AdaBridge-SR: Adaptive Bridge Matching for Real-World Image Super-Resolution

Jiangang Wang[1,2], Shangquan Sun[1], Aiping Zhang[1,2], Yuning Cui[3], and Wenqi Ren[1,2†]

[1] School of Cyber Science and Technology, Shenzhen Campus of Sun Yat-sen University, Shenzhen, 518107, China
[2] Shenzhen Key Laboratory of Adversarial Artificial Intelligence, Shenzhen, 518107, China
[3] School of Computation, Information and Technology, Technical University of Munich, 80333 Munich, Germany
wangjg33@mail2.sysu.edu.cn, renwq3@mail.sysu.edu.cn

Abstract. Despite their remarkable performance in real-world image super-resolution, diffusion models remain challenged by the perception–distortion trade-off between structural fidelity and perceptual realism. Existing methods typically rely on a globally predefined restoration trajectory, applying spatially uniform noise and a fixed temporal schedule. Such rigid trajectories often lead to a dilemma: reliable structures may be distorted, while severely degraded textures may become over-smoothed. To address this, we propose AdaBridge-SR, built upon a novel *spatio-temporal adaptive bridge matching* (ST-ABM) formulation that casts restoration as a controlled bridge between degraded and clean image distributions. ST-ABM decouples the bridge stochasticity into spatial and temporal controls, adaptively determining *where* to inject stochasticity and *when* to allocate generative capacity conditioned on the input degradation. We instantiate these controls with a lightweight Adaptive Bridge Controller that predicts a spatial stochasticity map and a time reparameterization. Consequently, AdaBridge-SR unifies deterministic restoration and stochastic exploration within a single model. Extensive experiments demonstrate that it achieves a superior perception–distortion balance, supporting efficient deterministic one-step inference and high-quality stochastic few-step bridge refinement. Our code and models are available at https://github.com/W-JG/AdaBridge-SR.

Keywords: Diffusion Models · Bridge Matching · Image Super-Resolution

1 Introduction

Diffusion models [12, 28, 29] have substantially improved perceptual quality in real-world image super-resolution (Real-ISR) [21, 32, 35, 36, 39]. However, they remain challenged by the perception–distortion trade-off [5, 39], in which

† Corresponding author.

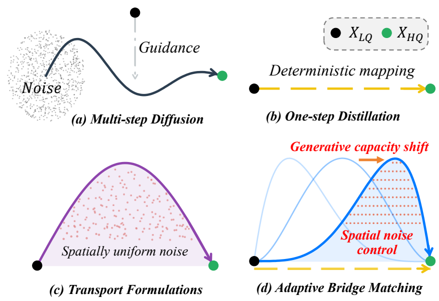

Fig. 1: Generative Real-ISR methods. Unlike existing methods (a–c) that rely on globally predefined restoration trajectories, our AdaBridge-SR (d) bridges degraded and clean image distributions via spatio-temporal controls.

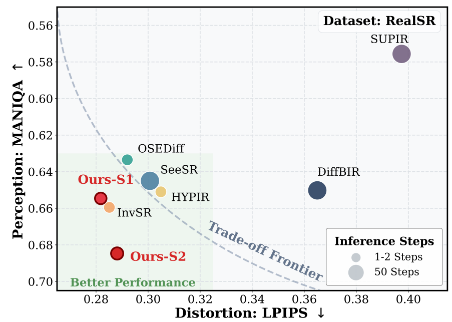

Fig. 2: Perception–distortion trade-off. AdaBridge-SR (Ours-S1/Ours-S2) achieves a superior balance (higher MANIQA, lower LPIPS) using only 1–2 inference steps.

enforcing structural fidelity often compromises perceptual realism, and vice versa. Specifically, structural fidelity requires strict adherence to the degraded input to preserve reliable structures, whereas perceptual realism often benefits from generative stochasticity to synthesize plausible textures [6, 18, 22, 33, 43]. Within a single image, relatively reliable regions, such as flat areas and well-preserved edges, favor deterministic restoration, whereas ambiguous or heavily degraded regions benefit from stochastic exploration. Furthermore, degradation severity varies substantially across images. These observations raise a central question for Real-ISR: **where and when should stochastic restoration be activated?**

Most existing methods leave this question implicit, relying instead on a globally predefined restoration trajectory. Existing generative Real-ISR methods mainly fall into three categories. Multi-step diffusion methods [17, 36, 39] typically rely on additional guidance modules [45] to guide a stochastic noise-to-image restoration trajectory conditioned on the degraded input (Fig. 1(a)). However, their spatially uniform noise injection can distort reliable structures and requires expensive iterative sampling. In contrast, one-step diffusion methods [9, 18, 35, 42] compress the diffusion trajectory into a single deterministic image-to-image mapping (Fig. 1(b)). While efficient, this direct mapping may sacrifice the stochasticity needed to model complex and uncertain degradations. Meanwhile, transport-based formulations, including flow and bridge matching [8, 15, 19, 20, 23, 27], provide a natural framework for restoration, as they directly model the transformation from degraded to clean image distributions. However, they typically assume spatially uniform noise injection and a fixed temporal schedule (Fig. 1(c)). Overall, applying a globally predefined restoration trajectory across all spatial locations and time steps often leads to a dilemma: it may either over-smooth degraded textures or introduce generative artifacts into reliable structures.

To address this dilemma, we propose **AdaBridge-SR**, built upon a novel *Spatio-Temporal Adaptive Bridge Matching* (ST-ABM) formulation (Fig. 1(d)). ST-ABM reformulates bridge matching for Real-ISR by generalizing a Brownian bridge into an input-conditioned controlled bridge process between degraded and clean image distributions. Crucially, ST-ABM decouples bridge stochasticity into two complementary controls: **where** to inject stochasticity and **when** to allocate generative capacity, thereby defining a dynamic, degradation-aware trajectory instead of a globally predefined restoration trajectory. We implement ST-ABM with a lightweight Adaptive Bridge Controller (ABC), which is conditioned on the degraded input and produces the spatial and temporal controls: (i) Spatial control: ABC predicts a spatial stochasticity map that injects stochasticity into uncertain regions for stochastic exploration, while encouraging deterministic restoration in reliable regions. (ii) Temporal control: ABC estimates a time reparameterization, allocating the generative capacity along the trajectory based on the overall degradation severity. This controlled bridge yields a unified restoration formulation: it adaptively determines where to activate stochastic exploration and when to allocate generative capacity. Under this unified formulation, a single AdaBridge-SR model supports efficient deterministic one-step inference and high-quality stochastic few-step bridge refinement (Fig. 2).

Our main contributions are threefold:

- We propose ST-ABM, a novel bridge matching formulation for Real-ISR that generalizes the Brownian bridge into an input-conditioned controlled bridge process, explicitly modeling where and when stochastic restoration should be activated for each input.
- We develop AdaBridge-SR by instantiating ST-ABM with a lightweight ABC that predicts spatial and temporal controls, enabling a unified restoration process that combines deterministic structure recovery and stochastic texture synthesis for adaptive refinement.
- Extensive experiments and ablations show that AdaBridge-SR achieves a strong perception–distortion balance and high computational efficiency, while supporting efficient deterministic one-step inference and high-quality stochastic few-step bridge refinement within a single model.

2 Related Work

2.1 Multi-step Diffusion-based Real-ISR Methods

Early generative Real-ISR methods [6,33,43] largely relied on generative adversarial networks (GANs) [10] to synthesize high-frequency textures. Recently, diffusion models [12,28,29] have achieved state-of-the-art performance by leveraging powerful generative priors [25]. To guide the restoration process, additional guidance modules [45] are commonly used in multi-step diffusion methods [17,36,39]. While these methods achieve strong perceptual realism, they typically require 20–200 iterative denoising steps, which limits their practical deployment. More importantly, these multi-step methods typically rely on a globally predefined noise-to-image denoising trajectory with spatially uniform noise injection. Such uniform stochasticity ignores the spatial heterogeneity inherent in real-world degradations. Consequently, stochasticity strong enough to synthesize textures in uncertain regions may distort reliable regions. Unlike these spatially uniform approaches, AdaBridge-SR introduces a spatial control mechanism that adaptively activates stochastic exploration in uncertain regions, while preserving structural fidelity in reliable regions.

2.2 One-step Diffusion-based Real-ISR Methods

To reduce the inference cost of multi-step sampling, recent works compress diffusion trajectories into a single network evaluation. Distillation-based methods [9,35] train a student model to approximate a multi-step diffusion trajectory. Meanwhile, other works [18, 30, 37, 42] use pretrained diffusion backbones as initialization for adversarial fine-tuning. These accelerated methods often achieve efficiency by collapsing stochastic sampling into a largely deterministic image-to-image mapping. However, this direct mapping may suppress the stochasticity needed to model complex and ambiguous degradations. Because real images contain both reliable structures and severely degraded textures, a deterministic image-to-image mapping may struggle to generate plausible details in highly uncertain regions. AdaBridge-SR addresses this by adaptively activating stochastic restoration, unifying deterministic structure recovery in reliable regions and stochastic texture synthesis in uncertain regions.

2.3 Transport-based Real-ISR Methods

Transport-based methods provide a principled alternative by learning dynamics that map one distribution to another. Flow matching [19, 23] formulates generation as solving deterministic ordinary differential equations (ODEs), enabling efficient few-step generation. In contrast, bridge matching [3, 4, 8, 15, 20, 27] introduces stochasticity through bridge processes (e.g., Brownian bridges [8]). In Real-ISR, methods such as ResShift [41] model transport between degraded and clean image distributions. Despite these advances, most existing flow- and bridge-based methods typically rely on spatially uniform stochasticity and a fixed temporal schedule, which may adapt poorly to varying degradation severity across real-world inputs. A closely related method, UPSR [44], introduces spatial uncertainty weighting to modulate stochasticity. However, it still follows a fixed temporal schedule. AdaBridge-SR extends these fixed-trajectory transport formulations into a spatio-temporally adaptive formulation. AdaBridge-SR supports efficient deterministic one-step inference and high-quality stochastic few-step bridge refinement within a single model, with spatial stochasticity and temporal allocation modulated according to the input degradation.

3 Methodology

3.1 Preliminaries: Bridge Matching

Bridge Matching (BM) [8, 15, 20, 27] formulates paired generation as learning a stochastic bridge that transports a source distribution π_0 to a target distribution

π_1. In paired image restoration, π_0 corresponds to degraded samples and π_1 to clean samples. Given an endpoint pair (x_0, x_1) with $x_0 \sim \pi_0$ and $x_1 \sim \pi_1$, BM constructs a stochastic interpolant $\{x_t\}_{t\in[0,1]}$ satisfying $x_{t=0} = x_0$ and $x_{t=1} = x_1$. A common choice is the Gaussian bridge family:

$$x_t = (1-t)x_0 + tx_1 + \sigma_t\,\epsilon, \qquad \epsilon \sim \mathcal{N}(0, \mathbf{I}), \tag{1}$$

where σ_t controls the time-dependent noise amplitude and satisfies $\sigma_0 = \sigma_1 = 0$ for endpoint consistency. A widely used instance is the Brownian bridge with a symmetric schedule:

$$\sigma_t = \sigma\sqrt{t(1-t)}, \qquad \sigma > 0, \tag{2}$$

where σ is a global noise-scale parameter. Eq. (1) defines the conditional distribution:

$$x_t \mid (x_0, x_1) \sim \mathcal{N}\big((1-t)x_0 + tx_1,\ \sigma_t^2\mathbf{I}\big), \tag{3}$$

whose variance peaks at $t = 0.5$ and vanishes at the endpoints, naturally enforcing endpoint consistency.

Classical BM is often parameterized by a time-dependent conditional velocity field that transports samples along the bridge. For paired restoration, a convenient endpoint-consistent target, defined for $t \in [0, 1)$, is

$$u^*(x_t, t; x_0, x_1) = \frac{x_1 - x_t}{1-t}, \tag{4}$$

which guarantees $x_1 = x_t + (1-t)u^*(x_t, t; x_0, x_1)$. This endpoint identity motivates our endpoint-prediction parameterization, where the model predicts the clean endpoint rather than explicitly regressing the velocity field.

Limitations for Real-ISR. Despite its effectiveness, standard BM is misaligned with Real-ISR due to two rigid assumptions. (i) *Spatially uniform stochasticity*: $\sigma_t^2\mathbf{I}$ applies identical noise to all pixels, ignoring spatially varying degradation. (ii) *Fixed temporal allocation*: the symmetric schedule enforces maximal stochasticity at $t = 0.5$, limiting input-adaptive redistribution of modeling capacity along the trajectory. These limitations motivate an adaptive bridge with spatial and temporal controllability.

3.2 Spatio-Temporal Adaptive Bridge Matching (ST-ABM)

To address the above limitations, we propose ST-ABM, which generalizes BM into an input-conditioned *controlled bridge* between a degraded endpoint x_0 and a clean endpoint x_1. ST-ABM introduces two complementary controls conditioned on x_0: (i) *where* to inject stochasticity via a stochasticity map $\mathbf{M} = \mathcal{M}(x_0) \in [0, 1]^{H\times W\times 1}$, and (ii) *when* to allocate stochasticity via a time reparameterization $\tau = \Phi(t; x_0) \in [0, 1]$.

Controlled bridge. For clarity, we formulate the bridge in the intrinsic time variable τ, while the external time t enters only through $\tau = \Phi(t; x_0)$. Let $\Sigma(x_0) = \mathrm{Diag}(\mathrm{vec}(\mathbf{M}))$ denote the diagonal diffusion operator induced by $\mathbf{M}$, with $\mathbf{M}$ broadcast to all channels when needed, and let $\beta > 0$ denote a global noise-amplitude scale. We define the controlled bridge family as the conditional Gaussian:

$$x_\tau \mid (x_0, x_1) \sim \mathcal{N}\Big(\mu_\tau(x_0, x_1),\ \beta^2 \tau(1-\tau)\Sigma(x_0)\Sigma(x_0)^\top\Big), \quad (5)$$
$$\mu_\tau(x_0, x_1) = (1-\tau)x_0 + \tau x_1.$$

Equivalently, sampling can be written as:

$$x_\tau = (1-\tau)x_0 + \tau x_1 + \beta\sqrt{\tau(1-\tau)}\,\Sigma(x_0)\epsilon, \qquad \epsilon \sim \mathcal{N}(0, \mathbf{I}). \quad (6)$$

Spatio-temporal variance factorization. Conditioned on the endpoints (x_0, x_1), ST-ABM yields an explicit decoupling of bridge stochasticity:

$$\mathrm{Var}(x_\tau \mid x_0, x_1) = \beta^2\tau(1-\tau)\Sigma(x_0)\Sigma(x_0)^\top, \quad (7)$$

demonstrating that $\Sigma(x_0)$ modulates stochasticity *spatially* (where to explore), while τ shifts stochasticity *temporally* (when to reach high-variance states).

From Controlled Interpolant to a Controlled Bridge SDE. Eq. (6) can be interpreted as a Brownian-bridge-like process defined in intrinsic time τ, followed by the time reparameterization $\tau = \Phi(t; x_0)$.

Intrinsic-time bridge dynamics. In intrinsic time $\tau \in [0, 1)$, we consider the controlled bridge stochastic differential equation (SDE):

$$\mathrm{d}x_\tau = u^*(x_\tau, \tau; x_0, x_1)\,\mathrm{d}\tau + \beta\Sigma(x_0)\,\mathrm{d}W_\tau, \qquad u^*(x_\tau, \tau; x_0, x_1) = \frac{x_1 - x_\tau}{1-\tau}, \quad (8)$$

where W_τ is a standard Wiener process. The velocity target implies the endpoint identity $x_1 = x_\tau + (1-\tau)u^*(x_\tau, \tau; x_0, x_1)$ and aligns with the bridge constraint. With the initial condition $x_{\tau=0} = x_0$, solving this linear SDE recovers the Gaussian bridge in Eq. (5).

External-time controlled SDE via time change. Applying the monotone time change $\tau = \Phi(t; x_0)$ with $\dot{\tau}(t) = \frac{d\tau}{dt} > 0$ for $t > 0$ gives $d\tau = \dot{\tau}(t)dt$ and $dW_\tau = \sqrt{\dot{\tau}(t)}dW_t$.

For the time-changed process, denoted again by x_t for simplicity, we obtain

$$\mathrm{d}x_t = \dot{\tau}(t)\,u^*(x_t, \tau(t); x_0, x_1)\,\mathrm{d}t + \beta\Sigma(x_0)\sqrt{\dot{\tau}(t)}\,\mathrm{d}W_t. \quad (9)$$

Eq. (9) makes the control roles explicit: $\Sigma(x_0)$ gates diffusion *spatially*, while $\dot{\tau}(t)$ rescales both drift and diffusion *temporally*.

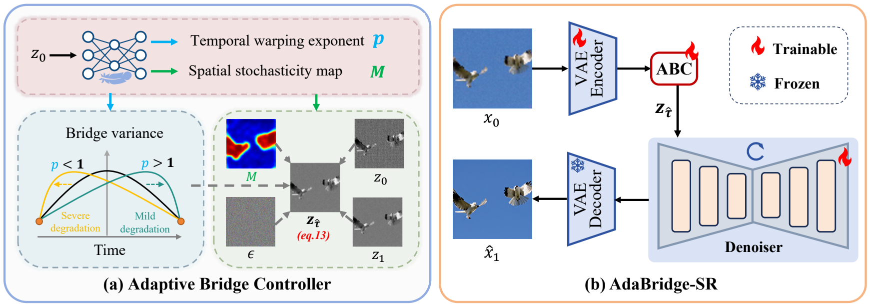

Fig. 3: Overall framework of AdaBridge-SR. (a) The lightweight Adaptive Bridge Controller (ABC) predicts spatio-temporal bridge controls from the input. It estimates a temporal warping exponent p to allocate generative capacity along the bridge trajectory according to the input degradation severity and a stochasticity map $\mathbf{M}$ to activate stochastic exploration spatially. (b) The AdaBridge-SR architecture operates in the latent space. During training, the trainable ABC predicts the spatio-temporal controls from the degraded latent z_0 to construct the controlled bridge latent $z_{\hat{\tau}}$ in Eq. (13), which is then processed by the trainable denoising backbone to refine the clean-endpoint estimate.

ODE vs. SDE under a unified formulation. In regions where $\mathcal{M}(x_0) \to \mathbf{0}$ (thus $\Sigma(x_0) \to \mathbf{0}$), Eq. (9) collapses into a deterministic ODE update that preserves reliable structures. Where $\mathcal{M}(x_0)$ is large, the process becomes an SDE that injects generative stochasticity to explore plausible textures in uncertain regions. This unified view directly answers the Real-ISR question: *where and when should stochastic restoration be activated?*

3.3 Adaptive Bridge Controller (ABC)

ST-ABM requires two input-conditioned controls: a spatial stochasticity map and a temporal schedule modulation. We instantiate both with a lightweight ABC C_ψ (illustrated in Fig. 3(a)), implemented as a single network with two prediction heads. Given the degraded endpoint x_0, ABC outputs:

$$(\mathbf{M}, p) = C_\psi(x_0), \qquad \mathbf{M} \in [0,1]^{H\times W\times 1}, \qquad p \in [p_{\min}, p_{\max}], \tag{10}$$

where $0 < p_{\min} < p_{\max}$. The map head obtains $\mathbf{M}$ by a sigmoid activation followed by mean-normalization and clamping, which stabilizes its scale and avoids degenerate all-zero/all-one solutions. The temporal head maps its raw prediction to $[p_{\min}, p_{\max}]$ by a bounded sigmoid-affine transform, ensuring a monotone time warp.

Spatial control (where). The stochasticity map $\mathbf{M}$ specifies the local noise amplitude of the controlled bridge. Small values of $\mathbf{M}$ suppress stochastic perturbations and favor deterministic structure preservation, whereas large values allow stronger stochastic exploration in ambiguous or severely degraded regions.

Temporal control (when). We operate on a discrete timestep set $\mathcal{T}$ with $t_{\max} = \max \mathcal{T}$ and define a base mixing coefficient $\gamma(t) = t/t_{\max} \in [0,1]$. ABC warps the temporal allocation using a power law:

$$\hat{\tau}(t) = \Phi(t; x_0) = \gamma(t)^p, \tag{11}$$

so that $p < 1$ allocates stochasticity earlier for severely degraded inputs, while $p > 1$ delays stochasticity to emphasize late-stage refinement on mild degradations.

3.4 Model Training and Objectives

To inherit strong generative priors for Real-ISR, we build AdaBridge-SR upon a pretrained latent diffusion backbone (illustrated in Fig. 3(b)). In practice, we instantiate the same ST-ABM formulation in the latent space. Let $\mathcal{E}$ and $\mathcal{D}$ denote the VAE encoder and decoder, respectively. Given a paired training sample (x_0, x_1), we encode

$$z_0 = \mathcal{E}(x_0), \qquad z_1 = \mathcal{E}(x_1). \tag{12}$$

From this point on, the variables in the bridge are latent: x_0, x_1, x_τ in the abstract formulation are replaced by $z_0, z_1, z_{\hat{\tau}}$. Accordingly, the ABC is conditioned on the degraded latent z_0 and predicts the latent-domain controls $(\mathbf{M}, p) = C_\psi(z_0)$.

During training, we sample a base mixing coefficient s from a discrete timestep set that includes $s = 0$, and obtain the controlled intrinsic time $\hat{\tau} = s^p$. The corresponding controlled bridge latent is constructed as

$$z_{\hat{\tau}} = (1 - \hat{\tau})z_0 + \hat{\tau} z_1 + \beta\, \mathbf{M} \odot \sqrt{\hat{\tau}(1 - \hat{\tau})}\epsilon, \qquad \epsilon \sim \mathcal{N}(0, \mathbf{I}), \tag{13}$$

where β is the global noise-amplitude scale defined in Eq. (6). The controlled bridge latent $z_{\hat{\tau}}$ is then used as the bridge input to the denoising backbone, which follows the endpoint-prediction parameterization and outputs the clean latent $\hat{z}_1$ as the endpoint estimate of the bridge.

Stochasticity map supervision. To explicitly guide *where* stochastic exploration should occur, we supervise the stochasticity map $\mathbf{M}$ using a normalized residual proxy. Rather than relying on absolute errors, which vary drastically across different inputs, we encourage the network to capture relative uncertainty within each sample. Specifically, we first compute the channel-averaged absolute latent difference $e_z = \frac{1}{C_z} \sum_{c=1}^{C_z} |z_0^c - z_1^c|$, where C_z is the number of latent channels. We then normalize it by its spatial mean $\bar{e}_z = \mathrm{mean}(e_z)$ and clip the values to define the target map $\mathbf{R}$:

$$\mathbf{R} = \mathrm{clip}\left(\frac{e_z}{\bar{e}_z + \delta}, 0, 1\right), \qquad \mathcal{L}_{\mathrm{map}} = \mathrm{SmoothL1}(\mathbf{M}, \mathbf{R}), \tag{14}$$

where $\delta = 10^{-6}$ is a small constant for numerical stability. This formulation encourages $\mathbf{M}$ to highlight locally challenging regions while assigning smaller stochasticity to relatively reliable regions.

EMA-based temporal supervision. To guide *when* stochasticity is allocated along the bridge, we dynamically estimate the relative degradation severity of each input. Let $r = \mathrm{mean}(|z_0 - z_1|)$ denote the global latent residual strength of a given sample. To decouple this metric from dataset-level variation, we maintain an exponential moving average (EMA), denoted as $\hat{r}$, of the batch-wise mean residual. We define a relative severity score $\rho = \mathrm{clip}(r/(\hat{r} + \delta), 0, 2)$. The target temporal exponent p^* is then linearly mapped from ρ to a bounded valid range $[p_{\min}, p_{\max}]$:

$$p^*(\rho) = p_{\mathrm{mid}} - p_{\mathrm{half}} \cdot (\rho - 1), \qquad \mathcal{L}_{\mathrm{temp}} = \mathrm{SmoothL1}(p, p^*(\rho)), \tag{15}$$

where $p_{\mathrm{mid}} = (p_{\min} + p_{\max})/2$ and $p_{\mathrm{half}} = (p_{\max} - p_{\min})/2$. Consequently, severely degraded inputs ($\rho > 1$) receive a smaller exponent to shift the high-variance portion of the bridge earlier, whereas mildly degraded inputs ($\rho < 1$) delay it for late-stage refinement.

Overall objective. The final training objective is defined as

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{map}} \mathcal{L}_{\mathrm{map}} + \lambda_{\mathrm{temp}} \mathcal{L}_{\mathrm{temp}}. \tag{16}$$

Here $\mathcal{L}_{\mathrm{rec}}$ denotes the main endpoint restoration supervision for the denoising backbone, consisting of standard pixel reconstruction, perceptual, and adversarial terms applied to the predicted clean endpoint, following established paradigms [18, 37, 42] in generative image super-resolution. The overall objective jointly optimizes restoration fidelity, spatial stochasticity calibration, and degradation-aware temporal modulation.

3.5 Unified Inference: One-Step to Few-Step

AdaBridge-SR supports efficient deterministic one-step inference and high-quality stochastic few-step bridge refinement in a single model. Given the input latent $z_0 = \mathcal{E}(x_0)$, the inference follows a two-mode strategy, where N denotes the number of inference steps.

(i) One-step inference ($N = 1$): The denoising backbone performs a single deterministic forward pass from z_0 to directly predict the clean latent $\hat{z}_1^{(1)}$.

(ii) Few-step bridge refinement ($N > 1$): For bridge refinement, the ABC is evaluated *only once* to extract the controls $\mathbf{M}$ and p, which are reused across all refinement steps. We initialize $\hat{z}_1^{(1)}$ from the one-step prediction. For each refinement step $k = 2, \ldots, N$, we set $s_k = (k-1)/N$ and $\hat{\tau}_k = s_k^p$. Treating the previous prediction $\hat{z}_1^{(k-1)}$ as a dynamic clean endpoint, we construct the intermediate bridge latent as

$$z_{\hat{\tau}_k}^{(k)} = (1 - \hat{\tau}_k) z_0 + \hat{\tau}_k \hat{z}_1^{(k-1)} + \beta \mathbf{M} \odot \sqrt{\hat{\tau}_k (1 - \hat{\tau}_k)}\, \epsilon_k, \epsilon_k \sim \mathcal{N}(0, \mathbf{I}). \tag{17}$$

The denoising backbone then refines the clean-endpoint estimate $\hat{z}_1^{(k)}$ using $z_{\hat{\tau}_k}^{(k)}$. For few-step refinement, this recurrence dynamically injects spatially modulated noise to recover complex textures while preserving deterministic structures. The final latent $\hat{z}_1^{(N)}$ is decoded into pixel space as $\hat{x}_1^{(N)} = \mathcal{D}(\hat{z}_1^{(N)})$.

4 Experiments

4.1 Experimental Settings

Datasets. *(i) Training dataset.* Following previous work, we adopt the LSDIR dataset [16] for training. The corresponding low-quality (LQ) images are synthesized from their high-quality (HQ) counterparts using the complex degradation model in Real-ESRGAN [33], strictly maintaining identical degradation parameters for a fair comparison [32, 37]. *(ii) Test datasets.* We evaluate our method on four datasets, including RealSR [6], DRealSR [34], RealLQ250 [2], and DIV2K-Val [1]. For RealSR [6] and DRealSR [34], we follow the evaluation protocol in StableSR [32] for $\times 4$ super-resolution, where the image resolutions range from 128×128 to 512×512. This identical protocol is also applied to the synthetic DIV2K-Val [1] dataset.
Comparison Methods. We compare our approach with representative methods, including multi-step diffusion methods such as DiffBIR [17], SeeSR [36], and SUPIR [39], as well as one-step diffusion methods including OSEDiff [35], InvSR [40], and HYPIR [18]. For fairness, all competing methods are reproduced using their publicly available official implementations and pretrained models.
Evaluation Metrics. We evaluate all methods using both full-reference (FR) and no-reference (NR) image quality metrics. For FR evaluation, we adopt PSNR and SSIM, which are computed on the Y channel in the YCbCr color space to measure reconstruction fidelity, as well as LPIPS [46] to quantify perceptual similarity. The NR metrics include MUSIQ [14], MANIQA [38], and CLIPIQA+ [31], which estimate perceptual quality without reference images. These metrics reflect image quality from multiple perspectives. In this work, all metrics are computed using the PyIQA [7] toolkit.
Implementation Details. Our implementation is built upon SD2.1-base [26]. Similar to previous work [35, 42], we optimize LoRA [13] adapters in both the VAE encoder and the U-Net network. The LoRA rank is set to 16 for the VAE and 32 for the U-Net. The ABC utilizes a ResNet-based architecture [11]. The ABC is trained jointly with the LoRA adapters. We train the model using the AdamW [24] optimizer with a learning rate of 5×10^{-5}, a batch size of 1, and gradient accumulation of 4. Training runs for 8000 steps on four RTX 4090 GPUs. For fair comparison, we align the restoration supervision $\mathcal{L}_{\text{rec}}$ with OMGSR [37].

4.2 Comparison with State-of-the-Art Methods

Quantitative Comparison. To evaluate the effectiveness of the proposed method, we compare AdaBridge-SR against recent state-of-the-art methods on four widely used datasets. The quantitative results on RealSR, DRealSR, and DIV2K-Val are presented in Table 1. The evaluation metrics are divided into FR indicators for fidelity and NR indicators for perceptual quality. As the table illustrates, the one-step variant of the proposed model, denoted as Ours-S1, achieves strong structural fidelity across all three datasets. Specifically, Ours-S1

Table 1: Quantitative comparison of different super-resolution methods. The best and second-best results are highlighted in **red** and blue, respectively.

Dataset	Method	Steps	Fidelity			Perceptual		
			PSNR↑	SSIM↑	LPIPS↓	MUSIQ↑	MANIQA↑	CLIPIQA+↑
RealSR [6]	SeeSR [36]	50	25.1478	0.7210	0.3007	**69.8141**	0.6450	0.6911
	DiffBIR [17]	50	24.8336	0.6501	0.3650	69.1780	0.6502	**0.7234**
	SUPIR [39]	50	25.1184	0.6523	0.3974	58.8225	0.5756	0.6144
	OSEDiff [35]	1	25.1511	**0.7341**	0.2920	69.0960	0.6335	0.6966
	HYPIR [18]	1	22.9001	0.6859	0.3049	66.2074	0.6510	0.6805
	InvSR [40]	1	24.4480	0.7138	0.2851	68.3372	0.6595	0.6870
	Ours-S1	1	**25.2152**	0.7326	**0.2818**	67.6153	0.6546	0.6756
	Ours-S2	2	24.8815	0.7183	0.2881	69.2031	**0.6847**	0.7034
DRealSR [34]	SeeSR [36]	50	28.0722	0.7684	0.3174	65.0845	0.6052	0.6793
	DiffBIR [17]	50	25.9021	0.6244	0.4669	**66.1343**	0.6221	**0.7173**
	SUPIR [39]	50	26.8640	0.6464	0.4593	52.7430	0.5242	0.5861
	OSEDiff [35]	1	27.9220	0.7835	**0.2967**	64.6881	0.5898	0.6822
	HYPIR [18]	1	26.0742	0.7285	0.3282	61.3903	0.6024	0.6457
	InvSR [40]	1	26.4022	0.7152	0.3516	65.1059	0.6275	0.6818
	Ours-S1	1	**28.2226**	**0.7845**	0.2969	62.7943	0.6073	0.6596
	Ours-S2	2	27.8302	0.7637	0.3103	65.3485	**0.6436**	0.6849
DIV2K-Val [1]	SeeSR [36]	50	23.6780	0.6043	0.3193	68.6715	0.6222	0.7042
	DiffBIR [17]	50	23.1447	0.5441	0.3668	**69.8692**	0.6440	**0.7340**
	SUPIR [39]	50	22.9361	0.5260	0.3809	63.7793	0.6043	0.6784
	OSEDiff [35]	1	23.7235	0.6109	0.2942	67.9659	0.6131	0.6945
	HYPIR [18]	1	22.2203	0.5763	0.3077	66.2385	0.6244	0.6886
	InvSR [40]	1	23.2892	0.5921	0.3167	68.6748	0.6296	0.7046
	Ours-S1	1	**23.9717**	**0.6145**	**0.2926**	65.4500	0.6210	0.6605
	Ours-S2	2	23.5063	0.5958	0.2954	67.7713	**0.6556**	0.6942

Table 2: Quantitative comparison on the RealLQ250 dataset. Since ground-truth high-quality images are unavailable for this dataset, only NR perceptual metrics are reported. The best and second-best results are highlighted in **red** and blue, respectively.

Metric	Multi-step methods			One-step methods			Ours	
	SeeSR [36]	DiffBIR [17]	SUPIR [39]	OSEDiff [35]	HYPIR [18]	InvSR [40]	**Ours-S1**	**Ours-S2**
MUSIQ↑	**70.3776**	67.5317	62.3367	69.5520	69.7529	65.5552	67.3177	69.8798
MANIQA↑	0.5927	0.5877	0.5792	0.5782	0.6075	0.5719	0.5836	**0.6218**
CLIPIQA+↑	0.7033	0.6918	0.6170	0.6981	**0.7195**	0.6696	0.6750	0.7071

obtains the highest PSNR and the best or second-best LPIPS, outperforming most one-step and multi-step baselines in fidelity.

Furthermore, the perception-distortion trade-off is effectively navigated by extending the inference trajectory to merely two steps. As shown in Table 1, Ours-S2 exhibits highly competitive perceptual quality. While multi-step methods such as DiffBIR [17] and SeeSR [36] require 50 inference steps to achieve high scores in MUSIQ and CLIPIQA+, they severely compromise structural fidelity, as evidenced by their degraded PSNR and high LPIPS values. Conversely, Ours-S2 achieves the best MANIQA across all three datasets and remains competitive in MUSIQ and CLIPIQA+, while maintaining a clear LPIPS advantage over these multi-step methods. This indicates that the adaptive temporal scheduling

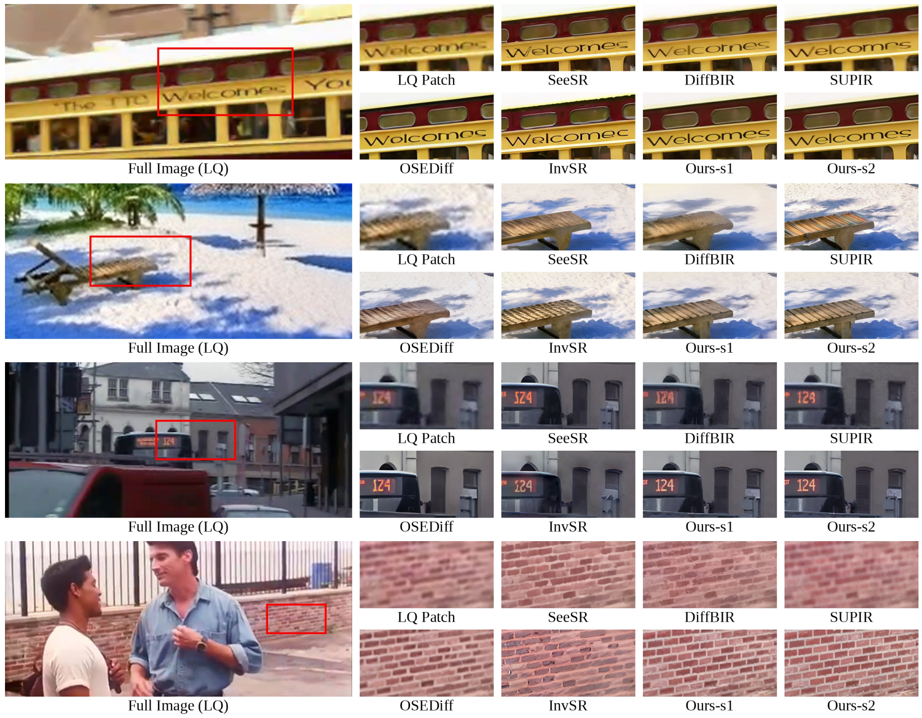

Fig. 4: Qualitative comparison of different super-resolution methods. **Zoom in for details.**

of the model appropriately allocates generative capacity, enabling the synthesis of realistic textures.

To further validate the generalization capability of the proposed approach on unknown real-world degradations, we present the evaluation results on the RealLQ250 dataset in Table 2. Since ground-truth images are unavailable for this dataset, only NR metrics are utilized for assessment. The results indicate that AdaBridge consistently delivers top-tier visual realism, achieving the highest MANIQA score and the second-best performance in both MUSIQ and CLIPIQA+. The stable performance of the proposed method on real-world inputs confirms that the learned spatio-temporal bridge accurately models complex degradations and synthesizes plausible details without relying on excessive iterative sampling.

Qualitative Comparison. Fig. 4 presents visual comparisons between our AdaBridge-SR and state-of-the-art Real-ISR methods. The results clearly demonstrate how our proposed ST-ABM effectively alleviates the perception-distortion dilemma by explicitly controlling where and when to inject stochasticity. For reliable structures with strict semantic constraints, such as the text "Welcomes" (first row) and the bus number "124" (third row), existing diffusion-based methods frequently introduce severe generative artifacts or spelling distortions (e.g., OSEDiff and InvSR). In contrast, ABC accurately identifies these regions and enforces deterministic restoration, better preserving structural fidelity and re-

Table 3: Ablation study on the core components of the Adaptive Bridge Controller (ABC). Experiments are conducted under the 2-step inference setting. **M** denotes the spatial stochasticity map, and p denotes the temporal warping exponent in $\hat{\tau} = \gamma(t)^p$.

M	p	RealSR [6]					DRealSR [34]				
Spatial	Temporal	PSNR↑	SSIM↑	LPIPS↓	MUSIQ↑	MANIQA↑	PSNR↑	SSIM↑	LPIPS↓	MUSIQ↑	MANIQA↑
×	×	24.6818	0.7051	0.2893	69.2005	0.6853	27.4918	0.7424	0.3122	65.6529	0.6424
×	✓	24.5748	0.6983	0.3007	69.6987	**0.6886**	27.4454	0.7332	0.3375	**66.5795**	**0.6503**
✓	×	24.6160	0.7048	0.2906	**69.7787**	0.6873	27.4471	0.7416	0.3244	66.1481	0.6463
✓	✓	**24.8815**	**0.7182**	**0.2881**	69.2030	0.6847	**27.8302**	**0.7637**	**0.3103**	65.3485	0.6436

ducing hallucinated distortions. Conversely, for heavily degraded areas requiring stochastic exploration, such as the wooden chair slats (second row) and the brick wall (fourth row), global deterministic trajectories tend to over-smooth details, while predefined multi-step priors (e.g., SUPIR) often synthesize unnatural, over-sharpened patterns. Ours-S1 performs deterministic one-step inference, preserving reliable semantic structures with a single forward pass. With stochastic few-step bridge refinement, Ours-S2 selectively allocates generative capacity to uncertain regions, synthesizing realistic and plausible textures. Together, the two variants demonstrate the flexibility of AdaBridge-SR in balancing structural fidelity and perceptual realism under different inference budgets.

4.3 Ablation Studies

Effectiveness of Spatio-Temporal Decoupling. The core innovation of AdaBridge-SR is the decoupling of bridge stochasticity into spatial and temporal controls. To validate this, we deconstruct the ABC and evaluate its components using a 2-step inference budget, as detailed in Table 3. The baseline model (without **M** and p) acts as a standard bridge with uniform noise and fixed schedules. When only the temporal warping exponent p is introduced, we observe a slight increase in perceptual metrics (e.g., MUSIQ on DRealSR rises to 66.5795). However, this purely temporal adjustment causes a severe drop in structural fidelity (LPIPS degrades to 0.3375 on DRealSR). Crucially, the full model (ST-ABM) achieves the best fidelity-oriented trade-off when both components are integrated, yielding a substantial leap in fidelity (e.g., PSNR increases from 24.6818 to 24.8815 on RealSR) while maintaining competitive perceptual realism. This demonstrates that dynamic temporal allocation is only safely effective when rigorously constrained by spatial stochasticity modulation.

Effectiveness of Inference Steps. To investigate the effect of inference steps on restoration performance, an ablation study is conducted on the RealSR and DRealSR datasets. The quantitative results are visually summarized in Fig. 5. By varying the number of inference steps N from 1 to 4, a smooth and continuous transition between structural fidelity and perceptual quality is observed. At $N = 1$, the proposed model performs deterministic one-step inference, achieving strong structural preservation with low LPIPS scores. The fidelity at this stage is comparable to or better than the performance of the representative one-step baseline, OSEDiff. As the number of inference steps progressively increases, the

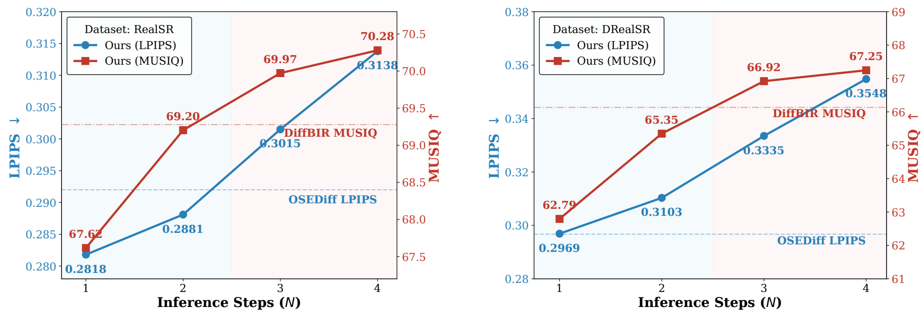

Fig. 5: Ablation study on the number of inference steps across the RealSR (left) and DRealSR (right) datasets. The dual-axis plots illustrate the continuous performance transition between structural fidelity (LPIPS) and perceptual realism (MUSIQ). Horizontal reference lines indicate the baseline performance of OSEDiff (1-step) and DiffBIR (50-step). Our method allows a flexible trade-off by simply varying N.

Table 4: Ablation study on the global noise-amplitude scale β under the 4-step inference setting. The proposed setting uses $\beta = 0.005$ (Ours).

β	RealSR [6]					DRealSR [34]				
	PSNR↑	SSIM↑	LPIPS↓	MUSIQ↑	MANIQA↑	PSNR↑	SSIM↑	LPIPS↓	MUSIQ↑	MANIQA↑
0	24.3200	0.6817	0.3284	69.7594	0.7003	27.0703	0.7086	0.3744	67.2754	0.6658
0.001	24.3614	0.6855	0.3288	70.6916	0.6992	**27.2931**	**0.7214**	0.3631	67.8448	0.6645
0.005	**24.4671**	**0.6884**	**0.3138**	70.2799	**0.7177**	27.2116	0.7176	**0.3548**	67.2462	0.6756
0.01	24.0997	0.6770	0.3275	**70.7515**	0.7131	26.9332	0.7035	0.3757	**68.0651**	**0.6770**

model enters a stochastic bridge refinement regime, where additional generative capacity is allocated to improve perceptual realism. This mechanism significantly enhances perceptual realism, as indicated by the monotonic rise in MUSIQ scores. Notably, at $N = 4$, the perceptual quality of the proposed approach reaches or exceeds the MUSIQ level of the 50-step DiffBIR baseline. These observations confirm that a single unified architecture of AdaBridge-SR effectively navigates the perception-distortion trade-off without necessitating separate training phases or distinct network weights for different inference budgets.

Effect of Global Noise-Amplitude Scale. To validate the spatial stochasticity modulation presented in Section 3.3, we ablate the global noise-amplitude scale β under a four-step inference budget. The results on RealSR and DRealSR are summarized in Table 4. When $\beta = 0$, the bridge trajectory reduces to a deterministic ODE, which limits plausible texture synthesis in severely degraded regions and leads to elevated LPIPS scores. Conversely, increasing β strengthens stochastic exploration. A large value (*e.g.*, $\beta = 0.01$) improves perceptual quality, but excessive stochasticity comes at the cost of full-reference fidelity. In contrast, the adopted setting $\beta = 0.005$ strikes a better balance between deterministic structure preservation and stochastic texture synthesis. It achieves the lowest LPIPS scores across both datasets while maintaining highly competitive perceptual realism. These observations substantiate the critical role of calibrated spatial noise modulation in navigating the perception–distortion trade-off.

Table 5: Average inference latency of AdaBridge-SR on an NVIDIA RTX 4090 GPU at 512×512 resolution. ABC overhead is the runtime of the Adaptive Bridge Controller.

Variant	Steps	Total (ms/img)	ABC Overhead (ms/img)	Overhead (%)
Ours-S1	1	96.38	0.00	0.00
Ours-S2	2	162.49	1.77	1.09

Inference Efficiency and Controller Overhead. We measure the average inference latency of AdaBridge-SR on a single NVIDIA RTX 4090 GPU at 512×512 resolution. As shown in Table 5, Ours-S1 processes one image in 96.38 ms without invoking the ABC, while Ours-S2 performs two-step bridge refinement in 162.49 ms. This confirms that the deterministic one-step mode remains highly efficient for latency-sensitive scenarios. In Ours-S2, the controller is evaluated once and takes only 1.77 ms, accounting for 1.09% of the total latency. Since the predicted controls are reused across refinement steps, the additional cost of adaptive bridge refinement remains marginal. These results show that the controller is lightweight and introduces negligible overhead for a few-step refinement.

5 Conclusion

In this paper, we propose AdaBridge-SR to address the fundamental perception–distortion trade-off in real-world image super-resolution. To overcome the limitations of globally predefined restoration trajectories, we introduce ST-ABM. Through a lightweight controller, ST-ABM explicitly *decouples* bridge stochasticity into spatial and temporal controls, dynamically deciding *where* to inject noise and *when* to allocate generative capacity. This formulation unifies deterministic restoration for reliable structures with stochastic exploration for heavily degraded regions. By varying the inference steps, the same trained model can smoothly adjust the balance between structural fidelity and perceptual realism. Extensive experiments demonstrate that AdaBridge-SR achieves a superior perception–distortion balance while supporting efficient deterministic one-step inference and high-quality stochastic few-step bridge refinement within a single model.

Acknowledgements

This work was supported by National Natural Science Foundation of China (No. U24B20175, No. 62322216), Scientific Research Innovation Capability Support Project for Young Faculty (No. SRICSPYF-ZY2025012), GBA Ascend Application Innovation Institute, Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) under Grant No. GML-ST-2026-04, and Shenzhen Science and Technology Program (No. SYSRD20250529113401002).

References

1. Agustsson, E., Timofte, R.: NTIRE 2017 challenge on single image super-resolution: Dataset and study. In: CVPR Workshops (2017)
2. Ai, Y., Zhou, X., Huang, H., Han, X., Chen, Z., You, Q., Yang, H.: DreamClear: High-capacity real-world image restoration with privacy-safe dataset curation. In: NeurIPS (2024)
3. Albergo, M., Boffi, N.M., Vanden-Eijnden, E.: Stochastic interpolants: A unifying framework for flows and diffusions. JMLR (2025)
4. Albergo, M.S., Vanden-Eijnden, E.: Building normalizing flows with stochastic interpolants. In: ICLR (2023)
5. Blau, Y., Michaeli, T.: The perception-distortion tradeoff. In: CVPR (2018)
6. Cai, J., Zeng, H., Yong, H., Cao, Z., Zhang, L.: Toward real-world single image super-resolution: A new benchmark and a new model. In: ICCV (2019)
7. Chen, C., Mo, J.: IQA-PyTorch: PyTorch toolbox for image quality assessment. [Online]. Available: `https://github.com/chaofengc/IQA-PyTorch` (2022), accessed 1 July 2026
8. De Bortoli, V., Thornton, J., Heng, J., Doucet, A.: Diffusion schrödinger bridge with applications to score-based generative modeling. In: NeurIPS (2021)
9. Dong, L., Fan, Q., Guo, Y., Wang, Z., Zhang, Q., Chen, J., Luo, Y., Zou, C.: TSD-SR: One-step diffusion with target score distillation for real-world image super-resolution. In: CVPR (2025)
10. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: NeurIPS (2014)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
12. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: NeurIPS (2020)
13. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-rank adaptation of large language models. In: ICLR (2022)
14. Ke, J., Wang, Q., Wang, Y., Milanfar, P., Yang, F.: MUSIQ: multi-scale image quality transformer. In: ICCV (2021)
15. Li, B., Xue, K., Liu, B., Lai, Y.K.: BBDM: Image-to-image translation with brownian bridge diffusion models. In: CVPR (2023)
16. Li, Y., Zhang, K., Liang, J., Cao, J., Liu, C., Gong, R., Zhang, Y., Tang, H., Liu, Y., Demandolx, D., Ranjan, R., Timofte, R., Van Gool, L.: LSDIR: A large scale dataset for image restoration. In: CVPR Workshops (2023)
17. Lin, X., He, J., Chen, Z., Lyu, Z., Dai, B., Yu, F., Qiao, Y., Ouyang, W., Dong, C.: DiffBIR: Toward blind image restoration with generative diffusion prior. In: ECCV (2024)
18. Lin, X., Yu, F., Hu, J., You, Z., Shi, W., Ren, J.S., Gu, J., Dong, C.: Harnessing diffusion-yielded score priors for image restoration. ACM TOG (2025)
19. Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: ICLR (2023)
20. Liu, G.H., Vahdat, A., Huang, D.A., Theodorou, E.A., Nie, W., Anandkumar, A.: I2SB: image-to-image Schrödinger bridge. In: ICML (2023)
21. Liu, H., Ren, W., Wang, R., Cao, X.: A super-resolution transformer fusion network for single blurred image. Journal of Image and Graphics (2022)
22. Liu, W., Diao, S., Xiao, Z., Chen, Y., Jia, W., Zhao, Y.: Enhancing sparse multi-view super-resolution with unified multi-plane image spatial representation. Visual Intelligence (2026)

23. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. In: ICLR (2023)
24. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: ICLR (2019)
25. Qin, X., Wang, Z., Li, F., Chen, H., Pei, R., Li, W., Cao, X.: CamEdit: Continuous camera parameter control for photorealistic image editing. In: NeurIPS (2025)
26. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR (2022)
27. Shi, Y., De Bortoli, V., Campbell, A., Doucet, A.: Diffusion schrödinger bridge matching. In: NeurIPS (2023)
28. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: ICML (2015)
29. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: ICLR (2021)
30. Sun, L., Wu, R., Ma, Z., Liu, S., Yi, Q., Zhang, L.: Pixel-level and semantic-level adjustable super-resolution: A dual-lora approach. In: CVPR (2025)
31. Wang, J., Chan, K.C.K., Loy, C.C.: Exploring CLIP for assessing the look and feel of images. In: AAAI (2023)
32. Wang, J., Yue, Z., Zhou, S., Chan, K.C., Loy, C.C.: Exploiting diffusion prior for real-world image super-resolution. IJCV (2024)
33. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In: ICCV Workshops (2021)
34. Wei, P., Xie, Z., Lu, H., Zhan, Z., Ye, Q., Zuo, W., Lin, L.: Component divide-and-conquer for real-world image super-resolution. In: ECCV (2020)
35. Wu, R., Sun, L., Ma, Z., Zhang, L.: One-step effective diffusion network for real-world image super-resolution. In: NeurIPS (2024)
36. Wu, R., Yang, T., Sun, L., Zhang, Z., Li, S., Zhang, L.: SeeSR: Towards semantics-aware real-world image super-resolution. In: CVPR (2024)
37. Wu, Z., Sun, Z., Zhou, T., Fu, B., Cong, J., Dong, Y., Zhang, H., Tang, X., Chen, M., Wei, X.: OMGSR: You only need one mid-timestep guidance for real-world image super-resolution. arXiv:2508.08227 (2025)
38. Yang, S., Wu, T., Shi, S., Lao, S., Gong, Y., Cao, M., Wang, J., Yang, Y.: MANIQA: multi-dimension attention network for no-reference image quality assessment. In: CVPR Workshops (2022)
39. Yu, F., Gu, J., Li, Z., Hu, J., Kong, X., Wang, X., He, J., Qiao, Y., Dong, C.: Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In: CVPR (2024)
40. Yue, Z., Liao, K., Loy, C.C.: Arbitrary-steps image super-resolution via diffusion inversion. In: CVPR (2025)
41. Yue, Z., Wang, J., Loy, C.C.: ResShift: Efficient diffusion model for image super-resolution by residual shifting. In: NeurIPS (2023)
42. Zhang, A., Yue, Z., Pei, R., Ren, W., Cao, X.: Degradation-guided one-step image super-resolution with diffusion priors. arXiv:2409.17058 (2024)
43. Zhang, K., Liang, J., Van Gool, L., Timofte, R.: Designing a practical degradation model for deep blind image super-resolution. In: ICCV (2021)
44. Zhang, L., You, W., Shi, K., Gu, S.: Uncertainty-guided perturbation for image super-resolution diffusion model. In: CVPR (2025)
45. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: ICCV (2023)
46. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR (2018)