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 image 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 Deep Restoration Priors (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 \( x \in \mathbb{R}^n \) from its corrupted observation

\[
y = Ax + e,
\]

where \( A \in \mathbb{R}^{m \times n} \) is a measurement operator and \( e \in \mathbb{R}^m \) is the noise. A common strategy for addressing inverse problems involves formulating them as an optimization problem

\[
\hat{x} \in \arg \min_{x \in \mathbb{R}^n} f(x) \quad \text{with} \quad f(x) = g(x) + h(x),
\]

where \( g \) is the data-fidelity term that measures the fidelity to the observation \( y \) and \( h \) is the regularizer that incorporates prior knowledge on \( x \). For example, common functionals in imaging inverse problems are the least-squares data-fidelity term \( g(x) = \frac{1}{2} \| Ax - y \|_2^2 \) and the total variation (TV) regularizer \( h(x) = \tau \| D x \|_1 \), where \( D \) 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., 2020). 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)

\[
x^k \leftarrow \text{prox}_{\sigma g}(z^k) \quad \text{with} \quad z^k \leftarrow x^{k-1} - \gamma \tau (x^{k-1} - D_\sigma(x^{k-1})),
\]

where \( D_\sigma \) 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 \( (1 - D_\sigma) \) in (3) corresponds to a gradient \( \nabla h \) 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; Kadkhodaei & 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 $x^0 \in \mathbb{R}^n$ and parameters $\gamma, \tau > 0$
2: for $k = 1, 2, 3, \ldots$ do
3: $z^k \leftarrow x^{k-1} - \gamma \tau G(x^{k-1})$ where $G(x) := x - R(Hx)$
4: $x^k \leftarrow \text{sprox}_{\gamma g}(z^k)$
5: end for

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

$$s = Hx + n \quad \text{with} \quad x \sim p_x, \quad n \sim \mathcal{N}(0, \sigma^2 I),$$

where $H \in \mathbb{R}^{p \times n}$ is a degradation operator, such as blur or downscaling, and $n \in \mathbb{R}^p$ is the additive white Gaussian noise (AWGN) of variance $\sigma^2$. The density $p_x$ 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 $H = I$, the restoration operator $R$ reduces to an AWGN denoiser used in the traditional PnP methods (Romano et al., 2017; Kadkhodaei & 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

$$\text{sprox}_{\gamma g}(z) := \text{sprox}_{\gamma H^T H}(z) = \arg \min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \| x - z \|^2_{H^T H} + \gamma g(x) \right\},$$

where $\| v \|_{H^T H} := v^T H^T H v$ denotes the weighted Euclidean seminorm of a vector $v$. When $H^T H$ 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 $H^T H$ 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 $x^* \in$
implies that the solutions of DRP satisfy the first-order conditions for the objective function \( h \) (Robbins, 1956; Miyasawa, 1961; Gribonval, 2011) to restoration operators. The theorem

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 \( p_y|x \), then the fixed-points of DRP can be interpreted as maximum-a-posteriori probability (MAP) solutions corresponding to the prior density \( p_u \). The density \( p_u \) is

\[ h(x) = -\tau \sigma^2 \log p_u(Hx), \quad x \in \mathbb{R}^n. \]  

Assumption 1. The prior density \( p_u \) is non-degenerate over \( \mathbb{R}^n \).

For example, a probability density \( p_u \) 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

\[ h(x) = -\tau \sigma^2 \log p_u(Hx), \quad x \in \mathbb{R}^n, \]  

where \( \tau \) is the parameter in Algorithm 1, \( p_u \) is the density of the observation (8), and \( \sigma^2 \) 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 \( p_u \) with a Gaussian density \( G_\sigma \) (Gribonval, 2011; Gribonval & Machart, 2013).

Assumption 2. The scaled proximal operator \( \text{prox}_{\tau g} \) is well-defined in the sense that there exists a solution to the problem (5) for any \( z \in \mathbb{R}^n \). 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, \( \text{prox}_{\tau g} \) 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 \( g(x) = \frac{1}{2} \| y - Ax \|^2_2 \). One can relax the assumption of convexity by considering \( g \) that is proper, closed, and coercive, in which case \( \text{prox}_{\tau g} \) 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 \( \text{prox}_{\tau g} \) 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 \( x^* \in \mathbb{R}^n \) of DRP satisfies

\[ 0 \in \partial g(x^*) + \nabla h(x^*), \]  

where \( h \) is given in (9).
related to the true prior $p_x$ through eq. (8), which implies that DRP has access to the prior $p_x$ through the restoration operator $\mathbb{R}$ via density $p_x$. As $H \rightarrow I$ and $\sigma \rightarrow 0$, the density $p_x$ approaches the prior distribution $p_x$.

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 $f(x) \geq f^*$ for all $x \in \mathbb{R}^n$.

**Assumption 4.** The function $h$ has a Lipschitz continuous gradient with constant $L > 0$. The degradation operator associated with the restoration network is such that $\lambda \geq H^\top H \geq \mu > 0$.

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 $\nabla h$ is Lipschitz continuous (Nesterov, 2004). It is also worth noting that the positive definiteness of $H^\top H$ 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 $w(x^k) \in \partial f(x^k)$ such that

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

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 $w(x^k) \rightarrow 0$ as $t \rightarrow \infty$. 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 $y = Ax + e$: (a) Image Deblurring and (b) Single Image Super Resolution (SISR). For both problems, we assume that $e$ is the additive white Gaussian noise (AWGN). We adopt the traditional $\ell_2$-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.
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.

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Datasets</th>
<th>SD-RED</th>
<th>PnP-ADMM</th>
<th>IRCNN+</th>
<th>DPIR</th>
<th>DRP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Set3c</td>
<td>27.14/0.121</td>
<td>29.11/0.099</td>
<td>28.14/0.086</td>
<td>29.53/0.110</td>
<td>30.69/0.056</td>
</tr>
<tr>
<td></td>
<td>Set5</td>
<td>29.78/0.160</td>
<td>32.31/0.137</td>
<td>29.46/0.130</td>
<td>32.38/0.152</td>
<td>32.79/0.109</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>25.78/0.344</td>
<td>28.90/0.287</td>
<td>26.86/0.307</td>
<td>28.86/0.307</td>
<td>29.10/0.246</td>
</tr>
<tr>
<td></td>
<td>McMaster</td>
<td>29.69/0.167</td>
<td>32.20/0.136</td>
<td>29.15/0.153</td>
<td>32.42/0.151</td>
<td>32.79/0.107</td>
</tr>
</tbody>
</table>

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× and 3× super resolution with the Gaussian blur kernel of standard deviation 2.0. Average distance $\|x^k - x^{k-1}\|_2^2$ and PSNR relative to the groundtruth are plotted, with shaded areas indicating the standard deviation of these metrics across all test images.
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 \( K^T K + \gamma H^T H \). As bicubic SR model is adopted as prior, the \( R(Hx) \) in Step 3 of Algorithm 1 performs a bicubic downsampling of the intermediate image \( x^{k-1} \) 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-tuned the hyper-parameter \( \tau \), 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 denoising 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× prior, (ii) using only 2× prior, and (iii) the use of the prior refinement strategy, which combines both the 2× and 3× 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.

\[
spro_{x,q}(z) = \left( K^T K + \gamma H^T H \right)^{-1} [K^T y + \gamma H^T H z].
\]
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 $3 \times$ prior, (ii) use of only $2 \times$ prior, and (iii) use of the prior-refinement strategy to leverage both $3 \times$ and $2 \times$ 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 $\sigma$ 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 $A = SK$, where $K$ is convolution with the blur kernel $k$ and $S$ performs standard $d$-fold down-sampling with $d^2 = n/m$. The scaled proximal operator $\text{sprox}_{\lambda g}$ in (5) with data-fidelity term $g(x) = \frac{1}{2} \| y - SKx \|_2^2$ can be write as:

$$\text{sprox}_{\lambda g}(z) = (K^T S^T SK + \gamma H^T H)^{-1}[K^T S^T y + \gamma H^T Hz],$$

where $H$ is the bicubic downsampling operator. Similarly to deblurring in Section 5.2, we use CG to efficiently compute (11). We adjust the hyper-parameter $\gamma$, $\tau$, 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 $25 \times 25$ Gaussian blur kernels, each with distinct standard deviations ($1.6$ and $2.0$), and for two distinct downsampling factors ($2 \times$ and $3 \times$), incorporating an AWGN vector $e$ 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 $2 \times$ and $3 \times$ SISR. Figure 4 shows the visual reconstruction results for the same downsampling factors. Table 2 summarizes the PSNR 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.
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 $\ell_1$-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 $H$ (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 $H$ for a given measurement operator $A$. 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


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

\[
0 \in \partial g(x^*) + \nabla h(x^*),
\]

where \( h \) is given in (9).

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

\[
x^* = \text{sprox}_{\gamma g}(x^* - \gamma \tau \mathbb{G}(x^*)) = \arg \min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \| x - (x^* - \gamma \tau \mathbb{G}(x^*)) \|_2^2 \mathbb{H}^T \mathbb{H} + \gamma g(x) \right\}, \tag{12}
\]

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

\[
0 \in \partial g(x^*) + \tau \mathbb{H}^T \mathbb{G}(x^*). \tag{13}
\]

On the other hand, the gradient of \( p_s \), defined in (8), can be expressed as

\[
\nabla_s p_s(s) = \int \left( \frac{1}{\sigma^2} (\mathbb{H} x - s) \right) \mathbb{G}_s(s - \mathbb{H} x) p_x(x) \, dx = \frac{1}{\sigma^2} (\mathbb{H} \mathbb{R}(s) - s p_s(s)), \tag{14}
\]

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

\[
\mathbb{H} \mathbb{R}(s) - s = \sigma^2 \nabla \log p_s(s), \quad s \in \mathbb{R}^p. \tag{15}
\]

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

\[
\mathbb{H}^T \mathbb{G}(x) = \mathbb{H}^T \mathbb{H} (x - \mathbb{R}(\mathbb{H} x)) = -\sigma^2 \nabla \log p_s(\mathbb{H} x) = \frac{1}{\tau} \nabla h(x). \tag{16}
\]

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

\[
0 \in \partial g(x^*) + \nabla h(x^*).
\]

\( \square \)

A.2 Proof of Theorem 2

**Theorem.** Run DRP 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 \( w(x^k) \in \partial f(x^k) \) such that

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

where \( C > 0 \) is an iteration independent constant.
**Proof.** Consider the iteration \( k \geq 1 \) of DRP
\[
x^k = \text{sprox}_g (x^{k-1} - \gamma \tau G(x^{k-1})) \quad \text{with} \quad G := x - R(Hx),
\]
where \( R \) is the MMSE restoration operator specified in (7). This implies that \( x^k \) minimizes
\[
\varphi(x) := \frac{1}{2\gamma}(x - x^{k-1})^T H^T H(x - x^{k-1}) + \left[\gamma H^T H g(x^{k-1})\right]^T (x - x^{k-1}) + g(x)
\]
\[
= \frac{1}{2\gamma}(x - x^{k-1})^T H^T H(x - x^{k-1}) + \nabla h(x^{k-1})^T (x - x^{k-1}) + g(x),
\]
where in the second inequality we used eq. (16) from the proof in Supplement A.1. By evaluating \( \varphi \) at \( x^k \) and \( x^{k-1} \), we obtain the following useful inequality
\[
g(x^k) \leq g(x^{k-1}) - \frac{1}{2\gamma}(x^k - x^{k-1})^T H^T H(x^k - x^{k-1}) - \nabla h(x^{k-1})^T(x^k - x^{k-1}). \tag{17}
\]
On the other hand, from the \( L \)-Lipschitz continuity of \( \nabla h \), we have the following bound
\[
h(x^k) \leq h(x^{k-1}) + \nabla h(x^{k-1})^T(x^k - x^{k-1}) + \frac{L}{2} \|x^k - x^{k-1}\|_2^2. \tag{18}
\]
By combining eqs. (17) and (18), we obtain
\[
f(x^k) \leq f(x^{k-1}) - \frac{1}{2}(x^k - x^{k-1})^T \left[\frac{1}{\gamma} H^T H - L I\right] (x^k - x^{k-1})
\]
\[
\leq f(x^{k-1}) - (\alpha - 1) \frac{L}{2} \|x^k - x^{k-1}\|_2^2, \tag{19}
\]
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 \( \varphi \), we also have
\[
0 \in H^T H(x^k - x^{k-1} + \gamma \tau G(x^{k-1})) + \gamma \partial g(x^k)
\]
\[
\Leftrightarrow \frac{1}{\gamma} H^T H(x^k - x^{k-1}) \in \partial g(x^k) + \nabla h(x^{k-1}),
\]
where we used eq. (16) from Supplement A.1. This directly implies that the following inclusion
\[
w(x^k) := \frac{1}{\gamma} H^T H(x^k - x^{k-1}) + \nabla h(x^k) - \nabla h(x^{k-1}) \in \partial f(x^k)
\]
The norm of the subgradient \( w(x^k) \) can be bounded as follows
\[
\|w(x^k)\|_2 \leq \frac{1}{\gamma} \|H^T H(x^k - x^{k-1})\|_2 + \|\nabla h(x^k) - \nabla h(x^{k-1})\|_2
\]
\[
\leq L (\alpha (\lambda/\mu) + 1) \|x^k - x^{k-1}\|_2, \tag{20}
\]
where we used the Lipschitz constant of \( \nabla h \), \( \gamma = \mu/(\alpha L) \), and \( \lambda \geq \mu > 0 \) defined in Assumption 4.

By combining eqs. (19) and (20), we obtain the following inequality
\[
\|w(x^k)\|_2^2 \leq A_1 \|x^k - x^{k-1}\|_2^2 \leq A_2 (f(x^{k-1}) - f(x^k)), \tag{21}
\]
where \( A_1 := L^2(\alpha (\lambda/\mu) + 1)^2 > 0 \) and \( A_2 := 2A_1/(L(\alpha - 1)) > 0 \). Hence, by averaging over \( t \geq 1 \) iterations, we can directly get the desired result
\[
\min_{1 \leq k \leq t} \|w(x^k)\|_2^2 \leq \frac{1}{t} \sum_{k=1}^{t} \|w(x^k)\|_2^2 \leq \frac{A_2 (f(x^0) - f^*)}{t}. \tag{22}
\]
This implies that \( w(x^k) \to 0 \) as \( t \to \infty. \) \qed
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 $y = x + e$, where $e$ is AWGN with the standard deviation $\sigma$ and $x$ 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 $H$ corresponding to bicubic downsampling. The scaled proximal operator $\text{sprox}_\lambda g$ in (5) with data-fidelity term $g(x) = \frac{1}{2} \|y - x\|^2_2$ can be written as

$$\text{sprox}_\gamma g(z) := (I + \gamma H^T H)^{-1}[y + \gamma H^T z],$$

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.](image-url)
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.
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.

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Datasets</th>
<th>GS-PnP</th>
<th>DPIR</th>
<th>DRP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Set3c</td>
<td>29.53</td>
<td>29.78</td>
<td>30.69</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>28.86</td>
<td>28.70</td>
<td>29.10</td>
</tr>
<tr>
<td></td>
<td>Set3c</td>
<td>27.52</td>
<td>27.32</td>
<td>27.89</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>27.44</td>
<td>27.52</td>
<td>27.46</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>SR Kernels</th>
<th>Datasets</th>
<th>GS-PnP</th>
<th>DPIR</th>
<th>DRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2×</td>
<td>Set3c</td>
<td>28.23</td>
<td>28.18</td>
<td>29.26</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>26.19</td>
<td>26.80</td>
<td>27.41</td>
</tr>
<tr>
<td></td>
<td>Set3c</td>
<td>26.20</td>
<td>26.64</td>
<td>27.77</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>25.18</td>
<td>25.84</td>
<td>26.84</td>
</tr>
<tr>
<td>3×</td>
<td>Set3c</td>
<td>26.79</td>
<td>26.98</td>
<td>26.98</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>26.77</td>
<td>26.98</td>
<td>27.18</td>
</tr>
<tr>
<td></td>
<td>Set3c</td>
<td>26.30</td>
<td>26.39</td>
<td>26.60</td>
</tr>
</tbody>
</table>

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 widely-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)).

<table>
<thead>
<tr>
<th>Kernel</th>
<th>DPS</th>
<th>DRP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>29.61</td>
<td>34.61</td>
</tr>
<tr>
<td></td>
<td>28.80</td>
<td>33.05</td>
</tr>
</tbody>
</table>

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× 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 deblurring and SISR settings.

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Datasets</th>
<th>PnP-ADMM</th>
<th>DPIR</th>
<th>DRP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Set3c</td>
<td>29.11/0.925/0.099</td>
<td>29.53/0.930/0.110</td>
<td>30.69/0.940/0.056</td>
</tr>
<tr>
<td></td>
<td>Set5</td>
<td>32.31/0.90/0.137</td>
<td>32.38/0.901/0.152</td>
<td>32.79/0.902/0.109</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>28.90/0.832/0.287</td>
<td>28.86/0.836/0.307</td>
<td>29.10/0.837/0.246</td>
</tr>
<tr>
<td></td>
<td>McMaster</td>
<td>32.20/0.895/0.136</td>
<td>32.42/0.900/0.151</td>
<td>32.79/0.901/0.107</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SR Kernel</th>
<th>Datasets</th>
<th>PnP-ADMM</th>
<th>DPIR</th>
<th>DRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2×</td>
<td>Set3c</td>
<td>27.05/0.888/0.157</td>
<td>27.52/0.901/0.159</td>
<td>27.89/0.908/0.088</td>
</tr>
<tr>
<td></td>
<td>Set5</td>
<td>30.77/0.873/0.181</td>
<td>30.94/0.880/0.188</td>
<td>31.04/0.875/0.143</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>27.45/0.783/0.360</td>
<td>27.52/0.787/0.376</td>
<td>27.46/0.773/0.296</td>
</tr>
<tr>
<td></td>
<td>McMaster</td>
<td>30.50/0.863/0.185</td>
<td>30.78/0.871/0.193</td>
<td>30.79/0.865/0.141</td>
</tr>
<tr>
<td>3×</td>
<td>Set3c</td>
<td>25.86/0.861/0.187</td>
<td>26.80/0.884/0.171</td>
<td>27.41/0.891/0.107</td>
</tr>
<tr>
<td></td>
<td>Set5</td>
<td>30.06/0.854/0.205</td>
<td>30.36/0.862/0.241</td>
<td>30.42/0.859/0.163</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>26.88/0.749/0.393</td>
<td>26.98/0.757/0.401</td>
<td>26.98/0.745/0.327</td>
</tr>
<tr>
<td></td>
<td>McMaster</td>
<td>29.53/0.836/0.219</td>
<td>29.87/0.847/0.215</td>
<td>30.03/0.845/0.167</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SR Kernel</th>
<th>Datasets</th>
<th>PnP-ADMM</th>
<th>DPIR</th>
<th>DRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2×</td>
<td>Set3c</td>
<td>25.85/0.865/0.173</td>
<td>26.64/0.883/0.169</td>
<td>27.77/0.896/0.090</td>
</tr>
<tr>
<td></td>
<td>Set5</td>
<td>30.09/0.857/0.195</td>
<td>30.39/0.863/0.202</td>
<td>30.83/0.868/0.149</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>26.78/0.751/0.378</td>
<td>26.80/0.755/0.394</td>
<td>27.18/0.763/0.299</td>
</tr>
<tr>
<td></td>
<td>McMaster</td>
<td>29.52/0.839/0.208</td>
<td>29.82/0.848/0.210</td>
<td>29.92/0.841/0.164</td>
</tr>
<tr>
<td>3×</td>
<td>Set3c</td>
<td>24.87/0.834/0.216</td>
<td>25.84/0.863/0.195</td>
<td>26.84/0.879/0.108</td>
</tr>
<tr>
<td></td>
<td>Set5</td>
<td>29.26/0.836/0.229</td>
<td>29.70/0.847/0.225</td>
<td>29.88/0.847/0.162</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>26.28/0.722/0.420</td>
<td>26.39/0.730/0.422</td>
<td>26.60/0.735/0.324</td>
</tr>
<tr>
<td></td>
<td>McMaster</td>
<td>28.72/0.814/0.246</td>
<td>29.11/0.827/0.236</td>
<td>29.47/0.832/0.167</td>
</tr>
</tbody>
</table>

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.

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 $2 \times$ SISR task. Specifically, 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.

<table>
<thead>
<tr>
<th>SISR Kernel</th>
<th>Datasets</th>
<th>BSRDM</th>
<th>BSRGAN</th>
<th>DRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set3c</td>
<td></td>
<td>23.51/0.825/0.139</td>
<td>22.69/0.818/0.102</td>
<td>29.26/0.915/0.079</td>
</tr>
<tr>
<td>Set5</td>
<td></td>
<td>27.80/0.827/0.179</td>
<td>27.07/0.806/0.124</td>
<td>31.47/0.876/0.120</td>
</tr>
<tr>
<td>CBSD68</td>
<td></td>
<td>25.66/0.718/0.378</td>
<td>25.51/0.723/0.225</td>
<td>28.12/0.799/0.128</td>
</tr>
<tr>
<td>McMaster</td>
<td></td>
<td>27.43/0.804/0.226</td>
<td>26.78/0.785/0.132</td>
<td>31.39/0.873/0.122</td>
</tr>
<tr>
<td>Set3c</td>
<td></td>
<td>22.34/0.768/0.228</td>
<td>22.62/0.801/0.123</td>
<td>27.41/0.891/0.107</td>
</tr>
<tr>
<td>Set5</td>
<td></td>
<td>26.68/0.791/0.246</td>
<td>26.88/0.789/0.153</td>
<td>30.42/0.859/0.163</td>
</tr>
<tr>
<td>CBSD68</td>
<td></td>
<td>24.86/0.670/0.451</td>
<td>25.24/0.695/0.265</td>
<td>26.98/0.745/0.327</td>
</tr>
<tr>
<td>McMaster</td>
<td></td>
<td>26.47/0.768/0.294</td>
<td>26.69/0.770/0.157</td>
<td>30.03/0.845/0.167</td>
</tr>
</tbody>
</table>

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 show 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 deblurring and SISR settings.

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Datasets</th>
<th>DPIR (SwinIR)</th>
<th>DPIR (DRUNet)</th>
<th>DRP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Set3c</td>
<td>30.66/0.932/0.094</td>
<td>29.53/0.930/0.110</td>
<td>30.69/0.940/0.056</td>
</tr>
<tr>
<td></td>
<td>Set5</td>
<td>32.35/0.888/0.161</td>
<td>32.38/0.901/0.152</td>
<td>32.79/0.902/0.109</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>27.81/0.769/0.360</td>
<td>28.86/0.836/0.307</td>
<td>29.10/0.837/0.246</td>
</tr>
<tr>
<td></td>
<td>McMaster</td>
<td>31.10/0.859/0.154</td>
<td>32.42/0.900/0.151</td>
<td>32.79/0.901/0.107</td>
</tr>
<tr>
<td></td>
<td>Set3c</td>
<td>27.77/0.896/0.143</td>
<td>27.52/0.901/0.159</td>
<td>27.89/0.908/0.088</td>
</tr>
<tr>
<td></td>
<td>Set5</td>
<td>30.85/0.866/0.185</td>
<td>30.94/0.880/0.188</td>
<td>31.04/0.875/0.143</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>27.29/0.764/0.378</td>
<td>27.52/0.787/0.376</td>
<td>27.46/0.773/0.296</td>
</tr>
<tr>
<td></td>
<td>McMaster</td>
<td>30.69/0.859/0.189</td>
<td>30.78/0.871/0.193</td>
<td>30.79/0.865/0.141</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>SISR</th>
<th>Kernel</th>
<th>Datasets</th>
<th>DPIR(SwinIR)</th>
<th>DPIR(DRUNet)</th>
<th>DRP (SR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2×</td>
<td>Set3c</td>
<td>28.90/0.911/0.099</td>
<td>28.18/0.908/0.137</td>
<td>29.26/0.915/0.079</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Set5</td>
<td>31.38/0.872/0.122</td>
<td>31.42/0.883/0.175</td>
<td>31.47/0.876/0.120</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>27.92/0.783/0.349</td>
<td>27.97/0.801/0.356</td>
<td>28.12/0.799/0.285</td>
<td></td>
</tr>
<tr>
<td></td>
<td>McMaster</td>
<td>31.34/0.870/0.127</td>
<td>31.16/0.875/0.179</td>
<td>31.39/0.873/0.122</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Set3c</td>
<td>26.75/0.877/0.147</td>
<td>26.80/0.884/0.171</td>
<td>27.41/0.891/0.107</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Set5</td>
<td>30.23/0.852/0.240</td>
<td>30.36/0.862/0.241</td>
<td>30.42/0.859/0.163</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>26.83/0.739/0.413</td>
<td>26.98/0.757/0.401</td>
<td>26.98/0.745/0.327</td>
<td></td>
</tr>
<tr>
<td></td>
<td>McMaster</td>
<td>29.96/0.842/0.207</td>
<td>29.87/0.844/0.215</td>
<td>30.03/0.845/0.167</td>
<td></td>
</tr>
<tr>
<td>3×</td>
<td>Set3c</td>
<td>26.81/0.883/0.136</td>
<td>26.64/0.883/0.169</td>
<td>27.77/0.898/0.090</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Set5</td>
<td>30.35/0.855/0.159</td>
<td>30.39/0.863/0.202</td>
<td>30.83/0.868/0.149</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>26.79/0.742/0.352</td>
<td>26.80/0.755/0.394</td>
<td>27.18/0.763/0.299</td>
<td></td>
</tr>
<tr>
<td></td>
<td>McMaster</td>
<td>29.94/0.845/0.193</td>
<td>29.82/0.848/0.210</td>
<td>29.92/0.841/0.164</td>
<td></td>
</tr>
</tbody>
</table>

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 how DRP can outperform SOTA PnP methods that use denoisers 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 \( \sigma \) 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 without DRP does not work.

B.8.2 Single Image Super Resolution Using Deblurring Prior

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

\[
\text{sprox}_{\gamma g}(z) = \left( K^T S^T SK + \gamma H^T H \right)^{-1} \left[ K^T S^T y + \gamma H^T Hz \right],
\]

where \( H \) is the convolution operator with blur kernel \( k \). Similarly to deblurring in Section 5.3, we use CG to efficiently compute (11). We adjust the hyper-parameter \( \gamma \), \( \tau \) 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.

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Datasets</th>
<th>Deblur</th>
<th>DPIR</th>
<th>DRP (SR)</th>
<th>DRP (Deblur)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2×</td>
<td>Set3c</td>
<td>13.63/0.300/0.640</td>
<td>28.18/0.908/0.137</td>
<td>29.26/0.915/0.079</td>
<td>29.13/0.921/0.082</td>
</tr>
<tr>
<td></td>
<td>Set5</td>
<td>14.10/0.172/0.814</td>
<td>31.42/0.883/0.175</td>
<td>31.47/0.876/0.120</td>
<td>31.81/0.883/0.127</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>13.69/0.152/0.849</td>
<td>27.97/0.801/0.356</td>
<td>28.12/0.799/0.285</td>
<td>28.27/0.704/0.283</td>
</tr>
<tr>
<td></td>
<td>McMaster</td>
<td>14.31/0.158/0.844</td>
<td>31.16/0.875/0.179</td>
<td>31.39/0.873/0.122</td>
<td>31.52/0.881/0.129</td>
</tr>
<tr>
<td>2×</td>
<td>Set3c</td>
<td>7.85/0.090/0.956</td>
<td>26.80/0.884/0.171</td>
<td>27.41/0.891/0.107</td>
<td>27.52/0.893/0.101</td>
</tr>
<tr>
<td></td>
<td>Set5</td>
<td>8.77/0.052/1.000</td>
<td>30.36/0.862/0.241</td>
<td>30.42/0.859/0.163</td>
<td>30.73/0.864/0.175</td>
</tr>
<tr>
<td></td>
<td>CBSD68</td>
<td>8.58/0.040/1.000</td>
<td>26.98/0.757/0.401</td>
<td>26.98/0.745/0.327</td>
<td>27.19/0.755/0.338</td>
</tr>
<tr>
<td></td>
<td>McMaster</td>
<td>9.26/0.048/1.000</td>
<td>29.87/0.844/0.215</td>
<td>30.03/0.845/0.167</td>
<td>30.16/0.849/0.179</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Seed ID</th>
<th>Kernel</th>
<th>DPIR</th>
<th>DRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID = 0</td>
<td></td>
<td>29.53/0.930/0.110</td>
<td><strong>30.69/0.940/0.056</strong></td>
</tr>
<tr>
<td>ID = 10</td>
<td></td>
<td>29.50/0.930/0.113</td>
<td><strong>30.65/0.940/0.056</strong></td>
</tr>
<tr>
<td>ID = 100</td>
<td></td>
<td>29.53/0.930/0.112</td>
<td><strong>30.67/0.941/0.058</strong></td>
</tr>
<tr>
<td>ID = 0</td>
<td></td>
<td>27.52/0.901/0.159</td>
<td><strong>27.89/0.908/0.088</strong></td>
</tr>
<tr>
<td>ID = 10</td>
<td></td>
<td>27.51/0.901/0.156</td>
<td><strong>27.85/0.907/0.085</strong></td>
</tr>
<tr>
<td>ID = 100</td>
<td></td>
<td>27.52/0.900/0.158</td>
<td><strong>27.88/0.909/0.086</strong></td>
</tr>
</tbody>
</table>

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 $\ell_1$-norm or SSIM.

We retrained the SwinIR model for bicubic SR using the same setting as in Section 5.1, but with the $\ell_1$-norm loss instead of $\ell_2$. 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.

<table>
<thead>
<tr>
<th>Kernel</th>
<th>DPIR</th>
<th>DRP (SR$_{\ell_1}$)</th>
<th>DRP (SR$_{\ell_2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>29.53/0.930/0.110</td>
<td><strong>30.67/0.941/0.052</strong></td>
<td><strong>30.69/0.940/0.056</strong></td>
</tr>
<tr>
<td></td>
<td>27.52/0.901/0.159</td>
<td><strong>27.85/0.909/0.086</strong></td>
<td><strong>27.89/0.908/0.088</strong></td>
</tr>
</tbody>
</table>

Table 13: PSNR↑/SSIM↑/LPIPS↓ performance of DRP(SR$_{\ell_1}$), DRP(SR$_{\ell_2}$) and DPIR for image deblurring on the Set3c dataset using two blur kernels. The best and second best results are highlighted.