PLUG-IN IMAGE QUALITY CONTROL FOR POSTERIOR DIFFUSION SUPER-RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion-based super-resolution (SR) has shown remarkable progress, mainly through prior-guided approaches that require explicit degradation models or semantic priors. While posterior diffusion SR avoids these assumptions by directly learning from LR–HR pairs, it still suffers from numerical errors during sampling and lacks plug-and-play mechanisms for quality control. We introduce the first plug-and-play framework for posterior diffusion SR, enabling pretrained models to support controllable quality without retraining. Our numerical analysis reveals that discretization errors are a key bottleneck in posterior SR. We prove that these errors can be equivalently expressed as gradients of KL divergence, unifying numerical error correction with image based classifier guidance. This provides a principled explanation of fidelity degradation and a new lens for posterior diffusion trajectories. In principle, these errors can be corrected to improve fidelity when reference supervision is available, offering a new theoretical understanding of posterior diffusion trajectories. In real-world SR, we further show that our image based guidance offers a controllable trade-off between fidelity and perception, delivering perceptual sharpness competitive with state-of-the-art prior-based models. Experiments confirm that our method consistently improves perceptual quality, while also validating the theoretical link between numerical errors and fidelity in posterior SR. These results position our work as a new direction for posterior diffusion models, bridging probabilistic analysis with practical deployment.

1 INTRODUCTION

Image super-resolution (SR) is a fundamental problem in low-level vision, aiming to recover high-resolution (HR) images from their low-resolution (LR) counterparts. Classical approaches include regression-based methods Mehta et al. (2023); Wang et al. (2023a); Lim et al. (2017); Liang et al. (2021a) and adversarial training with GANs Ledig et al. (2017); Wang et al. (2021; 2018), but these methods still struggle with the intrinsic ill-posedness of SR.

Diffusion models Ho et al. (2020); Song et al. (2021b) have recently emerged as a powerful generative framework, leveraging a physics-driven probabilistic process that excels at reconstructing high-frequency details in ill-posed image restoration problems Saharia et al. (2022b); Wang et al. (2023c); Li et al. (2022a); Kawar et al. (2022). While diffusion models were initially criticized for their high computational cost, subsequent advances such as improved samplers Song et al. (2021a); Yue et al. (2023) and distillation-based acceleration Wu et al. (2024a); Dong et al. (2024); Chen et al. (2025; 2024c); Yue et al. (2025); Lin et al. (2024); Wu et al. (2024b); Yang et al. (2024) have reduced the sampling steps to single or few iterations, making runtime comparable to regression-based SR. Moreover, semantic priors such as CLIP have been employed to guide perceptual quality.

Despite these advances, most existing works adopt a *diffusion prior* paradigm: pretrained generative priors with explicit or learned degradation models Saharia et al. (2021); Li et al. (2022c). However, diffusion priors require an accurate likelihood formulation of the degradation process, which is often intractable beyond simplistic assumptions such as Gaussian blur or bicubic downsampling. While higher-order solvers Lu et al. (2022; 2023); Zheng et al. (2023) mitigate discretization error in the generative trajectory, they remain dependent on assumed likelihoods that do not align well with real-world degradations, especially those introduced by optical lenses.

Table 1: Prior vs. Posterior Diffusion SR

Aspect	Diffusion Prior SR	Posterior Diffusion SR
Training	Learn prior $p(x)$ + likelihood $p(y\|x)$	Directly learn $p(x\|y)$ from LR–HR pairs
Assumption	need explicit degradation model. Bicubic, Gaussian)	Covered by Dataset
Plug-and-play	CLIP Radford et al. (2021)	Ours: first plug-in model
Distillation	Actively used	Not Active

In contrast, *posterior learning* approaches define the degradation process directly through LR-HR training pairs Cai et al. (2019a); Wang et al. (2021); Bhat et al. (2021). This formulation bypasses the need for explicit likelihood modeling and naturally inherits the legacy of regression-based SR datasets. However, posterior diffusion models remain underexplored: unlike diffusion priors, no plug-and-play modules for image quality control have been established on pretrained posterior models.

From a numerical analysis perspective, diffusion samplers typically employ first-order forward discretization, which is computationally efficient but induces significant numerical error. Existing efforts Lu et al. (2022; 2023); Zheng et al. (2023) target discretization in diffusion priors, but posterior models have not been systematically analyzed in this regard. Notably, Yue et al. (2023) adopts a single-order solver, which remains susceptible to discretization error. For general image generation, Li et al. Li & van der Schaar (2024) analyzed cumulative error between forward and backward processes, but the role of discretization error in SR has yet to be explicitly studied.

Our contribution. The contributions of this work are summarized as follows:

- We propose the first *plug-and-play module* for posterior diffusion SR, enabling pretrained models to support image quality control without retraining.
- We theoretically derive that discretization errors in first-order posterior models can be corrected by incorporating second-order differential terms, providing a principled explanation of fidelity degradation.
- We further propose an *image-based classifier guidance formulation* for posterior diffusion and prove that it is mathematically equivalent to the second-order correction term, thereby unifying numerical analysis with guidance-based conditioning.
- While explicit fidelity correction requires HR references and is thus of limited use in blind SR, we show that the same formulation can be adapted in a sign-flipped manner to enhance perceptual quality in reference-free scenarios.
- This dual view—fidelity-oriented correction under supervision and perception-oriented control without supervision—establishes a unified framework for posterior SR.
- Extensive analysis and experiments validate that our approach bridges the gap between numerical error theory and real-world perceptual enhancement in posterior learning models.

2 Related Works

Image Super-Resolution Since the emergence of deep neural network-based image super-resolution Liang et al. (2021a); Zhang et al. (2021); Dong et al. (2011); Gu et al. (2015); Zamir et al. (2022), significant advancements have been made in this field. Generative adversarial networks (GANs) Goodfellow et al. (2014) have played a crucial role in enhancing details in super-resolution outputs Wang et al. (2021; 2018). By leveraging adversarial learning, GANs improve texture sharpness and realism in super-resolved images. Recently, diffusion models Saharia et al. (2022b); Li et al. (2022a); Kawar et al. (2022); Yue et al. (2023); Wang et al. (2023c) have emerged as a powerful alternative for image super-resolution. Originally developed for novel image synthesis, these models have demonstrated exceptional capabilities in texture restoration and fine-detail reconstruction. Unlike GANs, diffusion models progressively refine images from noise, providing a more stable and controlled SR framework. Studies have shown that diffusion-based SR outperforms GAN-based approaches, particularly in preserving high-frequency details and reducing artifacts Rombach et al. (2022a). Furthermore, developments in VQGAN Esser et al. (2021) and stable diffusion models Rombach et al. (2022a) have significantly influenced image synthesis Saharia et al. (2022a), expanding beyond SR into text-conditioned image generation. These advancements have played a pivotal role in integrating SR capabilities within broader generative frameworks.

Performance Enhancement Diffusion models for image super-resolution face an inherent challenge of high computational overhead Saharia et al. (2022b); Ho et al. (2020). Various efforts have been made to address this issue Song et al. (2021a); Lu et al. (2022; 2023); Zheng et al. (2023); Wu et al. (2024a); Wang et al. (2024a). Starting with DDIM, subsequent works on DPM solvers have significantly reduced the number of iteration steps from thousands to just tens. Residual-based approaches in image synthesis Liu et al. (2024) and image super-resolution Yue et al. (2023) have demonstrated stable performance even with fewer than ten steps. More recently, techniques such as model distillation Wang et al. (2024a) and single-step SR Wu et al. (2024a) have emerged, further pushing the boundaries of efficiency in diffusion-based SR.

Error Estimation Reducing the number of iteration steps inherently increases step intervals in finite difference methods, making diffusion models more susceptible to numerical errors Strang (2007); Peter & Eckhard (1992); Hochbruck & Ostermann (2005). Consequently, research has been conducted to analyze and mitigate numerical errors and their effects Li & van der Schaar (2024); Li et al. (2024). Trajectory analysis Chen et al. (2024b;a) is also effective for detailed analysis of numerical errors. In Li & van der Schaar (2024), the authors investigated errors in pretrained diffusion models by analyzing discrepancies between forward and reverse processes. They defined **modular errors** between these processes and extended their analysis to **cumulative errors** throughout the entire process. Similarly, Li et al. (2024) examined how iterative inference steps exacerbate exposure bias due to training-inference discrepancies and proposed a method to mitigate this issue without requiring DPM retraining.

High Order Differential Equation Solver Diffusion based super resolution initially began by learning the posterior distribution.Li et al. (2022a); Saharia et al. (2022c). However, the diffusion posterior sampling(DPS) approach has gained attention for its efficiency , as it enables sampling using a pre-trained prior without requiring additional training costs. Most existing higher-order ODE/SDE solvers (Heun, RK2/3, DPM-Solver-2/3, etc.) have been developed in the context of DDPM-like models, where the prior is learned and the sampling process is derived from a score-based or probability-flow ODE interpretation Song et al. (2021a); Karras et al. (2022); Lu et al. (2022; 2023); Zheng et al. (2023); Liu et al. (2022). These works typically apply numerical integration of a pre-trained prior distribution, thus enabling explicit higher-order schemes for fast sampling. However, compared to the DPS approach, which reuses a pre-trained model trained with a distribution based loss, the posterior learning using pixel-based loss is still considered advatangeous in terms of final image quality. Yue et al. (2023) Thus, our setting focuses on a **learned-posterior** approach rather than a learned-prior approach, where the solver is based on the first-order Euler-type update. The higher-order solvers introduced in prior-based diffusion works may not be readily applicable to the learned posterior method without re-architecting the model to fit a score-based ODE sampling paradigm.

3 Plug and Play for Posterior Diffusion Model

3.1 Background for Numerical Accuracy

A Taylor series expansion can be written as

$$x(t+h) = x(t) + h\,\nabla x(t) \;+\; \frac{1}{2}\,h^2\,\nabla^2 x(t) \;+\; \mathcal{O}(h^3).$$

In a first-order difference scheme, we usually adopt

$$x(t+h) = x(t) + h\,\nabla x(t) + DE,$$

where DE is the leading-order discretization error, whose dominant term is $\frac{1}{2}\,h^2\,\nabla^2 x(t)$. If we explicitly include this second-order term in the difference scheme, the accuracy increases. Most existing diffusion models, however, only consider a first-order discretization, whereas in this paper, we adopt a second-order discretization that accounts for $\frac{1}{2}\,h^2\,\nabla^2 x(t)$. Higher-order discretization methods and their errors have also been analyzed in Lu et al. (2022; 2023); Zheng et al. (2023). While conventional formulas rely on starting the restoration from white Gaussian noise (thus requiring many diffusion steps), recent works on one-step or few-step diffusion models employ large step intervals, which can cause large discretization errors. Therefore, it becomes necessary to correct such discretization errors.

DPM-Solver Revisited. Consider a random variable x_0 sampled from $q_0(x_0)$. The forward SDE on the interval $[0, T]$ is

$$d\mathbf{x}_t = f(t)\,\mathbf{x}_t\,dt + g(t)\,d\mathbf{w}_t, \quad \mathbf{x}_0 \sim q_0(\mathbf{x}_0),$$

where $\mathbf{w}_t \in \mathbb{R}^D$ is a standard Wiener process. Song et al. (2021b) show that its reverse process from T to 0, given the marginal $q_T(x_T)$, is

$$d\mathbf{x}_t \;=\; \big[f(t)\,\mathbf{x}_t - g^2(t)\,\nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t)\big]\,dt \;+\; g(t)\,d\bar{\mathbf{w}}_t, \quad \mathbf{x}_T \sim p_T(\mathbf{x}_T),$$

where $\bar{\mathbf{w}}_t$ is a standard Wiener process in reverse time. From the above, Lu et al. (2022; 2023) transform it into a probability flow ODE for faster sampling:

$$d\mathbf{x}_t \;=\; \big[f(t)\,\mathbf{x}_t \;-\; \tfrac{1}{2}g^2(t)\,\nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t)\big]\,dt. \tag{1}$$

To estimate the score function $\nabla_x \log q_t(x_t)$, DPMs use a neural network $\epsilon_\theta(x_t, t)$. The parameter θ is optimized by minimizing

$$\mathcal{L}(\theta) \;=\; \int_0^T \mathbb{E}_{q_t(\mathbf{x}_t)} \big\|\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) + \nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t)\big\|_2^2\,dt,$$

yielding

$$d\mathbf{x}_t \;=\; f(t)\,\mathbf{x}_t + g^2(t)\,\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\,dt.$$

We can generate samples by numerically solving this ODE from T down to 0. Given an initial value x_s at time $s > 0$, Lu et al. (2022) show that the solution x_t for $t \in [0, s]$ can be written in integral form as

$$\mathbf{x}_t \;=\; \frac{\alpha_t}{\alpha_s}\,\mathbf{x}_s \;+\; \alpha_t \int_{\lambda_s}^{\lambda_t} e^{-\lambda}\,\boldsymbol{\epsilon}_\theta(\mathbf{x}_\lambda)\,d\lambda, \tag{2}$$

which in discrete form approximates

$$\mathbf{x}_t \;=\; \frac{\alpha_t}{\alpha_s}\,\mathbf{x}_s \;-\; \alpha_t \sum_{n=0}^{k-1} \boldsymbol{\epsilon}_\theta^{(n)}(\mathbf{x}_{\lambda_t}) \int_{\lambda_s}^{\lambda_t} F(\lambda)\,d\lambda, \tag{3}$$

where $F(\lambda)$ collects certain integral factors (see Lu et al. (2022)), and $\epsilon_\theta^{(n)}(\cdot)$ is the n-th order derivative for $n \le k-1$. For $k \ge 2$, they approximate these derivatives using intermediate points (Runge–Kutta). Although evaluating the integral more finely can reduce the discretization error, in practice it is typically done via finite-difference approaches like Runge–Kutta Hochbruck & Ostermann (2005). In our paper, however, we propose directly computing the second-order derivatives by taking gradients of the neural network model.

3.2 Background for Inverse Problem Solver

Super-resolution (SR) is an ill-posed inverse problem, where the goal is to recover the high-resolution (HR) image $\mathbf{x}$ from its degraded low-resolution (LR) observation $\mathbf{y}$. The forward measurement process can be expressed as $\mathbf{y} = \boldsymbol{A}(\mathbf{x}) + \mathbf{n}, \quad \mathbf{y}, \mathbf{n} \in \mathbb{R}^n,\ \mathbf{x} \in \mathbb{R}^d$ where $\boldsymbol{A}(\cdot) : \mathbb{R}^d \to \mathbb{R}^n$ is the degradation operator (e.g., bicubic downsampling, blur, or camera pipeline), and $\mathbf{n}$ denotes measurement noise. From a Bayesian perspective, the posterior distribution is given by

$$p(\mathbf{x}|\mathbf{y}) \;=\; p(\mathbf{y}|\mathbf{x})\,p(\mathbf{x})/p(\mathbf{y})$$

where $p(\mathbf{x})$ the prior over natural images, $p(\mathbf{y}|\mathbf{x})$ the likelihood for degradation, $p(\mathbf{x}|\mathbf{y})$ the posterior. In **Diffusion Prior** methods Saharia et al. (2021); Li et al. (2022b); Wang et al. (2023c); Wu et al. (2024b), the diffusion model serves as the image prior $p(\mathbf{x})$. The likelihood $p(\mathbf{y}|\mathbf{x})$ is typically assumed to be a Gaussian degradation model: $p(\mathbf{y}|\mathbf{x}) \;=\; \mathcal{N}(\mathbf{y}; A(\mathbf{x}), \sigma^2 I)$. Conditioning the diffusion process on $\mathbf{y}$ yields a posterior-guided reverse ODE:

$$d\mathbf{x}_t \;=\; \Big[f(t)\,\mathbf{x}_t \;-\; \tfrac{1}{2}g^2(t)\,\nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t|\mathbf{y})\Big]\,dt, \tag{4}$$

where the score term $\nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t|\mathbf{y})$ incorporates the assumed likelihood. Thus, the accuracy of prior-based SR crucially depends on the correctness of the degradation model.

Posterior learning approaches Yue et al. (2023) directly learn the conditional distribution $p(\mathbf{x}|\mathbf{y})$ from paired LR–HR data, bypassing the need for an explicit likelihood model. In this case, the degradation process $\boldsymbol{A}(\cdot)$ does not need to be analytically specified; it is implicitly encoded through the training dataset. The training objective is to maximize the conditional log-likelihood: $\mathcal{L}(\theta) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}}\big[\log p_\theta(\mathbf{x}|\mathbf{y})\big]$. This makes posterior approaches attractive for real-world SR, where the degradation is complex or unknown. However, posterior SR is sensitive to discretization errors during sampling and currently lacks plug-and-play quality control mechanisms, which are more natural in prior-based frameworks.

3.3 Theoretical Foundation

Error estimation in diffusion models has been addressed by Lu et al. (2022; 2023); Zheng et al. (2023); Li & van der Schaar (2024); Li et al. (2024). Most prior work focuses on measuring the difference between forward and reverse processes or errors from sampling intervals but their restoration has been less studied. Our interest lies in the discretization errors of finite-difference methods Strang (2007) and SDEs Peter & Eckhard (1992). Below, we define how we measure discretization error and Kullback–Leibler(KL)-based divergence error in our diffusion model for image restoration.

Assumption 1. (Definition of Error in the Diffusion Path) In Li & van der Schaar (2024), they defined the error in path of diffusion trajectory. The modular error $\mathbf{E}_t^{\text{modular}}$ measures the accuracy that every module maps its input to the output. $\mathbf{E}_t^{\text{modular}} = \mathbb{E}\big[D_{\text{KL}}\big(p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t) \,\|\, q(\mathbf{x}_{t-1} \mid \mathbf{x}_t)\big)\big]$ The cumulative error $\mathbf{E}_t^{\text{cumu}}$measures the amount of error which are accumulated for sequentially running the first $T-1$ denoising modules. $\mathbf{E}_t^{\text{cumu}} = D_{\text{KL}}\big(p_\theta(\mathbf{x}_{t-1}) \,\|\, q(\mathbf{x}_{t-1})\big)$ Here $p_\theta(\cdot)$ and $q(\cdot)$ denote the model distribution in the backward process and reference distribution in the forward process, following the notation of Li & van der Schaar (2024).

Assumption 2. (Predefined Variance Schedule) As in DDPM or DDIM, we assume each timestep t has a predefined fixed variance σ_t^2 that applies uniformly across temporal coordinates. This allows us to simplify the SDE-based error formulation into a constant-variance expression at each step.

Definition 3.2.1 (Second-order derivative as the numerical error) We define the *numerical error* (DE) of the first-order derivative solution as

$$\mathbf{x}_t^{\text{Exact}} - \mathbf{x}_t^{\text{Euler(1st)}} \approx \nabla_t\big[f(t)\,\mathbf{x}_t - \tfrac{1}{2}g^2(t)\,\nabla_x \log p_t(\mathbf{x}_t)\big].$$

It is the local truncation error of numerically integrating the backward ODE by Euler's method.

Lemma 3.2.1 (KL gradient in the Gaussian case) For two normal distributions $\mathcal{P} : \mathcal{N}(\mathbf{x}_t, \sigma_1^2)$, $\mathcal{Q} : \mathcal{N}(\boldsymbol{\mu}_t, \sigma_2^2)$, this shows that the difference between two Gaussian distributions can be expressed as a distance between their means.

$$\nabla D_{\text{KL}}\big(\mathcal{P}||\mathcal{Q})\big) \approx \tfrac{(\mathbf{x}_t - \boldsymbol{\mu}_t)}{\sigma_2^2}.$$

Proposition 3.2.1 (Constant Variance Leads to a Simpler Form). Let σ^2 be a predefined constant over the spatial dimensions. Then, the gradient with respect to t of the modified score function from Definition 3.2.1 satisfies

$$\nabla_t\big[f(t)\,\mathbf{x}_t - \tfrac{1}{2}g^2(t)\,\nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t)\big] \approx A \cdot \tfrac{(\mathbf{x}_t - \boldsymbol{\mu}_t)}{\sigma^2} + B.$$

where $A, B \in \mathbb{R}$ is the optimization constant to minimize discretization error. With constant variance, the numerical error can be found as a linear equation of the distance of two distributions.

Theorem 3.2.1 (Approximation Error and KL Gradient Scaling). For a given step t, the difference between the exact solution $\mathbf{x}^{\text{Exact}}$ and the Euler discretization $\mathbf{x}^{\text{Euler}}$ can be expressed as

$$\mathbf{x}^{\text{Exact}} - \mathbf{x}^{\text{Euler}} \approx A \cdot \nabla D_{\text{KL}}(\mathcal{P} \,\|\, \mathcal{Q}) + B.$$

where $\mathcal{P}$ is discrete distribution, $\mathcal{Q}$ is continuous exact distribution. Here, A, B can be derived exactly through linear regression between latant spaces of SR and HR . Intuitively, this tell us that discretization error at each step behaves like the KL gradient between the continuous and discrete processes.

Discussion. The detailed proofs of the lemma, proposition, and theorem are provided in the supplemental material. In summary, starting from Assumption 2, we derive Theorem 3.2.1, which establishes that the numerical error—characterized by a second-order derivative—can be formulated as the Kullback–Leibler (KL) divergence between distribution from the discrete process and the continuous process. Assumption 1 defined that cumulative error in a path is defined as divergence of the forward and backward process. Likewise, we assume that exact distribution is substituted with the distribution of the forward process and Euler distribution is with the distribution of the backward process. Theorem 3.2.1 define the error in a single step depends linearly on the gradient of KL divergence of the forward and backward process.

The distribution of the continuous process becomes the reference of the discrete process, which can be substituted with the forward process of HR image because it has better accuracy than the reverse process starting from LR image. Specifically, in the context of a conditional diffusion framework, the backward process is initialized from the low-resolution (LR) image, while the forward process originates from the high-resolution (HR) image. Consequently, we demonstrate that the second-order numerical discretization error exhibits a linear dependency on the KL divergence between the latent representations of the HR and LR images, as governed by the conditional posterior learning model.

In the preceding paragraphs, we defined how the second-order numerical error can be evaluated and corrected. The next step is to formulate the theorectical framework for appyling these equations to the existing posterior model. From Yue et al. (2023), the reverse process for estimating the posterior distribution $p(\mathbf{x}_0 \mid \mathbf{y}_0)$ is defined by:

$$p(\mathbf{x}_0 \mid \mathbf{y}_0) = \int p(\mathbf{x}_T \mid \mathbf{y}_0) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y}_0)\, \mathrm{d}\mathbf{x}_{1:T}. \tag{5}$$

where x_0 is the data distribution and y_0 is a measurement or LR. From above equation, we need to incorporate numerical error correction term from Theorem 3.2.1 into the above equation. For this purpose, we provide an additional LR image as a condition for inference to the existing conditional diffusion model. To achieve this, we extend the ResShift formula with the explicit guided diffusion approach from Dhariwal & Nichol (2021). In Eq. equation 5, x_0 and y_0 represent the HR and LR images, respectively.

$$p(\mathbf{x}_0 \mid \mathbf{y}_0, \mathbf{c}) = \int p(\mathbf{x}_T \mid \mathbf{y}_0, \mathbf{c}) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y}_0, \mathbf{c})\, \mathrm{d}\mathbf{x}_{1:T}, \tag{6}$$

where $\mathbf{c}$ is an additional LR condition.

Proposition 3.3.1 (Guidance for LR-based Conditioning) The target distribution $p(\mathbf{x}_0 \mid \mathbf{y}_0, \mathbf{c})$ with LR guidance is given by

$$\mathcal{N}\Big(\boldsymbol{\mu} \;+\; \Sigma\, \nabla p_\theta(\mathbf{c} \mid \mathbf{x}_t, \mathbf{y}_0, t),\; \Sigma\Big), \tag{7}$$

where $\mathbf{c}$ indicates the additional image guidance.

Remark 3.3.1 (Implementation and Partial Guidance). From Theorem 3.2.1 and Proposition 3.3.1, one obtains:

$$\nabla_{\mathbf{x}_t} \log p\big(\mathbf{c} \mid \mathbf{x}_t, \mathbf{y}_0\big) = \nabla D_{\mathrm{KL}}\big(p_\theta(\mathbf{x}_t) \,\|\, p_\theta(\mathbf{c})\big), \tag{8}$$

for an image-based guidance approach. This remark connects numerical theory to practice: numerical correction and classifier guidance have the equivalence.

Discussion. With Remark 3.3.1, we derive a theoretical basis for using an input image as an additional condition alongside the conventional LR input. In practical scenarios, the conventional condition is passed through a noise-added pipeline as part of the diffusion process, while the newly introduced condition is incorporated through a noise-free path. This dual-conditioning strategy allows us to correct the discretization error accumulated during few-step sampling. Please note that the numerical error correction term is equivalent to our image-based classifier guidance term, up to a constant. This demonstrates that numerical error correction can be realized through image-based classifier guidance, thereby establishing a plug-and-play model. More detailed description about practical usage will be introduced in the next section.

3.4 IMAGE CONTROL VIA IMAGE-BASED GUIDANCE

Based on previous formulations, the second-order numerical error can be represented by the gradient of the KL divergence between the SR and HR distributions. This implies that the second-order numerical error can be linearly regressed using the gradient of the KL divergence, thereby making numerical error restoration theoretically feasible. Simultaneously, from Remark 3.3.1, we defined a method how to supply additional image to the diffusion pipeline as the classifier condition where it is also the gradient of the KL divergence of latant vector and classifier condition. Therefore, if we want to restore numerical error of the posterior model with HR image or LR image, we just need to supply HR image or LR image as classifier condition.

Reference Region The following procedure demonstrates how to estimate the constants A and B by measuring the difference in the latent space between SR and HR within the diffusion model and regressing it against the divergence between the two distributions.

$$\Delta = \mu_t(\text{HR}) - \mu_t(\text{SR}) \tag{9}$$

$$A, B = \arg\min_{a\in\mathbb{R},\, b\in\mathbb{R}} \|\Delta - (a \cdot \nabla_x D_{\text{KL}}\left(\text{SR} \,\|\, \text{HR}\right) + b)\|_2^2 \tag{10}$$

$$\mu_{\text{new}}(\text{SR}) = \mu_{\text{old}}(\text{SR}) + B + A \cdot \nabla_x D_{\text{KL}}\left(\text{SR} \,\|\, \text{HR}\right) \tag{11}$$

where $A, B \in \mathbb{R}$. $\nabla_x D_{\text{KL}}\left(\text{HR} \,\|\, \text{SR}_t\right)$ follows the trajectory of the surface between SR and HR.Chen et al. (2024b) Therefore, it can restore truncation error in Theorem 3.2.1. This will be validated in Figure 1, 9, 10.

Non-Reference Region In real-world blind super-resolution, it is not feasible to access HR references, and thus the difference between HR and SR (Eq. 9) as well as the corresponding trajectory (Eq. 10) are unavailable. To address this, we propose to estimate the parameters A and B from regions where LR–HR pairs exist, and then apply them to LR-only regions. Since Eq. 10 does not exist in such settings, we instead utilize $\nabla_x D_{\text{KL}}(\text{SR}|\text{LR})$. While Eq. 10 corresponds to the gradient at each time step along the SR–HR trajectory, if the LR–SR trajectory resided in the extrapolated space of the SR–HR trajectory, one could substitute $D_{\text{KL}}(\text{SR}|\text{HR})$ with $D_{\text{KL}}(\text{SR}|\text{LR})$ and compute A and B to construct the correction term. Our validation experiments in Fig. 5 confirmed that they are well aligned in the same trajectory, particularly in the pretrained model of Yue et al. (2023). Instead, we propose a scheme that approximates numerical errors using $D_{\text{KL}}(\text{SR}|\text{LR})$. This scheme is summarized in Algorithm 1 and Algorithm 2.

The process of determining parameters A and B can be regarded as a calibration step, where multiple LR–HR image pairs are used to identify numerical errors and extract the corresponding parameters, which are then provided during inference. Linear regression over multiple images failed to yield optimized parameters for both A and B. Consequently, we abandon the estimation of B and instead focus solely on finding A, which minimizes numerical errors across multiple images and is supplied during inference. More details about how to derive algorithm 1 and algorithm 2 will be provided in Fig. 4 and in section 6.

4 EXPERIMENT

Testing Environment To validate our theoretical framework, we conducted experiments using the DIV2K validation dataset Agustsson & Timofte (2017) , the RealSR dataset Cai et al. (2019b), and Flickr30k Dataset Young et al. (2014). Since datasets contain images of varying sizes, we cropped 128×128 center region and further divided them into four 128×128 patches. We used 100 images for DIV2K and RealSR for generation of quantitative table. Calibration of Guidance scale is performed over Imagenet Dataset Deng et al. (2009) which is used in training the pretrained model of Yue et al. (2023). We conducted all experiments on Nvidia H100 machine.

Metrics for comparison For a comprehensive image quality assessment, we employ both full-reference and no-reference metrics. Reference-based fidelity measures: PSNR and SSIM Wang et al. (2004) / Reference-based perceptual quality measures: LPIPS Zhang et al. (2018) and DISTS Ding et al. (2020) / No-reference image quality measures: NIQE Zhang et al. (2015), MANIQA Yang et al. (2022), and CLIPIQA Wang et al. (2023b) / Image distribution-based metric: FID Heusel et al. (2017), which evaluates the distance between restored images and ground truth.

Algorithm 1 Calibrate Guidance Scale

Require: A set of LR–HR pairs $(\mathbf{LR}_i, \mathbf{HR}_i)$ model parameters θ, candidate scale set $\{\alpha_t\}$
Ensure: A chosen guidance scale α_t^*.

for $i = 1$ to n **do**
  $\mathcal{Z}_T^{(\mathrm{HR})} \leftarrow \mathrm{ForwardODE}(\mathbf{HR}_i)$
  $x_T \leftarrow \mathbf{LR}_i$
  for $t = T$ down to 1 **do**
    $(\mu, \Sigma) \leftarrow \mu_\theta(x_t),\ \Sigma_\theta(x_t)$
    $\Delta = \mu_\theta(\mathcal{Z}_T^{(\mathrm{HR})}) - \mu$
    $G_{\mathrm{LR}} = \nabla_{x_t} D_{\mathrm{KL}}\big(p_\theta(x_t) \parallel \mathbf{LR}_i\big)$
    $L_i(\alpha_t) = \big\|\Delta\big\| - \big\|\alpha_t \cdot G_{\mathrm{LR}}\big\|$
  end for
end for
$\alpha_t^* = \underset{\alpha_t}{\mathrm{argmin}} \sum_{i=1}^{n} L_i(\alpha_t).$
return α_t^*

Algorithm 2 Inference with LR Guidance

Require: LR image (or condition) c, guidance scale α_t^*, diffusion model parameters θ,
Ensure: Reconstructed SR image x_0.

$x_T \leftarrow \mathbf{LR}_i$
for $t = T$ down to 1 **do**
  $(\mu, \Sigma) \leftarrow \mu_\theta(x_t),\ \Sigma_\theta(x_t)$
  $x_{t-1} \leftarrow$
  $\mathcal{N}\Big(\mu + \alpha_t^* \Sigma \nabla_{x_t} D_{\mathrm{KL}}\big(p_\theta(x_t) \parallel \mathbf{LR}_i\big)$
end for
return x_0

Figure 1: **Reference Region**. Ours (s4) and Ours (s15) represent ResShift (s4 and s15) with the proposed correction module in Eq. 11. In the region with reference overlap, the KL divergence gradient between HR and SR is used to restore the numerical error. This demonstrates that the proposed method effectively corrects the error in the overlapped region. After numerical error correction, architectural features such as the text on travel signboards were successfully restored. In particular, the s15 model achieved restoration results nearly indistinguishable from the HR image because its truncation error is already lower than that of the s4 model.

Quantitative Comparison with State-of-the-art methods To validate the effectiveness of our proposed method, we conducted quantitative and qualitative comparisons against various state-of-the-art (SOTA) methods, including: SwinIR Liang et al. (2021b), Real-ESRGAN Wang et al. (2021), StableSR Wang et al. (2023c), ResShift Yue et al. (2023), PASD Yang et al. (2024), DiffBIR Lin et al. (2024), SinSR Wang et al. (2024b), OseDiff Wu et al. (2024a), InvSR Yue et al. (2025). Our comparative analysis includes regression-based, GAN-based, and diffusion-based super-resolution methods. In Table 2, the best and second-best results are highlighted in red and blue, respectively, while regression- and GAN-based methods were excluded from this ranking. To the best of our knowledge, ResShift Yue et al. (2023) is currently the only posterior-based super-resolutin model. Thereby, we adopt ResShift Yue et al. (2023) as our backbone framework. Other methods based on pretrained diffusion prior such as Rombach et al. (2022b) are PASD Yang et al. (2024), StableSR Wang et al. (2023c), and DiffBIR Lin et al. (2024). They incorporate additional neural networks into the diffusion prior model. This could lead to overlapping effects between our theoretical enhancement and their inherent framework improvements. From Table 2, our algorithm achieves SOTA performance in non-reference metrics such as MANIQA, CLIPIQA, and FID, as well as in perception-related reference metrics such as LPIPS while minimizing degradation in reference based metrics. This demonstrates that our method effectively restores the perceptual quality of posterior models. Our method is implemented as a plug-in module integrated into the ResShift-4s or ResShift-15s framework. As such, its performance fundamentally converges to the quality of the base model. Due to discretization error, the image quality of ResShift-4s is inherently lower than that of the ResShift-15s model and can be identified in Figure 2. In the HR ground truth, all wires

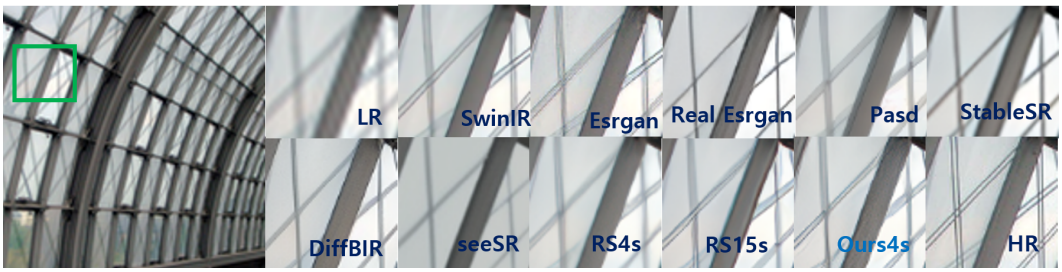


Figure 2: **Non-Reference Region** Qualitative comparisons in the non-reference region. Ours shows perceptual quality enhancement in the non-overlapped region with the reference according to the Remark 3.3.1

Table 2: Quantitative comparison with contemporary state-of-the-art image super resolution algorithms for non-overlapped region. 's' denotes the number of diffusion steps. The best and second-best results of each metric are highlighted in red and blue, respectively.

Datasets	Methods	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	NIQE↓	MANIQA↑	CLIPIQA↑	FID↓
RealSR	SwinIR	26.401153	0.822698	0.198882	0.212861	5.533595	0.482269	0.520561	34.319923
	Realesrgan	24.824283	0.690881	0.228614	0.211228	5.554380	0.515528	0.602108	68.559638
	StableSR(s200)	23.141011	0.700490	0.251224	0.218401	5.993226	0.445875	0.619377	66.324215
	DiffBIR(s50)	24.044400	0.680713	0.274067	0.221713	5.538441	0.618653	0.707664	61.714618
	PASD(s20)	24.608645	0.712806	0.242913	0.209351	4.795048	0.540107	0.625039	52.826575
	Resshift(s15)	24.932412	0.734073	0.198175	0.187559	5.319726	0.557401	0.677602	46.756755
	Resshift(s4)	25.262886	0.729898	0.207532	0.193287	5.930618	0.481287	0.646861	47.521572
	SinSR(s1)	24.814703	0.732119	0.264237	0.204537	5.189266	0.504561	0.639777	49.655490
	Osediff(s1)	23.221523	0.658442	0.279372	0.221037	5.464938	0.477634	0.661837	69.084736
	InvSR(s1)	23.163527	0.660025	0.274930	0.226382	4.827873	0.588372	0.704826	45.806334
	Ours(s4, HR)	27.563256	0.859821	0.169382	0.232162	5.158372	0.553831	0.651693	35.052837
	Ours(s4, LR)	24.537216	0.730573	0.201882	0.188932	5.366422	0.649319	0.757744	44.208402
DIV2k	SwinIR	25.524262	0.756364	0.196358	0.253459	6.574358	0.236129	0.369175	66.022553
	Realesrgan	25.017905	0.729274	0.208930	0.231028	4.377671	0.323017	0.440141	64.281207
	StableSR(s200)	25.089883	0.723934	0.230532	0.241219	6.143113	0.334181	0.402819	63.208495
	DiffBIR(s50)	24.952325	0.664533	0.245898	0.222816	5.170970	0.519603	0.694852	64.320329
	PASD(s20)	25.268907	0.730030	0.239548	0.222048	5.840237	0.409314	0.503350	66.624089
	Resshift(s15)	25.195770	0.682318	0.236521	0.230849	6.443043	0.367619	0.565044	69.377387
	Resshift(s4)	25.195317	0.682465	0.210933	0.236403	6.325634	0.324337	0.526237	68.193004
	SinSR(s1)	25.323979	0.700209	0.226813	0.238258	6.067024	0.359430	0.557111	87.956015
	Osediff(s1)	24.794832	0.714938	0.2194821	0.213805	5.859347	0.427836	0.689721	67.801938
	InvSR(s1)	24.143282	0.670584	0.2415381	0.253827	5.981039	0.471184	0.691187	74.120972
	Ours(s4,LR)	24.910596	0.682398	0.215568	0.235888	6.024669	0.475017	0.707421	68.141436

are structured as dual wires, but among the SR results, only SwinIR, ResShift-15s, and our method successfully reconstruct the dual-wire structure. Naturally, a 15-step model is expected to suffer less from discretization error than a 4-step model (as long as the model is properly trained). Nonetheless, despite operating with only four steps, our method succeeds in reconstructing the dual-wire structure by effectively restoring its discretization error. Our framework is the first to provide a controllable knob for navigating this trade-off without retraining.

Ablation Study, Qualitative Results is available in section 6 and 8 in the supplemental material.

5 Conclusion

In this work, we introduced a plug-and-play module for posterior learning-based super-resolution models. From a theoretical perspective, we derived that discretization errors inherent in first-order ODE formulations can be corrected by incorporating second-order differential terms. We further showed that this correction can be formulated within the conditional diffusion framework, and that the resulting expression naturally coincides with the structure of image-based classifier guidance. This provides a new theoretical understanding of posterior diffusion trajectories.

From a practical perspective, we demonstrated that applying the correction term can reduce numerical errors in pretrained models without retraining. While fidelity enhancement through this correction requires HR supervision and is thus limited in blind SR scenarios, we proposed instead to exploit LR-based classifier guidance in a sign-flipped manner to enhance perceptual quality. This design enables posterior SR models to achieve perceptual results competitive with state-of-the-art methods, thereby bridging theoretical insight with practical applicability.

REFERENCES

Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPRW*, July 2017.

Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. In *CVPR*, 2021.

Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *ICCV*, pp. 3086–3095, 2019a.

Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *ICCV*, pp. 3086–3095, 2019b.

Bin Chen, Gehui Li, Rongyuan Wu, Xindong Zhang, Jie Chen, Jian Zhang, and Lei Zhang. Adversarial diffusion compression (adc) for real-world isr. In *CVPR*, 2025.

Defang Chen, Zhenyu Zhou, Jian-Ping Mei, Chunhua Shen, Chun Chen, and Can Wang. A geometric perspective on diffusion models. In *ICLR*, May 2024a.

Defang Chen, Zhenyu Zhou, Can Wang, Chunhua Shen, and Siwei Lyu. On the trajectory regularity of ode-based diffusion sampling. In *ICML*, July 2024b.

Junyang Chen, Jinshan Pan, and Jiangxin Dong. Faithdiff: Faithful and controllable image super-resolution with diffusion models. 2024c.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, pp. 248–255, 2009.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *NeurIPS*, volume 34, pp. 8780–8794, 2021.

Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Image quality assessment: Unifying structure and texture similarity. In *IEEE TPAMI*, volume 44, pp. 2567–2581, 2020.

Linwei Dong, Qingnan Fan, Yihong Guo, Zhonghao Wang, Qi Zhang, Jinwei Chen, Yawei Luo, and Changqing Zou. Tsd-sr: One-step diffusion with target score distillation for real-world image super-resolution. 2024.

Weisheng Dong, Lei Zhang, Guangming Shi, and Xiaolin Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE TIP*, 20(7):1838–1857, 2011.

Patrick Esser, Robin Rombach, and Björn Ommer. Taming transformers for high-resolution image synthesis. In *CVPR*, pp. 12873–12883, 2021.

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. In *NeurIPS*, volume 2, pp. 2672–2680, 2014.

Shuhang Gu, Wangmeng Zuo, Qi Xie, Deyu Meng, Xiangchu Feng, and Lei Zhang. Convolutional sparse coding for image super-resolution. In *ICCV*, pp. 1823–1831, 2015.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NeurIPS*, 2017.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, volume 33, pp. 6840–6851, 2020.

M. Hochbruck and A. Ostermann. Explicit exponential runge-kutta methods for semilinear parabolic problems. In *SIAM Journal on Numerical Analysis*, volume 43, pp. 1069–1090, 2005.

et al. Joram Soch. Statproofbook/statproofbook.github.io: The book of statistical proofs. 2024.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Proc. NeurIPS*, 2022.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *NeurIPS*, 2022.

Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 2017.

Junyong Lee, Myeonghee Lee, Sunghyun Cho, and Seungyong Lee. Reference-based video super-resolution using multi-camera video triplets. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Haoying Li, Yifan Yang, Meng Chang, Shiqi Chen, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. Srdiff: Single image super-resolution with diffusion probabilistic models. *Neurocomputing*, 479:47–59, 2022a.

Haoying Li, Yifan Yang, Meng Chang, Shiqi Chen, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. Srdiff: Single image super-resolution with diffusion probabilistic models. *Neurocomputing*, 479:47–59, 2022b.

Haoying Li, Yifan Yang, Meng Chang, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. Srdiff: Single image super-resolution with diffusion probabilistic models. volume 479, pp. 47–59, 2022c.

Mingxiao Li, Tingyu Qu, Ruicong Yao, Wei Sun, and Marie-Francine Moens. Alleviating exposure bias in diffusion models through sampling with shifted time steps. In *ICLR*, May 2024.

Yangming Li and Mihaela van der Schaar. On error propagation of diffusion models. In *ICLR*, May 2024.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV-W*, pp. 1833–1844, 2021a.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV-W*, pp. 1833–1844, 2021b.

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPR*, 2017.

Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, and Chao Dong. Diffbir: Towards blind image restoration with generative diffusion prior, 2024.

Jiawei Liu, Qiang Wang, Huijie Fan, Yinong Wang, Yandong Tang, and Liangqiong Qu. Residual denoising diffusion models. In *CVPR*, 2024.

Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo numerical methods for diffusion models on manifolds. In *ICLR*, 2022.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 5 steps. In *NeurIPS*, 2022.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. In *ICLR*, 2023.

Nancy Mehta, Akshay Dudhane, Subrahmanyam Murala, Syed Waqas Zamir, Salman Khan, and Fahad Shahbaz Khan. Gated multi-resolution transfer network for burst restoration and enhancement. In *CVPR*, 2023.

E. Kloeden Peter and Platen Eckhard. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *CVPR*, 2021.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pp. 10684–10695, 2022a.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pp. 10684–10695, 2022b.

Chitwan Saharia, Jonathan Ho, William Chan, David J. Fleet, Mohammad Norouzi, and Tim Salimans. Image super-resolution via iterative refinement. In *CVPR*, 2021.

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

Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE TPAMI*, 2022b.

Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE TPAMI*, 2022c.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021a.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021b.

Gilbert Strang. *Computational Science and Engineering*. Wellesley-Cambridge Press, 2007.

Hang Wang, Xuanhong Chen, Bingbing Ni, Yutian Liu, and Jinfan Liu. Omni aggregation networks for lightweight image super-resolution. In *CVPR*, 2023a.

Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023b.

Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *arXiv preprint arXiv:2305.07015*, 2023c.

Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *ECCV-W*, 2018.

Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *ICCV-W*, pp. 1905–1914, 2021.

Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C. Kot, and Bihan Wen. Sinsr: Diffusion-based image super-resolution in a single step. In *CVPR*, 2024a.

Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C Kot, and Bihan Wen. Sinsr: diffusion-based image super-resolution in a single step. In *CVPR*, pp. 25796–25805, 2024b.

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

Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. 2024a.

Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *CVPR*, pp. 25456–25467, 2024b.

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *CVPR*, pp. 1191–1200, 2022.

Tao Yang, Rongyuan Wu, Peiran Ren, Xuansong Xie, and Lei Zhang. Pixel-aware stable diffusion for realistic image super-resolution and personalized stylization. In *ECCV*, 2024.

Peter Young, Alice Lai, Micah Hodosh, and Julia Hockenmaier. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. In *TACL*, volume 2, pp. 67–78, 2014.

Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. In *NeurIPS*, 2023.

Zongsheng Yue, Kang Liao, and Chen Change Loy. Arbitrary-steps image super-resolution via diffusion inversion. In *CVPR*, 2025.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022.

Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *ICCV*, pp. 4791–4800, 2021.

Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. In *IEEE TIP*, volume 24, pp. 2579–2591, 2015.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pp. 586–595, 2018.

Kaiwen Zheng, Cheng Lu, Jianfei Chen, and Jun Zhu. Dpm-solver-v3: Improved diffusion ode solver with empirical model statistics. In *NeurIPS*, 2023.

CONTENTS

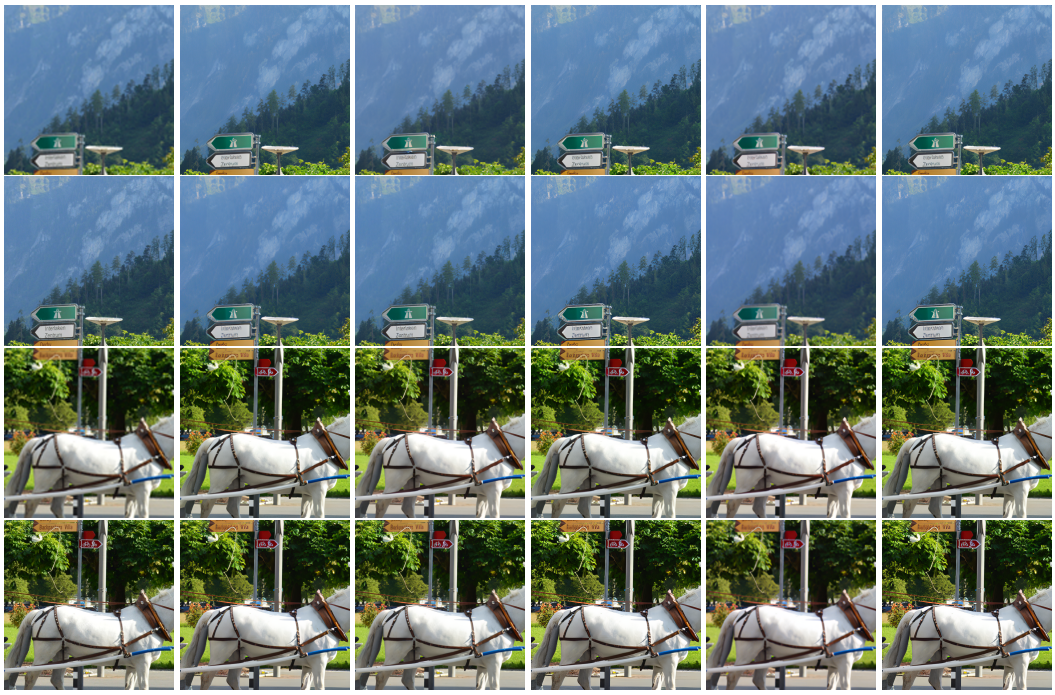
6 NUMERICAL ERROR RESTORATION: REFERENCE-BASED AND REFERENCE-FREE APPROACHES

6.1 RESTORATION IN REFERENCE REGION AND NON-REFERENCE REGION

We first estimate correction parameters A and B in reference regions by linear regression of numerical errors and apply them via classifier guidance. As shown in Table 2, fidelity enhancement (e.g., Ours(s=4, HR)) requires guidance forms such as option 1 or 5 in Fig. 4. However, in non-reference regions, only options 3/7 (LR-only regression) can be used, which yield overly smoothed results. Fig. 5 shows that cosine similarity between LR–SR and SR–HR trajectories remains below 0.8 in most regions, confirming misalignment. Thus, fidelity-oriented correction is infeasible in non-reference regions. Exploring pretrained models with better trajectory alignment is left as future work.

Fidelity vs. Perception Trade-off Numerical correction improves reference-based metrics (PSNR, SSIM, LPIPS) but not perceptual quality. To address this, we flip the sign of the correction term. As seen in Fig. 4, cases 2/4/6/8 do not increase fidelity but yield sharper perceptual appearance, even for LR-only guidance (cases 6/8). In contrast, direct fidelity correction (cases 1/5) recovers small details (e.g., text) but increases blur. We therefore propose to use sign-flipped image-based classifier guidance as a practical plug-in for perceptual enhancement in non-reference regions.

Calibration Strategy In practice, estimating both A and B is unnecessary; A alone suffices. Fig. **??** shows error curves when sweeping A from -300 to 300 in an $AX + B$ correction form (with $B = 0$).



LR	SR(4s)	1.$A\nabla D_{KL}(\mathrm{SR}(4s) \mid \mathrm{HR}) + B$	2.$-A\nabla D_{KL}(\mathrm{SR}(4s) \mid \mathrm{HR}) - B$	3.$A\nabla D_{KL}(\mathrm{SR}(4s) \mid \mathrm{LR}) + B$	4.$-A\nabla D_{KL}(\mathrm{SR}(4s) \mid \mathrm{LR})$
HR	SR(15s)	5.$A\nabla D_{KL}(\mathrm{SR}(15s) \mid \mathrm{HR}) + B$	6.$-A\nabla D_{KL}(\mathrm{SR}(15s) \mid \mathrm{HR}) - B$	7.$A\nabla D_{KL}(\mathrm{SR}(15s) \mid \mathrm{LR}) + B$	8.$-A\nabla D_{KL}(\mathrm{SR}(15s) \mid \mathrm{LR}) - B$

Figure 4: Reference region restoration result by various guidance schemes. The guidance schemes applied to the images above are summarized in the table below. Detailed description is available in section 6.1

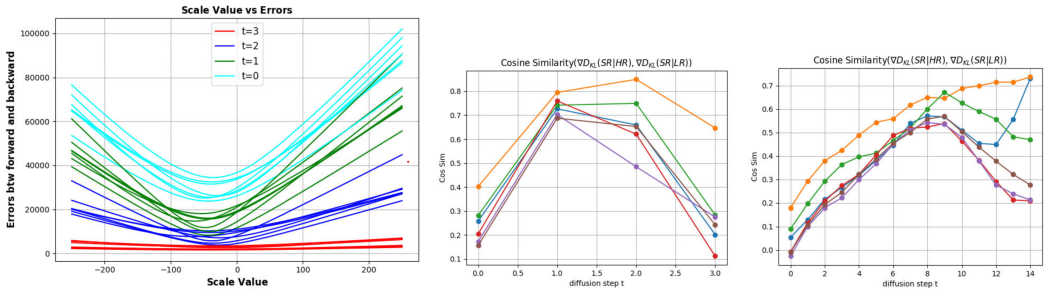


Figure 5: (a) The range of scale values used to compensate for the error in the forward and backward processes varies slightly depending on the camera type. The error was measured using ResShift Yue et al. (2023) with step=4. As the step point progresses from $t = 3$ to $t = 0$, the error accumulates. Each line corresponds to the error vector from an image. (b) cosine similarity between the trajectory between $\nabla D_{KL}(SR|HR)$ and $\nabla D_{KL}(SR|LR)$ for step 4 and (c) for step 15 pretrained models of Yue et al. (2023)

Most images exhibit a clear error minimum, which can be transferred from reference to non-reference regions. Fig. 7 further illustrates that perceptual quality improves with calibrated A until an optimal value, beyond which artifacts increase. Accurate guidance scale calibration is therefore crucial.

Practical Implications Correction via $\nabla D_{KL}(\mathrm{SR}\|\mathrm{LR})$ quantitatively improves fidelity but degrades perceptual plausibility (Fig. 4). For practical deployment, our plug-in sacrifices fidelity in favor of perceptual enhancement, as confirmed by improved MTF (Fig. 2) and perceptual scores (Table 2, Figs. 11, 12). Reference-based SR systems (e.g., Lee et al. (2022)) suggest that overlap–nonoverlap calibration is possible, but real smartphone images suffer from heavy postprocessing. Developing a raw-image pipeline to fully validate this strategy is left as future work.

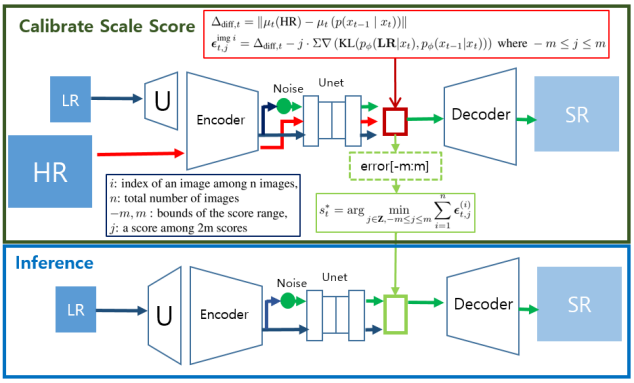


Figure 6: Block diagram for inferring the non-overlapped region using optimized parameters extracted from the overlapped region. A score S^* means the scale score calibrated in Algorithm 1 stage. Calibrated scale score is delivered to Inference pipeline which is described in Algorithm 2.

Table 2: Computation costs of existing methods to process 1000 tiles(64x64) with from DIV2K dataset.

	SwinIR	Esrgan	Realesrgan	StableSR	DiffBir	SeeSR	PASD	ResShift(s15)	ResShift(s4)	SinSR	OseDiff	InvSR	Ours(s4)
Time(s)	191	23	12	148	1948	994	3105	808	229	2877	181	115	235

6.2 Performance Comparison

Table 2 reaffirms the value of our approach. Although ResShift s4 version has more numerical errors than s15 version and makes it less favorable in image quality metrics, it has the advantages in the limited runtime environment. However, after applying our method, the image quality metrics become comparable to those of s15 version, while the speed remains at the s4 level. Therefore, our approach serves as an option for maintaing optimal image quality while satisfying limited runtime constraints.

6.3 Ablation Study

In Figure 8, we present the error plots for the s4 and s15 version. The overall error reduction is highly sensitive to the choice of optimized parameters: negative values of the parameters tend to increase cumulative error, while appropriately optimized parameters significantly reduce it.

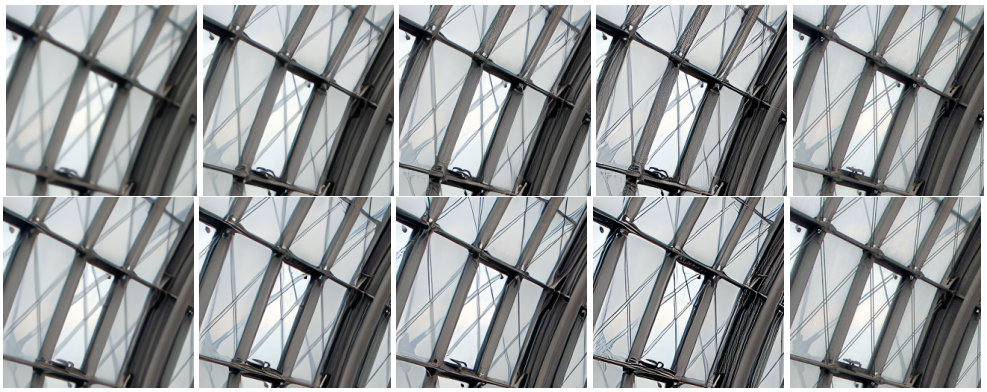


Figure 7: Non-overlap region. Image quality comparison for various scale value. The images from the left to the right have scale values such as (A^*), $(-0.5 \cdot A^*)$, $(-A^*)$,$(-2 \cdot A^*)$,HR. The first row uses step 4 and the second row uses step 15

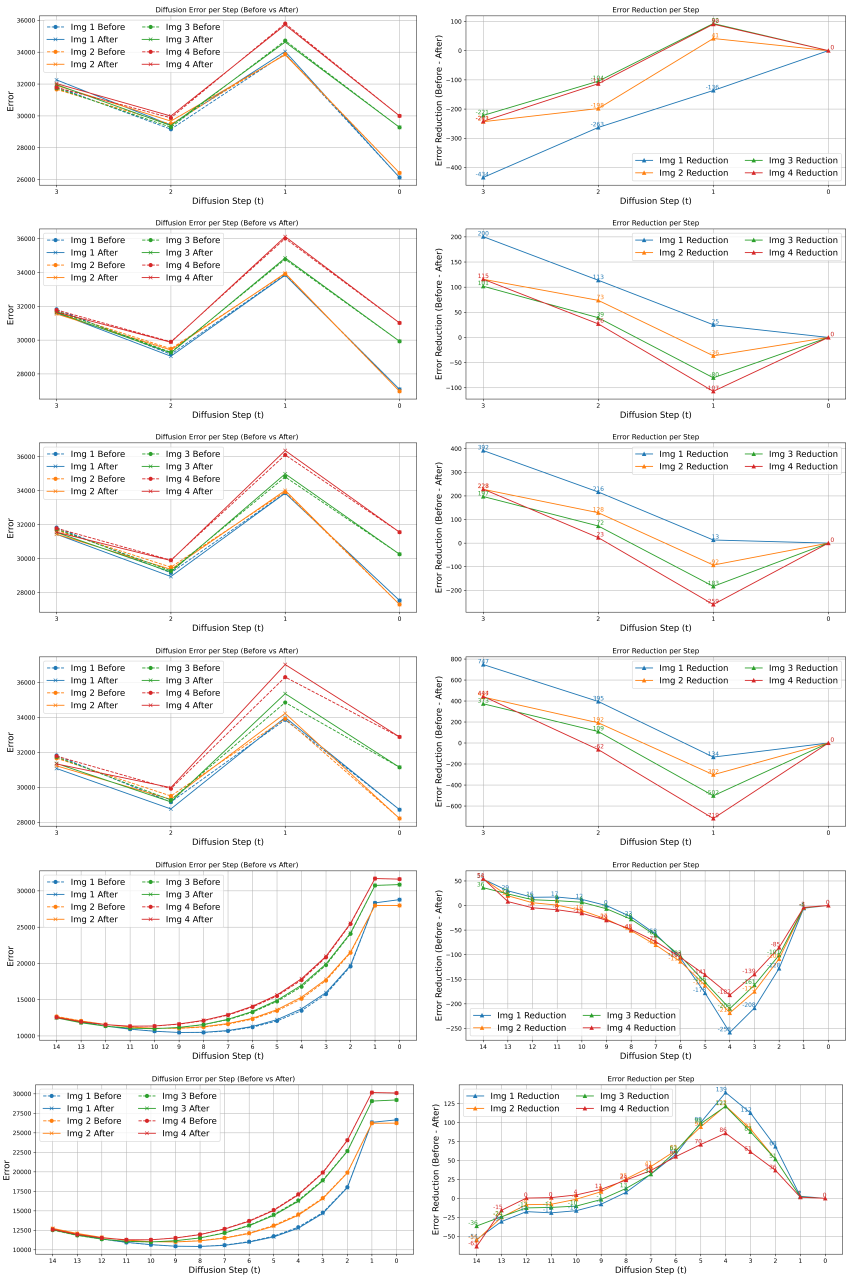


Figure 8: Non-overlap region. Cumulative Error at each diffusion step. Left Column denotes $\mathbf{E}_t^{\text{cumu}}$ and Right Column denotes $\mathbf{E}_t^{\text{cumu}}$(before)-$\mathbf{E}_t^{\text{cumu}}$(after). According to the row sequence: (1)$-A^*$ (2) $0.5 \cdot A^*$ (3) $1 \cdot A^*$ (4) $2 \cdot A^*$ (5)$-A^*$ (6) A^*

7 MATHEMATICAL DETAILS

Lemma 3.2.1 (KL gradient in the Gaussian case) For two normal distributions $\mathcal{P} : \mathcal{N}(\mathbf{x}, \sigma_1^2)$, $\mathcal{Q} : \mathcal{N}(\boldsymbol{\mu}_t, \sigma_2^2)$

$$\nabla D_{\mathrm{KL}}\big(\mathcal{P}||\mathcal{Q})\big) \approx \frac{(\mathbf{x} - \boldsymbol{\mu}_t)}{\sigma_2^2}.$$

Proof. From Joram Soch (2024), Kullback-Leibler divergence for two normal distributions $\mathcal{P}$ and $\mathcal{Q}$ is as follows:

$$\mathrm{KL}[\mathcal{P}||\mathcal{Q}] = \frac{1}{2}\left(\log\frac{\sigma_2^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_2^2} + \frac{(\mathbf{x} - \boldsymbol{\mu}_2)^2}{2\sigma_2^2}\right)$$

We assume same variance in two distributions. For two distributions $D_{KL}(\mathcal{N}(\mathbf{x}, \sigma_1^2), \mathcal{N}(\boldsymbol{\mu}_t, \sigma_2^2))$, we get their KL divergence as follows.

$$\nabla_{\mathbf{x}}\mathrm{KL} = \nabla_{\mathbf{x}}\left[\mathrm{KL}[\mathcal{P}||\mathcal{Q}]\right] = \frac{\mathbf{x} - \boldsymbol{\mu}_t}{\sigma_2^2}.$$

Definition 3.2.1 (Second-order derivative as the numerical error) We define the *numerical error* (DE) of the first-order derivative solution as

$$\mathbf{x}_t^{\mathrm{Exact}} \;-\; \mathbf{x}_t^{\mathrm{Euler(1st)}} \;\approx\; \nabla_t\big[f(t)\,\mathbf{x}_t - \tfrac{1}{2}g^2(t)\,\nabla_x \log p_t(\mathbf{x}_t)\big].$$

Proof of Definition 3.2.1 From Eq. 1, discretizing with a first-order Euler step of size Δt:

$$\mathbf{x}_{t+\Delta t}^{(\mathrm{Euler})} = \mathbf{x}_t + \Delta t\,\big[f(t)\,\mathbf{x}_t - \tfrac{1}{2}g^2(t)\,\nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t)\big].$$

Comparing to the exact solution

$$\mathbf{x}_{t+\Delta t}^{(\mathrm{Exact})} = \mathbf{x}_t + \int_t^{t+\Delta t} \big[f(s)\,\mathbf{x}_s \;-\; \tfrac{1}{2}g^2(s)\,\nabla_{\mathbf{x}} \log p_s(\mathbf{x}_s)\big]\,ds,$$

Taylor expansion of the above equation yields

$$\begin{aligned}&\mathbf{x}(t) \approx \mathbf{x}_s + (t-s)V + \frac{1}{2}(t-s)^2\nabla_t V + \mathcal{O}((t-s)^3),\\ &\text{where } V = f(t)\mathbf{x}_t - \tfrac{1}{2}g^2(t)\nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t).\end{aligned} \tag{12}$$

A second-order (i.e., $\mathcal{O}(\Delta t^2)$) discretization error term appears upon Taylor expansion.

$$\begin{aligned}\mathbf{x}_{t+\Delta t}^{(\mathrm{Exact})} &= \mathbf{x}_t + \Delta t V + \frac{1}{2}(\Delta t)^2\nabla_t V + \mathcal{O}((\Delta t)^3)\\ \mathbf{x}_{t+\Delta t}^{(\mathrm{Euler})} &= \mathbf{x}_t + \Delta t V + \mathcal{O}((\Delta t)^2)\end{aligned} \tag{13}$$

Subtraction of two above equations yields

$$\mathrm{DE} = \mathbf{x}_{t+\Delta t}^{(\mathrm{Exact})} - \mathbf{x}_{t+\Delta t}^{(\mathrm{Euler})} \;\approx\; \tfrac{1}{2}\,(\Delta t)^2\,\nabla_t V \;\approx\; \nabla_t\big[f(t)\mathbf{x}_t - \tfrac{1}{2}g^2(t)\nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t)\big].$$

Proposition 3.2.1 (Constant Variance Leads to a Simpler Form). Let σ^2 be a predefined constant over the spatial dimensions. Then, the gradient with respect to t of the modified score function from Definition 3.2.1 satisfies

$$\nabla_t\big[f(t)\,\mathbf{x}_t - \tfrac{1}{2}g^2(t)\,\nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t)\big] \approx A \cdot \frac{(\mathbf{x}_t - \boldsymbol{\mu}_t)}{\sigma^2} + B.$$

where $A, B \in \mathbb{R}$ is the optimization constant to minimize discretization error.

Proof We apply the total derivative:

$$\frac{d}{dt}\left[f(t)\mathbf{x}(t) - \tfrac{1}{2}g^2(t)\nabla_{\mathbf{x}}\log p_t(\mathbf{x})\right] = \frac{d}{dt}\left[f(t)\mathbf{x}(t)\right] - \frac{d}{dt}\left[\tfrac{1}{2}g^2(t)\nabla_{\mathbf{x}}\log p_t(\mathbf{x})\right].$$

First, since $\mathbf{x} = \mathbf{x}(t)$, the chain rule gives:

$$\frac{d}{dt}[f(t)\mathbf{x}(t)] = f'(t)\mathbf{x} + f(t)\frac{d\mathbf{x}}{dt} = f'(t)\mathbf{x} + f(t)\left(f(t)\mathbf{x} - \tfrac{1}{2}g^2(t)\nabla_{\mathbf{x}}\log p_t(\mathbf{x})\right).$$

Now compute the second term:

$$\frac{d}{dt}[g^2(t)\nabla_{\mathbf{x}}\log p_t(\mathbf{x})] = 2g(t)g'(t)\nabla_{\mathbf{x}}\log p_t(\mathbf{x}) + g^2(t)\frac{\partial}{\partial t}\left(\nabla_{\mathbf{x}}\log p_t(\mathbf{x})\right).$$

We now assume $p_t(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_t, \sigma_t^2 I)$, so that

$$\nabla_{\mathbf{x}}\log p_t(\mathbf{x}) = -\frac{1}{\sigma_t^2}(\mathbf{x} - \boldsymbol{\mu}_t),$$

and hence:

$$\frac{\partial}{\partial t}\nabla_{\mathbf{x}}\log p_t(\mathbf{x}) = -\left[\frac{d}{dt}\left(\frac{1}{\sigma_t^2}\right)(\mathbf{x} - \boldsymbol{\mu}_t) - \frac{1}{\sigma_t^2}\boldsymbol{\mu}_t'\right].$$

Now substitute all terms:

$$\begin{aligned}\nabla_t\left[f(t)\mathbf{x} - g^2(t)\nabla_{\mathbf{x}}\log p_t(\mathbf{x})\right] &= f'(t)\mathbf{x} + f^2(t)\mathbf{x} - f(t)g^2(t)\nabla_{\mathbf{x}}\log p_t(\mathbf{x}) \\ &\quad - g(t)g'(t)\nabla_{\mathbf{x}}\log p_t(\mathbf{x}) - \tfrac{1}{2}g^2(t)\left[-\frac{d}{dt}\left(\frac{1}{\sigma_t^2}\right)(\mathbf{x} - \boldsymbol{\mu}_t) + \frac{1}{\sigma_t^2}\boldsymbol{\mu}_t'\right]\end{aligned}$$

Substitute the gradient expression:

$$\nabla_{\mathbf{x}}\log p_t(\mathbf{x}) = -\frac{1}{\sigma_t^2}(\mathbf{x} - \boldsymbol{\mu}_t),$$

we collect terms as:

$$\begin{aligned}\nabla_t\left[f(t)\mathbf{x} - g^2(t)\nabla_{\mathbf{x}}\log p_t(\mathbf{x})\right] &= (f'(t) + f^2(t))\mathbf{x} + \left[\frac{f(t)g^2(t) + 2g(t)g'(t)}{\sigma_t^2}\right](\mathbf{x} - \boldsymbol{\mu}_t) \\ &\quad + \left[g^2(t)\frac{d}{dt}\left(\frac{1}{\sigma_t^2}\right)(\mathbf{x} - \boldsymbol{\mu}_t)\right] \quad + \frac{g^2(t)}{\sigma_t^2}\boldsymbol{\mu}_t' \\ &= \mathsf{A}(t)\cdot\mathbf{x} + \mathsf{B}(t)\cdot(\mathbf{x} - \boldsymbol{\mu}_t) + g^2(t)\frac{d}{dt}\left(\frac{\mathbf{x}}{\sigma_t^2}\right) + \mathsf{C}(t)\end{aligned}$$

where $\mathsf{A}, \mathsf{B}, \mathsf{C}$ denote the aggregated terms consisting of corresponding constant components. By substituting $\frac{d}{dt}\left(\frac{\mathbf{x}}{\sigma_t^2}\right)$ with the following equation,

$$\frac{d}{dt}\left(\frac{\mathbf{x}}{\sigma_t^2}\right) = \frac{f(t)}{\sigma_t^2}\mathbf{x} + \frac{\mathbf{x} - \boldsymbol{\mu}_t}{\sigma_t^4} - \frac{2\cdot\mathbf{x}}{\sigma_t^2} = \mathsf{D}(t)(\mathbf{x} - \boldsymbol{\mu}_t) + \mathsf{E}(t)$$

Now, we have the decomposition:

$$\nabla_t\left[f(t)\mathbf{x} - g^2(t)\nabla_{\mathbf{x}}\log p_t(\mathbf{x})\right] = \alpha(t)(\mathbf{x} - \boldsymbol{\mu}_t) + \beta(t) \approx A\cdot(\mathbf{x} - \boldsymbol{\mu}_t) + B$$

Theorem 3.2.1 (Approximation Error and KL Gradient Scaling). For a given step t, the difference between the exact solution $\mathbf{x}^{\text{Exact}}$ and the Euler discretization $\mathbf{x}^{\text{Euler}}$ can be expressed as

$$\mathbf{x}^{\text{Exact}} - \mathbf{x}^{\text{Euler}} \approx A\cdot\nabla D_{\text{KL}}(\mathcal{P}\,\|\,\mathcal{Q}) + B.$$

where $\mathcal{P}$ is discrete distribution, $\mathcal{Q}$ is continuous exact distribution. Here, A, B can be derived exactly through linear regression between latant spaces of SR and HR .

Proof. From Definition 3.2.1 and Proposition 3.2.1, we obtain

$$\mathbf{x}^{\text{Exact}} - \mathbf{x}^{\text{Euler}} \approx \frac{\mathbf{x} - \boldsymbol{\mu}_t}{\sigma_t^2}.$$

Meanwhile, from Corollary 3.2.1, it follows that

$$\nabla_{\mathbf{x}} D_{\text{KL}}(\mathcal{P} \,\|\, \mathcal{Q}) \approx \frac{\mathbf{x} - \boldsymbol{\mu}_t}{\sigma^2}.$$

Thus, we conclude that

$$\mathbf{x}^{\text{Exact}} - \mathbf{x}^{\text{Euler}} \approx A \cdot \nabla D_{\text{KL}}(\mathcal{P} \,\|\, \mathcal{Q}) + B,$$

for an appropriate scaling factor S^*.

Proposition 3.3.1 (Guidance for LR-based Conditioning) The target distribution with LR guidance is given by

$$\mathcal{N}\Big(\boldsymbol{\mu} \;+\; \Sigma\, \nabla p_\theta(\mathbf{c} \mid \mathbf{x}_t, \mathbf{y}_0, t),\; \Sigma\Big), \tag{14}$$

where $\mathbf{c}$ indicates the LR guidance.

Proof First, we enumerates derivation process of ResShift algorithm Yue et al. (2023) here.

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_0) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_{t-1} + \alpha_t e_0, \kappa^2 \alpha_t I), \tag{15}$$

where $e_0 = \mathbf{y}_0 - \mathbf{x}_0$ is the residual between the c and HR images, and κ is a hyper-parameter controlling the noise variance.The reverse process aims to estimate the posterior distribution $p(\mathbf{x}_0|\mathbf{y}_0)$ by follows:

$$p(\mathbf{x}_0|\mathbf{y}_0) = \int p(\mathbf{x}_T|\mathbf{y}_0) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_0) \mathrm{d}\mathbf{x}_{1:T}, \tag{16}$$

The targeted distribution is as follows:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_0) = \mathcal{N}\left(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, \mathbf{y}_0, t), \Sigma(t)\right) \tag{17}$$

$$\Sigma(t) = \kappa^2 \frac{\eta_{t-1}}{\eta_t} \alpha_t I. \tag{18}$$

$$\mu_\theta(\mathbf{x}_t, \mathbf{y}_0, t) = \frac{\eta_{t-1}}{\eta_t} \mathbf{x}_t + \frac{\alpha_t}{\eta_t} f_\theta(\mathbf{x}_t, \mathbf{y}_0, t), \tag{19}$$

where $\alpha_t = \eta_t - \eta_{t-1}$ for $t > 1$ and $\alpha_1 = \eta_1$.
In the previous section, we proposed conditional diffusion process as follows:

$$p(\mathbf{x}_0|\mathbf{y}_0, \mathbf{c}) = \int p(\mathbf{x}_T|\mathbf{y}_0, \mathbf{c}) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_0, \mathbf{c}) \mathrm{d}\mathbf{x}_{1:T}, \tag{20}$$

In the above equation, $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_0, \mathbf{c})$ can be simplified.

$$\begin{aligned} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_0, c) &= \frac{p(\mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{y}_0, c)}{p(\mathbf{x}_t, \mathbf{y}_0, c)} = \frac{p(\mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{y}_0, c)}{p(\mathbf{c}|\mathbf{x}_t, \mathbf{y}_0) p(\mathbf{x}_t, \mathbf{y}_0)} \\ &= \frac{p(\mathbf{c}|\mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{y}_0) p(\mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{y}_0)}{p(\mathbf{c}|\mathbf{x}_t, \mathbf{y}_0) p(\mathbf{x}_t, \mathbf{y}_0)} = \frac{p(\mathbf{c}|\mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{y}_0) p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_0) p(\mathbf{x}_t, \mathbf{y}_0)}{p(\mathbf{c}|\mathbf{x}_t, y) p(\mathbf{x}_t, \mathbf{y}_0)} \\ &= \frac{p(\mathbf{c}|\mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{y}_0) p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_0)}{p(\mathbf{c}|\mathbf{x}_t, \mathbf{y}_0)} = \frac{p(\mathbf{c}|\mathbf{x}_{t-1}, \mathbf{y}_0) p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_0)}{p(\mathbf{c}|\mathbf{x}_t)} \end{aligned} \tag{21}$$

The $p(\mathbf{c}|\mathbf{x}_t)$ term can be treated as a constant since it does not depend on $\mathbf{x}_{t-1}$. from Markov Chain property, $p(\mathbf{y}_0|\mathbf{x}_t, \mathbf{x}_{t+1}) = p(\mathbf{y}_0|\mathbf{x}_t)$ and $p(\mathbf{c}|\mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{y}_0) = p(\mathbf{c}|\mathbf{x}_{t-1}, \mathbf{y}_0)$ is derived. Now, we conclude as follows:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_0, \mathbf{c}) = Z p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_0) p(\mathbf{c}|\mathbf{x}_{t-1}, \mathbf{y}_0) \tag{22}$$

we already knows $p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_0)$ and we need to handle $p(\mathbf{c}|\mathbf{x}_{t-1}, \mathbf{y}_0)$. For now, we write it as $p(\mathbf{c}|\mathbf{x}_t, \mathbf{y}_0)$ in convenience. $\log p(\mathbf{c}|\mathbf{x}_t, \mathbf{y}_0)$ can be approximated by using a Taylor expansion around $\mathbf{x}_t = \mu$.

$$\log p(\mathbf{c}|\mathbf{x}_t, \mathbf{y}_0) \approx \log p(\mathbf{c}|\mathbf{x}_t, \mathbf{y}_0)|_{\mathbf{x}_t=\mu} + (\mathbf{x}_t - \mu)\nabla_{\mathbf{x}_t} \log p(\mathbf{c}|\mathbf{x}_,\mathbf{y}_0)|_{\mathbf{x}_t=\mu}$$

Let $G = \nabla_{\mathbf{x}_t} \log p(\mathbf{c}|\mathbf{x}_t, \mathbf{y}_0)|_{\mathbf{x}_t=\mu}$. From Eq. equation 18 equation 19, we aleady got Σ_θ and μ.

$$\begin{aligned}\log p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_0, \mathbf{c}) &\approx -\frac{1}{2}(\mathbf{x}_t - \mu)^T\Sigma^{-1}(\mathbf{x}_t - \mu) + (\mathbf{x}_t - \mu)G + C \\ &= -\frac{1}{2}(\mathbf{x}_t - \mu - \Sigma G)^T\Sigma^{-1}(\mathbf{x}_t - \mu - \Sigma G) + C\end{aligned}$$

Throughout all the above equations, the original $\mathbf{c}$ image can be fed by the condition $\mathcal{N}(\mu + \Sigma\nabla p_\theta(\mathbf{c}|\mathbf{x}_t, \mathbf{y}_0, t), \Sigma)$.

Remark 3.3.1 (Implementation and Partial Guidance). From Theorem 3.2.1 and Proposition 3.2.2, one obtains:

$$\nabla_{\mathbf{x}_t} \log p\big(\mathbf{c} \mid \mathbf{x}_t, \mathbf{y}_0\big) = \nabla D_{\mathrm{KL}}\big(p_\theta(\mathbf{x}_t) \,\|\, p_\theta(\mathbf{c})\big),$$

for an image-based guidance approach. Here we assume $\mathbf{x}_{\mathrm{LR}}$ is the vector at $\mathbf{x}_T$ prior to the diffusion process.

Proof. According to Corollary 3.2.1, let $\mathcal{P} = \mathcal{N}(\mathbf{x}, \sigma_1^2)$ and $\mathcal{Q} = \mathcal{N}(\boldsymbol{\mu}_t, \sigma_2^2)$. Then

$$\nabla D_{\mathrm{KL}}(\mathcal{P} \,\|\, \mathcal{Q}) \;\approx\; \frac{\mathbf{x} - \boldsymbol{\mu}_t}{\sigma^2}.$$

Suppose a final distribution is $\mathcal{N}(\mathbf{x} - \boldsymbol{\mu} - \Sigma G,\ \Sigma)$. We introduce two distributions $\mathcal{P}' = \mathcal{N}(\mathbf{x} - \boldsymbol{\mu}_t, \sigma^2)$ for $p(\mathbf{x}_0 \mid \mathbf{y}_0)$, and $\mathcal{Q}' = \mathcal{N}(\Sigma G, \sigma^2)$ for $p(\mathbf{c} \mid \mathbf{y}_0)$. Their KL divergence leads to

$$\nabla D_{\mathrm{KL}}(\mathcal{P}' \,\|\, \mathcal{Q}') \;\approx\; \frac{\mathbf{x} - \boldsymbol{\mu}_t - \Sigma G}{\sigma^2}.$$

Hence,

$$\nabla_{\mathbf{x}_t} \log p\big(\mathbf{c} \mid \mathbf{x}_t, \mathbf{y}_0\big) = \nabla D_{\mathrm{KL}}\big(p_\theta(\mathbf{x}_t) \,\|\, p_\theta(\mathbf{c})\big)$$

is justified.

8 Additional Experimental Result

8.1 Additional Experiment of Numerical Error Restoration in reference region

Fig. 9 and Fig. 10 provides additional experimental proof to support the validness of numerical error restoration by Eq. 11 in reference region.

8.2 Additional Experiment of Numerical Error Restoration out of reference region

Fig. 11 , Fig. 12 provides additional experimental proof to support the validness of perceptual quality enhancement by Algorithm 1 and Algorithm 2 in non-reference region.

8.2.1 Future works

Beyond super-resolution, our framework is applicable to a broader class of inverse problems such as image deblurring, denoising, and compression artifact removal. Since posterior diffusion directly leverages LR–HR supervision, future work can readily reuse existing regression-based datasets, opening a path toward a general posterior diffusion paradigm where numerical analysis and plug-and-play conditioning jointly enable controllable restoration across diverse domains.

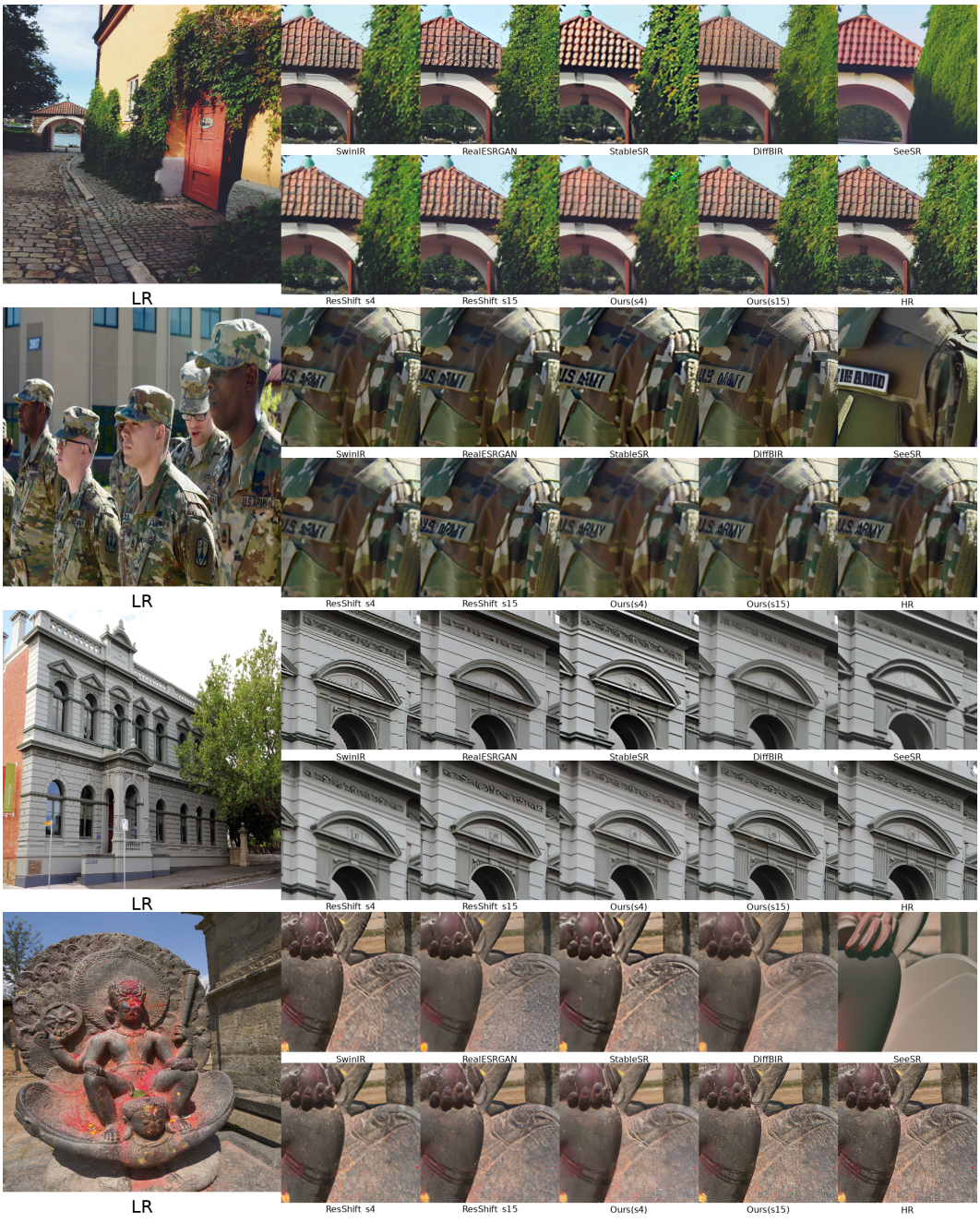


Figure 9: Experiment for Reference region. Qualitative comparison on the Flickr30k dataset Young et al. (2014). Ours (s4) and Ours (s15) represent ResShift (s4 and s15) with the proposed restoration in Eq. 11. Please note that most of the collapsed patterns and text that failed to be reconstructed in the baseline have been successfully restored.

Figure 10: Experiment for Reference region. Qualitative comparison on the Flickr30k dataset Young et al. (2014). Ours (s4) and Ours (s15) represent ResShift (s4 and s15) with the proposed restoration in Eq. 11. While the collapsed patterns were successfully restored, various fine textures originally present on the owl and the wall were rather simplified.

Figure 11: Experiment for Non-Reference region. Qualitative comparisons of image quality across various algorithms on the DIV2K dataset Agustsson & Timofte (2017). Our method demonstrates outstanding performance in enhancing the restoration of unstructured objects that lack clear semantic information. Note the significant perceptual quality enhancement compared to the baseline ResShift network.

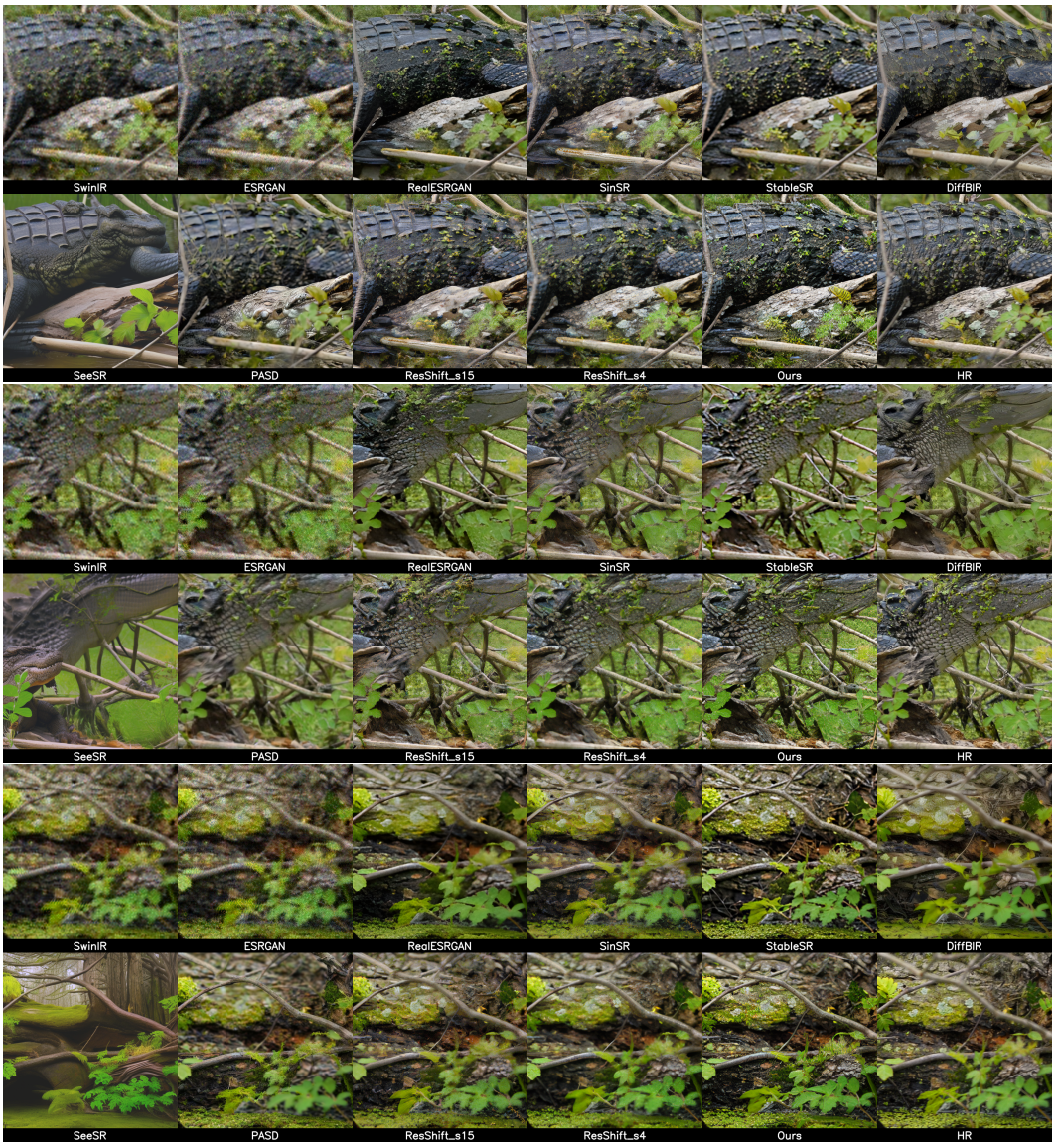


Figure 12: Experiment for Non-Reference region. Qualitative comparisons of image quality across various algorithms on the DIV2K dataset Agustsson & Timofte (2017). Note that the perceptual quality is improved.