CN 51-1346/O4 ISSN 1003-501X (印刷版) ISSN 2094-4019 (网络版)

光电工程
Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

结合极化自注意力和Transformer的结直肠息肉分割方法

谢斌，刘阳倩，李俞霖

引用本文:

谢斌，刘阳倩，李俞霖. 结合极化自注意力和Transformer的结直肠息肉分割方法[J]. 光电工程，2024， **51**(10): 240179.

Xie B, Liu Y Q, Li Y L. Colorectal polyp segmentation method combining polarized self-attention and Transformer[J]. *Opto-Electron Eng*, 2024, **51**(10): 240179.

https://doi.org/10.12086/oee.2024.240179

收稿日期: 2024-07-30; 修改日期: 2024-09-18; 录用日期: 2024-09-19

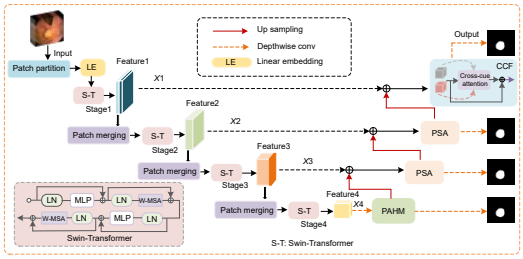
Opto-Electronic Engineering
光 电 工 程

Article

2024 年, 第 51 卷, 第 10 期

DOI: 10.12086/oee.2024.240179 CSTR: 32245.14.oee.2024.240179

结合极化自注意力和 Transformer 的结直肠息肉分割方法



谢 斌[1]，刘阳倩[1*]，李俞霖[2]

[1]江西理工大学信息工程学院，江西 赣州 341000；
[2]江西理工大学电气与自动化学院，江西 赣州 341000

摘要：针对传统结直肠息肉图像分割方法存在的目标分割不够精确、对比度不足，以及边缘细节模糊等问题，文中结合极化自注意力和 Transformer 提出了一种新的结直肠息肉图像分割方法。首先，设计了一种改进的相位感知混合模块，通过动态捕捉 Transformer 结直肠息肉图像的多尺度上下文信息，以使目标分割更加精确。其次，在新方法中引入了极化自注意力机制，实现了图像的自我注意力强化，使得到的图像特征可以直接用于息肉分割任务中，以达到提高病灶区域与正常组织区域对比度的目的。另外，利用线索交叉融合模块加强动态分割时对图像几何结构的捕捉能力，以达到提升结果图像边缘细节的目的。实验结果表明，文中提出的方法不仅能够有效地提升结直肠息肉分割的精确度和对比度，并且还能够较好地克服分割图像细节模糊的问题。在数据集 CVC-ClinicDB、Kvasir 、CVC-ColonDB 和 ETIS-LaribPolypDB 上的测试结果表明，文中所提新方法能够取得更好的分割效果，其 Dice 相似性指数分别为 0.946、0.927、0.805 和 0.781。

关键词：结直肠息肉；Transformer；相位感知模块；极化自注意力模块

中图分类号：R735.34; TP391.41 **文献标志码：**A

谢斌，刘阳倩，李俞霖. 结合极化自注意力和 Transformer 的结直肠息肉分割方法 [J]. 光电工程，2024，**51**(10): 240179
Xie B, Liu Y Q, Li Y L. Colorectal polyp segmentation method combining polarized self-attention and Transformer[J]. *Opto-Electron Eng*, 2024, **51**(10): 240179

Colorectal polyp segmentation method combining polarized self-attention and Transformer

Xie Bing[1], Liu Yangqian[1*], Li Yuling[2]

[1]School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China;
[2]School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China

Abstract: A new colorectal polyp image segmentation method combining polarizing self-attention and Transformer is proposed to solve the problems of traditional colorectal polyp image segmentation such as insufficient target segmentation, insufficient contrast and blurred edge details. Firstly, an improved phase sensing hybrid module is designed to dynamically capture multi-scale context information of colorectal polyp images in Transformer to make

收稿日期：2024-07-30；修回日期：2024-09-18；录用日期：2024-09-19
基金项目：国家自然科学基金资助项目 (61972264)；江西理工大学博士启动基金 (20520010058)
*通信作者：刘阳倩，1974718266@qq.com。

target segmentation more accurate. Secondly, the polarization self-attention mechanism is introduced into the new method to realize the self-attention enhancement of the image, so that the obtained image features can be directly used in the polyp segmentation task to improve the contrast between the lesion area and the normal tissue area. In addition, the cue-cross fusion module is used to enhance the ability to capture the geometric structure of the image in dynamic segmentation, so as to improve the edge details of the resulting image. The experimental results show that the proposed method can not only effectively improve the precision and contrast of colorectal polyp segmentation, but also overcome the problem of blurred detail in the segmentation image. The test results on the data sets CVC-ClinicDB, Kvasir, CVC-ColonDB and ETIS-LaribPolypDB show that the proposed method can achieve better segmentation results, and the Dice similarity index is 0.946, 0.927, 0.805 and 0.781, respectively.

Keywords: colorectal polyp; Transformer; phase sensing module; polarized self-attention module

1 引　言

现阶段，结直肠癌的发病率和死亡率长期保持高位，晚期结直肠癌的死亡率高达 90%，是最常见的恶性肿瘤之一[1]。目前，定期进行结肠镜检查是预防和发现结直肠癌最有效的方法[2]。医生通常借助图像分割来大致确定结直肠息肉的病灶区域，从而给出诊断结果。然而，结直肠道的病理特征十分复杂，不同时期的结直肠息肉大小不一，边界模糊且病灶组织与正常组织相似度高[3]，这给结直肠息肉的图像分割带来诸多挑战。现阶段，基于结直肠息肉的图像分割主要可以分为传统方法和深度学习方法[4]。

文献 [5] 中，Vala 等人提出了一种基于阈值的图像分割方法 (Otsu)，Otsu 能够自适应地确定分割阈值，并根据灰度值的大小将图像分成前景和背景两部分，达到图像分割的目的。Otsu 方法简单易实现，且计算量小，但对噪声和光照比较敏感，通常只能分割细小的目标，因而不适用于结直肠息肉的分割。为了使分割目标更加多样性，文献 [6] 中 Vincent 等人提出了一种基于拓扑学的图像分割方法。该方法将图像看作一个三维地形，用像素点的灰度值表示海拔高度，并以此来将图像划分成不同的区域。虽然取得了较文献 [5] 方法更好的分割表现，但是该方法对于细节丰富的图像容易出现过度分割的问题。为此，文献 [7] 中 Canny 等人提出了一种基于边缘检测的分割方法。借助高斯滤波来平滑目标图像，然后利用高低阈值算法来检测和连接边缘以达到分割图像的目的。但是 Canny 算法的高低阈值通常需要用户主观设定，其在结直肠息肉分割方面的自适应能力较差。另外，文献 [8] 利用“基于标记”的分水岭算法对目标图像进行自适应分割处理，取得了更好的效果。由于息肉图像的病理特征十分复杂且形态各异，上述传统方法[5-8]大都缺乏对目标图像重要特征的自动提取能力，其分割的准确度和泛化能力相对偏低。

近年来，随着深度学习技术的迅速发展，其在获取图像重要特征方面的优点被部分学者用于结直肠息肉图像的分割。Ali 等人[9] 提出了一种基于 ResNet 的深度可分离卷积神经网络 (CNN) 并将其应用于结直肠息肉的分割，一定程度上解决了传统方法存在的准确度低和泛化能力差的问题。由于 CNN 在处理目标图像时池化层容易丢失信息，对图像局部与整体之间的信息关联性处理能力相对较弱，容易造成分割结果出现较明显的误差。为此，Dosovitskiy 等人[10] 在计算机视觉任务中构建了一个新的 Transformer 架构 (vision Transformer, ViT)，直接在非重叠的固定长度块上进行图像分类，并建立了全局信息链，增强了提取病变特征的能力。虽然在图像分类的精度上有了较大提升，但是 ViT 的计算开销相对较大，导致其在执行密集视觉任务时效率较低。为了适应密集视觉任务，Wang 等人[11] 提出了金字塔视觉 Transformer 架构 (pyramid vision Transformer, PVT)，采用渐进式缩小金字塔的策略来以更小的代价处理高分辨率图像。然而，该方法并没有很好地考虑图像的局部联系，导致 PVT 不能很好地获取图像的多尺度特征。为了得到更好的效果，Wu 等人[12] 将反向空间注意力机制引入金字塔 Swin Transformer[13] 编码器中，并设计了多尺度通道注意模块，更有效地提取和聚合多尺度特征信息，提高网络学习和提取息肉各种形态特征的能力。虽然文献 [12] 取得了较传统深度学习方法[9-11] 更好的效果，但是存在目标分割不够精确、对比度不足以及边缘细节模糊等问题。

针对上述问题，文中提出了一种结合极化自注意

力和 Transformer 的结直肠息肉分割方法。首先，利用改进的相位感知混合模块动态捕捉结直肠息肉图像的多尺度上下文信息，动态调制特征图在不同阶段之间振幅和相位的关系，以解决目标分割不够精确的问题。其次，引入极化自注意力机制，采用极化滤波的思想，同时在其正交方向上保持高分辨率，充分考虑像素的回归，实现目标图像的自我注意力强化，把得到的图像特征直接用于息肉分割任务，从而提高病灶区域与正常组织的对比度。最后，通过线索交叉融合模块加强对图像几何结构的捕捉能力，让图像的几何一致性从静态区域到单目深度的动态区域传播，以解决分割时可能出现的边缘细节模糊问题。与现有的结直肠息肉分割方法相比，本文提出的方法在多个公开数据集上的实验结果表明，息肉的分割精度得到提升。

2 网络整体架构

2.1 总体结构

结直肠息肉图像分割时存在目标分割不够精确、对比度不足，以及边缘细节模糊等问题，严重削弱病灶区域特征间的关联性，致使结直肠息肉图像分割时出现边缘细节缺失和病灶区域误分割。为了缓解上述问题，文中提出了一种结合极化自注意力和 Transformer 的结直肠息肉分割方法，其网络结构如图 1 所示。结直肠息肉图像的分割主要包括三个阶段: 1) 通过分段编码器获得四个不同阶段 (stage1-4) 的特征图 (feature 1-4)； 2) 将特征图依次通过解码器的三个模块，分别是相位感知混合模块[14] (phase-aware hybrid module, PAHM) 、极化自注意模块[15] (polarized self-attention, PSA) 和线索交叉融合模块[16] (cross-cue fusion module, CCF)； 3) 得到模型的结直肠息肉分割结果。

在 1) 阶段，将结直肠息肉源图像输入到补丁分割 (patch partition) 模块中进行 4×4 分块，并将所有分块送入分段编码器，获得四个不同阶段的特征图，其大小分别为原图的 1/4、1/8、1/16、1/32。需要说明的是，分段编码器由 4 个完全一致的 Swin-Transformer 模块 (S-T) 组成，Stage1 利用线性嵌入结构 (linear embedding) 将特征维度变为预先设置好的值 C，后面 3 个 Stage 是先通过补丁合并 (patch merging) 将所有 4×4 的块进行合并，再把特征维度扩展为原来的两倍。

在 2) 阶段，设计了一种改进的相位感知混合模块 PAHM，用于动态捕捉各阶段跨层次交互信息，以提升结直肠息肉分割的精确度；并结合极化自注意模块 PSA 来进一步提升 PAHM 模块输出特征图的内分辨率，实现图像的自我注意力强化，从而提高病灶区域与正常组织的对比度。另外，文中利用线索交叉融合模块 CCF 将 Transformer 输出的单帧特征 $X1$ 与 PSA 模块输出的多尺度特征进行融合，以加强对图像几何结构的捕捉能力，从而达到保持结直肠息肉分割

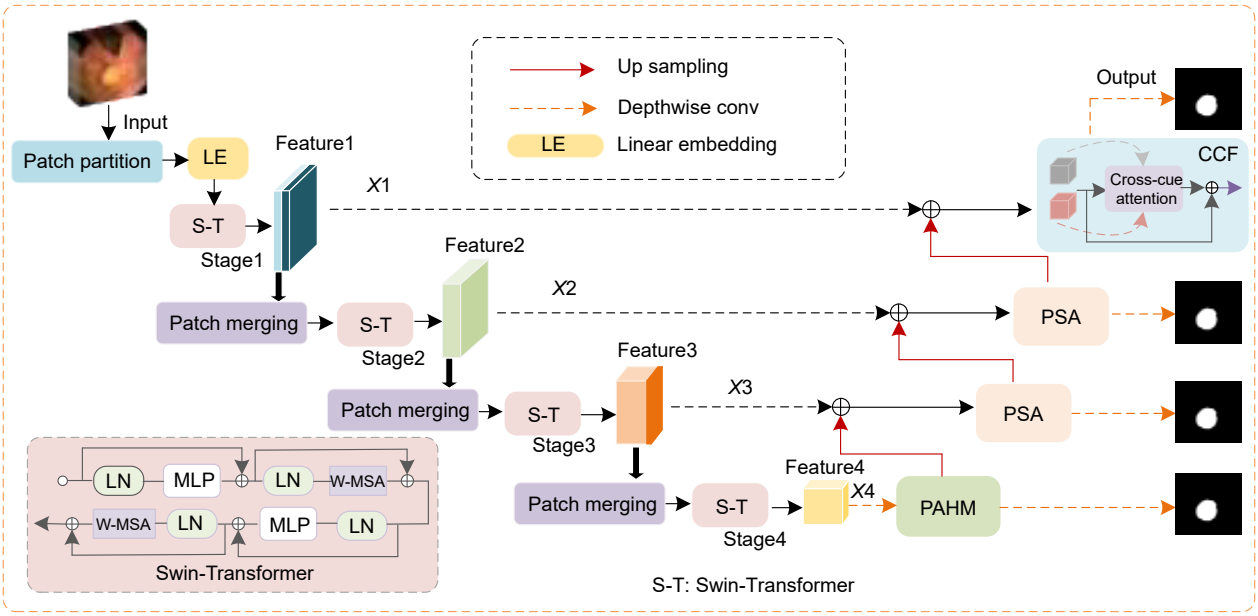


图 1 结合极化自注意力和 Transformer 的结直肠息肉分割网络

Fig. 1 Colorectal polyp segmentation network combining polarized self-attention and Transformer

结果图像细节的目的。

在 3) 阶段，利用深度可分离卷积 (depthwise conc) 分别对解码器三个模块的通道数进行优化调整，最后得到输出的结直肠息肉分割结果图。

2.2 相位感知混合模块

现有的 MLP 模型直接采用固定的权重来聚合图像块，模型在输出特征图时容易忽略来自不同相位层次图像块的语义信息，进而在结直肠息肉分割时容易影响目标图像的分割精度。为了更好地利用不同相位层次图像块的语义信息提高目标图像的分割精度，文中设计了一种改进的相位感知混合模块，具体结构如图 2 所示。

首先，将输入图像分割成多个图像块，然后将每个图像块x_j都视为带有幅值X_j和相位θ_j的波$\tilde{X}_j$，即

$$\begin{cases} \tilde{X}_j = X_j e^{i\theta_j} = X_j \cos\theta_j + iX_j \sin\theta_j \\ X_j = A - FC(x_j, W^c) \\ \theta_j = P - FC(x_j, W^\theta) \end{cases}, \quad (1)$$

式中：$j = 1,2,\cdots,N$为图像块的序号，X_j为图像块x_j通过全连接模块A - FC得到的幅值信息 (W^c为权值参数)，θ_j为图像块x_j通过全连接模块P - FC得到的相位信息 (W^θ为权值参数)。

其次，利用混合机制 Mix 将$\tilde{X}_j$的实部和虚部进行特征聚合得到具有相位差的波信号$\tilde{Z}$，再通过全连接模块T - FC得到更具有表达能力的输出$\tilde{O}_j$，即

$$\tilde{O}_j = T - FC\left(\tilde{Z}_j, W^t\right), j = 1,2,\cdots,N, \quad (2)$$

式中：W^t是图像块混合权值。

最后，为了避免总信息损失，提高结直肠息肉图像特征信息的复用率，文中将初始输入特征$\tilde{X}_j$与$\tilde{O}_j$相融合，得到最终的特征图输出Z_{out}。

表 1 所示为文中模型加入 PAHM 模块前后在 CVC-ClinicDB 和 CVC-ColonDB 数据集上的测试结果。其中，N1 为文中模型未加入 PAHM 模块得到的测试结果，N4 为文中模型加入 PAHM 模块得到的测试结果。图 3 所示为文中模型加入 PAHM 模块前后得到的结直肠息肉分割结果图像。其中，图 3(a) 为结直肠息肉原图像，图 3(b) 为权威专家标注的金标签图像，图 3(c) 为文中模型加入 PAHM 模块得到的分割结果，图 3(d) 为文中模型未加入 PAHM 模块得到的分割结果。

由表 1 和图 3 所示结果可以看出，N4 方法在 CVC-ClinicDB 和 CVC-ColonDB 数据集上的 Dice 指数提升最高，分别为 0.4% 和 0.5%；图 3(d) 存在明显的细节模糊问题，容易导致结直肠息肉分割的精确度降低，图 3(c) 的分割结果明显更为接近金标签，说

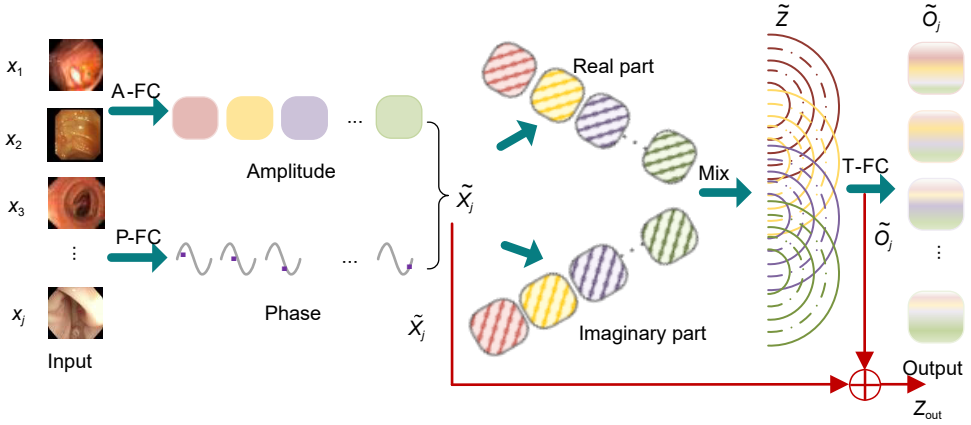


图 2 相位感知混合模块

Fig. 2 Phase-aware hybrid module

表 1 有/无 PAHM 在 CVC-ClinicDB 和 CVC-ColonDB 上的对比

Table 1 Comparison with/without PAHM on CVC-ClinicDB and CVC-ColonDB

Dataset	Method	Dice	MIoU	SE
CVC-ClinicDB	N1	0.942	0.898	0.950
	N4	**0.946**	**0.901**	**0.951**
CVC-ColonDB	N1	0.800	0.727	0.819
	N4	**0.805**	**0.729**	**0.822**

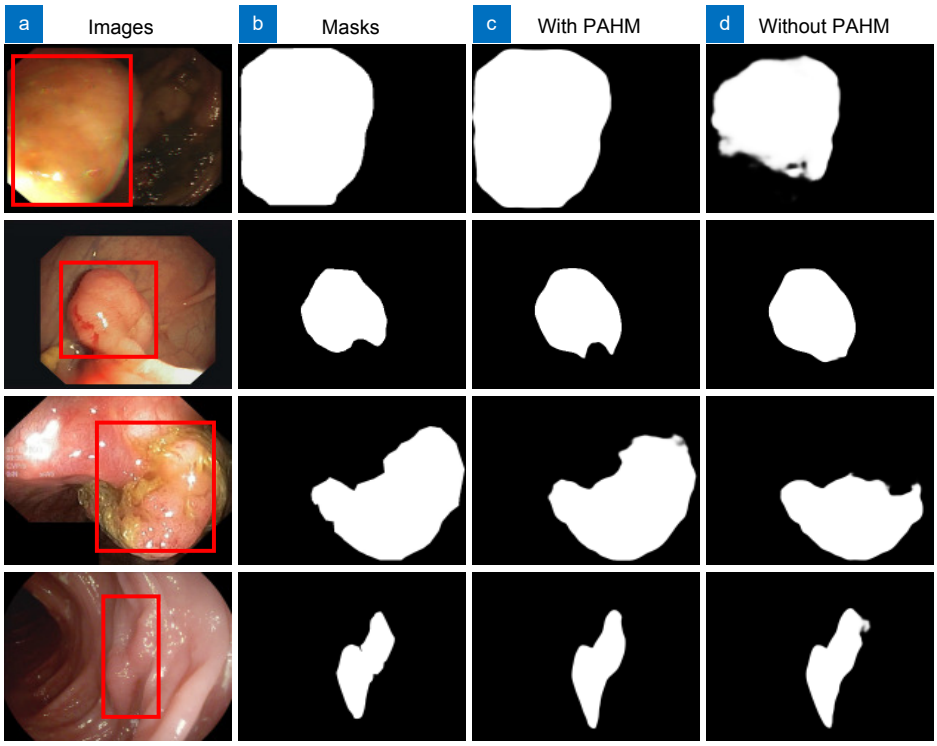


图 3 有/无 PAHM 得到的分割结果

Fig. 3 Segmentation results obtained with/without PAHM

明 PAHM 模块的加入能够达到提升分割精确度的目的，这主要是因为 PAHM 模块可以更充分地利用结直肠息肉不同相位层次特征图的语义信息。

2.3 极化自注意力模块

结直肠息肉分割作为一种细粒度的视觉任务，需要充分考虑每一个像素点的特征，且估计像素语义时依赖高分辨率的输入。传统结直肠息肉分割方法不能充分提取像素点的特征，往往容易导致分割结果图像出现对比度不足的问题。因此文中在 Stage2 和 Stage3 处引入能够充分考虑像素点特征的极化自注意力模块 (PSA) 以解决分割结果对比度不足的问题。PSA 由通道自注意力模块 (channel-only self-attention, CSA) 和空间自注意力模块 (spatial-only self-attention, SSA) 构成，结构如图 4 所示。

1) 通道自注意力模块

通道自注意力 (CSA) 可以帮助网络获取不同通道之间的关联性，进而能够更充分地捕获输入特征的深层语义信息。CSA 首先将输入特征 X (C×H×W) 转换成子特征图 V_{ch} ($C/2$×H×W) 和 Q_{ch} (1×H×W) 矩阵，其中 C 为通道数，H、W 为特征图的尺寸。其次，将 V_{ch} 和 Q_{ch} 的通道数调整得到 V_{ch}' ($C/2$×HW) 和 Q_{ch}' (HW×1×1) 并进行矩阵相乘。最后，将相乘结果输入由 Conv1×1、特征维度标准化 (LayerNorm，LN) 和 Sigmoid 模块组成的卷积映射层，将通道维度重新变为 C，得到通道极化注意力图 $A^{ch}(X)$，即

$$A^{ch}(X) = F_{SG}[W_{z|\theta_1}((\sigma_1(W_v(X)) \times F_{SM}(\sigma_2(W_q(X))))], \quad (3)$$

式中：W_v，W_z 和 W_q 分别为标准的 1×1 卷积，σ_1 和 σ_2 为对特征图进行重塑操作，$F_{SM}(X)$ 为 Softmax 函数运算。另外，为了保留更丰富的特征信息，文中将 $A^{ch}(X)$ 与初始输入特征 X 相乘得到通道自注意力输出 Z^{ch}，即

$$Z^{ch} = A^{ch}(X) \odot^{ch} X . \quad (4)$$

2) 空间自注意力模块

空间自注意力 (SSA) 是在自注意力的基础上增加了空间信息处理能力，可以利用输入特征之间的空间位置关系，更充分地捕获局部细节信息。

SSA 首先将输入特征 X 转换成子特征图 Q_{sp} ($C/2$×1×1) 和 V_{sp} ($C/2$×H×W)，目的是在减少参数量的同时整合空间信息以提高网络的鲁棒性。其次，为了更充分地处理不同区域的空间信息，将 Q_{sp} 和 V_{sp} 的空间维度调整成 V_{sp}' ($C/2$×HW) 和 Q_{sp}' (1×$C/2$)，并进行矩阵乘法。最后，为了更充分地保留特征图的原始信息，增加复用率，将相乘结果输入由 Reshape 模块和 Sigmoid 模块组成的特征映射层以还原特征图

的空间维度，得到空间极化注意力图$A^{sp}(X)$。具体公式为

$$A^{sp}(X)=F_{SG}[\sigma_3(F_{SM}(\sigma_1(F_{GP}(W_q(X))))\times\sigma_2(W_v(X)))], \tag{5}$$

式中：W_v和W_q分别为标准的 1×1 卷积，σ_1、σ_2和σ_3为对特征图进行重塑操作，$F_{SM}(X)$为Softmax函数操作，$F_{GP}(\cdot)$为全局池化操作。

另外，为了保留更丰富的特征信息，文中将$A^{sp}(X)$与输入特征 X 相乘得到极化注意力的输出Z^{sp}。

$$Z^{sp}=A^{sp}(X)\odot^{sp}X. \tag{6}$$

由图 4 可知，输入特征信号 X 经过 CSA 和 SSA 的处理，PSA 的最终输出为

$$PSA(X)=Z^{sp}(Z^{ch})=A^{sp}(A^{ch}(X)\odot^{ch}X)\odot^{sp}A^{ch}(X)\odot^{ch}X, \tag{7}$$

表 2 所示为文中模型加入 PSA 模块前后在 CVC-ClinicDB 和 CVC-ColonDB 数据集上的测试结果。其中，N2 为文中模型未加入 PSA 模块得到的测试结果，N4 为文中模型加入 PSA 模块得到的测试结果。图 5 所示为文中模型加入 PSA 模块前后的结直肠息肉分割结果图像。其中，图 5(a) 为结直肠息肉原图像；图 5(b) 为权威专家标注的金标签图像；图 5(c) 为加入 PSA 模块得到的分割结果；图 5(d) 为未加入 PSA 模块得到的分割结果。

由表 2 和图 5 所示结果可以看出：N4 方法在 CVC-ClinicDB 和 CVC-ColonDB 数据集上的 MIoU 指数提升幅值最高，分别为 1.0% 和 1.8%；未加入 PSA 模块得到的分割图结果图 5(d) 存在对比度明显不足的问题，而图 5(c) 所示结果非常接近金标签，说明加入 PSA 模块可以较好地提升分割结果的对比度，这主要是由于 PSA 模块采用了极化滤波的思想，能够更好地提取结直肠息肉像素点的特征，因此在进行结直肠息肉分割时可以有效地提升分割结果图像的对比度。

2.4 线索交叉融合模块

传统的多帧深度估计方法依赖单帧多视角的几何一致性获得高精度结果。然而，在应用于动态环境如结直肠息肉检测时，由于肠道经常发生蠕动，因此结直肠息肉图像的几何一致性容易被动态影响从而导致

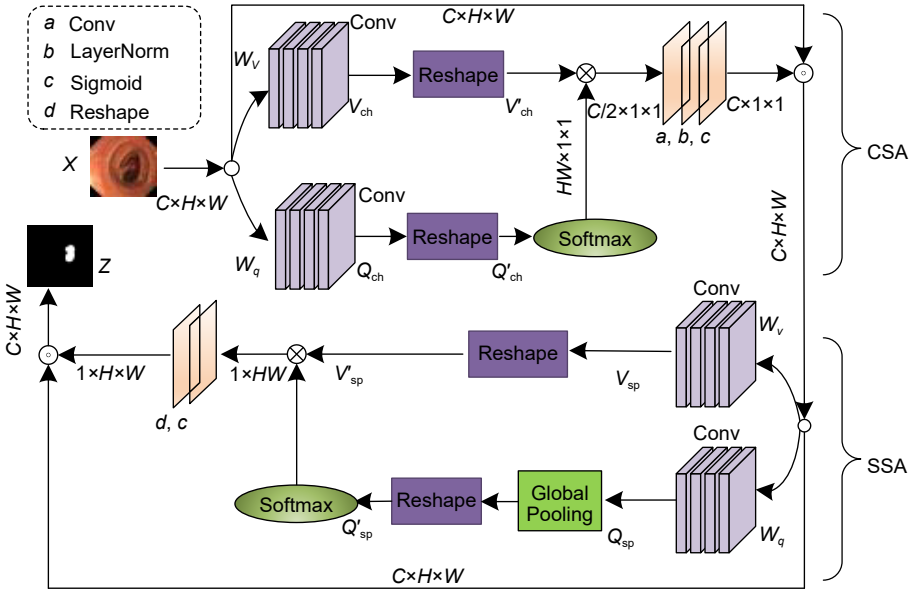


图 4 极化自注意模块

Fig. 4 Polarized self-attention module

表 2 有/无 PSA 在 CVC-ClinicDB 和 CVC-ColonDB 上的对比

Table 2 Comparison with/without PSA on CVC-ClinicDB and CVC-ColonDB

Dataset	Method	Dice	MIoU	SE
CVC-ClinicDB	N2	0.937	0.881	0.946
	N4	**0.946**	**0.901**	**0.951**
CVC-ColonDB	N2	0.788	0.711	0.813
	N4	**0.805**	**0.729**	**0.822**

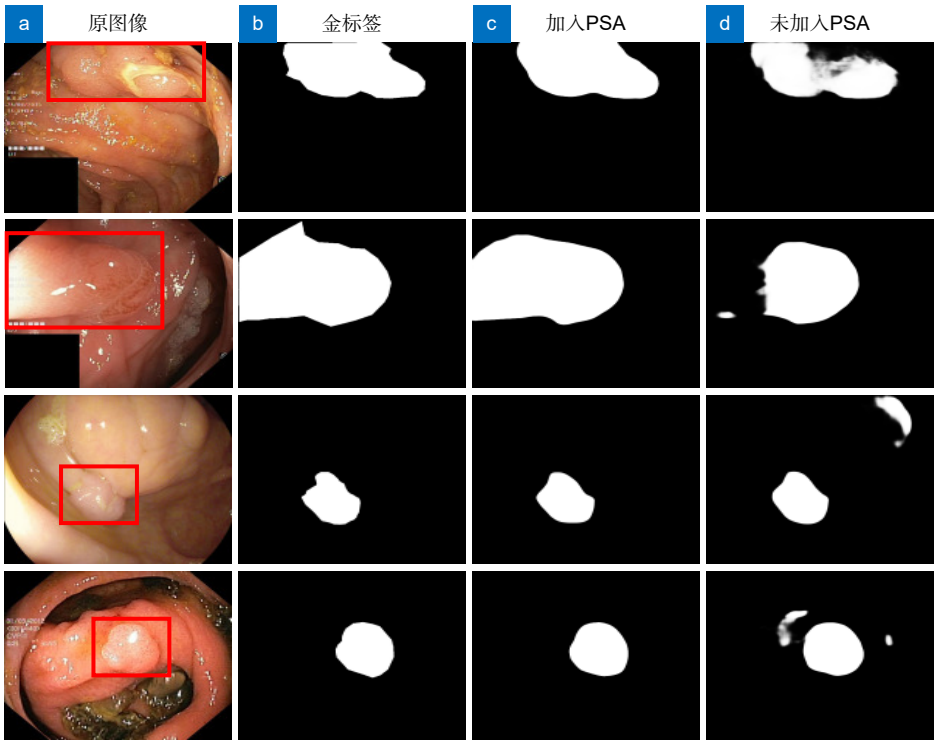


图 5 有/无 PSA 得到的分割结果

Fig. 5 Segmentation results with or without PSA

分割结果出现边缘细节模糊的问题。为了让单/多帧两个线索的互补特性相互提升，文中引入线索交叉融合模块 (CCF) 以更好地解决分割结果边缘细节模糊的问题。CCF 的具体架构图如图 6 所示。

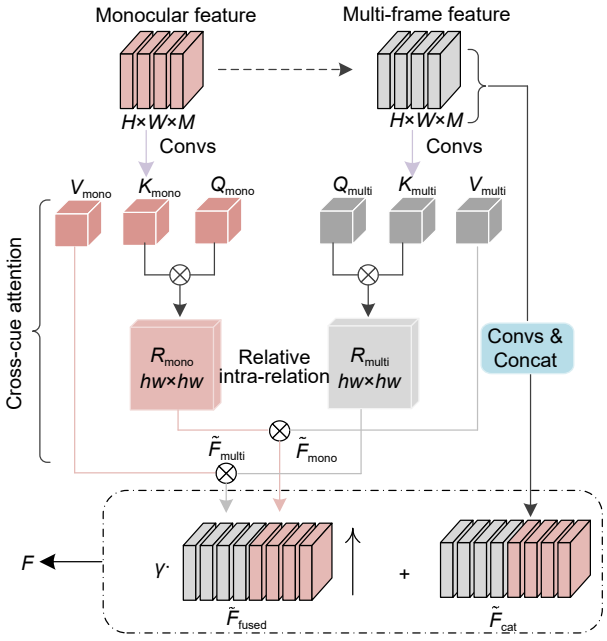


图 6 线索交叉融合模块

Fig. 6 Cross-cue fusion module

首先，给定输入的单帧特征C_{mono}和多帧特征C_{muliti}，通过下采样得到其对应的深度特征F_{mono}、F_{muliti}，随后将F_{mono}和F_{muliti}输入给线索交叉注意模块 (cross-cue fusion module, CCA) 将二者互相增强。CCA 是通过提取其它交叉线索的相对内部关系来增强每个深度线索的几何信息，并以无显式分割公式获取增强后的特征$\tilde{F}_{\text{mono}}$、$\tilde{F}_{\text{multi}}$。

$$\begin{cases}\tilde{F}_{\text{mono}} = CCA_{\text{mono}}(F_{\text{multi}}, F_{\text{mono}}) \\ \tilde{F}_{\text{multi}} = CCA_{\text{multi}}(F_{\text{mono}}, F_{\text{multi}})\end{cases}. \tag{8}$$

其次，将增强后的特征$\tilde{F}_{\text{mono}}$、$\tilde{F}_{\text{multi}}$连接以产生融合特征$\tilde{F}_{\text{fused}}$。为了保留初始深度线索的细节信息，通过*Convs&Concat*处理输入的单/多帧深度线索，并添加残差连接。最终的交叉线索计算式为

$$F_{\text{cat}} = Cat(Conv(C_{\text{multi}})), Conv(C_{\text{mono}}), \tag{9}$$

$$F = \gamma \tilde{F}_{\text{fused}} \uparrow + F_{\text{cat}}, \tag{10}$$

式中：γ是加权因子，“↑”表示上采样操作。

表 3 所示为文中模型加入 CCF 模块前后在 CVC-ClinicDB 和 CVC-ColonDB 数据集上的测试结果。其中，N3 为文中模型未加入 CCF 模块得到的测试结果，N4 为文中模型加入 CCF 模块得到的测试结果。图 7

表 3 有/无 CCF 在 CVC-ClinicDB 和 CVC-ColonDB 上的对比

Table 3 Comparison with/without CCF on CVC-ClinicDB and CVC-ColonDB

Dataset	Method	Dice	MIoU	SE
CVC-ClinicDB	N3	0.942	0.894	0.949
	N4	**0.946**	**0.901**	**0.951**
CVC-ColonDB	N3	0.751	0.684	0.777
	N4	**0.805**	**0.729**	**0.822**

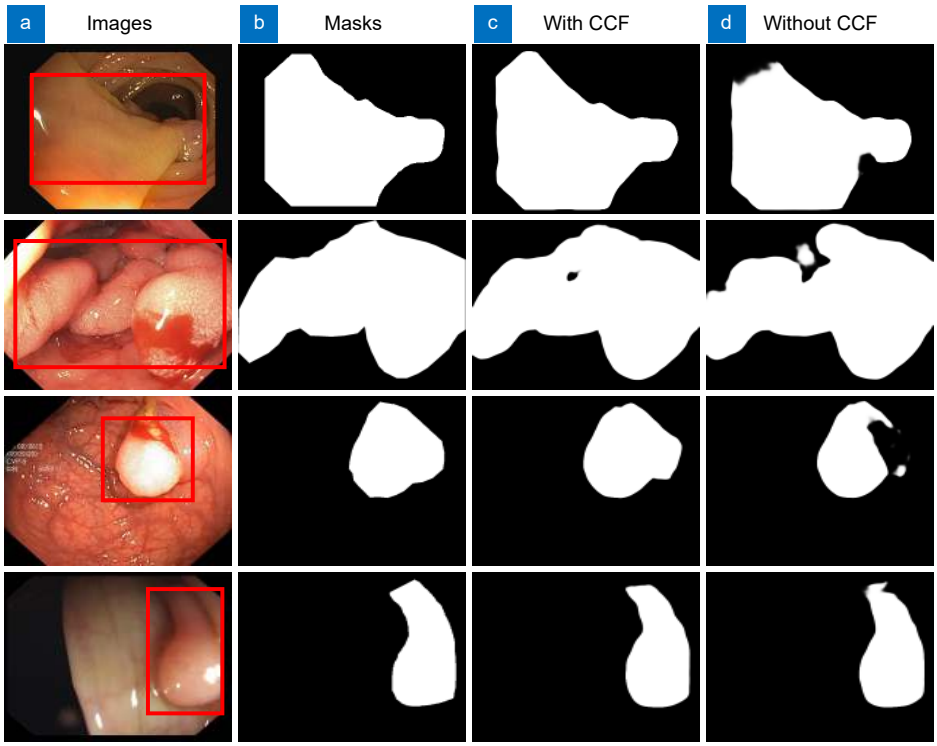


图 7 有/无 CCF 得到的分割结果

Fig. 7 Segmentation results obtained with or without CCF

所示为文中模型加入 CCF 模块前后的结直肠息肉分割结果图像。其中，图 7(a) 为结直肠息肉原图像；图 7(b) 为权威专家标注的金标签图像；图 7(c) 为文中模型加入 CCF 模块得到的分割结果；图 7(d) 为文中模型未加入 CCF 模块得到的分割结果。

由表 3 和图 7 所示结果可以看出：N4 方法在 CVC-ClinicDB 和 CVC-ColonDB 数据集上的 Dice 指数和 MioU 指数均有大幅提升；图 7(d) 的分割结果其边缘细节比较模糊，这是主要是由于息肉图像的单帧和多帧线索不能较好融合所导致的。相较而言，图 7(c) 的分割结果更加接近金标签，这说明引入了 CCF 模块后文中方法能够更好地捕捉结直肠息肉的边缘细节。

3 实验结果与分析

3.1 实验设置与环境

文中是基于开源的 Pytorch 框架下实现的，所有实验均在 Windows 11 操作系统进行，实验所使用的 CPU 是 Inter Core i5-13600，内存大小为 16 GB，GPU 为 NVIDIA GeForce RTX 4070Ti，显存大小为 12 GB。模型采用业内公认的交叉熵合并比损失函数，初始学习率为 le-5，学习速率衰减率为 0.1，衰减周期设置为 50，同时模型采用自调整矩阵估计优化器，动量大小设置为 0.9，批次处理数据给定为 6，模型全部实验训练 100 个 epoch，并且使用 {0.75，1，1.25}的多尺度训练策略。

3.2 损失函数

大多学者在研究图像分割领域时会采用加权二进制交叉熵损失函数来提高分割模型的准确性，但在结直肠息肉分割这类特定任务中，由于分割时目标区域通常较为细小，损失函数难以收敛或性能不稳定，导致训练效果不佳。而加权 IoU 损失函数通过计算预测边界框与真实边界框的相交并集来衡量模型对真实样本的接近程度，从而减轻了小目标区域带来的误差。因此，文中结合两者优点，使用加权二进制交叉熵函数和加权 IoU 损失函数组成新的损失函数来测评本文网络。具体表达式为

$$L_{\mathrm{BCE}}=-\frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(1+\lambda\beta_{ij})\sum_{i=1}^{q}\varphi(g_{ij}=l)\log P(p_{ij}=l|\alpha)}{\sum_{i=1}^{H}\sum_{j=1}^{W}\lambda\beta_{ij}}, \tag{11}$$

$$L_{\mathrm{IoU}}=1-\frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(g_{ij}\times p_{ij})(1+\lambda\beta_{ij})}{\sum_{i=1}^{H}\sum_{j=1}^{W}(g_{ij}+p_{ij}-g_{ij}\times p_{ij})\times(1+\lambda\beta_{ij})}, \tag{12}$$

$$L=L_{\mathrm{BCE}}+L_{\mathrm{IoU}}, \tag{13}$$

式中：λ为超参数；β_{ij}为权重值，范围为$(0,1)$，该值越大像素和周围区域像素值差距越大；$l\in(0,1)$用于区分病灶区域与非病灶区域；$\varphi(\cdot)$为标记像素类别的指数函数；$P(p_{ij}=l|\alpha)$为预测结果的概率值。

3.3 数据集

为了评估文中网络模型的性能，采用 CVC-ClinicDB[17]，Kvasir[18]，CVC-ColonDB[19] 和 ETIS-LaribPolypDB[20] (简称 ETIS) 四个公开数据集 (如表 4 所示) 对该网络进行测试，由此来验证算法的有效性。其中 CVC-ClinicDB 数据集在 2015 年医学图像计算机与计算机辅助干预国际会议上发布，Kvasir 数据集是由挪威奥斯陆医学院内窥镜专家团采集并标注，ETIS 数据集是来自于 2017 年 MIC-CAI 结直肠息肉挑战赛，CVC-ColonDB 数据集是由美国梅奥诊所发布，图片是从直肠镜检查中随机抽取的短视频中生成。实验是从 Kvasir 和 CVC-ClinicDB 这两个数据集中随机挑选，其中 90% 的图片构成训练集，剩下 10% 的图片与 CVC-ColonDB 和 ETIS 数据集一起构成测试集，用以评估该网络模型的预测、学习和泛化能力。由于四个数据集选取的图片分辨率尺寸各有不同，为了使实验的训练和测试进行的更加顺利方便，文中统一将分辨率调整为 352×352。

表 4 实验参数设置

Table 4 Experimental parameter settings

Dataset	Traindata	Testdata	Picture size/pixel
CVC-ClinicDB	550	62	352×352
Kvasir	900	100	352×352
ETIS-LaribPolypDB	0	196	352×352
CVC-ColonDB	0	380	352×352

3.4 评价指标

文中在测试时所使用的数据集、学习率、和优化策略均相同，同时采用 Dice 相似性系数、平均交并比 (mean intersection over union，MIoU) 、精确度 (precision，PC) 、召回率 (recall ratio，RC) 、F2 得分和平均绝对误差 (mean absolute error，MAE) 来对结肠息肉的分割性能和结果进行评估。相应的计算公式分别为

$$Dice=\frac{2|M\cap N|}{|M|+|N|}, \tag{14}$$

$$MIoU=\frac{|M\cap N|}{|M|+|N|-|M\cap N|}, \tag{15}$$

$$PC=\frac{TP}{TP+FP}, \tag{16}$$

$$RC=\frac{TP}{TP+FN}, \tag{17}$$

$$F2=\frac{5\times RC\times PC}{4\times PC+RC}, \tag{18}$$

$$MAE=\frac{1}{Z}\sum|N-M|, \tag{19}$$

其中：M 为预测的输出图像，N 为权威专家标注的金标签图像，TP为预测结果中正确分类的前景像素数量，FP为预测结果中被错误识别成背景像素的数量，FN为预测结果中被错误识别进而分类为前景像素的数量，Z 为图像中的像素点个数。

由式 (14) - (19) 可知，Dice 指标越高表明分割结果与标准预测结果的一致性越高，即分割效果越好；MIoU 表示预测图像和金标签的交集比例平均求和，MIoU 指标越高说明分割结果越贴合金标签。PC 为精确度，表示在预测为正的样本中有多少是准确的。RC 为召回率，表示在所有实际为正类的样本中被预测为正类的比例。F2 得分是召回率和精确度的综合体现，F2 指标得分越高，说明能更精确地将病变对

象从背景区域中分割出来。MAE 为平均绝对值误差，表示预测值与观测值之间绝对误差的平均值。

3.5 网络性能对比实验

为了进一步验证所提 TPSA-Net 分割结直肠息肉图像的性能，文中将其分别与传统基于 CNN 的经典医学图像分割网络 U-Net[21]、PraNet[22]、EU-Net[23]、DCRNet[24] 和近年来基于 Transformer 的医学图像分割网络 SSFormer-S[25]、MSRAFormer 等进行了比较。

表 5 给出了 TPSA-Net 与其它六种网络[12,21-25] 在 CVC-ClinicDB 和 Kvasir-SEG 两个数据集的测试结果。由表 5 可以看出，U-Net、EU-Net、PraNet 和 DCRNet 的效果相对不够理想，这主要是由于这 4 种网络均是在 CNN 基础框架上构建的，主要侧重于局部特征的提取，而对全局特征的提取能力则较为薄弱。虽然基于 Transformer 框架的 MSRAFormer、SSFormer-S 取得了较 CNN 基础框架方法[20-23] 更好的效果，但是由于没有充分考虑不同特征层语义信息之间的联系，因此其分割效果还有待提高。相较而言，文中所提 TPSA-Net 在 Dice 和 MIoU 两个评价指标方面均取得了最优。其中，基于 CVC-ClinicDB 数据集的 Dice 和 MIoU 分别为 0.946 和 0.901，与 SSFormer-S 相比分别提升 2.7% 和 2.9%；基于 Kvasir-SEG 数据集的 Dice 和 MIoU 分别为 0.927 和 0.880，相较于经典传统医学图像分割网络 U-Net 分别提升 12.4% 和 14.5%。这主要是由于文中所提 TPSA-Net 能够利用相位感知混合模块更精准地捕捉跨层次交互信息，因此取得更好的分割效果。

另外，图 8 给出了表 5 所示实验的可视化结果图，从上到下依次为原图像 (image) 、金标签 (masks) 、U-Net、PraNet、EU-Net、DCRNet、SSFormer-S、MSRAFormer 和 TPSA-Net 的分割结果。由图 8 可以看出，基于 CNN 框架的 U-Net、EU-Net 和 DCR-Net 基础网络分割效果较差，在第 3 行和第 6 行均出现了大量的伪影；MSRAFormer 和 SSFormer-S 虽然取得了较 CNN 基础框架方法更好的分割效果，但是相比之下，文中所提网络 TPSA-Net 无论是在边缘细节方面还是分割精确度方面明显都更胜一筹。

为了进一步说明所提 TPSA-Net 的性能，文中在 CVC-ColonDB 和 ETIS-LaribPolypDB 两个底层噪声较多数据集上也进行了测试，结果如表 6 和图 9 所示。由表 6 所示结果可以看出，文中所提 TPSA-Net 所有评价指标均取得最优。其中，基于 CVC-ColonDB 数据集的 Dice、MIoU 和 F2 得分分别为 0.805、0.729 和 0.806，与 MSRAFormer 网络相比分别提升 4%、3.4% 和 3.4%；基于 ETIS-LaribPolypDB 数据集的 Dice、MIoU 和 F2 得分为 0.781、0.706 和 0.807，相较于先进网络算法 SSFormer-S 分别提升 1.1%、1.1% 和 6.4%。这主要是由于 TPSA-Net 利用极化自注意力模块较好地获取了结直肠息肉图像中像素级语义信息间的关联性，从而有效地提升了分割结果的对比度。

表 5 不同算法在 CVC-ClinicDB 和 Kvasir 上的对比

Table 5 Comparison of different algorithms on CVC-ClinicDB and Kvasir

Dataset	Method	Dice	MIoU	SE	PC	F2	MAE
CVC-ClinicDB	U-Net	0.822	0.756	0.836	0.835	0.828	0.020
	PraNet	0.902	0.850	0.911	0.905	0.901	0.009
	EU-Net	0.905	0.849	0.956	0.881	0.927	0.011
	DCRNet	0.899	0.847	0.912	0.893	0.907	0.010
	SSFormer-S	0.919	0.872	0.903	0.939	0.908	0.007
	MSRAFormer	0.934	0.884	0.950	0.924	0.944	0.007
	Ours	**0.946**	**0.901**	**0.957**	**0.943**	**0.949**	**0.005**
Kvasir	U-Net	0.821	0.747	0.855	0.856	0.828	0.055
	PraNet	0.901	0.841	0.910	0.916	0.903	0.030
	EU-Net	0.911	0.858	0.931	0.912	0.919	0.028
	DCRNet	0.889	0.823	0.903	0.902	0.892	0.034
	SSFormer-S	0.925	0.876	0.917	0.944	0.921	0.020
	MSRAFormer	0.919	0.870	0.921	0.938	0.918	0.020
	Ours	**0.927**	**0.880**	**0.932**	**0.950**	**0.923**	**0.020**

表 6 所示结果说明文中所提网络 TPSA-Net 相比其它网络的分割性能更优。

另外，图 9 给出了表 6 所示实验的可视化结果图，从上到下依次为原图像 (image) 、金标签 (masks) 、U-Net、 PraNet、 EU-Net、 DCRNet、 SSFormer-S、MSRAFormer 和 TPSA-Net 分割结果。由图 9 可以看出，U-Net、EU-Net、DCR-Net 和 PraNet 网络出现了伪影明显和分割结果全黑的问题；MSRAFormer 和

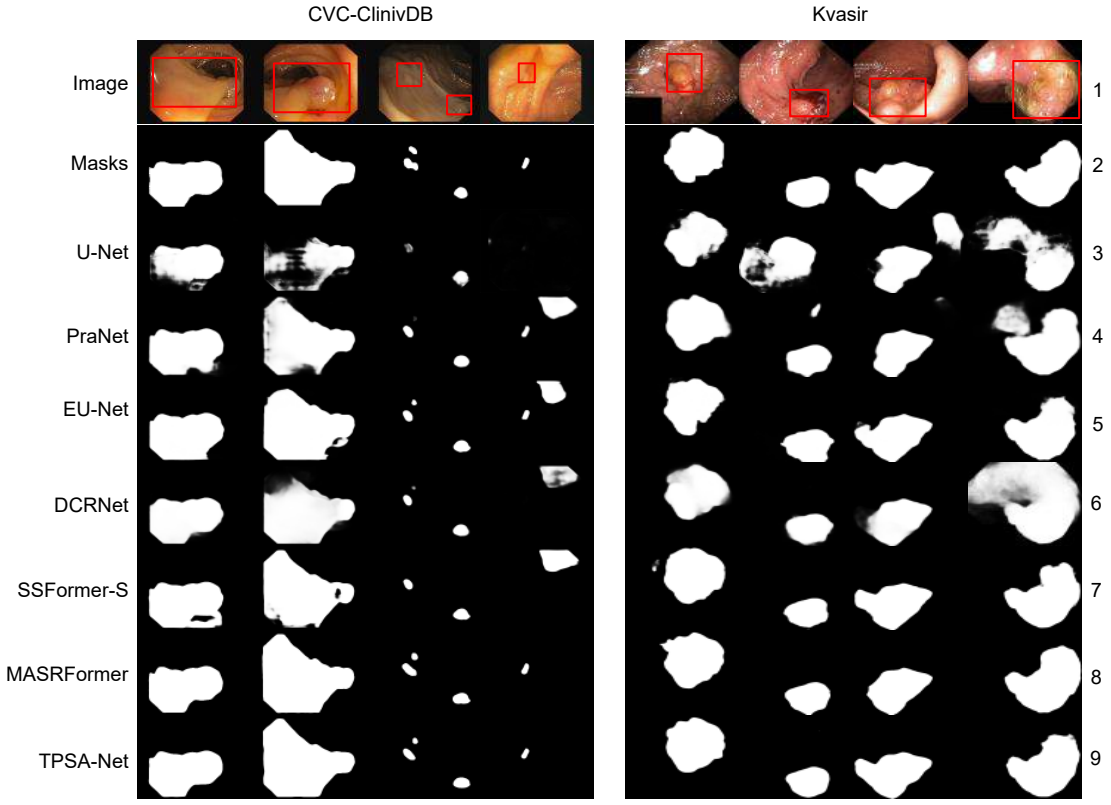


图 8 不同网络模型在 CVC-ClinicDB 和 Kvasir 数据集上的分割结果

Fig. 8 Visualization of segmentation results of different network models on CVC-ClinicDB and Kvasir datasets

表 6 不同算法在 CVC-ColonDB 和 ETIS-LaribPolypDB 上的对比

Table 6 Comparison of different algorithms on CVC-ColonDB and ETIS-LaribPolypDB

Dataset	Method	Dice	MIoU	SE	PC	F2	MAE
CVC-ColonDB	U-Net	0.512	0.438	0.524	0.621	0.510	0.059
	PraNet	0.717	0.641	0.740	0.755	0.716	0.044
	EU-Net	0.756	0.683	0.848	0.756	0.789	0.043
	DCRNet	0.707	0.632	0.777	0.719	0.723	0.051
	SSFormer-S	0.775	0.698	0.776	0.836	0.767	0.034
	MSRAFormer	0.765	0.695	0.801	0.870	0.772	0.031
	Ours	**0.805**	**0.729**	**0.878**	**0.872**	**0.806**	**0.025**
ETIS-LaribPolypDB	U-Net	0.406	0.334	0.482	0.439	0.428	0.037
	PraNet	0.631	0.567	0.689	0.628	0.649	0.030
	EU-Net	0.690	0.611	0.871	0.637	0.749	0.065
	DCRNet	0.548	0.484	0.744	0.504	0.600	0.095
	SSFormer-S	0.770	0.695	0.856	0.744	0.782	0.017
	MSRAFormer	0.749	0.674	0.821	0.787	0.782	0.012
	Ours	**0.781**	**0.706**	**0.874**	**0.808**	**0.807**	**0.011**

SSFormer-S 仍存在分割不精确和边缘细节模糊的情况。相比之下，文中提出的 TPSA-Net 网络利用线索交叉融合模块从动态和静态两个层面融合多尺度特征，有效减少了分割结果边缘不清晰的问题。值得说明的是，相比于 MSRAFormer，本文方法的参数量降低了 21%，且单论迭代时长 (179 round/s) 优于 MSRAFormer 网络 (199 round/s)，这说明设计的相位感知混合模块在不增加参数量和计算复杂度的条件下提升了网络的分割性能。

表 5、图 8、表 6 和图 9 的实验结果说明，文中所提 TPSA-Net 不论是在数据的分割精度还是在图像分割的可视化效果方面都更胜一筹。

3.6 消融实验

为了更好地说明所提模块 CCF、PAHM 和 PSA 对整体模型分割性能的影响，文中在 Kvasir 和 ETIS 数据集上进行了消融实验，结果如表 7 所示。需要说明的是，M4 方法为文中所提网络，加粗为最优值。

由表 7 可知，M4 方法在 Kvasir 数据集上的 Dice 指数和 MIou 指数比 M1 方法分别提升 0.8% 和 0.7%，在 ETIS 数据集上也同样提升，这说明 CCF 能够将单帧线索与多帧线索整合更精准，提高网络分割效果。M2 方法由于没有 PAHM 模块发挥作用，M4 方法在 Kvasir 数据集和 ETIS 数据集的 Dice 指数提升最高，分别为 0.9% 和 0.4%，由此可见，PAHM 可以

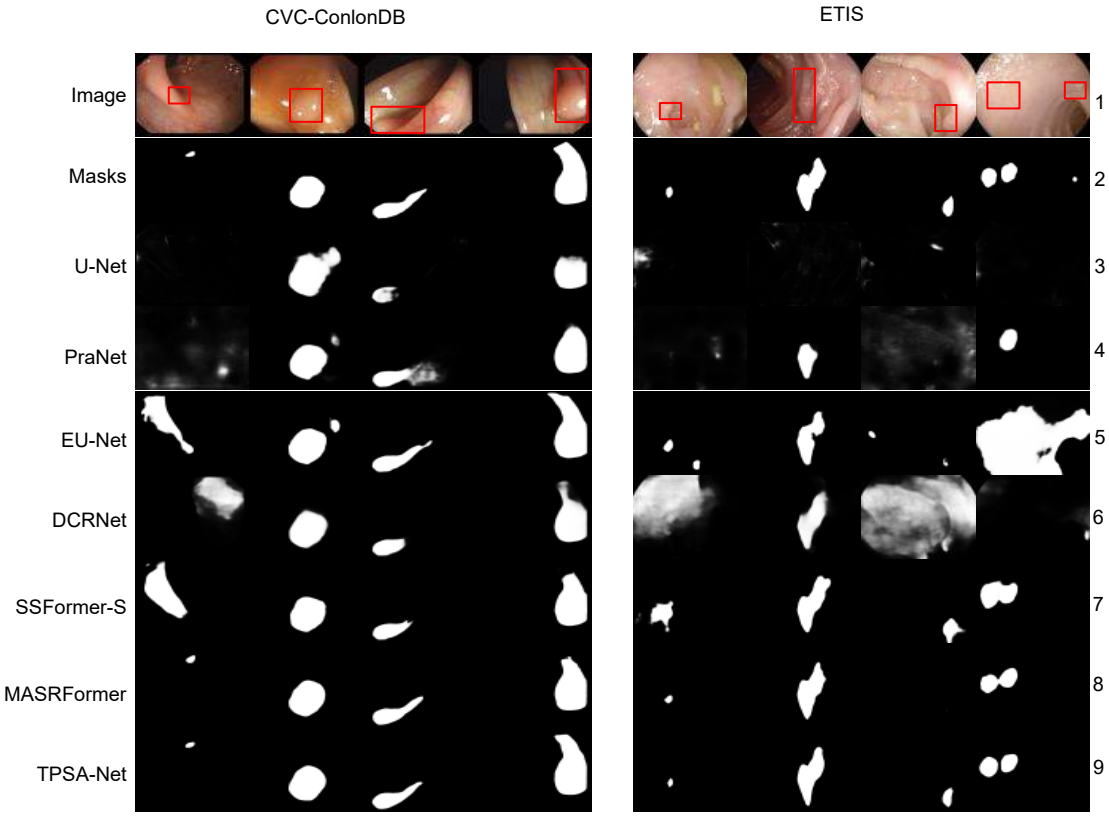


图 9 不同网络模型在 CVC-ColonDB 和 ETIS 上的分割结果

Fig. 9 Visualization of segmentation results of different network models on CVC-ColonDB and ETIS

表 7 各模块在 Kvasir 和 EITS 数据集上的消融研究

Table 7 Ablation of each module on Kvasir and EITS datasets

Method	CCF	PAHM	PSA	Kvasir				ETIS			
				Dice	MIoU	SE	F2	Dice	MIoU	SE	F2
M1	×	√	√	0.919	0.873	0.919	0.920	0.744	0.674	0.803	0.769
M2	√	×	√	0.918	0.872	0.913	0.914	0.740	0.672	0.802	0.767
M3	√	√	×	0.924	0.876	0.924	0.918	0.756	0.681	0.836	0.792
M4	√	√	√	**0.927**	**0.880**	**0.926**	**0.923**	**0.781**	**0.706**	**0.874**	**0.807**

有效聚合多尺度上下文信息，提升网络 Dice 值。M4 在 Kvasir 数据集和 ETIS 数据集的 MIoU 指数分别比 M3 高 0.4% 和 2.5%，这说明极化自注意力模块可以加深特征通道之间依赖性，同时减少空间信息缺失，提高网络的精确度。M4 方法将所用模块应用于结直肠息肉图像分割，指标达到最优。表 7 所示结果说明了网络各模块的实际作用。

4 结　论

针对传统方法在结直肠息肉分割时存在的目标分割不够精确、对比度不足，以及边缘细节模糊等问题，文中提出了一种结合极化自注意力和 Transformer 的结直肠息肉分割网络 TPSA-Net。一方面，改进了相位感知混合模块，能够动态捕捉结直肠息肉图像不同层次的多尺度上下文信息，以提高目标分割的精确度。另一方面，引入极化自注意力模块，强化了图像的自我注意力，以提高病灶区域与正常组织区域的对比度。最后，借助线索交叉融合模块加强对图像几何结构的动态捕捉能力，以解决结直肠息肉分割时边缘细节模糊的问题。在四个数据集上的 Dice 相似性指数分别为 0.946、0.927、0.805 和 0.781。大量实验结果表明，相比于传统方法而言，文中所提结直肠息肉分割网络能够得到质量更好的分割结果。如何利用深度学习技术研究更加简单、高效的结直肠息肉分割方法是今后的重点。

参考文献

[1] Liang H, Cheng Z M, Zhong H Q, et al. A region-based convolutional network for nuclei detection and segmentation in microscopy images[J]. *Biomed Signal Process Control*, 2022, **71**: 103276.

[2] Jha D, Smedsrud P H, Johansen D, et al. A comprehensive study on colorectal polyp segmentation with ResUNet++, conditional random field and test-time augmentation[J]. *IEEE J Biomed Health Inform*, 2021, **25**(6): 2029−2040.

[3] Li W S, Zhao Y H, Li F Y, et al. MIA-Net: multi-information aggregation network combining transformers and convolutional feature learning for polyp segmentation[J]. *Knowl-Based Syst*, 2022, **247**: 108824.

[4] Ding J H, Yuan M H. A multi-target semantic segmentation method for millimetre wave SAR images based on a dual-branch multi-scale fusion network[J]. *Opto-Electron Eng*, 2023, **50**(12): 230242.
丁俊华, 袁明辉. 基于双分支多尺度融合网络的毫米波 SAR 图像多目标语义分割方法[J]. 光电工程, 2023, **50**(12): 230242.

[5] Vala M H J, Baxi A. A review on Otsu image segmentation algorithm[J]. *Int J Adv Res Comput Eng Technol*, 2013, **2**(2): 387−389.

[6] Vincent L, Soille P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations[J]. *IEEE Trans Pattern Anal Mach Intell*, 1991, **13**(6): 583−598.

[7] Canny J. A computational approach to edge detection[J]. *IEEE Trans Pattern Anal Mach Intell*, 1986, **PAMI-8**(6): 679−698.

[8] Liang Y B, Fu J. Watershed algorithm for medical image segmentation based on morphology and total variation model[J]. *Int J Patt Recogn Artif Intell*, 2019, **33**(5): 1954019.

[9] Ali S M F, Khan M T, Haider S U, et al. Depth-wise separable atrous convolution for polyps segmentation in gastro-intestinal tract[C]//*Proceedings of the Working Notes Proceedings of the MediaEval 2020 Workshop*, 2021.

[10] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale[C]//*Proceedings of the 9th International Conference on Learning Representations*, 2021.

[11] Wang W H, Xie E Z, Li X, et al. Pyramid vision transformer: a versatile backbone for dense prediction without convolutions[C]//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*, 2021: 548–558. https://doi.org/10.1109/ICCV48922.2021.00061.

[12] Wu C, Long C, Li S J, et al. MSRAformer: multiscale spatial reverse attention network for polyp segmentation[J]. *Comput Biol Med*, 2022, **151**: 106274.

[13] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*, 2021: 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986.

[14] Tang Y H, Han K, Guo J Y, et al. An image patch is a wave: phase-aware vision MLP[C]//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 10925–10934. https://doi.org/10.1109/CVPR52688.2022.01066.

[15] Liu H J, Liu F Q, Fan X Y, et al. Polarized self-attention: towards high-quality pixel-wise regression[Z]. arXiv: 2107.00782, 2021. https://arxiv.org/abs/2107.00782.

[16] Li R, Gong D, Yin W, et al. Learning to fuse monocular and multi-view cues for multi-frame depth estimation in dynamic scenes[C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 21539–21548. https://doi.org/10.1109/CVPR52729.2023.02063.

[17] 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]. *Comput Med Imaging Graph*, 2015, **43**: 99−111.

[18] Amsaleg L, Huet B, Larson M, et al. Proceedings of the 27th ACM international conference on multimedia[C]. New York: ACM Press, 2019.

[19] 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.

[20] Tajbakhsh N, Gurudu S R, Liang J M. Automated polyp detection in colonoscopy videos using shape and context information[J]. *IEEE Trans Med Imaging*, 2016, **35**(2): 630−644.

[21] 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*, 2015: 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

[22] Fan D P, Ji G P, Zhou T, et al. PraNet: parallel reverse attention network for polyp segmentation[C]//*Proceedings of the 23rd International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020: 263–273. https://doi.org/10.1007/978-3-030-59725-2_26.

[23] Patel K, Bur A M, Wang G H. Enhanced U-Net: a feature enhancement network for polyp segmentation[C]//*Proceedings of 2021 18th Conference on Robots and Vision*, 2021: 181–188. https://doi.org/10.1109/CRV52889.2021.00032.

[24] Yin Z J, Liang K M, Ma Z Y, et al. Duplex contextual relation network for polyp segmentation[C]//*Proceedings of 2022 IEEE 19th International Symposium on Biomedical Imaging*, 2022: 1–5. https://doi.org/10.1109/ISBI52829.2022.9761402.

[25] Wang J F, Huang Q M, Tang F L, et al. Stepwise feature fusion: local guides global[C]//*Proceedings of the 25th International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2022: 110–120. https://doi.org/10.1007/978-3-031-16437-8_11.

作者简介

谢斌 (1977-)，男，博士，研究生导师，主要研究方向为深度学习、图像处理。

E-mail：xiebin-66@163.com

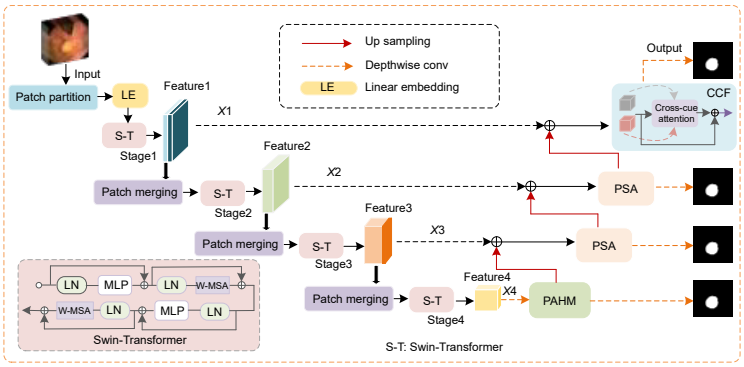
【通信作者】刘阳倩 (1999-)，女，硕士研究生，主要研究方向为深度学习、图像处理。

E-mail: 1974718266@qq.com

扫描二维码，获取PDF全文

Colorectal polyp segmentation method combining polarized self-attention and Transformer

Xie Bing[1], Liu Yangqian[1*], Li Yuling[2]



Colorectal polyp segmentation network combining polarized self-attention and Transformer

Overview: Among malignant diseases, colorectal cancer is one of the most common cancers in life, and its morbidity and mortality have been high. Therefore, it is urgent to develop an automatic recognition and automatic segmentation algorithm for colorectal polyp image segmentation to help doctors improve the efficiency of diagnosing patients. However, the traditional colorectal polyp segmentation method requires manual extraction of lesion features and the integration strategy will over-rely on the experience of the implementor. Therefore, the traditional colorectal polyp segmentation method is prone to problems such as inaccurate target segmentation, insufficient contrast and blurred edge details during segmentation. In order to solve the problems existing in the traditional method, In this paper, a new colorectal polyp segmentation network TPSA-Net, which combines polarized self-attention and Transformer, is proposed. Firstly, in order to make better use of the semantic information of image blocks at different phase levels to improve the segmentation accuracy of target images, an improved phase sensing hybrid module is designed in this paper, which can dynamically capture multi-scale context information at different levels of colorectal polyp images to improve the accuracy of target segmentation. Secondly, the polarization self-attention module is introduced to fully consider the characteristics of pixels and strengthen the self-attention of the image, so as to improve the contrast between the lesion area and the normal tissue area. Finally, the dynamic capturing ability of the geometric structure of the image was enhanced by the cross-fusion module of the clues, and the complementary characteristics of the two clues in single/multi-frame were improved to solve the problem of blurred edge details during colorectal polyp segmentation. Experiments were conducted on four datasets, CVC-ClinicDB, Kvasir, CVC-ColonDB and ETIS-LaribPolypDB, and the Dice similarity index was 0.946, 0.927, 0.805 and 0.781, respectively. Compared with U-Net, the traditional medical image segmentation network was improved by 12.4%, 14.5%, 29.3% and 37.5 respectively. The average MIou intersection ratio index was 0.901, 0.880, 0.729 and 0.706, respectively, which had certain application value in the diagnosis of colorectal polyps. A large number of experimental results show that the TPSA-Net method proposed in this paper can not only effectively improve the accuracy and contrast of colorectal polyp segmentation, but also overcome the problem of blurred detail in the segmentation image. How to use deep learning technology to research more simple and efficient colorectal polyp segmentation methods is the future focus.

Xie B, Liu Y Q, Li Y L. Colorectal polyp segmentation method combining polarized self-attention and Transformer[J]. *Opto-Electron Eng*, 2024, **51**(10): 240179; DOI: 10.12086/oee.2024.240179

Foundation item: Project supported by National Natural Science Foundation of China (61972264), and Jiangxi University of Science and Technology PhD Start-up Fund (20520010058)

[1]School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China; [2]School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China

* E-mail: 1974718299@qq.com