This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Fast Image Super-Resolution via Consistency Rectified Flow

Jiaqi Xu[1,2*], Wenbo Li[2†], Haoze Sun[2], Fan Li[2], Zhixin Wang[2], Long Peng[2], Jingjing Ren[3], Haoran Yang[1], Xiaowei Hu[4], Renjing Pei[2†], and Pheng-Ann Heng[1]
[1]The Chinese University of Hong Kong [2]Huawei Noah's Ark Lab
[3]HKUST (GZ) [4]South China University of Technology

Abstract

Diffusion models (DMs) have demonstrated remarkable success in real-world image super-resolution (SR), yet their reliance on time-consuming multi-step sampling largely hinders their practical applications. While recent efforts have introduced few- or single-step solutions, existing methods either inefficiently model the process from noisy input or fail to fully exploit iterative generative priors, compromising the fidelity and quality of the reconstructed images. To address this issue, we propose FlowSR, a novel approach that reformulates the SR problem as a rectified flow from low-resolution (LR) to high-resolution (HR) images. Our method leverages an improved consistency learning strategy to enable high-quality SR in a single step. Specifically, we refine the original consistency distillation process by incorporating HR regularization, ensuring that the learned SR flow not only enforces self-consistency but also converges precisely to the ground-truth HR target. Furthermore, we introduce a fast-slow scheduling strategy, where adjacent timesteps for consistency learning are sampled from two distinct schedulers: a fast scheduler with fewer timesteps to improve efficiency, and a slow scheduler with more timesteps to capture fine-grained texture details. Extensive experiments demonstrate that FlowSR achieves outstanding performance in both efficiency and image quality.

1. Introduction

Real-world image super-resolution (SR) aims to reconstruct high-resolution (HR) images from their low-resolution (LR) counterparts while simultaneously removing unknown degradations. With the remarkable success of diffusion models (DMs) [10, 37], SR methods leveraging these models—particularly those built on powerful text-to-image (T2I) models like Stable Diffusion (SD) [31]—have demonstrated outstanding performance [42, 49, 58]. However, these diffusion-based SR approaches require an iterative reverse sampling process that gradually refines noisy inputs into HR outputs. Despite advancements such as efficient ordinary differential equation (ODE) solvers like DDIM [36], the slow inference speed remains a bottleneck, limiting their practicality in real-world applications.

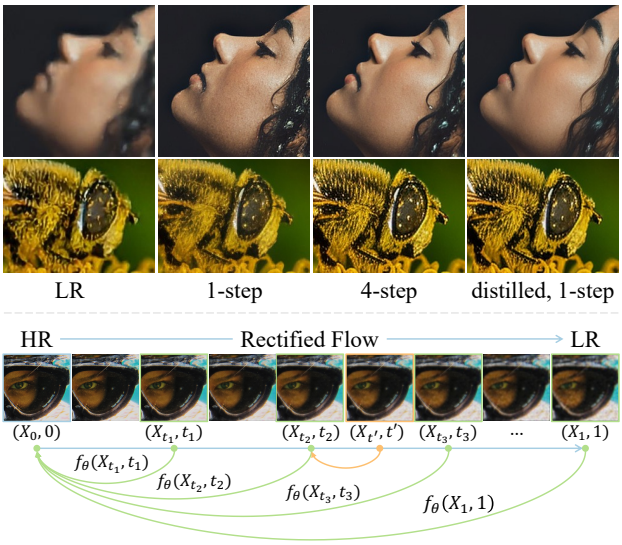

Figure 1. Our consistency SR flow model achieves high-quality single-step inference (distilled, 1-step) by distilling from multi-step higher-quality sampling process (top). We achieve this by formulating SR as a rectified flow that bridges LR and HR images, combined with improved consistency learning (bottom).

Recently, few-step or single-step SR methods derived from diffusion models are designed [3, 45, 48, 59, 61]. Several studies focus on designing more efficient diffusion processes for SR. ResShift [59] speeds up diffusion by shifting the LR–HR residual into the Markov chain and reduces sampling to 15 steps, while DoSSR [3] introduces a diffusion process more compatible with pre-trained DMs. However, they rely on DDPM [10], which uses a more curved transition trajectory and starts with a noise-perturbed LR image that loses crucial information, ultimately limiting performance. In parallel, another line of work tar-

*Part of work done as an intern at Huawei.
†Corresponding author.

gets single-step SR by learning the LR–HR mapping directly. For example, SinSR [45] learns one-step SR prediction by leveraging the teacher's output from ResShift [59] and ground-truth HR as target, whereas OSEDiff adapts pre-trained SD for SR and refines one-step prediction with the VSD loss [46]. However, these methods fail to fully harness the advantages of iterative generative modeling, which naturally facilitates high-quality texture synthesis.

In this work, we present FlowSR, which unifies rectified flow with consistency models (CMs) to enable efficient, single-step image super-resolution. As illustrated in Fig. 1, we reformulate SR as a rectified flow [21], which establishes a simple and straight ODE-based mapping between LR and HR images. By explicitly modeling this trajectory, FlowSR learns fine-grained LR-to-HR transformations for better SR quality and facilitates fast sampling. Building upon this formulation, we further leverage consistency models [38] to enhance single-step inference. Enforcing consistency across points on the same SR flow trajectory distills multi-step higher-quality restoration into fewer steps that reach to the same HR result.

However, the naive consistency distillation (CD) objective in CMs is suboptimal for SR flow, where the SR task demands both high quality and high fidelity. While CD enforces self-consistency across discretization steps, there is no guarantee that the final distillation target aligns well with the ground-truth HR. To address this, we propose HR-regularized consistency learning, which explicitly requires the model's predictions to match real HR images. This additional constraint mitigates teacher-induced errors in the distillation target and enhances the reconstruction of fine-grained details. To further improve efficiency and robustness of the consistency SR flow model, we introduce a fast-slow time scheduling strategy. Rather than sampling distillation timestep pairs from the two boundaries of a discretized interval, we sample adjacent timesteps from distinct "fast" and "slow" schedulers. The fast scheduler uses fewer timesteps to facilitate efficient inference, while the slow scheduler employs more granular steps to maintain alignment with the SR flow objectives. This mixed sampling introduces large and flexible jumps with mild perturbations into the HR regularization steps, allowing for a broader coverage of SR flow trajectories.

To train our consistency SR flow model, we first fine-tune a pre-trained SD model to align with the SR flow objective, followed by consistency SR flow distillation. Additionally, we incorporate a GAN [7] loss and an image quality alignment loss, with the latter promoting desirable text-described attributes in the restored images, thereby improving the SR quality.

As illustrated in Fig. 1, our model leverages the proposed consistency SR flow distillation to transfer the quality improvement typically achieved with multiple sampling steps into a single step. Experiments on real-world datasets demonstrate the superiority of FlowSR. Our core contribution lie in a new paradigm to solve the one-step SR problem. First, we explore efficient flow modeling for SR and identify potential challenges when incorporating consistency models. Second, we introduce several techniques, including HR regularization in consistency learning and a fast-slow time scheduling to enhance one-step SR performance.

2. Related Work

2.1. Image Super-Resolution

Real-world image super-resolution (SR) aims to reconstruct high-resolution (HR) images from low-resolution (LR) inputs, tackling complex and unknown degradations like noise, blur, and compression artifacts. Since Dong et al. [5], numerous deep learning-based methods have been proposed to address the SR problem from different perspectives, including network design [18], model training [15], and degradation simulation [28, 44, 62]. Recently, diffusion models [10, 37] have achieved remarkable success in image generation [30, 31], which inspires their application to SR and demonstrate significant advancements [32, 42].

Diffusion Model-Based SR Conditioned on the LR image, diffusion model-based SR iteratively denoises towards the target HR image. The LR condition can be leveraged either by input concatenation [32] or via adapters [26, 63]. StableSR [42] based on Stable Diffusion [31], uses a trainable time-aware encoder to incorporates LR. Subsequent improvements are achieved through enhanced LR conditioning in DiffBIR [19], semantics-aware prompts in SeeSR [49], reference image generation in CoSeR [39], and scaled-up training in SUPIR [58]. However, these DM-based SR approaches, starting from noise, suffer from low inference speed, typically requiring 50-200 sampling steps. In contrast, ResShift [59] constructs an efficient diffusion model by shifting the residual between HR and LR. Denoising from the noised LR reduces the sampling steps to 15 but it still suffers from limited restoration quality.

Single-Step / Few-Step Image SR Fast diffusion model-based SR methods require only one or a few sampling steps during inference, typically achieved through knowledge distillation [23] or adversarial training [15, 50]. SinSR [45] derives a deterministic sampling process for ResShift and reduces its number of sampling steps to one using distillation. OSEDiff [48] applies variational score distillation [46, 57] to enhance one-step restoration quality by minimizing the KL-divergence between the distribution of its generated outputs and that of the pre-trained T2I model. AddSR [51] achieves 4-step inference by tailoring adversarial diffusion distillation [34] for SR. Similarly, other recent methods, accelerate the inference by using distillation [6, 8, 27] or GAN [2, 16, 60, 61]. On the other hand, more

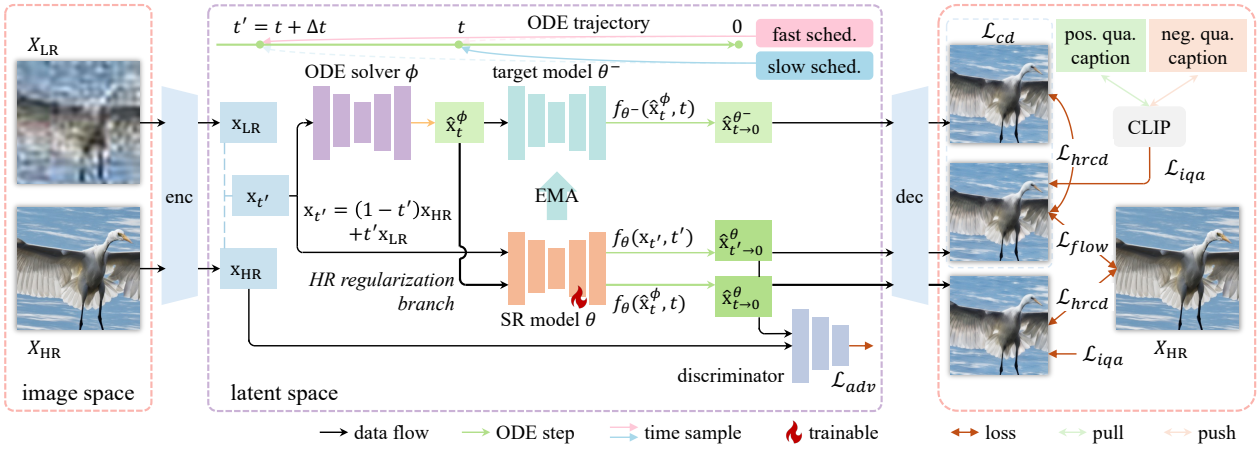

Figure 2. Overview of the training process. The consistency SR flow distills multi-step, high-quality SR capability into single-step inference, while HR regularization ensures the model converges to the high-resolution target. Times t, t' are sampled alternately from fast and slow schedulers. Note that the flow loss $\mathcal{L}_{flow}$ and the consistency loss $\mathcal{L}_{hrcd}$ are computed for different samples within each batch.

efficient SR diffusion processes are constructed using domain shift [3] or flow matching [35], which allow few-step sampling. Yet, they still suffer from inefficient modeling designs and suboptimal performance.

2.2. Consistency Models

Consistency Models (CMs) [38] learn to map any point on the ODE trajectory to its origin, which enable single-step generation and allow trade-offs between quality and computation through multi-step sampling. CMs can be implemented for modern T2I models in latent space [24]. Consistency Trajectory Models (CTMs) [14] mitigate the multi-step sampling issues in CMs by learning an any-to-any mapping from initial points to final points on the ODE trajectory. The Phased Consistency Model [40] partitions the entire ODE trajectory into multiple sub-trajectory phases and ensures consistency within each phase, while PeRFlow [52] straightens the sub-trajectories using the Reflow [22] operation. Additionally, Consistency Flow Matching [53] further enforces velocity field consistency.

3. Preliminaries

3.1. Rectified Flow

Rectified flow [21] is an ODE-based generative model that constructs a simple, straight trajectory to transform samples between two distributions: π_0 (*e.g.*, data) and π_1 (*e.g.*, noise). Rectified flow defines a linear interpolation path between observed samples $X_0 \sim \pi_0$ and $X_1 \sim \pi_1$ as: $X_t = tX_1 + (1-t)X_0$ with time $t \in [0, 1]$. The core idea is to learn a velocity field $v_\theta(X_t, t)$, parameterized by a neural network with weights θ, that matches the derivative of this trajectory. This is achieved by optimizing the following objective: $\min_\theta \int_0^1 \mathbb{E}_{X_0\sim\pi_0, X_1\sim\pi_1}[\|v_\theta(X_t, t) - (X_1 - X_0)\|^2]dt$, which encourages the learned velocity field to align with the direction of the straight path $X_1 - X_0$. Once trained, samples from π_1 can be transformed into π_0 by solving the ODE: $dX_t = v_\theta(X_t, t)dt,\ X_1 \sim \pi_1$. A key advantage of rectified flow lies in its straight trajectories. By design, the optimal velocity field corresponds to a constant-speed flow between X_0 and X_1, enabling efficient and stable sampling with only a few ODE steps.

3.2. Consistency Models

Consistency models [38] are a class of generative models that enable high-quality sample generation with few computational steps. These models learn to map any point (X_t, t) along a probability flow ODE trajectory directly to its origin X_ϵ, where ϵ is a fixed small positive number denoting the trajectory's start. This capability is formalized through the self-consistency property: for any pair of points (X_t, t) and $(X_{t'}, t')$ on the same ODE trajectory, the model $f_\theta(\cdot, t)$ is trained to satisfy: $f_\theta(X_t, t) = f_\theta(X'_t, t') = X_\epsilon$ for all $t, t' \in [\epsilon, 1]$. To ensure consistency during training, the boundary condition $f_\theta(X_\epsilon, \epsilon) = X_\epsilon$ is enforced. Consistency models eliminate trajectory drift by mapping all intermediate states to a consistent origin, avoiding errors from iterative denoising steps. This enables them to generate high-fidelity samples even with large step sizes, such as directly mapping from $t = 1$ to the origin in a single step.

4. Methodology

We leverage the generative capability of rectified flow for high-quality SR and explore consistency models to enable fast inference with fewer sampling steps (*i.e.*, one step), without compromising quality. We define a rectified flow, termed SR flow, for image super-resolution in Sec. 4.1. We then introduce the improved consistency SR flow learning method in Sec. 4.2 and detail the training in Sec. 4.3.

4.1. Rectified Flow for SR

SR Flow We consider Rectified Flow naturally aligns with image super-resolution, where high-resolution (HR) images $X_{\text{HR}} \sim \pi_0$ and their low-resolution (LR) counterparts $X_{\text{LR}} \sim \pi_1$. We define the forward process as a straight path via a linear interpolation between them:

$$X_t = (1-t)X_{\text{HR}} + tX_{\text{LR}}, \tag{1}$$

here time $t \in [0,1]$ with $t = 0$ corresponding to $X_{\text{HR}} \in \mathbb{E}^{H\times W\times 3}$ and $t = 1$ corresponding to $X_{\text{LR}} \in \mathbb{E}^{H\times W\times 3}$, where X_{LR} is upscaled to match the spatial dimensions of X_{HR}. The learning objective of the SR neural network v_θ is to regress the conditional vector fields [20, 21] by following the direction $X_{\text{LR}} - X_{\text{HR}}$:

$$\mathbb{E}_{t,X_t}\|v_\theta(X_t,t) - (X_{\text{LR}} - X_{\text{HR}})\|_2^2. \tag{2}$$

Intuitively, this SR flow establishes a smooth transition between HR and LR. During training, the SR flow model v_θ implicitly learns to invert degradations (*e.g.*, blur, noise) and recover high-frequency details through intermediate refinements such as edge sharpening and texture synthesis. At inference, HR images are reconstructed via reverse sampling along the learned trajectory. Starting from X_{LR}, we solve the reverse ODE using numerical methods like Euler solver: $X_{t'} = X_t + \Delta t \cdot v_\theta(X_t, t)$ from $t = 1$ to $t = 0$. Notably, unlike diffusion-based SR methods that corrupt LR inputs with noise [3, 35, 59], our transition process directly starts from LR, which preserves most of the structural information in the LR image, enabling stable and efficient sampling.

Faster Inference with SR Flow A key advantage of SR flow is its flexibility in sampling steps. It supports faster inference by reducing the number of sampling steps, requiring as few as a single step that directly maps the LR image to its corresponding HR output: $\hat{X}_{\text{HR}} = X_{\text{LR}} - 1 \cdot v_\theta(X_{\text{LR}}, 1)$. This is more easily achievable than the noise-to-image mapping or initializing from noisy inputs in ResShift [59], thanks to the strong correlation between LR and HR images. Additionally, SR flow also supports any number of sampling steps and produces better visual quality when using iterative multi-step sampling, as illustrated in Fig. 1.

4.2. Consistency SR Flow

We enhance single-step super-resolution in SR flow models by distilling multi-step sampling capability (*e.g.*, four steps) into one via consistency learning (Fig. 1). To ensure accurate HR reconstruction, we incorporate the HR target explicitly into consistency learning. Additionally, a fast-slow time sampling schedule is introduced to improve efficiency and robustness. Fig. 2 shows the overall training process.

Consistency Distillation in SR Flow We start with consistency distillation (CD) to distill the capability of a (pre-trained) teacher SR flow model $v_\phi(X,t)$, enabling high-quality SR in fewer inference steps. Given timesteps t and

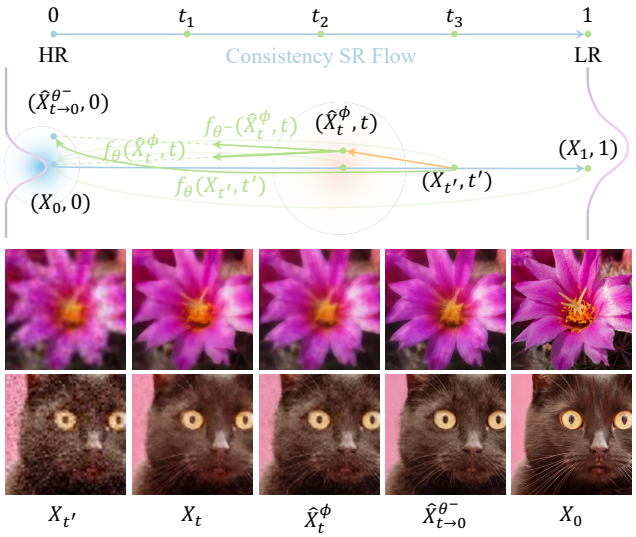

Figure 3. Illustration of HR-regularized consistency learning. Approximation errors may cause the distillation target $\hat{X}^{\theta^-}_{t\to 0}$ to deviate from the true high-quality HR target. HR regularization corrects this by enforcing HR alignment under mild perturbations.

$t' = t + \Delta t$ (where $\Delta t > 0$), we first sample $X_{t'}$ from the forward degradation process (Eq. (1)). The teacher model then estimates $\hat{X}^{\phi}_t$ from its degraded counterpart $X_{t'}$:

$$\hat{X}^{\phi}_t = X_{t'} - \Delta t \cdot v_\phi(X_{t'}, t'), \tag{3}$$

where Euler solver is used. The consistency function $f_\theta(X_t, t)$ for deriving the origin is defined by:

$$f_\theta(X_t, t) = X_t - t \cdot v_\theta(X_t, t), \tag{4}$$

which maps any degraded input X_t at timestep t to the HR image domain. The consistency distillation in SR flow learns a vector field v_θ by minimizing the distance between the prediction of the network $\hat{X}^{\theta}_{t'\to 0} = f_\theta(X_{t'}, t')$ and the distillation target $\hat{X}^{\theta^-}_{t\to 0} = f_{\theta^-}(\hat{X}^{\phi}_t, t)$:

$$\mathcal{L}_{cd} = \mathbb{E}[d(f_\theta(X_{t'}, t'), f_{\theta^-}(\hat{X}^{\phi}_t, t))], \tag{5}$$

where $d(\cdot,\cdot)$ is a distance metric, and θ^- denotes the target model parameters (*e.g.*, exponential moving average (EMA) of θ). This distillation loss enables the newly trained model, f_θ, to generate high-quality samples in a single step.

HR-Regularized Consistency Learning Although the original CD objective theoretically ensures that all trajectory points map to the same origin, it does not explicitly restrict alignment between the generated samples and the corresponding HR targets. This is because velocity field approximation errors from the teacher model v_ϕ and the target model v_{θ^-} can propagate into the distillation target $f_{\theta^-}(\hat{X}^{\phi}_t, t)$, as shown in Fig. 3.

To address this, we introduce an extra HR regularization into Eq. (5), which directly enforces the student's

predictions $\hat{X}^{\theta}_{t\to 0} = f_\theta(\hat{X}^{\phi}_t, t)$ (thus distillation target $f_{\theta^-}(\hat{X}^{\phi}_t, t)$) to match the ground-truth HR images X_0:

$$\mathcal{L}_{hrcd} = \mathbb{E}[d(f_\theta(X_{t'}, t'), f_{\theta^-}(\hat{X}^{\phi}_t, t)) + d(f_\theta(\hat{X}^{\phi}_t, t), X_0)]. \quad (6)$$

By training with real HR targets, consistency learning reduces reliance on the imperfect teacher and target model. This dual consistency objective ensures that f_θ maintains both trajectory consistency and fidelity to HR data. Note that the input $(\hat{X}^{\phi}_t, t)$ can be viewed as a perturbed version of sampled (X_t, t) from the forward process in Eq. (1), as visualized in Fig. 3. This perturbation resembles signal noise and distribution shifts, thereby improving the robustness of the trained model.

Fast-Slow Time Scheduling To further improve the efficiency and robustness of consistency learning, we propose sampling pairs of timesteps, t and t', from two distinct time schedulers: a "fast" scheduler with relatively few timesteps (*e.g.*, 4) and a "slow" scheduler with more granular timesteps (*e.g.*, 1000). The slow scheduler, a standard practice for diffusion or flow models, provides fine-grained learning signals to keep intermediate velocity field predictions well-aligned with the SR flow objectives (Eq. (2)). In contrast, the fast scheduler emphasizes larger jumps along the trajectory, reflecting more desired one-step inference.

During training, we randomly select one scheduler (fast or slow) to first sample $t + \Delta t$, then sample t in its adjacent region from the other scheduler, and compute $(\hat{X}^{\phi}_t, t)$ accordingly using Eq. (3), as illustrated in Fig. 3. This procedure allows flexible and larger jumps Δt through the fast sampler and relaxes the requirement for accurately estimating of X_t from $X_{t+\Delta t}$ by running one discretization step of a numerical ODE solver [38]. We hypothesize that introducing mild perturbations and diversity to $\hat{X}^{\phi}_t$, deviating it from the sampled point X_t, enhances the SR model's robustness and mitigates distribution shifts, particularly when combined with HR regularization.

4.3. Training Consistency SR Flow

SR Flow in Image Space We use Stable Diffusion [31] as the backbone for our SR model, comprising a VAE to map images X into latent representations x, along with a UNet θ adapted for velocity field learning. While the conditional flow matching loss in Eq. (2) can be formulated in latent space, our early experiments observed that such a latent-space objective does not consistently maintain a favorable balance between fidelity and visual quality over varying numbers of inference steps (*e.g.*, from one to four).

In practice, we find that loss in the image space is more effective. Specifically, during SR flow training, we first predict the velocity prediction $v_\theta(\mathrm{x}_t, t)$ in the latent space and compute the HR latent representation $\hat{\mathrm{x}}_{\mathrm{HR}} = f_\theta(\mathrm{x}_t, t)$ (see Eq. (4)). Next, we decode $\hat{\mathrm{x}}_{\mathrm{HR}}$ into the SR image $\hat{X}_{\mathrm{HR}}$ using the VAE's decoder. Last, we compare $\hat{X}_{\mathrm{HR}}$ to the ground-truth HR image X_{HR} using an l_2 loss and a perceptual (LPIPS [64]) loss:

$$\mathcal{L}_{flow} = \|\hat{X}_{\mathrm{HR}} - X_{\mathrm{HR}}\|_2^2 + \lambda_p \,\mathrm{LPIPS}(\hat{X}_{\mathrm{HR}}, X_{\mathrm{HR}}), \quad (7)$$

where $\lambda_p = 2$ is set to weight the LPIPS term.

Adversarial Loss We incorporate an adversarial GAN objective to further enhance SR quality, particularly for texture synthesis. Specifically, a discriminator $\mathcal{D}$ is trained to distinguish between real HR images and those restored by our model. We use the pre-trained diffusion model as a feature extractor in the latent space and attach several additional discriminator heads, following [33, 40, 56]. The hinge loss is adopted for training:

$$\mathcal{L}_{adv} = \mathbb{E}\big[\max(0, 1 - \mathcal{D}(\mathrm{x}_{\mathrm{HR}})) + \max(0, 1 + \mathcal{D}(\hat{\mathrm{x}}_{\mathrm{HR}}))\big]. \quad (8)$$

Note that the input to $\mathcal{D}$ is the noised latent with diffusion time t (omitted here for simplicity). The SR model is then trained to minimize this adversarial objective.

Image Quality Alignment Loss To further align the SR outputs with desirable image quality attributes (*e.g.*, high-resolution, sharp, detailed), we propose an image quality alignment loss based on a text-image contrastive loss. Specifically, we use large vision-language models to generate a positive quality caption c_{pos} (*e.g.*, "image quality is good, with clear details and vibrant colors") from the HR image and a negative quality caption c_{neg} (*e.g.*, "image quality is poor, blurry, and noisy") from its LR counterpart. We then encode these captions using the text encoder $\mathcal{E}_T$ of CLIP [29], while the SR result $\hat{X}$ is encoded via CLIP's image encoder $\mathcal{E}_I$ (ViT-L/14 is used). Image quality alignment loss encourages $\hat{X}$ to be close to c_{pos} and far from c_{neg}:

$$\mathcal{L}_{iqa} = -\log \frac{\exp(\mathrm{sim}(\hat{X}, c_{pos}))}{\exp(\mathrm{sim}(\hat{X}, c_{pos})) + \exp(\mathrm{sim}(\hat{X}, c_{neg}))}, \quad (9)$$

where $\mathrm{sim}(X, c) = \cos(\mathcal{E}_I(X), \mathcal{E}_T(c))$. Intuitively, this loss implicitly guides the network to improve perceptual fidelity and clarity by bridging the gap between human-perceived "good" and "bad" image attributes.

Putting Things Together We first pre-train the SR flow model using the basic conditional flow loss in Eq. (7). Then we continue with the consistency SR flow training by fine-tuning it. The overall loss function is defined as:

$$\mathcal{L} = \mathcal{L}_{flow} + \lambda_{cd}\mathcal{L}_{hrcd} + \lambda_{adv}\mathcal{L}_{adv} + \lambda_{iqa}\mathcal{L}_{iqa}, \quad (10)$$

where λ_{cd}, λ_{adv}, and λ_{iqa} are hyperparameters.

Table 1. Quantitative comparison with state-of-the-art DM-based SR methods on RealSR [1] and DRealSR [47]. The best and second-best results are highlighted in **bold** and underlined.

Datasets	Methods	#Steps	PSNR ↑	SSIM ↑	LPIPS ↓	DISTS ↓	FID ↓	NIQE ↓	MUSIQ ↑	MANIQA ↑	CLIPIQA ↑
RealSR	StableSR [42]	200	24.70	0.7085	0.3018	0.2288	128.51	5.91	65.78	0.6221	0.6178
	DiffBIR [19]	50	24.75	0.6567	0.3636	0.2312	128.99	5.53	64.98	0.6246	0.6463
	SeeSR [49]	50	25.18	0.7216	0.3009	0.2223	125.55	5.41	**69.77**	0.6442	0.6612
	PASD [55]	20	25.21	0.6798	0.3380	0.2260	124.29	5.41	68.75	**0.6487**	0.6620
	ResShift [59]	15	26.31	0.7421	0.3460	0.2498	141.71	7.26	58.43	0.5285	0.5444
	SinSR [45]	1	**26.28**	0.7347	0.3188	0.2353	135.93	6.29	60.80	0.5385	0.6122
	OSEDiff [48]	1	25.15	0.7341	0.2921	0.2128	123.49	5.65	69.09	0.6326	0.6693
	DoSSR [3]	1	25.67	0.7387	0.3356	0.2741	153.25	11.19	62.69	0.5243	0.6259
	FlowSR	1	25.54	**0.7434**	**0.2728**	**0.2013**	**112.60**	**5.28**	69.22	0.6486	**0.6701**
DRealSR	StableSR [42]	200	28.03	0.7536	0.3284	0.2269	148.98	6.52	58.51	0.5601	0.6356
	DiffBIR [19]	50	26.71	0.6571	0.4557	0.2748	166.79	6.31	61.07	0.5930	0.6395
	SeeSR [49]	50	28.17	0.7691	0.3189	0.2315	147.39	6.40	64.93	0.6042	0.6804
	PASD [55]	20	27.36	0.7073	0.3760	0.2531	156.13	**5.55**	64.87	0.6169	0.6808
	ResShift [59]	15	28.46	0.7673	0.4006	0.2656	172.26	8.12	50.60	0.4586	0.5342
	SinSR [45]	1	28.36	0.7515	0.3665	0.2485	170.57	6.99	55.33	0.4884	0.6383
	OSEDiff [48]	1	27.92	0.7835	**0.2968**	0.2165	135.30	6.49	64.65	0.5899	0.6963
	DoSSR [3]	1	**28.55**	**0.7991**	0.3353	0.2801	166.19	12.25	56.72	0.4623	0.5739
	FlowSR	1	28.50	0.7859	0.2975	**0.2115**	**130.30**	6.13	**65.46**	**0.6172**	**0.7074**

5. Experiments

5.1. Experimental Settings

Implementation Details Our model is based on the pre-trained SD 2.1-base [31]. During training, we fine-tune only the U-Net using a LoRA [11] rank of 32. The learning rate is set to 2e-5, with a training patch size of 512×512 and a batch size of 16. The training process consists of two stages: we firstly train the SR flow model for 10k iterations, followed by consistency learning for an additional 20k iterations. The loss weights $\lambda_{cd}, \lambda_{adv}, \lambda_{iqa}$ are set to 0.1, 0.05, and 0.1, respectively.

Data We train our model using LSDIR [17] and the first 10K face images from FFHQ [12]. The LR-HR training pairs are synthesized using the degradation pipeline of Real-ESRGAN [44]. To generate image quality captions for HR and LR images, we use Qwen2-VL [43]. We adopt the real-world test set of StableSR [42] for evaluation and comparison. The test sets include RealSR [1] and DRealSR [47].

Compared Methods We compare our method against several state-of-the-art diffusion-based SR approaches, including multi-step StableSR [42], DiffBIR [19], SeeSR [49], PASD [55], ResShift [59], and one-step SinSR [45], OSEDiff [48], and DoSSR [3].

Evaluation Metrics We evaluate our model using both reference-based and no-reference metrics. For fidelity, we report PSNR and SSIM (on the Y channel in YCbCr space), and for perceptual quality, we use LPIPS [64] and DISTS [4]. FID [9] is reported to compare the distribution of restored images with the ground truth. For no-reference evaluation, we use NIQE [25], MUSIQ [13], MANIQA [54], and CLIPIQA [41].

5.2. Comparison with State-of-the-Art Methods

Quantitative Comparisons Table 1 shows that our approach achieves superior or at least competitive performance compared to recent state-of-the-art methods. Compared with multi-step diffusion-based methods such as SeeSR [49] and PASD [55], our FlowSR obtains equivalent or better no-reference metrics, such as MUSIQ, MANIQA, and CLIPIQA, and surpasses these methods in fidelity-based scores, *e.g.* PSNR and LPIPS, while reducing inference to a single step. Compared to efficient diffusion-based SR methods such as ResShift [59], our method demonstrates a clear advantage. Against the strong DDPM-based one-step competitor OSEDiff [48], FlowSR achieves superior performance across nearly all metrics. Overall, these results demonstrate that our flow-based SR method effectively balances fidelity and perceived quality while showing superiority over compared methods.

Qualitative Comparisons Fig. 4 shows visual comparisons of two real-world examples. Multi-step DM-based methods, such as DiffBIR and PASD, often generate rich but inaccurate textures, while one-step methods like SinSR and OSEDiff tend to produce blurry or less detailed results. In contrast, our method generates faithful SR results, effectively recovering structures such as the stone breakwater and the textures of the cloth.

5.3. Ablation Studies

Effects of SR Flow We first study the flow formulation for image super-resolution. Our SR flow is a straight path between $\pi_0 = X_{\text{HR}}$ and $\pi_1 = X_{\text{LR}}$. We consider alternative formulations starting from noise, where $\pi_1 = \mathcal{N}(0, 1)$, as used in pre-trained T2I models, or starting from noised

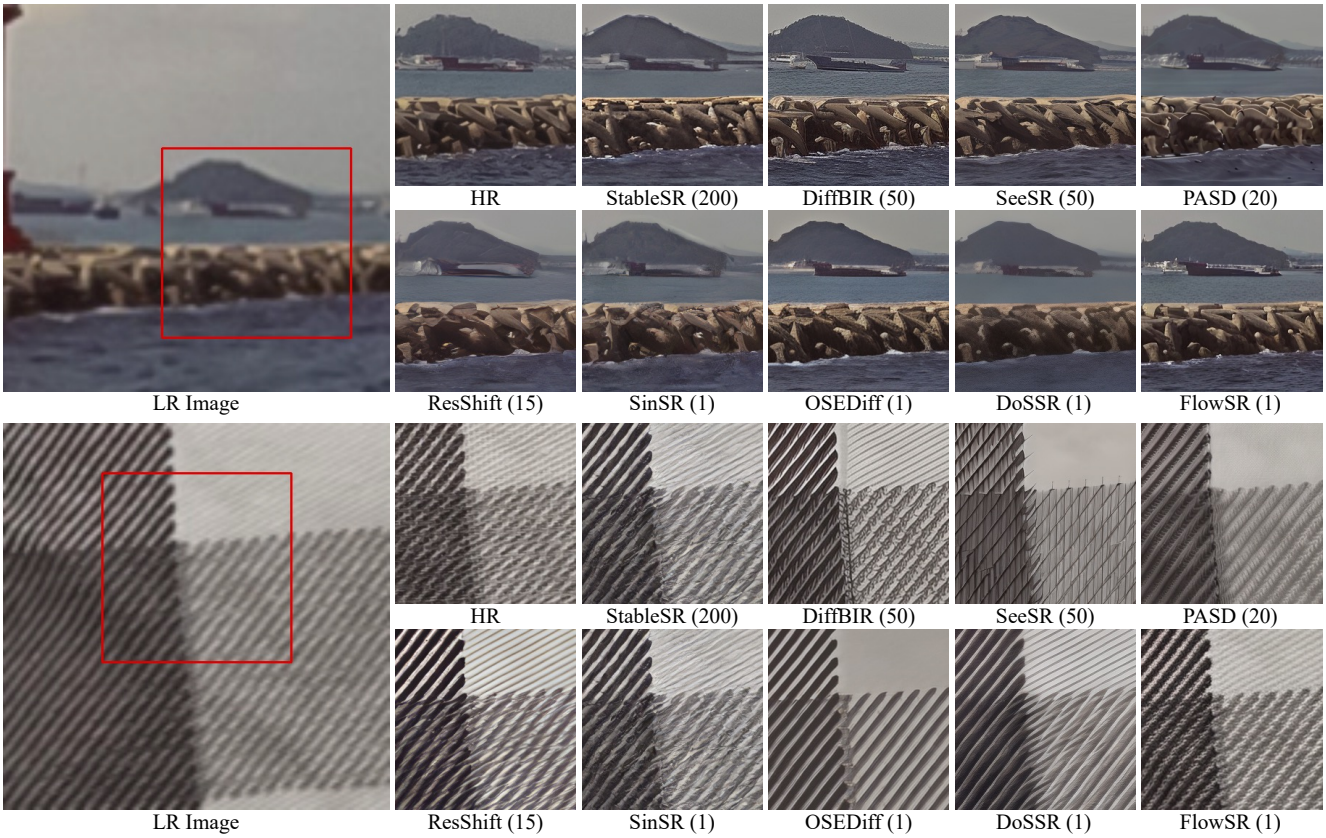

Figure 4. Visual comparisons of different SR methods on real-world examples. The number of sampling steps are indicated in parentheses. Please zoom in for a better view.

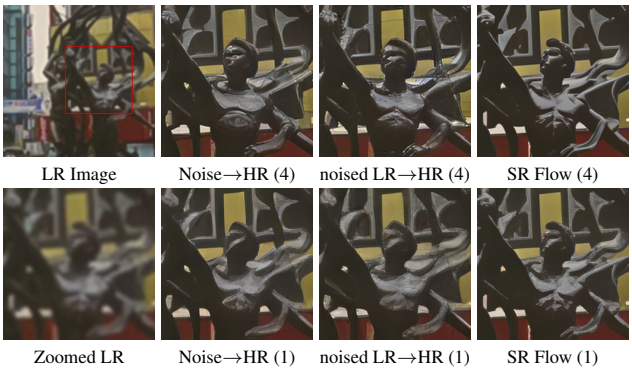

Figure 5. Ablation of SR flow. Our SR flow model produces more accurate SR results with improved structure with more steps.

LR, where $\pi_1 = X_{\mathrm{LR}} + \mu\epsilon$, with $\epsilon \sim \mathcal{N}(0,1)$, as proposed in ResShift [59]. For these alternatives, we concatenate X_t with X_{LR} as the input to provide the LR condition. The trained models are evaluated on the RealSR [1] and DRealSR [47] datasets, as shown in Table 2. We observe that with more inference steps, the image quality (MUSIQ) generally improves, while the fidelity (PSNR) slightly decreases. Fig. 5 presents the visual comparisons, where our SR flow model produces more accurate predictions, with further structure and overall quality enhancement as additional sampling steps (*i.e.*, 4 steps) are applied. Notably, these results demonstrate that the model trained using our SR flow clearly outperforms other transition variants in both single-step and multi-step inference settings.

Table 2. Ablation of SR flow. SR flow mapping LR to HR outperforms alternative flow formulations that start from noise or ResShift [59]-style noised LR on RealSR and DRealSR.

#Steps	Method	RealSR		DRealSR	
		PSNR ↑	MUSIQ ↑	PSNR ↑	MUSIQ ↑
4	Noise→HR	23.83	66.01	26.75	62.76
	noised LR→HR	24.47	65.15	27.74	61.76
	SR Flow (ours)	**25.00**	**68.09**	**27.98**	**65.05**
1	Noise→HR	24.82	65.17	28.07	59.34
	noised LR→HR	24.67	65.25	28.03	60.64
	SR Flow (ours)	**25.51**	**67.40**	**28.65**	**62.53**

Effects of consistency learning Next, we fine-tune the pre-trained SR flow model using the consistency objective described in Sec. 4.2. As shown in Table 3, enforcing self-consistency $\mathcal{L}_{cd}$ across neighboring time steps stabilizes intermediate representations and often reduces LPIPS. However, when the teacher's predictions deviate from the true HR manifold, these errors may propagate to the student model, leading to diminished IQA scores. We address this by introducing the HR regularization term in Eq. (6), which

LR Image | SR Flow | w/ $\mathcal{L}_{cd}$ | w/ $\mathcal{L}_{hrcd}$

Figure 6. Ablation of consistency learning. Our HR-regularized consistency learning effectively reduces distortions in SR outputs while producing high-quality and sharper results (zoom in).

Table 3. Ablation of consistency learning. Fine-tuned SR flow models with consistency objectives are evaluated on DRealSR.

#Steps	Method	PSNR ↑	LPIPS ↓	MUSIQ ↑	MANIQA ↑	CLIPIQA ↑
4	SR Flow	27.98	0.2763	65.05	0.6296	0.6681
	w/ $\mathcal{L}_{cd}$	28.64	0.2654	61.82	0.6042	0.6180
	w/ $\mathcal{L}_{hrcd}$	28.44	0.2799	65.71	0.6374	0.6709
1	SR Flow	28.65	0.2821	62.53	0.5895	0.6545
	w/ $\mathcal{L}_{cd}$	28.59	0.2609	61.69	0.5949	0.6391
	w/ $\mathcal{L}_{hr}$	28.85	0.2805	62.46	0.5871	0.6589
	w/ $\mathcal{L}_{hrcd}$	28.62	0.2830	64.77	0.6156	0.6748

directly aligns student predictions $f_\theta(\hat{X}_t^\phi, t)$ with real HR images, mitigating the teacher's approximation errors. This extra ground-truth constraint helps the model recover fine textural details, which boosts IQA metrics (*e.g.*, MUSIQ and MANIQA) while preserves perceptual fidelity. Consistency learning enhances the robustness of flow-based inference by distilling multi-step capabilities into a single step, allowing for faster sampling without sacrificing quality. Additionally, HR regularization alone (w/ $\mathcal{L}_{hr}$) provides minimal IQA gains. This highlights the importance of consistency learning and its synergy with HR regularization for acceleration in SR flow. Fig. 6 presents visual comparisons, showing that consistency learning helps resolve distorted textures in the baseline SR flow model. Notably, HR-regularized consistency learning generates more realistic and sharper results compared to its baseline.

Analysis of Fast-Slow Time Sampling We evaluate the effectiveness of fast-slow time sampling by comparing it to the N-interval scheduler [38] and by varying the number of fast-scheduler timesteps, while fixing the slow scheduler fixed at 1000 timesteps. When the fast scheduler uses only one timestep, we set $t + \Delta t = 1$ and then sample t from the slow scheduler (see Eq. (3)). As shown in Table 4, increasing Δt (thus reducing the number of intervals) within a reasonable range improves IQA metrics. A plausible explanation is that the slightly larger perturbations introduced at intermediate predictions $\hat{X}_t^\phi$ from X_t encourage the model to be more robust to distribution shifts; see Fig. 3. This, in turn, reduces variance in its estimates of

LR Image | w/o $\mathcal{L}_{hrcd}$ | w/o $\mathcal{L}_{adv}$ | w/o $\mathcal{L}_{iqa}$ | Ours

Figure 7. Ablation of training loss.

Table 4. Ablation of fast-slow time sampling. We compare our fast-slow sampling approach with the N-interval method on DRealSR. The number of N and fast-scheduler timesteps are indicated in parentheses.

Method	PSNR ↑	LPIPS ↓	MUSIQ ↑	MANIQA ↑	CLIPIQA ↑
N-Interval (50)	28.61	0.2807	62.48	0.5914	0.6599
N-Interval (18)	28.74	0.2792	62.90	0.5972	0.6605
N-Interval (4)	28.66	0.2818	63.03	0.5972	0.6617
Slow Only (1000)	28.78	0.2801	62.72	0.5916	0.6586
Fast-Slow (8)	28.54	0.2760	64.01	0.6098	0.6634
Fast-Slow (4)	28.62	0.2830	64.77	0.6156	0.6748
Fast-Slow (1)	28.79	0.2835	63.01	0.5948	0.6605

Table 5. Ablation of training loss on DRealSR.

Method	PSNR ↑	LPIPS ↓	MUSIQ ↑	MANIQA ↑	CLIPIQA ↑
w/o $\mathcal{L}_{hrcd}$	28.52	0.3013	64.42	0.6061	0.6954
w/o $\mathcal{L}_{adv}$	28.45	0.2961	65.29	0.6154	0.7099
w/o $\mathcal{L}_{iqa}$	28.64	0.2958	64.83	0.6111	0.7012
Ours	28.50	0.2975	65.46	0.6172	0.7074

the ODE trajectory's starting point, *i.e.*, HR, resulting in sharper restorations. Moreover, compared to the N-interval strategy or slow-only scheduling, our fast-slow scheduler further boosts IQA metrics while maintaining comparable fidelity. This is because the slow scheduler ensures fine-grained SR flow estimation, whereas the fast scheduler enables efficient sampling. Based on these findings, we set the fast scheduler to 4 timesteps throughout our experiments.

Influence of loss function In addition to flow and consistency learning, we incorporate a GAN loss and an image quality alignment loss to further enhance SR performance, as detailed in Eq. (10). As demonstrated in Table 5 and Fig. 7, these losses all improve no-reference metrics and visual quality. The ablation highlights that eliminating any of these losses results in reduced SR quality, with consistency learning contributing a substantial improvement.

6. Conclusion

This paper presents FlowSR, a new approach for efficient one-step image super-resolution. FlowSR reformulates SR as a rectified flow, leveraging the strengths of iterative generative modeling. To enable high-quality single-step inference, we incorporate consistency learning and devise HR regularization to address its distillation target drifting issue. Additionally, a fast-slow time scheduling strategy is designed to enhance the efficiency and robustness of the consistency SR flow model. FlowSR contributes to the advancement of efficient real-world SR applications.

Acknowledgement

The work described in this paper was supported in part by the National Key R&D Program of China (Grant No. 2023YFE0202700), Research Grants Council of the Hong Kong Special Administrative Region, China, under Project 14200824; and by the Hong Kong Innovation and Technology Fund, under Project MHP/092/22.

References

[1] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *ICCV*, 2019. 6, 7

[2] Bin Chen, Gehui Li, Rongyuan Wu, Xindong Zhang, Jie Chen, Jian Zhang, and Lei Zhang. Adversarial diffusion compression for real-world image super-resolution. In *CVPR*, 2025. 2

[3] Qinpeng Cui, Yixuan Liu, Xinyi Zhang, Qiqi Bao, Zhongdao Wang, Qingmin Liao, Li Wang, Tian Lu, and Emad Barsoum. Taming diffusion prior for image super-resolution with domain shift sdes. *NeurIPS*, 2024. 1, 3, 4, 6

[4] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Image quality assessment: Unifying structure and texture similarity. *TPAMI*, 2020. 6

[5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014. 2

[6] Linwei Dong, Qingnan Fan, Yihong Guo, Zhonghao Wang, Qi Zhang, Jinwei Chen, Yawei Luo, and Changqing Zou. Tsd-sr: One-step diffusion with target score distillation for real-world image super-resolution. In *CVPR*, 2025. 2

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *NeurIPS*, 2014. 2

[8] Xiao He, Huaao Tang, Zhijun Tu, Junchao Zhang, Kun Cheng, Hanting Chen, Yong Guo, Mingrui Zhu, Nannan Wang, Xinbo Gao, et al. One step diffusion-based super-resolution with time-aware distillation. *arXiv preprint arXiv:2408.07476*, 2024. 2

[9] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *NeruIPS*, 2017. 6

[10] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeruIPS*, 2020. 1, 2

[11] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *ICLR*, 2022. 6

[12] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019. 6

[13] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *ICCV*, 2021. 6

[14] Dongjun Kim, Chieh-Hsin Lai, Wei-Hsiang Liao, Naoki Murata, Yuhta Takida, Toshimitsu Uesaka, Yutong He, Yuki Mitsufuji, and Stefano Ermon. Consistency trajectory models: Learning probability flow ode trajectory of diffusion. In *ICLR*, 2024. 3

[15] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 2017. 2

[16] Jianze Li, Jiezhang Cao, Zichen Zou, Xiongfei Su, Xin Yuan, Yulun Zhang, Yong Guo, and Xiaokang Yang. Distillation-free one-step diffusion for real-world image super-resolution. *arXiv preprint arXiv:2410.04224*, 2024. 2

[17] Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, et al. Lsdir: A large scale dataset for image restoration. In *CVPR*, 2023. 6

[18] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV Workshop*, 2021. 2

[19] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. Diffbir: Toward blind image restoration with generative diffusion prior. In *ECCV*, 2024. 2, 6

[20] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *ICLR*, 2023. 4

[21] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *ICLR*, 2023. 2, 3, 4

[22] Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, et al. Instaflow: One step is enough for high-quality diffusion-based text-to-image generation. In *ICLR*, 2024. 3

[23] Eric Luhman and Troy Luhman. Knowledge distillation in iterative generative models for improved sampling speed. *arXiv preprint arXiv:2101.02388*, 2021. 2

[24] Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, and Hang Zhao. Latent consistency models: Synthesizing high-resolution images with few-step inference. *arXiv preprint arXiv:2310.04378*, 2023. 3

[25] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 2012. 6

[26] Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, and Ying Shan. T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. In *AAAI*, 2024. 2

[27] Mehdi Noroozi, Isma Hadji, Brais Martinez, Adrian Bulat, and Georgios Tzimiropoulos. You only need one step: Fast super-resolution with stable diffusion via scale distillation. In *ECCV*, 2024. 2

[28] Long Peng, Wenbo Li, Renjing Pei, Jingjing Ren, Jiaqi Xu, Yang Wang, Yang Cao, and Zheng-Jun Zha. Towards realistic data generation for real-world super-resolution. In *ICLR*, 2025. 2

[29] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, 2021. 5

[30] Jingjing Ren, Wenbo Li, Haoyu Chen, Renjing Pei, Bin Shao, Yong Guo, Long Peng, Fenglong Song, and Lei Zhu. Ultrapixel: Advancing ultra high-resolution image synthesis to new peaks. *NeurIPS*, 2024. 2

[31] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022. 1, 2, 5, 6

[32] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *TPAMI*, 2022. 2

[33] Axel Sauer, Frederic Boesel, Tim Dockhorn, Andreas Blattmann, Patrick Esser, and Robin Rombach. Fast high-resolution image synthesis with latent adversarial diffusion distillation. In *SIGGRAPH Asia*, 2024. 5

[34] Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *ECCV*, 2024. 2

[35] Johannes Schusterbauer, Ming Gui, Pingchuan Ma, Nick Stracke, Stefan Andreas Baumann, Vincent Tao Hu, and Björn Ommer. Boosting latent diffusion with flow matching. In *ECCV*, 2024. 3, 4

[36] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021. 1

[37] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021. 1, 2

[38] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *ICML*, 2023. 2, 3, 5, 8

[39] Haoze Sun, Wenbo Li, Jianzhuang Liu, Haoyu Chen, Renjing Pei, Xueyi Zou, Youliang Yan, and Yujiu Yang. Coser: Bridging image and language for cognitive super-resolution. In *CVPR*, 2024. 2

[40] Fu-Yun Wang, Zhaoyang Huang, Alexander Bergman, Dazhong Shen, Peng Gao, Michael Lingelbach, Keqiang Sun, Weikang Bian, Guanglu Song, Yu Liu, et al. Phased consistency models. *NeurIPS*, 2024. 3, 5

[41] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *AAAI*, 2023. 6

[42] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *IJCV*, 2024. 1, 2, 6

[43] Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, et al. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *arXiv preprint arXiv:2409.12191*, 2024. 6

[44] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *ICCV Workshop*, 2021. 2, 6

[45] Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C Kot, and Bihan Wen. Sinsr: diffusion-based image super-resolution in a single step. In *CVPR*, 2024. 1, 2, 6

[46] Zhengyi Wang, Cheng Lu, Yikai Wang, Fan Bao, Chongxuan Li, Hang Su, and Jun Zhu. Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. *NeurIPS*, 2023. 2

[47] Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In *ECCV*, 2020. 6, 7

[48] Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. *NeurIPS*, 2024. 1, 2, 6

[49] Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *CVPR*, 2024. 1, 2, 6

[50] Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion gans. In *ICLR*, 2022. 2

[51] Rui Xie, Ying Tai, Chen Zhao, Kai Zhang, Zhenyu Zhang, Jun Zhou, Xiaoqian Ye, Qian Wang, and Jian Yang. Addsr: Accelerating diffusion-based blind super-resolution with adversarial diffusion distillation. *arXiv preprint arXiv:2404.01717*, 2024. 2

[52] Hanshu Yan, Xingchao Liu, Jiachun Pan, Jun Hao Liew, Qiang Liu, and Jiashi Feng. Perflow: Piecewise rectified flow as universal plug-and-play accelerator. *NeurIPS*, 2024. 3

[53] Ling Yang, Zixiang Zhang, Zhilong Zhang, Xingchao Liu, Minkai Xu, Wentao Zhang, Chenlin Meng, Stefano Ermon, and Bin Cui. Consistency flow matching: Defining straight flows with velocity consistency. *arXiv preprint arXiv:2407.02398*, 2024. 3

[54] Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *CVPR*, 2022. 6

[55] Tao Yang, Rongyuan Wu, Peiran Ren, Xuansong Xie, and Lei Zhang. Pixel-aware stable diffusion for realistic image super-resolution and personalized stylization. In *ECCV*, 2024. 6

[56] Tianwei Yin, Michaël Gharbi, Taesung Park, Richard Zhang, Eli Shechtman, Fredo Durand, and Bill Freeman. Improved distribution matching distillation for fast image synthesis. *NeurIPS*, 2024. 5

[57] Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Fredo Durand, William T Freeman, and Taesung Park. One-step diffusion with distribution matching distillation. In *CVPR*, 2024. 2

[58] Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In *CVPR*, 2024. 1, 2

[59] Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-

resolution by residual shifting. *NeurIPS*, 2023. 1, 2, 4, 6, 7

[60] Zongsheng Yue, Kang Liao, and Chen Change Loy. Arbitrary-steps image super-resolution via diffusion inversion. In *CVPR*, 2025. 2

[61] Aiping Zhang, Zongsheng Yue, Renjing Pei, Wenqi Ren, and Xiaochun Cao. Degradation-guided one-step image super-resolution with diffusion priors. *arXiv preprint arXiv:2409.17058*, 2024. 1, 2

[62] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *ICCV*, 2021. 2

[63] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *ICCV*, 2023. 2

[64] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 5, 6