E-mail: csa@iscas.ac.cn http://www.c-s-a.org.cn Tel: +86-10-62661041

用于 3D 器官图像分割的波随机自注意力编码器^①

周 迪¹, 刘 豪², 程远志^{2,3}, 李 辉¹, 刘晓亚¹

¹(青岛科技大学 数据科学学院, 青岛 266061) ²(青岛科技大学 信息科学技术学院, 青岛 266061) ³(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001) 通信作者: 程远志, E-mail: yzcheng2007@163.com

摘 要: 在光谱三维 CT 数据中, 传统卷积的全局特征捕捉能力不足, 而全尺度的自注意力机制则需要大量的计算 资源. 为了解决这一问题, 本文引入一种新视觉注意力范式 (wave self-attention, WSA). 相比于 ViT 技术, 该机制使 用更少的资源获得同等的自注意力信息. 此外, 为更充分地提取器官间的相对依赖关系并提高模型的鲁棒性和执行 速度, 本文为 WSA 机制设计了一种即插即用的模块——波随机编码器 (wave random encoder, WRE). 该编码器能 够生成一对互逆的非对称全局 (局部) 位置信息矩阵. 其中, 全局位置矩阵用来对波特征进行全局性的随机取样, 局 部位置矩阵则用于补充因随机取样而丢失的局部相对依赖. 本文在标准数据集 Synapse 和 COVID-19 的肾脏和肺 实质的分割任务上进行实验. 结果表明, 本文方法在精度、参数量和推理速率方面均超越了 nnFormer、Swin-UNETR 等现有模型, 达到了 SOTA 水平.

关键词: 医学影像; 图像分割; 波自注意力机制; 波随机编码器

引用格式:周迪,刘豪,程远志,李辉,刘晓亚.用于 3D 器官图像分割的波随机自注意力编码器.计算机系统应用,2025,34(2):84-91. http://www.c-s-a. org.cn/1003-3254/9768.html

Wave Random Self-attention Encoder for 3D Organ Image Segmentation

ZHOU Di¹, LIU Hao², CHENG Yuan-Zhi^{2,3}, LI Hui¹, LIU Xiao-Ya¹

¹(School of Data Science, Qingdao University of Science and Technology, Qingdao 266061, China) ²(School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China) ³(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: In spectral 3D CT data, the traditional convolution has a poor ability to capture global features, and the fullscale self-attention mechanism consumes large resources. To solve this problem, this study introduces a new visual attention paradigm, the wave self-attention (WSA). Compared with the ViT technology, this mechanism uses fewer resources to obtain the same amount of self-attention information. In addition, to more adequately extract the relative dependency among organs and to improve the robustness and execution speed of the model, a plug-and-play module, the wave random-encoder (WRE), is designed for the WSA mechanism. The encoder is capable of generating a pair of mutually inverse asymmetric global (local) position information matrices. The global position matrix is used to globally conduct random sampling of the wave features, and the local position matrix is used to complement the local relative dependency lost due to random sampling. In this study, experiments are performed on the task of segmenting the kidney and lung parenchyma in the standard datasets Synapse and COVID-19. The results show that this method outperforms existing models such as nnFormer and Swin-UNETR in terms of accuracy, the number of parameters, and inference rate, arriving at the SOTA level.

Key words: medical imaging; image segmentation; wave self-attention (WSA) mechanism; wave random encoder (WRE)



① 基金项目: 国家重点研发计划 (2023YFF0612102); 青岛市重点科技攻关及产业化示范项目 (23-7-2-qljh-4-gx, 24-1-2-qljh-19-gx) 收稿时间: 2024-07-13; 修改时间: 2024-08-13; 采用时间: 2024-09-03; csa 在线出版时间: 2024-12-19 CNKI 网络首发时间: 2024-12-20

⁸⁴ 系统建设 System Construction

CT 影像中的器官分割是临床医学中人体器官诊疗的一项关键任务. 传统专家手动勾勒器官的方法是费时且繁琐的, 而现有的自动化分割方法速度慢且分割边缘模糊, 因此快速且精确的器官分割方法成为医学图像研究中的重点. 相较于二维平面图像, 三维CT 图像分割任务存在更多的挑战: (1) 更高的维度会导致全局特征信息量的爆炸式增长; (2) 更大规模的数据会产生更巨大的数据量和计算量; (3) 相对统一的数据形状和分布导致其存在难发现的全局相对依赖.

在器官图像分割任务中最经典的是 U-Net^[1]模型, 该模型在多尺度编码器和多尺度解码器之间使用跳跃 连接来补充多尺度特征之间的特征损失,但结构冗余 严重,局部信息被多尺度特征稀释,容易出现欠分割问 题. V-Net^[2]进一步优化了 U-Net, 通过更复杂的特征提 取块和跳跃连接,提高了器官分割精度,但图像全局特 征提取不够充分. 图像注意力机制 (vision Transformer, ViT)^[3]的出现使医疗影像的分割精度得到了进一步的 提高, 如基于信道注意力的 MA-Unet^[4]、用注意力层 替换编码层的 U-Net^[5]和注意力模块替换卷积块的 Swin-Unet^[6]. 但简单的分块处理并不能压缩数据, 这导 致注意力机制需要存储大量的中间变量到显存中执行 大规模矩阵运算,从而使计算量和显存使用量大幅增加. 奇异值分解技术 (singular value decomposition, SVD)^[7] 有着除噪和降维的作用,多应用于线性特征提取[8-10]和 二维图像压缩^[11-13].本文将 SVD、自注意力机制 (selfattention)和 3D 体素 (Voxel)处理相融合,并引入了适 用于高维特征的波自注意力机制 (wave self-attention, WSA). WSA 既解决了全局特征信息量爆炸式增长的 问题,也大大降低了自注意力机制的数据量和计算量.

在全局相对依赖的提取上,常规方法是采用更大的卷积核^[14]和更深的层数^[15],但这样的方法会消耗大量的计算资源和更多的存储空间,He等人^[16]提出了 MAE 自编码器,通过图像的随机部分予以屏蔽 (Mask) 再重建丢失的像素的方法,从而降低特征规模,提取像 素之间的随机性依赖关系,并实现了类似于 dropout 的 效果,降低模型的过拟合问题.但是,MAE 模块需要对 输入的特征进行预训练操作,这要求输入特征是确定 的已知图像,但在训练阶段每个中间层的输入特征均 不确定,而确定的输入特征意味模型不能被修改,MAE 模块也失去了作用,这从根本上限制了 MAE 作为即插 即用的中间层模型.为解决这一问题,本文设计了波随 机编码器 (wave random encoder, WRE), 将随机掩码矩 阵采用传统算法转换为随机编码矩阵 EN 和随机解码 矩阵 DE, 解码矩阵 DE 仅通过屏蔽的特征位置进行像 素复原, 可以在训练的任意时刻完成特征的尺度复原 任务和多尺度残差连接的工作.

综上所述, 波自注意力机制的引入可以降低参与 计算的数据规模并提高分割精度; 采用波随机编码器 的方法可以作为即插即用的方式嵌入到现有的任意模 型中, 作为中间层提取更加丰富的随机性依赖关系, 并 一定程度上避免模型的过拟合问题. 基于这些方法, 本 文贡献如下工作.

(1)引入了波自注意力机制,通过特征降维的方式 生成波特征.从而,降低数据量级,减少资源消耗,增强 模型的全局特征提取能力.

(2) 解决 MAE 自编码器无法作为即插即用的中间 层的问题, 提出了波自编码器, 能在降低数据规模的同 时, 促使波自注意力机制捕捉更多的随机性依赖关系, 提高模型的器官定位和形状捕捉能力.

(3) 在肺实质及肾脏等器官上进行了实验, 验证了 该方法在多样化的器官大小和形状中分割任务的高效 性与精确性, 表现性能超越已知模型.

1 基本原理

1.1 波自注意力机制 WSA

奇异值分解技术 (singular value decomposition, SVD) 对任意一个矩阵 $H \in \mathbb{R}^{m \times n}$,都有正交矩阵 $U = (u_1, u_2, \dots, u_m) \in \mathbb{R}^{m \times m}$ 及正交矩阵 $V = (v_1, v_2, \dots, v_n) \in \mathbb{R}^{n \times n}$,使得 矩阵H满足:

$$H = U\Sigma V^{\mathrm{T}} \tag{1}$$

其中, $\Sigma = (\text{diag}(\alpha_1, \alpha_2, \dots, \alpha_r), 0)$ ($\Sigma \in \mathbb{R}^{m \times n}$), $r = \min(m, n)$, 其中 $\alpha_1 \ge \alpha_2 \ge \dots \ge \alpha_r > 0$, α_i ($i = 1, 2, \dots, r$)为矩阵 *H* 的 奇异值.

然而,由于 3D 体素特征 $F \in R^{B \times C \times D \times H \times W}$ 的特征维 度共包含:批次 (batchsize, *B*)、通道数 (channel, *C*)、 深度 (deep, *D*)、高度 (height, *H*)、宽度 (width, *W*),直 接采用奇异值分解技术会造成大量的资源浪费.请注 意,由于 SVD 可以对高维特征进行降噪、降维处理. 而且处理后的特征具有跟波类似的性质,其能描述数 据中的复杂、多维动态特性,类似于物理波的传播特 性.波在传播过程中会涉及多个参数 (如振幅、频率、

相位), 这与处理多维数据 (具有不同特征的多个参数) 的情况相似. 以此本文将其特征称为波, 后续提出的自注意力机制和编码器均是以特征波为处理对象并因此命名. 波自注意力机制 (WSA) 对特征进行如下的方式处理:首先, 将特征维度划分并压缩成 $H^d \in \mathbb{R}^{H \times W}$ 、 $H^h \in \mathbb{R}^{D \times W}$ 、 $H^w \in \mathbb{R}^{H \times D}$ 的3组维度组合; 然后, 对 3D 特征的3组维度组合分别运用式 (2) 对提取3组特征波 S和3组转换矩阵 P 进行处理; 最后, 将通道数合并至批次维度得到多维波特征 $S^L \in \mathbb{R}^{BC \times L \times q}$ (L=D+H+W, q 代表波特征规模).

$$S = U^{\mathrm{T}} H V$$

$$P = \{F V \Sigma, \Sigma U^{\mathrm{T}} F\}$$
(2)

除了得到3个特征波以外还会得到3组降维矢量 U(V),分别让3组矢量信息经过残差连接的双卷积层 将得到3组升维矢量Û(Ŷ). Û(Ŷ)用于将波注意力SA^L ∈ R^{BC×L×q}重建为体素注意力,具体参考式(3).

$$VA = \sum_{L} \hat{U}(SA^{L})\hat{V}^{\mathrm{T}}/3 \tag{3}$$

自注意力机制 (self-attention, SA) 拥有优秀的捕获 长距离依赖性和适应性的能力,可以很好地实现上下 文联系. ViT 的方法是将特征分块,分块数量作为句子 长度,块内的数据作为单词大小也是 SA 的输入. WSA 则是将波合并张量X ∈ R^{BC×L×q} 作为新的输入, L 和 q 分别表示句子长度和单词大小. 如图 1 所示, WSA 将 多维混合波特征 S 使用全连接映射为 O、K、V, 之后 进行注意力权重计算和 V矩阵聚合,提取波内自注意 力特征SA ∈ R^{BC×L×q}.图1中,绿红蓝色的是特征波,彩 色的是注意力波.



图 1 波自注意力机制 WSA

注意力层输出的特征图维度随输入数据的维度增 大而增大,因此用奇异值分解技术 (SVD) 融合学到的 深度特征,降低特征维度,去除对输入大小的限制.全 连接层网络容易产生过拟合现象,为了有效避免过拟 合和加快训练过程,本文在全连接层之前采用波随机 编码器 (WRE) 随机处理特征减少过拟合,改进深度神 经网络的泛化误差.

1.2 波随机编码器 WRE

掩码自编码器 (mask auto-encoder, MAE) 通过随 机压缩的极少的特征点来还原全局特征, 拥有优秀的 自编码能力, 并可以很好地实现上下文联系和特征压 缩. 基于 MAE 的思想, 使用底层算法实现波随机编码 器 (WRE). 本文的 WRE 模块解决了 MAE 机制不能作 为即插即用的中间层插入现有的深度模型中. 如图 2 所示, WRE 模块可以直接插入到以波的形式存在的任 意特征阶段进行随机编码. 图 2 中, 绿色为宽 W, 红色 为高 H, 蓝色为深 D.



图 2 WRE 嵌入 WSA 的工作示意图

SVD 处理后的 3 组波输出的每个特征张量*X* € ℝ^{BC×L×q}和体素特征保留率γ作为波随机编码器 WRE 的输入,其中 BC 表示波数, L 表示波长度, q 表示波特 征规模. 三维的线性波特征在进行随机编码后特征保 留比更大,且线性波特征不易观察对比编码中随机采 样的效果. 如图 3 所示,本文随机选择了一个样本 CT 采用 WRE 算法进行随机编解码操作.

图 3 中展示了体素保留率 γ∈{0.1³, 0.2³, 0.3³, 0.4³,
 0.5³}下编解码效果,在最低保留率 0.1³的编码特征形 变程度最大,虽然丢失了部分边缘信息,但器官形状和

位置得到了最大程度上的还原,在保留了特征的全局 信息的同时,起到了特征形变的效果.



在将特征向量 $V_1 \in \mathbb{R}^n$, $V_2 \in \mathbb{R}^m$ (0 < m < n)的压缩 和复原的过程中, 需要对特征向量进行如式 (4) 所示的 两个非方阵变换 (列最简矩阵 $A \in \mathbb{R}^{nm}$ 和矩阵 $B \in \mathbb{R}^{nm}$) 以改变特征向量的长度.

$$\begin{cases} V_2 = V_1 A\\ V_1 = V_2 B \end{cases}$$
(4)

要实现这种可复原的压缩,它的基本要求是E = AB,即A、B互为逆矩阵.但是,在矩阵理论体系中,非方阵无逆矩阵.对于列最简的矩阵 $A \in \mathbb{R}^{nm}$,存在矩阵 $B \in \mathbb{R}^{mn}$,使E = AB,那么称B为A的列最简近似逆.因此,对于列最简矩阵的近似逆本文进行了算法 1.

算法 1. WRE 模块的列最简近似逆求解算法

1. 生成掩码向量 $v=\{1\cdots 1\}\cup\{0\cdots 0\}\cup\{1\}\in \mathbb{R}^L$,其中1的数量为 $L\sqrt[3]{\gamma}(0\leqslant\gamma\leqslant 1);$
2. 生成全 0 的编码矩阵A∈ℝ ^{B×L×L ∛} 7和解码矩阵B∈ℝ ^{B×L ∛7×L} ;
3. 遍历波数量 <i>i</i> ∈{0,1,…, <i>B</i> }:
4. 打乱v中除首尾元素的位置;
5. 初始化保留数索引 k←0, 前向遍历向量 v 波长度 j∈{0,1,…,L}:
6. 如果是 1, 则 $k \leftarrow k+1, p \leftarrow l, A_{ijk} \leftarrow 1, A_{ikj} \leftarrow 1;$
7. 否则,如果当前位置是波尾,则B _{ikj} ←1;
8. 否则, <i>B_{ikj}</i> ←1/(<i>l</i> − <i>p</i>);
9. 初始化保留数索引 k←L ∛y, 后向遍历向量v波长度 j∈{L,…,1,0}:
10. 如果是 1,则 $k \leftarrow k+1, p \leftarrow l, A_{ijk} \leftarrow 1, B_{ikj} \leftarrow 1$;
11. 否则,如果当前位置是波尾,则B _{ikj} ←1;
12. 否则, <i>B_{ikj}</i> ←1/(<i>l</i> - <i>p</i>);
13. 初始化保留数索引 <i>k</i> ←L ∛y, 后向遍历向量v波长度 j∈{L,…,1,0}:
14. 如果是 1,则 <i>k</i> ← <i>k</i> −1, <i>q</i> ← <i>l</i> ;
15. 否则,如果当前位置是波首,则B _{ikj} ←1;
16. 否则:
17. $B_{ikj} \leftarrow 1/(q-k);$
18. $sum \leftarrow B_{ikj} + B_{i(k-1)j};$
19. $B_{ikj} \leftarrow B_{ikj} / sum, B_{i(k-1)j} \leftarrow B_{i(k-1)j} / sum.$



图 4 通过 WRE 编码矩阵进行波纹特征随机编码过程

如图 5 所示, 若采用 WRE 得到的随机波纹特征输入到自注意力机制, 计算出 3 组不同空间维度的随机 波注意力*RSA^{Ly}* ∈ ℝ^{BC×L ∛ӯ×q}. 则需用编码矩阵 *A* 分别 对应的解码矩阵 $B \in \mathbb{R}^{B \times L \sqrt[3]{y \times L}}$ 和随机波注意力进行多 维矩阵乘法. 从而得到重建的波注意力 $SA^{Ly} \in \mathbb{R}^{B \times L \times q}$.



图 5 通过 WRE 解码矩阵的随机波纹注意解码过程

1.3 模型框架 WRAUNet

为更方便地嵌入 WSA/WRE 模块并提高模型收敛 速度,本文以 3D U-Net^[17]为基线模型提出了 WRAUNet 模型,全部调整方案如下.

(1) 将所有的双层卷积块替换成单层卷积块+WSA 模块 (WRE 嵌入到 WSA 模块中).

(2)所有在解码器需要与编码器残差连接时将通 道拼接替换成对应位置加法.

(3) 在解码器的每个尺度后连接一个卷积层并插入深监督层 (deep supervision).

(4) 中间卷积层的特征通道全部调整为 32 通道, 固定标签种类为 2 类 (前景,背景).

1.4 损失函数

基于 *DSC* (dice similarity coefficient) 评价指标的 损失函数 *DL* 常应用于医学图像分割中. *DSC* 和 *DL* 的 计算如式 (5) 所示.

$$\begin{cases} DSC(P,G) = \frac{2||P \cap G||_2}{||P||_2 + ||G||_2} \\ DL(P,G) = 1 - DSC(P,G) \end{cases}$$
(5)

其中, ||·||₂为矩阵的 2 范数, 用来计算矩阵每个元素的算数平方和, *G* 是金标准 ground truth, *P* 是模型的输出结果.

WRAUNet 使用了深监督技术会在多个深监督层 产生不同的损失,因此模型使用了如式 (6) 所示的整体 混合损失 *Loss*,其中λ为约束低尺度深监督层的损失影 响因子,它随着 Epoch 次数的迭代逐渐减小.

$$\begin{cases} DS(P,G) = \lambda \cdot \sum_{i} DL(P_{i},G) \\ Loss = DL(P_{0},G) + DS(P,G) \\ = DL(P_{0},G) + \lambda \cdot \sum_{i} DL(P_{i},G) \end{cases}$$
(6)

2 实验

2.1 实验配置

本文在 Linux 系统上用 Torch 实现, 所有实验均

在 Intel(R) Core(TM) i9-10900K CPU@3.70 GHz 处理器、24 GB NVIDIA GeForce RTX 3090 服务器上进行. 实验统一将数据集分辨率设置为 256. 在神经网络训练阶段,使用 Adam 优化算法^[18]更新网络权重,初始学习率设置为 0.0001,衰减系数为 0.2,率衰条件为 2 轮检验集 Loss 不下降,终止条件为 10 轮检验集 Loss 不下降. 对比模型的参数均为默认值.

2.2 模型评价指标

为了衡量模型的效率和精度,实验中共使用了 4 种模型性能评价指标:模型的参数量(parameters, PRM)、计算量(floating point operations, FLOPs)、精 确度(*DSC*)和每片时间(per slice time, PST).其中, PRM 是指网络模型中需要训练的参数总数, FLOPs 是指模 型中浮点运算次数, PST 是指模型每处理一张切片的 平均时长.模型的4项性能评价指标中, *DSC* 的数值越 大说明模型越好,其余3项数值越小说明模型越好.除 4项性能评价指标以外,本文对消融实验的结果,单独 采用了消融建议指数(suggestion factor, *SF*),作为综合 4种性能评价指标的消融系数,系数越高代表消融情况 越好. *SF* 的计算方法如式(7)所示,其中 M 为4项指 标正负相关性系数, *E*^v 为类间期望, *Z*^e 为类内标准化:

$$\begin{cases} \mathbf{M} = \begin{bmatrix} 1 & -0.2 & -0.2 & -0.2 \end{bmatrix} \\ P = 10 \times E^{\nu} (\mathbf{M} \times Z^{c}(data)) \\ SF = e^{P} / \sum e^{P} \end{cases}$$
(7)

2.3 数据库介绍

肾脏数据库来源于 Synapse^[19]多器官分割数据集, 包括 30 个训练样例,每个样例均包含几十组腹部临床 轴位 CT 图,共含 870 张有标签的三维 CT 数据,切片 数量共 109 591.本研究在 30 例中选取每例前 23 组做 训练集,后 6 组作检验集,共得到训练集 CT 数 690、 切片数 86917,检验集 CT 数 180、切片数 22 674,以此 为基础进行了大样本的训练.并用 Dice 指标,对腹部 的左肾和右肾进行评价.

肺实质数据库来源于 COVID-19^[20]肺部病变数据 集,本文在该数据库上进行大目标器官的边缘精准性 评估.该数据集由与确诊的 COVID-19 受试者相关的 20 幅轴向体积 CT 扫描组成,总共包含 3138 张二维肺 部切片,由放射科专家标记、分割和验证.本文从原始 的 4 个医疗单位数据库中,选取每单位前 3 张用作训 练集,后 2 张作为检验集.共得到 12 个训练集 CT 和 8个检验集 CT, 评价指标采用 Dice 系数.

2.4 消融实验

波随机编码器 WRE 依赖于波自注意力机制 WSA,并作为其中间的一个波变换过程.因此,本文将 两个模块的消融组合分为3种:WRE+WSA、WSA、 无.通过这种组合方式来分析模块间的关联关系以及 模块消融情况.如表1所示,两模块全部消融时,模型 的参数量、计算量、平均切片推理时长最短,但分割 的准确率最低、消融建议指数最低;只消融 WSA 模 块时模型的准确率最高;两个模块均不消融时的消融 建议指数最高.这是因为 WSA 模块补充了原始模型 中缺少的全局自注意力信息,WRE 模块拥有对波特 征随机压缩的效果,既降低了参数量、计算量、平均 切片推理时长,也减少了模型的过拟合问题,提高了 模型的泛化能力.

表 1	WSA 和	WRE	的模块消融结果
-----	-------	-----	---------

WSA	WRE	SF (%)	DSC (%)	PRM (M)	FLOPs (G)	PST (ms)		
—	_	4.00	85.13	0.52	76.57	0.64		
\checkmark	_	24.15	89.2	8.47	77.46	9.39		
\checkmark	\checkmark	71.85	89.2	8.47	77.08	5.32		

波注意力机制 WSA 存在代表波特征规模的参数 q, 波随机编码器 WRE 存在代表体素保留率的参数 y, 两个参数的取值会直接影响到模型性能. 如表 2 所示, 文中对两个参数的取值分别进行了参数规模的消融实 验. 参数 q 越大模型的分割准确率越高, 但模型参数 量、计算和切片的平均推理时间呈现指数式增长; 参 数 y 越大模型的分割准确率越高, 但计算和切片的平 均推理时间呈现对数式增长. 两个参数的消融建议指 数分别在 q =3 和 y=0.2 时取得最大值, 表示在当前数 据集和当前模型下, 两个参数分别设置为 3、0.2 时模 型性能最好.

2.5 对比实验

本研究在肾脏数据库上对比了经典方法 3D U-Net、 最新的多阶段模型 nnU-Net、MFU-Net 以及最新的注 意力分割方法 Swin UNETR、nnFormer. 该研究从这 些方法的原论文中收集了在 Synapse 上的左肾脏 DSC 得分. 对于未公开源代码的方法,按照论文描述在 Torch 中进行模型结构的复现. 最后采用 ptflop 工具对全部 模型进行参数量、计算量以及单张切片的平均推理时 长的统计. 肾脏分割实验的结果如表 3 所示, WRAUNet 的 DSC 得分达到 89.2%, 肾脏分割精度更高. 相比于新 的多阶段方法,本文方法在精准度,参数量,推理速度 上均取得 SOTA (state-of-the-art,最先进的)效果,计算 量略高于 nnU-Net;相比于新的注意力方法,本文方法 在 4 种性能指标上均取得 SOTA 效果.这是因为,本文 采用了特征波化的方式来插入自注意力机制,即解决 了非注意力方法中的全局信息不充分的问题,也解决 了现有注意力方法参数量和计算量大的问题.此外,波 随机编码器的引入,使波注意力模块进一步降低了计 算量、提高了推理速度,并且使模型拥有了更高的鲁 棒性和泛化能力.

Parametric	-	Wave c	haracteristic	c scale q			Vo	oxel feature	retention rat	eγ	
metrics	1	2	3	4	5	0.05	0.1	0.2	0.4	0.6	0.8
DSC (%)	88.93	89.15	89.2	89.2	89.21	89.11	89.18	89.2	89.2	89.2	89.2
PRM (M)	2.69	4.86	8.47	13.52	20.01	8.47	8.47	8.47	8.47	8.47	8.47
FLOPs (G)	76.63	76.8	77.08	77.48	77.99	76.89	76.97	77.08	77.22	77.31	77.39
PST (ms)	3.34	3.81	5.32	25.75	93.6	3.98	4.72	5.32	6.75	7.6	7.82
SF (%)	0.23	28.67	51.3	16.95	2.85	0.13	16.61	45.48	18.79	10.9	8.09

表 2 模块相关参数消融结果

表 3 肾脏分割对比实验结果						
Model	DSC (%)	PRM (M)	FLOPs (G)	PST (ms)		
3D U-Net ^[17]	77.77	2126.77	90.3	19.53		
nnU-Net ^[21]	85.57	1079.30	62.4	9.92		
MFU-Net ^[22]	84.63	134.12	34.22	14.32		
nnFormer ^[23]	87.64	702.78	300.1	6.48		
Swin UNETR ^[24]	83.28	61.98	394.84	15.67		
W/P ALINAT	80.2	8 17	77.08	5 32		

本文方法分割出的肾脏器官精度更高. 如图 6 所 示, 绿色边界为左肾分割的金标准 (ground truth, GT),

红色边界为本文方法和对比方法的推理结果 (predict outcomes, PO). 其中, 第1列为分割金标准, 第2列为 本文方法, 第3-6列为对比方法. 为了更严谨的展示其 分割效果, 本文从3组扫描断层 (包括横断面 Transverse section、冠状面 Coronal section 和矢状面 Median sagittal section) 中选出器官占比较高的切片用于展示. 结果表明, 本文方法在器官的主体形状以及器官边界 均有很好的分割效果, 极大程度上描述器官形状信息, 并给出了更加清晰的分割边界.



图 6 3 种扫描层下的分割结果

为了更科学地评价本文方法的有效性,本模型与 国内外的 5 种最新的肺实质分割方法进行了对比.与 肝脏中的对比实验相同,对不含代码的方法进行模型 复现,和含代码的方法共同统计前 3 组评价指标 (PRM、 FLOPs 和 PST).实验结果如表 4 所示,本文方法的 DSC 精度仅比 Mahmoudi 等人^[25]的两阶段方法低 0.1 个百分点.同时,本文方法的参数量、计算量和平均切 片推理速度均优于其他已有方法.这是因为,本文方法 的 WSA 模块和 WRE 模块,提取到了更多局部细节和 整体差异,即提高了分割结果的定位精度和形状细节.

为了更直观的分析本文方法的肺实质分割实验, 如图 7 所示,绿色边界为左肾分割的金标准 (GT),红色

边界为本文方法和对比方法的推理结果 (PO). 其中, 第1列为分割金标准, 第2列为本文方法, 第3-7列为 对比方法. 结果显示, WRAUNet 可以从原始 CT 中提 取出粗略的肺实质区域, 保留肺实质及肺实质边缘的 局部背景, 屏蔽长距离干扰数据. 并且可以对提取的肺 实质及其周边区域进行边缘优化, 精准剔除非肝脏区 域的假阳性体素.

表 4 双肺分割对比实验结果						
Model	PRM (M)	FLOPs (G)	DSC (%)	PST (ms)		
文献[26]	1079.30	62.41	97.0	9.92		
文献[27]	2126.77	90.38	95.6	19.53		
文献[28]	43.21	7.85	96.1	5.35		
文献[29]	33.17	6.21	96.0	4.62		
文献[25]	4253.54	180.76	98.0	39.06		
WRAUNet	20.24	5.47	97.1	3.30		



图 7 双肺分割结果可视化 (绿: GT; 红: PO)

3 总结

本文提出了波自注意机制 (WSA), 用于解决在提 升捕捉全局特征的能力的同时而又不会消耗大量资源 的问题. 另外根据 CT 图像中, 器官的位置和形状具有 相对的依赖性的特点, 本文设计了一种波随机编码器 (WRE)来捕捉器官的相对依赖性. 在肾脏和肺的数据 集上进行广泛实验, 本文方法在精准度, 参数量, 推理 速度上均取得 SOTA 效果. 由于 WSA 和 WRE 不仅可 以降低参与计算的数据规模, 不会消耗大量资源而且 能提升捕捉全局特征的能力, 提高分割精度. 不仅可以 应用在医学分割领域, 也可以用于数据规模大, 参数复 杂的问题的相关领域. 未来的研究可以探索波注意机 制和波自动编码器在脑部、心脏等复杂结构中的应用, 进一步提升疾病检测和诊断的精度.

参考文献

1 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234–241. [doi: 10.1007/978-3-319-24574-4 28]

- 2 Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. Proceedings of the 4th International Conference on 3D Vision (3DV). Stanford: IEEE, 2016. 565–571. [doi: 10.1109/3DV.2016.79]
- 3 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations. ICLR, 2021.
- 4 Cai YT, Wang Y. MA-Unet: An improved version of Unet based on multi-scale and attention mechanism for medical image segmentation. Proceedings of the 3rd International Conference on Electronics and Communication; Network and Computer Technology. Harbin: SPIE, 2022. 121670X. [doi: 10.1117/12.2628519]
- 5 Hatamizadeh A, Tang YC, Nath V, *et al.* UNETR: Transformers for 3D medical image segmentation. Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2022. 1748–1758.
- 6 Cao H, Wang YY, Chen J, *et al.* Swin-Unet: Unet-like pure Transformer for medical image segmentation. Proceedings of the 2023 European Conference on Computer Vision. Tel Aviv: Springer, 2023. 205–218. [doi: 10.1007/978-3-031-25066-8_9]
- 7 Butanovs E, Zolotarjovs A, Kuzmin A, et al. Nanoscale X-

ray detectors based on individual CdS, SnO₂ and ZnO nanowires. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2021, 1014: 165736. [doi: 10. 1016/j.nima.2021.165736]

- 8 Peken T, Adiga S, Tandon R, *et al.* Deep learning for SVD and hybrid beamforming. IEEE Transactions on Wireless Communications, 2020, 19(10): 6621–6642. [doi: 10.1109/ TWC.2020.3004386]
- 9 Bendjador H, Deffieux T, Tanter M, *et al.* The SVD beamformer: Physical principles and application to ultrafast adaptive ultrasound. IEEE Transactions on Medical Imaging, 2020, 39(10): 3100–3112. [doi: 10.1109/TMI.2020.2986830]
- 10 Abdelwahab KM, Abd El-Atty SM, El-Shafai W, et al. Efficient SVD-based audio watermarking technique in FRT domain. Multimedia Tools and Applications, 2020, 79(9): 5617–5648. [doi: 10.1007/s11042-019-08023-z]
- 11 Diwakar M, Kumar P, Singh P, et al. An efficient reversible data hiding using SVD over a novel weighted iterative anisotropic total variation based denoised medical images. Biomedical Signal Processing and Control, 2023, 82: 104563. [doi: 10.1016/j.bspc.2022.104563]
- 12 Bhatti A, Ishii T, Kanno N, *et al.* Region-based SVD processing of high-frequency ultrafast ultrasound to visualize cutaneous vascular networks. Ultrasonics, 2023, 129: 106907. [doi: 10.1016/j.ultras.2022.106907]
- 13 Shi MW, Zhang F, Wang SW, *et al.* Detail preserving image denoising with patch-based structure similarity via sparse representation and SVD. Computer Vision and Image Understanding, 2021, 206: 103173. [doi: 10.1016/j.cviu. 2021.103173]
- 14 Ding XH, Zhang XY, Han JG, *et al.* Scaling up your kernels to 31×31: Revisiting large kernel design in CNNs. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 11953–11965. [doi: 10.1109/CVPR52688.2022.01166]
- 15 He KM, Zhang XY, Ren SQ, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1026–1034. [doi: 10.1109/ICCV.2015.123]
- 16 He KM, Chen XL, Xie SN, *et al.* Masked autoencoders are scalable vision learners. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 15979–15988. [doi: 10.1109/ CVPR52688.2022.01553]
- 17 Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. Proceedings of the 19th International Conference on Medical Image Computing and Computer-assisted

Intervention. Athens: Springer, 2016. 424–432. [doi: 10. 1007/978-3-319-46723-8_49]

- 18 Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv:1412.6980, 2014.
- 19 Clark K, Vendt B, Smith K, *et al.* The cancer imaging archive (TCIA): Maintaining and operating a public information repository. Journal of Digital Imaging, 2013, 26(6): 1045–1057. [doi: 10.1007/s10278-013-9622-7]
- 20 Yang XY, He XH, Zhao JY, *et al*. COVID-CT-dataset: A CT scan dataset about COVID-19. arXiv:2003.13865, 2020.
- 21 Isensee F, Jaeger PF, Kohl SAA, et al. nnU-Net: A selfconfiguring method for deep learning-based biomedical image segmentation. Nature Methods, 2021, 18(2): 203–211. [doi: 10.1038/s41592-020-01008-z]
- 22 Yang AS, Xu LL, Qin N, et al. MFU-Net: A deep multimodal fusion network for breast cancer segmentation with dual-layer spectral detector CT. Applied Intelligence, 2024, 54(5): 3808–3824. [doi: 10.1007/s10489-023-05090-6]
- Zhou HY, Guo JS, Zhang YH, *et al.* nnFormer: Volumetric medical image segmentation via a 3D Transformer. IEEE Transactions on Image Processing, 2023, 32: 4036–4045. [doi: 10.1109/TIP.2023.3293771]
- 24 Hatamizadeh A, Nath V, Tang YC, *et al.* Swin UNETR: Swin Transformers for semantic segmentation of brain tumors in MRI images. Proceedings of the 7th International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer, 2022. 272–284. [doi: 10.1007/978-3-031-08999-2_22]
- 25 Mahmoudi R, Benameur N, Mabrouk R, et al. A deep learning-based diagnosis system for COVID-19 detection and pneumonia screening using CT imaging. Applied Sciences, 2022, 12(10): 4825. [doi: 10.3390/app12104825]
- 26 Ma J, Wang Y, An X, *et al.* Toward data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation. Medical Physics, 2021, 48(3): 1197–1210. [doi: 10.1002/mp.14676]
- 27 Müller D, Soto-Rey I, Kramer F. Robust chest CT image segmentation of COVID-19 lung infection based on limited data. Informatics in Medicine Unlocked, 2021, 25: 100681. [doi: 10.1016/j.imu.2021.100681]
- 28 Alirr OI. Automatic deep learning system for COVID-19 infection quantification in chest CT. Multimedia Tools and Applications, 2022, 81(1): 527–541. [doi: 10.1007/s11042-021-11299-9]
- 29 Punn NS, Agarwal S. CHS-Net: A deep learning approach for hierarchical segmentation of COVID-19 via CT images. Neural Processing Letters, 2022, 54(5): 3771–3792. [doi: 10. 1007/s11063-022-10785-x]

(校对责编:张重毅)