STOCHASTIC DEEP RESTORATION PRIORS FOR IMAGING INVERSE PROBLEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Deep neural networks trained as image denoisers are widely used as priors for solving imaging inverse problems. While Gaussian denoising is thought sufficient for learning image priors, we show that priors from deep models pre-trained as more general restoration operators can perform better. We introduce *Stochastic deep Restoration Priors (ShaRP)*, a novel method that leverages an ensemble of such restoration models to regularize inverse problems. ShaRP improves upon methods using Gaussian denoiser priors by better handling structured artifacts and enabling self-supervised training even without fully sampled data. We prove ShaRP minimizes an objective function involving a regularizer derived from the score functions of minimum mean square error (MMSE) restoration operators, and theoretically analyze its convergence. Empirically, ShaRP achieves state-of-the-art performance on tasks such as magnetic resonance imaging reconstruction and single-image super-resolution, surpassing both denoiser- and diffusion-model-based methods without requiring retraining.

1 INTRODUCTION

Many problems in computational imaging, biomedical imaging, and computer vision can be viewed as *inverse problems*, where the goal is to recover an unknown image from its noisy and incomplete measurements. Inverse problems are typically ill-posed, thus requiring additional prior information for accurate image reconstruction. While many approaches have been proposed for implementing image priors, the current research focuses on methods based on deep learning (DL) (McCann et al., 2017; Ongie et al., 2020; Kamilov et al., 2023; Wen et al., 2023).

Deep neural networks trained as image denoisers are widely-used for specifying image priors for solving *general* inverse problems (Romano et al., 2017; Kadkhodaie & Simoncelli, 2021; Zhang et al., 2022). The combination of pre-trained Gaussian denoisers with measurement models has been shown to be effective in many inverse problems, including image super-resolution, deblurring, and medical imaging (Metzler et al., 2018; Zhang et al., 2017; Meinhardt et al., 2017; Dong et al., 2019; Zhang et al., 2019; Wei et al., 2020; Zhang et al., 2022) (see also the recent reviews (Ahmad et al., 2020; Kamilov et al., 2023)). This success has led to active research on novel methods based on denoiser priors, their theoretical analyses, statistical interpretations, as well as connections to related approaches such as score matching and diffusion models (Venkatakrishnan et al., 2013; Chan et al., 2017; Romano et al., 2017; Buzzard et al., 2018; Reehorst & Schniter, 2019; Sun et al., 2019; Sun et al., 2019; Ryu et al., 2019; Xu et al., 2020; Liu et al., 2021; Cohen et al., 2021a; Hurault et al., 2022a;b; Laumont et al., 2022; Gan et al., 2023a; Renaud et al., 2024b).

The mathematical relationship between denoising and the score function (the gradient of the log of the image distribution), known as the Tweedie's formula (Robbins, 1956; Efron, 2011) seemingly implies that Gaussian denoising alone might be sufficient for learning priors, independent of the specific characteristics of an inverse problem. This view overlooks an important point that Gaussian-denoiser networks are suboptimal when when used as priors to restore images degraded by other, non-Gaussian artifacts. One potential approach to mitigate this issue is to explore a broader class of restoration models that are better tailored to handle non-Gaussian artifacts inherent to inverse problems. However, there is limited research exploring whether priors based on pre-trained restoration models can outperform those based on Gaussian denoisers. In this paper, we present evidence that priors derived from deep models pre-trained as general restoration operators can surpass

those trained exclusively for Gaussian denoising. We introduce a novel framework called ***Stochastic deep Restoration Priors (ShaRP)***, which provides a principled approach to integrate an ensemble of general restoration models as priors to regularize inverse problems. ShaRP is conceptually related to several recent papers exploring priors specified using other types of image restoration operators, such as, for example, image super-resolution models (Zhang et al., 2019; Liu et al., 2020; Gilton et al., 2021b; Hu et al., 2024c). The key benefit of ShaRP relative to prior work (Hu et al., 2024c) lies in its versatility, enabling seamless integration of a wide-range of restoration models trained on multiple degradation types (e.g., various undersampling masks in MRI or blur kernels in image deblurring). By using more versatile restoration models, ShaRP improves upon traditional methods using Gaussian denoiser priors and restoration priors in two key ways: (a) ShaRP improved performance by using restoration models better suited to mitigating the structured artifacts that arise during inference, which are often linked to the characteristics of the underlying inverse problem. (b) The restoration models in ShaRP can sometimes be directly trained in a self-supervised manner without fully sampled measurement data.

We present new theoretical and numerical results highlighting the potential of using an ensemble of restoration models as image priors. Our first theoretical result introduces a novel notion of regularization for inverse problems corresponding to the average of likelihoods associated with the degraded observations of an image. The proposed regularizer has an intuitive interpretation as promoting solutions whose multiple degraded observations resemble realistic degraded images. We show that ShaRP seeks to minimize an objective function containing this regularizer. Our second theoretical result analyzes the convergence of ShaRP iterations when using both exact and inexact minimum mean squared error (MMSE) restoration operators. Numerically, we show the practical relevance of ShaRP by applying it to MRI reconstruction with varying undersampling patterns and rates, using a fixed-rate pre-trained MRI reconstruction network as a prior. We also show that ShaRP can use a pre-trained image deblurring model to perform single image super-resolution (SISR). Our numerical experiments show that ShaRP effectively adapts the pre-trained restoration model as a prior, outperforming existing methods based on image denoisers and diffusion models, and achieving state-of-the-art results. Our experiments additionally highlight the benefit of using restoration models as priors by considering a setting where only undersampled and noisy MRI data is available for pre-training the prior. In such cases, self-supervised training of a restoration model is feasible, whereas training a Gaussian denoiser requires fully sampled data.

2 Background

Inverse Problems. Many computational imaging tasks can be formulated as inverse problems, where the goal is to reconstruct an unknown image $\boldsymbol{x} \in \mathbb{R}^n$ from its corrupted measurement

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

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a measurement operator and $\boldsymbol{e} \in \mathbb{R}^m$ is the noise. A common approach to addressing inverse problems is to formulate them as an optimization problem

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x} \in \mathbb{R}^n}{\operatorname{arg\,min}} f(\boldsymbol{x}) \quad \text{with} \quad f(\boldsymbol{x}) = g(\boldsymbol{x}) + h(\boldsymbol{x}) \,, \tag{2}$$

where g is the data-fidelity term that quantifies the fit to the measurement $\boldsymbol{y}$ and h is a regularizer that incorporates prior information on $\boldsymbol{x}$. For instance, typical functions used in imaging inverse problems are the least-squares term $g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2$ and the total variation (TV) regularizer $h(\boldsymbol{x}) = \tau \|\boldsymbol{D}\boldsymbol{x}\|_1$, where $\boldsymbol{D}$ is the image gradient and $\tau > 0$ is a regularization parameter.

Deep Learning. DL has emerged as a powerful tool for addressing inverse problems (McCann et al., 2017; Ongie et al., 2020; Wen et al., 2023). Instead of explicitly defining a regularizer, DL methods use deep neural networks (DNNs) to map the measurements to the desired images (Wang et al., 2016; Jin et al., 2017; Kang et al., 2017; Chen et al., 2017; Delbracio et al., 2021; Delbracio & Milanfar, 2023). Model-based DL (MBDL) is a widely-used sub-family of DL algorithms that integrate physical measurement models with priors specified using CNNs (see reviews by Ongie et al. (2020); Monga et al. (2021)). The literature of MBDL is vast, but some well-known examples include plug-and-play priors (PnP), regularization by denoising (RED), deep unfolding (DU), compressed sensing using generative models (CSGM), and deep equilibrium models (DEQ) (Bora et al., 2017; Romano et al., 2017; Zhang & Ghanem, 2018; Hauptmann et al., 2018; Gilton et al.,

2021a; Liu et al., 2022; Hu et al., 2024d). These approaches come with different trade-offs in terms of imaging performance, computational and memory complexity, flexibility, need for supervision, and theoretical understanding.

Denoisers as Priors. Score-based models (SBMs) are a powerful subset of DL methods for solving inverse problems that use deep Gaussian denoisers as imaging priors. Plug-and-Play (PnP) methods can be viewed as SBMs that incorporate denoisers within iterative optimization algorithms (see recent reviews (Ahmad et al., 2020; Kamilov et al., 2023)). These approaches construct a cost function by combining an explicit likelihood with a score function implicitly defined by the denoiser prior. Over the past few years, numerous variants of PnP have been developed (Venkatakrishnan et al., 2013; Romano et al., 2017; Metzler et al., 2018; Zhang et al., 2017; Meinhardt et al., 2017; Dong et al., 2019; Zhang et al., 2019; Wei et al., 2020; Hurault et al., 2022a), which has motivated an extensive research into their theoretical properties and empirical effectiveness (Chan et al., 2017; Buzzard et al., 2018; Ryu et al., 2019; Sun et al., 2019; Tirer & Giryes, 2019; Teodoro et al., 2019; Xu et al., 2020; Sun et al., 2021; Cohen et al., 2021b; Hu et al., 2022; Laumont et al., 2022; Hurault et al., 2022b; Gan et al., 2023a; Cohen et al., 2021a; Fang et al., 2024; Renaud et al., 2024b; Hu et al., 2024a; Renaud et al., 2024a; Terris et al., 2024). Diffusion Models (DMs) represent another category of SBMs; they are trained to learn the score function of the underlying probability distribution governed by stochastic differential equations (SDEs) (Ho et al., 2020; Song et al., 2021). Once trained, these models can be used as powerful priors for inverse problems by leveraging their learned score functions. Specifically, pre-trained DMs facilitate posterior sampling by guiding the denoising process to generate data consistent with observed measurements. This approach enables DMs to address inverse problems, often achieving impressive perceptual performance even for highly ill-posed inverse problems (Chung et al., 2023; Zhu et al., 2023; Wang et al., 2023; Feng et al., 2023; Sun et al., 2024; Wu et al., 2024; Song et al., 2024; Hu et al., 2024b; Alçalar & Akçakaya, 2024; Zhao et al., 2024; Rout et al., 2024; Bai et al., 2024).

Restoration Networks as Priors. In addition to denoiser-based methods, recent work has also considered using restoration models as implicit priors for solving inverse problems (Zhang et al., 2019; Liu et al., 2020; Gilton et al., 2021b; Hu et al., 2024c). It has been observed that pre-trained restoration models can be effective priors for addressing unseen inverse problems, sometimes surpassing traditional denoiser-based approaches (Hu et al., 2024c). However, existing methods present two main limitations. First, restoration models considered so far have relied on a fixed prior tailored to a specific degradation. Although deep restoration models can be trained in various settings—such as different blur kernels for image deblurring or diverse undersampling masks for MRI reconstruction—current approaches do not leverage this capability, limiting their robustness to diverse artifacts. Second, prior work has not explored the potential of learning restoration priors directly from undersampled measurements, without access to fully sampled data. Unlike Gaussian denoisers, training without fully sampled data is natural for restoration models (Yaman et al., 2020; Liu et al., 2020; Tachella et al., 2022; Chen et al., 2022; Millard & Chiew, 2023; Gan et al., 2023b). It is also worth highlighting the related work that has explored using corrupt measurements for training Ambient DMs (Daras et al., 2023; Aali et al., 2024). Ambient DMs seek to sample from $p_{\boldsymbol{x}}$ using DMs trained directly on undersampled measurements. Thus, during inference Ambient DMs assume access to the image prior $p_{\boldsymbol{x}}$, while ShaRP only assumes access to the ensemble of likelihoods of multiple degraded observations.

Our contribution. **(1)** We propose ShaRP, a new framework for solving inverse problems leveraging a set of priors implicit in a pre-trained deep restoration network. ShaRP generalizes Regularization by Denoising (RED) (Romano et al., 2017) and Stochastic Denoising Regularization (SNORE) (Renaud et al., 2024b) by using more flexible restoration operators and extends Deep Restoration Priors (DRP) (Hu et al., 2024c) by using multiple restoration priors instead of relying on a single one. **(2)** We introduce a novel regularization concept for inverse problems that encourages solutions that produce degraded versions closely resembling real degraded images. For example, our regularizer favors an MR image solution only if its various degraded versions are consistent with the characteristics of actual degraded MR images. **(3)** We show that ShaRP minimizes a composite objective that incorporates our proposed regularizer. We provide a theoretical analysis of its convergence for both exact and approximate MMSE restoration operators. **(4)** We implement ShaRP with both supervised and self-supervised restoration models as priors and test it on two inverse problems: compressed sensing MRI (CS-MRI) and single-image super-resolution (SISR). Our results highlight the capability of restoration models to achieve state-of-the-art performance. Notably, in the MRI context,

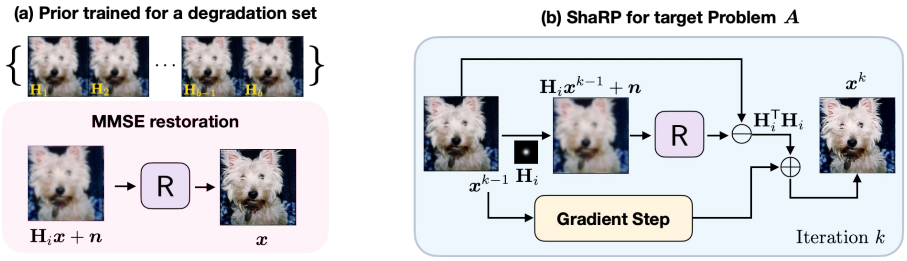

Figure 1: A restoration network trained on a set of tasks $\{\mathbf{H}_i\}$ can be used as a prior within ShaRP to address different target inverse problems without the need for retraining.

Algorithm 1 Stochastic deep Restoration Priors (ShaRP)

input: Initial value $\boldsymbol{x}^0 \in \mathbb{R}^n$, $\gamma > 0$, $\sigma > 0$, and $\tau > 0$
for $k = 1, 2, 3, \ldots$ **do**
  Sample a degradation operator: $\mathbf{H} \sim p(\mathbf{H})$
  $\boldsymbol{s} \leftarrow \mathbf{H}\boldsymbol{x}^{k-1} + \boldsymbol{n}$ with $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$
  $\boldsymbol{x}^k \leftarrow \boldsymbol{x}^{k-1} - \gamma\hat{\nabla} f(\boldsymbol{x}^{k-1})$
    where $\hat{\nabla} f(\boldsymbol{x}^{k-1}) := \nabla g(\boldsymbol{x}^{k-1}) + (\tau/\sigma^2)\mathbf{H}^{\mathsf{T}}\mathbf{H}(\boldsymbol{x}^{k-1} - \mathsf{R}(\boldsymbol{s}, \mathbf{H}))$
end for

we show that restoration networks trained directly on subsampled and noisy MRI data can serve as effective priors, a scenario where training traditional Gaussian denoisers is infeasible.

3 Stochastic Deep Restoration Priors

ShaRP is presented in Algorithm 1. It considers a prior based on a deep restoration model $\mathsf{R}(\boldsymbol{s}, \mathbf{H})$ pre-trained using the family of degradation operators, such as blur kernels or MRI masks. More specifically, the deep restoration model R is trained to solve the following set of restoration problems

$$\boldsymbol{s} = \mathbf{H}\boldsymbol{x} + \boldsymbol{n} \quad \text{with} \quad \boldsymbol{x} \sim p_{\boldsymbol{x}}, \quad \mathbf{H} \sim p(\mathbf{H}), \quad \boldsymbol{n} \sim \mathcal{N}(0, \sigma^2\mathbf{I}), \tag{3}$$

where $\boldsymbol{n}$ is the AWGN vector with variance σ^2 and $p_{\boldsymbol{x}}$ denotes the probability distribution of the target images, and $p(\mathbf{H})$ is the probability density of considered degradation operators. Importantly, the restoration problems (3) are used exclusively for training R and do not need to match the target inverse problem (1), which involves the measurement operator $\boldsymbol{A}$.

Our analysis below shows that $\widehat{\nabla} f$ corresponds to a stochastic approximation of an objective function of form $f = g + h$. Similar to traditional stochastic gradient methods, ShaRP can be implemented using various selection strategies for the degradation operators.

Each iteration of ShaRP has an intuitive interpretation, where the next solution is obtained by combining the gradient of the data-fidelity term ∇g and the residual of restored image corresponding to the selected degradation operator. When $\mathbf{H} = \mathbf{I}$, then the restoration prior reduces to the Gaussian denoiser, and ShaRP can be viewed as an instance of the Regularization by Denoising (RED) method (Romano et al., 2017) and Stochastic Denoising Regularization (SNORE) (Renaud et al., 2024b). On the other hand, there is a single $\mathbf{H}$, ShaRP can be viewed as the instance of the Deep Restoration Priors (DRP) method (Hu et al., 2024c). Thus, ShaRP can be viewed as a generalization of both that can account for a wide-range of degradation operators.

4 Theoretical Analysis of ShaRP

We present two key theoretical results on ShaRP. The first introduces a closed-form regularizer minimized by ShaRP. The second examines convergence of ShaRP when using inexact MMSE operators, highlighting its stable convergence even with approximate MMSE estimators.

Consider a restoration model that perform MMSE estimation of $\boldsymbol{x} \in \mathbb{R}^n$ for problems (3)

$$\mathsf{R}^*(\boldsymbol{s}, \mathbf{H}) = \mathbb{E}\left[\boldsymbol{x}|\boldsymbol{s}, \mathbf{H}\right] = \int \boldsymbol{x}\, p(\boldsymbol{x}|\boldsymbol{s}, \mathbf{H})\, \mathrm{d}\boldsymbol{x} = \frac{1}{p(\boldsymbol{s}|\mathbf{H})} \int \boldsymbol{x}\, G_\sigma(\boldsymbol{s} - \mathbf{H}\boldsymbol{x}) p_{\boldsymbol{x}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}. \tag{4}$$

where we used the probability density of the observation $\boldsymbol{s}$ conditioned on the operator $\mathbf{H}$

$$p(\boldsymbol{s}|\mathbf{H}) = \int G_\sigma(\boldsymbol{s} - \mathbf{H}\boldsymbol{x}) p_{\boldsymbol{x}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}. \tag{5}$$

The function G_σ in (5) denotes the Gaussian density function with the standard deviation $\sigma > 0$.

We propose the ShaRP regularizer

$$h(\boldsymbol{x}) = \tau \mathbb{E}_{\boldsymbol{s} \sim G_\sigma(\boldsymbol{s} - \mathbf{H}\boldsymbol{x}), \mathbf{H} \sim p(\mathbf{H})} \left[-\log p(\boldsymbol{s}|\mathbf{H})\right], \tag{6}$$

where $\tau > 0$ is the regularization parameter and $p(\mathbf{H})$ is the distribution of considered degradations. The regularizer h is minimized if degraded versions of $\boldsymbol{x}$ are highly probable in the distribution $p(\boldsymbol{s}|\mathbf{H})$, where $\mathbf{H}$ is sampled from $p(\mathbf{H})$. In other words, a solution $\widehat{\boldsymbol{x}}$ is considered good if its degraded versions $\mathbf{H}\widehat{\boldsymbol{x}}$ match the degraded versions $\mathbf{H}\boldsymbol{x}$ of clean images $\boldsymbol{x} \sim p_{\boldsymbol{x}}$, for all $\mathbf{H} \sim p(\mathbf{H})$. The key benefit of the proposed regularizer in (6) lies in its versatility, enabling seamless integration of a wide-range of degradation operators within a unified formulation. In particular, this formulation remains compatible with Gaussian denoisers, as $\mathbf{H} = \mathbf{I}$ can always be incorporated into $p(\mathbf{H})$.

We are now ready to state our first theoretical result.

Theorem 1. *Assume that the prior density $p_{\boldsymbol{x}}$ is non-degenerate over $\mathbb{R}^n$ and let R^* be the MMSE restoration operator* (4) *corresponding to the restoration problems* (3)*. Then, we have that*

$$\nabla h(\boldsymbol{x}) = \frac{\tau}{\sigma^2} \left(\mathbb{E}_{\boldsymbol{s} \sim G_\sigma(\boldsymbol{s} - \mathbf{H}\boldsymbol{x}), \mathbf{H} \sim p_{\mathbf{H}}} \left[\mathbf{H}^{\mathsf{T}} \mathbf{H} (\boldsymbol{x} - \mathsf{R}^*(\boldsymbol{s}, \mathbf{H})) \right] \right), \tag{7}$$

where h is the ShaRP regularizer in (6)*.*

The proof is provided in the appendix. Note that the expression within the square parenthesis in (7) matches the ShaRP update in Line 4 of Algorithm 1, which directly implies that ShaRP using the exact MMSE restoration operator R^* is a stochastic gradient method for minimizing $f = g + h$, where g is the data-fidelity term and h is the ShaRP regularizer in (6).

In practical scenarios, the learned restoration model may be imperfect, meaning it cannot be considered a perfect MMSE estimator. To demonstrate that ShaRP can effectively integrate such non-ideal restoration models into its framework as priors while still ensuring stable convergence, we provide the following proof. We now present the convergence analysis of ShaRP under a restoration operator R that *approximates* the true MMSE restoration operator R^*. For a given degraded observation $\boldsymbol{s} = \mathbf{H}\boldsymbol{x} + \boldsymbol{n}$ with $\mathbf{H} \sim p_{\mathbf{H}}$ and $\boldsymbol{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$, we define the stochastic gradient used by ShaRP

$$\widehat{\nabla} f(\boldsymbol{x}) = \nabla g(\boldsymbol{x}) + \widehat{\nabla} h(\boldsymbol{x}) \quad \text{with} \quad \widehat{\nabla} h(\boldsymbol{x}) \coloneqq \frac{\tau}{\sigma^2} \mathbf{H}^{\mathsf{T}} \mathbf{H} (\boldsymbol{x} - \mathsf{R}(\boldsymbol{s}, \mathbf{H})). \tag{8}$$

Since R is an inexact MMSE restoration operator, we also define the bias vector

$$\boldsymbol{b}(\boldsymbol{x}) = \frac{\tau}{\sigma^2} \mathbb{E}_{\boldsymbol{s} \sim G_\sigma(\boldsymbol{s} - \mathbf{H}\boldsymbol{x}), \mathbf{H} \sim p_{\mathbf{H}}} \left[\mathbf{H}^{\mathsf{T}} \mathbf{H} (\mathsf{R}^*(\boldsymbol{s}, \mathbf{H}) - \mathsf{R}(\boldsymbol{s}, \mathbf{H})) \right], \tag{9}$$

which quantifies the average difference between the exact and inexact MMSE restoration operators. Our analysis requires three assumptions that jointly serve as sufficient conditions for our theorem.

Assumption 1. *The function f has a finite minimum $f^* > -\infty$ and the gradient ∇f is Lipschitz continuous with constant $L > 0$.*

This is a standard assumption used in the analysis of gradient-based algorithms (see (Nesterov, 2004), for example). It is satisfied by a large number of functions, including the traditional least-squares data-fidelity function.

Assumption 2. *The stochastic gradient has a bounded variance for all $\boldsymbol{x} \in \mathbb{R}^n$, which means that there exists a constant $\nu > 0$ such that*

$$\mathbb{E}\left[\left\| \widehat{\nabla} f(\boldsymbol{x}) - \mathbb{E}\left[\widehat{\nabla} f(\boldsymbol{x}) \right] \right\|_2^2 \right] \leq \nu^2,$$

where expectations are with respect to $\mathbf{H} \sim p_{\mathbf{H}}$ and $\boldsymbol{s} \sim G_\sigma(\boldsymbol{s} - \mathbf{H}\boldsymbol{x})$.

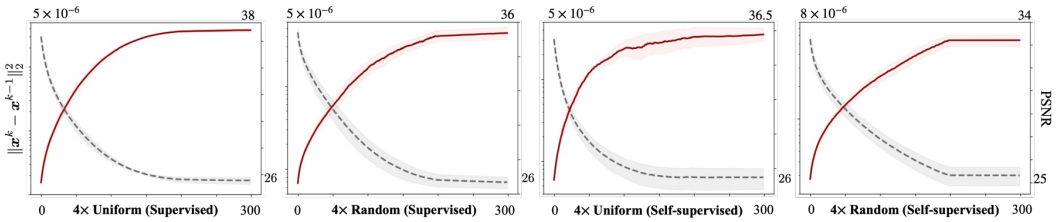

Figure 2: Convergence of ShaRP for $4\times$ accelerated MRI reconstruction on the fastMRI dataset. **(a)-(b)** depict the convergence behavior of ShaRP using restoration operators trained in a supervised manner, while **(c)-(d)** correspond to those trained in a self-supervised manner. The plots illustrate the average distance $\|\boldsymbol{x}^k - \boldsymbol{x}^{k-1}\|_2^2$ and PSNR relative to the ground truth, as a function of the iteration number, with shaded regions representing the standard deviation. Note the stable convergence of ShaRP with both types of priors.

This is another standard assumption extensively used in the analysis of online or stochastic optimization algorithms (Bertsekas, 2011; Ghadimi & Lan, 2016).

Assumption 3. *The bias $\boldsymbol{b}(\boldsymbol{x})$, as defined in (9), is bounded, which means that there exists $\varepsilon > 0$ such that for all $\boldsymbol{x} \in \mathbb{R}^n$*

$$\|\boldsymbol{b}(\boldsymbol{x})\|_2 \leq \varepsilon.$$

Note that our only assumption on the bias is that it is bounded, which is a relatively mild assumption in the context of biased stochastic gradient methods (Demidovich et al., 2023).

Proposition 1. *Run ShaRP for $t \geq 1$ iterations using the step-size $0 < \gamma \leq 1/L$ under Assumptions 1-3. Then, the sequence $\boldsymbol{x}^k$ generated by ShaRP satisfies*

$$\mathbb{E}\left[\frac{1}{t}\sum_{k=1}^{t}\|\nabla f(\boldsymbol{x}^{k-1})\|_2^2\right] \leq \frac{2}{\gamma t}(f(\boldsymbol{x}^0) - f^*) + \gamma L \nu^2 + \varepsilon^2.$$

The proof is provided in the appendix. This proposition shows that *in expectation*, ShaRP minimizes the norm of the gradient ∇f up to an error term that has two components, $\gamma L \nu^2$ and ϵ^2. Since the first component depends on γ, it can be made as small as desired by controlling the step-size γ. The second component only depends on the magnitude of the bias ε, which, in turn, directly depends on the accuracy of the restoration operator relative to the true MMSE restoration operator R^*. Note that the goal of Theorem 1 is not to provide the general analysis of biased SGD, which has been extensively studied elsewhere (see for example (Demidovich et al., 2023)), but rather show the stability of ShaRP using restoration networks that do not correspond to ideal MMSE estimators.

5 Numerical Results

We numerically validate ShaRP on two inverse problems of the form $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$: (*Compressive Sensing MRI (CS-MRI)* and (b) *Single Image Super Resolution (SISR)*. In both cases, $\boldsymbol{e}$ represents additive white Gaussian noise (AWGN). For the data-fidelity term in eq. (2), we use the ℓ_2-norm loss for both problems. Quantitative performance is evaluated using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). Additionally, for the SISR task, we include the Learned Perceptual Image Patch Similarity (LPIPS) metric to evaluate perceptual quality. Additional numerical results are provided in the supplementary material.

5.1 CS-MRI setting

The measurement of CS-PMRI can be modeled as $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{F}\boldsymbol{S}\boldsymbol{x} + \boldsymbol{e}$, where $\boldsymbol{P}$ is the k-space subsampling pattern, $\boldsymbol{F}$ is the Fourier transform operator, $\boldsymbol{S} = (\boldsymbol{S}_1, \cdots, \boldsymbol{S}_{n_c})$ are the multi-coil sensitivity maps, and $\boldsymbol{e}$ is the noise vector.

Dataset. We simulated multi-coil subsampled measurements using T2-weighted human brain MRI data from the open-access fastMRI dataset, which comprises 4,912 fully sampled multi-coil slices

for training and 470 slices for testing. Each slice has been cropped into a complex-valued image with dimensions 320×320. The coil sensitivity maps for each slice are precomputed using the ESPIRiT algorithm (Uecker et al., 2014). We simulated a Cartesian sampling pattern that subsamples along the k_y dimension while fully sampling along the k_x dimension.

Ensemble of Restoration Priors for CS-MRI. Recent methods, such as InDI (Delbracio & Milanfar, 2023) and I^{2}SB (Liu et al., 2023), introduce controllable processes for training an ensemble of restoration priors, where each prior functions as an MMSE restoration operator tailored to a specific setting. Building on this approach, we trained an $8\times$ uniform subsampling CS-MRI model with 8 distinct masks as our restoration prior. Similar to InDI, we decompose the original MRI degradation operator $\boldsymbol{M}$ into several convex combinations of the original task $\boldsymbol{M}$ and the identity mapping $\mathbf{I}$, represented by the new degradation operator $\mathbf{H}_\alpha = (1-\alpha)\mathbf{I} + \alpha\boldsymbol{M}$, with α controlling the degradation level. By selecting a range of α values, we create an ensemble of restoration tasks. Training the restoration network R to handle all these tasks allows it to function as an ensemble of MMSE restoration operators, $\mathsf{R}(\boldsymbol{s}, \mathbf{H}_\alpha) = \mathbb{E}\left[\boldsymbol{x}|\boldsymbol{s}, \mathbf{H}_\alpha\right]$. We used the MSE loss to train the restoration model.

Training restoration priors without groundtruth. When fully-sampled ground truth images are not available for training restoration priors, MRI restoration priors can be trained in self-supervised manner (Yaman et al., 2020; Millard & Chiew, 2023; Gan et al., 2023b; Hu et al., 2024d). In self-supervised training, rather than using the ground-truth image as the label, a separate subsampled measurement serves as the label. In such cases, we can train our priors using a weighted ℓ_2 loss function, following the self-supervised approach in (Gan et al., 2023b). We thus train the $8\times$ uniform subsampling CS-MRI model to handle eight distinct restoration operators, each corresponding to a different sampling mask.

Additional details on supervised and self-supervised restoration network training and our CS-MRI sampling masks are in Section B.1 of the appendix.

With the pre-trained $8\times$ models as ensembles of restoration priors, we evaluate ShaRP's performance across a variety of configurations, including two sub-sampling rates ($4\times$ and $6\times$), two mask types (uniform and random), and three noise levels ($\sigma = 0.005$, 0.01, and 0.015).

	4× Uniform						6× Uniform					
Noise level	$\sigma = 0.005$		$\sigma = 0.010$		$\sigma = 0.015$		$\sigma = 0.005$		$\sigma = 0.010$		$\sigma = 0.015$	
Metrics	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Zero-filled	26.93	0.848	26.92	0.847	26.90	0.848	22.62	0.728	22.60	0.726	22.59	0.721
TV	31.17	0.923	31.08	0.921	30.91	0.915	25.00	0.806	24.94	0.803	24.33	0.755
PnP-FISTA	35.88	0.938	31.14	0.894	30.32	0.846	26.30	0.822	25.97	0.786	25.46	0.747
PnP-ADMM	35.76	0.941	32.36	0.878	30.66	0.838	26.13	0.808	25.90	0.776	25.51	0.742
DRP	35.52	0.936	32.32	0.914	30.57	0.901	29.51	0.872	28.52	0.882	28.35	0.876
DPS	32.62	0.888	31.39	0.870	30.29	0.856	30.53	0.862	29.41	0.843	28.63	0.830
DDS	35.21	0.937	35.03	0.935	34.51	0.925	31.02	0.889	30.84	0.888	30.79	0.888
ShaRP	**37.59**	**0.963**	**35.81**	**0.951**	**34.92**	**0.942**	**33.42**	**0.940**	**32.86**	**0.932**	**32.09**	**0.922**

Table 1: Quantitative comparison of ShaRP with several baselines for CS-MRI using uniform masks at undersampling rates of 4 and 6 on fastMRI dataset. The **best** and second best results are highlighted. Notably, ShaRP outperforms SOTA methods based on denoisers and diffusion models.

Baselines. ShaRP was compared against several baseline methods, including denoiser-based approaches (PnP-FISTA (Kamilov et al., 2017), PnP-ADMM (Chan et al., 2017)) and diffusion model-based methods (DPS (Chung et al., 2023), DDS (Chung et al., 2024)). To highlight the advantages of using a stochastic set of restoration operators, we also compared ShaRP with the DRP method (Hu et al., 2024c), which applies only a single restoration operator. Additional details related to the baseline methods can be found in Section B.1 of the appendix.

Results with supervised MMSE Restoration operator. Figure 2 illustrates the convergence behavior of ShaRP on the test set with an acceleration factor of $R = 6$ and additional noise $\sigma = 0.01$. Table 1 provides a quantitative comparison of reconstruction performance across different acceleration factors and noise levels using a uniform sub-sampling mask. In all configurations, ShaRP consistently outperforms the baseline methods. The use of a set of restoration operators clearly

enhances ShaRP's performance, highlighting the effectiveness of employing multiple operators to maximize the regularization information provided by the restoration model. Figure 3 presents visual reconstructions for two test scenarios, where ShaRP accurately recovers fine brain details, particularly in the zoomed-in regions, while baseline methods tend to oversmooth or introduce artifacts. These results highlight ShaRP's superior ability to manage structured artifacts and preserve fine details, outperforming both denoiser-based and diffusion model-based methods.

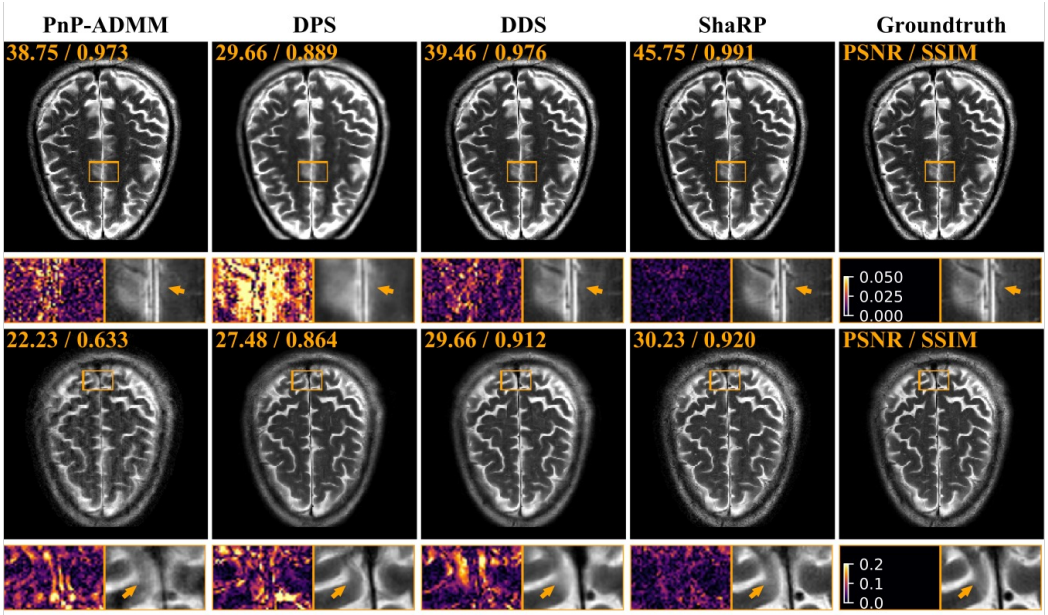

Figure 3: Visual comparison of ShaRP with baseline methods on CS-MRI. The top row shows results for a $4\times$ random mask with noise $\sigma = 0.005$, and the bottom row for a $6\times$ random mask with noise $\sigma = 0.015$. PSNR and SSIM values are in the top-left corner of each image. Error maps and zoomed-in areas highlight differences. Notably, ShaRP with stochastic priors outperforms state-of-the-art methods using denoiser and diffusion model priors.

Results with self-supervised MMSE Restoration operator. We further evaluate ShaRP's performance using an restoration model, learned in a self-supervised manner, as introduced in (Gan et al., 2023b). In this setting, we compare ShaRP against two classical methods for CS-MRI reconstruction without groundtruth: TV (Block et al., 2007) and GRAPPA (Griswold et al., 2002) and a recent state-of-the-art self-supervised deep unrolling method: SPICER (Hu et al., 2024d). As shown in Table 2, even in scenarios where only incomplete measurements ($8\times$ subsampled measurement) are available, ShaRP can effectively apply a self-supervised trained restoration prior to various reconstruction tasks. ShaRP using self-supervised restoration prior even outperforms DPIR and DPS that use Gaussian denoisers trained using fully-sampled ground truth images (see Table 4 in the appendix). Note that given only undersampled measurements, training Gaussian denoisers is not feasible.

5.2 Single Image Super Resolution (SISR)

The measurement operator in SISR can be written as $\boldsymbol{A} = \boldsymbol{S}\boldsymbol{K}$, where $\boldsymbol{K}$ represents convolution with the blur kernel, and $\boldsymbol{S}$ performs standard d-fold down-sampling. In our experiments, we use two Gaussian blur kernels $\boldsymbol{k}$, each with distinct standard deviations (1.25 and 1.5), and with down-sampling factor of 2. Both noisy and noise-free cases are considered to evaluate the noisy robustness of ShaRP. We randomly selected 100 images from the ImageNet test set, as provided in DiffPIR[1].

Ensemble of Restoration Priors for Image Deblurring. Following the approach used to train our CS-MRI restoration prior, we decompose the original deblurring task represented by the Gaussian blur operator $\boldsymbol{K}$ into convex combinations of the original task and the identity mapping I. This

[1] https://github.com/yuanzhi-zhu/DiffPIR/tree/main/testsets

	4× Random						6× Random					
Noise level	$\sigma = 0.005$		$\sigma = 0.010$		$\sigma = 0.015$		$\sigma = 0.005$		$\sigma = 0.010$		$\sigma = 0.015$	
Metrics	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PnP-ADMM	28.83	0.842	28.39	0.816	27.70	0.786	25.59	0.776	25.19	0.740	24.93	0.728
ADMM-TV	28.14	0.866	28.06	0.863	27.96	0.859	24.55	0.782	24.33	0.750	24.28	0.736
GRAPPA	28.09	0.792	25.39	0.699	23.94	0.649	25.67	0.737	23.72	0.646	22.51	0.595
SPICER	31.87	0.901	31.67	0.889	31.50	0.887	30.18	0.871	30.05	0.863	30.01	0.860
ShaRP$^{\text{self}}$	**33.87**	**0.909**	**33.64**	**0.900**	**33.21**	**0.892**	**30.87**	**0.899**	**30.36**	**0.890**	**30.21**	**0.875**

Table 2: PSNR (dB) and SSIM values for ShaRP with a self-supervised pre-trained restoration operator, compared to several baselines for CS-MRI with random undersampling masks at rates of 4 and 6 on the fastMRI dataset. The **best** and second best results are highlighted. For reference, the highlighted row presents a PnP method using a Gaussian denoiser, which requires fully sampled data for training. Note the excellent performance of ShaRP even using priors trained without fully-sampled ground-truth data.

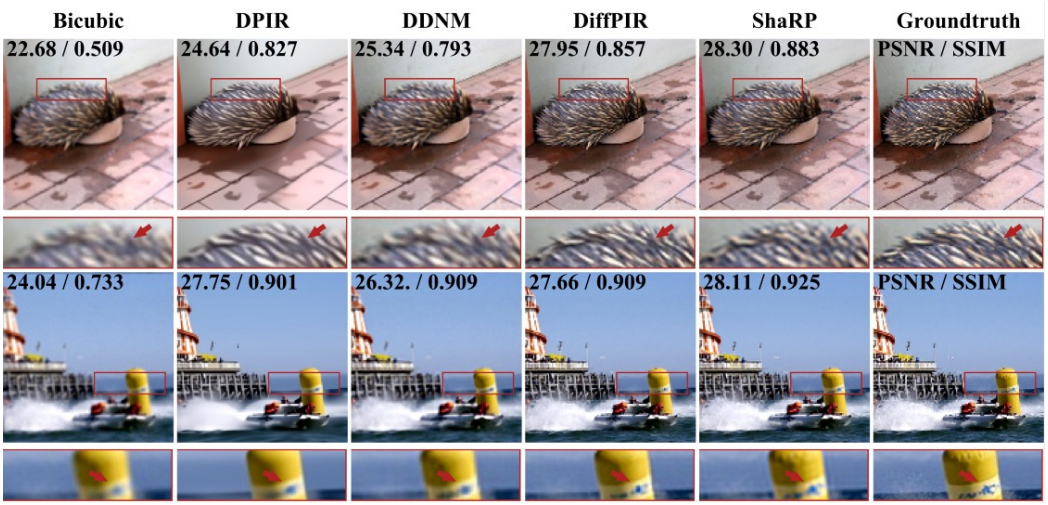

Figure 4: Visual comparison of ShaRP with several well-known methods on 2× SISR. The top row shows results for SISR with gaussian blur kernel with $\sigma = 1.25$, while the bottom row shows results for SISR with gaussian blur kernel with $\sigma = 1.5$. The quantities in the top-left corner of each image provide PSNR and SSIM values for each method. The squares at the bottom of each image visualize the zoomed area in the image.

results in a new degradation operator defined as $\mathbf{H}_\alpha = (1-\alpha)\mathsf{I} + \alpha\boldsymbol{K}$, where α controls the degradation level. By varying α, we generate multiple degradation operators, allowing us to train the restoration network R to handle all these operators. This training enables R to function as an ensemble of MMSE restoration operators, expressed as $\mathsf{R}(\boldsymbol{s}, \mathbf{H}_\alpha) = \mathbb{E}[\boldsymbol{x} \mid \boldsymbol{s}, \mathbf{H}_\alpha]$, where $\boldsymbol{s}$ is the degraded image and $\boldsymbol{x}$ is the original image. The original deblurring operator $\boldsymbol{K}$ corresponds to convolution with a Gaussian blur kernel of size 31×31 and standard deviation 3. More details on the deblurring restoration network training are in Section B.2 of the Appendix.

Baselines. We compared ShaRP with several baseline methods, including DPIR (Zhang et al., 2022), DPS (Chung et al., 2023), DDNM (Wang et al., 2023), DDRM (Kawar et al., 2022) and DiffPIR (Zhu et al., 2023). DPIR represents the state-of-the-art (SOTA) PnP method that uses pre-trained denoisers as priors to address SISR. In contrast, DPS, DDNM, and DiffPIR use different sampling strategies to leverage pre-trained unconditional diffusion models for solving SISR. More baseline details can be found in Section B.2 of the Appendix.

Results on SISR with deblurring prior. Figure 4 shows the visual reconstruction results for two settings with different blur kernels. As demonstrated, ShaRP successfully recovers most features and

maintains high data consistency with the available measurements. Table 3 provides a quantitative comparison of ShaRP against other baseline methods, evaluated across various blur kernels and noise levels. ShaRP achieves the highest PSNR and SSIM values but ranks second in perceptual performance (LPIPS). This is consistent with the SOTA perceptual performance of DMs on SISR. However, note how the use of a deblurring prior within ShaRP enables it to recover fine details, ensuring overall competitiveness of the perceptual quality of the ShaRP solutions.

												
Noise level	Noiseless			$\sigma = 0.01$			Noiseless			$\sigma = 0.01$		
Metrics	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
DPIR	28.10	0.809	0.305	28.05	0.807	0.308	27.90	0.803	0.314	27.87	0.800	0.314
DDNM	27.53	0.786	0.240	27.49	0.784	0.246	27.02	0.764	0.264	27.01	0.763	0.267
DPS	24.68	0.661	0.395	24.60	0.657	0.399	24.50	0.657	0.403	24.44	0.655	0.406
DiffPIR	28.92	0.852	**0.152**	28.63	0.839	**0.169**	28.59	0.834	**0.172**	28.02	0.819	**0.185**
DDRM	28.20	0.845	0.161	28.11	0.832	0.188	27.93	0.826	0.188	27.67	0.817	0.193
DRP	29.28	0.868	0.207	28.87	0.848	0.248	28.24	0.836	0.235	28.01	0.820	0.278
ShaRP	**30.09**	**0.891**	0.179	**29.03**	**0.852**	0.223	**29.28**	**0.872**	0.209	**28.06**	**0.821**	0.268

Table 3: Quantitative comparison of ShaRP with several baselines for SISR based on two different blur kernels on ImageNet dataset. The **best** and second best results are highlighted. Notably, ShaRP outperforms SOTA methods based on denoisers and diffusion models.

6 Conclusion

The work presented in this paper proposes a new ShaRP method for solving imaging inverse problems by using pre-trained restoration network as a prior, presents its theoretical analysis, and applies the method to two well-known inverse problems. Unlike previous approaches that relied on Gaussian denoisers or a single restoration prior, our method uses a set of restoration priors, each corresponding to different degradation settings. The numerical validation shows that ShaRP benefits from stochastically using multiple degradation priors, leading to better results. A key conclusion is the potential effectiveness of exploring priors beyond those defined by traditional Gaussian denoisers and a specific restoration operator.

Ethics Statement

To the best of our knowledge this work does not give rise to any significant ethical concerns.

References

A. Aali, G. Daras, B. Levac, S. Kumar, Al. G Dimakis, and J. I Tamir. Ambient diffusion posterior sampling: Solving inverse problems with diffusion models trained on corrupted data. *arXiv preprint arXiv:2403.08728*, 2024.

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

Y. U. Alçalar and M. Akçakaya. Zero-shot adaptation for approximate posterior sampling of diffusion models in inverse problems. *arXiv preprint arXiv:2407.11288*, 2024.

W. Bai, S. Chen, W. Chen, and H. Sun. Blind inversion using latent diffusion priors. *arXiv preprint arXiv:2407.01027*, 2024.

D. P. Bertsekas. Incremental proximal methods for large scale convex optimization. *Math. Program. Ser. B*, 129:163–195, 2011.

K. T. Block, M. Uecker, and J. Frahm. Undersampled radial MRI with multiple coils. iterative image reconstruction using a total variation constraint. *Magn. Reson. Med.*, 2007.

A. Bora, A. Jalal, E. Price, and A. G. Dimakis. Compressed sensing using generative priors. In *Int. Conf. Mach. Learn.*, pp. 537–546, August 2017.

G. T. Buzzard, S. H. Chan, S. Sreehari, and C. A. Bouman. Plug-and-play unplugged: Optimization free reconstruction using consensus equilibrium. *SIAM J. Imaging Sci.*, 11(3):2001–2020, Sep. 2018.

S. H. Chan, X. Wang, and O. A. Elgendy. Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. *IEEE Trans. Comp. Imag.*, 3(1):84–98, Mar. 2017.

D. Chen, J. Tachella, and M. Davies. Robust equivariant imaging: a fully unsupervised framework for learning to image from noisy and partial measurements. In *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR)*, 2022.

H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans. Med. Imag.*, 36(12): 2524–2535, Dec. 2017.

H. Chung, J. Kim, M. T. Mccann, M. L. K., and J. C. Ye. Diffusion posterior sampling for general noisy inverse problems. In *Int. Conf. on Learn. Represent.*, 2023.

H. Chung, S. Lee, and J. C. Ye. Decomposed diffusion sampler for accelerating large-scale inverse problems. In *Proc. Int. Conf. on Learn. Represent. (ICLR)*, 2024.

R. Cohen, Y. Blau, D. Freedman, and E. Rivlin. It has potential: Gradient-driven denoisers for convergent solutions to inverse problems. In *Proc. Adv. Neural Inf. Process. Syst. 34*, 2021a.

R. Cohen, M. Elad, and P. Milanfar. Regularization by denoising via fixed-point projection (red-pro). *SIAM Journal on Imaging Sciences*, 14(3):1374–1406, 2021b.

G.s Daras, K. Shah, Y. Dagan, A. Gollakota, A. Dimakis, and A. Klivans. Ambient diffusion: Learning clean distributions from corrupted data. In *Proc. Adv. Neural Inf. Process. Syst.*, 2023.

M. Delbracio and P. Milanfar. Inversion by direct iteration: An alternative to denoising diffusion for image restoration. *Trans. on Mach. Learn. Research*, 2023. ISSN 2835-8856.

M. Delbracio, H. Talebei, and P. Milanfar. Projected distribution loss for image enhancement. In *2021 Int. Conf. on Comput. Photography (ICCP)*, pp. 1–12, 2021.

Y. Demidovich, G. Malinovsky, I. Sokolov, and P. Richtarik. A guide through the zoo of biased SGD. In *Proc. Adv. Neural Inf. Process. Syst.*, 2023.

W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu. Denoising prior driven deep neural network for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 41(10):2305–2318, Oct 2019.

B. Efron. Tweedie's formula and selection bias. *J. Amer. Statistical Assoc.*, 106(496):1602–1614, 2011.

Z. Fang, S. Buchanan, and J. Sulam. What's in a prior? Learned proximal networks for inverse problems. *Proc. Int. Conf. on Learn. Represent. (ICLR)*, 2024.

B. T Feng, J. Smith, M. Rubinstein, H. Chang, K. L Bouman, and W. Freeman. Score-based diffusion models as principled priors for inverse imaging. In *Proc. IEEE Int. Conf. Comp. Vis. (ICCV)*, pp. 10520–10531, 2023.

W. Gan, S. Shoushtari, Y. Hu, J. Liu, H. An, and U. S. Kamilov. Block coordinate plug-and-play methods for blind inverse problems. In *Proc. Adv. Neural Inf. Process. Syst.*, 2023a.

W. Gan, C. Ying, P. E. Boroojeni, T. Wang, C. Eldeniz, Y. Hu, J. Liu, Y. Chen, H. An, and U. S Kamilov. Self-supervised deep equilibrium models with theoretical guarantees and applications to mri reconstruction. *IEEE Trans. Comp. Imag.*, 2023b.

S. Ghadimi and G. Lan. Accelerated gradient methods for nonconvex nonlinear and stochastic programming. *Math. Program. Ser. A*, 156(1):59–99, Mar. 2016.

D. Gilton, G. Ongie, and R. Willett. Deep equilibrium architectures for inverse problems in imaging. *IEEE Trans. Comput. Imag.*, 7:1123–1133, 2021a.

D. Gilton, G. Ongie, and R. Willett. Model adaptation for inverse problems in imaging. *IEEE Trans. Comput. Imag.*, 7:661–674, July 2021b.

M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Reson. in Med.*, 47(6):1202–1210, 2002.

A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, and S. Arridge. Model-based learning for accelerated, limited-view 3-D photoacoustic tomography. *IEEE Trans. Med. Imag.*, 37(6):1382–1393, 2018.

J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Proc. Adv. Neural Inf. Process. Syst.*, 33:6840–6851, 2020.

J. Hu, W. Gan, Z. Sun, H. An, and U. Kamilov. A plug-and-play image registration network. In *Proc. Int. Conf. on Learn. Represent. (ICLR)*, 2024a.

J. Hu, B. Song, X. Xu, L. Shen, and J. A Fessler. Learning image priors through patch-based diffusion models for solving inverse problems. *arXiv preprint arXiv:2406.02462*, 2024b.

Y. Hu, J. Liu, X. Xu, and U. S Kamilov. Monotonically convergent regularization by denoising. In *Int. Conf. on Image Processing (ICIP)*, pp. 426–430, 2022.

Y. Hu, M. Delbracio, P. Milanfar, and U. Kamilov. A Restoration Network as an Implicit prior. In *Proc. Int. Conf. on Learn. Represent. (ICLR)*, 2024c.

Y. Hu, W. Gan, C. Ying, T. Wang, C. Eldeniz, J. Liu, Y. Chen, H. An, and U. S Kamilov. Spicer: Self-supervised learning for mri with automatic coil sensitivity estimation and reconstruction. *Magn. Reson. Med.*, 2024d.

S. Hurault, A. Leclaire, and N. Papadakis. Gradient step denoiser for convergent plug-and-play. In *Proc. Int. Conf. on Learn. Represent. (ICLR)*, Kigali, Rwanda, May 1-5, 2022a.

S. Hurault, A. Leclaire, and N. Papadakis. Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization. In *Int. Conf. Mach. Learn.*, pp. 9483–9505, Kigali, Rwanda, 2022b. PMLR.

K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process.*, 26(9):4509–4522, Sep. 2017.

Z. Kadkhodaie and E. P. Simoncelli. Stochastic solutions for linear inverse problems using the prior implicit in a denoiser. In *Proc. Adv.Neural Inf. Process. Syst. 34*, pp. 13242–13254, December 6-14, 2021.

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

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

E. Kang, J. Min, and J. C. Ye. A deep convolutional neural network using directional wavelets for low-dose x-ray CT reconstruction. *Med. Phys.*, 44(10):e360–e375, 2017.

B. Kawar, M. Elad, S. Ermon, and J. Song. Denoising diffusion restoration models. *Proc. Adv. Neural Inf. Process. Syst.*, 35:23593–23606, 2022.

R. Laumont, V. De Bortoli, A. Almansa, J. Delon, A. Durmus, and M. Pereyra. Bayesian imaging using plug & play priors: When Langevin meets Tweedie. *SIAM J. Imaging Sci.*, 15(2):701–737, 2022.

J. Liang, J. Cao, G. Sun, K. Zhang, L. Van G., and R. Timofte. Swinir: Image restoration using swin transformer. In *Proc. IEEE Int. Conf. Comp. Vis. (ICCV)*, pp. 1833–1844, 2021.

G. Liu, A. Vahdat, D. Huang, E. A Theodorou, W. Nie, and A. Anandkumar. I2sb: image-to-image schrödinger bridge. In *Proc. Int. Conf. Machine Learning (ICML)*, pp. 22042–22062, 2023.

J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov. RARE: Image reconstruction using deep priors learned without ground truth. *IEEE J. Sel. Topics Signal Process.*, 14(6):1088–1099, Oct. 2020.

J. Liu, S. Asif, B. Wohlberg, and U. S. Kamilov. Recovery analysis for plug-and-play priors using the restricted eigenvalue condition. In *Proc. Adv. Neural Inf. Process. Syst. 34*, pp. 5921–5933, December 6-14, 2021.

J. Liu, X. Xu, W. Gan, S. Shoushtari, and U. S. Kamilov. Online deep equilibrium learning for regularization by denoising. In *Proc. Adv. Neural Inf. Process. Syst.*, 2022.

M. T. McCann, K. H. Jin, and M. Unser. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Process. Mag.*, 34(6):85–95, 2017.

T. Meinhardt, M. Moeller, C. Hazirbas, and D. Cremers. Learning proximal operators: Using denoising networks for regularizing inverse imaging problems. In *Proc. IEEE Int. Conf. Comp. Vis.*, pp. 1799–1808, Venice, Italy, Oct. 22-29, 2017.

C. Metzler, P. Schniter, A. Veeraraghavan, and R. Baraniuk. prDeep: Robust phase retrieval with a flexible deep network. In *Proc. 36th Int. Conf. Mach. Learn.*, pp. 3501–3510, Jul. 10–15 2018.

C. Millard and M. Chiew. A theoretical framework for self-supervised mr image reconstruction using sub-sampling via variable density noisier2noise. *IEEE Trans. Comp. Imag.*, 2023.

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

Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, 2004.

G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett. Deep learning techniques for inverse problems in imaging. *IEEE J. Sel. Areas Inf. Theory*, 1(1):39–56, May 2020.

E. T. Reehorst and P. Schniter. Regularization by denoising: Clarifications and new interpretations. *IEEE Trans. Comput. Imag.*, 5(1):52–67, March 2019.

M. Renaud, J. Liu, V. D. Bortoli, A. Almansa, and U. Kamilov. Plug-and-play posterior sampling under mismatched measurement and prior models. In *Proc. Int. Conf. on Learn. Represent. (ICLR)*, 2024a.

M. Renaud, J. Prost, A. Leclaire, and N. Papadakis. Plug-and-play image restoration with stochastic denoising regularization. In *Int. Conf. Machine Learning (ICML)*, 2024b.

H. Robbins. An empirical bayes approach to statistics. In *Proc. 3rd Berkeley Symp. Math. Statist. Probab., 1956*, volume 1, pp. 157–163, 1956.

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

L. Rout, Y. Chen, A. Kumar, C. Caramanis, S. Shakkottai, and W. Chu. Beyond first-order tweedie: Solving inverse problems using latent diffusion. In *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR)*, pp. 9472–9481, 2024.

E. K. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin. Plug-and-play methods provably converge with properly trained denoisers. In *Proc. Int. Conf. Machine Learning (ICML)*, volume 97, pp. 5546–5557, 2019.

B. Song, S. M. Kwon, Z. Zhang, X. Hu, Q. Qu, and L. Shen. Solving inverse problems with latent diffusion models via hard data consistency. In *Proc. Int. Conf. on Learn. Represent. (ICLR)*, 2024.

Y. Song, Jascha S., Diederik P K., Abhishek K., Stefano E., and Ben P. Score-based generative modeling through stochastic differential equations. In *Proc. Int. Conf. on Learn. Represent. (ICLR)*, 2021.

A. Sriram, J. Zbontar, T. Murrell, A. Defazio, C. L. Zitnick, N. Yakubova, F. Knoll, and P. M. Johnson. End-to-end variational networks for accelerated MRI reconstruction. *arXiv*, 2004.06688, 2020.

Y. Sun, B. Wohlberg, and U. S. Kamilov. An online plug-and-play algorithm for regularized image reconstruction. *IEEE Trans. Comput. Imag.*, 5(3):395–408, September 2019.

Y. Sun, S. Xu, Y. Li, L. Tian, B. Wohlberg, and U. S. Kamilov. Regularized Fourier ptychography using an online plug-and-play algorithm. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, pp. 7665–7669, Brighton, UK, May 12-17, 2019.

Y. Sun, Z. Wu, B. Wohlberg, and U. S. Kamilov. Scalable plug-and-play ADMM with convergence guarantees. *IEEE Trans. Comput. Imag.*, 7:849–863, July 2021.

Yu Sun, Zihui Wu, Yifan Chen, Berthy T Feng, and Katherine L Bouman. Provable probabilistic imaging using score-based generative priors. *IEEE Trans. Comp. Imag.*, 2024.

J. Tachella, C. Dongdong, and M. Davies. Unsupervised learning from incomplete measurements for inverse problems. In *Proc. Adv. Neural Inf. Process. Syst.*, pp. 4983–4995, 2022.

A. M. Teodoro, J. M. Bioucas-Dias, and M. A. T. Figueiredo. A convergent image fusion algorithm using scene-adapted Gaussian-mixture-based denoising. *IEEE Trans. Image Process.*, 28(1):451–463, January 2019.

M. Terris, T. Moreau, N. Pustelnik, and J. Tachella. Equivariant plug-and-play image reconstruction. In *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR)*, pp. 25255–25264, 2024.

T. Tirer and R. Giryes. Image restoration by iterative denoising and backward projections. *IEEE Trans. Image Process.*, 28(3):1220–1234, Mar. 2019.

M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig. ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magn. Reson. Med.*, 71(3):990–1001, 2014.

S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg. Plug-and-play priors for model based reconstruction. In *Proc. IEEE Global Conf. Signal Process. and Inf. Process.*, pp. 945–948, Dec. 3-5, 2013.

S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang. Accelerating magnetic resonance imaging via deep learning. In *Proc. Int. Symp. Biomedical Imaging (ISBI)*, pp. 514–517, April 2016. doi: 10.1109/ISBI.2016.7493320.

Y. Wang, J. Yu, and J. Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *Proc. Int. Conf. on Learn. Represent. (ICLR)*, 2023.

K. Wei, A. Aviles-Rivero, J. Liang, Y. Fu, C.-B. Schönlieb, and H. Huang. Tuning-free plug-and-play proximal algorithm for inverse imaging problems. In *Int. Conf. Machine Learning (ICML)*, 2020.

B. Wen, S. Ravishankar, Z. Zhao, R. Giryes, and J. C. Ye. Physics-driven machine learning for computational imaging. *IEEE Signal Process. Mag.*, 40(1):28–30, 2023.

Z. Wu, Y. Sun, Y. Chen, B. Zhang, Y. Yue, and K. L Bouman. Principled probabilistic imaging using diffusion models as plug-and-play priors. *arXiv preprint arXiv:2405.18782*, 2024.

Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *Proc. IEEE Int. Conf. Comp. Vis. (ICCV)*, pp. 13095–13105, 2023.

X. Xu, Y. Sun, J. Liu, B. Wohlberg, and U. S. Kamilov. Provable convergence of plug-and-play priors with mmse denoisers. *IEEE Signal Process. Lett.*, 27:1280–1284, 2020.

B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magn. Reson. Med.*, 84(6):3172–3191, Jul. 2020.

J. Zhang and B. Ghanem. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1828–1837, 2018.

K. Zhang, W. Zuo, S. Gu, and L. Zhang. Learning deep CNN denoiser prior for image restoration. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 3929–3938, July 21-26, 2017.

K. Zhang, W. Zuo, and L. Zhang. Deep plug-and-play super-resolution for arbitrary blur kernels. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1671–1681, Long Beach, CA, USA, June 16-20, 2019.

K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Trans. Patt. Anal. and Machine Intell.*, 44(10):6360–6376, October 2022.

J. Zhao, B. Song, and L. Shen. Cosign: Few-step guidance of consistency model to solve general inverse problems. *arXiv preprint arXiv:2407.12676*, 2024.

Y. Zhu, K. Zhang, J. Liang, J. Cao, B. Wen, R. Timofte, and L. Van G. Denoising diffusion models for plug-and-play image restoration. In *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR)*, pp. 1219–1229, 2023.

A THEORETICAL ANALYSIS OF SHARP

A.1 PROOF OF THEOREM 1

Theorem. *Assume that the prior density $p_{\boldsymbol{x}}$ is non-degenerate over $\mathbb{R}^n$ and let R^* be the MMSE restoration operator* (4) *corresponding to the restoration problems* (3)*. Then, we have that*

$$\nabla h(\boldsymbol{x}) = \frac{\tau}{\sigma^2}\left(\mathbb{E}_{\boldsymbol{s}\sim G_\sigma(\boldsymbol{s}-\mathbf{H}\boldsymbol{x}),\mathbf{H}\sim p_{\mathbf{H}}}\left[\mathbf{H}^{\mathsf{T}}\mathbf{H}(\boldsymbol{x}-\mathsf{R}^*(\boldsymbol{s},\mathbf{H}))\right]\right),$$

where h is the ShaRP regularizer in (6)*.*

Proof. The ShaRP regularizer $h(\boldsymbol{x})$ is defined as

$$\begin{aligned} h(\boldsymbol{x}) &= \tau\mathbb{E}_{\boldsymbol{s}\sim G_\sigma(\boldsymbol{s}-\mathbf{H}\boldsymbol{x}),\mathbf{H}\sim p_{\mathbf{H}}}\left[-\log p(\boldsymbol{s}|\mathbf{H})\right] \\ &= -\tau\int p(\mathbf{H})\left[\int G_\sigma(\boldsymbol{s}-\mathbf{H}\boldsymbol{x})\log p(\boldsymbol{s}|\mathbf{H})\,\mathrm{d}\boldsymbol{s}\right]\mathrm{d}\mathbf{H}, \end{aligned} \tag{10}$$

where G_σ is the Gaussian probability density with variance σ^2 and $p(\boldsymbol{s}|\mathbf{H})$ is the likelihood function for the degraded observation given the operator $\mathbf{H}$. The expectation over $p(\mathbf{H})$ accounts for the randomness of the restoration operator $\mathbf{H}$.

We start by relating the MMSE restoration operator to the score of the degraded observation

$$\nabla p(\boldsymbol{s}|\mathbf{H}) = \frac{1}{\sigma^2}\int(\mathbf{H}\boldsymbol{x}-\boldsymbol{s})\,G_\sigma(\boldsymbol{s}-\mathbf{H}\boldsymbol{x})p_{\boldsymbol{x}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x},$$

where $p_{\boldsymbol{x}}$ is the prior. By using the definition of the MMSE estimator, we obtain the relationship

$$\nabla\log p(\boldsymbol{s}|\mathbf{H}) = \frac{1}{\sigma^2}\left(\mathbf{H}\mathsf{R}^*(\boldsymbol{s},\mathbf{H})-\boldsymbol{s}\right). \tag{11}$$

Consider the function inside the parenthesis in the expression for the ShaRP regularizer (10)

$$\rho(\boldsymbol{z}) \coloneqq (G_\sigma * \log p_{\boldsymbol{s}|\mathbf{H}})(\boldsymbol{z}) = \int G_\sigma(\boldsymbol{z}-\boldsymbol{s})\log p(\boldsymbol{s}|\mathbf{H})\,\mathrm{d}\boldsymbol{s},$$

where $\boldsymbol{z}$ has the same dimensions as $\boldsymbol{s}$ and $*$ denotes convolution. The gradient of ρ is given by

$$\begin{aligned} \nabla\rho(\boldsymbol{z}) &= (\nabla G_\sigma * \log p_{\boldsymbol{s}|\mathbf{H}})(\boldsymbol{z}) = (G_\sigma * \nabla\log p_{\boldsymbol{s}|\mathbf{H}})(\boldsymbol{z}) \\ &= \frac{1}{\sigma^2}\int G_\sigma(\boldsymbol{z}-\boldsymbol{s})\left[\mathbf{H}\mathsf{R}^*(\boldsymbol{s},\mathbf{H})-\boldsymbol{s}\right]\mathrm{d}\boldsymbol{s} \\ &= \frac{1}{\sigma^2}\left(\mathbf{H}\int \mathsf{R}^*(\boldsymbol{s},\mathbf{H})G_\sigma(\boldsymbol{z}-\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}-\boldsymbol{z}\right) \end{aligned}$$

where we used (11). By using $\boldsymbol{z}=\mathbf{H}\boldsymbol{x}$, we write the gradient with respect to $\boldsymbol{x}$

$$\nabla_{\boldsymbol{x}}\rho(\mathbf{H}\boldsymbol{x}) = \frac{1}{\sigma^2}\mathbf{H}^{\mathsf{T}}\mathbf{H}\left(\int \mathsf{R}^*(\boldsymbol{s},\mathbf{H})G_\sigma(\boldsymbol{s}-\mathbf{H}\boldsymbol{x})\,\mathrm{d}\boldsymbol{s}-\boldsymbol{x}\right)$$

By using this expression in (10), we obtain the desired result

$$\begin{aligned} \nabla h(\boldsymbol{x}) &= -\frac{\tau}{\sigma^2}\left[\int p(\mathbf{H})\int G_\sigma(\boldsymbol{s}-\mathbf{H}\boldsymbol{x})\left(\mathbf{H}^{\mathsf{T}}\mathbf{H}(\mathsf{R}^*(\boldsymbol{s},\mathbf{H})-\boldsymbol{x})\right)\mathrm{d}\boldsymbol{s}\,\mathrm{d}\mathbf{H}\right] \\ &= \frac{\tau}{\sigma^2}\mathbb{E}_{\boldsymbol{s}\sim G_\sigma(\boldsymbol{s}-\mathbf{H}\boldsymbol{x}),\mathbf{H}\sim p_{\mathbf{H}}}\left[\mathbf{H}^{\mathsf{T}}\mathbf{H}(\boldsymbol{x}-\mathsf{R}^*(\boldsymbol{s},\mathbf{H}))\right]. \end{aligned}$$

□

A.2 PROOF OF THEOREM 2

Theorem. *Run ShaRP for $t\geq 1$ iterations using the step-size $0<\gamma\leq 1/L$ under Assumptions 1-3. Then, the sequence $\boldsymbol{x}^k$ generated by ShaRP satisfies*

$$\mathbb{E}\left[\frac{1}{t}\sum_{k=1}^{t}\|\nabla f(\boldsymbol{x}^{k-1})\|_2^2\right] \leq \frac{2}{t}(f(\boldsymbol{x}^0)-f^*)+\gamma L\nu^2+\varepsilon^2.$$

Proof. First note that from the definition of the bias in eq. (9), we have that

$$\mathbb{E}\left[\widehat{\nabla} f(\boldsymbol{x}^{k-1}) \,|\, \boldsymbol{x}^{k-1}\right] = \nabla f(\boldsymbol{x}^{k-1}) + \boldsymbol{b}(\boldsymbol{x}^{k-1}), \tag{12}$$

where the expectation is with respect to $\boldsymbol{s} \sim G_\sigma(\boldsymbol{s} - \mathbf{H}\boldsymbol{x}^{k-1})$ and $\mathbf{H} \sim p_{\mathbf{H}}$. In order to simplify the notation, we will drop these subscripts from the expectations in the analysis below.

Consider the iteration $k \geq 1$ of ShaRP with inexact MMSE operator

$$\begin{aligned} f(\boldsymbol{x}^k) &\leq f(\boldsymbol{x}^{k-1}) + \nabla f(\boldsymbol{x}^{k-1})^{\mathsf{T}}(\boldsymbol{x}^k - \boldsymbol{x}^{k-1}) + \frac{L}{2}\|\boldsymbol{x}^k - \boldsymbol{x}^{k-1}\|_2^2 \\ &= f(\boldsymbol{x}^{k-1}) - \gamma \nabla f(\boldsymbol{x}^{k-1})^{\mathsf{T}} \widehat{\nabla} f(\boldsymbol{x}^{k-1}) + \frac{L\gamma^2}{2}\|\widehat{\nabla} f(\boldsymbol{x}^{k-1})\|^2, \end{aligned}$$

where in the first line we used the Lipschitz continuity of ∇f. By taking the expectation with respect to $\boldsymbol{s} \sim G_\sigma(\boldsymbol{s} - \mathbf{H}\boldsymbol{x}^{k-1})$ and $\mathbf{H} \sim p_{\mathbf{H}}$ on both sides of this expression, we get

$$\begin{aligned} \mathbb{E}[f(\boldsymbol{x}^k)|\boldsymbol{x}^{k-1}] &\leq f(\boldsymbol{x}^{k-1}) - \gamma \nabla f(\boldsymbol{x}^{k-1})^{\mathsf{T}}(\nabla f(\boldsymbol{x}^{k-1}) + \boldsymbol{b}(\boldsymbol{x}^{k-1})) + \frac{L\gamma^2}{2}\mathbb{E}\left[\|\widehat{\nabla} f(\boldsymbol{x}^{k-1})\|_2^2 | \boldsymbol{x}^{k-1}\right] \\ &\leq f(\boldsymbol{x}^{k-1}) - \frac{\gamma}{2}\|\nabla f(\boldsymbol{x}^{k-1})\|_2^2 + \frac{\gamma}{2}\|\boldsymbol{b}(\boldsymbol{x}^{k-1})\|_2^2 \\ &\quad + \frac{L\gamma^2}{2}\left(\mathbb{E}\left[\|\widehat{\nabla} f(\boldsymbol{x}^{k-1})\|_2^2 | \boldsymbol{x}^{k-1}\right] - \left(\mathbb{E}[\widehat{\nabla} f(\boldsymbol{x}^{k-1})|\boldsymbol{x}^{k-1}]\right)^2\right) \\ &\leq f(\boldsymbol{x}^{k-1}) - \frac{\gamma}{2}\|\nabla f(\boldsymbol{x}^{k-1})\|_2^2 + \frac{\gamma\varepsilon^2}{2} + \frac{L\gamma^2\nu^2}{2}. \end{aligned}$$

In the second row, we completed the square, applied eq. (12), and used the assumption that $\gamma \leq 1/L$. In the third row, we used the variance and bias bounds in Assumptions 2 and 3. By rearranging the expression, we get the following bound

$$\|\nabla f(\boldsymbol{x}^{k-1})\|_2^2 \leq \frac{2}{\gamma}\left(f(\boldsymbol{x}^{k-1}) - \mathbb{E}[f(\boldsymbol{x}^k)|\boldsymbol{x}^{k-1}]\right) + L\gamma\nu^2 + \varepsilon^2$$

By taking the total expectation, averaging over t iterations, and using the lower bound f^*, we get the desired result

$$\mathbb{E}\left[\frac{1}{t}\sum_{k=1}^{t}\|\nabla f(\boldsymbol{x}^{k-1})\|_2^2\right] \leq \frac{2}{\gamma t}(f(\boldsymbol{x}^0) - f^*) + L\gamma\nu^2 + \varepsilon^2.$$

□

A.3 Proof of Theorem 3:

In this section, we present a theorem that establishes the feasibility of learning an MMSE estimator from undersampled MRI measurements. The measurement model for CS-PMRI can be expressed as: $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{F}\boldsymbol{S}\boldsymbol{x} + \boldsymbol{e}$, where $\boldsymbol{P}$ is the k-space subsampling pattern, $\boldsymbol{F}$ is the Fourier transform operator, $\boldsymbol{S} = (\boldsymbol{S}_1, \cdots, \boldsymbol{S}_{n_c})$ are the multi-coil sensitivity maps, and $\boldsymbol{e}$ is the noise vector.

To show that an MMSE estimator can be learned from undersampled MRI data, we need the following assumption.

Assumption 4. $\mathbb{E}_{\boldsymbol{P}}[\boldsymbol{P}^{\mathsf{T}}\boldsymbol{P}]$ has a full rank and $\boldsymbol{F}\boldsymbol{S}$ is an orthogonal matrix, where the expectation is taken over $p_{\boldsymbol{P}}$.

This assumption implies that the *union* of all sampling matrices $\boldsymbol{P}$ spans the complete measurement domain, even though each individual $\boldsymbol{P}$ may remain undersampled. This property can be achieved by incorporating an additional weight $\boldsymbol{W}$ into the loss function, where: $\boldsymbol{W} = \boldsymbol{P}'\overline{\boldsymbol{W}}(\boldsymbol{P}'\overline{\boldsymbol{W}})^{\mathsf{T}} \in \mathbb{R}^{m\times m}$ denotes a subsampled variant of $\overline{\boldsymbol{W}}$ given $\boldsymbol{P}'$. It is worth noting that normalizing the coil sensitivities, such that $\boldsymbol{S}^{\mathsf{T}}\boldsymbol{S} = \mathsf{I}$, is a standard practice. This assumption is the same as provided in previous work (Gan et al., 2023b).

Proposition 2. *When Assumption 2 is satisfited,*

$$\mathbb{E}\left[\boldsymbol{M'}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{M'}\right] = \boldsymbol{I}\ ,$$

where $\boldsymbol{M'} = \boldsymbol{P'FS}$ and the expectation is with respect to $p_{\boldsymbol{M}}$. This proof is the same as provided in previous work (Gan et al., 2023b).

Theorem 2. *Under Assumption 4, the MMSE estimator* R *learned using the weighted self-supervised loss (ℓ_{self}) is equivalent to its supervised counterpart (ℓ_{sup}). Specifically, we have:*

$$\mathsf{R}_{\ell_{\mathsf{self}}}(\boldsymbol{\theta}) = \mathsf{R}_{\ell_{\mathsf{sup}}}(\boldsymbol{\theta})\ . \tag{13}$$

where

$$\ell_{\mathsf{sup}} = \mathbb{E}\left[\frac{1}{2}\left\|\overline{\boldsymbol{x}} - \boldsymbol{x}\right\|_2^2\right] \tag{14}$$

and

$$\ell_{\mathsf{self}} = \mathbb{E}\left[\frac{1}{2}\left\|\boldsymbol{M'}\overline{\boldsymbol{x}} - \boldsymbol{y'}\right\|_{\boldsymbol{W}}^2\right]\ . \tag{15}$$

The vector $\overline{\boldsymbol{x}} = \mathsf{R}(\boldsymbol{y})$ is MMSE estimation of R *for $\boldsymbol{y}$.*

Proof. For simplicity, we define $\boldsymbol{M'} = \boldsymbol{P'FS}$. Note that the self-supervised loss involves the term $\boldsymbol{M'}\bar{\boldsymbol{x}} - \boldsymbol{y'}$, where $\boldsymbol{y'} = \boldsymbol{M'}\boldsymbol{x} + \boldsymbol{e'}$

$$\boldsymbol{M'}\bar{\boldsymbol{x}} - \boldsymbol{y'} = \boldsymbol{M'}(\bar{\boldsymbol{x}} - \boldsymbol{x}) - \boldsymbol{e'}. \tag{16}$$

Thus, the self-supervised loss becomes:

$$\ell_{\text{self}} = \mathbb{E}\left[\frac{1}{2}\|\boldsymbol{M'}(\bar{\boldsymbol{x}} - \boldsymbol{x}) - \boldsymbol{e'}\|_{\boldsymbol{W}}^2\right]. \tag{17}$$

Expanding the squared term:

$$\begin{aligned}\|\boldsymbol{M'}(\bar{\boldsymbol{x}} - \boldsymbol{x}) - \boldsymbol{e'}\|_{\boldsymbol{W}}^2 &= \|\boldsymbol{M'}(\bar{\boldsymbol{x}} - \boldsymbol{x})\|_{\boldsymbol{W}}^2 - 2(\boldsymbol{M'}(\bar{\boldsymbol{x}} - \boldsymbol{x}))^{\mathsf{T}}\boldsymbol{W}\boldsymbol{e'} + \|\boldsymbol{e'}\|_{\boldsymbol{W}}^2 \\ &= (\bar{\boldsymbol{x}} - \boldsymbol{x})^{\mathsf{T}}\boldsymbol{M'}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{M'}(\bar{\boldsymbol{x}} - \boldsymbol{x}) - 2(\boldsymbol{M'}(\bar{\boldsymbol{x}} - \boldsymbol{x}))^{\mathsf{T}}\boldsymbol{W}\boldsymbol{e'} + \|\boldsymbol{e'}\|_{\boldsymbol{W}}^2.\end{aligned}$$

So that

$$\begin{aligned}&\mathbb{E}\left[\|\boldsymbol{M'}(\bar{\boldsymbol{x}} - \boldsymbol{x}) - \boldsymbol{e'}\|_{\boldsymbol{W}}^2\right] \\ &= \mathbb{E}\left[(\bar{\boldsymbol{x}} - \boldsymbol{x})^{\mathsf{T}}\boldsymbol{M'}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{M'}(\bar{\boldsymbol{x}} - \boldsymbol{x})\right] - \mathbb{E}\left[2(\boldsymbol{M'}(\bar{\boldsymbol{x}} - \boldsymbol{x}))^{\mathsf{T}}\boldsymbol{W}\boldsymbol{e'}\right] + \mathbb{E}\left[\|\boldsymbol{e'}\|_{\boldsymbol{W}}^2\right]. \\ &= \mathbb{E}\left[\|\overline{\boldsymbol{x}} - \boldsymbol{x}[\|_2^2\right] + \text{constant},\end{aligned}$$

where the first term equals to $\mathbb{E}\left[\|\overline{\boldsymbol{x}} - \boldsymbol{x}[\|_2^2\right]$ due to the Proposition 1 that $\mathbb{E}\left[\boldsymbol{M'}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{M'}\right] = \mathbf{I}$; the second term equals to zero because $\boldsymbol{e'}$ is zero-mean and independent of $\boldsymbol{M'}$ and $\boldsymbol{x}$; The third term, $\|\boldsymbol{e'}\|_{\boldsymbol{W}}^2$, does not depend on $\boldsymbol{x}$ and contributes a constant that does not affect the optimization for training the MMSE estimator R.

□

B EXPERIMENT DETAILS

B.1 IMPLEMENTATION DETAILS OF CS-MRI TASKS

Subsampling pattern for CS-MRI. In this paper, we explored two types of subsampling patterns for MRI reconstruction tasks. All undersampling masks were generated by first including a set number of *auto-calibration signal (ACS)* lines, ensuring a fully-sampled central k-space region.

Figure 5 illustrates the k-space trajectories for both random and uniform (equidistant) subsampling at acceleration factors of 4, 6, and 8. Notably, different patterns were used for training and testing. During training, our restoration prior was only trained on a uniform mask with a subsampling rate of 6. However, for inference, we employed both uniform and random masks at subsampling rates of 4 and 6, creating a mismatch between the pre-trained restoration prior and the test configurations.

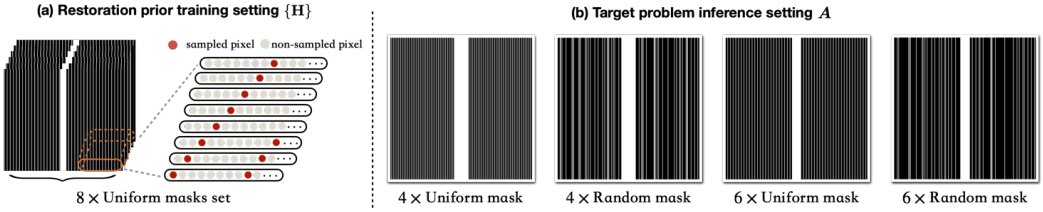

Figure 5: Illustration of the undersampling masks used for CS-MRI in this work. **(a)** The eight different $8\times$ uniform masks used for training the restoration prior. **(b)** The inference setting for ShaRP, demonstrating how the prior trained on the masks in **(a)** can be applied to solve other problems without retraining.

Algorithm 2 Supervised Training of CS-MRI Restoration Network

Require: dataset: $p(\boldsymbol{x})$, sampling operator set: $\{\boldsymbol{M}_1, \boldsymbol{M}_1, \cdots, \boldsymbol{M}_1\}$, Restoration model: $\mathsf{R}_\theta(\cdot, \alpha)$
  repeat:
    $\boldsymbol{x} \sim p(\boldsymbol{x})$, $\boldsymbol{M} \sim \{\boldsymbol{M}_1, \boldsymbol{M}_2, \cdots, \boldsymbol{M}_8\}$, $\boldsymbol{e} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, $\alpha \sim \mathcal{U}([0,1])$
    $\boldsymbol{y} = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{e}$
    $\mathsf{min}_\theta \left\| \mathsf{R}_\theta\left((1-\alpha)\boldsymbol{x} + \alpha\boldsymbol{M}^\mathsf{T}\boldsymbol{y}; \alpha\right) - \boldsymbol{x} \right\|_2^2$
  until converge

B.1.1 IMPLEMENTATION OF SUPERVISED PRIOR FOR CS-MRI

Models training for supervised case. We use the same U-Net architecture as employed in the official implementation of DDS[2] for $\mathsf{R}(\cdot;\alpha)$. For the supervised learning case, we select 1,000 different α values to train the model, following the α schedule outlined by I^{2}SB (Liu et al., 2023). The model is trained with Adam optimizer with a learning rate of 5×10^{-5}. As shown in Algorithm 2, we train our supervised learning model using eight different masks for $8\times$ uniform sampling CS-MRI reconstruction. In the pseudocode, $\{\boldsymbol{M}_1, \boldsymbol{M}_2, \cdots, \boldsymbol{M}_8\}$ represent the eight different MRI degradation operators, each defined by a unique sampling pattern, as shown in Figure 5 (a). This results in a total of 8,000 possible combinations of α values and sampling masks, effectively creating an ensemble of restoration priors during training.

Inference with a Subset of the Ensemble (Supervised Case). During inference, to simplify computation and focus on the most effective priors, we use only a subset of the supervised trained ensemble. Specifically, we fix the α value to a particular choice (e.g., $\alpha = 0.5$) and use the 8 different sampling masks $\{\boldsymbol{M}_1, \boldsymbol{M}_2, \cdots, \boldsymbol{M}_8\}$, resulting in 8 restoration priors.

Step size and regularization parameter. To ensure fairness, for each problem setting, each method—both proposed and baseline—is fine-tuned for optimal PSNR using 10 slices from a validation set separate from the test set. The same step size γ and regularization parameter τ are then applied consistently across the entire test set.

[2] https://github.com/HJ-harry/DDS

Baseline details. We compare ShaRP with several variants of denoiser- and diffusion model-based methods. For denoiser-based approaches, we include PnP-FISTA (Kamilov et al., 2023), PnP-ADMM (Chan et al., 2017). PnP-FISTA and PnP-ADMM correspond to the FISTA and ADMM variants of PnP, both utilizing AWGN denoisers built on DRUNet (Zhang et al., 2022). For diffusion model-based methods, we compare with DPS (Chung et al., 2023) and DDS (Chung et al., 2024), which use pre-trained diffusion models as priors and apply different posterior sampling strategies to address general inverse problems. We use the same pre-trained diffusion model configuration as outlined in the DDS paper. For all baseline methods, we fine-tuned their parameters to maximize the PSNR value. Notably, both the DRUNet denoiser and the diffusion model were trained using the same dataset employed for training our restoration prior. For a fair comparison, the diffusion model pre-trained for DDS and DPS use the same network architecture as our restoration network . All models are trained from scratch on the fastMRI training set, following the architecture settings provided in DDS[3]. We also compared with method that also use the deep restoration prior to solve general inverse problem: DRP (Hu et al., 2024c). For DRP, we utilize the same pre-trained restoration network as in ShaRP. However, instead of employing a set of degradation priors, DRP uses a single fixed prior. For a fair comparison, we selected the optimal fixed prior—defined by a fixed α and subsampling mask—based on PSNR performance on the validation set, and applied it accordingly.

B.1.2 Implementation of Self-Supervised Prior for CS-MRI

Algorithm 3 Self-Supervised Training of CS-MRI Restoration Network

Require: dataset: $p(\boldsymbol{y}_i, \boldsymbol{M}_i, \boldsymbol{y}_j, \boldsymbol{M}_j)$, Restoration model: $\mathsf{R}_\theta(\cdot)$
  repeat:
    $\boldsymbol{y}_i, \boldsymbol{M}_i, \boldsymbol{y}_j, \boldsymbol{M}_j \sim p(\boldsymbol{y}_i, \boldsymbol{M}_i, \boldsymbol{y}_j, \boldsymbol{M}_j)$, $\boldsymbol{e} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$
    $\mathsf{min}_\theta \left\| \boldsymbol{M}_j \mathsf{R}_\theta \left(\boldsymbol{M}_i^\mathsf{T} (\boldsymbol{y}_i + \boldsymbol{e}) \right) - \boldsymbol{y}_j \right\|_{\boldsymbol{W}}^2$
  until converge

Models training for (Self-Supervised Case). For self-supervised training, the ground truth reference $\boldsymbol{x}$ is not available as a label. Instead, as shown in Algorithm 3, we work with pairs of subsampled measurements, y_i and y_j, along with their corresponding sampling operators, $\boldsymbol{M}_i$ and $\boldsymbol{M}_j$. These paired measurements exhibit significant overlap within the shared *auto-calibration signal (ACS)* region, which increases the weighting of these overlapping k-space regions. Following the approach proposed by SSDEQ (Gan et al., 2023b), we introduce a diagonal weighting matrix $\boldsymbol{W}$ to account for the oversampled regions in the loss function. By incorporating this weighted loss, we are able to train our MMSE restoration operator using incomplete measurements alone. Furthermore, unlike the supervised case where we use the combination of α values to form an ensemble, in the self-supervised setting, we construct the ensemble using only eight different sampling masks across the entire dataset.

Inference Using All Restoration Priors (Self-Supervised Case). During inference in the self-supervised setting, we utilize all 8 restoration priors corresponding to the different sampling masks. By incorporating the entire ensemble, we fully leverage its capacity to remove the artifacts and enhance reconstruction performance.

Step size and regularization parameter. To ensure fairness, for each problem setting, each method—both proposed and baseline—is fine-tuned for optimal PSNR using 10 slices from a validation set separate from the test set. The same step size γ and regularization parameter τ are then applied consistently across the entire test set.

Baseline details. In the self-supervised setting, we compared ShaRP with two widely used traditional methods: TV (Block et al., 2007) and GRAPPA (Griswold et al., 2002), both of which address the restoration problem without requiring fully-sampled references. Additionally, we included SPICER (Hu et al., 2024d), a recent state-of-the-art self-supervised deep unrolling method designed for MRI reconstruction using only pairs of undersampled measurements. To ensure consistency, we trained the SPICER model on the same amount of paired data used for training our restoration prior in the $8\times$ uniform CS-MRI setting and applied it to other CS-MRI configurations.

[3] https://github.com/HJ-harry/DDS

B.2 IMPLEMENTATION DETAILS OF SISR TASKS

Algorithm 4 Gaussian Deblurring Restoration network training

Require: dataset:$p(\boldsymbol{x}, \boldsymbol{y})$, Gaussian blur operator: $\boldsymbol{K}$, $\mathsf{R}_\theta(\cdot, \alpha)$
  repeat:
    $\boldsymbol{x} \sim p(\boldsymbol{x})$, $\boldsymbol{e} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, $\alpha \sim \mathcal{U}([0,1])$
    $\min_\theta \|\mathsf{R}_\theta\left((1-\alpha)\boldsymbol{x} + \alpha\boldsymbol{K}\boldsymbol{x}; \alpha\right) - \boldsymbol{x}\|_2^2$
  until converge

Restoration Model training. We use the same U-Net architecture as the Gaussian deblurring model provided by I^{2}SB[4]. Utilizing the pre-trained checkpoints from their repository, we fine-tune our model accordingly. Specifically, we align with their codebase and configure the model type to OT-ODE to satisfy our MMSE restoration operator assumption.

To create an ensemble of restoration priors, we consider a family of degradation operators that are convex combinations of the identity mapping $\mathbf{I}$ and the Gaussian blur operator $\boldsymbol{K}$. The blurring operator $\boldsymbol{K}$ corresponds to convolution with a Gaussian blur kernel of size 31×31 and standard deviation 3. Specifically, we define the degradation operator as $\mathbf{H}_\alpha = (1-\alpha)\mathsf{I} + \alpha\boldsymbol{K}$, where $\alpha \in [0,1]$ controls the degradation level. By varying α, we generate multiple degradation operators, allowing us to train the restoration network R to handle all these operators, expressed as $\mathsf{R}(\boldsymbol{s}, \mathbf{H}_\alpha) = \mathbb{E}\left[\boldsymbol{x}|\boldsymbol{s}, \mathbf{H}_\alpha\right]$, where $\boldsymbol{s}$ is the degraded image and $\boldsymbol{x}$ is the original image.

We select 1,000 different α values from the interval $[0,1]$, following the α schedule outlined by I^{2}SB (Liu et al., 2023). This results in 1,000 different degradation operators $\mathbf{H}_\alpha$, effectively creating an ensemble of restoration priors during training. The model is trained using the Adam optimizer with a learning rate of 5×10^{-5}.

Inference with a Subset of the Ensemble. During inference, to simplify computation and focus on the most effective priors, we use only a subset of the supervised trained ensemble. Specifically, we select 6 α values, resulting in 6 restoration priors.

Step size and regularization parameter. To ensure fairness, for each problem setting, each method—both proposed and baseline—is fine-tuned for optimal PSNR using 5 images from a validation set separate from the test set. The same step size γ and regularization parameter τ are then applied consistently across the entire test set.

Baseline details. We compare ShaRP against several denoiser- and diffusion model-based methods. For denoiser-based approaches, we evaluate DPIR (Zhang et al., 2022), which relies on half-quadratic splitting (HQS) iterations with DRUNet denoisers. For diffusion model-based methods, we compare with DPS (Chung et al., 2023), DDNM (Wang et al., 2023), and DiffPIR (Zhu et al., 2023). These methods all use the same pre-trained diffusion models as priors, but each employs a distinct posterior sampling strategy to solve general inverse problems. We specifically use the pre-trained diffusion model from DiffPIR. We also compared with method that also use the deep restoration prior to solve general inverse problem: DRP (Hu et al., 2024c). For DRP, we utilize the same pre-trained deblurring network as in ShaRP. However, instead of employing a set of degradation priors, DRP uses a single fixed prior. For a fair comparison, we selected the optimal fixed prior—defined by a fixed α based on PSNR performance on the validation set, and applied it accordingly. For all baselines, we fine-tuned their parameters to maximize PSNR performance. Notably, the diffusion model backbone for all diffusion-based baselines was trained on the same dataset used to train our restoration prior.

[4] https://github.com/NVlabs/I2SB

C ADDITIONAL RESULTS FOR CS-MRI

C.1 PERFORMANCE OF SHARP FOR RANDOM SUBSAMPLING SETTING

Due to space constraints, we present only the quantitative performance for the uniform subsampling setting in the main paper. In this section, we further evaluate ShaRP's performance on random subsampling setting, with two sub-sampling rates ($4\times$ and $6\times$), and three noise levels ($\sigma = 0.005$, 0.01, and 0.015).

Table 4 provides a quantitative comparison of reconstruction performance across different acceleration factors and noise levels using a uniform sub-sampling mask. In all configurations, ShaRP consistently outperforms the baseline methods. The use of a set of restoration operators clearly enhances ShaRP's performance, highlighting the effectiveness of employing multiple operators to maximize the regularization information provided by the restoration model. Figure 6 presents visual reconstructions for two test scenarios, where ShaRP accurately recovers fine brain details, particularly in the zoomed-in regions, while baseline methods tend to oversmooth or introduce artifacts. These results highlight ShaRP's superior ability to manage structured artifacts and preserve fine details, outperforming both denoiser-based and diffusion model-based methods.

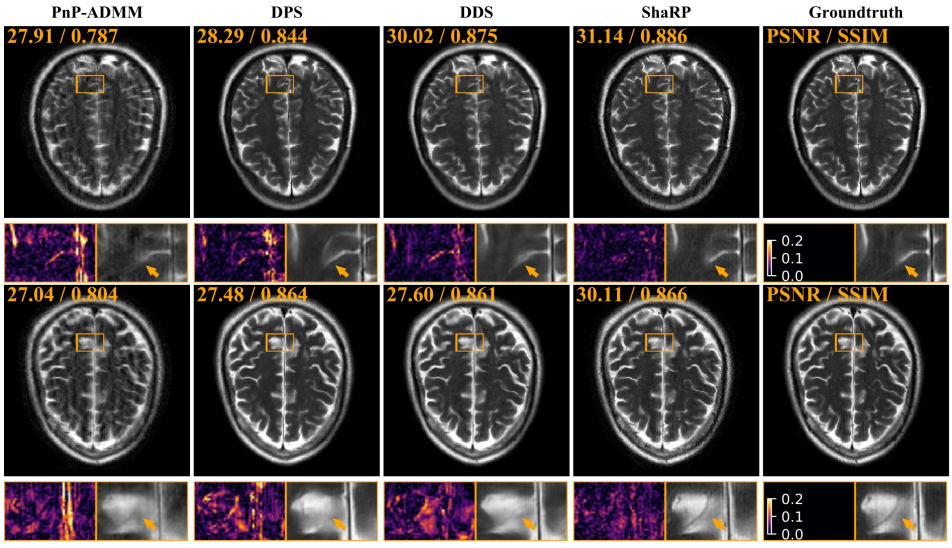

Figure 6: Visual comparison of ShaRP with baseline methods on CS-MRI for $6\times$ random sampling mask with noise $\sigma = 0.015$. PSNR and SSIM values are in the top-left corner of each image. Error maps and zoomed-in areas highlight differences. Notably, ShaRP with stochastic priors outperforms state-of-the-art methods using denoiser and diffusion model priors.

	4× Random						6× Random					
Noise level	$\sigma = 0.005$		$\sigma = 0.010$		$\sigma = 0.015$		$\sigma = 0.005$		$\sigma = 0.010$		$\sigma = 0.015$	
Metrics	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Zero-filled	25.83	0.815	25.81	0.812	25.76	0.807	22.68	0.724	22.67	0.722	22.67	0.719
TV	28.14	0.866	28.06	0.863	27.96	0.859	24.55	0.782	24.33	0.750	24.28	0.736
PnP-FISTA	29.31	0.863	28.40	0.817	27.49	0.799	26.01	0.797	25.63	0.756	24.94	0.717
PnP-ADMM	28.83	0.842	28.39	0.816	27.70	0.786	25.59	0.776	25.19	0.740	24.93	0.728
DRP	29.97	0.880	29.37	0.839	28.31	0.794	26.98	0.866	26.78	0.853	26.49	0.821
DPS	31.72	0.874	30.45	0.857	29.50	0.843	30.32	0.856	29.36	0.824	27.99	0.810
DDS	32.41	0.910	32.37	0.906	32.25	0.901	30.59	0.876	30.35	0.874	30.31	0.879
ShaRP	**34.66**	**0.949**	**33.57**	**0.920**	**33.18**	**0.931**	**31.53**	**0.924**	**31.46**	**0.918**	**31.45**	**0.914**

Table 4: Quantitative comparison of ShaRP with several baselines for CS-MRI using random masks at undersampling rates of 4 and 6 on fastMRI dataset. The **best** and second best results are highlighted. Notably, ShaRP outperforms SOTA methods based on denoisers and diffusion models.

C.2 PERFORMANCE OF ADDITIONAL BASELINE METHODS ON MATCHED AND MISMATCHED SETTINGS

In this section, we highlight an important observation: pre-trained restoration networks typically exhibit poor generalization to mismatched settings. We chose two commonly used methods (SwinIR (Liang et al., 2021) and E2E-VarNet (Sriram et al., 2020)) for the specific setting of CS-MRI. We trained them on the same $8\times$ uniform subsampling setting as our restoration prior and directly applied them to solve both matched and mismatched problems, as ShaRP did. As shown in the Table 5, the baseline method's performance drops significantly under mismatched conditions, whereas ShaRP maintains stable performance and convergence guarantees. This demonstrates ShaRP's ability to adapt pre-trained restoration models as priors and use it to solve problems under mismatched settings. As shown in the Figure 7, due to the mismatched settings, the two baseline methods suffer from over-smoothing, lack important details, and exhibit artifacts, whereas ShaRP still provides high-quality reconstruction performance. This indicates that ShaRP can balance data fidelity and the artifact removal capabilities of the prior model, leading to an artifact-free reconstruction that preserves important details.

Settings	$4\times$ Uniform		$4\times$ Random		$6\times$ Uniform		$6\times$ Random		$8\times$ Uniform	
Metrics	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SwinIR	24.78	0.849	25.09	0.841	29.55	0.907	27.98	0.819	29.37	0.898
E2E-VarNet	35.40	0.957	33.48	0.945	32.79	0.936	31.02	0.913	**32.59**	**0.919**
ShaRP	**37.59**	**0.963**	**34.66**	**0.949**	**33.42**	**0.940**	**31.53**	**0.924**	32.37	0.907

Table 5: Quantitative comparison of ShaRP with task-specific baselines trained on the $8\times$ uniform mask. Baselines perform well in matched settings (highlighted in the table) but show a significant drop under mismatched conditions. In contrast, ShaRP remains robust, handling both matched and mismatched scenarios effectively.

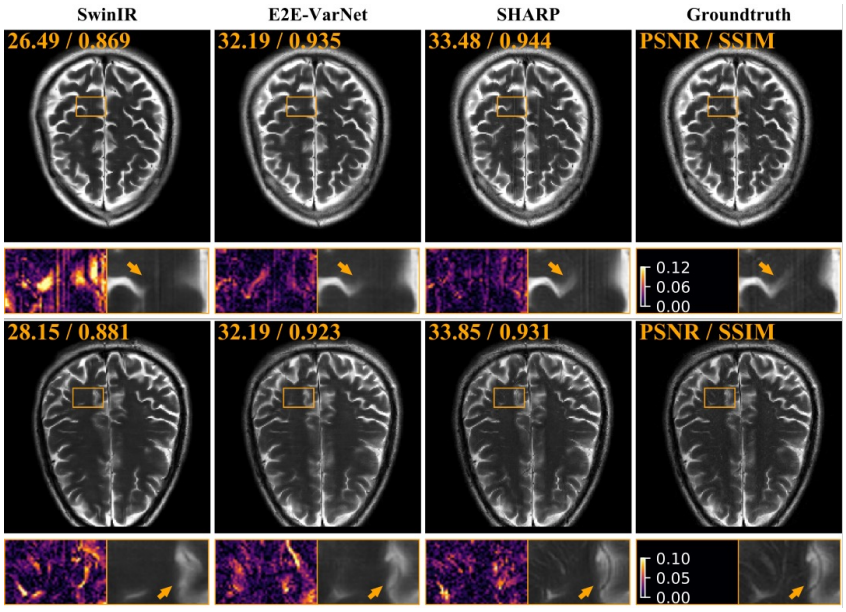

Figure 7: Visual comparison of ShaRP with task-specific baseline methods on CS-MRI for $6\times$ random sampling mask with noise $\sigma = 0.015$. PSNR and SSIM values are in the top-left corner of each image. Error maps and zoomed-in areas highlight differences. Notably, ShaRP with stochastic priors outperforms state-of-the-art methods using denoiser and diffusion model priors.

D ADDITIONAL VISUAL RESULTS FOR SISR

In this section, we present additional visual results to numerical comparisons for the SISR task.

D.1 ADDITIONAL VISUAL RESULTS AGAINST BASELINES

As illustrated in Figure 8 and Figure 9, ShaRP outperforms all baseline approaches under both blur kernel settings, achieving higher PSNR and SSIM values. Moreover, we maintain superior data consistency with the measurements while achieving enhanced perceptual quality. The use of an ensemble of deblurring priors enables our method to recover fine details at varying corruption levels, contributing to the improved performance.

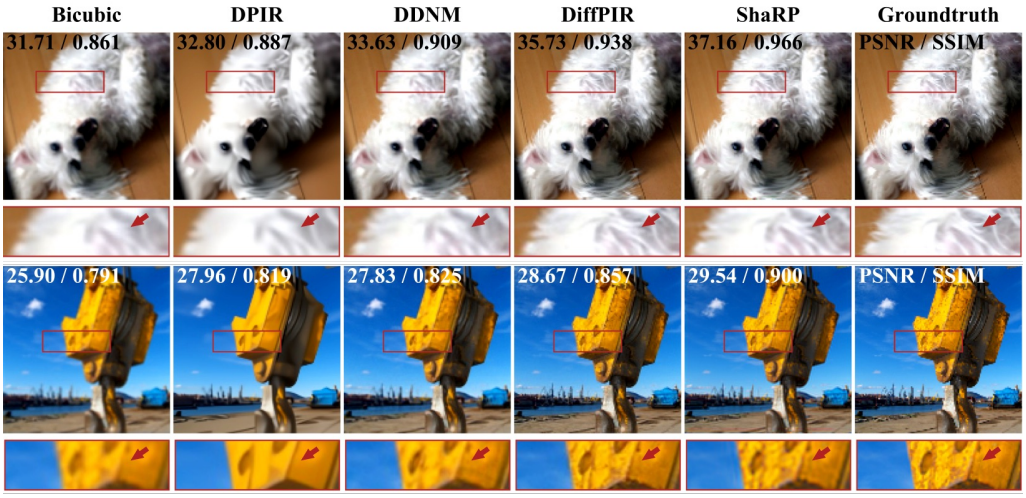

Figure 8: Visual comparison of ShaRP with several well-known methods on $2\times$ SISR with gaussian blur kernel with $\sigma = 1.5$. The quantities in the top-left corner of each image provide PSNR and SSIM values for each method. The squares at the bottom of each image visualize the zoomed area in the image.

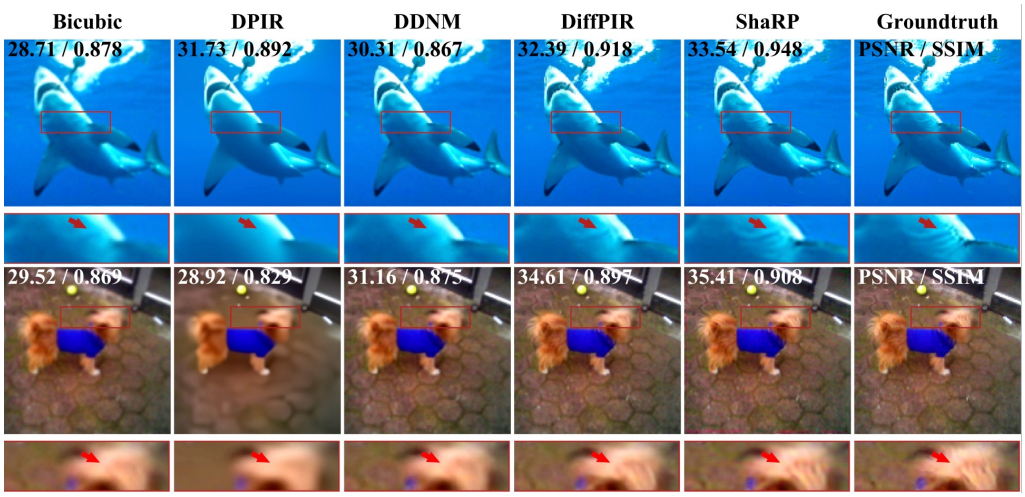

Figure 9: Visual comparison of ShaRP with several well-known methods on $2\times$ SISR with gaussian blur kernel with $\sigma = 1.5$. The quantities in the top-left corner of each image provide PSNR and SSIM values for each method. The squares at the bottom of each image visualize the zoomed area in the image.

D.2 ADDITIONAL VISUAL RESULTS AGAINST DRP

To further emphasize the necessity and advantages of using an ensemble of deblurring priors, as opposed to a fixed prior like in DRP (Hu et al., 2024c), we provide additional visual comparison results. As shown in Figure 10, ShaRP consistently recovers finer details, resulting in improved PSNR and SSIM scores, along with enhanced perceptual performance.

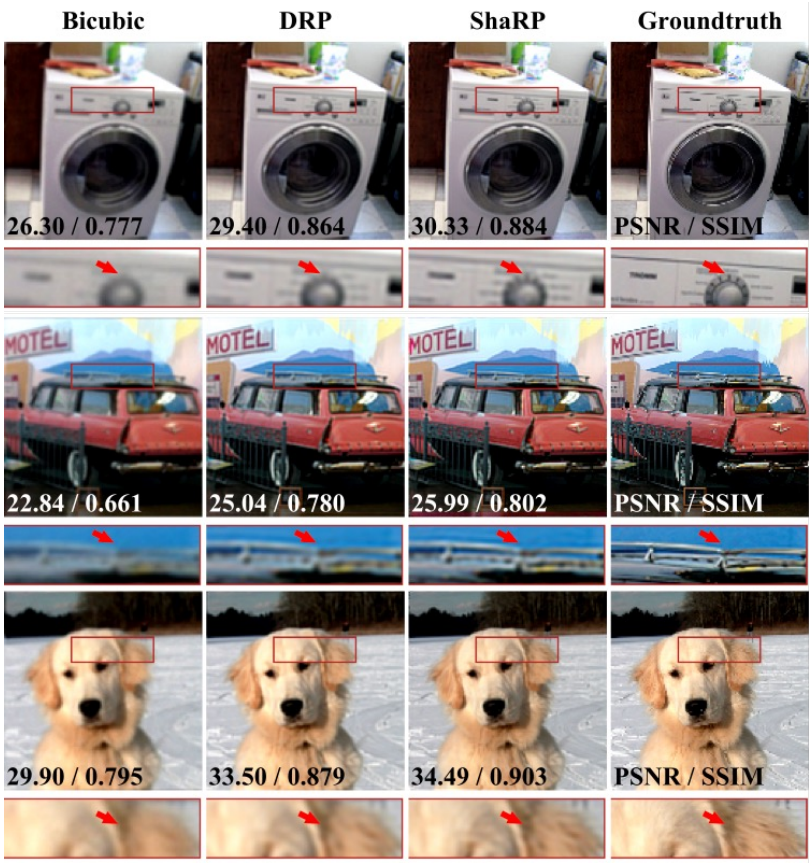

Figure 10: Visual comparison of ShaRP with DRP on $2\times$ SISR with gaussian blur kernel with $\sigma = 1.5$. The quantities in the bottom-left corner of each image provide PSNR and SSIM values for each method. The squares at the bottom of each image visualize the zoomed area in the image.

D.3 Additional Comparison against DDRM and DiffIR

To further evaluate ShaRP's performance against state-of-the-art diffusion-based methods, we included two additional baselines for comparison: DDRM (Kawar et al., 2022) and DiffIR (Xia et al., 2023). The experiment setting is $2\times$ SISR task with gaussian blur kernel with $\sigma = 1.25$ on ImageNet dataset. For DDRM, we utilized the same pre-trained unconditional diffusion backbone as DiffPIR, DDNM, and DDS, but followed the sampling procedure outlined in their original paper. For DiffIR, we directly used the provided checkpoint from the authors.

Metrics	PSNR	SSIM	LPIPS
DPIR	28.10	0.809	0.305
DDNM	27.53	0.786	0.240
DPS	24.68	0.661	0.395
DiffPIR	28.92	0.852	**0.152**
DiffIR	25.79	0.812	0.180
DDRM	28.20	0.845	0.161
DRP	29.28	0.868	0.207
ShaRP	**30.09**	**0.891**	0.179

Table 6: Quantitative comparison of ShaRP with several additional baselines for $2\times$ SISR with gaussian blur kernel with $\sigma = 1.25$ on ImageNet dataset. The **best** and second best results are highlighted. Notably, ShaRP outperforms SOTA methods based on denoisers and diffusion models.

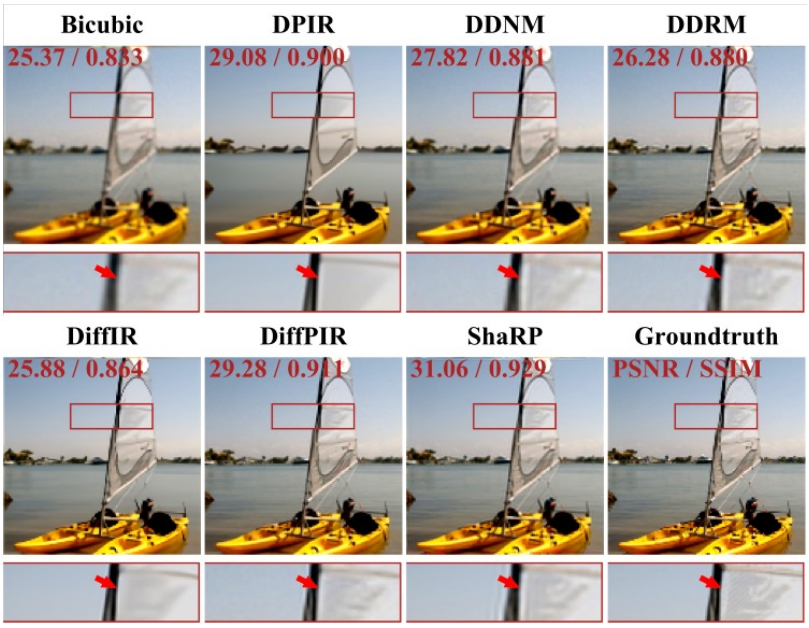

Figure 11: Visual comparison of ShaRP with additional baselines on $2\times$ SISR with gaussian blur kernel with $\sigma = 1.25$. The quantities in the bottom-left corner of each image provide PSNR and SSIM values for each method. The squares at the bottom of each image visualize the zoomed area in the image.

E ADDITIONAL EXPERIMENTS

In this section, we include two additional ablation studies to further highlights ShaRP's capability to leverage restoration priors for solving general inverse problems, as well as to evaluate its performance under different hyperparameter settings.

E.1 ABLATION STUDY ON USING SR PRIOR FOR CS-MRI TASK

To demonstrate the flexibility of our approach in integrating diverse restoration models to address general inverse problems, we conducted an additional ablation study using a pre-trained super-resolution network as a prior for solving the CS-MRI problem.

Algorithm 5 MRI Super Resolution network training

Require: dataset:$p(\boldsymbol{x}, \boldsymbol{y})$, $4\times$ bicubic downsampling operator: $\boldsymbol{K}$, $\mathsf{R}_\theta(\cdot, \alpha)$
  repeat:
    $\boldsymbol{x} \sim p(\boldsymbol{x})$, $\boldsymbol{e} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, $\alpha \sim \mathcal{U}([0,1])$
    $\min_\theta \left\| \mathsf{R}_\theta\left((1-\alpha)\boldsymbol{x} + \alpha \boldsymbol{D}^\mathsf{T}\boldsymbol{D}\boldsymbol{x}; \alpha\right) - \boldsymbol{x}\right\|_2^2$
  until converge

Models training for MRI-SR We use the same U-Net architecture as employed in the official implementation of DDS[5] for $\mathsf{R}(\cdot;\alpha)$. To create an ensemble of restoration priors, we consider a family of degradation operators that are convex combinations of the identity mapping $\mathbf{I}$ and the Gaussian blur operator $\boldsymbol{D}$. The $4\times$ bicubic downsampling operator $\boldsymbol{D}$ corresponds to bicubic downsample with factor equals to 4. Specifically, we define the degradation operator as $\mathbf{H}_\alpha = (1-\alpha)\mathsf{I} + \alpha\boldsymbol{D}^\mathsf{T}\boldsymbol{D}$, where $\alpha \in [0,1]$ controls the degradation level. By varying α, we generate multiple degradation operators, allowing us to train the restoration network R to handle all these operators, expressed as $\mathsf{R}(\boldsymbol{s}, \mathbf{H}_\alpha) = \mathbb{E}[\boldsymbol{x}|\boldsymbol{s}, \mathbf{H}_\alpha]$, where $\boldsymbol{s}$ is the degraded image and $\boldsymbol{x}$ is the original image.

We select 1,000 different α values from the interval $[0,1]$, following the α schedule outlined by I^{2}SB (Liu et al., 2023). This results in 1,000 different degradation operators $\mathbf{H}_\alpha$, effectively creating an ensemble of restoration priors during training. The model is trained using the Adam optimizer with a learning rate of 5×10^{-5}.

Using MRI-SR model as prior for CS-MRI task. During inference, to simplify computation and focus on the most effective priors, we use only a subset of the ensemble. Specifically, we select 6 α values, resulting in 6 restoration priors.

As shown in Table 7, under the $4\times$ uniform mask setting, employing the pre-trained MRI-SR model as prior allows ShaRP to outperform denoiser- and diffusion-based approaches. However, its performance remains inferior to ShaRP with a mismatched CS-MRI-specific prior. In the $4\times$ random mask setting, ShaRP with the pre-trained MRI-SR model as prior continues to surpass PnP-based methods that utilize a denoiser prior but performs worse than approaches based on diffusion models. Notably, ShaRP with a mismatched CS-MRI-specific prior consistently delivers the best performance.

Tasks	Metrics	PnP-FISTA	PnP-ADMM	DPS	DDS	ShaRP$_{\text{CS}}$	ShaRP$_{\text{SR}}$
4x Uniform	PSNR	35.88	35.76	32.62	35.21	**37.59**	35.91
	SSIM	0.938	0.941	0.888	0.937	**0.961**	0.943
4x Random	PSNR	29.31	28.83	31.72	32.41	**34.66**	30.91
	SSIM	0.863	0.842	0.874	0.910	**0.949**	0.905

Table 7: Quantitative comparison of ShaRP against baselines for CS-MRI reconstruction using $8\times$ CS-MRI and $4\times$ super-resolution priors, evaluated on the fastMRI dataset. Results are reported for both uniform and random undersampling masks at a 4x undersampling rate. The **best** and second best results are highlighted.

[5] https://github.com/HJ-harry/DDS

E.2 Ablation Study on the Influence of Hyperparameters

To investigate the impact of the hyperparameters α and b on ShaRP's performance, we conducted an ablation study under the $4\times$ CS-MRI setting with random sampling masks. The parameter α determines the selection of the specific restoration prior, while b controls the number of restoration priors used in the ensemble. Details about these hyperparameters can be found in Section B.1. Specifically, we analyzed how varying the values of α and b influenced reconstruction performance. This analysis provides valuable insights into ShaRP's sensitivity to these parameters and their roles in achieving optimal results.

As shown in Figure 12 and Figure 13, increasing the value of b, which corresponds to using more restoration priors in the ensemble, generally improves ShaRP's reconstruction performance. Similarly, Figure 14 demonstrates the influence of α on performance. A very small α fails to provide sufficient regularization to constrain the solution, while an excessively large α overly restricts the model, leading to a decline in performance. These findings highlight the importance of appropriately tuning α and b to balance flexibility and regularization for optimal results.

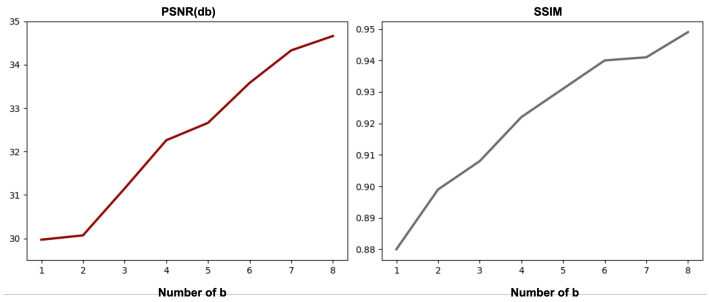

Figure 12: Performance comparison of ShaRP's CS-MRI reconstruction at $4\times$ acceleration with varying numbers of restoration priors, b. Left: PSNR vs. b; Right: SSIM vs. b. ShaRP with $b = 8$ consistently achieves superior results, highlighting the performance improvements gained by incorporating more restoration priors into ShaRP.

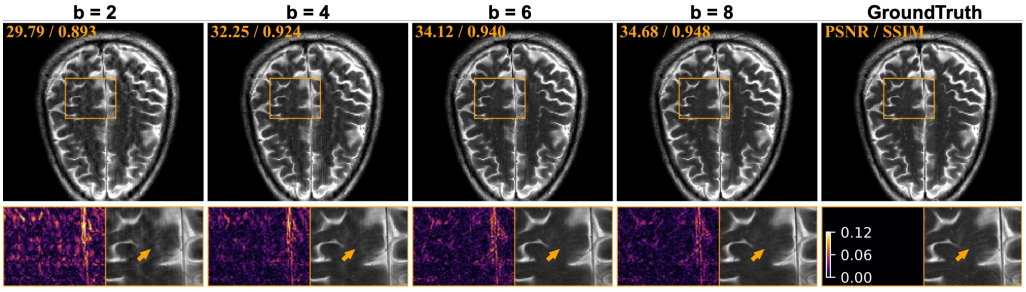

Figure 13: Visual comparison of ShaRP with varying amounts of restoration priors, denoted by b, in the ensemble. The PSNR and SSIM values for each method are shown in the top-left corner of each image. Zoomed-in regions, highlighted as squares at the bottom of each image, provide a closer look at key details. Notably, increasing the number of restoration priors in the ensemble enhances visual performance by effectively reducing artifacts and capturing finer details.

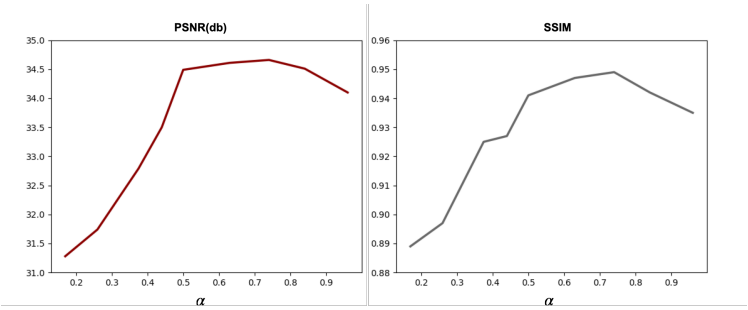

Figure 14: Performance comparison of ShaRP's CS-MRI reconstruction at $4\times$ acceleration with varying α. Left: PSNR vs. α; Right: SSIM vs. α.