ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 39,No. 1,Jan. 2024,pp. 223－235
DOI:10. 16337/j. 1004-9037. 2024. 01. 020

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于多重注意力和schatten-*p*范数的息肉分割网络

李　苏[1,2]，刘国奇[1,2]，刘　栋[1,2]，赵曼琪[1,2]

(1. 河南师范大学计算机与信息工程学院,新乡 453007;2. 河南师范大学河南省教育人工智能与个性化学习重点实验室,新乡 453007)

摘　要：自动准确的息肉定位分割方法可以在结直肠癌病变早期及时地发现息肉,大大降低癌变几率。编解码结构作为近年来息肉分割中最主流的网络结构,已经得到了很大的改进,如提高模型捕获全局上下文特征和局部特征的能力,使用深层特征对浅层解码做指导。但是息肉形状和大小不一,在编码时,由于卷积特性容易过于陷入局部信息挖掘,而失去远程信息依赖关系;还有一些息肉图像存在对比度低、空间复杂的特性,导致息肉与背景两者极易混淆。本文提出了基于多重注意力和schatten-*p*范数的息肉分割网络。其中,轴向多重注意力模块利用轴向注意力补充图像中的远程上下文关系,同时补充对边缘、背景信息的关注以实现特征互补,在注意全局特征的同时加强对局部细节特征的捕捉;利用矩阵奇异值和矩阵隐含信息的关联性,引入schatten-*p*范数作约束,从矩阵角度分析数据,辅助模型辨别前景和背景。通过设置大量实验,证明了本文提出方法的有效性,并且MASNet在Kvasir-SEG数据集上对比不同的方法,取得了较好的分割结果。

关键词：息肉分割;卷积;注意力;schatten-*p*　范数

中图分类号：TP391　　**文献标志码**:A

Polyp Segmentation Network Based on Multiple Attention and schatten-*p* Norm

LI Su[1,2], LIU Guoqi[1,2], LIU Dong[1,2], ZHAO Manqi[1,2]

(1. College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China;2. Henan Key Laboratory of Educational Artificial Intelligence and Personalized Learning, Henan Normal University, Xinxiang 453007,China)

Abstract: Automatic and accurate polyp localization and segmentation methods can detect polyps in a timely manner in the early stage of colorectal cancer lesions, greatly reducing the risk of cancer transformation. The encoder-decoder architecture, as the most mainstream network structure in polyp segmentation in recent years, has been greatly improved, such as improving the model's ability to capture global contextual and local features, and using deep features to guide shallow decoding. However, polyps vary in shape and size, and due to their convolutional nature, they are prone to getting too caught up in local information mining and losing remote information dependencies during encoding. Some polyp images also have low contrast and complex spatial characteristics, which makes it easy to confuse the polyp with the background. Based on this, this paper proposes a polyp segmentation network based on multiple attention and schatten-*p* norm(MASNet). Among them, the axial multiple attention module utilizes axial attention to supplement remote contextual relationships in the image, while also paying attention to

基金项目:国家自然科学基金(61901160, U1904123)。
收稿日期:2023-05-10;**修订日期**:2023-09-05

boundary and background information to achieve feature complementarity. It enhances the capture of local detail features while paying attention to global features. By utilizing the correlation between matrix singular values and matrix implicit information, the schatten-p norm is introduced as a constraint to analyze the data from a matrix perspective and assist the model in distinguishing foreground and background. By setting up a large number of experiments, the effectiveness of the proposed method is proven, and MASNet achieves the best segmentation results by comparing different advanced methods on the Kvasir-SEG dataset.

Key words: polyp segmentation; convolution; attention; schatten-p norm

引　言

结直肠癌,作为全球第三大癌症,有80%～95%是由结直肠息肉发展而来[1]。如果能在早期发现息肉,对病人进行诊断和治疗,可以大大降低其向结直肠癌病变的几率。目前,结直肠镜检查是最常用的检查手段,但这一过程需要人工操作,耗费大量人力,成本高昂,并且有较高的误诊率[2]。让机器提供可信的预测能帮助医生更好地诊断病变区域,减少对癌症信号的忽视,具有重要的现实意义。然而,对息肉的准确分割一直是一个具有挑战性的任务,在结肠镜检测的过程中,结直肠息肉图像中病灶区域和周围粘液存在对比度低、与周围粘膜之间边界的模糊和形状不规则等复杂特性,都会导致息肉的分割不准确,甚至漏检。卷积神经网络在大多数图像处理任务上都表现出了优异的性能,随着深度学习的发展,一系列卷积神经网络的变体被应用于息肉分割并取得了不错的效果。目前先进的分割方法几乎都以UNet[3]框架引入的编解码结构作为基本设计。该结构主要有编码器和解码器组成。编码器是有多个卷积块构成,通过卷积缩小特征映射尺度,扩大接受范围。解码器则是逐层上采样,保持神经元感受野的同时逐步将特征映射恢复到原图分辨率。编解码器之间还存在跳跃连接,以弥补编码过程中可能会丢失的信息。UNet++[4]通过改善编解码器之间的跳跃连接层,致力于消除编码层和解码层的语义分歧。ResUNet[5]使用残差单元来改善原始卷积块的学习性能,达到更好的提取特征的作用。PraNet[6]使用并行部分解码器(Parallel partial decoder,PPD)来聚合高层中的特征,对高级特征使用反向注意力(Reverse attention,RA)模块来挖掘边界线索,建立区域与边界的关系,指导更浅层的解码,校准预测。CaraNet[7]在PraNet的基础上,改进了RA模块,引入了轴向注意力来补充图像内部的远程上下文信息,在小目标数据集上起到了不错的效果。ACSNet[8]在编码器最高层使用全局上下文模块(Global context module,GCM)提取全局上下文信息,用局部上下文模块(Local context attention module,LCA)提取边缘细节信息,两组信息流被送入自适应选择模块(Adaptive selection module,ASM)与解码器上采样的非局部信息连接并重新校准权重,让网络基于上下文自动选择,达到全局上下文和局部上下文信息的互相补充以获得更好的分割预测。DCRNet[9]是一个双重上下文关系网络,用来捕捉图像内和图像间的上下文关系。该方法认为不应该仅考虑图像内部的上下文关系,还基于息肉多样性导致的分割困难,设计了一种情景存储器,从整个数据集的整体角度来探索图像间的上下文相关性。

全局和局部上下文特征的结合可以为解码过程提供不错的引导信号,但LCA只是补充了边缘信息,忽视了图像内部的长距离依赖关系,对一些背景复杂的息肉不能很好地辨别目标和背景的关系。而且本文认为,网络模型的构建决定了学习能力的上限,损失函数则是为发挥这份能力指引正确的方向,一个合适的损失函数能指引网络取得更好的分割结果。由于在计算机的视野中,图像数据被表示为矩阵形式,所以采用矩阵的理论与方法也可以对图像进行分析处理[10-12]。基于此,本文提出了一种基于多重注意力和schatten-p范数的息肉分割网络(Polyp segmentation network based on multiple attention and schatten-p norm, MASNet)。

1　网络模型

图1是MASNet网络模型,其中:$Loss_{s\text{-}p}$为schatten-p范数最大化约束,即秩最大化;LIB(Loss in the batch)为批次内损失模块。该模型基于编解码体系结构,它总共包含5层,左侧编码器采用的是ResNet34[13]对结肠镜图像进行特征提取,缩小特征映射尺度,聚焦目标区域,并且每层后紧跟一个通道注意力模块(Efficient channel attention,ECA)[14]校准通道方向上的特征响应。同样地,右侧是解码器,进行上采样操作,逐步将特征映射恢复到原图分辨率。本文提出了一个轴向多重注意力(Axial multiple attention, AMA)模块加强网络对边界困难区域的信息挖掘,对编码块(Encoder-block,E-Block)获得的特征用轴向注意力补充分析位置信息,避免过度关注局部而丢失远程上下文信息,同时利用从上一层解码块(Decoder-block,D-Block)获得的预测输出作为指导映射,引导加强网络对边界、背景的关注。GCM被放置在编码器最深层来捕获全局上下文特征,并级联到每层编码器前的ASM中,与解码器的上层输出特征和经过AMA后提取到的特征,共同在ASM中融合,基于上下文选择性地重新校准通道方向上的权重,然后传递到下一个D-Block块。按照文献[9]实现GCM和ASM。

在解码过程中,解码器每层会生成一个不同分辨率的预测图,分别由下采样至相同大小的Ground Truth监督。特别地,在1/16、1/8 Ground Truth下采样层,本文构造了一个新的损失函数,利用schatten-p范数对矩阵的奇异值约束,增强网络对决策边界处目标和背景的判别能力。

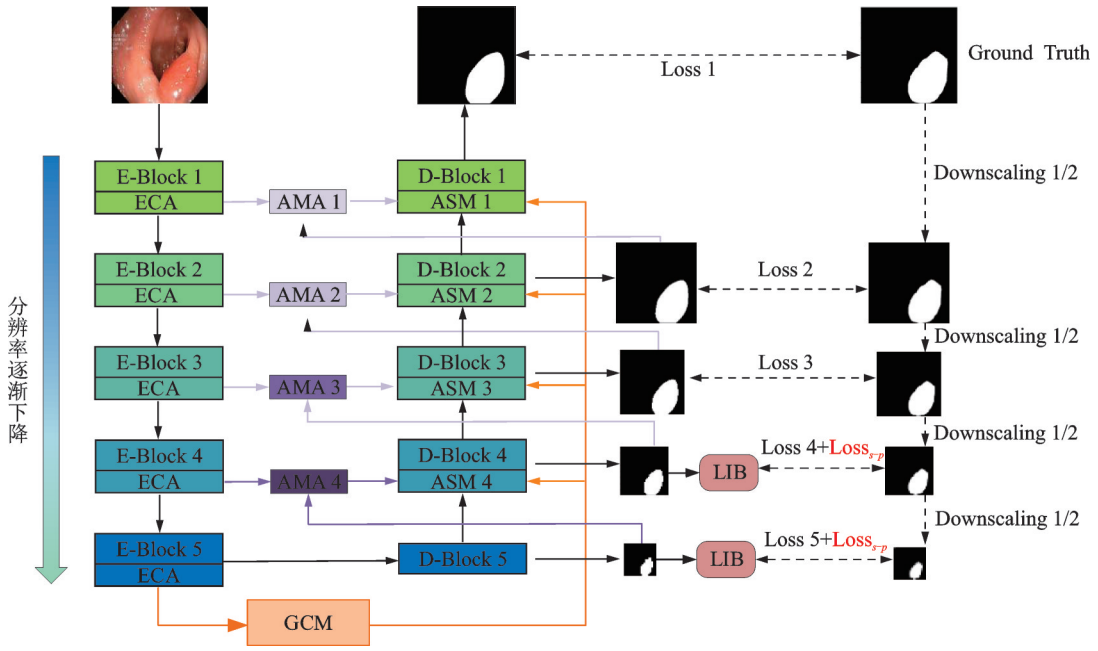

图1　MASNet结构图

Fig.1　Overall architecture of MASNet

1.1　轴向多重注意力

随着层数的增加,卷积神经网络(Convolutional neural network, CNN)会更多地关注局部特征而丢失远程上下文信息,缺乏对图像中的长距离依赖关系的理解。所以在本文中引入了轴向注意力[15]来补充分析定位信息,如图2所示,轴向注意力将自注意力的二维扫描,分为沿高度轴和沿宽度轴的两个一维注意力,降低了计算量的同时还能注意更大区域的内在位置关系。为了在解码时获得更全面的特

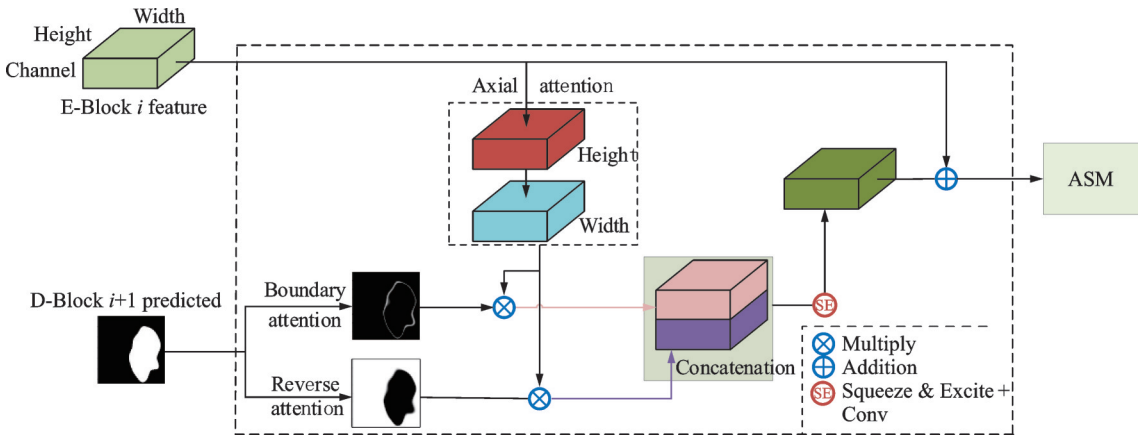

图2 AMA模块结构图

Fig.2 Structure diagram of AMA

征，使用深层解码侧的输出预测做指导，关注边界和背景区域，以实现分层特征互补和预测精化，更加关注不确定和复杂的区域。

自注意力[16]捕捉特征内部相关性的实现如式(1~2)所示，将原矩阵 $\boldsymbol{X}$ 与3个可学习矩阵相乘，映射到向量空间，分别得到查询向量 $\boldsymbol{Q}$、键向量 $\boldsymbol{K}$、值向量 $\boldsymbol{V}$。通过将 $\boldsymbol{Q}$ 与 $\boldsymbol{K}$ 的转置做点积，计算两者的相似性，从几何角度看，点积反应了两个向量在方向上的相似度，结果越大越相似。之后用Softmax函数对结果归一化，得到注意力得分，与值向量 $\boldsymbol{V}$ 相乘，即得到经过注意力机制加权的表示。

$$\boldsymbol{Q}=\boldsymbol{X}\times\boldsymbol{W}_Q,\quad \boldsymbol{K}=\boldsymbol{X}\times\boldsymbol{W}_K,\quad \boldsymbol{V}=\boldsymbol{X}\times\boldsymbol{W}_V \tag{1}$$

$$\mathrm{Att}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_K}}\right)\boldsymbol{V} \tag{2}$$

式中：$\boldsymbol{W}_Q$、$\boldsymbol{W}_K$、$\boldsymbol{W}_V$ 为3个均可学习的查询矩阵(Query)、键矩阵(Key)以及值矩阵(Value)；d_K 为 $\boldsymbol{K}$ 的维度，除以 d_K 是为了防止 $\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}$ 值过大，保持梯度稳定。

轴向注意力是基于自注意力的一种改进，如图3所示，它将传统的自注意力计算方式分为了分别沿高度轴和宽度轴的两次一维扫描。在计算时，将输入通过核大小1×1的卷积分别得到 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$，将通道维度(C)分别和宽度维度(W)、高度维度(H)压缩到一个长为 $W\times C$ 或 $H\times C$ 的维度，不仅减小了计算量，并且糅合了不同通道间的信息关系，能在更大区域内捕获长依赖关系。归一化方式上，Softmax函

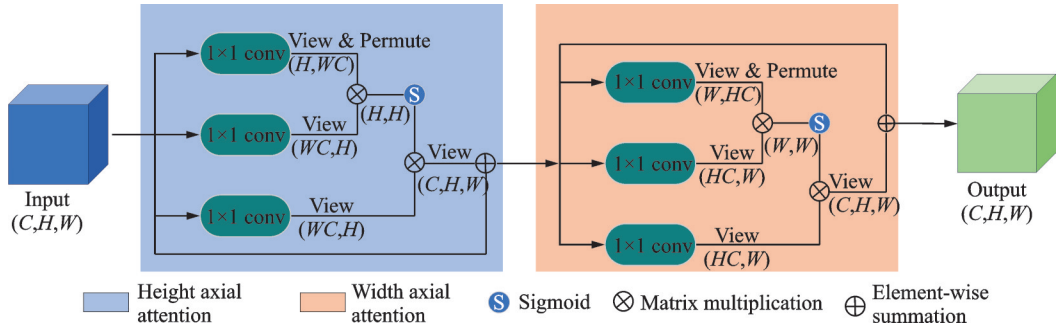

图3 轴向注意力结构图

Fig.3 Architecture of axial attention

数的输出是所有元素权重之和归为1的概率分布,Sigmoid函数的输出是伯努利分布,返回每个元素是前景的概率值,为了与后续的特征图更好地加权,在这里选用Sigmoid函数。

用解码侧D-Block $i+1$层生成的预测图做引导,利用式(3~4)可以得到权重图$\mathrm{Att}_{\mathrm{b}}^{i}$、$\mathrm{Att}_{\mathrm{r}}^{i}$,分别是边界注意图[9]、反转注意图[7](如图2左下所示)。

$$\mathrm{Att}_{i}^{j}=1-\frac{\left|P_{i+1}^{j}-T\right|}{\max(T,1-T)} \tag{3}$$

$$\mathrm{Att}_{i}^{j}=1-\mathrm{Sigmoid}(P_{i+1}^{j}) \tag{4}$$

式中:i表示编解码结构的第i层;$P^{j}\in[0,1]$为预测图P中的第j个元素的值;$\mathrm{Att}^{j}\in[0,1]$为计算之后的第j个位置的值;T为阈值,对应位置的预测值越接近阈值T,就可获得越大的关注度,在本文中将T设置为0.5,用于聚焦目标域和背景域交界处的边界域。

具体地,AMA模块被放到编解码结构的$\{i|1,2,3,4\}$层捕获特征,AMA_{i}对来自E-Block i层的特征图Map^{i}做轴向注意,补充图像中的远程依赖关系,得到特征图$\mathrm{Map}_{\mathrm{A}}^{i}$,将其与$\mathrm{Att}_{\mathrm{b}}^{i}$、$\mathrm{Att}_{\mathrm{r}}^{i}$分别相乘,这有助于确定每个特征图需要注意哪一部分特征。再将两者按通道方向连接,通过Squeeze-and-Excite[17]重新校准通道方向上的权重,并送入卷积层,以将通道映射为原图Map^{i}一样大小,其作为一个多重注意后的权重图与原图相加,以实现对不同特征的精细捕捉。可将AMA模块的具体实现描述为

$$\mathrm{AMA}^{i}=\mathrm{Conv}(\mathrm{SE}(\mathrm{Concat}(\mathrm{Map}_{\mathrm{A}}^{i}\otimes\mathrm{Att}_{\mathrm{b}}^{i},\mathrm{Map}_{\mathrm{A}}^{i}\otimes\mathrm{Att}_{\mathrm{r}}^{i})))\oplus\mathrm{Map}^{i} \tag{5}$$

1.2　schatten-p范数

范数主要是对矩阵和向量的一种描述,在机器学习中,经常使用范数约束数据之间的关系,让模型具有人们想要的特性,比如低秩、稀疏、平滑。其中一个流行的范数是核范数,核范数是矩阵所有奇异值σ的和。矩阵的奇异值往往对应着矩阵中隐含的重要信息,且重要性和奇异值大小正相关,隐含的重要信息在分割预测中可以理解为不同层次的特征。基于此,本文针对目标与背景边界处相似、难以区分的问题,引入了另一个流行的schatten-p范数约束奇异值,相比于核范数,它多了一个参数p作幂指数,使用起来更加灵活,对矩阵$\boldsymbol{X}$,则有

$$\|\boldsymbol{X}\|_{s\text{-}p}=\left(\sum_{i=1}^{\mathrm{rank}(\boldsymbol{X})}\sigma_{i}^{p}(\boldsymbol{X})\right)^{\frac{1}{p}} \tag{6}$$

式中:$\|\boldsymbol{X}\|_{s\text{-}p}$表示矩阵$\boldsymbol{X}$的schatten-$p$范数;$\sigma_{i}$为矩阵的第$i$个奇异值,可以看出,当$p=1$时,schatten-$p$范数即为核范数,当$0<p<1$,相比于核范数,schatten-$p$范数可削弱较大奇异值在目标函数中的比重。

图4是按照不同的参数p削弱主要特征和次要特征的差距之后再组合的结果,图例是经过归一化处理的灰度图。$p=2$时,增强了较大奇异值的比重,导致低层次特征的作用被忽视,缺少了精细化信息;而当$0<p<1$时,削弱了主要特征,增强了次要特征的精细化表达,使得图像的细节信息更加突出。挖掘次要特征所隐含信息的前提是保全主要特征的地位,本文最初以$p=0.5$为实验,并于后续实验中

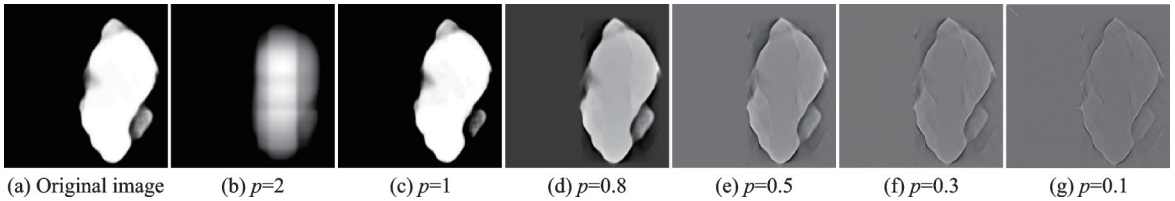

图4　不同p值下的矩阵重组

Fig.4　Matrix reorganization with different p

调整参数p以达到最好效果。

本文认为高级特征中的低层次信息对确定边界是有用的，可以通过参数p影响不同奇异值的比重，改变不同层次特征的重要性，达到削弱主要特征、增强次要特征的作用，让网络对特征有新的认知，加强网络对于决策边界附近特征的学习，从而更准确地判别目标与背景。故应用到网络中，构建如下损失函数

$$\mathrm{Loss}_{s\text{-}p}=-\|\boldsymbol{X}\|_{s\text{-}p} \tag{7}$$

$$\mathrm{Loss}_{4,5}=\sum_{i=4}^{5}(\mathrm{Loss}_{\mathrm{seg}}(G_i,P_i)+\mathrm{Loss}_{s\text{-}p}(P_i)) \tag{8}$$

式中：$\mathrm{Loss}_{\mathrm{seq}}$为图像中常用的损失函数；$G_i$、$P_i$分别是解码块的第$i$层的Ground Truth和预测图Prediction。由于解码到浅层时，特征已经相对稳健，且预测图的分辨率较大，进行奇异值分解时计算量巨大，本文只在第5层和第4层解码层添加schatten-p范数约束，减少计算量的同时以反馈上层解码块更多的高级层次特征，上层接受多方信息来源后，在$\mathrm{Loss}_{\mathrm{seg}}$的监督下融合得到更健全的特征，以达到更好的约束网络对于决策边界附近特征的学习。

1.3 批次内损失

由于息肉图像的多样性，同等看待所有图像的特征可能对一些复杂图像并不友好，受通道注意力机制与DCRNet探索图像间关系的启发，本文提出了LIB模块，在计算损失时，为每次参与迭代的图像分配不同权重，引起网络不同的关注度，更好地促进梯度流动。权重系数是一组可学习的参数，由网络自主学习。如图5所示，作为输入的batchsize张预测图，经全局平均池化将每张图像压缩为一个特征值，即$\{P_b|b=1,2,\cdots,\mathrm{batchsize}\}$，然后送入两次batchsize长度的全连接层，具有更多的非线性，可以更好地拟合不同特征性质的复杂关联，最后的全连接层输出经过Softmax激活可得一组概率值之和为1的权重系数$\{n_b|b=1,2,\cdots,\mathrm{batchsize}\}$，反映不同图像的重要程度。使用获得的权重系数对损失函数重新加权，能够根据输入特征调整各图像的损失重要性，使网络能够较好地拟合对不同图像间的特征分布，达到对整体更好的分割效果。

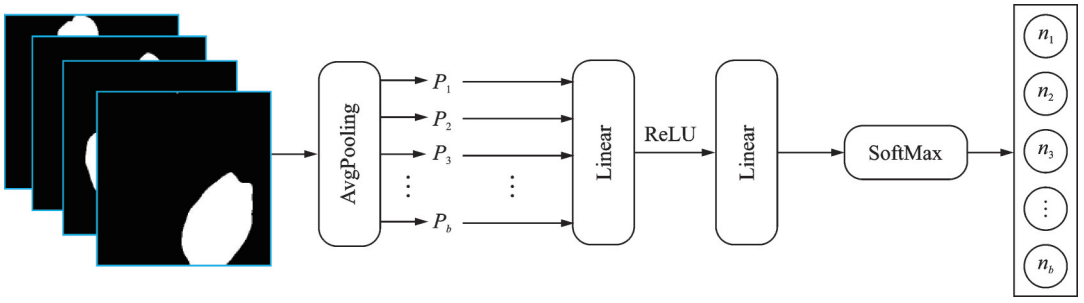

图5 LIB结构图

Fig.5 Structural diagram of LIB

1.4 损失函数

为了提高模型稳健性，本文在损失函数方面，如式(9)所示，选择二值交叉熵(Binary cross entropy，BCE)损失逐像素点拉近，Dice损失从全局上考察，两者搭配使用从互补的角度更好地考察图像。当前后景分布极不均衡(背景过大，息肉过小)时，即使全判断为背景，BCE损失也会因过多的背景像素被准确辨别而认为分割效果好，而Dice损失却不受背景影响。当前景内容不均衡(大息肉，小息肉)时，Dice损失会趋向于学习大的目标，而BCE损失则保证了对小样本的学习。

$$\mathrm{Loss}_{\mathrm{seg}}=\mathrm{Loss}_{\mathrm{BCE}}+\mathrm{Loss}_{\mathrm{Dice}} \tag{9}$$

$$\mathrm{Loss}_{\mathrm{BCE}}=-\frac{1}{n}\sum_{1}^{n}(G_n\times \ln P_n+(1-G_n)\times \ln(1-P_n)) \tag{10}$$

式中:$G_n\in[0,1]$表示Ground Truth的第n个像素值;$P_n\in[0,1]$表示预测图的第n个像素值。

$$\mathrm{Loss}_{\mathrm{Dice}}=1-\frac{2|G\cap P|}{|G|+|P|} \tag{11}$$

式中:$|G\cap P|$表示Ground Truth和预测图的交集;$|G|$、$|P|$分别表示其元素和。

为充分训练底层特征,以便用schatten-p范数约束底层预测以反馈上层更好的特征,本文选用深监督策略来优化梯度更新过程,在5层解码侧分别有下采样至相同大小的Ground Truth进行监督,全文的损失表述为

$$\mathrm{Loss}_{\mathrm{total}}=\sum_{i=1}^{5}\mathrm{Loss}_{\mathrm{seg}}(G_i,P_i)+\sum_{i=4}^{5}\mathrm{Loss}_{s\text{-}p}(P_i) \tag{12}$$

2　实验分析

2.1　实验数据

本文选择了Kvasir-SEG数据集[18]对MASNet进行评估和比较。该数据集包括1 000张胃肠道息肉图像和相应的Ground Truth,如图6所示,每张图像至少包括1个息肉区域,数据集中的图像分辨率范围从332像素×482像素到1 920像素×1 072像素,目标息肉的形状、大小不一,背景复杂多样。实验时按照8∶1∶1的比例,将1 000张图像随机划分为训练集、验证集和测试集。

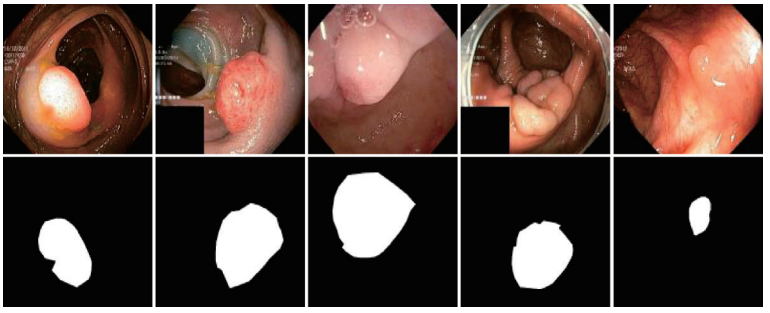

图6　Kvasir-SEG的部分图像与对应的Ground Truth

Fig.6　Partial images and corresponding Ground Truth of Kvasir-SEG

2.2　实验细节

在网络训练期间,统一将图片大小调整为320像素×320像素以规范批次、减小计算量。对其采用水平和垂直翻转、旋转、缩放、移动和随机裁剪等方法作为数据增强策略,以减少过拟合风险,提高模型的稳健性。使用随机梯度下降(Stochastic gradient descent,SGD)优化器对模型进行优化,设动量为0.9,权重衰减率为0.000 5。初始学习率init_lr设为0.001,后续学习率lr随当前训练轮次epoch增大而减小,具体策略为

$$\mathrm{lr}=\mathrm{init\_lr}\times\left(1-\frac{\mathrm{epoch}}{n\mathrm{epoch}}\right)^{0.9} \tag{13}$$

所有实验都在NVIDIA GeForce RTX GPU上使用PyTorch[19]框架实现,$n$epoch为150,batchsize设为8。

2.3 评价指标

本文选择了多个评价指标从不同方面来评估网络的稳健性，包括召回率(Recall, Rec)、特异度(Specificity, Spec)、精确率(Precision, Prec)、Dice系数(Dice coefficient, Dice)、息肉交并比(Intersection-over-union for polyp, IoUp)、背景交并比(IoU for background, IoUb)、平均交并比(Mean IoU, mIoU)、准确率(Accuracy, Acc)。计算方法如下

$$\mathrm{Rec}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{14}$$

$$\mathrm{Spec}=\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \tag{15}$$

$$\mathrm{Prec}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{16}$$

$$\mathrm{Dice}=\frac{2*\mathrm{TP}}{2*\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \tag{17}$$

$$\mathrm{IoUp}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \tag{18}$$

$$\mathrm{IoUb}=\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{19}$$

$$\mathrm{mIoU}=\frac{\mathrm{IoUp}+\mathrm{IoUb}}{2.0} \tag{20}$$

$$\mathrm{Acc}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{21}$$

式中：TP和TN分别表示被正确分割的息肉区域和背景区域；FP代表将背景错认为息肉的区域；FN表示将息肉错认为背景的区域。

2.4 实验结果

在Kvasir-SEG上，本文将MASNet与UNet、ResUnet、PraNet、CaraNet、ACSNet、DCRNet、TransFuse[20]和Polyp-PVT[21]八个先进网络进行了比较，其中UNet、ResUnet是医学图像分割中比较经典的网络模型，PraNet、CaraNet、ACSNet和DCRNet是编解码结构中比较先进有效的网络，TransFuse是融合了transformer和CNN优势的经典有效网络，Polyp-PVT是基于金字塔结构Transformer的先进网络。

实验结果如表1所示，各个网络模型在不同指标下的最优数据经过了加粗处理，可以看出，MASNet在Dice、IoUp、IoUb、mIoU、Acc五项指标上均取得了最高值，尤其是Dice、IoUp两个重量级指标上领先其

表1 各个网络模型在Kvasir-SEG上的对比结果

Table 1 Comparison results of various network on Kvasir-SEG %

网络模型	Rec	Spec	Prec	Dice	IoUp	IoUb	mIoU	Acc
UNet	77.75	98.24	86.08	78.25	69.05	93.42	81.23	94.51
ResUnet	73.24	95.91	73.26	68.24	55.61	90.18	72.90	91.50
PraNet	89.15	98.30	93.24	88.77	83.26	96.28	89.77	97.10
CaraNet	84.48	97.40	83.87	81.83	73.22	93.94	83.58	95.04
ACSNet	92.02	97.53	91.40	90.07	84.68	95.76	90.22	96.83
DCRNet	90.12	98.72	**94.16**	90.41	85.05	96.37	90.71	97.12
TransFuse	89.25	**99.23**	94.09	90.58	85.01	96.67	90.85	97.34
Polyp-PVT	**92.29**	98.16	92.91	90.89	86.17	96.26	91.22	96.97
MASNet	91.99	98.91	93.61	**92.00**	**86.89**	**96.87**	**91.88**	**97.63**

他网络,并且另3项指标也稳居前三。综合来看,MASNet获得了比其他先进网络更好的性能。

为了更直观地体现网络效能差异性,本文将各网络对相同图像的分割预测图进行可视化比较,如图7所示,蓝色线是Ground Truth在原图相同位置的边界,红色线是预测图在原图的边界,最右侧1列是本文MASNet网络的分割结果。从前两行的分割结果对比中可以看到,MASNet对边界的处理具备不错的效果,第3、4行表现出对小目标也有不错的识别能力,尤其第4行的息肉与其他息肉特征迥异,导致只有DCRNet和MASNet两个考虑图像间关系的网络相对准确地找到了息肉区域。最后两行的息肉对比度低、边缘很模糊,只有MASNet对两张都有不错的分割。综合来看,MASNet在其他几个网络处理不佳的图像上,也能有不错的分割预测。

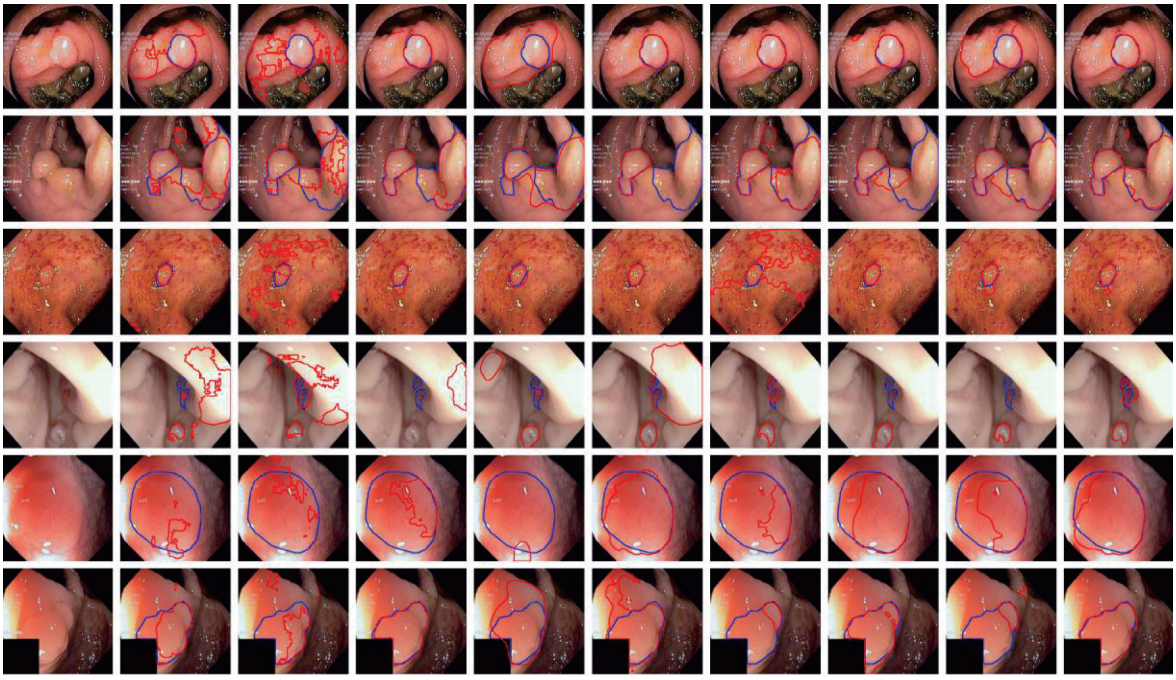

图7　各个网络分割效果可视化

Fig.7　Visualization segmentation effect of each network

2.5　消融实验

为了证明模块的有效性,本文将MASNet模型去除AMA、schatten-p范数、LIB之后的网络作为基准网络(Baseline),然后逐个添加模块证明有效性。实验结果如表2所示,加粗为最优值。

表2　Kvasir-SEG上的消融结果

Table 2　Results of ablation on Kvasir-SEG　%

网络模型	Rec	Spec	Prec	Dice	IoUp	IoUb	mIoU	Acc
Baseline	89.91	98.55	91.42	89.15	83.36	96.12	89.74	96.83
Baseline + AMA	90.81	98.82	**96.20**	90.80	85.32	96.61	90.97	97.29
Baseline + $\text{Loss}_{s\text{-}p}$	91.49	98.52	92.96	90.91	85.43	96.40	90.91	97.25
Baseline + AMA + $\text{Loss}_{s\text{-}p}$	91.71	**99.04**	93.04	91.52	86.62	**96.99**	91.81	97.62
Baseline + AMA + $\text{Loss}_{s\text{-}p}$ +LIB	**91.99**	98.91	93.61	**92.00**	**86.89**	96.87	**91.88**	**97.63**

研究结果发现，AMA模块在Dice表现上有1.65%的提升，schatten-p范数的引入，在Dice上提升了1.76%，添加了AMA和Loss$_{s\text{-}p}$之后，提升了2.37%，本文的最终网络MASNet在Dice表现上获得了92.00%的得分，比Baseline提升了2.85%。

此外，部分分割结果可视化如图8、9所示，蓝色轮廓为Ground Truth，红色轮廓是预测结果。从图8可以看出，在背景空间复杂、对比度低的图像上，通过本文构造的损失函数的约束，使模型对背景和目标有了不错的鉴别能力，对一些复杂图像的分割有很好的提升效果。图9是AMA的消融结果，这个模块是为了补充局部上下文特征，主要是捕获边缘细节信息，从图中可以看出，添加了AMA之后的模型，网络对边界的判断更为准确，精细度上更贴近Ground Truth。

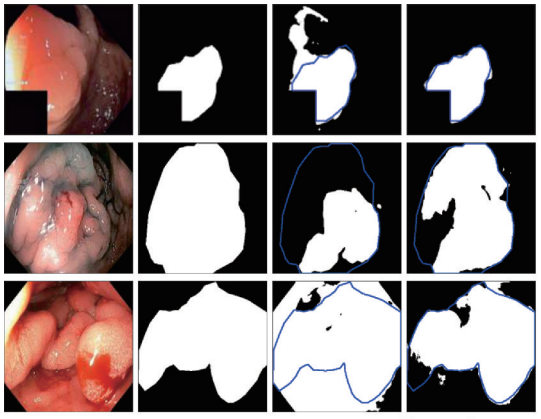

图8 Loss$_{s\text{-}p}$的消融结果可视化

Fig.8 Visualization of ablation results for Loss$_{s\text{-}p}$

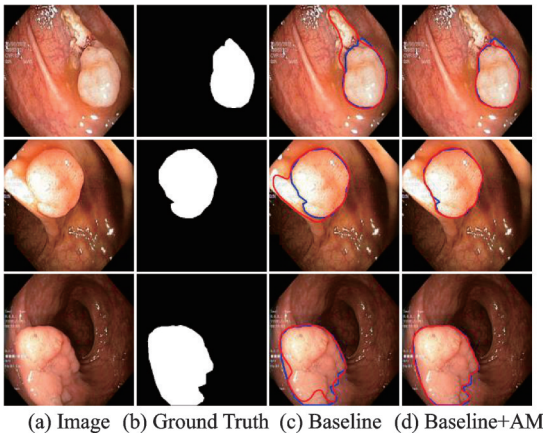

图9 AMA模块的消融结果可视化

Fig.9 Visualization of ablation results for AMA

2.6 泛化实验

为了验证模型的泛化性能，防止过拟合，本文将各网络在Kvasir-SEG数据集上训练得到的权重放在了CVC-300[22]、CVC-ClinicDB[23]、CVC-ColonDB[24]和ETIS-LaribPolypDB[25]四个其他的息肉数据集上进行测试，测试集的数量及图像尺寸大小如表3所示。

表3 泛化实验数据集信息

Table 3 Information of generalized experimental datasets

数据集	图像数量/幅	图像大小/(像素×像素)
CVC-300	60	574×500
CVC-ClinicDB	62	384×288
CVC-ColonDB	380	574×500
ETIS-LaribPolypDB	196	1 225×996

各数据集的图像来自不同的结肠镜检查序列，使得采集到的息肉图像不仅色彩、背景空间特征不同，而且息肉的形状、大小上均有差别，图10展示了这4个数据集内的图像，蓝色轮廓为息肉区域。其中，CVC-300中的息肉大都背景单一、形状圆

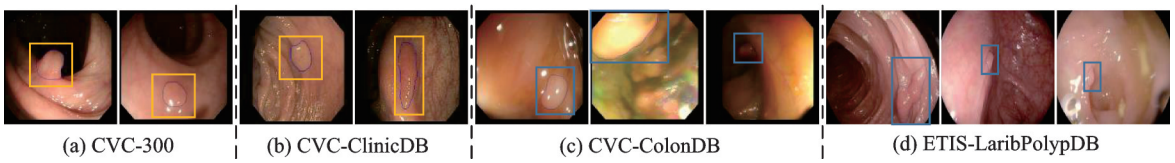

图10 不同数据集的息肉图像对比

Fig.10 Comparison of polyp images from different datasets

润但边缘模糊;CVC-ClinicDB中的息肉部分背景复杂、形状多样;CVC-ColonDB中有的息肉图像中粘液过多、背景肮脏,并且还有一部分小型息肉;而ETIS-LaribPolypDB中的息肉不仅边缘模糊、有粘液残留,而且大部分都是很小型的息肉,该测试集是4个数据集中与训练集特征差异最大、最难以分割的。

使用Dice、IoUp、Acc三个评价指标对模型性能进行评价,测试结果如表4所示,其中加粗标注为最优值。可以看出,MASNet与其他模型相比,在4个数据集上的测试结果基本上都稳居前二,说明MASNet也有不错的泛化性能。网络的好坏也取决于训练集的质量,训练集决定了网络学习知识的来源,经背景复杂、息肉形状多样性的Kvasir-SEG训练集训练出的网络,在特征相似的测试集CVC-300和CVC-ClinicDB上表现都不错。而在小型息肉占大部分的CVC-ColonDB和ETIS-LaribPolypDB测试集上,由于训练集Kvasir-SEG中绝大多数为大、中息肉,这种特征局限性导致训练出的所有网络泛化都表现不佳,Dice得分更是无一超过75%。

表4　泛化实验结果

Table 4　Results of generalization experiment　%

数据集	评价指标	UNet	ResUnet	PraNet	CaraNet	ACSNet	DCRNet	TransFuse	Polyp-PVT	MASNet
CVC-300	Dice	56.72	20.55	83.80	81.06	85.26	71.87	82.19	82.25	**85.32**
	IoUp	48.63	13.01	76.53	71.99	**78.45**	64.82	72.62	74.93	77.88
	Acc	98.13	96.40	99.14	98.51	**99.18**	98.11	98.68	96.88	99.14
CVC-ClinicDB	Dice	56.24	46.79	72.41	68.43	74.48	71.74	**79.70**	73.88	79.32
	IoUp	46.42	36.22	65.82	59.34	67.04	63.64	71.17	66.70	**72.44**
	Acc	94.66	93.66	97.16	96.47	96.07	96.11	96.92	92.51	**97.39**
CVC-ColonDB	Dice	41.97	32.34	60.25	58.70	63.85	59.17	**73.69**	69.30	70.48
	IoUp	34.96	22.67	54.31	51.02	57.17	52.32	**64.46**	62.39	63.35
	Acc	94.21	92.98	95.91	95.76	95.73	94.71	96.38	87.67	**96.42**
ETIS-LaribPolypDB	Dice	32.14	24.69	46.45	41.10	37.18	58.17	**64.97**	61.46	62.45
	IoUp	26.55	18.14	42.85	34.05	52.67	51.34	**56.39**	55.45	55.15
	Acc	96.60	95.27	**98.30**	97.17	94.53	96.37	96.53	85.37	95.58

2.7　模型参数比较

表5显示了各模型的参数量和浮点计算次数。其中,计算量与图像大小有关,如TransFuse、Polyp-PVT因使用了transformer结构,对输入大小有要求的则按照原文设置。除此外,均设为320像素×320像素大小。

表5　各模型参数量和计算量对比

Table 5　Params and GFLOPs of each model

参数	UNet	ResUnet	PraNet	CaraNet	ACSNet	DCRNet	TransFuse	Polyp-PVT	MASNet
Params/10^6	31.04	13.04	30.5	44.59	29.45	28.73	26.17	25.11	30.77
FLOPs/10^9	85.53	126.54	10.87	17.97	17.98	14.27	**8.65**	10.02	20.65

3　结束语

本文提出的基于多重注意力和schatten-p范数的息肉分割网络MASNet,采用了传统的编解码体系结构,使用多重注意力模块AMA扩大感受野的同时加强了对局部细节信息的捕捉,提供了更好的局

部上下文特征,引入的schatten-p范数加强了网络对目标和背景的鉴别能力,还利用了LIB模块让网络感知不同图像的重要性,以达到对数据集更好的分割效果。并且通过数据集CVC-300、CVC-ClinicDB、CVC-ColonDB和ETIS-LaribPolypDB验证了网络模型的稳健性。横向对比其他网络,泛化性能保持在正常水平,但是纵向对比自身,泛化结果并不尽如人意,未来将尝试其他方法提高模型的分割效果和泛化性能。

参考文献:

[1] XI Y, XU P. Global colorectal cancer burden in 2020 and projections to 2040[J]. Translational Oncology, 2021, 14(10): 101174.

[2] VAN RIJN J C, REITSMA J B, STOKER J, et al. Polyp miss rate determined by tandem colonoscopy: A systematic review [J]. Official Journal of the American College of Gastroenterology(ACG), 2006, 101(2): 343-350.

[3] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]// Proceedings of Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer International Publishing, 2015: 234-241.

[4] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKHSH 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: 4th International Workshop, DLMIA 2018, and 8th International Workshop. Granada, Spain: Springer International Publishing, 2018: 3-11.

[5] 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.

[6] 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.

[7] LOU A, GUAN S, LOEW M. CaraNet: Context axial reverse attention network for segmentation of small medical objects[J]. Journal of Medical Imaging, 2023, 10(1): 014005.

[8] ZHANG R, LI G, LI Z, et al. Adaptive context selection for polyp segmentation[C]//Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention. Lima, Peru: Springer International Publishing, 2020: 253-262.

[9] YIN Z, LIANG K, MA Z, et al. Duplex contextual relation network for polyp segmentation[C]//Proceedings of 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). [S.l.]: IEEE, 2022: 1-5.

[10] LIU J, MUSIALSKI P, WONKA P, et al. Tensor completion for estimating missing values in visual data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(1): 208-220.

[11] ONGIE G, WILLETT R, NOWAK R D, et al. Algebraic variety models for high-rank matrix completion[C]//Proceedings of the 34th International Conference on Machine Learning. New York: ICML, 2017: 2691-2700.

[12] FAN J, DING L, CHEN Y, et al. Factor group-sparse regularization for efficient low-rank matrix recovery[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. San Diego: NIPS, 2019: 5104-5114.

[13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 770-778.

[14] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]// Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Virtual: IEEE, 2020: 11531-11539.

[15] WANG H, ZHU Y, GREEN B, et al. Axial-deeplab: Stand-alone axial-attention for panoptic segmentation[C]//Proceedings of European Conference on Computer vision. Cham: Springer International Publishing, 2020: 108-126.

[16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. San Diego: NIPS, 2017: 6000-6010.

[17] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[18] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Kvasir-SEG: A segmented polyp dataset[C]//Proceedings of MultiMedia Modeling: 26th International Conference. [S.l.]: Springer International Publishing, 2020: 451-462.

[19] PASZKE A, GROSS S, MASSA F, et al. PyTorch: An imperative style, high-performance deep learning library[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. San Diego: NIPS, 2019: 8026-8037.

[20] ZHANG Y, LIU H, HU Q. Transfuse: Fusing transformers and CNNs for medical image segmentation[C]//Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Strasbourg, France: Springer International Publishing, 2021: 14-24.

[21] DONG B, WANG W, FAN D, et al. Polyp-PVT: Polyp segmentation with Pyramid vision transformers[EB/OL]. (2021-01-11)[2022-05-08]. https://arXiv.org/abs/2108.06932, 2021.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

作者简介:

李苏(2000-),**通信作者**,男,硕士研究生,研究方向:医学图像分割,E-mail: 2574625635@qq.com.

刘国奇(1984-),男,博士,副教授,研究方向:计算机视觉、图像分割、机器学习。

刘栋(1976-),男,教授,研究方向:教育数据挖掘和复杂网络分析。

赵曼琪(1998-),女,硕士研究生,研究方向:计算机视觉、图像分割。

(编辑:刘彦东)