Multi-weather Image Restoration via Domain Translation
Supplementary Material

Overview

The supplementary material contains:

- Details of training losses
- Details of domain translation architecture: Figure S1
- Ablation study with and without MFE module for Domain Translation: Figure S2
- Restoration network ablation study: Figure S3 and Table S1
- Domain features alignment orders: Figure S4

1 Details of Training Losses

1.1 Domain Translation

The weather degraded image (I) is used as input to the domain translation network which produces the domain translated output ($I_{\mathbb{D}\epsilon(\mathbb{D}_1,\mathbb{D}_2,\mathbb{D}_3)}$). These translated images are similar to hazy (I_H), rain with veil (I_{Rv}), and snow with veil (I_{Sv}), respectively. Adversarial learning is adopted for domain translation network training. Along with $\mathbb{L}_1$ loss, adversarial loss ($\mathbb{L}_A$) is calculated as:

$$\mathbb{L}_A = \max_{\mathbf{D}_d} \min_{\mathbf{G}_d} \mathbb{E}[log(D_d(I, I_{\mathbb{T}}))] + \mathbb{E}[log(1 - D_d(I, G_d(I)))] \tag{1}$$

where, D_d and G_d are the domain-translation discriminator and generator respectively, $I_{\mathbb{T}\epsilon(H,Rv,Sv)}$ are the original target domains of respective translated images $G_d(I) = I_{\mathbb{D}\epsilon(\mathbb{D}_1,\mathbb{D}_2,\mathbb{D}_3)}$. To guide the network for textural and structural information, the perceptual loss ($\mathbb{L}_{V_d}$) is calculated between the restored and target image by passing them through the pre-trained VGG19 model [1] as:

$$\mathbb{L}_V = \sum_{s=1}^{S} \|\psi_s(I_{\mathbb{T}}) - \psi_s(I_{\mathbb{D}})\|_1 \tag{2}$$

where, ψ_s are the feature maps ($s\epsilon(1; S)$) of the VGG19 model. Further, the contrastive loss [2] is used to maximize the difference between the generated image (restored image) and the weather degraded input in a common latent feature space, and minimize the difference between generated image and ground-truth image in the common latent feature space. The contrastive loss ($\mathbb{L}_c$) is defined as:

$$\mathbb{L}_c(I, I_{\mathbb{D}}, I_{\mathbb{T}}) = \sum_{s=1}^{S} \beta_i * \frac{\|\psi_i(I_{\mathbb{T}}) - \psi_i(I_{\mathbb{D}})\|_1}{\|\psi_i(I) - \psi_i(I_{\mathbb{D}})\|_1} \tag{3}$$

where, β_i* is the weight given to each i^{th} layer ψ_i of a pretrained model of VGG-19. Further, the style of target image is matched with translated image through mean and standard deviation as:

$$\mathbb{L}_s = \sum_{s=1}^{S} \|\sigma(\psi_s(I_{\mathbb{T}})) - \sigma(\psi_s(I_{\mathbb{D}}))\|_1 + \sum_{s=1}^{S} \|\mu(\psi_s(I_{\mathbb{T}})) - \mu(\psi_s(I_{\mathbb{D}}))\|_1 \tag{4}$$

where, σ and μ represents mean and standard deviation of target and translated image. Therefore, the total loss ($\mathbb{L}_d$) is:

$$\mathbb{L}_d(\mathbb{D}_i, \mathbb{T}_i) = \lambda_1 \mathbb{L}_{1_d} + \lambda_2 \mathbb{L}_{A_d} + \lambda_3 \mathbb{L}_{V_d} + \lambda_4 \mathbb{L}_{C_d} + \lambda_5 \mathbb{L}_{s_d} \tag{5}$$

where, the weights $\lambda_{k\epsilon(1,5)}$ for each loss are set empirically ($\lambda_1 = 1$, $\lambda_2 = 0.01$, $\lambda_3 = 0.2$, $\lambda_4 = 0.2$, $\lambda_5 = 0.2$). The domain translation architecture is trained for each domain ($I_{\mathbb{D}_1}$, $I_{\mathbb{D}_2}$, $I_{\mathbb{D}_3}$) separately with this provided setting.

1.2 Image Restoration Losses

The weather degraded image (I) is used as input and output of the network is restored image (I_g). The proposed network is required to generate the restored image similar to the respective clean image (I_c). The $\mathbb{L}_1$ loss is used to optimize the network for better reconstruction. The adversarial loss is the min-max problem between generator and discriminator, respectively and given as:

$$\mathbb{L}_A = \underset{\mathbf{D}_r}{max}\ \underset{\mathbf{G}_r}{min}\ \mathbb{E}[log(D_r(I, I_c))] + \mathbb{E}[log(1 - D_r(I, G_r(I, I_{\mathbb{D}})))] \tag{6}$$

where, D_r and G_r are the restoration discriminator and generator respectively. To guide the network for textural and structural information, the perceptual loss ($\mathbb{L}_V$) is calculated between the restored and target image by passing them through the pre-trained VGG19 model [1] as:

$$\mathbb{L}_V = \sum_{s=1}^{S} \|\psi_s(I_c) - \psi_s(I_g)\|_1 \tag{7}$$

where, ψ_s are the feature maps ($s\epsilon(1; S)$) of the VGG19 model. Further, the contrastive loss [2] is used to maximize the difference between the generated image (restored image) and the weather degraded input in a common latent feature space, and minimize the difference between generated image and ground-truth image in the common latent feature space. The contrastive loss ($\mathbb{L}_c$) is defined as:

$$\mathbb{L}_c(I, I_g, I_T) = \sum_{s=1}^{S} \beta_i * \frac{\|\psi_i(I_c) - \psi_i(I_g)\|_1}{\|\psi_i(I) - \psi_i(I_g)\|_1} \tag{8}$$

where, β_i* is the weight given to each i^{th} layer ψ_i of a pretrained model of VGG-19. So, the total loss ($\mathbb{L}_r$) is:

$$\mathbb{L}_r(I_g, I_c) = \lambda_1 \mathbb{L}_1 + \lambda_2 \mathbb{L}_A + \lambda_3 \mathbb{L}_V + \lambda_4 \mathbb{L}_c \tag{9}$$

where, the weights $\lambda_{k\epsilon(1,4)}$ for each loss are set empirically ($\lambda_1 = 1$, $\lambda_2 = 0.01$, $\lambda_3 = 0.4$, $\lambda_4 = 0.2$).

2 Details of domain translation architecture

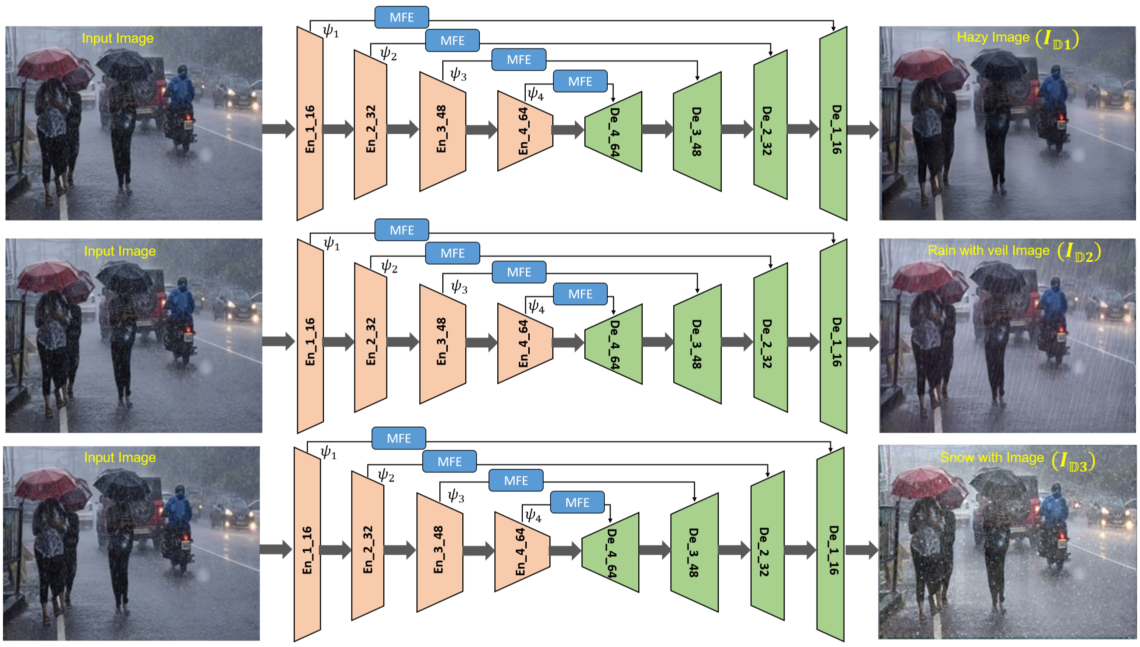

Figure S 1: Overview of the domain translation architecture.

3 Ablation study with and without MFE module for Domain Translation

We provided the visual results of domain translation with and without multi-attentive feature extraction in Figure S2 for your reference.

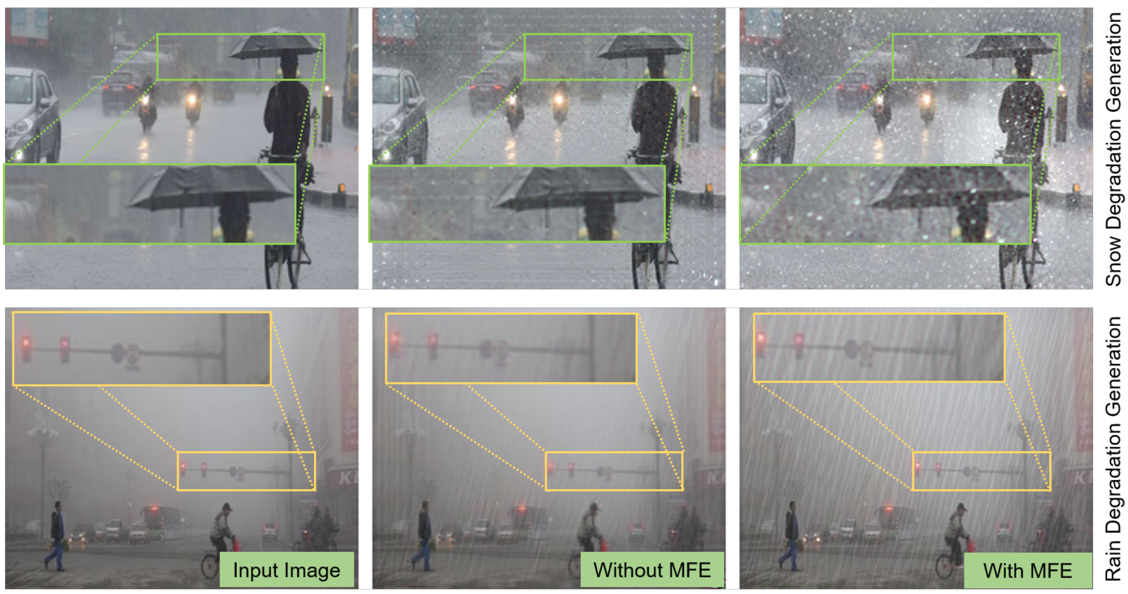

Figure S 2: Ablation study on domain translation with and without MFE module (*see zoomed version of patch*).

4 Restoration network ablation study analysis on SOTS and CSD database

Table S 1: Ablation study on SOTS and CSD databases

Module	SOTS	CSD
Baseline (BS)	28.63/0.921	28.32/0.879
BS + IL (**Net-1**)	33.23/0.934	29.01/0.899
BS + IL + PMDA (**Net-2**)	35.46/0.978	31.76/0.932
BS + IL + PMDA + CMA (**Proposed**)	**36.26/ 0.987**	**32.95/ 0.940**

Figure S 3: Visual results for the ablation study (see Table R2 for combination).

5 Ablation study with domain feature alignment orders

We can see that the HRS ($Haze \rightarrow Rain \rightarrow Snow$) alignment order helps to achieve better restoration as compared to RSH ($Rain \rightarrow Snow \rightarrow Haze$) and SHR ($Snow \rightarrow Haze \rightarrow Rain$) alignment orders. Based on our experience, aligning the features in the order of degradations that are increasingly "locally distinct" leads to better restoration. The haze are the least locally distinct as they usually come as a global patch; even though rain have long streaks, they are more locally distinct than haze, and the snow are the most locally distinct due to being small patches-like degradation.

Figure S 4: Visual results on domain features different alignment orders for image restoration.

References

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[2] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, "Contrastive learning for compact single image dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 551–10 560.

[3] A. Kulkarni, P. W. Patil, S. Murala, and S. Gupta, "Unified multi-weather visibility restoration," *IEEE Transactions on Multimedia*, 2022.

[4] W.-T. Chen, Z.-K. Huang, C.-C. Tsai, H.-H. Yang, J.-J. Ding, and S.-Y. Kuo, "Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 653–17 662.

[5] J. M. J. Valanarasu, R. Yasarla, and V. M. Patel, "Transweather: Transformer-based restoration of images degraded by adverse weather conditions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2353–2363.

[6] W.-T. Chen, H.-Y. Fang, C.-L. Hsieh, C.-C. Tsai, I. Chen, J.-J. Ding, S.-Y. 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*, 2021, pp. 4196–4205.

[7] R. Li, L.-F. Cheong, and R. T. Tan, "Heavy rain image restoration: Integrating physics model and conditional adversarial learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1633–1642.

[8] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018.