Self-supervised Dynamic Heterogeneous Degradation Modeling for Unified Zero-Shot Image Restoration

Supplementary Material

1. Additional Finding Analysis

1.1. Degradation Dataset Construction

We constructed degraded datasets by introducing noise, haze, and low light to clean images from three datasets: Kodak24 [2], the HSTS subset of RESIDE [4], and LOLv1 [8]. The degradation processes are implemented as follows:

- **Noise:** Gaussian noise is added to clean images which are normalized to the range $[0, 1]$, as defined below:

$$x_{lq} = \text{clip}(x_{hq} + N(0, \sigma^2)), \tag{1}$$

  where σ controls the noise intensity (*Light*: $\sigma = 20/255$, *Heavy*: $\sigma = 50/255$), and clip$(\cdot)$ ensures that pixel values remain within the valid range [0, 1].
- **Haze:** hazy images are generated based on the atmospheric scattering model:

$$\mathbf{x}_{lq}(\mathbf{p}) = \mathbf{x}_{hq}(\mathbf{p}) \cdot t(\mathbf{p}) + \mathbf{A} \cdot (1 - t(\mathbf{p})), \tag{2}$$

  where $t(\mathbf{p}) = \exp(-\beta \cdot d(\mathbf{p}))$. Here, $\mathbf{A}$ denotes the global atmospheric light, which is fixed as $\mathbf{A} = 1$ for all images, and β controls the haze density (*Light*: $\beta \in [0.02, 0.05]$, *Heavy*: $\beta \in [0.06, 0.09]$). The variable $d(\mathbf{p})$ represents a pseudo depth map that simulates the scene distance at pixel position $\mathbf{p} = (i, j)$, defined as:

$$d(i, j) = \frac{\sqrt{(i - c_x)^2 + (j - c_y)^2}}{s}, \tag{3}$$

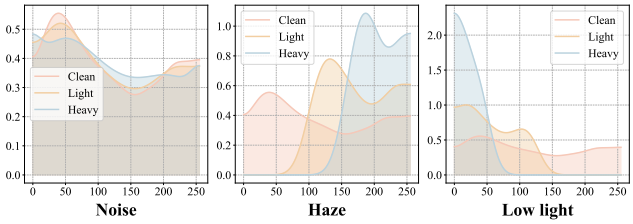
  where (c_x, c_y) denotes the image center and s is the normalization scale. For images with height row and width col, we set the scale as $s = \sqrt{\max(\text{row}, \text{col})}$ and the center as $(c_x, c_y) = (\text{row}/2, \text{col}/2)$.
- **Low Light:** low-light images are generated by linearly scaling the brightness of clean images in the HSV color space. Given a clean image $\mathbf{x}_{hq}$ with HSV channels $(\mathbf{H}, \mathbf{S}, \mathbf{V})$, the low-light image $\mathbf{x}_{lq}$ is obtained as:

$$\begin{aligned} \mathbf{V}_{lq} &= \text{clip}(\gamma \cdot \mathbf{V}_{hq}), \\ \mathbf{x}_{lq} &= \text{HSV2RGB}(\mathbf{H}, \mathbf{S}, \mathbf{V}_{lq}), \end{aligned} \tag{4}$$

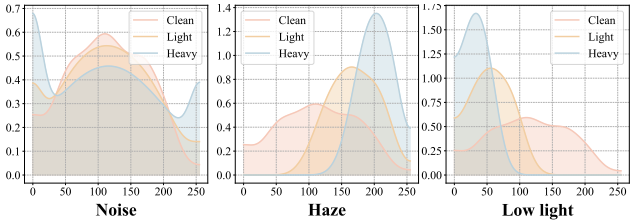
  where $\gamma \in (0, 1)$ is a scaling factor controlling the illumination level. Specifically, $\gamma = 0.7$ for light degradation and $\gamma = 0.3$ for heavy degradation.

1.2. Visualization Results on Other Datasets

In the main paper, we focused on statistical analyses on the Kodak24 dataset. This section extends our observations to the HSTS subset of RESIDE and the LOLv1 datasets. As shown in Fig. 1, heterogeneous degradations induce consistently strong divergences and systematic shifts in pixel-domain distributions across all datasets due to the distinct underlying mechanisms. Fig. 2 demonstrates that our physically coherent degradation representation provides a unified and physically grounded parameterization of degraded images. This enables homogeneous modeling of heterogeneous degradations and further supports the consistency of the findings reported in the main paper.



(a) Distribution of the HSTS dataset for image dehazing.



(b) Distribution of the LOLv1 dataset for low-light enhancement.

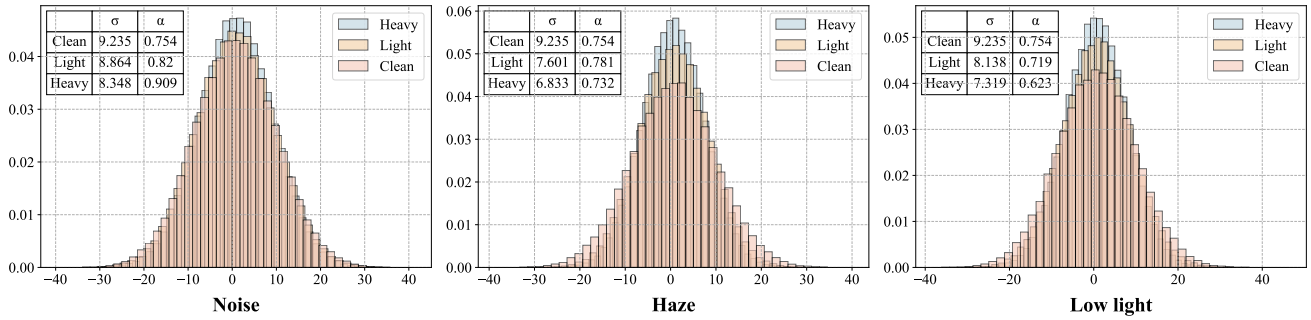
Figure 1. **Pixel-domain distributions across various degradation types and intensities.** The x-axis and y-axis represent pixel values and probability density, respectively. Each degradation type is analyzed across three intensities: clean, light, and heavy.

Task	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
Low-light enhancement	1e-5	1e0	2e-3	5e-5	1e-2	1e-2
Image dehazing	1e-3	1e0	5e-3	5e-5	1e-2	1e-3
Image denoising	1e-3	1e0	5e-4	5e-5	1e-5	1e-5

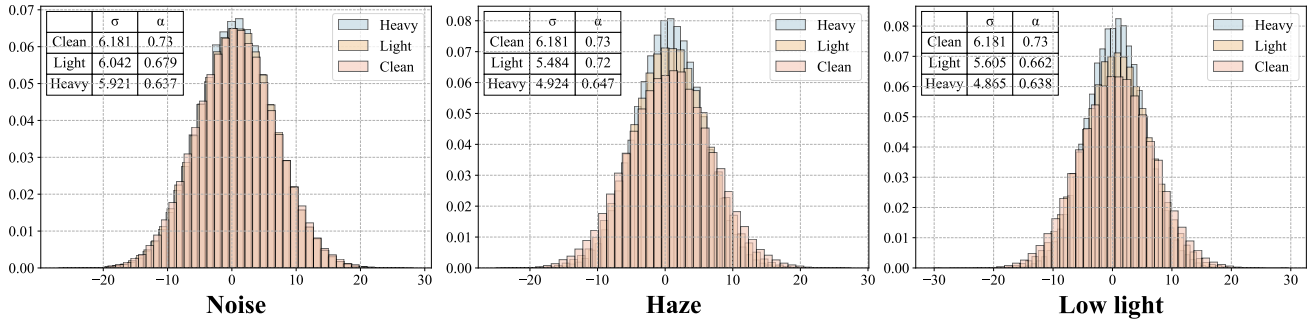
Table 1. **Loss weighting factors for different restoration tasks.**

2. More Implementation Details

In all experiments, we adopt the pre-trained Stable Diffusion model and set the timestep to $T = 1000$. The diffusion inference process is divided into three stages. In the first stage, $t \in [1000, 700)$, we train PCDM exclusively using the Adam optimizer with a learning rate of 1×10^{-5}, enabling accurately modeling the physically



(a) Distribution of the HSTS dataset for image dehazing.



(b) Distribution of the LOLv1 dataset for low-light enhancement.

Figure 2. **Homogeneous degradation distribution.** The x-axis represents the latent representation, while the y-axis denotes the probability density. Parameters σ and α define the dispersion and tail behavior of the distribution, respectively.

Methods	Property			LOLv1					LOLv2				
	B	U	Z	PSNR↑	SSIM↑	LPIPS↓	PI↓	NIQE↓	PSNR↑	SSIM↑	LPIPS↓	PI↓	NIQE↓
Supervised All-in-one													
AirNet [5]	✓	✗	✗	7.07	0.121	0.813	8.06	8.29	9.62	0.181	0.534	8.13	9.30
PromptIR [7]	✓	✗	✗	7.73	0.171	0.549	9.70	11.09	9.73	0.189	0.534	9.61	11.61
DiffUIR [9]	✓	✗	✗	**21.36**	**0.907**	**0.125**	**4.68**	**5.95**	**26.14**	**0.898**	**0.114**	**5.26**	**7.34**
Posterior sampling													
GDP [1]	✗	✓	✓	16.86	0.689	0.299	4.67	5.91	15.32	0.597	0.337	5.13	8.18
TAO [3]	✓	✓	✓	17.42	0.792	0.320	6.03	7.74	16.78	0.747	**0.314**	6.41	9.44
LD-RPS [6]	✓	✓	✓	17.26	0.797	0.291	5.02	5.79	18.22	0.744	0.335	5.05	6.03
Ours	✓	✓	✓	**18.21**	**0.823**	**0.241**	**4.65**	**5.47**	**19.20**	**0.761**	0.332	**4.98**	**5.72**

Table 2. **Quantitative results of low-light enhancement on the LOLv1 and LOLv2 datasets.** The best results within each category are boldfaced. Methods are grouped by three properties: B (task-blind), U (unsupervised), and Z (zero-shot), reflecting their adaptability.

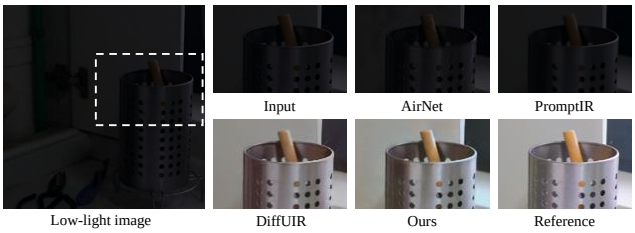


Figure 3. **Qualitative performance of low-light enhancement on the LOLv1 dataset, corresponding to the main paper.**

coherent degradation transformation. During the second stage, $t \in [700, 150)$, we update the posterior sampling trajectory by incorporating the degradation alignment loss J_{deg} along with additional image distance terms. In the final stage, $t \in [150, 0)$, we further introduce the image quality term Q to enhance the perceptual fidelity of the restored output. The corresponding loss weighting factors, which are fine-tuned for each restoration task to ensure optimal parameter settings, are summarized in Tab. 1.

For the dynamic quality-refinement strategy, the refinement depth is set to an intermediate diffusion step, specifically $t' = 500$. In the mixed-degradation experiments, we

Methods	Property			HSTS		
	B	U	Z	PSNR↑	SSIM↑	LPIPS↓
Supervised All-in-one						
AirNet [5]	✓	✗	✗	24.37	0.899	0.059
PromptIR [7]	✓	✗	✗	25.67	0.907	0.048
DiffUIR [9]	✓	✗	✗	**26.88**	**0.914**	**0.045**
Posterior sampling						
GDP [1]	✗	✓	✓	12.57	0.703	0.164
TAO [3]	✓	✓	✓	15.66	0.775	0.216
LD-RPS [6]	✓	✓	✓	20.48	0.804	0.173
Ours	✓	✓	✓	**21.51**	**0.820**	**0.163**

Table 3. **Quantitative results of image dehazing on the HSTS subset of the RESIDE dataset.** The best results are boldfaced.

Methods	Property			Kodak24		
	B	U	Z	PSNR↑	SSIM↑	LPIPS↓
Supervised All-in-one						
AirNet [5]	✓	✗	✗	29.94	0.834	0.114
PromptIR [7]	✓	✗	✗	**30.88**	**0.873**	**0.113**
DiffUIR [9]	✓	✗	✗	22.86	0.789	0.219
Posterior sampling						
GDP [1]	✗	✓	✓	22.37	0.715	0.244
TAO [3]	✓	✓	✓	27.12	0.768	0.222
LD-RPS [6]	✓	✓	✓	27.66	0.830	0.176
Ours	✓	✓	✓	**28.51**	**0.845**	**0.155**

Table 4. **Quantitative results of image denoising on the Kodak24 dataset.** The best results are highlighted in bold.

Figure 4. **Qualitative performance of image dehazing on the HSTS dataset, corresponding to the main paper.**

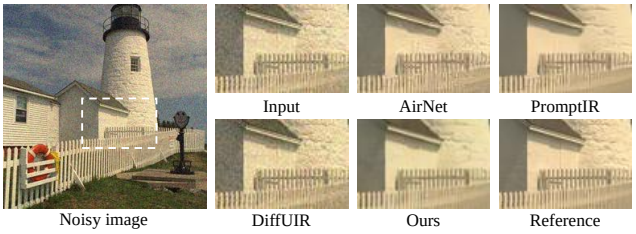


Figure 5. **Qualitative performance of image denoising on the Kodak24 dataset, corresponding to the main paper.**

construct datasets by introducing light-level atmospheric-scattering haze and Gaussian noise into the low-light images of the LOLv1 dataset, following the degradation procedures described above. For a comprehensive and fair evaluation, we include three categories of comparison baselines: supervised all-in-one methods, task-specific methods, and zero-shot posterior-sampling methods.

3. Additional Experimental Results

Single Degradation. The main paper presents comparisons of our method with task-specific and posterior sampling approaches on low-light enhancement, image dehazing, and image denoising tasks. In this section, we provide the evaluation results for supervised all-in-one methods, including AirNet [5], PromptIR [7], and DiffUIR [9], which correspond to the results reported in the main paper. As shown in Tabs. 2 to 4, our method achieves quantitative results comparable to these supervised all-in-one methods on several metrics. Qualitative results in Figs. 3 to 5 illustrate that our method delivers visual fidelity comparable to, and even surpassing, that of supervised all-in-one models. We further provide additional visual comparisons in Figs. 6 to 8. Compared with other baselines, our method achieves state-of-the-art performance while avoid common adverse effects, such as over-smoothing, color distortions, and structural loss, thereby effectively balancing between degradation removal and visual fidelity.

Moreover, due to the limited generalization ability of supervised methods, they often struggle to handle unseen degradations during training. The poor performance of AirNet and PromptIR on low-light enhancement, as well as the suboptimal results of DiffUIR on image denoising, further highlight the superiority of our zero-shot framework.

Mixed Degradation. To intuitively illustrate the robustness under mixed degradations, we present the visual results of our UP-ZeroIR alongside those of other posterior-sampling methods. As shown in Fig. 9 and Fig. 10, our method simultaneously handles low light, noise, and haze under the combined effects of multiple degradations, significantly improving image quality and naturalness for robust restoration.

4. Discussion

Model Efficiency. Compared with supervised methods, our method eliminates the computational burden of training and reduces the reliance on paired datasets, thereby substantially lowering the overall resource consumption. Our method achieves a lightweight design with only 1M parameters and leverages a physically coherent degradation model to adaptively handle various degradations. For a fair comparison, we evaluate our approach against the state-of-the-art LD-RPS, as both methods are built on the pre-trained Stable Diffusion model. With the proposed dynamic

quality-refinement strategy, our method facilitates a more efficient diffusion inference process, achieving an approximate 30% reduction in inference time.

Task Adaptability. Our method demonstrates strong generalizability and can be effectively transferred to other image restoration tasks. As analyzed in the main paper, heterogeneous degradations can be universally parameterized as a physically consistent distribution in the latent diffusion space. During the diffusion inference, our method adaptively optimizes the degradation distribution of the input image, progressively converging to the clean image, regardless of the type of degradation. This enables our UP-ZeroIR to adapt to a variety of restoration tasks while maintaining superior and robust restoration performance. Furthermore, our proposed physically coherent degradation model operates as a plug-and-play solution, demonstrating its significant potential for unified degradation modeling and alignment across a variety of image-related tasks.

References

[1] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9935–9946, 2023. 2, 3

[2] Rich Franzen. Kodak lossless true color image suite, 1999. 1

[3] Yuanbiao Gou, Haiyu Zhao, Boyun Li, Xinyan Xiao, and Xi Peng. Test-time degradation adaptation for open-set image restoration. In *International Conference on Machine Learning*, pages 16167–16177, 2024. 2, 3

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

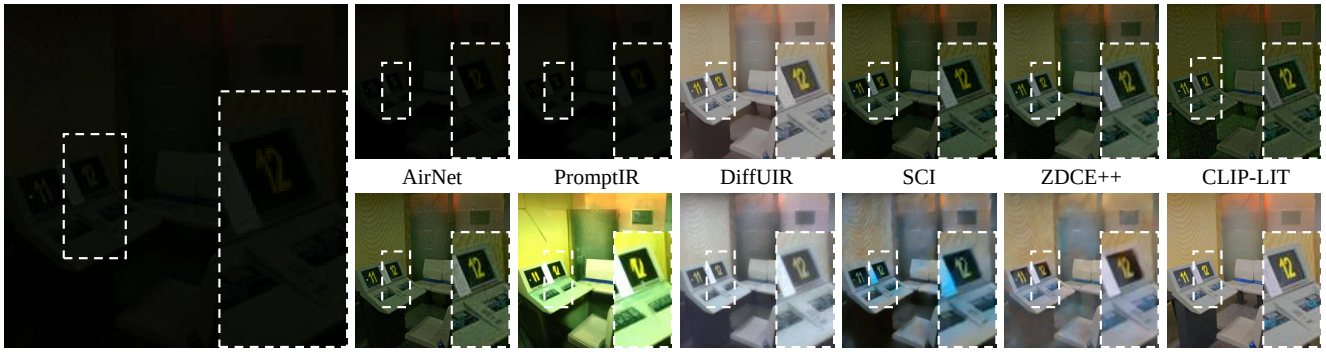
[5] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17452–17462, 2022. 2, 3

[6] Huaqiu Li, Yong Wang, Tongwen Huang, Hailang Huang, Haoqian Wang, and Xiangxiang Chu. Ld-rps: Zero-shot unified image restoration via latent diffusion recurrent posterior sampling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025. 2, 3

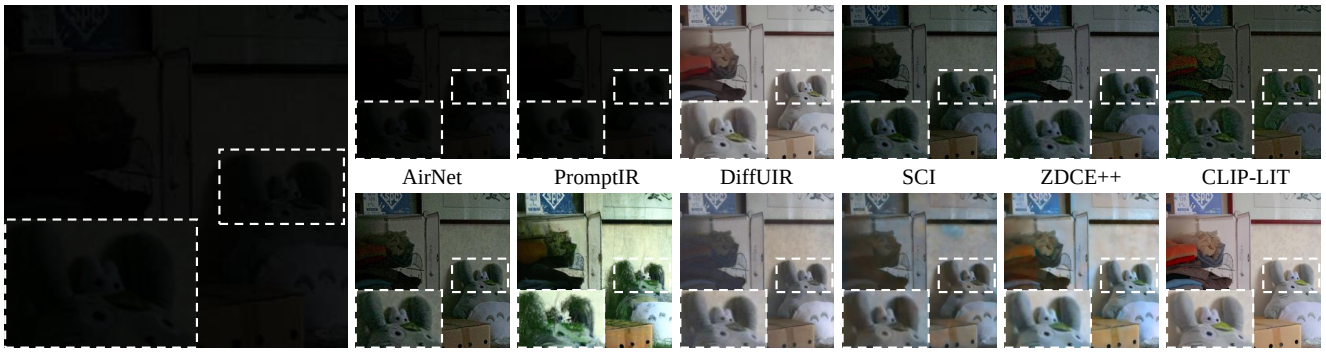
[7] Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Khan. Promptir: Prompting for all-in-one image restoration. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 2, 3

[8] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *BMVC*, 2018. 1

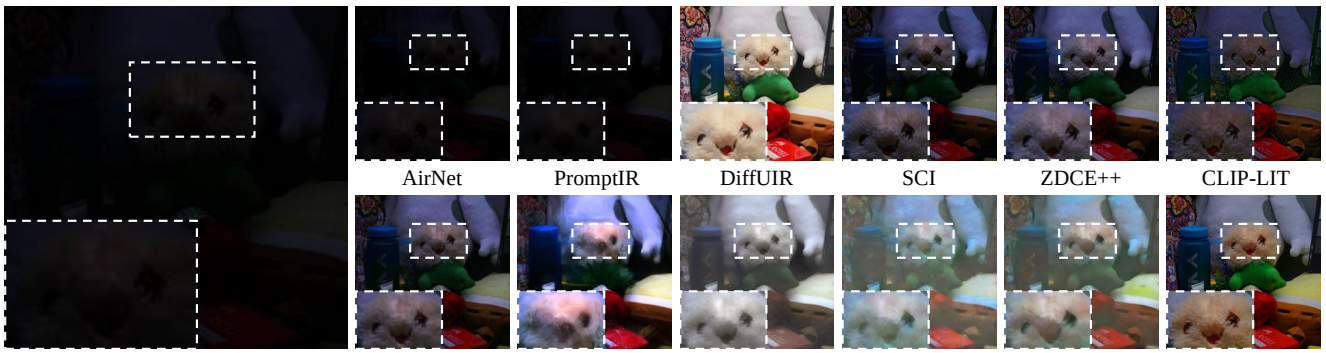
[9] Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jian Zhang, Jian-Fang Hu, and Wei-Shi Zheng. Selective hourglass mapping for universal image restoration based on diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 25445–25455, 2024. 2, 3



(a) LOLv1: 665

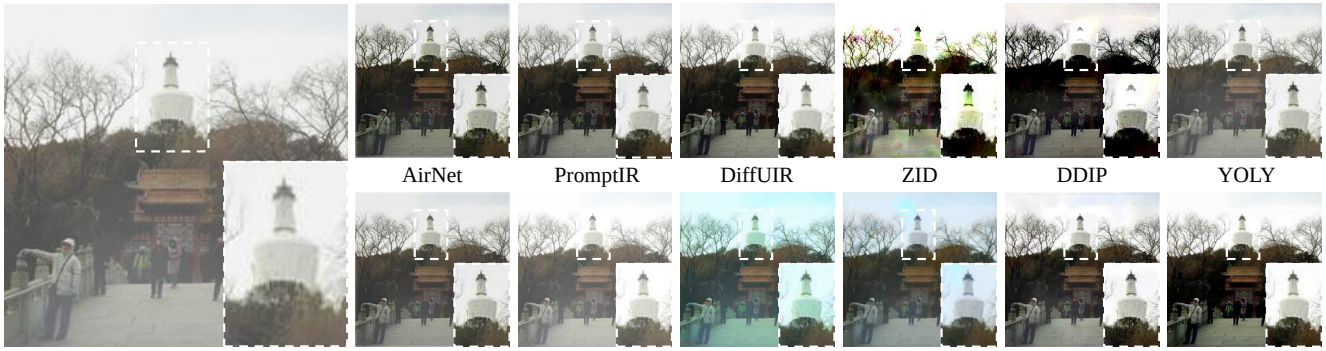


(b) LOLv1: 23

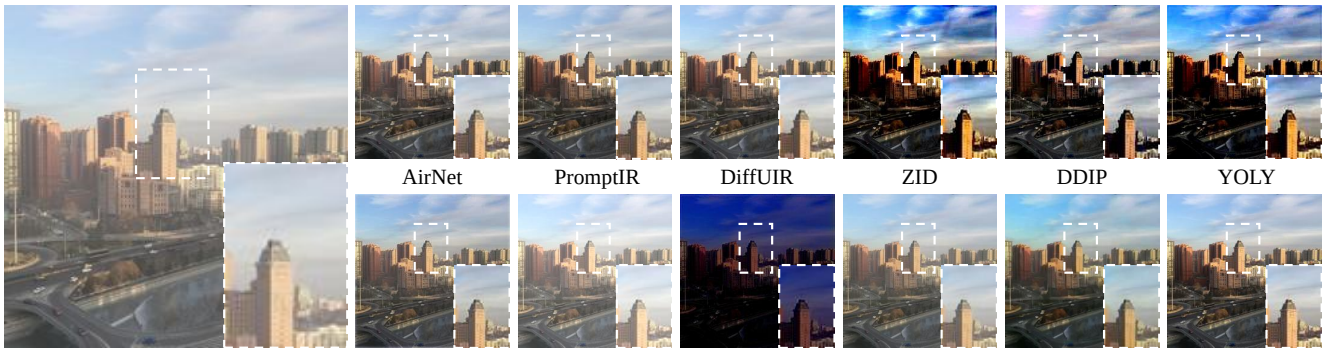


(c) LOLv1: 493

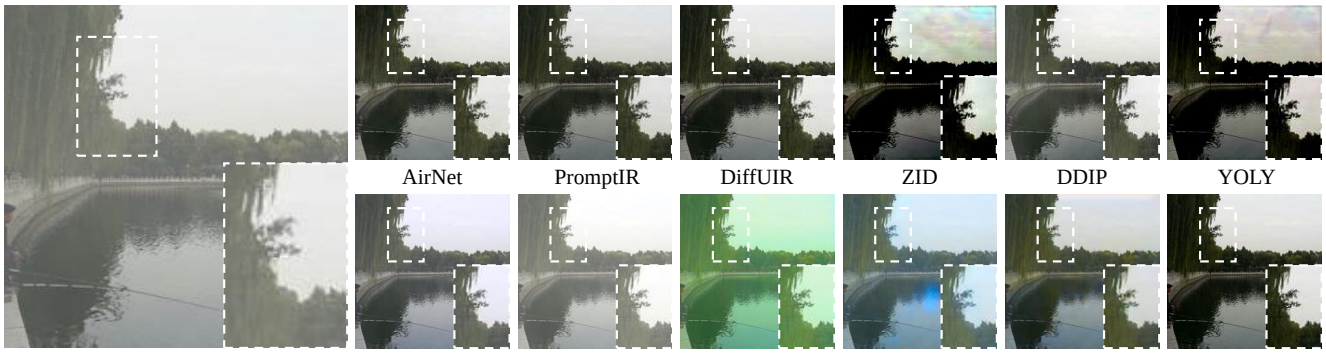
Figure 6. **Supplementary qualitative results of low-light enhancement on the LOLv1 dataset.**



(a) HSTS: 5576

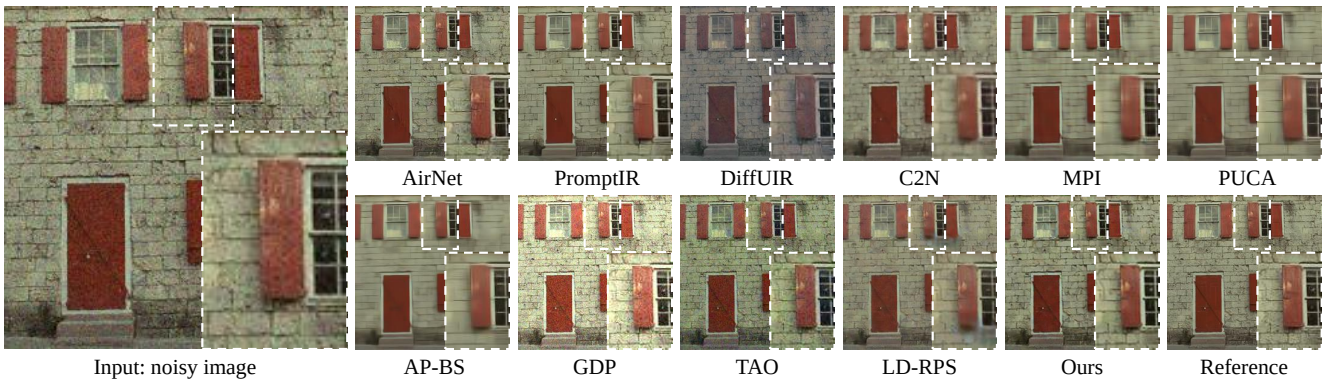


(b) HSTS: 5920

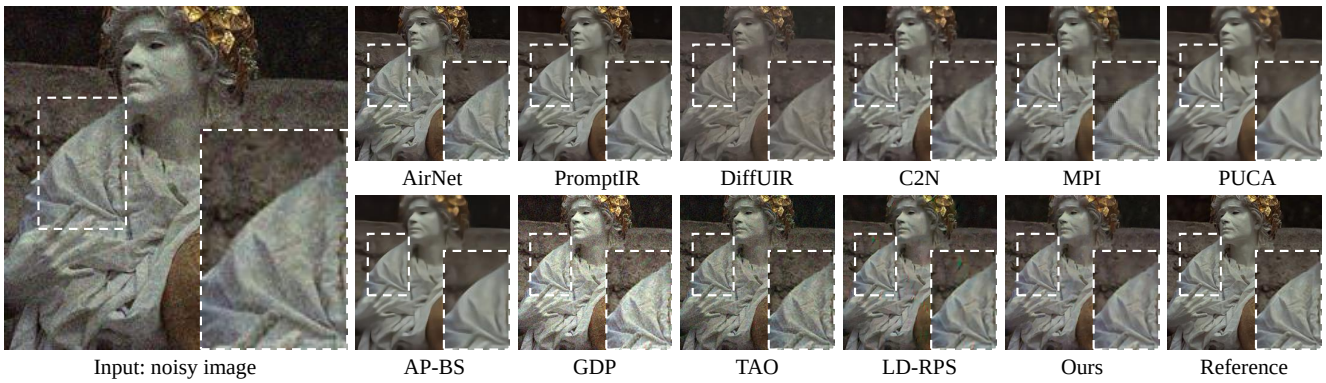


(c) HSTS: 3146

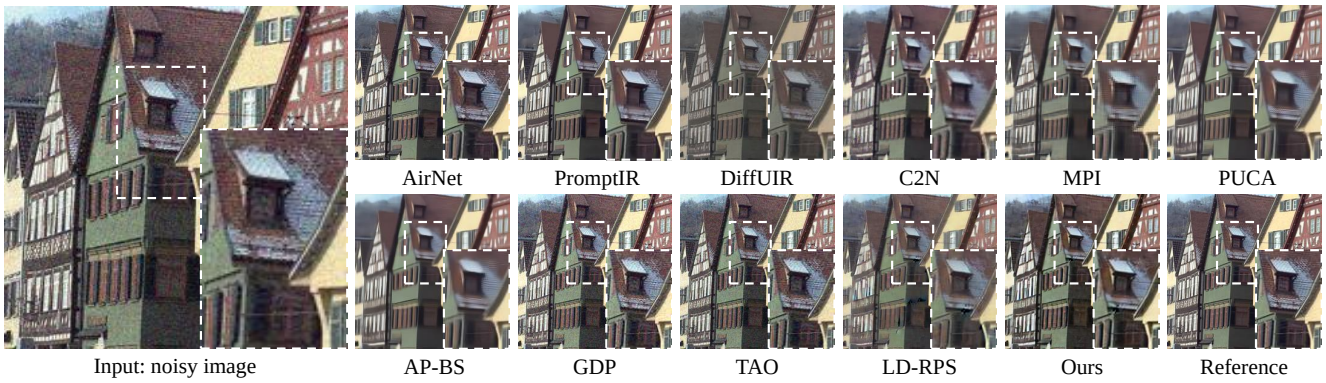
Figure 7. **Supplementary qualitative performance of image dehazing on the HSTS dataset.**



(a) Kodak24: kodim01



(b) Kodak24: kodim17

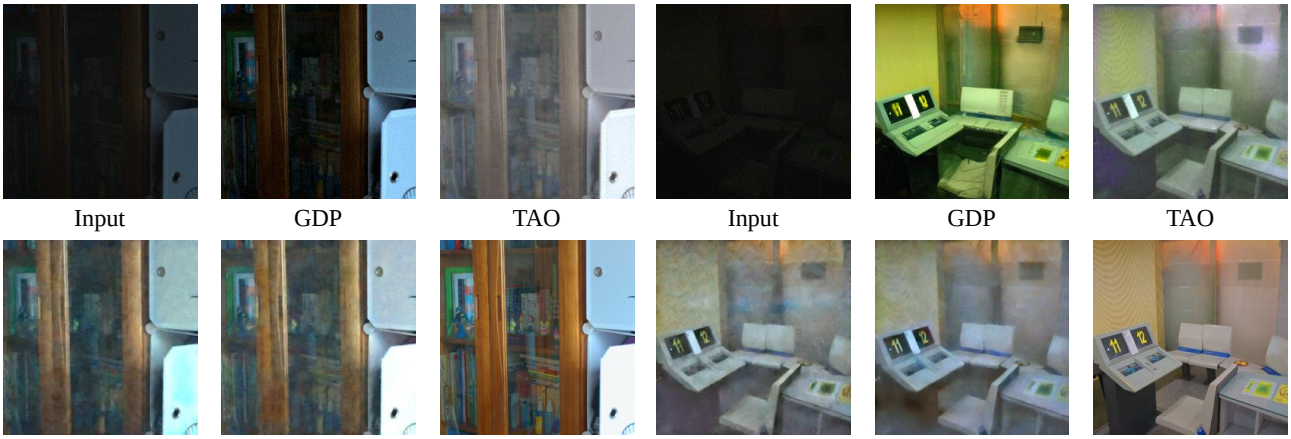


(c) Kodak24: kodim08

Figure 8. **Supplementary qualitative performance of image denoising on the Kodak24 dataset.**

(a) LOLv1: 1 (b) LOLv1: 665

Figure 9. **Visual comparisons on the Low-light + Noise scenario.**



(a) LOLv1: 1 (b) LOLv1: 665

Figure 10. **Visual comparisons on the Low-light + Haze + Noise scenario.**