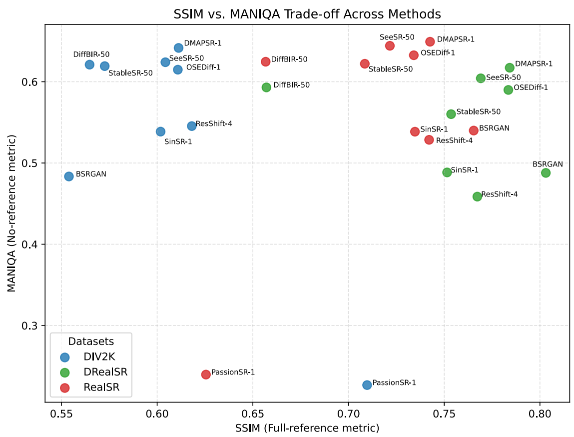

Figure 7: Scatter plot illustrating model performance, where SSIM (full-reference metric) is shown on the x-axis and MANIQA (no-reference metric) on the y-axis. This visualization enables a fair comparison across methods and clearly highlights the trade-off characteristics between reference-based fidelity and perceptual quality.

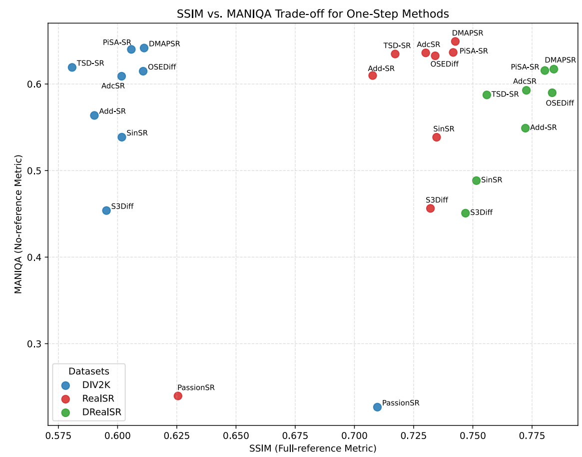

Figure 8: Scatter plot illustrating model performance of one step methods, where SSIM (full-reference metric) is shown on the x-axis and MANIQA (no-reference metric) on the y-axis. This visualization enables a fair comparison across methods and clearly highlights the trade-off characteristics between reference-based fidelity and perceptual quality.

Supplementary Material

In the supplementary material, we provide the following additional details:

- Scatter plot illustrating model performance in Fig.7 and Fig.8.
- A complete proof of the posterior energy formulation presented in the main paper.
- MAP estimation with MRF in a multi-step Diffusion model
- Additional qualitative results to further support our findings.

Proof of the MRF energy of the posterior:

Theorem: Let the prior distribution $\mathbb{P}(X=\omega)$ be a Gibbs distribution defined over a neighborhood $\{S,\mathcal{G}\}$ with corresponding energy $\mathcal{E}$ and potential$\{V_C\}$: $\mathbb{P}(X=\omega)=e^{-\mathcal{E}(\omega)}/Z$, $\mathcal{E}(\omega)=\sum_C V_C(\omega)$, where $\omega=(f,l)$. Then, for any fixed observation g, the posterior distribution $\mathbb{P}(X=\omega|G=g)$ is also a Gibbs distribution, defined over the neighborhood system $\{S,\mathcal{G}^P\}$, with the posterior energy function given by:

$$\mathcal{E}^P(f,l)=\mathcal{E}(f,l)+\frac{1}{2\sigma^2}||\mu-\Phi(g,\mathcal{D}_\psi(H(F)))||^2 \tag{14}$$

where $\mathcal{G}^P$ denotes the extended neighborhood system defined as:

$$\mathcal{G}_s^P=\begin{cases}\mathcal{G}_s, & \text{if } s\in D_m\\ \mathcal{G}_s\cup\mathcal{H}_s^2\setminus\{s\}, & \text{if } s\in Z_m\end{cases} \tag{15}$$

proof: We start with the definition of the degradation operator, $G=\phi(H(F))\odot\mathcal{N}$, where $\mathcal{N}\sim\mathcal{N}(\mu,\sigma^2)$ is additive white Gaussian noise, assumed independent of the MRF $\{S,\mathcal{G}\}$. The operation $\odot$ is assumed to be invertible, such that we can write $\mathcal{N}=\Phi(G,\mathcal{D}_\psi(H(F)))=\{\Phi_s, s\in Z_m\}$. Let $\mathcal{H}_s, s\in Z_m$ denote the set of pixels that affect the blurred image $H(F)$ at s. For instance $\mathcal{H}_s$ can be a 3×3 window centered at s. $\{\Phi_s, s\in Z_m\}$ depends only on g_s and $\{f_t: t\in\mathcal{H}_s\}$. Because of the shift invariance of H, these neighborhoods satisfy $\mathcal{H}_{r+s} = s + \mathcal{H}_r$ where $\mathcal{H}_r \subseteq Z_m, s + r \in Z_m$ and $s + \mathcal{H}_r$ intersects Z_m. If $\{\mathcal{H}_s\}$ is symmetric such that $r \in \mathcal{H}_0 \implies -r \in \mathcal{H}_0$, then the collection $\{\mathcal{H}_s \backslash \{s\}, s \in Z_m\}$ forms a valid neighborhood system over Z_m. Let $\mathcal{H}^2$ define the second-order neighborhood system as:

$$\mathcal{H}_s^2 = \cup_{r \in \mathcal{H}_s} \mathcal{H}_r, s \in Z_m \tag{16}$$

Then $\{\mathcal{H}_s^2 \setminus \{s\}, s \in Z_m\}$ also defines a neighborhood system. We define the posterior neighborhood system $\{\mathcal{G}^P = \mathcal{G}_s^P, s \in S\}$ as,

$$\mathcal{G}_s^P = \begin{cases} \mathcal{G}_s, & \text{if } s \in D_m \\ \mathcal{G}_s \cup \mathcal{H}_s^2 \setminus \{s\}, & \text{if } s \in Z_m \end{cases} \tag{17}$$

Applying Bayes' rule, we express the posterior as:

$$\mathbb{P}(X = \omega | G = g) = \frac{\mathbb{P}(G = g | X = \omega) \cdot \mathbb{P}(X = \omega)}{\mathbb{P}(G = g)} \tag{18}$$

$\forall\ \omega = (f, l)$ and each g. Assuming $\mathbb{P}(X = \omega) = e^{-\mathcal{E}(\omega)}/Z$, the likelihood term becomes:

$$\begin{aligned} \mathbb{P}(G = g | X = \omega) &= \mathbb{P}(\mathcal{D}_\psi(H(X)) \odot \mathcal{N} = g | X = \omega) \\ &= \mathbb{P}(\mathcal{N} = \Phi(g, \mathcal{D}_\psi(H(X)))) \\ &= (2\pi\sigma^2)^{-M/2} \exp\{-\frac{1}{2\sigma^2} ||\mu - \Phi||^2\} \end{aligned} \tag{19}$$

Again, $\mathbb{P}(X = \omega | G = g) = e^{-\mathcal{E}^P(\omega)}/Z^P$.

Case $s \in Z_m$: The term Φ does not cancel out. $\Phi(g, \mathcal{D}_\psi(H(F))) = \{\Phi_s, s \in Z_m\}$.

Taking Eq 18 and Eq. 19 we can write,

$$\mathbb{P}(X = \omega \mid \mathcal{N} = \Phi) \propto \mathbb{P}(\mathcal{N} = \Phi \mid X = \omega) \cdot \mathbb{P}(X = \omega). \tag{20}$$

Taking the negative logarithm, the posterior energy becomes:

$$\mathcal{E}^P(f, l) = -\log \mathbb{P}(\mathcal{N} = \Phi \mid X = \omega) - \log \mathbb{P}(X = \omega). \tag{21}$$

From Eq. 19 we get,

$$-\log \mathbb{P}(\mathcal{N} = \Phi \mid X = \omega) = \frac{1}{2\sigma^2} \sum_{r \in Z_m} \left(\Phi_r(g_r; f_t,\, t \in \mathcal{H}_r) - \mu\right)^2 + \text{const.} \tag{22}$$

Combining both terms, the full posterior energy becomes:

$$\mathcal{E}^P(f, l) = \sum_C V_C(f, l) + \frac{1}{2\sigma^2} \sum_{r \in Z_m} \left(\Phi_r(g_r; f_t,\, t \in \mathcal{H}_r) - \mu\right)^2. \tag{23}$$

$$\begin{aligned} \mathbb{P}(F_s = f_s \mid F_r = f_r,\ r \neq s,\ r \in Z_m,\ L = l,\ G = g) &= \frac{e^{-\mathcal{E}^P(\omega)}/Z^P}{\sum_{f_s} e^{-\mathcal{E}^P(\omega)}/Z^P} = \frac{e^{-\mathcal{E}(\omega)}/Z}{\sum_{f_s} e^{-\mathcal{E}(\omega)}/Z} \\ &= \frac{e^{-\mathcal{E}(f,l) - \frac{1}{2\sigma^2} \sum_{r \in Z_m} (\Phi_r - \mu)^2}}{\sum_{f_s} e^{-\mathcal{E}(f,l) - \frac{1}{2\sigma^2} \sum_{r \in Z_m} (\Phi_r - \mu)^2}} \end{aligned} \tag{24}$$

$$\Rightarrow \quad \mathcal{E}^P(f,l) = \sum_{C:s\in C} V_C(f,l) + \frac{1}{2\sigma^2}\sum_{r:s\in\mathcal{H}_r}\left(\Phi_r(g_r; f_t,\ t\in\mathcal{H}_r)-\mu\right)^2$$
$$+ \sum_{C:s\notin C} V_C(f,l) + \frac{1}{2\sigma^2}\sum_{r:s\notin\mathcal{H}_r}\left(\Phi_r(g_r; f_t,\ t\in\mathcal{H}_r)-\mu\right)^2 \tag{25}$$

It can be seen that the last two terms in 25 does not involve f_s and the ratio in 24 depends only on the first two terms of 25. The first two terms depends only on the coordinate (f,l) for the sites in $\mathcal{G}_s\{s\in C \implies C\subseteq\mathcal{G}_s\}$ and the second term only on the sites in $=\cup_{r:s\in\mathcal{H}_r}\mathcal{H}_s = \cup_{r\in\mathcal{H}_s}\mathcal{H}_r = \mathcal{H}_s^2$. Hence we can say, $\mathcal{G}_s^P = \mathcal{G}_s \cup \mathcal{H}_s^2\backslash\{s\}$.

Case $s\in D_m$:

$$\mathbb{P}(L_s = l_s | L_r = l_r, r\neq s, r\in D_m, F=f, G=g)$$
$$= \frac{e^{-\mathcal{E}^P(\omega)}/Z^P}{\sum_{l_s} e^{-\mathcal{E}^P(\omega)}/Z^P} = \frac{e^{-\mathcal{E}(\omega)}/Z}{\sum_{l_s} e^{-\mathcal{E}(\omega)}/Z}$$

The sum extends over all possible values of L_s Hence we can say, $\mathcal{G}_S^P = \mathcal{G}_S$.

Thus, the posterior energy becomes,

$$\mathcal{E}^P(f,l) = \mathcal{E}(f,l) + \frac{1}{2\sigma^2}||\mu - \Phi(g, \mathcal{D}_\psi(H(F)))||^2 \tag{26}$$

Corollary: It can be observed that the second term is strictly positive. To generalize this further, we note that this term can be interpreted as a discrepancy measure between the likelihood and the prior. While the KL divergence is a common choice—being strictly positive—other discrepancy measures may also be employed. To demonstrate the similarity between the second term and the KL divergence, we proceed as follows:

$$\mathbb{P}(G=g|X=\omega) = \frac{1}{(2\pi\sigma^2)^{M/2}}\exp\{-\frac{1}{2\sigma^2}||\mu - \Phi(g,\mathcal{D}_\psi(H(F)))||^2\}$$
$$\implies \log\mathbb{P}(G=g|X=\omega) = -\frac{1}{2\sigma^2}||\mu - \Phi(g,\mathcal{D}_\psi(H(F)))||^2 + C$$

Where $C = -\frac{M}{2}\log(2\pi\sigma^2)$ as X is independent of $\mathcal{N}$. Now, taking $\langle h(X)\rangle = E_X[h(X)]$ we can write the above as,

$$\mathrm{KL}[\mathbb{P}(X=\omega)||\mathbb{P}(G=g|X=\omega)] = \langle\log\mathbb{P}(X=\omega)\rangle - \langle\log\mathbb{P}(G=g|X=\omega)\rangle$$
$$= \frac{1}{2\sigma^2}\langle||\mu - \Phi(g,\mathcal{D}_\psi(H(F)))||^2\rangle + C - \langle -\frac{U(\omega)}{Z}\rangle$$
$$= \frac{1}{2\sigma^2}\langle||\mu - \Phi(g,\mathcal{D}_\psi(H(F)))||^2\rangle + C$$

Hence, effectively we can write the posterior energy as,

$$\mathcal{E}^P(f,l) = \mathcal{E}(f,l) + \mathrm{KL}[\mathbb{P}_{\mathbf{Y}|\mathbf{X}}(g|f,l)||\mathbb{P}_{\mathbf{X}}(f,l)] \tag{27}$$

MAP Estimation with MRF in a Multi-Step Diffusion Model

A T-step diffusion model defines a sequence of latent variables

$$x_0 \to x_1 \to \cdots \to x_T,$$

generated by the forward noising process

$$q(x_t \mid x_{t-1}) = \mathcal{N}(\alpha_t x_{t-1},\ (1-\alpha_t)\mathrm{I})\,.$$

A multi-step reverse sampler learns the reverse Markov chain

$$\mathbb{P}_\theta(x_{t-1} \mid x_t) = \mathcal{N}\big(\mu_\theta(x_t,t),\, \sigma_t^2 \mathrm{I}\big)\,,$$

where the mean is determined using the learned score:

$$\mu_\theta(x_t,t) = \frac{1}{\alpha_t}\left(x_t - \frac{1-\bar{\alpha}_t}{1-\alpha_t}\,\epsilon_\theta(x_t,t)\right).$$

Multi-step diffusion is equivalent to MAP with a Gaussian prior

In the generative model, the reverse conditional satisfies

$$\mathbb{P}(x_{t-1} \mid x_t) \propto \exp\!\left(-\frac{1}{2\sigma_t^2}\|x_{t-1} - \mu_\theta(x_t,t)\|^2\right).$$

For a given observation $\mathbf{Y}$ and unknown clean image $\mathbf{X}$, classical MAP solves:

$$\hat{\mathbf{X}}_{\mathrm{MAP}} = \arg\max_{\mathbf{X}} \left[\log \mathbb{P}(\mathbf{Y} \mid \mathbf{X}) + \log \mathbb{P}(\mathbf{X})\right].$$

Identifying $\mathbf{X} = x_0$ and $\mathbb{P}(\mathbf{X}) = \mathbb{P}(x_0)$, diffusion models define a hierarchical prior:

$$\mathbb{P}(x_0) = \int \mathbb{P}(x_T) \prod_{t=1}^{T} \mathbb{P}_\theta(x_{t-1} \mid x_t)\; dx_{1:T}.$$

Therefore, the multi-step MAP objective is

$$\hat{x}_0 = \arg\max_{x_0} \left[\log \mathbb{P}(\mathbf{Y} \mid x_0) + \sum_{t=1}^{T} \log \mathbb{P}_\theta(x_{t-1} \mid x_t)\right]. \tag{28}$$

The score network is the gradient of the log-prior

From DDPM:

$$\nabla_{\mathbf{x}_t} \log \mathbb{P}_t(x_t) = -\frac{1}{1-\bar{\alpha}_t}\;\epsilon_\theta(x_t,t).$$

Thus diffusion implicitly learns

$$\nabla_{x_0} \log \mathbb{P}(x_0),$$

which is the exact prior gradient appearing in MAP.

MAP combines:

$$\nabla_{x_0} \log \mathbb{P}(x_0) \quad \text{(diffusion score)} \qquad \text{and} \qquad \nabla_{x_0} \log \mathbb{P}(\mathbf{Y} \mid x_0) \quad \text{(likelihood)}.$$

Replacing diffusion's prior with an MRF prior. We use a structured MRF prior:

$$\log \mathbb{P}(\mathbf{X}) \propto -\mathcal{E}_{\mathrm{MRF}}(\mathbf{X}).$$

Thus we can replace or augment the diffusion prior term $\log \mathbb{P}_\theta(x_{t-1} \mid x_t)$ with $-\lambda\,\mathcal{E}_{\mathrm{MRF}}(x_t)$.

Multi-step Reverse With MAP Regularization

The DDPM update,

$$x_{t-1} = \mu_\theta(x_t, t) + \sigma_t z.$$

Under MAP regularization with MRF, the update becomes:

$$x_{t-1} = \mu_\theta(x_t, t) - \eta_t \nabla_{x_t} \mathcal{E}_{\text{MRF}}(x_t) + \sigma_t z. \tag{29}$$

More compactly:

$$x_{t-1} = x_{t-1}^{\text{DDPM/DDIM}} - \eta_t \nabla_{x_t} \mathcal{E}_{\text{MRF}}(x_t).$$

Single-step diffusion as a limiting case

Our one-step estimator:

$$x_0' = x_T + \sigma_T g_\phi(x_T)$$

is the $T \to 1$ collapse of equation 29.

If

$$\eta = \sum_{t=1}^{T} \eta_t,$$

then the MAP update becomes:

$$x_0 \approx x_T - \eta \nabla_{x_T} \mathcal{E}_{\text{MRF}}(x_T).$$

Our learned enhancer satisfies:

$$g_\phi(x_T) \approx -\nabla_{x_T} \mathcal{E}_{\text{MRF}}(x_T).$$

Thus:

$$x_0' = x_T + \sigma_T g_\phi(x_T)$$

is a learned proximal/MAP update.

In summary, A T-step diffusion model defines the prior

$$\mathbb{P}(x_0) = \int \mathbb{P}(x_T) \prod_{t=1}^{T} \mathbb{P}_\theta(x_{t-1} \mid x_t)\, dx_{1:T},$$

and its reverse process

$$\mathbb{P}_\theta(x_{t-1} \mid x_t) = \mathcal{N}\big(\mu_\theta(x_t, t),\, \sigma_t^2 I\big)$$

approximates $\nabla_{x_0} \log \mathbb{P}(x_0)$.

Thus MAP inference with a likelihood term becomes:

$$\hat{x}_0 = \arg\max_{x_0} \left[\log \mathbb{P}(\mathbf{Y} \mid x_0) + \sum_{t=1}^{T} \log \mathbb{P}_\theta(x_{t-1} \mid x_t) \right].$$

Substituting the Gaussian form yields the multi-step MAP update:

$$x_{t-1} = \mu_\theta(x_t, t) - \eta_t \nabla_{x_t} \mathcal{E}_{\text{MRF}}(x_t) + \sigma_t z.$$

Our single-step model is a limiting case of this regularized reverse diffusion chain.

Table 7: Ablation on the number of MAP-regularized reverse diffusion steps. Increasing the number of steps improves fidelity by repeatedly applying the MAP-corrected update (Eq. 29), while fewer steps favor realism and efficiency.

Steps	PSNR ↑	LPIPS ↓	MUSIQ ↑	CLIPIQA ↑	SSIM ↑	DISTS ↓	NIQE ↓	MANIQA ↑
1	28.32	0.2957	64.97	0.6975	0.7842	0.2096	6.3245	0.6172
2	28.35	0.2954	65.02	0.6982	0.7851	0.2087	6.3217	0.6175
3	28.12	0.2972	65.14	0.6992	0.7847	0.2092	6.3198	0.6162
4	28.15	0.2961	65.21	0.6986	0.7855	0.2099	6.3157	0.6181
5	28.25	0.2941	66.12	0.7005	0.7844	0.2093	6.3106	0.6185

Additional results:

We present additional comparative results with existing diffusion model-based methods in Fig. 9. Our method demonstrates superior performance, particularly in recovering fine structures such as artificial flower petals, leaf textures, and cloth patterns, under both ground truth and non-ground truth scenarios.

Multistep Formulation: As analyzed in the previous section, MAP regularization only needs to be applied to the intermediate latent states $\{x_t\}$ that participate in the reverse diffusion process. In particular, the MAP-corrected update (Eq. equation 29) adjusts the DDPM/DDIM mean by incorporating the gradient of the MRF energy, thereby refining the predicted sample at each step. To flexibly evaluate the effect of MAP regularization across different diffusion horizons, we apply the correction term at several pre-selected timesteps, corresponding to varying numbers of sampling iterations. Once trained, the starting timestep for the reverse chain can be freely chosen at inference time, providing a controllable trade-off between fidelity (larger number of MAP-corrected steps) and realism (fewer steps, closer to the learned prior), analogous to the behavior observed in classical multi-step diffusion samplers. The total number of steps used during inference is determined jointly by the chosen starting timestep and the skipping stride of the accelerated DDIM-style sampler. For clarity, we report detailed results for 1 to 5 MAP-corrected steps in Table 7, reflecting a practical trade-off between performance and computational budget.

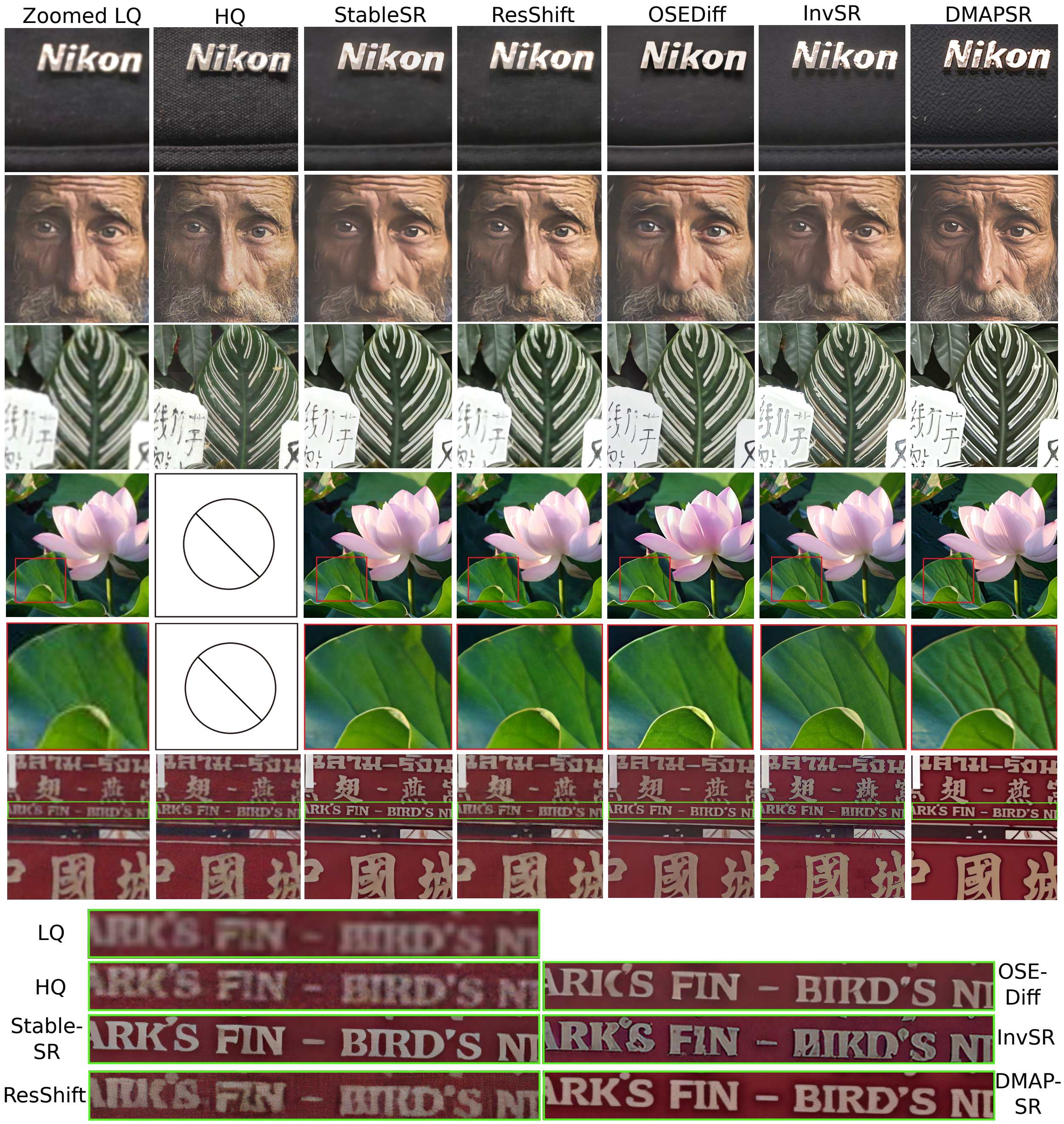

Figure 9: Qualitative visual comparisons of Real-ISR methods are presented. Note that the third example lacks a corresponding high-quality ground truth image. Please zoom in for a clearer view.