Ultrasound Beamforming using MobileNetV2

Sobhan Goudarzi
Electrical and Computer Engineering Department
Concordia University
Montreal, QC, Canada
s_oudarz@ece.concordia.ca

Amir Asif
Electrical Engineering and Computer Science
York University
Ontario, Canada
asif@eecs.yorku.ca

Hassan Rivaz
Electrical and Computer Engineering Department
Concordia University
Montreal, QC, Canada
hrivaz@ece.concordia.ca

***Abstract*—In the past few years, deep learning has disrupted several low-level medical imaging tasks such as reconstruction of Computed Tomography (CT) and Magnetic Resonance (MR) images. In this work, we propose a novel deep learning-based approach for the low-level task of ultrasound image reconstruction from the pre-beamformed channel data. More specifically, we adapt MobileNetV2 to train a model that mimics Minimum Variance Beamforming (MVB). Results confirm that the proposed method takes much less time to reconstruct images with a quality similar to one achieved by applying MVB directly. The current paper is a part of our submission to Challenge on Ultrasound Beamforming with Deep Learning (CUBDL) announced by 2020 IEEE International Ultrasonics Symposium (IUS).**

***Index Terms*—Adaptive beamforming, ultrasound imaging, minimum variance, deep learning.**

I. Introduction

Beamforming is an essential step in the ultrasound image formation pipeline, which can be applied in both transmission and receive steps. In receive beamforming, the main goal is to extract the highest quality of spatial map of the target echogenicity. Adaptive methods estimate the apodization weights from echo traces acquired by transducer elements and have a better performance compared to non-adaptive methods wherein a predefined set of weights are used.

Minimum Variance Beamforming (MVB) is one of the best adaptive methods that performs well regardless of the imaging settings [1]. MVB, however, is computationally very expensive mainly because of the covariance matrix estimation step. Therefore, speeding up MVB is of crucial importance to make it applicable online [2]–[4].

Recently, deep learning (DL) has been proposed for ultrasound image reconstruction [5]–[9]. There are a variety of approaches to accomplish this task. More specifically, DL can be designed to complete a single, few, or all of the reconstruction steps. Another advantage is that DL can simultaneously fulfill another objective such as speckle reduction or super-resolution in the reconstruction process. Nevertheless, the design of ultrasound image reconstruction using deep learning entails several challenges. The scarcity of training data as well as lack of ground truth are among the main limitations. Moreover, the changes in imaging settings cause a large domain shift in the high dimensional input space of DL, limiting its generalization.

Funded by Natural Sciences and Engineering Research Council of Canada (NSERC) grants RGPIN-2020-04612 and RGPIN-2017-06629.

Herein, we strive to address all of the aforementioned challenges. In essence, the proposed approach is designed to approximate MVB. As MVB can be summarized in a set of closed-form mathematical steps, we can calculate the desired output for any input. As such, we do not have the problem of domain shift or lack of ground truth. Furthermore, the proposed method does not need many training images since MVB works on each pixel separately meaning that each pixel is a sample in the training process.

We also consider the fact that all mathematical transformations, including DL, cannot generate new information that is not present in the input data. Therefore, necessary preprocessing steps are applied to raw Radio-Frequency (RF) channel data before feeding it to the network, and the network input contains all required information for estimating the result of MVB. More specifically, first, IQ demodulation is applied to the RF channel data since MVB requires complex signals to compute complex weights allowing for beampatterns that are asymmetrical around the center of the beam. Second, time delays are compensated to reduce the load on the network. Finally, the F-number is fixed for all image depths in order to make the image quality uniform.

MobileNetV2 is used as the network structure since it is a leading architecture for networks with low computational complexity and memory requirement. This is of critical importance for commercial success of deep learning beamforming given the very large ultrasound frame-rate and limited computational resources, especially in mobile ultrasound devices.

The proposed method is trained on a set of public datasets available in the ultrasound toolbox [10]. The proposed approach has been accepted for presentation during the Challenge on Ultrasound Beamforming with Deep Learning (CUBDL) at the 2020 IEEE International Ultrasonics Symposium (IUS) [11], [12]. The results presented here only correspond to the training step because the test data is not released to the participants while the challenge is ongoing.

II. Method

Consider an ultrasound array that transmits a pulse into the domain with a sound speed of c. Regardless of the transmission technique, let us assume n elements record the

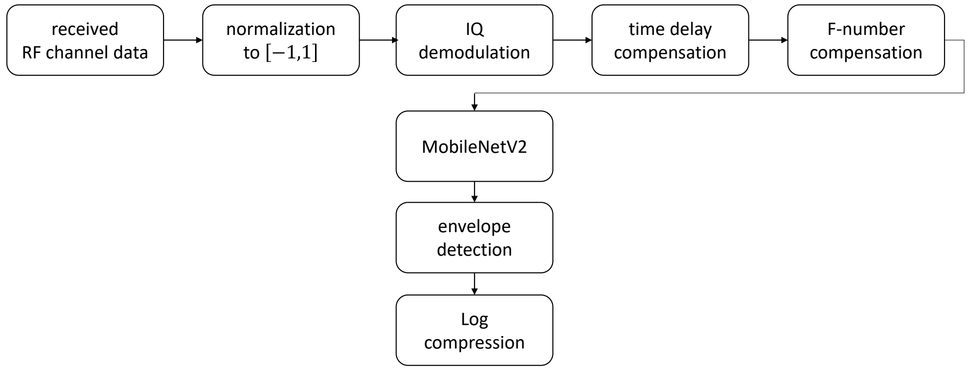

Fig. 1. Diagram of the proposed method.

backscattered signals, denoted by $\mathbf{h}_i(t)$. d_t is defined as the transmission distance from the origin of the transmitted pulse to an arbitrary point (z, x) in the region-of-interest (ROI), and d_r is the receiving distance from (z, x) to the location of element i. The RF data corresponding to (z, x) in $\mathbf{h}_i(t)$ can be found by applying the propagation delay as follows (hereafter, capital and bold font variables represent matrices and vectors, respectively):

$$\tau = \frac{d_t + d_r}{c} \Longrightarrow r_i = \mathbf{h}_i(t) \mid_{t=\tau}, \tag{1}$$

where r_i is the RF data of channel i, which corresponds to the point (z, x) in the ROI. S is defined as the resulting image of ROI and each pixel (z, x) of S can be obtained through a weighted summation of RF data corresponding to receiving elements as follows:

$$s = \sum_{i=0}^{n-1} \mathbf{w}(i) r_i, \tag{2}$$

where $\mathbf{w}$ is the apodization window of length n. Eq. 2 can be vectorized in the following form:

$$s = \mathbf{w}^H \mathbf{r}, \tag{3}$$

The goal of receive beamforming is to estimate the apodization window in order to reconstruct a high-quality ultrasound image which is a spatial map of the target echogenicity.

A. Minimum variance beamforming

In Capon's MVB, data dependent apodization weights $\mathbf{w}$ are estimated while a unity gain is maintained in the steering direction [1]. The corresponding minimization problem is as follows:

$$\min_{\mathbf{w}} E\{\mid s \mid^2\} = \mathbf{w}^H R\, \mathbf{w} \quad s.t. \quad \mathbf{w}^H \mathbf{a} = 1 \tag{4}$$

where $R = E\{\mathbf{r}\mathbf{r}^H\}$ is the spatial covariance matrix and E refers to the expectation operation. For delayed signals, the steering vector $\mathbf{a} = 1$. Eq. 4 can be solved using the method of Lagrange multipliers, and the estimated apodization vector is as follows [1]:

$$\mathbf{w}_{MV} = \frac{R^{-1}\mathbf{a}}{\mathbf{a}^H R^{-1} \mathbf{a}} \tag{5}$$

The estimation of R is made robust with temporal averaging over $2k+1$ samples and averaging over subarrays of length l as follows [1]:

$$\widetilde{R}(z, x) = \frac{\sum_{j=-k}^{k} \sum_{i=0}^{n-l} \overline{\mathbf{r}}_i^H (z-j, x)}{(2k+1)(n-l+1)} \tag{6}$$

where:

$$\overline{\mathbf{r}}_i(z, x) = \left[r_i(z, x), r_{i+1}(z, x), ..., r_{i+l-1}(z, x)\right]^T \tag{7}$$

A diagonal loading factor is added to the covariance matrix for numerical stability by $\widehat{R}(z, x) = \widetilde{R}(z, x) + \epsilon I$, where I is the identity matrix and:

$$\epsilon = \frac{\Delta}{l} trace(\widetilde{R}(z, x)) \tag{8}$$

The result of subarray averaging is a vector of length l. Finally, each point (z, x) of S using MVB can be computed as follows:

$$s_{MV} = \frac{1}{n-l+1} \sum_{i=0}^{n-l} \mathbf{w}_{MV}^H \overline{\mathbf{r}}_i, \tag{9}$$

B. The proposed method

The proposed receive beamforming approach can be summarized in a few steps presented in Fig. 1. It has to be mentioned that the input to the network is supposed to be within the $\left[-1, 1\right]$ range otherwise RF channel data has to be scaled proportionally.

Each pixel of the image is reconstructed separately as is the case for MVB. The network's input is a $2 \times m \times n$ matrix in which first the two channels are the real and imaginary parts of IQ data, n is the number of channels and m is the length of the window considered for temporal averaging to preserve the speckle statistics. The network output is a two-dimensional vector containing real and imaginary parts of the beamformed data. The network is designed to estimate the apodization window and apply Eq. 9 on the input IQ channel data. After reconstructing the whole output IQ data, it is subjected to envelope detection and log compression in order to obtain the final B-mode ultrasound image.

As mentioned before, MobileNetV2 [13] is used as the network structure. The MobileNetV2 architecture is based

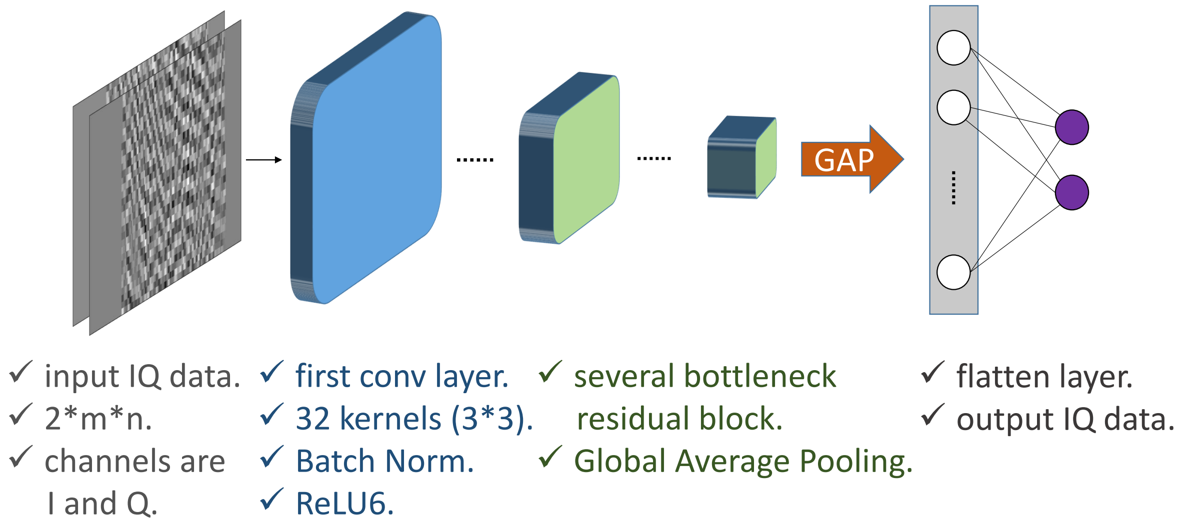

Fig. 2. Overview of the MobileNetv2 architecture. GAP refers to Global Average Pooling.

on using depth-wise separable convolution building blocks. Moreover, it contains linear bottlenecks between the layers as well as shortcut connections between the bottlenecks. An overview of the MobileNetv2 architecture is shown in Fig. 2. More details regarding MobileNetV2 can be found in [13].

C. Training

As the trained network has to be able to generalize across the range of parameters provided in the CUBDL Data Guide, the network is trained with a variety of imaging settings such as the acquisition center frequency, sampling frequency, transducer shape, and number of transducer elements. More specifically, the training data contains one image collected with a phased array probe and 10 images collected with linear array probes. Among the second group, 2 images are from focused imaging dataset and 8 are from coherent plane-wave compounding (CPWC) dataset. 2 of CPWC data are collected with an Alpinion scanner (Seoul, South Korea) using a L3-8 probe and the other 6 are collected with Verasonics Vantage 256 platform (Kirkland, WA, USA) and the linear L11-4v probe. All these datasets are publicly available through UltraSound ToolBox (USTB) [10]. The network's output for each image is reconstructed using the MVB code provided by USTB. The number of input channels (n) is different based on the probe specifications, while the length of window for temporal averaging (m) is set to 32 for all datasets.

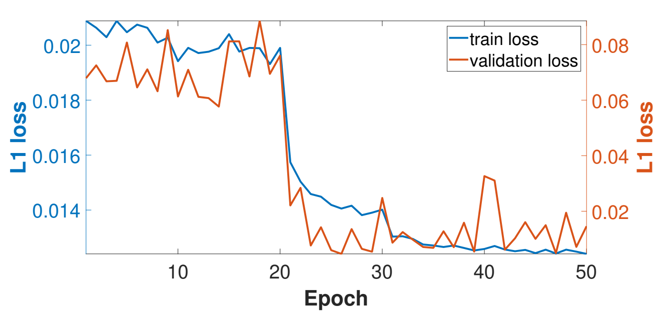

Fig. 3. Plot of the training and validation losses during training.

The model is implemented using PyTorch library. The batch size is 50, and AdamW optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ is used. The learning rate is linearly decayed from 10^{-3} to 5×10^{-5} during 50 epochs. Fig. 3 illustrates the training history of the network. As the network works on each pixel of the image separately, the total size of the input-output pairs is more than 1 million. 80 percent of the data is used for training and 20 percent for validation. Fig. 3 confirms that the weights are not overfitted to the training dataset.

III. Results and Discussion

In this section, the results on training and validation datasets are presented. As mentioned before, we do not have any test data because they are not released to the participants while the challenge is ongoing. Fig. 4 shows the results on CPWC datasets containing Simulation Resolution (SR), Experimental Resolution (ER), Simulation Contrast (SC), and Experimental Contrast (EC) images. Fig. 4 confirms that the proposed approach provides images of a better quality than Delay And Sum (DAS) beamformer and similar to MVB. This comes with a large gain in speed: MVB takes 4.05 min for the reconstruction of EC image while the proposed method takes 0.67 min. Although it is still far from real-time performance, the current paper can be considered as a proof of concept that DL can be used to speed-up the classical approaches. Moreover, our proper design confirms that deep learning can be used with a wide variety of imaging settings which is one of the main problems limiting the medical practical applications of deep learning.

IV. Conclusion

In this paper, a deep learning framework for ultrasound beamforming has been presented. The proposed approach is

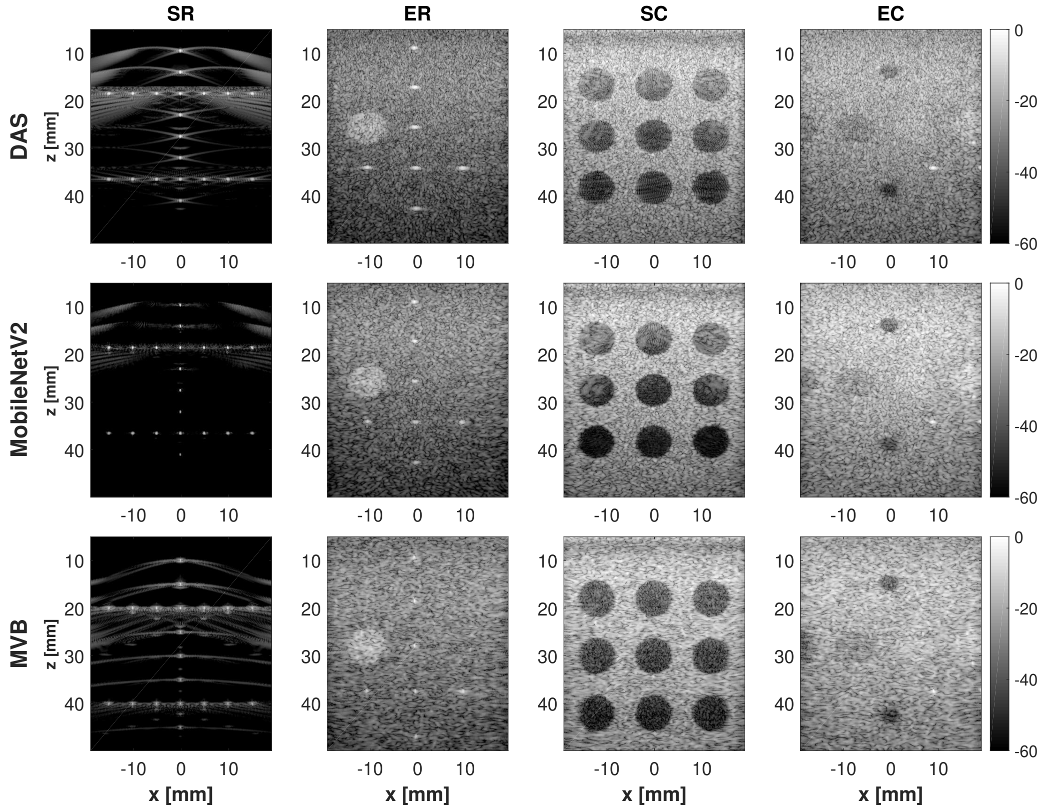

Fig. 4. Beamforming results on the single 0° plane wave. Columns indicate different image data sets while rows correspond to beamforming methods.

based on MobileNetV2 structure and works on IQ channel data to reconstruct each pixel of the final image separately. The training results confirm that deep learning can be used as a general beamformer working on a wide variety of imaging settings.

V. Acknowledgment

The authors would like to thank the organizers of CUBDL for providing the chance of scientific competition as well as the organizers of ultrasound toolbox for providing publicly available codes and data.

References

[1] J. Synnevag, A. Austeng, and S. Holm, "Benefits of minimum-variance beamforming in medical ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 9, pp. 1868–1879, 2009.

[2] J. Chen, J. Chen, R. Zhuang, and H. Min, "Multi-operator minimum variance adaptive beamforming algorithms accelerated with gpu," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2020.

[3] Avinash S. Vaidya and M.B. Srinivas, "A low-complexity and robust minimum variance beamformer for ultrasound imaging systems using beamspace dominant mode rejection," *Ultrasonics*, vol. 101, pp. 105979, 2020.

[4] S. Afrakhteh and H. Behnam, "Low-complexity adaptive minimum variance ultrasound beam-former based on diagonalization," *Biomedical Signal Processing and Control*, vol. 62, pp. 102110, 2020.

[5] S. Goudarzi, A. Asif, and H. Rivaz, "Multi-focus ultrasound imaging using generative adversarial networks," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 1118–1121.

[6] S. Khan, J. Huh, and J. C. Ye, "Adaptive and compressive beamforming using deep learning for medical ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 8, pp. 1558–1572, 2020.

[7] B. Luijten, R. Cohen, F. J. De Bruijn, H. A. W. Schmeitz, M. Mischi, Y. C. Eldar, and R. J. G. Van Sloun, "Adaptive ultrasound beamforming using deep learning," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2020.

[8] A. C. Luchies and B. C. Byram, "Assessing the robustness of frequency domain ultrasound beamforming using deep neural networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, pp. 1–1, 2020.

[9] S. Goudarzi, A. Asif, and H. Rivaz, "Fast multi-focus ultrasound image recovery using generative adversarial networks," *IEEE Transactions on Computational Imaging*, pp. 1–1, 2020.

[10] A. Rodriguez-Molares, O. M. H. Rindal, O. Bernard, A. Nair, M. A. Lediju Bell, H. Liebgott, A. Austeng, and L. Lovstakken, "The ultrasound toolbox," in *2017 IEEE International Ultrasonics Symposium (IUS)*, 2017, pp. 1–4.

[11] M. A. L. Bell, J. Huang, D. Hyun, Y. C. Eldar, R. J. G. van Sloun, and M. Mischi, "Challenge on ultrasound beamforming with deep learning (cubdl)," in *2020 IEEE International Ultrasonics Symposium (IUS)*, 2020, pp. 1–1.

[12] "Challenge on ultrasound beamforming with deep learning (cubdl)," IEEE Dataport, 2019. [Online]. Available:http://dx.doi.org/10.21227/f0hn-8f92.

[13] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.