

Single-Shot CNN-Based Ultrasound Imaging with Sparse Linear Arrays

Dimitris Perdios*, Manuel Vonlanthen*, Florian Martinez*, Marcel Arditi*, and Jean-Philippe Thiran*[†]

*Signal Processing Laboratory 5 (LTS5), École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland

[†]Department of Radiology, University Hospital Center (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland

Abstract—Sparse arrays are a topic of high interest within the ultrasound (US) imaging community, because of their promising ability to reduce costs, complexity, energy consumption, and data transfer requirements of US systems, thus addressing the main challenges of 3-D and portable 2-D systems. Undersampling a transducer array usually results in a significant increase in imaging artifacts, caused primarily by higher grating lobe (GL) levels. Thus, state-of-the-art sparse arrays design strategies focus on avoiding GLs, while compromising on the resulting image resolution and uniformity. In this work, we investigated the applicability of convolutional neural network (CNN)-based image reconstruction, having recently proven its potential in reducing GL artifacts, for reconstructing images from single unfocused acquisitions using uniformly undersampled linear array configurations on receive. The proposed reconstruction method consists of first computing a low-quality estimate from the undersampled single-shot acquisitions using a delay-and-sum (DAS) algorithm, followed by applying a real-time-capable CNN, trained specifically to reduce diffraction artifacts. Experiments were conducted within a simulation environment, in the context of plane wave imaging on a numerical test phantom dedicated to US image quality assessment. The proposed approach achieved an image comparable or better to that obtained from conventional DAS beamforming using the full array with uniformly undersampled arrays up to a factor of three, demonstrating a promising potential for sparse array imaging in general.

Index Terms—Convolutional neural networks, deep learning, grating lobes, image reconstruction, image restoration, sparse arrays, ultrafast ultrasound imaging.

I. INTRODUCTION

Sparse array design and image formation have been studied extensively during the past decades, mainly to reduce costs and complexity of large-channel-count 3-D ultrasound (US) scanners [1], and more recently for low-power portable ones [2]. Real-time 3-D US imaging relies on 2-D transducer arrays with thousands of elements, rendering both the fabrication process and element control involved very challenging. The use of sparse 2-D arrays has proven to be a promising solution to reduce the complexity of 3-D US systems and is still heavily studied [3]. On the other hand, the recent advent of low-power, portable US systems [4] has led to increased interest in 2-D imaging using sparse 1-D arrays [2], not only to reduce the costs and complexity, but also to keep the energy consumption and data transfers minimal.

Many sparse array design strategies have been studied with the goal of maintaining a high image quality, while reducing the number of array elements deployed as low as possible. While “naively” undersampling a transducer array uniformly keeps both the main lobe shape and side lobe (SL) levels largely intact, the interferences between element channels can lead to devastating grating lobe (GL) artifacts, dominating a large portion of the resulting image. Thus, most design strategies focus on mitigating GL artifacts, while making compromises on the resulting resolution, SL levels, and image uniformity. Some consider random element positioning (or subsampling) to break the periodicity of the array [5], others leverage probabilistic optimization techniques (e.g. simulated annealing) under some arbitrary cost function [3] or adopt different transmit and receive element arrangements with the aim of cancelling out their respective GLs [2], [6].

In [7], we proposed reconstructing US images from single unfocused acquisitions, by first computing a low-quality image estimate with a backprojection-based delay-and-sum (DAS) algorithm and then restoring a high-quality image using a convolutional neural network (CNN), specifically trained to reduce diffraction artifacts. The CNN has proven a strong ability to remove GL artifacts while properly restoring previously shadowed tissue structures of interest.

In this contribution, we investigated, in a numerical environment, the applicability of our CNN-based image reconstruction method [7] to the case of ultrafast plane-wave (PW) imaging using undersampled arrays. To ensure GL-free transmit wavefronts, we considered unapodized single PW insonifications using the full array on transmit and uniformly undersampled arrays on receive. As this simple transmit scheme could be implemented in the front-end, receive-only sparsity is sufficient for the sought-after data-rate reduction. In contrast to most sparse array design approaches, we considered uniform undersampling for its simplicity and its ability to maintain proper main-lobe and SL characteristics, and relied on a trained CNN to reduce the resulting GL artifacts significantly.

II. MATERIALS AND METHODS

A. Imaging Configurations

As in [7], we considered a US acquisition system composed of the 9L-D transducer (GE Healthcare, Chicago, Illinois, USA) and the Vantage 256 system (Verasonics, Kirkland, WA, USA). Relevant system specifications are summarized in Table I. The 9L-D is a linear array transducer containing

This work was supported in part by the Swiss National Science Foundation under Grant 205320_175974 and Grant 206021_170758. (Corresponding author: Dimitris Perdios.) (email: dimitris.perdios@epfl.ch.)

TABLE I
ACQUISITION SYSTEM SPECIFICATIONS

Parameter	Value
Center frequency	5.3 MHz
Bandwidth	75 %
Aperture	43.93 mm
Element number	192
Pitch	230 μ m
Element width ^a	207 μ m
Element height	6 mm
Elevation focus	28 mm
Transmit frequency	5.208 MHz
Excitation cycles ^b	1
Sampling frequency	20.833 MHz

^aGuessed (no official data available).

^bSingle excitation cycle with equalization pulses.

TABLE II
TRANSMIT AND RECEIVE ARRAY CONFIGURATIONS

Parameter	UQ ^a	LQ	U2	U3	U4	U6
Tx element number	383	192	192	192	192	192
Tx pitch	0.39 λ	0.78 λ	0.78 λ	0.78 λ	0.78 λ	0.78 λ
Rx element number	383	192	96	64	48	32
Rx pitch	0.39 λ	0.78 λ	1.56 λ	2.34 λ	3.12 λ	4.68 λ
Tx-Rx scheme	383 SA	1 PW	1 PW	1 PW	1 PW	1 PW

^aSimulation exclusive.

192 elements with a center frequency of 5.3 MHz and a bandwidth of 75 % (at -6 dB). An excitation composed of a single-cycle tri-state waveform of 67 % duty cycle centered at 5.208 MHz, with leading and trailing equalization pulses of quarter-cycle durations and opposite polarities, was used for all pulse-echo measurements. The received echo signals were sampled at 20.833 MHz, guaranteeing a Nyquist sampling rate up to a bandwidth of 200 %. We assumed a speed of sound of 1540 m/s (typical for soft tissue), resulting in an element spacing (i.e. pitch) of more than half a wavelength (i.e. $\sim 0.78\lambda$) at the effective frequency. Thus, even when using all transducer elements on receive, images reconstructed from unfocused acquisitions with a DAS algorithm will inevitably be contaminated by GL artifacts, especially at shallow depths.

Based on this acquisition system, we considered full-aperture unapodized single-PW insonifications with normal incidence and different uniformly undersampled receive configurations, namely by a factor of 2, 3, 4, and 6 (U2–U6). For comparison purposes, we also considered single-PW insonifications with full-aperture receive, named low-quality (LQ) to keep the same terminology as in [7]. As it will serve as reference for CNN trainings (Section II-B), we also introduce here the ultra-high-quality (UQ) configuration, which relies on synthetic aperture (SA) imaging from a spatially oversampled version of the transducer array with 383 elements and a halved pitch of $\sim 0.39\lambda$ (i.e. simulation exclusive), guaranteeing GL-free images of gold-standard quality. All configurations considered are summarized in Table II.

From the acquisitions obtained with each configuration considered, full-view images were reconstructed using the

backprojection-based DAS algorithm introduced in [7]. Analytic in-phase quadrature (IQ) images were reconstructed on a $\lambda/4 \times \lambda/8$ (Cartesian) grid, with a width spanning the 9L-D aperture (Table I) and a depth ranging from 1 mm to 60 mm. The image grid resolution guaranteed Nyquist sampling for the radio frequency (RF) content of US images in both lateral and axial dimensions, and resulted in images of size 596×1600 pixels.

B. CNN-Based Image Reconstruction Method

To obtain higher quality images from single-shot unfocused acquisitions using sparse array configurations, we deployed the two-step CNN-based image reconstruction method proposed in [7], briefly summarized hereafter. In a first step, a (vectorized) low-quality estimate $\tilde{\mathbf{x}} \in \mathbb{R}^n$ is computed from the (vectorized) transducer elements measurements $\mathbf{y} \in \mathbb{R}^m$, obtained from a single unfocused insonification, by means of a backprojection-based DAS operator $\mathbf{D}: \mathbb{R}^m \rightarrow \mathbb{R}^n$ as $\tilde{\mathbf{x}} = \mathbf{D}\mathbf{y}$. In this study, the low-quality estimates corresponded to either of the U2 to U6 images described in Section II-A. The low-quality estimate $\tilde{\mathbf{x}}$ is then fed to a CNN $f_{\theta}: \mathbb{R}^n \rightarrow \mathbb{R}^n$, with parameters θ , trained to reduce diffraction artifacts while restoring the underlying image content of interest and preserving proper speckle patterns, thereby recovering a high-quality estimate as $\hat{\mathbf{x}} = f_{\theta}(\tilde{\mathbf{x}})$.

The deployed network architecture is a residual CNN with multi-scale and multi-channel filtering properties, adapted from the popular U-Net. It is composed of 2-D convolutional layers (CLs) and rectified linear units (ReLUs) arranged in symmetric downsampling and upsampling paths and includes several adaptations such as the use of residual convolutional blocks (RCBs) and additive intrinsic skip connections [7]. For this study, we deployed the best-performing CNN architecture analyzed in [7] that still maintains real-time capabilities, namely using 16 initial expansion channels.

Trainings were performed precisely as detailed in [7], namely in a supervised manner using a dataset composed of 30 000 simulated image pairs (i.e. input and reference). The mean signed logarithmic absolute error (MSLAE) loss [7], which accounts for both the high dynamic range (HDR) and the RF property of US images, was minimized using the well-known Adam optimizer over 500 000 iterations, using a learning rate of 5×10^{-5} and a batch size of two. For each input image type, a dedicated CNN was trained using a specific simulated-images training dataset composed of low-quality input images (U2–U6) and corresponding reference images (UQ). An in-house 3-D spatial impulse response (SIR) simulator, validated against the well-known Field II simulator [8], was used to generate the element raw-data for both input and reference images. The numerical phantoms were composed of random scatterers with a density that ensured fully-developed speckle patterns throughout the resulting images. The resulting training images are characterized by overlapping ellipsoidal zones of random size, position, and orientation, with mean echogenicities spanning an 80-dB range.

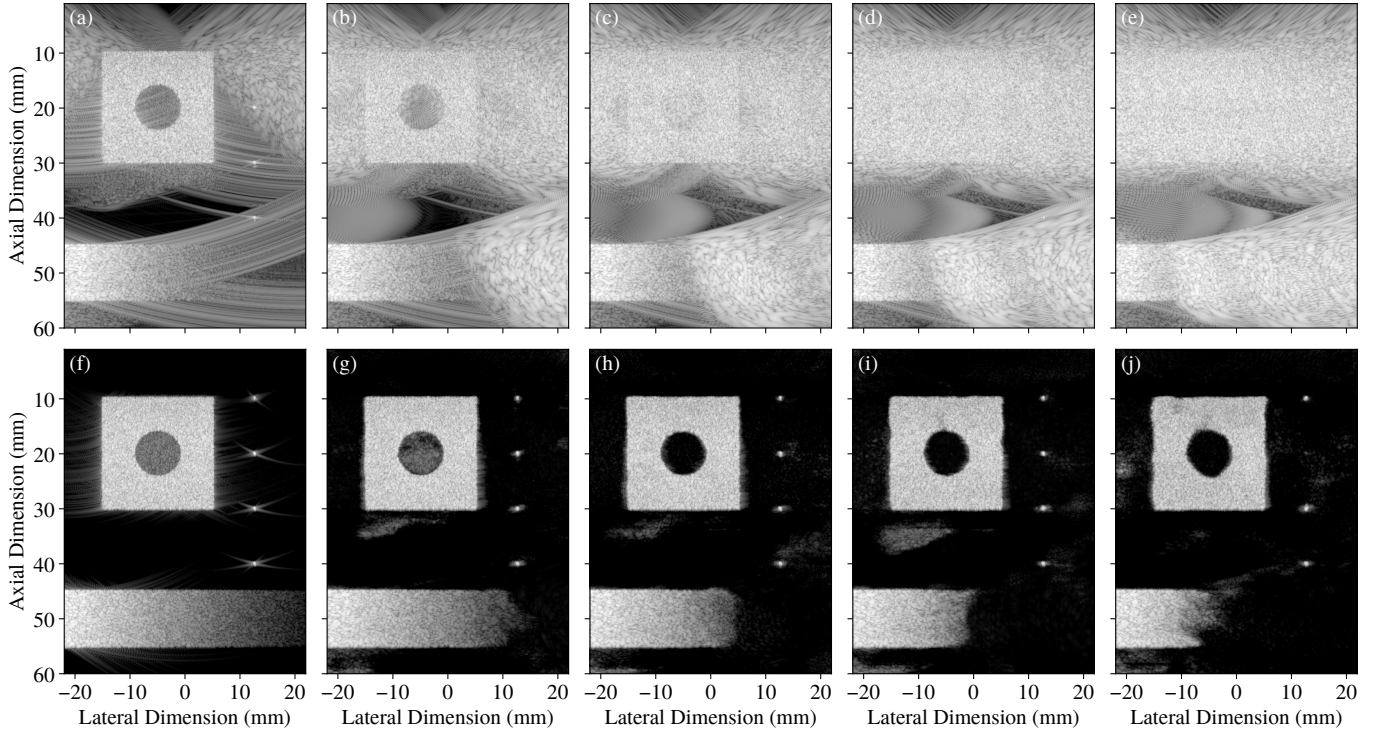


Fig. 1. B-mode image representations (98-dB range) of a numerical test phantom sample, reconstructed using each imaging configuration considered (Table II): (a) low-quality (LQ) configuration (i.e. full array on receive) and (f) ultra-high-quality (UQ) configuration (i.e. reference image), providing image quality comparisons; (b) to (e) U2, U3, U4, and U6 uniformly undersampled configurations (by a factor of 2, 3, 4, and 6 on receive only); (g) to (j) images recovered from U2, U3, U4, and U6 using the proposed approach with the corresponding trained CNN.

C. Numerical Experiment

To evaluate the image quality achieved using CNN-based reconstruction (Section II-B) on single-shot PW acquisitions with the different uniformly undersampled receive configurations considered (U2–U6, Section II-A), we relied on the numerical test phantom deployed in [7], previously adapted from [9] to the US acquisition system under study. The test phantom is composed of tissue-mimicking echogenic zones embedded in an anechoic background [Fig. 1(f)], and allows to compute a wide range of image quality metrics. It comprises a cylindrical inclusion embedded in a block of square section with a prescribed contrast (C) of 36 dB, a rectangular region with a log-linear lateral echogenicity gradient ranging from 30 dB to -50 dB, and four bright reflectors p_0 , p_1 , p_2 , and p_3 at increasing depths. Similar to the numerical phantoms used for training purposes (Section II-B), echogenic zones were filled with densely-populated random scatterers to guarantee fully-developed speckle patterns.

A total of 300 statistically independent realizations of the test phantom were simulated for each undersampled configuration considered with and without applying the corresponding trained CNN, as well as for the comparative methods LQ and UQ. Image quality was assessed using a subset of the quality metrics from the numerical test phantom, namely using the structural similarity (SSIM) against UQ reference images as a global metric (evaluated on log-compressed B-mode representations), the C achieved between the block and the inclusion, the

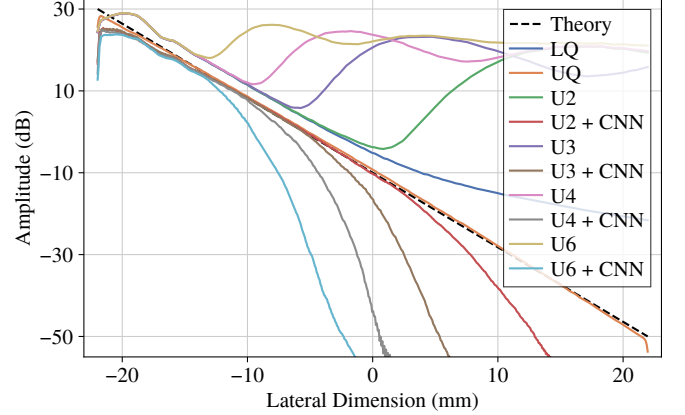


Fig. 2. Mean amplitude responses (averaged along the axial dimension) of the lateral gradient zone in the numerical phantom [Fig. 1(f)], averaged over 300 independent realizations.

lateral full width at half maximum (FWHM) at each bright reflector position, and the linearity preservation. All metrics were computed precisely as described in [7] and averaged over the 300 statistically independent realizations.

III. RESULTS AND DISCUSSION

Fig. 1 shows example images of a test phantom sample (from the 300 realizations) for each imaging configuration considered (Table II), and all resulting image quality metrics, computed as described in Section II-C, are provided in Table III.

TABLE III
NUMERICAL TEST PHANTOM METRICS

Metric ^a	LQ	UQ	U2	U2 + CNN	U3	U3 + CNN	U4	U4 + CNN	U6	U6 + CNN
SSIM	0.31	× ^b	0.12	0.68	0.07	0.62	0.03	0.56	−0.00	0.56
C (dB)	−27.43	−35.25	−15.28	−43.05	−2.96	−70.34	−0.75	−74.07	−0.52	−73.80
FWHM lat.	p_0 (μm)	276.6	202.2	× ^c	265.9	× ^c	256.9	× ^c	269.0	× ^c
	p_1 (μm)	336.2	242.7	339.5	273.7	× ^c	304.4	× ^c	× ^c	× ^c
	p_2 (μm)	388.6	280.5	× ^c	299.7	× ^c	316.1	× ^c	× ^c	× ^c
	p_3 (μm)	446.6	322.4	× ^c	359.8	× ^c	376.9	× ^c	410.8	× ^c

^a Metrics were averaged over 300 independent realizations.

^b SSIM metric was computed against UQ.

^c At least one image sample resulted in an undetectable reflector.

The impact of increased uniform undersampling can be analyzed in Fig. 1(a) to 1(e), where both a significant increase in the area covered by GL artifacts and a decrease of the signal-to-artifact ratio within zones contaminated by these artifacts can be observed with the increase of the undersampling factor. Furthermore, comparing images obtained from the undersampled configurations, in zones free of GL artifacts (e.g. highly echogenic portion of the rectangular gradient), [Fig. 1(b) to 1(e)] to the LQ image which uses the full array [Fig. 1(a)], one can note that both main-lobe and SL characteristics are not effected by the uniform undersampling.

Overall, the CNN-based image reconstruction demonstrated a high efficiency at recovering many structures previously hidden by the significant GL artifacts arising from the undersampled configurations. For instance, one can appreciate the recovery performance of the bright reflectors which were almost undetectable [e.g. compare Fig. 1(d) and Fig. 1(i)], or the recovery of the square block which progressively disappears below GL artifacts as the undersampling factor increases. This is confirmed by the increase in SSIM, for instance improved from 0.07 to 0.62 in the U3 case, even surpassing the SSIM obtained for LQ significantly.

However, these recovery performances were limited to a certain signal-to-artifact ratio of the input images, below which the CNNs generally removed both the GL artifacts and the underlying signal of interest. Because of decreasing signal-to-artifact ratios with increased undersampling, a progressively larger portion of the image could not be recovered accurately and the image quality achieved thus decreased steadily (Table III). This behavior can be well observed when comparing the recovery of the circular inclusion from the U2 and U3 images [Fig. 1(g) and 1(h)]. While the CNN was able to (partially) recover the structures within the circular inclusion in the U2 case, the inclusion was fully thresholded by the CNN from an undersampling factor of three (U3–U6).

The structures that could be recovered by the CNN, namely in zones of the input images with a signal-to-artifact ratio sufficiently high before restoration, were recovered with high-quality speckle patterns, preserved log-linearity (see Fig. 2) and, in the case of the point scatterers, with a lateral resolution comparable or better than the one achieved for LQ (FWHM in Table III). The analysis of the mean amplitude responses within the lateral gradient zone (Fig. 2) can be used to estimate

the signal-to-artifact ratio below which the CNN is not capable to recover the underlying signal of interest with the prescribed log-linear amplitude. For instance, in the U3 case, the restored mean amplitude (U3 + CNN) starts to deviate from UQ at approximately −5 mm, where the input signal-to-artifact ratio (between LQ and U3) is of approximately −3 dB.

IV. CONCLUSION

We showed that the proposed CNN-based image reconstruction method can increase the image quality achieved from single PW acquisitions using uniformly undersampled linear arrays on receive. Indeed, the overall quality of the restored images is competitive to images reconstructed by a DAS algorithm from single PW acquisitions on the full array up to an undersampling factor of three, demonstrating a high potential in the context of sparse array imaging. As observed in [7], the performances can be further improved by optimizing the CNN architecture, its training process, or simply increasing its capacity. To evaluate the potency of the studied approach further, additional studies, including *in vitro* and *in vivo* data should be conducted.

REFERENCES

- [1] K. E. Thomenius, “Evolution of ultrasound beamformers,” in *1996 IEEE Ultrason. Symp. Proc.*, vol. 2, 1996, pp. 1615–1622.
- [2] J. H. Song, J. Lee, S. Yeo, G.-D. Kim, and T.-K. Song, “An analytical approach to designing optimal sparse 1-D phased arrays for handheld ultrasound imaging,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 7, pp. 1354–1365, Jul. 2020.
- [3] E. Roux, F. Varray, L. Petrusca, C. Cachard, P. Tortoli, and H. Liebgott, “Experimental 3-D ultrasound imaging with 2-D sparse arrays using focused and diverging waves,” *Sci. Rep.*, vol. 8, no. 1, p. 9108, Dec. 2018.
- [4] P. A. Hager and L. Benini, “LightProbe: A digital ultrasound probe for software-defined ultrafast imaging,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 4, pp. 747–760, Apr. 2019.
- [5] R. E. Davidsen, J. A. Jensen, and S. W. Smith, “Two-dimensional random arrays for real time volumetric imaging,” *Ultrasonic Imaging*, vol. 16, no. 3, pp. 143–163, Jul. 1994.
- [6] S. S. Brunke and G. R. Lockwood, “Broad-bandwidth radiation patterns of sparse two-dimensional vernier arrays,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 44, no. 5, pp. 1101–1109, Sep. 1997.
- [7] D. Perdios, M. Vonlanthen, F. Martinez, M. Arditi, and J.-P. Thiran, “CNN-based image reconstruction method for ultrafast ultrasound imaging,” Aug. 2020. [Online]. Available: <http://arxiv.org/abs/2008.12750>
- [8] J. A. Jensen, “FIELD: A program for simulating ultrasound systems,” in *10th Nord. Conf. Biomed. Imaging*, vol. 4, no. Supplement 1, 1996, pp. 351–353.
- [9] O. M. H. Rindal, A. Austeng, A. Fatemi, and A. Rodriguez-Molares, “The effect of dynamic range alterations in the estimation of contrast,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 7, pp. 1198–1208, Jul. 2019.