

级联结构网络在肺癌患者颈部淋巴结超声图像分割中的应用价值

宫霞 赵富强 吴卫华

摘要 为了准确诊断肺癌转移,本文应用深度学习技术对肺癌患者颈部淋巴结超声图像病灶区域进行分割,提出了一种用于超声图像分割的级联注意力 UNet 网络,该级联结构是将注意力 UNet 与 EfficientNet 相结合的二阶段分割网络,第一阶段为粗分割,第二阶段为细分割,编码器采用 EfficientNet-B5 作为主干网,图像多尺度输入;提出了适用于小目标、小样本场景的新损失函数;试验结果表明,本文提出的级联结构网络在肺癌患者颈部淋巴结超声图像分割中网络性能优异,Dice 系数达到 0.95,较其他 UNet 方法具有更优的分割性能。

关键词 图像分割;超声图像;注意力机制;级联

[中图分类号]R445.1

[文献标识码]A

Application value of cascade structure network in ultrasound image segmentation of cervical lymph nodes in patients with lung cancer

GONG Xia, ZHAO Fuqiang, WU Weihua

Department of Ultrasound, Affiliated Chest Hospital of Shanghai Jiao Tong University, Shanghai 200030, China

ABSTRACT In order to precisely diagnose the metastasis of lung cancer, a deep learning technology to segment cervical lymph node in ultrasound images of patients with lung cancer was applied in this paper, and proposed a cascade attention UNet network for ultrasound image segmentation. The cascade structure was a two-stage segmentation network combining attention UNet and EfficientNet. The first stage was coarse segmentation and the second stage was fine segmentation. The encoder adopted EfficientNet-B5 as the backbone network. The multi-scale features of the image were taken as the input. A new loss function was proposed, which was suitable for small target and few-shot scenarios. The experimental results showed that the proposed cascade structure had excellent network performance in cervical lymph node ultrasonic image segmentation, and the Dice coefficient was 0.95, which had better segmentation performance than other UNet methods.

KEY WORDS Image segmentation; Ultrasound image; Attention mechanism; Cascade

在传统的医学影像分析中,超声具有成本低、辐射小等优点,应用广泛,但其图像较 CT、MRI 图像的分辨率低,分析相关医疗影像较耗时、耗力,且主观性强。近年来随着深度学习技术发展,使得计算机分析医疗影像成为可能,大量的图像分类、图像分割和目标检测等试验^[1-4]表明,计算机应用于智能医疗影像后其分析速度和精度均优于临床医师。深度学习中的数据增强、迁移学习和对抗网络等技术可辅助临床医师进行超声图像分析,对提高分析图像的效率 and 准确性,尤其在区分肺癌转移性淋巴结病变和良恶性病变方面具有重要的临床研究意

义。肺癌转移非常关键的一项鉴别诊断是锁骨上淋巴结转移,通过计算机辅助诊断肺癌患者颈部淋巴结超声图像,分割出淋巴结,进而对淋巴结进行良恶性分类,不但在理论上丰富了现有的研究方法,其在临床上也有很好的应用前景。

在超声图像分割中,应用最广的深度学习模型是 UNet^[1] 网络,其变形版本包括:V-Net^[2]、mu-Net^[3]、M-Net^[4]等,但它们均未采用多尺度训练,在医疗图像小目标和小样本分割上精度较低。UNet 改进的网络结构包括:Attention UNet^[5]、UNet++^[6]、UNet3+^[7]等,这些网络在分割精度上得到了很大提高,注意力

基金项目:上海市卫生健康委员会科研课题面上项目(201940494)

作者单位:200030 上海市胸科医院 上海交通大学附属胸科医院超声科(宫霞、吴卫华);上海慧邸智能科技中心(赵富强)

通讯作者:吴卫华,Email: fqzhao@126.com

机制、多尺度训练、特征融合和混合损失函数等技术值得借鉴。王波等^[8]将注意力机制应用于 UNet 网络,提出了一种基于改进 UNet 网络的甲状腺结节超声图像分割方法,Attention 模块对边缘轮廓分割细化;Transformer 在 NLP 上取得了很大成功,近期有学者将其引入到目标检测和图像分割中,提出了 Swin Transformer^[9],其结合 UNet 提出了 TransUNet^[10]和 Swin-UNet^[11]。Swin Transformer 根据网络大小,可分为 Tiny、Small、Base 和 Large 等级别,网络参数和结构依次增大,所需训练硬件资源也相应增加,这使得 Transformer 在临床应用受到限制。为此,本文提出了基于级联结构改进注意力 UNet 网络,改进了混合损失函数,图像多尺度输入进行训练和推理,更适合小目标的分割,针对病灶边缘区域的分割更加清晰。

一、UNet 与 EfficientNet 网络

超声图片输入 UNet 网络后,先通过左侧的下采样,即 4 组卷积,每次下采样图片缩小 2 倍;然后特征图上采样,对应右侧 4 组反卷积,左侧每层特征图与右侧对应层拼接生成上采样的输入,保留了该尺度图像特征。

深度学习中的 Backbone 网络大多采用 ResNet 结构,很好地解决了网络的深度问题,但对网络的宽度却无能为力,因此,谷歌 2019 年提出了 EfficientNet^[12] 网络。EfficientNet 包括 B0~B7,训练的模型也依次增大,该网络结构能够兼顾网络深度、宽度和分辨率,使模型的准确率、鲁棒性等相关指标均得到了很大提高。

二、级联注意力 UNet(CAUNet)

Cascade R-CNN^[13]是级联多个检测网络进行模型训练,检测的正负样本通过 IOU 值限定,多个网络级联使得网络的检测结果得到优化和提高。与 Cascade 级联不同,本文提出的 CAUNet(Cascade Attention UNet)网络结构是将注意力 UNet 与 EfficientNet 结合进行改进,级联结构如图 1 所示。改进的注

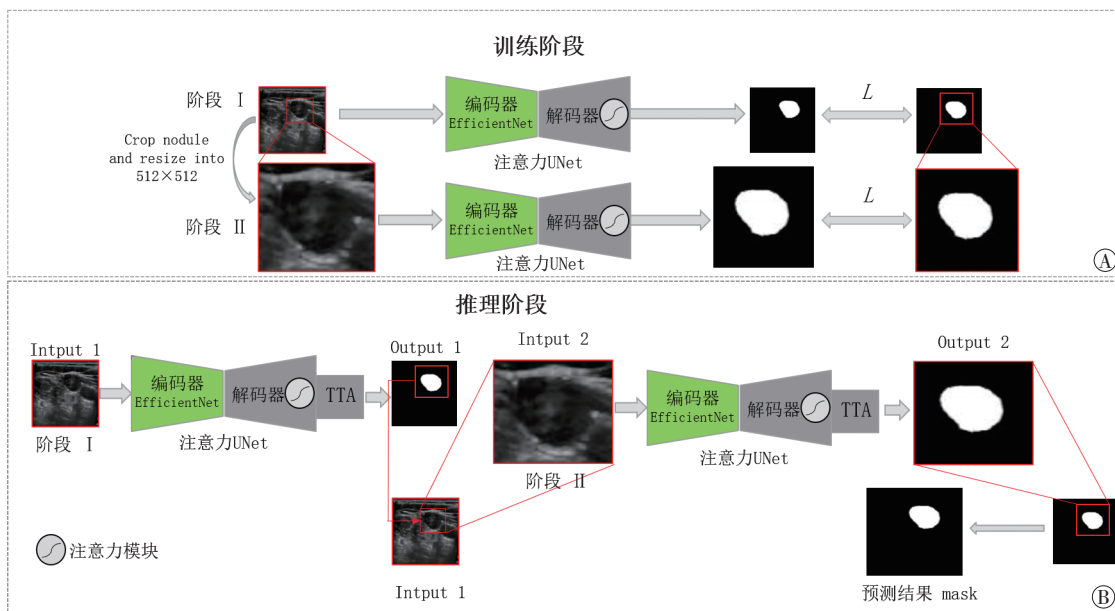
意力 UNet 包括编码器和解码器两部分;其中编码器采用 EfficientNet-B5 作为 Backbone,更利于提取超声图像病灶区域特征,第一阶段分割网络和第二阶段分割网络下采样均采用该编码器结构,上采样是增加了注意力模块的解码器,构成二阶段级联分割网络。在训练阶段,将采集的超声图像数据预处理后输入到第一个分割网络(级联的阶段 I),CAUNet 利用改进的损失函数对网络进行训练,该网络提供颈部淋巴结结节的粗略分割;在第一阶段粗略分割的基础上,继续训练二次分割网络(级联的阶段 II),该网络进行颈部淋巴结结节的精细分割。但训练的输入数据与第一阶段不同,首先对一阶段训练的图像去除非病灶区域,针对 mask 掩码轮廓取外接矩形,裁剪出该区域图像;然后将该图像尺度放大到 512×512(像素),输入到二阶段网络进行训练。

在推理阶段,图片多尺度输入,一阶段同样先进行整张图片分割,然后将分割病灶区域放大,再进行二阶段的精细分割。推理阶段如果使用 TTA 技术,病灶区域分割效果更优。注意力 UNet 为图 1 网络结构中“编码器和解码器”部分,UNet 下采样部分不变,上采样部分引入注意力模块,其详细结构组成见图 2。输入编码器的颈部淋巴结超声图片采用多尺度信息,输入图像分为 4 个尺度,每个尺度在本层使用 EfficientNet-B5 作为 Backbone,提取的特征分别作为下采样和右侧注意力门的输入,将图像的底层 feature map 与高层 feature map 融合,即 g 与 x' 依次进行加、ReLU、 ψ 、Sigmoid、重采样、乘等操作,对病灶区域的边界分割准确率提高更有效。

三、损失函数改进

1.Generalized Dice 损失函数

Dice loss 在评价小目标病灶区域分割时的误差较大,原因如下:如模型对超声图像中较小的病灶区域预测错误,则 Dice loss 值会出现幅度比较大的波动,反向传播的梯度变化相应增



A: 训练阶段;B: 推理阶段
图 1 CAUNet 网络结构图

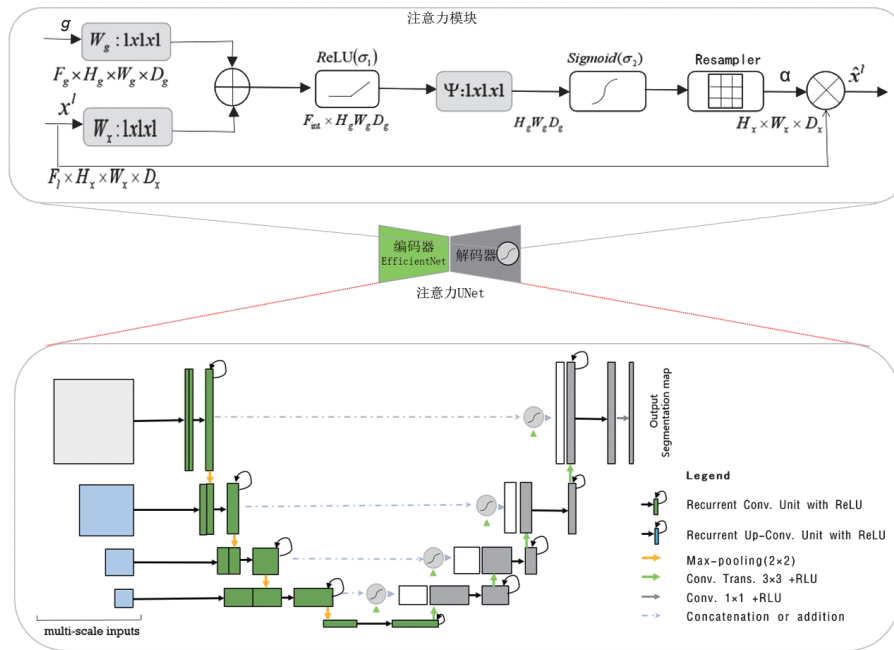


图2 注意力UNet网络结构图

大,致使模型训练稳定差。当超声图像病灶分割有多个区域时,使用Generalized Dice loss 计算公式如下:

$$L_{GDice} = 1 - \sum_k \frac{2\omega_k \sum_i p(k,i)g(k,i)}{\sum_i p^2(k,i) + \sum_i g^2(k,i)} \quad (1)$$

其中, N 为像素个数, $\omega_k = \frac{N - N_k}{N}$ 为第 k 个类别的权重, $\sum_k \omega_k = 1$ 。该权重的引入可以增加病灶区域所占比重。

2. 多尺度结构相似度损失函数

由于超声图像清晰度相对较低,分割颈部淋巴结病灶区域的边界模糊,为了解决该问题,Wang等^[14]提出了多尺度结构相似度损失函数(Multi-Scale Structural Similarity Index, MS-SSIM),模糊边界的权重更大。本文级联结构模型同样对模糊边界更加关注,如超声图像区域分布差异越大,多尺度结构相似度损失值越大。病灶区域分割预测结果 p 与裁剪了标签 mask g 对应病灶区域的外接正方形($N \times N$)公式分别表示为:

$$p = \{p; j = 1, \dots, N^2\} \quad (2)$$

$$g = \{g; j = 1, \dots, N^2\} \quad (3)$$

多尺度结构相似度 loss 定义如公式(4)所示:

$$L_{ms-ssim} = 1 - \prod_{m=1}^M \left(\frac{2\mu_p \mu_g + C_1}{\mu_p^2 + \mu_g^2 + C_1} \right)^{\beta_m} \left(\frac{2\sigma_{pg} + C_2}{\sigma_p^2 + \sigma_g^2 + C_2} \right)^{\gamma_m} \quad (4)$$

其中, μ_p 为 p 的均值, μ_g 为 g 的均值, σ_p 为 p 的方差, σ_g 为 g 的方差, σ_{pg} 为协方差, M 为尺度数, β_m 、 γ_m 分别为在 m 尺度时 p 、 g 的权重, C_1 、 C_2 为取值很小的常数。

结合 Generalized Dice 损失函数^[15]、正则化损失函数、MS-SSIM 损失函数 $L_{ms-ssim}$ 和 Focal 损失函数 L_{Focal} ^[16], 定义本文总损失函数为:

$$L = L_{GDice} + L_{Rg} + L_{ms-ssim} + \lambda L_{Focal} \quad (5)$$

其中, λ 为 L_{Focal} 损失函数所占权重, 其取值为 0、0.1、0.5 或 1; L_{Rg} 表示正则化损失函数。公式(5)损失函数有效避免在数据不平衡时, 组合 loss 退化为 Dice loss, 而且对超声图像小目标病灶区域有较好的分割效果, 病灶边界更加清晰^[17]。

四、数据增强方法

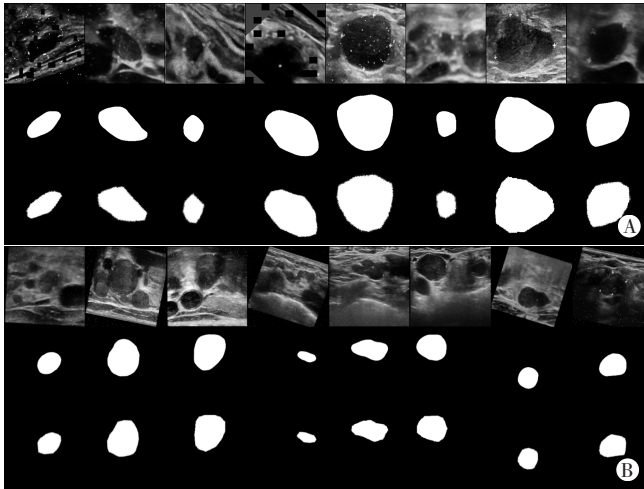
当收集的肺癌患者淋巴结超声图像数据较少时, 通常采用数据增强(Data Augmentation)技术, 该技术可以有效弥补数据不足来训练模型, 防止过拟合。本文采用的数据增强方法包括: 水平翻转、垂直翻转、随机剪切、随机仿射变换、中心裁剪和随机旋转($<90^\circ$ 、 90° 、 180° 和 270°); 此外, 还通过调整对比度、模糊(高斯、平均值、中值)、高斯噪声和 Cutout 等数据增强方法, 得到增强后的图片共 4545 张。

试验环境是基于深度学习 Pytorch 框架, GPU 服务器配置为: 操作系统为 Ubuntu 18.04, CPU 为 Intel i7-8700K, 内存 64 GB; 显卡为英伟达 RTX 1080Ti、显卡显存 11 GB。为了提高分割模型的泛化能力, 本文采用了 TTA 技术, 包括用于垂直翻转、水平翻转、 90° 和 180° 旋转。第 60 个 Epoch 训练和验证结果见图 3。

五、实证分析

1. 数据来源

本文数据来源于上海市胸科医院超声科 440 例肺癌患者的 510 张淋巴结超声图像, 其中男 240 例, 淋巴结图像 260 张; 女 200 例, 淋巴结图像 250 张。采集的每张超声图像均由专业超声医师扫描完成, 并标注病灶长径和短径, 用于训练的超声图像病灶区域由 3 名经验丰富的超声医师使用 Labelme 共同标注, 诊断结果均与病理结果相对应, 确保数据的正确可靠; 患者的淋巴结均进行针吸细胞学检查及细针穿刺活检。两个阶段使用的数据不同, 先去掉参与训练图像的不相关区域, 在第一阶段, 即上述采集的超声图像, 原图和掩码图的分辨率均为



A: 训练结果; B: 验证结果

图3 模型训练和验证分割结果

256×256, 三通道读取; 在第二阶段, 病灶区域图像与掩码图均相应放大, 分辨率为 512×512, 掩码图为二值化图像, 病灶区域为白色, 背景为黑色。

2. 模型测试与结果分析

将采集的颈部淋巴结超声图像数据集按照 8:1:1 划分为训练集、验证集和测试集, 采用 Adam 优化器。使用 Dice 系数对超声图像的分割结果进行评估。通过图 4, 5 分析发现, UNet++ I 和 II 阶段损失函数收敛较 CAUNet 网络慢, 第二阶段的 loss 值较第一阶段更低; 由图 6 可知 UNet++ Dice 系数低于 0.92, 本文提出的级联二阶段 CAUNet 网络 Dice 系数达到了 0.95, 且第二阶段优于第一阶段, 说明第二阶段对关键病灶区域的放大训练和分割是有效的。可见, 级联二阶段 CAUNet 网络分割效果优于 UNet++, 将提高颈部淋巴结良恶性分类的准确率。测试集结果见图 7, 从左往右依次为: 原图、预测结果、标注结果。与其他 UNet 网络比较见表 1, 该表列出了不同网络结构平均交并比 (IOU)、Dice 系数和曲线下面积 (AUC) 3 个指标比较, 验证了本文提出的方法对病灶区域分割最优, 上级联 CAUNet 在 TTA 技术下平均 IOU 较 UNet 网络提升了 14%, Dice 系数提升了 10%, AUC 提升了 3%, 而 UNet3+ 和 Swin-UNet 的 Dice 系数均为 0.93。同样, 与其余对比算法相比, 本文的模型在各项指标上也均有较大改进。本文 Swin-UNet 采用 Base 级网络结构, 跳跃连接数目为 3。

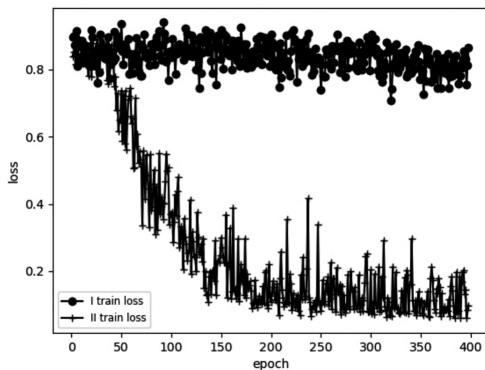


图4 UNet++ I 和 II 阶段损失函数

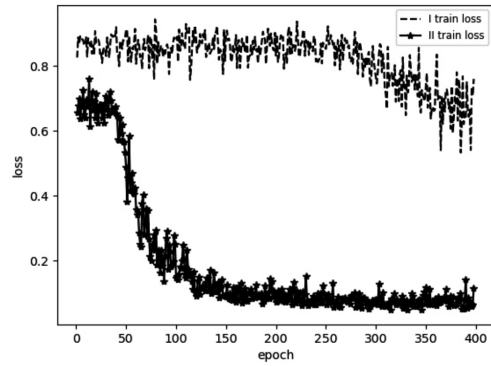


图5 CAUNet I 和 II 损失函数

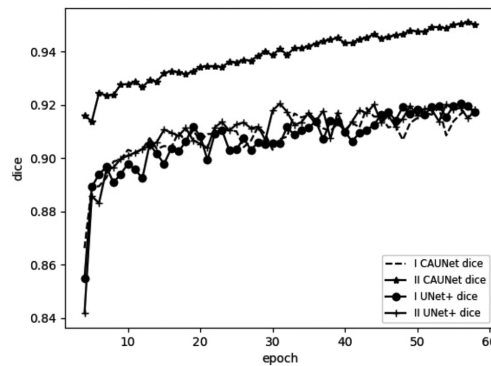


图6 CAUNet与UNet++网络 I、II 阶段Dice系数

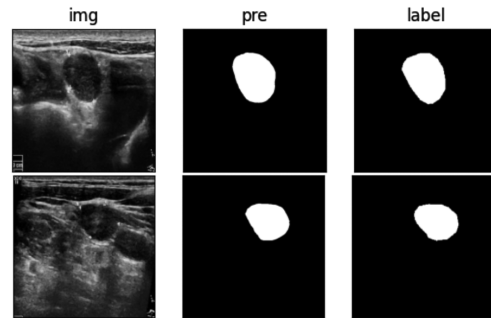


图7 淋巴结病灶分割

表1 不同网络结构量化指标比较

网络结构	平均IOU	Dice系数	AUC
UNet	0.78	0.85	0.96
AUNet	0.88	0.91	0.98
R2AttUNet	0.87	0.90	0.97
UNet++	0.89	0.92	0.97
UNet3+	0.90	0.93	0.97
Swin-UNet	0.89	0.93	0.97
Ours			
I	0.86	0.90	0.95
II (TTA)	0.92	0.95	0.99

IOU: 交并比; AUC: 曲线下面积

工具箱 torchstat.stat, 可以计算训练模型的算法复杂度、参数量 (可训练参数量和总参数量)、GPU 服务器占用量等, 得到的不同网络结构参数量见表 2。CAUNet 网络是两个阶段训练和推理, 因此, 较 UNet 和注意力 UNet 网络在参数和计算资源消耗上均要大, 如需进一步提高实时性, 可通过增加 GPU 数量

表2 不同网络结构参数量比较

网络结构	参数量(个)	GFLOPs
UNet	3 352 802	31.42
AUNet	3 440 395	32.25
R2AttUNet	39 442 990	267.23
UNet++	47 176 132	113.85
Ours	24 436 369	7.84

实现。

本文结果表明,CAUNet网络能够快速、准确分割肺癌患者颈部淋巴结超声图像。图像分割结果的敏感性为95%,特异性为96%,准确率为92%,AUC为0.99,表明本文提出的新模型是有效的,有助于区分肺癌患者转移性淋巴结与良性淋巴结。

六、结论

本文提出了基于注意力UNet网络级联EfficientNet-B5的二阶段超声图像分割模型,针对颈部淋巴结超声图像进行增强,多尺度训练;为了提高病灶小目标分割性能和小样本训练模型,提出了新的损失函数,有效防止网络的过拟合,网络的鲁棒性得到提高;采用了数据增强扩充训练数据,试验结果表明该网络提高了Dice系数。今后需进一步深入探讨的内容包括:①对颈部淋巴结超声图像分割出来的区域,需要分类网络进一步预测淋巴结病灶区域的良恶性;②在本文基础上,进一步优化损失函数、对DeepLabv3+网络采用级联结构深入研究;③将GAN和UNet结合,提出生成对抗UNet,可应用于本文的图像数据增强,改善模型的性能。

参考文献

- [1] Ronneberger O, Fischer P, Brox T. UNet: convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-assisted Intervention. IEEE, 2015: 234-241.
- [2] Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016: 565-571.
- [3] Seo H, Huang C, Bassenne M, et al. Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images[J]. IEEE Trans Med Imag, 2019, 39(5): 1316-1325.
- [4] Mehta R, Sivaswamy J. M-Net: a convolutional Neural network for deep brain structure segmentation[C]// IEEE International Symposium on Biomedical Imaging. IEEE, 2017: 437-440.
- [5] Schlemper J, Ozan O, Schaap M, et al. Attention gated networks: learning to leverage salient regions in medical images[J]. Med Image Anal, 2019, 53(1): 197-207.
- [6] Zhou Z, Siddiquee MMR, Tajbakhsh N, et al. UNet++: a nested UNet architecture for medical image segmentation [C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham, 2018: 3-11.
- [7] Huang H, Lin L, Tong R, et al. UNet 3+: a full-scale connected unet for medical image segmentation[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020: 1055-1059.
- [8] 王波, 李梦翔, 刘侠. 基于改进U-Net网络的甲状腺结节超声图像分割方法[J]. 电子与信息学报, 2022, 44(2): 514-522.
- [9] Liu Z, Lin Y, Cao Y, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows [C]//Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2021: 10012-10022.
- [10] Attia R. Wildfire segmentation using deep vision transformers [J]. Remote Sensing, 2021, 13(7): 2072-4292.
- [11] Liang J, Cao J, Sun GK, et al. SwinIR: Image restoration using swin transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2021: 1833-1844.
- [12] Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks [C]//UNet: Convolutional networks for biomedical image segmentation, 2019: 6105-6114.
- [13] Cai Z, Vasconcelos N. Cascade R-CNN: delving into high quality object detection [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.
- [14] Wang Z, Bovik AC, Sheikh HR, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Trans Image Process, 2004, 13(4): 600-612.
- [15] Jadon S. A survey of loss functions for semantic segmentation [C]//2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology. IEEE, 2020: 27-29.
- [16] Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection [J]. IEEE Trans Pattern Anal Mach Intell, 2020, 42(2): 318-327.
- [17] Li L, Zhao X, Lu W. Deep learning for variational multimodality tumor segmentation in PET/CT [J]. Neuro Computing, 2020, 392(1): 277-295.

(收稿日期: 2022-03-03)