A RESTORATION NETWORK AS AN IMPLICIT PRIOR

Yuyang Hu[1], **Mauricio Delbracio**[2], **Peyman Milanfar**[2], **Ulugbek S. Kamilov**[1,2]
[1]Washington University in St. Louis, [2]Google Research
{h.yuyang, kamilov}@wustl.edu,
{mdelbra,milanfar}@google.com

ABSTRACT

Image denoisers have been shown to be powerful priors for solving inverse problems in imaging. In this work, we introduce a generalization of these methods that allows any image restoration network to be used as an implicit prior. The proposed method uses priors specified by deep neural networks pre-trained as general restoration operators. The method provides a principled approach for adapting state-of-the-art restoration models for other inverse problems. Our theoretical result analyzes its convergence to a stationary point of a global functional associated with the restoration operator. Numerical results show that the method using a super-resolution prior achieves state-of-the-art performance both quantitatively and qualitatively. Overall, this work offers a step forward for solving inverse problems by enabling the use of powerful pre-trained restoration models as priors.

1 INTRODUCTION

Many problems in computational imaging, biomedical imaging, and computer vision can be formulated as *inverse problems*, where the goal is to recover a high-quality images from its low-quality observations. Imaging inverse problems are generally ill-posed, thus necessitating the use of prior models on the unknown images for accurate inference. While the literature on prior modeling of images is vast, current methods are primarily based on *deep learning (DL)*, where a deep model is trained to map observations to images (Lucas et al., 2018; McCann et al., 2017; Ongie et al., 2020).

Image denoisers have become popular for specifying image priors for solving inverse problems (Venkatakrishnan et al., 2013; Romano et al., 2017; Kadkhodaie & Simoncelli, 2021; Kamilov et al., 2023). Pre-trained denoisers provide a convenient proxy for image priors that does not require the description of the full density of natural images. The combination of state-of-the-art (SOTA) deep denoisers with measurement models has been shown to be effective in a number of inverse problems, including image super-resolution, deblurring, inpainting, microscopy, and medical imaging (Metzler et al., 2018; Zhang et al., 2017b; 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 (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., 2023).

Despite the rich literature on the topic, the prior work has narrowly focused on leveraging the statistical properties of denoisers. There is little work on extending the formalism and theory to priors specified using other types of image restoration operators, such as, for example, deep image super-resolution models. Such extensions would enable new algorithms that can leverage SOTA pre-trained restoration networks for solving other inverse problems. In this paper, we address this gap by developing the ***D**eep **R**estoration **P**riors (**DRP**)* methodology that provides a principled approach for using restoration operators as priors. We show that when the restoration operator is a *minimum mean-squared error (MMSE)* estimator, DRP can be interpreted as minimizing a composite objective function that includes log of the density of the degraded image as the regularizer. Our interpretation extends the recent formalism based on using MMSE denoisers as priors (Bigdeli et al., 2017; Xu et al., 2020; Kadkhodaie & Simoncelli, 2021; Laumont et al., 2022; Gan et al., 2023). We present a theoretical convergence analysis of DRP to a stationary point of the objective function

under a set of clearly specified assumptions. We show the practical relevance of DRP by solving several inverse problems by using a super-resolution network as a prior. Our numerical results show the potential of DRP to adapt the super-resolution model to act as an effective prior that can outperform image denoisers. This work thus addresses a gap in the current literature by providing a new principled framework for using pre-trained restoration models as priors for inverse problems.

All proofs and some details that have been omitted for space appear in the supplementary material.

2 Background

Inverse Problems. Many imaging problems can be formulated as inverse problems that seek to recover an unknown image $\boldsymbol{x} \in \mathbb{R}^n$ from from its corrupted observation

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

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 strategy for addressing inverse problems involves formulating them as an optimization problem

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

where g is the data-fidelity term that measures the fidelity to the observation $\boldsymbol{y}$ and h is the regularizer that incorporates prior knowledge on $\boldsymbol{x}$. For example, common functionals in imaging inverse problems are the least-squares data-fidelity term $g(\boldsymbol{x}) = \frac{1}{2}\left\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\right\|_2^2$ and the total variation (TV) regularizer $h(\boldsymbol{x}) = \tau \left\|\boldsymbol{D}\boldsymbol{x}\right\|_1$, where $\boldsymbol{D}$ is the image gradient, and $\tau > 0$ a regularization parameter.

Deep Learning. DL is extensively used for solving imaging inverse problems (McCann et al., 2017; Lucas et al., 2018; Ongie et al., 2020). Instead of explicitly defining a regularizer, DL methods often train convolutional neural networks (CNNs) to map the observations 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 (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., 2021; Liu et al., 2022). 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. PnP (Venkatakrishnan et al., 2013; Sreehari et al., 2016) is one of the most popular MBDL approaches for inverse problems based on using deep denoisers as imaging priors (see recent reviews (Ahmad et al., 2020; Kamilov et al., 2023)). For example, the proximal-gradient method variant of PnP can be written as (Hurault et al., 2022a)

$$\boldsymbol{x}^k \leftarrow \mathsf{prox}_{\gamma g}(\boldsymbol{z}^k) \quad \text{with} \quad \boldsymbol{z}^k \leftarrow \boldsymbol{x}^{k-1} - \gamma\tau(\boldsymbol{x}^{k-1} - \mathsf{D}_\sigma(\boldsymbol{x}^{k-1})), \tag{3}$$

where D_σ is a denoiser with a parameter $\sigma > 0$ for controlling its strength, $\tau > 0$ is a regularization parameter, and $\gamma > 0$ is a step-size. The theoretical convergence of PnP methods has been established for convex functions g using monotone operator theory (Sreehari et al., 2016; Sun et al., 2019; Ryu et al., 2019), as well as for nonconvex functions based on interpreting the denoiser as a MMSE estimator (Xu et al., 2020) or ensuring that the term $(\mathsf{I} - \mathsf{D}_\sigma)$ in (3) corresponds to a gradient ∇h of a function h parameterized by a deep neural network (Hurault et al., 2022a;b; Cohen et al., 2021a). Many variants of PnP have been developed over the past few years (Romano et al., 2017; Metzler et al., 2018; Zhang et al., 2017b; Meinhardt et al., 2017; Dong et al., 2019; Zhang et al., 2019; Wei et al., 2020), which has motivated an extensive research on its theoretical properties (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; Hurault et al., 2022a; Laumont et al., 2022; Hurault et al., 2022b; Gan et al., 2023).

This work is most related to two recent PnP-inspired methods using restoration operators instead of denoisers (Zhang et al., 2019; Liu et al., 2020). Deep plug-and-play super-resolution (DPSR) (Zhang et al., 2019) was proposed to perform image super-resolution under arbitrary blur kernels by using a bicubic super-resolver as a prior. Regularization by artifact removal (RARE) (Liu et al., 2020)

was proposed to use CNNs pre-trained directly on subsampled and noisy Fourier data as priors for magnetic resonance imaging (MRI). These prior methods did not leverage statistical interpretations of the restoration operators to provide a theoretical analysis for the corresponding PnP variants.

It is also worth highlighting the work of Gribonval and colleagues on theoretically exploring the relationship between MMSE restoration operators and proximal operators (Gribonval, 2011; Gribonval & Machart, 2013; Gribonval & Nikolova, 2021). Some of the observations and intuition in that prior line of work is useful for the theoretical analysis of the proposed DRP methodology.

Our contribution. ***(1)*** Our first contribution is the new method DRP for solving inverse problems using the prior implicit in a pre-trained deep restoration network. Our method is a major extension of recent methods (Bigdeli et al., 2017; Xu et al., 2020; Kadkhodaie & Simoncelli, 2021; Gan et al., 2023) from denoisers to more general restoration operators.***(2)*** Our second contribution is a new theory that characterizes the solution and convergence of DRP under priors associated with the MMSE restoration operators. Our theory is general in the sense that it allows for nonsmooth data-fidelity terms and expansive restoration models. ***(3)*** Our third contribution is the implementation of DRP using the popular SwinIR (Liang et al., 2021) super-resolution model as a prior for two distinct inverse problems, namely deblurring and super-resolution. We publicly share our implementation that shows the potential of using restoration models to achieve SOTA performance.

3 Deep Restoration Prior

Image denoisers are currently extensively used as priors for solving inverse problems. We extend this approach by proposing the following method that uses a more general restoration operator.

Algorithm 1 Deep Restoration Priors (DRP)

1: **input:** Initial value $\boldsymbol{x}^0 \in \mathbb{R}^n$ and parameters $\gamma, \tau > 0$
2: **for** $k = 1, 2, 3, \ldots$ **do**
3: $\quad \boldsymbol{z}^k \leftarrow \boldsymbol{x}^{k-1} - \gamma\tau \mathsf{G}(\boldsymbol{x}^{k-1})$ where $\mathsf{G}(\boldsymbol{x}) := \boldsymbol{x} - \mathsf{R}(\mathbf{H}\boldsymbol{x})$
4: $\quad \boldsymbol{x}^k \leftarrow \mathsf{sprox}_{\gamma g}(\boldsymbol{z}^k)$
5: **end for**

The prior in Algorithm 1 is implemented in Line 3 using a deep model $\mathsf{R} : \mathbb{R}^p \to \mathbb{R}^n$ pre-trained to solve the following restoration problem

$$\boldsymbol{s} = \mathbf{H}\boldsymbol{x} + \boldsymbol{n} \quad \text{with} \quad \boldsymbol{x} \sim p_{\boldsymbol{x}}, \quad \boldsymbol{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}), \tag{4}$$

where $\mathbf{H} \in \mathbb{R}^{p \times n}$ is a degradation operator, such as blur or downscaling, and $\boldsymbol{n} \in \mathbb{R}^p$ is the *additive white Gaussian noise (AWGN)* of variance σ^2. The density $p_{\boldsymbol{x}}$ is the prior distribution of the desired class of images. Note that the restoration problem (4) is only used for training R and doesn't have to correspond to the inverse problem in (1) we are seeking to solve. When $\mathbf{H} = \mathbf{I}$, the restoration operator R reduces to an AWGN denoiser used in the traditional PnP methods (Romano et al., 2017; Kadkhodaie & Simoncelli, 2021; Hurault et al., 2022a). The goal of DRP is to leverage a pre-trained restoration network R to gain access to the prior.

The measurement consistency is implemented in Line 4 using the *scaled* proximal operator

$$\mathsf{sprox}_{\gamma g}(\boldsymbol{z}) := \mathsf{prox}_{\gamma g}^{\mathbf{H}^\top \mathbf{H}}(\boldsymbol{z}) = \underset{\boldsymbol{x} \in \mathbb{R}^n}{\arg\min} \left\{ \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{z}\|_{\mathbf{H}^\top \mathbf{H}}^2 + \gamma g(\boldsymbol{x}) \right\}, \tag{5}$$

where $\|\boldsymbol{v}\|_{\mathbf{H}^\top \mathbf{H}} := \boldsymbol{v}^\top \mathbf{H}^\top \mathbf{H} \boldsymbol{v}$ denotes the weighted Euclidean seminorm of a vector $\boldsymbol{v}$. When $\mathbf{H}^\top \mathbf{H}$ is positive definite and g is convex, the functional being minimized in (5) is strictly convex, which directly implies that the solution is unique. On the other hand, when g is not convex or $\mathbf{H}^\top \mathbf{H}$ is positive semidefinite, there might be multiple solutions and the scaled proximal operator simply returns one of the solutions. It is also worth noting that (5) has an efficient solution when g is the least-squares data-fidelity term (see for example the discussion in (Kamilov et al., 2023) on efficient implementations of proximal operators of least-squares).

The fixed points of Algorithm 1 can be characterized for subdifferentiable g (see Chapter 3 in (Beck, 2017) for a discussion on subdifferentiability). When DRP converges, it converges to vectors $\boldsymbol{x}^* \in$

$\mathbb{R}^n$ that satisfy (see formal analysis in Supplement A.1)

$$\mathbf{0} \in \partial g(\boldsymbol{x}^*) + \tau \mathbf{H}^\mathsf{T}\mathbf{H}\mathsf{G}(\boldsymbol{x}^*) \tag{6}$$

where ∂g is the subdifferential of g and G is defined in Line 3 of Algorithm 1. As discussed in the next section, under additional assumptions, one can associate the fixed points of DRP with the stationary points of a composite objective function $f = g + h$ for some regularizer h.

4 Convergence Analysis of DRP

In this section, we present a theoretical analysis of DRP. We first provide a more insightful interpretation of its solutions for restoration models that compute MMSE estimators of (4). We then discuss the convergence of the iterates generated by DRP. Our analysis will require several assumptions.

We will consider restoration models that perform MMSE estimation of $\boldsymbol{x} \in \mathbb{R}^n$ for the problem (4)

$$\mathsf{R}(\boldsymbol{s}) = \mathbb{E}\left[\boldsymbol{x}|\boldsymbol{s}\right] = \int \boldsymbol{x} p_{\boldsymbol{x}|\boldsymbol{s}}(\boldsymbol{x};\boldsymbol{s})\,\mathsf{d}\boldsymbol{x} = \int \boldsymbol{x} \frac{p_{\boldsymbol{s}|\boldsymbol{x}}(\boldsymbol{s};\boldsymbol{x})p_{\boldsymbol{x}}(\boldsymbol{x})}{p_{\boldsymbol{s}}(\boldsymbol{s})}\,\mathsf{d}\boldsymbol{x}. \tag{7}$$

where we used the probability density of the observation $\boldsymbol{s} \in \mathbb{R}^p$

$$p_{\boldsymbol{s}}(\boldsymbol{s}) = \int p_{\boldsymbol{s}|\boldsymbol{x}}(\boldsymbol{s};\boldsymbol{x})p_{\boldsymbol{x}}(\boldsymbol{x})\,\mathsf{d}\boldsymbol{x} = \int G_\sigma(\boldsymbol{s} - \mathbf{H}\boldsymbol{x})p_{\boldsymbol{x}}(\boldsymbol{x})\,\mathsf{d}\boldsymbol{x}. \tag{8}$$

The function G_σ in (8) denotes the Gaussian density function with the standard deviation $\sigma > 0$.

Assumption 1. *The prior density $p_{\boldsymbol{x}}$ is non-degenerate over $\mathbb{R}^n$.*

For example, a probability density $p_{\boldsymbol{x}}$ is degenerate over $\mathbb{R}^n$, if it is supported on a space of lower dimensions than n. Our goal is to establish an explicit link between the MMSE restoration operator (7) and the following regularizer

$$h(\boldsymbol{x}) = -\tau\sigma^2 \log p_{\boldsymbol{s}}(\mathbf{H}\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^n, \tag{9}$$

where τ is the parameter in Algorithm 1, $p_{\boldsymbol{s}}$ is the density of the observation (8), and σ^2 is the AWGN level used for training the restoration network. We adopt Assumption 1 to have a more intuitive mathematical exposition, but one can in principle generalize the link between MMSE operators and regularization beyond non-degenerate priors (Gribonval & Machart, 2013). It is also worth observing that the function h is infinitely continuously differentiable, since it is obtained by integrating $p_{\boldsymbol{x}}$ with a Gaussian density G_σ (Gribonval, 2011; Gribonval & Machart, 2013).

Assumption 2. *The scaled proximal operator $\mathsf{sprox}_{\gamma g}$ is well-defined in the sense that there exists a solution to the problem* (5) *for any $\boldsymbol{z} \in \mathbb{R}^n$. The function g is subdifferentiable over $\mathbb{R}^n$.*

This mild assumption is needed for us to be able to run our method. There are multiple ways to ensure that the scaled proximal operator is well defined. For example, $\mathsf{sprox}_{\gamma g}$ is always well-defined for any g that is proper, closed, and convex (Parikh & Boyd, 2014). This directly makes DRP applicable with the popular least-squares data-fidelity term $g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$. One can relax the assumption of convexity by considering g that is proper, closed, and coercive, in which case $\mathsf{sprox}_{\gamma g}$ will have a solution (see for example Chapter 6 of (Beck, 2017)). Note that we do not require the solution to (5) to be unique; it is sufficient for $\mathsf{sprox}_{\gamma g}$ to return one of the solutions.

We are now ready to theoretically characterize the solutions of DRP.

Theorem 1. *Let R be the MMSE restoration operator* (7) *corresponding to the restoration problem* (4) *under Assumptions 1-3. Then, any fixed-point $\boldsymbol{x}^* \in \mathbb{R}^n$ of DRP satisfies*

$$\mathbf{0} \in \partial g(\boldsymbol{x}^*) + \nabla h(\boldsymbol{x}^*),$$

where h is given in (9).

The proof of the theorem is provided in the supplement and generalizes the well-known *Tweedie's formula* (Robbins, 1956; Miyasawa, 1961; Gribonval, 2011) to restoration operators. The theorem implies that the solutions of DRP satisfy the first-order conditions for the objective function $f = g + h$. If g is a negative log-likelihood $p_{\boldsymbol{y}|\boldsymbol{x}}$, then the fixed-points of DRP can be interpreted as *maximum-a-posteriori probability (MAP)* solutions corresponding to the prior density $p_{\boldsymbol{s}}$. The density $p_{\boldsymbol{s}}$ is

related to the true prior $p_{\boldsymbol{x}}$ through eq. (8), which implies that DRP has access to the prior $p_{\boldsymbol{x}}$ through the restoration operator R via density $p_{\boldsymbol{s}}$. As $\mathbf{H} \to \mathbf{I}$ and $\sigma \to 0$, the density $p_{\boldsymbol{s}}$ approaches the prior distribution $p_{\boldsymbol{x}}$.

The convergence analysis of DRP will require additional assumptions.

Assumption 3. *The data-fidelity term g and the implicit regularizer h are bounded from below.*

This assumption implies that there exists $f^* > -\infty$ such that $f(\boldsymbol{x}) \geq f^*$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

Assumption 4. *The function h has a Lipschitz continutous gradient with constant $L > 0$. The degradation operator associated with the restoration network is such that $\lambda \succeq \mathbf{H}^\mathsf{T}\mathbf{H} \succeq \mu > 0$.*

This assumption is related to the implicit prior associated with a restoration model and is needed to ensure the monotonic reduction of the objective f by the DRP iterates. As stated under eq. (9), the function h is infinitely continuously differentiable. We additionally adopt the standard optimization assumption that ∇h is Lipschitz continuous (Nesterov, 2004). It is also worth noting that the positive definiteness of $\mathbf{H}^\mathsf{T}\mathbf{H}$ in Assumption 4 is a relaxation of the traditional PnP assumption that the prior is a denoiser, which makes our theoretical analysis a significant extension of the prior work (Bigdeli et al., 2017; Xu et al., 2020; Kadkhodaie & Simoncelli, 2021; Gan et al., 2023).

We are now ready to state the following results.

Theorem 2. *Run DRP for for $t \geq 1$ iterations under Assumptions 1-4 using a step-size $\gamma = \mu/(\alpha L)$ with $\alpha > 1$. Then, for each iteration $1 \leq k \leq t$, there exists $\boldsymbol{w}(\boldsymbol{x}^k) \in \partial f(\boldsymbol{x}^k)$ such that*

$$\frac{1}{t}\sum_{k=1}^{t} \|\boldsymbol{w}(\boldsymbol{x}^k)\|_2^2 \leq \frac{C(f(\boldsymbol{x}^0) - f^*)}{t},$$

where $C > 0$ is an iteration independent constant.

The exact expression for the constant C is given in the proof. Theorem 2 shows that the iterates generated by DRP satisfy $\boldsymbol{w}(\boldsymbol{x}^k) \to \mathbf{0}$ as $t \to \infty$. Theorems 1 and 2 do not explicitly require convexity or smoothness of g, and non-expansiveness of R. They can thus be viewed as a major generalization of the existing theory from denoisers to more general restoration operators.

5 Numerical Results

We now numerically validate DRP on several distinct inverse problems. Due to space limitations in the main paper, we have included several additional numerical results in the supplementary material.

We consider two inverse problems of form $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$: (a) *Image Deblurring* and (b) *Single Image Super Resolution (SISR)*. For both problems, we assume that $\boldsymbol{e}$ is the additive white Gaussian noise (AWGN). We adopt the traditional ℓ_2-norm loss as the data-fidelity term in (2) for both problems. We use the Peak Signal-to-Noise Ratio (PSNR) for quantitative performance evaluation.

In the main manuscript, we compare DRP with several variants of denoiser-based methods, including SD-RED (Romano et al., 2017), PnP-ADMM (Chan et al., 2017), IRCNN (Zhang et al., 2017b), and DPIR (Zhang et al., 2022). SD-RED and PnP-ADMM refer to the steepest-descent variant of RED and the ADMM variant of PnP, both of which incorporate AWGN denoisers based on DnCNN (Zhang et al., 2017a). IRCNN and DPIR are based on half-quadratic splitting (HQS) iterations that use the IRCNN and the DRUNet denoisers, respectively.

In the supplement, we present several additional comparisons, namely: (a) evaluation of the performance of DRP on the task of image denoising; (b) additional comparison of DRP with the recent provably convergent variant of PnP called gradient-step plug-and-play (GS-PnP) (Hurault et al., 2022a); (c) comparison of DRP with the diffusion posterior sampling (DPS) (Chung et al., 2023) method that uses a denoising diffusion model as a prior; (d) illustration of the improvement of DRP using SwinIR as a prior over the direct application of SwinIR on SR using the Gaussian kernel; (e) presentation of quantitative results using SSIM and LPIPS metrics; (f) additional comparison with DPIR using with SwinIR trained as denoiser; (g) additional evaluation of DRP with a non-MMSE restoration priors and other MMSE restoration priors; and (h) evaluation of robustness of our numerical results to random seed.

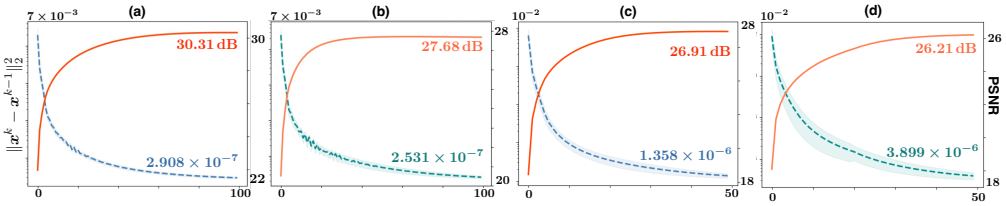

Figure 1: Illustration of the convergence behaviour of DRP for image deblurring and single image super resolution on the Set3c dataset. ***(a)-(b)***: Deblurring with Gaussian blur kernels of standard deviations 1.6 and 2.0. ***(c)-(d)***: $2\times$ and $3\times$ super resolution with the Gaussian blur kernel of standard deviation 2.0. Average distance $\|\boldsymbol{x}^k - \boldsymbol{x}^{k-1}\|_2^2$ and PSNR relative to the groundtruth are plotted, with shaded areas indicating the standard deviation of these metrics across all test images.

Kernel	Datasets	SD-RED	PnP-ADMM	IRCNN+	DPIR	DRP
	Set3c	27.14/0.101	29.11/0.084	28.14/0.084	29.53/0.093	**30.69/0.047**
	Set5	29.78/0.109	32.31/0.095	29.46/0.097	32.38/0.106	**32.79/0.075**
	CBSD68	25.78/0.254	28.90/0.229	26.86/0.240	28.86/0.246	**29.10/0.198**
	McMaster	29.69/0.119	32.20/0.098	29.15/0.111	32.42/0.110	**32.79/0.075**
	Set3c	25.83/0.140	27.05/0.137	26.58/0.138	27.52/0.137	**27.89/0.075**
	Set5	28.13/0.161	30.77/0.131	28.75/0.152	30.94/0.137	**31.04/0.102**
	CBSD68	24.43/0.371	27.45/0.293	25.97/0.323	**27.52**/0.307	27.46/**0.245**
	McMaster	28.71/0.162	30.50/0.140	28.27/0.159	30.78/0.146	**30.79/0.105**

Table 1: PSNR↑/LPIPS↓ comparison of DRP and several SOTA methods for solving inverse problems using denoisers on image deblurring with the Gaussian blur kernels of standard deviation 1.6 and 2.0 on Set3c, Set5, CBSD68 and McMaster datasets. The **best** and second best results are highlighted. Note how DRP can outperform SOTA PnP methods that use denoisers as priors.

5.1 Swin Transformer based Super Resolution Prior

Super Resolution Network Architecture. We pre-trained a $q\times$ super resolution model R_q using the SwinIR (Liang et al., 2021) architecture based on Swin Transformer. Our training dataset comprised both the DIV2K (Agustsson & Timofte, 2017) and Flick2K (Lim et al., 2017) dataset, containing 3450 color images in total. During training, we applied $q\times$ bicubic downsampling to the input images with AWGN characterized by standard deviation σ randomly chosen in [0, 10/255]. We used three SwinIR SR models, each trained for different down-sampling factors: $2\times$, $3\times$ and $4\times$.

Prior Refinement Strategy for the Super Resolution prior. Theorem 1 suggests that as $\mathbf{H} \to \mathbf{I}$, the prior in DRP converges to $p_{\boldsymbol{x}}$. This process can be approximated for SwinIR by controlling the down-sampling factor q of the SR restoration prior $\mathsf{R}_q(\cdot)$. We observed through our numerical experiments that gradual reduction of q leads to less reconstruction artifacts and enhanced fine details. We will denote the approach of gradually reducing q as *prior refinement strategy*. We initially set q to a larger down-sampling factor, which acts as a more aggressive prior; we then reduce q to a smaller value leading to preservation of finer details. This strategy is conceptually analogous to the gradual reduction of σ in the denoiser in the SOTA PnP methods such as DPIR (Zhang et al., 2022).

5.2 Image Deblurring

Image deblurring is based on the degradation operator of the form $\boldsymbol{A} = \boldsymbol{K}$, where $\boldsymbol{K}$ is a convolution with the blur kernel $\boldsymbol{k}$. We consider image deblurring using two 25×25 Gaussian kernels (with the standard deviations 1.6 and 2) used in (Zhang et al., 2019), and the AWGN vector $\boldsymbol{e}$ corresponding to noise level of 2.55/255. For fair comparison, we use the official implementations provided by each baseline method and use the same random seed to ensure consistency of random noise for all methods. The restoration model used as a prior in DRP is SwinIR introduced in Section 5.1, so that the operation $\mathbf{H}$ corresponds to the standard bicubic downsampling. The scaled proximal operator

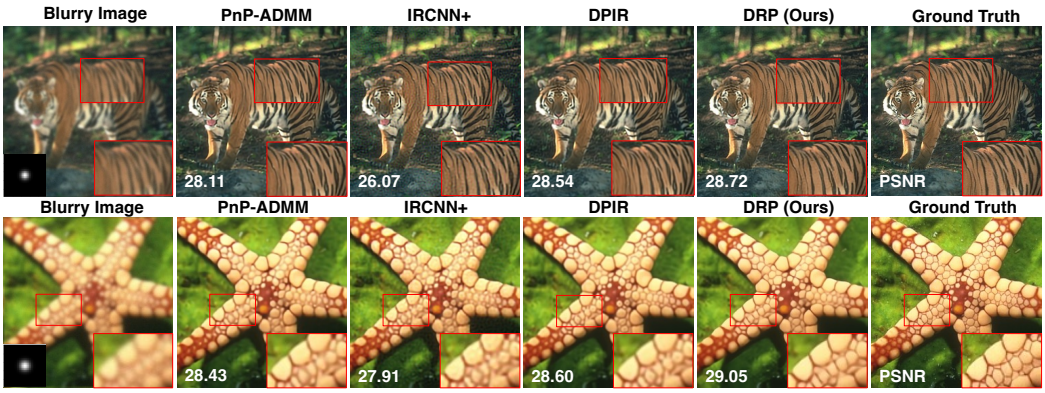

Figure 2: Visual comparison of DRP with several well-known methods on Gaussian deblurring of color images. The top row shows results for a blur kernel with a standard deviation (std) of 1.6, while the bottom row shows results for another blur kernel with std = 2. The squares at the bottom-left corner of blurry images show the blur kernels. Each image is labeled by its PSNR in dB with respect to the original image. The visual differences are highlighted in the bottom-right corner. Note how DRP using restoration prior improves over SOTA methods based on denoiser priors.

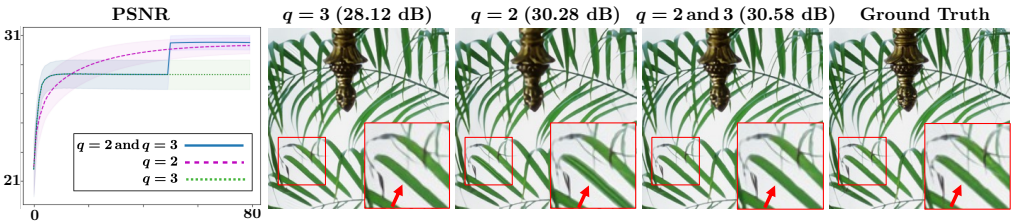

Figure 3: Illustration of the impact of different SR factors in the prior used within DRP for image deblurring. We show three scenarios: (i) using only $3\times$ prior, (ii) using only $2\times$ prior, and (iii) the use of the *prior refinement strategy*, which combines both the $2\times$ and $3\times$ priors. ***Left***: Convergence of PSNR against the iteration number for all three configurations. ***Right***: Visual illustration of the final image for each setting. The visual difference is highlighted by the red arrow in the zoom-in box. Note how the reduction of q can lead to about 0.3 dB improvement in the final performance.

$\mathsf{sprox}_{\lambda g}$ in (5) with data-fidelity term $g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2$ can be written as

$$\mathsf{sprox}_{\gamma g}(\boldsymbol{z}) = (\boldsymbol{K}^\top \boldsymbol{K} + \gamma \mathbf{H}^\top \mathbf{H})^{-1}[\boldsymbol{K}^\top \boldsymbol{y} + \gamma \mathbf{H}^\top \mathbf{H}\boldsymbol{z}]. \tag{10}$$

We adopt a standard approach of using a few iterations of the conjugate gradient (CG) method (see for example (Aggarwal et al., 2019)) to implement the scaled proximal operator (10) by avoiding the direct inversion of $(\boldsymbol{K}^\top \boldsymbol{K} + \gamma \mathbf{H}^\top \mathbf{H})$. As bicubic SR model is adopted as prior, the $\mathsf{R}(\mathbf{H}\boldsymbol{x})$ in Step 3 of Algorithm 1 performs a bicubic downsampling of the intermediate image $\boldsymbol{x}^{k-1}$ and inputs it into the bicubic SR SwinIR. In each DRP iteration, we run three steps of a CG solver, starting from a warm initialization from the previous DRP iteration. We fine-turned the hyper-parameter γ, τ and SR restoration prior rate q to achieve the highest PSNR value on the Set5 dataset and then apply the same configuration to the other three datasets.

Figure 1 (a)-(b) illustrates the convergence behaviour of DRP on the Set3c dataset for two blur kernels. Table 1 presents the quantitative evaluation of the reconstruction performance on two different blur kernels, showing that DRP outperforms the baseline methods across four widely-used datasets. Figure 2 visually illustrates the reconstructed results on the same two blur kernels. Note how DRP can reconstruct the fine details of the tiger and starfish, as highlighted within the zoom-in boxes, while all the other baseline methods yield either oversmoothed reconstructions or noticeable artifacts. These results show that DRP can leverage SwinIR as an implicit prior, which not only ensures stable convergence, but also leads to competitive performance when compared to denoisers priors.

SR	Kernel	Datasets	SD-RED	PnP-ADMM	IRCNN+	DPIR	DRP
2×		Set3c	27.01/0.119	27.88/0.114	27.48/0.119	28.18/0.117	**29.26/0.066**
		Set5	28.98/0.133	31.41/0.119	29.47/0.144	31.42/0.124	**31.47/0.083**
		CBSD68	26.11/0.365	28.00/0.274	26.66/0.336	27.97/0.286	**28.12/0.230**
		McMaster	28.70/0.197	30.98/0.129	29.11/0.183	31.16/0.134	**31.39/0.089**
		Set3c	25.20/0.170	25.86/0.163	25.92/0.149	26.80/0.149	**27.41/0.093**
		Set5	28.57/0.188	30.06/0.152	28.91/0.178	30.36/0.153	**30.42/0.118**
		CBSD68	25.77/0.439	26.88/0.326	26.06/0.411	26.98/0.330	**26.98/0.273**
		McMaster	28.15/0.224	29.53/0.170	28.41/0.229	29.87/0.167	**30.03/0.127**
3×		Set3c	25.50/0.156	25.85/0.151	25.72/0.155	26.64/0.147	**27.77/0.077**
		Set5	28.75/0.159	30.09/0.144	29.14/0.157	30.39/0.149	**30.83/0.107**
		CBSD68	25.69/0.407	26.78/0.312	26.01/0.372	26.80/0.325	**27.18/0.248**
		McMaster	28.38/0.203	29.52/0.160	28.53/0.199	29.82/0.162	**29.92/0.126**
		Set3c	24.55/0.187	24.87/0.192	24.87/0.178	25.84/0.171	**26.84/0.093**
		Set5	28.19/0.193	29.26/0.172	28.37/0.197	29.70/0.169	**29.88/0.118**
		CBSD68	25.40/0.450	26.28/0.353	25.56/0.432	26.39/0.352	**26.60/0.270**
		McMaster	27.79/0.255	28.72/0.196	27.85/0.248	29.11/0.187	**29.47/0.127**

Table 2: PSNR↑/LPIPS↓ comparison of DRP and several baselines for SISR on Set3c, Set5, CBSD68 and McMaster datasets. The **best** and second best results are highlighted. Note the excellent quantitative performance of DRP, which suggests the potential of using general restoration models as priors.

Figure 3 illustrates the impact of the *prior-refinement strategy* described in Section 5.1. We compare three settings: (i) use of only 3× prior, (ii) use of only 2× prior, and (iii) use of the prior-refinement strategy to leverage both 3× and 2× priors. The subfigure on the left shows the convergence of DRP for each configuration, while the ones on the right show the final imaging quality. Note how the reduction of q leads to better performance, which is analogous to what was observed with the reduction of σ in the SOTA PnP methods (Zhang et al., 2022).

5.3 Single Image Super Resolution

We apply DRP using the bicubic SwinIR prior to Single Image Super Resolution (SISR) task. The measurement operator in SISR can be written as $\boldsymbol{A} = \boldsymbol{S}\boldsymbol{K}$, where $\boldsymbol{K}$ is convolution with the blur kernel $\boldsymbol{k}$ and $\boldsymbol{S}$ performs standard d-fold down-sampling with $d^2 = n/m$. The scaled proximal operator $\mathsf{sprox}_{\lambda g}$ in (5) with data-fidelity term $g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{K}\boldsymbol{x}\|_2^2$ can be write as:

$$\mathsf{sprox}_{\gamma g}(\boldsymbol{z}) = (\boldsymbol{K}^\mathsf{T}\boldsymbol{S}^\mathsf{T}\boldsymbol{S}\boldsymbol{K} + \gamma\mathbf{H}^\mathsf{T}\mathbf{H})^{-1}[\boldsymbol{K}^\mathsf{T}\boldsymbol{S}^\mathsf{T}\boldsymbol{y} + \gamma\mathbf{H}^\mathsf{T}\mathbf{H}\boldsymbol{z}], \tag{11}$$

where $\mathbf{H}$ is the bicubic downsampling operator. Similarly to deblurring in Section 5.2, we use CG to efficiently compute (11). We adjust the hyper-parameter γ, τ, and the SR restoration prior factor q for the best PSNR performance on Set5, and then use these parameters on the remaining datasets.

We evaluate super-resolution performance across two 25×25 Gaussian blur kernels, each with distinct standard deviations (1.6 and 2.0), and for two distinct downsampling factors (2× and 3×), incorporating an AWGN vector $\boldsymbol{e}$ corresponding to noise level of 2.55/255. For fair comparison, we use the official implementations provided by each baseline method and use the same random seed to ensure consistency of random noise for all methods.

Figure 1 (c)-(d) illustrates the convergence behaviour of DRP on the Set3c dataset for 2× and 3× SISR. Figure 4 shows the visual reconstruction results for the same downsampling factors. Table 2 summarizes the PSNR and LPIPS values achieved by DRP relative to other baseline methods when applied to different blur kernel and downsampling factors on four commonly used datasets.

It is worth highlighting that the SwinIR model used in DRP was pre-trained for the bicubic super-resolution task. Consequently, the direct application of the pre-trained SwinIR to the setting considered in this section leads to the suboptimal performance due to mismatch between the kernels used. See Supplement B.4 to see how DRP improves over the direct application of SwinIR.

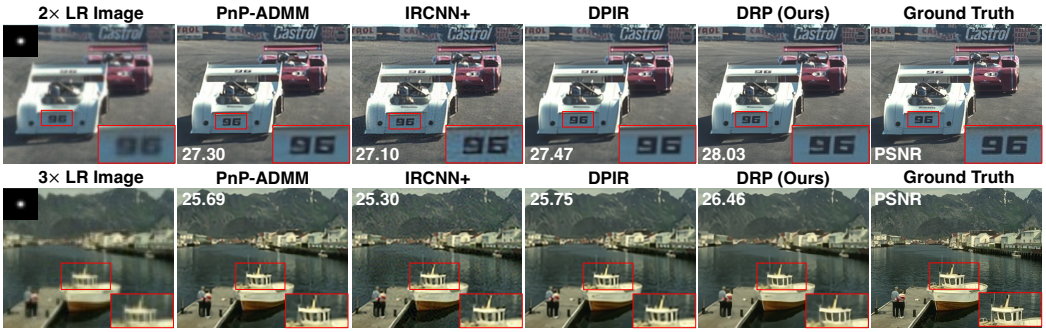

Figure 4: Visual comparison of DRP and several well-known methods on single image super resolution. The top row displays performances for $2\times$ SR, while the bottom row showcases results for $3\times$ SR. The lower-left corner of each low-resolution image shows the blur kernels. Each image is labeled by its PSNR in dB with respect to the original image. The visual differences are highlighted by the boxes in the bottom-right corner. Note the excellent performance of the proposed DRP method using the SwinIR prior both visually and in terms of PSNR.

6 Conclusion

The work presented in this paper proposes a new DRP method for solving imaging inverse problems by using pre-trained restoration operators as priors, presents its theoretical analysis in terms of convergence, and applies the method to two well-known inverse problems. The proposed method and its theoretical analysis extend the recent work using denoisers as priors by considering more general restoration operators. The numerical validation of DRP shows the improvements due to the use of learned SOTA super-resolution models. One conclusion of this work is the potential effectiveness of going beyond priors specified by traditional denoisers.

Limitations and Future Work

The work presented in this paper comes with several limitations. The proposed DRP method uses pre-trained restoration models as priors, which means that its performance is inherently limited by the quality of the pre-trained model. As shown in this paper, pre-trained restoration models provide a convenient, principled, and flexible mechanism to specify priors; yet, they are inherently self-supervised and their empirical performance can thus be suboptimal compared to priors trained in a supervised fashion for a specific inverse problem. Our theory is based on the assumption that the restoration prior used for inference performs MMSE estimation. While this assumption is reasonable for deep networks trained using the MSE loss, it is not directly applicable to denoisers trained using other common loss functions, such as the ℓ_1-norm or SSIM. Finally, as is common with most theoretical work, our theoretical conclusions only hold when our assumptions are satisfied, which might limit their applicability in certain settings.

Our work opens several interesting directions for future research. First, the implicit prior in 9 can be seen as an analysis prior with a transform $\mathbf{H}$ (Elad et al., 2007; Selesnick & Figueiredo, 2009), which suggests a possibility of considering broader class of linear transforms for priors. Second, the excellent performance of the restoration priors beyond denoisers lead to an interesting open question: what is the optimal linear transform $\mathbf{H}$ for a given measurement operator $\boldsymbol{A}$. Third, while in this paper we considered non-blind inverse problems, the extension of DRP to blind inverse problems would be an interesting future direction of research (Gan et al., 2023).

Reproducibility Statement

We have provided the anonymous source code in the supplementary materials. The included README.md file contains detailed instructions on how to run the code and reproduce the results reported in the paper. The pseudo-code of DRP is outlined in Algorithm 1. The complete proofs and technical details for our theoretical analysis can be found in the supplement.

ETHICS STATEMENT

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

REFERENCES

H. K. Aggarwal, M. P. Mani, and M. Jacob. Modl: Model-based deep learning architecture for inverse problems. *IEEE Trans. Med. Imag.*, 38(2):394–405, Feb. 2019. ISSN 1558-254X. doi: 10.1109/TMI.2018.2865356.

E. Agustsson and R. Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR) workshops*, July 2017.

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.

H. Attouch, J. Bolte, and B. F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Math. Program. Ser. A*, 137(1):91–129, 2013.

A. Beck. *First-Order Methods in Optimization*. MOS-SIAM Series on Optimization. SIAM, 2017.

A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009.

S. A. Bigdeli, M. Zwicker, P. Favaro, and M. Jin. Deep mean-shift priors for image restoration. *Proc. Adv. Neural Inf. Process. Syst.*, 30, 2017.

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

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. URL https://openreview.net/forum?id=OnD9zGAGT0k.

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.

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. URL https://openreview.net/forum?id=VmyFF5lL3F. Featured Certification.

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

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.

M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3):947–968, 2007.

W. Gan, S. Shoushtari, Y. Hu, J. Liu, H. An, and U. S. Kamilov. Block coordinate plug-and-play methodsfor blind inverse problems. In *Proc. Adv. Neural Inf. Process. Syst. 36*, 2023.

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

R. Gribonval. Should penalized least squares regression be interpreted as maximum a posteriori estimation? *IEEE Trans. Signal Process.*, 59(5):2405–2410, May 2011.

R. Gribonval and P. Machart. Reconciling "priors" & "priors" without prejudice? In *Proc. Adv. Neural Inf. Process. Syst. 26*, pp. 2193–2201, Lake Tahoe, NV, USA, December 5-10, 2013.

R. Gribonval and M. Nikolova. On Bayesian estimation and proximity operators. *Appl. Comput. Harmon. Anal.*, 50:49–72, January 2021.

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.

S. Hurault, A. Leclaire, and N. Papadakis. Gradient step denoiser for convergent plug-and-play. In *Int. Conf. on Learn. Represent.*, 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. doi: 10.1109/TIP.2017.2713099.

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, 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.

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.

B. Lim, S. Son, H. Kim, S. Nah, and K. Mu L. Enhanced deep residual networks for single image super-resolution. In *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR) workshops*, pp. 136–144, 2017.

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.*, New Orleans, LA, 2022.

A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos. Using deep neural networks for inverse problems in imaging: Beyond analytical methods. *IEEE Signal Process. Mag.*, 35(1):20–36, January 2018.

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, Stockholmsmässan, Stockholm Sweden, Jul. 10–15 2018.

K. Miyasawa. An empirical bayes estimator of the mean of a normal population. *Bull. Inst. Internat. Statist*, 38:1–2, 1961.

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.

N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):123–231, 2014.

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

H. Robbins. An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. on Math. Statist. and Prob., Vol. 1 (Univ. of Calif. Press, 1956)*, 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.

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. 36th Int. Conf. Mach. Learn.*, volume 97, pp. 5546–5557, Long Beach, CA, USA, Jun. 09–15 2019.

I. W. Selesnick and M. A. T. Figueiredo. Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors. In *Proc. SPIE 7446, 74460D*, 2009. doi: doi: 10.1117/12.826663.

S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, G. T. Buzzard, L. F. Drummy, J. P. Simmons, and C. A. Bouman. Plug-and-play priors for bright field electron tomography and sparse interpolation. *IEEE Trans. Comput. Imaging*, 2(4):408–423, Dec. 2016.

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. doi: 10.1109/ICASSP.2019.8683057.

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.

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.

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

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, Austin, TX, USA, 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*, 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. *arXiv preprint arXiv:2212.00490*, 2022.

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 *Proc. 37th Int. Conf. Mach. Learn.*, 2020.

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.

Z. Yue, Q. Zhao, J. Xie, L. Zhang, D. Meng, and K. K Wong. Blind image super-resolution with elaborate degradation modeling on noise and kernel. In *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR)*, pp. 2128–2138, 2022.

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

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

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

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.*, pp. 1671–1681, Long Beach, CA, USA, June 16-20, 2019.

K. Zhang, L. Van G., and R. Timofte. Deep unfolding network for image super-resolution. In *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR)*, pp. 3217–3226, 2020.

K. Zhang, J. Liang, L. Van G., and R. Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proc. IEEE Int. Conf. Comp. Vis. (ICCV)*, pp. 4791–4800, 2021.

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.

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.

This supplement to the main paper presents the details of our mathematical analysis and provides additional numerical results. The mathematical analysis builds on two distinct lines of work: (a) relationship between the MMSE denoising and the probability density of the noisy image (Gribonval, 2011; Gribonval & Machart, 2013; Bigdeli et al., 2017; Kadkhodaie & Simoncelli, 2021); (b) analysis of the proximal-gradient method (Beck & Teboulle, 2009; Attouch et al., 2013; Parikh & Boyd, 2014; Hurault et al., 2022a). Our results can be viewed as an extension of this line of research to priors specified using more general restoration operators.

The structure of this supplementary document is as follows. In Section A.1, we provide formal characterization of the fixed-points of DRP. In Section A.2, we prove the convergence of DRP under additional assumptions on g and R. In Section B, we provide a number of interesting numerical results that didn't make it to the main paper due to space. In particular, we evaluate the AWGN denoising performance of DRP, compare it with another recent PnP method called GS-PnP, and show how DRP enables adaptation of SwinIR beyond problems it was trained on.

A Theoretical Analysis of DRP

A.1 Proof of Theorem 1

Theorem. *Let* R *be the MMSE restoration operator* (7) *corresponding to the restoration problem* (4) *under Assumptions 1-3. Then, any fixed-point* $\boldsymbol{x}^* \in \mathbb{R}^n$ *of DRP satisfies*

$$\mathbf{0} \in \partial g(\boldsymbol{x}^*) + \nabla h(\boldsymbol{x}^*),$$

where h *is given in* (9).

Proof. First note that any fixed point $\boldsymbol{x}^* \in \mathbb{R}^n$ of the DRP method can be expressed as

$$\boldsymbol{x}^* = \mathsf{sprox}_{\gamma g}(\boldsymbol{x}^* - \gamma\tau\mathsf{G}(\boldsymbol{x}^*)) = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min}\left\{\frac{1}{2}\|\boldsymbol{x} - (\boldsymbol{x}^* - \gamma\tau\mathsf{G}(\boldsymbol{x}^*))\|^2_{\mathbf{H}^\mathsf{T}\mathbf{H}} + \gamma g(\boldsymbol{x})\right\}, \tag{12}$$

where we used the definition of the scaled proximal operator. From the optimality conditions of the scaled proximal operator, we then get

$$\mathbf{0} \in \partial g(\boldsymbol{x}^*) + \tau\mathbf{H}^\mathsf{T}\mathbf{H}\mathsf{G}(\boldsymbol{x}^*). \tag{13}$$

On the other hand, the gradient of $p_{\boldsymbol{s}}$, defined in (8), can be expressed as

$$\nabla_{\boldsymbol{s}} p_{\boldsymbol{s}}(\boldsymbol{s}) = \int \left(\frac{1}{\sigma^2}\left(\mathbf{H}\boldsymbol{x} - \boldsymbol{s}\right)\right) G_\sigma(\boldsymbol{s} - \mathbf{H}\boldsymbol{x}) p_{\boldsymbol{x}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \frac{1}{\sigma^2}\left(\mathbf{H}\mathsf{R}(\boldsymbol{s}) - \boldsymbol{s}p_{\boldsymbol{s}}(\boldsymbol{s})\right), \tag{14}$$

where we used the gradient of the Gaussian density with respect to $\boldsymbol{s}$ and the definition of the MMSE restoration operator in eq. (7). By rearranging the terms, we obtain the following relationship

$$\mathbf{H}\mathsf{R}(\boldsymbol{s}) - \boldsymbol{s} = \sigma^2\nabla\log p_{\boldsymbol{s}}(\boldsymbol{s}), \quad \boldsymbol{s}\in\mathbb{R}^p. \tag{15}$$

By using the definitions $\mathsf{G}(\boldsymbol{x}) = \boldsymbol{x} - \mathsf{R}(\mathbf{H}\boldsymbol{x})$ and $h(\boldsymbol{x}) = -\tau\sigma^2\log p_{\boldsymbol{s}}(\mathbf{H}\boldsymbol{x})$, with $\boldsymbol{x}\in\mathbb{R}^n$, in (15), we obtain the following generalization of the well-known Tweedie's formula

$$\mathbf{H}^\mathsf{T}\mathbf{H}\mathsf{G}(\boldsymbol{x}) = \mathbf{H}^\mathsf{T}\mathbf{H}\left(\boldsymbol{x} - \mathsf{R}(\mathbf{H}\boldsymbol{x})\right) = -\sigma^2\nabla\log p_{\boldsymbol{s}}(\mathbf{H}\boldsymbol{x}) = \frac{1}{\tau}\nabla h(\boldsymbol{x}). \tag{16}$$

By combining(12) and (16), we directly obtain the desired result

$$\mathbf{0} \in \partial g(\boldsymbol{x}^*) + \nabla h(\boldsymbol{x}^*).$$

□

A.2 Proof of Theorem 2

Theorem. *Run DRP for for* $t \geq 1$ *iterations under Assumptions 1-4 using a step-size* $\gamma = \mu/(\alpha L)$ *with* $\alpha > 1$. *Then, for each iteration* $1 \leq k \leq t$, *there exists* $\boldsymbol{w}(\boldsymbol{x}^k) \in \partial f(\boldsymbol{x}^k)$ *such that*

$$\min_{1\leq k\leq t}\|\boldsymbol{w}(\boldsymbol{x}^k)\|_2^2 \leq \frac{1}{t}\sum_{k=1}^{t}\|\boldsymbol{w}(\boldsymbol{x}^k)\|_2^2 \leq \frac{C(f(\boldsymbol{x}^0) - f^*)}{t},$$

where $C > 0$ *is an iteration independent constant.*

Proof. Consider the iteration $k \geq 1$ of DRP

$$\boldsymbol{x}^k = \mathsf{sprox}_{\gamma g}\left(\boldsymbol{x}^{k-1} - \gamma\tau\mathsf{G}(\boldsymbol{x}^{k-1})\right) \quad \text{with} \quad \mathsf{G} \coloneqq \boldsymbol{x} - \mathsf{R}(\mathbf{H}\boldsymbol{x}),$$

where R is the MMSE restoration operator specified in (7). This implies that $\boldsymbol{x}^k$ minimizes

$$\begin{aligned}\varphi(\boldsymbol{x}) &\coloneqq \frac{1}{2\gamma}(\boldsymbol{x}-\boldsymbol{x}^{k-1})^\mathsf{T}\mathbf{H}^\mathsf{T}\mathbf{H}(\boldsymbol{x}-\boldsymbol{x}^{k-1}) + \left[\tau\mathbf{H}^\mathsf{T}\mathbf{H}\mathsf{G}(\boldsymbol{x}^{k-1})\right]^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^{k-1}) + g(\boldsymbol{x}) \\ &= \frac{1}{2\gamma}(\boldsymbol{x}-\boldsymbol{x}^{k-1})^\mathsf{T}\mathbf{H}^\mathsf{T}\mathbf{H}(\boldsymbol{x}-\boldsymbol{x}^{k-1}) + \nabla h(\boldsymbol{x}^{k-1})^\mathsf{T}(\boldsymbol{x}-\boldsymbol{x}^{k-1}) + g(\boldsymbol{x}),\end{aligned}$$

where in the second inequality we used eq. (16) from the proof in Supplement A.1. By evaluating φ at $\boldsymbol{x}^k$ and $\boldsymbol{x}^{k-1}$, we obtain the following useful inequality

$$g(\boldsymbol{x}^k) \leq g(\boldsymbol{x}^{k-1}) - \frac{1}{2\gamma}(\boldsymbol{x}^k-\boldsymbol{x}^{k-1})^\mathsf{T}\mathbf{H}^\mathsf{T}\mathbf{H}(\boldsymbol{x}^k-\boldsymbol{x}^{k-1}) - \nabla h(\boldsymbol{x}^{k-1})^\mathsf{T}(\boldsymbol{x}^k-\boldsymbol{x}^{k-1}). \tag{17}$$

On the other hand, from the L-Lipschitz continuity of ∇h, we have the following bound

$$h(\boldsymbol{x}^k) \leq h(\boldsymbol{x}^{k-1}) + \nabla h(\boldsymbol{x}^{k-1})^\mathsf{T}(\boldsymbol{x}^k-\boldsymbol{x}^{k-1}) + \frac{L}{2}\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|_2^2. \tag{18}$$

By combining eqs. (17) and (18), we obtain

$$\begin{aligned}f(\boldsymbol{x}^k) &\leq f(\boldsymbol{x}^{k-1}) - \frac{1}{2}(\boldsymbol{x}^k-\boldsymbol{x}^{k-1})^\mathsf{T}\left[\frac{1}{\gamma}\mathbf{H}^\mathsf{T}\mathbf{H} - L\mathbf{I}\right](\boldsymbol{x}^k-\boldsymbol{x}^{k-1}) \\ &\leq f(\boldsymbol{x}^{k-1}) - (\alpha-1)\frac{L}{2}\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|_2^2,\end{aligned} \tag{19}$$

where we used $\gamma = \mu/(\alpha L)$ with $\alpha > 1$ and $\mu > 0$ defined in Assumption 4.

On the other hand, from the optimality conditions for φ, we also have

$$\begin{aligned}&\mathbf{0} \in \mathbf{H}^\mathsf{T}\mathbf{H}(\boldsymbol{x}^k-\boldsymbol{x}^{k-1}+\gamma\tau\mathsf{G}(\boldsymbol{x}^{k-1})) + \gamma\partial g(\boldsymbol{x}^k) \\ \Leftrightarrow \quad &\frac{1}{\gamma}\mathbf{H}^\mathsf{T}\mathbf{H}(\boldsymbol{x}^k-\boldsymbol{x}^{k-1}) \in \partial g(\boldsymbol{x}^k) + \nabla h(\boldsymbol{x}^{k-1}),\end{aligned}$$

where we used eq. (16) from Supplement A.1. This directly implies that the following inclusion

$$\boldsymbol{w}(\boldsymbol{x}^k) \coloneqq \frac{1}{\gamma}\mathbf{H}^\mathsf{T}\mathbf{H}(\boldsymbol{x}^k-\boldsymbol{x}^{k-1}) + \nabla h(\boldsymbol{x}^k) - \nabla h(\boldsymbol{x}^{k-1}) \in \partial f(\boldsymbol{x}^k)$$

The norm of the subgradient $\boldsymbol{w}(\boldsymbol{x}^k)$ can be bounded as follows

$$\begin{aligned}\|\boldsymbol{w}(\boldsymbol{x}^k)\|_2 &\leq \frac{1}{\gamma}\|\mathbf{H}^\mathsf{T}\mathbf{H}(\boldsymbol{x}^k-\boldsymbol{x}^{k-1})\|_2 + \|\nabla h(\boldsymbol{x}^k) - \nabla h(\boldsymbol{x}^{k-1})\|_2 \\ &\leq L\left(\alpha(\lambda/\mu)+1\right)\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|_2,\end{aligned} \tag{20}$$

where we used the Lipschitz constant of ∇h, $\gamma = \mu/(\alpha L)$, and $\lambda \geq \mu > 0$ defined in Assumption 4.

By combining eqs. (19) and (20), we obtain the following inequality

$$\|\boldsymbol{w}(\boldsymbol{x}^k)\|_2^2 \leq A_1\|\boldsymbol{x}^k-\boldsymbol{x}^{k-1}\|_2^2 \leq A_2(f(\boldsymbol{x}^{k-1}) - f(\boldsymbol{x}^k)), \tag{21}$$

where $A_1 \coloneqq L^2(\alpha(\lambda/\mu)+1)^2 > 0$ and $A_2 \coloneqq 2A_1/(L(\alpha-1)) > 0$. Hence, by averaging over $t \geq 1$ iterations, we can directly get the desired result

$$\min_{1\leq k\leq t}\|\boldsymbol{w}(\boldsymbol{x}^k)\|_2^2 \leq \frac{1}{t}\sum_{k=1}^{t}\|\boldsymbol{w}(\boldsymbol{x}^k)\|_2^2 \leq \frac{A_2(f(\boldsymbol{x}^0)-f^*)}{t}. \tag{22}$$

This implies that $\boldsymbol{w}(\boldsymbol{x}^k) \to \mathbf{0}$ as $t \to \infty$. □

B Additional Numerical Results

B.1 Image denoising

In this subsection, we show the performance of DRP on Gaussian image denoising. The corresponding measurement model is $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}$, where $\boldsymbol{e}$ is AWGN with the standard deviation σ and $\boldsymbol{x}$ is the unknown clean image. We use the same SwinIR SR model, introduced in 5.1, as the prior for denoising. The degradation model in the SwinIR prior is the operation $\mathbf{H}$ corresponding to bicubic downsampling. The scaled proximal operator $\mathsf{sprox}_{\lambda g}$ in (5) with data-fidelity term $g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2$ can be written as

$$\mathsf{sprox}_{\gamma g}(\boldsymbol{z}) \coloneqq (\mathbf{I} + \gamma \mathbf{H}^{\mathsf{T}}\mathbf{H})^{-1}[\boldsymbol{y} + \gamma \mathbf{H}^{\mathsf{T}}\mathbf{H}\boldsymbol{z}], \tag{23}$$

which can be efficiently implemented using CG, as in Section 5.2.

We compare DRP with one of SOTA denoising model DRUNet (Zhang et al., 2019) on noise level ($\sigma = 0.1$). Figure 5 and Figure 6 illustrate the visual performance of DRP on the Set5 and CBSD68 datasets, respectively. Figure 7 further explores the impact of using different SR factors q as priors, elucidating how these choices influence the visual quality of denoising.

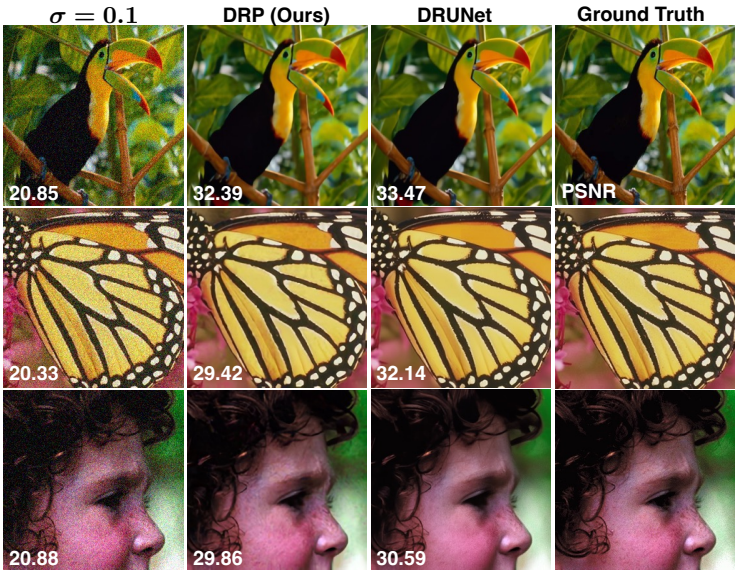

Figure 5: Illustration of denoising results of DRP on Set5 dataset with noise level $\sigma = 0.1$. Each image is labeled by its PSNR (dB) with respect to the original image.

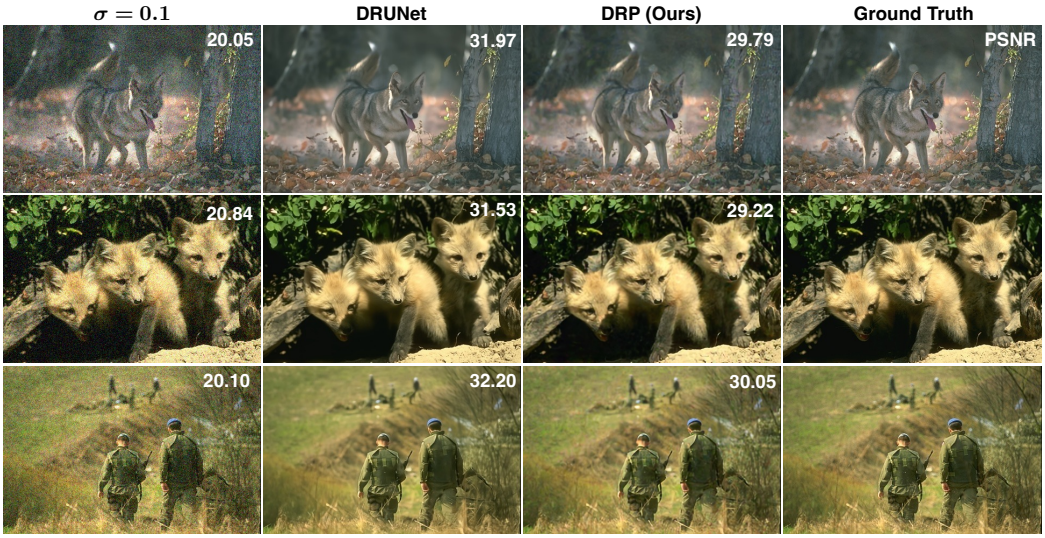

Figure 6: Illustration of denoising results of DRP on CBSD68 dataset with noise level $\sigma = 0.1$. Each image is labeled by its PSNR (dB) with respect to the original image.

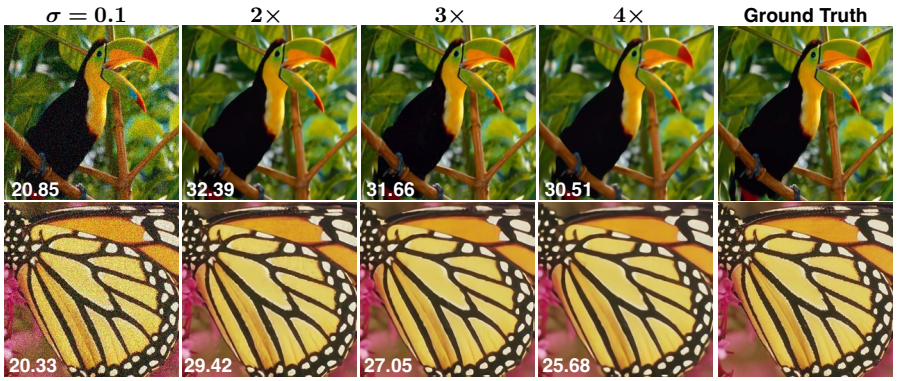

Figure 7: Illustration of denoising results of DRP on Set5 dataset with three SR level prior ($2\times$, $3\times$ and $4\times$). Each image is labeled by its PSNR (dB) with respect to the original image.

B.2 Comparison With GS-PnP

In this subsection, we will compare DRP with the recent gradient-step denoiser PnP method (GS-PnP) (Hurault et al., 2022a). These comparisons were not included in the main paper due to space, but are provided here for completeness. GS-PnP provides comparable performance on image deblurring and single image super resolution as DPIR (Zhang et al., 2019), but comes with theoretical convergence guarantees.

Table 3 shows that DRP outperforms both DPIR and GS-PnP on image deblurring in most settings in terms of PSNR. Similarly, Table 4 shows that DRP can achieve better SISR performance in terms of PSNR compared to both methods. Figure 8 provides additional visual results on SISR showing that DRP can recover intricate details and sharpen features.

Kernel	Datasets	GS-PnP	DPIR	DRP
	Set3c	29.53	29.78	**30.69**
	CBSD68	28.86	28.70	**29.10**
	Set3c	27.52	27.32	**27.89**
	CBSD68	27.44	**27.52**	27.46

Table 3: PSNR performance of GS-PnP, DPIR, and DRP for image deblurring on Set3c and CBSD68 datasets on two blur kernels. The **best** and second best results are highlighted.

SR	Kernels	Datasets	GS-PnP	DPIR	DRP
2×		Set3c	28.23	28.18	**29.26**
		CBSD68	28.03	27.97	**28.12**
		Set3c	26.19	26.80	**27.41**
		CBSD68	26.79	26.98	**26.98**
3×		Set3c	26.20	26.64	**27.77**
		CBSD68	26.77	26.80	**27.18**
		Set3c	25.18	25.84	**26.84**
		CBSD68	26.30	26.39	**26.60**

Table 4: PSNR performance of GS-PnP, DPIR, and DRP for 2× and 3× SISR on the Set3c and CBSD68 datasets, using two blur kernels. The **best** and second best results are highlighted.

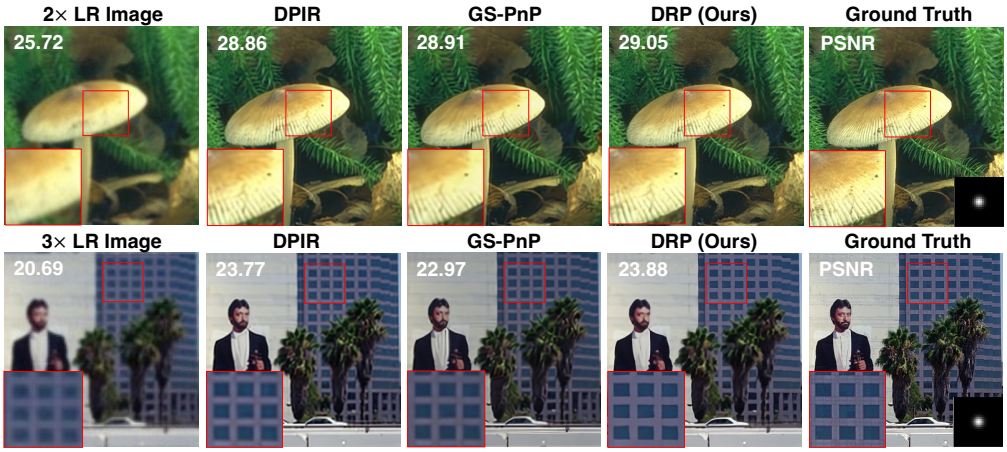

Figure 8: Illustration of SISR results of DRP compared with two SOTA denoiser based PnP method DPIR (Zhang et al., 2019) and GS-PnP (Hurault et al., 2022a). The top row displays the performance for 2× SISR task, while the bottom row showcases results for 3× SISR task. Each image is labeled by its PSNR (dB) with respect to the original image, and the visual difference is highlighted by the boxes in the left bottom corner.

B.3 Comparison with Diffusion Posterior Sampling

There is a growing interest in using denoisers within diffusion models for solving inverse problems (Zhu et al., 2023; Wang et al., 2022; Chung et al., 2023). One of the most wiedely-adopted diffusion model in this context is the *diffusion posterior sampling (DPS)* method from (Chung et al., 2023), which integrates pre-trained denoisers and measurement models for posterior sampling. One may argue that DPS is related to PnP due to the use of image denoisers as priors. In this section, we present results comparing DRP with DPS for deblurring human faces. We used the public implementation of DPS on the GitHub page that uses the prior specifically trained on human face image dataset (Chung et al., 2023). DRP uses the same SwinIR model trained on general image datasets (see Section 5.1). DPS and DRP are related but very different classes of methods. While DPS seeks to use denoisers to generate perceptually realistic solutions to inverse problems, DRP enables the adaptation of pre-trained restoration models as priors for solving other inverse problems.

Table 5 presents PSNR results obtained by DPS and DRP for human face deblurring. While we omitted the visual results from the paper for the privacy reasons, we will be happy to provide them if requested by the reviewers. Overall, DPS achieves more perceptually realistic images, while DRP achieves higher PSNR and more closely matches the ground truth images. This is not surprising when considering the generative nature of DPS. A similar observation is available in the original DPS publication, which reported better PSNR and SSIM performance of PnP-ADMM relative to DPS on SISR and deblurring (see Supplement E in (Chung et al., 2023)).

Kernel	DPS	DRP
	29.61	**34.61**
	28.80	**33.05**

Table 5: PSNR performance of DPS and DRP for image deblurring on three sample images from FFHQ validation set (provided in the DPS GitHub project) with two blur kernels. The **best** results are highlighted.

B.4 PERFORMANCE OF SWINIR TRAINED FOR BICUBIC SR ON A MISMATCHED SISR TASK

In this section, we make a noteworthy point: the SwinIR SR network we used as a prior in our DRP method is specifically trained for the bicubic SR task. Its direct application to the SISR task (which is SR under Gaussian blur kernels and additive white Gaussian noise) leads to sub-optimal performance. This implies that our DRP method has the capacity to use a mismatched restoration model as an implicit prior, effectively adapting it for other image restoration tasks.

Figure 9 present qualitative and quantitative PSNR results on the Set3c dataset. Note how the direct use of SwinIR trained for bicubic SR does poorly on the SISR task, while using it within our DRP method as a prior leads to the SOTA performance.

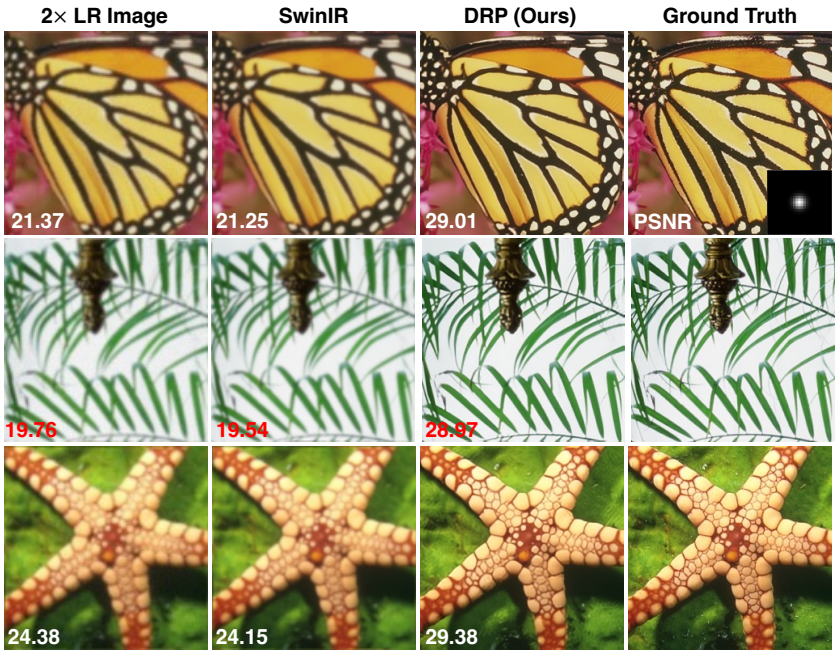

Figure 9: Illustration of $2\times$ SISR using DRP compared with the directly use of SwinIR trained for bicubic SR. Each image is labeled by its PSNR (dB) with respect to the ground truth.

B.5 Additional Evaluation using SSIM and LPIPS Quality Metrics

In this section, we present results using two popular quality metrics—SSIM and LPIPS—for evaluating image restoration. We compare the proposed DRP method using SwinIR trained for bicubic SR as a prior against PnP-ADMM and DPIR. The results below complement those using PSRN in Table 1 and Table 2 in the main paper. As can be seen in Table 6 and Table 7, DRP always achieves better LPIPS and in most cases achieves better SSIM in all the debluring and SISR settings.

Kernel	Datasets	PnP-ADMM	DPIR	DRP
	Set3c	29.11/0.925/0.084	29.53/0.930/0.093	**30.69/0.940/0.047**
	Set5	32.31/0.900/0.095	32.38/0.901/0.106	**32.79/0.902/0.075**
	CBSD68	28.90/0.832/0.229	28.86/0.836/0.246	**29.10/0.837/0.198**
	McMaster	32.20/0.895/0.098	32.42/0.900/0.110	**32.79/0.901/0.075**
	Set3c	27.05/0.888/0.137	27.52/0.901/0137	**27.89/0.908/0.075**
	Set5	30.77/0.873/0.131	30.94/**0.880**/0.137	**31.04**/0.875/**0.102**
	CBSD68	27.45/0.783/0.293	**27.52**/**0.787**/0.307	27.46/0.773/**0.245**
	McMaster	30.50/0.863/0.140	30.78/**0.871**/0.146	**30.79**/0.865/**0.105**

Table 6: PSNR↑/SSIM↑/LPIPS↓ of DRP and several SOTA PnP methods on image deblurring with the Gaussian blur kernels of standard deviations 1.6 and 2.0 on Set3c, Set5, CBSD68, and McMaster datasets. The **best** and second best results are highlighted. Note how DRP using image SR prior can outperform SOTA PnP methods that use denoisers as priors.

SR	Kernel	Datasets	PnP-ADMM	DPIR	DRP
2×		Set3c	27.88/0.903/0.114	28.18/0.908/0.117	**29.26/0.915/0.066**
		Set5	31.41/0.879/0.119	31.42/**0.883**/0.124	**31.47**/0.876/**0.083**
		CBSD68	28.00/0.796/0.274	27.97/**0.801**/0.286	**28.12**/0.799/**0.230**
		McMaster	30.98/0.869/0.129	31.16/**0.875**/0.134	**31.39**/0.873/**0.089**
		Set3c	25.86/0.861/0.163	26.80/0.884/0.149	**27.41/0.891/0.093**
		Set5	30.06/0.854/0.152	30.36/**0.862**/0.153	**30.42**/0.859/**0.118**
		CBSD68	26.88/0.749/0.326	26.98/**0.757**/0.330	**26.98**/0.745/**0.273**
		McMaster	29.53/0.836/0.170	29.87/**0.847**/0.167	**30.03**/0.845/**0.127**
3×		Set3c	25.85/0.865/0.151	26.64/0.883/0.147	**27.77/0.898/0.077**
		Set5	30.09/0.857/0.144	30.39/0.863/0.149	**30.83/0.868/0.107**
		CBSD68	26.78/0.751/0.312	26.80/0.755/0.325	**27.18/0.763/0.248**
		McMaster	29.52/0.839/0.160	29.82/**0.848**/0.162	**29.92**/0.841/**0.126**
		Set3c	24.87/0.834/0.192	25.84/0.863/0.171	**26.84/0.879/0.093**
		Set5	29.26/0.836/0.172	29.70/0.847/0.169	**29.88/0.847/0.118**
		CBSD68	26.28/0.722/0.353	26.39/0.730/0.352	**26.60/0.735/0.270**
		McMaster	28.72/0.814/0.196	29.11/0.827/0.187	**29.47/0.832/0.127**

Table 7: PSNR↑/SSIM↑/LPIPS↓ comparison of DRP and several baselines for SISR on Set3c, Set5, CBSD68, and McMaster datasets. The **best** and second best results are highlighted. Note the excellent quantitative performance of DRP, which suggests the potential of using general restoration models as priors.

B.6 COMPARISON ON SISR TASK WITH ADDITIONAL BASELINES

In this section, We will compare our proposed DRP method with some other baseline methods on 2× SISR task. Specificity, we will compare it with DIP prior-based methods BSRDM (Yue et al., 2022) and pure learning-based method BSRNet (Zhang et al., 2021) on the same setting as we shown in Section 5.3. As observable in Table 8, DRP achieves better performance compared to both of them.

SISR	Kernel	Datasets	BSRDM	BSRGAN	DRP
2×		Set3c	23.51/0.825/0.122	22.69/0.818/0.087	**29.26/0.915/0.066**
		Set5	27.80/0.827/0.138	27.07/0.806/0.096	**31.47/0.876/0.083**
		CBSD68	25.66/0.718/0.319	25.51/0.723/**0.194**	**28.12/0.799**/0.230
		McMaster	27.43/0.804/0.187	26.78/0.785/0.107	**31.39/0.873/0.089**
		Set3c	22.34/0.768/0.204	22.62/0.801/0.108	**27.41/0.891/0.093**
		Set5	26.68/0.791/0.206	26.88/0.789/0.123	**30.42/0.859/0.118**
		CBSD68	24.86/0.670/0.395	25.24/0.695/**0.234**	**26.98/0.745**/0.273
		McMaster	26.47/0.768/0.255	26.69/0.770/0.131	**30.03/0.845/0.127**

Table 8: PSNR↑/SSIM↑/LPIPS↓ comparison of DRP and several baselines for SISR on Set3c, Set5, CBSD68 and McMaster datasets. The **best** and second best results are highlighted. Note the excellent quantitative performance of DRP, which suggests the potential of using general restoration models as priors.

B.7 Comparison with DPIR (SwinIR): DPIR equipped with SwinIR denoiser

In this section,we will conduct an ablation study to shown the enhanced performance of DRP over denoiser-prior methods that do not originate from the SwinIR structure. To support this claim, we substitute the DRUNet denoiser utilized in DPIR with the SwinIR denoiser (Liang et al., 2021), called DPIR(SwinIR), aligning it with the same network structure employed in our SwinIR SR prior. As shown in table 9 and table 10, DRP can achieve better performance than both DPIR(SwinIR) and DPIR(DRUNet) in most of the debluring and SISR settings.

Kernel	Datasets	DPIR (SwinIR)	DPIR (DRUNet)	DRP
	Set3c	30.66/0.932/0.066	29.53/0.930/0.093	**30.69/0.940/0.047**
	Set5	32.35/0.888/0.103	32.38/0.901/0.106	**32.79/0.902/0.075**
	CBSD68	27.81/0.769/0.276	28.86/0.836/0.246	**29.10/0.837/0.198**
	McMaster	31.10/0.859/0.132	32.42/0.900/0.110	**32.79/0.901/0.075**
	Set3c	27.77/0.896/0.124	27.52/0.901/0137	**27.89/0.908/0.075**
	Set5	30.85/0.866/0.133	30.94/**0.880**/0.137	**31.04**/0.875/**0.102**
	CBSD68	27.29/0.764/0.308	**27.52**/**0.787**/0.307	27.46/0.773/**0.245**
	McMaster	30.69/0.859/0.143	30.78/**0.871**/0.146	**30.79**/0.865/**0.105**

Table 9: PSNR (dB) of DRP and several SOTA methods for solving inverse problems using denoisers on image deblurring with the Gaussian blur kernels of standard deviation 1.6 and 2.0 on Set3c, Set5, CBSD68 and McMaster datasets. The **best** and second best results are highlighted. Note how DRP can outperform SOTA PnP methods that use denoisers as priors.

SISR	Kernel	Datasets	DPIR(SwinIR)	DPIR(DRUNet)	DRP (SR)
2×		Set3c	28.90/0.911/0.099	28.18/0.908/0.117	**29.26/0.915/0.066**
		Set5	31.38/0.872/0.122	31.42/**0.883**/0.124	**31.47**/0.876/**0.083**
		CBSD68	27.92/0.783/0.280	27.97/**0.801**/0.286	**28.12**/0.799/**0.230**
		McMaster	31.34/0.870/0.127	31.16/**0.875**/0.134	**31.39**/0.873/**0.089**
		Set3c	26.75/0.877/0.147	26.80/0.884/0.149	**27.41/0.891/0.093**
		Set5	30.23/0.852/0.148	30.36/**0.862**/0.153	**30.42**/0.859/**0.118**
		CBSD68	26.83/0.739/0.330	26.98/**0.757**/0.330	**26.98**/0.745/**0.273**
		McMaster	29.96/0.842/0.139	29.87/0.844/0.167	**30.03/0.845/0.127**
3×		Set3c	26.81/0.883/0.136	26.64/0.883/0.147	**27.77/0.898/0.077**
		Set5	30.35/0.855/0.141	30.39/0.863/0.149	**30.83/0.868/0.107**
		CBSD68	26.79/0.742/0.320	26.80/0.755/0.325	**27.18/0.763/0.248**
		McMaster	**29.94**/0.845/0.154	29.82/**0.848**/0.162	29.92/0.841/**0.126**
		Set3c	25.66/0.853/0.175	25.84/0.863/0.171	**26.84/0.879/0.093**
		Set5	29.55/0.837/0.164	29.70/0.847/0.169	**29.88/0.847/0.118**
		CBSD68	26.29/0.715/0.352	26.39/0.730/0.352	**26.60/0.735/0.270**
		McMaster	29.20/0.825/0.180	29.11/0.827/0.187	**29.47/0.832/0.127**

Table 10: PSNR↑/SSIM↑/LPIPS↓ comparison of DRP and several baselines for SISR on Set3c, Set5, CBSD68 and McMaster datasets. The **best** and second best results are highlighted. Note the excellent quantitative performance of DRP, which suggests the potential of using general restoration models as priors.

B.8 Evaluation of DRP using Image Deblurring Network as a Prior

In this section, we show that DRP can be used with restoration priors and network architectures (beyond the SwinIR-based bicubic SR prior). To that end, we evaluate DRP under the deep unfolding network for image deblurring as a prior for solving Single Image Super Resolution (SISR) task.

B.8.1 Restoration Prior based on Deblurring Deep Unfolding Network

Deep Unfolding Network Architecture. We pre-trained a deblurring model using the USR-Net (Zhang et al., 2020) architecture based on deep unfolding network. Our training dataset consists of both the DIV2K (Agustsson & Timofte, 2017) and Flick2K (Lim et al., 2017) datasets, containing 3,450 color images in total. During training, we applied synthesized blur kernels to the input images, introducing Additive White Gaussian Noise (AWGN) characterized by σ randomly chosen in [0, 5/255]. This process uses the same synthesized blur kernels as detailed in (Zhang et al., 2020).

The restoration network is trained only for the deblurring task, which limits its performance on the SISR task. The direct application of the deblur prior on SISR is presented under the name "Deblur" in Table 11. Note how the direc use of the deblurring network on SISR without DRP does not work.

B.8.2 Single Image Super Resolution using Deblurring Prior

The measurement operator in SISR can be written as $\boldsymbol{A} = \boldsymbol{S}\boldsymbol{K}$, where $\boldsymbol{K}$ is convolution with the blur kernel $\boldsymbol{k}$ and $\boldsymbol{S}$ performs standard d-fold down-sampling with $d^2 = n/m$. The scaled proximal operator $\mathsf{sprox}_{\lambda g}$ in (5) with data-fidelity term $g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{K}\boldsymbol{x}\|_2^2$ can be write as:

$$\mathsf{sprox}_{\gamma g}(\boldsymbol{z}) = (\boldsymbol{K}^\top \boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{K} + \gamma \mathbf{H}^\top \mathbf{H})^{-1}[\boldsymbol{K}^\top \boldsymbol{S}^\top \boldsymbol{y} + \gamma \mathbf{H}^\top \mathbf{H}\boldsymbol{z}], \tag{24}$$

where $\mathbf{H}$ is the convolution operator with blur kernel $\boldsymbol{k}$. Similarly to deblurring in Section 5.3, we use CG to efficiently compute (11). We adjust the hyper-parameter γ, τ and then use these parameters on the remaining datasets.

We refer to DRP using the Deblurring Prior as DRP (Deblur) and DRP using the Super-Resolution Prior as DRP (SR). Table 11 provides a comprehensive quantitative evaluation of the reconstruction performance for the $2\times$ Super-Resolution (SISR) task, using two distinct blur kernels. The results show that DRP (Deblur) outperforms DPIR and achieves a comparable performance as DRP (SR) across four datasets.

	Kernel	Datasets	Deblur	DPIR	DRP (SR)	DRP (Deblur)
$2\times$		Set3c	13.63/0.300/0.640	28.18/0.908/0.117	**29.26**/0.915/**0.066**	29.13/**0.921**/0.066
		Set5	14.10/0.172/0.814	31.42/0.883/0.124	31.47/0.876/**0.083**	**31.81**/**0.883**/0.122
		CBSD68	13.69/0.152/0.849	27.97/0.801/0.286	28.12/0.799/0.230	**28.27**/**0.704**/**0.224**
		McMaster	14.31/0.158/0.844	31.16/0.875/0.134	31.39/0.873/**0.089**	**31.52**/**0.881**/0.091
		Set3c	7.85/0.09/0.956	26.80/0.884/0.149	27.41/0.891/0.093	**27.52**/**0.893**/**0.092**
		Set5	8.77/0.052/1.00	30.36/0.862/0.153	30.42/0.859/**0.118**	**30.73**/**0.864**/0.124
		CBSD68	8.58/0.040/1.00	26.98/**0.757**/0.330	26.98/0.745/**0.273**	**27.19**/0.755/0.281
		McMaster	9.26/0.048/1.00	29.87/0.844/0.167	30.03/0.845/**0.127**	**30.16**/**0.849**/0.134

Table 11: PSNR↑/SSIM↑/LPIPS↓ comparison of two DRP variants and several baselines for $2\times$ SISR on Set3c, Set5, CBSD68, and McMaster datasets. The **best** and second best results are highlighted. Note how the deblurring model doesn't generalize well on SISR without DRP; while its integration as a prior into DRP yields SOTA performance.

B.9 EVALUATION OF DIFFERENT RANDOM SEEDS

In this section, we present results illustrating the influence of different random seeds on the performance. We use three different random seeds for the image deblurring task on the Set3c dataset. For each seed, DRP is compared with the best baseline method DPIR on the image deblurring task.

Seed ID	Kernel	DPIR	DRP
ID = 0		29.53/0.930/0.093	**30.69/0.940/0.047**
ID = 10		29.50/0.930/0.092	**30.65/0.940/0.047**
ID = 100		29.53/0.930/0.093	**30.67/0.941/0.047**
ID = 0		27.52/0.901/0.137	**27.89/0.908/0.075**
ID = 10		27.51/0.901/0.135	**27.85/0.907/0.071**
ID = 100		27.52/0.900/0.136	**27.88/0.909/0.073**

Table 12: PSNR↑/SSIM↑/LPIPS↓ performance of DPIR and DRP for image deblurring on the Set3c dataset with two blur kernels and three different random seeds. The **best** results are highlighted.

B.10 NUMERICAL EVALUATION USING NON-MMSE PRIORS

In this section, we show that although our theory relies on the assumption that the restoration prior used for inference performs MMSE estimation, in practice, DRP work even for priors trained with other loss functions, such as the ℓ_1-norm or SSIM.

We retrained the SwinIR model for bicubic SR using the same setting as in Section 5.1, but with the ℓ_1-norm loss instead of ℓ_2. Table 13 shows that there is not much difference between the two. This suggests that DRP can be stable in practice even without the MMSE assumption.

Kernel	DPIR	DRP (SR_{ℓ_1})	DRP (SR_{ℓ_2})
	29.53/0.930/0.093	30.67/**0.941/0.043**	**30.69**/0.940/0.047
	27.52/0.901/0.137	27.85/**0.909/0.073**	**27.89**/0.908/0.075

Table 13: PSNR↑/SSIM↑/LPIPS↓ performance of DRP(SR_{L1}), DRP(SR_{L2}) and DPIR for image deblurring on the Set3c dataset using two blur kernels. The **best** and second best results are highlighted.