

一种基于 U<sup>2</sup> 型网络的三维肾肿瘤图像分割方法

李思源, 李强, 关欣\*

天津大学微电子学院, 天津 300072

**摘要** 肾肿瘤危害极大,严重影响人类身体健康,对肾肿瘤进行早期检测和诊断有助于患者的治疗与康复。为高效地从腹部电子计算机断层扫描(CT)图像中分割出肾脏及肾肿瘤图像,提出一种基于三维 U<sup>2</sup>型网络(U<sup>2</sup>-Net)的分割方法。首先,将二维 U<sup>2</sup>-Net 升级并调整网络深度、损失函数与深度监督策略;其次,为了增强解码端的特征表达能力,提出残差特征强化模块,对解码端特征图进行通道与空间域的强化;最后,为提高网络对全局信息的提取能力,提出基于全局特征的多头自注意力模块,计算特征图所有体素点间的长期依赖关系,获取丰富的三维医学图像上下文信息。所提方法在 KiTS19 官方数据集上的实验结果表明,平均 Dice 值为 0.9008,参数量为 4.60 MB,与现有方法相比,所提方法能够在参数量较小的前提下,取得较好的分割精度,对小内存嵌入式肾脏和肾肿瘤图像分割系统具有很高的应用价值。

**关键词** 医学图像分割;卷积神经网络;残差特征强化;多头自注意力

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP231203

A 3D Renal Tumor Image Segmentation Method Based on U<sup>2</sup>-Net

Li Siyuan, Li Qiang, Guan Xin\*

School of Microelectronics, Tianjin University, 300072, China

**Abstract** Renal tumors pose great harm to and seriously affect human health. Early detection and diagnosis of renal tumors can help patients' treatment and recovery. To efficiently segment kidneys and tumors from abdominal computed tomography (CT) images, this paper proposes a method based on 3D U<sup>2</sup>-Net. First, we upgrade the 2D U<sup>2</sup>-Net, adjust loss function, depth of network, and supervision strategy. To improve the feature expression ability of the decoder, we propose a residual feature enhancement module, which enhances the channel and spatial domain of the feature map at the decoder. To further improve the model's ability to extract global information, we propose a multi-head self-attention module based on global features, which calculates the long-term dependencies between all voxel points in the feature map and obtains more contextual information of 3D medical images. The method is tested on the official KiTS19 dataset and the results show that the average Dice value is 0.9008 and the parameter quantity is 4.60 MB. Compared with existing methods, our method can achieve better segmentation accuracy with small parameter quantity, and has great application value for small memory embedded system used for kidneys and tumors image segmentation.

**Key words** medical image segmentation; convolutional neural network; residual feature enhancement; multi-head self-attention

## 1 引言

肾细胞癌简称肾癌,起源于肾实质泌尿系统的恶性肿瘤,每年约有 40 万名患者被确诊为肾癌<sup>[1]</sup>,其严重危害人类身体健康。电子计算机断层扫描(CT)技术由于成像清晰、密度分辨率高<sup>[2]</sup>,目前已成为肾肿瘤诊断的主要技术手段之一。从 CT 图像中将肾脏和肾肿瘤与其余背景进行区分,能够协助医生进行诊断,帮助患者尽早获取诊断结果并接受治疗。患者的腹部 CT

图像包含肾脏及其他腹部器官,病灶区域噪声较多,且肾脏及肾肿瘤的存在和表现形式具有较大的个体差异<sup>[3]</sup>,这为准确分割图像造成较大困难。

基于边缘检测<sup>[4]</sup>、阈值<sup>[5]</sup>和区域<sup>[6]</sup>的传统图像分割方法利用图像不同区块灰度值不同的特性,能够对简单图像进行较好的分割,但缺少对图像空间信息的利用。基于机器学习的分割方法如聚类<sup>[7]</sup>和支持向量机<sup>[8]</sup>等,能对有限的数据进行学习,得到分割模型并用于待测数据分割,进一步提升了性能,但多数算法复杂

收稿日期: 2023-04-28; 修回日期: 2023-05-19; 录用日期: 2023-05-30; 网络首发日期: 2023-06-10

通信作者: \*guanxin@tju.edu.cn

度较高。基于深度学习和卷积神经网络(CNN)的分割方法近年来在医学图像分割领域表现优异,如全卷积神经网络(FCN)<sup>[9]</sup>、U型网络(U-Net)<sup>[10]</sup>、V型网络(V-Net)<sup>[11]</sup>等,在图像分割任务中取得不错的分割效果。FCN将CNN中所有的全连接层改为卷积层以实现任意尺寸的输入,通过跳跃连接并融合高层语义信息与浅层纹理信息,构建出性能较为优秀的端到端分割系统。U-Net在FCN的基础上提出编码解码结构,所搭建的完全对称的特征提取网络在分割任务上表现优异,是医学图像分割的最佳解决方案之一。V-Net将编码路径上的最大池化下采样替换为卷积操作,并引入残差连接,该网络能够对三维医学图像进行较好分割。

在此基础上,各种变体结构和功能模块相继被提出以获取更好的分割性能。密集跳跃连接网络<sup>[12]</sup>将编解码路径中不同尺度的特征图融合,避免训练过程中的特征丢失问题。改进的三级级联网络<sup>[13]</sup>将肿瘤分割任务细化为3个阶段,通过对上一阶段的输出子图进行更精细的分割,在简单模型中获得更高的分割性能。Zuo等<sup>[14]</sup>利用注意力机制增强三维语义分割网络对全局上下文信息的获取能力,并结合空间金字塔采样来减少计算量。多视图卷积<sup>[15]</sup>的分割方法利用伪三维卷积从3个维度提取特征并结合组卷积降低计算量,实现了对参数的轻量级处理。尽管基于CNN的分割方法能够在各项分割任务中取得不错的效果,但能够从腹部CT图像中较好地分割出肾脏及肾肿瘤图像的方法很少,无法大规模应用于临床实践。

为提高肾脏及肾肿瘤图像分割精度并减少网络参数量,本文提出一种基于全局特征多头自注意力与残差特征强化的三维U<sup>2</sup>-Net(Transformer-based residual feature enhancement U<sup>2</sup>-Net, TransRFE-U<sup>2</sup>-Net)用于肾脏及肾肿瘤图像的分割。首先,选取分割性能较好的二维U<sup>2</sup>-Net<sup>[16]</sup>为基准网络并升维以适应三维输入数据,为避免过度下采样引起特征信息的丢失,减少了主干与嵌套子结构中下采样的次数,调整深度监督策略,在训练阶段取解码端各级输出做损失计算,而在预测阶段只取最后一级输出,减少模型推理时间。其次,为提高解码端特征表达能力并减弱目标区域周围的噪声干扰,提出残差特征强化模块,在解码端对特征图进行空间与通道上的特征强化,使网络能更好地聚焦图像特征。最后,针对CNN感受野有限的问题,聚焦Transformer模型<sup>[17]</sup>对全局信息的提取能力,提出一种基于全局特征的多头自注意力模块,在网络编码末端将特征图转换为序列,生成查询( $Q$ )、索引( $K$ )、内容( $V$ )矩阵进行多头自注意力计算,获取特征图所有体素点间的长期依赖关系,提升网络对肾脏及肾肿瘤图像的分割精度。在改进网络结构的同时,引入对数据集的预处理和后处理方案,并采用骰子损失函数(Dice Loss)和交叉熵损失函数的组合形式作为目标

优化函数进行训练。

## 2 TransRFE-U<sup>2</sup>-Net方法原理

### 2.1 升维U<sup>2</sup>-Net

U<sup>2</sup>-Net最开始被用于显著性目标检测任务<sup>[18]</sup>,在取得高性能的同时较主流方法有更少的计算量。在二维U<sup>2</sup>-Net的基础上升维得到三维U<sup>2</sup>-Net,其基本结构如图1所示,外部结构参考了U-Net的编解码结构,编解码各路径下的每一层为一个结构块,每个结构块内嵌套了一个子U-Net结构,即整个模型是一个两级嵌套的U型结构,每一个结构块称为Residual U-block(RSU)。三维U<sup>2</sup>-Net编码路径中有4次下采样,共有5级编码器(En\_1~En\_5),在解码路径中对应4次上采样,共有4级解码器(De\_1~De\_4)。在编码阶段,特征图每通过一个RSU后,采用最大池化实现下采样,经下采样后的特征图尺寸降低为原特征图的1/2。在与每个编码阶段相对应的解码阶段,图像每通过一个RSU后,采用线性插值实现上采样,经上采样后的特征图尺寸恢复为原特征图的2倍。

模块细节上,En\_1~En\_4,De\_1~De\_4分别包含4、3、2和1次下采样,RSU只对最高分辨率特征图使用两次卷积,其余分辨率特征图均为一次,特征图经过RSU处理前后的尺寸一致。En\_5与前4级RSU不同,其内部不含下采样操作,而是采用膨胀卷积扩大感受野,在整个过程中使特征图尺寸保持不变。在网络的解码端,对De\_1~De\_4以及En\_5的中间输出结果进行深度监督,使用线性插值上采样将图像恢复为原图大小,在通道拼接后缩减为目标类别数以实现分割功能,根据损失函数求取分割结果与原始标签的差距,并反向传播实现迭代训练与参数优化,深度监督的损失Sup 0为Sup 1~Sup 5的通道叠加。在对称的编解码器之间添加跳跃连接,融合浅层与深层特征信息,编码端通道数设定为64,RSU与外部下采样不改变通道数,解码端经跳跃连接后通道数增加为128,再经过卷积调整为64。在改进的三维U<sup>2</sup>-Net中使用神经网络单元激活(Leaky ReLU)函数<sup>[19]</sup>和实例标准化策略,在小样本输入前提下缓解梯度消失问题,提升模型拟合能力。

### 2.2 残差特征强化模块

因医学图像目标区域周围存在干扰性的组织与器官,且拍摄过程由于设备或患者原因易造成伪影,对目标区域特征的提取和恢复存在一定的困难。所提网络通过多级解码模块获取输出,因此解码端的特征表达能力和恢复能力对分割精度有重要影响。受文献<sup>[20]</sup>的启发,提出一种在解码端应用的残差特征强化模块,结合改进的三维U<sup>2</sup>-Net,加强解码路径中特征的表达能力,提升网络聚焦感兴趣区域的能力,进一步提升模型的分割性能。

残差特征强化解码模块的结构示意图如图2所示

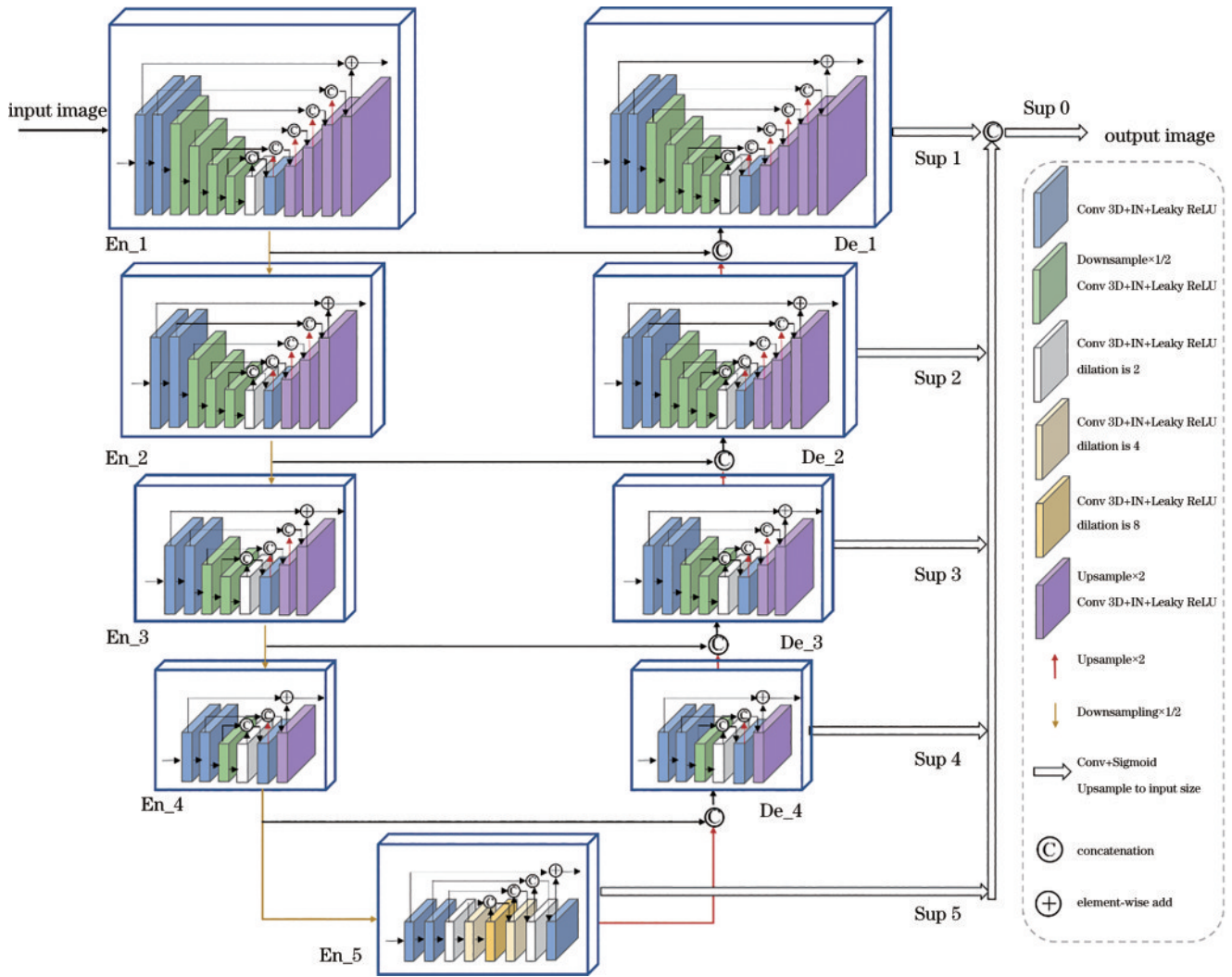


图 1 三维 U<sup>2</sup>-Net 网络结构

Fig. 1 Network structure of 3D U<sup>2</sup>-Net

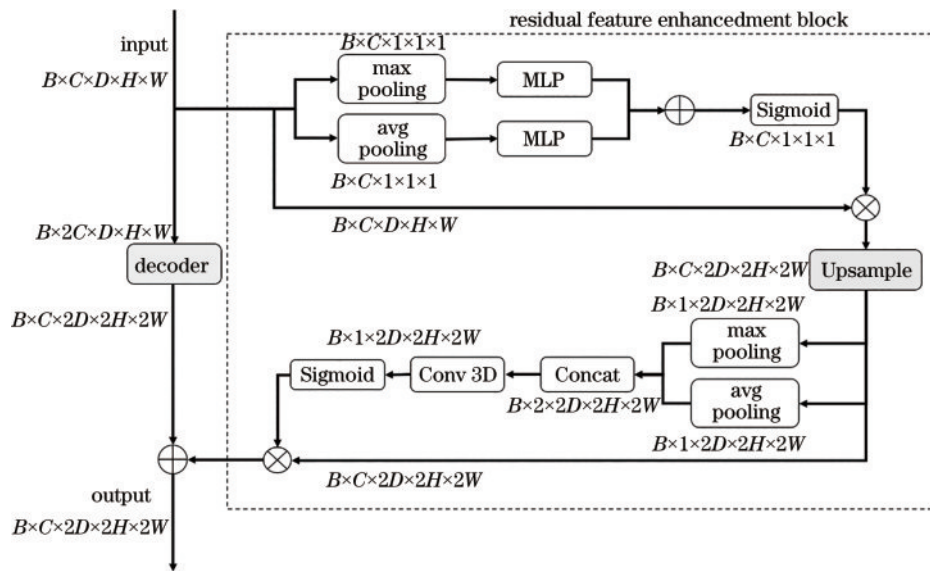


图 2 残差特征强化解码模块

Fig. 2 Residual feature enhancement decoder block

示,特征图  $X \in \mathbb{R}^{B \times C \times D \times H \times W}$  ( $B$  为批量大小,  $C$  为通道数,  $D$ 、 $H$ 、 $W$  分别为特征图的深度、高度与宽度), 通过编码端处理后进入解码器 De\_4, 原解码模块将输入特征图与跳层连接的特征图融合后, 通过三线性插值和一个  $3 \times 3 \times 3$  卷积扩大图像尺寸并减少通道数。改进的网络在该解码模块的基础上添加残差分支, 当特征图进入分支时, 分别通过基于 3 个维度的全局最大池化模块和全局平均池化模块, 得到两个尺寸为  $B \times C \times 1 \times 1 \times 1$  的特征图, 通过线性感知单元 (MLP) 进行全连接处理, 此时的特征图已经在 3 个维度上按照全局池化标准对权重进行调整和修正, 恢复尺寸后所提网络将更容易聚焦感兴趣区域, 使用 Sigmoid 函数激活后作残差连接并进行与原解码模块相同的上采样操作。为强化特征图在空间域的表达能, 引入空间域的特征强化模块, 将上采样后的特征图分别通过基于通道的全局池化模块, 得到两个尺寸为  $B \times 1 \times 2D \times 2H \times 2W$  的特征图, 维度拼接后通过卷积层降维为 1, 使用 Sigmoid 函数激活并与第二阶段的残差输出相乘, 得到最终的分支强化特征图, 空间域强化模块聚合了所有通道的信息, 加强了所提网络对肾脏和肾肿瘤前景信息的关注。特征强化模块用于解码路径中的每一级解码模块, 原解码模块共有 4 级, 因此特征强化模块也有 4 级。

### 2.3 基于全局特征的多头自注意力模块

为了强化所提网络对医学图像上下文关联信息的提取, 提出基于全局特征的多头自注意力模块 (Global

feature-based multi-head self-attention, GF-MSA), 添加在 RFE-U<sup>2</sup>-Net 的编码末端, 描述了其具体计算流程, 如图 3 所示。自注意力模块处理的对象是一维序列数据, 进行图像分割任务时, 需将编码末端得到的整个三维特征图展平成一维序列。具体实现为将三维特征图均分为  $N$  个图像块 (patch), 再将每个图像块展平成序列  $X_p$ , 通过线性投影  $E$  将每一个序列映射到  $C$  维嵌入空间, 并添加位置编码信息  $E_{pos}$ , 嵌入处理结束后的所有图像序列  $X_{GF}^{(l-1)}$  可表示为

$$X_{GF}^{(l-1)} = [X_p^{(1)}E; X_p^{(2)}E; \dots; X_p^{(N)}E] + E_{pos}, \quad (1)$$

式中:  $X_{GF}^{(l-1)} \in \mathbb{R}^{N \times C}$ ;  $N$  为图像切块处理后得到的块数, 即序列长度 (将深层特征图均分为  $1 \times 1 \times 1$  大小的图像块,  $N$  的大小为深层特征图尺寸的乘积);  $C$  为通道数;  $l-1$  为图 3 中的级序数。图 3 中 GF-MSA 对序列  $X_{GF}^{(l-1)}$  作自注意力计算, 获取原特征图中各个子图像块间的上下文相关性, 通过多组待学习的  $W_q$ 、 $W_k$ 、 $W_v$  矩阵与单个序列相乘得到该序列的多组  $Q$ 、 $K$ 、 $V$  向量。注意力系数矩阵  $\text{Attention}(Q, K, V)$  可表示为

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (2)$$

式中:  $d_k$  为序列维度;  $K$ 、 $V$  来自同一个序列,  $Q$  来自另一个序列。通过式 (2) 计算得到多组其余序列对当前序列的注意力系数矩阵, 将当前序列用其余序列线性表示, 当前序列包含了全局信息, 多头自注意力模块会对所有序列做上述计算。

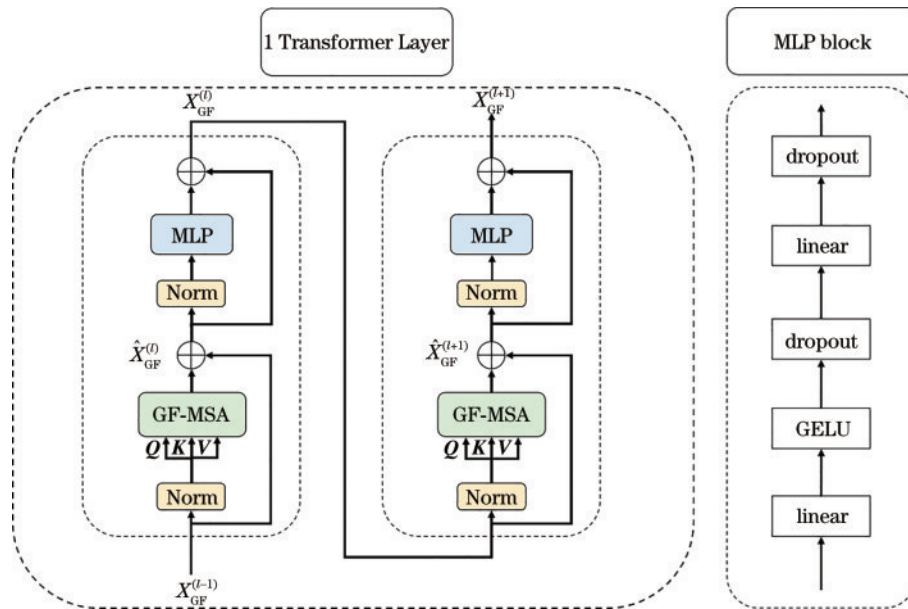


图 3 多头自注意力模块计算流程图

Fig. 3 Calculation process of multi-head self-attention block

图 3 中, MLP 包含两层全连接层、两层 dropout 层与一个激活函数层 GELU<sup>[21]</sup>, 对完成自注意力计算后的序列数据进行全局关系处理, 每一个 GF-MSA 与 MLP 均采用残差连接并进行层归一化。第一级 GF-MSA 与

MLP 的输出作为第二级的输入, 每一级的待学习矩阵  $W_q$ 、 $W_k$ 、 $W_v$  均不同, 为体现 GF-MSA 的输入输出关系与参数关系, 将两级 GF-MSA 组合为一个 Transformer Layer。最终的多头自注意力计算公式可表示为

$$\hat{X}_{GF}^{(l)} = GF - MSA \left[ \text{Norm} \left( X_{GF}^{(l-1)} \right) \right] + X_{GF}^{(l-1)}, \quad (3)$$

$$X_{GF}^{(l)} = \text{MLP} \left[ \text{Norm} \left( \hat{X}_{GF}^{(l)} \right) \right] + \hat{X}_{GF}^{(l)}, \quad (4)$$

$$\hat{X}_{GF}^{(l+1)} = GF - MSA \left[ \text{Norm} \left( X_{GF}^{(l)} \right) \right] + X_{GF}^{(l)}, \quad (5)$$

$$X_{GF}^{(l+1)} = \text{MLP} \left[ \text{Norm} \left( \hat{X}_{GF}^{(l+1)} \right) \right] + \hat{X}_{GF}^{(l+1)}. \quad (6)$$

将编码末端特征图展平为序列进行自注意力计

算,以及重塑尺寸为原特征图的数据,处理步骤如图 4 所示,将其命名为 Trans-block,其输入为最后一层编码器输出,输出端紧接基于残差特征强化模块的解码器,选择 4 层 Transformer Layers(共 8 级 GF-MSA)并通过  $1 \times 1 \times 1$  卷积将输入序列通道调整为 128,注意力计算结束再调整至原特征图通道数。

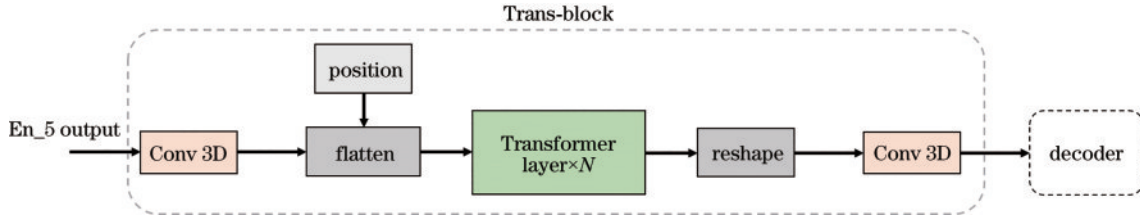


图 4 多头自注意力数据处理步骤

Fig. 4 Multi-head self-attention data processing procedure

### 2.4 所提网络结构

所提 TransRFE-U<sup>2</sup>-Net 的整体网络结构如图 5 所示,输入数据经编码路径提取特征后,通过 Trans-block 进行全局依赖关系的提取,然后通过残差特征强化模块结合的解码模块,恢复特征图尺寸,编码模块的具体操作和跳跃连接方式与 3D U<sup>2</sup>-Net 一

致。TransRFE-U<sup>2</sup>-Net 加强了解码端的特征表达能力,网络在解码输出时更能关注肾脏和肾肿瘤的前景信息,通过添加多头自注意力模块,加强了对深层特征图全局体素点间关系的提取,对上下文关联密切的医学三维图像具有较好的分割作用。

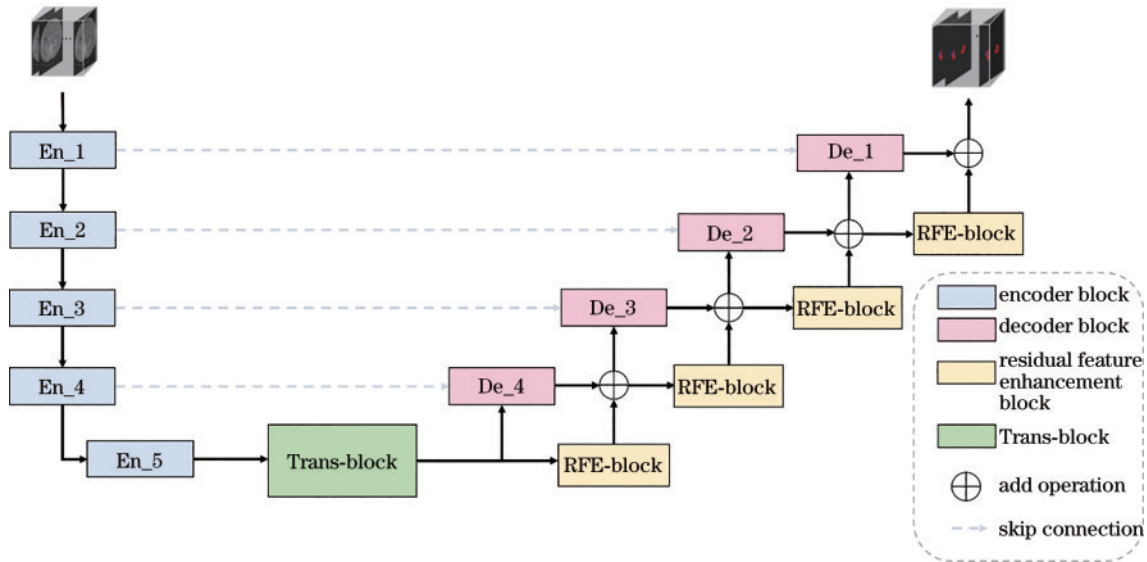


图 5 TransRFE-U<sup>2</sup>-Net 结构

Fig. 5 Structure of TransRFE-U<sup>2</sup>-Net

## 3 实验及结果分析

### 3.1 KiTS19 数据集

使用 KiTS2019 肾脏及肾肿瘤分割挑战赛的官方数据集 KiTS19<sup>[22]</sup> (亦可参见网址 <https://kits19.grand-challenge.org/data/>) 进行网络训练和推理。数据集包含 300 例肾癌患者的 CT 图像,其中 210 例由专家进行肾脏及肾肿瘤标签划分,按 8:2 的比例分为训练集和验证集进行训练,测试集包含 90 例患者图像,其标签数据未公开,训练集中每一例包含一个分辨率

为  $D \times 512 \times 512$  的三维数据 ( $D$  为深度维,各个病例不同),每个病例图像由于拍摄环境因素影响体素间距不同,而数据的各向异性会破坏三维卷积的优势。首先对数据进行预处理,将所有数据重采样至  $3.22 \times 1.62 \times 1.62$  的体素间距,并将强度值窗位设定为  $[-20, 300]$ <sup>[23]</sup>。KiTS19 带标签的横截面可视化示意如图 6 所示,图 6(a)~图 6(c) 分别为 3 个不同病例,可见数据集中不同病例的肾脏与肿瘤形状与大小各有差异。在推理结束后的分割图像上进行连通域后处理,排除明显错误的假阳性点。将最终的预测结果提交至肾肿瘤分割

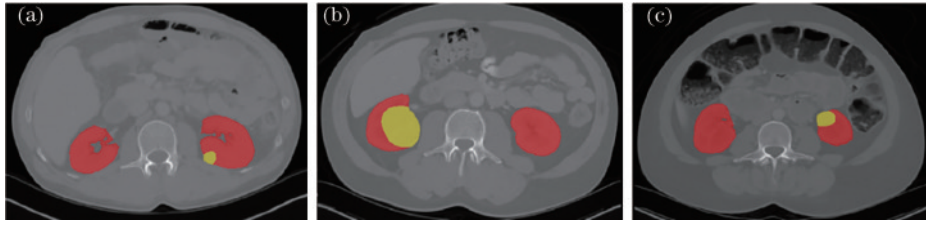


图 6 KiTS19数据集可视化。(a)病例1;(b)病例2;(c)病例3

Fig. 6 Visualization of KiTS19 dataset. (a) Case 1; (b) case 2; (c) case 3

挑战赛官网,获取 Dice 值作为实验结果以评估所提方法优劣。

### 3.2 实验设备及参数设置

实验在 NVIDIA GeForce RTX 2080 Ti 11 GB GPU 上进行训练和推理,网络使用基于 Python 语言的 PyTorch 深度学习开发框架实现,具体环境为 Ubuntu 18.04, Python 3.9, PyTorch 1.12。训练开始时,对所有需要学习的数据进行随机初始化,受图形处理器(GPU)内存限制,设定批量大小为 2,将图像裁剪至大小为  $80 \times 160 \times 160$ ,该裁剪尺寸能够在保留必要的图像细节信息的同时提供足够多的上下文信息。由于训练数据较少,采用数据增强技术增加训练样本数量,降低因训练集数据不足导致的过拟合风险,按随机数概率在  $[-15^\circ, 15^\circ]$  范围内旋转,在  $[0.85, 1.25]$  范围内进行尺寸缩放<sup>[24]</sup>。所提网络使用随机梯度下降(SGD)优化器优化参数,优化器参数设置为 0.9,初始学习率为 0.01,采用 Poly 策略调整训练时的学习率,为使模型能够充分收敛,设置网络最大训练轮数为 500。实验采用 Dice 损失函数与交叉熵损失函数的组合形式  $L_{\text{comp}}$  进行训练,  $L_{\text{comp}}$  可表示为

$$L_{\text{comp}} = -\frac{1}{N} \sum_{i=1}^N \left\{ (\ln p_i) t_i + [\ln(1-p_i)](1-t_i) \right\} - \frac{2 \sum_{i=1}^N p_i t_i + \epsilon}{\sum_{i=1}^N p_i^2 + \sum_{i=1}^N t_i^2 + \epsilon}, \quad (7)$$

式中:  $N$  为总样例数;  $p_i, t_i$  分别为预测图与标签图中的目标像素点。实验使用 Dice 相似系数作为评价指标,取值范围为  $[0, 1]$ , 该值越接近 1 表示分割精度越高, Dice 相似系数  $f_{\text{Dice}}$  可表示为

$$f_{\text{Dice}} = \frac{2f_{\text{TP}}}{f_{\text{FP}} + 2f_{\text{TP}} + f_{\text{FN}}}, \quad (8)$$

式中:  $f_{\text{TP}}, f_{\text{FP}}, f_{\text{FN}}$  分别为分割结果中的像素为真阳性、假阳性、假阴性的数量。将预测结果提交至肾肿瘤分割挑战赛官网,分别得到肾脏、肾肿瘤以及平均的 Dice 值,以评价分割性能。

### 3.3 实验结果分析

为验证所提 TransRFE-U<sup>2</sup>-Net 的有效性,以三维 U<sup>2</sup>-Net 为基准网络设计对比实验,分别使用加入残差特征强化模块的网络(RFE-U<sup>2</sup>-Net)、加入全局特征多头自注意力的网络(Trans-U<sup>2</sup>-Net)以及融合所有模块的网络(TransRFE-U<sup>2</sup>-Net)对 KiTS19 数据集进行训练与推理,对比各自的肾脏、肾肿瘤和平均分割精度,同时比较各方法的参数量与计算量。实验在预处理与后处理方案上进行,并使用组合的损失函数进行训练,所提方法在 KiTS19 测试集上的消融实验结果如表 1 所示。

由表 1 可知,升维后的三维 U<sup>2</sup>-Net 在肾脏和肾肿瘤上的 Dice 值分别为 0.9658 和 0.7983, 平均 Dice 值为 0.8821, 参数量为 2.10 MB, 浮点计算量(Floating point operations, FLOPs)为 350.25 GB。引入残差特征强化模块后,肾脏和肾肿瘤精度分别提升 0.02、0.64 个百分点,参数量和计算量分别增加 0.53 MB 和 2.14 GB, 实验结果表明,基于残差特征强化的网络较原网络在解码端具有更强的特征表达能力,强化网络对肾脏和肾肿瘤前景信息能够进行更精确的分割。在原网络中引入多头自注意力模块后,肾脏和肾肿瘤的分割精度分别提升 0.27、3.07 个百分点,参数量增加了 1.88 MB, 对肿瘤的分割精度具有较大提升,可见多头自注意力模块能够获取更多的医学图像上下文信息,对肿瘤小目标区域进行更好的分割。将两种模块融合后,平均精度达到 0.9008, 较基准网络提高了 1.87 个百分点,

表 1 所提方法在 KiTS19 测试集上的消融实验结果

Table 1 Results of ablation experiments of our methods on the KiTS19 test dataset

Model	Parameter quantity /MB	FLOPs /GB	Dice score		
			Kidney	Tumor	Average
3D U <sup>2</sup> -Net	2.10	350.25	0.9658	0.7983	0.8821
RFE-U <sup>2</sup> -Net	2.63	352.39	0.9660	0.8047	0.8854
Trans-U <sup>2</sup> -Net	3.98	352.60	0.9685	0.8290	0.8987
TransRFE-U <sup>2</sup> -Net	4.60	352.87	0.9706	0.8311	0.9008

证明所提方法在引入较少的参数量和计算量的情况下能大幅提升性能。

由于测试集没有标签数据,对验证集分割结果进行可视化处理。图 7 为从验证集中随机挑选的两个病例的三维 U<sup>2</sup>-Net 和 TransRFE-U<sup>2</sup>-Net 的可视化分割结果,与专家手工分割的标签图像进行对比,图 7(a)~图 7(d)分别为单个案例的横截面、矢状面、冠状面和三维立体图。观察得知,三维 U<sup>2</sup>-Net 和 TransRFE-U<sup>2</sup>-Net 对肾脏和肾肿瘤均能进行较好的分割,在三个

切面和三维立体图中,两种方法对肾脏的分割结果与标签非常相近,对应于较高的肾脏分割精度;就肿瘤分割结果而言,与专家手动分割标签相比,三维 U<sup>2</sup>-Net 分割出的肿瘤偏小且形状存在一定差异,TransRFE-U<sup>2</sup>-Net 的肿瘤分割结果得到进一步改善,与标签相似度更高,说明了残差特征强化模块和基于全局特征的多头自注意力模块的有效性,特征表达能力的提高和更多空间信息的提取更有利于提升对肿瘤小目标区域的分割精度。

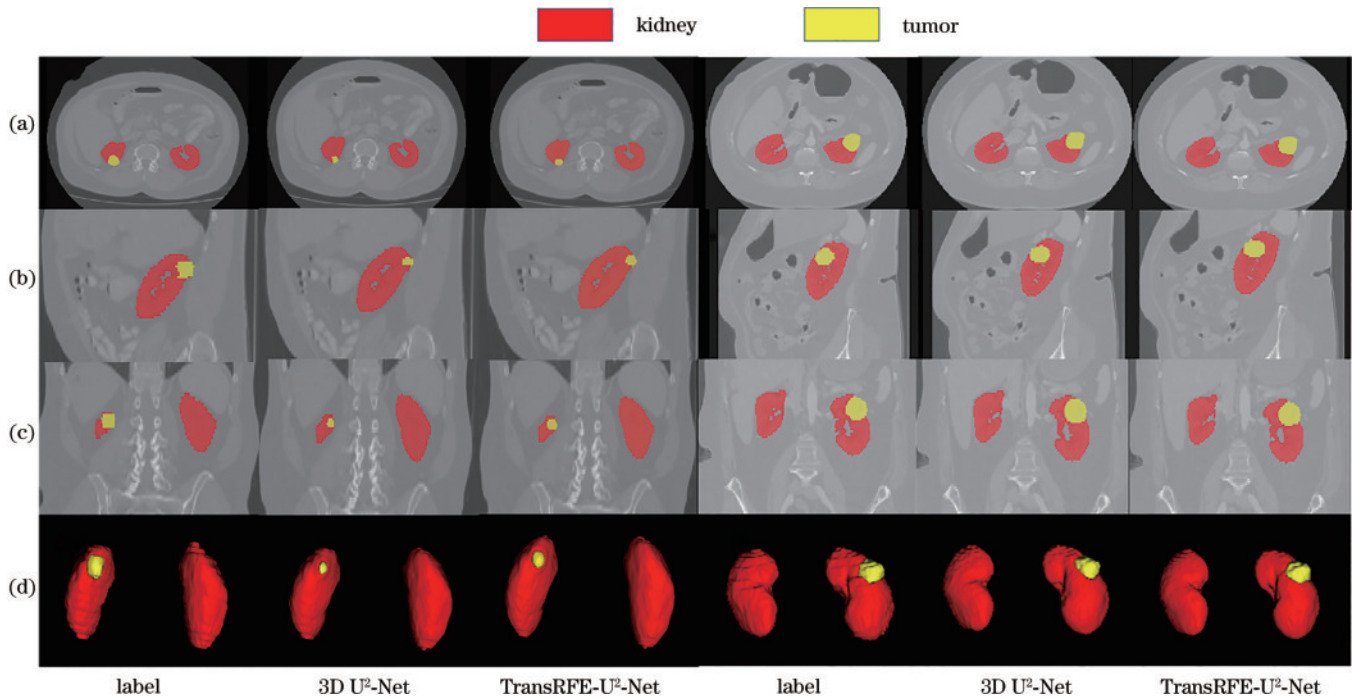


图 7 分割结果可视化。(a)横截面;(b)矢状面;(c)冠状面;(d)三维立体图

Fig. 7 Visualization of segmentation results. (a) Cross section; (b) sagittal plane; (c) coronal plane; (d) 3D stereogram

将所提方法与其他已发表方法在 KiTS19 测试集上的分割性能进行比较。除对比分割精度外,还对比了各方法的阶段数与同输入条件下的参数量。在已发表方法中,Kang 等<sup>[25]</sup>在三维 CNN 中引入长短期记忆模块提取图像的空间信息,Guo 等<sup>[26]</sup>在 U-Net 中为每个卷积层添加残差连接并在解码端引入注意力模块,Zhao 等<sup>[27]</sup>在网络中融入多尺度监督策略,对解码端多级输出进行损失评估,上述 3 种方法与所提方法均为单阶段端到端的语义分割方法。Zhang 等<sup>[28]</sup>在 FCN 的基础上进行二级级联分割,先进行肾脏粗定位,进而在裁剪出的目标区域进行肾脏和肾肿瘤的细化分割,Hou 等<sup>[29]</sup>使用级联的三阶段自引导网络实现由粗到细的分割,并利用第三阶段肾肿瘤的细化分割结果与第二阶段前景细化分割结果进行校准,加强对肾肿瘤小目标区域的识别,多阶段方法设计思想与单阶段端到端的网络模型不同,其涉及的流程复杂、且需训练多级分割网络,参数量较大。

表 2 为所提方法与上述方法的性能对比,表 2 中,Stage 一栏表示方法的阶段数,Single 表示单阶段方

法,Double 与 Triple 分别表示双阶段方法与三阶段方法。由表 2 可知,所提方法在测试集上肾脏分割精度为 0.9706,肾肿瘤分割精度为 0.8311,平均分割精度为 0.9008,参数量仅有 4.60 MB,与其余单阶段端到端的方法相比,在肾肿瘤分割精度、平均分割精度上有很大优势,且参数量更少,在端到端分割系统应用中有更好的参考价值。在多阶段方法中,Zhang 等<sup>[28]</sup>通过双阶段级联网络获得较高的肾脏分割精度,参数量为 85.54 MB,Hou 等<sup>[29]</sup>提出的三阶段级联分割网络获得优秀的肾肿瘤分割精度和平均分割精度,总参数量为 271.35 MB。多阶段方法需要训练多级网络,参数量较大,训练的负载较高,推理成本较大,而所提方法只需要训练一个端到端的网络模型,且肾肿瘤分割精度优于双阶段方法,肾脏分割精度优于三阶段方法,平均分割性能较优,能够在参数量较小的前提下获得较高的分割精度,在一些需要迅速为病人提供诊断结果的场景下和移动端、小内存的嵌入式医疗设备系统中具有更高的应用价值。

表 2 所提方法与其余方法在 KiTS19 测试集上的分割性能对比

Table 2 Comparison between our method and other methods on the KiTS19 test dataset

Method	Stage	Parameter quantity /MB	Dice score		
			Kidney	Tumor	Average
Method in Ref. [25]	Single	19.55	0.9639	0.7891	0.8765
Method in Ref. [26]	Single	17.39	0.9619	0.7675	0.8647
Method in Ref. [27]	Single	29.65	0.9740	0.8180	0.8960
Method in Ref. [28]	Double-stage	85.54	0.9742	0.8306	0.9024
Method in Ref. [29]	Triple-stage	271.35	0.9674	0.8454	0.9064
Ours	Single	4.60	0.9706	0.8311	0.9008

## 4 结 论

提出一种 TransRFE-U<sup>2</sup>-Net 用于肾脏及肾肿瘤医学图像分割。首先将 U<sup>2</sup>-Net 升维,用于处理三维输入数据,为避免过度下采样引起的信息丢失,减少了外部与嵌套结构中的下采样次数,优化激活函数为 Leaky ReLU,并用实例标准化替代批量标准化以处理小批量输入样本,实验证明升维后的三维 U<sup>2</sup>-Net 具有较好的分割精度与较小的参数量,将其作为基准网络并进行后续优化。在多级解码端融合残差特征强化分支,增强了解码端特征图的前景特征表达能力,减少了目标区域周围噪声对分割性能的影响。针对 CNN 感受野有限的问题,参考 Transformer 的多头自注意力结构,提出基于全局特征的多头自注意力模块,将深层特征图转换为类 Transformer 的输入序列形式,计算各个子序列间的相互依赖关系,对应于各个子图像块间的上下文关联(当子图像块取 1×1×1 单位大小,则对应于体素与体素间的依赖关系),提高模型对全局信息的把控能力。此外,针对医学图像数据的各向异性,采用重采样和强度裁剪对图像进行预处理,针对分割结果中产生的假阳性点,使用连通域分析法对图像进行后处理。将 Dice 损失函数与交叉熵损失函数结合,组成新的联合损失函数进行网络训练,加快训练时的收敛速度并提升分割精度,最终在 KiTS19 测试集上取得了 0.9706 和 0.8311 的肾脏及肾肿瘤分割精度。实验结果表明:所提单阶段端到端的分割方法优于多数肾脏及肾肿瘤图像分割方法,且具有较少的参数量,在小内存嵌入式系统等应用场景下具有较高的参考价值。

## 参 考 文 献

- [1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: A Cancer Journal for Clinicians, 2018, 68(6): 394-424.
- [2] 姜宁宇, 张鹏, 廖良忠, 等. CT 与 MRI 对肾脏囊性病变诊断与鉴别诊断的临床价值探讨[J]. 中外医学研究, 2021, 19(28): 79-82.
- [3] 杨勃, 赵于前, 张帆, 等. 基于空间模糊 C 均值和图割的肝脏 CT 序列图像分割[J]. 激光与光电子学进展, 2022, 59(12): 1217002.
- [4] Ma W Y, Manjunath B S. EdgeFlow: a technique for boundary detection and image segmentation[J]. IEEE Transactions on Image Processing, 2000, 9(8): 1375-1388.
- [5] Batenburg K J, Sijbers J. Optimal threshold selection for tomogram segmentation by projection distance minimization [J]. IEEE Transactions on Medical Imaging, 2009, 28(5): 676-686.
- [6] Ugarriza L G, Saber E, Vantaram S R, et al. Automatic image segmentation by dynamic region growth and multiresolution merging[J]. IEEE Transactions on Image Processing, 2009, 18(10): 2275-2288.
- [7] Sangewar S, Peshattiwari A A, Alagdeve V, et al. Liver segmentation of CT scan images using K means algorithm [C]//2013 International Conference on Advanced Electronic Systems (ICAES), September 21-23, 2013, Pilani, India. New York: IEEE Press, 2013: 6-9.
- [8] Zhang N, Su R, Lebonvallet S, et al. Multi-kernel SVM based classification for brain tumor segmentation of MRI multi-sequence[C]//2009 16th IEEE International Conference on Image Processing (ICIP), November 7-10, 2009, Cairo. New York: IEEE Press, 2010: 3373-3376.
- [9] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [10] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [11] Milletari F, Navab N, Ahmadi S A. V-net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International



- Conference on 3D Vision (3DV), October 25-28, 2016, Stanford, CA, USA. New York: IEEE Press, 2016: 565-571.
- [12] 牟海维, 郭颖, 全星慧, 等. 基于改进 U-Net 的磁共振成像脑肿瘤图像分割[J]. 激光与光电子学进展, 2021, 58(4): 0410022.  
Mu H W, Guo Y, Quan X H, et al. Magnetic resonance imaging brain tumor image segmentation based on improved U-net[J]. Laser & Optoelectronics Progress, 2021, 58(4): 0410022.
- [13] 褚晶辉, 黄凯隆, 吕卫. 一种级联改进 U-Net 网络的脑肿瘤分割方法[J]. 激光与光电子学进展, 2021, 58(8): 0810020.  
Chu J H, Huang K L, Lü W. A method for brain tumor segmentation using cascaded modified U-net[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0810020.
- [14] 左蒙, 刘意杨, 崔好, 等. 基于稀疏卷积和注意力机制的点云语义分割方法[J]. 激光与光电子学进展, 2023, 60(20): 2015002.  
Zuo M, Liu Y Y, Cui H, et al. Semantic segmentation method of point clouds based on sparse convolution and attention mechanism[J]. Laser & Optoelectronics Progress, 2023, 60(20): 2015002.
- [15] 单成响, 李镔, 关欣. 多视图卷积轻量级脑肿瘤分割算法[J]. 激光与光电子学进展, 2023, 60(10): 1010018.  
Shan C X, Li Q, Guan X. Lightweight brain tumor segmentation algorithm based on multi-view convolution [J]. Laser & Optoelectronics Progress, 2023, 60(10): 1010018.
- [16] Qin X B, Zhang Z C, Huang C Y, et al. U<sup>2</sup>-Net: going deeper with nested U-structure for salient object detection [J]. Pattern Recognition, 2020, 106: 107404.
- [17] Vaswani A, Shazeer N, Parmar N, et al. Attention is all You need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, California, USA. New York: ACM Press, 2017: 6000-6010.
- [18] Wang L J, Lu H C, Ruan X, et al. Deep networks for saliency detection via local estimation and global search [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3183-3192.
- [19] Zhang Y, Hua Q L, Xu D, et al. A complex-valued convolutional neural network with different activation functions in polarimetric SAR image classification[C]//2019 International Radar Conference (RADAR), September 23-27, 2019, Toulon, France. New York: IEEE Press, 2020.
- [20] Zhao Z X, Chen K X, Yamane S. CBAM-UNet: easier to find the target with the attention module "CBAM" [C]//2021 IEEE 10th Global Conference on Consumer Electronics (GCCE), October 12-15, 2021, Kyoto, Japan. New York: IEEE Press, 2021: 655-657.
- [21] Hendrycks D, Gimpel K. Gaussian error linear units (GELUs) [EB/OL]. (2016-06-27) [2023-02-03]. <https://arxiv.org/abs/1606.08415>.
- [22] Sathianathen N, Heller N, Tejpal R, et al. Automatic segmentation of kidneys and kidney tumors: the KiTS19 international challenge[J]. Frontiers in Digital Health, 2022(3): 797607.
- [23] Hsiao C H, Lin P C, Chung L A, et al. A deep learning-based precision and automatic kidney segmentation system using efficient feature pyramid networks in computed tomography images[J]. Computer Methods and Programs in Biomedicine, 2022, 221: 106854.
- [24] Isensee F, Jaeger P F, Kohl S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nature Methods, 2021, 18(2): 203-211.
- [25] Kang L, Zhou Z Q, Huang J J, et al. Renal tumors segmentation in abdomen CT Images using 3D-CNN and ConvLSTM[J]. Biomedical Signal Processing and Control, 2022, 72: 103334.
- [26] Guo J N, Zeng W, Yu S, et al. RAU-net: U-net model based on residual and attention for kidney and kidney tumor segmentation[C]//2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), January 15-17, 2021, Guangzhou, China. New York: IEEE Press, 2021: 353-356.
- [27] Zhao W S, Jiang D H, Queralta J P, et al. MSS U-Net: 3D segmentation of kidneys and tumors from CT images with a multi-scale supervised U-Net[J]. Informatics in Medicine Unlocked, 2020, 19: 100357.
- [28] Zhang Y, Wang Y X, Hou F, et al. Cascaded volumetric convolutional network for kidney tumor segmentation from CT volumes[EB/OL]. (2019-10-05) [2023-02-03]. <https://arxiv.org/abs/1910.02235>.
- [29] Hou X S, Xie C M, Li F Y, et al. A triple-stage self-guided network for kidney tumor segmentation[C]//2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), April 3-7, 2020, Iowa City, IA, USA. New York: IEEE Press, 2020: 341-344.