

基于双分支特征提取的息肉图像分割网络

林元杰¹, 刘巧红^{2*}, 韩啸翔¹, 陈柯炎¹, 张维坤¹

¹上海理工大学健康科学与工程学院, 上海

²上海健康医学院医疗器械学院, 上海

收稿日期: 2023年10月27日; 录用日期: 2024年1月3日; 发布日期: 2024年1月11日

摘要

结肠息肉分割是从结肠息肉图像中提取病理信息的关键步骤, 对结直肠癌的诊断和治疗具有重要意义。针对结肠息肉分割中形状大小不一、病灶组织与背景差异性难以区分等问题, 本文提出了一种基于CNN和Swin Transformer的双分支特征提取的医学息肉分割网络(DST-Net), 其充分考虑了卷积神经网络和Transformer在提取局部特征和全局特征各自的优势。DST-Net是一种编码器-解码器架构, 首先设计了基于VGG和空洞卷积(AC)的两种编码器分别提取局部边界特征和多尺度特征; 接下来在编解码器中间的底部模块使用两个连续的Swin Transformer模块, 充分利用Transformer的远程依赖关系进一步加强网络的全局特征提取能力; 最后, 在编解码器之间的跳过连接中使用了通道注意力模块(CAB), 以更加关注可疑和复杂的区域。在CVC-Clinic DB和Kvasir两个公开的息肉数据集上验证了所提出方法, 结果表明该模型优于现有的其他方法, 可以准确有效地实现结肠息肉的分割任务。

关键词

医学图像分割, CNN Transformer, 特征融合, 特征提取

Polyp Image Segmentation Network Based on Double Branch Feature Extraction

Yuanjie Lin¹, Qiaohong Liu^{2*}, Xiaoxiang Han¹, Keyan Chen¹, Weikun Zhang¹

¹School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai

²School of Medical Instruments, Shanghai University of Medicine and Health Sciences, Shanghai

Received: Oct. 27th, 2023; accepted: Jan. 3rd, 2024; published: Jan. 11th, 2024

Abstract

Colon polyp segmentation is a key step to extract pathological information from colon polyp im-

*通讯作者。

文章引用: 林元杰, 刘巧红, 韩啸翔, 陈柯炎, 张维坤. 基于双分支特征提取的息肉图像分割网络[J]. 软件工程与应用, 2024, 13(1): 1-10. DOI: 10.12677/sea.2024.131001

ages, which is of great significance for the diagnosis and treatment of colorectal cancer. Aiming at the problems such as different shapes and sizes in colon polyp segmentation and difficult to distinguish lesion tissue and background differences, this paper proposes a Dual-branch feature extraction medical polyp segmentation network (DST-Net), which fully considers the advantages of convolutional neural network and Transformer in extracting local features and global features. DST-Net is an encoder-decoder architecture. Firstly, two encoders based on VGG and void convolution (AC) are designed to extract local boundary features and multi-scale features respectively. Next, the bottom module in the middle of the codec uses two continuous Swin Transformer modules to make full use of the remote dependency of Transformer to further strengthen the global feature extraction capability of the network. Finally, the Channel Attention Module (CAB) is used in skip connections between codecs to pay more attention to suspicious and complex areas. The proposed method was validated on two publicly available polyp datasets, CVC-Clinic DB and Kvasir. The results show that the model is superior to other existing methods and can accurately and effectively achieve the task of colon polyp segmentation.

Keywords

Polyp Segmentation, CNN Transformer, Feature Extraction, Feature Fusion

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

结直肠癌(Colorectal cancer, CRC)是最常见的癌症之一,据相关统计,在2022年,中国有592,232例新发病例和309,114例死亡[1]。息肉作为最明显的CRC前体之一,可以通过筛查发现并在其发展为癌症之前移除。目前,内镜筛查作为筛查结直肠相关病变的金标准,可以提供结直肠息肉的外观和位置信息,帮助医生早期诊断和治疗。然而,医生的主观性和经验,以及高昂的人工成本会导致临床效果不佳。因此,计算机辅助诊断有利于克服这些局限性,减轻医生负担,提高诊断效率,防止误诊漏诊。目前人工智能研究已经取得了很好的进展,但由于结肠息肉颜色形态不一、病灶组织与背景差异性难以区分等问题,导致息肉分割的学习模型很难去鉴别,在息肉的自动分割上仍具有挑战。

近年来,卷积神经网络(Convolutional Neural Network, CNN)已经成为医学图像分割的主流[2] [3] [4]。以CNN构建的息肉图像分割算法,能够很好的发挥其局部提取空间和通道特征的能力[5]。U-Net [6]在编码器中通过卷积层提取特征,并在对称解码路径中恢复空间信息。U-Net的U型结构已成为医学图像分割的一种经典网络架构。进一步将U-Net++ [7]和ResUNet++ [8]引入到息肉分割中,以减少U型网络编码器和解码器之间的语义差距,并产生一些可接受的结果。这些CNN所采用的编码器-解码器架构可以有效地消除传统手工制作的基于特征方法的局限性,但是CNN提取特征的时候容易忽略全局上下文信息,无法学习远程依赖关系。息肉的形态不一且与边界周围内在联系紧密,导致卷积网络在特征提取时会忽视一些信息或产生不准确的预测。针对CNN的缺点,引入注意力机制可以有效抑制息肉背景噪声,提升分割的准确率。Transformer有很好的全局上下文建模能力,能够有效的学习特征的远程依赖性。但其对于局部低层次细节信息获取不充分,因此会导致分割结果粗糙。目前,在息肉分割任务中去使用CNN和Transformer结合的方法越来越流行了。ColonFormer [9]包含一个Transformer编码器和一个CNN解码器,用于高效准确的息肉分割。SwinE-Net [10]结合了基于CNN的EffcentNet和基于vit的Swin

Transformer [11], 通过应用多重扩张卷积、多特征聚合和专注的反卷积来分割息肉。这些 CNN-Transformer 结构通常利用 CNN 特征的详细高分辨率空间信息和 Transformer 的全局上下文来提高息肉分割性能。

基于上述讨论, 本文提出了一种基于 CNN 和 Swin Transformer 的双分支特征提取的医学图像分割网络(Dual branch feature extraction network based on CNN and Swin Transformer, DST-Net)。与几种最先进的网络相比, DST-Net 的主要创新内容如下:

1) DST-Net 包含一个双分支特征提取路径编码器, 其中一个是 VGG 编码器分支, 另一个是 AC (Atrous convolution, 空洞卷积) [12]编码器分支。该双分支编码器可以同时提取局部边界特征和上下文多尺度特征, 提高了网络在息肉分割中的特征表达能力。

2) 在编码器和解码器中间的底部模块融入两个连续的 Swin Transformer 模块, 充分利用 Transformer 的远程依赖关系进一步加强网络的全局特征提取能力。

3) 在第一个 U 型的跳跃连接引用通道注意力模块, 能精确定位息肉区域并抑制息肉背景的干扰, 提高分割的精度。

在 2 个标准的息肉分割基准数据集(CVC-Clinic DB, Kvasir)上进行了实验, 并比较了 DST-Net 与其它息肉分割方法的有效性。大量的实验表明, 该网络能对息肉进行精准分割。

2. 方法

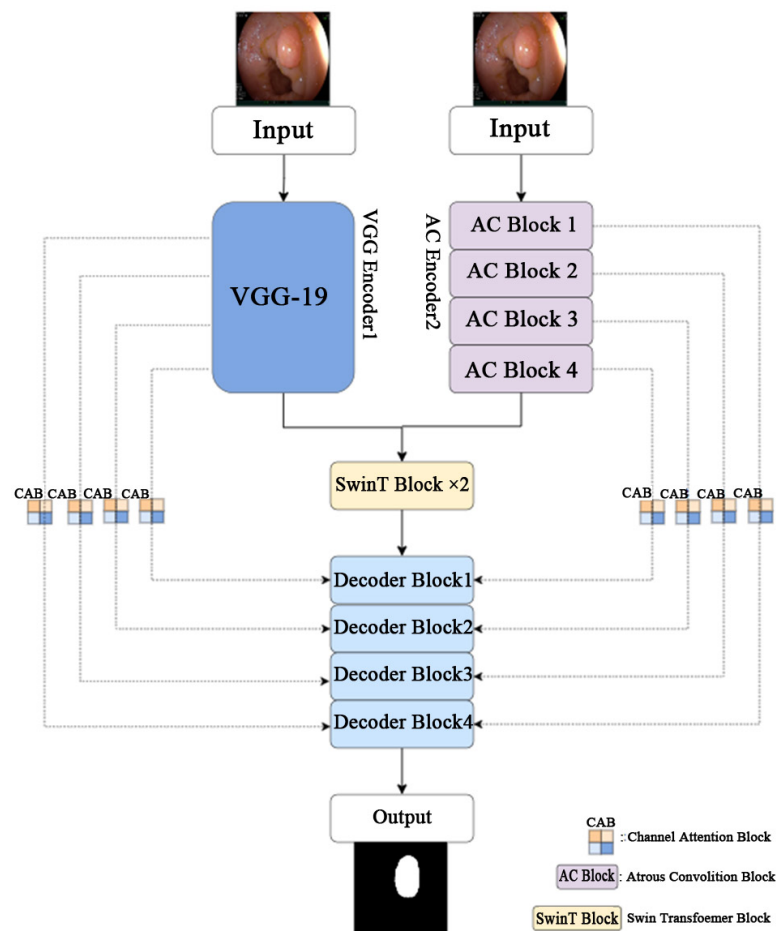


Figure 1. DST-Net network structure

图 1. DST-Net 网络结构

2.1. 网络架构框架

如图 1 所示, 整个框架分为使用基于 VGG 和空洞卷积的双分支特征提取路径编码器、Swin Transformer 底部模块和解码器。其中 VGG 编码器分支包含 5 个编码器块, 使用了预训练的 VGG-19 作为骨干网络, 经过 5 次下采样后, 最后将原图缩小了 32 倍。AC 编码器分支包含 5 个自定义的编码器块, 每个编码器块都是一个相同的残差结构, 包括 1 个空洞卷积模块加 BN 和激活函数, 和一个 2×2 的池化层。Swin Transformer 底部模块包含两个连续的 Swin Transformer 模块。解码器包含五个解码器子块, 由一个双线性插值上采样层和两个 3×3 的卷积层加 BN 组成。上采样的特征通道数为 256, 128, 64, 32, 16。在五次上采样中, 通过跳跃连接将双分支编码器不同层输出的低级语义特征使用通道注意力模块重塑后与上一层解码器子模块得到的高级语义特征拼接。

2.2. 空洞卷积模块(AC Block)

如图 2 所示, 空洞卷积模块对输入特征分别经过扩张率分别为 1, 6, 12, 18 的 3×3 空洞卷积、BN 和激活函数。然后对这四类特征进行拼接, 再经过 1×1 的卷积、BN 和激活函数处理。空洞卷积模块通过增强不同的感受野, 捕捉图像的上下文信息, 让提取到的特征具备上下文多尺度的特性。

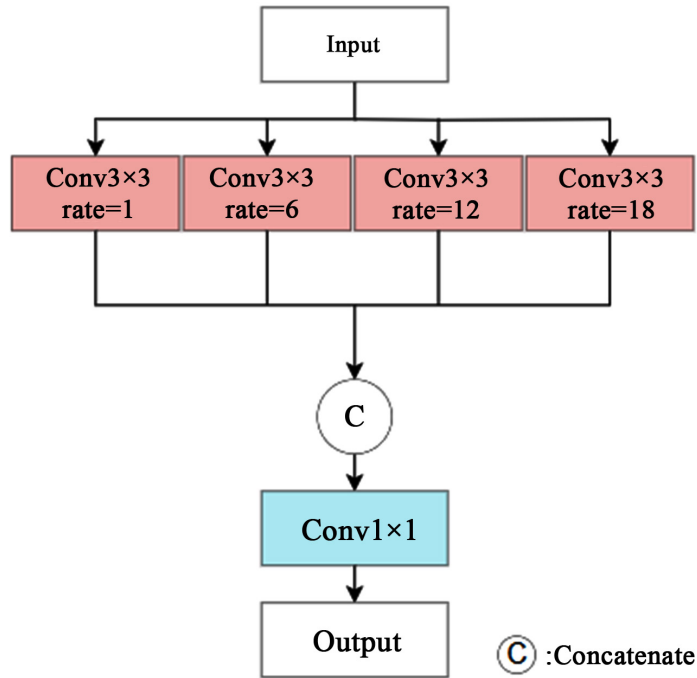


Figure 2. Atrous convolution module

图 2. 空洞卷积模块

其主要公式如下所示:

$$Output_{Branch_A} = \partial_{ReLU} \left(BN \left(AtrousConv_{1^{3 \times 3}} (Input) \right) \right) \quad (1)$$

$$Output_{Branch_B} = \partial_{ReLU} \left(BN \left(AtrousConv_{6^{3 \times 3}} (Output') \right) \right) \quad (2)$$

$$Output_{Branch_C} = \partial_{ReLU} \left(BN \left(AtrousConv_{12^{3 \times 3}} (Output') \right) \right) \quad (3)$$

$$Output_{Branch_D} = \hat{\sigma}_{ReLU} \left(BN \left(AtrousConv_{18}^{3 \times 3} (Output') \right) \right) \quad (4)$$

$$Output = \hat{\sigma}_{ReLU} \left(BN \left(Conv^{1 \times 1} \left(Conca \left(Output_{Branch_{A-D}} \right) \right) \right) \right) \quad (5)$$

上列公式中, $Input \in R^{H,W}$ 代表输入特征, $Conv^{1 \times 1}$ 代表 1×1 的普通卷积, BN 代表 Batch Normalization, $\hat{\sigma}_{ReLU}$ 代表 ReLU 函数, $Output_{Branch_{A-D}}$ 代表 A~D 4 个空洞卷积分支的输出, $AtrousConv_{1-18}^{3 \times 3}$ 代表 $rate = (1,6,12,18)$ 的 3×3 空洞卷积, $Conca$ 代表堆叠, \odot 代表哈达玛积, $Output \in R^{H,W}$ 代表最终输出。

2.3. Swin Transformer 底部模块

如图 3 所示, Swin Transformer 底部模块位于编码器和解码器中间, 将双分支特征提取路径编码器的输出特征经过两个连续的 Swin Transformer 模块, 利用 Swin Transformer 良好的全局上下文建模能力, 能够有效的学习特征的远程依赖性, 进一步增强了网络提取全局特征的能力。Swin Transformer 模块包括两个归一化层、一个窗口多头注意力模块和一个 MLP 层组成。Swin Transformer block 是成对使用的, 其中第一个 block 的多头注意力模块先使用一个 Window MSA (W-MSA) 结构, 第二个 block 使用 Shifted Window MSA (SW-MSA) 结构。两个连续的 Swin Transformer 模块流程公式可表述为:

$$\widehat{Output}^L = W_MSA \left(LN \left(Input^L \right) \right) + Input^L \quad (6)$$

$$Output^L = MLP \left(LN \left(\widehat{Output}^L \right) \right) + \widehat{Output}^L \quad (7)$$

$$\widehat{Output}^{L+1} = SW_MSA \left(LN \left(Output^L \right) \right) + Output^L \quad (8)$$

$$Output^{L+1} = MLP \left(LN \left(\widehat{Output}^{L+1} \right) \right) + \widehat{Output}^{L+1} \quad (9)$$

上列公式中, $Input^L$ 表示来自 L 层的 Swin Transformer 模块输入, $Output^{L+1}$ 代表 L + 1 层的 Swin Transformer 模块的输出, \widehat{Output}^L 和 \widehat{Output}^{L+1} 分别代表 L 层和 L + 1 层的多头注意力的输出与输入的相加。

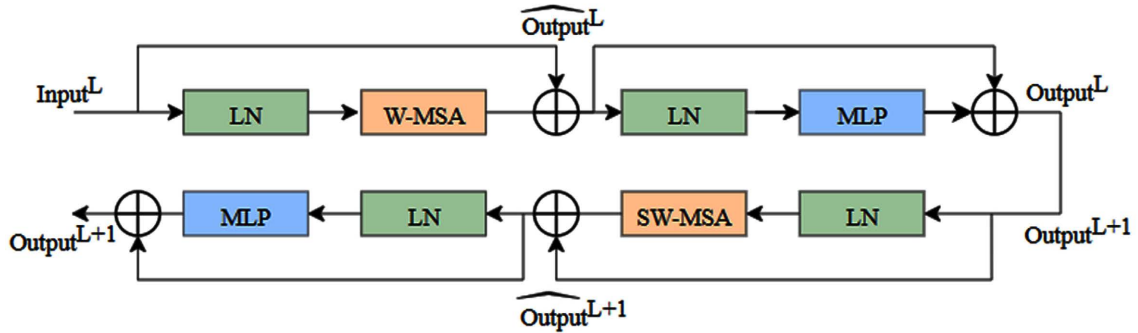


Figure 3. Two consecutive Swin Transformer modules

图 3. 两个连续的 Swin Transformer 模块

2.4. 通道注意力模块 (CAB)

如图 4 所示, CAB 对输入特征分别在通道域进行全局最大池化和全局平均池化。然后对这两类特征进行两个全连接层的处理, 得到两个值域分别为 0 和 1 的一维向量。将两个向量连接起来并与原始输入特征相乘, 最后再通过 Dense 层, 能够抑制背景的干扰, 增强感兴趣区域的特征响应。其主要公式如下所示。

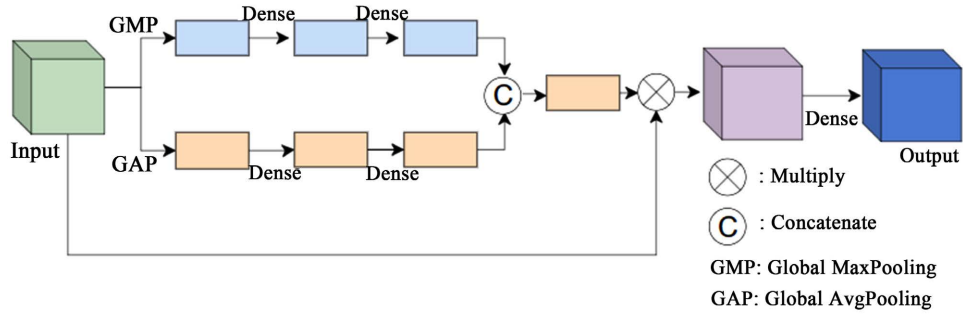


Figure 4. Channel attention module
图 4. 通道注意力模块

$$\widehat{Output}^{GMP} = Dense\left(Dense\left(F^{GMP}(Input)\right)\right) \quad (10)$$

$$\widehat{Output}^{GAP} = Dense\left(Dense\left(F^{GAP}(Input)\right)\right) \quad (11)$$

$$Output = Dense\left(\left(Concat\left(\widehat{Output}^{GMP}, \widehat{Output}^{GAP}\right)\right) \odot Input\right) \quad (12)$$

上列公式中， $Input \in R^{B \times H \times W \times C}$ 代表原始输入， \widehat{Output}^{GMP} 代表全局最大池化分支的输出， \widehat{Output}^{GAP} 代表全局平均池化分支的输出， F^{GMP} 代表全局最大池化函数， F^{GAP} 代表全局平均池化函数 Dense 代表全连接层，Concat 代表堆叠， \odot 代表哈达玛积， $Output \in R^{B \times H \times W \times C}$ 代表最终输出。

3. 实验

3.1. 数据集和对比模型

基于两个具有挑战性的公开息肉数据集进行评估，即 CVC-Clinic DB [13]，Kvasir [14]来验证提出模型的有效性。具体如下：

CVC-ClinicDB 数据集通过 31 个结肠镜检查视频中收集了 612 张息肉图像。

Kvasir 数据集通过内窥镜设备收集了 1000 张息肉图像。

为了评估网络的性能，将 DST-Net 与五个开源模型的分割结果进行对比，对比模型包括 U-Net、ResU-Net [15]、DoubleU-Net、PraNet [16]、SwinU-Net。为了公平的比较，所有的网络在相同的训练、验证和测试集上进行评估。

3.2. 实验设置

本文的实验基于 Ubuntu 18.04 LTS 64 位操作系统中的 python 3.7.0 和 Tensorflow + Keras 版本 2.5.0 环境。在训练过程中，优化器使用 Nadam，学习率 $lr_{base} = 1e-5$ ，损失函数使用二进制交叉熵。Batch size 设置为 4，迭代次数设置为 200。同时利用相应的回调函数，监控验证集损失，当损失 50 轮无法迭代时，停止迭代。实验借鉴了 PraNet 的训练设置，将 CVC-ClinicDB 和 Kvasir 的图像按照 8:1:1 随机分成训练集、验证集和测试集。所有方法都在带有 NVIDIA GTX3080Ti (8GB) 的服务器上进行。

3.3. 对比试验

本文使用 ClinicDB 和 Kvasir-SEG 数据集来评估所提出模型的学习能力。首先对训练数据集进行数据增强，通过水平翻转、垂直旋转、中心裁剪、弹性变换、高斯噪声和通道转置等方式对图像进行 12 次扩充。最终将训练数据扩充到 15660 张图片。从表 1 的结果可知，在 ClinicDB 数据集上，DST-Net 网络

在 Dice、mIoU 和 Precision 三项指标上均获得了最优的结果，分别达到了 0.929、0.879、0.971。其中在 Dice 和 mIoU 的评分上比次优模型 DoubleU-Net 高 3.0%和 2.5%，比 PraNet 高 3.1%和 3.0%。在 Recall 指标上 DST-Net 网络取得 0.840，只比最高的 DoubleU-Net 低 0.3%，高于 PraNet2.8%。在 Kvasir 数据集上，DST-Net 网络在 Dice、mIoU、Recall 和 Precision 四项指标上均获得了最优的结果，分别达到了 0.902、0.831、0.825 和 0.953。其中 Dic 和 mIoU 的评分上比次优模型 PraNet 高 0.5%和 0.01%，比 DoubleU-Net 高 0.07%和 1.4%。从对比实验结果可知，提出的模型优于现有的方法，证明了它通过统筹卷积和 Transformer 在局部和全局特征提取中各自的优势拥有了更好的学习能力和分割性能。图 5 给出在这两个数据集上 DST-Net 模型与比较模型的可视化结果。从图 5 中可知，DST-Net 的分割结果和 PraNet、DoubleU-Net 和 SwinU-Net 对比是最接近真实标签的。DST-Net 解决了 SwinU-Net 在局部边界特征提取能力上的不足，从图中可以明显看出 DST-Net 在边界细节上的处理更好。DST-Net 解决了 PraNet 和 DoubleU-Net 在全局上下文特征提取能力上的不足，DST-Net 对背景干扰的处理更好。DST-Net 在面对内窥镜采集和结肠镜检查视频这两种不同的采集方式依然能保持稳定的分割能力。

Table 1. Comparative experimental results

表 1. 对比实验结果

Dataset	Model	Dice	mIoU	Recall	Precision
CVC-ClinicDB	U-Net	0.773	0.753	0.731	0.956
	ResU-Net	0.768	0.732	0.740	0.942
	DoubleU-Net	0.899	0.854	0.843	0.965
	PraNet	0.898	0.849	0.812	0.967
	SwinU-Net	0.715	0.594	0.632	0.858
	Ours	0.929	0.879	0.840	0.971
Kvasir	U-Net	0.714	0.611	0.679	0.829
	ResU-Net	0.689	0.585	0.667	0.773
	DoubleU-Net	0.895	0.817	0.814	0.932
	PraNet	0.897	0.830	0.810	0.943
	SwinU-Net	0.629	0.474	0.612	0.677
	Ours	0.902	0.831	0.825	0.953

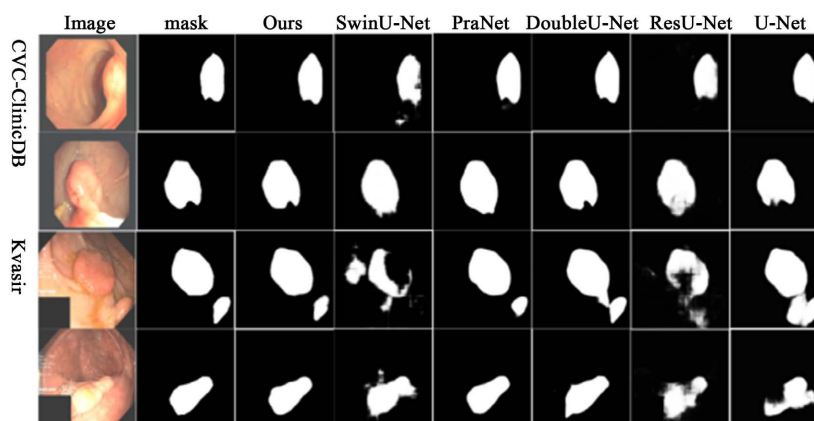


Figure 5. Comparison of experimental visualization

图 5. 对比试验可视化

3.4. 消融实验

为了证明该网络的有效性，进一步对 DST-Net 进行了消融实验。消融实验通过逐渐去除与 DST-Net 网络相关的组件来进行。

验证通道注意力模块的有效性。在 DST-Net 中去除通道注意力模块形成 DST-CAB 网络。如表 2 所示, DST-CAB 的 Dice、mIoU、Recall 和 Precision 四项指标在两个数据集上分别达到了 0.921、0.867、0.837、0.964 和 0.817、0.820、0.949。其中 Dice 得分分别低于 DST-Net 约 0.8% 和 0.6%, mIoU 约 1.2% 和 1.4%, 通过对比验证了通道注意力模块的有效性。

验证空洞卷积模块的有效性。在 DST-CA 中去除空洞卷积模块形成 DST-CAB-ACB 网络。如表 2 所示, DST-CAB-ACB 的 Dice、mIoU、Recall 和 Precision 四项指标在两个数据集上分别达到了 0.913、0.857、0.815、0.958 和 0.887、0.804、0.814、0.931。其中 Dice 得分分别低于 DST-CA 约 0.8% 和 0.9%, mIoU 约 1.0% 和 1.3%, 通过对比验证了空洞卷积模块的有效性。

验证 Swin Transformer 底部模块的有效性。在 DST-CAB-ACB 中去除 Swin Transformer 底部模块形成 DST-CAB-ACB-ST 网络。如表 2 所示, DST-CAB-ACB-ST 的 Dice、mIoU、Recall 和 Precision 四项指标在两个数据集上分别达到了 0.906、0.831、0.816、0.953 和 0.881、0.802、0.801、0.930。其中 Dice 得分分别低于 DST-CA-PIFM 约 0.7% 和 0.6%, mIoU 约 2.6% 和 0.2%, 通过对比验证了 Swin Transformer 底部模块的有效性。

此外, 通过图 6 消融实验可视化可以观察到 DST-Net 的预测结果是最接近真实标签的。DST-CAB 网络和 DST-CAB-ACB 网络的预测结果在边界的处理上稍差一点。DST-CAB-ACB-ST 网络的预测结果是图中最差的。

Table 2. Ablation results
表 2. 消融实验结果

Dataset	Model	Dice	mIoU	Recall	Precision
CVC-ClinicDB	DST-CA-AC-ST	0.906	0.831	0.816	0.953
	DST-CA-AC	0.913	0.857	0.815	0.958
	DST-CA	0.921	0.867	0.837	0.964
	Ours	0.929	0.879	0.840	0.971
Kvasir	DST-CA-AC-ST	0.881	0.802	0.801	0.930
	DST-CA-AC	0.887	0.804	0.814	0.931
	DST-CA	0.896	0.817	0.820	0.949
	Ours	0.902	0.831	0.825	0.953

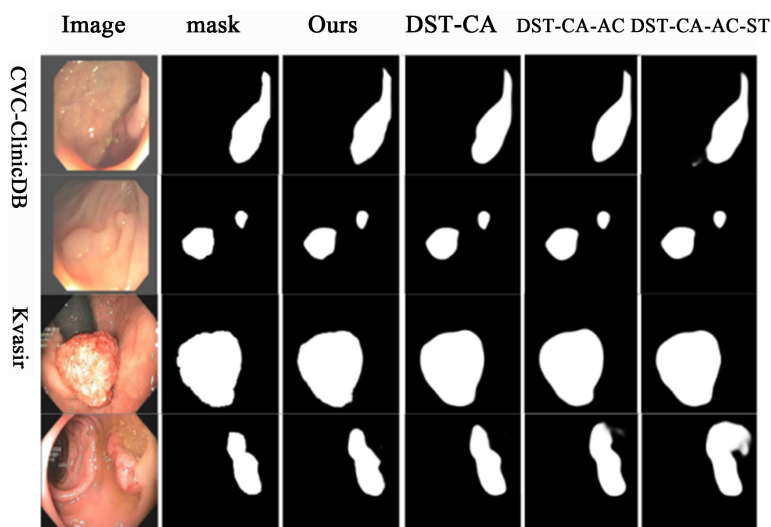


Figure 6. Ablation of experimental visualization
图 6. 消融试验可视化

4. 总结

本文提出了一种基于 CNN 和 Swin Transformer 的双分支特征提取的医学息肉分割网络(DST-Net), 通过统筹卷积神经网络和 Transformer 在提取局部特征和全局特征各自的优势, 能够有效提高息肉分割的精度。我们设计了结合 VGG 和空洞卷积(AC)的双分支编码器, 在 VGG 通过卷积运算提取局部细节特征的基础上, 利用空洞卷积更大的感受野去补充特征的上下文多尺度信息, 以此增强网络对息肉特征的表达能力。我们还设计了 Swin Transformer 底部模块, 通过 Transformer 的远程依赖关系进一步增强网络的全局语义特征提取能力。我们在跳过连接使用了通道注意力模块(CAB), 能精确定位息肉区域并抑制息肉背景的干扰, 提高分割的精度。本文提出的 DST-Net 在 CVC-Clinic DB 和 Kvasir 两个公开的息肉数据集上都达到了最先进的性能, 充分展示了强大的学习能力和出色的泛化能力。对预测结果的定性和定性分析表明, DST-Net 可以促进计算机辅助诊断和治疗的发展。在未来的工作中, 将考虑并推动轻量化模型进行更多的病灶分割。

基金项目

国家自然科学基金(61801288)。

参考文献

- [1] Xia, C., Dong, X., Li, H., *et al.* (2022) Cancer Statistics in China and United States, 2022: Profiles, Trends, and Determinants. *Chinese Medical Journal*, **135**, 584-590. <https://doi.org/10.1097/CM9.0000000000002108>
- [2] 张恒良, 李镡, 关欣. 一种改进的三维双路径脑肿瘤图像分割网络[J]. 光学学报, 2021, 41(3): 0310002.
- [3] 黄鸿, 彭超, 吴若愚, 等. 基于部分注释 CT 图像的自监督迁移学习肺结节分类[J]. 光学学报, 2020, 40(18): 1810003.
- [4] 李大湘, 张振. 基于改进 U-Net 视网膜血管图像分割算法[J]. 光学学报, 2020, 40(10): 1010001.
- [5] 王亚刚, 郝怡媛, 潘晓英. 改进 DeepLabv3 +网络的肠道息肉分割方法[J]. 计算机科学与探索, 2020, 14(7): 1243-1250.
- [6] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [7] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., *et al.* (2018) UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov, D., *et al.*, Eds., *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, Cham, 3-11. https://doi.org/10.1007/978-3-030-00889-5_1
- [8] Jha, D., Smedsrud, P.H., Riegler, M.A., *et al.* (2019) ResUNet++: An Advanced Architecture for Medical Image. 2019 *IEEE International Symposium on Multimedia (ISM)*, San Diego, 09-11 December 2019, 225-2255. <https://doi.org/10.1109/ISM46123.2019.00049>
- [9] Duc, N.T., Oanh, N.T., Thuy, N.T., *et al.* (2022) ColonFormer: An Efficient Transformer based Method for Colon Polyp Segmentation. *IEEE Access*, **10**, 80575-80586. <https://doi.org/10.1109/ACCESS.2022.3195241>
- [10] Park, K.-B. and Lee, J.Y. (2022) SwinE-Net: Hybrid Deep Learning Approach to Novel Polyp Segmentation Using Convolutional Neural Network and Swin Transformer. *Journal of Computational Design and Engineering*, **9**, 616-632. <https://doi.org/10.1093/jcde/qwac018>
- [11] Liu, Z., Lin, Y., Cao, Y., *et al.* (2021) Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 9992-10002. <https://doi.org/10.1109/ICCV48922.2021.00986>
- [12] Chen, L.C., Papandreou, G., Schroff, F., *et al.* (2017) Rethinking Atrous Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV 2018. Lecture Notes in Computer Science*, Springer, Cham, 833-851. https://doi.org/10.1007/978-3-030-01234-2_49
- [13] Bernal, J., Sánchez, F., Fernández-Esparrach, G., *et al.* (2015) WM-DOVA Maps for Accurate Polyp Highlighting in Colonoscopy: Validation vs. Saliency Maps from Physicians. *Computerized Medical Imaging and Graphics*, **43**, 99-111. <https://doi.org/10.1016/j.compmedimag.2015.02.007>

- [14] Jha, D., Smedsrud, P.H., Riegler, M.A., *et al.* (2020) Kvasir-SEG: A Segmented Polyp Dataset. In: Ro, Y., *et al.*, Eds., *MultiMedia Modeling, Lecture Notes in Computer Science*, Springer, Cham, 451-462. https://doi.org/10.1007/978-3-030-37734-2_37
- [15] Xiao, X., Lian, S., Luo, Z., *et al.* (2018) Weighted Res-UNet for High-Quality Retina Vessel Segmentation. 2018 *9th International Conference on Information Technology in Medicine and Education (ITME)*, Hangzhou, 19-21 October 2018, 327-331. <https://doi.org/10.1109/ITME.2018.00080>
- [16] Fan, D.P., Ji, G.P., Zhou, T., *et al.* (2020) PraNet: Parallel Reverse Attention Network for Polyp Segmentation. In: Martel, A.L., *et al.*, Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020. Lecture Notes in Computer Science*, Springer, Cham, 263-273. https://doi.org/10.1007/978-3-030-59725-2_26