

文章编号 1004-924X(2024)16-2523-14

渐进式 CNN-Transformer 语义补偿息肉分割网络

李大湘, 李登辉*, 刘 颖, 唐 垚

(西安邮电大学 通信与信息工程学院, 陕西 西安 710121)

摘要:针对结肠镜图像中息肉大小不一、形态复杂、息肉与黏膜界限不清导致分割精度较低的问题,提出了一个渐进式 CNN-Transformer 语义补偿息肉分割网络,以提高结肠镜息肉的分割精度。为了更好地利用来自 CNN 编码器的局部特征和来自 Transformer 编码器的全局特征,设计了一个同层特征交互耦合模块,通过分组交互耦合的方式在空间和通道两个维度上自适应融合来自 CNN 和 Transformer 编码器的特征;然后,针对解码过程中上采样导致的语义丢失问题,设计了一个基于 Query 的语义补偿模块,通过一组可学习的描述子渐进式地集成和分发图像语义,有效提升网络的特征判别能力;实验结果表明,所提网络在 CVC-ClinicDB, CVC-300, Kvasir 以及 CVC-ColonDB 公开数据集上, mDice 分别达到了 94.23%, 90.36%, 92.93%, 80.26%, mIoU 分别达到了 89.87%, 83.75%, 88.21%, 72.09%。与现有的分割网络相比,该网络能够在提升息肉分割有效性的同时保证其泛化性。

关键词:息肉分割;卷积神经网络;Transformer;特征交互;语义补偿

中图分类号: TP391.41 **文献标识码:** A **doi:** 10.37188/OPE.20243216.2523

Progressive CNN-transformer semantic compensation network for polyp segmentation

LI Daxiang, LI Denghui*, LIU Ying, TANG Yao

(College of Communication and Information Engineering, Xi'an University of Posts and Telecommunication, Xi'an 710121, China)

* Corresponding author, E-mail: ldh_wy0908@163.com

Abstract: In response to the problem of low segmentation accuracy in colon polyp images due to varying sizes, complex shapes, and unclear boundaries between polyps and mucosa, a progressive CNN-Transformer semantic compensation polyp segmentation network was proposed to improve the segmentation accuracy of colon polyps. In order to better utilize the local features from the CNN encoder and the global features from the Transformer encoder, a same-layer feature interaction coupling module was designed to adaptively fuse features from the CNN and Transformer encoders in both spatial and channel dimensions through grouped interaction coupling. Then, to address the issue of semantic loss caused by upsampling during the decoding process, a Query-based semantic compensation module was designed. This module gradually integrates and distributes image semantics through a set of learnable descriptors, effectively enhancing the network's feature discrimination capability. The experimental results show that the proposed network achieved mDice scores of 94.23%, 90.36%, 92.93%, and 80.26% on the CVC-ClinicDB,

收稿日期: 2024-03-15; 修订日期: 2024-06-12.

基金项目: 国家自然科学基金 (No. 62071379); 陕西省自然科学基金 (No. 2019JM-604)

CVC-300, Kvasir, and CVC-ColonDB public datasets, respectively. The mIoU scores are 89.87%, 83.75%, 88.21%, and 72.09%, respectively. Compared to existing segmentation networks, the proposed network can effectively improve the effectiveness of polyp segmentation while ensuring its generalization.

Key words: polyp segmentation; convolutional neural network; transformer; feature interaction; semantic compensation

1 引 言

结直肠癌(Colorectal Cancer, CRC)是一种危害人类健康的常见致命疾病。据统计,在全球范围内,CRC已经成为男性第三常见、女性第二常见的癌症^[1]。结直肠息肉作为结直肠癌的癌前病变,其可能发展为晚期结直肠癌,因此及时发现和治疗十分重要^[2]。结肠镜检查是息肉筛查的重要手段,可使结直肠癌发病率降低约40%,死亡率降低约60%^[3]。然而人工检查的误诊率高达25%,而漏诊的息肉会导致晚期结直肠癌,且其存活率不到10%。因此,设计计算机辅助诊断方法,对结肠镜图像中的息肉进行自动、准确地分割,以减轻医生的负担,提高息肉检出率,是非常必要和迫切的^[4]。

近年来,得益于卷积神经网络(Convolutional Neural Networks, CNN)在特征提取方面的巨大潜力,基于CNN的息肉分割取得了一系列研究成果,Fan等^[5]提出一种基于并行反向注意力的息肉分割网络,利用并行的部分解码器在高层聚合语义特征生成初始引导区域,然后使用反向注意模块挖掘边缘信息,在提高精度的同时保证了实时性;Tomar等^[6]提出了一种文本引导注意力网络,通过引入一个辅助分类任务来加权基于文本的嵌入,使网络能够学习额外的特征表示,从而提高了网络对息肉的适应能力;Yue等^[7]提出了一种新型边界约束网络,通过协同学习息肉和边界特征,获得了更好的分割性能;Yin等^[8]从整个数据集出发探索上下文相关性,提出了一个双工上下文关系网络来捕获图像内和交叉图像之间的上下文关系;Zhou等^[9]针对息肉大小多变和边界模糊提出了一种跨层特征聚合网络,该网络能同时利用边界信息和多层次的语义信息来准确分割息肉。尽管CNN在息肉分割中取得了良好的效果,但由于卷积操作的固有局部性,它

无法在高分辨率尺度上建模全局信息。

为提高模型的全局建模能力,许多研究都采用了Transformer的自注意力机制,Wang^[10],Zhou^[11]和Liang^[12]针对息肉形状多变、数据过拟合等问题设计了一种新的息肉分割网络,通过将金字塔Transformer结构引入到息肉分割任务中,提高了网络的泛化能力;Zhang等^[13]采用并行的CNN和Transformer编码器以较浅的方式分别提取局部细节和全局信息,并通过融合两个分支的多层次特征进行息肉分割,同时提高了分割精度和推理速度;Sanderson等^[14]针对输入和输出大小不一致的问题,提出了一种用于全尺寸分割的新架构,从而得到全尺度分割图;Kim等^[15]通过考虑特征图中的不确定区域,提出了一种基于不确定性增强的上下文注意力息肉分割网络;Li等^[16]为解决息肉分割不完整、定位不准确等问题提出了一种新的息肉分割网络,该网络通过集成边缘和语义细节,提高了小目标和不确定区域的分割精度。然而,上述方法要么是将CNN和自注意力机制结合在单一架构中用于息肉分割,要么是将CNN和Transformer进行简单融合之后用于息肉分割,未能针对局部特征和全局特征之间存在的差异性和互补性将两种特征充分融合,并且忽略了解码过程中由于上采样简单融合而产生的语义损失问题,因此未能取得令人满意的分割性能。

针对上述问题,本文结合CNN与Transformer各自的优势,设计了一种并行交互的混合网络结构,即渐进式CNN-Transformer语义补偿(Progressive CNN-Transformer Semantic Compensation, PCTSC)网络,以提高结直肠息肉分割精度,主要创新和贡献总结如下:(1)设计了一个面向结直肠息肉分割的PCTSC网络,它能够有效地融合Transformer分支捕获的全局特征和CNN分支提取的局部特征,从而更好地应对结

直肠息肉分割任务。(2)设计了一个同层特征交互耦合 (Same Layer Features Inter-action Coupling, SLFIC) 模块,通过空间分组交互耦合和通道分组交互耦合机制实现了局部特征与全局特征交互融合,能够在实现两种特征互补的同时保留它们的独特属性,从而更好地利用图像的结构和语义信息,获得更丰富的特征表示。(3)设计了一个基于 Query 的语义补偿 (Query-based Semantic Compensation, QBSC) 模块,通过一组可学习的语义描述子自下而上地学习与分发图像语义,不仅能缓解网络在解码过程中由于上采样而导致的语义损失问题,还能挖掘出更高级的语义线索以增强网络对息肉特征的判别能力。

2 所提模型

2.1 模型整体结构设计

本文设计的 PCTSC 网络整体架构如图 1 所示,主要由 Transformer 全局特征提取器、CNN 局部特征提取器和语义补偿渐进分割器三个部分组成。该网络采用 Mix Transformer-B2 (MiT-B2)^[17] 作为基于 Transformer 的全局特征提取器

的主干,轻型 EfficientNet-B3^[18] 作为基于 CNN 的局部特征提取器的主干。对于输入的结肠镜息肉图像 I ,首先使用 Transformer 编码器和 CNN 编码器提取全局特征和局部特征,对于 Transformer 分支,网络保留了 Mix Transformer-B2 每一层的输出特征图,为简单起见,将其记为 $\{F_1^G, F_2^G, F_3^G, F_4^G\}$;在 CNN 分支中,网络保留了 EfficientNet-B3 的四个高级特征图,其大小与来自 Transformer 分支的四个特征图对应一致,为简单起见,将它们记为 $\{F_1^L, F_2^L, F_3^L, F_4^L\}$;然后,在每个对等层上使用 SLFIC 模块,通过空间分组交互耦合和通道分组交互耦合两种方式自适应地融合 F_i^G 和 F_i^L ($i \in \{1, 2, 3, 4\}$),充分利用这两种异质特征,最终生成增强后的融合特征 MF_i ;接着,在各个相邻层之间采用 QBSC 模块,通过语义描述子集合 D_i ($i \in \{2, 3, 4\}$) 来实现语义信息的逐层传递,以补偿上采样过程中丢失的语义信息,最终得到语义补偿后的输出特征 CF_i ($i \in \{1, 2, 3\}$);最后,通过自上而下的深度监督^[19],得到一组由粗到细的预测分割图 O_i ($i \in \{1, 2, 3, 4\}$),并将 O_1 作为网络的最终预测图。

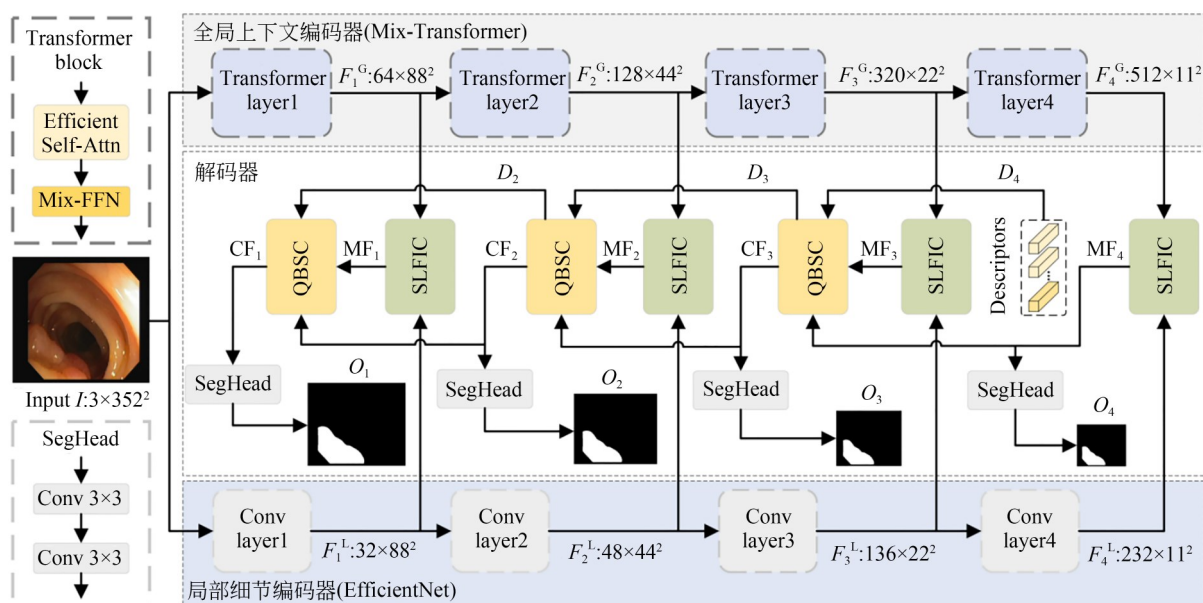


图 1 PCTSC 网络整体架构

Fig. 1 PCTSC network overall architecture

2.2 同层特征交互耦合模块

在基于 CNN 与 Transformer 的特征提取过

程中,由于 CNN 的卷积利于捕获图像的局部细节特征,而 Transformer 的自注意力利于建模图

像的全局上下文信息,因此,可以将两个分支得到的特征融合起来用于息肉分割。先前的融合方法大致可分为两类,第一类是采用拼接或者相加的方式将两种特征直接融合,例如FCBFormer^[14],CovTransNet^[20],这类方法不仅低效,且忽略了两种特征的差异性;第二类是利用注意力机制进行融合,例如TransFuse^[13],DFETC-Net^[21],即在各自分支上采用注意力机制突出有效信息

后进行融合,尽管这类方法保证了两种特征的独特属性,但不能实现有效地互补,降低了融合的有效性。针对上述问题,如图2所示,本节构造了两种交互注意力机制,设计了一个SLFIC模块,该模块通过对两种特征在各自的特征空间进行特征调整,使两种特征在实现互补的同时保留每种特征的独特属性,从而增强息肉的特征表示。

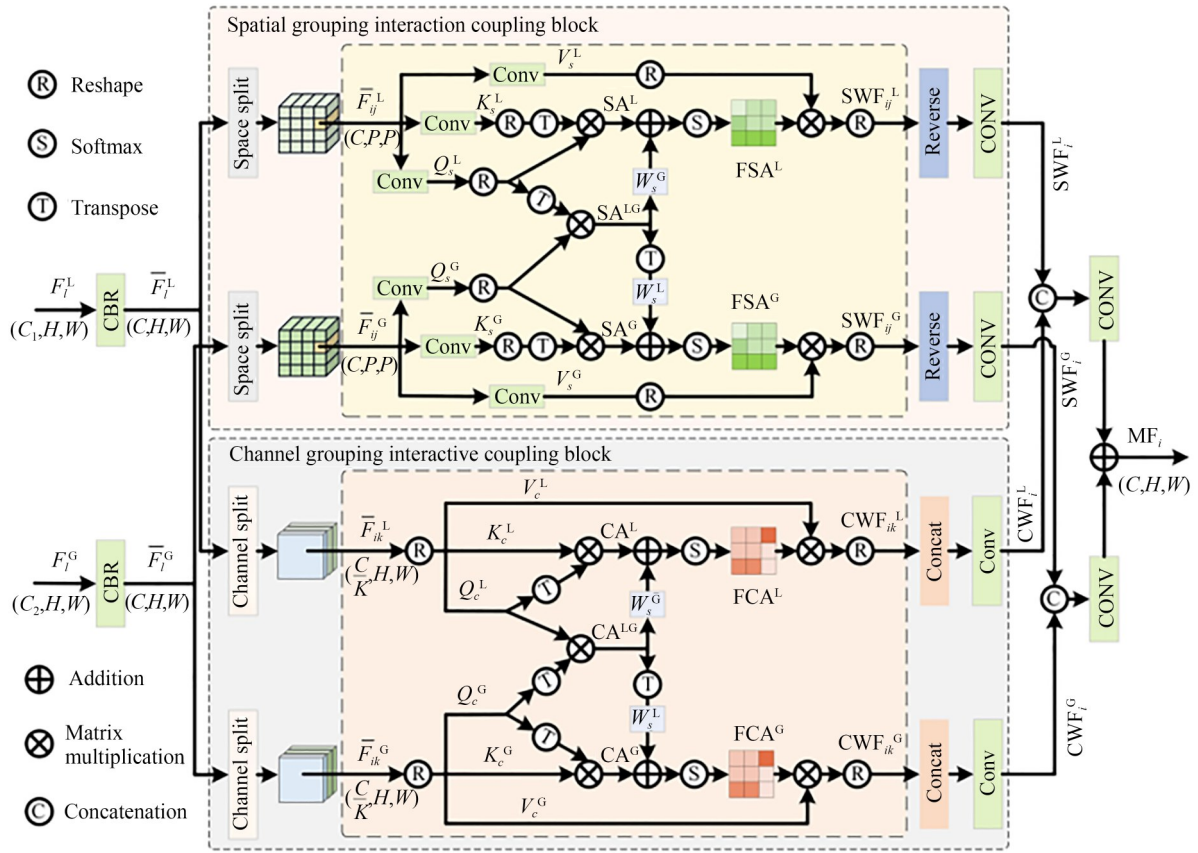


图2 同层特征交互耦合模块

Fig. 2 Same layer features interaction coupling module

设来自CNN与Transformer分支第 i ($i=1, 2, 3, 4$)层的局部与全局特征图谱分别记为 $F_i^L \in \mathbb{R}^{C_1 \times H \times W}$ 与 $F_i^G \in \mathbb{R}^{C_2 \times H \times W}$ 。首先,将它们分别通过一个卷积块进行通道校正,以确保两种输入特征的通道数一致,处理后的特征分别记为 $\bar{F}_i^L$ 和 $\bar{F}_i^G \in \mathbb{R}^{C \times H \times W}$,其中 C, H 和 W 分别表示特征图谱的通道数、高度和宽度,该过程可表示为:

$$\begin{cases} \bar{F}_i^L = \text{CBR}(F_i^L) \\ \bar{F}_i^G = \text{CBR}(F_i^G) \end{cases} \quad (1)$$

其中, $\text{CBR}(\cdot)$ 表示卷积块的处理函数,其包含

3×3 卷积、批量归一化和 ReLU 激活等处理方法。

然后,将 $\bar{F}_i^L$ 和 $\bar{F}_i^G$ 同时输入到空间分组交互耦合 (Spatial Grouping Interactive Coupling, SGIC) 块和通道分组交互耦合 (Channel Grouping Interactive Coupling, CGIC) 块中,以利用两种不同特征之间的互补性从不同角度自适应地将它们进行交互耦合,经过 SGIC 块处理后的特征分别记为 SWF_i^L 和 $SWF_i^G \in \mathbb{R}^{C \times H \times W}$, 经过 CGIC 块处理后的特征分别记为 CWF_i^L 和

$CWF_i^G \in \mathbf{R}^{C \times H \times W}$, 该过程可以描述为:

$$\begin{cases} SWF_i^L, SWF_i^G = SGIC(\bar{F}_i^L, \bar{F}_i^G) \\ CWF_i^L, CWF_i^G = CGIC(\bar{F}_i^L, \bar{F}_i^G) \end{cases} \quad (2)$$

最后,将交互耦合后的局部特征图谱 SWF_i^L 和 CWF_i^L 沿通道进行拼接,并采用一个 3×3 卷积进行融合,全局特征图谱 SWF_i^G 和 CWF_i^G 采用相同的操作进行处理,然后将得到的两组融合特征相加获得模块的最终输出 $MF_i \in \mathbf{R}^{C \times H \times W}$, 该过程可以记为:

$$MF_i = \text{Conv}_{3 \times 3}([\text{SWF}_i^L, \text{CWF}_i^L]) + \text{Conv}_{3 \times 3}([\text{SWF}_i^G, \text{CWF}_i^G]), \quad (3)$$

其中: $[\cdot]$ 表示拼接操作, $\text{Conv}_{3 \times 3}(\cdot)$ 表示 3×3 卷积的处理函数。

2.2.1 空间分组交互耦合块

尽管 $\bar{F}_i^L$ 和 $\bar{F}_i^G$ 是对同一张图像的两组互补性描述,但它们可能在大多数区域上存在相似,仅在某些区域存在差别。因此,它们在进行相互学习时,无需从整个特征图谱中进行探索,只需从对应的局部区域中进行挖掘学习。如图 2 上半部分所示,在位置注意力机制^[22]的启发下,构造了 SGIC 块用于对 $\bar{F}_i^L$ 和 $\bar{F}_i^G$ 进行空间重加权,以融入互补的上下文信息而增强各自的表示能力。为方便描述,这里以 $\bar{F}_i^L$ 的空间重加权为例说明空间分组交互耦合原理。

首先,将 $\bar{F}_i^L$ 沿空间维度划分为大小为 $P \times P$ 的窗口,每个窗口的特征可表示为 $\bar{F}_{ij}^L \in \mathbf{R}^{C \times P \times P}$,再采用三个不同的 1×1 卷积对 $\bar{F}_{ij}^L$ 进行处理,所得结果分别表示为局部查询 Q_s^L 、局部键 K_s^L 与局部值 V_s^L ,同时将全局特征 $\bar{F}_i^G$ 也划分为 $\bar{F}_{ij}^G \in \mathbf{R}^{C \times P \times P}$ 表示的窗口并通过一个 1×1 卷积对其进行处理,所得结果称为全局查询 Q_s^G ,即:

$$\begin{cases} Q_s^L = \text{Conv}_{1 \times 1}(\bar{F}_{ij}^L) \\ K_s^L = \text{Conv}_{1 \times 1}(\bar{F}_{ij}^L) \\ V_s^L = \text{Conv}_{1 \times 1}(\bar{F}_{ij}^L) \\ Q_s^G = \text{Conv}_{1 \times 1}(\bar{F}_{ij}^G) \end{cases}, \quad (4)$$

其中: $\text{Conv}_{1 \times 1}(\cdot)$ 表示 1×1 卷积的处理函数, j 表示第 j 个窗口。

然后,将 K_s^L 的重塑且转置与 Q_s^L 的重塑作矩阵乘法,从而得到局部初始空间注意力 $SA^L \in \mathbf{R}^{P^2 \times P^2}$,同时,为了在局部特征中补充全局信息,同理也利用 Q_s^L 与 Q_s^G 计算互引导空间注意力 $SA^{LG} \in \mathbf{R}^{P^2 \times P^2}$;接着,将 SA^{LG} 通过一个全连接

层后与 SA^L 相加,且经 Softmax 函数处理,从而获得融合空间注意力局部权重矩阵 $FSA^L \in \mathbf{R}^{P^2 \times P^2}$,再将 V_s^L 的重塑与其作矩阵乘法完成空间注意力重加权,每个窗口的空间重加权过程可总结如下:

$$\begin{cases} SA^L = \text{Trans}(\text{Rhp1} - D(K_s^L)) \times \text{Rhp1} - D(Q_s^L) \\ SA^{LG} = \text{Trans}(\text{Rhp1} - D(Q_s^L)) \times \text{Rhp1} - D(Q_s^G) \\ FSA^L = \text{Softmax}((SA^L + W_s^G SA^{LG})) \\ SWF_{ij}^L = \text{Rhp2} - D(\text{Rhp1} - D(V_s^L) \times FSA^L) \end{cases}, \quad (5)$$

其中: $SWF_{ij}^L \in \mathbf{R}^{C \times P \times P}$ 表示第 j 个窗口的全局信息引导空间注意力局部特征重加权处理结果, $\times$ 表示矩阵乘法, $\text{Rhp1}-D(\cdot)$ 表示将维度为 $\mathbf{R}^{D_1 \times D_2 \times D_3}$ 的矩阵重塑为 $\mathbf{R}^{D_1 \times D_2 \times D_3}$, $\text{Rhp2}-D(\cdot)$ 表示 $\text{Rhp1}-D(\cdot)$ 的逆操作, $\text{Trans}(\cdot)$ 和 $\text{Softmax}(\cdot)$ 分别表示转置和 Softmax 函数, W_s^G 表示一个线性权重矩阵。

最后,将重加权后的所有窗口恢复为大小为 $H \times W$ 的特征图谱,并通过一个 3×3 卷积进行处理,以平滑窗口间的接缝来获得最终输出 $SWF_i^L \in \mathbf{R}^{C \times H \times W}$ 。

以上描述了 $\bar{F}_i^L$ 空间重加权的细节, $\bar{F}_i^G$ 的空间重加权具有类似的过程,最终会得到输出特征 $SWF_i^G \in \mathbf{R}^{C \times H \times W}$ 。

2.2.2 通道分组交互耦合块

由于空间交互旨在为每个像素建立丰富的上下文关系,受通道注意力机制^[22]的启发,如图 2 下半部分所示,本节进一步引入 CGIC 块用于对 $\bar{F}_i^L$ 和 $\bar{F}_i^G$ 进行通道重加权,以利用不同特征的通道图之间的相互依赖关系,增强相互依赖的特征图谱,并改进特定语义的特征表示。为方便描述,这里以 $\bar{F}_i^L$ 的通道重加权为例说明通道分组交互耦合原理。

首先,将 $\bar{F}_i^L$ 沿通道维度均匀地划分为 K 组,每个分组的特征可记为 $\bar{F}_{ik}^L \in \mathbf{R}^{(C/K) \times H \times W}$,再对 $\bar{F}_{ik}^L$ 进行重塑操作,得到局部查询 Q_c^L 、局部键 K_c^L 与局部值 V_c^L ,同时将全局特征 $\bar{F}_i^G$ 划分为 K 组并将其分组特征 $\bar{F}_{ik}^G \in \mathbf{R}^{(C/K) \times H \times W}$ 重塑为全局查询 Q_c^G ,即:

$$\begin{cases} Q_c^L = K_c^L = V_c^L = \text{Rhp1} - D(\bar{F}_{ik}^L) \\ Q_c^G = \text{Rhp1} - D(\bar{F}_{ik}^G) \end{cases}, \quad (6)$$

其中: k 表示第 k 组。然后,将 K_c^L 与 Q_c^L 的转置作

矩阵乘法,得到局部初始通道注意力 $CA^L \in R^{(C/K) \times (C/K)}$,同时,可以利用 Q_c^L 与 Q_c^G 计算互引导通道注意力 $CA^{LG} \in R^{(C/K) \times (C/K)}$;接着,将 CA^{LG} 通过一个全连接层后与 CA^L 相加,且经 Softmax 函数处理,进而获得融合通道注意力局部权重矩阵 $FCA^L \in R^{(C/K) \times (C/K)}$,再将其与 V_c^L 作矩阵乘法完成通道注意力重加权,每个分组的通道重加权过程可描述如下:

$$\begin{cases} CA^L = K_c^L \times \text{Trans}(Q_c^L) \\ CA^{LG} = Q_c^L \times \text{Trans}(Q_c^G) \\ FCA^L = \text{Softmax}((CA^L + W_c^G CA^{LG})) \\ CWF_{ik}^L = \text{Rhp2} - D(FCA^L \times V_c^L) \end{cases} \quad (7)$$

其中, $CWF_{ik}^L \in R^{(C/K) \times H \times W}$ 表示第 k 个分组的全局特征引导通道注意力局部特征重加权处理结果。

最后,将重加权后的所有分组特征沿通道维度进行拼接,并通过一个 1×1 卷积进行处理,以融合不同组的信息并生成最终的输出 $CWF_i^L \in R^{C \times H \times W}$ 。

以上描述了 $\bar{F}_i^L$ 通道重加权的细节, $\bar{F}_i^G$ 的通道重加权具有类似的过程,最终会得到输出特征 $CWF_i^G \in R^{C \times H \times W}$ 。

2.3 基于 Query 的语义补偿模块

随着网络层数的增加,网络获得的语义信息越来越丰富,而空间信息越来越模糊。为了使网络获得准确、完整的分割结果,通常在解码过程中将深层语义信息传递到浅层,以实现语义信息与空间信息的融合。然而,现有的一些分割方法^[23-25]只是简单地将两种不同层次的特征融合在一起,这会导致信息多样性的丢失。为了避免上述问题,如图 3 所示,受基于 DETR^[26] 的方法的启发,本节构建了一个基于迭代式交叉注意力的语义补偿机制,设计了一个 QBSC 模块,该模块采用一组可学习的语义描述子集合从高层特征中进一步挖掘语义线索,并将其分发给上采样后的解码特征,从而缓解上采样简单融合所带来的语义损失问题,使网络获得更具判别性的输出特征。

以第 $i(i=1,2,3)$ 个 QBSC 模块为例来说明自底向上的语义补偿原理,设模块的输入分别为来自第 i 层的交互耦合特征 $MF_i \in R^{C \times H \times W}$ 、第 $i+1$ 层的语义补偿特征 $CF_{i+1} \in R^{C \times H/2 \times W/2}$ 和初始语义描述子集合 $D_{i+1} \in R^{N \times C}$, N 表示语义描述

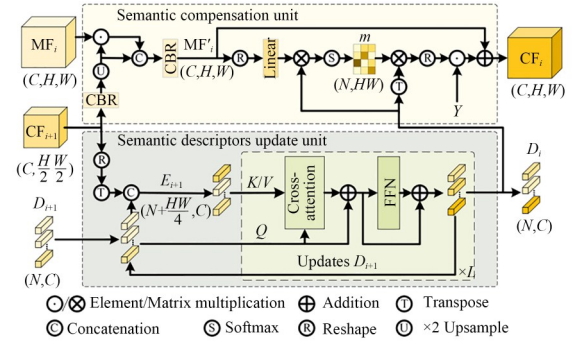


图 3 基于 Query 的语义补偿模块

Fig. 3 Query-based semantic compensation module

子数量。首先,将 CF_{i+1} 和 D_{i+1} 一起送入语义描述子更新 (Semantic Descriptor Update, SDU) 单元,生成包含更高级语义信息的语义描述子集合 $D_i \in R^{N \times C}$;接着,将 MF_i , CF_{i+1} 和 D_i 输入到语义补偿 (Semantic Compensation, SC) 单元中进行语义补偿,最终得到语义补偿特征 $CF_i \in R^{C \times H \times W}$,整个过程可表示为:

$$\begin{cases} D_i = \text{SDG}(D_{i+1}, CF_{i+1}) \\ CF_i = \text{SC}(MF_i, CF_{i+1}, D_i) \end{cases} \quad (8)$$

语义描述子更新单元: 为了得到包含高阶语义关系的语义描述子集合,需要对输入的 D_{i+1} 做进一步处理以实现 D_{i+1} 的更新。首先,将 CF_{i+1} 进行重塑与转置操作后与 D_{i+1} 沿空间维度进行拼接,得到 $E_{i+1} \in R^{(N+HW/4) \times C}$;接着,将 D_{i+1} 作为查询、 E_{i+1} 作为键和值一起送入多头交叉注意力层,多头机制可以从不同的子空间提取深层语义信息,从而产生更高级的语义描述子集合。经过多头交叉注意力层的一次迭代处理后,更新的查询将取代最初的查询,并同时更新键和值,迭代 L 次的过程可描述为:

$$\begin{cases} E_{i+1}^{[t]} = [D_{i+1}^{[t]}, \text{Trans}(\text{Rhp1} - D(CF_{i+1}))] \\ D_{i+1}^{[t+1]} = \text{MHCA}(D_{i+1}^{[t]}, E_{i+1}^{[t]}, E_{i+1}^{[t]}) \end{cases}, \quad 0 \leq t < L \quad (9)$$

其中: $\text{MHCA}(\cdot)$ 表示带有前馈网络的多头交叉注意力层的处理函数, t 为第 t 次迭代。特别地,对于初始迭代, $D_{i+1}^{[0]} = D_{i+1}$, $E_{i+1}^{[0]} = E_{i+1}$ 。

经过 L 次迭代后,最终得到语义描述子集合 $D_i \in R^{N \times C}$, 即:

$$D_i = D_{i+1}^{[L]}. \quad (10)$$

特别地, D_4 在网络中被作为网络参数随机初始化。

语义补偿单元:得到 D_i 后,分两步实现对 MF_i 的语义补偿。第一步,首先将 CF_{i+1} 通过一个 3×3 卷积块和上采样进行处理,得到一个 3×3 卷积块和上采样进行处理,得到 $CF'_{i+1} \in R^{C \times H \times W}$,然后将 CF'_{i+1} 与 MF_i 作元素乘法,以排除 MF_i 中模糊的图像结构和背景噪声,最后将其与 CF'_{i+1} 沿通道维度拼接并通过一个 3×3 卷积块进行处理,得到一个增强的特征图谱 $MF'_i \in R^{C \times H \times W}$,即:

$$\begin{cases} CF'_{i+1} = \text{up}(\text{CBR}(CF_{i+1})) \\ MF'_i = \text{CBR}([CF'_{i+1} \odot MF_i, CF'_{i+1}]) \end{cases}, \quad (11)$$

其中:up 表示上采样 2 倍, $\odot$ 表示元素乘法。特别地,在第三个 QBSC 模块中, $CF_4 = MF_4$ 。第二步,将 D_i 分发给 MF'_i 中的每个像素,以实现对 MF'_i 的语义补偿。具体来说,首先将 MF'_i 的重塑通过一个全连接层后与 D_i 作矩阵乘法,接着采用 softmax 函数对其进行归一化处理,得到相似度矩阵 $m \in R^{N \times HW}$,然后将 D_i 的转置与 m 作矩阵乘法以实现语义描述子的软分配,最后对其进行重塑并与 MF'_i 相加,得到最终的语义补偿特征 $CF_i \in R^{C \times H \times W}$,即:

$$\begin{cases} m = \text{softmax}\left(\frac{D_i \times (\text{Rhp1} - D(MF'_i)W^p)}{\sqrt{d}}\right) \\ CF_i = \gamma \cdot \text{Rhp2} - D((\text{Trans}(D_i) \times m)) + MF'_i \end{cases}, \quad (12)$$

其中: W^p 表示一个线性权重矩阵, d 表示 MF'_i 的通道数, γ 表示一个可学习的权重参数。

3 实验与结果分析

3.1 实验数据集

为了验证所提网络对结肠镜图像中息肉的分割性能,本文选择了 4 个公开的息肉分割数据集进行实验验证分析,包括 CVC-ClinicDB^[27], Kvasir-SEG^[28], CVC-300^[29] 和 CVC-ColonDB^[30], 这些数据集的简要介绍如表 1 所示。

其中,训练集包含来自 CVC-ClinicDB 的 550 个样本和来自 Kvasir-SEG 的 900 个样本,而测试集则包括 CVC-ClinicDB 和 Kvasir-SEG 的剩余样本,以及剩余两个数据集的所有样本。这种数据集划分策略的好处在于, CVC-ClinicDB 和 Kvasir-SEG 上的实验结果属于域内测试,能够反映每种分割方法的有效性;而剩余两个数据集上

表 1 数据集简介

Tab. 1 Introduction to the dataset

数据集	年份	原始分辨率	图像总数
CVC-ClinicDB	2015	384×288	612
Kvasir-SEG	2020	332×487~1 920×1 072	1 000
CVC-300	2017	574×500	60
CVC-ColonDB	2016	574×500	380

的实验结果属于域外测试,能够反映每种分割方法的泛化能力。

3.2 实验环境及参数设置

本文所用实验环境为: CPU: 15 vCPU AMD EPYC 7543 32-Core Processor; GPU: A40 (48 GB) * 1, Ubuntu20. 04 操作系统, Pytorch 1. 10. 0 深度学习框架, Python 3. 8 编程语言, Cuda 11. 3 环境。

网络训练的参数设置: 在 SLFIC 模块中,校正后的通道数 C 设置为 64, SGIC 块的窗口大小 P 设置为 3, CGIC 块的组数 K 设置为 16。在 QBSC 模块中,语义描述子的数量 N 设置为 32,语义描述子更新迭代次数 L 设置为 4,多头交叉注意力的头部数量设置为 8。在训练网络时,将输入图像分辨率统一压缩为 352×352 ,使用 Adam 优化器,采用加权 IOU 损失和加权 BCE 损失^[31] 作为网络的损失函数,学习率设定为 $1e-4$,训练 40 个 epoch, batch size 设置为 16,采用随机旋转、随机翻转、随机平移、亮度变化、不同比例(0. 75, 1, 1. 25)的多尺度调整等数据增强方法进行训练。

3.3 评价指标

本文采用六个常用的评价指标来评估 PCTSC 网络与其他对比方法的分割性能,包括平均 Dice 相似系数(mDice)、息肉平均 IoU(mIoUp)、S-度量(S_a)、最大 E-度量($E_{\phi}^{\max}$)、加权 F-度量(F_{β}^w)和平均绝对误差(MAE),它们的具体定义如下:

mDice 和 mIoUp 用于衡量模型的分割结果与真实结果之间的重叠程度,公式如下:

$$\text{mDice} = \frac{2TP}{FN + FP + 2TP}, \quad (13)$$

$$\text{mIoUp} = \frac{TP}{FN + FP + TP}, \quad (14)$$

其中:TP 表示真阳性,TN 表示真阴性,FP 表示假阳性,FN 表示假阴性。

S_a 用于评估两幅图像在结构信息方面的相似性,公式如下:

$$S_a = (1 - \alpha)S_s + \alpha S_l, \quad (15)$$

其中: S_s 表示结构相似性指数, S_l 表示亮度相似性指数, α 为权重超参数。

F_β^ω 是一种在不平衡类别分布情况下评估分割性能的指标,公式如下:

$$F_\beta^\omega = \frac{(1 + \beta^2)(P * R)}{\beta^2 P + R}, \quad (16)$$

其中: P 表示精度, R 表示召回率。

MAE 用于衡量模型的预测值与真实值之间的平均绝对差异,公式如下:

$$\text{MAE} = \frac{1}{n} \sum_i^n |y_i - x_i|, \quad (17)$$

其中: y_i 表示第 i 个样本的真实值。 x_i 表示第 i 个样本的预测值, n 表示样本数量。

E_ϕ 用于衡量局部像素值和图像均值之间的对齐性,公式如式(18)所示:

$$E_\phi = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \phi(i, j), \quad (18)$$

其中: $\phi(i, j)$ 表示局部像素值与图像均值之间的增强对齐矩阵, H 和 W 表示输入的高度和宽度。

3.4 综合对比实验

本节将所提出的 PCTSC 网络与 17 种最先进的医学图像分割方法进行比较,这些方法可以分为三大类,第一类是基于 CNN 的分割网络,分别包括 U-Net^[32], U-Net++^[33], CE-Net^[34], PraNet^[5], UACA-Net^[15], DCRNet^[8], CFA-Net^[9] 和 META-Unet^[35];第二类是基于 Transformer 的分割网络,分别包括 SSFormer^[10], MCE-Net^[16], CGMA-Net^[36] 和 Colonformer^[37];第三类是结合 CNN 和 Transformer 的分割网络,分别为 FCBFormer^[14], TransFuse^[13], DFETC-Net^[21], MC-DC^[38] 和 TGDAUNet^[39]。

3.4.1 定量分析

为了评价 PCTSC 网络的有效性,表 2 展示了各个分割网络在 CVC-ClinicDB 和 Kvasir-SEG 数据集上的分割对比结果,由于训练集来自这两个数据集,因此该表可以反映出各分割网络

表 2 不同方法在 CVC-ClinicDB 和 Kvasir-SEG 数据集上性能对比

Tab. 2 Performance comparison of different methods on CVC-ClinicDB and Kvasir-SEG datasets

Methods	CVC-ClinicDB						Kvasir-SEG					
	mDice	mIoUp	MAE	S_a	F_β^ω	E_ϕ	mDice	mIoUp	MAE	S_a	F_β^ω	E_ϕ
U-Net ^[32]	0.901	0.849	0.011	0.921	0.895	0.971	0.844	0.772	0.042	0.873	0.819	0.901
U-Net++ ^[33]	0.911	0.856	0.013	0.929	0.904	0.965	0.840	0.769	0.044	0.872	0.813	0.901
CE-Net ^[34]	0.914	0.867	0.015	0.933	0.911	0.965	0.894	0.835	0.034	0.903	0.877	0.908
PraNet ^[5]	0.930	0.887	0.008	0.943	0.919	0.964	0.894	0.842	0.033	0.901	0.881	0.938
UACA-Net ^[15]	0.937	0.889	0.008	0.941	0.929	0.982	0.906	0.855	0.026	0.914	0.897	0.948
DCRNet ^[8]	0.935	0.895	0.008	0.943	0.927	0.985	0.882	0.831	0.033	0.901	0.870	0.917
CFA-Net ^[9]	0.865	0.810	0.017	0.902	0.863	0.959	0.893	0.837	0.026	0.906	0.878	0.923
META-Unet ^[35]	0.912	0.861	0.009	0.929	0.907	0.975	0.898	0.843	0.029	0.911	0.884	0.926
SSFormer ^[10]	0.893	0.840	0.017	0.920	0.884	0.950	0.917	0.866	0.022	0.924	0.910	0.940
Colonformer ^[37]	0.930	0.891	0.007	0.946	0.920	0.966	0.922	0.877	0.021	0.929	0.908	0.948
MCE-Net ^[16]	0.925	0.877	0.008	0.939	0.917	0.973	0.912	0.865	0.023	0.920	0.903	0.935
CGMA-Net ^[36]	0.939	0.903	0.007	0.951	0.933	0.984	0.918	0.874	0.024	0.927	0.901	0.941
FCBFormer ^[14]	0.897	0.844	0.014	0.921	0.886	0.955	0.913	0.860	0.027	0.915	0.899	0.942
TransFuse ^[13]	0.880	0.821	0.017	0.913	0.869	0.952	0.872	0.809	0.035	0.890	0.855	0.923
DFETC-Net ^[21]	0.893	0.846	0.010	0.924	0.889	0.957	0.899	0.839	0.028	0.906	0.883	0.936
MC-DC ^[38]	0.930	0.881	0.011	0.937	0.927	0.981	0.907	0.853	0.027	0.915	0.898	0.937
TGDAUNet ^[39]	0.911	0.861	0.009	0.928	0.910	0.966	0.898	0.843	0.031	0.907	0.887	0.934
PCTSC	0.942	0.899	0.006	0.946	0.937	0.985	0.929	0.882	0.021	0.929	0.919	0.952

的有效性。由表 2 可以看出,基于 CNN 的分割网络在结肠息肉分割中表现一般,例如,在 CVC-ClinicDB 数据集上,表现最好的 UACANet 的 mDice, mIoUp, MAE, S_a , F_{β}^w 和 E_{ϕ} 分别只有 0.937, 0.889, 0.008, 0.941, 0.929 和 0.982, 相比之下,基于 Transformer 的分割网络要优于基于 CNN 的分割网络,例如,在 CVC-ClinicDB 数据集上,表现最好的 CGMA-Net 在各个指标上都要优于 UACANet;而在 Kvasir-SEG 数据集上,基于 Transformer 的任何一个分割网络都要优于基于 CNN 的分割网络。对于结合了 CNN 和 Trans-former 的分割网络,它们的分割性能并不理想,有些网络的分割性能甚至不如基于 CNN 的分割网络。而本文所提出的息肉分割网络

PCTSC 在两个数据集上都取得了较好的结果,六个评价指标都优于其他基于 CNN 的息肉分割网络,例如,与 PraNet 相比,它在 CVC-ClinicDB 的 mDice 值和 mIoUp 值分别提升了 1.2% 和 1.2%;在 Kvasir-SEG 的 mDice 和 mIoUp 分别提升了 3.5% 和 5.0%;另外,与采用 Transformer 作为骨干网的分割方法相比,PCTSC 网络仍然是领先的,例如,与表现最好的 CGMA-Net 相比,它在 CVC-ClinicDB 的 mDice 值提升了 0.3%, mIoUp 值略低于 CGMA-Net,在 Kvasir-SEG 的 mDice 和 mIoUp 分别提升了 1.1% 和 0.8%;而与其他结合了 CNN 和 Transformer 的分割网络相比,PCTSC 网络仍然是最优的。这说明 PCTSC 网络能够有效提高息肉分割性能。

表 3 不同方法在 CVC-300 和 CVC-ColonDB 数据集上性能对比

Tab. 3 Performance comparison of different methods on CVC-300 and CVC-ColonDB datasets

Methods	CVC-300						CVC-ColonDB					
	mDice	mIoUp	MAE	S_a	F_{β}^w	E_{ϕ}	mDice	mIoUp	MAE	S_a	F_{β}^w	E_{ϕ}
U-Net ^[32]	0.670	0.588	0.019	0.802	0.636	0.880	0.588	0.507	0.051	0.741	0.563	0.813
U-Net++ ^[33]	0.709	0.627	0.017	0.827	0.666	0.863	0.591	0.511	0.049	0.747	0.566	0.807
CE-Net ^[34]	0.893	0.825	0.008	0.930	0.870	0.960	0.732	0.655	0.047	0.824	0.710	0.858
PraNet ^[5]	0.891	0.828	0.008	0.934	0.869	0.952	0.713	0.641	0.042	0.816	0.693	0.845
UACANet ^[15]	0.879	0.802	0.009	0.918	0.844	0.950	0.734	0.662	0.037	0.828	0.721	0.885
DCRNet ^[8]	0.871	0.806	0.007	0.918	0.841	0.960	0.650	0.584	0.050	0.782	0.639	0.844
CFA-Net ^[9]	0.885	0.818	0.007	0.928	0.861	0.961	0.667	0.591	0.048	0.789	0.656	0.868
META-Unet ^[35]	0.892	0.819	0.008	0.925	0.868	0.965	0.688	0.614	0.044	0.800	0.672	0.853
SSFormer ^[10]	0.887	0.815	0.009	0.927	0.861	0.953	0.779	0.696	0.034	0.845	0.753	0.883
Colonformer ^[37]	0.875	0.804	0.008	0.921	0.848	0.949	0.807	0.731	0.031	0.867	0.784	0.901
MCE-Net ^[16]	0.890	0.824	0.010	0.928	0.868	0.955	0.791	0.711	0.032	0.851	0.775	0.903
CGMA-Net ^[36]	0.891	0.829	0.008	0.935	0.865	0.961	0.792	0.721	0.033	0.863	0.772	0.896
FCBFormer ^[14]	0.878	0.806	0.010	0.919	0.847	0.948	0.787	0.706	0.034	0.848	0.756	0.884
TransFuse ^[13]	0.850	0.777	0.012	0.913	0.817	0.921	0.734	0.646	0.036	0.825	0.707	0.869
DFETC-Net ^[21]	0.845	0.766	0.014	0.907	0.805	0.921	0.653	0.572	0.051	0.782	0.636	0.851
MC-DC ^[38]	0.888	0.821	0.008	0.933	0.862	0.947	0.778	0.691	0.030	0.846	0.754	0.894
TGDAUNet ^[39]	0.894	0.828	0.008	0.933	0.870	0.957	0.756	0.681	0.035	0.842	0.741	0.883
PCTSC	0.904	0.838	0.006	0.936	0.882	0.969	0.803	0.721	0.029	0.860	0.781	0.910

为了进一步评估各分割网络的泛化能力,在 CVC-300 和 CVC-ColonDB 数据集上也进行了对比实验,表 3 列出了各分割网络的实验结果,由于这两个数据集中的图像属于网络的未知样本,因此其实验结果能够反映训练后网络的泛化能力。由表 3 可以看出,与其他对比网络相比,PCTSC

网络具有良好的泛化能力。具体来说,它在两个数据集上的六个指标优于所有基于 CNN 以及结合 CNN 和 Trans-former 的所有分割网络,例如,在这两个数据集上,它的 mDice 值分别比 PraNet 高出 1.3% 和 9.0%, mIoUp 值分别高出 1.0% 和 8.0%,与 TGDAUNet 相比,它的 mDice 值分别

高出 1.0% 和 4.7%, mIoUp 值分别高出 1.0% 和 4.0%。相较于基于 Transformer 的息肉分割方法, 它在 CVC-300 数据集上的六个指标全部优于其他分割网络, 而在 CVC-ColonDB 数据集上, 它的某些指标仅略低于 ColonFormer, 这可能是由于该数据集与其他三个数据集的领域差距较大, 面临的挑战也多种多样, 从而导致分割性能不理想。尽管 ColonFormer 在该数据集上表现较好, 但在其他三个数据集上的表现一般, 远不如 PCTSC 网络。这表明 PCTSC 网络能够在提高分割有效性的同时保持良好的泛化能力, 在有效性和泛化性之间实现了更好的权衡。

3.4.2 定性分析

图 4 展示了 PCTSC 网络与 17 个对比网络在 8 张结肠镜图像上的测试结果, 这 8 张图像在息肉大小变化、边界模糊和背景干扰方面存在挑战。从图 4 可以看出, 基于 CNN 的分割网络并不能对结肠息肉进行有效分割, 例如, 在第二、五、六行, U-Net 和 U-Net++ 几乎未能分割出息肉, 而在其他行, 它们的分割结果也有很大残缺, CE-Net 由于引入了多尺度特征提取块从而能够感知出息肉位置, 但存在严重的过分割和欠分割问题; 其他分割网络虽然能够识别出息肉, 但其分

割结果仍然存在缺失, 例如, 在第二行和第六行, DCRNet 和 CFA-Net 虽然能够定位出息肉的位置, 但不能实现完整的分割, 这是因为当息肉与其周围的正常组织非常相似时, 由于卷积的局限性, 网络很难区分息肉与正常组织而导致分割不全; 基于 Transformer 的分割网络尽管都能识别出息肉, 但同时不能实现完整的息肉分割, 如第三、四、七、八行所示, 甚至出现误分割的情况, 如五、六行所示。尽管基于 CNN 和 Transformer 的分割网络将两者结合用于息肉分割, 但在这些具有挑战性的图像上的分割效果却不理想, 甚至有些方法不如基于 CNN 的方法, 这是由于这些方法只是将 CNN 与 Transformer 进行了简单的特征融合或相互嵌入, 未能充分结合两者的优势。相比之下, PCTSC 网络在这些图像上均能取得很好的分割效果, 例如, 在第五行和第六行, 息肉与背景存在很高的相似性, 其他方法要么忽略它, 要么错误地将正常组织识别为息肉, 而 PCTSC 网络基本上可以实现完整、准确的分割; 在第三、四、七、八行, PCTSC 网络相比于其他方法, 不仅能够检测出息肉的完整轮廓, 还能有效保存局部细节。总的来说, PCTSC 网络能够有效适应不同场景下的结肠息肉分割任务。

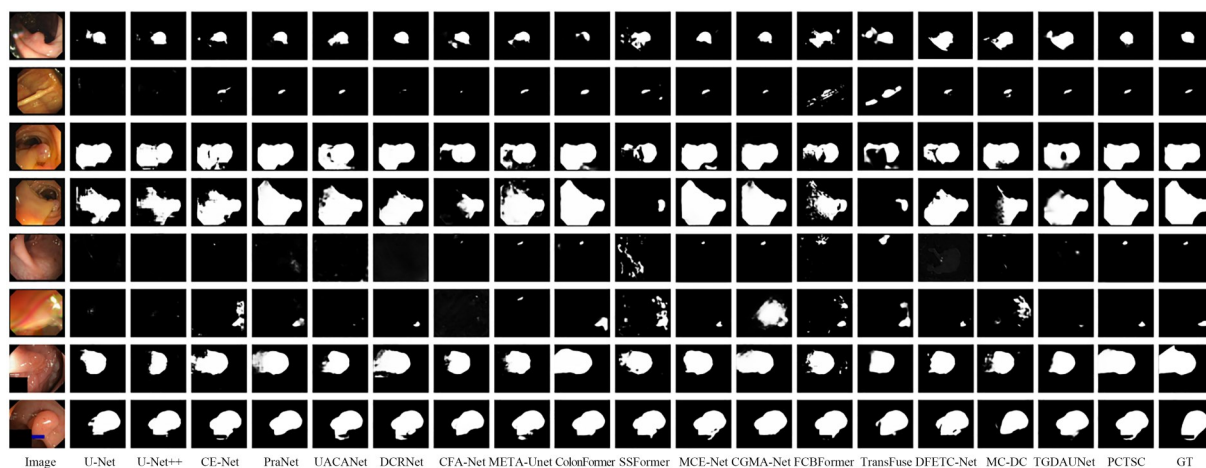


图 4 不同方法的分割结果对比

Fig. 4 Comparison of segmentation results of different methods

3.4.3 计算效率分析

本节将 PCTSC 网络与另外 17 种对比网络在计算效率方面进行了比较, 表 4 显示了各个网络的浮点运算数 (FLOPs) 和参数量 (Params)。

由表 4 可以看出, META-Unet 所需的 FLOPs 和 Params 是最少的, 但它的分割性能一般, 在定量和定性实验中远不如 PCTSC 网络, 如表 2~表 3、图 4 所示。PCTSC 网络的 FLOPs 和 Params 分

表 4 不同方法的计算效率比较

Tab. 4 Comparison of computational efficiency of different methods

Methods	FLOPs/G	Params/M
U-Net ^[32]	152.154	39.396
U-Net++ ^[33]	377.449	47.176
CE-Net ^[34]	16.828	29.003
PraNet ^[5]	13.149	30.498
UACANet ^[15]	59.646	67.109
DCRNet ^[8]	17.268	28.734
CFA-Net ^[9]	55.360	25.239
META-Unet ^[35]	9.716	21.696
SSFormer ^[10]	22.746	46.399
Colonformer ^[37]	22.982	52.945
MCE-Net ^[16]	39.212	58.459
CGMA-Net ^[36]	33.121	30.962
FCBFormer ^[14]	73.295	52.943
TransFuse ^[13]	21.788	26.174
DFETC-Net ^[21]	66.013	46.092
MC-DC ^[38]	126.950	100.979
TGDAUNet ^[39]	105.145	261.502
PCTSC	16.264	34.586

别为 16.264 G 和 34.586 M,所需的 FLOPs 和 Params 比一些基于 CNN 分割网络要少,例如 U-Net,UACANet。与基于 Transformer 的分割方法相比,PCTSC 网络所需的 FLOPs 是最少的,且所需的 Params 只比 CGMA-Net 略高。在基于 CNN 与 Transformer 的分割方法中,PCTSC 网络所需的 FLOPs 仍然是最少的。总的来说,PCTSC 网络在精度和计算效率之间得到了更适当的权衡。

3.5 消融实验

为了验证 PCTSC 网络中各模块在息肉分割任务中的有效性,设置了如下消融实验,在没有添加任何模块的基础模型(Baseline)上,分别应用 SLFIC 模块与 QBSC 模块。在相同数据集和实验设置条件下,逐步分析各模块的作用,验证它们的有效性,结果如表 5 所示。为了方便比较,本节只采用 mDice 和 mIoUp 进行比较。从表 5 的数据对比可以看出,只采用 Baseline 进行息肉分割所获得的分割性能最差,分别加入 SLFIC 模块和 QBSC 模块后,在各个数据集上均能获得比 Baseline 更好的分割性能,例如,在 CVC-ColonDB 数据集上,Baseline 的 mDice 和 mIoUp 值分别为 0.793 和 0.701,加入 SLFIC 模块后,mDice 和 mIoUp 分别提升了 2.2% 和 2.0%,这是由于 SLFIC 模块针对局部特征和全局特征的差异性和互补性进行特征交互耦合,将来自不同编码器的不同特征进行了充分融合,从而使网络获得了更加丰富的息肉特征表示。在 Kvasir-SEG 数据集上,加入 QBSC 模块之后,mDice 和 mIoUp 相比 Baseline 分别提升了 1.2% 和 1.8%,说明 QBSC 模块实现了高层语义信息与低层空间信息的有效结合,增强了网络的判别特征表示能力。值得注意的是,在 CVC-ColonDB 数据集上,使用 SLFIC 模块有时会获得比使用两个模块更好的性能,这可能是由于四个数据集具有较大的领域差距和多样化的挑战,而在该数据集中,设计的 SLFIC 模块可能比两个模块的组合更适合应对这些挑战。相较于其他情况,将全部模块嵌入后分割性能达到最佳,各指标提升比较明显,且两个模块具有一定的互补性,这表明 PCTSC 网络在结肠息肉分割任务中能够获得更好的分割性能。

表 5 各组成模块消融实验结果

Tab. 5 Results of ablation experiments for different modules

Methods	CVC-ClinicDB		Kvasir-SEG		CVC-300		CVC-ColonDB	
	mDice	mIoUp	mDice	mIoUp	mDice	mIoUp	mDice	mIoUp
Baseline	0.933	0.886	0.915	0.861	0.883	0.814	0.793	0.701
Baseline+SLFIC	0.939	0.893	0.922	0.873	0.902	0.837	0.815	0.734
Baseline+QBSC	0.934	0.890	0.927	0.879	0.903	0.837	0.798	0.721
PCTSC	0.942	0.899	0.929	0.882	0.904	0.838	0.803	0.721

4 结 论

为了进一步提高结肠镜图像中息肉的分割精度,本文设计了一个渐进式 CNN-Transformer 语义补偿息肉分割网络,首先,构建了一个同层特征交互耦合模块,采用空间分组交互耦合和通道分组交互耦合两种方式充分融合来自 CNN 编码器的局部特征和来自 Transformer 编码器的全局特征,增强了网络对局部特征和全局特征的联合表征能力;其次,为了缓解解码过程中的语义

丢失问题,提出了一个基于 Query 的语义补偿模块,渐进地实现不同层之间的语义传递,有效提升了网络对息肉特征的判别能力;在 CVC-ClinicDB、Kvasir-SEG、CVC-300 和 CVC-ColonDB 四个公开数据集上,mDice 分别达到了 94.2%,92.9%,90.4% 和 80.3%,mIoUp 分别达到了 89.9%,88.2%,83.8% 和 72.1%,能够有效提高结肠镜图像中息肉的分割精度。后续将进一步研究提升网络的泛化能力,优化网络结构,在保证有效性的同时提升泛化性。

参考文献:

- [1] WORLD HEALTH ORGANIZATION. International agency for research on cancer[J]. 2023.
- [2] 中华医学会肿瘤学分会早诊早治学组. 中国结直肠癌早诊早治专家共识(2023版)[J]. 中华医学杂志, 2023, 103(48): 3896-3908.
Expert consensus on the early diagnosis and treatment of colorectal cancer in China (2023 edition) [J]. *National Medical Journal of China*, 2023, 103(48): 3896-3908. (in Chinese)
- [3] SIEGEL R L, WAGLE N S, CERCEK A, *et al.* Colorectal cancer statistics, 2023[J]. *CA: A Cancer Journal for Clinicians*, 2023, 73(3): 233-254.
- [4] 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.
SUN F Y, WANG Q, (LÜ/LV/LU/LYU) Z W, *et al.* Review of application of deep learning in colon polyp segmentation[J]. *Computer Engineering and Applications*, 2023, 59(23): 15-27. (in Chinese)
- [5] FAN D P, JI G P, ZHOU T, *et al.* Pranet: Parallel Reverse Attention Network for Polyp Segmentation[C]. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer, 2020: 263-273.
- [6] TOMAR N K, JHA D, BAGCI U, *et al.* Tganet: Text-Guided Attention for Improved Polyp Segmentation[C]. WANG L, DOU Q, FLETCHER PT, *et al.* *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer, 2022: 151-160.
- [7] YUE G H, HAN W W, JIANG B, *et al.* Boundary constraint network with cross layer feature integration for polyp segmentation[J]. *IEEE Journal of Biomedical and Health Informatics*, 2022, 26(8): 4090-4099.
- [8] YIN Z J, LIANG K M, MA Z Y, *et al.* Duplex Contextual Relation Network for Polyp Segmentation[C]. 2022 *IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. Kolkata, India. IEEE, 2022: 1-5.
- [9] ZHOU T, ZHOU Y, HE K L, *et al.* Cross-level feature aggregation network for polyp segmentation [J]. *Pattern Recognition*, 2023, 140: 109555.
- [10] WANG J F, HUANG Q M, TANG F L, *et al.* Stepwise Feature Fusion: Local Guides Global [C]. WANG L, DOU Q, FLETCHER PT, *et al.* *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer, 2022: 110-120.
- [11] 周雪, 柏正尧, 陆倩杰, 等. 融合 Transformer 和轴向注意的结直肠息肉分割[J]. 计算机工程与应用, 2023, 59(11): 222-230.
ZHOU X, BAI Z Y, LU Q J, *et al.* Colorectal polyp segmentation combining pyramid vision transformer and axial attention[J]. *Computer Engineering and Applications*, 2023, 59(11): 222-230. (in Chinese)
- [12] 梁礼明, 何安军, 董信, 等. 融合 PVTv2 和多尺度边界聚合的结直肠息肉分割算法[J]. 计算机应用研究, 2023, 40(5): 1553-1558.
LIANG L M, HE A J, DONG X, *et al.* Colorectal polyp segmentation algorithm fusion PVTv2 and multiscale boundary aggregation[J]. *Application Research of Computers*, 2023, 40(5): 1553-1558. (in Chinese)
- [13] ZHANG Y D, LIU H Y, HU Q. Transfuse: fusing transformers and cnns for medical image seg-

- mentation[C]. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer, 2021: 14-24.
- [14] SANDERSON E, MATUSZEWSKI B J. FCN-Transformer feature fusion for polyp segmentation[C]. *Annual Conference on Medical Image Understanding and Analysis*. Cham: Springer, 2022: 892-907.
- [15] KIM T, LEE H, KIM D. UACANet: Uncertainty augmented context attention for polyp segmentation[C]. *Proceedings of the 29th ACM International Conference on Multimedia*. Virtual Event China. ACM, 2021: 2167-2175.
- [16] LI H, YANG L, MIAO J, *et al.* MCE-Net: polyp segmentation with multiple branch series-parallel attention and channel interaction via edge distribution guidance[J]. *Physics in Medicine & Biology*, 2023, 68(13): 135003.
- [17] 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.
- [18] TAN M, LE Q. Efficientnet: rethinking model scaling for convolutional neural networks[C]. *International conference on machine learning*. PMLR, 2019: 6105-6114.
- [19] LEE C Y, XIE S, GALLAGHER P, *et al.* Deeply-supervised nets[C]. *Artificial intelligence and statistics*. Pmlr, 2015: 562-570.
- [20] PENG Y J, ZHANG T, GUO Y F. CovTransNet: dual branch fusion network with transformer for COVID-19 infection segmentation[J]. *Biomedical Signal Processing and Control*, 2023, 80: 104366.
- [21] LIU Q H, LIN Y J, HAN X X, *et al.* Dual-branch feature extraction network combined with Transformer and CNN for polyp segmentation[J]. *International Journal of Imaging Systems and Technology*, 2024, 34(1): e22987
- [22] FU J, LIU J, TIAN H J, *et al.* Dual Attention Network For Scene Segmentation [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA. IEEE, 2019: 3141-3149.
- [23] XIA Y, YUN H J, LIU Y J, *et al.* MGCBFormer: the multiscale grid-prior and class-inter boundary-aware transformer for polyp segmentation[J]. *Computers in Biology and Medicine*, 2023, 167: 107600.
- [24] SRIVASTAVA A, CHANDA S, JHA D, *et al.* GMSRF-Net: an improved generalizability with global multi-scale residual fusion network for polyp segmentation[C]. 2022 *26th International Conference on Pattern Recognition (ICPR)*. Montreal, QC, Canada. IEEE, 2022: 4321-4327.
- [25] CHANG Q, AHMAD D, TOTH J, *et al.* ESF-PNet: efficient deep learning architecture for real-time lesion segmentation in autofluorescence bronchoscopic video[C]. *Medical Imaging 2023: Biomedical Applications in Molecular, Structural, and Functional Imaging*. February 19-24, 2023. San Diego, USA. SPIE, 2023.
- [26] CARION N, MASSA F, SYNNAEVE G, *et al.* End-To-End object detection with transformers [C]. VEDALDI A, BISCHOF H, BROX T, *et al.* *European Conference on Computer Vision*. Cham: Springer, 2020: 213-229.
- [27] 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.
- [28] JHA D, SMEDSRUD P H, RIEGLER M A, *et al.* Kvasir-seg: a segmented polyp dataset [C]. *MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5-8, 2020, Proceedings, Part II 26*. Springer International Publishing, 2020: 451-462.
- [29] 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.
- [30] TAJBAKHS N, GURUDU S R, LIANG J M. Automated polyp detection in colonoscopy videos using shape and context information [J]. *IEEE Transactions on Medical Imaging*, 2016, 35(2): 630-644.
- [31] WEI J, WANG S H, HUANG Q M. F³Net: fusion, feedback and focus for salient object detection [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 12321-12328.
- [32] RÖNNBERGER O, FISCHER P, BROX T.

- U-net: convolutional networks for biomedical image segmentation[C]. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer, 2015: 234-241.
- [33] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHS N, *et al.* UNet++: redesigning skip connections to exploit multiscale features in image segmentation[J]. *IEEE Transactions on Medical Imaging*, 2020, 39(6): 1856-1867.
- [34] GU Z W, CHENG J, FU H Z, *et al.* CE-net: context encoder network for 2D medical image segmentation[J]. *IEEE Transactions on Medical Imaging*, 2019, 38(10): 2281-2292.
- [35] WU H S, ZHAO Z B, WANG Z Z. META-unet: multi-scale efficient transformer attention unet for fast and high-accuracy polyp segmentation[J]. *IEEE Transactions on Automation Science and Engineering*, 2024, 21(3): 4117-4128.
- [36] ZHENG J W, YAN Y D, ZHAO L, *et al.* CG-MA-net: cross-level guidance and multi-scale aggregation network for polyp segmentation [J]. *IEEE Journal of Biomedical and Health Informatics*, 2024, 28(3): 1424-1435.
- [37] DUC N T, OANH N T, THUY N T, *et al.* ColonFormer: an efficient transformer based method for colon polyp segmentation[J]. *IEEE Access*, 2022, 10: 80575-80586.
- [38] JIANG X B, ZHU Y, LIU Y T, *et al.* MC-DC: an MLP-CNN based dual-path complementary network for medical image segmentation[J]. *Computer Methods and Programs in Biomedicine*, 2023, 242: 107846.
- [39] SONG P F, LI J J, FAN H, *et al.* TGDAUNet: transformer and GCNN based dual-branch attention UNet for medical image segmentation[J]. *Computers in Biology and Medicine*, 2023, 167: 107583.

作者简介:



李大湘(1974—),男,湖南怀化人,博士,硕士生导师,2005年、2011年于西北大学分别获得硕士和博士学位,主要从事遥感图像分类、目标检测与跟踪、医学图像识别、多实例学习、深度学习等方面的研究。E-mail: www_ldx@163.com

通讯作者:



李登辉(1999—),男,河南驻马店人,硕士研究生,2022年于塔里木大学获得学士学位,主要从事深度学习、医学图像分割的研究。E-mail: ldh_wy0908@163.com