LearnIR: Learnable Posterior Sampling for Real-World Image Restoration

Yihang Bao[1,†] Zhen Huang[2,†] Shanyan Guan[2] Songlin Yang[2] Yanhao Ge[2] Wei Li[2] Bukun Huang[3] Zengmin Xu[1,4,5,*]

[1]Guilin University of Electronic Technology, [2]vivo Mobile Communication Co., Ltd, [3]Zhejiang Gongshang University, [4]Center for Applied Mathematics of Guangxi (GUET), [5]Anview.ai

†Equal contributions, *Corresponding author: xzm@guet.edu.cn

Abstract

Image restoration in real-world conditions is highly challenging due to heterogeneous degradations such as haze, noise, shadows, and blur. Existing diffusion-based methods remain limited: conditional generation struggles to balance fidelity and realism, inversion-based approaches accumulate errors, and posterior sampling requires a known forward operator that is rarely available. We introduce **LearnIR**, a learnable diffusion posterior sampling framework that eliminates this dependency by training a lightweight model to directly predict gradient correction distributions, enabling *Diffusion Posterior Sampling Correction (DPSC)* that maintains consistency with the true image distribution during sampling. In addition, a *Dynamic Resolution Module (DRM)* dynamically adjusts resolution to preserve global structures in early stages and refine fine textures later, while avoiding the need for a pretrained VAE. Experiments on ISTD, O-HAZE, HazyDet, REVIDE, and our newly constructed FaceShadow dataset show that LearnIR achieves state-of-the-art performance in PSNR, SSIM, and LPIPS.

Architecture

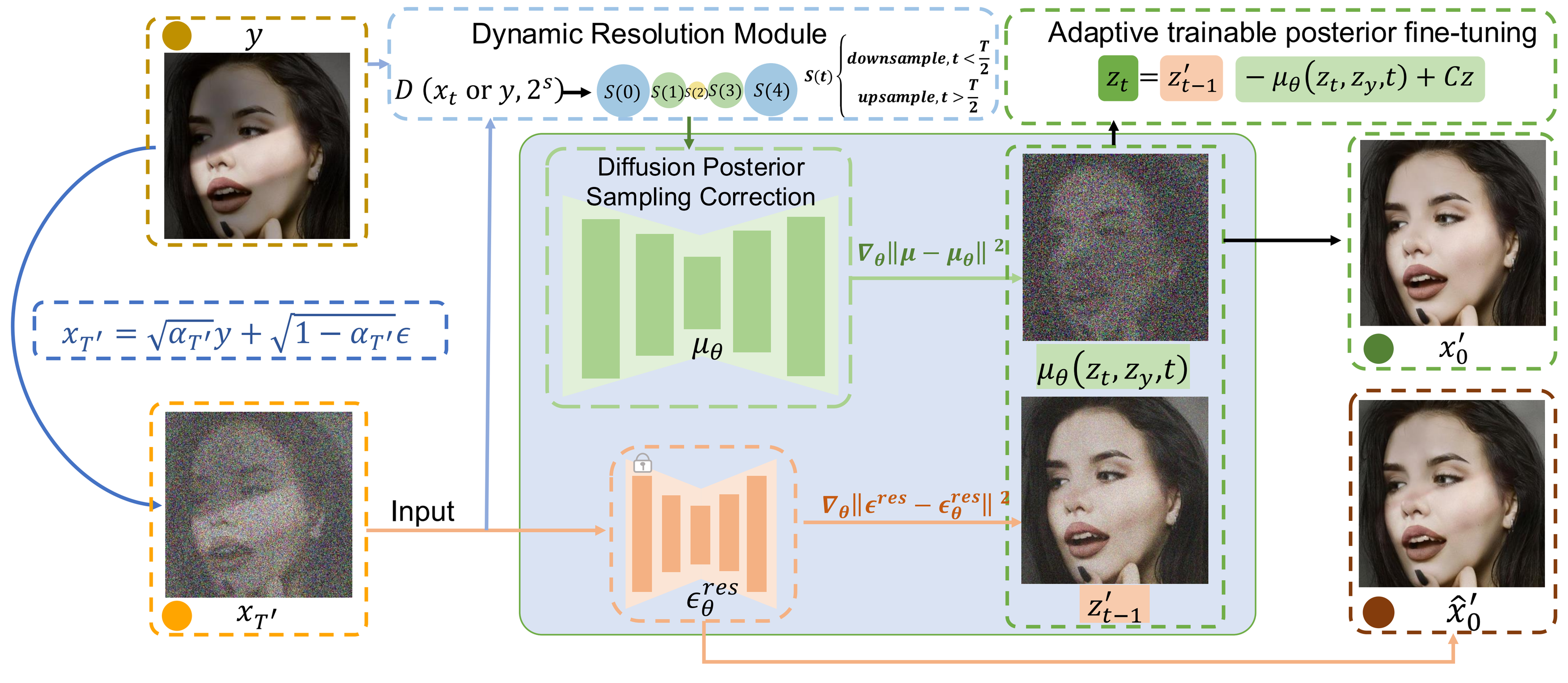

Figure 1. **Overview of the LearnIR Framework.** The smooth equivalence transformation identifies the optimal sampling step T' (blue line) to bypass the unstable intermediate states derived from observation y. The proposed framework consists of two key components: 1) The **Dynamic Resolution Module (DRM)** projects inputs into a time-dependent latent space via $\mathcal{D}(\cdot, 2^s)$ to balance structural coherence and detail recovery; 2) The **Diffusion Posterior Sampling Correction (DPSC)** acts as a plug-and-play regularization to analytically eliminate the trajectory inconsistency (structural bias).

Methodology

we present **LearnIR**, a robust framework for residual-based image restoration. As illustrated in Fig. 1, our approach integrates two complementary mechanisms:

1. **Dynamic Resolution Module (DRM)** that constructs a resolution-aware latent space to suppress noise and redundancy,
2. **Diffusion Posterior Sampling Correction (DPSC)** that operates within this space to align the diffusion trajectory with the true posterior.

Dynamic Resolution Module (DRM)

Drawing inspiration from multi-scale generation frameworks such as MDM [2] and Pixelflow [1], we introduce a Dynamic Resolution Module (DRM) to enable adaptive processing across spatial scales. Specifically, we define a time-dependent downsampling operator $\mathcal{D}(\cdot, s(t))$ that maps the clean image $\boldsymbol{x}_0$ and degraded image $\boldsymbol{y}$ into a variable-resolution latent space:

$$\boldsymbol{z}_0^{(t)} = \mathcal{D}(\boldsymbol{x}_0, s(t)), \quad \boldsymbol{z}_y^{(t)} = \mathcal{D}(\boldsymbol{y}, s(t)). \tag{1}$$

Within this space, we extend the residual learning formulation by defining the latent residual $\boldsymbol{R_z} = \boldsymbol{z}_y^{(t)} - \boldsymbol{z}_0^{(t)}$, and the corresponding forward diffusion process becomes:

$$q(\boldsymbol{z}_t \mid \boldsymbol{z}_0^{(t)}) = \mathcal{N}\left(\boldsymbol{z}_t;\ \sqrt{\bar{\alpha}_t}\boldsymbol{z}_0^{(t)} + (1-\sqrt{\bar{\alpha}_t})\boldsymbol{R_z},\ (1-\bar{\alpha}_t)\mathbf{I}\right). \tag{2}$$

The schedule $s(t)$ naturally forms a coarse-to-fine sampling trajectory: high-noise stages operate at lower resolutions to capture global structure, while later stages restore fine details at the native scale. This design allows LearnIR to effectively leverage degradation priors across scales without introducing additional trainable components.

Diffusion Posterior Sampling Correction (DPSC)

To correct posterior inconsistency in diffusion sampling, we propose Diffusion Posterior Sampling Correction (DPSC). Although the standard denoising objective

$$\mathcal{L}_{\text{denoise}} = \mathbb{E}_{\boldsymbol{z}_0^{(t)},\boldsymbol{\epsilon},t}\|\boldsymbol{\epsilon} - \epsilon_\theta(\boldsymbol{z}_t, t)\|_2^2 \tag{3}$$

ensures accurate noise prediction, it does not guarantee that the learned reverse posterior $p_\theta(\boldsymbol{z}_{t-1} \mid \boldsymbol{z}_t)$ matches the forward posterior $q(\boldsymbol{z}_{t-1} \mid \boldsymbol{z}_t, \boldsymbol{z}_0^{(t)})$.

In the DRM latent space, this posterior discrepancy can be written as

$$\nabla_{\boldsymbol{z}_t} \log p(\boldsymbol{z}_y^{(t)} \mid \boldsymbol{z}_t) \propto \boldsymbol{z}_{t-1}^{\text{pred}} - \boldsymbol{z}_{t-1}^{\text{forward}}. \tag{4}$$

Since $\boldsymbol{\epsilon}$ is Gaussian and the remaining terms are deterministic functions of $(\boldsymbol{z}_t, \boldsymbol{z}_y^{(t)}, t)$, the offset follows

$$\boldsymbol{z}_{t-1}^{\text{pred}} - \boldsymbol{z}_{t-1}^{\text{forward}} \sim \mathcal{N}\left(\boldsymbol{\mu}(\boldsymbol{z}_t, \boldsymbol{z}_y^{(t)}, t), \sigma^2(\boldsymbol{z}_t, \boldsymbol{z}_y^{(t)}, t)\mathbf{I}\right). \tag{5}$$

The analytic posterior mean is

$$\boldsymbol{\mu}(\boldsymbol{z}_t, \boldsymbol{z}_y^{(t)}, t) = \left(\frac{1}{\sqrt{\alpha_t}} - \frac{2\sqrt{\bar{\alpha}_{t-1}} - 1}{\sqrt{\bar{\alpha}_t}}\right)\boldsymbol{z}_t - (1-\sqrt{\bar{\alpha}_{t-1}})\boldsymbol{z}_y^{(t)} - \frac{1-\alpha_t}{\sqrt{\alpha_t(1-\bar{\alpha}_t)}}\boldsymbol{\epsilon}_\theta^{\text{res}}(\boldsymbol{z}_t, \boldsymbol{z}_y^{(t)}, t), \tag{6}$$

with variance

$$\sigma^2(\boldsymbol{z}_t, \boldsymbol{z}_y^{(t)}, t) = \left(\sqrt{\frac{(1-\bar{\alpha}_{t-1})(1-\alpha_t)}{1-\bar{\alpha}_t}} - \sqrt{1-\bar{\alpha}_{t-1}} + \frac{(2\sqrt{\bar{\alpha}_{t-1}}-1)(1-\alpha_t)}{\sqrt{\bar{\alpha}_t(1-\bar{\alpha}_t)}}\right)^2. \tag{7}$$

We learn to predict this posterior correction and regularize its mean via

$$\mathcal{L}_{\text{consistency}} = \mathbb{E}\left\|\boldsymbol{\mu}(\boldsymbol{z}_t, \boldsymbol{z}_y^{(t)}, t) - \hat{\boldsymbol{\mu}}_\theta(\boldsymbol{z}_t, \boldsymbol{z}_y^{(t)}, t)\right\|_2^2, \tag{8}$$

leading to

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{denoise}} + \lambda\mathcal{L}_{\text{consistency}}, \tag{9}$$

Experiments

We conduct experiments on standard shadow removal and dehazing benchmarks, as well as our newly constructed FaceShadow dataset. Due to space limitations, we present the experimental results on the FaceShadow dataset in this poster.

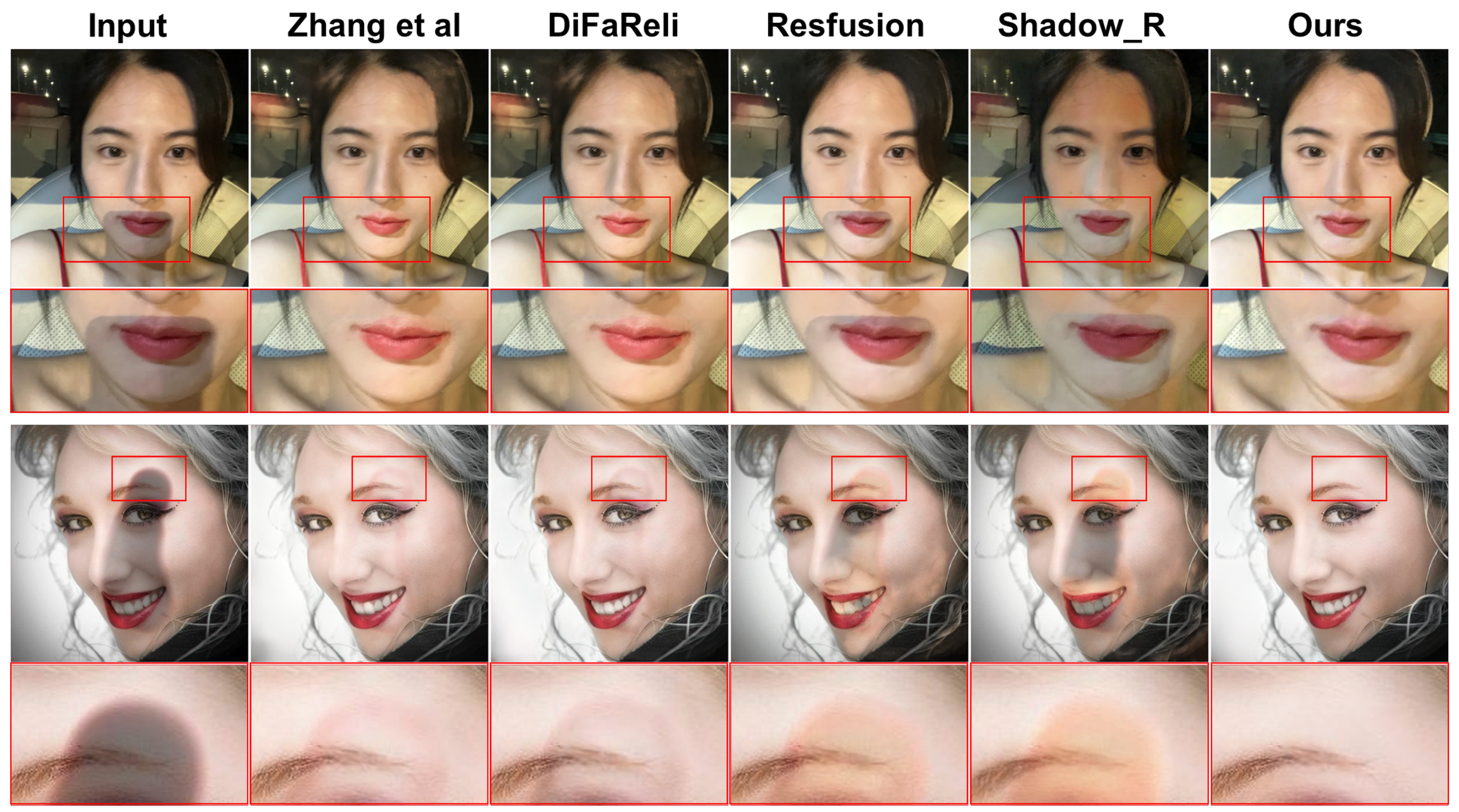

Figure 2. Visual comparisons on 1,000 real-world face shadow images from the FaceShadow test set.

Figure 3. **Real face shadow data** Samples of real-world facial images with shadows, collected from the Internet. These images feature diverse identities and scene conditions, demonstrating realistic and challenging shadow patterns.

Overall, the visual comparisons demonstrate that **LearnIR** achieves highly competitive performance in real-world facial shadow restoration. Our method effectively removes complex and spatially varying shadows while preserving identity-specific details, skin texture, and structural consistency. Even under challenging illumination conditions, LearnIR produces visually natural and artifact-free results, highlighting its strong robustness and generalization ability for real-world portrait restoration.

References

[1] Shoufa Chen, Chongjian Ge, Shilong Zhang, Peize Sun, and Ping Luo. Pixelflow: Pixel-space generative models with flow. *arXiv preprint arXiv:2504.07963*, 2025.

[2] Jiatao Gu, Shuangfei Zhai, Yizhe Zhang, Josh Susskind, and Navdeep Jaitly. Matryoshka diffusion models. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2024.