Neural Maximum-a-Posteriori Beamforming for Ultrasound Imaging

Luijten, Ben; Ossenkoppele, Boudewine W. ; de Jong, Nico; Verweij, Martin D.; Eldar, Yonina C.; Mischi, Massimo; van Sloun, Ruud J.G.

DOI
10.1109/ICASSP49357.2023.10096073

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10096073

NEURAL MAXIMUM-A-POSTERIORI BEAMFORMING FOR ULTRASOUND IMAGING

Ben Luijten[1], Boudewine W. Ossenkoppele[2], Nico de Jong[2,3], Martin D. Verweij[2,3]
Yonina C. Eldar[4], Massimo Mischi[1], Ruud J.G. van Sloun[1,5]

1: Dept. of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands.
2: Dept. of Imaging Physics, Delft University of Technology, The Netherlands
3: Dept. of Cardiology, Erasmus MC Rotterdam, The Netherlands
4: Weizmann Institute of Science, Rehovot, Israel
5: Philips Research, Eindhoven

ABSTRACT

Ultrasound imaging is an attractive imaging modality due to its low-cost and real-time feedback, although it often falls short in image quality compared to MRI and CT imaging. Conventional ultrasound image reconstruction, such as Delay-and-Sum beamforming, is derived from maximum-likelihood estimation. As such, no prior information is exploited in the image formation process, which limits potential image quality. Maximum-a-posteriori (MAP) beamforming aims to overcome this issue, but often relies on rough approximations of the underlying signal statistics. Deep learning based reconstruction methods have demonstrated impressive results over the past years, but often lack interpretability and require vast amounts of data.

In this work we present a neural MAP beamforming technique, which efficiently combines deep learning in the MAP beamforming framework. We show that this model-based deep learning approach can achieve high-quality imaging, improving over the state-of-the-art, without compromising the real-time abilities of ultrasound imaging.

***Index Terms*—** Ultrasound, Beamforming, Deep-Learning, Probabilistic modelling

1. INTRODUCTION

Ultrasound imaging is a cost-effective, portable and highly interactive imaging modality that allows visualisation of soft tissue and blood flow inside the body, while being low risk due to the absence of ionizing radiation. As a result ultrasound is used in many medical specialties and different stages of medical treatment, from initial diagnostics, to planning of procedures, monitoring of interventions, and treatment follow-up.

During signal acquisition, an ultrasound transducer is used to transmit and receive acoustic waves. In B-mode (brightness mode) imaging, the received time-domain signals are transformed into a spatial mapping (i.e. beamforming), such that an image of the acoustic reflectivity of the insonified tissue can be reconstructed. Although this signal processing has traditionally been implemented in hardware, a contemporary approach involves the use of software-based processing. This alternative methodology affords the implementation of more complex and adaptive algorithms, yet incurs a substantial computational cost.

The majority of conventional beamforming methods, such as Delay-and-Sum (DAS) [1] or Minimum-Variance (MV) [2, 3], do not exploit prior information about the distribution of the underlying tissue reflectivity, which limits achievable image quality. For comparison, incorporating such prior information is standard in high-quality MRI and CT image reconstruction. Few ultrasound reconstruction methods exist that do exploit prior information, but typically assume very simplified signal models (e.g. independent Gaussians per pixel) [4, 5]. Alternatively, several methods have been proposed that exploit more complex signal structures, however these are often prohibitively slow and complex for real-time imaging [6, 7, 8].

Here we present a novel approach, neural MAP beamforming, which is derived from probabilistic modelling of the beamforming process. Neural MAP provides a way to replace critical parts of MAP estimation by learning them through model-based neural networks [9, 10], aiming to achieve high image quality with a low computational cost. We will first provide an overview of conventional ML and MAP beamforming methods, after which we will detail on the presented neural MAP beamformer. We compare our method with DAS and a deep learning based beamformer [11, 12], and evaluate Peak Signal to Noise Ratio (PSNR) based on varying levels of input noise.

This work was supported in part by the Dutch Research Council (NWO) and Philips Research through the research programme "High Tech Systems and Materials (HTSM)" under Project 17144, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant No. 101000967), and the Israel Science Foundation (grant No. 536/22).

2. BACKGROUND

2.1. Maximum-Likelihood Beamforming

Many classical beamforming methods can be derived from the narrowband linear measurement model with additive noise given by

$$\mathbf{y} = \mathbf{a}_\theta x + \mathbf{n}, \tag{1}$$

where $\mathbf{y} \in R^L$ is a vector of L channel signals, x is the tissue reflectivity, $\mathbf{a}_\theta \in R^L$ is a steering vector and $\mathbf{n} \in R^L$ additive noise. Note that the model above assumes independent sources x (i.e. each pixel is independent), as is typical assumed in beamforming setups. In the broadband pulse-echo ultrasound imaging setting, we first perform a TOF-correction (time-alignment). As a result, the steering vector then simplifies to $\mathbf{a}_\theta = \mathbf{1}$.

In beamforming we aim to find the most likely candidate $\hat{x}$ (per pixel) that explains our measurements $\mathbf{y}$. Without a prior on x, i.e. x is assumed to be a deterministic variable with random noise, we can find the ML estimator. As such we aim to maximize the probability $p(\mathbf{y}|x)$. Maximizing the log-likelihood, or equivalently minimizing the negative log-likelihood, yields

$$\arg\max_x \log p(x|\mathbf{y}) = \arg\min_x (\mathbf{y}-\mathbf{1}x)^T \mathbf{C}^{-1} (\mathbf{y}-\mathbf{1}x), \tag{2}$$

where $\mathbf{C}$ is the noise covariance matrix. This estimator assumes independent sources (pixels in the ultrasound setup), ignoring the strong dependency between pixels in images. We will show in later sections that we can go beyond this, and include spatial information in our estimator.

Standard Delay-and-Sum beamforming can be derived from the ML estimate for x under the assumption of Gaussian noise, where $\mathbf{n} \sim N(0, \sigma^2 I)$ and $p(\mathbf{y}|x) \sim N(\mathbf{1}x, \sigma^2 I)$. From (2) we can then derive the closed-form solution

$$\hat{x} = \frac{1}{L} \mathbf{1}^T \mathbf{y}. \tag{3}$$

Here, L denotes the number of array elements. Simply put, this corresponds to taking the mean over the individual channel signals for every pixel or focus point.

For colored noise $\mathbf{n} \sim N(0, \mathbf{C}^{-1})$, the ML solution corresponds to the weighed sum

$$\hat{x} = (\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{C}^{-1} \mathbf{y} = \mathbf{w}_{\text{MV}}^T \mathbf{y}, \tag{4}$$

where $\mathbf{w}_{\text{MV}}^T$ is a vector of (adaptive) apodization weights. This is also known as Minimum Variance beamforming [2, 3]. Generally speaking, the noise covariance is not known, and is estimated from data following $E[\mathbf{y}\mathbf{y}^T]$. However, this process is error prone, which can strongly affect image quality and may yield an unstable matrix inversion. To overcome these instability issues, several methods have been suggested (e.g. spatial smoothing) [3], however these add computational overhead and can compromise resolution.

2.2. Maximum-a-Posterori Beamforming

If we have a prior $p(x)$, we can maximize the posterior probability

$$p(x|\mathbf{y}) \sim p(\mathbf{y}|x) p(x), \tag{5}$$

which yields the maximum-a-posteriori (MAP) estimate. We can include this prior knowledge on the distribution of x by adding a regularization term to (2) such that

$$\arg\max_x \log p(x|\mathbf{y}) = (\mathbf{y}-\mathbf{1}x)^T \mathbf{C}^{-1} (\mathbf{y}-\mathbf{1}x) - \log p(x). \tag{6}$$

For $x \sim N(0, \sigma_x^2 I)$, and under the same likelihood model as (3) and (4), the general solution is given by

$$\hat{x} = \frac{\sigma_x^2}{\sigma_x^2 + \mathbf{w}_{\text{MV}}^T \mathbf{C} \mathbf{w}_{\text{MV}}} \mathbf{w}_{\text{MV}}^T \mathbf{y} \tag{7}$$

which is known as Wiener beamforming [4]. This general case equates to a postfilter (scaling) of the MV beamformer derived in (4). Similarly, an iterative solution to MAP beamforming, iMAP [5], was proposed by Chernyakova *et al.*, which aims to improve the reconstruction of x by iterating on (7). In iMAP both x and $\mathbf{n}$ are assumed to be white Gaussian variables, where the noise power is estimated through

$$\{\hat{\sigma}_x^2, \hat{\sigma}_n^2\}_{(t)} = \left\{ \hat{x}_t^2, \frac{1}{L} ||\mathbf{y} - \hat{x}_{(t)} \mathbf{1}||^2 \right\}. \tag{8}$$

However, for many relevant prior distributions, the MAP estimator has no closed form solution. Instead, the MAP estimators are typically obtained through iterative methods, such as proximal gradient descent. Proximal gradient descent alternates between a data-consistency (DC) step, which is based on the data likelihood model, followed by a prior (proximal) step. In the context of beamforming, we propose to write each iteration k as:

$$\tilde{x}_{k+1} = x_k + 2\mu \mathbf{1}^T \mathbf{C}^{-1} (\mathbf{y} - \mathbf{1}x_k) \tag{9}$$

$$\hat{x}_{k+1} = \text{Prox}(\tilde{x}_{k+1}) \tag{10}$$

where $\text{Prox}(\cdot)$ denotes the proximal operator operating on the beamformed image domain, and μ is a step size for the gradient of (4).

3. NEURAL MAP BEAMFORMING

Like in MV and Wiener beamforming, direct implementation of (9) requires estimation of the signal statistics for each pixel, i.e. the inverse covariance matrix. In neural MAP beamforming we propose to circumvent this slow and often unstable procedure by learning parts of the DC step with neural networks. Furthermore, to model the proximal operators of more complex/rich prior distributions we propose to use neural networks in the proximal step (10). As a general case we can

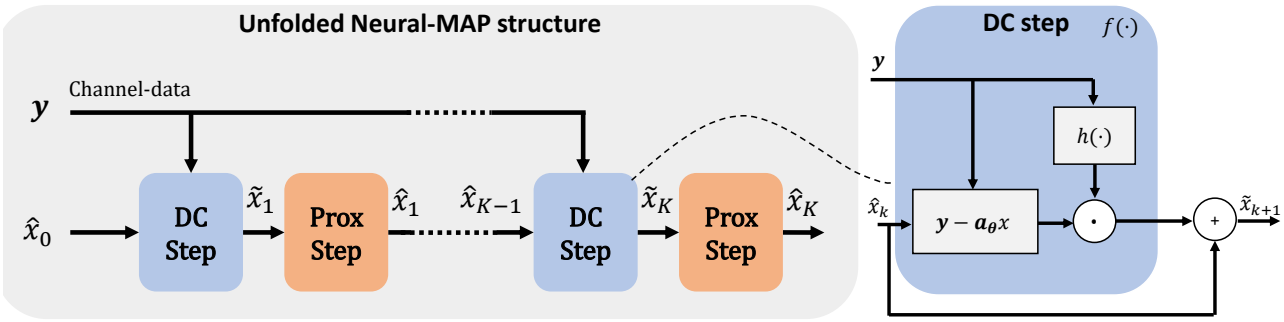

Fig. 1. (Left) Schematic overview of neural MAP beamforming, comprising a fixed number of DC and prior steps. On the right a detailed diagram of the DC step is given, which inherits its structure from adaptive beamforming.

write this as

$$\tilde{x}_{k+1} = x_k + f_{\theta,k}(x_k, \mathbf{y}) \tag{11}$$

$$\hat{x}_{k+1} = g_{\theta,k}(x_k, \mathbf{y}) \tag{12}$$

where θ constitutes the trainable weights. This iterative process can then be unfolded [13] into a forward neural network comprising K folds (iterations) of alternating DC and prior steps, to yield a fixed complexity model. A schematic overview of this is given in Figure 1. In the following sections we will detail on the specific architectures of f_θ and g_θ.

3.1. Data-consistency step

For the DC step we use the approach presented in [11] [12], which provides a model-based deep learning approach to adaptive beamforming (ABLE), such that

$$\hat{x}_{\text{ABLE}} = h_\theta(\mathbf{y})\mathbf{y} \tag{13}$$

where $h_\theta(\cdot)$ is a fully connected neural network that predicts a set of adaptive apodization weights. This can be seen as an approximation to the MV estimator in (4), where $\mathbf{w}_{\text{MV}} \approx h_\theta(\mathbf{y})$. Similarly, in neural MAP we estimate (11) through

$$\tilde{x}_{k+1} = \hat{x}_k + h_{\theta,k}(\mathbf{y})(\mathbf{y} - \mathbf{1}x_k), \tag{14}$$

multiplying the residual $\mathbf{y} - \mathbf{1}x_k$ with a set of content adaptive apodization weights. Note that for a single DC step, and initializing $x_0 = 0$, this is equivalent to (13).

To extend the field of view, we implement $h_\theta(\cdot)$ as a 2D convolutional network comprising 4 convolutional layers, with a kernel size of 3×3. We set the number of filters in the first and the last layer to 128, corresponding the the number of array elements in our data, and introduce a latent space of only 32 filters for the inner layers. Between each convolutional layer, an Antirectifier activation was used, as in [12]. In this work we consider 4 folds (K = 4) which do not share network weights, such that each instance $h_{\theta,k}$ is optimized for update step k.

3.2. Proximal step

Assuming sparsity, a commonly used proximal step is the soft-thresholding operator:

$$\hat{x}_{k+1} = \mathcal{T}_\lambda(\tilde{x}_{k+1}) \tag{15}$$

with threshold parameter λ. In the context of specific applications, such as ultrasound-localization microscopy (the detection of sparsely distributed microbubbles), such a prior was shown appropriate [14]. However, in general, the probability density function describing ultrasound images is complex and strongly entangled across pixels. Deriving an explicit proximal step is therefore non-trivial.

We therefore instead consider learning a general non-linear transformation function $\mathcal{F}$ which transforms the beamformed images to a sparse domain. Following the approach by Zhang et al. [15], we define a function $\mathcal{F}(\cdot)$, and its inverse $\widetilde{\mathcal{F}}(\cdot)$ which map x to and from a sparse domain, such that

$$\hat{x}_k = \widetilde{\mathcal{F}}(\mathcal{T}_\lambda(\mathcal{F}(\tilde{x}_k))). \tag{16}$$

We can learn such a mapping from data by approximating this with a neural network $\mathcal{F}_{\theta,k}(\cdot)$ and $\widetilde{\mathcal{F}}_{\theta,k}(\cdot)$. To promote symmetry between these transformations, ($\mathcal{F}_{\theta,k} \circ \mathcal{F}_{\theta,k} \approx \mathcal{I}$), we apply a symmetry loss during training defined as

$$\mathcal{L} = ||x - \widetilde{\mathcal{F}}_{\theta,k}(\mathcal{F}_{\theta,k}(x))||_2^2 \tag{17}$$

In this work, $\mathcal{F}_{\theta,k}$ and $\widetilde{\mathcal{F}}_{\theta,k}$ comprise two 2D convolutional layers with 11×11 kernels, a filter size of 2, and Antirectifier activation.

3.3. Training details

For training, a dataset of *in-vivo* ultrasound images was acquired using the Vantage system (Verasonics Inc., WA, USA) with a linear probe comprising 128 array elements (L11-4v), and a center frequency of 6.25 MHz. A total of 810 images

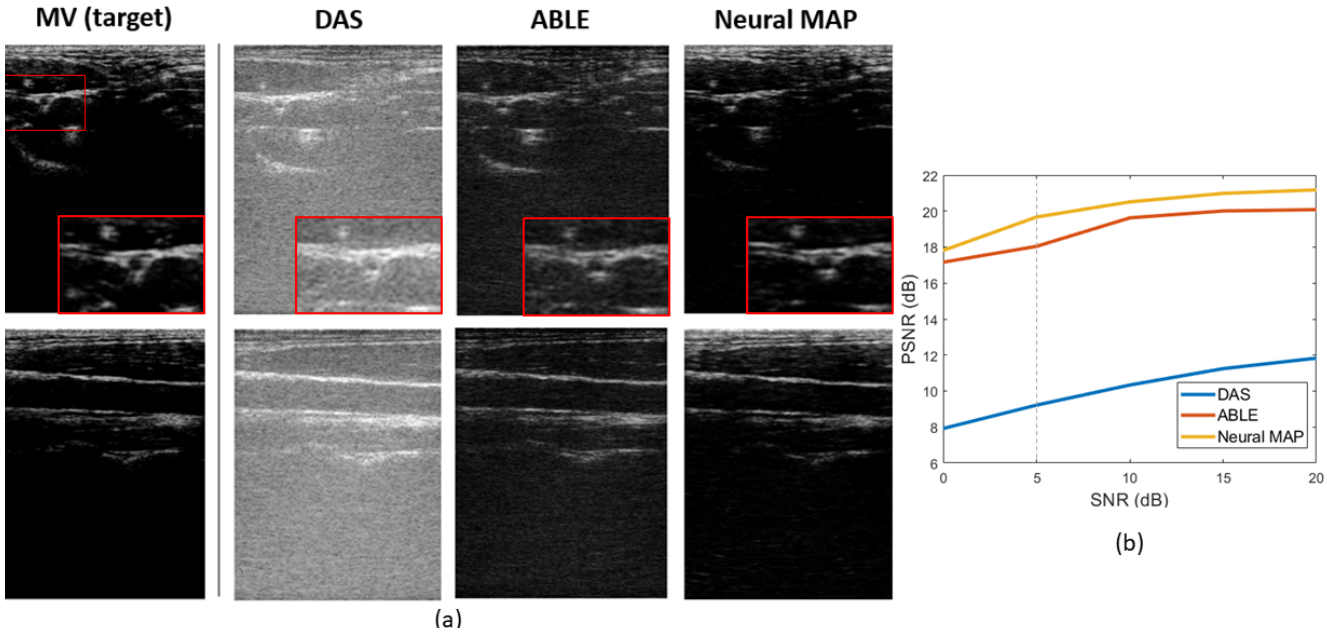

Fig. 2. (a) *In-vivo* carotid artery cross-section (top), and longitudinal cross-section (bottom) for DAS, ABLE and neuralMAP, with noise added to the input (5dB SNR). MV images, without noise, are included in the left column as a ground truth. (b) PSNR plotted against input SNR level. The red line indicates the SNR of the images in (a).

were used for training, acquired at 11 equispaced plane wave (PW) angles between -18 and 18 degrees. As an input to our model, we provide only the RF data of a single (center) PW angle. However, the multi-angle acquisitions were used to generate high-quality training targets by means of MV beamforming following the method described in [12]. Neural MAP was implemented in Tensorflow (Google, CA, USA), and trained on a GPU workstation with an Nvidia RTX 3080 Ti. The training loss was calculated using the signed-mean-square-logarithmic-error, first proposed in [11], which works well for high-dynamic range, modulated data, such as ultrasound RF signals.

4. RESULTS AND DISCUSSION

After training neural MAP on *in-vivo* data for 100 epochs, we tested the model on a separately acquired *in-vivo* dataset. We compared neural MAP against DAS (reference) and ABLE (state-of-the-art). To evaluate robustness under the influence of noise, we added white Gaussian noise to the input, such that we achieved an SNR ranging from 0dB to 20dB compared to the original signal. The image reconstructions for each method are visualized in Fig. 2a, for an SNR of 5dB. As a reference, the MV beamformed images were included, corresponding to our training targets. In Fig. 2b we plot the PSNR of each method for different levels of input noise calculated over the test set. Note that ABLE and neural MAP were only trained on the original uncorrupted inputs, and not optimized for each noise level.

A significant improvement in contrast and resolution can be seen in both ABLE and neural MAP as compared to DAS. However, neural MAP allows for better noise suppression compared to ABLE, which can be explained by the fact that the proximal step acts as a denoiser. So far we have not yet performed an exhaustive search to optimize network hyperparameters (e.g. kernel sizes), as such future work may further improve the reported results.

In the current implementation, we achieved an inference time of 30-35ms per frame, which indicates the feasibility of real-time imaging. We expect that optimizations will further reduce computation time, thereby improving the framerate.

5. CONCLUSION

This work presents a model-based beamforming approach that leverages deep learning techniques and prior knowledge of the signal model. Our experimental results demonstrate that the proposed method achieves enhanced contrast on *in-vivo* data in the presence of varying levels of interference, surpassing the current state-of-the-art in this field. Moreover, the achieved inference time suggests that the proposed method is a feasible solution for real-time ultrasound imaging applications.

6. REFERENCES

[1] Vincent Perrot, Maxime Polichetti, François Varray, and Damien Garcia, "So you think you can das? a viewpoint on delay-and-sum beamforming," *Ultrasonics*, vol. 111, pp. 106309, 2021.

[2] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug 1969.

[3] Johan Fredrik Synnevag, Andreas Austeng, and Sverre Holm, "Adaptive beamforming applied to medical ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 54, no. 8, pp. 1606–1613, 2007.

[4] Carl-Inge Colombo Nilsen and Sverre Holm, "Wiener beamforming and the coherence factor in ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, no. 6, pp. 1329–1346, 2010.

[5] T. Chernyakova, D. Cohen, M. Shoham, and Y. C. Eldar, "iMAP beamforming for high quality high frame rate imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, pp. 1–1, 2019.

[6] Teodora Szasz, Adrian Basarab, and Denis Kouamé, "Beamforming through regularized inverse problems in ultrasound medical imaging," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 63, no. 12, pp. 2031–2044, 2016.

[7] Ece Ozkan, Valery Vishnevsky, and Orcun Goksel, "Inverse problem of ultrasound beamforming with sparsity constraints and regularization," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 65, no. 3, pp. 356–365, 2017.

[8] Erasmia Lyka, Christian M. Coviello, Catherine Paverd, Michael D. Gray, and Constantin C. Coussios, "Passive acoustic mapping using data-adaptive beamforming based on higher order statistics," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2582–2592, 2018.

[9] Vishal Monga, Yuelong Li, and Yonina C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021.

[10] Nir Shlezinger, Yonina C. Eldar, and Stephen P. Boyd, "Model-based deep learning: On the intersection of deep learning and optimization," 2022.

[11] Ben Luijten, Regev Cohen, Frederik J De Bruijn, Harold AW Schmeitz, Massimo Mischi, Yonina C Eldar, and Ruud JG Van Sloun, "Deep learning for fast adaptive beamforming," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1333–1337.

[12] Ben Luijten, Regev Cohen, Frederik J De Bruijn, Harold AW Schmeitz, Massimo Mischi, Yonina C Eldar, and Ruud JG Van Sloun, "Adaptive ultrasound beamforming using deep learning," *IEEE Transactions on Medical Imaging*, 2020.

[13] Karol Gregor and Yann LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th international conference on international conference on machine learning*, 2010, pp. 399–406.

[14] Ruud JG Van Sloun, Regev Cohen, and Yonina C Eldar, "Deep learning in ultrasound imaging," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 11–29, 2019.

[15] Jian Zhang and Bernard Ghanem, "Ista-net: Interpretable optimization-inspired deep network for image compressive sensing," in *CVPR*, 2018, pp. 1828–1837.