

Performance Improvement of Single Plane-Wave Imaging Using U-Net and Discrete Wavelet Transform

Hiromi Shidara* Kanta Miura* Takuro Ishii[†] Koichi Ito* Takafumi Aoki* Yoshifumi Saijo[‡] and Jun Ohmiya[§]

* Graduate School of Information Sciences, Tohoku University, Japan

E-mail: shidara@aoki.ecei.tohoku.ac.jp

[†] Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Japan

[‡] Graduate School of Biomedical Engineering, Tohoku University, Japan

[§] Konica Minolta, Inc., Japan

Abstract—Single Plane-Wave Imaging (SPWI), which transmits a single plane wave, can acquire ultrasound images at more than 1,000 fps, although it has poor lateral resolution and contrast. Some methods have been proposed to improve the quality of ultrasound images acquired by SPWI using deep learning, however, the quality is lower than that of compounded images, which are composed of multiple SPWI images. In addition, the RF signal is used as an input, which is computationally expensive. In this paper, we propose a method to improve the performance of SPWI using U-Net and the Discrete Wavelet Transform (DWT). The proposed method uses In-phase and Quadrature (IQ) data as the input and output of U-Net and loss functions that take into account the characteristics of the RF signal to improve the quality of images, and also uses IQ data after DWT to reduce the computational complexity and the inference time. Through a set of experiments using our ultrasound image dataset, we demonstrate the effectiveness of the proposed method.

I. INTRODUCTION

Ultrasound imaging can acquire images in real time with a small and portable device and with low impact on the human body compared to Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). Therefore, ultrasound imaging is utilized for analyzing and evaluating the cardiovascular system and observing muscular tissue [1]. In standard ultrasound imaging, ultrasound images are obtained by sequentially transmitting more than 100 focused ultrasound beams, resulting in speeds limited to a few tens of fps [2], [3]. Since detailed tissue observation [4]–[7] and blood flow evaluation [8]–[11] require ultrasound images at high frame rates, Plane-Wave Imaging (PWI), which uses plane waves to acquire ultrasound images, has been used [12].

PWI can acquire images at a higher frame rate than standard ultrasound imaging by transmitting plane waves of the same width as the field of view several times. Single Plane-Wave Imaging (SPWI), which acquires images by transmitting a single plane wave, can archive speeds of more than 1,000 fps, while there is a problem that the lateral resolution and contrast of the images are degraded [13]. Coherent Plane-Wave-Compounding (CPWC) is applied to SPWI to address the above problems [14]. CPWC acquires a compound image

by transmitting a plane wave multiple times in different directions while controlling the delay of the ultrasound elements, and then coherently adding the received Radio Frequency (RF) signals. CPWC can improve the lateral resolution and contrast of images, although the frame rate is lower than that of SPWI due to the need to transmit a plane wave multiple times. To address the trade-off between frame rate and image quality in CPWI, several methods have been proposed to improve the quality of ultrasound images obtained by SPWI comparable to that of compound images by deep learning [15], [16]. Li et al. [15] proposed a method using U-Net [17], which is commonly used in image segmentation and image generation. U-Net in this method consists of VGG-13 [18] as an encoder and a network of four inverse convolution layers as a decoder. This U-Net is trained to minimize the L1 loss between the compound images obtained by CPWC and the ultrasound images generated by U-Net. Since the L1 loss between images is used, the high-frequency components of the compound image cannot be represented sufficiently, resulting in the generation of blurred ultrasound images. Perdios et al. [16] also proposed a U-Net based method. Mean Signed Logarithmic Absolute Error (MSLAE) is used as a loss function to take into account the wide dynamic range of ultrasound images and to preserve the characteristics of the RF signal. On the other hand, it is necessary to set an appropriate threshold to train U-Net with MSLAE, however, the details of the threshold determination are not available, making the method unreproducible. Both methods use RF signals as input, which have higher resolution than images, and are computationally expensive, therefore, the real-time capability of ultrasound imaging may be compromised.

To address the above problems, we propose a method to improve the quality of ultrasound images acquired by SPWI using U-Net and to reduce the computational cost by using Discrete Wavelet Transform (DWT). The proposed method uses In-phase and Quadrature (IQ) data, which consists of the real part of the beamformed RF signal and the imaginary part of the Hilbert transform of the RF signal, as the input and output of U-Net to represent the high frequency component

of the ultrasound signals. Since the effect of the point spread of the envelope signal, which is the amplitude of the IQ data, increases with the transmission distance of the ultrasound [19], we introduce loss functions that take into account the effect of the point spread, which is close to the envelope signal, and the frequency characteristics of the RF signals. We also reduce the computational complexity and inference time by inputting the IQ data after DWT into U-Net. Since 2D data can be decomposed into four sub-bands (LL , LH , HL , HH) by DWT, DWT has been applied to image coding and texture identification based on multiple resolution analysis [20]. In the field of ultrasound imaging, DWT has been also used to reduce speckle noise in ultrasound images [21] and to automatic diagnosis of Graves' disease [22]. Since DWT can reduce the data size by half while keeping the information in the 2-D data, the proposed method can reduce the computational complexity and inference time by inputting the IQ data after DWT. The effectiveness of the proposed method is demonstrated through a set of experiments using the ultrasound image dataset created by the authors.

II. METHOD

In this section, we describe the details of the network architecture and loss functions for the proposed method to improve the quality of ultrasound images acquired by SWPI.

A. Network Architecture

The proposed method employs an encoder-decoder model based on U-Net [17] as shown in Fig. 1. The proposed method replaces the encoder of U-Net with ResNet-34 [23] and uses a decoder consisting of inverse convolution layers as in U-Net. To clarify the difference between the original U-Net and U-Net used in the proposed method, U-Net used in the proposed method is denoted as ResU-Net in the following. We use IQ data as the input and output of ResU-Net instead of ultrasound images to take into account the high-frequency components of ultrasound images. The use of IQ data makes it possible to take into account point the effect of point spread of the envelope signal acquired by CPWC and frequency characteristics of the RF signal. For IQ data with $H \times W$, 2D DWT is applied to each of the real and imaginary parts, which is decomposed into four sub-bands, i.e., LL , LH , HL , and HH , and the size of IQ data becomes $\frac{H}{2} \times \frac{W}{2}$ for each sub-band. The four sub-bands are concatenated in the channel direction and the data with $8 \times \frac{H}{2} \times \frac{W}{2}$ is used as the input to ResU-Net. The IQ data is output by applying 2D IDWT to the data with $8 \times \frac{H}{2} \times \frac{W}{2}$ obtained at the final layer of ResU-Net and reconstructing it to the original size of $2 \times H \times W$.

B. Loss Functions

The training of the proposed method employs the three loss functions based on the IQ data, the envelope signal, and the frequency characteristics. In the calculation of the loss functions, the IQ data after compounding are used as the ground truth. In the following, we denote $r(n_1, n_2)$ and $i(n_1, n_2)$ as the real and imaginary parts of the ground truth

IQ data of an $N_1 \times N_2$ matrix, respectively. Let $\hat{r}(n_1, n_2)$ and $\hat{i}(n_1, n_2)$ denote the real and imaginary parts of IQ data output by ResU-Net, respectively. The loss based on IQ data, L_{IQ} , is defined by

$$L_{IQ} = \frac{1}{N_1 N_2} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \left\{ \left| \hat{r}(n_1, n_2) - r(n_1, n_2) \right| + \left| \hat{i}(n_1, n_2) - i(n_1, n_2) \right| \right\}. \quad (1)$$

The loss based on the envelope signal, L_{Env} , is defined by

$$L_{Env} = \frac{1}{N_1 N_2} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \left\{ \left| \sqrt{\hat{r}(n_1, n_2)^2 + \hat{i}(n_1, n_2)^2} - \sqrt{r(n_1, n_2)^2 + i(n_1, n_2)^2} \right| \right\}. \quad (2)$$

The use of L_{Env} makes it possible to achieve a similar effect of point spread to the envelope signal obtained by CPWC. Let $A(n_1, k_2)$ denote the amplitude and $\theta(n_1, k_2)$ denote the phase obtained by 1D discrete Fourier transform of the real part of the IQ data in the column direction, i.e., n_2 , respectively. The loss based on the frequency characteristics, $L_{Fourier}$, is defined by

$$L_{Fourier} = \frac{1}{N_1 K_2} \sum_{n_1=0}^{N_1-1} \sum_{k_2=0}^{K_2-1} \left\{ \lambda_A \left| \hat{A}(n_1, k_2) - A(n_1, k_2) \right| + \lambda_\theta \left(1 - \cos(\hat{\theta}(n_1, k_2) - \theta(n_1, k_2)) \right) \right\}, \quad (3)$$

where $A(n_1, k_2)$ and $\theta(n_1, k_2)$ are the amplitude and phase spectrum calculated from the real part of IQ data after compounding, respectively, $\hat{A}(n_1, k_2)$ and $\hat{\theta}(n_1, k_2)$ are the amplitude and phase spectrum calculated from the real part of IQ data output by ResU-Net, respectively, k_2 is the index of discrete frequency, and λ_A and λ_θ are hyperparameters that adjust the balance between amplitude and phase spectrum, respectively. The use of $L_{Fourier}$ makes it possible to close the frequency spectrum of the output by ResU-Net to the frequency spectrum of the ground truth of the RF signal. The total loss function, L , used in training is defined by

$$L = L_{IQ} + \lambda_{Env} L_{Env} + \lambda_{Fourier} L_{Fourier}, \quad (4)$$

where λ_{Env} and $\lambda_{Fourier}$ are hyperparameters that adjust the balance among loss functions.

III. EXPERIMENTS AND DISCUSSION

This section describes a performance evaluation of the proposed method for improving the quality of SPWI.

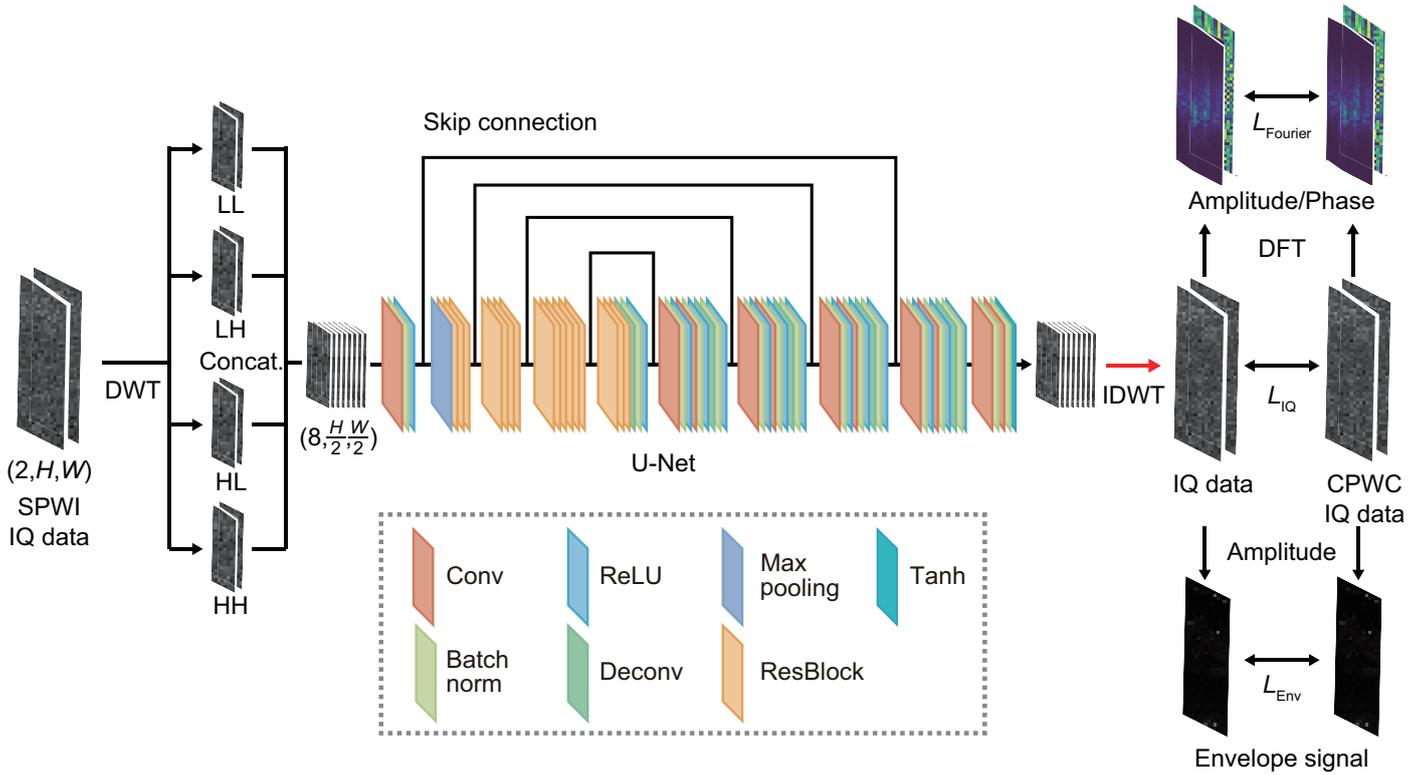


Fig. 1. Overview of the proposed method and loss functions used in training.

TABLE I
CONFIGURATION OF SPWI/CPWC PAIRS FOR TRAINING, VALIDATION,
AND TEST IN OUR DATASET.

Target	Train	Val	Test	Total
Breast ultrasound examination phantom	2,800	400	800	4,000
Image quality assurance phantom	400	—	1,000	1,400
Subject #1	500	—	—	500
Subject #2	—	200	200	400
Subject #3	—	—	100	100
Total	3,700	600	2,100	6,400

A. Dataset

In the experiments, we use our SPWI/CPWC dataset¹. This dataset contains 6,400 pairs consisting of IQ data acquired by SPWI and IQ data acquired by CPWC as shown in Table I. The targets are the breast ultrasound examination phantom (BP), the image quality assurance phantom (QAP), and the cervical regions of three healthy subjects (Subject). The IQ data of CPWC used as ground truth in the experiments are obtained by coherently adding 75 beamformed IQ data acquired by varying the transmission angle of the plane wave in the range of $-16 \sim 16$ degrees.

B. Experimental Condition

The dataset is separated into 3,700 pairs for training, 600 pairs for validation, and 2,100 pairs for test as shown in Table I. The IQ data acquired by SPWI is padded to 384×768 and

normalized to have each element in the range $[-1, 1]$. We set $\lambda_A = 0.014$, $\lambda_\theta = 0.026$, $\lambda_{Env} = 0.560$, and $\lambda_{Fourier} = 0.440$, which are optimized by Optuna [24]. AdamW [25] is used as the optimizer, the learning rate is $1e-4$, the batch size is 16, and the number of epochs is 300.

We compare the performance of the proposed method (“ResU-Net”) with that of the conventional method [15] (“Li et al.”), the proposed method with the original U-Net (“U-Net”), and the proposed method without DWT (“ResU-Net w/o DWT”). Since the code for the conventional method is not publicly available, we implemented it according to [15] and trained it under the same experimental conditions as the proposed method. We employ “U-Net” in the Pytorch implementation² and evaluate it in the same environment as the other methods. The experimental conditions for “U-Net” are the same as for the proposed method, except for the encoder architecture. The difference between “ResU-Net w/o DWT” and “ResU-Net” is whether DWT is applied to IQ data or not. We use the discrete Haar wavelet transform as DWT in the experiments. All methods are implemented using Pytorch 1.13.1 and evaluated on an AMD EPYC 7502 32-core Processor and an A100 (80GB).

C. Evaluation Metrics

In this experiment, we evaluate image quality, lateral resolution, and contrast to compare the quality improvement of each

¹https://github.com/gsaiki/Improvement_of_Ultrasound_Image_Quality

²https://pytorch.org/hub/mateuszbeda_brain-segmentation-pytorch_unet/

TABLE II

EXPERIMENTAL RESULTS FOR EVALUATING IMAGE QUALITY, WHERE THE VALUES IN BOLD INDICATE THE HIGHEST VALUES FOR EACH METRIC.

Method	PSNR [dB] \uparrow			SSIM \uparrow			LPIPS \downarrow		
	BP	QAP	Subject	BP	QAP	Subject	BP	QAP	Subject
SPWI (Input)	17.25	16.05	14.74	0.217	0.218	0.222	0.437	0.436	0.447
Li et al. [15]	20.28	19.03	18.93	0.310	0.313	0.260	0.635	0.590	0.635
U-Net [17]	14.50	13.13	14.80	0.115	0.086	0.110	0.625	0.624	0.628
ResU-Net w/o DWT	17.17	17.62	16.78	0.314	0.351	0.243	0.461	0.430	0.485
ResU-Net	17.68	17.76	17.24	0.301	0.336	0.223	0.469	0.435	0.459

TABLE III

EXPERIMENTAL RESULTS FOR EVALUATING LATERAL RESOLUTION AND CONTRAST, WHERE THE VALUES IN BOLD INDICATE THE HIGHEST VALUES FOR EACH METRIC.

Method	FWHM_t [%] \uparrow	ΔFWHM [μm] \uparrow	CR [dB] \uparrow	GCNR \uparrow
SPWI (Input)	76.38	—	4.95	0.379
CPWC (GT)	93.17	—	7.34	0.521
Li et al. [15]	83.27	-36.28	6.31	0.612
U-Net [17]	92.32	217.841	4.41	0.410
ResU-Net w/o DWT	92.83	133.375	8.35	0.477
ResU-Net	89.98	93.17	7.54	0.445

TABLE IV

COMPARISON OF FLOPS, THE NUMBER OF PARAMETERS, AND INFERENCE TIME OF EACH METHOD.

Method	FLOPS	# of parameters	Inference time [ms]	
			GPU	CPU
Li et al. [15]	139.57B	13.40M	54.64	487.5
U-Net [17]	217.11B	31.04M	55.45	740.8
ResU-Net w/o DWT	32.49B	24.35M	54.49	183.0
ResU-Net	8.53B	24.37M	55.09	53.12

method. For evaluating the image quality, we employ Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM) [26], and Learned Perceptual Image Patch Similarity (LPIPS) [27]. Higher values for PSNR and SSIM indicate higher image quality, while lower values for LPIPS indicate higher image quality. LPIPS is known as an image quality evaluation metric that is closer to human perception compared to PSNR and SSIM [27]. For evaluating the lateral resolution, we employ FWHM_t and ΔFWHM , which are based on the Full Width at Half Maximum (FWHM). For the 4,949 wire targets in QAP, FWHM is the width of the pixel that is -6dB from the peak value of the point spread function in the horizontal direction. FWHM_t is the ratio of the number of wire targets whose FWHM is below the threshold t [μm] out of 4,949 wire targets. In the experiments, we set $t = 1,580\mu\text{m}$ to take into account the width of the adjacent wire targets. ΔFWHM is the difference of FWHM between the single plane-wave image and the generated image for a wire target whose FWHM is less than the threshold t [μm] in the single plane-wave image, the compound image, and the generated image. For evaluating the contrast, we employ Contrast Ratio (CR) and Generalized Contrast-to-Noise Ratio (GCNR) [28]. In the experiments, CR and GCNR are calculated for the surrounding regions of 813 grayscale targets in QAP, and their average are used for evaluation.

D. Results and Discussion

Table II shows the results of quantitative evaluation of image quality for each method. The conventional method has the highest PSNR. Since PSNR evaluates image quality based on the root-mean-square of the differences in the pixel areas of the generated and compound images, the conventional method, which is trained using the L1 loss between the images, has the advantage. For SSIM, ResU-Net w/o DWT is the highest for BP and QAP, and the conventional method is the highest for Subject. For LPIPS, ResU-Net w/o DWT is the lowest for BP and QAP, and ResU-Net is the lowest for Subject. Comparing U-Net and ResU-Net w/o DWT, ResNet is suitable as an encoder for improving the quality of SPWI, since ResU-Net w/o DWT is better on all metrics. ResU-Net w/o DWT and ResU-Net exhibit similar performance on all metrics. Considering that LPIPS is the closest to human subjective evaluation among the three evaluation metrics [27], ResU-Net w/o DWT and ResU-Net can produce ultrasound images with quality close to that of compound images. Table III shows the results of the quantitative evaluation of the lateral resolution and contrast for each method. The conventional method is lower than the other methods in all metrics except for GCNR. ResU-Net w/o DWT performs slightly better than ResU-Net on all metrics. Table IV shows FLOPS, the number of parameters, and the inference time on GPU and CPU for each method. ResU-Net w/o DWT and ResU-Net achieve fewer FLOPs than the other methods. ResU-Net reduces FLOPS to about $\frac{1}{4}$ compared to ResU-Net w/o DWT. The inference time on GPU is similar for all methods, while ResU-Net achieves the shortest inference time on CPU at 53.12 ms. ResU-Net is particularly suitable for general ultrasound systems, which are often configured with a CPU. From the above results, we have demonstrated that the loss functions that take into account the characteristics of the RF signal can produce high-quality ultrasound images, and that the use of DWT can reduce the computational complexity and inference time.

IV. CONCLUSION

We proposed a method for improving the performance of SPWI using U-Net and DTW. The proposed method uses IQ data as the input and output of U-Net and to represent the high frequency component of the ultrasound signals and loss functions that take into account the effect of the point spread and the frequency characteristics of the RF signals. We also reduce the computational complexity and inference time by inputting the IQ data after DWT into U-Net. We demonstrated

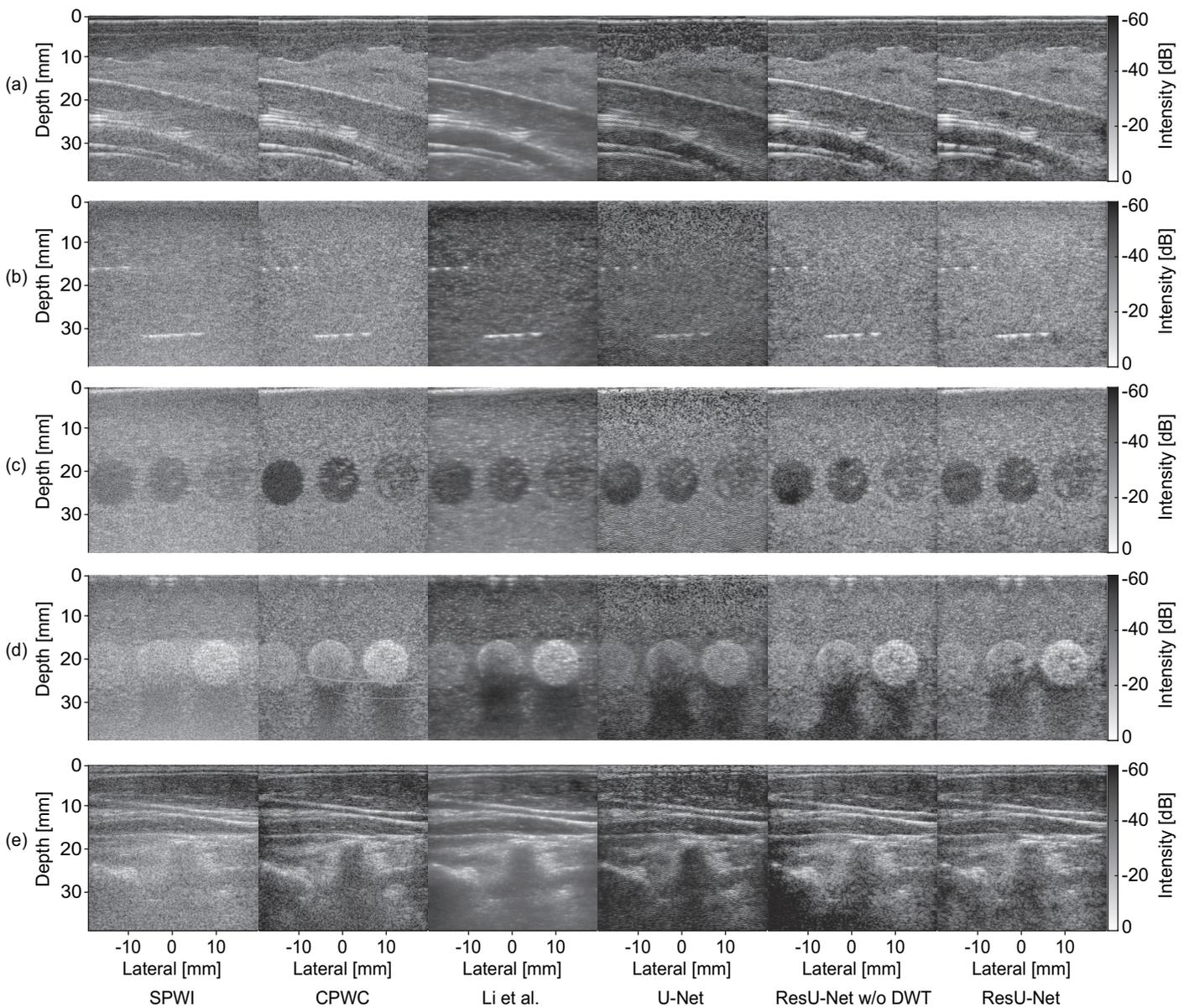


Fig. 2. Examples of ultrasound images generated by each method: (a) BP, (b) wire target in QAP, (c) grayscale target in QAP, (d) grayscale target in QAP, and (e) Subject.

the effectiveness of the proposed method through a set of experiments using our image dataset. In the future work, we will apply the ultrasound images whose quality is improved by the proposed method to the analysis of dynamic organs such as the heart.

REFERENCES

- [1] T. L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out*. Academic Press, 2004.
- [2] P. Hoskins, K. Martin, and A. Thrush, *Diagnostic Ultrasound: Physics and Equipment*. CRC Press, 2019.
- [3] J.-Y. Lu, H. Zou, and J. F. Greenleaf, "Biomedical ultrasound beam forming," *Ultrasound in Medicine and Biology*, vol. 20, no. 5, pp. 403–428, Apr. 1994.
- [4] J. Bercoff, M. Tanter, and M. Fink, "Supersonic shear imaging: A new technique for soft tissue elasticity mapping," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 51, no. 4, pp. 396–409, Apr. 2004.
- [5] M. Fink and M. Tanter, "Multiwave imaging and super resolution," *Physics Today*, vol. 63, no. 2, pp. 28–33, Feb. 2010.
- [6] E. Macé, G. Montaldo, I. Cohen, M. Baulac, M. Fink, and M. Tanter, "Functional ultrasound imaging of the brain," *Nature Methods*, vol. 8, pp. 662–664, Jul. 2011.
- [7] T. Deffieux, C. Dmené, and M. Tanter, "Functional ultrasound imaging: A new imaging modality for neu-

- rosience,” *Neuroscience*, vol. 474, pp. 110–121, Oct. 2021.
- [8] J. Udesen, F. Gran, K. L. Hansen, J. A. Jensen, C. Thomsen, and M. B. Nielsen, “High frame-rate blood vector velocity imaging using plane waves: Simulations and preliminary experiments,” *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 55, no. 8, pp. 1729–1743, Aug. 2008.
- [9] K. L. Hansen, J. Udesen, F. Gran, J. A. Jensen, and M. Bachmann Nielsen, “In-vivo examples of flow patterns with the fast vector velocity ultrasound method,” *Ultraschall in der Medizin*, vol. 30, no. 5, pp. 471–477, Sep. 2009.
- [10] J. Bercoff, G. Montaldo, T. Loupas, *et al.*, “Ultrafast compound doppler imaging: Providing full blood flow characterization,” *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 58, no. 1, pp. 134–147, Jan. 2011.
- [11] I. K. Elroll, A. Swillens, P. Segers, T. Dahl, H. Torp, and L. Lovstakken, “Simultaneous quantification of flow and tissue velocities based on multi-angle plane wave imaging,” *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 60, no. 4, pp. 727–738, Apr. 2013.
- [12] O. Coutere, M. Fink, and M. Tanter, “Ultrasound contrast plane wave imaging,” *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 59, no. 12, pp. 2676–2683, Dec. 2012.
- [13] M. Tanter and M. Fink, “Ultrafast imaging in biomedical ultrasound,” *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 61, no. 1, pp. 102–119, Jan. 2014.
- [14] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, “Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography,” *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 3, pp. 489–506, Mar. 2009.
- [15] Z. Li, A. Wiacek, and M. A. L. Bell, “Beamforming with deep learning from single plane wave RF data,” *Proc. IEEE Int’l Ultrasonics Symp.*, pp. 1–4, Sep. 2020.
- [16] D. Perdios, M. Vonlanthen, F. Martinez, M. Arditì, and J.-P. Thiran, “CNN-based image reconstruction method for ultrafast ultrasound imaging,” *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 4, pp. 1154–1168, Nov. 2022.
- [17] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” *Proc. Int’l Conf. Medical Image Computing and Computer Assisted Intervention*, vol. 9351, pp. 234–241, Oct. 2015.
- [18] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *Proc. Int’l. Conf. Learning Representations*, pp. 1–14, May 2015.
- [19] C. Chen, H. H. G. Hansen, G. A. G. M. Hendriks, J. Menssen, J.-Y. Lu, and C. L. de Korte, “Point spread function formation in plane-wave imaging: A theoretical approximation in fourier migration,” *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 2, pp. 296–307, Feb. 2020.
- [20] T. Guo, T. Zhang, E. Lim, M. Lopez-Benitez, F. Ma, and L. Yu, “A review of wavelet analysis and its applications: Challenges and opportunities,” *IEEE Access*, vol. 10, pp. 58 869–58 903, 2022.
- [21] P. Namsopa, S. Koonkarnkhai, P. Kovintavewat, H. Dinsakul, and S. Suwansawang, “A study and exploration of discrete wavelet transform for speckle noise reduction in ultrasound images,” *Proc. Int’l Technical Conf. Circuits/Systems, Computers, and Communications*, pp. 1–5, Jun. 2023.
- [22] S. -. Yu, J. -. He, Y. Pan, Y. Wang, J. Wen, and R. Yan, “The application of texture classification network with hybrid adaptive wavelet in graves’ disease ultrasound diagnosis,” *Proc. Int’l Conf. Wavelet Analysis and Pattern Recognition*, pp. 38–43, Jul. 2023.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 770–778, Jun. 2016.
- [24] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” *Proc. ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, pp. 2623–2631, Jul. 2019.
- [25] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” *Proc. Int’l Conf. Learning Representations*, pp. 1–10, 2019.
- [26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [27] R. Zhang, P. Isola, A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, pp. 586–595, Jun. 2018.
- [28] A. Rodriguez-Molares, O. M. H. Rindal, J. D’hooge, *et al.*, “The generalized contrast-to-noise ratio: A formal definition for lesion detectability,” *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 4, pp. 745–759, Apr. 2020.