Ground-Truth Free Meta-Learning for Deep Compressive Sampling

Xinran Qin[1,2] Yuhui Quan[1,2*] Tongyao Pang[3] Hui Ji[3]
[1]School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China
[2] Pazhou Lab, Guangzhou 510335, China
[3] Department of Mathematics, National University of Singapore, 119076, Singapore
csqinxinran@gmail.com, csyhquan@scut.edu.cn, matpt@nus.edu.sg, matjh@nus.edu.sg [†]

Abstract

Compressive sampling (CS) is an efficient technique for imaging. This paper proposes a ground-truth (GT) free meta-learning method for CS, which leverages both external and internal deep learning for unsupervised high-quality image reconstruction. The proposed method first trains a deep neural network (NN) via external meta-learning using only CS measurements, and then efficiently adapts the trained model to a test sample for exploiting sample-specific internal characteristic for performance gain. The meta-learning and model adaptation are built on an improved Stein's unbiased risk estimator (iSURE) that provides efficient computation and effective guidance for accurate prediction in the range space of the adjoint of the measurement matrix. To improve the learning and adaption on the null space of the measurement matrix, a modified model-agnostic meta-learning scheme and a null-space consistency loss are proposed. In addition, a bias tuning scheme for unrolling NNs is introduced for further acceleration of model adaption. Experimental results have demonstrated that the proposed GT-free method performs well and can even compete with supervised methods.

1. Introduction

CS is an imaging technique that captures an image by collecting a limited number of measurements using a specific measurement matrix. This technique has a wide range of applications, including magnetic resonance (MR) imaging, computed tomography (CT), and many others. One main challenges in CS is reconstructing an image from its limited measurements. Let $\boldsymbol{x} \in \mathbb{R}^N$ denote an image of interest, $\boldsymbol{y} \in \mathbb{C}^M$ its CS measurement vector with $M \ll N$, and $\boldsymbol{\Phi}$ the measurement matrix. Then the CS reconstruction needs to solve an under-determined system:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{\epsilon} \in \mathbb{C}^M$ denotes measurement noise. As $\boldsymbol{\Phi}$ is under-determined, a direct inversion suffers from solution ambiguity and noise sensitivity.

Supervised learning is one popular approach for CSR, which trains an end-to-end NN for CSR over a dataset with pairs of a GT image and its measurements; see *e.g.* [10, 41, 44, 46]. However, supervised CSR has two limitations. Firstly, it is often expensive or even impractical to collect a sufficient number of GT images in practice. Consequently, the NN trained using a limited amount of external training data may not generalize well. Secondly, the training data may be biased or insufficient to cover the full range of patterns and characteristics of the images being reconstructed, resulting in poor generalization performance.

Recently, many works (*e.g.* [5, 6, 26, 49]) have sought to address the first limitation of supervised CSR by developing *GT-free external learning* techniques, which train an end-to-end NN without accessing GT images. Quite a few methods along this line are based on the *Stein's unbiased risk estimator* (SURE) [36]. While they have shown promise, there remains a noticeable performance gap between these GT-free techniques and existing supervised methods.

To address the second limitation of supervised CSR, there are also some works on *self-supervised internal learning*, which train a deep NN model directly on a test sample to exploit sample-specific internal characteristics; see *e.g.* [14, 29, 38]. These methods are based on the *deep image prior* (DIP) [37] induced by the structures of a convolutional NN (CNN). The downside of these methods is their computational efficiency. Since each test example requires learning an individual model, processing a large number of test samples can be time-consuming and overwhelming.

Motivated by the pros and cons of existing works on external and internal learning, this paper aims at developing a GT-free joint external and internal learning method for CSR that takes the full advantages of both:

*Corresponding author: Yuhui Quan

†This work was supported by Natural Science Foundation of Guangdong Province (Grant No. 2022A1515011755 and 2023A1515012841), and Singapore MOE AcRF Tier 1 Grant (WBS No. A-8000981-00-00).

- Same as existing external learning methods, it trains a model using an external dataset in an offline manner. The advantage is that the resulting model trained without GTs can compete well against its supervised counterparts.
- Same as existing internal learning methods, it is GT-free and its model is adaptive to test samples to reduce dataset bias. The advantage is that it is much more computationally efficient when processing many test samples.

This paper proposes a GT-free meta-learning method for CSR that achieves these goals. Meta-learning is about "learning to learn" which develops a learning model that can learn new concepts or fast adapt to new environments with efficient updates to an existing model. Different from existing works which access GT images for meta-learning (*e.g*. [20, 39]), we consider a GT-free environment where only measurement data is available during training. In the training stage, our proposed method trains a deep NN model so that not only it can reconstruct high-quality images, but also its weights are suitable for efficient test-time model adaption (*i.e*. internal learning). Then, in the test stage, the meta-trained model can be efficiently adapted to each test sample for further performance improvement.

Since no GT images are exposed to meta-learning and model adaptation, the key challenge is how to train the model to make accurate predictions solely based on measurement data. Motivated by SURE-based unsupervised learning, we propose a SURE-motivated self-supervised loss function, called iSURE ("i" for "improved"). Similar to SURE, iSURE provides an unbiased estimate of the mean squared error in $\mathrm{range}(\mathbf{\Phi}^{\mathrm{H}})$ (range space of adjoint of $\mathbf{\Phi}$), but using noisier input. The gradients derived from iSURE result in a more computationally efficient iteration scheme than existing SURE-based loss functions. Furthermore, the noise injection in iSURE also helps to alleviate potential overfitting during model training and adaptation.

Built upon the iSURE, an unsupervised training scheme is developed with the integration of *model-agnostic meta-learning* (MAML) [11]. MAML is a gradient-based meta-learning algorithm, which trains a model such that a small number of gradient updates on it will give a model to perform well on a new task. In our method, each target task is defined as the self-supervised reconstruction on a test sample via iSURE. Similar to other SURE variants, the iSURE defined on $\mathrm{range}(\mathbf{\Phi}^{\mathrm{H}})$ does not address the solution ambiguity caused by non-empty $\mathrm{null}(\mathbf{\Phi})$ (null space of $\mathbf{\Phi}$). Fortunately, the DIP from a CNN imposes implicit regularization on the output, partially addressing such ambiguity. To further reduce ambiguity, an ensemble-based sub-process is introduced into MAML. The basic idea is that the CNN with re-corrupted inputs in iSURE is likely to generate estimates with some degree of diversity on $\mathrm{null}(\mathbf{\Phi})$, and the ensemble of these estimates will refine the prediction in $\mathrm{null}(\mathbf{\Phi})$.

The adaption that only updates the gradients via iSURE ignores the influence on the prediction on $\mathrm{null}(\mathbf{\Phi})$. Thus, an additional loss on the prediction consistency in $\mathrm{null}(\mathbf{\Phi})$ between the original and adapted models is adopted, which provides guidance on reducing the solution ambiguity in $\mathrm{null}(\mathbf{\Phi})$ during adaption. To further accelerate the adaptation process, we use an unrolling CNN and propose a *bias-tuning* scheme that only adapts the bias parameters of the unrolling CNN. The motivation stems from the similarity between the bias parameters in an unrolling CNN and the threshold values in a shrinkage-based scheme for CSR.

To summarize, this paper proposes a GT-free meta-learning method for CSR, with the following contributions:

- An iSURE loss function for GT-free meta-learning and model adaption, which results in a more efficient and effective scheme for mitigating overfitting compared to existing SURE-based loss functions;
- A meta-learning process based on MAML and iSURE for learning CSR from external GT-less data, with an ensemble sub-process for improving null-space learning;
- A model adaption scheme using iSURE and null-space consistency, which exploits internal characteristics of a test sample for more accurate prediction;
- A bias-tuning scheme for accelerating the adaption of unrolling CNNs, with little reconstruction accuracy loss.

The experiments demonstrate that the proposed method outperforms GT-free learning methods and competes with supervised methods, while being faster than internal methods.

2. Related Work

Deep external learning for CSR Supervised methods use an external dataset to train a deep NN for CSR, with a focus on architecture design. One popular design is the physics-aware CNNs that incorporate CS models via deep unrolling, often implemented by replacing prior-related operations in an iterative algorithm using convolutional blocks, *e.g*., [10, 12, 15, 27, 40, 43–47]. To avoid collecting paired data in supervised learning, the plug-and-play methods (*e.g*. [17]) call some pre-trained denoiser(s) in a deep unrolling model, while the generative methods (*e.g*. [16, 22]) call generative models pre-trained on domain-specific images for estimation. Nevertheless, these methods still rely on latent images.

There is an increasing interest in GT-free external learning. Xia *et al*. [42] used measurement samples collected by two sensors to derive a self-supervised consistency loss. Working on single measurement samples, Metzler *et al*. [26] as well as Zhussip *et al*. [49] proposed to train with SURE-based denoisers. Chen *et al*. [5, 6] proposed the equivalent imaging framework which showed improvement over SURE-based methods. Quan *et al*. [33] proposed a joint self-supervised learning and model adaption scheme which leads to further improvement.

Deep internal learning for CSR Methods along this line express the target image by an untrained CNN fit on the test sample. Their foundation is DIP [37] that demonstrates the inductive biases of CNNs towards regular image structures over random patterns, with additional regularizations used, *e.g.*, the explicit ones via a specific CNN structure [14, 29] or the implicit ones via random sampling [38]. While internal learning shows promising performance by exploiting the characteristics of a test sample, its per-sample training is computationally expensive, which is addressed in our method by meta-learning and model adaption. Though model adaption has been investigated in [2, 13, 33], these works left a lot of room for improvement in both computational efficiency and prediction accuracy. Recently, for efficiency improvement, Mohan *et al*. [28] proposed the *gain-tuning*, re-weighting of feature maps, for fast adaption of denoising models. However, gain-tuning does not work well for our problem, and instead we propose a bias-tuning scheme which is more effective.

Meta-learning for image restoration and reconstruction There is increasing attention in applying meta-learning for fast model adaption in image recovery, and MAML is often used for its simplicity; see *e.g*. denoising [20], deblurring [7] and super-resolution [21, 31, 35]. One related work to ours is [39] that applies meta-learning for CSR of videos. All these works are supervised and need GTs for training.

SURE and its variants for image recovery reconstruction The gSURE [9] generalized the SURE as a regularization term in $\mathrm{range}(\boldsymbol{\Phi}^{\mathrm{H}})$ to solving inverse problems. The eSURE [50] extended SURE to paired noisy images for training a denoising NN. In the context of unsupervised deep learning, Metzler *et al*. [26] applied SURE to image denoising and CSR. The EnSURE [3] applies SURE with multiple measurement matrices for having an unbiased estimate beyond $\mathrm{range}(\boldsymbol{\Phi}^{\mathrm{H}})$ via ensemble. Quan *et al*. [33] revealed the connection between SURE and the R2R [30], a self-supervised loss for denoising and other image recovery tasks [32, 48]. Compared to gSURE, our iSURE additionally injects noise into input data, which leads to a more efficient iterative scheme. In comparison to EnSURE, our method only assumes a single measurement matrix, and we exploit noise injection for self-ensemble during learning.

3. Methodology

Consider a GT-free training dataset $\mathbb{D} := \{\boldsymbol{y}\}_{\boldsymbol{y}}$ with the measurements $\boldsymbol{y}$ generated by some measurement matrix $\boldsymbol{\Phi}$. Denote a CNN with weights $\boldsymbol{\omega}$ by

$$\mathcal{F}_{\boldsymbol{\omega}}(\cdot,\cdot) : (\mathbb{C}^M, \mathbb{C}^{M\times N}) \to \mathbb{R}^N, \tag{2}$$

which reconstructs latent images from their measurements and measurement matrix. Our proposed method has two steps: 1) Using unsupervised meta-learning to train $\mathcal{F}_{\boldsymbol{\omega}}$ on $\mathbb{D}$ for better test-time model adaptation; and 2) Running $\mathcal{F}_{\boldsymbol{\omega}}$ on all test samples with $\boldsymbol{\omega}$ fine-tuned (adapted) to each one.

3.1. iSURE Loss

Both the meta-learning and model adaption are based on a self-supervised loss motivated by SURE. For Gaussian white noise $\boldsymbol{\epsilon} \in \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$, a direct adoption of SURE (*e.g.* gSURE [9]) to our problem can be expressed by the loss

$$\begin{aligned}\ell^{\mathrm{SURE}}(\boldsymbol{\omega}; \boldsymbol{y}, \boldsymbol{\Phi}) := \|\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}, \boldsymbol{\Phi}) - \boldsymbol{y}\|_2^2 \\ + 2\sigma^2 \mathrm{tr}\Big(\boldsymbol{\Phi}(\partial\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}, \boldsymbol{\Phi})/\partial\boldsymbol{y})\Big),\end{aligned} \tag{3}$$

where $\mathrm{tr}(\cdot)$ denotes matrix trace and $\big(\partial\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}, \boldsymbol{\Phi})/\partial\boldsymbol{y}\big)_{ij} = \partial\big(\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}, \boldsymbol{\Phi})\big)_i/\partial\boldsymbol{y}_j$. The SURE loss involves a derivative term $\partial\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}, \boldsymbol{\Phi})/\partial\boldsymbol{y}$, which is approximated by the Monte-Carlo (MC) simulation [34] (termed MC-SURE) in exiting works, as there is no explicit form for a general NN $\mathcal{F}_{\boldsymbol{\omega}}$ with a complicated structure and it is quite expensive to compute the 2nd-order derivative $\partial\mathcal{F}^2_{\boldsymbol{\omega}}(\boldsymbol{y}, \boldsymbol{\Phi})/\partial\boldsymbol{y}\partial\boldsymbol{\omega}$ in back-propagation during training.

The iSURE loss we propose is defined as

$$\begin{aligned}\ell^{\mathrm{iSURE}}(\boldsymbol{\omega}; \boldsymbol{y}, \boldsymbol{\Phi}, \boldsymbol{\epsilon}') := \|\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y} + \boldsymbol{\epsilon}', \boldsymbol{\Phi}) - \boldsymbol{y}\|_2^2 \\ + 2\sigma^2 \mathrm{tr}\big(\boldsymbol{\Phi}(\partial\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y} + \boldsymbol{\epsilon}', \boldsymbol{\Phi})/\partial\boldsymbol{y})\big),\end{aligned} \tag{4}$$

where $\boldsymbol{\epsilon}'$ is (complex) Gaussian white noise independent from measurement noise $\boldsymbol{\epsilon}$. Under Gaussian white noise assumption on $\boldsymbol{\epsilon}$, the training with $\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\ell^{\mathrm{iSURE}}$ can be simplified into an efficient gradient update shown in Theorem 1 (with the proof given in our supplement).

Theorem 1. *Let $\boldsymbol{J}_{\boldsymbol{\omega}}$ be the Jacobian matrix w.r.t. $\boldsymbol{\omega}$,* i.e.*, $\boldsymbol{J}_{\boldsymbol{\omega}}\mathcal{F}_{\boldsymbol{\omega}} = \partial\mathcal{F}_{\boldsymbol{\omega}}/\partial\boldsymbol{\omega}$, and $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{\epsilon}$. Assume $\boldsymbol{\epsilon}, \boldsymbol{\epsilon}' \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$ are independent. Then, we have*

$$\begin{aligned}\nabla_{\boldsymbol{\omega}}\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\ell^{iSURE}(\boldsymbol{\omega}; \boldsymbol{y}, \boldsymbol{\Phi}, \boldsymbol{\epsilon}') = 2\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\big[\boldsymbol{J}^H_{\boldsymbol{\omega}}\big(\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y} + \boldsymbol{\epsilon}', \boldsymbol{\Phi})\big) \\ \boldsymbol{\Phi}^H\big(\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y} + \boldsymbol{\epsilon}', \boldsymbol{\Phi}) - \boldsymbol{y} + \boldsymbol{\epsilon}'\big)\big].\end{aligned} \tag{5}$$

Now the handling of $\partial\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}, \boldsymbol{\Phi})/\partial\boldsymbol{y}$ is avoided in (5), while the Jacobian term in (5) can be implemented with standard back prorogation. Similar to gSURE [9], iSURE is a weak form of SURE to provide accurate estimation in $\mathrm{range}(\boldsymbol{\Phi}^{\mathrm{H}})$. By the range-null space decomposition: $\mathbb{C}^N = \mathrm{range}(\boldsymbol{\Phi}^{\mathrm{H}}) \oplus \mathrm{null}(\boldsymbol{\Phi})$, the remaining is to recover the information in $\mathrm{null}(\boldsymbol{\Phi})$. It can be addressed to certain degree by the implicit regularization effect of a CNN structure [18, 37] and by the noise injection scheme in (4) which is a common practice in deep learning. The noise-injection scheme also allows us to design an ensemble-based subprocess to improve learning on $\mathrm{null}(\boldsymbol{\Phi})$ in the following.

3.2. Unsupervised Training with Meta-Learning

Before meta-learning is applied, the model $\mathcal{F}_{\boldsymbol{\omega}}$ is pre-trained on $\mathbb{D}$ with an iSURE-based loss for sufficient epochs

to achieve reasonable performance. This helps to improve the efficiency of the subsequent meta-learning and alleviate the possible instability of MAML. Based on MAML, we propose a more pertinent meta-learning scheme. Recall that MAML learns the optimal initial state of a model such that the base-learner can adapt to a new task effectively and rapidly. Following MAML, our meta-learning scheme includes an inner loop and an outer loop. At the beginning of each outer iteration, several task batches are sampled. The sequential inner loop updates model weights using gradient descent for each batch task individually. After the inner loop is completed, the losses of all tasks along with their updated weights are combined to perform a meta-gradient update on original model parameters. The complete training process is described in Algorithm 1.

In our MAML setting, each task is the self-supervised image reconstruction on a test sample based on the iSURE loss. Consider a specific task with the training samples $\mathbb{Y}_j$ drawn from $\mathbb{D}$, where $\mathbb{Y}_j$ contains only one sample theoretically but we use a small batch of samples in practice for faster convergence. The training loss is then given by

$$\mathcal{L}^{\text{iSURE}}_{\mathbb{Y}_j}(\boldsymbol{\omega}) := \mathbb{E}_{\boldsymbol{y}\in\mathbb{Y}_j,\boldsymbol{\epsilon}'\sim\mathcal{N}(\mathbf{0},\sigma^2\boldsymbol{I})}\,\ell^{\text{iSURE}}(\boldsymbol{\omega};\boldsymbol{y},\boldsymbol{\Phi},\boldsymbol{\epsilon}'). \quad (6)$$

At each inner loop, we first copy current parameters $\boldsymbol{\omega}$ to $\bar{\boldsymbol{\omega}}$ to form a detached model, and then update $\bar{\boldsymbol{\omega}}$ via $\mathcal{L}^{\text{iSURE}}_{\mathbb{Y}_j}$ for Q steps with a step size γ. At the end of the inner loop, we conduct one-step gradient descent with the loss:

$$\mathcal{L}^{\text{ensem}}_{\mathbb{Y}_j}(\boldsymbol{\omega};\bar{\boldsymbol{\omega}}) := \mathbb{E}_{\boldsymbol{y}\in\mathbb{Y}_j}\ell^{\text{ensem}}(\boldsymbol{\omega};\boldsymbol{y},\boldsymbol{\Phi},\bar{\boldsymbol{\omega}}), \quad (7)$$

where ℓ^{ensem} is the *ensemble loss* defined by

$$\|\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y},\boldsymbol{\Phi}) - \mathbb{E}_{\boldsymbol{\epsilon}''\sim\mathcal{N}(\mathbf{0},\sigma^2\boldsymbol{I})}(\mathcal{F}_{\bar{\boldsymbol{\omega}}}(\boldsymbol{y}+\boldsymbol{\epsilon}'',\boldsymbol{\Phi}))\|_2^2. \quad (8)$$

This step is additionally introduced to MAML for better addressing the ambiguity on $\text{null}(\boldsymbol{\Phi})$. The second term in (8) is the average of multiple estimates from $\mathcal{F}_{\bar{\boldsymbol{\omega}}}$ with noise-perturbed inputs, which is likely to cancel the prediction errors of these estimates to some degree. As a result, the loss (7) may improve the NN's prediction, particularly on $\text{null}(\boldsymbol{\Phi})$. Then the adapted parameters are calculated by

$$\boldsymbol{\omega}_j = \boldsymbol{\omega} - \alpha\nabla_{\boldsymbol{\omega}}\mathcal{L}^{\text{ensem}}_{\mathbb{Y}_j}(\boldsymbol{\omega};\bar{\boldsymbol{\omega}}), \quad (9)$$

where $\alpha\in\mathbb{R}^+$ controls the rate of inner update.

Then, the outer iteration updates model parameters by

$$\boldsymbol{\omega} \leftarrow \boldsymbol{\omega} - \beta\nabla_{\boldsymbol{\omega}}\sum\nolimits_j \mathcal{L}^{\text{iSURE}}_{\mathbb{Y}_j}(\boldsymbol{\omega}_j), \quad (10)$$

with $\beta\in\mathbb{R}^+$ controlling the rate of outer update. This step is to achieve minimal test errors for the tasks.

3.3. Testing with Nullspace-Consistent Adaptation

Let $\mathcal{F}_{\widetilde{\boldsymbol{\omega}}}$ denote the meta-learned model. One may adapt $\mathcal{F}_{\widetilde{\boldsymbol{\omega}}}$ to a test sample $\boldsymbol{y}^\star$ via minimizing $\mathcal{L}^{\text{iSURE}}$. However, as

Algorithm 1: GT-Free Meta-Learning for CSR

Input: Measurement dataset $\mathbb{D}$; Initial $\boldsymbol{\omega}$
Required: Learn_Rates α,β,γ; Inner_Iter_Num Q
1. Pre-training: update $\boldsymbol{\omega}$ with an iSURE-based loss
2. **while** not done **do**
3. Sample sets of measurements $\{\mathbb{Y}_j\}_j$ from $\mathbb{D}$
4. **for each** $\mathbb{Y}_j$ **do**
5. Initialize $\bar{\boldsymbol{\omega}}$ with $\boldsymbol{\omega}$
6. **for** $q=1,\cdots,Q$
7. $\bar{\boldsymbol{\omega}} \leftarrow \bar{\boldsymbol{\omega}} - \gamma\nabla_{\bar{\boldsymbol{\omega}}}\mathcal{L}^{\text{iSURE}}_{\mathbb{Y}_j}(\bar{\boldsymbol{\omega}})$
8. **end**
9. $\boldsymbol{\omega}_j = \boldsymbol{\omega} - \alpha\nabla_{\boldsymbol{\omega}}\mathcal{L}^{\text{ensem}}_{\mathbb{Y}_j}(\boldsymbol{\omega};\bar{\boldsymbol{\omega}})$
10. **end**
11. Update $\boldsymbol{\omega} \leftarrow \boldsymbol{\omega} - \beta\nabla_{\boldsymbol{\omega}}\sum_j \mathcal{L}^{\text{iSURE}}_{\mathbb{Y}_j}(\boldsymbol{\omega}_j)$
13. end

Algorithm 2: Testing with Model Adaption

Input: Measurements $\boldsymbol{y}^\star$; Meta-trained $\widetilde{\boldsymbol{\omega}}$
Required: Learn_Rate γ; Epoch_Num S
1. Initialize $\boldsymbol{\omega}^\star$ with $\widetilde{\boldsymbol{\omega}}$
2. **for** $s=1,\cdots,S$
3. $\boldsymbol{\omega}^\star \leftarrow \boldsymbol{\omega}^\star - \nabla_{\boldsymbol{\omega}^\star}(\gamma\mathcal{L}^{\text{iSURE}}_{\{\boldsymbol{y}^\star\}}(\boldsymbol{\omega}^\star) + \lambda\mathcal{L}^{\text{NC}}_{\boldsymbol{y}^\star}(\boldsymbol{\omega}^\star;\widetilde{\boldsymbol{\omega}}))$
4. **end**
5. Compute $\boldsymbol{x}^\star = \mathbb{E}_{\boldsymbol{\epsilon}'\sim\mathcal{N}(\mathbf{0},\sigma^2\boldsymbol{I})}\mathcal{F}_{\boldsymbol{\omega}^\star}(\boldsymbol{y}^\star+\boldsymbol{\epsilon}',\boldsymbol{\Phi})$

$\mathcal{L}^{\text{iSURE}}$ only considers the predition w.r.t. $\text{range}(\boldsymbol{\Phi}^{\text{H}})$, such an adaption may bring negative effects to the reconstruction w.r.t. $\text{null}(\boldsymbol{\Phi})$. Therefore, we propose a *null-space consistency* (NC) loss to mitigate those effects, which measures the prediction difference on $\text{null}(\boldsymbol{\Phi})$ between the original and adapted models:

$$\mathcal{L}^{\text{NC}}_{\boldsymbol{y}^\star}(\boldsymbol{\omega};\boldsymbol{\Phi},\widetilde{\boldsymbol{\omega}}) := \|\boldsymbol{P}_{\text{null}(\boldsymbol{\Phi})}(\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}^\star,\boldsymbol{\Phi}) - \mathcal{F}_{\widetilde{\boldsymbol{\omega}}}(\boldsymbol{y}^\star,\boldsymbol{\Phi}))\|_2^2, \quad (11)$$

where $\boldsymbol{P}_{\text{null}(\boldsymbol{\Phi})} = \boldsymbol{I} - \boldsymbol{\Phi}^\dagger\boldsymbol{\Phi}$ is the null-space projection with the pseudo inverse $\boldsymbol{\Phi}^\dagger$. Concretely, we first initialize $\boldsymbol{\omega}^\star$ by $\widetilde{\boldsymbol{\omega}}$ and then adapt the model to $\boldsymbol{y}^\star$ using $\mathcal{L}^{\text{iSURE}}_{\{\boldsymbol{y}^\star\}}$ and $\mathcal{L}^{\text{NC}}_{\boldsymbol{y}^\star}$ with the weights $\gamma,\lambda\in\mathbb{R}^+$, leading to the gradient update:

$$\boldsymbol{\omega}^\star \leftarrow \boldsymbol{\omega}^\star - \gamma\nabla_{\boldsymbol{\omega}^\star}\mathcal{L}^{\text{iSURE}}_{\{\boldsymbol{y}^\star\}}(\boldsymbol{\omega}^\star) - \lambda\nabla_{\boldsymbol{\omega}^\star}\mathcal{L}^{\text{NC}}_{\boldsymbol{y}^\star}(\boldsymbol{\omega}^\star,\widetilde{\boldsymbol{\omega}}), \quad (12)$$

for S steps, where S is a much smaller number than that in training. It can be seen that the above scheme is expected to adapt the model w.r.t. $\text{range}(\boldsymbol{\Phi}^{\text{H}})$ while maintaining the prediction accuracy on $\text{null}(\boldsymbol{\Phi})$ by pulling the prediction back. When the noise level is unknown, we estimate σ in iSURE with $\|\mathcal{F}_{\widetilde{\boldsymbol{\omega}}}(\boldsymbol{y}^\star,\boldsymbol{\Phi}) - \boldsymbol{y}^\star\|$ from the pre-trained model.

To consist with the iSURE training and the sub-task implied in $\mathcal{L}^{\text{ensem}}_{\mathbb{Y}}$, we adopt an ensemble inference scheme:

$$\boldsymbol{x}^\star := \mathbb{E}_{\boldsymbol{\epsilon}'\sim\mathcal{N}(\mathbf{0},\sigma^2\boldsymbol{I})}\mathcal{F}_{\boldsymbol{\omega}^\star}(\boldsymbol{y}^\star+\boldsymbol{\epsilon}',\boldsymbol{\Phi}). \quad (13)$$

Noise	CS Ratio	Non-Learning		Unsupervised	Internal		Unsupervised + Internal		Supervised	
		ZF	SparseMRI	REI	BNN	ASGLD	DDSSL	MetaCS	ADMMNet	Supervised
w/o	50%	37.07/.89	39.93/.94	41.15/.95	40.24/.93	41.60/.95	42.53/.97	**43.92/.98**	43.00/.97	44.09/.98
	40%	35.14/.85	38.51/.93	39.06/.95	38.63/.91	39.29/.95	41.00/.96	**41.24/.97**	41.56/.96	42.52/.98
	30%	33.01/.80	37.72/.91	37.84/.92	37.14/.89	38.11/.93	38.47/.94	**39.59/.95**	39.84/.93	40.70/.94
	20%	30.41/.72	35.46/.89	36.07/.90	35.54/.88	36.08/.90	36.73/.92	**37.12/.93**	37.17/.93	37.58/.94
w/	50%	30.87/.66	31.54/.67	34.91/.90	34.69/.89	34.51/.89	35.35/.91	**36.23/.93**	35.71/.91	36.44/.93
	40%	30.81/.68	31.24/.69	34.45/.90	33.32/.86	33.71/.87	34.83/.90	**35.53/.91**	35.31/.91	35.64/.91
	30%	30.39/.67	30.55/.68	33.64/.90	31.86/.85	32.25/.87	34.51/.91	**35.25/.92**	34.82/.90	35.18/.91
	20%	29.18/.64	29.74/.64	33.04/.88	31.29/.87	32.16/.88	33.79/.90	**34.34/.91**	34.04/.89	34.32/.90

Table 1. Mean PSNR(dB)/SSIM of MR image reconstruction on MRI150 dataset. **Boldfaced**: best of all compared **GT-free** methods.

Noise	CS Ratio	Non-Learning		Unsupervised	Internal		Unsupervised + Internal		Supervised	
		ZF	SparseMRI	REI	BNN	ASGLD	DDSSL	MetaCS	ADMMNet	Supervised
w/o	1/3	28.20/.72	34.58/.92	36.02/.94	35.58/.94	35.79/.94	36.40/**.95**	**36.54/.95**	35.31/.94	37.08/.97
	1/4	26.23/.65	32.31/.90	33.15/.91	34.08/.95	34.38/.95	34.66/.95	**34.85/.96**	33.70/.93	35.67/.98
	1/5	25.00/.61	30.72/.86	32.34/.90	32.28/.92	32.40/.92	33.83/**.95**	**34.08/.95**	32.32/.92	35.26/.96
w/	1/3	27.38/.66	29.32/.69	30.15/.74	29.58/.82	29.85/.84	31.68/.86	**31.79/.91**	30.94/.83	31.68/.91
	1/4	25.82/.62	25.93/.65	29.75/.75	29.47/.87	29.61/.87	31.04/.88	**31.17/.90**	29.95/.82	31.02/.90
	1/5	24.75/.58	24.88/.59	29.67/.77	28.38/.84	28.45/.84	30.44/**.87**	**30.56/.87**	28.81/.81	30.47/.87

Table 2. Mean PSNR(dB)/SSIM of MR image reconstruction on ADNI dataset. **Boldfaced**: best of all compared **GT-free** methods.

It is actually done by averaging the predictions w.r.t. different instances of $\boldsymbol{\epsilon}'$ (only 3 instances suffice in practice). See Algorithm 2 for a description of the whole test process.

3.4. Unrolling CNN with Bias-Tuning

Similar to [8], the physics-aware CNN for meta-learning is constructed by unrolling the proximal gradient descent of the regularized optimization problem: $\min_{\boldsymbol{x}} \ \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \psi(\boldsymbol{x})$. The unrolling leads to the iteration: $\boldsymbol{x}^{(k)} = \text{prox}_{\psi}^{\rho}(\boldsymbol{x}^{(k-1)} - \rho\boldsymbol{\Phi}^{\text{H}}(\boldsymbol{\Phi}\boldsymbol{x}^{(k-1)} - \boldsymbol{y}))$, where $\text{prox}_{\psi}^{\rho}(\boldsymbol{x}) := \text{argmin}_{\boldsymbol{z}} \frac{1}{2\rho}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 + \psi(\boldsymbol{z})$ is the proximal operator of the prior-inducing term ψ. The CNN $\mathcal{F}_{\boldsymbol{w}}$ is then constructed by replacing $\text{prox}_{\psi}^{\rho}$ with a standard residual convolutional block. It contains K blocks in total. Each block acts as one iteration and contains 6 convolutional layers; see our supplement for an illustration. The learnable parameters $\boldsymbol{\omega}$ then consists of the rate ρ (initialized to 0.5), as well as the weights (initialized by Xavier) and biases (initialized to 0) inside all convolutional layers.

We propose to perform adaption only on the bias parameters of the unfolding CNN, with other parameters frozen. The rational is, viewing the unrolling CNN as an advanced iterative shrinkage process, the biases play a similar role as the thresholds of shrinkage. In traditional iterative shrinkage approaches for CSR, such thresholds are critical to the performance while the filters can be fixed (*e.g.* wavelet filters), and their optimal values vary much across different samples. Thus, adapting the biases suffice to yield good performance. Since the number of biases is much less than the number of weights of convolutional kernels, our bias tuning scheme can further accelerate the adaption process.

4. Experiments

Out proposed method is named MetaCS and evaluated on reconstruction of natural images and MR images. Its configuration is as follows: 1) NN: $K = 10$; 2) Training: Adam is called with 500 epochs and 200 epochs respectively for pre-training, with a fixed learning rate of 10^{-4}, and then called with 50 epochs for meta-learning, where $\alpha, \beta, \gamma = 10^{-4}, \lambda = 5 \times 10^{-5}, Q = 50$; 3) Adaption: $S = 200$; and 4) Ensemble: only three instances are used.

4.1. MR Image Reconstruction

Two datasets are used for evaluation: MRI150 [46] and ADNI [23]. The MRI150 dataset contains 100 MR images for training and 50 test, where the k-space measurements are taken by applying a fixed radial mask to the Fourier coefficients with sampling ratio $r = 20\%, 30\%, 40\%, 50\%$ respectively. The ADNI dataset contains 300 MR images for training and 21 for test, where radial and Gaussian masks of ratios $r = \frac{1}{5}, \frac{1}{4}, \frac{1}{3}$ are applied respectively. Two cases of noise are considered on both datasets: the noiseless case with $\boldsymbol{\epsilon} = \boldsymbol{0}$ and the noisy case with $\boldsymbol{\epsilon} = \boldsymbol{\Phi}(\boldsymbol{\epsilon}_1 + i\boldsymbol{\epsilon}_2)$ where $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2 \sim \mathcal{N}(\boldsymbol{0}, (0.1\max(\boldsymbol{x}))^2\boldsymbol{I})$. In the noiseless case, we set $\sigma = 2.55$ for iSURE. We train models separately for different settings and different datasets, and each image is called only once to generate measurements. The MetaCS is compared with some GT-free methods of different kinds, including ZF [4], SparseMRI [25], REI [6], BNN [29], ASGLD [38], and DDSSL [33]. For DDSSL which is also a

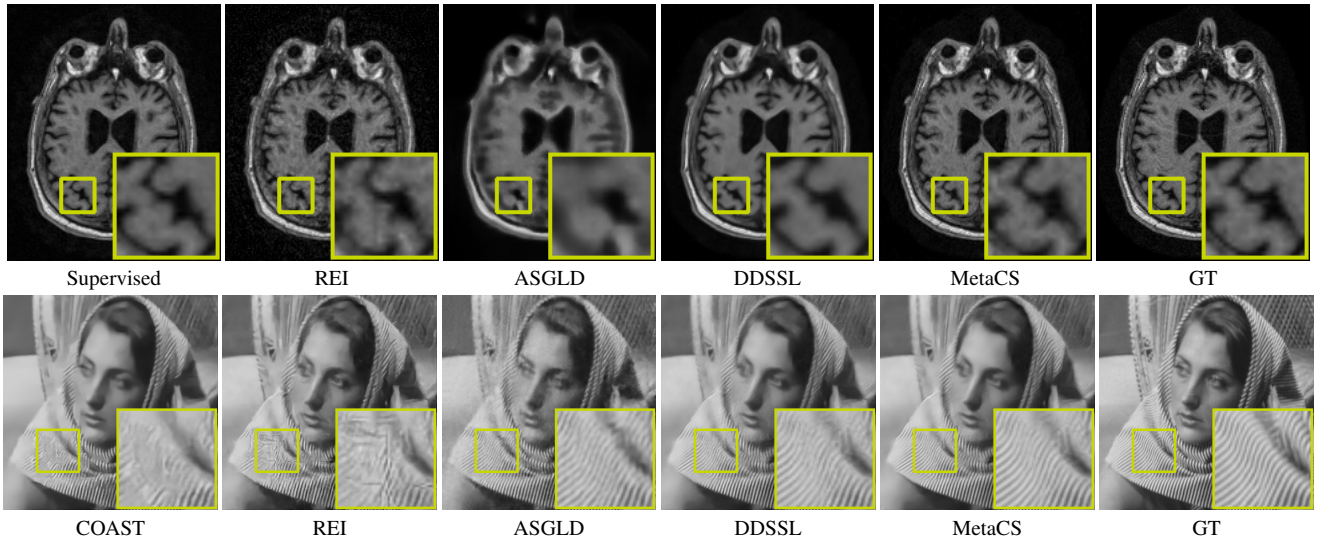


Figure 1. Visual inspection on reconstructed image. Top: reconstructed MR images with radial mask and sampling ratio 1/4 in noisy setting. Bottom: reconstructed natural images reconstruction from Gaussian measurements of sampling ratio 25% in noisy setting.

Noise	CS Ratio	Non-Learning	Unsupervised		Internal		Unsupervised+Internal		Supervised		
		TVAL3	LSURE	REI	BNN	ASGLD	DDSSL	MetaCS	Supervised	COAST	SSLIP
w/o	40%	31.21/.75	33.30/95	35.63/.95	35.71/.95	35.87/.95	37.18/**.96**	**37.25/.96**	36.02/.96	36.94/.96	33.73/.93
	25%	27.63/.62	31.31/.90	31.11/.90	32.30/.92	33.06/.92	33.28/**.94**	**33.52/.94**	32.44/.92	33.85/.94	30.42/.89
	10%	22.45/.38	25.00/.65	22.79/.64	27.49/.83	**28.15**/.83	27.57/**.84**	28.02 **/.84**	26.94/.82	28.34/.84	25.02/.75
	1%	14.90/.06	-	15.16/.32	16.69/.40	17.32/.41	17.37/.41	**17.63/.42**	17.48/.42	17.57/.42	15.72/.41
w/	40%	26.66/.72	28.73/.81	28.99/.81	30.39/**.88**	31.11/.90	31.58/**.88**	**31.64/.88**	31.14/.89	31.16/.89	30.58/.89
	25%	24.75/.67	28.14/.82	28.08/.81	28.67/.84	29.35/.85	29.61/**.87**	**29.71/.87**	29.49/.86	29.37/.86	28.71/.85
	10%	21.02/.54	23.31/.64	22.26/.66	25.23/.76	26.06/.76	26.10/**.78**	**26.17/.78**	25.03/.70	25.81/.78	24.48/.73
	1%	13.23/.05	-	14.75/.29	16.03/.32	16.74/.33	16.88/.34	**17.01/.35**	16.86/.34	16.92/.35	15.21/.35

Table 3. Mean PSNR(dB)/SSIM of natural image reconstruction on Set11 dataset. **Boldfaced**: best of all compared **GT-free** methods.

joint GT-free learning and model adaption method, we reduce its block number to have a similar model size as our CNN for a fair comparison. In addition, we introduce two supervised methods for the performance reference. One is ADMMNet [44] and the other is the supervised counterpart of MetaCS trained with the same data but with GTs.

The quantitative results on the two datasets are listed in Table 1 and Table 2, respectively, with some reconstructed images shown in Figure 1 and our supplement. It can be seen that on both datasets, MetaCS consistently outperforms both the internal learning methods including BNN and ASGLD and the recent GT-free external learning method REI. The PSNR gain by MetaCS is noticeable overall and significant in some settings. Such impressive performance is due to that MetaCS simultaneously exploits external knowledge of training datasets and internal properties of test samples. MetaCS also shows superior performance over DDSSL, the latest joint GT-free learning and adaption method. While GT-free methods are not expected to beat supervised methods, the results also show that MetaCS competes against its supervised counterpart in many settings. Our MetaCS even outperforms the supervised ADMMNet in many cases, which is probably due to the use of meta-learning and the difference in NN structures.

4.2. Natural Image Reconstruction

Following [19,46], the measurements are taken by a row-orthogonalized block Gaussian matrix $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ with the ratio $r = M/N$ set to $40\%, 25\%, 10\%, 1\%$, respectively. Both noiseless ($\boldsymbol{\epsilon}$=$\mathbf{0}$) and noisy ($\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, 10^2\boldsymbol{I})$) cases are employed. We train an individual model for each setting. The 88912 image blocks of size 33×33 (*i.e.* $N = 1089$) provided by [19] are used to generate 88912 measurement samples for unsupervised meta-learning. The Set11 [46] dataset is used for test, each image of which is cropped into non-overlapping blocks to generate measurements. In addition to BNN, AGSLD, REI and DDSSL, our MetaCS is compared with LSURE [26], another GT-free learning. We also introduce COAST [45], SSLIP [17], and the supervised counterpart of MetaCS for performance reference.

The results in terms of PSNR and SSIM are summarized in Table 3. Some reconstructed images are shown in Fig-

ure 1. Compared to BNN and AGSLD, two internal learning methods, MetaCS performs noticeably better in almost all cases. Compared to LSURE and REI, two unsupervised external learning methods trained on the same dataset as ours, MetaCS also shows superior performance. Compared to DDSSL, the latest model adaption method, MetaCS outperforms it in all settings in terms of PSNR. Again, MetaCS remains competitive to its supervised counterpart and even outperforms other supervised methods in some settings. All these results have demonstrated the benefits of joint external and internal learning via the meta-learning of MetaCS.

4.3. Ablation Studies

We conduct ablation studies on the ADNI dataset and present the results in Table 4. First, we study the effectiveness of the meta-learning by disabling it and training the NN until convergence. We can see that the meta-learning always brings performance gain which varies in different settings. Figure 2 displays the overall PSNR trace over iterations in the adaptation procedure w/ and w/o meta-learning. The curves indicated by w/ meta-learning are always above and increase faster than those indicated by w/o meta-learning, which implies that the meta-learning leads to faster and more effective model adaptation. Note that the meta-learning has already yielded a better model at the beginning, as our modified MAML benefits not only the adaption but also the unsupervised training. We further skip the adaptation process in the test phase (see "w/o Adaption"). Obviously, the adaptation brings noticeable improvement.

Next, we study the effectiveness of the components of meta-learning in MetaCS. A standard MAML is used to replace our modified one. We can see that the modified MAML leads to higher PSNR gain than the standard one. The standard MAML even performs a bit worse than "w/o Meta-Learning" in some cases. This is because the standard MAML does not address the learning ambiguity in $\mathrm{null}(\mathbf{\Phi})$ well. Further, we replace the iSURE loss with the gSURE-related loss (3) solved by MC approximation, and replace the ensemble inference by a single inference without noise injection. As a result, a significant PSNR drop is observed.

Last, the effectiveness of the ingredients in the adaption process is studied. We disable the NC loss during adaption by setting $\lambda = 0$, which leads to a certain PSNR decrease. We also use all model parameters beyond the biases for meta-learning and adaption; see "All weights". The resulting PSNR performance is slightly higher or lower than that purely adjusting biases for adaption. This indicates our bias tuning scheme suffices for a good adaption. Further, following gain-tuning [28], we conduct adaption only on the weights of feature maps. The resulting PSNR gain is not big and much less than that only using bias-tuning.

Regarding the inference process, the ensemble does not contribute a big part to the performance (not shown in Table 4). However, the ensemble inference that provides a number of instances can be utilized for uncertainty quantization of CSR, a feature welcomed in scientific imaging and medical imaging. See our supplement for more details. To conclude, the ablation studies have demonstrated the effectiveness of each key component in MetaCS.

	Method	w/o Noise			w/ Noise		
		r = 1/5	1/4	1/3	r = 1/5	1/4	1/3
Training	iSure → gSURE	29.13	32.57	34.29	28.56	29.17	29.14
	w/o Meta-Learning	32.93	33.97	35.68	29.52	30.03	30.34
	Standard MAML	32.86	34.01	35.73	29.63	30.08	30.42
Test	w/o Adaption	33.63	34.40	35.89	29.60	30.78	31.36
	w/o NC	33.71	34.65	36.40	30.41	31.01	31.66
	All weights	34.02	34.82	36.58	30.60	31.19	31.74
	Gain-tuning	33.68	34.58	36.19	30.35	30.83	31.37
	MetaCS	34.00	34.85	36.54	30.56	31.17	31.79

Table 4. PSNR(dB) results in ablation studies on ADNI dataset with radial masks.

4.4. More Analysis

Computational complexity See Table 5 for the running time of MetaCS in natural image reconstruction, with the comparison to four GT-free methods. With an additional adaptation process in the test phase, MetaCS is slower than the SURE-based external learning method LSURE. However, the time cost caused by adaptation is acceptable, compared to the internal learning methods BNN and ASGLD. While BNN and ASGLD show superior performance to LSURE in previous experiments, their time cost is very high due to millions of iterations. Our MetaCS is much faster than BNN and ASGLD while providing overall superior performance. Compared to the adaption method DDSSL, our MetaCS takes only around $1/7$ time, which benefits from the acceleration by meta-learning and bias tuning. The comparison regarding number of model parameters is given in Table 6. It can be seen that our model is relatively small, and the bias parameters are much less than that of the whole model, which thus leads to noticeable acceleration.

CS Ratio	LSURE	BNN	ASGLD	DDSSL	MetaCS
40%	0.45	282.54	178.01	5.72	0.89
25%	0.44	294.32	178.32	5.67	0.87
10%	0.44	298.93	189.40	5.50	0.86

Table 5. Running time (minutes) of different methods for reconstructing the images in Set11, tested on a TITAN RTX GPU.

COAST	SSLIP	LSURE	ASGLD	DDSSL	MetaCS / Bias-only
1.12	0.67	0.38	2.19	0.67	0.3756 / 0.0013

Table 6. Comparison in number of model parameters (M).

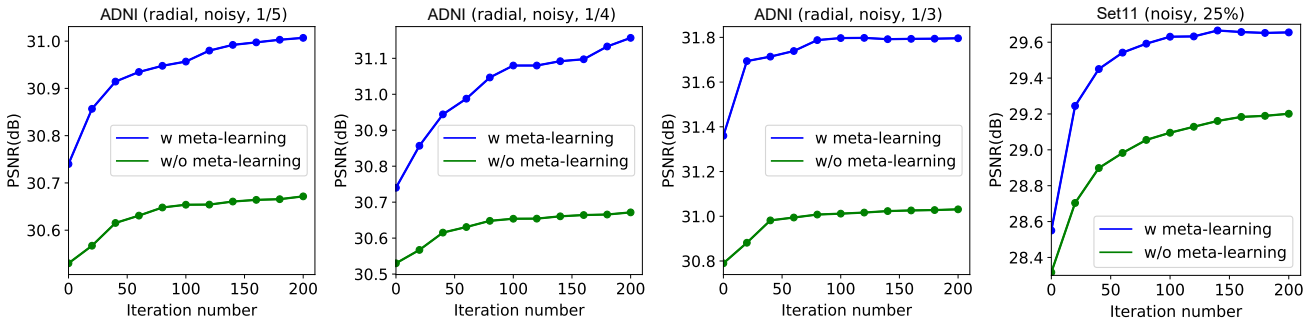


Figure 2. Performance curves vs. iteration number during adaptation on ANDI dataset (left three) and Set11 dataset (last one).

Adaption across different measurement matrices We also evaluate the performance of MetaCS for the case where the measurement matrix differs between training and test. Table 7 presents the results on the ADNI dataset, with the comparison to REI. We can see that when the performance of both methods drops in dealing with an unseen mask. However, the PSNR drop of MetaCS is much smaller than REI. Particularly, when training on a Gaussian mask of ratio 1/4 and testing on a Gaussian mask of ratio 1/3, the performance of MetaCS is nearly unchanged. This is attributed to the effectiveness of the meta-learning and model adaption. Also note that with some PSNR drop, MetaCS is still very competitive against the supervised methods in Table 2.

Test Mask	Training = Test		Gauss, r=1/4		PSNR Drop	
	REI	MetaCS	REI	MetaCS	REI	MetaCS
Gauss, r=1/3	36.72	37.87	36.38	37.80	-0.34	-0.07
Gauss, r=1/5	34.04	35.52	33.63	35.47	-0.41	-0.05
Radial, r=1/4	33.15	34.46	32.62	34.84	-0.53	-0.01

Table 7. PSNR(dB) comparison in adaption across different measurement matrices in noiseless setting.

Joint training across different noise levels We train a single model across different noise levels for natural images, with noise level randomly drawn from $[1, 10]$. We re-train REI and DDSSL for comparison and test on three noise levels. See Table 8 for the comparison. All the methods have a performance drop compared to their individually-trained versions, as shown in the case of noise level 10. For all the three tested noise levels, our MetaCS performs the best.

Method	Joint training			Individual training
	$\sigma = 1$	5	10	10
REI	30.29/.86	28.16/.82	27.45/.80	28.08/.81
DDSSL	31.24/.89	30.73/.88	29.29/.86	29.61/**.87**
MetaCS	**31.89/.91**	**31.66/.90**	**29.64/.87**	**29.71/.87**

Table 8. PSNR(dB)/SSIM on Set11 (25%) in joint training.

Handling unknown noise types in real scenarios Follow [24], we conduct performance evaluation using the subset (500/25 for training/test) of the MICCAI2013 dataset [1] with Rician noise to simulate unknown noise types in real scenarios. See Table 9 for the results and comparison. Our MetaCS outperforms other GT-free methods.

REI	AGLSD	DDSSL	MetaCS	ADMMNet	Supervised
27.32/.74	27.18/.73	28.55/.83	**28.79/.84**	27.81/.81	29.42/.86

Table 9. PSNR(dB)/SSIM results in CS-MRI with an unknown noise type. **Boldfaced**: best of all compared **GT-free** methods.

Performance of using a larger model We enlarge the model of MetaCS by using K=20 blocks and compare its performance to the original DDSSL model. See our supplement for the results. Our MetaCS still performs better in both reconstruction accuracy and complexity.

5. Conclusion and Discussion

This paper developed an effective GT-free meta-learning method for CSR, which can leverage external measurement data for unsupervised end-to-end NN training as well as exploit internal characteristics of test samples via model adaption for performance gain. The meta-learning was achieved through the iSURE loss for efficient range-space learning and the modified MAML algorithm for improved null-space learning. The model adaption was achieved by the null-space consistency loss for better null-space prediction and the bias-tuning scheme on an unrolling CNN for further acceleration. The effectiveness of the proposed method has been demonstrated in extensive experiments on two tasks. The techniques developed above can be extended to solving other ill-posed image reconstruction problems, which will be investigated in our future work.

One limitation of our proposed method is that it assumes the availability of the statistical characteristics of measurement noise. Indeed, this limitation also exists in existing GT-free methods, both external and internal, and even some supervised ones. Our future work will also investigate how to improve GT-free meta-learning and model adaption for data with unknown noise distributions.

References

[1] Miccai 2013 grand challenge dataset. https://wiki.cancerimagingarchive.net/display/public/nci-miccai+2013+grand+challenges+in+image+segmentation, 2013. 8

[2] Hemant Kumar Aggarwal and Mathews Jacob. Model adaptation for image reconstruction using generalized stein's unbiased risk estimator. *arXiv preprint arXiv:2102.00047*, 2021. 3

[3] Hemant Kumar Aggarwal, Aniket Pramanik, and Mathews Jacob. Ensure: Ensemble stein's unbiased risk estimator for unsupervised learning. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1160–1164. IEEE, 2021. 3

[4] Matt A Bernstein, Sean B Fain, and Stephen J Riederer. Effect of windowing and zero-filled reconstruction of mri data on spatial resolution and acquisition strategy. *Journal of Magnetic Resonance Imaging*, 14(3):270–280, 2001. 5

[5] Dongdong Chen, Julián Tachella, and Mike E Davies. Equivariant imaging: Learning beyond the range space. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 1, 2

[6] Dongdong Chen, Julián Tachella, and Mike E Davies. Robust equivariant imaging: a fully unsupervised framework for learning to image from noisy and partial measurements. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 5647–5656, 2022. 1, 2, 5

[7] Zhixiang Chi, Yang Wang, Yuanhao Yu, and Jin Tang. Test-time fast adaptation for dynamic scene deblurring via meta-auxiliary learning. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 9137–9146, 2021. 3

[8] Sören Dittmer, Tobias Kluth, Peter Maass, and Daniel Otero Baguer. Regularization by architecture: A deep prior approach for inverse problems. *Journal of Mathematical Imaging and Vision*, 62, 04 2020. 5

[9] Yonina C Eldar. Generalized SURE for exponential families: Applications to regularization. *IEEE Transactions on Signal Processing*, 57(2):471–481, 2008. 3

[10] Chun-Mei Feng, Zhanyuan Yang, Geng Chen, Yong Xu, and Ling Shao. Dual-octave convolution for accelerated parallel mr image reconstruction. In *Proceedings of AAAI Conference on Artificial Intelligence*, 2021. 1, 2

[11] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Procedings of International Conference on Machine Learning*, pages 1126–1135. PMLR, 2017. 2

[12] Davis Gilton, Greg Ongie, and Rebecca Willett. Neumann networks for linear inverse problems in imaging. *IEEE Transactions on Computational Imaging*, 6:328–343, 2019. 2

[13] Davis Gilton, Gregory Ongie, and Rebecca Willett. Model adaptation for inverse problems in imaging. *IEEE Transactions on Computational Imaging*, 7:661–674, 2021. 3

[14] Reinhard Heckel. Regularizing linear inverse problems with convolutional neural networks. *Advances in Neural Information Processing Systems Worshop*, 2019. 1, 3

[15] Kyong Hwan Jin, Michael T. McCann, Emmanuel Froustey, and Michael A. Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transacitons on Image Processing*, 26:4509–4522, 2017. 2

[16] Maya Kabkab, Pouya Samangouei, and Rama Chellappa. Task-aware compressed sensing with generative adversarial networks. In *Proceedings of AAAI Conference on Artificial Intelligence*, volume 32, 2018. 2

[17] Zahra Kadkhodaie and Eero Simoncelli. Stochastic solutions for linear inverse problems using the prior implicit in a denoiser. *Advances in Neural Information Processing Systems*, 34:13242–13254, 2021. 2, 6

[18] Hirofumi Kobayashi, Ahmet Can Solak, Joshua Batson, and Loic A Royer. Image deconvolution via noise-tolerant self-supervised inversion. *arXiv preprint arXiv:2006.06156*, 2020. 3

[19] 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*, pages 449–458, 2016. 6

[20] Seunghwan Lee, Donghyeon Cho, Jiwon Kim, and Tae Hyun Kim. Self-supervised fast adaptation for denoising via meta-learning. *arXiv preprint arXiv:2001.02899*, 2020. 2, 3

[21] Suyoung Lee, Myungsub Choi, and Kyoung Mu Lee. Dynavsr: Dynamic adaptive blind video super-resolution. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2093–2102, 2021. 3

[22] Zhongnian Li, Tao Zhang, Peng Wan, and Daoqiang Zhang. Segan: structure-enhanced generative adversarial network for compressed sensing mri reconstruction. In *Proceedings of AAAI Conference on Artificial Intelligence*, volume 33, pages 1012–1019, 2019. 2

[23] J. Liu, T. Kuang, and X. Zhang. Image reconstruction by splitting deep learning regularization from iterative inversion. In *Procedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 224–231. Springer, 2018. 5

[24] Risheng Liu, Yuxi Zhang, Shichao Cheng, Xin Fan, and Zhongxuan Luo. A theoretically guaranteed deep optimization framework for robust compressive sensing mri. In *Proceedings of AAAI Conference on Artificial Intelligence*, volume 33, pages 4368–4375, 2019. 8

[25] Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007. 5

[26] C. Metzler, A. Mousavi, R. Heckel, and R. Baraniuk. Unsupervised learning with stein's unbiased risk estimator. *arXiv preprint arXiv:1805.10531*, 2018. 1, 2, 3, 6

[27] Christopher A Metzler, Ali Mousavi, and Richard G Baraniuk. Learned d-amp: Principled neural network based compressive image recovery. *Advances in Neural Information Processing Systems*, 2017. 2

[28] Sreyas Mohan, Joshua L Vincent, Ramon Manzorro, Peter Crozier, Carlos Fernandez-Granda, and Eero Simoncelli. Adaptive denoising via gaintuning. *Advances in Neural In-*

formation Processing Systems, 34:23727–23740, 2021. 3, 7

[29] Tongyao Pang, Yuhui Quan, and Hui Ji. Self-supervised bayesian deep learning for image recovery with applications to compressive sensing. In *Proceedings of European Conference on Computer Vision*, 2020. 1, 3, 5

[30] Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. Recorrupted-to-recorrupted: unsupervised deep learning for image denoising. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2043–2052, 2021. 3

[31] Seobin Park, Jinsu Yoo, Donghyeon Cho, Jiwon Kim, and Tae Hyun Kim. Fast adaptation to super-resolution networks via meta-learning. *Proceedings of European Conference on Computer Vision*, 5, 2020. 3

[32] Yuhui Quan, Zhile Chen, Tongyao Pang, and Hui Ji. Unsupervised deep learning for phase retrieval via teacher-student distillation. In *Proceedings of AAAI Conference on Artificial Intelligence*, 2023. 3

[33] Yuhui Quan, Qin Xinran, Tongyao Pang, and Hui Ji. Dual-domain self-supervised learning and model adaption for deep compressive imaging. In *Proceedings of European Conference on Computer Vision*, 2022. 2, 3, 5

[34] S. Ramani, T. Blu, and M. Unser. Monte-carlo sure: A black-box optimization of regularization parameters for general denoising algorithms. *IEEE Transacitons on Image Processing*, 17(9):1540–1554, 2008. 3

[35] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 3516–3525, 2020. 3

[36] Charles M Stein. Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, pages 1135–1151, 1981. 1

[37] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018. 1, 3

[38] Weixi Wang, Ji Li, and Hui Ji. Self-supervised deep image restoration via adaptive stochastic gradient langevin dynamics. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1989–1998, 2022. 1, 3, 5

[39] Zhengjue Wang, Hao Zhang, Ziheng Cheng, Bo Chen, and Xin Yuan. Metasci: Scalable and adaptive reconstruction for video compressive sensing. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 2083–2092, 2021. 2, 3

[40] Kaixuan Wei, Angelica Aviles-Rivero, Jingwei Liang, Ying Fu, Carola-Bibiane Schönlieb, and Hua Huang. Tuning-free plug-and-play proximal algorithm for inverse imaging problems. In *Procedings of International Conference on Machine Learning*, pages 10158–10169. PMLR, 2020. 2

[41] Yan Wu, Mihaela Rosca, and Timothy Lillicrap. Deep compressed sensing. In *Procedings of International Conference on Machine Learning*, pages 6850–6860. PMLR, 2019. 1

[42] Zhihao Xia and Ayan Chakrabarti. Training image estimators without image ground-truth. *Advances in Neural Information Processing Systems*, 2019. 2

[43] Jinxi Xiang, Yonggui Dong, and Yunjie Yang. Fista-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. *IEEE Transactions on Medical Imaging*, 40(5):1329–1339, 2021. 2

[44] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. Admm-csnet: A deep learning approach for image compressive sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(3):521–538, 2019. 1, 2, 6

[45] Di You, Jian Zhang, Jingfen Xie, Bin Chen, and Siwei Ma. Coast: Controllable arbitrary-sampling network for compressive sensing. *IEEE Transacitons on Image Processing*, 30:6066–6080, 2021. 2, 6

[46] Jian Zhang and Bernard Ghanem. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1828–1837, 2018. 1, 2, 5, 6

[47] Zhonghao Zhang, Yipeng Liu, Jiani Liu, Fei Wen, and Ce Zhu. Amp-net: Denoising-based deep unfolding for compressive image sensing. *IEEE Transacitons on Image Processing*, 30:1487–1500, 2021. 2

[48] Huan Zheng, Tongyao Pang, and Hui Ji. Unsupervised deep video denoising with untrained network. In *Proceedings of AAAI Conference on Artificial Intelligence*, 2023. 3

[49] 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 Conference on Computer Vision and Pattern Recognition*, pages 10255–10264, 2019. 1, 2

[50] Magauiya Zhussip, Shakarim Soltanayev, and Se Young Chun. Extending stein's unbiased risk estimator to train deep denoisers with correlated pairs of noisy images. *Advances in Neural Information Processing Systems*, 32, 2019. 3

Ground-Truth Free Meta-Learning for Deep Compressive Sampling (Supplemental Material)

1. Network Structure

See Figure 1 for the detailed structure of our bias-adaptive unrolling CNN. Our method only adjusts the biases of all convolutional layers except for the first and last ones during adaption.

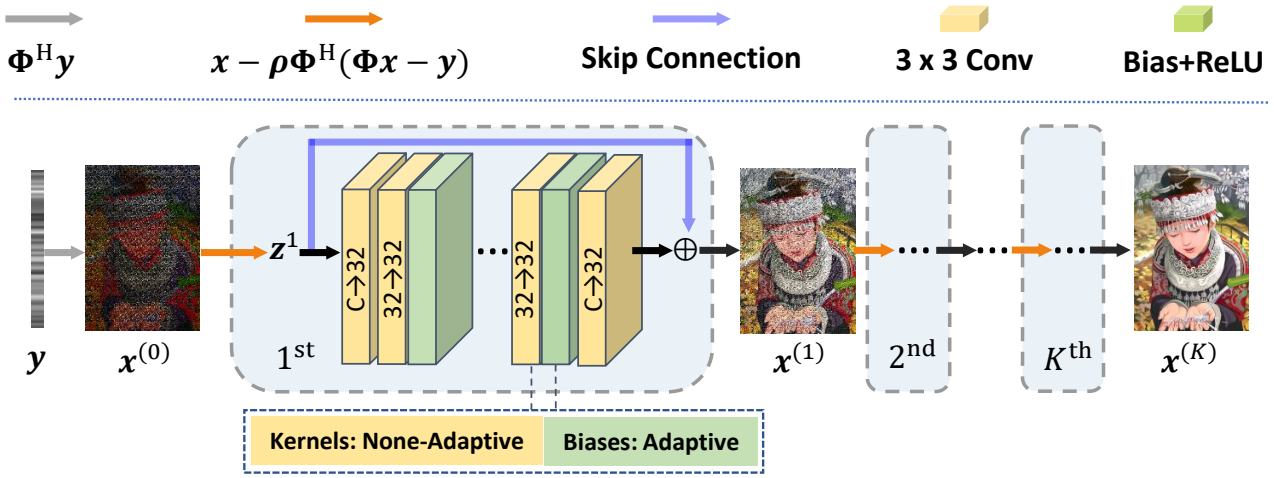


Figure 1. Illustration of our bias-adaptive unrolling CNN. It contains K blocks. Each block is composed of six convolution layers with the kernel size of 3×3, and with a ReLU activation equipped in the second to forth layers. Only the biases of all convolutional layers except for the first and last ones are adjusted during adaption.

2. Proof of Theorem 1

Proof.

$$\nabla_{\boldsymbol{\omega}}\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\ell^{\mathrm{iSURE}}(\boldsymbol{\omega};\boldsymbol{y},\boldsymbol{\Phi},\boldsymbol{\epsilon}') = \nabla_{\boldsymbol{\omega}}\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[\|\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})-\boldsymbol{y}\|_2^2 + 2\sigma^2\mathrm{tr}\Big(\boldsymbol{\Phi}\frac{\partial\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})}{\partial\boldsymbol{y}}\Big)\Big] \tag{1}$$

Let $\boldsymbol{z} = \boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})$ which is a function of $\boldsymbol{y}+\boldsymbol{\epsilon}'$. We have $\partial\boldsymbol{z}/\partial\boldsymbol{\epsilon}' = \partial\boldsymbol{z}/\partial\boldsymbol{y}$ and

$$\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[\mathrm{tr}\Big(\boldsymbol{\Phi}\frac{\partial\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})}{\partial\boldsymbol{y}}\Big)\Big] = \mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[\sum_{i=1}^{M}\frac{\partial\boldsymbol{z}_i}{\partial\boldsymbol{\epsilon}_i'}\Big] = \mathbb{E}_{\boldsymbol{y}}\Big[\sum_{i=1}^{M}\int_{-\infty}^{+\infty}\frac{\partial\boldsymbol{z}_i}{\partial\boldsymbol{\epsilon}_i'}\phi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i'\Big], \tag{2}$$

where $\phi_\sigma(\cdot):\mathbb{R}\to\mathbb{R}$ is the 1D Gaussian p.d.f. with standard deviation σ. Using $\nabla\phi_\sigma(x) = -\frac{1}{\sigma^2}x\phi_\sigma(x)$, we have that

$$\begin{aligned}\int_{-\infty}^{+\infty}\frac{\partial\boldsymbol{z}_i}{\partial\boldsymbol{\epsilon}_i'}\phi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i' &= \boldsymbol{z}_i\phi_\sigma(\boldsymbol{\epsilon}_i')|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty}\boldsymbol{z}_i\nabla\phi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i' = \boldsymbol{z}_i\phi_\sigma(\boldsymbol{\epsilon}_i')|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty}\frac{1}{\sigma^2}\boldsymbol{z}_i\boldsymbol{\epsilon}_i'\phi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i' \\ &= \int_{-\infty}^{+\infty}\frac{1}{\sigma^2}\boldsymbol{z}_i\boldsymbol{\epsilon}_i'\phi_\sigma(\boldsymbol{\epsilon}_i')d\boldsymbol{\epsilon}_i = \frac{1}{\sigma^2}\mathbb{E}_{\boldsymbol{\epsilon}_i'}\boldsymbol{z}_i\boldsymbol{\epsilon}_i'.\end{aligned} \tag{3}$$

Substituting Eq. (3) back into Eq. (2), we have

$$\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[\mathrm{tr}\Big(\boldsymbol{\Phi}\frac{\partial\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})}{\partial\boldsymbol{y}}\Big)\Big]=\mathbb{E}_{\boldsymbol{y}}\Big[\sum_{i=1}^{M}\frac{1}{\sigma^2}\mathbb{E}_{\boldsymbol{\epsilon}_i'}\boldsymbol{z}_i\boldsymbol{\epsilon}_i'\Big]=\frac{1}{\sigma^2}\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[(\boldsymbol{\epsilon}')^{\mathrm{H}}\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})\Big]. \tag{4}$$

Finally, substituting Eq. (4) back into Eq. (1) gives us

$$\begin{aligned}\nabla_{\boldsymbol{\omega}}\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\ell^{\mathrm{iSURE}}(\boldsymbol{\omega};\boldsymbol{y},\boldsymbol{\Phi},\boldsymbol{\epsilon}')=&\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[2\boldsymbol{J}_{\boldsymbol{\omega}}^{\mathrm{H}}\Big(\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})\Big)\boldsymbol{\Phi}^{\mathrm{H}}\Big(\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})-\boldsymbol{y}\Big)\Big]+2\nabla_{\boldsymbol{\omega}}\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[(\boldsymbol{\epsilon}')^{\mathrm{H}}\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})\Big]\\=&\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[2\boldsymbol{J}_{\boldsymbol{\omega}}^{\mathrm{H}}\Big(\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})\Big)\boldsymbol{\Phi}^{\mathrm{H}}\Big(\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})-\boldsymbol{y}\Big)\Big]+2\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[\boldsymbol{J}_{\boldsymbol{\omega}}^{\mathrm{H}}\Big(\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})\Big)\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\epsilon}'\Big]\\=&2\,\mathbb{E}_{\boldsymbol{y},\boldsymbol{\epsilon}'}\Big[\boldsymbol{J}_{\boldsymbol{\omega}}^{\mathrm{H}}\Big(\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})\Big)\boldsymbol{\Phi}^{\mathrm{H}}\Big(\boldsymbol{\Phi}\mathcal{F}_{\boldsymbol{\omega}}(\boldsymbol{y}+\boldsymbol{\epsilon}',\boldsymbol{\Phi})-\boldsymbol{y}+\boldsymbol{\epsilon}'\Big)\Big].\end{aligned} \tag{5}$$

The proof is done. □

Remark 1. *Our theorem and proof consider the case that both the measurements and the measurement noise are real-valued. During implementation, we treat them as real variables with double dimensions, and all expectations involving these complex-valued variables are calculated using real calculus. The proof may be extended to the complex-valued case using Wirtinger derivatives.*

3. Additional Quantitative Experiments

3.1. MR Image Reconstruction

We also use Gaussian masks of ratios $\frac{1}{5}$, $\frac{1}{4}$, $\frac{1}{3}$ for the experiments on the ADNI dataset. See Table 1 for the results. Other experimental settings are kept unchanged. It can be seen that the proposed MetaCS performs the best overall among all GT-free methods and performs competitively with the supervised methods.

Noise	CS Ratio	Non-Learning		Unsupervised	Internal		Unsupervised + Internal		Supervised	
		ZF	SparseMRI	REI	BNN	ASGLD	DDSSL	MetaCS	ADMMNet	Supervised
w/o	1/3	27.22/.71	34.93/.93	37.34/.94	37.60/.94	37.79/.94	37.81/.95	**37.87/.96**	38.22/.98	38.95/.98
	1/4	26.16/.68	32.79/.90	36.08/.94	36.10/.95	36.07/.95	36.68/**.96**	**36.71/.96**	35.94/.96	37.42/.97
	1/5	25.66/.66	31.69/.89	34.35/.92	33.81/.93	34.40/.92	35.43/**.94**	**35.52/.94**	34.81/.96	36.09/.95
w/	1/3	26.60/.65	27.91/.69	29.95/.80	29.46/.75	29.80/.76	31.59/.87	**31.76/.88**	31.04/.86	31.56/.88
	1/4	25.79/.64	27.42/.69	29.87/.79	29.20/.77	29.61/.78	30.91/.86	**31.38/.87**	30.92/.86	31.23/.87
	1/5	25.40/.63	26.97/.68	29.21/.77	29.17/.76	29.45/.75	29.39/.86	**30.95/.87**	30.81/.86	31.02/.88

Table 1. Mean PSNR(dB)/SSIM results of natural image reconstruction. **Boldfaced**: best of all compared **GT-free** methods.

3.2. Natural Image Reconstruction

We also test MetaCS with natural image reconstruction on BSD68 [1]. The results are reported in Table 2. Again, MetaCS performs the best overall among all GT-free methods and performs competitively with the supervised methods.

Noise	CS Ratio	Non-Learning	Unsupervised		Internal		Unsupervised+Internal		Supervised		
		TVAL3	LSURE	REI	BNN	ASGLD	DDSSL	MetaCS	Supervised	COAST	SSLIP
w/o	40%	29.39/.86	31.87/.90	31.79/.90	31.28/.90	31.36/.90	32.53/**.92**	**32.63/.92**	32.17/.92	33.02/.92	30.72/.88
	25%	26.48/.77	28.73/.84	28.45/.82	28.63/.84	29.51/.84	29.47/**.86**	**29.67/.86**	29.36/.85	30.07/.87	28.26/.81
	10%	22.49/.58	23.07/.65	23.11/.63	25.24/.71	25.51/.70	**26.03/.72**	25.88/**.72**	25.32/.71	26.25/.72	24.72/.66
w/	40%	26.15/.68	27.73/.77	28.05/.79	28.13/.81	28.75/.81	29.21/**.84**	**29.49/.84**	26.86/.72	28.98/.83	28.47/.83
	25%	24.75/.67	28.14/.82	28.08/.81	28.67/.84	29.35/.85	29.61/**.87**	**29.71/.87**	29.49/.86	29.37/.86	28.71/.85
	10%	22.03/.52	23.54/.60	22.37/.60	23.79/.64	24.56/.64	24.63/.67	**24.67/.68**	23.86/.60	24.02/.67	24.25/.67

Table 2. Mean PSNR(dB)/SSIM results of natural image reconstruction. **Boldfaced**: best of all compared **GT-free** methods.

3.3. Performance Using a Larger Model

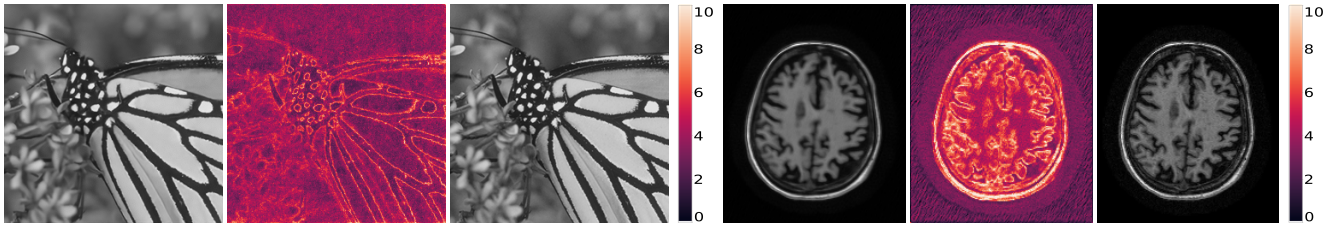
We enlarge the model size of MetaCS to 0.75M (versus 0.79M of the original DDSSL) by setting $K = 20$ while keeping other settings unchanged. See Table 3 for its results with comparison to the original DDSSL (without reducing its model but quoting results from its paper). Clearly, our MetaCS still performs better in both reconstruction accuracy and complexity.

Noise	Method	MRI - MRI150			MRI - ADNI			Natural Images			Time
		$r = 20\%$	30%	40%	r =1/5	1/4	1/3	$r = 10\%$	25%	40%	(min.)
w/o	DDSSL	36.75	38.48	41.03	33.90	34.67	36.43	27.59	33.41	37.19	7.28
	Ours	**37.15**	**39.63**	**41.26**	**34.11**	**34.86**	**36.59**	**28.06**	**33.58**	**37.36**	**1.81**
w/	DDSSL	33.82	34.56	34.90	30.49	31.08	31.71	26.12	29.84	31.94	7.28
	Ours	**34.38**	**35.27**	**35.62**	**30.62**	**31.21**	**31.84**	**26.25**	**29.89**	**31.97**	**1.81**

Table 3. PSNR results of original DDSSL and enlarged MetaCS. The time is reported on natural image reconstruction.

4. Uncertainty Quantization

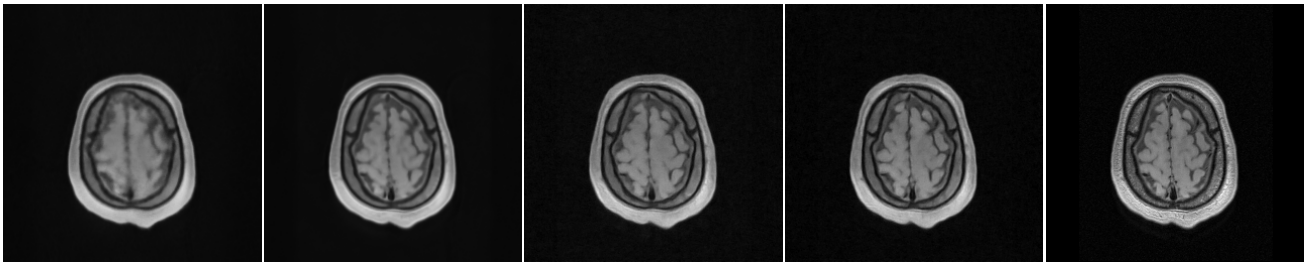
Uncertainty quantification on reconstructed images is an important feature often requested in scientific imaging and medical imaging. Recall that the ensemble inference provides a number of instances, which, as a byproduct, can be used for uncertainty quantization of CSR. Concretely, we calculate the standard deviation of $\mathcal{F}_{\boldsymbol{\omega}^*}(\boldsymbol{y}^* + \boldsymbol{\epsilon}', \boldsymbol{\Phi})$ over $\boldsymbol{\epsilon}'$ as the uncertainty map. Figure 2 visualizes such maps on two samples. We can see that flat regions have lower uncertainty while structure regions (*e.g.* edges and textures) have higher uncertainty.



Reconstructed image | Uncertainty map | GT | Reconstructed image | Uncertainty map | GT

Figure 2. Visualization of uncertainty maps of noisy CSR using 10 instances. Up: Natural image, r=40%; bottom: MR image, r=1/3.

5. Visual Comparison on Reconstructed Images



REI | ASGLD | DDSSL | MetaCS | GT

Figure 3. Visual results of MR image reconstruction with unknown noise type.

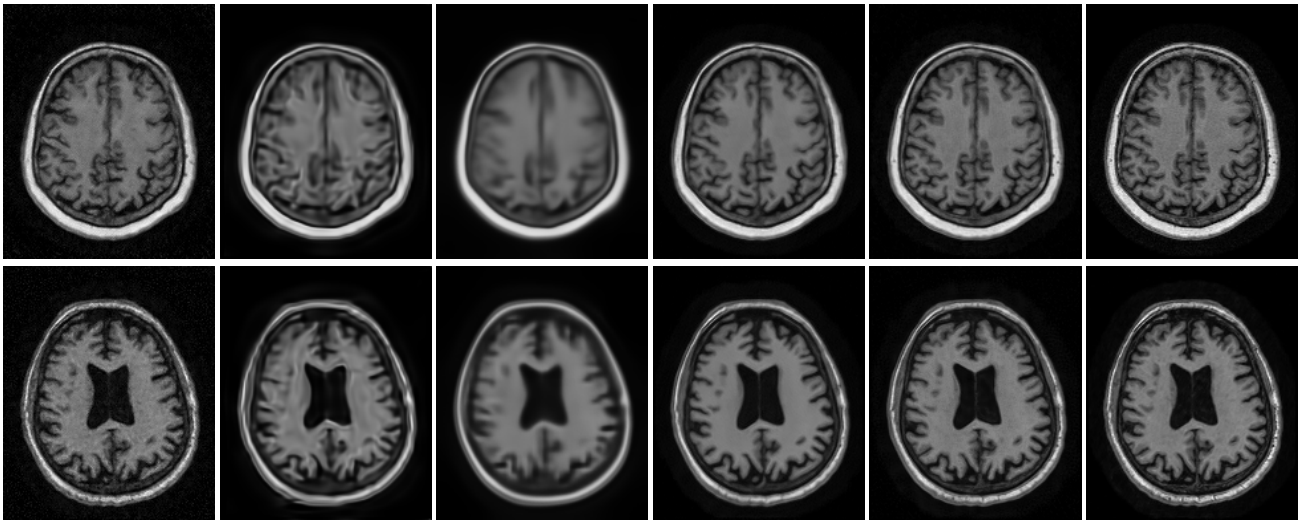


Figure 4. Visual results of MR image reconstruction with radial mask (the first row) and 2D Gaussian mask (the second row) of sampling ratio $1/4$ in noisy setting.

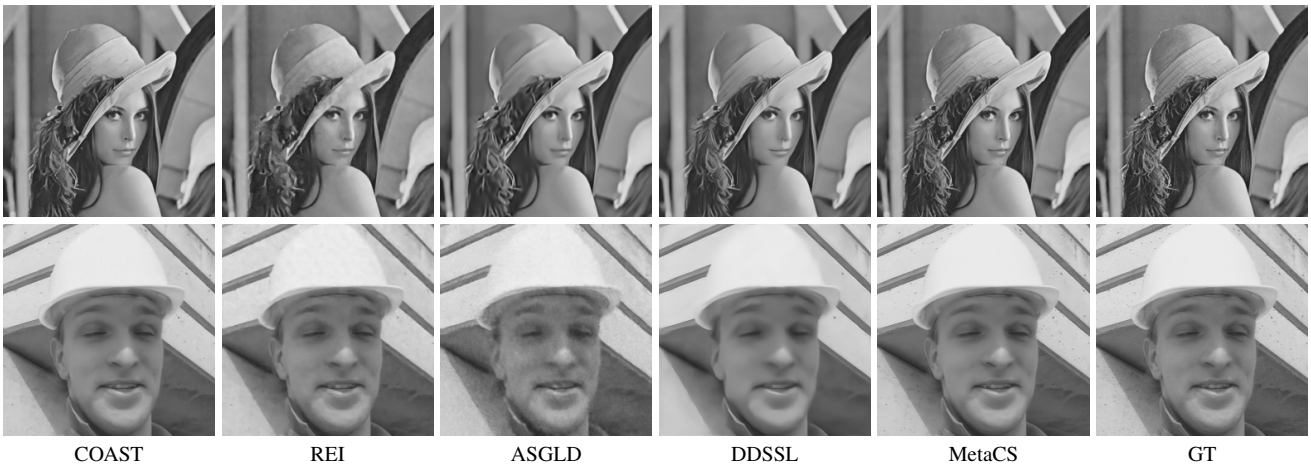


Figure 5. Visual results of natural image reconstruction from Gaussian measurements of sampling ratio 40% in noisy setting (second row).

References

[1] 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 International Conference on Computer Vision*, volume 2, pages 416–423. IEEE, 2001. 2