

Supplementary Material for Unified Removal of Raindrops and Reflections: A New Benchmark and A Novel Pipeline

Xingyu Liu^{1,2}, Zewei He^{1,2}, Yu Chen^{1,2}, Chunyu Zhu³, Zixuan Chen⁴,
Xing Luo², and Zhe-Ming Lu^{1,2}

¹ Huanjiang Laboratory, Zhuji, China

² School of Aeronautics and Astronautics, Zhejiang University, Hangzhou, China

³ Hangzhou Institute of Technology, Xidian University, Hangzhou, China

⁴ The Chinese University of Hong Kong, Hong Kong, China
{xingyu_liu, zeweihe}@zju.edu.cn

Project page: <https://xingyuliu00.github.io/diffur3/>

A Category distribution diagram

Our RDRF dataset covers a diverse range of urban scene categories with varying semantic attributes. As shown in Fig. 1, the category distribution diagram reveals a rich diversity of semantic groups, along with a relatively balanced proportion across them, thereby providing a solid foundation for the UR³ task.

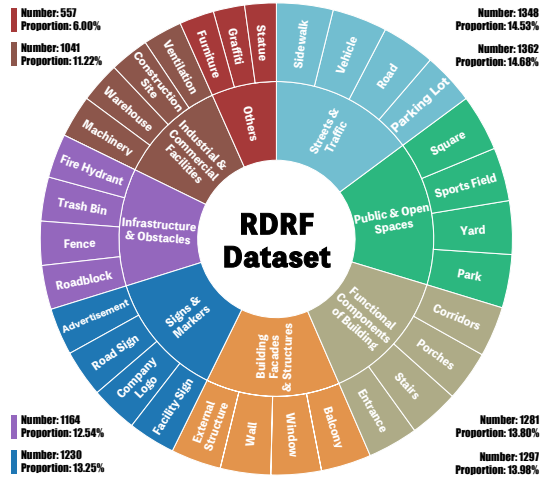


Fig. 1: The category distribution diagram illustrates the hierarchical structure of our dataset. The inner ring represents the distribution of eight primary categories and their proportions, while the outer ring showcases the corresponding representative scenarios for each category.

✉ Corresponding author (zeweihe@zju.edu.cn).

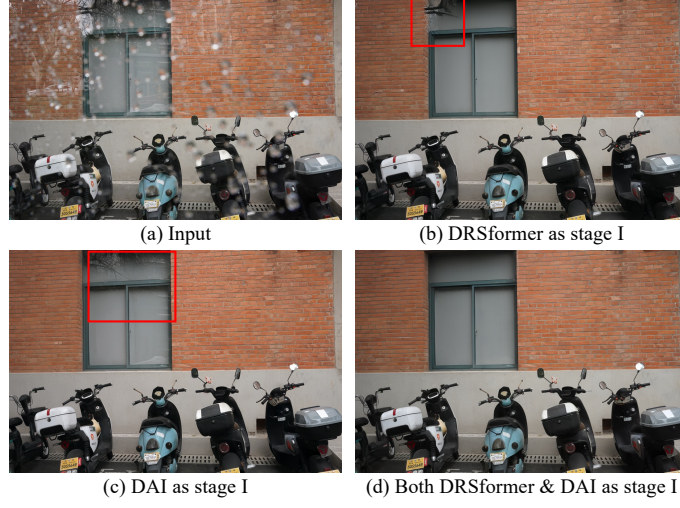


Fig. 2: Experimental results with different settings of stage I.

B Discussion on stage I

Our DiffUR³ pipeline represents a highly flexible framework. Although it needs a condition image generated by Stage I (i.e., DRSformer) during the training phase, the testing phase does not necessarily have to adhere to this specific condition, and instead, another condition image generated by other models (e.g., AGAN [5], DAI [2]) can be utilized. This indicates that the DiffUR³ has already learned the capability to extract valuable information from the condition image. Fig. 2 shows the experimental results with different stage I. By replacing DRSformer with DAI during the **inference** phase, our DiffUR³ can output similar result (Fig. 2 (c)). Note this DAI model even has not been re-trained on our RDRF dataset.

We also extend our DiffUR³ to let it incorporate multiple stage I. As shown in Fig. 3, we re-design the gate block so that it can jointly process several condition latents extracted from diverse stage I. Specifically, we first send condition latents c_{lq} , c_{s1} , c_{s2} to modulate block to perform feature modulation separately, producing modulated features $\hat{c}_{s1}$, $\hat{c}_{s2}$ and $\hat{c}_{lq}$. The obtained $\hat{c}_{s1}$, $\hat{c}_{s2}$ and $\hat{c}_{lq}$ are then concatenated together along the channel dimension and fed into a spatial-attention module, the output is normalized by a softmax operation to produce the spatial weight $\alpha_i \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8}}$ ($i = 1, 2, 3$). The formulations are as follows:

$$\begin{aligned}
 \alpha &= SpAttn([\hat{c}_{s1}, \hat{c}_{s2}, \hat{c}_{lq}]), \\
 \alpha_1, \alpha_2, \alpha_3 &= Softmax(\alpha), \\
 \bar{c}_{s1} &= \alpha_1 \cdot \hat{c}_{s1}, \\
 \bar{c}_{s2} &= \alpha_2 \cdot \hat{c}_{s2}, \\
 \bar{c}_{lq} &= \alpha_3 \cdot \hat{c}_{lq}.
 \end{aligned} \tag{1}$$

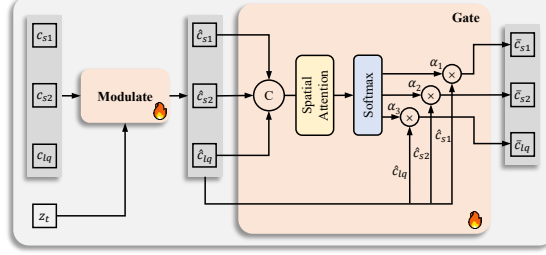


Fig. 3: Detail of re-designed Gate block for integrating multiple stage I models. The newly designed gate block aims to leverage the complementary advantages of multiple stage I models, since single stage I may not be able to provide comprehensive degradation-aware cues.

where $\bar{c}_{s1}$, $\bar{c}_{s2}$, $\bar{c}_{lq}$ are the output condition latent variables, α_1 , α_2 , and α_3 are the corresponding attention weight slices. In our case, c_{s1} and c_{s2} represent condition latents generated from stage I results of DRSformer and DAI. This model achieves better performance (Fig. 2 (d)). Multiple stage I can provide more useful information. To incorporate more additional stage I models, we can simply append their condition latents to the concatenation and increase the output channels of the spatial-attention module accordingly. We will further delve into an in-depth exploration of this research direction.

C Color correction visualization

Due to variations in display devices and viewing conditions, color shifts may not always be clearly perceived in the visual results. To provide a clearer comparison, we further visualize color error maps between the restored images and the ground truth. The color error map measures the mean deviation of hue and saturation in the HSV color space. As shown in Fig. 4, the proposed color normalization effectively reduces the color errors observed before correction, while the wavelet correction introduces additional errors.

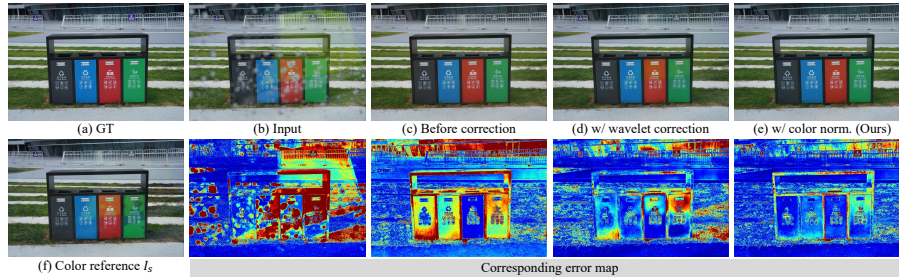


Fig. 4: Visualization of color error maps.

D Generalization to single-degradation dataset

We also apply our DiffUR³ on raindrop-only dataset (i.e., Qian’s dataset [5]). It contains 861 image pairs for training and another 58 pairs for testing. The benchmark results on the testing image pairs (i.e., Test-a) are listed in Table 1. We observe that our DiffUR³ ranks first in terms of no-reference image quality assessment metrics, which aligns closely with human visual perception (some visual results are illustrated in Fig. 5). It should be noted that Qian’s dataset has a relatively small number of samples (861 pairs) and relatively low resolutions (mostly merely 720×480), which is detrimental to the training of diffusion models and may lead to sub-optimal results. During the training, we have to reduce the patch size to accommodate Qian’s dataset.

Table 1: Benchmark results on Test-a dataset. We report PSNR, SSIM and three no-reference image quality assessment metrics (i.e., MUSIQ, CLIPQA+, HyperIQA) to perform comprehensive comparisons. The **bold** and underline indicate the best and second best.

Method	Test-a				
	PSNR $\uparrow$	SSIM $\uparrow$	MUSIQ $\uparrow$	CLIPQA+ $\uparrow$	HyperIQA $\uparrow$
AGAN	31.57	0.9023	70.52	<u>0.6691</u>	0.6650
Histoformer	33.06	0.9441	70.66	0.6530	0.6695
WeatherDiff ₆₄	30.71	0.9312	<u>71.60</u>	0.6578	<u>0.6771</u>
DiffUR ³	<u>32.33</u>	<u>0.9339</u>	72.13	0.6873	0.7064

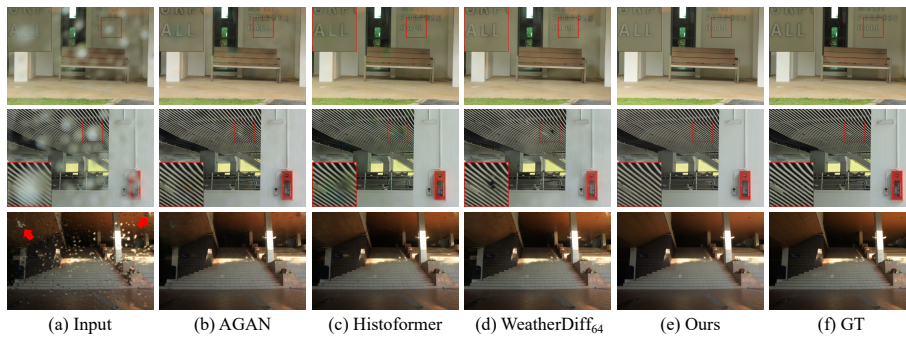


Fig. 5: Visual results of various methods on Test-a dataset. Please check and zoom in on screen for a better view.

E Downstream application

To further demonstrate the downstream applicability of our DiffUR³, we employ Google Vision API¹ to test whether our outputs can improve the object detection performance. Specifically, we conduct object detection on degraded images and our restored results. As shown in Fig. 6, the original inputs yield missing or incorrect object annotations due to the visual interference of raindrops and reflections. In contrast, our DiffUR³ effectively assist the detector in recognizing the omitted objects. This proves that our method not only improves perceptual image quality but also enhances performance in high-level vision tasks, validating its potential for real-world applications.

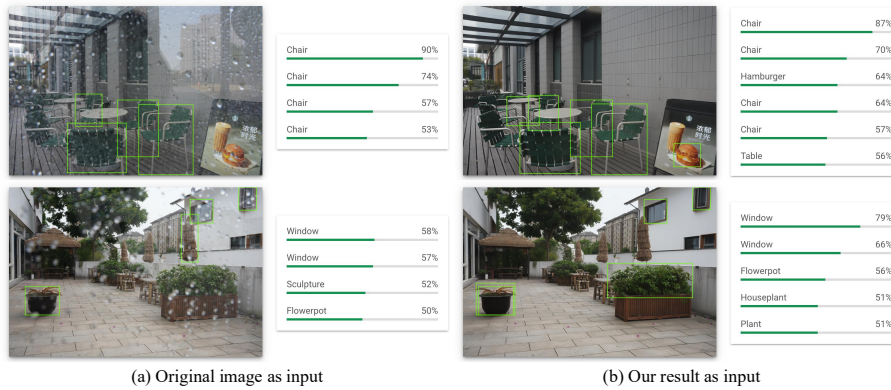


Fig. 6: We conduct object detection on original images and our restored results.

F RDRF-wild

Images collected in real-world scenarios often exhibit lower quality than those captured in controlled environments, making them more suitable for evaluating model robustness in practical settings. We capture a set of images from actual driving environments, namely **RDRF-wild**. Some samples are exhibited in Fig. 7. It comprises 41 images at present, and will continue to be expanded. We commit to making this dataset publicly available upon acceptance of our paper. We hope this data can contribute to the advancement and development of this task, and benefit the entire community. Quantitative comparisons with the competitors are shown in Table 2. Our DiffUR³ ranks first among all three metrics.

¹ <https://cloud.google.com/vision>.

**Fig. 7:** Some visual samples from our RDRF-wild.**Table 2:** Benchmark results on our RDRF-wild dataset.

Type	Method	Venue	RDRF-wild		
			MUSIQ $\uparrow$	CLIPQA $\uparrow$	HyperQA $\uparrow$
Raindrop removal	AGAN [5]	CVPR2018	66.89	0.5954	0.6328
	UMAN [6]	TIP2021	66.98	0.5933	0.6322
Reflection removal	RDNet [8]	CVPR2025	64.93	0.5762	0.6007
	DSIT [3]	NeurIPS2024	65.57	0.5718	0.5980
	DAI [2]	AAAI2026	68.05	0.6176	0.6240
Cascaded	AGAN+RDNet	-	67.92	0.6210	0.6562
	RDNet+AGAN	-	67.96	0.6132	0.6464
	AGAN+DSIT	-	68.31	0.6287	0.6468
	DSIT+AGAN	-	68.44	0.6183	0.6449
All-in-One	Histoformer [7]	ECCV2024	66.20	0.5809	0.6215
	WeatherDiff ₆₄ [4]	TPAMI2024	65.72	0.6085	0.6019
	DiffUIR [9]	CVPR2024	61.06	0.5175	0.5629
	UniRestore [1]	CVPR2025	57.31	0.4868	0.4453
Re-trained All-in-One	Histoformer [†]	-	65.02	0.5818	0.5947
	WeatherDiff ₆₄ [†]	-	64.99	0.5921	0.5634
	DiffUIR [†]	-	65.92	0.5794	0.5809
	UniRestore [†]	-	65.68	0.5695	0.5271
	Stage I	-	67.83	0.5945	0.6268
Ours	DiffUR ³	-	70.28	0.6604	0.6705

G Additional visual results

Some visual results on our RDRF-wild dataset are shown in Fig. 8. The results demonstrate that our method effectively removes raindrops and reflections, significantly enhancing image quality, which further validates the generalization capability of our DiffUR³ pipeline. We also provide additional visual results of various methods on our RDRF-testing in Fig. 9.

H Limitations

Although our method achieves promising restoration performance, its computational efficiency remains the primary limitation. Our model requires 3.67s to process a 512×512 image on an A6000 GPU. We compare the inference time of our DiffUR³ with several existing methods in Table 3. In our original implementation, a spaced DDPM sampling schedule with 50 steps is adopted. As shown

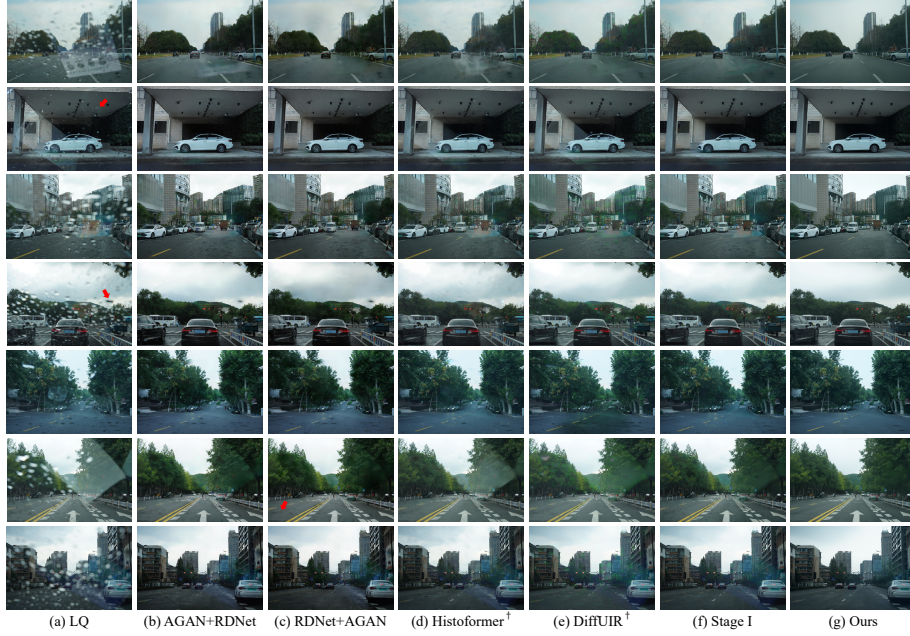


Fig. 8: More visual results of various methods on our RDRF-wild dataset. Please check and zoom in on screen for a better view.

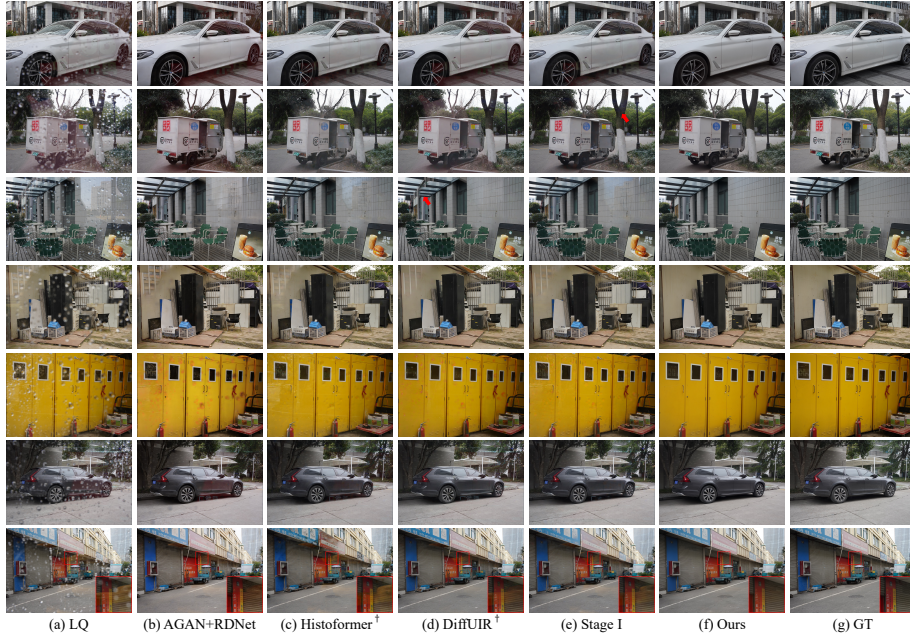


Fig. 9: More visual results of various methods on our RDRF-testing dataset. Please check and zoom in on screen for a better view.

in Fig. 10, our performance metrics are not sensitive to the number of sampling steps (when it exceeds 10). Comparable results can be achieved with merely 10 steps, which reaches a trade-off between the performance and speed. In future work, we will explore more efficient diffusion models and sampling strategies to accelerate inference speed.

Table 3: Inference time comparisons with several existing methods. Subscripts denote number of steps.

Method	AGAN+RDNet	AGAN+DSIT	Histoformer	WeatherDiff ₆₄	DiffUIR	UniRestore	Ours ₅₀	Ours ₁₀
Time (s)	0.52	0.60	0.52	58.74	0.88	0.76	3.67	0.73

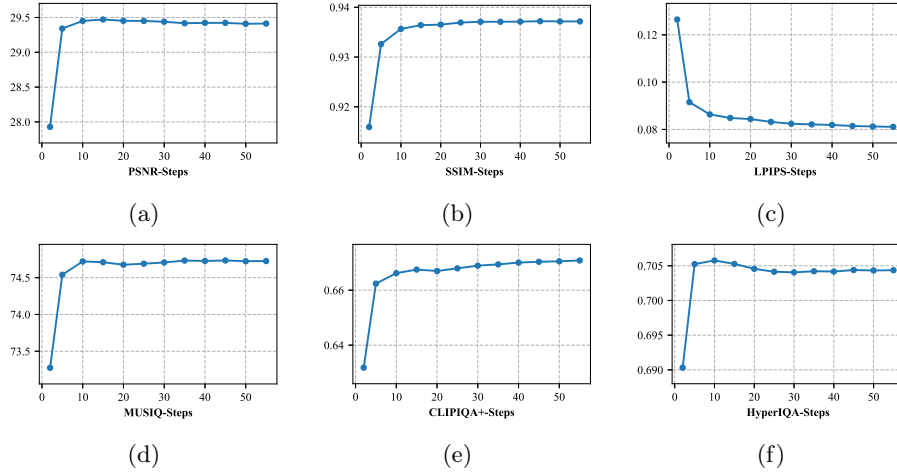


Fig. 10: Metrics *vs.* steps.

References

1. Chen, I., Chen, W.T., Liu, Y.W., Chiang, Y.C., Kuo, S.Y., Yang, M.H., et al.: Unirestore: Unified perceptual and task-oriented image restoration model using diffusion prior. In: CVPR. pp. 17969–17979 (2025)
2. Hu, J., Yang, C., Zhou, Z., Fang, J., Yang, X., Tian, Q., Shen, W.: Dereflexion Any Image with Diffusion Priors and Diversified Data. In: AAAI. pp. 4860–4868 (2026)
3. Hu, Q., Wang, H., Guo, X.: Single image reflection separation via dual-stream interactive transformers. In: NeurIPS. vol. 37, pp. 55228–55248 (2024)

4. Özdenizci, O., Legenstein, R.: Restoring Vision in Adverse Weather Conditions With Patch-Based Denoising Diffusion Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(8), 10346–10357 (2023)
5. Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J.: Attentive Generative Adversarial Network for Raindrop Removal from a Single Image. In: *CVPR*. pp. 2482–2491 (2018)
6. Shao, M.W., Li, L., Meng, D.Y., Zuo, W.M.: Uncertainty Guided Multi-Scale Attention Network for Raindrop Removal from a Single Image. *IEEE Transactions on Image Processing* **30**, 4828–4839 (2021)
7. Sun, S., Ren, W., Gao, X., Wang, R., Cao, X.: Restoring Images in Adverse Weather Conditions via Histogram Transformer. In: *ECCV*. pp. 111–129 (2024)
8. Zhao, H., Li, M., Hu, Q., Guo, X.: Reversible decoupling network for single image reflection removal. In: *CVPR*. pp. 26430–26439 (2025)
9. Zheng, D., Wu, X.M., Yang, S., Zhang, J., Hu, J.F., Zheng, W.s.: Selective hourglass mapping for universal image restoration based on diffusion model. In: *CVPR*. pp. 25445–25455 (2024)