A Fully Convolutional Neural Network for Beamforming Ultrasound Images

Arun Asokan Nair*, Mardava Rajugopal Gubbi*, Trac Duy 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*—Plane wave ultrasound imaging is one of the fastest ultrasound methods available to reduce latency for ultrasound-based robotic tracking tasks. However, the presence of acoustic clutter and speckle in the images can confuse robotic tracking algorithms. In addition, multiple plane wave insonification angles are often necessary to generate good-quality images, which further reduces the speed of the tracking process. To overcome these challenges, we are exploring deep learning as a method to extract pertinent information directly from raw radiofrequency channel data to locate targets of interest from a single plane wave insonification. Particularly, in this work, we trained a deep convolutional neural network (CNN) with 50,000 Field-II simulations corresponding to a single cyst in tissue insonified by one plane wave transmitted at 0 degrees. The simulated cyst radius, axial and lateral positions were varied, along with the simulated tissue sound speed. The output of the training process is an interpretable segmentation mask that is free of clutter and speckle, which we call a CNN-Based image. An additional dataset of 100 simulations was created and two cyst targets in an experimental phantom were imaged to test our approach. The Dice Similarity Coefficient (DSC), representing the overlap between the true cyst location and the cyst location in the CNN-Based image, was 0.91 for simulated data and 0.74 for experimental data. The network was generally sensitive to cyst radius, with mean DSCs increasing from 0.91 to 0.97 when the cyst radius was $\geq$ 5 mm. A robot controlled ultrasound probe enabled volumetric reconstruction of CNN-Based images, revealing the three-dimensional structure of the two cysts in the phantom. These results demonstrate that a deep neural network trained exclusively with simulated data can generalize to experimental data, which is promising for the development of deep learning methods as an alternative to traditional ultrasound beamforming for robotic tracking tasks.**

***Index Terms*—Deep Learning, Ultrasound Image Formation, Beamforming, Image Segmentation, Machine Learning**

I. Introduction

Ultrasound-based tracking typically necessitates the use of high frame rate acquisition methods to minimize latency. Traditionally, plane wave ultrasound imaging is employed when speed is required, as plane wave imaging is capable of displaying images at frame rates on the order of thousands of frames per second [1]. One limitation of plane wave imaging is that the quality of images acquired is negatively affected by image degradations like speckle and clutter. In order to address these challenges, multiple independent acquisitions are obtained, each yielding a different speckle realization. The images from these acquisitions are then averaged [2] through spatial compounding. While spatial compounding improves image quality, the maximum achievable imaging frame rate suffers, resulting in a trade-off between image quality and acquisition speed.

Fundamentally, this trade-off is directly related to the image formation pipeline. First, beamforming is applied to the reflected signals received by the ultrasound probe with the goal of exploiting the diversity of information received across the elements of the sensor array to achieve beam directionality and focusing [3]. Following beamforming, envelope detection, log compression, and filtering are applied to obtain an ultrasound image that is interpretable by humans. In addition, the formation of a B-mode ultrasound image is the first step for ultrasound-based robotic tracking systems. Post processing and segmentation algorithms are then applied to the beamformed B-mode image to extract the locations of targets of interest, which may then be used as an input for robotic control.

Beamforming is a model-based approach with a rich history and a solid mathematical basis. However, model-free machine learning approaches have recently achieved much success, with deep convolutional neural networks (CNNs) being especially successful. CNNs have outperformed traditional model-based imaging methods in a myriad of fields from natural image classification [4] to cancer detection [5]. They have also been successfully applied to imaging tasks in the ultrasound domain, like real-time prostate segmentation [6], suppression of off-axis scattering [7], simulating ultrasound images from given echogenicity maps [8], and enhancing ultrasound image quality [9] and resolution [10].

Previous work from our group [11] introduced and demonstrated the possibility of using deep neural networks to directly segment targets of interest from the raw radiofrequency (RF) channel data, prior to applying time delays or any other traditional beamforming steps. In doing so, our proposed plan bypasses many of the intermediary steps required for image-guided robotic tracking tasks (e.g., beamforming and segmentation, as illustrated in Fig. 1), which has the potential to reduce algorithm complexity and processing times. This paper expands our previous work by including multiple hypoechoic cysts and demonstrating the ability to detect experimental targets after training with simulated data. We also integrate our approach with a robot-controlled ultrasound probe to image multiple slices of an anechoic and a -6dB hypoechoic

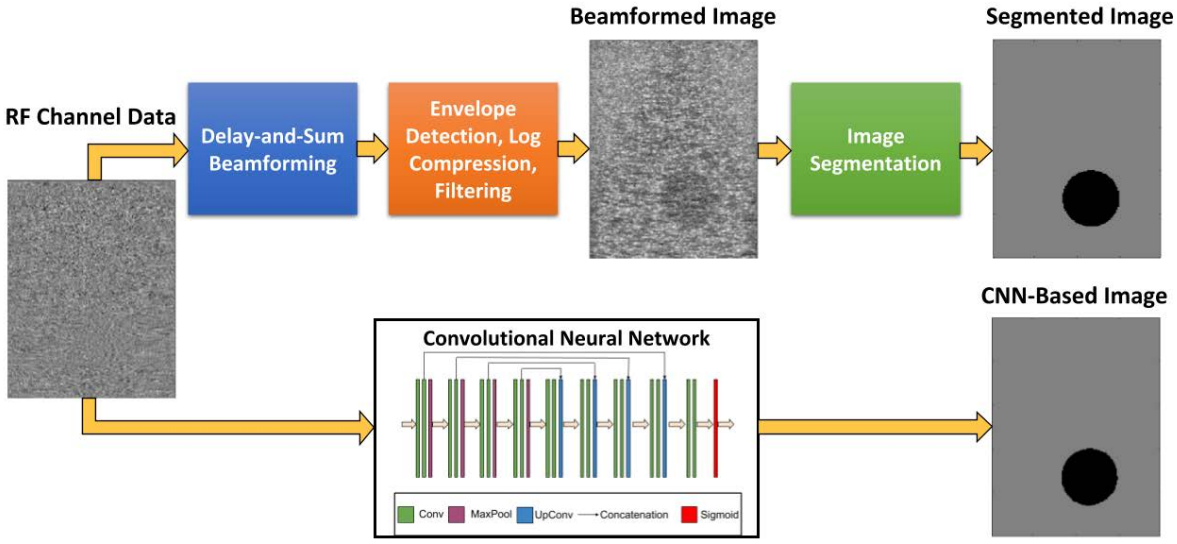

Fig. 1: An overview of our proposed pipeline. Traditionally, for ultrasound-based robotic tracking tasks, beamforming, envelope detection, log compression, filtering and segmentation are performed in sequence on input RF channel data to generate information for the robotic tracking tasks. Instead, our proposed approach uses a deep convolutional neural network to directly locate and segment cyst targets from surrounding tissue using a single plane wave insonification.

cylinder in a phantom, and combine the resulting predictions to form a volumetric reconstruction of the cylinders. Network performance is assessed both qualitatively and quantitatively as a function of multiple parameters of interest.

II. Methods

We trained our network using 50,000 Field-II [12] simulations of a single cyst in normal tissue, with 32,000 simulations corresponding to an anechoic cyst and the remaining 18,000 simulations corresponding to a -6dB hypoechoic cyst. Additional cyst parameters that were varied are listed in Table I.

TABLE I: Field-II cyst simulation parameters

Parameter	Min	Max	Increment
Radius (r)	2 mm	8 mm	1 mm
Speed of Sound (c)	1440 m/s	1640 m/s	10 m/s
Lateral position of cyst center (x)	-15 mm	15 mm	2.5 mm
Axial position of cyst center(z)	35 mm	75 mm	2.5 mm

Each simulation corresponds to a single plane wave insonification at 0 degrees. The parameters of an Alpinion L3-8 linear array transducer were used for the simulations in order to transfer the learned network to real data acquired using an L3-8 transducer connected to an Alpinion E-Cube 12R ultrasound research scanner.

To test the network, an additional test set of 100 Field-II simulations corresponding to parameters randomly chosen from within the range of our training data was constructed. In addition, anechoic and -6dB hypoechoic cysts at a depth of 40 mm in a CIRS 054GS phantom were imaged. These phantom data were used to test generalization of a network trained exclusively on simulated data to experimental phantom data. For the simulations, the ground truth segmentation mask was known a priori, while the ground truth for the phantom data was acquired by manually labeling regions in the beamformed ultrasound image as cyst or tissue.

The L3-8 ultrasound transducer was then attached to the end effector of a Sawyer robot (Rethink Robotics, Boston, MA) to enable volumetric image acquisition. Channel data from 20 individual slices of the anechoic and -6dB hypoechoic cylinders in the CIRS 054GS experimental phantom were acquired with 1 mm increments in the elevation direction. CNN-Based images from each slice were combined in MATLAB to generate a volumetric reconstruction of the experimental data, revealing the two cylindrical structures in the phantom.

The architecture used for the neural network is based on the U-Net [13] architecture commonly used for image segmentation. The neural network has a VGG-13 [14] encoder, with BatchNorm [15] layers used to speed up convergence. The network was trained to minimize the mean square difference between the true segmentation mask and the neural network prediction using the Adam [16] optimization algorithm, with training run for 80 epochs.

The final layer of the network is a sigmoid layer which outputs a scalar value between 0 and 1 for each pixel. This output is interpreted as a per-pixel measure of how confident the network is that the given pixel is a cyst or normal tissue.

An output value of 0 corresponds to a 100% confidence that the pixel is normal tissue, and an output value of 1 corresponds to a 100% confidence that the pixel is a cyst pixel.

To evaluate the output of the network, the Dice Similarity Coefficient (DSC) metric was used. The DSC is defined as:

$$DSC(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|} \quad (1)$$

where X is the true segmentation mask and Y is the predicted segmentation mask. The DSC is thus limited to be a scalar between 0 and 1, with a higher DSC score corresponding to a better result.

III. Results & Discussion

A. Simulation Results

When the trained network was applied to the simulated test set, an average DSC score of 0.91 was obtained with a standard deviation of 0.11. The radii of each simulated cyst was rounded to the nearest integer mm, to group the DSC results as a function of radius, as shown in Fig. 2. Note that the mean DSC for each integer radius monotonically increases as a function of radius from a mean of 0.62 at 2 mm radius to a mean of 0.98 at 8 mm radius. The standard deviation of the mean DSC decreases from 0.12 to 0.01 as the radius increases. Thus, network performance appears to be degraded in the presence of small cysts. However, this apparent degradation occurs because the DSC score penalizes errors in the position and shape of smaller cysts more than it penalizes larger cysts.

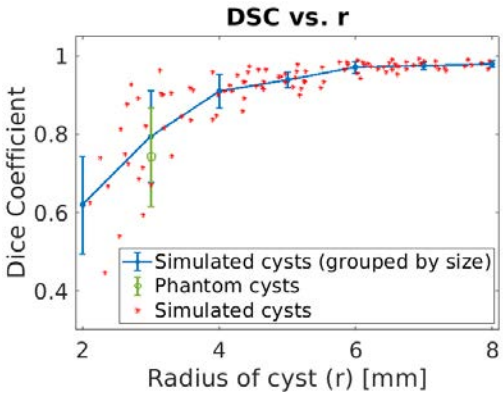

Fig. 2: Performance variation of the trained network as a function of cyst radius. Mean DSC increases and standard deviation decreases as cyst radius increases.

Because the presence of small cysts causes a noticeable degradation in the network performance, the DSC as a function of the remaining variables (i.e., c, z, x, and intrinsic contrast) is reported in Fig. 3 for both the entire dataset and after restricting consideration to cysts with radii $\geq$ 5mm. After limiting the DSC based on cyst radius, it is apparent that the DSC results slightly degrade as intrinsic contrast increases from anechoic to -6dB. Otherwise, there is no apparent correlation between DSC and the remaining parameters in the simulated data.

B. Phantom Results

For the anechoic and -6dB hypoechoic targets in the CIRS phantom, measured performance is similar to that of simulated cysts of similar size, as shown in Fig. 2. In particular, the average DSC score was 0.74 with standard deviation of 0.13. Note that the network was not retrained to obtain these phantom results. One difference between simulated and experimental data is that the Field-II simulations do not model tissue attenuation, which occurs in experimental data. To address attenuation changes during testing, the top and bottom halves of the channel data were normalized and processed separately. In addition, a morphological filter was applied to the final predicted mask in order to suppress small false positives.

C. Volumetric Reconstruction

The 20 slices acquired with robotic control are shown in Fig. 4. Each slice was passed through the network and processed to produce a prediction slice. The prediction slices were then stacked into a cube, and a volumetric segmentation of cyst versus tissue was obtained. Finally, this volumetric segmentation was compared to the hand annotated ground truth segmentation. When compared to the ground truth, the mean errors in cyst radius (r), cyst lateral position (x), and cyst axial position (z) for the anechoic cyst were 0.37 mm, 0.57 mm, and 0.75 mm, respectively. Similarly, the mean errors in r, x, and z for the -6dB hypoechoic cyst were 0.57 mm, 0.59 mm, and 0.90 mm, respectively.

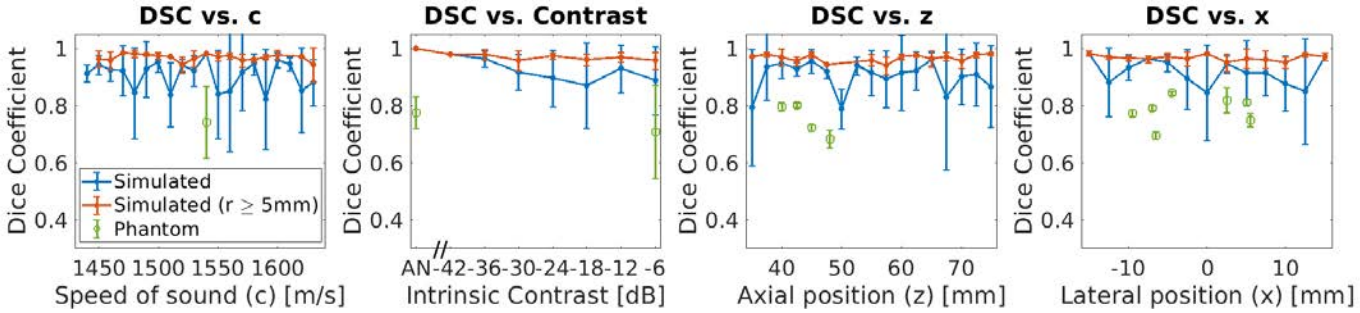

Fig. 3: Performance of the network applied to the test data. Speed of sound (c), intrinsic cyst contrast (where AN = anechoic), axial position of cyst center (z), and lateral position of cyst center (x) were varied in turn aggregating over all other parameters, and the mean Dice similarity coefficient (DSC) $\pm$ one standard deviation are reported for each.

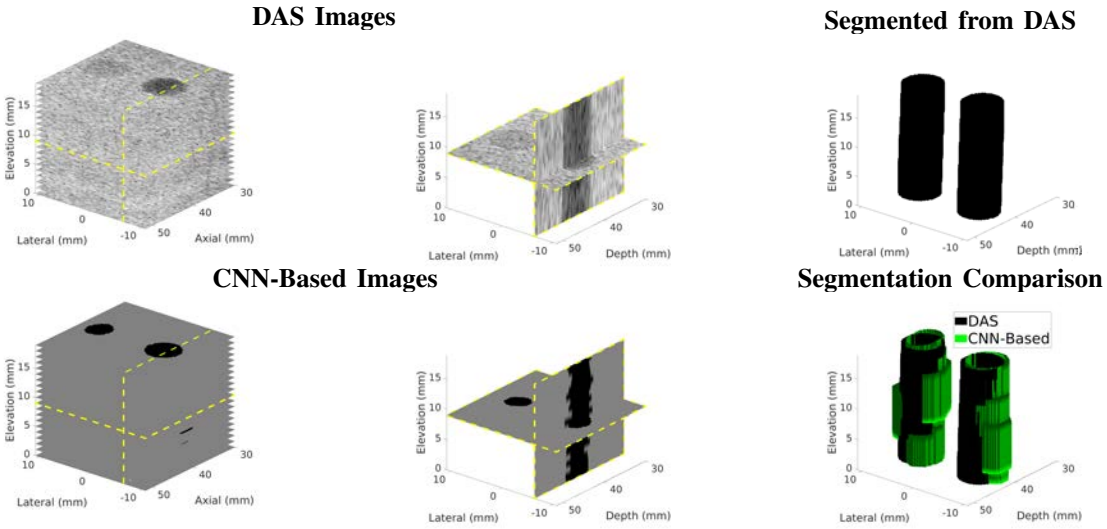

Fig. 4: Volumetric reconstruction of an anechoic and -6 dB hypoechoic cylinder in a CIRS 054GS phantom. Twenty slices in the elevation direction were obtained with an elevation spacing of 1 mm. Predictions were obtained from each image slice, and these predictions were stacked to obtain a volumetric segmentation that was then compared to segmentations based on the ultrasound delay-and-sum (DAS) B-mode image.

IV. Conclusion

A fully convolutional neural network was trained to extract tissue information directly from raw RF channel data prior to the application of any delays to account for time of arrival differences. The overall goal is to produce a segmentation mask that can be directly used as an input for robotic tracking algorithms. The network was successfully transferred to experimental data after training exclusively with simulated cysts. Finally, the network was integrated with a robot and used to construct a volumetric segmentation of a pair of cylinders in the CIRS 054GS phantom. This work demonstrates the promise of using deep learning as a novel ultrasound image formation methodology to directly process RF channel data and obtain image segmentations for robotic tracking tasks.

Acknowledgments

This work is supported by NIH Trailblazer Award R21 EB025621. The authors thank Alycen Wiacek for assistance with the volumetric reconstruction and visualization methods.

References

[1] Mickael Tanter and Mathias Fink, "Ultrafast imaging in biomedical ultrasound," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 61, no. 1, pp. 102–119, 2014.

[2] Robert R Entrekin, Bruce A Porter, Henrik H Sillesen, Anthony D Wong, Peter L Cooperberg, and Cathy H Fix, "Real-time spatial compound imaging: application to breast, vascular, and musculoskeletal ultrasound," in *Seminars in Ultrasound, CT and MRI*. Elsevier, 2001, vol. 22, pp. 50–64.

[3] Kai E Thomenius, "Evolution of ultrasound beamformers," in *Ultrasonics Symposium, 1996. Proceedings., 1996 IEEE*. IEEE, 1996, vol. 2, pp. 1615–1622.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[5] Dezső Ribli, Anna Horváth, Zsuzsa Unger, Péter Pollner, and István Csabai, "Detecting and classifying lesions in mammograms with deep learning," *Scientific reports*, vol. 8, no. 1, pp. 4165, 2018.

[6] Emran Mohammad Abu Anas, Parvin Mousavi, and Purang Abolmaesumi, "A deep learning approach for real time prostate segmentation in freehand ultrasound guided biopsy," *Medical image analysis*, 2018.

[7] Adam C Luchies and Brett C Byram, "Deep neural networks for ultrasound beamforming," *IEEE transactions on medical imaging*, vol. 37, no. 9, pp. 2010–2021, 2018.

[8] Francis Tom and Debdoot Sheet, "Simulating patho-realistic ultrasound images using deep generative networks with adversarial learning," in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 2018, pp. 1174–1177.

[9] Sanketh Vedula, Ortal Senouf, Alex M Bronstein, Oleg V Michailovich, and Michael Zibulevsky, "Towards ct-quality ultrasound imaging using deep learning," *arXiv preprint arXiv:1710.06304*, 2017.

[10] Yeo Hun Yoon, Shujaat Khan, Jaeyoung Huh, and Jong Chul Ye, "Deep learning in rf sub-sampled b-mode ultrasound imaging," *arXiv preprint arXiv:1712.06096*, 2017.

[11] Arun Asokan Nair, Trac D Tran, Austin Reiter, and Muyinatu A Lediju Bell, "A deep learning based alternative to beamforming ultrasound images," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 3359–3363.

[12] 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*. Citeseer, 1996.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas 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.

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

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

[16] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.