Residual Diffusion Bridge Model for Image Restoration

Supplementary Material

A. Doob's h transform

Theorem 1 *For a given SDE:*

$$\mathrm{d}\mathbf{x}_t = \mathbf{f}\left(\mathbf{x}_t, t\right)\mathrm{d}t + g_t\mathrm{d}\mathbf{w}_t, \qquad \mathbf{x}_0 \sim p\left(\mathbf{x}_0\right), \tag{A.1}$$

For a fixed $\mathbf{x}_T$, the evolution of conditional probability $p(\mathbf{x}_t \mid \mathbf{x}_T)$ follows:

$$\mathrm{d}\mathbf{x}_t = \left[\mathbf{f}(\mathbf{x}_t, t) + g_t^2\mathbf{h}(\mathbf{x}_t, t, \mathbf{x}_T, T)\right]\mathrm{d}t + g_t\mathrm{d}\mathbf{w}_t, \qquad \mathbf{x}_0 \sim p\left(\mathbf{x}_0 \mid \mathbf{x}_T\right), \tag{A.2}$$

where $\mathbf{h}(\mathbf{x}_t, t, \mathbf{x}_T, T) = \nabla_{\mathbf{x}_t}\log p(\mathbf{x}_T \mid \mathbf{x}_t)$.

Proof: In theory, $p(\mathbf{x}_t \mid \mathbf{x}_0)$ and $p(\mathbf{x}_T \mid \mathbf{x}_t)$ satisfy the Kolmogorov Forward Equation (KFE) and Kolmogorov Backward Equation (KBE), respectively [55], as formulated below:

$$\frac{\partial}{\partial t}p(\mathbf{x}_t \mid \mathbf{x}_0) = -\nabla_{\mathbf{x}_t}\cdot\left[\mathbf{f}(\mathbf{x}_t, t)p(\mathbf{x}_t \mid \mathbf{x}_0)\right] + \frac{1}{2}g_t^2\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_t \mid \mathbf{x}_0), \tag{A.3}$$

$$-\frac{\partial}{\partial t}p(\mathbf{x}_T \mid \mathbf{x}_t) = \mathbf{f}(\mathbf{x}_t, t)\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t) + \frac{1}{2}g_t^2\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t). \tag{A.4}$$

Using Bayes' rule, we have:

$$\begin{aligned} p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T) &= \frac{p(\mathbf{x}_T \mid \mathbf{x}_t, \mathbf{x}_0)p(\mathbf{x}_t \mid \mathbf{x}_0)}{p(\mathbf{x}_T \mid \mathbf{x}_0)} \\ &= \frac{p(\mathbf{x}_T \mid \mathbf{x}_t)p(\mathbf{x}_t \mid \mathbf{x}_0)}{p(\mathbf{x}_T \mid \mathbf{x}_0)} \end{aligned} \tag{A.5}$$

Therefore, the derivative of conditional transition probability $p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)$ with time follows:

$$\begin{aligned} \frac{\partial}{\partial t}p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T) &= \frac{p(\mathbf{x}_t \mid \mathbf{x}_0)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\frac{\partial}{\partial t}p(\mathbf{x}_T \mid \mathbf{x}_t) + \frac{p(\mathbf{x}_T \mid \mathbf{x}_t)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\frac{\partial}{\partial t}p(\mathbf{x}_t \mid \mathbf{x}_0) \\ &= \frac{p(\mathbf{x}_t \mid \mathbf{x}_0)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\left[-\mathbf{f}(\mathbf{x}_t, t)\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t) - \frac{1}{2}g_t^2\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t)\right] \\ &\quad + \frac{p(\mathbf{x}_T \mid \mathbf{x}_t)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\left\{-\nabla_{\mathbf{x}_t}\cdot\left[\mathbf{f}(\mathbf{x}_t, t)p(\mathbf{x}_t \mid \mathbf{x}_0)\right] + \frac{1}{2}g_t^2\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_t \mid \mathbf{x}_0)\right\} \\ &= -\left[\frac{p(\mathbf{x}_t \mid \mathbf{x}_0)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\mathbf{f}(\mathbf{x}_t, t)\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t) + \frac{p(\mathbf{x}_T \mid \mathbf{x}_t)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\mathbf{f}(\mathbf{x}_t, t)\nabla_{\mathbf{x}_t}p(\mathbf{x}_t \mid \mathbf{x}_0)\right. \\ &\quad \left. + \frac{p(\mathbf{x}_T \mid \mathbf{x}_t)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}p(\mathbf{x}_t \mid \mathbf{x}_0)\nabla_{\mathbf{x}_t}\cdot\mathbf{f}(\mathbf{x}_t, t)\right] \\ &\quad + \frac{1}{2}g_t^2\left[\frac{p(\mathbf{x}_T \mid \mathbf{x}_t)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_t \mid \mathbf{x}_0) - \frac{p(\mathbf{x}_t \mid \mathbf{x}_0)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t)\right] \\ &= -\left[\mathbf{f}(\mathbf{x}_t, t)\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T) + p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)\cdot\nabla_{\mathbf{x}_t}\mathbf{f}(\mathbf{x}_t, t)\right] \\ &\quad + \frac{1}{2}g_t^2\left[\frac{p(\mathbf{x}_T \mid \mathbf{x}_t)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_t \mid \mathbf{x}_0) - \frac{p(\mathbf{x}_t \mid \mathbf{x}_0)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t)\right] \\ &= -\nabla_{\mathbf{x}_t}\cdot\left[\mathbf{f}(\mathbf{x}_t, t)p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)\right] \\ &\quad + \frac{1}{2}g_t^2\left[\frac{p(\mathbf{x}_T \mid \mathbf{x}_t)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_t \mid \mathbf{x}_0) - \frac{p(\mathbf{x}_t \mid \mathbf{x}_0)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t)\right] \end{aligned} \tag{A.6}$$

For the second term, we have:

$$
\begin{aligned}
&\frac{1}{2}g_t^2\left[\frac{p(\mathbf{x}_T \mid \mathbf{x}_t)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_t \mid \mathbf{x}_0) - \frac{p(\mathbf{x}_t \mid \mathbf{x}_0)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t)\right] \\
=&\frac{1}{2}g_t^2\left[\frac{p(\mathbf{x}_T \mid \mathbf{x}_t)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_t \mid \mathbf{x}_0) + \frac{1}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t)\cdot\nabla_{\mathbf{x}_t}\, p(\mathbf{x}_t \mid \mathbf{x}_0)\right. \\
&\left.+\frac{1}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t)\cdot\nabla_{\mathbf{x}_t}\, p(\mathbf{x}_t \mid \mathbf{x}_0) + \frac{p(\mathbf{x}_t \mid \mathbf{x}_0)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t)\right] \\
&- g_t^2\left[\frac{1}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t)\cdot\nabla_{\mathbf{x}_t}\, p(\mathbf{x}_t \mid \mathbf{x}_0) + \frac{p(\mathbf{x}_t \mid \mathbf{x}_0)}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t)\right] \\
=&\frac{1}{2}g_t^2\left[\frac{1}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot[p(\mathbf{x}_T \mid \mathbf{x}_t)\nabla_{\mathbf{x}_t}p(\mathbf{x}_t \mid \mathbf{x}_0)] + \frac{1}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot[p(\mathbf{x}_t \mid \mathbf{x}_0)\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t)]\right] \\
&- g_t^2\frac{1}{p(\mathbf{x}_T \mid \mathbf{x}_0)}\nabla_{\mathbf{x}_t}\cdot[p(\mathbf{x}_t \mid \mathbf{x}_0)\nabla_{\mathbf{x}_t}p(\mathbf{x}_T \mid \mathbf{x}_t)] \\
=&\frac{1}{2}g_t^2\left[\nabla_{\mathbf{x}_t}\cdot[p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t \mid \mathbf{x}_0)] + \nabla_{\mathbf{x}_t}\cdot[p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_T \mid \mathbf{x}_t)]\right] \\
&- g_t^2\nabla_{\mathbf{x}_t}\cdot[p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_T \mid \mathbf{x}_t)] \\
=&\frac{1}{2}g_t^2\left[\nabla_{\mathbf{x}_t}\cdot[p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)]\right] - g_t^2\nabla_{\mathbf{x}_t}\cdot[p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_T \mid \mathbf{x}_t)] \\
=&\frac{1}{2}g_t^2\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T) - g_t^2\nabla_{\mathbf{x}_t}\cdot[p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_T \mid \mathbf{x}_t)]
\end{aligned}
\tag{A.7}
$$

Bring it back to (A.6):

$$
\begin{aligned}
\frac{\partial}{\partial t}p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T) &= -\nabla_{\mathbf{x}_t}\cdot[\mathbf{f}(\mathbf{x}_t, t)p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)] + \frac{1}{2}g_t^2\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T) \\
&\quad - g_t^2\nabla_{\mathbf{x}_t}\cdot[p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_T \mid \mathbf{x}_t)] \\
&= -\nabla_{\mathbf{x}_t}\cdot\left[[\mathbf{f}(\mathbf{x}_t, t) + g_t^2\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_T \mid \mathbf{x}_t)]p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)\right] + \frac{1}{2}g_t^2\nabla_{\mathbf{x}_t}\cdot\nabla_{\mathbf{x}_t}p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)
\end{aligned}
\tag{A.8}
$$

This is the definition of FP equation of conditional transition probability $p(\mathbf{x}_t \mid \mathbf{x}_0, \mathbf{x}_T)$, which represents the evolution that follows the SDE:

$$
\mathrm{d}\mathbf{x}_t = \left[\mathbf{f}(\mathbf{x}_t, t) + g_t^2\nabla_{\mathbf{x}_t}\log p(\mathbf{x}_T \mid \mathbf{x}_t)\right]\mathrm{d}t + g_t\mathrm{d}\mathbf{w}_t \tag{A.9}
$$

This concludes the proof of the **Theorem 1** in Sec. 3.1.

B. Mean-Reverting Ornstein–Uhlenbeck Process

Theorem 2 *The SDE formulation of the Ornstein–Uhlenbeck process with its predefined coefficients θ_t, σ_t is:*

$$
d\mathbf{x}_t = \theta_t(\boldsymbol{\mu} - \mathbf{x}_t)dt + \sigma_t dw_t, \tag{B.1}
$$

where $\boldsymbol{\mu}$ represents the mean value that $\mathbf{x}_t$ will approximate at $t = T$. The solution of OU process can be calculated as:

$$
\mathbf{x}_t = \boldsymbol{\mu} + (x_0 - \boldsymbol{\mu})e^{-\int_0^t \theta_s ds} + e^{-\int_0^t \theta_s ds}\int_0^t \sigma_s e^{\int_0^s \theta_u du}dw_s, \tag{B.2}
$$

Proof. We define a surrogate differentiable function $\psi(\mathbf{x}, t) = \mathbf{x}e^{\int_0^t \theta_z dz} = \mathbf{x}e^{\overline{\theta}_t}$ and expand it by $Itô$ formula:

$$
\begin{aligned}
d\psi(\mathbf{x}, t) &= \frac{\partial\psi}{\partial t}(\mathbf{x}, t)dt + \frac{\partial\psi}{\partial\mathbf{x}}(\mathbf{x}, t)d\mathbf{x} + \frac{1}{2}\frac{\partial^2\psi}{\partial\mathbf{x}^2}(\mathbf{x}, t)d\mathbf{x}^2 \\
&= \theta_t\mathbf{x}e^{\overline{\theta}_t}dt + e^{\overline{\theta}_t}(\theta_t(\boldsymbol{\mu} - \mathbf{x})dt + \sigma_t dw_t) \\
&= \boldsymbol{\mu}\theta_t e^{\overline{\theta}_t}dt + \sigma_t e^{\overline{\theta}_t}dw_t
\end{aligned}
\tag{B.3}
$$

Then, we can solve $\mathbf{x}_t$ conditioned on $\mathbf{x}_s$ where $s < t$, as:

$$\psi(\mathbf{x}_t, t) - \psi(\mathbf{x}_s, s) = \int_s^t \boldsymbol{\mu}\theta_z e^{\overline{\theta}_z} dz + \int_s^t \sigma_z e^{\overline{\theta}_z} dw_z, \tag{B.4}$$

$$\mathbf{x}_t e^{\overline{\theta}_t} - \mathbf{x}_s e^{\overline{\theta}_s} = \boldsymbol{\mu}(e^{\overline{\theta}_t} - e^{\overline{\theta}_s}) + \int_s^t \sigma_z e^{\overline{\theta}_z} dw_z, \tag{B.5}$$

$$\mathbf{x}_t = \boldsymbol{\mu} + (\mathbf{x}_s - \boldsymbol{\mu}) e^{-\overline{\theta}_{s:t}} + \int_s^t \sigma_z e^{-\overline{\theta}_{z:t}} dw_z. \tag{B.6}$$

where $-\overline{\theta}_{s:t} = -\int_s^t \theta_z dz$, and thus we complete the proof. The expectation and variance of Eq. (B.6) can be rewritten:

$$E[x_t] = \boldsymbol{\mu} + (x_s - \boldsymbol{\mu}) e^{-\overline{\theta}_{s:t}}, \tag{B.7}$$

$$Var[x_t] = \int_s^t \sigma_z^2 e^{-2\overline{\theta}_{z:t}} dz, \tag{B.8}$$

This concludes the derivations in Sec. 3.2.

C. RDBM Formulation

Proposition 1: Let $\mathbf{x}_t$ be a finite random variable governed by the generalized OU process, with terminal condition $\mathbf{x}_T = \boldsymbol{\mu}$. The evolution of its marginal distribution $p(\mathbf{x}_t \mid \mathbf{x}_T)$ satisfies the following SDE under a fixed drift-to-diffusion coefficient ratio λ:

$$d\mathbf{x}_t = \theta_t \coth(\overline{\theta}_{t:T})(\boldsymbol{\mu} - \mathbf{x}_t)dt + \sqrt{2\boldsymbol{\pi}^2 \lambda \theta_t} dw_t, \tag{C.1}$$

where $\overline{\theta}_{t:T} = \int_t^T \theta_z dz$ and $\boldsymbol{\pi} \in \mathbb{R}$ is the predefined parameter.
Proof: First, we define a generalized OU process with the properties of mean-reverting:

$$d\mathbf{x}_t = \theta_t(\boldsymbol{\mu} - \mathbf{x}_t)dt + \boldsymbol{\pi}\sigma_t dw_t. \tag{C.2}$$

In our formulation, $\boldsymbol{\pi} = \boldsymbol{\mu} - \mathbf{x}_0$ is considered as the residual of given distributions. When $\boldsymbol{\pi} = 1$ or $\boldsymbol{\pi} = 0$, Eq. (C.2) can degenerate to other bridge models, as discussed in Suppl. F. Here, we solve this SDE step by step, akin to Suppl. B. First, we define a surrogate differentiable function $\psi(\mathbf{x}, t) = \mathbf{x} e^{\int_0^t \theta_z dz} = \mathbf{x} e^{\overline{\theta}_t}$ and expand it by $Itô$ formula:

$$\begin{aligned} d\psi(\mathbf{x}, t) &= \frac{\partial \psi}{\partial t}(\mathbf{x}, t)dt + \frac{\partial \psi}{\partial \mathbf{x}}(\mathbf{x}, t)d\mathbf{x} + \frac{1}{2}\frac{\partial^2 \psi}{\partial \mathbf{x}^2}(\mathbf{x}, t)d\mathbf{x}^2 \\ &= \theta_t \mathbf{x} e^{\overline{\theta}_t} dt + e^{\overline{\theta}_t}(\theta_t(\boldsymbol{\mu} - \mathbf{x})dt + \boldsymbol{\pi}\sigma_t dw_t) \\ &= \boldsymbol{\mu}\theta_t e^{\overline{\theta}_t} dt + \boldsymbol{\pi}\sigma_t e^{\overline{\theta}_t} dw_t \end{aligned} \tag{C.3}$$

Then, we can solve $\mathbf{x}_t$ conditioned on $\mathbf{x}_s$ where $s < t$, as:

$$\psi(\mathbf{x}_t, t) - \psi(\mathbf{x}_s, s) = \int_s^t \boldsymbol{\mu}\theta_z e^{\overline{\theta}_z} dz + \int_s^t \boldsymbol{\pi}\sigma_z e^{\overline{\theta}_z} dw_z, \tag{C.4}$$

$$\mathbf{x}_t e^{\overline{\theta}_t} - \mathbf{x}_s e^{\overline{\theta}_s} = \boldsymbol{\mu}(e^{\overline{\theta}_t} - e^{\overline{\theta}_s}) + \int_s^t \boldsymbol{\pi}\sigma_z e^{\overline{\theta}_z} dw_z, \tag{C.5}$$

$$\mathbf{x}_t = \boldsymbol{\mu} + (\mathbf{x}_s - \boldsymbol{\mu}) e^{-\overline{\theta}_{s:t}} + \int_s^t \boldsymbol{\pi}\sigma_z e^{-\overline{\theta}_{z:t}} dw_z. \tag{C.6}$$

where $-\overline{\theta}_{s:t} = -\int_s^t \theta_z dz$. The expectation and variance of Eq. (C.6) can be written as below:

$$E[\mathbf{x}_t] = \boldsymbol{\mu} + (\mathbf{x}_s - \boldsymbol{\mu}) e^{-\overline{\theta}_{s:t}}, \tag{C.7}$$

$$Var[\mathbf{x}_t] = \boldsymbol{\pi}^2 \int_s^t \sigma_z^2 e^{-2\overline{\theta}_{z:t}} dz, \tag{C.8}$$

To derive the analytical form of Eq. (C.8), we assume that $\lambda = \frac{\sigma_t^2}{2\theta_t}$ is pre-defined stationary variance, and obtain:

$$Var[\mathbf{x}_t] = \lambda \boldsymbol{\pi}^2 \int_s^t 2\theta_z e^{-2\overline{\theta}_{z:t}} dz = \lambda \boldsymbol{\pi}^2 (1 - e^{-2\overline{\theta}_{s:t}}), \tag{C.9}$$

We can conclude that:

$$p(\mathbf{x}_t|\mathbf{x}_s) \sim \mathcal{N}(\boldsymbol{\mu} + (\mathbf{x}_s - \boldsymbol{\mu}) e^{-\overline{\theta}_{s:t}}, \lambda \boldsymbol{\pi}^2 (1 - e^{-2\overline{\theta}_{s:t}})), \tag{C.10}$$

To ensure that the final state of time point $t = T$ conforms to the distribution of low-quality image $\mathbf{x}_T = \boldsymbol{\mu} \sim p_{LQ}(\boldsymbol{x})$, we leverage the Doob's h transform by modifying the forward SDE from Eq. (C.11) to Eq. (C.12):

$$d\mathbf{x}_t = \mathbf{f}(\mathbf{x}, t)dt + g(t)dw_t, \tag{C.11}$$

$$d\mathbf{x}_t = [\mathbf{f}(\mathbf{x}, t) + g(t)^2 \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_T|\mathbf{x}_t)]dt + g(t)dw_t, \tag{C.12}$$

where term $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_T|x_t)$ can be calculated by setting $s = 0, t = T$ in Eq. (C.10):

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_T|\mathbf{x}_t) = (\boldsymbol{\mu} - \mathbf{x}_t) \frac{e^{-2\overline{\theta}_{t:T}}}{\lambda \boldsymbol{\pi}^2 (1 - e^{-2\overline{\theta}_{t:T}})}. \tag{C.13}$$

The mean-reverting OU process turns into a mean-arriving process, which can be formulated as:

$$\begin{aligned}
d\mathbf{x}_t &= (\theta_t + \frac{\sigma_t^2 e^{-2\overline{\theta}_{t:T}}}{\lambda(1 - e^{-2\overline{\theta}_{t:T}})})(\boldsymbol{\mu} - \mathbf{x}_t)dt + \boldsymbol{\pi}\sigma_t dw_t, \\
&= \theta_t (1 + \frac{2e^{-2\overline{\theta}_{t:T}}}{1 - e^{-2\overline{\theta}_{t:T}}})(\boldsymbol{\mu} - \mathbf{x}_t)dt + \boldsymbol{\pi}\sigma_t dw_t \\
&= \theta_t (\frac{1 + e^{-2\overline{\theta}_{t:T}}}{1 - e^{-2\overline{\theta}_{t:T}}})(\boldsymbol{\mu} - \mathbf{x}_t)dt + \boldsymbol{\pi}\sigma_t dw_t, \\
&= \theta_t \coth(\overline{\theta}_{t:T})(\boldsymbol{\mu} - \mathbf{x}_t)dt + \boldsymbol{\pi}\sigma_t dw_t, (\sigma_t^2 = 2\lambda\theta_t), \\
&= \theta_t \coth(\overline{\theta}_{t:T})(\boldsymbol{\mu} - \mathbf{x}_t)dt + \sqrt{2\boldsymbol{\pi}^2 \lambda \theta_t} dw_t,
\end{aligned} \tag{C.14}$$

Eq. (C.14) can be converted into an analytical formula as follows. First, we substitute $y_t = \mathbf{x}_t - \boldsymbol{\mu}$, then the SDE of y_t becomes:

$$dy_t = -\theta_t \coth(\overline{\theta}_{t:T}) y_t dt + \sqrt{2\boldsymbol{\pi}^2 \lambda \theta_t} dw_t, \tag{C.15}$$

Second, we introduce $\Psi_t = \exp(\int_0^t \theta_s \coth(\overline{\theta}_{s:T})ds)$ as the integrating factor and expand $\Psi_t y_t$ by Itô formula:

$$d(\Psi_t y_t) = \Psi_t dy_t + y_t d\Psi_t + d\Psi_t dy_t, \tag{C.16}$$

Since Ψ is a deterministic function, it satisfies $d\Psi = \Psi\theta_t \coth(\overline{\theta}_{t:T})$. $d\Psi dy_t$ produces $(dt)^2, dtdw_t$, which are the higher order infinitesimal of dt and can be omitted. Thus, we obtain:

$$d(\Psi_t y_t) = \Psi_t(-\theta_t \coth(\overline{\theta}_{t:T}) y_t dt + \sqrt{2\boldsymbol{\pi}^2 \lambda \theta_t} dw_t) + \Psi_t \theta_t \coth(\overline{\theta}_{t:T}) y_t dt = \Psi_t \sqrt{2\boldsymbol{\pi}^2 \lambda \theta_t} dw_t, \tag{C.17}$$

Furthermore, we integrate both sides of Eq. (C.17):

$$\Psi_t y_t = y_0 + \int_0^t \Psi_s \sqrt{2\boldsymbol{\pi}^2 \lambda \theta_s} dw_s. \tag{C.18}$$

Consequently, we have:

$$\mathbf{x}_t = \boldsymbol{\mu} + (\mathbf{x}_0 - \boldsymbol{\mu}) \Psi_t^{e^{-\int_0^t \theta_s \coth(\overline{\theta}_{s:T})ds}} + \int_0^t \sqrt{2\boldsymbol{\pi}^2 \lambda \theta_s} e^{-\int_s^t \theta_z \coth(\overline{\theta}_{z:T})dz} dw_s. \tag{C.19}$$

We next analyze the analytical formulation of Ψ_t. Considering the internal integral $\int_0^t \theta_s \coth(\overline{\theta}_{s:T})ds$ at first, we set $u = \overline{\theta}_{s:T}$ satisfying $du = -\theta_s ds$:

$$\int_0^t \theta_s \coth(\overline{\theta}_{s:T})ds = -\int_{\overline{\theta}_{0:T}}^{\overline{\theta}_{t:T}} \coth(u)du = -\ln|\sinh(u)|\Big|_{\overline{\theta}_{0:T}}^{\overline{\theta}_{t:T}} = \ln|\frac{\sinh(\overline{\theta}_{0:T})}{\sinh(\overline{\theta}_{t:T})}|, \tag{C.20}$$

Therefore, the analytical expression of Ψ_t is:

$$\Psi_t = \frac{\sinh(\overline{\theta}_{0:T})}{\sinh(\overline{\theta}_{t:T})}. \tag{C.21}$$

Finally, we can compute the closed-form of $\mathbf{x}_t$ in Eq. (C.19):

$$\mathbf{x}_t = \boldsymbol{\mu} + (\mathbf{x}_0 - \boldsymbol{\mu})\frac{\sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{0:T})} + \int_0^t \sqrt{2\boldsymbol{\pi}^2\lambda\theta_s}\frac{\sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{s:T})}dw_s. \tag{C.22}$$

Eq. (C.22) preserves the properties of diffusion bridge models, whose initial state $\mathbf{x}_0$ and final state $\mathbf{x}_T$ are determined. The formulation of variance can be further simplified as follows:

$$\begin{aligned} Var[x_t] = \int_0^t 2\boldsymbol{\pi}^2\lambda\theta_s(\frac{\sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{s:T})})^2 ds &= 2\boldsymbol{\pi}^2\lambda\sinh^2(\overline{\theta}_{t:T})\int_{\overline{\theta}_{0:T}}^{\overline{\theta}_{t:T}} -\frac{du}{\sinh^2(u)} = 2\boldsymbol{\pi}^2\lambda\sinh^2(\overline{\theta}_{t:T})\coth(u)\Big|_{\overline{\theta}_{0:T}}^{\overline{\theta}_{t:T}} \\ &= 2\boldsymbol{\pi}^2\lambda\sinh^2(\overline{\theta}_{t:T})(\coth(\overline{\theta}_{t:T}) - \coth(\overline{\theta}_{0:T})) \\ &= 2\boldsymbol{\pi}^2\lambda\sinh^2(\overline{\theta}_{t:T})(\frac{\sinh(\overline{\theta}_{0:T} - \overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{0:T})\sinh(\overline{\theta}_{t:T})}) \\ &= 2\boldsymbol{\pi}^2\lambda\frac{\sinh(\overline{\theta}_{0:t})\sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{0:T})} \end{aligned} \tag{C.23}$$

This concludes the derivations in Sec. 4.1. The expectation and variance of Eq. (C.22) are summarized as follows:

$$E[\mathbf{x}_t] = \boldsymbol{\mu} + (\mathbf{x}_0 - \boldsymbol{\mu})\frac{\sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{0:T})}, \tag{C.24}$$

$$Var[\mathbf{x}_t] = 2\boldsymbol{\pi}^2\lambda\frac{\sinh(\overline{\theta}_{0:t})\sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{0:T})}. \tag{C.25}$$

D. In-depth analysis of $\boldsymbol{\pi}$ selection

Rethinking the diffusion process. Mainstream diffusion models perturb the entire image with Gaussian noise and then perform pixel-wise reconstruction, aiming to handle noise corruption and recover high-quality information in parallel. For global degradations (e.g., low-light, noise), this approach achieves favorable performance by leveraging the known distribution of noise to guide the restoration of missing details. However, for mask-based degradations (e.g., rain, snow), only the degraded regions require restoration, while unaffected areas remain nearly identical to high-quality images. This approach introduces additional task complexity, which not only enables recovery of degraded regions, but also simultaneously compromises the quality of intact areas through redundant reconstruction. Moreover, severely degraded regions (with limited preserved information) benefit from enhanced noise perturbation to facilitate reconstruction, while mildly degraded regions require noise suppression to retain valid information. Drawing from the above analyses, $\boldsymbol{\pi}$ should possess weighted masking properties, effectively equivalent to the image residual $\boldsymbol{\pi} = \mathbf{x}_T - \mathbf{x}_0$.

Power analysis. Eq. (C.22) reveals that the diffusion process is determined by two terms given the final state $\mathbf{x}_T = \boldsymbol{\mu}$. The power ratio between residual component and noise component at pixel i, j can be defined as residual-to-noise ratio $R(t, i, j)$:

$$R(t,i,j) = \frac{(\mathbf{x}_T(i,j) - \mathbf{x}_0(i,j))^2}{2\boldsymbol{\pi}^2(i,j)\lambda}\frac{(\frac{\sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{0:T})})^2}{\frac{\sinh(\overline{\theta}_{0:t})\sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{0:T})}} = \frac{(\mathbf{x}_T(i,j) - \mathbf{x}_0(i,j))^2}{2\boldsymbol{\pi}^2(i,j)\lambda}\frac{\sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{0:t})\sinh(\overline{\theta}_{0:T})}, \tag{D.1}$$

The first part is determined by predefined parameters $\boldsymbol{\pi}, \lambda$ given initial and final states. The second part is entirely determined by the sequence of θ_t values, which approaches infinity at time 0 and converges to infinitesimal at time T. If $\boldsymbol{\pi}$ is a globally

predefined parameter, when pixel residual $(\mathbf{x}_T(i,j) - \mathbf{x}_0(i,j))$ approaches zero, $R(t,i,j) \to \infty$. In this context, the high-quality regions are disrupted by noise and cannot be perfectly reconstructed due to the predicted error. Besides, the refinement of low-quality regions with varying degradation degrees is dominated by their respective residual magnitudes. To make the $R(t,i,j)$ smooth, we leverage the setting of $\boldsymbol{\pi} = \mathbf{x}_T - \mathbf{x}_0$, and obtain:

$$R(t,i,j) = R(t) = \frac{\sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{0:t})\sinh(\overline{\theta}_{0:T})}. \tag{D.2}$$

Let us check the monotonic properties of $R(t)$ by its logarithm derivatives:

$$A(t) = \sinh(\overline{\theta}_{t:T}), \qquad B(t) = \sinh(\overline{\theta}_{0:t}), \qquad C = \sinh(\overline{\theta}_{0:T}) \tag{D.3}$$

$$A'(t) = \cosh(\overline{\theta}_{t:T}) \cdot \frac{d}{dt}\overline{\theta}_{t:T} = -\theta_t \cosh(\overline{\theta}_{t:T}), B'(t) = \cosh(\overline{\theta}_{0:t}) \cdot \frac{d}{dt}\overline{\theta}_{0:t} = \theta_t \cosh(\overline{\theta}_{t:T}), \tag{D.4}$$

$$\frac{dR(t)}{dt} = \frac{A'(t)B(t) - A(t)B'(t)}{B^2(t)C} = -\theta_t \frac{\cosh(\overline{\theta}_{t:T})\sinh(\overline{\theta}_{0:t}) + \sinh(\overline{\theta}_{t:T})\cosh(\overline{\theta}_{0:t})}{\sinh^2(\overline{\theta}_{0:t})\sinh(\overline{\theta}_{0:T})} = -\frac{\theta_t}{\sinh^2(\overline{\theta}_{0:t})}. \tag{D.5}$$

For the typical condition that $\theta_t \geq 0$, and $\sinh^2(\overline{\theta}_{t:T}) \geq 0$, we can conclude that $R(t)$ is a monotonically decreasing function starting from $R(0) \to \infty$ to $R(T) \to 0$, as $\frac{d}{dt}R(t) \leq 0$. It can be observed that if $\boldsymbol{\pi} = \mathbf{x}_T - \mathbf{x}_0$ is set, $R(t)$ decreases evenly for each pixel without being affected by image contents. Hence, we set $\boldsymbol{\pi}$ as residual component in Sec. 4.1.

E. Process of Reverse Inference

For simplicity, we use Θ_t and Σ_t to represent the coefficients in Eq. (C.24) and Eq. (C.25), respectively. We have:

$$\Theta_t \equiv \frac{\sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{0:T})}, \quad \Sigma_t \equiv 2\lambda \frac{\sinh(\overline{\theta}_{0:t})\sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{0:T})} \tag{E.1}$$

DDPM Reverse Process. Leveraging the properties of Bayesian formula, we obtain:

$$p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0, \mathbf{x}_T) = \frac{p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0, \mathbf{x}_T)p(\mathbf{x}_{t-1}|\mathbf{x}_0, \mathbf{x}_T)}{p(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T)}, \tag{E.2}$$

$$\mathbf{x}_{t-1} = \boldsymbol{\mu} + (\mathbf{x}_0 - \boldsymbol{\mu})\Theta_{t-1} + \boldsymbol{\pi}\Sigma_{t-1}\epsilon_{t-1}, \tag{E.3}$$

$$\mathbf{x}_t = \boldsymbol{\mu} + (\mathbf{x}_0 - \boldsymbol{\mu})\Theta_t + \boldsymbol{\pi}\Sigma_t\epsilon_t. \tag{E.4}$$

Eliminating the variable $\mathbf{x}_0$, we have:

$$\mathbf{x}_t = \boldsymbol{\mu} + \Theta_t \frac{\mathbf{x}_{t-1} - \boldsymbol{\mu} - \Sigma_{t-1}\epsilon_{t-1}}{\Theta_{t-1}} + \boldsymbol{\pi}\Sigma_t\epsilon_t \tag{E.5}$$

$$= \boldsymbol{\mu} + \frac{\Theta_t}{\Theta_{t-1}}(\mathbf{x}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\pi}\sqrt{\Sigma_t^2 - \frac{\Theta_t^2}{\Theta_{t-1}^2}\Sigma_{t-1}^2}\epsilon \tag{E.6}$$

Back to Eq. (E.2), we have:

$$\begin{aligned}
&\log p(\mathbf{x}_{t-1}|\mathbf{x}_0, \mathbf{x}_t, \mathbf{x}_T) = \log p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0, \mathbf{x}_T) + \log p(\mathbf{x}_{t-1}|\mathbf{x}_0, \mathbf{x}_T) - \log p(\mathbf{x}_t|\mathbf{x}_0, \mathbf{x}_T) \\
&\propto -\frac{1}{2\boldsymbol{\pi}^2}\left[\frac{(\mathbf{x}_t - \boldsymbol{\mu} - \frac{\Theta_t}{\Theta_{t-1}}(\mathbf{x}_{t-1} - \boldsymbol{\mu}))^2}{\Sigma_t^2 - \frac{\Theta_t^2}{\Theta_{t-1}^2}\Sigma_{t-1}^2} + \frac{(\mathbf{x}_{t-1} - \boldsymbol{\mu} - (\mathbf{x}_0 - \boldsymbol{\mu})\Theta_{t-1})^2}{\Sigma_{t-1}^2} - \frac{(\mathbf{x}_t - \boldsymbol{\mu} - (\mathbf{x}_0 - \boldsymbol{\mu})\Theta_t)^2}{\Sigma_t^2}\right] \\
&= -\frac{1}{2\boldsymbol{\pi}^2}\left[\frac{(\mathbf{x}_{t-1} - \boldsymbol{\mu} - \frac{\Theta_{t-1}}{\Theta_t}(\mathbf{x}_t - \boldsymbol{\mu}))^2}{\frac{\Theta_{t-1}^2}{\Theta_t^2}\Sigma_t^2 - \Sigma_{t-1}^2} + \frac{(\mathbf{x}_{t-1} - \boldsymbol{\mu} - (\mathbf{x}_0 - \boldsymbol{\mu})\Theta_{t-1})^2}{\Sigma_{t-1}^2} + C\right] \\
&= -\frac{1}{2\boldsymbol{\pi}^2}\left[\frac{(\mathbf{x}_{t-1}^2 - 2(\boldsymbol{\mu} + \frac{\Theta_{t-1}}{\Theta_t}(\mathbf{x}_t - \boldsymbol{\mu}))\mathbf{x}_{t-1} + (\boldsymbol{\mu} + \frac{\Theta_{t-1}}{\Theta_t}(\mathbf{x}_t - \boldsymbol{\mu}))^2}{\frac{\Theta_{t-1}^2}{\Theta_t^2}\Sigma_t^2 - \Sigma_{t-1}^2}\right. \\
&\qquad\qquad \left. + \frac{(\mathbf{x}_{t-1}^2 - 2(\boldsymbol{\mu} + (\mathbf{x}_0 - \boldsymbol{\mu})\Theta_{t-1})\mathbf{x}_{t-1} + (\boldsymbol{\mu} + (\mathbf{x}_0 - \boldsymbol{\mu})\Theta_{t-1})^2}{\Sigma_{t-1}^2} + C\right]
\end{aligned} \tag{E.7}$$

Furthermore, all the terms not related to $\mathbf{x}_{t-1}$ are categorized as C:

$$\begin{aligned}\log p(\mathbf{x}_{t-1}|\mathbf{x}_0,\mathbf{x}_t,\mathbf{x}_T) = -\frac{1}{2\boldsymbol{\pi}^2}\Bigg[&(\frac{1}{\frac{\Theta_{t-1}^2}{\Theta_t^2}\Sigma_t^2-\Sigma_{t-1}^2}+\frac{1}{\Sigma_{t-1}^2})\mathbf{x}_{t-1}^2 - 2\big[\Sigma_{t-1}^2(\boldsymbol{\mu}+\frac{\Theta_{t-1}}{\Theta_t}(\mathbf{x}_t-\boldsymbol{\mu}))\\ &+(\frac{\Theta_{t-1}^2}{\Theta_t^2}\Sigma_t^2-\Sigma_{t-1}^2)(\boldsymbol{\mu}+(\mathbf{x}_0-\boldsymbol{\mu})\Theta_{t-1})\big]\mathbf{x}_{t-1}+C.\Bigg]\end{aligned} \tag{E.8}$$

We can reformulate the Eq. (E.8) in Gaussian distribution format:

$$Var[\mathbf{x}_{t-1}] = \boldsymbol{\pi}^2(\frac{1}{\frac{\Theta_{t-1}^2}{\Theta_t^2}\Sigma_t^2-\Sigma_{t-1}^2}+\frac{1}{\Sigma_{t-1}^2})^{-1} = \boldsymbol{\pi}^2\frac{\Sigma_{t-1}^2(\Theta_{t-1}^2\Sigma_t^2-\Theta_t^2\Sigma_{t-1}^2)}{\Theta_{t-1}^2\Sigma_t^2} \tag{E.9}$$

$$\begin{aligned}E[\mathbf{x}_{t-1}] &= \big[\Sigma_{t-1}^2(\boldsymbol{\mu}+\frac{\Theta_{t-1}}{\Theta_t}(\mathbf{x}_t-\boldsymbol{\mu}))+(\frac{\Theta_{t-1}^2}{\Theta_t^2}\Sigma_t^2-\Sigma_{t-1}^2)(\boldsymbol{\mu}+(\mathbf{x}_0-\boldsymbol{\mu})\Theta_{t-1})\big]\cdot\frac{\Sigma_{t-1}^2(\Theta_{t-1}^2\Sigma_t^2-\Theta_t^2\Sigma_{t-1}^2)}{\Theta_{t-1}^2\Sigma_t^2}\\ &= \frac{\Sigma_{t-1}^2(\Theta_{t-1}^2\Sigma_t^2-\Theta_t^2\Sigma_{t-1}^2)}{\Theta_t^2}[\boldsymbol{\mu}+(\mathbf{x}_0-\boldsymbol{\mu})\Theta_{t-1}]+\frac{\Sigma_{t-1}^4(\Theta_{t-1}^2\Sigma_t^2-\Theta_t^2\Sigma_{t-1}^2)}{\Theta_{t-1}\Sigma_t\Theta_t}\boldsymbol{\pi}\epsilon_t\end{aligned} \tag{E.10}$$

DDIM Reverse Process. A common forward process in our framework can be determined as follows:

$$\mathbf{x}_{t-1} = \boldsymbol{\mu}+(\mathbf{x}_0-\boldsymbol{\mu})\Theta_{t-1}+\boldsymbol{\pi}\Sigma_{t-1}\epsilon_{t-1}, \tag{E.11}$$

$$\mathbf{x}_t = \boldsymbol{\mu}+(\mathbf{x}_0-\boldsymbol{\mu})\Theta_t+\boldsymbol{\pi}\Sigma_t\epsilon_t. \tag{E.12}$$

We assume the reverse process follows a Gaussian distribution:

$$\begin{aligned}\mathbf{x}_{t-1} &= \kappa_t\mathbf{x}_t+\eta_t\boldsymbol{\mu}+\gamma_t\mathbf{x}_0+\dot{\sigma}_t\boldsymbol{\pi}\epsilon_t\\ &= \kappa_t(\boldsymbol{\mu}+(\mathbf{x}_0-\boldsymbol{\mu})\Theta_t+\boldsymbol{\pi}\Sigma_t\epsilon_t)+\eta_t\boldsymbol{\mu}+\gamma_t\mathbf{x}_0+\dot{\sigma}_t\boldsymbol{\pi}\epsilon_t\\ &= (\kappa_t+\eta_t-\kappa_t\Theta_t)\boldsymbol{\mu}+(\kappa_t\Theta_t+\gamma_t)\mathbf{x}_0+\boldsymbol{\pi}(\kappa_t^2\Sigma_t^2+{\dot{\sigma}_t}^2)^{\frac{1}{2}}\epsilon_t,\end{aligned} \tag{E.13}$$

we have:

$$\kappa_t+\eta_t-\kappa_t\Theta_t = 1-\Theta_{t-1}, \tag{E.14}$$

$$\kappa_t\Theta_t+\gamma_t = \Theta_{t-1}, \tag{E.15}$$

$$\Sigma_{t-1}^2 = \kappa_t^2\Sigma_t^2+{\dot{\sigma}_t}^2. \tag{E.16}$$

By setting $\dot{\sigma}_t = 0$:

$$\begin{aligned}\kappa_t = \frac{\Sigma_{t-1}}{\Sigma_t}, \gamma_t &= \Theta_{t-1}-\Theta_t\frac{\Sigma_{t-1}}{\Sigma_t}\\ \eta_t &= 1-\Theta_{t-1}-(1-\Theta_t)\frac{\Sigma_{t-1}}{\Sigma_t},\end{aligned} \tag{E.17}$$

substituting into Eq. (E.13):

$$\begin{aligned}\mathbf{x}_{t-1} &= \frac{\Sigma_{t-1}}{\Sigma_t}\mathbf{x}_t+(1-\Theta_{t-1}-(1-\Theta_t)\frac{\Sigma_{t-1}}{\Sigma_t})\boldsymbol{\mu}+(\Theta_{t-1}-\Theta_t\frac{\Sigma_{t-1}}{\Sigma_t})\mathbf{x}_0\\ &= \frac{\Sigma_{t-1}}{\Sigma_t}\mathbf{x}_t+(1-\frac{\Sigma_{t-1}}{\Sigma_t}-(\Theta_{t-1}-\Theta_t\frac{\Sigma_{t-1}}{\Sigma_t}))\boldsymbol{\mu}+(\Theta_{t-1}-\Theta_t\frac{\Sigma_{t-1}}{\Sigma_t})\mathbf{x}_0\\ &= \frac{\Sigma_{t-1}}{\Sigma_t}\mathbf{x}_t+(1-\frac{\Sigma_{t-1}}{\Sigma_t})\boldsymbol{\mu}+(\Theta_{t-1}-\Theta_t\frac{\Sigma_{t-1}}{\Sigma_t})(\mathbf{x}_0-\boldsymbol{\mu})\\ &= \frac{\Sigma_{t-1}}{\Sigma_t}\mathbf{x}_t+(1-\frac{\Sigma_{t-1}}{\Sigma_t})\boldsymbol{\mu}+(\Theta_{t-1}-\Theta_t\frac{\Sigma_{t-1}}{\Sigma_t})(\frac{\mathbf{x}_t-\boldsymbol{\mu}-\boldsymbol{\pi}\Sigma_t\epsilon_t}{\Theta_t})\\ &= \boldsymbol{\mu}+\frac{\Theta_{t-1}}{\Theta_t}(\mathbf{x}_t-\boldsymbol{\mu})-\boldsymbol{\pi}(\frac{\Theta_{t-1}}{\Theta_t}\Sigma_t-\Sigma_{t-1})\epsilon_t\end{aligned} \tag{E.18}$$

This concludes the derivations in Sec. 4.2.

F. Connections Among Existing Diffusion Bridge

Suppose the high-quality image $\mathbf{x}$ is sampled from the data distribution $p_{HQ}(\boldsymbol{x})$ and the paired degraded image $\boldsymbol{\mu}$ is sampled from prior distribution $p_{LQ}(\boldsymbol{x})$. We redefine the generalized meaning-reverting process as:

$$d\mathbf{x}_t = \theta_t(\boldsymbol{\mu} - \mathbf{x}_t)dt + \boldsymbol{\pi}\sigma_t d\omega_t, \tag{F.1}$$

where $\boldsymbol{\pi}$ is a predefined value. By applying the Doob's h-transform, we can establish the bridge SDE that connects the paired distribution under a fixed drift-to-diffusion coefficient ratio $\lambda = \sigma_t^2/(2\theta_t)$:

$$d\mathbf{x}_t = \theta_t \coth(\overline{\theta}_{t:T})(\boldsymbol{\mu} - \mathbf{x}_t)dt + \sqrt{2\boldsymbol{\pi}^2\lambda\theta_t}d\omega_t, \tag{F.2}$$

F.1. Connections to Variance-Exploding (VE) and Variance-Preserving (VP) SDEs

SMLD [63] primarily introduces two mainstream diffusion formulations, namely VP and VE. For a given generalized OU process in Eq. (F.1), there exists relationships:

$$\begin{aligned}
\lim_{\theta_t\to 0}^{\boldsymbol{\pi}=1} \text{Eq. (F.1)} &= \lim_{\theta_t\to 0}^{\boldsymbol{\pi}=1}\{d\mathbf{x}_t = \theta_t(\boldsymbol{\mu} - \mathbf{x}_t)dt + \boldsymbol{\pi}\sigma_t d\omega_t\} \\
&= \lim_{\theta_t\to 0}^{\boldsymbol{\pi}=1}\{d\mathbf{x}_t = \sigma_t d\omega_t\} \\
&= \text{VE},
\end{aligned} \tag{F.3}$$

where σ_t can be any noise schedule. Besides, we have:

$$\begin{aligned}
\lim_{\boldsymbol{\mu}\to 0,\theta_t\to\sigma_t^2}^{\boldsymbol{\pi}=1} \text{Eq. (F.1)} &= \lim_{\boldsymbol{\mu}\to 0,\theta_t\to\sigma_t^2}^{\boldsymbol{\pi}=1}\{d\mathbf{x}_t = \theta_t(\boldsymbol{\mu} - \mathbf{x}_t)dt + \boldsymbol{\pi}\sigma_t d\omega_t\} \\
&= \lim_{\boldsymbol{\mu}\to 0,\theta_t\to\sigma_t^2}^{\boldsymbol{\pi}=1}\{d\mathbf{x}_t = \theta_t\boldsymbol{\mu}dt - \theta_t\mathbf{x}_t dt + \sigma_t d\omega_t\} \\
&= \lim_{\boldsymbol{\mu}\to 0,\theta_t\to\sigma_t^2}^{\boldsymbol{\pi}=1}\{d\mathbf{x}_t = -\frac{1}{2}\sigma_t^2\mathbf{x}_t dt + \sigma_t d\omega_t\} \\
&= \text{VP},
\end{aligned} \tag{F.4}$$

where we set the $\theta_t \to \sigma_t^2$, implying $\lambda = \frac{1}{2}$. On this basis, DDBM [85] further extends such diffusion configuration to bridge models, which are also special cases of our formulation under specific configurations:

$$\begin{aligned}
\lim_{\theta_t\to 0,\sigma_t^2\to C}^{\boldsymbol{\pi}=1} \text{Eq. (F.2)} &= \lim_{\theta_t\to 0,\sigma_t^2\to C}^{\boldsymbol{\pi}=1}\{d\mathbf{x}_t = \theta_t \coth(\overline{\theta}_{t:T})(\boldsymbol{\mu} - \mathbf{x}_t)dt + \sqrt{2\pi^2\lambda\theta_t}d\omega_t\} \\
&= \lim_{\theta_t\to 0,\sigma_t^2\to C}^{\boldsymbol{\pi}=1}\{d\mathbf{x}_t = \frac{\sigma_t^2}{\sigma_T^2 - \sigma_t^2}(\boldsymbol{\mu} - \mathbf{x}_t)dt + \sigma_t d\omega_t\} \\
&= \text{VE Bridge},
\end{aligned} \tag{F.5}$$

where C denotes a constant and $\lambda = \frac{\sigma_t^2}{2\theta_t} \to \infty$. We utilize the following approximation:

$$\lim_{\theta_t\to 0} \theta_t \coth(\overline{\theta}_{t:T}) = \theta \coth(\theta(T-t)) = \frac{1}{T-t}. \tag{F.6}$$

VP bridge drives terminate state towards a zero-mean Gaussian distribution, which satisfies:

$$\begin{aligned}
\lim_{\boldsymbol{\mu}\to 0,\theta_t\to\sigma_t^2}^{\boldsymbol{\pi}=1} \text{Eq. (F.2)} &= \lim_{\boldsymbol{\mu}\to 0,\theta_t\to\sigma_t^2}^{\boldsymbol{\pi}=1}\{d\mathbf{x}_t = \theta_t \coth(\overline{\theta}_{t:T})(\boldsymbol{\mu} - \mathbf{x}_t)dt + \sqrt{2\pi^2\lambda\theta_t}d\omega_t\} \\
&= \lim_{\boldsymbol{\mu}\to 0,\theta_t\to\sigma_t^2}^{\boldsymbol{\pi}=1}\{d\mathbf{x}_t = -\sigma_t^2 \coth(\overline{\sigma}_{t:T}^2)\mathbf{x}_t dt + \sigma_t d\omega_t\} \\
&= \text{VP Bridge},
\end{aligned} \tag{F.7}$$

F.2. Connections to Brownian Bridge SDEs

Brownian bridge is a fundamental architecture for diffusion model, which are widely adopted in BBDM [30], I^{2}SB [35]. By setting $\theta_t \to 0$ with condition $2\lambda\theta_t = 1 = \sigma_t^2$, we can derive the Brownian Bridge as formulated below:

$$\begin{aligned}\lim_{\theta_t\to 0,\sigma_t^2\to 1}^{\boldsymbol{\pi}=1} \text{Eq. (F.2)} &= \lim_{\theta_t\to 0,\sigma_t^2\to 1}^{\boldsymbol{\pi}=1} \{d\mathbf{x}_t = \theta_t \coth(\bar{\theta}_{t:T})(\boldsymbol{\mu}-\mathbf{x}_t)dt + \sqrt{2\pi^2\lambda\theta_t}d\omega_t\} \\ &= \lim_{\theta_t\to 0,\sigma_t^2\to 1}^{\boldsymbol{\pi}=1} \{d\mathbf{x}_t = \frac{\boldsymbol{\mu}-\mathbf{x}_t}{T-t}dt + d\omega_t\} \\ &= \text{Brownian Bridge},\end{aligned} \tag{F.8}$$

where the corresponding expectation and variance are:

$$\mathbf{x}_t = \boldsymbol{\mu} + (\mathbf{x}_0-\boldsymbol{\mu})(1-\frac{t}{T}) + \int_0^t \frac{T-t}{T-s}dw_s, \tag{F.9}$$

$$E[\mathbf{x}_t] = \boldsymbol{\mu} + (\mathbf{x}_0-\boldsymbol{\mu})(1-\frac{t}{T}), \tag{F.10}$$

$$Var[\mathbf{x}_t] = t(1-\frac{t}{T}), \tag{F.11}$$

F.3. Connections to Flow Matching

Flow-based generative models [33, 38] design a deterministic probability path that linearly interpolates between a prior and the data distribution, and then directly learn a time-dependent vector field whose integral trajectories realize this path. By discarding the stochastic noise ($\sigma_t = 0$) and adopting the Brownian bridge configuration, Eq. (F.2) can be transformed into:

$$\begin{aligned}\lim_{\theta_t\to 0}^{\boldsymbol{\pi}=0} \text{Eq. (F.2)} &= \lim_{\theta_t\to 0}^{\boldsymbol{\pi}=0} \{d\mathbf{x}_t = \theta_t \coth(\bar{\theta}_{t:T})(\boldsymbol{\mu}-\mathbf{x}_t)dt + \sqrt{2\pi^2\lambda\theta_t}d\omega_t\} \\ &= \lim_{\theta_t\to 0}^{\boldsymbol{\pi}=0} \{d\mathbf{x}_t = \frac{\boldsymbol{\mu}-\mathbf{x}_t}{T-t}dt\} \\ &= \text{Flow Matching},\end{aligned} \tag{F.12}$$

whose trajectories satisfy:

$$\mathbf{x}_t = \boldsymbol{\mu} + (\mathbf{x}_0-\boldsymbol{\mu})(1-\frac{t}{T}). \tag{F.13}$$

F.4. Connections to OU Bridge SDEs

Eq. (F.2) can be transformed into naive OU bridge [75] by setting $\pi = 1$ to recover global noise perturbation:

$$\begin{aligned}\lim_{\theta_t,\lambda}^{\boldsymbol{\pi}=1} \text{Eq. (F.2)} &= \lim_{\theta_t,\lambda}^{\boldsymbol{\pi}=1} \{d\mathbf{x}_t = \theta_t \coth(\bar{\theta}_{t:T})(\boldsymbol{\mu}-\mathbf{x}_t)dt + \sqrt{2\pi^2\lambda\theta_t}d\omega_t\} \\ &= \lim_{\theta_t,\lambda}^{\boldsymbol{\pi}=1} \{d\mathbf{x}_t = \frac{\boldsymbol{\mu}-\mathbf{x}_t}{T-t}dt\} \\ &= \text{OU Bridge},\end{aligned} \tag{F.14}$$

F.5. Connections to Stochastic Interpolants

Stochastic interpolants [1] define a unified framework for flows and diffusions, which can be expressed as:

$$\mathbf{x}_t = I(t,\mathbf{x}_0,\mathbf{x}_T) + \gamma(t)z, t\in[0,T], \tag{F.15}$$

whose boundary conditions are $I(0,\mathbf{x}_0,\mathbf{x}_T)=\mathbf{x}_0$ and $I(T,\mathbf{x}_0,\mathbf{x}_T)=\mathbf{x}_T$. Eq. (F.2) describes our probability path as:

$$E[\mathbf{x}_t] = \boldsymbol{\mu} + (\mathbf{x}_0-\boldsymbol{\mu})\frac{\sinh(\bar{\theta}_{t:T})}{\sinh(\bar{\theta}_{0:T})}, \tag{F.16}$$

$$Var[\mathbf{x}_t] = 2\boldsymbol{\pi}^2\lambda\frac{\sinh(\bar{\theta}_{0:t})\sinh(\bar{\theta}_{t:T})}{\sinh(\bar{\theta}_{0:T})}. \tag{F.17}$$

Hence, the above process can be regarded as stochastic interpolants. The derivative of $I(t, \mathbf{x}_0, \mathbf{x}_T)$ to time t is fixed as:

$$\partial_t I(t, \mathbf{x}_0, \mathbf{x}_T) = \partial_t \frac{\sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{0:T})}(\mathbf{x}_0 - \boldsymbol{\mu}), \tag{F.18}$$

and $\gamma(t)$ with boundary conditions $\gamma(0) = \gamma(T) = 0$ is:

$$\gamma(t)^2 = 2\boldsymbol{\pi}^2 \lambda \frac{\sinh(\overline{\theta}_{0:t}) \sinh(\overline{\theta}_{t:T})}{\sinh(\overline{\theta}_{0:T})}. \tag{F.19}$$

These relationships are summarized in Tab. 1 in Sec. 4.3.

G. Training Objective

Proposition 3 *Let $\mathbf{x}_t$ be a finite random variable described by the given residual diffusion bridge in Eq. (F.2). For a fixed final state $\mathbf{x}_T = \boldsymbol{\mu}$. the expectation of log-likelihood $\mathbb{E}_{p(\mathbf{x}_0)}[\log p_\theta(\mathbf{x}_0|\boldsymbol{\mu})]$ possesses an Evidence Lower Bound (ELBO):*

$$ELBO = \mathbb{E}_{p(\mathbf{x}_0)}\left[\mathbb{E}_{p(\mathbf{x}_1|\mathbf{x}_0,\boldsymbol{\mu})}[\log p_\theta(\mathbf{x}_0|\mathbf{x}_1, \mathbf{x}_T)] - \sum_{t>1} \mathbb{E}_{p(\mathbf{x}_t|\mathbf{x}_0,\boldsymbol{\mu})}[D_{KL}(p(\mathbf{x}_{t-1}|\mathbf{x}_0, \mathbf{x}_t, \mathbf{x}_T))\|p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_T)]\right] \tag{G.1}$$

Assuming $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_T)$ follows a Gaussian distribution with a constant variance $\mathcal{N}(\boldsymbol{\mu}_{\theta,t-1}, \sigma^2_{\theta,t-1} I)$, maximizing the ELBO is equivalent to minimizing:

$$\mathcal{L} = \mathbb{E}_{t,\mathbf{x}_0,\mathbf{x}_t,\mathbf{x}_T}\left[\frac{1}{2\sigma^2_{\theta,t-1}}\|\boldsymbol{\mu}_{t-1} - \boldsymbol{\mu}_{\theta,t-1}\|^2\right] \tag{G.2}$$

where μ_{t-1} is the expectation at time $t-1$ and $\mu_{\theta,t-1}$ is predicted by a neural network parameterized by θ.

Proof. For the conditional marginal likelihood of the data $\mathbf{x}_0$, we have

$$p_\theta(\mathbf{x}_0|\boldsymbol{\mu}) = \int p_\theta(\mathbf{x}_{0:T}|\boldsymbol{\mu}) d\mathbf{x}_{1:T} = \int \frac{p_\theta(\mathbf{x}_{0:T}|\boldsymbol{\mu})}{p(\mathbf{x}_{1:T}|\mathbf{x}_0, \boldsymbol{\mu})} p(\mathbf{x}_{1:T}|\mathbf{x}_0, \boldsymbol{\mu}) d\mathbf{x}_{1:T} \tag{G.3}$$

To maximize Eq. (G.3), we leverage the property of Jensen's inequality:

$$\log p_\theta(\mathbf{x}_0|\boldsymbol{\mu}) \geq \mathbb{E}_{p(\mathbf{x}_{1:T}|\mathbf{x}_0,\boldsymbol{\mu})}\left[\log \frac{p_\theta(\mathbf{x}_{0:T}|\boldsymbol{\mu})}{p(\mathbf{x}_{1:T}|\mathbf{x}_0, \boldsymbol{\mu})}\right] = \mathbb{E}\left[\log p_\theta(\mathbf{x}_T|\boldsymbol{\mu}) + \log \frac{p_\theta(\mathbf{x}_{0:T-1}|\boldsymbol{\mu})}{p(\mathbf{x}_{1:T}|\mathbf{x}_0, \boldsymbol{\mu})}\right] \tag{G.4}$$

$$= \mathbb{E}\left[\log p_\theta(\mathbf{x}_T|\boldsymbol{\mu}) + \sum_{t\geq 1} \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \boldsymbol{\mu})}{p(\mathbf{x}_t|\mathbf{x}_{t-1}, \boldsymbol{\mu})}\right] \tag{G.5}$$

$$= \mathbb{E}\left[\log p_\theta(\mathbf{x}_T|\boldsymbol{\mu}) + \sum_{t>1} \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \boldsymbol{\mu})}{p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0, \boldsymbol{\mu})} + \log \frac{p_\theta(\mathbf{x}_0|\mathbf{x}_1, \boldsymbol{\mu})}{p(\mathbf{x}_1|\mathbf{x}_0, \boldsymbol{\mu})}\right] \tag{G.6}$$

$$= \mathbb{E}\left[\log p_\theta(\mathbf{x}_T|\boldsymbol{\mu}) + \sum_{t>1} \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \boldsymbol{\mu})}{p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0, \boldsymbol{\mu})} \cdot \frac{p(\mathbf{x}_{t-1}|\mathbf{x}_0, \boldsymbol{\mu})}{p(\mathbf{x}_t|\mathbf{x}_0, \boldsymbol{\mu})} + \log \frac{p_\theta(\mathbf{x}_0|\mathbf{x}_1, \boldsymbol{\mu})}{p(\mathbf{x}_1|\mathbf{x}_0, \boldsymbol{\mu})}\right] \tag{G.7}$$

$$= \mathbb{E}\left[\log p_\theta(\mathbf{x}_T|\boldsymbol{\mu}) + \sum_{t>1} \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \boldsymbol{\mu})}{p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0, \boldsymbol{\mu})} + \sum_{t>1} \log \frac{p(\mathbf{x}_{t-1}|\mathbf{x}_0, \boldsymbol{\mu})}{p(\mathbf{x}_t|\mathbf{x}_0, \boldsymbol{\mu})} + \log \frac{p_\theta(\mathbf{x}_0|\mathbf{x}_1, \boldsymbol{\mu})}{p(\mathbf{x}_1|\mathbf{x}_0, \boldsymbol{\mu})}\right] \tag{G.8}$$

$$= \mathbb{E}\left[\log p_\theta(\mathbf{x}_T|\boldsymbol{\mu}) + \sum_{t>1} \log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \boldsymbol{\mu})}{p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0, \boldsymbol{\mu})} + \log \frac{p(\mathbf{x}_1|\mathbf{x}_0, \boldsymbol{\mu})}{p(\mathbf{x}_T|\mathbf{x}_0, \boldsymbol{\mu})} + \log \frac{p_\theta(\mathbf{x}_0|\mathbf{x}_1, \boldsymbol{\mu})}{p(\mathbf{x}_1|\mathbf{x}_0, \boldsymbol{\mu})}\right] \tag{G.9}$$

$$= \mathbb{E}\left[\log \frac{p_\theta(\mathbf{x}_T|\boldsymbol{\mu})}{p(\mathbf{x}_T|\mathbf{x}_0, \boldsymbol{\mu})}\right] + \sum_{t>1} \mathbb{E}\left[\log \frac{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \boldsymbol{\mu})}{p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0, \boldsymbol{\mu})}\right] + \mathbb{E}\left[\log p_\theta(\mathbf{x}_0|\mathbf{x}_1, \boldsymbol{\mu})\right] \tag{G.10}$$

$$= \mathbb{E}_{p(\mathbf{x}_1|\mathbf{x}_0,\boldsymbol{\mu})}\left[\log p_\theta(\mathbf{x}_0|\mathbf{x}_1, \boldsymbol{\mu})\right] - \sum_{t>1} \mathbb{E}_{p(\mathbf{x}_t|\mathbf{x}_0,\boldsymbol{\mu})}\left[D_{KL}(p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0, \boldsymbol{\mu})\|p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_T, \boldsymbol{\mu}))\right]. \tag{G.11}$$

Accordingly,

$$
\begin{aligned}
& D_{KL}\left(p\left(\mathbf{x}_{t-1} \mid \mathbf{x}_0, \mathbf{x}_t, \mathbf{x}_T\right) || p_\theta\left(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_T\right)\right) \\
= & \mathbb{E}_{p(\mathbf{x}_{t-1}|\mathbf{x}_0,\mathbf{x}_t,\mathbf{x}_T)}\left[\log \frac{\frac{1}{\sqrt{2\pi}\sigma_{t-1}} e^{-(x_{t-1}-\boldsymbol{\mu}_{t-1})^2/2\sigma_{t-1}^2}}{\frac{1}{\sqrt{2\pi}\sigma_{\theta,t-1}} e^{-(x_{t-1}-\boldsymbol{\mu}_{\theta,t-1})^2/2\sigma_{\theta,t-1}^2}}\right] \\
= & \mathbb{E}_{p(\mathbf{x}_{t-1}|\mathbf{x}_0,\mathbf{x}_t,\mathbf{x}_T)}\left[\log \sigma_{\theta,t-1} - \log \sigma_{t-1} - (x_{t-1}-\boldsymbol{\mu}_{t-1})^2/2\sigma_{t-1}^2 + (x_{t-1}-\boldsymbol{\mu}_{\theta,t-1})^2/2\sigma_{\theta,t-1}^2\right] \\
= & \log \sigma_{\theta,t-1} - \log \sigma_{t-1} - \frac{1}{2} + \frac{\sigma_{t-1}^2}{2\sigma_{\theta,t-1}^2} + \frac{(\boldsymbol{\mu}_{t-1}-\boldsymbol{\mu}_{\theta,t-1})^2}{2\sigma_{\theta,t-1}^2}
\end{aligned}
\tag{G.12}
$$

Ignoring unlearnable constant, the training objective that involves minimizing the negative ELBO is :

$$
\mathcal{L} = \mathbb{E}_{t,\mathbf{x}_0,\mathbf{x}_t,\mathbf{x}_T}\left[\frac{1}{2\sigma_{\theta,t-1}^2}\|\boldsymbol{\mu}_{t-1} - \boldsymbol{\mu}_{\theta,t-1}\|^2\right]. \tag{G.13}
$$

By substituting Eq. (E.10) into Eq. (G.13) yields the equivalent loss:

$$
\mathcal{L} = \mathbb{E}_{t,\mathbf{x}_0,\mathbf{x}_t,\mathbf{x}_T}\left[C_\theta\|\boldsymbol{\pi}\epsilon_{t-1} - \boldsymbol{\pi}\epsilon_{\theta,t-1}\|^2\right]. \tag{G.14}
$$

Where C_θ are corresponding weights. This concludes the proof of the **Proposition 3** in Sec. 4.2.

H. More Experiments

H.1. Summary about the Datasets

We evaluate the proposed method on five natural image restoration tasks, including deraining, low-light enhancement, desnowing, dehazing, and deblurring. We select the most widely used datasets for each task, as summarized in Tab. S1.

Table S1. Summary of the image restoration datasets utilized in this paper.

Task	Dataset	Synthetic/Real	Train samples	Test samples
Deraining	DID [80]	Synthetic	-	1,200
	Rain13K [20]	Synthetic	13,711	-
	Rain_100 [74]	Synthetic	-	200
	DeRaindrop [52]	Real	861	307
	GT-Rain [4]	Real	26,125	2,100
	RealRain-1k [32]	Real	1792	448
Low-light Enhancement	LOL [70]	Real	485	15
	MEF [45]	Real	-	17
	VE-LOL-L [36]	Synthetic/Real	900/400	100/100
	NPE [68]	Real	-	8
	DICM [27]	Real	-	64
Desnowing	CSD [6]	Synthetic	8,000	2,000
	Snow100K-Real [39]	Real	-	1,329
Dehazing	SOTS [28]	Synthetic	-	500
	ITS_v2 [28]	Synthetic	13,990	-
	D-HAZY [10]	Synthetic	1,178	294
	NH-HAZE [3]	Real	-	55
	Dense-Haze [2]	Real	-	55
	NHRW [81]	Real	-	150
Deblur	GoPro [49]	Synthetic	2,103	1,111
	RealBlur [54]	Real	3,758	980

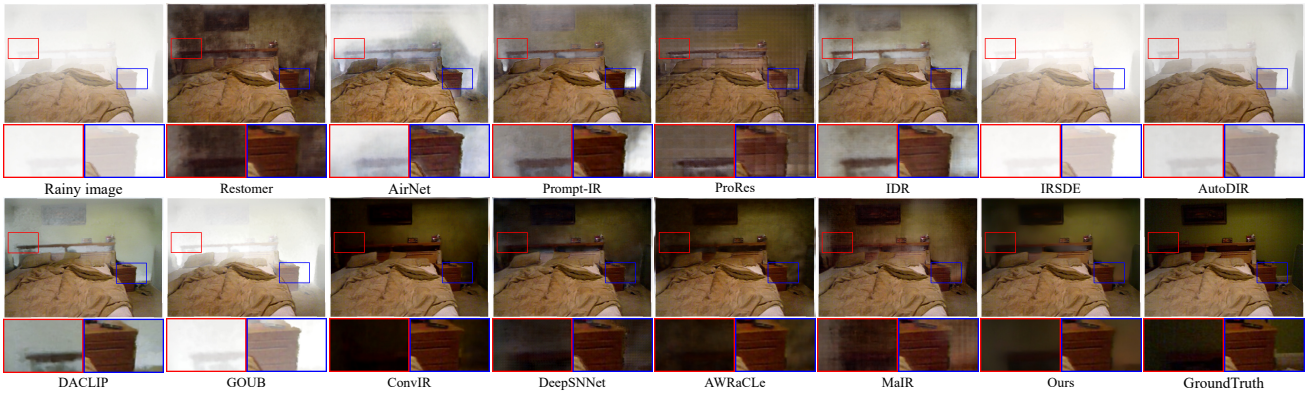

Figure S1. Visualization comparison with state-of-the-art methods on dehazing. Zoom in for best view.

Figure S2. Visualization comparison with state-of-the-art methods on deblurring. Zoom in for best view.

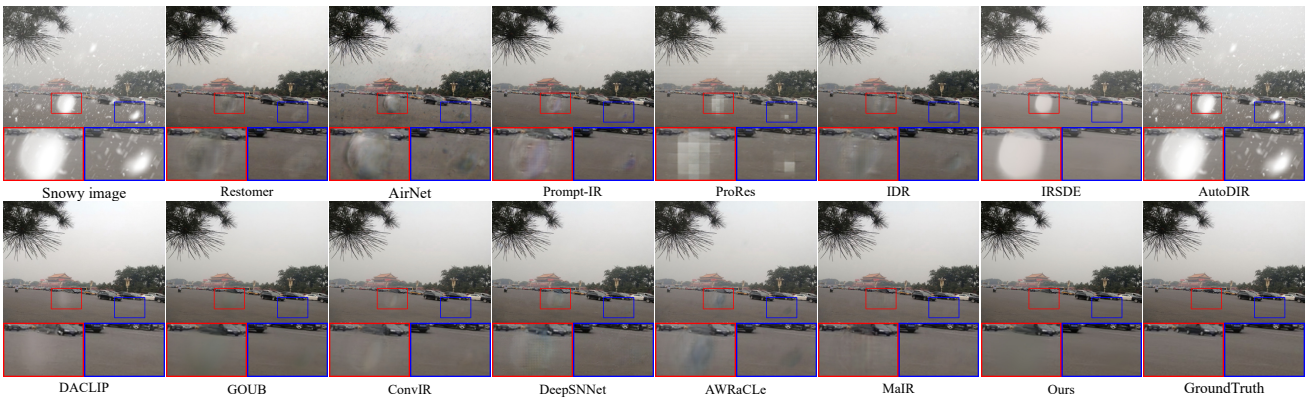

Figure S3. Visualization comparison with state-of-the-art methods on desnowing. Zoom in for best view.

H.2. More Visual Comparisons on Image Restoration

We show the visualization results of other degradation categories in Fig. S1, Fig. S2, Fig. S3, and Fig. S4, to further demonstrate our superiority. Evidently, our method generates more stable image samples with high fidelity than other universal image restoration methods. Benefiting from the adaptivity of residual bridge score matching, we achieve the outstanding reconstruction of the missing details and preserve undegraded regions well.

H.3. More Visual Comparisons on Real-world Scene Generalization

Known task generalization. We randomly select 20 samples for each task to conduct the non-reference assessment, as presented in Tab. 8. Furthermore, to fully demonstrate that our method can handle the real-world restoration tasks, we have generalized all well-optimized models to five known tasks within real-world scenarios. Visual comparisons are displayed in Fig. S5, Fig. S6, Fig. S7, Fig. S8, and Fig. S9, respectively. Clearly, our method produces the highest-quality restored images.
Unknown task generalization. Unknown task image restoration is performed on both POLED and TOLED [86]. Visual

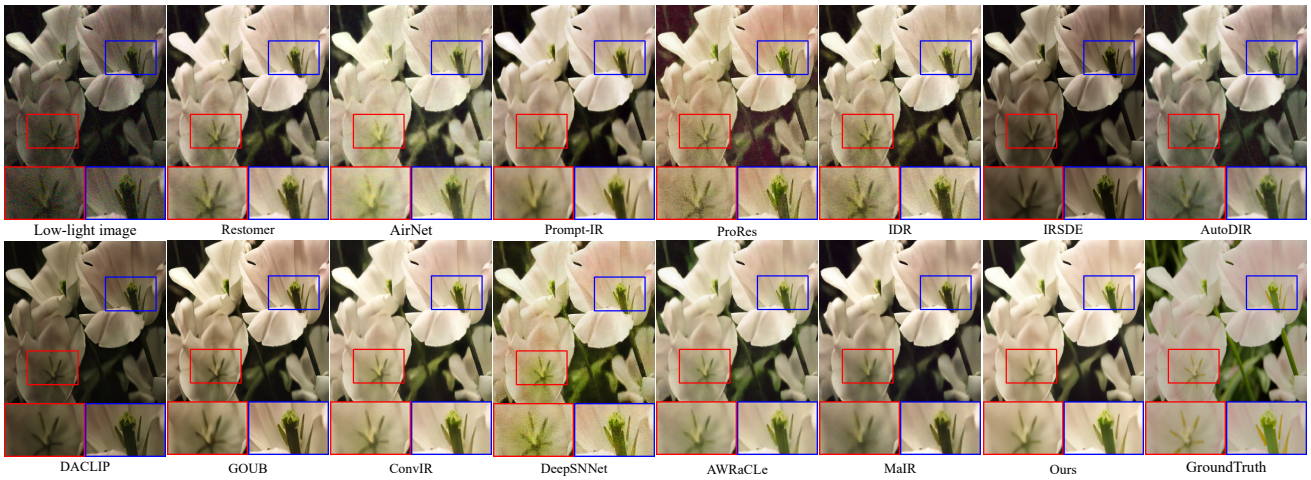

Figure S4. Visualization comparison with state-of-the-art methods on low-light enhancement. Zoom in for best view.

Figure S5. Visualization comparison of deblurring task in real-world scenarios. Zoom in for best view.

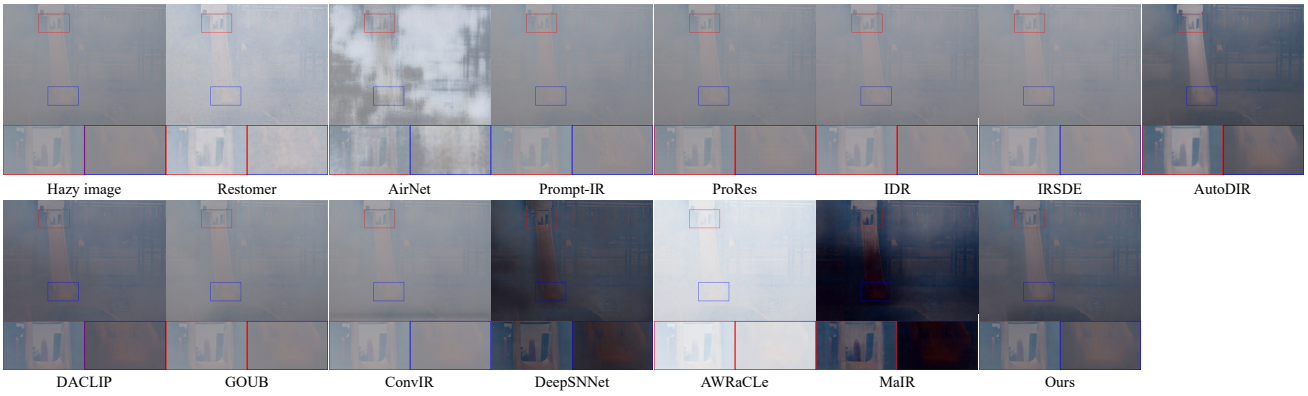

Figure S6. Visualization comparison of dehazing task in real-world scenarios. Zoom in for best view.

comparisons on the POLED dataset are provided in Fig. S10. The results show that our method can generalize to real-world scenes and achieve competitive visual results.

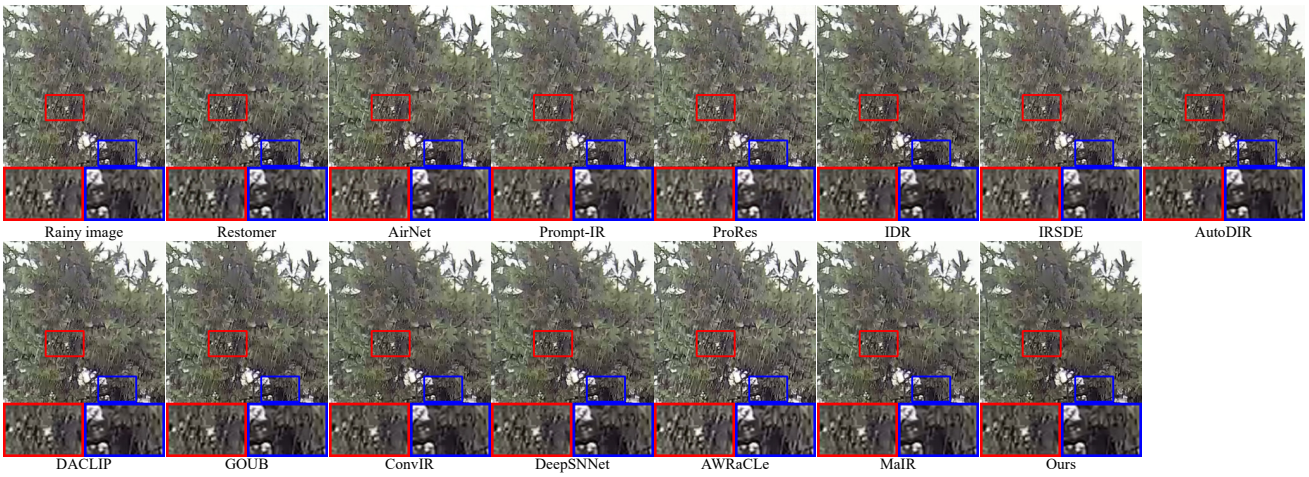

Figure S7. Visualization comparison of deraining task in real-world scenarios. Zoom in for best view.

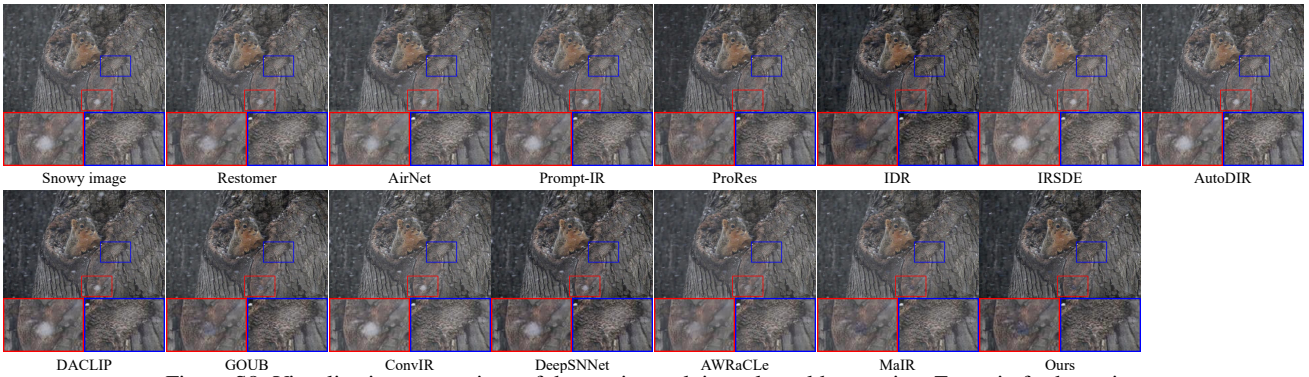

Figure S8. Visualization comparison of desnowing task in real-world scenarios. Zoom in for best view.

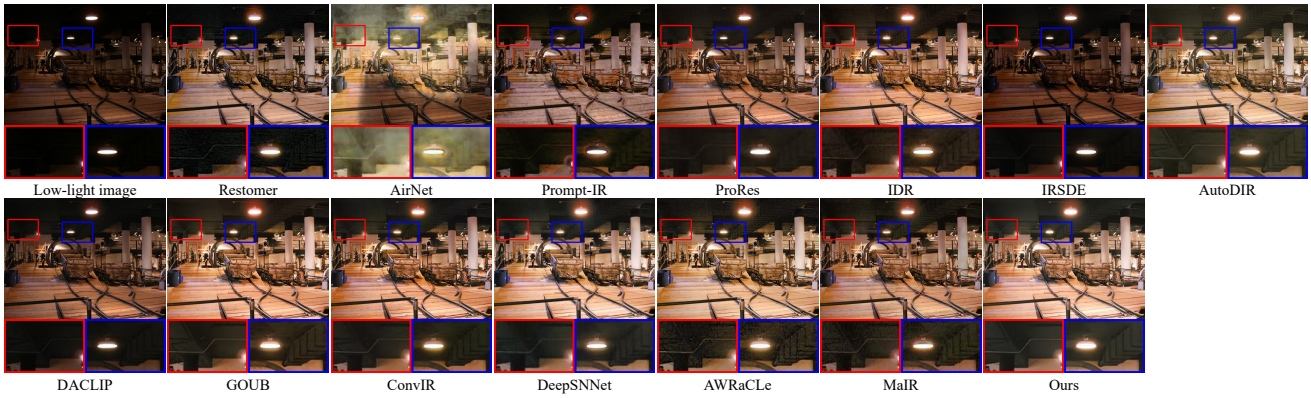

Figure S9. Visualization comparison of low-light enhancement task in real-world scenarios. Zoom in for best view.

H.4. More Visual Comparisons on Image Translation and Inpainting

To further show the visual advantages of our approach across tasks, we present additional comparisons for image translation (Fig. S11) and image inpainting (Fig. S12). In image translation, our method better preserves semantic and structural consistency, produces more faithful colors, sharper edges, and richer details. In image inpainting, it synthesizes textures and boundaries highly consistent with the surrounding context while avoiding oversmoothing and texture drift. Overall, our qualitative results show clearer details, stronger global consistency, and fewer visual artifacts than competing methods.

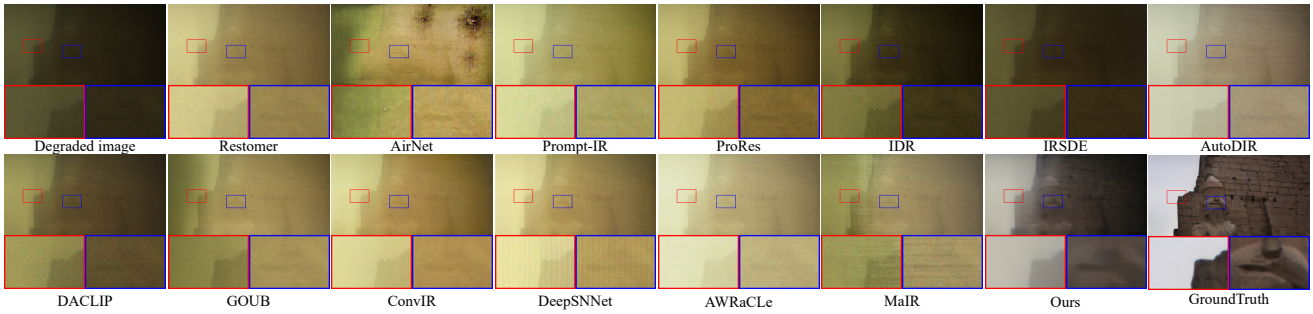

Figure S10. Visualization results of zero-shot generalization in real-world POLED dataset. Zoom in for best view.

Figure S11. Visualization results of image translation. Zoom in for best view.

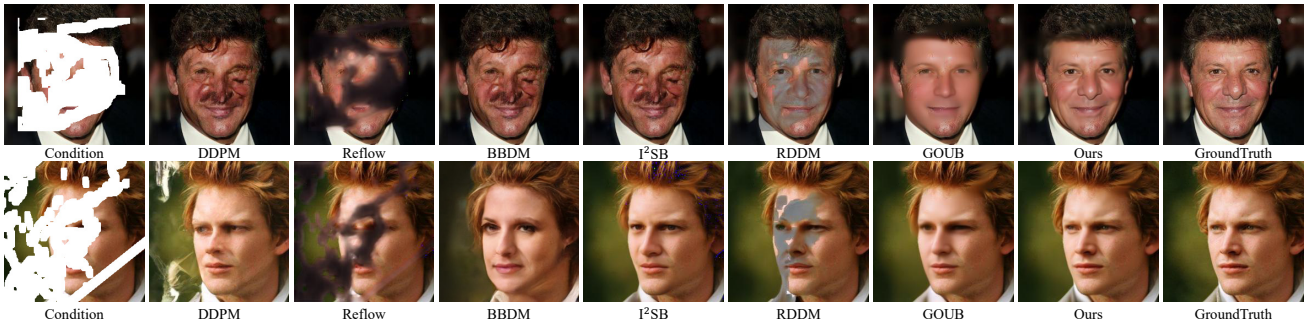

Figure S12. Visualization results of image inpainting. Zoom in for best view.

H.5. Efficiency Comparison

Our mixed dataset consists of images with resolutions ranging from 256 to 1024 pixels. Accordingly, we evaluate model efficiency under three representative resolution settings, as summarized in Tab. S2. Evidently, our methods are moderately efficient with reasonable resource consumption. Overall, RDBM strikes a balance between efficiency and performance.

I. Discussions, Limitations, and Future Work

Limitations and broader impact. The main challenge lies in fully exploring the connections between the data and prior distribution to modify the diffusion process. Although we have theoretically proposed a general and analytical formulation for diffusion bridge models, our core analysis assumes a fixed drift-to-diffusion coefficient ratio $\lambda = \sigma_t^2/(2\theta_t)$ to admit closed-form solutions of SDEs. In the fields of image restoration, translation and inpainting where the data and prior distributions share semantic or structural affinity, our method is highly flexible and robust with competitive performance. However, it may be sub-optimal when applied to the generative tasks, where the distributions lack direct correspondence. Despite current limitations, we believe our unified model offers a strong foundation for diffusion bridge models.

Future Work. Future work could be explored in several promising directions. (1) With the rise of high-resolution imagery (e.g., 4K, 8K), developing multi-dimensional latent diffusion bridge models is crucial to address the computational demands. (2) Exploring more efficient network architectures to reduce memory usage and enhance efficiency. (3) Expanding the model

Table S2. Efficiency comparisons among universal methods. '—' means out of memmory.

Resolution	256×256			512×512			1024×1024		
Method	Mem.(G)	Time(s)	FPS	Mem.(G)	Time(s)	FPS	Mem.(G)	Time(s)	FPS
Restomer [78]	1.959	0.105	9.563	6.670	0.381	2.622	25.419	1.773	0.564
AirNet [29]	1.039	0.194	5.159	3.480	0.738	1.355	11.244	20.499	0.049
Prompt-IR [51]	2.544	0.111	8.981	7.255	0.399	2.508	26.005	1.845	0.542
ProRes [44]	2.027	0.318	3.149	2.514	0.766	1.305	6.025	1.715	0.583
IDR [82]	1.340	0.052	19.253	4.313	0.136	7.373	16.110	0.615	1.626
IRSDE [41]	1.554	5.017	0.199	2.743	18.493	0.054	9.997	72.289	0.014
AutoDIR [21]	7.023	6.266	0.160	11.021	11.986	0.083	—	—	—
DA-CLIP [42]	2.119	2.585	0.387	6.775	7.937	0.126	58.548	60.893	0.016
GOUB [75]	1.554	4.996	0.200	2.868	18.442	0.054	10.122	72.239	0.014
ConvIR [11]	0.708	0.035	28.570	1.184	0.055	18.020	2.809	0.192	5.202
DeepSNNet [12]	0.862	0.100	9.974	0.989	0.102	9.801	1.364	0.267	3.749
AWRaCLe [53]	1.929	0.101	9.922	4.264	0.354	2.822	13.608	1.374	0.728
MaIR [31]	2.091	1.297	0.771	6.593	4.744	0.211	24.593	18.029	0.055
RDBM-T	0.813	0.418	2.392	1.907	1.621	0.617	13.407	6.648	0.150
RDBM-S	0.825	0.431	2.322	1.921	1.648	0.607	13.421	6.775	0.148
RDBM-B	1.124	0.480	2.081	2.186	1.926	0.519	14.938	7.872	0.127
RDBM-L	1.150	0.504	1.986	2.307	1.982	0.505	15.059	8.135	0.123

capacity and datasets to strengthen restoration performance and generalization. (4) Designing adaptive learning rate schedules or applying model distillation to reduce sampling steps and improve restoration quality.

Reproducibility. Source code is provided in supplementary materials and will be released upon camera-ready submission.