A Dual-Branch Multidomain Feature Fusion Network for Axial Super-Resolution in Optical Coherence Tomography

Quanqing Xu, Xiang He, Muhao Xu, Kaixuan Hu, and Weiye Song

***Abstract*—High-resolution retinal optical coherence tomography(OCT) images are crucial for the diagnosis of numerous retinal diseases, but images acquired by narrow bandwidth OCT devices suffer from axial resolution degradation and are difficult to support disease diagnosis. Deep learning-based methods can enhance the axial resolution of OCT images, but most methods focus on improving the model architecture, the potential of fully exploiting the fusion of spatial and frequency domain information for image reconstruction has not been fully explored. This paper proposes a Dual-branch Multidomain Feature Fusion Network (MDFNet). The core module of the model consists of a parallel Enhanced Multi-scale Spatial Feature module and an Auxiliary Frequcy Feature module to provide non-interfering dual-domain feature information to improve the reconstruction effect. MDFNet achieved the best performance in the tests of mouse retina and human retina datasets, outperforming the state-of-the-art (SOTA) algorithms by 0.11 dB and 0.18 dB respectively. In addition, the results of this method performed best in the retinal layer segmentation test.**

***Index Terms*—Axial super-resolution, multidomain feature fusion, Optical Coherence Tomography.**

I. Introduction

OCT is an imaging technique based on the principle of optical coherence. Due to its high resolution, non-invasive nature, and ability to produce three-dimensional images, OCT has become one of the most successful imaging technologies in clinical practice. It has shown important value in different fields [1], [2], [3]. Particularly, OCT technology can generate high-resolution retinal B-scan images along the depth of tissue, which are used to diagnose various retinal diseases [4], [5]. High-resolution OCT images, regarded as the gold standard for diagnosing ophthalmic diseases, place stringent demands on the axial resolution of the equipment. According to the $\Delta z = \frac{2\lambda^2 \ln 2}{\pi \Delta\lambda}$, increasing the spectral bandwidth $\Delta\lambda$ and utilizing shorter central wavelengths λ are two primary approaches to enhance resolution. For instance, Nishizawa et al. [6] achieved 3-micron axial resolution in air using a supercontinuum light source with a spectral bandwidth of 140 nm. Similarly, Song et al. developed VN-OCT with a visible light source, which demonstrated higher sensitivity in distinguishing between normal and suspect eyes [7]. However, upgrading hardware resources to achieve these improvements is associated with increased costs, limiting researchers' capacity for in-depth OCT technology development, such as image denoising and image segmentation [8], [9], [10].

In recent years, deep learning-based axial super-resolution for OCT images has made remarkable progress. Yuan et al. employed a generative adversarial network to extract asymmetric spatial features, achieving significant improvements in axial super-resolution [11]. Song et al. utilized a multi-residual block serial structure to deeply mine spatial features, producing outstanding results in fingerprint OCT image reconstruction [12]. Wang et al. designed an effective pure complex-valued network (CVSR) using phase and amplitude information from OCT signals to achieve axial super-resolution [13]. Similarly, Liang et al. applied Gaussian windowing to spectral bandwidth data for training a cGAN model, achieving notable results across multiple datasets [14]. Li et al. enhanced both axial and lateral resolution by adjusting scale factors to limit data quantities in the spatial and spectral domains, collecting data pairs for training [15]. Yao et al. also achieved high-resolution hyperspectral image reconstruction by leveraging cumulative attention blocks (CAB) to extract features from non-local spatial-spectral details [16].

Although the use of convolutional networks to extract features from the spatial domain has been widely studied and applied, high-frequency information in image reconstruction tasks can be easily masked by a large number of redundant spatial domain features [17]. Additionally, spatial-domain convolution relies on local receptive fields for feature extraction and cannot capture the periodic characteristics of signals across different frequencies as effectively as frequency-domain features [18]. When images are transformed into the frequency domain, the model can more easily distinguish between high- and low-frequency information [19], balancing their contributions to feature representation and mitigating issues of missing critical details. Consequently, dual-domain learning models decouple features along different dimensions, improving feature separability and image reconstruction performance, an approach that has shown positive impact across various fields [20], [21]. Based on these

Received 12 September 2024; revised 15 November 2024; accepted 23 November 2024. Date of publication 28 November 2024; date of current version 7 January 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62205181, in part by the Natural Science Foundation of Shandong Province under Grant ZR2022QF017, and in part by the Shandong Province Outstanding Youth Science Fund Project (Overseas) under Grant 2023HWYQ-023. The associate editor coordinating the review of this article and approving it for publication was Dr. Yu Liu. *(Corresponding author: Weiye Song.)*

The authors are with the School of Mechanical Engineering, and Key Laboratory of High Efficiency and Clean Mechanical Manufacture, Ministry of Education, Shandong University, Jinan 250061, China (e-mail: xuquanqing@mail.sdu.edu.cn; he_xiang@mail.sdu.edu.cn; ujnmhxu@hotmail.com; hukaixuan@mail.sdu.edu.cn; songweiye@sdu.edu.cn).

Digital Object Identifier 10.1109/LSP.2024.3509337

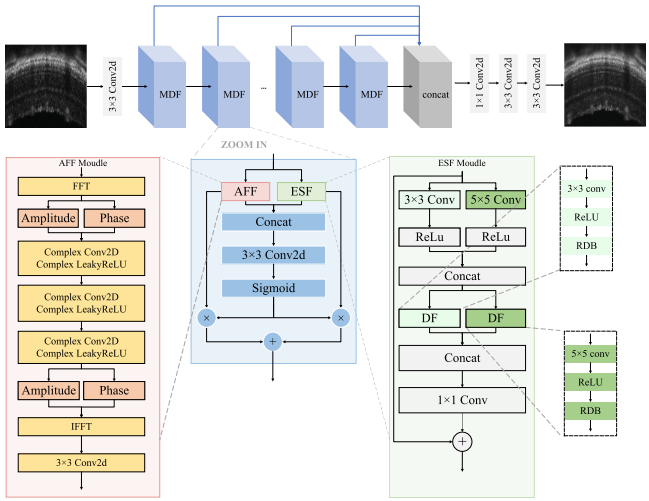

Fig. 1. The network architecture of the MDFNet.

observations, this paper proposes a Dual-branch Multidomain Feature Fusion Network(MDFNet) for axial super-resolution of OCT images. This method enhances the guidance of frequency-domain information during reconstruction by decoupling and aggregating spatial and frequency domain features, thereby overcoming the limitations of high-frequency information perception in the spatial domain. Furthermore, to avoid the limitations of single-scale feature extraction in the spatial domain, we employ a multi-scale feature extraction scheme, which balances fine-grained modeling with structural stability in images, and enhances model robustness across different resolutions—a technique used similarly in other research [22]. In summary, the main contributions of this paper are as follows:

a) The Enhanced Multi-scale Spatial Feature(ESF) module is used to further extract deep image feature information in the spatial domain under different receptive fields and achieve effective expression of spatial features.
b) The Auxiliary Frequency Feature(AFF) module converts image spatial information into frequency domain information and further extracts features from pixel signals decomposed into different frequencies, enriching reconstruction information.
c) The Feature Fusion module aggregates non-interfering spatial and frequency domain information and provides it to the reconstruction part.

II. Methods

A. Overall Framework

The overall structure of MDFNet is shown in Fig. 1. MDFNet consists of a shallow feature extraction part, a deep feature extraction part, and a reconstruction part. The input image $I \in \mathbb{R}^{H\times W\times 1}$ undergoes low-level feature extraction through 3×3 convolutional layers to obtain $F_0^{H\times W\times 64}$. The deep feature extraction component consists of N (where N=6) Dual-branch MDF blocks. The number of channels of all MDF modules is maintained at 64. Each deep feature F_i is channel-cascaded, where the MDF Block mainly extracts non-interfering information through the parallel operation of the ESF module and the AFF module and then aggregates the dual-domain information. The reconstruction part includes a 1×1 convolutional layer for channel compression and two 3×3 convolutional layers for reconstructing the image.

B. Mutildomain Feautre Extraction Blocks

MDF block employs a Dual-branch structure, consisting of a Enhanced Multi-scale Spatial Feature Module that includes multi-scale feature extraction and Auxiliary Frequcy Feature Module based on complex convolutions. The results from the parallel processing of the dual branches are fed into a feature fusion module for information integration, with the final output serving as the input for the subsequent module.

1) Enhanced Multi-Scale Spatial Feature Module: This work designs a ESF module that extracts and cross-fuses information at two scales, as shown in the ESF module of Fig. 1. First, the input image is processed in parallel through a 5×5 convolution layer and a 3×3 convolution layer (both using the ReLU activation function), and the resulting features are merged. The feature maps are then input into DF modules with receptive fields defined by convolution kernels of size $i\times i$ ($i=3,5$). The DF modules (as shown in Fig. 1, with receptive fields of 3 and 5 on the left and right, respectively) consist of a convolution layer with the corresponding receptive field, a ReLU activation function, and a 4-layer RDB (Residual Dense Block) module [23]. The output feature maps are merged and passed through a 1×1 convolution layer for bypass feature sharing, and the final output is obtained by residual addition. The specific operation process is described as follows:

$$f = \text{cat}(\text{ReLU}(\text{Conv}_3(I)), \text{ReLU}(\text{Conv}_5(I))) \quad (1)$$

$$F = I \oplus \text{Conv}_1(\text{cat}(F_{\text{DF3}}(f), F_{\text{DF5}}(f))) \quad (2)$$

where I is the input image, $Conv_k$ denotes a convolutional layer with a kernel size of k(k=1,3,5),ReLU represents the ReLU activation function, cat denotes channel concatenation,F_{DFi} denotes the feature extraction module with a receptive field of $i\times i$(i=3,5),and $\oplus$ represents residual addition.

2) Auxiliary Frequcy Feature Module: This module is designed to extract frequency features from images, addressing the issue of overly simplistic spatial information and providing more non-structural information for image reconstruction. The Frequency Domain Extraction Module primarily processes the amplitude and phase maps obtained from the Fourier transform using complex convolutions [24], as illustrated in the AFF module in Fig. 1. In this process, the input image is first transformed into the phase image I_P and the amplitude image I_A through the Fast Fourier Transform (FFT) and then converted into real values.

The two images are processed through three layers of complex convolutions and LeakyReLU activation functions, enhancing the extracted phase and amplitude information using complex convolutions, thereby enhancing the model's ability to learn frequency distributions in reconstruction tasks. During the complex convolution process, I_A and I_P are multiplied by the convolution kernels K_A and K_P, respectively, and then subtracted. Simultaneously, I_A and I_P are multiplied by the convolution kernels K_A and K_P, respectively, and then added. Finally, the two results are reshaped into complex numbers. The specific operation flow is as follows:

$$I * K = (I_A + jI_P) * (K_A + jK_P) \quad (3)$$

where I denotes the input image, K represents complex convolution,I_A denotes the amplitude image,I_P denotes the phase image,K_A denotes the real part convolution,K_P denotes the imaginary part convolution, and j represents the imaginary unit, satisfying $j^2 = -1$.

The enhanced frequency spectrum is then subjected to an inverse Fourier transform, and finally, a 3×3 convolution is applied to obtain the output feature map with frequency domain information.

3) Multi-Domain Feature Aggregation Module: This part first performs channel concatenation of the spatial features and frequency domain features. Then, a 3×3 convolutional layer is used to further refine the composite features, followed by a sigmoid activation function to obtain the result F_C. Finally, half of the feature channels in F_C are multiplied pixel-wise with F_s and F_f respectively, and then added pixel-wise. The resulting output is used as the input for the next part. The processing flow is as follows:

$$f = Sigmoid(Conv3(cat(F_{FF}, F_{SF}))) \tag{4}$$

$$F_c = F_f \otimes f(0 : N/2) \oplus F_s \otimes f(N/2 : N) \tag{5}$$

where, $\otimes$ denotes pixel-wise addition, and $\oplus$ denotes pixel-wise multiplication.

III. Experiment

A. Data Collection and Supplementary Details

The previously reported device [25] was used to collect mouse retinal OCT images, and the human retinal dataset was collected by a custom-built near-infrared OCT device (spectral bandwidth of 96 nm, central wavelength of 812 nm). All datasets include 1000 B-scan images from different eyes, and the datasets are divided into 8:1:1 ratios. All images are cropped to a size of 384×384 to remove unnecessary information. We use a Gaussian window to center-crop the original spectrum at a ratio of 1/2, 1/3, and 1/4 to simulate the reduction in optical resolution of the OCT system, forming grayscale images with different axial resolutions. All training was optimized using the Adam optimizer. The initial learning rate was set to 5e-5, and other parameters remained default. The batch size was set to 2, and the learning rate was halved every 20 epochs to promote convergence, for a total of 250 epochs. The L1 loss function was used as the training loss. Peak signal-to-noise ratio (PSNR), root mean square error (RMSE), and structural similarity index measure (SSIM) were used to evaluate network performance.

B. Experimental Results

We compared MDFNet with multiple super-resolution (SR) networks [23], [26], [27], [28], [29], [30], [31], [32], [33] to verify the model performance. The size of each image remained unchanged during training and testing, so the scaling factor in the upsampling layer of the above network was set to 1. In the mouse retina dataset, Fig. 2 shows the results of MDFNet compared with other methods. In the magnified rectangular range, except for Fig. 2(e)–(g), and (k), the other reconstruction results are blurred in the boundary area between RPE, OS and ELM and lack high-frequency details. Although the image quality of Fig. 2(e)–(g) has been improved, the ability to express the hierarchical structure has not yet reached the level of Fig. 2(k). Fig. 2(k) is closer to the real image in terms of the comprehensive performance of brightness, contrast and structural distortion model, and the degree of restoration in low-light areas is better.

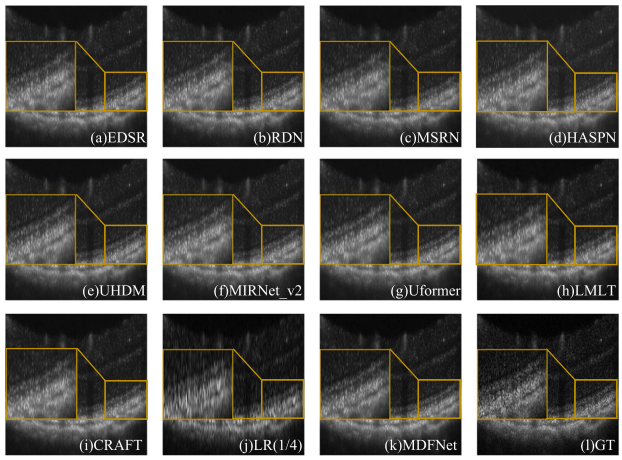

Fig. 2. Super-resolution results of the MDFNet on mouse retina datatest.

TABLE I
Evaluation Metrics Provided by Different Methods on the Mouse Retina Dataset

Method \ Metrics	1/2			1/3			1/4		
	PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑	RMSE↓
MSRN	31.1900	0.9014	5.5909	27.8896	0.7765	7.0905	26.5362	0.6836	7.4198
MIRNetV2	31.0157	0.8984	5.6362	27.8849	0.7738	6.8848	26.5790	0.6811	7.3236
Uformer	30.6525	0.8921	5.7443	27.6904	0.7664	6.9380	26.3810	0.6731	7.3700
UHDM	30.7522	0.8945	5.7198	27.7573	0.7695	6.9152	26.5450	0.6805	7.3319
EDSR	31.0759	0.8996	5.6335	27.8170	0.7736	6.9157	26.4519	0.6797	7.3677
RDN	31.2025	0.9018	5.5836	27.9245	0.7772	6.8810	26.5523	0.6857	7.3390
LMLT	30.7265	0.8932	5.7441	27.5864	0.7630	6.9776	26.1685	0.6611	7.4323
HASPN	29.9022	0.8233	6.1547	27.1618	0.6806	7.5115	25.5081	0.6054	8.1471
CRAFT	30.9931	0.8976	5.6508	26.1851	0.6794	7.3480	26.4646	0.6754	7.3535
MDFNet	**31.2586**	**0.9021**	**5.5826**	**27.9688**	**0.7787**	**6.8643**	**26.6591**	**0.6866**	**7.3115**

As shown in Table I, MDFNet achieves the highest PSNR and SSIM values and the lowest RMSE in objective evaluation metrics compared to other methods. It consistently ensures the best performance of these objective metrics on datasets with different axial resolutions. In addition, even on the 1/3 and 1/4 datasets where the image axial resolution is more blurred, MDFNet still maintains the best qualitative results.

The effectiveness of this method is further verified on the human retina dataset. In the enlarged rectangular frame, in Fig. 3(a)–(d), the boundary between the RPE layer and the OS layer is blurred, and the pixel transition speed between the retinal layers is inconsistent with the boundary change trend of the GT image. The image quality in Fig. 3(d) and (f) is improved, the reconstructed image is too smooth, the overall image is not clear enough, and the high-frequency details are missing. In addition, except for Fig. 3(d) and (k), the pixel signal of the OPL is lost in other results, resulting in the inability to observe the OPL layer. Although the OPL layer can be observed in Fig. 3(d), the image quality caused by excessive smoothing is not as good as that in Fig. 3(k). The OPL layer signal in Fig. 3(k) is well

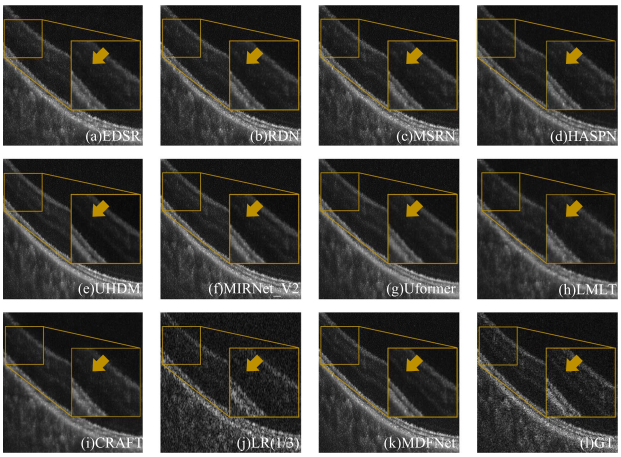

Fig. 3. Super-resolution results of the MDFNet on human retina datatest.

TABLE II
EVALUATION METRICS PROVIDED BY DIFFERENT METHODS ON THE HUMAN RETINA DATASET

Method \ Metrics	1/2			1/3			1/4		
	PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑	RMSE↓
MSRN	25.8295	0.7066	7.6940	23.2416	0.4491	8.4320	23.3269	0.4026	8.4797
MIRNetV2	25.8264	0.6992	7.7101	23.4597	0.4532	8.3975	23.4942	0.4046	8.4550
Uformer	25.3039	0.6628	7.8738	23.0675	0.4147	8.4674	23.2615	0.3864	8.4939
UHDM	25.5981	0.6818	7.7898	23.3665	0.4261	8.4355	23.4735	0.4004	8.4605
EDSR	25.6738	0.6982	7.7447	23.1354	0.4440	8.4523	23.1903	0.3965	8.5013
RDN	25.9172	0.7100	7.6684	23.2952	0.4515	8.4261	23.3458	0.4041	8.4792
LMLT	25.3798	0.6778	7.8276	22.8922	0.4177	8.5031	23.0102	0.3768	8.5370
HASPN	25.0786	0.6364	8.2272	21.8304	0.4018	9.1802	22.8280	0.3166	8.9115
CRAFT	25.6098	0.6884	7.7664	23.1773	0.4332	8.4553	23.2828	0.3909	8.4877
MDFNet	**26.0046**	**0.7125**	**7.6512**	**23.5809**	**0.4629**	**8.3765**	**23.4993**	**0.4061**	**8.4533**

TABLE III
ABLATION EXPERIMENTS ON 1/2 MOUSE DATASET

baseline	ESF	AFF	PSNR↑	SSIM↑	RMSE↓
✓			31.1900	0.9014	5.5909
✓	✓		31.1924	0.9016	5.5869
✓		✓	31.2053	0.9015	5.5886
✓	✓	✓	**31.2586**	**0.9021**	**5.5826**

restored, which is consistent with the structure in GT. In addition to qualitative results, the effect of image reconstruction can also be reflected by objective evaluation indicators. As shown in Table II, the MDFNet designed in this paper achieved the best performance in PSNR, RMSE and SSIM indicators.

IV. DISCUSSION

In the ablation experiments, we tested the baseline, the replacement with the improved ESF module, the addition of the AFF module, and the simultaneous addition of both the ESF and AFF modules. As shown in Table III, the evaluation metrics for each method showed improvement, with the best results achieved when both modules were added simultaneously.

We evaluate the performance of different reconstruction results in the retinal segmentation task. The test results of ReLayNet [34] and LightReSeg [10] show that the reconstruction results of MDFNet perform best and achieve the highest mean

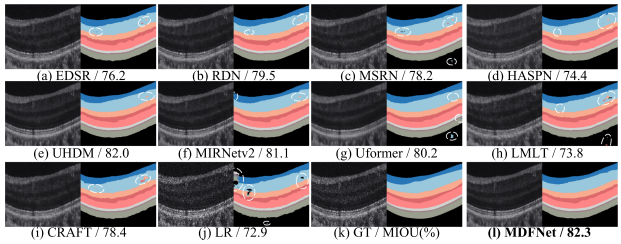

Fig. 4. Segmentation results of different reconstruction methods on ReLayNet. Objective metrics use the averaged MIOU.

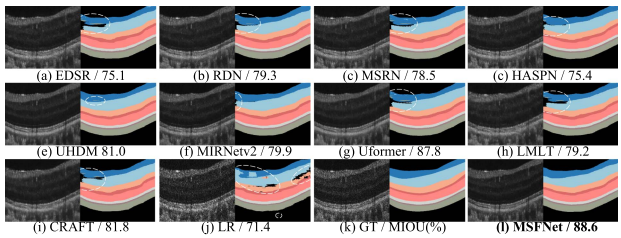

Fig. 5. Segmentation results of different reconstruction methods on LightReSeg. Objective metrics use the averaged MIOU.

intersection over union(MIOU), as shown in Figs. 4 and 5. We believe that the frequency domain information provides more unstructured information, which helps the network model learn the sharp edge signals between retinal layers to enhance the segmentation task.

MDFNet completes large image reconstruction with FLOPs of 3.21×10^{12} and takes 1.96 seconds to complete the inference of a single medical image reconstruction. The number of parameters used to store information such as weights and biases is 21.76×10^{6}.

V. CONCLUSION

This work proposes a Dual-branch feature fusion network (MDFNet) for achieving axial super-resolution in OCT imaging. The method uses a Dual-branch architecture to simultaneously provide spatial and frequency information, enhancing model performance by incorporating features from different dimensions. A multi-scale feature extraction module is employed for spatial feature extraction, while complex convolution is used to extract frequency-domain features from the phase and amplitude maps after Fourier transformation. The method is evaluated on two different OCT retina datasets, and the reconstruction results outperform other methods in terms of stability and detail clarity, with significant improvements in objective metrics. Additionally, MDFNet's performance remains excellent in segmentation tests. The computational complexity of the model has room for improvement. The computational burden may be caused by the deeper network structure and a large number of convolution operations. Therefore, methods such as depthwise separable convolutions or efficient transformers may offer more promising results. In the future, we plan to collect more disease images and explore powerful models that can reconstruct complex and abnormal structure images. In addition, integrating spectral information as prior knowledge of the model will also become an interesting research direction in the future.

REFERENCES

[1] M. Everett, S. Magazzeni, T. Schmoll, and M. Kempe, "Optical coherence tomography: From technology to applications in ophthalmology," *Transl. Biophotonics*, vol. 3, no. 1, 2021, Art. no. e202000012.

[2] H. G. Bezerra, M. A. Costa, G. Guagliumi, A. M. Rollins, and D. I. Simon, "Intracoronary optical coherence tomography: A comprehensive review: Clinical and research applications," *JACC: Cardiovasc. Interv.*, vol. 2, no. 11, pp. 1035–1046, 2009.

[3] J. Welzel, "Optical coherence tomography in dermatology: A review," *Skin Res. Technol.: Rev. Art.*, vol. 7, no. 1, pp. 1–9, 2001.

[4] N. Hassan, "OCT biomarkers in wet type age related macular degeneration," *Acta Ophthalmologica*, vol. 102, 2024, Art. no. 9994098.

[5] Y. Cui et al., "Quantitative assessment of OCT and OCTA parameters in diabetic retinopathy with and without macular edema: Single-center cross-sectional analysis," *Front. Endocrinol.*, vol. 14, 2024, Art. no. 1275200.

[6] N. Nishizawa and M. Yamanaka, "OCT: Ultrahigh resolution optical coherence tomography at visible to near-infrared wavelength region," in *Multidisciplinary Computational Anatomy: Toward Integration of Artificial Intelligence With MCA-Based Medicine*. Berlin, Germany: Springer, 2022, pp. 305–313.

[7] W. Song, S. Zhang, N. Sadlak, M. G. Fiorello, M. Desai, and J. Yi, "Visible and near-infrared optical coherence tomography (vnoct) in normal, suspect, and glaucomatous eyes," *Proc. SPIE*, vol. 11623, 2021, Art. no. 1162311.

[8] Q. Xie et al., "Multi-task generative adversarial network for retinal optical coherence tomography image denoising," *Phys. Med. Biol.*, vol. 68, no. 4, 2023, Art. no. 045002.

[9] X. He et al., "Exploiting multi-granularity visual features for retinal layer segmentation in human eyes," *Front. Bioeng. Biotechnol.*, vol. 11, 2023, Art. no. 1191803.

[10] X. He et al., "Light-weight retinal layer segmentation with global reasoning," *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 2520214.

[11] Z. Yuan, D. Yang, W. Wang, J. Zhao, and Y. Liang, "Self super-resolution of optical coherence tomography images based on deep learning," *Opt. Exp.*, vol. 31, no. 17, pp. 27566–27581, 2023.

[12] Z. Song, Y. Lin, L. Xiong, and Z. Li, "Super-resolution algorithm for the characterization of sweat glands in fingerprint OCT images," *J. Opt. Soc. Amer. A*, vol. 40, no. 11, 2023, Art. no. 2520214.

[13] L. Wang, S. Chen, L. Liu, X. Yin, G. Shi, and J. Mo, "Axial super-resolution optical coherence tomography via complex-valued network," *Phys. Med. Biol.*, vol. 68, no. 23, 2023, Art. no. 235016.

[14] K. Liang et al., "Resolution enhancement and realistic speckle recovery with generative adversarial modeling of micro-optical coherence tomography," *Biomed. Opt. Exp.*, vol. 11, no. 12, pp. 7236–7252, 2020.

[15] X. Li et al., "Multi-scale reconstruction of undersampled spectral-spatial OCT data for coronary imaging using deep learning," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 12, pp. 3667–3677, Dec. 2022.

[16] Z. Yao, S. Liu, X. Yuan, and L. Fang, "Specat: Spatial-spectral cumulative-attention transformer for high-resolution hyperspectral image reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 25368–25377.

[17] Z. Zhou, A. He, Y. Wu, R. Yao, X. Xie, and T. Li, "Spatial-frequency dual progressive attention network for medical image segmentation," 2024, *arXiv:2406.07952* .

[18] S. Zhou, X. Duan, and J. Zhou, "Human pose estimation based on frequency domain and attention module," *Neurocomputing*, vol. 604, 2024, Art. no. 128318.

[19] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.

[20] Y. Xiao, Q. Yuan, K. Jiang, Y. Chen, Q. Zhang, and C.-W. Lin, "Frequency-assisted Mamba for remote sensing image super-resolution," 2024, *arXiv:2405.04964*.

[21] X. Li, Z. Dong, H. Liu, J. J. Kang-Mieler, Y. Ling, and Y. Gan, "Frequency-aware optical coherence tomography image super-resolution via conditional generative adversarial neural network," *Biomed. Opt. Exp.*, vol. 14, no. 10, pp. 5148–5161, 2023.

[22] K. Jiang et al., "Multi-scale progressive fusion network for single image deraining," in *Proc. 2020 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8343–8352.

[23] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2018, pp. 2472–2481.

[24] C. Trabelsi et al., "Deep complex networks," 2017, *arXiv:1705.09792*.

[25] W. Yi, K. Hu, Y. Wan, F. Wu, and W. Song, "A high-speed near-infrared optical coherence tomography angiography system for mouse retina," *J. Lumin.*, vol. 270, 2024, Art. no. 120550.

[26] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 136–144.

[27] R. Davis, B. Lang, and N. Gautam, "Modeling utilitarian-hedonic dual mediation (UHDM) in the purchase and use of games," *Internet Res.*, vol. 23, no. 2, pp. 229–256, 2013.

[28] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general u-shaped transformer for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 17683–17693.

[29] S. W. Zamir et al., "Learning enriched features for fast image restoration and enhancement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 1934–1948, Feb. 2023.

[30] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 517–532.

[31] A. Li, L. Zhang, Y. Liu, and C. Zhu, "Feature modulation transformer: Cross-refinement of global representation via high-frequency prior for image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 12514–12524.

[32] J. Kim, J. Nang, and J. Choe, "LMLT: Low-to-high multi-level vision transformer for image super-resolution," 2024, *arXiv:2409.03516*.

[33] Z. Guo and Z. Zhao, "Hybrid attention structure preserving network for reconstruction of under-sampled OCT images," 2024, *arXiv:2406.00279*.

[34] A. G. Roy et al., "Relaynet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed. Opt. Exp.*, vol. 8, no. 8, pp. 3627–3642, 2017.