Siamese Cooperative Learning for Unsupervised Image Reconstruction from Incomplete Measurements

Yuhui Quan, Xinran Qin, Tongyao Pang, and Hui Ji

Abstract—Image reconstruction from incomplete measurements is one basic task in imaging. While supervised deep learning has emerged as a powerful tool for image reconstruction in recent years, its applicability is limited by its prerequisite on a large number of latent images for model training. To extend the application of deep learning to the imaging tasks where acquisition of latent images is challenging, this paper proposes an unsupervised deep learning method that trains a deep model for image reconstruction with the access limited to measurement data. We develop a Siamese network whose twin sub-networks perform reconstruction cooperatively on a pair of complementary spaces: the null space of the measurement matrix and the range space of its pseudo inverse. The Siamese network is trained by a self-supervised loss with three terms: a data consistency loss over available measurements in the range space, a data consistency loss between intermediate results in the null space, and a mutual consistency loss on the predictions of the twin sub-networks in the full space. The proposed method is applied to four imaging tasks from different applications, and extensive experiments have shown its advantages over existing unsupervised solutions.

Index Terms—Image reconstruction, Unsupervised learning, Deep learning, Siamese neural networks.

1 Introduction

Image reconstruction from incomplete linear measurements finds its applications in many imaging tasks, such as medical imaging [1], [2], multi-spectral imaging [3], [4], and high-speed imaging [5]. It can be formulated as solving an ill-posed linear inverse problem:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{A} \in \mathbb{C}^{d\times D}$ ($d \ll D$) denotes the so-called measurement matrix which models the forward acquisition process, $\boldsymbol{x} \in \mathbb{R}^D$ the latent image to reconstruct, $\boldsymbol{n} \in \mathbb{C}^d$ the measurement noise, and $\boldsymbol{y} \in \mathbb{C}^d$ the collected incomplete measurements. As the system in (1) is under-determined, a direct inversion is not unique and is sensitive to measurement noise. Over last decades, regularization has been one prominent approach to solving ill-posed inverse problems, which addresses solution ambiguity and noise sensitivity by imposing certain priors on latent images for reconstruction. While these regularization-based methods saw their success in certain applications, *e.g.*, sparsity-based ℓ_1-norm regularization for compressed-sensing (CS) medical imaging [6], [7], the need for more accurate and faster reconstruction with fewer measurements remains in practice.

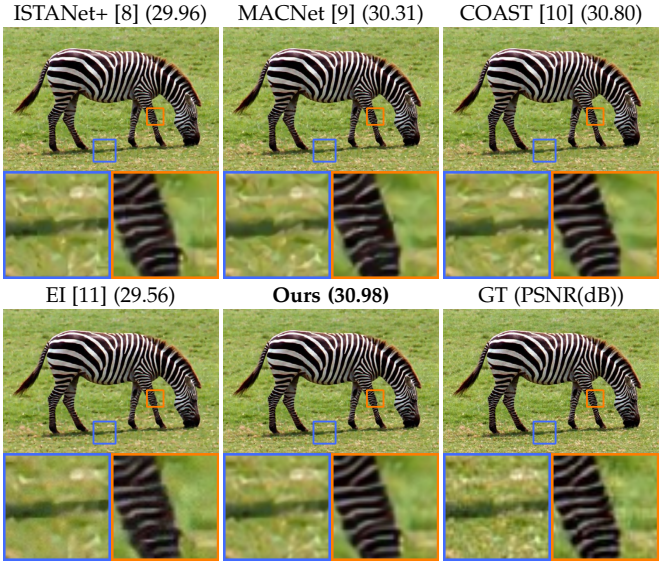


Fig. 1. Natural image reconstruction results with sensing ratio 0.25. The ISTANet+, MACNet and COAST are supervised methods, while EI and Ours are unsupervised methods.

In recent years, deep learning has been one main driving force in many fields, including image reconstruction; see *e.g.* [12], [13], [8], [14], [15], [9], [3], [16], [10], [17], [18], [2]. The paradigm of most existing deep learning-based methods is training a deep neural network (DNN) model over a paired dataset of latent images and their measurements. While the resulting models show impressive performance in terms of both reconstruction accuracy and running-time efficiency, such an advantage requires access to a large paired dataset. There are also some weakly-supervised methods

- *Yuhui Quan and Xinran Qin were with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510000, China. (email: csyhquan@scut.edu.cn; csqinxinran@gmail.com)*
- *Tongyao Pang and Hui Ji were with the Department of Mathematics, National University of Singapore 119076, Singapore. (e-mail: matpt@nus.edu.sg; matjh@nus.edu.sg)*
- *Corresponding author: Xinran Qin.*
- *This work was supported in part by National Natural Science Foundation of China under Grant 62372186, in part by Natural Science Foundation of Guangdong Province under Grants 2022A1515011755 and 2023A1515012841, in part by Fundamental Research Funds for the Central Universities under Grant x2js-D2230220, and in part by Singapore MOE AcRF Tier 1 under Grant A-8000981-00-00.*

(*e.g.* [19]) that utilize unpaired latent images and measurements. Nevertheless, the prerequisite of latent images still limits their wider applications in many domains, *e.g.*, medicine and science, where acquisition of latent images is challenging. The challenges are multifaceted. For instance, transient phenomena such as fleeting cellular reactions and rapid astronomical events, are elusive. Many subjects at microscopic scales are hard to image with traditional techniques. Some methods, especially in material science, may alter the very sample under observation. Obstacles like low signal-to-noise ratios, intricate data processing needs, ethical and safety concerns in medical scenarios, limited access to advanced imaging tools, rigorous sample preparation requirements, and fundamental physical imaging limits further compound these challenges.

To bypass the challenges raised by the prerequisite of domain-specific latent images, there has been an increasing interest in leveraging DNN models pre-trained on the images from another domain; see *e.g.*, plug-and-play methods [20], [21], [22] and generative methods [23], [24]. However, the performance will not be satisfactory when the images of the two domains differ much in structure. For instance, a model trained over natural images in digital photography may not generalize well to medical images of internal organs or scientific images of biological molecules.

1.1 Motivation and Aim

Unsupervised deep learning in the absence of ground-truth (GT) images is receiving increasing attention in many image reconstruction and recovery tasks. Recent studies [25], [26], [27], [28], [29], [11] have shown that it is possible to train a DNN for image reconstruction without using GT images. The key part is how to mitigate the over-fitting that can occur when training with only the measurements with noisy and partial information of GT images. Some methods [25], [28] require multiple acquisitions of the same image, which is inconvenient or impossible in practice. There are also studies [30], [31], [4] employing an untrained DNN as a generative image prior for image reconstruction, without using any external latent image in self-supervised learning. However, in many cases, their performance is not comparable to their supervised counterparts. In addition, they use an online learning scheme which trains different DNN models for different test samples. Such an online scheme is computationally overwhelming for large-scale data and not qualified for real-time applications.

The limitations of the aforementioned unsupervised deep learning methods, in terms of both reconstruction accuracy and testing-time efficiency, motivated us to develop an alternative with the following desired features:

1) Training neither on latent images nor on multiple acquisitions of the same image;
2) An offline training scheme that addresses the testing-time efficiency of online self-supervised learning schemes;
3) Competitive performance against existing supervised learning-based methods to meet practical needs.

See Fig. 1 for an illustration of the images reconstructed by the proposed unsupervised deep learning approach and several existing methods for CS-based image acquisition, a task to reconstruct an image from a small number of measurements sensed with some specific sensing matrix [6].

1.2 Basic Idea

Before proceeding, we first introduce some notions and basics on matrices. Let $\boldsymbol{I}_m \in \mathbb{R}^{m\times m}$ denote the $m \times m$ identity matrix. For a matrix $\boldsymbol{T} \in \mathbb{C}^{M\times N}$, its conjugate transpose is denoted by $\boldsymbol{T}^* \in \mathbb{C}^{N\times M}$. Its range space $\mathcal{R}(\boldsymbol{T})$ and null space $\mathcal{N}(\boldsymbol{T})$ are defined by

$$\mathcal{R}(\boldsymbol{T}) \triangleq \{\boldsymbol{T}\boldsymbol{u} : \boldsymbol{u} \in \mathbb{C}^N\} \subseteq \mathbb{C}^M, \tag{2}$$

$$\mathcal{N}(\boldsymbol{T}) \triangleq \{\boldsymbol{u} \in \mathbb{C}^N : \boldsymbol{T}\boldsymbol{u} = \boldsymbol{0}\} \subseteq \mathbb{C}^N. \tag{3}$$

For a measurement matrix $\boldsymbol{A} \in \mathbb{C}^{d\times D}$ with full row rank, its Moore–Penrose inverse is defined as $\boldsymbol{A}^\dagger = \boldsymbol{A}^*(\boldsymbol{A}\boldsymbol{A}^*)^{-1} \in \mathbb{C}^{D\times d}$ such that $\boldsymbol{A}\boldsymbol{A}^\dagger = \boldsymbol{I}_d$. Then, the Euclidean space $\mathbb{C}^D$ has the following orthogonal decomposition:

$$\mathbb{C}^D = \mathcal{R}(\boldsymbol{A}^\dagger) \oplus \mathcal{N}(\boldsymbol{A}). \tag{4}$$

In other words, any $\boldsymbol{x} \in \mathbb{C}^D$ can be uniquely decomposed as $\boldsymbol{x} = \boldsymbol{x}^+ + \boldsymbol{x}^\perp$ where $\boldsymbol{x}^+ \in \mathcal{R}(\boldsymbol{A}^\dagger)$ and $\boldsymbol{x}^\perp \in \mathcal{N}(\boldsymbol{A})$.

For a training sample $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$ without the GT $\boldsymbol{x}$, it only provides noisy measurements of $\boldsymbol{x}$ in $\mathcal{R}(\boldsymbol{A}^\dagger)$. Recall that the training is about tuning the DNN model so that it can predict $\boldsymbol{x}$ well from $\boldsymbol{y}$. Thus, there are two questions to answer when training a DNN using only $\boldsymbol{y}$.

1) How to handle measurement noise in the training sample $\boldsymbol{y}$ so that the reconstructed $\boldsymbol{x}$ in $\mathcal{R}(\boldsymbol{A}^\dagger)$ is accurate?
2) How to train the DNN to reconstruct $\boldsymbol{x}^\perp$ when no information of $\boldsymbol{x}$ in $\mathcal{N}(\boldsymbol{A})$ provided by the training sample?

To answer these two questions, we propose a Siamese network with twin DNNs $(\mathcal{M}_\mathcal{R}, \mathcal{M}_\mathcal{N})$ constructed via unrolling the proximal gradient iterative algorithm. The twin DNNs focus on the reconstruction of $\boldsymbol{x}$ in $\mathcal{R}(\boldsymbol{A}^\dagger)$ and in $\mathcal{N}(\boldsymbol{A})$, respectively. That is, the outputs of the twin DNNs, $\boldsymbol{x}_\mathcal{R}$ and $\boldsymbol{x}_\mathcal{N}$ are expected to satisfy

$$\boldsymbol{x}_\mathcal{R} \approx \boldsymbol{x} + \boldsymbol{u}, \boldsymbol{u} \in \mathcal{N}(\boldsymbol{A}), \tag{5}$$

$$\boldsymbol{x}_\mathcal{N} \approx \boldsymbol{x} + \boldsymbol{v}, \boldsymbol{v} \in \mathcal{R}(\boldsymbol{A}^\dagger). \tag{6}$$

Then, the final prediction is defined by

$$\widehat{\boldsymbol{x}} = (\boldsymbol{A}^\dagger\boldsymbol{A})\boldsymbol{x}_\mathcal{R} + (\boldsymbol{I} - \boldsymbol{A}^\dagger\boldsymbol{A})\boldsymbol{x}_\mathcal{N}, \tag{7}$$

which will eliminate $\boldsymbol{u}, \boldsymbol{v}$ to have an accurate estimate of $\boldsymbol{x}$.

The answer to the first question is a self-supervised data consistency loss, which is the extension of the Recorrupted2Recorrupted (R2R) self-supervised Gaussian denoising network [32] from denoising to image reconstruction. It is shown in this paper that the proposed self-supervised loss provided a loss equivalent to its supervised counterpart in the range space $\mathcal{R}(\boldsymbol{A}^\dagger)$. That is, the proposed self-supervised loss can well simulate the loss defined on the noise-free measurements $\boldsymbol{A}\boldsymbol{x} \in \mathcal{R}(\boldsymbol{A}^\dagger)$. As a result, this loss enables us to train $\mathcal{M}_\mathcal{R}$ to predict $\boldsymbol{x}$ in $\mathcal{R}(\boldsymbol{A}^\dagger)$ without the impact from the measurement noise, despite that only noisy measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$ are available

The answer to the second question is motivated by the Noise2Noise (N2N) training [33], which uses paired noisy images of the same scene to simulate the training of a denoiser over noisy/clean image pairs. For the prediction of $\boldsymbol{x}^\perp$ in $\mathcal{N}(\boldsymbol{A})$, we simulate a N2N training on $\mathcal{M}_\mathcal{N}$ using multiple estimates from $\mathcal{M}_\mathcal{R}$. The idea is based on the observation that the DNN architecture has certain regularization effect on its output, called deep image prior

(see *e.g.* [34], [35]). Then the predictions obtained by $\mathcal{M}_{\mathcal{R}}$ can be considered as fine initial estimates with relatively in-significant error in $\mathcal{N}(\boldsymbol{A})$. In addition, empirically, the random noise injected into the the inputs of $\mathcal{M}_{\mathcal{R}}$ in the R2R loss diversifies the outputs such that the predictions of $\mathcal{M}_{\mathcal{R}}$ in two adjacent steps can provide different views of $\boldsymbol{x}$ with small correlations. Intuitively, we treat the projections of these predictions onto $\mathcal{N}(\boldsymbol{A})$ as the diverse observations of $\boldsymbol{x}^{\perp}$ with weakly-correlated noise and use them as paired data to train $\mathcal{M}_{\mathcal{N}}$ in a N2N manner.

The twin DNNs $\mathcal{M}_{\mathcal{R}}$ and $\mathcal{M}_{\mathcal{N}}$ share an optimization unrolling network architecture but with two different embedded matrices: a measurement-space projection matrix and a null-space projection matrix. The weight sharing between the twin DNNs allow them to cooperate with each other during training, while the different embedding matrices make them more effective for image reconstruction in their own subspaces. To enable better interaction between the reconstructions from the twin DNNs, *i.e.*, $\boldsymbol{x}_{\mathcal{R}}$ and $\boldsymbol{x}_{\mathcal{N}}$, we introduce a mutual loss that measures the distance between them for joint training. To conclude, we propose a DNN with a Siamese structure for cooperative image reconstruction in two subspaces: $\mathcal{R}(\boldsymbol{A}^{\dagger})$ and $\mathcal{N}(\boldsymbol{A})$. It is trained by a self-supervised loss with three terms defined only based on $\boldsymbol{y}$. which respectively accounts for the reconstruction in $\mathcal{R}(\boldsymbol{A}^{\dagger})$, $\mathcal{N}(\boldsymbol{A})$, and a fusion in the full image space $\mathbb{C}^D$.

1.3 Contributions

This paper proposes an unsupervised deep learning method for image reconstruction from incomplete measurements, where only requires unorganized training samples of measurements. There are two technical contributions:

1) A Siamese network for cooperative image reconstruction from incomplete noisy measurements, defined in an orthogonal decomposition of the image space.
2) A self-supervised loss function that allows effectively training the Siamese network for accurate image reconstruction, in the absence of latent images.

The experiments on four image reconstruction tasks show that our proposed unsupervised learning method provides better performance than existing unsupervised methods and is computationally more efficient than online self-supervised (test-time learning) methods. Moreover, the proposed method competes well with recent supervised learning methods, making it very attractive for real-world applications where acquisition of latent images is challenging. All our codes will be published once the paper is accepted.

The preliminary results of this paper appeared in a conference paper [36] which introduced a self-supervised loss defined in two domains: the measurement domain related to range space $\mathcal{R}(\boldsymbol{A}^{\dagger})$ and the image domain related to full space $\mathbb{R}^D$. In [36], loss functions defined on the two domains respectively are both based on an extended R2R loss where a flipping-sign scheme is used for the noise simulation on the R2R loss in the image domain. This paper extends the conference paper [36] in several aspects:

1) A novel Siamese DNN structure using twin sub-networks embedded with two different measurement matrices for collaborative prediction in $\mathcal{R}(\boldsymbol{A}^{\dagger})$ and $\mathcal{N}(\boldsymbol{A})$.
2) A self-supervised training loss defined in $\mathcal{N}(\boldsymbol{A})$ which avoids noise simulation done in the image-space loss of [36], with a consistency loss between two $\mathcal{R}(\boldsymbol{A}^{\dagger})$ and $\mathcal{N}(\boldsymbol{A})$ for mutual information fusion.
3) Additional application of the proposed method to hyperspectral imaging in the experiments.

Extensive experiments show that the proposed method outperforms its preliminary version [36] overall, indicating the significance of the extensions introduced in this paper.

2 Literature Review

There is abundant literature on non-learning-based methods for image reconstruction from incomplete measurements. The prominent one is regularization-based methods which impose some pre-defined image priors on latent images to guide the reconstruction process in a variational form. Due to space limitation, the following review on related works focuses on deep learning-based methods.

Supervised deep learning from paired data Most existing deep learning-based methods fall into this category, concentrating on architecture design. Early studies (*e.g.* [12]) used off-the-shelf DNNs. Recent methods (*e.g.* [37], [13], [20], [8], [14], [38], [39], [18], [3], [16], [10], [2], [40]) adopted deep algorithm unrolling to embed measurement physics into the DNN architecture. It is done by unrolling the computational scheme of some traditional image reconstruction method and casting the denoising-related operators into pre-trained or learnable DNN blocks. The DNN adopted in our method is also designed based on deep algorithm unrolling.

Some studies [41], [9], [15] constructed DNNs using a different strategy. One closely-related work is the deep decomposition network (DDN) proposed by Chen *et al.* [41], a light-weight yet efficient DNN performing range-null space decomposition-based reconstruction with a two-stage scheme. The DDN showed effectiveness in supervised learning but cannot be directly applied in our unsupervised learning scheme. In comparison, our DNN has a Siamese structure inspired by (4) to facilitate unsupervised learning. Recently, there is an increasing interest in developing supervised DNNs from different aspects, *e.g.*, improvement of computational scalability [42], [43], training of universal models [44], and exploitation of additional sources [45].

Unsupervised deep learning over organized measurement data To overcome the lack of latent images, a few methods base the model training on multiple samples of measurements captured from the same image but via different sensing matrices, *e.g.*, a pair of measurement matrices are adopted in [27] with a consistency loss, and multiple measurement matrices are adopted in [28] with an adversarial loss. While multiple acquisitions encode sufficient additional information of latent images for overcoming the solution ambiguity in unsupervised learning, this specific data acquisition manner is of limited practicability.

Unsupervised deep learning over unorganized measurement data There is an increasing interest in learning with unorganized samples of measurements only. In [25], [26], the Stein's unbiased estimator (SURE) is combined with the denoiser-approximate message passing (DAMP) network for unsupervised learning. The prediction on $\mathcal{R}(\boldsymbol{A}^{\dagger})$

is learned with SURE, while the ambiguity in $\mathcal{N}(\boldsymbol{A})$ is handled by the inherent implicit regularization from the DNN structure. In comparison, our method not only uses a different training loss for $\mathcal{R}(\boldsymbol{A}^\dagger)$, but also introduces an additional training scheme for $\mathcal{N}(\boldsymbol{A})$.

Very recently, Chen *et al.* [11] proposed an effective framework called equivariant imaging (EI) for unsupervised image reconstruction, focusing on addressing the learning ambiguity in $\mathcal{N}(\boldsymbol{A})$. Based on the equivariance presented in latent images, the idea of EI is to move the component of the DNN's prediction in $\mathcal{R}(\boldsymbol{A}^\dagger)$ of one sample to $\mathcal{N}(\boldsymbol{A})$ by some transformations such as translations and rotations to which $\mathcal{R}(\boldsymbol{A}^\dagger)$ are not invariant. This forms effective pseudo supervision on $\mathcal{N}(\boldsymbol{A})$ for another equivariant sample. A noise-robust version of EI, namely REI, is proposed in [46] to improve the robustness to measurement noise by combining EI with SURE. Compared with EI and REI, our method employs a different loss inspired by N2N to reduce the prediction error in $\mathcal{N}(\boldsymbol{A})$, while EI and REI assume the equivalence of training data to resolve the ambiguity from $\mathcal{N}(\boldsymbol{A})$. Also, the collaborative prediction by the Siamese network structure further reduces noise/error in $\mathcal{R}(\boldsymbol{A}^\dagger)$ and $\mathcal{N}(\boldsymbol{A})$. Moreover, the noise-resistant loss function for the learning on $\mathcal{R}(\boldsymbol{A}^\dagger)$ of the proposed method is more computationally effiicent than the SURE loss used in REI which requires expensive Monte Carlo sampling.

In our preliminary work [36], the R2R-extended loss is employed for both the treatment of measurement noise in $\mathcal{R}(\boldsymbol{A}^\dagger)$ and the treatment of prediction error in the full image space $\mathbb{R}^D$. As the distribution of prediction error is unknown, a sign-flipping scheme is adopted to simulate the noise from this unknown distribution. Such a noise simulation scheme leaves a lot of room for improvement. In this paper, while still adopting the R2R-extended loss for handling measurement noise in $\mathcal{R}(\boldsymbol{A}^\dagger)$, a simpler N2N-inspired loss is used for partially addressing prediction error in the complementary null space $\mathcal{N}(\boldsymbol{A})$. Then, the prediction error is further reduced by using a Siamese DNN with twin sub-networks that collaborate the predictions in $\mathcal{R}(\boldsymbol{A}^\dagger)$ and $\mathcal{N}(\boldsymbol{A})$, together with an additional consistency loss between them for information refinement in training.

Deep generative learning on a single test sample of measurements The methods [47], [30], [31], [4] in this category use a DNN with a certain structure to parameterize the latent image and train it as a generator with some data consistency loss on the test sample. Van *et al.* [47] used a regularized GAN, Pang *et al.* [30] used a Bayesian DNN with random weights, Sun *et al.* [31] used an attentive DNN, and Meng *et al.* [4] used an unrolling-based DNN. Different from ours, these methods suffer from high computational costs, as they have to train individual DNNs for different test samples.

Siamese learning Siamese DNNs have been extensively studied for high-level and middle-level vision tasks, often used with contrastive learning; see *e.g.* [48], [49]. This paper is one of the few works to apply Siamese DNNs to low-level vision tasks, which is for cooperative learning in two spaces.

3 PRELIMINARIES

Before proceeding, we first provide some preliminaries on two existing unsupervised denoising methods related to our proposed approach: N2N [33] and R2R [32].

The N2N was proposed to train a denoising DNN using paired noisy images of the same scene with independent noise. Given paired noisy images

$$\hat{\boldsymbol{i}} = \boldsymbol{x} + \hat{\boldsymbol{n}}, \quad \text{and} \quad \tilde{\boldsymbol{i}} = \boldsymbol{x} + \tilde{\boldsymbol{n}}, \tag{8}$$

the N2N loss is defined by

$$L_{\text{N2N}} \triangleq \mathbb{E}_{\hat{\boldsymbol{i}},\tilde{\boldsymbol{i}}} \|\mathcal{F}_{\boldsymbol{\theta}}(\hat{\boldsymbol{i}}) - \tilde{\boldsymbol{i}}\|_2^2, \tag{9}$$

where $\mathcal{F}_{\boldsymbol{\theta}}$ denotes the DNN for denoising. Its connection to the loss in supervised learning defined over noisy/clean pair $(\boldsymbol{i}, \boldsymbol{x})$, $\mathbb{E}_{\hat{\boldsymbol{i}}} \|\mathcal{F}_{\boldsymbol{\theta}}(\hat{\boldsymbol{i}}) - \boldsymbol{x}\|_2^2$, could be seen by expanding (9):

$$L_{\text{N2N}} \triangleq \mathbb{E}_{\hat{\boldsymbol{i}},\tilde{\boldsymbol{i}}} \big[\|\mathcal{F}_{\boldsymbol{\theta}}(\hat{\boldsymbol{i}}) - \boldsymbol{x}\|_2^2 + 2\tilde{\boldsymbol{n}}^\top \mathcal{F}_{\boldsymbol{\theta}}(\hat{\boldsymbol{i}}) + \|\tilde{\boldsymbol{n}}\|_2^2\big]. \tag{10}$$

Note that the second term satisfies

$$\mathbb{E}_{\hat{\boldsymbol{i}},\tilde{\boldsymbol{i}}} 2\tilde{\boldsymbol{n}}^\top \mathcal{F}_{\boldsymbol{\theta}}(\hat{\boldsymbol{i}}) = 2\mathbb{E}_{\boldsymbol{x}} \big[\mathbb{E}_{\tilde{\boldsymbol{n}},\hat{\boldsymbol{n}}|\boldsymbol{x}} \tilde{\boldsymbol{n}} \mathcal{F}_{\boldsymbol{\theta}}(\boldsymbol{x} + \hat{\boldsymbol{n}})\big] = 0, \tag{11}$$

if $\tilde{\boldsymbol{n}}$ and $\hat{\boldsymbol{n}}$ are conditionally independent given $\boldsymbol{x}$ and $\mathbb{E}_{\tilde{\boldsymbol{n}}|\boldsymbol{x}} \tilde{\boldsymbol{n}} = \boldsymbol{0}$. In addition, the last term $\mathbb{E}_{\hat{\boldsymbol{i}},\tilde{\boldsymbol{i}}} \|\tilde{\boldsymbol{n}}\|_2^2$ is a constant irrelevant to $\boldsymbol{\theta}$. Therefore, for the pair $(\hat{\boldsymbol{i}}, \tilde{\boldsymbol{i}})$ defined by (8) where $\hat{\boldsymbol{n}}$ and $\tilde{\boldsymbol{n}}$ are conditionally independent given $\boldsymbol{x}$ and $\mathbb{E}_{\tilde{\boldsymbol{n}}|\boldsymbol{x}} \tilde{\boldsymbol{n}} = \boldsymbol{0}$, we have then

$$L_{\text{N2N}} = \mathbb{E}_{\hat{\boldsymbol{i}},\boldsymbol{x}} \|\mathcal{F}_{\boldsymbol{\theta}}(\hat{\boldsymbol{i}}) - \boldsymbol{x}\|_2^2 + \text{const.}. \tag{12}$$

In other words, the N2N loss for denoising defined over the pairs of noisy/noisy images can simulate well the loss over pairs of noisy/clean images seen in supervised learning.

The R2R is inspired by N2N and further relaxes the requirements on training data, from paired noisy images to unorganized noisy images. Consider a singe noisy image $\boldsymbol{i} = \boldsymbol{x} + \boldsymbol{n}$, R2R constructs paired noisy images $(\hat{\boldsymbol{i}}, \tilde{\boldsymbol{i}})$ via

$$\hat{\boldsymbol{i}} = \boldsymbol{i} + \boldsymbol{D}^\top \boldsymbol{n}', \ \tilde{\boldsymbol{i}} = \boldsymbol{i} - \boldsymbol{D}^{-1} \boldsymbol{n}', \tag{13}$$

where $\boldsymbol{D}$ can be any invertible matrix (often set to a diagonal matrix for simplicity), and $\boldsymbol{n}'$ denotes simulated noise. Denote the noise in the generated noisy images $\hat{\boldsymbol{i}}$ and $\tilde{\boldsymbol{i}}$ by $\hat{\boldsymbol{n}} = \hat{\boldsymbol{i}} - \boldsymbol{x} = \boldsymbol{n} + \boldsymbol{D}^\top \boldsymbol{n}'$ and $\tilde{\boldsymbol{n}} = \tilde{\boldsymbol{i}} - \boldsymbol{x} = \boldsymbol{n} - \boldsymbol{D}^{-1} \boldsymbol{n}'$. Suppose that the simulated noise $\boldsymbol{n}'|\boldsymbol{x}$ and the image noise $\boldsymbol{n}|\boldsymbol{x}$ are independent and identically distributed (i.i.d.) Gaussian noise, then $\hat{\boldsymbol{n}}|\boldsymbol{x}$ and $\tilde{\boldsymbol{n}}|\boldsymbol{x}$ also follow normal distributions with zero covariance, *i.e.*, $\hat{\boldsymbol{n}}|\boldsymbol{x}$ and $\tilde{\boldsymbol{n}}|\boldsymbol{x}$ are independent. Thus, the R2R training over the synthesized noisy image pairs $(\hat{\boldsymbol{i}}, \tilde{\boldsymbol{i}})$ mimics the N2N training over a pair of images with independent noise, which connects to the supervised loss in the form of (12); see Proposition 1.

Proposition 1 ([32]). *Consider $\boldsymbol{i} = \boldsymbol{x} + \boldsymbol{n}$, where $\boldsymbol{n}|\boldsymbol{x}$ follows a zero-mean normal distribution. Suppose $\boldsymbol{n}'|\boldsymbol{x}$ is i.i.d. to $\boldsymbol{n}|\boldsymbol{x}$. The R2R loss is defined by*

$$L_{R2R}(\boldsymbol{\theta}) \triangleq \mathbb{E}_{\boldsymbol{i},\boldsymbol{n}'} \|\mathcal{F}_{\boldsymbol{\theta}}(\boldsymbol{i} + \boldsymbol{D}^\top \boldsymbol{n}') - (\boldsymbol{i} - \boldsymbol{D}^{-1} \boldsymbol{n}')\|_2^2 \tag{14}$$

for any invertible matrix $\boldsymbol{D}$ and it holds that

$$L_{R2R}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{n},\boldsymbol{n}'} \|\mathcal{F}_{\boldsymbol{\theta}}(\boldsymbol{i} + \boldsymbol{D}^\top \boldsymbol{n}') - \boldsymbol{x}\|_2^2 + const. \tag{15}$$

4 Siamese DNN for Cooperative Learning

Our proposed approach is based on the well-known Range-Null orthogonal decomposition.

Definition 1 (Range-Null orthogonal decomposition)**.** *For a sensing matrix $\boldsymbol{A} \in \mathbb{C}^{d\times D}$ with row full rank, define*

$$\boldsymbol{P}_{\mathcal{R}} = \boldsymbol{A}^{\dagger}\boldsymbol{A}, \quad \boldsymbol{P}_{\mathcal{N}} = (\boldsymbol{I}_D - \boldsymbol{A}^{\dagger}\boldsymbol{A}), \tag{16}$$

The matrices, $\boldsymbol{P}_{\mathcal{R}}$ and $\boldsymbol{P}_{\mathcal{N}}$, are the operators of the orthogonal projection of $\mathbb{C}^D$ to two subspaces, $\mathcal{R}(\boldsymbol{A}^{\dagger})$ and $\mathcal{N}(\boldsymbol{A})$, respectively. For any $\boldsymbol{x} \in \mathbb{C}^D$, there exists an unique decomposition: $\boldsymbol{x} = \boldsymbol{x}^{+} + \boldsymbol{x}^{\perp}$ with $\langle \boldsymbol{x}^{\dagger}, \boldsymbol{x}^{\perp}\rangle = 0$, defined by

$$\boldsymbol{x}^{+} = \boldsymbol{P}_{\mathcal{R}}\boldsymbol{x}, \quad \boldsymbol{x}^{\perp} = \boldsymbol{P}_{\mathcal{N}}\boldsymbol{x}. \tag{17}$$

In supervised learning, we can access the pair $(\boldsymbol{y}, \boldsymbol{x})$ to have a loss function that measures the reconstruction error. Then the DNN is trained to minimize that loss so that it can reconstruct $\boldsymbol{x}$ well from $\boldsymbol{y}$. In the absence of image $\boldsymbol{x}$, the available measurement $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$ only provides the information of $\boldsymbol{x}$ in $\mathcal{R}(\boldsymbol{A}^{\dagger})$. No information of $\boldsymbol{x}$ in $\mathcal{N}(\boldsymbol{A})$ is available. In other words, we only have a noisy observation of $\boldsymbol{x}^{+}$ while knowing nothing about $\boldsymbol{x}^{\perp}$. To develop an effective scheme for the training with the access limited to a set of $\boldsymbol{y}$, we need to address two challenges: how to handle the noise $\boldsymbol{n}$ in $\boldsymbol{y}$ to have an accurate estimate of $\boldsymbol{x}^{+}$, and how to regularize the reconstruction of $\boldsymbol{x}^{\perp}$ in training.

Cooperative reconstruction using Siamese network As the information of $\boldsymbol{x}^{\dagger}$ and $\boldsymbol{x}^{\perp}$ is provided differently in $\boldsymbol{y}$, we propose to separately reconstruct them in the DNN, yet with certain interaction between the two estimators. Our solution is adopting a DNN with Siamese structure, called *SiamNet* (Siamese Network), which has twin networks defined in the complementary spaces $(\mathcal{R}(\boldsymbol{A}^{\dagger}), \mathcal{N}(\boldsymbol{A}))$ and cooperating with each other for the final image reconstruction. See Fig. 2 for the illustration of the proposed SiamNet.

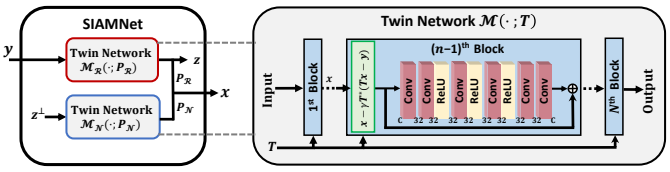


Fig. 2. Diagram of SiamNet for unsupervised image reconstruction. The feature channel numbers are displayed below the layers.

The twin DNNs in SiamNet, denoted by $(\mathcal{M}_{\mathcal{R}}, \mathcal{M}_{\mathcal{N}})$, share the same network structure as well as the trainable parameters, but take the inputs in different spaces (*i.e.* that captured by different measurement matrices). In a quick glance, $\mathcal{M}_{\mathcal{R}}$ and $\mathcal{M}_{\mathcal{N}}$ cannot be a Siamese pair as the space of their inputs are different. We address this by using deep algorithm unrolling for the network architecture design, which embeds the measurement matrices, $\boldsymbol{P}_{\mathcal{R}}$ and $\boldsymbol{P}_{\mathcal{N}}$, into the twin networks. Concretely, we have

$$\mathcal{M}_{\mathcal{R}}(\cdot;\boldsymbol{\omega}) : \mathcal{R}(\boldsymbol{A}^{\dagger}) \to \mathbb{C}^D, \tag{18}$$

$$\mathcal{M}_{\mathcal{N}}(\cdot;\boldsymbol{\omega}) : \mathcal{N}(\boldsymbol{A}) \to \mathbb{C}^D, \tag{19}$$

both of which are trained to predict the latent image from their input.

Implementation of SiamNet via deep algorithm unrolling Consider a penalized least squares problem:

$$\min_{\boldsymbol{x}} \ \|\bar{\boldsymbol{y}} - \boldsymbol{T}\boldsymbol{x}\|_2^2 + f(\boldsymbol{x}), \tag{20}$$

where f is a regularization term derived from some prior imposed on latent images. This problem is often solved by some iterative algorithms. Deep optimization unrolling forms a CNN by casting the steps related to the regularization term into learnable network blocks. Consider the proximal gradient method for solving (20):

$$\boldsymbol{x}_n = \mathrm{Prox}_{\gamma f}(\boldsymbol{x}_{n-1} - \gamma\boldsymbol{T}^{*}(\boldsymbol{T}\boldsymbol{x}_{n-1} - \bar{\boldsymbol{y}})), \tag{21}$$

for $n = 1, 2, \cdots$, where $\boldsymbol{x}^0 = \boldsymbol{T}^{*}\boldsymbol{y}$, $\gamma \in \mathbb{R}^{+}$ is a hyper-parameter, and the proximal operator $\mathrm{Prox}_{\gamma f}(\cdot)$ is given by

$$\mathrm{Prox}_{\gamma f}(\boldsymbol{x}) \triangleq \operatorname*{argmin}_{\boldsymbol{x}'} \|\boldsymbol{x} - \boldsymbol{x}'\|_2^2 + 2\gamma f(\boldsymbol{x}'). \tag{22}$$

Then we obtain a network $\mathcal{M}(\cdot;\boldsymbol{T})$ via replacing the proximal operator at each iteration by a residual block with six convolutional layers, where the middle four are individually equipped with a rectified linear unit (ReLU). As a result, the whole network $\mathcal{M}$ predicts $\boldsymbol{x}$ from $\bar{\boldsymbol{y}}$ using a number of stages. In each stage, it first performs the calculation of (21) and then passes the result to a residual block; see Fig. 2 for more details. Finally, the SiamNet is defined by

$$\mathcal{M}_{\mathcal{R}}(\cdot) = \mathcal{M}(\cdot;\boldsymbol{\omega}, \boldsymbol{P}_{\mathcal{R}}); \quad \mathcal{M}_{\mathcal{N}}(\cdot) = \mathcal{M}(\cdot;\boldsymbol{\omega}, \boldsymbol{P}_{\mathcal{N}}). \tag{23}$$

In other words, the twin networks in SiamNet are based on the same unrolling network $\mathcal{M}$ but embedded with two different measurement matrices respectively, which thus can handle the measurement data from $\boldsymbol{P}_{\mathcal{R}}$ and $\boldsymbol{P}_{\mathcal{N}}$ respectively.

In $\mathcal{M}(\cdot;\boldsymbol{T})$, the operator $\mathrm{prox}_{\gamma f}$ functions in the image space, and so are those residual blocks in the network. Thus, their dimensions do not vary according to the measurement matrix $\boldsymbol{T}$. In addition, the information incorporated in $\mathrm{prox}_{\gamma f}$ is about the assumed or learned image properties and has nothing to do with $\boldsymbol{T}$. That is, the physics of $\boldsymbol{T}$ is not embedded into the learnable part but only the non-learnable part of the unrolling network. Thus, the learnable modules can share their weights by the twin DNNs despite that they handle different measurement matrices: $\boldsymbol{P}_{\mathcal{R}}$ and $\boldsymbol{P}_{\mathcal{N}}$.

Remark 1. *The SiamNet is a pair of sub-networks with shared weights. The rationale of such a structure is two-fold: reduction in model size for alleviating possible overfitting and collaborating the constructions in two complementary subspaces. Recall that the learnable blocks in both sub-networks are about neuralizing proximal operators, which learn image priors for removing artifacts from previous estimates. Thus, the blocks in both sub-networks should be consistent in terms of their functions, which is ensured by weight sharing. As shown in the experiments in Section 6.5, such a structure does bring noticeable performance gain.*

5 Unsupervised Training and Inference

With the ideas discussed in Section 1.2, the following loss function is proposed for unsupervised training of SiamNet:

$$L_{\mathrm{All}} \triangleq L_{\mathcal{R}} + \alpha \cdot L_{\mathcal{N}} + \beta \cdot L_{\mathrm{C}}, \tag{24}$$

where $\alpha, \beta \in \mathbb{R}^{+}$ are two hyper-parameters, $L_{\mathcal{R}}$ is the loss measuring the reconstruction error in $\mathcal{R}(\boldsymbol{A}^{\dagger})$, $L_{\mathcal{N}}$ is the

loss for the reconstruction learning in $\mathcal{N}(\boldsymbol{A})$, and L_{C} is a mutual consistency loss to enable the interaction between twin networks for further reduction of solution ambiguity.

Learning image reconstruction in $\mathcal{R}(\boldsymbol{A}^\dagger)$ via $L_{\mathcal{R}}$ The measurement $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$ provides a noisy measurement of $\boldsymbol{x}^+$. By the definition of $\boldsymbol{x}^+$, we have

$$\boldsymbol{A}^\dagger \boldsymbol{y} = \boldsymbol{A}^\dagger \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^\dagger \boldsymbol{n} = \boldsymbol{x}^+ + \boldsymbol{A}^\dagger \boldsymbol{n}. \tag{25}$$

In other words, $\boldsymbol{x}^+$ can be constructed by removing the noise $\boldsymbol{A}^\dagger \boldsymbol{n}$ in $\boldsymbol{A}^\dagger \boldsymbol{y}$. Inspired by R2R, we propose the following loss for training $\mathcal{M}_{\mathcal{R}}$:

$$L_{\mathcal{R}} \triangleq \mathbb{E}_{\boldsymbol{y},\boldsymbol{n}'} \|\boldsymbol{P}_{\mathcal{R}} \mathcal{M}_{\mathcal{R}}(\boldsymbol{z}^+ + \boldsymbol{A}^\dagger \boldsymbol{n}') - \boldsymbol{z}^+ + \boldsymbol{A}^\dagger \boldsymbol{n}'\|_2^2, \tag{26}$$

where $\boldsymbol{z}^+ = \boldsymbol{A}^\dagger \boldsymbol{y}$, and $\boldsymbol{n}'$ denotes the newly-simulated noise. Note that the theoretical justification provided in [32] for Proposition 1 is only applicable to Gaussian noise, as its proof for the independence of $\hat{\boldsymbol{n}}$ and $\tilde{\boldsymbol{n}}$ is based on the specific properties of Gaussian noise with zero covariance. In real-world cases, measurement noise can be correlated Gaussian or non-Gaussian. Then, the theoretical connection of the R2R loss to its supervised counterpart does not hold anymore in such cases. In this paper, we revisit the R2R loss from a different perspective, extending the applicability of the R2R loss to the more general noise.

Proposition 2. *Consider $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}+\boldsymbol{n}$. For the loss function $L_{\mathcal{R}}$ defined by* (26) *where $\boldsymbol{n}'|\boldsymbol{x}^+$ and $\boldsymbol{n}|\boldsymbol{x}^+$ are i.i.d.. Then, we have*

$$L_{\mathcal{R}} = \mathbb{E}_{\boldsymbol{y},\boldsymbol{n}'} \|\boldsymbol{P}_{\mathcal{R}} \mathcal{M}_{\mathcal{R}}(\boldsymbol{z}^+ + \boldsymbol{A}^\dagger \boldsymbol{n}') - \boldsymbol{x}^+\|_2^2 + \text{const.}.$$

Proof. Recall that $\boldsymbol{z}^+ = \boldsymbol{A}^\dagger \boldsymbol{y} = \boldsymbol{x}^+ + \boldsymbol{A}^\dagger \boldsymbol{n}$. Rewrite $L_{\mathcal{R}}$ as

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{x}^+,\boldsymbol{n},\boldsymbol{n}'} \|\boldsymbol{P}_{\mathcal{R}} \mathcal{M}_{\mathcal{R}}(\boldsymbol{z}^+ + \boldsymbol{A}^\dagger \boldsymbol{n}') - \boldsymbol{z}^+ + \boldsymbol{A}^\dagger \boldsymbol{n}'\|_2^2 \\
=&\mathbb{E}_{\boldsymbol{x}^+,\boldsymbol{n},\boldsymbol{n}'} \|\boldsymbol{P}_{\mathcal{R}} \mathcal{M}_{\mathcal{R}}(\boldsymbol{x}^+ + \boldsymbol{A}^\dagger(\boldsymbol{n}+\boldsymbol{n}')) - \boldsymbol{x}^+ + \boldsymbol{A}^\dagger(\boldsymbol{n}'-\boldsymbol{n})\|_2^2 \\
=&\mathbb{E}_{\boldsymbol{x}^+,\boldsymbol{n},\boldsymbol{n}'} \Big\{ \|\boldsymbol{P}_{\mathcal{R}} \mathcal{M}_{\mathcal{R}}(\boldsymbol{z}^+ + \boldsymbol{A}^\dagger \boldsymbol{n}') - \boldsymbol{x}^+\|_2^2 + \|\boldsymbol{A}^\dagger(\boldsymbol{n}'-\boldsymbol{n})\|_2^2 \\
&+ 2\big(\boldsymbol{P}_{\mathcal{R}} \mathcal{M}_{\mathcal{R}}(\boldsymbol{x}^+ + \boldsymbol{A}^\dagger(\boldsymbol{n}+\boldsymbol{n}')) - \boldsymbol{x}^+\big)^\top \boldsymbol{A}^\dagger(\boldsymbol{n}'-\boldsymbol{n}) \Big\}.
\end{aligned} \tag{27}$$

The second term in the last line of (27) is a constant independent of the network parameters. As for the third term, since $\boldsymbol{n}|\boldsymbol{x}^+$ and $\boldsymbol{n}'|\boldsymbol{x}^+$ are i.i.d., by decomposing the joint distribution and switching $\boldsymbol{n}$ and $\boldsymbol{n}'$, we can obtain

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{x}^+} \mathbb{E}_{\boldsymbol{n}|\boldsymbol{x}^+} \mathbb{E}_{\boldsymbol{n}'|\boldsymbol{x}^+} \big(\boldsymbol{P}_{\mathcal{R}} \mathcal{M}_{\mathcal{R}}(\boldsymbol{x}^+ + \boldsymbol{A}^\dagger(\boldsymbol{n}+\boldsymbol{n}')) - \boldsymbol{x}^+\big)^\top \boldsymbol{A}^\dagger \boldsymbol{n}' \\
=\ &\mathbb{E}_{\boldsymbol{x}^+} \mathbb{E}_{\boldsymbol{n}|\boldsymbol{x}^+} \mathbb{E}_{\boldsymbol{n}'|\boldsymbol{x}^+} \big(\boldsymbol{P}_{\mathcal{R}} \mathcal{M}_{\mathcal{R}}(\boldsymbol{x}^+ + \boldsymbol{A}^\dagger(\boldsymbol{n}+\boldsymbol{n}')) - \boldsymbol{x}^+\big)^\top \boldsymbol{A}^\dagger \boldsymbol{n}.
\end{aligned} \tag{28}$$

Thus the third term is zero. Finally, we have

$$L_{\mathcal{R}} = \mathbb{E}_{\boldsymbol{y},\boldsymbol{n}'} \|\boldsymbol{P}_{\mathcal{R}} \mathcal{M}_{\mathcal{R}}(\boldsymbol{z}^+ + \boldsymbol{A}^\dagger \boldsymbol{n}') - \boldsymbol{x}^+\|_2^2 + \text{const.} \tag{29}$$

The proof is done. □

It is noted that different from [32], the justification of R2R loss on more general noise is only based on the symmetry of $\boldsymbol{n}$ and $\boldsymbol{n}'$, not specific properties of normal distribution. As the loss defined in Proposition 2 differs from the one defined in our preliminary work [36], the proof used in [36] needs to be revised to justify the loss function used in this paper.

Now, Proposition 2 says that $L_{\mathcal{R}}$ is an unbiased estimator to the supervised loss defined over GT images. With such a theoretical guarantee, the training of $\mathcal{M}(\cdot; \boldsymbol{P}_{\mathcal{R}})$ under the loss $L_{\mathcal{R}}$ will lead to an efficient estimator of $\boldsymbol{x}^+$ from its noisy version $\boldsymbol{A}^\dagger \boldsymbol{y}$, which has comparable performance to the supervised learning on the range space $\mathcal{R}(\boldsymbol{A}^\dagger)$.

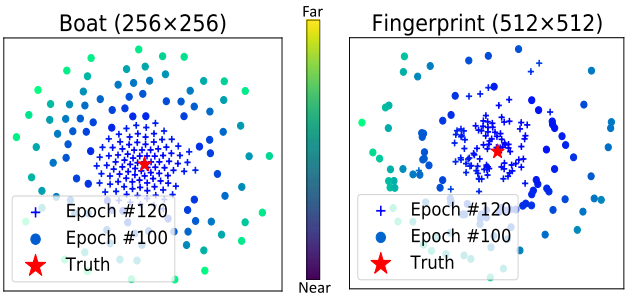


Fig. 3. Visualization (t-SNE) of the distribution of $\{\boldsymbol{z}_k^\perp\}$, produced by our model trained in different epochs. The instances are scattered around the GT. Their residuals to GT become smaller along the training process.

Learning image reconstruction in $\mathcal{N}(\boldsymbol{A})$ via $L_{\mathcal{N}}$ To tackle the issue that no information of $\boldsymbol{x}^\perp$ is available in training data, we utilize the intermediate reconstruction result from $\mathcal{M}_{\mathcal{R}}$ to form weak supervision for the prediction on $\mathcal{N}(\boldsymbol{A})$. Concretely, the loss $L_{\mathcal{N}}$ is defined as

$$L_{\mathcal{N}} \triangleq \mathbb{E}_{\boldsymbol{y},\boldsymbol{n}_1',\boldsymbol{n}_2'} \|\boldsymbol{P}_{\mathcal{N}} \mathcal{M}_{\mathcal{N}}(\boldsymbol{z}_1^\perp) - \boldsymbol{z}_2^\perp\|_2^2, \tag{30}$$

where $\boldsymbol{z}_k^\perp = \boldsymbol{P}_{\mathcal{N}} \mathcal{M}_{\mathcal{R}}(\boldsymbol{A}^\dagger \boldsymbol{y} + \boldsymbol{A}^\dagger \boldsymbol{n}_k')$ and $\boldsymbol{n}_k'$ is drawn from a normal distribution, for $k = 1, 2$. The motivation comes from the N2N training [33] which utilizes pairs of noisy images to mimic the supervised learning over the noisy/clean image pairs, provided that the noises in each noisy image pair are i.i.d., as shown in (12).

More specifically, the pair $(\boldsymbol{z}_1^\perp, \boldsymbol{z}_2^\perp)$ in (30) is generated by $\mathcal{M}_{\mathcal{R}}$ and used to mimic the N2N training for $\mathcal{M}_{\mathcal{N}}$. With the pairs of independent random noise $(\boldsymbol{n}_1', \boldsymbol{n}_2')$ injected into the inputs, the corresponding outputs, $\mathcal{M}_{\mathcal{R}}(\boldsymbol{A}^\dagger \boldsymbol{y} + \boldsymbol{A}^\dagger \boldsymbol{n}_1')$ and $\mathcal{M}_{\mathcal{R}}(\boldsymbol{A}^\dagger \boldsymbol{y} + \boldsymbol{A}^\dagger \boldsymbol{n}_2')$, are likely to have weak correlation, due to high non-linearity and redundancy of a DNN-based mapping. Therefore, one can have many instances of $\boldsymbol{z}_1^\perp, \boldsymbol{z}_2^\perp$ whose residuals to $\boldsymbol{x}^\perp$ have sufficient statistical independence. By viewing $\boldsymbol{z}_1^\perp, \boldsymbol{z}_2^\perp$ as the noisy versions of $\boldsymbol{x}^\perp$ and their residuals as noise, the loss $L_{\mathcal{N}}$ coincides with L_{N2N} and thus can approximate the supervised loss in $\mathcal{N}(\boldsymbol{A})$ well when correlation of the residuals of $\boldsymbol{z}_1^\perp, \boldsymbol{z}_2^\perp$ to $\boldsymbol{x}^\perp$ are sufficiently weak. It is empirically observed the assumption on the weak correlation holds true. See Fig. 3 for a demonstration.

Improvement via L_{C} The mutual consistency loss L_{C} measures the distance between the estimates of two DNNs:

$$L_{\mathrm{C}} \triangleq \mathbb{E}_{\boldsymbol{y}} \|\mathcal{M}_{\mathcal{R}}(\boldsymbol{z}^+) - \mathcal{M}_{\mathcal{N}}(\boldsymbol{z}^\perp)\|_2^2, \tag{31}$$

where $\boldsymbol{z}^\perp = \boldsymbol{P}_{\mathcal{N}} \mathcal{M}_{\mathcal{R}}(\boldsymbol{A}^\dagger \boldsymbol{y})$. In the ideal case, $L_{\mathcal{R}}$ and $L_{\mathcal{N}}$ approximate the supervised training in $\mathcal{R}(\boldsymbol{A}^\dagger)$ and $\mathcal{N}(\boldsymbol{A})$ respectively. Suppose the training under these two losses leads to perfect reconstruction in $\mathcal{R}(\boldsymbol{A}^\dagger)$ and $\mathcal{N}(\boldsymbol{A})$, *i.e.*

$$\mathcal{M}_{\mathcal{R}}(\boldsymbol{z}^+) = \boldsymbol{x} + \boldsymbol{u}, \boldsymbol{u} \in \mathcal{N}(\boldsymbol{A}), \tag{32}$$

$$\mathcal{M}_{\mathcal{N}}(\boldsymbol{z}^\perp) = \boldsymbol{x} + \boldsymbol{v}, \boldsymbol{v} \in \mathcal{R}(\boldsymbol{A}^\dagger). \tag{33}$$

Then we have that

$$\mathcal{M}_{\mathcal{R}}(\boldsymbol{z}^+) - \mathcal{M}_{\mathcal{N}}(\boldsymbol{z}^\perp) = \boldsymbol{u} - \boldsymbol{v}. \tag{34}$$

Minimizing (31) yields $\boldsymbol{u} - \boldsymbol{v} = \boldsymbol{0}$ and thus $\boldsymbol{u} = \boldsymbol{v} = \boldsymbol{0}$ due to $\boldsymbol{u} \perp \boldsymbol{v}$. In other words, the loss L_{C} allows the twin networks in SiamNet to interact with each other, so as to further reduce the reconstruction error during learning.

Inference Given a sample of measurements $\boldsymbol{y}$ for test, the image reconstruction is done by

$$\boldsymbol{x}^* = \boldsymbol{P}_{\mathcal{R}}\mathcal{M}_{\mathcal{R}}(\boldsymbol{A}^\dagger \boldsymbol{y}) + \boldsymbol{P}_{\mathcal{N}}\mathcal{M}_{\mathcal{N}}(\boldsymbol{P}_{\mathcal{N}}\mathcal{M}_{\mathcal{R}}(\boldsymbol{A}^\dagger \boldsymbol{y})). \quad (35)$$

That is, the SiamNet first reconstructs an image from $\boldsymbol{z}$ via $\mathcal{M}_{\mathcal{R}}$, then applies $\mathcal{M}_{\mathcal{R}}$ to refine the null-space component, and finally infers the image via fusing the range-space and null-space components from $\mathcal{M}_{\mathcal{R}}$ and $\mathcal{M}_{\mathcal{N}}$ respectively.

Remark 2. *Our approach is readily applicable to test-time model adaptation, rooting in the self-supervised nature of the loss function L_{All}. That is, as L_{All} can function without using the GT, it can be directly employed for model fine-tuning on a test sample.*

6 Performance Evaluation

The experiments are conducted with four imaging applications: CS-based acquisition for natural images with Gaussian measurements, CS-MRI (magnetic resonance imaging) with Fourier measurements, sparse-view computed tomography (CT) imaging with CT measurements, and hyperspectral (HS) imaging with snapshot measurements.

In the training of SiamNet, we initialize the ρ by 0.5, all convolution kernels by Xavier [50], and all biases by 0. The α, β in L_{All} of (24) are set to common values such that the magnitudes of $L_{\mathcal{R}}, L_{\mathcal{N}}, L_{\mathcal{C}}$ are at the same order. Concretely, they are fixed to $\alpha = 0.1, \beta = 0.05$ through all settings. The injected noise $\boldsymbol{n}'$ in $L_{\mathcal{R}}$ is drawn from the normal distribution $N(\boldsymbol{0}, \frac{2}{255}\boldsymbol{I})$ when the measurement noise is little. For Gaussian noise, it is drawn from the distribution of measurement noise. For Poisson noise, the injected noise is generated approximately by $\boldsymbol{n}' = \boldsymbol{y}' - \boldsymbol{y}, \boldsymbol{y}' \sim \mathrm{Poisson}_\gamma(\boldsymbol{y})$, which works fine empirically. The $\boldsymbol{n}_1', \boldsymbol{n}_2'$ in $L_{\mathcal{N}}$ are drawn from $N(\boldsymbol{0}, \frac{2}{255}\boldsymbol{I})$. The Adam optimizer is applied with learning rate 10^{-4}. Same as many existing works, for each measurement matrix we train an individual model of SiamNet. To simulate a realistic unsupervised learning setting, throughout all experiments, for each latent image of a given dataset, we only generate one sample of measurements for training.

The results of compared methods through all experiments are quoted from existing works whenever possible, or produced by officially released codes/models otherwise. For the closely-related work EI [11] and REI [46], we replaced its original U-Net with one of the twin networks in our SiamNet and retrained it using the same training samples as ours for better results. The default random shift is used as the EI transformations during its training. See supplemental material for more details on how the results of the compared methods are obtained. The reconstruction performance is evaluated by peak signal-to-noise ratio (PSNR) and structural similarity (SSIM).

6.1 Evaluation on CS-based Image Acquisition

Experimental settings of image acquisition vary in existing works. Following the setting of [12], [8], the sensing matrix applied to image blocks is generated by entry-wisely sampling from i.i.d. normal distribution with subsequent row-wise orthogonalization. The sampling ratio $r = \frac{d}{D}$ is set to $0.4, 0.1, 0.25$ respectively. The dataset of image blocks (33×33) from [12] is used to form the training samples of measurements. Both noiseless setting and noisy setting are considered. In the noisy case, the measurements within pixel range $[0, 255]$ for both training and test are contaminated by the additive Gaussian white noise drawn from $N(\boldsymbol{0}, \frac{10}{255})$. For SiamNet, we use 20 blocks and 500 epochs.

The BSD68 [51] and Set11 [8] are used as the test datasets. Each image on the datasets is cropped into non-overlapping blocks to generate measurements. The image blocks reconstructed from these measurements are concatenated back as an image for comparison to GT. Recently, there are some studies [18], [9], [15] using full images instead of image blocks for training, so as to reduce the block artifacts in reconstruction caused by non-overlapping block partition. For a fair comparison to those methods, we also train another model of SiamNet on the measurements of those full images of [18], with the same training strategy as before.

For comparison, we select DAMP [37], ISTANet+ [8], NNet [52], DPANet [15], SLPI [22] MACNet [9], AMP-Net [18], FSTANet [2], COAST [10], CS-DIP [47], BNN [30], L-SURE [25], EI [11], REI [46] and DDSSL [36]. Note that the DDSSL model without test-time adaption is used and the unrolled block number in its DNN is set to the same as ours for a fair comparison. The quantitative results are listed in

TABLE 1
Mean PSNR(dB)/SSIM of reconstructed images in natural image acquisition. The best results among all compared methods and among all unsupervised methods are **boldfaced** and underlined respectively.

	Method	BSD68			Set11		
		r = 0.4	r = 0.25	r = 0.1	r = 0.4	r = 0.25	r = 0.1
noiseless case	DAMP	28.17/.79	25.63/.70	21.94/.52	33.56/.93	28.46/.85	22.64/.60
	ISTANet+	32.17/.92	29.29/.85	25.29/.70	36.02/**.96**	32.44/.92	26.49/.80
	NNet	28.84/.85	26.42/.78	23.44/.64	29.51/.85	26.57/.78	22.99/.66
	DPANet*	31.33/.91	29.00/.83	25.97/.61	35.04/.95	31.74/.92	26.99/.84
	SLPI*	30.72/.88	28.27/.81	24.72/.66	33.73/.93	30.42/.89	25.02/.75
	MACNet*	31.47/.91	29.42/.85	25.80/.70	35.34/.95	32.91/.92	27.68/.82
	AMPNet*	32.81/.92	29.86/.86	25.33/.70	36.71/**.96**	32.90/**.93**	27.35/.82
	FISTANet	32.25/.92	29.18/.85	25.09/.69	36.24/.95	32.60/**.93**	26.94/.81
	COAST*	**32.93/.93**	**30.07/.87**	**26.28/.74**	**37.13/.96**	**33.85/.93**	**28.69/.86**
	CS-DIP	30.82/.87	27.87/.80	24.95/.69	33.44/.92	31.42/.91	27.34/.83
	BNN	31.28/.90	28.63/.84	25.24/.71	35.71/.95	32.30/.92	27.49/.83
	L-SURE	31.84/.90	28.86/.84	23.15/.65	33.19/.94	31.25/.90	24.92/.65
	EI	31.68/.90	28.42/.82	23.24/.63	35.46/.95	31.01/.90	22.74/.64
	DDSSL*	32.10/.80	29.12/.85	25.41/.70	35.89/.85	32.26/.92	26.80/.81
	SiamNet	32.48/.92	28.86/.83	25.93/.72	36.41/.95	32.42/.92	27.28/.83
	SiamNet*	32.68/.92	29.87/.86	26.00/.72	36.65/**.96**	33.22/**.93**	27.81/.84
noisy case	DAMP	26.55/.72	24.87/.65	21.70/.51	29.25/.86	26.35/.80	20.84/.58
	ISTANet+	28.98/.83	27.26/.77	23.86/.60	31.09/.89	29.20/.86	24.55/.70
	DPANet*	28.98/**.84**	27.24/.76	24.34/.63	30.01/.89	29.25/.86	25.21/.75
	SLPI*	28.47/.83	26.91/.75	24.25/.67	30.57/.89	28.71/.78	24.51/.71
	MACNet*	28.92/**.84**	27.57/**.78**	24.63/.65	30.34/.89	29.31/.86	25.56/.76
	AMPNet*	**29.12/.84**	27.51/**.78**	24.57/.64	31.11/.90	29.34/.86	25.31/.75
	FISTANet	28.53/.82	27.30/.77	24.34/.63	30.64/.88	29.37/.86	24.92/.74
	COAST*	28.98/.83	27.56/.77	24.86/.67	31.09/.89	29.50/**.87**	25.84/**.78**
	CS-DIP	25.24/.64	24.07/.59	22.46/.51	28.87/.83	27.36/.79	24.19/.68
	BNN	28.13/.81	26.47/.75	23.79/.64	30.39/.88	28.67/.84	25.23/.76
	L-SURE	27.60/.77	26.85/.73	23.55/.60	29.81/.84	28.35/.82	23.17/.64
	REI	28.04/.79	27.04/.75	22.32/.60	28.99/.81	28.08/.81	22.26/.66
	DDSSL*	28.37/.80	27.32/.76	24.72/.66	31.42/.90	29.19/.86	25.48/.74
	SiamNet	28.58/.81	27.39/.77	24.81/.67	31.14/.90	29.28/.86	25.85/**.78**
	SiamNet*	28.65/.82	**27.64/.78**	**24.91/.68**	**31.49/.91**	**29.52/.87**	**25.86/.78**

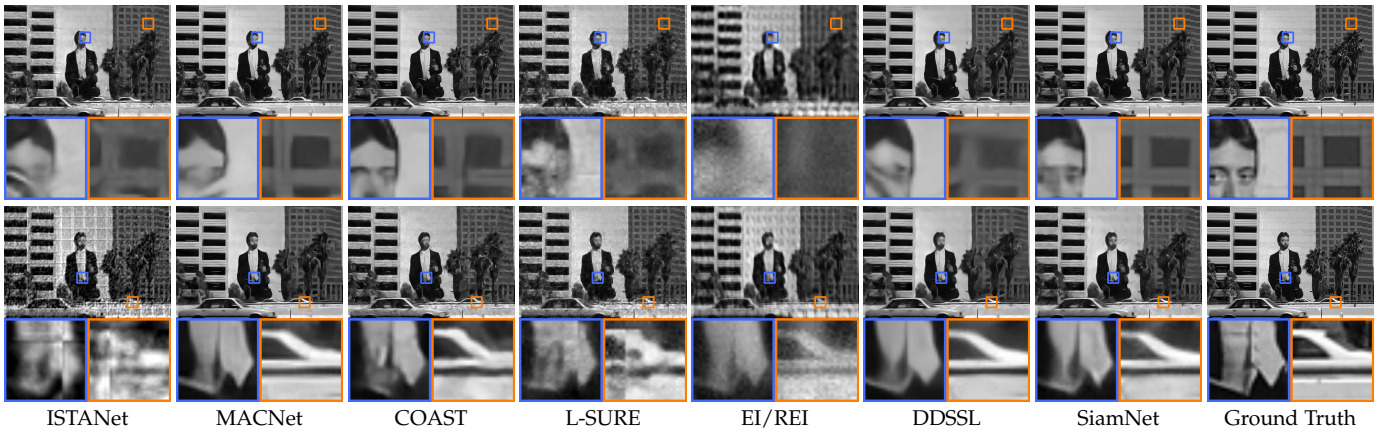


Fig. 4. Images reconstructed by selected methods on a sample image from BSD in natural image acquisition with sampling ratio 0.10 in the noiseless case (upper row) and noisy case (bottom row).

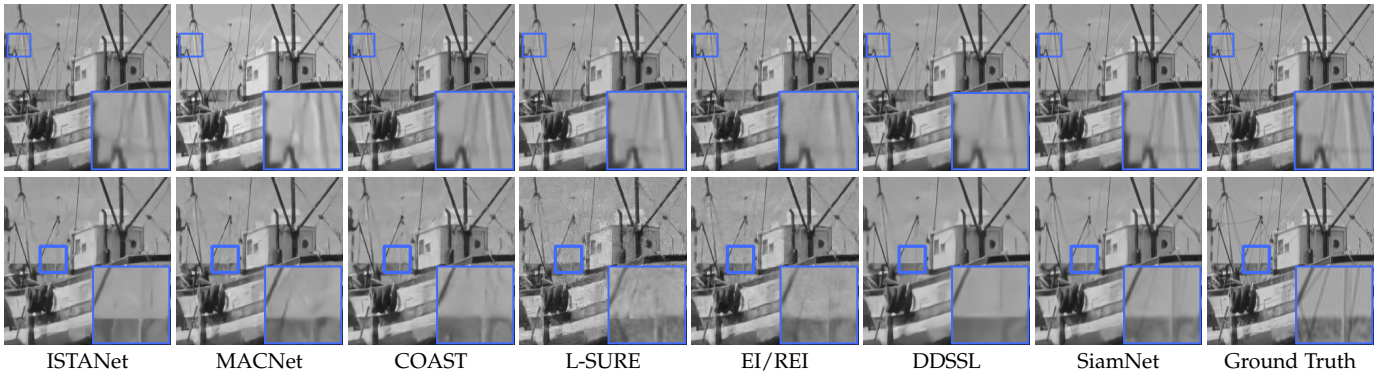


Fig. 5. Images reconstructed by selected methods on a sample image from Set11 in natural image acquisition with sampling ratio 0.25 in the noiseless case (upper row) and noisy case (bottom row).

Table 1, where we mark the models trained with full images by *. With the powerful data adaptivity of deep learning, SiamNet significantly outperformed DAMP, the only non-learning method among the compared ones.

The CS-DIP, BNN, L-SURE, EI/REI and DDSSL are unsupervised reconstruction methods. Both SiamNet and SimaNet* outperformed CS-DIP and L-SURE by a large margin, while outperforming BNN and EI/REI noticeably, through all the settings. The inferior performance of EI/REI may be caused by the insufficient equivariance existing in training images. Particularly, EI did not perform not well at the low sampling ratio $r = 0.1$, as the information of the image in range space is too little for EI/REI to train the null-space reconstruction. In addition, the PSNR advantage of SiamNet over EI/REI is enlarged for noisy data, which shows the effectiveness of our proposed noise-aware self-supervised loss effectiveness. Further, SiamNet* also outperformed DDSSL* across all settings, which demonstrated the improvement brought by the extensions in this paper.

In comparison to the supervised methods, SiamNet performed much better than some of them (*e.g.* NNet), and it competed well against those recent methods, *e.g.*, DPANet, MACNet, AMPNet and FISTNet. Compared with the top performer in supervised methods, COAST, there is a performance gap for SiamNet, but not dramatic, *e.g.*, no more than 0.3dB in PSNR in half cases. See also Fig. 5, 4 for the visual inspection on the reconstruction results of some selected methods. In summary, the results in natural image acquisition have demonstrated the effectiveness of SiamNet.

6.2 Evaluation on CS-MRI

This experiment adopts the setting used by [53], [30]. The sensing matrix is defined as the down-sampling on Fourier domain with some predefined down-sampling pattern. The measurement noise is synthesized by generating a Gaussian white noise in image domain which is then multiplied by the sensing matrix. Two types of down-sampling patterns of sampling ratio $r = 1/5, 1/4, 1/3$ respectively are used: Gaussian patterns and radial patterns. The MR images from Alzheimer's Disease Neuroimaging Initiative are used for both training and test. The standard deviation of the Gaussian noise is set to 10% of the maximum pixel value of the MR image. For SiamNet, we set the block number to 12 and the epoch number to 2000. The 300 samples of measurements generated from 300 MR images are used for unsupervised learning. The test is done on 21 MR images.

The proposed method is compared with SparseCS [7], SNALE [53], ADMMNet [14], DDN [41], MACNet [54], CS-DIP [47], BNN [30], EI [11] and DDSSL [36]. The DDSSL model without test-time adaption is used for a fair comparison. The quantitative results are listed in Table 2. Our SiamNet noticeably outperformed the non-learning method SparseSC and other unsupervised learning methods including CS-DIP, BNN, EI and DDSSL, across all settings.

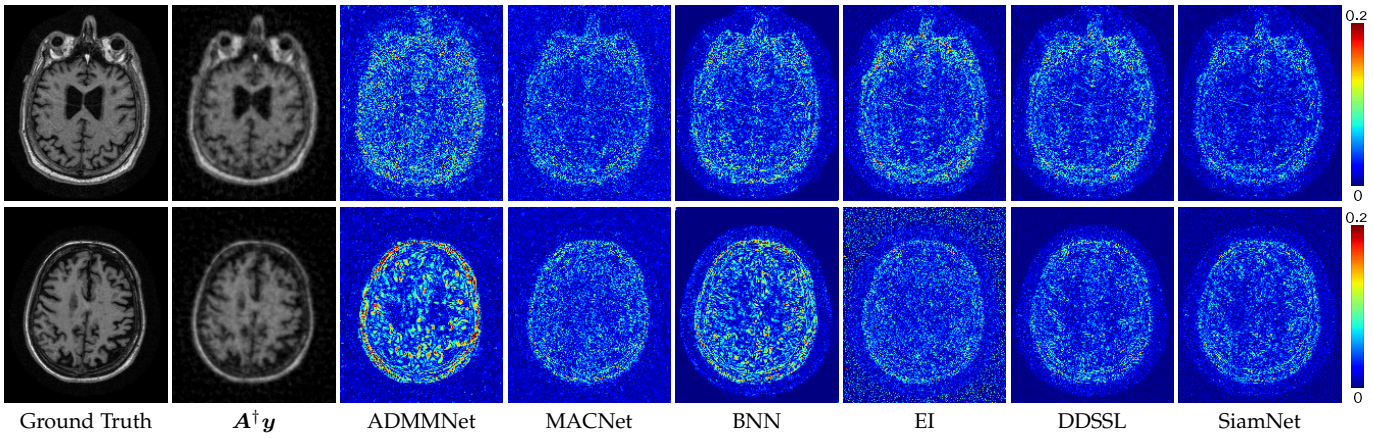


Fig. 6. Visualization of point-wise error maps of selected methods in MRI with a radial mask of sampling ratio 1/4 in the noisy case.

TABLE 2
Mean PSNR(dB)/SSIM of reconstructed images in MRI. The best results among all compared methods and among unsupervised methods are **boldfaced** and underlined respectively.

Method	Gaussian Pattern			Radial Pattern		
	$r = 1/5$	$r = 1/4$	$r = 1/3$	$r = 1/5$	$r = 1/4$	$r = 1/3$
SparseCS	24.97/.51	24.92/.49	24.91/.47	25.96/.61	26.50/.60	26.50/.60
SNALE	25.97/.67	26.15/.67	26.41/.62	25.98/.68	26.38/.66	26.70/.65
ADMMN.	25.42/.61	25.84/.60	26.14/.56	25.44/.73	25.96/.61	26.50/.60
DDN	29.94/.84	30.91/.84	31.31/.85	29.75/.82	30.92/.85	31.31/.85
MACNet	30.81/.89	30.92/.90	**31.74/.91**	30.07/.86	30.70/.85	**31.93/.92**
CS-DIP	23.90/.63	24.72/.72	25.53/.70	23.42/.63	24.89/.73	25.36/.75
BNN	29.17/.86	29.20/.86	29.46/.87	28.38/.84	29.47/.87	29.58/.86
EI	29.10/.71	29.82/.72	29.96/.72	29.67/.74	29.71/.75	30.12/.77
DDSSL	30.26/.85	30.65/.86	30.87/.86	30.16/.85	30.62/.86	30.66/.85
SiamNet	**30.94/.90**	**31.23/.91**	31.55/.90	**30.66/.89**	**30.96/.90**	31.78/.91

In comparison to the supervised learning-based methods, the SiamNet competes well against the top performer, *i.e.* MACNet, and even outperformed it on some settings. Such an advantage comes from both the effectiveness of our unsupervised training scheme and the cooperative architecture of SiamNet. See Fig. 6 for the visualization of point-wise reconstruction error maps of some selected methods. To conclude, both the quantitative and qualitative results have demonstrated the effectiveness of our method.

6.3 Evaluation on Sparse-View CT Imaging

The experiment follows the protocol of [46]. The sensing matrix $\boldsymbol{A}$ is defined based on a nonlinear forward model $\boldsymbol{A} = I_0 e^{-\text{randon}(\boldsymbol{x})}$ with Radon transform where $I_0 = 10^5$ denotes X-ray source intensity. The filtered back projection is used to approximate $\boldsymbol{A}^\dagger$. The evaluation is conducted with the CT100 dataset [55] which consists of the middle slice of 100 CT images of size 128×128 taken from 69 different patients. The measurements of 90 samples are used for training and the remaining ones are for test. We evaluate the noiseless case and noisy cases respectively. The setting of the noisy case follows [46]. Considering quantum and electronic noise are two major noise sources in X-ray CT scanners and for normal clinical exposures, the measurements are modeled as the sum of a Poisson distribution representing photon counting statistics and an independent Gaussian distribution representing additive electronic noise: $\boldsymbol{y} = \boldsymbol{z} + \boldsymbol{n}$ where $\boldsymbol{z} \sim \text{Poisson}(\frac{\boldsymbol{Ax}}{\gamma})$ and $\boldsymbol{n} \sim \mathcal{N}(\sigma \boldsymbol{I})$ with $\gamma = 1$ and $\sigma = 30$. The injected noise for $L_{\mathcal{R}}$ is also generated as a Poisson-Gaussian mixture. The hyper-parameters of SiamNet are set the same as that in MRI.

The proposed method is compared with FBP [56], FISTANet [2], CS-DIP [47], BNN [30], EI [11], REI [46] and DDSSL [36]. The DDSSL model without test-time adaption is used for a fair comparison. See Table 3 for a quantitative comparison. The SiamNet still leads to better reconstruction than the unsupervised learning-based methods CS-DIP, BNN, EI, REI and DDSSL in the noisy setting. In the noiseless setting, SiamNet performs slightly worse than DDSSL with a minor PSNR gap of 0.03dB, while much better than CS-DIP, BNN and REI. In addition, its performance is close to that of FISTANet, a supervised trained DNN. See Fig. 7 for the point-wise reconstruction error maps of the compared methods, where SiamNet achieved higher visual quality than other unsupervised learning-based methods.

TABLE 3
Mean PSNR (dB)/SSIM of reconstructed images in CT imaging. The best results among all compared methods and among unsupervised methods are **boldfaced** and underlined respectively.

Noisy	FISTANet	CS-DIP	BNN	REI	DDSSL	SiamNet
No	**41.75/.984**	38.94/.971	39.87/.973	40.64/.971	41.35/.979	41.32/.981
Yes	**36.15/.923**	28.17/.724	29.87/.764	34.82/.916	35.27/.919	35.60/.921

6.4 Evaluation on HS Imaging

We follow [16] for the experimental setting. The measurement matrix corresponds to a coded aperture snapshot spectral imaging system, capturing a three-dimensional HS image into a two-dimensional snapshot via mixing different wavelength signals modulated by a physical mask and a disperser. For a K-band HS image, the measurement matrix is composed by concatenating K shifted diagonal matrices: $\boldsymbol{A} = [\boldsymbol{S}_1(\text{diag}(\boldsymbol{a}_1)), ..., \boldsymbol{S}_K(\text{diag}(\boldsymbol{a}_K))]$, where $\boldsymbol{S}_K$ is a row-wise shifting operator with the offset determined by the system's dispersion characteristics, $\boldsymbol{a}_k(i) = C(\boldsymbol{p}_i, \lambda_k)h(\lambda_k)$ denotes the element corresponding to the spatial position $\boldsymbol{p}_i$ and wavelength λ_k, $C(\cdot, \cdot)$ represents the spatial-spectral

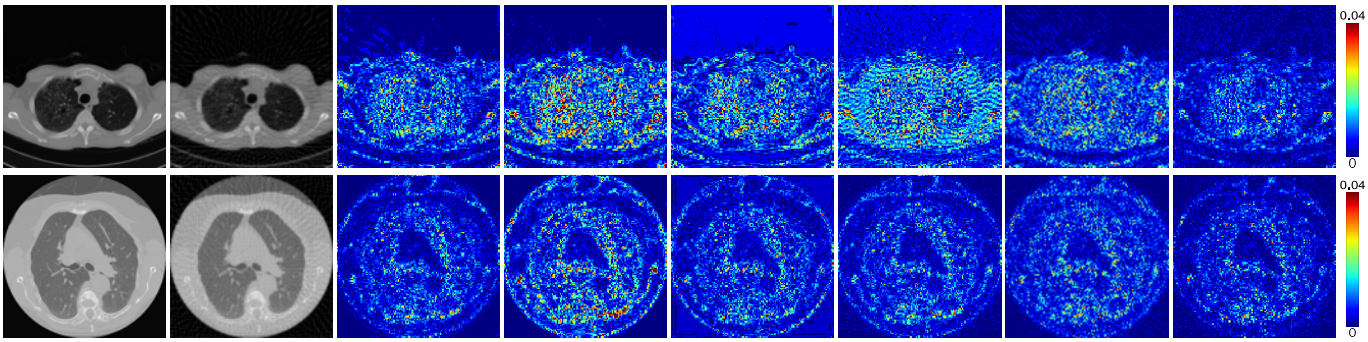


Fig. 7. Visualization of point-wise error maps of compared methods in CT imaging.

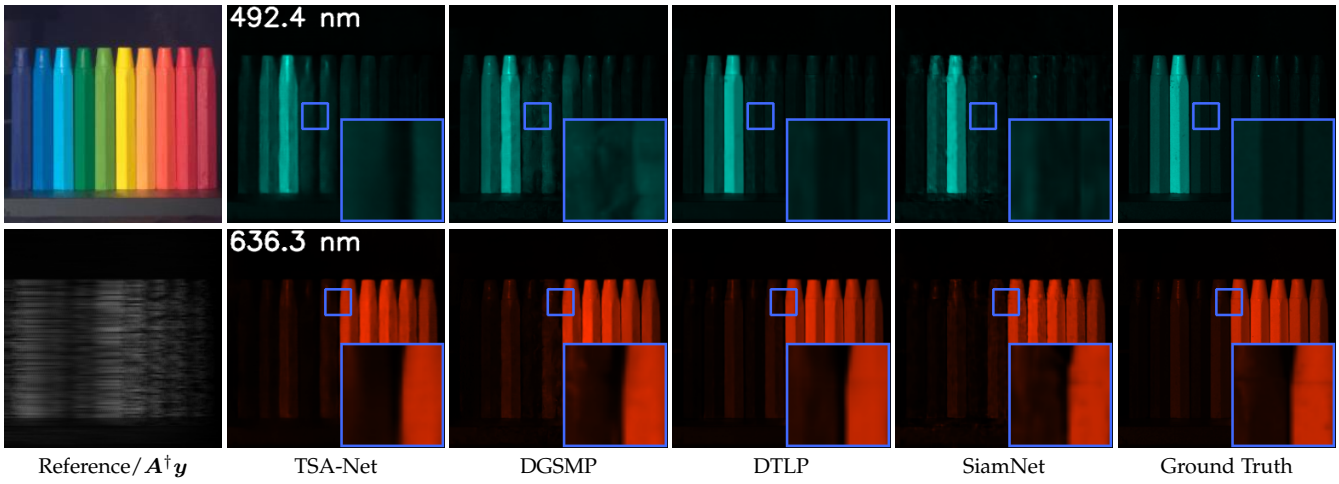


Fig. 8. Visualization of reconstructed HS images of selected competitive methods in hyperspectral imaging, in terms of 2/28 spectral channels.

modulation pattern determined by the coded aperture, and the function $h(\cdot)$ represents the spectral filter response function accounting for the spectral encoding in the system. The CAVE dataset [57] is used for training and the KAIST dataset [58] consisting of 10 scenes of spatial size 256×256 is used for test. Both the training and test data have 28 spectral bands ranging from 450nm to 650nm. For training SiamNet, we use the same hyper-parameters as that in CS-MRI.

The methods for the comparison with SiamNet include HSSP [59], DNU [3], TSANet [17], DGSMP [16], DTLP [60] and EI [11] and DDSSL [36]. The DDSSL model without test-time adaption is used for a fair comparison. The quantitative results are listed in Table 4. See also Fig. 8 for a visual inception of the reconstructed HS images. While the SiamNet has a certain performance gap from DTLP, the top-performing supervised DNN, it performs better than other supervised DNNs including HSSP, DNU, TSANet and DGSMP. In comparison to the unsupervised competitors EI and DDSSL, the SiamNet outperformed them noticeably. Particularly, DDSSL does not work well in this experiment, which may be due to the fact that its sign-flipping-based scheme failed to provide good samples from the distribution of prediction error for R2R in the full image space. In contrast, the N2N-inspired loss adopted in the proposed method does not require sampling the unknown distribution of prediction error. To conclude, all above results have demonstrated the effectiveness of the proposed method.

TABLE 4
Mean PSNR (in dB; upper row) and SSIM (bottom row) of reconstructed images in HS imaging. The best results among all compared methods and among unsupervised methods are **boldfaced** and underlined respectively.

HSSP	DNU	TSANet	DGSMP	DTLP	EI	DDSSL	SiamNet
30.35	30.74	31.46	32.63	**33.98**	30.91	30.15	<u>33.24</u>
0.85	0.86	0.89	0.92	**0.94**	0.87	0.82	<u>0.92</u>

6.5 Ablation Study

Loss function The following ablation studies on the loss function are conducted, with results summarized in Table 5.

1) We train the network $\mathcal{M}_\mathcal{R}$ using only the loss $L_\mathcal{R}$ and use its output for prediction. The resulting performance is noticeably worse, which verifies the benefit of the dual-space cooperative learning in SiamNet. Note that the resulting performance is not very bad. This is probably because the $\mathcal{M}_\mathcal{R}$ trained with $L_\mathcal{R}$ can predict well on $\mathcal{R}(\boldsymbol{A}^\dagger)$ and meanwhile there is certain implicit regularization effect from the network architecture so that the ambiguity caused by $\mathcal{N}(\boldsymbol{A})$ is moderately reduced.
2) We replace the loss $L_\mathcal{R}$ by the one similar to DIP [34], [35]: $L_\mathcal{R}^- = \|\boldsymbol{A}^\dagger\boldsymbol{y} - \mathcal{M}_\mathcal{R}(\boldsymbol{A}^\dagger\boldsymbol{y})\|_2^2$, and retrain SiamNet. The performance drops significantly, which is due to the reconstruction in $\mathcal{R}(\boldsymbol{A}^\dagger)$ becomes worse and it also

invalidates the training of $\mathcal{M}_{\mathcal{N}}$ via $L_{\mathcal{N}}$. Such results have demonstrated the effectiveness of $L_{\mathcal{R}}$.

3) We retrain the SiamNet by removing the loss L_{C}. The performance has a certain drop, which implies that L_{C} can moderately reduce solution ambiguity in training.
4) We retrain the SiamNet in a supervised manner: the networks $\mathcal{M}_{\mathcal{R}}, \mathcal{M}_{\mathcal{N}}$ are supervised by the range-space and null-space components of the GT image respectively, using the same paired data as those supervised methods. Only minor improvement is observed, which indicates the effectiveness of our unsupervised training scheme.

TABLE 5
Results of ablation studies on loss functions in terms of mean PSNR(dB), tested in MRI.

Method	Gaussian Pattern			Radial Pattern		
	$r = 1/5$	$r = 1/4$	$r = 1/3$	$r = 1/5$	$r = 1/4$	$r = 1/3$
Train with $L_{\mathcal{R}}$	29.98	30.25	30.48	29.64	30.49	30.91
Train with $L_{\mathcal{R}}^{-}$	20.61	20.77	20.87	19.71	19.83	19.98
Train w/o $L_{\mathcal{C}}$	30.76	31.04	31.33	30.18	30.74	31.39
Supervised	31.23	31.45	31.86	30.94	31.54	31.95
Original	30.94	31.23	31.55	30.66	30.96	31.78

DNN structure We also conduct ablation studies on the SiamNet's structure, with results summarized in Table 6.

1) We evaluate the performance of using the output of $\mathcal{M}_{\mathcal{R}}$ and $\mathcal{M}_{\mathcal{N}}$ respectively as the reconstruction result. A noticeable improvement is observed from the combination of two twin networks over the single one, which verified the effectiveness of our inference scheme for SiamNet.
2) We disable weight sharing between $\mathcal{M}_{\mathcal{R}}$ and $\mathcal{M}_{\mathcal{N}}$ and retrain the model. We can see that weight sharing leads to a remarkable performance improvement. Recall that the benefits of weight sharing are two-fold: reduction of parameter number and additional regularization. Both are important for resolving the possible overfitting in unsupervised learning. Note that both twin DNNs contain the learnable modules acting as proximal operators in the unrolled algorithm, whose roles are to remove artifacts from the image estimated in the previous stage. Thus, they may share image features/priors to benefit each other. As a result, weight sharing of the twin DNNs brings certain regularization and robustness.

TABLE 6
Results of ablation studies on network structure in terms of mean PSNR(dB), tested on natural image acquisition.

Method	$\sigma = 0$			$\sigma = 10$		
	$r = 0.4$	0.25	0.1	$r = 0.4$	0.25	0.1
Inference from $\mathcal{M}_{\mathcal{R}}$	32.23	29.43	25.81	28.16	27.34	24.43
Inference from $\mathcal{M}_{\mathcal{N}}$	31.66	29.12	25.73	28.03	27.16	24.22
Nonshared weights	31.01	28.12	24.98	28.14	27.21	24.37
Original	32.68	29.87	26.00	28.67	27.73	24.92

Reconstruction performance in $\mathcal{R}(\boldsymbol{A}^{\dagger})$ and $\mathcal{N}(\boldsymbol{A})$ By projecting the reconstructed images onto $\mathcal{R}(\boldsymbol{A}^{\dagger})$ and $\mathcal{N}(\boldsymbol{A})$ using $\mathcal{P}_{\mathcal{R}}$ and $\mathcal{P}_{\mathcal{N}}$ respectively, we assess the reconstruction performance in both the subspaces, in terms of mean squared error (MSE). Table 7 lists these results of SiamNet of its leading competitor DDSSL. As the measurements contain no image information in $\mathcal{N}(\boldsymbol{A})$, both methods exhibit higher MSE in $\mathcal{N}(\boldsymbol{A})$ than in $\mathcal{R}(\boldsymbol{A}^{\dagger})$. When the sampling ratio increases, the dimension of $\mathcal{N}(\boldsymbol{A})$ decreases, leading to decreased MSE in $\mathcal{N}(\boldsymbol{A})$. Conversely, the MSE in $\mathcal{R}(\boldsymbol{A}^{\dagger})$ increases due to the impact from the amplification of noise $\boldsymbol{A}^{\dagger}\boldsymbol{n}$. Importantly, the results indicate that SiamNet consistently produces less reconstruction error across both subspaces, demonstrating its advantage over DDSSL.

TABLE 7
Reconstruction MSE in $\mathcal{R}(\boldsymbol{A}^{\dagger})$ and $\mathcal{N}(\boldsymbol{A})$ in noisy CS-based natural image acquisition.

Method	MSE in $\mathcal{R}(\boldsymbol{A}^{\dagger})$			MSE in $\mathcal{N}(\boldsymbol{A})$		
	$r = 0.1$	$r = 0.25$	$r = 0.4$	$r = 0.1$	$r = 0.25$	$r = 0.4$
DDSSL	4.84	8.37	12.22	216.83	70.33	37.49
SiamNet	4.11	6.66	11.13	187.45	59.03	33.35

6.6 Complexity Comparison

The model size comparison in terms of the number of model parameters is given in Table 8 for different methods in natural image acquisition. The model size of our SiamNet is in the middle level among all the compared methods. Since we use EI with our network architecture, the model size of EI is the same as that of SiamNet. For CS-MRI, sparse-view CT imaging and HS imaging, the model sizes of SiamNet are the same, and thus we compare it with other models in all those three tasks together See Table 9 where our SiamNet shows good performance while maintaining a moderate number of parameters. See also Table 10 for the comparison on inference time. The time cost of our SiamNet is also in the middle level among all compared methods and is much less than BNN, an online self-supervised method.

TABLE 8
Number of parameters (Billion) of different models in CS-based natural image acquisition.

ISTANet+	NNet	DPANet*	SLPI	MACNet*	AMPNet*	FISTANet
0.33	7.75	9.32	0.66	0.47	0.58	0.08

COAST*	CS-DIP	BNN	L-SURE	DDSSL	EI	SiamNet
1.12	2.21	2.54	0.37	1.11	0.93	0.93

TABLE 9
Number of parameters (Billion) of different models in MRI, sparse-view CT imaging, and HS imaging.

DDN	MAC-Net	FISTANet	DNU	TSANet	DGSM
35.08	0.47	0.08	0.41	87.66	3.76

DTLP	CS-DIP	BNN	DDSSL	EI	SiamNet
3.19	2.21	2.54	0.52	0.44	0.44

6.7 Application to Test-Time Adaption

As mentioned in Remark 2, our approach is also applicable to test-time adaption, with a similar spirit to [36]. This is

TABLE 10
Per-sample inference time and PSNR on BSD68 with sampling ratio 0.25. The test is run on a single NVIDIA RTX 3090 GPU.

Metric	ISTANet+	DPANet	MACNet	COAST	BNN	L-SURE	SiamNet
Time(s)	0.026	0.043	0.164	0.050	1318	3.027	0.068
PSNR(dB)	29.29	29.00	29.42	30.07	28.63	28.86	29.87

done by model fine-tuning on each test sample via the loss L_{All} of (24) with the same hyper-parameters as those used in training. Table 11 shows the results of SiamNet with and without test-time adaption (denoted by "TA") in noisy setting respectively. The DDSSL is used for comparison, and we also report the results of both its adapted and non-adapted version. For a fair comparison, we set the iteration number of test-time adaption of SiamNet the same as that of DDSSL. We can see that the test-time adaption further improves the performance of SiamNet. Interestingly, the performance of SiamNet without adaption is already close to that of DDSSL with adaption. After adaption, SiamNet performs better than the adapted DDSSL. The overall improvement of SiamNet and DDSSL via test-time adaption is very close to each other, which is 0.44dB vs. 0.55dB in CS-MRI and 0.57dB vs. 0.46dB in natural image acquisition.

TABLE 11
PSNR(dB) comparison of DDSSL and SiamNet in MRI and natural image acquisition w/o or w/ test-time adaption, under the noisy settings.

Method	MRI (Radial Pattern)			CS acquisition on Set11		
	r=1/5	r=1/4	r=1/3	r=0.4	r=0.25	r=0.1
DDSSL	30.16	30.62	30.66	31.42	29.19	25.48
DDSSL-TA	30.44	31.04	31.68	32.05	29.53	26.13
SiamNet	30.66	30.96	31.78	31.49	29.52	25.86
SiamNet-TA	31.03	31.39	32.30	32.06	29.97	26.48

Figure 9 illustrates the PSNR increment against the number of iterations during test-time model adaptation. Notably, while SiamNet and DDSSL show similar PSNR increment trends, SiamNet exhibits a marginally faster speed. Moreover, in the early stage, the performance of SiamNet noticeably surpasses that of DDSSL. Remarkably, even without adaptation, SiamNet competes effectively with the adapted DDSSL. Moreover, as shown in Table 12, when both methods use the same number of iterations, the adaptation time required by SiamNet takes roughly 60% of that required by DDSSL. These results highlight the superior efficiency of SiamNet over DDSSL in test-time model adaptation.

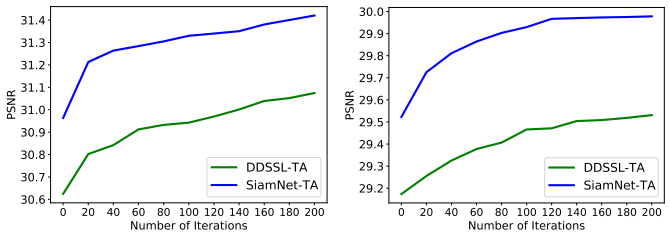


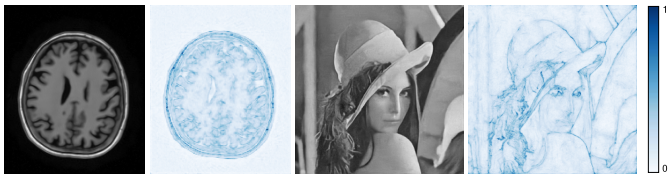
Fig. 9. Performance curves against the number of iterations during model adaptation for CS-MRI (left) and CS image acquisition (right).

TABLE 12
Per-sample adaption time (seconds) of DDSSL and SiamNet in two applications with noiseless settings.

Method	CS-MRI (Radial pattern)			CS-Acquisition (Set11)		
	r=1/5	r=1/4	r=1/3	r=0.4	r=0.25	r=0.1
DDSSL	14.4	14.5	14.7	51.2	51.5	51.6
SiamNet	8.5	8.6	8.6	31.4	31.8	31.7

6.8 Uncertainty Quantization

In both scientific and medical imaging, uncertainty quantification on reconstructed images is desired. With our approach, uncertainty quantification can be straightforwardly done by determining the point-wise standard deviation of the results reconstructed using randomized noise re-corruption. See Fig. 10 for two demonstrations in CS-MRI and CS-based image acquisition respectively. Initially, we produce 50 image estimates by re-corrupting the input by random noise in a manner consistent with our training process. Subsequently, we compute the uncertainty map using the pixel-wise standard deviation across these estimates. We can see that flatter areas typically show lower uncertainty, whereas regions with more-complex structures, such as edges and textures, exhibit higher uncertainty.



CS-MRI CS-based image acquisition

Fig. 10. Uncertainty map visualization on two cases. Left: CS-MRI with a sampling rate r=1/4 and radial pattern under noisy condition; Right: CS-based natural image acquisition with r=0.25 under noisy condition.

7 Conclusion

In contexts such as medical and scientific imaging, where GT images can be scarce, sub-optimal or compromised in quality, unsupervised learning emerges as a vital complementary approach to traditional deep supervised learning. This paper introduced an unsupervised deep learning methodology for image reconstruction from incomplete measurements. Our method leverages a self-supervised training paradigm on an unrolling-based Siamese DNN, addressing the inherent ill-posedness with two self-supervised loss functions and a mutual consistency loss. We validated the effectiveness and efficiency of our method across four imaging tasks. In future, we would like to extend our approach to handle unknown noise and other inverse problems, as well as combine it with supervised learning to develop semi-supervised image reconstruction methods.

References

[1] R. Liu, Y. Zhang, S. Cheng, X. Fan, and Z. Luo, "A theoretically guaranteed deep optimization framework for robust compressive sensing mri," in *Proceedings of AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 4368–4375.

[2] J. Xiang, Y. Dong, and Y. Yang, "Fista-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1329–1339, 2021.

[3] L. Wang, C. Sun, M. Zhang, Y. Fu, and H. Huang, "Dnu: deep non-local unrolling for computational spectral imaging," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1661–1671.

[4] Z. Meng, Z. Yu, K. Xu, and X. Yuan, "Self-supervised neural networks for spectral snapshot compressive imaging," *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[5] C. Deng, Y. Zhang, Y. Mao, J. Fan, J. Suo, Z. Zhang, and Q. Dai, "Sinusoidal sampling enhanced compressive camera for high speed imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 4, pp. 1380–1393, 2021.

[6] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[7] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[8] J. Zhang and B. Ghanem, "Ista-net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1828–1837.

[9] J. Chen, Y. Sun, Q. Liu, and R. Huang, "Learning memory augmented cascading network for compressed sensing of images," in *Proceedings of European Conference on Computer Vision*. Springer, 2020, pp. 513–529.

[10] D. You, J. Zhang, J. Xie, B. Chen, and S. Ma, "Coast: Controllable arbitrary-sampling network for compressive sensing," *IEEE Transactions on Image Processing*, vol. 30, pp. 6066–6080, 2021.

[11] D. Chen, J. Tachella, and M. E. Davies, "Equivariant imaging: Learning beyond the range space," in *Proceedings of IEEE/CVF International Conference on Computer Vision*, 2021.

[12] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 449–458.

[13] C. A. Metzler, A. Mousavi, and R. G. Baraniuk, "Learned d-amp: Principled neural network based compressive image recovery," *Proceedings of Annual Conference on Neural Information Processing Systems*, 2017.

[14] Y. Yang, J. Sun, H. Li, and Z. Xu, "Admm-csnet: A deep learning approach for image compressive sensing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 3, pp. 521–538, 2019.

[15] Y. Sun, J. Chen, Q. Liu, B. Liu, and G. Guo, "Dual-path attention network for compressed sensing image reconstruction," *IEEE Transactions on Image Processing*, vol. 29, pp. 9482–9495, 2020.

[16] T. Huang, W. Dong, X. Yuan, J. Wu, and G. Shi, "Deep gaussian scale mixture prior for spectral compressive imaging," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 16 216–16 225.

[17] Z. Meng, J. Ma, and X. Yuan, "End-to-end low cost compressive spectral imaging with spatial-spectral self-attention," in *Proceedings of European Conference on Computer Vision*. Springer, 2020, pp. 187–204.

[18] Z. Zhang, Y. Liu, J. Liu, F. Wen, and C. Zhu, "Amp-net: Denoising-based deep unfolding for compressive image sensing," *IEEE Transactions on Image Processing*, vol. 30, pp. 1487–1500, 2021.

[19] M. Kabkab, P. Samangouei, and R. Chellappa, "Task-aware compressed sensing with generative adversarial networks," in *Proceedings of AAAI Conference on Artificial Intelligence*, 2018.

[20] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, "Denoising prior driven deep neural network for image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 10, pp. 2305–2318, 2018.

[21] Y. Fu, T. Zhang, L. Wang, and H. Huang, "Coded hyperspectral image reconstruction using deep external and internal learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3404–3420, 2022.

[22] Z. Kadkhodaie and E. Simoncelli, "Stochastic solutions for linear inverse problems using the prior implicit in a denoiser," *Proceedings of Advances in Neural Information Processing Systems*, vol. 34, 2021.

[23] L. Zhaoqiang, L. Jiulong, S. Ghosh, H. Jun, and J. Scarlett, "Generative principal component analysis," *Proceeding In International Conference on Learning Representations*, vol. abs/2203.09693, 2021.

[24] J. Liu and Z. Liu, "Non-iterative recovery from nonlinear observations using generative models," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2022, pp. 233–243.

[25] C. Metzler, A. Mousavi, R. Heckel, and R. Baraniuk, "Unsupervised learning with stein's unbiased risk estimator," *arXiv preprint arXiv:1805.10531*, 2018.

[26] M. Zhussip, S. Soltanayev, and S. Chun, "Training deep learning based image denoisers from undersampled measurements without ground truth and without image prior," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 255–10 264.

[27] Z. Xia and A. Chakrabarti, "Training image estimators without image ground-truth," *Proceedings of Annual Conference on Neural Information Processing Systems*, pp. 2436–2446, 2019.

[28] E. K. Cole, J. M. Pauly, S. S. Vasanawala, and F. Ong, "Unsupervised mri reconstruction with generative adversarial networks," *arXiv preprint arXiv:2008.13065*, 2020.

[29] A. A. Hendriksen, D. M. Pelt, and K. J. Batenburg, "Noise2inverse: Self-supervised deep convolutional denoising for tomography," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1320–1335, 2020.

[30] T. Pang, Y. Quan, and H. Ji, "Self-supervised bayesian deep learning for image recovery with applications to compressive sensing," in *Proceedings of European Conference on Computer Vision*, 2020.

[31] Y. Sun, Y. Yang, Q. Liu, and M. Kankanhalli, "Unsupervised spatial-spectral network learning for hyperspectral compressive snapshot reconstruction," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[32] T. Pang, H. Zheng, Y. Quan, and H. Ji, "Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[33] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," in *Procedings of International Conference on Machine Learning*, 2018, pp. 2965–2974.

[34] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.

[35] S. Dittmer, T. Kluth, P. Maass, and D. Otero Baguer, "Regularization by architecture: A deep prior approach for inverse problems," *Journal of Mathematical Imaging and Vision*, vol. 62, 04 2020.

[36] Y. Quan, Q. Xinran, and H. Pang, Tongyaoand Ji, "Dual-domain self-supervised learning and model adaption for deep compressive imaging," in *Proceedings of European Conference on Computer Vision*, 2022.

[37] C. A. Metzler, A. Maleki, and R. Baraniuk, "From denoising to compressed sensing," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5117–5144, 2016.

[38] H. Zheng, F. Fang, and G. Zhang, "Cascaded dilated dense network with two-step data consistency for mri reconstruction," *Proceedings of Annual Conference on Neural Information Processing Systems*, 2019.

[39] Y. Wu, M. Rosca, and T. Lillicrap, "Deep compressed sensing," in *Procedings of International Conference on Machine Learning*. PMLR, 2019, pp. 6850–6860.

[40] D. You, J. Xie, and J. Zhang, "Ista-net¡sup¿++¡/sup¿: Flexible deep unfolding network for compressive sensing," in *IEEE International Conference on Multimedia and Expo*, 2021, pp. 1–6.

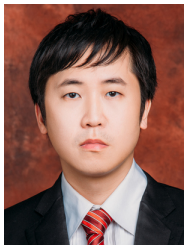
[41] D. Chen and M. E. Davies, "Deep decomposition learning for inverse imaging problems," in *Proceedings of European Conference on Computer Vision*, 2020.

[42] K. Xu, Z. Zhang, and F. Ren, "Lapran: A scalable laplacian pyramid reconstructive adversarial network for flexible compressive sensing reconstruction," in *Proceedings of European Conference on Computer Vision*, 2018, pp. 485–500.

[43] W. Shi, F. Jiang, S. Liu, and D. Zhao, "Scalable convolutional neural network for image compressed sensing," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 290–12 299.

[44] ——, "Image compressed sensing using convolutional neural network," *IEEE Transactions on Image Processing*, vol. 29, pp. 375–388, 2019.

[45] B. Zhou and S. K. Zhou, "Dudornet: Learning a dual-domain recurrent network for fast mri reconstruction with deep t1 prior," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4273–4282.
[46] D. Chen, J. Tachella, and M. E. Davies, "Robust equivariant imaging: a fully unsupervised framework for learning to image from noisy and partial measurements," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.
[47] D. Van Veen, A. Jalal, M. Soltanolkotabi, E. Price, S. Vishwanath, and A. G. Dimakis, "Compressed sensing with deep image prior and learned regularization," *arXiv preprint arXiv:1806.06438*, 2018.
[48] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr, "Fully-convolutional siamese networks for object tracking," in *Proceedings of European Conference on Computer Vision*. Springer, 2016, pp. 850–865.
[49] G. Koch, R. Zemel, R. Salakhutdinov *et al.*, "Siamese neural networks for one-shot image recognition," in *Proceedings of IEEE/CVF International Conference on Computer Vision Worshop*, vol. 2. Lille, 2015.
[50] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.
[51] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings of IEEE International Conference on Computer Vision*, vol. 2. IEEE, 2001, pp. 416–423.
[52] D. Gilton, G. Ongie, and R. Willett, "Neumann networks for linear inverse problems in imaging," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 328–343, 2019.
[53] J. Liu, T. Kuang, and X. Zhang, "Image reconstruction by splitting deep learning regularization from iterative inversion," in *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*. Springer, 2018, pp. 224–231.
[54] C.-M. Feng, Z. Yang, G. Chen, Y. Xu, and L. Shao, "Dual-octave convolution for accelerated parallel mr image reconstruction," in *Proceedings of AAAI Conference on Artificial Intelligence*, 2021.
[55] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle *et al.*, "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.
[56] K. H. Jin, M. T. McCann, E. Froustey, and M. A. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, pp. 4509–4522, 2017.
[57] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar, "Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum," *IEEE Transactions on Image Processing*, vol. 99, Mar 2010.
[58] I. Choi, M. Kim, D. Gutierrez, D. Jeon, and G. Nam, "High-quality hyperspectral reconstruction using a spectral prior," *ACM Transactions on Graphics*, vol. 36, no. 6, pp. 218:1–13, 2017.
[59] L. Wang, C. Sun, Y. Fu, M. H. Kim, and H. Huang, "Hyperspectral image reconstruction using a deep spatial-spectral prior," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8032–8041.
[60] S. Zhang, L. Wang, L. Zhang, and H. Huang, "Learning tensor low-rank prior for hyperspectral image reconstruction," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 12 006–12 015.



Yuhui Quan received the Ph.D. degree in Computer Science from South China University of Technology in 2013. He worked as the postdoctoral research fellow in Mathematics at National University of Singapore from 2013 to 2016. He is currently an associate professor in Computer Science at South China University of Technology. His research interests include image reconstruction, computational photography, and unsupervised learning.

Xinran Qin received the M.Sc degree in Computer Science from South China University of Technology in 2023. She is currently a Ph.D candidate in Computer Science. She is working on image reconstruction, compressed sensing, and unsupervised deep learning.

Tongyao Pang is currently a research fellow at the Department of Mathematics, National University of Singapore. She obtained her Ph.D. degree in Mathematics from National University of Singapore at 2019 and Bachelor degree from Peking University at 2014. Her research interests lie primarily in Bayesian statistics, deep learning and image processing.

Hui Ji received the Ph.D. degree in Computer Science from the University of Maryland, College Park in 2006, and joined National University of Singapore as a faculty member ever since. Currently, he is an associate professor in mathematics at National University of Singapore. His research interests include computational harmonic analysis, inverse problems, computational vision, and machine learning.