

中图法分类号: TP183 文献标识码: A 文章编号: 1006-8961(2020)10-2195-11

论文引用格式: Hu Y S, Qin P L, Zeng J C, Chai R and Wang L F. 2020. Ultrasound thyroid segmentation based on segmented frequency domain and local attention. Journal of Image and Graphics 25(10): 2195-2205 [胡屹杉, 秦品乐, 曾建潮, 柴锐, 王丽芳. 2020. 结合分段频域和局部注意力的超声甲状腺分割. 中国图象图形学报 25(10): 2195-2205] [DOI: 10.11834/jig.200230]

结合分段频域和局部注意力的超声甲状腺分割

胡屹杉^{1 2 3}, 秦品乐^{1 2 3}, 曾建潮^{1 2 3}, 柴锐^{1 2 3}, 王丽芳^{1 2 3}

1. 中北大学山西省医学影像与数据分析工程研究中心, 太原 030051; 2. 中北大学大数据学院, 太原 030051;
3. 中北大学山西省医学影像人工智能工程技术研究中心, 太原 030051

摘要: 目的 超声检查是诊断甲状腺疾病的主要影像学方法之一,但由于超声图像中斑点强度具有随机性、组织器官复杂等问题,导致甲状腺在不同数据源间的形态、大小和纹理差异性较大,容易导致观察者视觉疲劳。针对甲状腺超声成像存在斑点强度随机性以及周边组织复杂性的问题,为了更准确地描述出器官与病理性病变的解剖边界,提出一种基于频域增强和局部注意力机制的甲状腺超声分割网络。方法 针对原始数据采用高通滤波器获取高低频段的图像信息,整合高频段细节特征与低频段边缘特征,增强图像前背景的对比度,降低图像间的差异性。根据卷积网络中网络深度所提取特征信息量的不同,采用局部注意力机制对高低维特征信息进行自适应激活,增强低维特征的细节信息,弱化对非目标区域的关注,增强高维特征的全局信息,弱化冗余信息对网络的干扰,增强前背景分类以及对非显著性目标检测的能力。采用金字塔级联空洞卷积获取不同感受野的特征信息,解决数据源间图像差异较大的问题。结果 实验结果表明,本文方法在 11 ~ 16 MHz 时采集的 16 个手绘甲状腺超声公开数据集中,通过 10 折交叉验证显示准确率为 0.989,召回率为 0.849,精准率为 0.940, Dice 系数为 0.812,效果优于当前其他医学图像分割网络。通过消融实验,证明本文的几个模块对超声图像分割确实具有一定的提升效果。结论 本文所提分割网络,结合深度学习模型及传统图像处理模型的优点,能较好地处理超声图像随机斑点并且提升非显著性组织分割效果。

关键词: 图像分割; 频域分析; 注意力机制; 空洞卷积; 超声影像

Ultrasound thyroid segmentation based on segmented frequency domain and local attention

Hu Yishan^{1 2 3}, Qin Pinle^{1 2 3}, Zeng Jianchao^{1 2 3}, Chai Rui^{1 2 3}, Wang Lifang^{1 2 3}

1. Shanxi Medical Imaging and Data Analysis Engineering Research Center, North University of China, Taiyuan 030051, China;
2. College of Big Data, North University of China, Taiyuan 030051, China; 3. Shanxi Medical Imaging Artificial Intelligence Engineering Technology Research Center, North University of China, Taiyuan 030051, China

Abstract: **Objective** Ultrasound is a main imaging method used for the diagnosis of thyroid diseases. It is convenient for the diagnosis of medical results through the real-time study of its internal anatomical structure. In computer vision, the segmentation of image tissue and organ is the pre background classification of the pixels in the image. The final segmentation image boundary is the combination of the target pixels. The research on medical image segmentation has received much

收稿日期: 2020-05-30; 修回日期: 2020-07-07; 预印本日期: 2020-07-14

基金项目: 山西省工程技术研究中心建设项目(201805D121008); 山西省研究生创新项目(2020SY381); 中北大学研究生科技立项项目(20191634)

Supported by: Construction Project of Engineering Technology Research Center of Shanxi Province (201805D121008); Innovation Project of Graduate Students of Shanxi Province (2020SY381)

attention , which is mainly divided into two ideas , where the first idea is to obtain the target area by analyzing the pixel value of a given image through computer vision technology. However , the generalization ability of the given image analysis is poor , and the segmentation effect is unremarkable because of the interference of random noise in the ultrasonic image. The second idea is to use deep learning for obtaining the target area through the background information before deep convolution classification. However , the target area may be insignificant using the depth learning model because of the complexity of tissue and organs , the evident surrounding tissues , and the lack of background information before the image , making the abstract features obtained by the depth network mostly the surrounding non target area and causing the segmentation effect of the original target unideal. A thyroid image is different in shape , size , and texture among different data sources. To solve the two problems , a thyroid ultrasound segmentation network based on frequency domain enhancement and local attention mechanism is proposed to solve the problem of random noise interference and insignificant target. **Method** First , high and low pass filters are used to obtain the image information of high- and low-frequency bands , and the detail features of high frequency band and the edge feature of low frequency band are integrated to enhance the contrast of background and reduce the difference between images. Second , a local attention mechanism is used to adaptively activate the high- and low-dimensional feature information in accordance with the different information amounts of the features extracted by the network depth in the convolution network. This mechanism can enhance the detailed information of low-dimensional features , weaken the attention to nontarget areas , enhance the global information of high-dimensional features , and weaken the interference of redundant information on the network , thereby enhancing the ability of background classification and nonsignificant target detection. Finally , a pyramid cascading hole is used , and convolution is utilized to obtain the feature information of different receptive fields and solve the problem of large image difference between data sources. In the training process , a mixed loss function is used to regress the network training effect , and pixel level loss (binary cross entropy) and image similarity loss (structural similarity) can better evaluate the segmentation prediction results. This paper uses the ResNet34 network , which is trained in advance to fine tune , to train the model of the network. The training set adopts the open data set of the network and selects approximately 3 500 images through the screening of appropriate images. During the training , one NVIDIA P100 graphics processing unit (GPU) server is used , the network training of approximately 10 epochs can achieve a better and stable effect , and the total training time is approximately 120 min. **Result** Experimental results show that the accuracy of the proposed method is 0.989 , the recall rate is 0.849 , the specificity is 0.94 , and the Dice coefficient is 0.812 , which is better than the current methods of medical image segmentation network , such as U-Net and CE-Net network , and is more accurate and special in the effect of ultrasound thyroid image segmentation. A significant improvement is found in heterosexuality and is better than the evaluation result for the network using the same dataset , such as sumNet. At the same time , the ablation experiments show that the proposed modules have a certain improvement effect on ultrasound image segmentation. **Conclusion** The proposed segmentation model combined with the advantages of deep learning model and traditional image processing model can better deal with ultrasound image random spots and improve the results of non-significant tissue segmentation.

Key words: image segmentation; frequency domain analysis; attention mechanism; dilate convolution; ultrasound image

0 引 言

甲状腺是人体内分泌系统的腺体 , 相关疾病大多涉及甲状腺形状和大小随时间发生的变化 , 因此甲状腺的分割和体积重建对于诊断相关疾病至关重要。测量甲状腺体积是诊断其病理变化的常规做法 , 也是评判其病理变化的可靠指标 , 而就常规的甲状腺检测而言 , 超声 (ultrasound , US) 检查是主要检测方法。超声是一种无创伤、方便和安全的检查手

段。通过超声图像 , 可以找到甲状腺结节的位置 , 测量其大小 , 并确保病患在手术前获得有关甲状腺结节的大量信息 , 从而决定是否需要手术。

随着医疗影像技术的不断进展 , 计算机辅助诊断 (computer-aided design , CAD) 已可以在一定程度上辅助医生进行诊断分析。其中 , 医学图像分割研究受到了较多关注。彭文献等人 (2017) 结合灰度共生矩阵和灰度梯度矩阵的统计学特征识别 CT (computed tomography) 上的甲状腺结节。迟剑宁等人 (2018) 融合深度网络和浅层纹理特征诊断甲状

腺结节癌变情况。Quan 等人(2013)将 Ncut(normalized cut) 方法应用于合并过分割的区域,处理超声图像的强噪声、斑点及弱边缘问题。Zhuang 等人(2019)采用基于超声的模糊增强技术与 FBM(fractional Brownian motion) 模型相结合,增强其超声图像的对比度与分割效果。Yan 等人(2017)使用不同计算机图像处理方法来实现超声血管腔壁的腔壁分离。然而在临床中,存在多种干扰因素,如超声反馈的斑点强度出现随机模式,成像和成像过程中呼吸和其他身体运动(如医生徒手探头扫描部位以及病人的身体运动等)引入的伪影,不同成像机器根据反馈信号成像结果不同等,这些因素往往导致超声图像分割结果并不理想。

随着深度学习不断发展,采用神经网络建立医学影像的分割模型得到广泛应用。Goyal 等人(2020)使用全卷积网络(fully convolutional network, FCN) 对黑色素瘤、角膜以及良性病变进行训练,使得网络能够同时对 3 种目标进行分割。Tran(2017)首次提出用于心脏磁共振图像(magnetic resonance imaging, MRI) 左、右心室分割的 FCN,采用全卷积网络的方法有效提高了对于 MRI 图像的心脏分割能力。Poudel 等人(2017)将类似 U-Net 的网络结构与门控循环单元(gated recurrent unit, GRU) 结合,提出 RFCN(recurrent fully convolutional networks),该网络模型利用 2D 切片间的空间依赖性,改善了对心尖部分左心室外膜的分割。Alom 等人(2018)将 U-Net、残差网络结构与循环神经网络结构结合,提出了 RU-Net(residual U-Net) 和 R2U-Net(recurrent residual U-Net),并使用该网络进行视网膜图像中的血管分割、皮肤癌分割和肺损伤分割。Nandamuri 等人(2019)通过一种全卷积神经网络 SUMNet 实现对超声甲状腺的快速分割。Lian 等人(2017)使用神经网络获取包含胆结石的目标区域,再使用区域生长的方法修正分割结果。

上述方法通过卷积神经网络提取图像特征信息,通过分析图像内部的深层隐性信息划分目标区域与非目标区域,使其在生物医学图像分割任务上实现较高的性能。然而对于超声甲状腺图像来说,由于超声成像斑点强度的随机性和低信噪比导致结构间的对比度降低,同时甲状腺组织周边器官的干扰,导致目标区域分割不理想。本文针对该问题,采用影像组学的方法,探索深度学习对甲状腺超声图

像处理方面的有效性,通过对超声影像的纹理特征分析,引入图像频域知识,利用卷积网络提取高低维特征信息量不同的特点,以及空洞卷积获取更多感受野,从而增强获取高维语义信息的能力,设计特征增广模块与局部注意力模块。本文的主要工作有:

1) 设计了一个图像增广模块,结合传统超声图像的纹理分析特点与频域图像的知识,对于既定的甲状腺图像进行低频与高频信号处理,从而获取较好的边缘信息与纹理信息,改善不同超声数据源间差异性较大的问题。

2) 设计了一个级联多尺度空洞卷积金字塔特征提取结构。利用多尺度空洞卷积获取不同感受野下的特征信息,在保证分辨率的情况下,利用金字塔结构整合不同感受野下的特征信息,增强网络对不同尺度目标特征的语义信息提取,有效解决分割目标大小不统一、纹理信息复杂的问题。

3) 设计了一个局部注意力模块,采用空间注意力机制与通道注意力机制,针对网络各层中所提取特征信息量的不同,采用局部引入的方式,激活特征响应增强图像中关键信息的获取,提高网络对于前背景区域的分辨能力,提升非显著目标区域的分割能力。

4) 以 ResNet34 作为网络特征提取器,再采用对称形式构建特征解码器,采用 U-Net 的编码器—解码器结构,构建单阶段的端到端网络。

1 相关工作

医学图像分割研究受到了极大关注,主要分为两种思路:1) 通过计算机视觉技术分析既定图像的像素值,从而获取目标区域;2) 采用深度学习,通过深度卷积层分类前背景信息,从而获取目标区域。

1.1 图像频域分析

分析超声图像像素值获取目标区域,除了对像素值灰度值信息的分析外,还提取甲状腺纹理进行分析(Poudel 等 2019),建立图像信号参数化模型,根据不同建模纹理频带之间的能量比来区分甲状腺和非甲状腺纹理。纹理分析是超声图像分析中用于结构分割和组织分类的主要方法,它不局限于传统图像的灰度值信息分析,开始逐渐转向图像频域信息的分析,以频域信息提取纹理特征,从而进一步提升分割效果。

1.2 神经网络高低维度信息分析

随着卷积神经网络层数的增加,不同深度的特征对于图像内容的抽象程度不同,根据特征信息量的不同(Zhao 和 Wu 2019),特征逐渐由低级表示演化成高级表示,其低维度特征更多是空间信息,包含图像的纹理及空间位置信息。高维特征包含更多是全局图像的语义信息,通过将高维特征采用上采样的方式进行降维操作,可视化后显示更多的是目标图像的近似区域。对于 U-Net 分割网络,在使用低维特征还原细节特征时,应该保证高低层特征在保留自身的情况下进行互相细化,避免混合。

2 本文方法

2.1 网络结构

本文提出基于频域增强和局部注意力机制的甲状腺分割网络,结构如图 1 所示,采用 U-Net 网络的编码-解码结构,其编码器采用 ResNet34 预训练模型的卷积层参数进行迁移学习提取图像纹理特征,通过添加级联多尺度空洞卷积,保证图像特征分辨率,并获取不同感受野的上下文信息。然后根据不同高低维度特征信息量包含内容的不同,解码器采用编码器最大池化索引引导还原高维特征的细节及位置信息,采用反卷积操作还原低维像素点间的上下文语义信息,同时采用通道注意力机制和空间注意力机制对其特征信息进行自适应激活响

应,增强目标非显著差异性。最后将响应激活的编码器特征采用跳层连接添加到特征还原模块,实现网络整体结果的残差连接,还原图像细节信息,提高分割区域的完整性与精准度。

本文网络结构包含 3 部分:频域特征增广(frequency-domain feature augmentation, FFA) 模块、级联多尺度空洞卷积金字塔增广(cascade multi-scale dilated convolution pyramid feature augmentation module, CMDP) 模块和局部注意力(local attention module, LA) 模块。频域特征增广模块通过对图像高低频域过滤整合频域信息,解决超声图像数据源间数据差异较大以及图像边缘模糊,纹理信息复杂的问题。级联多尺度空洞卷积金字塔增广模块在保证图像分辨率的同时,通过获取不同感受野下的特征信息,获取更宽泛的语义特征信息,增强网络对于不同尺度形态目标的分割能力。局部注意力模块根据高低维度特征信息包含纹理特征信息不同,分别采用通道注意力与空间注意力,增强对关键信息的获取,并提升网络对非显著性目标区域的分割能力。

2.2 频域特征增广模块

图像分割任务主要基于纹理分析原理划分目标区域,通过纹理特征描述图像属性,量化纹理信息,测量图像像素间的平滑度、规则性及粗糙度,然后采用分类或聚类的方法提取特征信息,估计其图像特征及图像相似度信息,从而达到图像分割的目的。在超声图像中,由于随机斑点和伪影的引入,其图像

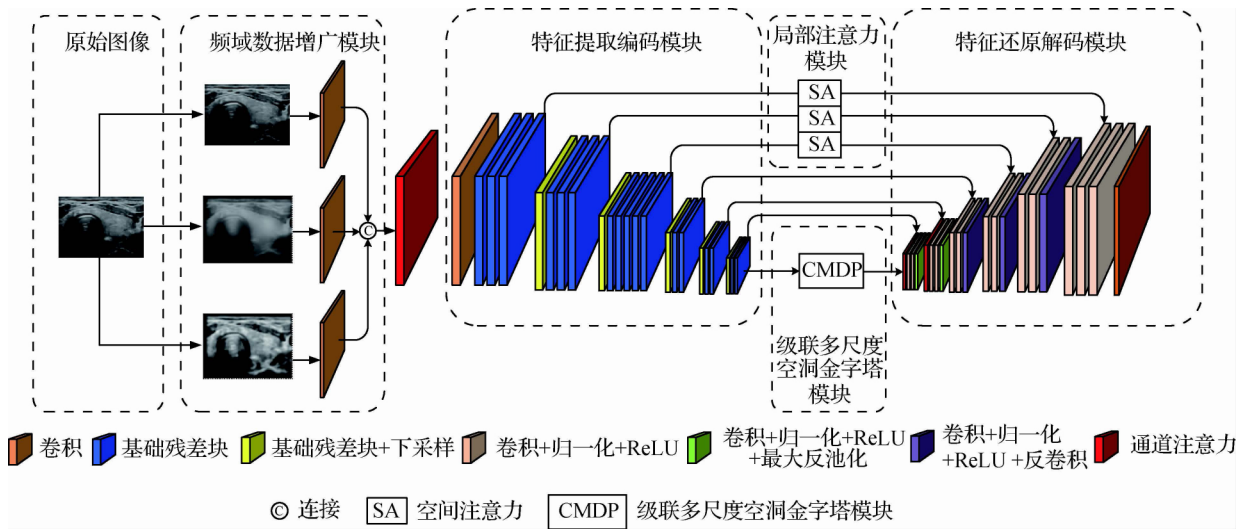


图 1 网络结构图

Fig. 1 Overall network structure

纹理信息模糊以及边界不明显,导致图像分割不理想,因此超声图像的纹理分析来应不局限于像素值灰度值信息的分析,还可引入图像频域信息来分析图像纹理信息。通过构建图像信号参数化模型,可根据不同频带之间的能量比来区分超声组织部分(Illanes等2019)。如图2所示,分析原始超声图像并将其灰度值信息转换为信号信息,通过高低通滤波器过滤频域信号,可观察出图像频域中的高频分

量对应图像的细节信息(图2(c)为周边信号增强),而图像低频分量对应图像的整体轮廓信息(图2(b)为中心信号增强)。如图3所示,整合高低频域图像信息及原始超声图像,获取超声增强图像,通过频域增广原始超声图像特征,加强图像纹理特征及边缘信息,同时采用直方图均衡法,增强图像各组织间的对比度,降低图像间的差异性,增强网络对于超声数据的训练的适配性。

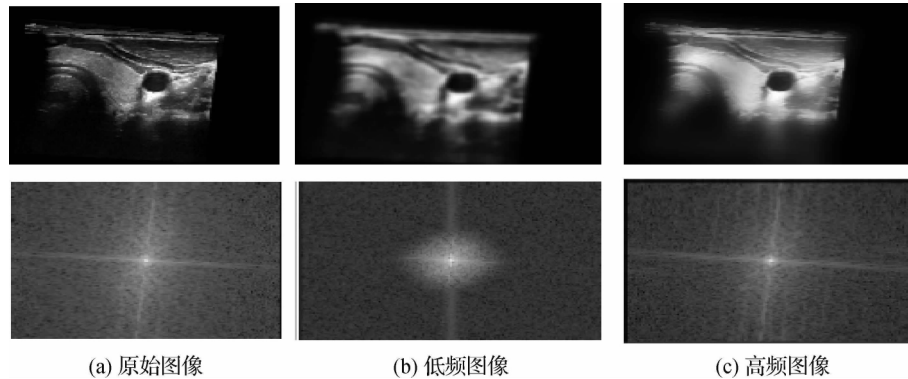


图2 高低频域分析

Fig. 2 High and low frequency domain analysis((a) original; (b) low frequency; (c) high frequency)

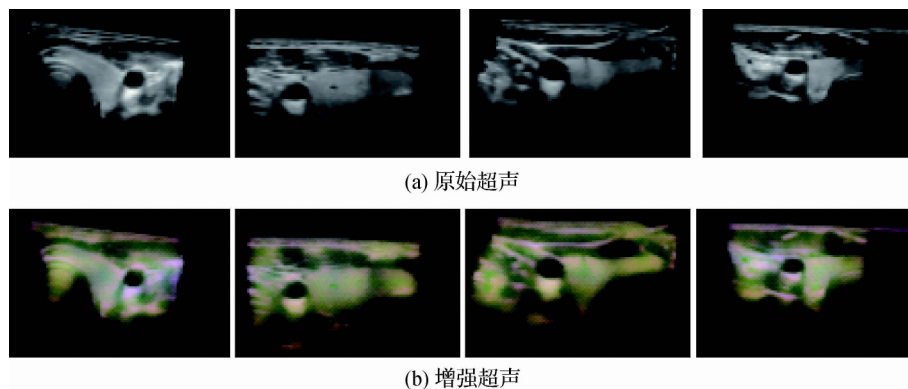


图3 原始超声与增强超声图像对比

Fig. 3 Comparison of original ultrasound and enhanced ultrasound images((a) original; (b) enhanced)

2.3 级联多尺度空洞卷积金字塔增广模块

图像分割任务需获取具有高分辨率的高级语义信息,即对于编码器而言应尽可能获取含有较多上下文语义的特征信息。由于特征提取时受限于有限的感受野,导致学习到的特征无法包含丰富的上下文信息来准确检测尺度、形状和位置等多变的目标(李晓光等2019)。如图4所示,不同空洞率下的空洞卷积的感受野大小不同(Wang等2018)。通过组合不同空洞率的空洞卷积,可获取不同感受野信息,如图5所示。采用空洞率为1、3、5的空洞卷积逐

步叠加可获取空洞率为3、9、19的感受野,通过级联不同空洞率的空洞卷积获取不同感受野下的特征信息,参考Inception结构,整合不同感受野下的全局特征信息,便于网络获取各种形态大小的对象的特征。

2.4 局部注意力模块

由于超声甲状腺图像周边的组织复杂,可能导致周边组织较为显著地影响甲状腺分割效果。因此,需较好地划分前背景信息,并增强方法的非显著性目标检测能力。解码器将高维全局上下文语义特征信息逐步还原细节纹理及边缘信息,在此过程中

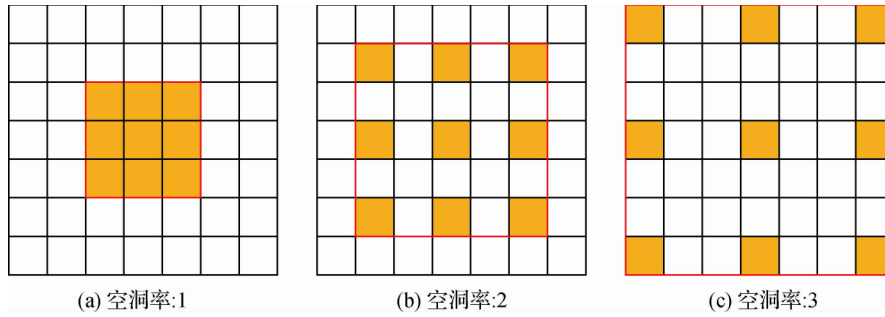


图4 不同空洞率的空洞卷积感受野对比

Fig. 4 Comparison of receptive fields of dilated convolutions with different dilated rates
((a) dilated rate is 1; (b) dilated rate is 2; (c) dilated rate is 3)

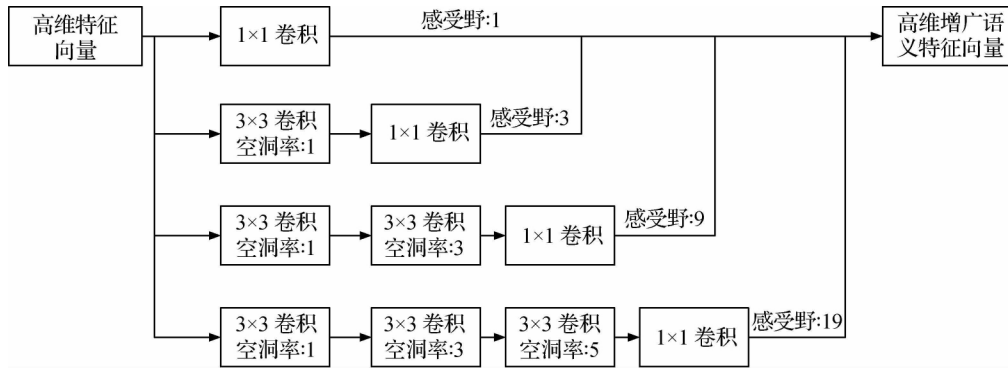


图5 级联多尺度空洞卷积金字塔结构

Fig. 5 Cascaded multi-scale dilated convolution pyramid structure

采用跳层连接引导编码器特征还原细节信息,但在编码器中其不同深度的特征对于图像内容的抽象程度不同,对于网络前几层而言,其更多地获取图像的边界纹理信息(Qin等2020)。如果跳层连接只是简单的特征拼接或特征相加来合并高级语义信息和低级的细节信息,会传递嘈杂的信息降低分割效果,因此高低层特征之间应在保留自身的情况下进行互相细化。随着网络层数的加深(Zhao和Wu2019),特征的尺寸逐步缩小,通道数逐步叠加,由低级演化成高级,其低维度特征更多是空间信息,包含图像的纹理及空间位置信息。高维度特征更多是语义信息,包含全局图像的内在含义。因此,对于低维度特征,需筛选过滤特征中无关区域的纹理信息,通过引入空间注意力机制,关注各个特征所提取出的感兴趣区域,通过与真实框的对比,响应激活目标区域的特征区域,弱化其他非目标区域(显著性的干扰区域),从而增强图像的对于目标区域的特征提取。高维度特征包含的主要是图像语义信息,通过引入通道注意力机制,关注各个特征层间的关联性,即通过响应图像的全局信息,对其高维度特征进行激活响应,从语义的角度,加强网络对于图像目标的分类判断,降低相似

性组织的干扰。因此,本文设计局部注意力模块,就其高低维度特征采用不同的注意力机制进行响应激活映射,在降低网络参数的同时,尽可能获取更为有效的特征信息,便于后期解码模块对于图像特征还原细节时,降低其他显著性特征的干扰。

2.4.1 通道注意力机制

为了反映不同特征通道对于关键信息响应程度的不同,采用通道注意力机制增强其对关键信息通道的响应程度。用 $f^h \in \mathbf{R}^{W \times H \times C}$ 表示长、宽、深度分别为 W 、 H 、 C 维的高维度语义特征, $\mathbf{R}^{W \times H}$ 表示 $W \times H$ 维空间, $\nu^h \in \mathbf{R}^C$ 表示深度为 C 维的高维语义向量。如图6所示,采用全局平均池化与全局最大池化压缩反馈其特征在通道的数值信息,即 $f^h \in \mathbf{R}^{W \times H \times C}$ 采用全局平均池化(Zhou等2016)和全局最大池化的方式压缩 $\mathbf{R}^{W \times H}$ 上的特征信息,叠加通道维度上的特征向量 $\nu^h \in \mathbf{R}^C$,该特征向量即可反映各通道间对于不同特征区域的响应程度,为了限制模型复杂度和辅助泛化,Hu等人(2018)采用两个连续的全连接层计算并完全捕获通道的可信度,之后采用Sigmoid激活函数,对映射到 $[0, 1]$ 的编码信道特征向量进行归一化处理。

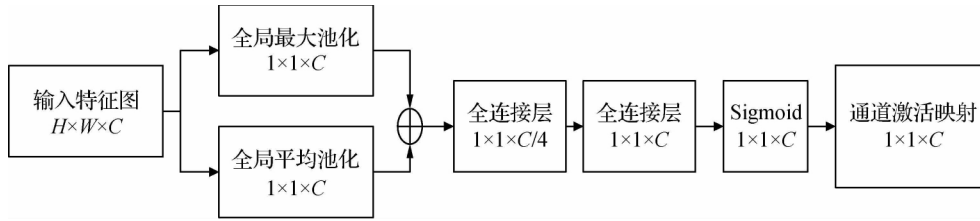


图6 通道注意力结构

Fig. 6 Channel attention structure

将特征向量经过两个全连接层再响应激活获取到不同通道的可信度 CA , 即

$$CA = F(\nu^h) = \sigma(f_c(\delta(f_c(\nu^h)))) \quad (1)$$

式中, σ 表示 Sigmoid 激活函数, f_c 表示全连接层 (fully convolutional, FC), δ 表示 ReLU 函数, 获取到的 CA 即为反馈不同通道特征对于关键信息的响应值, 最终将通道特征响应值与原特征相乘, 即可获取增强目标区域的显著性特征通道。

2.4.2 空间注意力机制

空间注意力主要是关注特征图上的关键信息, 对于复杂环境的图像信息, 应尽量减少其他显著性目标区域的影响, 同时需增强图像边缘信息, 关注显著性物体与背景间的详细边界。因此设计了针对图像边缘信息及感兴趣区域增强的空间注意力模块。用 $f^1 \in \mathbf{R}^{W \times H \times C}$ 表示长、宽、深度分别为 W, H, C 维的低维度特征, 如图7所示。由于全局平均池化 (global average pooling, GAP) 是将位置的响应都赋予相同的权重值, 可能导致对目标区域的高估。而对于全局最大池化, 只有最大响应位置的值被考虑,

其关注对象大多无法完整拓展。因此, 对于低维特征信息 $f^1 \in \mathbf{R}^{W \times H \times C}$ 而言, 采用全局最大池化 (global max pooling, GMP) 获取其关键位置的响应区域, 为弥补其他关注对象的拓展区域及边界纹理信息, Peng 等人 (2017) 同时为了获取空间关注点及显著性边缘信息, 应用一个内核为 $1 \times K$, 另一个为 $K \times 1$ 的卷积层, 通过叠加响应区域以及关注点的边缘信息 C_1 与 C_2 , 然后采用 Sigmoid 激活函数, 对映射到 $[0, 1]$ 的编码空间特征采取归一化处理措施, 计算为

$$C_1 = \text{conv}_2(\text{conv}_1(f^1)) \quad (2)$$

$$C_2 = \text{conv}_1(\text{conv}_2(f^1)) \quad (3)$$

$$SA = F(f^1) = \sigma(C_1 + C_2 + f^{\text{GMP}}) \quad (4)$$

式中, conv_1 和 conv_2 分别表示 $1 \times K$ 和 $K \times 1$ 卷积核, f^{GMP} 表示将特征采用全局最大池化压缩维度的特征向量, K 为常数。实验结果表明, 设置 $K = 5$, 获得的 SA 即为反馈各特征图的显著性边缘的响应值, 将空间特征响应值与原特征相乘, 即可获取最终对于目标区域的感兴趣区域及边缘增强特征信息。

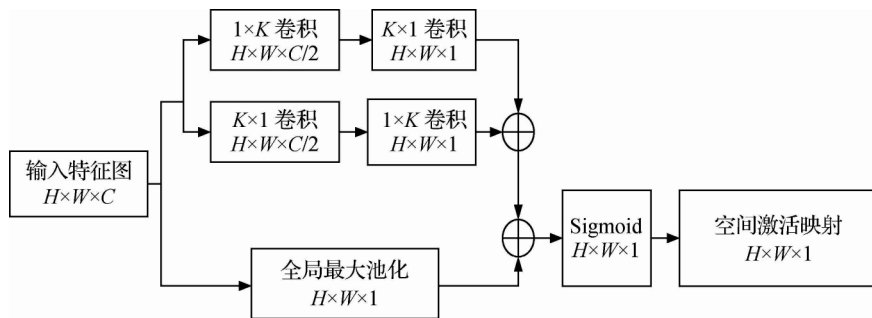


图7 空间注意力结构

Fig. 7 Spatial attention structure

本文设计的空间注意力机制在弥补全局最大池化细节信息不足的同时, 增强图像对于显著性区域的边界信息获取, 更好地获取图像的边界信息。同时增强显著性目标区域的全局上下文信息, 从而增强网络对于非显著性目标区域的分类。

3 损失函数

图像分割更多的是逐个判断像素是否为目标像素点, 即为一个逐像素间的分类问题。通常使用交

叉熵损失函数来评估图像分割的效果,但对超声图像而言,由于随机斑点和伪影问题,其关注的区域所占比例差异性较大,且边界不清晰导致其人工标注的真实框可能存在一定误差,交叉熵对此情况的分割没有较好的评判训练效果。因此,本文引入像素细节损失与轮廓相似损失,采用交叉熵损失(binary cross entropy loss, BCE)与结构相似性损失(structural similarity, SSIM) (Wang 等 2003),能够更好地分割显著性对象区域并且对分割对象边缘信息有引导作用。如下所示,定义损失 L 为

$$L = \alpha L_{\text{BCE}} + (1 - \alpha) L_{\text{SSIM}} \quad (5)$$

式中,超参数 α 为平衡系数,用于平衡细节损失及轮廓损失对最终结构的影响,本文中选取 $\alpha = 0.6$ 。 L_{BCE} 与 L_{SSIM} 分别定义为 BCE 损失和 SSIM 损失。BCE 损失通过对比前背景像素的平均加权,加速细节信息的收敛。其定义为

$$L_{\text{BCE}} = - \sum [G(r, c) \log(S(r, c)) + (1 - G(r, c)) \log(1 - S(r, c))] \quad (6)$$

式中, $G(r, c) \in \{0, 1\}$ 表示第 (r, c) 像素的真实值,其标签值为 0 或 1, $S(r, c)$ 表示预测目标的可能性。

SSIM 损失通过分配较高权重给边界邻域,使边界的损失权重增加,边界上的预测概率与前景其余部分相同。用 $x = \{x_j | j = 1, \dots, N^2\}$ 与 $y = \{y_j | j = 1, \dots, N^2\}$ 表示从预测图像 S 和二值真实图像 G 上分别裁取的两个大小为 $N \times N$ 的相关补丁的向量,其中 x_j 与 y_j 表示第 j 个像素的值, N^2 表示 $N \times N$, j 取 $1 \sim N^2$ 的个数。SSIM 损失定义为

$$L_{\text{SSIM}} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (7)$$

式中, μ_x 和 μ_y 分别为 x, y 的均值, σ_x 和 σ_y 分别为 x, y 的标准方差, σ_{xy} 为它们的协方差, C_1 和 C_2 为设置的常量参数。

4 实验与结果分析

本文使用的数据集主要为公开获得的数据集 (Wunderling 等 2017), 其中包括以 11 ~ 16 MHz 探针成像的 16 位健康人的, 通过手绘获得的甲状腺超声图像。在本次实验中使用 3 424 幅甲状腺图像, 每

幅图像上所标注的轮廓信息都较为完整。为满足实际的临床医学图像条件, 仅采用随机左右翻转图像作为数据增广方法。训练时采用 1 台 NVIDIA P100 GPU 服务器, ResNet34 为骨干网络前期特征提取的迁移学习网络, 通过 Fine-tune 方法调整特征提取参数, 该网络训练约 10 个 epoch 就可达到较好的平稳效果, 训练总时长为 120 min 左右。本文采用以下定量指标说明分割的准确度及精度: 准确度 (accuracy), 召回率 (recall), 精准率 (precision), Dice 分数 (Dice coefficient)。通过上述指标评估本文模型和其他模型的性能, 同时对特征增广模块和注意力模块的结构合理性进行消融实验, 以验证结构设计的合理性。

4.1 方法对比实验

首先对本文方法和其他医学图像分割算法进行性能对比实验。本文采用传统的基于 ResNet34 为特征提取器的 U-Net 网络作为实验基准, 与多种医学图像分割任务中适应性强的 CE-Net (Gu 等, 2019)、BASNet (Qin 等 2019)、nnU-Net (Isensee 等, 2018) 和针对超声甲状腺分割的 SUMNet (Nandamuri 等 2019) 进行对比实验。如表 1 所示, 相对于其他算法, 本文算法在各个评估指标下均有较好的提升, 尤其在 Dice 系数上有明显的提升, 本文算法对于超声甲状腺图像分割问题有较好的解决效果。

表 1 不同方法的实验结果对比
Table 1 Comparison of experimental results of different methods

对比方法	准确度	召回率	精准率	Dice
U-Net	0.962	0.793	0.911	0.751
Ce-Net	0.979	0.841	0.932	0.798
BSANet	0.976	0.838	0.931	0.773
nnU-Net	0.972	0.834	0.926	0.769
SUMNet	0.988	0.846	0.933	0.796
本文	0.989	0.849	0.941	0.812

注: 加粗字体为每列最优值。

4.2 消融实验

为验证本文设计的特征增广模块的合理性及有效性, 采用对比消融实验, 所有实验训练配置均相同。针对本文频域特征增广模块 (FFA), 级联多尺度空洞金字塔模块 (CMDP) 及局部注意力模块 (LA), 采用消融实验证明该结构在实验中的有效性

及合理性。本文设计6个对比实验,基于ResNet34为特征提取器的U-Net网络为基础网络,依次叠加上上述3个模块,如表2所示。使用频域数据增广和高维语义增广在一定程度上可以提升网络对超声图像的分割能力,在准确度、特异性及召回率上也均改进了初始网络U-Net的分割结果。本文所设计的局部注意力模块能提升网络分割准确度与精准率,表明其对网络关键特征的提取有一定的引导意义,同时也可以进一步优化高维语义增广和频域数据增广所增加的有效精准的特征信息,有效解决图像数据差异性 & 图像非显著分割的问题。

表2 网络结构消融对比

Table 2 Comparison of network structure ablation

对比方法	准确度	召回率	精准率	Dice
U-Net	0.962	0.793	0.912	0.751
U-Net + FFA	0.975	0.834	0.933	0.783
U-Net + CMDP	0.977	0.839	0.931	0.778
U-Net + LA	0.981	0.836	0.936	0.792
U-Net + FFA + LA	0.982	0.836	0.937	0.801
U-Net + CMDP + LA	0.984	0.838	0.937	0.805
U-Net + FFA + CMDP + LA	0.989	0.849	0.94	0.812

注:加粗字体表示每列最优结果。

4.3 损失函数对比实验

为验证本文所提损失函数的有效性,采用像素细节损失和轮廓相似损失结合的混合损失函数和交叉熵损失函数进行对比实验。如图8所示,混合损

失函数在保留原有交叉熵损失函数的精准性上,引入相似性轮廓进一步优化分割边缘,保证分割边缘较为平滑,降低因超声图像随机斑点和人为标注所引起的误差。

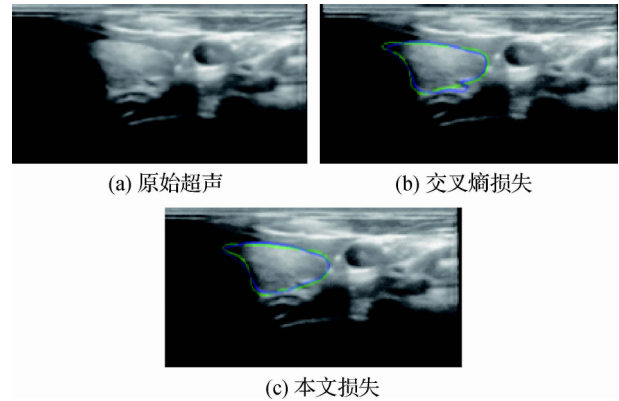


图8 损失函数分割结果对比

Fig. 8 Comparison of segmentation results of loss function
((a) original ultrasound image; (b) BCE loss; (c) ours)

4.4 测试实验结果

如图9所示,对比下列图像分割结果。第1行为甲状腺组织明显显著;第2行为甲状腺组织非显著,周边组织明显;第3行为甲状腺边缘模糊且形态大小差异变化较大的分割结果。通过和CE-Net, BASNet, SUMNet方法的结果对比可以看出,本文方法在甲状腺组织显著的情况下细化分割结果,同时对于非显著目标分割来说其分割效果较好,能较好地应对甲状腺边界模糊和适应形态差异变化大的问题,解决存在显著性周边组织干扰的问题。

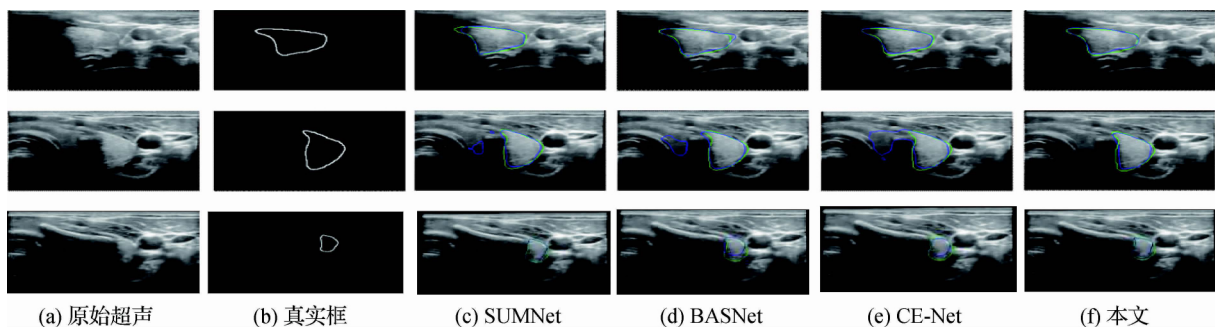


图9 不同算法分割组织结果对比

Fig. 9 Comparison of segmentation results of different algorithms

((a) original ultrasound images; (b) ground truth; (c) SUMNet; (d) BASNet; (e) CE-Net; (f) ours)

5 结 论

针对现有医学分割算法对超声甲状腺图像适应

性差和非显著性目标分割不理想的问题,提出一种基于高低频域分析和高低维度特征分层处理的超声甲状腺分割模型。对于既定的超声图像采用高低频域信息进行特征融合,增强图像的对对比度,同时减少

数据源间的差异性,根据不同网络深度的网络层所提取的特征信息量的不同,采用局部注意力机制,优化浅层特征的细节信息和深层特征的全局信息,增强网络对于前背景的分离能力,从而提高图像非显著目标的分割结果。在测试阶段,为了避免超声图像随机噪点和人工标注引起的误差,算法结合像素级损失和轮廓相似性损失用以更新网络训练的分割效果,获取更为精确和平滑的轮廓结果。实验结果表明,在公开的基准数据集上,依据4种性能评价指标,本文方法与目前的医学图像分割网络相比在分割的精度和准确性上有较好的性能。在主观视觉上,本文能在优化超声甲状腺分割图像精度的同时较好地处理非显著性目标的分割任务。

由于甲状腺贴于喉和气管颈段的两侧,且分布在气管的两侧,其周边组织差异性较大,对于过小的甲状腺组织其分割效果有所局限。在今后的研究中,将考虑利用多尺度目标检测及边界检测模块提升模型对于小物体图像的检测能力。

参考文献(References)

- Alom M Z, Yakopcic C, Taha T M and Asari V K. 2018. Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net) // Proceedings of NAECON 2018 IEEE National Aerospace and Electronics Conference. Dayton: IEEE: 228-233 [DOI: 10.1109/NAECON.2018.8556686]
- Chi J N, Yu X S and Zhang Y F. 2018. Thyroid nodule malignantrisk detection in ultrasound image by fusing deep and texture features. *Journal of Image and Graphics*, 23(10): 1582-1593 (迟剑宁,于晓升,张艺菲. 2018. 融合深度网络和浅层纹理特征的甲状腺结节癌变超声图像诊断. *中国图象图形学报*, 23(10): 1582-1593) [DOI: 10.11834/jig.180232]
- Goyal M, Yap M H and Hassanpour S. 2020. Multi-class semantic segmentation of skin lesions via fully convolutional networks[EB/OL]. [2020-03-13]. <https://arxiv.org/pdf/1711.10449.pdf>
- Gu Z W, Cheng J, Fu H Z, Zhou K, Hao H Y, Zhao Y T, Zhang T Y, Gao S H and Liu J. 2019. CE-Net: context encoder network for 2D medical image segmentation. *IEEE Transactions on Medical Imaging*, 38(10): 2281-2292 [DOI: 10.1109/TMI.2019.2903562]
- Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks // Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]
- Illanes A, Esmaili N, Poudel P, Balakrishnan S and Friebe M. 2019. Parametrical modelling for texture characterization-a novel approach applied to ultrasound thyroid segmentation. *PLoS One*, 14(1): e0211215 [DOI: 10.1371/journal.pone.0211215]
- Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger P F, Kohl S, Waserthaler J, Köhler G, Norajitra T, Wirtk S and Maier-Hein K H. 2018. nnU-Net: self-adapting framework for U-Net-based medical image segmentation[EB/OL]. [2018-09-27]. <https://arxiv.org/pdf/1809.10486.pdf>
- Li X G, Fu C P, Li X L and Wang Z H. 2019. Improved faster R-CNN algorithm for multi-scale target detection. *Journal of Computer-Aided Design and Computer Graphics*, 31(7): 1095-1101 (李晓光,付陈平,李晓莉,王章辉. 2019. 面向多尺度目标检测的改进 Faster R-CNN 算法. *计算机辅助设计与图形学学报*, 31(7): 1095-1101) [DOI: 10.3724/SP.J.1089.2019.17283]
- Lian J, Ma Y D, Ma Y R, Shi B, Liu J Z, Yang Z and Guo Y N. 2017. Automatic gallbladder and gallstone regions segmentation in ultrasound image. *International Journal of Computer Assisted Radiology and Surgery*, 12(4): 553-568 [DOI: 10.1007/s11548-016-4515-z]
- Nandamuri S, China D, Mitra P and Sheet D. 2019. SUMNet: fully convolutional model for fast segmentation of anatomical structures in ultrasound volumes // The 16th IEEE International Symposium on Biomedical Imaging. Venice: IEEE: 1729-1732 [DOI: 10.1109/ISBI.2019.8759210]
- Peng C, Zhang X Y, Yu G, Luo G M and Sun J. 2017. Large kernel matters-improve semantic segmentation by global convolutional network // Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 1743-1751 [DOI: 10.1109/CVPR.2017.189]
- Peng W X, Liu C B, Xia S E, Chen Y H and Liu R. 2017. Statistic texture feature based thyroid nodule recognition on CT images. *Space Medicine and Medical Engineering*, 30(4): 258-262 (彭文献,刘晨彬,夏顺仁,陈益红,刘蕊. 2017. 基于 CT 图像统计纹理特征的甲状腺结节识别技术. *航天医学与医学工程*, 30(4): 258-262) [DOI: 10.16289/j.cnki.1002-0837.2017.04.005]
- Poudel P, Illanes A, Ataide E J G, Esmaili N, Balakrishnan S and Friebe M. 2019. Thyroid ultrasound texture classification using autoregressive features in conjunction with machine learning approaches. *IEEE Access*, 7: 79354-79365 [DOI: 10.1109/ACCESS.2019.2923547]
- Poudel R, Lamata P and Montana G. 2017. Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation // Zuluaga M A, Bhatia K, Kainz B, Moghari M H and Pace D F, eds. *Reconstruction, Segmentation, and Analysis of Medical Images*. Cham: Springer: 83-94 [DOI: 10.1007/978-3-319-52280-7_8]
- Qin P L, Wu K, Hu Y S, Zeng J C and Chai X F. 2020. Diagnosis of benign and malignant thyroid nodules using combined conventional ultrasound and ultrasound elasticity imaging. *IEEE Journal of Biomedical and Health Informatics*, 24(4): 1028-1036 [DOI: 10.1109/JBHI.2019.2950994]
- Qin X B, Zhang Z C, Huang C Y, Gao C, Dehghan M and Jagersand M.

2019. BASNet: boundary-aware salient object detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) . Las Vegas: IEEE: 7479-7489 [DOI: 10.1109/CVPR.2019.00766]
- Quan L ,Zhang D ,Yang Y ,Liu Y and Qin Q Q. 2013. Segmentation of tumor ultrasound image via region-based Ncut method. Wuhan University Journal of Natural Sciences ,18(4) : 313-318 [DOI: 10.1007/s11859-013-0934-8]
- Tran P V. 2017. A fully convolutional neural network for cardiac segmentation in short-axis MRI [EB/OL]. [2020-05-01]. <https://arxiv.org/pdf/1604.00494.pdf>
- Wang P Q ,Chen P F ,Yuan Y ,Liu D ,Huang Z H ,Hou X D and Cottrell G. 2018. Understanding convolution for semantic segmentation//Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision (WACV 2018) . Lake Tahoe: IEEE: 1451-1460 [DOI: 10.1109/WACV.2018.00163]
- Wang Z ,Simoncelli E P and Bovik A C. 2003. Multiscale structural similarity for image quality assessment//Proceedings of the 37th Asilomar Conference on Signals , Systems and Computers. Pacific Grove: IEEE: 1398-1402 [DOI: 10.1109/ACSSC. 2003.1292216]
- Wunderling T ,Golla B ,Poudel P ,Arens C ,Friebe M and Hansen C. 2017. Comparison of thyroid segmentation techniques for 3D ultrasound//Proceedings Volume 10133 , Medical Imaging 2017: Image Processing. Orlando: SPIE ,10133: 1013317 [DOI: 10.1117/12.2254234]
- Yan J Y ,Lv D and Cui Y Y. 2017. A novel segmentation approach for intravascular ultrasound images. Journal of Medical and Biological Engineering ,37(3) : 386-394 [DOI: 10.1007/s40846-017-0233-5]
- Zhao T and Wu X Q. 2019. Pyramid feature attention network for saliency detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) . Long Beach: IEEE: 3080-3089 [DOI: 10.1109/CVPR.2019.00320]
- Zhou B L ,Khosla A ,Lapedriza A ,Oliva A and Torralba A. 2016. Learning deep features for discriminative localization//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . Las Vegas: IEEE: 2921-2929 [DOI: 10.1109/CVPR.2016.319]
- Zhuang Z M ,Lei N H ,Raj A N J and Qiu S M. 2019. Application of fractal theory and fuzzy enhancement in ultrasound image segmentation. Medical and Biological Engineering and Computing ,57(3) : 623-632 [DOI: 10.1007/s11517-018-4907-z]

作者简介



胡屹杉,1996年生,男,硕士研究生,主要研究方向为计算机视觉、医学影像分析。
E-mail: hu996716708@gmail.com



曾建潮,通信作者,男,教授,主要研究方向为复杂系统的维护决策和健康管理。
E-mail: zjc@nuc.edu.cn

秦品乐,男,教授,主要研究方向为医学影像分析、机器视觉。
E-mail: QPL@nuc.edu.cn

柴锐,男,讲师,主要研究方向为医学影像处理。
E-mail: chairui20@126.com

王丽芳,女,副教授,主要研究方向为机器视觉、医学图像处理。
E-mail: wsm2004@nuc.edu.cn