

RDG-Net:基于双阶段解码器的结直肠息肉图像分割模型

谭山湖¹, 郭小燕¹, 魏伟一²

1. 甘肃农业大学信息科学技术学院, 甘肃 兰州 730070; 2. 西北师范大学计算机科学与工程学院, 甘肃 兰州 730070

【摘要】基于深度学习的息肉图像分割可以有效帮助医生评估癌前病变,本文针对结直肠息肉图像中息肉边界不清晰时分割效果不佳、对新样本泛化能力不足的问题,提出一种基于双阶段解码器的结直肠息肉图像分割模型RDG-Net。该模型采用Res2Net-50作为编码器以提高图像分割精度。解码器分为两个阶段,第一阶段利用4层多尺度特征聚合模块整合不同阶段编码器提取的特征,第二阶段通过3层并行卷积融合模块增强解码器第一阶段输出的图像特征并解码至更高分辨率作为模型的最终输出结果。采用CVC-ClinicDB和Kvasir-SEG数据集的训练集数据进行模型训练,并采用CVC-ClinicDB与Kvasir-SEG数据集以及未参与训练的CVC300和ETIS-LaribPolypDB数据集分别对模型进行测试。测试结果显示,CVC-ClinicDB与Kvasir-SEG数据集上准确率、精确度、召回率、Dice系数、交并比和F2的平均值分别为98.41%、94.25%、92.62%、93.42%、87.69%、92.93%,CVC300和ETIS-LaribPolypDB数据集上各评价指标的平均结果分别为99.05%、87.79%、89.13%、88.39%、79.33%、88.82%。实验结果表明RDG-Net模型在结直肠息肉区域的分割任务中表现出色,在新数据集上表现出较好的泛化能力。

【关键词】图像分割;结直肠息肉;多尺度特征聚合;并行卷积

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2025)01-0052-07

RDG-Net: a colorectal polyp image segmentation model based on a dual-stage decoder

TAN Shanhu¹, GUO Xiaoyan¹, WEI Weiyi²

1. College of Information Science and Technology, Gansu Agricultural University, Lanzhou 730070, China; 2. College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China

Abstract: Deep learning-based polyp image segmentation is helpful for assessing precancerous lesions. A colorectal polyp image segmentation model (RDG-Net) based on a dual-stage decoder is proposed to addresses the issues of poor segmentation performance due to unclear boundaries in colorectal polyp images and insufficient generalization ability for new samples. The model uses Res2Net-50 as the encoder to enhance the precision of image segmentation. The decoder has two stages: the first stage utilizes a 4-layer multi-scale feature aggregation module to integrate features extracted by the encoder at different stages, while the second stage enhances the image features output by the first stage of the decoder through a 3-layer parallel convolution fusion module and decodes them to a higher resolution as the model's final output. The model is trained using the CVC-ClinicDB and Kvasir-SEG training datasets, and tested using the CVC-ClinicDB and Kvasir-SEG test datasets, as well as the CVC300 and ETIS-LaribPolypDB datasets that are not involved in the training. The test results show that the proposed method has an average accuracy, precision, recall rate, Dice similarity coefficient, intersection over union and F2 score of 98.41%, 94.25%, 92.62%, 93.42%, 87.69% and 92.93% on the CVC-ClinicDB and Kvasir-SEG datasets, while 99.05%, 87.79%, 89.13%, 88.39%, 79.33% and 88.82% on the CVC300 and ETIS-LaribPolypDB datasets, respectively, demonstrating that RDG-Net model performs well in colorectal polyp region segmentation and has a high generalization performance on new datasets.

Keywords: image segmentation; colorectal polyp; multi-scale feature aggregation; parallel convolution

【收稿日期】2024-09-18

【基金项目】国家自然科学基金(62363031)

【作者简介】谭山湖,硕士研究生,研究方向:深度学习、图像处理,E-mail: 964913737@qq.com

【通信作者】郭小燕,博士,教授,研究方向:深度学习、图像处理,E-mail: guoxy@gsau.edu.cn

前言

现代医学影像分析中,结直肠息肉的准确检测和分割对于早期诊断治疗结直肠癌至关重要。结直肠癌是全球最常见的癌症之一,对人类健康构成严重威胁^[1-2]。这种疾病通常始于大肠内壁的息肉,随着时间推移可能发展为恶性肿瘤。结肠镜检查是发

现结肠息肉最可靠的诊断方法,及早发现并手术切除息肉,可以有效预防其发生癌变。息肉的形状多样、大小不一、颜色丰富,这为自动化检测与图像分割技术带来了挑战^[3-4]。通过图像分割技术,医生能从复杂的医学影像中精确识别出息肉的形状、大小和具体位置,这对于准确评估息肉的性质(良性或恶性)极为关键。传统的图像处理技术由于缺乏对复杂背景和息肉形态差异的适应性识别,难以达到临床应用的精度要求。随着深度学习技术在医疗领域的应用,为医学图像分析提供了新的解决方案^[5]。卷积神经网络(CNN)以其强大的图像特征提取能力被研究者们广泛用于医学图像分割任务中。U-Net模型通过编码器-解码器结构和跳跃连接策略,为后续分割模型的设计提供了灵感,已在多个医学图像分割任务中显示出卓越性能^[6-8]。CE-Net结合了预处理模块、特征提取模块和后处理模块,通过引入残差模块、空间金字塔池化等机制,有效提升了医学图像分割的准确性和鲁棒性^[9]。VMN采用交互式三维医学图像分割技术,利用外部记忆组件存储历史分割信息,有效处理用户交互,显著提高了分割精度^[10]。DSCA-Net采用双阶段编码注意力机制,通过特征融合和多尺度特征提取,显著提升了核分割的精度^[11]。UNETR++^[12]是对UNETR^[13]的改进,采用一种高效配对注意力(EPA)模块,结合了空间和通道注意力,在多个基准数据集上表现出色。

综合以上研究发现,深度学习技术在医学图像检测与医学影像分割任务中已取得了一定的研究成果,然而面临形状和大小多变的息肉分割任务,依然存在分割精度不足、息肉边缘分割不清晰的问题。在最近的研究中,有研究者针对以上问题提出了相应的改进策略,ACSNet^[14]使用基于上下文的编码器框架,自适应地关注相关特征以实现更精确的分割,这种方法利用局部上下文注意力(Local Context Attention, LCA)、全局上下文模块(Global Context Module, GCM)和自适应选择模块(Adaptive Selection Module, ASM)来处理息肉外观的变化,显著提高了分割精度,尤其是在处理形状和大小多变的息肉时表现出色,但该模型的泛化能力欠佳。MSNet^[15]凭借其创新的减法单元和多尺度特征融合策略,在结直肠息肉分割任务中展现了强大的潜力和实际应用价值。然而,其泛化能力较弱和模型参数量较高可能限制其在某些应用场景中的广泛使用。PraNet^[16]采用一种反向注意力机制,通过擦除前景对象的方式与不同层级的特征进行融合以捕获其他模型经常遗漏的关键边界信息,在精度、实时性方面表现出色,但同样有着较高的计算成本且对未见

样本分割能力不足。CFA-Net^[17]提出了一种基于跨层特征聚合网络的结直肠息肉分割方法,通过边界预测网络和两流结构分割网络,结合跨层特征融合模块和边界聚合模块有效提升了分割精度,但其泛化能力依然不足。以上模型虽从不同程度提高率结肠分割精度,但大多数模型对特定样本具有依赖性,对于差异较大的息肉样本识别效果差,因此普遍存在泛化能力弱的问题。

针对以上问题本文提出一种基于双阶段解码器模型(Recurrent Dual-Stage Decoder Network, RDG-Net),通过多尺度特征聚合和并行卷积融合模块有效提高直肠息肉特征提取能力以及模型的泛化能力。RDG-Net与其他模型相比具有较强的分割性能,尤其在新的数据集上表现优越,实验表明这种方法在保证识别效果的同时增强了模型的泛化能力。

1 方法

本文提出的结直肠息肉图像分割模型(RDG-Net)采用编码器-解码器结构设计,如图1所示,采用Res2Net-50^[18]作为编码器的骨干网络,利用多尺度特征融合机制有效提高信息传递效率、减少信息丢失。解码器包括两个阶段:第一阶段为四层多尺度特征聚合模块用于融合不同阶段的上下文信息。该模块采用深度可分离卷积(Depthwise Separable Convolution, DSC)代替传统卷积以减少计算量^[19];引入多层感受野模块(RFB)和全局注意力机制(GAM)聚焦于息肉区域提高关键特征的提取能力^[20-21]。第二阶段为3层并行卷积融合模块,每一层包括4个并行子路径,分别模拟不同尺寸的卷积核以增强模型对不同大小息肉的识别能力。这4条路径的输出按位相加并进行融合,以综合各自特征,强化模型的表达能力。

1.1 编码器 Res2Net-50

编码器采用Res2Net-50,主要包括5个级别(level),如图1中Res2Net Block部分所示。Res2Net^[18]是对ResNet^[22]的扩展,将ResNet结构中 3×3 卷积替换为多尺度融合结构,通过分割通道、通道卷积、通道融合以达到不同尺度卷积的效果,通道分割及卷积过程如下所示:

$$y_i = \begin{cases} x_i & i = 1 \\ k_i(x_i + y_{i-1}) & 2 \leq i \leq 4 \end{cases} \quad (1)$$

其中, $x_i(i \in \{1, 2, 3, 4\})$ 代表分割的4个通道, k 为 3×3 的卷积操作, $y_i(i \in \{1, 2, 3, 4\})$ 为每个通道经过卷积和通道融合后的结果。为了有效融合各尺度特征,将各通道的卷积结果通过 1×1 的卷积操作和通道级拼接组合(cat)起来,如式(2)所示:

$$\text{output} = \text{Conv}(\text{cat}(y_1, y_2, y_3, y_4)) \quad (2)$$

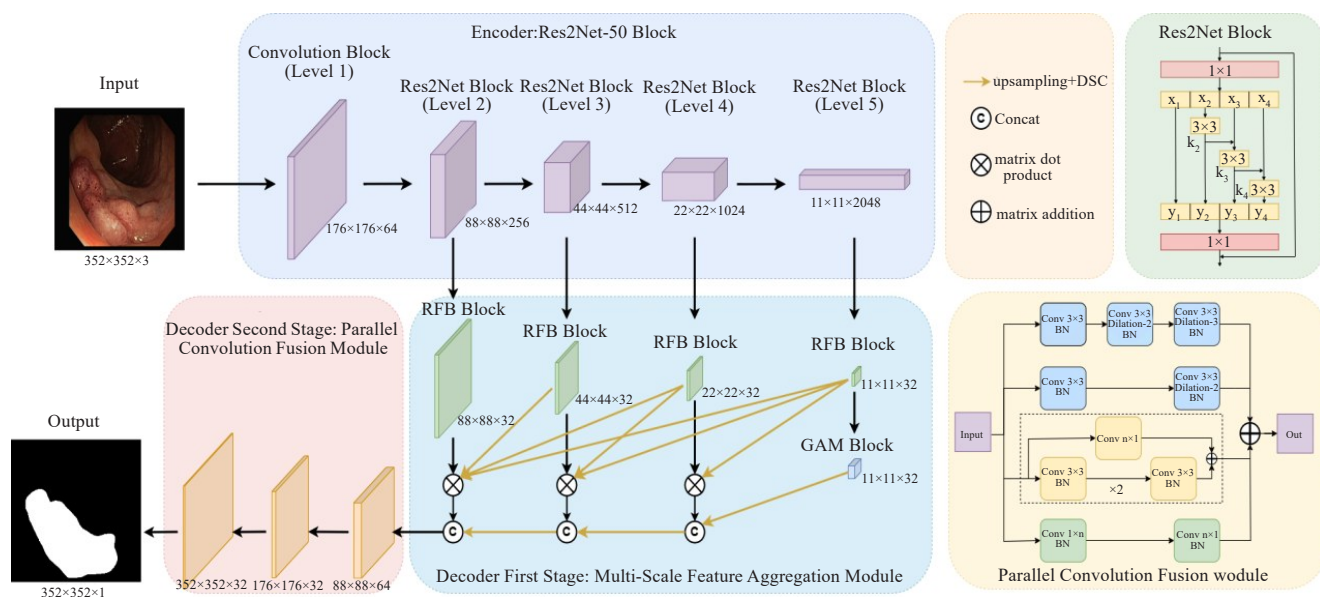


图1 本文模型 RDG-Net
Figure 1 RDG-Net model

其中, output 为 Res2Net 的模块输出结果, Conv 为 1×1 的卷积, 这种在单一模块内部进行的细粒度尺度融合增强了网络在同一层次上处理多尺度特征的能力, 有利于捕捉小息肉的细微特征, 提升对关键特征的识别和提取能力。

1.2 解码器

解码器分为两个阶段: 多尺度特征聚合 (Multi-Scale Feature Aggregation, MFA) 模块和并行卷积融合 (Parallel Convolution Fusion Module, PCF) 模块。

1.2.1 MFA 为了减轻模型训练时梯度消失问题, 增强模型信息传播能力, 提升训练速度, 结合多层感受野 (Receptive Field Block, RFB) 模块、全局注意力机制 (Global Attention Mechanism, GAM)、深度可分离卷积和密集连接结构^[23], 将编码器的 level2-level5 的输出与 RFB 各模块级联, RFB 模块通过不同大小的卷积核和多尺度卷积组合来模拟视网膜中感受野的结构^[20,24], 使模块能够在不同的尺度上感知图像的细节和上下文信息, 关注重点区域, 进而更准确地识别和分割息肉。第五级编码器模块在 RFB 模块后加入一个 GAM 模块。GAM 通过加权全局的特征响应来实现, 能够提升模型对关键特征的敏感性和区分能力, 以增强对息肉特征的关注, 提升图像处理能力。该模块采用密集连接结构, 通过特征逐元素相乘并与上一级特征在通道维度上拼接以达到特征聚合的效果。特征聚合后通过深度可分离卷积处理, 然后进入下一级的特征聚合。这种机制通过整合图像的颜色、纹理和边缘信息, 增强了模型对息肉图像结构和内容的识别能力。深度可分离卷积则是有效减少

了模型的参数量和计算成本, 并且能够分步骤提取空间特征和处理通道间的关系。MFA 模块的具体实现过程如公式 (3) 所示:

$$\begin{cases} \text{Mout}_i = \text{DSC}(\text{cat}(\prod_{i=1}^5 x_i, \text{Mout}_{i+1})), i \in [2, 4] \\ \text{Mout}_i = \text{GAM}(x_i), i = 5 \end{cases} \quad (3)$$

其中, Mout_i 为第 i 个多尺度特征聚合模块的输出; DSC 为深度可分离卷积; cat 为通道级拼接; x_i 为编码器模块的输出经过第 i 个 RFB 模块处理的结果, GAM 为全局注意力模块。

1.2.2 PCF 在 MFA 模块完成后特征图尺寸被解码至 88×88 , 随后送入 PCF 模块解码至与原输入图像相同的尺寸。该模块包含 4 个并行子路径。第一和第二子路径分别采用不同参数的扩张卷积, 其中第一条子路径模拟 15×15 大小的核, 第二条子路径模拟 7×7 大小的核。这种设计能扩大感受野, 更好地捕捉图像的上下文信息, 进而提升对图像中大尺寸结构的识别能力。第三条子路径采用残差结构的卷积模块增强特征传递, 减少训练过程中的信息损失, 提高网络的深度和稳定性。最后一条子路径采用 $1 \times n$ 和 $n \times 1$ 的卷积核序列以捕捉图像中的细长结构。4 条子路径的输出进行按位相加处理以实现多尺度特征的有效融合, 具体公式实现如下所示:

$$\begin{cases} Z_1 = \text{CD}_3(\text{CD}_2(C_3(x))) \\ Z_2 = \text{CD}_2(C_3(x)) \\ v_3 = C_3(C_3(x)) + C_1(x) \\ Z_3 = v_3(v_3(x)) \\ Z_4 = C_{n \times 1}(C_{1 \times n}(x)) \end{cases} \quad (4)$$

$$\text{out} = Z_1 + Z_2 + Z_3 + Z_4 \quad (5)$$

其中, $Z_i(i \in \{1, 2, 3, 4\})$ 表示 PCF 模块中的 4 个子路径输出的结果; CD_2 表示卷积核为 3, 扩展系数为 2 的扩张卷积; CD_3 表示卷积核为 3, 扩展系数为 3 的扩张卷积; C_3 表示卷积核为 3 的卷积; C_1 表示卷积核为 1 的卷积; $C_{n \times 1}$ 和 $C_{1 \times n}$ 分别表示 $n \times 1$ 和 $1 \times n$ 的卷积核序列, out 表示 PCF 模块的输出。

这种设计的优势在于能深入解析第一阶段输出的图像特征,并在第二阶段的解码过程中实现特征的有效互补,从而确保在解码过程中不丢失关键的边界信息,不仅优化了特征传递的效率,还增强了模型对复杂图像结构的识别能力。

2 实验结果与分析

2.1 实验数据集

为验证模型对于息肉图像分割的有效性,实验采用 Kvasir-SEG^[25]、CVC-ClinicDB^[26]、ETIS-LaribPolypDB^[27]和 CVC-300^[28] 4 个公共结直肠息肉图像数据集,分别由 1000、612、196、60 张结直肠息肉图像和分割掩码组成。本实验取 Kvasir-SEG 中的 900 张图像数据和 CVC-ClinicDB 中的 550 张图像数据,取 90% 作为模型的训练数据,10% 作为验证数据,将 Kvasir-SEG 剩余的 100 张图像数据、CVC-ClinicDB 剩余的 62 张图像数据与从未参与训练的 ETIS-LaribPolypDB 和 CVC-300 数据集分别进行测试。

2.2 评价指标

本文实验采用的评价指标为 Dice 相似系数(DSC)、交并比(Intersection over Union, IoU)、准确率(Accuracy)、精确度(Precision)、召回率(Recall)和 F2 分数,其计算公式分别表示为:

$$\text{DSC} = \frac{2 \cdot |A \cap B|}{|A| + |B|} \quad (6)$$

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (7)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (8)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (9)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (10)$$

$$\text{F2} = \frac{5 \cdot \text{Precision} \cdot \text{Recall}}{4 \cdot \text{Precision} + \text{Recall}} \quad (11)$$

2.3 实验环境及参数设置

本实验的模型基于 pytorch 框架实现,硬件配置包括 Intel(R) Xeon(R) Gold 5218 CPU @ 2.30 GHz 和 GPU 为英伟达的 tesla T4 GPU(16 G 显存)。优化器使用 Adam,初始学习率设置为 0.001,使用周期性动态衰减策略。Batch size 设置为 16,训练总轮数为

200。模型输入的图像尺寸为 352×352,采用多尺度训练策略{0.75, 1, 1.25}。为了使训练更快收敛,本文将 Res2Net-50 的预训练权重加入到训练过程中,损失函数将 IoU 损失和二元交叉熵(BCE)损失进行组合,以改善小目标分割精度较低和类别不平衡问题,如式(12)~(14)所示:

$$\mathcal{L}_{\text{IoU}} = 1 - \frac{\sum_{i \in I} y_i \hat{y}_i}{\sum_{i \in I} y_i + \hat{y} - y_i \hat{y}_i} \quad (12)$$

$$\mathcal{L}_{\text{BCE}} = -\sum_{i \in I} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (13)$$

$$\mathcal{L} = \mathcal{L}_{\text{IoU}} + \mathcal{L}_{\text{BCE}} \quad (14)$$

2.4 RDG-Net 消融实验

为验证本文模型设计模块的有效性,本实验在训练方法和参数设置相同的环境下设计消融实验,并在 4 个数据集上分别验证。本文设计了 4 个消融模型,分别为模型 A:将 RDG-Net 中解码器第一阶段中的 GAM 去除;模型 B:在模型 A 的基础上将解码器第一阶段中的深度可分离卷积替换为普通卷积;模型 C:在模型 B 的基础上将解码器第二阶段的 PCF 模块去除;模型 D:在模型 B 的基础上将编码器替换为 ResNet-50,实验结果如表 1 所示。

模型 C 与模型 B 相比,在 Kvasir-SEG、ETIS-LaribPolypDB 和 CVC-300 数据集测试的结果中 DSC 和 IoU 分别比模型 B 降低了 0.54%、7.4%、2.25% 和 0.92%、10.51%、3.71%,表明解码器第二阶段的 PCF 模块在一定程度上可以对第一阶段输出的特征有效地解码至更高分类率的特征图,尤其是对于未参与训练的两个数据集性能大幅度提升。模型 B 与模型 A 相比,在 Kvasir-SEG 和 CVC-ClinicDB 数据集测试的结果中 DSC 和 IoU 分别比模型 A 降低了 0.02%、0.21%、0.03%、0.39%,但在 ETIS-LaribPolypDB 和 CVC-300 数据集中模型 A 的 DSC 和 IoU 略低于模型 B,表明使用深度可分离卷积替换传统卷积后,模型的参数量得到降低并且模型的性能几乎没有受到影响。模型 A 与 RDG-Net 相比,在 4 个数据集测试的结果中模型 A 的 DSC 和 IoU 分别比 RDG-Net 降低了 0.35%、0.13%、1.88%、1.77% 和 0.59%、0.24%、2.81%、2.97%。表明 GAM 模块可以有效帮助模型更关注息肉区域。模型 D 与 RDG-Net 相比,各项评判标准中均略低,表明与 ResNet 相比,Res2Net 更能有效地提取息肉的关键信息。

2.5 不同模型对比试验

为进一步验证 RDG-Net 的性能,在相同实验环境和参数设置的情况下与目前主流的分割模型 Unet++^[29]、segFormer^[30]和 LinkNet^[31]分别在 4 个数据集上进行性能对比,实验结果如表 2 所示。

表1 消融实验(%)
Table 1 Ablation study (%)

测试集	模型	准确率	精确度	召回率	DSC	IoU	F2
Kvasir-SEG	模型A	97.56	93.52	89.96	91.70	84.68	90.65
	模型B	97.61	95.82	87.89	91.68	84.65	89.37
	模型C	97.35	91.33	90.95	91.14	83.73	91.03
	模型D	97.16	88.88	92.69	90.75	83.06	91.90
	RDG-Net	97.66	93.76	90.40	92.05	85.27	91.05
CVC-ClinicDB	模型A	99.14	94.57	94.75	94.66	89.87	94.71
	模型B	99.11	94.31	94.58	94.45	89.48	94.53
	模型C	99.30	95.01	96.34	95.67	91.70	96.07
	模型D	99.18	93.89	96.07	94.97	90.43	95.63
	RDG-Net	99.16	94.74	94.84	94.79	90.11	94.82
ETIS-LaribPolypDB	模型A	98.53	85.43	81.62	83.48	71.65	82.35
	模型B	98.67	87.01	83.02	84.97	73.87	83.79
	模型C	97.93	76.42	78.75	77.57	63.36	78.27
	模型D	98.07	77.21	81.41	79.25	65.64	80.53
	RDG-Net	98.70	87.16	83.63	85.36	74.46	84.31
CVC-300	模型A	99.27	86.57	92.92	89.64	81.22	91.58
	模型B	99.37	88.40	93.54	90.90	83.32	92.46
	模型C	99.19	84.15	93.66	88.65	79.61	91.59
	模型D	99.19	83.46	94.87	88.80	79.86	92.34
	RDG-Net	99.40	88.41	94.63	91.41	84.19	93.32

表2 不同模型对比实验(%)
Table 2 Comparative experiments with different models (%)

测试集	模型	准确率	精确度	召回率	DSC	IoU	F2
Kvasir-SEG	Unet++	96.13	90.91	82.45	86.47	76.17	84.01
	segFormer	95.78	87.74	83.55	85.60	74.82	84.36
	LinkNet	97.50	95.40	87.55	91.30	84.00	89.01
	RDG-Net	97.66	93.76	90.40	92.05	85.27	91.05
CVC-ClinicDB	Unet++	98.63	95.59	86.93	91.05	83.58	88.53
	segFormer	98.59	90.79	91.77	91.28	83.95	91.57
	LinkNet	99.27	95.89	95.03	95.46	91.32	95.20
	RDG-Net	99.16	94.74	94.84	94.79	90.11	94.82
ETIS-LaribPolypDB	Unet++	96.80	66.69	58.68	62.43	45.38	60.13
	segFormer	94.39	43.83	84.29	57.67	40.52	71.15
	LinkNet	98.01	77.98	78.31	78.15	64.13	78.24
	RDG-Net	98.70	87.16	83.63	85.36	74.46	84.31
CVC-300	Unet++	98.35	75.65	75.32	75.49	60.63	75.39
	segFormer	97.09	53.99	91.08	67.80	51.28	80.08
	LinkNet	99.25	85.24	94.13	89.46	80.94	92.21
	RDG-Net	99.40	88.41	94.63	91.41	84.19	93.32

由表2可以看出在Kvasir-SEG和CVC-ClinicDB数据集中,RDG-Net的准确率、精确度、召回率、DSC、IoU和F2的平均值分别为98.41%、94.25%、92.62%、93.42%、87.69%、92.93%。其中,在CVC-ClinicDB数据集中召回率,DSC、IoU、F2评价指标略低于LinkNet,但在ETIS-LaribPolypDB和CVC-300两个新的数据集中展现了优势,准确率、精确率、召回率、DSC、IoU、F2指标的均值分别为99.05%、87.79%、89.13%、88.39%、79.33%、88.32%,各项评价标准均高于对比模型,如图2所示,在新数据集上的表现表明RDG-Net与其他模型相比在拥有较强性能的同时也具有更强的泛化能力。

图3可以明显看出,使用测试数据集和ETIS-LaribPolypDB和CVC-300这两个未参与训练的数据集对不同模型进行测试,Unet++、segFormer、LinkNet

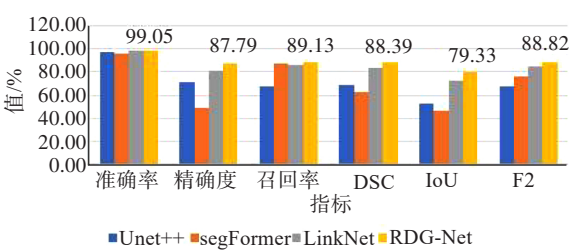


图2 RDG-Net在新数据集上的表现
Figure 2 Performance of RDG-Net on new datasets

这3个模型对数据中的小息肉分割能力较差,对中大息肉的边界分割也同样不理想。RDG-Net对测试数据和新数据中不管是对小息肉还是对中大息肉的边界部分都能够有效地分割出来。

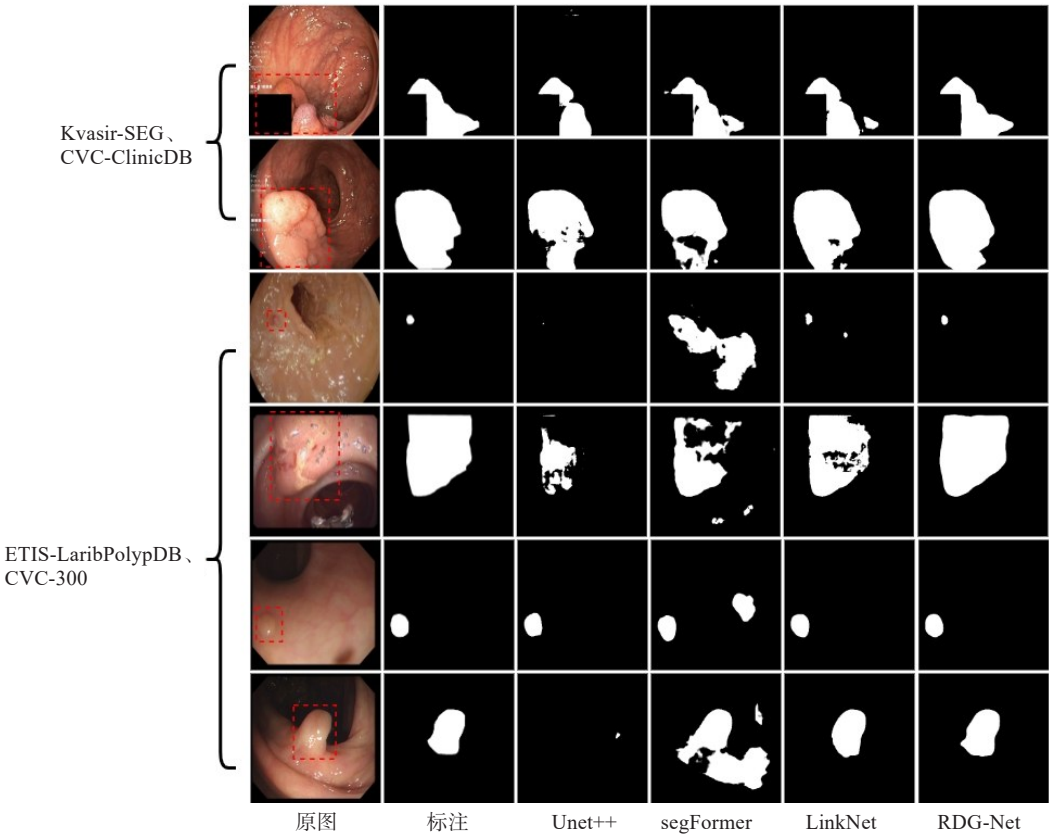


图3 不同模型在ETIS-LaribPolypDB和CVC-300数据集中部分分割效果
Figure 3 Some segmentation results of different models on the ETIS-LaribPolypDB and CVC-300 datasets

3 结论

本文针对结直肠息肉图像中息肉边界不清晰、形状大小各异、图像对比度低且特征不明显等问题,提出了一种基于Res2Net-50的双阶段解码器结直肠息肉图像分割模型RDG-Net。RDG-Net采用编码器-

解码器结构,其中编码器部分使用Res2Net-50作为骨干网络,以其卓越的特征提取能力有效地提取息肉特征。在解码器第一阶段,RDG-Net引入了特征聚合模块,充分融合不同层级的息肉特征。同时,RFB模块和GAM的加入强化了对不同大小息肉特征的提取,尤其是对较小且难以关注的息肉特征。

解码器第二阶段的PCF模块则充分利用第一阶段解码器与编码器聚合的特征,通过不同核心的卷积获取多尺度信息并进行融合,确保在解码过程中不丢失关键的边界信息,实现特征互补的效果。在Kvasir-SEG、CVC-ClinicDB、ETIS-LaribPolypDB和CVC-300 4个数据集上的验证实验结果表明,RDG-Net能够有效且精准地分割息肉区域,辅助医生进行临床诊断并提升诊断效率。尽管RDG-Net模型性能表现优秀,但仍存在进一步优化与提升的空间。可以考虑引入更先进的注意力机制或特征融合策略,以进一步挖掘图像中的潜在信息,提高模型对复杂病例的处理能力。加强跨数据集的实验验证,以检验模型在不同医疗环境下的泛化能力,对于推动其在实际临床中的应用具有重要意义。

【参考文献】

- [1] Arnold M, Sierra MS, Laversanne M, et al. Global patterns and trends in colorectal cancer incidence and mortality[J]. Gut, 2017, 66(4): 683-691.
- [2] Tian JB, Miao XP, Lin DX. Recent advances in genetic epidemiology of colorectal cancer in Chinese population[J]. Progress China Epidemiol, 2023(1): 187.
- [3] Mamonov AV, Figueiredo IN, Figueiredo PN, et al. Automated polyp detection in colon capsule endoscopy[J]. IEEE Trans Med Imaging, 2014, 33(7): 1488-1502.
- [4] Bernal J, Sánchez J, Vilariño F. Towards automatic polyp detection with a polyp appearance model[J]. Pattern Recognit, 2012, 45(9): 3166-3182.
- [5] Qureshi I, Yan JH, Abbas Q, et al. Medical image segmentation using deep semantic-based methods: a review of techniques, applications and emerging trends[J]. Inf Fusion, 2023, 90: 316-352.
- [6] 周金治, 胡震, 郭莉莉, 等. 基于GAN-DAUnet的肝脏CT图像肿瘤分割算法[J]. 中国医学物理学杂志, 2023, 40(8): 971-976.
Zhou JZ, Hu Z, Guo LL, et al. Liver CT image tumor segmentation algorithm based on GAN-DAUnet[J]. Chinese Journal of Medical Physics, 2023, 40(8): 971-976.
- [7] 魏坤, 沈记全, 赵艳梅. MAUNet:用于皮肤病变分割的轻量级模型[J]. 中国医学物理学杂志, 2023, 40(5): 555-561.
Wei K, Shen JQ, Zhao YM. MAUNet: a lightweight model for skin lesion segmentation[J]. Chinese Journal of Medical Physics, 2023, 40(5): 555-561.
- [8] 邱敬, 马帅, 王国栋, 等. 基于改进Unet与动态阈值可变FCMSPCNN的医学图像分割[J]. 中国医学物理学杂志, 2023, 40(3): 328-335.
Di J, Ma S, Wang GD, et al. Medical image segmentation using improved Unet combined with dynamic threshold changed FCMSPCNN[J]. Chinese Journal of Medical Physics, 2023, 40(3): 328-335.
- [9] Gu ZW, Cheng J, Fu HZ, et al. CE-net: context encoder network for 2D medical image segmentation[J]. IEEE Trans Med Imaging, 2019, 38(10): 2281-2292.
- [10] Zhou TF, Li LL, Bredell G, et al. Volumetric memory network for interactive medical image segmentation[J]. Med Image Anal, 2023, 83: 102599.
- [11] Ye ZW, Hu B, Sui HG, et al. DSCA-Net: double-stage codec attention network for automatic nuclear segmentation[J]. Biomed Signal Process Control, 2024, 88, Part A: 105569.
- [12] Shaker A, Maaz M, Rasheed H, et al. UNETR++: delving into efficient and accurate 3D medical image segmentation[J]. IEEE Trans Med Imaging, 2024, 43(9): 3377-3390.
- [13] Hatamizadeh A, Tang YC, Nath V, et al. UNETR: transformers for 3D medical image segmentation[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ, USA: IEEE, 2022: 1748-1758.
- [14] Zhang RF, Li GB, Li Z, et al. Adaptive context selection for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2020. Cham: Springer International Publishing, 2020: 253-262.
- [15] Zhao XQ, Zhang LH, Lu HC. Automatic polyp segmentation via multi-scale subtraction network[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer International Publishing, 2021: 120-130.
- [16] Fan DP, Ji GP, Zhou T, et al. PraNet: parallel reverse attention network for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2020. Cham: Springer International Publishing, 2020: 263-273.
- [17] Zhou T, Zhou Y, He KL, et al. Cross-level feature aggregation network for polyp segmentation[J]. Pattern Recognit, 2023, 140: 109555.
- [18] Gao SH, Cheng MM, Zhao K, et al. Res2Net: a new multi-scale backbone architecture[J]. IEEE Trans Pattern Anal Mach Intell, 2021, 43(2): 652-662.
- [19] Howard AG, Zhu ML, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17). <https://arxiv.org/abs/1704.04861>.
- [20] Liu ST, Huang D, Wang YH. Receptive field block net for accurate and fast object detection[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 404-419.
- [21] Liu YC, Shao ZR, Hoffmann N. Global attention mechanism: retain information to enhance channel-spatial interactions[EB/OL]. (2021-12-10). <https://arxiv.org/abs/2112.05561>.
- [22] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [23] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 2261-2269.
- [24] Wandell BA, Winawer J. Computational neuroimaging and population receptive fields[J]. Trends Cogn Sci, 2015, 19(6): 349-357.
- [25] Jha D, Smedsrud PH, Riegler MA, et al. Kvasir-SEG: a segmented polyp dataset[C]//MultiMedia Modeling. Cham: Springer International Publishing, 2020: 451-462.
- [26] Bernal J, Sánchez FJ, Fernández-Esparrach G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians[J]. Comput Med Imaging Graph, 2015, 43: 99-111.
- [27] Silva J, Histace A, Romain O, et al. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer[J]. Int J Comput Assist Radiol Surg, 2014, 9(2): 283-293.
- [28] Vázquez D, Bernal J, Sánchez FJ, et al. A benchmark for endoluminal scene segmentation of colonoscopy images[J]. J Healthc Eng, 2017, 2017: 4037190.
- [29] Zhou ZW, Rahman Siddiquee MM, Tajbakhsh 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 International Publishing, 2018: 3-11.
- [30] Xie EZ, Wang WH, Yu ZD, et al. SegFormer: simple and efficient design for semantic segmentation with transformers[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2024: 12077-12090.
- [31] Chaurasia A, Culurciello E. LinkNet: Exploiting encoder representations for efficient semantic segmentation[C]//2017 IEEE Visual Communications and Image Processing (VCIP). Piscataway, NJ, USA: IEEE, 2017: 1-4.

(编辑:薛泽玲)