잠재 흐름 매칭을 통한 실세계 초해상도용 연속 열화 데이터셋 생성

김현재 [1], 김동진 [2], 진유진 [3], 김태현 [4]
[1] 한양대학교 컴퓨터소프트웨어학과 석사과정
[2] 한양대학교 컴퓨터소프트웨어학과 박사과정
[3] 한양대학교 AI 응용학과 석사과정
[4] 한양대학교 컴퓨터소프트웨어학과 교수

khj1112@hanyang.ac.kr, dongjinkim@hanyang.ac.kr, eugenebori@hanyang.ac.kr, taehyunkim@hanyang.ac.kr

Continuous Real-World Degradation Dataset Generation via Latent Flow Matching for Image Super-Resolution

Hyeon-Jae Kim[1], Dong-Jin Kim[2], Eugene Jin [3], Tae-Hyun Kim[4]
[1]Dept. of Computer Science, Han-Yang University
[2]Dept. of Computer Science, Han-Yang University
[3]Dept. of Artificial Intelligence Application, Han-Yang University
[4]Dept. of Computer Science, Han-Yang University

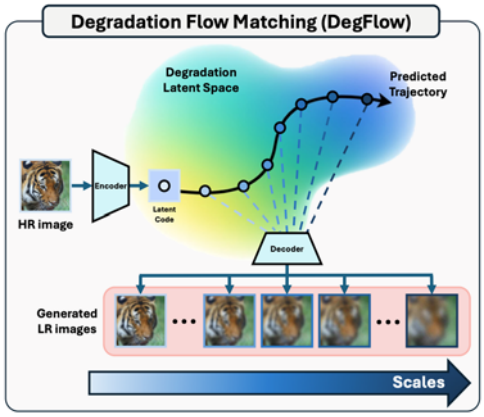
Abstract

While deep learning-based super-resolution (SR) methods have shown impressive outcomes with synthetic degradation scenarios such as bicubic downsampling, they frequently struggle to perform well on real-world images that feature complex, non-linear degradations like noise, blur, and compression artifacts. Recent efforts to address this issue have involved the painstaking compilation of real low-resolution and high-resolution (HR) image pairs, usually limited to several specific downscaling factors. To address these challenges, our work introduces a novel framework capable of synthesizing authentic LR images from a single given HR image by leveraging the latent degradation space with flow matching. Our approach generates LR images with realistic artifacts at unseen degradation levels, which facilitates the creation of large-scale, real-world SR training datasets. Qualitative assessments verify that our synthetic LR images accurately replicate real-world degradations.

1. INTRODUCTION

Image super-resolution (SR) models excel on synthetic LR–HR pairs yet falter in the wild because bicubic-style degradations ignore real camera pipelines. Hand-crafted mixtures help but still miss real statistics, and physically captured datasets are costly and limited to a few discrete magnifications [1]. Recent learned degraders synthesize extra LR from small real pairs, but most lack explicit scale control or require paired LR at multiple scales, limiting arbitrary-scale SR [2].

We propose DegFlow, a degradation modeling framework that learns from a small number of discrete scales and then synthesizes realistic LR images at continuous, unseen scales using only an HR input at test time. As illustrated in Fig. 1, a residual autoencoder maps images to a compact latent preserving high-frequency structure, and a scale-conditioned latent flow-matching network learns a smooth trajectory fitted with a natural cubic spline through the training scales. At inference, we transport the HR latent to any target point on this trajectory and decode it to LR with realistic degradations. This yields more realistic degradations than hand-crafted pipelines and prior learned degraders, while providing explicit, continuous scale control that benefits both fixed- and arbitrary-scale SR models.



(Fig. 1) DegFlow generates real-world LR images across continuous scales by modeling degradation trajectories in a learned latent space. The generated LR images are used to train arbitrary SR models for high-quality restoration.

2. PROPOSED METHOD

2.1 Overview

DegFlow follows a two-stage pipeline: in the first stage we train a residual autoencoder (RAE) to obtain compact but detail-preserving latents; in the second stage we learn a latent flow that connects latents of discrete degradation levels along a smooth, time-parameterized trajectory. At test time, an HR image is encoded to a latent, evolved along the learned flow to an arbitrary timestep corresponding to the desired scale, and decoded to the LR image.

2.2 Residual Autoencoder (RAE)

Let $I \in \mathbb{R}^{C\times H\times W}$ denote an HR or LR image. The encoder E_θ maps I to a compact latent $z = E_\theta(I) \in \mathbb{R}^{Cr^2\times\frac{H}{r}\times\frac{W}{r}}$, where r is the spatial compression factor. To mitigate detail loss resulted from this compression, we propagate multi-scale encoder features to the decoder through residual skip connection as:

$$\hat{I} = D_\theta(z;\, \mathcal{H}_{HR}),$$

where $\mathcal{H}_{HR} = \{h_{HR}^{(l)}\}_{l=1}^{L}$ is the hidden features on multiple scales, and $h_{HR}^{(l)}$ denotes the hidden feature at scale level l among L scales.

We train the RAE with an L1 reconstruction loss applied to both HR and LR inputs, while the decoder exclusively receives HR feature skips to preserve high-frequency structure[4].

2.3 Latent Flow Matching (LFM)

Timestamping degradation. Suppose the training set provides discrete scales $S = \{s_k\}_{k=1}^{m}$ (e.g.,$\{1,2,4\}$). We map each s_k to a normalized timestamp $t_k \in [0,1]$ via min–max normalization, e.g., $s = 1 \mapsto t = 0$, $s = 4 \mapsto t = 1$. We encode each image at scale s_k to obtain $z_{t_k} = E_\theta(I_{s_k})$.

Spline trajectory. To connect $\{z_{t_k}\}$ we adopt a natural cubic spline trajectory $\mu_t(\epsilon)$ defined piecewise on $[t_k,\ t_{k+1}]$ with coefficients solving a tridiagonal system that enforces continuity of the function and its first two derivatives, and natural boundary conditions $\mu''_{t_1}(\varepsilon) = \mu''_{t_m}(\varepsilon) = 0$[5].

Conditional flow matching. We learn a velocity field $v_\phi(x,t)$ to follow the spline's derivative $\mu'_t(\varepsilon)$ using the conditional flow-matching objective[6]:

$$\mathcal{L}_{CFM} = \mathbb{E}_{t\sim\mathcal{U}[0,1],x\sim\sim p_t(x|\epsilon),\epsilon\sim q(\epsilon)}\|v\phi(x,t) - \mu'_t(\varepsilon)\|_2^2,$$

This objective is tractable under a deterministic path with zero intermediate variance, so the model directly regresses the target velocity along the spline.

Taylor-guided perceptual supervision. Intermediate timesteps (e.g., scales 1.532× or 3.361×) lack ground-truth LR. We approximate the latent at such an intermediate t by third-order Taylor extrapolation toward the next trained level t_{k+1}, decode it, and apply an image-domain LPIPS loss to encourage perceptual realism around that neighborhood[7].

4. EXPERIMENTS

4.1 Implementation Details

RAE. The RAE is trained using the Adam optimizer to minimize the reconstruction loss. Training continues for 200k iterations with a cosine-annealed learning rate schedule, decaying from 1×10^{-4} to 1×10^{-7}. Each mini-batch contains 16 randomly cropped 256 ×256 patches with random horizontal and vertical flips for data augmentation.

LFM. The LFM network uses the Adam optimizer to minimize the CFM and LPIPS losses over 400k iterations. A cosine-annealed learning rate schedule decays from 2×10^{-4} to 1×10^{-7}, with mini-batches of 32 randomly cropped 256 ×256 patches and random flips.

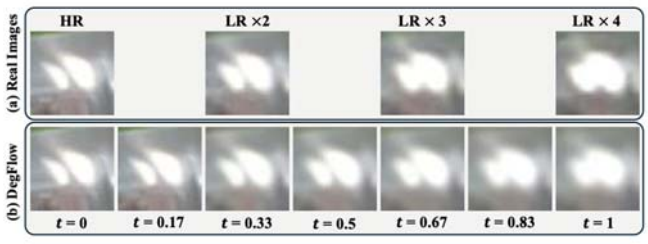
4.2 Datasets

Training. In all experiments, DegFlow is trained on the RealSR-V2 dataset, which contains paired images at degradation levels ×1, ×2, and ×4 from two DSLR camera models: Canon and Nikon. We train on Canon-train dataset and generate LR images from HR images of Nikon-train dataset. to test the robustness of our method.

Evaluation. SR performance is evaluated on real-world benchmarks: RealSR. This dataset collectively covers a wide range of camera, scene, and degradation characteristics, offering a comprehensive generalization evaluation.

4.3 LR Image Generation Results

We first demonstrate the capability of our DegFlow to model continuous real-world degradations in latent space. In Fig. 2 (a), we display RealSR dataset images at discrete scales (HR, ×2, ×3, ×4). Fig. 2 (b) shows LR images generated by DegFlow at uniformly spaced timesteps $0 \le t \le 1$, using the model trained on degradation levels S= {1,2,4}. Our approach achieves smooth and physically consistent transitions between scales, and the synthesized images exhibit gradual variations in blur and detail loss. These transitions closely match the characteristics of both seen levels (×2, ×4) and unseen levels (×3), indicating that DegFlow successfully learns a scale-continuous degradation manifold.



(Fig. 2) Visualization of continuous degradation. (a) Real images from the RealSR dataset at discrete scales (HR, ×2, ×3, ×4). (b) DegFlow-generated intermediate degradations at evenly spaced timesteps $0 \le t \le 1$

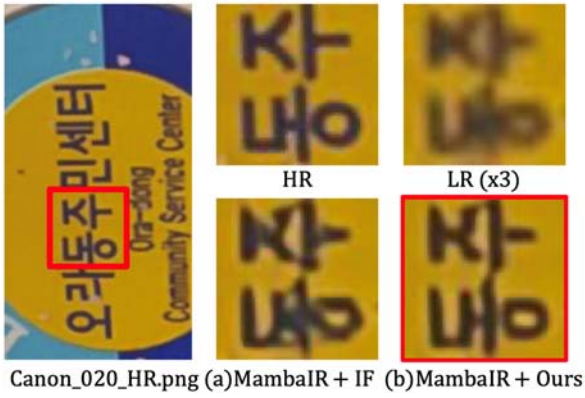
4.4 Super-resolution results

Tab. 1 demonstrates quantitative results on the RealSR ×3 test set, comparing three models: SwinIR, HAT and MambaIR. First, the oracle setting is established by training on RealSR (×3), which directly matches the target degradation level. Next, SR models are trained on RealSR (×2, ×4), without using the target degradation level. Finally, SR models undergo training using synthetic LR images generated by our model (Ours), ranging from ×2 to ×4. Notably, our model is trained on only ×2 and ×4 RealSR datasets, synthesizing intermediate scales including the target scale (×3). In the results, the SR models trained with our synthesized dataset consistently outperform those trained on RealSR (×2, ×4) achieving higher PSNR and SSIM values and better LPIPS. These results validate the effectiveness of our continuous degradation modeling generating realistic and scale-continuous LR images, enabling SR networks to generalize more effectively to unseen target scales.

Displayed in Fig. 3, the fixed-scale SR (MambaIR) is presented, where our generated LR image evidently contribute to enhanced SR results over InterFlow.

<Tab. 1> Fixed-scale SR results on RealSR (×3). Best and second-best highlighted in bold and underline.

Model	Train set	Metric		
		PSNR↑	SSIM↑	LPIPS↓
SwinIR	RealSR(x3)	30.69	0.8647	0.3217
	RealSR(x2, x4)	30.23	0.8597	0.3255
	Ours(x2~ x4)	**30.78**	**0.8658**	<u>0.3193</u>
HAT	RealSR(x3)	30.71	0.8645	0.3221
	RealSR(x2, x4)	30.39	0.8607	0.3248
	Ours(x2~ x4)	**30.86**	**0.8668**	<u>0.3186</u>
MambaIR	RealSR(x3)	30.62	0.8636	0.3208
	RealSR(x2, x4)	30.39	0.8660	0.3240
	Ours(x2~ x4)	**30.86**	**0.8686**	<u>0.3152</u>



(Fig. 3) Qualitative comparisons on RealSR dataset (×3). Fixed scale SR results (MambaIR) with and without our synthetic dataset are compared. IF indicates InterFlow.

5. CONCLUSION

We introduce DegFlow, a novel continuous degradation modeling framework for real-world super-resolution. Unlike previous methods that rely on handcrafted degradation pipelines or require paired low-resolution inputs for generation, DegFlow learns a degradation manifold in latent space from only discrete real-world HR–LR pairs and synthesizes realistic degradations at arbitrary, unseen scales using only high-resolution images. By combining a residual autoencoder with latent flow matching, DegFlow effectively captures the nonlinear geometry of real-world degradations while maintaining explicit degradation level control. Experiments show that SR networks trained on our synthetic datasets consistently outperform those trained with existing generation methods in both fidelity and perceptual quality.

REFERENCES

[1] Chen, Y. et al. Learning continuous image representation with local implicit image function. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 8628–8638. (2021)

[2] Wolf, V. et al. Deflow: Learning complex image degradations from unpaired data with conditional flows. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 94–103. (2021)

[3] Zhang, Y. et al. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European conference on computer vision (ECCV), 286–301.(2018)

[4] Luo, Z. et al. Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 1680–1691. (2023)

[5] De Boor, C. A practical guide to splines, volume 27. springer New York. (1978)

[6] Liu, X. et al. Flow straight and fast: Learning to generate and transfer data with rectified flow. In ICLR. (2023)

[7] Zhang, R. et al. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on computer vision and pattern recognition, 586–595. (2018)