PFESA: FFT-based Parameter-Free Edge and Structure Attention for Medical Image Segmentation

Mingqian Li[1] , Zhiqian Yan[1] , Miaoning Yan[2] , Yaodong Liang[1] ,

Qingmao Zhang[1(✉)] , and Qiongxiong Ma[1(✉)]

[1] Guangdong Provincial Key Laboratory of Nanophotonic Functional Materials and Devices, School of Optoelectronic Science and Engineering, South China Normal University, Guangzhou, China
maqx@m.scnu.edu.cn

[2] Riton 3D., Ltd., Guangzhou, China

Abstract. The U-Net architecture remains pivotal in medical image segmentation, yet its skip connections often propagate redundant noise and compromise edge information. We propose a **P**arameter-**F**ree **E**dge and **S**tructure **A**ttention (**PFESA**) based on Fast Fourier Transform (FFT) to address these limitations. PFESA employs frequency-domain feature decoupling to separate high-frequency (edge details) and low-frequency (structural components) representations. Leveraging feature Signal-to-Noise Ratio(SNR) analysis, we devise dual attention paths: a High-frequency Edge Attention (EA) enhances gradient-sensitive regions to preserve anatomical contours, while a Low-frequency Structure Attention (SA) suppresses noise through energy redistribution. This frequency-aware attention mechanism enables adaptive feature refinement in skip connections without introducing trainable parameters. The parameter-free design ensures robustness against overfitting in medical datasets with scarce data. Extensive experiments on multi modal 2D/3D medical image datasets demonstrate PFESA's superiority over existing attention methods, achieving SOTA performance with statistically significant improvements in Dice Similarity Coefficient (DSC: +3.3% vs. baseline) and Hausdorff Distance metrics. Code is available at: https://github.com/59-lmq/PFESA.

Keywords: Medical Image Segmentation · Parameter-Free Attention · FFT-based Feature Decoupling.

1 Introduction

As a cornerstone of computer-aided diagnosis, medical image segmentation accuracy directly determines the reliability of lesion localization and surgical planning. U-Net [1] and its variants, which employ skip connections to integrate

multi-scale features, have become benchmark architectures in this domain. However, these direct feature concatenation mechanisms present three inherent limitations [2]: 1) Low signal-to-noise ratio (SNR) background noise in shallow encoder features interferes with semantic reconstruction in decoders; 2) Progressive downsampling induces attenuation of high-frequency edge information; 3) Existing fusion mechanisms lack interpretability for establishing correspondence between feature weights and anatomical structures.

To address these challenges, attention mechanisms enhance critical regions through dynamic feature weighting. SENet [3] models global dependencies via channel compression but introduces overfitting risks in small-sample scenarios through fully-connected layers. CBAM [4] combines spatial-channel attention at the cost of quadratic computational complexity growth with channel numbers. ECA-Net [5] reduces parameters through 1D convolution but remains data-dependent. Attention U-Net [2] selectively enhances salient features via dual attention in skip connections. While improving performance, such parameterized modules risk overfitting in medical images' data-scarce context [6] and obscure clinical credibility due to their black-box nature - clinicians cannot verify spatial correlations between attention weights and anatomical landmarks [7]. Recent parameter-free attention methods (e.g., SimAM [8]'s energy function-based attention) enhance generalizability but neglect the frequency characteristics of medical images, which high-frequency components encode anatomical edges while low-frequency components represent organ morphology. Conventional convolutions struggle to decouple frequency information due to limited receptive fields, whereas spectral processing enables physically interpretable feature enhancement through global analysis. Liu et al. [6] enhances ViT high frequencies via Laplacian operators but lost low-frequency structures. SF-Net [7] uses rectangular masks extracting frequencies after FFT, causing discontinuous/noisy high frequencies. WaveSNet [9] replaces downsampling with wavelets but ignores high-frequency in skip connections, introducing extra noise. PFD-Net [10] applies 1x1 convolutions in frequency domain without frequency ban analysis. MC-FDN [22] neglects frequency continuity via channel attention on real/imaginary components after FFT. SASAN [23] fuses denoised low frequencies but discarded high-frequency feature.

To overcome these limitations, we propose a novel **P**arameter-**F**ree **E**dge-**S**tructure **A**ttention (**PFESA**) mechanism based on Fast Fourier Transform (FFT). Our contributions include three key aspects: 1) We implements Gaussian filtering to separate high-frequency (edge details) and low-frequency (structural components) features, avoiding optimization bias from trainable filters, and decoupling frequency-domain feature; 2) We propose a SNR-Driven Parameter-Free Attention which enhances edge SNR through local energy maximization in high-frequency components and reinforces structural morphology via contrast-enhanced normalization in low-frequency components. The statistical computation process eliminates trainable parameters, ensuring mathematically traceable weight generation. Grad-CAM visualization further enhances interpretability; 3) When integrated into skip connections of both CNN-based and Transformer-

based models, our method demonstrates state-of-the-art performance across multiple 2D/3D multi-modal medical image datasets, confirming its robustness.
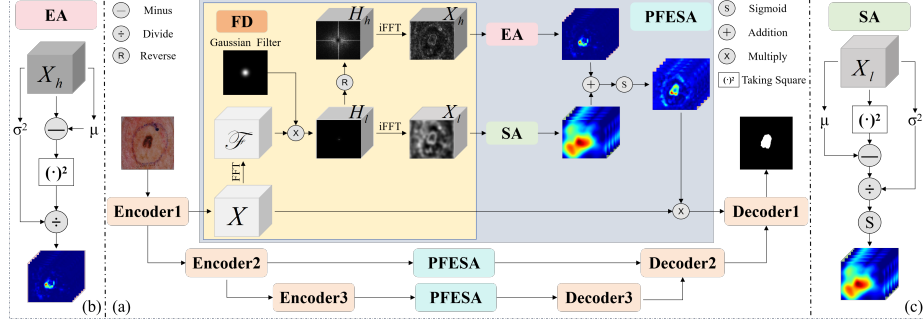

Fig. 1. An overview of the PFESA(a), which comprises Frequency-Domain **F**eature **D**ecouping, **E**dge **A**ttention(b) and **S**tructure **A**ttention(c).

2 Methodology

Fig. 1 illustrates the overall framework of the proposed PFESA and shows an example usage of it. Since it is a parameter-free plug-to-play attention module with the same channels input and output channels, PFESA can be plugged into any skip connection. In this section, We first introduce the Frequency-Domain Feature Decouping (Sect. 2.1) and then concentrate on the details of Parameter-Free SNR-Driven Attention (Sect. 2.2).

2.1 Frequency-Domain Feature Decouping

Medical image frequency components exhibit distinct separability in the spectral domain, which high-frequency elements consist edges and textures while low-frequency components contain organ morphology. Using 2D data as an example (with 3D cases following the same principle), we first transform input feature maps $X \in \boldsymbol{R}^{C \times H \times W}$ into the frequency domain $\mathcal{F}(X) \in R^{C \times H \times W}$ via Fast Fourier Transform (FFT). Gaussian filtering $G_\sigma(u,v) = \exp\left(-\frac{u^2+v^2}{2\sigma^2}\right)$ is then applied for frequency-band decoupling: $H_l = G_\sigma \odot \mathcal{F}(x), H_h = 1 - H_l$, where the parameter $\sigma = r \cdot \frac{\min(H,W)}{\max(H,W)}$, $r = 0.1$ controls the cutoff frequency (see Sect. 3.3 for ablation studies). The high-frequency features X_h and low-frequency features X_l are subsequently reconstructed through inverse FFT.

2.2 Parameter-Free SNR-Driven Attention

The signal-to-noise ratio (SNR) of medical image features can be mathematically defined as the ratio of signal energy to noise energy [11]: $SNR = \frac{E[S^2]}{E[N^2]} = \frac{\mu_S^2}{\sigma_N^2}$,

where σ_N^2 denotes noise variance and μ_s^2 approximates the signal energy intensity. This fundamental distinction necessitates differentiated computational frameworks for modeling SNR in high-frequency edge features versus low-frequency structural features.

Edge Attention(EA) High-frequency features S_h exhibit intense gradient variations where local energy (quantified as variance) significantly surpasses noise energy in anatomical edge regions of medical images. Assuming additive Gaussian noise distribution $N_h \sim N(0, \sigma_{n,h}^2)$, let $\sigma_{s,h}^2$ and $\sigma_{n,h}^2$ denote signal and noise variances respectively. In edge regions, $\sigma_{s,h}^2 \gg \sigma_{n,h}^2$ will be present. The edge SNR is formulated as:

$$SNR_{edge} = \frac{E[S_h^2]}{\sigma_{n,h}^2} \approx \frac{E[(X_h - \mu_h)^2]}{\sigma_{n,h}^2} \tag{1}$$

By substituting $\sigma_{n,h}^2$ with σ_h^2 (feature variance), we derive the Edge Attention (EA) map normalized to [0, 1].

$$EA = \frac{(X_h - \mu_h)^2}{\sigma_h^2}, \tag{2}$$

where σ_h^2 is feature variance. For high-frequency feature S_h, when $(X_h - \mu_h)^2 \approx \sigma_{s,h}^2, \sigma_h^2 \approx \sigma_{s,h}^2$ in edge regions, $EA \approx 1$; In noise-dominated regions, when $(X_h - \mu_h)^2 \approx \sigma_{n,h}^2, \sigma_h^2 \approx \sigma_{s,h}^2 + \sigma_{n,h}^2$, $EA \approx \frac{\sigma_{n,h}^2}{\sigma_{n,h}^2 + \sigma_{s,h}^2} \to 0$.

Structure Attention(SA) Low-frequency artifacts noise N_l in medical image typically exhibit energy distributions comparable to anatomical signals S_l [24]. To amplify their discriminability, we strategically enhance dominant regions through the squared mean-to-variance difference:

$$SNR_{structure} = \frac{E[S_l^2]}{\sigma_{n,l}^2} \approx \frac{E[(X_l^2 - \mu_l)]}{\sigma_{n,l}^2} \tag{3}$$

Substituting $\sigma_{n,l}^2$ with σ_l^2 and applying Sigmoid normalization yields the Structure Attention (SA):

$$SA = Sigmoid\left(\frac{X_l^2 - \mu_l}{\sigma_l^2}\right) \tag{4}$$

where σ_l^2 is feature variance. By applying Sigmoid for $SNR_{structure}$, SA is constrained to the range of [0.5, 1]. For low-frequency feature S_l, when $X_l^2 - \mu_l \gg 0, \sigma_l^2 \approx \sigma_{s,l}^2$, $SNR_{structure}$ and SA approaches 1; In noise-dominated regions, when $X_l^2 - \mu_l \approx 0$, $SNR_{structure}$ approaches 0 and SA approaches 0.5.

The high-frequency Edge Attention (EA) enhances boundary signal-to-noise ratio through localized energy normalization, while the low-frequency Structure Attention (SA) amplifies morphological discriminability via global variance contrast. These dual mechanisms operate complementarily in the spectral domain to collectively refine spatial-domain representations. The parameter-free design ensures mathematically traceable weight generation and anatomical coherence

with medical image characteristics, guaranteeing physically plausible and interpretable feature enhancement. The final output after PFESA is obtained by fusing EA and SA through a Sigmoid-activated fusion gate, followed by element-wise multiplication with input features:

$$A = Sigmoid(EA + SA), X_{out} = A \odot X \qquad (5)$$

3 Experiments

3.1 Datasets and Implementation

Datasets. We conduct extensive experiments on two 2D datasets (GlaS [12] and ISIC-2017 [13]) and two 3D datasets (LA [14] and Tooth [15]). GlaS is a gland segmentation dataset, including 165 Hematoxylin and Eosin (H&E) stained images of benign and malignant tissue. The number of training, valid and test images are 85, 20 and 60, respectively. ISIC-2017 is curated from the ISIC Challenge Dataset 2017 provided by the International Skin Imaging Collaboration (ISIC), comprising 2000 images for training, 150 images for validation and 600 images for test. LA is a left atrial segmentation dataset from 2018 Atrial Segmentation Challenge. It consists of 100 3D LGI-MRI images, with a resolution of $0.625 \times 0.625 \times 0.625$ mm. Following [16], we use 80 images for training and 20 images for testing. Tooth dataset is collected from a subset of a large scale dataset [15], which is used for segmentation and reconstruction of individual teeth and alveolar bone to aid in dental treatment. Following [17], we use 103 images for training and 26 images for testing.

Implementation Details. We implemented our framework using PyTorch. Training and inference of all models were performed on one NVIDIA GeForce RTX 4090. The input resolutions of 2D datasets and 3D datasets were 256^2 and 128^3, respectively. The loss function was DiceCE Loss. We ran 500 epochs for all 2D datasets and 10000 iterations for all 3D datasets. We used SGD optimizer, with initial learning rates of 0.1 and 0.01, decays of 1×10^{-5} and 1×10^{-4}, and momentum of 0.9 and 0.99 for 2D and 3D datasets, respectively. Performance was evaluated using the dice similarity coefficient(DSC,%), and hausdorff distance(HD, mm) for all the experiments.

3.2 Comparison with the State-of-the-Art Methods

The baseline frameworks comprise CNN-based models (U-Net [1] and 3D U-Net [18]) and Transformer-based frameworks (TransUNet [19] for 2D and UN-etr++ [20] for 3D). We compared our method against with parameterized attention mechanisms (CBAM [4], SE [3], ECA [5]) and parameter-free attention mechanisms (SimAM [8] and SIAM [21]). All attention modules are integrated into every skip connection of the baseline architectures. For a fair comparison,

Table 1. Comparison experimental results obtained from four datasets, highlighting the best scores with **bold**.

2D Methods	GlaS DSC↑	GlaS HD↓	ISIC-2017 DSC↑	ISIC-2017 HD↓	3D Methods	LA DSC↑	LA HD↓	Tooth DSC↑	Tooth HD↓
U-Net [1]	88.61	11.51	84.09	11.40	3D U-Net [18]	84.55	15.68	84.88	25.27
+CBAM [4]	87.31	16.23	83.43	10.67	+CBAM [4]	86.01	16.26	86.88	9.26
+SE [3]	84.21	17.49	83.05	12.01	+SE [3]	86.29	14.35	87.36	8.66
+ECA [5]	84.34	18.32	83.47	11.62	+ECA [5]	87.09	15.12	85.00	15.62
+SimAM [8]	87.44	12.07	83.87	11.77	+SimAM [8]	86.63	14.23	84.99	31.81
+SIAM [21]	88.05	14.66	83.79	12.43	+SIAM [21]	83.72	20.30	80.08	26.40
+PFESA(ours)	**90.18**	**8.13**	**84.25**	**10.12**	+PFESA(ours)	**87.17**	**12.91**	**88.21**	**8.22**
TransUNet [19]	88.83	13.76	84.55	10.16	UNetr++ [20]	89.78	6.75	86.27	2.47
+CBAM [4]	87.92	14.79	83.50	10.91	+CBAM [4]	89.47	7.82	84.20	4.99
+SE [3]	86.26	14.75	83.60	11.20	+SE [3]	89.40	10.13	87.30	2.38
+ECA [5]	88.06	13.25	84.44	10.36	+ECA [5]	89.52	11.98	86.97	2.40
+SimAM [8]	85.88	15.96	84.84	10.36	+SimAM [8]	90.09	6.74	87.22	2.34
+SIAM [21]	86.48	15.65	84.15	10.14	+SIAM [21]	88.65	7.41	81.44	9.45
+PFESA(ours)	**89.33**	**13.21**	**85.27**	**9.54**	+PFESA(ours)	**90.16**	**6.72**	**87.77**	**2.14**

we used publicly available codes for all methods. We reported segmentation performance solely on individual model accuracy, without using pre-training, model ensembling, or extra data.

Quantitative Analysis. Table 1 presents detailed quantitative comparisons, with **bold** indicating the best results. On the GlaS [12] dataset, other attention methods degrade performance versus baseline, which DSC reduced at least 0.56%, HD is increased by at least 0.56 mm, indicating that these methods are overfitting on scarce dataset. However, our method significantly reduces HD (reduced by 3.4 mm compared to U-Net [1] and 0.55 mm compared to TransUNet [19]) while improving DSC. In the ISIC-2017 [13] dataset, it is noted that except for the improvement of SimAM [8] added to TransUNet [19] over TransUNet [19], the DSC of other methods has decreased (DSC decreased by at least 0.11%). Our method has improved the performance of other methods (DSC increased by 0.16% compared to U-Net [1] and 0.72% compared to TransUNet [19]). It shows that our method can also be effective in the field of dermoscopic images and natural-like images. In the LA [14] dataset and Tooth [15] dataset, it can be found that when using 3D U-Net [18] as the baseline, our method has significantly decreased in HD compared to other methods (HD of the Tooth [15] dataset decreased from 25.27 mm of the baseline to 8.22 mm), which highlights the superiority of our performance. When using UNetr++ [20] as the baseline, it can be found that the baseline is better than our method in 3D U-Net [18] (DSC of the LA [14] dataset increased by 2.61%), but after adding our method, the indicator is still improved (DSC from 89.78% to 90.16%). The results show that our method can achieve SOTA level on both CNN-Based Model and Transformer-Based Model, also achieving SOTA on datasets of different modalities, which verifies the robustness of our method.
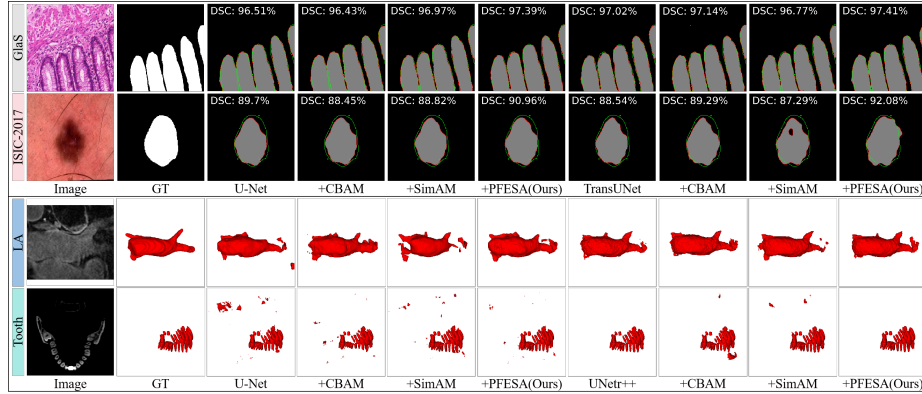
Fig. 2. Qualitative examples of four datasets. The first and second columns are the raw images and labels, and the other columns are the predictions of different methods.

Qualitative Analysis. Figure 2 shows some specific examples from the comparative experiments. Upon visually inspecting the results of other attention mechanisms, it is apparent that our method pays more attention to the boundary regions. Figure 3 presents the Grad-Cam visualizations of different layers of Skip Connections (SC-1, SC-2, SC-3) in TransUNet [19], applied to the GlaS [12] and ISIC-2017 [13] datasets using different attention mechanisms. It can be seen that in the lower layers with high background noise (SC-1), other methods lose edge information, while our method emphasizes feature edges and details. In the deeper layers (SC-2 and SC-3), while we focus on the edges, we also retain attention on the main structural features, whereas other methods only focus on certain regions of the target features. As shown in Table 1 and Figure 3, our method achieves lower HD values compared to other approaches and effectively preserves high-frequency edge information at skip connections, enhancing the feature signal-to-noise ratio, which intuitively supports the reliability and interpretability of Eq. 2 and Eq. 4.

3.3 Ablation Study

All ablation experiments were conducted on the ISIC-2017 [13] dataset using U-Net [1]. Table 2 demonstrates the impact of varying cutoff frequencies (r) on frequency-band separation and segmentation performance. As r increases from 0.1 to 0.5, the Hausdorff Distance (HD) progressively rises (from 10.12 mm to 11.02 mm), indicating that broader low-frequency bandwidth captured by the Gaussian filter reduces high-frequency signal energy. This degradation in the SNR of skip connection directly correlates with performance decline (DSC decreased by 0.35%). Table 3 shows the impact of different components (FD: Frequency-Domain Feature Decoupling, SA: Structure Attention, EA: Edge Attention) on segmentation performance. When only FD is used, the sum of high- and low-frequency features replaces the original skip connection features. The
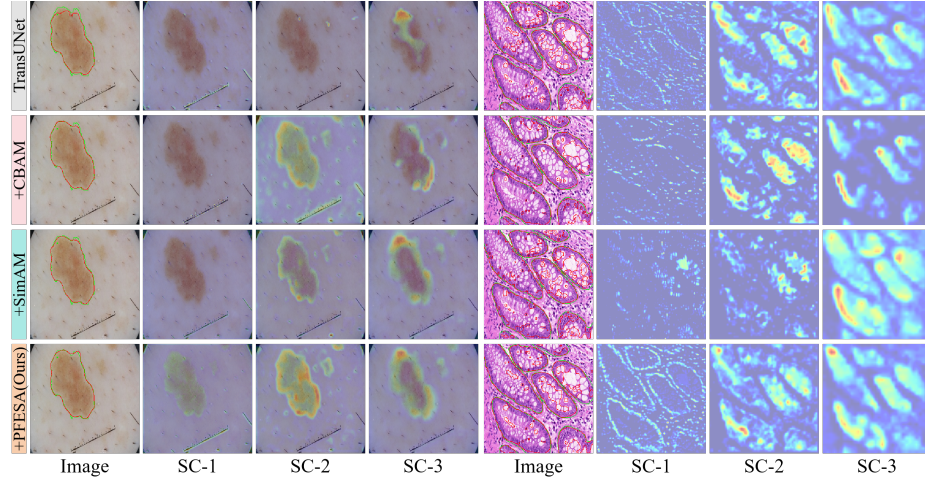
Fig. 3. Visualization of each skip connection feature using the trained TransUNet with different attention. The green and red line in Image represent Ground Truth and Prediction respectively. SC-1 means that low-level skip connection. SC-2 and SC-3 represent the high-level skip connection.

Table 2. Ablation of the r value.

r	DSC↑	HD↓
0.1	**84.25**	**10.12**
0.2	83.21	11.53
0.3	84.19	11.49
0.4	84.08	10.78
0.5	83.90	11.02

Table 3. Ablation of the each component.

FD	SA	EA	DSC↑	HD↓
			84.09	11.40
✓			84.14	11.10
✓	✓		84.17	11.02
✓		✓	84.21	10.95
✓	✓	✓	**84.25**	**10.12**

Table 4. Ablation of the Feature fusion strategies.

Method	DSC↑	HD↓
baseline	84.09	11.40
Replace	83.78	10.94
ADD	**84.26**	10.59
Mul	84.25	**10.12**

results in Table 3 show that with only FD, DSC increases by 0.05% and HD decreases by 0.3 mm compared to the baseline. When only SA is used, the model focuses more on structural information at the skip connections, reducing some redundant features. When only EA is used, HD decreases by 0.45 mm compared to the baseline, indicating that the model focuses more on edge information at the skip connections. Our method achieves the best results, with HD reduced by 1.28 mm and DSC increased by 0.14%, showing that the model, while emphasizing edge information at the skip connections, also captures structural information and suppresses background noise. Table 4 compares feature fusion strategies. Direct replacement slightly degrades DSC (-0.31%) but improves HD (-0.46mm), validating edge-specific enhancements. Additive fusion achieves peak DSC (84.26%) at the cost of compromised HD due to feature distribution distortion. Multiplicative fusion balances accuracy and robustness (DSC: 82.5%, HD: 10.12), preserving original feature statistics while integrating attention guidance.

4   Conclusion

In this study, we addresses noise interference and edge degradation in skip connections by proposing a Fast Fourier Transform (FFT)-based parameter-free edge and structural attention mechanism. Through frequency-domain decomposition, we disentangle edge and structural features, then construct SNR-driven attention modules to collaboratively suppress noise while preserving critical edge information. Experimental results demonstrate state-of-the-art performance across multi modal datasets with varying baselines, confirming the method's robustness. Our work establishes a novel frequency feature decoupling paradigm for medical image segmentation, offering fresh insights into architectural design through interpretable frequency-space operations.

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

References

1. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) Medical Image Computing and Computer-Assisted Intervention- MICCAI 2015. MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham., Munich (2015). https://doi.org/10.1007/978-3-319-24574-4_28
2. Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D.: Attention gated networks: Learning to leverage salient regions in medical images. Medical Image Analysis **53** 197–207 (2019).
3. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132–7141. IEEE, Salt Lake City, UT, USA (2018).
4. Woo, S., Park, J., Lee, JY., Kweon, I.S.: CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Proceedings of the European Conference on Computer Vision (ECCV) 2018, LNCS, vol. 11211, pp.3–19. Springer, Munich (2018). https://doi.org/10.1007/978-3-030-01234-2
5. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: ECA-Net: Efficient channel attention for deep convolutional neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11534–11542. IEEE, Seattle, WA, USA (2020).
6. Qiong, L., Chaofan, L., Jinnan, T., Liping, C., Jianxiang, S.: Medical image segmentation based on frequency domain decomposition SVD linear attention. Scientific Reports, **15**(1), 2833 (2025).

7. Zhou, Z., He, A., Wu, Y., Yao, R., Xie, X., Li, T.: Spatial-Frequency Dual Domain Attention Network For Medical Image Segmentation. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 4076–4081. IEEE, Lisbon, Portugal (2024).

8. Yang, L., Zhang, R. Y., Li, L., Xie, X.: Simam: A simple, parameter-free attention module for convolutional neural networks. In: Proceedings of the 38th International Conference on Machine Learning, pp. 11863–11874. PMLR (2021).

9. Li, Q., Shen, L.: Wavesnet: Wavelet integrated deep networks for image segmentation. In: Yu, S., et al. (eds.) Pattern Recognition and Computer Vision. PRCV 2022, LNCS, vol. 13537, pp. 325–337. Springer, Shenzhen (2022). https://doi.org/10.1007/978-3-031-18916-6_27

10. Yang, C., Zhang, Z.: Pfd-net: Pyramid fourier deformable network for medical image segmentation. Computers in Biology and Medicine **172** 108302 (2024).

11. Hoogeboom, E., Heek, J., Salimans, T.: simple diffusion: End-to-end diffusion for high resolution images. In: Proceedings of the 40th International Conference on Machine Learning, pp. 13213–13232. PMLR, Hawaii (2023).

12. Sirinukunwattana, K., Pluim, J. P., Chen, H., Qi, X., Heng, P. A., Guo, Y. B., ... Rajpoot, N. M.: Gland segmentation in colon histology images: The glas challenge contest. Medical image analysis **35** 489–502 (2017).

13. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). pp. 168–172. IEEE (2018)

14. Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., ... Zhao, J.: A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Medical image analysis **67** 101832 (2021).

15. Cui, Z., Fang, Y., Mei, L., Zhang, B., Yu, B., Liu, J., ... Shen, D.:A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images. Nature communications, **13**(1), 2096 (2022).

16. Zhou, Y., Huang, J., Wang, C., Song, L., Yang, G.: Xnet: Wavelet-based low and high frequency fusion networks for fully-and semi-supervised semantic segmentation of biomedical images. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 21085–21096. IEEE, Paris, France (2023).

17. Zhong, J., Tian, W., Xie, Y., Liu, Z., Ou, J., Tian, T., Zhang, L.: PMFSNet: Polarized multi-scale feature self-attention network for lightweight medical image segmentation. Computer Methods and Programs in Biomedicine **261** 108611 (2025).

18. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G., Wells, W. (eds.) Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham., Athens, Greece (2016). https://doi.org/10.1007/978-3-319-46723-8_49

19. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., ... Zhou, Y.: TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis, **97**, 103280 (2024).

20. Shaker, A. M., Maaz, M., Rasheed, H., Khan, S., Yang, M. H., Khan, F. S.: UN-ETR++: delving into efficient and accurate 3D medical image segmentation. IEEE Transactions on Medical Imaging **43**(9) (2024).
21. Han, G., Huang, S., Zhao, F., Tang, J.: SIAM: a parameter-free, spatial intersection attention module. Pattern Recognition **153** 110509 (2024).
22. Caixia, D., Duwei, D., Zongfang, L., Songhua, X.: A novel deep network with triangular-star spatial–spectral fusion encoding and entropy-aware double decoding for coronary artery segmentation. Information Fusion **112** 102561 (2024).
23. Xingru H., Jian H., Kai Z., Tianyun Z., Zhi L., Changpeng Y.: Sasan: Spectrum-axial spatial approach networks for medical image segmentation. IEEE Transactions on Medical Imaging **43**(8) 3044-3056 (2024).
24. Tanguay J., Kim J., Kim H.K., Iniewski K., Cunningham IA.: Frequency-dependent signal and noise in spectroscopic x-ray imaging. Medical Physics. **47**(7), 2881–2901 (2020)