

一种基于特征融合的息肉分割双解码模型

吴 港, 全海燕

(昆明理工大学信息工程与自动化学院, 昆明 650500)

摘要: 在结直肠癌的早期筛查中, 通过对结肠镜图像进行自动化的息肉检测和分割可以提高诊断效率和准确性。由于肠道内部环境的复杂性以及图像质量的限制, 自动化的息肉分割仍然是一个具有挑战性的问题。针对这一问题, 提出了一种基于Transformer和空洞卷积特征融合的息肉分割双解码模型(Dual decoded polyp segmentation model fusing Transformer and dilated convolution, FTDC-Net)。该模型以ResNet50作为编码器, 以便能够更好地提取图像深层次特征。使用Transformer编码模块, 它的自注意力(Self-attention)机制能够捕捉输入之间的长距离依赖关系, 模型中使用了不同的空洞卷积(Dilated-convolution)来扩大模型的感受野, 让模型能捕捉到结肠镜图像更大范围内的信息。本文网络模型的解码部分使用双解码结构, 包含一个自动编码器分支, 自动编码器可以重构输入, 另一个编码分支用于分割结果。模型中, 自动编码器的输出被用于生成一个注意力图作为注意力机制, 该图将被用于指导分割结果。在Kvasir-SEG和ETIS-LARIBPOLYPDB标准数据集上进行了实验验证, 实验结果表明FTDC-Net能有效地分割出结肠息肉, 相比目前主流息肉分割模型, 在各项评价指标上均取得了较高的提升。

关键词: 息肉分割; Transformer; 自注意力机制; ResNet50; 空洞卷积

中图分类号: TP391.41

文献标志码: A

A Double-Decoding Model for Polyp Segmentation Based on Feature Fusion

WU Gang, QUAN Haiyan

(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: In the early screening of colorectal cancer, diagnostic efficiency and accuracy can be improved by automated polyp detection and segmentation of colonoscopy images. Due to the complexity of internal environment of intestines and the limitation of image quality, automated polyp segmentation is still a challenging problem. Aiming at this problem, this paper proposes a dual-decoding model for polyp segmentation using Transformer and null convolution to achieve feature fusion (FTDC-Net). ResNet50 is used as an encoder in order to be able to better extract deep image features. The Transformer coding module is used, which has a self-attention mechanism to capture long distance dependencies between the inputs, and different dilated-convolutions are used in the model to expand the sensory field of the model to allow the model to capture a larger range of information in the colonoscopy image. The decoding part of the network model in this paper uses a dual-decoding structure, including an autoencoder branch that reconstructs the inputs and a coding branch for segmenting the results. The output of the autoencoder is

基金项目: 国家自然科学基金(61861023)。

收稿日期: 2023-08-28; 修订日期: 2024-01-27

used in the model to generate an attention map as an attention mechanism. This map will be used to guide the segmentation results. Experimental validation is carried out on the Kvasir-SEG and ETIS-LARIBPOLYPDB standard datasets, and experimental results show that FTDC-Net can effectively segment colon polyps, and achieves a high level of improvement in all evaluation metrics compared to the current mainstream polyp segmentation models.

Key words: polyp segmentation; Transformer; self-attention; ResNet50; dilated-convolution

引 言

在癌症早期预防的诸多领域中,结肠息肉的检测和分割显得尤为关键,因为结肠息肉是结直肠癌的重要前驱病变,及时发现和切除息肉可以显著降低患者结直肠癌的风险^[1]。结肠镜检查是最常见的检测方法,医生通过结肠镜可以直接观察到结肠内部,发现并切出息肉。然而,这种方法存在一些限制,如需要专业的操作技术,耗时较长,病人可能会有不适,而且可能会漏检小息肉或位置难以察觉的息肉。计算机辅助检测(Computer-aided detection, CAD)系统能够在结肠镜视频中实时分割出息肉的位置,为内窥镜医生提供决策支持^[2],这有助于降低息肉漏诊或误诊的可能性。然而,对于结肠息肉的准确分割仍然是一项具有挑战性的任务。一方面,结肠息肉的形状和大小差异大,颜色与周围组织相似,使得识别和定位困难;另一方面,由于结肠镜图像中的运动模糊、照明不均等因素,图像质量差异也为精确分割增加了复杂性。

传统的息肉分割方法主要是通过手动提取图像特征进行操作^[3-4]。例如,阈值分割需要预先确定亮度或颜色阈值,边缘检测则需要人工设定用于识别边缘的特定模式或纹理,而形态学分割则依赖于特定的形状和结构参数。如 Gupta 等^[5]采用一种基于边缘检测的分割方法,该方法通过找到一个初始点,将周围像素加入到同一区域的方式来完成分割,但是这种方法对初始点的选择非常敏感,并且可能会由于噪声或者非均匀照明导致过分割或者欠分割。Vala 等^[6]提出了一种基于阈值的分割方法,该方法根据像素强度来划分图像,简单快速,但在处理复杂图像时,特别是在光照不均或存在伪影的情况下,可能无法准确分割息肉。李梦等^[7]提出了基于局部熵的区域活动轮廓模型并将其应用于图像分割,这类方法对图像中的光照变化和噪声可能较为敏感,如果图像受到不同光照条件或包含噪声,模型的性能可能会下降,因为局部熵的计算可能受到这些因素的干扰。以上这些手动特征提取方法有两个主要的缺点:(1)这些方法需要领域专家具备深入的知识 and 理解,以便选择和提取合适的特征,这使得其应用具有一定的复杂性和专业性;(2)这些方法通常对图像的特定属性非常敏感,例如光照条件、视角和噪声等,这些因素会对手动提取的特征产生影响,从而影响分割的准确性。传统的图像分割方法在某些情况下能够取得良好的效果,但是在处理复杂和变化的图像条件,特别是在进行结肠息肉分割的任务时,其表现往往不足。这就需要寻找新的方法,以提高息肉分割的准确性和可靠性。

近年来,随着深度学习的快速发展,尤其是卷积神经网络(Convolutional neural networks, CNN)^[8]在图像处理领域的广泛应用,结肠息肉分割的研究和应用已经取得了显著的进步。深度学习方法,特别是基于深度卷积神经网络的分割方法,能够自动学习并提取图像的深层特征,而无需手动选择和提取特征,这极大地减少了人工干预的需求。此外,深度学习方法对图像的噪声、光照变化和视角等因素的适应性更强,分割的准确性和稳定性也得到了显著提高。目前,已经有许多基于深度学习的结肠息肉分割研究和应用报告^[9]。

Ronneberger 等的研究^[10]引入了 U-Net 网络,这是一种具有完全对称的编码器-解码器结构,因其在生物医学图像分割中的成功应用,许多研究者开始采用 U-Net 及其变体进行息肉分割。例如, Zhang

等^[11]发表了一种深度残差结构的U型网络,名为ResUNet,该网络将残差连接纳入U-Net的每个卷积模块,以提取更深层次的图像特征,并产生更精确的分割结果。Zhou等^[12]提出了一种称为UNet++的网络,它通过减小未知网络深度、重新设计跳跃连接,并设计了一种网络剪枝方案以增强UNet++的性能。Fan等^[13]设计了一种名为平行反向注意力网络(PraNet)的模型,用于精确分割息肉。Chen等^[14]提出了基于Transformer的医学图像分割框架TransUnet。Zhang等^[15]将CNN与Transformer以并行方式结合在一起,名为Transfuse。另外,Jha等^[16]提出了一种名为DoubleUNet的网络,这种网络通过级联两个变体U-Net结构,使得整个网络具有更强的特征提取能力和更大的感受野,并在网络中加入了SE(Squeeze-and-excite)和ASPP(Atrous spatial pyramid pooling)等附加模块,以获得边界更清晰的息肉分割结果。

在医学图像中,息肉可能具有不同的尺寸和形状,从微小的息肉到较大的息肉都可能存在。上述U-Net及其变体模型在处理不同尺寸和形状的息肉时存在几个问题。首先,由于其编码-解码结构采用固定尺寸的卷积和池化操作,可能导致分辨率不足的情况,特别是对于较小的息肉,模型的感受野有限,很容易忽略细小的细节信息,从而影响分割准确性。其次,编码器和解码器部分使用相同尺寸的卷积核,感受野固定,导致模型对不同尺寸的息肉处理不够灵活,难以适应不同尺度下的息肉。此外,它们模型中的卷积操作主要关注局部特征的提取,对于不同尺寸和形状的息肉的全局关联性理解能力较弱。这可能导致模型在分割过程中无法准确捕捉息肉与周围组织的边界和内部细节。最后,在特征提取和融合过程中缺乏适应性,无法根据不同尺寸的息肉自动调整特征的权重和重要性。这导致模型对不同尺寸的息肉处理效果不一致,难以实现精准的分割。因此,如何设计网络模型以提取更强大的多尺度上下文信息,尤其是在处理复杂的息肉分割任务时,仍然是图像分割研究领域的一大挑战。

针对以上问题,本文设计了一种深度网络模型FTDC-Net(Dual decoded polyp segmentation model fusing Transformer and dilated convolution),以ResNet50提取特征并引入Transformer编码器用于特征融合以充分利用编码器提取多尺度特征信息,采用多尺度的空洞卷积(Dilated-convolution)增加感受野,使模型能够更好地处理不同尺寸和形状的息肉,并结合自编码器和注意力机制进行特征融合和选择。本文模型不仅可以获取全局和局部的上下文信息,而且可以对特征进行有效的选择和重构,从而提高了分割的精度和稳定性。

1 FTDC-Net模型

1.1 整体框架

本文所提出的FTDC-Net模型总体架构如图1所示,它包括前向网络、2个不同的解码器和1个注意力映射模块。首先使用ResNet50作为前向网络,提取特征并生成有细节和深度的特征图。ResNet50的4个层分别生成了4个不同尺度的特征图,然后利用这些特征图进一步操作。

在跨尺度融合部分,本文将从ResNet50得到的最深层特征经过一个瓶颈(Bottleneck)层模块和一个空洞卷积(Dilated-convolution)模块处理,瓶颈层提取的是高级语义特征,帮助模型理解图像中的抽象模式和对象,空洞卷积模块接收来自ResNet50的具有较大感受野的特征,在更大的感受野范围内捕获上下文信息,有助于理解图像中的全局结构,然后将这两部分的输出进行拼接。这一操作不仅融合了两种不同类型的特征,还提高了特征的复杂性和丰富性,这两者的结合有助于模型更好地理解图像的语义和结构信息。此后,这一融合特征被送入两个解码器中完成输入图像的分割和重构。

在解码部分,本文设计了两个解码器,一个用于自编码任务,另一个用于分割任务。在自编码任务中,自编码器分支通过 1×1 卷积和Sigmoid激活函数生成注意力图,这些注意力图与分割解码器块的输出相乘作为下一个分割解码器块的输入,自编码器最终输出输入图像重建的灰度图。在分割解码器

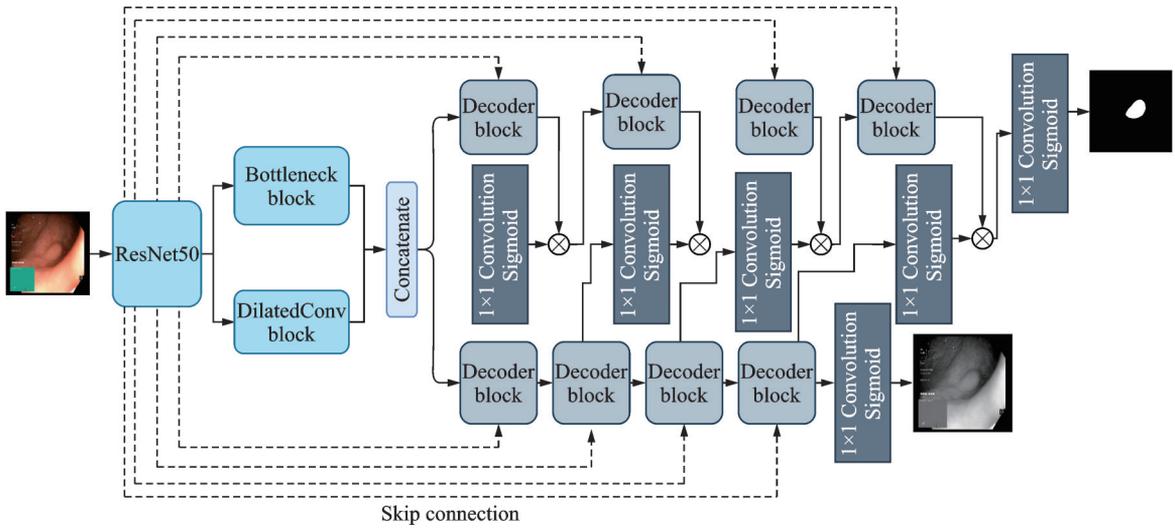


图1 FTDC-Net框架

Fig.1 FTDC-Net framework

中,使用4个解码模块(Decoder block)逐渐上采样并恢复图像的大小。在每一步,都利用了对应尺度的ResNet50特征图作为跳跃连接,与当前层的特征进行拼接,以帮助恢复图像的细节。

1.2 损失函数

FTDC-Net模型的损失函数可用二值交叉熵(Binary cross entropy, BCE)损失函数^[17]和Dice损失函数的结合表示为

$$\begin{cases} l_{BCE} = -[y_c \log P_c + (1 - y_c) \log (1 - P_c)] \\ l_{Dice} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \\ l = \lambda l_{BCE} + (1 - \lambda) l_{Dice} \end{cases} \quad (1)$$

式中: y_c 为输入图像的二值化像素值; p_c 为自编码器输出的概率值。在自编码任务中完成重建输入图像,而二进制交叉熵损失 l_{BCE} 则用于衡量模型的输出与输入之间的差异,模型将输入图像传递到编码器,然后将编码器的输出传递到解码器,最终产生一个重建图像。二进制交叉熵损失函数通过比较重建图像和原始输入图像之间的像素级别差异,来指导模型优化学习过程。

在分割任务中使用 l_{BCE} 和 l_{Dice} 损失的加权和 l ,可以综合考虑像素级别的分类准确性(通过 l_{BCE})和分割结果的相似性(通过 l_{Dice})。 l_{BCE} 在训练初期具有较好的稳定性,有助于加快模型的收敛速度,而 l_{Dice} 则更关注于像素级别的相似性,可以促使模型生成更平滑和连续的分割结果。其中 λ 表示权值,在实验中取值0.5, X 、 y_c 代表真实分割图像的像素标签, Y 、 p_c 表示模型的分割结果。 $|X \cap Y|$ 近似为预测图像的像素与真实标签图像像素之间的点乘,并将点乘结果相加。 $|X|$ 和 $|Y|$ 分别近似为它们各自对应图像中的像素相加。

1.3 瓶颈层模块

U-Net中的传统卷积结构在处理长距离依赖关系时效果有限,特别是对于大范围的上下文信息。使用瓶颈层模块引入Transformer结构,通过自注意力机制,能够更好地捕捉图像中不同区域之间的关联。通过Transformer编码层,瓶颈层模块实现了对输入特征图的全局信息建模,每个编码层内部包含

多个注意力头,允许模型关注输入中不同位置的内容。

瓶颈层模块在网络中通常被设计为具有较小通道数的中间层,即输入通道数和输出通道数较多的层通过瓶颈层的中间层进行维度压缩和扩展。这种设计有助于减少模型的计算复杂度和参数量,同时保持较高的表示能力。

在本文模型中,瓶颈层模块包含几个顺序操作,如图2所示。首先,使用 1×1 卷积层来降低输入特征图的维度,有效减少了输入通道的数量,这种维度的降低有助于减少计算复杂度和模型参数,同时保持关键特征。随着维度的降低,引入Transformer编码器层,用于捕捉特征之间的长距离依赖关系并建模特征之间的相互依赖。Transformer编码器层包含自注意力机制^[18],允许特征图中的每个位置与其他位置进行关注,有效捕捉全局的上下文信息。

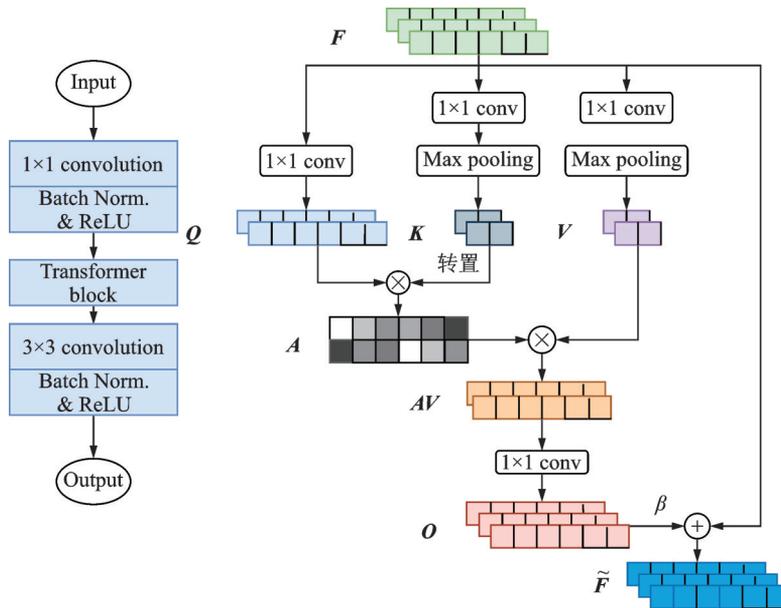


图2 瓶颈层模块

Fig.2 Bottleneck layer module

Transformer编码器层中的多头注意力由多个自注意力组成,图2展示了自注意力机制的结构示意图,包括输入特征图 F 、查询矩阵 Q 、键矩阵 K 、值矩阵 V 、注意力矩阵 A 、加权和矩阵 O 以及输出特征图 \tilde{F} 。该机制通过计算序列中每个元素与其他元素之间的相对关联度,为每个元素生成一个上下文相关的表示。自注意力机制由3个主要组件组成:查询(Query)、键(Key)和值(Value)。对于每个位置上的元素,通过线性变换得到对应的查询、键和值向量。然后,通过计算查询向量与所有位置的键向量之间的相似性得分,并经过softmax归一化,获得每个位置与其他位置的注意权重。最终,将这些权重应用于值向量,加权求和后得到当前位置的表示,具体可表示为

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{2}$$

式中 d 为 K 矩阵的列数,即向量维度。

通过利用Transformer编码器层的能力,瓶颈层模块可以有效地捕捉特征之间的复杂关系,并增强所提取特征的表示判别能力。最后,使用 3×3 卷积层来恢复特征图的空间维度,同时保持压缩的表示,

瓶颈层模块输出的特征图随后被送入后续层进行进一步处理和解码。

瓶颈层模块的引入显著提高了整体性能,通过维度的降低和相互依赖建模来提取和优化判别特征。通过压缩特征图并引入 Transformer 编码器层,瓶颈层模块确保了高效的信息流动,并促进了高级表示的学习,从而为成功分割息肉做出贡献。

1.4 空洞卷积模块

由于 FCN 在医学图像分割中通过池化增大感受野缩小图像尺寸,然后通过上采样还原图像尺寸,但是在这个过程中造成了精度的损失,为了减小这种损失就得去掉池化层,然而这样就会导致特征图感受野太小。为了解决这一问题, Yu 等^[19]提出了在卷积核内部设定扩张率 r 的空洞卷积,图 3 为 $r=2$ 时的空洞卷积示意图。本文引入了空洞卷积模块,模块采用了空洞卷积方式,旨在扩展感受野并增强特征的上下文信息。

由于卷积操作的局部性,以 U-Net 为基础的网络架构缺乏感受野的范围。对比普通卷积,本文采用空洞卷积模块,空洞卷积在保持计算效率的同时增加了感受野的范围。使用不同的扩张率,每个卷积核关注不同范围的像素,从而丰富了模型对于图像层次结构的理解,实现多尺度的感受野,有助于捕获图像中不同尺度的信息。

空洞卷积模块由一系列的卷积操作组成。首先,通过使用不同的扩张率(Dilation rate)配置多个空洞卷积层,每个空洞卷积层具有不同的感受野大小。这种配置允许每个空洞卷积层通过扩大感受野,从而更好地捕捉图像中的局部和全局上下文信息,空洞卷积感受野可表示为

$$r_n = r_{n-1} + (k - 1) \prod_{i=1}^{n-1} s_i \tag{3}$$

式中: r_n 表示本层感受野; s_i 表示第 i 层卷积的步长; k 表示卷积核大小。

在具体的实现中,空洞卷积模块包括 4 个空洞卷积层,每个层的扩张率逐渐增加。通过使用不同的扩张率,空洞卷积模块能够在不增加模型参数的情况下显著增加感受野,以获得更大范围的上下文信息。在每个空洞卷积层之后,采用了一个 1×1 的卷积层进行特征融合。通过将 4 个空洞卷积层的输出特征在通道维度上进行拼接,然后应用 1×1 卷积操作,将通道数压缩到与输出通道数相匹配。这种特征融合策略能够有效地整合多个不同感受野的特征,提高特征的表达能力和判别性。

通过引入空洞卷积模块,本文的模型能够在图像分割任务中更好地捕捉特征的上下文信息。通过扩展感受野和特征融合,空洞卷积模块能够提供更全面的视觉信息,并改善模型对图像细节和全局结构的理解能力。因此,空洞卷积模块在提高息肉分割性能方面具有重要的意义,并在本文模型中发挥着关键的作用。

1.5 残差模块

如图 4 所示,本文的残差模块由两个卷积层组成,每个卷积层后面都有 1 个批归一化层和 ReLU 激活函数。第 1 个卷积层采用 3×3 的卷积核对输入特征进行卷积操作;第 2 个卷积层则进一步对输出特

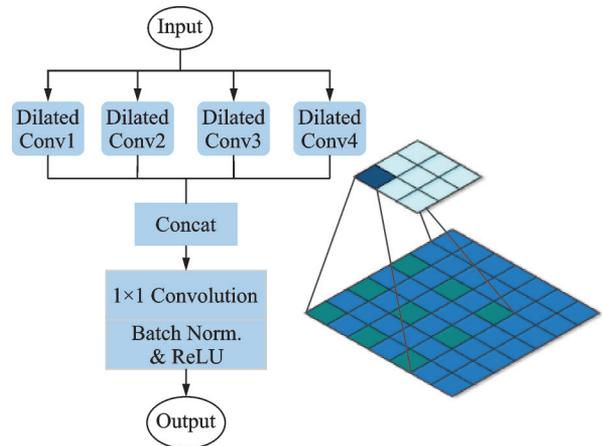


图 3 空洞卷积模块

Fig.3 DilatedConv module

征进行卷积。残差块的输入特征通过捷径连接(Shortcut connection)直接与残差块的输出特征相加,然后通过ReLU激活函数进行非线性变换。这种残差结构允许输入特征通过主路径和捷径路径进行信息传递。通过残差连接,模型可以学习到输入特征的变化部分,而无需关注完全重建输入特征的过程。这样的设计有助于解决梯度消失问题,使模型能够更深层地进行训练,并提升模型的表达能力。

因为本文模型较U-Net更复杂,为了缓解深度网络的梯度消失问题,通过在编码器和解码器中引入多个残差块,可以更好地提取低频信息、缓解语义信息缺失问题和梯度消失问题,从而提高图像分割的准确率。在本文模型中,残差结构被应用于编码器和解码器。这些残差结构的堆叠和组合使得模型能够更好地捕捉图像特征的细节和上下文信息,从而提高模型的性能和准确度。具体被应用于两个关键部分:瓶颈(Bottleneck)层模块和解码模块。在瓶颈层模块中,残差块通过残差连接将输入特征和卷积层的输出特征相加,以增强编码器的表达能力和信息传递效果。而在解码模块中,残差块接收上一层解码器的输出特征并与上采样后的特征进行连接,通过残差连接促进特征融合,有助于恢复细节信息和提高重建图像的质量。这种应用方式使得模型能够更好地传递和融合特征,提高模型的性能和重建效果。

1.6 解码模块

如图5所示,解码模块由1个上采样层和两个残差块组成。在解码器块之间,有1个上采样层用于将特征图的尺寸放大两倍。上采样后的特征图与来自编码器的跳跃连接(Skip connection)特征图进行拼接,以便融合更多的上下文信息。解码器块的目标是逐渐恢复图像的细节和空间分辨率。本文模型是1个双解码器结构,1个解码器作为自编码器(Autoencoder)部分,自编码器任务是通过编码和解码的过程重建输入图像,有助于学习输入图像中的有用特征,并在训练过程中减小输入图像与重构图像之间的差异,从而提高模型学习到的特征的代表能力。自编码任务可以帮助模型学习图像中的结构和纹理信息,从而提高图像分割的性能。通过自编码得到的注意力图与分割解码的输出相乘的过程可以看作是一种引入注意力机制,在这个过程中,自编码任务的学习过程使得模型能够关注输入图像中的重要区域。这些区域对于图像的重构是关键,将这种自编码任务学到的注意力模式与分割任务相结合,可以使得分割模型更加集中于输入图像中感兴趣的区域,从而提高分割的准确性。在本文模型中,自编码器的输出通过卷积层进行处理,生成一个掩码(Mask),将这个掩码和特征图进行元素级别的乘法作为一个注意力机制,掩码的值越大,对应的特征图的位置就越重要,掩码的值越小,对应的特征图的位置就越不重要,它与解码器共享相同的结构,但目标是通过重建输入图像来学习特征表示来辅助分割任务。

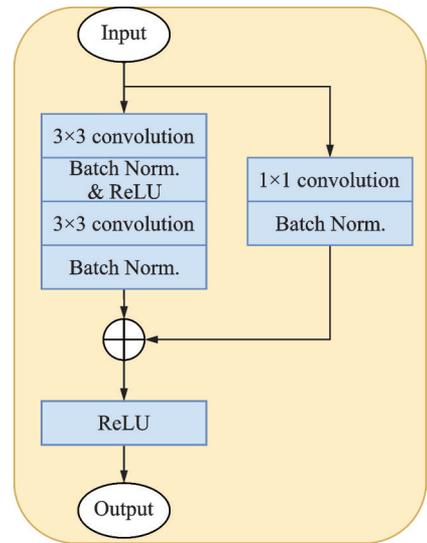


图4 残差模块

Fig.4 Residual module

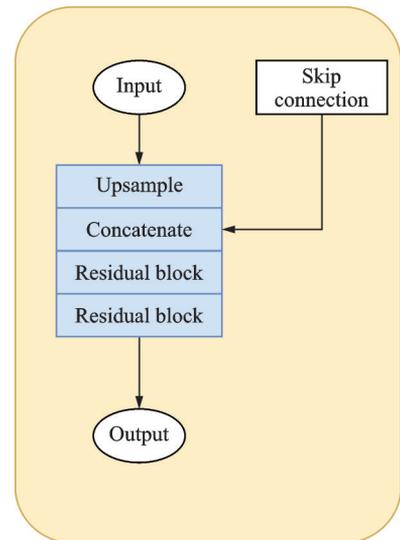


图5 解码器块

Fig.5 Decoder block

2 实验结果与分析

2.1 数据预处理

使用两个公开的息肉数据集来评估所提出的方法:Kvasir-SEG^[20]、ETIS-LARIBPOLYPDB^[21]数据集。Kvasir-SEG数据集由1000幅分辨率从332像素×487像素到1920像素×1072像素不等的息肉图像和对应的标签图像组成,ETIS-LARIBPOLYPDB数据集包含196张分辨率为1255像素×996像素的息肉图片和它所对应的标签。

为了增加模型的鲁棒性和泛化能力,先对原始数据集进行预处理,将原始图像数据进行数据增强,通过对原始图像进行旋转、裁剪、翻转、缩放、亮度调整和噪声添加等操作,生成具有不同视角、光照条件等特征的图像,生成更多的训练样本,可以有效地解决数据不平衡、过拟合和欠拟合等问题。通过数据增强,可以增加训练数据的数量和多样性,提高模型在实际场景中的表现和准确率。

将预处理后的图像输入模型,训练的批次大小为16,使用Adam优化器^[22],初始学习率设置0.0001,使用学习率衰减策略来调整。整个模型的训练和测试是在RTX 3090上使用PyTorch^[23]实现的,迭代100次。

2.2 评价指标

本文用来评估所提出方法的指标包括DSC分数^[24]、交并比(Intersection over union, IoU)^[25]、召回率(Recall)和精确率(Precision)^[26]。如果分割效果越好,这些指标数值就越高。DSC分数是精确率和召回率的调和平均值,综合衡量了模型的准确性和召回性能;IoU用于衡量预测掩码和真实掩码的重叠程度。DSC指数是计算预测掩码和真实掩码的交集面积除以它们的并集面积;Recall衡量模型能够正确检测正样本(目标区域)的能力,即模型成功找到的目标区域的比例;Precision衡量模型在预测为正样本的样本中的准确性,即正确预测为正样本的比例。DSC、IoU、Recall和Precision的计算公式分别为

$$\text{DSC} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (4)$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (5)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

式中:TP(True positive)表示预测为目标区域且与真实区域重叠的像素数;FP(False positive)表示预测为目标区域但与真实区域不重叠的像素数;FN(False negative)表示预测为背景区域但与真实区域重叠的像素数。

2.3 实验结果

使用4个评价指标(DSC、IoU、Precision和Recall)在Kvasir-SEG数据集和ETIS-LARIBPOLYPDB数据集上评估所提出的模型。为了验证所提出模型的有效性,使用这两个数据集与经典和先进的6种算法做实验并进行对比分析,包括U-Net^[10]、ResUNet^[11]、UperNet^[27]、SegResNet^[28]、Deeplabv3+^[29]和DDANet^[30]算法。表1和表2分别显示了所提出的方法在Kvasir-SEG、ETIS-LARIBPOLYPDB数据集上和各个算法的总体性能。如表1所示,在Kvasir-SEG数据集中,FTDC-Net的DSC、IoU、Precision和Recall分别达到了87.07%、80.55%、87.71%和90.82%。相比于对比模型,本文所提出的模型在

DSC、IoU、和 Recall 方面取得了最好的性能,但 Precision 指标欠佳,这表明所提出的模型具有更好的分割性能。与比较先进的 DDANet 相比,本文方法在 DSC、IoU、Precision、Recall 分别获得了 1.31%、2.55%、1.28%、2.02% 的性能提升。如表 2 所示,在 ETIS-LARIBPOLYPDB 数据集中,本文方法的 DSC、IoU、Precision 和 Recall 分别达到 82.77%、76.95%、85.73% 和 90.32%。相比于其他模型,本文所提出的模型在 DSC、IoU、和 Recall 方面取得了最好的性能,这表明它具有更好的分割性能。与比较先进的 DDANet 相比,本文方法在 DSC、IoU、Precision 和 Recall 分别获得了 2.01%、2.92%、2.1% 和 3.49% 的性能提升,其中本文模型和其他模型的参数量的大小如表 2 中 Params 所示。

表 1 本文方法与各模型在 Kvasir-SEG 数据集上的性能比较

Method	DSC	IoU	Precision	Recall
U-Net ^[10]	56.17	43.94	61.62	62.96
ResUNet ^[11]	69.87	57.76	77.12	69.83
UperNet ^[27]	70.70	64.83	81.96	70.61
SegResNet ^[28]	83.82	57.15	83.13	70.32
Deeplabv3++ ^[29]	69.04	75.81	91.25	83.78
DDANet ^[30]	85.76	78.00	86.43	88.80
MKDCNet ^[31]	79.70	73.62	85.07	81.15
TGANet ^[32]	85.23	79.31	89.55	88.23
FTDC-Net (本文方法)	87.07	80.55	87.71	90.82

表 2 本文方法与各模型在 ETIS-LARIBPOLYPDB 数据集上的性能比较

Table 2 Performance comparison of the proposed method and other models on the ETIS-LARIBPOLYPDB dataset

Method	DSC/%	IoU/%	Precision/%	Recall/%	Params/10 ⁶
U-Net ^[10]	50.11	39.34	88.12	60.36	26.36
ResUNet ^[11]	60.32	47.83	86.42	63.54	30.00
UperNet ^[27]	64.83	53.26	85.32	66.23	126.07
SegResNet ^[28]	78.88	52.45	83.73	69.62	53.55
Deeplabv3++ ^[29]	65.09	70.11	89.75	82.58	40.00
DDANet ^[30]	80.76	74.03	83.63	86.83	6.84
MKDCNet ^[31]	76.02	63.29	83.66	83.58	19.84
TGANet ^[32]	77.49	72.38	86.65	79.67	19.84
FTDC-Net(本文方法)	82.77	76.95	85.73	90.32	33.46

本文使用 Kvasir-SEG 数据集对多个模型进行了分割实验,并对实验结果(表 1)进行了比较。从图 6(a)中的分割结果可以观察到,与其他各个模型相比,本文模型在息肉图像边缘分割上具有明显的优势,在捕捉图像边缘和细节方面表现更加出色,在处理复杂结构和纹理丰富的区域时表现更加出色,分割结果更加准确和清晰。在更具挑战难度的 ETIS-LARIBPOLYPDB 数据集上的分割结果如图 6(b)所示,与其他模型相比,本文的模型在处理小尺寸目标时也表现出较好的能力,能够更好地捕捉小尺寸目标的特征,从而实现更精确的分割。在具有复杂结构和纹理的区域,其他模型可能存在分割模糊或者错漏的问题,而本文方法能够更好地保留细节信息并进行准确的分割。

综上所述,本文模型在 2 个数据集上各项评价指标均有所提高,分割结果图更加精准,表明该方法能有效地分割出结肠息肉,更具竞争力。

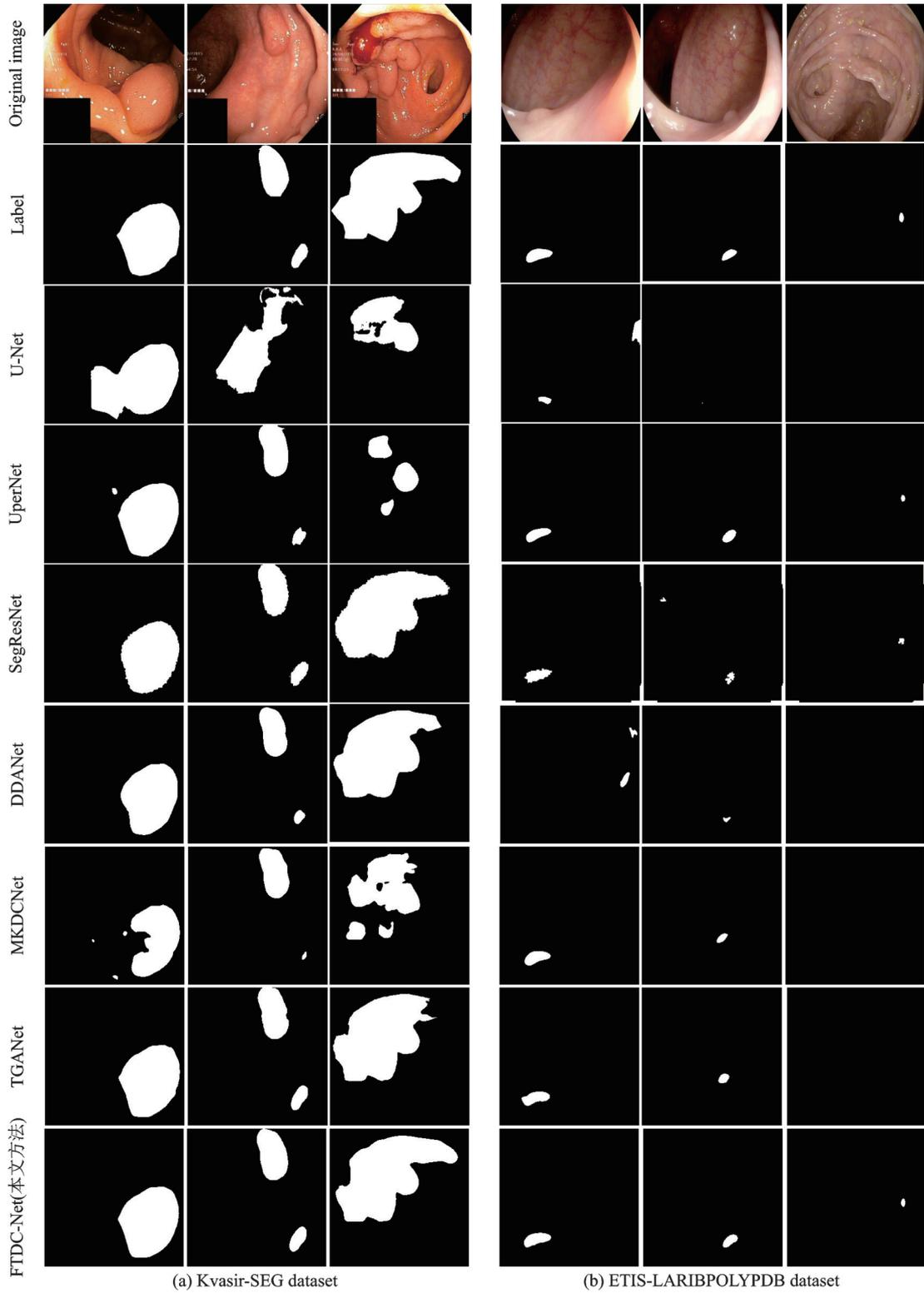


图 6 不同方法在 Kvasir-SEG 和 ETIS-LARIBPOLYPDB 数据集上的分割结果

Fig.6 Segmentation results of different methods on Kvasir-SEG and ETIS-LARIBPOLYPDB datasets

2.4 消融实验

为了验证本文方法模块的有效性和必要性,分别对Kvasir-SEG和ETIS-LARIBPOLYPDB数据集进行广泛的消融实验,对提出的FTDC-Net模型的各个模块进行深入分析。在本文中,将分别针对FTDC-Net模型的Transformer编码器块、空洞卷积模块、自编码器注意力机制进行消融实验。通过逐个添加这些模块并比较它们在性能上的影响,探索每个组件对模型性能的贡献,基准网络为以ResNet50前4层作为编码提取特征,后续通过解码器进行分割,即将ResNet作为Encoder替换U-Net原始结构(ResUNet)。TF为Transformer编码器块,DC为Dilated-convolution模块,DA为自编码器注意力机制。

消融实验结果如表3,4所示。从表3,4可以看出,在引入Transformer编码器块(+TF)、Dilated convolution模块(+DC)以及自编码器注意力机制(+DA)后,网络的性能得到了显著的提升。Transformer编码器块在息肉分割任务中具有积极影响,它采用自注意力机制,可以捕捉输入序列中的长距离依赖关系,有助于模型更好地理解息肉图像中不同位置之间的语义关联。这对于像息肉这样的结构性特征而言非常重要,因为息肉可能具有不同尺寸和形状,Transformer能够更好地处理这种变化,从而提高分割性能。Dilated-convolution模块也发挥着重要的作用。它通过增加卷积核的感受野,在不增加参数的情况下扩大了感受野的范围。这对于分割任务非常有用,因为息肉可能具有不同的大小,使用Dilated-convolution模块可以更好地捕捉不同尺度的结构信息,从而提高模型的分割准确性。自编码器注意力机制能够根据任务需求,学习图像的注意力热图,使得模型在关键区域更加集中。对于息肉分割任务而言,自编码器注意力机制可以帮助模型更关注息肉区域,并动态地调整注意力,以更准确地进行分割。这种注意力机制可以提高模型对息肉特征的感知能力,从而进一步提升分割性能。

从上述实验可以得出以下结论:在息肉分割任务中,Transformer编码器块、Dilated-convolution模块和自编码器注意力机制各自发挥着关键的作用。Transformer编码器块帮助模型处理不同尺度和形状的息肉,Dilated-convolution模块扩大感受野以捕获结构信息,自编码器注意力机制则调整注意力并增强对息肉区域的关注。因此,引入这些模块可以显著提高模型对息肉的分割性能。

3 结束语

为了解决由于息肉具有不同尺寸和形状而导致U-Net及其变体分割方法性能不佳的问题,本文提出了一种基于特征融合的息肉分割双解码模型。在传统的编码解码结构上,以ResNet50作为骨干网络,引入Transformer编码器用于特征融合以充分利用编码器提取的多尺度特征信息,采用多尺度的空洞卷积,增加感受野,使模型能够更好地处理大尺度目标,并在解码器中进行多分支的特征融合。同时,本文引入注意力机制来动态分配特征之间的相关权重,从而进一步提高分割性能。最终,该模型可以同时生成图像分割结果和图像重建结果。为了证明所提方法的有效性,在Kvasir-SEG和ETIS-

表3 所提模块在Kvasir-SEG数据集上的性能

Table 3 Performance of the proposed module on the Kvasir-SEG dataset %

Method	DSC	IoU	Precision	Recall
基准网络	79.41	39.34	81.09	83.46
+TF	84.32	78.11	84.28	88.59
+DC	86.33	79.26	86.92	89.73
+DA	87.07	80.55	87.71	90.82

表4 所提模块在ETIS-LARIBPOLYPDB数据集上的性能

Table 4 Performance of the proposed module on the ETIS-LARIBPOLYPDB dataset %

Method	DSC	IoU	Precision	Recall
基准网络	74.11	69.34	81.12	82.36
+TF	79.32	74.83	83.42	87.54
+DC	81.83	75.26	85.32	89.23
+DA	82.77	76.95	85.73	90.32

LARIBPOLYPDB数据集上做了大量的实验。在实验中,对该模型进行了广泛的评估,并与其他流行的医学图像分割模型进行比较。实验结果表明,本文模型在各项指标上均有显著优势。未来的工作可以通过进一步优化模型结构和参数来改进本文提出的模型,探索更多的方法,以进一步提升分割性能,提升医学图像分割的应用范围。

参考文献:

- [1] MAHMUD T, PAUL B, FATTAH S A. PolypSegNet: A modified encoder-decoder architecture for automated polyp segmentation from colonoscopy images[J]. *Computers in Biology and Medicine*, 2021, 128: 104119.
- [2] MORI Y, KUDO S. Detecting colorectal polyps via machine learning[J]. *Nature Biomedical Engineering*, 2018, 2(10): 713-714.
- [3] MAMONOV A V, FIGUEIREDO I N, FIGUEIREDO P N, et al. Automated polyp detection in colon capsule endoscopy[J]. *IEEE Transactions on Medical Imaging*, 2014, 33(7): 1488-1502.
- [4] TAJBAKSH 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.
- [5] GUPTA S, MAZUMDAR S G. Sobel edge detection algorithm[J]. *International Journal of Computer Science and Management Research*, 2013, 2(2): 1578-1583.
- [6] VALA H J, BAXI A. A review on Otsu image segmentation algorithm[J]. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 2013, 2(2): 387-389.
- [7] 李梦,詹毅,王艳.基于局部熵的区域活动轮廓图像分割模型[J].*数据采集与处理*,2023,38(3): 586-597.
LI Meng, ZHAN Yi, WANG Yan. Regional active contour image segmentation model based on local entropy[J]. *Journal of Data Acquisition and Processing*, 2023, 38(3): 586-597.
- [8] GU J, WANG Z, KUEN J, et al. Recent advances in convolutional neural networks[J]. *Pattern Recognition*, 2018, 77: 354-377.
- [9] YU L, CHEN H, DOU Q, et al. Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos[J]. *IEEE Journal of Biomedical and Health Informatics*, 2016, 21(1): 65-75.
- [10] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//*Proceedings of Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference*. Munich, Germany: Springer International Publishing, 2015: 234-241.
- [11] ZHANG Z, LIU Q, WANG Y. Road extraction by deep residual U-Net[J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(5): 749-753.
- [12] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKSH N, et al. Unet++: A nested U-net architecture for medical image segmentation[C]//*Proceedings of Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Granada, Spain: Springer International Publishing, 2018: 3-11.
- [13] FAN D P, JI G P, ZHOU T, et al. PraNet: Parallel reverse attention network for polyp segmentation[C]//*Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer International Publishing, 2020: 263-273.
- [14] CHEN J, LU Y, YU Q, et al. TransuNet: Transformers make strong encoders for medical image segmentation[EB/OL]. (2021-02-08). <https://doi.org/10.48550/arXiv.2102.04306>.
- [15] ZHANG Y, LIU H, HU Q. Transfuse: Fusing transformers and CNNs for medical image segmentation[C]//*Proceedings of Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference*. Strasbourg, France: Springer International Publishing, 2021: 14-24.
- [16] JHA D, RIEGLER M A, JOHANSEN D, et al. DoubleU-Net: A deep convolutional neural network for medical image segmentation[C]//*Proceedings of 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*. [S.l.]: IEEE, 2020: 558-564.
- [17] CRESWELL A, ARULKUMARAN K, BHARATH A A. On denoising autoencoders trained to minimise binary cross-

entropy[EB/OL]. (2017-08-28). <https://doi.org/10.48550/arXiv.1708.08487>.

- [18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of Advances in Neural Information Processing Systems. [S.l.]: [s.n.], 2017.
- [19] YU F, KOLTUN V, FUNKHOUSER T. Dilated residual networks[J]. IEEE Computer Society, 2017. DOI:10.1109/CVPR.2017.75.
- [20] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Kvasir-SEG: A segmented polyp dataset[C]//Proceedings of MultiMedia Modeling: 26th International Conference, MMM 2020. Daejeon, South Korea: Springer International Publishing, 2020: 451-462.
- [21] 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(2): 283-293.
- [22] KINGMA D P, BA J. Adam: A method for stochastic optimization[EB/OL]. (2014-12-22). <https://doi.org/10.48550/arXiv.1412.6980>.
- [23] PASZKE A, GROSS S, CHINTALA S, et al. Automatic differentiation in PyTorch[C]//Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS). Long Beach, CA, USA: [s.n.], 2017.
- [24] YAO A D, CHENG D L, PAN I, et al. Deep learning in neuroradiology: A systematic review of current algorithms and approaches for the new wave of imaging technology[J]. Radiology: Artificial Intelligence, 2020, 2(2): e190026.
- [25] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2019: 658-666.
- [26] GRAU J, GROSSE I, KEILWAGEN J. PRROC: Computing and visualizing precision-recall and receiver operating characteristic curves in R[J]. Bioinformatics, 2015, 31(15): 2595-2597.
- [27] XIAO T, LIU Y, ZHOU B, et al. Unified perceptual parsing for scene understanding[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: Springer, 2018: 418-434.
- [28] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [29] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.
- [30] TOMAR N K, JHA D, ALI S, et al. DDANet: Dual decoder attention network for automatic polyp segmentation[C]//Proceedings of International Conference on Pattern Recognition Workshops. [S.l.]: Springer International Publishing, 2021: 307-314.
- [31] TOMAR N K, SRIVASTAVA A, BAGCI U, et al. Automatic polyp segmentation with multiple kernel dilated convolution network[C]//Proceedings of 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS). [S.l.]: IEEE, 2022: 317-322.
- [32] TOMAR N K, JHA D, BAGCI U, et al. TGANet: Text-guided attention for improved polyp segmentation[C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022: 151-160.

作者简介:



吴港(2001-),男,硕士研究生,研究方向:图像分割, E-mail:1715174709@qq.com。



全海燕(1970-),通信作者,男,博士,副教授,硕士生导师,研究方向:信号处理、智能优化等, E-mail: quanhaiyan@163.com。