Wave self-attention mechanism for three-dimensional features

1st Hao Liu
College of Computer Science and Technology
Qingdao University of Science and Technology
Qingdao, China
hliu1997ac@163.com

2st Yuanzhi Cheng*
College of Computer Science and Technology
Qingdao University of Science and Technology
Qingdao, China
yzcheng2007@163.com

3st Hui Li
College of Computer Science and Technology
Qingdao University of Science and Technology
Qingdao, China
lipeilin1984xyz@163.com

***Abstract*—Accurately distinguishing organs from CT scans is a demanding yet indispensable task due to its complexity. A frequent quandary in various preceding studies is that we are required to utilize Visual Transformer (ViT) technology and expend significant resources to attain comprehensive self-attentive insights into characteristics. In this paper, we propose a new paradigm for integrating self-attention and feature dimensionality reduction into CT image segmentation, aiming at reduce resource consumption by compression of feature which is used for self-attention extraction. Specifically, we propose a novel module to extract condensed attentional features, wherein wave attention is used to precisely identify the focal areas of interest from the compressed features. In those identified regions, the self-attention with head fusion records the interrelatedness of spatial features over extended distances. Furthermore, an Parallel wave dimensionality reduction strategy is proposed to compression voxel characteristics and decompression wave characteristics by the generated approximation reciprocal matrix in a asymmetrical fashion. The experimental results of kidney segmentation on the Synapse benchmark dataset manifest that our approach achieves an absolute enhancement of 1.56% and 32.83% accuracy (DSC) and efficiency (PST) respectively, when compared to the existing state-of-the-art techniques.**

***Index Terms*—voxel processing, singular value decomposition, parallel wave dimension reduction, wave self-attention.**

I. Introduction

Organ segmentation in three-dimensional CT images is a pivotal undertaking in clinical medicine, essential for diagnosing and treating human organs. However, manual delineation of organs is laborious and existing automatic segmentation methods are slow and unreliable, making fast and effective organ segmentation methods an important focus of medical image research. 3D CT segmentation tasks pose more challenges than 2D images, including the higher dimensionality of global information and its extraction difficulty, as well as larger sample scales leading to increased consumption of space and computational resources. In organ segmentation tasks, the most classical model is U-Net [1], which uses skip links between multiscale encoders and multiscale decoders to compensate for the loss of features between multiscale features, but its structure is too simple, local information is diluted by multi-scale features, and under-segmentation and over-segmentation are prone to occur. Some methods [2]–[4] use more complex feature extraction blocks and skip links to improve organ segmentation accuracy, some other methods [5]–[10] optimize the segmentation results by cascading multiple UNET models. However, these methods only use convolution as the basic feature extraction layer, which leads to their insufficient use of global features. The introduction of the visual transformer (ViT) [11] has significantly enhanced the accuracy of organ segmentation by providing access to richer global information. Among the many methods of combining visual transformers with Unet, the typical methods are Ma-UNET [12] based on channel attention, UNETR [13] based on ViT encoder, and Swin-unet [14] based on window attention. In addition, some researchers [15]–[22] also used ViT technology to optimize Unet model from different angles. In fact, ViT essentially uses feature partitioning to process and transform data, and cannot compress features while reducing dimensions. As a result, the parameters and computation of the Self-Attention Mechanism (SA) have increased dramatically, and a large amount of video memory needs to be occupied. Singular value decomposition (SVD) [23] is a widespread matrix compression approach that can be leveraged for linear feature extraction [24]–[26], image dimensionality reduction and noise cancellation [27]–[29].However, SVD is only applicable to 2D features. If you use batch processing technology to apply it to 3D features, it will not only consume huge computing resources, but also lose the dependency between batch processing dimensions and other dimensions.

In summary, ViT can capture richer global information, thus improving segmentation accuracy; SVD can reduce the feature dimension easily and provide feature compression and denoising. However, when applied to 3D features, both of these technologies have a number of application problems. Based on these problems, we have proposed the following contributions.:

- By leveraging the singular value decomposition technique, we introduce a novel Parallel Wave Dimension

* Corresponding author: Yuanzhi Cheng (yzcheng2007@163.com)

Reduction (PWDR) methodology for 3D features, which is more computationally efficient for feature dimension reduction, thereby overcoming the limitation of integrating 3D modules with 1D modules.

- We introduce a novel Wave Self-Attention (WSA) mechanism based on the self-attention framework, which optimizes resource utilization to extract optimal global features. Our experimentation on kidney organs shows that this method improves accuracy and efficiency in segmenting small target organs.

II. Method

A. Parallel wave dimensionality reduction (PWDR)

Kaleman (1996) [23] proposed Singular Value Decomposition (SVD) for any matrix $H \in \mathbb{R}^{mn}$, all have orthogonal $U = (u_1, u_2, \cdots, u_m) \in \mathbb{R}^{mn}$ and orthogonal matrix $V = (v_1, v_2, \cdots, v_n) \in \mathbb{R}^{nn}$, so that the matrix H satisfies (1).

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

Where, $\Sigma = \mathrm{diag}(\alpha_1, \alpha_2, \cdots, \alpha_r) \in \mathbb{R}^{mn}, r = \min(m, n)$, where $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r \geq 0$ is the singular value of matrix H.

To reduce the waste of resources caused by the large data size of the 3D voxel features $F \in \mathbb{R}^{BCDHW}$ (consisting of batch size (B), number of channels (C), depth (D), height (H), and width (W)), we first use the global average function to compress the features into $H_d \in \mathbb{R}^{HW}, H_h \in \mathbb{R}^{DW}, H_w \in \mathbb{R}^{DH}$, a total of three groups of spatial characteristic matrices. Then, the singular value decomposition technique is used to obtain three sets of decomposition preserving the first q singular values in descending order.However, the global mean function results in the loss of the dependency between the batch, channel, and three spatial dimensions. To tackle this issue, we propose a Wave Self-Attention (WSA) mechanism that integrates the original features. It extracts three sets of parallel feature waves S and a feature recovery matrix P over the full dimension via (2), thus complementing the dependencies between multiple dimensions. Furthermore, a Conv layer with a kernel size of 3 is used to process the recovery matrix P, transforming its dependency from the feature space to a self-attention space.

$$\begin{aligned} S &= U^{\mathrm{T}} H V \\ \hat{U} &= \mathrm{Conv}(FVS) \\ \hat{V} &= \mathrm{Conv}(SU^{\mathrm{T}} F) \\ P &= \{\hat{U}\ , \hat{V}\} \end{aligned} \tag{2}$$

As illustrated in Figure 1, the proposed PWRD mechanism is used to process voxel features, resulting in three sets of parallel wave features $S^l \in \mathbb{R}^{BClq}, (l \in \{D, H, W\})$, and the combination of these three wave characteristics yields $S^L \in \mathbb{R}^{BCLq}, (L = D + H + W)$. Here, q denotes the number of retained singular values, i.e., the compression ratio of the features.

The proposed PWAD mechanism processes 3D features by extracting three sets of singular values Σ and orthogonal matrices $U(V)$. Specifically, Σ belongs to the vector spaces of D, H, and W, respectively, and encodes the 3D spatial information of the voxel features. To further enhance the self-attention features, we apply a Conv layer with a kernel size of 3 to convert the singular values Σ into three new orthogonal matrices $\hat{U}(\hat{V})$. Additionally, (3) is used to transform the wave self-attention feature $S^l_{attn} \in \mathbb{R}^{BClq}$ into a voxel feature V_{attn}.

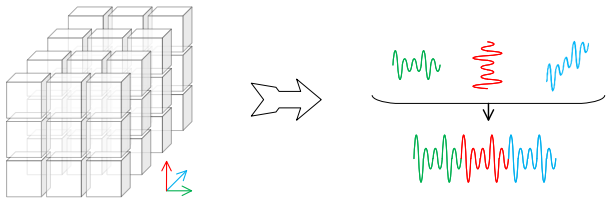

Fig. 1. Parallel wave dimensionality reduction technology for voxel features.

$$V_{attn} = \textstyle\sum_l \hat{U}(S^l_{attn})\hat{V}^T/3 \tag{3}$$

B. Wave self-attention mechanism (WSA)

To exploit the excellent ability of Self-Attention (SA) in capturing long-range dependencies and adaptations, and further achieve contextual connections, we integrate it with wave features and propose the Wave Self-Attention mechanism (WSA). Specifically, $X \in \mathbb{R}^{BCLq}$ is employed as the input characteristic of WSA, while $Y \in \mathbb{R}^{BCLq}$ represents the output characteristic of the WSA. Here, L and Cq respectively denote the sentence length and word size, so that $x_{ij} \in \mathbb{R}^B\ (0 \leq i < L,\ 0 \leq j < Cq)$ represents the element of the input tensor and $y_{ij} \in \mathbb{R}^B\ (0 \leq i < L,\ 0 \leq j < Cq)$ represents the element of the output tensor. The following equations provide further details of the wave attention calculation:

$$\begin{aligned} q_{ij} &= W^q x_{ij} \\ k_{ij} &= W^k x_{ij} \\ v_{ij} &= W^v x_{ij} \end{aligned} \tag{4}$$

$$\begin{aligned} s_{ij} &= \mathrm{Attnetion}(q_{ij}, k_{ab}) \\ &= \mathrm{softmax}\left(\frac{(W^q x_{ij})^{\mathrm{T}}(W^k x_{ab})}{\sqrt{d}}\right) \end{aligned} \tag{5}$$

$$\begin{aligned} y_{ij} &= \sum_{a,b \in N_k(i,j)} \mathrm{Attnetion}(q_{ij}, k_{ab}) v_{ab} \\ &= \sum_{a,b \in N_k(i,j)} \mathrm{softmax}\left(\frac{(W^q x_{ij})^{\mathrm{T}}(W^k x_{ab})}{\sqrt{d}}\right) v_{ab} \end{aligned} \tag{6}$$

Where, $\mathrm{W_q}$, $\mathrm{W_k}$, $\mathrm{W_v}$ represent the weight matrices, $N_k(i, j)$ denotes the local region with the center of (i, j) and the range of K, and s_{ij} stands for the self-attention weight in region $N_k(i, j)$.

As illustrated in Figure 2, the self-attention mechanism employs a fully connected layer to project the multi-dimensional mixed wave feature S^L onto Q, K, and V matrices, thus allowing for the computation of attention weights and matrix aggregation to generate the wave self-attention feature $S^L_{attn} \in \mathbb{R}^{BCLq}$ in an elegant and precise manner.

WSA has the ability to automatically adapt to channel and spatial dimensions, thus enabling the automatic search for

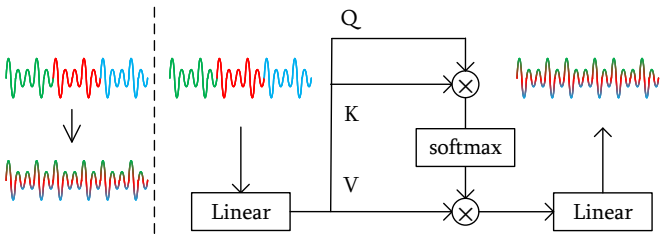

Fig. 2. wave self-attention mechanism.

global self-attention across the full dimensions. As presented in Figure 3, the attention representation at four different scales is depicted. The channel dimension and spatial dimension are independent variable axes, while the original input characteristic wave is the dependent variable axis. The wave self-attention forms an attention projection at the bottom of the characteristic wave, and the wave self-attention color bar is given on the right. As depicted in (a), WSA mainly exhibits channel self-focus in the low-scale feature layer, due to the large amount of global information stored in the channel dimension of the low-scale feature. Conversely, WSA primarily demonstrates spatial attention in the high-scale feature layer, as depicted in (d), as the channel dimension of the high-scale feature contains insufficient global information.

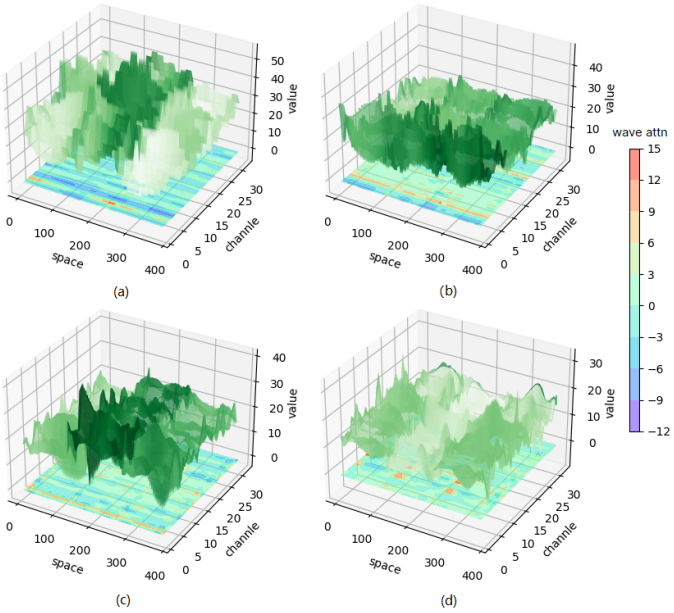

Fig. 3. Wave attention factor at multiple scales ($a \in \mathbb{R}^{8\cdot8\cdot8}, b \in \mathbb{R}^{16\cdot16\cdot16}, c \in \mathbb{R}^{32\cdot32\cdot32}, d \in \mathbb{R}^{64\cdot64\cdot64}$).

C. Model framework

To further improve the performance of WSA modules, we have applied Unet3d as the baseline model and made the following adjustments to better embed the WSA modules and improve model convergence:

- Replaced all Conv2 with ConvWSA modules (as illustrated in Figure 4);
- Replaced channel splicing at the residual connection of the decoder and encoder with matrix addition;
- Inserted a Deep Supervision (Sup) Layer after each scale of the decoder;
- Adjusted the number of feature channels of all M-servers to 32, which is only applicable to binary tasks.

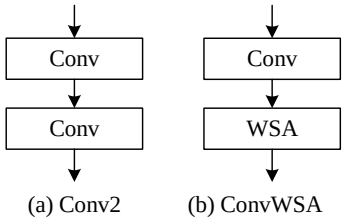

Fig. 4. Conv2 structure "a" and ConvWSA structure "b".

To depict the modifications more clearly, we show the U-Net model structure and WaveUNet model structure before and after the modifications in Figure reffig4. The monitoring layer inserted at the current position is denoted by "Sup", which is used to calculate the model loss and back propagation gradient.

D. loss function

Loss function based on DSC evaluation index $\mathcal{L}_{dsc}$

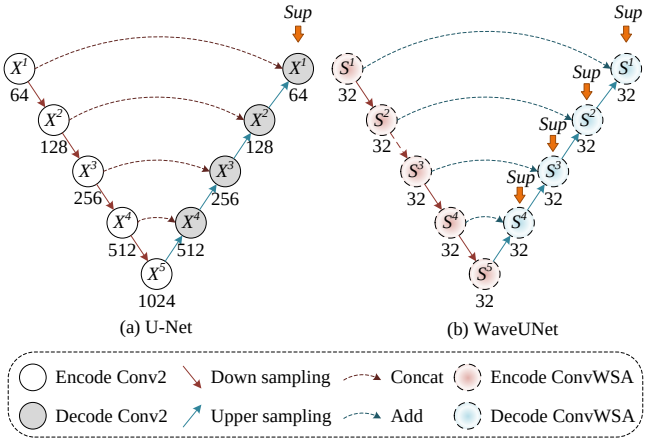

Fig. 5. Model frame structure diagram.

$$\begin{aligned} \mathrm{DSC}(P,G) &= \frac{2\|P\cap G\|_2}{\|P\|_2+\|G\|_2} \\ \mathcal{L}_{dsc}(P,G) &= 1-\mathrm{DSC}(P,G) \end{aligned} \tag{7}$$

Where $\|\ \|_2$ represents the 2-norm of the matrix, which is apply to compute the arithmetic sum of squares of all elements of the matrix. G is the gold standard Ground True, and P is the output result of the model. Since we employ the depth monitoring technology, The number of predictions increases

with the number of deep monitoring layers. Therefore, we use the mixed loss $\mathcal{L}$ as illustrated in (8).

$$\begin{aligned} \mathcal{L}_{dsc}^{\lambda}(P,G) &= \lambda \cdot \textstyle\sum_i \mathcal{L}_{dsc}(P_i,G) \\ \mathcal{L} &= \mathcal{L}_{dsc}(P_0,G) + \mathcal{L}_{dsc}^{\lambda}(P,G) \\ &= \mathcal{L}_{dsc}^{\lambda}(P_0,G) + \lambda \cdot \sum\nolimits_i \mathcal{L}_{dsc}^{\lambda}(P_i,G) \end{aligned} \tag{8}$$

Where λ represents the loss impact factor that restricts the low-scale deep monitoring layer and diminishes over time with Epoch iterations.

III. Experiment

A. Experimental configuration

For our model, we employed PyTorch on Linux and carried out all experiments on a computer with an Intel (R) Core (TM) i9-10900K CPU@3.70GHz processor, running on a server with 24GB of NVIDIA GeForce RTX 3090. During the neural network's training phase, the network weights were updated using Adam optimization, with an initial learning rate of 1e-4 and an attenuation coefficient of 0.2. The rate decay condition was determined by the lack of decrease in the loss of the 2-round test set, while the end condition was the lack of decrease in the loss of the 10-round test set. The parameters of the comparison model were the default values.

B. Model evaluation index

In order to comprehensively assess the efficiency and accuracy of our model, four performance evaluation metrics were employed in the experiments: Parameters (Para), Floating Point Operations (FlOPs), Dice Similarity Coefficient (DSC) and Per Slice Time (PST).PRM indicates the number of parameters to be trained in the network model, FLOPs denote the number of floating-point operations produced by processing $\mathbb{R}^{128\cdot128\cdot128}$ scale data, DSC measures the similarity between the predicted and true labels, and PST denotes the average length of time a model takes to process each slice. Among the four performance evaluation metrics, DSC is positively correlated with model performance, while the remaining metrics are inversely correlated. Apart from the aforementioned performance evaluation indexes, we also use the recommendation factor (SF) as a composite of the four performance evaluation indexes. The higher the SF, the better the comprehensive performance effect of the model. SF was calculated using the formula in (9).

$$\begin{aligned} \mathbf{M} &= \begin{bmatrix} 1 & -0.2 & -0.2 & -0.2 \end{bmatrix} \\ \mathrm{P} &= 10 \times \mathrm{E}^{v}(\mathbf{M} \times \mathrm{Z}^{c}(data)) \\ \mathrm{SF} &= e^{\mathrm{P}} / \textstyle\sum e^{\mathrm{P}} \end{aligned} \tag{9}$$

Where M is the positive and negative correlation coefficient of the four indicators, E^{v} is the inter-class expectation, and Z^{c} is the intra-class standardization.

C. Dataset

Our study utilized kidney segmentation data from the Synapse organ segmentation database, comprising of 30 cases with 29 clinical axial CT images of the abdomen, 870 3D CT with markers, and 109591 2D slices with markers. We randomly split 690 and 180 CT images for training and test sets, respectively.

D. Comparative experiment

Through our kidney segmentation experiment, we compared our method with typical methods from recent years and the results are displayed in TABLE I. Our method achieves 1.56% and 32.83% higher accuracy (DSC) and efficiency (PST) than the state-of-the-art. Moreover, compared to the multi-stage method, our method achieved SOTA in terms of accuracy, number of parameters, inference speed, and computational effort, except for a slightly higher computational effort than nnunt. Compared to other attention methods, our method achieved SOTA results on all four performance indicators, demonstrating that our method addresses the issue of insufficient global information in non-attention methods as well as the issue of extensive parameters and computations in existing attention methods. The high percentage of organs segmented in three scanning types (transverse section, coronal section, and sagittal section) is illustrated in Fig. 6. The first column is the ground truth, the second is our prediction, and the third to sixth are the comparison methods. We can see that the introduction of the WSA module results in better segmentation of the kidney-liver junction and organ corners than the baseline method and is no worse than other block-attention models. This suggests that WSA can capture better and sufficient global attention information to optimise the segmentation of low-contrast boundaries.

TABLE I
COMPARATIVE EXPERIMENTAL RESULTS OF KIDNEY SEGMENTATION

Head	***DSC***	***Para(M)***	***FlOPs(G)***	***PST(ms)***
Unet3d [1]	77.77	2126.77	722.4	18.53
Vnet [2]	78.9	45.61	755.32	24.21
MFUNet [5]	84.63	134.12	1095.04	34.32
nnunet [6]	85.57	1079.30	**62.4**	7.92
nnformer [9]	87.64	702.78	300.1	14.48
TransUNet [18]	81.87	324.02	2159.36	49.29
HiFormer [19]	82.5	25.51	1265.92	39.12
MT-UNet [10]	81.47	79.08	1432	39.29
Swin-unet [14]	83.28	196.06	502.4	19.67
LeVit-UNet-384s [21]	84.61	52.17	817.6	27.81
WaveUNet(our)	**89.2**	**8.47**	77.08	**5.32**

E. Ablation experiment

As shown in TABLE II, when the WSA module is abbreviated, there is a considerable decline in the accuracy of segmentation and the abbreviated recommendation index, though there is a slight reduction in the parameters, computation and average slicing inference time of the model. This indicates that the incorporation of the WSA module boosts the lacking global self-attentive information in the initial model, while curtailing the overfitting issue of the model and augmenting the generalization capacity of the model.

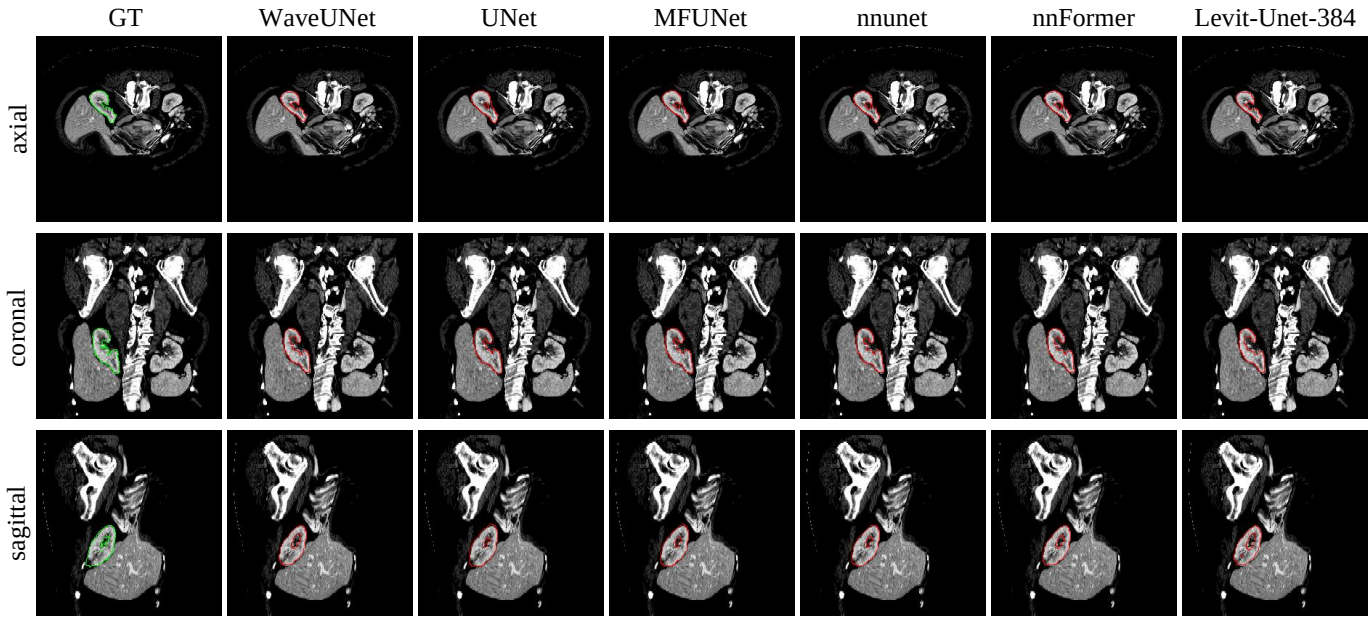

Fig. 6. Comparative experiment

TABLE II
TABLE II. MODULE ABLATION RESULTS OF WSA AND WRE

Head	*SF(%)*	*DSC*	*Para(M)*	*FlOPs(G)*	*PST(ms)*
—	11.92	85.13	0.52	76.57	0.64
√	**88.08**	89.2	8.47	77.46	9.39

F. Parameter Selection

The Wave Self-Attention Mechanism (WSA) utilizes a parameter q to characterize its compression ratio, which is proportional to the size of the compressed wave.Increasing q leads to improved segmentation accuracy, and calculation and slicing time have grown exponentially along with a growth in model parameters, as shown in TABLE III. The ablation index of the parameter reaches its peak when set to $q = 3$, showing that the model performs optimally when this value is set.

TABLE III
TABLE II. MODULE ABLATION RESULTS OF WSA AND WRE

Parametric metrics	**Table Column Head**				
	1	***2***	***3***	***4***	***5***
DSC(%)	88.93	89.15	89.2	89.2	**89.21**
Para(M)	**2.69**	4.86	8.47	13.52	20.01
FlOPs(G)	**76.63**	76.8	77.08	77.48	77.99
PST(ms)	**3.34**	3.81	5.32	25.75	93.6
SF(%)	0.23	28.67	**51.3**	16.95	2.85

IV. Conclusion

This paper presents a novel visual attention mechanism, Wave Self-Attention (WSA), which integrates the Parallel Wave Dimensionality Reduction (PWAD) technique to effectively capture global self-attentive information with fewer resources. This module can be seamlessly integrated with any convolutional model, enabling it to automatically equalize global self-attention in both the channel and spatial dimensions. Additionally, our experiments on kidney segmentation demonstrate that the WSA achieves state-of-the-art performance. We anticipate that more researchers will participate in the WSA research, continue to optimize its structure and apply it to additional 3D vision processing tasks.

References

[1] O. Ronneberger, P. Fischer and T Brox, "U-net: Convolutional networks for biomedical image segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015: 234-241.

[2] F. Milletari, N. Navab and S. A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016: 565-571.

[3] Y. Xu, S. Hou and X. Wang, "A Medical Image Segmentation Method Based on Improved UNet 3+ Network," Diagnostics, 2023, 13(3): 576.

[4] H. Huang, L. Lin and R. Tong, "Unet 3+: A full-scale connected unet for medical image segmentation," ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020: 1055-1059.

[5] S. W. Zamir, A. Arora and S. Khan, "Multi-stage progressive image restoration," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 14821-14831.

[6] F. Isensee, P. F. Jäger and S. A. A. Kohl, "Automated design of deep learning methods for biomedical image segmentation," arXiv preprint arXiv:1904.08128, 2019.

[7] X. M. Li, H. Chen and X. J. Qi, "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes," IEEE transactions on medical imaging, 2018, 37(12): 2663-2674.

[8] F. Özcan, O. N. Uçan and S. Karaçam, "Fully Automatic Liver and Tumor Segmentation from CT Image Using an AIM-Unet," Bioengineering, 2023, 10(2): 215.

[9] H. Y. Zhou,J. Guo and Y. Zhang, "nnformer: Interleaved transformer for volumetric segmentation," arXiv preprint arXiv:2109.03201, 2021.

[10] H. Wang, S. Xie and L. Lin, "Mixed transformer u-net for medical image segmentation," ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022: 2390-2394.
[11] A. Dosovitskiy, L. Beyer and A. Kolesnikov, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.
[12] Y. Cai and Y. Wang, "Ma-unet: An improved version of unet based on multi-scale and attention mechanism for medical image segmentation," Third International Conference on Electronics and Communication; Network and Computer Technology (ECNCT 2021). SPIE, 2022, 12167: 205-211.
[13] A. Hatamizadeh, Y. Tang and V. Nath, " Unetr: Transformers for 3d medical image segmentation," Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022: 574-584.
[14] H. Cao, Y. Wang and J. Chen, "Swin-unet: Unet-like pure transformer for medical image segmentation," Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III. Cham: Springer Nature Switzerland, 2023: 205-218.
[15] H. LI, J. ZOU and X. TAN, "Multi-attention fusion network for medical image segmentation," Journal of Computer Applications, 2022 , 42(9): 1-10.
[16] S. Zhang,Z. PENG and H. Li, "SAU-Net: Medical image segmentation method based on u-net and self-attention," Acta Electronica Sinica, 2022, 50(10): 1-10.
[17] Y. SUN, F. BI and Y. GAO, "A multi-attention unet for semantic segmentation in remote sensing images," Symmetry, 2022, 14(5): 906-906.
[18] J. Chen, Y. Lu and Q. Yu, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv preprint arXiv:2102.04306, 2021.
[19] M. Heidari, A. Kazerouni and M. Soltany, "Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation," arXiv preprint arXiv:2207.08518, 2022.
[20] Y. Li , W. Cai and Y. Gao, "More than encoder: Introducing transformer decoder to upsample," 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2022: 1597-1602.
[21] G. Xu, X. Wu and X. Zhang, "Levit-unet: Make faster encoders with transformer for medical image segmentation," arXiv preprint arXiv:2107.08623, 2021.
[22] Y. Wen, C. Liang and J. Lin, "ExSwin-Unet: An Unbalanced Weighted Unet with Shifted Window and External Attentions for Fetal Brain MRI Image Segmentation," Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III. Cham: Springer Nature Switzerland, 2023: 340-354.
[23] D Kalman." A singularly valuable decomposition: the SVD of a matrix". The college mathematics journal, 1996, 27(1): 2-23.
[24] T. Peken, S. Adiga and R. Tandon, "Deep learning for SVD and hybrid beamforming," IEEE Transactions on Wireless Communications, 2020, 19(10): 6621-6642.
[25] H. Bendjador, T. Deffieux and M. Tanter, "The SVD beamformer: Physical principles and application to ultrafast adaptive ultrasound," IEEE transactions on medical imaging, 2020, 39(10): 3100-3112.
[26] K. M. Abdelwahab, S. M. Abd El-atty and W. El-Shafai, "Efficient SVD-based audio watermarking technique in FRT domain," Multimedia Tools and Applications, 2020, 79: 5617-5648.
[27] M. Diwakar, P. Kumar and P. Singh, "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.
[28] A. Bhatti,T. Ishii and N. Kanno, "Region-based SVD processing of high-frequency ultrafast ultrasound to visualize cutaneous vascular networks," Ultrasonics, 2023, 129: 106907.
[29] M. Shi, F. Zhang and S. Wang, "Detail preserving image denoising with patch-based structure similarity via sparse representation and SVD," Computer Vision and Image Understanding, 2021, 206: 103173.