A DEEP LEARNING BASED ALTERNATIVE TO BEAMFORMING ULTRASOUND IMAGES

Arun Asokan Nair[⋆]*, Trac D. Tran*[⋆]*, Austin Reiter*[†]*, Muyinatu A. Lediju Bell*[⋆‡]

[⋆]Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, USA
[†]Department of Computer Science, Johns Hopkins University, Baltimore, USA
[‡]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, USA

ABSTRACT

Deep learning methods are capable of performing sophisticated tasks when applied to a myriad of artificial intelligent (AI) research fields. In this paper, we introduce a novel approach to replace the inherently flawed beamforming step during ultrasound image formation by applying deep learning directly to RF channel data. Specifically, we pose the ultrasound beamforming process as a segmentation problem and apply a fully convolutional neural network architecture to segment anechoic cysts from surrounding tissue. We train our network on a dataset created using the Field II ultrasound simulation software to simulate plane wave imaging with a single insonification angle. We demonstrate the success of our architecture in extracting tissue information directly from the raw channel data, which completely bypasses the beamforming step that would otherwise require multiple insonification angles for plane wave imaging. Our simulated results produce mean Dice coefficient of 0.98 ± 0.02, when measuring the overlap between ground truth cyst locations and cyst locations determined by the network. The proposed approach is promising for developing dedicated deep-learning networks to improve the real-time ultrasound image formation process.

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

1. INTRODUCTION

Medical ultrasound imaging uses high-frequency sound waves to image biological tissue. An ultrasound probe consisting of an array of elements transmits sound to a target region that travels through the body and encounters acoustic impedance mismatches that cause the waves to be reflected back to the probe [1]. Advantages of ultrasound imaging over other medical imaging modalities include real-time imaging capabilities, mobility, cost-effectiveness, and lack of harmful ionizing radiation [2]. Diagnostic applications of ultrasound include breast cancer screening [3], liver tumor detection and tracking [4] and blood vessel imaging [5].

This work is partially supported by the NSF under Grant CCF-1422995 and by the NIH under Grant R00 EB018994.

The ultrasound image formation process contains multiple steps after the reflected signals are received by the ultrasound probe. The first step is beamforming, typically performed in any array-based imaging method [6]. Beamforming is applied to sensor array data – i.e., radio frequency (RF) channel data – in order to achieve beam directionality and focusing. Beamforming is then followed by envelope detection, log compression, filtering, and other post-processing steps. One disadvantage of the beamforming step when applied to plane wave imaging is that multiple insonification angles are required to achieve reduced clutter and sufficient spatial resolution, which reduces the potential for higher frame rates [7]. Therefore, transmission of multiple plane waves is not ideal for achieving the highest frame rates possible when solving the inverse problem for both 2D and 3D plane wave imaging [8].

Seemingly unrelated to this particular challenge, deep neural networks (DNNs) have recently achieved state-of-the-art results in numerous AI tasks including image classification [9], image segmentation [10], automatic speech recognition [11] and gaming [12]. DNNs have also found applications in ultrasound imaging, including locating the standard plane in fetal ultrasound images [13], classifying liver [14] and breast lesions [15], and tracking the left ventricle endocardium in cardiac ultrasound images [16]. DNNs were recently applied directly to the RF ultrasound channel data to compress and recover ultrasound images [17] and to operate on sub-band ultrasound channel data after conversion to the frequency domain [18]. However, to the authors' knowledge, there are no applications of DNN to investigate a direct image-to-image transformation from RF channel data to an output representation understandable by a human, entirely bypassing both beamforming and other post processing steps.

This work is the first to extract image details directly from the received ultrasound channel data without beamforming. We achieve this goal by employing a U-Net [10] type image-to-image segmentation network that takes RF channel data as the input and learns a transformation to the segmentation mask of the scene. Three possible advantages include:

1. Speed - beamforming of plane wave data typically requires multiple insonification angles that are combined

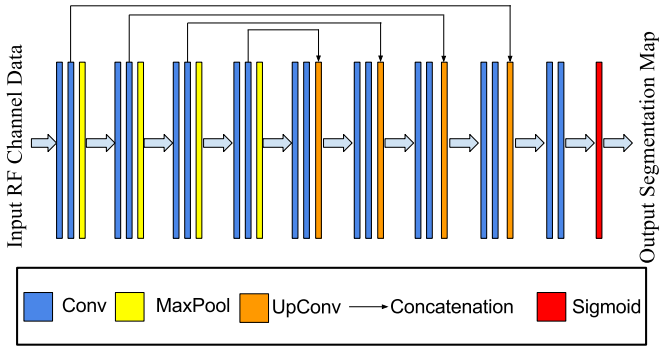

Fig. 1. Fully convolutional encoder-decoder architecture with skip connections for ultrasound image segmentation.

into an image with receive beamforming. We aim to reduce the number of transmissions required and thereby expect to increase the imaging speed beyond current capabilities with plane wave imaging.

2. Noise suppression - the trained neural network is taught to suppress typical artifacts that would be present in plane wave images created from a single insonificaton angle, such as acoustic clutter, which appears as a hazy structure that "fills in" anechoic regions [19, 20]

3. Accuracy - the beamforming process is only an approximate solution to the inverse problem that is not entirely accurate in the presence of multiple tissues with multiple varying acoustic properties. With enough training data, we expect the DNN to learn a better inversion function.

In Section 2 of this paper, we provide an overview of the neural network model we employ. In Section 3, we discuss details of the dataset we used to train our network along with the parameters we used for training. Section 4 details our achievements when applying this model to simulated anechoic cysts of varying sizes and locations and when embedded in tissues with varying sound speeds. Finally, we offer concluding remarks in Section 5.

2. ARCHITECTURE

Our neural network architecture is based on the widely used U-Net [10] segmentation network. The architecture, as seen from Fig. 1, is fully convolutional and has two major parts – a contracting encoder path and an expanding decoder path.

In the contracting encoder, we have convolutional (Conv) layers and max pooling (MaxPool) layers. For each convolutional (Conv) layer, we employ 3×3 convolutions with a stride of 1, zero padding the input in order to ensure the sizes of the input and output match. We use rectified linear units (ReLU) [21] as our non-linearity in the Conv layers. For the max-pooling layers, we employ a pool size of 2×2 with stride set to 2 in each direction as well. Each max pool layer thus has an output size half that of the input (hence the term 'contracting'). To offset this, we also increase the number of feature channels learned by 2 after every max pooling step.

In the expanding decoder, we have up-convolutional (UpConv) layers, also termed transposed convolutions in addition to regular convolutional layers. The UpConv layers reverse the reduction in size caused by the convolution and max pooling layers in the encoder by learning a mapping to an output size twice the size of the input. As a consequence, we also halve the number of feature channels learned in the output. The output of each UpConv layer is then concatenated with the features generated by the segment of the encoder corresponding to the same scale, before being passed to the next part of the decoder. The reason for this is two-fold: to explicitly make the network consider fine details at that scale that might have been lost during the down sampling process, and to allow the gradient to back-propagate more easily through the network through these 'skip' or 'residual' connections [22], reducing training time and training data requirements.

The final layer of the network is a 1×1 convolutional layer with a sigmoid non-linear function. The output is a per-pixel confidence value of whether the pixel corresponds to the cyst region (predict 1) or tissue region (predict 0) based on the learned multi-scale features. We train the network end-to-end, using the negative of a differentiable formulation of the Dice similarity coefficient (Eq. 1) as our training loss.

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

where X corresponds to vectorized predicted segmentation mask and Y corresponds to the vectorized ground truth mask.

3. EXPERIMENTAL SETUP

3.1. Field II Dataset

In order to train our network, we simulate a large dataset using the open-source Field II [23] ultrasound simulation software. All simulations considered a single, water-filled anechoic cyst in normal tissue with our region of interest maintained between -19.2 mm and +19.2 mm in the lateral direction and between 30 mm and 80 mm in the axial direction. The transducer was modeled after an Alpinion L3-8 linear array trans-

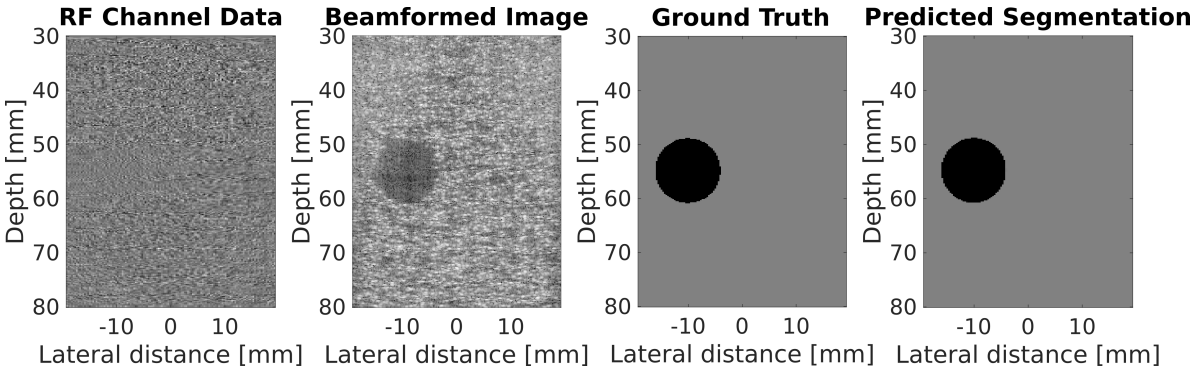

Fig. 2. Example of RF channel data that is typically beamformed to obtain a readable ultrasound image. A ground truth mask of the anechoic cyst location is compared to the the mask predicted by our neural network. Our network provides a clearer view of the cyst location when compared to the conventional ultrasound image created with a single plane wave transmission.

ducer with parameters provided in Table 1. Plane wave imaging was implemented [24] with a single insonification angle of 0°. RF channel data corresponding to a total of 21 different sound speeds (1440 m/s to 1640 m/s in increments of 10 m/s), 7 cyst radii (2 mm to 8 mm in increments of 1 mm), 13 lateral positions (-15 mm to 15 mm in steps of 2.5 mm) for the cyst center, and 17 axial locations (35 mm to 75 mm in steps of 2.5 mm) for the cyst center were considered, yielding a total of 32,487 simulated RF channel data inputs after 10,000 machine hours on a high performance cluster. We then performed a 80:20 split on this data, retaining 25,989 images as training data and using the remaining 6,498 as testing data. We further augmented only the training data by flipping it laterally to simulate imaging the same regions with the probe flipped laterally. We resized the original channel data from an initial dimensionality of 2440*128 to 256*128 in order to fit it in memory, and normalized by the maximum absolute value to restrict the amplitude range from -1 to +1.

Table 1. Ultrasound transducer parameters

Parameter	Value
Element number	128
Pitch	0.30 mm
Aperture	38.4 mm
Element width	0.24 mm
Transmit Frequency	8 MHz
Sampling Frequency	40 MHz

3.2. Network Implementation

All neural network code was written in the Keras API [25] on top of a TensorFlow [26] backend. Our network was trained for 20 epochs using the Adam optimizer [27] with a learning rate of $1e^{-5}$ on negative Dice loss (Eq. 1). Weights of all neurons in the network were initialized using the Glorot uniform initialization scheme. Mini-batch size was chosen to be 16 samples to attain a good trade-off between memory requirements and convergence speed. The training of the neural network was performed on an NVIDIA Tesla P40 GPU with 24 GB of memory.

4. RESULTS AND DISCUSSIONS

4.1. Qualitative Assessment

As visible from Fig. 2, deep learning enables a new kind of ultrasound image – one that does not depend on the classical method of beamforming. Using a fully convolutional encoder-decoder architecture, we extract details directly from the non human-readable RF channel data and produce a segmentation mask for the region of interest. This also allows us to overcome common challenges with ultrasound, like the presence of acoustic clutter when using a single insonification angle in plane wave imaging. We also ignore the presence of speckle, which provides better object detectability, although this feature can be considered a limitation for techniques that rely on the presence of speckle.

As a consequence, the final output image is more interpretable than the corresponding beamformed image created with a single plane wave insonification. In addition to requiring less time to create this image, thereby increasing possible real-time frame rates, this display method would require less expert training to understand. Our method can also serve as supplemental information to experts in the case of difficult-to-discern tissue features in traditional beamformed ultrasound images. In addition, it can also be employed as part of a real-time fully automated robotic tracking system [28].

4.2. Objective evaluation

To objectively assess the performance of the neural network, we employ four evaluation criteria:

1. **Dice score** - This is the loss metric that was used to train the neural network as described by Eq. 1. The mean Dice scores for the test data samples was evalu-

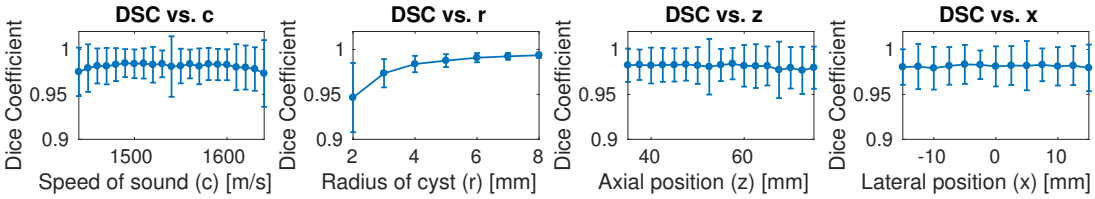

Fig. 3. Performance variation of the trained network versus different simulation conditions. We varied cyst radius (r), speed of sound (c), axial position of cyst center (z), and lateral position of cyst center (x), aggregating over all other parameters, and calculated the mean Dice similarity coefficient (DSC). The error bars show $\pm$ one standard deviation.

ated to be a promisingly high value of 0.9815 $\pm$ one standard deviation of 0.0222.

2. **Contrast** - Contrast is a common measure of image quality, particularly when imaging cysts. It measures the signal intensity ratio between two tissues of interest, in our case between that of the cyst and the tissue:

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

where S_i and S_o are the mean signal intensities inside and outside the cyst, respectively. This measurement provides quantitative insight into how discernible the cyst is from its surroundings. A major advantage of our approach to image formation is that segmentation into cyst and non-cyst regions is produced with high confidence, which translates to very high contrast. For the example images shown in Fig. 2, the cyst contrast in the conventionally beamformed ultrasound image is 10.15 dB, while that of the image obtained from the network outputs is 33.77 dB, which translates to a 23.62 dB improvement in contrast for this example. Overall, the average contrast for network outputs was evaluated to be 45.85 dB (when excluding results with infinite contrast due to all pixels being correctly classified).

3. **Recall** - Also known as specificity, recall is the fraction of positive examples that are correctly labeled as positive. For our network, we define a test example as correctly labeled if at least 75% of the cyst pixels were correctly labeled as belonging to a cyst. Our network yields a recall of 0.9977. This metric indicates that clinicians (and potentiality robots) will accurately detect at least 75% of cyst over 99% of the time.

4. **Time** - The time it takes to display our DNN-based images is related to our ability to increase the real-time capabilities of plane wave imaging. We processed the 6,498 test images in 53 seconds using the DNN, which translates to a frame rate of approximately 122.6 frames/s on our single-threaded CPU. Using the same data and computer, conventional beamforming took 3.4 hours, which translates to a frame rate of approximately 0.5 frames/s. When plane wave imaging is implemented on commercial scanners with custom computing hardware, the frame rates are more like 350 frames/s for 40 insonification angles [24]. However, we are only using one insonification angle, which indicates that our approach can reduce the acquisition time for plane wave imaging and still achieve real time frame rates while enhancing contrast.

4.3. Performance variations with simulation parameters

We evaluated the Dice coefficient produced by our network as functions of four simulation parameters: cyst radius (r), speed of sound (c), axial position of cyst center (z), and lateral position of cyst center (x). We calculated the average Dice coefficients when fixing the parameter of interest and averaging over all other parameters. The results are shown in Fig. 3.

In each case, the mean Dice coefficients was always greater than 0.94, regardless of variations in the four simulated parameters. Varations in Dice coefficients were most sensitive to cyst size. The Dice coefficients were lower for smaller cysts, with performance monotonically increasing as cyst size increased. This increase with size is likely a result of smaller cysts activating fewer neurons that the network can aggregate for a prediction, and also mirrors traditional ultrasound imaging, where cysts of smaller size are more difficult to discern [29]. Otherwise, the network appears to be more robust to changes in sound speed and the axial and lateral positions of the anechoic cyst.

5. CONCLUSIONS

This work is the first to demonstrate the feasibility of employing deep learning as an alternative to traditional ultrasound image formation and beamforming. Our network is a fully convolutional encoder-decoder that aggregates information learned from the input channel data at multiple scales in order to directly produce a segmentation map of tissue. As a consequence, not only would our approach be faster than traditional plane wave ultrasound imaging, but it also learns to recognize and suppress speckle and clutter noise. Future work includes training and testing with multiple cysts and point targets as well as extending the framework to an end-to-end DNN that can automatically identify, track, and recognize objects of interest. We also note that application to point targets has previously shown promise in related photoacoustic imaging deep learning methods [30, 31].

6. REFERENCES

[1] Philip ES Palmer et al., *Manual of diagnostic ultrasound*, World Health Organization, 1995.

[2] Thomas L Szabo, *Diagnostic ultrasound imaging: inside out*, Academic Press, 2004.

[3] Wendie A Berg et al., "Detection of breast cancer with addition of annual screening ultrasound or a single screening mri to mammography in women with elevated breast cancer risk," *Jama*, vol. 307, no. 13, pp. 1394–1404, 2012.

[4] V De Luca et al., "The 2014 liver ultrasound tracking benchmark," *Physics in medicine and biology*, vol. 60, no. 14, pp. 5571, 2015.

[5] Jukka T Salonen and Riitta Salonen, "Ultrasound b-mode imaging in observational studies of atherosclerotic progression.," *Circulation*, vol. 87, no. 3 Suppl, pp. II56–65, 1993.

[6] Barry D Van Veen and Kevin M Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE assp magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[7] Bruno Madore et al., "Accelerated focused ultrasound imaging," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 56, no. 12, 2009.

[8] Jean-François Cardoso and Antoine Souloumiac, "Blind beamforming for non-gaussian signals," in *IEE proceedings F (radar and signal processing)*. IET, 1993, vol. 140, pp. 362–370.

[9] Alex Krizhevsky et al., "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[10] 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.

[11] Geoffrey Hinton et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[12] David Silver et al., "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[13] Hao Chen et al., "Standard plane localization in fetal ultrasound via domain transferred deep neural networks," *IEEE journal of biomedical and health informatics*, vol. 19, no. 5, pp. 1627–1636, 2015.

[14] Kaizhi Wu et al., "Deep learning based classification of focal liver lesions with contrast-enhanced ultrasound," *Optik-International Journal for Light and Electron Optics*, vol. 125, no. 15, pp. 4057–4063, 2014.

[15] Jie-Zhi Cheng et al., "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans," *Scientific reports*, vol. 6, pp. 24454, 2016.

[16] Gustavo Carneiro and Jacinto C Nascimento, "Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 11, pp. 2592–2607, 2013.

[17] Dimitris Perdios et al., "A deep learning approach to ultrasound image recovery," in *IEEE International Ultrasonics Symposium*, 2017, number EPFL-CONF-230991.

[18] Adam Luchies and Brett Byram, "Deep neural networks for ultrasound beamforming," in *IEEE International Ultrasonics Symposium*, 2017.

[19] Sabine Huber, Monika Wagner, Michael Medl, and Heinrich Czembirek, "Real-time spatial compound imaging in breast ultrasound," *Ultrasound in medicine & biology*, vol. 28, no. 2, pp. 155–163, 2002.

[20] Muyinatu A Lediju et al., "Quantitative assessment of the magnitude, impact and spatial extent of ultrasonic clutter," *Ultrasonic imaging*, vol. 30, no. 3, pp. 151–168, 2008.

[21] Vinod Nair and Geoffrey E Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[23] 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.

[24] 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.

[25] François Chollet et al., "Keras," 2015.

[26] Martín Abadi et al., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

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

[28] Joshua Shubert and Muyinatu A Lediju Bell, "Photoacoustic based visual servoing of needle tips to improve biopsy on obese patients," in *IEEE International Ultrasonics Symposium*, 2017.

[29] Wendie A Berg et al., "Cystic breast masses and the acrin 6666 experience," *Radiologic Clinics of North America*, vol. 48, no. 5, pp. 931–987, 2010.

[30] Derek Allman, Austin Reiter, and Muyinatu A Lediju Bell, "A machine learning method to identify and remove reflection artifacts in photoacoustic channel data," in *IEEE International Ultrasonics Symposium*, 2017.

[31] Austin Reiter and Muyinatu A Lediju Bell, "A machine learning approach to identifying point source locations in photoacoustic data," in *Proc. of SPIE Vol*, 2017, vol. 10064, pp. 100643J–1.