Generative Diffusion Prior for Unified Image Restoration and Enhancement

Ben Fei\textsuperscript{1,2,*}, Zhaoyang Lyu\textsuperscript{2,*}, Liang Pan\textsuperscript{3}, Junzhe Zhang\textsuperscript{3}, Weidong Yang\textsuperscript{1,†}, Tianyue Luo\textsuperscript{1}, Bo Zhang\textsuperscript{2}, Bo Dai\textsuperscript{2,†}

\textsuperscript{1}Fudan University, \textsuperscript{2}Shanghai AI Laboratory, \textsuperscript{3}S-Lab, Nanyang Technological University

bfei21@fudan.edu.cn, wdyang@fudan.edu.cn, (lvzhaoyang,daibo)@pjlab.org.cn

Figure 1. Generative Diffusion Prior (GDP) is capable of generating high-fidelity restoration across various tasks. GDP gives faithful image recovery on (a) linear and multi-linear restoration. In addition, GDP also enables novel applications of (b) blind, non-linear, multiple-guidance, or any-size image, including low-light enhancement and HDR recovery.

Abstract

Existing image restoration methods mostly leverage the posterior distribution of natural images. However, they often assume known degradation and also require supervised training, which restricts their adaptation to complex real applications. In this work, we propose the Generative Diffusion Prior (GDP) to effectively model the posterior distributions in an unsupervised sampling manner. GDP utilizes a pre-train denoising diffusion generative model (DDPM) for solving linear inverse, non-linear, or blind problems. Specifically, GDP systematically explores a protocol of conditional guidance, which is verified more practical than the commonly used guidance way. Furthermore, GDP is strength at optimizing the parameters of degradation model during the denoising process, achieving blind image restoration. Besides, we devise hierarchical guidance and patch-based methods, enabling the GDP to generate images of arbitrary resolutions. Experimentally, we demonstrate GDP ‘s versatility on several image datasets for linear problems, such as super-resolution, deblurring, inpainting, and colorization, as well as non-linear and blind issues, such as low-light enhancement and HDR image recovery. GDP outperforms the current leading unsupervised methods on the diverse benchmarks in reconstruction quality and perceptual quality. Moreover, GDP also generalizes well for natural images or synthesized images with arbitrary sizes from various tasks out of the distribution of the ImageNet training set. The project page is available at https://generativediffusionprior.github.io/

1. Introduction

Image quality often degrades during capture, storage, transmission, and rendering. Image restoration and enhancement [42] aim to inverse the degradation and improve the image quality. Typically, restoration and enhancement tasks can be divided into two main categories: 1) Linear inverse problems, such as image super-resolution (SR) [24, 38], deblurring [36, 75], inpainting [88], colorization [37, 94], where the degradation model is usually linear and known; 2) Non-linear or blind problems [1], such as image low-light enhancement [39] and HDR image recovery [10, 79], where the degradation model is non-linear and unknown. For a specific linear degradation model, image restoration can be tackled through end-to-end supervised training of neural networks [16, 94]. Nonetheless, corrupted images in the real world often have multiple complex degradations [57], where fully supervised approaches suffer to
There is a surge of interest to seek for more general image priors through generative models [1, 21, 69], and tackle image restoration in an unsupervised setting [8, 19], where multiple restoration tasks of different degradation models can be addressed during inference without re-training. For instance, Generative Adversarial Networks (GANs) [20] that are trained on a large dataset of clean images learn rich knowledge of the real-world scenes have succeeded in various linear inverse problems through GAN inversion [21, 51, 58]. In parallel, Denoising Diffusion Probabilistic Models (DDPMs) [2, 7, 35, 67, 72, 77] have demonstrated impressive generative capabilities, level of details, and diversity on top of GAN [26, 61, 62, 71, 73, 76]. As an early attempt, Kawar et al. [31] explore pre-trained DDPMs with variational inference, and achieve satisfactory results on multiple restoration tasks, but their Denoising Diffusion Restoration Model (DDRM) leverages the singular value decomposition (SVD) on a known linear degradation matrix, making it still limited to linear inverse problems.

In this study, we take a step further and propose an efficient approach named Generative Diffusion Prior (GDP). It exploits a well-trained DDPM as effective prior for general-purpose image restoration and enhancement, using degraded image as guidance. As a unified framework, GDP not only works on various linear inverse problems, but also generalizes to non-linear, and blind image restoration and enhancement tasks for the first time. However, solving the blind inverse problem is not trivial, as one would need to concurrently estimate the degradation model and recover the clean image with high fidelity. Thanks to the generative prior in a pre-trained DDPM, denoising within the DDPM manifold naturally regularizes the realness and fidelity of the recovered image. Therefore, we adopt a blind degradation estimation strategy, where the degradation model parameters of GDP are randomly initialized and optimized during the denoising process. Moreover, to further improve the photorealism and image quality, we systematically investigate an effective way to guide the diffusion models. Specifically, in the sampling process, the pre-trained DDPM first predicts a clean image \( \hat{x}_0 \) from the noisy image \( x_t \) by estimating the noise in \( x_t \). We can add guidance on this intermediate variable \( \hat{x}_0 \) to control the generation process of the DDPMs. In addition, with the help of the proposed hierarchical guidance and patch-based generation strategy, GDP is able to recover images of arbitrary resolutions, where low-resolution images and degradation models are first predicted to guide the generation of high-resolution images. We demonstrate the empirical effectiveness of GDP by comparing it with various competitive unsupervised methods under the linear or multi-linear inverse problem on ImageNet [14], LSUN [89], and CelebA [30] datasets in terms of consistency and FID. Over the low-light [39] and NTIRE [59] datasets, we further show GDP results on non-linear and blind issues, including low-light enhancement and HDR recovery, superior to other zero-shot baselines both qualitatively and quantitively, manifesting that GDP trained on ImageNet also works on images out of its training set distribution.

Our contributions are fourfold: (1) To our best knowledge, GDP is the first unified problem solver that can effectively use a single unconditional DDPM pre-trained on ImageNet provide by [15] to produce diverse and high-fidelity outputs for unified image restoration and enhancement in an unsupervised manner. (2) GDP is capable of optimizing randomly initiated parameters of degradation that are unknown, resulting in a powerful framework that can tackle any blind image restoration. (3) Further, to achieve arbitrary size image generation, we propose hierarchical guidance and patch-based methods, greatly promoting GDP on natural image enhancement. (4) Moreover, the comprehensive experiments are carried out, different from the conventional guidance way, where GDP directly predicts the temporary output given the noisy image in every step, which will be leveraged to guide the generation of images in the next step.

2. Related works

Linear Inverse Image Restoration. Most diffusion models toward linear inverse problems have employed unconditional models for the conditional tasks [50, 74], where only one model needs to be trained. However, unconditional tasks tend to be more difficult than conditional tasks. Moreover, the multi-linear task is also a relatively under-explored subject in image restoration. For instance, [60, 90] train simultaneously on multiple tasks, but they mainly concentrate on the enhancement tasks like deblurring and so on. Some works have also handled the multi-scale super-resolution by simultaneous training over multiple degradations [34]. Here, we propose GDP as a single model for dealing with single linear inverse or multiple linear inverse tasks.

Non-linear Image Restoration. The non-linear image formation model provides an accurate description of several imaging systems, including camera response functions in high-dynamic-range imaging [63]. The non-linear image restoration model is more accurate but is often more computationally intractable. Recently, great attention has been paid to non-linear image restoration problems. For example, HDR-GAN [56] is proposed for synthesizing HDR images from multi-exposed LDR images, while Enlighten-GAN [28] is devised as an unsupervised GAN to generalize very well on various real-world test images. The diffusion models are rarely studied for non-linear image restoration.

Blind Image Restoration. Early supervised attempts [5, 25] tend to estimate the unknown point spread function. As an example, [33] designs a class of structured denoisers, and...
is a Markov chain that gradually corrupts data $x_t$ into simple noise distribution $x_T \sim p_{data}$ and recover data from noise, where $N$ is the Gaussian distribution. DDPMs mainly comprise the diffusion process and the reverse process. The Diffusion Process is a Markov chain that gradually corrupts data $x_0$ until it approaches Gaussian noise $p_{data}$ at $T$ diffusion time steps. Corrupted data $x_1, \ldots, x_T$ are sampled from data $p_{data}$ with a diffusion process, which is defined as Gaussian transition:

$$q(x_1, \ldots, x_T | x_0) = \prod_{t=1}^{T} q(x_t | x_{t-1}),$$

where $t$ denotes as diffusion step, $q(x_t | x_{t-1}) = N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$, and $\beta_t$ are fixed or learned variance schedule. An important property of the forward noise process is that any step $x_t$ can be sampled directly from $x_0$ through the following equation:

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1-\alpha_t}\epsilon,$$

where $\epsilon \sim N(0, I), \alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$.

Proved by Ho et al. [26], there is a closed form expression for $q(x_t | x_0)$. We can obtain $q(x_T | x_0) = N(x_T; \sqrt{\alpha_T}x_0, (1 - \bar{\alpha}_t)I)$. Herein, $\alpha_t$ goes to 0 with large $T$, and $q(x_T | x_0)$ is close to the latent distribution $p_{data}$. The Reverse Process is a Markov chain that iteratively denoises a sampled Gaussian noise to a clean image. Starting from noise $x_T \sim N(0, I)$, the reverse process from latent $x_T$ to clean data $x_0$ is defined as:

$$p_\theta(x_0, \ldots, x_{T-1} | x_T) = \prod_{t=1}^{T} p_\theta(x_{t-1} | x_t),$$

$$p_\theta(x_{t-1} | x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta I)$$

According to Ho et al. [26], the mean $\mu_\theta(x_t, t)$ is the target we want to estimate by a neural network $\theta$. The variance $\Sigma_\theta$ can be either time-dependent constants [26] or learnable parameters [55]. $\epsilon_\theta$ is a function approximator intended to predict $\epsilon$ from $x_t$ as follow:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{\beta_t}{\sqrt{1-\alpha_t}} \epsilon_\theta(x_t, t) \right)$$

In practice, $\tilde{x}_0$ is usually predicted from $x_t$, then $x_{t-1}$ is sampled using both $\tilde{x}_0$ and $x_t$ computed as:

$$\tilde{x}_0 = \frac{x_t}{\sqrt{\alpha_t}} - \sqrt{1-\alpha_t} \epsilon_\theta(x_t, t) \sqrt{\alpha_t}$$

$$q(x_{t-1} | x_t, \tilde{x}_0) = N(\tilde{x}_{t-1}; \tilde{\mu}_t(x_t, \tilde{x}_0), \tilde{\beta}_t I),$$

where $\tilde{\mu}_t(x_t, \tilde{x}_0) = \frac{\sqrt{\alpha_t-\beta_t}}{1-\alpha_t} \tilde{x}_0 + \frac{\sqrt{\alpha_t}}{1-\alpha_t} x_t$

and $\tilde{\beta}_t = \frac{1 - \alpha_{t-1}}{1 - \alpha_t} \beta_t$.

4. Generative Diffusion Prior

In this study, we aim to exploit a well-trained DDPM as an effective prior for unified image restoration and enhancement, in particular, to handle degraded images of a
wide range of varieties. In detail, assume degraded image \( y \) is captured via \( y = D(x) \), where \( x \) is the original natural image, and \( D \) is a degradation model. We employ statistics of \( x \) stored in some prior and search in the space of \( x \) for an optimal \( x \) that best matches \( y \), regarding \( y \) as corrupted observations of \( x \). Due to the limited GAN inversion performance and the restricted applications of previous works [31, 32, 58, 64] in Table 1, in this paper, we focus on studying a more generic image prior, i.e., the diffusion models trained on large-scale natural images for image synthesis. Inspired by the [4, 9, 12, 65, 68], the reverse denoising process of the DDPM can be conditioned on the guidance synthesis. Inspired by the [4, 9, 12, 65, 68], the reverse denoising process of the DDPM can be conditioned on the degradation model: Conditioner guided diffusion sampling on \( x_t \), given a diffusion model \((\mu_t(\cdot), \Sigma_t(\cdot))\), corrupted image conditioner \( y \).

### Algorithm 1: GDP-\( x_t \) with fixed degradation model: Conditioner guided diffusion sampling on \( x_t \), given a diffusion model \((\mu_t(\cdot), \Sigma_t(\cdot))\), corrupted image conditioner \( y \).

**Input:** Corrupted image \( y \), gradient scale \( s \), degradation model \( D \), distance measure \( L \), optional quality enhancement loss \( Q \), quality enhancement scale \( \lambda \).

**Output:** Output image \( x_0 \) conditioned on \( y \)

```
Sample \( x_{0} \) from \( N(0, I) \)
end
```

### Algorithm 2: GDP-\( x_0 \): Conditioner guided diffusion sampling on \( x_0 \), given a diffusion model \((\mu_t(\cdot), \Sigma_t(\cdot))\), corrupted image conditioner \( y \).

**Input:** Corrupted image \( y \), gradient scale \( s \), degradation model \( D_0 \) with randomly initialized parameters \( \phi \), learning rate \( l \) for optimizable degradation model, distance measure \( L \), optional quality enhancement loss \( Q \), quality enhancement scale \( \lambda \).

**Output:** Output image \( x_0 \) conditioned on \( y \)

```
Sample \( x_{0} \) from \( N(0, I) \)
end
```

### 4.1. Single Image Guidance

The super-resolution, inpainting, colorization, deblurring, and enlightening tasks use single-image guidance. The Influence of Variance \( \Sigma \) on the Guidance. In previous conditional diffusion models [15, 78], the variance \( \Sigma \) is applied for the mean shift in the sampling process, which is theoretically proved in the Appendix. In our work, we find that the variance \( \Sigma \) might exert a negative influence on the quality of the generated images in our experiments. Therefore, we remove the variance during the guided denoising process to improve our performance. With the absence of \( \Sigma \) and the fixed guidance scale \( s \), the guided denoising process can be controlled by the variable scale \( \tilde{s} \).

**Guidance on \( x_t \).** Further, as vividly shown in Fig. 2b, Algo. 1 and Algo. 1 in Appendix, this class of guided dif-

---

### Table 1. Comparison of different generative priors and regularization priors for image restoration and enhancement.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>GAN</td>
<td>MMSE</td>
<td>Laplacian-based</td>
<td>DDPM</td>
<td>DDPM</td>
</tr>
<tr>
<td>Linear</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Non-linear</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Blind</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### unconditioned distribution by \(-s\Sigma \nabla_x \mathcal{L}(\mathcal{D}(x_t), y) + \lambda s \nabla_x \mathcal{Q}(x_t)\). However, we find that the way of adding guidance [3] and the variance \( \Sigma \) negatively influence the reconstructed images.
fusion models is the commonly used one [15, 46, 84], where the guidance is conditioned on $x_t$, but with the absence of $\Sigma$, named GDP-$x_t$. However, this variant that applies the guidance on $x_t$ may still yield less satisfactory quality images. The intuition is $x_t$ is a noisy version with a specific noise magnitude, but $y$ is in general a corrupted image with no noise or noises of different magnitude. We lack reliable ways to define the distance between $x_t$ and $y$. A naive MSE loss or perceptual loss will make $x_t$ deviate from its original noise magnitude and result in low-quality image generation.

**Guidance on $\tilde{x}_0$.** To tackle the problem as mentioned above, we systematically study the conditional signal applied on $\tilde{x}_0$. Details, in the sampling process, the pre-trained DDPM usually first predicts a clean image $\tilde{x}_0$ from the noisy image $x_t$, by estimating the noise in $x_t$, which can be directly inferred when given $x_t$ by the Eq. 6 in every timestep $t$. Then the predicted $\tilde{x}_0$ together with $x_t$ are utilized to sample the next step latent $x_{t-1}$. We can add guidance on this intermediate variable $\tilde{x}_0$ to control the generation process of the DDPM. The detailed sampling process can be found in Fig. 2c and Algo. 2, where there is only one corrupted image.

**Known Degradation.** Several tasks [24, 36, 37, 88] can be categorized into the class that the degradation function is known. In detail, the degradation model for image deblurring and super-resolution can be formulated as $y = (x \otimes k) \downarrow_s$. It assumes the low-resolution (LR) image is obtained by first convolving the high-resolution (HR) image with a Gaussian kernel (or point spread function) $k$ to get a blurry image $x \otimes k$, followed by a down-sampling operation $\downarrow_s$ with scale factor $s$. The goal of image inpainting is to recover the missing pixels of an image. The corresponding degradation transform is to multiply the original image with a binary mask $m$: $\psi(x) = x \otimes m$, where $\otimes$ is Hadamard’s product. Further, image colorization aims at restoring a gray-scale image $y \in \mathbb{R}^{H \times W}$ to a colorful image with RGB channels $x \in \mathbb{R}^{3 \times H \times W}$. To obtain $y$ from the colorful image $x$, the degradation transform $\psi$ is a graying transform that only preserves the brightness of $x$.

**Unknown Degradation.** In the real world, many images undergo complicated degradations [92], where the degradation models or the parameters of degradation models are unknown [43, 81]. In this case, the original images and the parameters of degradation models should be estimated simultaneously. For instance, in our work, the low-light image enhancement and the HDR recovery can be regarded as tasks with unknown degradation models. Here, we devise a simple but effective degradation model to simulate the complicated degradation, which can be formulated as follows:

$$y = f x + \mathcal{M}, \quad (10)$$

where the light factor $f$ is a scalar and the light mask $\mathcal{M}$ is a vector of the same dimension as $x$. $f$ and $\mathcal{M}$ are unknown parameters of the degradation model. The reason that we can use this simple degradation model is that the transform between any pair of corrupted images and the corresponding high-quality image can be captured by $f$ and $\mathcal{M}$ as long as they have the same size. If they do not have the same size, we can first resize $x$ to the same size as $y$ and then apply this transform. It is worth noting that this degradation model is non-linear in general, since $f$ and $\mathcal{M}$ depend on $x$ and $y$. We need to estimate $f$ and $\mathcal{M}$ for every individual corrupted image. We achieve this by randomly initializing them and synchronously optimizing them in the reverse process of DDPMs as shown in Algo. 2.

**4.2. Extended version**

**Multi-images Guidance.** Under specific circumstances, there are several images could be utilized to guide the generation of a single image [56, 98], which is merely studied and much more challenging than single-image guidance. To this end, we propose the HDR-GDP for the HDR image recovery with multiple images as guidance, consisting of three input LDR images, i.e. short, medium, and long exposures. Similar to low-light enhancement, the degradation models are also treated as Eq. 10, where the parameters remain unknown that determine the HDR recovery is the blind problem. However, as shown in Fig. 2c and Algo. 3 in Appendix, in the reverse process, there are three corrupted images ($n = 3$) to guide the generation so that three pairs of blind parameters for three LDR images are randomly initiated and optimized.

**Restore Any-size Image.** Furthermore, the pre-trained diffusion models provide by [15] with the size of 256 are only able to generate the fixed size of images, while the sizes of images from various image restoration are diverse. Herein, we employ the patch-based method as [45] to tackle this problem. By the merits of this patch-based strategy (Fig. 6 and Algo. 4 in the Appendix), GDP can be extended to recover the images of arbitrary resolution to promote the versatility of the GDP.
Table 2. Quantitative comparison of linear image restoration tasks on ImageNet 1k [58]. GDP outperforms other methods in terms of FID and Consistency across all tasks.

<table>
<thead>
<tr>
<th>Method</th>
<th>4× Super-resolution PSNR↑ SSIM↑ Consistency↓ FID↓</th>
<th>Deblur PSNR↑ SSIM↑ Consistency↓ FID↓</th>
<th>25% Inpainting PSNR↑ SSIM↑ Consistency↓ FID↓</th>
<th>Colorization PSNR↑ SSIM↑ Consistency↓ FID↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP [58]</td>
<td>21.65 0.56 158.74 152.85</td>
<td>26.00 0.54 475.10 136.53</td>
<td>27.59 0.82 414.60 60.65</td>
<td>18.42 0.71 305.59 94.59</td>
</tr>
<tr>
<td>SNIPS [52]</td>
<td>22.38 0.66 21.38 154.43</td>
<td>24.73 0.69 60.11 17.11</td>
<td>17.55 0.74 587.90 103.50</td>
<td>-</td>
</tr>
<tr>
<td>RED [64]</td>
<td>24.18 0.71 27.57 98.30</td>
<td>21.30 0.58 63.20 69.55</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DDRM [31]</td>
<td>26.53 0.78 19.39 40.75</td>
<td>35.64 0.98 50.24 4.78</td>
<td>34.28 0.95 4.08 24.09</td>
<td>22.12 0.91 37.33 47.05</td>
</tr>
<tr>
<td>GDP-x0</td>
<td>24.27 0.67 80.32 64.67</td>
<td>23.86 0.75 54.08 5.00</td>
<td>31.06 0.93 8.80 30.24</td>
<td>21.30 0.86 75.24 66.43</td>
</tr>
<tr>
<td>GDP-x1</td>
<td>24.42 0.68 6.49 38.24</td>
<td>25.98 0.75 41.27 2.44</td>
<td>34.40 0.96 5.29 16.58</td>
<td>21.41 0.92 36.92 37.60</td>
</tr>
</tbody>
</table>

Table 3. Quantitative comparison of image enlighten task on LOL [83], VE-LOL-L [45], and LoLi-phone [39] benchmarks. Bold font indicates the best performance in zero-shot learning, and the underlined font denotes the best results in all models.

<table>
<thead>
<tr>
<th>Learning</th>
<th>Methods</th>
<th>LOL [83]</th>
<th>VE-LOL-L [45]</th>
<th>LoLi-Phone [39]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised</td>
<td>LLNet [47]</td>
<td>17.91 0.76 169.20 384.21 4.10</td>
<td>17.38 0.73 124.98 291.59 5.54</td>
<td>343.34 5.36</td>
</tr>
<tr>
<td></td>
<td>LightenNet [41]</td>
<td>10.29 0.45 90.91 273.21 7.09</td>
<td>13.26 0.57 82.26 199.45 7.29</td>
<td>500.22 6.63</td>
</tr>
<tr>
<td></td>
<td>Retinex-Net [83]</td>
<td>17.24 0.55 129.99 513.28 8.63</td>
<td>16.41 0.64 135.20 421.41 8.62</td>
<td>542.29 8.23</td>
</tr>
<tr>
<td></td>
<td>MBBLE [49]</td>
<td>17.90 0.77 122.69 175.10 8.39</td>
<td>15.95 0.70 105.74 114.91 7.45</td>
<td>137.34 6.46</td>
</tr>
<tr>
<td></td>
<td>KinD [97]</td>
<td>17.57 0.82 74.52 377.59 7.41</td>
<td>18.07 0.78 80.12 253.79 7.51</td>
<td>265.47 6.84</td>
</tr>
<tr>
<td></td>
<td>KinD++ [93]</td>
<td>17.60 0.80 100.15 712.12 7.96</td>
<td>16.80 0.74 101.23 421.79 7.98</td>
<td>382.51 7.71</td>
</tr>
<tr>
<td></td>
<td>TBFEN [48]</td>
<td>17.25 0.82 90.59 367.66 8.29</td>
<td>18.91 0.81 91.30 276.65 8.02</td>
<td>214.30 7.34</td>
</tr>
<tr>
<td></td>
<td>DSLR [41]</td>
<td>14.98 0.67 183.92 272.68 7.09</td>
<td>15.70 0.68 124.80 271.63 7.27</td>
<td>281.25 6.99</td>
</tr>
<tr>
<td>Unsupervised</td>
<td>EnlightenGAN [28]</td>
<td>17.44 0.74 82.60 379.23 8.78</td>
<td>17.45 0.75 86.51 311.85 8.27</td>
<td>373.41 7.26</td>
</tr>
<tr>
<td>Self-supervised</td>
<td>DRBN [87]</td>
<td>15.15 0.52 94.96 692.99 5.53</td>
<td>18.47 0.78 88.10 268.70 6.15</td>
<td>285.06 5.31</td>
</tr>
<tr>
<td>Zero-shot</td>
<td>ExCNet [93]</td>
<td>16.04 0.62 111.18 220.38 8.70</td>
<td>16.20 0.66 115.24 225.15 8.62</td>
<td>359.96 7.95</td>
</tr>
<tr>
<td></td>
<td>Zero-DCE [23]</td>
<td>14.91 0.70 81.11 245.54 8.84</td>
<td>17.84 0.73 85.72 194.10 8.12</td>
<td>214.30 7.34</td>
</tr>
<tr>
<td></td>
<td>Zero-DCE++ [40]</td>
<td>14.86 0.62 86.22 302.06 7.08</td>
<td>16.12 0.45 86.96 313.50 7.92</td>
<td>308.15 7.18</td>
</tr>
<tr>
<td></td>
<td>RRDNet [99]</td>
<td>11.37 0.53 89.09 127.22 8.17</td>
<td>13.99 0.58 83.41 94.23 7.36</td>
<td>92.73 7.20</td>
</tr>
<tr>
<td></td>
<td>GDP-εx</td>
<td>7.32 0.57 238.92 364.15 8.26</td>
<td>9.45 0.50 152.68 194.49 7.12</td>
<td>508.73 8.06</td>
</tr>
<tr>
<td></td>
<td>GDP-x0</td>
<td>13.93 0.63 75.16 110.39 6.47</td>
<td>13.04 0.55 78.74 79.08 6.47</td>
<td>75.29 6.35</td>
</tr>
</tbody>
</table>

Figure 4. Qualitative results of (a) 25% inpainting and (b) 4× super-resolution on CelebA [30].

5. Loss Function

In GDP, the loss function can be divided into two main parts: Reconstruction loss and quality enhancement loss, where the former aims to recover the information contained in the conditional signal while the latter is integrated to promote the quality of the final outputs.

**Reconstruction Loss.** The reconstruction loss can be MSE, structural similarity index measure (SSIM), perceptual loss, or other reconstructive loss. Here, we primarily choose MSE loss as our reconstruction loss.

**Quality Enhancement Loss.**

1) **Exposure Control Loss:** To enhance the versatility of GDP, an exposure control loss $L_{exp}$ [23] is employed to control the exposure level for low-light image enhancement, which is written as:

$$L_{exp} = \frac{1}{U} \sum_{k=1}^{U} |I_k - E|,$$

where $U$ stands for the number of non-overlapping local regions of size $8 \times 8$, and $R$ represents the average intensity value of a local region in the reconstructed image. Follow-
In this section, we systematically compare GDP, which uses a single unconditional DDPM pre-trained on ImageNet, with other methods of various image restoration and enhancement tasks, and ablate the effectiveness of the proposed design. We furthermore list details on implementation, datasets, evaluation, and more qualitative results for all tasks in Appendix.

6.1. Linear and Multi-linear Degradation Tasks

Aiming at quantifying the performance of GDP, we focus on the ImageNet dataset for its diversity. For each experiment, we report the average peak signal-to-noise ratio (PSNR), SSIM, and Consistency to measure faithfulness to the original image and the FID to measure the resulting image quality. GDP is compared with other unsupervised methods that can operate on ImageNet, including RED [64], DGP [58], SNIPS [32], and DDRM [31]. We evaluate all methods on the tasks of 4x super-resolution, deblurring, inpainting, and colorization on one validation set from each of the 1000 ImageNet classes, following [58]. Table 2 shows that GDP-x0 outperforms other methods in Consistency and FID. The only exception is that DDRM achieves better PSNR and SSIM than GDP, but it requires higher Consistency and FID [11, 13, 15, 17, 27, 68]. GDP produces high-quality reconstructions across all the tested datasets and problems, which can be seen in Appendix. As a posterior sampling algorithm, GDP can produce multiple outputs for the same input, as demonstrated in the colorization task in Fig. 3. Moreover, the unconditional ImageNet DDPMs can be used to solve inverse problems on out-of-distribution images with general content. In Figs. 4 and 5, and more illustrations in Appendix, we show GDP successfully restores 256 \times 256 images from USC-SIPI [82], LSUN [89], and CelebA [30], which do not necessarily belong to any ImageNet class. GDP can also restore the images under multi-degradation (Fig. 1 and Appendix).
6.2. Exposure Correction Tasks

Encouraged by the excellent performance on the linear inverse problem, we further evaluate our GDP on the low-light image enhancement, which is categorized into non-linear and blind issues. Following the previous works [39], the three datasets LOL [83], VE-LOL-L [45], and the most challenging LoLi-phone [39] are leveraged to test the capability of GDP on low-light enhancement. As shown in Table 3, our GDP-x₀ fulfills the best FID, lightness order error (LOE) [80], and perceptual index (PI) [54] across all the zero-shot methods under three datasets. The lower LOE demonstrates better preservation for the naturalness of lightness, while the lower PI indicates better perceptual quality. In Fig. 6 and Appendix, our GDP-x₀ yields the most reasonable and satisfactory results across all methods. For more control, by the merits of Exposure Control Loss, the brightness of the generated images can be adjusted by the well-exposedness level $E$ (Fig. 1 and Appendix)

6.3. HDR Image Recovery

To evaluate our model on the HDR recovery [39], we compare HDR-GDP-x₀ with the state-of-the-art HDR methods on the test images in the HDR dataset from the NTIRE2021 Multi-Frame HDR Challenge [59], from which we randomly select 100 different scenes as the validation. Each scene consists of three LDR images with various exposures and corresponding HDR ground truth. The state-of-the-art methods used for comparison include AHDR-Net [86], HDR-GAN [56], DeepHDR [85] and deep-high-dynamic-range [29]. The quantitative results are provided in Table 4, where HDR-GDP-x₀ performs best in PSNR, SSIM, LPIPS, and FID. As shown in Fig. 7 and Appendix, HDR-GDP-x₀ achieves a better quality of reconstructed images, where the low-light parts can be enhanced, and the over-exposure regions are adjusted. Moreover, HDR-GDP-x₀ recovers the HDR images with more clear details.

6.4. Ablation Study

The Effectiveness of the Variance $\Sigma$ and the Guidance Protocol. The ablation studies on the variance $\Sigma$ and two ways of guidance are performed to unveil their effectiveness. As shown in Table 5, the performance of GDP-x₁ and GDP-x₀ is superior to GDP-x₀ with $\Sigma$ and GDP-x₀ with $\Sigma$, respectively, verifying the absence of variance $\Sigma$ can yield better quality of images. Moreover, the results of GDP-x₀ and GDP-x₀ with $\Sigma$ are better than GDP-x₀ and GDP-x₀ with $\Sigma$, respectively, demonstrating the superiority of the guidance on x₀ protocol.

The Effectiveness of the Trainable Degradation and the Patch-based Tactic. Moreover, to validate the influence of trainable parameters of the degradation model and our patch-based methods, further experiments are carried out on the LOL [83] and NTIRE [59] datasets. Model A is designed to naively restore the images from patches and patches where the parameters are not related. ModelB is designed with fixed parameters for all patches in the images. As shown in Table 6, our GDP-x₀ ranks first across all models and obtains the best visualization results (Fig. 7 and Appendix), revealing the strength of our proposed hierarchical guidance and patch-based method.

7. Conclusion

In this paper, we propose the Generative Diffusion Prior for unified image restoration that can be employed to tackle the linear inverse, non-linear and blind problems. Our GDP is able to restore any-size images via hierarchical guidance and patch-based methods. We systematically studied the way of guidance to exploit the strength of the DDPM. The GDP is comprehensively utilized on various tasks such as super-resolution, deblurring, inpainting, colorization, low-light enhancement, and HDR recovery, demonstrating the capabilities of GDP on unified image restoration.

Acknowledgement. This project is funded in part by Shanghai AI Laboratory.
References


[38] Chongyi Li, Chunle Guo, Ling-Hao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A survey. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 44(01):1–1, 2021. 1, 2, 6, 7, 8


