Tuning-free Plug-and-Play Proximal Algorithm for Inverse Imaging Problems

Kaixuan Wei [1] Angelica Aviles-Rivero [2] Jingwei Liang [3] Ying Fu [1] Carola-Bibiane Schnlieb [3] Hua Huang [1]

Abstract

Plug-and-play (PnP) is a non-convex framework that combines ADMM or other proximal algorithms with advanced denoiser priors. Recently, PnP has achieved great empirical success, especially with the integration of deep learning-based denoisers. However, a key problem of PnP based approaches is that they require manual parameter tweaking. It is necessary to obtain high-quality results across the high discrepancy in terms of imaging conditions and varying scene content. In this work, we present a tuning-free PnP proximal algorithm, which can automatically determine the internal parameters including the penalty parameter, the denoising strength and the terminal time. A key part of our approach is to develop a policy network for automatic search of parameters, which can be effectively learned via mixed model-free and model-based deep reinforcement learning. We demonstrate, through numerical and visual experiments, that the learned policy can customize different parameters for different states, and often more efficient and effective than existing handcrafted criteria. Moreover, we discuss the practical considerations of the plugged denoisers, which together with our learned policy yield state-of-the-art results. This is prevalent on both linear and nonlinear exemplary inverse imaging problems, and in particular, we show promising results on Compressed Sensing MRI and phase retrieval.

1. Introduction

The problem of recovering an underlying unknown image $x \in \mathbb{R}^N$ from noisy and/or incomplete measured data $y \in \mathbb{R}^M$ is fundamental in computational imaging, in applications including magnetic resonance imaging (MRI) (Fessler, 2010), computed tomography (CT) (Elbakri & Fessler, 2002), microscopy (Aguet et al., 2008; Zheng et al., 2013), and inverse scattering (Katz et al., 2014; Metzler et al., 2017b). This image recovery task is often formulated as an optimization problem that minimizes a cost function, i.e.,

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \quad \mathcal{D}(x) + \lambda \mathcal{R}(x), \tag{1}$$

where $\mathcal{D}$ is a data-fidelity term that ensures consistency between the reconstructed image and measured data. $\mathcal{R}$ is a regularizer that imposes certain prior knowledge, *e.g.* smoothness (Osher et al., 2005; Ma et al., 2008), sparsity (Yang et al., 2010; Liao & Sapiro, 2008; Ravishankar & Bresler, 2010), low rank (Semerci et al., 2014; Gu et al., 2017) and nonlocal self-similarity (Mairal et al., 2009; Qu et al., 2014), regarding the unknown image. The problem in Eq. (1) is often solved by first-order iterative proximal algorithms, *e.g.* fast iterative shrinkage/thresholding algorithm (FISTA) (Beck & Teboulle, 2009) and alternating direction method of multipliers (ADMM) (Boyd et al., 2011), to tackle the nonsmoothness of the regularizers.

To handle the nonsmoothness caused by regularizers, first-order algorithms rely on the proximal operators (Beck & Teboulle, 2009; Boyd et al., 2011; Chambolle & Pock, 2011; Parikh et al., 2014; Geman, 1995; Esser et al., 2010) defined by

$$\text{Prox}_{\sigma^2\mathcal{R}}(v) = \underset{x}{\text{argmin}} \left(\mathcal{R}(x) + \frac{1}{2\sigma^2}\|x - v\|_2^2\right). \tag{2}$$

Interestingly, given the mathematical equivalence of the proximal operator to the regularized denoising, the proximal operators $\text{Prox}_{\sigma^2\mathcal{R}}$ can be replaced by any off-the-shelf denoisers $\mathcal{H}_\sigma$ with noise level σ, yielding a new framework namely plug-and-play (PnP) prior (Venkatakrishnan et al., 2013). The resulting algorithms, *e.g.* PnP-ADMM, can be written as

$$x_{k+1} = \text{Prox}_{\sigma_k^2\mathcal{R}}(z_k - u_k) = \mathcal{H}_{\sigma_k}(z_k - u_k), \tag{3}$$

$$z_{k+1} = \text{Prox}_{\frac{1}{\mu_k}\mathcal{D}}(x_{k+1} + u_k), \tag{4}$$

$$u_{k+1} = u_k + x_{k+1} - z_{k+1}, \tag{5}$$

where $k \in [0, \tau)$ denotes the k-th iteration, τ is the terminal time, σ_k and μ_k indicate the denoising strength (of the

[1]School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China [2]DPMMS, University of Cambridge, Cambridge, United Kingdom [3]DAMTP, University of Cambridge, Cambridge, United Kingdom.

denoiser) and the penalty parameter used in the k-th iteration respectively.

In this formulation, the regularizer $\mathcal{R}$ can be implicitly defined by a plugged denoiser, which opens a new door to leverage the vast progress made on the image denoising front to solve more general inverse imaging problems. To plug well-known image denoisers, *e.g.* BM3D (Dabov et al., 2007) and NLM (Buades et al., 2005), into optimization algorithms often leads to sizeable performance gain compared to other explicitly defined regularizers, *e.g.* total variantion. That is PnP as a stand-alone framework can combine the benefits of both deep learning based denoisers and optimization methods, e.g. (Zhang et al., 2017b; Rick Chang et al., 2017; Meinhardt et al., 2017). These highly desirable benefits are in terms of *fast and effective inference whilst circumventing the need of expensive network retraining whenever the specific problem changes.*

Whilst a PnP framework offers promising image recovery results, a major drawback is that its performance is highly sensitive to the internal parameter selection, which generically includes the penalty parameter μ, the denoising strength (of the denoiser) σ and the terminal time τ. The body of literature often utilizes manual tweaking e.g. (Rick Chang et al., 2017; Meinhardt et al., 2017) or handcrafted criteria e.g. (Chan et al., 2017; Zhang et al., 2017b; Eksioglu, 2016; Tirer & Giryes, 2018) to select parameters for each specific problem setting. However, manual parameter tweaking requires several trials, which is very cumbersome and time-consuming. Semi-automated handcrafted criteria (for example monotonically decreasing the denoising strength) can, to some degree, ease the burden of exhaustive search of large parameter space, but often leads to suboptimal local minimum. Moreover, the optimal parameter setting differs image-by-image, depending on the measurement model, noise level, noise type and unknown image itself. These differences can be noticed in the further detailed comparison in Fig. 1, where peak signal-to-noise ratio (PSNR) curves are displayed for four images under varying denoising strength.

This paper is devoted to addressing the aforementioned challenge – how to deal with the manual parameter tuning problem in a PnP framework. To this end, we formulate the internal parameter selection as a sequential decision-making problem. To do this, a policy is adopted to select a sequence of internal parameters to guide the optimization. Such problem can be naturally fit into a reinforcement learning (RL) framework, where a policy agent seeks to map observations to actions, with the aim of maximizing cumulative-reward. The reward reflects the *to do* or *not to do* events for the agent, and a desirable high reward can be obtained if the policy leads to a faster convergence and better restoration accuracy.

We demonstrate, through extensive numerical and visual experiments, the advantage of our algorithmic approach on Compressed Sensing MRI and phase retrieval problems. We show that the policy well approximates the intrinsic function that maps the input state to its optimal parameter setting. By using the learned policy, the guided optimization can reach comparable results to the ones using oracle parameters tuned via the inaccessible ground truth. An overview of our algorithm is shown in Fig. 2. Our contributions are as follows:

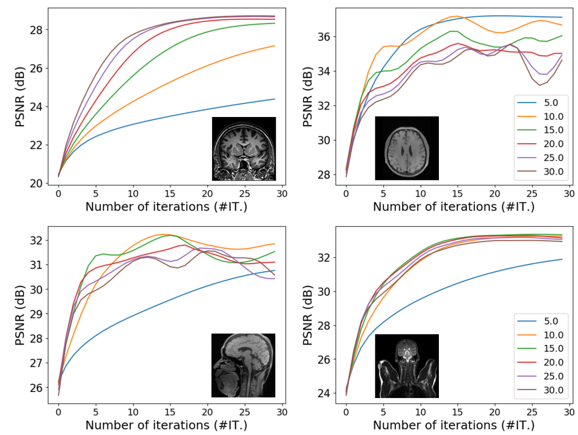

Figure 1. Compressed Sensing MRI using radial sampling pattern with 20% sampling rate, where PSNR curves of four medical images are displayed - using PnP-ADMM with different denoising strengths. Different images requires different denoising strengths to reach the optimal performance.

1. We present a tuning-free PnP algorithm that can customize parameters towards diverse images, which often demonstrates faster practical convergence and better empirical performance than handcrafted criteria.
2. We introduce an efficient mixed model-free and model-based RL algorithm. It can optimize jointly the discrete terminal time, and the continuous denoising strength/penalty parameters.
3. We validate our approach with an extensive range of numerical and visual experiments, and show how the performance of the PnP is affected by the parameters. We also show that our well-designed approach leads to better results than state-of-the-art techniques on compressed sensing MRI and phase retrieval.

2. Related Work

The body of literature has reported several PnP algorithmic techniques. In this section, we provide a short overview of these techniques.

Plug-and-play (PnP). The definitional concept of PnP was first introduced in (Danielyan et al., 2010; Zoran & Weiss,

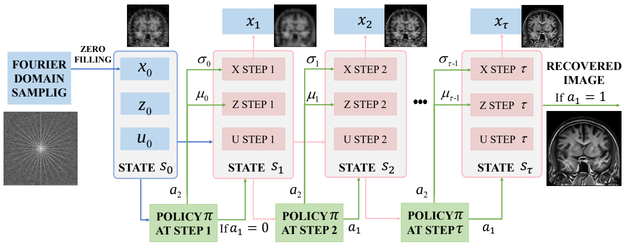

Figure 2. Overview of our tuning-free plug-and-play framework (taking CS-MRI problem as example).

2011; Venkatakrishnan et al., 2013), which has attracted great attention owing to its effectiveness and flexibility to handle a wide range of inverse imaging problems. Following this philosophy, several works have been developed, and can be roughly categorized in terms of four aspects, i.e., proximal algorithms, imaging applications, denoiser priors, and the convergence. **(i)** *proximal algorithms* include half-quadratic splitting (Zhang et al., 2017b), primal-dual method (Ono, 2017), generalized approximate message passing (Metzler et al., 2016b) and (stochastic) accelerated proximal gradient method (Sun et al., 2019a). **(ii)** *imaging applications* have such as bright field electronic tomography (Sreehari et al., 2016); diffraction tomography (Sun et al., 2019a); low-dose CT imaging (He et al., 2018); Compressed Sensing MRI (Eksioglu, 2016); electron microscopy (Sreehari et al., 2017); single-photon imaging (Chan et al., 2017); phase retrieval (Metzler et al., 2018); Fourier ptychography microscopy (Sun et al., 2019b); light-field photography (Chun et al., 2019); hyperspectral sharpening (Teodoro et al., 2018); denoising (Rond et al., 2016); and image processing – e.g. demosaicking, deblurring, super-resolution and inpainting (Heide et al., 2014; Meinhardt et al., 2017; Zhang et al., 2019a; Tirer & Giryes, 2018).

Moreover, **(iii)** *denoiser priors* include BM3D (Heide et al., 2014; Dar et al., 2016; Rond et al., 2016; Sreehari et al., 2016; Chan et al., 2017), nonlocal means (Venkatakrishnan et al., 2013; Heide et al., 2014; Sreehari et al., 2016), Gaussian mixture models (Teodoro et al., 2016; 2018), weighted nuclear norm minimization (Kamilov et al., 2017), and deep learning-based denoisers (Meinhardt et al., 2017; Zhang et al., 2017b; Rick Chang et al., 2017). Finally, **(iv)** *theoretical analysis on the convergence* include the symmetric gradient (Sreehari et al., 2016), the bounded denoiser (Chan et al., 2017) and the nonexpansiveness assumptions (Sreehari et al., 2016; Teodoro et al., 2018; Sun et al., 2019a; Ryu et al., 2019; Chan, 2019).

Differing from these aspects, in this work we focus on the challenge of parameter selection in PnP, where a bad choice of parameters often leads to severe degradation of the results (Romano et al., 2017; Chan et al., 2017). Unlike existing semi-automated parameter tuning criteria (Wang & Chan, 2017; Chan et al., 2017; Zhang et al., 2017b; Eksioglu, 2016; Tirer & Giryes, 2018), *our method is fully automatic and is purely learned from the data, which significantly eases the burden of manual parameter tuning.*

Automated Parameter Selection. There are some works that considering automatic parameter selection in inverse problems. However, the prior term in these works is restricted to certain types of regularizers, *e.g.* Tikhonov regularization (Hansen & Ołeary, 1993; Golub et al., 1979), smoothed versions of the ℓ_p norm (Eldar, 2008; Giryes et al., 2011), or general convex functions (Ramani et al., 2012). To the best of our knowledge, none of them can be applicable to the PnP framework with sophisticated non-convex and learned priors.

Deep Unrolling. Perhaps the most confusable concept to PnP in the deep learning era is the so-called deep unrolling methods (Gregor & LeCun, 2010; Hershey et al., 2014; Wang et al., 2016; Yang et al., 2016; Zhang & Ghanem, 2018; Diamond et al., 2017; Metzler et al., 2017a; Adler & Oktem, 2018; Dong et al., 2018; Xie et al., 2019), which explicitly unroll/truncate iterative optimization algorithms into learnable deep architectures. In this way, the penalty parameters (and the denoiser prior) are treated as trainable parameters, meanwhile the number of iterations has to be fixed to enable end-to-end training. *By contrast, our PnP approach can adaptively select a stop time and penalty parameters given varying input states, though using the off-the-shelf denoiser as prior.*

Reinforcement Learning for Image Recovery. Although Reinforcement Learning (RL) has been applied in a range of domains, from game playing (Mnih et al., 2013; Silver et al., 2016) to robotic control (Schulman et al., 2015), only few works have successfully employed RL to the image recovery tasks. Authors of that (Yu et al., 2018) learned a RL policy to select appropriate tools from a toolbox to progressively restore corrupted images. The work of (Zhang et al., 2019b) proposed a recurrent image restorer whose endpoint was dynamically controlled by a learned policy. In (Furuta et al., 2019), authors used RL to select a sequence of classic filters to process images gradually. The work of (Yu et al., 2019) learned network path selection for image restoration in a multi-path CNN. *In contrast to these works, we apply a mixed model-free and model-based deep RL approach to automatically select the parameters for the PnP image recovery algorithm.*

3. Tuning-free PnP Proximal Algorithm

In this work,we elaborate on our tuning-free PnP proximal algorithm, as described in (3)-(5). This section describes in detail our approach, which contains three main parts. Firstly, we describe how the automated parameter selection is driven.

Secondly, we introduce our environment model, and finally, we introduce the policy learning, which is guided by a mixed model-free and a model-based RL.

It is worth mentioning that our method is generic, and can be applicable to PnP methods derived from other proximal algorithms, *e.g.* forward backward splitting, as well. The reason is that these are distinct methods, they share the same fixed points as PnP-ADMM (Meinhardt et al., 2017).

3.1. RL Formulation for Automated Parameter Selection

This work mainly focuses on the automated parameter selection problem in the PnP framework, where we aim to select a sequence of parameters $(\sigma_0, \mu_0, \sigma_1, \mu_1, \cdots, \sigma_{\tau-1}, \mu_{\tau-1})$ to guide optimization such that the recovered image x^τ is close to the underlying image x. We formulate this problem as a Markov decision process (MDP), which can be addressed via reinforcement learning (RL).

We denote the MDP by the tuple $(\mathcal{S}, \mathcal{A}, p, r)$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, p is the transition function describing the environment dynamics, and r is the reward function. Specifically, for our task, $\mathcal{S}$ is the space of optimization variable states, which includes the initialization (x_0, z_0, u_0) and all intermedia results (x_k, z_k, u_k) in the optimization process. $\mathcal{A}$ is the space of internal parameters, including both discrete terminal time τ and the continuous denoising strength/penalty parameters (σ_k, μ_k). The transition function $p : \mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}$ maps input state $s \in \mathcal{S}$ to its outcome state $s' \in \mathcal{S}$ after taking action $a \in \mathcal{A}$. The state transition can be expressed as $s_{t+1} = p(s_t, a_t)$, which is composed of one or several iterations of optimization. On each transition, the environment emits a reward in terms of the reward function $r : S \times \mathcal{A} \mapsto \mathbb{R}$, which evaluates actions given the state. Applying a sequence of parameters to the initial state s_0 results in a trajectory T of states, actions and rewards: $T = \{s_0, a_0, r_0, \cdots, s_N, a_N, r_N\}$. Given a trajectory T, we define the return r_t^γ as the summation of discounted rewards after s_t,

$$r_t^\gamma = \sum_{t'=0}^{N-t} \gamma^{t'} r(s_{t+t'}, a_{t+t'}), \tag{6}$$

where $\gamma \in [0, 1]$ is a discount factor and prioritizes earlier rewards over later ones.

Our goal is to learn a policy π, denoted as $\pi(a|s) : \mathcal{S} \mapsto \mathcal{A}$ for the decision-making agent, in order to maximize the objective defined as

$$J(\pi) = \mathbb{E}_{s_0 \sim S_0, T \sim \pi} [r_0^\gamma], \tag{7}$$

where $\mathbb{E}$ represents expectation, s_0 is the initial state, and S_0 is the corresponding initial state distribution. Intuitively, the objective describes the expected return over all possible trajectories induced by the policy π. The expected return on states and state-action pairs under the policy π are defined by state-value functions V^π and action-value functions Q^π respectively, i.e.,

$$V^\pi(s) = \mathbb{E}_{T \sim \pi} [r_0^\gamma | s_0 = s], \tag{8}$$

$$Q^\pi(s, a) = \mathbb{E}_{T \sim \pi} [r_0^\gamma | s_0 = s, a_0 = a]. \tag{9}$$

In our task, we decompose actions into two parts: a discrete decision a_1 on terminal time and a continuous decision a_2 on denoising strength and penalty parameter. The policy also consists of two sub-policies: $\pi = (\pi_1, \pi_2)$, a stochastic policy and a deterministic policy that generate a_1 and a_2 respectively. The role of π_1 is to decide whether to terminate the iterative algorithm when the next state is reached. It samples a boolean-valued outcome a_1 from a two-class categorical distribution $\pi_1(\cdot|s)$, whose probability mass function is calculated from the current state s. We move forward to the next iteration if $a_1 = 0$, otherwise the optimization would be terminated to output the final state. Compared to the stochastic policy π_1, we treat π_2 deterministically, *i.e.* $a_2 = \pi_2(s)$ since π_2 is differentiable with respect to the environment, such that its gradient can be precisely estimated.

3.2. Environment Model

In RL, the environment is characterized by two components: the environment dynamics and reward function. In our task, the environment dynamics is described by the transition function p related to the PnP-ADMM. Here, we elucidate the detailed setting of the PnP-ADMM as well as the reward function used for training policy.

Denoiser Prior. Differentiable environment makes the policy learning more efficient. To make the environment differentiable with respect to π_2[1], we take a convolutional neural network (CNN) denoiser as the image prior. In practice, we use a residual U-Net (Ronneberger et al., 2015) architecture, which was originally designed for medical image segmentation, but was founded to be useful in image denoising recently. Besides, we incorporate an additional tunable noise level map into the input as (Zhang et al., 2018), enabling us to provide continuous noise level control (*i.e.* different denoising strength) within a single network.

Proximal operator of data-fidelity term. Enforcing consistency with measured data requires evaluating the proximal operator in (4). For inverse problems, there might exist fast solutions due to the special structure of the observation model. We adopt the fast solution if feasible (*e.g.* closed-form solution using fast Fourier transform, rather than the general matrix inversion) otherwise a single step of gradient descent is performed as an inexact solution for (4).

[1] π_1 is non-differentiable towards environment regardless of the formulation of the environment.

Transition function. To reduce the computation cost, we define the transition function p to involve m iterations of the optimization. At each time step, the agent thus needs to decide the internal parameters for m iterates. We set $m = 5$ and the max time step $N = 6$ in our algorithm, leading to 30 iterations of the optimization at most.

Reward function. To take both image recovery performance and runtime efficiency into account, we define the reward function as

$$r(s_t, a_t) = \zeta(p(s_t, a_t)) - \zeta(s_t) - \eta. \tag{10}$$

The first term, $\zeta(p(s_t, a_t)) - \zeta(s_t)$, denotes the PSNR increment made by the policy, where $\zeta(s_t)$ denotes the PSNR of the recovered image at step t. A higher reward is acquired if the policy leads to higher performance gain in terms of PSNR. The second term, η, implies penalizing the policy as it does not select to terminate at step t, where η sets the degree of penalty. A negative reward is given if the PSNR gain does not exceed the degree of penalty, thereby encouraging the policy to early stop the iteration with diminished return. We set $\eta = 0.05$ in our algorithm[2].

3.3. RL-based policy learning

In this section, we present a mixed model-free and model-based RL algorithm to learn the policy. Specifically, model-free RL (agnostic to the environment dynamics) is used to train π_1, while model-based RL is utilized to optimize π_2 to make full use of the environment model[3]. We apply the actor-critic framework (Sutton et al., 2000), that uses a policy network $\pi_\theta(a_t|s_t)$ (actor) and a value network $V^\pi_\phi(s_t)$ (critic) to formulate the policy and the state-value function respectively[4]. The policy and the value networks are learned in an interleaved manner. For each gradient step, we optimize the value network parameters ϕ by minimizing

$$L_\phi = \mathbb{E}_{s\sim D, a\sim\pi_\theta(s)}\left[\frac{1}{2}(r(s,a) + \gamma V^\pi_{\hat{\phi}}(p(s,a)) - V^\pi_\phi(s))^2\right], \tag{11}$$

where D is the distribution of previously sampled states, practically implemented by a state buffer. This partly serves as a role of the experience replay mechanism (Lin, 1992), which is observed to "smooth" the training data distribution (Mnih et al., 2013). The update makes use of a target value network $V^\pi_{\hat{\phi}}$, where $\hat{\phi}$ is the exponentially moving average of the value network weights and has been shown to stabilize training (Mnih et al., 2015).

[2]The choice of the hyperparameters m, N and η is discussed in the *suppl. material*.

[3]π_2 can also be optimized in a model-free manner. The comparison can be found in the Section 4.2.

[4]Details of networks are given in the *suppl. material*.

Table 1. Comparisons of different CNN-based denoisers: we show the results of (1) Gaussian denoising performance (PSNR) under noise level $\sigma = 50$; (2) the CS-MRI performance (PSNR) when plugged into the PnP-ADMM; (3) the GPU runtime (ms) of denoisers when processing an image with size 256×256.

Performance	DnCNN	MemNet	UNet
DENOISING PERF.	27.18	27.32	27.40
PNP PERF.	25.43	25.67	25.76
TIMES	8.09	64.65	5.65

The policy network has two sub-policies, which employs shared convolutional layers to extract image features, followed by two separated groups of fully-connected layers to produce termination probability $\pi_1(\cdot|s)$ (after softmax) or denoising strength/penalty parameters $\pi_2(s)$ (after sigmoid). We denote the parameters of the sub-polices as θ_1 and θ_2 respectively, and we seek to optimize $\theta = (\theta_1, \theta_2)$ so that the objective $J(\pi_\theta)$ is maximized. The policy network is trained using policy gradient methods (Peters & Schaal, 2006). The gradient of θ_1 is estimated in a model-free manner by a likelihood estimator, while the gradient of θ_2 is estimated relying on backpropagation via environment dynamics in a model-based manner. Specifically, for discrete terminal time decision π_1, we apply the policy gradient theorem (Sutton et al., 2000) to obtain unbiased Monte Carlo estimate of $\nabla_{\theta_1} J(\pi_\theta)$ using advantage function $A^\pi(s,a) = Q^\pi(s,a) - V^\pi(s)$ as target, i.e.,

$$\nabla_{\theta_1} J(\pi_\theta) = \mathbb{E}_{s\sim D, a\sim\pi_\theta(s)}\left[\nabla_{\theta_1} \log \pi_1(a_1|s)\, A^\pi(s,a)\right]. \tag{12}$$

For continuous denoising strength and penalty parameter selection π_2, we utilize the deterministic policy gradient theorem (Silver et al., 2014) to formulate its gradient, i.e.,

$$\nabla_{\theta_2} J(\pi_\theta) = \mathbb{E}_{s\sim D, a\sim\pi_\theta(s)}\left[\nabla_{a_2} Q^\pi(s,a) \nabla_{\theta_2}\pi_2(s)\right], \tag{13}$$

where we approximate the action-value function $Q^\pi(s,a)$ by $r(s,a) + \gamma V^\pi_{\hat{\phi}}(p(s,a))$ given its unfolded definition.

Using the chain rule, we can directly obtain the gradient of θ_2 by backpropagation via the reward function, the value network and the transition function, in contrast to relying on the gradient backpropagated from only the learned action-value function in the model-free DDPG algorithm (Lillicrap et al., 2016).

4. Experiments

In this section, we detail the experiments and evaluate our proposed algorithm. We mainly focus on the tasks of Compressed Sensing MRI (CS-MRI) and phase retrieval (PR), which are the representative linear and nonlinear inverse imaging problems respectively.

Table 2. Comparisons of different policies used in PnP-ADMM algorithm for CS-MRI on seven widely used medical images under various acceleration factors (x2/x4/x8) and noise level 15. We show both PSNR and the number of iterations (#IT.) used to induce results. * denotes to report the best PSNR over all iterations (*i.e.* with optimal early stopping). The best results are indicated by orange color and the second best results are denoted by blue color.

Policies	$\times 2$		$\times 4$		$\times 8$	
	PSNR	#IT.	PSNR	#IT.	PSNR	#IT.
handcrafted	30.05	30.0	27.90	30.0	25.76	30.0
handcrafted*	30.06	29.1	28.20	18.4	26.06	19.4
fixed	23.94	30.0	24.26	30.0	22.78	30.0
fixed*	28.45	1.6	26.67	3.4	24.19	7.3
fixed optimal	30.02	30.0	28.27	30.0	26.08	16.7
fixed optimal*	30.03	6.7	28.34	12.6	26.16	30.0
oracle	30.25	30.0	28.60	30.0	26.41	30.0
oracle*	30.26	8.0	28.61	13.9	26.45	21.6
model-free	28.79	30.0	27.95	30.0	26.15	30.0
Ours	30.33	5.0	28.42	5.0	26.44	15.0

4.1. Implementation Details

Our algorithm requires two training processes for: the denoising network and the policy network (and value network). For training the denoising network, we follow the common practice that uses 87,000 overlapping patches (with size 128×128) drawn from 400 images from the BSD dataset (Martin et al., 2001). For each patch, we add white Gaussian noise with noise level sampled from $[1, 50]$. The denoising networks are trained with 50 epoch using L_1 loss and Adam optimizer (Kingma & Ba, 2014) with batch size 32. The base learning rate is set to 10^{-4} and halved at epoch 30, then reduced to 10^{-5} at epoch 40.

To train the policy network and value network, we use the 17,125 resized images with size 128×128 from the PASCAL VOC dataset (Everingham et al., 2014). Both networks are trained using Adam optimizer with batch size 48 and 1500 iterations, with a base learning rate of 3×10^{-4} for the policy network and 10^{-3} for the value network. Then we set these learning rates to 10^{-4} and 3×10^{-4} at iteration 1000. We perform 10 gradient steps at every iteration.

For the CS-MRI application, a single policy network is trained to handle multiple sampling ratios (with x2/x4/x8 acceleration) and noise levels (5/10/15), simultaneously. Similarly, one policy network is learned for phase retrieval under different settings.

4.2. Compressed sensing MRI

The forward model of CS-MRI can be mathematically described as $y = \mathcal{F}_p x + \omega$, where $x \in \mathbb{C}^N$ is the underlying image, the operator $\mathcal{F}_p : \mathbb{C}^N \to \mathbb{C}^M$, with $M < N$, denotes the partially-sampled Fourier transform, and $\omega \sim \mathcal{N}(0, \sigma_n I_M)$ is the additive white Gaussian noise. The data-fidelity term is $\mathcal{D}(x) = \frac{1}{2}\|y - \mathcal{F}_p x\|^2$ whose proximal operator is given in (Eksioglu, 2016).

Denoiser priors. To show how denoiser priors affect the performance of the PnP, we train three state-of-the-art CNN-based denoisers, *i.e.* DnCNN (Zhang et al., 2017a), MemNet (Tai et al., 2017) and residual UNet (Ronneberger et al., 2015), with tunable noise level map. We compare both the Gaussian denoising performance and the PnP performance[5] using these denoisers. As shown in Table 1, the residual UNet and MemNet consistently outperform DnCNN in terms of denoising and CS-MRI. It seems to imply a better Gaussian denoiser is also a better denoiser prior for the PnP framework[6]. Since UNet is significantly faster than MemNet, we choose UNet as our denoiser prior.

Comparisons of different policies. We start by giving some insights of our learned policy by comparing the performance of PnP-ADMM with different polices: i) the handcrafted policy used in IRCNN (Zhang et al., 2017b); ii) the fixed policy that uses fixed parameters ($\sigma = 15$, $\mu = 0.1$); iii) the fixed optimal policy that adopts fixed parameters searched to maximize the average PSNR across all testing images; iv) the oracle policy that uses different parameters for different images such that the PSNR of each image is maximized and v) our learned policy based on a learned policy network to optimize parameters for each image. We remark that all compared polices are run for 30 iteration whilst ours automatically choose the terminal time.

To understand the usefulness of the early stopping mechanism, we also report the results of these polices with optimal early stopping[7]. Moreover, we analyze whether the model-based RL benefits our algorithm by comparing it with the learned policy by model-free RL whose π_2 is optimized using the model-free DDPG algorithm (Lillicrap et al., 2016).

The results of all aforementioned policies are provided in Table 2. We can see that the bad choice of parameters (see "fixed") induces poor results, in which the early stopping is quite needed to rescue performance (see "fixed*"). When the parameters are properly assigned, the early stopping would be helpful to reduce computation cost. Our learned policy leads to fast practical convergence as well as excellent performance, sometimes even outperforms the oracle policy tuned via inaccessible ground truth (in $\times 2$ case). We note this is owing to the varying parameters across iterations generated automatically in our algorithm, which yield extra flexibility than constant parameters over iterations. Besides, we find the learned model-free policy produces suboptimal

[5] We exhaustively search the best denoising strength/penalty parameters to exclude the impact of internal parameters.

[6] Further investigation of this argument can be found in the *suppl. material*.

[7] It should be noted some policies (*e.g.* "fixed optimal" and "oracle") requires to access the ground truth to determine parameters, which is generally impractical in real testing scenarios.

Table 3. Quantitative results (PSNR) of different CS-MRI methods on two datasets under various acceleration factors f and noise levels σ_n. The best results are indicated by orange color and the second best results are denoted by blue color.

DATASET	f	σ_n	TRADITIONAL		DEEP UNROLLING		PNP		
			RecPF	FCSA	ADMMNet	ISTANet	BM3D-MRI	IRCNN	Ours
Medical7	×2	5	32.46	31.70	33.10	34.58	33.33	34.67	34.78
		10	29.48	28.33	31.37	31.81	29.44	31.80	32.00
		15	27.08	25.52	29.16	29.99	26.90	29.96	30.27
	×4	5	28.67	28.21	30.24	31.34	30.33	31.36	31.62
		10	26.98	26.67	29.20	29.71	28.30	29.52	29.68
		15	25.58	24.93	27.87	28.38	26.66	27.94	28.43
	×8	5	24.72	24.62	26.57	27.65	26.53	27.32	28.26
		10	23.94	24.04	26.21	26.90	25.81	26.44	27.35
		15	23.18	23.36	25.49	26.23	25.09	25.53	26.41
MICCAI	×2	5	36.39	34.90	36.74	38.17	36.00	38.42	38.57
		10	31.95	30.12	34.20	34.81	31.39	34.93	35.06
		15	28.91	26.68	31.42	32.65	28.46	32.81	33.09
	×4	5	33.05	32.30	34.15	35.46	34.79	35.80	36.11
		10	30.21	29.56	32.58	33.13	31.63	32.99	33.07
		15	28.13	26.93	30.55	31.48	29.35	30.98	31.42
	×8	5	28.35	28.71	30.36	31.62	31.34	31.66	32.64
		10	26.86	27.68	29.78	30.54	29.86	30.16	30.89
		15	25.70	26.35	28.83	29.50	28.53	28.72	29.65

Table 4. Quantitative results of different PR algorithms on four CDP measurements and varying amount of Possion noise (large α indicates low sigma-to-noise ratio).

Algorithms	$\alpha = 9$ PSNR	$\alpha = 27$ PSNR	$\alpha = 81$ PSNR
HIO	35.96	25.76	14.82
WF	34.46	24.96	15.76
DOLPHIn	29.93	27.45	19.35
SPAR	35.20	31.82	22.44
BM3D-prGAMP	40.25	32.84	25.43
prDeep	39.70	33.54	26.82
Ours	40.33	33.90	27.23

denoising strength/penalty parameters compared with our mixed model-free and model-based policy, and it also fails to learn early stopping behavior.

Comparisons with state-of-the-arts. We compare our method against six state-of-the-art methods for CS-MRI, including the traditional optimization-based approaches (RecPF (Yang et al., 2010) and FCSA (Huang et al., 2010)), the PnP approaches (BM3D-MRI (Eksioglu, 2016) and IRCNN (Zhang et al., 2017b)), and the deep unrolling approaches (ADMMNet (Yang et al., 2016) and ISTANet (Zhang & Ghanem, 2018)). To keep comparison fair, for each deep unrolling method, only single network is trained to tackle all the cases using the same dataset as ours. Table 3 shows the method performance on two set of medical images, *i.e.* 7 widely used medical images (Medical7) (Huang et al., 2010) and 50 medical images from MICCAI 2013 grand challenge dataset[8]. The visual comparison can be found in Fig. 3. It can be seen that our approach significantly outperforms the state-of-the-art PnP method (IRCNN) by a large margin, especially under the difficult ×8 case. In the simple cases (*e.g.* ×2), our algorithm only runs 5 iterations to arrive at the desirable performance, in contrast with 30 or 70 iterations required in IRCNN and BM3D-MRI respectively.

[8] https://my.vanderbilt.edu/masi/

4.3. Phase retrieval

The goal of phase retrieval (PR) is to recover the underlying image from only the amplitude, or intensity of the output of a complex linear system. Mathematically, PR can be defined as the problem of recovering a signal $x \in \mathbb{R}^N$ or $\mathbb{C}^N$ from measurement y of the form $y = |Ax| + \omega$, where the measurement matrix A represents the forward operator of the system, and ω represents shot noise. We approximate it with $\omega \sim \mathcal{N}(0, \alpha|Ax|)$. The term α controls the sigma-to-noise ratio in this problem.

We test algorithms with coded diffraction pattern (CDP) (Cands et al., 2015). Multiple measurements, with different random spatial modulator (SLM) patterns are recorded. We model the capture of four measurements using a phase-only SLM as (Metzler et al., 2018). Each measurement operator can be mathematically described as $A_i = \mathcal{F}D_i, \quad i \in [1, 2, 3, 4]$, where $\mathcal{F}$ can be represented by the 2D Fourier transform and D_i is diagonal matrices with nonzero elements drawn uniformly from the unit circle in the complex planes.

We compare our method with three classic approaches (HIO (Fienup, 1982), WF (Candes et al., 2014), and DOLPHIn (Mairal et al., 2016)) and three PnP approaches (SPAR

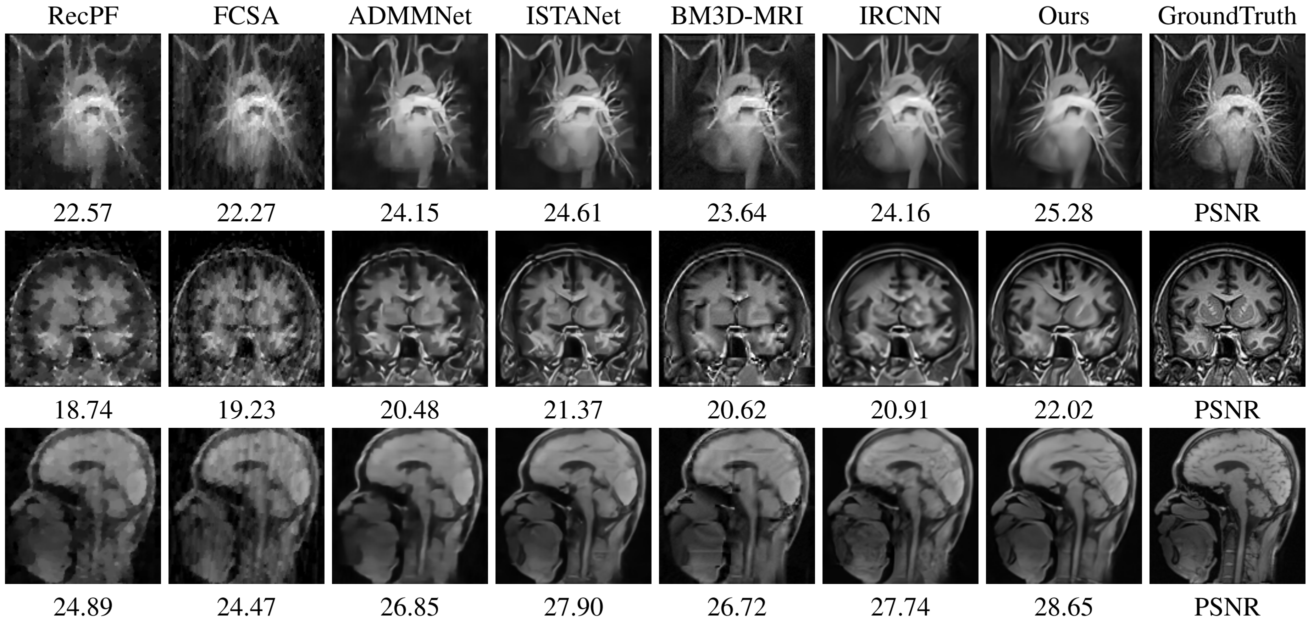

Figure 3. CS-MRI reconstruction results of different algorithms on medical images. (**best view on screen with zoom**).

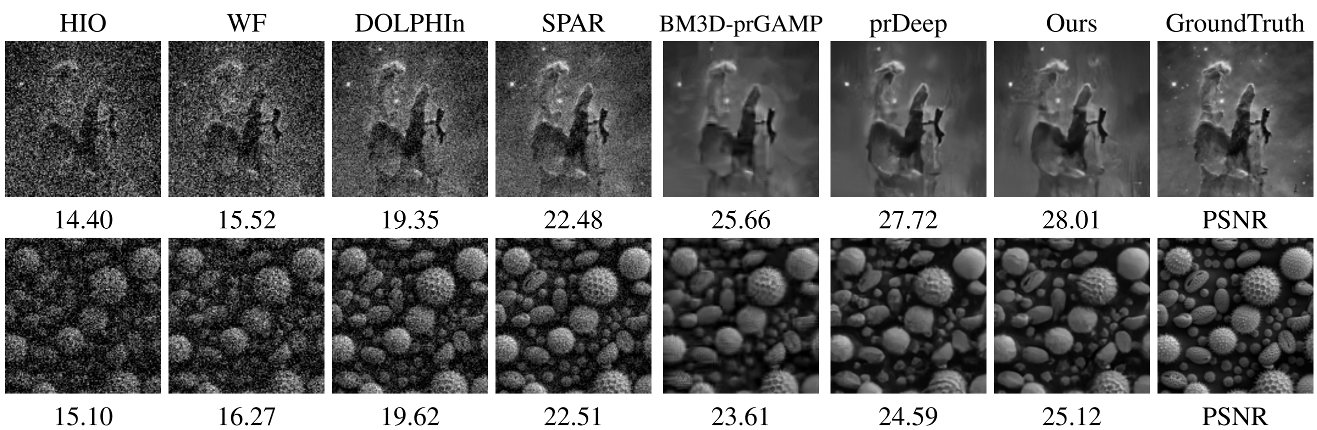

Figure 4. Recovered images from noisy intensity-only CDP measurements with seven PR algorithms. (**Details are better appreciated on screen.**).

(Katkovnik, 2017), BM3D-prGAMP (Metzler et al., 2016a) and prDeep (Metzler et al., 2018)). Table 4 and Fig. 4 summarize the results of all competing methods on twelve images used in (Metzler et al., 2018). It can be seen that our method still leads to state-of-the-art performance in this nonlinear inverse problem, and produces cleaner and clearer results than other competing methods.

5. Conclusion

In this work, we introduce RL into the PnP framework, yielding a novel tuning-free PnP proximal algorithm for a wide range of inverse imaging problems. We underline the main message of our approach *the main strength of our proposed method is the policy network, which can customize well-suited parameters for different images.* Through numerical experiments, we demonstrate our learned policy often generates highly-effective parameters, which even often reaches to the comparable performance to the "oracle" parameters tuned via the inaccessible ground truth.

References

Adler, J. and Oktem, O. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 37(6): 1322–1332, 2018.

Aguet, F., Van De Ville, D., and Unser, M. Model-based 2.5-d deconvolution for extended depth of field in brightfield microscopy. *IEEE Transactions on Image Processing*, 17 (7):1144–1153, 2008.

Beck, A. and Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

Buades, A., Coll, B., and Morel, J.-M. A non-local algorithm for image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 60–65, 2005.

Candes, E., Li, X., and Soltanolkotabi, M. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61, 07 2014.

Cands, E. J., Li, X., and Soltanolkotabi, M. Phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 39(2):277–299, 2015.

Chambolle, A. and Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

Chan, S. H. Performance analysis of plug-and-play admm: A graph signal processing perspective. *IEEE Transactions on Computational Imaging*, 5(2):274–286, 2019.

Chan, S. H., Wang, X., and Elgendy, O. A. Plug-and-play admm for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1):84–98, 2017.

Chun, I. Y., Huang, Z., Lim, H., and Fessler, J. A. Momentum-net: Fast and convergent iterative neural network for inverse problems. *arXiv preprint arXiv:1907.11818*, 2019.

Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8): 2080, 2007.

Danielyan, A., Katkovnik, V., and Egiazarian, K. Image deblurring by augmented lagrangian with bm3d frame prior. In *Workshop on Information Theoretic Methods in Science and Engineering*, pp. 16–18, 2010.

Dar, Y., Bruckstein, A. M., Elad, M., and Giryes, R. Postprocessing of compressed images via sequential denoising. *IEEE Transactions on Image Processing*, 25(7):3044–3058, 2016.

Diamond, S., Sitzmann, V., Heide, F., and Wetzstein, G. Unrolled optimization with deep priors. *arXiv preprint arXiv:1705.08041*, 2017.

Dong, W., Wang, P., Yin, W., Shi, G., Wu, F., and Lu, X. Denoising prior driven deep neural network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(10):2305–2318, 2018.

Eksioglu, E. M. Decoupled algorithm for mri reconstruction using nonlocal block matching model: Bm3d-mri. *Journal of Mathematical Imaging and Vision*, 56(3):430–440, 2016.

Elbakri, I. A. and Fessler, J. A. Segmentation-free statistical image reconstruction for polyenergetic x-ray computed tomography. In *IEEE International Symposium on Biomedical Imaging*, pp. 828–831, 2002.

Eldar, Y. C. Generalized sure for exponential families: Applications to regularization. *IEEE Transactions on Signal Processing*, 57(2):471–481, 2008.

Esser, E., Zhang, X., and Chan, T. F. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.

Everingham, M., Eslami, S., Van Gool, L., Williams, C., Winn, J., and Zisserman, A. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111, 01 2014.

Fessler, J. A. Model-based image reconstruction for mri. *IEEE Signal Processing Magazine*, 27(4):81–89, 2010.

Fienup, J. R. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15):2758–2769, 1982.

Furuta, R., Inoue, N., and Yamasaki, T. Fully convolutional network with multi-step reinforcement learning for image processing. In *AAAI Conference on Artificial Intelligence*, pp. 3598–3605, 2019.

Geman, D. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*, 4(7):932–946, 1995.

Giryes, R., Elad, M., and Eldar, Y. C. The projected gsure for automatic parameter tuning in iterative shrinkage methods. *Applied and Computational Harmonic Analysis*, 30 (3):407–422, 2011.

Golub, G. H., Heath, M., and Wahba, G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

Gregor, K. and LeCun, Y. Learning fast approximations of sparse coding. In *International Conference on Machine Learning (ICML)*, pp. 399–406, 2010.

Gu, S., Xie, Q., Meng, D., Zuo, W., Feng, X., and Zhang, L. Weighted nuclear norm minimization and its applications to low level vision. *International Journal of Computer Vision*, 121(2):183–208, 2017.

Hansen, P. C. and Ołeary, D. P. The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6):1487–1503, 1993.

He, J., Yang, Y., Wang, Y., Zeng, D., Bian, Z., Zhang, H., Sun, J., Xu, Z., and Ma, J. Optimizing a parameterized plug-and-play admm for iterative low-dose ct reconstruction. *IEEE Transactions on Medical Imaging*, 38(2): 371–382, 2018.

Heide, F., Steinberger, M., Tsai, Y.-T., Rouf, M., Pajak, D., Reddy, D., Gallo, O., Liu, J., Heidrich, W., Egiazarian, K., et al. Flexisp: A flexible camera image processing framework. *ACM Transactions on Graphics*, 33(6):231, 2014.

Hershey, J. R., Roux, J. L., and Weninger, F. Deep unfolding: Model-based inspiration of novel deep architectures. *arXiv preprint arXiv:1409.2574*, 2014.

Huang, J., Zhang, S., and Metaxas, D. Efficient mr image reconstruction for compressed mr imaging. *Medical Image Analysis*, 15:135–142, 2010.

Kamilov, U. S., Mansour, H., and Wohlberg, B. A plug-and-play priors approach for solving nonlinear imaging inverse problems. *IEEE Signal Processing Letters*, 24 (12):1872–1876, 2017.

Katkovnik, V. Phase retrieval from noisy data based on sparse approximation of object phase and amplitude. *arXiv preprint arXiv:1709.01071*, 2017.

Katz, O., Heidmann, P., Fink, M., and Gigan, S. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. *Nature Photonics*, 8(10):784, 2014.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Liao, H. Y. and Sapiro, G. Sparse representations for limited data tomography. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1375–1378. IEEE, 2008.

Lillicrap, T., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. Continuous control with deep reinforcement learning. *international conference on learning representations (ICLR)*, 2016.

Lin, L. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning*, 8(3):293–321, 1992.

Ma, S., Yin, W., Zhang, Y., and Chakraborty, A. An efficient algorithm for compressed mr imaging using total variation and wavelets. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8. IEEE, 2008.

Mairal, Julien, Tillmann, Andreas, M., Eldar, Yonina, and C. Dolphin-dictionary learning for phase retrieval. *IEEE Transactions on Signal Processing*, 2016.

Mairal, J., Bach, F. R., Ponce, J., Sapiro, G., and Zisserman, A. Non-local sparse models for image restoration. In *IEEE International Conference on Computer Vision (ICCV)*, volume 29, pp. 54–62, 2009.

Martin, D., Fowlkes, C., Tal, D., and Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 416–423, 2001.

Meinhardt, T., Moller, M., Hazirbas, C., and Cremers, D. Learning proximal operators: Using denoising networks for regularizing inverse imaging problems. In *IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

Metzler, C., Mousavi, A., and Baraniuk, R. Learned d-amp: Principled neural network based compressive image recovery. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 1772–1783. 2017a.

Metzler, C., Schniter, P., Veeraraghavan, A., et al. prdeep: Robust phase retrieval with a flexible deep network. In *International Conference on Machine Learning (ICML)*, pp. 3498–3507, 2018.

Metzler, C. A., Maleki, A., and Baraniuk, R. G. Bm3d-prgamp: Compressive phase retrieval based on bm3d denoising. In *IEEE International Conference on Image Processing*, 2016a.

Metzler, C. A., Maleki, A., and Baraniuk, R. G. From denoising to compressed sensing. *IEEE Transactions on Information Theory*, 62(9):5117–5144, 2016b.

Metzler, C. A., Sharma, M. K., Nagesh, S., Baraniuk, R. G., Cossairt, O., and Veeraraghavan, A. Coherent inverse scattering via transmission matrices: Efficient phase retrieval algorithms and a public dataset. In *IEEE International Conference on Computational Photography (ICCP)*, pp. 1–16, 2017b.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540): 529–533, 2015.

Ono, S. Primal-dual plug-and-play image restoration. *IEEE Signal Processing Letters*, 24(8):1108–1112, 2017.

Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling and Simulation*, 4(2):460–489, 2005.

Parikh, N., Boyd, S., et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.

Peters, J. and Schaal, S. Policy gradient methods for robotics. *International Conference on Intelligent Robots and Systems (IROS)*, pp. 2219–2225, 2006.

Qu, X., Hou, Y., Lam, F., Guo, D., Zhong, J., and Chen, Z. Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator. *Medical Image Analysis*, 18(6):843–856, 2014.

Ramani, S., Liu, Z., Rosen, J., Nielsen, J.-F., and Fessler, J. A. Regularization parameter selection for nonlinear iterative image restoration and mri reconstruction using gcv and sure-based methods. *IEEE Transactions on Image Processing*, 21(8):3659–3672, 2012.

Ravishankar, S. and Bresler, Y. Mr image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Transactions on Medical Imaging*, 30(5): 1028–1041, 2010.

Rick Chang, J. H., Li, C.-L., Poczos, B., Vijaya Kumar, B. V. K., and Sankaranarayanan, A. C. One network to solve them all – solving linear inverse problems using deep projection models. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.

Romano, Y., Elad, M., and Milanfar, P. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.

Rond, A., Giryes, R., and Elad, M. Poisson inverse problems by the plug-and-play scheme. *Journal of Visual Communication and Image Representation*, 41:96–108, 2016.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, 2015.

Ryu, E., Liu, J., Wang, S., Chen, X., Wang, Z., and Yin, W. Plug-and-play methods provably converge with properly trained denoisers. In *International Conference on Machine Learning (ICML)*, pp. 5546–5557, 2019.

Schulman, J., Levine, S., Abbeel, P., Jordan, M., and Moritz, P. Trust region policy optimization. In *International Conference on Machine Learning (ICML)*, pp. 1889–1897, 2015.

Semerci, O., Hao, N., Kilmer, M. E., and Miller, E. L. Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. *IEEE Transactions on Image Processing*, 23(4):1678–1693, 2014.

Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., and Riedmiller, M. Deterministic policy gradient algorithms. *International Conference on Machine Learning (ICML)*, 2014.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484, 2016.

Sreehari, S., Venkatakrishnan, S. V., Wohlberg, B., Buzzard, G. T., Drummy, L. F., Simmons, J. P., and Bouman, C. A. Plug-and-play priors for bright field electron tomography and sparse interpolation. *IEEE Transactions on Computational Imaging*, 2(4):408–423, 2016.

Sreehari, S., Venkatakrishnan, S., Bouman, K. L., Simmons, J. P., Drummy, L. F., and Bouman, C. A. Multi-resolution data fusion for super-resolution electron microscopy. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 88–96, 2017.

Sun, Y., Wohlberg, B., and Kamilov, U. S. An online plug-and-play algorithm for regularized image reconstruction. *IEEE Transactions on Computational Imaging*, 2019a.

Sun, Y., Xu, S., Li, Y., Tian, L., Wohlberg, B., and Kamilov, U. S. Regularized fourier ptychography using an online plug-and-play algorithm. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7665–7669, 2019b.

Sutton, R., Mcallester, D., Singh, S., and Mansour, Y. Policy gradient methods for reinforcement learning with function approximation. *Advances in Neural Information Processing Systems (NIPS)*, 2000.

Tai, Y., Yang, J., Liu, X., and Xu, C. Memnet: A persistent memory network for image restoration. In *IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

Teodoro, A. M., Bioucas-Dias, J. M., and Figueiredo, M. A. Image restoration and reconstruction using variable splitting and class-adapted image priors. In *IEEE International Conference on Image Processing*, pp. 3518–3522, 2016.

Teodoro, A. M., Bioucas-Dias, J. M., and Figueiredo, M. A. A convergent image fusion algorithm using scene-adapted gaussian-mixture-based denoising. *IEEE Transactions on Image Processing*, 28(1):451–463, 2018.

Tirer, T. and Giryes, R. Image restoration by iterative denoising and backward projections. *IEEE Transactions on Image Processing*, 28(3):1220–1234, 2018.

Venkatakrishnan, S. V., Bouman, C. A., and Wohlberg, B. Plug-and-play priors for model based reconstruction. In *IEEE Global Conference on Signal and Information Processing*, pp. 945–948, 2013.

Wang, S., Fidler, S., and Urtasun, R. Proximal deep structured models. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 865–873, 2016.

Wang, X. and Chan, S. H. Parameter-free plug-and-play admm for image restoration. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1323–1327, 2017.

Xie, X., Wu, J., Liu, G., Zhong, Z., and Lin, Z. Differentiable linearized admm. In *International Conference on Machine Learning (ICML)*, pp. 6902–6911, 2019.

Yang, J., Zhang, Y., and Yin, W. A fast alternating direction method for tvl1-l2 signal reconstruction from partial fourier data. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):288–297, 2010.

Yang, Y., Sun, J., Li, H., and Xu, Z. Deep admm-net for compressive sensing mri. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 10–18. 2016.

Yu, K., Dong, C., Lin, L., and Change Loy, C. Crafting a toolchain for image restoration by deep reinforcement learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2443–2452, 2018.

Yu, K., Wang, X., Dong, C., Tang, X., and Loy, C. C. Path-restore: Learning network path selection for image restoration. *arXiv preprint arXiv:1904.10343*, 2019.

Zhang, J. and Ghanem, B. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017a.

Zhang, K., Zuo, W., Gu, S., and Zhang, L. Learning deep cnn denoiser prior for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017b.

Zhang, K., Zuo, W., and Zhang, L. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018.

Zhang, K., Zuo, W., and Zhang, L. Deep plug-and-play super-resolution for arbitrary blur kernels. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019a.

Zhang, X., Lu, Y., Liu, J., and Dong, B. Dynamically unfolding recurrent restorer: A moving endpoint control method for image restoration. In *International Conference on Learning Representations (ICLR)*, 2019b.

Zheng, G., Horstmeyer, R., and Yang, C. Wide-field, high-resolution fourier ptychographic microscopy. *Nature Photonics*, 7(9):739, 2013.

Zoran, D. and Weiss, Y. From learning models of natural image patches to whole image restoration. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 479–486, 2011.