计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2024,33(4):133−142 [doi: 10.15888/j.cnki.csa.009465]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于局部-全局特征交互的双分支结肠息肉分割网络①

徐康业[1], 陈建平[2], 陈平华[1]

[1](广东工业大学 计算机学院, 广州 510006)
[2](肇庆市教育局, 肇庆 526020)
通信作者: 陈平华, E-mail: phchen@gdut.edu.cn

摘　要: 大小、形状、颜色、纹理的多变性以及肠壁分界模糊给结肠息肉的分割带来巨大挑战. 针对单分支网络连续采样操作造成部分细节信息丢失以及不同层次特征信息无法交互进而导致分割效果不佳的问题, 提出一种基于局部-全局特征交互的双分支结肠息肉分割网络. 网络采用 CNN 与 Transformer 双分支结构, 逐层捕获息肉局部细节特征与全局语义特征; 为充分利用不同层级、不同尺度特征信息的互补性, 利用深层语义特征对浅层细节特征的引导与增强, 设计特征协同交互模块, 动态感知并聚合跨层次特征交互信息; 为强化病变区域特征, 抑制背景噪声, 设计特征增强模块, 应用空间与通道注意力机制强化息肉病变区域特征, 同时采用结合注意力门的跳跃连接机制进一步突出边界信息, 提高边缘区域的分割精度. 实验表明, 所提出网络在多个息肉分割数据集上取得的 *mDice* 与 *mIoU* 分数均优于基线网络, 具有更高的分割准确率和稳定性.

关键词: 结肠息肉分割; 卷积神经网络; Transformer; 双分支结构; 协同交互; 多尺度特征

引用格式: 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. http://www.c-s-a.org.cn/1003-3254/9465.html

Two-branch Colon Polyp Segmentation Network Based on Local-global Feature Interaction

XU Kang-Ye[1], CHEN Jian-Ping[2], CHEN Ping-Hua[1]

[1](School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China)
[2](Zhaoqing Municipal Education Bureau, Zhaoqing 526020, China)

Abstract: The variability in size, shape, color, and texture, along with the blurring demarcation of the bowel wall, presents a significant challenge in colon polyp segmentation. The detail information loss and lack of interaction between different feature levels due to continuous sampling in single-branch networks lead to poor segmentation results. To address this problem, this study proposes a two-branch colon polyp segmentation network based on local-global feature interaction. The network utilizes a dual branch structure consisting of CNN and Transformer, systematically capturing the precise local details and the global semantic features of the polyp in each layer. To make full use of the complementary nature of feature information at different levels and scales, and to utilize the guidance and enhancement of shallow detailed features by deep semantic features, the paper designs the feature cooperative interaction module to dynamically sense and aggregate cross-level feature interaction information. To enhance the feature of the polyp lesion region while reducing background noise, the feature enhancement module utilizes spatial and channel attention mechanisms. Additionally, the skip-connection mechanism in conjunction with the attention gate further highlights boundary

① 基金项目: 广东省重点领域研发计划 (2019A050510041, 2020B0101100001)
收稿时间: 2023-10-16; 修改时间: 2023-11-15; 采用时间: 2023-12-05; csa 在线出版时间: 2024-03-04
CNKI 网络首发时间: 2024-03-08

information, resulting in improved edge region segmentation accuracy. Experiments show that the proposed network achieves better *mDice* and *mIoU* scores than the baseline network on multiple polyp segmentation datasets, with higher segmentation accuracy and stability.

Key words: colon polyp segmentation; convolutional neural network (CNN); Transformer; two-branch structure; cooperative interaction; multi-scale feature

结直肠癌, 是一种常见、致命的消化道恶性肿瘤, 其发病率和死亡率逐年上升且呈年轻化趋势. 结肠息肉易发生恶性病变, 诱发结直肠癌, 结肠镜检查能够有效筛查结肠息肉, 确定息肉位置以及外观信息, 帮助医生诊断, 实现早期治疗[1]. 然而, 临床实践中因息肉大小、形状、颜色、纹理各异, 只有经验丰富的医生才能准确诊断, 但受精力限制和情绪波动等影响, 诊断往往存在漏检或误检的情况. 为辅助诊断, 降低误诊概率, 研究一种能够准确分割结肠息肉的方法, 具有重要的临床意义与积极的应用前景.

结肠镜图像易受如光照、气泡、拍摄设备等因素影响, 出现模糊、伪影; 同时, 息肉外观各异, 边缘模糊, 分割难度大. 基于深度学习的医学图像分割方法发展至今, 涌现出各种模型结构针对各种医学场景与存在的问题提出了各种解决方法. Long 等[2]提出的全卷积神经网络 (FCN) 最早将卷积神经网络应用于图像分割任务, 实现像素级别的分类, 由此揭开了语义级别医学图像分割研究的序幕. Ronneberger 等[3]在 2015 年提出基于编码器-解码器结构的 U-Net 网络, 引入跳跃连接实现特征传递, 融合浅层与深层特征, 补充采样操作所损失的细节特征. 近年来, 研究人员围绕 U-Net 开展大量研究, 对网络结构进行诸多改进与扩展. 针对息肉分割场景, Fan 等[4]提出的 PraNet 通过反向注意力模块建立分割区域与边界的联系, 增强分割性能; 为消除息肉颜色的影响, Zhong 等[5]提出的 SANet 将图像内容与颜色解耦, 使模型更加关注息肉外观, 增强分割性能; 针对息肉定位不准确和边缘模糊的问题, Zhao 等[6]提出的 M^{2}SNet 通过具有不同感受野的金字塔减法单元捕捉多尺度息肉特征, 具有良好的分割精度和泛化能力.

2017 年, Google 提出应用于自然语言处理领域的 Transformer[7], 研究人员开始将 Transformer 引入视觉任务中. ViT (vision Transformer)[8]首次将 Transformer 应用于图像分类任务, 达到超越 ResNet[9]的精度, 具有更强的全局感知能力和适应性. 在医学图像分割领域, Chen 等[10]提出的 TransUNet 首次将 Transformer 应用于 U 型网络结构, 其在编码器深层阶段引入 Transformer 建立长距离特征关系, 提高分割性能. 针对息肉分割场景, Zhang 等[11]提出 TransFuse, 通过并行结构结合应用多种注意力机制的 BiFusion 模块捕获多层次特征, 在多个数据集上验证了方法有效性; Wang 等[12]提出由金字塔 Transformer 和多级特征聚合结构的渐进局部解码器构成的 SSFormer, 提高模型对局部特征的处理能力.

结肠息肉分割方法发展至今取得显著进步, 但仍存在分割边缘模糊、分割效果有待进一步提高的问题. 现有的分割方法采取单分支网络结构, 其连续的下采样、上采样操作会造成部分细节信息丢失, 同时, 不同层次的特征信息无法或不便进行交互, 带来了分割边界模糊、分割效果不佳的问题. 针对这些问题, 提出基于局部-全局特征交互的双分支结肠息肉分割网络, 主要包括以下工作.

(1) 针对单分支网络结构连续采样操作造成部分细节信息丢失和视野受限问题, 提出采用基于 CNN 的 CSPDarknet53[13]网络与基于 Transformer 的 PvTv2[14]网络的双分支网络结构, 逐层提取的息肉局部细节特征与全局语义特征, 扩大同层次特征视野, 提高对息肉边缘的判断与定位能力; 同时, 采用结合注意力门[15]的跳跃连接机制进一步突出边界信息, 增强边缘区域分割性能.

(2) 为挖掘不同层次特征的依赖关系, 设计特征协同交互模块, 动态感知并聚合跨层次的特征交互信息, 实现不同层次特征信息的交互, 应对息肉病变区域外观各异的挑战; 同时, 为强化特征, 抑制背景噪声, 设计特征增强模块, 应用注意力机制强化息肉区域特征信息, 提高对息肉的精确识别与定位.

(3) 相较于 U-Net、PraNet、TransFuse、M^{2}SNet 等现有模型, 本文所提出的模型在 Kvasir[16]、ClinicDB[17]、ColonDB[18]以及 ETIS[19]这 4 个结肠息肉分割数据集上

均取得更高的*mDice*与*mIoU*分数, 拥有更好的分割准确率和稳定性.

1 算法模型

大小、形状、颜色、纹理的多变性以及肠壁分界模糊给结肠息肉分割带来巨大挑战. 针对单分支网络连续采样操作导致部分细节信息丢失以及不同层次特征信息无法交互进而导致分割效果不佳的问题, 提出一种基于局部-全局特征交互的双分支结肠息肉分割网络, 模型总体结构如图 1 所示, 其编码阶段由局部特征提取模块 (local feature extraction module, LFEM)、全局特征提取模块 (global feature extraction module, GFEM), 特征协同交互模块 (feature cooperative interaction module, FCIM) 以及特征增强模块 (feature enhancement module, FEM) 组成, 解码阶段由上采样模块 (upsample module, UM) 组成.

编码阶段, 模型通过由 CSPDarknet53 构成的局部特征提取模块与 PvTv2 构成的全局特征提取模块并行提取多尺度的息肉局部细节特征与全局语义特征, 扩大同层次特征视野的同时减少因连续采样操作造成部分细节信息的丢失, 提高对息肉边缘的判断与定位能力; 随后, 为挖掘不同层次特征间的依赖关系, 设计特征协同交互模块, 应用哈达玛积、卷积与通道注意力, 结合深监督机制[20]动态感知并聚合跨层次的特征交互信息, 实现不同层次特征信息的交互, 应对息肉外观各异的挑战; 之后, 通过应用空间与通道注意力的特征增强模块强化病变区域特征, 抑制背景噪声, 提高对息肉的精确识别与定位; 同时, 应用结合注意力门 (attention gate, AG) 的跳跃连接机制进一步突出边界信息, 提高边缘区域的分割精度.

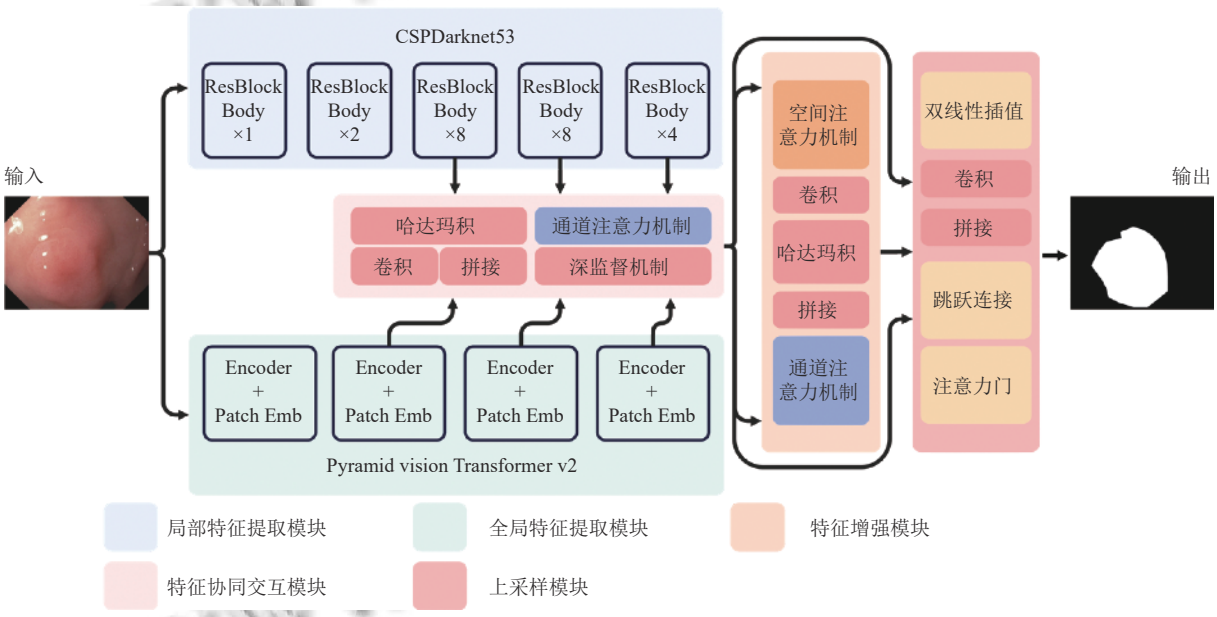

图 1 网络结构图

1.1 局部特征提取模块

CSPDarknet53 构成的局部特征提取模块通过分层次特征重塑图像结构, 提取不同尺度下的局部细节特征. CSPDarknet53 由应用Mish激活函数[21]的卷积层以及多个应用跨阶段局部 (cross stage partial, CSP) 网络的残差模块体 (ResBlock Body) 组成, 如图 2 所示, 其中×n表示该残差结构体中包含n个残差块 (ResBlock). 具体来说, 卷积层由卷积、批量归一化和Mish激活函数组成, 相比于*Sigmoid*[22]、*ReLU*[23]等函数, Mish具有无下界、有上界、非单调性和平滑性的特点, 能够更好地传播特征信息, 提高模型表达能力.

针对网络优化过程中因梯度重复导致的梯度消失问题, 残差模块体应用跨阶段合并思想, 将输入特征图经1×1卷积降采样后分别通过1×1卷积划分为残差卷积与跳跃连接两部分; 残差卷积部分的特征通过由1×1卷积、3×3卷积以及残差块 (ResBlock) 进行特征提取, 随后在调整通道数后与跳跃连接部分的特征拼接并通过1×1卷积得到输出特征; 残差模块体通过跨

阶段合并的结构分割梯度信息, 缓解梯度重复, 并结合残差连接实现同一层次特征融合, 进而减少因连续采样操作造成的部分细节信息丢失. 借助金字塔结构, CSPDarknet53 通过堆叠多个残差模块体以捕获多尺度局部细节特征, 并输出 3 种不同尺度的特征信息用于后续处理.

1.2 全局特征提取模块

全局特征提取模块 (如图 3) 采用 PvTv2 (pyramid vision Transformer v2) 进行全局关系建模, 逐层提取全局语义特征, 扩大同层次特征视野. PvTv2 包含 4 个阶段, 每个阶段均由嵌入层 (Patch Emb) 以及两个 Transformer 编码器 (Transformer encoder) 构成, 其在每一阶段应用渐进缩减策略缩小特征图分辨率以获得不同尺度的特征图. 具体来说, 在每一阶段, PvTv2 首先将输入特征图分解为多个4×4大小的图像块 (patch), 通过嵌入层将图像块拉直并通过线性映射降维; 随后, 将其与位置嵌入相加并通过 Transformer 编码器提取特征, 之后将其转换为三维形态送入下一阶段.

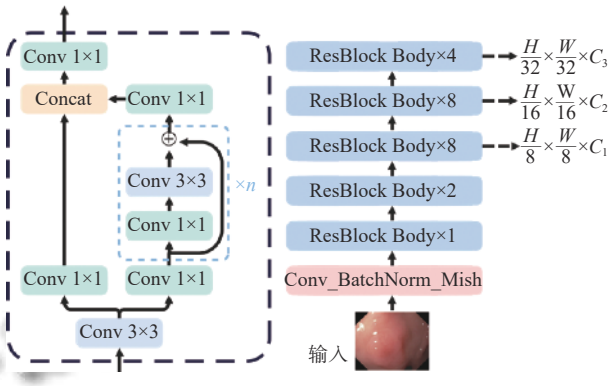

图 2 局部特征提取模块 CSPDarknet53

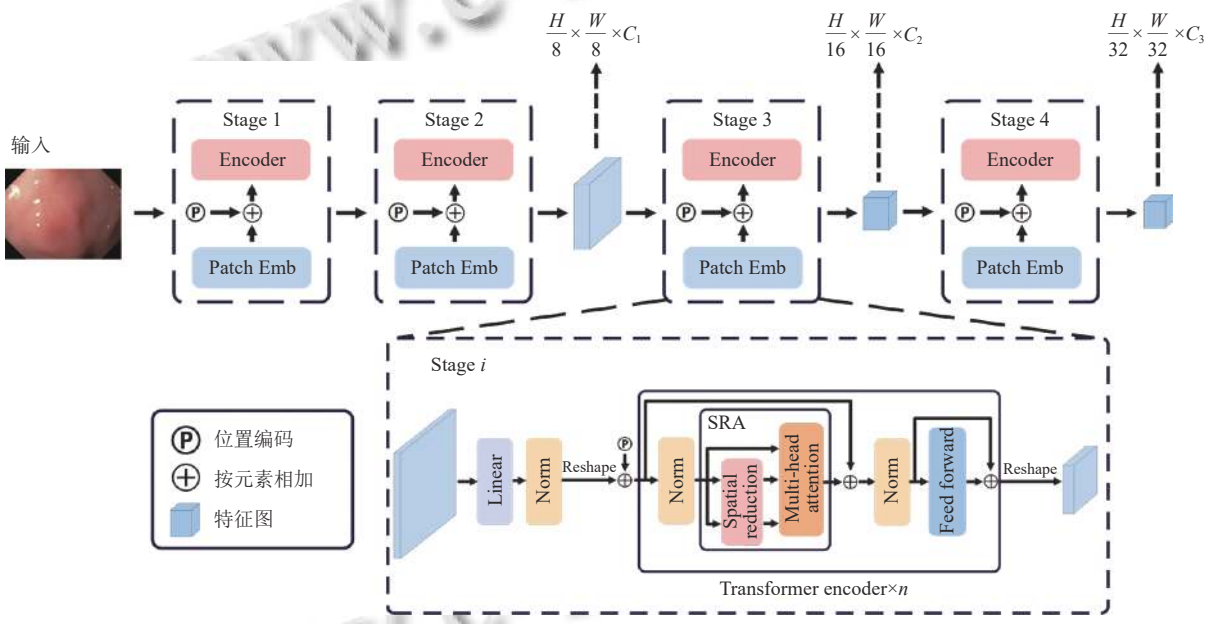

图 3 全局特征提取模块 PvTv2

Transformer 编码器由自注意力 (self attention) 层与结合深度可分离卷积的前馈网络 (feed forward network) 层组成, 深度可分离卷积包含深度卷积与逐点卷积, 分别对通道与空间上进行卷积操作, 能够避免特征混淆, 提高模型的计算效率; 同时, 自注意力层中的残差连接实现了同层次特征融合, 保留关键信息的同时减少因采用渐进缩减策略造成的部分信息丢失. 另外, 针对注意力机制存在的计算开销问题, PvTv2 在 Transformer 编码器中提出空间缩减注意力机制替代原有的多头注意力 (multi-head self attention) 机制, 以 *Query*、*Key*、*Value* 作为输入, 在计算前进行特征筛选, 通过平均池化缩小 *Key* 和 *Value* 的空间尺度, 提高计算效率. 针对息肉分割任务, PvTv2 输出 3 个尺度的全局特征用于后续处理, 缓解局部特征提取模块因建模远程关系不足导致的视野受限问题, 提高对息肉边缘的判断与定位能力.

1.3 特征协同交互模块

为充分利用不同层次、不同尺度特征的互补性, 利用深层语义特征对浅层细节特征的引导与增强, 如图 4 所示, 特征协同交互模块动态感知并聚合跨层次的特征交互信息, 实现不同层次特征信息的交互, 以应对息肉外观变化的挑战. 模块可分为 3 个阶段, 每阶段

中来自局部与全局特征提取模块的特征具有相同的空间与通道维度. 具体来说, 利用哈达玛积与卷积对同层次的局部与全局特征进行交互, 利用特征间的互补性实现有效整合, 得到交互后的特征T_1、T_2、T_3; 沿特征传播方向, 将T_1下采样至与T_2在空间、通道维度上一致后进行交互, 得到特征T_4; 随后, 将T_4下采样值与T_3在空间与通道上一致后进行交互, 进而聚合不同层次的特征交互信息, 挖掘不同层次特征的依赖关系, 利用深层语义特征对浅层细节特征进行引导, 并应用SENet[24]进行增强, 得到交互特征T_5; 最后, 将T_5分别与局部、全局特征提取模块输出的特征L_i与G_i进行拼接, 并通过卷积调整维度, 得到局部特征L与全局特征G; 通过逐层交互与传递不同层次、不同尺度的特征信息, 充分利用特征间的互补性, 增强网络对息肉区域的感知能力.

为促进特征的充分交互, 引入深监督机制, 结合由加权损失和加权损失构成的联合损失函数, 对局部特征L与全局特征G进行监督训练. 具体而言, 深监督机制将局部特征L与全局特征G进行损失计算, 使特征更早接受反馈信号并进行调整, 加速模型收敛; 同时, 动态调整不同层次特征的权重分配, 应对息肉病变区域外观各异的挑战, 为后续模块提供精确的上下文特征信息, 增强模型的鲁棒性与泛化能力.

1.4 特征增强模块

如图 4 所示, 特征增强模块应用空间与通道注意力强化息肉区域病变区域特征, 抑制背景噪声, 提高对息肉的精确识别与定位. 具体来说, 应用 SAM (spatial attention module)[25]对局部特征L进行空间注意力增强, 提取关键区域特征信息, 其计算推导式如式 (1) 和式 (2) 所示:

$$L_c = Concat(AvgPool(L), MaxPool(L)) \tag{1}$$

$$L_{\text{attn}} = L \times Sigmoid(Conv(L_c)) \tag{2}$$

如图 5 所示, SAM 将局部特征L进行最大池化$MaxPool$与平均池化$AvgPool$得到两个$1\times H\times W$的特征图, 拼接后通过卷积转变为单通道特征图, 再通过$Sigmoid$函数激活得到空间注意力特征图, 最后将其与原特征L相乘得到增强的局部特征L_{attn}.

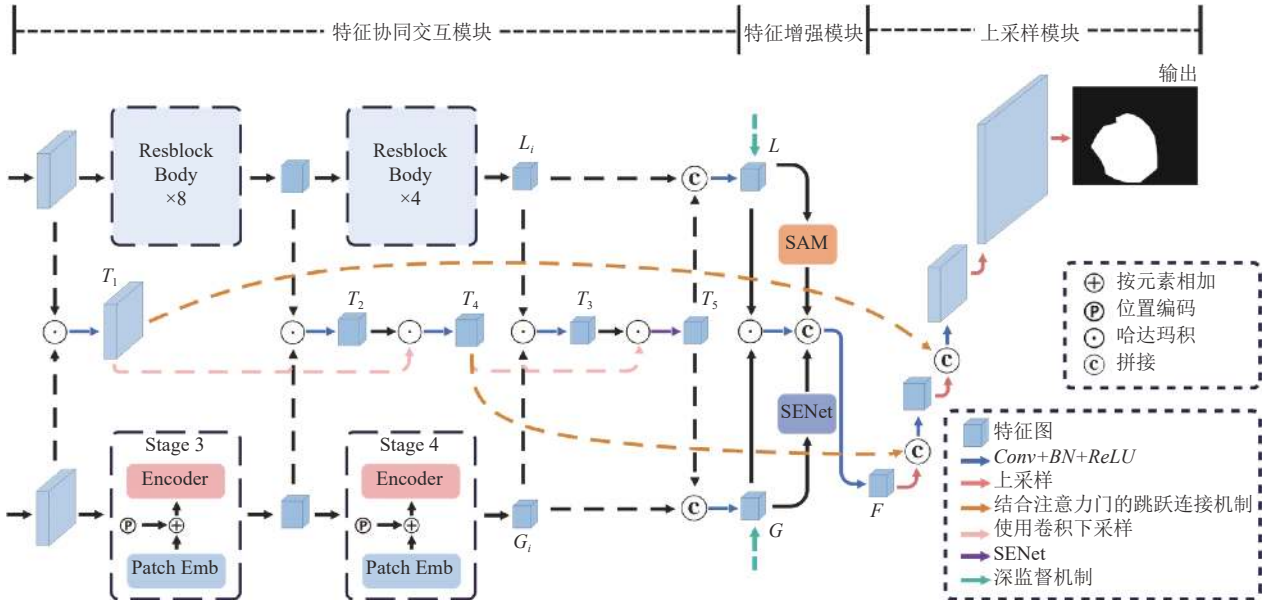

图 4 特征协同交互模块、特征增强模块、上采样模块

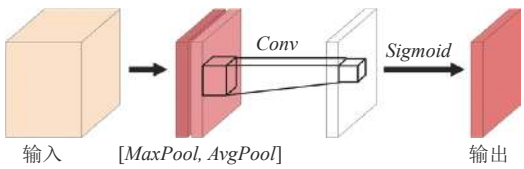

图 5 空间注意力机制 SAM

针对全局特征G, SENet (squeeze and excitation network) 使用通道注意力进行增强, 自适应强化重点通道特征信息, 其计算推导式如式 (3) 和式 (4) 所示:

$$G_a = GlobalAvgPool(G) \tag{3}$$

$$G_{attn} = G \times Sigmoid(Dense(G_a)) \tag{4}$$

如图 6 所示 (*表示逐通道乘法), SENet 将全局特征G进行全局平均池化$GlobalAvgPool$转变为$1\times1\times C$的特征图, 通过全连接层计算通道重要性, 由$Sigmoid$函数激活得到通道注意力特征图, 最后将其与原特征图G相乘得到增强的全局特征G_{attn}.

$$Z = Conv(HadamardProduct(L,G)) \tag{5}$$

$$F = Conv(Concat(L_{\text{attn}},G_{\text{attn}},Z)) \tag{6}$$

最后, 如式 (5) 与式 (6), 应用哈达玛积 $Hadamard\text{-}Product$ 与卷积将局部特征L与全局特征G进行交互, 得到特征Z, 并与L_{attn}和G_{attn}拼接, 通过卷积降维得到具有清晰轮廓与准确位置的息肉特征图F.

图 6 通道注意力机制 SENet

1.5 上采样模块

上采样模块应用结合注意力门的跳跃连接机制进一步突出边界信息, 提高边缘区域的分割精度. 如式 (7) 所示, 上采样模块将F进行上采样Up, 并与来自特征协同交互模块的特征T_1进行注意力门AG计算得到W_1, 后如式 (8) 所示, 与上采样后的F进行拼接并通过卷积得到新特征F, 将其与特征T_4重复上述步骤得到最终特征F; 最后, 通过两次上采样恢复至原图大小, 完成分割.

$$W_i = AG(Up(F),T_i),\ i=1,4 \tag{7}$$

$$F = Conv(Concat(Up(F),W_i)),\ i=1,4 \tag{8}$$

具体来说, 本模块采用双线性插值方法, 相比于反卷积、最近邻插值等方法, 其计算复杂度低, 能够保持分割图像的平滑性和连续性, 有助于恢复边缘细节. 注意力门的算法流程图如图 7 所示, 其将来自先前模块的特征T_1、T_4与上采样后的F作为输入; 其先将T_1与F分别卷积, 随后进行相加并激活, 再通过卷积与激活函数得到注意力特征图F_{attn}, 与F相乘得到新特征F, 从而利用深层语义特征信息筛选对分割任务有利的边缘细节特征, 使边缘区域得到聚焦; 之后, 应用相同流程对新特征F与T_4进行计算, 得到F.

1.6 损失函数

对于边缘信息要求高的息肉分割任务, 选择合适的损失函数是提高分割精度的关键. 传统IoU和BCE损失函数易忽略类别间的权重差异. 针对这一问题, 如式 (9) 所示, 本文采用加权IoU损失和加权BCE损失构建的联合损失函数结合深监督机制进行训练, 提高网络对边缘像素的识别能力, 细化边缘特征, 提升分割精度的同时缓解数据不均衡的问题.

$$Loss = Loss_{IoU} + Loss_{BCE} \tag{9}$$

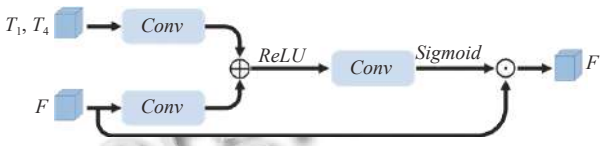

图 7 注意力门 AG

具体来说, 本文将加权IoU与加权BCE损失函数进行加和; 加权IoU损失函数是一种衡量分割模型准确性的指标, 如式 (10) 所示, 其衡量模型输出结果与真实标签间的重叠度, 为不同类别分配不同权重系数, 提高模型在少样本上的准确性.

$$Loss_{IoU} = 1 - \frac{\sum_{i=1}^{N} w_i \times y_i \times \hat{y}_i}{\sum_{i=1}^{N} w_i \times (y_i + \hat{y}_i - y_i \times \hat{y}_i)} \tag{10}$$

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.

加权损失BCE函数是一种分类损失函数, 如式 (11) 所示, 其为不同类别分配不同权重系数, 提高模型对重点类别的关注, 提升分割性能.

$$Loss_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} w_i \times [y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \tag{11}$$

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.

2 实验

本文配置的实验环境如下: CPU Intel i9 11900K/F, 内存 64 GB, GPU NVIDIA GeForce RTX 3090, 显存 24 GB, Python 3.8, Linux Ubuntu 操作系统, PyTorch 深度学习框架.

2.1 实验数据集

实验采用胃肠道疾病研究的医学图像公开数据集 Kvasir、ClinicDB、ColonDB 以及 ETIS. Kvasir 数据集共包含 1 000 张胃肠道息肉图片及其对应的分割标

签图, 涵盖息肉、溃疡、出血等症状且存在检查中易被忽略的直径小于 10 mm 的息肉; ClinicDB 与 ColonDB 数据集分别包含 612 张与 380 张图片及其对应的分割标签图; ETIS 数据集包含来自不同地区医疗中心的不同内镜设备中提取出来 196 张结肠息肉图片及其对应的分割标签图. 具体来说, 训练数据集由来自 Kvasir 数据集的 900 张图像及 ClinicDB 数据集的 550 张图像组成, 共包含 1 450 个样本及其对应的分割标签图; 测试数据集共由来自 Kvasir 数据集的 100 张图像, ClinicDB 数据集的 62 张图像, ColonDB 数据集的 380 张图像以及 ETIS 数据集的 196 张图像组成, 共包含 738 个样本及其对应的分割标签图; 消融实验测试数据集由来自 ETIS 数据集的 196 张图像及其对应的分割标签图组成.

如图 8 所示, 实验所使用的数据集在息肉大小、形状、纹理上存在多变性, 且肠壁分界模糊, 同时成像质量存在不稳定性, 受到如光照、气泡、拍摄条件等因素影响, 实际分割存在巨大挑战.

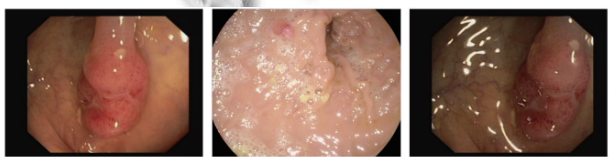

图 8 数据集图片

2.2 实验参数与评价指标

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[26]优化器, 共进行 100 轮次训练; 数据增强方法采用归一化与 0.75 倍、1 倍、1.25 倍的多尺度输入方法.

实验选取*mDice*与*mIoU*分数作为分割图像网络的评价指标. *mDice*, 即平均*Dice*分数, 是对所有测试样本的*Dice*分数取平均值. *Dice*函数用于计算两个样本的相似度, 其取值范围在[0,1]. *Dice*分数越大, 样本越相似. *Dice*函数如式 (12) 所示:

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \tag{12}$$

其中, A表示网络对结肠息肉图像分割的结果图像, B表示结肠息肉的标签图像. 假设测试样本的总数为n, *mDice*计算公式如式 (13) 所示:

$$mDice = \frac{1}{n}(Dice_1 + Dice_2 + \cdots + Dice_n) \tag{13}$$

*IoU*函数计算两个集合交集和并集的比值, 取值范围在[0,1]. *IoU*分数越大, 样本越相似. *IoU*函数如式 (14) 所示:

$$IoU = \frac{A \cap B}{A \cup B} \tag{14}$$

其中, A表示网络对结肠息肉图像分割的结果图像, B表示结肠息肉的标签图像. 假设测试样本的总数为n, *mIoU*即平均交并比, 是对所有测试样本的*IoU*取平均值. *mIoU*计算公式如式 (15) 所示:

$$mIoU = \frac{1}{n}(IoU_1 + IoU_2 + \cdots + IoU_n) \tag{15}$$

2.3 实验结果与分析

为验证模型的有效性与鲁棒性, 在使用相同测试数据集的基础上, 选取一系列模型开展对比实验.

如表 1 所示, 相较于 U-Net, PraNet 等 CNN 网络模型, 结合 CNN 与 Transformer 的 TransFuse, 采用不确定性增强的上下文注意力机制的 UACANet 以及多尺度减法网络 M^{2}SNet 等, 本文提出的模型在 4 个测试数据集上均取得更高*mDice*的与*mIoU*分数; 同时, 结合图 9 中部分模型的分割结果图进行比较, 可见, 本文所提出模型的分割区域完整, 边缘清晰, 拥有更好的分割准确率与稳定性.

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

Methods	Year	Kvasir		ClinicDB		ColonDB		ETIS	
		mDice	*mIoU*	*mDice*	*mIoU*	*mDice*	*mIoU*	*mDice*	*mIoU*
U-Net	2015	0.818	0.746	0.823	0.750	0.512	0.444	0.398	0.335
UNet++[27]	2018	0.821	0.743	0.794	0.729	0.483	0.410	0.401	0.344
SFA[28]	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[29]	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[30]	2021	0.912	0.859	0.926	0.880	—	—	0.766	0.689
LAPFormer-S[31]	2022	0.910	0.857	0.901	0.849	0.781	0.695	0.768	0.686
M^{2}SNet	2023	0.912	0.861	0.922	0.880	0.758	0.685	0.749	0.678
Ours	**2023**	**0.923**	**0.873**	**0.933**	**0.886**	**0.798**	**0.714**	**0.793**	**0.708**

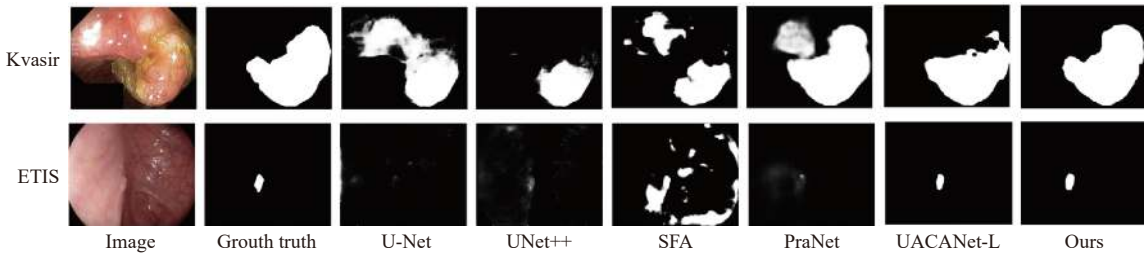

图 9 部分模型在 Kvasir、ETIS 数据集上的分割结果

2.4 消融实验结果与分析

为验证本文所提出的模块、方法对模型性能提升的有效性和必要性, 基于相同的实验环境、实验参数以及训练数据集, 开展相应的消融实验. 实验中, 本文将表 1 中所提出网络的分割指标作为比较基准, 分别进行 6 项消融实验, 使用一致的评价指标*mDice*与*mIoU*在 ETIS 数据集上进行测试, 结果如表 2 所示.

表 2 网络在 ETIS 数据集上的消融实验结果

Model	*mDice*	*mIoU*
Ours	**0.793**	**0.708**
w/o DS	0.780	0.699
w/o AG	0.768	0.680
w/o FEM	0.749	0.678
w/o AG, FCIM	0.754	0.669
w/o FEM, AG	0.741	0.655
w/o FEM, AG, FCIM	0.736	0.633

第 1 项消融实验单纯去除特征协同交互模块中的深监督机制, 网络在*mDice*与*mIoU*指标上有些许下降, 表明在模型中应用深监督机制能够促进特征间充分交互, 帮助模型动态调整不同层次特征的权重分配, 应对息肉病变区域外观各异的挑战, 为后续阶段提供更精确的特征信息, 提升模型鲁棒性.

息肉边缘的准确分割对于分割性能的提升十分重要. 第 2 项消融实验去除结合注意门的跳跃连接机制, 网络在*mDice*与*mIoU*指标上均有约 2% 的下降, 表明在解码器中应用结合注意力门的跳跃连接机制, 利用深层特征信息对分割任务有利的边缘细节特征, 使边缘区域特征得到聚焦, 提高分割精度.

第 3 项消融实验去除特征增强模块, 网络在*mDice*上出现约 4% 的明显下降; 可见, 特征增强模块发挥关键作用, 通过空间与通道注意力强化特征信息, 提供更全面、更准确的息肉病变区域特征, 获得具有清晰轮廓与准确位置的特征图, 提高网络对息肉的精确识别与定位.

第 4 项消融实验去除结合注意门的跳跃连接机制以及特征协同交互模块, 网络在*mDice*与*mIoU*指标上出现约 4% 的明显下降; 可见, 特征协同交互模块扮演关键角色, 其充分利用特征间的互补性, 以及深层语义特征对浅层细节特征的引导与增强, 动态感知并聚合跨层次的特征交互信息, 实现不同层次间的特征交互; 同时, 为上采样模块提供丰富的边缘细节信息, 提高分割精度.

第 5 项消融实验去除特征增强模块与结合注意力门的跳跃连接机制, 第 6 项消融实验在第 5 项实验的基础上去除特征协同交互模块. 相比前 4 项实验, 网络在与指标上均出现了更大程度的下降, 由此可见, 以上模块的有机结合十分重要; 具体而言, 特征协同交互模块动态感知并聚合跨层次的特征交互信息, 为后续特征增强模块提供更为丰富、适应息肉外观变化的特征信息, 并应用注意力机制进一步强化病变区域特征, 抑制背景噪声; 同时与结合注意力门的跳跃连接机制进一步细化边缘特征, 显著提升网络的分割性能与泛化性能.

3 结论与展望

大小、形状、颜色、纹理的多变性以及肠壁分界模糊给结肠息肉分割带来巨大挑战. 针对单分支网络连续采样操作导致部分细节信息丢失以及不同层次特征信息无法交互进而导致分割效果不佳的问题, 本文提出一种基于局部-全局特征交互的双分支结肠息肉分割网络. 网络在特征提取阶段构建并行网络结构, 使用 CSPDarknet53 网络与 PvTv2 网络逐层捕获息肉局部细节特征与全局语义特征, 扩大同层次特征视野的同时减少因连续采样操作造成的部分细节信息丢失, 提高对息肉边缘的判断与定位能力; 为充分利用不同层级、不同尺度特征信息的互补性, 利用深层语义特征对浅层细节特征的引导与增强, 提出特征协同交互模块动态感知并聚合跨层次交互信息, 实现不同层次

特征信息的交互, 以适应息肉外观变化的挑战; 设计特征增强模块, 应用空间与通道注意力机制强化息肉区域特征, 抑制背景噪声, 提高对息肉的精确识别与定位; 最后, 采用结合注意力门的跳跃连接机制进一步突出边缘信息, 增强边缘区域分割性能. 网络在 ClinicDB、Kvasir、ColonDB、ETIS 这 4 个息肉分割数据集上与主流基线网络进行对比实验以验证有效性; 同时, 在 ETIS 数据集上进行消融实验以验证所提出模块的有效性; 结果表明, 本文所提出网络能够有效提升息肉分割精度, 在临床诊断中具有一定的辅助应用价值, 但仍存在不足, 后续研究中将改进各个模块, 加强对小目标、低对比度场景下的息肉检测能力, 进一步提高分割准确性与稳定性.

参考文献

1 Zauber AG, Winawer SJ, O'Brien MJ, *et al*. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. The New England Journal of Medicine, 2012, 366(8): 687–696. [doi: 10.1056/NEJMoa1100370]

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 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.

4 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.

5 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.

6 Zhao XQ, Jia HP, Pang YW, *et al*. M^2SNet: Multi-scale in multi-scale subtraction network for medical image segmentation. arXiv:2303.10894, 2023.

7 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: ACM, 2017. 6000–6010.

8 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. ICLR, 2021.

9 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.

10 Chen JN, Lu YY, Yu QH, *et al*. TransUNet: Transformers make strong encoders for medical image segmentation. arXiv:2102.04306, 2021.

11 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.

12 Wang JF, Huang QM, Tang FL, *et al*. Stepwise feature fusion: Local guides global. Proceedings of the 25th International Conference on Medical Image Computing and Computer Assisted Intervention. Singapore: Springer, 2022. 110–120.

13 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.

14 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]

15 Schlemper J, Oktay O, Schaap M, *et al*. Attention gated networks: Learning to leverage salient regions in medical images. Medical Image Analysis, 2019, 53: 197–207. [doi: 10.1016/j.media.2019.01.012]

16 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.

17 Bernal J, Sánchez FJ, Fernández-Esparrach G, *et al*. WM-DOVA 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]

18 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]

19 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]

20 Lee CY, Xie SN, Gallagher PW, *et al*. Deeply-supervised nets. Proceedings of the 18th International Conference on Artificial Intelligence and Statistics. San Diego: AISTATS, 2015. 562–570.

21 Misra D. Mish: A Self Regularized non-monotonic activation function. Proceedings of the 31st British Machine Vision Conference. BMVC, 2020.

22 Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature, 1986, 323(6088): 533–536. [doi: 10.1038/323533a0]

23 Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. Fort Lauderdale: AISTATS, 2011. 315–323.

24 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.

25 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19.

26 Loshchilov I, Hutter F. Decoupled weight decay regularization. Proceedings of the 7th International Conference on Learning Representations. New Orleans: ICLR, 2019.

27 Zhou ZW, Rahman Siddiquee MM, Tajbakhsh N, *et al*. UNet++: A nested U-Net architecture for medical image segmentation. Proceedings of the 4th International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Granada: Springer, 2018. 3–11.

28 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.

29 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.

30 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.

31 Nguyen M, Bui TT, Van Nguyen Q, *et al*. LAPFormer: A light and accurate polyp segmentation Transformer. arXiv: 2210.04393, 2022.

(校对责编: 孙君艳)