



融合边界注意力的特征挖掘息肉小目标网络

刘国奇, 陈宗玉, 刘栋, 常宝方, 王佳佳

引用本文:

刘国奇, 陈宗玉, 刘栋, 常宝方, 王佳佳. 融合边界注意力的特征挖掘息肉小目标网络[J]. 智能系统学报, 2024, 19(5): 1092–1101.

LIU Guoqi, CHEN Zongyu, LIU Dong, et al. A small polyp objects network integrating boundary attention features[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(5): 1092–1101.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202306025>

您可能感兴趣的其他文章

基于反馈注意力机制和上下文融合的非模式实例分割

Feedback attention mechanism and context fusion based amodal instance segmentation
智能系统学报. 2021, 16(4): 801–810 <https://dx.doi.org/10.11992/tis.202007042>

双向特征融合与注意力机制结合的目标检测

Target detection based on bidirectional feature fusion and an attention mechanism
智能系统学报. 2021, 16(6): 1098–1105 <https://dx.doi.org/10.11992/tis.202012029>

基于注意力融合的图像描述生成方法

An image caption generation method based on attention fusion
智能系统学报. 2020, 15(4): 740–749 <https://dx.doi.org/10.11992/tis.201910039>

基于注意力机制的显著性目标检测方法

Salient object detection method based on the attention mechanism
智能系统学报. 2020, 15(5): 956–963 <https://dx.doi.org/10.11992/tis.201903001>

层次化双注意力神经网络模型的情感分析研究

Hierarchical double-attention neural networks for sentiment classification
智能系统学报. 2020, 15(3): 460–467 <https://dx.doi.org/10.11992/tis.201812017>

基于双向消息链路卷积网络的显著性物体检测

Salient object detection based on bidirectional message link convolution neural network
智能系统学报. 2019, 14(6): 1152–1162 <https://dx.doi.org/10.11992/tis.201812003>

DOI: 10.11992/tis.202306025

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20240828.1447.032>

融合边界注意力的特征挖掘息肉小目标网络

刘国奇^{1,2}, 陈宗玉^{1,2}, 刘栋^{1,2}, 常宝方¹, 王佳佳^{1,2}

(1. 河南师范大学 计算机与信息工程学院, 河南 新乡 453007; 2. 河南师范大学 河南省教育人工智能与个性化学习重点实验室, 河南 新乡 453007)

摘要: 从结肠图像中分割息肉小目标病变区域对于预防结直肠癌至关重要, 它可以为结直肠癌的诊断提供有价值的信息。然而目前现有的方法存在 2 个局限性: 一是不能稳健捕获全局上下文信息, 二是未能充分挖掘细粒度细节特征信息。因此, 提出融合边界注意力的特征挖掘息肉小目标网络 (transformer feature boundary network, TFB-Net)。该网络主要包括 3 个核心模块: 首先, 采用 Transformer 辅助编码器建立长程依赖关系, 补充全局信息; 其次, 设计特征挖掘模块进一步细化特征, 学习到更好的特征; 最后, 使用边界反转注意力模块加强对边界语义空间的关注, 提高区域辨别能力。在 5 个息肉小目标数据集上进行广泛实验, 实验结果表明 TFB-Net 具有优越的分割性能。

关键词: 息肉小目标分割; Transformer; 卷积神经网络; 特征挖掘; 注意力机制; 边界注意力; 语义信息; 全局特征
中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2024)05-1092-10

中文引用格式: 刘国奇, 陈宗玉, 刘栋, 等. 融合边界注意力的特征挖掘息肉小目标网络 [J]. 智能系统学报, 2024, 19(5): 1092-1101.

英文引用格式: LIU Guoqi, CHEN Zongyu, LIU Dong, et al. A small polyp objects network integrating boundary attention features[J]. CAAI transactions on intelligent systems, 2024, 19(5): 1092-1101.

A small polyp objects network integrating boundary attention features

LIU Guoqi^{1,2}, CHEN Zongyu^{1,2}, LIU Dong^{1,2}, CHANG Baofang¹, WANG Jiajia^{1,2}

(1. College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China; 2. Henan Key Laboratory of Educational Artificial Intelligence and Personalized Learning, Henan Normal University, Xinxiang 453007, China)

Abstract: Segmentation of small target lesion areas, such as polyps in colon images, is essential for the prevention and diagnosis of colorectal cancer. However, existing methods face two main limitations: either the global context information cannot be captured robustly or the fine-grained detail information cannot be fully mined. To address these issues, this study proposes TFB-Net, a feature mining network for small target polyps that integrates boundary attention. The network consists of three core modules: First, a Transformer is used to establish long-term dependencies and supplement global information. Second, the feature mining module is designed to further optimize and enhance the learned features. Finally, the boundary inversion attention module strengthens attention to the boundary semantic space, which consequently improves regional discrimination. Extensive experiments were conducted on five small polyp target datasets, and the results show that TFB-Net achieves superior segmentation performance.

Keywords: small polyp objects segmentation; Transformer; convolutional neural network; feature mining; attention mechanism; boundary attention; semantic information; global feature

结直肠癌作为世界第三大常见癌症, 已经严重威胁人们的健康甚至生命^[1]。研究表明, 结直

肠癌发现于早期, 死亡率仅为 10%^[2]。在结直肠癌病变初期, 息肉通常体积较小, 并且形状各异, 这大大增加了诊断难度。目前, 结肠镜检查是筛查和预防结直肠癌的主要技术。然而, 这种方法不仅耗费人力成本, 而且微小息肉很容易被漏

收稿日期: 2023-06-13. 网络出版日期: 2024-08-29.

基金项目: 国家自然科学基金项目 (61901160, U1904123).

通信作者: 刘国奇. E-mail: gqliu@htu.edu.cn.

©《智能系统学报》编辑部版权所有

检。因此,实现高效、准确的息肉小目标自动分割具有重要的临床意义^[3]。

与传统分割方法相比,基于编码器-解码器的深度学习方法^[4-5]可达到更高的分割精度。U-Net^[6]作为首个编码器-解码器结构分割网络,已成功应用于生物学医学图像分割领域。随后出现的 UNet++^[7]通过集成不同深度 U-Net 和一系列嵌套、密集跳跃连接,缩小收缩路径和扩展路径特征图间语义差距,从而灵活高效聚合不同语义尺度特征。而 UNet3+^[8]应用全尺度跳跃连接和深度监督策略,模型分割性能得到极大提高。随后相继提出 ResUNet(road extraction by deep residual U-net)^[9]以及 ResUNet++^[10]等,通过不断促进特征有效融合,这些网络进一步提高息肉分割精度,在许多医学图像分割任务中表现出良好性能。然而,这些方法捕获丰富全局上下文信息和细节信息的能力不足。

为捕获全局上下文信息,使用扩大感受野策略和 Transformer 模型 2 种优化方案。Dilated-SegNet(dilated segmentation network)^[11]通过将膨胀卷积模块的输出进行连接,分割息肉目标。MK-DCNet(multiple kernel dilated convolution network)^[12]利用多核膨胀卷积块扩展感受野,学习更稳定和异构的特征表示。HSNet(hybrid semantic network)^[13]以金字塔 Transformer 为编码器,通过交互机制来交换 2 种类型的语义信息,缩小高级和低级特征间差距;PVT(pyramid vision tranformer)^[14]以 Transformer 为编码器,不仅能够有效抑制噪声,还能显著增强特征表达能力。此外,为进一步增强特征提取能力,MIA-Net(multi-information aggregation network)^[15]利用 Transformer 和 CNNs(convolutional neural networks)双编码器架构,分别提取全局和局部特征;FAT-Net(feature adaptive transformers)^[16]作为特征自适应 Transformer 网络,为有效捕获长程依赖关系,集成一个额外的 Transformer 分支;Transfuse(fusing transformers)^[17]结合 CNNs 和 Transformer,有效捕获全局信息和低级空间细节特征;为捕获丰富全局上下文信息,MNFE-Net(multiscale nonlocal feature extraction network)^[18]基于 Transformer 和 CNNs 提出并行编码器结构,提高网络分割性能。以上 4 种模型均是以 Transformer 和 CNNs 作为双分支对输入图像进行编码。

为捕获更多细节特征信息,注意力机制在医学图像分割任务中的重要性已经凸显。ACSNet(adaptive context selection)^[19]提出局部上下文注意力(local context attention, LCA)模块,进一步加强

对复杂区域关注,补充局部细节信息;为捕获更多细节特征,PraNet(parallel reverse attention network)^[20]采用反转注意力(reverse attention, RA)模块对目标区域进行反转,逐步挖掘息肉区域信息;但在息肉小目标分割任务中,获取的全局信息和局部信息仍不够全面。为缓解以上问题,CaraNet(context axial reverse attention network)^[21]提出上下文轴向反转注意力网络来提高息肉小目标分割性能。该网络使用轻量级通道特征金字塔模块提取丰富特征信息;为获取更准确特征信息,设计轴向反转注意力模块对位置信息进行分析。为进一步提高息肉小目标分割性能,充分考虑卷积核较小的卷积运算和膨胀卷积的优势,并结合注意力机制,这能够获得息肉小目标更丰富的全局上下文信息和细粒度局部细节特征。

结合膨胀卷积和注意力机制虽然能够有效提高网络分割性能,但对于息肉小目标分割任务,仍存在漏检和误检问题,导致模型分割性能不理想。针对上述问题,本文提出融合 Transformer 辅助编码器和边界反转注意力的特征挖掘息肉小目标分割网络(transformer feature boundary network, TFB-Net),来促进结肠镜图像中息肉小目标分割。其中 Transformer 辅助编码器(Transformer auxiliary encoder, TAE)捕获输入图像全局上下文信息,利用自注意力机制建立像素间长程依赖关系,弥补 CNNs 中固有局限性;特征挖掘模块(feature mining module, FMM)将每层融合的特征进行深层次挖掘,进一步细化特征,学习特征表示;边界反转注意力(boundary reverse attention, BRA)模块用于加强对边缘关注,有效抑制不相关噪声,提升区域辨别能力。在 5 个可用息肉小目标数据集上对 TFB-Net 进行评估,大量的实验结果表明所提出的 TFB-Net 显著提高息肉小目标分割精度。综上所述,本文的贡献可归纳如下:

1) 提出一种新的融合 Transformer 的分割网络推进结肠息肉小目标分割,称为 TFB-Net。TFB-Net 利用 Transformer 辅助 CNNs 编码器进行特征提取。该网络可以补充 CNNs 编码器底部 3 层缺少的全局特征,从而提高息肉小目标分割精度。

2) 利用特征挖掘模块替代传统跳跃连接,恢复上采样过程中的细节信息,以不同感受野提取详细特征,针对不同层级采用不同特征挖掘模块,有效缩小编码器和解码器间语义鸿沟。

3) 采用边界反转注意力模块加强对特征边界区域关注,有效抑制不相关噪声,增强边界信息

处理能力,提高特征整体鲁棒性。

4) 在 5 个息肉小目标数据集上进行广泛实验,相比于 10 种先进分割网络,实验结果表明 TFB-Net 在息肉小目标分割上具有优越的分割性能,特别在 ETIS-LaribPolyPDB*数据集上,该网络在 Dice 和 IoU 指标上分别超过 CaraNet 网络 2.35 百分点和 3.91 百分点。

1 相关工作

Transformer 最初专注于自然语言处理 (natural language processing, NLP) 领域,由于自注意力机制在建模长程依赖关系的优异表现,Transformer 已经广泛应用在各种计算机视觉任务中。视觉 Transformer (vision Transformer, ViT)^[22] 首次将 Transformer 应用于图像分类任务,相比于卷积网络,通过应用 Transformer 获得令人满意的分类效果;语义分割 Transformer (semantic segmentation Transformer, SETR)^[23] 利用 Transformer 作为编码器,将输入图像编码为图像序列块,提取丰富全局上下文特征,获得较好的语义分割结果;数据高效图像 Transformer (data-efficient image Transformers, DeiT)^[24] 引入老师-学生策略进一步提高精度。

为学习更稳定的特征表示,PVT^[14] 将金字塔结构 Transformer 作为编码器,不仅减少整体计算量,而且获得较高输出分辨率。PVT 分割网络继承 Transformer 和 CNNs 优点,有效抑制图像中噪声,显著提高特征表达能力。相较于 PVT,Transfuse^[17] 在 PVT 基础上添加一条 CNNs 分支,以并

行分支结构结合 CNNs 和 Transformer,有效捕获全局上下文信息和低级空间细节特征;MGW-Net^[25] 采用可以捕捉多尺度信息的 PVT 编码器和实现不同层之间多尺度信息交互的并行多级特征融合解码器组成,进一步改善了分割性能,从而实现更高的分割精度。

为了充分利用自注意力机制和分层结构的优势,考虑在息肉小目标分割任务中使用金字塔 Transformer。不同于以往单一 Transformer 编码器和并行双分支编码器,考虑将金字塔 Transformer 作为辅助编码器,将其底部 3 层特征分别和 CNNs 中同级特征进行有效融合,增强特征稳健性,实现更好分割效果。

2 网络模型

本文提出融合 Transformer 和边界反转注意力的特征挖掘息肉小目标分割网络,称为 TFB-Net,其网络结构整体框架如图 1 所示。从图 1 中可得出 TFB-Net 网络结构主要包括 3 个核心模块,包括 Transformer 辅助编码器、特征挖掘模块以及边界反转注意力模块。Transformer 辅助编码器弥补 CNNs 编码器固有局限性,建立输入图像像素间长程依赖关系,补充全局特征信息;设计特征挖掘模块代替跳跃连接,融合不同尺度卷积和膨胀卷积操作生成更精细的特征图,进一步细化全局语义空间特征;在此基础上,利用边界反转注意力模块加强对边界区域关注,获取更多边界细节信息,减少特征噪声,使特征图更具稳健性。下面将逐一介绍各个模块。

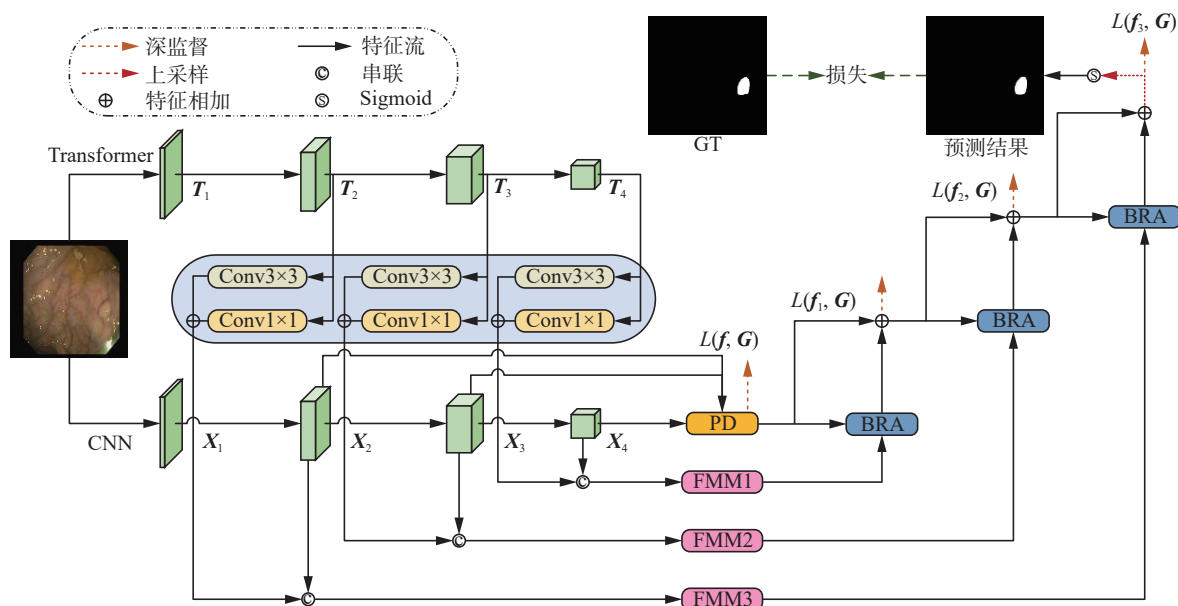


图 1 网络结构整体框架

Fig. 1 Overall framework of the network structure

2.1 Transformer 辅助编码器

相比于卷积神经网络, Transformer 通过建模长程依赖关系捕获全局上下文信息的极好表现使其频繁出现在各种计算机视觉任务中。因此, 本文考虑引入一个额外的 Transformer 作为辅助编码器与 CNNs 分支进行分层连接, 将局部特征信息和全局上下文信息相融合。此外, 由于 PVT 网络在多个密集检测任务中实现最先进的性能, 通过空间缩减序列有效降低计算负担, 所以 TFB-Net 采用预训练 PVT 对全局特征进行分层编码, 得到由粗到细的 4 层金字塔特征。

首先, 将输入图像 $X \in \mathbb{R}^{H \times W \times 3}$ 平均分成 $N = \frac{H}{P} \times \frac{W}{P}$, 其中 P 是每个图像块大小, N 是图像块总数, 为匹配 Transformer 层的输入, 通过对这些二维图像块应用线性映射函数转化为一维嵌入图像序列 $X_e \in \mathbb{R}^{N \times C_e}$, 输出维度为 C_e 。为保留图像原本的空间位置信息, 将其对应的相同维度可学习位置嵌入信息添加到生成的图像序列 $X_e \in \mathbb{R}^{N \times C_e}$ 中。最后将生成的图像嵌入 $X_l \in \mathbb{R}^{N \times C_e}$ 输入到 Transformer 模块中, Transformer 模块结构如图 2 所示。

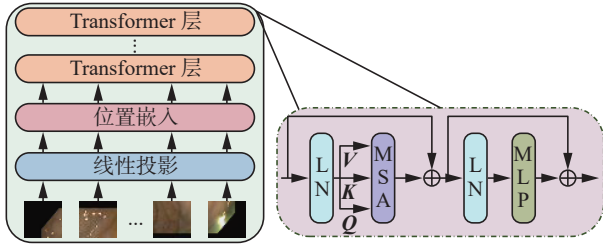


图 2 Transformer 模块结构

Fig. 2 Transformer module structure

1 个经典 Transformer 块通常包括 3 个关键部分: 多头自注意力、多层感知机和层归一化。其中高效的自注意力机制作为 Transformer 块的核心, 通过在每个层中聚合全局信息并行更新每个图像嵌入块状态。由生成的输出图像块嵌入 X_l 或

者前一个 Transformer 输出 X_m 计算得到图像块的 3 个序列向量, 包括查询向量 Q 、键向量 K 和值向量 V 。然后, 通过对 Q 和 K 进行点积运算来计算注意力得分。此外, 为确保所有值均为正数且相加为数值 1, 对生成的向量应用 Softmax 激活函数来归一化得分。最后, 将每个 V 与归一化得分相乘获得输出向量, 并对加权值 V 求和。自注意力的整个计算过程为

$$Q = X_m * W_Q, K = X_m * W_K, V = X_m * W_V$$

$$SA(X) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

式中: $W_Q \in \mathbb{R}^{C_e \times d_q}$ 、 $W_K \in \mathbb{R}^{C_e \times d_k}$ 和 $W_V \in \mathbb{R}^{C_e \times d_v}$ 是 3 个参数可学习的 Q 、 K 以及 V 。

多头自注意力具有 h 个独立自注意层, 将 Q 、 K 和 V 映射到 h 个头, 并投射到原始高维空间的不同子空间以并行方式进行自注意力计算。最后, 将不同子空间中用于计算自注意得分的多个头进行拼接。

从 TAE 中可得到 4 个不同层次语义特征 $\{T_i, i = 1, 2, 3, 4\}$, 并将其底部 3 层特征进行 1×1 和 3×3 卷积操作处理。进而与 CNNs 中底部 3 层特征 $\{X_i, i = 2, 3, 4\}$ 按通道进行连接分别得到 $\{C_i, i = 1, 2, 3\}$, 此时, TAE 辅助 CNNs 编码器在底部 3 层跳跃连接上获得全局上下文信息。

2.2 特征挖掘模块

在通道维度上直接融合 TAE 和 CNNs 编码器特征可能会导致输出特征不够细化。受跳跃连接启发, 本文设计了特征挖掘模块 (FMM), 将前期提取的特征 $\{C_i, i = 1, 2, 3\}$ 通过 FMM 进行优化并传递到对应的上采样过程中, 挖掘全局语义空间特征, 自适应匹配编码器和解码器之间的特征分布, 促进特征融合。特征挖掘模块具体结构如图 3 所示。针对息肉小目标分割问题, 膨胀率不宜设置过大, 故设置膨胀率为 2。

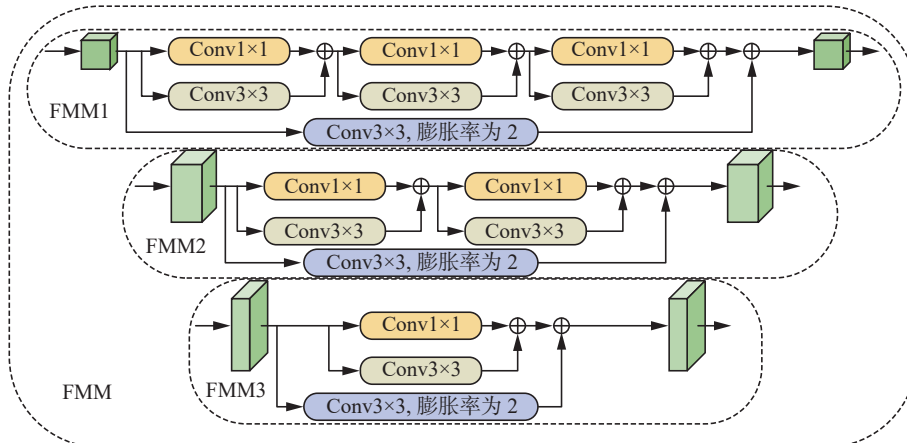


图 3 特征挖掘模块结构

Fig. 3 Feature mining module structure

针对不同层的特征表示具有差异性,在 FMM 模块中分别考虑 3 种方式以进一步精细化全局特征,进而学习更好特征表示,其中包括 FMM1、FMM2 以及 FMM3。将一个核大小为 3×3 的卷积操作和另一个核大小为 1×1 的卷积操作按元素进行相加,并将其设定为 Sub-FMM。由于编码器第 2 层特征图具有相对低级的特征,在 FMM3 中设置 1 个 Sub-FMM 来更好匹配编码器和解码器间特征分布。而编码器第 3 层和第 4 层特征图相对比较高级,因此,在 FMM2 和 FMM1 中分别设置 2 个和 3 个 Sub-FMM,缩小编码器和解码器间特征间隙。

前期融合的 3 层全局特征作为 FMM 模块的输入,经不同卷积计算和融合操作,可以进一步从原始全局特征中提取更详细特征。同时,在 FMM1、FMM2 以及 FMM3 的输入特征处分别应用 1 个核大小为 3×3 且膨胀率为 2 的卷积操作,通过按元素相加的方式分别与上述输出特征进行融合操作。

2.3 边界反转注意力模块

由于息肉小目标图像同样呈边缘模糊、成像亮度不一以及背景复杂等特点,受注意力机制启发,为提高特征鲁棒性,本文部署边界反转注意力 (BRA) 模块,有效增强边界细节信息,突出模型的抗噪能力。在上采样过程中,经 BRA 模块处理可以极大地恢复上采样过程中丢失细节,缩小编码器和解码器间语义鸿沟。边界反转注意力模块如图 4 所示。

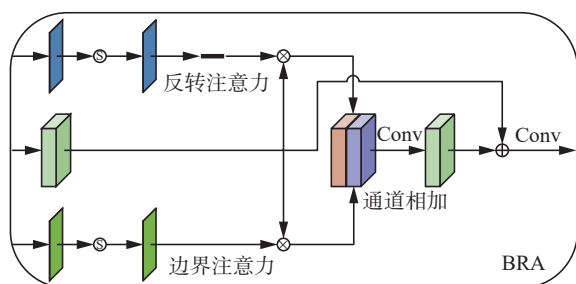


图 4 边界反转注意力模块

Fig. 4 Boundary inversion attention module

通过将输入特征分别与边缘注意力特征图和反转注意力特征图按通道执行相乘操作。将获得的 2 种注意力特征按通道维度进行拼接,并输入到卷积层,使通道变为原来的 $1/2$,接着与原输入特征执行按像素相加操作,这将有效提高区域辨别能力,从而实现更准确分割。从图 1 中可得到,通过将 $\{F_i, i=1,2,3\}$ 作为 3 个 BRA 模块的输入,最终得到 3 个输出特征 $\{B_i, i=1,2,3\}$ 。

2.4 损失函数

为缓解过拟合现象,获得更好性能,在损失函数中应用加权 IoU(intersection over union) 损失和加权二值交叉熵 (binary cross-entropy, BCE) 损失:

$$L = L_{IoU}^w + L_{BCE}^w$$

IoU^[26] 损失在处理边界上的突出表现已在许多挑战性医学图像分割任务中得到验证。此外,息肉小目标分割为二分类问题,而二值交叉熵损失主要适用于二分类图像分割任务。

为充分训练底层特征,在经 CNNs 编码器底部 3 层特征聚合得到的全局引导特征图和上采样过程生成的 3 层特征图中应用深监督策略,所有预测特征图上采样到与 G (ground truth) 相同分辨率大小。因此,最终训练损失计算公式为

$$L_{total} = \sum_{i=1}^3 L(G, f_i^{up}) + L(G, f)$$

3 实验与结果分析

3.1 数据集

本文利用 5 个息肉小目标数据集 (CVC-300*、CVC-ClinicDB*、Kvasir*、CVC-ColonDB*以及 ETIS-LaribPolyPDB*) 来验证 TFB-Net 的性能,具体如表 1 所示。为便于区分,在 5 个公共息肉数据集 (CVC-300^[27]、CVC-ClinicDB^[28]、Kvasir^[29]、CVC-ColonDB^[30] 和 ETIS-LaribPolyPDB^[31]) 名称后统一添加符号“*”分别代表 5 个息肉小目标数据集。

表 1 实验数据集

Table 1 Experimental datasets

数据集	总数量	训练集数量	测试集数量
CVC-300*	35	—	35
CVC-ClinicDB*	121	102	19
Kvasir*	77	67	10
CVC-ColonDB*	158	—	158
ETIS-LaribPolyPDB*	116	—	116

3.2 实现细节

在所有实验中,将所有图片大小统一调整为 352 像素 \times 352 像素。采用 Res2Net^[32] 作为主干编码器,使网络具有强大的多尺度表示能力。本文训练 TFB-Net 网络 100 个 epoch, batch size 设置为 8,部署系统环境为 Ubuntu 20.04、Python 3.9 以及 PyTorch 1.11,实验统一在单张 NVIDIA GeForce GTX 3090 GPU 上进行。

3.3 评价指标

为定量评价 TFB-Net 在息肉小目标数据集上分割性能,使用 5 个广泛使用的分割指标,包括

Dice 相似系数 (dice similarity coefficient, Dice)、交并比 (intersection over union, IoU)、灵敏度 (sensitivity, SE)、特异性 (specificity, SP) 以及准确率 (accuracy, ACC)。Dice 系数通常用来计算 2 个样本之间的相似度, 交并比通常用来表示预测特征图和 Ground Truth 之间的重叠率。计算方法分别为

$$I_{\text{Dice}} = \frac{2 \cdot N_{\text{TP}}}{2 \cdot N_{\text{TP}} + N_{\text{FP}} + N_{\text{FN}}}$$

$$I_{\text{IoU}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}} + N_{\text{FN}}}$$

$$I_{\text{SE}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}$$

$$I_{\text{SP}} = \frac{N_{\text{TN}}}{N_{\text{TN}} + N_{\text{FP}}}$$

$$I_{\text{Acc}} = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{TN}} + N_{\text{FP}} + N_{\text{FN}}}$$

式中: N_{TP} 和 N_{TN} 分别表示正确分割息肉小目标像素的数目, N_{FP} 代表将背景像素错误分割成息肉小目标的像素个数, N_{FN} 表示将息肉小目标像素错误分割为背景像素的个数。

3.4 消融研究

通过将 Transformer 辅助编码器、特征挖掘模块以及边界反转注意力模块按照不同组合方式替换到基准网络, 分别实现了 Baseline+TAE、Baseline+FMM、Baseline+BRA、Baseline+TAE+FMM、Baseline+TAE+BRA、Baseline+FMM+BRA 6 种方法, 在 5 个息肉小目标数据集上进行测试。实验结果如表 2 所示。可视化分割效果图如图 5 所示, 图 5(a) 表示输入图像, 图 5(b) 表示 Ground Truth, 图 5(c) 表示基准网络, 图 5(d)~(i) 依次代表上述 6 种分割方法, 图 5(j) 表示 TFB-Net。

表 2 消融研究结果
Table 2 Ablation study results

%

方法	CVC-300*		CVC-ClinicDB*		Kvasir*		CVC-ColonDB*		ETIS-LaribPolyPDB*	
	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
Baseline	81.50	72.35	77.72	68.93	78.77	70.03	71.16	61.34	68.08	58.12
Baseline+TAE	86.15	77.70	83.56	75.22	84.06	75.84	74.30	65.62	72.96	63.77
Baseline+FMM	86.42	78.41	83.44	75.89	83.67	76.01	74.81	66.25	72.60	64.10
Baseline+BRA	85.39	77.43	80.15	72.84	81.08	73.82	74.96	66.56	71.65	63.20
Baseline+TAE+FMM	86.34	77.77	83.42	75.53	84.04	76.24	76.95	67.97	72.85	64.44
Baseline+TAE+BRA	87.05	79.90	84.27	77.26	84.94	77.90	75.52	67.67	71.88	64.32
Baseline+FMM+BRA	86.11	79.18	83.81	76.87	84.26	77.28	75.81	67.73	73.68	65.89
Baseline+TAE+FMM+BRA	88.80	81.16	87.15	79.52	87.18	79.58	78.72	70.06	76.15	67.26

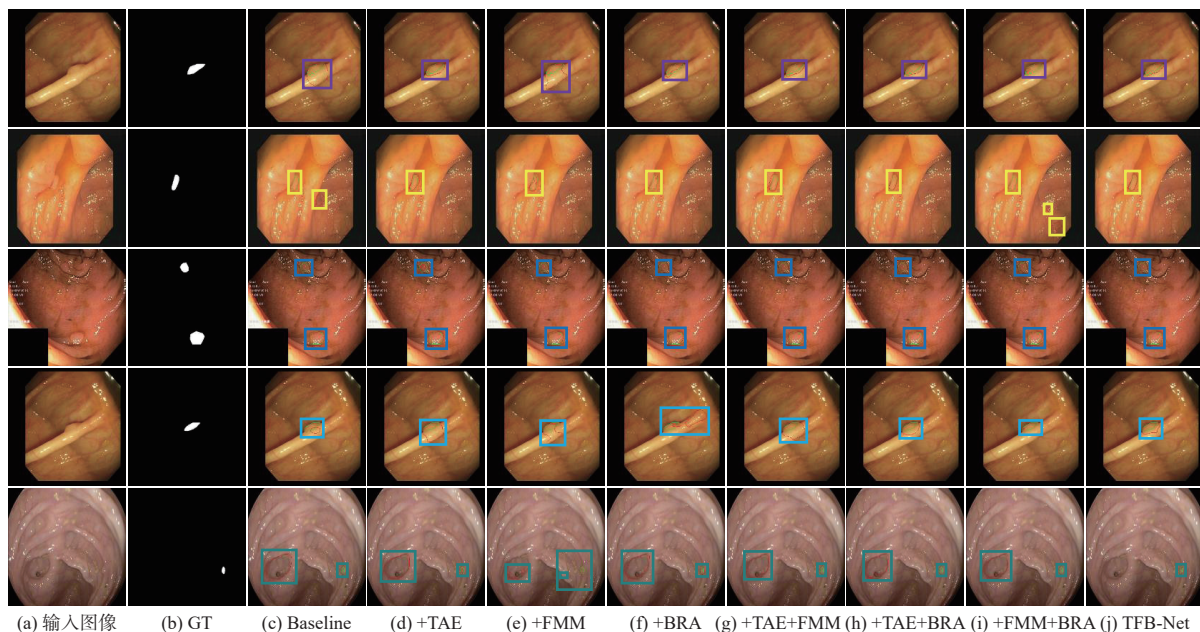


图 5 消融研究分割效果

Fig. 5 Segmentation effect of ablation study

从表 2 消融研究结果中可看出 TFB-Net 明显优于上述 6 种消融分割方法。从图 5 消融研究分割效果图可清晰看出 TFB-Net 的分割效果最佳。

单独模块有效性 将预训练金字塔 Transformer 作为辅助,与 Res2Net 共同构成编码器。从表 2 可以看出,添加 TAE 后网络性能在 5 个息肉小目标数据集上均有明显的提高。由于 transformer 更加关注全局信息,能够更准确地定位小目标息肉的位置。之后通过反转突出目标的边缘细节。在基线中添加 FMM 后,网络模型在 CVC-ClinicDB*数据集上 Dice 和 IoU 分别提高了 5.72 个百分点和 6.96 百分点。在基线中添加 BRA 后,其分割性能在 CVC-300*数据集上 Dice 和 IoU 分别提高了 3.89 个百分点和 5.08 百分点。

组合模块有效性 将 FMM 模块添加到 Baseline+TAE 中,由表 2 可得到在 CVC-ColonDB*数据集中,Baseline+TAE+FMM 比 Baseline+TAE 具有更好分割性能,其 Dice 和 IoU 分别提高 2.65 百分点和 2.35 百分点。同样在 CVC-ColonDB*数据集中,Baseline+TAE+FMM 比 Baseline+FMM 分割精度更高, Dice 和 IoU 分别提高 2.14 百分点和 1.72 百分点。由表 2 知相对于 Baseline+TAE, Baseline+TAE+BRA 在 CVC-ColonDB*数据集上性能得到

提升,其 Dice 和 IoU 分别提高了 1.22 百分点和 2.05 百分点。相比于 Baseline+BRA, Baseline+TAE+BRA 在 Clinic-DB*数据集上 Dice 和 IoU 分别提高 4.12 百分点和 4.42 百分点。通过图 5 可直观地看出,其在 CVC-300*、CVC-ClinicDB*以及 Kvasir*数据集上性能表现优于 Baseline+TAE 和 Baseline+BRA。通过表 2 数据对比 Baseline+FMM,可发现 Baseline+FMM+BRA 在 ETIS-Larib-PolyPDB*数据集上 Dice 和 IoU 分别提高了 1.08 百分点和 1.79 百分点。此外,在 Clinic-DB*数据集上,其 Dice 和 IoU 相比 Baseline+BRA 分别提高了 3.66 百分点和 4.03 百分点。

3.5 对比实验

将 TFB-Net 和 10 个先进网络模型进行对比,包括 U-Net^[6]、UNet++^[7]、UNet3+^[8]、PVT^[14]、FAT-Net^[16]、Transfuse^[17]、MNFE-Net^[18]、ACSNet^[19]、PraNet^[20] 以及 CaraNet^[21]。对比实验结果如表 3 所示。其可视化分割结果如图 6 所示,其中图 6(a) 代表输入图像,图 6(b) 表示 Ground Truth,图 6(c)~(l) 依次代表 U-Net、UNet++、UNet3+、Transfuse、PVT、FAT-Net、MNFE-Net、ACSNet、PraNet 以及 CaraNet,图 6(m) 表示 TFB-Net。绿色和红色轮廓曲线分别代表 Ground Truth 和分割结果。

表 3 对比实验结果

Table 3 Comparative experimental results

%

数据集	指标	U-Net	UNet++	UNet3+	Transfuse	PVT	FAT-Net	MNFE-Net	ACSNet	PraNet	CaraNet	TFB-Net
CVC-300*	Dice	53.46	70.93	72.51	84.58	75.59	79.82	85.04	84.74	86.72	84.62	88.80
	IoU	44.97	62.23	63.41	75.57	65.13	70.22	77.75	78.05	77.89	74.87	81.16
	ACC	98.81	99.17	99.26	99.39	98.98	99.38	99.31	99.44	99.44	99.35	99.56
	SE	47.67	72.98	68.78	96.37	85.58	80.41	97.07	93.46	97.57	97.83	96.81
	SP	99.93	99.65	99.86	99.44	99.27	99.72	99.36	99.59	99.47	99.36	99.62
Clinic-DB*	Dice	61.35	65.37	64.67	68.94	72.43	64.80	75.83	72.55	75.23	77.69	87.15
	IoU	56.59	59.62	58.66	61.62	63.42	58.30	67.20	66.21	67.52	68.42	79.52
	ACC	99.21	99.24	99.19	98.88	98.98	99.19	98.86	99.34	99.23	99.23	99.53
	SE	59.40	63.01	63.18	74.51	79.37	62.57	87.51	77.15	80.48	87.80	93.86
	SP	99.92	99.85	99.81	99.32	99.34	99.81	99.13	99.73	99.63	99.41	99.66
Kvasir*	Dice	63.58	61.03	64.98	78.81	72.03	59.52	80.14	83.45	78.73	78.74	87.18
	IoU	55.54	54.99	56.72	71.97	63.09	52.74	73.52	75.47	71.41	69.47	79.58
	ACC	98.97	98.66	99.05	99.17	98.98	99.01	99.25	99.36	99.20	99.25	99.52
	SE	56.42	60.35	58.43	80.89	78.39	54.87	80.04	80.39	81.57	88.23	93.29
	SP	99.74	99.43	99.75	99.51	99.36	99.77	99.58	99.70	99.54	99.43	99.65
Colon-DB*	Dice	28.19	34.99	31.86	66.88	60.06	46.30	73.01	67.26	71.08	76.51	78.72
	IoU	23.23	29.89	26.78	57.86	50.33	39.87	63.11	59.55	61.46	66.12	70.06
	ACC	98.77	98.92	98.88	99.19	98.72	99.13	99.01	99.30	99.27	99.30	99.40
	SE	24.77	33.63	28.86	72.63	65.71	43.81	84.61	68.81	75.48	84.77	82.99
	SP	99.94	99.86	99.94	99.50	99.22	99.87	99.19	99.73	99.61	99.50	99.67
ETIS-LaribPolyPDB*	Dice	28.14	35.58	28.12	57.66	52.55	28.89	60.86	63.25	63.38	73.8	76.15
	IoU	23.74	29.76	23.58	47.36	43.43	24.63	51.74	56.19	54.86	63.35	67.26
	ACC	99.13	99.13	99.12	98.83	98.63	99.14	98.52	99.42	99.27	99.26	99.37
	SE	25.79	35.84	26.49	74.51	59.41	26.74	81.00	68.13	72.25	84.38	82.66
	SP	99.93	99.80	99.92	99.06	99.11	99.90	98.68	99.71	99.52	99.43	99.60

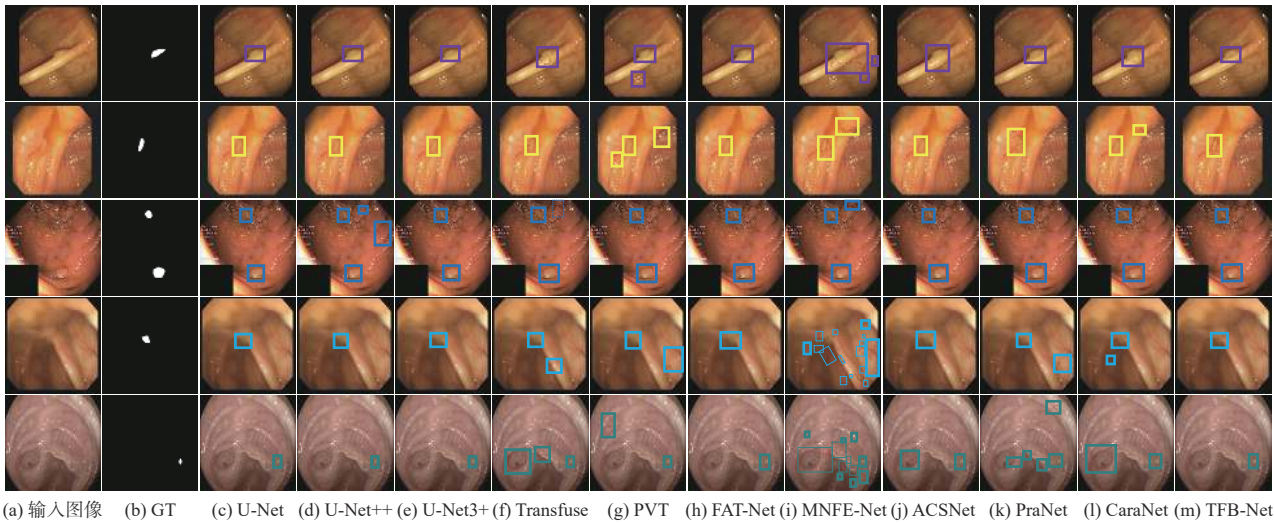


图 6 对比实验可视化分割效果
Fig. 6 Comparison experiment visual segmentation effect

从表 3 可以看出, 在 5 个息肉小目标数据集上, TFB-Net 在 Dice 和 IoU 指标上均获得最高值。TFB-Net 在 CVC-300*数据集上 Dice 值比 PraNet 高 2.08 百分点, 在 Kvasir*数据集上超过 ACSNet3.73 百分点, 在 CVC-ClinicDB*、CVC-ColonDB*和 ETIS-LaribPolyPDB*数据集中, 分别超越 CaraNet 网络 9.46 百分点、2.21 百分点和 2.35 百分点。从图 6 对比实验分割效果图可看出 TFB-Net 具有最佳分割效果, 尤其是在 CVC-

300*、CVC-ClinicDB*以及 ETIS-LaribPolyPDB*数据集上对比效果尤为明显。此外, 本文另外添加 3 个网络参数, 分别是参数数量 (parameters, Params)、浮点运算次数 (floating-point operations, FLOPs) 以及每秒传输帧数 (frames per second, FPS), 用于对比 TFB-Net 和其他网络在模型大小、复杂度和推理速度方面的性能, 具体如表 4 所示, TFB-Net 不仅在 5 个息肉小目标数据集上 Dice 和 IoU 指标均有显著提升, 并且降低了模型复杂度。

表 4 不同先进模型大小、复杂度和推理速度统计情况
Table 4 Statistics of size, complexity and inference speed of different advanced models

指标	U-Net	UNet++	UNet3+	Transfuse	PVT	FAT-Net	MNFE-Net	ACSNet	PraNet	CaraNet	TFB-Net
Params/ 10^6	31.04	47.18	26.97	26.17	25.11	29.62	58.16	29.45	30.50	44.59	69.78
FLOPs/ 10^9	103.49	377.45	377.64	8.65	10.02	42.80	9.42	21.75	13.15	21.75	31.55
FPS/(f/s)	99.72	38.38	29.13	169.56	71.73	176.36	104.93	74.77	116.81	80.26	64.20

4 结束语

本文提出融合 Transformer 和边界反转注意力的特征挖掘息肉小目标分割网络, 称为 TFB-Net。与传统卷积神经网络编码器不同, TFB-Net 以 Transormer 辅助 CNNs 编码。相比于普通卷积操作, Transormer 不仅利用强大的自注意力机制扩大感受野范围, 还能捕获丰富全局上下文信息。此外, 为获取更全面特征, 利用特征挖掘模块进一步学习稳定特征表示。为减少特征冗余, 设计反转注意力模块, 加强对特征边界的关注。本文在 5 个息肉小目标数据集 (CVC-300*、CVC-ClinicDB*、Kvasir*、CVC-ColonDB*以及 ETIS-LaribPolyPDB*) 上进行广泛实验, 实验结果表明, 相比于 10 种先进的分割网络, TFB-Net 具有出色的分割能力。在医学图像分割领域, 息肉小目标

分割俨然仍是当前一项富有挑战性的难题。未来的工作将仍致力于这一任务的研究, 并尝试使用其他方法降低模型的复杂度, 提高模型运行速度。

参考文献:

[1] XI Yue, XU Pengfei. Global colorectal cancer burden in 2020 and projections to 2040[J]. *Translational oncology*, 2021, 14(10): 101174.

[2] KOLLIGS F T. Diagnostics and epidemiology of colorectal cancer[J]. *Visceral medicine*, 2016, 32(3): 158–164.

[3] DAI Ying, CHEN Weimin, XU Xuanfu, et al. Factors affecting adenoma risk level in patients with intestinal polyp and association analysis[J]. *Journal of healthcare engineering*, 2022: 9479563.

[4] LI Zewen, LIU Fan, YANG Wenjie, et al. A survey of convolutional neural networks: analysis, applications, and

- prospects[J]. *IEEE transactions on neural networks and learning systems*, 2022, 33(12): 6999–7019.
- [5] SARVAMANGALA D R, KULKARNI R V. Convolutional neural networks in medical image understanding: a survey[J]. *Evolutionary intelligence*, 2022, 15(1): 1–22.
- [6] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234–241.
- [7] ZHOU Zongwei, RAHMAN SIDDIQUEE M M, TABAKHSH N, et al. UNet++: a nested U-net architecture for medical image segmentation[C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer, 2018: 3–11.
- [8] HUANG Huimin, LIN Lanfen, TONG Ruofeng, et al. UNet 3: a full-scale connected UNet for medical image segmentation[C]//2020 IEEE International Conference on Acoustics, Speech and Signal Processing. Barcelona: IEEE, 2020: 1055–1059.
- [9] ZHANG Zhengxin, LIU Qingjie, WANG Yunhong. Road extraction by deep residual U-net[J]. *IEEE geoscience and remote sensing letters*, 2018, 15(5): 749–753.
- [10] JHA D, SMEDSRUD P H, RIEGLER M A, et al. ResUNet: an advanced architecture for medical image segmentation[C]//2019 IEEE International Symposium on Multimedia. San Diego: IEEE, 2019: 225–2255.
- [11] TOMAR N K, JHA D, BAGCI U. DilatedSegNet: a deep dilated segmentation network for Polyp segmentation[C]//MultiMedia Modeling. Cham: Springer, 2023: 334–344.
- [12] TOMAR N K, SRIVASTAVA A, BAGCI U, et al. Automatic polyp segmentation with multiple kernel dilated convolution network[C]//2022 IEEE 35th International Symposium on Computer-Based Medical Systems. Shenzhen: IEEE, 2022: 317–322.
- [13] ZHANG Wenchao, FU Chong, ZHENG Yu, et al. HSNet: a hybrid semantic network for polyp segmentation[J]. *Computers in biology and medicine*, 2022, 150: 106173.
- [14] DONG Bo, WANG Wenhai, FAN Dengping, et al. Polyp-PVT: polyp segmentation with pyramid vision transformers[J]. *CAAI artificial intelligence research*, 2023: 9150015.
- [15] LI Weisheng, ZHAO Yinghui, LI Feiyan, et al. MIA-Net: multi-information aggregation network combining transformers and convolutional feature learning for polyp segmentation[J]. *Knowledge-based systems*, 2022, 247: 108824.
- [16] WU Huisi, CHEN Shihuai, CHEN Guilian, et al. FAT-Net: feature adaptive transformers for automated skin lesion segmentation[J]. *Medical image analysis*, 2022, 76: 102327.
- [17] ZHANG Yundong, LIU Huiye, HU Qiang. TransFuse: fusing transformers and CNNs for medical image segmentation[C]//Medical Image Computing and Computer Assisted Intervention. Cham: Springer, 2021: 14–24.
- [18] LIU Guoqi, WANG Jiajia, LIU Dong, et al. A multiscale nonlocal feature extraction network for breast lesion segmentation in ultrasound images[J]. *IEEE transactions on instrumentation and measurement*, 2023, 72: 1–12.
- [19] ZHANG Ruifei, LI Guanbin, LI Zhen, et al. Adaptive context selection for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention. Cham: Springer, 2020: 253–262.
- [20] FAN Dengping, JI Gepeng, ZHOU Tao, et al. PraNet: parallel reverse attention network for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention. Cham: Springer, 2020: 263–273.
- [21] LOU Ange, GUAN Shuyue, KO H, et al. CaraNet: context axial reverse attention network for segmentation of small medical objects[C]//Medical Imaging 2022: Image Processing. San Diego: SPIE, 2022: 81–92.
- [22] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. (2020–10–22)[2023–6–13]. <http://arxiv.org/abs/2010.11929>.
- [23] ZHENG Sixiao, LU Jiachen, ZHAO Hengshuang, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 6877–6886.
- [24] TOUVRON H, CORD M, DOUZE M, et al. Training data-efficient image transformers & distillation through attention[EB/OL]. (2020–12–23)[2023–06–13]. <https://arxiv.org/abs/2012.12877>.
- [25] 汪鹏程, 张波涛, 顾进广. 融合多尺度门控卷积和窗口注意力的结肠息肉分割[J]. *计算机系统应用*, 2024, 33(6): 70–80.
- WANG Pengcheng, ZHANG Botao, GU Jinguang. Colon polyp segmentation fusing multi-scale gate convolution and window attention[J]. *Computer systems and applications*, 2024, 33(6): 70–80.
- [26] QIN Xuebin, ZHANG Zichen, HUANG Chenyang, et al. BASNet: boundary-aware salient object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 7471–7481.
- [27] VÁZQUEZ D, BERNAL J, SÁNCHEZ F J, et al. A benchmark for endoluminal scene segmentation of colonoscopy images[J]. *Journal of healthcare engineering*, 2017: 4037190.
- [28] BERNAL J, SÁNCHEZ F J, FERNÁNDEZ-ESPAR-

RACH G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians[J]. *Computerized medical imaging and graphics*, 2015, 43: 99–111.

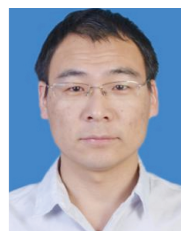
- [29] POGORELOV K, KRISTIN R R, GRIWODZ C, et al. Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection[C]//*Proceedings of the 8th ACM on Multimedia Systems Conference*. Taipei: ACM, 2017: 164–169.

- [30] BERNAL J, SÁNCHEZ J, VILARIÑO F. Towards automatic polyp detection with a polyp appearance model[J]. *Pattern recognition*, 2012, 45(9): 3166–3182.

- [31] SILVA J, HISTACE A, ROMAIN O, et al. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer[J]. *International journal of computer assisted radiology and surgery*, 2014, 9(2): 283–293.

- [32] GAO Shanghua, CHENG Mingming, ZHAO Kai, et al. Res2Net: a new multi-scale backbone architecture[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2021, 43(2): 652–662.

作者简介:



刘国奇, 副教授, 博士, 主要研究方向为图像分割、机器学习、偏微分方程。E-mail: gqliu@htu.edu.cn。



陈宗玉, 硕士研究生, 主要研究方向为图像分割、机器学习。E-mail: chenzongyu1010@163.com。



刘栋, 教授, 主要研究方向为教育数据挖掘和复杂网络分析。E-mail: liudong@htu.edu.cn。

第二届人工智能大模型技术高峰论坛

在 AI 大模型产品化、产业化阶段来临的背景下, 2024 第二届人工智能大模型技术高峰论坛如约起航, 将于 9 月 21 日在江苏南京举办。本届高峰论坛将汇聚 AI 大模型最新研究观点和学术成果, 探讨技术趋势和行业未来, 强化产学研协同、激发场景创新、加速应用落地。

届时, 政产学研各界嘉宾将齐聚高峰论坛开幕式, 见证重量级项目发布, 推动更多跨界合作与资源整合。特邀中外杰出科学家作主旨报告, 汇聚前沿智慧, 碰撞多元思想火花, 引领技术创新方向。特设多模态融媒体大模型、可信安全高效的大模型技术、大模型与具身智能、工业大模型与实体经济、生成式人工智能与新质生产力、人工智能助力科研范式升级换代、大模型 AI 应用技术和工程 7 大专题论坛, 促进知识共享与创新思维的碰撞, 促成跨领域合作解决实际问题。清华大学、东南大学、南京大学、武汉大学、西安交通大学、中山大学、浙江大学、重庆师范大学、中国科学院自动化研究所等多个高校及科研院所的研究成果集中呈现, 赋能大模型技术创新。华为、腾讯、科大讯飞等多家领军企业及创新企业分享大模型技术与行业实践经验, 加速新技术、新理念、新模式的传播与应用。来自各领域、各学科的专家学者将带来更多精彩报告, 凝聚行业精华, 全面覆盖技术前沿, 助力 AI 大模型应用突破。来自全国的部分智能科技从业者、学习者、研究者、教育者等观众到场参会学习交流, 紧跟行业发展动态, 拓展人脉资源, 共同面对挑战, 把握未来趋势。