Improving Lateral Resolution in 3-D Imaging With Micro-beamforming Through Adaptive Beamforming by Deep Learning

Ossenkoppele, Boudewine W.; Luijten, Ben; Bera, Deep; de Jong, Nico; Verweij, Martin D.; van Sloun, Ruud J.G.

DOI
10.1016/j.ultrasmedbio.2022.08.017

Publication date
2023

Document Version
Final published version

Published in
Ultrasound in Medicine and Biology

Ultrasound in Med. & Biol., Vol. 49, No. 1, pp. 237–255, 2023

0301-5629/$ - see front matter

https://doi.org/10.1016/j.ultrasmedbio.2022.08.017

● *Original Contribution*

IMPROVING LATERAL RESOLUTION IN 3-D IMAGING WITH MICRO-BEAMFORMING THROUGH ADAPTIVE BEAMFORMING BY DEEP LEARNING

Boudewine W. Ossenkoppele,* Ben Luijten,† Deep Bera,‡ Nico de Jong,*,§ Martin D. Verweij,*,§ and Ruud J.G. van Sloun†,¶

* Department of Imaging Physics, Delft University of Technology, Delft, The Netherlands; † Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands; ‡ Philips Research, Bangalore, India; § Department of Cardiology, Erasmus MC Rotterdam, Rotterdam, The Netherlands; and ¶ Philips Research, Eindhoven, The Netherlands

(*Received* 16 *February* 2022; *revised* 26 *July* 2022; *in final form* 28 *August* 2022)

Abstract—There is an increased desire for miniature ultrasound probes with small apertures to provide volumetric images at high frame rates for in-body applications. Satisfying these increased requirements makes simultaneous achievement of a good lateral resolution a challenge. As micro-beamforming is often employed to reduce data rate and cable count to acceptable levels, receive processing methods that try to improve spatial resolution will have to compensate the introduced reduction in focusing. Existing beamformers do not realize sufficient improvement and/or have a computational cost that prohibits their use. Here we propose the use of adaptive beamforming by deep learning (ABLE) in combination with training targets generated by a large aperture array, which inherently has better lateral resolution. In addition, we modify ABLE to extend its receptive field across multiple voxels. We illustrate that this method improves lateral resolution both quantitatively and qualitatively, such that image quality is improved compared with that achieved by existing delay-and-sum, coherence factor, filtered-delay-multiplication-and-sum and Eigen-based minimum variance beamformers. We found that only in silica data are required to train the network, making the method easily implementable in practice. (E-mail: b.w.ossenkoppele@tudelft.nl)

Key Words: **Volumetric imaging, Deep learning, Adaptive beamforming, Matrix transducers, Micro-beamforming.**

INTRODUCTION

Fast volumetric ultrasound imaging has become possible through the application of 2-D ultrasound arrays. Having ultrasound elements in two dimensions enables receive focusing in the azimuth and elevational directions as well as application of wide unfocused transmit fields, such that few transmissions are needed to insonify a large field of view (Shattuck et al. 1984; Provost et al. 2014). However, the use of unfocused transmit beams comes at the expense of image contrast and reduces spatial resolution. This can be mitigated by coherent compounding of multiple-angled unfocused transmissions or by use of multiple narrower diverging waves. However, both come at the expense of the desired high frame rate.

Address correspondence to: Boudewine Ossenkoppele, TU Delft, Imaging Physics, Lorentzweg 1, 2628 CJ Delft, The Netherlands. E-mail: b.w.ossenkoppele@tudelft.nl

Another concern in fast volumetric ultrasound imaging is that access to all transducer elements is usually not possible in matrix transducers because of the very large numbers of elements they need to consist of to allow full spatial sampling both in azimuth and in elevation. Transporting the data from all these elements to the back-end system becomes infeasible because of limitations in the number and bandwidth of cable connections. Sparse arrays, in which only a limited number of elements in a 2-D array are connected, have been proposed as a solution. However, reducing the number of active elements reduces image resolution, contrast and signal-to-noise ratio (SNR) with respect to fully populated matrix arrays. The cable connection and data rate problem of matrix transducers can be solved by employing micro-beamforming on in-probe application specific integrated circuits (ASICs) at the cost of image quality (Larson 1993; Savord and Solomon 2003; Blaak et al.

2009; Santos et al. 2016). This method reduces the number of channels by applying a set of pre-determined analogue delays to a sub-array of neighboring elements and subsequently summing the steered sub-array data into a single output. Data rate and channel reduction become an especially important concern for small-aperture devices that can be used from within the body.

In-body ultrasound transducers allow visualization of parts of the body where ultrasound imaging would otherwise be hampered by attenuation, aberration and possible shadowing of the overlying tissues. For instance, transesophageal echocardiography (TEE) probes are mounted at the tip of a gastric tube such that images can be made from the esophagus where the probe is located within millimeters of the heart (Seward et al. 1988). For patient safety and comfort, smaller TEE probes are desired (Spencer et al. 1997; Hastie et al. 2019), which constrains the size of the ultrasound array. For intracardiac echography (ICE) (Hijazi et al. 2009; Wildes et al. 2016), in which a catheter-based device is positioned inside the heart during interventional cardiology, the need for compact ultrasound arrays is especially apparent. Reducing the size of the imaging aperture, however, worsens the lateral resolution and consequently the image quality.

Satisfying the requirement of a small aperture, large field of view, volumetric image and high frame rate simultaneously results in trade-offs in the imaging transmit−receive scheme that sacrifice spatial resolution. As a result, the burden to achieve sufficient spatial resolution for in-body high-frame-rate 3-D ultrasound devices falls on the receive processing branch of the imaging chain.

The core step of the ultrasound image formation is beamforming, which transforms the received time-domain signals into an estimation of the acoustic reflectivity in the spatial domain. To improve image quality beyond the traditional low-complexity delay-and-sum (DAS) beamformer, various adaptive beamformers have been developed. the coherence factor (CF)−based beamformer (Mallart and Fink 1994) reduces image clutter by reducing clutter from phase aberrations and has low computational complexity. Unfortunately, it suffers from artifacts in low-SNR imaging scenarios (Nilsen and Holm 2010) and a degraded speckle pattern. The filtered delay multiply and sum (F-DMAS) beamformer (Matrone et al. 2015) involves a combinatorial coupling and multiplication of the received signals before summation. Prieur et al. (2018) reported that the increased sensitivity to coherence of the F-DMAS beamformer contributes to improvements in contrast ratio and the depiction of coherent structures, but also comes with the drawbacks of grainier appearance and dark regions that can appear around coherent structures.

The minimum variance (MV) beamformer improves resolution by using a data-adaptive apodization of the aperture that minimizes the output energy of the signal while constraining the response to have unit gain at the focal point of the receive beam (Synnevag et al. 2009). Further improving contrast without compromising the achieved resolution is possible with the Eigen-based minimum variance beamformer (EBMV) (Asl and Mahloojifar 2010). However, there are a number of drawbacks to the application of MV beamforming. First, the high computational burden resulting from the need to perform matrix inversion, and in the case of the EBMV beamformer also eigendecomposition, for every image pixel. This computational burden already limits practical application of EBMV-based methods for fast 2-D imaging. For fast 3-D imaging, where many more pixels need to be reconstructed to form a single frame, this becomes an even larger burden. Second, empirical tuning of parameters, such as diagonal loading and sub-aperture averaging, is needed to achieve a good result across different imaging scenarios (Lorenz and Boyd 2005). Third, images reconstructed with EBMV can suffer from dark region artifacts next to hyperechoic structures (Rindal et al. 2017). Finally, the performance of (EB)MV beamforming has not been examined yet in combination with micro-beamforming.

Recently, deep learning techniques have also been employed to improve ultrasound image quality (van Sloun et al. 2019, 2021). Some have applied a neural network to transform an already beamformed gray-scale ultrasound image of low quality to a high-quality image (Dietrichson et al. 2018; Zhou et al. 2018, 2020; Hyun et al. 2019; Goudarzi et al. 2020; Lu et al. 2020). The downside of using images as input data to the neural network is that a lot of the acquired information present in the radiofrequency (RF) data have been discarded in the image formation process. As a result, the data are not available to the neural network. Therefore, others have avoided this by implementing deep neural networks not after image formation, but to replace the beamformer partially (Luchies and Byram 2018, 2020; Yoon and Ye 2018; Luijten et al. 2020; Vignon et al. 2020) or as a whole (Nair et al. 2020) . Replacing the complete beamforming process with a neural network (Nair et al. 2018, 2020) means that the network also has to learn the geometric time-of-flight (TOF) correction. Therefore, many have instead implemented a neural network that replaces part of the beamforming process after the time-to-space migration (Yoon and Ye 2018; Luchies and Byram 2018, 2019, 2020; Luijten et al. 2020; Vignon et al. 2020; Huang et al. 2021; Mamistvalov et al. 2022).

In 2-D imaging, deep learning solutions that replace part of the beamforming process after TOF correction achieved improved resolution, while being data efficient

in training and computationally efficient in reconstructing unseen ultrasound data (Luijten et al. 2020; Vignon et al. 2020; Zhou et al. 2021). The adaptive beamforming by both the deep learning (ABLE) method (Luijten et al. 2020) and the delay and neural network (DANN) method (Vignon et al. 2020) employs a neural network on the channel data after time-of-flight correction, but replaces slightly different parts of the beamforming process. The ABLE method explicitly calculates pixelwise data-adaptive apodization weights from the channel data of that pixel. The DANN method instead uses convolutional layers to directly transform the TOF-corrected RF data into an axial line of pixel values, thereby replacing the beam-summing step. The method developed by Luchies and Byram (2018) also replaces the beam-summing step, but operates on the frequency domain instead. A frequency-specific deep neural network is trained for each discrete Fourier transform bin. Subsequently, data are transformed back to the time domain, and the filtered signals are summed along the aperture channels to form a beamformed image. Zhou et al. (2021) sought to maintain the advantage of image-based methods that are able to capture global features, as well as the advantages of access to full RF data. They combined a neural network that learns apodization weights with a neural network that works on an image-to-image basis in a hybrid approach. Their method is computationally efficient by reducing the needed pre- and post-processing steps such as envelope detection, but it loses some of the interpretability of ABLE where a generated pixel value can be directly linked to the angular response (beampattern) formed by the generated apodization weights.

Other approaches have implemented a neural network after combination of the channel data, but before the final image formation steps of envelope detection and log compression. These include methods that combine data from multiple transmissions, thereby replacing the compounding process (Chennakeshava et al. 2020; Lu et al. 2020). Applying deep learning in the beamforming process inherently requires access to the RF data, but gives the neural network access to a rich set of information that is not available when working on beamformed images. Methods that replace only part of the beamforming process, after time-to-space migration, require smaller amounts of training data and less trainable parameters.

Training a deep learning–based beamformer in a supervised fashion requires target data. In some scenarios, in vivo data can be used as part of the training set, for example, when deep learning is used to obtain the same high image quality as an existing beamformer at improved inference speeds or when the goal is to achieve the image quality of a full acquisition with an acquisition setup that is compressed or undersampled in some way. EBMV beamforming has been used as a method to obtain high-quality target images in 2-D imaging (Luijten et al. 2020; Zhou et al. 2021). In other scenarios, the training data consisted purely of simulations. One example includes training data where the target is based directly on the properties of the simulated medium (Nair et al. 2018; Youn et al. 2020). Other examples include a training target created by an ultrasound transducer with more desirable properties, such as a larger aperture (Vignon et al. 2020), or higher frequency without increased attenuation (Chennakeshava et al. 2020). An overview of deep learning methods for ultrasound image reconstruction more complete than described here can be found in van Sloun et al. (2021).

The ability of deep learning–based beamformers to improve resolution has not been observed for 3-D ultrasound imaging with matrix transducers in general or in combination with micro-beamforming in particular. This is not a trivial extension, as micro-beamforming prevents direct access to signals from all transducer elements by the back-end system. As a result, the beamformer must work with the lower spatial sampling and focusing errors contained in the micro-beamformed signals. Inspired by the improved image quality in 2-D imaging we hypothesize here that the lateral resolution of volumetric ultrasound images, acquired by on-chip micro-beamforming, can be improved by using a deep learning–based beamformer trained on target images formed by a larger aperture. As little research has applied (EB)MV beamforming in 3-D imaging with matrix transducers in general (Avanji et al. 2013) and for micro-beamformed data in particular, we chose to use training targets generated by a larger aperture and thereby avoid the limitations of existing adaptive beamformers. The ABLE beamformer is used as a starting point as it offers a data-efficient deep learning–based solution to beamforming. ABLE operates on a per-pixel basis. However, we hypothesize that the pixel-based approach might be suboptimal for handling the focusing errors present in micro-beamformed data and evaluate the effect of increasing the receptive field of the neural network to include the channel data of neighboring voxels. We moreover describe the results achieved by using only simulation data to train the neural network and compare these with the results achieved when the training set also includes in vitro and in vivo data.

METHODS

Network architecture

The deep learning–based adaptive beamformer (ABLE) developed by Luijten et al. (2020) is used as a basis for the neural network architecture used in this work. Thus, a neural network f_θ with a small number of

layers is used to calculate apodization weights for each voxel from the time-of-flight-corrected channel data. We maintain the per-pixel operation, where a beam-formed pixel value $V_{ABLE k,l,m}$ is obtained by per-pixel multiplication of the apodization weights with the input data. However, we extend the receptive field of the network spatially, from the TOF-corrected RF data of a single voxel $\left(\mathbf{Y} \in \mathbb{R}^{1\times C}\right)$ to a tensor that contains the RF data of a small group of voxels $\left(\mathbf{Y} \in \mathbb{R}^{Az\times El\times Ax\times C}\right)$:

$$V_{ABLE k,l,m} = \sum_{c=1}^{C}\left[f_\theta\left(Y_{k:k+Az-1,l:l+El-1,m:m+Ax-1}\right)Y_{k,l,m,c}\right] \quad (1)$$

$$V_{ABLE k,l,m} = \sum_{c=1}^{C}\left[f_\theta(\mathbf{Y})_{k,l,m,c}\, Y_{k,l,m,c}\right].$$

Here Az, El and Ax are the receptive fields of the neural network in the azimuth, elevation and axial directions, respectively, denoting the size of the group of neighboring voxels whose TOF-corrected RF data influence the computed apodization weights for a single voxel. We increase the receptive field because micro-beamformed data are sub-optimally focused. Even when the micro- and post-beamforming delays are applied perfectly, errors occur in TOF correction, as focusing has a range-dependent component, which can no longer be implemented accurately for each individual element after applying the static micro-beamforming delays and the summing operation. For some transducer elements, a slightly later or earlier sampling of the raw RF data is more accurate. The TOF-corrected data of neighboring pixels consist of raw RF data that have been sampled with slightly larger or smaller delays because of the slightly different distances between these voxels and the transducer elements. Therefore, we consider that the network might make use of the increased information that is available from the TOF data of a broader region of voxels to generate a better estimation of apodization weights.

The receptive field is extended by using convolutional layers. Each layer contains a number of 3-D convolutional kernels, as illustrated in Figure 1, which work along the spatial dimensions. The network remains fully connected along the channel dimension, across which the encoder–decoder structure is preserved. Zero padding is used to avoid compression in the spatial dimensions.

We evaluated the effect of receptive field size using four networks with different receptive field sizes, but the same number of layers. ABLE1 has a single voxel receptive field, similar to the original ABLE network. For the other three networks, we chose an axial receptive field size that extends 14 voxels (about 1.75 wavelengths). To see the impact of increasing the receptive field in the lateral direction, we set the azimuthal and elevational receptive fields of ABLE2, ABLE3 and ABLE4 to 2, 3

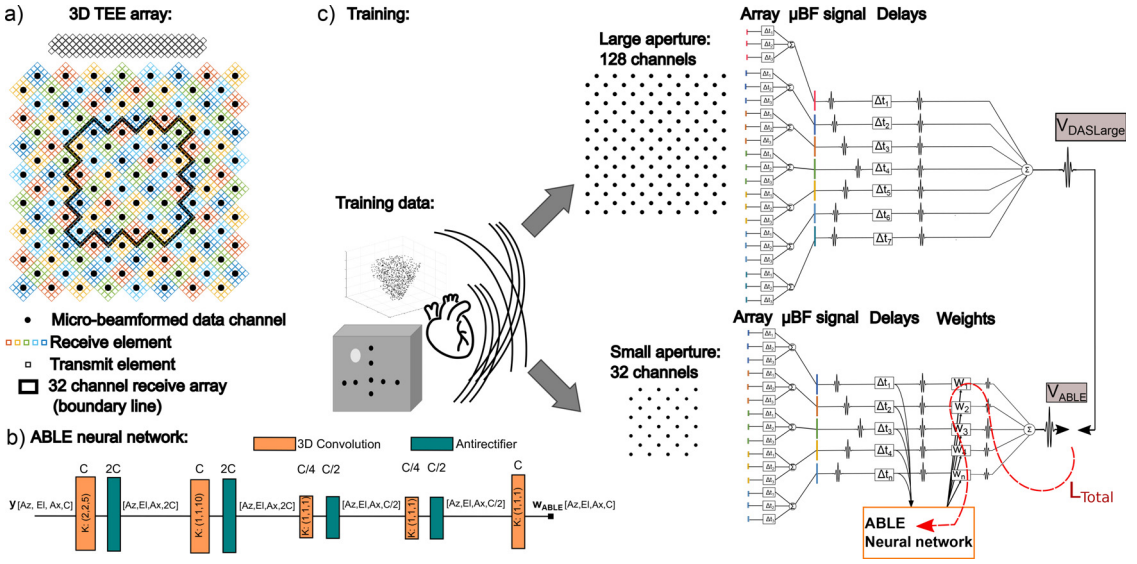

Fig. 1. (a) Layout of the 3-D TEE probe with split transmit–receive design. The large receive aperture of 2048 elements is combined into 128 data channels through micro-beamforming with 4 × 4 sub-arrays. (b) The modified adaptive beamforming by deep learning (ABLE) network consists of 3-D convolutional layers and antirectifier activation functions. The number of channels C = 32. The kernel size K is indicated in the convolutional layers. (c) Radiofrequency data of in silico, in vitro and/or in vivo data are acquired. All 128 channels are used to delay-and-sum (DAS) beamform target images. Time-of-flight-corrected RF data from the 32 center channels are used as input to the ABLE neural network. The network is trained by backpropagation based on the loss value between V_{ABLE} and the target $V_{DASLarge}$.

and 6 voxels, respectively (where one voxel has a width of 0.5°). The architecture of ABLE2 is illustrated in Figure 1 and all architectures are shown in the Appendix.

Identical to the original ABLE (Luijten et al. 2020), the antirectifier is chosen as the activation function:

$$g(\mathbf{z}) = \begin{bmatrix} max\left(0, \dfrac{\boldsymbol{z}-\mu_z}{\|\boldsymbol{z}-\mu_z\|_2}\right) \\ max\left(0, -\dfrac{\boldsymbol{z}-\mu_z}{\|\boldsymbol{z}-\mu_z\|_2}\right) \end{bmatrix} \tag{2}$$

as it prevents loss of the negative signal components, by concatenating the positive and negative components of the signal after l_2-normalization and then returning their absolute value. At the same time, it maintains the favorable properties of computational efficiency and resistance against vanishing gradients that ReLus have, and as the activation function is not bounded, it helps preserve the dynamic range of the input. Inspired by the lack of summation terms in the calculation of apodization weights by the minimum variance beamformer, we have not used bias terms in our network.

Training setup

We used supervised learning, with high-resolution training targets that were generated by DAS beamforming with a larger aperture array V_{DASLarge}. The loss function $\mathrm{L_{total}}$ consists of a weighted combination of an image loss $\mathrm{L_{SMSLE}}$ based on the pixel values and a loss based on the apodization weights $\mathrm{L_{unity}}$:

$$\begin{aligned} L_{total} = \frac{1}{I}\sum_{i=1}^{I} &\lambda L_{\mathrm{SMSLE}}\left(V_{\mathrm{ABLE}}^{(i)}, V_{\mathrm{DASLarge}}^{(i)}\right) \\ &+ (1-\lambda) L_{\mathrm{unity}}\left(V_{\mathrm{ABLE}}^{(i)}, V_{\mathrm{DASLarge}}^{(i)}\right) \end{aligned} \tag{3}$$

Here $V_{\mathrm{ABLE}}^{(i)}$ and $V_{\mathrm{DASLarge}}^{(i)}$ refer to respectively the network outputs and targets for a voxel i. To take into account the large dynamic range and both positive and negative components of the ultrasound signal, the image loss is computed as the signed-mean-squared-logarithmic error (SMSLE) between a target voxel (V_{DASLarge}) and the voxel calculated by ABLE (V_{ABLE}):

$$\begin{aligned} L_{\mathrm{SMSLE}}^{(i)} = &\\ &\frac{1}{2}\left\| \log_{10}\left(V_{\mathrm{ABLE}}^{(i)}\right)^{+} - \log_{10}\left(V_{\mathrm{DASLarge}}^{(i)}\right)^{+} \right\|_2^2 \\ &+\frac{1}{2}\left\| \log_{10}\left(V_{\mathrm{ABLE}}^{(i)}\right)^{-} - \log_{10}\left(V_{\mathrm{DASLarge}}^{(i)}\right)^{-} \right\|_2^2, \end{aligned} \tag{4}$$

where $(\cdot)^+$ and $(\cdot)^-$ denote the magnitudes of a positive and a negative value, respectively. At the same time, we promote a distortionless response with the $\mathrm{L_{unity}}$ loss term:

$$L_{\mathrm{unity}}^{(i)} = \left| \mathbf{1}^T \boldsymbol{w}_{\mathrm{ABLE}}^{(i)} - 1 \right|^2 \tag{5}$$

Stochastic optimization of eqn (3) was done using the Adam optimizer with a learning rate of $\alpha = 10^{-3}$. The exponential decay rates for the first and second moment, $\beta_1 = 0.9$ and $\beta_2 = 0.999$, were set according to the values suggested by Kingma and Ba (2015), and the constant for numerical stability was $\hat{\epsilon} = 10^{-7}$. All networks were trained for 400 epochs and evaluated based on the snapshot ensemble of the last five epochs. The network was implemented using Keras API with a Tensorflow (Google, Menlo Park, CA, USA) backend.

Training data acquisition

Data were acquired with a miniature transesophageal echocardiography (TEE) prototype transducer, described in detail in Bera et al. (2018). The full array consists of 2176 PZT elements, with a pitch of 181 μm and a 30 μm kerf, which was cut at a 45° angle with respect to the probe's center line. The aperture is split in a narrow transmit array of 128 elements, which are directly wired out to the external ultrasound system, and a larger receive array of 2048 elements (see Fig. 1). The signals of the receive array are micro-beamformed by the front-end ASIC in sub-arrays of 4 × 4 elements. The full receive aperture, consisting of 128 micro-beamformed data channels, was used to acquire the large-aperture target data, while the center 32 micro-beamformed data channels (see Fig. 1) were used to acquire the small-aperture input data from the same transmit firing made by the narrow transmit array.

Each 60° × 60° volume was acquired with 85 steered transmit−receive events. A wide transmit beam, produced from a virtual source located 100 mm behind the transducer, was steered to a combination of one of the 17 equally spaced azimuth angles between −24° and 24° degrees and one of the 5 equally spaced elevation directions between −20° and 20°. Micro-beamformed data sets were acquired by pre-steering the sub-arrays of the receive aperture to the transmission direction.

In vivo and in vitro

Micro-beamformed data sets were acquired with the TEE probe and a Verasonics ultrasound acquisition system, which sampled the data at 20 MHz. A commercial tissue phantom (multi-purpose multi-tissue ultrasound phantom 040-GSE, CIRS, Norfolk, VA, USA) was used to acquire images of wire targets, hyper- and hypoechoic cysts. In vivo images of the heart of an anesthetized adult pig were acquired through a hole in the chest wall and the diaphragm by an experienced cardiologist, as described in more detail in Bera et al. (2018). (This

experiment was approved by Erasmus MC Animal Experiments Committee Protocol 109-14-12.)

In silico

Simulated RF data were acquired by implementing the transducer geometry and imaging scheme in Field II (Jensen and Svendsen 1992; Jensen 1996). Two types of simulation phantoms were used. In the first, point scatterers were randomly distributed, with either 500 or 1000 scatterers in a volume. In the second type, cubes and spheres with fully developed speckle (10 scatterers per resolution cell) and varying average amplitude (0−60 dB) and size (2−6 mm) were distributed against a hypo-echoic background.

Training data preparation

All in silico, in vitro and in vivo data were split into training, validation and test data. Of the recorded porcine heart data, volumes acquired at different stages of the cardiac cycle were used in the training, validation and test data. Of the in vitro data, the volumes used in the training, validation and test data were acquired at different non-overlapping positions on the CIRS phantom. Of the in silico data, the validation volumes contained different random scatterer locations compared with the training volumes. The volume used in the test set was constructed to have a more structured pattern of scatterers than that in the training and validation data, to facilitate the evaluation of ultrasound imaging-related metrics. A single acquired volume of (TOF-corrected) RF data provides a large amount of training data for the network, as the loss is calculated per individual pixel and the receptive fields of the used neural networks are small. As a result, only a small number of volumetric acquisitions were needed to realize a sufficient amount of training data for the network. The training set consisted of five training volumes with a 60° × 60° opening angle and a depth varying between 4.6 and 8.0 cm. Three in silico training volumes were used in the first training set. In a second training set, an in vitro volume was added, and in a third training set, an in vivo volume was added as well. When training on a combination of in silico, in vitro and/or in vivo data it was ensured that there was an equal probability of sampling training data patches from each of these data types. For each training set, validation data were of the same types (in silico, in vitro, in vivo) as the training data. Each batch consisted of 16 randomly selected patches of training data. A patch consisted of an imaging region of 10 × 10 × 600 target pixels created by DAS beamforming, with the large array and the corresponding input TOF-corrected RF data recorded with the small array (input size single patch: 10 × 10 × 600 × 32). All values were pre-normalized with respect to the maximum of the volumetric image of which they are a part, such that all values are between −1 and +1.

Comparison with other beamformers

We compared the ABLE beamformers with the DAS beamformer as well as three adaptive beamformers: EBMV, CF and F-DMAS. For all methods, the same pre- and post-processing steps were used. Before beamforming, a bandpass filter with a 5-MHz center frequency and 100% bandwidth was applied to the acquired signals. After beamforming and envelope detection of the data from the 85 transmit−receive events, the final volume was obtained as an angle-weighted combination of overlapping sub-volumes (Bera et al. 2018). Finally, the data were log-compressed to gray-scale images with a dynamic range of 60 dB.

The first step of the beamforming process, the TOF correction, was also the same for all beamformers. On the basis of the geometric relation between the voxel location and transducer channel locations, the delay needed to focus the signal received by a channel c to an image point $\boldsymbol{r}$ is given by

$$\tau[x,y,z] = \frac{\| \boldsymbol{r}_{TX} - \boldsymbol{r} \|_2 + \| \boldsymbol{r}_c - \boldsymbol{r} \|_2}{v} \quad (6)$$

Here v is the estimated speed of sound in the medium, $\boldsymbol{r}_c$ is the location of the receive channel and $\boldsymbol{r}$ is the voxel location. $\boldsymbol{r}_{\mathrm{TX}}$ is the location of the transmit firing, which for our diverging wave transmissions is specified by the coordinates of the virtual source from which it emanates. For any focus point [x,y,z] and channel c, the delayed response, $\mathrm{y_c}$, is then given by

$$y_c[x,y,z] = x_c(\tau[x,y,z]), \quad (7)$$

where x_c is the received channel data.

Delay-and-sum

Image voxels $V_{\mathrm{DAS}}(x,\ y,\ z)$ are generated by summing across the channel dimension according to

$$V_{\mathrm{DAS}}[x,y,z] = \boldsymbol{w}_{\mathrm{DAS}}^T[x,y,z]\boldsymbol{y}[x,y,z], \quad (8)$$

where $\mathbf{y}$ is the delayed response for all channels. Here the weight vector $\mathbf{w}_{\mathrm{DAS}}$ consists of pre-determined data-independent weights, which can vary spatially. Here we used a boxcar window to emphasize resolution.

Minimum variance-based adaptive beamforming

The minimum variance distortionless response beamformer applies time-of-flight correction in the same way as the DAS beamformer, but aims to improve image resolution and contrast by replacing the static apodization weights, used in DAS, with a set of data-adaptive apodization weights. These apodization weights $\mathbf{w}_{\mathrm{MV}}$ are chosen to minimize the variance of the beam-formed signal, such that interfering signals and noises are rejected,

while maintaining a unity gain with respect to the signal from the beamformers' focusing point. The optimal apodization weights are found by minimizing

$$\min_{\boldsymbol{w}_{\mathrm{MV}}} \boldsymbol{w}_{\mathrm{MV}}^{H}\boldsymbol{R}\boldsymbol{w}_{\mathrm{MV}} \qquad s.t. \quad \boldsymbol{w}_{\mathrm{MV}}^{H}\boldsymbol{d} = 1, \tag{9}$$

which has an analytical solution (Capon 1969) given by

$$\boldsymbol{w}_{\mathrm{MV}} = \frac{\boldsymbol{R}^{-1}\boldsymbol{d}}{\boldsymbol{d}^{H}\boldsymbol{R}^{-1}\boldsymbol{d}}. \tag{10}$$

Here **d** is the so-called steering vector, which defines the signal that should be passed distortionless. For narrow-band applications this steering vector can be expressed as a complex exponential applying the required phase shifts; however, for broadband applications this is not possible and we work with signals that are already time-of-flight corrected; as such the steering vector simply becomes $\mathbf{d} = \mathbf{1}^{C}$ (Synnevag et al. 2009). **R** is the spatial covariance matrix:

$$\boldsymbol{R} = \mathbb{E}\left[\boldsymbol{y}^{h}\boldsymbol{y}\right] \tag{11}$$

In practice the spatial covariance matrix $\boldsymbol{R}[x,y,z]$ must be estimated from a single or a small number of temporal samples, as the received signals change rapidly in time. To increase the robustness of the estimation of the sample covariance matrix, we applied spatial smoothing (Evans et al. 1981) and diagonal loading as described in Synnevag et al. (2007). We used a diagonal loading of 0.001 and sub-array length of 16.

Eigen-based minimum variance beamformer (EBMV) applies an additional update step on the apodization weights to further improve contrast and resolution (Asl and Mahloojifar 2010). The weights of the EBMV beam former are calculated by taking the eigen-decomposition of **R** and projecting the signal subspace $\mathbf{E}_{\mathrm{signal}}$ on the weights determined by eqn (10):

$$\boldsymbol{w}_{\mathrm{EBMV}}[x,y,z] = \boldsymbol{E}_{\mathrm{signal}}[x,y,z]\boldsymbol{E}_{\mathrm{signal}}^{H}[x,y,z]\boldsymbol{w}_{\mathrm{MV}}[x,y,z]. \tag{12}$$

Coherence factor

The coherence factor weighting method also applies data-adaptive weights per pixel. However, in contrast to the EBMV, the calculated pixel-based weights are identical across channels. As such CF can also be viewed as a post-filter that is applied after a DAS (or another) beamformer. The CF weights are determined as the ratio between the coherent and incoherent energy received by the array (Mallart and Fink 1994):

$$w_{\mathrm{CF}}[x,y,z] = \frac{\left|\sum_{c=1}^{c=C} y[x,y,z,c]\right|^{2}}{\sum_{c=1}^{c=C}\left|y[x,y,z,c]^{2}\right|}. \tag{13}$$

Here C is the total number of channels in the aperture. The final voxel value is given by

$$V_{\mathrm{CF}}[x,y,z] = w_{\mathrm{CF}}[x,y,z]V_{\mathrm{DAS}}[x,y,z]. \tag{14}$$

Filtered delay multiply and sum

The F-DMAS beamformer applies a pairwise multiplication of the signals. The signed square root is first applied to the signal pairs to ensure a correct scaling of the final pixel value $V_{\text{F-DMAS}}$:

$$V_{\mathrm{F-DMAS}} = \sum_{i=1}^{C-1}\sum_{j=i+1}^{C} sign\left(y_i y_j\right)\sqrt{\left|y_i y_j\right|} \tag{15}$$

The resulting signal now has additional frequency components at zero and twice the center frequency; therefore, the signal is bandpass filtered around the second harmonic to remove the DC component (Matrone et al. 2015).

Algorithmic complexity

We exclude in the computational complexity calculation the part that is similar to all adaptive beamformers that were used: TOF correction, multiplication of the apodization weights and subsequent summation along the channels, envelope detection and log compression. We calculate the number of floating point operations (FLOPs) required for a single voxel, for the ABLE and EBMV beamformer. The version of the ABLE network that we used consisted of convolutional layers and antirectifiers. The number of FLOPs required for an antirectifier layer is four times the number of nodes in that layer. The computational complexity (F) of a convolutional layer depends on the number of nodes N in the layer and the number of elements in the convolutional kernel K. The total computational cost of the convolutional version of the ABLE network with L layers becomes

$$F_{\mathrm{ABLEconv}} = \sum_{L=1}^{L-1}\left[2N_{\mathrm{in}l}N_{\mathrm{out}l}K_l + 4N_{\mathrm{out}l}\right] + 2N_{\mathrm{in}_L}N_{\mathrm{out}_L}K_L \tag{16}$$

The number of nodes is linearly related to the number of channels C. Thus, the computational complexity is order $O(C^2)$.

The DAS and CF beamformer have a computational complexity of order $O(C)$ Although the computational complexity of the F-DMAS beamformer as stated in eqn (14) is of order $O(C^2)$ it can be reformulated such that it also reduces to order O(C) (Ramalli et al. 2017). The components dominating the computational complexity of the EBMV beamformer are the inversion and

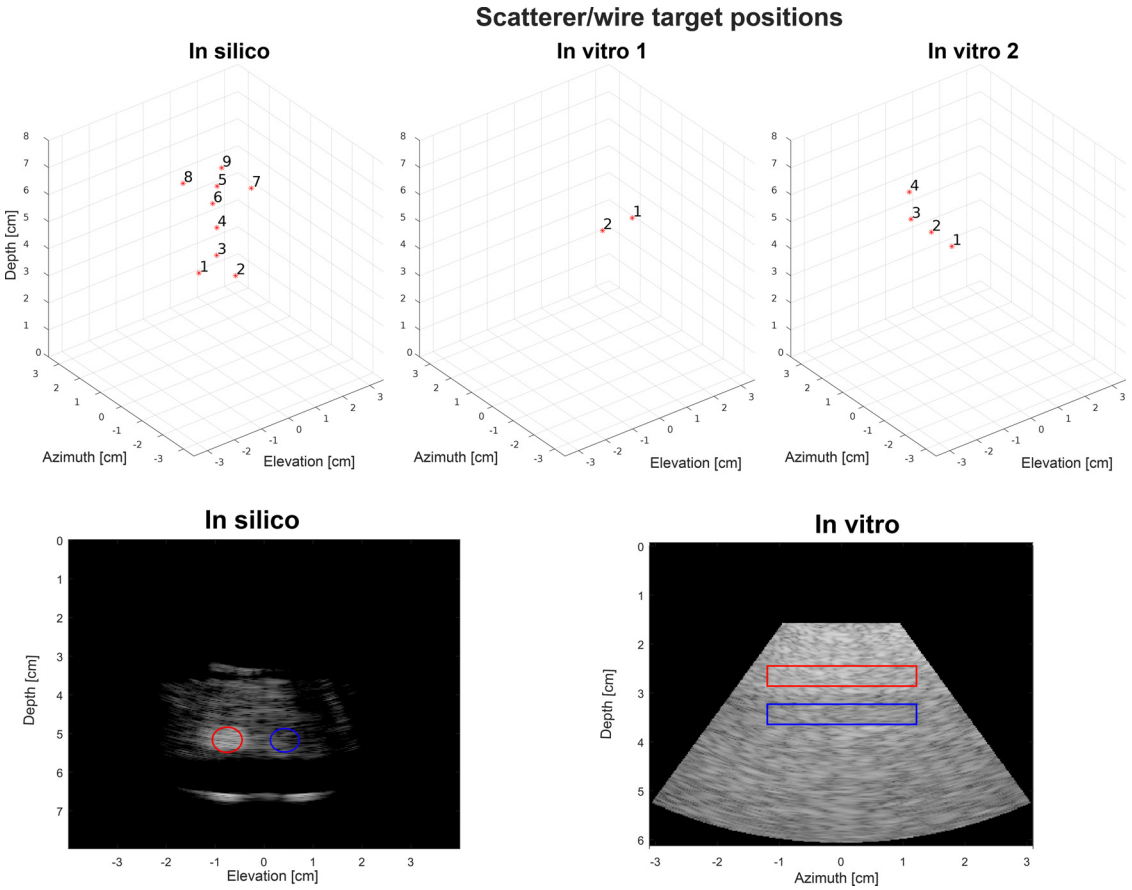

Fig. 2. (a) The nine scatterer locations used to evaluate the in silico resolution. (b) The six scatterer locations used to evaluate in vivo resolution. (c) The high- and low-intensity regions used to evaluate contrast metrics.

eigendecomposition of the covariance matrix R, which both use approximately C^3 FLOPS (Lorenz and Boyd 2005), making the computational complexity order O (C^3). The computational complexity of the EBMV beamfomer is calculated as described in Luijten et al. (2020).

Evaluation criteria

The full width at half-maximum (FWHM) of simulated point scatterers and in vitro wire targets (locations are illustrated in Fig. 2) is used to evaluate resolution in elevation and azimuth. Contrast is evaluated with the contrast ratio (CR) and contrast-to-noise ratio (CNR), as well as the generalized contrast-to-noise ratio (GCNR). A high-intensity region and low-intensity region (see Fig. 2) are selected in the in silico and in vitro images for evaluation. The CR and CNR are calculated as

$$\mathrm{CR} = \frac{|\mu_L - \mu_H|}{(\mu_L + \mu_H)/2} \tag{17}$$

$$\mathrm{CNR} = \frac{|\mu_L - \mu_H|}{\sqrt{\sigma_L^2 + \sigma_H^2}}, \tag{18}$$

where μ_L and μ_H are the mean voxel value of the low- and high-intensity regions ($\mathrm{V_L}$, $\mathrm{V_H}$), respectively, and σ_L and σ_H are the variances of the regions.

It is possible to change CR and CNR values with trivial dynamic range alterations, without image quality improvements that improve clinical value (Rindal et al. 2019). As such, we also include the GCNR metric, which has been found to be resistant to such effects (Rodriguez-Molares et al. 2020). The GCNR determines the success rate that can be expected from an ideal observer at the task of separating the pixels from two different regions, and values are always between 0 and 1. The GCNR is calculated as

$$\mathrm{GCNR} = 1 - \sum_{n=1}^{N} \min[h_n(V_L), h_n(V_H)], \tag{19}$$

where $\mathrm{h_L}$ and $\mathrm{h_H}$ represent the histogram of the voxel values of the low- and high-intensity regions,

respectively, and N is the number of bins in the histogram. Hyun et al. (2022) reported that it is important to choose a bin size that is not too fine or too coarse to obtain a valid estimate of the GCNR. We have chosen the number of bins as the square root of the number of pixels in the low- (or high-) intensity region and examined the histograms to check that the sampling was appropriate.

RESULTS

Elevation and azimuth slices of the 3-D images that were obtained for the small array by beamforming with DAS, F-DMAS, CF, EBMV and the ABLE method trained on in silico data are illustrated in Figure 3, next to DAS beamformed images of the larger array (DAS_{Large}). Additional image views from a single in vivo volume can be seen in Figure 4. The average FWHM of the scatterers in the in silico and in vitro data are summarized in Table 1. When averaging the FWHM across in silico and in vitro data and across both lateral directions, we see that the average FWHM of ABLE is only 62% of the average FWHM of DAS, 79% that of F-DMAS and 81% that of CF. The resolution improvement of ABLE is also clear from Figure 5, which illustrates C-planes of wire targets and simulated scatterers. The average FWHM of EBMV for the point scatterers considered is 3.7˚, which is smaller than the 6.3˚ achieved by ABLE when trained on *in silico* data and the 6.6˚ of DAS_{Large}. However, if we look closer at Figure 3, we see that the points lose consistency in their shape. Furthermore, the normalized beam profiles along in vitro and in silico scatterers (see Fig. 6) reveal that use of the EBMV beamformer results in an erratic

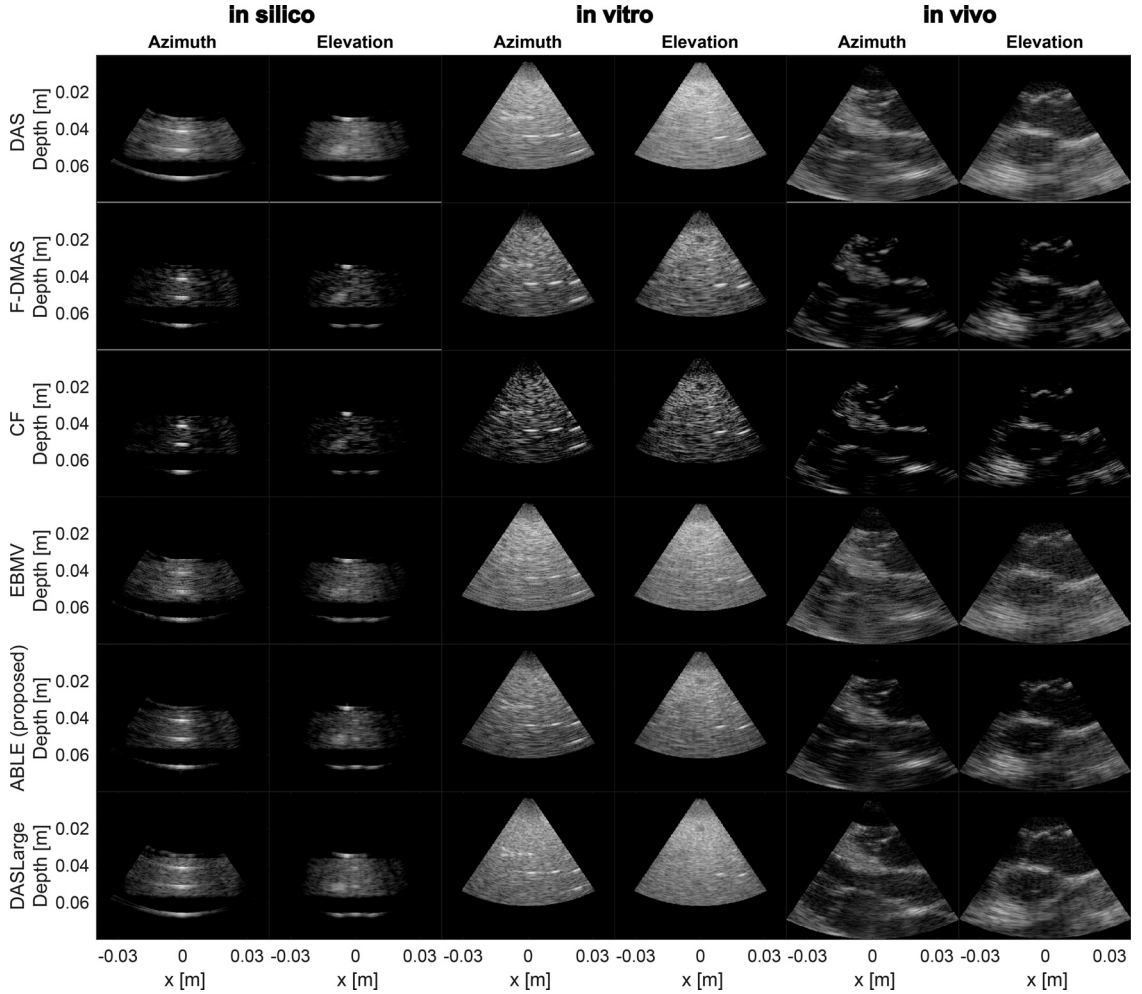

Fig. 3. Azimuth and elevation slices of an in silico phantom and in vitro phantom, as well as in vivo porcine heart data, for the delay and sum (DAS), coherence factor (CF), filtered delay multiply and sum (F-DMAS), Eigen-based minimum variance (EBMV) and proposed beamformer. In the in silico volume, simulated point scatterers are visible both in an anechoic region and surrounded by other scatterers forming an hypo-echoic region. The target DAS image formed by the larger array is shown on the last line. The images are for a dynamic range from −60 to 0 dB.

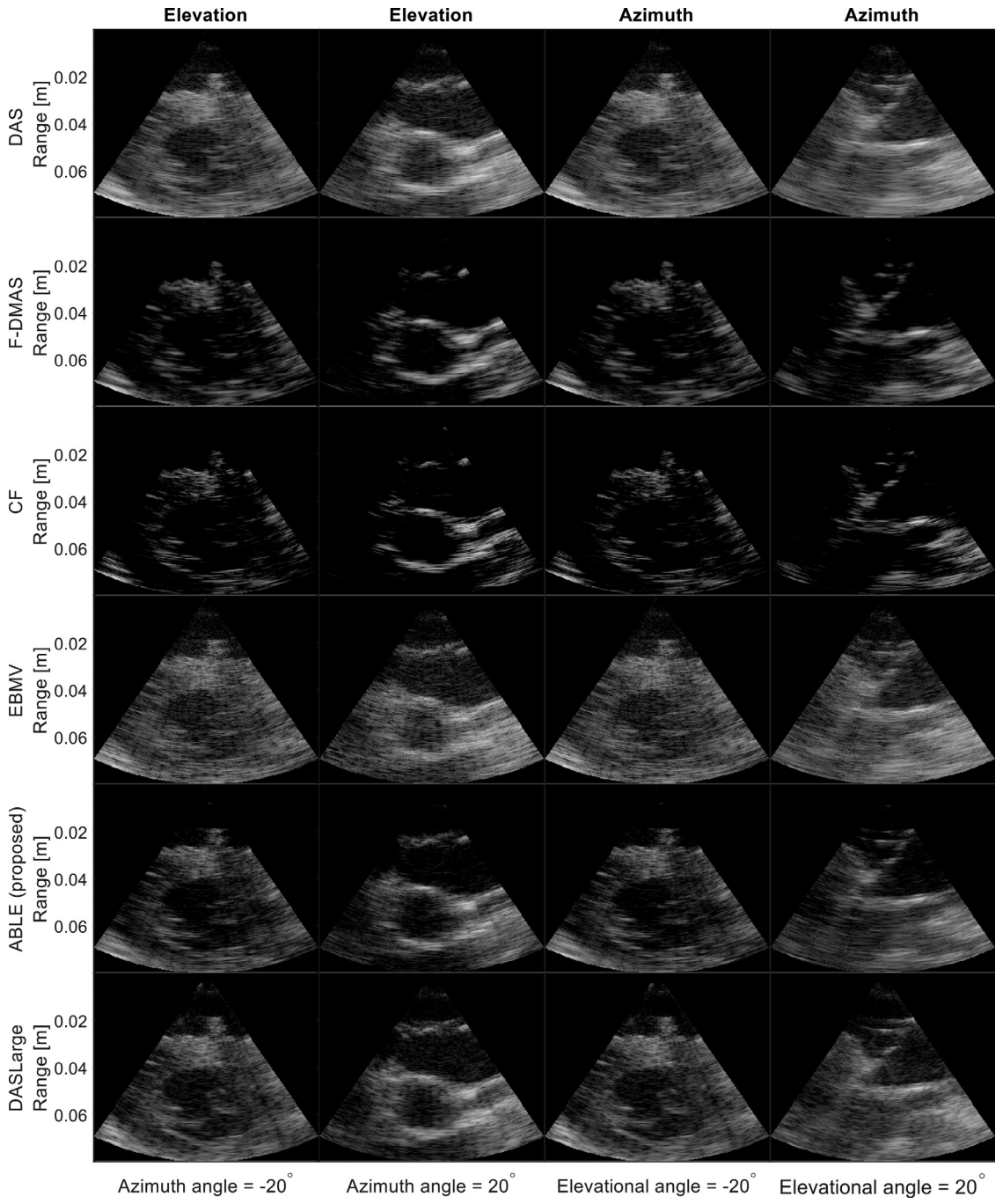

Fig. 4. Azimuth and elevation slices of an in vivo porcine heart for the delay and sum (DAS), coherence factor (CF), filtered delay multiply and sum (F-DMAS), Eigen-based minimum variance (EBMV) and proposed beamformer. The images are for a dynamic range from −60 to 0 dB.

beam profile. On the other hand, the beam profile reconstructed by ABLE reveals smooth behavior, similar to the DAS beam formers. Although ABLE also qualitatively improves resolution in the in vivo image (see Figs. 3 and 4), this is not the case for the EBMV beamformed image. Table 1 reveals that in vitro the GCNR of ABLE when trained on *in silico* data is 0.56, which is marginally lower than the GCNR of 0.59 for DAS, while in silico the GCNR of ABLE is 0.87, which is marginally higher than the 0.85 of DAS. The GCNR of F-DMAS and CF are close to those of DAS and ABLE in silico. However, on in vitro data, even though the CR is improved for F-DMAS and CF compared with DAS and ABLE, the CNR and GCNR are not.

Table 1. Image quality metrics

	DAS	F-DMAS	CF	EBMV	DAS_{Large}	ABLE (*in silico*)	ABLE (*in silico* + *vitro*)	ABLE (*in silico* + *vitro* + *vivo*)
FWHM El [°] *in silico*	8.1 ± 0.6	7.7 ± 0.8	6.7 ± 0.4	4.7 ± 1.8	5.6 ± 0.2	5.1 ± 0.3	5 ± 0.4	5.7 ± 0.6
FWHM Az [°] *in silico*	10.4 ± 0.7	4 ± 2.3	7.7 ± 0.6	2.3 ± 1	6.4 ± 0.5	6.6 ± 0.4	5.8 ± 0.8	6.7 ± 0.5
GCNR *in silico*	0.85	0.89	0.83	0.69	0.92	0.87	0.9	0.87
CNR *in silico*	1.67	1.38	1.27	1.04	1.5	1.58	1.64	1.62
CR *in silico*	1.44	1.69	1.8	1.25	1.51	1.55	1.52	1.47
FWHM El [°] *in vitro*	9.7 ± 1.3	9 ± 1.3	7.8 ± 0.8	4.3 ± 1.7	7.6 ± 1.2	6.1 ± 0.7	6.6 ± 0.5	7.4 ± 0.7
FWHM Az [°] *in vitro*	12.4 ± 1.1	11.1 ± 0.8	9 ± 0.3	3.3 ± 1.5	6.9 ± 0.5	7.3 ± 0.5	8.1 ± 0.5	8.4 ± 0.4
GCNR *in vitro*	0.59	0.45	0.34	0.38	0.61	0.56	0.54	0.53
CNR *in vitro*	1.11	0.75	0.57	0.68	1.1	0.96	1	0.95
CR *in vitro*	0.75	0.83	0.91	0.56	0.79	0.81	0.77	0.74

ABLE = adaptive beamforming by deep learning; Az = azimuth; CF = coherence factor; CNR = contrast-to-noise ratio; CR = contrast ratio; DAS = delay and sum; DAS_{Large} = DAS beamforming with a larger aperture array; EBMV = Eigen-based minimum variance; El = elevational; F-DMAS = filtered delay multiply and sum; FWHM = full width at half-maximum; GCNR = generalized contrast-to-noise ratio.

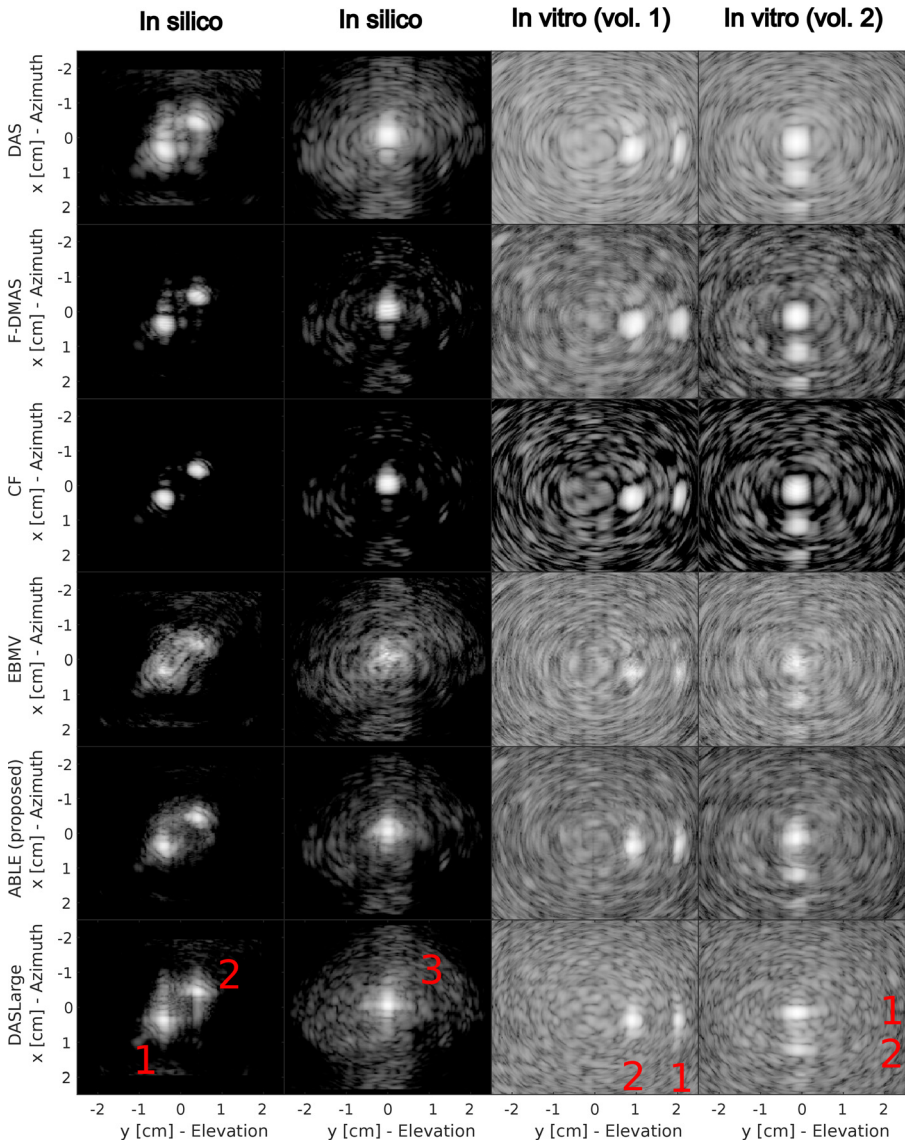

Fig. 5. C-Planes of an in silico phantom and in vitro phantom. The images are for a dynamic range from −60 to 0 dB.

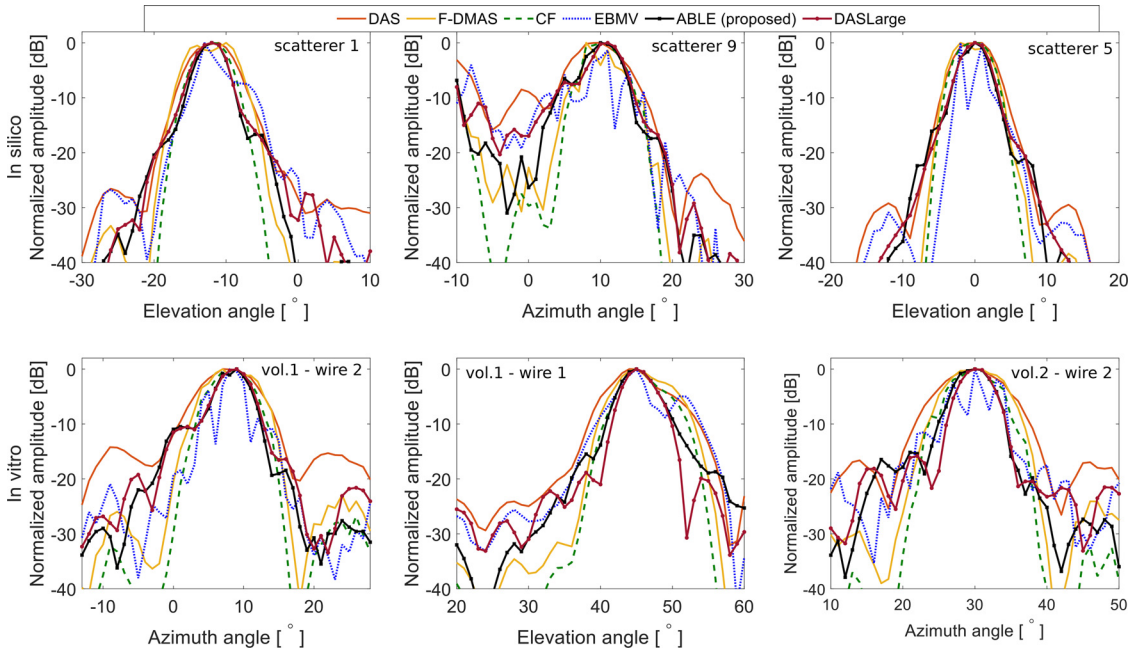

Fig. 6. Beam profiles around in vitro wire targets and in silico scatterers when beamformed with delay and sum (DAS), coherence factor (CF), filtered delay multiply and sum (F-DMAS), Eigen-based minimum variance (EBMV) and proposed adaptive beamforming by deep learning (ABLE) method for the small-array as well as large-array DAS target. (The probe was not rotated to produce the different imaging planes.)

Specifically, the GCNR of F-DMAS is only 0.45 and that of CF is only 0.34, which is lower than the GCNR of ABLE which is 0.56.

If we look closer at the speckle statistics in Figure 7, then we can see that after subtraction of the mean, the probability density function of the log-compressed pixel values is very similar for DAS and DAS_{Large}. The speckle distribution of CF is clearly skewed with a heavy tail, whereas that of F-DMAS is broader than that of DAS. Both the EBMV and ABLE images only have a slightly skewed distribution with respect to DAS. Figure 8 illustrates the apodization pattern calculated by ABLE for an image line where only background scattering is present and for an image line where a highly scattering wire target is present. ABLE clearly selects different apodization weights for the voxels aligned with the wire target than for the voxels aligned with background scattering. This confirms that ABLE, after training on large aperture targets, does not simply apply an advantageous manipulation of the dynamic range, but

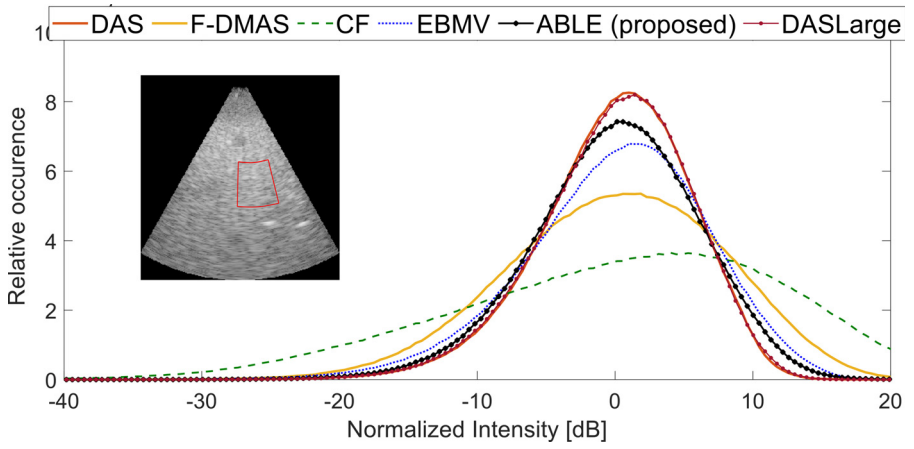

Fig. 7. Probability density function of the log-compressed pixel values after subtraction of the mean for the delay and sum (DAS), coherence factor (CF), filtered delay multiply and sum (F-DMAS), Eigen-based minimum variance (EBMV), proposed adaptive beamforming by deep learning (ABLE) method and larger array DAS target beamformers.

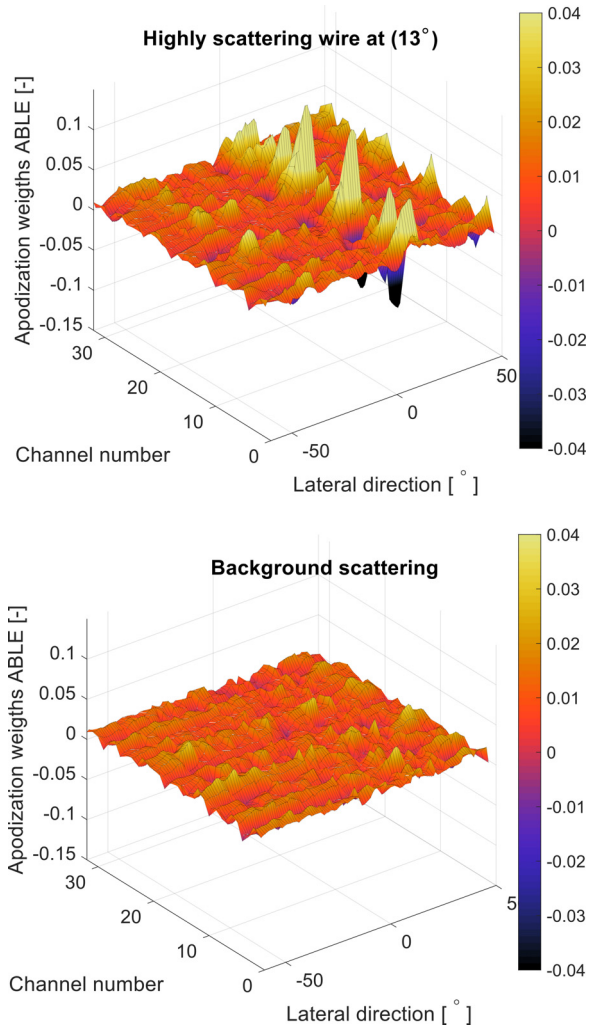

Fig. 8. Apodization weights that adaptive beamforming by deep learning (ABLE) assigns to the data channels for different image locations of the CIRS phantom, containing either a wire target or background scattering.

modifies the apodization weights and thereby the beampattern based on the received data.

Performance comparison modified ABLE: Training data

Figure 9 and the last three columns of Table 1 illustrate the performance of ABLE when the neural network is trained only on in silico data and when in vitro and in vivo data are added to the training set. The average FWHM of ABLE is 6.3˚ when only in silico training data are used, which is marginally better than the 6.4˚ that is achieved when in vitro data are added to the training set; further adding in vivo data to the training set worsens the average FWHM more significantly to 7.1˚. When ABLE is trained on in silico data, the average FWHM is 95% of the FWHM of the DAS_{Large} target, and the FWHM of ABLE remains marginally better than the target at 96% of the FWHM of DAS_{Large} when in vitro data are added to the training set of ABLE. A very slight improvement of in silico GCNR from 0.87 to 0.9 is visible when adding in vitro data to the training set, but adding in vivo data as well decreases the GCNR again to 0.87. Meanwhile a very slight decrease in the in vitro GCNR from 0.56 to 0.54 is visible when adding in vitro data to the training set, and adding in vivo data decreases it slightly further to 0.53.

Performance comparison ABLE: Receptive field size

Figure 10 illustrates the variability in the training and validation loss of ABLE2 trained on in silico for five training runs with a different set of training and validation data. The effect of the variability of five different training runs on the estimated FWHM can be seen in Figure 11, where the FWHM, GCNR and CNR are depicted for ABLE1−4. Figure 11 illustrates that increasing the receptive field of the neural network from a single voxel (ABLE1) to a larger field of view (ABLE2−4) improves performance in terms of in vitro and in silico lateral FWHM. The average FWHM in vitro is 6.7˚ for ABLE2 while it is 7.9˚ for ABLE1. Increasing the lateral receptive field from 2 to 3 or 6 voxels has no clear benefit on the lateral resolution achieved.

The effect of receptive field size on contrast is overall less clear (see Fig. 11). CNR and GCNR decrease slightly on in vitro data when increasing the receptive field from a single voxel, while increasing slightly on in silico data. Figure 12 illustrates that the increased performance of ABLE2 with respect to ABLE1 comes at the cost of greater algorithmic complexity. For the case of a 128-channel probe, the computational cost of ABLE2 would still be below that of EBMV. However, for 32 channels, the algorithmic cost of ABLE2 is about equal to that of EBMV.

DISCUSSION

In this study we used a modified ABLE beamformer and training data that consisted of RF signals as input and targets formed by a larger aperture matrix array. With the proposed method we were able to improve the lateral resolution of volumetric ultrasound images acquired by on-chip micro-beamforming, with respect to the traditional DAS beamformer. The average FWHM of ABLE decreased to 62% of the average FWHM of DAS. ABLE also improved lateral resolution, as measured by the FWHM with respect to the CF beamformer and F-DMAS beamformer, for which the average FWHM values were respectively 124% and 127% that of ABLE. The EBMV beamformer had on average an FWHM that was only 59% that of ABLE, but unlike ABLE, this came at the cost of erratic behavior of the beam profile

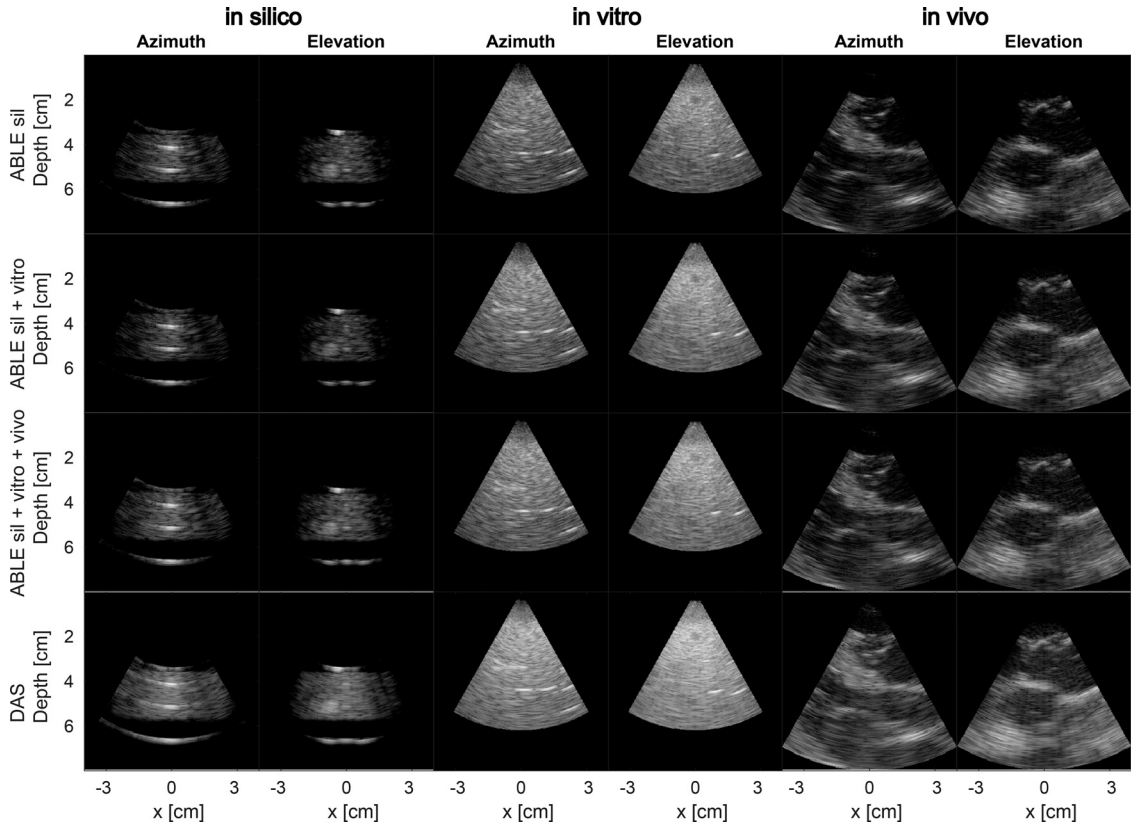

Fig. 9. Azimuth and elevation slices of an in silico phantom and in vitro phantom, as well as in vivo porcine heart data, for the adaptive beamforming by deep learning (ABLE) beamformer when trained on either in silico data, in silico + in vitro data or in silico + in vitro + in vivo data. In the bottom row is the delay and sum (DAS) beamformer for comparison. The images are for a dynamic range from −60 to 0 dB.

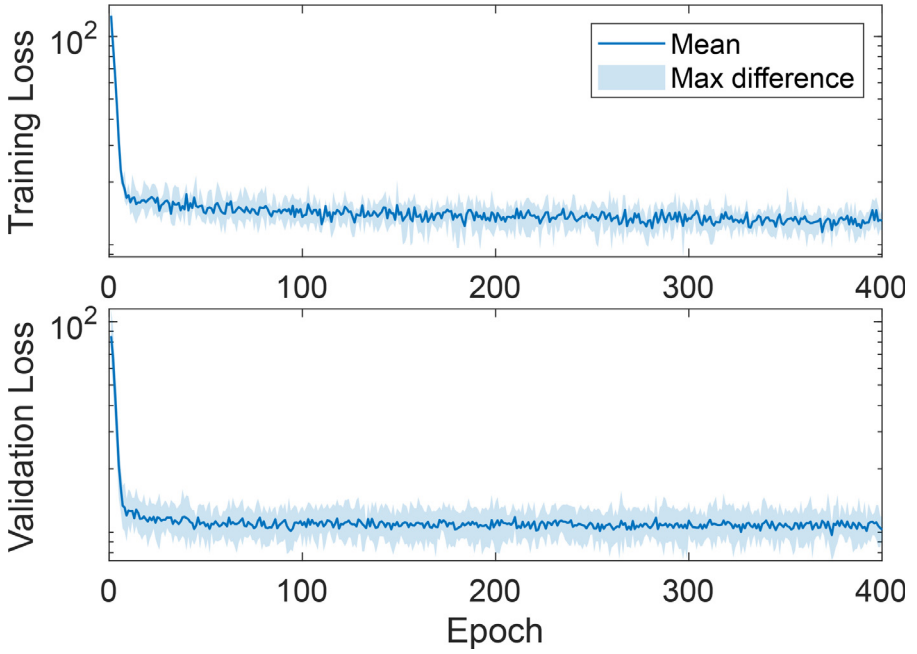

Fig. 10. Mean and maximum differences in training and validation loss when training ABLE2 five times on different training data sets.

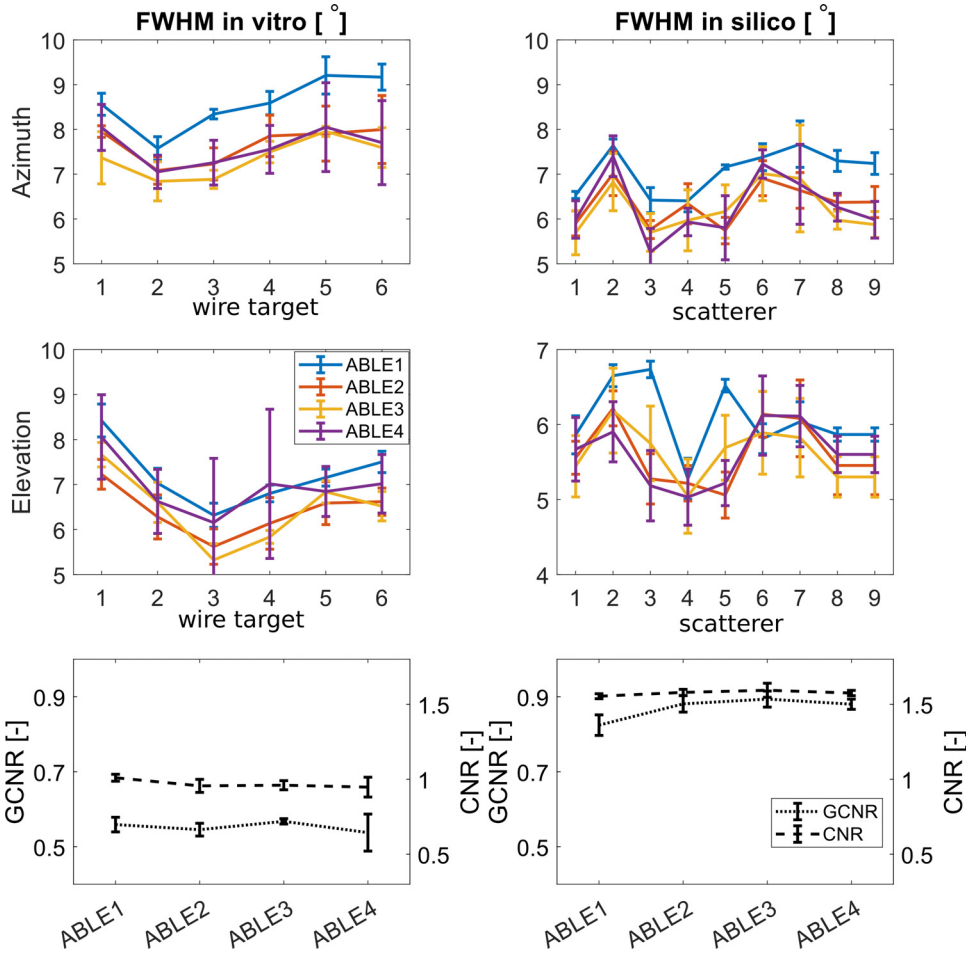

Fig. 11. In vitro full width at half-maximum (FWHM) achieved per scatterer and generalized contrast-to-noise ratio (GCNR) and contrast-to-noise ratio (CNR) as achieved by adaptive beamforming by deep learning (ABLE) beamformers with different receptive field sizes. (Full network architectures are given in Fig. 13.) Error bars indicate the standard deviation for five networks trained on five sets of training data. The mean GCNR and CNR are determined over a single hypo-echoic region.

around point scatterers and a severe reduction in contrast. We suspect that the undesirable performance of the EBMV beamformer is due to focusing inaccuracies resulting from the micro-beamforming step (Lok and Li 2018), needed to enable high-frame-rate read-out of the matrix array. Such inaccuracies may lead to incorrect estimation of signal statistics and, as a result, suboptimal prediction of apodization weights. In contrast, ABLE is trained over a wide range of input signals, which allows for a more robust modeling of the underlying system parameters.

ABLE slightly decreased in vitro contrast compared with DAS as measured by the GCNR, from 0.59 to 0.56, and slightly improved in silico contrast with respect to DAS from 0.85 to 0.87. Thus, ABLE improves the resolution achieved without affecting contrast. As the resolution improvement of ABLE can be achieved by purely training with in silico data, this makes the method easily implementable in practice. To achieve good contrast and avoid increased amplitude variation in background speckle, we needed to add densely scattering regions (10 scatterers per resolution cell) to the in silico training data instead of using only volumes with sparse scattering. Furthermore, the fact that the in silico data do not need to be designed to closely match the organ that is imaged, but consist of basic geometric shapes and point scatterers, gives some confidence that this approach is not just sufficient for this imaging scenario, but could generalize to other scenarios as well. Increasing the receptive field of the original ABLE method by replacing the fully connected layers of the neural network with convolutional layers led to an improvement in resolution, measured by a decrease in average FWHM from 71% that of DAS to 63% of the average FWHM of DAS. We presume that this small amount of spatial context is beneficial especially for volumetric imaging with micro-beamforming,

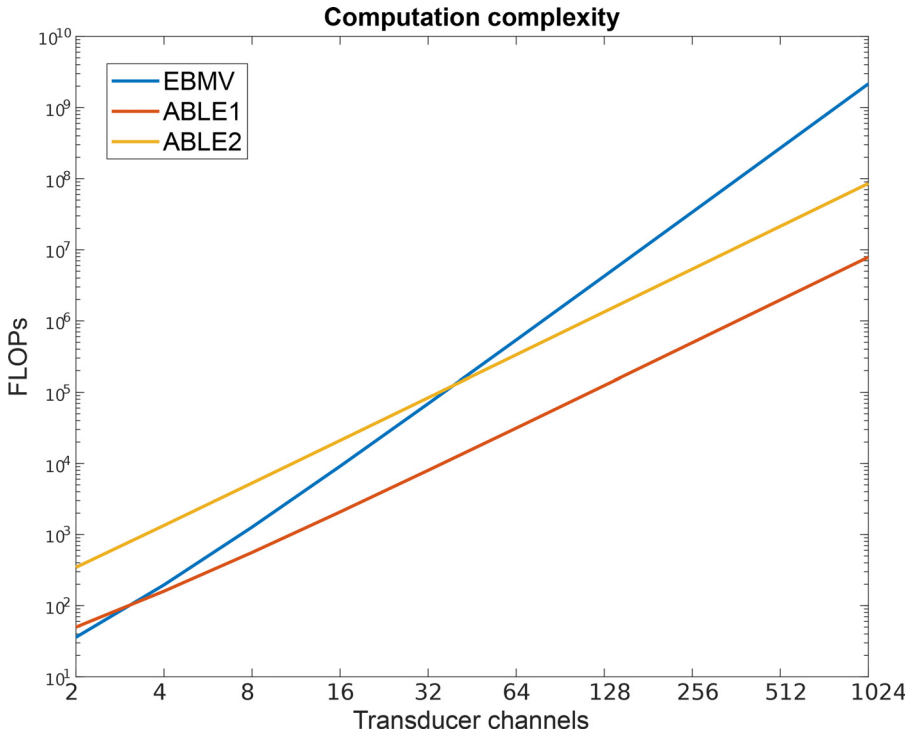

Fig. 12. Number of floating point operations (FLOPS) needed to calculate apodization weights for a single voxel with ABLE1, ABLE2 and Eigen-based minimum variance (EBMV). ABLE = adaptive beamforming by deep learning.

as TOF-corrected RF data of a single voxel will contain small receive focusing errors. Increasing the lateral receptive field size to a value larger than that of ABLE2 did not result in a clear performance improvement. The computational cost is increased by adding convolutional layers. The computational cost of ABLE1 with a single voxel receptive field is always below that of EBMV beamforming. For ABLE2, the computational cost is higher, but remains below that of EBMV beamforming for transducers with more than ∼32 channels. We did not add dropout layers, as the ABLE network in combination with the large array training target, exhibited no signs of overfitting on the training data for all network sizes discussed.

The method surpassed the image quality of the DAS, F-DMAS, CF and EBMV beamformers quantitatively and qualitatively. In silico and in vitro resolution measures reach at least the same level as the target images. When training occurred only on in silico data or on the combination of in silico and in vitro data, the FWHM was on average lower than that in the target images. A reason for this could be that the networks' bottleneck forces a compact representation of the channel data so that certain elements that are present in the target data, such as noise, cannot be accurately represented, thereby having a beneficial effect on the estimation of the apodization weights. However, further experiments would be needed to clarify this. The speckle size of ABLE is more similar to that achieved by the other beamformers than to that achieved by a larger-aperture DAS beamformer. Vignon et al. (2020) already described that learning speckle signals from elements further away from the original aperture is highly challenging if not infeasible. Unlike the spatial covariance of specular reflectors, the spatial covariance of the incoherent signal arising from the sub-resolution tissue scatterers decreases with increasing distance between observation points. As a result, when distance between elements increases, there is no relation between the speckle signals from the original array and those of a remote element in the target array, making the signal infeasible to estimate.

Previous work has indicated that ABLE in combination with an EBMV-based training target performs well across different 2-D ultrasound imaging modalities (Luijten et al. 2020). Here we not only state that extension to 3-D is possible, but we also demonstrate that the ABLE method is effective in a situation where a high-quality target from existing beamformers is not available, by using a larger aperture target. A different approach to improving 3-D ultrasound image quality with deep learning applied after beamforming is taken by Huh et al. (2021). They also used artificially generated targets to improve 3-D ultrasound image quality,

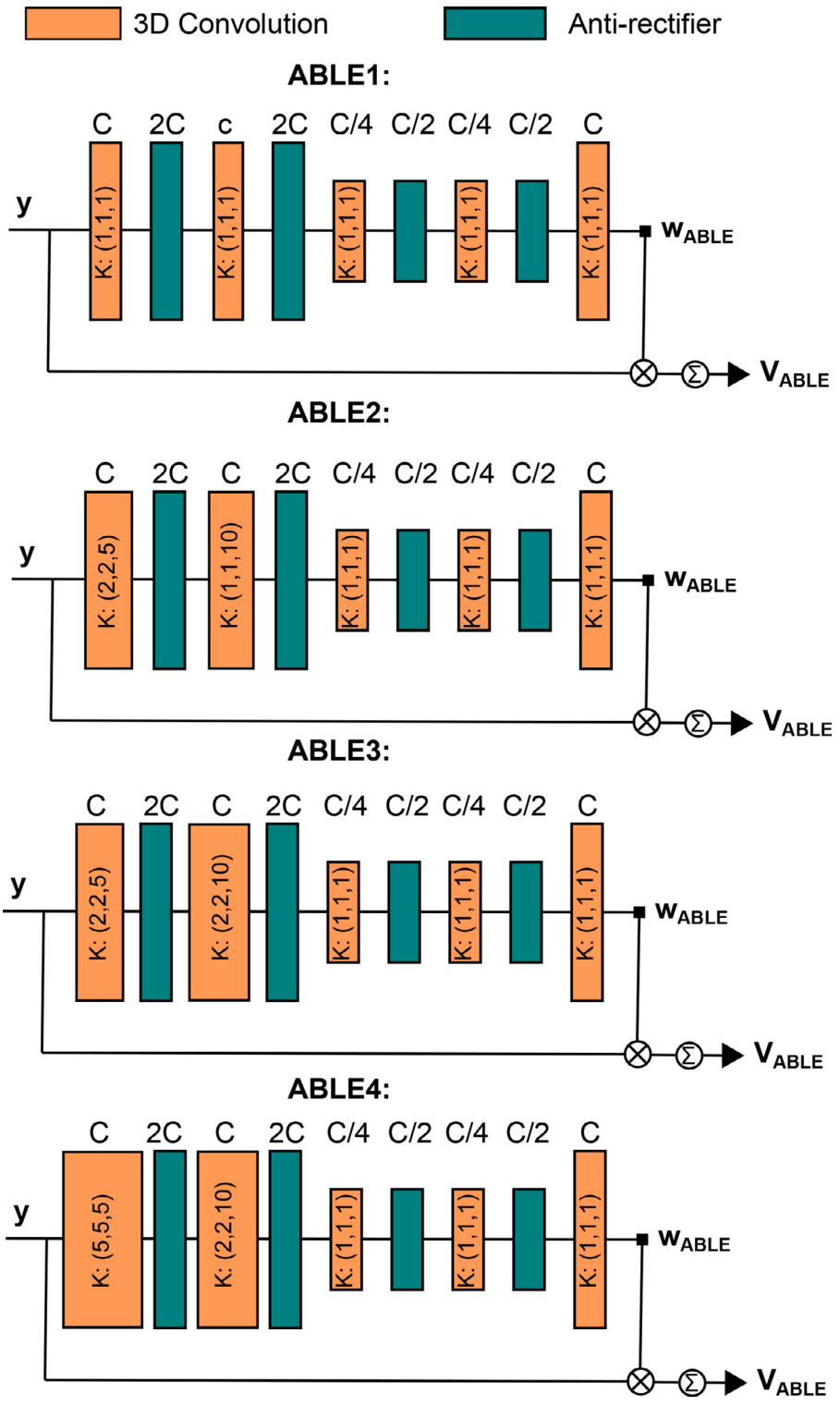

Fig. 13. Modified adaptive beamforming by deep learning (ABLE) networks, with different receptive field sizes. *K* indicates the kernel size in the three spatial directions (azimuth, elevation, axial). The number above each convolutional layer indicates the number of kernels in the layer. *C* is the number of receive channels of the transducer, which here is 32.

but instead used an unsupervised deep learning approach. A switchable CycleGan was used to learn a style transfer from an unpaired training set of low-quality 3-D ultrasound images and high-quality 2-D reference images. Unlike the matrix array used here, they used a 3-D mechanical scanning probe. Image quality improvements were assessed through evaluation by a clinician, making the methods difficult to compare.

Here we specifically focused on 3-D imaging with a matrix array that applies micro-beamforming, as this is a common strategy to solve the data rate and channel connection problem. However, we expect the ABLE method to also improve lateral resolution compared with DAS for fully populated arrays without micro-beamforming. For this case it would be interesting to evaluate whether a training target generated by a larger array or by EBMV beamforming, as was used for 2-D imaging (Luijten et al. 2020), would be a more suitable choice. For such an array with a very large number of channels, the advantages of computational cost when beamforming with ABLE1 or ABLE2 compared with EBMV would become especially apparent. Whether a receptive field larger than a single voxel would be beneficial here remains to be evaluated. The use of deep learning−based beamforming with an artificial array target could also be investigated for row−column and fixed-pattern sparse arrays. Furthermore, jointly learning a sparse sub-sampling pattern and beamforming algorithm could be explored for fully populated matrix arrays, as described for 1-D arrays in (Huijben et al. 2020). Finally, in this work we used real apodization weights. Giving the network more freedom by letting it learn complex apodization weights and working with IQ data can potentially further increase performance.

CONCLUSIONS

In this work we have reported that deep learning−based adaptive beamforming can be used to improve lateral resolution of volumetric ultrasound images acquired by on-chip micro-beamforming. By using only in silico training data consisting of simulated RF data and a target created by a larger aperture, we trained a deep learning−based adaptive beamformer. This beamformer improved lateral resolution, such that image quality was improved beyond that achieved by DAS, F-DMAS, CF and EBMV.

All architectures are provided in Figure 13.

CONFLICT OF INTEREST DISCLOSURE

The authors declare no conflicts of interest.

Acknowledgments—This work is part of the 3-D ICE Project (STW Project 14279), which is financed by the Netherlands Organization for Scientific Research (NWO).

REFERENCES

Asl BM, Mahloojifar A. Eigenspace-based minimum variance beamforming applied to medical ultrasound imaging. IEEE Trans Ultrason Ferroelectr Freq Control 2010;57:2381–2390.

Avanji SAI, Mahloojifar A, Asl BM. Adaptive 3D MV beamforming in medical ultrasound imaging. In: Proceedings, 2013 20th Iranian Conference on Biomedical Engineering, ICBME. New York. : IEEE; 2013. p. 81–86.

Bera D, van den Adel F, Radeljic-Jakic N, Lippe B, Soozande M, Pertijs MA, Verweij MD, Kruizinga P, Daeichin V, Vos HJ, van der Steen AF, Bosch JG, de Jong N. Fast volumetric imaging using a matrix transesophageal echocardiography probe with partitioned transmit−receive array. Ultrasound Med Biol 2018;44:2025–2042.

Blaak S, Yu Z, Meijer GCM, Prins C, Lancee CT, Bosch JG, De Jong N. Design of a micro-beamformer for a 2D piezoelectric ultrasound transducer. Proc IEEE Ultrason Symp 2009;1338–1341.

Capon J. High-resolution frequency-wavenumber spectrum analysis. Proc IEEE 1969;57:1408–1418.

Chennakeshava N, Luijten B, Drori O, Mischi M, Eldar YC, Van Sloun RJ. High resolution plane wave compounding through deep proximal learning. Proc IEEE Int Ultrason Symp 2020;1–4.

Dietrichson F, Smistad E, Ostvik A, Lovstakken L. Ultrasound speckle reduction using generative adversial networks. Proc IEEE Int Ultrason Symp 2018;55–58.

Evans J, Johnson JR, Sun D. High resolution angular spectrum estimation techniques for terrain scattering analysis and angle of arrival estimation. In: Proceedings, 1st IEEE ASSP Workshop on Spectral Estimation, McMaster University. Hamilton, ON, Canada. New York. : IEEE; 1981. p. 134–139.

Goudarzi S, Asif A, Rivaz H. Fast multi-focus ultrasound image recovery using generative adversarial networks. IEEE Trans Comput Imaging 2020;6:1272–1284.

Hastie J, Panzer OP, Weyker P, Flynn BC. Miniaturized echocardiography in the cardiac intensive care unit. J Cardiothorac Vasc Anesth 2019;33:1540–1547.

Hijazi ZM, Shivkumar K, Sahn DJ. Intracardiac echocardiography during interventional and electrophysiological cardiac catheterization. Circulation 2009;119:587–596.

Huang X, Lediju Bell MA, Ding K. Deep learning for ultrasound beamforming in flexible array transducer. IEEE Trans Med Imaging 2021;40:3178–3189.

Huh J, Khan S, Choi S, Shin D, Lee ES, Ye JC. Tunable image quality control of 3-D ultrasound using switchable CycleGAN. Available at: https://arxiv.org/abs/2112.02896. Accessed January 11, 2022.

Huijben IA, Veeling BS, Janse K, Mischi M, Van Sloun RJ. Learning subsampling and signal recovery with applications in ultrasound imaging. IEEE Trans Med Imaging 2020;39:3955–3966.

Hyun D, Brickson LL, Looby KT, Dahl JJ. Beamforming and speckle reduction using neural networks. IEEE Trans Ultrason Ferroelectr Freq Control 2019;66:898–910.

Hyun D, Kim GB, Bottenus N, Dahl JJ. Ultrasound lesion detectability as a distance between probability measures. IEEE Trans Ultrason Ferroelectr Freq Control 2022;69:732–743.

Jensen J. Field: A program for simulating ultrasound systems. Med Biol Eng Comput 1996;34:351–353.

Jensen A, Svendsen B. Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers. IEEE Trans Ultrason Ferroelectr Freq Control 1992;39:262–267.

Kingma DP, Ba JL. Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations. San Diego1–15. 7−9 MayICLR 2015 Conf Track Proc 2015.

Larson JD III. 2-D phased array ultrasound imaging system with distributed phasing. Patent EP-0430450-A2, Hewlett Packard Co. 1993.

Lok UW, Li PC. Microbeamforming with error compensation. IEEE Trans Ultrason Ferroelectr Freq Control 2018;65:1153–1165.
Lorenz RG, Boyd SP. Robust minimum variance beamforming. IEEE Trans Signal Process 2005;53:1684–1696.
Lu J, Millioz F, Garcia D, Salles S, Liu W, Friboulet D. Reconstruction for diverging-wave imaging using deep convolutional neural networks. IEEE Trans Ultrason Ferroelectr Freq Control 2020;67:2481–2492.
Luchies AC, Byram BC. Deep neural networks for ultrasound beamforming. IEEE Trans Med Imaging 2018;37:2010–2021.
Luchies AC, Byram BC. Training improvements for ultrasound beamforming with deep neural networks. Phys Med Biol 2019;64 045018.
Luchies AC, Byram BC. Assessing the robustness of frequency-domain ultrasound beamforming using deep neural networks. IEEE Trans Ultrason Ferroelectr Freq Control 2020;67:2321–2335.
Luijten B, Cohen R, De Bruijn FJ, Schmeitz HA, Mischi M, Eldar YC, Van Sloun RJ. Adaptive ultrasound beamforming using deep learning. IEEE Trans Med Imaging 2020;39:3967–3978.
Mallart R, Fink M. Adaptive focusing in scattering media through sound speed inhomogeneities: The van Cittert Zernike approach and focusing criterion. J Acoust Soc Am 1994;96:3721–3732.
Mamistvalov A, Amar A, Kessler N, Eldar YC. Deep-learning based adaptive ultrasound imaging from sub-Nyquist channel data. IEEE Trans Ultrason Ferroelectr Freq Control 2022;69:1638–1648.
Matrone G, Savoia AS, Caliano G, Magenes G. The delay multiply and sum beamforming algorithm in ultrasound B-mode medical imaging. IEEE Trans Med Imaging 2015;34:940–949.
Nair AA, Tran TD, Reiter A, Lediju Bell MA. A deep learning based alternative to beamforming ultrasound images. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York: IEEE; 2018. p. 3359–3363.
Nair AA, Washington KN, Tran TD, Reiter A, Lediju Bell MA. Deep learning to obtain simultaneous image and segmentation outputs from a single input of raw ultrasound channel data. IEEE Trans Ultrason Ferroelectr Freq Control 2020;67:2493–2509.
Nilsen CI, Holm S. Wiener beamforming and the coherence factor in ultrasound imaging. IEEE Trans Ultrason Ferroelectr Freq Control 2010;57:1329–1346.
Prieur F, Rindal OMH, Austeng A. Signal coherence and image amplitude with the filtered delay multiply and sum beamformer. IEEE Trans Ultrason Ferroelectr Freq Control 2018;65:1133–1140.
Provost J, Papadacci C, Arango JE, Imbault M, Fink M, Gennisson JL, Tanter M, Pernot M. 3D ultrafast ultrasound imaging in vivo. Phys Med Biol 2014;59:L1–L13.
Ramalli A, Dallai A, Bassi L, Scaringella M, Boni E, Hine GE, Matrone G, Savoia AS, Tortoli P. High dynamic range ultrasound imaging with real time filtered-delay multiply and sum beamforming. Proc IEEE Int Ultrason Symp 2017;1–4.
Rindal OMH, Rodriguez-Molares A, Austeng A. The dark region artifact in adaptive ultrasound beamforming. Proc IEEE Int Ultrason Symp 2017;1–4.
Rindal OMH, Austeng A, Fatemi A. Rodriguez-Molares A. The effect of dynamic range alterations in the estimation of contrast. IEEE Trans Ultrason Ferroelectr Freq Control 2019;66:1198–1208.
Rodriguez-Molares A, Rindal OMH, D'Hooge J, Masoy SE, Austeng A, Lediju Bell MA, Torp H. The generalized contrast-to-noise ratio: A formal definition for lesion detectability. IEEE Trans Ultrason Ferroelectr Freq Control 2020;67:745–759.
Santos P, Haugen GU, Lovstakken L, Samset E, D'hooge J. Diverging wave volumetric imaging using subaperture beamforming. IEEE Trans Ultrason Ferroelectr Freq Control 2016;63:2114–2124.
Savord B, Solomon R. Fully sampled matrix transducer for real time 3D ultrasonic imaging. Proc IEEE Ultrason Symp 2003;945–953.
Seward JB, Khandheria BK, Oh JK, Abel MD, Hughes RW, Edwards WD, Nichols BA, Freeman WK, Tajik AJ. Transesophageal echocardiography: Technique, anatomic correlations, implementation, and clinical applications. Mayo Clin Proc 1988;63:649–680.
Shattuck D, Weinshenker MD, Smith SW, von Ramm OT. Explososcan: A parallel processing technique for high-speed ultrasound imaging with linear phased arrays. J Acoust Soc Am 1984;75:1273–1282.
Spencer KT, Krauss D, Thurn J, Mor-Avi V, Poppas A, Vignon P, Connor BG, Lang RM. Transnasal transesophageal echocardiography. J Am Soc Echocardiogr 1997;10:728–737.
Synnevag JF, Austeng A, Holm S. Adaptive beamforming applied to medical ultrasound imaging. IEEE Trans Ultrason Ferroelectr Freq Control 2007;54:1606–1613.
Synnevag JF, Austeng A, Holm S. Benefits of minimum-variance beamforming in medical ultrasound imaging. IEEE Trans Ultrason Ferroelectr Freq Control 2009;56:1868–1879.
van Sloun RJG, Cohen R, Eldar YC. Deep learning in ultrasound imaging. Proc IEEE 2019;108:1–19.
van Sloun RJG, Ye JC, Eldar YC. Deep learning for ultrasound beamforming. arXiv preprint arXiv:2109.11431, 2021.
Vignon F, Shin JS, Meral FC, Apostolakis I, Huang SW, Robert JL. Resolution improvement with a fully convolutional neural network applied to aligned per-channel data. Proc IEEE Int Ultrason Symp 2020;1–4.
Wildes D, Lee W, Haider B, Cogan S, Sundaresan K, Mills DM, Yetter C, Hart PH, Haun CR, Concepcion M, Kirkhorn J, Bitoun M. 4-D ICE: A 2-D array transducer with integrated ASIC in a 10-Fr catheter for real-time 3-D intracardiac echocardiography. IEEE Trans Ultrason Ferroelectr Freq Control 2016;63:2159–2173.
Yoon YH, Ye JC. Deep Learning for Accelerated Ultrasound Imaging. IEEE International Conference on Acoustics, Speech and Signal Processing Proceedings (ICASSP). New York. : IEEE; 2018. p. 6673–6676.
Youn J, Ommen ML, Stuart MB, Thomsen EV, Larsen NB, Jensen JA. Detection and localization of ultrasound scatterers using convolutional neural networks. IEEE Trans Med Imaging 2020;39:3855–3867.
Zhou Z, Wang Y, Yu J, Guo Y, Guo W, Qi Y. High spatial temporal resolution reconstruction of plane-wave ultrasound images with a neural network. IEEE Trans Ultrason Ferroelectr Freq Control 2018;65:1983–1996.
Zhou Z, Wang Y, Guo Y, Jiang X, Qi Y. Ultrafast plane wave imaging with line-scan-quality using an ultrasound-transfer generative adversarial network. IEEE J Biomed Health Inform 2020;24:943–956.
Zhou Z, Guo Y, Wang Y. Ultrasound deep beamforming using a multi-constrained hybrid generative adversarial network. Med Image Anal 2021;71 102086.