Towards Real-World Adverse Weather Image Restoration: Enhancing Clearness and Semantics with Vision-Language Models

Supplementary Material

Jiaqi Xu[1], Mengyang Wu[1], Xiaowei Hu[2,⋆],
Chi-Wing Fu[1], Qi Dou[1], and Pheng-Ann Heng[1]

[1] The Chinese University of Hong Kong
[2] Shanghai Artificial Intelligence Laboratory

The supplementary material is structured into three main sections. First, it expands on the details of the methodology in Appendix A. Second, Appendix B delves into additional details regarding the datasets and the implementations. Finally, Appendix C presents supplementary experimental results, including more quantitative and qualitative analyses, and additional ablation studies.

A Network and Training Details

A.1 Network Structure

Our semi-supervised learning framework has compatibility with a variety of image restoration networks. We adopt MSBDN [17] as our backbone due to its balanced performance and rapid inference speed in our main study. MSBDN is originally an image dehazing network based on U-Net [34] architecture, which shows effectiveness in all-in-one adverse weather image restoration [5]. Three components comprise MSBDN, with an encoder, decoder, and in-between feature restoration module. The intermediate feature maps are of strides $1/1, 1/2, 1/4, 1/8, 1/16$ regarding the input image, with downsampling operations in the encoder and upsampling operations in the decoder. Several boosting and fusion techniques are proposed to enhance the image restoration capability. Please refer to MSBDN [17] for more details. Moreover, our proposed framework is also compatible with more advanced image restoration networks, *e.g.*, Restormer [51] and NAFNet [3]. We provide additional evidence for improving Restormer to restore real adverse weather images in Appendix C.10.

A.2 Description-Assisted Semantic Enhancement

In our approach, we employ vision-language models (VLMs) to provide a natural language description of the adverse weather image, encompassing rich semantic information about the scene and adverse weather conditions. Negative

⋆ Corresponding author (huxiaowei@pjlab.org.cn)

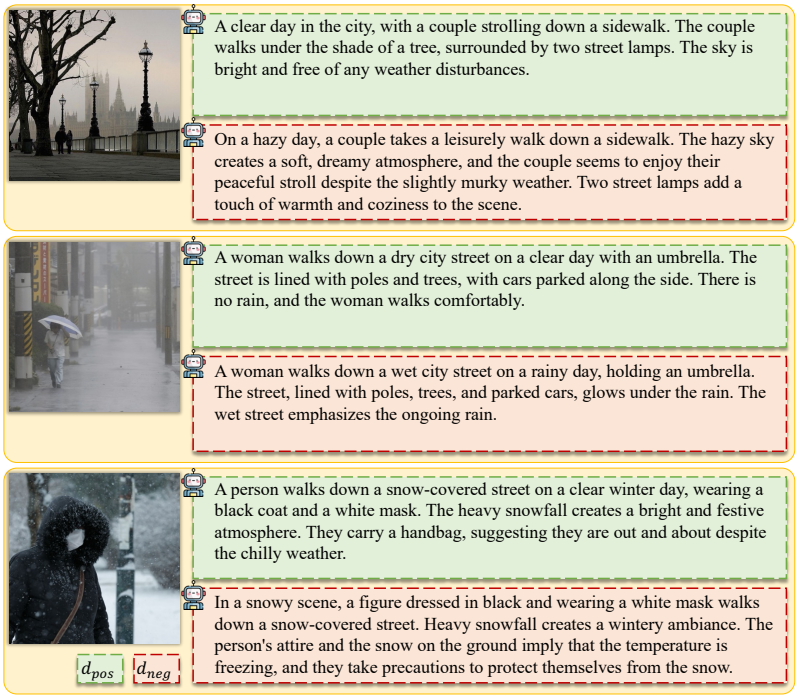

Fig. A: Examples of positive descriptions d_{pos} and negative descriptions d_{neg}.

descriptions d_{neg} pertain to the degraded images affected by adverse weather, while positive descriptions d_{pos} relate to their corresponding restored states. To achieve this, large VLMs, *e.g.*, LLaVA, are utilized to generate d_{neg}. Afterward, we prompt LLMs, *e.g.*, Llama, to produce d_{pos}. During implementation, LLaVA-v1.5 [23] and Llama 2 [37] are used for the description generation and conversion, respectively. We show examples of generated d_{pos} and d_{neg} in Fig. A.

A.3 VLM-based Visibility Assessment VLM-Vis

We employ a VLM-based image visibility assessment method to refine pseudo labels and evaluate restoration performance. For the VLM-Vis score computation, we utilize VLM experts, *i.e.*, a diverse ensemble of N VLM models [9, 23, 24, 35, 48]. For each image, we compute the VLM-based visibility score r_j^{vlm} for a given VLM j, following the procedure delineated in the method section. Recognizing that each VLM may yield scores within a distinct range, we standardize these by computing the minimum, $r_{min_j}^{vlm}$, and maximum $r_{max_j}^{vlm}$, score statistics across the dataset for each respective VLM. The final normalized visibility score, VLM-Vis, for an image is then determined using the following formula:

$$\text{VLM-Vis} = \frac{1}{N}\sum_{j=1}^{N}\frac{r_j^{vlm} - r_{min_j}^{vlm}}{r_{max_j}^{vlm} - r_{min_j}^{vlm}} \ . \tag{1}$$

Fig. B: Examples of the collected real rain and snow images.

B Datasets and Implementation Details

B.1 Dataset Details

Real data. We explore the real data to enhance image restoration performance across diverse adverse weather conditions in real-world settings. For the haze, we use the unannotated real-world hazy images (URHI) set of RESIDE [18], since it covers various haze levels and background scenes. We filter out the low-resolution images and retain the ones with noticeable haze effects, resulting in around 2,300 images. For the rain and snow categories, we choose to manually collect the real images to cover a wide range of scene and artifact levels since there exist limited real datasets with only narrow scene coverage [20] (*i.e.*, rain) or no suitable data available (*i.e.*, snow). Several searching keywords are first designed for image retrieval on the Internet, *e.g.*, *heavy rain*, *rainfall*, and *rain in* <*place*>, where <*place*> is the name of countries or cities. Then, we manually examine the retrieved images and keep only the images with legible rain and snow artifacts. In such a way, we build the real rain and snow datasets with diverse scenes and artifact levels of around 2,400 and 2,000 images, respectively. Note that these real images are also used to train the weather prompts together with the DF2K (DIV2K [36] and Flickr2K [21]) dataset. We show examples of the collected real rain and snow images in Fig. B.

Synthetic data. Following [38, 54], the (pseudo-)synthetic datasets [18, 19, 27, 32, 40, 54] are utilized in our semi-supervised learning as labeled data. Outdoor-Rain [19] is a synthetic rain dataset considering rain streaks and rain veiling effects. RainDrop [32] is for raindrop removal, capturing photos through glass with or without sprayed water. SPA [40, 54] is a pseudo-synthetic dataset that takes real rain images for the degraded ones, and the ground-truth counterparts are approximated by processing the rainy video to select the pixels without rain streaks. However, the obtained ground truth is pseudo-clear due to the approximation and is not effective for the removal of the rain veiling (haze) effect, which is common in heavy rain situations. OTS is the outdoor training set of RESIDE [18], which synthesizes the hazy image I based on the estimated depth maps d of the clear input J and the atmospheric scattering model:

Table A: Quantitative comparisons for rain and haze situations using B-FEN [44] and FADE [10]. **Bold** and underline indicate the best and second best, respectively.

Method	Restormer [51]	TransWeather [38]	TKL [5]	WeatherDiff [29]	WGWS-Net [54]	MWDT [30]	PromptIR [31]	DA-CLIP [28]	Our method
B-FEN ↑	0.330	0.329	0.332	0.329	<u>0.354</u>	0.325	0.330	0.327	**0.360**
FADE ↓	2.520	2.584	1.474	2.568	1.293	**0.978**	1.721	1.518	<u>1.120</u>

$I(x) = J(x)t(x) + A(1 - t(x)), t(x) = e^{-\beta d(x)}$, with haze amount variations β and environmental light variations A. Snow100K [27] is a synthetic snow dataset, which simulates the snowflakes using PhotoShop.

B.2 Other Implementation Details

We utilize existing deraining [4,7,14,15], dehazing [45,53], desnowing [11], general image restoration [51], and all-in-one adverse weather image restoration [5,28–31, 38,54] methods to produce the restoration results. The pseudo-label initialization is achieved using the VLM-based image assessment methods and majority voting among VLM experts, with image quality also considered.

During training, the learning rate is set to $1e-4$, and the cosine annealing schedule is adopted. Cropped regions (224×224) or (256×256) from images are used as patches for training.

C Additional Experimental Results

C.1 More Quantitative Results

In addition to the image quality assessment metrics and the proposed visibility assessment metric, VLM-Vis, detailed in the main text, we include further quantitative comparisons in our analysis. These comparisons utilize two additional metrics: B-FEN [44] and FADE [10]. B-FEN [44] is a neural network-based model that predicts the quality of images after deraining, which has been trained on a database of images assessed for deraining quality with mean opinion scores (MOS). On the other hand, FADE [10] estimates the visibility within foggy images by analyzing their statistical characteristics.

The comparative results are presented in Table A. As observed, our method outperforms others in terms of the B-FEN score, indicating superior deraining image quality. While our method achieves the second-best FADE score, narrowly trailing behind the MWDT [30] approach, it is important to note that MWDT tends to over-enhance the images, leading to notable color distortion. Furthermore, as reported in the main text, our method demonstrates the highest image quality assessment scores, and obtains a preference among users, suggesting an optimal balance between technical performance and visual appeal.

Furthermore, we evaluate PSNR and SSIM on paired "real" datasets [1,2,52] for comparisons; see Table B. Our method performs the best on most metrics.

Table B: Quantitative comparisons on paired O-HAZE [1], GT-RAIN [2], and WeatherStream [52] datasets using PSNR and SSIM.

Method	Restormer [51]	TransWeather [38]	TKL [5]	WeatherDiff [29]	WGWS-Net [54]	MWDT [30]	PromptIR [31]	DA-CLIP [28]	Our method
O-HAZE	18.07	14.71	18.77	17.48	16.54	15.32	18.20	17.48	**19.20**
	0.661	0.587	0.688	0.636	0.657	0.631	0.655	0.636	**0.693**
GT-RAIN	19.89	20.01	19.65	19.52	20.19	18.74	**21.01**	19.52	20.88
	0.696	0.682	0.664	0.686	0.690	0.675	0.704	0.653	**0.704**
Weather-Stream	20.30	20.63	20.13	20.35	20.82	18.31	**21.86**	19.85	21.11
	0.776	0.773	0.762	0.774	0.780	0.743	0.784	0.758	**0.788**

Table C: Comparisons on high-level vision tasks. Object detection results (mAP) for restored images are evaluated on the RIS [20] and RTTS [18] datasets.

Method	Restormer [51]	TransWeather [38]	TKL [5]	WeatherDiff [29]	WGWS-Net [54]	MWDT [30]	PromptIR [31]	DA-CLIP [28]	Our method
mAP (RIS)	0.218	0.201	0.204	0.213	0.175	0.216	0.212	<u>0.219</u>	**0.231**
mAP (RTTS)	0.504	0.472	0.506	0.491	0.497	<u>0.525</u>	0.513	0.511	**0.532**

Note that PSNR and SSIM are inadequate for assessing image restoration, as high scores may not align with good visual outcomes. Also, real images claimed in these datasets represent only a narrow slice of variability, and models trained on such datasets still struggle with real-world data, as shown in Fig. I.

C.2 More Qualitative Results

More visual comparisons on real images are shown in Figs. K to P. Our method effectively eliminates most weather-related artifacts, maintains the integrity of the image content, and yields visually pleasing restoration outcomes that are applicable to diverse real-world scenarios.

C.3 Downstream Applications

In this section, we evaluate the performance of our proposed method against others on high-level computer vision tasks, *e.g.*, object detection, under challenging weather conditions. For these experiments, we leverage two real adverse weather benchmark datasets with ground-truth annotations, RIS [20] and RTTS [18], for evaluation under rain and haze conditions, respectively. Specifically, RIS is the rain in surveillance set of the MPID [20] dataset, which provides images with rain veiling effects and includes annotated bounding boxes for object detection. RTTS is the real-world task-driven testing set of the well-known haze evaluation dataset, RESIDE [18]. A RetinaNet [22] model is employed for the evaluation of object detection performance. The detection model is applied to images that have been restored using various image restoration methods, and the resulting detection metrics are compared.

Fig. C: Visual comparisons on high-level vision tasks. Images are from RTTS [18] (first three rows) and RIS [20] (last two rows). Our method benefits object detection with fewer false negatives by providing clearer images with more visible regions.

Quantitative and visual results are presented in Table C and Fig. C, respectively. The results indicate that images restored using our method largely enhance the performance of high-level vision tasks, demonstrating a marked improvement over competing methods. Additionally, we conduct ablation experiments on our baseline model without semantic regularization, yielding mean Average Precision (mAP) scores of 0.217 for RIS [20] and 0.524 for RTTS [18], which is lower than the model trained with our full semi-supervised learning approach. These results underscore the contribution of semantic regularization to enhancing our model's performance in both image restoration and associated high-level vision tasks. Experiments on these datasets validate the efficacy of our approach in enhancing the robustness of vision systems under real adverse weather conditions by exploring the semantics in the images. Notably, the diminished performance observed on the RIS dataset can be attributed to factors beyond adverse weather, such as image blur and compression artifacts. Addressing these additional challenges to optimize image quality for high-level vision tasks is a direction for our future research.

C.4 Efficiency Comparison

Table D reports the inference FLOPs and runtimes (input size of 256×256) on a TITAN RTX GPU. Training time (estimated on two A40 GPUs) is also shown. Note that our method does not increase the backbone's inference time.

Table D: Inference and training efficiency comparisons.

Method	TransWeather	WeatherDiff	WGWS-Net	MWDT	PromptIR	DA-CLIP	Our method
FLOPs	4.7G	126.9T	71.0G	47.4G	172.7G	13.3T	19.6G
Inference	0.01s	17.56s	0.06s	0.05s	0.12s	3.77s	0.02s
Training	7h	90h	56h	71h	87h	105h	176h

Table E: Quantitative comparisons with methods using unlabeled data.

	NIMA ↑	MUSIQ ↑	CLIP-IQA ↑	LIQE ↑	Q-Align ↑	VLM-Vis ↑
SS-IRR [42]	**5.125**	56.99	0.383	2.397	3.322	0.334
DerainCycleGAN [43]	5.024	54.30	0.400	2.239	3.399	0.340
MOSS [15]	5.102	56.32	0.416	2.382	3.503	0.347
MUSS [14]	5.020	55.93	0.408	2.353	3.506	0.348
D4 [49]	5.073	54.82	0.405	2.163	3.169	0.344
NLCL [49]	5.069	57.43	0.363	2.152	3.253	0.330
Our method	5.084	**59.34**	**0.456**	**2.640**	**3.574**	**0.387**

C.5 Discussion on Utilizing Unlabeled Data

Leveraging the unlabeled data through semi-supervised or unsupervised learning boosts performance, particularly when the labeled data is limited or exhibits domain discrepancies. Existing works tackling adverse weather image restoration employ several strategies, such as Mean Teacher [14, 15, 26], cycle consistency [6, 8, 25, 43, 47], adversarial learning [46, 49, 50], and distribution regularization [42]. Some of them further incorporate disentanglement techniques tailored to address specific weather conditions. Note that the use of real unlabeled data extends benefits to other low-level vision tasks, including underwater image restoration [16]. However, these methods require specific imaging properties for the physical layer extraction and feature disentanglement, such as the rain [25,42,49] and haze [26,46,47] layers, which may not be readily adaptable to more complex all-in-one weather scenarios. Furthermore, certain studies mainly focus on utilizing unlabeled data to improve the performance of synthetic evaluations rather than addressing real-world scenarios [15,26]. In contrast, our semi-supervised learning framework is not confined to any specific weather condition. Instead, it capitalizes on the generalization capabilities of vision-language models for image clearness assessment and semantics regularization. This method is thus more versatile and better suited to a broad spectrum of complex and varied adverse weather conditions.

The quantitative comparisons with methods leveraging unlabeled data are shown in Table E. Results show that our method performs the best on most metrics. Further, we compare our method with a semi-supervised learning deraining method, MOSS [15], in Fig. D. Our method demonstrates superior performance in addressing rain streaks and substantially improves visibility under heavy rain conditions. Furthermore, our approach extends its applicability to a broader range of weather conditions, not limited to rain.

Fig. D: Visual comparisons with a semi-supervised deraning method.

Fig. E: Visual comparisons of the proposed weather prompt learning $\mathcal{L}_{wpl}$ with GANs. Three rows are the input, the results using GAN [41], and the results using $\mathcal{L}_{wpl}$.

C.6 Analysis on the Weather Prompt Learning

This section delves into a thorough analysis of the proposed weather prompt learning loss, denoted as $\mathcal{L}_{wpl}$. We initiate our discussion by comparing $\mathcal{L}_{wpl}$ with conventional adversarial learning approaches. Subsequently, we perform an in-depth implementation analysis of $\mathcal{L}_{wpl}$.

We compare our proposed weather prompt learning approach with conventional adversarial learning techniques, applying both methods to the set of images in clear, rain, haze, and snow conditions. The adversarial learning experiments leverage a U-Net-based discriminator, as detailed by Wang *et al.* [41]. The visual outcomes of this comparative study are illustrated in Fig. E. We observe that the adversarial approach, represented by GAN, displays limited capability in eliminating weather-induced distortions—this is particularly evident in images compromised by haze. We hypothesize that GANs struggle to initialize the notion of clarity within heavily degraded regions due to the inherent complexity of distinguishing between weather effects and image content. Conversely, the proposed weather prompt learning benefits from the expansive knowledge embedded in the large pre-trained vision-language model, which provides a more

Fig. F: Visual analysis of the pseudo-labels for the weather prompt learning with different CLIP [33] image encoders and position embedding strategies. Please zoom in for a better comparison.

nuanced understanding of favorable and adverse weather conditions. Hence, it translates to superior performance in weather artifact mitigation.

Besides, we have shown that the proposed weather prompt learning loss $\mathcal{L}_{wpl}$ effectively mitigates weather-related artifacts and enhances visual visibility in the processed images. Yet, in our preliminary experiments involving $\mathcal{L}_{wpl}$, we identified several key factors that significantly influence both the training process and the final restoration outcomes. These factors include the image encoder and the position embedding in the CLIP [33] model.

To demonstrate this, we employ several typical CLIP image encoders, such as ViT-B/32, ViT-L/14, and RN101, in monitoring the updated pseudo-labels. Our investigation is partly inspired by the findings from CLIP-IQA [39], which suggest that position embedding can influence image quality. To understand these effects, we examine variants both with and without position embedding.

As depicted in Fig. F, the choice of image encoder and position embedding strategy significantly impacts the visual quality of the pseudo-labels, consequently affecting the performance of the resulting restoration models. Our observations reveal that position embedding is particularly crucial for image encoders based on ViT [12]. Image encoders with position embedding are prone to generating strip-like noise, which manifests differently between ViT-B and ViT-L. Conversely, removing the position embedding mitigates such noise but can lead to severe color distortion. On the other hand, image encoders based on ResNet [13] seem less affected by these factors and consistently produce higher image quality, with fewer noise and color artifacts. Given these findings, we selected RN101 as the CLIP image encoder for our experiments. Further research is expected to uncover the intrinsic reasons within the CLIP model that lead to these behaviors in image manipulation.

Fig. G: Visual comparisons of pseudo-label selection by VLM and experts of VLMs.

Fig. H: Visual examples of pseudo-labels during training.

C.7 Impact of the VLM Experts

To counter potential biases that a single VLM might have towards certain image appearances, we employ a diverse ensemble of VLMs, each with different architectures and parameters, to serve as experts for image quality assessment. LLaVA-v1.5 [23] is the primary VLM used during training to assess image visibility. As previously shown, LLaVA-v1.5 is adept at selecting pseudo-labels with minimal weather-related artifacts, thereby improving the overall restoration performance. However, it occasionally encounters challenges with pseudo-label selection. To address this, we choose multiple VLMs as experts to refine the pseudo-labeling process. This panel includes LLaVA-v1.5 [23], LLaVA-v1.6 [24], mPLUG-Owl2 [48], InternVL [9], and Emu2 [35]. They collaborate during the pseudo-label initialization and update. As evident in Fig. G, the collective insights of these VLM experts help rectify the selections made by a single VLM, leading to enhanced training efficacy.

C.8 Training Progress

Figure H and Table F show how pseudo-labels improve progressively at each training round, continuously increasing in quality. The better pseudo-labels lead to improved overall training outcomes in turn.

C.9 Influence on Training Data

WeatherStream [52] attempted to compile a dataset of real degenerated images with corresponding ground truth for adverse weather image restoration. Nonetheless, this dataset was marred by low image quality issues, *e.g.*, compression artifacts and low resolution, primarily due to its reliance on compressed YouTube sources. Besides, the ground truth collected in this dataset is generated using some heuristics, *i.e.*, not the authentic, clear ones. We conduct experiments

Table F: Psdueo-label improvement with training.

	Input	Initial	Round 1	Round 2	Round 3	Round 4
MUSIQ ↑	49.25	51.81	53.20	53.68	54.03	54.26
CLIP-IQA ↑	0.547	0.621	0.642	0.657	0.665	0.670
VLM-Vis ↑	0.280	0.308	0.322	0.334	0.341	0.346

Table G: Comparisons of training using WeatherStream [52] data.

	NIMA ↑	MUSIQ ↑	CLIP-IQA ↑	LIQE ↑	Q-Align ↑	VLM-Vis ↑
WeatherStream	4.809	50.67	0.333	1.764	3.099	0.321
Our method	**5.084**	**59.34**	**0.456**	**2.640**	**3.574**	**0.387**

to validate the effectiveness of models trained using our method in restoring images affected by real adverse weather conditions compared to models trained on the WeatherStream dataset.

Table G and Figure I present the quantitative and the qualitative results, respectively. The model trained with the WeatherStream dataset is observed to frequently introduce significant noises, resulting in substantial degradation of image quality. Meanwhile, this model falls short of efficiently eliminating weather-related artifacts. This issue can be traced back to the inherent shortcomings of the WeatherStream ground truth collection process, which includes the use of unclear and noisy images for ground truth and the limitation to static scenes without dynamic elements. In contrast, our method advances the restoration of images under adverse weather by effectively leveraging unlabeled real-world data. It is a more practical solution for this problem and ensures that the restored images are clearer and more aesthetically pleasing.

C.10 More Advanced Backbone Network

Our proposed framework is also compatible with more advanced image restoration networks. This section briefly presents visual results for utilizing a more advanced network, Restormer [51], as the image restoration backbone.

As illustrated in Fig. J, the Restormer model, when trained using our method, significantly improves visibility in conditions of adverse weather relative to the baseline. Enhancing the granularity of image details by incorporating more powerful architectures is a direction for future research.

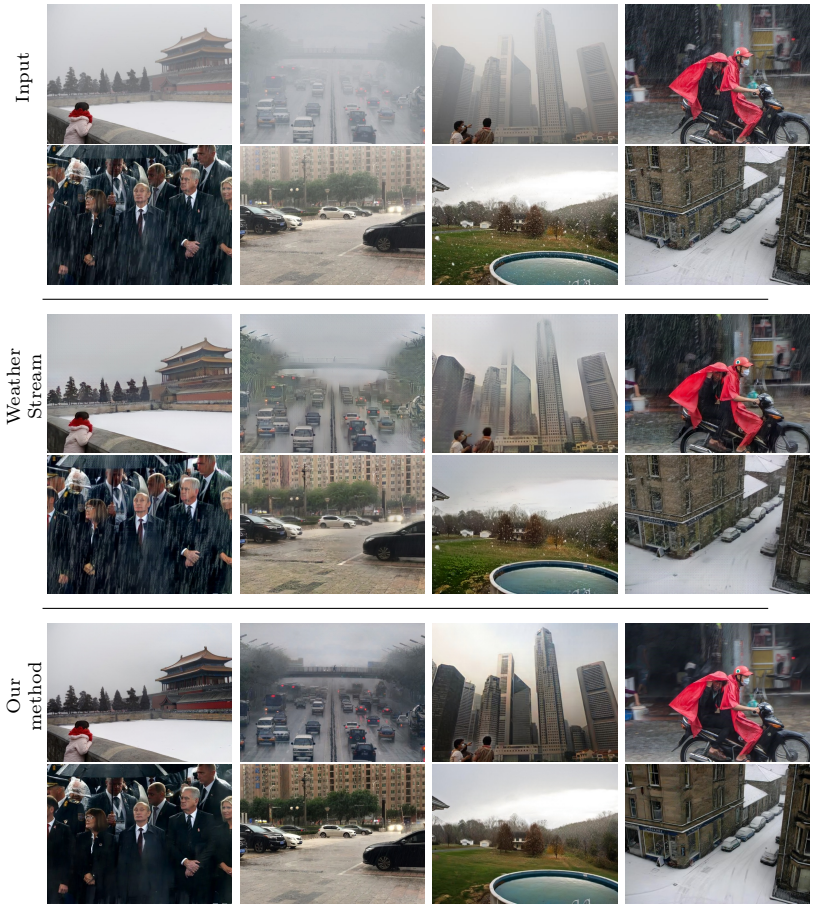

Fig. I: Visual comparisons of training using WeatherStream [52] data. The top parts are the input images, the middle parts are results from models trained using WeatherStream, and the bottom parts are the results of our method.

Fig. J: Visual comparisons of Restormer [51] using our semi-supervised learning. The last two rows are results from models trained without or with our method.

Fig. K: Visual comparisons on real-world images #1.

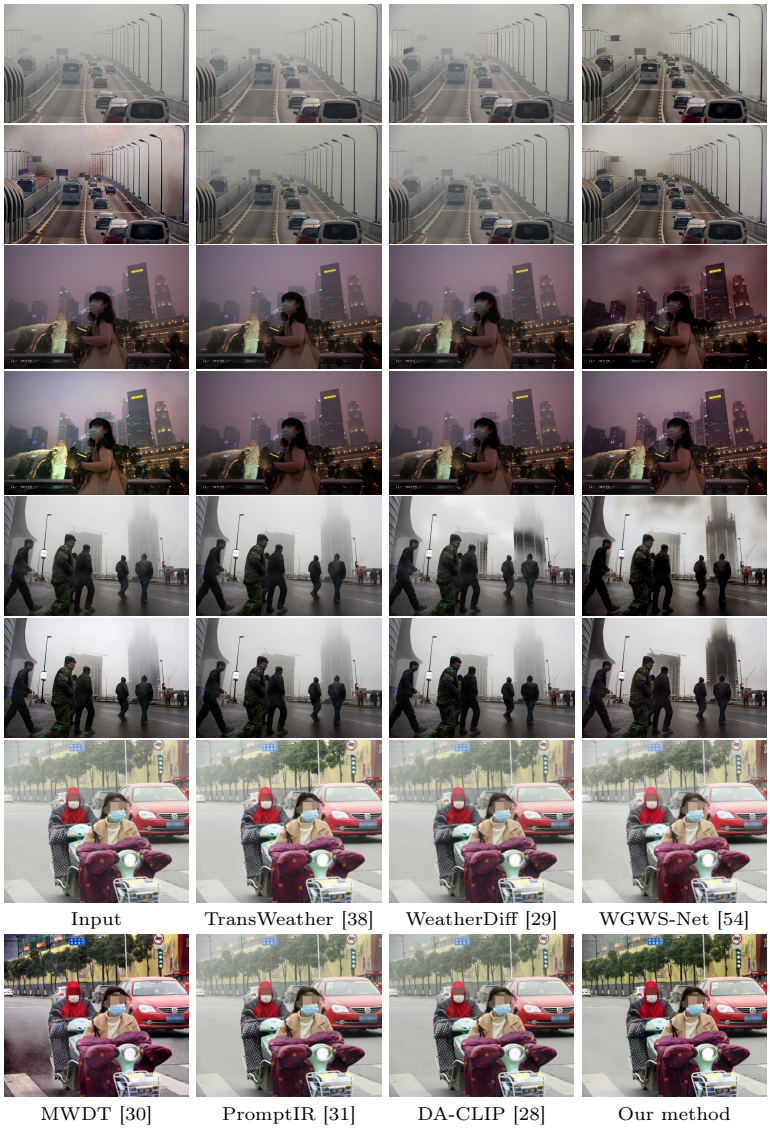

Fig. L: Visual comparisons on real-world images #2.

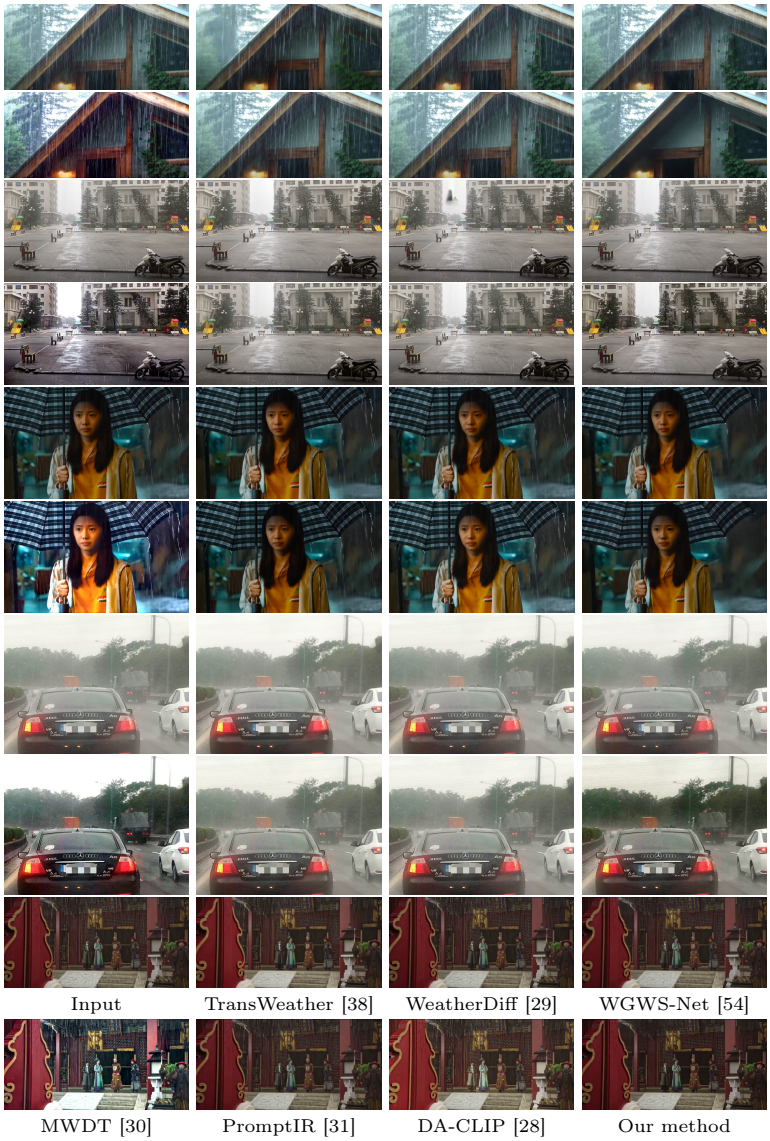

Fig. M: Visual comparisons on real-world images #3.

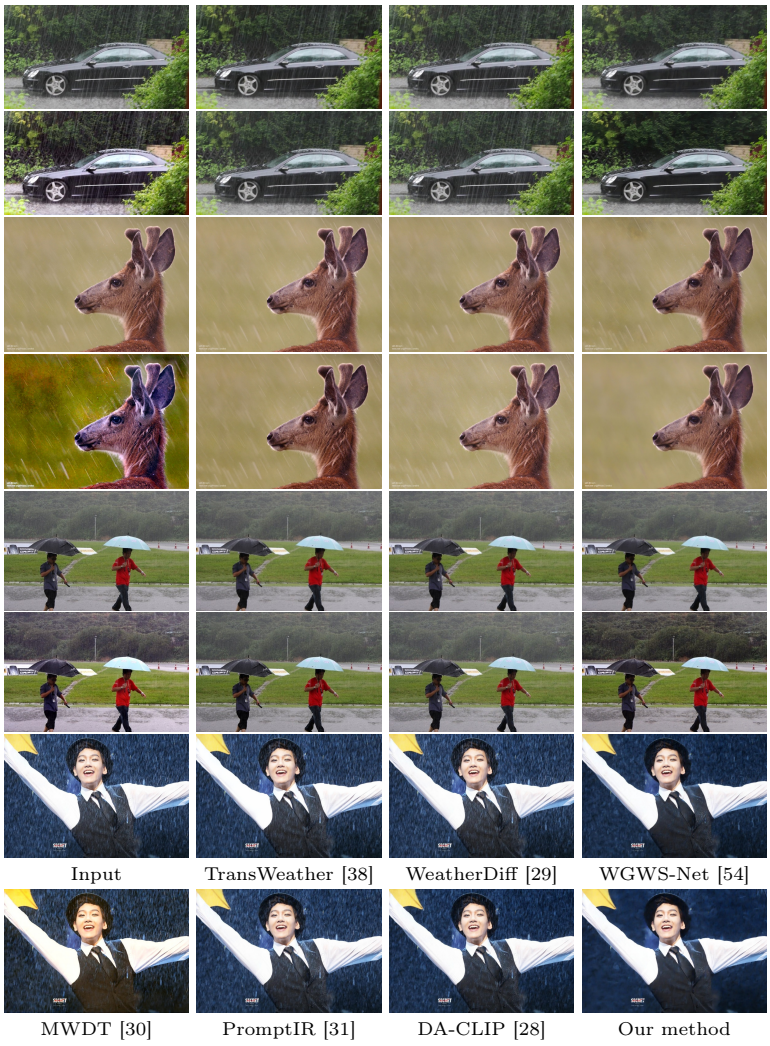

Fig. N: Visual comparisons on real-world images #4.

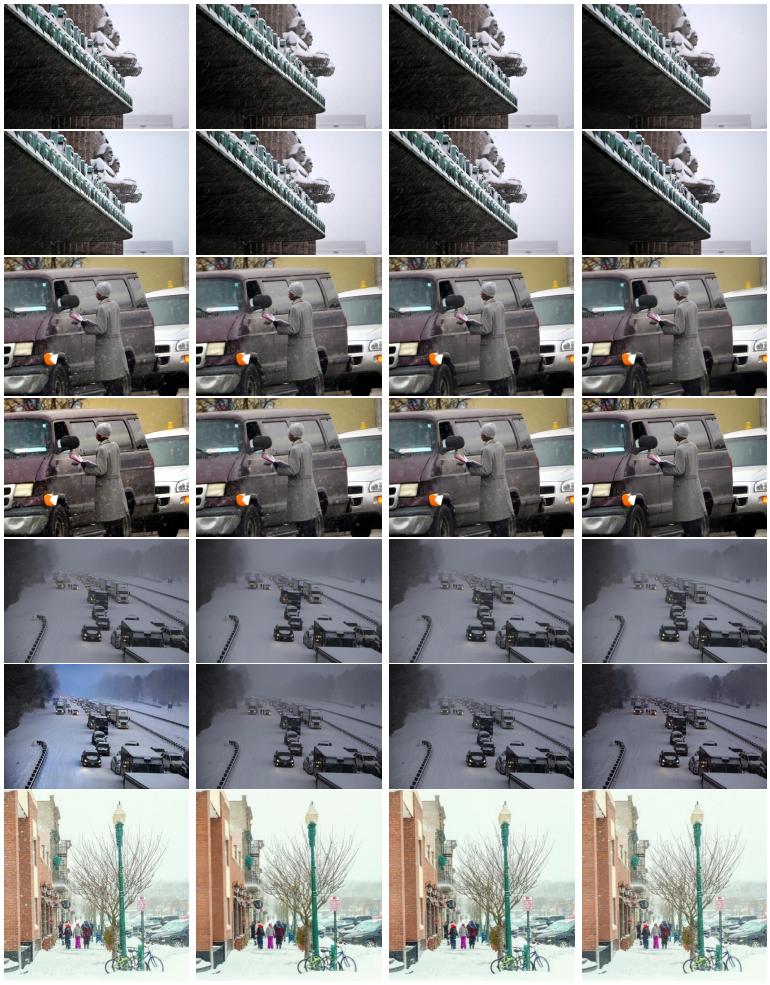

Fig. O: Visual comparisons on real-world images #5.

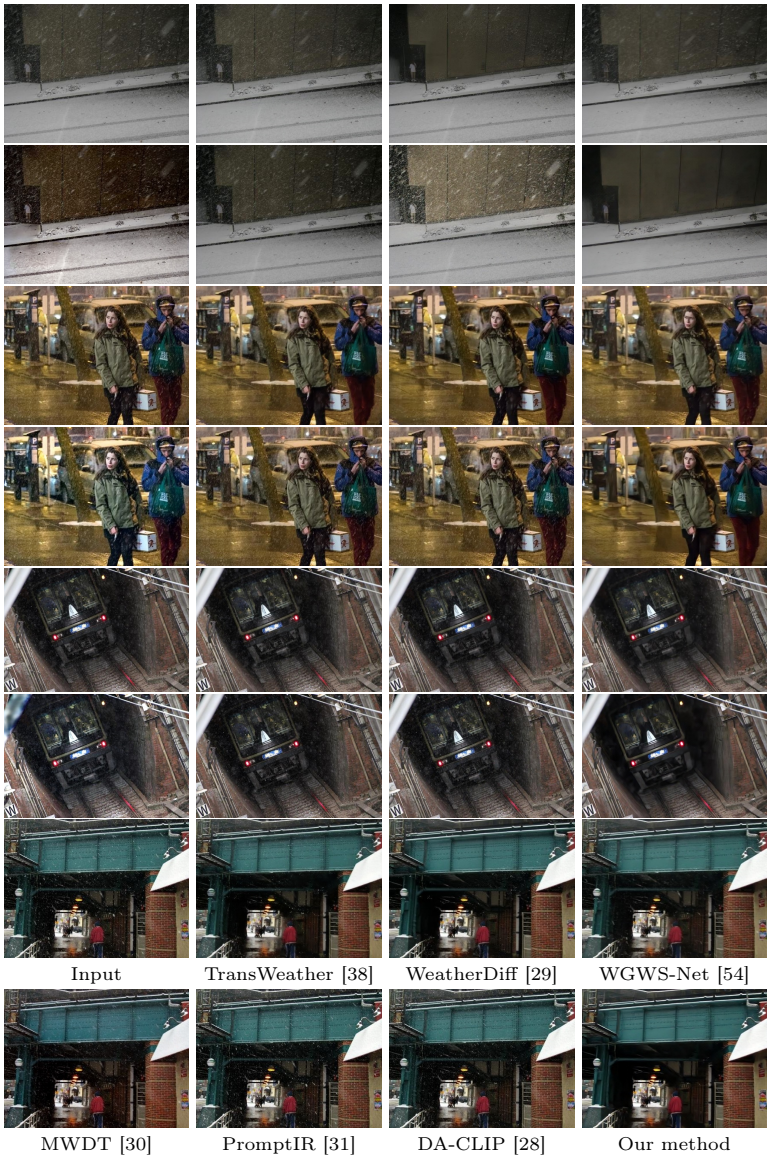

Fig. P: Visual comparisons on real-world images #6.

References

1. Ancuti, C.O., Ancuti, C., Timofte, R., De Vleeschouwer, C.: O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In: CVPRW (2018)
2. Ba, Y., Zhang, H., Yang, E., Suzuki, A., Pfahnl, A., Chandrappa, C.C., de Melo, C.M., You, S., Soatto, S., Wong, A., et al.: Not just streaks: Towards ground truth for single image deraining. In: ECCV (2022)
3. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: ECCV (2022)
4. Chen, S., Ye, T., Bai, J., Chen, E., Shi, J., Zhu, L.: Sparse sampling transformer with uncertainty-driven ranking for unified removal of raindrops and rain streaks. In: ICCV (2023)
5. Chen, W.T., Huang, Z.K., Tsai, C.C., Yang, H.H., Ding, J.J., Kuo, S.Y.: Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In: CVPR (2022)
6. Chen, X., Fan, Z., Li, P., Dai, L., Kong, C., Zheng, Z., Huang, Y., Li, Y.: Unpaired deep image dehazing using contrastive disentanglement learning. In: ECCV (2022)
7. Chen, X., Li, H., Li, M., Pan, J.: Learning a sparse transformer network for effective image deraining. In: CVPR (2023)
8. Chen, X., Pan, J., Jiang, K., Li, Y., Huang, Y., Kong, C., Dai, L., Fan, Z.: Unpaired deep image deraining using dual contrastive learning. In: CVPR (2022)
9. Chen, Z., Wu, J., Wang, W., Su, W., Chen, G., Xing, S., Zhong, M., Zhang, Q., Zhu, X., Lu, L., et al.: Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In: CVPR (2024)
10. Choi, L.K., You, J., Bovik, A.C.: Referenceless prediction of perceptual fog density and perceptual image defogging. TIP (2015)
11. Cui, Y., Ren, W., Cao, X., Knoll, A.: Focal network for image restoration. In: ICCV (2023)
12. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: ICLR (2021)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
14. Huang, H., Luo, M., He, R.: Memory uncertainty learning for real-world single image deraining. TPAMI (2022)
15. Huang, H., Yu, A., He, R.: Memory oriented transfer learning for semi-supervised image deraining. In: CVPR (2021)
16. Huang, S., Wang, K., Liu, H., Chen, J., Li, Y.: Contrastive semi-supervised learning for underwater image restoration via reliable bank. In: CVPR (2023)
17. Jiang, K., Wang, Z., Yi, P., Chen, C., Huang, B., Luo, Y., Ma, J., Jiang, J.: Multi-scale progressive fusion network for single image deraining. In: CVPR (2020)
18. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. TIP (2018)
19. Li, R., Cheong, L.F., Tan, R.T.: Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In: CVPR (2019)
20. Li, S., Araujo, I.B., Ren, W., Wang, Z., Tokuda, E.K., Junior, R.H., Cesar-Junior, R., Zhang, J., Guo, X., Cao, X.: Single image deraining: A comprehensive benchmark analysis. In: CVPR (2019)
21. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: CVPRW (2017)

22. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV (2017)
23. Liu, H., Li, C., Li, Y., Lee, Y.J.: Improved baselines with visual instruction tuning. In: CVPR (2024)
24. Liu, H., Li, C., Li, Y., Li, B., Zhang, Y., Shen, S., Lee, Y.J.: Llava-next: Improved reasoning, ocr, and world knowledge (January 2024), https://llava-vl.github.io/blog/2024-01-30-llava-next/
25. Liu, Y., Yue, Z., Pan, J., Su, Z.: Unpaired learning for deep image deraining with rain direction regularizer. In: ICCV (2021)
26. Liu, Y., Zhu, L., Pei, S., Fu, H., Qin, J., Zhang, Q., Wan, L., Feng, W.: From synthetic to real: Image dehazing collaborating with unlabeled real data. In: ACM MM (2021)
27. Liu, Y.F., Jaw, D.W., Huang, S.C., Hwang, J.N.: Desnownet: Context-aware deep network for snow removal. TIP (2018)
28. Luo, Z., Gustafsson, F.K., Zhao, Z., Sjölund, J., Schön, T.B.: Controlling vision-language models for universal image restoration. In: ICLR (2024)
29. Özdenizci, O., Legenstein, R.: Restoring vision in adverse weather conditions with patch-based denoising diffusion models. TPAMI (2023)
30. Patil, P.W., Gupta, S., Rana, S., Venkatesh, S., Murala, S.: Multi-weather image restoration via domain translation. In: ICCV (2023)
31. Potlapalli, V., Zamir, S.W., Khan, S., Khan, F.S.: Promptir: Prompting for all-in-one blind image restoration. In: NeruIPS (2023)
32. Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J.: Attentive generative adversarial network for raindrop removal from a single image. In: CVPR (2018)
33. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: ICML (2021)
34. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
35. Sun, Q., Cui, Y., Zhang, X., Zhang, F., Yu, Q., Wang, Y., Rao, Y., Liu, J., Huang, T., Wang, X.: Generative multimodal models are in-context learners. In: CVPR (2024)
36. Timofte, R., Agustsson, E., Van Gool, L., Yang, M.H., Zhang, L.: Ntire 2017 challenge on single image super-resolution: Methods and results. In: CVPRW (2017)
37. Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al.: Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288 (2023)
38. Valanarasu, J.M.J., Yasarla, R., Patel, V.M.: Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In: CVPR (2022)
39. Wang, J., Chan, K.C., Loy, C.C.: Exploring clip for assessing the look and feel of images. In: AAAI (2023)
40. Wang, T., Yang, X., Xu, K., Chen, S., Zhang, Q., Lau, R.W.: Spatial attentive single-image deraining with a high quality real rain dataset. In: CVPR (2019)
41. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: ICCVW (2021)
42. Wei, W., Meng, D., Zhao, Q., Xu, Z., Wu, Y.: Semi-supervised transfer learning for image rain removal. In: CVPR (2019)
43. Wei, Y., Zhang, Z., Wang, Y., Xu, M., Yang, Y., Yan, S., Wang, M.: Deraincyclegan: Rain attentive cyclegan for single image deraining and rainmaking. TIP (2021)

44. Wu, Q., Wang, L., Ngan, K.N., Li, H., Meng, F., Xu, L.: Subjective and objective de-raining quality assessment towards authentic rain image. TCSVT (2020)
45. Wu, R.Q., Duan, Z.P., Guo, C.L., Chai, Z., Li, C.: Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In: CVPR (2023)
46. Yang, X., Xu, Z., Luo, J.: Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In: AAAI (2018)
47. Yang, Y., Wang, C., Liu, R., Zhang, L., Guo, X., Tao, D.: Self-augmented unpaired image dehazing via density and depth decomposition. In: CVPR (2022)
48. Ye, Q., Xu, H., Ye, J., Yan, M., Hu, A., Liu, H., Qian, Q., Zhang, J., Huang, F.: mplug-owl2: Revolutionizing multi-modal large language model with modality collaboration. In: CVPR (2024)
49. Ye, Y., Yu, C., Chang, Y., Zhu, L., Zhao, X.l., Yan, L., Tian, Y.: Unsupervised deraining: Where contrastive learning meets self-similarity. In: CVPR (2022)
50. Yu, C., Chen, S., Chang, Y., Song, Y., Yan, L.: Both diverse and realism matter: Physical attribute and style alignment for rainy image generation. In: ICCV (2023)
51. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: CVPR (2022)
52. Zhang, H., Ba, Y., Yang, E., Mehra, V., Gella, B., Suzuki, A., Pfahnl, A., Chandrappa, C.C., Wong, A., Kadambi, A.: Weatherstream: Light transport automation of single image deweathering. In: CVPR (2023)
53. Zheng, Y., Zhan, J., He, S., Dong, J., Du, Y.: Curricular contrastive regularization for physics-aware single image dehazing. In: CVPR (2023)
54. Zhu, Y., Wang, T., Fu, X., Yang, X., Guo, X., Dai, J., Qiao, Y., Hu, X.: Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In: CVPR (2023)