2019 IEEE International Ultrasonics Symposium (IUS)
Glasgow, Scotland, October 6-9, 2019

One-Step Deep Learning Approach to Ultrasound Image Formation and Image Segmentation with a Fully Convolutional Neural Network

Arun Asokan Nair,* Trac D. Tran,* Austin Reiter,‡ and Muyinatu A. Lediju Bell*‡§
*Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD
‡Department of Computer Science, Johns Hopkins University, Baltimore, MD
§Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD

***Abstract*—Single plane wave imaging is well-suited to high frame rate imaging tasks (e.g., ultrasound based robotic tracking). However, suboptimal image quality is obtained when limited to a single plane wave transmission. To address this challenge, we propose to train deep neural networks (DNNs) as an alternative to delay-and-sum beamforming followed by segmentation. Our overall goal is to extract information directly from raw channel data prior to the application of time delays and to simultaneously generate both a segmentation map and an ultrasound B-mode image of anechoic cysts surrounded by tissue. A network trained with 17,676 Field II simulations was tested with both simulated and experimental phantom data sets that were not included during training (9,108 and 320 images, respectively). DNN results from simulated and phantom test sets produced similar dice similarity coefficients (DSC), contrast, tissue signal-to-noise ratios (SNR), and generalized contrast-to-noise ratios (GCNR). Similarity is reported as the mean $\pm$ standard deviation of these metrics for simulated and experimental test set results as follows:** 0.92 ± 0.13 **and** 0.92 ± 0.03 **DSC, respectively;** -39.56 ± 6.41 **dB and** -35.56 ± 3.81 **dB contrast, respectively;** 3.78 ± 1.08 **and** 4.53 ± 1.23 **SNR, respectively; and** 1.00 ± 0.01 **and** 1.00 ± 0.01 **GCNR, respectively. Thus, the DNNs successfully transferred feature representations learned from simulated data to experimental phantom data, highlighting the promise of this novel alternative to traditional ultrasound information extraction.**

***Index Terms*—Deep Learning, Neural Network, Fully Convolutional Neural Network, Beamforming, Image Segmentation**

I. Introduction

PLANE wave ultrasound imaging is ideal for ultrasound-based robotic tracking tasks due to its ability to achieve frame rates exceeding 18,000 Hz [1]. Ideally, a single plane wave transmission would provide sufficient image quality, but due to the presence of significant levels of acoustic clutter, a single plane wave transmission is often insufficient. Multiple plane wave transmissions with varying angles are typically needed for coherent plane wave compounding [2] to improve image quality, which limits the maximum possible frame rates of robotic tracking tasks. In addition, for robotic tracking tasks, the beamformed image is often provided as the input to an image segmentation algorithm that outputs a segmentation mask, which is used to command the robot or make a decision.

This work is supported by NIH Trailblazer Award R21 EB025621.

Recently, there has been growing interest in applying DNNs to augment or replace steps of the ultrasound image formation pipeline [3]–[11]. In this work, we improve our previous approaches to learning beamforming from single, unfocused plane wave channel data [5]–[7] by demonstrating the ability of our deep neural network (DNN) to produce B-mode images with improved contrast and smoother tissue texture compared to delay-and-sum beamforming, while concurrently extracting segmentation information directly from the raw data. We also introduce a novel method for enhancing simulated delay-and-sum (DAS) beamformed images using ground truth segmentation masks to produce enhanced beamformed images with superior contrast for network training. In addition, network training in this paper is performed in a purely supervised manner using a fully convolutional neural network (FCNN), making the network easier and faster to train when compared to the generative adversarial network (GAN) employed in our previous paper [7].

II. Methods

An illustration of the traditional beamforming approach compared to the proposed network architecture is presented in Fig. 1. Traditionally, raw ultrasound channel data is beamformed, envelope detected, log compressed, and filtered to produce a delay-and-sum (DAS) beamformed image. This DAS image then becomes the input to an algorithm that produces a segmented image for the robot to locate and track the target. Our goal with the proposed network is to replace the mathematical component of image formation and segmentation with a robot input obtained directly from raw ultrasound channel data. The network architecture was designed based on U-Net [12], possessing a single encoder adopting the VGG-13 [13] encoder with BatchNorm [14] layers to stabilize training and speed up convergence, followed by two decoders. One decoder generates a DNN image trained to match the DAS image, and the other decoder generates a DNN segmentation trained to match the true segmentation.

The Field II [15], [16] ultrasound simulator was used to generate 22,230 simulations of individual anechoic cysts surrounded by tissue. We employed simulations in our training

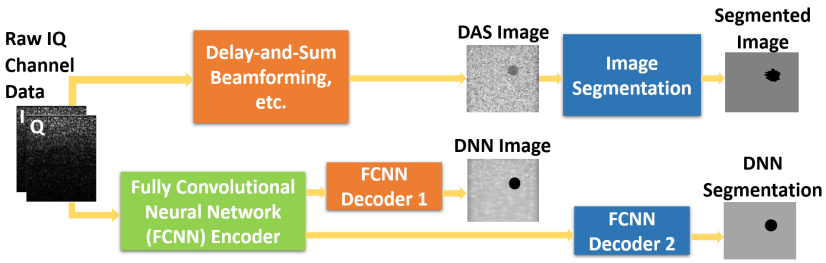

Fig. 1. An illustration of our proposed pipeline. For ultrasound-based robotic tracking tasks, traditionally raw channel data undergoes delay-and-sum beamforming followed by envelope detection, log compression and filtering to produce an interpretable delay-and-sum (DAS) beamformed image, which is then passed to a segmentation algorithm producing a segmented image on the basis of which the robot moves. Instead, a fully convolutional neural network architecture (FCNN) is developed with a single encoder and two decoders that directly outputs a DNN image from one decoder and a DNN segmentation, trained to match the true segmentation, from the second decoder using the raw channel data from a single plane wave insonification. In this figure, the input data is in-phase/quadrature (IQ) data presented stacked together as a tensor input, with I as one channel and Q as the other channel.

TABLE I
SIMULATED CYST DATA PARAMETERS

Parameter	Range	Increment
Radius (r)	2-8 mm	1-2 mm
Speed of Sound (c)	1420-1600 m/s	10 m/s
Lateral position of cyst center (x)	-16 mm - 0 mm	2 mm
Axial position of cyst center(z)	40-70 mm	2.5 mm

approach for two primary reasons. First, simulators enable the generation of large, diverse datasets that are required to train robust DNNs. Second, for segmentation tasks, simulators enable the specification of ground truth pixel labels, allowing one to avoid the expensive and time-consuming step of a human annotator to provide segmentation labels.

The simulated cyst radius, lateral and axial position, and speed of sound in the medium were varied using the range and increment sizes defined in Table I. The values of cyst radii were specifically 2, 3, 4, 6, and 8 mm. These cysts were contained within a cuboidal phantom volume located between an axial depth of 30 mm and 80 mm, with an axial width of 40 mm and an elevational thickness of 7 mm. Each simulation used a different seed for the random number generator, and therefore produced a unique speckle realization to model the diversity expected with real data. A total of 50,000 scatterers were contained within the simulated phantom to ensure fully developed speckle.

In each simulation, a single plane wave at normal incidence was simulated to insonify the region of interest. The simulated ultrasound probe matched the parameters of the Alpinion L3-8 linear array transducer, and its center was placed at the axial, lateral, and elevation center of the phantom (i.e., 0 mm, 0 mm, and 0 mm, respectively). The one exception to matching the real hardware system was a simulated sampling frequency of 100 MHz (rather than the 40 MHz sampling frequency of the Alpinion ultrasound scanner) in order to improve the Field II simulation accuracy [15], [16]. A total of 80% of the 22,230 simulated examples was reserved for training, and the remaining 20% was used for network testing. As cysts were purposely simulated to reside on the left side of the phantom (see Table I), the simulated channel data were flipped, and samples from the right side of simulated phantom were also trained and tested without explicitly simulating this possibility.

All input and output examples were normalized to facilitate training. For the input channel data, the input was normalized by the maximum absolute value in the input to ensure the normalized input lies between -1 and 1. The network was trained with the normalized input.

Channel data from a cross sectional slice of two anechoic cylinders in a CIRS 054GS phantom located at depths of 40 mm and 70 mm were acquired using an Alpinion L3-8 linear array ultrasound transducer attached to an Alpinion E-Cube 12R research scanner. Two independent 80-frame sequences were acquired. The anechoic targets were consistently in the left or right half of the image for each acquisition sequence, achieved by manually flipping the ultrasound probe. The channel data corresponding to each of the 80 frames in each sequence was flipped from left to right, producing a dataset consisting of 320 total images in order to test the generalizability of the trained networks. The ground truth for this phantom dataset was specified by manually annotating pixels in the beamformed ultrasound image as cyst or tissue. When quantitatively evaluating these phantom examples, the mean result for the two anechoic cysts in the same image is reported, unless otherwise stated.

The DNN results reported in this paper were trained using the following baseline settings. The Adam [17] optimizer used a learning rate of 10^{-5} for 25 epochs, where one epoch is defined as one pass over the entire training dataset (i.e., the entire training dataset is presented to the network once for training). The mini-batch size for the training dataset was set to 16. Training was parallelized across a set of 4 Tesla P40 GPUs. The input data format was IQ, with the I and Q components treated as two separate input feature channels. This training was performed over a combined dataset consisting of data with and without an attenuation coefficient of 0.5 dB/cm-MHz.

One known limitation of single plane wave transmissions is poor image contrast. Fig. 2 shows an example DAS beamformed image obtained from a single plane wave insonification of an anechoic cyst. The cyst is intrinsically anechoic, but the visualized cyst in the DAS beamformed image is not, as a

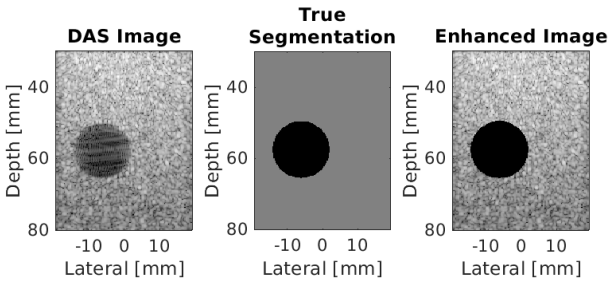

Fig. 2. From left to right, this example shows a simulated DAS beamformed ultrasound image, the known segmentation of the cyst from surrounding tissue, and the corresponding enhanced beamformed image.

result of acoustic clutter.

Our goal in using DNNs for ultrasound image formation is to obtain better quality images than that of DAS images, and not to simply replicate poor DAS image quality. Toward this end, the pixel labels obtained from the true segmentation mask are used to set the pixel values of the anechoic region to zero while preserving the pixel values of the surrounding tissue, removing the clutter and restoring the desired true anechoic appearance of the cyst [18]. All simulated DAS beamformed images are enhanced as described above, and the enhanced beamformed images are used to train the DNN.

The DNN was trained to minimize the sum of two losses:

1) Mean Absolute Error (L1Loss) between the predicted DNN image and the reference enhanced beamformed image, defined as:

$$\text{L1Loss}(D, E) = \frac{||D - E||_1}{N} \tag{1}$$

where D and E are the vectorized images, and N is the total number of image pixels.

2) Dice Similarity Coefficient Loss (DSCLoss) between the predicted DNN segmentation and the true segmentation, defined as:

$$\text{DSCLoss}(X, Y) = 1 - \text{DSC}(X, Y) = 1 - 2\frac{|X \cap Y|}{|X| + |Y|} \tag{2}$$

where X and Y are the vectorized binary segmentation masks.

The total network loss was the sum of the above two losses:

$$\text{Total Loss} = \text{L1Loss}(D, E) + \text{DSCLoss}(X, Y) \tag{3}$$

$$= \frac{||D - E||_1}{N} + 1 - 2\frac{|X \cap Y|}{|X| + |Y|} \tag{4}$$

As observed in our previous work [6], higher DSCs are achieved with larger cysts compared to smaller cysts. In addition, small cysts have greater potential to be missed, which is quantified as a DSC of approximately zero. Based on this knowledge, we prioritize a fair comparison of the multiple network parameters, which we define as a minimum DSC ≥ 0.05. Test cases that did not meet this basic criterion were excluded from the results reported in this paper. Overall, the network successfully detected the simulated cyst in $4,274$ out of $4,554$ test examples, with all missed cysts being of 2 mm radius. Note that none of the experimental phantom data met our exclusion criteria, thus all experimental test cases are included.

Contrast, signal-to-noise ratio (SNR), and generalized contrast-to-noise ratio (GCNR) [19] of DAS and DNN images were calculated using the following equations:

$$\text{Contrast} = 20\log_{10}\left(\frac{S_i}{S_o}\right) \tag{5}$$

where S_i and S_o represent the mean signals inside and outside the cyst, respectively.

$$\text{SNR} = \frac{S_o}{\sigma_o} \tag{6}$$

where σ_o is the standard deviation of signals outside the cyst.

$$\text{GCNR} = 1 - \sum_{x=0}^{1} \min_x\{p_i(x), p_o(x)\} \tag{7}$$

where $p_i(x)$ and $p_o(x)$ are the probability mass functions of the signal inside and outside the cyst, respectively.

III. Results

A. Simulation Results

Fig. 3 shows an example simulated test output. Left to right, this example shows simulated raw IQ channel data after applying envelope detection, a DAS beamformed ultrasound image, a DNN image produced by our network, the known segmentation of the cyst from surrounding tissue, the DNN segmentation predicted by our network, and the DNN segmentation overlaid on the true segmentation. The DNN segmentation produces a DSC of 0.98 and the DNN image produces a contrast of -44.91 dB, an SNR of 3.37, and a GCNR of 1.00, which is an improvement over the DAS image contrast of -16.27 dB, SNR of 1.69, and a GCNR of 0.83. Across the entire test, the DNN produced a mean $\pm$ one standard deviation DSC of 0.92 ± 0.13, contrast of -39.56 ± 6.41 dB, SNR of 3.78 ± 1.08, and GCNR of 1.00 ± 0.01, superior to DAS with contrast of -14.39 ± 2.52 dB, SNR of 1.82 ± 0.34, and GCNR of 0.80 ± 0.06.

B. Phantom Results

Fig. 4 shows an example phantom test case output. From left to right, this example shows raw phantom IQ channel data after applying envelope detection, a DAS beamformed ultrasound image, a DNN image produced by our network, the known segmentation of the cyst from surrounding tissue, the DNN segmentation predicted by our network, and the DNN segmentation overlaid on the true segmentation. The DNN segmentation produces a DSC of 0.88 and the DNN image produces a contrast of -34.77 dB, an SNR of 3.19, and a GCNR of 0.99, which is an improvement over the DAS image contrast of -16.05 dB, SNR of 1.90, and a GCNR of 0.99. Across the entire test set, the DNN produced a mean $\pm$ one standard deviation DSC of 0.92 ± 0.03, contrast

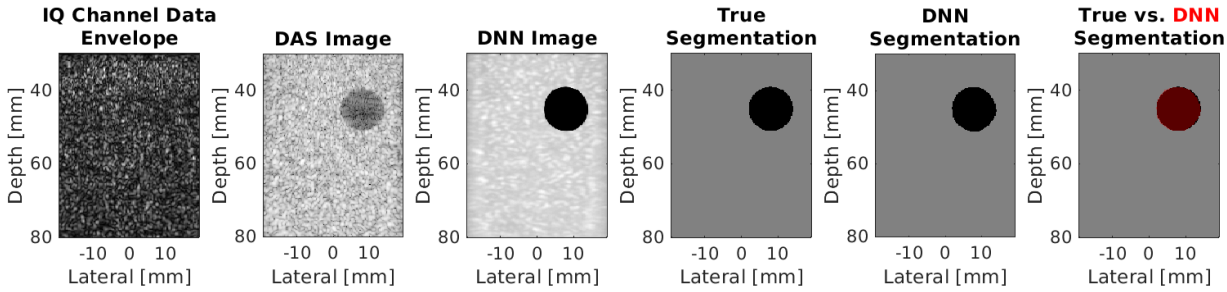

Fig. 3. From left to right, this example shows simulated raw IQ channel data after applying envelope detection, a DAS beamformed ultrasound image, a DNN image produced by our network, the known segmentation of the cyst from surrounding tissue, the DNN segmentation predicted by our network, and an image with a red transparent overlay of the DNN segmentation over the true segmentation.

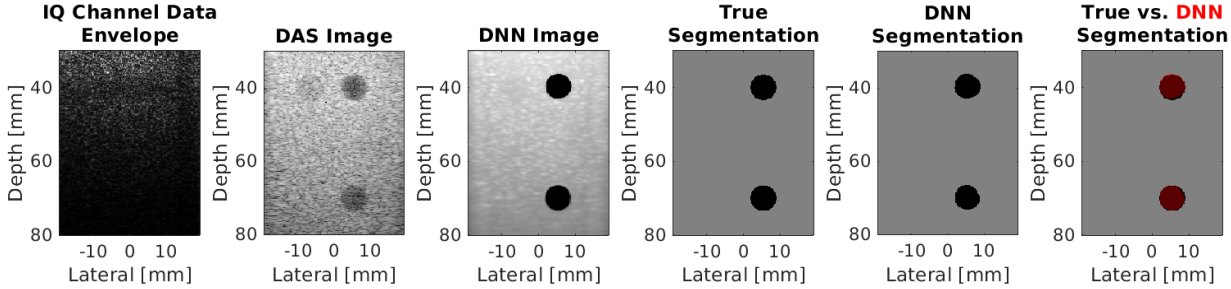

Fig. 4. From left to right, this example shows raw phantom IQ channel data after applying envelope detection, a DAS beamformed ultrasound image, a DNN image produced by our network, the known segmentation of the cyst from surrounding tissue, the DNN segmentation predicted by our network, and an image with a red transparent overlay of the DNN segmentation over the true segmentation.

of -35.56 ± 3.81 dB, SNR of 4.53 ± 1.23, and GCNR of 1.00 ± 0.01, superior to DAS with contrast of -14.17 ± 2.88 dB, SNR of 1.93 ± 0.12, and GCNR of 0.99 ± 0.00.

IV. Conclusion

This paper highlights our initial succes with using DNNs to directly extract information from raw single plane wave IQ channel data prior to the application of any time delays. Results are used to create high quality images and cyst segmentation masks from the same input data. The presented approach is a promising alternative to traditional DAS beamforming followed by segmentation for high-speed robotic ultrasound tasks that include detection and tracking of anechoic cysts.

References

[1] M. Tanter and M. Fink, "Ultrafast imaging in biomedical ultrasound," *IEEE TUFFC*, vol. 61, no. 1, pp. 102–119, 2014.

[2] R. R. Entrekin, B. A. Porter, H. H. Sillesen, A. D. Wong, P. L. Cooperberg, and C. H. Fix, "Real-time spatial compound imaging: application to breast, vascular, and musculoskeletal ultrasound," in *Seminars in Ultrasound, CT and MRI*, vol. 22, no. 1. Elsevier, 2001, pp. 50–64.

[3] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE TMI*, vol. 37, no. 9, pp. 2010–2021, 2018.

[4] B. Luijten, R. Cohen, F. J. de Bruijn, H. A. Schmeitz, M. Mischi, Y. C. Eldar, and R. J. van Sloun, "Deep learning for fast adaptive beamforming," in *IEEE ICASSP 2019*. IEEE, 2019, pp. 1333–1337.

[5] A. A. Nair, T. D. Tran, A. Reiter, and M. A. L. Bell, "A deep learning based alternative to beamforming ultrasound images," in *IEEE ICASSP 2018*. IEEE, 2018, pp. 3359–3363.

[6] A. A. Nair, M. R. Gubbi, T. D. Tran, A. Reiter, and M. A. L. Bell, "A fully convolutional neural network for beamforming ultrasound images," in *IEEE IUS 2018)*. IEEE, 2018, pp. 1–4.

[7] 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 *IEEE CISS 2019*. IEEE, 2019, pp. 1–6.

[8] S. Vedula, O. Senouf, G. Zurakhov, A. Bronstein, O. Michailovich, and M. Zibulevsky, "Learning beamforming in ultrasound imaging," *arXiv preprint arXiv:1812.08043*, 2018.

[9] D. Perdios, A. Besson, F. Martinez, M. Vonlanthen, M. Arditi, and J. Thiran, "On problem formulation, efficient modeling and deep neural networks for high-quality ultrasound imaging : Invited presentation," in *IEEE CISS 2019*, March 2019, pp. 1–4.

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

[11] W. Simson, M. Paschali, N. Navab, and G. Zahnd, "Deep learning beamforming for sub-sampled ultrasound data," in *IEEE IUS 2018*. IEEE, 2018, pp. 1–4.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.

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

[14] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

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

[16] 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.

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

[18] A. Luchies and B. Byram, "High dynamic range ultrasound beamforming using deep neural networks," in *Medical Imaging 2019: Ultrasonic Imaging and Tomography*, vol. 10955. International Society for Optics and Photonics, 2019, p. 109550P.

[19] A. Rodriguez-Molares, O. M. H. Rindal, J. D'hooge, S.-E. Måsøy, A. Austeng, and H. Torp, "The generalized contrast-to-noise ratio," in *IEEE IUS 2018*. IEEE, 2018, pp. 1–4.