

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2025)02-0589-12

论文引用格式: Meng X F, Zhang Z C, Yu C L and Zhang X Y. 2025. Colonoscopy polyp image segmentation method with boundary self-knowledge distillation. Journal of Image and Graphics, 30(02):0589-0600(孟祥福, 张智超, 俞纯林, 张霄雁. 2025. 结合边界自知识蒸馏的结肠镜息肉图像分割方法. 中国图象图形学报, 30(02):0589-0600)[DOI:10.11834/jig.240175]

结合边界自知识蒸馏的结肠镜息肉图像分割方法

孟祥福*, 张智超, 俞纯林, 张霄雁

辽宁工程技术大学电子与信息工程学院, 葫芦岛 125000

摘要: 目的 结肠镜技术在结肠息肉的早期检测中至关重要,但其依赖于操作员的专业技能和主观判断,因此存在局限性。现有的结肠息肉图像分割方法通常采用额外层和显式扩展网络结构,导致模型效率较低。此外,由于息肉与其周围粘膜之间的边界不清晰,现有模型对于息肉边界的分割效果并不理想。**方法** 提出了一种端到端的自知识蒸馏框架,专门用于结肠息肉图像分割。该框架将边界分割网络和息肉分割网络整合到一个统一的知识蒸馏框架中,以相互增强两个网络的性能。该框架采用专注于边界分割的模型作为教师网络,将息肉分割模型作为学生网络,两者共享一个特征提取模块,以促进更有效的知识传递。设计了一种反向特征融合结构,通过上采样和矩阵乘法聚合编码器深层特征,并利用反向浅层特征作为辅助信息,从而获得分割掩膜的全局映射。**结果** 通过在CVC-ClinicDB (colonoscopy videos challenge-clinicdatabase)、CVC-ColonDB (colonoscopy videos challenge-colondbase)、Kvasir以及HAM10000(human against machine with 10000 training images) 4个数据集上开展实验,与当前11种先进方法PraNet(parallel reverse attention network)和Polyp2Former(boundary guided network based on transformer for polyp segmentation)等进行比较,实验结果表明本文模型表现最佳,Dice相似性系数(Dice similarity coefficient, DSC)和平均交并比(mean intersection over union, mIoU)指标分别比现有最优模型提升了0.45%和0.68%。**结论** 本文模型适用于各种尺寸和形状的息肉分割,实现了准确的边界提取,并且具有推广到其他医学图像分割任务的潜力。本文代码可在<https://github.com/xiaoxiaotuo/BA-KD>下载。

关键词: 息肉分割;医学图像处理;深度学习;知识蒸馏;边界分割

Colonoscopy polyp image segmentation method with boundary self-knowledge distillation

Meng Xiangfu*, Zhang Zhichao, Yu Chunlin, Zhang Xiaoyan

School of Electronics and Information Engineering, Liaoning Technical University, Huludao 125000, China

Abstract: Objective Colorectal cancer remains a formidable global health challenge, underscoring the pressing need for early detection strategies to improve treatment outcomes. Among these strategies, colonoscopy stands out as a primary diagnostic tool, relying on the visual acumen of medical professionals to identify potentially cancerous abnormalities, such as polyps, within the colon and rectum. However, the effectiveness of colonoscopy is heavily contingent upon the skill and experience of the operator, leading to variability and limitations in detection rates across different practitioners and set-

收稿日期: 2024-03-29;修回日期: 2024-06-14;预印本日期: 2024-06-21

* 通信作者: 孟祥福 marxi@126.com

基金项目: 国家自然科学基金项目(61772249);辽宁省教育厅科学研究项目(LJ2019QL017, LJKZ0355)

Supported by: National Natural Science Foundation of China (61772249); Department of Education of Liaoning Province (LJ2019QL017, LJKZ0355)

tings. In response to these challenges, the integration of artificial intelligence and computer vision techniques has garnered increasing attention as a means to augment the accuracy and efficiency of colorectal cancer screening. Various algorithms have been developed to automatically segment colorectal images, with the overarching goal of precisely delineating polyps from the surrounding tissue. Despite advancements in this domain, many existing models confront inherent inefficiencies and limited effectiveness stemming from their intricate architectures and dependence on manual feature engineering.

Method This study proposes a novel end-to-end boundary self-knowledge distillation (BA-KD) framework, which aims to achieve precise polyp segmentation. In contrast to conventional methods, BA-KD seamlessly integrates boundary and polyp segmentation networks into a unified framework, facilitating effective knowledge transfer between the two domains. BA-KD represents a pioneering contribution in this field, aiming to harness the synergistic benefits of both boundary and polyp information for increased segmentation accuracy. The BA-KD framework comprises two interconnected branches: a boundary segmentation network serving as the teacher branch and a polyp segmentation network acting as the student branch. The inherent challenges associated with delineating polyp boundaries are addressed by introducing a boundary detection operator to automatically generate boundary masks, which are subsequently leveraged when training both branches. This approach not only enhances the segmentation performance of the student branch but also enriches the knowledge base of the teacher branch, thereby fostering mutual learning and refinement. A key distinguishing feature of BA-KD is the shared image feature extractors between the student and teacher branches, facilitating robust knowledge transfer across both domains. Two innovative structures, namely, reverse multilevel feature fusion (RMLF) and reverse feature fusion (RFM), are proposed to facilitate the effective fusion of feature information at various hierarchical levels. RMLF enables the integration of high-level features to generate a comprehensive global feature map, whereas RFM synergistically combines reverse shallow features with high-level features aggregated via RMLF to produce the final segmentation mask.

Result A comprehensive experimental validation of the BA-KD results is conducted against seven state-of-the-art methods across four different datasets: CVC-ClinicDB, CVC-ColonDB, Kvasir, and HAM10000. The comparative models include U-Net, Double-UNet, UNet++, TransFuse, PraNet, DuAT, RaBit, GroupSeg, and G-CASCADE. These models serve as benchmarks in the polyp segmentation and general medical image segmentation domains. The results on CVC-ClinicDB show that BA-KD demonstrates exceptional performance in terms of mSpe and mDSC, with values of 0.997 and 0.955, respectively. BA-KD outperforms all the competitors in terms of mDSC and mIoU, with improvements of 0.45% and 0.68%, respectively, over RaBit. For the CVC-ColonDB dataset, BA-KD outperforms all the other methods across all the evaluation metrics, achieving improvements of 2.20% in the mIoU and 1.51% in the mDSC compared with the optimal performance achieved by TransFuse. For the Kvasir dataset, BA-KD achieves an mIoU of 0.889 and an mDSC of 0.937, surpassing the best-performing RaBit by approximately 1.08% and 1.14%, respectively. Furthermore, the generalization ability of BA-KD is evaluated on other medical segmentation tasks via the HAM10000 dataset, which includes dermoscopic images from different populations. Compared with existing medical segmentation baselines, BA-KD excels in all metrics on HAM10000, achieving significant scores in mDSC (0.956 2) and mIoU (0.922 3), surpassing the best-performing Double UNet by 1.45% and 2.25%, respectively.

Conclusion The experimental results clearly demonstrate that BA-KD outperforms existing state-of-the-art segmentation methods, with substantial improvements in the Dice similarity coefficient (DSC) and mean intersection over union metrics (mIoU).

Key words: polyp segmentation; medical image processing; deep learning; knowledge distillation; boundary segmentation

0 引言

结直肠癌(colorectal cancer, CRC)是全球癌症相关死因的主要原因之一,并且呈现出更加高发趋势(Sung等, 2021),严重威胁到人民群众的生命健康(田传鑫和赵磊, 2021)。由于这一疾病的复杂性和

不可预测性,使得提高早期检测和诊断的准确性变得至关重要(Siegel等, 2019)。结肠镜检查能够直接观察结肠和直肠内部获取组织样本进行生物组织学检查,被广泛认为是结直肠癌早期筛查的金标准(Favoriti等, 2016)。然而,在结肠镜检查中,器官呈现复杂的拓扑结构,并可能会发生形态变化,同时存在消化物等因素,这造成了对操作者的主观性和专

业能力的高度依赖。因此,结肠镜检查的漏诊率高达6%~27%(Ahn等,2012)。为了应对这一挑战,将语义分割技术应用到该领域,语义分割能够将图像中的每个像素按其语义类别进行精确分类。在结肠癌筛查中,通过精确标记图像中的组织结构,使医生更容易检测病变或异常区域。

随着医学图像分析中深度学习技术的发展,已有多种方法应用于息肉分割(Fan等,2020;Zhang等,2020;Zhou等,2018;Shen等,2021),但是在实际应用中仍然存在一些挑战和限制(如图1所示),主要包括以下方面:1)结肠息肉的尺寸具有显著的多样性,从小型微小到大型突出,这种尺寸上的变化使得息肉在检测时容易因为大小差异而被忽视或误判;2)结肠息肉的形状极不规则,它们可能呈现出扁平、隆起和分支等多种形态,这种形态上的不规则性增加了诊断的难度和复杂性;3)结肠息肉总体呈现较为圆滑的外观,且其颜色与周围肠壁相似,因此在实际图像中息肉与其周围粘膜之间的边界往往表现得模糊难辨。现有分割方法对于不规则形状的息肉难以精准捕捉其边界,导致分割效果较差。同时,对于颜色和外观相似的结肠息肉和周围组织,现有方法在边界定位和分割上也存在一定困难。

为了解决上述问题,本文提出了端到端的边界

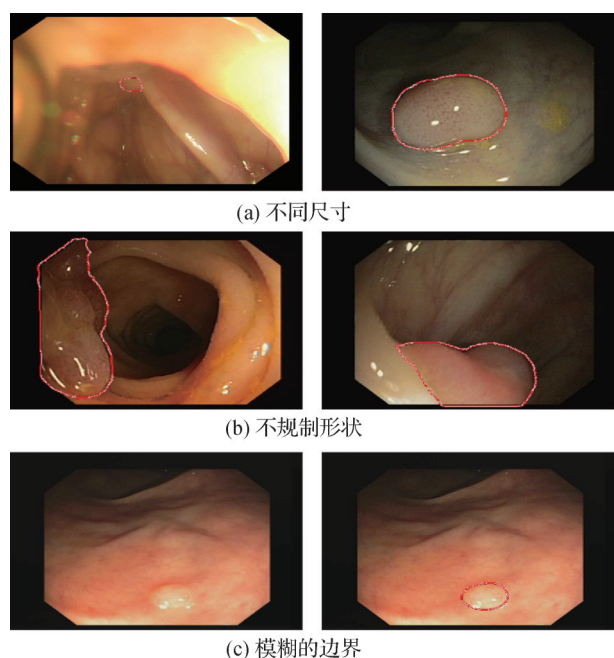


图1 息肉分割任务面临的挑战和局限

Fig. 1 Challenges and limitations of polyp segmentation tasks
((a) different sizes; (b) irregular shape; (c) fuzzy boundaries)

自知识蒸馏框架(boundary self-knowledge distillation,BA-KD),旨在实现对息肉的精准分割。BA-KD将边界分割网络和息肉分割网络整合到一个统一的知识蒸馏框架中,通过在两个网络之间有效的知识转移来相互增强性能。图2给出了BA-KD的处理流程,BA-KD包含一个专注于边界提取的分割网络作为教师分支,以及一个息肉分割网络作为学生分支。为了解决息肉分割中边界模糊问题,引入了边界引导解码器网络(Lin等,2023),利用边界检测算子(Kanopoulos等,1988)提取自动生成的边,需要注意的是,与一般的知识蒸馏方法(Park等,2019;Peng等,2019)不同,本模型不仅训练学生分支,还利用息肉边界掩膜来训练教师分支。

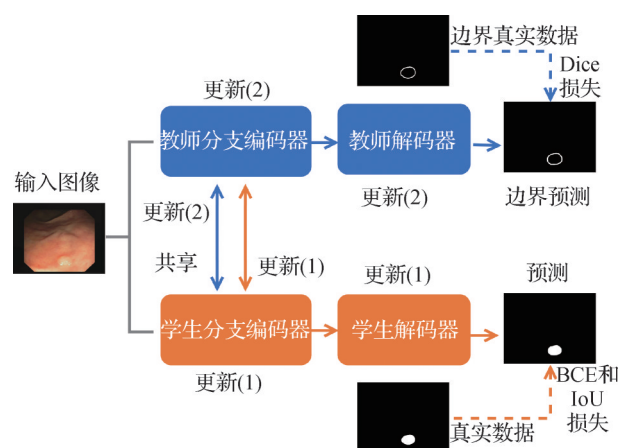


图2 BA-KD的处理流程

Fig. 2 BA-KD processing procedure

此外,模型通过共享学生和教师的图像特征提取器,进一步加强两个分支之间的知识传递。为了更有效地聚合不同尺度的特征信息,提出了反向多级特征融合结构(reverse multi-level fusion,RMLF)和反向特征融合结构(reverse fusion module,RFM),RMLF用于整合高级特征以获取全面上下文信息。为了在整合过程中既能保留关键信息又不丢失细节,设计了一种特定策略,首先聚合高级特征,然后执行取反操作,以保留关键信息之外的细节信息;接着,将得到的细节信息与原始高级特征再次进行聚合,这一设计的主要目标是使模型更深刻地理解图像,并生成准确的全局地图作为后续步骤的指导。在此基础上,RFM用于融合反向浅层特征和通过RMLF聚合得到的高级特征,以生成分割掩膜作为最终结果。

本文贡献概括如下:1)提出一个用于息肉分割的边界自知识蒸馏框架BA-KD。BA-KD将一个边界分割网络和一个息肉分割网络集成在一个知识蒸馏框架中,通过共享特征提取器实现教师和学生网络之间的双向知识转移,这一机制使得两个分割网络能够相互提高性能,从而在边界分割和息肉分割方面取得更好的效果。2)提出RMLF和RFM,这些结构用于整合高级特征,以获取更全面的上下文信息,从而协助模型更深刻地理解图像。3)BA-KD在3个结肠息肉数据集和1个皮肤镜图像数据集进行了大量实验,实验结果表明,BA-KD在结肠息肉分割任务和皮肤癌分割任务中都具有比当前先进的技术更好的性能,验证了BA-KD具有很强的可推广性。

1 相关工作

1.1 息肉分割

U-Net架构已广泛应用于息肉分割,通过各种改进来提高网络性能(Zhang等,2020;Zhao等,2021;Kim等,2021;Du等,2022)。MSNet(Zhao等,2021)通过在U-Net中结合减法单元和金字塔感受野获取多尺度差异信息,结合无需训练的LossNet全面监督,有效提高结肠镜图像中息肉的定位准确性。梁礼明等人(2023)提出一种融合Transformer和跨级相位感知技术的结肠息肉图像分割方法(Transformer and cross-level phase awareness fusion, TCPA-Net),克服了传统方法在空间归纳和全局上下文信息上的不足。通过分层编码、多尺度信息融合和边缘增强,提升了结肠息肉图像的分割精度。UACANet(Kim等,2021)通过在U-Net中引入多个级联的不确定性增强的上下文注意力模块(uncertainty augmented context attention, UACA),用于关注处理不确定性区域,通过学习残差特征图的方式提升结肠息肉分割的准确性。魏天琦和肖志勇(2022)提出了双重编—解码架构,通过上游网络生成注意力权重图,引导下游网络解码过程中的特征图,使模型更关注目标区域。同时,算法引入子空间通道注意力结构,有效细化分割边缘,增强了对小目标分割、边缘分割的准确率。基于Transformer架构的结肠息肉分割模型也逐渐受到关注,进一步拓展了分割方法的多样性。DuAT(dual-aggregation Transformer network)(Tang等,2022)通过引入全局到局部的空间聚

合(global-to-local spatial aggregation, GLSA)和选择性边界聚合(selective boundary aggregation, SBA)模块,解决了传统Transformer在医学图像分割中对关键局部细节的遗漏问题,尤其在处理小物体分割和模糊物体边界等具有挑战性的场景时,DuAT表现出了优越性能。

FCB-Former(Sanderson和Matuszewski,2022)则充分利用Transformer在主分支中提取分割最重要特征的优势,同时通过辅助的全卷积分支弥补其在完整尺寸预测方面的局限性,最终融合两个分支的特征提高了对结肠息肉分割的准确性。前面提到的工作主要使用了U-Net或Transformer架构,并结合了密集的跳跃连接和多尺度残差融合的方法,旨在提升模型性能。然而,密集的跳跃连接和多尺度残差融合方法虽然有助于捕获图像中的多层次信息,但也增加了模型的复杂性。此外,还可能导致模型更容易过拟合小规模或噪声数据,从而影响其泛化能力。特别是在处理结肠息肉时,由于其外观与周围组织相似,以及存在不规则形状和大小变异,模型在精确定位边界方面存在挑战。

1.2 知识蒸馏

知识蒸馏是一种在训练过程中将知识从预先训练的复杂教师网络传递到更简单的学生网络的方法。知识蒸馏的应用阶段,学生网络可以取代教师网络,从而在实际场景中达到更轻量级的模型效果(Chung等,2020;Park等,2019)。一种常用的将知识从教师网络传递到学生网络的方法是使用软标签或注意力权重(Hinton等,2015)。软标签包含概率分布信息,而注意力权重可使学生网络更专注于重要信息,从而实现模型的压缩和性能提升。

自知识蒸馏作为知识蒸馏的一个独特分支,其核心在于模型内部不同组件间知识的传递与共享。在这一过程中,模型的某一部分扮演教师角色,而另一部分则作为学生进行学习,通过内部组件间的相互学习,实现模型整体性能的提升(Caron等,2021)。近年来,自知识蒸馏技术得到了广泛发展,其独特之处在于构建学生网络时无需依赖预先训练的教师网络。与传统知识蒸馏方法不同,自知识蒸馏中的教师网络通常与学生网络结构相同,并通过基于动量的学生网络移动平均值进行实时更新。这一创新策略为学生网络训练提供了新视角,使性能提升更为高效和精准。基于自知识蒸馏理念,本文创新性地

提出了边界自知识蒸馏网络,并应用于结肠息肉分割任务,以优化分割效果。

2 方法

BA-KD的整体框架如图3所示,其中包括一个教师分支和一个学生分支。教师分支是一个边界分割网络,由特征提取编码器、Sobel边界检测算子(Kanopoulos等,1988)和普通卷积组成。学生分支是一个息肉分割模型,由一个特征提取编码器和反向RMLF和一组RFM组成。这两个分支共享同一个实例编码器。教师分支和学生分支交替优化,其

中教师分支的优化用图中的更新(2)表示,学生分支的优化用图中的更新(1)表示。两个分支共用相同的编码器进行图像特征提取,并且共享参数,以实现高效的知識转移。教师分支的解码器由边界检测算子Sobel和普通卷积组成,专注于提取边界信息,能够生成边界预测特征图;学生分支的解码器则采用了RMLF和一组RFM,旨在精确地分割结肠镜检查图像中的息肉。

教师和学生模型采用了交替训练策略(Qu等,2022),这使得教师不仅能够将边界信息传授给学生以指导息肉的分割,同时学生也能通过共享的编码器将其学到的知识传递给教师。

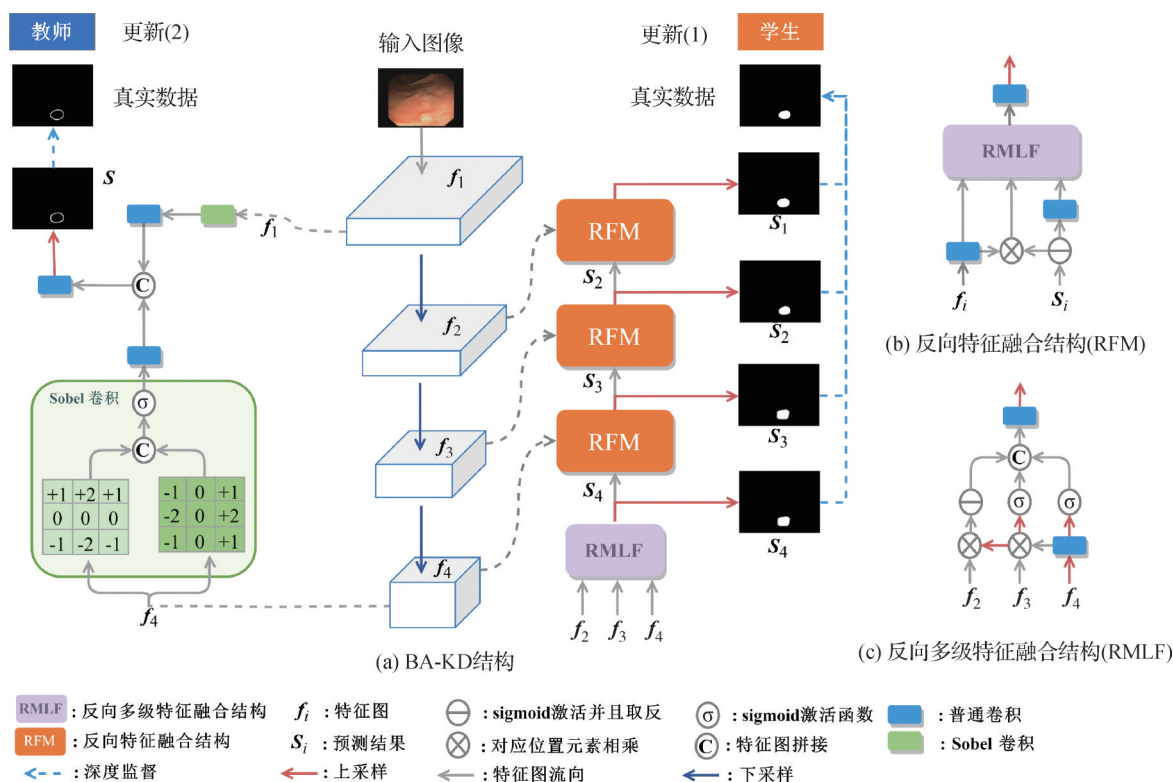


图3 BA-KD模型结构

Fig. 3 Structure of BA-KD model((a)BA-KD;(b)RFM;(c)RMLF)

2.1 特征提取编码器

BA-KD基于预训练的Res2Net50(Gao等,2021)作为骨干编码器网络构建,使用它的4个不同残差块来提取图像特征 $\{f_i, i=1, 2, 3, 4\}$,这些块按顺序用于辅助边界分割任务和主要的息肉分割任务。对于边界分割,以高级特征 f_4 和低级特征 f_1 作为输入,从中提取边界信息,并剔除了琐碎的与边界无关的信息。为了实现从息肉图像中分离出边界信息,使用Sobel算子在水平方向 G_x 和垂直方向 G_y 上提取

边界信息,以获取梯度图。对于息肉分割,首先将较为高级的特征 f_2, f_3 和 f_4 输入RMLF结构进行特征聚合,得到全局特征图,作为后续步骤的初始指导区域,这一步骤有助于粗略定位息肉区域,并提供了分割的起始点。接着,为了更精准地生成息肉分割的分割掩膜,引入了一组循环的RFM模块,用于融合区域级别和全局级别的特征之间的关系。通过这些RFM模块,实现了区域特征和全局特征之间的有机融合,使得模型能够更好地理解和捕捉息肉区域的

局部和整体特征。

2.2 教师分支解码器

教师分支是一个边界分割模型,用以将输入的息肉图像边界分割出来。首先使用编码器 Res2Net 提取图像特征 $\{f_i, i = 1, 2, 3, 4\}$,然后将低级特征 f_1 和高级特征 f_4 分别送入 Sobel 卷积,通过 Sobel 算子在水平方向 G_x 和垂直方向 G_y 上计算图像的梯度变化,进而提取出与边界相关的特征信息,并有效滤除与边界无关的区域噪声。最后,使用两个 3×3 参数固定卷积提取边界特征,并设置步幅为 1。Sobel 卷积的计算方法为

$$K_x = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, \quad K_y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (1)$$

式中, K_x 和 K_y 分别为 Sobel 卷积在水平方向和垂直方向上使用的 3×3 卷积核的参数。对输入特征图应用这两个特定卷积后,分别获得水平方向的梯度图 M_x 和垂直方向的梯度图 M_y 。接着,对得到的梯度图进行 1×1 大小的卷积操作,以确保通道数的一致性。最后,将这两个梯度图在通道维度进行拼接,并通过卷积操作得到最终的特征图。模型的输出受到地面真实边界图的监督,由此有效地消除了对象内部的边缘特征,得到了增强边缘的分割结果 S 。

教师分支的损失函数是 Dice Loss,这一损失函数通过衡量模型预测的前景与实际标签的相似度来指导训练。在训练过程中,优化模型参数以最小化 Dice Loss,有助于实现更精准的边界检测。其计算方法为

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|} \times 100\% \quad (2)$$

式中, X 和 Y 分别表示息肉分割预测结果和息肉分割真实标签的集合。

2.3 学生分支解码器

学生分支是一个息肉分割模型,由一个与教师分支共享的编码器 Res2Net、RMLF 模块,以及一系列连续的 RFM 模块。输入图像首先经过编码器进行特征提取,得到图像特征 $\{f_i, i = 1, 2, 3, 4\}$ 。接着,将高级特征 f_2 、 f_3 以及 f_4 输入 RMLF 结构进行特征聚合,生成全局特征图。由于低层特征 f_1 对预测性能贡献较低且由于其高空间分辨率需要大量计算,因此选择忽略低级特征 f_1 。

随后,全局特征图通过 RFM 逐渐与区域特征 f_4 、

f_3 和 f_2 融合,每个 RFM 的输出都会通过上采样操作以匹配区域特征的大小。通过区域和全局特征之间的持续融合,学生分支会逐渐适应息肉的大小可变性和不规则形状,从而实现精确的息肉分割。

学生分支同时使用二元交叉熵(binary cross entropy, BCE)损失和交并比(intersection over union, IoU)损失。损失函数 L 计算为

$$L = L_{BCE} + L_{IoU} \quad (3)$$

2.3.1 反向多级融合结构

反向多级融合 RMLF 模块的设计在部分解码器的基础上进行了改进(Wu 等, 2019)。该模块以高级特征 f_2 、 f_3 和 f_4 作为输入,通过 $PD = pd(f_2, f_3, f_4)$ 的计算,以获取全局映射 S_g 。随后,对 S_g 进行取反,得到除全局特征之外的细节信息。最后,将这些细节信息再次输入部分解码器 $pd(S_g, f_3, f_4)$ 进行二次聚合。这一设计有助于模型在保留全局特征的同时不丢失细微信息,进而更深入地理解图像。

2.3.2 反向融合结构

全局特征图 S_g 来自深度卷积神经网络(convolutional neural network, CNN)的最深层,但该特征图仅能粗略估计息肉的位置,难以传达息肉的详细信息。为了实现对息肉的精准分割,提出了一种渐进融合全局、区域和背景细节的策略。首先,设计了 3 个串行连接的 RFM 模块,其中 RFM 结构内嵌了一个 RMLF 模块用于特征融合。该模块接收 3 个输入,分别是上一层的分割结果(初始时采用 RMLF 输出的 S_g 作为指导信息)、局部特征 $\{f_i, i = 4, 3, 2\}$,以及反向注意力的运算结果(反向注意力能够从输入的局部特征中擦除现有估计的息肉区域,从而有序挖掘并补充额外的区域和细节信息)。然后,通过内嵌的 RMLF 对输入特征进行进一步的融合,最后通过上采样操作生成分割掩膜。随着输入特征信息的渐进融合,模型能有效纠正一些错位的预测结果。

3 实验

3.1 实验数据集

为评估所提模型 BA-KD 在结肠息肉分割任务中的效果和通用性,选择了 4 个公开的标准数据集,分别为 CVC-ClinicDB (Bernal 等, 2015)、CVC-ColonDB (Tajbakhsh 等, 2016)、Kvasir (Jha 等, 2020a) 以及 HAM10000 (Tschandl 等, 2018)。其中,前 3 个

数据集为息肉分割的基准数据集,以验证模型在息肉分割任务中的性能,最后一个数据集是皮肤镜图像分割数据集,旨在评估模型的泛化能力和可推广性。将数据集划分为 80%用于训练,10%用于验证,以及 10%用于测试。表 1 给出了实验中使用的 4 个数据集的细节信息,包括图像的数量、训练、测试和验证图像的数量。

表 1 数据集统计信息

Table 1 Statistic information of the datasets

数据集	图像数量	训练集	测试集	验证集
CVC-ClinicDB	612	490	61	61
CVC-ColonDB	380	304	38	38
Kvasir	1 000	800	100	100
HAM10000	10 015	8 012	1 001	1 002

3.2 实验环境及参数设置

本文使用 PyTorch 作为后端框架实现了所提出的模型,所有实验均在搭载 NVIDIA GeForce RTX 3090 GPU 的设备上完成。实验使用 Adam 优化器训练模型,设定初始学习率为 0.000 1,并每 50 轮进行一次 0.1 的学习率衰减。图像尺寸统一调整为 352 × 352 像素,批量大小为 16。教师分支和学生分支每经过 1 轮训练后同时更新,学生分支的更新触发后,教师分支再次进行更新。整体训练周期为 100 个 epoch。本文对每个数据集进行 5 次实验,通过计算平均值以得出最终结果。

3.3 评价指标

评价指标选择医学图像分割领域常用的度量标准,包括 Dice 相似性系数(Dice similarity coefficient, DSC)、平均交并比(mean intersection over union, mIoU)、敏感性(sensitivity, Sen)和特异性(specificity, Spe)。

3.4 不同算法的结果比较

对本文模型与 11 种最先进方法进行比较,包括 U-Net(Ronneberger 等, 2015)、Double-UNet(Jha 等, 2020b)、UNet++(Zhou 等, 2018)、TransFuse(Zhang 等, 2021)、GroupSeg(Zhang 等, 2023)、Dual encoded-decoded architecture(DualE-D)(魏天琦和肖志勇, 2022)、G-CASCADE(Rahman 和 Marculescu, 2024)、Polyp2Former(Huang 等, 2023)、RaBit(Thuan 等,

2023)、PraNet(Fan 等, 2020)和 DuAT(Tang 等, 2022)。这些算法是结肠息肉分割广泛采用的基线。具体量化结果见表 2—表 4。

表 2 显示了 BA-KD 在 CVC-ClinicDB 数据集的出色表现,尤其是在 Spe 方面达到了 0.997, DSC 为 0.955 5。BA-KD 在 DSC 和 mIoU 方面均优于所有对比模型,相比 Polyp2Former 提高了 0.3% 和 0.7%。

表 2 CVC-ClinicDB 数据集上不同算法的结果

Table 2 Results of different algorithms on the CVC-ClinicDB dataset

方法	DSC	mIoU	mSen	mSpe
U-Net	0.904	0.842	0.935	0.975
Double-UNet	0.880	0.801	0.946	0.986
UNet++	0.916	0.858	0.944	0.991
TransFuse	0.942	0.897	—	—
DuAT	0.934	0.879	0.974	0.992
PraNet	0.899	0.849	0.936	0.979
GroupSeg	0.938	0.890	—	—
DualE-D	0.942	0.894	—	—
G-CASCADE	0.947	0.902	—	—
Polyp2Former	0.953	0.910	—	—
本文	0.956	0.917	0.951	0.997

注:加粗字体表示各列最优结果,“—”表示参考文献未指出。

通过表 3 可以观察到 BA-KD 在 CVC-ColonDB 数据集上所有评估指标均超过其他模型。与表现最好

表 3 数据集上不同算法的结果

Table 3 Results of different algorithms on the CVC-ColonDB dataset

方法	DSC	mIoU	mSen	mSpe
U-Net	0.748	0.680	0.795	0.942
Double-UNet	0.746	0.657	0.837	0.957
UNet++	0.780	0.711	0.810	0.944
TransFuse	0.917	0.856	0.900	0.996
DuAT	0.850	0.759	0.940	0.988
PraNet	0.709	0.640	0.819	0.869
GroupSeg	0.831	0.747	—	—
G-CASCADE	0.826	0.746	—	—
RaBit	0.824	0.749	—	—
本文	0.931	0.875	0.956	0.997

注:加粗字体表示各列最优结果。“—”表示参考文献未指出。

的 TransFuse 相比，mIoU 和 DSC 提高 2.20% 和 1.51%。

从表 4 中可以发现，对于 Kvasir 数据集，BA-KD 实现了 0.889 的 mIoU 和 0.937 的 DSC，在效果上超过了最佳的 RaBit(Thuan 等,2023)约 1.08%

不同网络模型的分割对比结果如图 4 所示，边界分割结果的可视化如图 5 所示。

图 4 为 3 个息肉数据集在不同模型下的分割结果。可以发现尽管息肉的大小、颜色和形状存在差异，本文模型都能够精确地进行识别和分割。同样，即使息肉的颜色与周围组织相近或存在明显的色差，该模型也能通过学习到的特征来进行有效区分。

在图 5 中，为了直观比较不同分割方法的性能，将分割结果与原始息肉图像的真实标签进行叠加，展示了息肉边界与周围组织之间界限模糊的情况。大多数模型在这种情况下都难以划定一个清晰的边界。然而，本文模型在分割息肉轮廓方面表现出色，相较于其他方法，能够更加精确地捕捉和界定病变

表 4 Kvasir 数据集上不同算法的结果
Table 4 Results of different algorithms on the Kvasir dataset

方法	DSC	mIoU	mSen	mSpe
U-Net	0.777	0.681	0.825	0.949
Double-UNet	0.843	0.760	0.872	0.977
UNet++	0.851	0.769	0.875	0.978
TransFuse	0.920	0.870	—	—
DuAT	0.924	0.876	—	—
PraNet	0.898	0.840	0.915	0.948
GroupSeg	0.928	0.879	—	—
G-CASCADE	0.927	0.880	—	—
RaBit	0.927	0.879	—	—
本文	0.937	0.890	0.947	0.989

注：加粗字体表示各列最优结果。“—”表示参考文献未指出。

区域的边界。通过在 3 个数据集上的实验验证，本文提出的 BA-KD 展现出了显著的性能优势。

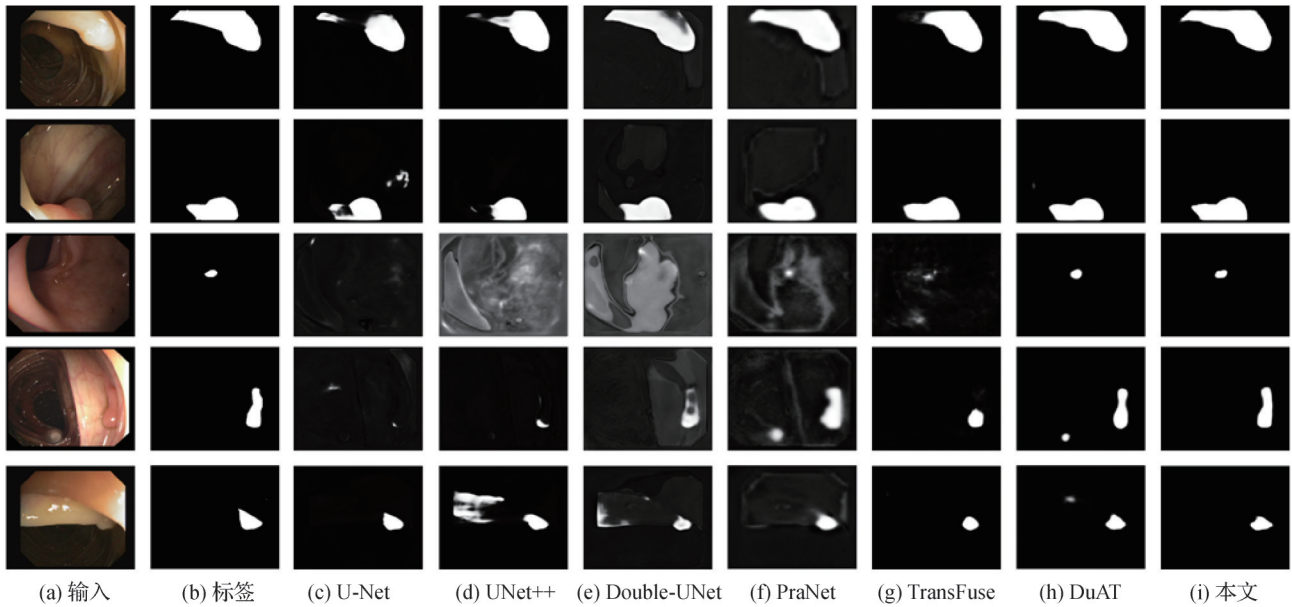


图 4 分割结果可视化

Fig. 4 Visualization of segmentation results

((a) inputs; (b) ground truths; (c) U-Net; (d) UNet++; (e) Double-UNet; (f) PraNet; (g) TransFuse; (h) DuAT; (i) ours)

BA-KD 通过教师端提取精确的边界信息，而学生端则专注于对息肉区域进行整体分割。本文采用的交替更新参数策略实现两者之间的有效互动与约束。这一策略使得教师端的参数优化能够引导学生端更精确地分割息肉边界，同时学生端也能反过来促进教师端对息肉边界的精准定位。通过这种相互

制约与促进的机制，BA-KD 模型在各项评估指标上均展现出卓越的性能，验证了该模型在息肉图像分割任务中的有效性。

3.5 消融实验

为了验证本文方法的有效性，在两个数据集上比较了 BA-KD 及其 3 种变体，对比结果见表 5。表

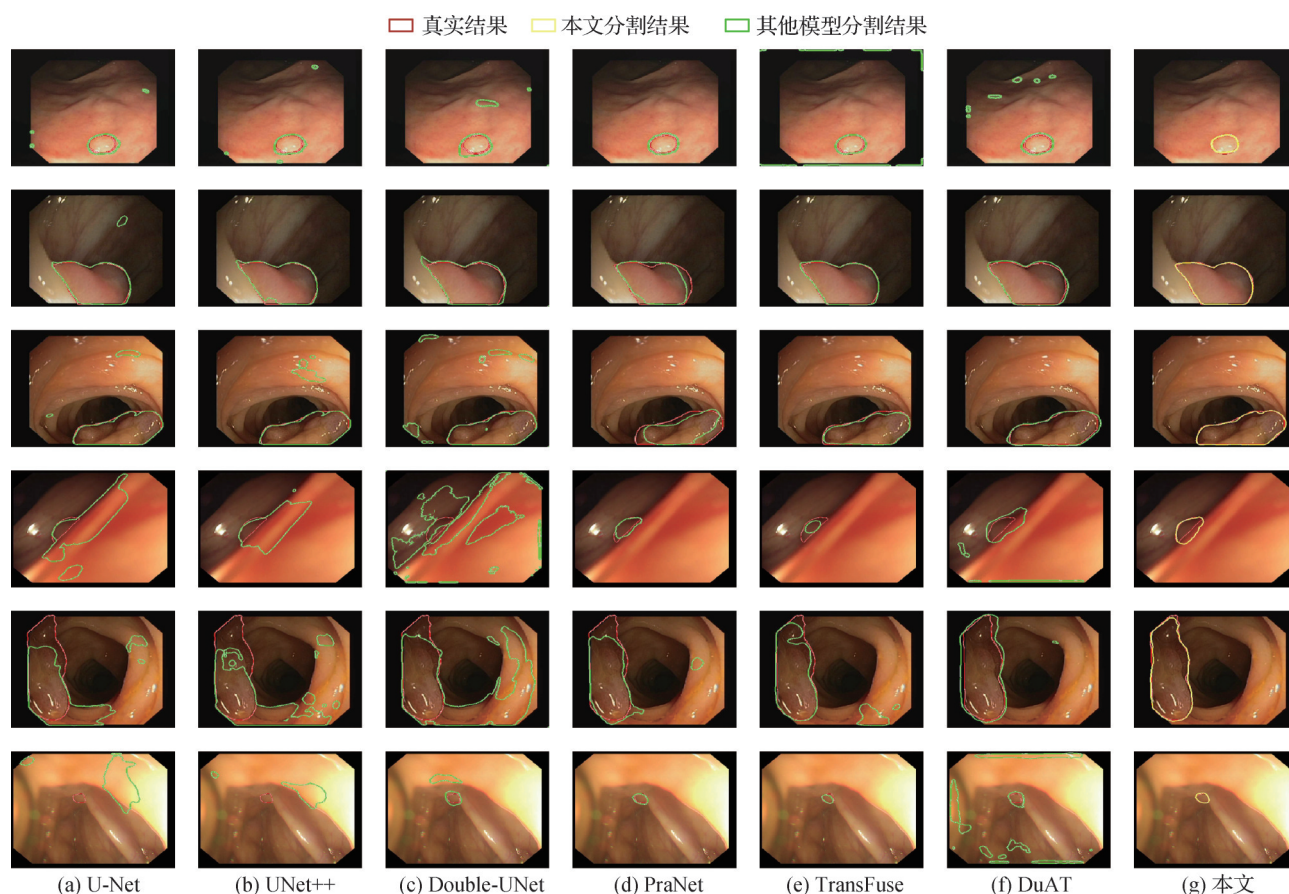


图5 边界分割结果可视化

Fig. 5 Visualization of boundary segmentation results

((a) U-Net; (b) UNet++; (c) Double-UNet; (d) PraNet; (e) TransFuse; (f) DuAT; (g) ours)

中,“S”表示学生分支,“T”表示教师分支,“T(1)”表示同时更新教师和学生分支的权重,“T(2)”表示在更新教师分支权重之前更新学生分支的权重。实验结果表明,边界自知识蒸馏方法显著提高了网络模型的效果。此外,两个分支中权重的更新策略也对性能产生影响。从表5还可看出,当学生分支的权重先于教师分支的权重更新时,可以获得最佳结果。

表5 CVC-ColonDB数据集上的消融实验

Table 5 Ablation experiments on the CVC-ColonDB dataset

方法	DSC	mIoU	mSen	mSpe
Backbone+RMLF	0.885	0.812	0.975	0.991
Backbone+S	0.917	0.864	0.945	0.993
Backbone+S+T(1)	0.921	0.862	0.936	0.996
Backbone+S+T(2)	0.931	0.875	0.956	0.997

注:加粗字体表示各列最优结果。

为了验证通过边界自知识蒸馏的方式准确划分息肉边界的效果,使用 Grad-CAM (Selvaraju 等, 2017)对提出模型进行了视觉分析。具体来说,本文选择息肉与正常组织相似的图像,并利用 Grad-CAM 可视化技术,验证模型对息肉图像中相关区域的关注。从图6可以看出,当模型不使用学生分支解码器时,注意力集中在息肉及正常组织区域,而在两者之间缺乏有效的区分。与之形成鲜明对比的是,当使用学生分支解码器时,模型显示对息肉边界信息的敏锐关注。通过对分割结果的观察,可以看出本文提出方法使模型得到更好的边界分割效果。

消融实验的可视化如图7所示,当仅使用学生分支训练模型时,模型难以准确确定息肉的范围和边界信息。然而,当使用学生和教师两个分支训练模型时,模型能够精确定位息肉。

3.6 泛化性实验

本文进一步评估了所提出模型在其他医疗分割任务(HAM10000,其中包含来自不同人群的皮肤镜

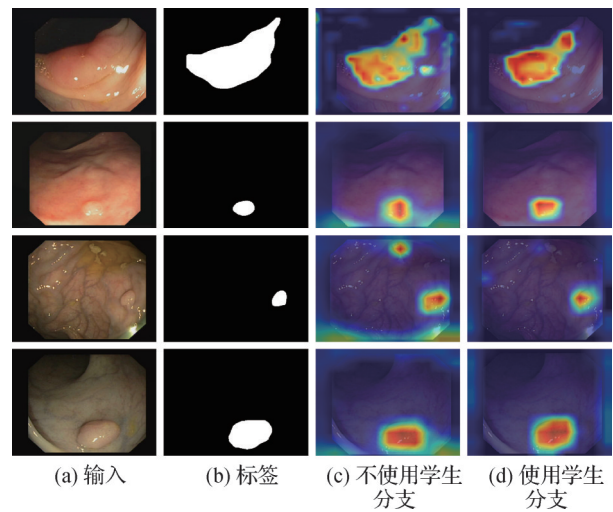


图6 特征可视化热力图

Fig. 6 Feature visualization heat map ((a) inputs; (b) ground truths; (c) without students branch; (d) with students branch)

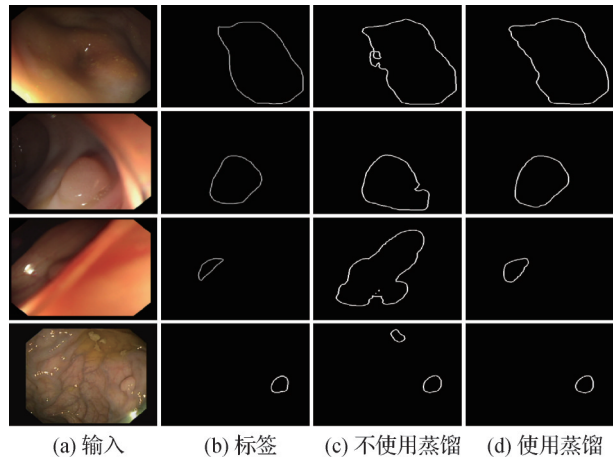


图7 消融实验的可视化结果

Fig. 7 Visual result of ablation study ((a) input; (b) ground truth; (c) without distillation; (d) with distillation)

图像)上的泛化能力。表6给出了本文模型与现有医疗分割基线模型的对比结果。可以看出,本文模型在HAM10000上的所有指标都表现出色。此外,BA-KD在平均DSC值(0.956 2)和mIoU(0.922 3)上取得了显著得分,与表现最好的DSCAU-Net相比分别显著提高了1.45%和2.25%。

4 结 论

本研究的主要目的是开发和验证一种新的边界知识蒸馏架构,以提高结肠息肉图像分割的准确性和效率。通过结合边界分割网络和息肉分割网络,提出了一个端到端的解决方案,更有效地处理息肉

表6 HAM10000数据集上不同算法的结果

Table 6 Results of different algorithms on the HAM10000 dataset

方法	DSC	mIoU	mSen	mSpe
U-Net	0.918	0.866	0.954	0.957
pspnet	0.943	0.902	0.955	0.975
Double-UNet	0.942	0.902	0.959	0.973
UNet++	0.941	0.899	0.955	0.975
EGEunet	0.863	0.778	0.909	0.944
TransFuse	0.939	0.896	0.960	0.972
DCSAUnet	0.933	0.887	0.949	0.972
本文	0.956	0.922	0.961 8	0.985

注:加粗字体表示各列最优结果。

的多样性和边界模糊问题。通过知识蒸馏,BA-KD利用教师网络的边界分割结果,约束学生网络的息肉边界分割。此外,本文提出的反向融合模块(RFM)能够整合全局和局部特征,进一步提升模型提取息肉特征的能力。在4个公开数据集上的实验结果显示,BA-KD在DSC和mIoU指标上均取得了显著提升。与现有的先进方法相比,BA-KD在所有评估指标上表现出更高的性能,在DSC和mIoU指标上分别提升了0.45%和0.68%。

尽管BA-KD表现出色,但仍存在一些不足。由于知识蒸馏过程中需要额外的计算资源,模型的训练时间较长。此外,对于一些极端复杂的息肉形态,模型的边界提取仍有进一步提升的空间。这些问题可能源于当前特征提取模块在处理极端复杂特征时的局限性。未来的研究方向包括:优化训练时间,研究更加高效的知识蒸馏策略,以减少训练时间;提升边界提取能力,进一步改进反向融合模块,以提高模型在极端复杂息肉边界上的分割能力;扩展应用范围,将BA-KD应用于其他类型的医学图像分割任务,验证其通用性和扩展性。

参考文献(References)

Ahn S B, Han D S, Bae J H, Byun T J, Kim J P and Eun C S. 2012. The miss rate for colorectal adenoma determined by quality-adjusted, back-to-back colonoscopies. Gut and Liver, 6(1): 64-70 [DOI: 10.5009/gnl.2012.6.1.64]

Bernal J, Sánchez F J, Fernández-Esparrach G, Gil D, Rodríguez C

- and Vilariño F. 2015. WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, 43: 99-111 [DOI: 10.1016/j.compmedimag.2015.02.007]
- Caron M, Touvron H, Misra I, Jegou H, Mairal J, Bojanowski P and Joulin A. 2021. Emerging properties in self-supervised vision transformers//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: IEEE: 9630-9640 [DOI: 10.1109/ICCV48922.2021.00951]
- Chung I, Park S U, Kim J and Kwak N. 2020. Feature-map-level online adversarial knowledge distillation//*Proceedings of 2020 International Conference on Machine Learning*. PMLR, 2020: 2006-2015
- Du X Q, Xu X B and Ma K P. 2022. ICGNet: integration context-based reverse-contour guidance network for polyp segmentation//*Proceedings of the 31st International Joint Conference on Artificial Intelligence*. Vienna, Austria: IJCAI: 877-883 [DOI: 10.24963/ijcai.2022/123]
- Fan D P, Ji G P, Zhou T, Chen G, Fu H Z, Shen J B and Shao L. 2020. PraNet: parallel reverse attention network for polyp segmentation//*Proceedings of the 23rd International Conference on Medical Image Computing and Computer-Assisted Intervention*. Lima, Peru: Springer: 263-273 [DOI: 10.1007/978-3-030-59725-2_26]
- Favoriti P, Carbone G, Greco M, Pirozzi F, Pirozzi R E M and Corcione F. 2016. Worldwide burden of colorectal cancer: a review. *Updates in Surgery*, 68(1): 7-11 [DOI: 10.1007/s13304-016-0359-y]
- Gao S H, Cheng M M, Zhao K, Zhang X Y, Yang M H and Torr P. 2021. Res2Net: a new multi-scale backbone architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(2): 652-662 [DOI: 10.1109/tpami.2019.2938758]
- Hinton G, Vinyals O and Dean J. 2015. Distilling the knowledge in a neural network [EB/OL]. [2024-03-29]. <https://arxiv.org/pdf/1503.02531.pdf>
- Huang X S, Huang J Z, Wang S, Wei Y G, An D and Liu J C. 2023. Polyp2Former: boundary guided network based on transformer for polyp segmentation//*Proceedings of 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Istanbul, Turkey: IEEE: 1971-1976 [DOI: 10.1109/BIBM58861.2023.10385257]
- Jha D, Smedsrud P H, Riegler M A, Halvorsen P, De Lange T, Johansen D and Johansen H D. 2020a. Kvasir-SEG: a segmented polyp dataset//*Proceedings of the 26th International Conference on Multimedia Modeling*. Daejeon, Korea (South): Springer: 451-462 [DOI: 10.1007/978-3-030-37734-2_37]
- Jha D, Riegler M A, Johansen D, Halvorsen P and Johansen H D. 2020b. 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, USA: IEEE: 558-564 [DOI: 10.1109/CBMS49503.2020.00111]
- Kanopoulos N, Vasanthavada N and Baker R L. 1988. Design of an image edge detection filter using the Sobel operator. *IEEE Journal of Solid-State Circuits*, 23(2): 358-367 [DOI: 10.1109/4.996]
- Kim T, Lee H and Kim D. 2021. UACANet: uncertainty augmented context attention for polyp segmentation//*Proceedings of the 29th ACM international conference on multimedia*: 2167-2175. [DOI: 10.1145/3474085.3475375]
- Liang L M, He A J, Zhu C K and Sheng X Q. 2023. Colorectal polyp segmentation method based on fusion of transformer and cross-level phase awareness. *Journal of Biomedical Engineering*, 40(2): 234-243 (梁礼明, 何安军, 朱晨锟, 盛校棋. 2023. 融合Transformer和跨级相位感知的结肠息肉分割方法. *生物医学工程学杂志*, 40(2): 234-243) [DOI: 10.7507/1001-5515.202211067]
- Lin Y, Zhang D, Fang X, Chen Y F, Cheng K T and Chen H. 2023. Rethinking boundary detection in deep learning models for medical image segmentation//*Proceedings of the 28th International Conference on Information Processing in Medical Imaging*. San Carlos de Bariloche, Argentina: Springer: 730-742 [DOI: 10.1007/978-3-031-34048-2_56]
- Park W, Kim D, Lu Y and Cho M. 2019. Relational knowledge distillation//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 3962-3971 [DOI: 10.1109/CVPR.2019.00409]
- Peng B Y, Jin X, Li D S, Zhou S F, Wu Y C, Liu J H, Zhang Z N and Liu Y. 2019. Correlation congruence for knowledge distillation//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 5006-5015 [DOI: 10.1109/ICCV.2019.00511]
- Qu L H, Luo X Y, Wang M N and Song Z J. 2022. Bi-directional weakly supervised knowledge distillation for whole slide image classification//*Proceedings of the 36th International Conference on Neural Information Processing Systems*. New Orleans, USA: Curran Associates Inc.: #1118
- Rahman M M, Marculescu R. 2024. G-CASCADE: efficient cascaded graph convolutional decoding for 2D medical image segmentation//*Proceedings of 2024 IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, USA: IEEE: 7713-7722 [DOI: 10.1109/WACV57701.2024.00755]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Sanderson E and Matuszewski B J. 2022. FCN-transformer feature fusion for polyp segmentation//*Proceedings of the 26th Annual Conference on Medical Image Understanding and Analysis*. Cambridge, UK:

- Springer: 892-907 [DOI: 10.1007/978-3-031-12053-4_65]
- Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 618-626 [DOI: 10.1109/ICCV.2017.74]
- Shen Y T, Jia X and Meng M Q H. 2021. HRENet: a hard region enhancement network for polyp segmentation//Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Strasbourg, France: Springer: 559-568 [DOI: 10.1007/978-3-030-87193-2_53]
- Siegel R L, Miller K D and Jemal A. 2019. Cancer statistics, 2019. CA: A Cancer Journal for Clinicians, 69(1): 7-34 [DOI: 10.3322/caac.21551]
- Sung H, Ferlay J, Siegel R L, Laversanne M, Soerjomataram I, Jemal A and Bray F. 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3): 209-249 [DOI: 10.3322/caac.21660]
- Tajbakhsh N, Gurudu S R and Liang J M. 2016. Automated polyp detection in colonoscopy videos using shape and context information. IEEE Transactions on Medical Imaging, 35(2): 630-644 [DOI: 10.1109/tmi.2015.2487997]
- Tang F L, Huang Q M, Wang J F, Hou X X, Su J L and Liu J X. 2022. DuAT: dual-aggregation transformer network for medical image segmentation [EB/OL]. [2024-03-29]. <https://arxiv.org/pdf/2212.11677.pdf>
- Thuan N H, Oanh N T, Thuy N T, Perry S and Sang D V. 2023. RaBit: an efficient transformer using bidirectional feature pyramid network with reverse attention for colon polyp segmentation [EB/OL]. [2024-03-29]. <https://arxiv.org/pdf/2307.06420.pdf>
- Tian C X and Zhao L. 2021. Epidemiological characteristics of colorectal cancer and colorectal liver metastasis. Chinese Journal of Cancer Prevention and Treatment, 28(13): 1033-1038 (田传鑫, 赵磊. 2021. 结直肠癌及结直肠癌肝转移流行病学特点. 中华肿瘤防治杂志, 28(13): 1033-1038) [DOI: 10.16073/j.cnki.cjcp.2021.13.12]
- Tschandl P, Rosendahl C and Kittler H. 2018. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data, 5(1): #180161 [DOI: 10.1038/sdata.2018.161]
- Wei T Q and Xiao Z Y. 2022. Dual encoded-decoded polyp segmentation method for gastroscopic images architecture. Journal of Image and Graphics, 27(12): 3637-3650 (魏天琦, 肖志勇. 2022. 双重编解码架构的肠胃镜图像息肉分割. 中国图象图形学报, 27(12): 3637-3650) [DOI: 10.11834/jig.210966]
- Wu Z, Su L and Huang Q M. 2019. Cascaded partial decoder for fast and accurate salient object detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Beach, USA: IEEE: 3902-3911 [DOI: 10.1109/CVPR.2019.00403]
- Zhang M W, Xia H Y and Tan Y M. 2023. GroupSeg: an efficient grouping transformer network for polyp segmentation//Proceedings of 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Istanbul, Türkiye: IEEE: 2380-2384 [DOI: 10.1109/BIBM58861.2023.10385401]
- Zhang R F, Li G B, Li Z, Cui S G, Qian D H and Yu Y Z. 2020. Adaptive context selection for polyp segmentation//Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention. Lima, Peru: Springer: 253-262 [DOI: 10.1007/978-3-030-59725-2_25]
- Zhang Y D, Liu H Y and Hu Q. 2021. TransFuse: fusing transformers and CNNs for medical image segmentation//Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Strasbourg, France: Springer: 14-24 [DOI: 10.1007/978-3-030-87193-2_2]
- Zhao X Q, Zhang L H and Lu H C. 2021. Automatic polyp segmentation via multi-scale subtraction network//Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Strasbourg, France: Springer: 120-130 [DOI: 10.1007/978-3-030-87193-2_12]
- Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2018. UNet++: a nested U-Net architecture for medical image segmentation//Proceedings of the 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018 on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Granada, Spain: Springer: 3-11 [DOI: 10.1007/978-3-030-00889-5_1]

作者简介

- 孟祥福,男,教授,主要研究方向为空间数据管理、医学影像分析和时序数据库系统。E-mail:marxi@126.com
- 张智超,男,硕士研究生,主要研究方向为医学图像处理。E-mail:1353925768@qq.com
- 俞纯林,男,硕士研究生,主要研究方向为医学图像处理。E-mail:yuchunlin627@126.com
- 张霄雁,女,副教授,主要研究方向为时空数据管理、推荐系统和人工智能。E-mail:zhangxiaoyan@lntu.edu.cn