Domain Expansion via Network Adaptation for Solving Inverse Problems

Anonymous submission

Abstract

Deep learning-based methods deliver state-of-the-art performance for solving inverse problems that arise in computational imaging. These methods can be broadly divided into two groups: (1) learn a network to map measurements to the signal estimate, which is known to be fragile; (2) learn a prior for the signal to use in an optimization-based recovery. Despite the impressive results from the latter approach, many of these methods also lack robustness to shifts in data distribution, measurements, and noise levels. Such domain shifts result in a performance gap and in some cases introduce undesired artifacts in the estimated signal. In this paper, we explore the qualitative and quantitative effects of various domain shifts and propose a flexible and parameter efficient framework that adapt pre-trained networks to such shifts. We demonstrate the effectiveness of our method for a number of natural image, MRI and CT reconstructions tasks under domain, measurement model, and noise-level shifts. Our experiments demonstrate that our method provides significantly better performance and parameter efficiency compared to existing domain adaptation techniques.

1 Introduction

Linear inverse problems arise in many real world applications. For instance, image enhancement and restoration tasks in denoising, deblurring, and super-resolution or medical image reconstruction from indirect (Fourier) measurements in computed tomography (CT) and magnetic resonance imaging (MRI). We can model such inverse problems as recovery of an unknown signal $\mathbf{x}$ from a set of measurements

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \eta, \tag{1}$$

where $\mathbf{y}$ represents measurements, $\mathbf{A}$ represents an $m \times n$ measurement matrix or forward operator, and η represents noise. The unknown signal and measurements can be real- or complex-valued. To recover $\mathbf{x}$, we can solve an optimization problem of the following form:

$$\min_{\mathbf{x}} g(\mathbf{x}) + h_\theta(\mathbf{x}), \tag{2}$$

where $g(\mathbf{x})$ is a data fidelity term (e.g., $g(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$), $h_\theta(\cdot)$ denotes a regularization function that enforces some prior constraint on the unknown signal, and θ denotes the regularization function parameters (Venkatakrishnan, Bouman, and Wohlberg 2013; Boyd et al. 2011). For instance, signal is sparse or low-rank in some representation space or belongs to a manifold of natural images (Candes, Romberg, and Tao 2006; Mairal et al. 2009; Lustig, Donoho, and Pauly 2007; Elad and Aharon 2006).

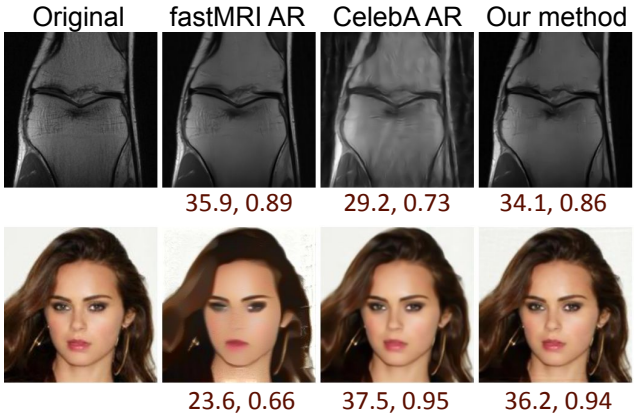

Figure 1: *Examples of image reconstruction under domain shifts. We use two artifact removal (AR) network trained for fastMRI and celebA images in deep unrolled framework. Second and third columns show both images reconstructed with fastMRI AR and celebA AR, respectively. Reconstruction quality degrades with domain shifts (PSNR and SSIM reported under each image). Our proposed network adaptation method, where we adapt celebA AR to recover an MR image (top row) and fastMRI AR to recover a celebA image (bottom row).*

In the deep learning era, we can recover $\mathbf{x}$ by either training a deep (reconstruction) network that maps measurements to the signal estimate or solving an iterative optimization problem (similar to the one in (2)) that can also be represented as an unrolled network (Ongie et al. 2020). While training a reconstruction network in an end-to-end manner is possible, it usually requires a large set of input-output training pairs $(\mathbf{y}, \mathbf{x})$. Furthermore, since these networks do not explicitly use the forward model in (1), they are known to be sensitive to small changes in the data distribution, measurement operators, and noise (Antun et al. 2020; Gottschling et al. 2020). Solving the optimization problem in (2) with an appropriate choice of regularization function $h(\cdot)$ is often considered a flexible and relatively robust option.

In recent years, deep networks are often used to represent $h(\cdot)$ instead of hand-designed functions (e.g., ℓ_1 norm

or total variation). For instance, deep unrolling (Gregor and LeCun 2010; Monga, Li, and Eldar 2021; Liu et al. 2021) and plug-and-play (PnP) (Venkatakrishnan, Bouman, and Wohlberg 2013; Sun, Wohlberg, and Kamilov 2019) methods use artifact-removal or image denoising networks that are trained to map a noisy or corrupted estimate of an image onto a clean image manifold (Ongie et al. 2020; Monga, Li, and Eldar 2021; Liu et al. 2021). Despite recent success of such deep unrolled or PnP methods, they are also sensitive to shifts in the data distribution (Darestani, Chaudhari, and Heckel 2021). Figure 1 illustrates this effect for deep unrolling with artifact removal (AR) networks under domain and forward model shifts. The fastMRI AR is trained while solving (2) for MR image reconstruction from radially under-sampled k-space measurements. CelebA AR is trained while solving (2) to reconstruct face images from random projections. Note that reconstructing MR images using the CelebA AR and vice versa results in a significant performance degradation (off-diagonal images in column 2 and 3).

In this paper, we propose a parameter-efficient method to adapt pretrained networks to multiple domains, measurement models, and noise with little to no drop in performance. In particular, we propose a domain-specific modulation of network weights using low-rank (or rank-one) factors. Given a single deep unrolled network, we learn a compact set of modulation parameters for each domain/measurement/noise setting, and adapt the weights of the network according to the specific problem at the inference time. In the remainder of the paper, we use the term domain shift and domain adaptation to refer to changes in data/measurement/noise distributions. Figure 2 illustrate our proposed unrolled multi-domain network that conditionally applies weight modulations to solve a diverse set of inverse problems. We present a set of experiments to demonstrate the effectiveness of our method in adapting the deep unrolled network for shifts in data distribution/domain ($\mathbf{x}$), forward models ($\mathbf{A}$), and noise levels (η). Last column in Fig. 1 presents an example of our method, where we adapt a pretrained celebA AR to fastMRI in first row and adapt fastMRI AR to celebA in second row. Our method recovers images that qualitatively and quantitatively resemble results of the networks trained for the correct domains. The number of parameters needed to adapt the pretrained network is less than 0.5% of the parameters in the pretrained network.

Our method can be viewed as an example of domain adaptation or domain expansion technique, where we update a network trained for a source domain to perform well on several target domains. Fine-tuning pre-trained networks is a widely used method for domain adaptation but suffers from catastrophic forgetting (McCloskey and Cohen 1989) and requires a large number of parameters for every new domain (Mallya, Davis, and Lazebnik 2018). Several parameter efficient domain adaptation techniques have been proposed in (Alanov, Titov, and Vetrov 2022; Kanakis et al. 2020). Our method resembles some of these methods in spirit and separates the network into shared and domain-specific modules. By limiting the number of parameters for the domain-specific modules, our method provides a parameter-efficient method to learn multiple tasks and domains. Furthermore, conditional computation is efficient during training and inference compared to independent networks (Riquelme et al. 2021).

Contributions. We summarize the contributions of this paper as follows.

- We proposed a simple parameter-efficient domain expansion technique to modulate weights of a pretrained network with rank-one factors. Our method expands the domain of the networks and adapts to a variety of data/-model shifts that arise in inverse problems.
- Our method requires a small number of domain-specific parameters (less than 0.5% of a single network) that can be stored separately from the shared network weights. This enables the network to continuously adapt to new domains without forgetting previous knowledge; therefore, we call it domain expansion.
- We present a detailed set of experiments that analyze the effects of domain, forward model, and noise-level shifts in natural and medical image reconstruction problems using deep unrolled methods.

2 Related Work

2.1 Inverse problems and structured priors.

Hand-crafted priors for inverse problems usually assume the signal is sparse in some transform domain. ℓ_1 norm has been widely used as a sparsity-promoting regularizer (Candès, Romberg, and Tao 2006; Donoho 2006). Total variation (TV) minimization is used as a regularization approach in (Beck and Teboulle 2009; Rudin, Osher, and Fatemi 1992) to solve denoising and deblurring problems. An iterative algorithm that minimizes the image total variation (TV) for CT reconstruction was proposed in (Sidky, Kao, and Pan 2006). These hand-crafted priors, however, have limited ability to represent the true underlying image and may lead to sub-optimal solutions (Qayyum et al. 2022).

2.2 Deep networks for inverse problems

Generative models learn to map a low-dimensional code into an image. Following (Bora et al. 2017), several methods have successfully applied generative networks as priors when solving inverse problems including MRI compressed sensing (Jalal et al. 2021), super-resolution (Menon et al. 2020), blind image deconvolution (Asim, Shamshad, and Ahmed 2020), and phase retrieval (Hand, Leong, and Voroninski 2018; Hyder et al. 2019).

End-to-end trained networks are purely data driven methods that learn to directly map measurements to signals. A denoising network that directly maps corrupted images to clean images was proposed in (Zhu et al. 2018). The method was applied to MRI measurements captured under different acquisition setups (Lehtinen et al. 2018). Other approaches such as (Sriram et al. 2020; Jin et al. 2017) use end-to-end networks to estimate artifact free signals from initial states.

Plug-and-play (PnP) methods are at the intersection of data driven and model based methods that alternatively

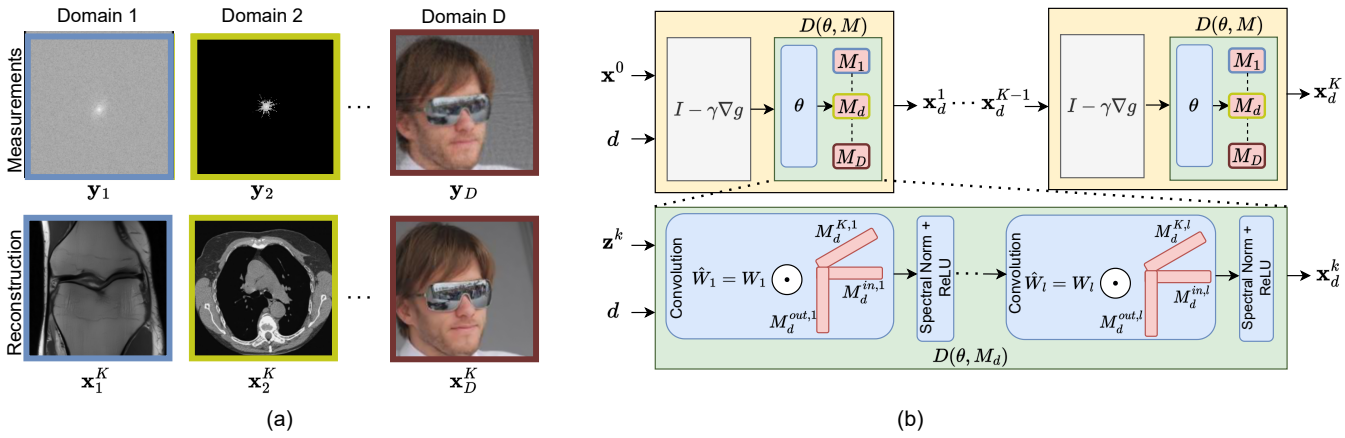

Figure 2: *Overview of our proposed factorized network adaptation. (a) Our network can receive measurements from different domains and our goal is to recover the underlying images (highlighted with measurements and corresponding reconstructions for three different domains). (b) Illustration of an unrolled network that has a pretrained denoiser network. Our main goal is to learn domain-specific rank-one factors that adapt the weights of the pretrained network at every layer. To reconstruct an image from given measurements, we select the domain-specific modulation weight and adapt the network.*

minimizes data consistency and regularization terms. PnP-ADMM (Venkatakrishnan, Bouman, and Wohlberg 2013) was the first plug-and-play iterative algorithm that used pretrained denoisers as priors. This method is based on the ADMM algorithm (Boyd et al. 2011). PnP-FISTA (Kamilov, Mansour, and Wohlberg 2017) is a PnP variant that replaces the proximal operator (Boyd et al. 2011) of the data fidelity with the gradient. These methods have been successfully applied to several inverse problems (Chan, Wang, and Elgendy 2016; Ahmad et al. 2020)

Deep unrolled networks learn the denoiser network in PnP algorithms in a supervised manner (Zhang and Ghanem 2018; Monga, Li, and Eldar 2021; Liu et al. 2021). These methods truncate the PnP algorithm for a fixed number of iterations and share the same network through the iterations. They perform updates using the reconstruction output of the final iteration. Deep unrolled methods show remarkable results in several inverse problems such as super-resolution (Zhang, Gool, and Timofte 2020), image restoration (Mou, Wang, and Zhang 2022), MRI (Jun et al. 2021) and CT (Wu, Kim, and Li 2019) reconstruction.

2.3 Domain expansion and adaptation

Developing a single network that can handle multiple domains as well as adapt to new target domains has been an active area of research. Deep neural networks can learn transferable features and fine-tuning to a new dataset improves generalization performance (Yosinski et al. 2014; Long et al. 2018). Despite its success, fine-tuning a network or parts of it force the network to lose previously learned domain or task, which requires storing multiple networks per domain and task. Parameter-efficient fine-tuning methods (Hu et al. 2021; He et al. 2023; Liu et al. 2022) propose networks that can achieve competitive performance to fully-tuned networks while requiring few number of additional parameters. Adapter-based techniques that learn efficient modules have been proposed in (Rebuffi, Bilen, and Vedaldi 2018; Li, Liu, and Bilen 2022; Chen et al. 2022). These modules are added to a pretrained network and enable it to adapt to new tasks. Other works employ sub-networks using binary masking for specific domains (Wortsman et al. 2020; Mallya, Davis, and Lazebnik 2018). Tuning specific layers such as the BatchNorm (Frankle, Schwab, and Morcos 2020), the final classification head (He et al. 2020), and LayerNorm (Basu et al. 2023) also provides effective adaptation. A related approach that scales and shifts features to achieve the performance of full-tuning was proposed in (Lian et al. 2022). In (Kanakis et al. 2020; Rosenfeld and Tsotsos 2018), a network reparametrization technique was proposed to learn shared and task-specific modules, enabling a single network to adapt to various settings. Hyperdomain Networks (Alanov, Titov, and Vetrov 2022) use modulated convolution to adapt generator networks to new domains. An adaptation method for shifts in domain and forward-models when solving inverse problems was proposed in (Gilton, Ongie, and Willett 2021). The method proposes a fine-tuning and regularization technique adopted from RED (Romano, Elad, and Milanfar 2017). Domain-specific batch normalization layers were proposed in (Karani et al. 2018) for a segmentation network that can handle brain MR scans across scanners and protocols. Unlike R&R (Romano, Elad, and Milanfar 2017), the method proposed in (Karani et al. 2018) can adapt to new domains without forgetting previous domains. Several test-time adaptation techniques have been proposed to close performance gaps resulting from domain shifts (Darestani, Liu, and Heckel 2022; Song, Shen, and Xing 2023; Goyal et al. 2022).

While many of these methods are proposed for purely data drive approaches, we focus on methods that fuse data-driven and model based techniques. In addition, our aim is to find parameter efficient domain adaptation techniques without introducing catastrophic forgetting.

3 Methods

In this section, we present details of our proposed domain expansion method for deep unrolling-based reconstruction.

We first briefly discuss deep unrolled networks (readers may refer to (Kamilov et al. 2023) for further details). Then we discuss how we adapt the network weights using rank-one factors to perform domain expansion/adaptation.

3.1 Deep unrolled network

A deep unrolled network in its simplest form represents a fixed number of iterations for solving the optimization problem in (2). Plug and play (PnP) methods based on accelerated proximal gradients (Venkatakrishnan, Bouman, and Wohlberg 2013; Parikh, Boyd et al. 2014; Kamilov et al. 2023) offer a flexible and efficient framework for solving such problems. Key steps of PnP with a deep denoiser at iteration k can be described as follows.

$$\mathbf{z}^k = \mathbf{x}^{k-1} - \gamma \nabla g(\mathbf{x}^{k-1}) \tag{3}$$

$$\mathbf{s}^k = \mathcal{D}(\mathbf{z}^k; \theta) \tag{4}$$

$$\mathbf{x}^k = \mathbf{s}^k + \beta_k(\mathbf{s}^k - \mathbf{s}^{k-1}), \tag{5}$$

where γ is the step size, superscript $k = 1, \ldots, K$ denotes iteration number, $\nabla g(\cdot)$ denotes gradient of data fidelity with respect to $\mathbf{x}$, $\mathcal{D}(\cdot; \theta)$ denotes a denoiser or artifact removal network with weights θ, $\beta_k = (q_{k-1} - 1)/q_k$, and $q_k = (1/2)(1 + \sqrt{1 + 4q_{k-1}^2})$. We can initialize the estimate as $\mathbf{x}^0 = \mathbf{A}^H\mathbf{y}$, where $\mathbf{A}^H$ denotes Hermitian transpose of the measurement operator. Similar to (Liu et al. 2021), we implement $\mathcal{D}$ as an artifact removal network: $\mathcal{D}(\mathbf{x}; \theta) = \mathbf{x} - \mathbf{f}(\mathbf{x}; \theta)$, where $\mathbf{f}$ is a DnCNN-based residual network (Zhang et al. 2017).

We can view each iteration of PnP as one layer of the unrolled network with predefined parameters. The output of an unrolled network with denoiser $\mathcal{D}(\cdot, \theta)$ and K iterations can be denoted as $\mathbf{x}^K(\theta)$. Since all operations are differentiable, we can further improve the performance by minimizing the reconstruction error on some training images with respect to θ. We can define such an optimization problem as

$$\min_{\theta} \sum_{\mathbf{x} \in \mathcal{X}} \mathcal{L}(\mathbf{x}, \mathbf{x}^K(\theta)), \tag{6}$$

where $\mathcal{X}$ denotes the set of training images.

3.2 Factorized network adaptation

Our method primarily adapts the denoiser in the unrolled network using domain/task-specific rank-one factors as the data, measurement, or noise distribution changes. We start with a pretrained network $\mathcal{D}(\cdot; \theta)$ with parameters θ. Then we learn domain-specific modulations, say $\{M_d\}_{d=1}^D$ for D domains. Each M_d represents a set of domain-specific modulation parameters that we use to adapt base network parameters to $\theta \odot M_d$, where $\odot$ represents element-wise multiplication. In practice, we do not create a new set of modulated weights; instead we keep the M_d and θ separate. We represent the domain-specific network for dth domain as $\mathcal{D}(\cdot, \theta, M_d)$ and the output of the unrolled network as $\mathbf{x}^K(\theta, M_d)$. To learn the modulation parameters for dth domain, we keep θ unchanged and solve the following optimization problem for M_i:

$$\min_{M_d} \sum_{\mathbf{x} \in \mathcal{X}_d} \mathcal{L}(\mathbf{x}, \mathbf{x}^K(\theta, M_d)). \tag{7}$$

Algorithm 1: Factorized network adaptation

Input: Training images $\mathbf{x} \in \mathcal{X}_d$ with measurements $\mathbf{y}$, and operator $\mathbf{A}$ for domain indicator d
Base network parameters θ, $\{\beta_k\}_{k \geq 0}$, γ, α
Output: Recovered image $\mathbf{x}^K$ and domain-specific M_d

$M_d \leftarrow$ `initialModulation`(d)
repeat
  for every $\mathbf{x} \in \mathcal{X}_d$ and $\mathbf{y}$
  initialize $\mathbf{x}^0 \leftarrow \mathbf{A}^H\mathbf{y}$
  for $k \in \{1, \ldots, K\}$ **do**
    $\mathbf{z}^k \leftarrow \mathbf{x}^{k-1} - \gamma \nabla g(\mathbf{x}^{k-1})$
    $\mathbf{s}^k \leftarrow \mathcal{D}(\mathbf{z}^k; \theta, M_d)$
    $\mathbf{x}^k \leftarrow \mathbf{s}^k + \beta_k(\mathbf{s}^k - \mathbf{s}^{k-1})$
  end for
  Calculate loss for all training samples in a minibatch and compute gradient w.r.t. M_d
  $M_d \leftarrow M_d - \alpha \nabla_{M_d} \sum_{\mathbf{x} \in \mathcal{X}_d} \mathcal{L}(\mathbf{x}^K, \mathbf{x})$
until Convergence of M_d
return $\mathbf{x}^K, M_d$

where $\mathcal{X}_d$ denotes the set of training images for the dth domain. A pseudocode for the factorized adaptation with unrolled network is provided in Algorithm 1.

Even though we do not explicitly discuss measurement operator $\mathbf{A}$ and noise η in the unrolled network, any mismatch between training and test time settings of domain, measurements, and noise can cause performance degradation. We can consider any variation in data, measurements, or noise as a new domain and use the same procedure described above to learn the domain-specific modulation parameters.

Rank-one factorization. Inspired by (Li et al. 2018; Hu et al. 2021), we assume the intrinsic dimension of the objective formulated in (7) is small. We parameterize M_d such that the number of parameters remains significantly smaller than the number of parameters in the original network θ.

To achieve the goal of parameter efficiency, we represent modulation weights for each layer as a rank-one tensor. Let us assume lth convolution layer has weights W^l with kernels of size $k \times k$ with C_{in} input and C_{out} output channels. We represent the modulation weights for dth domain and lth layer as an outer product of four vectors as $M_d^l = M_d^{1,l} \otimes M_d^{2,l} \otimes M_d^{3,l} \otimes M_d^{4,l}$, where $M_d^{1,l} \in \mathbb{R}^k, M_d^{2,l} \in \mathbb{R}^k, M_d^{3,l} \in \mathbb{R}^{C_{in}}, M_d^{4,l} \in \mathbb{R}^{C_{out}}$. Thus, we need $k + k + C_{in} + C_{out}$ parameters to adapt a layer with $k^2 C_{in} C_{out}$ parameters. We apply the rank-one factorization and modulation on the convolution layers as follows. For an input U with C_{in} channels, we can represent ith output channel of the convolution layer as

$$V(:, :, i) = \sum_{j=1}^{C_{in}} W^l(:, :, j, i) * U(:, :, j), \tag{8}$$

where $*$ represents 2D convolution. Modulated weights for domain d and layer l can be represented as $W_d^l = W^l \odot M_d^l$.

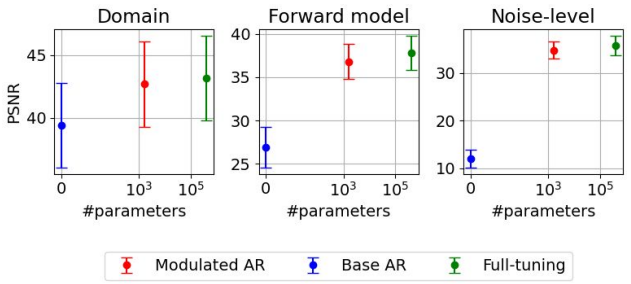

Figure 3: *Comparison of our modulated AR, fully-tuned AR, and the Base AR networks in terms of accuracy and number of additional parameters they require. Base AR requires no additional parameter and provides worst performance. Fully-tune AR provides best performance using a large number of parameters. Our proposed method, Modulated AR, shows performance comparable to Fully-tuned AR with a fraction of additional parameters.*

We can represent the convolution operation as

$$V(:,:,i) = M_d^{4,l}(i) \left[\sum_{j=1}^{C_{in}} \widetilde{W}^l(:,:,j,i) * \widetilde{U}(:,:,j) \right], \qquad (9)$$

where $\widetilde{W}^l(:,:,j,i) = W^l(:,:,j,i) \odot (M_d^{1,l} \otimes M_d^{2,l})$ represents a modulated version of (j,i) slice of weight tensor and $\widetilde{U}(:,:,j) = U(:,:,j) \odot M_d^{3,l}$ represents a modulated version of the jth input channel. In summary, even though we represent modulation weights are rank-one tensor, we do not need to modulate the weights of the base network. We can implement the same procedure by modulating input channels, 2D filters, and output channels.

4 Experiments and Results

We performed a number of experiments to analyse the effects of shifts in different parts of the inverse problem in (1). The shifts can occur in the data distribution $\mathbf{x}$, the forward model $\mathbf{A}$, and the measurement noise η. We test our proposed adaptation technique for all these shifts. In all our experiments, we start with a fixed base network, which we refer to as Base AR, and learn domain-specific rank-one modulations. Base AR is trained to reconstruct MR images from $4\times$ radially sub-sampled Fourier measurements without any measurement noise. Base AR uses spectral normalization proposed in (Miyato et al. 2018) along with the ReLU activation functions. We implement our AR network using a 12-layer DnCNN (Zhang et al. 2017) network. We will provide training details as well as hyper-parameters used in our experiments in the supplementary material.

4.1 Parameter efficiency for adaptation

Figure 3 compares the performance of a base network, full training, and our proposed modulation-based adaptation for shifts in data distribution/domain, forward model, and noise level. Base network does not require any additional parameter for different domains, but it provides worst performance. Full training learns a new network for every domain/distribution shifts and provides best performance, but at the expense of a large number of parameters per domain. Our proposed network adaptation approach requires a small number of parameters (nearly 1.6K additional parameters) and achieves performance close to Full training method. The additional parameters are unique for each domain and are stored separately from the base network. In this manner, the pre-trained model can be adapted to learn new domains while retaining previously learned knowledge.

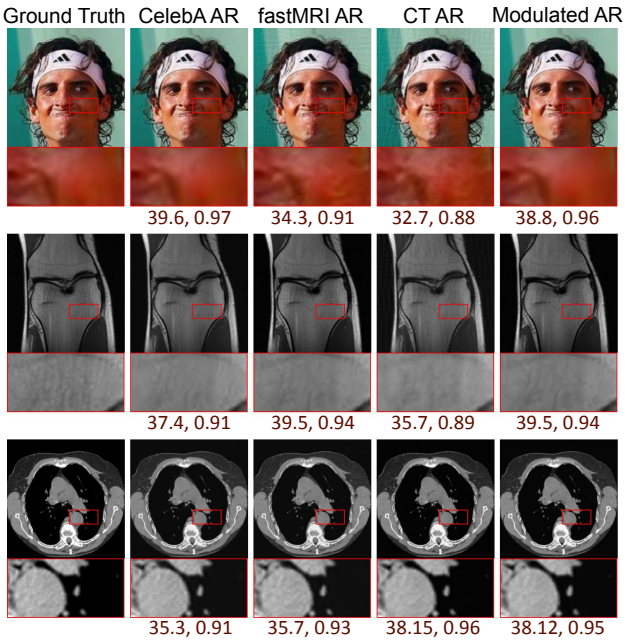

Figure 4: *We present sample ground truth images in the first column and reconstruction of these images using three AR networks trained on Face, MR, and CT images in the subsequent three columns. Our modulated AR, shown in the last column effectively removes this artifacts and closes the performance gap.*

Table 1: *Average PSNR of AR networks under domain shift. ARs trained for specific domain (MRI, CelebA, and CT ARs) do not perform well on out-of-domain samples. In contrast, our Modulated AR network, that applies learned modulations for each target domain, has the best average performance across all domains.*

Test domain	AR Trained on MRI	AR Trained on CelebA	AR Trained on CT	Modualted AR (Ours)
MRI	**40.93**	39.22	37.14	**40.93**
CelebA	40.34	**44.29**	35.44	42.97
CT	37.68	38.56	41.97	**42.25**
Avg	39.65	40.69	38.18	**42.05**

4.2 Domain shift

For experiments with domain/data distribution shifts in $\mathbf{x}$, we consider natural image, MRI, and CT scans. We use CelebA dataset (Liu et al. 2015) for natural images, NYU fastMRI dataset for (Zbontar et al. 2018) knee MRI scans, and a subset of TCGA-LUAD dataset (Clark et al. 2013) for CT scans. The first three columns of Table 1 show the performance single domain AR networks. We present the reconstruction PSNR of these AR networks evaluated under

Table 2: *Comparison of our method with existing domain adaptation techniques. We have highlighted the best-performing method in boldface and second best with an underscore. Additionally, we provide the count of additional parameters required by each method. Our modulated AR outperforms other methods and is comparable to full tuning.*

Target domain	Full-tuning	Supsup	RCM	Hyperdomain	Modulated AR (Ours)
	407k	407k	50.6k	0.7k	1.6k
CelebA	**44.29**	42.49	<u>43.88</u>	42.73	42.97
CT	<u>41.97</u>	40.57	40.99	41.57	**42.25**
Avg	**42.40**	41.33	41.93	41.74	<u>42.05</u>

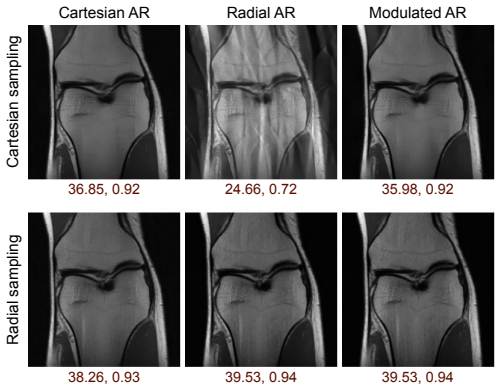

Figure 5: *Reconstruction results under sampling pattern shifts.*

the domain shifts. The last column shows the performance of our modulated network that uses weights of the Base AR trained on MRI and learned modulations for each target domain. Quantitatively we observe that performance drops as domains change (off-diagonal entries in columns 2,3,4). Our proposed method for modulated AR offers best overall performance. Figure 4 shows example reconstructed images for our domain shift experiments. Our modulated network effectively removes artifacts introduced by fastMRI AR and CT AR on CelebA images.

We compare our proposed approach with related domain adaptation techniques. Supsup (Wortsman et al. 2020) learns binary masks to find domain specific sub-networks. RCM (Kanakis et al. 2020) reparameterizes convolutions using domain-specific feature transformations. Hyperdomain (Alanov, Titov, and Vetrov 2022) learns domain-specific modulation for input channel of every convolution operation. Full-tuning retrains the entire network for each target domain and is typically considered to be an upper-bound. Table 2 shows comparison of these methods and our proposed method outperforms other adaptation techniques while requiring fewer additional trainable parameters.

4.3 Forward model shifts

To evaluate the performance with shifts in the forward model, $\mathbf{A}$, we consider sampling types, ratio, and patterns as domains that can induce shifts. The sampling type can be either Fourier or Gaussian sampling. In the case of Fourier sampling, we can have Cartesian, Radial, Gaussian, or Spiral patterns. The sampling ratio determines the rate at

Table 3: *Sampling pattern shift adaptation results. Modulated AR achieves competitive in-domain performance to ARs trained on specific pattern. Moreover, it shows an overall superior performance across all patterns.*

Test pattern	Radial AR	Cartesian AR	Gaussian AR	Spiral AR	Modulated AR (Ours)
Radial	**40.93**	37.75	40.55	40.83	**40.93**
Cartesian	29.74	**39.21**	28.39	29.12	37.10
Gaussian	41.91	40.19	42.05	42.04	**42.10**
Spiral	41.24	39.57	41.26	41.36	**41.38**
Avg	38.46	39.18	38.06	38.34	**40.38**

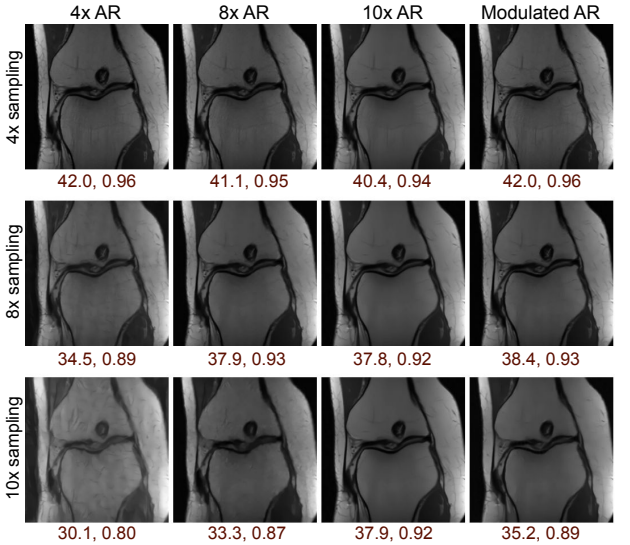

Figure 6: *Examples of image reconstruction images under sampling ratio shifts.*

which measurements are captured. We consider reconstruction from $4\times$, $8\times$ and $10\times$ under-sampled measurements. We will now examine the effects of each of these shifts and utilize our proposed method to adapt our Base AR.

Sampling pattern shifts. Table 3 shows the performance of AR networks trained on single sampling patterns when tested on all available patterns in the first four columns. The last column shows the performance of our modulated AR. We observed a significant performance drop when our Base AR was tested on samples from Cartesian samples. This drop is also evident qualitatively in Figure 5, where visible artifacts appear in the output. Our modulated AR successfully eliminates these artifacts and bridges the performance gap. Moreover, our method provides overall superior performance compared to networks trained for individual patterns.

Sampling ratio shifts. We compared the performance of different AR networks trained on three sampling ratios and presented the results in Table 4. The $4\times$ AR network exhibits poor performance when tested with $8\times$ and $10\times$ radially subsampled measurements. Additionally, the AR network trained on the $8\times$ ratio did not perform well with $10\times$ ratio. To address this, we applied our modulation technique to adapt the Base AR model to $8\times$ and $10\times$ sampling ratios. Figure 6 illustrates the reconstruction results of the networks

Table 4: *Sampling ratio shift adaptation results.*

Test ratio	4x AR	8x AR	10x AR	Modualted AR (Ours)
4x	**40.93**	40.23	39.61	**40.93**
8x	34.98	37.13	37.05	**37.32**
10x	31.00	33.63	**35.34**	34.73
Avg	35.64	37.00	37.33	**37.66**

Table 5: *Comparison of domain adaptation method under forward model shifts.*

Sampling shifts	Full-tuning	Supsup	RCM	Hyperdomain	Modulated AR (Ours)
	407k	407k	50.6k	0.7k	1.6k
Radial to Cartesian	**39.21**	36.37	36.52	36.57	37.10
4x to 10x	**35.34**	33.62	34.49	34.56	34.73
Fourier to Gaussian	**38.59**	36.31	38.49	38.45	38.55
Avg	**37.71**	35.43	36.50	36.53	36.79

trained at various sampling ratios, including our modulated network. On average, the modulated network outperforms AR networks trained on specific sampling ratios.

Comparison with existing domain adaptation methods. We now compare our method with some of the existing domain adaptation techniques under the forward model shifts discussed above. We report the average PSNR along with the number of trainable parameters with in each method in Table 5. Our proposed method outperforms all domain techniques and is only one dB less than full-tuning, which requires significantly larger number of parameters.

4.4 Noise-level shifts

Noise-level shifts can also cause significant performance degradation in AR networks. We model the noise as an additive Gaussian noise $\eta \sim \mathcal{N}(0, \sigma^2)$ and analyze the effects of different noise levels on the performance. Figure 7 shows sample reconstructed images with the Base AR, our Modulated AR, and AR networks trained for 10, 20, and 30 dB SNR. We observed that the Base AR is unable to reconstruct the MRI scans from the noisy measurements. This is also shown quantitatively in Table 4, where the performance of the Base AR is severely degraded in the presence of noise. The AR network trained on 10dB SNR performs well on higher noise settings but fails to recover fine details when tested with noise-free or low noise measurements. The last row of Figure 7 shows the $20\times$ amplified reconstruction residual of each model when reconstructing noise-free measurements. From this row, we can infer that AR networks trained on higher noise-levels fail to recover fine details when tested with lower noise-levels. To the contrary, our Modulated AR has the ability to reconstruct fine details when the measurement noise is low and maintains comparable performance to noise-specific AR as the noise level increases.

Table 7 reports comparison of our proposed method with related domain adaptation techniques. Although Full-tuning and Supsup (Wortsman et al. 2020) show slight performance improvement (less than 1 dB), they require a significant number of trainable parameters. Furthermore, Full-tuning does not have the ability to retrain previously learned knowledge. Our method achieves competitive performance

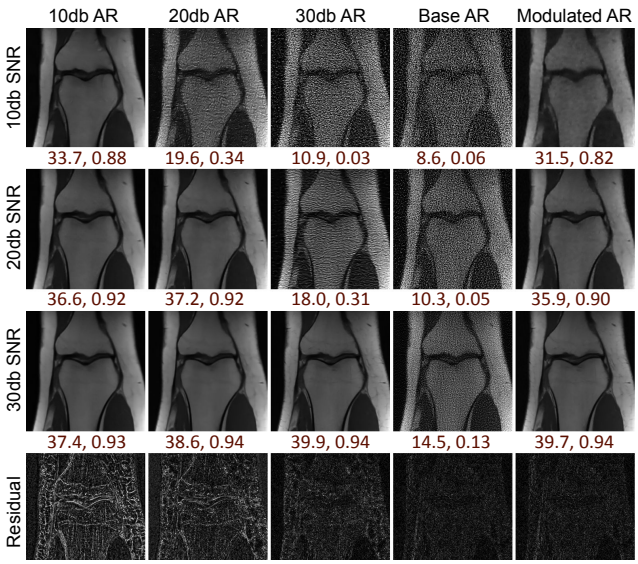

Figure 7: *Visual results of models trained at specific noise levels and our modulated network under measurement level. The last row shows the amplified residual of the reconstructed image under no measurement noise.*

Table 6: *Noise level shift adaptation results.*

Test SNR	10db AR	20db AR	30db AR	Base AR	Modualted AR (Ours)
10db	**33.37**	23.86	12.56	9.18	31.40
20db	35.39	**35.82**	22.86	11.20	35.09
30db	35.97	36.84	**38.04**	15.90	37.91
No noise	36.10	36.97	38.97	**40.93**	**40.93**
Avg	35.21	33.30	28.11	19.30	**36.33**

Table 7: *Comparison of various domain adaptation methods under noise level shifts*

Test SNR	Full-tuning	Supsup	RCM	Hyperdomain modulation	Modulated AR (Ours)
	407k	407k	50.6k	0.7k	1.6k
10db	**33.37**	31.94	32.80	29.11	31.40
20db	**35.82**	35.53	35.70	34.32	35.09
30db	38.04	37.75	**38.08**	37.58	37.91
Avg	**35.74**	35.07	35.53	33.67	34.80

to RCM (Kanakis et al. 2020) while requiring a fraction of the additional trainable parameters.

5 Conclusion

We proposed a simple and parameter-efficient method to adapt networks for domain adaptation and expansion. Our method uses a fixed base network and learns separate (domain-specific) rank-one modulation parameters. This capability allows our method to continually learn new domains while retaining previously acquired knowledge. We focused on shifts that arise in solving inverse problems for imaging, including shifts in data distribution, forward model, and noise level. We demonstrated the effectiveness of our approach in adapting to all these shifts.

References

Ahmad, R.; Bouman, C. A.; Buzzard, G. T.; Chan, S.; Liu, S.; Reehorst, E. T.; and Schniter, P. 2020. Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery. *IEEE signal processing magazine*, 37(1): 105–116.

Alanov, A.; Titov, V.; and Vetrov, D. P. 2022. Hyperdomainnet: Universal domain adaptation for generative adversarial networks. *Advances in Neural Information Processing Systems*, 35: 29414–29426.

Antun, V.; Renna, F.; Poon, C.; Adcock, B.; and Hansen, A. C. 2020. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proceedings of the National Academy of Sciences*, 117(48): 30088–30095.

Asim, M.; Shamshad, F.; and Ahmed, A. 2020. Blind image deconvolution using deep generative priors. *IEEE Transactions on Computational Imaging*, 6: 1493–1506.

Basu, S.; Massiceti, D.; Hu, S. X.; and Feizi, S. 2023. Strong Baselines for Parameter Efficient Few-Shot Finetuning. *arXiv e-prints*, arXiv–2304.

Beck, A.; and Teboulle, M. 2009. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE transactions on image processing*, 18(11): 2419–2434.

Bora, A.; Jalal, A.; Price, E.; and Dimakis, A. G. 2017. Compressed sensing using generative models. In *International conference on machine learning*, 537–546. PMLR.

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

Candès, E. J.; Romberg, J.; and Tao, T. 2006. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2): 489–509.

Candes, E. J.; Romberg, J. K.; and Tao, T. 2006. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8): 1207–1223.

Chan, S. H.; Wang, X.; and Elgendy, O. A. 2016. Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1): 84–98.

Chen, S.; Ge, C.; Tong, Z.; Wang, J.; Song, Y.; Wang, J.; and Luo, P. 2022. Adaptformer: Adapting vision transformers for scalable visual recognition. *Advances in Neural Information Processing Systems*, 35: 16664–16678.

Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; Tarbox, L.; and Prior, F. 2013. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *Journal of Digital Imaging*, 26(6): 1045–1057.

Darestani, M. Z.; Chaudhari, A. S.; and Heckel, R. 2021. Measuring robustness in deep learning based compressive sensing. In *International Conference on Machine Learning*, 2433–2444. PMLR.

Darestani, M. Z.; Liu, J.; and Heckel, R. 2022. Test-time training can close the natural distribution shift performance gap in deep learning based compressed sensing. In *International Conference on Machine Learning*, 4754–4776. PMLR.

Donoho, D. L. 2006. Compressed sensing. *IEEE Transactions on information theory*, 52(4): 1289–1306.

Elad, M.; and Aharon, M. 2006. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12): 3736–3745.

Frankle, J.; Schwab, D. J.; and Morcos, A. S. 2020. Training BatchNorm and Only BatchNorm: On the Expressive Power of Random Features in CNNs. In *International Conference on Learning Representations*.

Gilton, D.; Ongie, G.; and Willett, R. 2021. Model adaptation for inverse problems in imaging. *IEEE Transactions on Computational Imaging*, 7: 661–674.

Gottschling, N. M.; Antun, V.; Adcock, B.; and Hansen, A. C. 2020. The troublesome kernel: why deep learning for inverse problems is typically unstable. *arXiv preprint arXiv:2001.01258*.

Goyal, S.; Sun, M.; Raghunathan, A.; and Kolter, J. Z. 2022. Test Time Adaptation via Conjugate Pseudo-labels. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 6204–6218. Curran Associates, Inc.

Gregor, K.; and LeCun, Y. 2010. Learning fast approximations of sparse coding. In *Proceedings of the 27th international conference on international conference on machine learning*, 399–406.

Hand, P.; Leong, O.; and Voroninski, V. 2018. Phase retrieval under a generative prior. *Advances in Neural Information Processing Systems*, 31.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9729–9738.

He, X.; Li, C.; Zhang, P.; Yang, J.; and Wang, X. E. 2023. Parameter-Efficient Model Adaptation for Vision Transformers. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(1): 817–825.

Hu, E. J.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W.; et al. 2021. LoRA: Low-Rank Adaptation of Large Language Models. In *International Conference on Learning Representations*.

Hyder, R.; Shah, V.; Hegde, C.; and Asif, M. S. 2019. Alternating phase projected gradient descent with generative priors for solving compressive phase retrieval. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7705–7709. IEEE.

Jalal, A.; Arvinte, M.; Daras, G.; Price, E.; Dimakis, A. G.; and Tamir, J. 2021. Robust Compressed Sensing MRI with Deep Generative Priors. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in*

Neural Information Processing Systems, volume 34, 14938–14954. Curran Associates, Inc.

Jin, K. H.; McCann, M. T.; Froustey, E.; and Unser, M. 2017. Deep convolutional neural network for inverse problems in imaging. *IEEE transactions on image processing*, 26(9): 4509–4522.

Jun, Y.; Shin, H.; Eo, T.; and Hwang, D. 2021. Joint deep model-based MR image and coil sensitivity reconstruction network (joint-ICNet) for fast MRI. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5270–5279.

Kamilov, U. S.; Bouman, C. A.; Buzzard, G. T.; and Wohlberg, B. 2023. Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 40(1): 85–97.

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

Kanakis, M.; Bruggemann, D.; Saha, S.; Georgoulis, S.; Obukhov, A.; and Gool, L. V. 2020. Reparameterizing convolutions for incremental multi-task learning without task interference. In *European Conference on Computer Vision*, 689–707. Springer.

Karani, N.; Chaitanya, K.; Baumgartner, C.; and Konukoglu, E. 2018. A lifelong learning approach to brain MR segmentation across scanners and protocols. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 476–484. Springer.

Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; and Aila, T. 2018. Noise2Noise: Learning Image Restoration without Clean Data. In *International Conference on Machine Learning*, 2965–2974. PMLR.

Li, C.; Farkhoor, H.; Liu, R.; and Yosinski, J. 2018. Measuring the Intrinsic Dimension of Objective Landscapes. In *International Conference on Learning Representations*.

Li, W.-H.; Liu, X.; and Bilen, H. 2022. Cross-domain few-shot learning with task-specific adapters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7161–7170.

Lian, D.; Zhou, D.; Feng, J.; and Wang, X. 2022. Scaling & shifting your features: A new baseline for efficient model tuning. *Advances in Neural Information Processing Systems*, 35: 109–123.

Liu, J.; Asif, M. S.; Wohlberg, B.; and Kamilov, U. S. 2021. Recovery Analysis for Plug-and-Play Priors using the Restricted Eigenvalue Condition. In *Advances in Neural Information Processing Systems (NeurIPS)*. In press.

Liu, Y.-C.; MA, C.-Y.; Tian, J.; He, Z.; and Kira, Z. 2022. Polyhistor: Parameter-Efficient Multi-Task Adaptation for Dense Vision Tasks. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 36889–36901. Curran Associates, Inc.

Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep Learning Face Attributes in the Wild. In *Proceedings of International Conference on Computer Vision (ICCV)*.

Long, M.; Cao, Y.; Cao, Z.; Wang, J.; and Jordan, M. I. 2018. Transferable representation learning with deep adaptation networks. *IEEE transactions on pattern analysis and machine intelligence*, 41(12): 3071–3085.

Lustig, M.; Donoho, D.; and Pauly, J. M. 2007. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6): 1182–1195.

Mairal, J.; Bach, F.; Ponce, J.; and Sapiro, G. 2009. Online dictionary learning for sparse coding. In *Proceedings of the 26th annual international conference on machine learning*, 689–696.

Mallya, A.; Davis, D.; and Lazebnik, S. 2018. Piggyback: Adapting a single network to multiple tasks by learning to mask weights. In *Proceedings of the European conference on computer vision (ECCV)*, 67–82.

McCloskey, M.; and Cohen, N. J. 1989. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24, 109–165. Elsevier.

Menon, S.; Damian, A.; Hu, S.; Ravi, N.; and Rudin, C. 2020. Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, 2437–2445.

Miyato, T.; Kataoka, T.; Koyama, M.; and Yoshida, Y. 2018. Spectral Normalization for Generative Adversarial Networks. In *International Conference on Learning Representations*.

Monga, V.; Li, Y.; and Eldar, Y. C. 2021. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2): 18–44.

Mou, C.; Wang, Q.; and Zhang, J. 2022. Deep Generalized Unfolding Networks for Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17399–17410.

Ongie, G.; Jalal, A.; Metzler, C. A.; Baraniuk, R. G.; Dimakis, A. G.; and Willett, R. 2020. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1): 39–56.

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

Qayyum, A.; Ilahi, I.; Shamshad, F.; Boussaid, F.; Bennamoun, M.; and Qadir, J. 2022. Untrained neural network priors for inverse imaging problems: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Rebuffi, S.-A.; Bilen, H.; and Vedaldi, A. 2018. Efficient Parametrization of Multi-Domain Deep Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Riquelme, C.; Puigcerver, J.; Mustafa, B.; Neumann, M.; Jenatton, R.; Susano Pinto, A.; Keysers, D.; and Houlsby, N. 2021. Scaling vision with sparse mixture of experts. *Advances in Neural Information Processing Systems*, 34: 8583–8595.

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

Rosenfeld, A.; and Tsotsos, J. K. 2018. Incremental learning through deep adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 42(3): 651–663.

Rudin, L. I.; Osher, S.; and Fatemi, E. 1992. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1): 259–268.

Sidky, E. Y.; Kao, C.-M.; and Pan, X. 2006. 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.

Song, B.; Shen, L.; and Xing, L. 2023. PINER: Prior-Informed Implicit Neural Representation Learning for Test-Time Adaptation in Sparse-View CT Reconstruction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 1928–1938.

Sriram, A.; Zbontar, J.; Murrell, T.; Defazio, A.; Zitnick, C. L.; Yakubova, N.; Knoll, F.; and Johnson, P. 2020. End-to-end variational networks for accelerated MRI reconstruction. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23*, 64–73. Springer.

Sun, Y.; Wohlberg, B.; and Kamilov, U. S. 2019. An online plug-and-play algorithm for regularized image reconstruction. *IEEE Transactions on Computational Imaging*, 5(3): 395–408.

Venkatakrishnan, S. V.; Bouman, C. A.; and Wohlberg, B. 2013. Plug-and-play priors for model based reconstruction. In *2013 IEEE global conference on signal and information processing*, 945–948. IEEE.

Wortsman, M.; Ramanujan, V.; Liu, R.; Kembhavi, A.; Rastegari, M.; Yosinski, J.; and Farhadi, A. 2020. Supermasks in Superposition. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 15173–15184. Curran Associates, Inc.

Wu, D.; Kim, K.; and Li, Q. 2019. Computationally efficient deep neural network for computed tomography image reconstruction. *Medical physics*, 46(11): 4763–4776.

Yosinski, J.; Clune, J.; Bengio, Y.; and Lipson, H. 2014. How transferable are features in deep neural networks? *Advances in neural information processing systems*, 27.

Zbontar, J.; Knoll, F.; Sriram, A.; Murrell, T.; Huang, Z.; Muckley, M. J.; Defazio, A.; Stern, R.; Johnson, P.; Bruno, M.; Parente, M.; Geras, K. J.; Katsnelson, J.; Chandarana, H.; Zhang, Z.; Drozdzal, M.; Romero, A.; Rabbat, M.; Vincent, P.; Yakubova, N.; Pinkerton, J.; Wang, D.; Owens, E.; Zitnick, C. L.; Recht, M. P.; Sodickson, D. K.; and Lui, Y. W. 2018. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. arXiv:1811.08839.

Zhang, J.; and Ghanem, B. 2018. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *CVPR*, 1828–1837.

Zhang, K.; Gool, L. V.; and Timofte, R. 2020. Deep Unfolding Network for Image Super-Resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

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

Zhu, B.; Liu, J. Z.; Cauley, S. F.; Rosen, B. R.; and Rosen, M. S. 2018. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697): 487–492.