

一种新型编解码结构的结肠息肉分割算法研究

李筠, 汪芳, 杨海马, 宋夜夜

(上海理工大学光电信息与计算机工程学院, 上海 200093)

✉ lijuny@usst.edu.cn; wovf97@outlook.com; snowyhm@sina.com; syyhut@163.com



摘要:应用人工智能进行结肠息肉检测对医疗辅助系统十分重要,然而目前的分割模型存在准确率低、对部分样本细节难以识别的问题。因此,提出一种基于编解码结构的息肉分割模型。该模型采用跳跃轴向注意力解决梯度问题,利用适应联系训练弥补池化中信息丢失问题,使用双通道门控将低分辨率预测图细化为高分辨率显著图。在数据集 CVC-ClinicDB 与 Kvasir-SEG 上对该模型进行验证,从 mIoU(平均交并比系数)、Dice(Dice 相似系数)、Precision(准确率)和 Accuracy(正确率)四个指标上与同类深度学习分割算法进行对比,结果为该模型在 CVC-ClinicDB 上, $mIoU$ 为 0.903, $Dice$ 为 0.947, $Precision$ 为 0.933, $Accuracy$ 为 0.933, 在 Kvasir-SEG 上, $mIoU$ 为 0.763, $Dice$ 为 0.868, $Precision$ 为 0.857, $Accuracy$ 为 0.867, 均优于同类深度学习分割算法,验证了该模型对样本细节部分具有更好的分割效果。

关键词:跳跃轴向注意力;适应联系训练;双通道门控;息肉分割;深度学习

中图分类号:TP391 **文献标志码:**A

Research on a Novel Encoding and Decoding Structure for Colon Polyp Segmentation Algorithm

LI Jun, WANG Fang, YANG Haima, SONG Yeye

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

✉ lijuny@usst.edu.cn; wovf97@outlook.com; snowyhm@sina.com; syyhut@163.com

Abstract: The application of artificial intelligence for colon polyp detection is very important for medical assistance systems. However, current segmentation model has problems of low accuracy and difficulty in recognizing some sample details. Therefore, this paper proposes a polyp segmentation model based on encoding and decoding structure. In this model, skip axial attention is used to solve the gradient problem, adaptive connection training is used to compensate for information loss in pooling, and dual channel gating is used to refine the low resolution prediction map into a high-resolution saliency map. The model is verified on datasets CVC-ClinicDB and Kvasir-SEG and Compared with similar deep learning segmentation algorithms on four metrics: mIoU, Dice, Precision, and Accuracy. Results show that on CVC-ClinicDB, $mIoU$ is 0.903, $Dice$ is 0.947, $Precision$ is 0.933, and $Accuracy$ is 0.933. On Kvasir-SEG, $mIoU$ is 0.763, $Dice$ is 0.868, $Precision$ is 0.857, and $Accuracy$ is 0.867. All of the results are superior to similar deep learning segmentation algorithms, verifying that the proposed model has better segmentation performance for sample details.

Key words: skip axial attention; adaptive connection training; dual channel gating; polyp segmentation; deep learning

0 引言(Introduction)

随着生活条件的改善,人们的生活饮食结构也发生了变

化,流行病学研究显示不同地区居民的大肠息肉检出率在 10.25%~26.64%,并呈逐年上升趋势^[1]。由于大肠息肉的早

期症状不明显,不易被发现,所以结肠检查对结直肠癌的早期诊断和预防非常重要^[2]。为了解决医疗资源不足的问题,并且提高结肠检查的准确率,人们广泛运用高性能计算技术协助进行医疗诊断。

在图像分割的研究领域,ZHOU等^[3]在 U-Net 模型的基础上提出了 UNet+,将编码器和解码器通过一系列嵌套的密集跳过路径连接,从而缩小了编码器和解码器的特征映射之间的语义差距。FAN等^[4]提出使用并行的部分解码器组件获取全局特征图和递归反向注意模块,然后通过全局特征图和反向注意机制建立区域与边界的关系,提高了对息肉分割的准确性。YEUNG等^[5]采用双通道注意力,获取上下文的特征进行对比加权增强识别结果,弥补了传统空间卷积丢失相关细节特征的缺陷。

从上述研究内容可以看出,人们在基于 U-Net 模型的基础上提出了许多改进方案,但是这些改进方案中大部分忽略了在 U-Net 模型不断地编码解码的层次变换中出现了信息丢失,以及同一层次之间的编码器与解码器的联系,对一些畸形的不容易分割的息肉图像无法达到预期分割效果的问题。针对上述问题,本文在 U-Net 模型结构的基础上对编解码器结构模型进行优化,提出了一种新型结肠息肉图像分割模型。

1 模型结构(Model structure)

本文基于 U-Net 的编码器-解码器结构的模型基础提出了三种结构。

(1)轴向注意力机制的结合模块,弥补在网络层次加深后造成的梯度爆炸或者梯度消失的问题。同时,通过轴向注意力机制,保持了特征中较远距离的位置之间的联系。

(2)适应联系的训练。使用不同空洞膨胀率的空洞卷积弥补池化过程中的特征信息丢失问题。同时,采用自注意力模型弥补池化过程中空间结构的信息丢失问题。

(3)双通道注意力连接,挖掘特征图中目标区域的结构信息,将粗略和低分辨率的预测图细化为一个完整的包含目标区域和细节高分辨率的显著图。

1.1 模型结构图

如图 1 所示,本文所提研究模型基于 U-Net 模型的对称编解码器-解码器结构。编码器阶段,在进行每一层卷积运算之前加入跳跃轴向注意力模块,解决原编码器结构中存在的梯度问题;池化过程中,加入自适应联系训练,弥补池化过程中的信息丢失问题;解码器阶段,每层的输入特征与同层的编码器输出特征进行双通道注意力连接,保留目标区域信息。经过 4 层编解码器-解码器运算,得到输出结果。

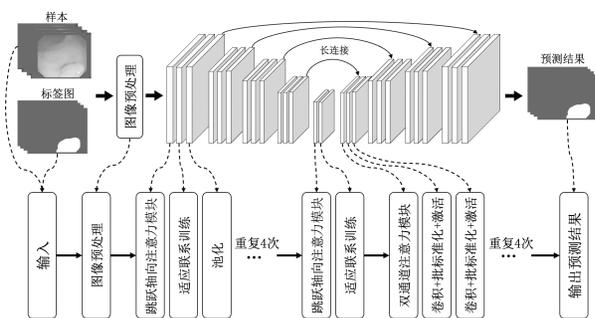


图 1 模型结构
Fig. 1 Model structure

1.2 跳跃轴向注意力

随着神经网络层次的增多,容易造成梯度爆炸和梯度消失的问题。梯度爆炸会导致神经网络的训练不稳定,无法获得有效的数据,而梯度消失会导致训练权重的更新缓慢甚至停滞。于是,本文提出跳跃轴向注意力机制解决梯度问题。跳跃轴向注意力模块结构如图 2 所示。

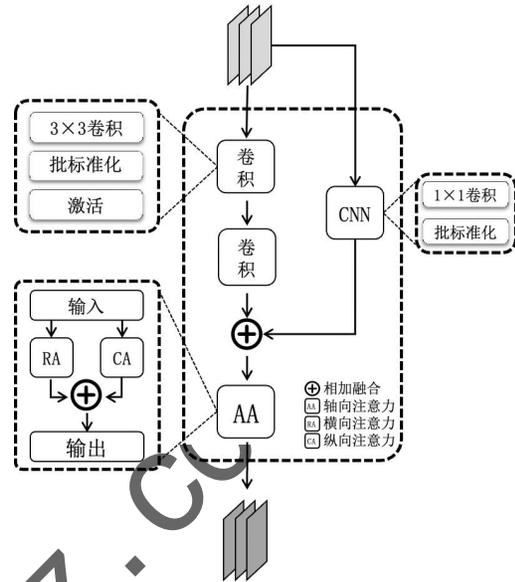


图 2 跳跃轴向注意力模块结构

Fig. 2 Skip axial attention module

(1)先将每一层的输入进行 2 次卷积、1 次批标准化及 1 次激活的运算,运算结果记为 R1。

(2)将每一层的输入进行一次 1x1 的卷积运算和批标准化,运算结果记为 R2。

(3)将 R1 与 R2 进行矩阵相加融合,运算结果记为 R3。

(4)将 R3 加入轴向注意力模块,轴向注意力即图 3 所示的横向注意力模块与图 4 所示的纵向注意力模块的并联合,图 3 中的 V、Q、K 分别代表值矩阵(Value Matrix)、查询矩阵(Query Matrix)和键矩阵(Key Matrix)。这些矩阵都是在训练过程中随机初始化的权重矩阵,并在梯度下降过程中进行优化。将两个注意力的运算结果进行矩阵相加融合,结果记为整体跳跃轴向注意力机制的输出。

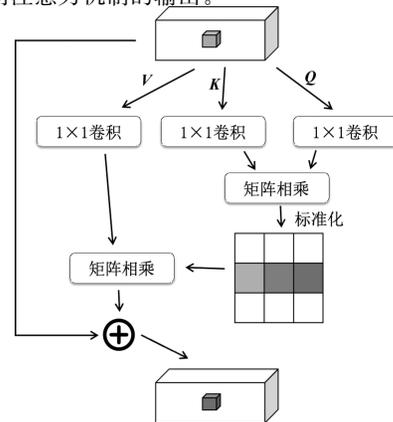


图 3 横向注意力结构

Fig. 3 Row attention structure

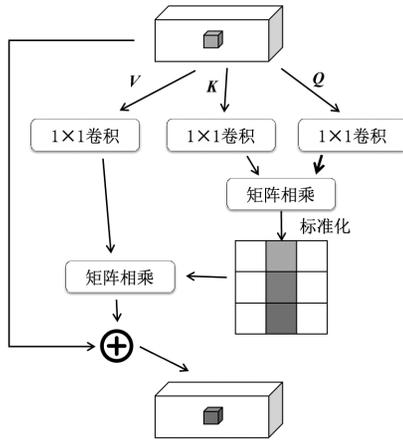


图4 纵向注意力结构
Fig. 4 Col attention structure

1×1 的卷积核提供了类似全连接的运算,有效地增加了网络的深度,保证输入尺寸不变,同时增强了非线性运算能力,有效地提高了整个网络的表达能力^[6]。轴向注意力将平面上的特征沿着横向和纵向进行平行的分解,将平面特征降为一维的线性特征,有效地降低了学习成本^[7]。

1.3 适应联系训练

在每一层运算结束后,模型会进行池化运算再进入下一层,这样的池化操作会丢失较多的空间结构信息,导致出现不同尺寸大小的图像、分割目标的尺寸相差过大,以及畸形或者尺寸较小的样本等现象,会导致模型泛化能力变弱,无法分割出复杂样本区域^[8]。于是,本文提出了适应联系训练用于抽取不同尺寸的样本关联,从而适应更多尺寸的样本。适应联系训练结构图如图5所示。

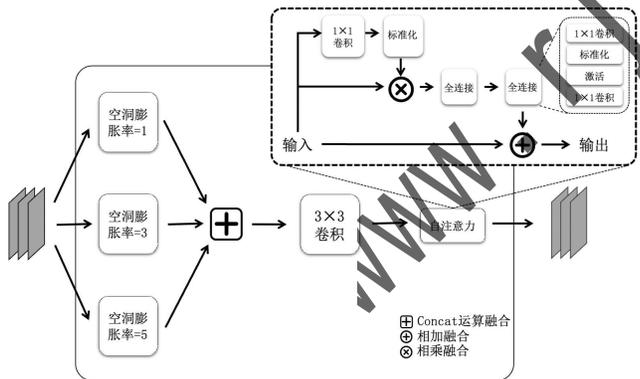


图5 适应联系训练结构图

Fig. 5 Diagram of adaptive connection training structure

- (1)对输入分别进行空洞膨胀率为1、3、5的空洞卷积,将输出记为 R_1, R_2, R_3 。
- (2)将 R_1, R_2, R_3 进行Concat运算融合,结果记为 R_4 。
- (3)将 R_4 进行一次 3×3 卷积运算,结果记为 R_5 。
- (4)将 R_5 加入如图5所示的自注意力模块,将输入特征复制为3份,即 I_1, I_2, I_3 ,对 I_1 进行 1×1 卷积及标准化操作,得到 R_6 ,将 R_6 与 I_2 进行相乘融合及两次全连接运算,得到 R_7 ,再将 R_7 与 I_3 进行相加融合,作为模块输出。

通过空洞卷积,保留了图像内部结构的特征。通过自注意力模型,将任意位置的信息关联,让模型在充分利用池化的操

作增强感受野优势的同时,也弥补了池化操作造成的信息丢失问题。

1.4 双通道注意力门控模块

解码阶段,在进行上采样时,通常会忽视编码器-解码器特有的对称结构信息,没有充分联系对应编码层输出所包含的信息,容易造成信息缺失^[9]。于是,本文改造了上采样阶段流程,具体流程如图6所示。

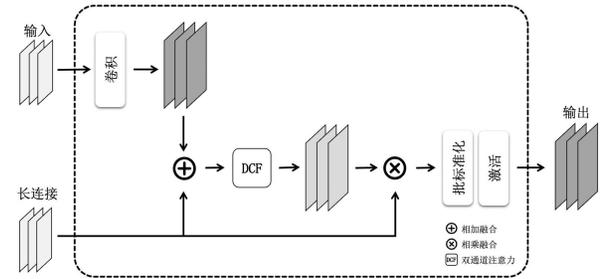


图6 双通道注意力门控模块

Fig. 6 Dual channel attention gating module

- (1)将前一层产生的输入特征进行上采样运算,运算结果记为 R_1 。
- (2)通过长连接将对应编码层的输出与 R_1 进行特征融合,结果记为 R_2 。
- (3)将 R_2 通过图7所示的双通道注意力模块,首先将模块输入与图8所示的通道注意力模型进行运算,其次与模块输入进行融合,再次与图9所示的空间注意力模型进行运算,最后与模块输入进行融合得到模块输出 R_3 。
- (4)将 R_3 进行批标准化和激活运算,作为整个模块输出。

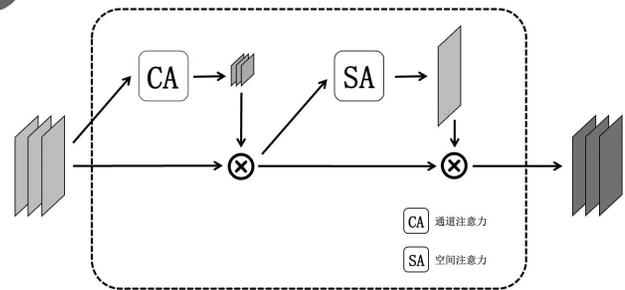


图7 双通道注意力结构图

Fig. 7 Diagram of dual channel attention structure

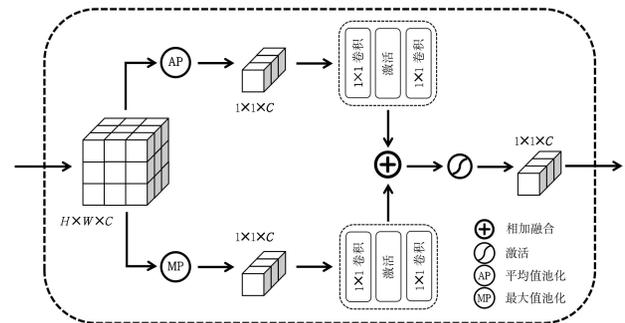


图8 通道注意力结构图

Fig. 8 Diagram of channel attention structure

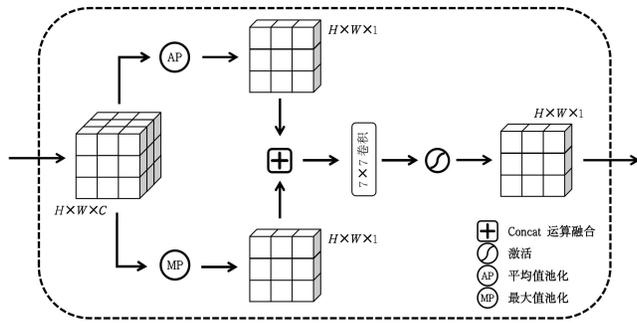


图9 空间注意力结构图

Fig.9 Diagram of spatial attention structure

1.5 损失函数设计

虽然在语义分割任务中经常使用交叉熵损失函数,但是在实验中,息肉区域相对背景区域小得多,实验过程中存在严重类别失衡问题,因此本实验采用骰子系数损失函数^[10]。其中, G_n 和 P_n 分别表示第 n 个像素的标记数据和第 n 个像素的预测数据, ϵ 是平滑常量, $\gamma\|\omega\|_2^2$ 表示用于避免过拟合的 $L2$ 正则化损失,具体如公式(1)所示:

$$Loss = 1 - \frac{\sum_{n=1}^N 2G_n P_n + \epsilon}{\sum_{n=1}^N G_n + P_n + \epsilon} + \gamma\|\omega\|_2^2 \quad (1)$$

2 实验过程(Experimentation)

2.1 数据集

如表1所示,本文所使用的数据集分别是:CVC-ClinicDB, Kvasir-SEG,其中CVC-ClinicDB包含612张样本数据,Kvasir-SEG包含1000张样本数据。将数据集分为8份训练集、1份验证集和1份测试集。训练集用于模型的训练,验证集用来进行模型泛化使用,将得到的预测图与标记图进行对比,得到评价指标得分,量化模型分割效果,测试集用来将得分最高的模型进行泛化,得到预测图。由于不同数据集的尺寸大小不一,所以训练前需要对数据集进行预处理工作,将样本图片尺寸统一以保证训练参数的一致性。

表1 实验数据集的构成

Tab.1 The composition of the experimental dataset

| 数据集 | 样本数量/张 | 尺寸(像素) | 样本 | 标签 |
|------------------------------|--------|----------|----|----|
| CVC-ClinicDB ^[11] | 612 | 384×288 | | |
| Kvasir-SEG ^[12] | 1000 | Variable | | |

2.2 评价指标

本文引入4个分界定义: TP (真阳性),即样本预测为正,标记为正,预测正确; FN (假阴性),即样本预测为负,标记为正,预测错误; FP (假阳性),即样本预测为正,标记为负,预测错误; TN (真阴性),即样本预测为负,标记为负,预测正确。同时,引入了4个评价指标量化检验所用模型的效果,具体的评价指标计算公式如下。

(1)Dice相似系数(Dice Similarity Coefficient):计算预测目标区域与实际目标区域的相似性。 $Dice$ 公式计算如下:

$$Dice = \frac{2TP}{FN + TP + TP + FP} \quad (2)$$

(2)平均交并比系数(mIoU):计算预测值和实际值两个集合的交集与并集的比值,结果的交并比系总和取平均值。

$mIoU$ 公式计算如下,其中 k 表示类别, $k+1$ 表示加上了背景类, i 表示真实值。

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{FN + TP + FP} \quad (3)$$

(3)准确率(Precision):计算机预测符合要求的正确识别物体的个数占总识别出的物体个数的百分数,准确率相关公式如下:

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

(4)正确率(Accuracy):计算机预测正确物体的个数所占所有样本个数的百分数,正确率计算公式如下:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

2.3 实验结果

如图10所示,与同类模型相比,本文所提出的模型具有更好的分割效果。在CVC-ClinicDB数据集上,本文实验的 $mIoU$ 和 $Dice$ 数值分别为0.903和0.947, $Precision$ 为0.933, $Accuracy$ 为0.933,对比其他组实验均有更好的效果,在CVC-ClinicDB数据集上不同模型的结果对比如表2所示。

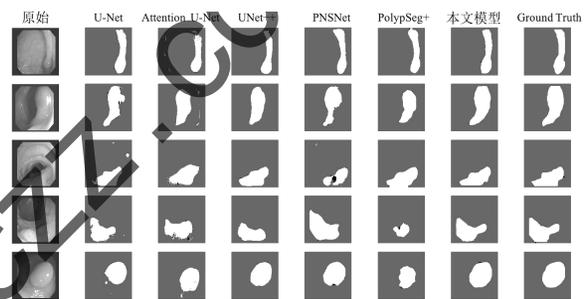


图10 模型在CVC-ClinicDB数据集上的部分分割结果对比图

Fig.10 Comparison of partial segmentation results of the model on the CVC-ClinicDB dataset

表2 在CVC-ClinicDB数据集上不同模型的结果对比

Tab.2 Comparison of results of different models on the CVC-ClinicDB dataset

| 模型 | $mIoU$ | $Dice$ | $Precision$ | $Accuracy$ |
|---------------------------------|--------------|--------------|--------------|--------------|
| U-Net ^[13] | 0.755 | 0.824 | 0.838 | 0.873 |
| Attention U-Net ^[14] | 0.858 | 0.911 | 0.919 | 0.912 |
| UNet++ ^[3] | 0.741 | 0.828 | 0.831 | 0.918 |
| PNSNet ^[15] | 0.800 | 0.873 | 0.894 | 0.923 |
| PolypSeg+ ^[10] | 0.884 | 0.921 | 0.907 | 0.923 |
| 本文模型 | 0.903 | 0.947 | 0.933 | 0.933 |

在Kvasir-SEG数据集上,本文实验的 $mIoU$ 和 $Dice$ 的指标分别为0.763和0.868, $Precision$ 为0.857, $Accuracy$ 为0.867,均对比实验组有更好的效果,在Kvasir-SEG数据集上不同模型的结果对比如表3所示。

表3 在Kvasir-SEG数据集上不同模型的结果对比

Tab.3 Comparison of results of different models on the Kvasir-SEG dataset

| 模型 | $mIoU$ | $Dice$ | $Precision$ | $Accuracy$ |
|---------------------------------|--------|--------|-------------|------------|
| U-Net ^[13] | 0.653 | 0.596 | 0.672 | 0.681 |
| Attention U-Net ^[14] | 0.670 | 0.767 | 0.811 | 0.823 |
| UNet++ ^[3] | 0.700 | 0.800 | 0.799 | 0.782 |

续表

| 模型 | mIoU | Dice | Precision | Accuracy |
|---------------------------|--------------|--------------|--------------|--------------|
| PNSNet ^[15] | 0.697 | 0.762 | 0.804 | 0.793 |
| PolypSeg+ ^[10] | 0.706 | 0.788 | 0.841 | 0.853 |
| 本文模型 | 0.763 | 0.868 | 0.857 | 0.867 |

2.4 消融实验

为了验证模型结构的合理性,在 CVC-ClinicDB 数据集上进行消融实验。Baseline 为骨干网络;SAA 为跳跃轴向注意力模块;ACT 为适应联系训练模块;DCG 为双通道注意力门控模块。本文设计了 8 组实验进行评估:①骨干网络;②骨干网络引入 SAA 模块;③骨干网络引入 ACT 模块;④骨干网络引入 DCG 模块;⑤骨干网络引入 SAA 和 ACT 模块;⑥骨干网络引入 ACT 和 DCG 模块;⑦骨干网络引入 SAA 和 DCG 模块;⑧本文模型。不同模块的消融实验结果对比如表 4 所示,在分别引入了三个模块后对比骨干网络均有显著提升,引入三个模块后,模型取得了较好的实验结果,证明了本文模型结构的合理性。

表 4 不同模块的消融实验结果对比

Tab.4 Comparison of ablation experimental results of different modules

| 方法 | mIoU | Dice | Precision | Accuracy |
|-----------------------------|--------------|--------------|--------------|--------------|
| Baseline | 0.755 | 0.824 | 0.838 | 0.873 |
| Baseline+SAA | 0.873 | 0.914 | 0.904 | 0.907 |
| Baseline+ACT | 0.793 | 0.874 | 0.861 | 0.884 |
| Baseline+DCG | 0.863 | 0.913 | 0.911 | 0.917 |
| Baseline+SAA+ACT | 0.892 | 0.923 | 0.922 | 0.896 |
| Baseline+ACT+DCG | 0.892 | 0.934 | 0.921 | 0.935 |
| Baseline+SAA+DCG | 0.894 | 0.932 | 0.927 | 0.918 |
| Baseline+SAA+ACT+DCG | 0.903 | 0.947 | 0.933 | 0.933 |

3 结论(Conclusion)

本文提出了以编码器-解码器结构模型为基础,通过采用 1 个跳跃连接模块和接入轴向注意力机制,解决了因神经网络的层次加深导致的梯度消失或者梯度爆炸的问题,更好地获取整个图像远近位置的联系。采用适应联系训练,有效地减少了池化过程中空间信息的丢失。采用双通道门控模块,保证了解码过程中空间信息和通道信息的完整性。经过对比实验验证了本文提出模型的效果和可行性。经过消融实验验证了本文结构的合理性。未来,需要丰富更多的数据集验证本文模型的效果,同时对模型进行工程类的嵌入开发,实现标准化输入与标准化输出,实现可以完成批量工业化处理图像的能力。

参考文献(References)

- [1] 李倩倩,王军,赵越,等. 结肠直肠息肉发生相关危险因素的研究现状[J]. 医学综述,2020,26(16):3196-3200.
- [2] 李可俊. LncRNA UCA1 促进结肠癌紫杉醇耐药的机制研究[D]. 广州:广州医科大学,2021.
- [3] ZHOU Z W, SIDDIQUEE M M R, TAJBAKSH N, et al. UNet++: a nested U-Net architecture for medical image segmentation[DB/OL]. (2018-07-18)[2023-03-16]. <https://doi.org/10.48550/arXiv.1807.10165>.
- [4] FAN D P, JI G P, ZHOU T, et al. PraNet: parallel reverse attention network for polyp segmentation [C]//Medical Image Computing and Computer Assisted Intervention. In-

ternational Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2020: 263-273.

- [5] YEUNG M, SALA E, SCHÖNLIEB C B, et al. Focus U-Net: a novel dual attention-gated CNN for polyp segmentation during colonoscopy[DB/OL]. (2021-06-22)[2023-03-16]. <https://doi.org/10.48550/arXiv.2105.07467>.
- [6] ZHANG X Q. Melanoma segmentation based on deep learning[J]. Computer Assisted Surgery, 2017, 22(sup1): 267-277.
- [7] ZHUANG X L, ZHANG T M. Detection of sick broilers by digital image processing and deep learning[J]. Biosystems Engineering, 2019, 179: 106-116.
- [8] WEI Y C, LIANG X D, CHEN Y P, et al. Learning to segment with image-level annotations[J]. Pattern Recognition, 2016, 59: 234-244.
- [9] JI Y Z, ZHANG H J, JONATHAN WU Q M. Salient object detection via multi-scale attention CNN[J]. Neurocomputing, 2018, 322: 130-140.
- [10] WU H S, ZHAO Z B, ZHONG J F, et al. PolypSeg+: a lightweight context-aware network for real-time polyp segmentation[J]. IEEE Transactions on Cybernetics, 2023, 53(4): 2610-2621.
- [11] BERNAL J, SÁNCHEZ F J, FERNÁNDEZ-ESPARACH 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.
- [12] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Kvasir-SEG: a segmented polyp dataset[C]//RO Y M, CHENG W H, KIM J, et al. MultiMedia Modeling. Cham: Springer, 2020: 451-462.
- [13] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [DB/OL]. (2015-05-18)[2023-03-16]. <https://doi.org/10.48550/arXiv.1505.04597>.
- [14] OKTAY O, SCHLEMPER J, LE FOLGOC L, et al. Attention U-Net: learning where to look for the pancreas [DB/OL]. (2018-05-20)[2023-03-16]. <https://doi.org/10.48550/arXiv.1804.03999>.
- [15] JI G P, CHOU Y C, FAN D P, et al. Progressively normalized self-attention network for video polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention. 24th International Conference. Strasbourg, France: Springer, 2021: 142-152.

作者简介:

李筠(1975-),女,博士,副教授。研究领域:数据处理及误差分析。

汪芳(1997-),女,硕士生。研究领域:图像处理。

杨海马(1976-),男,博士,副教授。研究领域:图像处理。本文通信作者。

宋夜夜(1998-),女,硕士生。研究领域:人工智能。