

Dual-Domain Self-Supervised Learning and Model Adaption for Deep Compressive Imaging

Yuhui Quan^{1,2*}, Xinran Qin^{1**}, Tongyao Pang², and Hui Ji^{2 ***}

¹ School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

² Department of Mathematics, National University of Singapore, 119076, Singapore
csyhquan@scut.edu.cn, csqinxinran@mail.scut.edu.cn,
matpt@nus.edu.sg, matjh@nus.edu.sg

Abstract. Deep learning has been one promising tool for compressive imaging whose task is to reconstruct latent images from their compressive measurements. Aiming at addressing the limitations of supervised deep learning-based methods caused by their prerequisite on the ground truths of latent images, this paper proposes an unsupervised approach that trains a deep image reconstruction model using only a set of compressive measurements. The training is self-supervised in the domain of measurements and the domain of images, using a double-head noise-injected loss with a sign-flipping-based noise generator. In addition, the proposed scheme can also be used for efficiently adapting a trained model to a test sample for further improvement, with much less overhead than existing internal learning methods. Extensive experiments show that the proposed approach provides noticeable performance gain over existing unsupervised methods and competes well against the supervised ones.

Keywords: Self-supervised deep learning, Compressed sensing, Model adaption, Compressive imaging, Image reconstruction

1 Introduction

Compressed Sensing (CS) provides an acceleration technique for imaging, with a broad spectrum of applications in different domains, such as computed tomography (CT) and magnetic resonance imaging (MRI) in medicine, as well as energy-efficient cameras, holography and scanning microscopy in science. In general, CS captures a small number of linear measurements of an image and then reconstructs the image from these measurements. Let $\mathbf{x} \in \mathbb{R}^N$ denote the

* Yuhui Quan is also with Pazhou Lab, Guangzhou 510335, China. He would like to thank the support in part by National Natural Science Foundation of China under Grant 61872151 and in part by Natural Science Foundation of Guangdong Province under Grant 2022A1515011755.

** Corresponding author: Xinran Qin

*** Hui Ji would like to thank the support in part by Singapore MOE AcRF under Grant R-146-000-315-114.

image of interest, $\mathbf{y} \in \mathbb{C}^M$ the measurements with $M \ll N$, and $\Phi \in \mathbb{C}^{M \times N}$ the sensing matrix. Then, compressive imaging (*i.e.* CS-based imaging) solves

$$\mathbf{y} = \Phi \mathbf{x} + \epsilon, \quad (1)$$

where $\epsilon \in \mathbb{C}^M$ denotes the measurement noise (*e.g.* Gaussian white noise). How to suppress the propagation of noise to the solution and how to resolve the solution ambiguity caused by the under-determinedness of Φ are two main concerns.

Deep learning has emerged as a promising tool for compressive imaging; see *e.g.*, [37, 24, 17, 42, 39, 4, 31, 45, 44, 43, 11, 40]. These methods train a deep neural network (NN) in a supervised manner, *i.e.*, using an external dataset with many ground truth (GT) images and their measurements. The NN's performance depends on both the size of training data and the coherence between training and test data. Unfortunately, in many data-limited domains, *e.g.*, MRI and CT in medicine, it is often very difficult or costly to collect sufficient GT images. Then, the possible bias in a limited amount of training data can lead to poor generalization performance of a pre-trained model, *e.g.*, novel pathology not present in training data might disappear in the reconstructed images of test data.

Recently, there has been an increasing interest in developing deep learning methods for compressive imaging with relaxed requirements on training data. Nevertheless, existing works along this line have various issues in practice. The weakly-supervised learning method [14] takes unpaired images and measurements for training, which still requires the access to GT images. The unsupervised learning methods [21, 2, 3] avoid accessing GT images. However, their performance is not competitive with their supervised counterparts. The internal learning methods [25, 41, 28] learn an NN from the test sample itself by exploiting its internal statistics, whereas the millions of gradient updates on each test sample result in high computational cost, especially when processing many samples. There are also some unsupervised methods working on specific settings of sensing matrices, *e.g.*, paired ones [35] or varying ones [21, 6]. The resulting specific sensing setups limit the wider adoption of these methods.

This paper is devoted to developing an self-supervised deep learning approach for compressive imaging, which enjoys competitive performance against existing supervised methods, while its training only requires measurements collected in a general sensing setup (*i.e.*, using a single fixed sensing matrix). Such an approach is applicable to data-limited environments and makes the sensing more flexible in practice than those using paired or varying sensing matrices.

Recall that the linear system in (1) is under-determined. It has a non-zero kernel (also called null space) defined by $\ker(\Phi) = \{\mathbf{x} | \Phi \mathbf{x} = \mathbf{0}\}$. Therefore, the measurement vector \mathbf{y} only provides the noisy information of image \mathbf{x} within $\text{im}(\Phi^H) = \ker(\Phi)^\perp$ where $\text{im}(\Phi^H)$ denotes the image (also called column space) of the adjoint operator of Φ , *i.e.*, its Hermitian transpose Φ^H . The remaining information of \mathbf{x} in $\ker(\Phi)$, is not available in \mathbf{y} . Then, two critical parts of an unsupervised approach are: (*a*) Training the NN to predict the image in $\text{im}(\Phi^H)$ when only noisy measurements are available; and (*b*) Training the NN to predict the image in $\ker(\Phi)$ when no information is available. Note that while the latter

one can be partially addressed if the training dataset contains measurements from varied Φ , it is not assumed in our setup as mentioned before.

In this work, those two parts are implemented by a self-supervised training scheme. Regarding the first part, as the measurement vector \mathbf{y} encodes the image \mathbf{x} by $\Phi\mathbf{x}$, we use \mathbf{y} as the “labels” for self-supervision and train the NN to predict an image \mathbf{x}' from \mathbf{y} such that $\Phi\mathbf{x}' \approx \mathbf{y}$ under some loss defined in the domain of measurements. Since \mathbf{y} can be noisy, the key is then how to make the training robust to the measurement noise, which is treated by extending the self-supervised loss introduced in the self-supervised Gaussian denoiser [26] to the measurement domain. Built on a double-head symmetric noise injection, where the injected noise is drawn from the same distribution as measurement noise, the proposed loss is equivalent, in terms of expectation, to the loss with noise-free labels $\Phi\mathbf{x}$ in the measurement domain. Such a loss enables us to train the NN to predict the GT in $\text{im}(\Phi^H)$, in analogy with supervised learning. We also show that the loss has a close connection to Stein’s unbiased risk estimator (SURE) [10, 21] in the presence of Gaussian noise.

Regarding the second part, the training scheme above does not address the solution ambiguity caused by $\ker(\Phi)$, as $\Phi(\mathbf{x} + \mathbf{n}) = \Phi(\mathbf{x}) \approx \mathbf{y}, \forall \mathbf{n} \in \ker(\Phi)$. While no information of \mathbf{x} in $\ker(\Phi)$ exists in \mathbf{y} , the prediction from an NN with specific structures, *e.g.*, a convolutional NN (CNN) or an unfolding-based NN, is an estimate biased to smooth and regular structures; see *e.g.* [7, 32]. Thus, one can utilize the intermediate prediction from an NN as another noisy observation of “labels” (GT images) for the self-supervision in $\ker(\Phi)$. The training in $\ker(\Phi)$ is implemented by defining a similar double-head symmetric noise-injected loss in image domain, which further refines the estimate from the NN regarding the information in both $\ker(\Phi)$ and $\text{im}(\Phi)$. Note that the “noise” in this case refers to the residual of the intermediate GT estimate, whose statistical distribution is complex and unknown. We introduce random sign flipping to simulate the samples drawn from the distribution of the residuals.

The above ideas lead to a dual-domain loss for self-supervised learning of compressive imaging, which brings noticeable performance gain over existing unsupervised methods. Motivated by the benefits of internal learning, we also apply the proposed loss to adapting a trained model to each test sample, which alleviates possible bias of training data and possible inconsistency of sensing matrices between training and test data. Such a model adaption scheme brings further performance gain. See below for the summary of our main contributions.

- A dual-domain self-supervised loss is proposed for handling possible ambiguity and overfitting when only noisy measurements are available in training.
- A model adaption scheme is proposed to exploit specific internal characteristics of a test sample for performance improvement. It also provides much higher computational efficiency than existing internal learning methods.
- A self-supervised learning approach for compressive imaging is introduced with a relaxed requirement on training data. Its performance is noticeably better than existing unsupervised methods and is competitive against state-of-the-art supervised methods.

2 Related Work

Image priors play an important role in compressive imaging for regularizing the solution of (1). Traditional methods use handcrafted image priors, such as transform-induced sparsity [18], non-local low-rankness [9], and the denoising prior from some manually-designed denoiser [22]. These methods require solving an optimization problem with some time-consuming iterative scheme.

Deep learning-based methods encode image priors into a measurements-to-image reconstructive NN. For supervised learning, the effort is mainly on the design of NN architectures; see *e.g.* [38, 29, 30, 45, 1, 8]. Among them, physics-aware NNs are the most prominent ones, which are often designed by an unfolding strategy that replaces the prior-related operations in a traditional iterative method by some NN blocks; see *e.g.* [24, 19, 39, 42, 44, 43, 40]. Dual-domain architectures are also superior in supervised methods; see *e.g.* [31, 45, 44]. In comparison to these works, the concept of dual domain used in the proposed approach is mainly for unsupervised training, rather than NN structures.

The prerequisite on many paired samples is one limitation of supervised methods. Plug-and-play methods (*e.g.* [34, 15, 12]) partially address this limitation by calling pre-trained models of denoising or generative NNs. These methods still require GT images in the target domain for model pre-training. When the target image is not in the domain of training data, the performance of these methods is likely to see a big drop. In the unpaired learning method [14], the prerequisite of GT images is still required.

To remove the prerequisite on GT images, there have been a few studies on unsupervised deep learning for compressive imaging. Xia *et al.* [35] used the measurement samples collected by two sensors. Cloe *et al.* [6] trained a generative adversarial NN using unpaired measurement samples collected by different sensing matrices. Both Metzler *et al.* [21] and Zhussip *et al.* [46] proposed to train a learned denoiser-approximate message passing NN with a SURE-based denoiser. The former addressed the ambiguity caused by the null space via using varied sensing matrices for collecting measurement data, while the latter mainly relies on the regularization provided by the SURE-based denoiser to resolve the ambiguity. Chen *et al.* [2, 3] proposed to exploit the equivariance present in latent images so that the missing null-space information of one sample can be supervised by the reconstructed range-space information of another sample. By assuming the equivariance, they showed that the unsupervised training with equivariant transforms is an unbiased estimator to the GT. In comparison, the proposed dual-domain training scheme uses the NN’s prediction to refine the learning in both spaces and leads to further improvement.

Instead of leveraging a dataset, internal learning exploits the implicit image priors encoded by an untrained CNN which enables one to exploit sample-specific statistics for good performance. It is GT-free as it only takes test samples for learning. Inspired by the dropout-based unsupervised denoiser [27], Pang *et al.* [25] utilized the model uncertainty of a Bayesian CNN to refine the result of internal learning. For accelerated MRI, Zalbagi *et al.* [41] used an under-parameterized CNN with few and simplified convolutional layers for addressing

the overfitting in internal learning. Wang *et al.* [33] proposed to run Monte Carlo sampling on the NN’s weights to approximate the Bayesian estimator of latent images. These methods suffer from high computational costs due to sample-wise model learning. In comparison, the proposed model adaption scheme exploits internal statistics of test data with much higher computational efficiency.

3 Dual-Domain Self-Supervised Training

Given a full row-rank matrix Φ , we have

$$\mathbb{C}^N = \text{im}(\Phi^H) \oplus \ker(\Phi). \quad (2)$$

Recall that there are two issues to address in unsupervised training: measurement noise in $\text{im}(\Phi^H)$ and solution ambiguity in $\ker(\Phi)$. Let $f_\Phi(\cdot; \omega) : \mathbb{C}^M \rightarrow \mathbb{R}^N$ denote a deep NN model parameterized by ω which predicts an image from input measurements with the sensing matrix Φ . We will omit the parameters ω or the subscript Φ if not causing notational confusion.

In supervised learning, by minimizing the error between the output $f(\mathbf{y})$ and GT \mathbf{x} , one can train f to simultaneously minimize the prediction errors in both $\text{im}(\Phi^H)$ and $\ker(\Phi)$. In unsupervised learning, as only \mathbf{y} is available, we only have the noisy information of \mathbf{x} in $\text{im}(\Phi^H)$. Then, the key is how to train f for prediction with small prediction errors in both $\text{im}(\Phi^H)$ and $\ker(\Phi)$. To address this, we propose a dual-domain self-supervised loss function $\mathcal{L}^{\text{Dual}}$:

$$\mathcal{L}^{\text{Dual}}(\omega) := \mathcal{L}^{\text{Measure}}(\omega) + \lambda \mathcal{L}^{\text{Image}}(\omega), \quad (3)$$

with a pre-defined hyper-parameter λ . The first part $\mathcal{L}^{\text{Measure}}$ concerns the prediction error in $\text{im}(\Phi^H)$. The second part $\mathcal{L}^{\text{Image}}$ measures the prediction error in image domain \mathbb{C}^N that covers the error in $\ker(\Phi)$.

3.1 Loss Function

For the prediction in $\text{im}(\Phi^H)$, we use the noisy measurement \mathbf{y} to train the NN. Motivated by [26], we run data augmentation on \mathbf{y} to eliminate the noise effect. Let $\mathbf{y}' = \mathbf{y} + \gamma$ denote an augmented version of \mathbf{y} with random noise γ . Consider

$$\Phi f_\Phi(\mathbf{y}') \rightarrow \mathbf{y} \approx \Phi \mathbf{x} + \epsilon. \quad (4)$$

Once the measurement noise ϵ can be handled by $\mathcal{L}^{\text{Measure}}$, f can learn accurate predictions in $\text{im}(\Phi^H)$, as if using the noise-free measurements $\Phi \mathbf{x}$.

Since \mathbf{y} carries no information about \mathbf{x} in $\ker(\Phi)$, we use the output of f , denoted by \mathbf{z} , as a noisy version of the GT image for supervising the training. As the structure of a CNN itself imposes certain implicit prior on its prediction [7, 32], \mathbf{z} can be viewed as a regularized solution that contains some information of \mathbf{x} in $\ker(\Phi)$. Similarly, consider an augmented version of \mathbf{z} : $\mathbf{z}' = \mathbf{z} + \mathbf{r}$ with random noise \mathbf{r} and the following scheme:

$$f_\Phi(\Phi \mathbf{z}') \rightarrow \mathbf{z} = \mathbf{x} + \mathbf{e}. \quad (5)$$

If the residual \mathbf{e} can be effectively handled by the loss $\mathcal{L}^{\text{Image}}$, then the ambiguity in null-space prediction can be effectively alleviated during learning. Note that while f is expected to learn accurate prediction in $\text{im}(\Phi^H)$ by the loss $\mathcal{L}^{\text{Measure}}$, the training with $\mathcal{L}^{\text{Image}}$ also provides certain refinement on its prediction in $\ker(\Phi)$ due to the implicit regularization of the NN structure.

To effectively handle the noise in \mathbf{y} and the error in \mathbf{z} , we define

$$\begin{aligned}\mathcal{L}^{\text{Measure}}(\boldsymbol{\omega}) &:= \mathbb{E}_{\mathbf{y} \sim \mathbb{Y}, \gamma} \|\Phi f_{\Phi}(\mathbf{y} + \gamma; \boldsymbol{\omega}) - \mathbf{y} + \gamma\|_2^2, \\ \mathcal{L}^{\text{Image}}(\boldsymbol{\omega}) &:= \mathbb{E}_{\mathbf{y} \sim \mathbb{Y}, \mathbf{r}} \|f_{\Phi}(\Phi(\mathbf{z} + \mathbf{r}); \boldsymbol{\omega}) - \mathbf{z} + \mathbf{r}\|_2^2,\end{aligned}\quad (6)$$

where $\mathbf{z} := f(\mathbf{y}; \boldsymbol{\omega}_0)$ ($\boldsymbol{\omega}_0$ is the NN parameters obtained in the previous epoch), \mathbb{Y} is the training dataset of measurement samples, and both γ and \mathbf{r} are injected random noises drawn from some distributions. The resulting self-supervised loss $\mathcal{L}^{\text{Dual}}$ is connected to its supervised counterpart as shown in Proposition 1 .

Proposition 1. Consider $\mathbf{y} = \Phi \mathbf{x} + \epsilon$. Let P_1 and P_2 denote the conditional distribution of ϵ and the residual $\mathbf{e} := f_{\Phi}(\mathbf{y}; \boldsymbol{\omega}_0) - \mathbf{x}$ on \mathbf{x} respectively, where $\boldsymbol{\omega}_0$ denotes the NN parameters obtained in the previous step, i.e., $\epsilon | \mathbf{x} \sim P_1$ and $\mathbf{e} | \mathbf{x} \sim P_2$. Assume the random noise γ (resp. \mathbf{r}) is drawn from P_1 (resp. P_2) and independent from ϵ (resp. \mathbf{e}) conditioned on \mathbf{x} . Then, the loss function $\mathcal{L}^{\text{Dual}}$ defined by (3) and (6) satisfies

$$\mathcal{L}^{\text{Dual}}(\boldsymbol{\omega}) = \mathbb{E}_{\mathbf{x}, \epsilon, \gamma} \|\Phi f_{\Phi}(\mathbf{y} + \gamma; \boldsymbol{\omega}) - \Phi \mathbf{x}\|_2^2 + \lambda \mathbb{E}_{\mathbf{x}, \epsilon, \mathbf{r}} \|f_{\Phi}(\Phi(\mathbf{z} + \mathbf{r}); \boldsymbol{\omega}) - \mathbf{x}\|_2^2 + C,$$

where $\mathbf{z} = f_{\Phi}(\mathbf{y}; \boldsymbol{\omega}_0)$ and C is a constant.

Proof. See supplementary materials for the proof. \square

Proposition 1 states both $\mathcal{L}^{\text{Measure}}$ and $\mathcal{L}^{\text{Image}}$ are the unbiased estimates of their supervised counterparts under certain conditions. Therefore, the proposed self-supervised loss enables us to train with only noisy measurements, provided that we can draw samples from P_1 and P_2 . Sampling from P_1 is easy when the distribution of measurement noise is known, e.g., Gaussian white noise. Sampling from P_2 is hard as its distribution is in general complex and unknown.

Generation of injected noise \mathbf{r} To sample from P_2 , one simple way is to approximate it using normal distribution $\mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I})$. Such a treatment certainly is sub-optimal and how to set the noise level σ_e is non-trivial either. We develop a sign-flipping-based scheme to generate the samples which empirically approximate the samples from P_2 better.

Let $\mathbf{z}' = f_{\Phi}(\Phi \mathbf{z} + \gamma)$ denote an intermediate prediction during training which is detached from back propagation. Suppose \mathbf{z}' is a good estimate of \mathbf{x} . Then the residual $\mathbf{e}' = \mathbf{z} - \mathbf{z}'$ can be viewed as an approximation to the residual \mathbf{e} . Using \mathbf{e}' as the injected noise is sub-optimal, as it remains correlated to \mathbf{e} . To reduce the correlation, we apply random sign-flipping to \mathbf{e}' so as to generate new samples $\{\mathbf{r}\}$, i.e., \mathbf{r} is generated by $\mathbf{r} = \mathbf{e}' \odot \mathbf{s}$, where \odot denotes element-wise product and $\mathbf{s}(i)$ takes values from $\{1, -1\}$ with probability 0.5 for all i . It is easy to show that $\mathbb{E}(\mathbf{r}^\top \mathbf{e}) = 0$ (see supplementary materials). Although the zero

covariance between \mathbf{r} and \mathbf{e} does not guarantee that they are independent, the correlation between them is likely to be reduced. Furthermore, if the distribution of \mathbf{e}' is symmetric w.r.t. the origin (which is observed empirically; see Fig. 1), flipping sign does not destroy the statistics, *i.e.*, \mathbf{r} follows the same distribution as \mathbf{e}' . In summary, the noises generated by the random sign-flipping simulate the samples from P_2 better with weak correlation to \mathbf{z} .

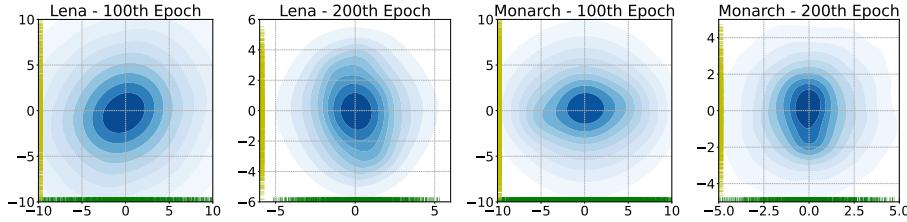


Fig. 1. Distribution of the residual $\mathbf{e} = \mathbf{z} - \mathbf{x}$ at different epochs. The results are computed in the configuration of Gaussian sensing matrices with $\epsilon \sim \mathcal{N}(\mathbf{0}, \frac{10}{255}\mathbf{I})$, using our NN trained with $\mathcal{L}^{\text{Measure}}$. Kernel density estimator is applied to the t-SNE projections of 1000 instances to show the distribution. It can be seen that \mathbf{e} are roughly symmetrically distributed around the origin.

Relation to SURE The proposed self-supervised loss is also closely related to the SURE-based loss. When applied to the measurements contaminated by Gaussian noise $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, the SURE loss takes the form:

$$\mathcal{L}^{\text{SURE}}(\boldsymbol{\omega}) := \|\Phi f_{\Phi}(\mathbf{y}; \boldsymbol{\omega}) - \mathbf{y}\|_2^2 + 2\sigma^2 \text{tr}\left(\Phi \frac{\partial f_{\Phi}(\mathbf{y}; \boldsymbol{\omega})}{\partial \mathbf{y}}\right). \quad (7)$$

By injecting noise into the NN input, we modify the above SURE loss to be

$$\mathcal{L}^{\text{SURE+}}(\boldsymbol{\omega}) := \|\Phi f_{\Phi}(\mathbf{y} + \epsilon'; \boldsymbol{\omega}) - \mathbf{y}\|_2^2 + 2\sigma^2 \text{tr}\left(\Phi \frac{\partial f_{\Phi}(\mathbf{y} + \epsilon'; \boldsymbol{\omega})}{\partial \mathbf{y}}\right), \quad (8)$$

where $\epsilon' \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ is i.i.d. to ϵ . Then see below for the connection to SURE+.

Proposition 2. Let $\mathbf{y} = \Phi \mathbf{x} + \epsilon \in \mathbb{R}^M$. Assume $\epsilon, \epsilon' \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ are independent from each other and \mathbf{x} . Then,

$$\mathbb{E}_{\mathbf{y}, \epsilon'} \mathcal{L}^{\text{SURE+}} = \mathbb{E}_{\mathbf{y}, \epsilon} \mathcal{L}^{\text{Measure}} - M\sigma^2. \quad (9)$$

Proof. See supplementary materials for the proof. \square

Proposition 2 shows that, for Gaussian noise, training with the loss $\mathcal{L}^{\text{Measure}}$ can be roughly viewed as a data-augmented version of SURE applied to \mathbf{y} . Different from SURE, no regularization term on partial derivatives are introduced in $\mathcal{L}^{\text{Measure}}$, which results in more efficient computation. Together with $\mathcal{L}^{\text{Image}}$, the proposed loss $\mathcal{L}^{\text{Dual}}$ noticeably outperformed SURE-based methods.

3.2 NN Architecture

Motivated by the studies in [7] which show that the unfolding of the proximal gradient descent algorithm (PGDA) has a certain regularization effect on the NN's prediction during unsupervised learning, we adopt a similar NN that unfolds the PGDA for the problem: $\min_{\mathbf{x}} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \psi(\mathbf{x})$, with the regularization function $\psi : \mathbb{R}^N \rightarrow \mathbb{R}$. The iterative scheme of PGDA reads

$$\mathbf{x}^{(k)} = \text{prox}_{\psi}(\mathbf{x}^{(k-1)} - \rho \Phi^H (\Phi \mathbf{x}^{(k-1)} - \mathbf{y})), \quad (10)$$

where prox_{ψ} is the proximal operator and ρ is a learnable step size. The NN is defined by replacing $\text{prox}_{\psi}^{\rho}$ with a convolutional block. It contains K phases, each of which mimics one iteration in (10). See Fig. 2 for the detailed structure.

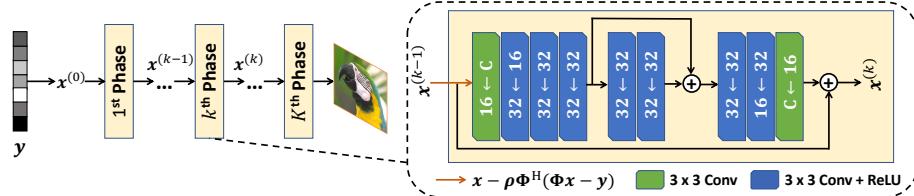


Fig. 2. NN architecture used in the proposed approach. The estimate $\mathbf{x}^{(0)}$ is initialized as the least-squares solution during training and initialized as the prediction from the pre-trained model during test-time adaption.

4 Inference with Test-Time Model Adaption

The proposed loss $\mathcal{L}^{\text{Dual}}$ not only can be used for unsupervised training, but also can be applied for adapting a deep model trained over an external dataset to unseen data. One such extension is adapting the model to a test sample under a specific sensing matrix (seen or unseen), so as to exploit sample-specific statistics for performance boost. Let $\widehat{\omega}$ denote the NN parameters of a trained model. For a test sample (\mathbf{y}^*, Φ^*) , we first initialize $\omega^* = \widehat{\omega}$ and then update it by the gradient of (3) on \mathbf{y}^* :

$$\omega^* := \omega^* - \tau \nabla_{\omega^*} \mathcal{L}_{\mathbf{y}^*}^{\text{Dual}}(\omega^*), \quad (11)$$

with T steps. Once done, the final prediction is given by

$$\mathbf{x}^* = f_{\Phi^*}(\mathbf{y}^*; \omega^*), \quad (12)$$

with the adapted NN parameters ω^* . See an algorithmic description of the whole adaption process in supplementary materials.

5 Experiments

The proposed approach is evaluated on three tasks compressive imaging in CT, MRI, and natural image reconstruction (NIR) respectively. The configuration is as follows: *(a)* NN: $K = 12$ for all tasks; *(b)* Initialization: Kaiming initialization is used for all learnable weights, except that the step size ρ and the bias terms are initialized by 0.5 and 0 respectively. *(c)* Training: We set $\lambda = 0.25$ in $\mathcal{L}^{\text{Dual}}$ and call the Adam optimizer with 500 epochs and with learning rate fixed at 10^{-4} ; and *(d)* Adaption: The Adam optimizer is used with $T = 200$ epochs and with fixed learning rate 10^{-4} . The performance of the proposed approach is evaluated with and without using test-time adaption respectively, which are denoted by Ours-TA and Ours-NA correspondingly. For the comparison with existing methods, we quote their results directly from the literature whenever possible; or we use their reproducible codes with efforts on hyper-parameter tuning-up to obtain the results; otherwise we leave them blank in tables.

5.1 Sparse-View CT from Randon Measurements

Following [2], the CT measurements are taken by the discrete Radon transform based sensing matrix without noise. The 100 real in-vivo CT images of size 128×128 from the CT100 dataset [5], collected from the cancer imaging archive which consists of the middle slice of CT images taken from 69 patients, are used for the experiment: 90 for training and 10 for test. The methods for comparison include *(a)* two supervised methods: FBP [13] and FISTANet [36]; *(b)* an internal learning method: BCNN [25]; and *(c)* a very recent state-of-the-art unsupervised method: EI [2]. Since EI is an unsupervised method whose focus is on the training scheme design, for a fair comparison, we replace its NN with ours and retrain it with the same data as ours throughout the experiments, which brings performance improvement to the original EI.

The quantitative comparison is presented in Table 1. Ours-TA is the best performer among all methods including both the internal learning and unsupervised ones. Its performance is close to that of the supervised-trained FISTANet. Even without adaption, Ours-NA still performs noticeably better than other GT-free methods. Such results have indicated that the proposed approach is very effective for unsupervised compressive imaging. See also Fig. 3 for a demonstration.

Table 1. Mean PSNR (dB) and SSIM values in sparse-view CT reconstruction. The best results among GT-dependent methods, GT-free methods, and all methods are marked in green, marked in blue, and underlined respectively.

Metric	FBP	FISTANet	BCNN	EI	Ours-NA	Ours-TA
PSNR	30.24	41.85	39.82	40.48	41.35	<u>41.99</u>
SSIM	-	0.983	0.970	0.968	0.979	<u>0.985</u>

5.2 MRI Reconstruction from k -Space Measurements

Following [17, 25], the k -space measurements in MRI are taken by $\Phi : \mathbf{x} \rightarrow \beta \odot (\mathbf{F}\mathbf{x})$, where β is a fixed binary mask for downsampling and \mathbf{F} denotes the discrete Fourier transform. The measurement noise is generated by $\epsilon = \Phi(\epsilon_1 + i\epsilon_2)$, where $\epsilon_1, \epsilon_2 \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. The 2D Gaussian masks and radial masks of different CS ratios are used for β . The images from the Alzheimer’s Disease Neuroimaging Initiative are used: 279 for training and 21 for test. Both the noiseless setting ($\sigma = 0$) and noisy setting ($\sigma = 0.1 \max(\mathbf{x})$) are considered. We train an individual model for each sensing matrix and each noise setting. In each setting, each image is called only once for measurement generation. Then we have 279 measurement samples in total for training. In addition to the BCNN and EI, the methods for comparison include (a) a classic non-learning method: SparseMRI [18]; (b) five supervised methods: GAN [17], ADMMNet [39], DDN [1], CDDN [44] MAC-Net [11]; and (c) a very recent internal learning method: ConvDec [41].

Table 2. Mean values of PSNR(dB) (even rows) and SSIM (odd rows) in MRI reconstruction. The best results among GT-dependent methods, GT-free methods, and all methods are marked in green, marked in blue, and underlined respectively.

Mask		Radial						2D Gaussian					
		0			0.1*MaxValue			0			0.1*MaxValue		
GT-Dependent	Noise Level	1/3	1/4	1/5	1/3	1/4	1/5	1/3	1/4	1/5	1/3	1/4	1/5
	GAN	34.49 0.94	32.29 0.90	30.10 0.84	26.72 0.75	25.55 0.74	25.02 0.73	35.12 0.94	32.96 0.91	31.79 0.89	26.49 0.76	26.31 0.75	25.79 0.75
	ADMMNet	35.31 0.94	33.70 0.93	32.32 0.92	26.50 0.60	25.97 0.61	25.44 0.59	36.34 0.95	34.95 0.94	33.82 0.93	26.17 0.56	25.82 0.60	25.40 0.61
	DDN	33.58 0.92	32.76 0.88	31.79 0.89	31.48 0.85	30.83 0.85	29.69 0.82	34.82 0.91	32.73 0.89	31.16 0.87	31.10 0.85	30.89 0.84	29.87 0.84
	CDDN	36.58 0.97	34.70 0.95	33.46 0.95	30.71 0.85	30.79 0.86	30.28 0.86	37.93 0.97	36.01 0.96	34.76 0.96	30.90 0.84	30.93 0.85	30.74 0.86
	MACNet	<u>36.70</u> <u>0.98</u>	<u>34.76</u> <u>0.97</u>	<u>33.42</u> <u>0.95</u>	<u>30.98</u> <u>0.87</u>	<u>30.97</u> <u>0.87</u>	<u>30.25</u> <u>0.86</u>	<u>38.22</u> <u>0.98</u>	<u>35.84</u> <u>0.96</u>	<u>34.81</u> <u>0.96</u>	<u>31.04</u> <u>0.86</u>	<u>30.92</u> <u>0.86</u>	<u>30.81</u> <u>0.86</u>
	SparseMRI	34.58 0.94	32.31 0.90	30.72 0.86	25.32 0.49	25.13 0.49	24.68 0.49	34.93 0.93	32.79 0.90	31.69 0.89	24.91 0.47	24.92 0.49	24.97 0.51
	ConvDec	35.02 0.94	33.99 0.94	32.14 0.91	29.00 0.85	28.27 0.83	27.48 0.71	36.77 0.94	35.82 0.96	32.92 0.91	28.94 0.85	28.66 0.84	28.24 0.84
	BCNN	35.58 0.94	34.08 <u>0.95</u>	32.28 0.92	29.58 0.85	29.47 0.87	28.38 0.84	37.60 0.94	36.10 0.95	33.81 0.93	29.46 0.86	29.20 0.85	29.17 0.85
GT-Free	EI	36.16 0.94	33.49 0.92	32.35 0.90	30.22 0.81	29.56 0.75	29.77 0.76	37.03 0.93	36.11 0.95	34.01 0.92	29.00 0.72	29.61 0.72	29.82 0.73
	Ours-NA	36.25 0.94	34.32 0.94	33.47 0.93	30.66 0.85	30.62 0.86	30.16 0.85	37.42 0.94	36.20 <u>0.96</u>	34.80 0.93	30.87 0.86	30.65 0.86	30.29 0.85
	Ours-TA	<u>36.43</u> <u>0.95</u>	<u>34.70</u> <u>0.95</u>	<u>33.97</u> <u>0.95</u>	<u>31.68</u> <u>0.86</u>	<u>31.04</u> <u>0.88</u>	<u>30.44</u> <u>0.87</u>	<u>37.82</u> <u>0.95</u>	<u>36.72</u> <u>0.96</u>	<u>35.46</u> <u>0.95</u>	<u>31.62</u> <u>0.87</u>	<u>31.30</u> <u>0.86</u>	<u>30.89</u> <u>0.86</u>

See Table 2 for quantitative comparison. Ours-TA outperforms other GT-free methods in all settings. By exploiting both the external knowledge from training data and the internal statistics of a test sample, Ours-TA even outperforms the supervised methods in many settings. Without model adaption, Ours-NA still performs better than other GT-free methods overall, and competes with the supervised methods. See Fig. 3 for the visual comparison of some reconstruction

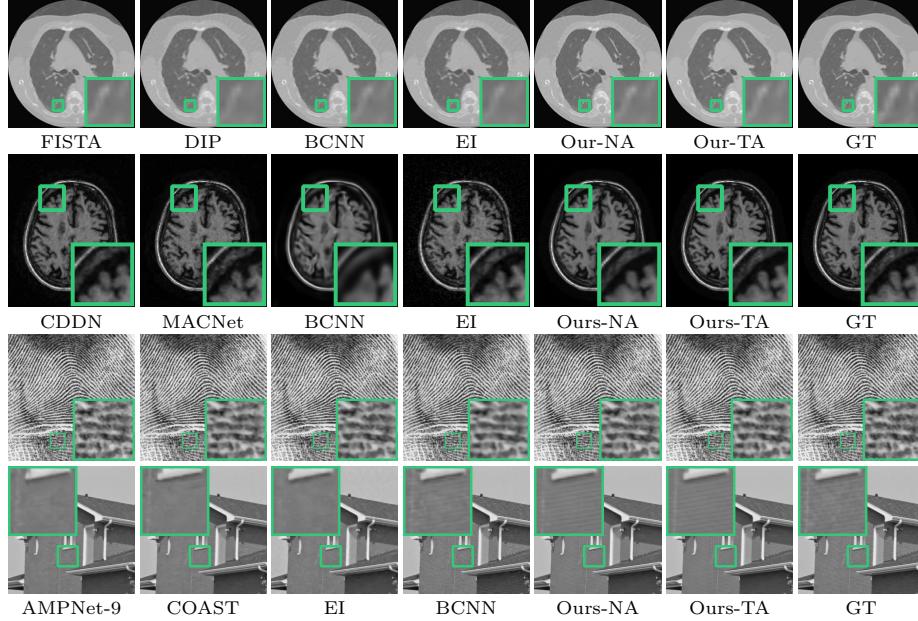


Fig. 3. Visual results. The first row: CT reconstruction. The second row: MRI reconstruction from noisy measurements with radial mask of CS ratio 25% The last two rows: NIR from noiseless Gaussian measurements of CS ratio 40% and 25%.

results, where our result is closer to the GT in comparison with other GT-free methods. Both quantitative and qualitative results have clearly verified the effectiveness of the proposed approach.

5.3 NIR from Blockwise Gaussian Measurements

Following [16, 42], the measurements are taken by a blockwise row-orthogonalized Gaussian matrix $\Phi \in \mathbb{R}^{M \times N}$ with different CS ratios defined by M/N . Both the noiseless setting with $\epsilon = \mathbf{0}$ and the noisy setting with $\epsilon \sim \mathcal{N}(\mathbf{0}, \frac{10}{255}\mathbf{I})$ are considered. Same as the previous experiment, we train an individual model for each sensing matrix and each noise setting. The 88912 image blocks of size 33×33 (*i.e.* $N = 1089$) provided by [16] are used to generate 88912 measurement samples for the training. Two datasets including Set11 [42] and BSD68 [20] are used for test, each image of which is cropped into non-overlapping blocks of size 33×33 for measurement generation. In addition to DDN, BCNN, EI (its robust version REI [3] is used for noisy cases), we also compare with (*a*) a non-learning method: DAMP [23]; (*b*) five supervised methods: ISTANet+ [42], DPANet [31], MACNet [4], AMPNet-9 [43], COAST [40]; (*c*) a plug-and-play method: SSLIP [15]; and (*d*) a recent unsupervised method: LDAMP-SURE [21].

The quantitative comparison is given in Table 3. Ours-TA is the best performer among the GT-free methods through all settings. Again, it is very com-

Table 3. Mean values of PSNR(dB) (even rows) and SSIM (odd rows) in NIR. The best results among GT-dependent methods, GT-free methods, and all methods are marked in green, marked in blue, and underlined respectively.

Dataset		Set11						BSD68					
Noise Level		0			10			0			10		
CS Ratio		0.40	0.25	0.10	0.40	0.25	0.10	0.40	0.25	0.10	0.40	0.25	0.10
GT-Dependent	ISTANet+	36.02 <u>0.96</u>	32.44 0.92	26.49 0.80	31.09 <u>0.89</u>	29.20 <u>0.86</u>	24.55 0.70	32.17 <u>0.92</u>	29.29 0.85	25.29 0.70	28.98 <u>0.83</u>	27.26 <u>0.77</u>	23.86 0.60
	DDN	27.38 0.80	26.18 0.78	23.30 0.70	25.64 0.73	24.39 0.69	17.22 0.35	28.84 0.85	26.42 0.78	23.44 0.64	21.98 0.49	22.99 0.52	20.61 0.46
	DPANet	35.04 0.95	31.74 0.92	26.99 <u>0.84</u>	30.17 <u>0.89</u>	29.31 <u>0.86</u>	25.13 0.75	31.43 0.91	29.21 0.83	25.88 0.61	28.95 0.81	27.25 0.76	<u>24.39</u> 0.63
	AMPNet-9	35.75 <u>0.96</u>	32.08 0.92	25.95 0.79	29.00 0.77	28.06 0.78	24.53 0.71	32.41 <u>0.92</u>	29.38 0.85	25.33 <u>0.70</u>	27.50 0.74	26.64 0.71	24.03 0.62
	COAST	36.94 <u>0.96</u>	33.85 0.93	28.34 0.80	31.16 <u>0.89</u>	29.37 0.86	25.71 0.78	33.02 <u>0.92</u>	30.07 <u>0.87</u>	26.25 0.70	29.15 0.83	27.43 0.77	24.02 0.60
	SSLIP	33.73 0.93	30.42 0.89	25.02 0.75	30.58 <u>0.89</u>	28.71 0.85	24.48 0.73	30.72 0.88	28.26 0.81	24.72 0.66	28.47 <u>0.83</u>	26.91 0.75	24.25 <u>0.67</u>
	DAMP	33.51 0.93	28.31 0.85	21.18 0.60	29.19 0.86	26.34 0.80	20.79 0.58	28.06 0.79	25.54 0.70	21.93 0.52	26.54 0.72	24.83 0.65	21.72 <u>0.51</u>
GT-Free	LDAMP-SURE	33.36 0.95	31.37 0.90	25.12 0.65	28.79 0.81	28.16 0.82	23.39 0.64	31.83 0.90	28.77 0.84	23.17 0.65	27.82 0.77	26.81 0.73	23.65 0.60
	BCNN	35.71 0.95	32.30 0.92	27.49 0.83	30.39 0.88	28.67 0.84	25.23 0.76	31.28 0.90	28.63 0.84	25.24 0.71	28.13 0.81	26.47 0.75	23.79 0.64
	EI/REI	35.63 0.95	31.11 0.90	22.79 0.64	28.86 0.76	28.10 0.76	22.25 0.60	31.79 0.90	28.45 0.82	23.11 0.63	28.07 0.68	27.02 0.72	22.34 0.59
	Ours-NA	36.37 0.95	32.70 0.93	26.89 0.82	31.42 0.90	29.17 0.86	25.41 0.78	32.17 0.91	28.97 0.86	25.61 0.70	28.24 0.81	27.23 0.77	24.10 0.66
	Ours-TA	<u>37.18</u> <u>0.96</u>	<u>33.41</u> <u>0.94</u>	<u>27.57</u> <u>0.84</u>	<u>31.94</u> <u>0.91</u>	<u>29.84</u> <u>0.87</u>	<u>26.06</u> <u>0.79</u>	<u>32.63</u> <u>0.92</u>	<u>29.66</u> <u>0.87</u>	<u>26.15</u> <u>0.72</u>	<u>29.32</u> <u>0.84</u>	<u>27.61</u> <u>0.78</u>	<u>24.63</u> <u>0.68</u>

petitive with the supervised methods and even outperformed them in nearly half settings. This is attributed to both the effectiveness of the unsupervised learning as well as the benefit from the model adaption. Without adaption, Ours-NA still performs competitively to supervised methods and is better than other GT-free methods. Particularly, with the dual-domain learning, it outperforms LDAMP-SURE significantly. We can observe a bigger improvement brought by model adaption in NIR than that in MRI. This is mainly because natural images exhibit larger variations between training and test samples, in comparison to the medical images in MRI. See also Fig. 3 for some visual comparison, where Ours-TA produces more details than other GT-free methods. To conclude, the results have again demonstrated the effectiveness of the proposed approach.

5.4 Additional Analysis on Test-Time Model Adaption

The benefit of the proposed model adaption scheme has been demonstrated by the result comparison between Our-TA and Ours-NA in Table 1,2,3. Those experiments use the same sensing matrix through training and test. We further examine the performance gain for the case when sensing matrices differ between training and test in MRI reconstruction. See Table 4 for the results. When dealing with Gaussian masks, the models trained on radial masks in Table 4 do not

perform as well as the models trained on Gaussian masks in Table 2. However, their performance is noticeably improved after the model adaption.

Table 4. Mean PSNR(dB) values of MRI reconstructions in cross-mask adaption test.

Train	Test	Gaussian, Ratio = 20%		Gaussian, Ratio = 25%	
		w/o adaption	w/ adaption	w/o adaption	w/ adaption
Radial, Ratio=25%	34.45/0.93	35.31/0.95	35.18/0.95	35.86/0.96	
Gaussian, Ratio=25%	34.80/0.94	35.46/0.95	36.39/0.96	36.72/0.96	

Table 5. Running time (in seconds) of different methods for reconstructing all images in BSD68. All the tests are conducted on an RTX 3090Ti GPU.

CS Ratio	ISTANet+	DDNet	DPANet	COAST	LDAMP-SURE	BCNN	Ours-NA	Ours-TA
40%	2.56	0.06	2.94	3.28	310.21	90013	2.91	2178
25%	2.61	0.07	2.93	3.31	321.56	89652	2.86	2136
10%	2.29	0.05	2.80	3.30	320.43	91089	2.79	2134

While the model adaption brings notable performance gain by exploiting the internal statistics of a test sample, it also has a much less computational overhead than the internal learning methods. See Table 5 for the testing time of some selected methods on all images of BSD68, with the comparison to ours. The speed of Ours-NA is not bad among the compared methods. With an additional model adaptation process in the inference phase, Ours-TA is slower than Ours-NA but is significantly faster than BCNN. We also select the supervised methods including COAST and ISTANet+ for comparison.

5.5 Ablation Studies

See Table 6 for the ablation studies and results, whose details are as follows.

Without $\mathcal{L}^{\text{Image}}$ We retrain the NN by removing the loss $\mathcal{L}^{\text{Image}}$ in (3). Noticeable performance degradation is observed when compared to Our-NA. This has indicated the effectiveness of $\mathcal{L}^{\text{Image}}$ for handling the null-space ambiguity in unsupervised learning. It is interesting to see that using only the measurement-domain loss $\mathcal{L}^{\text{Measure}}$ may yield reasonable performance. This is probably because the measurement-domain training together with the unfolding-based architecture has certain implicit bias to smoothness on the NN’s output, which resolves the null-space ambiguity to some degree.

Without $\mathcal{L}^{\text{Measure}}$ We remove the loss $\mathcal{L}^{\text{Measure}}$ in (3) and retrain the NN. It causes Ours-NA not to work. This is not surprising as the reconstruction from available measurements does play a critical role in the whole reconstruction and the effectiveness of the loss $\mathcal{L}^{\text{Image}}$ also relies on the success of the reconstruction in $\text{im}(\Phi^H)$ by sharing the NN weights during learning.

Without outer noise injection We only inject noise into the NN inputs \mathbf{y} and \mathbf{z} , and remove the outer noise injection in (3). A noticeable performance decrease is observed when compared to Ours-NA. The reason may be that the resulting loss function loses its connection to the supervised counterpart and leads to larger training error.

Using Gaussian noise for injected \mathbf{r} The sign-flipping-based generation scheme for injected noise is replaced by a simple scheme where $\mathbf{r} \sim \mathcal{N}(0, \alpha\mathbf{I})$ with α estimated from \mathbf{e}' . The PSNR results are fine but lower than that using the proposed generation scheme.

Supervised training We retrain our NN with a supervised ℓ_2 loss using the same paired data as those compared supervised methods. It leads to certain improvement over Ours-NA, but not significant. This demonstrates that the proposed dual-domain self-supervised loss is even as effective as the supervised loss, which coincides with Proposition 1.

Using other NN structures We replace the NN structure with that of the ISTANet+, a representative NN in supervised compressive imaging. The resulting performance is very close to the original one. We also apply the test-time model adaption to the supervised ISTANet+. There is also a noticeable improvement over the original one in Table 3. All these results suggest the possible applicability of both our training and adaption schemes to other NNs.

Table 6. Results of ablation studies in terms of mean PSNR(dB) values on Set11.

Noise Level		0			10		
CS Ratio		0.4	0.25	0.1	0.4	0.25	0.1
NA	w/o $\mathcal{L}^{\text{Image}}$	34.11	31.03	24.32	30.12	27.52	22.16
	w/o $\mathcal{L}^{\text{Measure}}$	16.67	15.09	11.25	15.21	14.16	10.32
	w/o Outer Injection	28.20	20.83	15.66	27.33	18.37	13.24
	Gaussian injected \mathbf{r}	35.67	31.65	26.18	31.39	29.31	24.81
	Supervised Training	36.98	32.93	27.02	31.57	29.60	26.55
TA	Ours-NA	36.37	32.70	26.89	31.42	29.17	25.41
	Ours-TA (ISTANet+)	36.67	33.23	27.18	31.02	29.31	25.85
	Adaptive ISTA-Net+	36.70	33.71	27.34	31.48	29.63	26.21
	Ours-TA	37.18	33.41	27.57	31.94	29.84	26.06

6 Conclusion

We proposed an unsupervised deep learning approach for compressive imaging. It is based on a dual-domain self-supervised training scheme which not only allows effective learning in both measurement domain and image domain without any GT image, but also allows test-time model adaption for enjoying the advantages from both external and internal learning. The effectiveness of the proposed approach is grounded by mathematical analysis and has been demonstrated by extensive experiments on three imaging tasks. The developed techniques can be extended to solving other ill-posed problems, which is our future work.

References

1. Chen, D., Davies, M.E.: Deep decomposition learning for inverse imaging problems. In: Proceedings of European Conference on Computer Vision (2020)
2. Chen, D., Tachella, J., Davies, M.E.: Equivariant imaging: Learning beyond the range space. In: Proceedings of International Conference on Computer Vision (2021)
3. Chen, D., Tachella, J., Davies, M.E.: Robust equivariant imaging: a fully unsupervised framework for learning to image from noisy and partial measurements. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2022)
4. Chen, J., Sun, Y., Liu, Q., Huang, R.: Learning memory augmented cascading network for compressed sensing of images. In: Proceedings of European Conference on Computer Vision. pp. 513–529. Springer (2020)
5. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al.: The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging* **26**(6), 1045–1057 (2013)
6. Cole, E.K., Pauly, J.M., Vasanawala, S.S., Ong, F.: Unsupervised MRI reconstruction with generative adversarial networks. arXiv preprint arXiv:2008.13065 (2020)
7. Diamond, S., Sitzmann, V., Heide, F., Wetzstein, G.: Unrolled optimization with deep priors. arXiv preprint arXiv:1705.08041 (2017)
8. Ding, Q., Chen, G., Zhang, X., Huang, Q., Ji, H., Gao, H.: Low-dose ct with deep learning regularization via proximal forward–backward splitting. *Physics in Medicine & Biology* **65**(12), 125009 (2020)
9. Dong, W., Shi, G., Li, X., Ma, Y., Huang, F.: Compressive sensing via nonlocal low-rank regularization. *IEEE Transactions on Image Processing* **23**(8), 3618–3632 (2014)
10. Eldar, Y.C.: Generalized sure for exponential families: Applications to regularization. *IEEE Transactions on Signal Processing* **57**(2), 471–481 (2008)
11. Feng, C.M., Yang, Z., Chen, G., Xu, Y., Shao, L.: Dual-octave convolution for accelerated parallel mr image reconstruction. In: Proceedings of AAAI Conference on Artificial Intelligence (2021)
12. Jalal, A., Karmalkar, S., Dimakis, A.G., Price, E.: Instance-optimal compressed sensing via posterior sampling. Proceedings of International Conference on Machine Learning (2021)
13. Jin, K.H., McCann, M.T., Froustey, E., Unser, M.A.: Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing* **26**, 4509–4522 (2017)
14. Kabkab, M., Samangouei, P., Chellappa, R.: Task-aware compressed sensing with generative adversarial networks. In: Proceedings of AAAI Conference on Artificial Intelligence (2018)
15. Kadkhodaie, Z., Simoncelli, E.: Stochastic solutions for linear inverse problems using the prior implicit in a denoiser. *Advances in Neural Information Processing Systems* **34** (2021)
16. Kulkarni, K., Lohit, S., Turaga, P., Kerviche, R., Ashok, A.: Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 449–458 (2016)

17. Liu, J., Kuang, T., Zhang, X.: Image reconstruction by splitting deep learning regularization from iterative inversion. In: Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 224–231. Springer (2018)
18. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine* **58**(6), 1182–1195 (2007)
19. Mardani, M., Sun, Q., Donoho, D., Popyan, V., Monajemi, H., Vasanawala, S., Pauly, J.: Neural proximal gradient descent for compressive imaging. *Advances in Neural Information Processing Systems* **31** (2018)
20. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of International Conference on Computer Vision. vol. 2, pp. 416–423. IEEE (2001)
21. Metzler, C., Mousavi, A., Heckel, R., Baraniuk, R.: Unsupervised learning with stein’s unbiased risk estimator. *arXiv preprint arXiv:1805.10531* (2018)
22. Metzler, C.A., Maleki, A., Baraniuk, R.: Bm3d-amp: A new image recovery algorithm based on bm3d denoising. In: Proceedings of International Conference on Image Processing. pp. 3116–3120. IEEE (2015)
23. Metzler, C.A., Maleki, A., Baraniuk, R.: From denoising to compressed sensing. *IEEE Transactions on Information Theory* **62**(9), 5117–5144 (2016)
24. Metzler, C.A., Mousavi, A., Baraniuk, R.G.: Learned D-AMP: Principled neural network based compressive image recovery. *Advances in Neural Information Processing Systems* (2017)
25. Pang, T., Quan, Y., Ji, H.: Self-supervised bayesian deep learning for image recovery with applications to compressive sensing. In: Proceedings of European Conference on Computer Vision (2020)
26. Pang, T., Zheng, H., Quan, Y., Ji, H.: Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2021)
27. Quan, Y., Chen, M., Pang, T., Ji, H.: Self2self with dropout: Learning self-supervised denoising from single image. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 1890–1898 (2020)
28. Quan, Y., Qin, X., Chen, M., Huang, Y.: High-quality self-supervised snapshot hyperspectral imaging. In: IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 1526–1530. IEEE (2022)
29. Shi, W., Jiang, F., Liu, S., Zhao, D.: Scalable convolutional neural network for image compressed sensing. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 12290–12299 (2019)
30. Shi, W., Jiang, F., Liu, S., Zhao, D.: Image compressed sensing using convolutional neural network. *IEEE Transacitons on Image Processing* **29**, 375–388 (2019)
31. Sun, Y., Chen, J., Liu, Q., Liu, B., Guo, G.: Dual-path attention network for compressed sensing image reconstruction. *IEEE Transactions on Image Processing* **29**, 9482–9495 (2020)
32. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 9446–9454 (2018)
33. Wang, W., Li, J., Ji, H.: Self-supervised deep image restoration via adaptive stochastic gradient langevin dynamics. In: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1989–1998 (2022)

34. Wei, K., Aviles-Rivero, A., Liang, J., Fu, Y., Schönlieb, C.B., Huang, H.: Tuning-free plug-and-play proximal algorithm for inverse imaging problems. In: Proceedings of International Conference on Machine Learning. pp. 10158–10169. PMLR (2020)
35. Xia, Z., Chakrabarti, A.: Training image estimators without image ground-truth. Advances in Neural Information Processing Systems (2019)
36. Xiang, J., Dong, Y., Yang, Y.: FISTA-Net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. IEEE Transactions on Medical Imaging **40**(5), 1329–1339 (2021)
37. Xin, B., Wang, Y., Gao, W., Wipf, D.: Maximal sparsity with deep networks? Advances in Neural Information Processing Systems (2016)
38. Xu, K., Zhang, Z., Ren, F.: Lapran: A scalable laplacian pyramid reconstructive adversarial network for flexible compressive sensing reconstruction. In: Proceedings of European Conference on Computer Vision. pp. 485–500 (2018)
39. Yang, Y., Sun, J., Li, H., Xu, Z.: ADMM-CSNet: A deep learning approach for image compressive sensing. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(3), 521–538 (2019)
40. You, D., Zhang, J., Xie, J., Chen, B., Ma, S.: Coast: Controllable arbitrary-sampling network for compressive sensing. IEEE Transactions on Image Processing **30**, 6066–6080 (2021)
41. Zalbagi Darestani, M., Heckel, R.: Accelerated MRI with un-trained neural networks. In: IEEE Transactions on Computational Imaging. vol. 7, pp. 724–733 (2021)
42. Zhang, J., Ghanem, B.: ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 1828–1837 (2018)
43. Zhang, Z., Liu, Y., Liu, J., Wen, F., Zhu, C.: AMP-Net: Denoising-based deep unfolding for compressive image sensing. IEEE Transactions on Image Processing **30**, 1487–1500 (2021)
44. Zheng, H., Fang, F., Zhang, G.: Cascaded dilated dense network with two-step data consistency for MRI reconstruction. Advances in Neural Information Processing Systems (2019)
45. Zhou, B., Zhou, S.K.: DuDoRNet: Learning a dual-domain recurrent network for fast MRI reconstruction with deep T1 prior. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 4273–4282 (2020)
46. Zhussip, M., Soltanayev, S., Chun, S.: Training deep learning based image denoisers from undersampled measurements without ground truth and without image prior. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 10255–10264 (2019)