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

Fast Diverging Wave Imaging Using Deep-Learning-Based Compounding

Jingfeng Lu*†, Fabien Millioz*, Damien Garcia*, Sebastien Salles*, Denis Friboulet*
* University of Lyon, CREATIS, CNRS UMR 5220, Inserm U1044, INSA-Lyon, University of Lyon 1, Lyon, France
†, METISlab, School of Instrumentation Science and Engineering, Harbin Institute of Technology, Harbin, China

***Abstract*—Diverging wave (DW) ultrasound imaging has become a very promising methodology for ultrafast cardiovascular imaging due to its high temporal resolution. However, if they are limited in number, DW transmits alter image quality compared with classical focused schemes. A conventional reconstruction approach consists in summing successive RF images coherently, at the expense of the frame rate. To deal with this limitation, we propose in this work a convolutional neural network (CNN) architecture for high-quality reconstruction of DW ultrasound images using a small number of transmissions. We experimentally demonstrate that the proposed method produces high-quality images using only three DWs, yielding an image quality equivalent to the one obtained with standard compounding of 31 DWs in terms of contrast and resolution.**

I. Introduction

Diverging wave imaging has drawn much interest in the research community due to the high temporal resolution. Compared with the conventional focused scheme where several narrow sectors of the entire image are reconstructed with successive transmissions, DW imaging can image the whole field of view with single diverging wave transmission [1], [2]. Therefore, DW imaging achieves a high frame-rate, facilitating capturing transient biological phenomena, e.g., cardiac motion [3]. However, images obtained from single DW emission suffer a poor image quality due to the unfocused beam. A standard reconstruction method performs coherent compounding [2] of multiple RF images from successive transmissions, at the expense of the frame rate. A trade-off needs to be made between image quality and frame rate since compounding of more beams produces images of higher quality but increases the acquisition time. Therefore, to achieve a high image quality while maintaining the frame rate of DW imaging is of great research interest.

Recently, deep learning approaches have achieved significant performance in various image reconstruction problems. Inspired by the tremendous success of deep learning, our group previously proposed to adopt convolutional neural works to reconstruct high-quality images using 3 plane wave (PW) transmissions [4]. Different from PW imaging, the image properties of DW imaging varies along image depth. We have indeed experimentally observed that directly applying a conventional CNN architecture to DW image reconstruction cannot produce a satisfactory performance. To deal with the spatially varying properties of DW images along depth, we adopted in this work an inception module [5] composed of the concatenation of multi-scale convolutional kernels. Incorporating the inception module aims at capturing different image features with multi-scale receptive fields.

II. Methods

Let x be the low-quality RF beamformed images of size $m \times w \times h$, where m is the number of DW acquisitions, w is the number of scan lines, and h is the length of each RF signal. Our work aims at reconstructing the high-quality image y of dimension $w \times h$ from input x. Standard compounding consists in summing all m DWs to obtain the high-quality image. Considering that there may be useful information which is not exploited by standard compounding, we employ a CNN with trainable parameters θ to learn the optimal mapping $f(\cdot)$ of $x \rightarrow y$.

The proposed network is a 2-D fully convolutional network with five hidden layers, whose architecture is described in Table I. Two types of basic building modules, including regular convolutional and inception module, are used to construct hidden layers. The inception module employed in our network is composed of four parallel paths. Each path performs convolution with different kernel size to the same input. The outputs of each path are stacked along the channel dimension as the output of the inception module. As the image properties of DW imaging vary along image depth, the multi-size convolution kernels contribute to extract different image features from multiple receptive field sizes. The activation function used in our network is a four-pieces maxout unit [6]. Maxout units are piecewise-linear convex functions, and a maxout network with more than two maxout units can approximate many general activation functions [7]. We exclude the pooling operation used in most popular CNN architectures to produce the feature maps with the same dimension. This guarantees that the spatial information is preserved at the same scale throughout the network, which may be beneficial for maintaining phase in RF signals.

III. Experiment

A. Data Set Acquisition

A Verasonics Vantage 256 equipped with ALT P4-2 probe was used to perform the data set acquisition. The samples were acquired by continuously moving the probe on the surface of the imaged objects, at an imaging rate of 50 frames/s and a packet size of 250 images. Each sample was acquired using 31 DWs with angles ranging between $\pm 30^\circ$, with an incremental step of 2°. For each DW transmission, the received

TABLE I
NETWORK ARCHITECTURE

block type	feature size channel × height × width	kernel size height × width	padding height × width	number of kernels	activation
inputs	$m \times h \times w$	-	-	-	-
convolution	$64 \times h \times w$	9×3	4×1	256	maxout 4
convolution	$32 \times h \times w$	17×5	8×2	128	maxout 4
convolution	$16 \times h \times w$	33×9	16×4	64	maxout 4
inception	$8 \times h \times w$	41×11	20×5	8	maxout 4
		49×13	24×6	8	maxout 4
		57×15	28×7	8	maxout 4
		65×17	32×8	8	maxout 4
convolution	$1 \times h \times w$	1×1	-	4	maxout 4

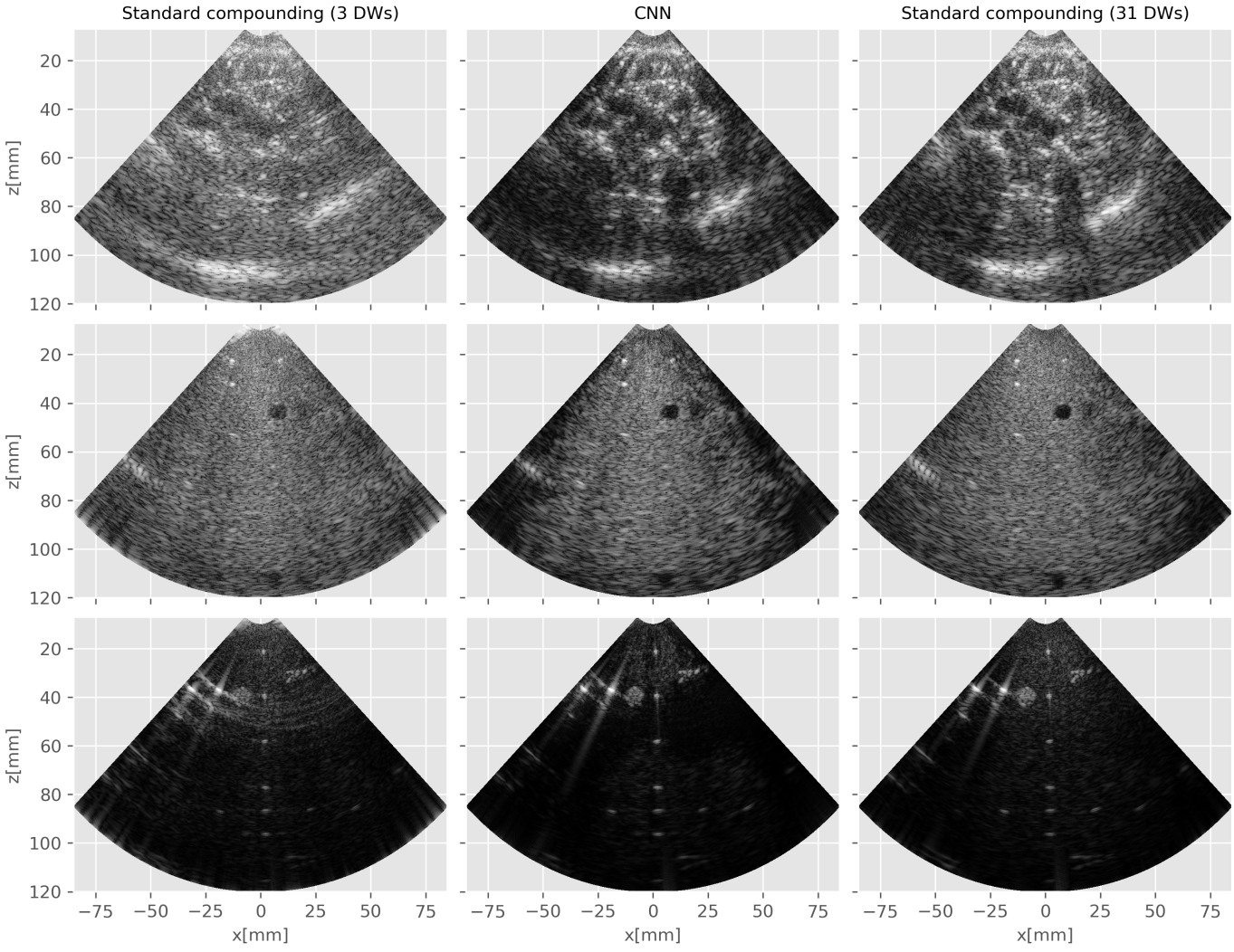

Fig. 1. B-mode images obtained using our network (middle column) and standard compounding of 3 DWs (left column) and 31 DWs (right column). Top to bottom: in-vivo tissues from the quadriceps femoris muscle; in-vitro tissues from the Gammex phantom; in-vitro tissues from the CIRS phantom.

RF signals were sampled at 11.9 MHz and beamformed with the conventional delay and sum method. Each RF beamformed image is of dimension 512×256, covering a polar region of size 12cm$\times 90^\circ$. The input images x were composed of a small subset of m = 3 DWs (-30°, 0°, and 30°), while the reference images y were the standard compounding of all n = 31 DWs. A total of 7000 samples were used in the experiment, and each was made up of low-quality images x and the

high-quality corresponding y. Specifically, 1500 samples were acquired from in-vivo tissues (quadriceps femoris muscle, phalanges of fingers, and liver region), and 5500 acquisitions were performed on in-vitro phantoms. From the 7000 samples of the entire set, 6000 samples were randomly selected as the training set. The remaining 1000 samples were used as the testing set for evaluation.

B. Training implementation

From the 6000 samples of the training set, 5000 samples were used for training the network, and the remaining 1000 samples were used as an independent validation set. In the training phase, the network weights were initialized with the Xavier initializer [8]. The Mean Squared Error (MSE) loss was minimized using mini-batch gradient descent with the Adam optimizer [9], and the batch size was set as 10. The initial learning rate was set as 1×10^{-4} and an early stopping strategy was used to adjust the learning rate. The learning rate was halved if there had been no decrease in the loss for 20 epochs, and 40 epochs without loss reduction would end the training. The training was performed using Pytorch [10] library on a NVIDIA Tesla V100 GPU with 32 Gb of memory, resulting in training times of about two days.

IV. Results

In the testing phase, the images reconstructed by our network were compared to those obtained from the compounding of 3 DWs and 31 DWs. As depicted in Fig. 1, the images reconstructed by our network using only three DWs (middle column) are visually very close to the reference (right column). The contrast and resolution are much improved compared with the images from the standard compounding of the same three DWs (left column). We use two evaluation indices for the quantitative evaluation of image quality: contrast ratio (CR) and the lateral resolution (LR), measured on test samples acquired from the Gammex phantom and CIRS phantom, respectively. As shown in Fig. 2 and Fig. 3, the CR and LR reached by our network is compared to those obtained from the standard compounding of an increasing number of DWs. The CR and LR of standard compounding (blue curves) tend to stabilize to an optimal value using more DWs. While using only three DWs, our network (orange lines) produced a CR equivalent to that of the standard compounding of about 23 DWs and an LR equivalent or even superior to the standard compounding of 31 DWs.

V. Conclusion

A deep-learning-based method is presented in this work for the reconstruction of diverging wave imaging. The proposed approach aims at learning a compounding operator to reconstruct high-quality images using a small number of DWs. The experiments are performed using a large number in vitro and in vivo samples. The experimental results demonstrated the effectiveness of the proposed method, producing an image quality equivalent to the one obtained with standard compounding of 31DWs in terms of contrast and resolution.

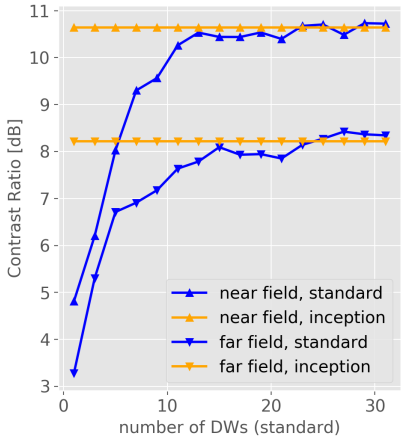

Fig. 2. CR reached by our network with three DWs ($-30°$, $0°$, and $30°$), compared to the standard compounding of an increasing number of DWs. CR is measured on two anechoic regions (the near field at 40mm depth and the far field at 120mm depth) of a B-mode image obtained from the Gammex phantom.

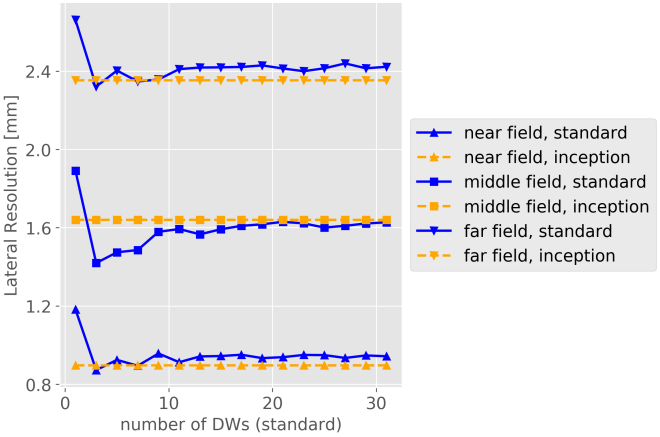

Fig. 3. LR reached by our network with three DWs ($-30°$, $0°$, and $30°$), compared to the standard compounding of an increasing number of DWs. LR is measured on 0.1 mm Nylon monofilaments (the near field at 20mm and 40 mm depth, the middle field at 60mm and 80mm depth, and the far field at 90mm and 100mm depth) of a B-mode image obtained from the CIRS phantom.

References

[1] H. Hasegawa and H. Kanai, "High-frame-rate echocardiography using diverging transmit beams and parallel receive beamforming," *Journal of medical ultrasonics*, vol. 38, no. 3, pp. 129–140, 2011.

[2] J. Porée, D. Posada, A. Hodzic, F. Tournoux, G. Cloutier, and D. Garcia, "High-frame-rate echocardiography using coherent compounding with doppler-based motion-compensation," *IEEE transactions on medical imaging*, vol. 35, no. 7, pp. 1647–1657, 2016.

[3] J. Faurie, M. Baudet, J. Porée, G. Cloutier, F. Tournoux, and D. Garcia, "Coupling myocardium and vortex dynamics in diverging-wave echocardiography," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 66, no. 3, pp. 425–432, 2018.

[4] M. Gasse, F. Millioz, E. Roux, D. Garcia, H. Liebgott, and D. Friboulet, "High-quality plane wave compounding using convolutional neural networks," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 64, no. 10, pp. 1637–1639, 2017.

[5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 1–9.

[6] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, "Maxout networks," *Computer Science*, pp. 1319–1327, 2013.

[7] H. Zhao, F. Liu, and L. Li, "Improving deep convolutional neural networks with mixed maxout units." *PloS one*, vol. 12, no. 7, 2017.

[8] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249–256.

[9] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *Advances in Neural Information Processing Systems Workshop*, 2017.

[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceeding of International Conference on Learning Represent*, 2015, pp. 1–41.