DTMIR-Pro: Domain Translation with Prompt-based Latent-Space Generalization for Multi-Weather Image Restoration

Supplementary Material

Overview of Supplementary Material

The supplementary material includes:

1. Key Differences between PIR [13] and the Proposed Method

The key differences between PIR [13] and the proposed method are highlighted in Table S 1. In brief, the proposed method handles multiple simultaneous weather degradations, and has better input adaptability in accordance to the weather prompts. Further, the proposed method has better performance in terms of computational complexity. Also, the quantitative metrics verify the effectiveness of the proposed method over PIR [13].

Table S 1. Key differences between existing method PIR [13] and proposed (Ours) method.

	PIR [13]	**Ours**
Tasks	Handles Multiple Individual Degradation Removal (where only single weather degradation is present in each image).	Handles Multiple Simultaneous Degradation Removal (where **more than one degradation** may be present in each image).
Prompts	Input Feature Attentive	Input Feature Attentive + **Weather Code Aware**
Prompt Position	Restoration Decoder	Domain Translation Encoder
Parameters (M)	35.5M	**3.9M**
GFLOPs	158.15	**28.60**
Runtime (sec/image)	0.15	**0.06**
PSNR on SOTS Dataset	30.58	**36.92**
SSIM on SOTS Dataset	0.974	**0.981**

2. Detailed Explanation of Dynamic Attention

The schematic of the proposed dynamic attention block is represented in Figure S 1. The proposed dynamic attention block consists of multi-head attention, where the queries (Q_d) are extracted using dynamic convolutions [3]. In the dynamic convolutions [3], initially, the attentive weights ($\pi_n(x), n \in [1, N]$) are extracted to be

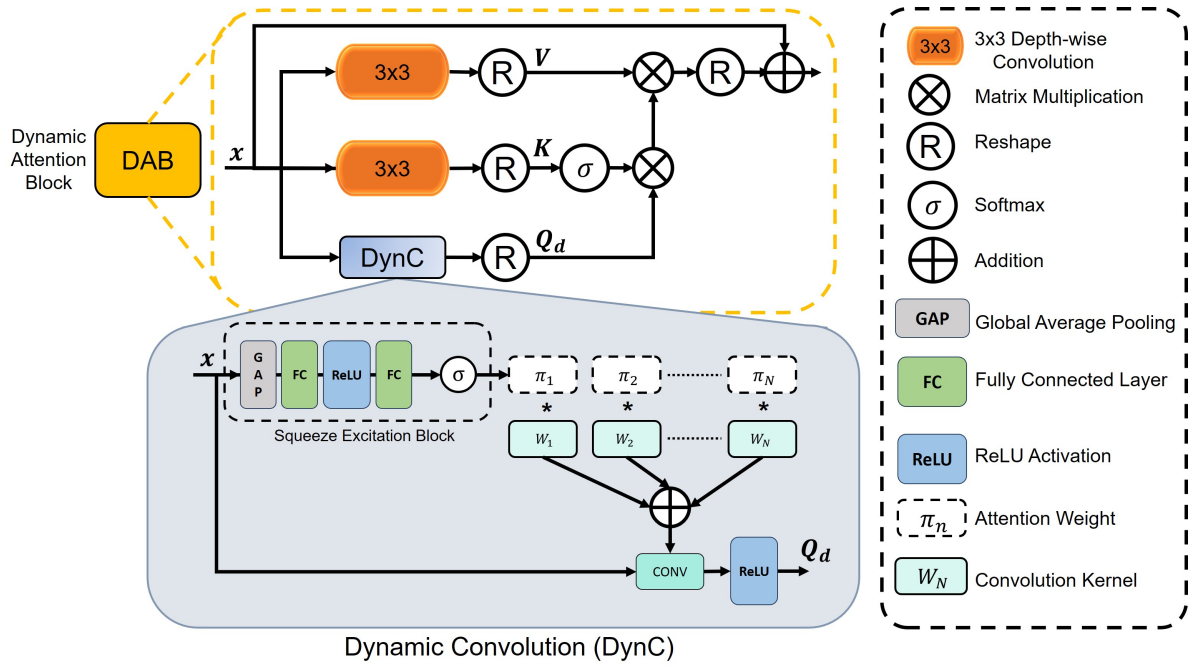

Figure S 1. Architecture of the proposed Dynamic Attention Block.

multiplied with convolution kernels (W_n). The attentive weights are obtained by passing the input features x through squeeze excitation block [4] consisting of a global average pooling layer, shrinking and expanding fully connected layers and ReLU activation function and finally, softmax activation for obtaining the attention weights (π_n). Each weight is then multiplied with respective convolution kernel (W_n). These kernels are then aggregated to get a convolution layer $(DynC)$:

$$DynC = \sum_{n=1}^{N} \pi_n(x) W_n \tag{1}$$

Further, for query extraction from input features x, this dynamic convolution is used as:

$$Q_d = ReLU(DynC(x)) \tag{2}$$

The keys (K), and the values (V) are calculated as $K = C_{3\times3}^{depth}(x)$, $V = C_{3\times3}^{depth}(x)$ respectively for input features x extracted using depthwise convolutions $(C_{3\times3}^{depth}(\cdot))$. Finally, the dynamic attention on input features (x), *i.e.,* $\mathcal{D}_{Att}(x)$ is calculated as:

$$\mathcal{D}_{Att}(x) = x + \sigma(Q_d \cdot K^T) \cdot V \tag{3}$$

where, $\sigma(\cdot)$ is the softmax activation.

3. Detailed Equations of Loss Functions

Here, we elaborate on the loss functions used to optimize the networks in the proposed method.

3.1. Domain Translation Loss Functions

For domain translation, we have the domain translated images $\mathbb{I}_i$, and domain translation ground truth $\mathbb{I}_{gt_i}^{DT}, i \in \{Haze, Rain, Snow\}$. With these, the contrastive loss $\mathcal{CL}_{DT}$ is calculated as:

$$\mathcal{CL}_{DT} = \frac{\left\|\phi_k\left(\mathbb{I}_i\right) - \phi_k\left(\mathbb{I}_{gt_i}^{DT}\right)\right\|_1}{\left\|\phi_k\left(\mathbb{I}_i\right) - \phi_k\left(\mathbb{I}_j\right)\right\|_1} \tag{4}$$

where, $i, j \in \{Haze, Rain, Snow\}$; $i \neq j$. $\phi_k(\cdot)$ is the k^{th} $(k \in \{3, 8, 15\})$ layer of pretrained VGG-16 model [16].

3.2. Multi-Weather Restoration Loss Functions

For multi-weather image restoration, we have the restored image $\mathbb{I}_R$, the restoration ground-truth $\mathbb{I}_{gt}^R$, and the domain translated images $\mathbb{I}_i, i \in \{Haze, Rain, Snow\}$. With these, the L_1 loss is calculated as:

$$L_1 = \left\|\mathbb{I}_R - \mathbb{I}_{gt}^R\right\|_1 \tag{5}$$

and the contrastive loss $\mathcal{CL}_R$ is calculated as:

$$\mathcal{CL}_R = \frac{\left\|\phi_k\left(\mathbb{I}_R\right) - \phi_k\left(\mathbb{I}_{gt}^R\right)\right\|_1}{\left\|\phi_k\left(\mathbb{I}_R\right) - \phi_k\left(\mathbb{I}_i\right)\right\|_1} \tag{6}$$

where, $\phi_k(\cdot)$ is the k^{th} $(k \in \{3, 8, 15\})$ layer of pretrained VGG-16 model [16].

4. Qualitative Results on Synthetic and Real-World Datasets

In this section, we provide more qualitative results on synthetic (SOTS [8], ORD [7], and CSD [1]), and real-world (RTTS [8], RID [9], and Snow-Realistic [10]) datasets for multi-weather image restoration.

- The qualitative results on SOTS dataset are provided in Figure S 2.
- The qualitative results on ORD dataset are provided in Figure S 3.
- The qualitative results on CSD dataset are provided in Figure S 4.
- The qualitative results on RTTS dataset are provided in Figure S 5.
- The qualitative results on RID dataset are provided in Figure S 6.
- The qualitative results on Snow-Realistic dataset are provided in Figure S 7.
- The qualitative results on night-time weather degraded images are provided in Figure S 8.

5. Domain Translation Results Comparison

As restoration performance depends on diversity of domains produced, domain translation network needs to generate a diverse range of domains in an effective manner. To verify this, we provide qualitative results in comparison with existing domain-translation based restoration method DMIR [12] in Figure S 9 and Figure S 10. As seen from the results, the proposed domain translation strategy produces more diverse weather degraded results as compared to DMIR [12], leading to a more robust restoration network.

6. Ablation on Feature Routing

We tested PSNR and MSE as routing metrics, but these pixel-wise measures are dominated by global intensity differences and often misidentify the degradation. SSIM, by jointly evaluating luminance, contrast, and structure, is more sensitive to local distortions and consistently selected the correct translation branch, improving no-reference scores. Table S 2 further confirms that SSIM routing outperforms PSNR and MSE routing across NIQE, Entropy, and BRISQUE, validating our choice.

Table S 2. Ablation on routing metrics for no-reference evaluation averaged on RTTS, RID, and Snow-Realistic datasets.

Routing Technique $\rightarrow$	**PSNR**	**MSE**	**SSIM**
Avg. NIQE ($\downarrow$)	5.132	4.982	**4.506**
Avg. Entropy ($\uparrow$)	7.042	7.188	**7.565**
Avg. BRISQUE ($\downarrow$)	24.893	23.671	**22.491**

7. Applications on Various Computer Vision Tasks

Restoration capability of any algorithm is utilized in many higher-level computer vision tasks. Results on applicability of the proposed method on tasks of depth estimation [14], object detection [15], and semantic segmentation [14] are given in Figure S 11. As seen from the figure, the performance of these applications is better when applied to restored outputs using our method as compared to degraded images.

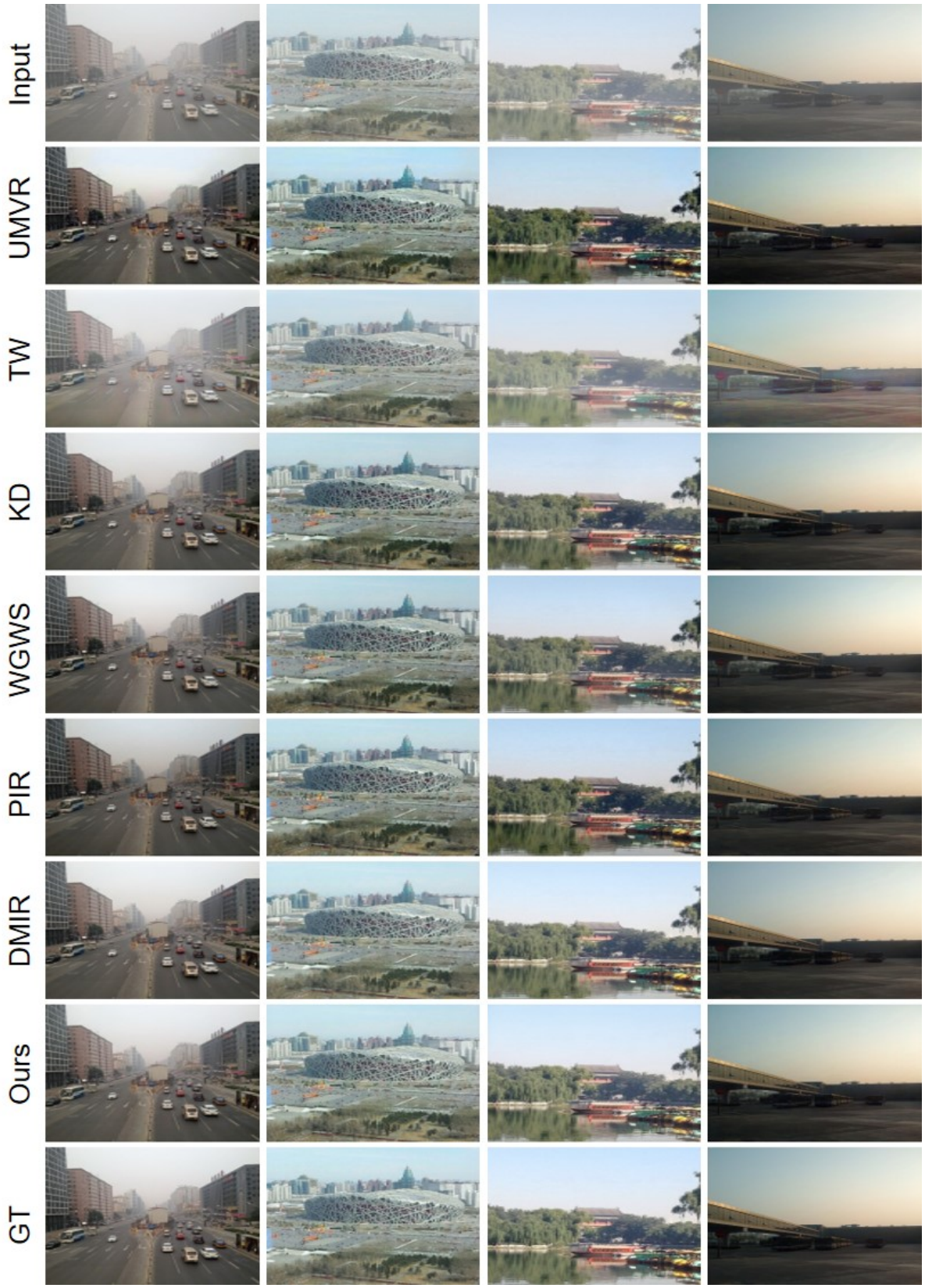

Figure S 2. Qualitative results comparison with the proposed method (Ours) and state-of-the-art methods UMVR [6], TW [5], KD [2], WGWS [17], PIR [13], and DMIR [12] on SOTS dataset.

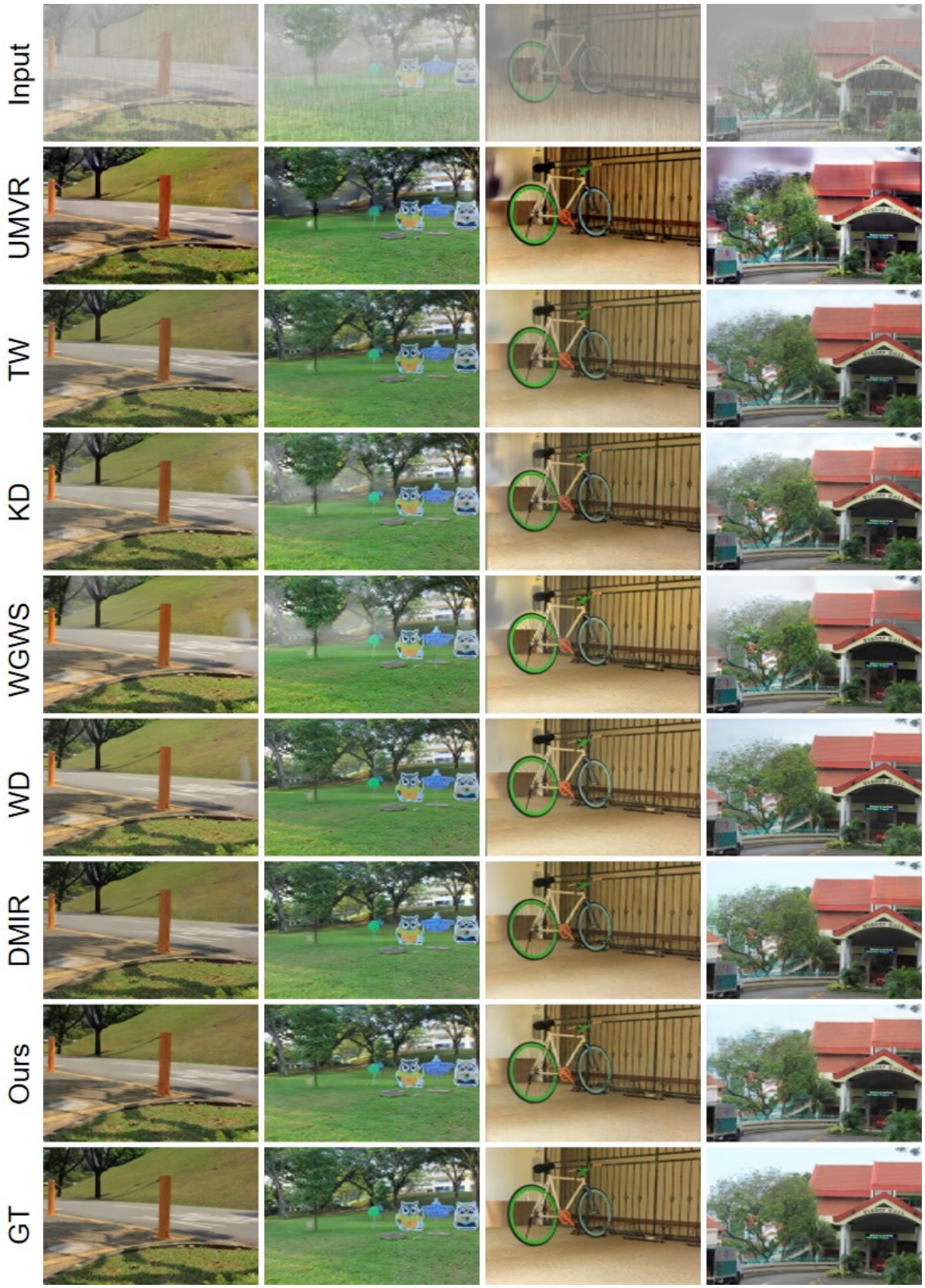

Figure S 3. Qualitative results comparison with the proposed method (Ours) and state-of-the-art methods UMVR [6], TW [5], KD [2], WGWS [17], WD [11], and DMIR [12] on ORD dataset.

Figure S 4. Qualitative results comparison with the proposed method (Ours) and state-of-the-art methods UMVR [6], TW [5], KD [2], and DMIR [12] on CSD dataset.

Figure S 5. Qualitative results comparison with the proposed method (Ours) and state-of-the-art methods UMVR [6], TW [5], KD [2], WGWS [17], WD [11], and DMIR [12] on real-world hazy images.

Figure S 6. Qualitative results comparison with the proposed method (Ours) and state-of-the-art methods UMVR [6], TW [5], KD [2], WGWS [17], WD [11], and DMIR [12] on real-world rainy images.

Figure S 7. Qualitative results comparison with the proposed method (Ours) and state-of-the-art methods UMVR [6], TW [5], KD [2], WGWS [17], WD [11], and DMIR [12] on real-world snowy images.

Figure S 8. Qualitative results comparison with the proposed method (Ours) and state-of-the-art methods UMVR [6], TW [5], KD [2], WGWS [17], WD [11], and DMIR [12] on real-world weather degraded night-time images.

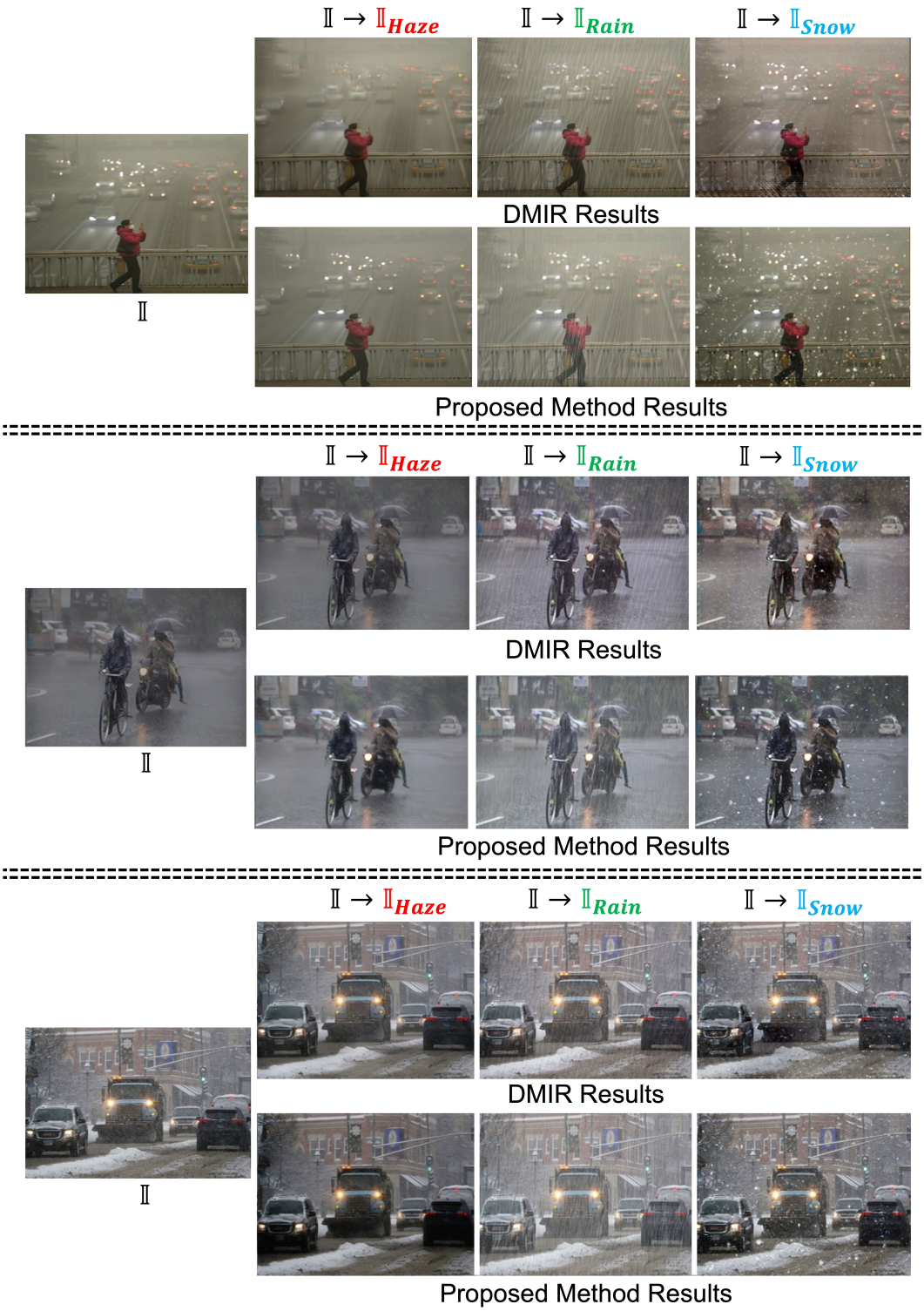

Figure S 9. Domain translation results comparison with DMIR [12] (*top row per image*) and the proposed method (*bottom row per image*). Here, $\mathbb{I}$ is the input weather degraded image, and $\mathbb{I}_{Haze}$, $\mathbb{I}_{Rain}$, $\mathbb{I}_{Snow}$ represent the domain translated images.

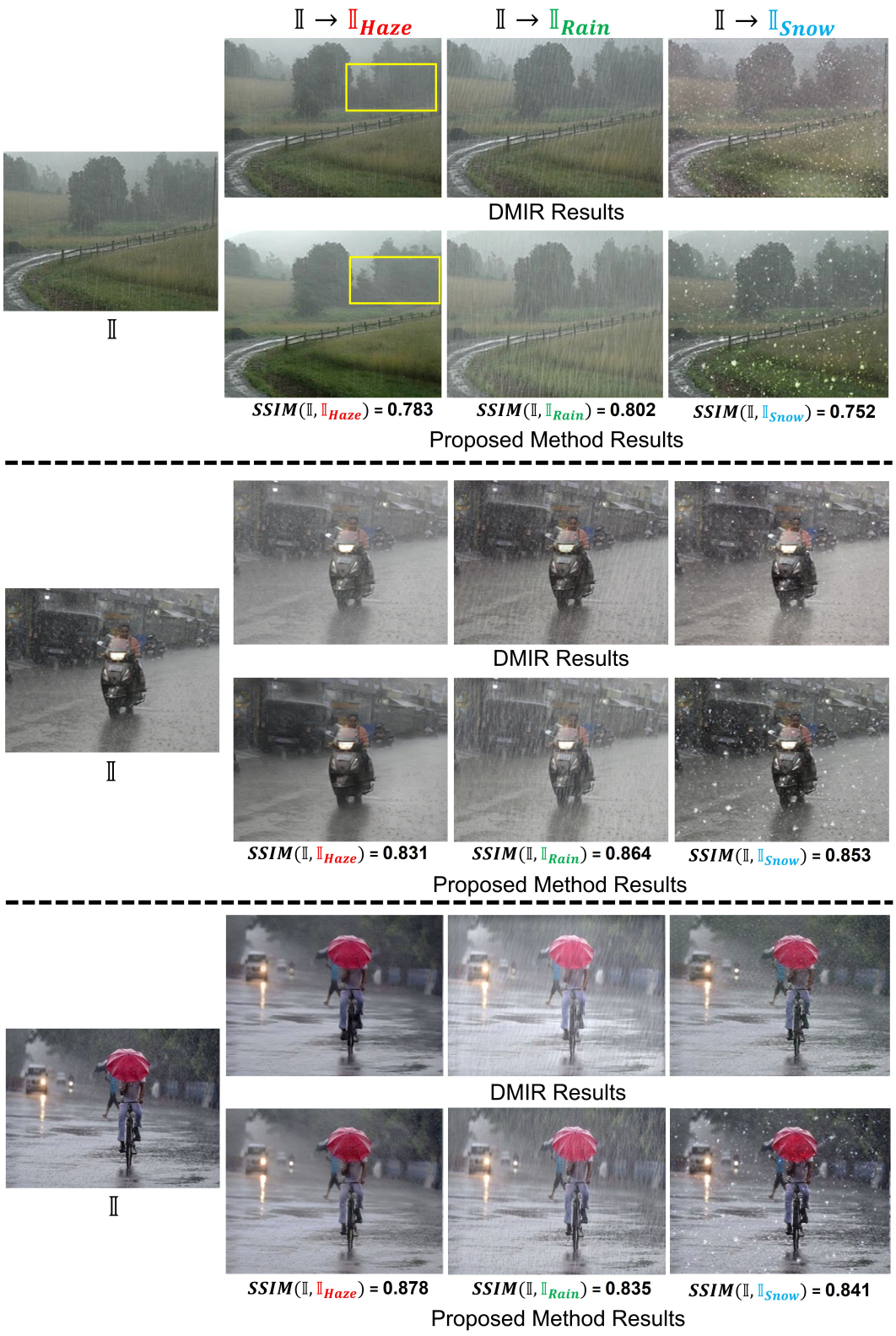

Figure S 10. Domain translation results comparison with DMIR [12] (*top row per image*) and the proposed method (*bottom row per image*). Here, $\mathbb{I}$ is the input weather degraded image, and $\mathbb{I}_{Haze}$, $\mathbb{I}_{Rain}$, $\mathbb{I}_{Snow}$ represent the domain translated images.

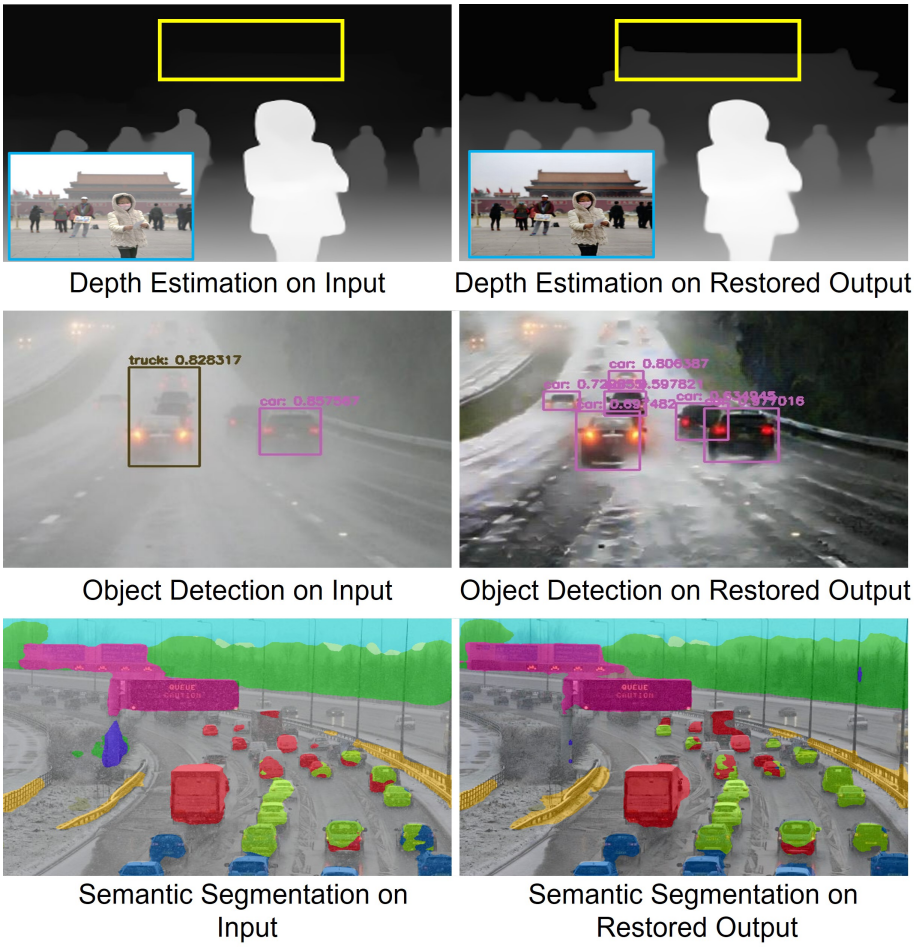

Figure S 11. Results analysis on depth estimation [14], object detection [15], and semantic segmentation [14].

References

[1] Wei-Ting Chen, Hao-Yu Fang, Cheng-Lin Hsieh, Cheng-Che Tsai, I Chen, Jian-Jiun Ding, Sy-Yen Kuo, et al. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4196–4205, 2021. 3

[2] Wei-Ting Chen, Zhi-Kai Huang, Cheng-Che Tsai, Hao-Hsiang Yang, Jian-Jiun Ding, and Sy-Yen Kuo. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17632–17641, 2022. 5, 6, 7, 8, 9, 10, 11

[3] Yinpeng Chen, Xiyang Dai, Mengchen Liu, Dongdong Chen, Lu Yuan, and Zicheng Liu. Dynamic convolution: Attention over convolution kernels. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11030–11039, 2020. 1

[4] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 2

[5] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M. Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2343–2353, 2022. 5, 6, 7, 8, 9, 10, 11

[6] Ashutosh Kulkarni, Prashant W. Patil, Subrahmanyam Murala, and Sunil Gupta. Unified multi-weather visibility restoration. *IEEE Transactions on Multimedia*, pages 1–13, 2022. 5, 6, 7, 8, 9, 10, 11

[7] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision*, pages 4770–4778, 2017. 3

[8] 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. 3

[9] Siyuan Li, Iago Breno Araujo, Wenqi Ren, Zhangyang Wang, Eric K Tokuda, Roberto Hirata Junior, Roberto Cesar-Junior, Jiawan Zhang, Xiaojie Guo, and Xiaochun Cao. Single image deraining: A comprehensive benchmark analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3838–3847, 2019. 3

[10] 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. 3

[11] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 6, 8, 9, 10, 11

[12] Prashant W. Patil, Sunil Gupta, Santu Rana, Svetha Venkatesh, and Subrahmanyam Murala. Multi-weather image restoration via domain translation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 21696–21705, October 2023. 3, 5, 6, 7, 8, 9, 10, 11, 12, 13

[13] Vaishnav Potlapalli, Syed Waqas Zamir, Salman Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one blind image restoration. *arXiv preprint arXiv:2306.13090*, 2023. 1, 5

[14] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12179–12188, 2021. 4, 14

[15] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 4, 14

[16] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 3

[17] Yurui Zhu, Tianyu Wang, Xueyang Fu, Xuanyu Yang, Xin Guo, Jifeng Dai, Yu Qiao, and Xiaowei Hu. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21747–21758, 2023. 5, 6, 8, 9, 10, 11