IEEE Open Journal of Signal Processing

Received 8 March 2023; revised 11 July 2023; accepted 11 July 2023. Date of publication 17 July 2023; date of current version 31 July 2023. The review of this article was arranged by Associate Editor Esteban Vera.

Digital Object Identifier 10.1109/OJSP.2023.3296036

Short Paper

Deep Equilibrium Learning of Explicit Regularization Functionals for Imaging Inverse Problems

ZIHAO ZOU[1], JIAMING LIU[2], BRENDT WOHLBERG[3], AND ULUGBEK S. KAMILOV[1,2]

[1] Department of Computer Science & Engineering, Washington University, St. Louis, MO 63130 USA
[2] Department of Electrical & Systems Engineering, Washington University, St. Louis, MO 63130 USA
[3] Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545 USA

CORRESPONDING AUTHOR: ULUGBEK S. KAMILOV (e-mail: kamilov@wustl.edu).

This work was supported in part by the NSF CAREER Award under Grant CCF-2043134 and in part by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under Grants 20200061DR and 20230771DI. *(Zihao Zou and Jiaming Liu contributed equally to this work.)*

ABSTRACT There has been significant recent interest in the use of deep learning for regularizing imaging inverse problems. Most work in the area has focused on regularization imposed *implicitly* by convolutional neural networks (CNNs) pre-trained for image reconstruction. In this work, we follow an alternative line of work based on learning *explicit* regularization functionals that promote preferred solutions. We develop the *Explicit Learned Deep Equilibrium Regularizer (ELDER)* method for learning explicit regularization functionals that minimize a mean-squared error (MSE) metric. ELDER is based on a regularization functional parameterized by a CNN and a deep equilibrium learning (DEQ) method for training the functional to be MSE-optimal at the fixed points of the reconstruction algorithm. The explicit regularizer enables ELDER to directly inherit fundamental convergence results from optimization theory. On the other hand, DEQ training enables ELDER to improve over existing explicit regularizers without prohibitive memory complexity during training. We use ELDER to train several approaches to parameterizing explicit regularizers and test their performance on three distinct imaging inverse problems. Our results show that ELDER can greatly improve the quality of explicit regularizers compared to existing methods, and show that learning explicit regularizers does not compromise performance relative to methods based on implicit regularization.

INDEX TERMS Computational imaging, imaging inverse problems, plug-and-play priors, deep learning.

I. INTRODUCTION

The recovery of an unknown image from a set of noisy measurements is one of the most widely-studied problems in computational imaging. The task is often formulated as an *inverse problem*, and solved by integrating the measurement model characterizing the response of the imaging instrument with a regularization functional imposing prior knowledge of the unknown image. Over the years, many regularizers have been proposed as image priors, including those based on transform-domain sparsity, low-rank penalty, and self-similarity. The focus in the area has recently shifted to methods based on deep learning (DL). Instead of explicitly defining a regularization functional, DL approaches for solving inverse problems learn a mapping from the measurements to the desired image by training a convolutional neural network (CNN) [1], [2].

Model-based DL (MBDL) has emerged as an alternative to the traditional DL, where the knowledge of the measurement model is combined with an image prior specified by a CNN (see reviews [3], [4]). For example, plug-and-play priors (PnP) is a widely-used MBDL regularisation method based on specifying the image prior using a pre-trained image denoiser [4], [5], [6], [7]. Deep unfolding (DU) is another MBDL approach based on interpreting a fixed-number of iterations of an algorithm as layers of a neural network and training it end-to-end in a supervised fashion [8]. DU architectures, however, are usually limited to a small number of unfolded iterations due to the high memory complexity of training. Deep equilibrium learning (DEQ) is an alternative to DU that enables training of very deep neural networks with a constant memory complexity in the number of iterations [9], [10], [11], [12].

Existing research on MBDL for inverse problems, including most of the work on PnP, DU, and DEQ, has largely focused on regularization *implicit* in pre-trained neural networks [13]. While this strategy has led to state-of-the-art performance, it requires stringent assumptions on the MBDL architecture to ensure algorithmic stability (see related reviews [4], [14]). For example, a common assumption used for proving stability of MBDL architectures is that the CNN is nonexpansive [10], [15], [16], [17], [18]. An

alternative line of work has explored MBDL with learned *explicit* regularization functionals parameterized by neural networks [19], [20], [21], [22], [23]. Explicit regularization functionals significantly simplify the convergence analysis due to the direct applicability of optimization theory. Throughout this article, we will use the term *explicit* to specify the direct availability of the regularization functional.

We develop *Explicit Learned Deep Equilibrium Regularizer (ELDER)* as a new method for learning explicitly defined regularization functionals for imaging inverse problems. This method seeks to learn a regularizer that achieves the smallest value of mean-squared error (MSE) for a given inverse problem. To this end, we parameterize the regularization functional by a CNN and train it end-to-end by using DEQ on the MSE loss. Similarly to existing approaches for learning explicit regularizers, ELDER directly inherits traditional concepts from optimization for parameter selection and convergence analysis. ELDER thus improves the stability and flexibility of DEQ by introducing a new forward pass based on minimizing an explicit objective function using line-search. On the other hand, ELDER outperforms existing PnP methods based on explicit regularization functionals due to its ability to optimize learned functionals near the fixed points of the reconstruction algorithm. We present numerical results on three imaging inverse problems: image super-resolution, image reconstruction from subsampled Fourier measurements, and image inpainting. We apply ELDER to optimize the weights of three parameterization approaches for regularization functionals—namely least squares residual (LSR), regularization by denoising (RED), and direct scalar-valued network (DSV)—and compare the effectiveness of all three as regularizers when trained to be MSE-optimal. We also show that ELDER does not compromise the imaging quality relative to methods based on implicit regularization. Our results show that ELDER achieves excellent imaging results, while also offering the potential for automatic step-size selection and algorithmic stability, even when using expansive update rules. In short, our work provides a method to learn state-of-the-art explicit regularizers for MBDL that preserve traditional tools from optimization theory.[1]

II. BACKGROUND

Inverse Problems: We consider the imaging inverse problem of recovering an unknown image $\boldsymbol{x} \in \mathbb{R}^n$ from noisy measurements $\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{e}$, where $\mathbf{A}$ is the *measurement operator* and $\boldsymbol{e} \in \mathbb{R}^m$ is additive white Gaussian noise (AWGN) vector. The problem is traditionally formulated as an optimization problem

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\, f(\boldsymbol{x}) \quad \text{with} \quad f(\boldsymbol{x}) = g(\boldsymbol{x}) + \tau h(\boldsymbol{x}), \tag{1}$$

where $\tau > 0$ is the regularization parameter, g is the *data-fidelity term* enforcing consistency of the solution with $\boldsymbol{y}$, and h is the *regularizer* enforcing prior knowledge of $\boldsymbol{x}$.

Model-based Optimization: Proximal algorithms are often used for solving problems of form (1) when h is nonsmooth (see the review [24]). One widely used family of proximal algorithms for imaging inverse problems are the proximal gradient method (PGM) [25]. PGM avoids differentiating h by using the *proximal operator*, which can be defined as

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

with $\tau > 0$, for any proper, closed, and convex function h [24]. Comparing (2) and (1), we see that the proximal operator can be interpreted as a MAP estimator for the problem

$$\boldsymbol{z} = \boldsymbol{x}_0 + \boldsymbol{w} \quad \text{where} \quad \boldsymbol{x}_0 \sim p_{\boldsymbol{x}_0}\,, \quad \boldsymbol{w} \sim \mathcal{N}(0, \tau \boldsymbol{I})\,, \tag{3}$$

by setting $h(\boldsymbol{x}) = -\log(p_{\boldsymbol{x}_0}(\boldsymbol{x}))$. Another less known but equally valid statistical interpretation of the proximal operator is as a minimum mean-squared error (MMSE) estimator [26].

PnP: PnP refers to a family of algorithms that integrate measurement operators and CNN denoisers for solving inverse problems (see the recent review [4]). Since the prior in PnP is implicit in the denoiser, it is common to interpret PnP as fixed-point iterations of some high-dimensional operators [27]. For example, given a denoiser $\mathsf{D}_{\boldsymbol{\theta}} : \mathbb{R}^n \to \mathbb{R}^n$ parameterized by a CNN with weights $\boldsymbol{\theta}$, the iterations of proximal gradient method (PGM) [25] variant of PnP can be written

$$\boldsymbol{x}^k = \mathsf{T}_{\boldsymbol{\theta}}(\boldsymbol{x}^{k-1}) \;\text{ with }\; \mathsf{T}_{\boldsymbol{\theta}} := \mathsf{D}_{\boldsymbol{\theta}}(\mathsf{I} - \gamma \nabla g)\,, \tag{4}$$

where g is the data-fidelity term in (1), I is the identity mapping, and $\gamma > 0$ is the step size.

DU and DEQ: DU is a DL paradigm that has gained popularity due to its ability to systematically connect iterative algorithms and deep neural network architectures (see reviews in [3], [28]). The DU training is usually performed by solving the optimization problem $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_i \|\mathsf{T}_{\boldsymbol{\theta}}(\boldsymbol{x}_i^k) - \boldsymbol{x}_i\|_2^2$ by a MSE loss. DEQ [9] is a related approach that enables training of infinite-depth, weight-tied networks by analytically backpropagating through the fixed points using implicit differentiation. The DEQ output is specified implicitly as a fixed point of an operator $\mathsf{T}_{\boldsymbol{\theta}}$ parameterized by weights $\boldsymbol{\theta}$

$$\bar{\boldsymbol{x}} = \mathsf{T}_{\boldsymbol{\theta}}(\bar{\boldsymbol{x}})\,. \tag{5}$$

The DEQ forward pass usually estimates $\bar{\boldsymbol{x}}$ in (5) by running a fixed-point iteration. The comparison of (4) and (5) highlights the connection between PnP and DEQ, which was recently explored [10] by using DEQ for end-to-end learning of the weights of the CNN prior $\mathsf{D}_{\boldsymbol{\theta}}$. There has been considerable interest in DEQ for imaging, including in MRI [10], [29], computed tomography (CT) [18] and video snapshot imaging [30]. The relationship between DEQ and bilevel learning in the context of inverse problems was recently explored in [12].

The convergence of forward iterations is essential for the stability and accuracy of DEQ. Similar to the theoretical analysis of PnP [15], [31], a sufficient condition to guarantee the convergence of the DEQ forward pass is to ensure that g is convex and the residual $\mathsf{R}_{\boldsymbol{\theta}} := \mathsf{I} - \mathsf{D}_{\boldsymbol{\theta}}$ of $\mathsf{D}_{\boldsymbol{\theta}}$ is *Lipschitz continuous* with constant $0 < \alpha \leq 1$ [10]

$$\|\mathsf{R}_{\boldsymbol{\theta}}(\boldsymbol{x}) - \mathsf{R}_{\boldsymbol{\theta}}(\boldsymbol{z})\|_2 \leq \alpha \|\boldsymbol{x} - \boldsymbol{z}\|_2, \quad \forall \boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^n. \tag{6}$$

Since most CNN architectures do not inherently satisfy this property, several methods have been proposed to train Lipschitz constrained CNNs [15], [32]. However, it has been observed that constraining CNNs to be Lipschitz continuous can negatively impact their performance [21], [23].

Learning Explicit Regularization Functionals: RED [19] is an early approach for specifying an explicit regularization functional by parameterizing it using an image denoiser $\mathsf{D}_{\boldsymbol{\theta}}$

$$h(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathsf{T}}(\boldsymbol{x} - \mathsf{D}_{\boldsymbol{\theta}}(\boldsymbol{x})). \tag{7}$$

When the denoiser is locally homogeneous and has a symmetric Jacobian [19], [27], the gradient of the RED regularizer h has a

[1] Our code is publicly available at https://github.com/wustl-cig/ELDER

simple expression

$$\nabla h(\boldsymbol{x}) = \boldsymbol{x} - \mathsf{D}_{\boldsymbol{\theta}}(\boldsymbol{x}), \tag{8}$$

which enables efficient minimization of the RED functional within (1). However, when these rather stringent conditions are not satisfied, RED does not correspond to an explicit regularizer, corresponding instead to an MBDL method with an implicit regularizer [27].

Other notable approaches for learning explicit regularization functionals include fields of experts (FoE) [33] and its bi-level counterpart [34], adversarial regularization [35] and its convex counterpart [36], network Tikhonov (NETT) [37], and total deep variation (TDV) [38]. These approaches are similar to ELDER in the sense that they seek to explicitly parametrize h using a neural network, but differ in the way the functional is learned. Another line of work has explored gradient-step denoisers for PnP based on the direct parameterization of regularization functionals, thus leading to explicit loss functions and convergence guarantees without Lipschitz constraints on the neural networks [20], [21], [39].

III. PROPOSED METHOD

We now present our proposed method. Unlike the traditional DEQ approach for inverse problems [10], the forward pass in our method is designed to minimize an explicit objective, where the regularization functional is parameterized by a CNN. Our method inherits the benefits of having an explicit regularizer, such as convergence without Lipschitz constraints on the CNN and the possibility of using line-search strategies for step-size selection. At the same time, our method leads to state-of-the-art explicit regularizers that are trained to be MSE optimal at the fixed-point of the inference algorithm.

A. ELDER FORWARD PASS

We consider an inverse problem with a regularization function $h_{\boldsymbol{\theta}} : \mathbb{R}^n \to \mathbb{R}$ parameterized by weights $\boldsymbol{\theta}$

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\, f_{\boldsymbol{\theta}}(\boldsymbol{x}) \quad \text{with} \quad f_{\boldsymbol{\theta}}(\boldsymbol{x}) = g(\boldsymbol{x}) + \tau h_{\boldsymbol{\theta}}(\boldsymbol{x}). \tag{9}$$

The forward pass of ELDER seeks to minimize $f_{\boldsymbol{\theta}}$ via the iterations

$$\boldsymbol{x}^k = \mathsf{T}_{\boldsymbol{\theta}}(\boldsymbol{x}^{k-1}) \;\text{ with }\; \mathsf{T}_{\boldsymbol{\theta}} := \operatorname{prox}_{\gamma g}(\mathsf{I} - \gamma\tau\nabla h_{\boldsymbol{\theta}}) \,, \tag{10}$$

where $\gamma > 0$ is the step-size. For a linear inverse problems with a ℓ_2-norm loss $g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \mathbf{A}\boldsymbol{x}\|_2^2$, the proximal operator has a closed-form solution

$$\operatorname{prox}_{\gamma g}(\boldsymbol{z}) = (\mathbf{I} + \gamma\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}(\boldsymbol{z} + \gamma\mathbf{A}^{\mathsf{H}}\boldsymbol{y}), \tag{11}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{A}^{\mathsf{H}}$ denotes the conjugate transpose of $\mathbf{A}$. In many applications, the proximal map (11) can be computed or approximated efficiently using general methods—such as conjugate gradient—or with specialized methods—such as when the forward model is a spatial blurring operator that can be computed using the fast Fourier transform (FFT) [4], [40], [41].

The step-size γ must be carefully selected to ensure the convergence of the algorithm in eq. (10). Since we have access to the explicit objective function, we can ensure convergence by adopting a line-search strategy. To that end, we use a backtracking line-search (BLS) to ensure the sufficient decrease condition (see Section IX-B in [42])

$$f(\boldsymbol{x}^{k-1}) - f(\boldsymbol{x}^k) \geq \frac{\rho}{\gamma}\|\boldsymbol{x}^k - \boldsymbol{x}^{k-1}\|_2^2, \quad 0 < \rho < \frac{1}{2}. \tag{12}$$

Our implementation starts with a large step-size $\gamma > 0$, which is subsequently reduced by a factor $0 < \beta < 1$ to ensure that (12) is satisfied.

B. PARAMETERIZATION OF REGULARIZATION FUNCTIONALS

Several approaches for explicitly parametrizing regularization functionals using CNNs have been previously proposed. We consider three well-known approaches and compare them in our numerical evaluations. All three regularizers are based on an operator $\mathsf{G}_{\boldsymbol{\theta}} : \mathbb{R}^n \to \mathbb{R}^n$ corresponding to an image-to-image CNN. Note that $\mathsf{G}_{\boldsymbol{\theta}}$ can be implemented using any differentiable CNN architecture.

Least-Squares Residual (LSR): Inspired by one of the formulations suggested in [19] (see Section V-B) and in the recent work on the gradient-step denoiser for PnP [21], we first consider a regularizer that explicitly quantifies the distance between the input and the output of $\mathsf{G}_{\boldsymbol{\theta}}$. More specifically, we consider

$$h_{\boldsymbol{\theta}}^{\mathsf{LSR}}(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x} - \mathsf{G}_{\boldsymbol{\theta}}(\boldsymbol{x})\|_2^2 \quad \text{and}$$
$$\nabla h_{\boldsymbol{\theta}}^{\mathsf{LSR}}(\boldsymbol{x}) = (\mathbf{I} - \mathbf{J}_{\mathsf{G}_{\boldsymbol{\theta}}}(\boldsymbol{x}))^{\mathsf{T}}(\boldsymbol{x} - \mathsf{G}_{\boldsymbol{\theta}}(\boldsymbol{x})), \tag{13}$$

where $\mathbf{J}_{\mathsf{G}_{\boldsymbol{\theta}}}$ is the Jacobian of $\mathsf{G}_{\boldsymbol{\theta}}$ with respect to $\boldsymbol{x}$. LSR can be interpreted as promoting solutions to the inverse problem that are near the fixed points of the CNN $\mathsf{G}_{\boldsymbol{\theta}}$.

(Real) RED Regularizer: While the RED gradient in (8) has been extensively used for solving inverse problems, the RED functional (7) has not been previously used as a regularizer. We explore the RED functional itself as a regularizer by considering the true gradient of $h_{\boldsymbol{\theta}}^{\mathsf{RED}}$ given by

$$\nabla h_{\boldsymbol{\theta}}^{\mathsf{RED}}(\boldsymbol{x}) = \boldsymbol{x} - \frac{1}{2}\mathsf{G}_{\boldsymbol{\theta}}(\boldsymbol{x}) - \frac{1}{2}[\mathbf{J}_{\mathsf{G}_{\boldsymbol{\theta}}}(\boldsymbol{x})]^{\mathsf{T}}\boldsymbol{x}, \tag{14}$$

where we do not require that the network $\mathsf{G}_{\boldsymbol{\theta}}$ is locally homogeneous and has a symmetric Jacobian.

Direct Scalar-Valued (DSV) Regularizer: Similar to [20] and [43], we can also directly sum the output of the CNN $\mathsf{G}_{\boldsymbol{\theta}}$ to obtain a scalar-valued neural network

$$h_{\boldsymbol{\theta}}^{\mathsf{DSV}}(\boldsymbol{x}) = \mathbf{1}^{\mathsf{T}}\mathsf{G}_{\boldsymbol{\theta}}(\boldsymbol{x}) \quad \text{and} \quad \nabla h_{\boldsymbol{\theta}}^{\mathsf{DSV}}(\boldsymbol{x}) = [\mathbf{J}_{\mathsf{G}_{\boldsymbol{\theta}}}(\boldsymbol{x})]^{\mathsf{T}}\mathbf{1}. \tag{15}$$

It is worth noting that for all the considered regularizers, we use conventional auto-differentiation tools to efficiently compute the Jacobian-vector products without actually storing the entire matrix.

C. JACOBIAN-FREE DEEP EQUILIBRIUM LEARNING

We train the regularizer $h_{\boldsymbol{\theta}}$ by minimizing the discrepancy between a fixed-point $\bar{\boldsymbol{x}} = \mathsf{T}_{\boldsymbol{\theta}}(\bar{\boldsymbol{x}})$ obtained via (10) and the ground-truth image $\boldsymbol{x}$ using MSE loss $\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{2}\|\bar{\boldsymbol{x}}(\boldsymbol{\theta}) - \boldsymbol{x}\|_2^2$. The DEQ backward pass produces gradients by implicitly differentiating through the fixed points

$$\nabla\mathcal{L}(\boldsymbol{\theta}) = \left[\mathbf{J}_{\mathsf{T}_{\bar{\boldsymbol{x}}}}(\boldsymbol{\theta})\right]^{\mathsf{T}}\left(\mathbf{I} - \mathbf{J}_{\mathsf{T}_{\boldsymbol{\theta}}}(\bar{\boldsymbol{x}})\right)^{-\mathsf{T}}(\bar{\boldsymbol{x}} - \boldsymbol{x}). \tag{16}$$

This converts the memory-intensive task of backpropagating through many iterations of $\mathsf{T}_{\boldsymbol{\theta}}$ to the problem of calculating an inverse Jacobian-vector product. Since inverting the Jacobian matrix in (16) can become computationally expensive, we introduce an approximation that replaces the inverse-Jacobian term in (16) with an identity as in [11]

$$\nabla\mathcal{L}(\boldsymbol{\theta}) \approx \nabla\mathcal{L}_{\mathsf{JFB}}(\boldsymbol{\theta}) = \left[\mathbf{J}_{\mathsf{T}_{\bar{\boldsymbol{x}}}}(\boldsymbol{\theta})\right]^{\mathsf{T}}(\bar{\boldsymbol{x}} - \boldsymbol{x}). \tag{17}$$

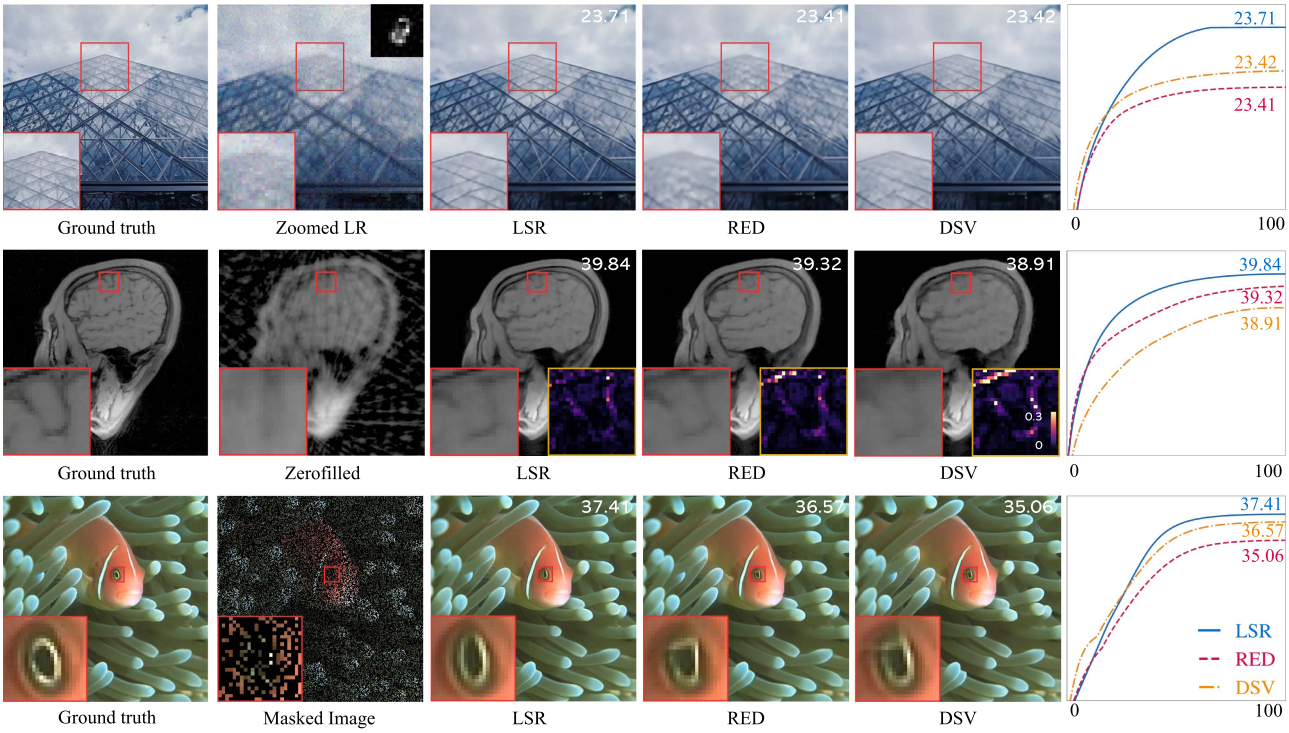

FIGURE 1. Visual illustration of the results obtained by ELDER using the three parameterization approaches for h_θ. We show results for three inverse problems: image super-resolution (top), reconstruction from Fourier samples (middle), and image inpainting (bottom). The rightmost panel plots the evolution of PSNR (dB) accross forward iterations. Note how LSR achieves an overall better performance compared to DSV and RED.

This *Jacobian-free approximation* significantly reduces the complexity of the backward pass without compromising the quality of DEQ.

D. CONVERGENCE THEORY

Since ELDER minimizes an explicit loss $f_\theta = g + \tau h_\theta$, it directly inherits traditional convergence results from optimization theory. For completeness, we state these convergence results using our notation.

Assumption 1: The function g is proper, convex, and lower semi-continuous. The function $h_\theta : \mathbb{R}^n \to \mathbb{R}$ is proper, lower semi-continuous, finite valued, and differentiable with L-Lipschitz gradient. The function f is bounded from below.

These assumptions on g and h_θ are standard, and are satisfied by a large number of functions used in the context of inverse problems (see, for example, the discussion in [21]).

Proposition 1: Run the PGM in (10) under Assumption 1 using $\gamma \in (0, 1/(\tau L)]$. Then, the sequence $\{f_\theta(\boldsymbol{x}^t)\}_{t\geq 1}$ monotonically decreases and $\nabla f_\theta(\boldsymbol{x}^t) \to \mathbf{0}$ as $t \to \infty$.

The proof is widely available in the existing analyses of PGM [21]. The importance of this result is that, unlike the traditional DEQ based on implicit regularization [10], the stability of the ELDER forward pass does not require any assumptions on the non-expansiveness of the CNN prior.

E. ADDITIONAL TECHNICAL DETAILS

Our method is compatible with any differentiable CNN architecture for implementing G_θ. We use the simplified DRUNet architecture [41] for ELDER and the traditional DEQ [10]. We have replaced the rectified linear unit (ReLU) activations with exponential linear unit (ELU) ones to ensure the smoothness of h_θ. We also limit the number of residual blocks at each scale to 2. Similar to [10], the CNN prior of ELDER is initialized using a pre-trained denoisers. Additionally, we follow [10], [44] in setting the convergence criterion in training to

$$\|\boldsymbol{x}^k - \boldsymbol{x}^{k-1}\|_2 / \|\boldsymbol{x}^{k-1}\|_2 < \epsilon\,, \tag{18}$$

where we choose $\epsilon = 10^{-2}$ for all experiments.

IV. NUMERICAL EVALUATION AND RESULTS

A. COMPARING PARAMETRIZATION STRATEGIES

We first compare the performance of the three parameterization strategies in Section III-B on image denoising. We pre-train all the regularizers by adopting the gradient-step denoising strategy from [21]. Pre-training explicit regularizers as denoisers is computationally useful for ELDER, since the denoisers can be used to initialize the DEQ learning. All three explicit denoisers are compared against DRUNet [41], which was shown to result in state-of-the-art PnP image restoration.

We employ the color image training dataset in [41]. During training we consider AWGN with standard deviation σ at a uniform random draw from the range $[0, 55/255]$. Table 1 presents the results of all denoisers at several noise levels ($\sigma \in \{5, 15, 25, 50\}$) on both color and gray images. The RED denoisier in the table refers to the gradient-step in (14). While all three explicit denoisers perform well at all noise levels, LSR most closely matches the performance of DRUNet. It is worth noting that LSR (with 2 residual blocks at each scale) uses fewer parameters than DRUNet (with 4 residual blocks). Note that the PSNR gap between DRUNet and LSR is within 0.15 dB, which implies that having explicit regularizers does not significantly impair performance (this was also observed in [21]).

TABLE 1. Average PSNR (dB) Results of Different Methods for Noise Levels 5, 15, 25 and 50 on BSD68, CBSD68 and Kodak24 Datasets. The Best and Second Best Results are Highlighted, Respectively

Datasets	Denoiser	Noise Level				Average
		5	15	25	50	
BSD68	DRUNet	**38.04**	**31.89**	**29.43**	**26.56**	31.48
	LSR	37.98	31.85	29.42	26.52	31.44
	RED	37.83	31.77	29.36	26.48	31.36
	DSV	37.63	31.52	29.10	26.18	31.11
CBSD68	DRUNet	**40.64**	**34.31**	**31.69**	**28.49**	33.78
	LSR	40.55	34.24	31.62	28.42	33.71
	RED	40.23	34.06	31.46	28.27	33.64
	DSV	40.24	34.05	31.45	28.25	33.50
Kodak24	DRUNet	**40.95**	**35.32**	**32.90**	**29.87**	34.76
	LSR	40.78	35.19	32.77	29.71	34.61
	RED	40.33	34.95	32.54	29.47	34.32
	DSV	40.43	34.92	32.53	29.46	34.34

TABLE 2. Average PSNR (dB) Values for Different Explicit CNN Regularizer Used in ELDER on SISR, CS-MRI and Inpainting

Regularizers	SISR		CS-MRI		Inpainting	
	x3	x4	10%	20%	50%	70%
LSR	25.53	25.22	35.14	38.68	32.30	28.82
RED	25.42	25.17	34.76	38.21	32.22	28.55
DSV	25.42	25,19	34.07	37.87	32.14	28.77

Fig. 1 and Table 2 present the results of comparing the three parameterization strategies within ELDER on the three inverse problems: single image super-resolution, reconstruction from Fourier measurements, and image inpainting (each problem is discussed in the dedicated section below). We observe that LSR leads to the best results, motivating its use as a primary parameterization strategy for ELDER.

B. SINGLE IMAGE SUPER-RESOLUTION

We consider measurements of form $\boldsymbol{y} = \mathbf{S}\mathbf{H}\boldsymbol{x} + \boldsymbol{e}$, where $\mathbf{H} \in \mathbb{R}^{n\times n}$ is the blurring matrix and $\mathbf{S} \in \mathbb{R}^{m\times n}$ performs standard d-fold down-sampling with $d^2 = n/m$. When the blur satisfies the circular boundary conditions, the blurring matrix and its conjugate transpose can be decomposed as $\mathbf{H} = \mathbf{F}^{\mathsf{H}}\mathbf{M}\mathbf{F}$ and $\mathbf{H}^{\mathsf{H}} = \mathbf{F}^{\mathsf{H}}\mathbf{M}^{\mathsf{H}}\mathbf{F}$, where $\mathbf{F}$ is the discrete Fourier transform (and $\mathbf{F}^{\mathsf{H}}$ its inverse, satisfying $\mathbf{F}\mathbf{F}^{\mathsf{H}} = \mathbf{F}^{\mathsf{H}}\mathbf{F} = \mathbf{I}$), and $\mathbf{M} \in \mathbb{C}^{n\times n}$ is a diagonal matrix whose diagonal elements are the Fourier coefficients of the blur. The proximal operator (11) of $g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \mathbf{S}\mathbf{H}\boldsymbol{x}\|_2^2$ has the closed-form solution

$$\mathrm{prox}_{\gamma g}(\boldsymbol{z}) = \boldsymbol{r} - \mathbf{F}^{\mathsf{H}}\left(\frac{\underline{\mathbf{M}}^{\mathsf{H}}\underline{\mathbf{M}}\mathbf{F}\boldsymbol{r}}{d^2\mathbf{I} + \gamma\underline{\mathbf{M}}\underline{\mathbf{M}}^{\mathsf{H}}}\right), \tag{19}$$

where $\boldsymbol{r} = \gamma\mathbf{H}^{\mathsf{H}}\mathbf{S}^{\mathsf{H}}\boldsymbol{y} + \boldsymbol{z}$. The matrix $\underline{\mathbf{M}} = [\mathbf{M}_1, \cdots, \mathbf{M}_{\mathbf{d^2}}] \in \mathbb{C}^{\mathbf{m}\times\mathbf{n}}$, and where the blocks $\mathbf{M}_i \in \mathbb{C}^{m\times n}$ satisfy $\mathbf{M} = \mathbf{diag}\{\mathbf{M}_1, \cdots, \mathbf{M}_{\mathbf{d^2}}\}$ a block-diagonal decomposition according to a $d \times d$ tiling in the Fourier domain (see Lemma 1 in [47]). Note that the inverse of diagonal matrix $d^2\mathbf{I} + \underline{\mathbf{M}}\underline{\mathbf{M}}^{\mathsf{H}}$ can be computed element-wise.

We verify the effectiveness of ELDER on SISR using a large variety of blur kernels and different down-sampling factors. We use 8 realistic camera shake kernels tested in [41], plus a 9×9 uniform kernel, and a 25×25 Gaussian kernel with standard deviation 1.6. All 10 kernels are presented in Table 3. We use the same dataset in Section IV-A for training ELDER. We train a single model on all blur kernels, each with downsampling factors of $d \in \{2, 3, 4\}$, to test the generalizability under different SISR settings. We set the number of forward-iterations to $K = 100$. At every training iteration, we initialize $\boldsymbol{x}_0$ with a shift-corrected bicubic interpolation of $\boldsymbol{y}$ [41]. Additionally, during training we use AWGN with random noise levels $\sigma \in [0, 10.0]/255$ to ensure robustness of our model to different noise perturbations.

We compare ELDER and DEQ against bicubic interpolation, RCAN [45], and state-of-the-art PnP methods IRCNN [46], DPIR [41], and GSPnP [21]. We use the publicly available implementations for all the baseline methods. RCAN refers to the PSNR oriented deep model based on bicubic degradation. GSPnP [21] is a PnP method using pre-trained denoising CNN as an explicit regularization functional, with an algorithmic update similar to the ELDER forward pass. DPIR uses the original DRUNet in Table 1, while GSPnP uses the LSR gradient-step denoiser. The regularization and step-size parameters of GSPnP, DEQ, and ELDER were optimized at test time using `fminbound` in `scipy.optimize`.

Table 3 summarizes the PSNR values achieved by ELDER and other methods when applied to $\{\sigma = 7.65, d = 3\}$ and $\{\sigma = 2.55, d = 4\}$ on the CBSD68 dataset. It is clear that both ELDER and DEQ outperform the other methods at all settings. ELDER also matches the overall performance of DEQ, while performing better for higher noise levels at $d = 3$. The excellent performance of ELDER relative to DEQ highlights that use of an explicit regularizer does not imply a compromise in performance. Fig. 2 visually compares ELDER against the baseline methods at scale factors $\times 3$ (top) and $\times 4$ (bottom), respectively. The enlarged regions in the images suggest that ELDER better recovers the fine details and sharper edges compared to DPIR and GSPnP, while providing same or better PSNR than DEQ. These results indicate that ELDER reaches state-of-the-art performance in PnP/DEQ SISR for a variety of kernels and noise levels.

Fig. 3 illustrate the convergence behavior of the forward pass of ELDER in terms of the objective $f(\boldsymbol{x}^k)$ (left) and the residual $\|\boldsymbol{x}^k - \boldsymbol{x}^{k-1}\|_2/\gamma$ (middle), when tested on a subset of 10 color images taken from the original CBSD68 dataset (CBSD10). Fig. 3 (right) shows the corresponding step-size γ used in the experiment. The optimal step-size on each test image is optimized for the best PSNR value. Note how ELDER using the backtracking line-search strategy enables automatic selection of the step-size parameter.

Fig. 4 compares the convergence of ELDER and GSPnP in terms of PSNR for SISR with scale factor 3. The figure also shows visual results for both methods at different iterations. Note how ELDER learns a regularizer that leads to an improved PSNR compared to GSPnP. Since both methods share the same parametrized regularization functional, the improvement is due to DEQ learning of the regularizer.

C. COMPRESSED SENSING MRI (CS-MRI)

MRI is a widely-used medical imaging technology that is limited by the low speed of data acquisition. CS-MRI seeks to address this limitation by recovering an image $\boldsymbol{x}^*$ from its sparsely-sampled Fourier measurements. We simulate a simplified noiseless single-coil CS-MRI using radial Fourier sampling. The measurement operator is thus $\mathbf{A} = \mathbf{B}\mathbf{F}$, where $\mathbf{B}$ is the diagonal sampling matrix with values in $\{0, 1\}$. The proximal operator of $g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \mathbf{B}\mathbf{F}\boldsymbol{x}\|_2^2$ has a closed-form

$$\mathrm{prox}_{\gamma g}(\boldsymbol{z}) = \mathbf{F}^{\mathsf{H}}\left(\frac{\gamma\mathbf{B}^{\mathsf{H}}\boldsymbol{y} + \mathbf{F}\boldsymbol{z}}{\mathbf{I} + \gamma\mathbf{B}^{\mathsf{H}}\mathbf{B}}\right). \tag{20}$$

We train ELDER using the brain dataset from [8], which consists of 800 slices of 256×256 training images and 50 slices of testing

TABLE 3. Average PSNR (dB) Results of Different Methods on CBSD68. The Best and Second Two Results are Highlighted, Respectively

Method	Scale & Noise	Blur Kernel										Average
												
Bicubic	$\sigma = 7.65$, $d = 3$	21.71	21.96	21.31	19.11	22.54	19.36	20.24	20.30	22.11	23.06	21.17
RCAN [45]		20.67	20.99	20.17	18.48	21.32	18.44	19.46	19.45	21.13	22.14	20.23
IRCNN [46]		24.34	24.36	24.98	24.13	25.04	24.60	24.49	24.51	24.11	24.99	24.56
DPIR [41]		24.50	24.63	25.51	24.22	25.75	24.99	24.92	24.97	24.38	25.86	24.97
GSPnP [21]		24.62	24.86	25.63	23.99	25.72	24.17	24.55	24.93	24.36	25.92	24.88
DEQ [10]		24.90	25.03	25.84	24.59	25.91	24.69	25.06	25.15	24.70	26.14	25.20
ELDER (Ours)		**25.15**	**25.31**	**25.97**	**25.01**	**26.11**	**25.65**	**25.53**	**25.49**	**24.84**	**26.24**	**25.53**
Bicubic	$\sigma = 2.55$, $d = 4$	21.60	22.04	21.12	19.40	22.31	19.70	20.46	20.71	22.28	22.64	21.23
RCAN [45]		20.84	21.50	20.04	18.98	21.32	18.77	19.91	20.11	21.84	21.91	20.52
IRCNN [46]		24.17	24.23	24.97	24.05	24.89	24.46	24.12	24.36	24.43	25.18	24.49
DPIR [41]		24.35	24.58	25.33	24.14	25.21	24.78	24.29	24.80	24.79	25.57	24.78
GSPnP [21]		24.03	24.25	25.09	22.88	24.89	23.55	23.21	23.88	24.72	25.41	24.19
DEQ [10]		24.85	24.86	25.62	24.93	**25.53**	**25.37**	**25.11**	**25.26**	**25.17**	**25.87**	**25.26**
ELDER (Ours)		**24.85**	**25.00**	**25.67**	**24.96**	25.45	25.29	25.05	25.18	25.07	25.74	25.22

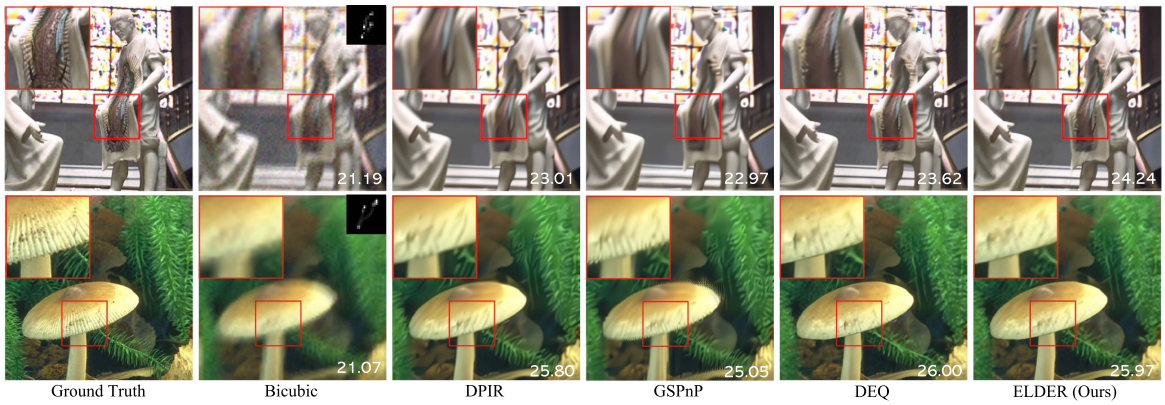

FIGURE 2. Visual comparison of ELDER against several well-known methods on the problem of single image super-resolution (SISR) with scale $d = 3$ (top) and $d = 4$ (bottom). The blur kernels are shown in the zoomed low-resolution (LR) images, respectively. Note how ELDER significantly improves over GSPnP, a recent method for learning explicit regularizers. Additionally, it performs better than DEQ and DPIR, both of which rely on implicit regularization specified by a CNN. This figure shows that ELDER learns explicit regularizers without compromising image quality.

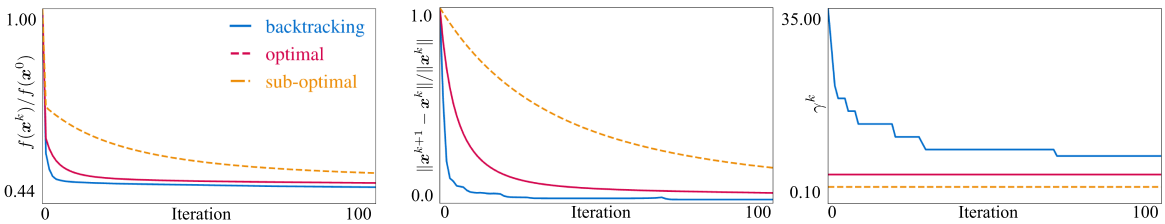

FIGURE 3. Illustration of the convergence of the ELDER forward iterations on SISR using three step-size selection strategies: (a) backtracking, (b) optimal, and (c) sub-optimal. Optimal is the idealized strategy that uses MSE with respect to the ground truth for step-size selection. Backtracking is a practical strategy that uses a backtracking line-search for automatic step-size selection. Suboptimal refers to a step-size selected as 10% of the optimal one. *Left*: Convergence of the objective. *Middle*: Convergence of the updates. *Right*: Step-size values across iterations. This figure illustrates that one of the key benefits of ELDER is that it can use backtracking for automatic step-size selection.

images. Both ELDER and conventional DEQ are trained on sampling ratios within [10, 20%]. At every training iteration, we use zero-filled image $\boldsymbol{x}_0 = \mathbf{A}^{\mathsf{H}}\boldsymbol{y}$ to initialize the forward pass with 100 iterations.

Table 4 presents PSNR values obtained by ELDER, publicly available implementations of several well-known methods, including TV [48], ADMM-Net [49], as well as IRCNN, DPIR, GSPnP, ISP [20], and DEQ. TV is solved using the accelerated proximal gradient descent method [48]. ADMM-Net is a DU method that trains both image transforms and shrinkage functions within the algorithm. ISP refers to an MBDL using explicit deep denoisers, similar to GSPnP. Overall, ELDER and DEQ achieve the best performances, indicating that having the explicit regularizer does not compromise

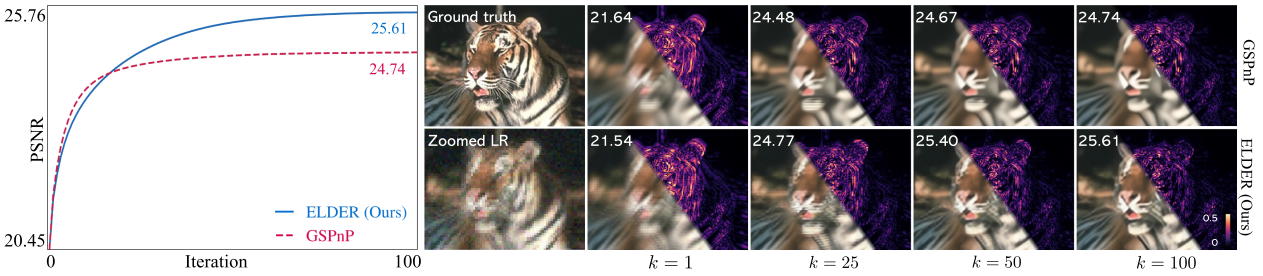

FIGURE 4. **Comparison between ELDER and GSPnP for SISR with $d = 3$. Both methods seek to learn explicit regularizers parametrized by the same CNN architecture, but are trained differently. *Left*: Evolution of the PSNR is plotted against iteration number for both methods. *Right*: Visual illustration of iterates obtained by GSPnP (top) and ELDER (bottom). Note how ELDER outperforms GSPnP due to its ability to learn regularizers that minimize MSE.**

TABLE 4. **Average PSNR (dB) and SSIM Values for Different Methods on CS-MRI**

CS Ratio	Metric	Method							
		TV [48]	ADMM-Net [49]	IRCNN [46]	DPIR [50]	ISP [20]	GSPnP [21]	DEQ [10]	ELDER
10%	PSNR	31.36	34.19	33.86	34.98	34.27	34.86	**35.16**	35.14
	SSIM	0.8200	0.8959	0.8865	0.9060	0.9007	0.9063	0.9114	**0.9116**
20%	PSNR	35.62	37.17	37.84	38.70	38.35	38.56	**38.88**	38.68
	SSIM	0.9121	0.9471	0.9411	0.9464	0.9414	0.9478	**0.9499**	0.9492
Average	PSNR	33.49	35.68	35.52	36.84	36.31	36.71	**37.02**	36.91
	SSIM	0.8660	0.9215	0.9138	0.9262	0.9211	0.9271	**0.9307**	0.9304

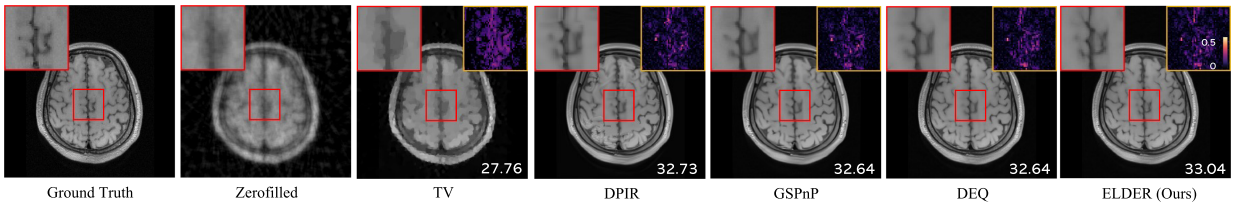

FIGURE 5. **Visual evaluation of several well-known methods on reconstruction of a brain image from its radial Fourier measurements at 10% sampling. Note how ELDER outperforms DPIR, GSPnP, and the traditional DEQ both quantitatively and visually.**

performance. Fig. 5 provides some visual examples at sampling ratio 10%, highlighting the imaging quality obtained by our method relative to several baseline methods. From the zoomed regions and the corresponding error maps, ELDER improves over DPIR and GSPnP due to the training of the explicit regularizer to be end-to-end MSE optimal.

D. IMAGE INPAINTING

We apply ELDER to image inpainting characterized by the measurement model $\boldsymbol{y} = \mathbf{P}\boldsymbol{x}$, where $\mathbf{A} = \mathbf{P}$ is a random diagonal matrix with $p \in [0, 1]$ denoting a probability of missing a pixel. We assume a noiseless setting. The data-fidelity term is the indicator function $g(\boldsymbol{x}) = \Pi_{\mathcal{C}}(\boldsymbol{x})$ for the set $\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{y} = \mathbf{P}\boldsymbol{x}\}$, which, by definition is 0 when $\boldsymbol{x} \in \mathcal{C}$ and $+\infty$ elsewhere. The proximal step (11) has a closed form solution

$$\text{prox}_{\gamma g}(\boldsymbol{z}) = \boldsymbol{y} + \boldsymbol{z} - \mathbf{P}\boldsymbol{z}. \tag{21}$$

At the kth forward iteration of ELDER, the proximal operator returns an image consisting of the network output at the missing pixels and measured pixels at the other locations.

We train our model under random sampling parameter $p \in [0.3, 0.7]$. To demonstrate the flexibility of ELDER, we consider $p = 0.5$ and $p = 0.7$ and compare to IRCNN, DPIR, GSPnP, and DEQ. Again, since IRCNN and DPIR lack the implementation for inpainting, we apply our data fidelity term on them, set the σ_K parameter to 3.0, and set the iteration number to 100. The PSNR performance is reported in Table 5, and visual results are provided in Fig. 6. These results indicate that, achieves better performance compared to GSPnP and nearly matches the performance of DEQ. Moreover, we can observe from the visual results that GSPnP smoothes out the fine details, while DPIR generates distortions. In contrast, ELDER can recover detail as well as avoid distortions.

TABLE 5. **Average PSNR (dB)/SSIM Values for Different Methods for Image Inpainting on CBSD68 Dataset**

Method	Probability of Masking		Average
	50%	70%	
IRCNN	31.61/0.9227	27.87/0.8558	29.77/0.8893
DPIR	31.72/0.9274	28.11/0.8601	29.92/0.8938
GSPnP	31.66/0.9263	27.98/0.8251	29.82/0.8757
DEQ	**32.79/0.9432**	**29.31/0.8820**	**31.05/0.9126**
ELDER (Ours)	32.30/0.9352	28.82/0.8700	30.56/0.9026

FIGURE 6. **Visual illustration of ELDER relative to DPIR, GSPnP and DEQ for image inpainting with masked probability of $p = 0.7$ on CBSD68. Note that ELDER provides substantial improvements over PnP image reconstruction methods, matching the performance of DEQ.**

V. CONCLUSION

ELDER is a new method for learning explicit regularization functionals for model-based deep learning in inverse problems. It parameterizes the regularizer using a CNN and learns its weights to minimize MSE values using DEQ. The benefit of having an explicit regularizer is that one directly inherits the fundamental results from optimization. This article shows that ELDER outperforms existing approaches for learning explicit regularizers for inverse problems. It also suggests that using an *explicit* regularization functional does *not* compromise imaging performance compared to the traditional DEQ methods based on *implicit* regularization.

REFERENCES

[1] 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.

[2] 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, Nov. 2017.

[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.*, vol. 40, no. 1, pp. 85–97, Jan. 2023.

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

[6] S. Sreehari et al., "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 408–423, Dec. 2016.

[7] A. Ebner and M. Haltmeier, "Plug-and-play image reconstruction is a convergent regularization method," 2022, *arXiv:2212.06881*.

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

[9] S. Bai, J. Z. Kolter, and V. Koltun, "Deep equilibrium models," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 690–701.

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

[11] S. W. Fung, H. Heaton, Q. Li, D. McKenzie, S. Osher, and W. Yin, "JFB: Jacobian-free backpropagation for implicit networks," in *Proc. AAAI Conf. Artif. Intell.*, 2022, pp. 6648–6656.

[12] D. Riccio, M. J. Ehrhardt, and M. Benning, "Regularization of inverse problems: Deep equilibrium models versus bilevel learning," 2022, *arXiv:2206.13193*.

[13] Z. Kadkhodaie and E. P. Simoncelli, "Stochastic solutions for linear inverse problems using the prior implicit in a denoiser," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 13242–13254.

[14] S. Mukherjee, A. Hauptmann, O. Öktem, M. Pereyra, and C.-B. Schönlieb, "Learned reconstruction methods with convergence guarantees: A survey of concepts and applications," *IEEE Signal Process. Mag.*, vol. 40, no. 1, pp. 164–182, Jan. 2023.

[15] E. K. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 5546–5557.

[16] Y. Sun, J. Liu, and U. S. Kamilov, "Block coordinate regularization by denoising," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 382–392.

[17] A. Pramanik and M. B. Zimmerman and M. Jacob, "Memory-efficient model-based deep learning with convergence and robustness guarantees," *IEEE Trans. Comput. Imag.*, vol. 9, pp. 260–275, 2023.

[18] J. Liu, X. Xu, W. Gan, S. Shoushtari, and U. S. Kamilov, "Online deep equilibrium learning for regularization by denoising," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2022, pp. 25363–25376.

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

[20] R. Cohen, Y. Blau, D. Freedman, and E. Rivlin, "It has potential: Gradient-driven denoisers for convergent solutions to inverse problems," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2021, pp. 18152–18164.

[21] S. Hurault, A. Leclaire, and N. Papadakis, "Gradient step denoiser for convergent plug-and-play," in *Proc. Int. Conf. Learn. Representations*, 2022.

[22] S. Hurault, A. Leclaire, and N. Papadakis, "Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization," in *Proc. IEEE Int. Conf. Mach. Learn. Appl.*, 2022, pp. 9483–9505.

[23] A. Goujon, S. Neumayer, P. Bohra, S. Ducotterd, and M. Unser, "A neural-network-based convex regularizer for image reconstruction," 2022, *arXiv:2211.12461*.

[24] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2014.

[25] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

[26] R. Gribonval, "Should penalized least squares regression be interpreted as maximum a posteriori estimation?," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2405–2410, May 2011.

[27] E. T. Reehorst and P. Schniter, "Regularization by denoising: Clarifications and new interpretations," *IEEE Trans. Comput. Imag.*, vol. 5, no. 1, pp. 52–67, Mar. 2019.

[28] 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.

[29] W. Gan et al., "Self-supervised deep equilibrium models for inverse problems with theoretical guarantees," 2022, *arXiv:2210.03837*.

[30] Y. Zhao, S. Zheng, and X. Yuan, "Deep equilibrium models for video snapshot compressive imaging," 2022, *arXiv:2201.06931*.

[31] P. L. Combettes and J. C. Pesquet, "Fixed point strategies in data science," *IEEE Trans. Signal Process.*, vol. 69, pp. 3878–3905, 2021.

[32] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *Proc. Int. Conf. Learn. Representations*, 2018.

[33] S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 860–867.

[34] Y. Chen, R. Ranftl, and T. Pock, "Insights into analysis operator learning: From patch-based sparse models to higher order MRFs," *IEEE Trans. Imag. Process.*, vol. 23, no. 3, pp. 1060–1072, Mar. 2014.

[35] S. Lunz, O. Öktem, and C.-B. Schönlieb, "Adversarial regularizers in inverse problems," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 8507–8516.
[36] S. Mukherjee, S. Dittmer, Z. Shumaylov, S. Lunz, O. Öktem, and C. Schönlieb, "Learned convex regularizers for inverse problems," 2020, *arXiv:2008.02839*.
[37] H. Li, J. Schwab, S. Antholzer, and M. Haltmeier, "NETT: Solving inverse problems with deep neural networks," *Inverse Problems*, vol. 36, no. 6, 2020, Art. no. 065005.
[38] E. Kobler, A. Effland, K. Kunisch, and T. Pock, "Total deep variation for linear inverse problems," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 7549–7558.
[39] R. Fermanian, M. L. Pendu, and C. Guillemot, "Learned gradient of a regularizer for plug-and-play gradient descent," 2022, *arXiv:2204.13940*.
[40] A. M. Teodoro, J. M. Bioucas-Dias, and M. Figueiredo, "A convergent image fusion algorithm using scene-adapted Gaussian-mixture-based denoising," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 451–463, Jan. 2019.
[41] 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.*, vol. 44, no. 10, pp. 6360–6376, Oct. 2022.
[42] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.
[43] S. Mukherjee, M. Carioni, O. Öktem, and C. B. Schönlieb, "End-to-end reconstruction meets data-driven regularization for inverse problems," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2021, pp. 21413–21425.
[44] S. Bai, V. Koltun, and J. Z. Kolter, "Neural deep equilibrium solvers," in *Proc. Int. Conf. Learn. Representations*, 2022.
[45] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 286–301.
[46] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3929–3938.
[47] N. Zhao, Q. Wei, A. Basarab, N. Dobigeon, D. Kouamé, and J.-Y. Tourneret, "Fast single image super-resolution using a new analytical solution for $\ell_2 - \ell_2$ problems," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3683–3697, Aug. 2016.
[48] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
[49] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 10–18.
[50] C. Zhang, Y. Liu, F. Shang, Y. Li, and H. Liu, "A novel learned primal-dual network for image compressive sensing," *IEEE Access*, vol. 9, pp. 26041–26050, 2021.