Deep Learning-based Universal Beamformer for Ultrasound Imaging

Shujaat Khan, Jaeyoung Huh, and Jong Chul Ye

Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea
{shujaat,woori93,jong.ye}@kaist.ac.kr

Abstract. In ultrasound (US) imaging, individual channel RF measurements are back-propagated and accumulated to form an image after applying specific delays. While this time reversal is usually implemented using a hardware- or software-based delay-and-sum (DAS) beamformer, the performance of DAS decreases rapidly in situations where data acquisition is not ideal. Herein, for the first time, we demonstrate that a single data-driven beamformer designed as a deep neural network can directly process sub-sampled RF data acquired at different sampling rates to generate high quality US images. In particular, the proposed deep beamformer is evaluated for two distinct acquisition schemes: focused ultrasound imaging and planewave imaging. Experimental results showed that the proposed deep beamformer exhibit significant performance gain for both focused and planar imaging schemes, in terms of contrast-to-noise ratio and structural similarity.

Keywords: Ultrasound imaging · Compressive sensing · Beamformer.

1 Introduction

Due to minimal invasiveness from non-ionizing radiations and excellent temporal resolution, ultrasound (US) is an indispensable tool for various clinical applications such as cardiac, fetal imaging, etc. The basic imaging principle of US imaging is based on the time-reversal [2,13], which is based on a mathematical observation that the wave operator is self-adjoint. For example, in focused B-mode US imaging, the return echoes from individual scan-line are recorded by the receiver channels, after which delay-and-sum (DAS) beamformer applies the time-reversal delay to the channel measurement and additively combines them for each time point to form images at each scan-line. Despite the simplicity, high-speed analog-to-digital converters (ADCs) and large number of receiver elements are often necessary in time reversal imaging to improve the image quality by reducing the side lobes which otherwise reduce image resolution and contrast.

To address this problem, various adaptive beamforming techniques have been developed over the several decades [12,16].

* Supported by organization x.
Recently, inspired by the tremendous success of deep learning, the authors in [4,3,8,10] use deep neural networks for the reconstruction of high-quality US images from limited number of received RF data. For example, the work in [3] uses deep neural network for coherent compound imaging from small number of plane wave illumination. In focused B-mode ultrasound imaging, [14] employs the deep neural network to interpolate the missing RF-channel data with multiline acquisition for accelerated scanning. In [8,10], the authors employ deep neural networks for the correction of blocking artifacts in multiline acquisition and transmission scheme. While these recent deep neural network approaches provide impressive reconstruction performance, the current design is not universal in the sense that the designed neural network cannot completely replace a DAS beamformer, since they are designed and trained for specific acquisition scenario.

Therefore, one of the most important contributions of this paper is to demonstrate that a single beamformer can generate high quality images robustly for various detector channel configurations and subsampling rates. The main innovation of our universal deep beamformer comes from one of the most exciting properties of deep neural network - exponentially increasing expressiveness with respect to the channel and depth [1]. Thanks to the expressiveness of neural networks, our novel deep beamformer can learn the mapping to images from various sub-sampled RF measurements, and exhibits superior image quality for all sub-sampling rates. Another amazing feature of the proposed network is that even though the network is trained to learn the mapping from the sub-sampled channel data to the B-mode images from full rate DAS images, the trained neural network can utilize the fully sampled RF data furthermore to improve the image contrast even for the full rate cases.

This paper is organized as follows. In Section 2 we describe the data set and experimental setup used in our study. The experimental results are discussed in Section 3 followed by the discussion and conclusion in Section 4.

2 Method

2.1 Dataset

For experimental verification, multiple RF data were acquired with the E-CUBE 12R US system (Alpinion Co., Korea). For data acquisition, we used a linear array transducer (L3-12H) with a center frequency of 8.48 MHz. The configuration of the probe is given in Table 1.

Using a linear probe, we acquired RF data from the carotid area of 10 volunteers. In focused mode imaging experiment the in-vivo data consists of 40 temporal frames per subject, providing 400 sets of Depth-Rx-TE data cube. The dimension of each Rx-TE plane was 64 × 96. A set of 30,000 Rx-TE planes was randomly selected from the 4 subjects datasets, and data cubes (Rx-TE-depth) are then divided into 25,000 datasets for training and 5000 datasets for validation. The remaining dataset of 360 frames was used as a test dataset. In plane wave imaging experiments, we acquire 109 frames, among which only 8
frames (images) from in-vivo data were used for training and 1 for validation purpose while remaining 100 were used as test dataset. Each US image RAW data consist of 31 PWs and 192-channels, and each frame have different depth ranges varying from 25-60 mm consist of 2000-9000 depth planes.

In addition, we acquired RF data from the ATS-539 multipurpose tissue mimicking phantom using focused and unfocused modes. These datasets were only used for test purpose and no additional training of CNN was performed on it. The phantom datasets were used to verify the generalization power of the proposed method.

2.2 RF sub-sampling scheme

For focused mode imaging, we generated six sets of sub-sampled RF data at different down-sampling rates. In particular, we use several subsampling cases using 64, 32, 24, 16, 8 and 4 Rx-channels. Since the active receivers at the center of the scan-line get RF data from direct reflection, two channels that are in the center of active transmitting channels were always included to improve the performance, and remaining channels were randomly selected from the total 64 active receiving channels. For each depth plane, a different sampling pattern (mask) is used.

For unfocused planar wave imaging, we generated six sets of sub-sampled RF data at different down-sampling rates. In particular, we used two subsampling schemes: variable down-sampling of RF-channel data pattern across the depth to reduce high data-rate and power requirements, and uniform sub-sampling of PWs angles to accelerate acquisition speed. Here we use the following subsampling cases: (1) 64, 32, 16, and 8 Rx-channels with 31 PWs. (2) 31, 11, 7, and 3 PWs with 64 Rx-channels.

2.3 Network architecture

In focused mode imaging, $3 \times 64 \times 96$ data-cube in the depth-Rx-TE sub-space was used for CNN training to generate a $2 \times 3 \times 96$ I and Q data in the depth-TE plane. The target IQ data is obtained from two output channels each representing real and imaginary parts. The proposed CNN consists of 27 convolution layers
Shujaat Khan, Jaeyoung Huh, and Jong Chul Ye

Fig. 1. Proposed CNN architecture for sub-sampled (a) focused US B-mode imaging. (b) planewave US B-mode imaging.

composed of a contracting path with concatenation, batch normalization, ReLU except for the last convolution layer. The first 26 convolution layers use $3 \times 3$ convolutional filters (i.e., the 2-D filter has a dimension of $3 \times 3$), and the last convolution layer uses a $1 \times 1$ filter and contract the $3 \times 64 	imes 96$ data-cube from depth-Rx-TE sub-space to $2 \times 3 \times 96$ IQ-depth-TE plane as shown in Figs. 1(a).

In plane wave imaging a multi-channel CNN was trained using $3 \times 31 \times 192$ data-cube in the depth-PW-Rx sub-space to generate a $1 \times 192$ RF sum data in the depth-TE plane. Three input channels were used to process three adjacent depth planes to generate target RF sum data of the central depth plane. The proposed CNN consists of 27 convolution layers composed of a contracting path with concatenation, batch normalization, and ReLU except for the last convolution layer. The first 26 convolution layers use $3 \times 3$ convolutional filters (i.e., the 2-D filter has a dimension of $3 \times 3$), and the last convolution layer uses a $1 \times 1$ filter and contract the $3 \times 31 \times 192$ data-cube from depth-PW-Rx sub-space to $1 \times 192$ depth-Rx plane as shown in Figs. 1(b).

Both networks were implemented with MatConvNet [9] in the MATLAB 2015b environment. Specifically, for network training, the parameters were estimated by minimizing the $l_2$ norm loss function using a stochastic gradient descent with a regularization parameter of $10^{-4}$. The learning rate started from $10^{-3}$ and gradually decreased to $10^{-5}$ in 200 epochs. The weights were initialized using Gaussian random distribution with the Xavier method [4].
3 Experimental Results

To quantitatively show the advantages of the proposed deep learning method, we used the contrast-to-noise ratio (CNR), generalized CNR (GCNR) [7], and structure similarity (SSIM).

![Focused B-mode imaging reconstruction results of standard DAS beam-former and the proposed method for in-vivo carotid region.](image)

**Fig. 2.** Focused B-mode imaging reconstruction results of standard DAS beam-former and the proposed method for in-vivo carotid region.

**Focused mode imaging** Figs. 2 show the results of an *in vivo* example for 64, 32, 24, 16, 8 and 4 Rx-channels down-sampling schemes. Since 64 channels are used as a full sampled data, this corresponds to 1×, 2×, 2.7×, 4×, 8× and 16× sub-sampling factors. The images are generated using the proposed DeepBF and the standard DAS beam-former method. Our method significantly improves the visual quality of the US images by estimating the correct dynamic range and eliminating artifacts for both sampling schemes. From difference images, it is evident that the quality degradation of images in DAS is higher than the DeepBF. Note that the proposed method successfully reconstruct both the near and the far field regions with equal efficacy, and only minor structural details are imperceivable. Furthermore, it is remarkable that the CNR and GCNR values are significantly improved by the DeepBF even for the fully sampled case (eg. from 1.69 to 2.16 in CNR and from 0.74 to 0.83 in GCNR), which clearly shows the advantages of the proposed method.
Table 2. Performance statistics on in vivo data for variable sampling pattern

| sub-sampling factor | CNR  | GCNR  | PSNR (dB) | SSIM  |
|--------------------|------|-------|-----------|-------|
|                    | DAS DeepBF | DAS DeepBF | DAS DeepBF | DAS DeepBF |
| 1                  | 1.38 | 1.40 | 0.04      | 0.00  |
| 2                  | 1.33 | 1.47 | 0.03      | 0.06  |
| 2.7                | 1.3  | 1.44 | 0.02      | 0.06  |
| 4                  | 0.25 | 1.38 | 0.06      | 0.04  |
| 8                  | 1.18 | 1.26 | 0.08      | 0.06  |
| 16                 | 1.12 | 1.17 | 0.06      | 0.08  |

We also compared the CNR, GCNR, PSNR, and SSIM distributions of reconstructed B-mode images obtained from 360 in-vivo test frames. Table 2 showed that the proposed deep beamformer consistently outperformed the standard DAS beamformer for all subsampling schemes and ratios. One big advantage of ultrasound image modality is its run-time imaging capability, which requires fast reconstruction time. Another important advantage of the proposed method is the run-time complexity. The average reconstruction time for each depth plane is around 4.8 milliseconds, which could be easily reduced by optimized implementation and reconstruction of multiple depth planes in parallel.

Fig. 3. Planewave B-mode imaging reconstruction results of standard DAS beamformer and the proposed method for: (a) in-vivo carotid region (b) tissue mimicking phantom.
Planewave US imaging  Figs. 3(a)(b) show the results in vivo and phantom image examples for different down-sampling schemes. The images are generated using the proposed DeepBF and the standard DAS beam-former method. Our method significantly improves the visual quality of the US images by estimating the correct dynamic range and eliminating artifacts for both sampling schemes. From zoomed region images, it can be seen that the quality of the DeepBF images is relatively unchanged for variable sampling scenarios. Note that the proposed method successfully reconstruction both the near and the far field regions with equal efficacy, and only minor structural details are imperceivable. In addition,

we compared the CNR, GCNR, and SSIM distributions of reconstructed B-mode images obtained from 100 invivo test frames. Our method shows significant performance gain in all measures. From Fig. 4 it is evident that the quality degradation of images in DAS is higher than the DeepBF. Furthermore, it is remarkable that the CNR value are significantly improved by the DeepBF even for the fully sampled case (e.g. from 2.85 to 2.86 in CNR), which clearly shows the advantages of the proposed method.

4 Conclusion

Herein, for the first time we demonstrated that a single universal deep beam-former trained using a purely data-driven way can be used for variable rate ultrasound imaging. Even for fully sampled data, the proposed method further improves the images. Moreover, CNR, GCNR, PSNR, and SSIM were significantly improved over standard DAS method across various subsampling schemes. The proposed schemes may substantially help in designing low-powered accelerated ultrasound imaging systems.

References

1. Arora, R., Basu, A., Mianjy, P., Mukherjee, A.: Understanding deep neural networks with rectified linear units. arXiv preprint arXiv:1611.01491 (2016)
2. Fink, M.: Time reversal of ultrasonic fields. i. basic principles. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 39(5), 555–566 (Sep 1992). [https://doi.org/10.1109/58.156174]

3. Gasse, M., Milloz, F., Roux, E., Garcia, D., Liebgott, H., Friboulet, D.: High-quality plane wave compounding using convolutional neural networks. IEEE transactions on ultrasonics, ferroelectrics, and frequency control 64(10), 1637–1639 (2017)

4. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. pp. 249–256 (2010)

5. Jensen, A.C., Austeng, A.: The iterative adaptive approach in medical ultrasound imaging. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 61(10), 1688–1697 (Oct 2014). [https://doi.org/10.1109/TUFFC.2014.006478]

6. Kim, K., Park, S., Kim, J., Park, S., Bae, M.: A fast minimum variance beamforming method using principal component analysis. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 61(6), 930–945 (June 2014). [https://doi.org/10.1109/TUFFC.2014.2989]

7. Rodriguez-Molares, A., Rindal, O.M.H., Dhooge, J., Másoy, S.E., Austeng, A., Torp, H.: The generalized contrast-to-noise ratio (2018)

8. Senouf, O., Vedula, S., Zurakhov, G., Bronstein, A., Zibulevsky, M., Michailovich, O., Adam, D., Blondheim, D.: High frame-rate cardiac ultrasound imaging with deep learning. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. pp. 126–134. Springer International Publishing, Cham (2018)

9. Vedaldi, A., Lenc, K.: Matconvnet: Convolutional neural networks for matlab. In: Proceedings of the 23rd ACM international conference on Multimedia. pp. 689–692. ACM (2015)

10. Vedula, S., Senouf, O., Zurakhov, G., Bronstein, A., Zibulevsky, M., Michailovich, O., Adam, D., Gaitini, D.: High quality ultrasonic multi-line transmission through deep learning. In: Knoll, F., Maier, A., Rueckert, D. (eds.) Machine Learning for Medical Image Reconstruction. pp. 147–155. Springer International Publishing, Cham (2018)

11. Vignon, F., Burcher, M.R.: Capon beamforming in medical ultrasound imaging with focused beams. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 55(3), 619–628 (March 2008). [https://doi.org/10.1109/TUFFC.2008.686]

12. Viola, F., Walker, W.F.: Adaptive signal processing in medical ultrasound beamforming. In: IEEE Ultrasonics Symposium, 2005. vol. 4, pp. 1980–1983 (Sep 2005). [https://doi.org/10.1109/ULTSYM.2005.1603264]

13. Wu, F., Thomas, J., Fink, M.: Time reversal of ultrasonic fields. ii. experimental results. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 39(5), 567–578 (Sep 1992). [https://doi.org/10.1109/58.156175]

14. Yoon, Y.H., Khan, S., Huh, J., Ye, J.C., et al.: Efficient b-mode ultrasound image reconstruction from sub-sampled rf data using deep learning. IEEE transactions on medical imaging (2018)