On Problem Formulation, Efficient Modeling and Deep Neural Networks for High-Quality Ultrasound Imaging

Invited Presentation

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

*Signal Processing Laboratory (LTS5), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

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

***Abstract*—Recently, many pulse-echo ultrasound (US) imaging methods have relied on the transmission of unfocused wavefronts. Such a strategy allows for very high frame rates at the cost of a degraded image quality. In this work, we present a regularized inverse problem approach and a highly efficient modeling of the physical measurement process to reconstruct high-quality US images from unfocused wavefronts. We compare it against a deep neural network (DNN) approach on the plane wave imaging challenge in medical ultrasound (PICMUS) and show that the use of carefully designed and trained DNN can overcome the limitations of standard image processing priors, which fail at capturing the very specific nature of US images accurately.**

***Index Terms*—Ultrasound imaging, inverse problems, image reconstruction, deep learning.**

I. Introduction

ULTRAFAST ultrasound (US) imaging has undoubtedly revolutionized the way of acquiring US images. Instead of transmitting a succession of focused beams for each image, ultrafast US imaging relies on the transmission of unfocused wavefronts, such as plane waves (PW) or diverging waves (DW). This strategy has allowed pulse-echo US imaging to reach frame rates of multiple kHz, opening up US to new diagnostic imaging modes, such as shear wave elastography, ultrafast vector flow and functional US imaging [1].

Compared to focused transmit beams, ultrafast US imaging generally suffers from a lower image quality. A popular approach to increase the quality of images obtained with ultrafast US imaging consists in averaging consecutive images reconstructed from differently steered wavefronts, in a process called coherent compounding [1], at the cost of reduced frame rates and increased data transfer rates.

Conventionally, US images are reconstructed by delay-and-sum (DAS) beamforming. Recently, many approaches have been proposed to increase the resulting image quality. Among those technique, the group of methods which express the image reconstruction process as an inverse problem is of particular interest [2]–[5]. These methods usually require a measurement model, as well as a prior for solving the image reconstruction problem using iterative algorithms. They mainly suffer from the heaviness of the measurement model involved in US imaging, especially when stored as a matrix, and from priors that do not accurately capture the specific statistics inherent to US images. Moreover, the iterative nature of the solvers involved in these approaches makes real-time imaging a serious challenge.

Recently, deep learning approaches have been shown to be highly promising for US image reconstruction and post-processing [6]–[10]. The potential of these methods is certainly immense. Yet, a clear understanding and robustness guarantees are still lacking, which hinder their adoption in specific fields such as medical imaging.

In this work, we briefly detail an efficient modeling of the physical measurement process involved in the problem of US image reconstruction. We present an iterative approach where imaging is performed by sparse regularization, and the sparsity prior is expressed in the sparsity averaging model. We also present a deep learning approach for solving the US image reconstruction problem. We compare both approaches on the plane wave imaging challenge in medical ultrasound (PICMUS) [11] metrics, and discuss their main advantages and drawbacks.

The remainder of this paper is organized as follows. Section II presents the problem formulation of US image reconstruction as well as an efficient modeling of the measurement process. Section III describes two approaches for solving this reconstruction problem. Experiments and results are detailed in Section IV, and concluding remarks are given in Section V.

II. Problem Formulation and Efficient Modeling

We consider the 2D pulse-echo US imaging configuration depicted in Fig. 1. An array of N_{el} transducer elements centered at positions $\{\boldsymbol{p}_i\}_{i=1}^{N_{el}}$ is used for transmitting an acoustic wavefront, such as a PW, in a domain of interest $\Omega \in \mathbb{R}^2 \backslash \{z \geq 0\}$ characterized by its tissue reflectivity function $\gamma(\boldsymbol{r})$ with $\boldsymbol{r} \in \Omega$, which accounts for the local fluctuations in acoustic impedance. These fluctuations result in backscatterred acoustic waves which are detected during a period T by the

This work was supported in part by the UltrasoundToGo RTD project (no. 20NA21_145911) funded by Nano-Tera.ch with Swiss Confederation financing. This work was also supported by the Swiss SNF project number 205320_175974.

978-1-7281-1151-3/19/$31.00 ©2019 IEEE

same array of transducers to form a set of measurements $\{m(\boldsymbol{p}_i,t)\}_{i=1}^{N_{el}}$, with $t \in [0,T]$. Using the spatial impulse response model and assuming the pulse-echo waveform to be a Dirac function, one can express the measurements recorded on the i-th element as

$$m(\boldsymbol{p}_i,t) = \int_{\boldsymbol{r}\in\Omega} o(\boldsymbol{p}_i,\boldsymbol{r})\delta(t-\tau(\boldsymbol{p}_i,\boldsymbol{r}))\gamma(\boldsymbol{r})d\boldsymbol{r}, \tag{1}$$

where $o(\boldsymbol{p}_i,\boldsymbol{r})$ accounts for the element directivity and decay of the acoustic wave [5]. Under an ideal wavefront assumption, the round-trip time-of-flight can be expressed as $\tau(\boldsymbol{p}_i,\boldsymbol{r}) = t_{\mathrm{Tx}}(\boldsymbol{r}) - t_{\mathrm{Rx}}(\boldsymbol{p}_i,\boldsymbol{r})$, where $t_{\mathrm{Rx}}(\boldsymbol{p}_i,\boldsymbol{r}) = \|\boldsymbol{p}_i - \boldsymbol{r}\|_2/c$, with c the speed of sound in the medium.

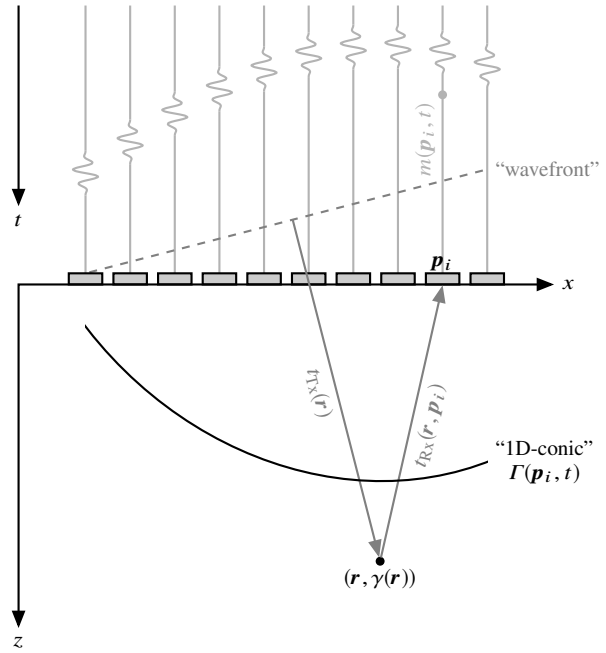

Fig. 1. Standard 2D ultrasound imaging configuration. (Adapted from [5]).

Equation (1) can be rewritten as the curvilinear integral

$$m(\boldsymbol{p}_i,t) = \int_{\boldsymbol{r}\in\Gamma(\boldsymbol{p}_i,t)} \frac{o(\boldsymbol{p}_i,\boldsymbol{r})}{|\nabla_{\boldsymbol{r}} g(\boldsymbol{p}_i,\boldsymbol{r},t)|}\gamma(\boldsymbol{r})d\sigma(\boldsymbol{r}), \tag{2}$$

where the implicit function $g(\boldsymbol{p}_i,\boldsymbol{r},t) = t - t_{\mathrm{Tx}}(\boldsymbol{r}) - t_{\mathrm{Rx}}(\boldsymbol{p}_i,\boldsymbol{r})$ and the curve $\Gamma(\boldsymbol{p}_i,t) = \{\boldsymbol{r}\in\Omega : g(\boldsymbol{p}_i,\boldsymbol{r},t) = 0\}$. It is interesting to note that these curves are, in this case of 2D propagation, 1D conics of different nature depending on the type of transmitted wavefront, and can therefore be efficiently parameterized. For example, in the case of a PW wavefront, the corresponding conic $\Gamma(\boldsymbol{p}_i,t)$ is a parabola.

After discretization, numerical integration and interpolation of (2) for every transducer element position $\boldsymbol{p}_i$, one can express the physical measurement process as

$$\boldsymbol{m} = \boldsymbol{H}\boldsymbol{\gamma} + \boldsymbol{n}, \tag{3}$$

where $\boldsymbol{\gamma} \in \mathbb{R}^{N_x N_z}$ and $\boldsymbol{m} \in \mathbb{R}^{N_{el}N_t}$ are discrete and vectorized representations of the US image and corresponding measurements, $\boldsymbol{H} : \mathbb{R}^{N_x N_z} \to \mathbb{R}^{N_{el}N_t}$ is the measurement model, and $\boldsymbol{n} \in \mathbb{R}^{N_{el}N_t}$ is the measurement noise, with N_t the number of samples recorded on each transducer element and $N_x N_z$ the total number of pixels. The inverse problem associated with (3) can therefore be defined as:

$$\text{Find } \boldsymbol{\gamma} \text{ from } \boldsymbol{m} = \boldsymbol{H}\boldsymbol{\gamma} + \boldsymbol{n}. \tag{4}$$

III. Image Reconstruction Approaches

In this section, we briefly describe two approaches for solving the image reconstruction problem defined in (4).

A. Sparse Regularization

In sparse regularization approaches, (4) is reformulated as the unconstrained minimization problem

$$\min_{\boldsymbol{\gamma}\in\mathbb{R}^{N_x N_z}} \frac{1}{2}\|\boldsymbol{H}\boldsymbol{\gamma} - \boldsymbol{m}\|_2^2 + \lambda\mathcal{R}(\boldsymbol{\gamma}), \tag{5}$$

where $\mathcal{R} : \mathbb{R}^{N_x N_z} \to \mathbb{R}^+$ is a convex regularization term encoding a prior knowledge on the unknown image $\boldsymbol{\gamma}$, and $\lambda \in \mathbb{R}^+$ is a regularization parameter. A solution to this problem can be found by the proximal gradient descent iteration [12]

$$\boldsymbol{\gamma}_{k+1} = \mathrm{prox}_{\lambda\tau\mathcal{R}}(\boldsymbol{\gamma}_k - \tau\boldsymbol{H}^*(\boldsymbol{H}\boldsymbol{\gamma}_k - \boldsymbol{m})), \tag{6}$$

where $\boldsymbol{H}^*$ is the adjoint of $\boldsymbol{H}$ and τ is the gradient step size. The proximity operator involved in (6) is defined as

$$\mathrm{prox}_{\mu\mathcal{R}}(\boldsymbol{z}) = \underset{\boldsymbol{\gamma}\in\mathbb{R}^{N_x N_z}}{\mathrm{argmin}} \|\boldsymbol{\gamma} - \boldsymbol{z}\|_2^2 + \mu\mathcal{R}(\boldsymbol{\gamma}), \tag{7}$$

for some $\mu \in \mathbb{R}^+$.

We express the regularization term as a sparsity prior in a transformed domain $\boldsymbol{\Psi}$, expressed as $\mathcal{R}(\boldsymbol{\gamma}) = \|\boldsymbol{\Psi}^*\boldsymbol{\gamma}\|_1$. For the sparsifying transform, we use a concatenation of 8 Daubechies (from 1 to 8) wavelet transforms expressed as

$$\boldsymbol{\Psi} = \frac{1}{\sqrt{q}}\left[\boldsymbol{\Psi}_1,\ldots,\boldsymbol{\Psi}_q\right]. \tag{8}$$

B. Neural Network

We also consider the approach proposed in our previous work [10] for solving the image reconstruction problem defined in (4). The strategy consists in first reconstructing a low-quality image $\tilde{\boldsymbol{\gamma}}$ using the classical DAS beamforming from a single PW insonification and second infer the low-quality image to a deep neural network (DNN) specifically trained for the purpose of reconstructing a high-quality image $\hat{\boldsymbol{\gamma}}$ from a low-quality image $\tilde{\boldsymbol{\gamma}}$. Let us define the DAS operator as $\boldsymbol{D} : \mathbb{R}^{N_{el}N_t} \to \mathbb{R}^{N_x N_z}$ and a DNN with trainable parameters $\boldsymbol{\theta}$ as $\boldsymbol{f}_{\boldsymbol{\theta}} : \mathbb{R}^{N_x N_z} \to \mathbb{R}^{N_x N_z}$. The proposed strategy can hence be expressed as

$$\hat{\boldsymbol{\gamma}} = \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{D}\boldsymbol{m}). \tag{9}$$

For the network architecture, we use an adaptation of the popular U-Net [13]. It is a residual convolutional neural network, such that $\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{\gamma}) = \boldsymbol{\gamma} + \boldsymbol{r}_{\boldsymbol{\theta}}(\boldsymbol{\gamma})$, capable of predicting the negative noise to be applied to a low-quality image in order to recover a high-quality image. It has an encoder-decoder structure, in which an input image first undergoes

a cascade of downsampling operations in the spatial dimension (together with channel expansion), followed by the corresponding upsampling operations (together with channel contraction). Internal skip connections are used to mitigate the loss of information in the downsampling path. For a detailed description of the network architecture, the reader may refer to [10], [13].

The training dataset is composed of low-quality and high-quality image pairs reconstructed by DAS from simulated data obtained from randomly generated tissue-mimicking phantoms. The low-quality images are reconstructed from a single PW insonification with normal incidence. The high-quality images are reconstructed from synthetic aperture (SA) measurements, considerer as the gold standard in the US community.

IV. Experiments and Results

A. Experimental Settings

We are interested in assessing the reconstructed image quality of the approaches described in Section III from a single PW insonification with normal incidence.

For the sparse regularization approach (SR), we use the popular fast iterative shrinkage thresholding algorithm (FISTA) to solve Problem (5). The algorithm is stopped when the mean squared error (MSE) of the difference between two consecutive solution candidates is smaller than 10^{-3}. The thresholding parameter λ is optimized by grid search.

For the neural network approach, we use the best performing network of our previous work [10]. It has been trained on 28 000 image pairs over 200 000 iterations using the well-know Adam optimizer with a learning rate of 5×10^{-5} and the MSE as the loss function.

The image reconstruction quality is assessed on the PICMUS dataset, and compared to the quality of images reconstructed by the classical DAS beamforming from 1 PW and 75 PWs insonifications. For each reconstructed image, we compute the corresponding B-mode image for visualization purpose by standard envelope detection, normalization and log-compression.

B. Performance Evaluation

We evaluate some of the image quality metrics available for the numerical PICMUS phantom[1]. In particular, we are interested in the contrast-to-noise ratio (CNR) and the averaged lateral and axial resolution (full width half maximum) in the near field (14 mm depth) and in the far field (45 mm depth). Two quality tests are also conducted, namely a speckle test and a linear intensity test. The results for the different methods are reported on Table I and Fig. 2 displays the corresponding B-mode images reconstructed using, from left to right, DAS with 1 PW, SR with 1 PW, U-Net with 1 PW, and DAS with 75 PWs.

It is clear that the NN approach largely surpasses the SR approach on the CNR metric. It even provides a better score than the reference image reconstructed by DAS beamforming from 75 PWs. This effect can be visually appreciated in the circular anechoic region on Fig. 2. One can note that the SR approach fail at entirely removing the serious level of sidelobes present in the image reconstructed by DAS beamforming with 1 PW, whereas the NN approach perfectly reconstructs this region. On the other, the SR method provides in general better scores on the resolution metrics compared to the NN approach. One should however note that the SR approach has a tendency to suffer from a little ripple effect around bright reflectors. This effect is particularly visible on the point reflectors located in the near-field of the numerical phantom (Fig. 2). Both the speckle and linear intensity tests are much better handled by the NN approach.

TABLE I
Image Quality Metrics Comparison on the PICMUS Numerical Phantom

Method	CNR [dB]	Lat. Res. [mm] 14 mm	Lat. Res. [mm] 45 mm	Ax. Res. [mm] 14 mm	Ax. Res. [mm] 45 mm	Speck. Tests	Lin. Test
1 PW + DAS	7.2	0.36	0.53	0.38	0.41	6/6	✓
1 PW + SR	10.6	0.23	0.31	0.33	0.37	3/6	✗
1 PW + U-Net	17.1	0.23	0.38	0.37	0.41	5/6	✓
75 PWs + DAS	16.4	0.32	0.46	0.39	0.40	6/6	✓

Figure 3 provides a visual comparison of an *in vivo* carotid B-mode images reconstructed using, from left to right, DAS with 1 PW, SR with 1 PW, U-Net with 1 PW, and DAS with 75 PWs. One can appreciate the overall image quality increase, mainly visible around the carotid artery walls, from both the SR and NN approach compared to the image reconstructed by DAS beamforming with 1 PW. The NN approach again seems to better perform on *in vivo* images, even though U-Net has been exclusively trained on simulated images. Both approaches fail at reconstructing some very fine details present in the reference image. It is however interesting to note that the NN approach seems capable of accurately reconstructing some details in the near-field region that were indistinguishable in the image reconstructed by DAS with a single PW.

One possible explanation for such an important difference in performance between the considered approaches is that, even though the sparsity averaging prior implemented in the SR method has proved to be one of the most efficient prior in the context of US image reconstruction [5], it cannot efficiently capture the very specific nature inherent to US images. However, it is worth mentioning that the NN approach does not include not explicitly incorporate a robust data-fidelity mechanism as the one included in the SR approach.

On the computational side, the NN approach has a clear advantage since it only requires the reconstruction of a low-quality image by DAS and a single inference, whereas the SR approach requires approximatively 100 iterations to converge.

V. Conclusion and Perspectives

In this study, we formulate the ultrasound (US) image reconstruction process as an inverse problem and present an efficient modeling of the measurement process in the context

[1]https://www.creatis.insa-lyon.fr/EvaluationPlatform/picmus/index.html

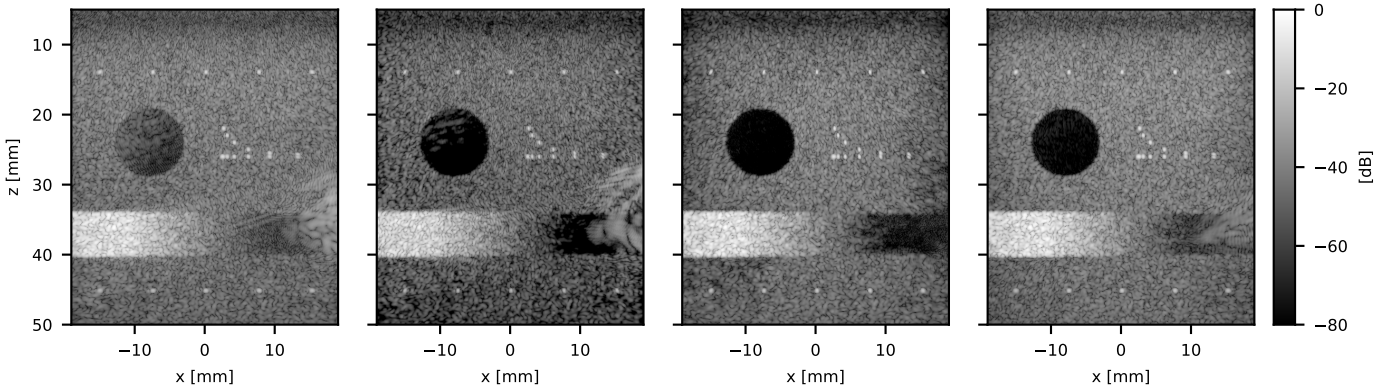

Fig. 2. B-mode images of the PICMUS numerical phantom reconstructed using (from left to right) DAS with 1 PW, SR with 1 PW, U-Net with 1 PW, and DAS with 75 PWs. The images are displayed on a dynamic range of 80 dB.

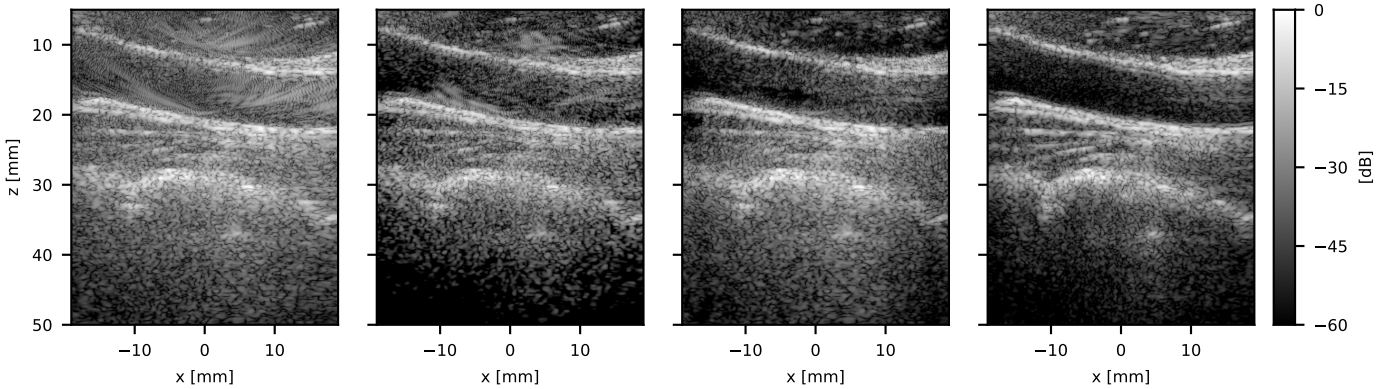

Fig. 3. B-mode images of a longitudinal *in vivo* carotid from the PICMUS dataset reconstructed using (from left to right) DAS with 1 PW, SR with 1 PW, U-Net with 1 PW, and DAS with 75 PWs. The images are displayed on a dynamic range of 60 dB.

ultrafast ultrasound (US) imaging. We propose two distinct approaches for solving such a problem, namely a sparse regularization approach with a prior expressed in the sparsity averaging model, and a deep neural network (DNN) approach. We compare both approaches on the plane wave imaging challenge (PICMUS) and demonstrate the overall superiority of the DNN.

In future works, we will investigate the underlying similarity of both approaches. Indeed, it is known that the delay-and-sum operator $\boldsymbol{D}$ involved in the DNN approach defined in (9) is similar to the adjoint operator of the measurement $\boldsymbol{H}^*$ and that the proximity operator involved in (6) is akin to a denoising operation. Hence, the proposed DNN approach may be interpreted to some extent as the first iteration of the proximal gradient descent iteration.

REFERENCES

[1] M. Tanter and M. Fink, "Ultrafast imaging in biomedical ultrasound," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 61, no. 1, pp. 102–119, 2014.

[2] M. F. Schiffner and G. Schmitz, "Fast pulse-echo ultrasound imaging employing compressive sensing," in *2011 IEEE Int. Ultrason. Symp.* IEEE, 2011, pp. 688–691.

[3] G. David, J.-L. Robert, B. Zhang, and A. F. Laine, "Time domain compressive beam forming of ultrasound signals," *J. Acoust. Soc. Am.*, vol. 137, no. 5, pp. 2773–2784, 2015.

[4] B. Byram, K. Dei, J. Tierney, and D. Dumont, "A model and regularization scheme for ultrasonic beamforming clutter reduction," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 62, no. 11, pp. 1913–1927, 2015.

[5] A. Besson, D. Perdios, F. Martinez, Z. Chen, R. E. Carrillo, M. Arditi, Y. Wiaux, and J.-P. Thiran, "Ultrafast ultrasound imaging as an inverse problem: Matrix-free sparse image reconstruction," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 65, no. 3, pp. 339–355, 2018.

[6] 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, 2017.

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

[8] S. Vedula, O. Senouf, A. M. Bronstein, O. V. Michailovich, and M. Zibulevsky, "Towards CT-quality ultrasound imaging using deep learning," pp. 1–4, 2017. [Online]. Available: http://arxiv.org/abs/1710.06304

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

[10] D. Perdios, M. Vonlanthen, A. Besson, F. Martinez, M. Arditi, and J.-P. Thiran, "Deep convolutional neural network for ultrasound image enhancement," in *2018 IEEE Int. Ultrason. Symp.*, 2018, pp. 1–4.

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

[12] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing." New York, NY, USA: Springer, 2011, pp. 185–212.

[13] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, 2017.