R2B-WFC Ultrasound Reconstruction: Wavelet Fourier Convolution-based Reconstruction from Radio Frequency to Image

Hyunsu Jeong[1,†[0000-0002-7446-0478]], Chiho Yoon[1,†[0000-0003-3971-3115]], Minsik Sung[1,†[0009-0008-1615-5177]], Kiduk Kim[2[0000-0002-9659-897X]], Dougho Park[3[0000-0002-1288-470X]], Chulhong Kim[1,*[0000-0001-7249-1257]]

[1] Graduate School of Artificial Intelligence (GSAI), Department of Electrical Engineering, Convergence IT Engineering, Mechanical Engineering, Medical Science and Engineering, Medical Device Innovation Center, and Convergence Science and Technology, Pohang University of Science and Technology (POSTECH), Pohang, South Korea
[2] Department of Convergence Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea
[3] Medical Research Institute, Pohang Stroke and Spine Hospital, Pohang, South Korea
chulhong@postech.edu

Abstract. Plane-wave ultrasound (PWUS) facilitates functional imaging through a high frame rate of a few thousand Hz. However, its application remains constrained due to the inferior B-mode image quality in comparison to conventional ultrasound imaging such as focused beam ultrasound (FBUS). In this paper, a data-driven approach is proposed through two steps to enhance the quality of PWUS images. In the first step, the unpaired neural Schrödinger bridge (UNSB) is employed to synthesize high-fidelity images that structurally correspond to the low-quality PWUS images. In the second step, our proposed model, R2B-WFC, is trained to reconstruct high-quality images from the PWUS radio frequency signals, incorporating a wavelet Fourier convolution (WFC) module. Multiple losses are also suggested, combining perceptual loss from a USNB pre-trained model and a Markovian discriminator to preserve high-frequency detail more effectively. As a result, Fréchet Inception Distance (FID), Kernel Inception Distance (KID), Learned Perceptual Image Patch Similarity (LPIPS), Feature Similarity Index Measure (FSIM), Signal to noise ratio (SNR), and Contrast Ratio (CR) scores were 136.32, 0.0356, 0.1956, 0.9514, 41.18 dB, and 27.48 dB, respectively. Compared to image-to-image translation methods, R2B-WFC from RF signal-to-image also shows faster inference time.

Keywords: Image reconstruction, Plane wave ultrasound image, Deep neural network, Spectral convolution, Schrödinger bridge

† These authors contributed equally to this work
* Corresponding author

1 Introduction

Ultrasound (US) imaging is a widely employed modality within the medical field, owing to its real-time performance, non-invasive approach, and cost-effectiveness. The ability to attain high frame rates has led to significant interest in plane wave ultrasound (PWUS) for functional imaging applications [1, 2]. The lower B-mode image quality of PWUS, resulting from its lower intensity than that of focused beam ultrasound (FBUS), restricts its clinical applicability [3]. Consequently, PWUS studies have been predominantly implemented in preclinical studies [4].

The advent of deep learning has prompted recent studies to enhance PWUS images from radio frequency (RF) signal. The studies demonstrated that data-driven approaches yielded superior US images in comparison to rule-based methodologies, such as the delay-and-sum (DAS) beamformer. Luijten et al. [5] replaced adaptive beamformer with a neural network with computational efficiency. Zhang et al. [6] proposed a neural network based on sparse regularization method with shorter reconstruction time. Lu et al. [7] showed a complex convolution neural network based on in-phase and quadrature components of RF signal.

Moreover, deep learning modules that learn features at the frequency level have recently gained attention. Fourier transforms allow the development of deep learning models to have non-local receptive fields by leveraging features that are obtained through spectral transformation [8, 9]. Fast Fourier Convolution (FFC) enable extracting local and global features through convolution and spectral transform [10, 11]. Wavelet transforms have also been employed to capture multiscale information in frequency domain [12, 13].

Our study proposes a data-centric US reconstruction method comprising two steps. In step 1, the unpaired neural Schrödinger bridge (UNSB) [14] is employed to synthesize high-fidelity images that maintain structural alignment with the low resolution of PWUS scans. A simple approach for enhancing image quality from RF signals is to train a deep learning model with high-fidelity synthesized images serving as the gold standard. However, it is inherently challenging to obtain structurally matched pairs of high-quality FBUS and low-quality PWUS images. To overcome the issue, we utilize the UNSB to synthesize high-fidelity images that structurally correspond to the low-quality PWUS images. In step 2, the proposed model is designed to reconstruct images of high quality from RF signals of PWUS. The model, which is referred to as R2B-WFC, uses wavelet Fourier convolution (WFC) modulation, which includes wavelet transform convolution and FFC, to convert the RF signals suitably into the B-mode images. It is demonstrated that R2B-WFC effectively facilitates the model's capacity to capture high-frequency characteristics in the synthesized images, accomplished through multiple losses. The losses encompass perceptual loss, Markovian discriminator loss, and reconstruction loss. Our proposed framework has the following contributions: 1) For the first time, we demonstrate that the whole process of conventional US reconstruction can be replaced by a single deep learning model, 2) R2B-WFC and multiple losses effectively transforms RF signals into images, 3) Compared to image to image translation, our model not only qualitatively and quantitatively improves the image quality, but also reduces the inference time.

2 Proposed Framework

The methodology in this study involves two distinct steps for the training of our model that reconstructs B-mode images from RF signals. As the first step, we synthesize high-fidelity images that structurally paired to low-quality PWUS images. At the second step, our proposed model, R2B-WFC, is trained to reconstruct the high-fidelity synthesized images from PWUS RF signals with multiple losses (Fig. 1).

2.1 Step 1: Image-to-Image translation for the high-fidelity Images

We synthesize the high-fidelity images from the low-quality PWUS images. Low-quality PWUS and high-quality FBUS images are obtained from the Verasonics device and GE, respectively. It is challenging to obtain paired ultrasound (US) images from two different devices because US devices can dynamically capture the motion of objects such as the heart or blood vessels through real-time acquisition. Therefore, we employ UNSB model—which is based on diffusion Schrödinger bridge model [15] — to transform the style of PWUS images into the FBUS style (Fig. 1).

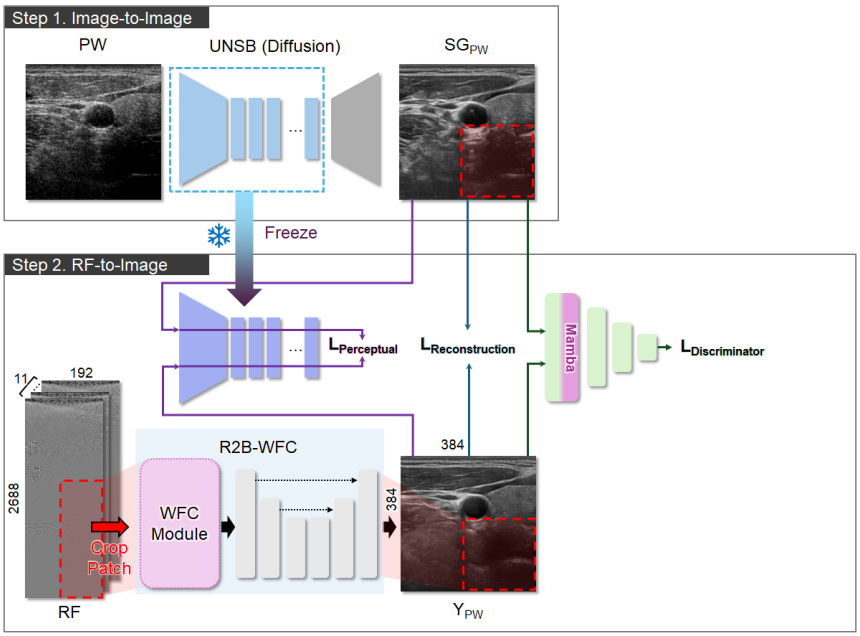

Fig. 1. Overview of our proposed training process.

2.2 Step 2: RF-to-Image reconstruction

We designed R2B-WFC model to reconstruct B-mode images from RF signals. High-fidelity images that are structurally matched with the low-quality PWUS images are synthesized in step 1 and served as the gold standard target SG_{PW} for training in step 2.

Multiple Loss: R2B-WFC adopts multiple losses for training (Fig. 1). First, L_{Rec} is defined as L1 loss between the synthesized gold standard SG_{PW} and Y_{PW}, a reconstructed B-mode image from the RF signal.

$$L_{Rec}(Y_{PW}, SG_{PW}) = |Y_{PW} - SG_{PW}| \tag{1}$$

Second, we incorporate Markovian discriminator loss which evaluates images on patch-by-patch basis. The localized evaluation allows for capturing and preserving fine details and textures—essentially, high-frequency information of B-mode images—ensuring sharp edges and intricate structures. While the discriminator D maximizes the loss, our model R2B-WFC that serves as a generator minimizes the loss. Mamba [16] with the state space models (SSMs) is attached on the first layer of the discriminator to refine high-frequency and edge features. Considering a discriminator D and the gold standard SG_{PW}, Markovian discriminator loss is calculated as follows:

$$L_{Adv}(X_{RF}, SG_{PW}) = \min_{R2B_{WFC}} \max_{D} \mathbb{E}[\log D(SG_{PW})] + \mathbb{E}[\log(1 - D(R2B_{WFC}(X_{RF})))] \tag{2}$$

, where X_{RF} is input RF signal. Third, we add perceptual loss [17] to achieve the style of the high-fidelity synthesized images at feature level. Considering that the UNSB model can already synthesize high-quality FBUS images, it can be assumed that the pre-trained encoder of UNSB model embeds the style of the synthesized images. Therefore, the perceptual loss L_{Per} is optimized by minimizing the difference between the features of reconstructed images Y_{PW} and target SG_{PW}. The perceptual function $\emptyset_j$ extracts features from the j-th layer of the encoder.

$$L_{Per}(Y_{PW}, SG_{PW}) = \frac{1}{C_j H_j W_j} \left\| \emptyset_j(Y_{PW}) - \emptyset_j(SG_{PW}) \right\|_2^2 \tag{3}$$

By setting $\lambda_1 = 1, \lambda_2 = 1$, and $\lambda_3 = 1$, the final objective function is defined by combining the three losses as follows:

$$L_{total} = \lambda_1 * L_{Rec} + \lambda_2 * L_{Adv} + \lambda_3 * L_{Per} \tag{4}$$

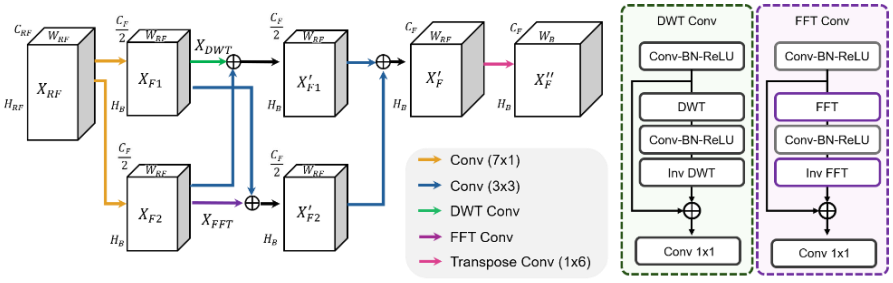

Fig. 2. WFC Module. FFT: Fast Fourier Transform, DWT: Discrete Wavelet Transform.

WFC module: WFC module processes RF signals enabling the U-Net to effectively reconstruct B-mode image $X_B \in \mathbb{R}^{1 \times H_B \times W_B}$, where $H_B \times W_B$ is set to the cropped patch size 192×192. As a first step, WFC crops 11×1344×96 size of RF signal,

resulting in $X_{RF} \in \mathbb{R}^{C_{RF} \times H_{RF} \times W_{RF}}$. Second, convolution filters split X_{RF} into two branches $X_{F1}, X_{F2} \in \mathbb{R}^{\frac{C_F}{2} \times H_B \times W_{RF}}$ where C_F is 16. The multiscale branch and global branch in the spectral domain utilize discrete wavelet transform (DWT) and fast Fourier transform (FFT), respectively. DWT conv is performed sequentially in the order of DWT, 1×1 convolution ($Conv$), batch normalization (BN), $ReLU$, and inverse DWT (Inv DWT). DWT decomposes X_{F1} into four frequency sub-bands, low-low (X_{LL}), low-high (X_{LH}), high-low (X_{HL}), and high-high (X_{HH}) band. 1×1 $Conv$ learns interactions among the sub-bands, while suppressing noise and enhancing key features such as edges and textures.

$$DWT(X_{F1}) = [X_{LL} \parallel X_{LH} \parallel X_{HL} \parallel X_{HH}] \in \mathbb{R}^{2C_F \times \frac{H_B}{2} \times \frac{W_{RF}}{2}} \tag{5}$$

$$\text{Inv } DWT(ReLU(BN(Conv_{1\times1}(DWT(X_{F1}))))) = X_{DWT} \in \mathbb{R}^{\frac{C_F}{2} \times H_B \times W_{RF}} \tag{6}$$

FFT conv also follows a sequential process such as FFT, $Conv$, BN, $ReLU$, and inverse FFT (Inv FFT). FFT produces real $X_R \in \mathbb{R}^{\frac{C_F}{2} \times \frac{H_B}{2} \times W_{RF}}$ and imaginary parts $X_I \in \mathbb{R}^{\frac{C_F}{2} \times \frac{H_B}{2} \times W_{RF}}$, which are then concatenated to form $X_R \parallel X_I \in \mathbb{R}^{C_F \times \frac{H_B}{2} \times W_{RF}}$. $Conv$ adjusts frequency bands of $X_R \parallel X_I$ which contains global information within a single value, acting as a filter to emphasize or suppress certain frequency components.

$$FFT(X_{F2}) = X_R \parallel X_I \in \mathbb{R}^{C_F \times \frac{H_B}{2} \times W_{RF}} \tag{7}$$

$$\text{Inv } FFT(ReLU(BN(Conv_{1\times1}(FFT(X_{F2}))))) = X_{FFT} \in \mathbb{R}^{\frac{C_F}{2} \times H_B \times W_{RF}} \tag{8}$$

X_{DWT} and X_{FFT} are respectively added to the feature map obtained from each branch's 3×3 $Conv$ layer to make $X'_{F1}, X'_{F2} \in \mathbb{R}^{\frac{C_F}{2} \times H_B \times W_{RF}}$. X'_{F1} and X'_{F2} pass through another 3×3 $Conv$, which doubles the channel size to yield C_F. The two feature maps are then added to make X'_F. Finally, transposed convolution expands the width of X'_F to match the width of the B-mode image, ensuring that the shape $\in \mathbb{R}^{C_F \times H_B \times W_B}$ aligns with the width and height of the B-mode image.

3 Experiments

PWUS data was acquired using the Vantage 256 system (Verasonics Inc., Kirkland, WA, USA), which spatially compounded of 11 steered PWs (evenly spaced from -10° to 10°). FBUS data was obtained from GE Logiq Fortis (GE Healthcare Inc., Chicago, IL, USA). From 47 volunteers, we gathered a total of 517 FBUS *in-vivo* images, along with 517 PWUS images and their corresponding RF data (101 Musculoskeletal (MSK), 262 carotids, and 154 thyroids). To quantitatively evaluate models on three anatomical regions, we used six metrics: Fréchet Inception Distance (FID) [18], Kernel Inception Distance (KID) [19], Learned Perceptual Image Patch Similarity (LPIPS) [20], Feature Similarity Index Measure (FSIM) [21], Contrast Ratio (CR), and Signal-to-Noise Ratio (SNR) [22].

3.1 Image-to-Image Results for high-fidelity Synthetized Images

To synthesize high-fidelity images from PWUS images, we compared four models, including CycleGAN [23], CUT [24], I2SB [25], and UNSB [14]. Qualitative results revealed that hallucination artifacts appeared in the Carotid and MSK classes (Fig. 3). The occurrence of these artifacts was especially identified in both conventional GAN models, including CycleGAN and CUT. Also, I2SB struggled to accurately reproduce the structural characteristics of the PWUS images. In contrast, UNSB effectively preserved the structural integrity of the input PWUS images while generating high fidelity that closely resembled the FBUS images. In quantitative result, UNSB achieved the best results on all metrics (Table 1).

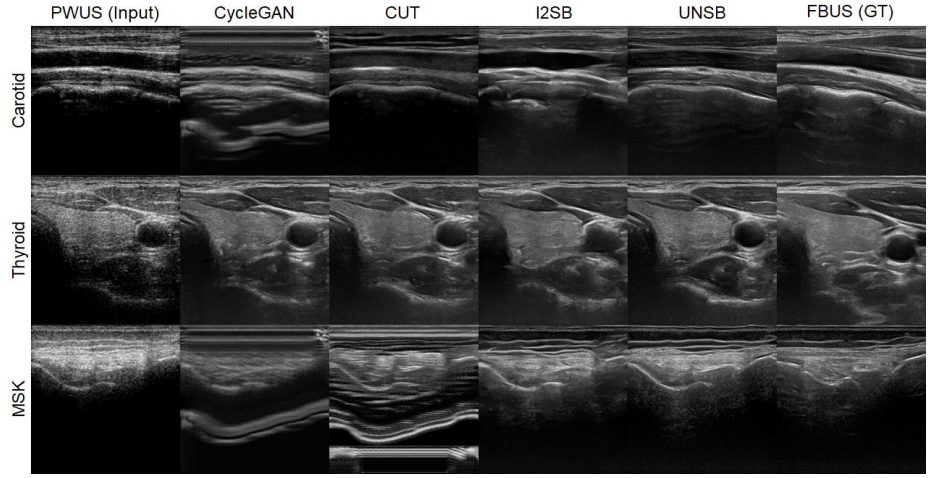

Fig. 3. Qualitative comparison of Image-to-Image results.

Table 1. Quantitative comparison of Image-to-Image translation results.

Method	FID↓	KID_{x100}↓	LPIPS↓	FSIM↑	CR_{dB} ↑	SNR_{dB}↑
PWUS	261.7638	22.5615	0.4035	0.9023	21.7583	27.4170
CycleGAN [23]	192.4531	5.0668	0.2987	0.9077	20.0888	31.7068
CUT [24]	180.6778	4.0367	0.3026	0.9060	21.4752	32.5420
I2SB [25]	167.0595	6.7854	0.2676	0.9168	16.2576	28.3448
USNB [14]	**142.9922**	**1.4826**	**0.2594**	**0.9141**	**29.1455**	**41.8104**

3.2 RF-to-Image Reconstruction Results

In this study, R2B-WFC was compared with other models that used a convolution kernel to transform the shape of RF signals into the shape of B-mode images and passed through different networks including CycleGAN [23], CUT [24], I2SB [25], and UNSB [14].

As illustrated in Fig. 4, other models, with the exception of R2B-WFC, were found to be inconsistent structures when compared to PWUS structures. In contrast, the R2B-WFC with RF signal processing module demonstrated effective preservation of structural integrity and superior performance, as evidenced by the highest scores on all metrics (Table 2). Notably, compared to other models, R2B-WFC showed better CR and SNR performance, representing that R2B-WFC preserved structural and high-frequency information well. The results indicated the difficulties inherent in directly converting large-scale RF data into images, thereby demonstrating the limitations of image-to-image translation models in the absence of RF signal processing techniques and appropriate loss functions. Furthermore, we conducted an ablation study on the impact of different module and loss combinations (Table 3). The WFC module for RF signal processing demonstrated consistent superiority over a simple convolution approach across all classes. It has been observed that the perceptual loss greatly leverages the Markovian discriminator. In addition, integrating the Mamba module into the discriminator's first layer has been shown to enhance the interaction between the discriminator and our reconstruction model, thereby leading to better SNR and KID.

R2B-WFC achieved the most efficient inference time of 0.0439s, outperforming image-to-image translation methods, with CycleGAN, CUT, I2SB, and UNSB achieving 0.1587s, 0.0552s, 18.105s, and 0.3133s, respectively (Fig. 5). While image-to-image translation models require a substantial amount of time due to the cascade approach, which involves image-to-image translation after conventional US reconstruction, our model achieved faster reconstruction by adopting an end-to-end approach that directly transforms RF signals into images. The comprehensive improvement demonstrated that the R2B-WFC model enhanced spatiotemporal resolution, enabling faster image reconstruction without sacrificing quality.

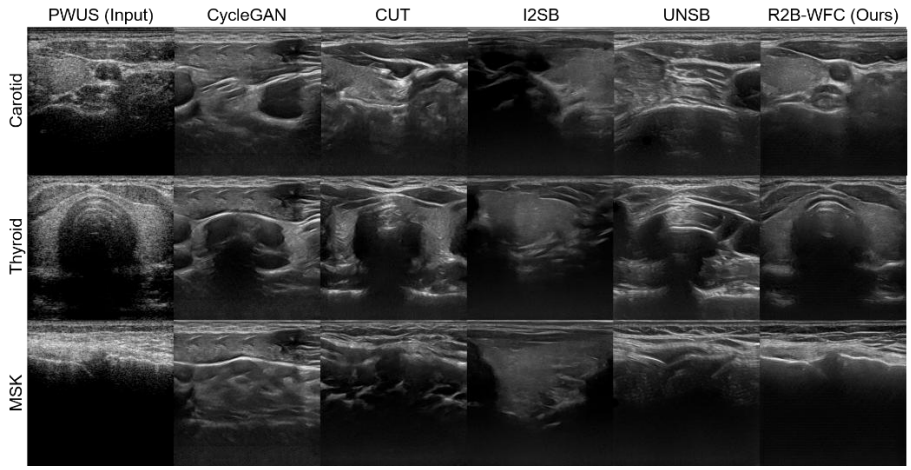

Fig. 4. Visualization of RF-to-Image results for qualitative comparison.

Table 2. Quantitative comparison of RF-to-Image results.

Method	FID↓	KID_{x100}↓	LPIPS↓	FSIM↑	CR_{dB}↑	SNR_{dB}↑
PWUS	270.5772	25.6898	0.3313	0.9424	21.7583	27.4170
CycleGAN [23]	205.1438	13.0466	0.3245	0.9015	4.3550	9.4933
CUT [24]	175.4161	6.5089	0.2864	0.9108	10.9040	18.5863
I2SB [25]	190.5750	9.6633	0.3496	0.8984	1.5929	11.1222
USNB [14]	185.3808	8.0504	0.2752	0.9133	8.9938	14.5603
R2B-WFC (Ours)	**136.3188**	**3.5608**	**0.1956**	**0.9514**	**27.4825**	**41.1804**

Table 3. Ablation study on different modules and loss functions for R2B-WFC. Mamba: discriminator with Mamba attached on the first layer of the network.

			Carotid		Thyroid		MSK	
Module	L_{Per}	L_{Adv}	KID_{x100}↓	SNR_{dB}↑	KID_{x100}↓	SNR_{dB}↑	KID_{x100}↓	SNR_{dB}↑
Conv	X	X	33.0346	37.4833	28.5498	34.4838	21.1940	33.7593
FFC	X	X	36.8563	41.1162	30.6652	37.7294	24.8195	34.8465
WFC	X	X	29.2349	41.4304	22.9669	36.1288	17.6179	34.0249
WFC	O	X	43.2559	38.3389	41.0451	31.7972	28.4508	35.6616
WFC	O	Basic	8.5622	37.9219	7.5946	35.6677	6.4702	38.1800
WFC	O	Mamba	**3.4890**	**44.1852**	**2.9644**	**40.2674**	**4.2289**	**39.0886**

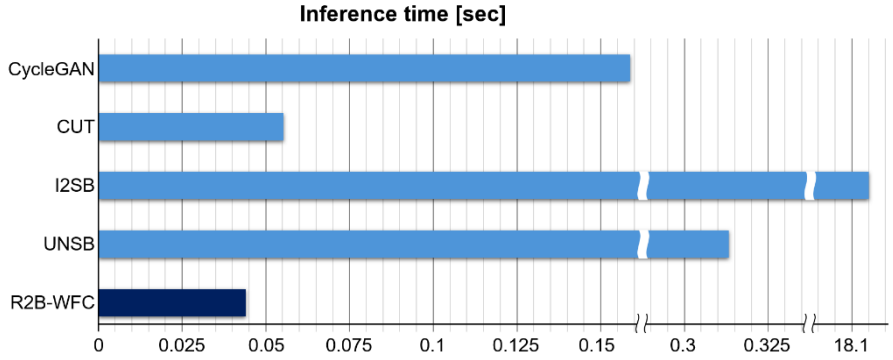

Fig. 5. Inference time comparison between the RF-to-Image method R2B-WFC and Image-to-Image translation methods.

4 Conclusion

In this study, we proposed a data-centric US reconstruction from RF signal to B-mode images. Using the high-fidelity synthesized images from UNSB, the proposed model preserved structural information and matches the style of high-quality images. It was demonstrated that R2B-WFC model successfully transforms RF signals into image

features. In addition, multiple losses facilitated high-fidelity reconstruction by leveraging shared representations across tasks such as Markovian discriminator loss, reconstruction loss, and perceptual loss. Unlike image-to-image translation models that applied translation after conventional US reconstruction, our model enabled faster reconstruction by directly converting RF signals into images through an end-to-end approach.

Acknowledgments. This work was supported by the following sources: the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2024-00415450); the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (2023R1A2C3004880); the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A03047902); the Commercialization Promotion Agency for R&D Outcomes (COMPA) funded by the Ministry of Science and ICT (MSIT) (No.2710006567); the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2019-II191906, Artificial Intelligence Graduate School Program(POSTECH)); BK21 FOUR program.

Disclosure of Interests. Chulhong Kim has financial interests in OPTICHO, which, however, did not support this work. All other authors declare no competing interest.

References

1. Demené, C., Deffieux, T., Pernot, M., Osmanski, B.F., Biran, V., Gennisson, J.L., Sieu, L.A., Bergel, A., Franqui, S., Correas, J.M., Cohen, I., Baud, O., Tanter, M.: Spatiotemporal Clutter Filtering of Ultrafast Ultrasound Data Highly Increases Doppler and fUltrasound Sensitivity. Ieee T Med Imaging 34, 2271-2285 (2015)
2. Oh, D., Lee, D.H.Y., Heo, J., Kweon, J., Yong, U.J., Jang, J., Ahn, Y.J., Kim, C.: Contrast Agent-Free 3D Renal Ultrafast Doppler Imaging Reveals Vascular Dysfunction in Acute and Diabetic Kidney Diseases. Adv Sci 10, (2023)
3. Zhou, Z.X., Wang, Y.Y., Guo, Y., Jiang, X.M., Qi, Y.X.: Ultrafast Plane Wave Imaging With Line-Scan-Quality Using an Ultrasound-Transfer Generative Adversarial Network. Ieee J Biomed Health 24, 943-956 (2020)
4. Renaudin, N., Demené, C., Dizeux, A., Ialy-Radio, N., Pezet, S., Tanter, M.: Functional ultrasound localization microscopy reveals brain-wide neurovascular activity on a microscopic scale. Nature Methods 19, 1004-1012 (2022)
5. Luijten, B., Cohen, R., de Bruijn, F.J., Schmeitz, H.A.W., Mischi, M., Eldar, Y.C., van Sloun, R.J.G.: Adaptive Ultrasound Beamforming Using Deep Learning. Ieee T Med Imaging 39, 3967-3978 (2020)
6. Zhang, J.K., He, Q., Xiao, Y., Zheng, H.R., Wang, C.Z., Luo, J.W.: Ultrasound image reconstruction from plane wave radio-frequency data by self-supervised deep neural network. Medical Image Analysis 70, (2021)
7. Lu, J.F., Millioz, F., Garcia, D., Salles, S., Ye, D., Friboulet, D.: Complex Convolutional Neural Networks for Ultrafast Ultrasound Imaging Reconstruction From In-Phase/Quadrature Signal. Ieee T Ultrason Ferr 69, 592-603 (2022)

8. Chi, L., Tian, G.Y., Mu, Y.D., Xie, L.X., Tian, Q.: Fast Non-Local Neural Networks with Spectral Residual Learning. Proceedings of the 27th Acm International Conference on Multimedia (Mm'19) 2142-2151 (2019)
9. Zhong, Z.S., Shen, T.C., Yang, Y.B., Zhang, C., Lin, Z.C.: Joint Sub-bands Learning with Clique Structures for Wavelet Domain Super-Resolution. Advances in Neural Information Processing Systems 31 (Nips 2018) 31, (2018)
10.Chi, L., Jiang, B., Mu, Y.: Fast fourier convolution. Advances in Neural Information Processing Systems 33, 4479-4488 (2020)
11.Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., Lempitsky, V.: Resolution-robust large mask inpainting with fourier convolutions. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 2149-2159. (2022)
12.Bae, W., Yoo, J., Chul Ye, J.: Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 145-153. (2017)
13.Gao, X., Qiu, T., Zhang, X., Bai, H., Liu, K., Huang, X., Wei, H., Zhang, G., Liu, H.: Efficient multi-scale network with learnable discrete wavelet transform for blind motion deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2733-2742. (2024)
14.Kim, B., Kwon, G., Kim, K., Ye, J.C.: Unpaired Image-to-Image Translation via Neural Schr\" odinger Bridge. arXiv preprint arXiv:2305.15086 (2023)
15.De Bortoli, V., Thornton, J., Heng, J., Doucet, A.: Diffusion schrödinger bridge with applications to score-based generative modeling. Advances in Neural Information Processing Systems 34, 17695-17709 (2021)
16.Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. arXiv preprint arXiv:2401.09417 (2024)
17.Johnson, J., Alahi, A., Li, F.F.: Perceptual Losses for Real-Time Style Transfer and Super-Resolution. Computer Vision - Eccv 2016, Pt Ii 9906, 694-711 (2016)
18.Hensel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. Advances in Neural Information Processing Systems 30 (Nips 2017) 30, (2017)
19.Bińkowski, M., Sutherland, D.J., Arbel, M., Gretton, A.: Demystifying mmd gans. arXiv preprint arXiv:1801.01401 (2018)
20.Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. 2018 Ieee/Cvf Conference on Computer Vision and Pattern Recognition (Cvpr) 586-595 (2018)
21.Zhang, L., Zhang, L., Mou, X.Q., Zhang, D.: FSIM: A Feature Similarity Index for Image Quality Assessment. Ieee T Image Process 20, 2378-2386 (2011)
22.Khan, S., Huh, J., Ye, J.C.: Adaptive and Compressive Beamforming Using Deep Learning for Medical Ultrasound. Ieee T Ultrason Ferr 67, 1558-1572 (2020)
23.Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. 2017 Ieee International Conference on Computer Vision (Iccv) 2242-2251 (2017)

24.Park, T., Efros, A.A., Zhang, R., Zhu, J.-Y.: Contrastive learning for unpaired image-to-image translation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16, pp. 319-345. Springer, (2020)
25.Liu, G.-H., Vahdat, A., Huang, D.-A., Theodorou, E., Nie, W., Anandkumar, A.: I^2SB: Image-to-Image Schrödinger Bridge. In: Andreas, K., Emma, B., Kyunghyun, C., Barbara, E., Sivan, S., Jonathan, S. (eds.) Proceedings of the 40th International Conference on Machine Learning, vol. 202, pp. 22042--22062. PMLR, Proceedings of Machine Learning Research (2023)