倪振鸿, 金晅宏. 基于混合注意力逐步解码的息肉图像分割[J]. 智能计算机与应用,2026,16(7):193-198. DOI:10.20169/j.issn.2095-2163.24092001

基于混合注意力逐步解码的息肉图像分割

倪振鸿, 金晅宏
(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘 要: 目前深度学习已成为息肉图像分割任务的热门方法。其中,由于 Transformer 能够通过自关注捕获远程依赖关系,在图像分割中取得了良好的效果,然而 Transformer 具有计算量大,无法学习像素间局部上下文等问题,而且结直肠息肉图像具有形状差异大、边界模糊等限制。本文针对这些挑战,提出了一种新的基于混合注意力逐步解码(HASD)的模型,模型结构主要包括:跳过连接增强模块、用于增强跳过连接中的局部细节特征;逐级融合模块、将经过上采样和跳过连接模块的多尺度特征相融合;注意力聚合模块、使用混合注意力聚合浅层和深层特征。本文在 5 个公开的息肉图像数据集上对所提出的模型进行评估,实验表明本文提出的新方法分割准确率优于先进方法。

关键词: 息肉图像分割; 多尺度特征融合; 混合注意力; Transformer; 跳过连接
中图分类号: TP391.41 **文献标志码**: A **文章编号**: 2095-2163(2026)07-0193-06

Polyp image segmentation with hybrid attention stepwise decoding

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

Abstract: Currently, deep learning has become a popular method for polyp image segmentation tasks. Transformer have achieved good results in image segmentation due to the ability to capture remote dependencies through self-attention. However, Transformer have many problems such as large computational complexity and inability to learn local context between pixels, and colorectal polyp images have limitations such as large shape differences and fuzzy boundaries. To address these challenges, this paper proposes a new model structure based on Hybrid Attention Stepwise Decoding (HASD). The specific structure mainly includes: skip connection enhancement module, which is used to enhance the local detail features in skip connections; a step-by-step fusion module, which is used to fuse multi-scale features that undergo upsampling and skip connection modules; Attention aggregation module, which uses hybrid attention to aggregate shallow and deep features. In this paper, the proposed model is evaluated on five public polyp image datasets. Experiments show that the accuracy of the new method proposed in this paper is better than the most advanced methods.
Key words: polyp image segmentation; multi-scale feature fusion; hybrid attention; Transformer; skip connection

0 引 言

结直肠癌是恶性肿瘤中第三高发病率和死亡率的癌症[1]。大部分的结直肠癌起源于结肠息肉,使用结肠镜检测是早期检查的标准方法,可以及时检测和切除息肉,大大降低结直肠癌的发病率。但人工结肠镜检查的准确性受限于操作者的水平及经验,部分息肉容易被漏检。最近有研究证明使用计算机辅助诊断对结肠息肉进行自动化准确分割[2],可以帮助临床医生更高效地定位息肉,提高诊断和治疗过程的效率。但是目前息肉分割算法的精度受到了息肉的形态多样以及其与周围粘膜之间边界模糊的限制,这些问题造成了息肉分割效果不佳。

最近,计算机视觉技术的发展促使了深度学习在息肉图像分割任务中衍生出多种模型,相比于传统的医学分割方法,基于深度学习的分割方法可以有效提高息肉分割的准确性和效率,其中基于卷积神经网络的方法已被广泛应用,特别是 U-Net[3],利

作者简介: 倪振鸿(1999—),男,硕士研究生。主要研究方向:图像处理。
通信作者: 金晅宏(1978—),女,副教授。主要研究方向:信号获取与处理,数字图像处理与技术应用,虚拟仪器技术及系统集成技术等。Email:xuanh_jin@usst.edu.cn。

收稿日期: 2024-09-20

用独特的跳过连接方案,通过将不同尺度的编码器特征与解码器中相应的特征图连接起来,可以有效地定位息肉边缘轮廓和提取息肉图像中的小结构,能够恢复编码器下采样过程中丢失的上下文信息。但是 U-Net 仍然存在不能建立任意像素之间的关联进而实现全局特征学习、还有跳过连接中直接连接编码器和解码器特征时引起的语义差距等问题。随着时间的推移,许多 U-Net 的变体被提出用于各种医学图像分割任务,如 UNet++[4]、UNet 3+[5]、nnU-Net[6]等,尽管这些算法在图像分割任务中表现出了显著改善,但是在图像中建模全局关系的能力依旧受到卷积算子局部性质的限制。

为了克服这一限制,许多研究尝试将 Transformer 结构引入到息肉分割任务中。Transformer 的自注意机制可以有效地弥补卷积的局限性,同时为了降低自注意机制的计算量和减少冗余,有些研究引入了 Vision Transformer[7]、Swin Transformer[8]等自注意力结构。尽管这些改进的模型有效地降低计算成本且提高了息肉图像分割的准确率,但是自注意力机制会破坏部分局部特征,限制了模型学习图像关键局部特征的能力。

所以为了高效结合 CNN 和 Transformer,充分保留全局和局部特征,最近,则相继提出了许多混合架构。首先 Segformer[9]、PVT[10]等部分模型尝试在 Transformer 编码器中嵌入卷积层来弥补缺点,但效果有限。还有很多模型[11]尝试在解码器中加入各种卷积模块,引导模型学习局部特征的同时融合多尺度特征,比如:U-Net v2[12]提出了一种新的跳过连接结构(SDI),通过每一层的特征映射 Hadamard 积[13]其他层的特征,进而增强每一层特征的语义信息和细节。Wang 等学者[14]提出 SSFormer,采用逐步自适应的方法,强调局部特征,减少注意力分散。Tang 等学者[15]提出 DUAT 网络,利用 GLSA 模块聚合全局和局部的空间信息,使用 SBA 模块结合浅层和深层特征。Rahman 等学者[16]提出级联注意解码器(CASCADE),分别使用注意门和卷积注意模块聚合多阶段特征,此后又提出基于图卷积的级联卷积注意解码器(GCASCADE)[17],使用图卷积操作来增强特征映射。

为了进一步高效结合卷积和注意力机制的优势,本文提出一种新的解码器结构。该解码器由用于增强局部特征的跳过连接增强模块、逐级融合模块和基于注意力机制的聚合模块组成。本文使用的跳过连接增强模块能够抑制背景噪声,定位关键信息和增强局部特征,解决了息肉图像对比度低的问题。多阶段逐级融合可以高效地引导不同深度的特征相互融合,聚合模块充分混合通道注意力权重和空间注意力权重,保证在融合浅层和深层特征的同时关注更多有用特征,防止丢失细节信息,提高分割性能。

1 方法

1.1 整体网络结构

为了提高息肉图像分割效果,本文提出了一种模型框架如图 1 所示,整体结构主要由编码器和解码器两个部分组成,模型输入为 3 × 352 × 352 的息肉图像。

在编码器部分,选择了 PVT 的改进版 PVT v2[18],从编码器的不同阶段获取到 4 层多尺度特征图,相关的输出特征图大小分别是原始图片的 $\frac{1}{4}$、$\frac{1}{8}$、$\frac{1}{16}$、$\frac{1}{32}$,第 1 层的浅层特征提供了详细的外观信息,而第 2、3、4 层的特征则提供了高级特征。解码器部分主要由跳过连接增强模块、逐级融合模块和注意力聚合模块组成。在跳过连接采用本文提出的新型卷积增强模块加强局部特征,提取出关键信息,其次引入逐级融合模块将深层不同尺度的特征融合,引导高低层特征信息互补,逐步补充细节。最后,引入新型的注意力聚合模块,将第 1 层浅层特征输出和融合后的深层特征进行聚合,充分利用混合注意力机制将浅层的边界信息和深层的语义信息结合,得到解码器的最终预测图。

1.2 卷积增强模块

文献[14]提出 Transformer 编码器中的全局特征不断融合会导致注意力分散的问题,为了进一步解决编码器中的这些问题并且加强局部特征,在跳过连接中本文采用了新型的卷积增强模块(CE),整个模块由 3×3 卷积模块和空间注意力模块组成,其中注意力机制有助于捕捉到关键的空间信息,抑制背景信息,此模块可以更有效地混合相邻 patch 的特征。整个模块如图 1(d)所示,卷积单元由 3×3 卷积算子、激活函数组成,空间注意力单元由通道维度平均池化和最大池化、7×7 卷积算子和激活函数组成。2 种单元分别采用不同尺度的卷积单元,可以从多个尺度上获取输入信息,提取出更丰富、更全面的特征图。整个模块公式如下:

$$CE_{234}(x) = \text{Convblock}(x) + \text{Spablock}(x) \quad (1)$$

其中，x 表示每一层跳过连接的输入张量，输入张量分别进入 2 个单元相加后输出；$CE_{234}(x)$ 表示第 2、3、4 层输入特征的卷积增强模块，该模块融合了 2 个独立单元的结果。卷积模块（Convblock）由激活函数 ReLU(·)和 $C_{3\times3}(\cdot)$ 组成；$C_{3\times3}(\cdot)$ 是卷积核为 3×3 且填充为 1 的卷积层，用于提取局部特征，模块如图 1(e)所示。Spablock 主要由空间注意力单元组成，具体公式如下：

$$\mathrm{Spablock}(x)=\mathrm{ReLU}(C_{1\times1}(\mathrm{Spa}(x))) \quad (2)$$

$$\mathrm{Spa}(x)=\sigma(C_{7\times7}(C_{\mathrm{M}}(x)+C_{\mathrm{A}}(x)))\times x \quad (3)$$

其中，$C_{\mathrm{M}}(\cdot)$ 和 $C_{\mathrm{A}}(\cdot)$ 分别表示通道维度的最大池化和平均池化；$\sigma(\cdot)$ 表示 Sigmoid 激活函数；$C_{1\times1}(\cdot)$ 表示卷积核为 1×1 且填充为 1 的卷积层，用于调整输入特征图的通道维度；$C_{7\times7}(\cdot)$ 表示卷积核为 7×7 且填充为 3 的卷积层，用于拼接最大池化和平均池化产生的特征图。卷积增强模块主要用于抑制背景噪声，强调图像中息肉周围的关键局部信息。

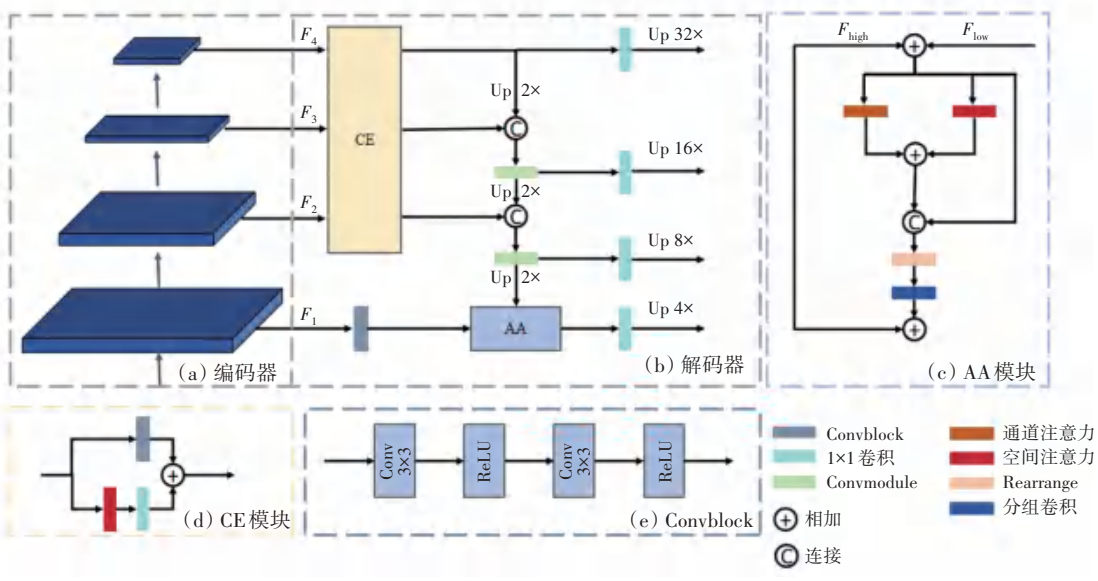


图 1　模型整体结构

Fig. 1　Overall model structure

1.3　逐级融合模块

逐级融合模块（SFM）用于将不同阶段的特征逐步融合，最后将融合后的特征图输入注意力聚合模块。具体公式如下：

$$\mathrm{SFM}(L_3,L_4)=\mathrm{ConvModule}(\mathrm{Concat}(L_3,\mathrm{Up}(L_4))) \quad (4)$$

其中，(L_3,L_4) 表示跳过连接输出的特征和解码器上一级融合的特征；Up 表示上采样操作；ConvModule 函数中包含了卷积、激活函数、批归一化操作。

该模块具体步骤为：首先，对当前层进行上采样操作和下一层特征维度匹配，然后在通道维度进行拼接，使用卷积操作融合 2 层特征，将融合后特征和下一级特征进行重复的相同操作。该模块不仅计算量小、而且可以充分引导编码器中不同深度间的信息交互，高效地融合高低层特征。

1.4　注意力聚合模块

通道注意力和空间注意力分别在通道和空间维度上计算注意力权重，能够获取不同通道和区域的重要程度，有效地针对不同任务抑制无关信息，提高了模型性能。文献[19]认为 2 个注意力权值间需要信息交换，提出了内容引导注意力方法（CGA），通过混合 2 个注意力权重生成初始 SIMs，再使用初始特征进一步细化，为每个通道分配单独的 SIM。

Transformer 编码器输出的特征中浅层特征有丰富的细节和清晰的边界信息，但高级语义信息较少。同时深层特征感受野更大，包含丰富的语义信息。然而因其感受野不同，直接融合可能会导致感受野失配和冗余。

为了解决这个问题，本文引入上文提及的 CGA 模块，并进行了进一步的改良和实验。具体来说，删除了其中一个多余的 Sigmoid 函数并在跳过连接阶段减去了浅层特征输入等，减少了浅层特征的冗余信息输入。模块的整体结构见图 1。在高效融合特征的同时解决了息肉图像边界模糊、对比度低等问题。

先将编码器输入的第一层特征通过卷积增强局部特征并减少通道数量，并将逐级融合模块输出通

过上采样模块与其尺寸匹配,2 个特征相加后送到注意力聚合模块(AA)中。模块又将特征分别送入 2 个注意力单元、即空间注意力(SA)和通道注意力(CA),此后相加计算出初始权重 S, 2 个单元公式如下:

$$\mathrm{Ca}(i)=C_{1\times1}(\mathrm{ReLU}(C_{1\times1}(S_{\mathrm{AP}}(i)))) \quad (5)$$

$$\mathrm{Sa}(i)=\sigma(C_{7\times7}(C_{\mathrm{M}}(i)+C_{\mathrm{A}}(i))) \quad (6)$$

其中, i 表示 2 个输入特征之和, S_{AP} 表示空间维度上自适应平均池化。为了降低计算复杂度,第 1 个 $C_{1\times1}$ 操作将通道数降低至八分之一,第 2 个 $C_{1\times1}$ 将通道数变回原先数量。

2 个模块通过简单的加法操作得到初始权重 S,最后使用输入特征在通道维度上引导生成最终权重值 W, 定义公式如下:

$$W=\sigma(\mathrm{GC}_{7\times7}(R_E(\mathrm{Concat}(i,S)))) \quad (7)$$

研究中将输入特征 i 和初始权重 S 在第二维度扩展,将两者拼接后使用重新排列函数 R_E 将张量变换为 $R^{(C\times2)\times H\times W}$, 在此基础上使用 7×7 分组卷积得到跟原先同样大小的特征,Sigmoid 激活函数用于将输出转换成 0~1 之间的权重。

经过上述单元得到最终权重后,将输入特征 F_{high} 和 F_{low} 加权求和,添加跳跃连接后经过 1×1 卷积投影得到输出特征 F, 公式表示为:

$$F=C_{1\times1}(WF_{\mathrm{high}}+(1-W)F_{\mathrm{low}}+F_{\mathrm{high}}) \quad (8)$$

总体而言,该模块基于混合注意力机制高效聚合浅层特征的边界信息和深层特征的语义信息,输出更清晰的息肉图像轮廓和更精准的位置信息。

2 实验

在本节中,首先介绍本文使用的数据集和评价指标,然后将提出的新型解码器的结果与先前的先进方法进行比较,以证明本文提出的模型的优越性,并且进行消融实验来评估本文提出的模型结构的有效性。

2.1 实验数据集和评价指标

研究使用 5 个公开的息肉图像分割数据集来评估所提的模型,这些数据集分别为 Kvasir-SEG[20]、CVC - ClinicDB[21]、ColonDB[22]、EndoScene[23] 和 ETIS-LaribDB[24],使用的训练方法与 PraNet[25] 相同,在 CVC-ClinicDB 和 Kvasir 数据集中取出 900 张和 548 张图像作为训练集。在剩余图像和另外 3 个数据集上进行测试模型的性能。实验中使用平均相似性系数(mDice)和平均交并比(mIoU)来评估模型。

2.2 实验环境

本文使用 PyTorch 在 RTX 3090 GPU 远程服务器上训练和测试所有模型。优化器使用 AdamW,初始学习率为 0.000 1,衰减率为 0.1,衰减周期为 40 个 epoch,最大训练周期为 100,批大小为 16(训练 GCASCADE 模型时为 4)。本文首先将图像大小调整为 352×352,应用多种数据增强手段增加数据多样性,使用 CASCADE 模型[16]中的训练策略,使用多尺度{0.75,1.00,1.25},梯度裁剪为 0.5,损失函数使用加权 IoU 和加权 BCE 的组合。

2.3 结果

本文在 5 个不同的息肉分割测试集上评估了本文所提的模型的性能和泛化性,除了 PraNet[25] 使用 ResNet 作为预训练编码器,其他模型使用了相同的预训练 Transformer 编码器 PVTv2[18],表 1 中显示了在 5 个数据集上与其他方法的比较结果。由表 1 可知,相比于现有模型,本文提出的模型泛化性能更好,除了 ClinicDB 数据集外,在 4 个数据集上的表现优于其他 5 种模型,mDice 分数分别提高了 0.25%、0.89%、1.08%、1.98%,mIoU 分数则分别提高了 0.31%、1.08%、0.95%、1.52%。图 2 展示了各种模型在 ETIS 数据集上的预测结果图,这些可视化图证明了本文所提的模型能够准确地定位到息肉边界,拥有很强的泛化能力和准确性。这些实验结果说明了提出的解码器结构可以有效地结合 Transformer 编码器和混合注意力的优势,每个输出特征能够获得更多的边界信息和细节信息,在未知数据集上具有更强的鲁棒性。

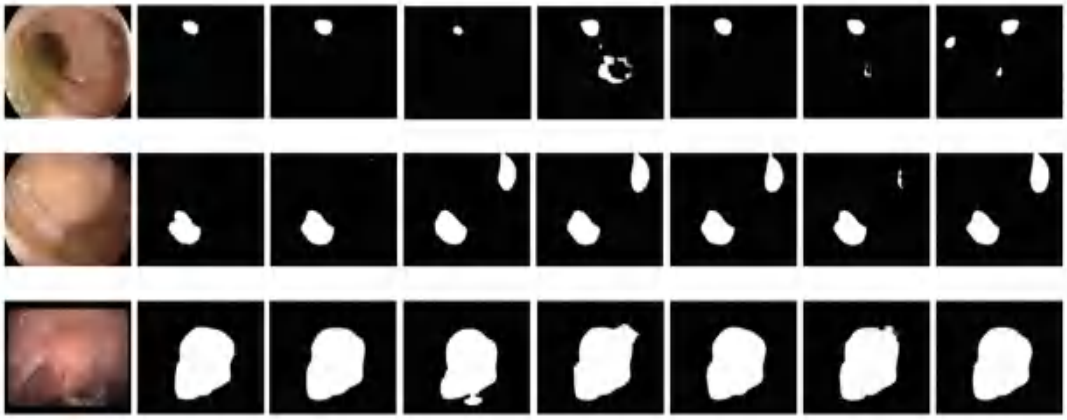
2.4 消融实验分析

本文在 5 个数据集上进行消融实验,以证明每个模块的有效性。训练和测试的配置和 2.2 节相同。实验结果见表 2,本文使用 PVTv2 作为基准模型,在模型中去除不同的模块并比较实验结果。表 2 中,“SF”、“CE”和“AA”分别表示逐级融合模块、卷积增强模块和注意力聚合模块,提出的标准模型为“Baseline+SF+CE+AA”。从表 2 中可以看出,在基准模型上增加本文研发的模块有助于提升模型性能,在增加“SF”基础上再加“AA”模块,除了在 Kvasir 数据集性能有所降低外,其他数据集性能都有提高。只有当 3 个模块同时存在时,整个模型的性能达到最高。

表 1　实验结果

Table 1　Experimental results

Dataset	Kvasir-SEG[20]		EndoScene[23]		CVC-ClinicDB[21]		ColonDB[22]		ETIS-LaribDB[24]	
	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU
PraNet	91.04	85.50	89.15	81.92	92.32	87.35	73.81	66.46	66.34	59.71
U-Net v2	91.83	86.59	89.47	82.52	93.18	88.33	81.69	73.51	78.21	70.07
SSFormer-S	92.19	87.34	89.20	82.10	93.77	89.01	81.12	73.07	79.12	71.54
CASCADE	92.34	87.36	89.44	82.70	94.25	89.65	81.65	73.02	79.08	71.30
GCASCADE	91.86	86.99	88.90	82.07	94.20	89.50	81.06	72.86	74.87	67.09
HASD	92.59	89.27	90.33	83.60	93.75	89.27	82.77	74.46	81.10	73.06



(a) 原图　(b) gt　(c) 本文模型　(d) PraNet　(e) GCASCADE　(f) U-Net v2　(g) SSFormer　(h) CASCADE

图 2　不同方法的分割结果

Fig. 2　Segmentation results of different methods

表 2　消融实验结果

Table 2　Results of ablation experiments

Components	Kvasir-SEG[20]		EndoScene[23]		CVC-ClinicDB[21]		ColonDB[22]		ETIS-LaribDB[24]	
	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU
Baseline	90.49	84.60	87.51	80.22	90.61	85.25	77.18	69.35	75.42	67.50
+SF	91.50	86.16	87.90	80.53	91.30	85.76	78.33	70.39	76.71	68.51
+SF+AA	90.77	85.19	88.34	81.12	91.94	86.80	79.07	70.70	78.12	70.31
+SF+CE	92.37	87.48	89.25	82.37	92.66	87.98	81.40	73.04	77.16	69.46
+SF+CE+AA	92.59	89.27	90.33	83.60	93.75	89.27	82.77	74.46	81.10	73.06

3　结束语

本文的主要成果有 3 个方面：

(1)本文提出了一种新的息肉分割模型、称为 HASD,不同于基于 U-Net 的方案,研究采用金字塔视觉 Transformer 作为编码器提取特征。

(2)在解码器中引入了 3 个新模块。具体来讲,在跳过连接中的卷积增强模块用于增强局部特征和统一特征图通道维度,逐级融合模块将输入的 3 个多尺度特征进行逐步融合,再由注意力聚合模块将浅层和深层特征聚合,使用混合注意力权重强调更多细节信息。

(3)本文将提出的模型在 5 个基准数据集上进行实验评估,实验表明该模型优于目前先进的方法。

参考文献

[1] 孙福艳，王琼，吕宗旺，等. 深度学习在结肠息肉分割中的应用综述[J]. 计算机工程与应用,2023,59(23):15-27.

[2] Quan S Y, Wei M T, Lee J, et al. Clinical evaluation of a real-time artificial intelligence-based polyp detection system: A US multi-center pilot study[J]. Scientific Reports, 2022, 12(1): 6598.

[3] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]// Proceedings of

the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015). Cham: Springer, 2015: 234-241.

[4] Zhou Z, Rahman S M M, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation [C]// Proceedings of 4th International Workshop on Deep Learning in Medical Image Analysis and 8th International Workshop on Multimodal Learning for Clinical Decision Support. Cham: Springer, 2018: 3-11.

[5] Huang H, Lin L, Tong R, et al. Unet 3+: A full-scale connected unet for medical image segmentation[C]// Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2020). Piscataway, NJ: IEEE, 2020: 1055-1059.

[6] Isensee F, Petersen J, Klein A, et al. nnU-Net: Self-adapting framework for u-net-based medical image segmentation[J]. arXiv preprint arXiv, 1809.10486, 2018.

[7] Dosovitskiy A. An image is worth 16×16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv, 2010.11929, 2020.

[8] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2021: 10012-10022.

[9] Xie E, Wang W, Yu Z, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers[J]. Advances in Neural Information Processing Systems, 2021, 34: 12077-12090.

[10] Wang W, Xie E, Li X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2021: 568-578.

[11] Dong B, Wang W, Fan D P, et al. Polyp-PVT: Polyp segmentation with pyramid vision transformers[J]. arXiv preprint arXiv, 2108.06932, 2021.

[12] Peng Y, Sonka M, Chen D Z. U-Net v2: Rethinking the skip connections of U-Net for medical image segmentation[J]. arXiv preprint arXiv, 2311.17791, 2023.

[13] Horadam K J. Hadamard matrices and their applications[M]. Princeton: Princeton University Press, 2012.

[14] Wang J, Huang Q, Tang F, et al. Stepwise feature fusion: Local guides global[C]//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2022: 110-120.

[15] Tang F, Xu Z, Huang Q, et al. DuAT: Dual-aggregation transformer network for medical image segmentation [C]// Proceedings of Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Cham: Springer, 2023: 343-356.

[16] Rahman M M, Marculescu R. Medical image segmentation via cascaded attention decoding[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway, NJ: IEEE, 2023: 6222-6231.

[17] Rahman M M, Marculescu R. G-CASCADE: Efficient cascaded graph convolutional decoding for 2D medical image segmentation [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway, NJ: IEEE, 2024: 7728-7737.

[18] Wang W, Xie E, Li X, et al. PVT V2: Improved baselines with pyramid vision transformer[J]. Computational Visual Media, 2022, 8(3): 415-424.

[19] Chen Z, He Z, Lu Z M. DEA-Net: Single imagedehazing based on detail-enhanced convolution and content-guided attention[J]. IEEE Transactions on Image Processing, 2024, 30: 443-447.

[20] Jha D, Smedsrud P H, Riegler M A, et al. Kvasir-SEG: A segmented polyp dataset [C]// Proceedings of the 26th International Conference on MultiMedia Modeling. Cham: Springer, 2020: 451-462.

[21] Bernal J, Sánchez F J, Fernández-Esparrach 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.

[22] Tajbakhsh N, Gurudu S R, Liang J. Automated polyp detection in colonoscopy videos using shape and context information[J]. IEEE Transactions on Medical Imaging, 2015, 35(2): 630-644.

[23] Vázquez D, Bernal J, Sánchez F J, et al. A benchmark for endoluminal scene segmentation of colonoscopy images [J]. Journal of Healthcare Engineering, 2017, 2017: 4037190.

[24] 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: 283-293.

[25] Fan D P, Ji G P, Zhou T, et al. PraNet: Parallel reverse attention network for polyp segmentation [C]//Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention. Cham: Springer, 2020: 263-273.