A Conditional Adversarial Network for Single Plane Wave Beamforming

Yaning Wang*, Kelley Kempski†, Jin U. Kang*, and Muyinatu A. Lediju Bell*,†
*Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD
†Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD

***Abstract*—Compared to traditional focused transmissions, plane wave (PW) ultrasound imaging could enable higher frame rates by over a hundred fold, which is clinically relevant to real-time applications and ultrafast imaging. However, along with reducing acquisition time, PW imaging is confounded by image quality degradation, acoustic clutter, and speckle noise. To tackle this problem, 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 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 plane wave image and cGAN-predicted data is 0.976, compared to 0.641 with DAS beamforming of a single PW insonification. These results demonstrate the feasibility of using this adversarial network to substitute traditional DAS beamforming in future applications.**

***Keywords*—Deep learning, conditional adversarial networks, beamforming, ultrafast ultrasound imaging**

I. Introduction

Improving frame rates and image quality are two essential components of medical ultrasound imaging improvements, with much attention dedicated to plane wave ultrasound imaging to maximize these criteria [1]. Compared to conventional focused insonification, PW imaging achieves high frame rates equal to thousands of images per second [2], which has promising potential for electromechanical wave imaging, high sensitivity Doppler imaging, and ultrasonic imaging of brain activity [3]. However, the combination of plane wave imaging and real-time imaging suffers from decreased resolution and contrast, thus reducing overall image quality. Several approaches have been attempted to eliminate the influence of speckle noise and acoustic clutter such as coherent plane-wave compounding [3]. In addition, minimum variance adaptive beamformer [4] and short-lag spatial coherence (SLSC) [5] have been explored to compensate image degradation and improve image resolution.

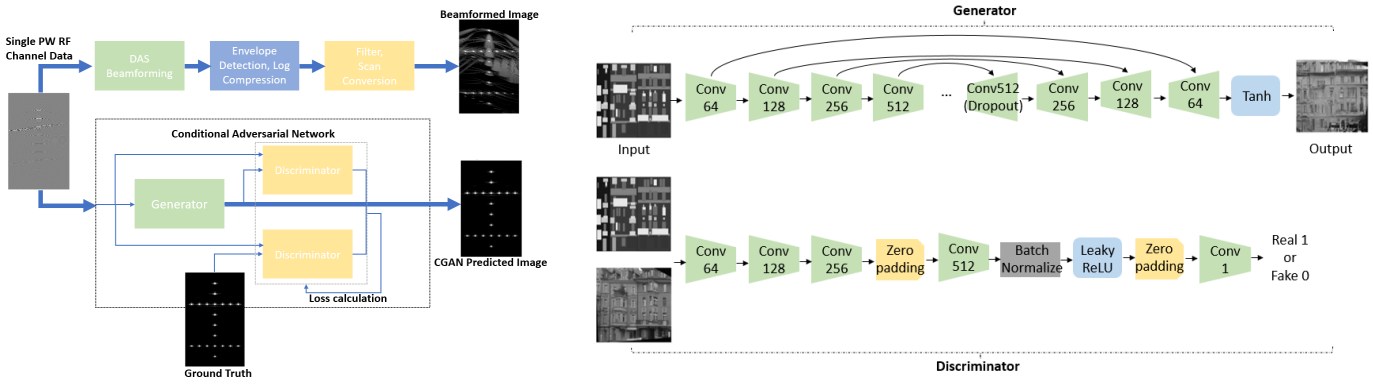
In contrast to conventional model-based mathematical technologies there has been a growing interest to apply convolutional neural networks (CNNs) to ultrasound beamforming. This application has already shown promise in medical image analysis with regard to feature extraction, classification, and restoration. 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 DAS beamforming. To minimize latency and computational cost, Nair *et al.* [2] trained a CNN model transforming raw RF channel data to manually segmented mask, excluding DAS and other post-processing methods (e.g., envelope detection, log compression, and filter, interpreted in Fig. 1a). This work has been expanded by simulating a single cyst in tissue insonified by single PW and adding robot controlled ultrasound probe enabled volumetric reconstruction [7]. A GAN was additionally explored as an ultrasound image formation method to match a DAS beamformed B-mode image and segment corresponding cyst from surrounding tissue [8]. In addition, deep neural networks (DNNs) were employed to obtain information directly from raw channel data and to simultaneously generate both a segmentation map and B-mode image for automated ultrasound tasks [9], [10].

In this paper, we demonstrate the use of a conditional generative adversarial network (cGAN) [11] for information translation from RF channel data directly to a B-mode ultrasound image. Specifically, we utilize cGAN with two discriminators. This adversarial network has already shown strong potential in computer graphics including translating images from aerial to map, or labels to facades. In this work, 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.

II. Methods

A. Dataset

We pre-trained our neural network using 400 pairs of CMP Facedes datasets [12] including the graphics transform from cartoon architectural labels to real photos, shown in Fig. 1b. 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 ultrasound toolbox



(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

(USTB) [13]. 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 longitudinal or cross-sections.

The simulation inclusions were generated through Field II [14], [15] and the transmission angle was zero degrees. The 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. 75 multi-plane wave DAS beamformed images were used as ground truth images for training. The ground truth images were post-processed using the steps described in the Fig. 1(a) pipeline. All training and ground truth images were normalized to the maximum signal intensity. In addition, the training and ground truth images were resized via downsampling and linear interpolation from approximately 1500×128 to 256×256. The dataset was divided into training (60%), validation (20%), and testing (20%).

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.

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

The generator attempts to produce predicted images that are close to the ground truth, such that the discriminator regards the images as real rather than fake. In contrast to unconditional GAN (uGAN) [16], 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 [17], 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 expansive path contains transposed convolutional layers, Batchnorm layers, Dropout layers, and ReLU layers for precise localization. Furthermore, there is

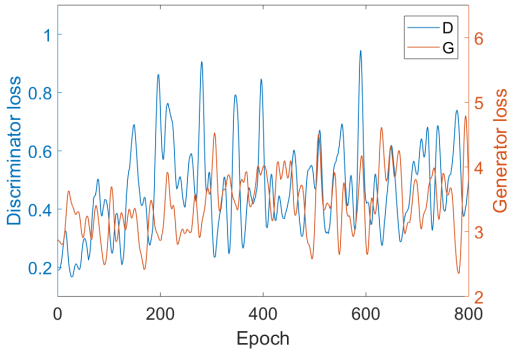


Fig. 2: cGAN loss curve on a pretraining dataset

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 images, and generated images. We compare its reconstruction accuracy when λ changes, where L1 loss alone ($\lambda = 0$) produce 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 [11]. We applied Adam optimization algorithm with 800 epochs for pre-training, 800 epochs for training, and 200 epochs for fine-tuning [18], 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 the 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 $H(X|Y)$ 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

A. B-mode image reconstruction

Fig. 2 shows the discriminator loss and the generator loss 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 tend to fluctuate competitively along with epochs. As learning progresses, the total L1 loss decreases.

Fig. 3(a-c) shows a single PW DAS beamformed image, a 75 PW DAS image, 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 to

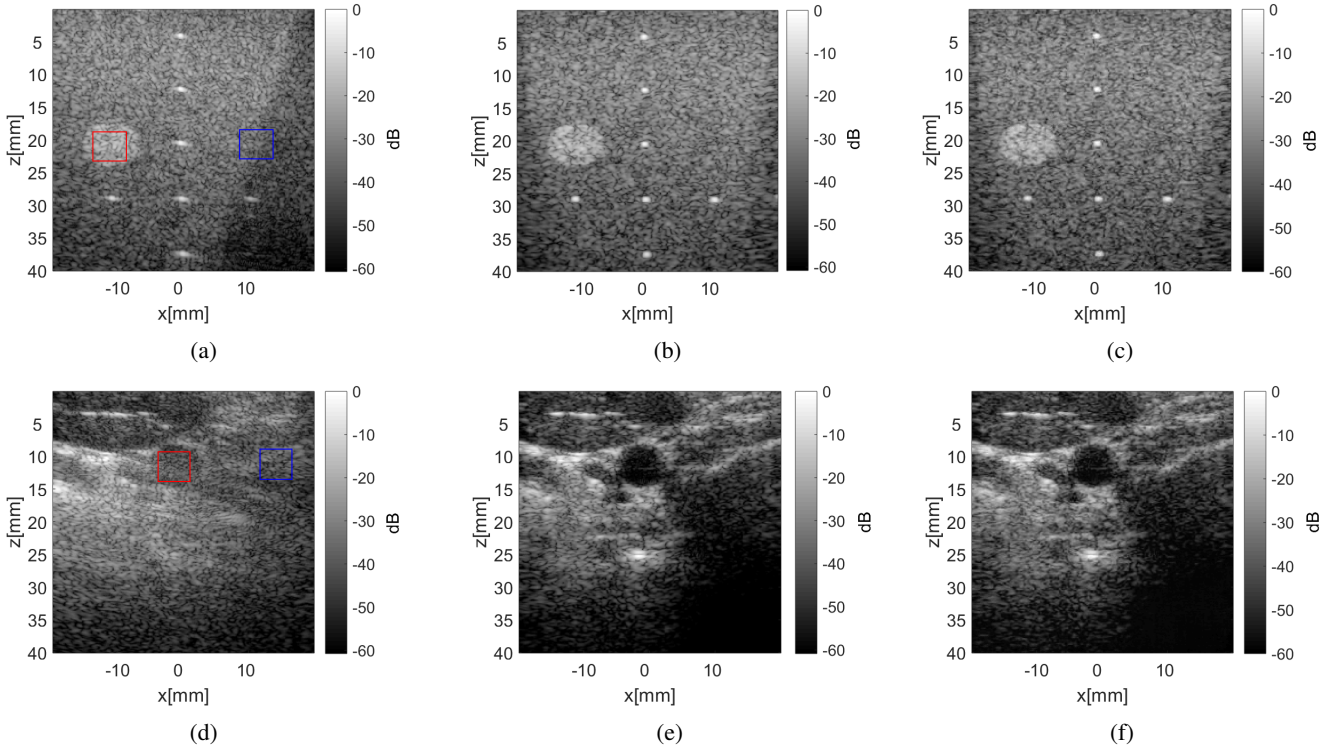


Fig. 3: B-mode images of (a-c) a CIRS phantom and (d-e) a carotid artery cross-section, produced with (a,d) a single plane wave insonification and DAS beamforming, (b,e) 75 multiple plane wave insonifications and DAS beamforming, and (c,f) a single plane wave insonification and the cGAN prediction. The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

right, respectively. Formed 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.

B. Performance evaluation

The image quality of the single PW insonification DAS and cGAN results were compared against the while multiple 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 enhancements with cGAN. The test dataset required 452ms 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 conditional adversarial networks. Compared to CNNs, cGAN employs two regressive discriminators to compute the image relevancy, which is more compatible 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 cGAN for future utilization, such as ultrafast imaging, real-time robotic tracking, and *in vivo* activity monitoring.

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] A. A. Nair, M. R. Gubbi, T. Duy Tran, A. Reiter, and M. A. Lediju Bell. A fully convolutional neural network for beamforming ultrasound images. In *2018 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4, 2018.

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

[4] 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.

[5] 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.

[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] A. A. Nair, K. N. Washington, T. D. Tran, A. Reiter, and M. A. L. Bell. Deep learning to obtain simultaneous image and segmentation outputs from a single input of raw ultrasound channel data. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, pages 1–1, 2020.

[10] A. A. Nair, T. D. Tran, A. Reiter, and M. A. Lediju Bell. One-step deep learning approach to ultrasound image formation and image segmentation with a fully convolutional neural network. In *2019 IEEE International Ultrasonics Symposium (IUS)*, pages 1481–1484, 2019.

[11] 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.

[12] 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.

[13] 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.

[14] Jørgen Arendt Jensen. Field: A program for simulating ultrasound systems. In *10TH NORDICBALTIC CONFERENCE ON BIOMEDICAL IMAGING, VOL. 4, SUPPLEMENT 1, PART 1:351–353*, pages 351–353, 1996.

[15] 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.

[16] 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.

[17] 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.

[18] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, 2016.