

引用格式: HUANG Hong, YANG Yichuan, WANG Long, et al. Boundary Perception Network for Pathological Image Segmentation[J]. Acta Photonica Sinica, 2024, 53(1):0111003

黄鸿, 杨沂川, 王龙, 等. 面向病理图像分割的边缘感知网络[J]. 光子学报, 2024, 53(1):0111003

# 面向病理图像分割的边缘感知网络

黄鸿<sup>1</sup>, 杨沂川<sup>1</sup>, 王龙<sup>1</sup>, 郑福建<sup>1</sup>, 吴剑<sup>2</sup>

(1 重庆大学 光电技术与系统教育部重点实验室, 重庆 400044)

(2 重庆大学附属肿瘤医院 头颈肿瘤中心, 重庆 400030)

**摘要:**提出了一种针对病理切片图像的端到端语义分割方法——边缘感知网络(BPNet),以提高病理图像分割精度。BPNet网络首先在解码器阶段增加边缘感知模块,改善网络对于病理图像边缘的特征信息提取能力。然后,采用自适应通道注意力模块弥补不同层次特征间的语义差距,进一步加强网络的特征聚合能力。在此基础上,设计了一种基于结构和边缘的联合损失函数,以实现最佳的病理图像分割结果。在GlaS和MoNuSeg两个公开病理数据集上的分割实验结果表明,所提方法的Dice系数得分在两个数据集上分别达到92.21%和81.18%,有效提升了病理图像的分割精度。

**关键词:**病理图像;自动分割;深度学习;边缘增强;联合损失函数

中图分类号: TP391.41

文献标识码: A

doi:10.3788/gzxb20245301.0111003

## 0 引言

在临床诊断中,对病理图像中的病灶区域进行精确分割和定量评估能够为后续的病理分析提供有价值的信息,在理解、诊断和治疗疾病方面发挥了关键作用<sup>[1]</sup>。然而,传统的手动分割方法耗时耗力,并且严重依赖病理学家的经验<sup>[2]</sup>。因此,高效的病理图像自动分割方法就显得尤其重要。

传统的图像分割方法主要基于手工特征,包括阈值分割法、图割法和分水岭法等<sup>[3]</sup>。典型的阈值分割法有大津法<sup>[4]</sup>,其主要原理是通过最大化类间方差来计算分割阈值,因其效率高而被广泛应用。分水岭算法<sup>[5]</sup>检测速度快且分割结果具有空间连续性,但该算法对噪声敏感且不适用于大区域分割。典型的图割法有Graphcut法<sup>[6]</sup>,其原理为利用图像纹理信息和边界信息先构建一张加权图,通过边权值的大小来筛选边,使划分的子图彼此尽量相互分离,从而实现分割的目的。该方法具有较好的泛化性,但对于相似目标分割能力差。上述传统图像分割方法主要适用于早期计算机辅助病理图像分割的研究中,过度依赖于手工特征的先验知识,而深层鉴别特征表达能力不足,导致其分割性能受限。

近年来,深度学习在自然图像处理任务中表现出了卓越的性能,并应用于生物医学图像分割。基于编码器-解码器结构的卷积神经网络(Convolutional Neural Network, CNN)因其可实现端到端的自动分割,在准确性以及效率方面相较于传统分割方法体现出明显优势<sup>[7]</sup>。U-Net<sup>[8]</sup>作为医学图像分割中广泛使用的编码器-解码器网络结构,其编码器捕获低级和高级特征,解码器利用上采样层恢复空间分辨率并通过跳跃连接实现特征融合来构建最终结果。该网络能够在病理图像数据缺乏的情况下完成图像分割任务,但由于跳跃连接导致了编码器与解码器之间的语义差距,限制了其分割性能。

为了克服上述限制,UNet++<sup>[9]</sup>在跳跃连接的过程中添加了一系列嵌套、密集的连接路径,以形成密集的编码-解码网络,并带有深监督实现对目标的分割。IBTEHAZ N等<sup>[10]</sup>提出一种MultiResUNet,观察到编码器特征和解码器特征之间可能存在语义差距,通过引入具有残差结构的Res-Path来改进跳跃连接,从而

基金项目:国家自然科学基金(No. 42071302),重庆市留学人员回国创业创新支持计划(No. cx2019144),重庆市科研机构绩效激励引导专项(No. cstc2021jxjl0064)

第一作者:黄鸿, hhuang@cqu.edu.cn

收稿日期:2023-07-04;录用日期:2023-08-01

<http://www.photon.ac.cn>

提升分割精度。OKTAY O等<sup>[11]</sup>提出 Attention U-Net对编码-解码块使用注意力门来突出特定局部区域的显著特征,同时抑制不相干区域。WANG JS等<sup>[12]</sup>提出了一种基于乳腺癌组织微阵列样本高光谱图像的癌巢分割方法,该方法结合了U-Net神经网络和无监督主成分分析的优点,在减少模型计算量的同时,提高了目标分割精度。黄鸿等<sup>[13]</sup>将不同深度解码特征重新进行组合并实现深度特征融合,提升了深层特征在预测中的作用,改善了分割性能。ZHAO BC等<sup>[14]</sup>提出具有渐进密集特征聚集的感知分割网络,通过提取病理图像的RGB信息、染色信息以及分割位置信息,使网络学习到更多有用的特征,提高病理图像分割精度。但是上述网络仅改变网络结构以增强特征间的交互,忽略了大部分病理图像病灶区域聚集、细胞形态差异过小的特点。

为了解决上述问题,ZAMORA W等<sup>[15]</sup>基于U-Net网络框架,通过图像前景、背景和细胞核边缘的距离信息来提取分割目标的形态特征,将形态学特征与全卷积网络模型结合,实现细胞核实例分割,且细胞核分割精度明显优于U-Net。FANG Y等<sup>[16]</sup>提出了一种具有面积和边界约束的选择性特征聚合网络用于息肉分割,考虑了区域和边界之间的相关性,并通过附加的边缘监督获得了良好的结果。FAN DP等<sup>[17]</sup>提出具有前景背景区分注意力模块作为补充特征的PraNet,在分割结果的边缘细化上取得了较大的突破。CHEN S等<sup>[18]</sup>提出了一种多边形上下文感知网络,通过对目标像素点集进行采样获取每个点集的质心与区域边界构成的多边形特征,利用置信度加权模块融合输出特征,并引入了形状感知损失,在一定程度上解决了细胞核分布拥挤和边界模糊的问题,提高了预测分割精度。ACSNet<sup>[19]</sup>利用编码器分支的局部和全局上下文特征为每个解码步骤提供具有空间注意力的引导信息,使网络在医学图像分割结果上取得了进一步的提升。由于CNN结构存在固有的归纳偏差,缺乏对图像远程依赖关系的学习理解,最近基于Transformer结构的网络也逐渐应用于医学图像分割。TransUNet<sup>[20]</sup>同时具有Transformer和U-Net的优点,一方面将CNN与Transformer的自编码器结合作为医学图像分割任务的强编码器提取全局上下文信息,另一方面解码器将编码特征上采样并与高分辨率的CNN特征图组合以提高定位精度,从而增强网络的分割性能。VALANARASU J等<sup>[21]</sup>提出了一种门控轴线注意力模型来扩展现有的Transformer模型,并提出LoGo训练策略,能够更好的训练医学图像,提高了分割精度。WANG H等<sup>[22]</sup>利用基于Transformer的通道变换器替代U-Net简单的跳跃连接,有效地将编码器多尺度信息交叉融合到解码器部分,分割精度明显优于U-Net。但上述深度学习方法仍存在以下两个问题:大多数网络更关注易于分割的区域,而忽略了分割困难的边缘像素点,限制了分割精度的提高;某些网络虽然提取了边缘信息作为补充,但并没有解决不同特征间语义层次不一致的问题,而利用通道注意力机制来弥补特征间语义差距的现有方法模型复杂度过高,在医学图像稀少的情况下网络模型拟合较慢。

针对上述问题,本文提出了基于边缘感知融合的病理图像分割网络(Boundary Perception Network, BPNNet)。该网络基于经典的编码-解码结构,并构建了边缘感知模块(Boundary Perception Module, BPM)以及自适应通道注意力模块(Adaptive Shuffle Channel Attention Module, ASCAM)来提升网络性能。BPM使边界位置信息深度参与到网络训练的优化过程中,大幅提升了网络分割结果的精细程度。ASCAM在增加极少参数量的情况下,自适应的捕捉了编码、边缘、解码三部分特征信息的交互关系,以弥补不同层次特征间的语义差距,捕获更复杂的通道相关性。此外,设计了边缘与结构的联合损失函数,以优化网络的目标定位和细节处理能力。与其他图像分割方法相比,本文方法在两个公共病理图像数据集上的实验结果都表现出更好的分割性能。

## 1 边缘感知网络

### 1.1 整体网络结构

针对病理图像病灶目标实例多、分布广以及边缘模糊的特性,本文提出了一种面向病理图像分割的边缘感知网络(BPNNet),该模型基于编码器-解码器结构,网络结构如图1所示。其中,编码过程采用基于EfficientNet-B4网络<sup>[23]</sup>的特征编码器。EfficientNet-B4属于EfficientNet系列的一部分,是一种高效的深度神经网络架构,它采用了复合缩放方法,即在网络的深度、宽度和分辨率三个维度上进行缩放,以实现高效的模型设计。本文通过EfficientNet-B4网络的特征编码器进行卷积下采样,提取图像的特征信息,在保持网络性能的同时,显著减少了计算复杂性和参数量。解码器包括解码块、边缘感知模块以及自适应通道注

注意力模块。解码块进行反卷积操作完成特征信息的解码过程。边缘感知模块(BPM)对每层解码器的输出进行边缘提取,再将边缘信息叠加到编码特征上输出边缘感知增强特征。自适应通道注意力模块(ASCAM)在增加极少参数量的情况下,自适应的捕捉编码、边缘、解码三部分特征信息的交互关系,弥补不同层次特征间的语义差距,捕获更复杂的通道相关性,以更好地恢复目标对象的细粒度细节。在此基础上,通过采用联合损失函数对网络进行深监督约束,首先将不同深度的解码特征进行激活输出,分别计算各层的层级监督损失  $\mathcal{L}_{ds_i}$ ,从解码器不同尺度优化隐藏层特征以获得更精确的分割结果;随后通过分割损失  $\mathcal{L}_1$  约束目标区域从而优化网络参数;同时,为了更大程度地提升边缘分割效果,训练过程中增加边缘损失约束  $\mathcal{L}_{edge}$ 。

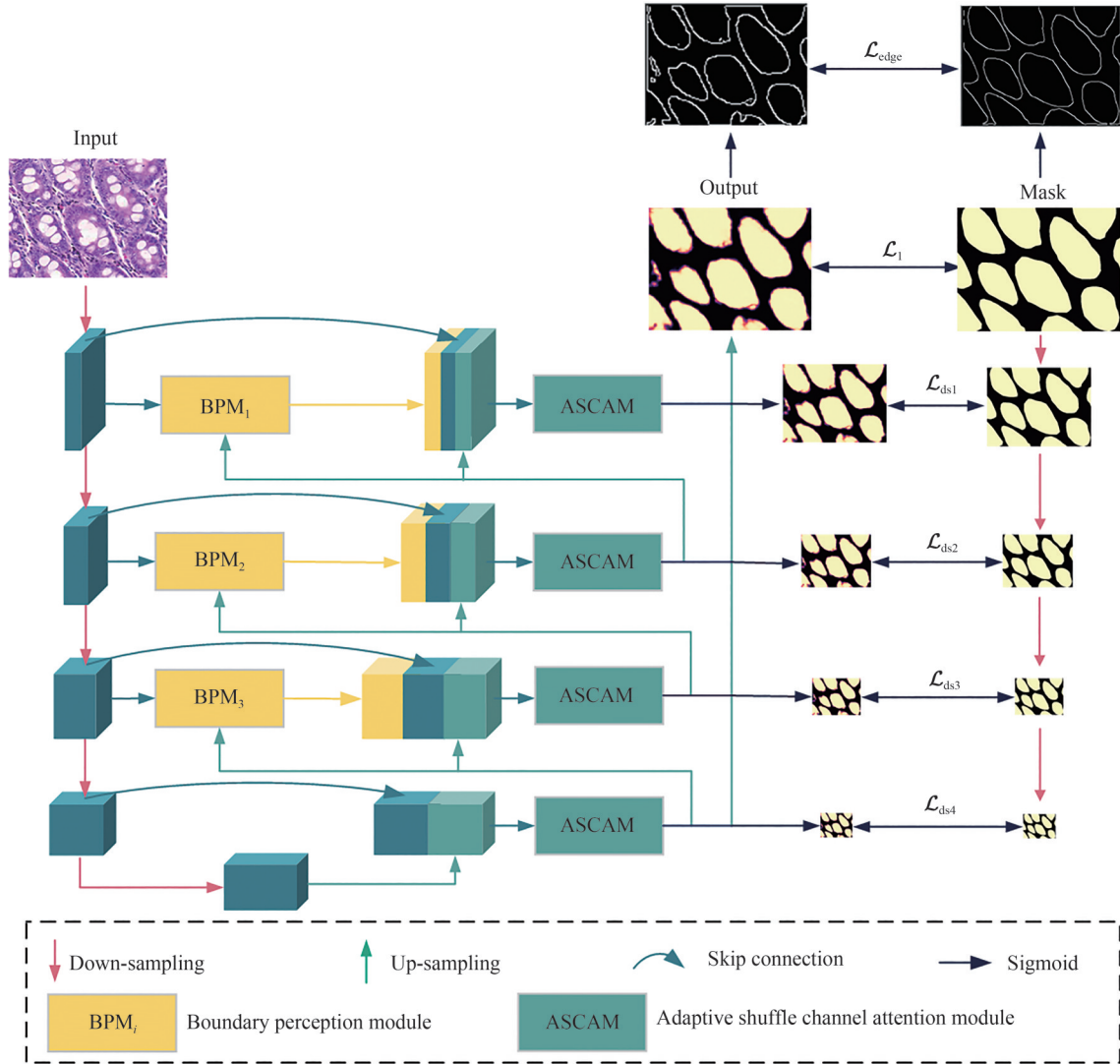


图1 BPNet整体网络结构  
Fig.1 The overall structure of the BPNet algorithm

### 1.2 边缘感知模块

本文设计的边缘感知模块(BPM<sub>i</sub>)具体结构如图2所示。该模块是一种新型边缘注意力方式,通过关注不确定性更高和更复杂的边缘区域,以实现特征互补和预测结果精细化。

在图2中,每个边缘感知模块的输入包含两部分:编码特征以及从解码器上一层网络生成的预测图。其中,编码特征是通过基于EfficientNet-B4网络的特征编码器对输入图像或上一层特征图进行卷积下采样生成的。而整体网络结构中第*i*个( $i = 1, 2, 3$ )边缘感知模块内对应的边缘注意力图由上一层网络预测图生成,表示为  $BP_i \in \mathbb{R}^{H_i \times W_i \times 1}$ ,其中  $H_i$  和  $W_i$  分别是输出特征图的高度和宽度。边缘注意力图的位置信息取值

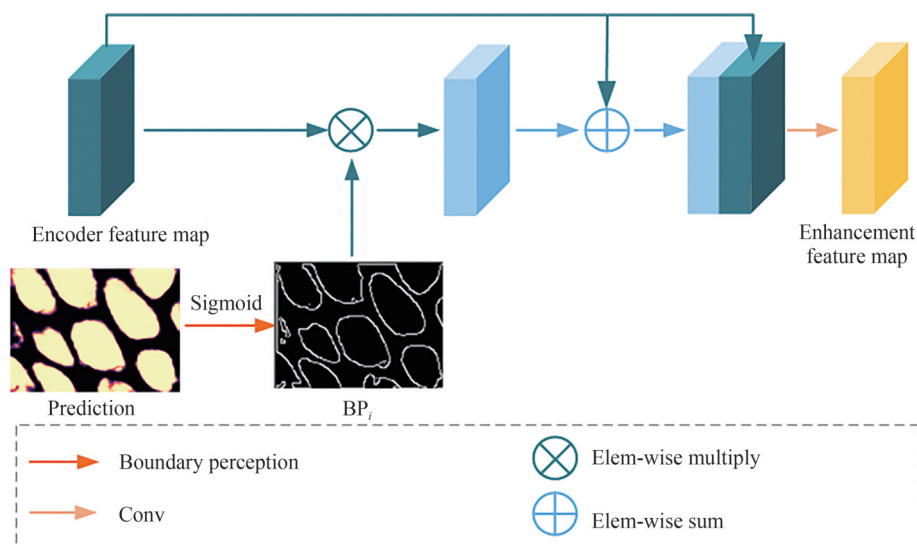


图2 边缘感知模块结构

Fig.2 The structure of attention boundary perception module

为  $j \in [1, 2, \dots, H_i \times W_i]$ , 因此第  $i$  个边缘感知模块内边缘注意力输出的位置特征值  $BP_i^j$  可以定义为

$$BP_i^j = \alpha \times \left( 1 - \cos \left( 2\pi \times \sigma \left( P_{i+1}^j \right) \right) \right) \quad (1)$$

式中,  $P_{i+1}^j$  表示自上而下第  $i+1$  个 ( $i=1, 2, 3$ ) 解码器块输出预测图的第  $j$  个像素值,  $\sigma$  表示 Sigmoid 激活, 因此  $\sigma(P_{i+1}^j)$  的取值范围是  $[0, 1]$ 。难以分类的像素具有接近 0.5 的激活值, 进而使得  $BP_i^j$  值偏高, 即注意力值偏高; 反之激活值越接近 0 或 1,  $BP_i^j$  值偏低, 即注意力值偏低。  $\alpha \in [0, 1]$ , 此处设置为 0.9, 以得到边缘区域更高的注意力分数。然后将得到的注意力图  $BP_i$  与编码特征按通道逐像素相乘, 再将所得结果与原编码特征逐像素相加, 这有助于网络确定需要关注的特征区域。最后边缘增强特征与编码特征进行特征通道拼接, 采用一次卷积操作将通道数减半以减少参数量, 输出最终增强的边缘感知特征图。

### 1.3 自适应通道注意力

为了更好地融合编码器、解码器和边缘感知模块语义不一致的特征, 本文使用改进的自适应通道注意力模块 (ASCAM), 通过有效学习特征间的通道交互信息从而提高特征表达能力, 该模块具体结构如图 3 所示。

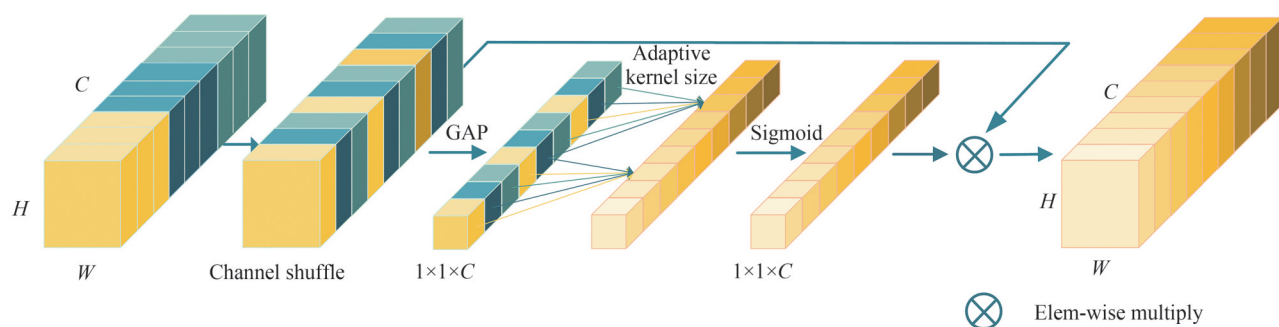


图3 自适应通道注意力模块结构

Fig.3 The structure of adaptive shuffle channel attention module

自适应通道注意力模块输入为第  $k$  层网络的编码器、解码器和边缘感知模块特征拼接得到的融合特征  $X^k \in \mathbb{R}^{H \times W \times C}$ 。首先对融合特征进行通道重排, 按顺序各取三个特征块的一个通道特征图依次排列组成新的特征块, 其目的是使后续卷积操作在不使用大卷积核的情况下能有效捕获不同特征块之间的信息交互。通道重排后的特征  $\mathbb{R}^{H \times W \times C}$  作为通道注意力处理的输入。之后空间压缩由一个全局平均池化 (Global Average Pooling, GAP) 进行, 产生特征向量  $G(x) \in \mathbb{R}^{1 \times 1 \times C}$

$$G(x) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X^k(i, j) \quad (2)$$

式中对输入特征  $X^k \in \mathbb{R}^{H \times W \times C}$  的  $C$  个通道的特征值分别求平均值,  $H, W$  分别为每个特征图的高度与宽度。

执行以上步骤后,通常采用一个多层感知器模块(Multiple Layer Perception, MLP)<sup>[24]</sup>计算通道权重系数。MLP主要由两个全连接层(Fully Connected Layer, FC)构成,其中首个FC的节点数通常利用降维操作降低模块的参数数量,第二个FC再将特征映射回特征空间。但这样的降维操作破坏了通道与权重参数之间一一对应的关系,进而影响最终的模型学习效果<sup>[25]</sup>。因此,本文提出的自适应通道注意力利用一维卷积捕获局部通道交互信息,显著降低通道注意力模块复杂度的同时兼顾了模块有效性。同时,为了最大程度地保证信息交互的通用性,需要确定相互作用的覆盖范围,即一维卷积核的尺寸。而网络不同层级的特征块中具有不同的通道数,需要通过人工调整优化卷积核大小。然而,通过手动调整核参数耗费大量计算资源,部分采用组卷积的改进模型证实,卷积操作的卷积核尺寸  $k$  与通道数量  $C$  之间存在正相关映射关系  $\varphi$ <sup>[26]</sup>,且网络中通道数量通常为2的指数,因此,将映射关系设置为如下所示的非线性映射关系

$$C = \varphi(k) = 2^{\gamma \times k - b} \quad (3)$$

因此,给定通道数  $C$ ,一维卷积核大小  $k$  可以自适应确定为

$$K = \varphi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (4)$$

式中,  $\lfloor t \rfloor_{\text{odd}}$  表示取  $t$  的绝对值相邻最近的奇数。本文中,  $\gamma$  的值设置为1,  $b$  设置为0。显然,通过非线性映射关系,高维通道的相互作用距离较长,而低维通道的相互作用距离较短,经计算自上往下的四个ASCAM模块自适应卷积核数量分别为1、3、5、7。在此基础上,利用Sigmoid函数进行激活得到通道注意力权值系数  $\beta \in [0, 1]^{1 \times 1 \times C}$ 。随后,将获得的通道注意力系数  $\beta$  与原始的输入映射  $X^k \in \mathbb{R}^{H \times W \times C}$  逐通道相乘,得到经过权值优化的层注意力融合特征。该模块可以有效地学习特征间的通道交互信息提高特征表达能力,与其他现有通道注意力相比,该方法以较低的模型复杂度获得了较好的性能。

#### 1.4 联合损失函数

本文提出边缘与结构联合损失函数,以优化网络的目标定位和细节处理能力。该联合损失函数包括三部分:分割损失  $\mathcal{L}_1$ , 层级监督损失  $\mathcal{L}_{\text{ds}}$  和边缘惩罚损失  $\mathcal{L}_{\text{edge}}$ 。其中,分割损失  $\mathcal{L}_1$  计算网络最终的分割预测结果与对应标签 Mask 之间的损失,在保证网络鲁棒性的同时有效提升训练速度,提高目标分割精度;层级监督损失  $\mathcal{L}_{\text{ds}}$  计算不同深度的解码特征进行激活输出后的预测值与对应标签 Mask 之间的损失,从解码器的不同尺度优化隐藏层特征以获得更精确的分割结果;边缘惩罚损失  $\mathcal{L}_{\text{edge}}$  计算网络最终的分割预测结果边缘与对应标签 Mask 边缘之间的损失(预测图像与标签的边缘均有 Sobel 算子提取),进而有效利用目标的边缘信息,约束网络预测的边缘精度,提升对边缘的分割效果。该联合损失函数  $\mathcal{L}_{\text{all}}$  表示为

$$\mathcal{L}_{\text{all}} = \alpha \mathcal{L}_1 + \beta \mathcal{L}_{\text{ds}} + \gamma \mathcal{L}_{\text{edge}} \quad (5)$$

式中,  $\alpha, \beta, \gamma$  分别为损失的权重系数,  $\alpha$  设置为2,  $\beta$  与  $\gamma$  均设置为1。

$\mathcal{L}_1$  是二元交叉熵损失(Binary Cross Entropy, BCE)与Dice损失的加权结果:  $\mathcal{L}_1 = \mathcal{L}_{\text{BCE}} + \mathcal{L}_{\text{Dice}}$ 。

$\mathcal{L}_{\text{BCE}}$  鲁棒性较好,但易受到医学图像类别不均衡的影响,定义为

$$\mathcal{L}_{\text{BCE}} = - \sum_i (y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)) \quad (6)$$

式中,  $y_i$  表示第  $i$  个像素的标签值,  $\hat{y}_i$  表示第  $i$  个像素的预测值。

$\mathcal{L}_{\text{Dice}}$  更关注目标区域,可以有效提升训练速度,提高目标分割效果,但稳定性较差,其表达式为

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2|y \cap \hat{y}|}{|y| + |\hat{y}|} \quad (7)$$

式中,  $y$  表示真实标注,  $\hat{y}$  表示预测结果。

$\mathcal{L}_{\text{ds}}$  从解码器不同尺度优化隐藏层特征以获得更精确的分割结果,属于深监督损失,每一层的  $\mathcal{L}_{\text{ds}}$  与  $\mathcal{L}_1$  计算方式相同,均为二元交叉熵损失与Dice损失的加权,定义为

$$\mathcal{L}_{ds} = \sum_{i=1}^4 \mathcal{L}_{ds_i}(\mathbf{y}_{down_i}, \mathbf{O}_i) = \sum_{i=1}^4 (\mathcal{L}_{BCE}(\mathbf{y}_{down_i}, \mathbf{O}_i) + \mathcal{L}_{Dice}(\mathbf{y}_{down_i}, \mathbf{O}_i)) \quad (8)$$

式中,  $\mathbf{y}_{down_i}$  和  $\mathbf{O}_i$  分别表示第  $i$  层下采样标签矩阵与第  $i$  层网络预测矩阵。

除了约束目标区域的  $\mathcal{L}_1$  与  $\mathcal{L}_{ds}$  损失外, 本文引入了增强边缘精度的  $\mathcal{L}_{edge}$  损失, 在网络训练过程中实时优化分割边缘, 该损失定义为预测结果边缘与标签边缘的二元交叉熵计算结果, 具体可表示为

$$\mathcal{L}_{edge} = \mathcal{L}_{BCE}(\mathbf{y}_b, \hat{\mathbf{y}}_b) \quad (9)$$

式中,  $\mathbf{y}_b$  表示标签结果的边缘,  $\hat{\mathbf{y}}_b$  表示预测结果的边缘。

## 2 实验结果及分析

在整个实验中, 实验硬件环境是基于 PANYAO 7048GR 服务器, CPU 为正式版 E5-2096 2.3G-3.6G 45M Cache22NM, 内存为 256G, 六张 NVIDIA GeForce TITAN RTX 显卡。软件环境是基于 Ubuntu 系统的 Tensorflow2.0 深度学习框架, CUDA 和 CUDNN 的版本分别为 10.1.243 和 7.6.5, 编程语言为 Python3.8。

### 2.1 实验数据集

1) MICCAI 2015 腺体分割挑战赛数据集 (GlaS)<sup>[27]</sup>。

GlaS 数据集病理图像采用苏木精和伊红 (Hematoxylin Eosin, H&E) 染色, 并使用蔡司 MIRAX MIDI 扫描仪对结直肠癌组织进行 20 倍放大成像, 原始图像大小不一, 大多数为  $775 \times 522$  大小的 RGB 图像。共包含 85 张训练数据以及 80 张测试数据。

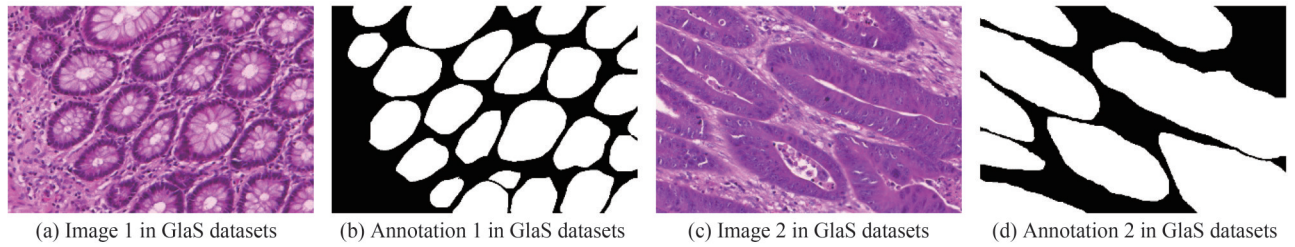


图4 GlaS 病理图像数据集  
Fig.4 The GlaS pathological image datasets

2) MoNuSeg 细胞核数据集<sup>[28-29]</sup>。

MoNuSeg 数据集由 40 倍放大下捕获的 H&E 染色组织图像形成。该数据集包含患者和多个器官的图像, 原始尺寸均为  $1000 \times 1000$ 。训练数据包含 30 张图像, 约 22 000 个细胞核核边界注释。测试数据包含 14 张图像, 其中核边界标注超过 7 000 个。

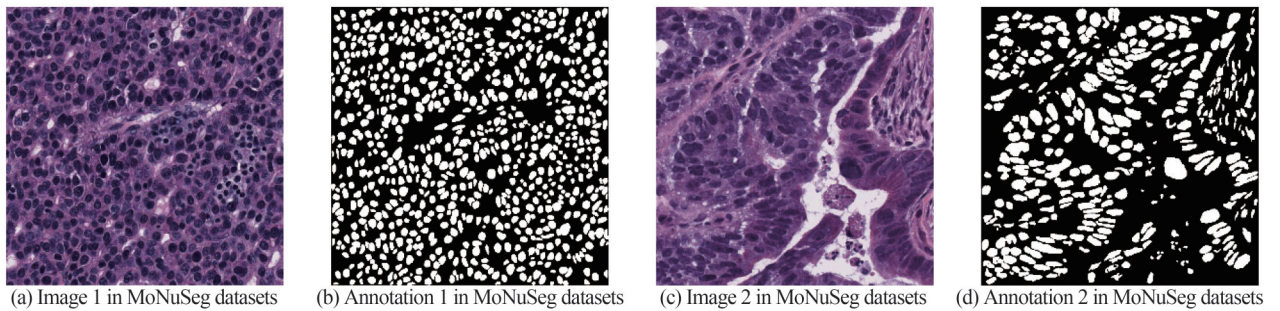


图5 MoNuSeg 细胞核数据集  
Fig.5 The MoNuSeg nuclei datasets

### 2.2 实验设置

两个实验数据集将训练数据按照训练集与测试集 4:1 划分。同时为了弥补训练数据不足造成过拟合, 实验中对训练集数据进行了水平翻转、垂直翻转两种在线数据增强方式。为尽量保留训练图像细节和减少模型计算量, 训练 BPNet 网络时两个数据集中图像均缩放为  $512 \times 512$  尺寸作为输入。模型训练采用随机梯

度下降(Stochastic Gradient Descent,SGD)作为模型优化器。训练初始学习率为0.01,训练 epoch 总数设置为200,当验证精度超过15个 epoch 不增加时,则将学习率更改为当前学习率的0.5,最小学习率为0.0005,当经过连续30个 epoch 验证精度没有提高时结束训练,批次大小统一设置为6。实验采用1.4节描述的边缘与结构联合损失函数,为保证实验可靠性,实验最终结果取5次重复实验平均值。

对于两个实验数据集,均采用Dice系数、交并比系数(Intersection over Union, IoU)、准确率(Accuracy, ACC)和精确率(Precision, PRE)四个评价指标。

$$\text{Dice} = \frac{2|y \cap \hat{y}|}{|y| + |\hat{y}|} \quad (10)$$

$$\text{IoU} = \frac{|y \cap \hat{y}|}{|y \cup \hat{y}|} \quad (11)$$

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (12)$$

$$\text{PRE} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (13)$$

式中, $y$ 表示标签, $\hat{y}$ 表示预测结果,TP、FP、TN、FN分别表示为真阳性、假阳性、真阴性、假阴性的对应像素预测结果,其中最终的预测结果为模型预测结果以0.5为阈值进行二值化后得出。

### 2.3 GlaS数据集实验结果

#### 2.3.1 BPNet算法和经典算法对比

为验证BPNet网络整体分割性能,将其与其他先进的方法进行对比。对比方法包括两种类型,四种基于CNN的方法:U-Net<sup>[8]</sup>、UNet++<sup>[9]</sup>、MultiResUNet<sup>[10]</sup>和 AttentionUNet<sup>[11]</sup>,三种基于Transformer的方法:TransUNet<sup>[20]</sup>、MedT<sup>[21]</sup>和 UCTransNet<sup>[22]</sup>。其中,MedT与 UCTransNet模型输入大小按照原始公布代码设置为 $224 \times 224$ ,其余网络模型输入设置为 $512 \times 512$ 。为了更加公平的比较,上述所有方法的其余设置均采用2.2节所提的实验设置,不同算法在GlaS数据集上的分割结果见表1,最佳分割结果采用加粗显示。

表1 不同算法在GlaS数据集上的实验结果(平均值±标准差)

Table 1 The experimental results with different methods on GlaS datasets (Mean ± Std)

Algorithm	Dice/%	IoU/%	ACC/%	PRE/%	Parameters/( $\times 10^6$ )
U-Net(2015)	83.40±3.26	71.84±5.86	83.04±3.83	79.99±6.51	31.0
UNet++(2018)	85.11±0.73	74.10±1.10	85.48±0.41	85.83±0.97	39.4
AttentionUNet(2018)	86.30±2.08	75.99±4.00	86.39±2.94	83.99±5.82	63.1
MultiResUNet(2020)	87.20±0.99	77.38±1.55	86.98±0.69	86.29±0.85	59.1
MedT(2021)	82.86±0.97	70.76±1.41	82.46±0.78	83.93±0.99	<b>7.0</b>
TransUNet(2021)	88.67±0.65	79.66±1.06	88.46±0.69	90.00±0.78	421.2
UCTransNet(2022)	89.39±0.68	80.83±1.11	87.93±0.93	87.00±1.65	65.5
<b>BPNet</b>	<b>92.21±0.19</b>	<b>85.55±0.33</b>	<b>92.14±0.15</b>	<b>92.07±0.93</b>	64.9

\* Parameters表示网络模型参数量

由表1可知,BPNet方法与现有方法相比在四个指标上具有明显优势。其中,U-Net作为经典的医学图像分割网络,提出使用跳跃连接融合细节特征,但其较弱的特征提取能力以及简单的特征融合方式导致分割性能不足。UNet++和MultiResUNet分别在UNet的基础上改进跳跃连接方式、增强特征融合使得分割效果有提升。而基于Transformer结构的TransUNet、UCTransNet网络通过利用自注意力机制编码远程依赖关系,在精度上相较于基于U-Net的模型精度有较大提升,但其参数量巨大,且网络训练拟合速度慢,在多个指标上仍略低于BPNet。MedT则由于其复杂的特征提取过程和训练策略,在数据量偏少的数据集上难以取得优势。BPNet因其加强了对于病理图像复杂区域的特征提取,取得最好的图像分割性能。

为了直观对比不同方法的预测结果,将模型分割结果进行可视化,如图6所示。可以看出,U-Net及其改进网络特征信息提取不够完全,导致存在欠分割的情况。基于Transformer系列网络的分割结果相较于U-Net系列网络有明显改善,但对于一部分相邻的病灶区域,其分割结果依旧存在粘连情况,说明上述基于Transformer的分割网络并没有针对边缘特征进行单独的网络改进。而BPNet拥有最佳的分割效果,尤其是在分割边缘保留了更多的细节信息,证明所提的边缘感知模块优化了网络对于边缘的分割效果。

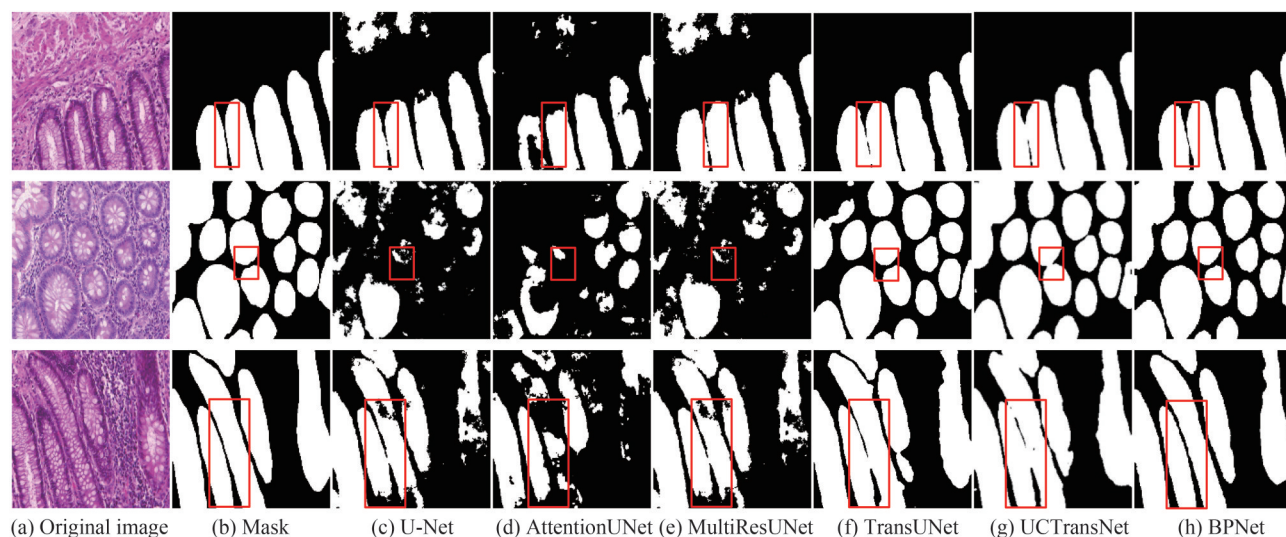


图6 不同算法在GlaS数据集上的分割结果

Fig.6 The segmentation results of different algorithms on GlaS datasets

### 2.3.2 算法消融实验

在此基础上,进行了消融实验,具体设置为:1) Baseline是以EfficientNet-B4网络作为编码器的U-Net,并使用ImageNet数据集预训练模型,训练过程中采用联合损失函数。2)在Baseline的基础上加入边缘感知模块,命名为Baseline+BPM。3)在Baseline+BPM的基础上增加自适应通道注意力模块,注意力模块的使用位置在各解码层对应的特征融合位置,实验中命名为Baseline+BPM+ASCAM,即BPNet。消融实验中所有模型训练设置与BPNet算法相同,取五次重复实验结果均值作为各模型得分,实验结果如表2所示。

表2 GlaS数据集上的消融实验结果(平均值±标准差)  
Table 2 The ablation experimental results on GlaS datasets (Mean ± Std)

Algorithm	Dice/%	IoU/%	ACC/%	PRE/%	Parameters/( $\times 10^6$ )
Baseline	87.09±1.97	78.10±3.33	88.65±2.16	88.54±3.10	63.3
Baseline+BPM	91.79±0.09	84.81±0.16	91.80±0.07	<b>92.87±0.30</b>	64.8
Baseline+BPM+ASCAM	<b>92.21±0.19</b>	<b>85.55±0.33</b>	<b>92.14±0.15</b>	92.07±0.93	64.9

由表2可知,Baseline在原始U-Net基础上更换有效编码器及使用联合损失函数使得网络的性能有一定的提升。同时,添加边缘感知模块后模型的预测分割精度提升明显,说明边缘感知模块有效地增强了网络对于边缘细节的预测。而自适应感知通道模块仅增加极少的参数量使得网络的分割结果进一步提升。

为体现各模块对分割结果的作用,图7展示了相应的消融实验实际分割效果。可以看出,在基础模型添加BPM后,相邻分割目标的粘连现象有明显改善,说明该模块加强了模型对于边缘的区分能力,但还存在定位不准确,分割目标缺失的问题。在上述模型添加ASCAM后,分割结果进一步完善,在保留精细边缘的同时准确定位目标。



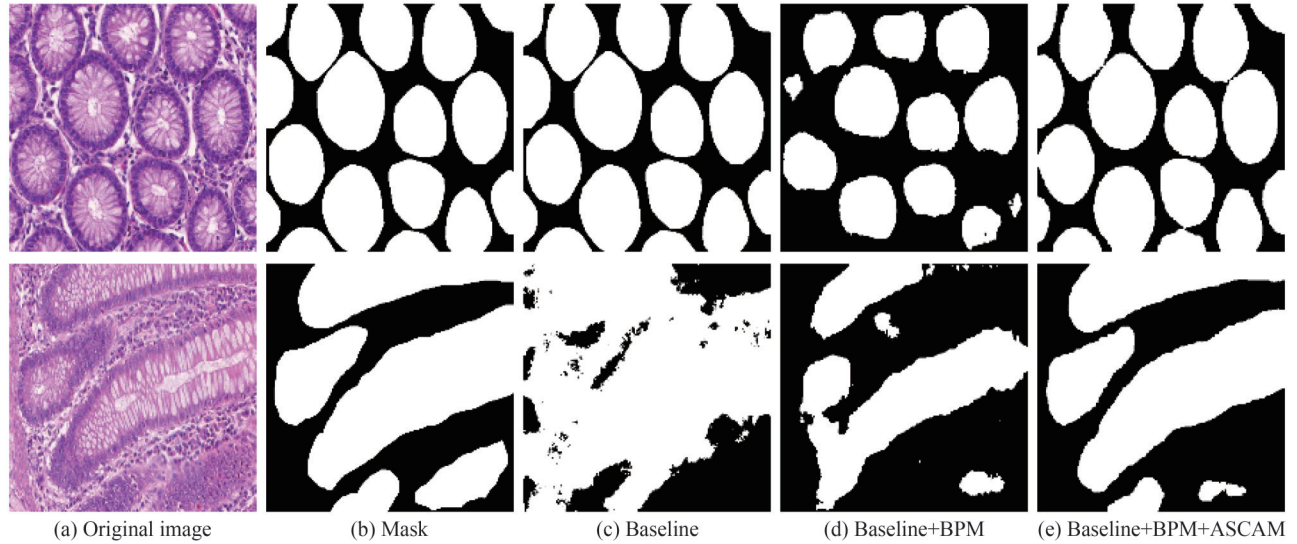


图7 GlaS数据集上的消融实验结果

Fig.7 The segmentation results of ablation experiments on GlaS datasets

## 2.4 MoNuSeg数据集实验结果

### 2.4.1 BPNet算法和经典算法对比

为了测试BPNet模型在不同数据集上的泛化性,在MoNuSeg数据集上进行了对比实验,实验设置同2.3.1节,实验结果见表3。

表3 不同算法在MoNuSeg数据集上的实验结果(平均值±标准差)

Table 3 The experimental results with different methods on MoNuSeg datasets (Mean ± Std)

Algorithm	Dice/%	IoU/%	ACC/%	PRE/%	Parameters/( $\times 10^6$ )
U-Net(2015)	77.17±2.84	62.90±3.79	89.90±1.62	65.01±3.68	31.0
UNet++(2018)	78.90±0.57	65.16±0.78	90.98±0.33	67.42±1.32	39.4
AttentionUNet(2018)	75.00±1.86	60.03±2.38	88.40±1.42	60.64±3.39	63.1
MultiResUNet(2020)	79.70±0.61	66.26±0.84	91.52±0.33	69.19±1.61	59.1
MedT(2021)	74.48±0.61	59.34±0.77	89.15±0.16	63.61±3.30	<b>7.0</b>
TransUNet(2021)	80.50±0.39	67.37±0.54	91.40±0.24	73.42±1.48	421.2
UCTransNet(2022)	79.08±0.79	65.58±1.00	91.02±0.53	70.38±1.13	65.5
<b>BPNet</b>	<b>81.18±0.44</b>	<b>68.34±0.65</b>	<b>92.50±0.24</b>	<b>75.46±2.15</b>	64.9

在表3中,UNet++和MultiResUNet相较于U-Net分别提升了约1.6%和2.5%的Dice系数,提升幅度小于GlaS数据集,这是由于MoNuSeg数据集的分割目标更加聚集复杂。而AttentionUNet的注意力模块并没有使网络在复杂的数据下获得最佳的融合特征,反而使得网络参数量增加导致过拟合,测试集得分相较于U-Net有所下降。基于Transformer的TransUNet和UCTransNet在得分上相较于基于U-Net的网络有明显提升,这是由于细胞核在整幅图像有重复性以及相似性,Transformer能够有效提取不同区域相似目标之间的关联特征,使分割效果提升,但MedT由于网络仅采用自注意力机制进行特征提取,在数据量偏少的情况下网络不能有效拟合,导致得分明显偏低。本文提出的BPNet在四个指标上均取得了最高得分,说明网络有效提取了复杂病理图像的细节信息。

各算法在测试集上的预测结果如图8所示。在图8中,各算法均出现了不同程度的过分割,主要体现在相邻细胞核的边界存在粘连,这种情况会极大地影响临床中后续的细胞计数。相较于对比算法,BPNet优化了边界的特征提取以及在训练过程中加入对边界效果的评价,使其取得了最好的分割效果。

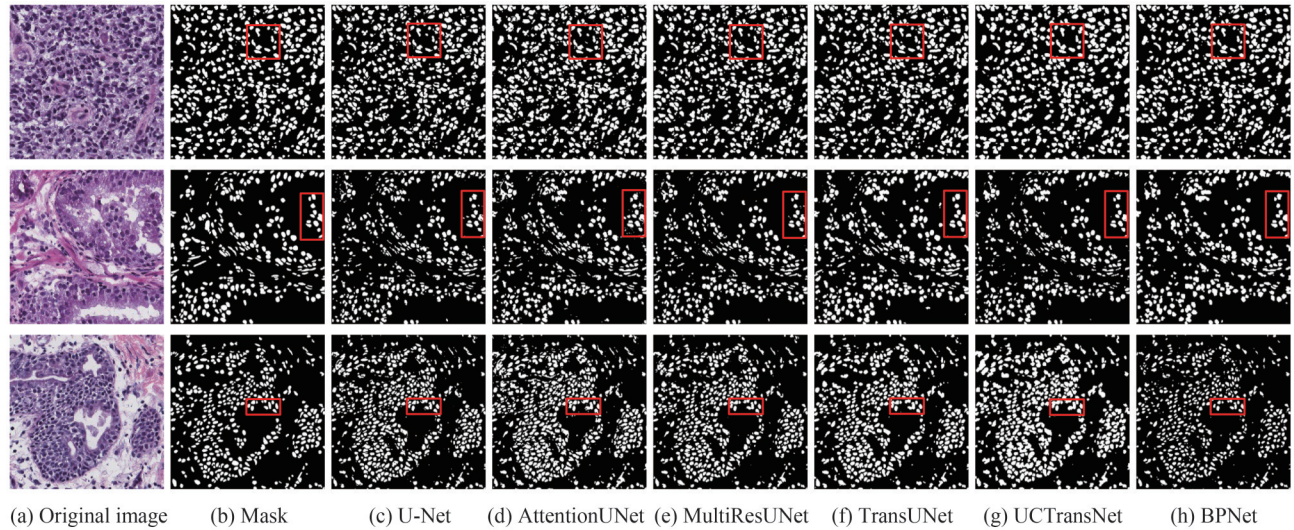


图8 不同算法在 MoNuSeg 数据集上的分割结果

Fig.8 The segmentation results of different algorithms on MoNuSeg datasets

#### 2.4.2 算法消融实验

在 MoNuSeg 数据集上的消融实验设置与 2.3.2 节一致,实验结果见表 4。

表 4 MoNuSeg 数据集上的消融实验结果(平均值±标准差)

Table 4 The ablation experimental results on MoNuSeg datasets (Mean ± Std)

Algorithm	Dice/%	IoU/%	ACC/%	PRE/%	Parameters/( $\times 10^6$ )
Baseline	77.74±0.73	63.59±0.98	91.47±0.06	73.38±2.74	63.3
Baseline+BPM	79.96±0.46	66.61±0.64	92.02±0.36	73.97±4.21	64.8
Baseline+BPM+ASCAM	<b>81.18±0.44</b>	<b>68.34±0.65</b>	<b>92.50±0.24</b>	<b>75.46±2.15</b>	64.9

从表 4 可以看出, Baseline 逐渐添加 BPM 以及 ASCAM 后各指标有了明显提升,说明采用边缘感知融合以及自适应通道注意力都能提高网络的病灶分割性能。结合表 3 的实验结果,可以发现添加 BPM 模块能够使目标的边界位置信息深度参与到网络训练的优化过程中,对于模型的分割性能提升尤为明显,这体现了边缘特征对于提高密集目标分割性能的重要性;而 ASCAM 模块在保证不增加网络模型复杂度的前提下,自适应的捕捉了编码、边缘、解码三部分特征信息的交互关系,弥补了不同层次特征间的语义差距,有效捕获了更复杂的通道相关性,进一步提升了分割精度。

### 3 结论

本文提出了一种改进的病理图像语义分割算法 BPNet。该算法首先利用解码特征图生成边缘增强特征,加强了网络对于实例多且分布广的病理图像的分割性能,并采用自适应通道注意力模块筛选重要特征,最后在输出层利用联合损失函数实现网络对于不同层级特征的信息捕获,使网络性能达到最优。在 GlaS 和 MoNuSeg 病理图像数据集上的实验结果表明,本文所提方法优于其他方法,获得了最好的分割效果,其 Dice 分割精度分别可达 92.21% 和 81.18%,尤其在边缘预测处理上具有明显的优势。此外,基于 Transformer 的网络能够有效提取图像不同区域相似目标之间的长距离依赖关系,提升分割性能。因此下一步将联合 Transformer 模型,在细化分割边缘的同时进一步加强网络的特征提取能力。

#### 参考文献

- [1] LITJENS G, KOOI T, BEJNORDI B E, et al. A survey on deep learning in medical image analysis[J]. Medical Image Analysis, 2017, 42: 60-88.
- [2] PRABHUA S, PRASADA K, ROBELS-KELLY A, et al. AI-based carcinoma detection and classification using histopathological images: A systematic review[J]. Computers in Biology and Medicine, 2022, 144: 105209.
- [3] RAMESH K, KUMAR G K, SWAPNA K, et al. A review of medical image segmentation algorithms[J]. EAI Endorsed

- Transactions on Pervasive Health and Technology, 2021, 7(27): e6.
- [4] OTSU N. A threshold selection method from gray-level histograms [J]. IEEE Transactions on Systems Man & Cybernetics, 2007, 9(1): 62-66.
- [5] VINCENT L, SOILLE P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(6): 583-598.
- [6] BOYKOV Y Y, JOLLY M P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images [C]. Proceedings of the Eighth IEEE International Conference on Computer Vision, Vancouver, British Columbia, Canada: ICCV, 2001: 105-112.
- [7] HUANG Xiaoming, HE Fuyun, TANG Xiaohu, et al. Review on applications of U-Net and its variants in medical image segmentation[J]. Chinese Journal of Biomedical Engineering, 2022, 41(5): 567-576.  
黄晓鸣,何富运,唐晓虎,等. U-Net及其变体在医学图像分割中的应用研究综述[J]. 中国生物医学工程学报, 2022, 41(5): 567-576.
- [8] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany: MICCAI, 2015: 234-241.
- [9] ZHOU Z, SIDDIQUEE M M R, TAJBAKHS N, et al. UNet++ : a nested U-Net architecture for medical image segmentation [C]. Proceedings of the 4th Deep Learning in Medical Image Analysis and multimodal learning for clinical decision support, Granada, Spain: DLMIA, 2018: 3-11.
- [10] IBTEHAZ N, RAHMAN M S. MultiResUNet: Rethinking the u-net architecture for multimodal biomedical image segmentation[J]. Neural Networks, 2020, 121: 74-87.
- [11] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-net: learning where to look for the pancreas [C]. 1st Conference on Medical Imaging with Deep Learning, Amsterdam, Netherlands: MIDL, 2018.
- [12] WANG J S, WANG Y, TAO X, et al. PCA-U-Net based breast cancer nest segmentation from microarray hyperspectral images[J]. Fundamental Research, 2021, 1(5): 631-640.
- [13] HUANG Hong, WANG Tao, LI Yuan, et al. Cancer pathological segmentation network based on depth feature fusion [J]. Acta Photonica Sinica, 2022, 51(3): 0310001.  
黄鸿,王涛,李远,等. 基于深度特征融合的癌症病理图像分割网络[J]. 光子学报, 2022, 51(3): 0310001.
- [14] ZHAO B C, CHEN X, LI Z, et al. Triple U-net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation[J]. Medical Image Analysis, 2020, 65: 101786.
- [15] ZAMORA W, MENDEZ M, CALDERON-RAMIREZ S, et al. Enforcing morphological information in fully convolutional networks to improve cell instance segmentation in fluorescence microscopy images [C]. Proceedings of the 16th International Work-Conference on Artificial Neural Networks, 2021, 1: 36-46.
- [16] FANG Y, CHEN C, YUAN Y, et al. Selective feature aggregation network with area-boundary constraints for polyp segmentation [C]. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China: MICCAI, 2019: 302-310.
- [17] FAN D P, JI G P, ZHOU T, et al. PraNet: parallel reverse attention network for polyp segmentation [C]. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru: MICCAI, 2020: 263-273
- [18] CHEN S, DING C, LIU M, et al. CPP-Net: context-aware polygon proposal network for nucleus segmentation [J]. IEEE Transactions on Image Processing, 2023, 32: 980-994.
- [19] ZHANG R, LI G, LI Z, et al. Adaptive context selection for polyp segmentation [C]. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru: MICCAI, 2020: 253-262.
- [20] CHEN J, LU Y, YU Q, et al. Transunet: transformers make strong encoders for medical image segmentation [J]. 2021, arXiv preprint arXiv: 2102.04306.
- [21] VALANARASU J, OZA P, HACIHSILOGLU I, et al. Medical transformer: gated axial-attention for medical image segmentation [C]. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Strasbourg, France: MICCAI, 2021: 36-46.
- [22] WANG H, CAO P, WANG J, et al. UCTransNet: rethinking the skip connections in u-net from a channel-wise perspective with transformer [C]. Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, Canada: AAAI, 2022, 36(3): 2441-2449.
- [23] TAN M X, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks [C]. Proceedings of the 36th International Conference on Machine Learning, California, UT, USA, ACML, 2019: 691-700.
- [24] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [25] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks,

- Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, W A, USA, CVPR, 2020: 11531-11539.
- [26] YANG Jun, LI Bozan. Semantic segmentation of 3D point cloud based on self-attention feature fusion group convolutional neural network[J]. Optics and Precision Engineering, 2022, 30(7): 840-853.  
杨军,李博赞.基于自注意力特征融合组卷积神经网络的三维点云语义分割[J].光学精密工程,2022,30(7):840-853.
- [27] SIRINUKUNWATTANA K, PLUIM J P, CHEN H, et al. Gland segmentation in colon histology images: The glas challenge contest[J]. Medical Image Analysis, 2021, 35: 489-502.
- [28] KUMAR N, VERMA R, SHARMA S, et al. A dataset and a technique for generalized nuclear segmentation for computational pathology[J]. IEEE Transactions on Medical Imaging, 2017, 36(7): 1550-1560.
- [29] KUMAR N, VERMA R, ANAND D, et al. A multi-organ nucleus segmentation challenge[J]. IEEE Transactions on Medical Imaging, 2019, 39(5): 1380-1391.

## Boundary Perception Network for Pathological Image Segmentation

HUANG Hong<sup>1</sup>, YANG Yichuan<sup>1</sup>, WANG Long<sup>1</sup>, ZHENG Fujian<sup>1</sup>, WU Jian<sup>2</sup>

(1 Key Laboratory of Optoelectronic Technology and System, Ministry of Education, Chongqing University, Chongqing 400044, China)

(2 Head and Neck Cancer Centre, Chongqing University Cancer Hospital & Chongqing Cancer Institute & Chongqing Cancer Hospital, Chongqing 400030, China)

**Abstract:** In clinical practice, segmentation and quantitative evaluation of target objects in pathological images provide valuable information for histopathological analysis, which is of great significance to auxiliary diagnosis and subsequent treatment. However, due to the dense distribution of cells and great morphological similarities between the cancer cells and normal cells, there are some challenges such as difficulty in feature extraction and unclear segmentation boundaries in the segmentation task of pathological images. At the same time, the traditional image segmentation methods are time-consuming and labor-intensive. They can only extract low level manual features, and the expression ability of deep discrimination features is insufficient, resulting in limited performance of traditional methods. Meanwhile, previous deep learning algorithms still suffer from two significant problems. Firstly, most networks ignore pixels that are difficult to segment, such as the boundaries of targets, which is particularly important for accurate segmentation. In addition, the problem of inconsistent semantic levels between different features are not solved, leading to low training efficiency. To address the above-mentioned problems, an end-to-end histopathological image segmentation network called Boundary Perception Network (BPNet) is proposed for improving the segmentation accuracy of histopathological images. Based on encoder-decoder structure, the encoder performs the convolutional downsampling operation to extract the feature information of the image through the Convolutional Neural Network (CNN). And the encoding process uses the feature encoder based on the EfficientNet-B4 network which is specifically used for pathological image segmentation. The decoder mainly consists of decoding blocks, Boundary Perception Module (BPM) and Adaptive Shuffle Channel Attention Module (ASCAM). In detail, the decoding block performs deconvolution operation to complete the decoding process of the feature information. Then, the BPM in the decoder stage aims to strengthen the ability of mining for difficult segmentation regions, so that the network focuses on the higher uncertainty as well as more complex edge regions, achieving feature complementarity and precision prediction results. For implementation, the BPM extracts the edge from the decoder output of each layer, and superimposes the edge information onto the encoded feature to strengthen the boundary feature information extracted from pathological images, outputting the enhanced edge perception feature map. Subsequently, the ASCAM is an improved channel attention module which is used to make up the semantic gap between different levels of features, extracted by encoder, decoder and BPM, so as to further strengthens the feature understanding ability of the BPNet. This module exploits adaptive kernel size one-dimensional convolution to capture the interactive information of local channels, at the same time ensures the efficiency and effectiveness of the training process. The obtained channel attention

coefficient is multiplied by the module input feature layer to obtain the fusion feature, helping effectively learn the channel interaction information between features to improve the feature representation ability. Furthermore, a joint loss function based on structure and boundary is designed to optimize the targeting and detail processing capabilities of this method, achieving the better segmentation result of pathological images. Experiments are carried out on the Gland segmentation (GlaS) and MoNuSeg dataset, respectively. Both of the two datasets are divided into 4:1 for training and validation. At the same time, in order to make up for the overfitting caused by the lack of training data, two kinds of online data enhancement methods of horizontal flipping and vertical flipping were carried out on the training set data in the experiment. And the four evaluation index, the Dice coefficient score, Intersection Over Union (IoU), Accuracy (ACC) and Precision (PRE), are used to evaluate the performance of this method proposed in this paper. The Dice coefficient score of the proposed method is 92.21% and 81.18%, the IoU is 85.55% and 68.34%, the ACC is 92.14% and 92.50%, the PRE is 92.07% and 75.46% on the GlaS and MoNuSeg datasets, respectively. Compared with the previous classical methods, such as U-Net, UNet++, MultiResUNet, TransUNet, UCTransNet and so on, the BpNet proposed gets the best segmentation result, especially retains more details in the segmentation boundary. Moreover, ablation experiments are carried out on the same two datasets for indicating the impacts of BPM and ASCAM. The results shows that the proposed BPM significantly optimizes the segmentation effect of the network for the edge, as well as the ASCAM makes up the semantic gap between features at different levels and further strengthens the feature understanding ability of the network. In conclusion, the BpNet proposed in this paper exploits BPM to generate edge enhancement feature maps, and uses ASCAM to seize crucial features. Finally, a joint loss function is used to capture the information of features at different levels in the output layer to achieve optimal segmentation performance. The experimental results have demonstrated that the effectiveness of each part of proposed method in the segmentation task of pathological images.

**Key words:** Pathological image; Automatic segmentation; Edge enhancement; Joint loss function; Deep learning

**OCIS Codes:** 100.2000; 170.4580; 100.4994; 100.4996