数据集包括来自15个内窥镜视频的380张白光图像, 清晰展示了结肠镜检查的真实临床情况, 其图像尺寸统一为574×500像素. 最具分割挑战性的ETIS数据集包含了196张尺寸为1225×966像素的高分辨率图像, 该数据集中存在难以分割的小息肉, 分割任务复杂性大. 对比实验中, ColonDB的380张图像和ETIS的196张图像均用于网络测试. 本文选取了3张息肉示例图, 分别来自ColonDB、Kvasir和Clinic数据集, 如图9所示.

### 2.2 实验参数设置

实验参数设置如下: 初始学习率为1E-4, 衰减率为0.1, 衰减周期为30轮次, 使用AdamW优化器, 共

进行100轮次训练; 数据增强方法采用归一化及0.75倍、1倍、1.25倍的多尺度输入方法.

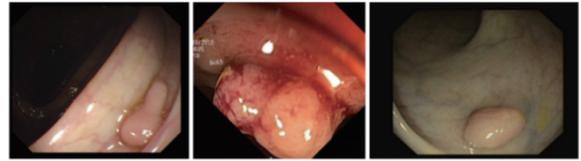


图9 数据集图片

### 2.3 实验评价指标与损失函数

为准确评估网络的有效性, 实验选取两个图像分割任务中常见且具有代表性的指标,  $mDice$  分数和  $mIoU$  分数, 两者数值越高, 代表息肉图像分割网络的分割准确率越高.  $mDice$  与  $mIoU$  分数表示对所有测试样本的  $Dice$  与  $IoU$  分数取平均值.  $Dice$  与  $IoU$  的具体计算公式分别如式(5)和式(6)所示, 两者通常作为评估图像分割任务中预测结果与真实标签之间匹配程度的重要指标.  $mDice$  与  $mIoU$  的具体计算公式如式(7)和式(8)所示. 其中  $A$  表示网络对结肠息肉图像分割的结果图,  $B$  表示对应的标签图像,  $n$  表示假设的测试样本总数<sup>[32]</sup>.

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (5)$$

$$IoU = \frac{A \cap B}{A \cup B} \quad (6)$$

$$mDice = \frac{1}{n}(Dice_1 + Dice_2 + \dots + Dice_n) \quad (7)$$

$$mIoU = \frac{1}{n}(IoU_1 + IoU_2 + \dots + IoU_n) \quad (8)$$

为提高网络对边界与重点区域的分割性能, 网络使用由加权  $IoU$  与加权 BCE 相加所构建的联合损失函数. 加权  $IoU$  与加权 BCE 分别如式(9)和式(10)所示. 其

中,  $N$ 表示像素的总数,  $y_i$ 表示第 $i$ 个像素的标签值,  $\hat{y}_i$ 表示第 $i$ 个像素的预测值.  $w_i$ 表示第 $i$ 个像素的权重系数.

$$Loss_{IoU} = 1 - \frac{\sum_{i=1}^N w_i \times y_i \times \hat{y}_i}{\sum_{i=1}^N w_i \times (y_i + \hat{y}_i - y_i \times \hat{y}_i)} \quad (9)$$

$$Loss_{BCE} = -\frac{1}{N} \sum_{i=1}^N w_i \times [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (10)$$

## 2.4 实验结果与分析

为验证所提出网络的有效性性与鲁棒性, 本文选取包括 U-Net、UNet++、PraNet、TransFuse、M<sup>2</sup>SNet<sup>[33]</sup> 等在内的一系列具有较好上下文聚合与表示能力的主流息肉图像分割网络作为对比网络, 基于相同的测试数据集, 开展对比实验.

定量分析结果如表 1 所示, 所提出的 MFCA-Net 在 4 个数据集上均取得了更高的  $mDice$  和  $mIoU$  分数.

实验结果表明, 所提出的上下文聚合框架可以高效利用局部与全局信息的推导与融合, 增强网络特征表示能力, 提高分割精度. 特别是对于因息肉区域小导致分割困难的 ETIS 数据集, MFCA-Net 与次优网络相比性能提升最为明显,  $mDice$  和  $mIoU$  分数均提高近 2 个百分点, 有力证明了所提出网络的有效性.

## 2.5 消融实验结果与分析

为探究所提出模块对网络整体性能的有效性与贡献程度, 基于相同的实验环境、实验数据集、实验参数设置, 设计了一系列消融实验. 实验结果如表 2 所示. 具体来说, 表 2 中, M1 表示仅使用 PvTv2 特征提取主干网络进行结肠息肉图像分割, 不添加其他任何模块; M2 表示在 M1 的基础上添加多尺度特征互补模块 (MFCM); M3 表示在 M2 的基础上添加全局信息增强模块 (GIEM); M4 表示在 M3 的基础上添加高级语义引导的上下文聚合模块 (HSCAM); M5 表示在 M4 的基础上添加边界感知模块 (BPM), 即所提出的完整网络 (MFCA-Net).

表 1 不同网络在 Kvasir、ClinicDB、ColonDB 和 ETIS 数据集上的对比实验结果

Networks	Year	Kvasir		ClinicDB		ColonDB		ETIS	
		$mDice$	$mIoU$	$mDice$	$mIoU$	$mDice$	$mIoU$	$mDice$	$mIoU$
UNet	2015	0.818	0.746	0.823	0.750	0.512	0.444	0.398	0.335
UNet++	2018	0.821	0.743	0.794	0.729	0.483	0.410	0.401	0.344
SFA <sup>[34]</sup>	2019	0.723	0.611	0.700	0.607	0.469	0.347	0.297	0.217
PraNet	2020	0.898	0.840	0.899	0.849	0.709	0.640	0.628	0.567
TransFuse-S	2021	0.918	0.868	0.918	0.868	0.773	0.696	0.733	0.659
EU-Net <sup>[35]</sup>	2021	0.908	0.854	0.902	0.846	0.756	0.681	0.687	0.609
SANet	2021	0.904	0.847	0.916	0.859	0.753	0.670	0.750	0.654
UACANet-L <sup>[36]</sup>	2021	0.912	0.859	0.926	0.880	—	—	0.766	0.689
LAPFormer-S <sup>[37]</sup>	2022	0.910	0.857	0.901	0.849	0.781	0.695	0.768	0.686
M <sup>2</sup> SNet	2023	0.912	0.861	0.922	0.880	0.758	0.685	0.749	0.678
IECFNet <sup>[38]</sup>	2024	0.907	0.856	0.924	0.873	0.775	0.632	0.707	0.632
Ours	2024	<b>0.920</b>	<b>0.868</b>	<b>0.935</b>	<b>0.886</b>	<b>0.798</b>	<b>0.712</b>	<b>0.789</b>	<b>0.703</b>

表 2 网络在 Kvasir、ClinicDB、ColonDB 和 ETIS 数据集上的消融实验结果

Network	Description	Kvasir		ClinicDB		ColonDB		ETIS	
		$mDice$	$mIoU$	$mDice$	$mIoU$	$mDice$	$mIoU$	$mDice$	$mIoU$
M1	PvTv2	0.887	0.824	0.877	0.817	0.780	0.689	0.728	0.627
M2	PvTv2+MFCM	0.906	0.850	0.925	0.875	0.788	0.703	0.762	0.671
M3	PvTv2+MFCM+GIEM	0.916	0.859	0.931	0.881	0.786	0.698	0.765	0.680
M4	PvTv2+MFCM+GIEM+HSCAM	0.917	0.861	0.934	0.884	0.792	0.706	0.776	0.690
M5	MFCA-Net: PvTv2+MFCM+GIEM+HSCAM+BPM	<b>0.920</b>	<b>0.868</b>	<b>0.935</b>	<b>0.886</b>	<b>0.798</b>	<b>0.712</b>	<b>0.789</b>	<b>0.703</b>

如表 2 所示, M1 在 4 个数据集上均取得较低的  $mDice$  和  $mIoU$  分数, 说明该网络对息肉区域的特征提取存在一定的局限性, 导致特征信息丢失较多, 分割精度较低. M2 在 M1 的基础上添加多尺度特征互补模块,

$mDice$  和  $mIoU$  分数均有所提升, 表明多尺度特征互补模块能够提取有效的多尺度局部上下文信息, 增强网络的特征表示能力, 减少息肉形态变化对分割结果的影响. M3 相比于 M2,  $mDice$  和  $mIoU$  分数有所提升,

说明全局信息增强模块能有效提高对息肉区域的精确定位,减少漏检情况发生.相比于M3, M4中的分割指标均呈上升趋势,说明多尺度特征互补模块和全局信息增强模块所提取的特征信息是充分的,且高级语义引导的上下文聚合模块能有效建立局部细节信息和全局语义信息的关联关系,提高分割精度. M5相比于M4,指标提升得最多,因为边缘感知模块能够有效提取边界信息,从而提升分割结果的连续性与完整性.综合来说,消融实验结果表明了所提方法的合理性和模块的有效性.

### 3 结论与展望

为提升结肠息肉图像分割的精确度,提出了一种基于多尺度特征和上下文聚合的结肠息肉分割网络MFCA-Net. MFCA-Net以Transformer架构为核心特征编码器,精心设计4个关键模块:多尺度特征互补模块通过融合与交互多尺度信息,有效捕获不同尺度下的局部上下文信息;全局信息增强模块捕获远距离依赖关系并提取关键的全局上下文信息;上下文聚合模块将局部与全局信息有效整合,以更准确还原目标息肉形态;边界感知模块专注于提高模型在边界定位方面的性能. MFCA-Net在Kvasir、ClinicDB、ColonDB和ETIS等4个结肠息肉图像数据集上进行了广泛的对比实验,以验证其有效性.同时,通过一系列消融实验,进一步证实了所提出模块的性能优势.

尽管MFCA-Net在总体上取得了成功,但其在处理如多个小息肉聚合这类复杂的分割情况时仍面临挑战.此外,由于个体情况与临床检查环境各异,结肠内部情况与所拍摄的息肉图像更是呈现复杂态势,存在光照多变或出现不同程度的遮挡等情况.本研究并未详细探究在影响息肉可见度因素下网络的泛化性和适用性.因此,未来研究的重点将探索有效提升不同光照条件下分割效果的方法,以及有效恢复遮挡信息的优化方法;同时,后续研究将探索建立更广泛的结肠息肉图像数据集,以进一步分析和增强所提出网络的泛化性和适用性.

#### 参考文献

- 1 Han BF, Zheng RS, Zeng HM, *et al.* Cancer incidence and mortality in China, 2022. *Journal of the National Cancer Center*, 2024, 4(1): 47–53. [doi: [10.1016/j.jncc.2024.01.006](https://doi.org/10.1016/j.jncc.2024.01.006)]
- 2 Long J, Shelhamer E, Darrell T. Fully convolutional networks

- for semantic segmentation. *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015. 3431–3440.
- 3 Brandao P, Mazomenos E, Ciuti G, *et al.* Fully convolutional neural networks for polyp segmentation in colonoscopy. *Proceedings of the Medical Imaging 2017: Computer-aided Diagnosis*. Orlando: SPIE, 2017. 101340F.
- 4 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention*. Munich: Springer, 2015. 234–241.
- 5 Zhou ZW, Siddiquee MMR, Tajbakhsh N, *et al.* UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 2020, 39(6): 1856–1867.
- 6 Jha D, Smedsrud PH, Riegler MA, *et al.* ResUNet++: An advanced architecture for medical image segmentation. *Proceedings of the 2019 IEEE International Symposium on Multimedia (ISM)*. San Diego: IEEE, 2019. 225–230.
- 7 Jha D, Riegler MA, Johansen D, *et al.* DoubleU-Net: A deep convolutional neural network for medical image segmentation. *Proceedings of the 33rd IEEE International Symposium on Computer-based Medical Systems (CBMS)*. Rochester: IEEE, 2020. 558–564.
- 8 Fan DP, Ji GP, Zhou T, *et al.* PraNet: Parallel reverse attention network for polyp segmentation. *Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention*. Lima: Springer, 2020. 263–273.
- 9 Zhong ZL, Lin ZQ, Bidart R, *et al.* Squeeze-and-attention networks for semantic segmentation. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020. 13062–13071.
- 10 刘方瑾. 基于卷积神经网络和Transformer的结肠息肉分割方法研究[硕士学位论文]. 烟台: 山东工商学院, 2023. [doi: [10.27903/d.cnki.gsdsg.2023.000048](https://doi.org/10.27903/d.cnki.gsdsg.2023.000048)]
- 11 Qiu ZH, Wang ZC, Zhang MM, *et al.* BDG-Net: Boundary distribution guided network for accurate polyp segmentation. *Proceedings of the Medical Imaging 2022: Image Processing*. San Diego: SPIE, 2022. 1203230.
- 12 Liu WK, Li ZG, Xia JA, *et al.* MCSF-Net: A multi-scale channel spatial fusion network for real-time polyp segmentation. *Physics in Medicine & Biology*, 2023, 68(17): 175041.
- 13 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *Proceedings of the 9th International Conference on Learning Representations*. OpenReview.net, 2021.
- 14 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Proceedings of the 31st International Conference on*

- Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 15 Chen JN, Lu YY, Yu QH, *et al.* TransUNet: Transformers make strong encoders for medical image segmentation. arXiv:2102.04306, 2021.
  - 16 Zhang YD, Liu HY, Hu Q. TransFuse: Fusing Transformers and CNNs for medical image segmentation. Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Strasbourg: Springer, 2021. 14–24.
  - 17 Wang WH, Xie EZ, Li X, *et al.* Pyramid vision Transformer: A versatile backbone for dense prediction without convolutions. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 548–558.
  - 18 Jha D, Smedsrud PH, Riegler MA, *et al.* Kvasir-SEG: A segmented polyp dataset. Proceedings of the 26th International Conference on MultiMedia Modeling. Daejeon: Springer, 2020. 451–462.
  - 19 Bernal J, Sánchez FJ, Fernández-Esparrach G, *et al.* WMDOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized Medical Imaging and Graphics, 2015, 43: 99–111. [doi: [10.1016/j.compmedimag.2015.02.007](https://doi.org/10.1016/j.compmedimag.2015.02.007)]
  - 20 Tajbakhsh N, Gurudu SR, Liang JM. Automated polyp detection in colonoscopy videos using shape and context information. IEEE Transactions on Medical Imaging, 2016, 35(2): 630–644. [doi: [10.1109/TMI.2015.2487997](https://doi.org/10.1109/TMI.2015.2487997)]
  - 21 Silva J, Histace A, Romain O, *et al.* Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. International Journal of Computer Assisted Radiology and Surgery, 2014, 9(2): 283–293. [doi: [10.1007/s11548-013-0926-3](https://doi.org/10.1007/s11548-013-0926-3)]
  - 22 Wang WH, Xie EZ, Li X, *et al.* PVT v2: Improved baselines with pyramid vision Transformer. Computational Visual Media, 2022, 8(3): 415–424. [doi: [10.1007/s41095-022-0274-8](https://doi.org/10.1007/s41095-022-0274-8)]
  - 23 Bentley PM, McDonnell JTE. Wavelet transforms: An introduction. Electronics & Communication Engineering Journal, 1994, 6(4): 175–186. [doi: [10.1049/ecej:19940401](https://doi.org/10.1049/ecej:19940401)]
  - 24 于晓, 林世基, 庄光耀, 等. 基于多梯度融合的污水域污染物边缘提取算法研究. 黑龙江工业学院学报(综合版), 2023, 23(6): 78–86. [doi: [10.16792/j.cnki.1672-6758.2023.06.010](https://doi.org/10.16792/j.cnki.1672-6758.2023.06.010)]
  - 25 Chen SH, Tan XL, Wang B, *et al.* Reverse attention for salient object detection. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 236–252.
  - 26 Sandler M, Howard A, Zhu ML, *et al.* MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4510–4520.
  - 27 Wang CC, He W, Nie Y, *et al.* Gold-YOLO: Efficient object detector via gather-and-distribute mechanism. Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2024. 2224.
  - 28 Wang HY, Kembhavi A, Farhadi A, *et al.* ELASTIC: Improving CNNs with dynamic scaling policies. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 2253–2262.
  - 29 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13708–13717.
  - 30 Guo MH, Lu CZ, Hou QB, *et al.* SegNeXt: Rethinking convolutional attention design for semantic segmentation. Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022. 84.
  - 31 Chen J, Yang LB, Wang H, *et al.* Road extraction from high-resolution remote sensing images via local and global context reasoning. Remote Sensing, 2023, 15(17): 4177. [doi: [10.3390/rs15174177](https://doi.org/10.3390/rs15174177)]
  - 32 徐康业, 陈建平, 陈平华. 基于局部-全局特征交互的双分支结肠息肉分割网络. 计算机系统应用, 2024, 33(4): 133–142. [doi: [10.15888/j.cnki.csa.009465](https://doi.org/10.15888/j.cnki.csa.009465)]
  - 33 Zhao XQ, Jia HP, Pang YW, *et al.* M<sup>2</sup>SNet: Multi-scale in multi-scale subtraction network for medical image segmentation. arXiv:2303.10894, 2023.
  - 34 Fang YQ, Chen C, Yuan YX, *et al.* Selective feature aggregation network with area-boundary constraints for polyp segmentation. Proceedings of the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention. Shenzhen: Springer, 2019. 302–310.
  - 35 Patel K, Bur AM, Wang GH. Enhanced U-Net: A feature enhancement network for polyp segmentation. Proceedings of the 18th Conference on Robots and Vision (CRV). Burnaby: IEEE, 2021. 181–188.
  - 36 Kim T, Lee H, Kim D. UACANet: Uncertainty augmented context attention for polyp segmentation. Proceedings of the 29th ACM International Conference on Multimedia. ACM, 2021. 2167–2175.
  - 37 Nguyen M, Bui TT, Van Nguyen Q, *et al.* LAPFormer: A light and accurate polyp segmentation Transformer. arXiv:2210.04393, 2022.
  - 38 Liu JQ, Zhang WW, Liu Y, *et al.* Polyp segmentation based on implicit edge-guided cross-layer fusion networks. Scientific Reports, 2024, 14(1): 11678. [doi: [10.1038/s41598-024-62331-5](https://doi.org/10.1038/s41598-024-62331-5)]

(校对责编: 王欣欣)