A Deep Learning Approach for Beamforming and Contrast Enhancement of Ultrasound Images in Monostatic Synthetic Aperture Imaging: A Proof-of-Concept

Edoardo Bosco, *Student Member, IEEE*, Edoardo Spairani, *Graduate Student Member, IEEE*, Eleonora Toffali, Valentino Meacci, *Member, IEEE*, Alessandro Ramalli, *Senior Member, IEEE*, and Giulia Matrone, *Member, IEEE*

***Abstract*—*Goal:* In this study, we demonstrate that a deep neural network (DNN) can be trained to reconstruct high-contrast images, resembling those produced by the multistatic Synthetic Aperture (SA) method using a 128-element array, leveraging pre-beamforming radiofrequency (RF) signals acquired through the monostatic SA approach. *Methods*: A U-net was trained using 27200 pairs of RF signals, simulated considering a monostatic SA architecture, with their corresponding delay-and-sum beamformed target images in a multistatic 128-element SA configuration. The contrast was assessed on 500 simulated test images of anechoic/hyperechoic targets. The DNN's performance in reconstructing experimental images of a phantom and different *in vivo* scenarios was tested too. *Results*: The DNN, compared to the simple monostatic SA approach used to acquire pre-beamforming signals, generated better-quality images with higher contrast and reduced noise/artifacts. *Conclusions*: The obtained results suggest the potential for the development of a single-channel setup, simultaneously providing good-quality images and reducing hardware complexity.**

***Index Terms*—Deep learning, monostatic configuration, synthetic aperture imaging, ULA-OP system, ultrasound beamforming.**

***Impact Statement*—The proposed deep-learning-based beamforming approach could allow for the realization of a very simple ultrasound imaging system, producing good-quality images and potentially achieving a 128:1 reduction of the transmission/reception channels number.**

Manuscript received 10 October 2023; revised 29 March 2024 and 8 May 2024; accepted 8 May 2024. Date of publication 15 May 2024; date of current version 7 June 2024. This work was supported in part by the Italian Ministry of Education, University and Research (PRIN 2020) under Grant 20205HFXE7 and in part by European Union - Next Generation EU, through the National Recovery and Resilience Plan, Investment 1.3 Enlarged Partnership 8, under Project Age-It, CUP: B83C22004800006. The review of this article was arranged by Editor Paolo Bonato. *(Edoardo Bosco and Edoardo Spairani contributed equally to this work.) (Corresponding author: Giulia Matrone.)*

Edoardo Bosco, Edoardo Spairani, Eleonora Toffali, and Giulia Matrone are with the Department of Electrical, Computer and Biomedical Engineering, University of Pavia, 27100 Pavia, Italy (e-mail: giulia.matrone@unipv.it).

Valentino Meacci and Alessandro Ramalli are with the Department of Information Engineering, University of Florence, 50134 Florence, Italy.

This article has supplementary downloadable material available at https://doi.org/10.1109/OJEMB.2024.3401098, provided by the authors.

Digital Object Identifier 10.1109/OJEMB.2024.3401098

I. Introduction

ULTRASOUND (US) imaging is a versatile and widely used medical imaging technique [1]. Considering that traditional US systems are known for their complex hardware, involving multiple transducers, advanced beamforming algorithms, and significant processing power, lately there has been an increasing focus on simplifying US imaging systems for cost reduction, improved portability, and enhanced accessibility, without sacrificing image quality and diagnostic capabilities [2].

Plane/diverging wave compounding [3] and Synthetic Aperture (SA) US imaging [4], [5] enable full dynamic synthetic focusing, at the expense of a higher processing load. SA imaging softens the requirements on the number of electronic transmission channels: its multistatic configuration, where one element at a time transmits and all the array elements are active during reception, is commonly referred to as Synthetic Transmit Aperture (STA), while the configuration where only a single element transmits/receives the echoes is usually denoted as monostatic Synthetic Aperture Focusing (SAF) [6]. Although SAF involves a lower processing load and requires a simpler system (ideally with a single channel only), reconstructed images have poor contrast when beamformed with the standard Delay and Sum (DAS) algorithm. Besides, the use of unfocused waves with single-element transmission/reception compromises the signal-to-noise ratio (SNR).

Deep learning (DL) has emerged as a powerful paradigm in the field of artificial intelligence [7], which has found great use in US medical imaging for classification, detection and segmentation of anatomical structures [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18].

While such techniques have been mostly applied to already beamformed images, DL has been also investigated for application to raw channel data and beamformed images for "front-end" tasks, like data compression and recovery [19], image

enhancement [20] and beamforming [21], [22], also coupled with segmentation [23] or speckle reduction [24].

Different studies focused on the application of neural networks to ultrafast and SA imaging, with the goal of simplifying the imaging system by minimizing the number of transmissions [25], [26] or samples in time and space [27], [28] required to obtain good-quality images, for example considering a reduced number of transducers only [27], [28], [29], [30].

In this paper, we propose a proof-of-concept of a DL-based approach to perform the reconstruction of B-mode images with improved SNR and contrast in monostatic SAF imaging, expanding on our preliminary study presented in [31]. The approach relies on the use of a U-Net, trained to learn the mapping between the raw radiofrequency (RF) data acquired using SAF and the corresponding DAS-beamformed image obtained with STA using a 128-element array. This could allow for the realization of a very simple system, with good-quality images, potentially achieving a 128:1 reduction of the required number of transmission/reception channels. The feasibility of a real-time ultra-portable, single-channel system was tested on a research scanner, connected to a host PC equipped with a mid-range graphics processing unit (GPU), which enabled a ∼30 Hz frame rate. The proposed proof-of-concept is a cost-effective leap forward in the democratization of US, for home- and self-monitoring (not diagnostic) ultra-portable systems without the need of specialized clinical personnel.

II. Materials and Methods

A. Synthetic Aperture Configurations

Two different SA configurations were considered. Monostatic SAF employs one transmitting element at a time in the transducer array. After each transmission, the same element is used to receive the backscattered echoes. Once all the N elements have transmitted/received the ultrasonic wave, signals are synthetically focused to generate the final image [4], [6]. If $s_{ii}(t)$ are the N RF echo signals, the beamformed value $p_{SAF}(x, z)$ in the image point with coordinates (x, z) is computed as:

$$p_{SAF}(x,z) = \sum_{i=1}^{N} s_{ii}(t - \tau_{ii}(x,z)), \tag{1}$$

where

$$\tau_{ii}(x,z) = \frac{2\sqrt{(x-x_i)^2 + z^2}}{c} \tag{2}$$

is the wave travel time from the i-th element with coordinates $(x_i, 0)$ to point (x, z) and vice-versa, while c is the speed of sound.

The second approach considered is STA [4], [6]. In this case, each time a single element is used to transmit, while all the N elements are used in reception. Synthetic focusing is performed as follows:

$$p_{STA}(x,z) = \sum_{i=1}^{N}\sum_{j=1}^{N} s_{ij}(t - \tau_{ij}(x,z)), \tag{3}$$

TABLE I
SIMULATION PARAMETERS

	Parameter	Value
Linear transducer array	Element width	0.215 mm
	Element height	6 mm
	Kerf	0.03 mm
	No. of elements	128
Excitation signal	Central frequency	5 MHz
	No. of sinusoidal cycles	4
	Speed of sound (c)	1540 m/s
	Sampling frequency (f_s)	20 MHz

TABLE II
NUMERICAL PHANTOM PROPERTIES (TRAINING SET)

	Parameter	Value/Range
Whole phantom	Start depth	z=5.5 mm
	Size (along x, y, z)	36×0.8×49.5 mm^3
Uniform background	Echogenicity	0 dB
Targets	Number	From 1 to 30, or 100
	Shape	Elliptical
	Echogenicity	-40/-20/-10/-5/5/10/20 dB
	Center coordinates	$x_c \in$[-13; 13] mm
		$z_c \in$[6; 55] mm
	Orientation	[-180°; 180°] (step=1°)
	First semi-axis	[2; 9] mm
	Second semi-axis	[2; 9] mm
Point-like scatterer	Number	1 (only in images with 1-30 ellipses, with a 35% probability)
	Position	Random (in the background)

where

$$\tau_{ij}(x,z) = \frac{\sqrt{(x-x_i)^2 + z^2} + \sqrt{(x-x_j)^2 + z^2}}{c} \tag{4}$$

is the travel time from the i-th transmitting element with coordinates $(x_i, 0)$ to point (x, z), and back to the j-th receiving element with coordinates $(x_j, 0)$.

Since more receiving elements are used in STA and synthetic two-way dynamic focusing is implemented, images show a higher SNR and contrast than those obtained with monostatic SAF; on the other hand, a more complex system with N channels is required and N^2 signals have to be beamformed.

B. Training and Test Set

The DNN was fed with a set of N pre-beamforming RF signals, obtained from monostatic SAF, and was trained to learn how to reconstruct the corresponding STA image. The input/output data size was 256×128.

We synthetically produced the training data through Field II [32], [33] in Matlab (The MathWorks, Natick, MA, USA), considering the scan parameters in Table I. The numerical phantoms consisted of a homogeneous background including elliptical targets (also overlapping) of different sizes and echogenicities (Table II). Two phantom subgroups were generated, the first comprising a random number (from 1 to 30) of targets and in some cases also a single highly-reflecting point scatterer, while

TABLE III
NUMERICAL PHANTOM PROPERTIES (TEST SET)

	Parameter	Value/Range
Whole phantom	Start depth	z=5.5 mm
	Size (along x, y, z)	36×0.8×49.5 mm^3
Uniform background	Echogenicity	0 dB
Targets	Number	1 or 2
	Shape	Circular
	Echogenicity	-30/+20 dB
	Radius	[2; 6] mm (step=1 mm)
	Center coordinates	$x_c \in$[-10; 10] mm (step=1 mm) $z_c \in$[13.5; 46.5] mm (step=1 mm)

the second one containing exactly 100 targets, in order to have a variable target density within the dataset.

The training dataset included 13600 pairs of SAF RF signals - STA target images. Moreover, data augmentation was applied by horizontally flipping both signals and images to obtain the full dataset (27200 input/output pairs).

The input SAF RF signals were band-pass filtered and then undersampled [34] considering 256 samples, as explained in the Supplementary Materials available online.

STA target images were instead obtained by DAS-beamforming the simulated RF signals (with $f_s = 20$ MHz), then through compounding and envelope detection; the generated images were undersampled (256 samples along the time/depth axis), normalized by the average value measured in a uniform speckle region, and logarithmically compressed in dB. Finally, a dynamic range of 60 dB (i.e., [−30; +30] dB) was considered.

Data were then normalized to facilitate the DNN learning process: RF signals were standardized with respect to the whole dataset, while target images were rescaled in the range [0; 1] by applying MinMax normalization [35].

The test set included both 500 simulated images of virtual phantoms, containing 1 or 2 circular non-overlapping targets (Table III) and experimental data. These latter were acquired on a commercial US phantom (model 040GSE, CIRS Inc., Norfolk, VA, USA) and a healthy volunteer, using the ULA-OP 256 research system [36], [37] with a linear array (LA533, Esaote s.p.a., Florence, Italy) working at 5 MHz. In this case, for each scenario, 20 consecutive frames were acquired and the corresponding 128 RF signals per frame were first band-pass filtered, undersampled (256 samples) and averaged to improve the SNR, before feeding them as input to the network. The *in vivo* protocol was approved by the institutional review board on ethics of the University of Florence (approval number 309/2024).

C. DNN Architecture

The developed architecture (see Fig. 1 and Supplementary Materials available online) is based on the U-Net [38], drawing inspiration from the DNN proposed by Nair et al. [23].

Training was conducted with a mini-batch approach, and a custom loss function was implemented combining the Mean Absolute Error (MAE) and Structural Similarity Index Measure

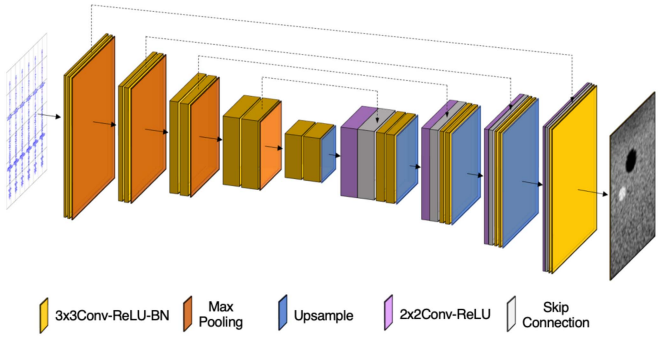

Fig. 1. Block diagram of the proposed convolutional neural network, inspired by the U-Net architecture implemented in [23].

(SSIM) [39]:

$$L = w_{SSIM} \cdot \left(1 - \frac{1}{M}\sum_{i=1}^{M} SSIM_i\right) + w_{MAE} \cdot \frac{1}{M}\sum_{i=1}^{M} MAE_i, \quad (5)$$

where M represents the batch size (16 in our case) and i denotes the i-th example within the batch. The inclusion of MAE allowed for a pixel-wise assessment of image dissimilarity, while the addition of SSIM provided a more perceptually-oriented comparison. The weights w_{SSIM} and w_{MAE} are hyper-parameters used to balance the contributions of the two terms. Following an empirical investigation, we determined that w_{SSIM}=1 and w_{MAE}=10 resulted in a satisfactory configuration.

Adam optimizer, with a constant learning rate of 1e-4, was employed. Early stopping with a patience of 10 epochs was used, which made training stop after 39 epochs.

The DNN was implemented using Tensorflow and Keras libraries, and trained on the Kaggle platform.

D. Performance Evaluation Metrics

To quantitatively compare the contrast performance of SAF, STA and the DNN, we computed the contrast ratio (CR) and generalized contrast-to-noise ratio (gCNR) [40]. Besides, we used the non-parametric Mann-Whitney U test, with a significance value α=0.05, to compare the median contrasts of targets in test images. The Peak Signal-to-Noise Ratio (PSNR) [23] was evaluated to assess the image reconstruction quality of SAF and DNN with respect to STA.

Further details are provided in the Supplementary Materials available online.

III. RESULTS

A. Results on Simulated Data

Fig. 2 illustrates two simulated test examples, comparing output images of DNN, SAF (obtained from the input RF signals through DAS beamforming) and STA (i.e., the target images). The DNN generates higher-quality images than SAF, accurately capturing both hyperechoic and anechoic targets, while in SAF

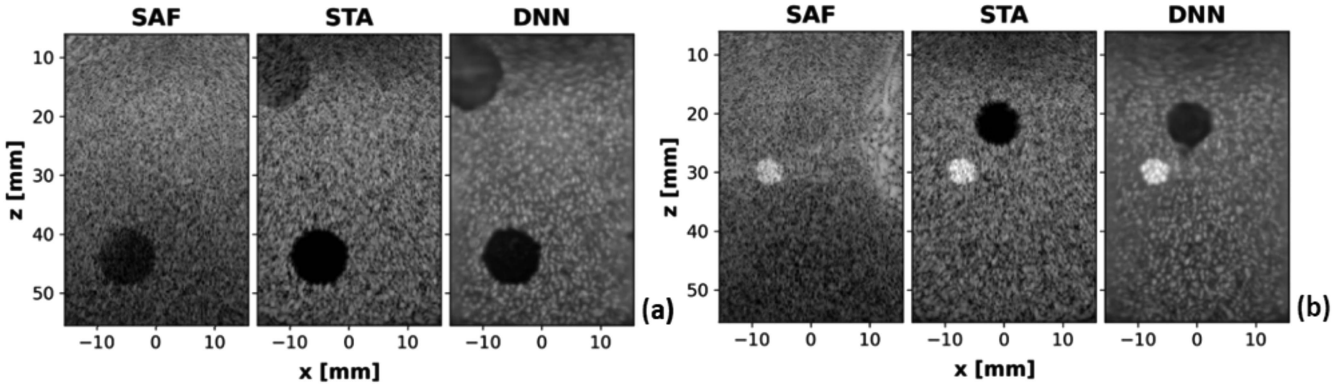

Fig. 2. Synthetic phantom images obtained with SAF+DAS, STA+DAS and with the DNN: (a) two anechoic targets and (b) an anechoic target plus a small hyperechoic target. Images are displayed over a 60 dB dynamic range (log-scale).

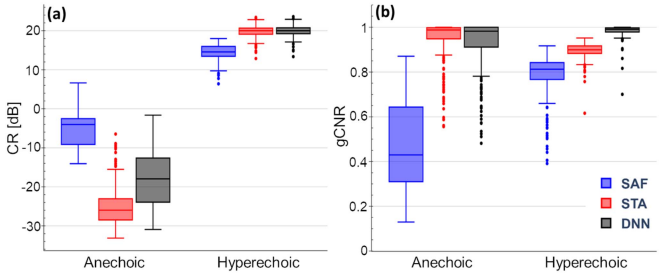

Fig. 3. Boxplots representing the CR and gCNR in simulated test images obtained with SAF+DAS beamforming (blue), STA+DAS beamforming (red) and with the DNN (black), grouped by target echogenicity.

the shallower anechoic targets are barely visible. Furthermore, the DNN has learnt to deal with images where targets with different echogenicity are present (Fig. 2(b)).

Fig. 3 shows the boxplots for the computed contrast metrics. CR and gCNR for the DNN's outputs demonstrate statistically significant improvements compared to SAF images, both for hyperechoic and anechoic targets. Besides, the CRs achieved by the DNN and STA are comparable for the hyperechoic targets, while for the anechoic ones the DNN shows worse performance than STA. Conversely, DNN achieves similar gCNR results compared to STA for anechoic targets and even outperforms it in the hyperechoic case. Further analyses and results are provided in the Supplementary Materials available online.

Finally, the average PSNR computed on the whole simulated test set is 18.86±0.95 dB for DNN vs. STA images, and 15.32±1.03 dB for SAF vs. STA images.

B. Results on Experimental Data

Fig. 4(a) and (b) shows example images obtained from RF signals acquired experimentally on the CIRS phantom. Fig. 4(a) shows two anechoic targets (radius=2.6 mm), at z=17.1 mm and 45.6 mm depth, and a highly reflecting scatterer. Fig. 4(b) shows a hyperechoic target (radius=3.9 mm) and a gray-scale target (with +6 dB natural contrast) at z=31.1 mm depth, with two sets of highly reflective scatterers at different depths.

The DNN accurately reconstructs the shallower anechoic target, which is not visible in the SAF case, but it struggles to reproduce the anechoic target at high depth, clearly visible in STA. Here, it is important to remember that the DNN was exclusively trained using artificially generated examples that did not account for acoustic attenuation effects, which impact the SNR for signals coming from deeper regions.

The following contrast values were achieved by the SAF, STA and DNN methods: for the shallower anechoic target, CR=−0.87 dB, −20.9 dB, and −31.8 dB, gCNR=0.41, 0.97 and 1, respectively; for the deeper target, CR=−6 dB, −21.9 dB and −8.5 dB, gCNR=0.52, 0.99 and 0.79, respectively.

Regarding the hyperechoic targets in Fig. 4(b), the DNN acceptably reconstructs both the circular target with higher echogenicity and the scatterers, also removing some of those artifacts at shallower depths in the SAF image. The CR achieved by the DNN is 13.1 dB, while that of SAF and STA images is 5.8 dB and 9.8 dB, respectively. The DNN also achieves a gCNR of 0.96, while that of SAF and STA images is 0.47 and 0.7, respectively. On the other hand, the less echogenic target (i.e., the one with +6 dB natural contrast at x=−4.05 mm, z=31.1 mm), only slightly glimpsed in the STA image (Fig. 4(b)), is hardly discernible in both SAF and DNN images.

Finally, Fig. 4 also shows some example images obtained from the pre-beamforming signals collected during *in vivo* acquisitions, i.e., a longitudinal (c) and transversal (d) section of the carotid artery, and images of a volunteer's abdomen (e) and thigh (f). Here, the DNN produces images with higher quality than those obtained through SAF, even if they seem to present a slight blur effect as compared to STA ones. Artifacts affecting SAF images at shallow depths appear significantly reduced in the DNN case; some artifacts are still present, especially in the deeper noisy regions and within the main targets. However, noise at the higher depths is visible in STA images too. The lumen of the vascular structures is generally well reconstructed, as well as the surrounding hyperechoic vessel walls and tissues. This is not true in SAF images, where, for example, the lumen of the carotid artery is barely visible.

The contrast was measured in the transversal carotid images, as an example, considering one region inside the carotid lumen and one in the adjacent thyroid, obtaining CRs of 3.5 dB, −21.9 dB, −15.2 dB, and gCNRs equal to 0.5, 0.95 and 1 with SAF, STA and the DNN, respectively.

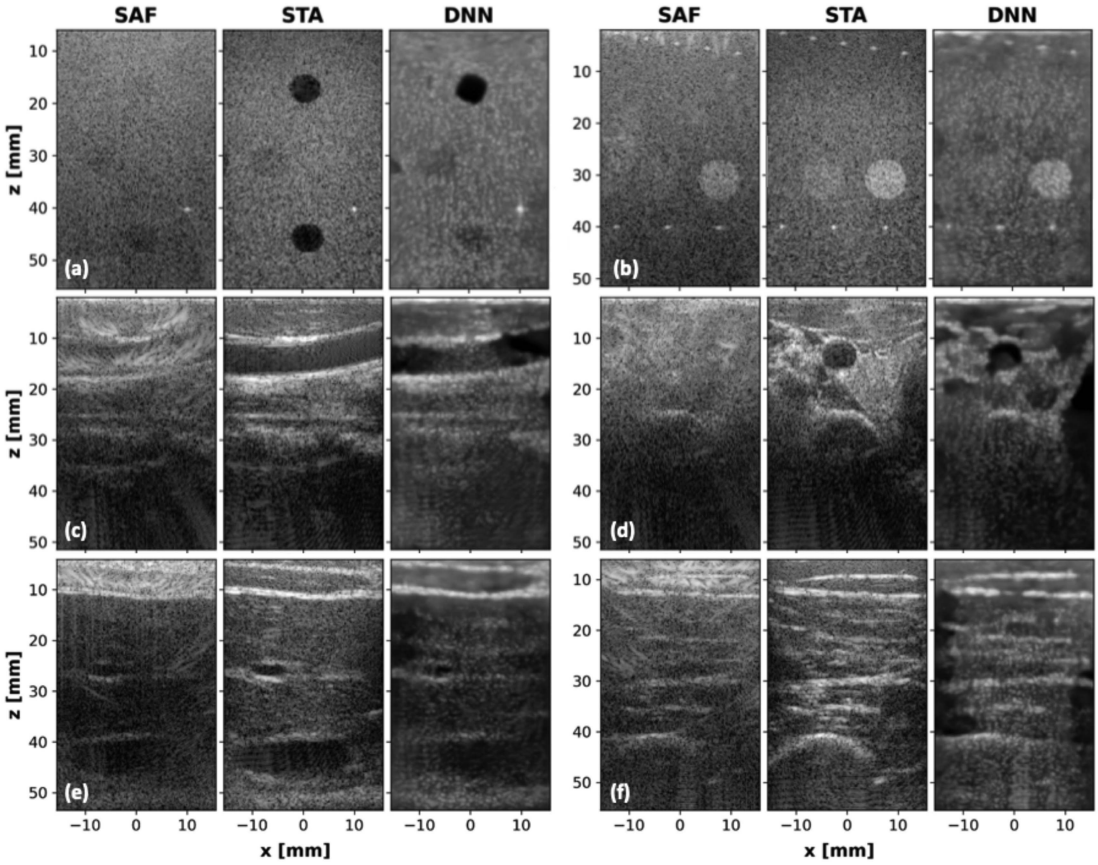

Fig. 4. Experimental phantom/*in vivo* images obtained with SAF+DAS, STA+DAS and the DNN, representing: (a) two anechoic targets with a single scatterer and (b) a hyperechoic target with several scatters in the CIRS phantom (in this last case, a +6 dB gray-scale target can be also noticed); (c) the longitudinal and (d) cross section of a carotid; (e) the abdomen and (f) thigh of a volunteer. Images are displayed over a 60 dB dynamic range (log-scale).

Comparisons between simulated and experimental results are provided in the Supplementary Materials available online.

IV. Discussion

This work proposes a U-Net architecture to implement the beam-formation process of signals acquired with a monostatic SAF approach, generating good-quality images, like those achievable with STA imaging using a 128-element array. The DNN consists of almost 15M parameters and, without any code/network optimization, it takes approximately 13 ms on average to process a single signal set. We preliminary assessed the feasibility of a real-time implementation by exploiting the ULA-OP 256 research scanner, connected to the LA533 probe and implementing the SAF scan sequence, for the experimental tests. A real-time frame rate of about 30 Hz was achieved. Details on this feasibility test, together with a video, are provided in the Supplementary Materials available online.

Other papers in the literature investigated the task of system requirements simplification using DL-based methods to generate enhanced images using a reduced set of elements [27], [28], [29], [30]. However, to the best of our knowledge, our work proposes an even stronger reduction of the channels number potentially required by the hardware system, i.e., from 128 to generate the target images, to 1 single channel only in the input data case. Furthermore, such an idea is tested by addressing the task of image reconstruction starting from raw pre-beamforming RF signals.

Simulation results demonstrate that the DNN is generally well able to reproduce anechoic and hyperechoic targets, with similar or higher gCNRs than those achieved by STA. Importantly, with the DNN contrast is significantly higher than that of SAF images on average and with reduced variability. Moreover, grating lobe artifacts, at the shallower depths in SAF images, are attenuated in DNN ones. The PSNR also confirms that the network is generally well able to reproduce images similar to the target ones.

Concerning the speckle, the DNN generally tends to make it more uniform, with some blurring effect mainly visible at the lower depths. This can primarily be attributed to the use of a suboptimal network architecture. Factors such as the feature map size in the bottleneck, influenced by the input image dimensions and number of pooling layers, and the receptive field size, determined by convolutional kernels, could all contribute to the loss of high-frequency information. Additionally, pixel-wise loss functions, such as MAE, can introduce blurring artifacts. To improve the speckle pattern, perceptual losses like SSIM could be prioritized or an adversarial term could be included. It should also be considered that in input SAF images, in the shallower regions, strong artifacts due to grating/side lobes are present,

which may affect the DNN ability to clearly reproduce the underlying structures, causing such blurring in output images.

Resolution was not quantitatively evaluated here, as our primary objective was to enhance contrast with respect to SAF. However, looking at the transversal profiles of simulated anechoic, hyperechoic, and point-like targets provided in the Supplementary Materials, it can be observed that lateral resolution is not worsened by the DNN as compared to SAF.

Experimental images further confirm that the DNN is well able to reproduce hyperechoic and anechoic structures, removing artifacts that affect SAF images, which can be clearly seen, for example, in the *in vivo* images at depths around 10 mm. The quality of *in vivo* STA images remains the highest. However, it should be considered that our work represents a proof-of-concept of a new approach that could simplify the US hardware requirements for ultra-portable systems, which are to be mainly meant for non-diagnostic applications and even for non-expert users.

Some limitations of the present study should be highlighted too. First, the network was trained on simulated data only, without additive noise, which is typically not Gaussian and complex to model, and without tissue attenuation effects, which would have required a $\times 30$ longer simulation time (1.7 hours vs. 3 minutes on the used PC), not feasible considering that 13600 datasets were generated. This is likely to explain why the network is not always able to well reproduce the structures of interest (especially anechoic regions) at the higher depths in real experimental cases, where such a phenomenon is non-negligible. However, in our case, it would have not been feasible to experimentally generate thousands of data (an order of magnitude which can be also found in other works in the literature on ultrasound image reconstruction applications [19], [21], [23]).

Besides, to generate training SAF and STA images, no apodization or other weighting techniques were applied, which could have further improved the quality of input images or target results. This was done to compare the different methods without introducing further degrees of freedom in the processing chain, while primarily focusing on demonstrating if generating images of acceptable quality with a single element/channel system could be feasible thanks to DNNs.

In the future, improvements could be achieved by modifying the architecture and training setup, for example by exploring alternative architectures offering a more comprehensive perspective of the data, or by optimizing training parameters, such as the selection of an adjustable weighting factor to combine the loss terms used.

Also, to address the challenge of limited generalization ability on experimental data, additional phantom and *in vivo* images could be incorporated into the training set. Regularization methods could be employed, leveraging a limited number of experimental examples to encourage the network to learn shared features present in both the training and test data. Finally, since our objective is to deploy the network on a portable device, it could be further optimized by reducing its size, thus making it lighter and more compact.

Potential changes can also be introduced by modifying the task itself. In this study, we selected images in a STA configuration as our targets, as STA provides high-quality images while keeping the single-element transmission approach used for the inputs. Nonetheless, different target images obtained with any other imaging modality providing high contrast/resolution could be used to improve the network outputs.

V. Conclusion

In conclusion, the proposed study demonstrates that, by implementing a DL-based beamforming strategy, a potential 128:1 reduction of the number of hardware channels in the US system could be possible, still achieving images with adequate quality, which could have a valuable impact on the democratization of US, for home- and self-monitoring (not diagnostic) ultra-portable systems without the need of specialized clinical personnel.

Author Contributions

E.B., E.S. and G.M. designed the study and wrote the paper. E.B., E.S. and E.T. created/managed the dataset, implemented the proposed methods and analysed the results. V.M. carried out the real-time tests and contributed to paper writing. A.R. acquired the experimental data, carried out the real-time tests and contributed to paper writing. G.M. conceived the idea, created/managed the dataset, and supervised the project. All the authors reviewed and approved the final manuscript.

Conflict of Interest

The authors declare no conflicts of interest.

References

[1] J. A. Jensen, "Medical ultrasound imaging," *Prog. Biophys. Mol. Biol.*, vol. 93, no. 1–3, pp. 153–165, Jan. 2007, doi: 10.1016/j.pbiomolbio.2006.07.025.

[2] J. M. Baran and J. G. Webster, "Design of low-cost portable ultrasound systems: Review," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2009, pp. 792–795, doi: 10.1109/IEMBS.2009.5332754.

[3] M. Tanter and M. Fink, "Ultrafast imaging in biomedical ultrasound," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 61, no. 1, pp. 102–119, Jan. 2014, doi: 10.1109/TUFFC.2014.2882.

[4] J. A. Jensen, S. I. Nikolov, K. L. Gammelmark, and M. H. Pedersen, "Synthetic aperture ultrasound imaging," *Ultrasonics*, vol. 44, pp. e5–e15, Dec. 2006, doi: 10.1016/j.ultras.2006.07.017.

[5] M. Jakovljevic, R. Michaelides, E. Biondi, D. Hyun, H. A. Zebker, and J. J. Dahl, "Adaptation of range-doppler algorithm for efficient beamforming of monostatic and multistatic ultrasound signals," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 69, no. 11, pp. 3165–3178, Nov. 2022, doi: 10.1109/TUFFC.2022.3205923.

[6] G. Matrone, A. S. Savoia, G. Caliano, and G. Magenes, "Depth-of-field enhancement in filtered-delay multiply and sum beamformed images using synthetic aperture focusing," *Ultrasonics*, vol. 75, pp. 216–225, Mar. 2017, doi: 10.1016/j.ultras.2016.11.022.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, May 2015, Art. no. 7553, doi: 10.1038/nature14539.

[8] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[9] S. Leclerc et al., "Deep learning for segmentation using an open large-scale dataset in 2D echocardiography," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2198–2210, Sep. 2019, doi: 10.1109/TMI.2019.2900516.

[10] S. Leclerc, T. Grenier, F. Espinosa, and O. Bernard, "A fully automatic and multi-structural segmentation of the left ventricle and the myocardium on highly heterogeneous 2D echocardiographic data," in *Proc. IEEE Int. Ultrason. Symp.*, 2017, pp. 1–4, doi: 10.1109/ULTSYM.2017.8092797.
[11] S. Y. Shin, S. Lee, I. D. Yun, S. M. Kim, and K. M. Lee, "Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images," *IEEE Trans. Med. Imag.*, vol. 38, no. 3, pp. 762–774, Mar. 2019, doi: 10.1109/TMI.2018.2872031.
[12] M. H. Yap et al., "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 4, pp. 1218–1226, Jul. 2018, doi: 10.1109/JBHI.2017.2731873.
[13] J. Ma, F. Wu, T. Jiang, J. Zhu, and D. Kong, "Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images," *Med. Phys.*, vol. 44, no. 5, pp. 1678–1691, May 2017, doi: 10.1002/mp.12134.
[14] J. Chi, E. Walia, P. Babyn, J. Wang, G. Groot, and M. Eramian, "Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network," *J. Digit. Imag.*, vol. 30, no. 4, pp. 477–486, Aug. 2017, doi: 10.1007/s10278-017-9997-y.
[15] F. T. Dezaki et al., "Deep residual recurrent neural networks for characterisation of cardiac cycle phase from echocardiograms," in *Deep Learning in Medical Image Analysis and Multimodal Learning For Clinical Decision Support*, M. J. Cardoso Eds., Cham, Switzerland: Springer, 2017, pp. 100–108, doi: 10.1007/978-3-319-67558-9_12.
[16] S. Roy et al., "Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2676–2687, Aug. 2020, doi: 10.1109/TMI.2020.2994459.
[17] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *Proc. IEEE*, vol. 108, no. 1, pp. 11–29, Jan. 2020, doi: 10.1109/JPROC.2019.2932116.
[18] B. Luijten, N. Chennakeshava, Y. C. Eldar, M. Mischi, and R. J. G. van Sloun, "Ultrasound signal processing: From models to deep learning," *Ultrasound Med. Biol.*, vol. 49, no. 3, Mar. 2023, Art. no. 3, doi: 10.1016/j.ultrasmedbio.2022.11.003.
[19] D. Perdios, A. Besson, M. Arditi, and J.-P. Thiran, "A deep learning approach to ultrasound image recovery," in *Proc. IEEE Int. Ultrason. Symp.*, 2017, pp. 1–4, doi: 10.1109/ULTSYM.2017.8092262.
[20] D. Perdios, M. Vonlanthen, A. Besson, F. Martinez, M. Arditi, and J.-P. Thiran, "Deep convolutional neural network for ultrasound image enhancement," in *Proc. IEEE Int. Ultrason. Symp.*, 2018, pp. 1–4, doi: 10.1109/ULTSYM.2018.8580183.
[21] 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.*, 2018, pp. 3359–3363, doi: 10.1109/ICASSP.2018.8461575.
[22] D. Hyun et al., "Deep learning for ultrasound image formation: CUBDL evaluation framework and open datasets," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 68, no. 12, pp. 3466–3483, Dec. 2021, doi: 10.1109/TUFFC.2021.3094849.
[23] A. A. Nair, K. N. Washington, T. D. Tran, A. Reiter, and M. A. Lediju Bell, "Deep learning to obtain simultaneous image and segmentation outputs from a single input of raw ultrasound channel data," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 12, pp. 2493–2509, Dec. 2020, doi: 10.1109/TUFFC.2020.2993779.
[24] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, "Beamforming and speckle reduction using neural networks," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 5, pp. 898–910, May 2019, doi: 10.1109/TUFFC.2019.2903795.
[25] 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, doi: 10.1109/TUFFC.2017.2736890.
[26] J. Lu, F. Millioz, D. Garcia, S. Salles, W. Liu, and D. Friboulet, "Reconstruction for diverging-wave imaging using deep convolutional neural networks," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 12, pp. 2481–2492, Dec. 2020, doi: 10.1109/TUFFC.2020.2986166.
[27] B. Luijten et al., "Adaptive ultrasound beamforming using deep learning," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 3967–3978, Dec. 2020, doi: 10.1109/TMI.2020.3008537.
[28] A. Mamistvalov, A. Amar, N. Kessler, and Y. C. Eldar, "Deep-learning based adaptive ultrasound imaging from sub-nyquist channel data," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 69, no. 5, pp. 1638–1648, May 2022, doi: 10.1109/TUFFC.2022.3160859.
[29] V. Kumar, P.-Y. Lee, B.-H. Kim, M. Fatemi, and A. Alizad, "Gap-filling method for suppressing grating lobes in ultrasound imaging: Experimental study with deep-learning approach," *IEEE Access*, vol. 8, pp. 76276–76286, 2020, doi: 10.1109/ACCESS.2020.2989337.
[30] F. Vignon, J. S. Shin, F. C. Meral, I. Apostolakis, S.-W. Huang, and J.-L. Robert, "Resolution improvement with a fully convolutional neural network applied to aligned per-channel data," in *Proc. IEEE Int. Ultrason. Symp.*, 2020, pp. 1–4, doi: 10.1109/IUS46767.2020.9251482.
[31] E. Toffali, E. Spairani, A. Ramalli, and G. Matrone, "Improving the quality of monostatic synthetic-aperture ultrasound imaging through deep-learning-based beamforming," in *Proc. IEEE Int. Ultrason. Symp.*, 2022, pp. 1–4, doi: 10.1109/IUS54386.2022.9958283.
[32] J. Jensen, "FIELD: A program for simulating ultrasound systems," *Med. Biol. Eng. Comput.*, vol. 34, pp. 351–352, Jan. 1996.
[33] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 39, no. 2, Mar. 1992, Art. no. 2, doi: 10.1109/58.139123.
[34] M. Mishali and Y. C. Eldar, "Sub-nyquist sampling," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 98–124, Nov. 2011, doi: 10.1109/MSP.2011.942308.
[35] S. G. Patro and K. K. Sahu, "Normalization: A preprocessing stage," *Int. Adv. Res. J. Sci., Eng. Technol.*, vol. 2, 2015, doi: 10.17148/IARJSET.2015.2305.
[36] E. Boni et al., "Architecture of an ultrasound system for continuous real-time high frame rate imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 64, no. 9, pp. 1276–1284, Sep. 2017, doi: 10.1109/TUFFC.2017.2727980.
[37] E. Boni et al., "ULA-OP 256: A 256-channel open scanner for development and real-time implementation of new ultrasound methods," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 63, no. 10, pp. 1488–1495, Oct. 2016, doi: 10.1109/TUFFC.2016.2566920.
[38] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi Eds., Cham, Switzerland: Springer, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.
[39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.
[40] A. Rodriguez-Molares et al., "The generalized contrast-to-noise ratio: A formal definition for lesion detectability," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 4, pp. 745–759, Apr. 2020, doi: 10.1109/TUFFC.2019.2956855.

Open Access funding provided by 'Università degli Studi di Pavia' within the CRUI CARE Agreement