Generative Diffusion Prior for Unified Image Restoration and Enhancement

Ben Fei[1,2,*], Zhaoyang Lyu[2,*], Liang Pan[3], Junzhe Zhang[3],
Weidong Yang[1,†], Tianyue Luo[1], Bo Zhang[2], Bo Dai[2,†]
[1] Fudan University, [2]Shanghai AI Laboratory, [3]S-Lab, Nanyang Technological University
bfei21@m.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

∗Equal contribution, †Corresponding author.

generalize.

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 $\tilde{\boldsymbol{x}}_0$ from the noisy image $\boldsymbol{x}_t$ by estimating the noise in $\boldsymbol{x}_t$. We can add guidance on this intermediate variable $\tilde{\boldsymbol{x}}_0$ 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 EnlightenGAN [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

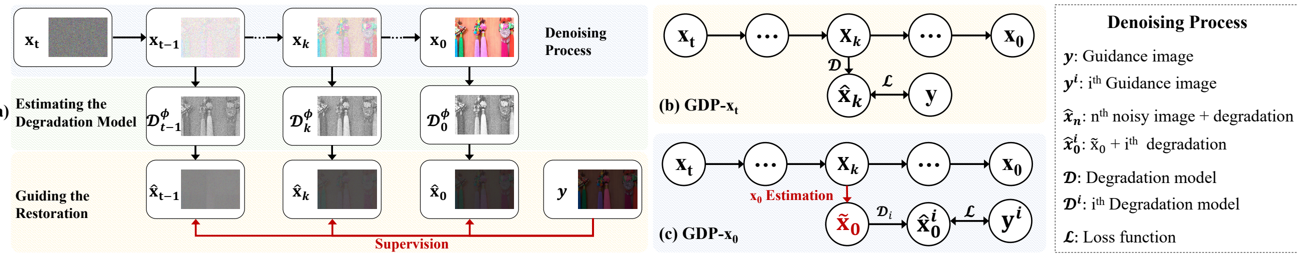

Figure 2. **Overview of our GDP for unified image recovery.** (a) Given a corrupted image $\boldsymbol{y}$ during inference, GDP systematically studies the reverse process from $\boldsymbol{x}_T$ to $\boldsymbol{x}_0$ guided by the $\boldsymbol{y}$. The guidance can be added on $\hat{\boldsymbol{x}}_0$ or $\hat{\boldsymbol{x}}_t$, leading to two variants of GDP. And $\hat{\boldsymbol{x}}_0$ and $\hat{\boldsymbol{x}}_t$ can be collectively expressed as $\hat{\boldsymbol{x}}_{\boldsymbol{t}}$. The supervision signal (Sec. 5) is applied between $\hat{\boldsymbol{x}}_{\boldsymbol{t}}$ and $\boldsymbol{y}$. GDP looks for an intermediate variable $\boldsymbol{x}_t$ and optimizes the degradation model $\{\mathcal{D}_\phi^i \mid i=1,2,\ldots,n\}$ that best reconstruct the image corresponding to $\boldsymbol{y}$ via gradient descent. Note that GDP is a generic image restoration method. We illustrate it with the low light enhancement example. (b) GDP-$\boldsymbol{x}_t$ adds guidance $\boldsymbol{x}_t$ in every time step (Algo. 1), while (c) GDP-$\boldsymbol{x}_0$ estimates the $\tilde{\boldsymbol{x}}_0$ given $\boldsymbol{x}_t$, then adds guidance on $\tilde{\boldsymbol{x}}_0$ to obtain $\hat{\boldsymbol{x}}_0$(Algo. 2). The number of guidance images $\{\boldsymbol{y}^i \mid i=1,2,\ldots,n\}$ and the degradation models $\{\mathcal{D}_\phi^i \mid i=1,2,\ldots,n\}$ are dependent on the tasks. For instance, $n=3$ for HDR recovery, while $n=1$ for other tasks.

[70] employs a fixed down-sampling operation to generate synthetic pairs during testing. However, these methods often are incapable of obtaining the parameters or distribution of the observed data due to the complicated degradation types. Another way to solve the blind image restoration is to utilize unsupervised learning methods [18, 100]. Following the CycleGAN [100], CinCGAN [91] and MCinCGAN [96] employ a pre-trained SR model together with cycle consistency loss to learn a mapping from the input image to high-quality image space. However, it still remains a challenge to exploit a unified architecture for blind image restoration. By the merits of the powerful GDP, these blind problems could also be solved by simultaneously estimating the recovered image and a specific degradation model.

3. Preliminary

Diffusion models [2, 22, 35, 66, 84] transform complex data distribution $\boldsymbol{x}_0 \sim p_{\text{data}}$ into simple noise distribution $\boldsymbol{x}_T \sim p_{\text{latent}} = \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and recover data from noise, where $\mathcal{N}$ 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 $\boldsymbol{x}_0$ until it approaches Gaussian noise p_{latent} at T diffusion time steps. Corrupted data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$ are sampled from data p_{data}, with a diffusion process, which is defined as Gaussian transition:

$$q\left(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_T \mid \boldsymbol{x}_0\right)=\prod_{t=1}^T q\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}\right), \tag{1}$$

where t denotes as diffusion step, $q\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}\right) = \mathcal{N}\left(\boldsymbol{x}_t ; \sqrt{1-\beta_t} \boldsymbol{x}_{t-1}, \beta_t \boldsymbol{I}\right)$, and β_t are fixed or learned variance schedule. An important property of the forward noising process is that any step $\boldsymbol{x}_t$ may be sampled directly from $\boldsymbol{x}_0$ through the following equation:

$$\boldsymbol{x}_t=\sqrt{\bar{\alpha}_t} \boldsymbol{x}_0+\sqrt{1-\bar{\alpha}_t} \boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I})$, $\alpha_t=1-\beta_t$ and $\bar{\alpha}_t=\prod_{i=1}^t \alpha_i$. Proved by Ho *et al.* [26], there is a closed form expression for $q\left(\boldsymbol{x}_t \mid \boldsymbol{x}_0\right)$. We can obtain $q\left(\boldsymbol{x}_t \mid \boldsymbol{x}_0\right) = \mathcal{N}\left(\boldsymbol{x}_t ; \sqrt{\bar{\alpha}_t} \boldsymbol{x}_0,\left(1-\bar{\alpha}_t\right) \boldsymbol{I}\right)$. Herein, $\bar{\alpha}_t$ goes to 0 with large T, and $q\left(\boldsymbol{x}_T \mid \boldsymbol{x}_0\right)$ is close to the latent distribution p_{latent}.

The Reverse Process is a Markov chain that iteratively denoises a sampled Gaussian noise to a clean image. Starting from noise $x_T \sim \mathcal{N}(0, \boldsymbol{I})$, the reverse process from latent $\boldsymbol{x}_T$ to clean data $\boldsymbol{x}_0$ is defined as:

$$\begin{aligned} & p_{\boldsymbol{\theta}}\left(\boldsymbol{x}_0, \cdots, \boldsymbol{x}_{T-1} \mid \boldsymbol{x}_T\right)=\prod_{t=1}^T p_{\boldsymbol{\theta}}\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t\right), \\ & p_{\boldsymbol{\theta}}\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t\right)=\mathcal{N}\left(\boldsymbol{x}_{t-1} ; \boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right), \Sigma_\theta \boldsymbol{I}\right) \end{aligned} \tag{3}$$

According to Ho *et al.* [26], the mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right)$ is the target we want to estimate by a neural network $\boldsymbol{\theta}$. The variance Σ_θ can be either time-dependent constants [26] or learnable parameters [55]. ϵ_θ is a function approximator intended to predict ϵ from $\boldsymbol{x}_t$ as follow:

$$\boldsymbol{\mu}_\theta\left(\boldsymbol{x}_t, t\right)=\frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta\left(\boldsymbol{x}_t, t\right)\right) \tag{4}$$

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

$$\tilde{\boldsymbol{x}}_0=\frac{\boldsymbol{x}_t}{\sqrt{\bar{\alpha}_t}}-\frac{\sqrt{1-\bar{\alpha}_t} \epsilon_\theta\left(\boldsymbol{x}_t, t\right)}{\sqrt{\bar{\alpha}_t}} \tag{5}$$

$$\begin{aligned} & q\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \tilde{\boldsymbol{x}}_0\right)=\mathcal{N}\left(\boldsymbol{x}_{t-1} ; \tilde{\boldsymbol{\mu}}_t\left(\boldsymbol{x}_t, \tilde{\boldsymbol{x}}_0\right), \tilde{\beta}_t \mathbf{I}\right), \\ \text{where} \quad & \tilde{\boldsymbol{\mu}}_t\left(\boldsymbol{x}_t, \tilde{\boldsymbol{x}}_0\right)=\frac{\sqrt{\bar{\alpha}_{t-1}} \beta_t}{1-\bar{\alpha}_t} \tilde{\boldsymbol{x}}_0+\frac{\sqrt{\alpha_t}\left(1-\bar{\alpha}_{t-1}\right)}{1-\bar{\alpha}_t} \boldsymbol{x}_t \\ \text{and} \quad & \tilde{\beta}_t=\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \beta_t \end{aligned} \tag{6}$$

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

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

Methods	DGP [58]	SNIPS [32]	RED [64]	DDRM [31]	GDP (Ours)
Prior	GAN	MMSE Gaussian denoiser	Laplacian-based regularization function	DDPM	DDPM
Linear	✓	✓	✓	✓	✓
Non-linear	✗	✗	✗	✗	✓
Blind	✗	✗	✗	✗	✓

wide range of varieties. In detail, assume degraded image $\boldsymbol{y}$ is captured via $\boldsymbol{y} = \mathcal{D}(\boldsymbol{x})$, where $\boldsymbol{x}$ is the original natural image, and $\mathcal{D}$ is a degradation model. We employ statistics of $\boldsymbol{x}$ stored in some prior and search in the space of $\boldsymbol{x}$ for an optimal $\boldsymbol{x}$ that best matches $\boldsymbol{y}$, regarding $\boldsymbol{y}$ as corrupted observations of $\boldsymbol{x}$. 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 degraded image $\boldsymbol{y}$. Specifically, the reverse denoising distribution $p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$ in Eq. 3 is adopted to a conditional distribution $p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{y})$. [15, 71] prove that

$$\begin{aligned} \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{y}) &= \log\left(p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)\, p(\boldsymbol{y}|\boldsymbol{x}_t)\right) + K_1 \\ &\approx \log p(\boldsymbol{r}) + K_2, \end{aligned} \tag{7}$$

where $\boldsymbol{r} \sim \mathcal{N}(\boldsymbol{r}; \boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) + \Sigma\boldsymbol{g}, \Sigma)$ and $\boldsymbol{g} = \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y} \mid \boldsymbol{x}_t)$, where $\Sigma = \Sigma_\theta(\boldsymbol{x}_t)$ for conciseness. K_1 and K_2 are constants, $p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)$ is defined by Eq. 3. $p(\boldsymbol{y} \mid \boldsymbol{x}_t)$ can be regarded as the probability that $\boldsymbol{x}_t$ will be denoised to a high-quality image consistent to $\boldsymbol{y}$. We propose a heuristic approximation of it:

$$p(\boldsymbol{y} \mid \boldsymbol{x}_t) = \frac{1}{Z} \exp\left(-\left[s\mathcal{L}(\mathcal{D}(\boldsymbol{x}_t), \boldsymbol{y}) + \lambda\mathcal{Q}(\boldsymbol{x}_t)\right]\right) \tag{8}$$

where $\mathcal{L}$ is some image distance metric, Z is a normalization factor, and s is a scaling factor controlling the magnitude of guidance. Intuitively, this definition encourages $\boldsymbol{x}_t$ to be consistent with the corrupted image $\boldsymbol{y}$ to obtain a high probability of $p(\boldsymbol{y} \mid \boldsymbol{x}_t)$. $\mathcal{Q}$ is the optional quality enhancement loss to enhance the flexibility of GDP, which can be used to control some properties (such as brightness) or enhance the quality of the denoised image. λ is the scale factor for adjusting the quality of images. The gradients of both sides are computed as:

$$\begin{aligned} &\log p(\boldsymbol{y} \mid \boldsymbol{x}_t) = -\log Z - s\mathcal{L}(\mathcal{D}(\boldsymbol{x}_t), \boldsymbol{y}) - \lambda\mathcal{Q}(\boldsymbol{x}_t) \\ &\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y} \mid \boldsymbol{x}_t) = -s\nabla_{\boldsymbol{x}_t}\mathcal{L}(\mathcal{D}(\boldsymbol{x}_t), \boldsymbol{y}) - \lambda\nabla_{\boldsymbol{x}_t}\mathcal{Q}(\boldsymbol{x}_t). \end{aligned} \tag{9}$$

where the distance metric $\mathcal{L}$ and the optional quality loss $\mathcal{Q}$ can be found in Sec. 5.

In this way, the conditional transition $p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y})$ can be approximately obtained through the unconditional transition $p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)$ by shifting the mean of the unconditional distribution by $-(s\Sigma\nabla_{\boldsymbol{x}_t}\mathcal{L}(\mathcal{D}(\boldsymbol{x}_t), \boldsymbol{y}) + \lambda\Sigma\nabla_{\boldsymbol{x}_t}\mathcal{Q}(\boldsymbol{x}_t))$ However, we find that the way of adding guidance [3] and the variance Σ negatively influence the reconstructed images.

Algorithm 1: GDP-$\boldsymbol{x}_t$ with fixed degradation model: Conditioner guided diffusion sampling on $\boldsymbol{x}_t$, given a diffusion model $(\mu_\theta(\boldsymbol{x}_t), \Sigma_\theta(\boldsymbol{x}_t))$, corrupted image conditioner $\boldsymbol{y}$.

Input: Corrupted image $\boldsymbol{y}$, gradient scale s, degradation model $\mathcal{D}$, distance measure $\mathcal{L}$, optional quality enhancement loss $\mathcal{Q}$, quality enhancement scale λ.
Output: Output image $\boldsymbol{x}_0$ conditioned on $\boldsymbol{y}$
Sample $\boldsymbol{x}_T$ from $\mathcal{N}(0, \mathbf{I})$
for *t from T to 1* **do**
  $\mu, \Sigma = \mu_\theta(\boldsymbol{x}_t), \Sigma_\theta(\boldsymbol{x}_t)$
  $\mathcal{L}^{total}_{\boldsymbol{x}_t} = \mathcal{L}(\boldsymbol{y}, \mathcal{D}(\boldsymbol{x}_t)) + \mathcal{Q}(\boldsymbol{x}_t)$
  Sample $\boldsymbol{x}_{t-1}$ by $\mathcal{N}\left(\mu + s\nabla_{\boldsymbol{x}_t}\mathcal{L}^{total}_{\boldsymbol{x}_t}, \Sigma\right)$
end
return $\boldsymbol{x}_0$

Algorithm 2: GDP-$\boldsymbol{x}_0$: Conditioner guided diffusion sampling on $\tilde{\boldsymbol{x}}_0$, given a diffusion model $(\mu_\theta(\boldsymbol{x}_t), \Sigma_\theta(\boldsymbol{x}_t))$, corrupted image conditioner $\boldsymbol{y}$.

Input: Corrupted image $\boldsymbol{y}$, gradient scale s, degradation model $\mathcal{D}_\phi$ with randomly initiated parameters ϕ, learning rate l for optimizable degradation model, distance measure $\mathcal{L}$, optional quality enhancement loss $\mathcal{Q}$, quality enhancement scale λ.
Output: Output image $\boldsymbol{x}_0$ conditioned on $\boldsymbol{y}$
Sample $\boldsymbol{x}_T$ from $\mathcal{N}(0, \mathbf{I})$
for *t from T to 1* **do**
  $\mu, \Sigma = \mu_\theta(\boldsymbol{x}_t), \Sigma_\theta(\boldsymbol{x}_t)$
  $\tilde{\boldsymbol{x}}_0 = \frac{\boldsymbol{x}_t}{\sqrt{\bar{\alpha}_t}} - \frac{\sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\boldsymbol{x}_t, t)}{\sqrt{\bar{\alpha}_t}}$
  $\mathcal{L}^{total}_{\phi, \tilde{\boldsymbol{x}}_0} = \mathcal{L}(\boldsymbol{y}, \mathcal{D}_\phi(\tilde{\boldsymbol{x}}_0)) + \mathcal{Q}(\tilde{\boldsymbol{x}}_0)$
  $\phi \leftarrow \phi - l\nabla_\phi\mathcal{L}^{total}_{\phi, \tilde{\boldsymbol{x}}_0}$
  Sample $\boldsymbol{x}_{t-1}$ by $\mathcal{N}\left(\mu + s\nabla_{\tilde{\boldsymbol{x}}_0}\mathcal{L}^{total}_{\phi, \tilde{\boldsymbol{x}}_0}, \Sigma\right)$
end
return $\boldsymbol{x}_0$

4.1. Single Image Guidance

The super-resolution, impainting, colorization, deblurring, and enlighting tasks use single-image guidance.

The Influence of Variance Σ on the Guidance. In previous conditional diffusion models [15, 78], the variance Σ 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 Σ 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 Σ and the fixed guidance scale s, the guided denoising process can be controlled by the variable scale $\hat{s}$.

Guidance on $\boldsymbol{x}_t$. Further, as vividly shown in Fig. 2b, Algo. 1 and Algo. 1 in Appendix, this class of guided diffusion models is the commonly used one [15, 46, 84], where the guidance is conditioned on $\boldsymbol{x}_t$ but with the absence of Σ, named GDP-x_t. However, this variant that applies the guidance on $\boldsymbol{x}_t$ may still yield less satisfactory quality images. The intuition is x_t is a noisy image 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 $\tilde{\boldsymbol{x}}_0$. Detailly, in the sampling process, the pretrained DDPM usually first predicts a clean image $\tilde{\boldsymbol{x}}_0$ from the noisy image $\boldsymbol{x}_t$ by estimating the noise in $\boldsymbol{x}_t$, which can be directly inferred when given $\boldsymbol{x}_t$ by the Eq. 6 in every timestep t. Then the predicted $\tilde{\boldsymbol{x}}_0$ together with $\boldsymbol{x}_t$ are utilized to sample the next step latent $\boldsymbol{x}_{t-1}$. We can add guidance on this intermediate variable $\tilde{\boldsymbol{x}}_0$ 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 $\boldsymbol{y} = (\boldsymbol{x} \otimes \mathbf{k}) \downarrow_{\mathbf{s}}$. 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) $\mathbf{k}$ to get a blurry image $\boldsymbol{x} \otimes \mathbf{k}$, followed by a down-sampling operation $\downarrow_{\mathrm{s}}$ with scale factor $\mathbf{s}$. 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 $\mathbf{m}$: $\psi(\boldsymbol{x}) = \boldsymbol{x} \odot \mathbf{m}$, where $\odot$ is Hadamard's product. Further, image colorization aims at restoring a gray-scale image $\boldsymbol{y} \in \mathbb{R}^{H \times W}$ to a colorful image with RGB channels $\boldsymbol{x} \in \mathbb{R}^{3 \times H \times W}$. To obtain $\boldsymbol{y}$ from the colorful image $\boldsymbol{x}$, the degradation transform ψ is a graying transform that only preserves the brightness of $\boldsymbol{x}$.

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:

$$\boldsymbol{y} = f\boldsymbol{x} + \boldsymbol{\mathcal{M}}, \tag{10}$$

where the light factor f is a scalar and the light mask $\boldsymbol{\mathcal{M}}$ is a vector of the same dimension as $\boldsymbol{x}$. f and $\boldsymbol{\mathcal{M}}$ 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 $\boldsymbol{\mathcal{M}}$ as long as they have the same size. If they do not have the same size, we can first resize $\boldsymbol{x}$ to the same size as $\boldsymbol{y}$ and then apply this transform. It is worth noting that this degradation model is non-linear in general, since f and $\boldsymbol{\mathcal{M}}$ depend on $\boldsymbol{x}$ and $\boldsymbol{y}$. We need to estimate f and $\boldsymbol{\mathcal{M}}$ 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.

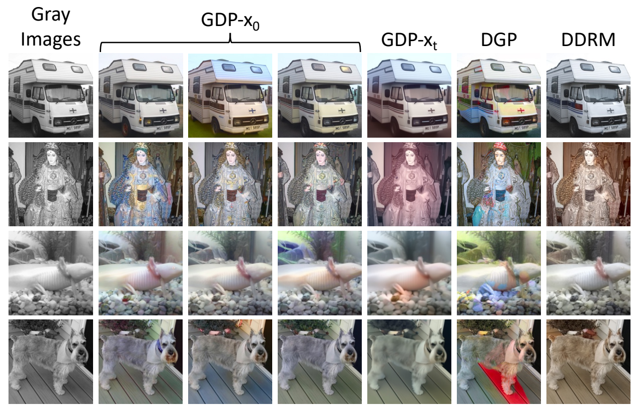

Figure 3. **Qualitative comparison of colorization results on ImageNet validation images**. GDP-x_0 generates various samples on the same input.

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.

Method	4× Super-resolution				Deblur				25% Inpainting				Colorization			
	PSNR ↑	SSIM ↑	Consistency ↓	FID ↓	PSNR ↑	SSIM ↑	Consistency↓	FID ↓	PSNR ↑	SSIM ↑	Consistency↓	FID ↓	PSNR ↑	SSIM ↑	Consistency ↓	FID ↓
DGP [58]	21.65	0.56	158.74	152.85	26.00	0.54	475.10	136.53	27.59	0.82	414.60	60.65	18.42	0.71	305.59	94.59
SNIPS [32]	22.38	0.66	21.38	154.43	24.73	0.69	60.11	17.11	17.55	0.74	587.90	103.50	-	-	-	-
RED [64]	24.18	0.71	27.57	98.30	21.30	0.58	63.20	69.55	-	-	-	-	-	-	-	-
DDRM [31]	**26.53**	**0.78**	19.39	40.75	**35.64**	**0.98**	50.24	4.78	34.28	0.95	**4.08**	24.09	**22.12**	0.91	37.33	47.05
GDP-x_t	24.27	0.67	80.32	64.67	25.86	0.75	54.08	5.00	31.06	0.93	8.80	20.24	21.30	0.86	75.24	66.43
GDP-x_0	24.42	0.68	**6.49**	**38.24**	25.98	0.75	**41.27**	**2.44**	**34.40**	**0.96**	5.29	**16.58**	21.41	**0.92**	**36.92**	**37.60**

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.

Learning	Methods	LOL [83]					VE-LOL-L [45]					LoLi-Phone [39]	
		PSNR ↑	SSIM↑	FID↓	LOE↓	PI↓	PSNR↑	SSIM↑	FID ↓	LOE↓	PI↓	LOE↓	PI↓
Supervised learning	LLNet [47]	17.91	0.76	169.20	384.21	4.10	17.38	0.73	124.98	291.59	5.54	343.34	5.36
	LightenNet [41]	10.29	0.45	90.91	273.21	7.09	13.26	0.57	82.26	199.45	7.29	500.22	6.63
	Retinex-Net [83]	17.24	0.55	129.99	513.28	8.63	16.41	0.64	135.20	421.41	8.62	542.29	8.23
	MBLLEN [49]	17.90	0.77	122.69	175.10	8.39	15.95	0.70	105.74	114.91	7.45	137.34	6.46
	KinD [97]	17.57	0.82	74.52	377.59	7.41	18.07	0.78	80.12	253.79	7.51	265.47	6.84
	KinD++ [95]	17.60	0.80	100.15	712.12	7.96	16.80	0.74	101.23	421.97	7.98	382.51	7.71
	TBFEN [48]	17.25	0.83	90.59	367.66	8.29	18.91	0.81	91.30	276.65	8.02	214.30	7.34
	DSLR [44]	14.98	0.67	183.92	272.68	7.09	15.70	0.68	124.80	271.63	7.27	281.25	6.99
Unsupervised learning	EnlightenGAN [28]	17.44	0.74	82.60	379.23	8.78	17.45	0.75	86.51	311.85	8.27	373.41	7.26
Self-supervised learning	DRBN [87]	15.15	0.52	94.96	692.99	5.53	18.47	0.78	88.10	268.70	6.15	285.06	5.31
Zero-shot learning	ExCNet [93]	**16.04**	0.62	111.18	220.38	8.70	16.20	0.66	115.24	225.15	8.62	359.96	7.95
	Zero-DCE [23]	14.91	**0.70**	81.11	245.54	8.84	**17.84**	**0.73**	85.72	194.10	8.12	214.30	7.34
	Zero-DCE++ [40]	14.86	0.62	86.22	302.06	7.08	16.12	0.45	86.96	313.50	7.92	308.15	7.18
	RRDNet [99]	11.37	0.53	89.09	127.22	8.17	13.99	0.58	83.41	94.23	7.36	92.73	7.20
	GDP-x_t	7.32	0.57	238.92	364.15	8.26	9.45	0.50	152.68	194.49	7.12	508.73	8.06
	GDP-x_0	13.93	0.63	**75.16**	**110.39**	**6.47**	13.04	0.55	**78.74**	**79.08**	**6.47**	**75.29**	**6.35**

(a) Inpainting (b) 4x super resolution

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,

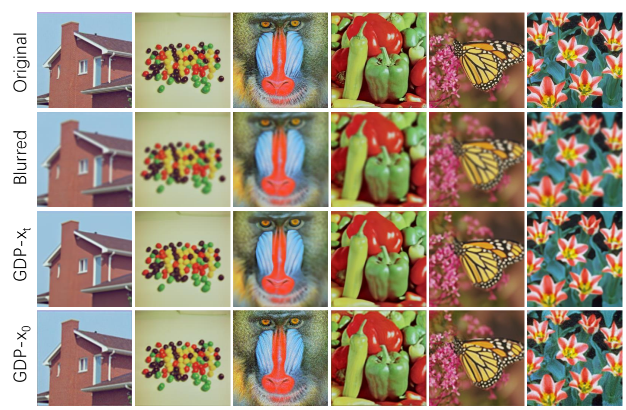

Figure 5. **Results of image deblurring task on 256 × 256 USC-SIPI images [82] using an ImageNet model.**

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:

$$L_{\text{exp}} = \frac{1}{U} \sum_{k=1}^{U} |R_k - E| \,, \tag{11}$$

where U stands for the number of non-overlapping local regions of size 8×8, and R represents the average intensity value of a local region in the reconstructed image. Follow-

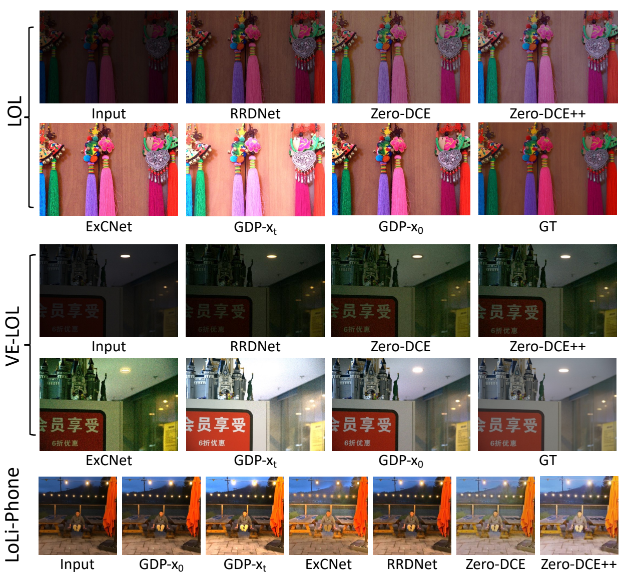

Figure 6. **Qualitative results of low-light enhancement on the LOL [83], VE-LOL [45], and LoLi-Phone [39] datasets.**

ing the previous works [52, 53], E is set as the gray level in the RGB color space. As expected, E can be adjusted to control the brightness in our experiments.

2) Color Constancy Loss: Following the Gray-World color constancy hypothesis [6], a color constancy loss L_{col} is exploited to correct the potential color deviations in the restored image and bridge the relations among the three adjusted channels in the colorization task, formulated as:

$$L_{\mathrm{col}} = \sum_{\forall(m,n)\in\varepsilon} (Y^m - Y^n)^2, \varepsilon = \{(R,G),(R,B),(G,B)\} \quad (12)$$

where Y^m is the average intensity value of m channel in the recovered image, (m,n) represents a pair of channels.

3) Illumination Smoothness Loss: To maintain the monotonicity relations between neighboring pixels in the optimized light mask $\mathcal{M}$, an illumination smoothness loss [23] is utilized for each light variance $\mathcal{M}$. The illumination smoothness loss $L_{tv_{\mathcal{M}}}$ is defined as:

$$L_{tv_{\mathcal{M}}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{c\in\xi}(|\nabla_h \mathcal{M}_n^c| + \nabla_v \mathcal{M}_n^c\ |)^2, \xi = \{R,G,B\}, \quad (13)$$

where N is iteration times, ∇_h and ∇_v are the horizontal and vertical gradient operations, respectively.

Specifically, the image colorization task uses color constancy loss to obtain more natural colors. The low-light enhancement requires color constancy loss for the same reason. In addition, the low-light enhancement task uses illumination smoothness loss to make the estimated light mask $\mathcal{M}$ smoother. Exposure control loss enables us to manually control the brightness of the restored image. The weights of the losses can be found in Appendix.

6. Experiments

In this section, we systematically compare GDP , which uses **a single unconditional DDPM pre-trained on ImageNet** provide by [15], 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.

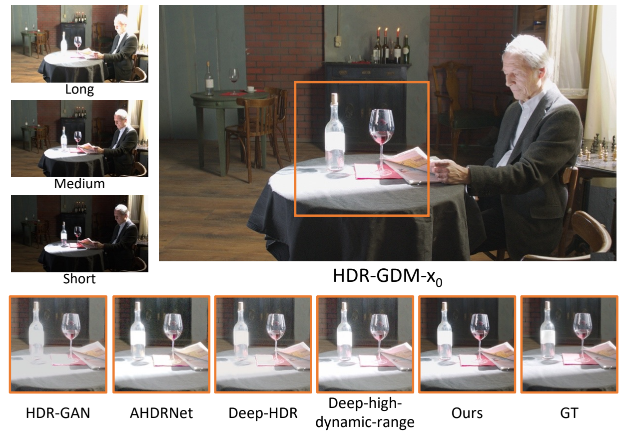

Figure 7. **Example from the NTIRE dataset [59].** We compare a set of patches cropped from the tone-mapped HDR images generated by state-of-the-art methods.

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 $4\times$ super-resolution, deblurring, impainting, and colorization on one validation set from each of the 1000 ImageNet classes, following [58]. Table 2 shows that GDP-x_0 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 256×256 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).

Table 4. **Quantitative comparison on the NTIRE dataset [59].**

Methods	PSNR↑	SSIM↑	LPIPS ↓	FID↓
AHDRNet [86]	18.72	0.58	0.39	81.98
HDR-GAN [56]	21.67	0.74	0.26	52.71
Deep-HDR [85]	21.66	0.76	0.26	57.52
Deep-high-dynamic-range [29]	21.33	0.71	0.26	51.92
GDP-x_t	19.36	0.65	0.30	63.89
GDP-x_0	**24.88**	**0.86**	**0.13**	**50.05**

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-x_0 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-x_0 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-x_0 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-x_0 performs best in PSNR, SSIM, LPIPS, and FID. As shown in Fig. 7 and Appendix, HDR-GDP-x_0 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-x_0 recovers the HDR images with more clear details.

6.4. Ablation Study

The Effectiveness of the Variance Σ and the Guidance Protocol. The ablation studies on the variance Σ and two ways of guidance are performed to unveil their effectiveness. As shown in Table. 5, the performance of GDP-x_t and GDP-x_0 is superior to GDP-x_t with Σ and GDP-x_0 with Σ, respectively, verifying the absence of variance Σ can yield better quality of images. Moreover, the results of GDP-x_0 and GDP-x_0 with Σ are better than GDP-x_t and GDP-x_t with Σ, respectively, demonstrating the superiority of the guidance on $\boldsymbol{x}_0$ protocol.

Table 5. **The ablation study on the variance Σ and the way of the guidance.**

Task	4× Super resolution				Deblur			
	PSNR	SSIM	Consistency	FID	PSNR	SSIM	Consistency	FID
GDP -x_t with Σ	22.86	0.60	88.37	68.04	22.06	0.57	69.46	80.39
GDP -x_0 with Σ	22.09	0.58	93.19	41.22	23.49	0.65	68.67	50.29
GDP -x_t	24.27	0.67	80.32	64.67	25.86	0.73	54.08	5.00
GDP -x_0	**24.42**	**0.68**	**6.49**	**38.24**	**25.98**	**0.75**	**41.27**	**2.44**

Task	25% Inpainting				Colorization			
	PSNR	SSIM	Consistency	FID	PSNR	SSIM	Consistency	FID
GDP -x_t with Σ	25.28	0.70	171.44	73.32	17.67	0.70	246.26	145.20
GDP -x_0 with Σ	24.58	0.75	65.59	22.77	21.28	0.91	66.57	38.39
GDP -x_t	31.06	0.93	8.80	20.24	21.30	0.86	75.24	66.43
GDP -x_0	**34.40**	**0.96**	**5.29**	**16.58**	**21.41**	**0.92**	**36.92**	**37.60**

Table 6. **The ablation study on the optimizable degradation and patch-based tactic.**

Methods	LOL					NTIRE			
	PSNR	SSIM	FID	LOE	PI	PSNR	SSIM	LPIPS	FID
Model A	11.05	0.49	156.51	707.57	8.61	24.12	0.67	0.32	86.69
Model B	9.01	0.37	355.99	969.89	9.04	9.83	0.04	1.02	253.11
GDP-x_t	7.32	0.57	238.92	364.15	8.26	19.36	0.65	0.30	63.89
GDP-x_0	**13.93**	**0.63**	**75.16**	**110.39**	**6.47**	**24.88**	**0.86**	**0.13**	**50.05**

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 devised 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-x_0 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

[1] Muhammad Asim, Fahad Shamshad, and Ali Ahmed. Blind image deconvolution using deep generative priors. *IEEE Transactions on Computational Imaging*, 6:1493–1506, 2020. 1, 2

[2] Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems*, 34:17981–17993, 2021. 2, 3

[3] Omri Avrahami, Dani Lischinski, and Ohad Fried. Blended diffusion for text-driven editing of natural images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18208–18218, 2022. 4

[4] Georgios Batzolis, Jan Stanczuk, Carola-Bibiane Schönlieb, and Christian Etmann. Conditional image generation with score-based diffusion models. *arXiv preprint arXiv:2111.13606*, 2021. 4

[5] Isabelle Begin and FR Ferrie. Blind super-resolution using a learning-based approach. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 2, pages 85–89. IEEE, 2004. 2

[6] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980. 7

[7] Ruojin Cai, Guandao Yang, Hadar Averbuch-Elor, Zekun Hao, Serge Belongie, Noah Snavely, and Bharath Hariharan. Learning gradient fields for shape generation. In *European Conference on Computer Vision*, pages 364–381. Springer, 2020. 2

[8] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network based noise modeling. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3155–3164, 2018. 2

[9] Nanxin Chen, Yu Zhang, Heiga Zen, Ron J Weiss, Mohammad Norouzi, and William Chan. Wavegrad: Estimating gradients for waveform generation. *arXiv preprint arXiv:2009.00713*, 2020. 4

[10] Xiangyu Chen, Yihao Liu, Zhengwen Zhang, Yu Qiao, and Chao Dong. Hdrunet: Single image hdr reconstruction with denoising and dequantization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 354–363, 2021. 1

[11] Yu Chen, Ying Tai, Xiaoming Liu, Chunhua Shen, and Jian Yang. Fsrnet: End-to-end learning face super-resolution with facial priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2492–2501, 2018. 7

[12] Jooyoung Choi, Jungbeom Lee, Chaehun Shin, Sungwon Kim, Hyunwoo Kim, and Sungroh Yoon. Perception prioritized training of diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11472–11481, 2022. 4

[13] Ryan Dahl, Mohammad Norouzi, and Jonathon Shlens. Pixel recursive super resolution. In *Proceedings of the IEEE international conference on computer vision*, pages 5439–5448, 2017. 7

[14] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 2

[15] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021. 2, 4, 5, 7

[16] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 1

[17] Alexey Dosovitskiy and Thomas Brox. Generating images with perceptual similarity metrics based on deep networks. *Advances in neural information processing systems*, 29, 2016. 7

[18] Wenchao Du, Hu Chen, and Hongyu Yang. Learning invariant representation for unsupervised image restoration. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pages 14483–14492, 2020. 3

[19] Majed El Helou and Sabine Süsstrunk. Bigprior: Toward decoupling learned prior hallucination and data fidelity in image restoration. *IEEE Transactions on Image Processing*, 31:1628–1640, 2022. 2

[20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 2

[21] Jinjin Gu, Yujun Shen, and Bolei Zhou. Image processing using multi-code gan prior. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3012–3021, 2020. 2

[22] Shuyang Gu, Dong Chen, Jianmin Bao, Fang Wen, Bo Zhang, Dongdong Chen, Lu Yuan, and Baining Guo. Vector quantized diffusion model for text-to-image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10696–10706, 2022. 3

[23] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1780–1789, 2020. 6, 7

[24] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018. 1, 5

[25] Yu He, Kim-Hui Yap, Li Chen, and Lap-Pui Chau. A soft map framework for blind super-resolution image reconstruction. *Image and Vision Computing*, 27(4):364–373, 2009. 2

[26] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 2, 3

[27] Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *J. Mach. Learn. Res.*, 23:47–1, 2022. 7

[28] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021. 2, 6

[29] Nima Khademi Kalantari, Ravi Ramamoorthi, et al. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.*, 36(4):144–1, 2017. 8

[30] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017. 2, 6, 7

[31] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *arXiv preprint arXiv:2201.11793*, 2022. 2, 4, 6, 7

[32] Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. *Advances in Neural Information Processing Systems*, 34:21757–21769, 2021. 4, 6, 7

[33] Rihuan Ke and Carola-Bibiane Schönlieb. Unsupervised image restoration using partially linear denoisers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5796–5812, 2021. 2

[34] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016. 2

[35] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in neural information processing systems*, 34:21696–21707, 2021. 2, 3

[36] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8878–8887, 2019. 1, 5

[37] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Learning representations for automatic colorization. In *European conference on computer vision*, pages 577–593. Springer, 2016. 1, 5

[38] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 1

[39] 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

[40] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *arXiv preprint arXiv:2103.00860*, 2021. 6

[41] Chongyi Li, Jichang Guo, Fatih Porikli, and Yanwei Pang. Lightennet: A convolutional neural network for weakly illuminated image enhancement. *Pattern recognition letters*, 104:15–22, 2018. 6

[42] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 1

[43] Jingyun Liang, Kai Zhang, Shuhang Gu, Luc Van Gool, and Radu Timofte. Flow-based kernel prior with application to blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10601–10610, 2021. 5

[44] Seokjae Lim and Wonjun Kim. Dslr: deep stacked laplacian restorer for low-light image enhancement. *IEEE Transactions on Multimedia*, 23:4272–4284, 2020. 6

[45] Jiaying Liu, Dejia Xu, Wenhan Yang, Minhao Fan, and Haofeng Huang. Benchmarking low-light image enhancement and beyond. *International Journal of Computer Vision*, 129(4):1153–1184, 2021. 5, 6, 7, 8

[46] Xihui Liu, Dong Huk Park, Samaneh Azadi, Gong Zhang, Arman Chopikyan, Yuxiao Hu, Humphrey Shi, Anna Rohrbach, and Trevor Darrell. More control for free! image synthesis with semantic diffusion guidance. *arXiv preprint arXiv:2112.05744*, 2021. 5

[47] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 6

[48] Kun Lu and Lihong Zhang. Tbefn: A two-branch exposure-fusion network for low-light image enhancement. *IEEE Transactions on Multimedia*, 23:4093–4105, 2020. 6

[49] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, volume 220, page 4, 2018. 6

[50] Chenlin Meng, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*, 2021. 2

[51] Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pages 2437–2445, 2020. 2

[52] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *15th Pacific Conference on Computer Graphics and Applications (PG'07)*, pages 382–390. IEEE, 2007. 7

[53] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. In *Computer graphics forum*, volume 28, pages 161–171. Wiley Online Library, 2009. 7

[54] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 8

[55] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021. 3

[56] Yuzhen Niu, Jianbin Wu, Wenxi Liu, Wenzhong Guo, and Rynson WH Lau. Hdr-gan: Hdr image reconstruction from multi-exposed ldr images with large motions. *IEEE Transactions on Image Processing*, 30:3885–3896, 2021. 2, 5, 8

[57] Gregory Ongie, Ajil Jalal, Christopher A Metzler, Richard G Baraniuk, Alexandros G Dimakis, and Rebecca Willett. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1):39–56, 2020. 1

[58] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. Exploiting deep generative prior for versatile image restoration and manipulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 2, 4, 6, 7

[59] Eduardo Pérez-Pellitero, Sibi Catley-Chandar, Ales Leonardis, and Radu Timofte. Ntire 2021 challenge on high dynamic range imaging: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 691–700, 2021. 2, 7, 8

[60] Guocheng Qian, Jinjin Gu, Jimmy S Ren, Chao Dong, Furong Zhao, and Juan Lin. Trinity of pixel enhancement: a joint solution for demosaicking, denoising and super-resolution. *arXiv preprint arXiv:1905.02538*, 1(3):4, 2019. 2

[61] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022. 2

[62] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In *International Conference on Machine Learning*, pages 8821–8831. PMLR, 2021. 2

[63] Renu M Rameshan, Subhasis Chaudhuri, and Rajbabu Velmurugan. High dynamic range imaging under noisy observations. In *2011 18th IEEE International Conference on Image Processing*, pages 1333–1336. IEEE, 2011. 2

[64] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017. 4, 6, 7

[65] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 4

[66] Dohoon Ryu and Jong Chul Ye. Pyramidal denoising diffusion probabilistic models. *arXiv preprint arXiv:2208.01864*, 2022. 3

[67] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10, 2022. 2

[68] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 4, 7

[69] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. Singan: Learning a generative model from a single natural image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4570–4580, 2019. 2

[70] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3118–3126, 2018. 3

[71] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015. 2, 4

[72] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 2

[73] Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. *Advances in neural information processing systems*, 33:12438–12448, 2020. 2

[74] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 2

[75] Maitreya Suin, Kuldeep Purohit, and AN Rajagopalan. Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3606–3615, 2020. 1

[76] Anwaar Ulhaq, Naveed Akhtar, and Ganna Pogrebna. Efficient diffusion models for vision: A survey. *arXiv preprint arXiv:2210.09292*, 2022. 2

[77] Arash Vahdat, Karsten Kreis, and Jan Kautz. Score-based generative modeling in latent space. *Advances in Neural Information Processing Systems*, 34:11287–11302, 2021. 2

[78] Jinyi Wang, Zhaoyang Lyu, Dahua Lin, Bo Dai, and Hongfei Fu. Guided diffusion model for adversarial purification. *arXiv preprint arXiv:2205.14969*, 2022. 4

[79] Lin Wang and Kuk-Jin Yoon. Deep learning for hdr imaging: State-of-the-art and future trends. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1

[80] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing*, 22(9):3538–3548, 2013. 8

[81] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1905–1914, 2021. 5

[82] Allan G Weber. The usc-sipi image database: Version 5. *http://sipi. usc. edu/database/*, 2006. 6, 7

[83] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 6, 7, 8

[84] Quanlin Wu, Hang Ye, and Yuntian Gu. Guided diffusion model for adversarial purification from random noise. *arXiv preprint arXiv:2206.10875*, 2022. 3, 5

[85] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. Deep high dynamic range imaging with large foreground motions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 117–132, 2018. 8

[86] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. Attention-guided network for ghost-free high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1751–1760, 2019. 8

[87] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3063–3072, 2020. 6

[88] Raymond A Yeh, Chen Chen, Teck Yian Lim, Alexander G Schwing, Mark Hasegawa-Johnson, and Minh N Do. Semantic image inpainting with deep generative models. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5485–5493, 2017. 1, 5

[89] Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*, 2015. 2, 7

[90] Ke Yu, Chao Dong, Liang Lin, and Chen Change Loy. Crafting a toolchain for image restoration by deep reinforcement learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2443–2452, 2018. 2

[91] Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang Lin. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 701–710, 2018. 3

[92] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4791–4800, 2021. 5

[93] Lin Zhang, Lijun Zhang, Xiao Liu, Ying Shen, Shaoming Zhang, and Shengjie Zhao. Zero-shot restoration of back-lit images using deep internal learning. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1623–1631, 2019. 6

[94] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer, 2016. 1

[95] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129(4):1013–1037, 2021. 6

[96] Yongbing Zhang, Siyuan Liu, Chao Dong, Xinfeng Zhang, and Yuan Yuan. Multiple cycle-in-cycle generative adversarial networks for unsupervised image super-resolution. *IEEE transactions on Image Processing*, 29:1101–1112, 2019. 3

[97] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. 6

[98] Zhuoran Zheng, Wenqi Ren, Xiaochun Cao, Xiaobin Hu, Tao Wang, Fenglong Song, and Xiuyi Jia. Ultra-high-definition image dehazing via multi-guided bilateral learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16180–16189. IEEE, 2021. 5

[99] Anqi Zhu, Lin Zhang, Ying Shen, Yong Ma, Shengjie Zhao, and Yicong Zhou. Zero-shot restoration of underexposed images via robust retinex decomposition. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2020. 6

[100] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 3