A Conditional Adversarial Network for Single Plane Wave Beamforming

Yaning Wang[1], Kelley Kempski[2], Jin U. Kang[1], and Muyinatu A. Lediju Bell[1,2,3]
[1]Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA
[2]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
[3]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

***Abstract*—Compared to traditional focused transmissions, plane wave (PW) ultrasound imaging has the potential to enable higher frame rates, which is clinically relevant to real-time applications and ultrafast imaging. However, when reducing the number of PW transmissions to reduce image formation times, PW imaging is confounded by image quality degradation, acoustic clutter, and speckle noise. To tackle this challenge, we present a deep learning-based method to analyze raw radiofrequency (RF) channel data acquired by the ultrasound probe and convert this signal to the final B-mode image, bypassing the traditional beamforming procedure. The deep learning architecture for this approach relies on a conditional generative adversarial network (cGAN), in which the generative model and classifying model work simultaneously to produce an indistinguishable output from a ground truth. The cGAN was trained to predict B-mode images that look like beamformed PW results after multiple insonifications. This network was trained and tested utilizing a publicly accessible PICMUS database composed of *in vivo* and *ex vivo* ultrasound inclusions with randomly distributed scatterers in various combinations. The proposed method produces signal-to-noise ratio (SNR) enhancements from 1.112 to 1.540 when compared with conventional delay-and-sum (DAS) beamforming of a single PW insonification. The cross-correlation coefficient between a 75 PW image and cGAN-predicted data was 0.976, an improvement over the 0.641 obtained when the 75 PW image was cross-correlated with a DAS PW image created after a single insonification. These results demonstrate the potential of generative adversarial networks to substitute traditional DAS beamforming in future applications.**

***Index Terms*—Deep Learning, Conditional Adversarial Networks, Beamforming, Ultrafast Ultrasound Imaging**

I. Introduction

Improving both frame rates and image quality are often two goals for next-generation medical ultrasound imaging systems, with much attention dedicated to plane wave ultrasound imaging to maximize these criteria [1]. Compared to conventional focused insonification, plane wave (PW) imaging achieves high frame rates equal to thousands of images per second, which has promising potential for electromechanical wave imaging, high sensitivity Doppler imaging, and ultrasonic imaging of brain activity [2]. However, reducing the number of plane wave transmissions to achieve even faster real-time imaging tends to suffer from decreased resolution and contrast, thus reducing overall image quality. In addition, several approaches have been attempted to eliminate the influence of speckle noise and acoustic clutter such as coherent plane-wave compounding [2], while minimum variance adaptive beamforming [3] and short-lag spatial coherence (SLSC) beamforming [4] have been explored to compensate image degradation and improve image resolution.

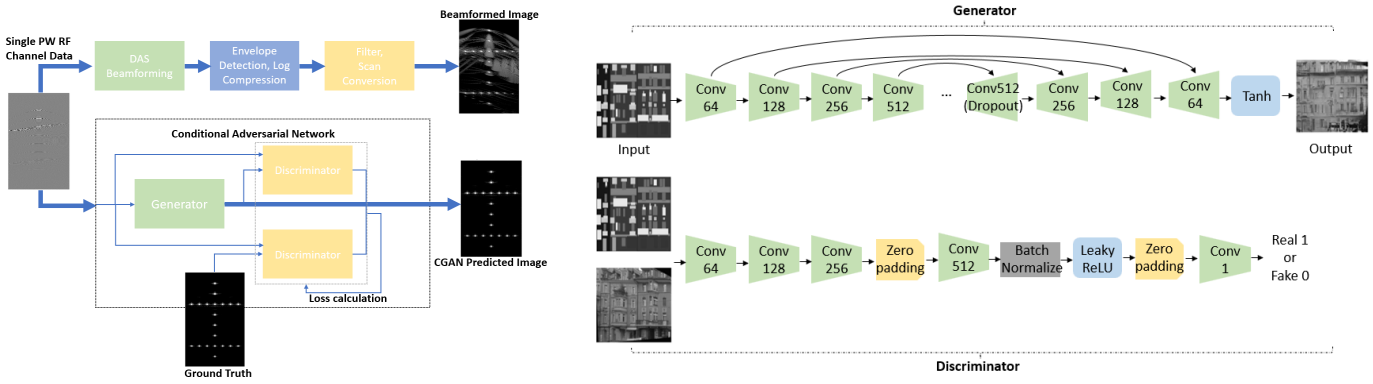
In contrast to conventional model-based mathematical techniques, there has been a growing interest in applying convolutional neural networks (CNNs) to ultrasound beamforming. This application has already shown promise in medical image analysis with regard to feature extraction, classification, and restoration [5]. Perdios *et al.* [6] used CNNs to learn the non-linear mapping between single and multiple plane wave reconstructed data to enhance image quality following delay-and-sum (DAS) beamforming. To minimize latency and computational cost, Nair *et al.* [7] trained a CNN model to transform raw RF channel data to a segmented mask, excluding DAS beamforming and other post-processing methods (e.g., envelope detection, log compression, filter, and scan conversion).

A generative adversarial network (GAN) was later explored to simultaneously create ultrasound images comparable to DAS beamformed images and segment corresponding cysts from surrounding tissue [8]. In this paper, we demonstrate the use of a conditional GAN (cGAN) [9] to directly transform RF channel data to a B-mode ultrasound image. Specifically, we utilize a cGAN with two discriminators. This adversarial network has already shown strong potential in computer graphics (including object transfiguration and image enhancement [10], [11]) and in ultrasound imaging [12]. In this paper, we employ RF channel data as an input signal for cGAN-based alternative method and evaluate the reconstruction results with the high quality image given by multiple PW transmissions. Results were submitted for comparison and evaluation in the Challenge on Ultrasound Beamforming with Deep Learning (CUBDL) [13], [14].

II. Methods

A. Dataset

We pre-trained our neural network using 400 pairs of CMP Facedes datasets [15] including the graphics transformation from cartoon architectural labels to real photos, shown in Fig. 1(b). Image pre-processing included grayscale conversion of colored images. We augmented the pre-training data by applying various geometric transformations (i.e., flipping, cutting, rotating) to increase the diversity of the training data. The training dataset was generated from 1500 single PW ultrasound images from PICMUS [1] and the MATLAB



(a) (b)

Fig. 1: (a) Overview of the proposed cGAN method to substitute conventional DAS beamforming. (b) Generator and discriminator architecture of the proposed cGAN. Black arrows represent concatenation.

TABLE I: PWI Configuration Parameters

Parameter	L11-4v
Element number	128
Pitch	300 μm
Center frequency	5.133 MHz
Bandwidth	67%
Element width	0.27 mm
Transmit frequency	5.208 MHz
Speed of sound	1540 m/s
Sampling frequency	20.832 MHz

ultrasound toolbox (USTB) [16]. The configuration parameters are described in Table I. To increase the robustness of our network, both simulation and experimental ultrasound images were used containing various distributions of 9 to 20 point scatters, 1 to 9 anechoic or hypoechoic cysts, or *in vivo* carotid artery longitudinal or circular cross-sections.

The simulations were generated using Field II [17], [18], and the transmission angle was zero degrees. Experimental PICMUS data [1] were acquired with a Verasonics Vantage 256 ultrasound system and a linear L11-4v ultrasound probe. To further improve the network robustness and avoid overfitting, -2 dB Gaussian white noise was added to each single plane wave RF signal. The ground truth images for training were formed after incorporating the 75 plane wave transmissions to create each corresponding DAS image. The ground truth images were post-processed using the steps described in the top of Fig. 1(a). All training and ground truth images were normalized to the maximum signal amplitude. In addition, the training and ground truth images were resized (using downsampling with linear interpolation) from approximately 1500 $\times$128 samples to 256$\times$256 samples. The entire dataset was divided into dedicated training (60%), validation (20%), and test (20%) datasets.

B. Network Architecture

The network architecture of cGAN is detailed in Fig. 1(b) including one generator and two discriminators. The objective of cGAN is to learn a nonlinear mapping from observed images (x) and randomly distributed noise (z) to an output image (y), given by the following equation.

$$L_{cGAN}(G,D) = E_{x,y}[logD(x,y)] + E_{x,z}[log(1-D(x,G(x,z)))] \quad (1)$$

where G is the generator which attempts to produce predicted images that are close to the ground truth, D is the discriminator which regards the images as real rather than fake, and E is the log probability of D. In contrast to unconditional GAN (uGAN) [19], cGAN has a symmetric network structure and the input signal is evaluated both by the generator and discriminator.

The architecture of the generator is based on U-Net [20], which is widely applied in medical image segmentation. The outline of the contraction path is composed of convolutional layers, Batchnorm layers, and Leaky ReLU layers aimed at capturing features. The expansion path contains transposed

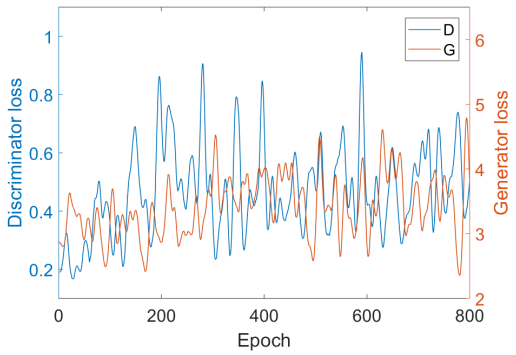


Fig. 2: cGAN performance on pretraining dataset for the discriminator loss (D) and generator loss (G).

convolutional layers, Batchnorm layers, Dropout layers, and ReLU layers for precise localization. In addition, there is a concatenation between the mirrored encoder and decoder blocks. The discriminator has an analogous design to the contraction path. For each loop, the discriminator works twice and calculates the corresponding cross-entropy loss between the input, ground truth, and generated images. An optimal parameter λ was chosen after training the network iteratively, while L1 loss alone ($\lambda = 0$) led to blurry results.

The total loss of the cGAN is described by the equation:

$$L_{total} = L_{cGAN} + \lambda L_{L1} \tag{2}$$

The initial value of λ was defined as 100 [9]. Adam optimization algorithm was applied with 800 epochs for pre-training, 800 epochs for training, and 200 epochs for fine-tuning, while the initial learning rate was 0.0002. The hardware support was Google Colab GPU with TensorFlow while the total runtime was 15 hours.

To illustrate the final reconstructed results, the network-generated images were compared to the corresponding ground truth images by calculating the mutual information, described by the following equation:

$$I(X;Y) = H(X,Y) - H(X|Y) - H(Y|X) \tag{3}$$

where $H(X)$ and $H(Y)$ are the marginal entropies, $H(X|Y)$ and $H(Y|X)$ are the conditional entropies. A high mutual information indicates a strong correlation between the two images. In addition, we also assess the beamformed image quality by calculating SNR, CNR, cross-correlation coefficient, and joint entropy of a single CIRS phantom image and a single *in vivo* image. For the CIRS phantom image, square regions of interest (ROI) with 5 mm side length were chosen in the center of the hyperechoic cyst for the target region with a background ROI laterally offset to the right by 20 mm. For the *in vivo* image, the target ROI was centered in the carotid artery, and the background ROI was laterally offset to the right by 15 mm.

III. Results

Fig. 2 shows the discriminator and generator losses as a function of the number of epochs for the pre-training dataset. Because neither the generator nor the discriminator computes the cross-entropy between predicted images and ground truth images directly, they each tended to fluctuate competitively along with epochs. As learning progressed, the total L1 loss decreased.

Fig. 3(a-c) shows a single PW DAS beamformed image, a 75 PW DAS images, and the cGAN-generated image for the CIRS phantom, from left to right, respectively. Similarly, Fig. 3(d-f) shows a single PW DAS beamformed image, 75 PW DAS image, and the cGAN-generated image for a circular cross-section of an *in vivo* carotid artery, from left

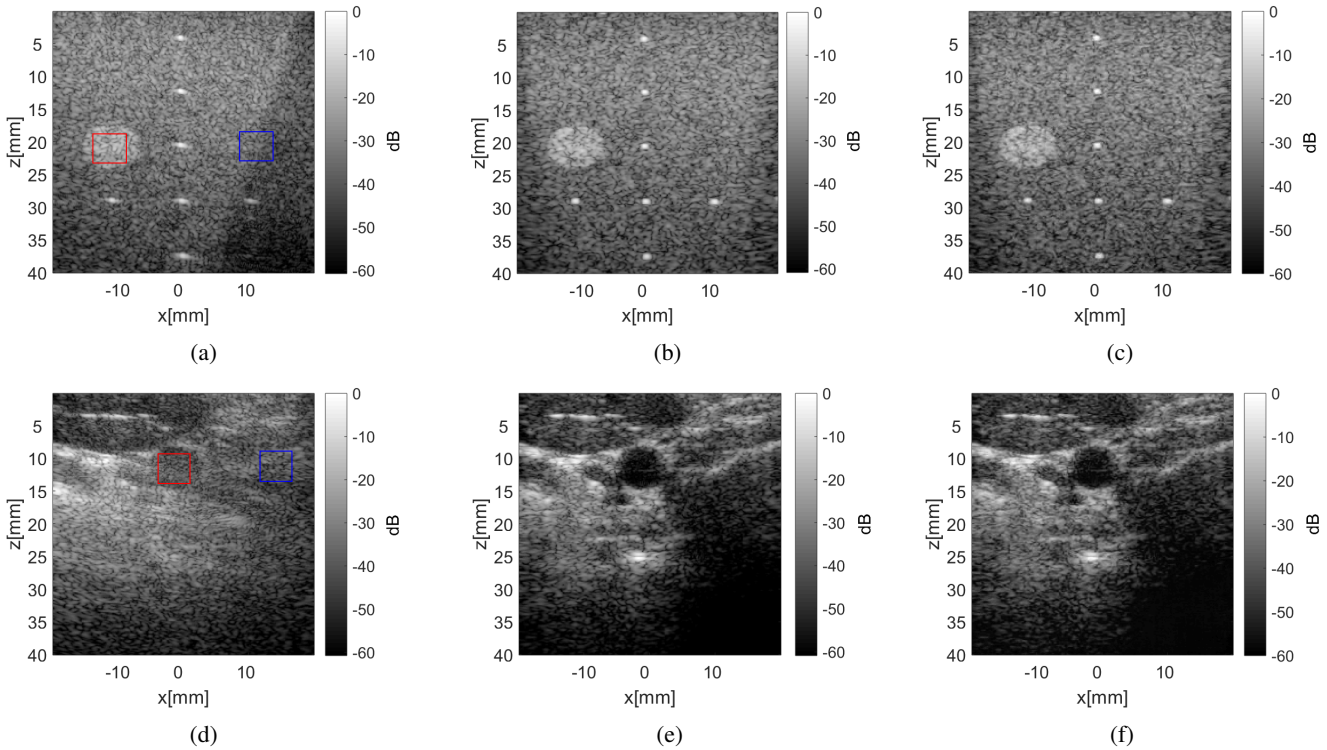


Fig. 3: B-mode images of (a-c) a CIRS phantom and (d-f) a carotid artery cross-section, produced with (a,d) a single plane wave insonification and DAS beamforming, (b,e) 75 multiple plane wave insonifications and DAS beamforming, and (c,f) a single plane wave insonification and the cGAN prediction. The target and background ROIs are outlined in red and blue, respectively.

TABLE II: Image Evaluation Results

Test case	Method	CNR	SNR	corr coef	joint entropy	mutual information
Carotid	single DAS	0.404	0.715	0.698	13.415	0.620
	multi DAS	1.565	1.558	1	6.652	6.652
	cGAN	1.600	1.606	0.987	11.278	2.530
CIRS phantom	single DAS	2.733	1.509	0.583	13.977	0.447
	multi DAS	1.594	1.463	1	7.138	7.138
	cGAN	1.715	1.474	0.966	11.978	2.229

to right, respectively. It is remarkable that with only 1 PW insonification, the images generated by the cGAN present less clutter and other artifacts in comparison to matched single PW DAS results.

The image quality of the single PW insonification DAS and cGAN results were compared against the multiple (i.e., 75) PW DAS results, which were used as the ground truth. Results are presented in Table II. Both *in vivo* and *ex vivo* RF channel data achieve considerable image quality enhancements with cGAN. The test dataset required 452 ms to generate one B-mode image of size 256×256.

IV. Discussion & Conclusion

We presented an alternative approach to process raw RF ultrasound PW channel data using a cGAN. This cGAN employs two regressive discriminators to compute the image relevancy, which is more compatible than conventional CNN U-Net designs to address complex information transfer. The cGAN successfully learned the nonlinear mapping between channel signal and high-quality B-mode PW images, without further post-processing (e.g., delays computation, envelope detection). We trained and tested our network with the PICMUS PW dataset [1] containing experimental and simulated cysts, resulting in cGANs that generated less artifacts with improved correlation coefficients, while preserving details present in DAS B-mode images created with multiple PW insonifications. The total computational cost and corresponding accuracy both show the promise of using and improving cGANs for future ultrasound-based tasks.

Acknowledgment

The authors acknowledge the support of NIH Trailblazer Award R21-EB025621.

References

[1] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. A. Jensen, and O. Bernard. Plane-wave imaging challenge in medical ultrasound. In *2016 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2016.

[2] M. Tanter and M. Fink. Ultrafast imaging in biomedical ultrasound. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 61(1):102–119, 2014.

[3] A. Austeng, C. C. Nilsen, A. C. Jensen, S. P. Näsholm, and S. Holm. Coherent plane-wave compounding and minimum variance beamforming. In *2011 IEEE International Ultrasonics Symposium*, pages 2448–2451, 2011.

[4] M. A. Lediju, G. E. Trahey, B. C. Byram, and J. J. Dahl. Short-lag spatial coherence of backscattered echoes: imaging characteristics. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 58(7):1377–1388, 2011.

[5] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen. Medical image classification with convolutional neural network. In *2014 13th International Conference on Control Automation Robotics Vision (ICARCV)*, pages 844–848, 2014.

[6] D. Perdios, M. Vonlanthen, A. Besson, F. Martinez, M. Arditi, and J. Thiran. Deep convolutional neural network for ultrasound image enhancement. In *2018 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2018.

[7] A. A. Nair, T. D. Tran, A. Reiter, and M. A. Lediju Bell. A deep learning based alternative to beamforming ultrasound images. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3359–3363, 2018.

[8] A. A. Nair, T. D. Tran, A. Reiter, and M. A. L. Bell. A generative adversarial neural network for beamforming ultrasound images : Invited presentation. In *2019 53rd Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6, 2019.

[9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[10] Xinyuan Chen, Chang Xu, Xiaokang Yang, and Dacheng Tao. Attention-gan for object transfiguration in wild images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[11] Jiahui Wei, Yuan-Ping Chen, and Peng Ren. Image enhancement with conditional adversarial networks. 2017.

[12] Ouwen Huang, Will Long, Nick Bottenus, Marcelo Lerendegui, Gregg E Trahey, Sina Farsiu, and Mark L Palmeri. Mimicknet, mimicking clinical image post-processing under black-box constraints. *IEEE Transactions on Medical Imaging*, 39(6):2277–2286, 2020.

[13] M.A.L. Bell, J. Huang, D. Hyun, Y. C. Eldar, R. van Sloun, and M. Mischi. Challenge on Ultrasound Beamforming with Deep Learning (CUBDL). In *Proceedings of the 2020 IEEE International Ultrasonics Symposium*, 2020.

[14] "Challenge on Ultrasound Beamforming with Deep Learning (CUBDL)". IEEE Dataport, 2019. [online]. Avaliable:http://dx.doi.org/10.21227/f0hn-8f92.

[15] Radim Tyleček and Radim Šára. Spatial pattern templates for recognition of objects with regular structure. In Joachim Weickert, Matthias Hein, and Bernt Schiele, editors, *Pattern Recognition*, pages 364–374, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[16] A. Rodriguez-Molares, O. M. H. Rindal, O. Bernard, A. Nair, M. A. Lediju Bell, H. Liebgott, A. Austeng, and L. Lvstakken. The ultrasound toolbox. In *2017 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2017.

[17] Jørgen Arendt Jensen. Field: A program for simulating ultrasound systems. In *10th Nordicbaltic Conference on Biomedical Imaging*, volume 4, pages 351–353, 1996.

[18] J. A. Jensen and N. B. Svendsen. Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 39(2):262–267, 1992.

[19] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.