FA-Diff: Leveraging Frequency-Aware RWKV Diffusion for OCT to Correlation-Stability Map Translation

Chuanhao Zhang[1], Yangxi Li[2], Jianping Song[3], Yingni Wang[1], Haojie Han[1], Canhong Xiang[3], Guochen Ning[1], Fang Chen[2], and Hongen Liao[1,2](✉)

[1] School of Biomedical Engineering, Tsinghua Medicine, Tsinghua University, Beijing, China
liao@tsinghua.edu.cn
[2] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
[3] Hepatopancreatobiliary Center, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing, China

Abstract. Optical coherence tomography (OCT) and its derived correlation stability (CS) maps provide high-resolution structural imaging and elastography information at the micrometer scale, which are critical for the rapid and accurate detection of malignant biliary strictures. However, acquiring intraoperative CS maps remains challenging due to the need for complex mechanical compression and the high costs associated with hardware modifications in conventional OCT systems. Moreover, existing deep learning methods for OCT-to-CS map translation often lack long-range modeling capability for recovering subtle deformation features and are limited in modeling the relative importance of different frequency components, which is essential for localizing high-frequency stiffness details. To address these limitations, we propose FA-Diff, a novel RWKV-based frequency-aware diffusion framework for elastography mapping. FA-Diff integrates a frequency-aware multi-granularity shift (FG-Shift) module and a frequency-adaptive sub-band normalization (FASN) mechanism to enhance elasticity-aware representation learning. Specifically, FG-Shift is incorporated into the RWKV architecture as a spatial-frequency adaptive scanning strategy to improve the modeling of vertical stiffness textures and enable effective global coarse-to-fine contextual aggregation. Meanwhile, FASN dynamically adjusts normalization parameters using inter-stage states to adaptively modulate wavelet sub-band coefficients. Comprehensive experiments on the clinical cholangiocarcinoma dataset demonstrate that FA-Diff consistently outperforms state-of-the-art CNN, Transformer, Mamba, and RWKV-based baselines, achieving superior reconstruction accuracy and perceptual quality.

Keywords: OCT to Stiffness Translation · Diffusion · RWKV.

1 Introduction

Optical coherence tomography (OCT) offers micrometer-scale structural imaging in distinguishing malignant biliary strictures. However, OCT B-scans alone exhibit limited sensitivity for malignancy detection [12]. Given the elevated stiffness of cholangiocarcinoma (CCA), quantitative elastography has demonstrated effectiveness in delineating malignant margins [14]. Following Zaitsev et al. [21], we adopt the correlation-stability (CS) map to characterize tissue biomechanical properties.

Despite its effectiveness, intraoperative acquisition of CS maps from OCT remains challenging. Accurate stiffness estimation typically requires localized mechanical compression, which is time-consuming and disruptive to surgical workflows [22]. Recent advances in deep learning-based medical image translation provide a promising alternative for virtual generation of quantitative tissue properties [23, 20, 24, 13, 7, 2]. While several studies estimate elastic wave velocity from phase-resolved OCT data, they do not operate directly on structural OCT B-scans [13, 23]. In contrast, we formulate OCT elastography as an image translation problem, enabling direct synthesis of CS maps from paired OCT–CS map supervision without additional hardware.

From an architectural perspective, existing CNN, Transformer and diffusion models tend to overemphasize low-frequency structural information, limiting their ability to model continuous fine-grained deformation patterns critical for elasticity modeling [10, 8, 1, 16]. Some works [24, 2] incorporate frequency-aware mechanisms to enhance localized high-frequency deformation details; however, they remain limited in capturing global structural context, which is essential for elasticity-related image translation. Although recent sequence models such as Mamba offer efficient global modeling capabilities, their reliance on sequence flattening disrupts inherent two-dimensional spatial dependencies [16]. In contrast, Receptance Weighted Key Value (RWKV) models provide stronger global contextual modeling while better preserving local spatial continuity, making them well suited for medical image translation [10, 8, 1].

Motivated by these observations, we propose the **F**requency-**A**ware RWKV **Diff**usion (FA-Diff) for OCT-to-CS map translation. FA-Diff explicitly integrates frequency-domain guidance into diffusion-based RWKV modeling to enhance elasticity-aware representation learning. A wavelet-guided stiffness interaction (WSI) module is introduced to facilitate effective fusion of low- and high-frequency information during the diffusion process. In addition, we design a frequency-aware multi-granularity shift (FG-Shift) to mimic tissue compression patterns and strengthen RWKV modeling of elastic geometric structures. We incorporate frequency-adaptive sub-band normalization (FASN), enabling adaptive modulation and fusion of wavelet sub-bands.

The main contributions of this work are as follows: (1) We propose FA-Diff, a frequency-aware RWKV-based diffusion framework for OCT-to-CS map translation. To the best of our knowledge, this is the first exploration of RWKV architectures for virtual OCT-based elasticity generation. (2) Extensive evaluation on a clinical CCA dataset demonstrates superior qualitative and quanti-

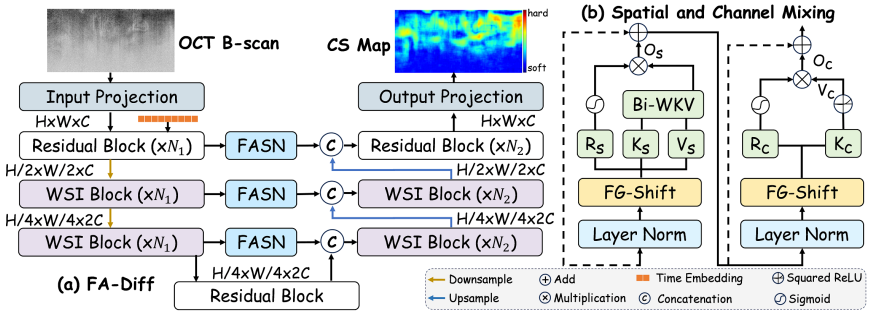

Fig. 1. Overview of the FA-Diff architecture. (a) U-Net based framework with WSI blocks in both encoder and decoder, and FASN modules for cross-level feature connection. (b) RWKV-style spatial and channel mixing within the WSI block.

tative performance, achieving 22% LPIPS improvement with 24% reduction in complexity.

2 Methods

An overview of FA-Diff is illustrated in Fig. 1. FA-Diff adopts an encoder-decoder architecture composed of multiple WSI blocks and residual blocks. Each WSI block incorporates the proposed FG-Shift, which integrates multi-level discrete wavelet transform (DWT) [2] with novel scanning scheme to enable adaptive coarse-to-fine interactions across spatial and frequency domains. In addition, FASN dynamically aggregates multi-stage features from all encoder levels, thereby enhancing tissue contrast and improving feature representation in the decoder.

First, the input projection applies down-sampling convolution and latent-space transformation to reduce computational cost. Given a paired B-scan and CS image inputs $I_{in}, I_{cs} \in \mathbb{R}^{H\times W\times C}$, where $H \times W$ indicates spatial resolution and C is the feature channel. First, the input convolution applies a 3×3 convolution followed by SiLU, and then a 3×3 convolution with stride 2 for down-sampling. Meanwhile, the inputs are passed through a pre-trained VQ-GAN [4] encoder to obtain latent embeddings x_0,y_0.

We perturb the latent residuals with Gaussian noise according to the diffusion timestep t, following [20]. Concretely, the diffusion process can be described as:

$$q(x_t|x_{t-1}, y_0) = \mathcal{N}(x_t; x_{t-1} + \alpha_t(y_0 - x_0), \kappa^2\alpha_t\mathbf{I}), t = 1, \ldots, T \quad (1)$$

where α_t is predefined according to noise schedule [20], κ is the hyper-parameter controlling the noise variance, and $\mathbf{I}$ is the identity matrix. While the output projection symmetrically reconstructs the output via convolutional layers and a VQGAN decoder. Here, we directly use the denoising diffusion probabilistic model (DDPM) residual block with time embedding in Fig.1 for residual diffusion [20].

2.1 Wavelet-Guided Stiffness Interaction (WSI)

1) **Spatial Mix and Channel Mix**: Inspired by RWKV-based models [8], the WSI block consists of a residual block followed by spatial and channel mixing modules (Fig.1). The residual output X_r is flattened into token sequences and fed into LayerNorm (LN) to produce $X_f \in \mathbb{R}^{T\times C}$ with $T = H \times W$, enabling spatial mixing and channel mixing to model long-range dependencies.

Specifically, the linear projections W_R, W_K and W_V produce receptance $R_s = FG(X_f)W_R$, key $K_s = FG(X_f)W_K$, and value $V_s = FG(X_f)W_V$, respectively. The FG-Shift operation $FG(\cdot)$ is applied in both the spatial mixing and channel mixing processes. The final output can be expressed as:

$$O_s = (\phi(R_s) \cdot \text{wkv}(K_s, V_s))W_o; O_c = (\phi(R_c) \cdot V_c)W_o \tag{2}$$

$$R_c = FG(X_c)W_R, K_c = FG(X_c)W_K, V_c = \gamma(K_c)W_v, X_c = \text{LN}(O_s + X_r) \tag{3}$$

where, wkv$(\cdot)$ is the bidirectional attention mechanism (Bi-WKV) [10], $\phi(\cdot)$ is the Sigmoid and W_o is the linear projection layer. The resulting features, with a residual connection, are then processed by channel mixing. $\gamma(\cdot)$ denotes the squared ReLU, W_v and W_o are the linear projection layers. O_s and O_c are the outputs of spatial mixing and channel mixing, respectively.

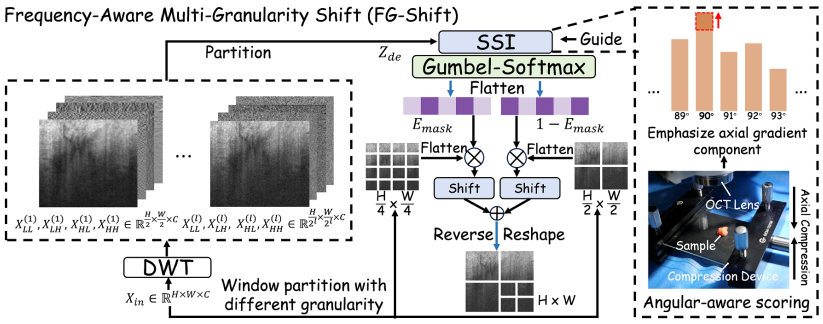

Fig. 2. Illustration of FG-Shift. Inputs are hierarchically partitioned into bi-level window quadrants from coarse to fine via the wavelet guidance.

2) **Frequency-aware multi-granularity shift (FG-Shift)**: Inspired by the granularity-aware scanning strategy [8, 10], the input X_{in} is partitioned into coarse sub-windows ($N_1 = 4$) and fine sub-windows ($N_2 = 16$) to capture spatial locality from coarse to fine. Specifically, a coarse partition is first applied to the DWT components $Z_{de} = \{X_{LL}^{(l)}, X_{LH}^{(l)}, X_{HL}^{(l)}, X_{HH}^{(l)}, X_{LL}^{(l-1)}, ..., X_{HH}^{(1)}\}$ with l levels, which are generated by multi-level DWT decomposition [2]. Subsequently, the SSI module is employed to quantify the richness of fine-grained features. The partitioning process is formulated as:

$$\left\{s_{N_1}^i | i = 0, 1, 2, 3\right\} = SSI\left(\text{Partition}(Z_{de}|Z_{N_1}^i)\right), \tag{4}$$

$$\left\{ Z_{N_2}^{(k,j)} | j = 0, 1, 2, 3 \right\} = \text{Partition}(Z_{N_1}^{k}) \tag{5}$$

where, $Z_{N_1}^{i}$ and $Z_{N_2}^{j}$ denote coarse- and fine-scale sub-windows, respectively. The TopK algorithm is applied to $s_{N_1}^{i}$ to select the quadrant with the highest energy, where $K = 1$. The selected coarse sub-window $Z_{N_1}^{k}$ is then further divided to obtain fine-grained sub-windows $Z_{N_2}^{(k,j)}$. Then, we use Gumbel-Softmax along with upsampling $\uparrow(\cdot)$ to construct the sequence mask E_{mask}:

$$E_{mask} = \uparrow (\text{Gumbel} - \text{Softmax}(\{s_{N_1}^{i} | i = 0, 1, 2, 3\})) \tag{6}$$

Element-wise multiplication with E_{mask} is then applied for index selection. The quad-directional token shift scanning [3] is applied once to the coarse sequences and twice to the fine sequences for coarse to fine scanning. Finally, the aggregated fine-grained partitioning result is obtained via the reverse transformation [10].

The angular-aware sparse selective interaction (SSI) is designed to mimic the axial compression process and guide elasticity reconstruction, as shown in Fig.2. SSI computes frequency importance by aggregating low- and high-frequency energy responses across multiple pyramid levels. At each level, channel-wise frequency energy is extracted and up-sampled to a common spatial resolution, and then combined with vertical angular responses. The SSI output $O_{ss} \in \mathbb{R}^{H \times W \times C_1}$ is defined as:

$$E(X) = \sum_{i=1}^{l} Softmax(\uparrow (\sum_{c} (X^{(i)})^2)), \tag{7}$$

$$O_{ss} = E(X_{LL})[1 + \tanh(E(X_{LH} + X_{HL} + X_{HH}) \sin(\pi/2) * s_y)] \tag{8}$$

where $*$ denotes the convolution, s_y is the Sobel operator used to enhance vertical angular responses, $E(\cdot)$ represents the energy mapping over C channels and l DWT levels. X_{LL}, X_{LH}, X_{HL} and X_{HH} denote the unions of all decomposition levels $\{X_{LL}^{(i)}\}$, $\{X_{LH}^{(i)}\}$, $\{X_{HL}^{(i)}\}$ and $\{X_{HH}^{(i)}\}$, respectively.

2.2 Frequency-Adaptive Sub-band Normalization (FASN)

FASN is proposed to dynamically modulate wavelet sub-bands via inter-stage states, thereby preserving fine-grained elasticity textures. Given the encoder output $M_t \in \mathbb{R}^{H_t \times W_t \times C_t}$ at stage $t \in 1, 2, 3$, we first decompose M_t into four wavelet sub-bands $\{X_k\}_{k=1}^{4}$ using DWT. For each sub-band X_k, global average pooling (GAP) is applied to extract a one-dimensional vector, which is then zero-padded to match the maximum channel dimension for dimensionality matching. The resulting vectors are stacked along the state dimension and fed into a multi-layer perceptron (MLP) with a tanh activation function, which is employed to yield the modulation descriptor X_k^{1D}.

We further introduce a variance-based modulation mechanism guided by frequency statistics to jointly model low- and high-frequency responses. For each sub-band, the modulation is defined as:

$$\hat{X}_k = \frac{X_k - \mu}{\sigma} \left(1 + \tanh\left(MLP(X_k^{1D})\right)\right) + \beta \tag{9}$$

where, μ and σ denote the mean and standard deviation of input feature X_k respectively, and β is a learnable shift parameter. Finally, the inverse DWT is adopted to $\{\hat{X}_k\}_{k=1}^4$ to reconstruct the features in the spatial domain.

3 Experiments

1) **Clinical Dataset**: The utilized OCT system with compression configuration [21, 6] incorporates a swept light source (SL131090, Thorlabs) paired with an interferometer (INT-MSI-1300B, Thorlabs). The imaging is accomplished by raster scanning the OCT beam across the sample using an objective lens with galvanometers (GVS002, Thorlabs). Using this system, we constructed a clinical dataset consisting of 10,834 paired OCT B-scans and CS maps acquired from 23 patients, with a spatial resolution of 512×512.
2) **Implementation Details**: FA-Diff was implemented using PyTorch and trained on an NVIDIA RTX 8000 GPU. We employed the Adam optimizer with a batch size of 64 and a crop size of 256×256. The model was trained for 50 epochs with an initial learning rate of 0.001. For the diffusion process, the number of sampling steps was set to $T = 4$ [20]. Optimization was performed using the composite loss function described in [8].
3) **Experimental Setup**: We evaluated model performance using widely adopted image quality assessment metrics, including PSNR, SSIM, LPIPS, and MS-SSIM [8, 17]. The average results from the 5-fold cross-validation are presented below. Data were split at the patient level; each fold included distinct patients during experiments. FA-Diff was compared with representative image translation methods, including BCI [11] , Li et al. [9], Yao et al. [19], and Ho et al. [5], as well as diffusion-based approaches such as WF-Diff [24] and ResShift [20]. In addition, the Mamba-based method WalMaFa [16] and RWKV-based models [18, 8] were included for comparison. As no existing methods are directly designed for OCT-to-CS map translation, we adapted all comparison models by modifying their input and output convolution accordingly.

3.1 Experiment Results

The quantitative comparison results presented in Table 1 demonstrate the superior performance of our method. FA-Diff consistently achieves the best results across all evaluation metrics, outperforming CNN, Transformer, Mamba, diffusion, and RWKV-based approaches in terms of PSNR, SSIM, MS-SSIM, and LPIPS under 5-fold cross-validation.

In particular, the notable improvements in SSIM and LPIPS indicate the model's enhanced capability to preserve structural contrast while effectively suppressing reconstruction artifacts, highlighting the effectiveness of the proposed FG-Shift and FASN modules. FS-RWKV achieves the second-best SSIM score,

Table 1. Quantitative comparison on the dataset. The highest scores are highlighted in bold. Suboptimal results are highlighted in underline.

Models	SSIM ↑	MS-SSIM ↑	PSNR ↑	LPIPS ↓	FLOPs	Params
BCI [11]	0.853 ± 0.04	0.857 ± 0.04	23.84 ± 2.45	0.173 ± 0.03	$224.99G$	$73.30M$
Li et al. [9]	0.611 ± 0.06	0.581 ± 0.07	18.40 ± 2.27	0.326 ± 0.04	$49.62G$	$11.19M$
Yao et al. [19]	0.846 ± 0.04	0.860 ± 0.04	23.93 ± 2.40	$\underline{0.113} \pm 0.03$	$18.15G$	$57.18M$
Ho et al. [5]	0.632 ± 0.05	0.573 ± 0.05	17.65 ± 1.68	0.220 ± 0.03	$103.32G$	$17.53M$
WF-Diff [24]	0.849 ± 0.04	0.857 ± 0.04	$\underline{24.08} \pm 2.46$	0.162 ± 0.03	$122.85G$	$57.28M$
ResShift [20]	0.856 ± 0.04	0.858 ± 0.04	23.57 ± 2.54	0.133 ± 0.03	$45.73G$	$107.19M$
WalMaFa [16]	0.804 ± 0.03	0.802 ± 0.04	22.44 ± 1.89	0.183 ± 0.03	$17.35G$	$16.09M$
Yang et al. [18]	0.859 ± 0.04	$\underline{0.862} \pm 0.04$	23.65 ± 2.53	0.164 ± 0.03	$22.29G$	$17.36M$
FS-RWKV [8]	$\underline{0.860} \pm 0.04$	0.856 ± 0.04	23.51 ± 2.54	0.131 ± 0.03	$46.25G$	$55.20M$
Ours	$\mathbf{0.874} \pm 0.03$	$\mathbf{0.878} \pm 0.03$	$\mathbf{24.16} \pm 2.66$	$\mathbf{0.101} \pm 0.02$	$34.77G$	$46.12M$

suggesting its ability to reconstruct texture-rich details; however, its performance remains inferior in terms of PSNR and overall perceptual quality. The generative model [19] exhibits higher MS-SSIM and lower LPIPS scores than the diffusion-based methods, indicating stronger structural consistency but limited fidelity. Compared with the recent RWKV-based model [8], our approach achieves an improvement of 0.6 dB in PSNR and a reduction of 0.03 in LPIPS, while simultaneously reducing the number of model parameters by 16%. Besides, we use paired student's t-test to evaluate the significant difference between FA-Diff and other models [15]. Under the 5% significance level, the metrics of each instance are statistically significant. The inference time of FA-Diff is $216ms$ per frame, enabling potential real-time intraoperative application.

Table 2. Ablation study on FASN and FG-Shift modules.

FASN	FG-Shift	SSIM ↑	MS-SSIM ↑	PSNR ↑	LPIPS ↓	FLOPs	Params
		0.848 ± 0.04	0.850 ± 0.05	22.91 ± 2.71	0.120 ± 0.03	$31.34G$	$36.43M$
✓		0.860 ± 0.04	0.861 ± 0.04	23.09 ± 2.41	0.118 ± 0.05	$31.35G$	$40.79M$
	✓	0.868 ± 0.03	0.869 ± 0.04	24.15 ± 2.54	0.114 ± 0.03	$34.76G$	$41.73M$

Beyond quantitative metrics, FA-Diff more effectively preserves elasticity-related structures, as illustrated in Fig. 3. It exhibits the lowest error variation among the compared methods and achieves higher reconstruction accuracy, particularly at the transitions between stiff and compliant regions.

3.2 Ablation Study

Table 2 presents the ablation results of the proposed FASN and FG-Shift modules. Removing either component results in consistent performance degradation

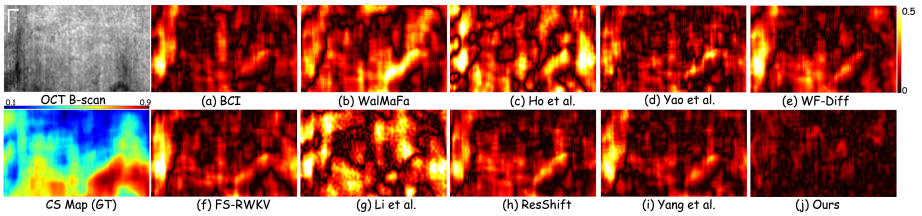

Fig. 3. Comparison of OCT-to-CS map translation methods. (a)-(j) show the reconstruction error maps, where higher brightness indicates larger MAE relative to the GT. Our method achieves markedly clearer and more continuous stiffness representations, effectively preserving fine structural details while substantially reducing artifacts that degrade the delineation performance of competing approaches. Scale bars: $200\mu m$.

across all evaluation metrics, including SSIM, MS-SSIM, PSNR, and LPIPS, indicating their complementary contributions. When FASN is removed, naïve skip connections implemented via feature concatenation are used to maintain structural continuity.

FG-Shift enhances stiffness-related texture modeling by facilitating global coarse-to-fine contextual aggregation, whereas FASN adaptively modulates wavelet sub-band responses through inter-stage, state-aware normalization. The full model achieves the best overall performance, demonstrating that effective global contextual modeling and fine-scale elasticity detail preservation are mutually reinforcing.

Table 3. Ablation study of RWKV scanning schemes.

Models	SSIM ↑	MS-SSIM ↑	PSNR ↑	LPIPS ↓	FLOPs	Params
Zigzag [1]	0.856 ± 0.04	0.857 ± 0.04	23.46 ± 2.51	0.129 ± 0.03	$35.10G$	$43.11M$
FSO-Shift [8]	0.860 ± 0.04	0.861 ± 0.04	23.31 ± 2.42	0.123 ± 0.03	$35.14G$	$43.29M$

Table 3 summarizes the ablation study of different scanning schemes. We compare the proposed FG-Shift with existing strategies, including Zigzag [1] and FSO-Shift [8]. The results show that replacing conventional sweep-based scanning with FG-Shift leads to a substantial performance improvement, yielding a 0.69 dB gain in PSNR while requiring fewer learnable parameters.

4 Conclusion

We propose FA-Diff, a frequency-aware RWKV-based diffusion framework for OCT-to-CS map translation that integrates the proposed FG-Shift and FASN modules to jointly enhance stiffness-sensitive feature modeling and fine-grained structural detail preservation, both of which are essential for cholangiocarcinoma

assessment. FA-Diff provides an alternative way for virtual optical coherence elastography, enabling direct inference of tissue elasticity from structural OCT images without additional hardware. Despite these advances, further validation across multi-center cohorts and heterogeneous imaging devices is required. Future work will focus on improving domain generalization, disease classification, and computational efficiency to support real-time deployment.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (U22A2051, 62271246), the National Key Research and Development Program of China (2022YFC2405200), the Science and Technology Commission of Shanghai Municipality (25ZR1402225, 24511104100), the National High Level Hospital Clinical Research Funding (2025-PUMCH-A-158).

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

References

1. Chen, T., Zhou, X., Tan, Z., Wu, Y., Wang, Z., Ye, Z., Gong, T., Chu, Q., Yu, N., Lu, L.: Zig-rir: Zigzag rwkv-in-rwkv for efficient medical image segmentation. IEEE Transactions on Medical Imaging (2025)
2. Du, P., Li, H., Xu, H., Jeon, P.B., Lee, D., Ji, D., Yang, R., Zhu, F.: Diffusion transformer meets multi-level wavelet spectrum for single image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 19700–19710 (2025)
3. Duan, Y., Wang, W., Chen, Z., Zhu, X., Lu, L., Lu, T., Qiao, Y., Li, H., Dai, J., Wang, W.: Vision-rwkv: Efficient and scalable visual perception with rwkv-like architectures. In: International Conference on Learning Representations. vol. 2025, pp. 83166–83182 (2025)
4. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12873–12883 (2021)
5. Ho, M.Y., Wu, C.M., Wu, M.S., Tseng, Y.J.: Every pixel has its moments: Ultra-high-resolution unpaired image-to-image translation via dense normalization. In: European Conference on Computer Vision. pp. 312–328. Springer (2024)
6. Kennedy, K.M., Chin, L., McLaughlin, R.A., Latham, B., Saunders, C.M., Sampson, D.D., Kennedy, B.F.: Quantitative micro-elastography: imaging of tissue elasticity using compression optical coherence elastography. Scientific Reports **5**(1), 15538 (2015)
7. Khosravi, P., Han, K., Wu, A.T., Rezvani, A., Feng, Z., Xie, X.: Xoct: Enhancing oct to octa translation via cross-dimensional supervised multi-scale feature learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 695–705. Springer (2025)
8. Lei, Y., Li, Z., Liu, Y., Chen, X., Pun, C.M.: Fs-rwkv: Leveraging frequency spatial-aware rwkv for 3t-to-7t mri translation. In: 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1–6. IEEE (2025)
9. Li, F., Hu, Z., Chen, W., Kak, A.: Adaptive supervised patchnce loss for learning h&e-to-ihc stain translation with inconsistent groundtruth image pairs. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 632–641. Springer (2023)

10. Li, X., He, X., Hu, T., Zhang, J., Zhou, M., Xie, C., Wang, Y., Huang, B.: Freq-rwkv: Granularity-aware spatial-frequency synergy via dual-domain recurrent scanning for pan-sharpening. In: Proceedings of the 33rd ACM International Conference on Multimedia. pp. 1890–1899 (2025)
11. Liu, S., Zhu, C., Xu, F., Jia, X., Shi, Z., Jin, M.: Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1815–1824 (2022)
12. Martinez, N.S., Trindade, A.J., Sejpal, D.V.: Determining the indeterminate biliary stricture: cholangioscopy and beyond. Current Gastroenterology Reports **22**(12), 58 (2020)
13. Mieling, R., Neidhardt, M., Behrendt, F., Latus, S., Heinemann, A., Ondruschka, B., Schlaefer, A.: A-scan sequence transformers for palpation with optical coherence elastography. Biomedical Optics Express **16**(5), 1925–1943 (2025)
14. Özdemir, M., Koç, U., Gökhan, M.B., Beşler, M.S.: Unveiling the potential of strain elastography in perihilar cholangiocarcinoma biopsies. Abdominal Radiology **49**(9), 3143–3148 (2024)
15. Sun, L., Wang, H., Xie, Q., Zheng, Y., Meng, D.: Encoding sampling pattern for robust and generalized mri reconstruction. Pattern Recognition p. 111772 (2025)
16. Tan, J., Pei, S., Qin, W., Fu, B., Li, X., Huang, L.: Wavelet-based mamba with fourier adjustment for low-light image enhancement. In: Proceedings of the Asian Conference on Computer Vision. pp. 3449–3464 (2024)
17. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003. vol. 2, pp. 1398–1402 (2003)
18. Yang, Z., Li, J., Zhang, H., Zhao, D., Wei, B., Xu, Y.: Restore-rwkv: Efficient and effective medical image restoration with rwkv. IEEE Journal of Biomedical and Health Informatics (2025)
19. Yao, Z., Luo, T., Dong, Y., Jia, X., Deng, Y., Wu, G., Zhu, Y., Zhang, J., Liu, J., Yang, L., et al.: Virtual elastography ultrasound via generative adversarial network for breast cancer diagnosis. Nature Communications **14**(1), 788 (2023)
20. Yue, Z., Wang, J., Loy, C.C.: Efficient diffusion model for image restoration by residual shifting. IEEE Transactions on Pattern Analysis and Machine Intelligence (2024)
21. Zaitsev, V.Y., Matveev, L., Gelikonov, G., Matveyev, A., Gelikonov, V.: A correlation-stability approach to elasticity mapping in optical coherence tomography. Laser Physics Letters **10**(6), 065601 (2013)
22. Zaitsev, V.Y., Matveyev, A.L., Matveev, L.A., Sovetsky, A.A., Hepburn, M.S., Mowla, A., Kennedy, B.F.: Strain and elasticity imaging in compression optical coherence elastography: the two-decade perspective and recent advances. Journal of Biophotonics **14**(2), e202000257 (2021)
23. Zhang, Y., Liao, J., Feng, Z., Yang, W., Perelli, A., Wang, Z., Li, C., Huang, Z.: Vp-net: an end-to-end deep learning network for elastic wave velocity prediction in human skin in vivo using optical coherence elastography. Frontiers in Bioengineering and Biotechnology **12**, 1465823 (2024)
24. Zhao, C., Cai, W., Dong, C., Hu, C.: Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8281–8291 (2024)