

## 基于卷积神经网络和Transformer的肝脏CT图像分割方法

胡晓阳, 李哲

沈阳理工大学自动化与电气工程学院, 辽宁 沈阳 110000

**【摘要】**针对现有的卷积神经网络在肝脏图像分割上精度较低的问题,提出了一种以U-Net网络模型为基础的分割算法。将多头自注意力机制引入到U-Net网络的跳跃连接中,在编码器部分使用空洞卷积,采用混合损失函数从而提高分割精度。在LITS数据集上通过实验结果表明,利用本文方法进行肝脏分割与传统U-Net方法相比Dice系数提升3.3%,平均交并比提升了2.4%,平均像素准确率提升了3.66%。

**【关键词】**卷积神经网络;肝脏;图像分割;多头自注意力机制;空洞卷积

**【中图分类号】**R318;TP391

**【文献标志码】**A

**【文章编号】**1005-202X(2023)04-0423-06

### Liver CT image segmentation method based on CNN and Transformer

HU Xiaoyang, LI Zhe

School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110000, China

**Abstract:** Aiming at the problem of low accuracy of the existing convolutional neural network in liver image segmentation, a segmentation algorithm based on U-Net model is presented. The segmentation accuracy is improved using multi-head self-attention mechanism which was introduced into the skip connection of U-Net, atrous convolution in the encoder, and mixed loss function. The experimental results on LITS data set show that the Dice coefficient, mean intersection over union and mean pixel accuracy of liver segmentation using the proposed method are improved by 3.3%, 2.4% and 3.66% as compared with traditional U-Net method.

**Keywords:** convolutional neural network; liver; image segmentation; multi-head self-attention mechanism; atrous convolution

### 前言

随着计算机技术的快速发展,出现了很多用于人体器官的扫描技术,比如计算机断层扫描技术(Computed Tomography, CT)和核磁共振成像技术(Magnetic Resonance Imaging, MRI)等,这些影像技术已成为了临床医生诊断过程中最常见的方法<sup>[1-2]</sup>,其中CT图像具有无创性和图像清晰等优点,方便医生为患者提供诊断并进行下一步的治疗<sup>[3-5]</sup>。近年来在医疗图像分割领域,基于深度学习的算法已经超过了许多传统算法并取得了卓越的成果<sup>[6]</sup>。利用深度学习技术对医学图像进行分割的技术已日趋成熟<sup>[7-9]</sup>,相较于传统算法,卷积神经网络不需要人工提

取特征或对图像进行过多的预处理设置大量的参数。因此在医学图像分割领域大量的研究人员开始从事深度学习技术的研究<sup>[10-12]</sup>。

医学图像分割就是从医学图像中分离出想要的目标,例如脏器、组织等,然后再经过数字图像技术实现可视化操作和图形引导手术等<sup>[13]</sup>。在医学图像分割任务中,由于医学图像边界的模糊性、纹理的不均匀性以及在没有领域知识的情况下分割区域的不确定性,使得获得准确的分割结果非常困难,即使是富有经验的专家,根据他们的经验标注的结果也略有不同<sup>[14-15]</sup>。

2015年Long等<sup>[16]</sup>提出了一种在卷积神经网络(Convolutional Neural Networks, CNN)基础上发展适用于图像语义分割任务的全卷积神经网络(Fully Convolutional Networks, FCN)模型。FCN不仅能将图像中的每个像素进行分类,而且还能确定像素的位置从而实现从图像类别分类到像素级别分类的进步。然而医学图像分割更加复杂,病灶边界更加难以确定,传统的FCN网络并不能满足实际的需求,因此在FCN的基础上,2015年Ronneberger等<sup>[17]</sup>提出用

**【收稿日期】**2022-12-14

**【基金项目】**辽宁省高校创新人才项目(LR2019058);辽宁省重点科技创新基地联合开放基金(2021-KF-12-05)

**【作者简介】**胡晓阳,博士,副教授,研究方向:飞行体运动控制、武器系统设计与仿真、智能图像识别等,E-mail: xiaoyang\_hu@163.com

于医学图像分割的经典网络模型 U-Net, U-Net 是一个典型编码-解码结构。U-Net 的编码器由卷积和池化组成进行下采样, 以此来获取图片中的语义信息, 然后通过解码器进行上采样并通过跳跃连接来融合解码器输出从而提高分割精度。在 U-Net 的思想基础上, 2018 年 Zhou 等<sup>[18]</sup>提出了 U-Net++, 通过更密集的跳跃连接来消除语义鸿沟。

随着深度学习技术的发展, 越来越多的科研工作者投入到深度学习的研究中, 一些用于提升网络精度的特殊模块被提出。如注意力门 (Attention Gate, AG) 模块, 2019 年 Zhang 等<sup>[19]</sup>在 U-Net 网络中加入 AG 机制对眼底图像进行分割。2020 年大量研究者将 Transformer 引入到图像分类任务中并取得了很好的效果, 此前 Transformer 已成为自然语言处理任务的实际标准, 但在计算机视觉中 CNN 一直占据主导地位, 因此 VIT (Vision Transformer) 的提出直接推动了后续的相关研究。虽然目前 Transformer 在许多视觉任务上取得了与 CNN 相当甚至更好的表现<sup>[20-22]</sup>, 有些工作甚至完全取代了 CNN, 但是这些应用 transformer 的网络都需要大量的数据集进行训练,

如果没有足够的数据集进行大量的训练, 基于 Transformer 的网络表现并不好, 因为 Transformer 结构与 CNN 相比缺少归纳偏置, 即一种先验知识, 所以这些 Transformer 结构的模型虽然高效, 但是需要巨大的硬件和数据支持。

因此越来越多研究人员开始研究视觉 Transformer 在图像领域中与 CNN 相结合的方法, 因为 Transformer 的优势在于通过注意力机制捕捉一些全局的信息用来建立远距离依赖从而提取更多的特征信息<sup>[23]</sup>, 但考虑到 CNN 在局部信息的提取能力和训练成本较低等优势, 本文提出了一种新型的基于 CNN 与 Transformer 相结合的分割方法 MixACU-Net 用于肝脏 CT 图像分割, 提高肝脏的分割效果。

## 1 方法

### 1.1 U-Net 模型

U-Net 网络经常被应用于医疗图像分割任务中, 其网络结构如图 1 所示。该网络是一种编码-解码结构。编码器用于提取图像的上下文信息, 解码器用来逐层恢复图像信息。

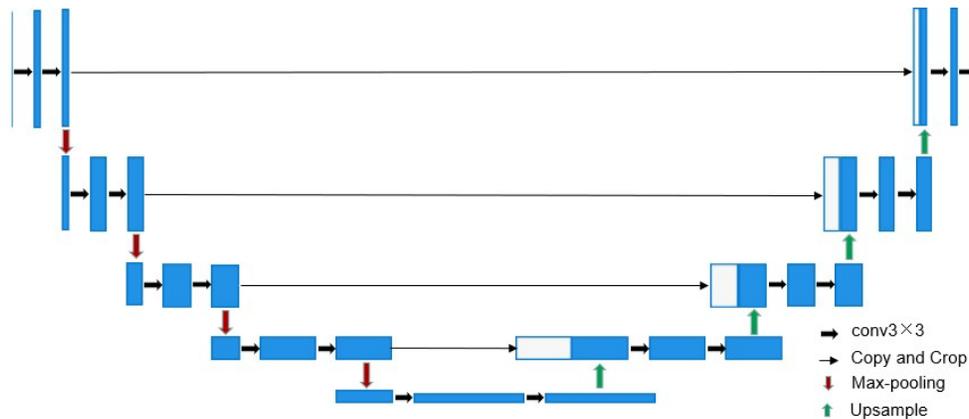


图 1 U-Net 模型  
Figure 1 U-Net model

### 1.2 MixACU-Net 模型

肝脏肿瘤一般体积较小, 所以要提高肝脏肿瘤的分割精度, 最简单的办法就是先将肝脏从 CT 图像中分割出来, 然后再进行肿瘤的分割。本文提出了一种编码-解码结构的肝脏图像分割模型 MixACU-Net, 网络结构如图 2 所示, 从 MixACU-Net 的结构图中可以看到此方法保留了经典 U-Net 的结构, 包括跳跃连接、尺寸为 3×3 的卷积模块、最大池化模块和反卷积模块, 但本文方法在经典 U-Net 中添加了多头注意力机制 (Multi-Head Self Attention, MHSA) 模块和空洞卷积 (Atrous Convolutions)。MHSA 模块被添加到跳

跃连接中, 用来捕捉全局特征。并且在编码器部分使用了空洞卷积, 在不降低图片分辨率的前提下扩大感受野, 从而提升分割的精度。

### 1.3 空洞卷积

空洞卷积又被称为扩张卷积 (Dilated Convolutions)<sup>[24]</sup>, 如图 3 所示。它经常被应用在语义分割与目标检测等任务中, 与普通的卷积操作不同的地方在于空洞卷积引入了一个被称为“扩张率 (Dilation Rate)”的新参数。空洞卷积的优势在于不改变图像分辨率的情况下通过空洞率来扩大卷积核的大小从而扩大感受野, 提高网络的分割精度。

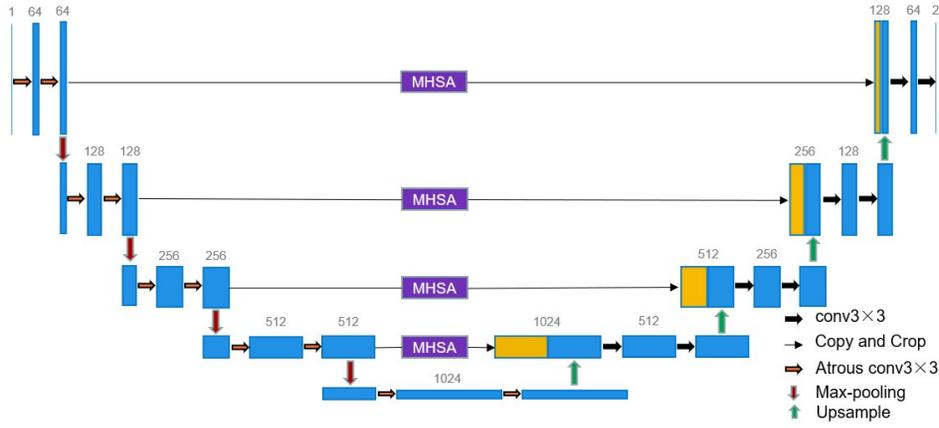


图2 MixACU-Net模型

Figure 2 MixACU-Net model

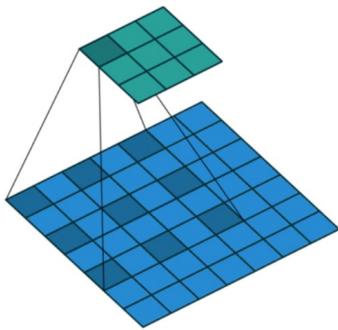


图3 空洞卷积示意图

Figure 3 Schematic diagram of atrous convolution

### 1.4 MHSA 模块

MHSA 与注意力机制不同的是, MHSA 模块对外部信息依赖的程度更低, 它更加关注数据内部或特征的内部联系。自注意力机制结构图如图4所示, 其中,  $R_h$ 、 $R_w$  分别代表高度和宽度的相对位置信息,  $q$ 、 $k$ 、 $v$ 、 $r$  分别代表查询(Query)、键(Key)、值(Value)和位置编码(Position Encodings)<sup>[25]</sup>。

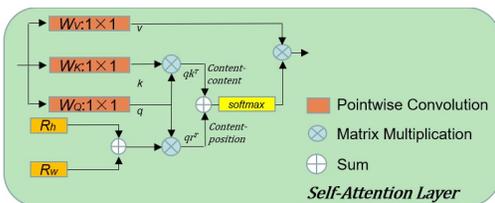


图4 自注意力机制

Figure 4 Self-attention mechanism

假设输入为  $X$ , MHSA 的计算如式(1)和式(2)所示:

$$O = \text{Soft max} \left( \frac{(X \cdot W_q)(X \cdot W_k)^T}{\sqrt{d_k}} \right) (X \cdot W_v) \quad (1)$$

$$\text{MHSA}(X) = \text{Concat} [O_1, \dots, O_n] W_o \quad (2)$$

其中,  $O$  为自注意力机制的输出,  $\text{MHSA}(X)$  为多头自注意力机制的输出,  $W_q$ 、 $W_v$ 、 $W_k$ 、 $W_o$  分别为线性映射的参数矩阵,  $d_k$  为向量  $q$  的维数,  $n$  为注意力头个数。

### 1.5 混合损失函数

考虑到类别不平衡的影响和网络在方向传播过程中出现梯度消失的问题, 本文采用二元交叉熵(Binary Cross Entropy, BCE)损失函数与Dice损失函数混合的损失函数, 函数如公式3所示:

$$\text{LOSS} = \alpha L_{\text{Dice}} + \beta L_{\text{BCE}} \quad (3)$$

其中,  $L_{\text{Dice}}$  和  $L_{\text{BCE}}$  分别代表Dice损失和BCE损失, 本文实验中  $\alpha$  和  $\beta$  分别取1和0.5。

## 2 实验

### 2.1 数据集的获取与处理

本次实验所采用的数据集来自LITS2017肿瘤分割挑战赛所提供的数据, 该数据集包含了具有多年临床经验的放射科医生标注的131例肝脏肿瘤患者腹部CT图像。因为LITS2017数据集的标注包括肿瘤和肝脏两部分, 但本次实验只进行肝脏部分的分割, 因此首先对数据集进行预处理, 通过阈值的方法将附着在肝脏上的肿瘤标签去除掉只保留背景和肝脏两类标签。预处理后的结果如图5所示。以nii格式存储的数据可以使用ITK-SNAP软件打开, 该软件的界面如图6所示。在软件的界面中可以看到CT序列的横断面, 矢状面以及冠状面。

### 2.2 评价指标

本次实验采用了Dice系数、平均交并比(Mean Intersection Over Union, MIOU)和类别平均像素准确率(Mean Pixel Accuracy, MPA)对肝脏的分割结果进行评价。

**2.2.1 Dice系数** Dice系数可以有效地反应网络的分

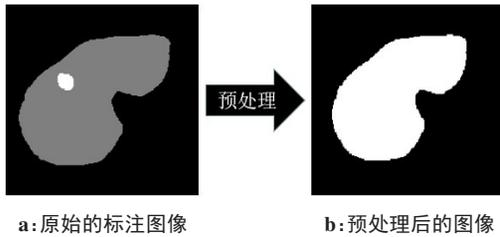


图5 标注图像预处理  
Figure 5 Labeled image preprocessing

割结果与真实标注标签重合的程度,是语义分割任务中最重要的评价指标之一,计算方法如公式(4)所示:

$$Dice = \frac{2TP}{FP + 2TP + FN} \quad (4)$$

其中,TP(True Positive)代表网络输出的预测结果为正例并且医生的标注结果也为正例的像素个数;FP(False Positive)代表网络输出的预测结果为正例但



图6 ITK-SNAP软件界面  
Figure 6 ITK-SNAP software interface

医生的标注为反例的像素个数;FN(False Negative)代表网络输出的预测结果为反例而医生的标注结果为正例的像素个数。

**2.2.2 MIOU** MIOU是语义分割问题中常用的分割指标之一,MIOU表示计算网络输出结果与实际标注结果中每个类别的交集与并集的比的平均值,MIOU越接近于1表示分割的效果越好,计算方法如公式(5)所示:

$$MIOU = \frac{1}{k + 1} \sum_{i=0}^k \frac{TP}{FN + FP + TP} \quad (5)$$

**2.2.3 MPA** MPA代表计算出每个类别的像素被正确分类的像素个数占总像素个数比例的平均值,计算方法如公式(6)所示:

$$MPA = \sum(P_i)/k \quad (6)$$

其中, $P_i$ 代表的是每个类别像素的准确率, $k$ 代表的是类别数。

### 2.3 实验细节

本次实验中使用深度学习框架Pytorch来实现网络模型,并使用1080Ti显卡来对网络进行训练,在训练中采用了Adam优化器来优化训练的模型,学习率设置为

0.003, Batch size设置为6, epoch设置为100,输入图像大小为256×256的单通道CT图像。MHSA模块的头数设置为4,空洞卷积空洞率依次设置为1、2。

为了验证本文所提出的MixACU-Net网络对肝脏的分割效果,在相同的条件下使用LITS2017肿瘤分割挑战赛提供的数据集进行测试,并与U-Net网络输出的结果进行比较。实验结果如表1所示,从表1可以看到本文所提出的MixACU-Net网络在Dice系数,MIOU和MPA上的表现全部优于U-Net网络。其中Dice系数相较于U-Net网络提高了3.3%,MIOU提升了2.4%,MPA提升了3.66%。综合来看本文所提出的MixACU-Net网络在LITS2017数据集上进行肝脏分割的表现优于U-Net网络。

表1 不同算法性能对比(%)

Table 1 Performance comparison between two algorithms (%)

模型	Dice	MIOU	MPA
U-Net	86.92	83.60	89.53
MixACU-Net	90.22	86.00	93.19

## 2.4 实验结果与分析

如图7所示,随机选择了3个测试集样本在不同网络所输出的预测结果与真实标注结果进行对比,

从图中可以更加清楚看到MixACU-Net网络的输出与U-Net网络的输出相比分割效果更好,结果更准确,边界特征更加明显。

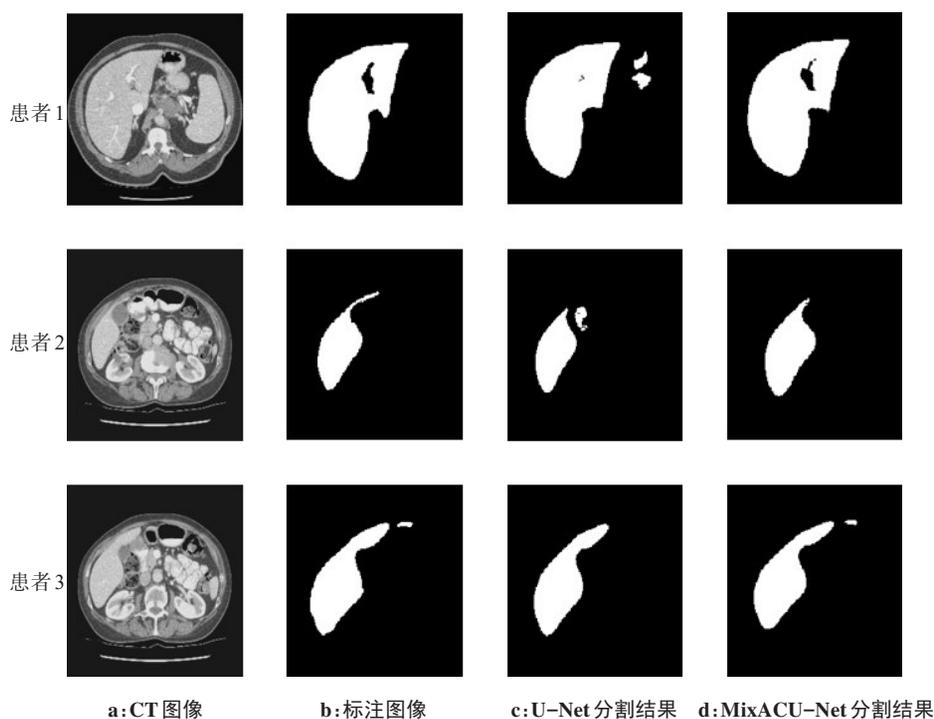


图7 实验结果对比图

Figure 7 Comparison of experimental results

## 2.5 消融实验

为了验证空洞卷积与MHSA在网络中的作用,进行了消融实验,实验过程中保证只会发生单一模块的变化。实验结果如表2所示。在去掉MHSA模块后只保留空洞卷积的情况下,与U-Net网络输出的分割结果相比,只保留了空洞卷积的MixACU-Net(无MHSA)精度仍有提升,这是因为增加空洞卷积在降低图像分辨率的情况下通过改变卷积核的大小扩大了感受野,提升了分割的效果,该结果符合预期。但与本文所提出的MixACU-Net模型分割效果相比仍有一些差距,说明MHSA捕捉长距离依赖的特点可以提升网络精度。

## 3 结论

为了提高对肝脏分割精度,本文提出了一种卷

积与Transformer相结合的医学图像分割网络,该网络利用Transformer可以捕捉长距离依赖的特点弥补了CNN在捕捉全局性特征的局限性。通过MHSA更好地进行全局特征的学习。与U-Net网络相比本文提出的网络无论在分割精度还是在边界轮廓的处理上都更加优秀。

## 【参考文献】

- [1] Haritaoglu I, Harwood D, Davis LS. W/sup 4: real-time surveillance of people and their activities[J]. IEEE Trans Pattern Anal, 2000, 22(8): 809-830.
- [2] 范同乐. 基于改进U-Net模型的肝脏CT图像分割方法研究及实现[D]. 保定: 河北大学, 2021.  
Fan TL. Research and application of liver CT images based on improved U-Net model[D]. Baoding: Hebei University, 2021.
- [3] 俞丽燕, 薛建辉, 吴春如. CT与MR影像学检查在肝细胞癌的临床诊断中的应用效果[J]. 影像研究与医学应用, 2022, 6(18): 162-164.  
Yu LY, Xue JH, Wu CR. Application of CT and MR imaging in clinical diagnosis of hepatocellular carcinoma [J]. Journal of Imaging Research and Medical Applications, 2022, 6(18): 162-164.
- [4] 翟颖. 肝细胞癌介入治疗后经增强CT和MRI影像诊断的对比分析[J]. 影像研究与医学应用, 2020, 4(7): 19-20.  
Zhai Y. Contrastive analysis of the diagnosis of hepatocellular carcinoma by enhanced CT and MRI after interventional therapy[J]. Journal of Imaging Research and Medical Applications, 2020, 4(7): 19-20.
- [5] 崔晓宁. CT与MRI增强扫描在肝细胞癌诊断中的应用效果分析[J]. 影像研究与医学应用, 2020, 4(2): 73-74.  
Cui XN. Application of CT and MRI enhanced scanning in diagnosis

表2 消融实验结果(%)

Table 2 Ablation study (%)

模型	Dice	MIOU	MPA
U-Net	86.92	83.60	89.53
MixACU-Net(无MHSA)	89.00	85.48	92.24
MixACU-Net	90.22	86.00	93.19

- of hepatocellular carcinoma [J]. *Journal of Imaging Research and Medical Applications*, 2020, 4(2): 73-74.
- [6] 周涛,董雅丽,刘珊,等.用于肺部肿瘤图像分割的跨模态多编码混合注意力U-Net[J].*光子学报*, 2022, 51(4): 376-392.  
Zhou T, Dong YL, Liu S, et al. Cross-modality multi-encoder hybrid attention U-Net for lung tumors images segmentation [J]. *Acta Photonica Sinica*, 2022, 51(4): 376-392.
- [7] 李昊轩.基于深度学习的医疗图像检测[J].*电子制作*, 2019(4): 53-55.  
Li HX. Medical image detection based on deep learning [J]. *Practical Electronics*, 2019(4): 53-55.
- [8] 艾飞玲,马圆,田思佳,等.深度学习在医学图像分析中的研究进展[J].*北京生物医学工程*, 2018, 37(4): 433-438.  
Ai FL, Ma Y, Tian SJ, et al. Research progress of deep learning in medical image analysis [J]. *Beijing Biomedical Engineering*, 2018, 37(4): 433-438.
- [9] Yang Q, Yan P, Zhang Y, et al. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss [J]. *IEEE Trans Med Imaging*, 2018, 37(6): 1348-1357.
- [10] 赵广文,王阳,杨晨.基于注意力与金字塔融合的医学图像分割仿真[J].*计算机仿真*, 2022, 39(2): 184-190.  
Zhao GW, Wang Y, Yang C. Simulation of medical image segmentation based on attention and pyramid fusion [J]. *Computer Simulation*, 2022, 39(2): 184-190.
- [11] Soltani-Nabipour J, Khorshidi A, Noorian B. Lung tumor segmentation using improved region growing algorithm [J]. *Nucl Eng Technol*, 2020, 52(10): 2313-2319.
- [12] Yan C, Sun ZQ, Tian EG, et al. Medical image segmentation model combining multi-branch structure and self-attention mechanism and its application [J]. *Journal of Chinese Computer Systems*, 2021, 42(11): 2375-2381.
- [13] 杨昌俊,杨新.基于图割与快速水平集的腹部CT图像分割[J].*CT理论与应用研究*, 2011, 20(3): 291-300.  
Yang CJ, Yang X. Abdominal CT image segmentation based on graph cuts and fast level set [J]. *CT Theory and Applications*, 2011, 20(3): 291-300.
- [14] 邓鸿,邓雅心,丁廷波,等.基于生成对抗网络的肝脏CT图像分割[J].*北京生物医学工程*, 2021, 40(4): 367-376.  
Deng H, Deng YX, Ding YB, et al. Liver CT image segmentation based on generative adversarial network [J]. *Beijing Biomedical Engineering*, 2021, 40(4): 367-376.
- [15] 王伟胜,骆嘉伟,林红利.医学图像计算机辅助诊断数据平台研究[J].*中国生物医学工程学报*, 2013, 32(1): 105-108.  
Wang WS, Luo JW, Lin HL. Computer-aided diagnosis data platform by using medical imaging [J]. *Chinese Journal of Biomedical Engineering*, 2013, 32(1): 105-108.
- [16] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015: 3431-3440.
- [17] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical imagesegmentation [C]//*International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, Cham, 2015: 234-241.
- [18] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, et al. U-Net++: a nested U-Net architecture for medical image segmentation [M]//*Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, Cham, 2018: 3-11.
- [19] Zhang S, Fu H, Yan Y, et al. Attention guided network for retinal image segmentation [C]//*International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2019: 797-805.
- [20] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers [C]//*Computer Vision-European Conference on Computer Vision (ECCV) 2020*. Springer International Publishing, 2020: 213-229.
- [21] Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, 2021: 10012-10022.
- [22] Zheng S, Lu J, Zhao H, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2021: 6881-6890.
- [23] 穆应晨,王学军,王怡雯,等.一种基于混合变压器的医学图像分割方法[J].*安徽工程大学学报*, 2022, 37(1): 39-45.  
Mu YC, Wang XJ, Wang YW, et al. A medical image segmentation method based on hybrid transformer [J]. *Journal of Anhui Polytechnic University*, 2022, 37(1): 39-45.
- [24] Chen LC, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//*Proceedings of the European Conference on Computer Vision (ECCV)*. 2018: 801-818.
- [25] Bello I, Oph B, Vaswani A, et al. Attention augmented convolutional networks [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, 2019: 3286-3295.

(编辑:薛泽玲)