Deep-Learning Based Adaptive Ultrasound Imaging from Sub-Nyquist Channel Data

Naama Kessler and Yonina C. Eldar

***Abstract*—Traditional beamforming of medical ultrasound images requires sampling rates significantly higher than the actual Nyquist rate of the received signals. This results in large amounts of data to store and process, translating to big, expensive and power-hungry devices. In light of the capabilities demonstrated by deep learning methods over the past years across a wide range of tasks, including medical imaging, it is natural to consider their ability to recover high-quality ultrasound images from partial data. Here, we propose an approach for reconstruction from temporally and spatially sub-sampled channel data. We begin by considering sub-Nyquist sampled data, time-aligned in the frequency domain and transformed back to the time domain with no additional recovery steps. This results in low resolution and corruption due to loss of frequencies and aliasing. The data is further sampled spatially to emulate acquisition from a sparse array. It is then given as input to an encoder-decoder convolutional neural network which is trained separately for each rate reduction, with targets generated from minimum-variance (MV) beamforming of the fully-sampled data. Our approach yields high-quality B-mode images, with higher resolution than previously proposed reconstruction approaches (NESTA) from compressed data as well as delay-and-sum beamforming (DAS) of the fully-sampled data. In terms of contrast, it is comparable to MV beamforming of the fully-sampled data. Moreover, prediction times are 10 times faster than NESTA's, thus enabling better, faster and more efficient imaging than is mostly used in clinical practice today.**

***Index Terms*—Compressed beamforming, ultrasound imaging, sub-Nyquist reconstruction.**

I. Introduction

OVER the past decades, ultrasound has become a preferred scanning modality in a variety of clinical scenarios due to its non-radiating and non-invasive nature, high availability and relatively low cost. In the scanning process, each line of the output image is acquired by transmitting a series of short acoustic pulses along a narrow beam from an array of transducer elements towards the imaged body. The pulses are scattered by acoustic impedance perturbations in the tissue, creating echos which are detected by the array elements. The received echos are then sampled and digitally aligned in time in order to compensate for delays originating from the array's geometry. Finally, the time-aligned signals are averaged to produce a B-mode image. This last step is referred to as beamforming.

This work was supported in part by the Igel Manya Center for Biomedical Engineering and Signal Processing, as well as the Benoziyo Endowment Fund for the Advancement of Science, the Estate of Olga Klein – Astrachan, and the European Union's Horizon 2020 research and innovation program under grant No. 646804-ERC-COG-BNYQ;

N. Kessler and Y. C. Eldar are with the Faculty of Math and CS, Weizmann Institute of Science, Rehovot, Israel, email:{kessler.naama, yonina.eldar}@weizmann.ac.il.

Performing beamforming in the time domain necessitates high sampling rates of the received signals. This requirement originates from the time-alignment step, in which sufficient delay resolution is obtained through over-sampling and interpolation. In practice, signals are sampled at rates 4 to 10 times higher than their Nyquist rate [1], which is considered as the minimum sampling rate to allow perfect reconstruction [2], [3]. This leads to vast amounts of data to transmit, store and process.

Data volume is crucial in particular in receive beamforming. Averaging the signals at this stage is performed using either fixed weights, as in Delay-And-Sum (DAS) beamforming [4], or adaptive weights, such as minimum-variance (MV) beamforming [5]. While the latter allows for better trade-off between the main lobe's width and side-lobes intensity, translating to improved resolution, its computational cost is high and increases with data size. Therefore, it is not applicable in real-time clinical applications where DAS is typically used, leading to degraded image quality. An alternative approach, proposed in [6], originates from a a statistical interpretation of the process. There, the signal of interest and interference are modeled as Gaussian random variables, and beamforming is viewed as maximum-a-posteriori (MAP) estimation of the signal. Since MAP implementation requires estimation of the signal and interference variances, an iterative process is proposed to evaluate the distribution parameters given the signal and vice versa. This approach has shown improvement over DAS in terms of contrast, but still is not comparable to MV in terms of resolution.

To circumvent the long processing time and the high computational cost, a variety of techniques have emerged which enable reconstruction of the beamformed signal from partial data. In [7], [8], the authors shift the process of time-alignment to the frequency domain by drawing a connection between the set of Fourier coefficients of the received signals pre-alignment and the set of Fourier coefficients of the beamformed signal, allowing to sample the former at their effective Nyquist rate. They consider also reduction to a sub-Nyquist rate by sampling only a subset of the Fourier coefficients of the received signals. Reconstruction is then performed using compressed sensing methods [9], [10], relying on the Finite-Rate-of-Innovation (FRI) structure of the beamformed signal [9], [11], [12].

In [13], a method is proposed to reduce the data and hardware burden by using sparse arrays, namely, only a subset of the receive elements are activated. Processing is then performed by convolutional beamforming, which is shown to preserve the array beampattern under appropriate conditions

on the chosen array [14]. More specifically, the achieved beampattern is equivalent to that of a virtual array given by the sum co-array of the sparse array. Thus, using a sparse array whose sum co-array includes a full uniform linear array, yields enhanced resolution and contrast from fewer transmitting elements. This method is integrated with Fourier domain beamforming in [15], allowing sub-sampling in both space and time and paving the way to wireless ultrasound.

The above mentioned works perform recovery from the partial samples by solving a minimization problem in an iterative manner. The process is repeated for every acquisition angle in the frame, resulting in long processing times. Moreover, fixed weighting is applied on the signals prior to their summation, which results in degraded image resolution in comparison to adaptive methods.

Inspired by the notable performance of deep learning over the past years across a wide range of fields and tasks [16]–[18], medical imaging included [19], [20], different uses of deep learning in ultrasound reconstruction have been investigated [21]. In [22], a deep-learning based MV beamformer is proposed, implemented with a fully-connected neural network over fully-sampled and spatially sub-sampled channel data. The objective is to ease the heavy computational burden of adaptive beamforming, as well as to improve performance by learning from samples. The papers [23]–[25] target the same problem using encoder-decoder architectures, while [26] considers the combination of deep neural networks and MV beamforming for contrast enhancement; It suggests an ensemble of networks operating in the frequency domain over frequency sub-bands, either before or after computation of the adaptive weights. In [27], the authors expand the problem to sub-sampling of transducer elements and transmission angles. They theoretically justify their approach by drawing a connection between the encoder-decoder architecture and low-rank Hankel matrix decomposition which models the problem. Other works that consider reconstruction from a partial set of plain-wave transmissions use fully convolutional networks [28], [29], encoder-decoder networks [30]–[33] and generative adversarial networks (GANs) [34], [35].

The aforementioned approaches propose fixed sampling schemes that do not depend on the transmitted pulses or the task in hand. This aspect is addressed in [36], where two concatenated models are proposed. The first model learns to sub-sample the data, while the second, whose architecture depends on the desired task, learns recovery. Both are jointly trained in an end-to-end fashion. The method is tested in several different sub-sampling tasks, including temporal sub-sampling of partial Fourier measurements and recovery of the original signal from them. However, this particular task is tested *in-silico* only over simulated random signal vectors; time-alignment of the sub-sampled signals, which is required for ultrasound image recovery, is not addressed, and neither is beamforming. Since traditional time-alignment is not possible over a sub-sampled grid, the application to recovery of temporally sub-sampled ultrasound data remains unexplored.

To the best of our knowledge, no other deep-learning based method has been proposed for the recovery of temporally partial ultrasound channel data. Here, we address this issue based on the results in [7] and [13]. The input to our model is partial Fourier measurements of the received signals, time-aligned in the frequency domain, as well as partial samples in both space and time which resemble acquisition from a sparse array; Those can be obtained by sub-Nyquist sampling implemented in hardware, as shown in [11], [37]. Using a convolutional neural network (CNN) trained with MV-beamformed targets, our model learns a full reconstruction to a high quality B-mode image. Despite the significant reduction in data volume, it outperforms DAS in terms of resolution and MV in terms of contrast, while both are evaluated over fully-sampled data. Moreover, it is almost 10 times faster than NESTA, the method used in [7] and [13] for reconstruction. Thus, our approach enables higher quality, faster and more efficient imaging than is mostly used in clinical practice today.

The remainder of this paper is organized as follows: Section 2 shortly reviews DAS and Fourier domain beamforming. Section 3 introduces our proposed method, which is verified in Sections 4, 5. Results are discussed in Section 6.

II. Existing Beamforming Methods

A. Delay-And-Sum beamforming

Consider a phased-array transducer of M elements aligned along the x-axis, where m_0 denotes the central element - the origin, and δ_m denotes the distance to the *m*th element.

The imaging cycle begins at time $t = 0$, as a short pulse is transmitted from the array in direction θ. Denote by $(x, z) = (ct\sin\theta, ct\cos\theta)$ the coordinates of the the pulse at time *t*, as it propagates through the tissue at speed c. Assume that a point reflector positioned at this location scatters the energy such that an echo is received by all array elements, at a time depending on their location. Beamforming is the operation of averaging the reflections while compensating for these differences in arrival time.

Let $\varphi_m(t)$ be the signal received by the *m*th element, and

$$\hat{\tau}(t;\theta) = t + \frac{d_m(t;\theta)}{c} \tag{1}$$

be its time of arrival, where

$$d_m(t;\theta) = \sqrt{(ct\cos\theta)^2 + (\delta_m - ct\sin\theta)^2} \tag{2}$$

is the distance traveled by the reflection to the element. Applying an appropriate delay to the *m*th signal results in its alignment to the origin m_0:

$$\hat{\varphi}_m(t;\theta) = \varphi_m(\tau_m(t;\theta)) \tag{3}$$

where

$$\tau_m(t;\theta) = \frac{1}{2}\left(t + \sqrt{t^2 - 4(\delta_m/c)\,t\sin\theta + 4(\delta_m/c)^2}\right). \tag{4}$$

The final beam is derived by averaging the aligned signals received by the entire array:

$$\Phi(t;\theta) = \frac{1}{M}\sum_{m=1}^{M}\hat{\varphi}_m(t;\theta). \tag{5}$$

B. Fourier Domain Beamforming

Applying appropriate time delays necessitates high sampling rates, 4-10 times higher than the Nyquist rate of the received signals. To address this problem, it was shown in [7] that beamforming can be implemented equivalently in the frequency domain, bypassing the need for oversampling as no shifts are actually performed.

Let $c[k]$ be the kth Fourier series coefficient of the beam $\Phi(t;\theta)$. Then one can write

$$c[k] = \frac{1}{M}\sum_{m=1}^{M}\hat{c}_m[k] \tag{6}$$

where $\hat{c}_m[k]$ is given by

$$\hat{c}_m[k] = \frac{1}{T}\int_0^T I_{[0,T_B(\theta)]}(t)\,\hat{\varphi}_m(t;\theta)\,e^{-i(2\pi/T)kt}dt. \tag{7}$$

Here, $I_{[0,T_B(\theta)]}$ is an indicator function for the beam's support and T is defined by the penetration depth of the transmitted pulse.

Following the derivation in [7], [8], $\hat{c}_m[k]$ can be expressed as

$$\hat{c}_m[k] = \sum_n c_m[k-n]\,Q_{k,m,\theta}[n] \tag{8}$$

where $c_m[k]$ are the Fourier coefficients of the signal received in the m^{th} element with no time-alignment applied to it, and $Q_{k,m,\theta}[n]$ are the Fourier coefficients of a distortion function $q_{k,m}(t;\theta)$, that effectively transfers the beamforming delays defined in (4) to the frequency domain. The function $q_{k,m}(t;\theta)$ depends on the geometry of the array alone, and therefore, its Fourier coefficients can be computed offline and stored in memory.

Since most of the energy of $\{Q_{k,m,\theta}[n]\}$ is concentrated around the DC component, the infinite sum in (8) can be approximated sufficiently with the finite sum

$$\hat{c}_m[k] \cong \sum_{n\in\nu(k)} c_m[k-n]\,Q_{k,m,\theta}[n] \tag{9}$$

where $\nu(k)$ depends on the decay properties of $\{Q_{k,m,\theta}[n]\}$.

Integrating (9) into (6) yields the beamformed signal in the frequency domain:

$$c[k] \cong \frac{1}{M}\sum_{m=1}^{M}\sum_{n\in\nu(k)} c_m[k-n]\,Q_{k,m,\theta}[n]\,. \tag{10}$$

and applying an inverse Fourier transform on $\{c[k]\}$ results in the beamformed signal in time. The relationship proves that the Fourier coefficients of the beam can be obtained as a linear combination of the Fourier coefficients of the non-delayed received signals. Therefore, it is possible to transfer the process of beamforming to the frequency domain while yielding similar results and avoiding over-sampling.

The required set of Fourier coefficients of the received signals can be obtained in hardware using low-rate sampling, significantly lower than the rate required for time-domain beamforming [11], [37]. Further reduction to a sub-Nyquist rate is achieved by obtaining only a subset of the coefficients, resulting in a subset of the beam's coefficients. In this scheme, however, inverse Fourier transform does not sufficiently recover the beamformed signal in time, and additional methods are required for full recovery. In practice, we obtained the required partial set of coefficients by digitally filtering the fully-sampled signals.

As is indicated in (10), Fourier domain beamforming incorporates DAS beamforming of the partial data and thus yields similar resolution. To improve on this, one would prefer to replace the fixed weighting with adaptive weights; However, those are calculated per scanned depth and transmission angle, and therefore cannot be calculated from partial measurements. To overcome this challenge, we introduce a deep-learning based approach, substituting adaptive beamforming on the sub-sampled grid by learning a direct transformation from it to a high-quality B-mode image. This allows the reduction of data volume, processing-time and hardware requirements on the one hand, while maintaining the quality of adaptive beamforming on the other hand.

III. Method

A. Data sub-sampling and pre-processing

Our pre-processing pipeline consists of digitally emulating temporal sub-sampling with two sampling factors, each constituting a distinct dataset, and applying time-of-flight (TOF) correction in the frequency domain. This is done by multiplication with the Fourier coefficients $\{Q_{k,m,\theta}[n]\}$ (9), following the method in [7]. From this point, we assume that we are given only the temporally sub-sampled data. Then, each dataset is transformed back to the time domain by restoring the negative spectrum of each signal (i.e. its transposed-conjugate, since the signals are real-valued), padding with an appropriate-sized vector of zeros to maintain the original resolution in time, and performing an inverse discrete Fourier transform. Another dataset is generated by spatially sub-sampling the denser sub-Nyquist dataset of the two, following the sparse arrays approach presented in [13]. However, no convolution is performed as opposed to [13], as it did not result in improved results.

Since no additional reconstruction is performed on the under-sampled data, transforming it back to the time domain introduces aliasing artifacts as well as degraded resolution. However, it allows us to model recovery and beamforming as a standard computer-vision problem, and to avoid the difficulties of designing a network that operates in the frequency domain and works with complex numbers. Our goal is to use this sub-sampled and degraded channel data to recover a high-quality B-mode image.

The pre-processed data, separated to three datasets according to sampling factor, is introduced to the network in 3D data cubes in order to exploit correlation in all three dimensions, i.e. between signals of neighboring elements and adjacent transmission angles, as well as in depth. Each sample consists of all scanned depths, per 3 consecutive transmission angles and all transducer elements; Slicing along the second

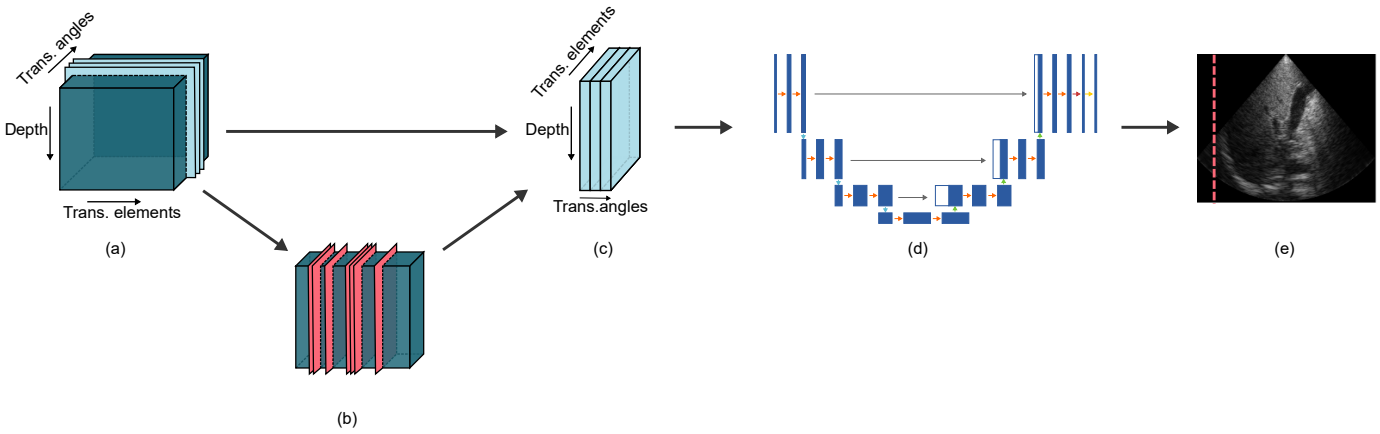


Fig. 1: Overview of the proposed method's pipeline. (a) Slicing the temporally sub-sampled data to blocks, each containing data from three consecutive transmission angles. (b) Spatial sub-sampling of 27 out of 64 transducer elements. (c) The resulting data cube, which formulates one sample in the training phase. (d) The data passes through a CNN (Fig. 2), which predicts one line of the final B-mode image (e).

dimension alleviates the memory requirements and reduces the amount of data which is required for training. Each target consists of the beamformed signal in the middle angle of the corresponding input sample (Fig. 1).

B. Network architecture

The desired output of our deep-beamformer is an enhanced B-mode image; Hence, we do not necessarily require an exact recovery of each sub-sampled signal prior to beamforming. In other words, we would rather let the network learn a non-exact recovery from the partial samples of each signal to the original signal, if such is able to produce a better "beamformed" image. To allow that relaxation, we tackle the recovery from partial samples and beamforming jointly in a single model, instead of handling each stage separately.

Our network is a modified UNet [38]. It consists of three blocks in the contractive and expansive paths, and one bottleneck block. Corresponding blocks are connected with skip connections, i.e. concatenation along the third dimension. Each block consists of two 3x3 convolutions, followed by either a max-pooling layer for a contractive block or an upsampling layer for an expansive block. Both operate along the first dimension. Reduction in the second dimension to a beamformed signal in one direction is achieved with a final summation layer (Fig. 2). Since channel data contains non-trivial negative values, we use Parametric ReLU [39] as an activation function to avoid many 'dying' nodes and significant loss of information.

C. Loss function

Following [22], we use signed-mean-squared-logarithmic-error (SMSLE) as a loss function in order to better represent deviation from the desirable beam given the large dynamic range of the RF data. The loss function is defined as

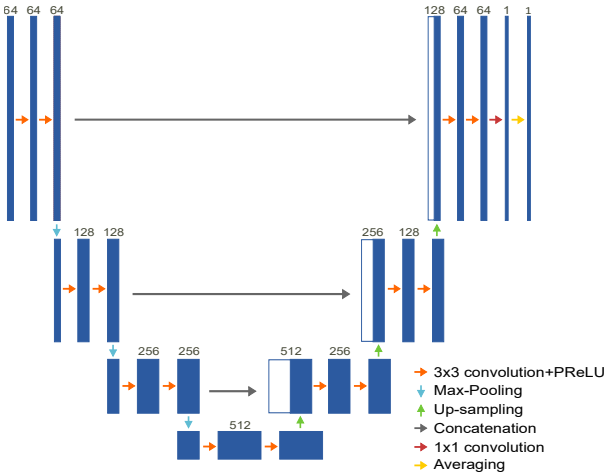


Fig. 2: Architecture of the proposed network. The number of channels is indicated above the blue boxes.

$$L_{SMSLE} = 0.5 \cdot \| \log_{10}(B^{+}_{Proposed} - \log_{10}(B^{+}_{MV}) \|_2^2 + 0.5 \cdot \| \log_{10}(B^{-}_{Proposed} - \log_{10}(B^{-}_{MV}) \|_2^2$$

where B^{+}, B^{-} are the positive and negative parts of the beamformed data, $B_{Proposed}$ is our network's prediction and B_{MV} is the MV target.

D. Metrics for evaluation

To the best of our knowledge, no other approach was proposed for deep-learning based reconstruction from temporally sub-sampled data. Therefore, we cannot directly compare our model to other state-of-the-art approaches. Instead, we concentrate on quantitative evaluation to demonstrate the benefits of our model in terms of resolution and contrast. Contrast-to-Noise-Ratio (CNR) is used to evaluate contrast, and Full-Width-at-Half-Maxima (FWHM), calculated over simulated

point scatterers and *in-vivo* data, is used to evaluate axial and lateral resolution. Overall similarity to the target image is evaluated using Structural Similarity Index (SSIM).

CNR is evaluated over beamformed phantom scans after envelope detection and logarithmic compression. It is calculated from two regions of different intensities in each image, namely, a simulated cyst and its background, and is given by

$$CNR = \frac{\mid \mu_c - \mu_b \mid}{\sigma_b}$$

where μ_c, μ_b, and σ_b are the means and standard deviation of the cyst and the background, respectively.

SSIM is evaluated for each sub-sampling factor separately, in reference to the fully-sampled MV beamformed target images, and is defined for two images x, y as

$$SSIM\left(x,y\right) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}.$$

Here, $\mu_x, \mu_y, \sigma_x, \sigma_y$ and σ_{xy} are the means, standard deviations and cross-correlation of the two images, calculated over a sliding window, and C_1, C_2 are constants meant to stabilize the division.

IV. Experimental Setup

In-vivo data for training and testing was acquired by scanning healthy volunteers, using a P4-2v Verasonics phased-array transducer with 64 elements. The dataset consists of organs of the abdominal cavity - liver, gallbladder, bladder, kidneys - and the Aorta. The carrier frequency was 2.7 MHz and the sampling rate was 10.9 MHz, which is twice the Nyquist rate, resulting in 1918 samples per image line. Two temporally sub-sampled datasets were generated from it using the described scheme: In the first, 400 samples per image line were sampled, resulting in sampling rate of 0.42 of the Nyquist rate and 5-fold reduction in comparison to the original data volume. In the second, 220 samples per image line were sampled, resulting in a sampling rate of 0.23 of the Nyquist rate and 9-fold reduction in comparison to the original data volume. A third dataset was generated by further sampling the denser sub-Nyquist dataset out of the two, omitting 37 out of the 64 transmitting elements, which results in 11-fold reduction in comparison to the original data volume.

Data from two volunteers was used for training and validation in an 80%-20% split, while data from the third volunteer was set aside for testing purposes. This results in 199 training frames, translating to 25,472 training samples. No organ-wise division was performed during training or testing, meaning that both stages operate on data from multiple organs. Targets for training were generated from MV beamforming of the fully-sampled data, time-aligned in the time domain.

For quantitative evaluation of contrast, tissue mimicking phantoms Gammex 403GSLE and 404GSLE were scanned by the 64-element phased array transducer P4-2v with the same transmission specifications as the *in-vivo* dataset.

The network was implemented with Keras, using Tensorflow backend. It was trained separately on each sampling factor for less than 200 epochs using Adam optimizer, with an initial learning rate of $3 \cdot 10^{-5}$. Weights were initialized using He Normal initialization [39].

V. Results

Table I presents quantitative evaluation of the lateral and axial resolution of the proposed method, over simulated point scatterers and *in-vivo* data. As can be seen, our method produces high resolution results on the *in-vivo* data, better than those achieved by NESTA, which is used in [7] and [13] for reconstruction, and even better than DAS beamforming of the fully-sampled data. In fact, they are comparable to MV-beamforming of the fully-sampled data. High axial resolution is indicated also by the results over the simulated frames. However, over the x9 sampling factor faded halos are detected along the axial direction, and lateral resolution is slightly degraded in comparison to DAS of the fully-sampled data; We believe that these results stem from the fact that the network was not trained on synthetic data. Beamformed images of simulated point scatterers, obtained by all approaches, are presented in Fig. 3.

TABLE I: Resolution Evaluation with Full Width at Half Maxima (FWHM) Parameter

	Simulated Scatterers		***In-Vivo* images**	
	lateral	**axial**	**lateral**	**axial**
DAS $_{\text{fully-sampled}}$	2.05	2.16	5.11	3.56
MV $_{\text{fully-sampled}}$	1.46	1.87	2.4	1.7
NESTA $_{\text{x5 reduction}}$	2.72	1.81	5.7	4.3
NESTA $_{\text{x9 reduction}}$	3.55	2.69	5.8	4.8
NESTA $_{\text{x11 reduction}}$*	1.96	1.99	3.89	2.95
Proposed $_{\text{x5 reduction}}$	**2.15**	**1.26**	**2.88**	**1.93**
Proposed $_{\text{x9 reduction}}$	**2.18**	**1.37**	**2.51**	**1.44**
Proposed $_{\text{x11 reduction}}$	**2.26**	**1.5**	**2.74**	**1.88**

* NESTA $_{\text{x11 reduction}}$ is applied after convolving the sub-sampled data according to the convolutional beamforming (COBA) framework, as opposed to the proposed method.

Table II presents quantitative evaluation of contrast and similarity to the target images, assessed on the phantom and *in-vivo* scans, respectively. As can be seen, our method yields the best SSIM score at the higher temporal sampling rate, with and without spatial sub-sampling, indicating good resemblance to the fully-sampled MV-beamformed targets. It also surpasses MV in terms of contrast, but is slightly degraded in comparison to DAS and NESTA. This may result from the improvement in resolution, since bright reflectors in the predicted images are more focused than in the DAS and NESTA images, leading to granularity. Comparison views of the beamformed cyst are presented in Fig. 5, whereas axial and lateral cross-sections are presented in Fig. 6. The strong resemblance to the target images indicates overall good generalization abilities, since training was not performed on in-silico data.

Fig. 4 presents temporally and spatially sub-sampled *In-vivo* data, beamformed by the proposed method and compared to the other approaches. Over the temporally denser datasets, our method yields results that are visually comparable to the target images in terms of both contrast and resolution.

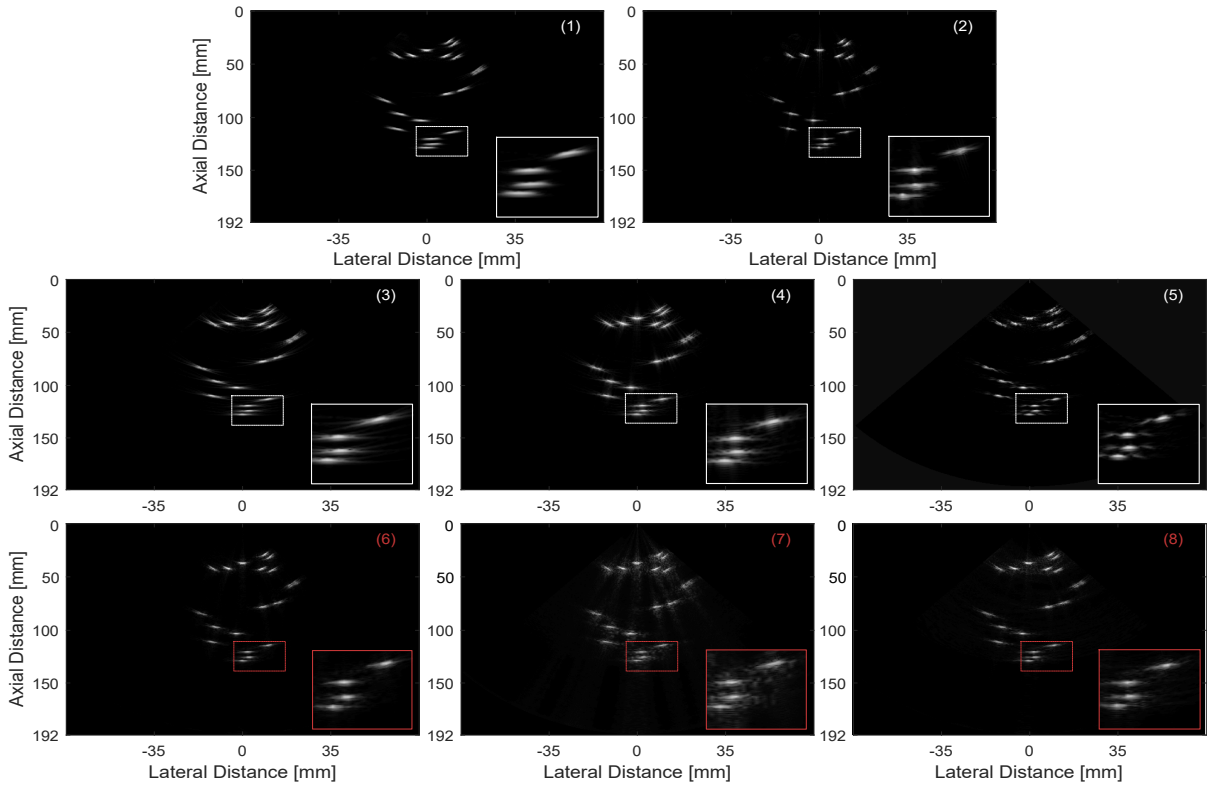


Fig. 3: Simulated point scatterers, obtained by (1) DAS (fully-sampled), (2) MV (fully-sampled), (3) NESTA (0.42 Nyquist, full array), (4) NESTA (0.23 Nyquist, full array), (5) NESTA (0.42 Nyquist, sparse array of 27 out of 64 elements), (6) proposed method (0.42 Nyquist, full array), (7) proposed method (0.23 Nyquist, full array), and (8) proposed method (0.42 Nyquist, sparse array of 27 out of 64 elements). In (5), NESTA is applied after convolving the sub-sampled data according to the convolutional beamforming (COBA) framework, as opposed to the proposed method.

TABLE II: Contrast and Structural Similarity Index (SSIM) Evaluation

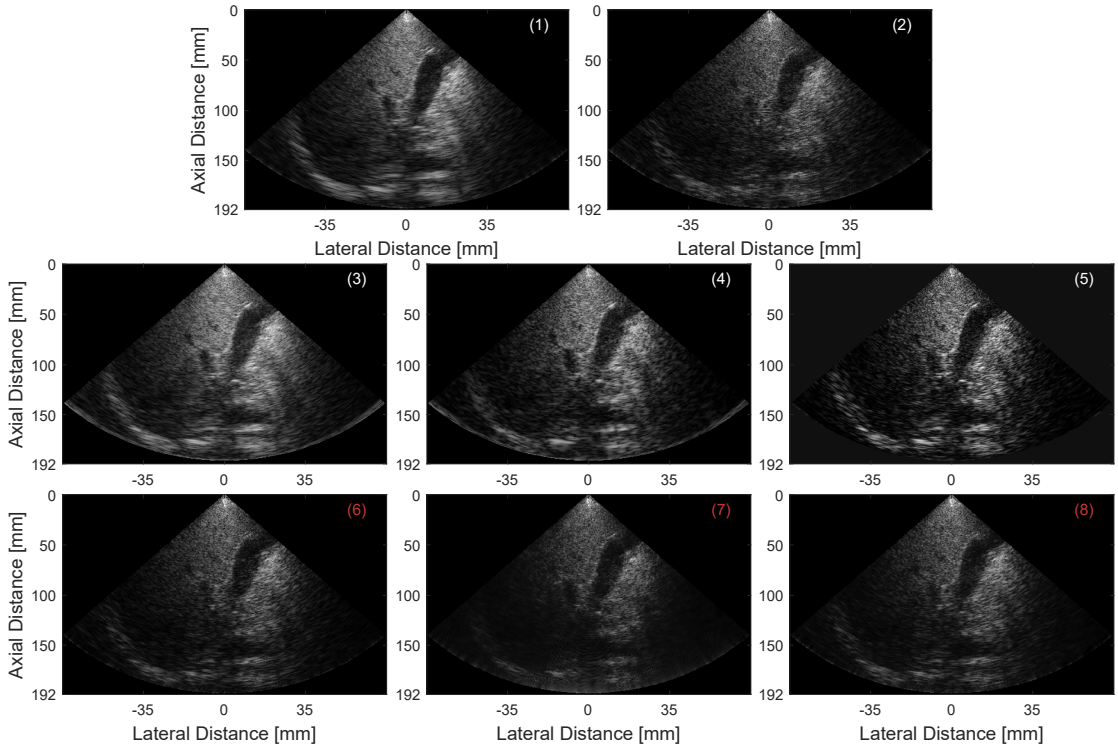
	CNR	**SSIM**
DAS $_{\text{fully-sampled}}$	3.16	0.87
MV $_{\text{fully-sampled}}$	1.94	1
NESTA $_{\text{x5 reduction}}$	3.24	0.72
NESTA $_{\text{x9 reduction}}$	2.67	0.74
NESTA $_{\text{x11 reduction}}$*	2.08	0.35
Proposed $_{\text{x5 reduction}}$	**2.1**	**0.9**
Proposed $_{\text{x9 reduction}}$	**2.1**	**0.8**
Proposed $_{\text{x11 reduction}}$	**1.98**	**0.88**

Moreover, there is good preservation of speckles and weaker reflections. In the further reduced temporal sampling rate, however, strong suppression of weaker reflections, degraded lateral resolution and slight aliasing artifacts are detected in some of the scans, especially in low SNR areas. The effect of these artifacts is alleviated when the network is trained over permuted inputs, such that its second dimension contains the transducer elements plane, its depth contains the transmission angles plane, and pooling is performed over both depth and elements dimensions. Similar improvement is detected over the simulated data, indicated by reduction of the halos along the axial direction. However, such adjustment does not address the suppression of weaker reflections. It has been shown in [7] and [15] that partial data sampled at this low rate can be reconstructed sufficiently; Therefore, we believe that introduction of single image super resolution (SISR) techniques into our model may improve its robustness at lower sampling rates. However, this remains outside the scope of this work.

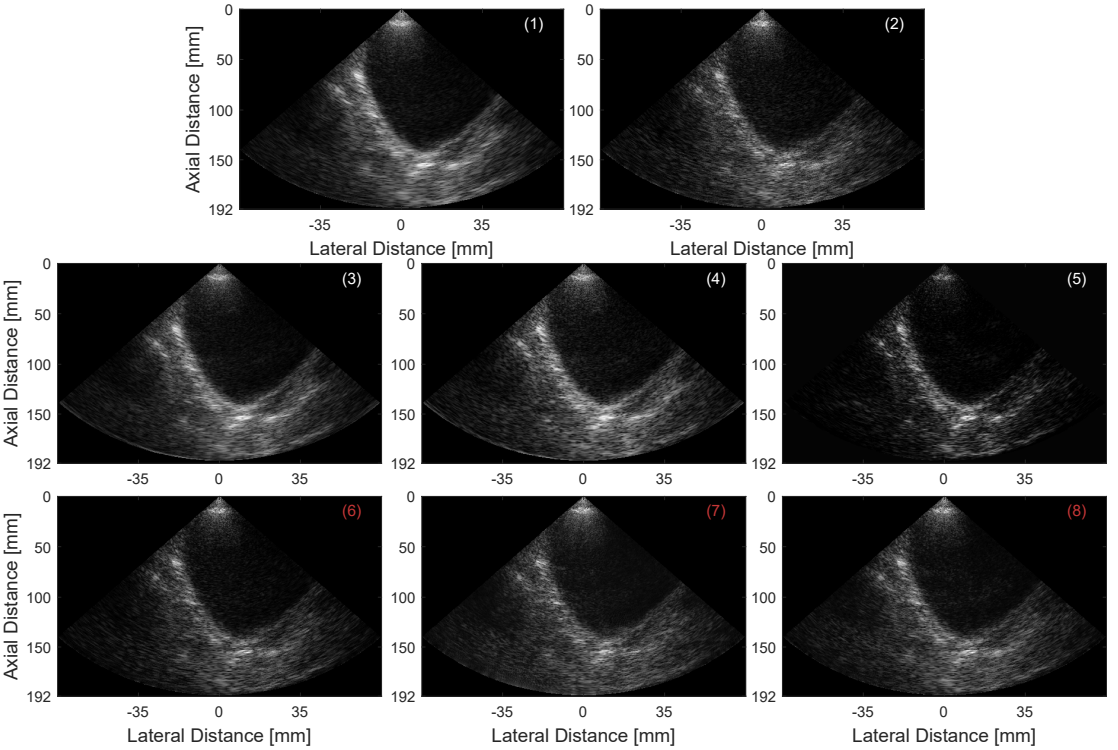
We further compare runtimes for NESTA and our approach. Timings were conducted for time-aligned and sub-sampled data at x5 factor, with NESTA implemented in Matlab and executed on a Dual 20-Core Intel Xeon E5 2698 v4 CPU @ 2.2 GHz, and the proposed method executed on a Tesla V100-SXM2-32GB GPU. In inference, our approach introduces an improvement factor of almost x10 over NESTA, making it more applicable to real-time applications.

VI. Conclusion

In this work we presented a deep-learning based method for high quality reconstruction of temporally and spatially sub-sampled channel data, obtained by the schemes presented in [7] and [13]. We have shown that an encoder-decoder CNN trained on multi-organ scans can be used to directly learn the production of high-quality B-mode images from



(a)



(b)

Fig. 4: Abdominal cavity images, obtained by (1) DAS (fully-sampled), (2) MV (fully-sampled), (3) NESTA (0.42 Nyquist, full array), (4) NESTA (0.23 Nyquist, full array), (5) NESTA (0.42 Nyquist, sparse array of 27 out of 64 elements), (6) proposed method (0.42 Nyquist, full array), (7) proposed method (0.23 Nyquist, full array), and (8) proposed method (0.42 Nyquist, sparse array of 27 out of 64 elements). In (5), NESTA is applied after convolving the sub-sampled data according to the convolutional beamforming (COBA) framework, as opposed to the proposed method.

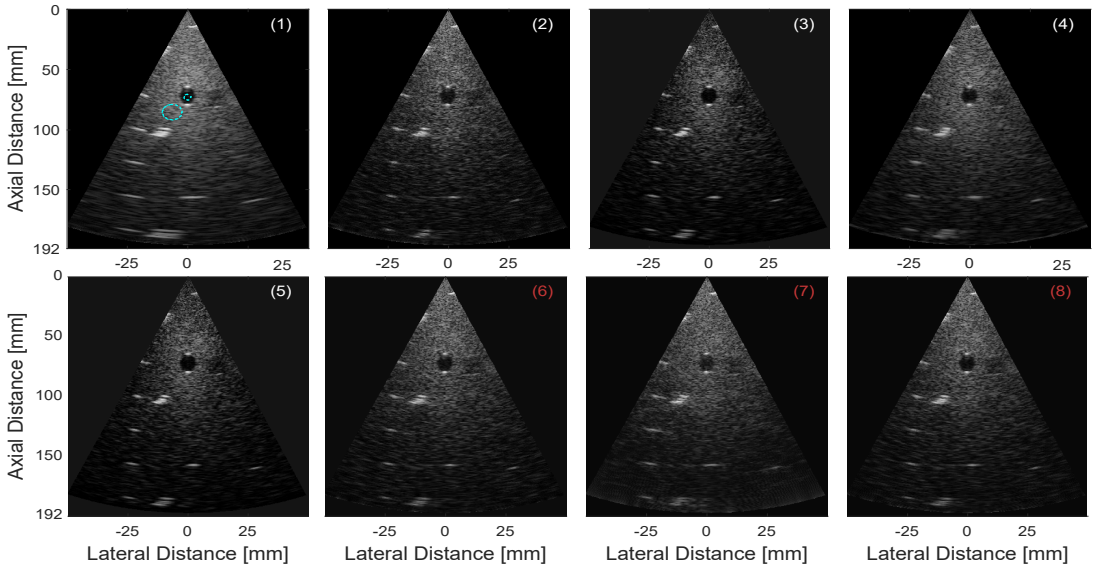
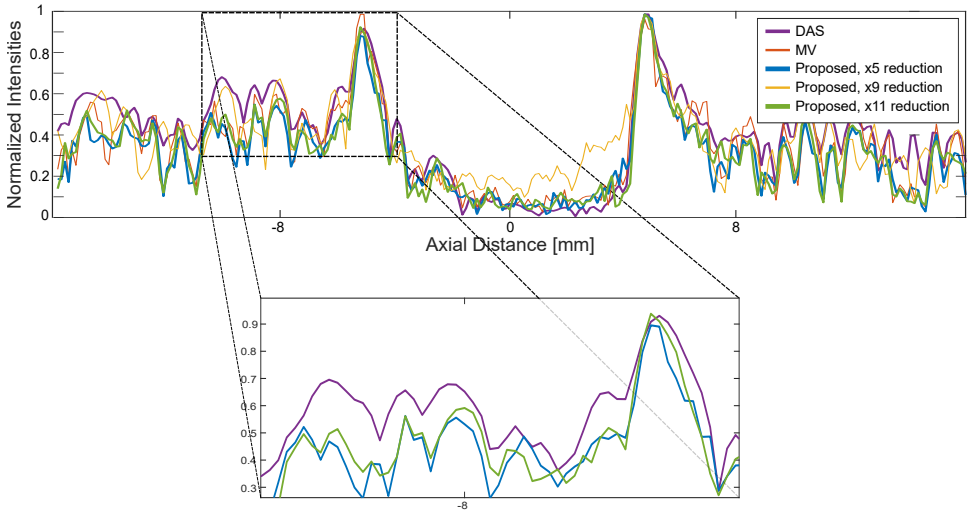
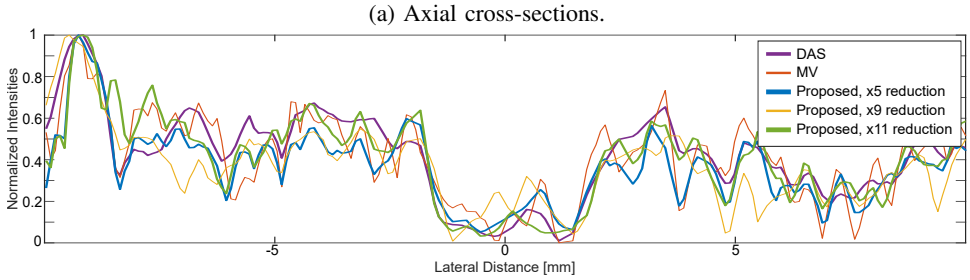


Fig. 5: Images of a simulated anechoic cyst phantom, obtained by (1) DAS (fully-sampled), (2) MV (fully-sampled), (3) NESTA (0.42 Nyquist, full array), (4) NESTA (0.23 Nyquist, full array), (5) NESTA (0.42 Nyquist, sparse array of 27 out of 64 elements), (6) proposed method (0.42 Nyquist, full array), (7) proposed method (0.23 Nyquist, full array) and (8) proposed method (0.42 Nyquist, sparse array of 27 out of 64 elements). In (5), NESTA is applied after convolving the sub-sampled data according to the convolutional beamforming (COBA) framework, as opposed to the proposed method. The dashed circles indicate the regions used for computing the CNRs.



(a) Axial cross-sections.



(b) Lateral cross-sections.

Fig. 6: Axial and lateral cross-sections of the beamformed cyst, obtained by DAS (fully-sampled), MV (fully-sampled) and the proposed method (sub-sampled temporally and spatially). As can be seen, the proposed method achieves higher resolution in comparison to DAS over the higher temporal sampling rate, with and without spatial sub-sampling.

the sub-sampled and degraded channel data, targeting both the recovery from partial frequency data and the learning of adaptive beamforming. Our method yields high resolution results, outperforming the iterative reconstruction method used in [7] and [13] in terms of resolution and runtime. More importantly, it outperforms DAS beamforming and is comparable to MV beamforming of the fully-sampled data, with a reduction of up to x5 in sampling rate and x2 in the number of transducer elements. These results indicate that our method can be plugged in to previously proposed schemes of sub-Nyquist and sparse-array ultrasound processing, to improve performance while significantly alleviating the computational burden and memory and hardware requirements of the data transmitting and processing.

ACKNOWLEDGMENT

The authors thank Dr. I. Aharony for performing the volunteer scans which provided the data for training and testing of the proposed method.

REFERENCES

[1] B. D. Steinberg, "Digital Beamforming in Ultrasound," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 39, no. 6, pp. 716–721, Nov. 1992.

[2] C. E. Shannon, "Communication in the Presence of Noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[3] Y. C. Eldar, "Signal spaces," in *Sampling Theory: Beyond Bandlimited Systems*, 1st ed. U.K.: Cambridge Univ. Press, 2015, ch. 4, pp. 98–99.

[4] K. E. Thomenius, "Evolution of ultrasound beamformers," in *IEEE Int. Ultrason. Symp (IUS)*, vol. 2, San Antonio, TX, USA, Nov. 1996, pp. 1615–1622.

[5] J. Capon, "High-Resolution Frequency-Wavenumber Spectrum Analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[6] T. Chernyakova, D. Cohen, M. Shoham, and Y. C. Eldar, "iMAP Beamforming for High Quality High Frame Rate Imaging," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 66, no. 12, pp. 1830–1844, Dec. 2018.

[7] T. Chernyakova and Y. Eldar, "Fourier-domain beamforming: The path to compressed ultrasound imaging," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 61, no. 8, pp. 1252–1267, Aug. 2014.

[8] N. Wagner, Y. C. Eldar, and Z. Friedman, "Compressed beamforming in ultrasound imaging," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4343–4657, Sept. 2012.

[9] Y. C. Eldar and G. Kutyniok, "Sampling at the rate of innovation: theory and applications," in *Compressed Sensing: Theory and Applications*, 1st ed. U.K.: Cambridge Univ. Press, 2012, ch. 4, pp. 148–209.

[10] Y. C. Eldar, "Compressed sensing," in *Sampling Theory: Beyond Bandlimited Systems*, 1st ed. U.K.: Cambridge Univ. Press, 2015, ch. 11, pp. 392–465.

[11] R. Tur, Y. C. Eldar, and Z. Friedman, "Innovation rate sampling of pulse streams with application to ultrasound imaging," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1827–1842, Apr. 2011.

[12] Y. C. Eldar, "Finite rate of innovation sampling," in *Sampling Theory: Beyond Bandlimited Systems*, 1st ed. U.K.: Cambridge Univ. Press, 2015, ch. 15, pp. 744–749.

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

[14] ——, "Sparse Array Design via Fractal Geometries," 2020. [Online]. Available: http://arxiv.org/abs/2001.01217

[15] A. Mamistvalov and Y. C. Eldar, "Sparse convolutional beamforming for wireless ultrasound," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 9254–9258.

[16] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[17] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016.

[18] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep Learning for Image Super-resolution: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020, DOI: 10.1109/TPAMI.2020.2982166.

[19] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[20] B. Sahiner *et al.*, "Deep learning in medical imaging and radiation therapy," *Med. Physics*, vol. 46, no. 1, pp. e1–e36, Jan. 2019.

[21] R. J. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *Proc. IEEE*, vol. 108, no. 1, pp. 11–29, Jan. 2019.

[22] B. Luijten *et al.*, "Adaptive ultrasound beamforming using deep learning," *IEEE Trans. Med. Imaging*, 2020, DOI: 10.1109/TMI.2020.3008537.

[23] S. Khan, J. Huh, and J. C. Ye, "Universal Deep Beamformer for Variable Rate Ultrasound Imaging," 2019. [Online]. Available: http://arxiv.org/abs/1901.01706

[24] ——, "Adaptive and Compressive Beamforming Using Deep Learning for Medical Ultrasound," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 67, no. 8, pp. 1558–1572, Aug. 2019.

[25] W. Simson *et al.*, "End-to-End Learning-Based Ultrasound Reconstruction," 2019. [Online]. Available: http://arxiv.org/abs/1904.04696

[26] R. Zhuang and J. Chen, "Deep learning based minimum variance beamforming for ultrasound imaging," in *Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis*, vol. 11798. Springer, 2019, pp. 83–91.

[27] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye, "Efficient B-mode Ultrasound Image Reconstruction from Sub-sampled RF Data using Deep Learning," *IEEE Trans. Med. Imaging*, vol. 38, no. 2, pp. 325–336, Feb. 2017.

[28] M. U. Ghani, F. Can Meral, F. Vignon, and J. L. Robert, "High Frame-Rate Ultrasound Imaging Using Deep Learning Beamforming," in *IEEE Int. Ultrason. Symp. (IUS)*, Glasgow, UK, 2019, pp. 295–298.

[29] 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.

[30] W. Simson, M. Paschali, N. Navab, and G. Zahnd, "Deep Learning Beamforming for Sub-Sampled Ultrasound Data," in *IEEE Int. Ultrason. Symp (IUS)*, Kobe, Japan, 2018, pp. 1–4.

[31] O. Senouf *et al.*, "High frame-rate cardiac ultrasound imaging with deep learning," in *Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, Granada, Spain, 2018, pp. 126–134.

[32] S. Vedula *et al.*, "High quality ultrasonic multi-line transmission through deep learning," in *Mach. Learn. for Med. Image Reconstruction (MLMIR)*, Granada, Spain, 2018, pp. 147–155.

[33] D. Perdios, M. Vonlanthen, A. Besson, F. Martinez, M. Arditi, and J. P. Thiran, "Deep Convolutional Neural Network for Ultrasound Image Enhancement," in *IEEE Int. Ultrason. Symp. (IUS)*, Kobe, Japan, 2018, pp. 1–4.

[34] A. A. Nair, T. D. Tran, A. Reiter, and M. A. Bell, "A Generative Adversarial Neural Network for Beamforming Ultrasound Images: Invited Presentation," in *Proc. 53rd Annu. Conf. Inf. Sci. Syst. (CISS)*, Baltimore, MD, USA, 2019, pp. 1–6.

[35] W. Choi, M. Kim, J. Haklee, J. Kim, and J. Beomra, "Deep CNN-Based Ultrasound Super-Resolution for High-Speed High-Resolution B-Mode Imaging," in *IEEE Int. Ultrason. Symp. (IUS)*, Kobe, Japan, 2018, pp. 1–4.

[36] I. A. M. Huijben, B. S. Veeling, K. Janse, M. Mischi, and R. J. G. van Sloun, "Learning Sub-Sampling and Signal Recovery with Applications in Ultrasound Imaging," 2019. [Online]. Available: http://arxiv.org/abs/1908.05764

[37] E. Baransky, G. Itzhak, N. Wagner, I. Shmuel, E. Shoshan, and Y. Eldar, "Sub-nyquist radar prototype: Hardware and algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 2, pp. 809–822, Apr. 2014.

[38] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, Munich, Germany, 2015, pp. 234–241.

[39] H. Kaiming, Z. Xiangyu, R. Shaoqing, and S. Jian, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, 2015, pp. 1026–1034.