Beamforming with deep learning from single plane wave RF data

Zehua Li*, Alycen Wiacek*, 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
‡Department of Computer Science, Johns Hopkins University, Baltimore, MD

***Abstract*—Deep learning approaches for improving ultrasound image reconstruction have proven successful in both experimental and clinical settings. In this paper, we present an autoencoder-based deep learning framework for ultrasound beamforming from the radio-frequency (RF) data received after a single plane wave transmission. Motivated by U-Net, the network consists of an encoder and a decoder. The network was trained and evaluated on simulated, phantom, and *in vivo* datasets. When tested on simulated data, the mean SNR, contrast, and gCNR of the learned image results were 3.16, -35.96 dB and 1.0 respectively, as well as a mean PSNR of 18.61 dB when compared to enhanced B-mode images. Each of these metrics outperformed the standard delay-and-sum (DAS) beamforming algorithm for the single plane wave image. In addition, the network was evaluated on an *in vivo* breast mass, achieving improved image quality compared to the corresponding single plane wave image. These results highlight the promise of exploring the proposed network to generate high quality ultrasound images from one plane wave, which could be applied to multiple ultrasound-based clinical tasks.**

***Index Terms*—Ultrasound, Deep Learning, Convolutional Neural Network, Single Plane Wave, Image Generation**

I. Introduction

Ultrasound is a widespread imaging modality in current clinical medicine, and it is one of the most commonly used diagnostic tools in specialties such as cardiology [1]. Compared with other commonly used imaging modalities, such as magnetic resonance imaging (MRI) and computed tomography (CT), ultrasound imaging is real-time, portable, inexpensive, and free of ionizing radiation risks [2]. Because of these significant advantages, ultrasound has drawn a large amount of research attention over the past decades and considerable effort has been made to improve the quality of ultrasound images, including resolution, contrast, and signal-to-noise ratios. One important method to improve image quality is beamforming. Some beamforming options include delay and sum (DAS) and short-lag spatial coherence (SLSC) imaging [3]. However, these options either suffer from poor image quality or high computational complexity, which is not compatible with fast real-time imaging.

Beamforming with deep learning has recently drawn a large amount of research interest. Many deep learning algorithms have been developed to increase the potential for high-quality, ultrafast ultrasound imaging in clinical applications. Of particular interest is single plane wave imaging, which is capable of displaying images at high frame rates at the expense of poor image quality. Conversely, to improve image quality, it is possible to acquire images with multiple plane waves, at the cost of speed. Instead, deep learning may potentially be utilized to form ultrasound images with image quality that is comparable to multi-plane wave images, while retaining high frame rates.

For example, Nair *et al.* [4] proposed a U-Net based deep convolutional network with a VGG-13 encoder to directly segment cyst targets and simultaneously display images using raw ultrasound data, with the potential to reduce computational costs by skipping intermediate beamforming steps. Hyun *et al.* [5] proposed a different CNN-based framework to beamform ultrasound channel signals into speckle-reduced images. Although these and other approaches have demonstrated the capability of using deep learning to improve ultrasound beamforming, such approaches emphasize image quality rather than speed. In this paper, we present an autoencoder-based deep neural network (DNN) for ultrasound beamforming [4], capable of converting the raw radio-frequency (RF) data from one single plane wave and producing an image comparable to multi-plane wave images. This network was submitted to the Challenge on Ultrasound Beamforming with Deep Learning (CUBDL) [6], [7] at the 2020 IEEE International Ultrasonics Symposium to evaluate image quality and speed.

II. Methods

A. Network Implementation

Motivated by the U-Net [8] and VGG-13 [9] architectures, and similar to [4], our network directly takes the radio-frequency (RF) data as input to generate high quality ultrasound images as the output. The complete DNN structure is shown in Fig. 1. There are two major parts of the DNN: (1) the encoder and (2) the decoder.

The encoder block employs convolutional, maxpooling, batch normalization, and leaky rectified linear unit (LeakyReLU) layers. The convolutional layers were implemented with a 3×3 kernel, stride of 1, and 1 point of zero padding for each dimension. Batch normalization layers were used to standardize the inputs of each layer, and also to reduce overfitting. The slope of the LeakyReLu was set to 0.2 [10]. The decoder block employs convolutional, up-convolutional, batch normalization, and rectified linear units (ReLU) layers. The kernel size of the up-convolutional layer was 2×2 with a stride of 2. The activation function of the last convolutional layer was hyperbolic tangent (tanh), which was used to enforce output values in the range [-1, 1].

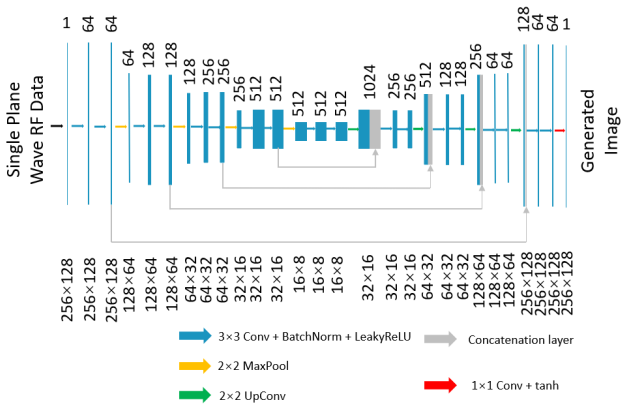

Fig. 1. DNN architecture for ultrasound image generation.

With the aim of generating high quality images, we used an L1 loss function (Mean Absolute Error) to train the network. The network was trained with the Adam optimization algorithm [11], a weight decay penalty of $1\text{x}10^{-4}$, and a learning rate of $2\text{x}10^{-4}$. Based on empirical testing, we chose a mini-batch size of 16. The network was written in Pytorch [12] and trained on 4 NVIDIA P40 GPUs in parallel.

B. Network Training Datasets

The network was trained and tested using a combination of three datasets: (1) the PICMUS (Plane-wave Imaging Challenge in Medical Ultrasound) dataset [13], (2) the Field II [14], [15] dataset from [4], and (3) an *in vivo* breast dataset from 4 different patients. The PICMUS training data [13] included two simulated datasets and two phantom datasets. For each example, a set of 75 steered plane waves with angles from -16° to 16° was either simulated or acquired. The input to the network was the raw RF data associated with one plane wave (i.e., any angle from -16° to 16°), and the ground truth was the corresponding image formed with 75 plane wave transmissions, delay-and-sum beamforming, and plane wave compounding. The RF test data set used plane wave transmission angles that were not included during training.

The Field II dataset was created by Nair *et al.* [4], and consists of a single anechoic cyst within a homogeneous tissue background. The raw single plane wave RF data were acquired from a single angle of 0° and used as the input to the network. To eliminate clutter within the anechoic cyst, the ground truth for the Field II dataset was an enhanced image as described in [4], where the pixel values of the anechoic regions were set to zero while preserving the pixel values of the surrounding tissue. Images were beamformed using the Ultrasound Toolbox [16] and displayed with 60 dB dynamic range after log compression.

The *in vivo* breast ultrasound dataset was created from 4 different patients with two orthogonal scans (i.e., radial and anti-radial) per patient. The data were acquired with an Alpinion ECUBE-12R research ultrasound scanner connected to either an L8-17 or L3-8 linear array ultrasound transducer. These *in vivo data* were acquired after informed consent and approval from the Johns Hopkins Medicine Institutional Review Board. The raw single plane wave RF data were acquired from different plane wave transmission angles within the range of -16° to 16°, and the ground truth was the same as described for the PICMUS dataset.

The complete dataset (i.e., the PICMUS, Field II, and *in vivo* breast datasets) consisted of 1,300 total training examples, 138 total validation examples, and 127 total testing examples. For the PICMUS and *in vivo* datasets, the channel data images were flipped, left to right, to augment the dataset. Preparation of the complete dataset for input to the network followed a series of preprocessing steps. First, the RF channel data were cropped in the axial direction according to the axial range of the ground truth image. Second, to reduce network computational complexity and speed up the training process, the amplitude of the final, log-compressed, 60-dB-dynamic-range, beamformed image was normalized to the range [0,1] (similar to the compression process reported in [4]). Therefore, the raw output of the network is considered to be compressed. Finally, the RF channel data corresponding to each single plane wave transmission was paired with the corresponding enhanced image and both were downsampled to size 256×128. The network was trained for 1000 epochs with the PICMUS, Field II, and breast datasets.

In order to reverse the compression and regain an image with values that are suitable for image quality assessments, the image was uncompressed with the following equation:

$$s = 10^{s_{dB}/20} \tag{1}$$

where s is the uncompressed pixel value, and s_{dB} is the compressed pixel with values ranging [0,1], as described in the preceding paragraph.

C. Image Quality Metrics

The following image quality metrics were used to evaluate and compare the image quality of the single plane wave, DNN, and ground truth images:

1) Signal-to-noise ratio (SNR): SNR is a measure of the smoothness of the tissue region and is defined as:

$$SNR = \frac{S_o}{\sigma_o} \tag{2}$$

where S_o and σ_o are the mean and standard deviation, respectively, of a background region of interest (ROI), measured for both the compressed and uncompressed DNN outputs.

2) Contrast: Contrast is a measure of the signal intensity ratio between the cyst and the tissue and is defined as:

$$Contrast = 20\log_{10}\left(\frac{S_o}{S_i}\right) \tag{3}$$

where S_i and S_o are the mean signal ampltudes within ROIs inside and outside cysts, respectively.

3) Generalized contrast-to-noise ratio (gCNR): gCNR is used to measure lesion detectability [17] and is defined as:

$$gCNR = 1 - \sum_{x=0}^{1} \min_{x}\{p_i(x), p_o(x)\} \tag{4}$$

where $p_i(x)$ and $p_o(x)$ are the probability mass functions of the ROIs that produce S_i and S_o, respectively.

4) Peak signal-to-noise ratio (PSNR): PSNR is a global measure of similarity between the ground truth image and either the single plane wave image or the DNN image. PSNR is defined as:

$$PSNR = 10\log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right) \tag{5}$$

where MAX is the maximum possible pixel value of the image and MSE is the mean square error between two images.

5) Full width at half maximum (FWHM): The lateral and axial FWHM of point targets were measured to quantify resolution.

III. Results and Discussion

Fig. 2(a) shows from left to right, respectively, corresponding single plane wave, DNN, and ground truth images. The top row shows one example from the Field II dataset, and the bottom row shows one example from the PICMUS dataset. When considering the images created from the Field II dataset (i.e., the top row of Fig. 2(a)), the background of the DNN image is smoother than that of the single-plane wave image, and the anechoic cyst contains less acoustic clutter in the DNN image than the single-plane wave image. The cyst in the DNN image is qualitatively more similar to the enhanced image than the single plane wave image, suggesting that with a goal of improved cyst visualization, our DNN can generate relatively high quality images with only one plane wave.

The PICMUS dataset (i.e., the bottom row of Fig. 2(a)) shows point targets and a hyperechoic cyst. The DNN image shows a blurred version of the hyperechoic cyst and point targets, likely due to the lack of hyperechoic targets during training. Quantitatively, the mean axial and lateral FWHM for the single plane wave image were 0.67 mm and 0.94 mm, respectively. The mean axial and lateral FWHM for the DNN image were 0.88 mm and 1.56 mm, respectively. Finally, the mean axial and lateral FWHM for the ground truth image were 0.64 and 0.67 mm, respectively. The increased FWHM in the DNN image is likely due to the lack of hyperechoic targets in the training data. Therefore, the resolution and performance of the DNN on hyperechoic targets can potentially be improved by including additional data with point and other hyperechoic targets during training.

Fig. 2(b) shows from left to right, respectively, a single plane wave image, the DNN testing result, and the corresponding ground truth image for one *in vivo* breast mass, which was not included during training. The cyst in the DNN image is significantly better visualized than the single plane wave image and shows a similar general position compared to the ground truth image. However, the DNN image is smoother, with less dynamic range compared to the ground truth image, and the cyst appears stretched in the DNN image compared to the ground truth. The mean uncompressed SNR values for the single plane, DNN, and ground truth images were 1.14, 2.30, and 1.30, respectively. The corresponding mean contrast values were 1.06 dB, -13.94 dB, and -14.18 dB, respectively (where a more negative contrast is better), and the corresponding mean gCNR values were 0.38, 0.99, and 1.0, respectively.

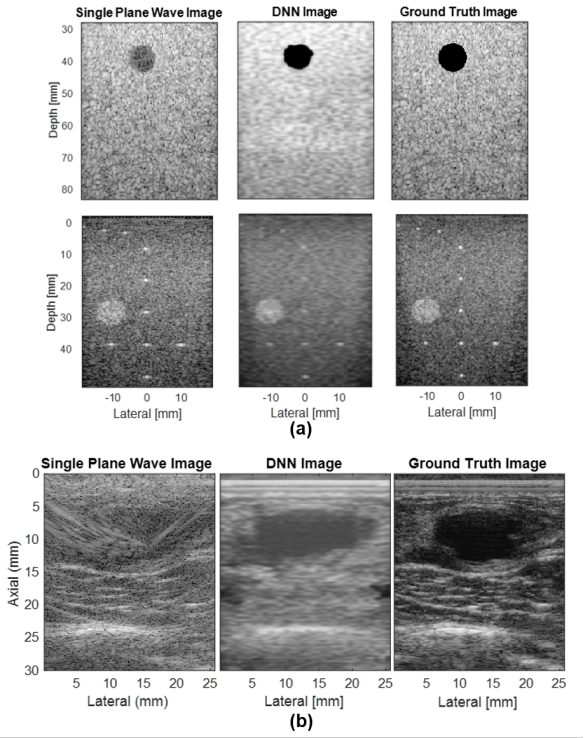

Fig. 2. (a) Example test results from the single plane wave DAS beamformed image, the DNN output image, and the ground truth image of (top) one example from the Field II dataset and (bottom) one example from the PICMUS dataset. (b) Example test results from the *in vivo* breast dataset. Each image is displayed with a 60 dB dynamic range.

These results demonstrate an improvement in SNR, contrast, and gCNR for the DNN image when compared to the single plane wave image. In addition, the mean PSNR values for the single plane wave and DNN images were 14.78 dB and 13.52 dB, respectively. The performance of the network on *in vivo* examples can potentially be further improved with the inclusion of additional *in vivo* data in the training dataset.

Fig. 3 shows the SNR, contrast, and gCNR from the Field II dataset as a function of the axial and lateral cyst position in the image for each of the single plane wave, DNN, and corresponding ground truth images. The SNR of the single plane wave images is not included, because it is the same as the enhanced image. The SNR of the compressed DNN output is higher than the SNR of the uncompressed DNN output across all lateral and axial positions. The mean compressed SNR of the DNN image in the lateral and axial directions were 14.91 and 12.91, respectively. The mean uncompressed SNR of the DNN image in the axial and lateral directions were 3.21 and 3.10, respectively, resulting in a global mean SNR of 3.16. Both the compressed and uncompressed DNN outputs are higher than the ground truth (i.e., the enhanced image) with mean SNR of 1.74 and 1.82 in the axial and lateral directions, respectively, which means the background of the DNN output image is smoother than that of the ground truth image.

The mean contrast of the uncompressed DNN image in the axial and lateral directions were -39.35 dB and -32.56 dB, respectively, resulting in a global mean contrast of -35.96 dB. For all lateral and axial cyst positions shown in Fig. 3,

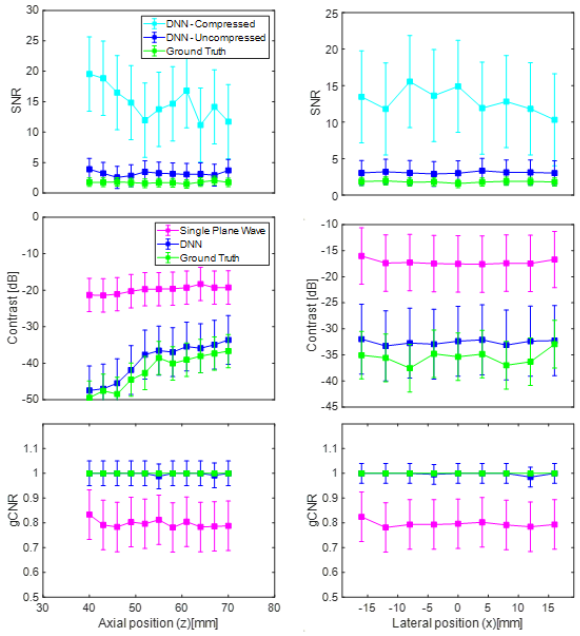

Fig. 3. Mean $\pm$ standard deviation of SNR, contrast, and gCNR as a function of axial (z) and lateral (x) cyst positions from the Field II dataset and measured from 127 testing examples. SNR was measured from both compressed and uncompressed DNN output images as described in Section II-B. The remaining metrics are measured from the uncompressed DNN output image.

the contrast of the DNN image is improved compared to the single plane wave image, and is within one standard deviation of the contrast of the ground truth image. The mean gCNR of the uncompressed DNN for all lateral and axial positions was approximately 1, which was larger than the gCNR of the single plane wave image with a mean gCNR of 0.80. The mean PSNR of the DNN image for all axial and lateral cyst positions was 18.61 dB indicating the similarity between the DNN image and the ground truth image across all cyst positions.

Overall, the DNN images show improved quantitative image quality results compared to the single plane wave images, although there is room for improvement, particularly for the hyperechoic and *in vivo* examples. Paired with the qualitatively improved images of anechoic cysts, these results suggest that the presented network could help to overcome the image quality trade-off of single plane wave imaging. In addition, because the proposed approach bypasses the need for multiple plane waves and plane wave compounding, this approach has the potential to be faster than a multi-plane wave approach.

IV. Conclusions

This paper describes a DNN architecture motivated by the U-Net and VGG-13 architectures. The DNN is capable of directly generating high quality ultrasound images from RF data, bypassing the delay-and-sum and plane wave compounding processes. The DNN output images generally achieve improved image quality metrics compared to single plane wave images, based on SNR, contrast, gCNR, and FWHM measurements. The DNN output images are also globally similar to enhanced images, indicated by a mean PSNR of 18.61 dB for the simulated test set and 13.52 dB for the *in vivo* test set. These results highlight the promise of performing additional explorations to generate high quality ultrasound images from a single plane wave transmission, which could be applied to multiple ultrasound-based clinical tasks.

Acknowledgments

This work is supported by NIH Trailblazer Award R21 EB025621.

References

[1] K. T. Spencer, B. J. Kimura, C. E. Korcarz, P. A. Pellikka, P. S. Rahko, and R. J. Siegel, "Focused cardiac ultrasound: recommendations from the american society of echocardiography," *Journal of the American Society of Echocardiography*, vol. 26, no. 6, pp. 567–581, 2013.

[2] R. Wang, Z. Fang, J. Gu, Y. Guo, S. Zhou, Y. Wang *et al.*, "High-resolution image reconstruction for portable ultrasound imaging devices," *EURASIP Journal on Advances in Signal Processing*, vol. 2019, 12 2019.

[3] 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*, vol. 58, no. 7, pp. 1377–1388, 2011.

[4] 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*, 2020.

[5] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, "Beamforming and speckle reduction using neural networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 5, pp. 898–910, 2019.

[6] "Challenge on Ultrasound Beamforming with Deep Learning (CUBDL)." [Online]. Available: http://dx.doi.org/10.21227/f0hn-8f92

[7] M. A. L. Bell, J. Huang, D. Hyun, Y. Eldar, R. van Sloun, and M. Mischi, "Challenge on Ultrasound Beamforming with Deep Learning (CUBDL)," in *2020 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2020.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2015, pp. 234–241.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[10] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134.

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

[12] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019, pp. 8026–8037.

[13] 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)*. IEEE, 2016, pp. 1–4.

[14] 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*, vol. 39, no. 2, pp. 262–267, 1992.

[15] J. A. Jensen, "Field: A program for simulating ultrasound systems," in *10TH Nordicbaltic Conference on Biomedical Imaging, vol. 4, Supplement 1, Part 1: 351–353*. Citeseer, 1996.

[16] A. Rodriguez-Molares, O. M. H. Rindal, O. Bernard, A. Nair, M. A. L. Bell, H. Liebgott *et al.*, "The ultrasound toolbox," in *2017 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2017, pp. 1–4.

[17] A. Rodriguez-Molares, O. M. H. Rindal, J. D'hooge, S.-E. Måsøy, A. Austeng, M. A. L. Bell *et al.*, "The generalized contrast-to-noise ratio: a formal definition for lesion detectability," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 4, pp. 745–759, 2019.