Resolving Endpoint Underfitting in Diffusion Bridges via Noise Alignment

Supplementary Material

7. Sampling Procedure

Since the training objective enables the prediction of the initial image X_0 from any noisy state X_t, we design the corresponding sampling distribution to ensure that samples progressively denoise along the interpolation path. Let $\{t_i\}_{i=0}^N$ represent a partition of the unit interval $[0,1]$, where $t_0 = 0$, $t_N = 1$, and $t_i < t_{i+1}$ for all $i = 0, \ldots, N$. Our sampling framework operates by coordinating the use of two distinct conditional distributions: the standard reverse transition $p(X_{t-1} \mid X_0^\epsilon, X_t)$ and the endpoint-conditioned distribution $p(X_{t-1} \mid X_0^\epsilon, \hat{X}_0, X_t)$. Specifically, we adopt a partitioned approach that applies different conditional distributions over distinct temporal regions. The standard reverse transition is used over the interval over the interval $[t_i, t_N]$, while the endpoint-conditioned distribution is employed over the complementary interval $[t_0, t_i]$. This formulation eliminates the intermediate step of solving for x_0^ϵ from the network output ϵ_θ. We refer to this sampling method as a two-stage sampling strategy, implemented as follows:

7.1. The First Stage

We use $X_{t_{i-1}} \sim p(X_{t_{i-1}} \mid X_0^\epsilon, X_{t_i})$ from $t = 1$ to $t = t_i$. The conditional probability distribution can be expressed in its reparameterized form as:

$$X_{t_{i-1}} = X_{t_i} - \epsilon_\theta \Delta_t + \sigma_t Z, \tag{15}$$

where $\sigma_t = \sqrt{\left[kt_{i-1}(1-t_{i-1})\right]^2 - \left[kt_{i-1}^\alpha t_i^{1-\alpha}(1-t_i)\right]^2}$, $\Delta_t = t_i^\alpha - t_{i-1}^\alpha$, $t_{i-1} \geq \frac{1-\alpha}{2-\alpha}$. It is important to note that this formulation is only valid within the interval $[\frac{1-\alpha}{2-\alpha}, 1]$, which constitutes the fundamental rationale for our two-stage sampling approach. Derivation of Eq. (15) can be found in Sec. 7.3.

7.2. The Second Stage

The direct application of Eq. (15) over the interval $[0, \frac{1-\alpha}{2-\alpha}]$ is prohibited due to the non-negativity requirement of its standard deviation term σ_t. To overcome this limitation, we introduce a hyperparameter w into the noise term of $X_{t_{i-1}} \sim p(X_{t_{i-1}} \mid X_0^\epsilon, \hat{X}_0, X_{t_i})$, yielding the modified expression:

$$X_{t_{i-1}} = a\epsilon_\theta - bX_{t_i} + c\hat{X}_0 + \sigma_{wt} Z, \tag{16}$$

where

$$\begin{aligned} a &= t_i^\alpha \left(w + t_{i-1}^\alpha - 1 - wt_i^\alpha\right), \\ b &= t_{i-1}^\alpha - 1 - wt_i^\alpha, \\ c &= t_{i-1}^\alpha - wt_i^\alpha, \\ \sigma_{wt} &= \sqrt{\left[kt_{i-1}(1-t_{i-1})\right]^2 - \left[kwt_i(1-t_i)\right]^2}. \end{aligned} \tag{17}$$

It should be noted that Eq. (16) remains valid over the entire interval $[0,1]$. This allows the value of t_i to be flexibly chosen (as long as $t_i \geq \frac{1-\alpha}{2-\alpha}$) in order to determine the specific subinterval over which Eq. (16) is applied. We set the partition point d between the two sampling stages to $\frac{1-\alpha}{2-\alpha}$ by default. The hyperparameter w controls the magnitude of σ_t, thereby determining the amount of stochastic noise introduced over the interval $[0, t_i]$. When $\sigma_t = 0$, the sampling process within this interval reduces to a deterministic trajectory. In our experiments, we set $w = \frac{t_{i-1}}{t_i}$ to evaluate the impact of varying noise levels on the final output. Derivation of Eq. (16) can be found in Sec. 7.3.

7.3. Derivation of Sampling Procedure

Proof. Our derivation leverages the reparameterization technique. For $s < t$, we define the process as:

$$X_s = X_t - \frac{X_t - X_0^\epsilon}{t^\alpha}\Delta_t + \sigma_t Z \tag{18}$$

$$= \frac{\Delta_t}{t^\alpha} X_0^\epsilon + \frac{t^\alpha - \Delta_t}{t^\alpha} X_t + \sigma_t Z. \tag{19}$$

According to the calculation rules of reparameterization, the coefficients of Gaussian noise in the left and right equations are equal, and we obtain the following equations:

$$\frac{\Delta_t}{t^\alpha} = 1 - t^\alpha, \tag{20}$$

$$\left[ks(1-s)\right]^2 = \left[\frac{t^\alpha - \Delta_t}{t^\alpha} \cdot kt(1-t)\right]^2 + \sigma_t^2. \tag{21}$$

Solving this system yields $\Delta_t = t^\alpha - s^\alpha$ and $\sigma_t = \sqrt{\left[ks(1-s)\right]^2 - \left[ks^\alpha t^{1-\alpha}(1-t)\right]^2}$.

The non-negativity constraint $\sigma_t \geq 0$ implies $s \geq \frac{1-\alpha}{2-\alpha}$, and thus $d \geq \frac{1-\alpha}{2-\alpha}$.

For the second stage from $t = d$ to $t = 0$, we introduce a hyperparameter w and define $\sigma_{wt} = \sqrt{\left[ks(1-s)\right]^2 - \left[wkt(1-t)\right]^2}$. Using reparameterization:

$$X_s = (1-s^\alpha)X_0^\epsilon + s^\alpha \hat{X}_0 + ks(1-s)Z \tag{22}$$

$$= (1-s^\alpha)X_0^\epsilon + s^\alpha \hat{X}_0 + \sigma_{wt} Z + wkt(1-t)Z, \tag{23}$$

Table 5. Comparison with PMRF in numerous image restoration tasks.

Method	JPEG Restoration						4× Super-Resolution						Deblurring					
	QF	NFE	FID↓	PSNR↑	SSIM↑	LPIPS↓	Filter	NFE	FID↓	PSNR↑	SSIM↑	LPIPS↓	Kernel	NFE	FID↓	PSNR↑	SSIM↑	LPIPS↓
PMRF	5	10	7.6	24.62	0.69	0.31	Pool	10	5.1	25.24	0.73	0.21	Uniform	10	6.6	27.84	0.82	0.20
NADB			6.9	24.45	0.69	0.30			5.3	24.75	0.71	0.23			4.8	27.70	0.81	0.18
PMRF		100	7.8	23.69	0.64	0.32		100	4.5	24.03	0.69	0.21		100	7.2	27.53	0.81	0.21
NADB			4.3	23.63	0.65	0.30			3.7	23.63	0.67	0.22			3.4	27.39	0.80	0.17

and substituting the identity $wkt(1-t)Z = w[X_t - (1-t^\alpha)X_0^\epsilon - t^\alpha \hat{X}_0]$, we rearrange to express X_s in terms of X_t, X_1 and Z:

$$X_s = [wt^\alpha - s^\alpha + 1]\, X_t + (s^\alpha - wt^\alpha)\, \hat{X}_0 + \sigma_{wt} Z \quad (24)$$

$$+ t^\alpha \left[s^\alpha - 1 + w(1-t^\alpha)\right] \cdot \frac{X_t - X_0^\epsilon}{t^\alpha}. \quad (25)$$

This completes the derivation. □

8. Proof of Theorem 2

Theorem 3. *Let ρ_0 be the target distribution, ρ_1 the degraded distribution, and $\hat{\rho}_0$ the output distribution of the mean network M_ϕ optimized via Eq. (11). The mean network reduces the distributional gap in the Wasserstein-2 sense:*

$$W_2(\rho_0, \hat{\rho}_0) \le W_2(\rho_0, \rho_1). \quad (26)$$

Proof. Given $(X_0, X_1) \sim (\rho_0, \rho_1)$ and $\hat{X}_0 = M_\phi(X_1) = \mathbb{E}(X_0 \mid X_1)$, there exits:

$$\mathbb{E}\left[\|X_0 - \hat{X}_0\|^2\right] \le \mathbb{E}\left[\|X_0 - X_1\|^2\right]. \quad (27)$$

Eq. (27) can be verified by expanding the square on the left-hand side. Using the definition of Wasserstein-2 distance and Eq. (27), we obtain:

$$W_2^2(\rho_0, \rho_1) = \inf_{\gamma \in \Pi(\rho_0, \rho_1)} \mathbb{E}_{(X_0, X_1)\sim\gamma}\left[\|X_0 - X_1\|^2\right] \quad (28)$$

$$= \mathbb{E}_{(X_0, X_1)\sim\gamma^*}\left[\|X_0 - X_1\|^2\right] \quad (29)$$

$$\ge \mathbb{E}_{(X_0, X_1)\sim\gamma^*}\left[\|X_0 - \mathbb{E}(X_0 \mid X_1)\|^2\right] \quad (30)$$

$$= \mathbb{E}_{(X_0, \hat{X}_0)\sim\pi^*}\left[\|X_0 - \hat{X}_0\|^2\right] \quad (31)$$

$$\ge \inf_{\pi \in \Pi(\rho_0, \hat{\rho}_0)} \mathbb{E}_{(X_0, \hat{X}_0)\sim\pi}\left[\|X_0 - \hat{X}_0\|^2\right] \quad (32)$$

$$= W_2^2(\rho_0, \hat{\rho}_0) \quad (33)$$

□

9. Hyperparameters in Eq. 5

This section outlines the selection of hyperparameters α and k in Eq. (5). The parameter k controls the overall scale of the stochastic term. We calibrate k so that the maximum amplitude of the stochastic term in Eq. (5) closely matches that of I2SB [21] in Eq. (2). This alignment follows the empirical setup of prior work, as neither our method nor I2SB [21] systematically investigates the effect of the stochastic term's magnitude on generation quality. A comprehensive analysis of this effect will be explored in future work.

The value of α in the mean term is theoretically constrained to the range $(0, 1)$, based on noise alignment considerations. We empirically determine its optimal value through systematic experiments, with quantitative results presented in Tab. 6. Considering the performance across all metrics, we select $\alpha = 0.4$.

10. Supplement the Experimental and Qualitative Results

10.1. Compare with PMRF

We conducted a systematic comparison between the NADB and the Flow Matching-based PMRF [26] method under three different types of noise, using identical network architectures and strictly adhering to the original hyperparameters. As shown in Tab. 5, the PMRF [26] method exhibits instability at higher NFE values. This result underscores the advantage of the diffusion bridge approach and highlights the importance of our research into the endpoint fitting problem.

Table 6. Hyperparameter tuning using JPEG restoration tasks, QF=5.

JPEG Restoration					
NFE	α	FID↓	PSNR↑	SSIM↑	LPIPS↓
10	0.3	8.2	24.60	0.70,	0.30
	0.4	6.9	24.45	0.69	0.30
	0.5	7.1	24.44	0.69	0.30
100	0.3	4.5	23.76	0.66	0.30
	0.4	4.3	23.63	0.65	0.30
	0.5	4.4	23.56	0.64	0.31

Table 7. Comparisons with additional baselines.

(a) Deblurring, kernel=uniform						(b) Edges2Handbags (256×256)			
Method	NFE	FID↓	PSNR↑	SSIM↑	LPIPS↓	Method	PSNR↑	SSIM↑	LPIPS↓
I3SB	10	6.5	24.17	0.65	0.32	NADB	19.25	0.740	0.231
NADB		4.8	27.70	0.81	0.18	RDBM	19.26	0.738	0.224
I3SB	100	4.9	22.70	0.58	0.33	GOUB	16.58	0.700	0.288
NADB		3.4	27.39	0.80	0.17	RDDM	14.66	0.645	0.256

(c) Comparison on DrealSR dataset, NFE=40					
Method	LPIPS↓	MUSIQ↑	MANIQA↑	ClipIQA↑	LIQE↑
NADB	0.315	71.480	0.483	0.576	3.610
DiT4SR	0.365	64.950	0.627	0.548	3.964

10.2. Additional Results

Tab. 7 provides additional image restoration, image translation, and super-resolution comparison experiments. Fig. 9 to 11 provide additional qualitative results on each restoration tasks, Fig. 12, 13 provide additional qualitative results on each translation tasks, and Fig. 14 provides additional examples comparing between NADB and I2SB w.r.t. various NFE sampling.

Figure 9. Generative processes of NADB on JPEG restoration tasks. **Top 4 rows**: QF=5. **Bottom 3 rows**: QF=10.

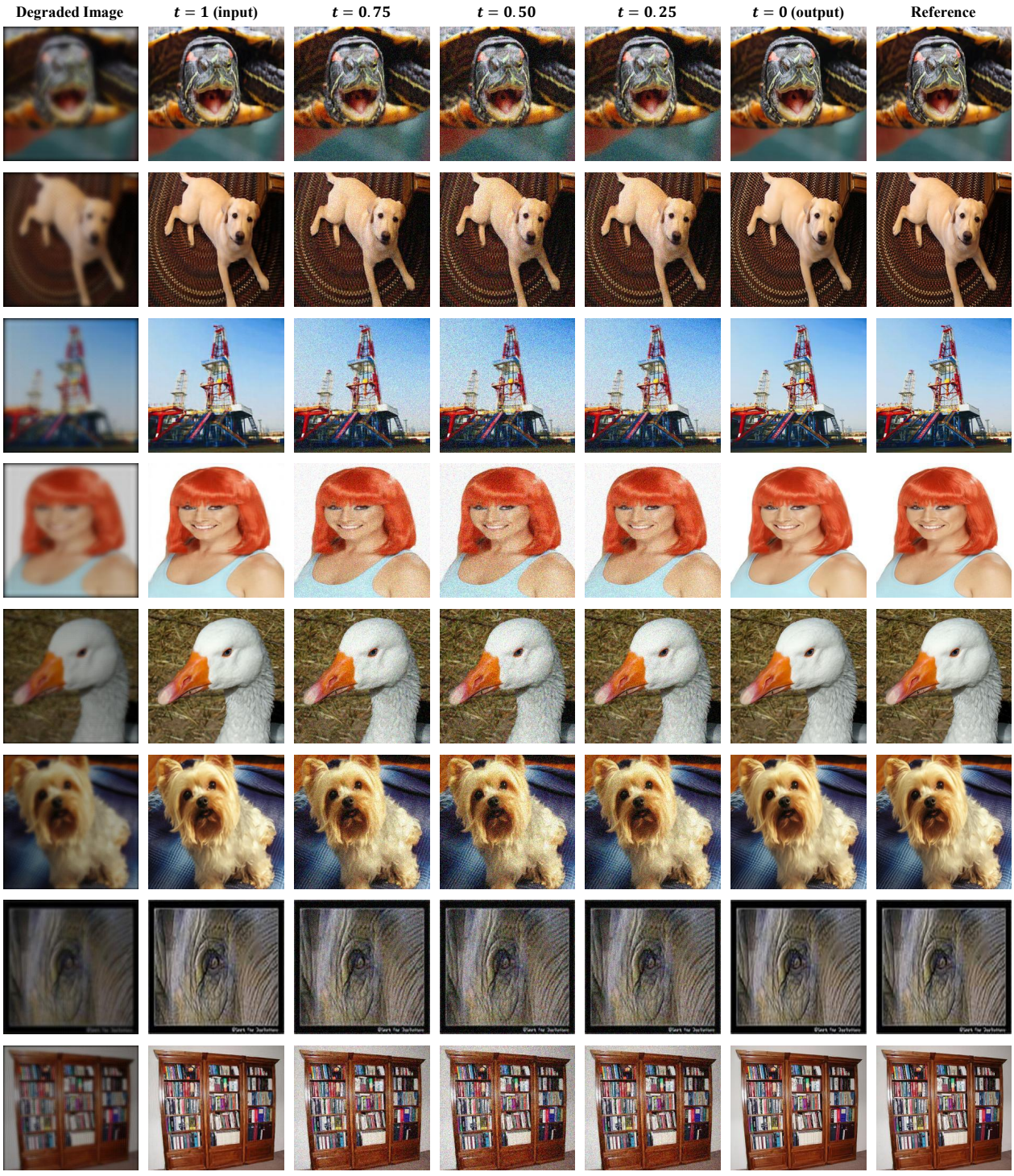

Figure 10. Generative processes of NADB on deblurring tasks. **Top 4 rows**: Uniform kernel. **Bottom 4 rows**: Gaussian kernel.

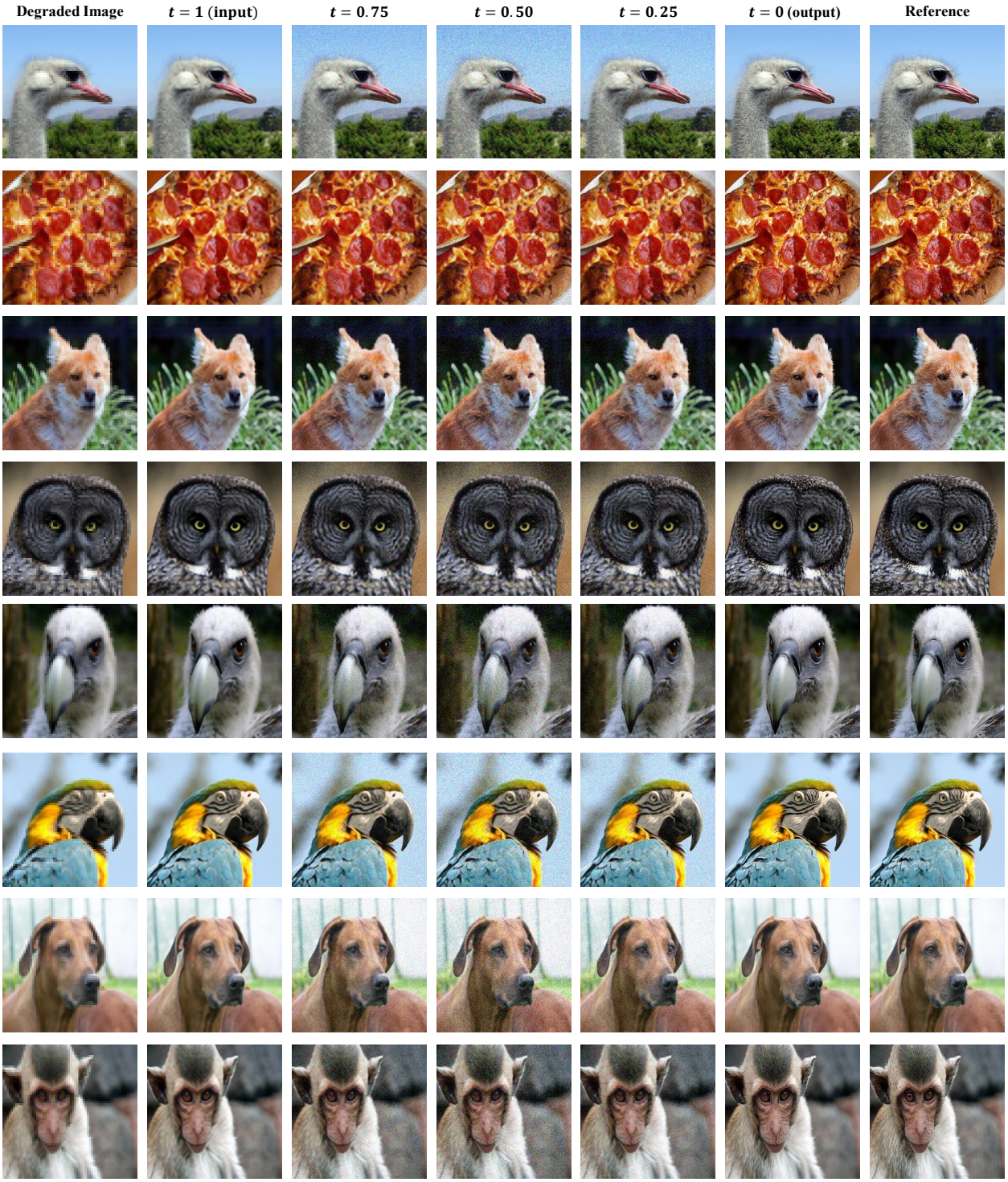

Figure 11. Generative processes of NADB on 4 × super-resolution tasks. **Top 4 rows**: Pool kernel. **Bottom 4 rows**: Bicubic kernel.

Figure 12. Generate results on edges2handbags dataset with resolution 64×64.

Figure 13. Generate results on edges2shoes dataset with resolution 64×64.

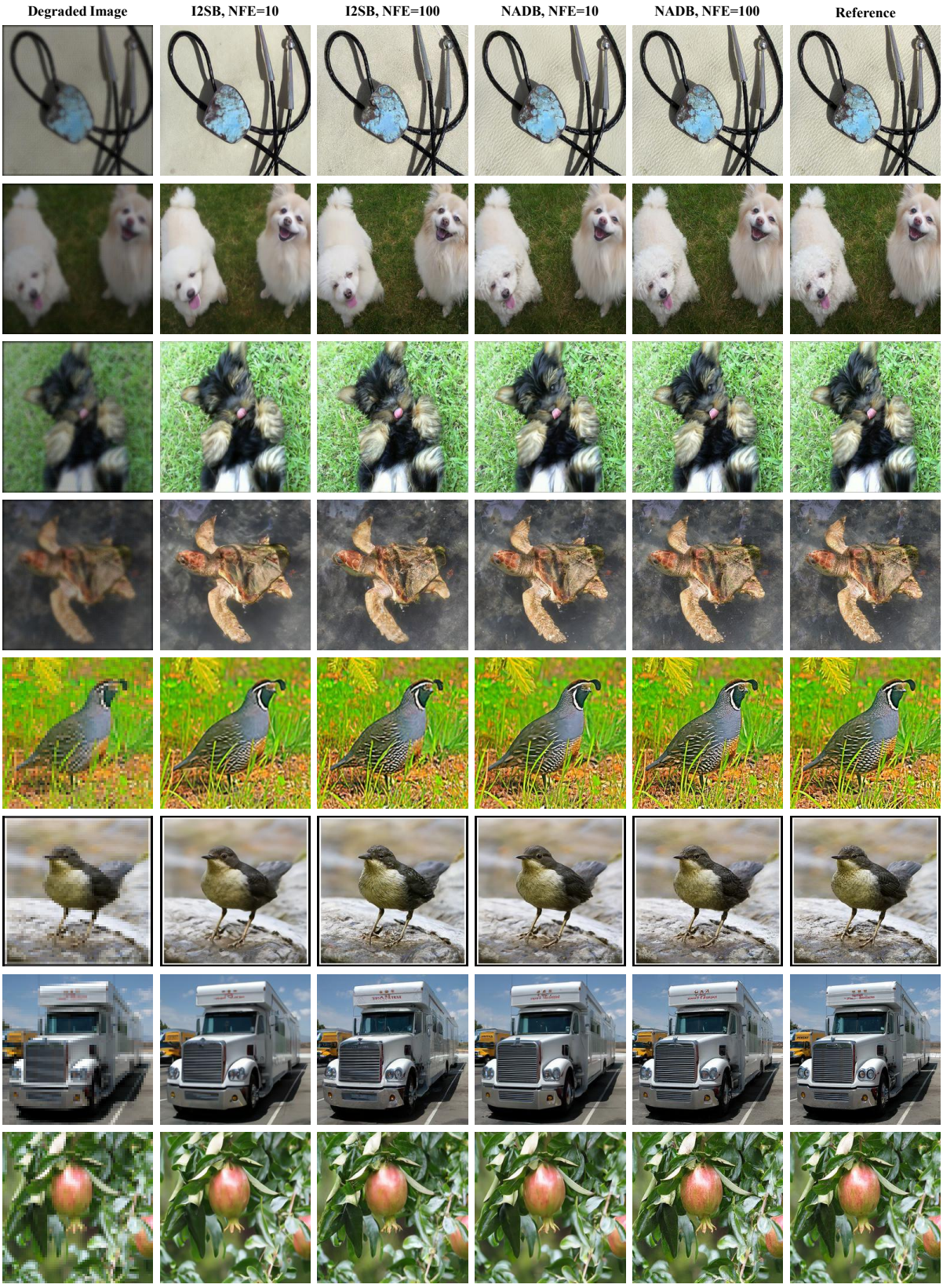

Figure 14. Additional qualitative comparison between NADB and I2SB. **Top 4 rows**: Deblurring tasks with gaussian kernel. **Bottom 4 rows**: 4 $\times$ super-resolution tasks with pool kernel.