Test-Time Model Adaptation for Image Reconstruction Using Self-supervised Adaptive Layers

Yutian Zhao[1], Tianjing Zhang[1], and Hui Ji[1]

Department of Mathematics, National University of Singapore, 119076, Singapore
{e0708171,tianjingzhang}@u.nus.edu, matjh@nus.edu.sg

Abstract. Image reconstruction from incomplete measurements is a basic task in medical imaging. While supervised deep learning proves to be a powerful tool for image reconstruction, it demands a substantial number of latent images for training. To extend the application of deep learning to medical imaging where collecting latent images poses challenges, this paper introduces an self-supervised test-time adaptation approach. The proposed approach leverages a pre-trained model on an external dataset and efficiently adapts it to each test sample for optimal generalization performance. Model adaption for an unrolling network is done with additional lightweight adaptive linear layers, enabling efficient alignment of testing samples with the distribution targeted in the pre-trained model. This approach is inspired by the connection between linear convolutional layer and Wiener filtering. Extensive experiments showed significant performance gain of the proposed method over other unsupervised methods and model adaptation techniques in two medical imaging tasks.

Keywords: Model adaptation, Image reconstruction, Self-supervised learning

1 Introduction

Image reconstruction (IR) from partial linear measurements plays a critical role in medical imaging. For instance, in key modalities such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), the reconstruction process is often framed as solving a linear inverse problem:

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

where $\mathbf{\Phi} \in \mathbb{C}^{M\times N}$ denotes the measurement matrix which models the forward acquisition process, $\mathbf{x} \in \mathbb{R}^N$ the latent image to reconstruct, $\epsilon \in \mathbb{C}^N$ the measurement noise, and $\mathbf{y} \in \mathbb{C}^M$ the collected incomplete measurements. The matrix $\mathbf{\Phi}$ is determined by the associated imaging system.

Generally, reconstructing image $\mathbf{x}$ from noisy observations $\mathbf{y}$ is ill-posed, as the system (1) is under-determined. Recent medical imaging exacerbates this by further reducing measurement. For example, compressed-sensing (CS) MRI [37]

decreases M to quicken acquisition, and sparse-view (SV) CT [32] decrease M to reduce patients' radiation exposure. These practices make the problem (1) increasingly susceptible to noise and solution ambiguity, thereby demanding more effective image reconstruction methods.

In recent years, deep learning has been an important tool for IR in medical imaging, including both CS-MRI and SV-CT; see *e.g.* [10, 13, 42–45, 48, 49]. The prevailing approaches are based on supervised learning, which fits a deep neural network (NN) model over the set of paired latent images and their measurements. The effectiveness of these models heavily relies on the size of the training dataset. However, the acquisition of ground-truth (GT) images can be very expensive and time-consuming in the medical field. Consequently, DNNs trained on limited datasets often suffer from unsatisfactory generalization performance. Furthermore, these datasets may be biased and not sufficiently diverse for robust image reconstruction, posing issues when new pathologies not present in the training data are encountered during testing.

1.1 Related deep learning for IR in the data-limited environments

To address potential biases and poor generalization due to limited training data, various methodologies have been developed. These can generally be divided into two categories: (1) *Ground truth-free self-supervised/unsupervised learning* and (2) *Unsupervised meta learning and Test-time model adaptation* (TTA).

Self-supervised and unsupervised learning methods: Current self-supervised or unsupervised learning techniques often utilize the deep image prior (DIP) [36], which suggests that an untrained convolutional NN (CNN) prefers regular structure over noise during training. Typically, these approaches are divided into two strategies: external learning, where an NN is trained without requiring GT [1, 2, 29, 51]; or internal learning [22, 38], which involves training a CNN on test samples to capture unique internal characteristics of those samples.

Unsupervised external learning tackles the absence of GT images by training the NN on unlabeled data. Generally, they suffer from the same issues as supervised learning: the possible poor generalization to test samples that deviate from the original distribution. Self-supervised internal learning allows for the exploitation of specific characteristics of each test sample. However, this method is very time-consuming, as it necessitates individual training on each test sample.

Unsupervised meta-learning and TTA for IR: Meta-learning is to train a model such that it can quickly adapt to new tasks or problems with minimal data. Most of meta learning for IR, *e.g.* [3, 24, 40, 50], are supervised, which requires GT images to fine-tune the NN. There are few works on unsupervised meta learning for IR. Qin *et al.* [25] proposed an unsupervised meta-learning for IR in CS-MRI, using a SURE-based self-supervised loss. TTA aims to adjust a pre-trained model at inference time, to align the model to test data, thereby improving its performance on this data (*e.g.* [6, 35]). Despite their practical benefits, there are few studies devoted to IR. MetaCS [25] introduced a self-supervised loss Based on R2R [23], DDSSL [29] formulated a dual-domain self-supervised loss, in the

context of IR in CS-MRI and SV-CT. While these self-supervised loss functions showed effectiveness in mitigating the issue caused by the absence of GT images, their performance still lags behind that of supervised methods.

1.2 Main idea

Motivated by the advantages of TTA methods for IR, this paper concentrates on developing an efficient unsupervised TTA method for IR. This method operates without needing access to ground-truth images during model adaptation. In contrast to existing works that emphasize defining self-supervised losses [25, 29], our focus is on network design to improve the performance of adaptation methods for IR. For clarification, the problem setting is outlined below.

Problem setting: *Consider a model, $\mathcal{F}(\cdot, \theta^*) : Y \to X$, pre-trained on a dataset $\{Y_j, X_j\}_j$, where $X \sim P_X$, the task is to efficiently adapt this pre-trained model $\mathcal{F}(\cdot, \theta^*)$ to a new test sample $\widetilde{Y}$, so that it performs optimally for this specific instance $\widetilde{Y}$, potentially at the expense of its performance on other examples.*

Our study is based on the the unrolling NNs for IR, *e.g.*, proximal gradient descent algorithm (PGDA) [7], derived from some iterative scheme for solving inverse imaging problems. The NN is concatenated with multiple stages that correspond to multiple iterations of the solver, each designed for updating image reconstruction from the preceding estimate and consequently refining it with a denoising NN. The entire network is optimized by minimizing the mean squared error (MSE) to the truth images of training data. To ensure the training stability, the denoising CNN within each stage is trained to minimize the MSE.

In an unrolling NN, the component that interacts with the target image distribution is the denoising NN. It functions as an estimator to the truth image from its noisy measurement(received from the latest update), since it's trained to minimize MSE to true images. For simplicity, we consider a linear CNN by omitting non-linear activation functions and assume data have zero mean, thereby removing the bias term. Consequently, the CNN architecture is simplified to a sequence of convolutions with small filters, equivalent to a single convolution with a very large filter. In image processing, a convolution-based optimal estimator is known as a *Wiener filter*, defined over the second moments of noise and image in the frequency domain. Assume the second moment of the test images is close to that of the training dataset, then applying an additional convolution on the Wiener filter designed for the original dataset can effectively approximate the Wiener filter for the new test sample.

Building on the insight above, we introduce a TTA method named *A*daptNet for IR. The main idea involves embedding simple, parameter-efficient convolution layers into the original unrolling network. This addition leads to $< 1\%$ increase in the model size in our implementation. Given a pre-trained model, during testing, updates are applied solely to these newly added parameters by minimizing an R2R-based self-supervised loss [23]. This tailored adaptation equips the original model to predict the latent image from the new test sample more accurately.

1.3 Contributions

In summary, our contributions are listed as follows:

- An understanding of model adaptation from the perspective Wiener filter for denoising. This provides insight and interpretation for our proposed method.
- An effective TTA method that incorporates a lightweight module, termed AdaptBlock, into the pre-trained model. This enables rapid and efficient adaptation to test samples with a significant performance boost.
- An self-supervised loss for guiding the model adaption in test time, built on the R2R loss for denoising [23].

Extensive experiments on MRI and CT image reconstruction showed the significant benefits of our proposed AdaptNet, over other unsupervised methods and model adaptation techniques:

- When adapting a model pre-trained on a dataset of similar medical images, our method significantly outperformed all existing model adaptation approaches, including both supervised and unsupervised methods.
- Even when utilizing a model initially trained on natural images, our method still excelled beyond other adaptation techniques that adapt the models trained with very related medical images. This shows our method's proficiency in transferring knowledge from natural to medical images.
- Our approach demonstrated its practical versatility by effectively adapting pre-trained models to test data with varying measurement matrices.

2 Related Work

GT-dependent supervised learning: The majority of deep learning methods for IR rely on supervised learning. These techniques primarily focus on network design, training an NN on some external datasets with GT images. Early work trains a CNN-based denoiser to reduce artifacts in outputs from other methods (*e.g.* [11]). A recent trend is the physics-aware unrolling NN, which unrolls some iterative solver for some variational model of IR, and replaces the prior-related operations by learnable NN blocks [8,9,44,45,48]. The performance of supervised learning heavily depends on both the quality and quantity of trained samples.

Deep Learning with Plug-and-Play prior: To address the challenges posed by limited training data, one strategy involves utilizing pre-trained models in IR. In [19, 31], pre-trained CNNs are used as the regularization step in an unrolling NN. Pre-trained generative adversarial network (GAN) models have also been used for regularization (*e.g.*, [15, 16]). More recently, pre-trained diffusion models [4,33,34], have been used for solving inverse imaging problems. The effectiveness of these methods depends on the alignment of the distribution of data for pre-training with that of test samples, and they face challenges when two distributions mismatch or when processing out-of-distribution test samples.

GT-free unsupervised external learning: The lack of GT images in many IR tasks has motivated the study of unsupervised learning methods. Unsupervised

external learning, for instance, involves training an NN on datasets without GT images. In compressive imaging, Metzler *et al.* [20] and Zhussip *et al.* [51] train the NN using a SURE-based denoiser. Chen *et al.* [1] exploits the equivalence of NN to use range-space information from training samples to supervise IR in null space. Quan *et al.* [29] utilized a dual-head injected loss for self-supervised learning. Quan *et al.* [30] proposed a Siamese cooperative learning scheme of self-supervised learning for IR. There are also other works on other IR tasks, *e.g.*, deblurring [28] and phase retrieval [27]. These approaches share the same issue as supervised learning: Poor generalization to test samples deviating from the distribution of training data and to out-of-distribution test samples.

Self-supervised internal learning: Self-supervised internal learning trains an untrained CNN on each test sample to predict the latent image. This approach is largely based on DIP [36], which enables using untrained CNNs for implicit regularization on IR. There are many works on self-supervised denoising tasks (*e.g.* [12,14,17,23,26,39]). Many methods have been extended to IR by adapting self-supervised loss to IR problems (*e.g.*) [25,29]). Also,Pang *et al.* [22] employed Bayesian CNNs and Wang *et al.* [38] applied the Monte Carlo sampling method to NN parameters to approximate Bayesian inference. As these methods are tailored to each test sample, they require training from scratch for each sample, leading to high computational costs.

Supervised and unsupervised Meta learning: In Meta-learning, a model is trained to quickly adapt to new tasks or problems with minimal data. In IR, meta-learning has been applied to various IR tasks (*e.g.* [3, 24, 40, 50]. Most of meta learning for IR are supervised, which requires GT images to fine-tune the NN. There are few works on unsupervised meta learning for IR. Qin *et al.* [25] proposed an unsupervised meta-learning for IR in CS-MRI, which is based on a SURE-based self-supervised loss defined on only measurements.

TTA for IR: TTA adjusts a pre-trained model in inference time. This adjustment is to align the model to the test data, thereby improving its performance on this data, even if it deviates from the training data (*e.g.* [6,35]). Many techniques have been developed for efficient fine-tuning [21, 25]. These include "Gain-tuning", adjusting the multiplicative scaling parameter for each channel, and "Bias-tuning", adjusting only the bias of the convolutional layer.

To overcome the lack of guidance from GT images during adaptation, several self-supervised losses have been proposed for TTA in IR. MetaCS [25] introduced a SURE-based self-supervised loss for unsupervised meta-learning in IR. Building on R2R [23], DDSSL [29] developed a dual-domain self-supervised loss, applicable to TTA for IR in CS-MRI and SV-CT. While these self-supervised loss functions [25, 29] have helped mitigate the absence of GT images during adaptation, their performance still lags behind that of supervised methods.

3 Methodology

This section introduces AdaptNet, an efficient unsupervised TTA method for IR. The main idea comes from recognizing that the crucial part associated with

the distribution of data/noise is the denoising CNN within an unrolling NN. Indeed, the optimal linear CNN without bias term, is trained to approximate the Wiener filter (the optimal linear estimator of GT images from their noisy measurements). Provided the distribution of test samples closely relates to that of the training data, adding a few convolution layers can efficiently adapt the corresponding Wiener filter to align the linear CNN with the new test samples.

3.1 Wiener filter within an iterative IR method

Consider the following IR problem:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \epsilon,$$

where $\mathbf{y}$ denotes the measurement, $\mathbf{x}$ the latent image, and ϵ the noise. An unrolling NN is based on some iterative scheme derived from a MAP estimator of the truth images from its partial measurement. Recall that, in the presence of i.i.d. white noise $\epsilon \sim \mathbb{N}(\mathbf{0}, \sigma^2\mathbf{I})$, the MAP estimator for $\mathbf{x} \sim p(\mathbf{x})$ is

$$\mathbf{x}_{\text{map}} = \underset{\mathbf{x}}{\text{argmax}} \log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x}) = \underset{\mathbf{x}}{\text{argmin}} \frac{1}{2\sigma^2}\|\mathbf{\Phi}\mathbf{x} - \mathbf{y}\|_2^2 + \log p(\mathbf{x}) \quad (2)$$

Among many solvers for (2), we adopt the PGDA method [7]: for $n = 1, 2, \ldots$,

$$\mathbf{z}^{(n)} = \mathbf{x}^{(n-1)} - \rho\mathbf{\Phi}^\top\left(\mathbf{\Phi}\mathbf{x}^{(n-1)} - \mathbf{y}\right), \qquad \text{(Updating)} \quad (3)$$

$$\mathbf{x}^{(n)} = \arg\min_{\mathbf{x}} \frac{1}{2\sigma^2}\left\|\mathbf{x} - \mathbf{z}^{(n)}\right\|_2^2 + \log p(\mathbf{x}); \qquad \text{(Denoising)} \quad (4)$$

Based on the iterative scheme above, one can have an unrolling NN with many stages. Within each stage, the denoising step is replaced by a learnable denoising NN. Let $\mathcal{F}(\cdot, \theta)$ denotes a NN with parameters θ. Then,

$$\mathbf{x}^{(n)} = \mathcal{F}(\mathbf{z}^{(n)}; \theta^{(n)}).$$

The prior distribution of images $p(\mathbf{x})$ is utilized within the denoising NN, $\mathcal{F}(\cdot, \theta)$. Training the NN to minimize prediction error means adjusting the parameters of $\mathcal{F}(\cdot, \theta)$ to approximate an minimum mean squared error (MMSE) estimator, predicting true 2D images $\mathbf{x}$ from their noisy counterparts $\mathbf{z} = \mathbf{x} + \mathbf{n}$ via minimizing the MSE:

$$\min_\theta \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x}), \mathbf{n}\sim p(\mathbf{n})}\{\|\mathcal{F}(\mathbf{z}, \theta) - \mathbf{x}\|_F^2\} \approx \min_\theta \sum_j \|\mathcal{F}(\mathbf{z}_j, \theta) - \mathbf{x}_j\|_F^2\}. \quad (5)$$

It remains an open problem to understand the statistical properties of a general CNN as a form of Bayesian inference. For ease of discussion, let's consider a CNN composed solely of linear convolution layers without non-linear activation functions and bias terms, which is expressed as

$$\mathcal{F}(\mathbf{z}; \theta) = W_\theta\mathbf{z} = (h_1 \otimes h_2 \otimes \cdots \otimes h_L) \otimes \mathbf{z} = h \otimes \mathbf{z}.$$

where $\otimes$ denotes the convolution operator, and h is the filter of large size, composed of many filters of small size. Indeed, the optimal filter h_w, known as Wiener filter, aims at minimizing the Bayesian risk:

$$h_w := \text{argmin}_h \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x}),\mathbf{n}\sim p(\mathbf{n})}\{\|h \otimes \mathbf{z} - \mathbf{x}\|_F^2\}.$$

The Wiener filter is defined by the discrete Fourier transform (DFT) denoted by $\widehat{(\cdot)}$. As $\widehat{(\cdot)}$ is unitary, we have, for each frequency $[k_1, k_2]$,

$$\widehat{h_w} := \text{argmin}_{\widehat{h}} \mathbb{E}\{\|\widehat{h} \odot \widehat{\mathbf{z}} - \widehat{\mathbf{x}}\|_F^2\} = \sum_{k_1}\sum_{k_2} \mathbb{E}\{|\widehat{h}[k_1,k_2]\widehat{\mathbf{z}}[k_1,k_2] - \widehat{\mathbf{x}}[k_1,k_2]|^2\}, \quad (6)$$

where $\odot$ denotes the element-wise multiplication. Suppose that the noise has zero mean: $\mathbb{E}\{\mathbf{n}\} = \mathbf{0}$. Then, the solution to (6) is

$$\widehat{h}_w[k_1,k_2] = \frac{\mathbb{E}_{\mathbf{x}}\{|\widehat{\mathbf{x}}[k_1,k_2]|^2\}}{\mathbb{E}_{\mathbf{x},\mathbf{n}}\{|\widehat{\mathbf{z}}[k_1,k_2]|^2\}} = \frac{\mathbb{E}_{\mathbf{x}}\{|\widehat{\mathbf{x}}[k_1,k_2]|^2\}}{\mathbb{E}_{\mathbf{x}}\{|\widehat{\mathbf{x}}[k_1,k_2]|^2\} + \mathbb{E}_{\mathbf{n}}\{|\widehat{\mathbf{n}}[k_1,k_2]|^2\}}$$

provided that $\mathbf{x}$ and $\mathbf{n}$ are independent. Similarily, for a specific test sample $\widetilde{\mathbf{z}} = \widetilde{\mathbf{x}} + \widetilde{\mathbf{n}}$, its **oracle wienter filter** that minimizes the risk: $\mathbb{E}_{\mathbf{n}}\|\widetilde{h} \otimes \widetilde{\mathbf{z}} - \widetilde{\mathbf{x}}\|_F^2$, is given by

$$\widehat{\widetilde{h}}_{oracle}[k_1,k_2] = \frac{|\widehat{\widetilde{\mathbf{x}}}[k_1,k_2]|^2}{|\widehat{\widetilde{\mathbf{x}}}[k_1,k_2]|^2 + \mathbb{E}_{\widetilde{\mathbf{n}}}\{|\widehat{\widetilde{\mathbf{n}}}[k_1,k_2]|^2\}}.$$

When pre-training a linear CNN on a training dataset, the goal is to approximate the optimal estimator for the image distribution, relying on the second moment of images and noise in the frequency domain. To approximate the estimator optimized for the test sample, an additional convolution layer can adjust the universal Wiener filter h_w to resemble the Wiener filter $\widetilde{h}_{oracle}$, specifically optimized for each individual test sample $\widetilde{\mathbf{x}}$. If the test sample's spectrum is close to the training image distribution's expectation, a convolution filter of small size can align the adjusted filter with the oracle Wiener filter for test sample. See Figure 1 for an illustration. This insight underlies our approach: adding a lightweight convolution layer to the pre-trained model to match the test sample's characteristics.

3.2 AdapNet for TTA

Unrolling NN with adaptive layer. Motivated by the discussion above, we present an unrolling NN where each iteration is attached by a single adaptive convolution block, namely AdaptBlock, which is specifically for efficient test-time adaption. The diagram of the unrolling NN is shown in Figure 2. The pre-trained model, locked in Figure 2, is a typical unrolling network where (4) is replaced by a learnable NN with multiple conv layers. In our implementation, we adopt the same structures as [29]. The NN is the unrolling of PGDA [7] with a total of 12 iterations. Within each iteration, the denoising sub-NN has 9 convolution layers with the middle 6 layers equipped with ReLU activation.

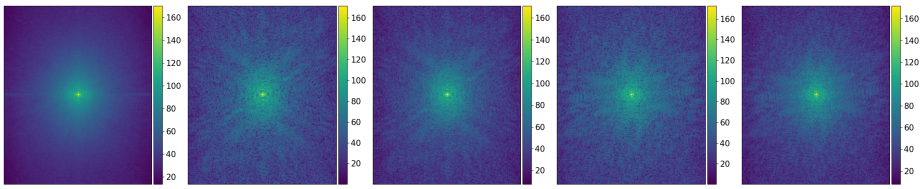

Fig. 1: The figure illustrates the log-spectrum of (**a**): the average over 20 training samples; (**b**)&(**d**): two testing samples; (**c**)&(**e**): the results after convolving testing sample in (**b**) and (**d**) with a filter of size 5×5, respectively. Both training samples and testing samples are randomly selected from ADNI dataset [16] without overlaps. It shows that when the spectrum of test sample is not too different from the average, the adjustment with a small filter effectively aligns the two spectra.

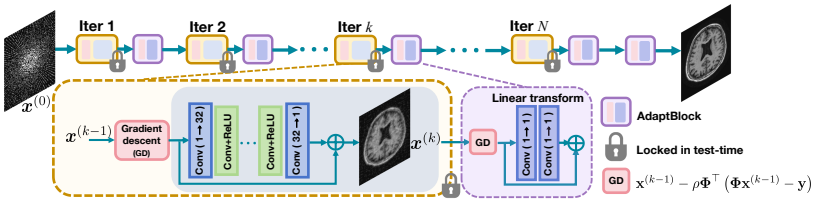

Fig. 2: The PGDA-based unrolling CNN with AdaptBlocks, which is attached to the end of the locked sub-NN from the pre-trained model in each iteration. Only these AdaptBlocks are fine-tuned in test time.

For the denoising sub-NN in each iteration, we concatenate an adjustable lightweight linear operator, termed AdaptBlock, to the end of the locked sub-NN. The AdaptBlock is a linear operator with one gradient descent step (3) and two convolution layers without ReLU activation. The convolutional layers take a single-channel input, apply a convolution with 3×3 filter, and produce a single-channel output. The convolution layers are initialized with 0 and the gradient descent step is initialized with 0.5.

Test-time adaption. Consider the AdaptNet $\mathcal{F}(\cdot; \theta^*, \phi)$:

$$\mathcal{F}(\cdot; \theta^*, \phi): \quad \mathbf{y} \to \mathbf{x},$$

where θ^* denotes the parameters of the pre-trained model locked in test time, and ϕ denotes the AdaptBlock parameters to be updated during test time. Given a test sample $\widetilde{\mathbf{y}}$, we may adapt the pre-trained model $\mathcal{F}(\cdot; \theta^*, \phi)$ to $\widetilde{\mathbf{y}}$ by minimizing some self-supervised loss. Recall that, the space $\mathbb{R}^N$ can be decomposed as:

$$\mathbb{R}^N = \text{range}(\mathbf{\Phi}^H) \oplus \text{null}(\mathbf{\Phi}).$$

where $\text{range}(\mathbf{\Phi}^H) = \{\mathbf{\Phi}^H\mathbf{y}, \mathbf{y} \in \mathbb{R}^M\}$ and $\text{null}(\mathbf{\Phi}) = \{\mathbf{x} \in \mathbb{R}^N : \mathbf{\Phi}\mathbf{x} = \mathbf{0}\}$.

For the supervision in range space, we proposed the following loss:

$$\mathcal{L}^{range}(\widetilde{\mathbf{y}}; \theta^*, \phi) = \mathbb{E}_{\epsilon' \sim \mathcal{N}(0, \sigma_1^2\mathbf{I})} \|\mathbf{\Phi}\mathcal{F}(\widetilde{\mathbf{y}} + \epsilon'; \theta^*, \phi) - \widetilde{\mathbf{y}} + \epsilon'\|^2 \tag{7}$$

where $\epsilon' \sim \mathcal{N}(0, \sigma_1^2 \mathbf{I})$ denote Gaussian noise added to the measurement, independent from ϵ and resampled in each iteration. The injection of the noise ϵ' is from the self-supervised R2R loss for image denoising in [23] and its extension to IR [29]. For the supervision in null space, the measurement provides no information of $\mathbf{x}$. Thus, we proposed to use the estimate from the pre-trained model as a noisy measurement of $\mathbf{x}$ in the null space, leading to the following loss:

$$\mathcal{L}^{null}(\widetilde{\mathbf{y}}; \theta^*, \phi) = \mathbb{E}_{\epsilon'' \sim p(\epsilon'')} \| (\mathbf{I} - \mathbf{\Phi}^H (\mathbf{\Phi}\mathbf{\Phi}^H)^{-1} \mathbf{\Phi}) \left[(\mathcal{F}(\mathbf{\Phi}(\mathbf{x}^* + \epsilon''), \theta^*, \phi)) - (\mathbf{x}^* - \epsilon'') \right] \|_2^2, \tag{8}$$

where $\mathbf{x}^* = \mathcal{F}(\widetilde{\mathbf{y}}, \theta^*)$ denote the estimate of $\mathbf{x}$ from the pre-trained model. In the loss above, $\mathbf{I} - \mathbf{\Phi}^H (\mathbf{\Phi}\mathbf{\Phi}^H)^{-1} \mathbf{\Phi}$ is the projection to null($\mathbf{\Phi}$), ϵ'' denote the additional injected noise, the same idea as $\mathcal{L}^{null}$. However, as the noise (the residual of $\mathbf{x}^*$) is highly correlated to the image, then we sample the noise from a Poisson distribution associated with image $\mathbf{x}$: $\epsilon'' \sim \text{Poisson}(\epsilon''; \mathbf{x}^*)$. The overall loss is then $\mathcal{L}^{range} + \lambda \mathcal{L}^{null}$ where $\lambda \in \mathbb{R}^+$ is a hyper-parameter. To conclude, in test time, a pre-trained model $\mathcal{F}(\cdot; \theta^*, \phi)$ is updated by gradient descent:

$$\phi \leftarrow \phi - \nabla_\phi \mathcal{L}^{range}(\mathbf{y}; \theta^*, \phi) - \lambda \nabla_\phi \mathcal{L}^{null}(\mathbf{y}; \theta^*, \phi) \tag{9}$$

for T iterations. Once done, we have a model $\mathcal{F}(\cdot; \theta^*, \phi^*)$ adapted to the input test sample $\widetilde{\mathbf{y}}$. Then, we apply it to predict the image associated with $\widetilde{\mathbf{y}}$ by:

$$\widetilde{\mathbf{x}}^* = \frac{1}{T} \sum_{i=1}^{T} \mathcal{F}(\widetilde{\mathbf{y}}_i + \epsilon_i'; \theta^*, \phi^*), \quad \epsilon_i' \sim \mathcal{N}(0, \sigma^2 \mathbf{I}).$$

4 Experiments

We evaluate the proposed framework on two medical image reconstruction tasks: CS-MRI and SV-CT IR. The AdaptNet is tested on the PGDA-based unrolling NN model, same as [29]. See the locked part of Figure 2 for the diagram of the pre-trained model. The learnable parameter ρ is initialized as 0.5, bias is initialized as 0) and the other weights ω use Xavier initialization.

Mask	CS Ratio	Classic	Unsupervised			Supervised				Model Adaptation		
		SparseMRI	REI	BNN	ASGLD	GAN	ADMMNet	Restormer	Pre-trained†	DDSSL	MetaCS	AdaptNet
Gaussian	1/3	27.91/.69	29.82/.73	29.46/.86	29.80/.76	26.49/.76	31.04/.86	30.27/.86	30.39/.89	31.62/.87	31.80/.89	**34.82/.91**
	1/4	27.42/.69	29.61/.72	29.20/.85	29.61/.78	26.31/.75	30.92/.86	30.45/.86	30.10/.89	31.30/.86	31.43/.87	**33.61/.91**
	1/5	26.97/.68	29.00/.72	29.17/.85	29.45/.75	25.79/.75	30.81/.86	30.10/.85	29.86/.88	30.89/.86	30.98/.87	**33.01/.90**
Radial	1/3	29.32/.69	30.22/.81	29.58/.82	29.85/.84	26.72/.75	30.94/.83	30.45/.83	30.35/.89	31.68/.86	31.84/.91	**34.13/.91**
	1/4	25.93/.65	29.56/.75	29.47/.87	29.61/.87	25.55/.74	29.95/.82	29.64/.84	29.82/.89	31.04/.88	31.21/.91	**32.88**/.90
	1/5	24.88/.59	29.77/.76	28.38/.84	28.45/.84	25.02/.73	28.81/.81	28.56/.83	29.38/.88	30.44/.87	30.62/.87	**31.68/.89**

Table 1: Mean PSNR(dB)/SSIM values of MRI reconstructions on ADNI datasets in noisy setting. Pre-trained† is the pre-trained model with the same backbone as DDSSL and MetaCS. **Boldfaced**/Underline: best/second best of all compared methods.

	Mask	Classic	Unsupervised			Supervised			Model Adaptation		
		SparseMRI	REI	ASGLD	Score-MRI	ADMMNet	Restormer	Pre-trained†	DDSSL	MetaCS	AdaptNet
×4	Poisson	28.32/.71	31.88/.83	30.68/.81	30.49/.76	33.87/.86	34.17/.86	33.74/.84	34.65/.87	34.88/.87	**36.51/.90**
	Uniform	27.83/.69	30.01/.80	29.94/.79	28.72/.73	32.44/.80	32.23/.81	32.13/.80	31.99/.78	32.28/.81	**33.01/.84**
	Gaussian	28.01/.70	31.56/.82	30.45/.81	29.96/.75	34.34/.86	34.12/.85	34.04/.84	33.56/.86	33.47/.86	**36.09/.89**
×8	Poisson	26.68/.67	29.76/.79	28.96/.77	29.15/.74	32.81/.82	32.79/.82	32.86/.82	31.56/.82	31.23/.82	**33.49/.86**
	Uniform	25.88/.64	27.66/.70	26.87/.68	27.08/.68	30.41/.78	30.33/.78	30.28/.78	29.07/.77	29.14/.77	**30.54/.79**
	Gaussian	26.86/.68	30.49/.81	30.11/.80	28.45/.71	33.25/.84	33.46/.84	33.65/.83	33.65/.84	33.78/.85	**35.40/.87**

Table 2: Mean PSNR(dB)/SSIM values of MRI reconstructions on fastMRI dataset in noisy setting. Pre-trained† is the pre-trained model with the same backbone as DDSSL and MetaCS. **Boldfaced** / Underline: best/second best of all compared methods.

Model pre-training: In pre-training, we used the Adam optimizer for both MRI and CT IR. MRI was trained for 200 epochs, and CT was trained for 500 epochs, with a constant learning rate of 5e-4 throughout all epochs.

Model test-time adaptation: Following the procedure in [25, 29], the adaptation was conducted over $T = 200$ iterations for MRI images and $T = 100$ iterations for CT images. For MRI, we quote the results from the paper [25]. For CT, we reported reproducible results with hyper-parameter tuning-up.

4.1 Experiments on CS-MRI

In CS-MRI, the measurements for IR are obtained by sampling the DFT (k-space) of the image. The acquisition process involves the forward operator $\mathbf{\Phi} := \mathrm{M} \circ \mathrm{F}(\mathbf{x})$, where M is the fixed mask and F represents the Fourier transform. In the noiseless scenario, $\epsilon = 0$. For noisy settings, measurement noise is modeled as $\epsilon = \mathbf{\Phi}(\epsilon_1 + i\epsilon_2)$ with $\epsilon_1, \epsilon_2 \sim \mathcal{N}(0, 0.1 \max(\mathbf{x})^2 \mathbf{I})$. The model is pre-trained separately for both noisy settings, and the results are quantitatively assessed using the metrics including PSNR and SSIM [41].

The datasets used in this study are from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [16] and NYU fastMRI Initiative [47]. The ADNI dataset contains 300 images for training and 21 for test, where Gaussian and Radial mask with ratio $r = \frac{1}{3}, \frac{1}{4}, \frac{1}{5}$ are applied. For the fastMRI dataset, we train and test on a subset of the single-coil measurements at 4× and 8× acceleration. There are 200 images, where the first 180 images are for training and the last 20 are for testing. The masks used include 2D Gaussian, Poisson, and 1D Uniform.

The methods for performance comparison are selected from different categories of IR methods, including BNN [23], REI [2], AGLSD [38], DDSSL [29], MetaCS [25], GAN [16], Score-MRI [5], ADMMNet [45] and Restormer [46]. Note that for a fairer comparison with MetaCS [25], we increase its model size to match ours. Quantitative results are detailed in Tables 1 and Table 2, and visual comparisons are available in Figure 3.

It is evident that supervised methods do not exhibit a performance advantage over GT-free unsupervised or self-supervised deep learning approaches. This is not surprising as the number of training samples used in supervised learning

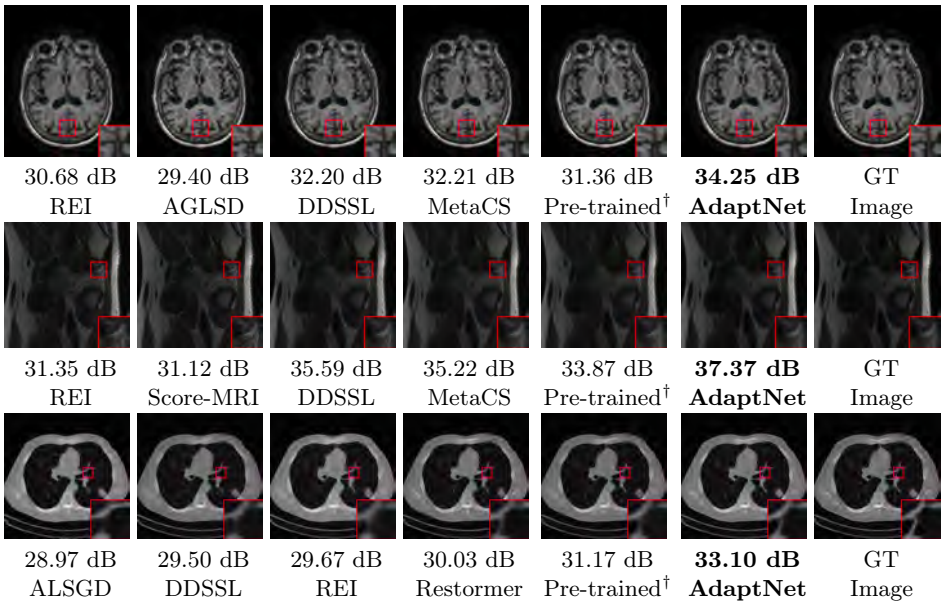

Fig. 3: Visual comparisons on reconstructed medical images. Top row: Images from ANDI dataset with Gaussian masks of CS ratio $r = 1/4$. Second row: Images from fastMRI dataset with Poisson mask of $4\times$ acceleration. Bottom rows: CT reconstruction with $I_0 = 1 \times 10^4$.

is limited, which does not comprehensively represent the distribution of target medical data. Consequently, the generalization of the trained model is not impressive. Generally, model adaptation yields the best performance. Its superiority over self-supervised methods stems from the simultaneous utilization of external training data via a pre-trained model and the exploitation of each test sample's unique characteristics through model adaptation.

Specifically, our proposed AdaptNet method demonstrates superior performance in most scenarios, highlighting its advantage over current model adaptation methods, particularly in challenging conditions with measurement noise. For noise-free cases, the pre-trained model already performs well, resulting in a relatively modest performance improvement through our proposed model adaptation. Refer to the supplementary material for more details.

4.2 Experiments on SV-CT

In sparse-view computed tomography (SV-CT) image reconstruction (IR), the measurement matrix is defined as $\mathbf{\Phi} := I_0 e^{-\text{radon}(\mathbf{x})}$, where radon refers to the sparse-view Radon transformation and I_0 denotes the X-ray source intensities. Considering the practical applications, our focus is mainly on IR for

noisy measurements. The noisy measurement is modeled as $\mathbf{y} = \mathbf{z} + \epsilon$, where $\mathbf{z} = \text{Poisson}(\mathbf{\Phi}(\mathbf{x}))$ and ϵ follows a normal distribution $\mathcal{N}(0, \sigma^2\mathbf{I})$.

For data synthesis, we adopt the reconstruction model from [2], where $\mathcal{F}(\mathbf{y}, \theta) = \mathbf{\Phi}_\theta \circ \text{iradon} \circ \log(I_0/\mathbf{y})$. MPG measurements with $\sigma = 30$ are generated in the same way as [2]. The radon function consists of 50 uniformly sampled views and four cases with varying X-ray source intensities $I_0 = 10^5, 5\times10^4, 10^4$, and 5×10^3 are considered. The methods for comparison includes Restormer [46], GAN [16],

Noise	I_0	Classic	Unsupervised			Supervised				Model Adaptation	
		FBP	REI	BNN	ASGLD	GAN	Restormer	FISTANet	Pre-trained†	DDSSL	AdaptNet
w/	1×10^5	23.22/.38	34.00/.91	32.78/.87	33.24/.86	30.21/.85	34.31/.91	<u>35.32/.91</u>	34.78/.92	33.71/.89	**36.61/.93**
	5×10^4	21.13/.30	33.68/.90	32.34/.87	32.84/.86	29.85/.83	33.83/.91	34.69/.91	34.68/.91	33.15/.87	**36.45/.93**
	1×10^4	14.87/.13	30.76/.85	29.49/.83	29.96/.83	26.89/.73	30.70/.85	31.47/.86	31.36/.86	30.02/.83	**32.58**/.86
	5×10^3	12.09/.08	28.50/.78	28.12/.75	28.67/.78	26.43/.71	30.17/.81	30.15/.84	<u>30.60/.84</u>	28.42/.77	**31.14**/.81

Table 3: Mean PSNR(dB)/SSIM values of CT reconstructions with different I_0. Pre-trained† here is the pre-trained model with the same backbone as DDSSL. **Boldfaced** / <u>Underline</u>: best / second best of all compared methods.

FISTANet [44], REI [2], AGLSD [38], BNN [22] and DDSSL [29]. Quantitative results are shown in Table 3, and visual comparisons in Figure 3. The same conclusion as in CS-MRI is made. Supervised learning methods lack performance advantages due to limited training samples. Model adaptation methods achieve the best performance. AdaptNet outperforms other model adaptation methods, especially with low X-ray source intensity.

Model	Params	Times	Radial			Gaussian			CT (view=50)		PSNR
	(k)	(s)	r =1/5	1/4	1/3	r =1/5	1/4	1/3	$I_0 = 10^4$	10^5	gain (Avg.)
Full-tuning	790 (100%)	14.29	30.14	31.19	32.41	31.15	32.30	33.14	32.35	35.78	-
Bias-tuning	0.768 (0.1%)	9.98	29.94	31.03	32.10	30.93	31.84	32.62	31.96	35.78	-0.28
Gain-tuning	3.1(3.39%)	8.67	29.77	30.98	32.01	30.87	31.76	32.56	32.17	35.59	-0.29
AdaptNet	0.234(0.03%)	11.69	31.68	32.88	34.13	33.01	33.61	34.82	32.58	36.61	+1.36

Table 4: Quantitative comparison in PSNR of applying different adaption techniques on the same pre-trained model for IR in CS-MRI.

4.3 Ablation Studies

Off-the-shelf model adaptions vs. AdaptNet: To assess the generalization and computational efficiency of AdaptNet, we compare it with full-tuning, which fine-tunes all pre-trained model weights in test time, and two popular efficient model adaptation techniques: bias-tuning and gain-tuning. The former fine-tunes the bias only, and the latter adjusts a single parameter (the gain) in each channel.

Refer to Table 4 for details. All methods adapt the same pre-trained model and utilize the same self-supervised loss. AdaptNet consistently outperforms both bias-tuning and gain-tuning, showing the effectiveness of applying a lightweight linear transform for quick distribution adjustment from trained to test data, supported by the mathematical understanding from the viewpoint of MMSE estimator. Also, AdaptNet updates the fewest weights. Indeed, as shown in Figure 4, AdaptNet requires only a fraction of the iterations (20 iterations) to achieve better performance than all other methods.

Model	Params	Radial			Gaussian		
		1/5	1/4	1/3	1/5	1/4	1/3
w/o tuning	-	29.38	29.82	30.35	29.86	30.10	30.39
AdaptNet-3	352(0.04%)	31.30	32.65	33.78	32.66	33.11	34.51
AdaptNet-2	234(0.03%)	31.68	32.88	34.13	33.01	33.61	34.82
AdaptNet-1	118(0.02%)	31.22	32.45	33.87	32.45	33.01	34.63

Table 5: Performance comparison in PSNR of the results from the AdaptBlock with different number of convolution layers.

Method	Radial			Gaussian		
r	1/5	1/4	1/3	1/5	1/4	1/3
w/o $\mathcal{L}^{null}$	31.58	32.55	33.59	32.52	33.31	34.18
AdaptNet	31.68	32.88	34.13	33.01	33.61	34.82

Table 6: Effect of using loss $\mathcal{L}^{null}$ during adaptation on ADNI dataset in noisy setting.

Dataset	FastMRI-Possion-x4			ADNI-Gaussian-$\frac{1}{4}$		ADNI-Radial-$\frac{1}{4}$	
size	200	400	800	100	300	100	300
w/o tuning	33.74	34.10	34.17	28.27	30.13	27.62	29.82
Full-tuning	35.58	35.60	35.64	31.97	32.30	31.25	31.19
AdaptNet	**36.51**	**36.58**	**36.72**	**32.62**	**33.31**	**31.91**	**32.88**

Table 7: Performance comparison across varying training dataset size in different MRI IR settings.

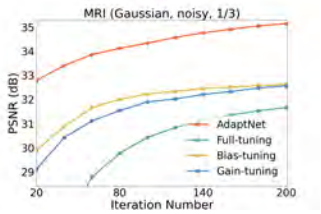

Fig. 4: PSNR curves vs. iteration number during adaptation for different methods.

Impact of the size of AdaptBlock: This experiment shows the relationship between AdaptBlock's model size and performance gains through model adaptation. Refer to Table 5 for the comparison with varying convolution layer used in AdaptBlock. Notably, AdaptNet-3 (3 layers) doesn't outperform AdaptNet-2 (2 layers). This could be due to the larger size of AdaptNet leading to overfitting on the single test sample. Moreover, AdaptNet-1 (1 layer) under-performs compared to AdaptNet-2, likely because its limited transformation capability is insufficient for transforming the training data's distribution.

Impact of using loss $\mathcal{L}^{null}$: To learn accurate prediction image in null space, we adopt the consistency loss $\mathcal{L}^{null}$ with Poisson noise injection. Table 6 shows that this loss leads to certain performance gains during the adaptation process.

Impact of training data size: To evaluate the performance gain of AdaptNet with respect to the model pre-trained with different training data sizes, we increased the number of training data from fastMRI to 400 and 800, keeping all other parameters unchanged, and reduced the number of images from ADNI to 100 from 300. Table 7 shows that AdaptNet consistently results in noticeable performance gains for the model pre-trained with various data sizes.

Method		Radial			Gaussian			PSNR
r		1/5	1/4	1/3	1/5	1/4	1/3	gap (avg.)
w/o tuning	AdaptNet*	28.52	29.00	29.68	28.99	29.49	29.78	-
	AdaptNet	29.38	29.82	30.35	29.86	30.13	30.41	-
w/ tuning	AdaptNet*	30.86	32.20	33.82	32.31	33.31	34.63	**+3.61**
	AdaptNet	31.68	32.88	34.13	33.01	33.61	34.82	+3.36

Table 8: Performance gain in PSNR when adapting a model pre-trained on natural images to MRI data.

4.4 Experiments on domain knowledge transfer

In this experiment, we consider adapting a model, pre-trained in one domain, such as natural images, to another, like medical images. There are two pre-trained models: one is trained on MRI data from previous experiments, and another on a natural images. We used 500 images from the Berkeley Segmentation Dataset (BSD) [18], cropping them into 1500 patches of 192×160 for size consistency with MRI images. The model adapted from MRI data is termed AdaptNet, while that from natural images is AdaptNet*. Table 8 compares the two. Initially, AdaptNet* does not perform as well as AdaptNet, which is expected due to the differences between natural images and MRI images. However, after model adaption by AdaptNet, the performance gap between the two narrows significantly, with AdaptNet* surpassing the pre-trained model of AdaptNet. This clearly indicates the effectiveness of AdaptNet in domain knowledge transfer.

4.5 More experiments and details in the supplementary file.

- The details of the derivations presented in Section 3 for completeness.
- Quantitative results on MRI reconstruction in the noiseless case.
- Visual inspection on more examples in CT and MRI.
- Studies on the capability of AdaptNet for adapting the pre-trained model to different measurement matrices.
- The effectiveness of AdaptNet for adapting a model pre-trained on MRI images to natural images.

5 Conclusion and Discussion

This paper introduces AdaptNet, a TTA approach for IR from incomplete measurements, inspired by the connection between linear denoising CNNs and Wiener filtering. By embedding a lightweight linear convolution layer into each stage of an unrolled NN, AdaptNet adapts a pre-trained model to each test sample for improved performance. Our experiments on two medical IR tasks, CS-MRI and SV-CT, confirm its effectiveness and demonstrate its capacity for processing out-of-distribution test samples. Future research will explore the theoretical foundations of model adaptation for broader estimators and its potential across more tasks and architectures.

References

1. Chen, D., Tachella, J., Davies, M.E.: Equivariant imaging: Learning beyond the range space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4379–4388 (2021) 2, 5
2. Chen, D., Tachella, J., Davies, M.E.: Robust equivariant imaging: a fully unsupervised framework for learning to image from noisy and partial measurements. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5647–5656 (2022) 2, 10, 12
3. Chi, Z., Wang, Y., Yu, Y., Tang, J.: Test-time fast adaptation for dynamic scene deblurring via meta-auxiliary learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9137–9146 (2021) 2, 5
4. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. arXiv preprint arXiv:2209.14687 (2022) 4
5. Chung, H., Ye, J.C.: Score-based diffusion models for accelerated mri. Medical Image Analysis **80**, 102479 (2022) 10
6. Darestani, M.Z., Liu, J., Heckel, R.: Test-time training can close the natural distribution shift performance gap in deep learning based compressed sensing. In: Proceedings of the 39th International Conference on Machine Learning. vol. 162, pp. 4754–4776. PMLR (2022) 2, 5
7. Diamond, S., Sitzmann, V., Heide, F., Wetzstein, G.: Unrolled optimization with deep priors. arXiv preprint arXiv:1705.08041 (2017) 3, 6, 7
8. Ding, Q., Chen, G., Zhang, X., Huang, Q., Ji, H., Gao, H.: Low-dose ct with deep learning regularization via proximal forward–backward splitting. Physics in Medicine & Biology **65**(12), 125009 (2020) 4
9. Ding, Q., Nan, Y., Gao, H., Ji, H.: Deep learning with adaptive hyper-parameters for low-dose ct image reconstruction. IEEE Transactions on Computational Imaging pp. 648–660 (2021) 4
10. Feng, C.M., Yang, Z., Chen, G., Xu, Y., Shao, L.: Dual-octave convolution for accelerated parallel mr image reconstruction. In: Proceedings of AAAI Conference on Artificial Intelligence. vol. 35, pp. 116–124 (2021) 2
11. Jin, K.H., McCann, M.T., Froustey, E., Unser, M.: Deep convolutional neural network for inverse problems in imaging. IEEE Transactions on Image Processing **26**(9), 4509–4522 (2017) 4
12. Laine, S., Karras, T., Lehtinen, J., Aila, T.: High-quality self-supervised deep image denoising. In: Advances in Neural Information Processing Systems. vol. 32. Curran Associates, Inc. (2019) 5
13. Lee, H., Lee, J., Kim, H., Cho, B., Cho, S.: Deep-neural-network-based sinogram synthesis for sparse-view ct image reconstruction. IEEE Transactions on Radiation and Plasma Medical Sciences **3**(2), 109–119 (2018) 2
14. Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., Aila, T.: Noise2noise: Learning image restoration without clean data. In: Proceedings of the 35th International Conference on Machine Learning. vol. 80, pp. 2965–2974. PMLR (2018) 5
15. Li, Z., Zhang, T., Wan, P., Zhang, D.: Segan: Structure-enhanced generative adversarial network for compressed sensing mri reconstruction. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 2128–2136 (2023) 4

16. Liu, J., Kuang, T., Zhang, X.: Image reconstruction by splitting deep learning regularization from iterative inversion. In: Medical Image Computing and Computer Assisted Intervention. pp. 224–231 (2018) 4, 8, 10, 12
17. Mansour, Y., Heckel, R.: Zero-shot noise2noise: Efficient image denoising without any data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14018–14027 (2023) 5
18. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings Eighth IEEE International Conference on Computer Vision. vol. 2, pp. 416–423 vol.2 (2001) 14
19. Meinhardt, T., Moller, M., Hazirbas, C., Cremers, D.: Learning proximal operators: Using denoising networks for regularizing inverse imaging problems. In: Proceedings of the IEEE International Conference on Computer Vision (2017) 4
20. Metzler, C.A., Mousavi, A., Heckel, R., Baraniuk, R.G.: Unsupervised learning with stein's unbiased risk estimator. arXiv preprint arXiv:1805.10531 (2018) 5
21. Mohan, S., Vincent, J.L., Manzorro, R., Crozier, P., Fernandez-Granda, C., Simoncelli, E.: Adaptive denoising via gaintuning. In: Advances in Neural Information Processing Systems. vol. 34, pp. 23727–23740. Curran Associates, Inc. (2021) 5
22. Pang, T., Quan, Y., Ji, H.: Self-supervised bayesian deep learning for image recovery with applications to compressive sensing. In: Proceedings of European Conference on Computer Vision. pp. 475–491 (2020) 2, 5, 12
23. Pang, T., Zheng, H., Quan, Y., Ji, H.: Recorrupted-to-recorrupted: unsupervised deep learning for image denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2043–2052 (2021) 2, 3, 4, 5, 9, 10
24. Park, S., Yoo, J., Cho, D., Kim, J., Kim, T.H.: Fast adaptation to super-resolution networks via meta-learning. In: Proceedings of European Conference on Computer Vision. pp. 754–769 (2020) 2, 5
25. Qin, X., Quan, Y., Pang, T., Ji, H.: Ground-truth free meta-learning for deep compressive sampling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9947–9956 (2023) 2, 3, 5, 10
26. Quan, Y., Chen, M., Pang, T., Ji, H.: Self2self with dropout: Learning self-supervised denoising from single image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020) 5
27. Quan, Y., Chen, Z., Pang, T., Ji, H.: Unsupervised deep learning for phase retrieval via teacher-student distillation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 2128–2136 (2023) 5
28. Quan, Y., Chen, Z., Zheng, H., Ji, H.: Learning deep non-blind image deconvolution without ground truths. In: Proceedings of European Conference on Computer Vision. pp. 642–659 (2022) 5
29. Quan, Y., Qin, X., Pang, T., Ji, H.: Dual-domain self-supervised learning and model adaption for deep compressive imaging. In: Proceedings of European Conference on Computer Vision. pp. 409–426 (2022) 2, 3, 5, 7, 9, 10, 12
30. Quan, Y., Qin, X., Pang, T., Ji, H.: Siamese cooperative learning for unsupervised image reconstruction from incomplete measurements. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(7), 4866–4879 (2024) 5
31. Romano, Y., Elad, M., Milanfar, P.: The little engine that could: Regularization by denoising (red). SIAM Journal on Imaging Sciences **10**(4), 1804–1844 (2017) 4
32. Sidky, E.Y., Kao, C.M., Pan, X.: Accurate image reconstruction from few-views and limited-angle data in divergent-beam ct. Journal of X-ray Science and Technology **14**(2), 119–139 (2006) 2

33. Song, J., Vahdat, A., Mardani, M., Kautz, J.: Pseudoinverse-guided diffusion models for inverse problems. In: International Conference on Learning Representations (2023) 4
34. Song, Y., Shen, L., Xing, L., Ermon, S.: Solving inverse problems in medical imaging with score-based generative models. In: International Conference on Learning Representations (2022) 4
35. Sun, Y., Wang, X., Liu, Z., Miller, J., Efros, A., Hardt, M.: Test-time training with self-supervision for generalization under distribution shifts. In: Proceedings of the 37th International Conference on Machine Learning. vol. 119, pp. 9229–9248. PMLR (2020) 2, 5
36. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9446–9454 (2018) 2, 5
37. Wakin, M.B., et al.: An introduction to compressive sampling. IEEE Signal Processing Magazine **25**(2), 21–30 (2008) 1
38. Wang, W., Li, J., Ji, H.: Self-supervised deep image restoration via adaptive stochastic gradient langevin dynamics. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1989–1998 (2022) 2, 5, 10, 12
39. Wang, Z., Liu, J., Li, G., Han, H.: Blind2unblind: Self-supervised image denoising with visible blind spots. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2027–2036 (2022) 5
40. Wang, Z., Zhang, H., Cheng, Z., Chen, B., Yuan, X.: Metasci: Scalable and adaptive reconstruction for video compressive sensing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2083–2092 (2021) 2, 5
41. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. vol. 2, pp. 1398–1402. Ieee (2003) 10
42. Wu, W., Hu, D., Niu, C., Yu, H., Vardhanabhuti, V., Wang, G.: Drone: Dual-domain residual-based optimization network for sparse-view ct reconstruction. IEEE Transactions on Medical Imaging **40**(11), 3002–3014 (2021) 2
43. Wu, Y., Rosca, M., Lillicrap, T.: Deep compressed sensing. In: International Conference on Machine Learning. pp. 6850–6860. PMLR (2019) 2
44. Xiang, J., Dong, Y., Yang, Y.: Fista-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. IEEE Transactions on Medical Imaging **40**(5), 1329–1339 (2021) 2, 4, 12
45. Yang, Y., Sun, J., Li, H., Xu, Z.: Admm-csnet: A deep learning approach for image compressive sensing. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(3), 521–538 (2018) 2, 4, 10
46. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5728–5739 (2022) 10, 12
47. Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M.J., Defazio, A., Stern, R., Johnson, P., Bruno, M., et al.: fastmri: An open dataset and benchmarks for accelerated mri. arXiv preprint arXiv:1811.08839 (2018) 10
48. Zhang, J., Ghanem, B.: Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1828–1837 (2018) 2, 4

49. Zhang, Z., Liang, X., Dong, X., Xie, Y., Cao, G.: A sparse-view ct reconstruction method based on combination of densenet and deconvolution. IEEE Transactions on Medical Imaging **37**(6), 1407–1417 (2018) 2
50. Zhu, M., Li, S., Li, D., Gao, Q., Bian, Z., Huang, J., Zeng, D., Ma, J.: Teacher-student network for ct image reconstruction via meta-learning strategy. In: IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). pp. 1–3 (2019) 2, 5
51. Zhussip, M., Soltanayev, S., Chun, S.Y.: Training deep learning based image denoisers from undersampled measurements without ground truth and without image prior. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10255–10264 (2019) 2, 5

Test-Time Model Adaptation for Image Reconstruction Using Self-supervised Adaptive Layers (Supplementary Materials)

Yutian Zhao[1], Tianjing Zhang[1], and Hui Ji[1]

Department of Mathematics, National University of Singapore, 119076, Singapore
{e0708171,tianjingzhang}@u.nus.edu, matjh@nus.edu.sg

1 Derivation of $\hat{h}_{\omega}$ in Section 3

Proof. Since

$$\hat{h}_{\omega} = \text{argmin}_{\widehat{h}} \mathbb{E}\{\|\widehat{h} \odot \widehat{\mathbf{z}} - \widehat{\mathbf{x}}\|_F^2\} = \text{argmin}_{\widehat{h}} \sum_{k_1} \sum_{k_2} \mathbb{E}\{|\widehat{h}[k_1, k_2]\widehat{\mathbf{z}}[k_1, k_2] - \widehat{\mathbf{x}}[k_1, k_2]|^2\} \tag{1}$$

Define

$$\mathbf{e}(k_1, k_2) = \mathbb{E}\{|\widehat{h}[k_1, k_2]\widehat{\mathbf{z}}[k_1, k_2] - \widehat{\mathbf{x}}[k_1, k_2]|^2\} \tag{2}$$

Expanding it further, we can get:

$$\begin{aligned}
\mathbf{e}(k_1, k_2) =& \mathbb{E}_{\mathbf{x},\mathbf{n}}\{|\widehat{h}[k_1, k_2](\widehat{\mathbf{x}}[k_1, k_2] + \widehat{\mathbf{n}}[k_1, k_2]) - \widehat{\mathbf{x}}[k_1, k_2]|^2\} \\
=& \mathbb{E}_{\mathbf{x},\mathbf{n}}\{|\widehat{\mathbf{x}}[k_1, k_2](\widehat{h}[k_1, k_2] - 1) + \widehat{h}[k_1, k_2]\widehat{\mathbf{n}}[k_1, k_2]|^2\} \\
=& (\widehat{h}[k_1, k_2] - 1)(\widehat{h}[k_1, k_2] - 1)^* \mathbb{E}_{\mathbf{x}}\{|\widehat{\mathbf{x}}[k_1, k_2]|^2\} \\
& - (\widehat{h}[k_1, k_2] - 1)\widehat{h}^*[k_1, k_2]\mathbb{E}_{\mathbf{x},\mathbf{n}}\{\widehat{\mathbf{x}}[k_1, k_2]\widehat{\mathbf{n}}^*[k_1, k_2]\} \\
& - (\widehat{h}[k_1, k_2] - 1)^* \widehat{h}[k_1, k_2]\mathbb{E}_{\mathbf{x},\mathbf{n}}\{\widehat{\mathbf{x}}^*[k_1, k_2]\widehat{\mathbf{n}}[k_1, k_2]\} \\
& + \widehat{h}[k_1, k_2]\widehat{h}^*[k_1, k_2]\mathbb{E}_{\mathbf{n}}\{|\widehat{\mathbf{n}}[k_1, k_2]|^2\}
\end{aligned}$$

where $*$ denotes the complex conjugation. Based on the assumption that $\mathbf{x}$ and $\mathbf{n}$ are independent and $\mathbb{E}[\mathbf{n}] = 0$, we have:

$$\mathbb{E}_{\mathbf{x},\mathbf{n}}\{\widehat{\mathbf{x}}^*[k_1, k_2]\widehat{\mathbf{n}}[k_1, k_2]\} = \mathbb{E}_{\mathbf{x},\mathbf{n}}\{\widehat{\mathbf{x}}[k_1, k_2]\widehat{\mathbf{n}}^*[k_1, k_2]\} = 0$$

Denote the real and imaginary part of $\hat{h}[k_1, k_2]$ as $\hat{h}_R[k_1, k_2]$ and $\hat{h}_I[k_1, k_2]$ respectively, then we calculate the Wirtinger derivative with respect to $\hat{h}[k_1, k_2]$ and let

$$\frac{\partial \mathbf{e}(k_1, k_2)}{\partial \widehat{h}[k_1, k_2]} = \frac{1}{2}\left(\frac{\partial}{\partial \widehat{h}_R} - i\frac{\partial}{\partial \widehat{h}_I}\right)\mathbf{e}(k_1, k_2) = 0$$

which leads to:

$$\begin{aligned}
&(\widehat{h}_R[k_1, k_2] - 1)\mathbb{E}_{\mathbf{x}}\{|\widehat{\mathbf{x}}[k_1, k_2]|^2\} + \widehat{h}_R[k_1, k_2]\mathbb{E}_{\mathbf{n}}\{|\widehat{\mathbf{n}}[k_1, k_2]|^2\} \\
&- i\widehat{h}_I[k_1, k_2]\mathbb{E}_{\mathbf{x}}\{|\widehat{\mathbf{x}}[k_1, k_2]|^2\} - i\widehat{h}_I[k_1, k_2]\mathbb{E}_{\mathbf{n}}\{|\widehat{\mathbf{n}}[k_1, k_2]|^2\} = 0
\end{aligned}$$

Recall that $\widehat{h}^*[k_1,k_2] = \widehat{h}_R[k_1,k_2] - i\widehat{h}_I[k_1,k_2]$, then

$$(\widehat{h}[k_1,k_2]-1)^*\mathbb{E}_{\mathbf{x}}\{|\widehat{\mathbf{x}}[k_1,k_2]|^2\} + \widehat{h}^*[k_1,k_2]\mathbb{E}_{\mathbf{n}}\{|\widehat{\mathbf{n}}[k_1,k_2]|^2\} = 0$$

Thus

$$\widehat{h}_\omega[k_1,k_2] = \frac{\mathbb{E}_{\mathbf{x}}\{|\widehat{\mathbf{x}}[k_1,k_2]|^2\}}{\mathbb{E}_{\mathbf{x}}\{|\widehat{\mathbf{x}}[k_1,k_2]|^2\} + \mathbb{E}_{\mathbf{n}}\{|\widehat{\mathbf{n}}[k_1,k_2]|^2\}} \tag{3}$$

The proof is done.

2 Model adaption across different measurement matrices

The effectiveness of the proposed model in transferring domain knowledge has been validated in the experiments presented in the main paper. In those experiments, the proposed method can effectively adapt a model pre-trained on natural images to medical imaging with significant performance gain.

In this section, we introduce additional ablation studies on its transferability. The experimental settings are as follows: For CS-MRI, different images are captured using various measurement matrices. Since these matrices are integrated into an unrolling network, it is intriguing to assess whether the proposed method can effectively adapt a model, initially trained with data obtained from one measurement matrix, to test data captured with a different measurement matrix.

Train \ Test	Gaussian, 1/5		Gaussian, 1/3	
	w/o adaption	w/ adaption	w/o adaption	w/ adaption
Gaussian, 1/5	29.86	33.01	30.43	34.98
Gaussian, 1/3	24.48	31.55	30.39	34.82

Table 1: Mean PSNR(dB) values of MRI reconstructions with Gaussian masks in cross-CS ratio adaption test.

Train \ Test	Radial, 1/5		Radial, 1/3	
	w/o adaption	w/ adaption	w/o adaption	w/ adaption
Radial, 1/5	29.38	31.68	30.40	34.34
Radial, 1/3	23.35	30.61	30.35	34.14

Table 2: Mean PSNR(dB) values of MRI reconstructions with Radial masks in cross-CS ratio adaption test.

Train \ Test	Gaussian, 1/4		Radial, 1/4	
	w/o adaption	w/ adaption	w/o adaption	w/ adaption
Gaussian, 1/4	30.10	33.61	28.77	32.37
Radial, 1/4	29.90	30.06	29.82	32.88

Table 3: Mean PSNR(dB) values of MRI reconstructions in cross-mask adaption test.

The first experiment evaluates the adaptability of the model when both the training and test data use the same type of mask but with varying sampling ratios. The results, presented in Table 1 for Gaussian masks and Table 2 for Radial masks on ADNI dataset, indicate that the model, when adapted using the proposed method, still achieves a noticeable performance improvement.

The second experiment assesses the model's adaptability when the training and testing data utilize different types of masks. The findings, detailed in Table 3, come to the same conclusion: the model, adapted through the proposed method, consistently leads to significant performance gains.

3 MRI reconstruction in noiseless setting

In this section, we present the CS-MRI reconstruction results of the ADNI dataset in a noise-free setting. Same as the noisy setting presented in the main paper, we utilized two types of down-sampling patterns with sampling ratios of r = $1/5$, $1/4$, and $1/3$, which are specifically Gaussian and Radial patterns. For the pre-trained stage, 300 MRI images are used, and the evaluation is conducted using a test set of 21 MRI images. Compared methods include one classic method SparseMRI [9], 3 unsupervised methods, *i.e.*, BNN [3], REI [1] and AGLSD [6], 2 supervised methods, *i.e.*, ADMMNet [7] and Restormer [8], and most related model adaption methods DDSSL [5], MetaCS [4]. In this case, the pre-trained model already performs well, resulting in a relatively modest performance improvement through our proposed model adaptation. However, according to the results presented in Table 4, our AdaptNet significantly outperformed the traditional SparseMRI method as well as the unsupervised methods, including REI, BNN, and ASGLD. When compared to supervised learning and model adaption approaches, AdaptNet was highly competitive. From Figure 1,2, AdaptNet also achieved higher visual quality than other compared methods.

4 More Visual Results

4.1 More Visual Comparisons with State-of-the-Art Methods of MRI Reconstruction

Additional visual results on MRI images are showcased in Figure 3,4,5. Specifically, we display more results under the noisy setting. For ADNI dataset, we consider Gaussian

Mask	CS Ratio	Classic	Unsupervised			Supervised			Model Adaptation		
		SparseMRI	REI	BNN	ASGLD	ADMMNet	Restormer	Pre-trained†	DDSSL	MetaCS	AdaptNet
Gaussian	1/3	34.93/.93	37.03/.93	37.60/.94	37.79/.94	38.22/.98	37.66/.97	<u>39.19/.97</u>	37.82/.95	37.88/.96	**39.64/.98**
	1/4	32.79/.90	36.11/.95	36.10/.95	36.07/.95	35.94/.96	35.53/.95	<u>36.80/.96</u>	36.72/.96	36.73/.96	**37.46/.97**
	1/5	31.69/.89	34.01/.92	33.81/.93	34.40/.92	34.81/.96	34.41/.94	<u>35.64/.95</u>	35.46/.95	35.55/.94	**35.87/.96**
Radial	1/3	34.58/.92	36.16/.94	35.58/.94	35.79/.94	35.31/.94	35.12/.93	<u>36.75/.96</u>	36.43/.95	36.59/.95	**37.02/.96**
	1/4	32.31/.90	33.49/.92	34.08/.95	34.38/.95	33.70/.93	33.35/.93	<u>34.90/.96</u>	34.70/.95	34.86/.96	**34.98/.96**
	1/5	30.72/.86	32.35/.90	32.28/.92	32.40/.92	32.32/.92	32.00/.93	33.59/.94	<u>33.97/.95</u>	**34.11/.95**	33.71/.94

Table 4: Mean PSNR(dB)/SSIM values of MRI reconstructions on ADNI datasets in noisless setting. Pre-trained† is the pre-trained model with the same backbone as DDSSL and MetaCS. **Boldfaced**/<u>Underline</u>: best/second best of all compared methods.

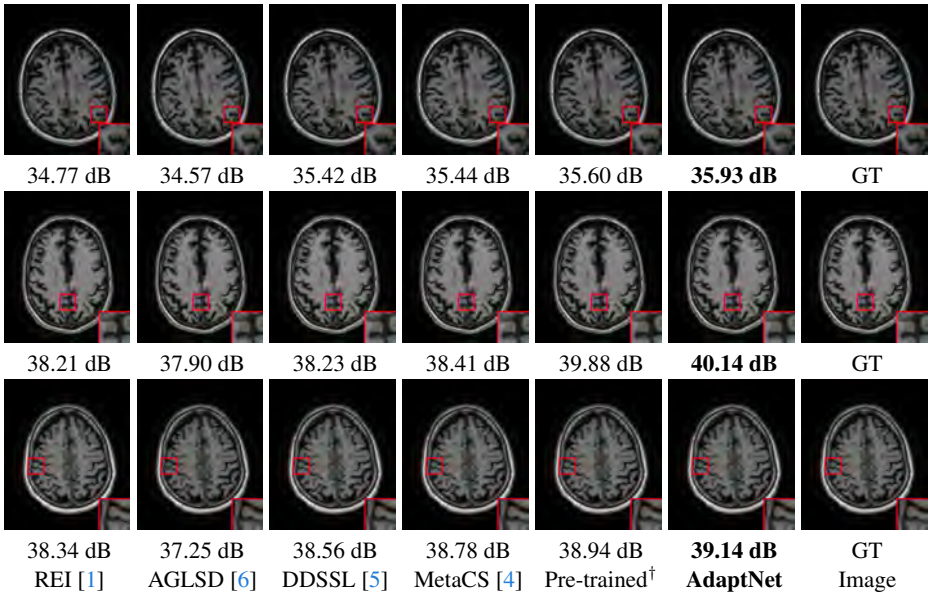

Fig. 1: Visual comparisons on reconstructed MRI images with Gaussian mask in noiseless setting. **Top row:** Sampling ratio $r = 1/5$. **Second row:** $r = 1/4$. **Bottom row:** $r = 1/3$

and Radial mask with varying sampling ratio $r = 1/5, 1/4, 1/3$. For fastMRI dataset, we compare against other methods under $8\times$ acceleration with Poisson, Gaussian and Uniform mask. As can be seen in Fig 5, our adaption process helps to preserve more details than the pre-trained model. Furthermore, these visual comparisons demonstrate the superior performance of our method compared to other methods in terms of visual quality. The results from our proposed method exhibit fewer artifacts than those from the other methods.

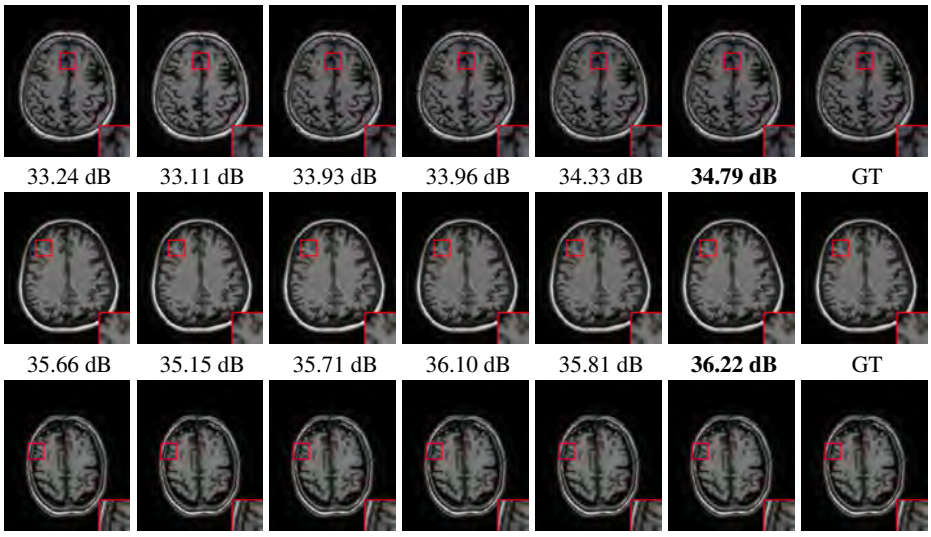

Fig. 2: Visual comparisons on reconstructed MRI images with Radial mask in noiseless setting. **Top row:** Sampling ratio $r = 1/5$. **Second row:** $r = 1/4$. **Bottom row:** $r = 1/3$

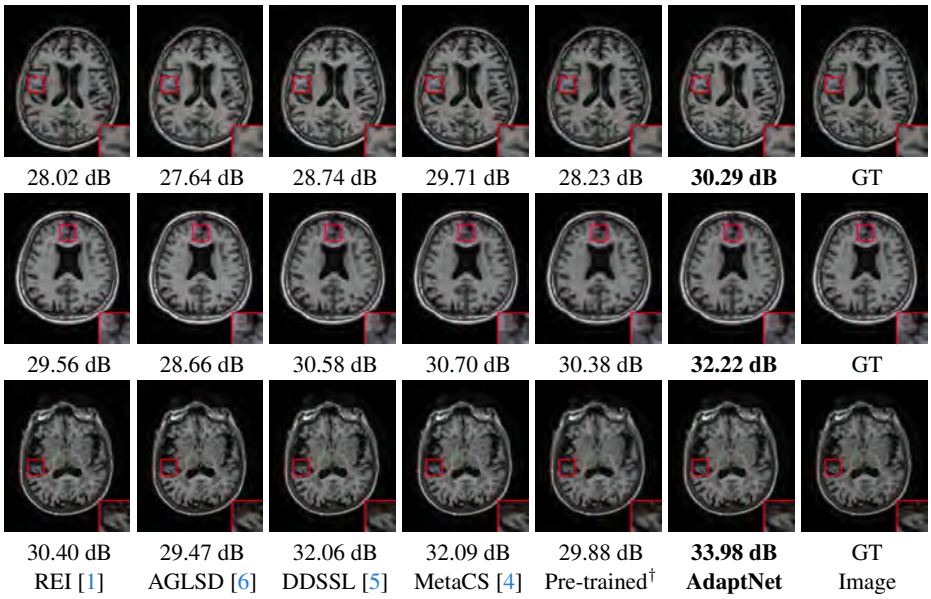

Fig. 3: Visual comparisons on reconstructed MRI images from ADNI dataset with Radial mask in noisy setting. **Top row:** Sampling ratio $r = 1/5$; **Second row:** $r = 1/4$; **Bottom row:** $r = 1/3$.

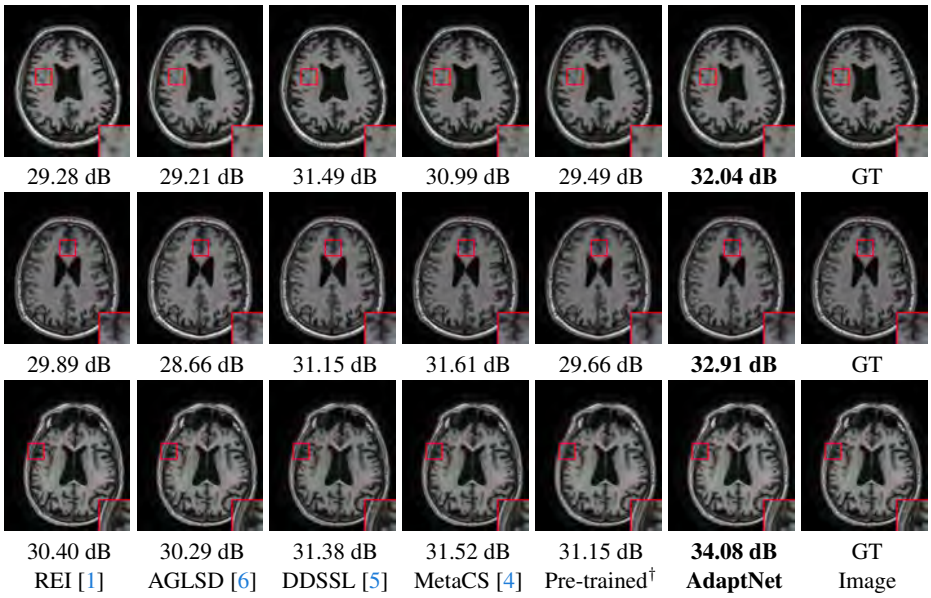

Fig. 4: Visual comparisons on reconstructed MRI images from ADNI dataset with Gaussian mask in noisy setting. **Top row:** Sampling ratio $r = 1/5$. **Second row:** $r = 1/4$. **Bottom row:** $r = 1/3$.

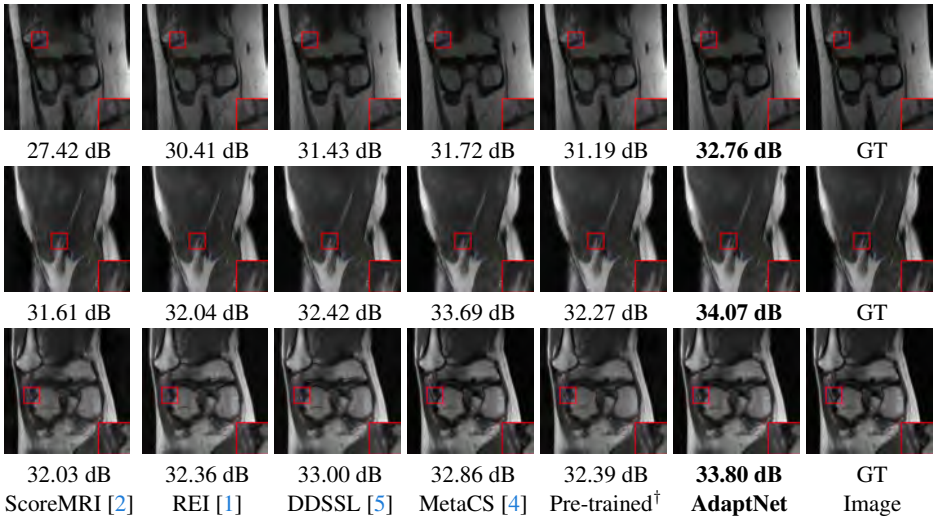

Fig. 5: Visual comparisons on reconstructed MRI images from fastMRI dataset of $8\times$ acceleration in the noisy setting. **Top row:** Poisson mask. **Second row:** Uniform Mask.**Bottom row:** Gaussian mask.

4.2 More Visual Comparisons with State-of-the-Art Methods of CT Reconstruction

In addition to those in the main paper, we provide extensive visual comparisons, we display more visual results for SV-CT images in Figure 6. Our proposed method still outperforms other methods, in terms of visual quality.

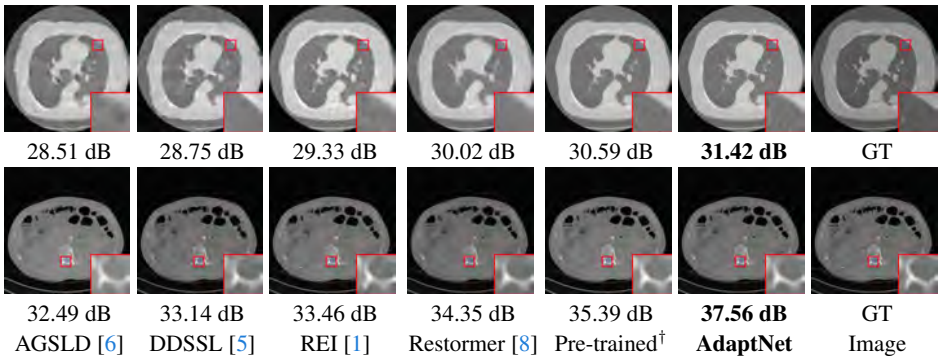

Fig. 6: Visual comparisons on reconstructed CT images. **Top row:** Images with $I_0 = 5 \times 10^3$. **Bottom row:** Images with $I_0 = 5 \times 10^4$.

Method		Radial			Gaussian		
r		1/5	1/4	1/3	1/5	1/4	1/3
MRI⟶ BSD's testing	w/o tuning	27.04	27.62	27.78	28.02	28.30	28.71
	w/. tuning	**28.11**	**29.32**	**29.88**	**29.95**	**30.52**	**31.38**
BSD's training⟶ BSD' testing	w/o tuning	28.16	28.76	29.57	29.16	29.56	29.95
	w/. tuning	**29.52**	**30.82**	**32.62**	**31.63**	**32.32**	**33.49**

Table 5: Performance gain in PSNR(dB) when adapting a model pre-trained on MRI images to natural testing samples.

5 More experiements on Domain Knowleged Tranfer

In this experiments, we evaluae the effectiveness of the AdaptNet on knowlege transfer from natural images to medical imaging. That is we consider adapting a model, pre-trained on natural images to test medical images. We split the BSD, natual image dataset, into 480 training data and 20 testing data and consider adapting natural images to a model pretrained on medical images. We considered two scenarios: (1) adapting

a model pre-trained on training data from BSD dataset to medical data. (2) adapting a model pre-trained on training data from BSD dataset to testing data also from BSD dataset. Table 5 demonstrates that, same as the domain knowlege transfer from natural images to medical images, our adaptation method resulted in substantial performance gains in PSNR value.

References

1. Chen, D., Tachella, J., Davies, M.E.: Robust equivariant imaging: a fully unsupervised framework for learning to image from noisy and partial measurements. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5647–5656 (2022) 3, 4, 5, 6, 7
2. Chung, H., Ye, J.C.: Score-based diffusion models for accelerated mri. Medical Image Analysis **80**, 102479 (2022) 6
3. Pang, T., Zheng, H., Quan, Y., Ji, H.: Recorrupted-to-recorrupted: unsupervised deep learning for image denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2043–2052 (2021) 3
4. Qin, X., Quan, Y., Pang, T., Ji, H.: Ground-truth free meta-learning for deep compressive sampling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9947–9956 (2023) 3, 4, 5, 6
5. Quan, Y., Qin, X., Pang, T., Ji, H.: Dual-domain self-supervised learning and model adaption for deep compressive imaging. In: Proceedings of European Conference on Computer Vision. pp. 409–426 (2022) 3, 4, 5, 6, 7
6. Wang, W., Li, J., Ji, H.: Self-supervised deep image restoration via adaptive stochastic gradient langevin dynamics. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1989–1998 (2022) 3, 4, 5, 6, 7
7. Yang, Y., Sun, J., Li, H., Xu, Z.: Admm-csnet: A deep learning approach for image compressive sensing. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(3), 521–538 (2018) 3
8. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5728–5739 (2022) 3, 7
9. Zhang, Z., Liang, X., Dong, X., Xie, Y., Cao, G.: A sparse-view ct reconstruction method based on combination of densenet and deconvolution. IEEE Transactions on Medical Imaging **37**(6), 1407–1417 (2018) 3