FlowIE: Efficient Image Enhancement via Rectified Flow

Supplementary Material

In the supplementary material, we provide a deeper exploration of insights and findings. In Section A, we present more implementation details regarding the training and evaluation of FlowIE. Section B delves into further discussions through a combination of quantitative analyses and qualitative experiments. In Section C, we show additional visualization results for Blind Face Restoration (BFR) and Blind Image Super-Resolution (BSR). Furthermore, Section D extends our investigations to encompass tasks such as single image deraining and dehazing. The source code for FlowIE is also provided in the zip file.

A. Detailed Implementations

To initialize our path estimator $\boldsymbol{v}_\theta$, we employ the text-to-image pretrained Stable Diffusion (SD 2.0-Base) [17], which offers ample generative priors suitable for various enhancement tasks. The input image $\boldsymbol{x} \in \mathbb{R}^{3\times512\times512}$ is encoded into the latent code $\boldsymbol{z} \in \mathbb{R}^{4\times64\times64}$ by the trained VQGAN. During the training of all tasks, we resize the input images to 512×512. For the images smaller than this size, we upsample them with the short side enlarged to 512 and crop them with a fixed-size bounding box. We train our FlowIE with 8 NVIDIA RTX 3090 GPUs. To maintain the pre-trained capability of the diffusion model, we utilize the LoRA [7] approach to unfreeze the linear layers of the cross-attention blocks in $\boldsymbol{v}_\theta$. We find that the small trainable parameters with a LoRA rank of 4 can significantly unleash the generative priors within the diffusion model and allow adaptation to various tasks. Another benefit of the partially unlocked models is preventing overfitting of the large diffusion model. To measure the throughput of FlowIE and other methods, we conduct evaluation experiments on the same dataset and using a single 3090 GPU.

B. More Discussions

Many-to-one mapping and result diversity. Compared with text-to-image generation, image enhancement tasks like BFR have more deterministic targets. Therefore, we employ the 'many-to-one' strategy for FlowIE during training to learn a direct mapping from noise to real data. However, it's crucial to clarify that FlowIE, being a probabilistic model like diffusion models, inherently yields diverse outcomes, especially for the inpainting task. As illustrated in Figure A, given the masked input (Col.1) and different initial noise $\boldsymbol{z}_0$, FlowIE generates various facial features (Col.2-5), encompassing variations in the shape of the nose, ears, and texture of the hair. Unlike rigid 'many-to-one' mapping often employed in GAN-based methods during inference, FlowIE embraces the generative capacity of diffusion models and enjoys the diversity of plausible results.

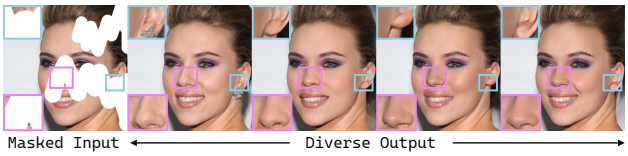

Figure A. **Diverse results of *FlowIE***. Our framework can generate various results with different initial noises.

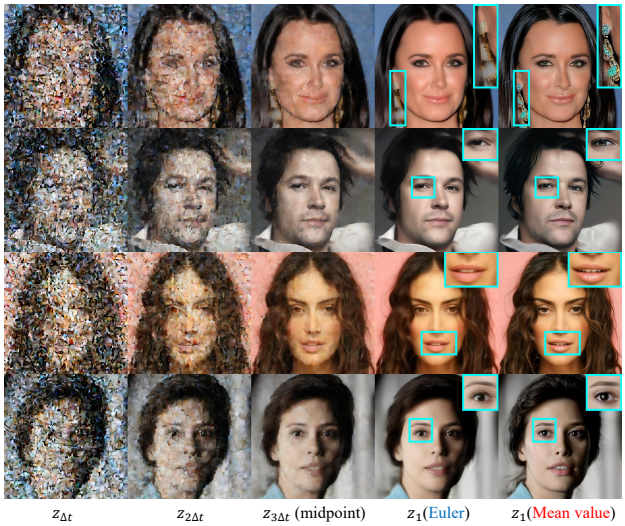

Figure B. **The visualization of the inference process.** FlowIE establishes straight-line paths from random noise to clean images. Through mean value sampling, we achieve clearer and more detailed results in fewer steps compared to the Euler method.

Visualization of different paths We showcase the visualization of each step in our inference process. Along the straight-line path, FlowIE adeptly generates high-quality (HQ) images from noise in less than 5 steps. As depicted in Figure B, the mean value sampling consistently yields clearer and more detailed results in fewer steps compared to the Euler method, highlighting its efficacy in enhancing the quality of the generated images.

About starting from $\boldsymbol{\tau}_\phi(\boldsymbol{z}_{\mathrm{LQ}})$. Since FlowIE predicts the path from random noise, switching the starting point to the coarse result $\boldsymbol{\tau}_\phi(\boldsymbol{z}_{\mathrm{LQ}})$ is indeed reasonable. Tuning and evaluation on the BFR task (shown in Table A) indicate a slightly worse FID compared with FlowIE. We attribute this result to the adjustment's reliance on initial results over pre-trained diffusion priors.

About artifacts in the first step. We acknowledge that extreme artifacts in the first step may result in failure cases.

Figure C. **Qualitative comparisons on CelebA-Test.** FlowIE produces high-quality results with rich details and maintains high identity similarity, even when confronted with severely degraded inputs, while previous methods exhibit visible artifacts or inconsistent faces.

Figure D. **Qualitative comparisons on real-world faces.** Our approach demonstrates credible enhancements on real-world faces, delivering high-fidelity and visually satisfying results. Compared to other methods, FlowIE showcases robustness in front of challenging cases.

Table A. **Ablation study about the starting point.** Latent initiation from $\boldsymbol{z}_0 = \boldsymbol{\tau}_\phi(\boldsymbol{z}_{\mathrm{LQ}})$ leads to worse FID.

Method	FID↓	
	CelebA	LFW
DiffBIR [11]	20.19	39.61
$\boldsymbol{z}_0 = \boldsymbol{\tau}_\phi(\boldsymbol{z}_{\mathrm{LQ}})$	19.87	38.80
FlowIE	**19.81**	**38.66**

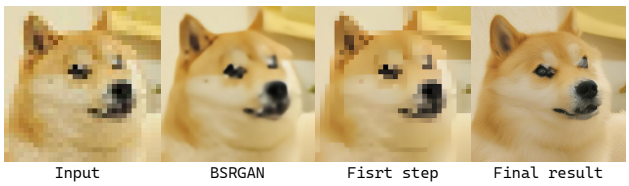

Figure E. **Failure case of *FlowIE*.** Our framework may give unsatisfying results when facing severe degradation.

In Figure E, the input undergoes challenging degradation (16× downsampling). Compared to GAN-based methods like BSRGAN [28] which introduce many artifacts and blur, FlowIE generates a cleaner image. However, the final result may still exhibit unrealistic eyes due to initial step artifacts.

Figure F. **Qualitative results in larger resolution.** The proposed FlowIE consistently delivers visually captivating results at higher resolutions.

About larger resolution. FlowIE demonstrates excellent scalability to process larger images. We can replace the original diffusion model (SD 2.0-base) with an enlarged version (SDXL) which generates 1024×1024 images by default and tune the FlowIE framework following the proposed method. As shown in Figure F, despite the limitation in training time, we still obtain satisfying outcomes with higher resolution (1024×1024).

Comparisons with diffusion models. Our proposed FlowIE mainly capitalizes on the powerful generation capability within the pre-trained diffusion model, which has demonstrated its versatility in various visual tasks. For example, DDVM [18] explicitly underscores the effectiveness of pre-trained priors in diffusion models for monocular depth estimation and SDEidt [14] focuses on image editing tasks like stroke-based editing. Additionally, [24] successfully achieves rapid image sampling by employing multi-modal denoising distributions and conditional GANs. Compared with these works, our FlowIE primarily harnesses the generative prior in diffusion models and employs a conditioned flow-based strategy to accelerate the sampling.

C. More Qualitative Comparisons

In this section, we provide additional visual comparisons on BFR and BSR with state-of-the-art methods. Our framework reliably demonstrates its ability to deliver robust and satisfying results in these challenging tasks, showcasing its efficacy across diverse image enhancement scenarios.

Blind Face Restoration. We conduct qualitative comparisons on both synthetic CelebA-Test [13] and in-the-wild LFW-Test [20], CelebChild-Test [20] and WIDER-Test [29]. Our comparisons involve recent state-of-the-art methods, including GPEN [25], GCFSR [6], GFP-GAN [20], VQFR [5], RestoreFormer [23], DMDNet [9], CodeFormer [29] and DiffBIR [11]. Visual results presented in Figure C and Figure D demonstrate that our FlowIE consistently produces visually pleasing outcomes on both synthetic and real-world datasets, affirming its effectiveness and robust performance in diverse scenarios.

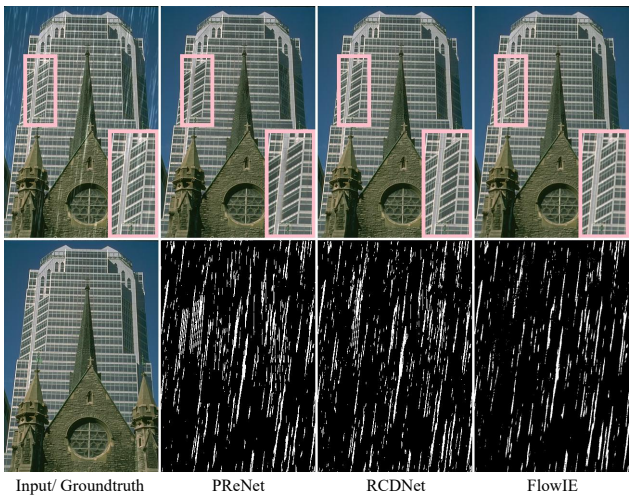

Figure G. **Single image Deraining via *FlowIE*.** Our framework adeptly identifies the rainy layers and proficiently restores the original images without complex task-specific designs.

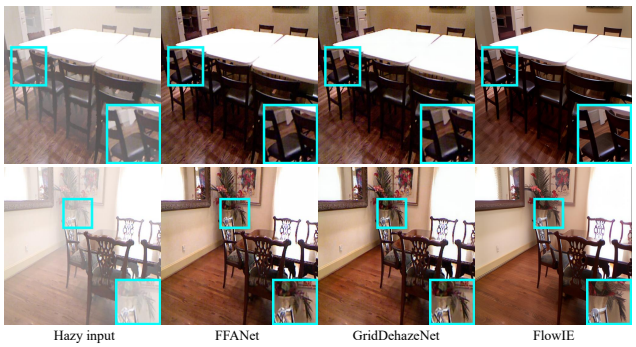

Figure H. **Single image dehazing via *FlowIE*.** Our framework effectively eliminates haze, enhancing the overall clarity of the images.

Blind Image Super-Resolution. For BSR, we also present additional results on RealSRSet [1] and our established Collect-100 dataset. We compare FlowIE with cutting-edge methods, including GAN-based Real-ESRGAN+ [21], BSRGAN [28], SwinIR-GAN [10], FeMaSR [2] and diffusion-based DDNM [22], GDP [3] and DiffBIR [11]. Figure I vividly illustrates the efficacy of FlowIE in generating visually appealing images with a commendable balance between realism and clarity.

D. More Extended Tasks

To showcase the versatility of our framework, we extend FlowIE to more tasks, specifically single image deraining and dehazing. The adaptation for these tasks involves a fine-tuning process with 15K steps on the respective datasets. Notably, we only use the single MSE loss for all tasks.

Deraining. We utilize RainTrainH [26], RainTrainL [26] and Rain12600 [4] for training and evaluate our framework

Figure I. **Qualitative comparisons on the real-world images.** FlowIE successfully enhances the LQ images through simultaneous upsampling, denoising, and deblurring, and provides rich details from the generative knowledge, leveraging generative knowledge to deliver high-quality outcomes with rich details.

on Rain-100L dataset [27]. We compare our results with PReNet [16] and RCDNet [19]. As shown in Figure G, FlowIE effectively separates the rainy layers and reconstructs the original clean images.

Dehazing. We employ the indoor part of the RESIDE dataset [8] for training and evaluate our framework on its test split. We compare the results with FFA-Net [15] and GridDehazeNet [12]. FlowIE demonstrates successful haze removal and enhances the clarity of the original images, as shown in Figure H.

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*, pages 3086–3095, 2019. 3

[2] Chaofeng Chen, Xinyu Shi, Yipeng Qin, Xiaoming Li, Xiaoguang Han, Tao Yang, and Shihui Guo. Real-world blind super-resolution via feature matching with implicit high-resolution priors. In *ACMMM*, pages 1329–1338, 2022. 3

[3] Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *CVPR*, pages 9935–9946, 2023. 3

[4] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *CVPR*, pages 3855–3863, 2017. 3

[5] Yuchao Gu, Xintao Wang, Liangbin Xie, Chao Dong, Gen Li, Ying Shan, and Ming-Ming Cheng. Vqfr: Blind face restoration with vector-quantized dictionary and parallel decoder. In *ECCV*, pages 126–143. Springer, 2022. 3

[6] Jingwen He, Wu Shi, Kai Chen, Lean Fu, and Chao Dong. Gcfsr: a generative and controllable face super resolution method without facial and gan priors. In *CVPR*, pages 1889–1898, 2022. 3

[7] 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. *arXiv preprint arXiv:2106.09685*, 2021. 1

[8] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *TIP*, 28(1):492–505, 2018. 4

[9] Xiaoming Li, Shiguang Zhang, Shangchen Zhou, Lei Zhang, and Wangmeng Zuo. Learning dual memory dictionaries for blind face restoration. *TPAMI*, 45(5):5904–5917, 2022. 3

[10] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV*, pages 1833–1844, 2021. 3

[11] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Ben Fei, Bo Dai, Wanli Ouyang, Yu Qiao, and Chao Dong. Diffbir: Towards blind image restoration with generative diffusion prior. *arXiv preprint arXiv:2308.15070*, 2023. 2, 3

[12] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *CVPR*, pages 7314–7323, 2019. 4

[13] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *ICCV*, pages 3730–3738, 2015. 3

[14] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. In *ICLR*, 2021. 3

[15] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for

single image dehazing. In *AAAI*, pages 11908–11915, 2020. 4

[16] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *CVPR*, pages 3937–3946, 2019. 4

[17] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pages 10684–10695, 2022. 1

[18] Saurabh Saxena, Charles Herrmann, Junhwa Hur, Abhishek Kar, Mohammad Norouzi, Deqing Sun, and David J Fleet. The surprising effectiveness of diffusion models for optical flow and monocular depth estimation. *NeurIPS*, 36, 2024. 3

[19] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. In *CVPR*, pages 3103–3112, 2020. 4

[20] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards real-world blind face restoration with generative facial prior. In *CVPR*, pages 9168–9178, 2021. 3

[21] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *ICCV*, pages 1905–1914, 2021. 3

[22] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *ICLR*, 2022. 3

[23] Zhouxia Wang, Jiawei Zhang, Runjian Chen, Wenping Wang, and Ping Luo. Restoreformer: High-quality blind face restoration from undegraded key-value pairs. In *CVPR*, pages 17512–17521, 2022. 3

[24] Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion gans. In *ICLR*, 2021. 3

[25] Tao Yang, Peiran Ren, Xuansong Xie, and Lei Zhang. Gan prior embedded network for blind face restoration in the wild. In *CVPR*, pages 672–681, 2021. 3

[26] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *CVPR*, pages 1357–1366, 2017. 3

[27] Wenhan Yang, Robby T Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. Joint rain detection and removal from a single image with contextualized deep networks. *TPAMI*, 42(6):1377–1393, 2019. 4

[28] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *ICCV*, pages 4791–4800, 2021. 2, 3

[29] Shangchen Zhou, Kelvin Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. *NeurIPS*, 35: 30599–30611, 2022. 3