Unsupervised Real-World Image Super-Resolution via Rectified Flow Degradation Modeling

Hongyang Zhou
University of Science and Technology Beijing
ilaopi@foxmail.com

Jingyan Qin
University of Science and Technology Beijing
qinjingyanking@foxmail.com

Liuling Chen
University of Science and Technology Beijing
2211703519@qq.com

Junyi He
University of Science and Technology Beijing
d202310391@xs.ustb.edu.cn

Xiaobin Zhu*
University of Science and Technology Beijing
zhuxiaobin@ustb.edu.c

Abstract

Unsupervised real-world image super-resolution (SR) faces critical challenges due to the complex, unknown degradation distributions in practical scenarios. Due to a significant domain gap, existing methods struggle to generalize from synthetic low-resolution (LR) and high-resolution (HR) image pairs to real-world data. In this paper, we propose an unsupervised real-world SR method based on rectified flow to capture and model real-world degradation effectively, synthesizing LR-HR training pairs with realistic degradation. Specifically, given unpaired LR and HR images, we propose a novel Rectified Flow Degradation Module (RFDM) that introduces degradation-transformed LR (DT-LR) images as intermediaries. By modeling the degradation trajectory continuously and invertibly, RFDM better captures real-world degradation and enhances the realism of generated LR images. Additionally, we propose a Fourier Prior Guided Degradation Module (FGDM) that leverages structural information embedded in Fourier phase components to ensure precise modeling of real-world degradation. Finally, the LR images are processed by both FGDM and RFDM, producing final synthetic LR images with real-world degradation. The synthetic LR images are paired with the given HR images to train the off-the-shelf SR networks. Extensive experiments show our method significantly boosts off-the-shelf approaches in real-world scenarios.

***Keywords:** Image Super-Resolution, Real-World, Unsupervised Learning, Flow Matching, Rectified Flow*

*Corresponding Author.

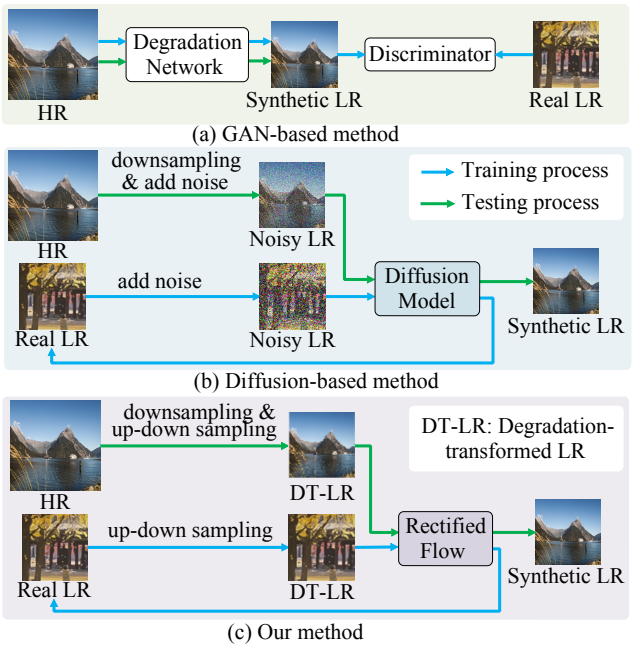


Figure 1. Illustration of different approaches for synthesizing realistic LR-HR pairs from unpaired LR-HR data.

1. Introduction

Image SR focuses on generating high-resolution (HR) images from low-resolution (LR) inputs, serving as a core task in low-level vision. While deep learning-based approaches [10, 23, 54, 12, 52] have shown strong performance under predefined degradations like Bicubic or Gaussian, their effectiveness often declines on real-world images due to the gap between synthetic training data and complex, unknown real-world degradation [25, 5].

Supervised real-world image SR methods rely on either real paired datasets [4, 44] or synthetic degradations [50, 40]. While real datasets provide realistic training data, they are costly to collect and often suffer from alignment and color issues [25, 5]. Synthetic approaches are more scalable but struggle to simulate accurate degradation [25]. Unsupervised methods address these limitations by utilizing unpaired LR-HR datasets to either train SR models directly or generate synthetic LR-HR pairs. The former [47, 30, 45] often struggles with capturing the true HR distribution due to training instability [25, 53]. The latter employs generative methods to synthesize LR-HR pairs, as shown in Fig. 1 (a) and (b), primarily using adversarial learning to capture real degradation or leveraging diffusion to learn real degradation from noisy LR images. For example, SynReal [46] uses diffusion to refine noisy LR inputs from HR images, but its reliance on noise may introduce distortions in training pairs. UDDM [7] combines adversarial learning and diffusion to synthesize LR, but it is limited by the instability of adversarial training. Additionally, using diffusion to learn real degradation from extremely downsampled LR images leads to significant information loss, impacting the quality of synthesized LR-HR pairs. Thus, constructing realistic LR-HR pairs from unpaired data remains a challenging task.

Recently, flow matching [2] has shown great promise in image restoration [60], especially with the development of rectified flow [26, 27]. Unlike diffusion-based methods [16] that rely on iterative denoising, rectified flow models a straight trajectory from a simple prior to the data distribution, leading to more efficient sampling and stable training. This makes it particularly suitable for the SR tasks, as it can more effectively capture the complex degradation in real-world LR images. However, its potential in SR remains largely unexplored, especially under unpaired settings where the absence of LR-HR pairs makes degradation modeling more challenging.

After repeated up-down sampling operations, LR images with different degradations transform into similar degradations. Based on this observation, we propose an unsupervised real-world image super-resolution method using a Rectified Flow-based Degradation Module (RFDM) to capture real-world degradation effectively. Specifically, as shown in Fig. 1 (c), we first apply repeated up-down sampling to real-world LR images to obtain degradation-transformed LR (DT-LR) images. RFDM then learns a continuous and invertible flow transformation from DT-LR to real-world LR images, allowing it to model complex real-world degradation using only real LR data. During inference, this learned transformation is applied to DT-LR images generated from HR images, ensuring that the generated LR images follow real-world degradation. In addition, we introduce a Fourier Prior Guided Degradation Module (FGDM) to enhance the accuracy of degradation modeling further. Leveraging the prior that degradation primarily affects the amplitude in the Fourier domain while the phase preserves structural information, FGDM refines the amplitude of DT-LR images using the phase of real LR images as structural guidance, thereby facilitating more precise degradation modeling in RFDM. Finally, the synthetic LR images generated by FGDM and RFDM, following real-world degradation distributions, are paired with HR images to train the SR models. In summary, the main contributions of this paper are three-fold:

- We propose an unsupervised real-world image super-resolution method via rectified flow degradation modeling, which effectively captures authentic degradation to synthesize realistic training data, thereby enhancing the off-the-shelf SR-Resolvers' performance.
- We propose a novel Rectified Flow-based Degradation Module (RFDM) that utilizes degradation-transformed LR (DT-LR) images as intermediaries to bridge unpaired LR-HR data, effectively modeling real-world degradation via Rectified Flow's ability to learn complex and invertible transformations.
- We propose a Fourier Prior Guided Degradation Module (FGDM) that leverages structural information embedded in Fourier phase components to ensure more precise modeling of real-world degradation in RFDM.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 provides a detailed description of the proposed methodology. The experimental setup, results, and ablation studies are presented in Section 4. We discuss the limitations of our approach in Section 5. Finally, Section 6 concludes the paper.

2. Related Work

2.1. Single Image Super-Resolution Methods

Early SR approaches such as SRCNN [10] employ shallow convolutional networks. Subsequent methods introduce residual learning [14], which enables the design of deeper architectures [17, 23, 55] with improved reconstruction capabilities. Attention-based models [54, 9, 32] are proposed to emphasize informative regions and refine feature representation. Transformer-based approaches [22, 8, 6] further improve performance by capturing long-range dependencies. More recently, MambaIR [13, 12, 58] leverages Vision Mamba [11, 49] to model global context efficiently with linear complexity. Despite these advances, these methods often generate perceptually smooth results lacking fine details. To address this, adversarial training [19, 40, 20] and diffusion-based methods [34, 39, 42, 48] are investigated to enhance perceptual quality. Though successful on synthetic degradation, they struggle with real-world scenarios.

2.2. Real-World Super-Resolution Methods

Supervised real-world SR methods either acquire real LR-HR pairs [4, 44] using specialized imaging systems, which capture authentic degradation but require costly hardware [5], or simulate real-world degradation by enumerating the degradation operations [50, 40, 39, 48, 21, 57], which offer scalability but suffer from inaccurate degradation simulation [25]. Unsupervised real-world SR methods typically utilize unpaired LR-HR images to implicitly learn degradation patterns and directly train SR models [3, 47, 30, 45, 36, 59]. In contrast, some methods first generate paired LR-HR data from unpaired images before training. For example, SynReal [46] trains a diffusion model on real LR images to iteratively add noise to HR inputs and denoise them, producing realistic LR-HR pairs. However, this process often introduces distortions in the generated pairs. UDDM [7] combines GANs and diffusion models to synthesize LR-HR pairs from extremely downsampled LR images. However, the loss of fine details and the instability of adversarial learning limit their quality.

2.3. Rectified Flow

Several recent studies have explored the use of Rectified Flow to improve the efficiency of generative modeling. Liu *et al.* [27] initially introduced the Rectified Flow framework, which learns to generate high-quality images by constructing a linear trajectory between two data distributions, thereby reducing the sampling process to only one or a few steps. InstaFlow [29] extends this concept to text-to-image generation, where rectifying the probability flow trajectories enables high-fidelity synthesis within a single forward pass. FlowGrad [28] further enhances controllability by enabling gradient backpropagation along the ODE trajectories, allowing users to manipulate the output of pre-trained Rectified Flow models. Recently, the Rectified Flow paradigm has also been adapted to low-level vision tasks for model acceleration [21, 60]. For example, FlowIE [60] formulates a conditional Rectified Flow to construct a linear many-to-one mapping for efficient image enhancement, while FluxSR [21] leverages the same principle to distill diffusion priors, achieving one-step SR. Rectified Flow has been rarely explored for the SR task, particularly in unpaired settings where missing LR-HR correspondences hinder effective degradation modeling.

3. Our Method

3.1. Overview

We aim to reduce the degradations gap between training and testing datasets so that SR models trained on synthetic datasets can be applied to real-world LR images. As shown in Fig. 2 and Algorithm 1, our method mainly consists of three components: training degradation modules, synthesizing LR-HR pairs, and training the SR model. Training degradation modules is a crucial component of our method and consists of two key parts: the Fourier Prior Guided Degradation Module (FGDM) and the Rectified Flow-based Degradation Module (RFDM). Specifically, as shown in Fig. 2 (a), we first train FGDM, which applies repeated up-down sampling to generate DT-LR images, enhances the amplitude components, and uses phase information to preserve structural details, facilitating the initial learning of realistic degradation. Building upon this foundation, we train RFDM, further refining the degradation modeling process by leveraging rectified flow to accurately capture real-world degradation transformations.

Once the degradation modules (FGDM and RFDM) are optimized, as shown in Fig. 2 (b), we synthesize LR-HR pairs by sequentially applying both modules. Specifically, the HR images are first downsampled using bilinear interpolation to obtain LR (bi) images. The FGDM then processes these LR (bi) images to introduce initial real-world degradations, followed by the RFDM, which further refines these degradations to match real-world conditions more closely. This pipeline effectively ensures that the SR model, which is trained on the synthesized LR-HR image pairs, can acquire and maintain strong generalization capability towards real-world LR inputs in diverse scenarios.

Algorithm 1 Training and Data Synthesis Pipeline

Input:
1. Unpaired LR-HR training data (X_{real}, Y)
2. SR model

Output:
1. Paired LR-HR training data (X_{syn}, Y)
2. Trained SR model

Networks:
1. FGDM (Fourier Prior Guided Degradation Module)
2. RFDM (Rectified Flow Degradation Module)

Training Degradation Modules:

Initialize parameters of FGDM, RFDM
Generate training data DT-LR images X_{dt} from real LR images X_{real}
Use X_{dt} and X_{real} to train FGDM
Use X_{dt}, X_{real}, and the trained FGDM to train RFDM
return trained FGDM, trained RFDM

Synthesizing LR-HR pairs:

Generate DT-LR images X_{dt} from HR images Y
Use X_{dt}, trained FGDM, and trained RFDM to synthesize real degraded LR images X_{syn}.
return paired training data (X_{syn}, Y)

Training SR Model:

Use paired training data (X_{syn}, Y) to train SR model
return trained SR model

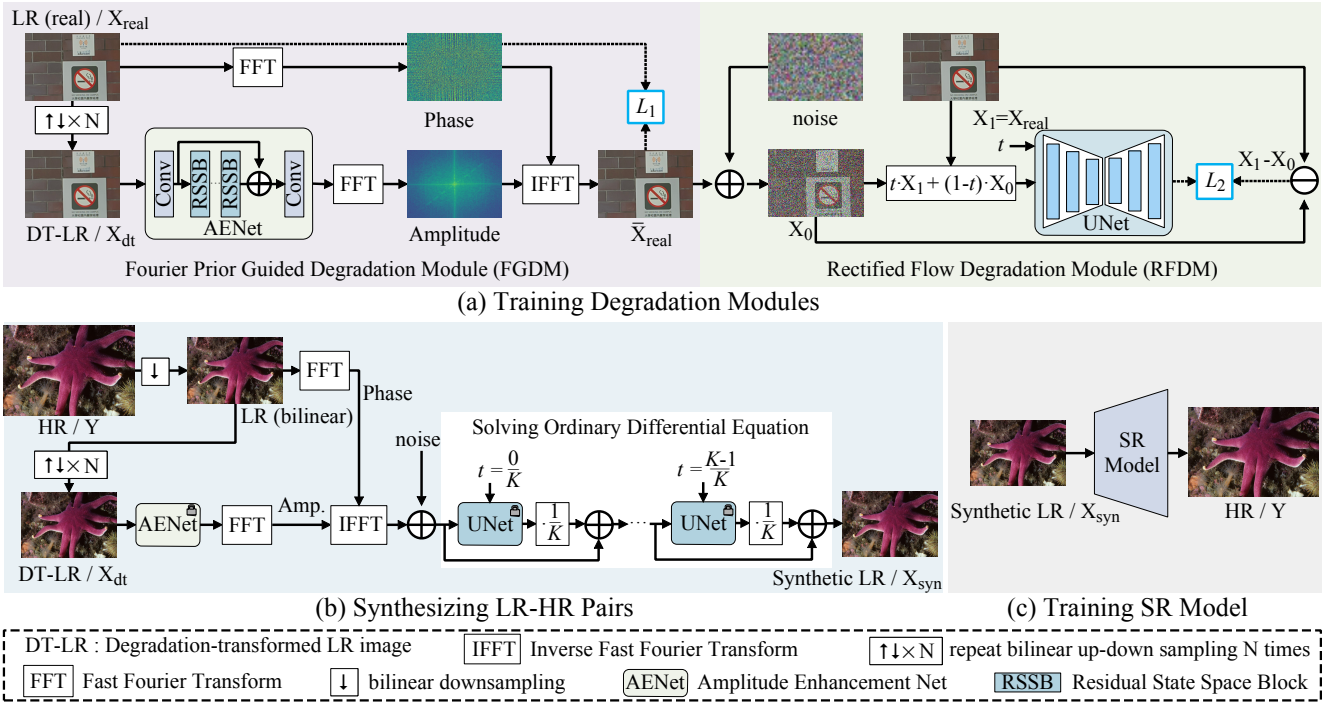


Figure 2. Overview of the proposed method. During the training phase, the Fourier Prior Guided Degradation Module (FGDM) and Rectified Flow Degradation Module (RFDM) are trained using only real-world LR images. In the synthesis phase, the trained models generate realistic pseudo LR images. Finally, the resulting LR-HR pairs can be used to train any SR model.

3.2. Rectified Flow Degradation Module

Flow matching [2, 24] formulates generative modeling as solving an ordinary differential equation (ODE):

$$dZ_t = v(Z_t, t)dt, \tag{1}$$

where v is a time-dependent velocity field that transforms samples from a source distribution P_{Z_0} to a target distribution P_{Z_1}. By integrating this ODE from P_{Z_0}, one can generate samples from P_{Z_1}. Since Eq. 1 may admit multiple valid solutions, flow matching aims to learn a unique v that ensures transformations between distributions. Rectified flow [27] defines a class of flow matching based on linear interpolation as:

$$Z_t = t \cdot Z_1 + (1 - t) \cdot Z_0. \tag{2}$$

Eq. 2 generates a constant velocity vector field $dZ_t = (Z_1 - Z_0)dt$. While this provides direct linear paths between P_{Z_0} and P_{Z_1}, it assumes access to Z_1 at all times $t < 1$, violating causality and limiting its applicability in generative modeling. To overcome this limitation, rectified flow adopts an alternative approach as:

$$v(Z_t, t) = \mathbb{E}[Z_1 - Z_0 \mid Z_t]. \tag{3}$$

Eq. 3 ensures a well-posed solution to the ODE in Eq. 1. Notably, solving Eq. 1 with v often approximates the optimal transport map from P_{Z_0} to P_{Z_1}, particularly when applied iteratively or when the marginals are close to the optimal transport plan [27, 38]. To estimate v, we can train v_θ with the loss criterion as:

$$\min_\theta \int_0^1 \mathbb{E}[||(Z_1 - Z_0) - v_\theta(Z_t, t)||^2]dt. \tag{4}$$

Rectified flow provides a solid foundation for learning real-world degradation from real LR images. By mapping samples from a simple distribution to a target distribution using causal velocity fields, it effectively models real degradation. This flow process allows us to better approximate the distribution of real-world degradation in LR images. As shown in Fig. 2 (a), after obtaining a preliminary real degraded LR $\bar{X}_{real}$ in stage 1, we train RFDM as follows:

$$\min_\theta \int_0^1 \mathbb{E}\left[\|(X_1 - X_0) - v_\theta(X_t, t)\|^2\right] dt, \tag{5}$$

$$X_t := t \cdot X_1 + (1 - t) \cdot X_0, \tag{6}$$

$$X_0 := \bar{X}_{real} + \lambda \cdot n, \quad n \sim \mathcal{N}(0, I), \tag{7}$$

where λ is a hyper-parameter that controls the Gaussian noise level. As shown by [2, 33], adding such a noise is critical when the source and target distributions lie on low and high dimensional manifolds, respectively. Specifically, it alleviates the singularities via learning a deterministic mapping between such distributions. We employ a UNet [35]

architecture to parameterize the velocity field v_θ, and optimize the objective function in Eq. 5 through L_2 loss.

As shown in Fig. 2 (b), to synthesize LR-HR pairs that better reflect realistic degradation, we start from initial LR images X_0 with preliminary real-world degradation derived from the corresponding HR images. We then numerically solve the ODE to simulate the degradation process and obtain realistically degraded LR images. Specifically, we solve the ODE using the Euler method with K discrete steps. Starting from the initial state X_0, the sample is then iteratively updated through the following process:

$$X_{\frac{i+1}{K}} = X_{\frac{i}{K}} + \frac{1}{K} v_\theta(X_{\frac{i}{K}}, \frac{i}{K}), i = 0, 1, ..., K-1. \quad (8)$$

After K steps, we obtain the final degraded images X_1, which are regarded as pseudo LR images that closely approximates the real-world degradation.

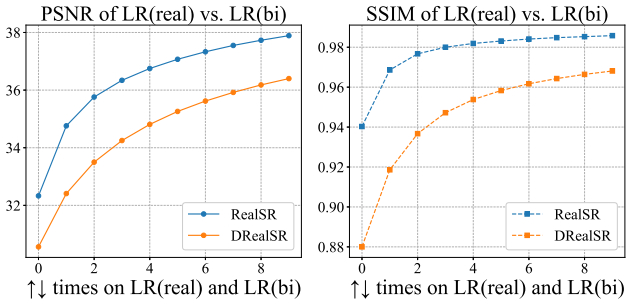


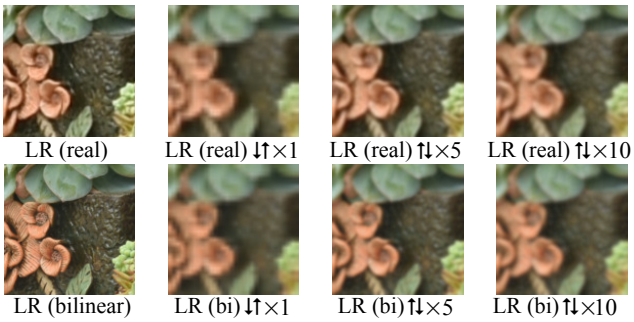
Figure 3. Illustration of degradation removal through repeated up-down sampling operation. PSNR/SSIM evaluation for different up-down sampling times for LR images on RealSR [4] and DRealSR [44] testing datasets.

3.3. Fourier Prior Guided Degradation Module

We observe that after repeated up-down sampling operations, LR images with different degradations transform into similar degradations. As shown in Fig. 3, we calculate the PSNR and SSIM between real-world degraded LR (real) images and bilinearly degraded LR (bi) images after different up-down sampling times. Experimental results on two real-world SR datasets, RealSR [4] and DRealSR [44], demonstrate that as the times of up-down sampling increases, both PSNR and SSIM steadily improve, suggesting that the degradations in LR images transforms towards similarity. Through repeated up-down sampling, we obtain degradation-transformed LR (DT-LR) images to bridge unpaired LR and HR images.

Furthermore, to provide additional validation, we compare the visual results of our approach with those produced by downsampling operation (down-up samping, ↓↑ ×1 in Fig. 4), as adopted in UDDM [7]. As shown in Fig. 4, downsampling step leads to significant loss of information in the LR images, which is observable on both the RealSR and DRealSR datasets. This information loss inevitably increases the difficulty of accurate restoration and ultimately limits the performance. In contrast, even after 10 iterations of up-down sampling, our method preserves considerable texture and structural information, thereby facilitating the learning of real-world degradation while still maintaining richer image details overall.

(a) Visual result of RealSR testing dataset



(b) Visual result of DRealSR testing dataset

Figure 4. Illustration of degradation removal through repeated up-down sampling operation. Visual results for different sampling times of LR images on RealSR [4] and DRealSR [44] datasets.

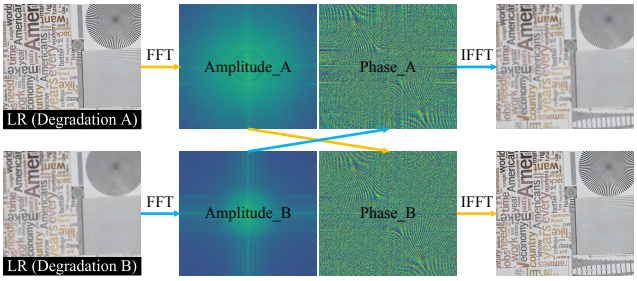


Figure 5. Illustration of amplitude and phase exchange in the Fourier domain. Swapping amplitude spectrum while keeping phase component fixed empirically shows that degradation information is predominantly contained in the amplitude.

While repeated up-down sampling helps transform similar degradations, it inevitably leads to information loss, limiting the effectiveness of RFDM in detailed recovery. To address this problem, we propose the Fourier Prior Guided Degradation Module (FGDM) for initial degradation mod-

Method	RealSR [4]				DRealSR [44]			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
SwinIR (Real-ESRGAN)	24.395	0.7760	0.3037	119.43	26.944	0.8308	0.3219	139.18
SwinIR (Syn-Real)	25.589	0.7687	0.3835	163.13	28.301	0.8309	0.3801	154.59
SwinIR (UDDM)	26.732	0.7913	0.2652	105.92	29.247	0.8386	0.2709	118.09
SwinIR **(Ours)**	**27.022**	**0.7981**	**0.2517**	**101.83**	**29.514**	**0.8409**	**0.2510**	**112.11**
Real-ESRGAN (Real-ESRGAN)	25.600	0.7587	0.2749	138.94	28.549	0.8043	0.2820	146.94
Real-ESRGAN (Syn-Real)	24.341	0.7370	0.3021	159.44	27.483	0.7899	0.3306	171.89
Real-ESRGAN (UDDM)	26.651	0.7769	0.2061	102.43	29.176	0.8032	0.2645	150.44
Real-ESRGAN **(Ours)**	**27.024**	**0.7932**	**0.1915**	**94.90**	**29.323**	**0.8051**	**0.2412**	**142.74**
StableSR (Real-ESRGAN)	24.629	0.7035	0.3014	133.92	27.846	0.7412	0.3337	152.62
StableSR (Syn-Real)	25.679	0.7302	0.3680	165.62	28.621	0.7952	0.3892	183.45
StableSR (UDDM)	26.820	0.7768	0.2514	128.11	29.678	0.8267	0.2567	140.55
StableSR **(Ours)**	**27.128**	**0.7798**	**0.2333**	**112.45**	**29.792**	**0.8313**	**0.2396**	**132.17**

Table 1. Quantitative comparisons of the SR performance of representative models (trained with distinct data generation methods) on RealSR and DRealSR datasets. The best results are highlighted in **bold**.

eling. As shown in Fig. 5, we transform two differently degraded LR images using the Fast Fourier Transform (FFT) to obtain their amplitude and phase spectra. By swapping their amplitude components and reconstructing the images via inverse FFT (IFFT), we observe that the degradations are also exchanged between the LR images. This result indicates that degradation information primarily resides in the amplitude, while structural information is preserved in the phase [31, 56]. Building on this Fourier prior, which indicates that the phase retains structural information with minimal influence from degradation, we leverage the phase from LR images to guide the DT-LR images in learning realistic degradation and compensating for their information loss.

The structural design of the FGDM is illustrated in Fig. 2. FGDM first applies the FFT to the DT-LR image to extract its amplitude and phase components. The amplitude component is enhanced via a dedicated Amplitude Enhancement Network (AENet), which primarily learns realistic degradation patterns from the amplitude, while the original phase from the LR image is retained to provide structural guidance and compensate for structural information loss in the DT-LR image. The enhanced amplitude and the guided phase are then combined through IFFT to reconstruct a preliminary LR image exhibiting realistic degradation. Our AENet mainly comprises convolutional layers and Residual State Space Blocks (RSSB) [13]. RSSB consists of two residual branches: a LayerNorm–VSSM branch for long-range dependency modeling and a LayerNorm–Conv–Channel Attention branch for local feature enhancement, with residual scaling applied to stabilize training. The channel width of each RSSB is set to 96. During training, the FGDM is optimized with an L1 loss function, which enforces pixel-level consistency.

4. Experiments

4.1. Implementation Details and Datasets

Implementation Details. To generate DT-LR images, we apply bilinear up-down sampling 10 times. The number of RSSB blocks in AENet is set to 6, and the hyperparameter λ in Eq. 7 is set to 0.1. We train our FGDM and RFDM with 1,000 epochs using the Adam optimizer [18] (β_1=0.9 and β_2=0.999). We apply a standard data augmentation scheme that comprises randomly rotating 90, 180, 270 degrees, and horizontally flipping. All experiments are conducted with PyTorch 2.2.1 on NVIDIA RTX 3090 GPUs.

Datasets. For training, we construct unpaired LR-HR pairs. The HR dataset includes DIV2K [1], Flickr2K [37], and OutdoorSceneTrain [41], while the LR dataset consists of training datasets from RealSR [4] and DRealSR [44]. For testing, we use the corresponding testing datasets of RealSR and DRealSR. To ensure fair comparison, we crop and evaluate the central region of each image. The resolutions of LR and HR images are 128×128 and 512×512, respectively. We employ peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) [43], learned perceptual image patch similarity (LPIPS) [51] (We use LPIPS-ALEX by default), and fréchet inception distance (FID) [15] to assess both the fidelity and perceptual quality of SR images. PSNR and SSIM are computed on the Y channel of the transformed YCbCr color space, while the other metrics are evaluated directly on the RGB channels.

4.2. Quantitative Evaluation

To evaluate the effectiveness of our method, we use the generated synthetic pairs to train several representative SR models in a supervised setting. For a fair comparison, we

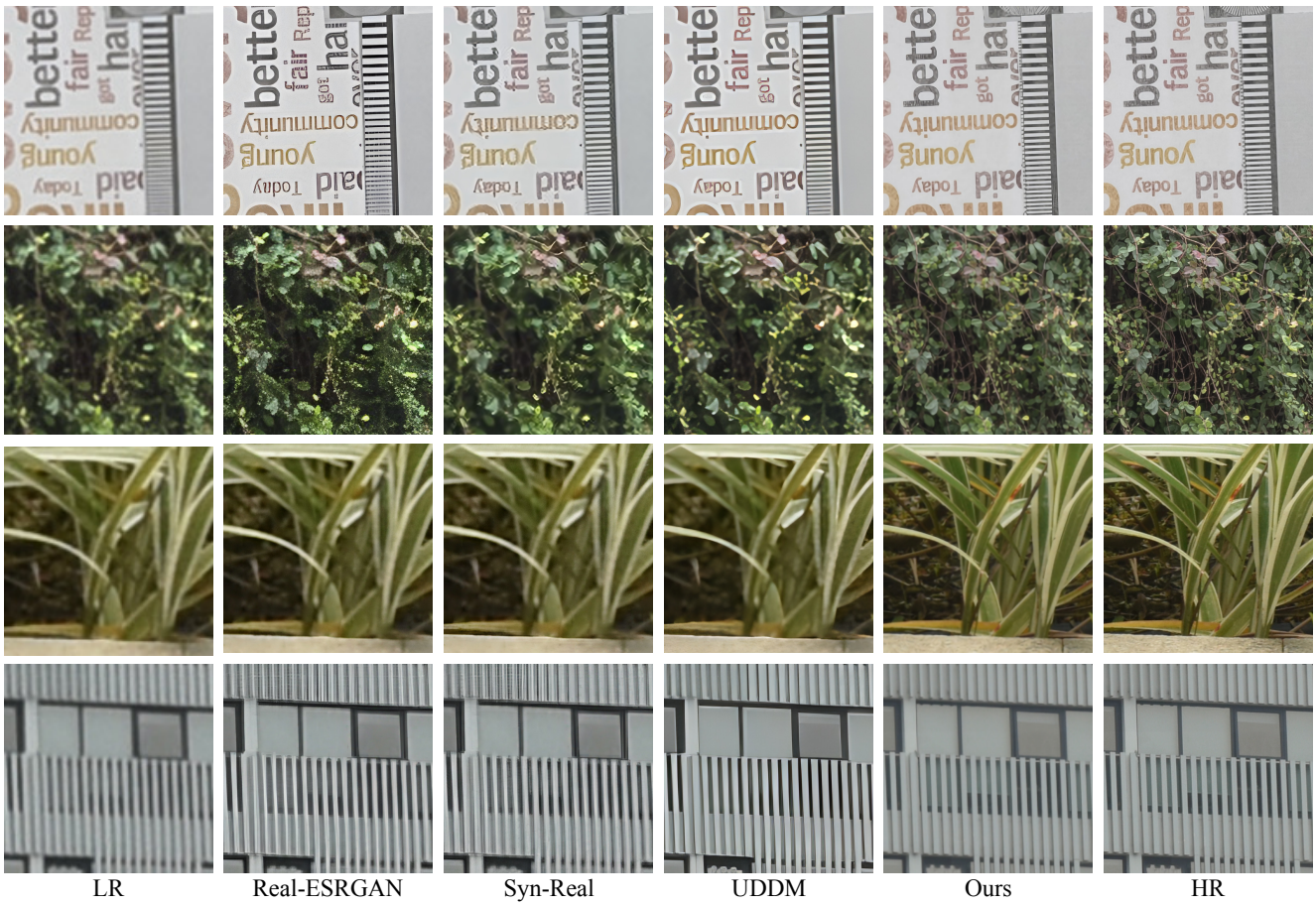
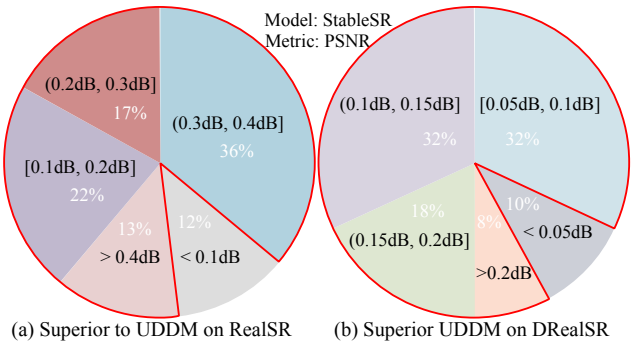


Figure 6. Qualitative results of SR models trained with different synthetic training pairs on the RealSR and DRealSR datasets.



(a) Superior to UDDM on RealSR (b) Superior UDDM on DRealSR

Figure 7. Distribution of per-image PSNR gains of our method compared with UDDM on RealSR and DRealSR.

adopt the same selection of SR architectures as in previous work [7], including SwinIR [22], Real-ESRGAN [40], and StableSR [39], which represent transformer-based, GAN-based, and diffusion-based frameworks, respectively. Additionally, we train these SR models using training pairs generated by other methods, such as Real-ESRGAN [40], Syn-Real [46], and UDDM [7], to evaluate the impact of different synthetic training pairs generation methods.

As listed in Tab. 1, the training data synthesized by our method significantly enhances the performance of SR methods on real-world datasets. Specifically, across all four evaluation metrics, our method achieves the highest results. For example, in terms of PSNR/SSIM, our approach outperforms UDDM by 0.29 dB/0.0068 and 0.267 dB/0.0023 on the RealSR and DRealSR datasets, respectively. Compared to StableSR, our method surpasses the Syn-Real approach by 0.1347/53.17 and 0.1496/51.28 for LPIPS/FID on RealSR and DRealSR datasets, respectively. Besides, as shown in Fig. 7, we perform a per-image statistical analysis on the RealSR and DRealSR datasets. The results indicate that on the RealSR dataset, our method achieves PSNR improvements exceeding 0.1 dB over UDDM on 88% of the test images, while on the DRealSR dataset, 90% of the images exhibit PSNR gains of more than 0.05 dB. These results validate the effectiveness of our method in improving SR performance in real-world scenarios.

4.3. Qualitative Evaluation

To demonstrate the effectiveness of our method in terms of visual performance, we trained SwinIR on datasets syn-

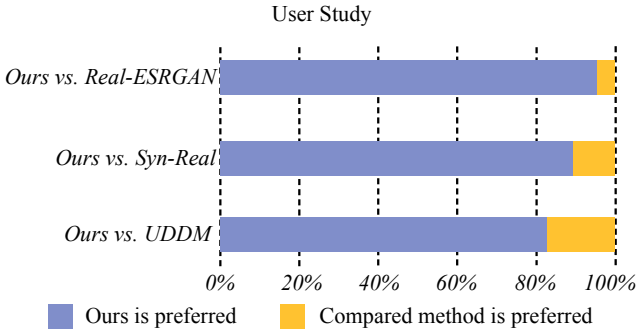


Figure 8. User study results on StableSR. The reported values indicate the preference rate of our method against other approaches.

thesized by different methods. We evaluated the performance on the RealSR and DRealSR datasets. As shown in Fig. 6, the first two rows display results on the RealSR dataset, while the last two rows show results on the DRealSR dataset. We observe that other methods suffer from structural distortions, especially in character and architectural line details, while our method produces fewer and less noticeable distortions. Regarding complex plant textures, the performance of other methods is less satisfactory: Real-ESRGAN introduces numerous artifacts, and Syn-Real and UDDM generate blurry results. In contrast, our method generates plant textures much closer to the HR images. These visual results thus validate the effectiveness of our proposed approach quite clearly.

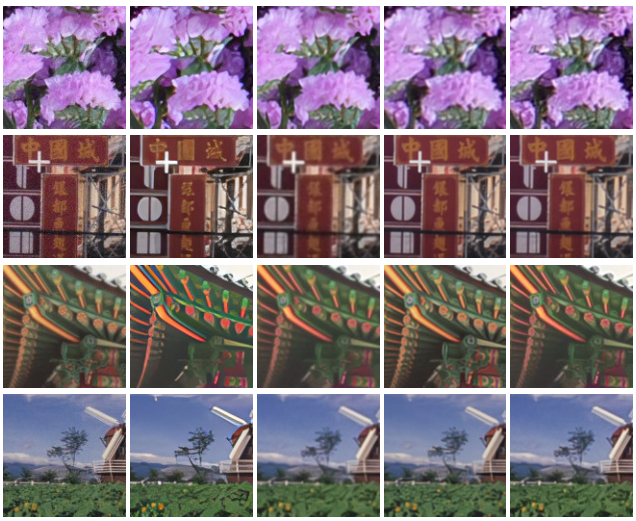


Figure 9. Qualitative comparisons for different synthetic LR images on RealSR and DRealSR datasets.

Besides, we conduct a user study to further assess the perceptual quality of our method in comparison with other approaches. In this study, each participant is presented with a series of image pairs, where one image is generated by our method and the other by a competing method, and is asked to choose the one that appears more realistic and visually pleasing. To ensure fairness, the order of image presentation is randomized to avoid bias. A total of 347 valid responses are collected from 54 participants with varying levels of image processing experience. As shown in Fig. 8, our method is preferred in 80% of the cases, demonstrating a clear perceptual advantage over existing methods. These results further validate the superior visual fidelity and effectiveness of our approach in producing realistic textures and natural image details.

To further validate the effectiveness of our method, we conducted a detailed visual comparison of LR images synthesized by different methods from HR images. As shown in Fig. 9, LR images generated by Real-ESRGAN contain certain unrealistic noise and artifacts, leading to a noticeable discrepancy compared with real-world LR images. The Syn-Real method, on the other hand, suffers from significant structural distortions. Similarly, UDDM faces challenges with the loss of structural textures, leading to blurred details in the synthesized LR images. In contrast, our method consistently generates LR images that are visually much closer to the real LR images, preserving both structural integrity and fine textures. This clearly demonstrates the superiority of our approach in producing more realistic LR images compared to prior methods.

FGDM	RFDM	RealSR	DRealSR
		PSNR/SSIM	PSNR/SSIM
✓		26.395/0.7784	28.752/0.8263
	✓	26.431/0.7791	28.761/0.8277
✓	✓	**27.022/0.7981**	**29.514/0.8409**

Table 2. Ablation study of FGDM and RFDM using the SwinIR method, with evaluation on RealSR and DRealSR.

4.4. Ablation Study

Effectiveness of FGDM and RFDM. To assess the effectiveness of the proposed FGDM and RFDM modules, we conduct ablation experiments, as listed in Tab. 2. When FGDM is used alone, it can only learn preliminary real-world degradation by leveraging Fourier priors to guide the degradation modeling. However, the synthesized training data still deviate from real degradation, leading to suboptimal performance. When RFDM is applied independently, although it achieves better results than FGDM, the model still struggles to reconstruct accurate details due to the lack of structural information in the DT-LR images. When the two modules are combined, the model attains the best performance. This result demonstrates that FGDM and RFDM are complementary and jointly contribute to improved reconstruction quality in a synergistic manner.

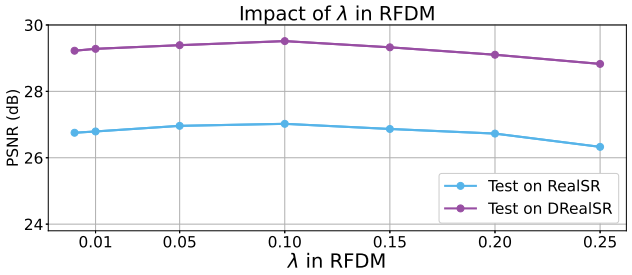


Figure 10. Ablation study of λ and flow steps K in RFDM. Testing on RealSR and DRealSR datasets using SwinIR.

About λ in RFDM. We further analyze the influence of the noise injection ratio λ in the RFDM module by varying its value across a range of settings. As depicted in Fig. 10, when $\lambda = 0$, no noise is introduced during the degradation rectification process, causing the rectified flow to degenerate into a deterministic mapping and thus bringing only marginal improvement. As λ increases, the injected noise encourages the model to explore a wider degradation space, which facilitates better flow refinement and enhances reconstruction performance. However, when λ exceeds 0.1, excessive noise leads to unstable flow estimation and noticeable image distortion, ultimately degrading the PSNR. Therefore, we set $\lambda = 0.1$ in our experiments to achieve a balanced trade-off between effectiveness and stability.

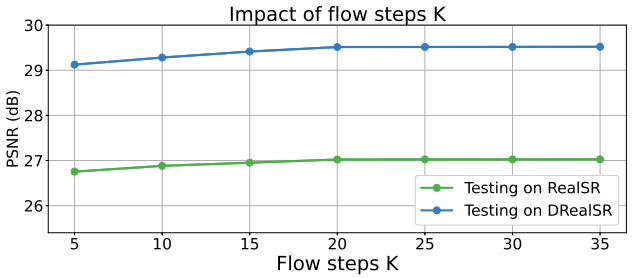


Figure 11. Ablation study of flow steps K in RFDM. Testing on RealSR and DRealSR datasets using SwinIR.

About flow steps K in RFDM. To investigate the influence of the flow step number K in the ODE-based degradation process during inference, we conduct experiments using the SwinIR backbone under different K settings. As illustrated in Fig. 11, the PSNR consistently improves as K increases on both testing datasets, indicating that a larger number of flow steps allows the model to perform a finer approximation of the degradation trajectory and thus capture more realistic degradation characteristics. However, when K exceeds 20, the performance gain becomes marginal while the computational cost continues to rise. Therefore, we fix $K = 20$ in all subsequent experiments to achieve a good and practical trade-off between reconstruction accuracy and inference efficiency for our model.

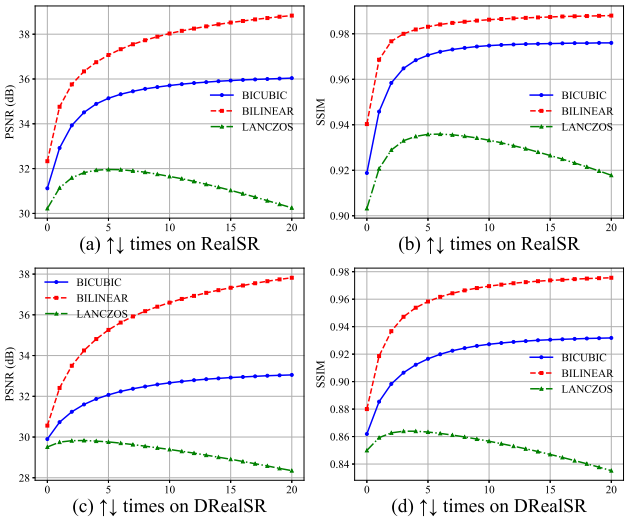
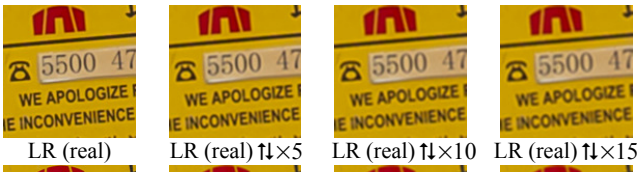


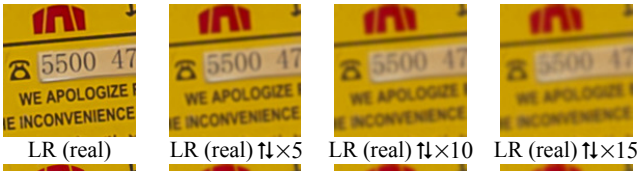
Figure 12. About different sampling methods, we first generate LR images and then perform different numbers of up-down sampling operations on both generated LR images and real LR images. Finally, we calculate the PSNR/SSIM.

About DT-LR images. To investigate the impact of different sampling strategies, we use three common methods: Bicubic, Bilinear, and Lanczos. First, we select a sampling strategy to generate LR images, then apply a different number of up-down sampling operations on both the generated LR images and the real LR images. Finally, we calculate the PSNR/SSIM between the two sets of images. As shown in Fig. 12, the results on the RealSR and DRealSR datasets reveal that with Lanczos, PSNR and SSIM initially improve with increasing up-down sampling operations, but quickly decline. In contrast, Bicubic shows slow improvement in PSNR and SSIM, without reaching the desired peak. This suggests that Lanczos and Bicubic cannot effectively transform LR images with different degradations into similar degradations. Bilinear, however, best meets our requirements. After only a few up-down sampling operations, the degradation of LR images gradually becomes consistent. Although PSNR and SSIM improve with more operations, LR information loss becomes significant. We selected 10 operations, balancing texture preservation and bridging unpaired LR and HR images.

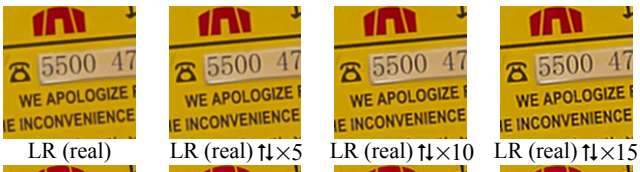
To further validate the effectiveness of the proposed DT-LR image selection, we provide visual comparisons under different downsampling schemes. As illustrated in Fig. 13 (a), the Bicubic method fails to maintain degradation consistency across LR images when the number of up-down sampling operations increases. In contrast, Fig. 13 (c) shows that the Lanczos kernel introduces pronounced ringing artifacts, and the discrepancy between real LR images and Lanczos-downsampled ones becomes more evident as the number of operations grows, thereby violating



(a) Using Bicubic as the sampling method



(b) Using bilinear as sampling method



(c) Using Lanczos as samping method

Figure 13. Visual comparison of LR images generated under different sampling strategies.

the degradation consistency requirement. Most importantly, Fig. 13 (b) reveals that our proposed method effectively preserves structural consistency even with an increasing number of up-down sampling operations, whereas alternative approaches suffer from significant LR information loss. We selected 10 operations, balancing texture preservation and bridging unpaired LR and HR images.

Method	RealSR	DRealSR
	PSNR/SSIM	PSNR/SSIM
FGDM (w/o FP)	26.712/0.7891	29.211/0.8335
FGDM (with FP)	**27.022/0.7981**	**29.514/0.8409**

Table 3. Ablation study of Fourier prior (FP) in FGDM using SwinIR method, with evaluation on RealSR and DRealSR.

About Fourier prior in FGDM. To verify the effectiveness of the Fourier prior, we conduct ablation experiments as presented in Tab. 3. In this experiment, we remove the Fourier prior from FGDM, where the DT-LR images are directly learned from the real LR images in the spatial domain without any frequency-domain guidance. The results demonstrate that incorporating the Fourier prior (FP) in FGDM significantly improves performance on both the RealSR and DRealSR datasets. Specifically, FGDM with FP achieves 27.022/0.7981 and 29.514/0.8409 in terms of PSNR/SSIM on RealSR and DRealSR, respectively, outperforming the variant without FP. This performance gap indicates that the absence of phase-related priors leads to severe loss of structural information in DT-LR images, making the degradation learning process more difficult. By contrast, the Fourier prior provides explicit phase guidance, which helps preserve structural integrity and facilitates more accurate degradation modeling. These results highlight the crucial role of Fourier priors in enhancing both the representational capacity and the robustness of FGDM.

Method	Parm.	FLOPS	Runtime
Syn-Real [46]	120.7M	2491G	2.112s
UDDM [7]	135.4M	2523G	2.443s
FGDM	953K	114G	0.612s
RFDM	68.3M	1562G	1.441s
FGDM + RFDM	69.4M	1676G	1.947s

Table 4. Ablation study of model complexity.

About model complexity. As shown in Tab. 4, FGDM is highly lightweight, containing only 953K parameters with 114G FLOPs and a runtime of 0.612s. RFDM accounts for most of the model complexity (68.3M parameters and 1562G FLOPs), which is expected since it is responsible for modeling complex real-world degradations via Rectified Flow. Nevertheless, when combining FGDM and RFDM, it remains more efficient than prior degradation modeling methods such as Syn-Real and UDDM.

Number	RealSR	DRealSR
	PSNR/SSIM	PSNR/SSIM
5	26.807/0.7935	29.331/0.8361
10	**27.022/0.7981**	**29.514/0.8409**
15	26.916/0.7964	29.482/0.8381
20	26.893/0.7940	29.355/0.8373

Table 5. Ablation study of number of up-down sampling operations of DT-LR images.

Number of up-down Sampling Operations. To further analyze the impact of the number of up-down sam-

pling operations (using the bilinear interpolation), we conduct controlled experiments with varying operation counts of 5, 10, 15, and 20. SwinIR is adopted as the baseline SR model, and evaluations are performed on both the RealSR and DRealSR testing datasets. As reported in Tab. 5, when the number of operations is relatively small (e.g., 5), the degradation of the generated LR images does not sufficiently align with that of real LR images. This misalignment leads to weaker regional consistency and suboptimal performance. Conversely, an excessive number of operations (e.g., 15 or 20) results in substantial information loss and structural degradation in the LR images, compromising the fidelity and representativeness of the synthetic data. Among all configurations, the setting with 10 up-down sampling operations yields the best balance between degradation alignment and information preservation, thereby effectively and simultaneously enhancing both reconstruction accuracy and visual realism.

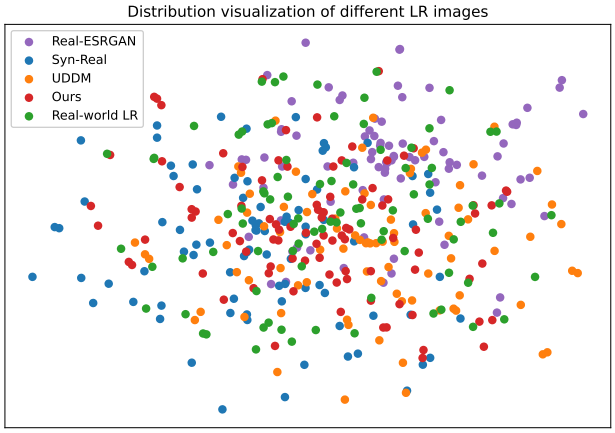


Figure 14. A visualization study on the statistical distribution of different synthetic LR images.

Distribution Analysis of Synthetic LR Images. Fig. 14 visualizes the t-SNE distributions of LR images generated by different methods and real-world LR images. Compared with Real-ESRGAN, Syn-Real, and UDDM, whose synthesized LR images exhibit noticeable distribution shifts, our method shows a much closer alignment and greater overlap with the real-world LR distribution. This indicates that our degradation modeling strategy better captures the statistical characteristics of real degradations, which helps explain the improved reconstruction performance.

5. Limitation

The first limitation of our method lies in its reliance on unpaired LR-HR images for training, while confronting a substantial data imbalance between the two domains. For HR images, we employ DIV2K (800 images), Flickr2K (2,560 images), and OutdoorSceneTraining (10,324 images). In contrast, the LR images are derived from real-world datasets, primarily RealSR (495 images) and DRealSR (840 images). This pronounced discrepancy leads to a significant shortage of LR data, which restricts the diversity of degradation patterns and scene coverage, thereby constraining the overall quality of our synthesized training pairs. It is important to note that scarcity of real-world data constitutes a common challenge in the SR field [5, 25]. Effectively addressing this limitation will constitute a central direction for our future work.

The second limitation concerns the generation process of DT-LR images. In our approach, we conduct an extensive experimental exploration of various combinations of sampling methods, scaling factors, and sampling counts through manual trial and error, which is both highly time-consuming and computationally demanding. Although the current strategy for producing DT-LR images may not represent the theoretically optimal solution, it has been empirically validated to outperform the methods employed in UDDM [7] and Syn-Real [46] in effectively bridging LR images with diverse degradations. In future work, we aim to establish a more efficient and principled framework for generating idealized DT-LR images, thereby providing a more robust and effective solution to the challenges encountered in real-world SR scenarios.

6. Conclusion

In this paper, we propose an unsupervised real-world image super-resolution via rectified flow degradation modeling, synthesizing LR-HR training pairs with realistic degradation. Our approach introduces two key modules: the Rectified Flow Degradation Module (RFDM) and the Fourier Prior Guided Degradation Module (FGDM). RFDM captures real-world degradation by modeling the degradation trajectory continuously and invertibly, using degradation-transformed LR (DT-LR) images as intermediaries to bridge unpaired LR-HR pairs. Meanwhile, FGDM leverages the structural information embedded in Fourier phase components to ensure a more precise degradation modeling. By utilizing both modules, we generate synthetic LR images that closely resemble real-world degradation, which are then paired with HR images for training the off-the-shelf SR network. Extensive experiments on diverse real-world datasets show that our method offers a robust and promising solution for enhancing SR performance in challenging real-world applications.

Acknowledgement

The research is supported by National Natural Science Foundation of China (62576031).

References

[1] E. Agustsson and R. Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 126–135, 2017. 6

[2] M. S. Albergo and E. Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *International Conference on Learning Representations*, 2023. 2, 4

[3] A. Bulat, J. Yang, and G. Tzimiropoulos. To learn image super-resolution, use a gan to learn how to do image degradation first. In *Proceedings of the European Conference on Computer Vision*, pages 185–200, 2018. 3

[4] J. Cai, H. Zeng, H. Yong, Z. Cao, and L. Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3086–3095, 2019. 2, 3, 5, 6

[5] H. Chen, X. He, L. Qing, Y. Wu, C. Ren, R. E. Sheriff, and C. Zhu. Real-world single image super-resolution: A brief review. *Information Fusion*, 79:124–145, 2022. 1, 2, 3, 11

[6] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao. Pre-trained image processing transformer. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12299–12310, 2021. 2

[7] Y. Chen, M. Yao, W. Li, R. Pei, J. Zhao, and W. Ren. Unsupervised diffusion-based degradation modeling for real-world super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 2348–2356, 2025. 2, 3, 5, 7, 10, 11

[8] Z. Chen, Y. Zhang, J. Gu, L. Kong, X. Yang, and F. Yu. Dual aggregation transformer for image super-resolution. In *International Conference on Computer Vision*, pages 12312–12321, 2023. 2

[9] T. Dai, H. Zha, Y. Jiang, and S.-T. Xia. Image super-resolution via residual block attention networks. In *IEEE/CVF International Conference on Computer Vision Workshops*, 2019. 2

[10] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, pages 184–199, 2014. 1, 2

[11] A. Gu and T. Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *Conference on Language Modeling*, 2024. 2

[12] H. Guo, Y. Guo, Y. Zha, Y. Zhang, W. Li, T. Dai, S.-T. Xia, and Y. Li. Mambairv2: Attentive state space restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 28124–28133, 2025. 1, 2

[13] H. Guo, J. Li, T. Dai, Z. Ouyang, X. Ren, and S.-T. Xia. Mambair: A simple baseline for image restoration with state-space model. In *European Conference on Computer Vision*, pages 222–241, 2024. 2, 6

[14] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 2

[15] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*, 30, 2017. 6

[16] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 2

[17] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016. 2

[18] D. P. Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[19] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4681–4690, 2017. 2

[20] M. Lee, S. Hyun, W. Jun, and J.-P. Heo. Auto-encoded supervision for perceptual image super-resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 17958–17968, 2025. 2

[21] J. Li, J. Cao, Y. Guo, W. Li, and Y. Zhang. One diffusion step to real-world super-resolution via flow trajectory distillation. In *The International Conference on Machine Learning*, 2025. 3

[22] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 1833–1844, 2021. 2, 7

[23] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 136–144, 2017. 1, 2

[24] Y. Lipman, R. T. Chen, H. Ben-Hamu, M. Nickel, and M. Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. 4

[25] A. Liu, Y. Liu, J. Gu, Y. Qiao, and C. Dong. Blind image super-resolution: A survey and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5461–5480, 2022. 1, 2, 3, 11

[26] Q. Liu. Rectified flow: A marginal preserving approach to optimal transport. *arXiv preprint arXiv:2209.14577*, 2022. 2

[27] X. Liu, C. Gong, and Q. Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023. 2, 3, 4

[28] X. Liu, L. Wu, S. Zhang, C. Gong, W. Ping, and Q. Liu. Flowgrad: Controlling the output of generative odes with gradients. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24335–24344, 2023. 3

[29] X. Liu, X. Zhang, J. Ma, J. Peng, et al. Instaflow: One step is enough for high-quality diffusion-based text-to-image generation. In *The International Conference on Learning Representations*, 2023. 3

[30] S. Maeda. Unpaired image super-resolution using pseudo-supervision. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 291–300, 2020. 2, 3

[31] H. Nehete, A. Monga, P. Kaushik, and B. K. Kaushik. Fourier prior-based two-stage architecture for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop*, pages 6014–6023, 2024. 6

[32] B. Niu, W. Wen, W. Ren, X. Zhang, L. Yang, S. Wang, K. Zhang, X. Cao, and H. Shen. Single image super-resolution via a holistic attention network. In *European Conference on Computer Vision*, pages 191–207, 2020. 2

[33] G. Ohayon, T. Michaeli, and M. Elad. Posterior-mean rectified flow: Towards minimum mse photo-realistic image restoration. In *International Conference on Learning Representations*, 2025. 4

[34] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2

[35] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 234–241, 2015. 4

[36] W. Sun and Z. Chen. Learning many-to-many mapping for unpaired real-world image super-resolution and downscaling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 3

[37] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 114–125, 2017. 6

[38] A. Tong, K. Fatras, N. Malkin, G. Huguet, Y. Zhang, J. Rector-Brooks, G. Wolf, and Y. Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, pages 1–34, 2024. 4

[39] J. Wang, Z. Yue, S. Zhou, K. C. Chan, and C. C. Loy. Exploiting diffusion prior for real-world image super-resolution. *International Journal of Computer Vision*, 132(12):5929–5949, 2024. 2, 3, 7

[40] X. Wang, L. Xie, C. Dong, and Y. Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1905–1914, 2021. 2, 3, 7

[41] X. Wang, K. Yu, C. Dong, and C. C. Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 606–615, 2018. 6

[42] Y. Wang, W. Yang, X. Chen, Y. Wang, L. Guo, L.-P. Chau, Z. Liu, Y. Qiao, A. C. Kot, and B. Wen. Sinsr: diffusion-based image super-resolution in a single step. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25796–25805, 2024. 2

[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 6

[44] P. Wei, Z. Xie, H. Lu, Z. Zhan, Q. Ye, W. Zuo, and L. Lin. Component divide-and-conquer for real-world image super-resolution. In *European Conference on Computer Vision*, pages 101–117, 2020. 2, 3, 5, 6

[45] Y. Wei, S. Gu, Y. Li, R. Timofte, L. Jin, and H. Song. Unsupervised real-world image super resolution via domain-distance aware training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13385–13394, 2021. 2, 3

[46] T. Yang, P. Ren, L. Zhang, et al. Synthesizing realistic image restoration training pairs: A diffusion approach. *arXiv preprint arXiv:2303.06994*, 2023. 2, 3, 7, 10, 11

[47] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 701–710, 2018. 2, 3

[48] Z. Yue, K. Liao, and C. C. Loy. Arbitrary-steps image super-resolution via diffusion inversion. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 23153–23163, 2025. 2, 3

[49] J. Zhang, M. Gao, W. Li, D. Fang, and C. Li. Visual state space model for image super-resolution. *IEEE Transactions on Instrumentation and Measurement*, 2024. 2

[50] K. Zhang, J. Liang, L. Van Gool, and R. Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4791–4800, 2021. 2, 3

[51] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 6

[52] S.-X. Zhang, X. Zhu, L. Chen, J.-B. Hou, and X.-C. Yin. Arbitrary shape text detection via segmentation with probability maps. *IEEE transactions on pattern analysis and machine intelligence*, 45(3):2736–2750, 2022. 1

[53] W. Zhang, X. Li, X. Chen, X. Zhang, Y. Qiao, X.-M. Wu, and C. Dong. Seal: A framework for systematic evaluation of real-world super-resolution. In *International Conference on Learning Representations*, 2023. 2

[54] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision*, pages 286–301, 2018. 1, 2

[55] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018. 2

[56] C. Zhao, W. Cai, C. Dong, and C. Hu. Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8281–8291, 2024. 6

[57] H. Zhou, X. Zhu, Z. Han, and X.-C. Yin. Real-world image super-resolution via spatio-temporal correlation network. In *IEEE International Conference on Multimedia and Expo*, pages 1–6, 2021. 3

[58] H. Zhou, X. Zhu, J. Qin, Y. Xu, R. M. Cesar-Jr, and X.-C. Yin. Multi-scale texture fusion for reference-based image super-resolution: New dataset and solution. *International Journal of Computer Vision*, 133(10):6971–6992, 2025. 2

[59] H. Zhou, X. Zhu, J. Zhu, Z. Han, S.-X. Zhang, J. Qin, and X.-C. Yin. Learning correction filter via degradation-adaptive regression for blind single image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12365–12375, 2023. 3

[60] Y. Zhu, W. Zhao, A. Li, Y. Tang, J. Zhou, and J. Lu. Flowie: Efficient image enhancement via rectified flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13–22, 2024. 2, 3