Image Restoration with Mean-Reverting Stochastic Differential Equations

Ziwei Luo [1] **Fredrik K. Gustafsson** [1] **Zheng Zhao** [1] **Jens Sjölund** [1] **Thomas B. Schön** [1]

Abstract

This paper presents a stochastic differential equation (SDE) approach for general-purpose image restoration. The key construction consists in a mean-reverting SDE that transforms a high-quality image into a degraded counterpart as a mean state with fixed Gaussian noise. Then, by simulating the corresponding reverse-time SDE, we are able to restore the origin of the low-quality image without relying on any task-specific prior knowledge. Crucially, the proposed mean-reverting SDE has a closed-form solution, allowing us to compute the ground truth time-dependent score and learn it with a neural network. Moreover, we propose a maximum likelihood objective to learn an optimal reverse trajectory that stabilizes the training and improves the restoration results. The experiments show that our proposed method achieves highly competitive performance in quantitative comparisons on image deraining, deblurring, and denoising, setting a new state-of-the-art on two deraining datasets. Finally, the general applicability of our approach is further demonstrated via qualitative results on image super-resolution, inpainting, and dehazing. Code is available at `https://github.com/Algolzw/image-restoration-sde`.

1. Introduction

Diffusion models have shown impressive performance in various image generation tasks, based on modeling a diffusion process and then learning its reverse (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song & Ermon, 2019; 2020; Song et al., 2021a;b;c; Rombach et al., 2022; Rissanen et al., 2022). Among the commonly used formulations (Yang et al., 2022), we adopt that of using the diffusion models defined via stochastic differential equations (SDEs, Song et al., 2021b;c). This entails gradually diffusing images towards a pure noise distribution using an SDE, and then generating samples by learning and simulating the corresponding reverse-time SDE (Anderson, 1982). The essence is training a neural network to estimate the score function of the noisy data distributions (Song & Ermon, 2019).

Image restoration is the general task of restoring a high-quality image from a degraded low-quality version. Common specific examples include image deraining (Li et al., 2019; Ren et al., 2019), deblurring (Nah et al., 2017; Zhang et al., 2020), denoising (Zhang et al., 2017a; 2018a), and super-resolution (Dong et al., 2015; Lugmayr et al., 2020; Luo et al., 2022a), just to mention a few. Image restoration has a rich history (Hunt, 1973; Andrews, 1974; Sezan & Tekalp, 1990; Banham & Katsaggelos, 1997) and remains an active topic within computer vision where learning-based approaches have a prominent role (Zhang & Zuo, 2017; Zhang et al., 2017b; Wang et al., 2022; Xiao et al., 2022).

Diffusion models have recently been applied to different image restoration tasks. Saharia et al. (2022b;a) train diffusion models which are conditioned on the low-quality images, while Lugmayr et al. (2022) utilize a pretrained unconditional model together with a modified generative process. Others explicitly treat image restoration as an inverse problem, assuming that the degradation and its parameters are known at test-time (Kawar et al., 2021; Chung et al., 2023; Kawar et al., 2022). These methods all employ the standard forward process, which diffuses images to pure noise. The reverse (generative) processes are thus initialized with sampled noise of high variance, which can result in poor restoration of the ground truth high-quality image. A number of experiments have shown that diffusion models can produce better perceptual scores, but often perform unsatisfactory in terms of some pixel/structure based distortion criteria (Saharia et al., 2022b; Li et al., 2022; Kawar et al., 2021).

To address this issue, we propose to solve the image restoration problem using a *mean-reverting* SDE. As illustrated in Figure 1, this adapts the forward process such that it models the image degradation itself, from a high-quality image to its low-quality counterpart. By simulating the corresponding reverse-time SDE, high-quality images can be restored. Importantly, no task-specific prior knowledge is required to model the image degradation at test time, just a set of image

[1]Department of Information Technology, Uppsala University, Sweden. Correspondence to: Ziwei Luo <ziwei.luo@it.uu.se>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

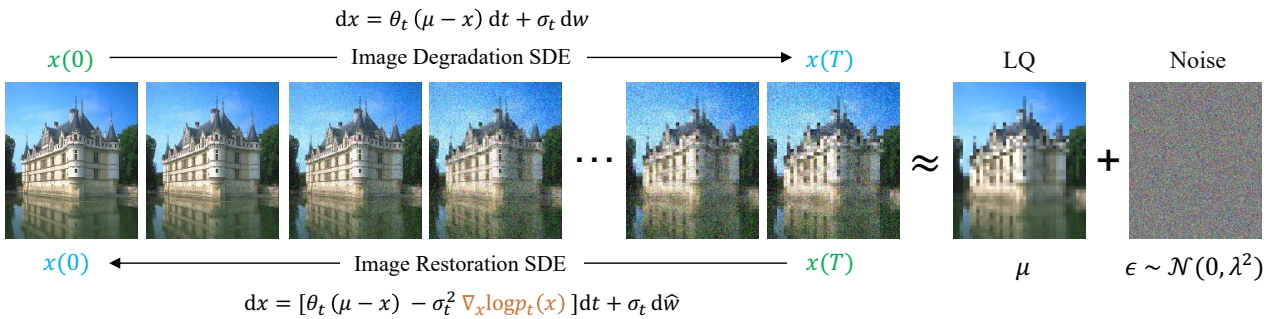


Figure 1. An overview of our proposed construction, where a mean-reverting SDE (3) is used for image restoration. The SDE models the degradation process from a high-quality image $x(0)$ to its low-quality counterpart μ, by diffusing $x(0)$ towards a noisy version $\mu + \epsilon$ of the low-quality image. By simulating the corresponding reverse-time SDE, high-quality images can then be restored.

pairs for training. Our main contributions are as follows:

- We propose a general-purpose image restoration approach using a mean-reverting SDE that directly models the image degradation process. Our formulation has a closed-form solution that enables us to compute the ground truth time-dependent score function and train a neural network to estimate it.
- We propose a simple alternative loss function for training the neural network, based on maximizing the likelihood of the reverse-time trajectory. The loss is demonstrated to stabilize training and consistently improve the image restoration performance compared to the common score matching objective.
- We demonstrate the general applicability of our proposed approach by applying it to six diverse image restoration tasks: image deraining, deblurring, denoising, super-resolution, inpainting and dehazing.
- Our approach achieves highly competitive restoration performance in quantitative comparisons on image deraining, deblurring and denoising, setting a new state-of-the-art on two deraining datasets.

2. Background

In this section, we briefly review the key concepts underlying SDE-based diffusion models and show the process of generating samples with reverse-time SDEs. Let p_0 denote the initial distribution that represents the data, and $t \in [0, T]$ denote the continuous time variable. We consider a diffusion process $\{x(t)\}_{t=0}^{T}$ defined by an SDE of the form,

$$\mathrm{d}x = f(x, t)\,\mathrm{d}t + g(t)\,\mathrm{d}w, \quad x(0) \sim p_0(x), \tag{1}$$

where f and g are the drift and dispersion functions, respectively, w is a standard Wiener process, and $x(0) \in \mathbb{R}^d$ is an initial condition. Typically, the terminal state $x(T)$ follows a Gaussian distribution with fixed mean and variance. The general idea is to design such an SDE that gradually transforms the data distribution into fixed Gaussian noise (Song et al., 2021c; Lu et al., 2022; De Bortoli et al., 2022).

We can then reverse the process to sample data from noise by simulating the SDE backward in time (Song et al., 2021c). Anderson (1982) shows that a reverse-time representation of the SDE (1) is given by

$$\mathrm{d}x = \Big[f(x, t) - g(t)^2\,\nabla_x \log p_t(x)\Big]\,\mathrm{d}t + g(t)\,\mathrm{d}\hat{w}, \tag{2}$$

where $x(T) \sim p_T(x)$. Here, $\hat{w}$ is a reverse-time Wiener process and $p_t(x)$ stands for the marginal probability density function of $x(t)$ at time t. The score function $\nabla_x \log p_t(x)$ is in general intractable and thus SDE-based diffusion models approximate it by training a time-dependent neural network $s_\theta(x, t)$ under a so-called score matching objective (Hyvärinen, 2005; Song et al., 2021c).

3. Method

The key idea of our proposed image restoration approach is to combine a mean-reverting SDE with a maximum likelihood objective for neural network training. We thus refer to it as an *Image Restoration Stochastic Differential Equation* (IR-SDE). We begin by describing the forward and reverse processes of the mean-reverting SDE, and adapt previously described, score-based, training methods to estimate this SDE. Then, we describe and contrast this with our proposed loss function based on a maximum likelihood objective.

3.1. Forward SDE for Image Degradation

We construct a special case of the SDE (1) whose score function is analytically tractable, as follows:

$$\mathrm{d}x = \theta_t\,(\mu - x)\,\mathrm{d}t + \sigma_t\,\mathrm{d}w, \tag{3}$$

where μ is the state mean, and θ_t, σ_t are time-dependent positive parameters that characterize the speed of the mean-

reversion and the stochastic volatility, respectively. There is a lot of freedom when it comes to choosing θ_t and σ_t and, as we will see in Section 5.3, the choice can have a significant impact on the resulting restoration performance.

In general, μ and the starting state $x(0)$ can be set to any pair of different images. The forward SDE (3) then transfers one image to the other as a kind of noisy interpolation. To carry out image degradation, we let $x(0)$ and μ be the ground truth high-quality (HQ) image and its degraded low-quality (LQ) counterpart, respectively (see Figure 1). It is worth noting that while μ thus depends on $x(0)$ (as they are paired HQ-LQ images of the same object or scene), $x(0)$ is independent of the Brownian motion and the SDE is therefore still valid in the Itô sense.

For our SDE (3) to have a closed-form solution, we set $\sigma_t^2 \,/\, \theta_t = 2\,\lambda^2$, where λ^2 is the stationary variance. With this, we have the following:

Proposition 3.1. *Suppose that the SDE coefficients in (3) satisfy $\sigma_t^2 \,/\, \theta_t = 2\,\lambda^2$ for all times t. Then, given any starting state $x(s)$ at time $s < t$, the solution to the SDE is*

$$x(t) = \mu + \big(x(s) - \mu\big)\,\mathrm{e}^{-\bar{\theta}_{s:t}} + \int_s^t \sigma_z\,\mathrm{e}^{-\bar{\theta}_{z:t}}\,\mathrm{d}w(z), \quad (4)$$

where $\bar{\theta}_{s:t} \coloneqq \int_s^t \theta_z\,\mathrm{d}z$ is known and the transition kernel $p(x(t) \mid x(s)) = \mathcal{N}\big(x(t) \mid m_{s:t}(x(s)), v_{s:t}\big)$ is a Gaussian with mean $m_{s:t}$ and variance $v_{s:t}$ given by:

$$\begin{aligned} m_{s:t}(x(s)) &\coloneqq \mu + (x(s) - \mu)\,\mathrm{e}^{-\bar{\theta}_{s:t}}, \\ v_{s:t} &\coloneqq \int_s^t \sigma_z^2\,\mathrm{e}^{-2\bar{\theta}_{z:t}}\,\mathrm{d}z \\ &= \lambda^2\left(1 - \mathrm{e}^{-2\,\bar{\theta}_{s:t}}\right). \end{aligned} \quad (5)$$

The proof is provided in Appendix A. To simplify the notation when the starting state is $x(0)$, we substitute $\bar{\theta}_{0:t}, m_{0:t}, v_{0:t}$ with $\bar{\theta}_t, m_t, v_t$, respectively. Then we have the distribution of $x(t)$ at any time t conditioned on the initial state, given by

$$\begin{aligned} p_t(x) &= \mathcal{N}\big(x(t) \mid m_t(x), v_t\big), \\ m_t(x) &\coloneqq \mu + (x(0) - \mu)\,\mathrm{e}^{-\bar{\theta}_t}, \\ v_t &\coloneqq \lambda^2\left(1 - \mathrm{e}^{-2\,\bar{\theta}_t}\right). \end{aligned} \quad (6)$$

Note that as $t \to \infty$, the mean m_t converges to the low-quality image μ and the variance v_t converges to the stationary variance λ^2 (hence the qualifier "mean-reverting"). In other words, the forward SDE (3) diffuses the high-quality image into a low-quality image with fixed Gaussian noise.

3.2. Reverse-Time SDE for Image Restoration

To recover the high-quality image from the terminal state $x(T)$, we reverse the SDE (3) according to (2) to derive an image restoration SDE (IR-SDE), given by

$$\mathrm{d}x = \big[\theta_t\,(\mu - x) - \sigma_t^2\,\nabla_x \log p_t(x)\big]\,\mathrm{d}t + \sigma_t\,\mathrm{d}\hat{w}. \quad (7)$$

At test time, the only unknown part is the score $\nabla_x \log p_t(x)$ of the marginal distribution at time t. But during training, the ground truth, high-quality image $x(0)$ is available and thus we can train a neural network to estimate the conditional score $\nabla_x \log p_t(x \mid x(0))$. Specifically, we can use (6) to compute the ground truth score as

$$\nabla_x \log p_t(x \mid x(0)) = -\frac{x(t) - m_t(x)}{v_t}. \quad (8)$$

This is analogous to the standard denoising score-matching which also computes the ground truth score based on a clean image and its noisy counterpart (Hyvärinen, 2005). Moreover, if we reparameterize $x(t) = m_t(x) + \sqrt{v_t}\,\epsilon_t$, where ϵ_t is a standard Gaussian noise $\epsilon_t \sim \mathcal{N}(0, I)$, we can obtain the score directly in terms of the noise by

$$\nabla_x \log p_t(x \mid x(0)) = -\frac{\epsilon_t}{\sqrt{v_t}}. \quad (9)$$

Then, we follow the common practice of approximating the noise using a noise network (Ho et al., 2020), i.e. a conditional time-dependent neural network $\tilde{\epsilon}_\phi(x(t), \mu, t)$ which takes both state x, condition μ, and time t as input and outputs pure noise. Such a network can be trained with the following objective similar to that used in DDPM (Ho et al., 2020):

$$L_\gamma(\phi) \coloneqq \sum_{i=1}^{T} \gamma_i\,\mathbb{E}\Big[\big\|\tilde{\epsilon}_\phi(x_i, \mu, i) - \epsilon_i\big\|\Big], \quad (10)$$

where $\gamma_1, \ldots, \gamma_T$ are positive weights and $\{x_i\}_{i=0}^T$ denotes the discretization of the diffusion process. Once trained, we can use the network $\tilde{\epsilon}_\phi$ to generate high-quality images by sampling a noisy state x_T and iteratively solving the IR-SDE (7) with a numerical scheme, such as Euler–Maruyama or Milstein's method (Mil'stein, 1975).

3.3. Maximum Likelihood Learning

Despite the fact that the objective in (10) offers a simple way to learn the score, we empirically found that the training often becomes unstable when applied to the complicated degradations encountered in image restoration. We conjecture that this difficulty stems from trying to learn the instantaneous noise at a given time. We therefore propose an alternative maximum likelihood objective, based on the idea of trying to find the optimal trajectory $x_{1:T}$ given the high-quality image x_0. Note that this objective is not proposed to learn a more accurate score function. Instead, it is used to stabilize training and recover more accurate images.

Specifically, we want to maximize the likelihood $p(x_{1:T} \mid x_0)$ which can be factorized according to

$$p(x_{1:T} \mid x_0) = p(x_T \mid x_0) \prod_{i=2}^{T} p(x_{i-1} \mid x_i, x_0), \tag{11}$$

where $p(x_T \mid x_0) = \mathcal{N}(x_T; m_T(x_0), v_T)$ is the low-quality image distribution. Then the reverse transition can be derived from Bayes' rule (Lindholm et al., 2022):

$$p(x_{i-1} \mid x_i, x_0) = \frac{p(x_i \mid x_{i-1}, x_0) p(x_{i-1} \mid x_0)}{p(x_i \mid x_0)}. \tag{12}$$

Since all distributions are Gaussians that can be computed from Proposition 3.1, it is natural to directly find an optimal reverse state that minimizes the negative log-likelihood:

$$x_{i-1}^* = \arg\min_{x_{i-1}} \Big[-\log p\big(x_{i-1} \mid x_i, x_0\big) \Big], \tag{13}$$

where we let x_{i-1}^* represent the ideal state reversed from x_i. To simplify the notation, we let $\theta_i' := \int_{i-1}^{i} \theta_t dt$. By solving for the above objective, we have the following:

Proposition 3.2. *Given an initial state x_0, for any state x_i at discrete time $i > 0$, the optimum reversing solution x_{i-1}^* in (13) for IR-SDE is given by:*

$$\begin{aligned} x_{i-1}^* = & \frac{1 - \mathrm{e}^{-2\,\bar{\theta}_{i-1}}}{1 - \mathrm{e}^{-2\,\bar{\theta}_i}} \mathrm{e}^{-\theta_i'} (x_i - \mu) \\ & + \frac{1 - \mathrm{e}^{-2\,\theta_i'}}{1 - \mathrm{e}^{-2\,\bar{\theta}_i}} \mathrm{e}^{-\bar{\theta}_{i-1}} (x_0 - \mu) + \mu. \end{aligned} \tag{14}$$

The proof is provided in Appendix A. Note that we can also use this objective to derive the mean of DDPM[1]. Then we choose to optimize the noise network $\tilde{\epsilon}_\phi(x_i, \mu, i)$ to make the IR-SDE reverse as the optimal trajectory, as

$$J_\gamma(\phi) := \sum_{i=1}^{T} \gamma_i \, \mathbb{E}\Big[\big\| \underbrace{x_i - (\mathrm{d}x_i)_{\tilde{\epsilon}_\phi}}_{\text{reversed } x_{i-1}} - x_{i-1}^* \big\| \Big], \tag{15}$$

where $(\mathrm{d}x)_{\tilde{\epsilon}_\phi}$ denotes the reverse-time SDE in (7) and its score is predicted by the noise network $\tilde{\epsilon}_\phi$. Note that the expectation of the martingale $\int_0^t \sigma_s \, \mathrm{d}\hat{w}(s)$ is zero, implying that we only have to consider the drift part in $(\mathrm{d}x)_{\tilde{\epsilon}_\phi}$.

4. Experiments

We experimentally evaluate our proposed IR-SDE method on three popular image restoration tasks: image deraining, deblurring and denoising. We compare IR-SDE to the prevailing approaches in their respective fields. The performance of a CNN baseline is also reported in each subsection.

[1]Please refer to Appendix C for details.

Table 1. Quantitative comparison between the proposed IR-SDE with other image deraining approaches on the Rain100H test set.

METHOD	DISTORTION		PERCEPTUAL	
	PSNR↑	SSIM↑	LPIPS↓	FID↓
JORDER	26.25	0.8349	0.197	94.58
PRENET	29.46	0.8990	0.128	52.67
MPRNET	30.41	0.8906	0.158	61.59
MAXIM	30.81	0.9027	0.133	58.72
CNN-BASELINE	29.12	0.8824	0.153	57.55
IR-SDE	**31.65**	**0.9041**	**0.047**	**18.64**

Table 2. Quantitative comparison between the proposed IR-SDE with other image deraining approaches on the Rain100L test set.

METHOD	DISTORTION		PERCEPTUAL	
	PSNR↑	SSIM↑	LPIPS↓	FID↓
JORDER	36.61	0.9735	0.028	14.66
PRENET	37.48	0.9792	0.020	10.98
MPRNET	36.40	0.9653	0.077	26.79
MAXIM	38.06	0.9770	0.048	19.06
CNN-BASELINE	33.17	0.9583	0.068	27.32
IR-SDE	**38.30**	**0.9805**	**0.014**	**7.94**

The CNN baseline takes a low-quality image as input and directly outputs a high-quality version. It uses the same network architecture as our IR-SDE, but is trained by minimizing the L_1 loss between outputs and ground truth images. In addition, we further propose a special SDE and an ordinary differential equation (ODE) to address the Gaussian denoising task. For all tasks, the Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al., 2018b) and Fréchet inception distance (FID) (Heusel et al., 2017) are reported to measure the perceptual discrepancy and visual effect. The PSNR and SSIM (Wang et al., 2004) are also provided to measure the pixel/structure similarity. Furthermore, we qualitatively illustrate the proposed method on image super-resolution, inpainting, and dehazing tasks. This shows that our method generalizes well to various image restoration problems, and *the only change required for each task was to change the dataset*. Implementation details are provided in Appendix D. For each of the six image restoration tasks, additional qualitative results are also found in Appendix E.

4.1. Image Deraining

We evaluate IR-SDE on two synthetic raining datasets: Rain100H (Yang et al., 2017) and Rain100L (Yang et al., 2017). The former has 1 800 pairs of images with/without rain for training, and 100 pairs for testing. The latter has 200 pairs for training and 100 pairs for testing. In this task, we report PSNR and SSIM scores on the Y channel (YCbCr space) similar to existing deraining methods (Ren et al., 2019; Zamir et al., 2021). Moreover, we compare our methods with several state-of-the-art deraining approaches

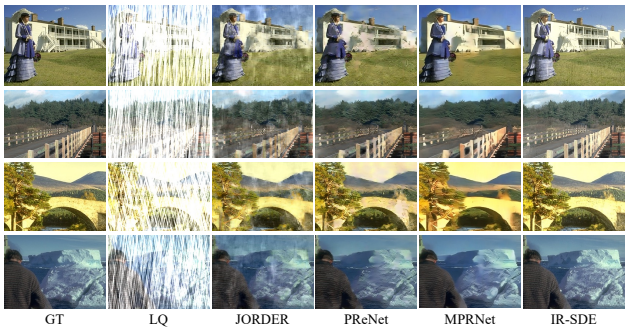


Figure 2. Visual results of our IR-SDE method and other *deraining* approaches on the Rain100H dataset.

Table 3. Quantitative comparison between the proposed IR-SDE with other image deblurring approaches on the GoPro test set.

METHOD	DISTORTION		PERCEPTUAL	
	PSNR↑	SSIM↑	LPIPS↓	FID↓
DEEPDEBLUR	29.08	0.9135	0.135	15.14
DEBLURGAN	28.70	0.8580	0.178	27.02
DEBLURGAN-V2	29.55	0.9340	0.117	13.40
DBGAN	31.18	0.9164	0.112	12.65
MAXIM	**32.86**	**0.9403**	0.089	11.57
CNN-BASELINE	28.87	0.8469	0.225	23.09
IR-SDE	30.70	0.9010	**0.064**	**6.32**

such as JORDER (Yang et al., 2019), PReNet (Ren et al., 2019), MPRNet (Zamir et al., 2021), and MAXIM (Tu et al., 2022). Note that achieving state-of-the-art performance on a specific task is not the main focus of this paper. Similar to other diffusion approaches, we will place more attention on the perceptual scores.

The quantitative comparisons on the two raining datasets are shown in Tables 1 and 2. The proposed IR-SDE achieves the best performance in all metrics. In particular, the perceptual scores (LPIPS and FID) of the IR-SDE are markedly better than those of the other approaches. Based on these scores and the visual comparison in Figure 2, we conclude that IR-SDE clearly produces the most realistic and high-fidelity results. Moreover, the CNN-baseline model only outperforms JORDER. Our method significantly improves its performance without changing the network structure, which further illustrates the superiority of the proposed method.

4.2. Image Deblurring

We evaluate the deblurring performance of IR-SDE on the public GoPro dataset (Nah et al., 2017) which contains 2 103 image pairs for training and 1 111 image pairs for testing. Note that the blurry images in GoPro are collected by averaging multiple sharp images captured by a high-speed video camera. Compared with other synthetic blurry images from blur kernels, the GoPro dataset contains more realistic blur and is much more complex.

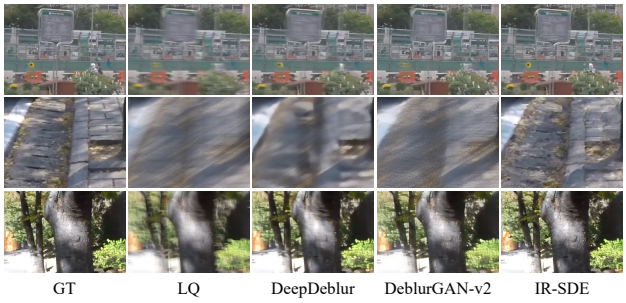


Figure 3. Visual results of our IR-SDE method compared to other *deblurring* approaches on the GoPro dataset.

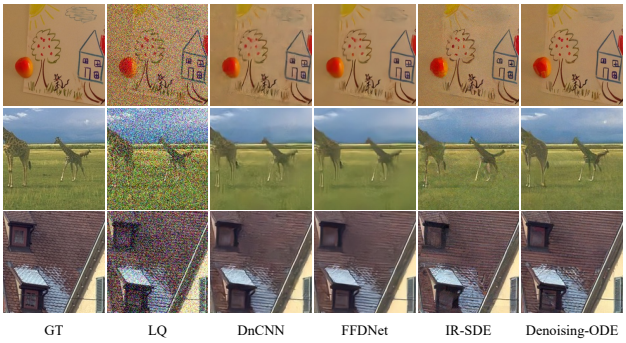


Figure 4. Visual results of our methods with other *denoising* approaches. The total timesteps of IR-SDE is fixed to 100, while the Denoising ODE only requires *22* steps to recover the clean image.

Table 3 summarizes the quantitative results of image deblurring. For comparison, we report four milestone deblurring approaches: DeepDeblur (Nah et al., 2017), DeblurGAN (Kupyn et al., 2018), DeblurGAN-v2 (Kupyn et al., 2019), DBGAN (Zhang et al., 2020), and MAXIM (Tu et al., 2022). Our method surpasses DeblurGAN-v2 by 1.15 dB in terms of PSNR and achieves the best perceptual performance overall. This indicates that the sharp images produced by IR-SDE look more realistic than other GAN-based methods and are still consistent with the ground truths. Moreover, our method significantly improves the CNN-baseline without changing its network structure, which also illustrates the superiority of our method. The visual comparison in Figure 3 shows that our method is able to handle difficult blurring cases and produces mostly clear and visually appealing results.

4.3. Gaussian Image Denoising

Recall that the Wiener process in the SDE is a Gaussian process. Hence, we introduce a *Denoising-SDE* – which is a special case of the IR-SDE in (3) and (7) – such that we can carry out the denoising computations with fewer time steps, by setting the clean image as the mean $\mu = x_0$ for all times t. Thus we can regard any noisy image as an intermediate state and directly reverse it to a clean image. Moreover,

Table 4. Denoising results on different test sets with noise level $\sigma = 25$. Note that the total steps of IR-SDE is 100, while the Denoising SDE/ODE only require *22* steps to recover the clean image. We provide more details and results in Appendix B and E, respectively.

METHOD	MCMASTER				KODAK24				CBSD68			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
DNCNN	31.52	0.8692	0.101	59.16	32.02	0.8763	0.129	41.96	31.24	0.8830	0.109	43.51
FFDNET	32.36	**0.8861**	0.103	63.84	32.13	**0.8779**	0.140	44.57	**31.22**	**0.8821**	0.121	49.64
CNN-BASELINE	31.79	0.8697	0.122	66.47	32.73	0.8666	0.161	45.81	30.74	0.8661	0.162	56.64
IR-SDE	29.48	0.8052	0.071	44.77	28.99	0.7772	0.106	35.19	28.09	0.7866	0.101	36.49
DENOISING-SDE	28.98	0.7512	0.088	45.84	28.55	0.7247	0.130	36.18	27.65	0.7457	0.131	39.25
DENOISING-ODE	**32.39**	0.8791	**0.055**	**34.66**	**32.14**	0.8739	**0.078**	**21.47**	31.14	0.8777	**0.074**	**28.71**

Table 5. Comparison of our method with DDRM (Kawar et al., 2022) on Gaussian image denoising, super-resolution, and face inpainting. We use the CBSD68, DIV2K, and CelebA-HQ datasets for task evaluations, respectively. Note that DDRM requires that the degradation parameters are known and can be composed by SVD. Moreover, all images are center cropped with size 256×256.

METHOD	IMAGE DENOISING				SUPER-RESOLUTION				FACE INPAINTING			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
DDRM	29.57	0.8484	0.135	58.99	24.35	0.5927	0.364	78.71	27.16	0.8893	0.089	37.02
CNN-BASELINE	29.69	0.8529	0.170	59.99	**26.64**	**0.6729**	0.389	133.95	**29.22**	**0.9218**	0.065	38.35
OURS	**31.01**	**0.8746**	**0.069**	**29.72**	25.90	0.6570	**0.231**	**45.36**	28.37	0.9166	**0.046**	**25.13**

since there is only Gaussian noise on the clean image, it is reasonable to derive a denoising ordinary differential equation (ODE) that shares the same marginal probability as the SDE (Song et al., 2021c) but can perform denoising without introducing additional noise from a Wiener process. This *Denoising-ODE* is given by,

$$\mathrm{d}x = \left[\theta_t\left(\mu - x\right) - \frac{1}{2}\,\sigma_t^2\,\nabla_x \log p_t(x)\right]\mathrm{d}t. \tag{16}$$

Theoretically, we can use (16) to solve the Gaussian denoising problem deterministically. The main difference between Denoising SDE and ODE is the stochastic term (i.e., Wiener process). In Appendix B, we provide a detailed deduction for the Denoising SDE/ODE and show that we can derive an appropriate denoising step to improve the sample efficiency.

To evaluate the image denoising performance, we train our models on 8 294 high-quality images collected from the DIV2K (Agustsson & Timofte, 2017), Flickr2K (Timofte et al., 2017), BSD500 (Arbelaez et al., 2010), and Waterloo Exploration datasets (Ma et al., 2016). Then all models are evaluated on the McMaster (Zhang et al., 2011), Kodak24 (Franzen, 1999), and CBSD68 (Martin et al., 2001) datasets. To show that our methods are in line with the state-of-the-art, we compare to the methods of (Zhang et al., 2017a) and (Zhang et al., 2018a), which we call DnCNN and FFDNet, respectively.

The numerical results for the three test datasets are reported in Table 4. The IR-SDE has a high perceptual performance but its fidelity scores (i.e., PSNR and SSIM) are worse than other CNN-based methods, and the same goes for Denoising-SDE. The reason may be that the diffusion process is not identifiable from the Gaussian noises, because the Denoising-ODE, which does not have the stochastic term, has significantly better PSNR on all datasets. The visual comparisons are shown in Figure 4. One can see that CNN-based methods very often produce over-smoothed images. Although IR-SDE and Denoising-ODE both generate realistic results, those of the Denoising-ODE are less noisy. We also compare Denoising-ODE with the recent diffusion method DDRM (Kawar et al., 2022) on cropped images in Table 5, achieving superior performance across all metrics.

Figure 5. Visual results of our IR-SDE method with EDSR on the DIV2K validation dataset for *super-resolution*. The LQ images are bicubicly upsampled to have the same size as GT images.

4.4. Qualitative Experiments

In this section, we further demonstrate the general applicability of our proposed IR-SDE method by performing

Figure 6. Illustration of the visual process of the reverse-time image restoration. The top row is the example of denoising ODE, in which the state $x(T)$ is exactly the LQ image. Middle and bottom rows are examples of IR-SDE for deraining and deblurring, respectively.

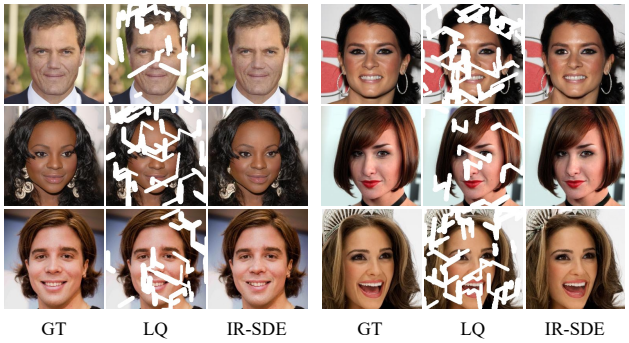


Figure 7. Visual results of *inpainting* on the CelebA-HQ dataset.

Figure 8. Visual results of *dehazing* on the SOTS indoor dataset.

qualitative experiments on image super-resolution, inpainting, and dehazing. The training settings for these experiments are the same as those of the previous sections. For super-resolution and inpainting, we also compare our quantitative results with DDRM (Kawar et al., 2022) to show the superiority of our method.

Super-Resolution We first experiment on single image super-resolution, which is a fundamental and challenging task in computer vision. Our IR-SDE is trained and evaluated on the DIV2K (Agustsson & Timofte, 2017) dataset. As an additional preprocessing step, all the low-resolution images are bicubicly re-scaled to be of the same size as the corresponding high-resolution images. Figure 5 shows the qualitative results on the DIV2K validation dataset. Compared to the L_2 trained EDSR (Lim et al., 2017) model, our IR-SDE is able to restore images that have rich details and are visually clear and realistic. Here we also provide the quantitative comparison with another diffusion-based model DDRM (Kawar et al., 2022) in Table 5.

Face Inpainting Inpainting is the task of filling new content to missing regions of an image. We select the CelebA-HQ (Karras et al., 2018) dataset to train and test the IR-SDE on this task. Here we set the mask to be unknown. The inpainted regions must harmonize with the rest such that the overall face is semantically reasonable and has a natural appearance. Visual examples of face inpainting are illustrated in Figure 7. As can be seen, the proposed IR-SDE demonstrates a strong generative capability in restoring masked areas while it at the same time maintains consistency with the original image. Moreover, the quantitative comparison with DDRM (Kawar et al., 2022) is shown in Table 5.

Dehazing Image dehazing is often an important prerequisite for improving the robustness of other high-level vision tasks. Note that DDRM requires that the degradation parameters are known and can be composed by SVD, and therefore not can be applied to dehazing. In contrast, our method is

flexible to deal with all kinds of tasks. We train the IR-SDE on the RESIDE (Li et al., 2018) Indoor Training Set (ITS) and test it on the Synthetic Objective Testing Set (SOTS). As shown in Figure 8, our IR-SDE successfully restores haze-free indoor scenes from the low-quality and low-contrast inputs. The quantitative results are shown in Appendix E.

5. Discussion and Analysis

In this section, we first give an in-depth analysis of the reverse-time restoration process of the IR-SDE, and then study two important components (maximum likelihood objective and theta schedule) and the limitations in more detail.

5.1. Reverse-Time Restoration Process

For IR-SDE, the terminal state x_T is generally obtained by adding noise to the degraded low-quality image. To restore a high-quality image, both the degradation and the noise thus have to be gradually removed. But how are these two different corruptions handled in the reverse-time process?

To analyze this we provide a few concrete restoration examples in Figure 6. Note that the top row of Figure 6 shows the denoising case by Denoising-ODE, where the noisy image is considered to be an intermediate state and the only aim is to gradually remove the Gaussian noise to recover the clean image. For other image restoration cases, we find that the IR-SDE tends to assign a higher priority to handle the original degradation and only performs Gaussian denoising in the last few steps. As illustrated for the image deraining and deblurring cases in Figure 6, most of the degradation (rain and blur) has been removed already in the middle timesteps.

In addition, we show the performance curves of the IR-SDE (with cosine schedule) when it comes to deblurring a single image in Figure 9. As can be seen, the deblurring performance (in terms of PSNR and LPIPS) increases after running 20 steps and then converges in the last few steps.

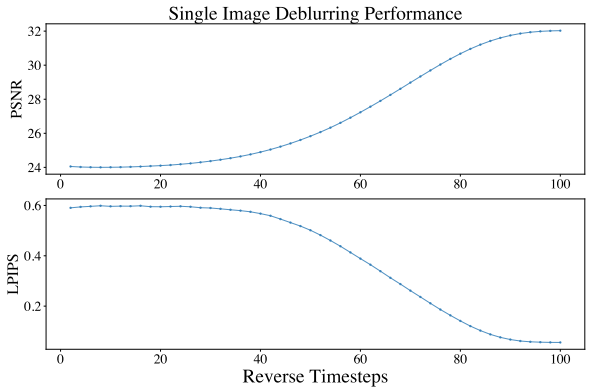


Figure 9. Performance curves of the reverse-time deblurring process. Curves are computed from the deblurring case in Figure 6.

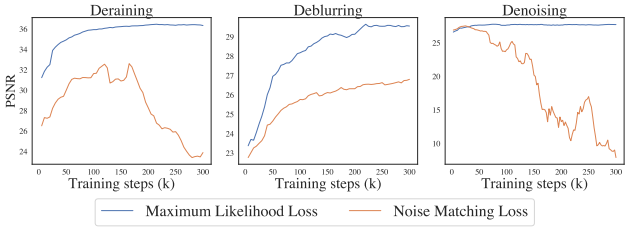


Figure 10. Training curves of the IR-SDE on various tasks. Our proposed maximum likelihood-based loss function stabilizes the training and improves the restoration performance.

5.2. Maximum Likelihood Objective

A key improvement of our IR-SDE method compared to other diffusion models, which directly learn the noise/score, is that we learn an optimal reverse-time trajectory from x_T to x_0 based on the maximum likelihood objective in (15). Here we show that this objective results in more stable training, which in turn improves the restoration performance, as illustrated in Figure 10. The PSNR when training with a noise-matching objective fluctuates and even deteriorates over time in the deraining and denoising tasks. While the training still works for deblurring, the performance is clearly inferior to the proposed maximum likelihood objective.

5.3. Time-Varying Theta Schedules

It is notable that our IR-SDE has two time-varying parameters θ_t and σ_t, which we set to be constrained by the stationary variance of x_t as $\sigma_t^2 \,/\, \theta_t = 2\,\lambda^2$ for all timesteps. Since λ is fixed as the noise level applied to the LQ image, we can simply adjust θ to construct different noise schedules in IR-SDE. As shown in Figure 11, we explore three different schedules for how to vary θ: constant, linear, and cosine (see Appendix D for details). When θ is constant, the IR-SDE simplifies to the Ornstein–Uhlenbeck (OU) process (Gillespie, 1996) which is widely used to solve mean-reverting problems. The linear/cosine schedules are widely-used in existing diffusion probabilistic models (Ho et al., 2020; Nichol & Dhariwal, 2021). We use their flipped version for θ_t such that the diffusion coefficient σ_t smoothly changes to a maximum value as $t \to \infty$. It is observed that all schedules work well for the deraining task, and that the cosine schedule performs significantly better than others.

5.4. Limitations and Future Works

We have shown the usefulness of our method on various image restoration tasks. However, it is also important to acknowledge one potential limitation: the exponential term in (6) for v_t leads to an overly smooth variance change in the last few steps (see Figure 12). In that area, the neighboring states (x_i, x_{i-1}) have quite similar appearances thus making learning difficult, especially when the maximum likelihood

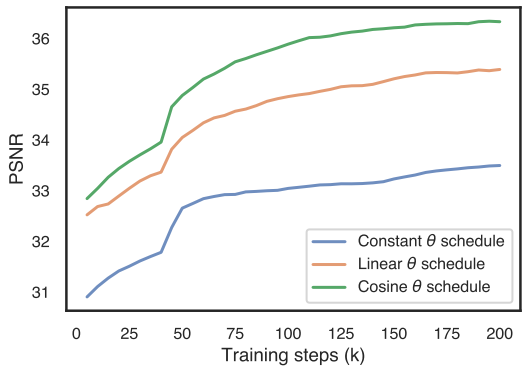


Figure 11. Training curves of different θ schedules on the image deraining task. Note that when θ is constant, the forward IR-SDE (3) simplifies to the OU process.

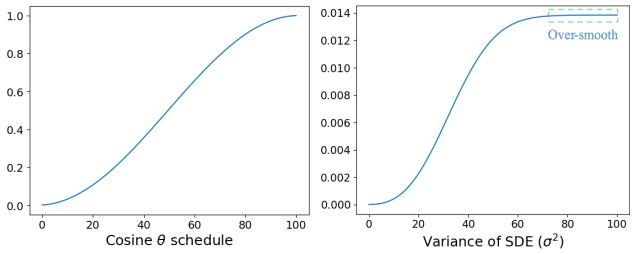


Figure 12. Cosine θ schedule and the variance of the forward SDE. The variance changes over-smoothly in the last few steps.

loss (which optimizes the difference between states) is used. In our future work, we will explore alternative theta schedules to alleviate this problem.

Moreover, it is worth noting that we can generalize the choice of the SDE, hence the conditional score, by using Tweedie's formula, see Kim & Ye (2021, Table 1) and Kim et al. (2022). As an example, if we choose the SDE to be a geometric Brownian motion, then the score in Equation (8) corresponds to that of an exponential distribution.

6. Related Work

Image restoration is an active research topic within computer vision (Zhang & Zuo, 2017; Zhang et al., 2017b; Wang et al., 2022; Xiao et al., 2022). The most common approach is to train some type of deep learning model to solve image restoration tasks in a supervised manner (Zamir et al., 2021). Various CNN-based architectures have been proposed (Zamir et al., 2021; Chen et al., 2022), and recently the use of transformers has also been extensively explored (Liang et al., 2021; Zamir et al., 2022; Luo et al., 2022b). These methods all entail training a neural network to directly predict high-quality images from given low-quality ones. In contrast, our proposed IR-SDE approach gradually restores a given low-quality image by simulating the reverse-time SDE (7) for multiple steps. While this increases the computational cost, it also enables a more accurate restoration of the ground truth. Recently, Refusion (Luo et al., 2023) extends the IR-SDE with a U-Net based latent framework to accelerate inference.

Most similar to the IR-SDE is the work of Welker et al. (2022b) and Richter et al. (2022), in which a mean-reverting SDE is applied to the speech processing tasks of speech enhancement and speech dereverberation. They use a mean-reverting SDE similar to (3) but with a different σ_t and a constant θ, i.e. a standard OU process. Also, they did not set the stationary variance condition. In a concurrent work by Welker et al. (2022a), they extend the idea to JPEG artifact removal, where they introduce another version of their SDE with a linear θ scheduler. As shown in Section 5.3, both of these are outperformed by our cosine θ scheduler. Moreover, Welker et al. (2022b); Richter et al. (2022); Welker et al. (2022a) all use the standard score matching objective, while we introduce an alternative maximum likelihood-based loss function that stabilizes training and improves the restoration performance. Finally, we demonstrate the general applicability of our approach by applying it to six diverse image restoration tasks.

7. Conclusion

We have presented a mean-reverting SDE-based method that is applicable to a wide class of image restoration tasks. Importantly, our SDE has a closed-form solution that enables us to compute the ground truth time-dependent score function and to train a neural network to estimate it. In addition, we have proposed a maximum likelihood-based loss objective, which significantly stabilizes the neural network training and consistently improves the restoration performance. The experiments performed on six diverse image restoration tasks demonstrate the wide applicability and highly competitive restoration performance of our proposed approach. Future directions include exploring techniques for optimizing the θ schedule and sampling procedures which can decrease the computational cost at test time.

Acknowledgements

This research was supported by the *Wallenberg AI, Autonomous Systems and Software Program (WASP)* funded by the Knut and Alice Wallenberg Foundation; by the project *Deep Probabilistic Regression – New Models and Learning Algorithms* (contract number: 2021-04301) funded by the Swedish Research Council; and by the *Kjell & Märta Beijer Foundation*. The computations were enabled by the *Berzelius* resource provided by the Knut and Alice Wallenberg Foundation at the National Supercomputer Centre. We also thank Daniel Gedon for providing helpful feedback.

References

Agustsson, E. and Timofte, R. Ntire 2017 challenge on single image super-resolution: dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 126–135, 2017.

Anderson, B. D. O. Reverse-time diffusion equation models. *Stochastic Processes and Their Applications*, 12(3):313–326, 1982.

Andrews, H. C. Digital image restoration: a survey. *Computer*, 7(5):36–45, 1974.

Arbelaez, P., Maire, M., Fowlkes, C., and Malik, J. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(5):898–916, 2010.

Banham, M. R. and Katsaggelos, A. K. Digital image restoration. *IEEE Signal Processing Magazine*, 14(2): 24–41, 1997.

Chen, L., Chu, X., Zhang, X., and Sun, J. Simple baselines for image restoration. In *Proceedings of the 17th European Conference on Computer Vision (ECCV)*, pp. 17–33. Springer, 2022.

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., and Ye, J. C. Diffusion posterior sampling for general noisy inverse problems. In *Proceedings of the 11th International Conference on Learning Representations (ICLR)*, 2023.

De Bortoli, V., Mathieu, E., Hutchinson, M., Thornton, J., Teh, Y. W., and Doucet, A. Riemannian score-based generative modeling. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, 2022.

Dong, C., Loy, C. C., He, K., and Tang, X. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 38(2):295–307, 2015.

Franzen, R. Kodak lossless true color image suite. *source: http://r0k. us/graphics/kodak*, 4(2), 1999.

Gillespie, D. T. Exact numerical simulation of the Ornstei–Uhlenbeck process and its integral. *Physical Review E*, 54(2):2084, 1996.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 2017.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 6840–6851, 2020.

Hunt, B. R. The application of constrained least squares estimation to image restoration by digital computer. *IEEE Transactions on Computers*, 100(9):805–812, 1973.

Hyvärinen, A. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4):695–709, 2005.

Karras, T., Aila, T., Laine, S., and Lehtinen, J. Progressive growing of GANs for improved quality, stability, and variation. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2018.

Kawar, B., Vaksman, G., and Elad, M. SNIPS: Solving noisy inverse problems stochastically. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 21757–21769, 2021.

Kawar, B., Elad, M., Ermon, S., and Song, J. Denoising diffusion restoration models. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, 2022.

Kim, K. and Ye, J. C. Noise2Score: Tweedie's approach to self-supervised image denoising without clean images. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, 2021.

Kim, K., Kwon, T., and Ye, J. C. Noise distribution adaptive self-supervised image denoising using Tweedie distribution and score matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2008–2016, June 2022.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In *Proceedings of International Conference on Learning Representations*, 2014.

Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., and Matas, J. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8183–8192, 2018.

Kupyn, O., Martyniuk, T., Wu, J., and Wang, Z. Deblurgan-v2: deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8878–8887, 2019.

Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., and Wang, Z. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1): 492–505, 2018.

Li, H., Yang, Y., Chang, M., Chen, S., Feng, H., Xu, Z., Li, Q., and Chen, Y. SRDIFF: single image super-resolution with diffusion probabilistic models. *Neurocomputing*, 479:47–59, 2022.

Li, S., Araujo, I. B., Ren, W., Wang, Z., Tokuda, E. K., Junior, R. H., Cesar-Junior, R., Zhang, J., Guo, X., and Cao, X. Single image deraining: A comprehensive benchmark analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3838–3847, 2019.

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., and Timofte, R. SwinIR: image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1833–1844, 2021.

Lim, B., Son, S., Kim, H., Nah, S., and Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 136–144, 2017.

Lindholm, A., Wahlström, N., Lindsten, F., and Schön, T. B. *Machine learning: a first course for engineers and scientists*. Cambridge University Press, 2022.

Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., and Zhu, J. Dpm-solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, 2022.

Lugmayr, A., Danelljan, M., Van Gool, L., and Timofte, R. SRFlow: learning the super-resolution space with normalizing flow. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., and Van Gool, L. Repaint: inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11461–11471, 2022.

Luo, Z., Huang, H., Yu, L., Li, Y., Fan, H., and Liu, S. Deep constrained least squares for blind image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17642–17652, 2022a.

Luo, Z., Li, Y., Cheng, S., Yu, L., Wu, Q., Wen, Z., Fan, H., Sun, J., and Liu, S. Bsrt: Improving burst super-resolution with swin transformer and flow-guided deformable alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 998–1008, 2022b.

Luo, Z., Gustafsson, F. K., Zhao, Z., Sjölund, J., and Schön, T. B. Refusion: enabling large-size realistic image restoration with latent-space diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2023.

Ma, K., Duanmu, Z., Wu, Q., Wang, Z., Yong, H., Li, H., and Zhang, L. Waterloo exploration database: new challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26(2):1004–1016, 2016.

Martin, D., Fowlkes, C., Tal, D., and Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the 18th IEEE International Conference on Computer Vision (ICCV)*, volume 2, pp. 416–423. IEEE, 2001.

Mil'stein, G. N. Approximate integration of stochastic differential equations. *Theory of Probability & Its Applications*, 19(3):557–562, 1975.

Nah, S., Hyun Kim, T., and Mu Lee, K. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3883–3891, 2017.

Nichol, A. Q. and Dhariwal, P. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning (ICML)*, pp. 8162–8171. PMLR, 2021.

Ren, D., Zuo, W., Hu, Q., Zhu, P., and Meng, D. Progressive image deraining networks: a better and simpler baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3937–3946, 2019.

Richter, J., Welker, S., Lemercier, J.-M., Lay, B., and Gerkmann, T. Speech enhancement and dereverberation with diffusion-based generative models. *arXiv preprint arXiv:2208.05830*, 2022.

Rissanen, S., Heinonen, M., and Solin, A. Generative modelling with inverse heat dissipation. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2022.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, 2022.

Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., and Norouzi, M. Palette: Image-to-image diffusion models. In *Proceedings of ACM SIGGRAPH Conference*, pp. 1–10, 2022a.

Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., and Norouzi, M. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2022b.

Sezan, M. I. and Tekalp, A. M. Survey of recent developments in digital image restoration. *Optical Engineering*, 29(5):393–404, 1990.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning (ICML)*, pp. 2256–2265. PMLR, 2015.

Song, J., Meng, C., and Ermon, S. Denoising diffusion implicit models. In *Proceedigns of International Conference on Learning Representations (ICLR)*, 2021a.

Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Song, Y. and Ermon, S. Improved techniques for training score-based generative models. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 12438–12448, 2020.

Song, Y., Durkan, C., Murray, I., and Ermon, S. Maximum likelihood training of score-based diffusion models. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 1415–1428, 2021b.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021c.

Timofte, R., Agustsson, E., Van Gool, L., Yang, M.-H., and Zhang, L. NTIRE 2017 challenge on single image super-resolution: methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 114–125, 2017.

Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., and Li, Y. MAXIM: Multi-axis MLP for image processing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5769–5780, 2022.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., and Li, H. Uformer: a general U-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17683–17693, 2022.

Welker, S., Chapman, H. N., and Gerkmann, T. DriftRec: Adapting diffusion models to blind image restoration tasks. *arXiv preprint arXiv:2211.06757*, 2022a.

Welker, S., Richter, J., and Gerkmann, T. Speech enhancement with score-based generative models in the complex STFT domain. *ISCA Interspeech*, 2022b.

Xiao, J., Fu, X., Wu, F., and Zha, Z.-J. Stochastic window transformer for image restoration. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Shao, Y., Zhang, W., Cui, B., and Yang, M.-H. Diffusion models: a comprehensive survey of methods and applications. *arXiv preprint arXiv:2209.00796*, 2022.

Yang, W., Tan, R. T., Feng, J., Liu, J., Guo, Z., and Yan, S. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1357–1366, 2017.

Yang, W., Tan, R. T., Feng, J., Guo, Z., Yan, S., and Liu, J. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(6):1377–1393, 2019.

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M.-H., and Shao, L. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14821–14831, 2021.

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., and Yang, M.-H. Restormer: efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5728–5739, 2022.

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

Zhang, K., Zuo, W., Gu, S., and Zhang, L. Learning deep CNN denoiser prior for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3929–3938, 2017b.

Zhang, K., Zuo, W., and Zhang, L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018a.

Zhang, K., Luo, W., Zhong, Y., Ma, L., Stenger, B., Liu, W., and Li, H. Deblurring by realistic blurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2737–2746, 2020.

Zhang, L. and Zuo, W. Image restoration: from sparse and low-rank priors to deep priors. *IEEE Signal Processing Magazine*, 34(5):172–179, 2017.

Zhang, L., Wu, X., Buades, A., and Li, X. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. *Journal of Electronic Imaging*, 20 (2):023016, 2011.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595, 2018b.

Zhang, Y., Li, D., Shi, X., He, D., Song, K., Wang, X., Qin, H., and Li, H. Kbnet: Kernel basis network for image restoration. *arXiv preprint arXiv:2303.02881*, 2023.

A. Proofs

Proposition 3.1. *Suppose that the SDE coefficients in (3) satisfy $\sigma_t^2 \,/\, \theta_t = 2\,\lambda^2$ for all times t. Then, given any starting state $x(s)$ at time $s < t$, the solution to the SDE is*

$$x(t) = \mu + \big(x(s) - \mu\big)\,\mathrm{e}^{-\bar{\theta}_{s:t}} + \int_s^t \sigma_z\,\mathrm{e}^{-\bar{\theta}_{z:t}}\,\mathrm{d}w(z), \tag{17}$$

where $\bar{\theta}_{s:t} \coloneqq \int_s^t \theta_z\,\mathrm{d}z$ is known and the transition kernel $p(x(t) \mid x(s)) = \mathcal{N}\big(x(t) \mid m_{s:t}(x(s)), v_{s:t}\big)$ is a Gaussian with mean $m_{s:t}$ and variance $v_{s:t}$ given by:

$$\begin{aligned} m_{s:t}(x_s) &\coloneqq \mu + (x(s) - \mu)\,\mathrm{e}^{-\bar{\theta}_{s:t}}, \\ v_{s:t} &\coloneqq \int_s^t \sigma_z^2\,\mathrm{e}^{-2\bar{\theta}_{z:t}}\,\mathrm{d}z \\ &= \lambda^2\left(1 - \mathrm{e}^{-2\,\bar{\theta}_{s:t}}\right). \end{aligned} \tag{18}$$

Proof. Recall the SDE that we want to solve:

$$\mathrm{d}x = \theta_t\,(\mu - x)\,\mathrm{d}t + \sigma_t\,\mathrm{d}w, \tag{19}$$

where θ_t and σ_t are two time-dependent positive functions. Also recall that when the starting state is $x(0)$, we substitute $\bar{\theta}_{0:t}$ with $\bar{\theta}_t$ for notation simplicity. To solve the SDE above, let us define a surrogate differentiable function

$$\psi(x,t) = x\,\mathrm{e}^{\bar{\theta}_t}, \tag{20}$$

and by Itô's formula (in the differential form), we have

$$\begin{aligned} \mathrm{d}\psi(x,t) = &\frac{\partial\psi}{\partial t}(x,t)\,\mathrm{d}t + \frac{\partial\psi}{\partial x}(x,t)\mathbf{f}(x,t)\,\mathrm{d}t \\ &+ \frac{1}{2}\frac{\partial^2\psi}{\partial x^2}(x,t)g(t)^2\,\mathrm{d}t \\ &+ \frac{\partial\psi}{\partial x}(x,t)g(t)\,\mathrm{d}\mathrm{w}_t. \end{aligned} \tag{21}$$

By substituting $\mathbf{f}(x,t)$ and $g(t)$ with the drift and the diffusion functions in (19), we obtain

$$\mathrm{d}\psi(x,t) = \mu\theta_t\mathrm{e}^{\bar{\theta}_t}\,\mathrm{d}t + \boldsymbol{\sigma}_t\mathrm{e}^{\bar{\theta}_t}\,\mathrm{d}\mathrm{w}_t \tag{22}$$

Note here $\mathrm{d}\bar{\theta}_t = \mathrm{d}\int_0^t \theta_z\,\mathrm{d}z = \theta_t$. Then we can solve $x(t)$ conditioned on $x(s)$, as

$$\psi(x,t) - \psi(x,s) = \int_s^t \mu\theta_z\mathrm{e}^{\bar{\theta}_z}\,\mathrm{d}z + \underbrace{\int_s^t \boldsymbol{\sigma}_z\mathrm{e}^{\bar{\theta}_z}\mathrm{d}\mathrm{w}(z)}_{\sim\mathcal{N}\left(0,\,\int_s^t \boldsymbol{\sigma}_z^2\mathrm{e}^{2\bar{\theta}_z}\mathrm{d}z\right)}. \tag{23}$$

Since that θ_t and $\boldsymbol{\sigma}_t$ are scalar-valued, we can analytically compute the two integrals above and then obtain

$$x(t)\mathrm{e}^{\bar{\theta}_t} - x(s)\mathrm{e}^{\bar{\theta}_s} = \mu(\mathrm{e}^{\bar{\theta}_t} - \mathrm{e}^{\bar{\theta}_s}) + \int_s^t \boldsymbol{\sigma}_z\mathrm{e}^{\bar{\theta}_z}\mathrm{d}\mathrm{w}(z), \tag{24}$$

By dividing $\mathrm{e}^{\bar{\theta}_t}$ to both sides, we have

$$x(t) = \mu + \big(x(s) - \mu\big)\,\mathrm{e}^{-\bar{\theta}_{s:t}} + \int_s^t \sigma_z\,\mathrm{e}^{-\bar{\theta}_{z:t}}\,\mathrm{d}w_z, \tag{25}$$

which is the solution to the given SDE.

Moreover, we show that the integral term $\int_s^t \sigma_z \, \mathrm{e}^{-\bar{\theta}_{z:t}} \, \mathrm{d}w(z)$ in (25) is actually a Gaussian noise $\mathcal{N}\left(0, \int_s^t \boldsymbol{\sigma}_z^2 \mathrm{e}^{-2\bar{\theta}_{z:t}} \mathrm{d}z\right)$ such that its variance can be solve by

$$\int_s^t \boldsymbol{\sigma}_z^2 \mathrm{e}^{-2\bar{\theta}_{z:t}} \mathrm{d}z = \frac{\sigma_t^2}{2\theta_t}\mathrm{e}^0 - \frac{\sigma_s^2}{2\theta_s}\mathrm{e}^{-2\bar{\theta}_{s:t}} = \lambda^2\left(1 - \mathrm{e}^{-2\,\bar{\theta}_{s:t}}\right), \tag{26}$$

under the condition of $\sigma_t^2 \,/\, \theta_t = 2\,\lambda^2$ for all times t. By reparameterizing the solution in (25) we arrive at the following transition

$$p(x(t) \mid x(s)) = \mathcal{N}\big(x(t) \mid m_{s:t}(x(s)), v_{s:t}\big), \tag{27}$$

where

$$\begin{aligned} m_{s:t}(x_s) &\coloneqq \mu + (x(s) - \mu)\,\mathrm{e}^{-\bar{\theta}_{s:t}}, \\ v_{s:t} &\coloneqq \int_s^t \sigma_z^2 \,\mathrm{e}^{-2\bar{\theta}_{z:t}} \,\mathrm{d}z = \lambda^2\left(1 - \mathrm{e}^{-2\,\bar{\theta}_{s:t}}\right). \end{aligned} \tag{28}$$

Thus we complete the proof. □

Proposition 3.2. *Given an initial state x_0, for any state x_i at discrete time $i > 0$, the optimum reversing solution for $x_i \to x_{i-1}$ in IR-SDE is:*

$$x_{i-1}^* = \frac{1 - \mathrm{e}^{-2\,\bar{\theta}_{i-1}}}{1 - \mathrm{e}^{-2\,\bar{\theta}_i}}\mathrm{e}^{-\theta_i^{'}}(x_i - \mu) + \frac{1 - \mathrm{e}^{-2\,\theta_i^{'}}}{1 - \mathrm{e}^{-2\,\bar{\theta}_i}}\mathrm{e}^{-\bar{\theta}_{i-1}}(x_0 - \mu) + \mu. \tag{29}$$

Proof. Recall that the transition distribution $p(x_i \mid x_0)$ and $p(x_i \mid x_{i-1}, x_0)$ can be known as stated in Proposition (3.1). Our objective is to minimize the following negative log likelihood (NLL) to obtain a theoretically optimum path for the reverse SDE:

$$x_{i-1}^* = \arg\min_{x_{i-1}}\Big[-\log p\big(x_{i-1} \mid x_i, x_0\big)\Big]. \tag{30}$$

By using Bayes' rule, we have

$$\begin{aligned} -\log p\big(x_{i-1} \mid x_i, x_0\big) &= -\log \frac{p(x_i \mid x_{i-1}, x_0)p(x_{i-1} \mid x_0)}{p(x_i \mid x_0)} \\ &\propto -\log p\big(x_i \mid x_{i-1}, x_0\big) - \log p\big(x_{i-1} \mid x_0\big) \end{aligned} \tag{31}$$

where all transitions are tractable. Then we can directly solve the NLL in (30) by computing its gradient and setting it to be zero:

$$\begin{aligned} \nabla_{x_{t-1}^*}\left\{-\log p\big(x_{i-1}^* \mid x_i, x_0\big)\right\} &\propto -\nabla_{x_{i-1}^*}\log p\big(x_i \mid x_{i-1}^*, x_0\big) - \nabla_{x_{i-1}^*}\log p\big(x_{i-1}^* \mid x_0\big) \\ &= -\frac{\mathrm{e}^{-\theta_i^{'}}(x_i - \mu - (x_{i-1}^* - \mu)\mathrm{e}^{-\theta_i^{'}})}{1 - \mathrm{e}^{-2\,\theta_i^{'}}} + \frac{x_{i-1}^* - \mu - (x_0 - \mu)\mathrm{e}^{-\bar{\theta}_{i-1}}}{1 - \mathrm{e}^{-2\,\bar{\theta}_{i-1}}} \\ &= \frac{(x_{i-1}^* - \mu)\mathrm{e}^{-2\theta_i^{'}}}{1 - \mathrm{e}^{-2\,\theta_i^{'}}} + \frac{(x_{i-1}^* - \mu)}{1 - \mathrm{e}^{-2\,\bar{\theta}_{i-1}}} - \frac{(x_i - \mu)\mathrm{e}^{-\theta_i^{'}}}{1 - \mathrm{e}^{-2\,\theta_i^{'}}} - \frac{(x_0 - \mu)\mathrm{e}^{-\bar{\theta}_{i-1}}}{1 - \mathrm{e}^{-2\,\bar{\theta}_{i-1}}} \\ &= \frac{(x_{i-1}^* - \mu)(1 - \mathrm{e}^{-2\,\bar{\theta}_i})}{(1 - \mathrm{e}^{-2\,\theta_i^{'}})(1 - \mathrm{e}^{-2\,\bar{\theta}_{i-1}})} - \frac{(x_i - \mu)\mathrm{e}^{-\theta_i^{'}}}{1 - \mathrm{e}^{-2\,\theta_i^{'}}} - \frac{(x_0 - \mu)\mathrm{e}^{-\bar{\theta}_{i-1}}}{1 - \mathrm{e}^{-2\,\bar{\theta}_{i-1}}} = 0, \end{aligned} \tag{32}$$

where $\theta_i^{'} = \int_{i-1}^{i} \theta_t dt$. Since (32) is linear we get

$$x_{i-1}^* = \frac{1 - \mathrm{e}^{-2\,\bar{\theta}_{i-1}}}{1 - \mathrm{e}^{-2\,\bar{\theta}_i}}\mathrm{e}^{-\theta_i^{'}}(x_i - \mu) + \frac{1 - \mathrm{e}^{-2\,\theta_i^{'}}}{1 - \mathrm{e}^{-2\,\bar{\theta}_i}}\mathrm{e}^{-\bar{\theta}_{i-1}}(x_0 - \mu) + \mu, \tag{33}$$

which completes the proof (the second-order derivative is a positive constant, i.e. x_{i-1}^* is indeed the optimal point). □

B. Denoising SDE/ODE for Gaussian Denoising

Here we provide details for the Denoising SDE/ODE as it is a special case of the IR-SDE, in the way of denoting μ with the HR clean image

$$\mu \coloneqq x_0. \tag{34}$$

Then we have a simplified transition kernel $p(x_i \mid x_0)$ given by

$$p_{noise}(x_i|x_0) = \mathcal{N}(m_i(x_0), v_i), \tag{35}$$

where

$$m_i := x_0, \quad v_i = \lambda^2 \left(1 - \mathrm{e}^{-2\,\bar{\theta}_i}\right). \tag{36}$$

Correspondingly, and the optimum path from $x_i \to x_{i-1}$ becomes

$$x_{i-1}^* = \frac{1 - \mathrm{e}^{-2\bar{\theta}_{i-1}}}{1 - \mathrm{e}^{-2\bar{\theta}_i}} \mathrm{e}^{-\theta_i}(x_i - x_0) + x_0. \tag{37}$$

Recall the reverse-time version of the IR-SDE:

$$\mathrm{d}x = \left[\theta_t\,(\mu - x) - \sigma_t^2\,\nabla_x \log p_t(x)\right]\mathrm{d}t + \sigma_t\,\mathrm{d}\hat{w}, \tag{38}$$

and the sampling strategy of x_i:

$$x_i = m_i + \sqrt{v_i}\,\epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, I). \tag{39}$$

We can then approximate $\mu - x$ from (39) and combine it with (9) to rewrite (38) to the Denoising SDE:

$$\mathrm{d}x = -\frac{1}{2}\boldsymbol{\sigma}(t)^2(1 + \mathrm{e}^{-2\bar{\theta}_t})\nabla_{x_t} \log p_t(x_t)\mathrm{d}t + \boldsymbol{\sigma}(t)\mathrm{d}\bar{\mathbf{w}}. \tag{40}$$

In addition, (Song et al., 2021c) states that there exists a deterministic process that shares the same marginal probability densities as the IR-SDE. Once we have the score, we can also recover images through a deterministic trajectory, as the *probability flow ODE* (Song et al., 2021c):

$$\mathrm{d}x = \left[\theta_t\,(\mu - x) - \frac{1}{2}\,\sigma_t^2\,\nabla_x \log p_t(x)\right]\mathrm{d}t. \tag{41}$$

In this denoising case, the corresponding ODE for (40) is

$$\mathrm{d}x = -\frac{1}{2}\boldsymbol{\sigma}_t^2 \mathrm{e}^{-2\bar{\theta}_t}\nabla_x \log p_t(x)\mathrm{d}t. \tag{42}$$

Moreover, once we know the real noise level σ_{real} of an image, we can easily derive a appropriate timestep t^* such that the variance of $p_{noise}(x_i|x_0)$ happens to be the noise level:

$$\sigma_{\text{real}}^2 = v_t = \lambda^2 \left(1 - \mathrm{e}^{-2\,\bar{\theta}_t}\right). \tag{43}$$

By solving the $\bar{\theta}_t$ we have

$$t^* = \arg\min_t \|\bar{\theta}_t - \frac{1}{2\Delta t}\log(1 - \frac{\sigma_{\text{real}}^2}{\lambda^2})\|. \tag{44}$$

where Δt denotes the time interval. Based on it, our method is able to process arbitrary noise levels and can start denoising from middle states, which is more practical and improves sample efficiency.

C. Relationship between Maximum Likelihood Objective and DDPM

To further illustrate the maximum likelihood objective, here we also apply it to the Denoising Diffusion Probabilistic Models (DDPM) (Ho et al., 2020) to mathematically show the connection with diffusion models.

Consider the diffusion process in DDPM:

$$\begin{aligned} q(x_t \mid x_{t-1}) &= \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I), \\ q(x_t \mid x_0) &= \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I), \end{aligned} \tag{45}$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$. Its reverse transition distribution can be derived from Bayes' rule:

$$q(x_{i-1} \mid x_i, x_0) \propto q(x_i \mid x_{i-1}, x_0) q(x_{i-1} \mid x_0). \tag{46}$$

Then we can minimize its negative log-likelihood (NLL) to get the optimal x^*_{t-1}, as our Proposition 3.2. More specifically, set the gradient of the NLL to zero:

$$\begin{aligned} &\nabla_{x^*_{t-1}} \left\{ -\log q\big(x^*_{i-1} \mid x_i, x_0\big) \right\} \\ &= \frac{1-\bar{\alpha}_t}{\beta_t(1-\bar{\alpha}_{t-1})} x^*_{t-1} - \Big(\frac{\sqrt{\alpha_t}}{\beta_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}} x_0\Big) = 0. \end{aligned} \tag{47}$$

Thus, x^*_{t-1} has an optimal value that minimizes the NLL:

$$x^*_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} x_0, \tag{48}$$

which is exactly the reverse mean of DDPM (Eq. (7) in their paper), that guarantees the learning for reverse process.

D. Additional Implementation Details

For all experiments, we use the same noise network: a U-Net similar to DDPM (Chung et al., 2023) but removes all group normalization layers and self-attention layers for inference efficiency. To handle different image sizes, we pad all inputs to make sure outputs could have the same sizes as inputs. The CNN-baseline uses the same network but directly input the low-quality image and output the high-quality image. The stationary variance λ^2 is set to 10 (over 255) and we use only 100 steps for all experiments since the forward process of IR-SDE could be non-Markov (conditioning on x_0), as in DDIM (Song et al., 2021a).

For most tasks, we set the training patch-size to be 128×128 and use a batch size of 16. We use Adam (Kingma & Ba, 2014) optimizer with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The total training steps are fixed to 500 thousand and the initial learning rate set to 10^{-4} and decays half per 200 thousand iterations. All of our models are trained on an A100 GPU with 40GB memory for about 1.5 days (400 000 iterations), the same as for the CNN-baseline.

In addition, we define our θ schedule to be the flipped version to the cosine noise schedule in (Nichol & Dhariwal, 2021):

$$\theta_t = 1 - \frac{f(t)}{f(0)}, \quad f(t) = \cos\left(\frac{t/T + s}{1+s} \cdot \frac{\pi}{2}\right)^2, \tag{49}$$

where $s = 0.008$ as the same as in (Nichol & Dhariwal, 2021). This cosine θ schedule is also visually shown in Figure 12. Once the θ function is determined, we can compute the corresponding diffusion coefficient σ_t by the following stationary condition $\frac{\sigma_t^2}{2\,\theta_t} = \lambda^2$. Practically, we approximate $\bar{\theta}_t$ using a discrete form as $\bar{\theta}_t \approx \sum_{i=1}^{t} \theta_i \Delta t$. To alleviate the over-smooth problem as mentioned in Section 5.4, we let the exponential term $\mathrm{e}^{-\bar{\theta}_T}$ to be a smaller value $\delta = 0.005$ instead of zero and then Δt can also be computed by $\Delta t = -\log \delta / \sum_{i=1}^{T} \theta_i$.

E. Additional Experimental Results

Here we show more detailed results for each task. Note that all reported results of the comparison methods are obtained from using their official codes and pretrained models.

Comparison of losses. We first give the final quantitative results of learning deraining task with the proposed maximum likelihood loss (15) and with the noise matching loss (10) in Table 6. The results show that the maximum likelihood significantly improves the performance over all criteria, which is consistent to Section 5.2 and the result in Figure 10.

Additional quantitative results. Here we give the quantitative results of dehazing in Table 7. Here we only compare with the CNN-baseline since DDRM (Kawar et al., 2022) requires the degradation parameters to be known, which limits its application on image dehazing. The comprehensive results of image denoising on three different test sets over different noise levels are given by Tables 9, 10, and 11. Note we add a SOTA denoising method KBNet (Zhang et al., 2023) on the CBSD68 dataset. The proposed Denoising-ODE has the best perceptual performance for all scenes.

Model complexities. We also provide the comparison of computational efficiency and model complexity in Table 8. our model only slightly increases the parameters and flops of CNN-baseline, while other SOTA methods have to rely on huge computation operations (MAXIM) or complex network structures (KBNet) to improve their performances. We also need to mention that the reverse process involves repeated network evaluations that increases the inference time and computational cost, which is a common limitation of diffusion models. But for training, we only need to sample noises and learn them, which usually converges quickly.

Additional qualitative results. We provide additional results on each task. Specifically, Figure 13, Figure 14, Figure 15, Figure 16, Figure 17, and Figure 18 illustrate visual results on denosing, deraining, deblurring, super-resolution, inpainting, and dehazing, respectively. In most tasks, the results produced by our method are sharper and more realistic. Please zoom in for the best view.

Table 6. Analysis of different losses on deraining task.

METHOD	RAIN100H DATASET				RAIN100L DATASET			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
MAXIMUM LIKELIHOOD LOSS	**30.75**	**0.9027**	**0.048**	**19.76**	**38.30**	**0.9805**	**0.014**	**7.94**
NOISE MATCHING LOSS	23.59	0.7373	0.221	91.49	31.81	0.9313	0.107	52.64

Table 7. Quantitative results of dehazing on the SOTS indoor dataset.

METHOD	DISTORTION		PERCEPTUAL	
	PSNR↑	SSIM↑	LPIPS↓	FID↓
CNN-BASELINE	29.78	0.9683	0.037	34.77
OUR METHOD	**34.14**	**0.9886**	**0.012**	**6.06**

Table 8. Comparison of the number of parameters and model computational efficiency.

METHOD	MAXIM	KBNET	CNN-BASELINE	OURS
#PARAMETERS	14.1M	118.5M	33.8M	34.2M
FLOPS	216G	68.7G	98.0G	98.3G

Table 9. Quantitative results of image denoising on McMaster (Zhang et al., 2011) test set.

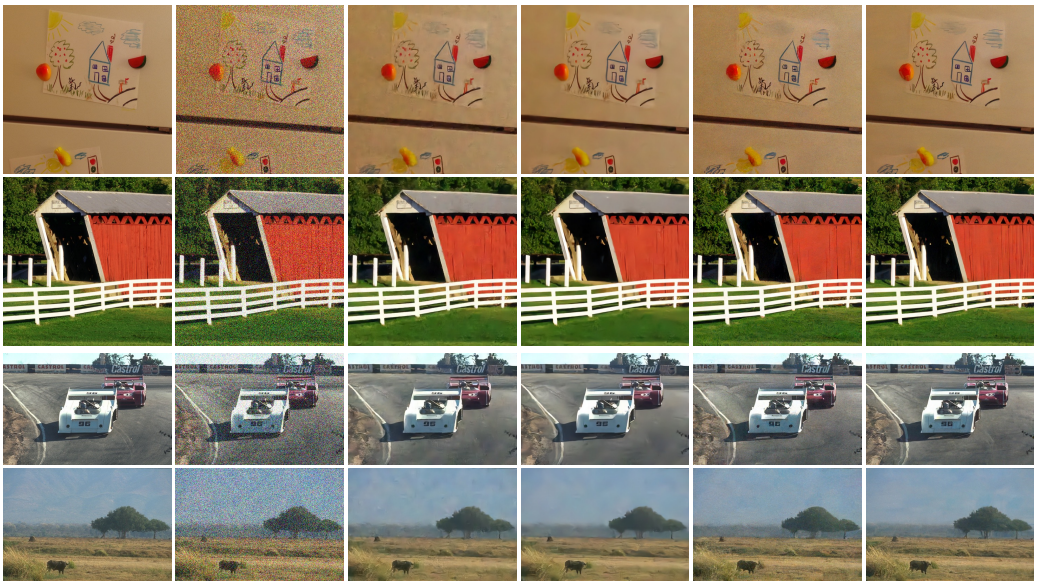
METHOD	$\sigma = 15, t^* = 15$				$\sigma = 25, t^* = 22$				$\sigma = 50, t^* = 39$			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
DNCNN	33.45	0.9035	0.068	37.14	31.52	0.8692	0.101	59.16	28.62	0.7986	0.173	107.31
FFDNET	34.66	**0.9216**	0.065	39.37	32.36	**0.8861**	0.103	63.84	**29.19**	**0.8149**	0.183	118.38
CNN-BASELINE	33.51	0.8978	0.089	43.90	31.79	0.8697	0.122	66.47	29.15	0.8122	0.160	93.68
IR-SDE	31.95	0.8600	0.038	23.97	29.48	0.8052	0.071	44.77	27.14	0.7549	0.151	97.53
DENOISING-ODE	**34.80**	0.9188	**0.036**	**22.03**	**32.39**	0.8791	**0.055**	**34.66**	29.03	0.7911	**0.091**	**63.84**
DENOISING-SDE	31.18	0.8195	0.049	29.21	28.98	0.7512	0.088	45.84	25.85	0.6272	0.173	92.19

Table 10. Quantitative results of image denoising on Kodak24 (Franzen, 1999) test set.

METHOD	$\sigma = 15, t^* = 15$				$\sigma = 25, t^* = 22$				$\sigma = 50, t^* = 39$			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
DNCNN	34.48	0.9189	0.083	21.71	32.02	0.8763	0.129	41.96	28.83	0.7908	0.229	83.27
FFDNET	34.63	**0.9215**	0.085	21.57	32.13	**0.8779**	0.140	44.57	**28.98**	**0.7942**	0.255	89.69
CNN-BASELINE	33.92	0.9090	0.110	24.52	32.73	0.8666	0.161	45.81	28.89	0.7904	0.223	66.01
IR-SDE	31.85	0.8603	0.057	15.25	28.99	0.7772	0.106	35.19	26.83	0.7190	0.208	70.96
DENOISING-ODE	**34.64**	0.9184	**0.050**	**13.74**	**32.14**	0.8739	**0.078**	**21.47**	28.75	0.7746	**0.134**	**45.96**
DENOISING-SDE	30.89	0.8099	0.074	21.09	28.55	0.7247	0.130	36.18	25.46	0.5788	0.249	75.33

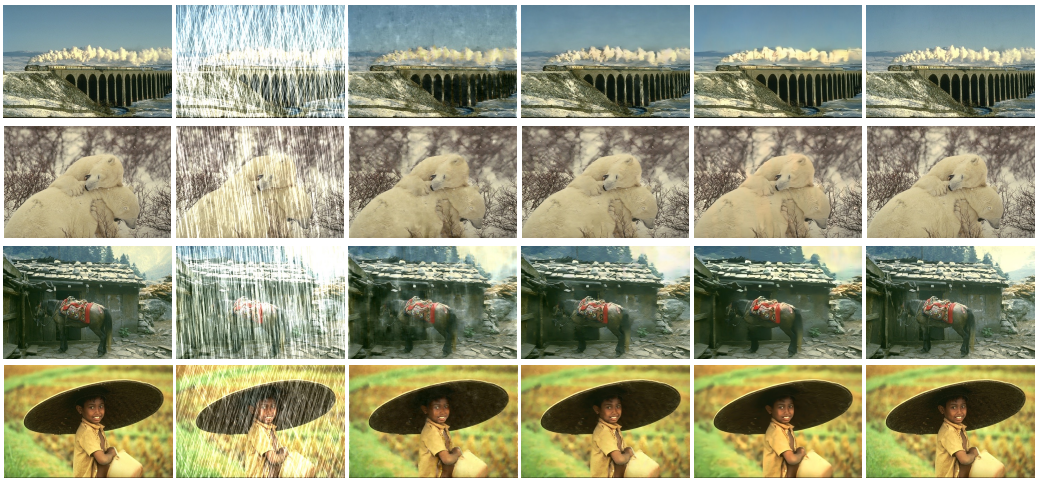
Table 11. Quantitative results of image denoising on CBSD68 (Martin et al., 2001) test set.

METHOD	$\sigma = 15, t^* = 15$				$\sigma = 25, t^* = 22$				$\sigma = 50, t^* = 39$			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
DNCNN	**33.90**	0.9289	0.063	25.59	31.24	0.8830	0.109	43.51	27.95	**0.7896**	0.210	84.56
FFDNET	33.88	**0.9290**	0.065	27.24	31.22	0.8821	0.121	49.64	**27.97**	0.7887	0.244	98.76
KBNET	-	-	-	-	**31.71**	**0.8923**	0.098	37.86	-	-	-	-
CNN-BASELINE	33.02	0.9139	0.098	31.99	30.74	0.8661	0.162	56.64	27.84	0.7827	0.232	78.51
IR-SDE	31.04	0.8708	0.055	21.56	28.09	0.7866	0.101	36.49	25.54	0.6894	0.219	97.95
DENOISING-ODE	33.80	0.9251	**0.042**	**16.71**	31.14	0.8777	**0.074**	**28.71**	27.59	0.7733	**0.138**	**50.46**
DENOISING-SDE	30.15	0.8270	0.078	25.32	27.65	0.457	0.131	39.25	24.37	0.5875	0.243	84.87



GT LQ DnCNN FFDNet IR-SDE Denoising ODE

Figure 13. Visual results of our methods with other denosing approaches on Denoising dataset. The noise level $\sigma = 50$.



GT LQ JORDER PReNet MPRNet IR-SDE

Figure 14. Visual results of our methods with other deraining approaches on Rain100H dataset.

GT LQ DeblurGAN-v2 IR-SDE

Figure 15. Visual results of on Deblurring task.

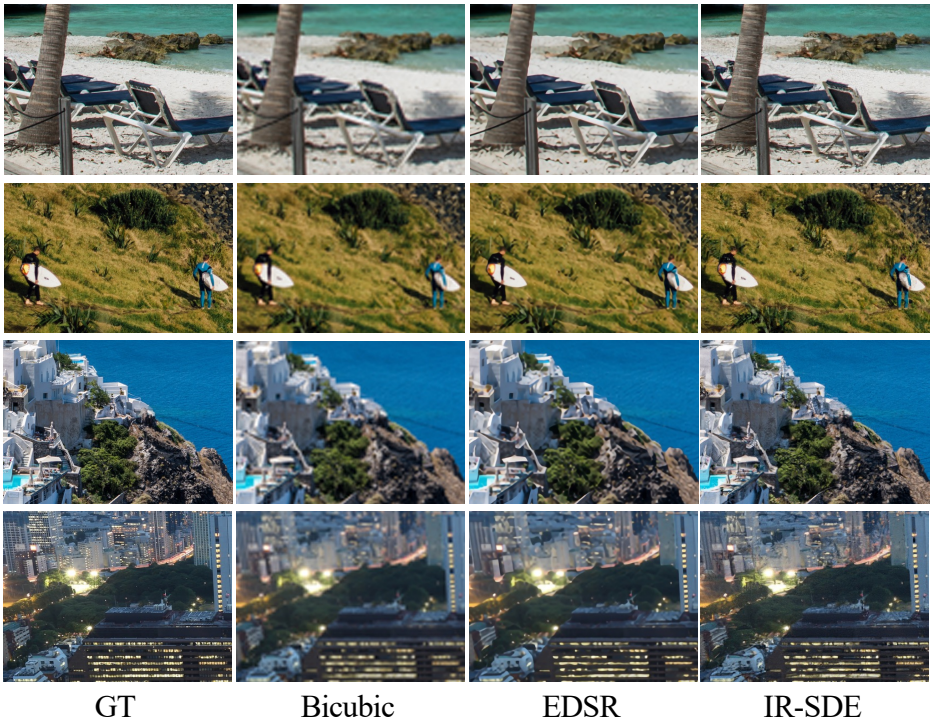


Figure 16. Visual results of our methods on super-resolution task.

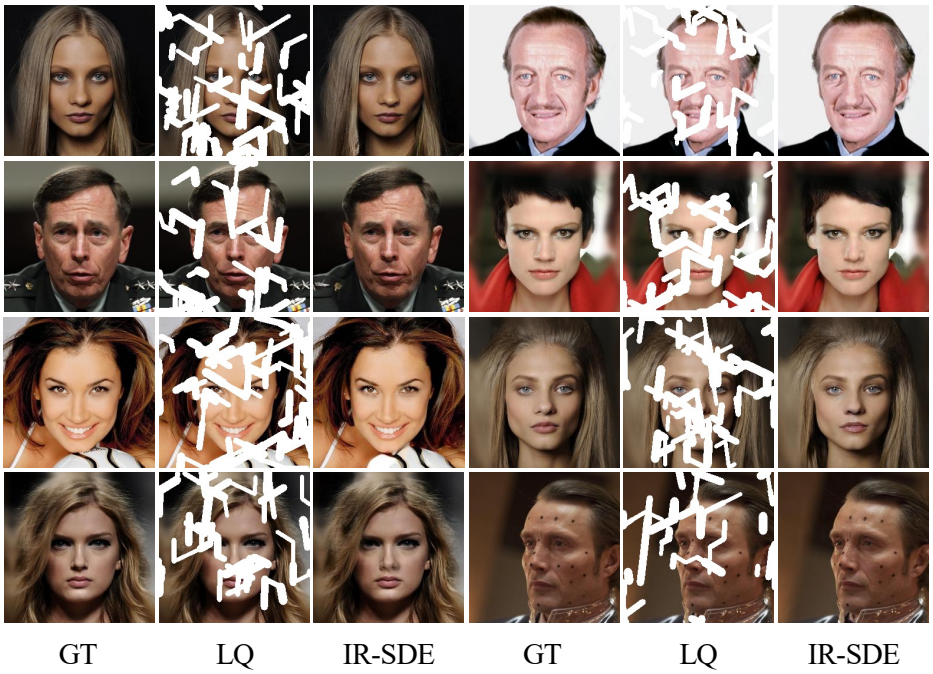


Figure 17. Visual results of our method on inpainting task.

GT LQ IR-SDE GT LQ IR-SDE

Figure 18. Visual results of our method on dehazing task.