

GenDeg: Diffusion-based Degradation Synthesis for Generalizable All-In-One Image Restoration

Supplementary Material

S1. Overview

In this supplementary, we begin by showcasing samples from the GenDS dataset along with dataset statistics, followed by comparing diffusion model conditioning with and without μ and σ . Next, we examine the influence of structure correction, S , on the quality of the generated images. We then provide more details about the dataset filtering strategy discussed in Sec. 3.2, followed by examples showcasing the effect of 3×3 convolutions in the decoder of the Swin-model for suppressing patch-border artifacts discussed in Sec. 3.3. Subsequently, we provide implementation details and a comprehensive overview of the dataset used for training and out-of-distribution (OoD) testing. We then discuss related works on diffusion models and limitations of our method. Finally, we present detailed quantitative results and additional qualitative comparisons.

To summarize, the supplementary discusses the following:

1. Samples from the GenDS dataset
2. Conditioning with and without μ, σ
3. Impact of structure correction S
4. Dataset filtering thresholds
5. Comparison of 1×1 and 3×3 convolutions in the Swin decoder
6. Implementation details
7. Training and OoD datasets
8. Related works on diffusion models
9. Limitations and scope for future work
10. Detailed quantitative results
11. Additional qualitative results

S2. Samples from the GenDS dataset

The GenDS dataset comprises a total of 783861 samples, with 224580 samples from existing datasets and 559281 samples generated using the GenDeg model (Sec. 3.1). Fig. S1 shows some generated examples from the GenDS dataset, where the left image is the ground truth and the right image is its degraded version synthesized by GenDeg. For rain, snow and low-light, the ground truth is the clean image reconstructed via the VAE encoding-decoding process, as mentioned in Section 3.1.

S3. Conditioning with and without μ, σ

As discussed in Sec. 3, not conditioning on μ and σ for degradation synthesis results in lack of diversity in the generated degradations. Fig. S2 illustrates examples of degradations generated without conditioning on μ and σ . These

degradations lack the diversity seen in those generated with μ and σ conditioning, as shown in Figure S1. Specifically, the haze appears excessively thick, the rain is either too intense or too light, and the snow is very faint.

S4. Impact of structure correction S

The structure correction module, S , discussed in Sec. 3.1 aims to reverse the structural distortions introduced during the VAE encoding-decoding process. To illustrate its effectiveness, Figure S3 shows examples of generated haze (row 1) and raindrop samples (row 2) before and after applying S . The module corrects distortions in the text caused by the VAE encoding-decoding process, thereby improving alignment between the degraded and clean images. The third row in Fig. S3 demonstrates the impact of S on snowy samples. While S successfully preserves structural details, it does not maintain the fine details of the snow, resulting in a blurred appearance of the snow. We observed similar effects for rain and low-light. Hence, we apply S only to haze, motion blur and raindrop samples, as mentioned in Sec. 3.1.

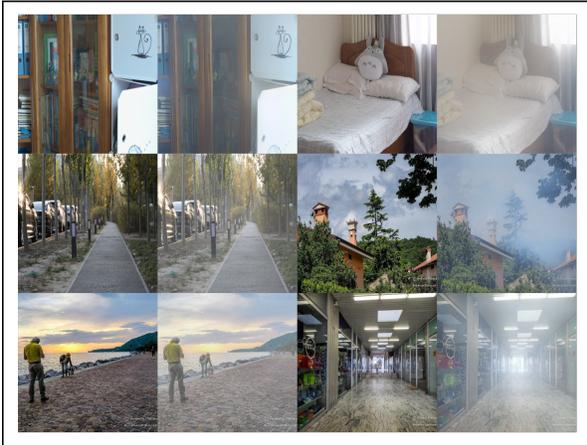
S5. Dataset filtering thresholds

After synthesizing samples using GenDeg, we filter out poor quality samples based on a mean degradation intensity based threshold. Specifically, we calculate the mean intensity, μ_{filter} , as the average absolute difference between each generated sample and its corresponding clean image. If μ_{filter} exceeds a certain threshold, T , for a given degradation, the sample is discarded. The threshold values used for each degradation are as follows: $T = 0.3$ for haze, $T = 0.23$ for rain, $T = 0.45$ for snow, $T = 0.07$ for motion blur, and $T = 0.1$ for raindrop. The thresholds were chosen by visual inspection for each degradation. Using this method, we filtered out approximately 50000 low-quality samples.

S6. Comparison of 1×1 and 3×3 convolutions in the Swin decoder

We observed that employing 3×3 convolutions in the decoder of the Swin based model significantly mitigates patch border artifacts [24]. To validate this, we trained the Swin model on the GenDS dataset using a decoder with 1×1 convolutions instead of the proposed 3×3 convolutions. Evaluation on the O-Haze [2] dataset yielded LPIPS/FID scores of 0.186/84.30, which are substantially worse than the scores obtained with 3×3 convolutions in the decoder

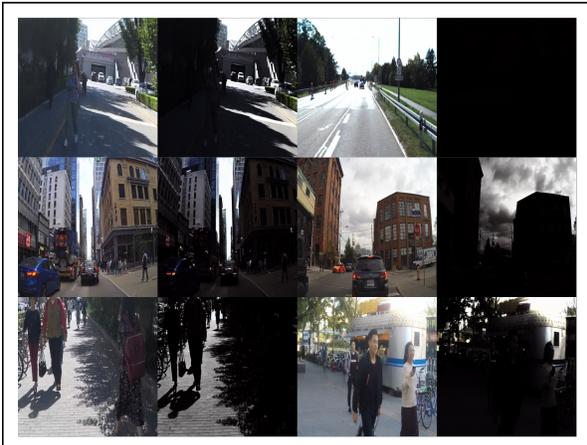
Haze



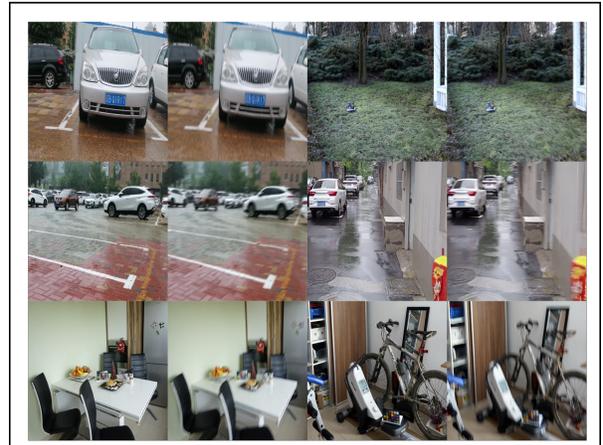
Raindrop



Low-light



Motion blur



Rain



Snow

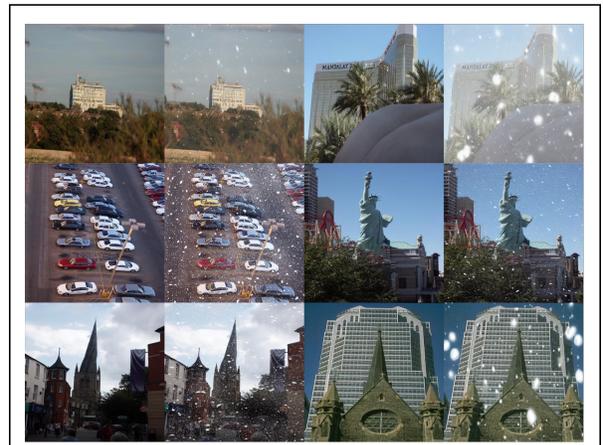


Figure S1. Samples from the GenDS dataset for each degradation type generated using the GenDeg model. The left image is the input clean image, and the right image is its synthesized degraded version.



Figure S2. Samples generated by the diffusion model trained without conditioning on μ and σ . Samples exhibit limited diversity with degradations having either very high intensity or being very faint.



Figure S3. Effect of S on haze, raindrop and snow samples. S corrects the structural distortions introduced by the VAE encoding-decoding process.

(0.165/74.6). Fig. S4 highlights the presence of patch border artifacts when 1×1 convolutions are used. The artifacts

are effectively removed with 3×3 convolutions, as shown in the zoomed-in patches. Furthermore, the decoder with 3×3 convolutions delivers superior dehazing performance, demonstrating its effectiveness.

S7. Implementation details

In this section, we provide various implementation details for training GenDeg and the restoration networks.

GenDeg. We use the InstructPix2pix [5] codebase to train our GenDeg framework, closely following their training strategies. The diffusion model is trained for a total of 60 epochs with a batch size of 512, while the structure correction module (S) is trained for 8 epochs with a batch size of 8, after training the diffusion model. For degradation generation, we use an image guidance scale of $s_I = 1.5$ and a text guidance scale of $s_T = 7.5$. The guidance scales correspond to those used in InstructPix2pix.

Restoration models. All restoration networks are trained for a total of 50 epochs using the AdamW optimizer with an initial learning rate of 2×10^{-4} and a Cosine annealing learning rate scheduler with linear warmup for 1 epoch. PromptIR and NAFNet are trained with batch sizes of 64 while the Swin model is trained with a batch size of 48. All models were optimized using $L1$ loss.

S8. Training and OoD datasets

We now describe the datasets used to train our degradation generator, GenDeg, and the image restoration models. We also detail the out-of-distribution (OoD) test sets employed for evaluating model generalization. Within-distribution testing is conducted on the test splits of the training datasets unless specified otherwise. For each type of degradation—haze, rain, snow, motion blur, low-light, and raindrops—we utilize both existing datasets and our synthesized data to train the restoration models. Dataset details are given below where R (in brackets ()) indicates real dataset while S indicates synthetic dataset.

1. Haze

Training Datasets - DenseHaze [3] (R) comprising 55 images of which we use 45 for training and 10 for testing, NH-Haze [4] (R) comprising 55 images of which we use 45 for training and 10 for testing, I-Haze [1] (R) comprising 30 images of which we use 25 for training and 5 for testing, RESIDE [20] (S) comprising 72135 images for training and 500 images from the SOTS [20] dataset for testing, and FoggyCityscapes [16] (S) comprising 8925 images for training and 4575 images for testing, totaling 81175 training samples. We further augment these with 113748 hazy images synthesized by GenDeg.

OoD Test Sets - O-Haze [2] (R) comprising 45 images and REVIDE [50] (R) comprising 284 test images.

2. Rain

Training Datasets - Real rain split of RainDS [33] (R) comprising 150 images for training and 98 images for testing, RealRain1K [23] (R) comprising 2100 images for training and 300 images for testing, ORD [22] (S) comprising 8250 images for training and 750 for testing, Rain13K [47] (S) comprising 13711 images for training and 4298 for testing, Rain1400 [14] (S) with 12600 images for training and 1400 for testing, and SPAC [9] (S) comprising 3124 images for training and 1690 images for testing, totaling 39935 training samples. The images in the ORD dataset are degraded by a mix of haze and rain. These are augmented with an additional 99753 rainy images synthesized by GenDeg.

OoD Test Sets - We evaluate on LHP-Rain [15] (R) comprising 1000 test images and the synthetic rain split of RainDS (S) with 200 test images.

3. Snow

Training Datasets - SnowCityscapes [48] (S) comprising 6000 images for both training and testing, CSD [12] (S) containing 8000 images for training 2000 for testing, and Snow100K [26] (S) with 50000 images for training and 16801 images from the Snow100k-L set for testing, totaling 64000 training samples. The images in the CSD dataset are degraded by a mix of haze and snow. These datasets are augmented with an additional 60516 snowy images synthesized by GenDeg.

OoD Test Sets - RSVD [8] (S) with 3558 test samples. The images in the RSVD dataset are degraded by a mix of haze and snow. We could not acquire the RealSnow [52] dataset due to technical difficulties.

4. Motion Blur

Training Datasets - HIDE [39] (R) with 6397 images for training and 2025 for testing, RealBlur [35] (R) comprising 3758 images for training and 980 for testing, and REDS [29] (R) with 24000 training images and 3000 test images, totaling 34155 samples for training. We augment these with an additional 79256 blurry images synthesized by GenDeg.

OoD Test Sets - GoPro [28] (R) dataset with 1111 test images.

5. Low-Light

Training Datasets - LOLv2 [42] (R, S) with 1589 training images and 200 testing images and SID [7] (R) comprising 1865 images for training and 598 for testing (Sony images only), totaling 3454 training samples. These are augmented with an additional 114053 low-light images synthesized by GenDeg.

OoD Test Sets - LOLv1 [41] (R) with 15 testing samples and SICE [6] (R) with 925 testing samples of which we only use the low-light images (images with index less than 4).

6. Raindrop

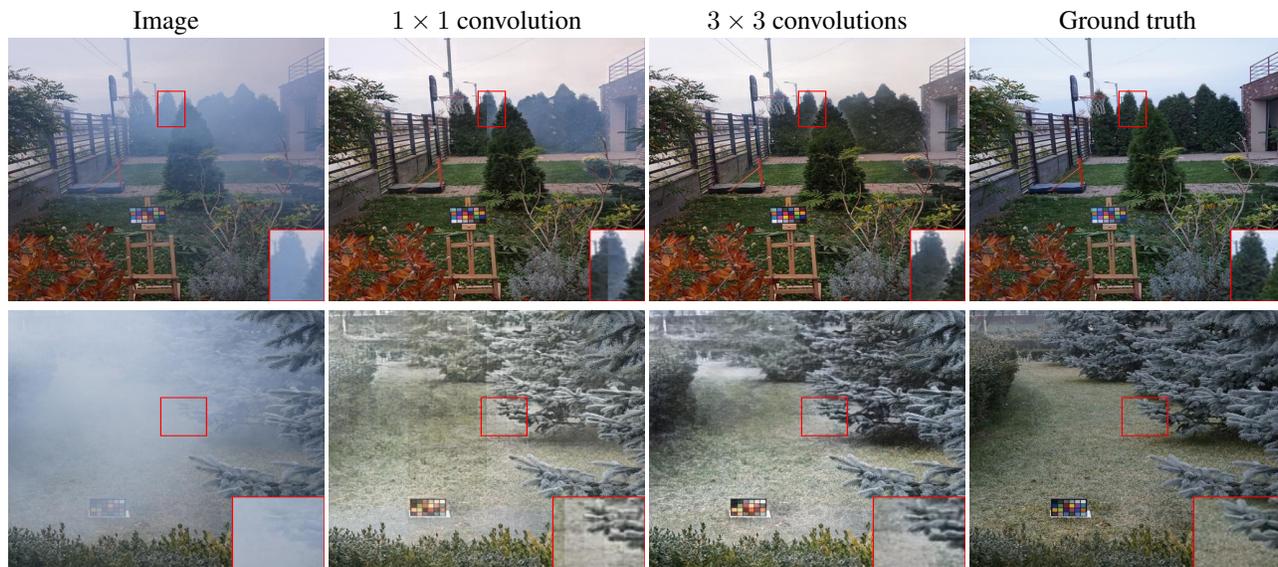


Figure S4. Impact of using 1×1 convolutions in the decoder of the Swin-based model instead of 3×3 convolutions. 3×3 convolutions in the decoder effectively mitigate the patch border artifacts, as shown by the zoomed-in patches.

Training Datasets - Raindrop dataset [32] (R) with 861 training images and 58 testing images, and the synthetic raindrop split of RainDS [33] (S) with 1000 training and 200 testing samples, totaling 1861 training samples. These are augmented with an additional 91955 raindrop images synthesized by GenDeg.

OoD Test Sets - Real raindrop split of RainDS [33] (R) with 98 samples for testing.

S9. Related works on diffusion models

Diffusion models have gained significant attention as generative models, with them being the current state-of-the-art data generators [17, 40]. In the diffusion process, two processes are learnt. The forward process iteratively adds noise to the data while the denoising (backward) process learns to model the reverse process and gradually denoises the data. Based on this principle, Denoising Diffusion Probabilistic Models (DDPMs) [17] were proposed that could generate high fidelity images from noise. Various recent advances in diffusion models have extended DDPMs to complex tasks like text-to-image synthesis [36], conditional synthesis [5] and audio generation [11], among others.

S10. Limitations and scope for future work

Despite our generated data significantly enhancing the out-of-distribution (OoD) performance of restoration models, we acknowledge some limitations with our data and pipeline. Our approach relies on the Stable Diffusion 1.5 [36] while more recent versions, such as Stable Diffusion 2.0, 3.0 and SDXL [30], have since emerged. An in-

teresting research direction would be to explore the domain gap effects and performance improvements offered by these newer versions. Additionally, our structure correction module, S , does not perform optimally for snow, rain and low-light (as discussed in Secs. S4 and 3.1). Future research could focus on developing improved correction modules to handle these specific degradations more effectively, to mitigate the structural distortions in the generated degraded images. Furthermore, scaling the synthetic data further and analyzing its impact could be beneficial, as our dataset is still relatively small compared to those used to train foundation models for low-level vision tasks, such as SAM [19] and Depth Anything [43]. Overall, leveraging diffusion models to generate large-scale synthetically degraded images to improve the generalization of image restoration models is a promising research direction.

S11. Detailed quantitative results

In the main paper, we provided out-of-distribution (OoD) performance of methods in terms of LPIPS and FID metrics in Table 2. Subsequently, we provided the mean within-distribution performance of methods for each degradation with LPIPS and FID metrics (Table 3). This section presents more detailed quantitative evaluations using PSNR, SSIM, LPIPS and FID metrics for within-distribution comparisons, and PSNR and SSIM metrics for OoD comparisons.

The within-distribution comparisons are provided in detail for each test set described in Sec. S8. Tables S1, S2, S3, S4, S5, S6 summarize the within-

distribution performances for haze, rain, snow, motion blur, low-light and raindrop degradations, respectively. Comparisons with state-of-the-art (SOTA) methods are included if the method is trained for that degradation. From the tables, substantial improvements can be observed for most cases when training with the GenDS dataset for haze, low-light, and raindrop degradations, while competitive performance is maintained for other degradations. Note that some within-distribution datasets might be OoD for SOTA approaches as they were not retrained on all within-distribution datasets. The specific within-distribution datasets for each SOTA approach are given in Table S8.

Table S7 reports PSNR and SSIM metrics for the OoD test sets (LPIPS and FID were provided in Table 2 of the main paper). Observe that training with the GenDS dataset yields substantial improvements in most cases, particularly in the SSIM scores. However, PSNR is more sensitive to color shifts, which can lower its values despite overall enhancements, as explained in Sec. 4.1. Furthermore, the SSIM improvement for NAFNet and NAFNet GenDS for the real raindrop split of RainDS dataset is marginal. However, the LPIPS and FID scores in Table 2 of the main paper reveal more pronounced improvements. The improvement in LPIPS and FID scores aligns well with the qualitative results for this dataset as shown in Fig. 5 of the main paper and Fig. S6. As LPIPS and FID metrics better correlate with perceptual quality, they are more reliable indicators than PSNR and SSIM [25, 36, 38, 49] for assessing OoD performance.

We also provide quantitative results for both within-distribution and OoD test sets in the form of radar plots in Fig. S5 for ease of viewing.

S12. Additional qualitative results

In this section, we provide additional qualitative results. Fig. S6 presents qualitative comparisons of three top-performing image restoration models- PromptIR [31], Swin model (Sec. 3.3) and NAFNet [10], evaluated on OoD test sets when trained with and without the GenDS dataset. It can be observed that training with the GenDS dataset results in improved performance. Figure S7 illustrates the within-distribution performance of the same models, showing that training with the GenDS dataset yields significant improvements for haze, low-light, and raindrop degradations, while performance remains nearly identical for other degradation types. This indicates that the GenDS dataset does not degrade within-distribution performance.

Finally, Figs. S8, S9 and S10 provide qualitative comparisons of top-performing approaches PromptIR, the Swin model, and NAFNet trained with the GenDS dataset against several SOTA approaches, namely, DiffUIR [51], Diff-Plugin [25], InstructIR [13], and AutoDIR [18]. Diff-Plugin used for qualitative comparisons is the publicly available

pre-trained model. Fig. S8 shows comparisons for haze, low-light, rain and motion blur, Fig. S9 for snow and Fig. S10 for raindrop removal. These comparisons are split across multiple figures because not all SOTA methods are trained for every degradation task. The models trained on GenDS dataset deliver consistently good OoD performance across all degradations whereas each SOTA approach tends to perform well only for specific degradation types. It is important to note that the primary aim of our approach is not to compete directly with SOTA models, but rather demonstrate the performance differences observed for OoD testing when models are trained with and without the proposed GenDS dataset. Furthermore, training SOTA models on our GenDS dataset may also boost their OoD performance.

References

- [1] Cosmin Ancuti, Codruta O Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: A dehazing benchmark with real hazy and haze-free indoor images. In *Advanced Concepts for Intelligent Vision Systems: 19th International Conference, ACIVS 2018, Poitiers, France, September 24–27, 2018, Proceedings 19*, pages 620–631. Springer, 2018. 4, 13
- [2] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 754–762, 2018. 6, 1, 4, 10, 17
- [3] Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *2019 IEEE international conference on image processing (ICIP)*, pages 1014–1018. IEEE, 2019. 4, 13
- [4] Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 444–445, 2020. 8, 4, 9, 13
- [5] Tim Brooks, Aleksander Holynski, and Alexei A Efros. Instructpix2pix: Learning to follow image editing instructions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18392–18402, 2023. 4, 5
- [6] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018. 6, 4, 10, 17
- [7] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3291–3300, 2018. 4, 16
- [8] Haoyu Chen, Jingjing Ren, Jinjin Gu, Hongtao Wu, Xuequan Lu, Haoming Cai, and Lei Zhu. Snow removal in video: A new dataset and a novel method. In *2023 IEEE/CVF In-*

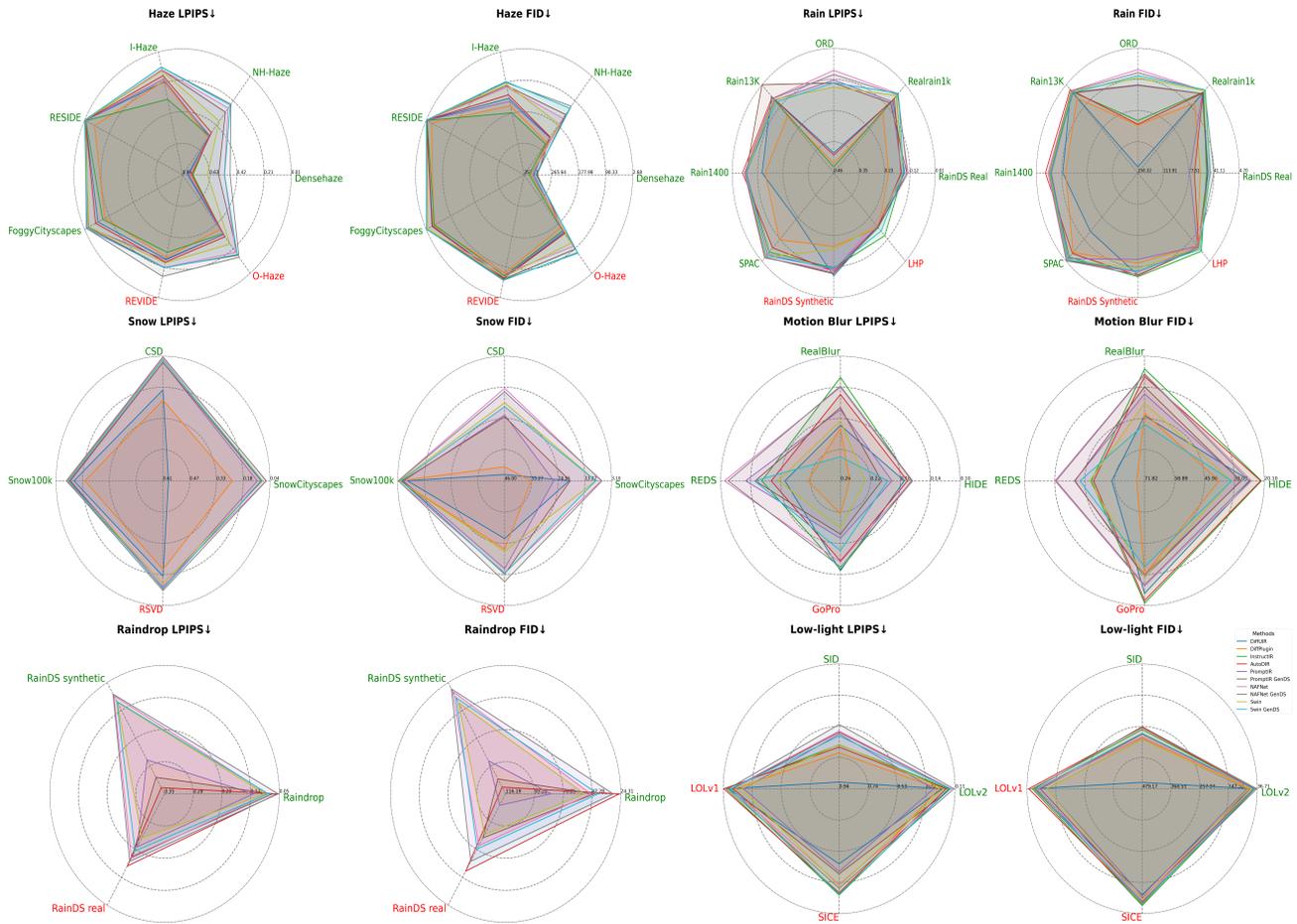


Figure S5. Comparison of NAFNet [10], PromptIR [31], and Swin-transformer models using LPIPS and FID metrics, trained with and without our GenDS dataset. Performance is evaluated on within-distribution (green) and OoD (red) test sets. The figure also includes the performance of existing state-of-the-art (SOTA) All-In-One Restoration (AIOR) approaches, namely, DiffUIR [51], Diff-Plugin [25], InstructIR [13], and AutoDIR [18]. Training with the GenDS dataset significantly enhances OoD performance. Legend is located in the lower-right corner. Metric values decrease outward.

ternational Conference on Computer Vision (ICCV), pages 13165–13176. IEEE, 2023. 6, 4, 8, 11, 17

[9] Jie Chen, Cheen-Hau Tan, Junhui Hou, Lap-Pui Chau, and He Li. Robust video content alignment and compensation for rain removal in a cnn framework. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6286–6295, 2018. 4, 14

[10] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022. 3, 5, 6, 7, 13, 14, 15, 16, 17

[11] Nanxin Chen, Yu Zhang, Heiga Zen, Ron J Weiss, Mohammad Norouzi, and William Chan. Wavegrad: Estimating gradients for waveform generation. In *International Conference on Learning Representations*, 2021. 5

[12] Wei-Ting Chen, Hao-Yu Fang, Cheng-Lin Hsieh, Cheng-Che Tsai, I-Hsiang Chen, Jian-Jiun Ding, and Sy-Yen Kuo. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and

contradict channel loss. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4176–4185, 2021. 4, 15

[13] Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *European Conference on Computer Vision*, pages 1–21. Springer, 2025. 1, 3, 6, 7, 10, 13, 14, 15, 16, 17

[14] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3855–3863, 2017. 4, 14, 17

[15] Yun Guo, Xueyao Xiao, Yi Chang, Shumin Deng, and Luxin Yan. From sky to the ground: A large-scale benchmark and simple baseline towards real rain removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12097–12107, 2023. 2, 6, 4, 8, 10, 17

[16] Martin Hahner, Dengxin Dai, Christos Sakaridis, Jan-Nico Zaech, and Luc Van Gool. Semantic understanding of

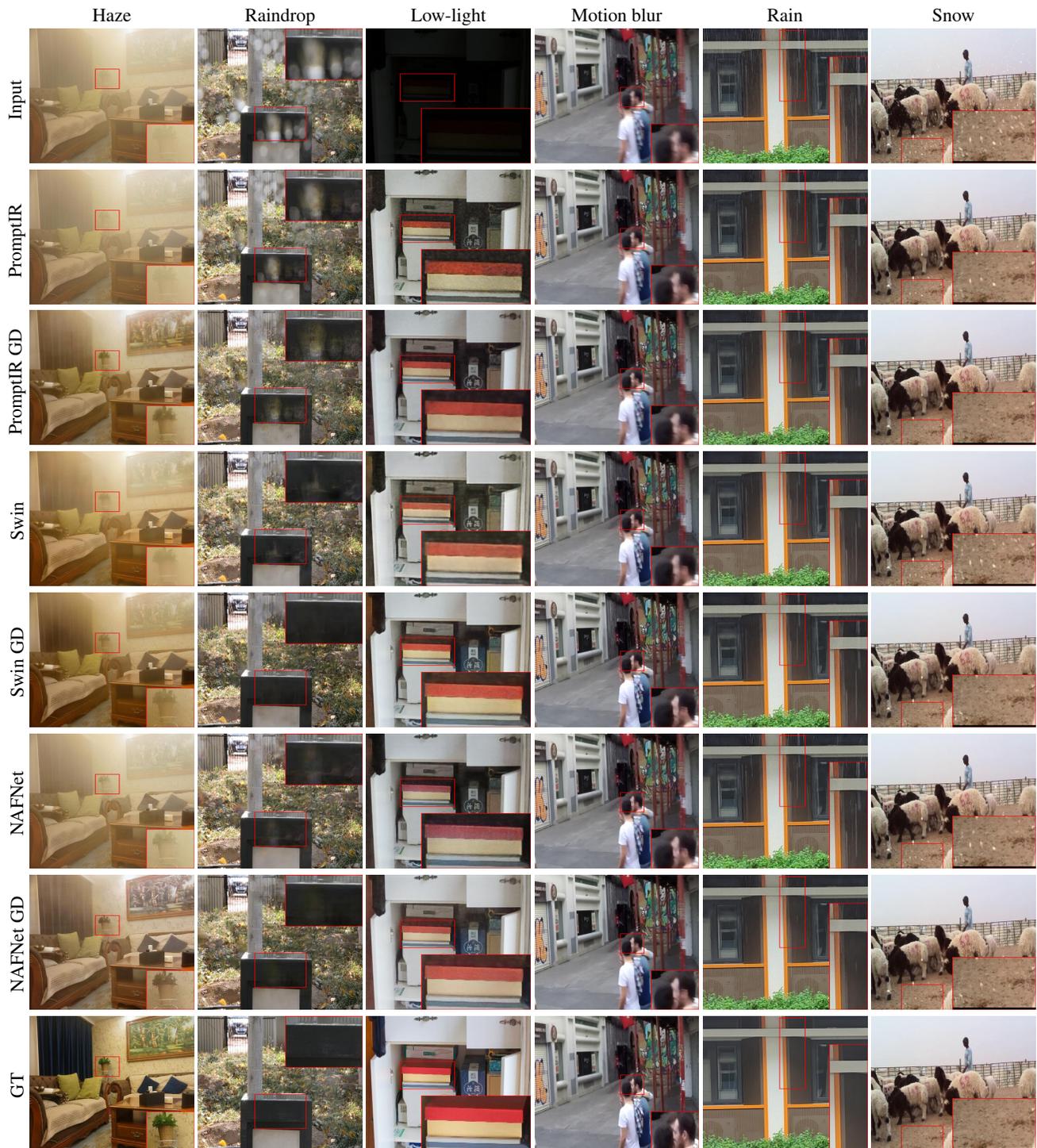


Figure S6. Additional qualitative comparisons of image restoration models (PromptIR, NAFNet and the Swin model) trained with and without our GenDS dataset. The suffix GD represents training with the GenDS dataset. Comparisons are on OoD test sets (Haze: REV-IDE [50], Raindrop: RainDS [33], Low-light: LOLv1 [41], Motion blur: GoPro [28], Rain: LHP [15] and Snow: RSVD [8]). Training with the GenDS dataset improves OoD performance. Zoomed-in patches are provided for viewing fine details.

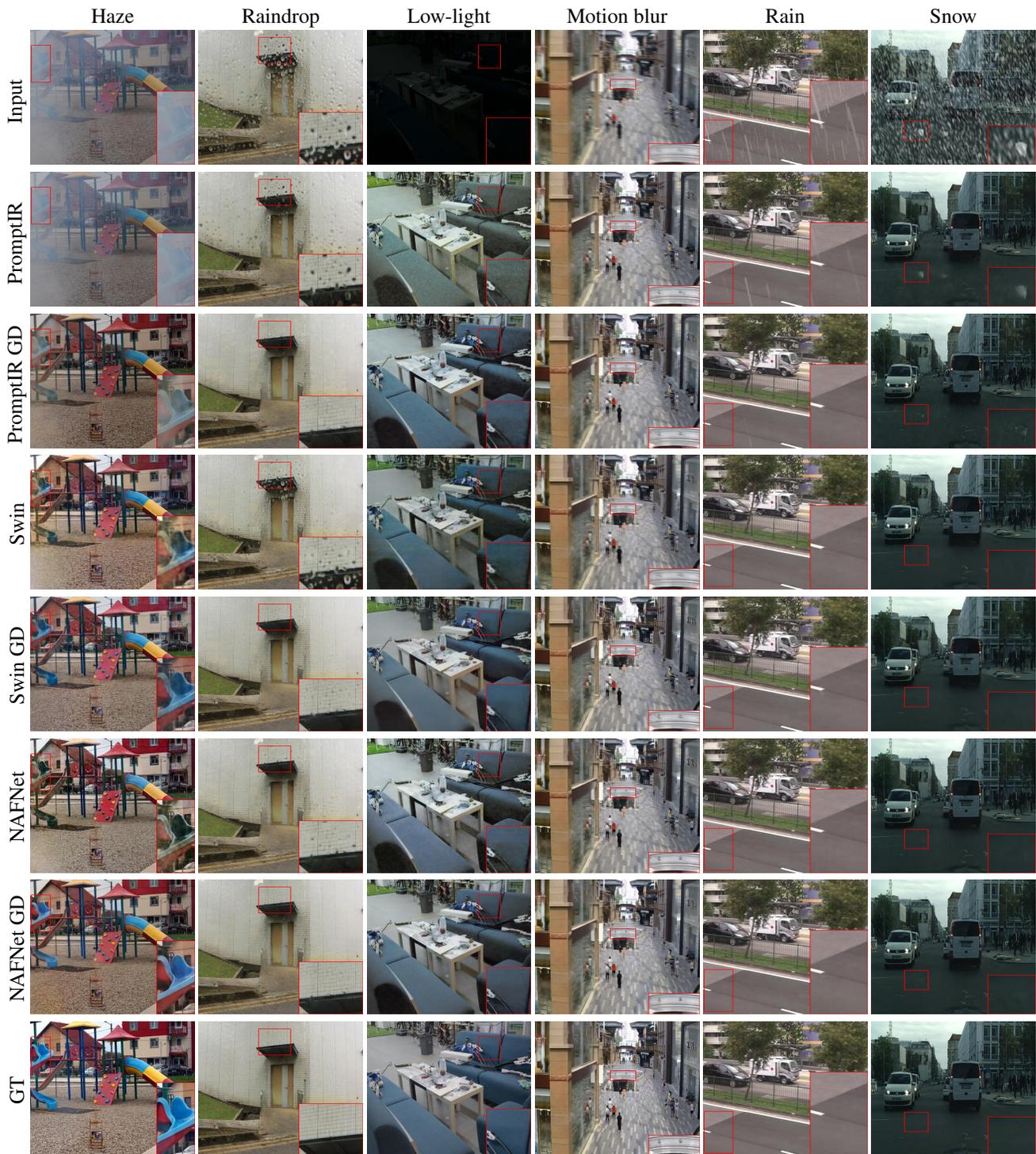


Figure S7. Qualitative comparisons of image restoration models on within-distribution test sets when trained with and without our GenDS dataset. The suffix GD represents training with the GenDS dataset. Images are from the following test sets - Haze: NH-Haze [4], Raindrop: Raindrop [32], Low-light: LOLv2 [42], Motion blur: HIDE [39] and Snow: SnowyCityscapes [48]. Zoomed-in patches are provided for viewing fine details.

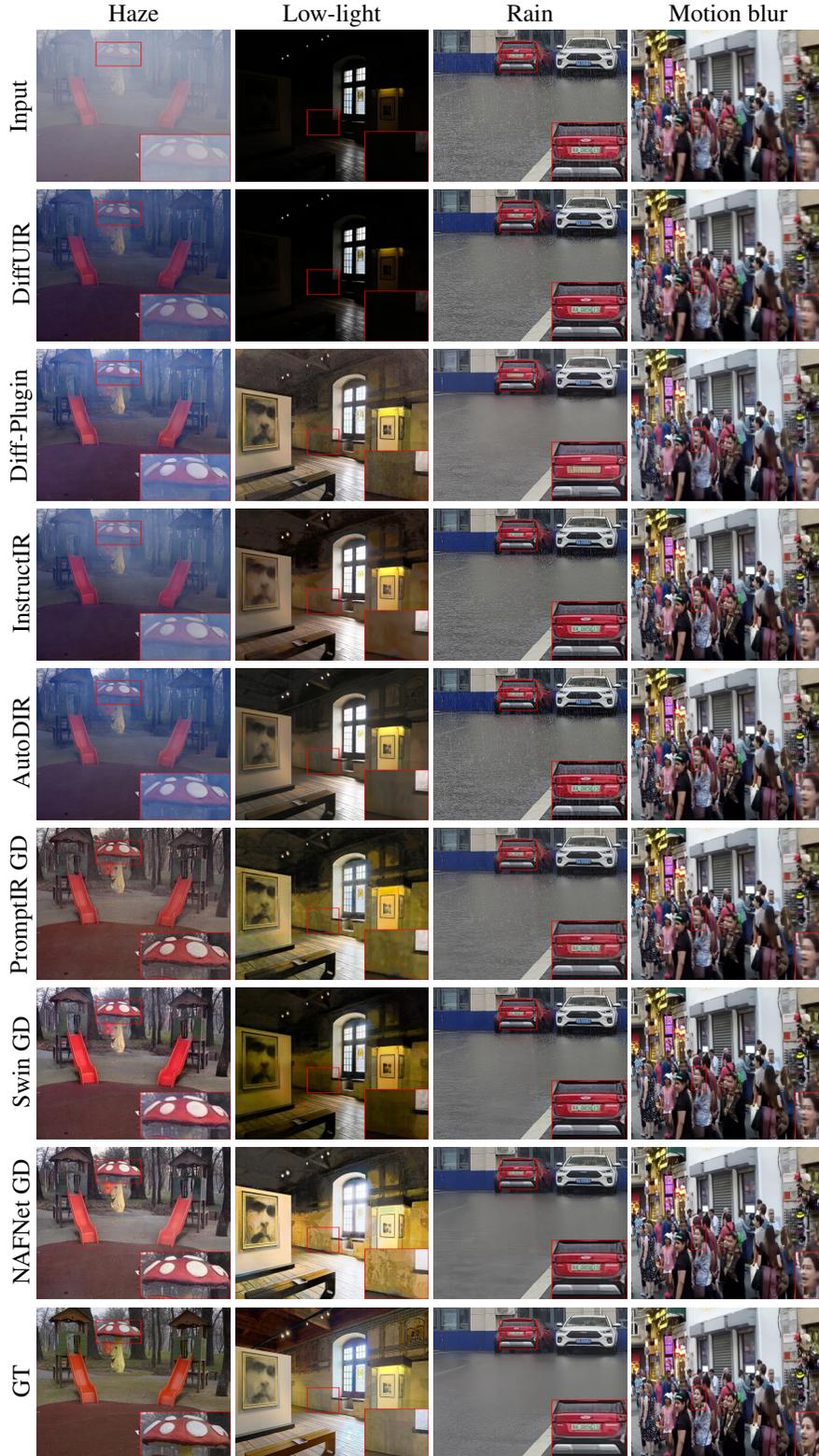


Figure S8. Qualitative comparisons for OoD performance of image restoration models (PromptIR, NAFNet and the Swin model) trained with our GenDS dataset, and SOTA AIOR approaches, namely, DiffUIR [51], Diff-Plugin [25], InstructIR [13] and AutoDIR [18]. Images are from the following datasets- Haze: O-Haze [2], Low-light: SICE [6], Rain: LHP [15] and Motion Blur: Go Pro [28].



Figure S9. Qualitative comparisons of OoD performance of image restoration models (PromptIR, NAFNet and the Swin model) trained with our GenDS dataset, and SOTA AIOR approaches, namely, DiffUIR [51], Diff-Plugin [25] for the task of desnowing on the RSVD [8] dataset.

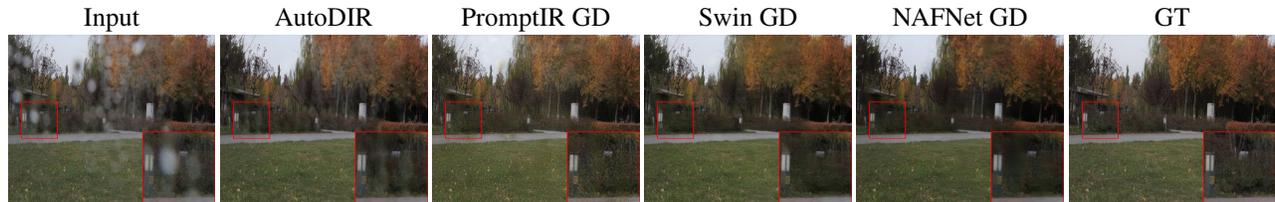


Figure S10. Qualitative comparisons for OoD performance of image restoration models (PromptIR, NAFNet and the Swin model) trained with our GenDS dataset, and AutoDIR, a SOTA AIOR approach, for the task of raindrop removal on the RainDS [33] dataset.

- 3675–3681, 2019. 4, 13
- [17] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2020. Curran Associates Inc. 5
- [18] Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. *arXiv preprint arXiv:2310.10123*, 2023. 1, 3, 5, 6, 7, 10, 13, 14, 15, 16, 17
- [19] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023. 2, 5
- [20] 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, 2019. 1, 2, 4, 13, 17
- [21] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*, pages 19730–19742. PMLR, 2023. 4
- [22] Ruoteng Li, Loong Fah Cheong, and Robby T. Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1633–1642, 2019. 4, 14
- [23] Wei Li, Qiming Zhang, Jing Zhang, Zhen Huang, Xinmei Tian, and Dacheng Tao. Toward real-world single image deraining: A new benchmark and beyond. *arXiv preprint arXiv:2206.05514*, 2022. 2, 4, 14
- [24] Yihao Liu, Jingwen He, Jinjin Gu, Xiangtao Kong, Yu Qiao, and Chao Dong. Degae: A new pretraining paradigm for low-level vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23292–23303, 2023. 5, 1
- [25] Yuhao Liu, Zhanghan Ke, Fang Liu, Nanxuan Zhao, and Rynson WH Lau. Diff-plugin: Revitalizing details for diffusion-based low-level tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4197–4208, 2024. 1, 3, 6, 7, 10, 11, 13, 14, 15, 16, 17
- [26] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27(6): 3064–3073, 2018. 1, 4, 15, 17
- [27] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for universal image restoration. *arXiv preprint arXiv:2310.01018*, 3(8), 2023. 1, 2, 3, 6, 7, 8, 13, 14, 15, 16, 17
- [28] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3883–3891, 2017. 1, 2, 6, 4, 8, 10, 17
- [29] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019. 4, 15
- [30] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023. 5

- [31] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36, 2024. 1, 3, 5, 6, 7, 13, 14, 15, 16, 17
- [32] Rui Qian, Robby T. Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for rain-drop removal from a single image. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 5, 9, 16, 17
- [33] Ruijie Quan, Xin Yu, Yuanzhi Liang, and Yi Yang. Removing raindrops and rain streaks in one go. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9147–9156, 2021. 6, 4, 5, 8, 11, 14, 16, 17
- [34] Stephan R Richter, Hassan Abu AlHajja, and Vladlen Koltun. Enhancing photorealism enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2): 1700–1715, 2022. 3
- [35] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 184–201. Springer, 2020. 4, 15
- [36] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2, 4, 5, 6
- [37] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22500–22510, 2023. 3
- [38] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):4713–4726, 2022. 6
- [39] Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. Human-aware motion deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5572–5581, 2019. 1, 2, 4, 9, 15
- [40] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, page 2256–2265. JMLR.org, 2015. 5
- [41] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 6, 4, 8, 17
- [42] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 4, 9, 16
- [43] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10371–10381, 2024. 2, 5
- [44] Rajeep Yasarla and Vishal M Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8405–8414, 2019. 1
- [45] Rajeep Yasarla and Vishal M Patel. Confidence measure guided single image de-raining. *IEEE Transactions on Image Processing*, 29:4544–4555, 2020. 1
- [46] Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Fredo Durand, William T Freeman, and Taesung Park. One-step diffusion with distribution matching distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6613–6623, 2024. 3
- [47] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, 2021. 1, 4, 14, 17
- [48] Kaihao Zhang, Rongqing Li, Yanjiang Yu, Wenhan Luo, and Changsheng Li. Deep dense multi-scale network for snow removal using semantic and depth priors. *IEEE Transactions on Image Processing*, 30:7419–7431, 2021. 4, 9, 15
- [49] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 6
- [50] Xinyi Zhang, Hang Dong, Jinshan Pan, Chao Zhu, Ying Tai, Chengjie Wang, Jilin Li, Feiyue Huang, and Fei Wang. Learning to restore hazy video: A new real-world dataset and a new method. In *CVPR*, pages 9239–9248, 2021. 6, 4, 8, 17
- [51] 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*, 2024. 1, 3, 6, 7, 10, 11, 13, 14, 15, 16, 17
- [52] Y. Zhu, T. Wang, X. Fu, X. Yang, X. Guo, J. Dai, Y. Qiao, and X. Hu. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21747–21758, 2023. 4

Table S1. Quantitative comparisons of various models using PSNR (\uparrow), SSIM (\uparrow), LPIPS (\downarrow) and FID (\downarrow) metrics on within-distribution haze datasets. The format of metrics is PSNR/SSIM/LPIPS/FID. PromptIR [31], NAFNet [10], the Swin model, DA-CLIP [27] and Diff-Plugin [25] are trained with and without the GenDS dataset. The table also includes the performance of existing state-of-the-art (SOTA) All-In-One Restoration (AIOR) approaches, namely, DiffUIR [51], Diff-Plugin [25], InstructIR [13] and AutoDIR [18]. (R) indicates real image dataset and (S) indicates synthetic image dataset. Diff-Plugin[#] is the publicly available pre-trained model.

Method	Densehaze [3] (R)	NH-Haze [4] (R)	I-Haze [1] (R)	RESIDE [20] (S)	FoggyCityscapes [16] (S)
DiffUIR	10.20/0.359/0.799/322.58	12.59/0.498/0.507/220.10	14.02/0.732/0.208/135.65	32.94/0.956/0.013/3.71	18.56/0.834/0.120/25.28
Diff-Plugin [#]	10.54/0.363/0.763/326.64	12.19/0.333/0.523/237.72	18.36/0.742/0.203/156.08	23.23/0.765/0.091/21.40	16.92/0.728/0.178/28.77
InstructIR	10.52/0.436/0.759/336.47	12.52/0.502/0.511/243.45	16.05/0.777/0.328/176.45	26.90/0.952/0.017/5.62	16.42/0.791/0.161/32.08
AutoDIR	11.69/0.381/0.770/310.78	13.00/0.519/0.483/218.52	16.86/0.814/0.168/125.15	30.98/0.974/0.013/3.81	19.04/0.846/0.100/23.12
PromptIR	11.11/0.376/0.789/328.71	12.62/0.494/0.513/221.45	17.36/0.806/0.191/143.74	29.13/0.967/0.016/5.28	28.10/0.961/0.039/6.08
PromptIR GenDS	16.25/0.490/0.561/312.94	16.07/0.668/0.311/138.88	20.07/0.857/0.129/99.44	30.87/0.977/0.014/5.14	29.08/0.964/0.038/6.34
Swin	16.24/0.458/0.647/332.50	16.49/0.635/0.386/172.37	21.87/0.874/0.135/95.23	36.16/0.987/0.008/2.96	28.62/0.956/0.044/6.57
Swin GenDS	17.90/0.530/0.516/315.05	20.31/0.742/0.251/117.14	22.05/0.883/0.110/87.39	38.08/0.989/0.007/2.81	33.45/0.972/0.027/3.82
NAFNet	16.99/0.517/0.518/327.77	18.11/0.714/0.276/152.80	21.33/0.867/0.127/104.00	32.31/0.981/0.011/3.26	33.40/0.979/0.021/3.30
NAFNet GenDS	17.67/0.545/0.474/316.06	19.10/0.743/0.242/109.51	20.57/0.869/0.109/89.96	34.80/0.986/0.008/2.88	32.55/0.977/0.024/3.76
DA-CLIP	10.57/0.348/0.8299/327.52	12.74/0.455/0.544/224.42	16.99/0.737/0.2160/146.69	24.91/0.943/0.025/5.80	19.01/0.863/0.1090/31.32
DA-CLIP GenDS	13.25/0.404/0.7019/232.87	15.17/0.568/0.3914/160.4	18.97/0.777/0.1878/135.08	29.73/0.972/0.012/4.00	25.00/0.942/0.045/13.21
Diff-Plugin	9.13/0.341/0.796/341.31	13.10/0.331/0.4586/206.70	16.83/0.707/0.2239/129.80	21.56/0.750/0.0997/21.86	19.57/0.776/0.1225/18.67
Diff-Plugin GenDS	9.33/0.359/0.7199/331.67	14.24/0.357/0.3942/156.41	16.74/0.718/0.1972/121.15	23.35/0.771/0.0908/20.91	22.12/0.798/0.1094/9.94

Table S2. Quantitative comparisons of various models using PSNR (\uparrow), SSIM (\uparrow), LPIPS (\downarrow) and FID (\downarrow) metrics on within-distribution rain datasets. The format of metrics is PSNR/SSIM/LPIPS/FID. PromptIR [31], NAFNet [10], the Swin model, DA-CLIP [27] and Diff-Plugin [25] are trained with and without the GenDS dataset. The table also includes the performance of existing state-of-the-art (SOTA) AIOR approaches, namely, DiffUIR [51], Diff-Plugin [25], InstructIR [13] and AutoDIR [18]. (R) indicates real image dataset and (S) indicates synthetic image dataset. Diff-Plugin[#] is the publicly available pre-trained model.

Method	RainDS [33] (R)	Realrain1k [23] (R)	ORD [22] (S)	Rain13K [47] (S)	Rain1400 [14] (S)	SPAC [9] (S)
DiffUIR	25.70/0.802/0.145/49.36	27.02/0.905/0.088/18.23	16.82/0.644/0.385/143.1	31.03/0.904/0.069/14.10	27.76/0.899/0.141/42.54	23.70/0.803/0.223/53.81
Diff-Plugin [#]	21.39/0.526/0.215/60.65	24.14/0.660/0.150/35.00	13.39/0.470/0.423/94.62	21.71/0.617/0.169/26.86	22.27/0.623/0.155/46.61	24.46/0.733/0.118/18.86
InstructIR	25.41/0.789/0.151/51.93	30.71/0.939/0.067/16.70	13.37/0.565/0.436/89.21	29.56/0.885/0.088/16.06	30.97/0.907/0.071/25.09	31.97/0.951/0.055/11.23
AutoDIR	24.87/0.773/0.161/69.21	30.04/0.924/0.079/18.21	14.29/0.594/0.394/93.56	30.52/0.895/0.073/12.65	31.18/0.913/0.055/17.64	31.69/0.941/0.077/17.49
PromptIR	24.28/0.759/0.157/75.87	32.85/0.922/0.095/19.47	25.53/0.863/0.124/47.54	29.85/0.884/0.089/17.45	30.69/0.901/0.073/26.23	33.07/0.958/0.045/9.14
PromptIR GenDS	25.36/0.788/0.135/50.35	32.02/0.917/0.103/19.54	26.45/0.866/0.135/48.05	30.17/0.888/0.010/18.41	31.11/0.908/0.072/24.82	34.32/0.964/0.030/6.54
Swin	25.63/0.793/0.151/47.10	33.26/0.936/0.065/14.47	27.39/0.859/0.150/39.30	30.89/0.898/0.085/15.84	31.16/0.909/0.070/23.71	33.82/0.959/0.031/6.62
Swin GenDS	25.97/0.794/0.149/46.46	34.15/0.951/0.054/13.25	28.31/0.867/0.134/36.88	30.62/0.894/0.098/17.96	30.98/0.907/0.071/25.07	33.37/0.957/0.037/7.29
NAFNet	25.73/0.798/0.134/45.92	34.98/0.951/0.055/14.71	29.18/0.896/0.089/29.65	31.22/0.904/0.078/14.48	31.73/0.917/0.063/21.36	34.59/0.966/0.026/4.94
NAFNet GenDS	25.94/0.800/0.130/41.22	34.28/0.947/0.059/14.46	29.02/0.893/0.104/33.43	31.23/0.906/0.079/15.22	31.56/0.916/0.066/122.80	34.78/0.967/0.029/5.66
DA-CLIP	22.46/0.660/0.1672/55.98	24.71/0.871/0.1087/21.90	20.80/0.778/0.1494/43.72	25.71/0.822/0.1062/32.00	26.28/0.841/0.0884/28.10	28.24/0.926/0.065/18.23
DA-CLIP GenDS	23.32/0.652/0.1941/61.30	28.28/0.883/0.1017/20.28	23.75/0.821/0.1124/38.60	28.38/0.852/0.0887/26.74	28.92/0.866/0.0764/25.24	32.57/0.944/0.0389/10.89
Diff-Plugin	21.14/0.522/0.2078/58.28	23.88/0.667/0.1421/33.94	18.05/0.570/0.2070/56.78	21.75/0.620/0.1645/26.32	22.23/0.624/0.1542/44.51	24.39/0.739/0.1127/18.47
Diff-Plugin GenDS	21.70/0.530/0.1989/54.25	24.33/0.676/0.1395/32.81	17.18/0.555/0.2268/58.42	21.89/0.621/0.1647/26.54	22.40/0.626/0.1544/45.32	24.72/0.741/0.1047/18.01

Table S3. Quantitative comparisons of various models using PSNR (\uparrow), SSIM (\uparrow), LPIPS (\downarrow) and FID (\downarrow) metrics on within-distribution snow datasets. The format of metrics is PSNR/SSIM/LPIPS/FID. PromptIR [31], NAFNet [10], the Swin model, DA-CLIP [27] and Diff-Plugin [25] are trained with and without the GenDS dataset. The table also includes the performance of existing state-of-the-art (SOTA) AIOR approaches, namely, DiffUIR [51] and Diff-Plugin [25]. (R) indicates real image dataset and (S) indicates synthetic image dataset. Diff-Plugin[#] is the publicly available pre-trained model.

Method	SnowCityscapes [48] (S)	CSD [12] (S)	Snow100k [26] (S)
DiffUIR	12.42/0.330/0.583/20.16	17.57/0.790/0.196/43.81	28.76/0.869/0.138/7.05
Diff-Plugin [#]	22.25/0.694/0.242/35.35	15.61/0.635/0.243/41.14	21.02/0.611/0.196/4.10
PromptIR	30.16/0.918/0.109/26.31	27.62/0.919/0.071/23.34	27.86/0.863/0.120/5.39
PromptIR	31.44/0.934/0.089/19.73	28.03/0.920/0.070/23.92	28.51/0.874/0.115/4.92
GenDS			
NAFNet	33.67/0.953/0.063/7.23	31.92/0.947/0.044/14.13	30.01/0.893/0.096/3.25
NAFNet	33.02/0.950/0.061/6.96	31.39/0.942/0.048/15.36	29.78/0.891/0.100/3.38
GenDS			
Swin	32.38/0.934/0.081/9.41	32.01/0.935/0.058/19.13	29.34/0.880/0.104/3.87
Swin GenDS	32.17/0.929/0.085/9.46	31.95/0.933/0.059/20.52	29.23/0.881/0.107/4.34
DA-CLIP	25.42/0.849/0.1574/53.67	22.72/0.872/0.092/32.25	24.31/0.803/0.1350/44.92
DA-CLIP	31.51/0.937/0.057/22.88	27.15/0.916/0.057/21.92	27.16/0.853/0.1037/34.65
GenDS			
Diff-Plugin	23.05/0.718/0.1932/15.63	20.21/0.699/0.1559/30.66	21.38/0.617/0.1855/7.95
Diff-Plugin	22.19/0.716/0.1823/13.45	19.08/0.691/0.1606/31.69	20.43/0.613/0.1866/7.96
GenDS			

Table S4. Quantitative comparisons of various models using PSNR (\uparrow), SSIM (\uparrow), LPIPS (\downarrow) and FID (\downarrow) metrics on within-distribution motion blur datasets. The format of metrics is PSNR/SSIM/LPIPS/FID. PromptIR [31], NAFNet [10], the Swin model, DA-CLIP [27] and Diff-Plugin [25] are trained with and without the GenDS dataset. The table also includes the performance of existing state-of-the-art (SOTA) AIOR approaches, namely, DiffUIR [51], Diff-Plugin [25], InstructIR [13] and AutoDIR [18]. (R) indicates real image dataset and (S) indicates synthetic image dataset. Diff-Plugin[#] is the publicly available pre-trained model.

Method	HIDE [39] (R)	RealBlur [35] (R)	REDS [29] (R)
DiffUIR	27.17/0.854/0.174/26.23	26.15/0.721/0.188/44.80	26.84/0.825/0.185/57.64
Diff-Plugin [#]	21.40/0.658/0.247/39.68	23.62/0.728/0.192/43.65	21.47/0.628/0.216/68.40
InstructIR	27.50/0.859/0.165/21.37	27.09/0.845/0.127/25.39	26.93/0.831/0.154/48.82
AutoDIR	27.03/0.862/0.170/21.67	24.77/0.770/0.149/28.58	27.23/0.846/0.167/49.77
PromptIR	26.55/0.837/0.190/30.07	28.13/0.845/0.168/35.78	29.45/0.867/0.133/41.52
PromptIR	26.70/0.845/0.204/30.29	28.44/0.857/0.165/32.81	29.58/0.871/0.146/41.70
GenDS			
NAFNet	27.40/0.864/0.164/25.66	28.93/0.876/0.140/27.41	30.94/0.898/0.105/33.10
NAFNet	27.53/0.866/0.163/26.68	29.03/0.879/0.138/27.68	30.82/0.896/0.110/33.59
GenDS			
Swin	25.91/0.829/0.226/34.35	28.15/0.851/0.182/39.60	29.23/0.866/0.176/48.14
Swin GenDS	25.89/0.830/0.196/34.24	28.44/0.854/0.228/48.39	29.01/0.866/0.145/43.88
DA-CLIP	22.03/0.743/0.2139/49.65	24.50/0.719/0.1928/45.40	24.37/0.779/0.1884/55.06
DA-CLIP	23.95/0.778/0.1620/39.37	27.52/0.825/0.1236/31.36	27.04/0.805/0.1342/42.74
GenDS			
Diff-Plugin	22.18/0.694/0.1673/27.61	25.40/0.764/0.1302/28.89	22.80/0.647/0.1445/56.56
Diff-Plugin	22.68/0.706/0.1599/25.74	25.81/0.772/0.1248/26.61	22.78/0.649/0.1414/55.62
GenDS			

Table S5. Quantitative comparisons of various models using PSNR (\uparrow), SSIM (\uparrow), LPIPS (\downarrow) and FID (\downarrow) metrics on within-distribution low-light datasets. The format of metrics is PSNR/SSIM/LPIPS/FID. PromptIR [31], NAFNet [10], the Swin model, DA-CLIP [27] and Diff-Plugin [25] are trained with and without the GenDS dataset. The table also includes the performance of existing state-of-the-art (SOTA) AIOR approaches, namely, DiffUIR [51], Diff-Plugin [25], InstructIR [13] and AutoDIR [18]. (R) indicates real image dataset and (S) indicates synthetic image dataset. Diff-Plugin[#] is the publicly available pre-trained model.

Method	LOLv2 [42] (R, S)	SID [7] (R)
DiffUIR	20.27/0.826/0.204/64.17	9.86/0.061/0.899/456.35
Diff-Plugin [#]	18.00/0.644/0.226/62.83	12.47/0.365/0.703/298.52
InstructIR	23.99/0.857/0.156/46.28	12.84/0.377/0.649/269.21
AutoDIR	19.94/0.800/0.176/48.02	13.45/0.457/0.665/262.27
PromptIR	17.25/0.710/0.264/76.65	17.67/0.517/0.579/303.10
PromptIR	21.90/0.868/0.142/51.75	18.31/0.550/0.566/285.43
GenDS		
NAFNet	22.54/0.866/0.140/49.89	18.13/0.557/0.558/294.83
NAFNet	22.87/0.887/0.120/38.55	18.64/0.586/0.514/258.75
GenDS		
Swin	21.97/0.844/0.195/70.92	17.12/0.517/0.647/304.38
Swin GenDS	22.32/0.869/0.146/49.53	16.18/0.483/0.591/282.73
DA-CLIP	12.29/0.601/0.2635/76.16	11.21/0.131/0.814/340.79
DA-CLIP	16.31/0.692/0.235/63.16	10.84/0.146/0.633/301.16
GenDS		
Diff-Plugin	16.80/0.594/0.2906/85.27	13.02/0.288/0.6424/250.00
Diff-Plugin	17.96/0.630/0.246/77.87	12.69/0.291/0.666/234.50
GenDS		

Table S6. Quantitative comparisons of various models using PSNR (\uparrow), SSIM (\uparrow), LPIPS (\downarrow) and FID (\downarrow) metrics on within-distribution raindrop datasets. The format of metrics is PSNR/SSIM/LPIPS/FID. PromptIR [31], NAFNet [10] the Swin model, DA-CLIP [27] and Diff-Plugin [25] are trained with and without the GenDS dataset. The table also includes the performance of an existing state-of-the-art (SOTA) AIOR approach, namely, AutoDIR [18]. (R) indicates real image dataset and (S) indicates synthetic image dataset.

Method	Raindrop [32] (R)	RainDS [33] (S)
AutoDIR	30.10/0.924/0.058/25.53	20.22/0.795/0.333/110.66
PromptIR	27.04/0.885/0.120/80.07	21.76/0.852/0.259/88.90
PromptIR	29.15/0.908/0.071/45.40	20.91/0.829/0.305/103.91
GenDS		
NAFNet	29.61/0.914/0.083/48.30	28.23/0.924/0.088/31.52
NAFNet	30.33/0.922/0.056/30.42	28.26/0.927/0.084/29.43
GenDS		
Swin	28.74/0.903/0.089/53.30	26.98/0.893/0.105/40.88
Swin GenDS	28.89/0.901/0.081/42.54	26.87/0.893/0.104/36.72
DA-CLIP	25.35/0.861/0.086/43.79	23.98/0.888/0.1175/37.86
DA-CLIP	28.44/0.891/0.0658/32.36	25.23/0.907/0.1212/39.70
GenDS		
Diff-Plugin	21.95/0.675/0.1716/71.91	21.21/0.635/0.1990/49.38
Diff-Plugin	25.03/0.733/0.1117/46.76	22.48/0.683/0.1550/38.29
GenDS		

Table S7. Quantitative comparisons of NAFNet [10], PromptIR [31], the Swin-transformer, DA-CLIP [27] and Diff-Plugin [25] models using PSNR and SSIM metrics (higher is better), trained with and without our GenDS dataset. Performance is evaluated on OoD test sets. The table also includes the performance of existing state-of-the-art (SOTA) AIOR approaches, namely, DiffUIR [51], Diff-Plugin [25], InstructIR [13] and AutoDIR [18]. (R) indicates real images and (S) indicates synthetic images. '-' indicates that the method was not trained for that degradation. Diff-Plugin[#] is the publicly available pre-trained model.

Method	REVIDE [50]	O-Haze [2]	RainDS [33]	LHP [15]	RSVD [8]	GoPro [28]	LOLv1 [41]	SICE [6]	RainDS [33]
Degradation Type	Haze (R)	Haze (R)	Rain (S)	Rain (R)	Snow (S)	Motion Blur (R)	Low-light (R)	Low-light (R)	Raindrop (R)
DiffUIR	17.26/0.792	16.59/0.705	30.85/0.897	26.71/0.832	21.60/0.823	29.17/0.864	21.65/0.836	10.00/0.367	-
Diff-Plugin [#]	17.45/0.728	15.79/0.471	22.04/0.635	26.02/0.735	18.92/0.662	21.76/0.633	19.38/0.713	17.59/0.611	-
InstructIR	16.51/0.831	16.56/0.709	30.24/0.879	28.93/0.871	-	28.26/0.870	22.81/0.836	17.58/0.750	-
AutoDIR	16.31/0.782	17.57/0.731	29.14/0.858	28.44/0.841	-	27.07/0.828	20.53/0.850	15.37/0.685	23.33/0.754
PromptIR	17.56/0.786	16.46/0.701	29.70/0.871	25.85/0.835	20.08/0.838	26.98/0.828	20.53/0.768	12.42/0.438	20.94/0.709
PromptIR GenDS	19.20/0.827	22.40/0.849	30.00/0.884	26.15/0.840	21.77/0.858	27.28/0.843	21.38/0.812	11.41/0.466	22.04/0.718
Swin	18.44/0.812	20.12/0.802	28.50/0.863	28.97/0.863	22.05/0.838	26.79/0.843	20.07/0.784	15.27/0.667	22.34/0.726
Swin GenDS	18.69/0.831	20.60/0.843	29.34/0.868	29.30/0.869	23.07/0.863	26.09/0.830	24.27/0.842	16.05/0.678	22.87/0.740
NAFNet	18.71/0.819	19.85/0.814	28.49/0.887	27.41/0.835	21.32/0.848	28.14/0.868	22.37/0.826	14.03/0.592	22.91/0.745
NAFNet GenDS	20.96/0.871	20.51/0.849	30.40/0.891	27.53/0.848	22.81/0.873	27.80/0.864	22.82/0.838	14.65/0.614	23.31/0.747
DA-CLIP	16.98/0.758	16.04/0.609	24.74/0.764	25.64/0.838	20.72/0.788	22.87/0.763	10.94/0.474	12.04/0.499	20.99/0.622
DA-CLIP GenDS	17.78/0.810	19.94/0.700	25.50/0.758	28.74/0.851	21.42/0.802	25.55/0.795	11.21/0.481	12.02/0.506	21.39/0.621
Diff-Plugin	17.84/0.748	16.36/0.464	22.00/0.637	25.10/0.723	19.12/0.663	22.23/0.654	17.77/0.651	18.60/0.588	18.47/0.453
Diff-plugin GenDS	17.14/0.765	16.54/0.460	21.89/0.636	26.05/0.739	19.39/0.674	22.84/0.674	18.29/0.676	19.47/0.623	19.54/0.489

Table S8. Within-distribution datasets of SOTA AIOR approaches.

Method	Within-distribution Datasets
DiffUIR [51]	RESIDE [20] (Haze), Rain13K [47] and Rain1400 [14] (Rain), Snow100k [26] (Snow), GoPro [28] (Motion blur), LOLv1 [41] (low-light)
Diff-Plugin [25]	RESIDE [20] (Haze), Rain13K [47] and Rain1400 [14] (Rain), Snow100k [26] (Snow), GoPro [28] (Motion blur), LOLv1 [41] (low-light)
InstructIR [13]	RESIDE [20] (Haze), Rain13K [47] and Rain1400 [14] (Rain), GoPro [28] (Motion blur), LOLv1 [41] (low-light)
AutoDIR [18]	RESIDE [20] (Haze), Rain13K [47] and Rain1400 [14] (Rain), GoPro [28] (Motion blur), LOLv1 [41] (low-light) and Raindrop [32] (Raindrop)