DEEP LEARNING FOR FAST ADAPTIVE BEAMFORMING

Ben Luijten[⋆]*, Regev Cohen*[†]*, Frederik J. de Bruijn*[‡]*, Harold A.W. Schmeitz*[‡]*,*
Massimo Mischi[⋆]*, Yonina C. Eldar*[†] *and Ruud J.G. van Sloun*[⋆]

[⋆]1: Dept. of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands.
[†]2: Dept. of Electrical Engineering, Technion Israel Institute of Technology, Haifa, Israel.
[‡]3: Philips Research, Eindhoven

ABSTRACT

The real-time nature that makes diagnostic ultrasonography so appealing to clinicians imposes strong constraints on the computational complexity of image reconstruction algorithms. As such, these typically rely on traditional delay-and-sum beamforming, a low-complexity approach that unfortunately comes at the cost of reduced image quality as compared to more advanced and content-adaptive beamformers. Here, we propose a model-aware deep learning strategy to ultrasound image reconstruction, which leverages knowledge of minimum variance beamforming while exploiting the efficiency of deep neural networks. Our approach yields high quality images with strong contrast at real-time reconstruction rates. The neural network is trained using *in vivo* and simulated radio frequency channel data of a single plane wave transmit, and corresponding high-quality minimum-variance beamformed reconstructions. Performance is benchmarked using simulated acquisitions from the PICMUS [1] dataset, demonstrating the convincing generalizability and image quality of the proposed beamformer.

Index Terms— Deep Learning, Ultrasound, Adaptive Beamforming, Plane Wave Imaging

1. INTRODUCTION

Conventional ultrasound image reconstruction often relies on delay-and-sum (DAS) beamforming because of its low computational complexity. While its reconstruction speed facilitates real-time imaging, DAS beamforming unfortunately provides lower contrast and resolution as compared to more advanced methods, mostly due to lack of content-adaptive array apodization. To compensate for this relatively poor performance, an image is usually obtained by compounding multiple acquisitions or using focused scan lines, albeit at the expense of temporal resolution.

Adaptive beamforming strategies improve on this by determining optimal apodization weights based on the input signal. A well known adaptive beamformer is the minimum variance (MV) beamformer, in which the apodization weights are continuously optimized to minimize the variance of the received signals after apodization, while maintaining unity gain in the desired direction. This process effectively suppresses the power of interfering signals from undesired directions that typically lead to cluttered images. Although MV beamforming has indeed shown to significantly improve resolution and contrast compared to DAS [2], it is also notoriously slow, relying on the computationally demanding inversion of an $n \times n$ spatial covariance matrix, having a complexity of n^3 [3]. This poses significant challenges on its real-time implementation.

Recently, developments in deep learning have spurred revolutionary breakthroughs in domains ranging from computer vision to natural language processing. Deep neural networks (DNNs) are trained to perform advanced tasks based on large amounts of data. DNNs consist of many layers of interconnected artificial neurons, which on their own perform only simple operations, but when combined they are universal function approximators [4]. Once trained, inference is typically fast, especially on GPU accelerated systems. DNNs have proven to be very successful in image classification [5], segmentation [6], speech recognition [7] and language translation [8].

Naturally, these techniques are also receiving significant attention in medical imaging. Although the focus has mainly been on solving image analysis tasks such as classification and segmentation [9], more recent developments exploit deep learning for the image reconstruction process itself, finding application in X-ray CT, MRI, PET and photoacoustic tomography [10]. In ultrasound, DNNs have been used for interpolating missing radio frequency (RF) data [11], reducing off-axis scattering [12], compounding of plane waves [13] and super-resolution microscopy [14]. However, these methods still rely on traditional techniques that either limit image quality, or acquisition and reconstruction time. In [15], the beamforming step was circumvented altogether, achieving direct segmentation of cyst phantoms from simulated RF channel data. For most clinical applications however, a B-mode image remains desired, though at a higher resolution and contrast than that achieved with conventional DAS beamforming.

In this paper, we propose a new beamforming strategy that

Table 1: Transducer Parameters

Parameter	Value
Elements	128
Pitch	0.30 mm
Aperture	38.4 mm
Transmit Frequency	6.25 MHz
Sampling Frequency	25 MHz

exploits deep learning to reconstruct high-quality B-mode images from RF channel data, while significantly reducing the computational time and complexity as compared to adaptive beamformers. Instead of treating the DNN as a black box, we apply a model-based approach by basing its architecture on the MV beamforming. We can divide the full adaptive image reconstruction process into three steps: 1) time-of-flight correction, 2) adaptation and application of the apodization weights, and 3) envelope detection. Here we focus on the computationally most expensive of the three: calculation of the adaptive apodization weigths.

2. TRAINING DATA

2.1. Data Acquisition

In order to reconstruct high quality images from RF data, DNNs rely on large amounts of training data in order to converge to a correct and generalized mapping from input to target. Using the Vantage system (Verasonics Inc., WA, USA) in combination with the L11-4v linear transducer, of which the parameters are shown in Table 1, 1000 *in vivo* scans of the carotid artery and wrist were acquired where each plane wave acquisition contains 128 RF lines comprising 2048 axial samples. Additionally, 100 simulated scans of randomly placed point scatterers were generated, aiming to improve resolution by providing more sparse training data.

2.2. Adaptive Beamforming

Target images are created by beamforming the recorded datasets. Applying pixel-wise time-of-flight correction to the received RF signal, relative to the transducer geometry, results in a data array $P(x,z)$ of size N, where N corresponds to the number of receiving elements. A dynamically expanding receive aperture with focus number $f\# = 1.0$ and angular apodization were applied to suppress signals that are outside the receptive area [16]. A coherently compounded DAS image can be formed by summing the array element signals, resulting in a single value for each pixel. Instead, tTo yield high-image-quality training targets, we adopt the Eigen Based Minimum Variance Beamformer [17], providing content adaptive apodization weights $\mathbf{w}$. An adaptively beamformed pixel can be constructed by multiplying the contributions of the receiving channels and summing the results, which can be described as:

$$\mathbf{P}_{\text{BF}}(x,y) = \sum_{n=0}^{N-1} \mathbf{w_n}(x,y)\mathbf{P_n}(x,y), \tag{1}$$

where $\mathbf{P_{BF}(x,y)}$ represents the beamformed pixel intensity. To find an optimal set of weights, the minimization problem

$$\begin{aligned} &\min_{w} \mathbf{w^H R w} \\ &s.t. \quad \mathbf{w^H a = 1} \end{aligned} \tag{2}$$

is solved, where $\mathbf{R}$ denotes the covariance matrix calculated over the receiving array elements and $\mathbf{a}$ a steering vector. Since the data is already time-of-flight corrected by applying delays, $\mathbf{a}$ is a vector of ones with length N. Solving (2) yields the following closed form solution:

$$\mathbf{w} = \frac{\mathbf{R^{-1}a}}{\mathbf{a^H R^{-1} a}}. \tag{3}$$

To circumvent the potentially unstable numerical inversion of $\mathbf{R}$, primarily caused by the correlated nature of the RF input signals, the covariance matrix is estimated by applying spatial smoothing using a subaperture $\mathbf{y_l}$ with length L:

$$\hat{\mathbf{R}}(x,z) = \frac{\sum_{l=0}^{N-L} \mathbf{y_l}[x,z]\mathbf{y_l}^H[x,z]}{N-L+1}. \tag{4}$$

Finally, diagonal loading is applied to further improve stability. Assuming that the received signal is a combination of the desired signal and noise, the Eigen-Decomposition of $\mathbf{R}$ is taken and the signal subspace $\mathbf{E_{signal}}$, composed of the dominant eigenvectors, is projected on the weight vector to obtain the final weights:

$$\mathbf{w}_{\text{EBMV}} = \mathbf{E}_{\text{signal}}\mathbf{E}_{\text{signal}}^{\mathbf{H}}\mathbf{w}. \tag{5}$$

For visualization, the beamformed signals are then envelope detected and subsequently logarithmically compressed.

3. NEURAL NETWORK

3.1. Network Architecture

Based on the the minimum variance beamformer, we train a neural network to adaptively calculate apodization weights corresponding to the input, thereby replacing the traditional, computationally expensive, adaptive processor. We reshape the time aligned RF data as visualized in Fig. 1, and process every pixel independently. Similar to (2) the output weights produced by the network are constrained by implementing a loss function $\mathcal{L}_{\text{unity}}$, penalizing deviations from unity gain. Finally, multiplying $\mathbf{w}$ with the RF input, and subsequently summing the result provides a beamformed pixel.

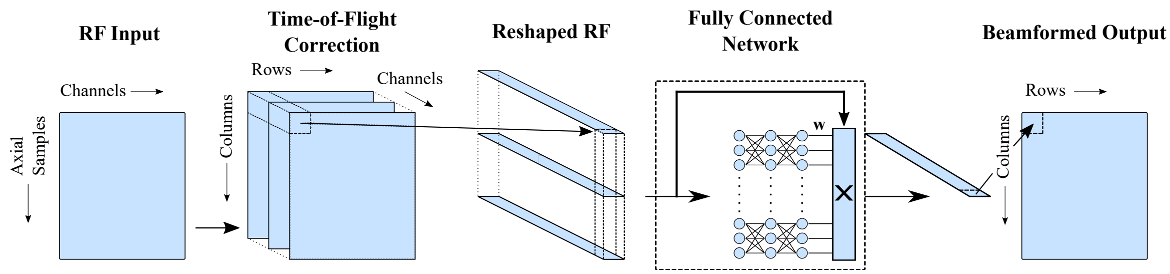

Fig. 1: Schematic overview of the beamforming process using a neural network.

The proposed network consists of 4 fully connected (FC) layers comprising 128 nodes for the outer layers and 32 for inner layers, as indicated in figure Fig. 2. This dimensionality reduction forces the network to find a more compact representation of the data which helps in noise suppression. Between every FC layer, dropout is applied with a probability of 0.2. The network is implemented in Python using the Keras API with a Tensorflow (Google, CA, USA) backend. For training the Adam optimizer was used with a learning rate of 0.001, stochastically optimizing across a batch of pixels belonging to a single image.

3.2. Antirectifier Activation

As of today, the rectified linear unit (ReLU) is the most commonly used activation in DNNs because of its computational efficiency, ability to provide sparse representations and largely avoiding vanishing gradients due to its positive unbounded output [18]. Such a non-linearity may however not be appropriate when dealing with RF input data, as it inherently leads to many 'dying' nodes, impairing the training process. In contrast, a hyperbolic tangent activation is able to preserve negative values. It is however bounded between -1 and 1, and therefore tends to saturate quickly for signals with a large dynamic range, resulting in a vanishing gradient during back propagation. This behavior becomes especially problematic in DNNs because of the substantial amount of consecutive activations [19]. Instead, we propose to use an Antirectifier layer [1], which combines a sample wise ℓ_2 normalization with two ReLU activations, thereby concatenating the positive and the negative part of the input. This operation can be described as:

$$g(x) = \begin{bmatrix} \max\left(0, \frac{x-\mu_x}{\|x-\mu_x\|_2}\right) \\ \max\left(0, -\frac{x-\mu_x}{\|x-\mu_x\|_2}\right) \end{bmatrix}, \tag{6}$$

where μ_x denotes the mean of x. The Antirectifier effectively introduces non-linearity, while preserving negative signal components as well as the dynamic range of the input.

[1] François Chollet, Antirectifier, Github, https://github.com/keras-team/keras/blob/master/examples/antirectifier.py

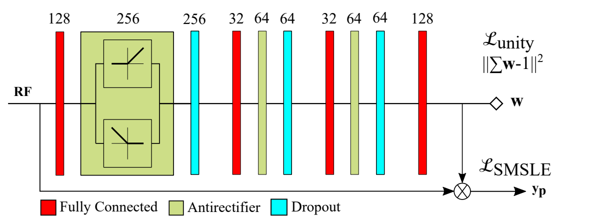

Fig. 2: Schematic overview of the proposed neural network. Above each layer the number of output nodes is indicated.

3.3. Computational Complexity

When looking at the computational complexity of the DNN, we observe that every FC layer requires $2n_in_{i+1} + n_{i+1}$ floating-point-operations (FLOPs) to compute [20], where n_i and n_{i+1} respectively specify the amount of input and output nodes of the layer. ReLU activation, involving a comparison and multiplication, requires 2 FLOPs. Consequently, for our network with L layers the number of required FLOPs amounts:

$$F = \underbrace{2n_0n_1 + n_1}_{\text{input}} + \sum_{i=1}^{L-1} \underbrace{4n_in_{i+1} + n_{i+1}}_{\text{FC layers}} + \underbrace{4n_{i+1}}_{\text{activation}}. \tag{7}$$

3.4. Logarithmic Loss

Conventionally, enveloped signals are logarithmically compressed to create a visually more appealing and insightful image. Consequently, typical mean-square-error or mean-absolute-error metrics penalize errors for high intensity pixels more stringently than for low intensity pixels. We therefore introduce a loss function, the signed-mean-squared-logarithmic-error (SMSLE), which better reflects deviations from desired image properties. The beamformed data is split into a positive (y^+) and negative (y^-) part on which the mean-squared-logarithmic-error is calculated. The total loss is the sum of these two contributions:

$$\mathcal{L}_{\text{SMSLE}} = \frac{1}{2}\|\log_{10}(y_p^+) - \log_{10}(y_t^+)\|_2^2 \quad + \\ \frac{1}{2}\|\log_{10}(-y_p^-) - \log_{10}(-y_t^-)\|_2^2. \tag{8}$$

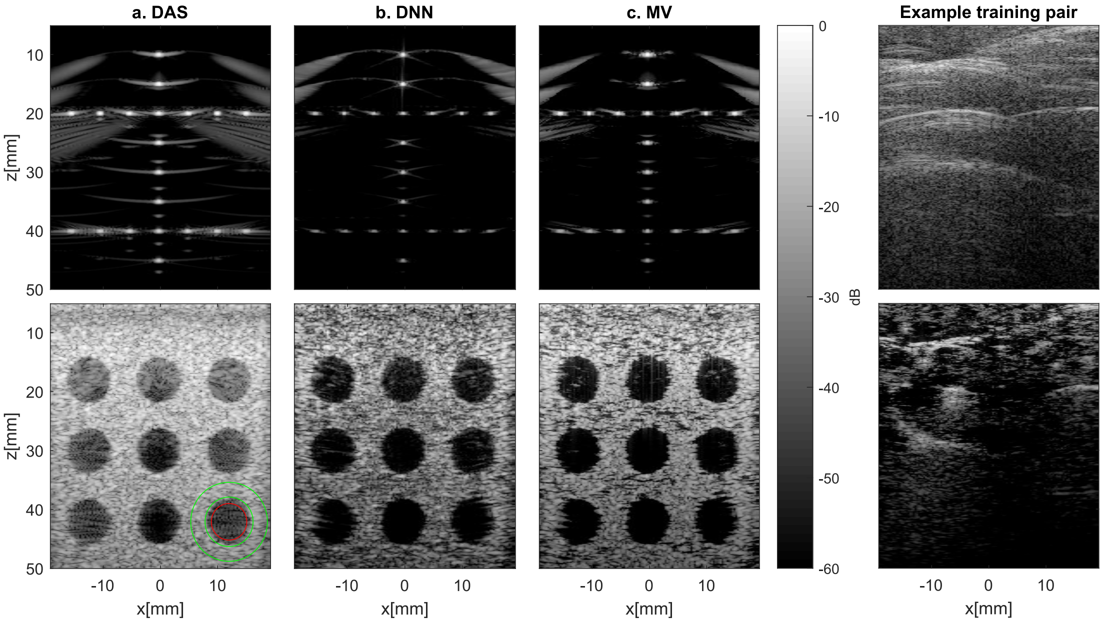

Fig. 3: Left: Comparison of performance between a) Delay-and-sum (DAS) beamforming with Hanning apodization, b) Deep neural network-based (DNN) adaptive beamforming, and c) Minimum variance (MV) beamforming. Right: Example of an *in vivo* carotid artery training pair with on top the RF input and on the bottom the MV reconstructed target image.

Table 2: Resolution and contrast metrics

Parameter	DAS	DNN	MV
$\mathrm{FWHM}_{\mathrm{lat}}$ (mm)	0.846	0.704	0.778
$\mathrm{FWHM}_{\mathrm{ax}}$ (mm)	0.431	0.342	0.434
CNR (dB)	10.96	11.48	12.45

4. RESULTS AND DISCUSSION

After training on *in vivo* data and the simulated point scatterers, the network was validated on unseen simulated images of the PICMUS dataset in order to compare resolution and contrast, as shown in Fig. 3. To this end the PICMUS data, originally sampled at 20.832 MHz, was resampled to match the *in vivo* training data, resulting in some artifacts.

Resolution was assessed by evaluating the averaged full-width-at-half-maxima (FWHM) of all point scatterers. Contrast was estimated using the averaged contrast-to-noise ratio (CNR) of the anechoic cysts, defined as:

$$CNR = 20\log_{10}\left(\frac{|\mu_{in}-\mu_{out}|}{\sqrt{(\sigma_{in}^2+\sigma_{out}^2)/2}}\right), \tag{9}$$

where μ_{in} and μ_{out} represent the mean intensity of the inner (red) and outer (green) regions, respectively, and σ_{in}^2 and σ_{out}^2 the variance of the inner and outer regions. An example of these regions of interest is indicated in Fig 3a. The resulting metrics are shown in Table 2. We observe that the proposed DNN beamformer is able to generate a high contrast image comparable to the MV target, with significantly less clutter. Additionally we see that both adaptive techniques (MV and DNN) show an increase in CNR and resolution compared to DAS, with the DNN even outperforming the MV target on the latter, likely due to its ability of incorporating a generalizable prior in the beamforming process by averaging statistics of the training data. The CNR improvement however is less pronounced since the more sparse speckle pattern of the MV beamformer is penalized by this metric. Naturally, training on higher quality images allows for improved network performance.

From (7) we can determine that the proposed network requires 74,656 FLOPs whereas 2,097,152 = 128^3 FLOPs are required for regularized MV beamforming, thus only 3.6% of the number of operations are needed. In practice this has led to a mean reconstruction time of 0.4 seconds for the DNN, as opposed to 160 seconds for the MV beamformer.

5. CONCLUSION

In this work we demonstrated how deep learning can be used to improve over conventional beamforming methods. Specifically we show that a surprisingly compact, model-based DNN is able to enable the reconstruct high-quality ultrasound images, comparable to those obtained using a state-of-the-art adaptive beamformer, yet at a drastically lower reconstruction time. This permits a real-time implementation of adaptive beamforming in ultrasound systems.

6. REFERENCES

[1] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. A. Jensen, and O. Bernard, "Plane-wave imaging challenge in medical ultrasound," in *2016 IEEE International Ultrasonics Symposium (IUS)*, Sept. 2016, pp. 1–4.

[2] I. K. Holfort, *Adaptive Beamforming for Medical Ultrasound Imaging*, Ph.D. thesis, Technical University of Denmark, DTU Electrical Engineering, 2009.

[3] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Processing*, vol. 53, no. 5, pp. 1684–1696, 2005.

[4] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, Jan. 2012, vol. 25, pp. 1097–1105.

[6] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. G. Rodríguez, "A review on deep learning techniques applied to semantic segmentation," *arXiv: 1704.06857*, 2017.

[7] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, 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.

[8] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean, "Google's neural machine translation system: Bridging the gap between human and machine translation," *arXiv: 1609.08144*, 2016.

[9] H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[10] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1289–1296, 2018.

[11] Y. H. Yoon and J. C. Ye, "Deep learning for accelerated ultrasound imaging," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 6673–6676.

[12] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Transactions on Medical Imaging*, vol. 37, no. 9, pp. 2010–2021, Sept. 2018.

[13] 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, Oct. 2017.

[14] R. J. G. van Sloun, O. Solomon, M. Bruce, Z. Z. Khaing, H. Wijkstra, Y. C. Eldar, and M. Mischi, "Super-resolution ultrasound localization microscopy through deep learning," *arXiv: 1804.07661*, 2018.

[15] A. A. Nair, T. D. Tran, A. Reiter, and M. A. L. Bell, "A deep learning based alternative to beamforming ultrasound images," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 3359–3363.

[16] A. Rodriguez-Molares, H. Torp, B. Denarie, and L. Løvstakken, "The angular apodization in coherent plane-wave compounding [correspondence]," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 62, no. 11, pp. 2018–2023, Nov. 2015.

[17] A. M. Deylami, J. A. Jensen, and B. M. Asl, "An improved minimum variance beamforming applied to plane-wave imaging in medical ultrasound," in *2016 IEEE International Ultrasonics Symposium (IUS)*, Sept. 2016, pp. 1–4.

[18] Y. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[19] 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*, May 2010, pp. 249–256.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004, pp. 661-683.