Adaptive Ultrasound Beamforming Using Deep Learning

Ben Luijten, *Student Member, IEEE*, Regev Cohen, *Student Member, IEEE*,
Frederik J. de Bruijn, Harold A. W. Schmeitz, Massimo Mischi, *Senior Member, IEEE*,
Yonina C. Eldar, *Fellow, IEEE*, and Ruud J. G. van Sloun, *Member, IEEE*

***Abstract*—Biomedical imaging is unequivocally dependent on the ability to reconstruct interpretable and high-quality images from acquired sensor data. This reconstruction process is pivotal across many applications, spanning from magnetic resonance imaging to ultrasound imaging. While advanced data-adaptive reconstruction methods can recover much higher image quality than traditional approaches, their implementation often poses a high computational burden. In ultrasound imaging, this burden is significant, especially when striving for low-cost systems, and has motivated the development of high-resolution and high-contrast adaptive beamforming methods. Here we show that deep neural networks, that adopt the algorithmic structure and constraints of adaptive signal processing techniques, can efficiently learn to perform fast high-quality ultrasound beamforming using very little training data. We apply our technique to two distinct ultrasound acquisition strategies (plane wave, and synthetic aperture), and demonstrate that high image quality can be maintained when measuring at low data-rates, using undersampled array designs. Beyond biomedical imaging, we expect that the proposed deep learning based adaptive processing framework can benefit a variety of array and signal processing applications, in particular when data-efficiency and robustness are of importance.**

***Index Terms*—Ultrasound, adaptive beamforming, deep learning.**

I. INTRODUCTION

THE reconstruction of high-quality images from measured sensor data is essential for diagnostic imaging, facilitating the timely diagnosis and treatment of life threatening diseases and aiding in the personalization of patient care. Unfortunately, current high-quality imaging technologies have major drawbacks: magnetic resonance imaging (MRI) is highly expensive, and X-ray computed tomography (CT) induces harmful ionizing radiation.

Conversely, ultrasound imaging has been an invaluable diagnostic tool due to its low cost, portability and generally causing little discomfort to the patient by being minimally invasive and free of ionizing radiation. Additionally, its high degree of interactability is unique among the various imaging options, enabled by real-time imaging using efficient image reconstruction algorithms. Unfortunately this technique is also known for reduced image quality as compared to other imaging modalities.

In ultrasound imaging, focused sonification strategies have been the standard for many years, where narrow scan lines are sequentially fired by applying focusing time delays to the active transducer elements. The resulting backscattered echoes are then processed to reconstruct a sub-region of the image. The total scan time consequently increases linearly with the amount of acquired scan lines and the physical time-of-flight, which is governed by the speed of sound and desired imaging depth. This poses a trade-off between the achievable frame-rate and lateral image resolution. As a result, this limits lateral resolution in applications where high frame rates are favored.

Recently, there has been a shift towards high-frame-rate unfocused sonification techniques such as Plane Wave (PW) and Synthetic Aperture (SA) imaging. Due to their unfocused transmission, these methods rely more heavily on digital reconstruction algorithms, where dynamic focusing is applied upon reception. In commercial devices, to maintain real-time reconstruction rates, this digital processing is done using a low-complexity, static delay-and-sum (DAS) beamformer, which applies predetermined delays and weights to the channel signals, after which the individual contributions are summed to yield a beamformed signal. This inherently poses a compromise between main lobe width and side lobe intensity, or equivalently, resolution and contrast. The latter is especially affected during unfocused transmissions, where clutter is received from a wide insonified area.

Manuscript received April 17, 2020; revised July 1, 2020; accepted July 3, 2020. Date of publication July 10, 2020; date of current version November 30, 2020. This work was supported in part by the Dutch Research Council (NWO) and Philips Research through the research programme "High Tech Systems and Materials (HTSM)" under Project 17144. *(Corresponding author: Ben Luijten.)*

Ben Luijten and Massimo Mischi are with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands (e-mail: w.m.b.luijten@tue.nl; m.mischi@tue.nl).

Regev Cohen is with the Department of Electrical Engineering, Technion Institute of Technology, Haifa 32000, Israel (e-mail: regev.cohen@gmail.com).

Frederik J. de Bruijn and Harold A. W. Schmeitz are with Philips Research, 5656 AE Eindhoven, The Netherlands (e-mail: frits.de.bruijn@philips.com; harold.schmeitz@philips.com).

Yonina C. Eldar is with the Faculty of Math and Computer Science, Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: yonina.eldar@weizmann.ac.il).

Ruud J. G. van Sloun is with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands, and also with Philips Research, 5656 AE Eindhoven, The Netherlands (e-mail: r.j.g.v.sloun@tue.nl).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TMI.2020.3008537.

Digital Object Identifier 10.1109/TMI.2020.3008537

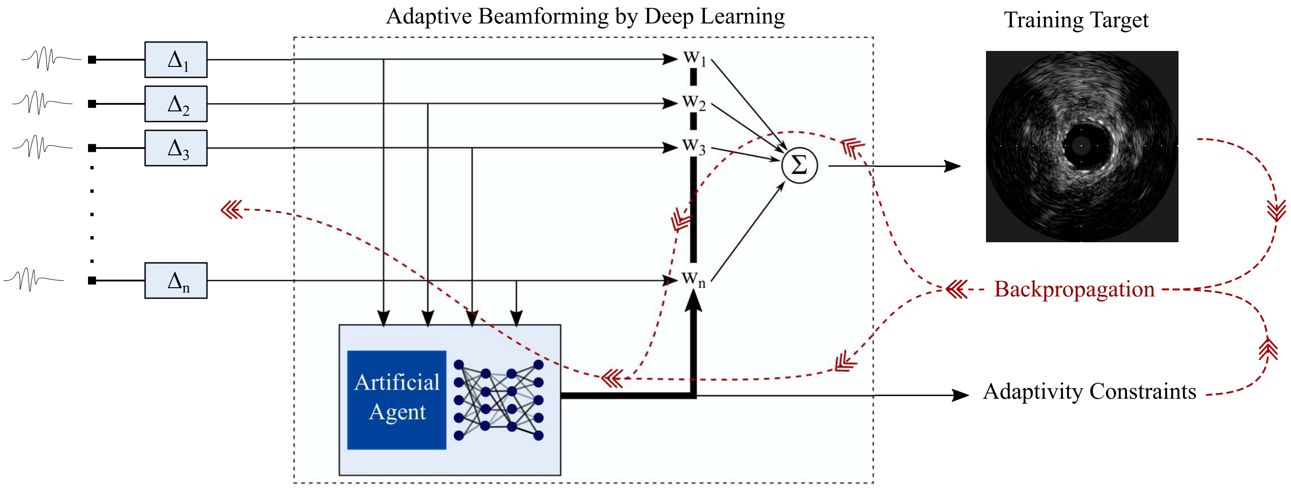

Fig. 1. Schematic overview of ABLE. Using the time aligned array signals as input, an artificial agent calculates an optimal set of apodization weights. The time-aligned array signals are subsequently multiplied by the predicted weights and summed to yield a beamformed output signal. This optimality is learned through training by backpropagating over a training target and adaptivity constraints.

Considerable improvements in image quality have been achieved through data-adaptive beamforming methods, of which Minimum Variance (MV) Distortionless Response [1], [2] is one of the most common approaches. This technique achieves increased resolution without compromising contrast by adaptively placing sidelobes in directions where little energy is received. However, MV beamforming is computationally much more demanding than DAS, resulting in a significantly longer reconstruction time and thereby depriving it of the interactability that makes ultrasound imaging so appealing compared to e.g. MRI and CT. While significant progress has been made to decrease the computational time of MV beamforming algorithms [3], [4], real-time implementation remains a major challenge.

Several lower complexity solutions have been proposed to improve image quality. Coherence Factor (CF) weighing [5] aims to reduce clutter by multiplying the DAS beamformed output with a coherence factor, defined as the ratio between the coherent and incoherent energy across the aperture. This method however suffers from artifacts when the SNR is low, due to the reduced accuracy of the coherent energy estimation [6]. For broad insonification techniques such as PW and SA imaging, this is especially problematic. In [7], Chernyakova *et. al.* proposed an iterative maximum-a-posteriori (iMAP) estimator, which provides a statistical interpretation to beamforming by treating the signal of interest and interference as uncorrelated Gaussian random variables. The method has shown to achieve higher contrast as compared to other low-complexity beamformers described in the literature. A different approach based on the convolution of the delayed RF signals, named COBA, was presented in [8]. This technique has led to better lateral resolution and contrast, while only slightly increasing complexity, as it can be implemented efficiently using the fast Fourier transform.

Meanwhile, deep learning has emerged as a popular tool to solve inverse problems, leading to revolutionary breakthroughs in domains ranging from computer vision to natural language processing. Deep neural networks (DNNs) are trained to perform advanced tasks, typically 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 are universal function approximators [9]. Once trained, inference is typically fast, especially on GPU accelerated systems. DNNs have proven to be very successful in tasks such as image classification [10], image segmentation [11] and speech recognition [12].

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 [13], more recent developments exploit deep learning for the image reconstruction process itself, finding application in X-ray CT, MRI, PET and photoacoustic tomography [14]. An overview of the application of deep learning in various ultrasound imaging tasks can be found in [15]. DNNs have been used for interpolating missing radio frequency (RF) data [16], reducing off-axis scattering [17], compounding of plane waves [18] and super-resolution microscopy [19]. In [20], the beamforming step was circumvented altogether, achieving direct segmentation of cyst phantoms from simulated RF channel data. For many clinical applications however, a brightness-mode (B-mode) image is desired, though at a higher resolution and contrast than that achieved with conventional DAS beamforming. In [21], an end-to-end convolutional neural network (CNN) based pipeline is proposed for data consisting of multiple transmissions. This method has shown improved image quality over DAS, without the need for conventional beamforming steps.

Nonetheless, deep learning strategies are strongly data dependent, relying on vast and diverse training datasets accompanied by a sufficiently large network in order to find a correct and generalized mapping from input to output. Even after convergence, it is hard to prove the robustness of such a network for new, unforeseen data, a trait that is undesirable in a clinical setting where a predictable system response is of vital importance for correct diagnosis of diseases.

Rather than relying on a large general-purpose network in conjuncture with abundant training data, we here atim to learn the parameters of a network that is designed to fulfill a specific subtask, which we then train within the full system by taking an end-to-end approach. This structure constrains the trainable degrees of freedom, and allows for a more compact architecture that is data efficient in training. We here propose Adaptive Beamforming by deep LEarning (ABLE), a method that exploits the algorithmic structure of adaptive beamforming, with a neural network adaptively computing a set of optimal image reconstruction parameters given the received RF data (see Fig. 1) [22]. In this case, these parameters constitute the content-adaptive apodization weights, as in MV beamforming. However, the network is not directly trained to reproduce the values of target apodization weights. Instead, it is trained to reproduce the downstream system output, using the associated beamformed pixel values as training target. Compared to other deep neural networks, adopting this model based structure results in a relatively small amount of trainable parameters, and consequently a lower computational complexity.

We demonstrate the performance of ABLE on two widely different ultrasound imaging modalities: PW (linear array, external) and SA (circular array, intravascular) imaging. Intravascular ultrasound imaging (IVUS) is an invaluable tool for e.g. interventional cardiography. Due to the strong constraints on size and power consumption of the cathether-based probe, this method is particularly susceptible to noise and grating lobe artifacts, making it an excellent candidate to test the merit of our adaptive deep-learning-based beamformer. We then show that ABLE can also learn to cope with subsampled SA transmit designs. The image quality of ABLE is compared against three existing beamformers: DAS, iMAP and MV. Furthermore we assess the computational complexity of these methods to highlight its implication on real-time performance.

While the scope of this work concerns ultrasound beamforming, we emphasize that this method is generally applicable in other applications requiring content-adaptive prediction of signal processing parameters.

The remainder of this paper is organised as follows. In Section II we describe existing beamforming methods. In Section III we introduce ABLE and elaborate on its design and training strategy. Data acquisition, validation of image quality, complexity analysis and subsampling are discussed in Section IV. In Section V, we present experimental results, which are discussed in Section VI. We conclude in Section VII.

II. EXISTING BEAMFORMERS

A. Delay and Sum Beamforming

In ultrasound beamforming, the received channel signals are first time-of-flight corrected by applying delays according to the geometry of the transducer and imaging scene. Pixel-wise time delays for each array channel can be calculated as:

$$\Delta[x,z] = \Delta[\mathbf{r}] = \frac{\|\mathbf{r}_{TX} - \mathbf{r}\|_2 + \|\mathbf{r}_{RX} - \mathbf{r}\|_2}{c}, \tag{1}$$

where $\Delta[\mathbf{r}]$ is the required channel delay to focus to an imaging point $\mathbf{r}$, $\mathbf{r}_{TX}$ and $\mathbf{r}_{RX}$ are the positions of the transmitting and receiving element in the array, respectively, and c is the speed of sound in the medium. By dynamically focusing to different imaging points, corresponding to pixels in the final image, a pixel-wise response can be generated for every receiving array element $n \in \{1, \ldots, N\}$, given by

$$\mathbf{y}_n[x,z] = \mathbf{x}_n[\Delta_n[x,z]], \tag{2}$$

where $\mathbf{x}_n[t]$ denotes the channel signal for the n^{th} element.

A DAS-beamformed image $\mathbf{P}_{\text{DAS}}[x,z]$ is constructed by multiplying the focused contributions of each receiving channel with predetermined (but possibly spatially varying) weights $\mathbf{w}[x,z]$, and summing the result:

$$\mathbf{P}_{\text{DAS}}[x,z] = \mathbf{w}^{\text{H}}[x,z]\mathbf{y}[x,z]. \tag{3}$$

Often, these weights are set according to a boxcar ($\mathbf{w} = \mathbf{1}^N$), Hanning or tapered cosine window, to either optimize resolution or contrast.

B. Iterative Maximum-a-Posteriori

The iMAP algorithm [7] provides a reweighing of the DAS beamformed output for $\mathbf{w} = \mathbf{1}^N$, calculated in two iterative steps. First, the signal and noise variances are estimated by

$$\{\hat{\sigma}_s^2, \hat{\sigma}_n^2\}_t = \left\{\hat{\mathbf{P}}^2_{\text{iMAP},t}, \frac{1}{N}\|\mathbf{y} - \hat{\mathbf{P}}_{\text{iMAP},t}\mathbf{1}\|^2\right\}, \tag{4}$$

and initializing $\hat{\mathbf{P}}_{\text{iMAP},(0)} = \mathbf{1}^{\text{H}}\mathbf{y}$, where $(\cdot)^{\text{H}}$ stands for the Hermitian transpose. Using these variances, a MAP estimate of the beamformed signal is given by

$$\hat{\mathbf{P}}_{\text{iMAP},(t+1)} = \frac{\hat{\sigma}_{s,t}}{\hat{\sigma}^2_{n,t} + N\hat{\sigma}^2_{s,t}}\mathbf{1}^{\text{H}}\mathbf{y}. \tag{5}$$

Equation (4) and (5) are iterated until a stopping criterion is met. In the original iMAP paper, the authors observed little to no improvement after 2 iterations. As such, throughout the rest of the paper we will consider iMAP2, i.e. the algorithm using 2 iterations.

C. Adaptive Eigen-Based Minimum Variance Beamforming

The MV beamformer provides a set of content-adaptive apodization weights for each pixel; these weights are selected to minimize output power (variance) of the beamformed signal, subject to a distortionless response in the desired direction. Accordingly, the following minimization problem is solved for every pixel in the image:

$$\begin{aligned}&\min_{\mathbf{w}} \mathbf{w}^{\text{H}}[x,z]\mathbf{R}[x,z]\mathbf{w}[x,z]\\ &s.t.\ \mathbf{w}^{\text{H}}[x,z]\mathbf{a} = 1\end{aligned} \tag{6}$$

where $\mathbf{R}$ denotes the sample covariance matrix calculated across the delayed receiving array signals, and $\mathbf{a}$ is a steering vector. Since the data is already time-of-flight corrected, we select $\mathbf{a} = \mathbf{1}^N$. Solving (6) yields the following closed form solution:

$$\mathbf{w}_{\text{MV}}[x,z] = \frac{\mathbf{R}^{-1}[x,z]\mathbf{a}}{\mathbf{a}^{\text{H}}\mathbf{R}^{-1}[x,z]\mathbf{a}}. \tag{7}$$

With the aim of avoiding a potentially unstable numerical inversion of $\mathbf{R}$, the covariance matrix is estimated by applying spatial smoothing using a subaperture $\tilde{\mathbf{y}}_l[x,z] \in \mathbb{R}^L$ of the full response $\mathbf{y}[x,z]$:

$$\hat{\mathbf{R}}[x,z] = \frac{\sum_{l=0}^{N-L} \tilde{\mathbf{y}}_l[x,z]\tilde{\mathbf{y}}_l^{\mathrm{H}}[x,z]}{N-L+1}. \tag{8}$$

where

$$\mathbf{y}_l[x,z] = \Big[y_l[x,z], \quad y_{l+1}[x,z], \quad \ldots \quad y_{l+L-1}[x,z] \Big] \tag{9}$$

To further improve stability, diagonal loading is applied by adding a factor ϵ to the diagonal of $\hat{\mathbf{R}}$ determined by the trace operator as:

$$\epsilon = D \cdot trace(\hat{\mathbf{R}}), \tag{10}$$

where D is the diagonal loading factor. The resulting solution for $\mathbf{w}$ corresponds to regularized MV and is computationally dominated by the inversion of the $N \times N$ covariance matrix $\mathbf{R}$.

Additional improvements in image contrast can be achieved by employing an Eigen-Based Minimum Variance Beamformer (EBMV) [23], at the cost of higher computational complexity. Assuming that the received signal is a combination of desired signal components and added noise, we can take the eigen-decomposition of $\mathbf{R}$ and project the signal subspace $\mathbf{E}_{\mathrm{signal}}$, composed of the dominant, desired eigenvectors, on the weight vector to obtain the updated weights:

$$\mathbf{w}_{\mathrm{EBMV}} = \mathbf{E}_{\mathrm{signal}}\mathbf{E}_{\mathrm{signal}}^{\mathrm{H}}\mathbf{w}. \tag{11}$$

This procedure is repeated for each pixel to obtain a EBMV-beamformed image:

$$\mathbf{P}_{\mathrm{EBMV}}[x,z] = \mathbf{w}_{\mathrm{EBMV}}^{\mathrm{H}}[x,z]\mathbf{y}_l[x,z]. \tag{12}$$

III. ABLE: ADAPTIVE BEAMFORMING BY DEEP LEARNING

A. Network Architecture

ABLE's architecture is inspired by the MV beamformer, with a neural network that adaptively calculates apodization weights according to the input channel data. Thus, the traditional computationally expensive adaptive processor is replaced with a more computationally efficient artificial agent in the form of a neural network, as shown in Fig. 1. This specific design allows the artificial agent to adaptively tune a set of image reconstruction parameters, in this case a set of real valued the apodization weights, within a predictable beamforming structure. A beamformed pixel is obtained by multiplying the time-of-flight corrected array channel data for that pixel with the apodization weights as produced by the neural network, $f_\theta : \mathbb{R}^N \rightarrow \mathbb{R}^N$, such that:

$$\mathbf{P}_{\mathrm{ABLE}}[x,z] = f_\theta(\mathbf{y}[x,z])^{\mathrm{H}}\mathbf{y}[x,z]. \tag{13}$$

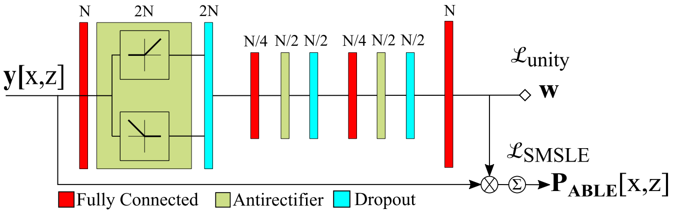

Fig. 2. Schematic overview of the proposed neural network. Above each layer the number of output nodes is denoted as a function of the aperture size *N*. Note that the antirectifier outputs double the amount of nodes of the preceding fully connected layers.

While convolutional neural networks have demonstrated their efficiency in image processing tasks, they also rely on the assumption that operations are spatially invariant. For pixel-based ultrasound beamforming, this can not be equivalently assumed, as the focus point changes, depending on the image location. Instead, we opt for a compact fully connected (FC) network which operates on a per-pixel basis. As such, the neural network is nested in an otherwise conventional beamforming structure.

The proposed network f_θ consists of 4 fully connected (FC) layers comprising N nodes for the outer layers, corresponding to the receiving aperture size, and $N/4$ nodes for the inner layers, as indicated in Fig. 2. This encoder-decoder structure introduces a compact latent space which forces the network to find a compact representation of the data, aiding in suppression of noise and incoherent signals. Between every FC layer, dropout is applied, randomly dropping nodes with a probability of 0.2, to prevent overfitting.

B. Activation Functions

Rectified linear units (ReLUs) are the most commonly used activations in DNNs because of their computational efficiency, ability to provide sparse representations and largely avoiding vanishing gradients due to a positive unbounded output [24]. However, when dealing with RF data, containing positive and negative values, such a non-linearity may not be appropriate, 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 [25].

Instead, we propose to use an antirectifier layer,[1] which combines a sample wise ℓ_2 normalization with the concatenation of the positive and the negative part of the input. This operation is described as:

$$g(\mathbf{x}) = \begin{bmatrix} \max\left(0, \frac{\mathbf{x}-\mu_x}{\|\mathbf{x}-\mu_x\|_2}\right) \\ \max\left(0, -\frac{\mathbf{x}-\mu_x}{\|\mathbf{x}-\mu_x\|_2}\right) \end{bmatrix}, \tag{14}$$

where μ_x denotes the mean of $\mathbf{x}$. The antirectifier effectively introduces non-linearity, while preserving the contributions of negative signal components as well as the dynamic range of the input.

C. Training Strategy

Our network is implemented in Python using the Keras API with a Tensorflow (Google, CA, USA) backend on a compute

---

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

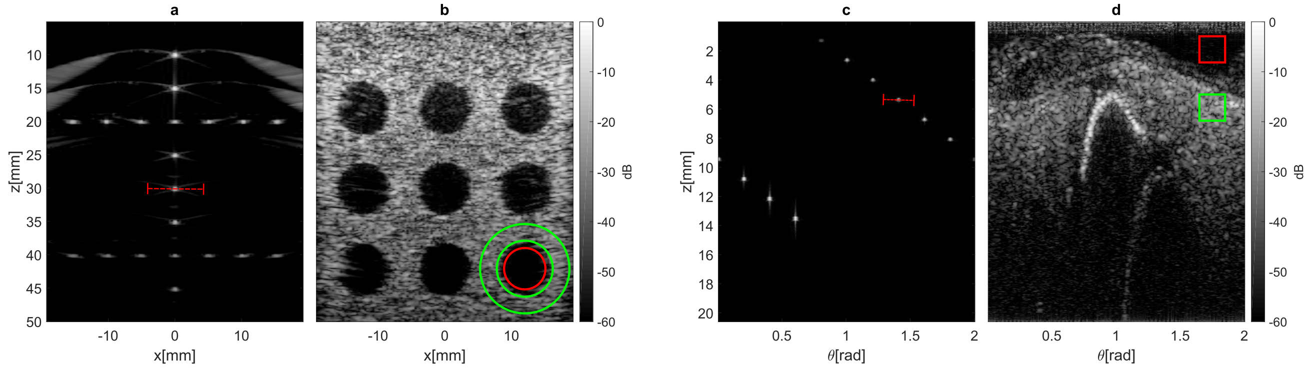

Fig. 3. Images used for the performance evaluation of the beamforming methods. ABLE's reconstructions are shown for: a) Simulated point scatterers (plane wave, linear array) b) Simulated anechoic cysts (plane wave, linear array) c) Simulated point scatterers (synthetic aperture, IVUS circular array) d) *In-vivo* coronary artery (synthetic aperture, IVUS circular array). For contrast evaluation, high and low intensity regions are indicated in green and red, respectively.

cluster with an Nvidia Tesla K80 graphics card. For training, the Adam optimizer was used with a learning rate of 0.001, stochastically optimizing the loss using batches that each contain all pixels belonging to a single image. A randomized training strategy was employed, where the data sequence supplied to the network was randomized at each epoch. The total training loss comprised an image loss $\mathcal{L}_{\text{SMSLE}}$, and an apodization-weight penalty $\mathcal{L}_{\text{unity}}$, leading to:

$$\mathcal{L}_{\text{total}} = \lambda \mathcal{L}_{\text{SMSLE}} + (1-\lambda)\mathcal{L}_{\text{unity}}, \tag{15}$$

where λ is a parameter between 0 and 1 which determines the weight ratio of the losses. We detail on the individual loss functions in the following.

1) Image Loss: The image loss is designed to promote similarity between the target image and the image produced using ABLE. Given the large dynamic range of ultrasound backscattering, ultrasound images are typically logarithmically compressed to create a visually more appealing and insightful image. Consequently, conventional loss metrics (e.g. mean-square-error (MSE) or mean-absolute-error (MAE)) do not provide a balanced representation of deviation from desireable image properties. To better reflect this, we introduce the signed-mean-squared-logarithmic error (SMSLE) as a loss function. The beamformed data is split into a positive ($\mathbf{P}^+$) and negative ($\mathbf{P}^-$) part on which the mean-squared-logarithmic-error is calculated. The total loss is the sum of these two contributions, resulting in:

$$\begin{aligned}\mathcal{L}_{\text{SMSLE}} = \frac{1}{2}\|\log_{10}(\mathbf{P}^+_{\text{ABLE}}) - \log_{10}(\mathbf{P}^+_{\text{MV}})\|_2^2 \\ + \frac{1}{2}\|\log_{10}(-\mathbf{P}^-_{\text{ABLE}}) - \log_{10}(-\mathbf{P}^-_{\text{MV}})\|_2^2. \end{aligned}\tag{16}$$

2) Weight Penalty: The direct access to the generated apodization weights permits additional penalties to be invoked during training. Inspired by MV beamforming, we promote a distortionless response by penalizing deviations from unity gain using:

$$\mathcal{L}_{\text{unity}} = |\mathbf{1}^T f_\theta(\mathbf{y}[x,z]) - 1|^2. \tag{17}$$

IV. EXPERIMENT SETUP

A. Data Acquisition

For training, we acquired both PW (linear array) and SA (IVUS) datasets, each comprising 900 *in-vivo* frames of the carotid artery and wrist of a single person, and a coronary artery of a single patient, respectively. In addition to this, 100 simulated frames of point scatterers were generated for each acquisition method using Field II [26]. These point scatterers were randomly distributed over the entire field of view of the transducer. For the PW data, that is from ± 19 mm laterally measured from the center of the transducer, and up to a depth of 63 mm. For the SA IVUS data, the points were generated over the full rotation, up to a depth of 50 mm).

In-vivo PW images were acquired using the Vantage system (Verasonics Inc., WA, USA) in combination with the L11-4v linear transducer. SA IVUS images were recorded using the EEP PV014P circular transducer (Philips Volcano, USA). In this particular system, each transmit is received by an aperture of 14 elements. During beamforming, a virtual aperture is constructed by combining the received signals of multiple transmits and receives. The transducer parameters for the respective acquisition methods are indicated in Table I.

For both imaging methods a separate dataset of 100 image frames each was used for testing purposes. For the PW case, this included recordings of the carotid artery and wrist of the same subject as for the training set, although they were different views and acquired 4 months later. The hold-out test set was supplemented with simulated PW images from the PICMUS dataset [27]. The SA test set comprised frames of an *in-vitro* wire-target phantom, an arterial stent, and a *in-vivo* coronary artery. Model development and hyperparameter tuning was performed on an early train/validation split of the train set. However, the model was tested on the later acquired test set, without additional fine tuning.

For each imaging mode, we retrain ABLE on 1000 training images (90% *in-vivo* and 10% simulations) to map time-of-flight corrected RF input signals to corresponding high-quality beamformed targets. Ultimately this yields two beamformer implementations, for PW and SA, respectively. For the

TABLE I
TRANSDUCER PARAMETERS

Parameter	Plane wave	Synthetic aperture
Elements	128	64
Aperture size	128	105 (virtual)
Pitch	0.300 mm	0.057 mm
Transmit Frequency	6.25 MHz	20MHz
Sampling Frequency	25 MHz	100MHz

TABLE II
EBMV PARAMETERS

Parameter	Plane wave	Synthetic aperture
Diagonal loading (D)	0.01	1
Array size (N)	128	105
Sub array size (L)	32	105

generation of targets we employ an EBMV beamformer on the RF channel data, of which the beamforming parameters are provided in Table II. Notably, the pixel-based structure of ABLE results in over 260,000 training data samples per image, with each sample being the set of (time-of-flight-corrected) channel signals corresponding to a specific pixel.

B. Validation

Resolution was assessed by evaluating the Full-Width-at-Half-Maxima (FWHM) of simulated point scatterers, which were calculated for both the lateral and the axial direction. For the circular-array based SA scans, lateral resolution was evaluated in the polar domain. An example of these simulated point-scatterers can be seen in Fig. 3a and 3c.

Contrast of the images was measured by the contrast-to-noise ratio (CNR) between a low and high intensity region, defined as:

$$\mathrm{CNR} = 20\log_{10}\left(\frac{|\mu_{low} - \mu_{high}|}{\sqrt{(\sigma_{low}^2 + \sigma_{high}^2)/2}}\right), \tag{18}$$

where μ_{low} and μ_{high} represent the mean intensity of the low and high intensity regions, respectively, and σ_{low}^2 and σ_{high}^2 the variance of the low and high intensity regions, respectively. The CNR was evaluated on simulated anechoic cysts for PW imaging with a linear array, and an *in-vivo* coronary artery for SA IVUS imaging (see Fig. 3b and 3d).

Finally we compare the mean-absolute-error (MAE), between the target images (EBMV) and the other imaging modalities, for all test images. It should be noted that this metric does not provide an accurate indication of resolution or contrast, but quantitatively measures similarity with our training targets.

C. Algorithmic Complexity

Here we assess the computational complexity of the adaptive beamformers described in Section II, in terms of floating-point-operations (FLOPs). Since all beamformers require time-of-flight correction, multiplication and summation of the array signals, as in DAS, these terms are disregarded. In other words, only the adaptive complexity is taken into account.

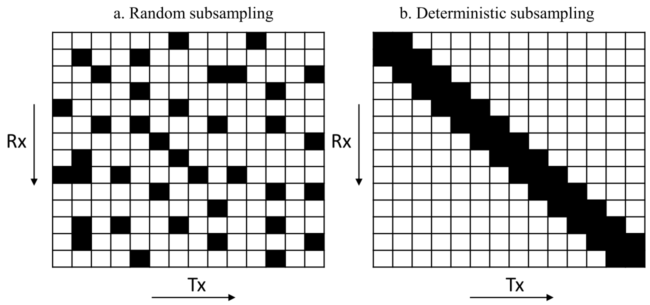
a. Random subsampling
b. Deterministic subsampling

Fig. 4. Visualization of the proposed subsampling schemes. Black boxes indicate active elements. a) Random subsampling, b) Deterministic subsampling.

1) iMAP: Following equation (4) and (5), we observe that for every iMAP iterations an estimate of signal and noise variances, and an update of the MAP estimate is required. Correspondingly, the total number of computations amounts to

$$C_{\mathrm{iMAP}} = I(\underbrace{3N+2}_{\text{variance estimation}} + \underbrace{3}_{\text{MAP estimate}}) = I(3N+5) \tag{19}$$

where I denotes the number of iterations. This yields a complexity of the order $\mathcal{O}(N)$.

2) MV Beamforming: The complexity of MV beamforming is largely dependent on the inversion of the covariance matrix $\mathbf{R}$, which requires approximately N^3 FLOPs [28]. Additionally, the calculation of the weight vector $\mathbf{w}$ as in (7) requires a vector-vector and a vector-scalar product, and two vector-matrix products. Consequently, the total number of FLOPs needed can be determined by:

$$C_{\mathrm{MV}} = \underbrace{N^3}_{\text{matrix inversion}} + \underbrace{2N^2 + 3N}_{\text{weight calculation}}. \tag{20}$$

The EBMV beamformer relies on an eigen-decomposition of R, which similarly adds approximately N^3 FLOPs, and a projection of the dominant eigenvectors as described in (11). This results in a total number of FLOPs

$$C_{\mathrm{EBMV}} = C_{\mathrm{MV}} + \underbrace{N^3}_{\text{eigen decomposition}} + \underbrace{kN^2 + N^2}_{\text{subspace projection}}. \tag{21}$$

In our implementation we project approximately 50% of the eigenvectors, i.e. $k = 0.5$. As such, the computational complexity of these MV methods is of order $\mathcal{O}(N^3)$.

3) ABLE: We consider the complexity of ABLE during inference (i.e. a forward pass). Every FC layer can be described as a matrix-vector product with an added bias $\mathbf{Ax}+\mathbf{b}$, where $\mathbf{A}$ is the incidence matrix containing the weights of the layer, $\mathbf{b}$ is the bias vector, and $\mathbf{x}$ the input vector. Given that $\mathbf{A}$ is an $N_i \times N_{i+1}$ matrix, where N_i and N_{i+1} respectively denote the amount of input and output nodes of the layer, and similarly $\mathbf{b}$ has N_{i+1} elements, we can determine that a total of $2N_iN_{i+1}+N_{i+1}$ FLOPs are required for a single FC layer [29].

ReLU activation, involving a comparison and multiplication, requires 2 FLOPs. As a result, for the antirectifier, which

TABLE III
RESOLUTION AND CONTRAST METRICS

Plane wave (linear array)

	DAS	iMAP2	ABLE	EBMV
FWHM_{lat} (mm)	0.846 ± 0.10	0.730 ± 0.12	0.704 ± 0.16	0.778 ± 0.14
FWHM_{ax} (mm)	0.431 ± 0.006	0.430 ± 0.008	0.218 ± 0.069	0.434 ± 0.013
CNR (dB)	11.35 ± 1.45	9.59 ± 2.81	11.91 ± 1.03	12.67 ± 1.04
MAE	9.86 ± 4.15	3.37 ± 1.41	3.29 ± 1.48	-

Synthetic Aperture (circular array)

	DAS	iMAP2	ABLE	EBMV
FWHM_{lat} (rad)	0.127 ± 0.003	0.094 ± 0.004	0.055 ± 0.009	0.087 ± 0.008
FWHM_{ax} (mm)	0.217 ± 0.021	0.173 ± 0.039	0.118 ± 0.026	0.046 ± 0.025
CNR (dB)	11.21 ± 3.42	10.09 ± 3.00	13.98 ± 1.10	13.62 ± 0.8109
MAE	19.58 ± 0.53	6.05 ± 1.77	0.65 ± 0.18	-

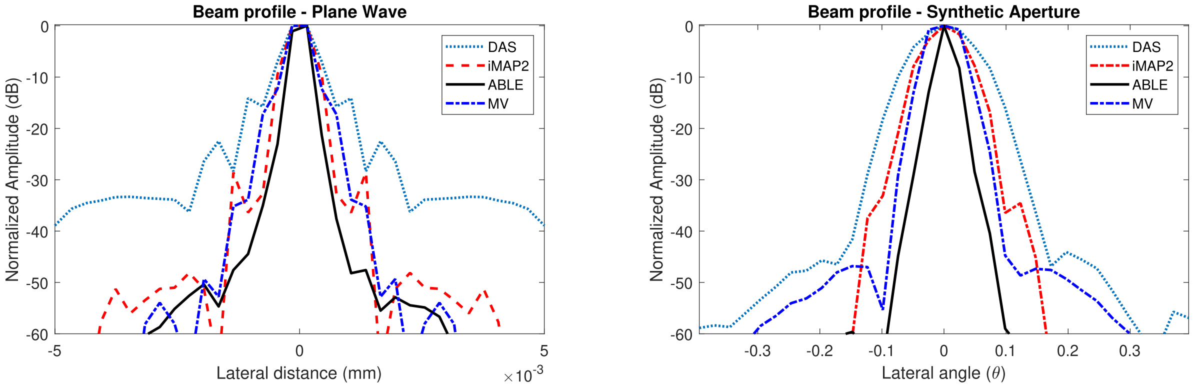

Fig. 5. Normalized beamprofiles along a simulated pointscatterer for (left) Plane-wave imaging and (right) Synthetic Aperture (IVUS) imaging.

concatenates two ReLU activations, 4 FLOPs are needed per node. Since dropout is not active during inference, it does not add to the complexity of the deployed network. In total, for our L-layer network, the number of required FLOPs amounts to

$$C_{ABLE} = \underbrace{2N_0N_1 + N_1}_{\text{input}} + \sum_{i=1}^{L-1} \underbrace{4N_iN_{i+1} + N_{i+1}}_{\text{FC layers}} + \underbrace{4N_i}_{\text{activation}} . \quad (22)$$

Altogether, we observe ABLE's complexity to be $\mathcal{O}(N^2)$.

D. Subsampling

Finally we evaluate ABLE's ability to reconstruct images from a subset of transducer elements in SA IVUS, using two different schemes (illustrated in Fig. 4). In the first, a subset of array elements is selected randomly from the full receiving array. In the second, the receiving elements that are close to the transmitting element are chosen, thereby favouring specular reflections. We train the network to reconstruct images with subsampled array signals, at rates of 25% and 50%, as input, and the fully sampled EBMV reconstructions as target.

During initial experiments we observed that ABLE was not able to adequately reconstruct images from such acquisitions directly. We therefore designed a two-stage variant of the original ABLE architecture, which first performs beamforming on the received signals belonging to each single transmit separately, and subsequently performs weighted compounding across transmits with a second instance of the neural network. In doing so, we create a network that is twice as deep as the single stage network, but with a lower amount of trainable parameters and computational cost due to the absence of interconnections between signals belonging to different transmits in the first beamforming stage. As such the complexity is still of $\mathcal{O}(N^2)$ and thus inference is relatively fast. The resulting reduction in degrees of freedom further constrains the network to the desired beamforming task, and in practice aided the reconstruction from undersampled array data.

V. RESULTS

The obtained axial resolution, lateral resolution, CNR and MAE for the beamforming methods, as described in Section II, are given in Table III, showing that ABLE outperforms conventional DAS and iMAP2 beamforming for both PW (linear array) and SA (IVUS) imaging. Their beam profiles are visualized in Fig. 5. Interestingly, ABLE even yields better lateral resolution than its train target, EBMV, displaying the power of mean-error optimization across many examples.

Figure 6 shows the PW images obtained using DAS, iMAP2, ABLE and EBMV beamforming. From the simulated point scatters one can see that DAS beamforming yields wider and more pronounced sidelobes than iMAP2, ABLE and EBMV. The *in-vivo* images show strong clutter suppression for ABLE and EBMV compared to DAS. Figure 7a exemplifies the apodization weights calculated by ABLE along a lateral line containing a set of equally spaced point-scatterers

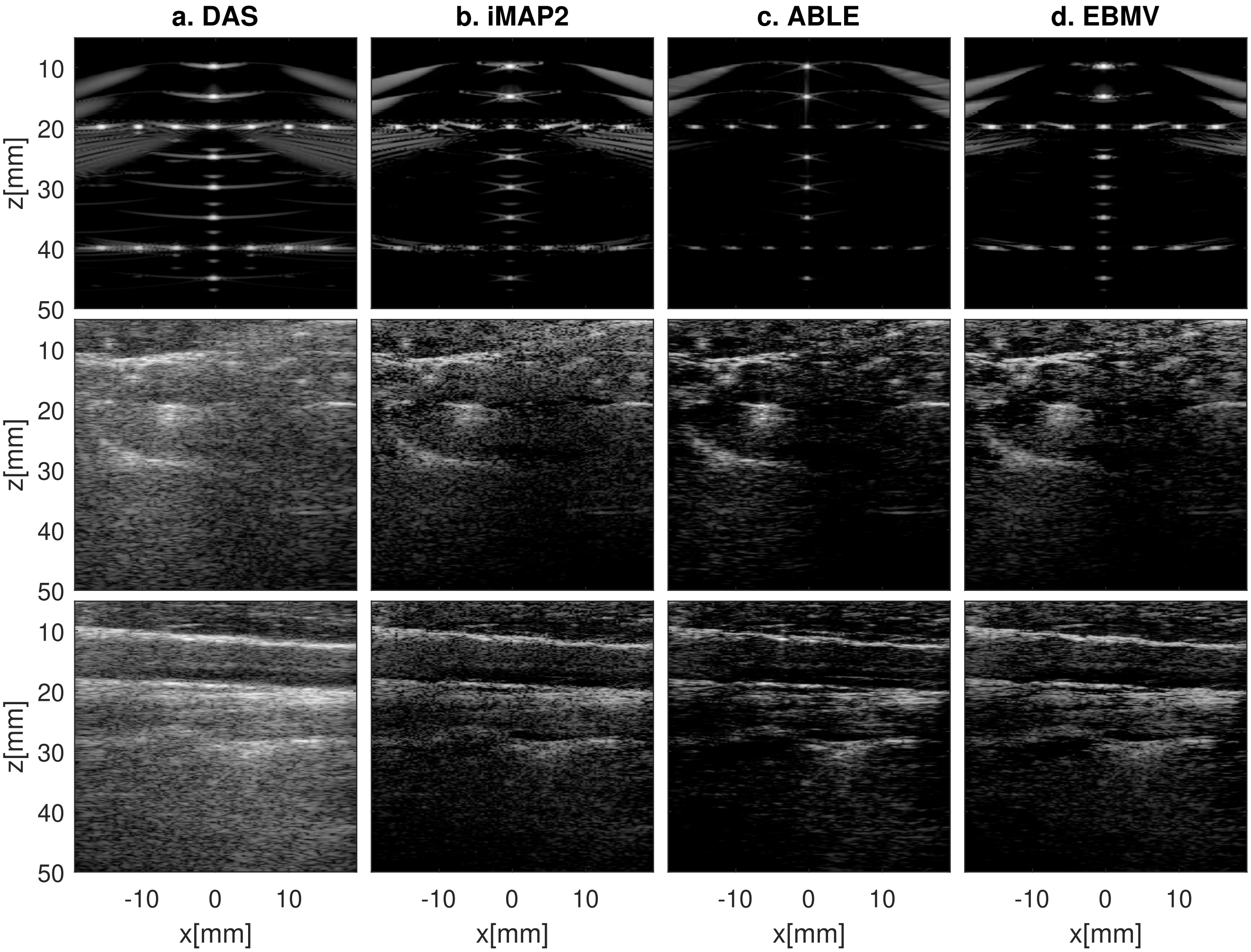

Fig. 6. Single-plane-wave *in-vivo* images using: a) Delay-and-sum (DAS) beamforming with Hanning apodization, b) iMAP2 beamforming c) Adaptive beamforming by deep learning (ABLE), and d) Eigen-Based Minimum variance (EBMV) beamforming. Images are logarithmically compressed with a dynamic range of 60dB. From top to bottom we have simulated point scatterers, carotid artery cross-section, and a carotid artery longitudinal cross-section.

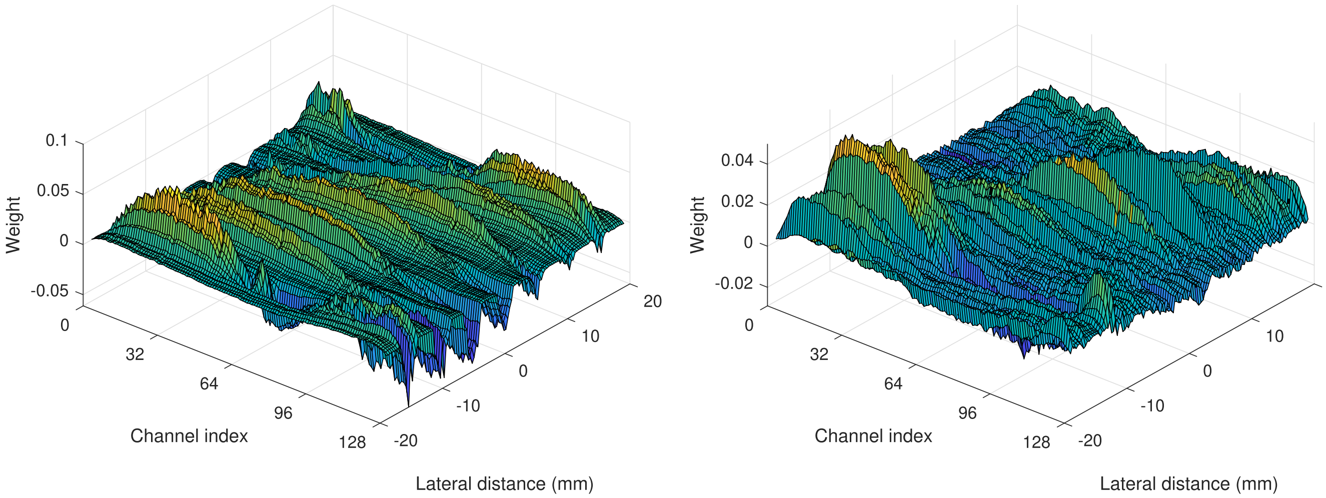

Fig. 7. Visualization of the set of predicted apodization weights for: a) Laterally spaced point-scatterers at a depth of 20 mm, b) Carotid artery cross-section at a depth of 13 mm.

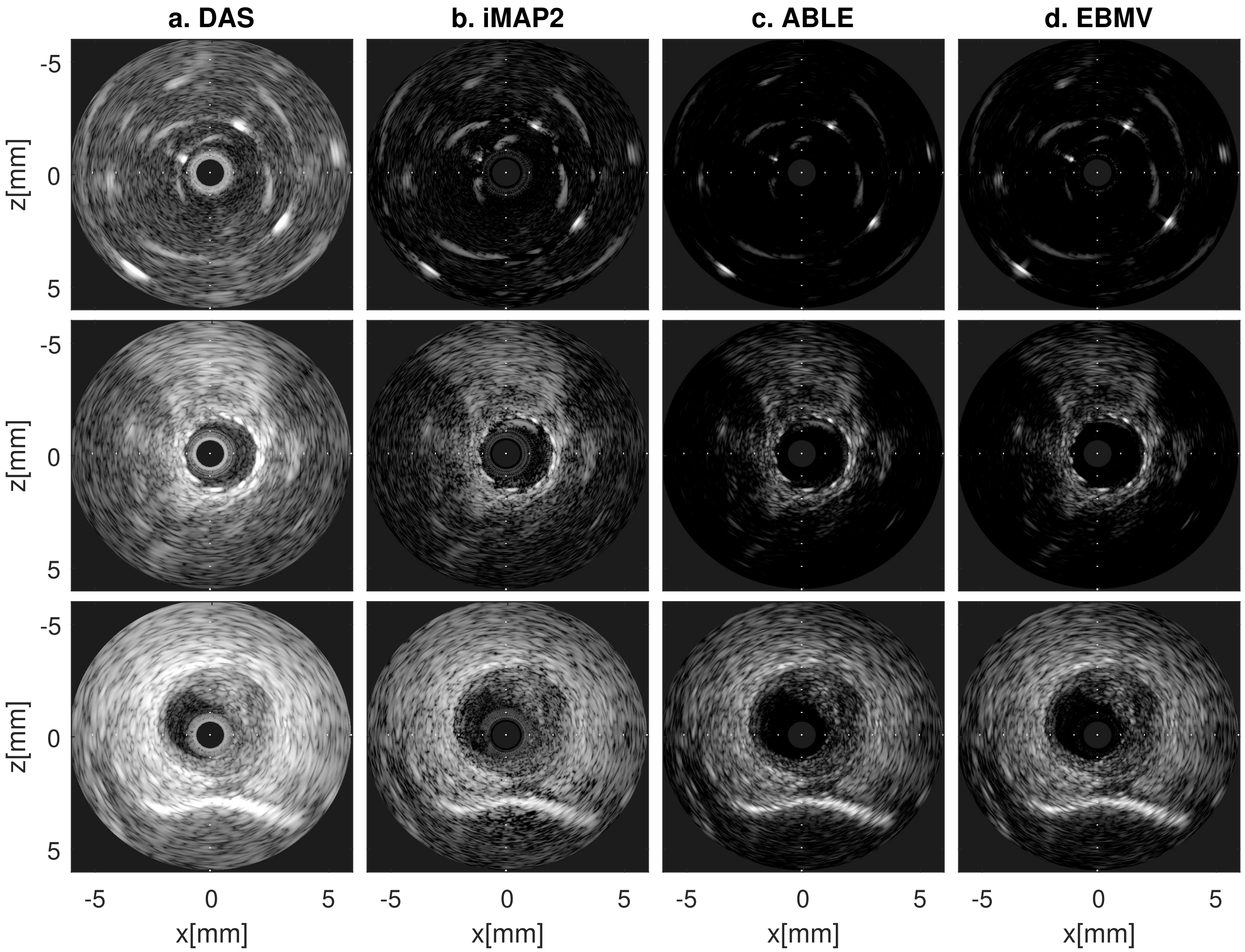

Fig. 8. Reconstructed Synthetic Aperture *in-vivo* images using: a) Delay-and-sum (DAS) beamforming with Hanning apodization, b) iMAP2 beamforming c) Adaptive beamforming by deep learning (ABLE), and d) Eigen-Based Minimum variance (EBMV) beamforming. Images are logarithmically compressed with a dynamic range of 60dB. From top to bottom the we have a wire phantom, an arterial stent, and a coronary artery with plaque.

(see Fig. 3a). The interpretable adaptivity of the apodization, and its relation with the spatial location of the point-scatterers is noteworthy. The reconstructed *in-vivo* SA IVUS images are given in Fig. 8. We again observe a notable reduction in sidelobe intensity near the simulated point scatterers for ABLE and EBMV, as compared to DAS. Consequently, in the *ex-vivo* arterial stent image, we see a more pronounced distinction between tissue and the stent struts. Substructures in the coronary plaque also become more clearly visible due to reduced clutter and improved resolution.

Next, we asses the computational complexity of the beamformers. From (22), we determine that ABLE requires 71,232 and 47,930 FLOPs to calculate the apodization weights for the linear array ($N = 128$) and the IVUS circular array ($N = 105$), respectively. This is well below the 6,186,599 and 414,994 FLOPs required for EBMV beamforming on those systems. In practice this led to a speed up of about a factor 400.

Finally, we evaluate ABLE on two subsampling schemes, using the the two-stage ABLE network. Beamformed images using both subsets (random and deterministic) with 50% and 25% of the transmit-receive pairs active are given in Fig. 9.

VI. DISCUSSION

For both PW and SA imaging, ABLE yields high-contrast, high-resolution imaging that is qualitatively comparable to the MV target with significantly less clutter than DAS. Quantitatively, MAE metrics indicate that ABLE converges well to EBMV beamformed targets. Furthermore we see that all adaptive techniques (iMAP2, EBMV and ABLE) display an increase in CNR and resolution compared to DAS, where ABLE even manages to outperform its train target (EBMV) on the latter, likely due to its ability of incorporating a generalizable prior in the beamforming process by averaging statistics of the training data. Additionally the encoder-decoder structure we employed here forces the network to find a compact representation of the information contained in the channel signals. This effectively regularizes the predicted apodization weights, as can be seen in Fig 7.

Another property we observe in ABLE, is its ability to easily generalize to various types of data, which were not included in the training set. The currently employed structure, in which only adaptive apodization weights are predicted, is a key factor in this robust predictions, as it lead to a predictable linear beamforming structure. While this arguably is a desired

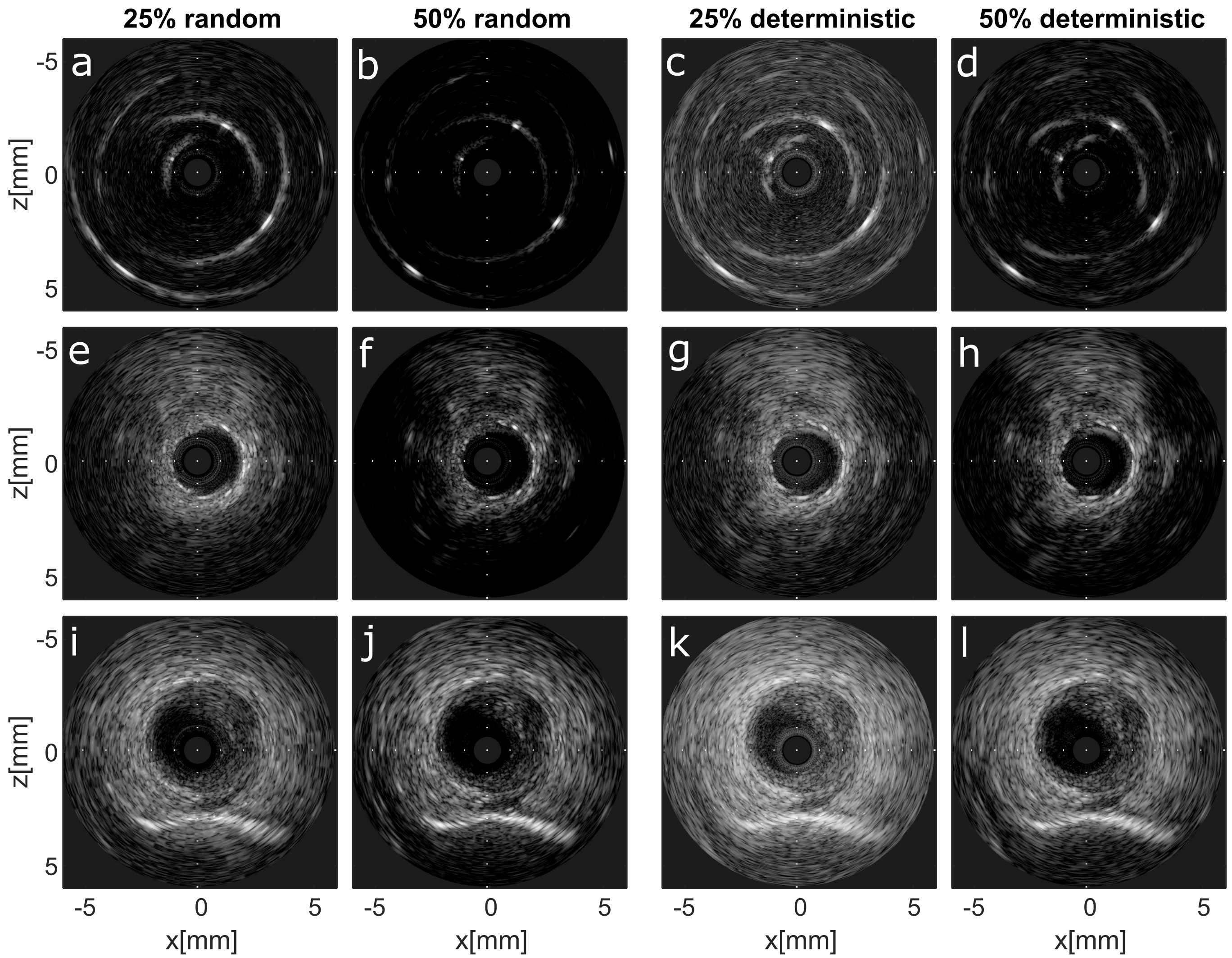

Fig. 9. IVUS imaging by two-stage ABLE from subsampled channel data at rates of 25% and 50% for a random and deterministic scheme. a-d) Reconstructions of simulated point scatterers, e-h) Reconstructions of an *ex-vivo* arterial stent and, i-l) Reconstructions of an *in-vivo* coronary artery.

property, especially in a clinical context, it also restricts the network in terms of potential image quality. To overcome these limitations, future work may include the exploration of more non-linear, yet interpretable, beamforming structures.

For IVUS imaging, adaptive EBMV beamforming yields a strong suppression of typical ringdown artifacts caused by residual vibrations after transmission. ABLE learns this behavior from the MV beamformer, in which these artifacts, caused by high-intensity yet mostly incoherent signal components, are minimized by optimizing the weights such that the signals destructively interfere with each other. Strong scatterers that were previously obscured by the ringdown become visible due to the highly correlated nature of their echoes.

After training the proposed network on subsampled data we again see that for both the random and deterministic strategies, ABLE is capable of reconstructing high-contrast images. However when comparing to the full-array reconstructions, some noticable artifacts are visible, especially at 25% channel subsampling. We see that these artifacts predominantly emerge in regions where mainly speckle is present. This is likely due to the (colored) noise-like properties of speckle, which are challenging to capture with only a limited amount of receiving channels. In contrast, strong point-scatterers remain clearly visible, relevant for applications such as arterial stent detection.

When comparing the two subsampling strategies at equal subsampling rates, we observe noticable differences in the reconstructed images. In the randomly subsampled images we notice a wide sidelobe pattern. Yet, grating lobes appear less pronounced. A seemingly better suppression of noise and clutter in (mostly) empty regions leads to an overall better contrast than the deterministic method.

We here propose to leverage neural networks to learn fast approximations of adaptive algorithms. To justify this, we can asses the algorithmic complexity of these methods, which in turn is strongly correlated with their respective reconstruction times and furthermore constitutes a system-independent measure. In Fig. 10 the number of required FLOPs required

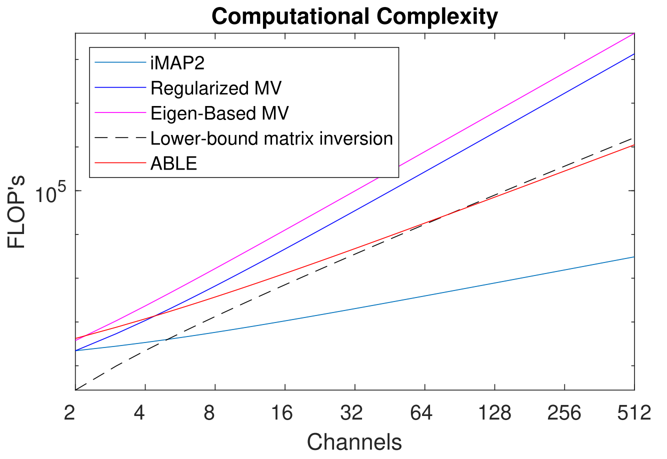

Fig. 10. Number of floating-point-operations (FLOPs) for the evaluated beamforming methods as a function of the number of receiving input channels.

for regularized MV, EBMV, iMAP2 and ABLE is plotted as a function of the input array size. Here we see that the iMAP beamformer provides a low complexity alternative compared to beamforming methods relying on the prediction of content-adaptive signal weights, however at the cost of reduced image quality. One can observe that for all array sizes, our deep learning approach provides a lower computational cost than the MV based methods relying on matrix inversion. While there are several methods to increase the efficiency of matrix inversion [30], [31], the best proven lower bound has indicated a minimum computational cost of $N^2 \log(N)$ [32], also shown in Fig. 10. This indicates that the deep learning approach severely reduces computational burden, in particular for a large number of transducer elements. Arrays consisting of more than 64 elements are very common for ultrasound imaging, whereas 3D ultrasound systems can contain over a 1000 elements.

In this work, ABLE was applied pixel-wise, i.e. its spatial receptive field was only 1 pixel. This facilitates generalization of our neural-network-based beamformer by both limiting the number of free model parameters and drastically increasing the number of available training data: each pixel in the image becomes a training sample. At the same time this restricts the network from leveraging any knowledge about the location and context around that pixel. While this also holds for MV beamforming, it may hamper its ability to suppress artifacts such as grating lobes. This leaves ample room for future development in the direction of more complex and spatially aware deep-learning-based adaptive beamformers, in which pixel locations or neighboring pixels are incorporated in the adaptive parameter prediction.

While in this work, we focus on real valued RF signals, we foresee that ABLE is also applicable for demodulated signals, operating on in-phase and quadrature (IQ) data, instead of real valued RF data. Operating on IQ data is common for MV beamforming and allows for reduced data rates. At the same time the transition to complex apodization weights provides an additional degree of freedom, by allowing phase manipulations.

Other than adaptive beamforming, there is a wide range of other applications that require the estimation of content-adaptive signal processing parameters, such as echo cancellation, source localization, directional communication, and linear prediction. One specific example is adaptive spectral Doppler [33], in which a set of optimal matched filter coefficients is calculated for high-quality spectral estimation. Due to their similarity, we expect such applications to similarly benefit from exploiting model-based, end-to-end trained, deep learning structures.

VII. CONCLUSION

In this work, we demonstrated how deep learning can be used to improve upon conventional beamforming methods. Specifically, we show that a compact and model-based architecture, which we term ABLE (adaptive beamforming by deep learning), enables the reconstruction of high-quality ultrasound images for multiple imaging systems. Its resolution and contrast are up to par with a complex state-of-the-art adaptive beamformer, yet at a substantially lower reconstruction time and computational burden. This paves the way towards a real-time implementation of adaptive beamforming in ultrasound systems.

REFERENCES

[1] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[2] I. K. Holfort, "Adaptive beamforming for medical ultrasound imaging," Ph.D. dissertation, Dept. Elect. Eng., Tech. Univ. Denmark, Lyngby, Denmark, 2009.

[3] K. Kim, S. Park, J. Kim, S.-B. Park, and M. Bae, "A fast minimum variance beamforming method using principal component analysis," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 61, no. 6, pp. 930–945, Jun. 2014.

[4] M. Bae, S. B. Park, and S. J. Kwon, "Fast minimum variance beamforming based on legendre polynomials," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 63, no. 9, pp. 1422–1431, Sep. 2016.

[5] R. Mallart and M. Fink, "Adaptive focusing in scattering media through sound–speed inhomogeneities: The van cittert zernike approach and focusing criterion," *J. Acoust. Soc. Amer.*, vol. 96, no. 6, pp. 3721–3732, Dec. 1994, doi: 10.1121/1.410562.

[6] C.-I.-C. Nilsen and S. Holm, "Wiener beamforming and the coherence factor in ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 57, no. 6, pp. 1329–1346, Jun. 2010.

[7] T. Chernyakova, D. Cohen, M. Shoham, and Y. C. Eldar, "IMAP beamforming for high-quality high frame rate imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 12, pp. 1830–1844, Dec. 2019.

[8] R. Cohen and Y. C. Eldar, "Sparse convolutional beamforming for ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 12, pp. 2390–2406, Dec. 2018.

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

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Stateline, NV, USA, Dec. 2012, pp. 1097–1105.

[11] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," 2017, *arXiv:1704.06857*. [Online]. Available: http://arxiv.org/abs/1704.06857

[12] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[13] 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 Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[14] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1289–1296, Jun. 2018.

[15] R. JG van Sloun, R. Cohen, and Y. C Eldar, "Deep learning in ultrasound imaging," 2019, *arXiv:1907.02994*. [Online]. Available: http://arxiv.org/abs/1907.02994

[16] Y. H. Yoon and J. C. Ye, "Deep learning for accelerated ultrasound imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 6673–6676.

[17] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2010–2021, Sep. 2018.

[18] M. Gasse, F. Millioz, E. Roux, D. Garcia, H. Liebgott, and D. Friboulet, "High-quality plane wave compounding using convolutional neural networks," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 64, no. 10, pp. 1637–1639, Oct. 2017.

[19] R. J. G. van Sloun *et al.*, "Super-resolution ultrasound localization microscopy through deep learning," 2018, *arXiv:1804.07661*. [Online]. Available: http://arxiv.org/abs/1804.07661

[20] A. A. Nair, T. D. Tran, A. Reiter, and M. A. Lediju Bell, "A deep learning based alternative to beamforming ultrasound images," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 3359–3363.

[21] S. Khan, J. Huh, and J. C. Ye, "Deep learning-based universal beamformer for ultrasound imaging," in *Proc. Med. Image Comput. Comput. Assist. Intervent.*, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds. Cham, Switzerland: Springer, 2019, pp. 619–627.

[22] B. Luijten *et al.*, "Deep learning for fast adaptive beamforming," in *Proc. ICASSP - IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1333–1337.

[23] A. M. Deylami, J. A. Jensen, and B. M. Asl, "An improved minimum variance beamforming applied to plane-wave imaging in medical ultrasound," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2016, pp. 1–4.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[25] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Int. Conf. Artif. Intell. Statist.*, May 2010, pp. 249–256.

[26] J. A. Jensen, "Field: A program for simulating ultrasound systems," *Med. Biol. Eng. Comput.*, vol. 4, no. 1, pp. 351–353, 1996.

[27] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. A. Jensen, and O. Bernard, "Plane-wave imaging challenge in medical ultrasound," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2016, pp. 1–4.

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

[29] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, U.K.: Cambridge Univ. Press, 2004, pp. 661–683.

[30] V. Strassen, "Gaussian elimination is not optimal," *Numerische Math.*, vol. 13, no. 4, pp. 354–356, Aug. 1969, doi: 10.1007/BF02165411.

[31] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions," in *Proc. 19th Annu. ACM Symp. Theory Comput.*, New York, NY, USA, 1987, pp. 1–6, doi: 10.1145/28395.28396.

[32] R. Raz, "On the complexity of matrix product," in *Proc. 34th Annu. ACM Symp. Theory Comput.*, New York, NY, USA, 2002, pp. 144–151, doi: 10.1145/509907.509932.

[33] F. Gran, A. Jakobsson, and J. A. Jensen, "Adaptive spectral Doppler estimation," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 56, no. 4, pp. 700–714, Apr. 2009.