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/ICASSP49357.2023.10096199

ROBUSTNESS OF DEEP EQUILIBRIUM ARCHITECTURES TO CHANGES IN THE MEASUREMENT MODEL

Junhao Hu , Shirin Shoushtari*, Zihao Zou, Jiaming Liu, Zhixin Sun, and Ulugbek S. Kamilov*

Computational Imaging Group (CIG), Washington University in St. Louis, MO, USA
* These authors contributed equally.

ABSTRACT

Deep model-based architectures (DMBAs) are widely used in imaging inverse problems to integrate physical measurement models and learned image priors. Plug-and-play priors (PnP) and deep equilibrium models (DEQ) are two DMBA frameworks that have received significant attention. The key difference between the two is that the image prior in DEQ is trained by using a specific measurement model, while that in PnP is trained as a general image denoiser. This difference is behind a common assumption that PnP is more robust to changes in the measurement models compared to DEQ. This paper investigates the robustness of DEQ priors to changes in the measurement models. Our results on two imaging inverse problems suggest that DEQ priors trained under mismatched measurement models outperform image denoisers.

Index Terms— Computational imaging, inverse problems, plug-and-play priors, deep equilibrium models.

1. INTRODUCTION

Deep learning (DL) has become a popular data-driven strategy for solving imaging inverse problems—such as image denoising, deblurring, super-resolution, and reconstruction—by training deep neural net architectures to map noisy measurements to the desired images [1]. Among various DL architectures for inverse problems, deep model-based architectures (DMBAs) have received significant attention due to their ability to integrate physical measurement models and image priors specified as convolutional neural nets (CNN). Well-known strategies for designing DMBAs include plug-and-play priors (PnP), regularization by denoising (RED), deep unfolding (DU), and deep equilibrium architectures (DEQ) (see review papers [2–4]). DMBAs can be systematically obtained from model-based iterative algorithms by parametrizing the regularization step as a CNN and training it to adapt to the empirical distribution of desired images.

Current DMBA strategies can be conceptually divided into two categories. The first category consists of models that rely on image priors trained independently of the measurement model. PnP and RED are two well-known frameworks that specify image priors using image denoisers trained to remove additive white Gaussian noise (AWGN) in the first category [5–7]. DU and DEQ are two well-known frameworks in the second category where the image prior is trained to be end-to-end optimal for a specific inverse problem [8–10]. Since all DMBA categories use the knowledge of the measurement model during inference, it is commonly accepted that DMBAs are more robust than generic CNNs to changes in the measurement model [3]. On the other hand, since the image prior in PnP/RED is independent of the specific measurement model, it is a common assumption that PnP/RED are more robust than DU/DEQ to changes in the measurement models. Model adaptation [11] has been proposed to investigate and improve image reconstruction with mismatched measurement model within PnP frameworks. Despite the rich literature on DMBAs, the robustness of DEQ to changes in the measurement models has never been systematically compared to PnP.

This paper addresses this gap by comparing DEQ to PnP. Both frameworks can be viewed as *implicit neural networks* with potentially an infinite number of layers [10]. While the image priors in PnP are AWGN denoisers, those in DEQ are *artifact removal (AR)* operators trained end-to-end using specific measurement models. We consider two distinct inverse problems: (a) compressive sensing magnetic resonance imaging (CS-MRI) [12] and (b) image super-resolution with known blur kernels [13]. We use the same deep architecture for both PnP and DEQ derived from the well-known model-based iterative algorithms. Our results suggest that contrary to common intuition, AR priors trained using mismatched measurement models within DEQ can perform better relative to the pure AWGN priors in PnP. We observe that mismatched AR priors outperform AWGN priors on average by 1.84 dB in CS-MRI and 0.23 dB in image super-resolution.

This paper is based upon work supported by the NSF CAREER award under grant CCF-2043134

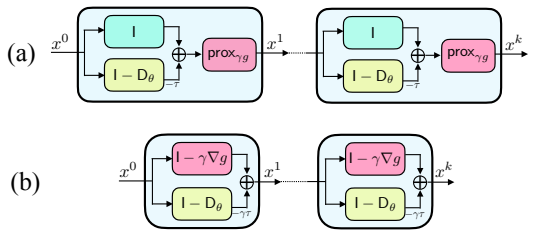

Fig. 1. Schematic illustration of two DMBAs for inverse problems: (a) PnP-PGM and (b) SD-RED. While both architectures have the same set of fixed points, their difference is in the use of the data-fidelity term, namely PnP-PGM uses $\mathrm{prox}_{\gamma g}$ and SD-RED uses ∇g.

2. INVERSE PROBLEMS

Recovering an unknown image $\boldsymbol{x}^* \in \mathbb{R}^n$ from its sub-sampled and noisy measurements

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{e}, \tag{1}$$

is an inverse problem, often formulated as a regularized optimization

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\, f(\boldsymbol{x}) \quad \text{with} \quad f(\boldsymbol{x}) = g(\boldsymbol{x}) + h(\boldsymbol{x}), \tag{2}$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is the measurement model that characterizes the response of a physical system, $\boldsymbol{e}$ is the AWGN, g is the data-fidelity term that measures the consistency of the solution with $\boldsymbol{y}$, and h is a regularizer that enforces prior knowledge on $\boldsymbol{x}$. For example, two traditional data-fidelity and regularization terms are the least-squares function $g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$ and the total variation (TV) function $h(\boldsymbol{x}) = \tau\|\boldsymbol{D}\boldsymbol{x}\|_1$, where $\tau > 0$ is the regularization parameter and $\boldsymbol{D}$ is an image gradient [14].

When the function g or h in (2) is nonsmooth, the optimization problem is often solved using a proximal algorithm. Given a proper, closed, and convex function g, both of these algorithms avoid differentiating it by relying on the proximal operator

$$\mathrm{prox}_{\gamma g}(\boldsymbol{z}) := \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\left\{\frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \gamma g(\boldsymbol{x})\right\}, \tag{3}$$

where the parameter $\gamma > 0$ is analogous to step-size.

3. DEEP MODEL-BASED ARCHITECTURES

We now present the details of DMBAs used in this paper. We first present two deep architectures and then discuss two training approaches for the image priors.

3.1. Plug-and-Play Architectures

We adopt *steepest descent* variant of RED (SD-RED) and *proximal gradient method* version of PnP (PnP-PGM) for CS-MRI and image super-resolution experiments, respectively. The SD-RED updates can be expressed as

$$\begin{aligned} \boldsymbol{x}^k &= \mathsf{T}_{\text{SD-RED}}(\boldsymbol{x}^{k-1}) \\ \text{with} \quad \mathsf{T}_{\text{SD-RED}}(\boldsymbol{x}) &= \boldsymbol{x} - \gamma(\nabla g(\boldsymbol{x}) + \tau(\boldsymbol{x} - \mathsf{D}(\boldsymbol{x}))), \end{aligned} \tag{4}$$

where $\gamma > 0$ is the step-size, D is the CNN prior, and $\tau > 0$ is the regularization parameter. Note that if the conditions on D from [6] are satisfied then $\tau(\boldsymbol{x} - \mathsf{D}(\boldsymbol{x}))$ is interpretable as a gradient of some convex regularizer h. Note that depending on the training procedure (see Section 3.2), the CNN prior can correspond to an AWGN denoiser or AR operator.

Instead of using the gradient ∇g, PnP-PGM performs a proximal update $\mathrm{prox}_{\gamma g}$ with respect to the data-fidelity term g. This update has a closed-form solution for image super-resolution, which makes it preferable for that problem. The update rule for PnP-PGM can be summarized as

$$\begin{aligned} \boldsymbol{x}^k &= \mathsf{T}_{\text{PnP-PGM}}(\boldsymbol{x}^{k-1}) \\ \text{with} \quad \mathsf{T}_{\text{PnP-PGM}}(\boldsymbol{x}) &= \mathrm{prox}_{\gamma g}(\boldsymbol{x} - \gamma\tau(\boldsymbol{x} - \mathsf{D}(\boldsymbol{x}))), \end{aligned} \tag{5}$$

where $\tau > 0$ is the regularization parameter, and $\gamma > 0$ is the step-size. For linear inverse problems, the proximal operator corresponding to $g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$ has the following closed-form solution

$$\mathrm{prox}_{\gamma g}(\boldsymbol{z}) = (\gamma\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A} + \boldsymbol{I})^{-1}(\boldsymbol{z} + \gamma\boldsymbol{A}^{\mathsf{T}}\boldsymbol{y}), \tag{6}$$

which can be efficiently evaluated in the Fourier domain for image super-resolution [15].

It is straightforward to verify that both methods have the same set of fixed points $\overline{\boldsymbol{x}} \in \mathrm{Fix}(\mathsf{T}) := \{\boldsymbol{x} \in \mathbb{R}^n : \mathsf{T}(\boldsymbol{x}) = \boldsymbol{x}\}$ that balance the measurement model and learned prior model.

3.2. Training Image Priors

The traditional PnP strategy considers the following image denoising problem

$$\boldsymbol{z} = \boldsymbol{x}_0 + \boldsymbol{w} \quad \boldsymbol{x}_0 \sim p_{\boldsymbol{x}_0}\,, \quad \boldsymbol{w} \sim \mathcal{N}(0, \sigma^2\boldsymbol{I})\,, \tag{7}$$

and trains D as a CNN that maps $\boldsymbol{z}$ to $\boldsymbol{x}_0$. Since the training does not use the measurement model $\boldsymbol{A}$, the prior is viewed as a generic image prior usable in multiple applications.

DEQ seeks to train D by minimizing the discrepancy between the fixed-point $\overline{\boldsymbol{x}} = \mathsf{T}_{\boldsymbol{\theta}}(\overline{\boldsymbol{x}})$ and the ground truth image $\boldsymbol{x}^*$

$$\ell(\boldsymbol{\theta}) = \frac{1}{2}\|\overline{\boldsymbol{x}}(\boldsymbol{\theta}) - \boldsymbol{x}^*\|_2^2. \tag{8}$$

The gradient of the loss with respect to $\boldsymbol{\theta}$ can be computed using implicit differentiation at the fixed-points

$$\nabla\ell(\boldsymbol{\theta}) = (\nabla_{\boldsymbol{\theta}}\mathsf{T}_{\boldsymbol{\theta}}(\overline{\boldsymbol{x}}))^{\mathsf{T}}\left(\mathsf{I} - \nabla_{\boldsymbol{x}}\mathsf{T}_{\boldsymbol{\theta}}(\overline{\boldsymbol{x}})\right)^{-\mathsf{T}}(\overline{\boldsymbol{x}} - \boldsymbol{x}^*), \tag{9}$$

where ℓ is given in (8) and I is the identity mapping. Since DEQ includes the information of the measurement model $\boldsymbol{A}$

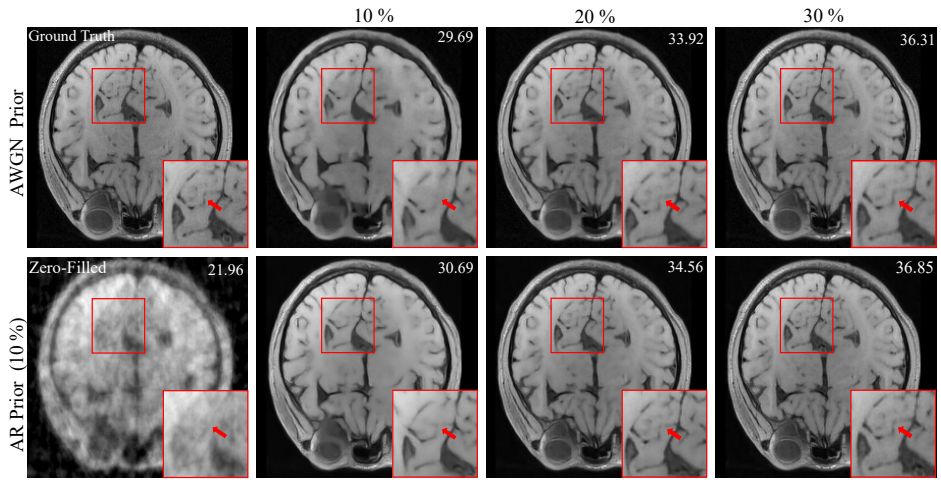

Fig. 2. The comparison of AWGN (PnP) and AR (DEQ) priors on CS-MRI image reconstruction at 10%, 20%, and 30% radial Fourier sampling. Since the AR prior was trained at 10% sampling rate, it is a mismatched prior for 20% and 30% sampling scenarios. Note how the mismatched AR prior outperforms AWGN prior in every scenario, showing its robustness to measurement model mismatch.

(embedded in the operator T), the corresponding AR operator can be viewed as a problem-specific image prior.

We consider image priors trained via DEQ using a "mismatched" measurement operator $\boldsymbol{A}'$ and applied at inference time using the true measurement operator $\boldsymbol{A}$. Accordingly, we train SD-RED by replacing ∇g in (4) with $\nabla g'(\boldsymbol{x}) = \boldsymbol{A}'^{\mathsf{T}}(\boldsymbol{A}'\boldsymbol{x} - \boldsymbol{y})$. Similarly, PnP-PGM is trained by replacing $\text{prox}_{\gamma g}$ in (6) by a mismatched update rule

$$\text{prox}_{\gamma g'}(\boldsymbol{z}) = (\gamma \boldsymbol{A}'^{\mathsf{T}} \boldsymbol{A}' + \boldsymbol{I})^{-1}(\boldsymbol{x} + \gamma \boldsymbol{A}'^{\mathsf{T}} \boldsymbol{y}). \qquad (10)$$

4. NUMERICAL RESULTS

Our numerical results evaluate the robustness of DEQ priors to changes in the measurement operator $\boldsymbol{A}$ in the context of two imaging problems: CS-MRI and image super-resolution.

4.1. Compressive Sensing MRI

In CS-MRI, the goal is to recover an image $\boldsymbol{x}^*$ from its sparsely-sampled Fourier measurements. We simulate a single-coil CS-MRI with radial Fourier Sampling. The measurement model $\boldsymbol{A}$ in CS-MRI is $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{F}$, where $\boldsymbol{F}$ is the Fourier transform and $\boldsymbol{P}$ is a diagonal sampling matrix.

Image priors were trained using the brain dataset [16], where we used 10 slices of 256×256 images as the test images. The AWGN denoisers for PnP correspond to a DnCNN architecture [17] trained using noise levels $\sigma \in \{1, 2, 3, 5, 7, 8, 10, 12, 15\}$. For each experiment, we select the denoiser that achieves the highest PSNR. We trained AR operators for DEQ using different CS ratios (m/n) with the same DnCNN architecture. The AR priors were initialized using the pre-trained AWGN denoiser with $\sigma = 5$. Nesterov [18] and Anderson accelerations [19] were used in the forward pass and the backward pass during DEQ training.

Table 1. Average PSNR (dB) for CS-MRI.

CS ratio	10%	20%	30%	Avg
AWGN Prior	31.37	35.52	37.88	34.92
Mismatched AR Prior	**33.05**	**37.48**	**39.77**	**36.76**

Fig. 2 shows results on a test image from [16] at three sampling ratios: 10%, 20%, and 30%. The AR prior in the reconstruction was trained using the measurement model corresponding to 10% sampling. Therefore, it is a mismatched prior for performing inference at 20% and 30% sampling. Despite the mismatch, the AR prior significantly outperforms the AWGN prior in all considered scenarios. Table 1 reports all the comparisons between the mismatched AR priors and AWGN priors in CS-MRI. Note that the mismatched AR operator outperforms the AWGN prior in all experiments.

Fig. 3 illustrates the impact of using an inaccurate measurement model *during inference*. The results are obtained by using the measurement model at 20% sampling for reconstructing from data corresponding to 10% sampling. Fig. 3 shows the results obtained using matched and mismatched measurement models at inference for both AWGN and AR priors. We observe a severe performance drop due to the usage of an inaccurate measurement model during inference, which highlights the importance of the measurement models in DMBAs.

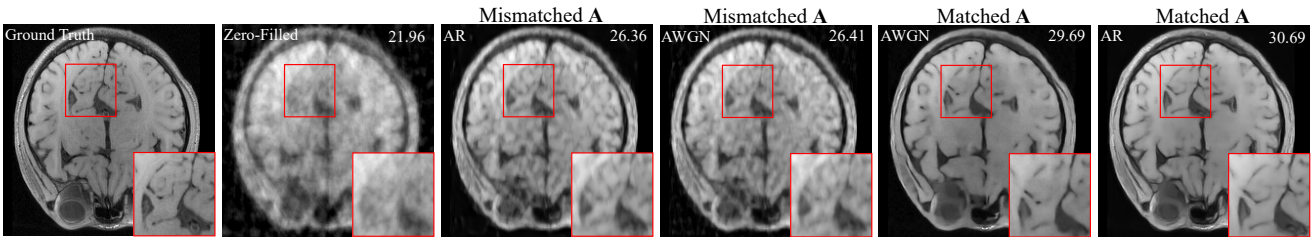

Fig. 3. Illustration of using mismatched measurement models *during inference*. The mismatched setting is obtained by using the measurement model at 20% sampling for reconstructing from data corresponding to 10% sampling. Note the dramatic performance drop due to the usage of mismatched measurement models during inference for both AWGN and AR priors.

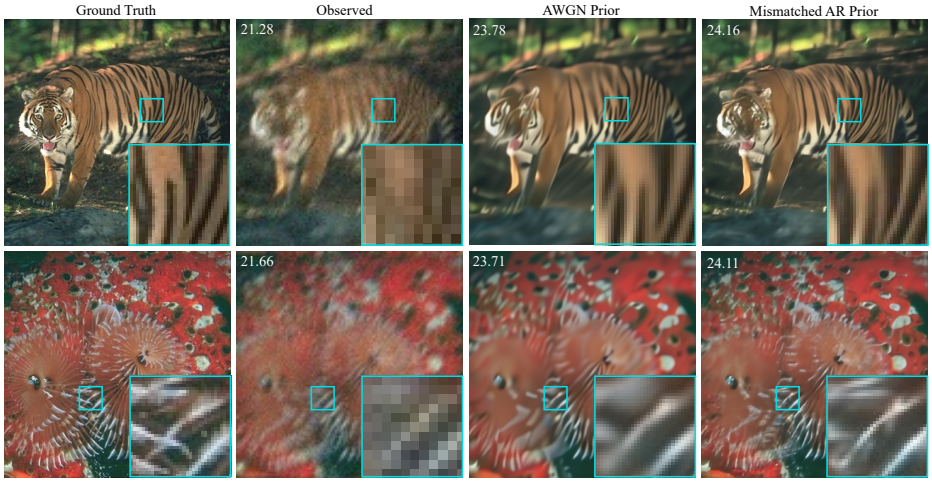

Fig. 4. Image super-resolution on two images using an AWGN prior and a mismatched AR prior. The measurement model corresponds to scaling of $3\times$ superresolution with a input noise level of $\nu = 0.03$. Note how the mismatched AR prior outperforms the AWGN prior.

Table 2. Average PSNR (dB) for image super-resolution.

Blur kernel	kernel 1	kernel 2	kernel 3	Avg
AWGN Prior	24.63	24.22	25.75	25.12
Mismatched AR Prior	**24.86**	**24.57**	**25.90**	**25.35**

4.2. Image Super-Resolution

The measurement model in image super-resolution corresponds to $\boldsymbol{A} = \boldsymbol{SH}$, where $\boldsymbol{H}$ is the convolution with an anti-aliasing kernel, $\boldsymbol{S}$ is the standard s-fold downsampling matrix of size $m \times n$, and $n = s^2 \times m$. The priors were trained using color image dataset in [20]. CBSD68 dataset proposed in [21] was used for inference. Three Gaussian blur kernels (kernels (b), (d), and (e) from [22]) were used to downsample images at scale $s = 3$ for inference, and jointly, at scale $s = 2$ and $s = 4$ during training of the DEQ prior. The CNN priors architectures correspond to U-net [23].

Fig. 4 illustrates results on two subsampled images at scale $s = 3$. Table 2 reports the results over all the test images for 3 different Gaussian blur kernels. Note how regardless of the blur kernel, mismatched AR priors outperform AWGN priors.

5. CONCLUSION

This work investigates the robustness of CNN priors trained using the DEQ framework to changes in the measurement operators. To that end, we compare image priors obtained via DEQ to PnP, where image priors are characterized using general AWGN denoisers. We show on two imaging inverse problems that DEQ priors outperform traditional PnP despite DEQ using different measurement operators at training and testing. Our results suggest the robustness of the image priors trained using DEQ to moderate shifts in the measurement operators, thus complementing the recent theoretical analysis of DMBAs under mismatched priors in [24]. In future works, we plan to broaden the perspective regarding the robustness of PnP frameworks by investigating (a) the performance of DEQ priors beyond moderate shift in measurement model to fundamentally different measurement models (i.e. DEQ prior trained on one task and tested on other tasks), and (b) deep unfolding against DEQ priors with mismatched measurement model.

References

[1] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 85–95, 2017.

[2] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, "Using deep neural networks for inverse problems in imaging: Beyond analytical methods," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 20–36, Jan. 2018.

[3] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett, "Deep learning techniques for inverse problems in imaging," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 39–56, May 2020.

[4] U. S. Kamilov, C. A. Bouman, G. T Buzzard, and B. Wohlberg, "Plug-and-play methods for integrating physical and learned models in computational imaging," *IEEE Signal Process. Mag.*, 2022.

[5] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," *IEEE Glob. Conf. Signal Inf. Process.*, pp. 945–948, 2013.

[6] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM J. Imaging Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.

[7] U. S. Kamilov, H. Mansour, and B. Wohlberg, "A plug-and-play priors approach for solving nonlinear imaging inverse problems," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1872–1876, December 2017.

[8] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 394–405, Feb. 2019.

[9] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, Mar. 2021.

[10] D. Gilton, G. Ongie, and R. Willett, "Deep equilibrium architectures for inverse problems in imaging," *IEEE Trans. Comput Imaging.*, vol. 7, pp. 1123–1133, 2021.

[11] Davis Gilton, Gregory Ongie, and Rebecca Willett, "Model adaptation for inverse problems in imaging," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 661–674, July 2021.

[12] M. Lustig, D. L Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing mri," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, 2008.

[13] M. S. C. Almeida and M. A. T. Figueiredo, "Deconvolving images with unknown boundaries using the alternating direction method of multipliers," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3074–3086, August 2013.

[14] A. Beck and M. Teboulle, "Fast gradient-based algorithm for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, November 2009.

[15] M. V. Afonso, J. M.Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, September 2010.

[16] J. Zhang and B. Ghanem, "Ista-net: Interpretable optimization-inspired deep network for image compressive sensing," *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1828–1837, 2018.

[17] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, 2017.

[18] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, vol. 87, Springer Science & Business Media, 2003.

[19] D. G. Anderson, "Iterative procedures for nonlinear integral equations," *J. ACM*, vol. 12, no. 4, pp. 547–560, 1965.

[20] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021.

[21] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," *IEEE Int. Conf. Comput. Vis.*, vol. 2, pp. 416–423, 2001.

[22] S. Hurault, A. Leclaire, and N. Papadakis, "Gradient step denoiser for convergent plug-and-play," *Int. Conf. Learn. Represent.*, 2021.

[23] J. Liu, X. Xu, W. Gan, S. Shoushtari, and U. S. Kamilov, "Online deep equilibrium learning for regularization by denoising," *J. Inf. Process. Syst.*, 2022.

[24] S. Shoushtari, J. Liu, Y. Hu, and U. S. Kamilov, "Deep model-based architectures for inverse problems under mismatched priors.," 2022, arXiv:2207.13200.