Allo{SR}^{2}: Rectifying One-Step Super-Resolution to Stay Real via Allomorphic Generative Flows

Zihan Wang[1⋆], Xudong Huang[2⋆], Junbo Qiao[1],
Wei Li[2†], Jie Hu[2], Xinghao Chen[2], and Shaohui Lin[1,3†]

[1] East China Normal University [2] Huawei Foundation Model Dept.
[3] Key Laboratory of Advanced Theory and Application in Statistics and Data Science

Abstract. Real-world image super-resolution (Real-SR) has been revolutionized by leveraging the powerful generative priors from Diffusion Models (DMs) and Flow Matching (FM). However, existing one-step methods typically replace Gaussian noise with degraded low-resolution (LR) latents at initialization, introducing a substantial distribution shift that further leads to trajectory deviation and prior collapse under extreme acceleration. To overcome these limitations, we propose Allo{SR}^{2}, a novel FM-based framework that rectifies one-step SR flows via allomorphic generative flows to maintain high-fidelity generative realism. Specifically, we utilize SNR-Guided Trajectory Initialization to identify a statistically aligned intermediate state along the pre-trained path to integrate LR representations into the generative flow. To ensure a stable, low-curvature path for one-step inference, we propose Flow-Anchored Trajectory Consistency (FATC), which explicitly regularizes the velocity field of the underlying probability flow. Furthermore, we develop Allomorphic Trajectory Matching (ATM), a self-adversarial distillation strategy that jointly models the SR flow and the generative flow within a unified velocity field, enabling one-step Real-SR while preserving the generative prior. Extensive experiments on both synthetic and real-world benchmarks demonstrate that Allo{SR}^{2} achieves state-of-the-art performance in one-step Real-SR, offering a superior balance between fidelity and realism while maintaining extreme efficiency.

Keywords: Super-Resolution · Generative Prior · Allomorphic Flow

1 Introduction

Image super-resolution (SR) [9, 10, 23, 47, 58] serves as a fundamental task in low-level vision, aiming to reconstruct high-resolution (HR) images from their low-resolution (LR) observations. While traditional SR methods [9,10] achieved remarkable fidelity under predefined synthetic degradations (*e.g.*, bicubic downsampling), they frequently produce over-smoothed results in practical scenarios. Consequently, the field has shifted toward real-world SR (Real-SR) [40,58], which tackles complex and unknown degradations present in real-world images.

⋆ Equal contribution. † Corresponding author.

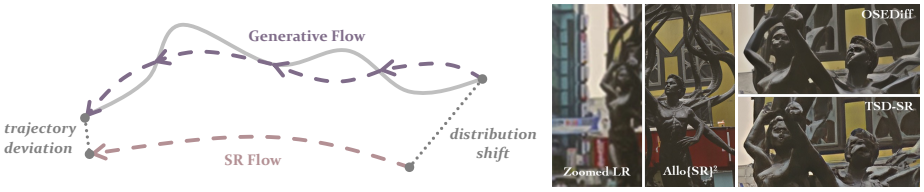

Fig. 1: (Left) Existing methods directly replace Gaussian noise with LR latents, leading to a distribution shift and trajectory deviation from the pre-trained generative path. (Right) Compared to other DM- and FM-based one-step methods like OSEDiff [46] and TSD-SR [11], our approach preserves the generative prior more effectively, recovering realistic textures in severely degraded regions where others suffer from prior collapse.

Early Real-SR methods leveraged Generative Adversarial Networks (GANs) [12, 16, 63] to recover high-frequency details by pushing the reconstructed outputs toward the natural image manifold. However, GAN-based methods [22, 40, 58] are inherently prone to mode collapse and often introduce synthetic artifacts. Recently, Diffusion Models (DMs) [14, 30, 31, 35, 36] and Flow Matching (FM) paradigms [3, 4, 25–27] have revolutionized image generation, demonstrating unprecedented capabilities in synthesizing realistic visuals by progressively transforming Gaussian noise into high-quality images. Driven by this success, an emerging trend in Real-SR adapts these pre-trained DM- and FM-based models as powerful generative priors [24, 39, 47, 52], leveraging their rich, encapsulated knowledge of natural image statistics to reconstruct highly realistic HR details. Nevertheless, their inherent iterative sampling process typically requires tens to hundreds of function evaluations to synthesize a single image, resulting in prohibitive computational latency for practical deployment.

To break this efficiency bottleneck, numerous attempts [11, 37, 42, 46, 49, 55, 57, 59] have been made to compress these continuous-time trajectories into a few steps or even one step. For instance, ResShift [57] and its extension [59] propose replacing Gaussian noise with noised LR images as the initial state of the diffusion process, thereby significantly shortening the required reverse path. SinSR [42] further reformulates this framework as an ordinary differential equation (ODE) and directly distills it into one step. Some other works [49, 55] extend Consistency Models to Real-SR by enforcing consistency along the generative trajectory to enable one-step reconstruction. Beyond these approaches, another active line of research explores Score Distillation [11, 37, 46, 55], which aligns the reconstructed image distribution with the pre-trained generative prior by minimizing the Kullback–Leibler (KL) divergence.

While these few-step and one-step approaches significantly improve efficiency, seamlessly transferring pre-trained generative prior to one-step Real-SR under such extreme acceleration remains a critical challenge. As illustrated in Fig. 1 (left), existing methods typically adapt generative models by directly replacing the initial Gaussian noise with degraded LR latent representations. This naive substitution induces a substantial *distribution shift*, as the latent distribution of

LR images is fundamentally misaligned with the Gaussian prior expected by the pre-trained network. Under one-step inference, this misalignment, compounded by the absence of explicit velocity field constraints, leads to a severe *trajectory deviation*. Consequently, the model suffers from *prior collapse*, where it fails to fully exploit its generative prior, resulting in either over-smoothed outputs or unnatural artifacts, as shown in Fig. 1 (right). Although recent efforts [48, 56] attempt to mitigate this distribution shift via learnable noise injection or VAE encoder fine-tuning to align LR latents with intermediate states of the generative flow, they still treat the underlying generative model merely as a static backbone. We argue that rather than treating the pre-trained network as an isolated backbone, the SR task should integrate it as a dynamic anchor within a unified flow framework. Our key insight is that image generation and restoration share an *allomorphic* nature: they represent structurally analogous trajectories converging toward the same target manifold of realistic HR images. While the generative flow provides an optimal, distortion-free trajectory, the SR flow serves as its localized, condition-constrained counterpart.

Inspired by this observation, we propose a novel FM-based Real-SR method, called **Allo{SR}**^{2}, which rectifies one-step **S**uper-**R**esolution to **S**tay **R**eal via **Allo**morphic generative flows. First, we rethink the injection point of the LR representations from a mathematical standpoint. Through Signal-to-Noise Ratio (SNR) analysis, we demonstrate that the LR latent features exhibit the highest statistical alignment with an intermediate state along the pre-trained generative trajectory. Consequently, this state naturally serves as the optimal initialization for the local SR flow. Instead of standard conditioned fine-tuning, we construct an asymmetric dual-path framework that concurrently models a global generative flow (from noise to HR) and a local SR flow (from LR to HR). By sharing weights and enforcing a cross-trajectory score matching constraint between corresponding timesteps, the pre-trained generative flow serves as a dynamic anchor. It continuously transfers its distortion-free prior to rectify the SR flow, effectively preventing prior collapse and explicitly matching the distribution of SR outputs to the natural image manifold. Remarkably, this structural rectification mitigates over-smoothed reconstructions while minimizing trajectory curvature, flattening the integration path to enable high-fidelity, photorealistic Real-SR in a single function evaluation.

Our main contributions are summarized as follows:

- We propose Allo{SR}^{2}, a novel FM-based framework that rectifies one-step SR flow through the allomorphic nature between generation and restoration.
- We utilize SNR-Guided Trajectory Initialization paired with Flow-Anchored Trajectory Consistency (FATC) to mitigate the distribution shift and trajectory deviation, ensuring a stable, low-curvature SR flow.
- We develop Allomorphic Trajectory Matching (ATM), a self-adversarial distillation strategy that effectively transfers generative prior to the SR flow, preventing prior collapse while preserving perceptual realism.

2 Related Work

2.1 Real-World Image Super-Resolution

Image super-resolution (SR) [9,10,23,47,58] aims to recover high-resolution (HR) details from degraded low-resolution (LR) observations. Early SR methods [9,10] primarily utilize simple synthetic degradations to generate training pairs, which often fail to generalize to complex and stochastic degradations in real world.

To bridge this gap, Generative Adversarial Networks (GANs) [12,16,63] were introduced to perform real-world SR (Real-SR) with more complex degradations. Pioneering work [20] demonstrates that combining adversarial objectives with perceptual losses yields superior visual quality compared to traditional pixel-wise constraints. Building upon this, subsequent research further improves the realism and diversity of synthesized training data through explicit degradation modeling. Specifically, BSRGAN [58] employs randomly shuffled degradation operators to better approximate real-world conditions, while Real-ESRGAN [40] further extends this paradigm via a high-order degradation modeling process. Despite these advancements, GAN-based Real-SR methods still suffer from the instability of adversarial training and the risk of mode collapse, which are highly prone to generating unnatural visual artifacts. Although several refined approaches [22, 41, 62] partially mitigate these issues, the inherent limitations of GANs continue to constrain the synthesis of highly natural and photorealistic details, ultimately bounding their generative capacity.

With the success of Diffusion Models (DMs) [14, 36], recent studies have extensively harnessed generative priors to address the Real-SR task. To leverage the pre-trained Stable Diffusion model [32], StableSR [39] and DiffBIR [24] inject LR representations as conditional guidance via specialized feature modulation, successfully recovering high-frequency details. Beyond low-level conditioning, semantic information has also been incorporated to enhance restoration quality. For example, SeeSR [47] and PASD [52] introduce semantic guidance and pixel-aware cross-attention, respectively, enabling more effective modeling of fine-grained structures and degradation-insensitive features. Meanwhile, ResShift [57] improves sampling efficiency by reformulating the diffusion process in the residual between HR and LR images. Nevertheless, these DM-based methods are fundamentally constrained by the iterative reverse denoising process, resulting in prohibitive inference latency compared to GAN-based methods.

2.2 Acceleration of Diffusion and Flow Matching

Diffusion Models (DMs) [14,30,31,35,36] have achieved remarkable success across various generative tasks by utilizing a parameterized Markov chain to bridge the natural image manifold with the standard Gaussian prior. Building upon this continuous-time formulation, the Flow Matching (FM) paradigm [3,4,25] learns a time-dependent velocity field to establish deterministic probability flows between distributions. Recent variants, such as Rectified Flow [26, 27], further straighten the interpolation trajectories to improve sampling efficiency.

Despite their strong generative performance, the iterative sampling process necessitates tens to hundreds of function evaluations during inference, resulting in substantial computational overhead. This has motivated extensive research on accelerating the sampling process. Early efforts, such as Progressive Distillation [29, 33], iteratively halve the inference steps of student models through a multi-stage distillation. However, the compounding errors across these sequential distillation stages may lead to degraded generation quality in the final student model. Meanwhile, Consistency Models [18,28,34,50] enable one-step or few-step generation by enforcing self-consistency along the sampling trajectory, ensuring that intermediate states map to a shared origin. More recently, Score Distillation approaches [43,53,54] leverage pre-trained diffusion models as implicit score functions to supervise a one-step generator by minimizing the Kullback–Leibler (KL) divergence between the generated distribution and the target distribution. Expanding this paradigm, a recent advancement [7] introduces a dual-flow matching framework that removes the need for auxiliary networks, thereby reducing memory overhead and enabling scalable training on large models.

Driven by these advancements, an emerging line of research seeks to adapt these highly efficient frameworks to the Real-SR task. SinSR [42] reformulates the diffusion trajectory into a deterministic sampling process, distilling the inference process into a single step. Alternatively, OSEDiff [46] leverages Variational Score Distillation (VSD) [43,54] to inherit the generative capacity of pre-trained multi-step prior within a one-step generator. To mitigate the artifacts and instability caused by VSD during early training stages, TSD-SR [11] introduces Target Score Distillation (TSD), which improves optimization stability and leads to more robust convergence. Beyond Score Distillation, CTMSR [55] formulates a deterministic PF-ODE mapping from noisy LR to HR images and employs consistency training combined with a distribution trajectory matching loss to achieve high-fidelity one-step generation.

3 Preliminaries

3.1 Flow Matching

Flow Matching provides an intuitive paradigm for generative modeling, aiming to learn a time-dependent velocity field that defines a probability path between a complex target distribution and a simple prior (*e.g.*, a standard Gaussian). Formally, given a real data sample $\boldsymbol{x} \sim p_{\text{data}}(\boldsymbol{x})$ and a noise sample $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, the probability path is typically parameterized as:

$$\boldsymbol{x}_t = a_t \boldsymbol{x} + b_t \boldsymbol{\epsilon}, \quad t \in [0, 1] \tag{1}$$

where a_t and b_t are predefined scalar schedules satisfying boundary conditions $a_0 = 1, b_0 = 0$ and $a_1 = 0, b_1 = 1$. The corresponding velocity field $\boldsymbol{v}_t$ is defined as the time derivative of the path:

$$\boldsymbol{v}_t = \frac{\mathrm{d}\boldsymbol{x}_t}{\mathrm{d}t} = a_t' \boldsymbol{x} + b_t' \boldsymbol{\epsilon}. \tag{2}$$

This represents the conditional velocity, denoted as $\boldsymbol{v}_t(\boldsymbol{x}_t|\boldsymbol{x})$. In particular, adopting the linear schedule $a_t = 1 - t$ and $b_t = t$ yields a straight-line interpolation between $\boldsymbol{x}$ and $\boldsymbol{\epsilon}$. Under this parameterization, the conditional velocity reduces to the time-independent form $\boldsymbol{v}_t = \boldsymbol{\epsilon} - \boldsymbol{x}$.

Since an intermediate state $\boldsymbol{x}_t$ does not uniquely identify its originating data sample $\boldsymbol{x}$, Flow Matching essentially models the expectation over all possibilities, known as the marginal velocity $\boldsymbol{v}(\boldsymbol{x}_t, t) = \mathbb{E}[\boldsymbol{v}_t(\boldsymbol{x}_t|\boldsymbol{x})|\boldsymbol{x}_t]$. Directly optimizing this marginal velocity is intractable due to the unknown data density $p_{\text{data}}(\boldsymbol{x})$. Conditional Flow Matching (CFM) [25] avoids this difficulty by minimizing a tractable surrogate loss between a neural network $\boldsymbol{v}_\theta$ and the conditional velocity:

$$\mathcal{L}_{\text{CFM}} = \mathbb{E}\left[\left\|\boldsymbol{v}_\theta(\boldsymbol{x}_t, t) - \boldsymbol{v}_t(\boldsymbol{x}_t|\boldsymbol{x})\right\|^2\right]. \tag{3}$$

Crucially, minimizing $\mathcal{L}_{\text{CFM}}$ is theoretically equivalent to matching the intractable marginal velocity field $\boldsymbol{v}(\boldsymbol{x}_t, t)$, thereby enabling $\boldsymbol{v}_\theta$ to accurately govern the distribution transformation during inference.

3.2 Score Distillation

To alleviate the costly iterative sampling of DM- and FM-based models, Score Distillation has emerged as a widely adopted paradigm for one-step or few-step acceleration. The core principle is to distill the rich distributional knowledge from a pre-trained multi-step teacher model into a one-step generator G_θ.

Mathematically, let $p_\theta(\boldsymbol{x})$ denote the distribution generated by the one-step generator, and $p_{\text{data}}(\boldsymbol{x})$ denote the target data distribution captured by the multi-step teacher model. Recent Score Distillation methods, such as Distribution Matching Distillation (DMD) [53, 54], seek to align these two distributions by minimizing the Kullback-Leibler (KL) divergence:

$$\min_\theta \mathcal{D}_{\text{KL}}\left(p_\theta(\boldsymbol{x}) \| p_{\text{data}}(\boldsymbol{x})\right). \tag{4}$$

Directly evaluating this objective requires access to the underlying probability densities, which are generally unavailable in high-dimensional spaces. Fortunately, optimizing the generator only requires the gradient of the divergence with respect to the generator parameters θ:

$$\nabla_\theta \mathcal{D}_{\text{KL}} = \mathbb{E}\left[\left(\nabla_{\boldsymbol{x}} \log p_\theta(\boldsymbol{x}) - \nabla_{\boldsymbol{x}} \log p_{\text{data}}(\boldsymbol{x})\right) \frac{\partial G_\theta}{\partial \theta}\right]. \tag{5}$$

Here, $\nabla_{\boldsymbol{x}} \log p_{\text{data}}(\boldsymbol{x})$ is the score of the target distribution, which is typically approximated using the teacher model pre-trained on real data. Meanwhile, $\nabla_{\boldsymbol{x}} \log p_\theta(\boldsymbol{x})$ represents the score of the generated distribution and is estimated by an auxiliary network trained on fake samples produced by the generator G_θ. By minimizing the discrepancy between these two score functions, Score Distillation aligns the generated distribution with the target distribution directly. Consequently, the student generator can inherit the teacher's generative capability without explicitly reproducing the multi-step sampling trajectory, enabling high-quality generation in a single forward pass.

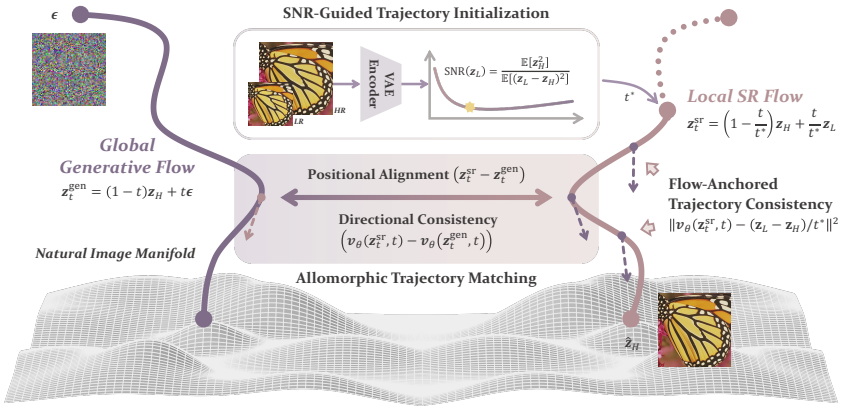

Fig. 2: Overview of Allo{SR}^{2}. (a) SNR-Guided Trajectory Initialization identifies the optimal initializing timestep to mitigate the distribution shift between LR latents and the Gaussian prior. (b) Flow-Anchored Trajectory Consistency (FATC) enforces a low-curvature SR flow via supervision on the instantaneous velocity. (c) Allomorphic Trajectory Matching (ATM) matches the scores from the local SR flow and the global generative flow, ensuring the SR outputs stay within the natural image manifold.

4 Methodology

In this section, we present Allo{SR}^{2}, as illustrated in Fig. 2. Our framework is motivated by the allomorphic nature between generation and restoration, where the SR flow can be viewed as a localized, condition-constrained counterpart of the generative flow. We first identify the optimal embedding timestep via SNR analysis to integrate LR representations into the pre-trained generative flow (Sec. 4.1). To mitigate the trajectory deviation, we introduce a Flow-Anchored Trajectory Consistency (FATC) to regularize the velocity field of the underlying probability flow (Sec. 4.2). Furthermore, we propose Allomorphic Trajectory Matching (ATM) to align the local SR flow with the global generative flow, thereby enabling efficient one-step SR while preserving the generative prior (Sec. 4.3).

4.1 SNR-Guided Trajectory Initialization

In standard image generation, a continuous-time trajectory transports a noise sample $\boldsymbol{\epsilon}$ to the HR latent $\boldsymbol{z}_H$. In contrast, the LR latent $\boldsymbol{z}_L$, as a degraded state of $\boldsymbol{z}_H$, exhibits a substantial distribution shift from the Gaussian prior. Directly replacing the initial noise with $\boldsymbol{z}_L$ may disrupt the pre-trained generative prior. To reconcile this, we interpret $\boldsymbol{z}_L$ as an intermediate state composed of the HR latent $\boldsymbol{z}_H$ and the degradation residual $\Delta\boldsymbol{z} = \boldsymbol{z}_L - \boldsymbol{z}_H$. This motivates identifying an optimal embedding timestep where the theoretical noise level of the pre-trained trajectory best matches the degradation severity of $\boldsymbol{z}_L$.

Following recent advances [48], we adopt the Signal-to-Noise Ratio (SNR) as a metric to evaluate latent similarity, as it naturally connects the physical degradation of the image to the decreasing noise schedule of the generative trajectory. Given a dataset of N paired LR-HR latents $\{(\boldsymbol{z}_L^{(i)}, \boldsymbol{z}_H^{(i)})\}_{i=1}^N$ encoded by the fixed pre-trained VAE encoder, we formulate the degraded SNR as the ratio of the signal variance to the residual variance. The optimal embedding timestep is then determined by searching over the schedule $t \in \{1, \dots, T\}$ to minimize the discrepancy between the theoretical trajectory SNR and the degraded SNR:

$$t^* = \arg\min_t \frac{1}{N} \sum_{i=1}^{N} \left| \frac{(1-t)^2 \cdot \mathbb{E}[\|\boldsymbol{z}_H^{(i)}\|^2]}{t^2} - \frac{\mathbb{E}[\|\boldsymbol{z}_H^{(i)}\|^2]}{\mathbb{E}[\|\boldsymbol{z}_L^{(i)} - \boldsymbol{z}_H^{(i)}\|^2]} \right|. \tag{6}$$

By explicitly initializing the SR trajectory at t^*, we establish a physically grounded starting state that directly bypasses the highly noisy integration region $t \in (t^*, 1]$. Crucially, this initialization forces the subsequent SR flow to closely adhere to the original generative velocity field, thereby preventing catastrophic forgetting of the pre-trained prior during fine-tuning.

4.2 Flow-Anchored Trajectory Consistency

Once the optimal initial timestep t^* is determined, our framework can reconstruct HR latent in a single functional evaluation. Specifically, we formulate the one-step reverse integration as a first-order Euler approximation:

$$\hat{\boldsymbol{z}}_H = \boldsymbol{z}_L - t^* \boldsymbol{v}_\theta(\boldsymbol{z}_L, t^*), \tag{7}$$

where $\boldsymbol{v}_\theta(\boldsymbol{z}_L, t^*)$ denotes the predicted velocity field at the starting state. To supervise the training process, we employ a reconstruction loss that integrates a latent-wise MSE loss and a pixel-wise perceptual loss based on LPIPS:

$$\mathcal{L}_{\text{Rec}} = \mathcal{L}_{\text{MSE}}(\hat{\boldsymbol{z}}_H, \boldsymbol{z}_H) + \lambda \mathcal{L}_{\text{LPIPS}}(D(\hat{\boldsymbol{z}}_H), D(\boldsymbol{z}_H)). \tag{8}$$

Although Eq. 7 facilitates a direct mapping, training solely on endpoint objectives often forces the network to fit a highly complex, non-linear function via a shortcut. In the absence of explicit supervision on the instantaneous velocity field of the underlying probability flow, the learned trajectory is prone to significant deviation from the optimal generative path. To mitigate this, we introduce Flow-Anchored Trajectory Consistency (FATC) by interpolating between $\boldsymbol{z}_H$ and $\boldsymbol{z}_L$. We define the rescaled intermediate state $\boldsymbol{z}_t^{\text{sr}}$ as:

$$\boldsymbol{z}_t^{\text{sr}} = \left(1 - \frac{t}{t^*}\right) \boldsymbol{z}_H + \frac{t}{t^*} \boldsymbol{z}_L, \quad t \in [0, t^*]. \tag{9}$$

Since $\boldsymbol{z}_L$ represents the state specifically at t^*, the interpolation schedule is normalized by t^* to ensure temporal alignment with the pre-trained flow. During

training, we anchor the network to the underlying flow by minimizing the velocity discrepancy across these intermediate states:

$$\mathcal{L}_{\text{FATC}} = \mathbb{E}\left\| v_\theta(\boldsymbol{z}_t^{\text{sr}}, t) - \frac{\boldsymbol{z}_L - \boldsymbol{z}_H}{t^*} \right\|^2. \tag{10}$$

By enforcing this velocity-level supervision, the SR model adheres to a low-curvature, linear trajectory. This consistency maximizes the fidelity of the one-step approximation and ensures the SR flow remains within the support of the pre-trained probability density.

4.3 Allomorphic Trajectory Matching

While the FATC in Eq. 10 regularizes the velocity field, fine-tuning pre-trained generative model on limited LR-HR data pairs often precipitates prior collapse. In this regime, the model sacrifices its generative richness to overfit the specific degradations of the training pairs. To circumvent this, we propose Allomorphic Trajectory Matching (ATM). Rather than adapting the generative prior to the SR task, ATM aligns the SR trajectory with the intrinsic generative flow.

Inspired by the self-adversarial framework of TwinFlow [7], we construct two allomorphic, parameter-shared trajectories within a unified velocity field: a generative flow and an SR flow. By sharing weights across both tasks, the model is forced to learn a unified latent representation where SR transitions are perceived as a subset of the broader generative probability flow. To ensure the SR results remain within the support of the natural image manifold, we treat the generative flow as a dynamic positive guidance. We minimize the Kullback-Leibler (KL) divergence between the distribution of reconstructed SR latents p_{sr} and the prior distribution p_{gen} maintained by the generative flow:

$$\mathcal{L}_{\text{ATM}} = \mathcal{D}_{\text{KL}}\left(p_{\text{sr}} \parallel p_{\text{gen}}\right) = \mathbb{E}\left[\log p_{\text{sr}}(\hat{\boldsymbol{z}}_H) - \log p_{\text{gen}}(\hat{\boldsymbol{z}}_H)\right]. \tag{11}$$

Following the Score Distillation paradigm, the gradient of $\mathcal{L}_{ATM}$ is approximated via the score discrepancy between the allomorphic paths:

$$\begin{aligned} \nabla_\theta \mathcal{L}_{\text{ATM}} &= \mathbb{E}\left[\left(\nabla_{\hat{\boldsymbol{z}}_H} \log p_{\text{sr}}(\hat{\boldsymbol{z}}_H) - \nabla_{\hat{\boldsymbol{z}}_H} \log p_{\text{gen}}(\hat{\boldsymbol{z}}_H) \right) \frac{\partial \hat{\boldsymbol{z}}_H}{\partial \theta} \right] \\ &= \mathbb{E}\left[w(t) \left[\underbrace{(\boldsymbol{z}_t^{\text{sr}} - \boldsymbol{z}_t^{\text{gen}})}_{\text{Position}} - \underbrace{t\left(v_\theta(\boldsymbol{z}_t^{\text{sr}}, t) - v_\theta(\boldsymbol{z}_t^{\text{gen}}, t)\right)}_{\text{Direction}} \right] \frac{\partial \hat{\boldsymbol{z}}_H}{\partial \theta} \right], \end{aligned} \tag{12}$$

where $w(t)$ is a time-dependent weighting function. since the SR schedule is rescaled by t^* in Eq. 10, the intermediate states $\boldsymbol{z}_t^{\text{sr}}$ and $\boldsymbol{z}_t^{\text{gen}}$ exhibit theoretical proximity. As derived in Eq. 12, the first term performs positional alignment by minimizing the distance between allomorphic states, while the second term ensures directional consistency. By minimizing $\mathcal{L}_{\text{ATM}}$, we achieve a better alignment between SR outputs and the natural image distribution.

This derivation demonstrates that by matching the scores of the SR flow to the generative flow, we effectively perform self-adversarial alignment. The SR task is no longer a simple regression; it becomes a specialized navigation within the generative manifold. Consequently, the shared weights learn to bridge the degradation gap while preserving global generative capabilities, achieving a superior balance between fidelity and realism.

5 Experiments

5.1 Experimental Settings

Training Datasets Following [46, 47], we train our model on LSDIR [21] and the first 10K face images from FFHQ [16], totaling 95K images. The degradation pipeline of Real-ESRGAN [40] is used to synthesize LR-HR training pairs.

Testing Datasets We evaluate our method on a synthetic dataset DIV2K-Val [1] and three real-world datasets, including RealSR [6], DRealSR [45] and RealLQ250 [2]. We crop 512×512 patches from DIV2K-Val and apply the Real-ESRGAN degradation pipeline to downsample them to 128×128. For RealSR and DRealSR, which provide paired GT images, we apply center-cropping to the LR images to a resolution of 128×128. Since RealLQ250 lacks corresponding GT images, we perform no cropping and maintain the original 256×256 resolution. All experiments are conducted with the scaling factor of $\times 4$.

Compared Methods We compare Allo{SR}^{2} with other DM- and FM-based models, including multi-step models: StableSR [39], SeeSR [47], ResShift [57]; and one-step models: SinSR [42], OSEDiff [46], TSD-SR [11], and CTMSR [55].

Evaluation Metrics To comprehensively evaluate the performance of all methods, we employ a series of full-reference (FR) and no-reference (NR) metrics. For FR evaluation, PSNR and SSIM [44] reflect fidelity, while LPIPS [61] and DISTS [8] measure perceptual quality. FID [13] further evaluates the distance of distributions between GT and reconstructed images. For NR evaluation, we adopt four blind image quality assessment metrics, namely NIQE [60], MUSIQ [17], MANIQA [51], and CLIPIQA [38], to evaluate visual naturalness and structural integrity of the images.

Implementation Details We utilize the pre-trained weights of FLUX.1-dev[1] as our foundation generative model. To facilitate efficient parameter adaptation, LoRA [15] is integrated into both the VAE encoder and the Diffusion Transformer (DiT) components with a rank of $r = 64$. The framework is optimized using the AdamW optimizer [19] with a learning rate of 5×10^{-5} and a global batch size of 16, for a total of 10K iterations.

[1] `https://github.com/black-forest-labs/flux`

Table 1: Quantitative comparison among the state-of-the-art DM- and FM-based Real-SR methods on both synthetic and real-world datasets. The best and second-best results are highlighted in **bold** and underline, respectively.

Methods	StableSR [39]	SeeSR [47]	ResShift [57]	SinSR [42]	OSEDiff [46]	TSD-SR [11]	CTMSR [55]	Allo{SR}^{2}
Steps	200	50	15	1	1	1	1	1
DIV2K-Val								
PSNR↑	23.31	23.71	24.69	24.43	23.72	23.02	**24.87**	22.98
SSIM↑	0.5728	0.6045	0.6175	0.6012	0.6108	0.5808	**0.6349**	0.6123
LPIPS↓	0.3129	0.3207	0.3374	0.3262	0.2941	**0.2673**	0.3011	0.2854
DISTS↓	0.2138	0.1967	0.2215	0.2066	0.1976	**0.1821**	0.2102	0.1964
FID↓	**24.67**	25.83	36.01	35.45	26.32	29.16	25.07	24.94
NIQE↓	4.76	4.82	6.82	6.02	4.71	4.32	5.3036	**4.18**
MUSIQ↑	65.63	68.49	60.92	62.80	67.97	71.69	66.59	**72.00**
MANIQA↑	0.6188	0.6239	0.5450	0.5395	0.6148	0.6192	0.5146	**0.6432**
CLIPIQA↑	0.6682	0.6857	0.6089	0.6499	0.6683	**0.7416**	0.6602	0.6837
RealSR								
PSNR↑	24.69	25.33	**26.31**	26.30	25.15	24.81	26.00	24.97
SSIM↑	0.7052	0.7273	0.7411	0.7354	0.7341	0.7172	**0.7549**	0.7432
LPIPS↓	0.3091	0.2985	0.3489	0.3212	0.2921	**0.2743**	0.2905	0.2755
DISTS↓	0.2167	0.2213	0.2498	0.2346	0.2128	**0.2104**	0.2204	0.2162
FID↓	127.20	125.66	142.81	137.05	123.49	114.45	134.97	**112.88**
NIQE↓	5.76	5.38	7.27	6.31	5.65	5.13	5.60	**5.08**
MUSIQ↑	65.42	69.37	58.10	60.41	69.09	**71.19**	64.52	70.11
MANIQA↑	0.6211	0.6439	0.5305	0.5389	0.6326	0.6347	0.5269	**0.6725**
CLIPIQA↑	0.6195	0.6594	0.5450	0.6204	0.6693	**0.7160**	0.6391	0.6750
DRealSR								
PSNR↑	28.04	28.26	28.45	28.41	27.92	27.77	**28.68**	27.95
SSIM↑	0.7460	0.7698	0.7632	0.7495	**0.7835**	0.7559	0.7830	0.7694
LPIPS↓	0.3354	0.3197	0.4073	0.3741	0.2968	**0.2967**	0.3242	0.2989
DISTS↓	0.2287	0.2306	0.2700	0.2488	0.2165	0.2136	0.2360	**0.2128**
FID↓	147.03	149.86	175.92	177.05	135.30	134.98	161.24	**132.48**
NIQE↓	6.51	6.52	8.28	7.02	6.49	5.91	6.21	**5.82**
MUSIQ↑	58.50	64.84	49.86	55.34	64.65	66.62	59.93	**66.95**
MANIQA↑	0.5602	0.6026	0.4573	0.4898	0.5899	0.5874	0.4866	**0.6264**
CLIPIQA↑	0.6171	0.6672	0.5259	0.6367	0.6963	**0.7344**	0.6451	0.6883
RealLQ250								
NIQE↓	4.62	4.44	5.33	5.82	3.97	**3.49**	4.58	3.59
MUSIQ↑	63.52	59.27	57.53	63.73	69.55	72.09	68.00	**72.51**
MANIQA↑	0.6733	**0.7037**	0.6572	0.5161	0.5782	0.5829	0.5078	0.6303
CLIPIQA↑	0.5160	0.7013	0.6129	0.6990	0.6725	**0.7221**	0.6706	0.7055

5.2 Comparison Results

Quantitative Comparisons. As shown in Tab. 1, Allo{SR}^{2} achieves an excellent balance between fidelity and realism across both multi-step and one-step methods. It achieves the best performance on most NR metrics across all four benchmarks. In particular, it significantly outperforms other one-step methods on MANIQA, which highly correlates with human visual preference. Moreover, the superior FID demonstrates that the proposed ATM effectively aligns the restored distribution with the natural image manifold. Although Allo{SR}^{2} does not achieve the highest scores on PSNR and SSIM, this is consistent with the well-known perception-distortion trade-off in Real-SR [5]. Since these two metrics emphasize pixel-wise correspondence, they tend to penalize realistic tex-

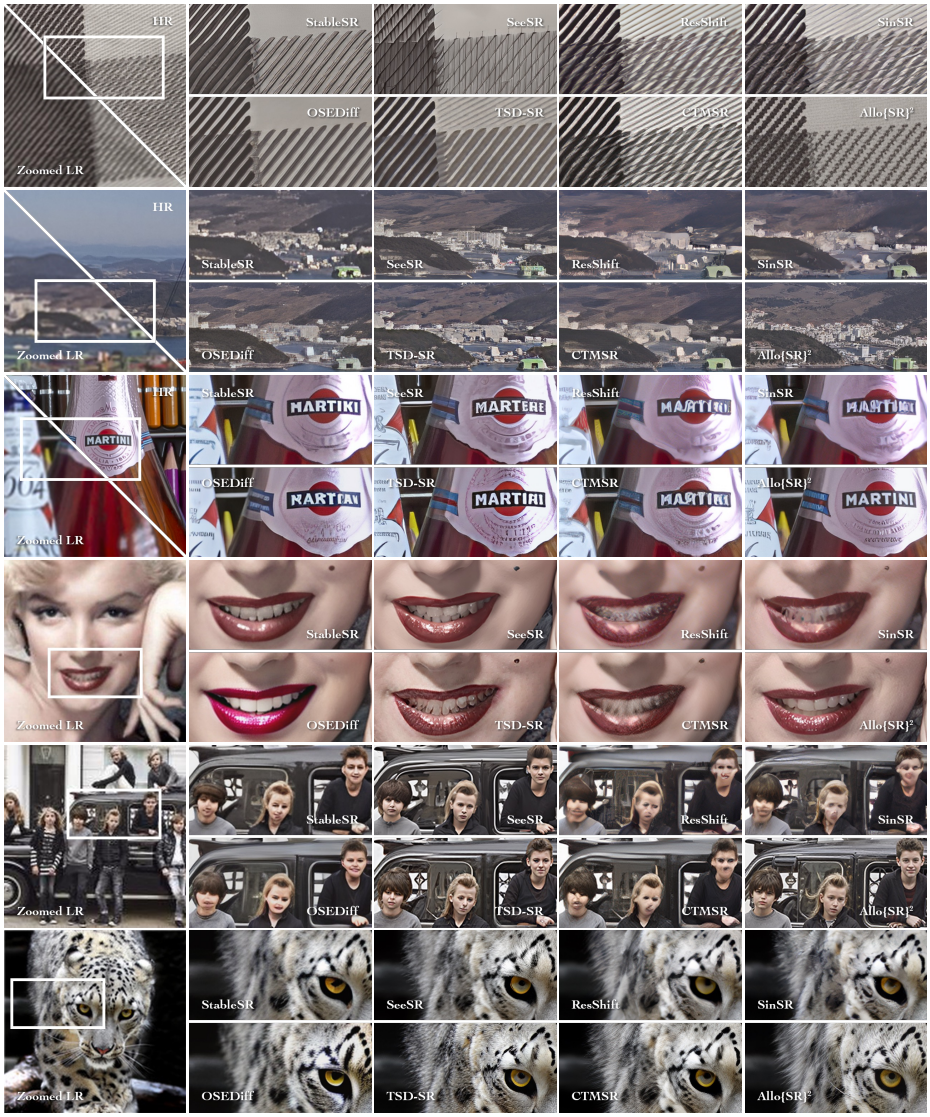

Fig. 3: Qualitative comparisons of different DM- and FM-based Real-SR methods.

tures while favoring over-smoothed reconstructions. Overall, the advantages of Allo{SR}^{2} manifest in two key dimensions. Regarding efficiency, our method achieves high-quality one-step restoration that reduces the number of function evaluations by 15× to 200× compared with representative multi-step methods. Regarding effectiveness, it consistently outperforms existing one-step methods across most metrics, demonstrating that rectifying the SR trajectory via our allomorphic design effectively preserves the generative prior.

Qualitative Comparisons. Fig. 3 presents visual comparisons on several challenging cases. Our method consistently produces more faithful and realistic restorations than existing one-step approaches, even surpassing some multi-step approaches. Specifically, it effectively recovers fine-grained textures and structural details, including the intricate needlework and thread patterns on fabrics as well as the sharp contours of buildings, avoiding over-smoothed results observed in competing methods. For text restoration, although pre-trained generative models inherently possess strong text-synthesis capability, existing one-step Real-SR methods often suffer from prior collapse during task-specific fine-tuning, resulting in truncated or distorted characters. In contrast, Allo{SR}^{2} faithfully reconstructs the complete "MARTINI" text with clear boundaries, demonstrating superior preservation of the pre-trained generative prior. On face restoration examples, our method accurately restores natural facial expressions and realistic dental structures. However, many baseline methods introduce hallucinations or artifacts due to the misalignment between LR latents and generative prior. Moreover, Allo{SR}^{2} successfully synthesizes high-frequency fur textures with sharp yet natural appearances, indicating the efficacy of our method in capturing complex natural image statistics.

5.3 Ablation Study

To validate the effectiveness of the core components in Allo{SR}^{2}, we conduct ablation studies on RealSR, DRealSR, and RealLQ250. Specifically, we investigate the impact of the selected initial timestep, the proposed training losses, and the resolution scaling. The quantitative results are summarized in Tab. 2.

Justification of the Initial Timestep. Most existing one-step methods employ a heuristic initial timestep during both training and inference, *e.g.*, $t = 999$ in OSEDiff [46] and $t = 1$ in PiSA-SR [37]. However, these choices are suboptimal for Real-SR. Specifically, within pre-trained generative prior, $t = 999$ corresponds to Gaussian noise, whereas $t = 1$ corresponds to a latent that is already close to the clean image. Neither matches the latent distribution of LR images, resulting in an unnecessary distribution shift at initialization. To validate the SNR-Guided Trajectory Initialization, we compare the selected timestep t^* against these two heuristic choices. The selected t^* consistently achieves the best overall performance across both FR and NR metrics, confirming that a statistically aligned state provides an optimal initialization for one-step Real-SR. Furthermore, by reducing the distribution shift, t^* accelerates training convergence, requiring only 7k iterations compared with 9k iterations for other settings.

Effectiveness of the Training Losses. The proposed training objectives play complementary roles in optimizing the one-step SR trajectory. $\mathcal{L}_{\text{FATC}}$ linearizes the SR flow, enabling a one-step Euler integration to better approximate the underlying probability flow, while $\mathcal{L}_{\text{ATM}}$ aligns the SR flow with the allomorphic generative flow through self-adversarial distillation. Ablating $\mathcal{L}_{\text{FATC}}$ introduces

Table 2: Ablation studies of initial timestep, training losses, and resolution scaling.

Datasets	Settings	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	FID↓	NIQE↓	MUSIQ↑	MANIQA↑	CLIPIQA↑
RealSR	Selected t^*	24.97	0.7432	0.2755	0.2162	112.88	5.08	70.11	0.6725	0.6750
	$t = 999$	24.16	0.7385	0.2869	0.2201	113.62	5.14	69.75	0.6713	0.6708
	$t = 1$	24.74	0.7434	0.2741	0.2156	112.89	5.27	68.90	0.6618	0.6681
DRealSR	Full	27.95	0.7694	0.2989	0.2128	132.48	5.82	66.95	0.6264	0.6883
	w/o $\mathcal{L}_{\text{FATC}}$	27.88	0.7641	0.3106	0.2105	121.75	6.04	63.49	0.6192	0.6357
	w/o $\mathcal{L}_{\text{ATM}}$	28.68	0.7717	0.3008	0.2003	139.74	6.19	62.77	0.6049	0.6147
RealLQ250	512×512	–	–	–	–	–	3.59	72.51	0.6303	0.7055
	$1k \times 1k$	–	–	–	–	–	3.52	72.93	0.6497	0.7105

severe discretization errors during one-step inference, leading to degraded perceptual quality. Removing $\mathcal{L}_{\text{ATM}}$ reveals a classic perception-distortion trade-off: while pixel-wise fidelity metrics like PSNR and SSIM improve, the perceptual metrics deteriorate substantially. This phenomenon confirms that without the guidance of the allomorphic generative flow, the model tends to overfit the specific degradations of the training pairs, yielding over-smoothed results devoid of rich generative details. The combination of $\mathcal{L}_{\text{FATC}}$ and $\mathcal{L}_{\text{ATM}}$ allows Allo{SR}^{2} to achieve a superior perception-distortion trade-off.

Scaling to Higher Resolutions. Following previous one-step Real-SR methods, we train and evaluate Allo{SR}^{2} at a resolution of 512×512 to ensure a fair comparison in Tab. 1. For the 1024×1024 images in RealLQ250, we employ a tiled inference strategy. Since the pre-trained FLUX model is originally trained at a resolution of 1024×1024, we conduct an additional experiment by training and evaluating Allo{SR}^{2} at its native resolution. The 1024×1024 setting consistently improves all metrics over the 512×512 setting. This validates that Allo{SR}^{2} effectively benefits from higher-resolution training and scales well to the native resolution of the pre-trained generative prior.

6 Conclusion

In this paper, we present Allo{SR}^{2}, a novel FM-based framework that redefines one-step Real-SR as a specialized navigation within a generative manifold. Inspired by the allomorphic nature between generation and restoration, our approach jointly models them within a unified flow framework. Specifically, SNR-Guided Trajectory Initialization reduces the distribution shift between LR latents and the Gaussian prior, while Flow-Anchored Trajectory Consistency (FATC) and Allomorphic Trajectory Matching (ATM) jointly mitigate trajectory deviation and preserve the generative prior under extreme acceleration. Extensive experiments demonstrate that Allo{SR}^{2} achieves a superior balance between fidelity, realism, and efficiency, offering a promising direction toward efficient, high-quality Real-SR.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 62572193), the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science, Ministry of Education, and the Fundamental Research Funds for the Central Universities. We also appreciate Ashin and his brand STAYREAL for inspiring the name of our method.

References

1. Agustsson, E., Timofte, R.: NTIRE 2017 challenge on single image super-resolution: Dataset and study. In: CVPRW (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.S., 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. Cheng, Z., Sun, P., Li, J., Lin, T.: Twinflow: Realizing one-step generation on large models with self-adversarial flows. In: ICLR (2026)
8. Ding, K., Ma, K., Wang, S., Simoncelli, E.P.: Image quality assessment: Unifying structure and texture similarity. TPAMI (2022)
9. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: ECCV (2014)
10. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. TPAMI (2016)
11. 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)
12. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., Bengio, Y.: Generative adversarial nets. In: NeurIPS (2014)
13. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: NeurIPS (2017)
14. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: NeurIPS (2020)
15. 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)
16. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: CVPR (2019)
17. Ke, J., Wang, Q., Wang, Y., Milanfar, P., Yang, F.: MUSIQ: multi-scale image quality transformer. In: ICCV (2021)
18. Kim, D., Lai, C., Liao, W., Murata, N., Takida, Y., Uesaka, T., He, Y., Mitsufuji, Y., Ermon, S.: Consistency trajectory models: Learning probability flow ODE trajectory of diffusion. In: ICLR (2024)

19. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2015)
20. Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A.P., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network. In: CVPR (2017)
21. Li, Y., Zhang, K., Liang, J., Cao, J., Liu, C., Gong, R., Zhang, Y., Tang, H., Liu, Y., Demandolx, D., Ranjan, R., Timofte, R., Gool, L.V.: LSDIR: A large scale dataset for image restoration. In: CVPRW (2023)
22. Liang, J., Zeng, H., Zhang, L.: Details or artifacts: A locally discriminative learning approach to realistic image super-resolution. In: CVPR (2022)
23. Liang, J., Cao, J., Sun, G., Zhang, K., Gool, L.V., Timofte, R.: Swinir: Image restoration using swin transformer. In: ICCVW (2021)
24. 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)
25. Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: ICLR (2023)
26. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. In: ICLR (2023)
27. Liu, X., Zhang, X., Ma, J., Peng, J., Liu, Q.: Instaflow: One step is enough for high-quality diffusion-based text-to-image generation. In: ICLR (2024)
28. Luo, S., Tan, Y., Huang, L., Li, J., Zhao, H.: Latent consistency models: Synthesizing high-resolution images with few-step inference. arXiv:2310.04378 (2023)
29. Meng, C., Rombach, R., Gao, R., Kingma, D.P., Ermon, S., Ho, J., Salimans, T.: On distillation of guided diffusion models. In: CVPR (2023)
30. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: ICML (2021)
31. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: ICCV (2023)
32. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR (2022)
33. Salimans, T., Ho, J.: Progressive distillation for fast sampling of diffusion models. In: ICLR (2022)
34. Song, Y., Dhariwal, P., Chen, M., Sutskever, I.: Consistency models. In: ICML (2023)
35. Song, Y., Ermon, S.: Generative modeling by estimating gradients of the data distribution. In: NeurIPS (2019)
36. 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)
37. 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)
38. Wang, J., Chan, K.C.K., Loy, C.C.: Exploring CLIP for assessing the look and feel of images. In: AAAI (2023)
39. Wang, J., Yue, Z., Zhou, S., Chan, K.C.K., Loy, C.C.: Exploiting diffusion prior for real-world image super-resolution. IJCV (2024)
40. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: ICCVW (2021)
41. Wang, X., Yu, K., Dong, C., Tang, X., Loy, C.C.: Deep network interpolation for continuous imagery effect transition. In: CVPR (2019)

42. Wang, Y., Yang, W., Chen, X., Wang, Y., Guo, L., Chau, L., Liu, Z., Qiao, Y., Kot, A.C., Wen, B.: Sinsr: Diffusion-based image super-resolution in a single step. In: CVPR (2024)
43. Wang, Z., Lu, C., Wang, Y., Bao, F., Li, C., Su, H., Zhu, J.: Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. In: NeurIPS (2023)
44. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. TIP (2004)
45. 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)
46. Wu, R., Sun, L., Ma, Z., Zhang, L.: One-step effective diffusion network for real-world image super-resolution. In: NeurIPS (2024)
47. Wu, R., Yang, T., Sun, L., Zhang, Z., Li, S., Zhang, L.: Seesr: Towards semantics-aware real-world image super-resolution. In: CVPR (2024)
48. 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)
49. Xu, J., Li, W., Sun, H., Li, F., Wang, Z., Peng, L., Ren, J., Yang, H., Hu, X., Pei, R., Heng, P.: Fast image super-resolution via consistency rectified flow. In: ICCV (2025)
50. Yang, L., Zhang, Z., Zhang, Z., Liu, X., Xu, M., Zhang, W., Meng, C., Ermon, S., Cui, B.: Consistency flow matching: Defining straight flows with velocity consistency. arXiv:2407.02398 (2024)
51. 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: CVPRW (2022)
52. Yang, T., Wu, R., Ren, P., Xie, X., Zhang, L.: Pixel-aware stable diffusion for realistic image super-resolution and personalized stylization. In: ECCV (2024)
53. Yin, T., Gharbi, M., Park, T., Zhang, R., Shechtman, E., Durand, F., Freeman, W.: Improved distribution matching distillation for fast image synthesis. In: NeurIPS (2024)
54. Yin, T., Gharbi, M., Zhang, R., Shechtman, E., Durand, F., Freeman, W.T., Park, T.: One-step diffusion with distribution matching distillation. In: CVPR (2024)
55. You, W., Zhang, M., Zhang, L., Zhou, X., Shi, K., Gu, S.: Consistency trajectory matching for one-step generative super-resolution. In: ICCV (2025)
56. Yue, Z., Liao, K., Loy, C.C.: Arbitrary-steps image super-resolution via diffusion inversion. In: CVPR (2025)
57. Yue, Z., Wang, J., Loy, C.C.: Resshift: Efficient diffusion model for image super-resolution by residual shifting. In: NeurIPS (2023)
58. Zhang, K., Liang, J., Gool, L.V., Timofte, R.: Designing a practical degradation model for deep blind image super-resolution. In: ICCV (2021)
59. Zhang, L., You, W., Shi, K., Gu, S.: Uncertainty-guided perturbation for image super-resolution diffusion model. In: CVPR (2025)
60. Zhang, L., Zhang, L., Bovik, A.C.: A feature-enriched completely blind image quality evaluator. TIP (2015)
61. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR (2018)
62. Zhang, Y., Ji, B., Hao, J., Yao, A.: Perception-distortion balanced ADMM optimization for single-image super-resolution. In: ECCV (2022)
63. Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV (2017)