

Sparse Reconstruction of Optical Doppler Tomography with Alternative State Space Model and Attention

Zhenghong Li, Jiaxiang Ren, Wensheng Cheng, Yanzuo Liu, Congwu Du, Yingtian Pan, and Haibin Ling

Stony Brook University
{zhenghli, hling}@cs.stonybrook.edu

Abstract. Optical coherence Doppler tomography (ODT) is an emerging blood flow imaging technique. The fundamental unit of ODT is the 1D depth-resolved trace named raw A-scans (or A-line). A 2D ODT image (B-scan) is formed by reconstructing a cross-sectional flow image via Doppler phase-subtraction of raw A-scans along B-line. To obtain a high-fidelity B-scan, densely sampled A-scans are required currently, leading to prolonged scanning time and increased storage demands. Addressing this issue, we propose a novel sparse ODT reconstruction framework with an Alternative State Space Attention Network (ASSAN) that effectively reduces raw A-scans needed. Inspired by the distinct distributions of information along A-line and B-line, ASSAN applies 1D State Space Model (SSM) to each A-line to learn the intra-A-scan representation, while using 1D gated self-attention along B-line to capture the inter-A-scan features. In addition, an effective feedforward network based on sequential 1D convolutions along different axes is employed to enhance the local feature. In validation experiments on real animal data, ASSAN shows clear effectiveness in the reconstruction in comparison with state-of-the-art reconstruction methods. The source code is available at: <https://github.com/ZhenghLi/ASSAN>.

Keywords: Optical Doppler Tomography · Sparse Reconstruction · State Space Model.

1 Introduction

Optical Doppler Tomography (ODT), *a.k.a.* Doppler Optical Coherence Tomography, is an emerging technique for blood flow visualization and quantification [4, 16, 39]. It acquires high-resolution and high-contrast tomographic images [17] and has many potential clinical applications such as vascular disease monitoring [16]. To detect the vessels with slow blood flows, dense scanning is usually employed [32] in ODT since the sensitivity to the motion depends on the sampling time [29]. Therefore, the traditional pipeline often suffers from long scanning time and serious storage burden. For example, for a 3D ODT volume of 500 2D B-scans, each raw B-scan takes nearly 4s and 55MB in our experiments.

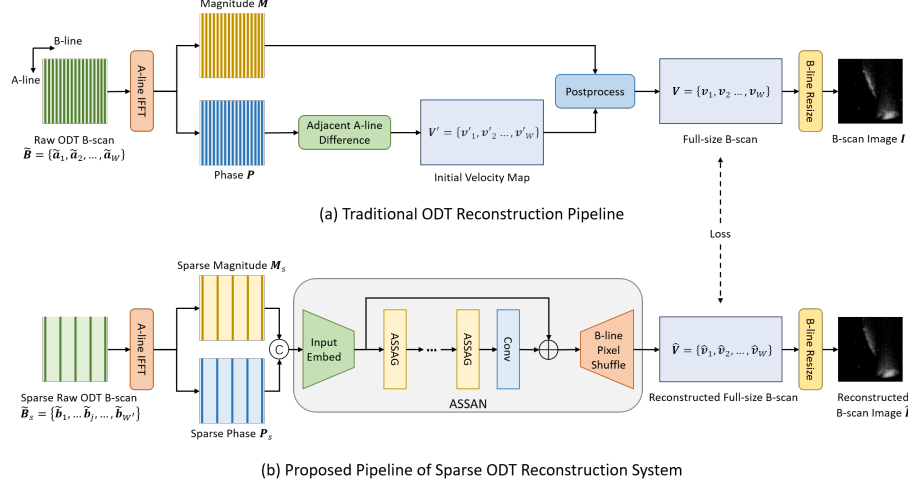


Fig. 1. ODT reconstruction. For sparse reconstruction (b), the sparse raw B-scan \tilde{B}_s is downsampled from \tilde{B} with a stride δ . \tilde{B}_s is forwarded to our Alternative State Space Attention Network (ASSAN) to reconstruct the B-scan image (details in Sec. 2.1).

In a typical ODT image reconstruction pipeline (Fig. 1(a)), the basic scan of raw ODT signals is along the A-line (depth) called raw A-scan encoding the depth information. Sliding the scanning sensor along the B-line (width) obtains a 2D signal called raw B-scan. Then, the final B-scan is achieved after a series of magnitude-phase processing as an indication of blood flow.

To solve the dense scanning issue, we propose, for the first time, a deep learning-based sparse ODT reconstruction pipeline (Fig. 1(b)). The new pipeline requires much fewer A-scans, thus reducing the scanning time and storage burdens. There have been many works on sparse reconstruction based on compressive sensing techniques [23, 38], which are applied in various domains, such as magnetic resonance imaging [24, 34], radar [26, 33], and OCT [7]. However, adapting these techniques to sparse ODT reconstruction is non-trivial and unexplored. Besides, simply applying generic image feature extraction methods for this task may only get suboptimal performance, due to the fundamental difference between natural images and ODT B-scans: information distribution along A-line differs from that along B-line. The A-line is depth-correlated and thus heterogeneous, while the B-line is nearly homogeneous. Besides, the sparse sampling is constrained to the B-line, and no downsampling happens to A-scans. Particularly, since A-line is long, to handle the full-depth information, we propose to use Mamba [8], a recently introduced State Space Model (SSM) [9, 10, 15], which can capture long-range dependency with linear complexity.

Motivated by the above analysis, we develop a novel Alternative State Space Attention Network (ASSAN) (Figure 1(b)) for sparse ODT reconstruction, with five main components: (1) an input block for initial encoding, (2) A-line state

space blocks to handle the depth-correlated A-line information, (3) novel B-line gated attention blocks to model the inter-A-scan features along each B-line, (4) an enhancement feedforward network to learn local features, and (5) a B-line pixel shuffle layer to upscale the B-line for reconstruction.

With all these components, ASSAN effectively captures the unique characteristics of ODT B-scans for accurate reconstruction. To our best knowledge, this is the first attempt for sparse ODT reconstruction and the first report of combining Mamba and self-attention modules to image reconstruction tasks. Experiments demonstrate that our method has exhibited promising results in comparison with state-of-the-arts image reconstruction methods. Besides, the proposed framework is expected to significantly save scanning time (critical to studying flow dynamics to various brain stimulation) and memory while keeping good results. To summarize, our main contributions include:

1. the first sparse reconstruction pipeline for ODT image generation to save the scanning time and storage,
2. a novel alternative SSM attention model along different axes for image reconstruction,
3. an effective gated self-attention model to refine the modeling of inter-A-scan features along the B-line, and
4. the effective sequential Conv1D-based FFN to enhance local features.

2 Alternative State Space Attention Network

2.1 Problem Formulation

Our task is to reconstruct an ODT B-scan image $\hat{\mathbf{I}}$ from the sparse raw B-scans $\tilde{\mathbf{B}}_s$ to speed up scanning and save storage. The traditional ODT reconstruction pipeline with dense scanning is shown in Fig. 1(a). The basic scan of ODT is called raw A-scan $\tilde{\mathbf{a}}_i$, encoding the depth (A-line) information in the frequency domain. A 2D raw signal named raw B-scan $\tilde{\mathbf{B}} = \{\tilde{\mathbf{a}}_i\}_{i=1}^W$, containing W raw A-scans, is acquired by sliding the OCT probing beam along the B-line (width). 1D Inverse Fast Fourier Transform (IFFT) is applied to each raw A-scan to decode the magnitude and phase along the depth. The 1D phase difference of successive A-scans represents the Doppler frequency shift [39], estimating the initial blood flow velocity map \mathbf{V}' . The magnitude is usually used as a mask for denoising in post-process to get the reconstructed full-size B-scan $\mathbf{V} = \{\mathbf{v}_i\}_{i=1}^W$. Due to dense sampling, the aspect ratio of \mathbf{V} is often too small. Finally, a B-line resize operation is used to shrink the width of \mathbf{V} to obtain the final B-scan image \mathbf{I} .

To solve the dense scanning issue of the traditional pipeline, we propose a deep learning-based sparse reconstruction pipeline for ODT imaging using fewer raw A-scans, as illustrated in Fig. 1(b). The goal is to reconstruct a high-quality ODT image $\hat{\mathbf{I}}$ from sparsely sampled raw A-scans stored in $\tilde{\mathbf{B}}_s = \{\tilde{\mathbf{b}}_j = \tilde{\mathbf{a}}_{1+(j-1)\delta}\}_{j=1}^{W'}$, where δ is sampling stride along B-line, $W' = W/\delta$ is the number of sparsely sampled raw A-scans. To maximize the usage of available data for training, we let the network predict the full-size B-scan $\hat{\mathbf{V}} = \{\hat{\mathbf{v}}_i\}_{i=1}^W$ and then resize the B-line of $\hat{\mathbf{V}}$ as the post-process to get the final result $\hat{\mathbf{I}}$.

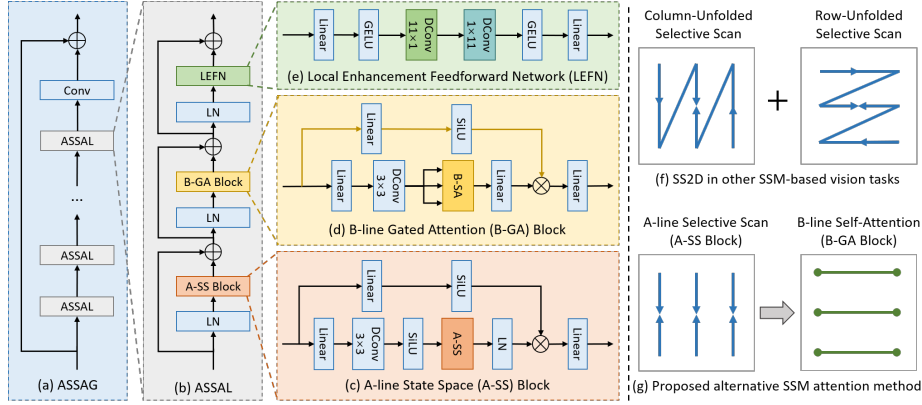


Fig. 2. Network blocks and SSM-based scanning strategies. Blocks: (a) Alternative State Space Attention Group (ASSAG), (b) Alternative State Space Attention Layer (ASSAL), (c) A-line State Space (A-SS) Block, (d) B-line Gated Attention (B-GA) Block, and (e) Local Enhancement Feedforward Network (LEFN). Scans: (f) Cross-scan 2D Selective Scan methods [11, 21] unfold the image to sequences along rows and columns, and scan them via SSM in parallel. (g) Our method first bidirectionally scans each A-line (column) via SSM, and then applies self-attention to each B-line (row).

2.2 Overall Pipeline of ASSAN

The overall pipeline of our method is presented in Fig. 1(b). The input to the whole system is a sparsely sampled raw ODT B-scan $\hat{\mathbf{B}}_s$. Following the traditional pre-process, 1D IFFT is applied to each raw A-scan $\hat{\mathbf{b}}_j$ in $\hat{\mathbf{B}}_s$ to transform the information from the frequency domain representation to the spatial magnitude \mathbf{M}_s and phase \mathbf{P}_s responses along the depth. To encourage rich interaction of the complementary information in magnitude and phase, we concatenate them along the channel dimension as the input to the ASSAN (denoted as \mathbf{X}_{in}). The network input is first encoded via a convolution layer and then processed by a sequence of G Alternative State Space Attention Groups (ASSAG). To maximize the utilization of the available supervision data, a B-line Pixel Shuffle (B-PS) layer is employed to upscale the B-line by rearranging the channel and width dimension to predict the full-size B-scan $\hat{\mathbf{V}}$. The network can be summarized as:

$$\begin{aligned} \mathbf{X}_0 &= \text{Conv}(\mathbf{X}_{in}), \mathbf{X}_g = \text{ASSAG}_g(\mathbf{X}_{g-1}), \quad g = 1, 2, \dots, G \\ \hat{\mathbf{V}} &= \text{B-PS}(\mathbf{X}_0 + \text{Conv}(\mathbf{X}_G)). \end{aligned} \quad (1)$$

We employ the pixel L_2 loss as the optimization function. Finally, the B-line of the predicted $\hat{\mathbf{V}}$ is resized to get the final B-scan image $\hat{\mathbf{I}}$ for visualization.

2.3 Alternative State Space Attention Group

The Alternative State Space Attention Group (ASSAG), shown in Fig. 2(a), consists of a sequence of Alternative State Space Attention Layers (ASSAL) and

a convolution layer for deep feature extraction. The ASSAL’s architecture and feature processing order are illustrated in Fig. 2(b) and Fig. 2(g). To model the different distributions along the A-line and B-line, we alternatively apply the A-line State Space (A-SS) Block and the B-line Gated Attention (B-GA) Block to model the A-line and B-line respectively. Then we employ a Local Enhancement Feedforward Network (LEFN) to enhance local features via Conv1Ds along different axes. LayerNorm (LN) is employed for normalization.

A-line State Space Block (Fig. 2(c)). Since an ODT A-scan is correlated with the depth, and the signals of large and/or vertically growing vessels may span long, we employ Mamba [8] to model full-depth information for its strong sequence modeling capability with linear complexity. In contrast, as widely applied self-attention’s complexity is quadratic to the sequence length and the depth of an A-scan is typically much longer than a usual window attention [19] (e.g., 512 vs 64), self-attention is not suitable to model full depth A-scans.

Unlike most visual Mamba approaches [11, 21], which directly model the entire image via 2D Selective Scan (SS2D) (Fig. 2(f)), we introduce a novel A-line Selective Scan (A-SS) mechanism tailored to A-scans (Fig. 2(g) left). Specifically, A-SS conducts 1D bidirectional scans on each A-line in parallel using Mamba’s 1D selective scan. Compared to SS2D, which unfolds feature maps into sequences by concatenating rows or columns—potentially disrupting depth information—A-SS more effectively models the depth-correlated intra-A-scan features. Given an input \mathbf{X} , an A-SS Block can be represented as:

$$\begin{aligned} \mathbf{X}_1 &= \text{SiLU}(\text{Linear}(\mathbf{X})), \mathbf{X}_2 = \text{LN}(\text{A-SS}(\text{SiLU}(\text{DConv}(\text{Linear}(\mathbf{X}))))), \\ \mathbf{X}_{\text{out}} &= \text{Linear}(\mathbf{X}_1 \odot \mathbf{X}_2). \end{aligned} \quad (2)$$

where SiLU [6] is the Sigmoid-weighted Linear Unit activation function, DConv denotes the depth-wise convolution, and \odot denotes the element-wise product.

B-line Gated Attention Block (Fig. 2(d)). We first introduce a B-line Self-Attention (B-SA) module to model inter-A-scan features, which processes each B-line via 1D self-attention [31] in parallel (Fig. 2(g) right). Since vessels can span long along B-line, capturing long-range information is desired for inter-A-scan modeling. Unlike the depth-correlated A-line, B-line information is nearly homogeneous, making self-attention well-suited. Besides, as sequentially applying bidirectional 1D Mamba across different axes severely affects the throughput [21], using B-SA instead of Mamba also enhances runtime.

However, simply applying the B-SA may not yield optimal performance, as it ignores the distinct behaviors of deep features along this axis. Therefore, uniformly processing B-line features that encode both magnitude and phase information is suboptimal. To further boost the B-line modeling, in addition to the B-SA, we propose a gated attention mechanism to account for the distinct properties of magnitude and phase along the B-line. This mechanism addresses the issue by processing the features in two branches. In one branch, B-SA is employed to model the inter-A-scan features along the B-line, similar to phase processing in the traditional pipeline. In the other branch, the input feature map is transformed into a score mask to refine the features learned by B-SA,

analogous to the traditional processing of magnitude. Although there have been some works on gated attention [12, 35], none of them explores the application to low-level vision tasks. Given an input \mathbf{X} , a B-GA Block can be represented as:

$$\begin{aligned}\mathbf{X}_1 &= \text{SiLU}(\text{Linear}(\mathbf{X})), \mathbf{X}_2 = \text{Linear}(\text{B-SA}(\text{DConv}(\text{Linear}(\mathbf{X})))), \\ \mathbf{X}_{\text{out}} &= \text{Linear}(\mathbf{X}_1 \odot \mathbf{X}_2).\end{aligned}\quad (3)$$

Local Enhancement Feedforward Network (Fig. 2(e)). To further enhance the local information, we introduce the Local Enhancement Feedforward Network (LEFN) by inserting Conv1Ds along different axes into a multi-layer perception (MLP) with GELU [13] activations. Because the above A-SS Block and B-GA Block primarily focus on long-range features, a convolution-based FFN with a large local receptive field could help enhance the local features. Inspired by [30], instead of directly using a large Conv2D kernel, we factorize the Conv2D kernel into two Conv1D kernels along different axes. This way, LEFN not only follows the separate processing of different axes but also brings a large receptive field with fewer parameters.

3 Experiments

3.1 Datasets and Implementation Details

Datasets. The datasets in previous related works [14, 17] contain only 6 volumes in total. Therefore, we collected two datasets, Mouse Cortex Dataset 2021 (MCD-21) and Mouse Cortex Dataset 2023 (MCD-23), each gathered by different experts on different groups of mice. MCD-21 includes 16 volumes (10 for training and 6 for testing) from awake and anesthetized mice. MCD-23 consists of 13 volumes (8 for training and 5 for testing) from anesthetized mice. The train and test volumes are from different mice to better test the generalization capability of the models. Each volume includes 500 densely scanned raw B-scans, and each raw B-scan consists of 12600 raw A-scans. The depth of the IFFT-related pre-processed input to the networks is 512. For practical evaluation, we calculate the PSNR and SSIM for each final B-scan image and crop out deep background regions in the test images to minimize the influence of background noise.

Implementation details. We conduct experiments with the sparse sampling factors 2 and 4 (denoted as $\times 2$ and $\times 4$). In our ASSAN, we employ 4 ASSAGs, each with 6 ASSALs. The feature embedding dimension is set to 60. The size of a pre-processed sparse B-scan patch is 512×64 . The batch size is set to 8. We employ Adam as the optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.99$. The number of training iterations is set to 200K. The learning rate scheduler is CosineAnnealingLR [22] whose initial and minimum learning rates are $2e - 4$ and $1e - 6$.

Since there is no previous work for sparse ODT reconstruction, we compare ASSAN with single-image super-resolution methods [1–3, 5, 11, 18–20, 25, 27, 36, 37, 40] as they also upsample the inputs. We compare with five transformer-based state-of-the-arts: SwinIR [19], HAT [2], SRFormer [40], DAT [3], CFAT [27], and one SSM-based: MambaIR [11]. We apply the same input embedding and replace their upsampling layer [28] with our B-PS layer to make them applicable.

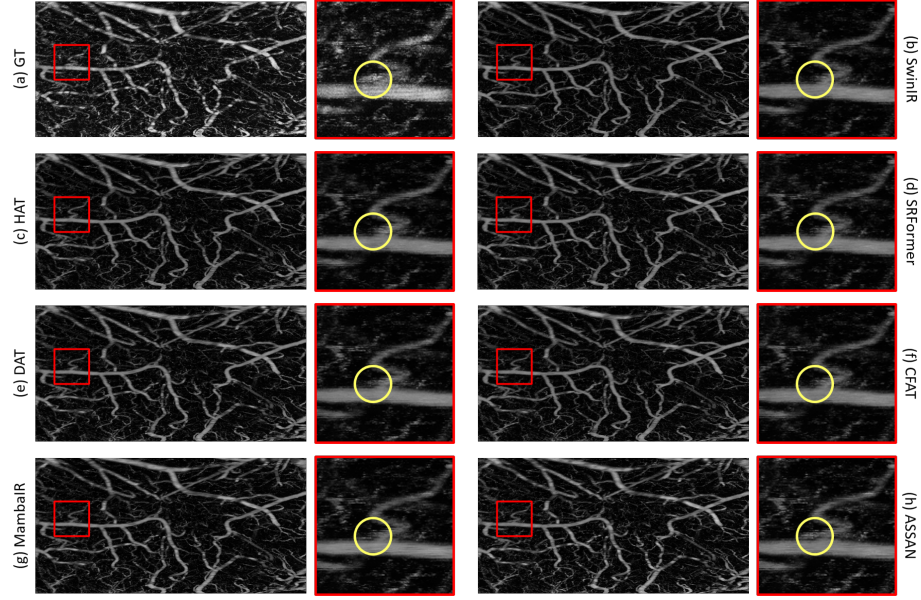


Fig. 3. Qualitative results of $\times 4$ sparsity on MCD-23. Regions of interest are zoomed in. Challenging cases are marked by circles.

3.2 Experimental Results

Quantitative results. Table 1 summarizes the experimental results. Most of the transformer-based methods do not perform well, with the exception of DAT, which also uses channel-wise self-attention to capture global information. However, because channel-wise self-attention is not directly based on spatial features, it lacks accuracy. Though MambaIR outperforms transformer-based methods, its results are still suboptimal due to the SS2D scanning method, which can confuse A-line and B-line information. In contrast, our ASSAN separates A-line and B-line processing with alternative SSM and attention along different axes, thus achieving superior performance compared to other methods.

Qualitative results. Following the common 3D vasculature visualization [14, 16, 17], we present Maximum Intensity Projection (MIP) images for qualitative comparison in Fig. 3, comparing the MIP images of the volumes in MCD-23 $\times 4$ sparsity. The primary evaluation factor of sparse ODT reconstruction is the reconstructed vessel structures. Fig. 3 shows that, in most challenging cases, only our ASSAN successfully reconstructs the vessels. These results demonstrate the advantage of the proposed ASSAN over the other compared methods.

3.3 Ablation Study

Scanning methods. Table 2 (left) presents results using different scanning methods on $\times 4$ sparsity. SS2D (Fig. 2(f)) refers to the method in MambaIR [11].

Table 1. Comparison of sparse ODT reconstruction results.

Method	x2 Sparsity				x4 Sparsity			
	MCD-21		MCD-23		MCD-21		MCD-23	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SwinIR [19]	35.55	0.9068	35.63	0.9275	33.35	0.8770	33.25	0.8948
HAT [2]	35.59	0.9103	35.74	0.9284	33.43	0.8801	33.24	0.8954
SRFormer [40]	35.54	0.9078	35.81	0.9284	33.19	0.8788	33.29	0.8953
DAT [3]	36.09	0.9144	36.48	0.9323	33.73	0.8841	33.71	0.9025
CFAT [27]	35.62	0.9105	35.80	0.9287	33.34	0.8800	33.28	0.8960
MambaIR [11]	36.41	0.9166	36.73	0.9335	33.85	0.8872	34.03	0.9048
ASSAN (Ours)	36.66	0.9185	37.33	0.9363	34.08	0.8911	34.37	0.9080

Table 2. Ablation study on scanning methods and feedforward networks.

Comparison of scanning methods					Comparison of feedforward networks				
Method	MCD-21		MCD-23		Method	MCD-21		MCD-23	
	PSNR	SSIM	PSNR	SSIM		PSNR	SSIM	PSNR	SSIM
SS2D	33.85	0.8874	33.94	0.9051	w/o FFN	34.01	0.8894	34.29	0.9075
A-SS → B-SS	33.89	0.8883	34.13	0.9063	MLP	33.99	0.8880	34.28	0.9073
A-SS → B-SA	33.92	0.8884	34.25	0.9068	CAB	34.03	0.8895	34.28	0.9076
B-GA → A-SS	33.98	0.8890	34.25	0.9074	LEFN2D	34.03	0.8895	34.30	0.9073
A-SS → B-GA	34.01	0.8894	34.29	0.9075	LEFN	34.08	0.8911	34.37	0.9080

A-SS→B-SS indicates that B-line Selective Scan (B-SS) is stacked after A-SS. A-SS→B-SA applies the A-SS Block followed by B-SA, while A-SS → B-GA applies the A-SS Block followed by B-GA Block with the gated design. The results validate the design of our scanning strategy.

Feedforward Network. The results for different FFNs are shown in Table 2 (right). We compare the proposed LEFN with MLP [19], CAB [11, 37], and LEFN2D (5×5 Conv2D). As MLP only performs channel mixing and CAB focuses on global information via channel attention, neither works well. LEFN2D also has limited impact due to its small receptive field and lack of explicit separation between A-line and B-line modeling. In contrast, the proposed LEFN effectively improves performance even with fewer parameters than LEFN2D.

4 Conclusion

We develop a novel ODT sparse reconstruction pipeline, named Alternative State Space Attention Network (ASSAN), to reduce scanning time and storage demands. To effectively model the unique information distributions along different axes in ODT B-scans, ASSAN processes each A-line through a state space model, followed by a gated self-attention block for modeling each B-line. A sequential Conv1D-based feedforward network is used to enhance local features. Extensive experiments indicate the effectiveness of the proposed pipeline.

Acknowledgments. This work was supported in part by NIH Grants 2R01 DA029718 (CD, YP), R21DA042597 (CD, YP), and NIH’s Intramural Program of NIAAA (NDV).

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

References

1. Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., Gao, W.: Pre-trained image processing transformer. In: CVPR. pp. 12299–12310 (2021)
2. Chen, X., Wang, X., Zhou, J., Qiao, Y., Dong, C.: Activating more pixels in image super-resolution transformer. In: CVPR. pp. 22367–22377 (2023)
3. Chen, Z., Zhang, Y., Gu, J., Kong, L., Yang, X., Yu, F.: Dual aggregation transformer for image super-resolution. In: ICCV. pp. 12312–12321 (2023)
4. Chen, Z., Zhang, J.: Doppler optical coherence tomography. *Optical Coherence Tomography: Technology and Applications* pp. 621–651 (2008)
5. Dai, T., Cai, J., Zhang, Y., Xia, S.T., Zhang, L.: Second-order attention network for single image super-resolution. In: CVPR. pp. 11065–11074 (2019)
6. Elfving, S., Uchibe, E., Doya, K.: Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural networks* **107**, 3–11 (2018)
7. Fang, L., Li, S., Cunefare, D., Farsiu, S.: Segmentation based sparse reconstruction of optical coherence tomography images. *TMI* **36**(2), 407–421 (2016)
8. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
9. Gu, A., Goel, K., Re, C.: Efficiently modeling long sequences with structured state spaces. In: ICLR (2021)
10. Gu, A., Johnson, I., Goel, K., Saab, K., Dao, T., Rudra, A., Ré, C.: Combining recurrent, convolutional, and continuous-time models with linear state space layers. *NIPS* **34**, 572–585 (2021)
11. Guo, H., Li, J., Dai, T., Ouyang, Z., Ren, X., Xia, S.T.: Mambair: A simple baseline for image restoration with state-space model. In: ECCV (2024)
12. Han, D., Wang, Z., Xia, Z., Han, Y., Pu, Y., Ge, C., Song, J., Song, S., Zheng, B., Huang, G.: Demystify mamba in vision: A linear attention perspective. *NIPS* **37**, 127181–127203 (2025)
13. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415 (2016)
14. Jiang, Z., Huang, Z., Qiu, B., Meng, X., You, Y., Liu, X., Liu, G., Zhou, C., Yang, K., Maier, A., et al.: Comparative study of deep learning models for optical coherence tomography angiography. *Biomedical optics express* **11**(3), 1580–1597 (2020)
15. Kalman, R.E.: A new approach to linear filtering and prediction problems (1960)
16. Leitgeb, R.A., Werkmeister, R.M., Blatter, C., Schmetterer, L.: Doppler optical coherence tomography. *Progress in retinal and eye research* **41**, 26–43 (2014)
17. Li, A., Du, C., Volkow, N.D., Pan, Y.: A deep-learning-based approach for noise reduction in high-speed optical coherence doppler tomography. *Journal of biophotonics* **13**(10), e202000084 (2020)

18. Li, Y., Fan, Y., Xiang, X., Demandolx, D., Ranjan, R., Timofte, R., Van Gool, L.: Efficient and explicit modelling of image hierarchies for image restoration. In: CVPR. pp. 18278–18289 (2023)
19. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: ICCV. pp. 1833–1844 (2021)
20. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: CVPR Workshops. pp. 136–144 (2017)
21. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Liu, Y.: Vmamba: Visual state space model. arXiv preprint arXiv:2401.10166 (2024)
22. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. In: ICLR (2016)
23. Marques, E.C., Maciel, N., Naviner, L., Cai, H., Yang, J.: A review of sparse recovery algorithms. IEEE access **7**, 1300–1322 (2018)
24. Ning, L., Laun, F., Gur, Y., DiBella, E.V., Deslauriers-Gauthier, S., Megherbi, T., Ghosh, A., Zucchelli, M., Menegaz, G., Fick, R., et al.: Sparse reconstruction challenge for diffusion mri: Validation on a physical phantom to determine which acquisition scheme and analysis method to use? Medical image analysis **26**(1), 316–331 (2015)
25. Niu, B., Wen, W., Ren, W., Zhang, X., Yang, L., Wang, S., Zhang, K., Cao, X., Shen, H.: Single image super-resolution via a holistic attention network. In: ECCV. pp. 191–207. Springer (2020)
26. Potter, L.C., Schniter, P., Ziniel, J.: Sparse reconstruction for radar. In: Algorithms for Synthetic Aperture Radar Imagery XV. vol. 6970, pp. 9–23. SPIE (2008)
27. Ray, A., Kumar, G., Kolekar, M.H.: Cfat: Unleashing triangular windows for image super-resolution. In: CVPR. pp. 26120–26129 (2024)
28. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: CVPR. pp. 1874–1883 (2016)
29. Spaide, R.F., Fujimoto, J.G., Waheed, N.K., Sadda, S.R., Staurengi, G.: Optical coherence tomography angiography. Progress in retinal and eye research **64**, 1–55 (2018)
30. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: CVPR. pp. 2818–2826 (2016)
31. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. NIPS **30** (2017)
32. Wang, R.K., Jacques, S.L., Ma, Z., Hurst, S., Hanson, S.R., Gruber, A.: Three dimensional optical angiography. Optics express **15**(7), 4083–4097 (2007)
33. Wei, S.J., Zhang, X.L., Shi, J., Xiang, G.: Sparse reconstruction for sar imaging based on compressed sensing. Progress in electromagnetics research **109**, 63–81 (2010)
34. Yang, A.C., Kretzler, M., Sudarski, S., Gulani, V., Seiberlich, N.: Sparse reconstruction techniques in magnetic resonance imaging: methods, applications, and challenges to clinical adoption. Investigative radiology **51**(6), 349–364 (2016)
35. Yang, S., Wang, B., Shen, Y., Panda, R., Kim, Y.: Gated linear attention transformers with hardware-efficient training. In: ICML (2024)
36. Zhang, L., Li, Y., Zhou, X., Zhao, X., Gu, S.: Transcending the limit of local window: Advanced super-resolution transformer with adaptive token dictionary. In: CVPR. pp. 2856–2865 (2024)
37. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: ECCV. pp. 286–301 (2018)

38. Zhang, Z., Xu, Y., Yang, J., Li, X., Zhang, D.: A survey of sparse representation: algorithms and applications. *IEEE access* **3**, 490–530 (2015)
39. Zhao, Y., Chen, Z., Saxer, C., Xiang, S., de Boer, J.F., Nelson, J.S.: Phase-resolved optical coherence tomography and optical doppler tomography for imaging blood flow in human skin with fast scanning speed and high velocity sensitivity. *Optics letters* **25**(2), 114–116 (2000)
40. Zhou, Y., Li, Z., Guo, C.L., Bai, S., Cheng, M.M., Hou, Q.: Srformer: Permuted self-attention for single image super-resolution. In: *ICCV*. pp. 12780–12791 (2023)