

显著性引导及不确定性监督的深度编解码网络^{*}

王 雪^{1,2}, 李占山^{1,2}, 陈海鹏^{1,2}



¹(吉林大学 计算机科学与技术学院, 吉林 长春 130012)

²(符号计算与知识工程教育部重点实验室(吉林大学), 吉林 长春 130012)

通信作者: 陈海鹏, E-mail: chenhp@jlu.edu.cn

摘要: 基于 U-Net 的编码-解码网络及其变体网络在医学图像语义分割任务中取得了卓越的分割性能。然而, 网络在特征提取过程中丢失了部分空间细节信息, 影响了分割精度。另一方面, 在多模态的医学图像语义分割任务中, 这些模型的泛化能力和鲁棒性不理想。针对以上问题, 提出一种显著性引导及不确定性监督的深度卷积编解码网络, 以解决多模态医学图像语义分割问题。该算法将初始生成的显著图和不确定概率图作为监督信息来优化语义分割网络的参数。首先, 通过显著性检测网络生成显著图, 初步定位图像中的目标区域; 然后, 根据显著图计算不确定分类的像素点集合, 生成不确定概率图; 最后, 将显著图和不确定概率图与原图像一同送入多尺度特征融合网络, 引导网络关注目标区域特征的学习, 同时增强网络对不确定分类区域和复杂边界的表征能力, 以提升网络的分割性能。实验结果表明, 所提算法能够捕获更多的语义信息, 在多模态医学图像语义分割任务中优于其他的语义分割算法, 并具有较好的泛化能力和鲁棒性。

关键词: 编码-解码网络; 显著图; 不确定概率图; 医学图像语义分割; 多模态

中图法分类号: TP18

中文引用格式: 王雪, 李占山, 陈海鹏. 显著性引导及不确定性监督的深度编解码网络. 软件学报, 2022, 33(9): 3165–3179. <http://www.jos.org.cn/1000-9825/6624.htm>

英文引用格式: Wang X, Li ZS, Chen HP. Deep Encoder-decoder Network with Saliency Guidance and Uncertainty Supervision. Ruan Jian Xue Bao/Journal of Software, 2022, 33(9): 3165–3179 (in Chinese). <http://www.jos.org.cn/1000-9825/6624.htm>

Deep Encoder-decoder Network with Saliency Guidance and Uncertainty Supervision

WANG Xue^{1,2}, LI Zhan-Shan^{1,2}, CHEN Hai-Peng^{1,2}

¹(College of Computer Science and Technology, Jilin University, Changchun 130012, China)

²(Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education (Jilin University), Changchun 130012, China)

Abstract: The encoder-decoder network based on U-Net and its variants have achieved excellent performance in semantic segmentation of medical images. However, some spatial details are lost during feature extraction, which affects the accuracy of segmentation, and the generalization ability and robustness of these models are unsatisfactory. Therefore, this study proposes a deep convolutional encoder-decoder network with saliency guidance and uncertainty supervision to solve the semantic segmentation problem in multimodal medical images. In this method, the initially generated saliency map and the uncertainty probability map are used as the supervised information to optimize the parameters of the semantic segmentation network. Specifically, the saliency map is generated by the saliency detection network to preliminarily locate the target region in an image, and on this basis, the set of pixel points with uncertain classification is calculated to generate the uncertainty probability map. Then, the two maps are sent into the multi-scale feature fusion network together with the original image to guide the network to focus on the learning of the features in the target region and to enhance the

* 基金项目: 国家重点研发计划(2018YFB0804202, 2018YFB0804203); 国家自然科学基金区域联合基金(U19A2057); 国家自然科学基金面上项目(61876070); 吉林省科技发展计划(20190303134SF)

本文由“融合媒体环境下的媒体内容分析与信息服务技术”专题特约编辑汪萌教授、张勇东教授、俞俊教授以及张伟高级工程师推荐。

收稿时间: 2021-06-30; 修改时间: 2021-08-15; 采用时间: 2022-01-14; jos 在线出版时间: 2022-02-22

representational capacity of regions with uncertain classification and complex boundaries. In this way, the segmentation performance of the network can be improved. The experimental results reveal that the proposed method can capture more semantic information and outperforms existing semantic segmentation methods in semantic segmentation of multimodal medical images, with strong generalization capability and robustness.

Key words: encoder-decoder network; saliency map; uncertainty probability map; semantic segmentation of medical image; multimodal

随着人工智能等新一代信息技术的发展,现代医学和信息技术的高度融合,促进了医学影像成像设备的发展,进而涌现了大量不同模态的影像数据。如何运用视觉智能算法对这些数据进行智能分析,辅助医生临床诊断和治疗,是计算机视觉领域的研究热点之一。医学图像语义分割是计算机辅助诊断的关键技术。作为医学图像分析理解的基础,医学图像语义分割被广泛应用于各种智能诊断任务中,如数字病理图像诊断和分析^[1]、皮肤镜下的皮肤癌诊断^[2,3],以及结肠直肠癌诊断^[4]等。

近年来,深度学习技术在计算机视觉领域的各应用中取得了显著的效果,如语义分割、目标检测、图像分类等。其中,全卷积网络(fully convolutional network, FCN)^[5]在分割中展现出巨大的潜力,为语义分割提供了新的思路和方向,并被广泛应用于医学图像语义分割任务中。然而,由于医学图像的多模态性及复杂性,不同模态下的语义分割仍然面临巨大的挑战。如图1所示,在不同模态下的分割任务中,因目标区域尺度变化大、形状不规则、边界模糊、对比度低,以及亮度不均等多种复杂情况,难以对病灶和组织区域进行精准分割。

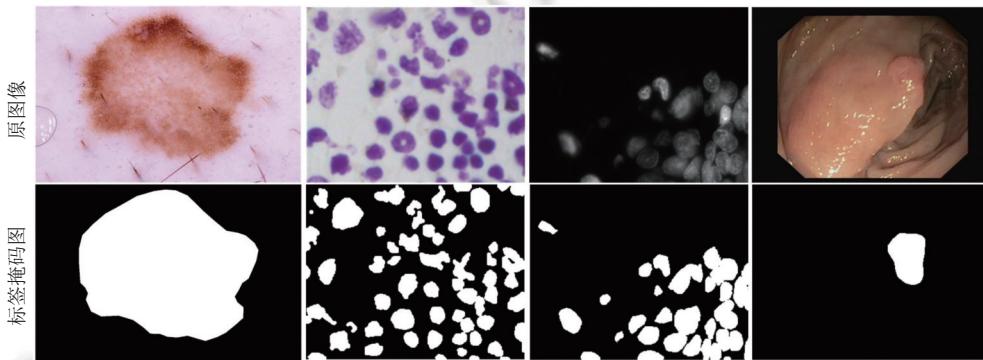


图 1 不同模态下的医学图像语义分割任务

基于 U-Net 的编码-解码网络^[6]通过在编解码路径中增加跳跃连接融合上下文的语义特征,有效提升了病灶分割的精度。一些方法^[7,8]在网络模型中增加注意力机制,使网络聚焦于目标区域特征的学习,同时抑制无关特征的干扰,进一步提升了语义分割的性能。此外,一些方法^[3,9-11]结合局部和全局语义信息,通过融合不同尺度和不同级别上下文特征的方法,提高了医学图像中尺度变换大、形状不规则等病灶区域的分割精度。然而,在多模态的医学图像语义分割任务中,上述全自动分割方法的精度、泛化能力和鲁棒性离临床诊断应用还具有一定的距离。基于此,一些专家学者提出了人机交互的分割方式。Wang 等人^[12]提出了基于深度学习的交互式分割方法,将用户交互信息转换为测地线距离图,通过不断地修正网络初始分割结果,在 2D MRI 胎盘分割和 3D MRI 脑肿瘤分割中取得了很好的分割效果。Wang 等人^[13]提出了 BIFSeg (bounding box and image-specific fine-tuning-based segmentation) 分割方法。该方法通过用户提供的边界框定位目标区域,将其作为网络的输入得到初始分割结果。然后对网络进行图像特异性微调,以提高分割精度。与传统的交互式分割方法相比,该方法在减少用户交互次数和用户时间的前提下获得了更好的分割效果。

上述方法虽然在医学影像分割任务中取得了不错的分割效果,但仍有一些不足之处。主要体现在两个方面:(1)基于 U-Net 的编码-解码网络及其变体,一方面在特征提取过程中丢失了部分空间细节信息,而对于语义分割来说,这种细粒度的信息至关重要。另一方面,在不同模态的医学图像语义分割任务中,这些模型的泛化能力和鲁棒性的表现不理想。(2)交互式的分割方法将用户提供的交互信息转化为基于目标和背景的距离图,并与原图像共同作为网络的输入,引导网络学习图像语义特征。该类方法虽然提升了网络模型在不同医学图像语义分割任务中

的分割精度, 以及泛化能力和鲁棒性, 但需要人为多次手动提供具有先验知识的交互信息。

针对上述问题, 本文充分结合 U-Net 编码-解码网络和交互式分割中用户提供的交互信息在语义分割中的优点, 构建了基于显著性引导和不确定性监督的医学图像语义分割算法。该算法将显著性检测网络生成的显著图和不确定概率图作为网络的输入, 引导网络关注目标区域特征的学习, 同时增强网络对不确定分类区域和复杂边界的表征能力, 以提升网络在不同模态医学图像语义分割任务中的分割性能。本文的主要贡献如下。

(1) 提出了一种显著性引导和不确定性监督的深度卷积编解码网络, 用来解决多模态下的复杂医学图像语义分割问题。

(2) 提出了一种多模态医学图像显著性检测算法, 通过多尺度特征的融合生成显著图, 并计算显著图中不确定分类的像素点集合, 生成不确定区域的概率图, 作为监督信息送入分割网络, 引导网络关注目标区域特征的学习, 同时增强网络对不确定分类区域和复杂边界的表征能力。

(3) 采取联合优化显著性检测网络和语义分割网络参数的方法, 既提升了显著性目标检测的性能, 避免了非目标区域的干扰, 同时也提高了网络的分割精度。

(4) 实验结果表明, 本文提出的显著性引导和不确定性监督的深度卷积编解码网络在多个模态医学图像数据集上, 与其他算法对比, 取得了更优的分割结果。

1 相关工作

1.1 基于深度学习的医学图像语义分割

近年来, 随着卷积神经网络的发展, 基于深度学习的方法因其在图像语义特征学习上的杰出表现, 在医学影像分析中得到了广泛的关注。目前, 基于深度学习的语义分割方法主要包括全卷积神经网络 FCN、编码-解码网络、循环神经网络 (recurrent neural networks, RNN) 和生成对抗网络 (generative adversarial network, GAN) 等。

2015 年, Long 等人提出了 FCN 网络^[5], 实现了任意尺寸图像端到端的像素级语义分割, 为医学影像的分割提供了新思路。例如, 全卷积神经网络 FCN 在结肠镜下息肉分割中的应用^[14]。Zhou 等人^[15]在初步预测的分割掩码上应用 FCN 的定点模型缩小输入区域, 实现了腹部 CT 图像中胰腺的精准分割。Zhang 等人^[16]利用堆叠的全卷积网络 (coarse-to-fine stacked fully convolutional nets, CFS-FCN), 实现了超声图像中淋巴结的自动分割。

由于 FCN 在下采样过程中采取池化操作, 导致部分空间信息丢失, 且在分割中未充分考虑上下文信息, 致使分割结果不够精细。为解决这些问题, 编码-解码网络引入医学影像分割中。其中, 典型的编码-解码网络 U-Net^[6]引入了 U 型的编码-解码结构, 并在每层的编解码中加入了跳跃连接, 将浅层的细粒度信息与深层的语义信息进行融合, 在生物医学图像语义分割任务中取得了很好的性能。

RNN 在全局特征建模和上下文信息保存方面具有一定的优势。因此, 在进行医学图像语义分割时, 一些专家学者将 RNN 的思想嵌入到网络中。Bai 等人^[17]将 FCN 和 RNN 结合起来进行医学图像语义分割。为加强特征传播和鼓励特征的重用, BCDU-Net 网络^[3]充分结合了 U-Net、双向 ConvLSTM 和密集卷积操作的优势, 在医学图像语义分割任务中取得了很好的效果。虽然 RNN 在图像的上下文特征学习中增强了网络的表征能力, 但由于 RNN 中预定义了单元序列输入的拓扑结构^[18], 上下文特征的学习具有一定的局限性, 导致分割结果因目标结构的变化而存在较大的差异。

此外, 基于 GAN 方法是图像语义分割中的另一种思路, 它将网络分成生成器网络和判别器网络两个部分。生成器网络和判别器网络在不断地对抗学习中优化网络参数, 逐步提高分割精度。例如, Xue 等人^[19]利用对抗学习的方法实现皮肤病灶区域及脑部肿瘤的分割。

1.2 编码-解码网络

目前, 基于编码-解码网络的方法在医学图像语义分割中取得了很好的分割效果, 受到了众多研究者的关注, 并在此结构上提出了一些改进方法^[20]。一些方法^[7,8,21]针对医学图像目标区域的复杂特征, 提出注意力机制, 通过在网络结构的不同位置加入 Attention 模块, 增强目标区域特征的学习, 同时抑制无关特征, 显著提升了目标区域的分割精度。一些方法^[9-11,22]从上下文特征融合的角度出发, 将不同尺度的空间信息、不同级别的语义信息以及不

同模态的特征进行融合,提升了网络的分割性能。为了提高 U-Net 网络在各种分割任务中的泛化能力,Jha 等人^[23]提出了 DoubleU-Net 网络,该网络将两个 U-Net 架构相互叠加进行组合,在不同模态的医学图像语义分割任务中获得了很好的分割结果。

1.3 显著性目标检测

显著性目标检测的目的是突出图像中视觉上显著的目标区域。近年来,基于卷积神经网络的显著性目标检测模型被广泛应用于计算机视觉领域,如图像理解^[24]、目标检测^[25]、语义分割^[26]等。Li 等人^[27]构建了基于全卷积神经网络的显著目标检测和语义分割框架。该模型采用数据驱动的策略有效学习了显著性对象的语义属性。Liu 等人^[28]提出了 DHSNet 网络,从全局粗略预测到分层逐步细化显著图的细节,实现端到端的深层次显著性检测。Wang 等人^[29]提出了一种自适应显著性切分方法实现显著目标的分割。此外,文献^[30,31]通过结合局部和全局上下文信息,以及聚合多级特征等方法增强特征表示能力,实现显著性目标检测。

现有的显著性目标检测方法在自然图像的显著性特征提取方面取得了一定的成果。为了有效提取医学图像中目标区域的显著特征,进而准确定位目标区域,实现目标区域的精准分割,受自然图像显著特征提取思想的启发,本文结合各模态下的医学图像特点,有效融合浅层空间细节信息和深层语义特征,提出了适用于不同模态下的医学图像显著特征学习网络,并作为引导信息进一步细化网络的分割结果。

2 显著性引导及不确定性监督的深度编解码网络

本文提出的显著性引导及不确定性监督的深度卷积编解码网络模型主要包括 3 个模块: 显著性引导模块(saliency-guided module, SGM)、不确定性监督模块(uncertainty-supervised module, USM) 和语义分割模块。总体网络模型结构如图 2 所示。首先,通过显著性引导模块,即图中的显著性检测网络,初步学习图像目标区域的显著性特征,生成显著图。然后,通过不确定性监督模块,计算显著图中不确定分类的像素点集合,生成不确定区域的概率图。最后,将显著图和不确定概率图作为监督信息,与输入图像一同送入分割网络,引导网络生成最终的分割结果图。

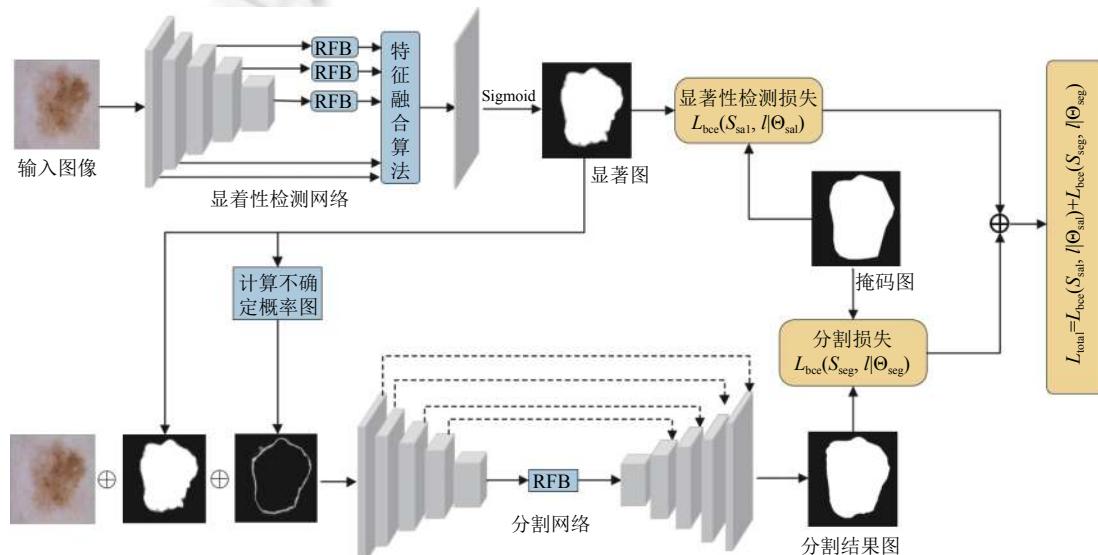


图 2 显著性引导及不确定性监督的深度编解码网络结构

2.1 显著性引导模块

该模块基于人类视觉特征,通过网络学习视觉注意区域的特征,计算图像区域的显著性,进而自动捕获图像中感兴趣的目标区域,作为后续分割任务的引导图。Wu 等人^[32]利用级联部分解码网络来获取自然图像的显著图,该网络结构包含两个分支,即注意力分支和检测分支,并在每个分支中采用部分解码模块,丢弃了预训练模型中前两

层的浅层特征, 以降低模型的复杂度。受该方法的启发, 本文提出了多层语义信息感知的显著图生成方法, 从而快速精准的定位医学图像中待分割的目标区域。考虑到 Wu 等人的方法采用级联的网络增加了模型的复杂度, 本文采用注意力检测生成显著图, 从而有效降低模型的复杂度。因不同的成像原理, 医学图像与自然图像存在较大的差别, 如医学图像中病灶区域和组织器官的边界模糊、形状不规则、尺度变化大、边界对比度低等, 浅层特征包含的边界和位置等细粒度信息对于目标区域的边界处分割至关重要。因此, 为了同时汇聚浅层的丰富空间信息和深层的全局语义信息, 本文的主干网络采用 ImageNet 数据集上预训练的 VGG16 网络, 通过加载网络模型及参数初步学习输入图像的语义特征, 同时将深层的语义特征通过感受野块 (receptive field block, RFB) 汇聚到解码层, 并与浅层特征进行融合, 以提升显著图在边界处的预测效果。其网络结构如图 3 所示。

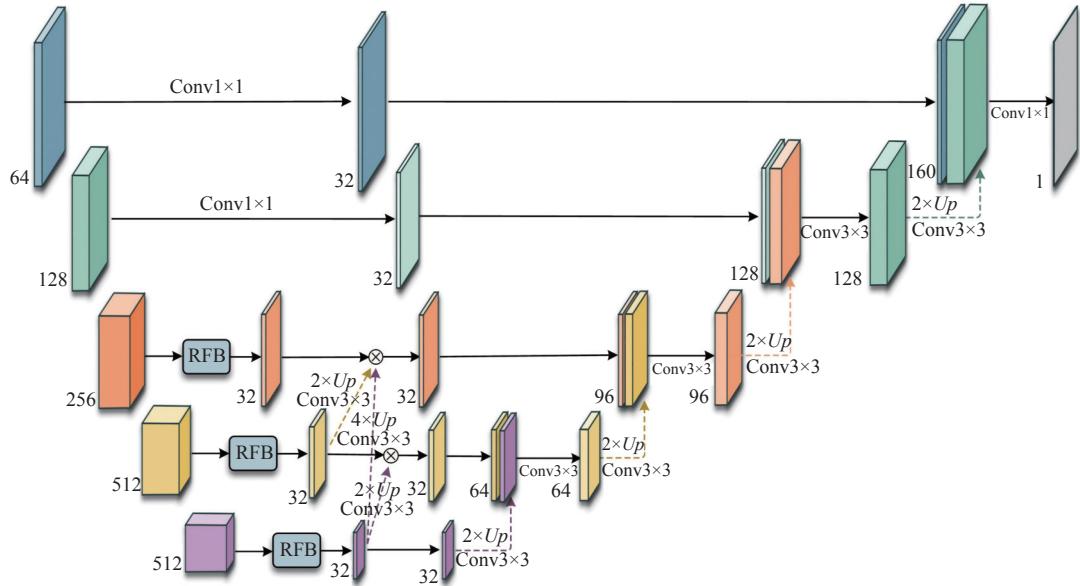


图 3 显著性检测网络结构图

对于输入图像 \mathbf{X} , 其大小尺寸为 $H \times W$, 这里我们将 VGG16 网络编码模块中每一层的特征图定义为 $\{f_i^e, i=1, \dots, 5\}$, 经过大池化操作, 对应的特征图尺寸大小为 $[H/2^{i-1}, W/2^{i-1}]$ 。编码网络中较浅层的特征图包含了更多的空间细节信息, 但缺少全局语义信息; 而较深层的特征图突出显示了显著性区域的位置, 同时包含更多的全局语义信息, 但却丢失了部分空间细节信息。为了弥补这一缺陷, 在解码过程中, 我们采用特征融合算法, 融合了所有浅层和深层的特征信息。为了在深层的特征图中获取更多的全局语义信息, 在编码的第 3、4 和 5 层, 引入了 RFB 模块^[32,33], 该模块的灵感来源于 Inception 结构^[34], 采用不同尺度的卷积核进行多分支的卷积操作, 从多个尺度提取特征, 进而获取丰富的上下文信息。同时, 在不增加模型参数量的前提下, 为扩大卷积感受野, 这里采用空洞卷积的思想^[35,36]。RFB 的结构如图 4 所示。该模块首先以 4 个平行的分支分别从不同尺度提取上下文信息, 扩大特征提取的感受野, 然后将各分支的输出进行拼接, 并经过一个 3×3 的卷积操作减少特征图的通道数。最后, 通过短连接操作与原特征图进行融合。

在解码过程的特征融合算法中, 为了解决不同层之间存在的语义鸿沟问题, 我们采取每一层特征图与更深层的语义特征图进行像素级的乘操作, 其目的是在浅层特征图中融合更多的深层语义信息。具体操作定义如下:

$$f_i = \begin{cases} Conv_1(f_i^e), i \in [1, 2] \\ f_i^{rfb} \odot \prod_{k=i+1}^5 Conv_2(Up(f_k)), i \in [3, 4] \\ f_i^{rfb}, i = 5 \end{cases} \quad (1)$$

其中, $Conv_1$ 和 $Conv_2$ 分别表示卷积核大小为 1×1 和 3×3 的卷积操作, f_i^{rb} 为经过 RFB 模块输出的特征图, $Up(\cdot)$ 表示采样因子为 2^{k-i} 的上采样操作。经过浅层与深层的特征信息融合后, 解码网络每层输出的特征图定义如下:

$$f_i^d = \begin{cases} Conv_1(f_i \cdot Conv_2(Up(f_{i+1}^d))), i=1 \\ Conv_2(f_i \cdot Conv_2(Up(f_{i+1}^d))), i \in [2, 4] \\ f_i, i=5 \end{cases} \quad (2)$$

其中, $Conv_1$ 和 $Conv_2$ 分别表示卷积核大小为 1×1 和 3×3 的卷积操作, $Up(\cdot)$ 表示采样因子为 2 的上采样操作, \cdot 表示特征图拼接操作。 f_i 为经公式 (1) 特征融合后输出的特征图。最后, 经过 Sigmoid 函数生成显著图。

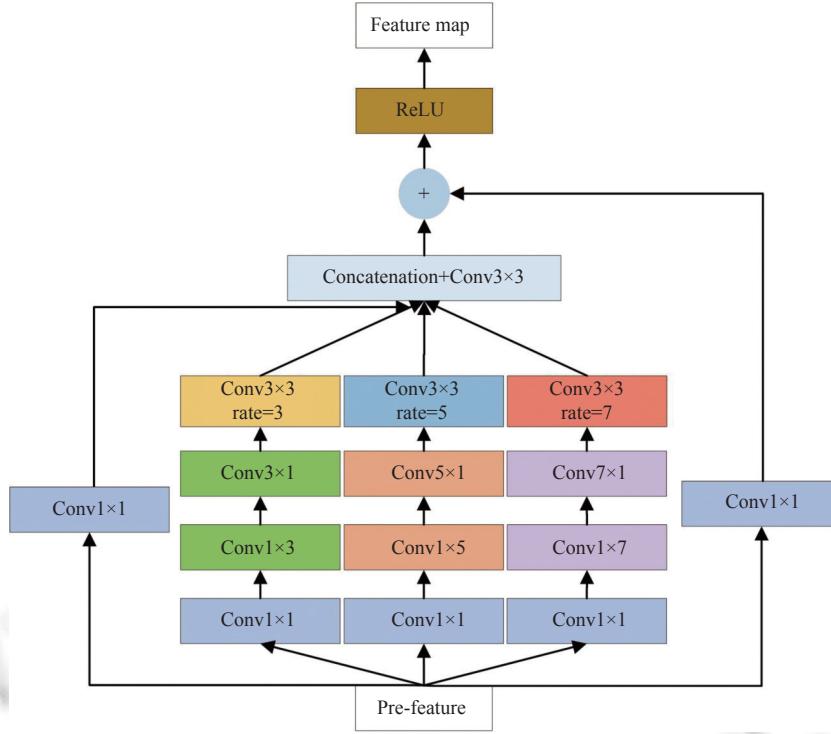


图 4 RFB 模块结构图

2.2 不确定性监督模块

尽管显著性引导模块生成了良好的显著图, 但因医学图像的模态多样性, 组织器官和病灶区域的形状不规则、尺度变化大、周围组织强度不均匀, 以及边界模糊等影响因素, 给医学图像显著性检测带来了一定的挑战, 算法对部分复杂样本的学习仍然存在局限性。因此, 如何让网络更多地关注那些不确定分类区域, 加强不确定分类区域和复杂边界的特征表示能力, 成为医学图像精准分割的关键。为了解决这一问题, 在文献 [37] 的启发下, 本文提出了像素级的不确定性监督模块, 将其作为一种软约束条件, 与原图和显著图一同送入网络训练, 有效引导语义分割网络学习图像的语义特征, 进而提升网络的分割性能。不确定概率图 $U_{i,j}$ 的计算如下:

$$U_{i,j} = 1 - \frac{|S_{i,j} - 0.5|}{0.5} \quad (3)$$

其中, $S_{i,j} \in (0, 1)$ 为显著图中像素点 (i, j) 的概率值。显著性引导模块的最后一层经过 Sigmoid 分类层, 输出为显著图中各像素点的概率值。我们认为预测概率值越接近阈值 0.5, 其对应位置的像素点分类预测越模糊。因此, 本文通过计算不确定概率图, 赋予显著图中那些不确定分类的像素点更大的权重。

2.3 语义分割模块

因 U-Net 网络在医学图像语义分割领域取得了卓越的分割性能, 本文分割网络模型以 U-Net 跳跃连接-编码-

解码网络结构为主干网络, 网络的输入除了原图像, 额外增加了原图像对应的显著图和不确定概率图两个通道。由于 U-Net 网络在编码路径中逐层采用固定尺寸的卷积核学习图像中的抽象语义特征, 而对于医学图像中那些尺度变化大和形状不规则, 以及边界处对比度低的目标区域, 如肿瘤、细胞和组织器官等, U-Net 网络不能鲁棒地学习其不同尺度的上下文信息, 导致分割性能的不稳定。因此, 本文在编码底层引入了多尺度特征融合模块, 以不同尺度的感受野学习深层语义特征图中丰富的上下文信息, 进一步提升网络的分割性能。本文的多尺度特征融合模块采用显著性检测网络中的 RFB 块。

2.4 损失函数

本文采用联合训练的方式, 分别在显著性检测网络和语义分割网络输出的概率图 $\{S_{\text{sal}}, S_{\text{seg}}\}$ 上计算损失。两部分的损失函数均采用像素级交叉熵损失。总的损失函数 L_{total} 定义为:

$$L_{\text{total}} = L_{\text{bce}}(S_{\text{sal}}, l|\Theta_{\text{sal}}) + L_{\text{bce}}(S_{\text{seg}}, l|\Theta_{\text{seg}}) \quad (4)$$

其中, l 为标签图, L_{bce} 为二分类交叉熵损失, 定义如下:

$$L_{\text{bce}}(\Theta) = - \sum_{j=1}^N \sum_{c \in \{0,1\}} \delta(l^j = c) \log p(S^j = c|\Theta) \quad (5)$$

其中, N 为像素个数, δ 为指标函数, j 为像素坐标, $\Theta = \{\Theta_{\text{sal}}, \Theta_{\text{seg}}\}$ 为概率图 $S = \{S_{\text{sal}}, S_{\text{seg}}\}$ 的参数集。训练过程中, 显著性检测网络和语义分割网络两部分联合优化参数。一方面, 显著性检测网络为语义分割部分提供精确的目标定位, 引导语义分割网络得到更精确的分割结果。另一方面, 语义分割网络反向传播过程中, 可以作为显著性检测网络的辅助损失, 有助于显著性检测网络聚焦于显著性目标区域。两部分的联合优化, 既提升了显著性目标检测的性能, 避免了非目标区域的干扰, 同时也提高了网络的分割精度。

3 实验结果及分析

3.1 实验设置

3.1.1 数据集

本文共使用 4 个公开的医学图像数据集来验证本文算法的有效性。数据集涵盖了不同模态下的医学图像语义分割任务, 包括皮肤镜下的皮肤癌病灶分割、显微镜成像下的细胞核分割和结肠镜下的息肉分割。数据集的具体细节信息描述如表 1 所示。其中, 显微镜成像下的细胞核分割数据集 2018 Data Science Bowl 的下载网址为: <https://www.kaggle.com/c/data-science-bowl-2018>.

表 1 数据集细节描述

数据集	图片数量	图片大小	模态
ISIC 2018 ^[38]	2 594	变化尺寸	皮肤镜图像
2018 Data Science Bowl	670	256×256	混合模态
CVC-ClinicDB ^[39]	612	384×288	结肠镜图像
Kvasir-SEG ^[40]	1 000	变化尺寸	结肠镜图像

3.1.2 评价指标

本文使用医学图像语义分割中常用的几种评估指标来验证算法的性能, 主要包括: $F1$ 分数、召回率 (recall, Rec)、准确率 (accuracy, Acc)、特异性 (specificity, Spec)、精确率 (precision, Prec)、Dice 系数和平均交并比 (mean IoU, mIoU)。

3.1.3 网络参数设置

本文实验采用开源的 PyTorch 框架, 训练和测试在 Ubuntu 16.04 LTS 64 位操作系统上进行, 内存为 64 GB DDR 4 MHz, 显卡采用 NVIDIA GeForce GTX 1080 Ti 独立显卡。为了验证算法的性能, 这里未采用图像预处理和后处理操作, 所有图像在送入模型前只进行了尺寸大小的调整。模型训练过程中, 损失函数采用二分类交叉

熵损失, 优化算法采用 Adam, 初始学习率为 0.000 1, 批处理大小为 4, 网络的训练迭代次数 epoch 为 100 次。当训练损失每 30 次迭代保持不变时, 我们采取学习率下降 10% 的更新策略。

3.2 实验结果分析

3.2.1 与其他算法的实验对比结果

(1) ISIC 2018 数据集上的实验对比结果

ISIC 2018 为皮肤癌病灶分割数据集, 该数据集共 2594 张图片, 包含了不同尺度大小、不同形状和不同颜色的皮肤病变区域。按照参考文献 [3] 的方法, 本文将 1815 张图片作为训练集, 259 张图片作为验证集, 剩下的 520 张图片用作测试集。在训练的过程中, 我们将图片尺寸调整为 512×512 的大小送入网络模型。表 2 给出了本文算法与其他算法在该数据集上的实验对比结果。从表 2 中可以看出, 本文算法除了在特异性 Spec 指标上略低于最优结果, 其他指标均优于其他算法。其中, F1 分数、召回率 Rec 和准确率 Acc 分别为 0.8803、0.8868 和 0.9609, 相比于其他算法的最优结果分别提高了 0.72%、2.18% 和 2.39%。在特异性 Spec 指标上本文取得的结果是 0.9802, 与最优结果仅差 0.18%。因特异性指标主要强调背景区域的分割精度, 而在皮肤癌病灶分割中, 我们更关注皮肤病变区域的分割精度, 因此本文算法的整体分割性能优于其他算法。本文的可视化结果如图 5 所示。从视觉分析上, 我们可以观察到本文算法的分割结果在不同尺度病灶区域、复杂背景遮挡、病灶区域与周围背景对比度低, 以及病灶区域边界模糊等情况下, 均取得了最优的分割结果。

表 2 ISIC 2018 数据集上的定量比较结果

方法	F1	Rec	Spec	Acc
U-Net ^[6]	0.8163	0.8192	0.9741	0.9391
BCDU-Net ($d=1$) ^[3]	0.8470	0.7830	0.9800	0.9360
BCDU-Net ($d=3$) ^[3]	0.8510	0.7850	0.9820	0.9370
FANet ^[41]	0.8731	0.8650	0.9611	0.9351
本文算法	0.8803	0.8868	0.9802	0.9609

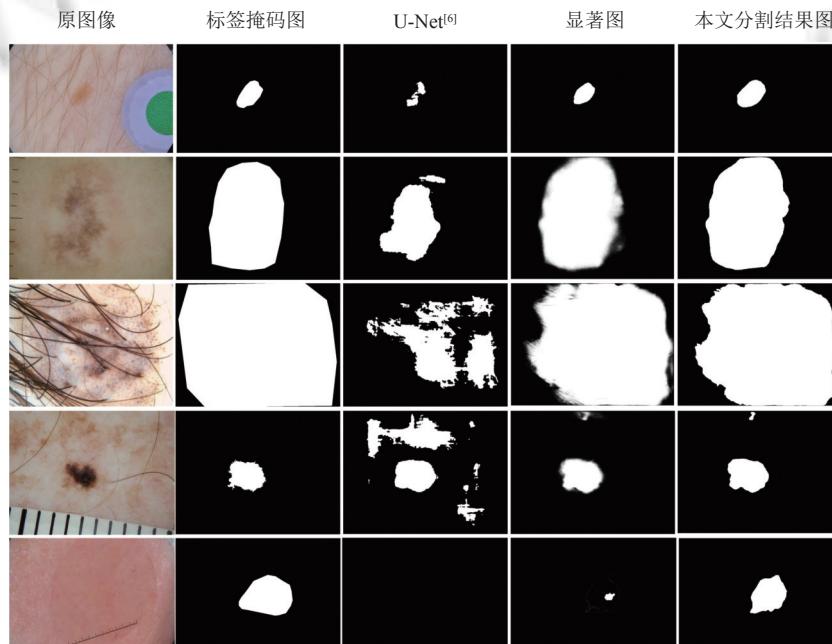


图 5 ISIC 2018 数据集上的可视化分割结果

(2) 2018 Data Science Bowl 数据集上的实验对比结果

2018 Data Science Bowl 为显微镜成像下的细胞核分割数据集。该数据集共 670 张图片，包含了大量不同细胞类型、不同放大倍数和不同成像方式的可变形状的细胞核。按照文献 [23] 划分数据集的方法，本文将数据集划分为训练集、验证集和测试集。在训练的过程中，我们将图片尺寸调整到 256×256 的大小送入网络模型。本文算法与其他算法在 2018 Data Science Bowl 数据集上的实验对比结果如表 3 所示。从表 3 可以看出，本文算法在 Dice 和 mIoU 评价指标上取得了最优结果，分别为 0.9209 和 0.8572，比最优的分割结果分别提升了 0.33% 和 0.03%。因细胞核图像中类别不平衡因素，导致召回率 Rec 和精确率 Prec 取得次优结果。与其他算法相比，本文在召回率 Rec 和精确率 Prec 之间取得了最佳的平衡使得 Dice 和 mIoU 指标取得最优值，总体分割性能优于其他算法。图 6 展示了本文算法的可视化分割结果。从视觉分析上，我们可以观察到，本文算法在不同类型的细胞核图像分割中取得了很好的分割效果。例如，在第 1 行和第 2 行的细胞核分割中，本文算法避免了细胞核粘连现象；在第 3、4、5 和第 6 行的不同成像方式细胞核分割中，因细胞核亮度不均、对比度低、边界模糊等因素，本文算法能够有效排除视觉相似区域的干扰，一定程度避免了欠分割和过分割现象。

表 3 2018 Data Science Bowl 数据集上的定量比较结果

方法	Dice	mIoU	Rec	Prec
U-Net ^[6]	0.9098	0.8372	0.8904	0.9164
DoubleU-Net ^[23]	0.9133	0.8407	0.6407	0.9496
FANet ^[41]	0.9176	0.8569	0.9222	0.9194
本文算法	0.9209	0.8572	0.9116	0.9347

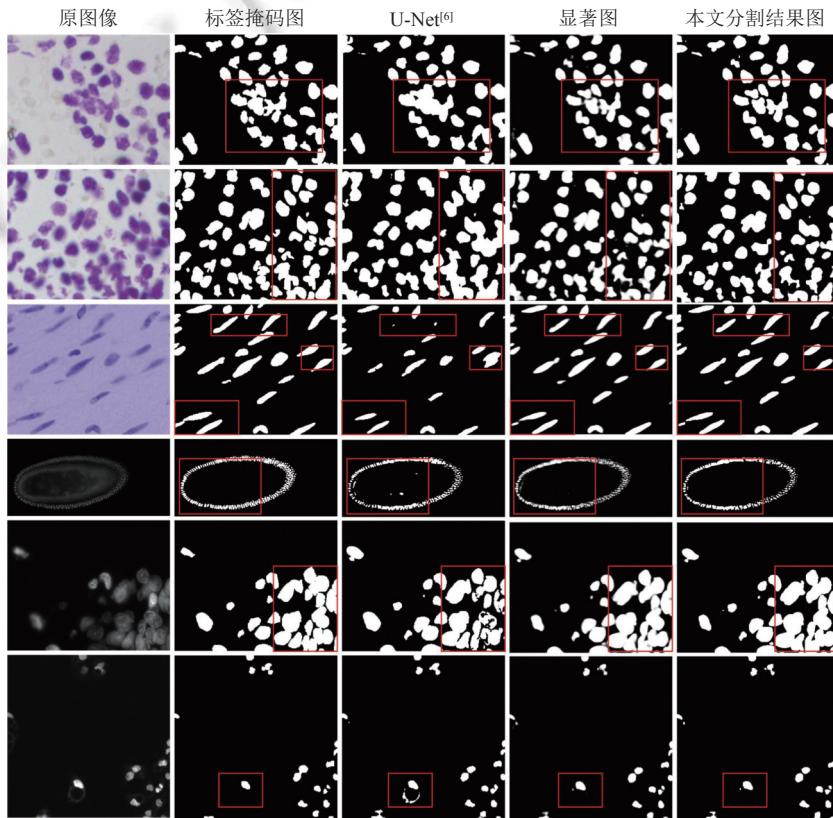


图 6 2018 Data Science Bowl 数据集上的可视化分割结果

(3) 息肉分割数据集上的实验对比结果

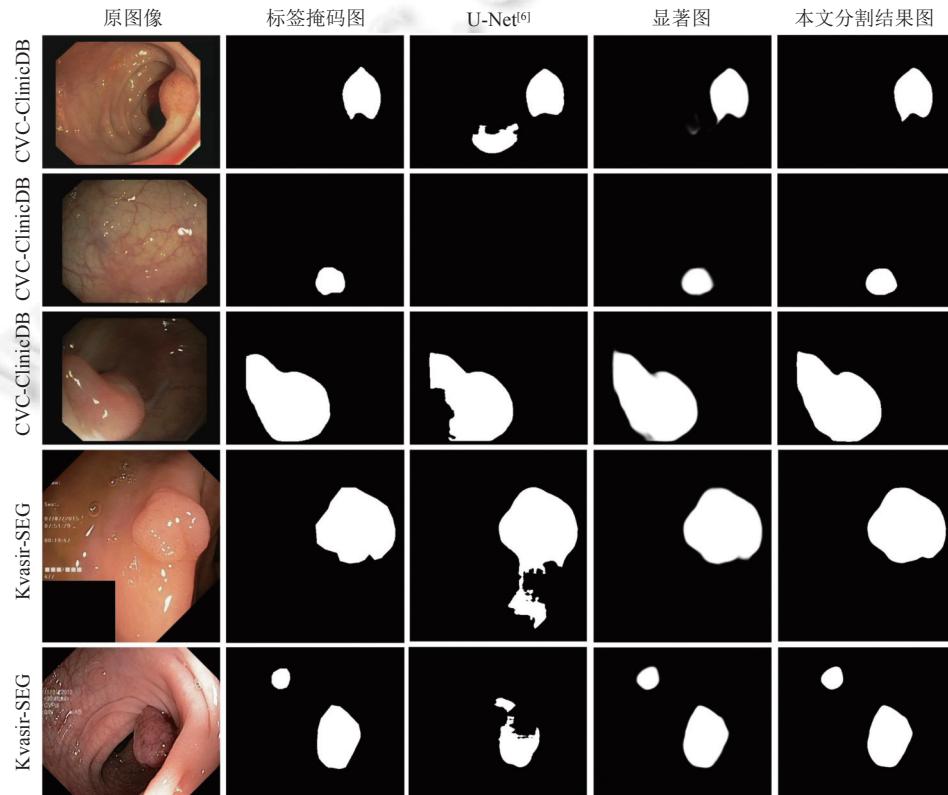
本文分别在结肠镜下的 CVC-ClinicDB 和 Kvasir-SEG 数据集上进行息肉分割实验。其中，CVC-ClinicDB 数据集包含 612 张图片，尺寸大小为 384×288 ；Kvasir-SEG 数据集包含 1000 张不同尺寸的图片。我们按照文献 [4] 的方法划分数据集，将其按照 80%、10% 和 10% 的比例划分为训练集、验证集和测试集。在训练的过程中，我们统一将图片尺寸调整到 256×256 的大小，并送入网络模型。**表 4** 和 **表 5** 分别给出了本文算法与其他算法在两个数据集上的定量对比结果。从**表 4** 和 **表 5** 可以看出，本文算法在 Dice、平均交并比 mIoU、召回率 Rec 和精确率 Prec 这 4 个评价指标上均取得了最优的分割结果。**图 7** 展示了本文算法在两个数据集上的可视化分割结果。通过观察，本文算法在多种情况下，如不同大小、形状、颜色和纹理的息肉分割中，以及息肉与其周围粘膜的模糊边界处，均能够精确地定位和分割息肉组织。

表 4 CVC-ClinicDB 数据集上的定量比较结果

方法	Dice	mIoU	Rec	Prec
U-Net ^[6]	0.8123	0.7476	0.8256	0.8388
ResUNet-mod ^[42]	0.7788	0.4545	0.6683	0.8877
ResUNet++ ^[4]	0.7955	0.7962	0.7022	0.8785
本文算法	0.9136	0.8619	0.9147	0.9223

表 5 Kvasir-SEG 数据集上的定量比较结果

方法	Dice	mIoU	Rec	Prec
U-Net ^[6]	0.8116	0.7217	0.7949	0.8726
ResUNet-mod ^[42]	0.7909	0.4287	0.6909	0.8713
ResUNet++ ^[4]	0.8133	0.7927	0.7064	0.8774
本文算法	0.8766	0.8077	0.8588	0.9347

**图 7** CVC-ClinicDB 和 Kvasir-SEG 数据集上的可视化分割结果

3.2.2 网络模型有效性消融实验

本节主要针对本文的显著性引导模块、不确定性监督模块和语义分割模块在网络模型中的有效性进行了消融实验研究。首先，显著性引导模块中，我们对原文献 [32] 的显著性检测网络进行了改进，在原文献的注意力分支基础上，本文在解码结构中增加了前两层的浅层特征融合，其目的是保留更多的空间细粒度信息，突出目标区域边

界信息。本文显著性检测网络采用 VGG16 作为主干网络, 编码部分采用 ImageNet 数据集上预训练的网络参数模型。图 8 展示了本文算法与原算法在 4 个不同模态的数据集上生成显著图的可视化对比结果。通过观察, 因本文在解码结构中融合了浅层的特征, 在边界模糊、背景遮挡、对比度低、目标区域颜色变化以及亮度不均等多种复杂情况下, 能够更准确地突出显著性目标和细化目标的边界。

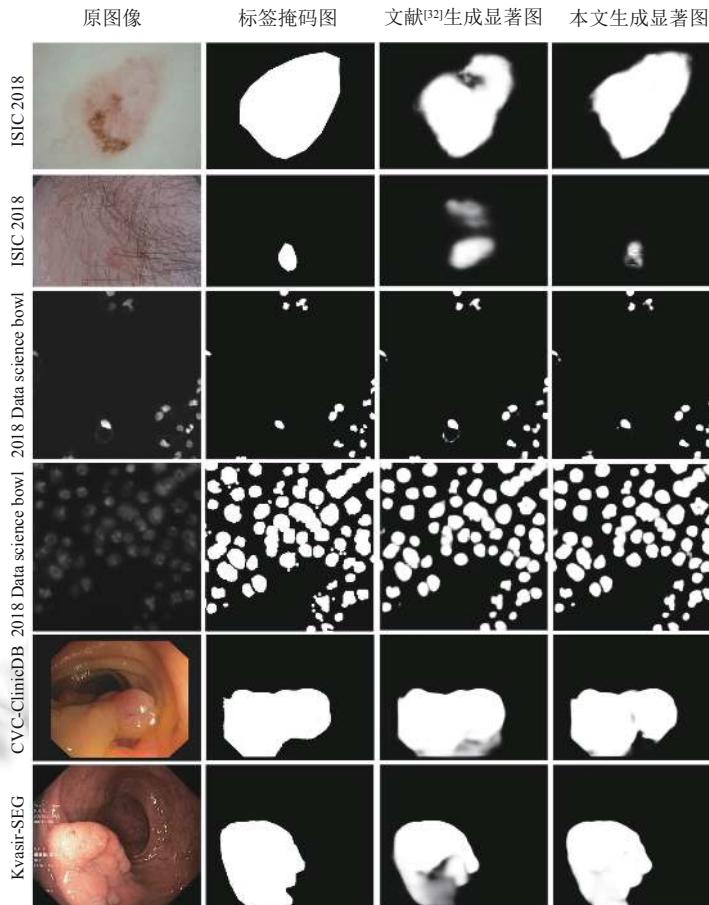


图 8 生成显著图的可视化对比结果

进一步地, 本文评估了各模块在网络模型中的贡献和分割性能。本文以 U-Net 网络为分割的主干网络, 我们采取各模块组合的操作来验证相应模块的有效性。图 9 中给出了各模块组合后在 4 个不同模态数据集上的对比结果。其中, 各模块的组合说明如下。

- SGM+U-Net: 该组合表示在主干网络 U-Net 基础上加入显著性引导模块, 其目的是通过显著性检测初步定位目标区域, 消除背景区域的干扰, 并作为输入信息引导网络分割。
- SGM+USM+U-Net: 该组合表示在显著性引导模块的基础上增加了不确定性监督模块, 其目的是加强不确定分类区域和复杂边界的特征表示能力, 使网络更关注不确定分类区域的学习。
- SGM+USM+RFB-U-Net: 该组合为本文最终的网络结构, 这里将前两个模块得到的显著图和不确定概率图作为监督信息与原图一并送入分割网络, 同时, 在 U-Net 编码底层加入了多尺度特征融合模块 RFB, 以不同尺度的感受野学习深层语义特征图中丰富的上下文信息, 以进一步提升网络的分割性能。

如图 9 所示, 通过逐步组合各模块, 各评价指标值在 4 个不同模态的数据集上大多数均不同程度依次得到了提升。其中, SGM+U-Net 因在 U-Net 主干网络加入显著性引导模块, 为网络提供辅助监督信息, 细化了网络分割结

果,在 4 个数据集上各评价指标均得到了提升; SGM+USM+U-Net 通过在显著性引导模块的基础上增加了不确定性监督模块,增强了网络对不确定区域和复杂边界特征的学习能力。从柱状图中可以看出,4 个不同模态的数据集上各评价指标均高于 U-Net 和 SGM+U-Net,也进一步验证了不确定性监督模块的有效性。此外,因本文在 U-Net 编码底层加入了多尺度特征融合模块,从多个尺度学习深层语义特征图的上下文信息,SGM+USM+RFB-U-Net 组合的网络结构在 4 个不同模态的数据集上均取得了最优的分割结果。因此,消融实验结果进一步验证了本文中各模块的有效性。

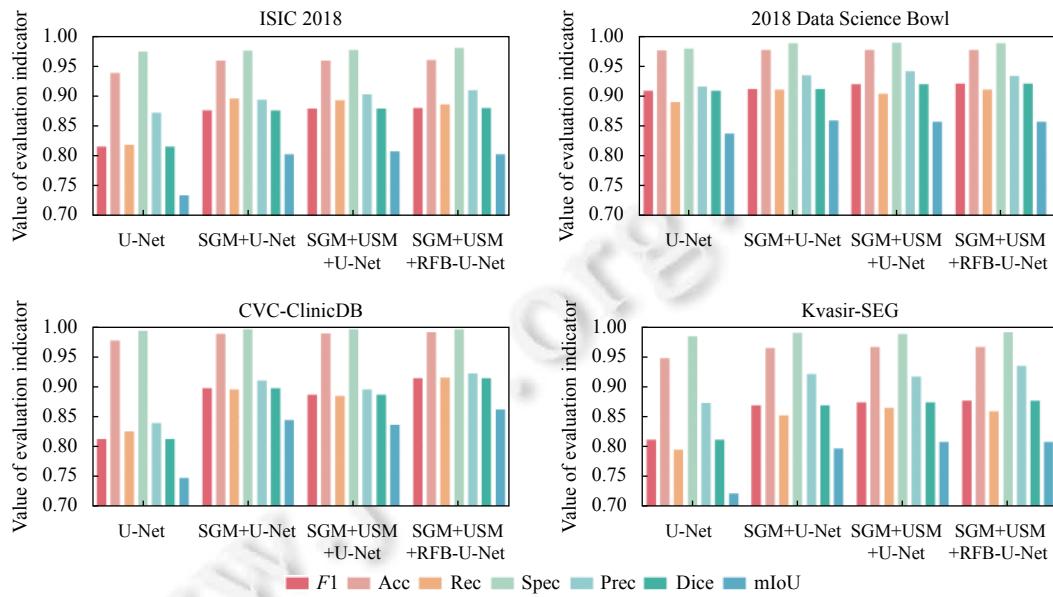


图 9 各评价指标在不同数据集上的消融实验对比结果

4 总 结

本文提出一种显著性引导及不确定性监督的深度卷积编解码网络,用来解决不同模态下的复杂医学图像语义分割问题。其中,显著性检测模块通过汇聚浅层的丰富空间信息和深层的全局语义信息,有效地挖掘出图像中的目标区域,提升了显著图的预测结果。不确定性监督模块通过计算不确定概率图,引导网络更加关注不确定区域的特征学习,增强了不确定区域和复杂边界的表征能力。语义分割模块通过在编码底层增加多尺度感受野特征提取模块,有效提升了网络全局上下文信息的感知能力,进一步提升了模型的分割效果。此外,本文采取联合优化显著性检测网络和语义分割网络参数的方法,既提升了显著性目标检测的性能,避免了非目标区域的干扰,同时也提高了网络的分割精度。本文通过消融实验证实了各模块在语义分割中的有效性。实验表明,本文算法在多个不同模态的医学图像数据集上的分割性能均优于其他算法。我们相信本文算法也可推广到其他医学图像语义分割任务中。

医学图像的模态多样性和复杂性,类别不均衡,样本量少和数据质量等问题,仍然给医学图像语义分割带来了更多的挑战。如何针对现有问题,设计更具泛化能力的网络模型,进一步提升医学图像的分割精度,成为未来研究的方向。

References:

- [1] Song J, Xiao L, Lian ZC, Cai ZY, Jiang GP. Overview and prospect of deep learning for image segmentation in digital pathology. Ruan Jian Xue Bao/Journal of Software, 2021, 32(5): 1427–1460 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6205.htm> [doi: 10.13328/j.cnki.jos.006205]
- [2] Ibtehaz N, Rahman MS. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural

- Networks, 2020, 121: 74–87. [doi: [10.1016/j.neunet.2019.08.025](https://doi.org/10.1016/j.neunet.2019.08.025)]
- [3] Azad R, Asadi-Aghbolaghi M, Fathy M, Escalera S. Bi-directional convLSTM U-Net with densely connected convolutions. In: Proc. of the IEEE/CVF Int'l Conf. on Computer Vision Workshop. Seoul: IEEE, 2019. 406–415. [doi: [10.1109/ICCVW.2019.00052](https://doi.org/10.1109/ICCVW.2019.00052)]
- [4] Jha D, Smedsrød PH, Riegler MA, Johansen D, De Lange T, Halvorsen P, Johansen HD. ResUNet++: An advanced architecture for medical image segmentation. In: Proc. of the 2019 IEEE Int'l Symp. on Multimedia (ISM). San Diego: IEEE, 2019. 225–230. [doi: [10.1109/ISM46123.2019.00049](https://doi.org/10.1109/ISM46123.2019.00049)]
- [5] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015. 3431–3440. [doi: [10.1109/CVPR.2015.7298965](https://doi.org/10.1109/CVPR.2015.7298965)]
- [6] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Proc. of the 18th Int'l Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Munich: Springer, 2015. 234–241. [doi: [10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28)]
- [7] Gu R, Wang GT, Song T, Huang R, Aertsen M, Deprest J, Ourselin S, Vercauteren T, Zhang ST. CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation. *IEEE Trans. on Medical Imaging*, 2021, 40(2): 699–711. [doi: [10.1109/TMI.2020.3035253](https://doi.org/10.1109/TMI.2020.3035253)]
- [8] Wang CY, Wang YL, Liu YF, He ZF, He R, Sun ZN. ScleraSegNet: An attention assisted U-Net model for accurate sclera segmentation. *IEEE Trans. on Biometrics, Behavior, and Identity Science*, 2020, 2(1): 40–54. [doi: [10.1109/TBIM.2019.2962190](https://doi.org/10.1109/TBIM.2019.2962190)]
- [9] Gu ZW, Cheng J, Fu HZ, Zhou K, Hao HY, Zhao YT, Zhang TY, Gao SH, Liu J. CE-Net: Context encoder network for 2D medical image segmentation. *IEEE Trans. on Medical Imaging*, 2019, 38(10): 2281–2292. [doi: [10.1109/TMI.2019.2903562](https://doi.org/10.1109/TMI.2019.2903562)]
- [10] Zhou ZW, Siddiquee MMR, Tajbakhsh N, Liang JM. UNet++: A nested U-Net architecture for medical image segmentation. In: Proc. of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Granada: Springer, 2018. 3–11. [doi: [10.1007/978-3-030-00889-5_1](https://doi.org/10.1007/978-3-030-00889-5_1)]
- [11] Huang HM, Lin LF, Tong RF, Hu HJ, Zhang QW, Iwamoto Y, Han XH, Chen YW, Wu J. UNet 3+: A full-scale connected UNet for medical image segmentation. In: Proc. of the 2020 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP). Barcelona: IEEE, 2020. 1055–1059. [doi: [10.1109/ICASSP40776.2020.9053405](https://doi.org/10.1109/ICASSP40776.2020.9053405)]
- [12] Wang GT, Zuluaga MA, Li WQ, Pratt R, Patel PA, Aertsen M, Doel T, David AL, Deprest J, Ourselin S, Vercauteren T. DeepGeoS: A deep interactive geodesic framework for medical image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2019, 41(7): 1559–1572. [doi: [10.1109/TPAMI.2018.2840695](https://doi.org/10.1109/TPAMI.2018.2840695)]
- [13] Wang GT, Li WQ, Zuluaga MA, Pratt R, Patel PA, Aertsen M, Doel T, David AL, Deprest J, Ourselin S, Vercauteren T. Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE Trans. on Medical Imaging*, 2018, 37(7): 1562–1573. [doi: [10.1109/TMI.2018.2791721](https://doi.org/10.1109/TMI.2018.2791721)]
- [14] Brandao P, Mazomenos E, Ciuti G, Caliò R, Bianchi F, Menciassi A, Dario P, Koulaouzidis A, Arezzo A, Stoyanov D. Fully convolutional neural networks for polyp segmentation in colonoscopy. In: Proc. of the SPIE 10134, Medical Imaging 2017: Computer-Aided Diagnosis. Orlando: SPIE, 2017. 101340F. [doi: [10.1117/12.2254361](https://doi.org/10.1117/12.2254361)]
- [15] Zhou YY, Xie LX, Shen W, Wang Y, Fishman EK, Yuille AL. A fixed-point model for pancreas segmentation in abdominal ct scans. In: Proc. of the 20th Int'l Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Quebec City: Springer, 2017. 693–701. [doi: [10.1007/978-3-319-66182-7_79](https://doi.org/10.1007/978-3-319-66182-7_79)]
- [16] Zhang YZ, Ying MTC, Yang L, Ahuja AT, Chen DZ. Coarse-to-fine stacked fully convolutional nets for lymph node segmentation in ultrasound images. In: Proc. of the 2016 IEEE Int'l Conf. on Bioinformatics and Biomedicine (BIBM). Shenzhen: IEEE, 2016. 443–448. [doi: [10.1109/BIBM.2016.7822557](https://doi.org/10.1109/BIBM.2016.7822557)]
- [17] Bai WJ, Suzuki H, Qin C, Tarroni G, Oktay O, Matthews PM, Rueckert D. Recurrent neural networks for aortic image sequence segmentation with sparse annotations. In: Proc. of the 21st Int'l Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI). Granada: Springer, 2018. 586–594. [doi: [10.1007/978-3-030-00937-3_67](https://doi.org/10.1007/978-3-030-00937-3_67)]
- [18] Van Den Oord A, Kalchbrenner N, Kavukcuoglu K. Pixel recurrent neural networks. In: Proc. of the 33rd Int'l Conf. on Machine Learning. New York: PMLR, 2016. 1747–1756.
- [19] Xue Y, Xu T, Zhang H, Long LR, Huang XL. SegAN: Adversarial network with multi-scale L_1 loss for medical image segmentation. *Neuroinformatics*, 2018, 16(3): 383–392. [doi: [10.1007/s12021-018-9377-x](https://doi.org/10.1007/s12021-018-9377-x)]
- [20] Yin XH, Wang YC, Li DY. Survey of medical image segmentation technology based on U-Net structure improvement. *Ruan Jian Xue Bao/Journal of Software*, 2021, 32(2): 519–550 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6104.htm> [doi: [10.13328/j.cnki.jos.006104](https://doi.org/10.13328/j.cnki.jos.006104)]
- [21] Ni ZL, Bian GB, Zhou XH, Hou ZG, Xie XL, Wang C, Zhou YJ, Li RQ, Li Z. RAUNet: Residual attention U-Net for semantic

- segmentation of cataract surgical instruments. In: Proc. of the 26th Int'l Conf. on Neural Information Processing. Sydney: Springer, 2019. 139–149. [doi: [10.1007/978-3-030-36711-4_13](https://doi.org/10.1007/978-3-030-36711-4_13)]
- [22] Dolz J, Ayed IB, Desrosiers C. Dense multi-path U-Net for ischemic stroke lesion segmentation in multiple image modalities. In: Proc. of the 4th Int'l Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Granada: Springer, 2018. 271–282. [doi: [10.1007/978-3-030-11723-8_27](https://doi.org/10.1007/978-3-030-11723-8_27)]
- [23] Jha D, Riegler MA, Johansen D, Halvorsen P, Johansen HD. DoubleU-Net: A deep convolutional neural network for medical image segmentation. In: Proc. of the 33rd IEEE Int'l Symp. on Computer-Based Medical Systems (CBMS). Rochester: IEEE, 2020. 558–564. [doi: [10.1109/CBMS49503.2020.00111](https://doi.org/10.1109/CBMS49503.2020.00111)]
- [24] Zhu JY, Wu JJ, Xu Y, Chang E, Tu ZW. Unsupervised object class discovery via saliency-guided multiple class learning. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2015, 37(4): 862–875. [doi: [10.1109/TPAMI.2014.2353617](https://doi.org/10.1109/TPAMI.2014.2353617)]
- [25] Ren ZX, Gao SH, Chia LT, Tsang IWH. Region-based saliency detection and its application in object recognition. IEEE Trans. on Circuits and Systems for Video Technology, 2014, 24(5): 769–779. [doi: [10.1109/TCSVT.2013.2280096](https://doi.org/10.1109/TCSVT.2013.2280096)]
- [26] Wei YC, Feng JS, Liang XD, Cheng MM, Zhao Y, Yan SC. Object region mining with adversarial erasing: A simple classification to semantic segmentation approach. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017. 6488–6496. [doi: [10.1109/CVPR.2017.687](https://doi.org/10.1109/CVPR.2017.687)]
- [27] Li X, Zhao LM, Wei LN, Yang MH, Wu F, Zhuang YT, Ling HB, Wang JD. DeepSaliency: Multi-task deep neural network model for salient object detection. IEEE Trans. on Image Processing, 2016, 25(8): 3919–3930. [doi: [10.1109/TIP.2016.2579306](https://doi.org/10.1109/TIP.2016.2579306)]
- [28] Liu N, Han JW. DHSNet: Deep hierarchical saliency network for salient object detection. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 678–686. [doi: [10.1109/CVPR.2016.80](https://doi.org/10.1109/CVPR.2016.80)]
- [29] Wang YT, Ren TW, Zhong SH, Liu Y, Wu GS. Adaptive saliency cuts. Multimedia Tools and Applications, 2018, 77(17): 22213–22230. [doi: [10.1007/s11042-018-5859-y](https://doi.org/10.1007/s11042-018-5859-y)]
- [30] Luo ZM, Mishra A, Achkar A, Eichel J, Li SZ, Jodoin PM. Non-local deep features for salient object detection. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 6593–6601. [doi: [10.1109/CVPR.2017.698](https://doi.org/10.1109/CVPR.2017.698)]
- [31] Zhang PP, Wang D, Lu HC, Wang HY, Ruan X. Amulet: Aggregating multi-level convolutional features for salient object detection. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision (ICCV). Venice: IEEE, 2017. 202–211. [doi: [10.1109/ICCV.2017.31](https://doi.org/10.1109/ICCV.2017.31)]
- [32] Wu Z, Su L, Huang QM. Cascaded partial decoder for fast and accurate salient object detection. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 3902–3911. [doi: [10.1109/CVPR.2019.00403](https://doi.org/10.1109/CVPR.2019.00403)]
- [33] Liu ST, Huang D, Wang YH. Receptive field block net for accurate and fast object detection. In: Proc. of the 15th European Conf. on Computer Vision (ECCV). Munich: Springer, 2018. 404–419. [doi: [10.1007/978-3-030-01252-6_24](https://doi.org/10.1007/978-3-030-01252-6_24)]
- [34] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: Proc. of the 31st AAAI Conf. on Artificial Intelligence. San Francisco: ACM, 2017. 4278–4284. [doi: [10.5555/3298023.3298188](https://doi.org/10.5555/3298023.3298188)]
- [35] Holschneider M, Kronland-Martinet R, Morlet J, Tchamitchian P. A Real-time Algorithm for Signal Analysis with the help of the Wavelet Transform. In: Combes JM, Grossmann A, Tchamitchian P, eds. Wavelets. Berlin: Springer, 1990. 286–297. [doi: [10.1007/978-3-642-75988-8_28](https://doi.org/10.1007/978-3-642-75988-8_28)]
- [36] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834–848. [doi: [10.1109/TPAMI.2017.2699184](https://doi.org/10.1109/TPAMI.2017.2699184)]
- [37] Zhang RF, Li GB, Li Z, Cui SG, Qian DH, Yu YZ. Adaptive context selection for polyp segmentation. In: Proc. of the 23rd Int'l Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI). Lima: Springer, 2020. 253–262. [doi: [10.1007/978-3-030-59725-2_25](https://doi.org/10.1007/978-3-030-59725-2_25)]
- [38] Codella NCF, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, Kalloo A, Liopyris K, Mishra NK, Kittler H, Halpern A. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In: Proc. of the 15th IEEE Int'l Symp. on Biomedical Imaging (ISBI 2018). Washington: IEEE, 2018. 168–172. [doi: [10.1109/ISBI.2018.8363547](https://doi.org/10.1109/ISBI.2018.8363547)]
- [39] Bernal J, Sánchez FJ, Fernández-Esparrach G, Gil D, Rodríguez C, Vilariño F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized Medical Imaging and Graphics, 2015, 43: 99–111. [doi: [10.1016/j.compmedimag.2015.02.007](https://doi.org/10.1016/j.compmedimag.2015.02.007)]
- [40] Jha D, Smedsrød PH, Riegler MA, Halvorsen P, De Lange T, Johansen D, Johansen HD. Kvasir-SEG: A segmented polyp dataset. In: Proc. of the 26th Int'l Conf. on Multimedia Modeling. Daejeon: Springer, 2020. 451–462. [doi: [10.1007/978-3-030-37734-2_37](https://doi.org/10.1007/978-3-030-37734-2_37)]
- [41] Tomar NK, Jha D, Riegler MA, Johansen HD, Johansen D, Rittscher J, Halvorsen P, Ali S. FANet: A feedback attention network for

- improved biomedical image segmentation. arXiv: 2103.17235, 2021.
- [42] Zhang ZX, Liu QJ, Wang YH. Road extraction by deep residual U-Net. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 749–753. [doi: [10.1109/LGRS.2018.2802944](https://doi.org/10.1109/LGRS.2018.2802944)]

附中文参考文献:

- [1] 宋杰, 肖亮, 练智超, 蔡子贊, 蒋国平. 基于深度学习的数字病理图像分割综述与展望. 软件学报, 2021, 32(5): 1427–1460. <http://www.jos.org.cn/1000-9825/6205.htm> [doi: [10.13328/j.cnki.jos.006205](https://doi.org/10.13328/j.cnki.jos.006205)]
- [20] 殷晓航, 王永才, 李德英. 基于U-Net结构改进的医学影像分割技术综述. 软件学报, 2021, 32(2): 519–550. <http://www.jos.org.cn/1000-9825/6104.htm> [doi: [10.13328/j.cnki.jos.006104](https://doi.org/10.13328/j.cnki.jos.006104)]



王雪(1982—), 女, 博士生, 主要研究领域为计算机视觉, 图像处理, 多媒体融合.



陈海鹏(1978—), 男, 博士, 教授, CCF 专业会员, 主要研究领域为机器学习, 视觉推理.



李占山(1966—), 男, 博士, 教授, CCF 专业会员, 主要研究领域为约束优化与约束求解, 机器学习, 图像处理.