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

High Frame-Rate Ultrasound Imaging Using Deep Learning Beamforming

Muhammad Usman Ghani
Department of Electrical Computer Engineering
Boston University
Boston, MA, USA
mughani@bu.edu

F. Can Meral
Philips Research North America
Cambridge, MA, USA
can.meral@philips.com

Francois Vignon
Philips Research North America
Cambridge, MA, USA
francois.vignon@philips.com

Jean-luc Robert
Philips Research North America
Cambridge, MA, USA
jean-luc.robert@philips.com

***Abstract*— Unfocused transmit beams such as diverging waves (DW) and coherent compounding are essential in achieving higher volumetric frame rates in 3D ultrasound imaging. However, image quality loss that comes with the use of DW becomes an issue, especially when the number of transmits is small. We propose a deep learning beamforming method for eliminating some of the artifacts associated with DW imaging. We train a convolutional neural network to map the non-linear transformation between the aligned per-channel data from 11 DW transmits, before compounding, to the compounded per-channel data from 51 transmits. We include additional terms in our loss function such as the beamsum value and log-detected image pixel value to guide the learning in the desired direction. The neural network is trained and tested on simulation and in-vivo data. The final network successfully suppresses acoustic artifacts such as side lobe and clutter in the images obtained with 11 DW transmits.**

Keywords—.echocardiography, beamforming, fast volumetric imaging, convolutional neural networks, deep learning, loss function.

I. Introduction

High-frame rate 3D ultrasound (US) imaging is of high clinical importance for Echocardiography applications. Data acquisition is the limiting factor with conventional focused transmit beams in realizing high-frame rate. Coherent compounding of unfocused transmit patterns such as diverging wave (DW) imaging has been used to improve temporal resolution. These transmission techniques provide larger coverage and allow high-frame rate imaging with fewer acquisitions at the cost of reduced image quality [1][2].

Ultrasound image formation can be considered as an inverse problem, in which the imaged medium is estimated using the data that is acquired through insonifications of the media. Data driven machine-learning methods, such as deep-learning, are shown to be more effective than the analytical methods, which consisted of handcrafted algorithms that require precise tuning of parameters, for inverse problems [3]. There has been an ever-increasing interest in applying deep learning to the problems of computer vision and image processing. Deep learning methods are gaining popularity in ultrasound signal processing, beamforming and image processing.

Following are some of the recent efforts in which neural network and deep learning methods are applied to beamforming, ultrasound image processing and image quality improvement problems. Luchies *et al.* [4] used fully connected neural networks to perform improved beamforming. Their network and learning task is intended to form images from the main lobe of the ultrasound beam, which they called 'acceptance region', while suppressing signals from other directions, i.e. 'rejection region', to improve quality. They trained multiple networks operating at different frequency bands, using the Fourier transform (FT) of the aligned per-channel data from simulations as inputs; and final beamsummed signal, if the signals originate from the acceptance region; or zero, if they come from the rejection region, as the output. In [5] Gasse *et al.*, used neural networks on the beam-summed radio-frequency (RF) data to improve the image quality of plane wave images acquired at three different transmit angles by learning a mapping to the images obtained through compounding thirty-one transmits. Authors trained a convolutional neural network (CNN) using data acquired on healthy volunteers and imaging phantoms. Similar problem was addressed in [6], where the authors trained a U-Net architecture [7] from the input-output pair of low quality images reconstructed from a single plane wave transmit, and synthetic aperture images reconstructed from full transmit and receive dataset. They used simulation data to train their networks and showed improved results on in-vivo test cases. Yoon *et al.* [8] used fully convolutional neural networks to restore images from subsampled (in receive or transmit and receive) RF-data by interpolating the missing receive elements or transmit events. They trained their network using in-vivo data acquired from volunteers. Vedula *et al.* [9] and Senouf *et al.* [10] also used a U-Net architecture to solve several problems associated with high frame rate ultrasound imaging techniques, such as multiline transmission (MLT) and multiline acquisition (MLA). In these studies, the inputs to the network was delayed element-wise MLT and MLA in-phase/quadrature (I/Q) data and outputs were corresponding single-line transmission or single-line acquisition data, respectively. They used in-vivo cardiac data and phantom data for training and testing the network performance.

Motivated by the success of these early investigations, we explore the potential of CNN's applied to ultrasound beamforming process to improve image quality. We train

CNN's to learn the non-linear mapping between aligned per-channel RF-data acquired with fewer DW transmissions and RF-data from larger number of DW acquisitions. We choose to investigate the per-channel domain learning as opposed to beamsummed RF-data as it was reported in the literature for similar problems. Our motivation is to exploit the acoustic artifact structure in per-channel RF-data. We train the CNN to learn such structure in data obtained from an ultrasound pulse sequence with reduced number of transmissions and estimate the channel data with reduced artifacts, which is obtained from a pulse sequence of higher number of transmissions. The inputs to the network consists of per-channel data, after transmit and receive related delays are applied, for outputs we used a combination of per-channel RF-data, beam-summed RF-data and log-detected line data from the large number of transmits sequence.

II. Methods

A. Data

We demonstrate the image quality reduction problem with diverging waves on a 1-dimensional phased array. A cardiac phased array (S5-1, Philips Healthcare, Andover MA) and sector scan geometry is used for all the experiments.

Our Deep Learning Beamforming (DLB) approach begins with per-channel time-delayed RF-data from 11 Diverging wave transmits -prior to compounding- and learns a mapping to the target data of 51 transmits. All the diverging wave transmits are focused at 50 mm behind the transducer surface and uniformly spaced to span the 90° scan geometry. We simulate three random phantoms using Field II [11] to generate the training data. We initially simulate and generate per-channel receive data for the all 51 transmission angles; data for 11 DW transmit case is obtained by downsampling the fully sampled transmit space. The per-channel data was downsampled to 8 MHz after applying the delays. These data were used for training and testing the neural networks. We also collected in-vivo data from healthy volunteers for training and test purposes. The final training dataset consisted of simulation and in-vivo data. Another dataset corresponding to a single B-mode image is used to fine-tune the network trained on phantom dataset.

The input data is divided into overlapping patches of size [13,80,11] in fast-time, elements and transmits to be compounded. Fast time samples correspond to about four wavelengths. The corresponding outputs are a size [1,80] vector for compounded channel data and scalars for the beamsummed channel data and log-detected envelope data. Because each training sample corresponds to a single pixel of the ultrasound image, a single frame provides around 120,000 samples for training. Input data is normalized by dividing each patch by the patch RMS, such that the post normalization patch values have a unity variance.

B. CNN Architecture

A fully convolutional neural network is implemented in Tensorflow framework. The architecture consists of two layers with 3D convolutions, and remaining four layers with 2D convolutions. Each layer is followed by a leaky rectified linear unit (ReLU) activation function with the exception of last layer, whose activation is linear. Details of the network are given in Table 1. We experimented with number of 3D and 2D convolutional layers keeping the overall number layers constant since there is a trade-off between computational complexity and network expressive power. We started with five layers involving 3D convolutions and reduced to two layers with minimal image quality degradation while achieving significant computational performance gains

TABLE I. CNN Architecture

Layer number	Convolution type	Number of filters	Filter size
1	3D	64	7x3x15
2	3D	32	5x3x15
3	2D	16	3x15
4	2D	8	3x15
5	2D	4	3x15
6	2D	1	3x15

Our architecture consist of multiple outputs in order to be able to define a combined loss function. The loss function include terms from per-channel data, prior to the summation; post channel-sum RF-data; and log-compressed envelope data. The mean-square error loss function (MSE) is used for network training. The compound loss is written as:

$$Loss = a \times MSE(C, \hat{C}) + b \times MSE(S, \hat{S}) + c \times MSE(I, \hat{I}) + d \times \sum_{i=1}^{N} \ell_2(W_i) \quad (1)$$

where C, S and I stand for compounded channel data, beamsummed channel data, and log-detected envelope data, respectively. Here ˆ indicates the predicted values as opposed to true values. *'a, b, c* and *d'* are the hyperparameters used to accommodate the scale difference between different losses and to fine tune the network performance. Finally, we use an l_2-norm penalty across the network weights to regularize the network and eliminate overfitting. Typical values for the hyperparameters are shown in table 2. Although using the individual components of the loss term provide reasonable training performance; in our experiments, using the compound loss not only leads to better performance but also smoother and faster training convergence.

TABLE II. Loss Function Hyperparameters

a	b	c	d
10	10^{-1}	1	10^{-4}

We used the Adam optimizer [12], with an initial learning rate of 1e-3, decaying exponentially every 10 epochs. We further tuned the network parameters based on its performance on the validation data. We target 1000 epochs for training our network. Additionally, performance on validation dataset is used as a stopping criterion for training, the final network is trained for 630 epochs.

III. Results and Discussion

Our initial experiments with simulated phantom and in-vivo cardiac datasets show promising results in reducing clutter and enhancing signal quality. B-mode image reconstructions for a simulated phantom using conventional delay-and-sum (DAS) approach and our DLB approach are presented in Fig 1. The phantom results clearly indicate that the DLB successfully eliminates the sidelobes, which are present in the DAS case. Because sidelobes are apparent in the per-channel RF-data as tilted wavefronts our initial hypothesis of 'exploiting the acoustic artifact structure in per-channel RF-data' is proven to be correct. However, eliminating these acoustic artifacts does not necessarily improve the image quality significantly. In-vivo results are more subtle with evident clutter reduction in cardiac chambers. In both cases DLB with 11 transmits improves the image quality by suppressing the artifacts, which are absent in 51 transmit case. However, additional information that is present in 51 transmit images is not recovered by DLB processing.

The pixel-level training strategy we employ in this study offers a number of advantages. First, it enables us to extract a large amount of data from a single image. Second, it reduces the number of training samples required to avoid overfitting. Third advantage is that it implicitly prohibits the network to learn dataset biases as it would have if we trained the network using an image or patch-based approach.

We also observed that the newly introduced compound loss term guides the network during training, through several local minima, which individual loss terms were not able to move out of and converge to. This is inferred by calculating an image MSE between the ground truth image and images predicted from networks trained with different loss functions. The image predicted by the network with the compound loss function has the lowest image MSE. This is counterintuitive because typically when the network is constrained more, e.g. with a compound loss function, the final MSE is higher than the MSE's of networks trained with a single loss term. However, this shows that the individual components of the compound loss term work in cooperation.

Additionally, the network we use is relatively small compared to others used for similar beamforming tasks (i.e. U-Net). Our motivation for keeping the network light is to develop an understanding of the network learning capability for a specific problem, i.e. learning the acoustic artifacts in the per-channel signal and eliminating them from the resulting image pixel value. There are additional benefits of using a smaller network such as less data necessary for training, shorter training time and shorter inference time. It is worth mentioning here that our network has not 'seen' a single cardiac image, therefore has not learned any structural cardiac information that could be beneficial in forming images of the heart but on the other hand the network has no bias for any anatomy, disease or clinical condition.

IV. Conclusions

In this work, we have presented a deep learning beamformer (DLB) implemented for high-frame rate echocardiography with limited DW transmissions. Our simulation and in-vivo results demonstrate that DLB improves image quality by eliminating the sidelobes that are absent in images with larger number of transmits. Although they can be implemented successfully as presented in this work, the clinical impact and diagnostic validity of neural networks for signal processing tasks such as beamforming still needs to be addressed. Future work needs to address these problems as well as to investigate more elaborate network architectures and loss functions.

References

[1] M. Couade *et al.*, "Ultrafast imaging of the heart using circular wave synthetic imaging with phased arrays," 2009 IEEE International Ultrasonics Symposium, Rome, 2009, pp. 515-518.

[2] G. Montaldo, M. Tanter, J. Bercoff, N. Benech and M. Fink, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 56, no. 3, pp. 489-506, March 2009

[3] A. Lucas, M. Iliadis, R. Molina and A. K. Katsaggelos, "Using Deep Neural Networks for Inverse Problems in Imaging: Beyond Analytical Methods," in IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 20-36, Jan. 2018.

[4] A. Luchies and B. Byram, "Deep neural networks for ultrasound beamforming," 2017 IEEE International Ultrasonics Symposium (IUS), Washington, DC, 2017, pp. 1-1.

[5] M. Gasse, F. Millioz, E. Roux, D. Garcia, H. Liebgott and D. Friboulet, "High-Quality Plane Wave Compounding Using Convolutional Neural Networks," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 64, no. 10, pp. 1637-1639, Oct. 2017.

[6] D. Perdios, M. Vonlanthen, A. Besson, F. Martinez, M. Arditi and J. Thiran, "Deep Convolutional Neural Network for Ultrasound Image Enhancement," 2018 IEEE International Ultrasonics Symposium (IUS), Kobe, 2018, pp. 1-4.

[7] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation," In International Conference on Medical image computing and computer-assisted intervention, Munich, Oct. 2015, (pp. 234-241). Springer, Cham.

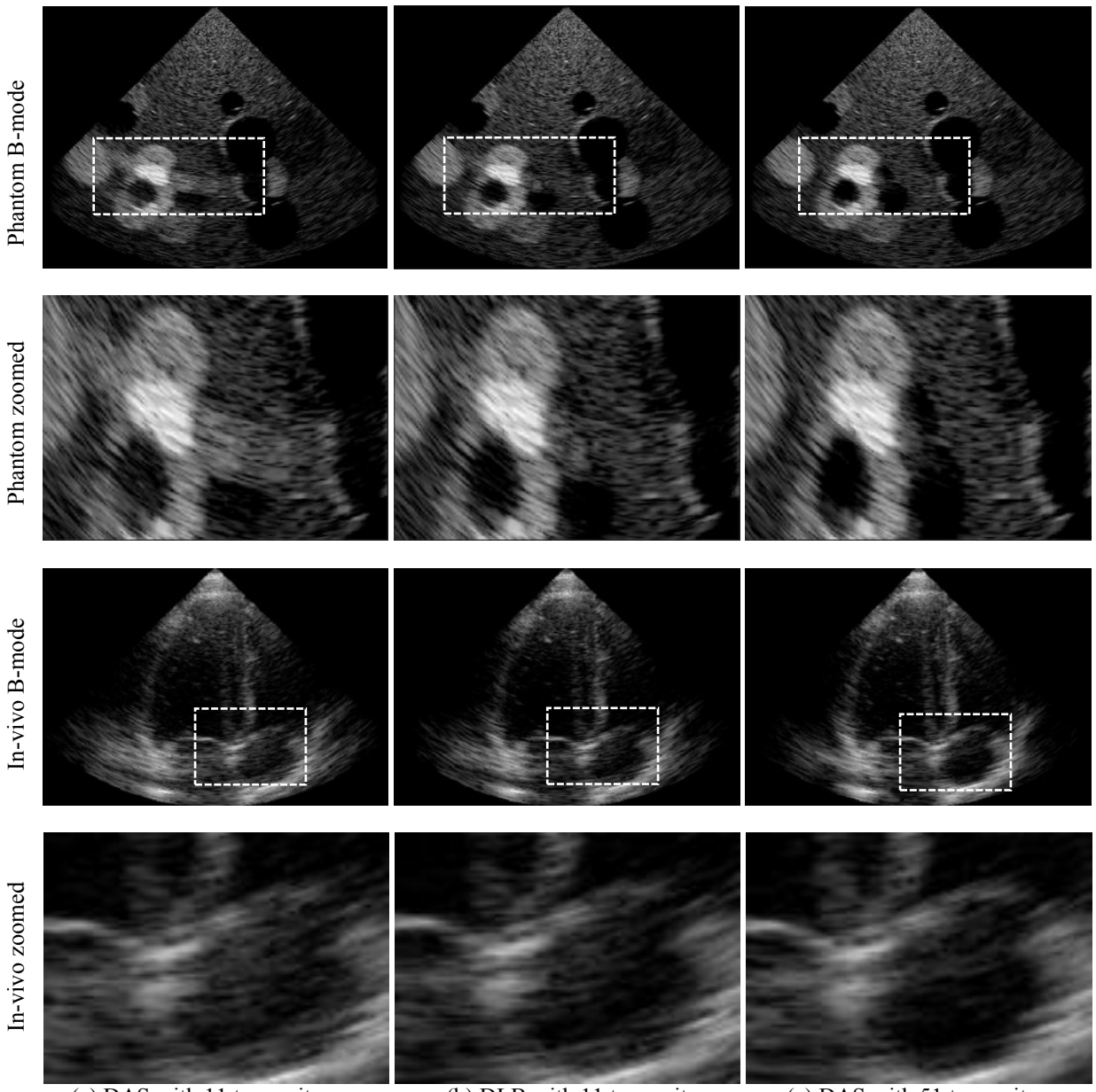
[8] Y. H. Yoon, S. Khan, J. Huh and J. C. Ye, "Efficient B-Mode Ultrasound Image Reconstruction From Sub-Sampled RF Data Using Deep Learning," in IEEE Transactions on Medical Imaging, vol. 38, no. 2, pp. 325-336, Feb. 2019.

[9] S. Vedula *et al.*, "High quality ultrasonic multi-line transmission through deep learning," in: Knoll F., Maier A., Rueckert D. (eds) Machine Learning for Medical Image Reconstruction. MLMIR 2018. Lecture Notes in Computer Science, vol 11074. Springer, Cham

[10] O. Senouf *et al.*, High frame-rate cardiac ultrasound imaging with deep learning. In: Frangi A., Schnabel J., Davatzikos C., Alberola-López C., Fichtinger G. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, vol 11070. Springer, Cham

[11] J.A. Jensen, "Field: A program for simulating ultrasound systems", Paper presented at the 10th Nordic-Baltic Conference on Biomedical Imaging Published in Medical & Biological Engineering & Computing, 1996, pp. 351-353, Volume 34, Supplement 1, Part 1

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



(a) DAS with 11 transmits (b) DLB with 11 transmits (c) DAS with 51 transmits

Figure 1: First two rows presents B-mode US image reconstructions and zoomed patches for a phantom, and bottom two rows present B-mode US image and zoomed patches for an in-vivo dataset. All images are displayed in [0, 60] dB scale. Zoomed patches are presented to accentuate the differences.