Multi-task learning for ultrasound image formation and segmentation directly from raw *in vivo* data

Manish Bhatt[1]*, Arun Asokan Nair[1]*, Kelley M. Kempski[2], Muyinatu A. Lediju Bell[1,2,3]
[1]*Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, United States*
[2]*Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, United States*
[3]*Department of Computer Science, Johns Hopkins University, Baltimore, MD, United States*

***Abstract*—Deep neural networks have demonstrated potential to both create images and segment structures of interest directly from raw ultrasound data in one step, through an end-to-end transformation. Building on previous work from our group, subaperture beamformed IQ data from *in vivo* breast cyst data was the input to a custom network that outputs parallel B-mode and cyst segmentation images. Our new model includes bright point and line targets during training to overcome the limited field of view challenges encountered with our previous deep learning models, which were purely trained using simulations of cysts and homogeneous tissue structures. This new network resulted in cyst contrast values of -33.07 ± 10.79 dB, -32.09 ± 0.04 dB, and -15.95 ± 12.04 dB for simulated, phantom, and *in vivo* data, respectively, which is an improvement over the contrast of corresponding delay and sum (DAS) images (i.e., -17.37 ± 6.06 dB, -17.14 ± 0.16 dB, and -14.80 ± 1.30 dB for simulated, phantom, and *in vivo*, respectively). Higher dice similarity coefficients (DSCs) were obtained with *in vivo* data with the new network (0.83 ± 0.01) when compared to our previous model (0.63 ± 0.03), and fewer false positives were encountered. This work demonstrates the feasibility of using multi-task learning to simultaneously form a B-mode image and cyst segmentation with a wider field of view that is appropriate for *in vivo* breast imaging. These results have promising implications for multiple tasks, including emphasizing or de-emphasizing structures of interest for diagnostic, interventional, automated, and semi-automated decision making.**

***Index Terms*—Ultrasound imaging, deep learning, beamforming, image segmentation, *in vivo* breast imaging**

I. Introduction

Advantages of ultrasound imaging include its portability, cost effectiveness, and use of non-ionizing radiation, which makes it a widely used imaging technique for diagnosing various diseases, including breast cancer [1]. However, ultrasound imaging is known to suffer from image interpretation challenges that arise due to speckle, clutter, and inefficient filtering, often leading to indecisive segmentations. Our group previously demonstrated that deep learning-based techniques have the potential to overcome these challenges [2]. In particular, deep neural networks (DNN) were implemented to perform end-to-end transformation tasks directly from subaperture beamformed in-phase and quadrature (IQ) channel data while also overcoming the challenges of acoustic clutter and inefficient filtering [2]–[5].

One of the challenges with this end-to-end transformation is the absence of receive focusing delays when learning information directly from raw channel data. The DNN input often consists of data dimensions of time vs. channels while the imaging output requires data to be displayed as depth vs. width. To overcome this challenge, Nair *et al.* [2] developed a DNN architecture that has the ability to map the temporal data recorded on multiple channels to a single pixel in the lateral image direction, thus performing the task of transforming raw IQ data to final output images. The ability to map temporal recordings to pixel locations enables the DNN to take advantage of lower spatial frequencies available with raw, complex, baseband, IQ data.

This end-to-end transformation model [2] performed multi-task learning (a term described in more detail in [6]) by successfully reconstructing B-mode ultrasound images in addition to improving breast cyst segmentations. However, outstanding limitations were encountered when testing on *in vivo* data, including misidentifying hypoechoic regions in the image as anechoic cysts, resulting in a limited field of view for successful implementation. This paper addresses these limitations by incorporating more representative features during the network training process. Our results highlight the *in vivo* success of task-specific DNNs in support of our larger vision to provide multiple outputs from a single input of subaperture beamformed IQ data with applications to improving automated and semi-automated ultrasound-based decision making.

II. Materials and Methods

A. Deep neural network architecture

A DNN based on the U-Net architecture [7] was modeled to reconstruct and segment cysts from *in vivo* breast tissue. After performing receive delays to construct a 3D data tensor (depth × scanlines × 128 elements), we define one subaperture as the summation of the delayed data received by 8 adjacent elements. The network input of 16 subapertures of beamformed IQ data (i.e., 2 IQ channels per subaperture, 32 IQ channels total) produced two parallel outputs after multi-task learning: (1) an interpretable B-mode image that emphasizes structures of interest and de-emphasizes less important surrounding structures for the proposed task and (2) a segmented cyst image. The network consisted of one encoder and two decoders to extract information directly from subaperture beamformed IQ data received after a single plane wave insonification. The first decoder generated a DNN image similar to a delay and sum (DAS) beamformed image, and the

*These authors contributed equally.

978-1-7281-5448-0/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: Johns Hopkins University. Downloaded on January 14,2021 at 16:13:28 UTC from IEEE Xplore. Restrictions apply.

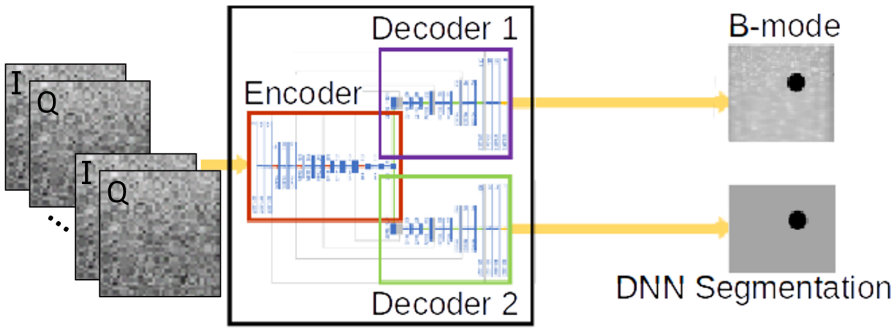

Fig. 1. Illustration of the DNN architecture employed in this study. The input is in-phase/quadrature (IQ), subaperture beamformed ultrasound data that simultaneously outputs both a DNN image and a DNN segmentation directly from raw ultrasound channel data received after a single plane wave insonification.

second decoder generated a DNN segmentation of the cyst. The employed DNN architecture is summarized in Fig. 1.

The dimensions of the focused downsampled subaperture IQ channel data (I_{fds}) were of size $d \times w \times q_s$, where d is the length of IQ signal in the depth direction, w is the image width (which is same as the number of transducer element receive channels in our implementation), and q_s has twice the number of subapertures, each representing the in-phase or quadrature component of the recording. The network produces predictions of a DNN beamformed image and a corresponding cyst segmentation map, each with dimensions $d \times w$. A fully convolutional neural network was employed to learn the optimal mapping of $I_{fds} \rightarrow y$, where y is the ground truth reference for the optimal mapping. This reference consisted of a true segmentation map, S_t, and the corresponding true enhanced beamformed image, E. Thus, y describes the tuple (E, S_t). Additional details about this network architecture are available in Section II-E of [2].

B. Numerical simulations

The neural network trained by Nair *et al.* [2] and described in Section II-A (named CystNet1 hereafter) did not include bright point and line targets in the training dataset and, thus, displayed limitations while performing *in vivo* imaging tasks. Therefore, we extended the training dataset to include numerical simulations containing bright point and line targets. This new network is named CystNet2.

The dataset for DNN learning was created using numerical simulations (created with Field II software [8], [9]) of individual, anechoic, cylindrical cysts, positioned at 30-80 mm depth, within cuboidal phantoms of 50 mm axial depth, 40 mm lateral width, and 7 mm elevation thickness. A total of 66,690 of these simulations were created to include a combination of these cyst targets, as well as point targets and bright line targets. The simulated medium had cyst radii ranging 2-8 mm, axial positions ranging 40-70 mm, and lateral positions ranging -16 to 0 mm. The speed of sound ranged 1420-1600 m/s. A random number generator with a unique seed was employed to generate unique speckle realizations for each phantom and thereby model the diversity expected within clinical datasets. A total of 50,000 scatterers were modeled to ensure fully developed speckle. The simulated dataset was divided into training (80%), testing (10%), and validation (10%) sets.

C. Phantom and in vivo breast data acquisition

In addition to simulated data, phantom and *in vivo* ultrasound data were also acquired for network testing. The phantom data was acquired from two anechoic cylindrical inclusions in a CIRS 054GS phantom. These inclusions were located at depths of 40 mm and 70 mm. An Alpinion E-Cube 12R research scanner was used to acquire the channel data with an Alpinion L3-8 linear ultrasound transducer. Two independent 80-frame sequences were acquired. All phantom channel data were flipped to augment the dataset, resulting in 320 total experimental phantom images.

The two *in vivo* breast cysts described in [2] were additionally tested. The first cyst (denoted as cyst #1) was imaged with an 80 frame plane wave imaging sequence using the same ultrasound equipment described in the preceding paragraph. The second cyst (denoted as cyst #2) was imaged with a 10 frame ultrasound focused transmissions sequence using an Alpinion L8-17 probe. These *in vivo* images were acquired with Johns Hopkins Medicine Institutional Review Board approval and informed consent. The acquired datasets were transformed into focused subaperture beamformed IQ data to use as input to the DNN [2]. The ground truth segmentations for these datasets were manually created from B-mode images. All *in vivo* channel data were flipped to augment the dataset, resulting in a total of 180 images.

D. Image quality metrics

Contrast, signal-to-noise ratio (SNR), and generalized contrast-to-noise ratio (gCNR) [10] were utilized to evaluate DNN performance, using the following equations:

$$Contrast = 20 \log_{10} \frac{S_i}{S_o} \tag{1}$$

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

$$gCNR = 1 - \sum_{x=0}^{1} \min(p_i(x), p_o(x)) \tag{3}$$

Authorized licensed use limited to: Johns Hopkins University. Downloaded on January 14,2021 at 16:13:28 UTC from IEEE Xplore. Restrictions apply.

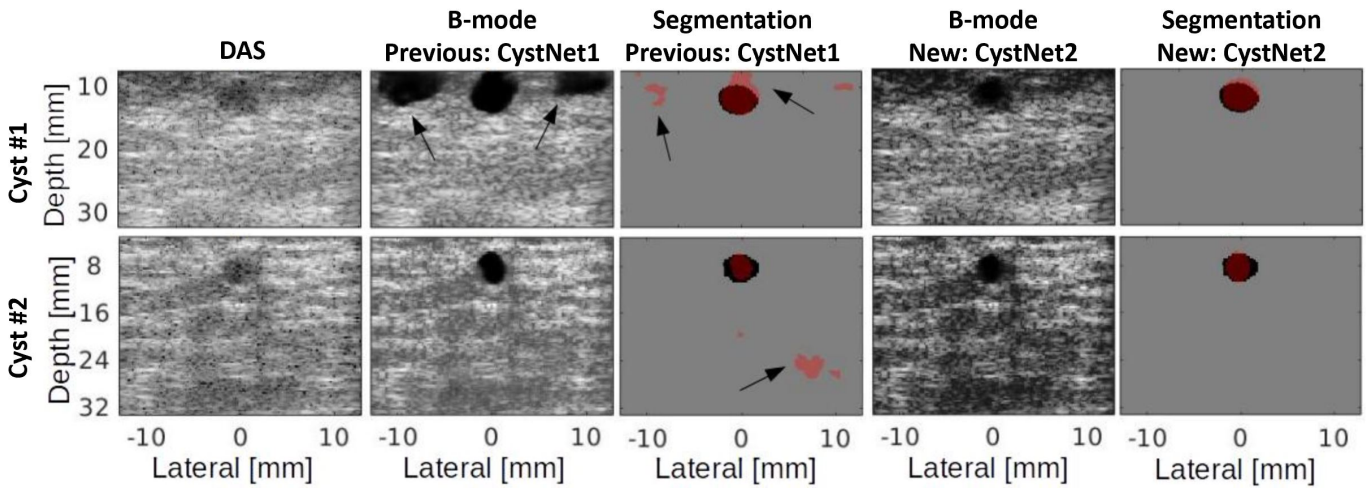

Fig. 2. Image formation and segmentation results for *in vivo* breast cyst #1 and cyst #2. The delay and sum (DAS) B-mode image, DNN reconstruction and cyst segmentation for the previous network (i.e., CystNet1), and the DNN and cyst segmentation outputs for the proposed network (i.e., CystNet2) are shown. Arrows point to network failure.

where S_i and S_o are the mean signals inside and outside the cyst region, respectively, σ_o is the standard deviation of signals outside the cyst region, and $p_i(x)$ and $p_o(x)$ represent the probability mass functions of the signal inside and outside the cyst region, respectively.

Peak signal to noise ratio (PSNR) was used to quantify the similarity of a generated DNN image to its reference DAS beamformed image, using the following equation:

$$PSNR(D,R) = 10\log_{10}\frac{MAX_R^2}{MSE} \tag{4}$$

where D is the output DNN B-mode image, R is the normalized reference image, MSE is the mean square error between D and R. MAX_R denotes the maximum absolute pixel value of the reference image R, which is equal to 1 here.

The Dice similarity coefficient (DSC) was calculated to quantify overlap between two segmentation masks [2], [11], expressed as:

$$DSC(S_p,S_t) = 2\frac{|S_p \cap S_t|}{|S_p| + |S_t|} \tag{5}$$

where S_p is the predicted DNN segmentation and S_t is the true segmentation. A perfect DNN segmentation produces a DSC of 1.

III. Results and Discussion

Figure 2 (top) displays B-mode and cyst segmentation images for cyst #1 obtained using CystNet1 and CystNet2. As observed in previous work [2], CystNet1 produced false positives in reconstructed images, which led to a limited field of view. CystNet2 successfully overcomes these challenges and provides cleaner and robust B-mode reconstruction and cyst segmentation. Higher DSCs were obtained with *in vivo* cyst segmentation performed with CystNet2 (0.83 $\pm$ 0.01) when compared to the segmentation performed with CystNet1 (0.63 $\pm$ 0.03). The image quality metrics for this dataset presented in Table I indicate that CystNet2 generally achieves either similar or improved image quality compared to both CystNet1 and DAS imaging.

Figure 2 (bottom) presents B-mode and cyst segmentation results for cyst #2 achieved using CystNet1 and CystNet2. Similar to the results obtained when testing with cyst #1, CystNet1 produced false positives in the cyst segmentation image, and CystNet2 overcame this and associated limited field of view challenges discussed previously. Higher dice similarity coefficients (DSCs) were obtained with the segmentation performed with CystNet2 (0.79 $\pm$ 0.01) when compared to segmentation performed with CystNet1 (0.36 $\pm$ 0.07). A comparison of the image quality metrics for this dataset (listed in Table I) reveals that CystNet2 generally achieves

TABLE I
QUANTITATIVE COMPARISON OF CYSTNET2 WITH CYSTNET1 [2] AND DELAY AND SUM BEAMFORMING (DAS)

	Cyst #1			Cyst #2		
	CystNet2	CystNet1	DAS	CystNet2	CystNet1	DAS
DSC	**0.83 $\pm$ 0.01**	0.63 $\pm$ 0.03	-	**0.79 $\pm$ 0.01**	0.36 $\pm$ 0.07	-
Contrast [dB]	-15.95 $\pm$ 2.04	**-28.84 $\pm$ 1.98**	-14.80 $\pm$ 1.30	-24.21 $\pm$ 4.97	**-32.56 $\pm$ 4.16**	-21.33 $\pm$ 2.18
SNR	**5.88 $\pm$ 0.49**	5.18 $\pm$ 1.17	0.95 $\pm$ 0.15	9.81 $\pm$ 0.19	**9.85 $\pm$ 0.14**	1.06 $\pm$ 0.11
gCNR	**0.61 $\pm$ 0.02**	0.48 $\pm$ 0.18	0.61 $\pm$ 0.05	0.83 $\pm$ 0.04	**0.89 $\pm$ 0.02**	0.75 $\pm$ 0.10
PSNR [dB]	12.02 $\pm$ 0.03	**16.66 $\pm$ 0.13**	-	11.10 $\pm$ 0.07	**18.69 $\pm$ 0.14**	-

Authorized licensed use limited to: Johns Hopkins University. Downloaded on January 14,2021 at 16:13:28 UTC from IEEE Xplore. Restrictions apply.

either better or comparable performance to CystNet1 and DAS images.

The mean $\pm$ standard deviation of the contrast of CystNet2 ultrasound images created from simulated and phantom test data were -33.07 ± 10.79 dB and -32.09 ± 0.04 dB, respectively, which is an improvement over the contrast of corresponding DAS images (i.e., -17.37 ± 6.06 dB and -17.14 ± 0.16 dB, respectively). CystNet2 generated simulation and phantom test images with mean $\pm$ standard deviation SNR values of 2.01 ± 0.63 and 1.78 ± 0.01, respectively, compared to 1.95 ± 0.30 and 1.96 ± 0.01 of the DAS beamformed images, respectively. These results support previous reports, demonstrating that DNN-based end-to-end transformation models have the potential to provide superior contrast and similar SNR when compared to conventional DAS beamforming [2]–[4].

IV. Conclusion

The work presented in this paper demonstrates the ability of DNNs to perform simultaneous image formation and cyst segmentation, using a multi-task learning approach, producing a wide field of view for *in vivo* breast images when incorporating a range of representative training features that include cysts, tissue, lines, and points. We envisage that this work will enhance the potential of ultrasound imaging for multiple diagnostic tasks. Tasks requiring emphasis or de-emphasis of specific structures are particularly well suited to benefit from the proposed approach.

Acknowledgments

The authors acknowledge funding support from NIH Trailblazer Award R21-EB025621.

References

[1] Kevin M Kelly, Judy Dean, W Scott Comulada, and Sung-Jae Lee. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *European Radiology*, 20(3):734–742, 2010.

[2] Arun Asokan Nair, Kendra N Washington, Trac D Tran, Austin Reiter, and Muyinatu A Lediju 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.

[3] 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)*, pages 3359–3363. IEEE, 2018.

[4] Arun Asokan Nair, Trac D Tran, Austin Reiter, and Muyinatu A Lediju Bell. One-step deep learning approach to ultrasound image formation and image segmentation with a fully convolutional neural network. In *2019 IEEE International Ultrasonics Symposium (IUS)*, pages 1481–1484. IEEE, 2019.

[5] Alycen Wiacek, Eduardo González, and Muyinatu A Lediju Bell. Coherenet: A deep learning architecture for ultrasound spatial correlation estimation and coherence-based beamforming. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2020.

[6] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv:1706.05098*, 2017.

[7] 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*, pages 234–241. Springer, 2015.

[8] 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.

[9] Jørgen Arendt Jensen and Niels Bruun Svendsen. Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 39(2):262–267, 1992.

[10] Alfonso Rodriguez-Molares, Ole Marius Hoel Rindal, Jan D'hooge, Svein-Erik Måsøy, Andreas Austeng, Muyinatu A Lediju Bell, and Hans Torp. The generalized contrast-to-noise ratio: a formal definition for lesion detectability. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(4):745–759, 2019.

[11] Kelly H Zou, Simon K Warfield, Aditya Bharatha, Clare MC Tempany, Michael R Kaus, Steven J Haker, William M Wells III, Ferenc A Jolesz, and Ron Kikinis. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. *Academic Radiology*, 11(2):178–189, 2004.

Authorized licensed use limited to: Johns Hopkins University. Downloaded on January 14,2021 at 16:13:28 UTC from IEEE Xplore. Restrictions apply.