High-quality Reconstruction of Plane-wave Imaging Using Generative Adversarial Network

Xi Zhang[1], Jing Liu[1], Qiong He[1], Heye Zhang[2], Jianwen Luo[1]
[1]Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China
[2]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
Email: luo_jianwen@tsinghua.edu.cn

***Abstract*—Coherent plane wave compounding (PWC) using tens of steered plane waves (PWs) can obtain high-quality ultrasound images but reduces the gain in frame rate. Recently a new strategy using convolutional neural network (CNN) was proposed to recover high-quality images from only 3 PWs. Considering the excellent performance of generative adversarial network (GAN) in image reconstruction, we propose to use GAN to reconstruct high-quality ultrasound images from 3 PWs. Phantom and *in vivo* experiments are performed. The results of GAN using 3 PWs, in terms of contrast ratio and lateral resolution are competitive with those of CNN using 3 PWs and coherent compounding using 31PWs, which demonstrates the feasibility of this method.**

Keywords—contrast, frame rate, generative adversarial network, lateral resolution, plane wave compounding

I. Introduction

Ultrasound imaging is a widely-used modality in medical diagnosis because it is nonionizing, noninvasive, potable and cost-effective. In clinical applications, classical focused transmissions are used to form a number of scan lines, which limits the frame rate to tens of frames per second (fps). With the development of ultrasound imaging, high frame rate is demanded in real-time 3-D imaging, shear wave tracking and motion estimation of cardiovascular tissues. Plane-wave imaging (PWI) was proposed, which can illuminate the region of interest with a single one transmission, leading to a frame rate at thousands of fps [1].

However, PWI obtains low quality images with low contrast ratio (CR), contrast-to-noise ratio (CNR) and lateral resolution (LR) because of the lack of focusing. Coherent plane-wave compounding (PWC) with multi-angle transmissions was proposed to produce high-quality image, at the cost of a decreased frame rate [1]. In order to find a better trade-off between the image quality and frame rate, convolutional neural network (CNN) was recently proposed to reconstruct high-quality images with only 3 PWs [2], competing with the standard compounding of 31 PWs in terms of CR, CNR and LR.

Recently, generative adversarial network (GAN) [3] is frequently used in a wide variety of applications such as image generation [4], representation learning [5], image manipulation [6], object detection [7] and video applications [8]. For example, in photo-realistic single image super-solution, GAN has achieved splendid results, because it has advantages in reconstructing realistic texture details compared with CNN [9]. Therefore, we investigate the feasibility of GAN in recovering high-quality ultrasound images from a small number of PWs though phantom and *in vivo* experiments.

This study was supported in part by the National Key R&D Program of China (2016YFC0102200) and the National Natural Science Foundation of China (81471665, 81561168023, 81561168023 and 61871251).

II. Methods

We aimed to adopt GAN to reconstruct a high-quality ultrasound image I^H from the input I^L which consists of a small number of PWs. I^{CH} is the standard compounding images which are only used as reference (i.e., label) during the training process. Here I^L is a real-valued tensor of size $W \times H \times C$, while I^H and I^{CH} are of size $W \times H$. W denotes the width of the beamformed image, H denotes the height of the beamformed image and C denotes the number of the beamformed images corresponding to different PWs. Considering that there may be useful information which is not exploited by standard compounding from I^L for reconstructing a high-quality image I^H, we tried to learn from data with an adequate model [2].

A GAN consists of two networks: the generator network G_{θ_G} parametrized by θ_G and the discriminator network D_{θ_D} parametrized by θ_D. Both the generator and the discriminator are feed-forward CNN. The objective of GAN is to train a generator G_{θ_G} with the goal of confusing a discriminator D_{θ_D}. And the discriminator D_{θ_D} is trained to distinguish the reference image I^{CH} from the recovered high-quality image I^H. Therefore, we optimized the D_{θ_D} and G_{θ_G} to solve the following problem [9]:

$$\min_{\theta_G}\max_{\theta_D} \mathrm{E}_{I^H\sim p_{train(I^H)}}\left[log D_{\theta_D}(I^H)\right] + \mathrm{E}_{I^L\sim p_{G(I^L)}}\left[log(1-D_{\theta_D}(G_{\theta_G}(I^L))\right] \quad (1)$$

In our GAN, we trained a generator network G_{θ_G} to recover I^H from the input I^L. Assuming I_n^L with the corresponding $I_n^{CH}(n=1,\dots,N)$ are the training images, we solved the following equation [9]:

$$\hat{\theta}_G = \arg\min_{\theta_G}\frac{1}{N}\sum_{n=1}^{N} l_{sum}(G_{\theta_G}(I_n^L), I_n^{CH}) \quad (2)$$

where l_{sum} is a loss function. In this study, it is a combination of two loss components, which will be discussed in detail in Section II.E. Meanwhile, we also trained a discriminator network to distinguish the reconstructed images

from the real high-quality images. We thus aimed to minimize the binary cross entropy:

$$\hat{\theta}_D = \arg\max_{\theta_D} \frac{1}{N} \sum_{n=1}^{N} (\log (D_{\theta_D}(I_n^{CH})) + \log (1 - D_{\theta_D}(G_{\theta_G}(I_n^L))) \quad (3)$$

A. Residual Block

We adopted residual blocks in our GAN. The structure of residual block is shown in Fig. 1. The residual block can be expressed as [10]:

$$y = \sigma(\mathcal{F}(x, \{W_i, b_i\}) + x) \quad (4)$$

where x and y are the input and output of the residual block, respectively. In Fig. 1, the function $\mathcal{F}(x, \{W_i, b_i\})$ can be defined as $\mathcal{F} = W_2\sigma(W_1 x + b_1)$. Here σ denotes the activation function, such as ReLU (the rectified linear unit) and Leaky ReLU (the leaky rectified linear unit). b_1 is the bias. The operation $\mathcal{F}(x, \{W_i, b_i\}) + x$ is realized by a shortcut connection and element-wise addition. Compared with other deep learning networks with the same number of parameters, depth and width, the shortcut connections cannot increase either the number of parameters or the computational complexity. Furthermore, the shortcut connections can also address vanishing/exploding gradients which limit the design of very deep networks [10].

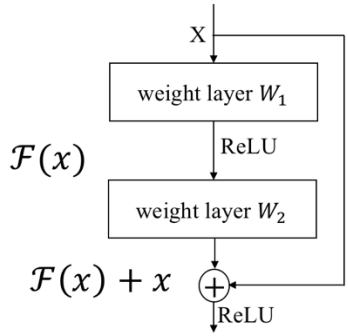

Fig. 1. The structure of residual block.

B. ReLU And Leaky ReLU

In this study, we used ReLU and Leaky ReLU as the activation function defined as [11]:

$$f(x) = \begin{cases} x, & x < 0 \\ ax, & x \geq 0 \end{cases} \quad (5)$$

Here x is the input of the nonlinear activation function f. a is a coefficient controlling the slope of the negative part, and was set as a small fixed value of 0.1. When a is equal to 0, it represents ReLU. In a very deep networks, the usage of Leaky ReLU can address the zero gradients [11]. Moreover, it can expedite the convergence of very deep models. Fig. 2 shows the shapes of ReLU and Leaky ReLU.

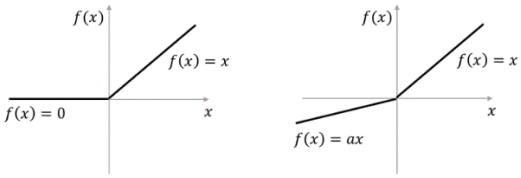

Fig. 2. The active function of ReLU vs. Leaky ReLU. respectively.

C. Batch Normalization

In this study, we used the Batch Normalization (BN) transform before the nonlinearity, which is expressed as [12]:

$$z = \sigma(\mathrm{BN}(Wx + b)) \quad (6)$$

where W and b are the learned parameters of the model, and $\sigma(\cdot)$ is the nonlinearity activation function. BN can stabilize learning by normalizing the input to each unit to have zero mean and unit variance. Generally, it can address problems caused by poor initialization. An example of the problems is that the optimal results get stuck in poor local minima [12]. It can also avoid vanishing/exploding gradients, which helps gradient flow in deep models.

D. Adversarial Network Architecture

In this study, we followed the architecture of GAN proposed in [9] and slightly modified to adapt to our problem. On one hand, considering the excellent performance of GAN, this method can potentially obtain better result. On the other hand, it conforms to the architecture guidelines for stable GAN proposed in [13].

Specifically, the framework of the GAN model we used is illustrated in Fig. 3. In the generative network G_{θ_G}, we mainly used 50 residual blocks each of which consisted of two convolutional layers with 5×3 kernels and 64 features maps followed by batch-normalization layers, and ReLU was used as the activation function. As for the discriminative network D_{θ_D}, it had eight convolutional layers with the filter kernels increasing from 64 to 512 by a factor of 2. The last 512 features produced by the eighth layers were followed by two dense layers, which was used to obtain the possible classification.

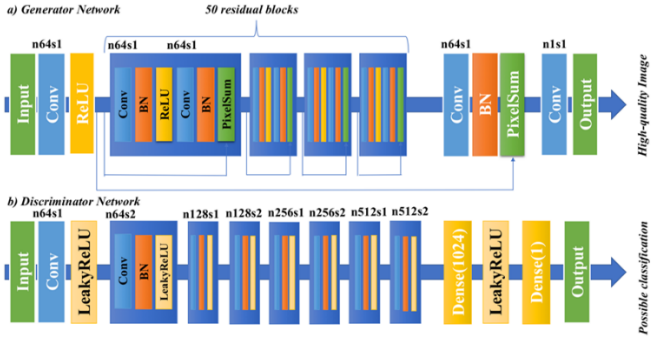

Fig. 3. (a) Generator network. (b) Discriminator network. n64s1 denotes 64 feature maps (n) and stride 1 (s) for each convolutional layer.

E. Loss Function

The design of the loss function l_{sum}, a combination of the generator loss l_p and the weighted adversarial loss l_d, is critical to the performance of the GAN. By adding the adversarial loss l_d, the images with high texture details can be produced by GAN [9].

$$l_{sum} = l_p + \lambda l_d \tag{7}$$

where λ is weighted number. Here we took a pixel-wise MSE (mean squared error) loss as the generator loss, which is defined as [9]:

$$l_p = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} (I_{x,y}^{CH} - G_{\theta_G}(I^L)_{x,y})^2 \tag{8}$$

The adversarial loss l_d is defined as:

$$l_d = \sum_{n=1}^{N} -\log D_{\theta_D}(G_{\theta_G}(I^L)) \tag{9}$$

Here $D_{\theta_D}(G_{\theta_G}(I^L))$ means the possibility of $G_{\theta_G}(I^L)$ being a real high-quality image. Using $-\log D_{\theta_D}(G_{\theta_G}(I^L))$ instead of $-\log[1 - D_{\theta_D}(G_{\theta_G}(I^L))]$ can improve the optimization process [9].

F. Data Acquisition

A Vantage 256 system (Verasonics Inc., Redmond, WA, USA) equipped with an L10-5 linear array (f_0 = 7.5 MHz) was moved evenly on the surface of the imaging objects to acquired 6,530 frames of channel data at a frame rate of 50 Hz, each of which contained 31 steered PWs (-15° to 15° with 1° steps). 2,030 frames of channel data were acquired from a tissue-mimicking phantom (model 040GSE, CIRS, Norfolk, VA, USA), 500 from swine muscles *ex vivo*, 2,000 from carotid artery of a healthy human subject and 2,000 from brachioradialis of another healthy human subject *in vivo*. We first acquired 2,000 frames from the phantom, which were used for training. Then we randomly acquired 30 frames from the same phantom, which were used for testing. Moreover, 500 frames were acquired from swine muscles for training. 1,500 frames were acquired from the longitudinal section of the carotid artery of a human subject for training, and 500 frames were acquired from the cross section of the carotid artery for testing. Finally, 2,000 frames were acquired from the brachioradialis of another human subject for training. All the channel data were beamformed with delay-and-sum algorithm.

In summary, 6,000 frames were used as the training data, including the input I^L and label I^{CH}. I^L included three images of beamformed RF data from steering angles of 0° and ±15°, respectively. I^{CH} was the beamformed RF data obtained by PWC with 31 PWs (PWC-31). The remaining 530 frames were used as the testing data, with the three images of beamformed RF data corresponding to the above 3 PWs as the input.

G. Training Details

Our GAN was trained by minimizing the loss function via stochastic gradient descent with Adam optimizer [9] with β_1 = 0.9. The training was performed on a PC workstation (CPU//RAM) equipped with a Tesla P100 GPU (NVIDIA, Santa Clara, CA). Its learning rate was 0.0001 with no update. The value of a in Leaky ReLU was a fixed value of 0.1. Considering that the generator G_{θ_G} and the discriminator D_{θ_D} cannot converge synchronously, we pre-trained both of them with the same parameters, which stabilized the training process and could be seen as initialization to avoid undesired local optimum. The whole training process took about 15 days with 3.6 $\times 10^6$ update iterations. During testing time, we turned the batch normalization update off to obtain an output that deterministically depends only on the input [9]. Moreover, we also trained the convolutional network (CNN) proposed in [2] for comparison. The training process of CNN took about 2 days with 2.4 $\times 10^6$ iterations.

III. Results

The images obtained by GAN with 3 PWs were compared with those by CNN with 3 PWs, and PWC with 3 PWs (PWC-3) and 31 PWs (PWC-31), respectively. The full width at half maximum (FWHM) of a wire target at 40 mm depth (yellow arrow in Fig. 4(a)), the contrast ratio (CR) and contrast-to-noise ratio (CNR) between a cystic region and the background (red circles in Fig. 4(a), r = 1 mm) in the phantom were quantified.

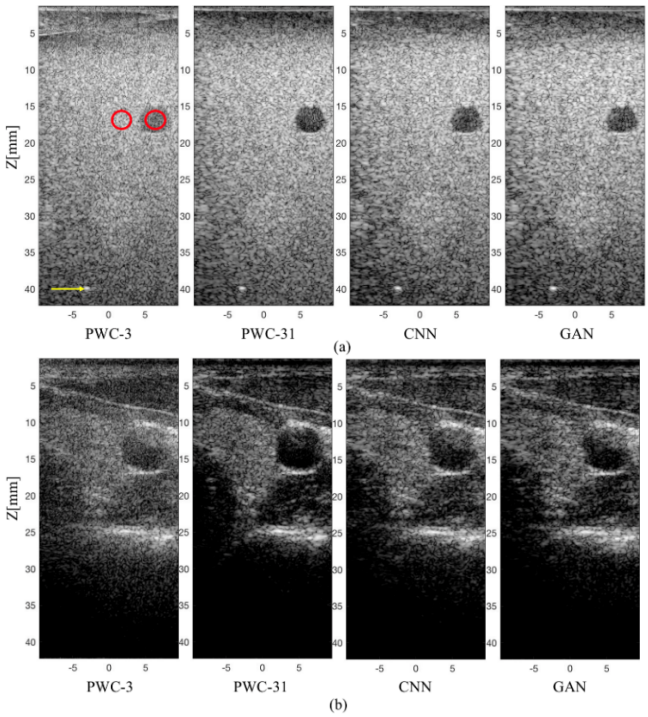

Fig. 4. B-mode images of (a) the phantom and (b) carotid artery using PWC-3, PWC-31, CNN (3 PWs) and GAN (3 PWs), respectively.

The B-mode images of the phantom and carotid artery show that GAN obtains better performance than PWC-3 (Figs. 4(a) and 4(b)), and obtains similar images to CNN and PWC-31. Quantitatively, as shown in TABLE I, the FWHMs of GAN, PWC-3 and PWC-31 are close, which are smaller than that of CNN. The CR of GAN is higher than those of PWC-3 and CNN, and is slightly lower than that of PWC-31. The CNR of GAN is close to that of CNN, lower than that of

PWC-31, and higher than that of PWC-3. The results show that GAN can reconstruct high-quality ultrasound images from 3 PWs.

TABLE I. The FWHM, CR and CNR of Different Methods

	PWC-3	PWC-31	CNN	GAN
FWHM (mm)	0.50	0.53	0.56	0.53
CR (dB)	10.23	19.63	19.08	19.46
CNR (dB)	1.30	2.43	2.26	2.25

IV. Discussion

The performance of the GAN may be further improved by optimizing its architecture and the loss function. More residual blocks could be used to obtain a deeper network, which may achieve better performance but at high computational cost [15]. Moreover, the learned feature representations in the discriminator could be used as the basis for the reconstruction objective to obtain better results in terms of visual fidelity [16].

3 PWs were used for the input of the GAN in this study. In the future, we could investigate whether only 1 PW can reconstruct high-quality ultrasound image.

V. Conclusion

We have developed a generative adversarial network (GAN) with residual blocks that can reconstruct high-quality ultrasound image from transmission of 3 plane waves. The results show that GAN is a promising method in ultrasound image reconstruction.

References

[1] 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. Ultrason. Ferroelectr. Freq. Control*, vol. 56, no. 3, pp. 489-506, 2009.

[2] M. Gasse, F. Millioz, E. Roux, D. Garcia, H. Liebgott and D. Friboulet, "High-quality plane wave compounding using convolutional neural networks," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 64, no. 10, pp. 1637-1639, 2017.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative adversarial nets," in *Proc. Advances in Neural Information Processing Systems*, 2014, pp. 2672-2680.

[4] M. Arjovsky, S. Chintala and L. Bottou, "Wasserstein gan," in *Proc. International Conference on Machine Learning (ICML)*, 2017, pp. 214-223

[5] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford and X. Chen, "Improved techniques for training gans," in *Proc. Advances in Neural Information Processing Systems*, 2016, pp. 2234-2242.

[6] J.-Y. Zhu, P. Krähenbühl, E. Shechtman and A. A. Efros, "Generative visual manipulation on the natural image manifold," in *Proc. European Conference on Computer Vision*, 2016, pp. 597-613.

[7] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng and S. Yan, "Perceptual generative adversarial networks for small object detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1951-1959

[8] M. Mathieu, C. Couprie and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," *arXiv preprint arXiv:1511.05440*, 2015.

[9] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 105-114.

[10] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

[11] K. He, X. Zhang, S. Ren and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proc. IEEE International Conference on Computer Vision*, 2015, pp. 1026-1034.

[12] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network. training by reducing internal covariate shift," in *Proc. International Conference on Machine Learning*, 2015, pp. 448-456.

[13] A. Radford, L. Metz and S. Chintala, "Unsupervised representation. learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[14] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1-9.

[16] A. B. L. Larsen, S. K. Sønderby, H. Larochelle and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," *arXiv preprint arXiv:1512.09300*, 2015.