UWT-Net: Mining low-frequency feature information for medical image segmentation

Pengcheng Zhang[1][0009−0000−2707−1149], Xiaocao Ouyang(✉)[2][0000−0002−4626−4373], and Ran Peng[1][0000−0002−0272−2926]

[1] College of Information Engineering, Sichuan Agricultural University, Ya'an, 625014, China. 2024319026@stu.sicau.edu.cn

[2] School of Computing and Artificial Intelligence, Southwestern University of Finance and Economics, Chengdu, 611130, China. oyxiaocao@swufe.edu.cn

Abstract. Medical image segmentation is the core technology of precision medicine, which can improve diagnostic accuracy, optimize treatment plans, and enhance research efficiency. U-Net is a classical and fundamental model in this field. Because of its excellent architecture, Transformer and MLP have been fused on top of it in subsequent work, all with good results. Each of these methods has advantages, but none further explores the image's low-frequency feature information. The low-frequency feature information reflects the overall structure and contour of the image and provides key background and boundary information for image segmentation. To address this problem, we explore the potential of Wavelet Convolutions for medical segmentation tasks by proposing a novel feature extraction block: the Image Multi-frequency Feature Information Extraction (IMFIE) block. The IMFIE block can effectively extract both high-frequency and low-frequency feature information from images by combining Wavelet Convolutions. This approach takes full advantage of their excellent ability to mine and utilize low-frequency information in images while expanding the receptive field at a low cost. We propose a novel model, UWT-Net, which leverages the IMFIE block and reconstructs the classical U-Net. Experiments on three public pathology image datasets show that the proposed method outperforms the state-of-the-art baseline U-KAN. Code is available at https://github.com/zpc2002zpc/UWT-Net.git.

Keywords: Medical image segmentation · Low-frequency feature information · U-Net · WTConv

1 Introduce

Medical image segmentation is an important research direction in medical image processing, which is extremely important in many aspects of clinical medicine, medical research, and medical equipment development. With the rapid development of deep learning and the popularization of computer-aided diagnosis, more and more research has focused on using deep learning for medical image segmentation [19].

U-Net is a milestone in this area [15], which uses a symmetric encoder-decoder architecture for end-to-end training and utilizes hopping to fuse feature information at different scales. Nested Skip Connections and Dense Feature Fusion proposed by U-Net++ [21] significantly improve the performance of U-Net by enhancing the effect of feature fusion and being able to utilize multi-scale features more effectively, respectively. More CNN-based improvements continue to advance the field [2,9,8]. However, CNNs are struggling to model global dependencies, resulting in suboptimal performance.

With the emergence and development of Transformer [18], related research has received attention. Att-Unet [13] effectively meets the needs of organ segmentation by introducing Attention Gates, which enable the network to automatically focus on the target region while suppressing irrelevant background information. TransUNet [5] and Swin-Unet [4] further applied Transformer to medical segmentation and achieved remarkable results. The attention mechanism can effectively compensate for the disadvantage of CNNs. Also, the work of combining CNNs and MLP has had remarkable results [17,11].

The Wavelet Transform has also given rise to a number of results in medical image segmentation. Integration of wavelet transform and converter modules into a modified U-Net architecture significantly improves the accuracy and model robustness of nasopharyngeal cancer image segmentation [20]. Spectral U-Net employs DTCWT and iDTCWT for down-sampling and up-sampling, respectively [14]. Wavelet transform is further applied to modify pooling methods [6].

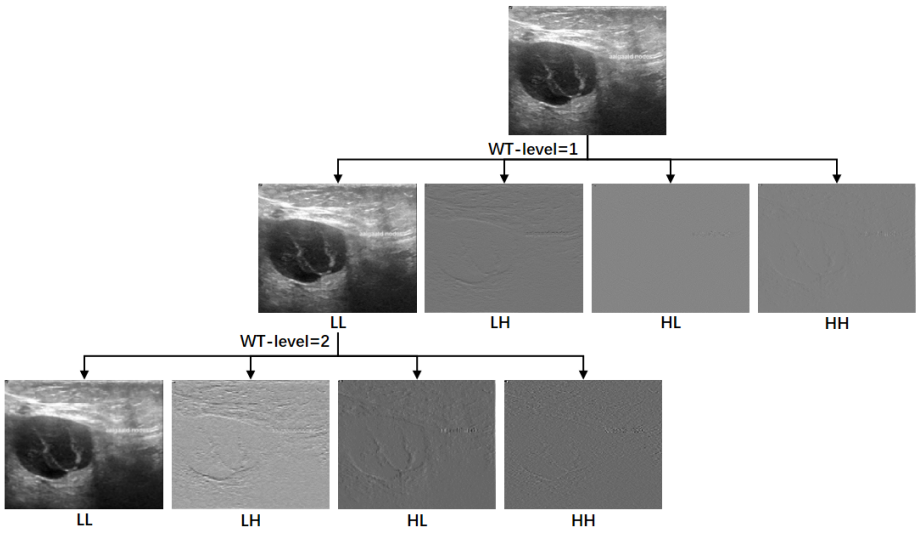

Fig. 1. Visualization of deep low-frequency feature information. Each level contains one low-frequency and three high-frequency wavelet subbands.

However, existing methods have demonstrated insufficient attention to low-frequency information in the images and do not further mine and utilize the low-frequency information. Fig. 1 visualizes the potential feature information embedded in low-frequency images. Low-frequency feature information mainly contains information about the global context and spatial structure of an image. In medical image segmentation, this information helps the model to understand the overall shape, location, and spatial relationship of organs or tissues, thus providing important contextual clues for the segmentation task. Furthermore, in the process of image downsampling, high-frequency information is easily lost, resulting in blurred boundaries. The retention and utilization of low-frequency feature information can reduce this loss of boundary information and thus improve the segmentation accuracy. Moreover, low-frequency information is relatively less sensitive to noise, so in medical image segmentation, utilizing low-frequency information can enhance the robustness of the model to noise. In conclusion, exploring in depth and utilizing this information wisely can be of immense help in medical image segmentation tasks.

In this work, we notice the above problem and, inspired by Wavelet Convolutions [7], we propose an Image Multi-Frequency Information Extraction (IMFIE) block. This block combines the advantages of standard convolution and WTConv, which can fully exploit and utilize low-frequency feature information to achieve more accurate medical image segmentation.

Our contributions can be summarized as follows: (1) We propose UWT-Net model that verifies the importance of image low-frequency feature information for medical image segmentation. (2) We design an Image Multi-Frequency Feature Information Extraction (IMFIE) block that can effectively extract image features at different frequencies, rationally utilizing the image low-frequency feature information that people overlook. (3) Experiments demonstrate that UWT-Net achieves state-of-the-art performance, with multiple sizes of models to meet different clinical needs.

2 Method

Fig. 2 illustrates UWT-Net's structure, which employs an encoder-decoder architecture centered on the IMFIE block. In the encoder, the IMFIE block extracts features and reduces resolution by half, while in the decoder, it restores resolution by doubling it. Feature connections between the encoder and decoder enable feature stitching, enhancing the model's performance.

2.1 Wavelet Convolutions

Wavelet Convolutions(WTConv) employ the Haar WT because it is efficient and straightforward [7]. The 2D WTConv uses four sets of filters to perform deep convolution:

$$F_{LL} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, F_{HH} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, F_{HL} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, F_{LH} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}. \tag{1}$$

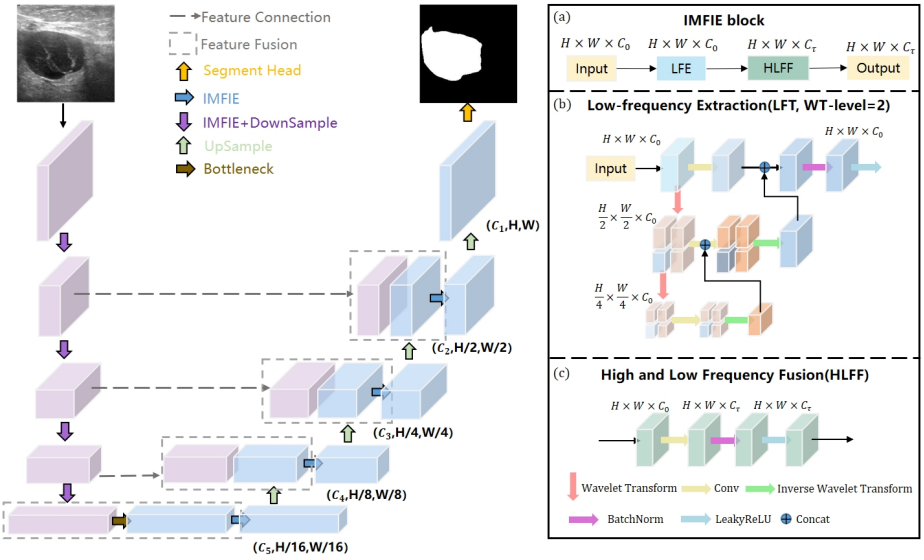

Fig. 2. Overview of UWT-Net. (a) IFMIE block, which contains Low-frequency Extraction (LFE) and High and Low Frequency Fusion (HLFF); (b) LFE mainly consists of Wavelet Convolutions, BatchNorm, and LeakyReLU; (c) HLFF is performed by standard convolution, BatchNorm, and LeakyReLU.

Note that F_{HH}, F_{HL}, and F_{LH} are a set of high-pass filters, and F_{LL} is a low-pass filter. After the convolution process, the output of each channel has four channels, and the spatial feature dimensions in these four channels are halved.

$$[N_{HH}, N_{HL}, N_{LH}, N_{LL}] = Conv\left([F_{HH}, F_{HL}, F_{LH}, F_{LL}], N\right), \tag{2}$$

where N_{HH}, N_{HL}, N_{LH} are N's horizontal, vertical, and diagonal high-frequency components, while N_{LL} is its low-frequency component. Since F_{HH}, F_{HL}, F_{LH}, F_{LL} form an orthonormal basis, the transposed convolution can realize the Inverse Wavelet Transform (IWT):

$$N = Conv{-}transposed\left([F_{HH}, F_{HL}, F_{LH}, F_{LL}], [N_{HH}, N_{HL}, N_{LH}, N_{LL}]\right). \tag{3}$$

The cascade wavelet decomposition is then given by recursively decomposing the low-frequency component, which increases the frequency resolution and decreases the spatial resolution of the lower frequencies.

$$N_{HH}^{(i)}, N_{HL}^{(i)}, N_{LH}^{(i)}, N_{LL}^{(i)} = WT\left(N_{LL}^{(i-1)}\right), \tag{4}$$

where $N_{LL}^{(0)} = N$ and i is the current level. The Convolution in the Wavelet Domain is implemented as follows: first, using the Wavelet Transform (WT) to filter and downscale the input lower- and higher-frequency content. Then, a small-kernel depth-wise convolution is performed on the different frequency

maps, and finally, using the Inverse Wavelet Transform (IWT) to construct the output.

$$M = IWT(Conv(W, WT(N))). \tag{5}$$

N is the input tensor, and W is the weight tensor of a k × k depth-wise kernel with four times as many input channels as N.

Next, take the WT-level=1 combined operation as an example, and increase it further using the same cascade principle from Eq.(5). The process is given by:

$$N_H^{(i)}, N_{LL}^{(i)} = WT\left(N_{LL}^{(i-1)}\right), \tag{6}$$

$$M_H^{(i)}, M_{LL}^{(i)} = Conv\left(W^{(i)}, \left(N_H^{(i)}, N_{LL}^{(i)}\right)\right), \tag{7}$$

where $N_{LL}^{(0)}$ is the input of the layer, and $N_H^{(i)}$represents all three high-frequency maps of level i described in Section 2.1.

Since the WT and its inverse are linear operations, we can combine the outputs of the different frequencies in the following way:

$$L^{(i)} = IWT\left(M_{LL}^{(i)} + L^{(i+1)}, M_H^{(i)}\right). \tag{8}$$

Results in the summation of the different levels' convolutions, where $L^{(i)}$ is the aggregated outputs from level i onward. These two outputs of different-sized convolutions are summed as the output.

2.2 UWT-Net Architecture

IMFIE Each block is constructed of the components as follows: a Wavelet Convolution layer (WTConv), a standard convolution layer (Conv), a batch normalization layer (BN), and a LeakyReLU activation function(LR). Formally, given an image $I \in R^{H_0 \times W_0 \times C_0}$.The output of each IMFIE block can be elaborated as follows:

$$O = LR(BN(WTConv(I))), \tag{9}$$

$$O_\tau = LR(BN(Conv(O))), \tag{10}$$

after IMFIE block, $O_\tau \in R^{H' \times W' \times C_\tau}$. Where C_τ is determined by the scale of the model, with the specific parameters set in Section 3.3.

UWT-Net Encoder In the encoder, the image data is downsampled by Maxpool (MP) after IMFIE to achieve feature resolution halving:

$$E_\tau = LR(MP(IMFIE(E_{\tau-1}))). \tag{11}$$

Bottleneck One standard convolution layer is used to construct the bottleneck to learn the deep feature representation. In the bottleneck, the feature dimension and resolution are kept unchanged.

UWT-Net Decoder the feature data E'_τ at layer τ in the encoder is spliced with the feature data D'_τ in the decoder in the channel dimension, after IMFIE and Upsample(US), to realize the doubling of the feature resolution and halving of the number of channels:

$$D_\tau = Cat(E'_\tau, D'_\tau), \tag{12}$$

$$D_{\tau-1} = LR(US(IMFIE(D_\tau))). \tag{13}$$

The final segmentation map can be derived from the output feature maps $D_0 \in R^{H_0 \times W_0 \times C_Y}$ at layer-0, where C_Y is the number of semantic categories and T denotes the ground-truth segmentation. As a result, the segmentation loss can be:

$$L_{Seg} = CE(T, UWT - Net(I)), \tag{14}$$

where CE denotes the pixel-wise cross-entropy loss. Overall, such a network design can effectively make up for the weakening of low-frequency information by standard convolution while inheriting the advantages of U-Net. By fusing the low-frequency and high-frequency information, the model can better balance the global structure and local details. The increase of receptive field brought by IMFIE can further improve the feature extraction ability and generalization ability of the model, which enhances the performance of the model.

3 Experiment

3.1 Dataset and Implementation Details

To validate the effectiveness, robustness, and generalizability of the proposed model, we conducted experiments on three publicly available datasets: BUSI [1], GlaS [16], and CVC-ClinicDB [3]. For a fair comparison, we use the same experimental setup as prior works [10].

The experiments were run using Pytorch on NVIDIA A100-PCIE-40GB GPU. UWT-Net was trained with an Adam optimizer with a learning rate of 1e-3, and we used a cosine annealing learning rate scheduler with a minimum learning rate of 1e-4. The loss function was a combination of binary cross entropy (BCE) and dice loss. We trained the model for 400 epochs in total. We use various metrics such as IoU and F1 Score to compare the output segmentation images both qualitatively and quantitatively. To account for the limited data size of the datasets, we repeated this process three times and reported the average and standard deviation of the results [10]. For all three datasets, we only applied vanilla data augmentations, including random rotation and flipping, and the batch size was set to 8. We randomly split each dataset into 80% training and 20% validation subsets. None of the experiments used any pre-trained weights or post-processing methods.

Table 1. Performance comparison of various methods on BUSI, GlaS, CVC datasets, and average results.

Methods	BUSI		GlaS	
	IoU	F1	IoU	F1
U-Net [15](MICCAI'15)	63.90±0.65	77.50±0.55	87.49±0.89	93.04±0.61
Att-Unet [13](MIDL'18)	64.19±0.38	78.26±0.34	88.05±0.17	93.42±0.09
U-Net++ [21](MICCAI'18)	57.41±4.77	72.11±3.90	87.07±0.76	92.96±0.44
U-NeXt [17](MICCAI'22)	60.19±0.72	74.94±0.59	84.32±0.34	91.48±0.20
Rolling-UNet [11](AAAI'24)	63.74±0.32	77.64±0.19	87.74±0.18	93.46±0.10
U-Mamba [12](arXiv'24)	61.81±3.24	75.55±3.01	87.01±0.39	93.02±0.24
U-KAN [10](AAAI'25)	63.38±2.83	76.40±2.90	87.64±0.32	93.37±0.16
UWT-Net (Ours)	**66.57±0.56**	**79.65±0.56**	**88.26±0.29**	**93.76±0.16**
Methods	CVC		Average	
	IoU	F1	IoU	F1
U-Net [15](MICCAI'15)	83.71±0.48	91.05±0.31	78.36±0.67	87.20±0.49
Att-Unet [13](MIDL'18)	83.57±0.54	90.94±0.32	78.60±0.36	87.54±0.25
U-Net++ [21](MICCAI'18)	84.61±1.47	91.53±0.88	76.36±2.33	85.53±1.74
U-NeXt [17](MICCAI'22)	74.25±0.54	84.92±0.35	72.92±0.53	83.78±0.38
Rolling-UNet [11](AAAI'24)	81.31±0.78	89.48±0.49	77.60±0.43	86.86±0.26
U-Mamba [12](arXiv'24)	84.79±0.58	91.63±0.39	77.87±1.47	86.73±1.25
U-KAN [10](AAAI'25)	85.05±0.53	91.88±0.29	78.69±1.27	87.22±1.15
UWT-Net (Ours)	**86.08±2.46**	**92.46±1.50**	**80.30±1.10**	**88.62±0.74**

3.2 Performance Comparison

We evaluated our approach against seven baseline methods. These methods include U-Net [15] and U-Net++ [21], which are based entirely on traditional convolutional neural network designs, Att-Unet [13], which is based on attention mechanisms, and the efficient transformer variant, U-Mamba [12]. We also evaluated performance against U-NeXt [17], which combines convolution and MLP, Rolling-UNet [11], and the state-of-the-art, U-KAN [10]. The results in Table 1 indicate that our UWT-Net outperforms all other methods[3], which shows that our method has good generalization ability and robustness.

3.3 Ablation Study

We performed a variety of ablation studies to thoroughly evaluate the proposed UWT-Net framework and validate the performance under different settings.

Table 2. Performance with and without IMFIE block across different datasets.

Feature extraction block	BUSI		GlaS		CVC	
	IoU	F1	IoU	F1	IoU	F1
standard convolution	62.28	75.86	83.58	91.37	80.11	88.72
IMFIE	**67.14**	**80.17**	**88.54**	**93.92**	**87.63**	**93.38**

[3] Results of U-Net++, U-Mamba, and U-KAN are referenced from U-KAN [10]

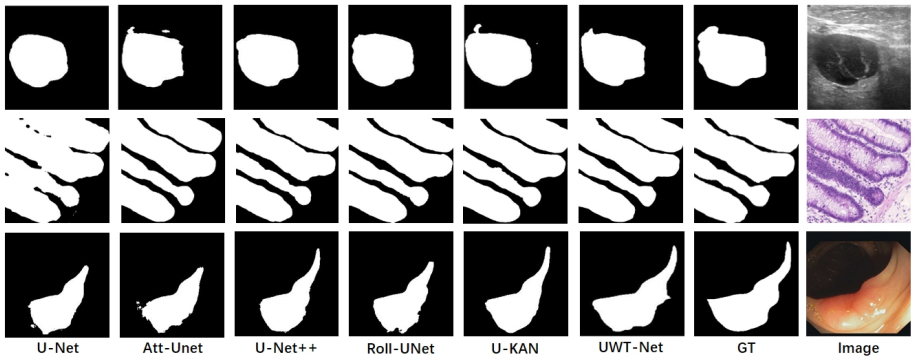

Fig. 3. The results of the proposed UWT-Net against all the compared methods over all the benchmarking datasets.

Table 3. Performance with different WT-levels across average segmentation results.

WT-levels	Average	
	IoU	F1
Level=1	79.56±0.83	88.23±0.54
Level=2	**80.30±1.10**	**88.62±0.74**
Level=3	80.02±0.60	88.46±0.36

We replaced the introduced IMFIE block with standard convolution block. The results in Table 2 highlight the effectiveness of the IMFIE block. This confirms the usefulness of low-frequency feature information in medical images for medical segmentation tasks. The results in Table 3 show that deeper mining of low-frequency information is more effective in improving segmentation, further demonstrating the potential value of low-frequency information in images.

Table 4. Performance comparison with different model scales across average segmentation results.

Model Scale	Average		Efficiency	
	IoU	F1	Gflops	Params(M)
UWT-Net-S	79.38±1.08	88.02±0.75	8.37	7.71
UWT-Net	80.30±1.10	88.62±0.74	32.44	29.54
UWT-Net-L	80.61±0.36	89.03±0.28	127.70	155.56

We performed an ablation study on UWT-Net variants, including UWT-Net-S ([32, 64, 128, 256, 512] channels) and UWT-Net-L ([128, 256, 512, 1024, 2048] channels), compared to our default model UWT-Net ([64, 128, 256, 512, 1024] channels). Results in Table 4 show that performance improves with model size. To balance performance and computational cost, we selected the default

model UWT-Net. Our UWT-Net achieves leading accuracy across all scales, demonstrating its potential to meet diverse clinical needs.

4 Conclusion

In this study, we explore the potential of underutilized low-frequency features in medical imaging and propose an innovative block IMFIE and a novel model, UWT-Net. We perform empirical evaluations of our method under three medical image segmentation tasks. The results demonstrate that our model can effectively mine and utilize the overlooked low-frequency feature information and fuse it with the high-frequency information in medical images, improving segmentation effects. UWT-Net has the potential to be used in a wide range of medical image segmentation tasks.

Acknowledgments. This research was supported by the National Natural Science Foundation of China (Grant No:62406259).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in brief **28**, 104863 (2020)
2. Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K.: Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955 (2018)
3. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized medical imaging and graphics **43**, 99–111 (2015)
4. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision. pp. 205–218. Springer (2022)
5. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
6. El-Khamy, S., El-Bana, S., Al-Kabbany, A., Elragal, H.: Toward better semantic segmentation by retaining spectral information using matched wavelet pooling. Neural Computing and Applications pp. 1–18 (2025)
7. Finder, S.E., Amoyal, R., Treister, E., Freifeld, O.: Wavelet convolutions for large receptive fields. In: European Conference on Computer Vision. pp. 363–380. Springer (2024)
8. Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.W., Wu, J.: Unet 3+: A full-scale connected unet for medical image segmentation. In: ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). pp. 1055–1059. IEEE (2020)

9. Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., De Lange, T., Halvorsen, P., Johansen, H.D.: Resunet++: An advanced architecture for medical image segmentation. In: 2019 IEEE international symposium on multimedia (ISM). pp. 225–2255. IEEE (2019)
10. Li, C., Liu, X., Li, W., Wang, C., Liu, H., Liu, Y., Chen, Z., Yuan, Y.: U-kan makes strong backbone for medical image segmentation and generation. arXiv preprint arXiv:2406.02918 (2024)
11. Liu, Y., Zhu, H., Liu, M., Yu, H., Chen, Z., Gao, J.: Rolling-unet: Revitalizing mlp's ability to efficiently extract long-distance dependencies for medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 3819–3827 (2024)
12. Ma, J., Li, F., Wang, B.: U-mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722 (2024)
13. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
14. Peng, Y., Sonka, M., Chen, D.Z.: Spectral u-net: Enhancing medical image segmentation via spectral decomposition. arXiv preprint arXiv:2409.09216 (2024)
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
16. Valanarasu, J.M.J., Oza, P., Hacihaliloglu, I., Patel, V.M.: Medical transformer: Gated axial-attention for medical image segmentation. In: Medical image computing and computer assisted intervention–MICCAI 2021: 24th international conference, Strasbourg, France, September 27–October 1, 2021, proceedings, part I 24. pp. 36–46. Springer (2021)
17. Valanarasu, J.M.J., Patel, V.M.: Unext: Mlp-based rapid medical image segmentation network. In: International conference on medical image computing and computer-assisted intervention. pp. 23–33. Springer (2022)
18. Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems (2017)
19. Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., Nandi, A.K.: Medical image segmentation using deep learning: A survey. IET image processing **16**(5), 1243–1267 (2022)
20. Zeng, Y., Li, J., Zhao, Z., Liang, W., Zeng, P., Shen, S., Zhang, K., Shen, C.: Wetunet: Wavelet integrated efficient transformer networks for nasopharyngeal carcinoma tumor segmentation. Science Progress **107**(2), 00368504241232537 (2024)
21. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. pp. 3–11. Springer (2018)