

面向尿路结石CT图像的DAST-UNet 双重注意力自动分割模型

孙海刚 张叶飞 刘彦斌

摘要 目的 腹部CT图像中尿路结石存在形态多样、边界模糊及与周围组织对比度低等问题,给自动分割任务带来了显著挑战,尤其在小目标检测与复杂背景干扰下,传统方法往往难以兼顾精度与鲁棒性。为此,本文提出一种融合双重注意力机制的DAST-UNet模型,用于实现尿路结石的高精度自动分割。**方法** 所提模型以Swin-Unet为主干结构,设计了CASI(channel and spatial interactive)模块与TDA(token dependency attention)模块,分别从通道-空间维度和局部-全局语义关系两个层面增强模型特征表达能力。整体采用对称的编码器-解码器架构,通过多尺度特征融合与逐步上采样策略提升分割性能。其中,CASI模块增强模型对目标区域的关注并有效抑制背景噪声,TDA模块则通过加强Token之间的依赖关系,有效捕捉细粒度结构特征,优化上下文信息建模。**结果** 在构建的尿路结石CT图像数据集上,DAST-UNet在多个关键性能指标上均优于Res Unet、SegNet与Swin-Unet等经典模型,最终在测试集上获得Dice系数80.34%、灵敏度80.21%、准确率89.91%和IoU 67.14%。训练过程中模型损失函数快速收敛,验证集loss与训练集趋势一致,表明网络训练过程稳定,未出现明显过拟合。消融实验进一步验证了CASI和TDA模块对性能提升的具体贡献,Swin-Unet+CASI模型在Dice、灵敏度等指标上有明显改进,继续引入TDA后性能进一步提升,最终DAST-UNet在所有组合中表现最佳。**结论** 本文提出的DAST-UNet模型在结构上实现了对局部细节与全局上下文的协同建模,有效提升了CT图像中尿路结石的自动分割精度。实验结果充分验证了双重注意力机制对小目标识别与复杂背景处理的积极作用,为相关临床应用提供了一种可靠且具有推广价值的自动化解决方案。

关键词 结石分割;CT图像;DAST-UNet;CASI;TDA

DOI: 10.3969/j.issn.1002-3208.2026.02.003.

中图分类号 R318.04 **文献标志码** A **文章编号** 1002-3208(2026)02-0127-10

本文著录格式 孙海刚,张叶飞,刘彦斌. 面向尿路结石CT图像的DAST-UNet双重注意力自动分割模型[J]. 北京生物医学工程, 2026, 45(2): 127-136. SUN Haigang, ZHANG Yefei, LIU Yanbin. Automatic segmentation of urinary stones in CT images using DAST-UNet with dual attention mechanism[J]. Beijing Biomedical Engineering, 2026, 45(2): 127-136.

Automatic segmentation of urinary stones in CT images using DAST-UNet with dual attention mechanism

SUN Haigang, ZHANG Yefei, LIU Yanbin

Taiyuan Central Hospital, Taiyuan 030009

Corresponding author: SUN Haigang (E-mail: sunhaigang2025@163.com)

【Abstract】 Objective Urinary tract stones in abdominal CT images present significant challenges for automatic segmentation, particularly due to their diverse morphologies, blurred boundaries, and low contrast with surrounding tissues. Traditional methods often struggle to achieve a balance of accuracy and robustness,

particularly when detecting small objects and encountering complex background interference. Therefore, a DAST-UNet model is proposed that incorporates a dual attention mechanism to achieve high-precision automatic segmentation of urinary tract stones. **Methods** The proposed model utilizes the Swin-Unet backbone architecture and incorporates a CASI (channel and spatial interactive) module and a TDA (Token dependency attention) module to enhance the model's feature representation capabilities from the channel-spatial dimension and the local-global semantic relationship level, respectively. A symmetric encoder-decoder architecture is employed to improve segmentation performance through multi-scale feature fusion and a progressive upsampling strategy. The CASI module enhances the model's focus on the target region and effectively suppresses background noise, while the TDA module effectively captures fine-grained structural features and optimizes contextual information modeling by strengthening dependencies between Tokens. **Results** On a constructed urinary tract calculi CT image dataset, DAST-UNet outperforms classic models such as Res Unet, SegNet, and Swin-Unet across multiple key performance metrics, ultimately achieving a Dice coefficient of 80.34%, sensitivity of 80.21%, accuracy of 89.91%, and intersection over union (IoU) of 67.14% on the test set. The model loss function converges rapidly during training, with the validation set loss trend consistent with the training set, indicating that the network training process is stable and free of significant overfitting. Ablation experiments further validate the specific contributions of the CASI and TDA modules to performance improvement. The Swin-Unet + CASI model shows significant improvements in metrics such as Dice and sensitivity. The introduction of TDA further improves performance, resulting in DAST-UNet performing the best among all combinations. **Conclusions** The proposed DAST-UNet model achieves a coordinated modeling of local details and global context, effectively improving the accuracy of automatic segmentation of urinary tract calculi in CT images. The experimental results fully verify the positive effect of the dual attention mechanism on small target recognition and complex background processing, and provide a reliable and scalable automated solution for related clinical applications.

【Keywords】 urinary stone segmentation; CT image; DAST-UNet; CASI; TDA

0 引言

尿路结石是临床上最常见的泌尿系统疾病之一,通常起源于肾脏,但可沿尿路发生在输尿管、膀胱及尿道等部位。患者常伴有腰痛、血尿等症状,严重时可导致尿路梗阻,影响肾功能。因此,精准高效地检测和识别尿路结石对于临床诊断和治疗方案的制定至关重要。

随着尿路结石发病率的上升,诸如平扫CT、增强CT、磁共振成像(MRI)和超声等成像方法在确保准确及时诊断中起着关键作用^[1-3]。目前,CT扫描因其高分辨率和对软组织的良好对比度,被广泛应用于泌尿系结石的诊断。然而,当结石的灰度值与

周围组织相近时,形态不规则且大小差异显著,在CT扫描图像中快速识别结石对于医生而言依旧是一项繁琐而困难的任务。

近年来,人工智能和深度学习技术的进步,为医学图像分割提供了新的解决方案。Babajide等^[4]分析了U-Net在自动识别结石和测量结石方面的性能,验证了机器学习算法在结石识别方面的可行性和有效性。Huang等^[5]使用Inception-ResNetV2模型和ResNet混合U-net模型实现了肾结石轮廓的识别。Kilic等^[6]使用YOLOv4和Mask R-CNN架构设计了检测肾结石的算法模型,以协助临床专家完成肾结石检测。Shetty等^[7]提出了一种使用卷积神经网络(convolutional neural network, CNN)在医学成像中自动诊断尿路结石的方法。Preedanani等^[8]提出了一种基于MultiResUnet的算法模型,可以完成从全腹部图像中分割出结石。Li等^[9]借助现有模型3D U-Net、Res Unet、SegNet、DeepLabV3+和UNETR

基金项目:国家区域医疗中心科技创新计划项目(202212)资助

作者单位:太原市中心医院(太原 030009)

通信作者:孙海刚,副主任医师。E-mail:sunhaigang2025@163.com

等模型,在腹部CT数据集中验证了肾脏和肾结石检测的可能性。尽管CNN模型在图像特征提取方面具有良好表现,但其局部感受野有限,难以有效捕捉小目标、区分结构相似的组织区域,且易受背景噪声干扰。近年来,Transformer架构凭借其自注意力机制在建模长距离依赖关系和全局特征提取方面表现出色,逐渐应用于医学影像分析。Choi等^[10]借助ViT模型,应用于CT图像中实现尿路结石的有效检测。Islam等^[11]利用ViT和可解释的迁移学习模型,从CT图中自动检测肾囊肿、结石和肿瘤。然而,Transformer架构在复杂背景下仍可能丢失局部细节,影响对小目标的精准分割。因此,在CT图像中尿路结石的多样性及其与周围组织的低对比度,使得结石分割任务面临精度不足的挑战。

为提升CT图像中复杂结石区域的识别精度与鲁棒性,本文提出了一种基于双重注意力机制的DAST-UNet (dual attention swin Transformer Unet, DAST-UNet)模型。DAST-UNet模型融合Swin-Unet主干、CASI(channel attention spatial intersection)模块和TDA(token dual attention)模块,旨在优化尿路结石分割的特征表达能力,提高复杂背景下的分割精度与鲁棒性。其中,Swin-Unet主干网络采用分层Swin Transformer作为编码器,通过滑动窗口机制提取局部与全局信息,以增强对多尺度特征的处理能力。CASI模块结合通道注意力与空间交集注意力,有效抑制背景噪声,增强模型对结石区域的关注度,并提升对视觉相似目标的区分能力。TDA模块通过建模Token间和Token内的注意力关系,进一步优化特征表达,并通过跳跃连接保留编码器中的自注意力特性,补偿深层网络中的信息丢失问题。

1 基于DAST-UNet的自动分割算法

本文提出的基于双注意力机制的DAST-UNet模型,结合Swin-Unet主干、CASI模块和TDA模块,实现对腹部CT图像中尿路结石的自动分割。DAST-UNet的模型结构如图1所示,该模型采用编码器-解码器架构,其中编码器负责提取并编码前景与背景特征;解码器则通过多尺度特征融合与上采样策略,逐步恢复图像细节与空间分辨率。

1.1 Swin-Unet

Swin-Unet是一种基于Swin Transformer的对称

编码-解码结构^[12],结合滑动窗口注意力与跳跃连接,增强特征提取与上下文建模能力,如图2所示。

编码器将图像划分为 4×4 不重叠块,特征维度为48,经线性嵌入投影至维度C后,输入多个Swin Transformer模块与Patch Merging层,提取分层特征并完成下采样。解码器结构对称,由Swin Transformer与Patch Expanding层组成,通过跳跃连接融合多尺度特征,补偿空间信息损失。Patch Expanding实现逐步上采样,最后一层执行 $4 \times$ 上采样,将特征图恢复至原始分辨率,并通过线性投影输出像素级分割结果。

在骨干网络中,Swin Transformer通过窗口划分提取局部特征,并借助窗口移位策略增强全局建模能力。每个模块包括LayerNorm、多头自注意力(multi-head self-attention,MSA)、残差连接及含GELU激活的双层多层感知机(multi-layer perceptron,MLP)结构,如图3所示。

在连续的Swin Transformer块中,首先应用基于窗口的多头自注意力(W-MSA)模块以处理局部信息;随后,在下一个Transformer模块中,采用基于移位窗口的多头自注意力模块,以增强全局上下文信息的建模能力。窗口划分与移位窗口相结合的机制,提升Swin Transformer对长距离依赖关系的建模能力。Swin Transformer块可以表示为式(1)。

$$\begin{aligned}\hat{z}^l &= W - \text{MSA}(\text{LN}(z^{l-1})) + z^{l-1} \\ z^l &= \text{MLP}(\text{LN}(\hat{z}^l)) + \hat{z}^l \\ \hat{z}^{l+1} &= SW - \text{MSA}(\text{LN}(z^l)) + z^l \\ z^{l+1} &= \text{MLP}(\text{LN}(\hat{z}^{l+1})) + \hat{z}^{l+1}\end{aligned}\quad (1)$$

式中: $\hat{z}^l$ 和 z^l 分别表示 l^{th} 个块中(S)W-MSA模块MLP模块的输出,自注意力的计算如式(2)所示。

$$\text{Attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V \quad (2)$$

式中: $Q, K, V \in R^{M \times d}$ 分别表示查询(Query)、键(Key)和值(Value)矩阵; M 和 d 分别表示窗口中patch的数量和query、key的维度;矩阵 B 中的值来自偏置矩阵 $\hat{B} \in R^{(2M-1) \times (2M+1)}$ 。

1.2 CASI模块

为解决CT图像的空间非均匀性问题,本文在网络结构中引入了CASI模块。CASI模块由空间交集(spatial intersection, SI)模块和通道注意(channel

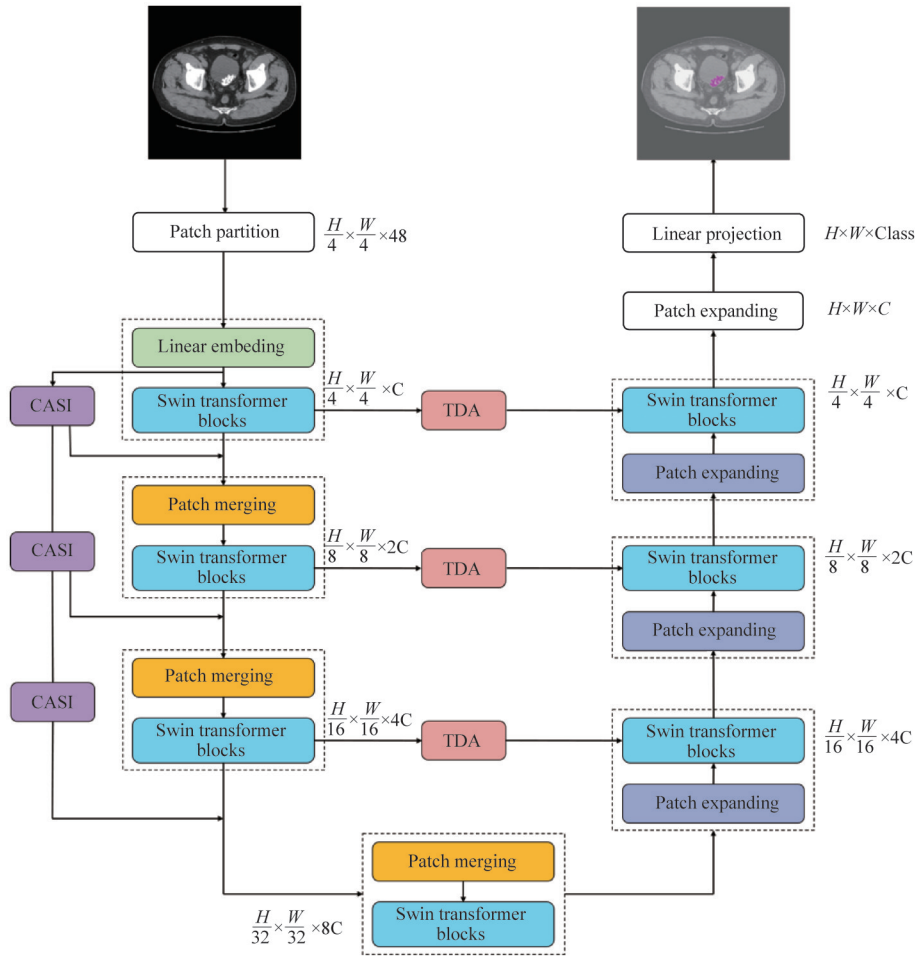


图1 DAST-Unet系统框架图

Figure 1 DAST-Unet system framework

attention, CA)模块组成。

SI模块通过扩展感受野内的平均池化抑制噪声,增强结石前景与背景对比,缓解Swin Transformer的全局信息损失。为提升结石与相似组织的区分能力,CASI模块引入CA模块,融合多波段特征,优化通道表示,并细化空间注意力权重,从而协同建模局部与全局信息,提升分割精度。CASI的结构如图4所示。

附加分支由三个级联的CASI模块组成,分别对应主干网络的三个阶段,以增强特征提取能力和全局上下文信息建模。CASI分支与主干网络的第一编码阶段共享相同的输入,表示为 $X \in \mathbb{R}^{L \times C}$ 。在进入第一个CASI模块之前,输入图像首先被重新塑形为二维特征图 $X_{\text{CASI-}r} \in \mathbb{R}^{H \times W \times C}$ 。

注意力建模由两级卷积块构成,结合 3×3 与

1×1 卷积,并配合批量归一化(batch normalization, BN)层和ReLU激活函数层,增强特征表达与训练稳定性。

第一个卷积块的输出经过 2×2 平均池化后,作为输入传递至后续的CASI模块。第二个卷积块的输出则同时作为CA模块和SI模块的输入。其中,在SI模块中,对特征图 $X_{\text{CASI-conv2}} \in \mathbb{R}^{H \times W \times C}$ 进行行池化操作,以进一步增强空间信息提取能力,具体计算如式(3)所示。

$$S_w = \text{AvgPool}_{\text{row}}(X_{\text{CASI-conv2}}) \quad (3)$$

S_w 的形状为 $H \times 1 \times C$ 。类似于行池化,列池化计算每列的平均值,使得特征图在列方向上进行压缩。经过列池化操作后,特征图的尺寸从 $H \times W \times C$ 变为 $1 \times W \times C$,如式(4)所示。

$$S_h = \text{AvgPool}_{\text{column}}(X_{\text{CASI-conv2}}) \quad (4)$$

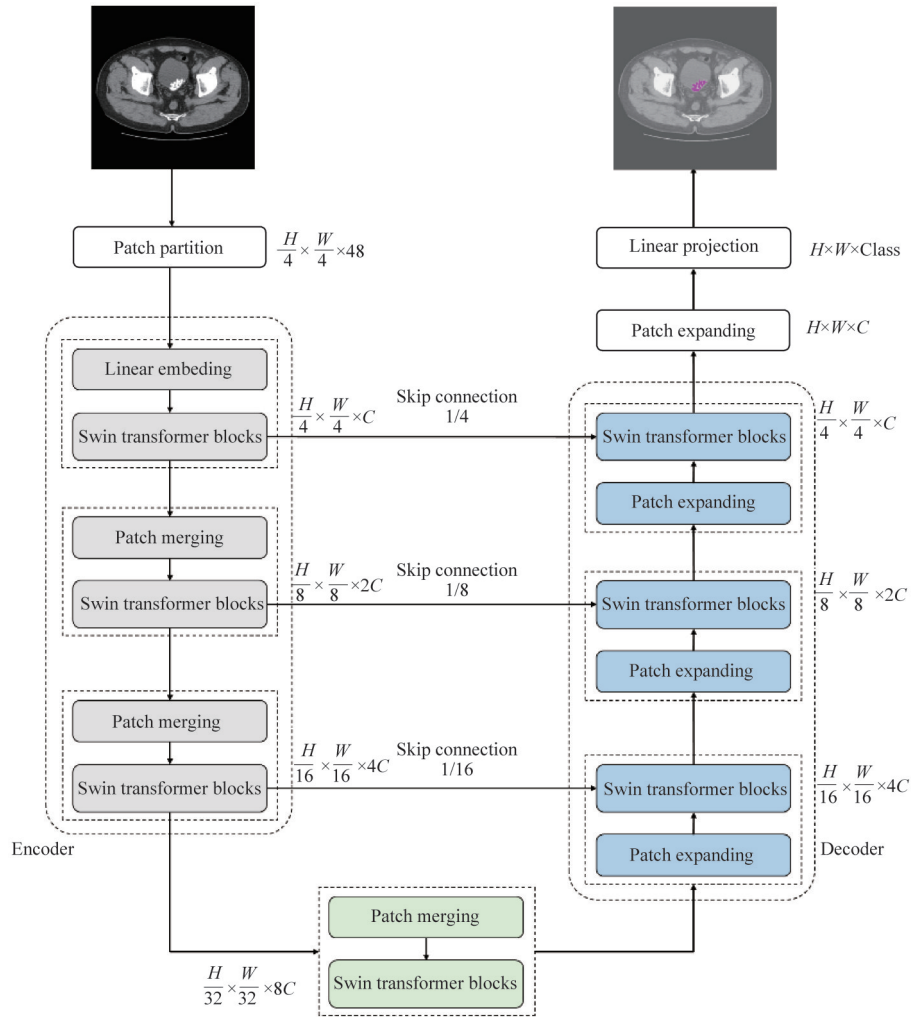


图 2 Swin-Unet 系统框架图

Figure 2 Swin-Unet system framework

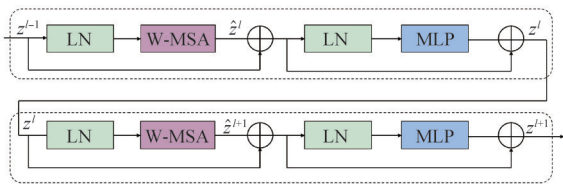


图 3 Swin Transformer 模块

Figure 3 Swin Transformer module

SI 模块通过行池化和列池化后进行矩阵乘法得到输出值,从而融合水平与垂直方向的信息,最终生成一个形状为 $H \times W \times C$ 的特征矩阵。空间权重特征图 W_s 通过对该矩阵应用 Sigmoid 函数进行归一化,以确保权重值位于 $[0, 1]$ 之间,具体计算如式 (5) 所示。

$$W_s = \text{Sigmoid}(S_w \times S_h) \quad (5)$$

CA 模块同样以 $X_{\text{CASI-conv2}}$ 的特征图作为输入,如式 (6) 所示。先进行全局平均池化(global average pooling, GAP)提取通道响应,再归一化生成注意力权重 W_c ,用于强化关键通道特征。

$$W_c = \text{Sigmoid}(\text{AvgPool}_{\text{global}}(X_{\text{CASI-conv2}})) \quad (6)$$

通道注意力权重 W_c 用于更新空间注意力权重 W_s ,如式 (7) 所示,最终生成 $H \times W \times C$ 加权特征图 W_{CASI} ,确保其与骨干 Swin Transformer 块中的特征图 X_{Trans} 对齐。通过将 X_{Trans} 和 W_{CASI} 进行逐元素乘法,得到整个编码器阶段的最终输出,增强网络对目标区域的关注,提高特征表达能力,并优化结石分割任务的性能。

$$\begin{aligned} W_{\text{CASI}} &= W_c \odot W_s \\ X_{\text{t-C}} &= X_{\text{trans}} \odot W_s \end{aligned} \quad (7)$$

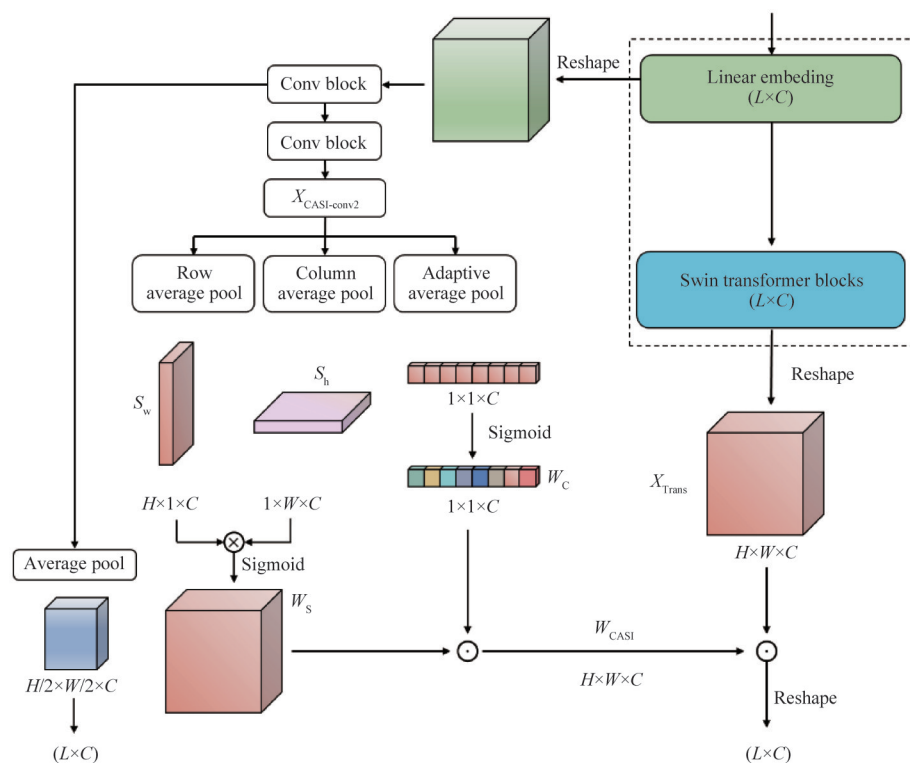


图4 CASI的结构图

Figure 4 CASI structure diagram

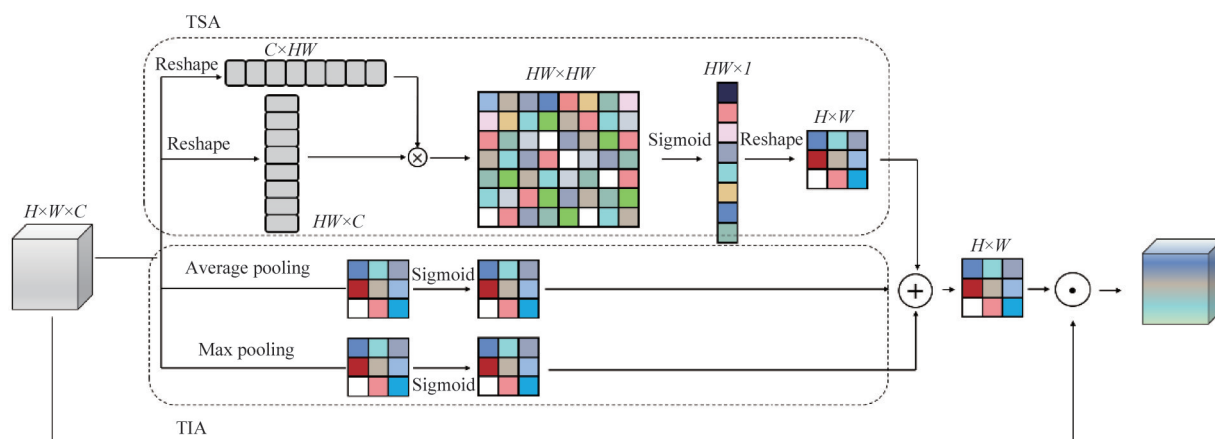


图5 TDA 结构图

Figure 5 TDA structure diagram

式中: $\odot$ 表示实数逐元素乘法。CASI模块在空间与通道维度扩展注意力范围,增强复杂结石特征建模,提升识别与分割精度。同时,其独立分支结构减少冗余计算,提升网络效率与泛化能力。

1.3 TDA 模块

DAST-UNet 设计了基于 TDA 的跳跃连接,以融合分层注意力信息^[13],其在模型中的位置如图 1 中

所示。TDA 模块采用双分支结构,保持 Token 自注意力特性,强化特征一致性与关键信息表达,抑制无关干扰。其仅含两个超参数 α 和 β ,在提升鲁棒性的同时保持高效计算能力。TDA 结构如图 5 所示。

每个 TDA 模块以编码器阶段 1 到阶段 3 的输出特征图 $x^{(s)}$ 作为输入,其中 $s=1,2,3$ 。首先,TDA 模块

对输入特征图 $x^{(s)}$ 进行重塑为矩阵形式,以便计算标记相似度 $S^{(s)}$ 。定义通道向量化运算符如式(8)所示。

$$\text{vec}: R^{\frac{H}{2^{s+1}} \times \frac{W}{2^{s+1}} \times 2^{s-1}} \rightarrow R^{\frac{HW}{2^{2s+2}} \times 2^{s-1}} \quad (8)$$

则标记相似度 $S^{(s)}$ 可表示为式(9)。

$$S_{(i,j)}^{(s)} = \text{vec}\left(x_{(i)}^{(s)}\right)g\left(\text{vec}\left(x_{(j)}^{(s)}\right)\right)^T \quad (9)$$

式中: $S^{(s)}$ 的第 (i,j) 个元素表示 $x^{(s)}$ 中第 i 个 Token 和第 j 个 Token 之间的相似度,其中 $1 \leq i, j \leq HW/2^{2s+2}$ 。进一步计算 Token 相似度注意力(token similarity attention, TSA),该注意力机制通过对 $S^{(s)}$ 进行逐行平均池化,测量一个 Token 与其他 Token 之间的相似度。编码器第 s 阶段的 TSA 特征图 $M_{\text{TSA}}^{(s)}$ 计算如式(10)所示。

$$M_{\text{TSA}}^{(s)} = \text{reshape}\left(\sigma\left(\text{AvgPool}\left(S^{(s)}; 2\right)\right); \frac{H}{2^{s+1}}, \frac{W}{2^{s+1}}\right) \in R^{\frac{H}{2^{s+1}} \times \frac{W}{2^{s+1}}} \quad (10)$$

式中: $\text{reshape}(x; h, w)$ 将 HW 维向量 x 映射为 $H \times W$ 维矩阵,作为 vec 的逆运算。符号 σ 表示 Sigmoid 激活函数,确保输出的注意力权重在 $[0, 1]$ 之间,以增强模型对关键特征的关注度。

此外, $\text{AvgPool}(X; 2)$ 表示沿矩阵 X 的列维度进行平均池化运算,计算公式如式(11)所示。

$$\text{AvgPool}(X; 2) = \frac{1}{\text{size}(X; 2)} Xg1 \quad (11)$$

式中: 1 为全 1 向量, $\text{size}(X; 2)$ 表示 X 在第二维度(列维度)的大小。该操作用于提取每个 Token 相对于所有其他 Token 的全局相似度信息,以优化注意力机制的计算过程。

与 $S^{(s)}$ 不同, TSA 特征图 $M_{\text{TSA}}^{(s)}$ 中的每个元素表示一个 Token 的显著性或权重,其值通过聚合该 Token 与所有其他 Token 的相似性来计算。通过这种方式, TSA 从全局角度提取了 CT 图像中与结石区分度较低的特征,使模型能够更全面地建模图像中的信息分布。

此外,引入了标记重要性注意力(token importance attention, TIA),以进一步捕获 CT 图像中最具辨别性的结石特征。TIA 机制基于每个像素在通道维度上的最大池化(max pooling)和平均池化(average pooling),以增强结石区域的显著性,提高

模型对关键区域的聚焦能力。对于 $s=1, 2, 3$, 编码器第 s 阶段的 TIA 特征图 $M_{\text{TIA}}^{(s)}$ 的定义如式(12)所示。

$$M_{\text{TIA}}^{(s)} = \alpha \sigma\left(\text{AvgPool}\left(x^{(s)}; 3\right)\right) + (1 - \alpha) \sigma\left(\text{MaxPool}\left(x^{(s)}; 3\right)\right) \quad (12)$$

式中: $\text{AvgPool}(x; 3)$ 和 $\text{MaxPool}(x; 3)$ 分别表示对三阶张量 x 在通道维度上的平均池化和最大池化操作;正则化因子 $0 < \alpha < 1$ 用于平衡池化操作中的信息聚合和抑制过拟合。通过这种方式, TIA 特征图从局部视角捕捉了每个 Token 在通道维度上的显著性,有效增强了对结石区域和其他重要特征的关注。

最终,通过将 TSA 和 TIA 图合并,得到输入特征 $x^{(s)}$ 的 TDA 图。需要注意的是, $M_{\text{TSA}}^{(s)}$ 值越大,表示每个标记相对于所有其他标记的相似度越大。因此,为了表示整体 Token 注意力,使用全 1 矩阵对 $M_{\text{TSA}}^{(s)}$ 进行减法运算,以去除背景信息并强调重要区域。最终, TDA 图 $M_{\text{TDA}}^{(s)}$ 通过式(13)表示。

$$M_{\text{TDA}}^{(s)} = \beta(1 - M_{\text{TSA}}^{(s)}) + (1 - \beta)M_{\text{TIA}}^{(s)} \quad (13)$$

式中: $0 < \beta < 1$ 是正则化因子。最后,每个 TDA 模块的输出由式(14)得到。

$$f_{\text{TDA}}(x^{(s)}) = M_{\text{TDA}}^{(s)} \odot x^{(s)} \quad (14)$$

式中: $\odot$ 表示沿 $x^{(s)}$ 的每个通道进行的元素乘法; $f_{\text{TDA}}(x^{(s)})$ 作为连续 swin-Transformer 块的输入。

2 实验

2.1 数据集

本研究使用了来自北京大学第一医院太原医院/太原市中心医院的腹部 CT 图像数据集,包括 1 072 张含结石图像和 400 张不含结石的图像。数据集中的结石分布部位多样,包括左肾、右肾、输尿管和膀胱等,所有图像均已匿名化处理。所有样本均用于 k 倍交叉验证的训练和测试。

为解决数据规模有限与类别不平衡问题,本研究采用裁剪、旋转、缩放及亮度、对比度、色相等多种图像增强策略,提升模型泛化能力。最终共生成 4 612 张增强图像,并标注 5 237 个结石,构建用于训练与验证的实验数据集,结石示例如图 6 所示。

2.2 模型训练和评估

本研究在 NVIDIA Tesla V100(16GB) GPU 上,

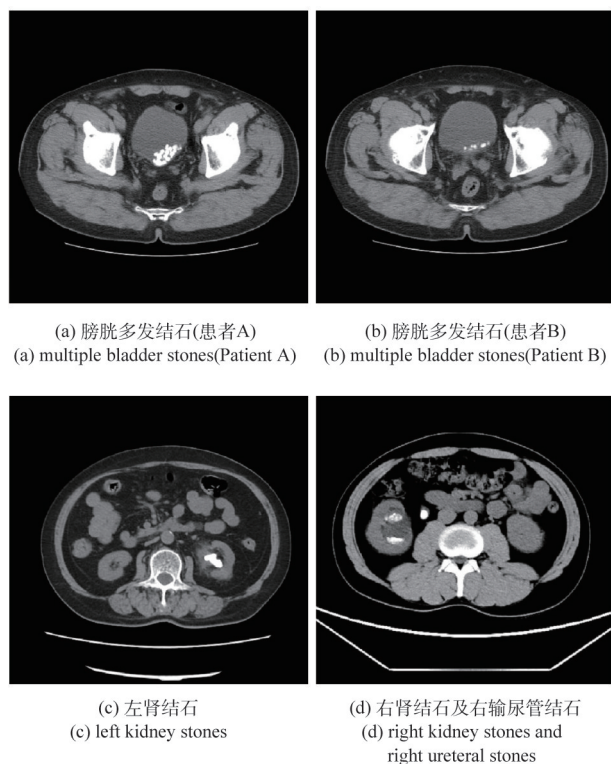


图 6 CT 图中结石示例图

Figure 6 Example of stones in CT images

基于 PyTorch (Python 3.9) 完成模型训练。采用 Adam 优化器, 初始学习率为 0.001, 训练集设为 shuffle 模式, 训练轮数设为 60 轮, 以兼顾计算效率与模型收敛性能。

在每个训练周期结束后, 完成验证, 并保存验证准确率最高的权重文件, 以便后续测试使用。为了进行有效训练, 选择交叉熵损失 (cross-entropy loss, CE) 作为损失函数, 具体公式如式 (15) 所示:

$$L_{CE} = - \sum_i^h \sum_j^w \sum_c^C y_{i,j}^c \times \lg \frac{e^{\hat{y}_{i,j}^c}}{\sum_c e^{\hat{y}_{i,j}^c}} \quad (15)$$

式中: $y_{i,j}^c$ 表示图像中 (i,j) 位置上的像素真实标签的 one hot 编码; $\hat{y}_{i,j}^c$ 模型为预测该像素属于结石的概率。最终的损失是所有像素损失的平均值。

为了全面评估模型的性能, 使用 Dice 系数、特异性、敏感性、准确度和 IoU 等常见的评估指标。所有这些指标均基于混淆矩阵中的 4 个基本元素: 真阳性 (TP)、假阳性 (FP)、真阴性 (TN) 和假阴性 (FN), 定义如式 (16) 所示:

$$\begin{aligned} \text{Dice} &= \frac{2TP}{2TP + FP + FN} \\ \text{Specificity} &= \frac{TN}{TN + FP} \\ \text{Sensitivity} &= \frac{TP}{TP + FN} \\ \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ \text{IoU} &= \frac{TP}{TP + FP + FN} \end{aligned} \quad (16)$$

式中: TP 表示正确预测为正类的样本数; FP 表示被错误预测为正类的负类样本数; TN 表示被正确预测为负类的样本数; FN 被错误预测为负类的正类样本数。

2.3 模型训练与验证

DAST-UNet 模型在训练过程中训练集与验证集损失值随 Epoch 变化的趋势如图 7 所示。蓝色曲线为训练损失, 橙色曲线为验证损失。可以观察到, 模型在初始阶段损失迅速下降, 约在第 60 轮后趋于稳定, 表明网络已基本收敛。训练与验证损失整体趋势一致, 二者数值接近, 未出现明显偏差, 说明模型在未见数据上保持了良好的泛化能力, 训练过程稳定有效, 参数设置合理。

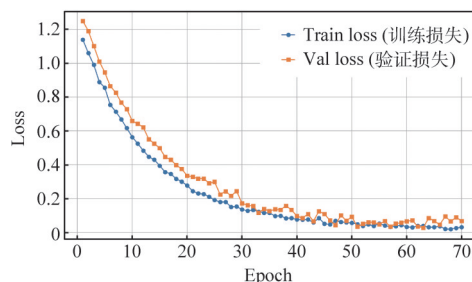


图 7 Loss 曲线

Figure 7 Loss curve

2.4 对比实验

为评估 DAST-UNet 性能, 本文与 SegNet、Res Unet、Swin-Unet 在自建数据集上进行了对比实验, 结果如表 1 所示, 可视化结果如图 8 所示。DAST-UNet 在所有指标上均优于其他模型, Dice 达 80.34%、IoU 为 67.14%, 显著优于 Res Unet (Dice: 72.45%, IoU: 56.8%)、SegNet (Dice: 71.92%, IoU: 56.15%) 和 Swin-Unet (Dice: 76.17%, IoU: 67.14%)。

Res Unet 基于 U-Net 架构, 擅长提取局部特征, 但在应对小而不规则的尿路结石时精度不足, 敏

表1 不同模型的对比结果(单位:%)

Table 1 Comparison results of different models(unit:%)

模型	准确性	特异性	灵敏性	Dice	IoU
SegNet	84.67	99.67	72.32	71.92	56.15
Res Unet	85.92	99.73	73.21	72.45	56.8
Swin-Unet	86.76	99.85	75.64	76.17	61.51
DAST-UNet	89.91	99.97	80.21	80.34	67.14

敏感性和Dice系数较低。相比之下,SegNet采用编码器-解码器结构与最大池化操作,提升了背景分割能力,但对小目标仍欠缺表达,尽管性能略优于Res Unet,仍难以在复杂背景下保持高敏感性。

Swin-Unet融合窗口自注意力机制,较传统CNN在全局与局部特征提取上更具优势,提升了敏感性和Dice系数。但在复杂背景下,受限于窗口划分,局部细节易丢失,小目标分割效果仍有限。

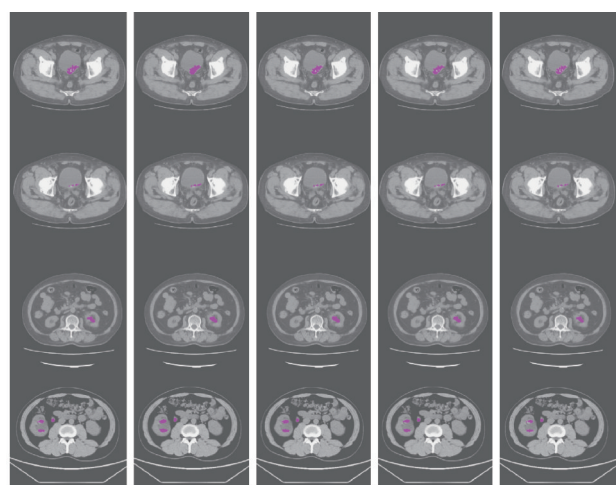
DAST-UNet模型通过引入双重注意力机制(CASI和TDA模块)显著提升了分割性能。CASI模块融合空间与通道注意力,有效增强了对尿路结石区域的关注,抑制背景噪声干扰,尤其在小目标与复杂背景的识别中表现出色。TDA模块则通过强化Token之间的依赖关系,进一步优化了特征表达能力与上下文建模效果。在此基础上,DAST-UNet在测试中整体优于其他对比模型,灵敏性和Dice系数均达到较高水平,展现出其在结石边缘细节和小目标精准识别方面的优势。此外,模型特异性高达99.97%,主要得益于网络对非目标区域特征的准确建模以及强背景抑制能力,有效降低了假阳性率,表明模型在复杂医学图像中具备出色的负类识别能力和稳健性。

2.5 消融实验

为了验证所提出改进策略的有效性,通过消融实验评估了不同改进模块对DAST-UNet模型性能的影响,具体实验结果如表2所示。

从表2可见,Swin-Unet在尿路结石分割中具备一定的全局与局部特征提取能力,准确率为86.76%,灵敏性为75.64%,Dice系数为76.17%。但由于其依赖窗口自注意力机制,易在复杂背景下丢失小目标信息,影响分割精度。为提升模型性能,引入CASI模块以融合空间与通道注意力,有效增强对结石区域的关注并抑制背景干扰。加入

CASI后,模型准确率提升至87.56%,灵敏性和Dice系数分别提升至77.32%和77.28%,验证了CASI在提升特征表达与小目标分割方面的有效性。进一步引入TDA模块,该模块通过加强Token之间的相互依赖关系,进一步提升了模型的细节捕捉能力。Swin-Unet+TDA在灵敏性(79.16%)和Dice系数(79.23%)上的表现较Swin-Unet+CASI有所提高,且准确率也达到了88.32%。这表明,TDA模块在提升全局特征建模能力、处理复杂背景和小目标分割任务中起到了重要作用。



(a) 真实值 (b) SegNet (c) Res Unet (d) Swin-Unet (e) DAST-UNet
(a) ground truth (b) SegNet (c) Res Unet (d) Swin-Unet (e) DAST-UNet

图8 不同模型的图像分割结果(紫色:结石分割区域)

Figure 8 Image segmentation results of different models
(Purple: Stone segmentation area)

表2 基于DAST-UNet算法模型的消融实验(单位:%)

Table 2 Ablation experiments based on the DAST-UNet algorithm model(unit:%)

模型	准确性	特异性	灵敏性	Dice	IoU
Swin-Unet	86.76	99.85	75.64	76.17	61.51
Swin-Unet+CASI	87.56	99.88	77.32	77.28	62.97
Swin-Unet+TDA	88.32	99.92	79.16	79.23	65.6
DAST-UNet	89.91	99.97	80.21	80.34	67.14

最终,DAST-Unet结合了CASI模块和TDA模块,在所有性能指标上表现最佳,特别是在灵敏性(80.21%)和Dice系数(80.34%)方面,较其他模型有显著提高。其高灵敏性和高Dice系数证明了双重注意力机制对细节和小目标精准识别的极大贡献。

3 讨论与结论

本文提出了一种基于双重注意力机制的 DAST-UNet 模型,用于尿路结石 CT 图像的自动分割。该模型融合了 Swin-Unet 主干与新设计的 CASI 和 TDA 模块,在特征表达与小目标识别方面进行了结构性改进。实验结果表明, DAST-UNet 在 Dice 系数(80.34%)和灵敏性(80.21%)等指标上显著优于现有主流模型,验证了所提方法在结石区域分割任务中的有效性和优势。

通过消融实验进一步分析了各模块对模型性能的具体贡献。在 Swin-Unet 基础上引入 CASI 模块后,准确率、灵敏性与 Dice 系数分别提升至 87.56%、77.32% 和 77.28%,表明空间与通道注意力的融合有效增强了模型对关键区域的聚焦能力,并减少了背景噪声干扰。继续加入 TDA 模块后,模型在细节建模方面进一步增强, Swin-Unet+TDA 在灵敏性(79.16%)和 Dice 系数(79.23%)上均高于仅引入 CASI 的模型,且 IoU 提升至 65.60%。最终,结合两者的 DAST-UNet 在所有评价指标上均取得最优表现,说明双重注意力机制在提升小目标检测精度与复杂背景处理能力方面具有显著效果。

参考文献

- [1] Yildirim K, Bozdag PG, Talo M, et al. Deep learning model for automated kidney stone detection using coronal CT images[J]. Computers in Biology and Medicine, 2021, 135: 104569.
- [2] Asif S, Zheng X, Zhu Y. An optimized fusion of deep learning models for kidney stone detection from CT images[J]. Journal of King Saud University-Computer and Information Sciences, 2024, 36(7): 102130.
- [3] El Beze J, Mazeaud C, Daul C, et al. Evaluation and understanding of automated urinary stone recognition methods[J]. BJU International, 2022, 130(6): 786-798.
- [4] Babajide R, Lembrikova K, Ziemba J, et al. Automated machine learning segmentation and measurement of urinary stones on CT scan[J]. Urology, 2022, 169: 41-46.
- [5] Huang ZH, Liu YY, Wu WJ, et al. Design and validation of a deep learning model for renal stone detection and segmentation on kidney-ureter-bladder images[J]. Bioengineering, 2023, 10(8): 970.
- [6] Kilic U, Karabey Aksakalli I, Tumuklu Ozyer G, et al. Exploring the effect of image enhancement techniques with deep neural networks on direct urinary system (DUSX) images for automated kidney stone detection[J]. International Journal of Intelligent Systems, 2023, 2023: 3801485.
- [7] Shetty M, Shetty SB, Sequeria GP, et al. Kidney stone detection using CNN[C]//2024 Second International Conference on Data Science and Information System (ICDSIS). Hassan, India: IEEE Press, 2024: 1-6.
- [8] Preedan W, Suzuki K, Kondo T, et al. Improvement of urinary stone segmentation using GAN-based urinary stones inpainting augmentation[J]. IEEE Access, 2022, 10: 115131-115142.
- [9] Li D, Xiao C, Liu Y, et al. Deep segmentation networks for segmenting kidneys and detecting kidney stones in unenhanced abdominal CT images[J]. Diagnostics, 2022, 12(8): 1788.
- [10] Choi HS, Kim JS, Whangbo TK, et al. Improved detection of urolithiasis using high-resolution computed tomography images by a vision transformer model[J]. International Neurourology Journal, 2023, 27(Suppl 2): S99-S103.
- [11] Islam MN, Hasan M, Hossain MK, et al. Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography[J]. Scientific Reports, 2022, 12(1): 11440.
- [12] Yao J, Jin S. Multi-category segmentation of Sentinel-2 images based on the Swin UNet method[J]. Remote Sensing, 2022, 14: 3382.
- [13] Du J, Guan K, Zhou Y, et al. Parameter-free similarity-aware attention module for medical image classification and segmentation[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, 7(3): 845-857.

(2025-06-04 收稿, 2025-08-16 修回)