MOERL: When Mixture-of-Experts Meet Reinforcement Learning for Adverse Weather Image Restoration

Supplementary Material

Overview

In this document, we first provide more details of the Transformer layer. Then we conduct an additional ablation study to evaluate the effectiveness of the proposed MOE layer. Finally, we provide more visual comparisons to demonstrate our method's qualitative performance.

1. More details of Transformer layer

The MOE block combines a MOE layer and multiple Transformer units [8]. Each Transformer unit consists of a multi-dconv head transposed attention (MDTA) mechanism and a gated-dconv feed-forward network (GDFN). MDTA performs self-attention along the channel dimension, reducing computational complexity while preserving feature effectiveness. Given input features $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$, layer normalization is applied, followed by 1×1 and 3×3 depth-wise convolutions to generate the Query ($\mathbf{Q}$), Key ($\mathbf{K}$), and Value ($\mathbf{V}$). These tensors compute a channel-wise attention matrix to enhance feature interaction. This process can be expressed as

$$\begin{aligned} \mathrm{Att}(\hat{\mathbf{Q}}, \hat{\mathbf{K}}, \hat{\mathbf{V}}) &= \mathrm{softmax}\left(\frac{\hat{\mathbf{Q}}\hat{\mathbf{K}}^\top}{\alpha}\right)\hat{\mathbf{V}}, \\ \hat{\mathbf{X}} &= W_{1\times1}\,\mathrm{Att}(\hat{\mathbf{Q}}, \hat{\mathbf{K}}, \hat{\mathbf{V}}) + \mathbf{X}, \end{aligned} \tag{1}$$

where $\hat{\mathbf{X}}$ denotes the output features. GDFN adopts a gating mechanism to propagate useful information selectively. Specifically, given input $\mathbf{X}$, convolutions and element-wise operations controlled by a GELU activation function $\phi(\cdot)$ refine the feature representation. The process of GDFN is shown as

$$\begin{aligned} \mathrm{Gate} &= \phi\left(W^1_{3\times3}W^1_{1\times1}(LN(\mathbf{X}))\right) \odot \left(W^2_{3\times3}W^2_{1\times1}(LN(\mathbf{X}))\right), \\ \hat{\mathbf{X}} &= W_{1\times1}\mathrm{Gate}(\mathbf{X}) + \mathbf{X}, \end{aligned} \tag{2}$$

where LN is the normalization, $\hat{\mathbf{X}} \in \mathbb{R}^{C\times H\times W}$ represents the output features, Gate$(.)$ refers to the gated mechanism. $W_{1\times1}(.)$ and $W_{3\times3}(.)$ denote 1×1 convolution and 3×3 depth-wise convolution. $\odot$ is the element-wise multiplication operation. Together, MDTA and GDFN enable efficient feature extraction and transformation, ensuring robust restoration performance in adverse weather conditions.

2. Additional Ablation Study

Table 1 presents the results of an ablation study on the number of spatial and channel experts in the MOE layer, highlighting the trade-off between performance and computational efficiency. **Variant1**, with 4 spatial and 4 channel experts, achieves a PSNR of 32.65 and an SSIM of 0.9427, using 12.62M parameters and 68.58G FLOPs. Increasing the experts to 6 each in **Variant2** improves performance slightly to a PSNR of 32.76 and an SSIM of 0.9428, with a moderate increase in complexity (12.78M parameters and 69.57G FLOPs). **Variant3**, with 8 experts of each type, achieves the best performance (32.78 PSNR, 0.9430 SSIM) but incurs the highest computational cost (12.94M parameters and 70.58G FLOPs). Our chosen configuration, with 2 spatial experts and 4 channel experts, strikes the best balance, achieving a PSNR of 32.70 and an SSIM of 0.9428 while maintaining the lowest parameter count (12.50M) and computational demand (68.35G). These results confirm that our design achieves competitive performance with optimized efficiency.

Table 1. Ablation study settings. The MACs of each model is measured on 256×256 image.

Models	Spatial experts	Channel experts	PSNR/SSIM	Params.	FLOPs
Variant1	4	4	32.65/0.9427	12.62M	68.58G
Variant2	6	6	32.76/0.9428	12.78M	69.57G
Variant3	8	8	32.78/0.9430	12.94M	70.58G
Ours	2	4	32.70/0.9428	12.50M	68.35G

3. More Visual Results

This section presents additional visual comparisons in different datasets, including Snow100K-S [3], Snow100K-L [3], Outdoor-Rain [2], RainDrop [5], and real-world images from Snow100K [3]. Specifically, the comparison methods are TransWeather [7], Chen *et al.* [1], WGWS-Net [9], WeatherDiff$_{64}$ [4], and Histoformer [6].

Visual Results on Snow100K-S: Figures 1, 2.

Visual Results on Snow100K-L: Figures 3, 4.

Visual Results on Outdoor-Rain: Figures 5 6.

Results on RainDrop: Figures 7, 8.

Visual Results on real-world images: Figures 9, 10.

References

[1] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 2

[2] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 2, 4

[3] 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. 2, 4, 5

[4] 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*, 45(8):10346–10357, 2023. 2

[5] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2482–2491, 2018. 2, 5

[6] Shangquan Sun, Wenqi Ren, Xinwei Gao, Rui Wang, and Xiaochun Cao. Restoring images in adverse weather conditions via histogram transformer. In *Proceedings of European Conference on Computer Vision*, pages 111–129, 2025. 2

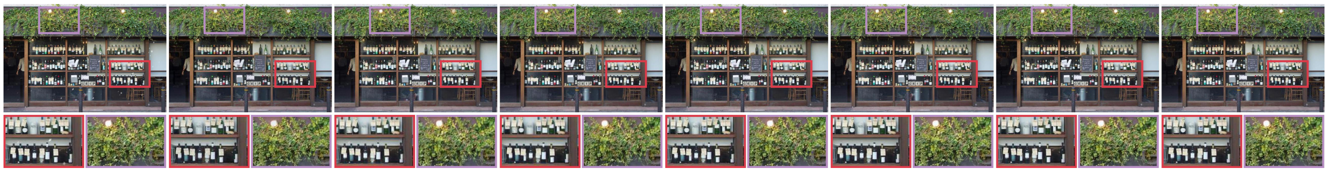
[7] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2353–2363, 2022. 2

[8] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 2

[9] 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 Conference on Computer Vision and Pattern Recognition*, 2023. 2

Input TransWeather Chen *et al.* WGWSNet WeatherDiff Histoformer Ours GT

Figure 1. Visual comparison on Snow100K-S dataset [3].

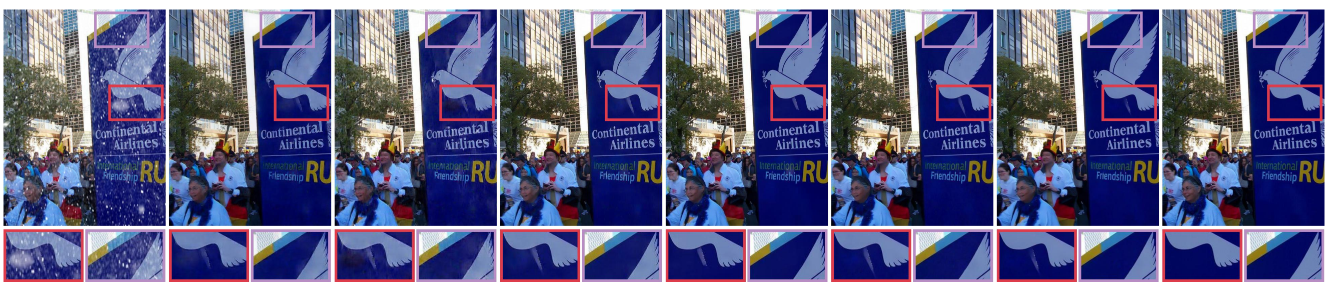


Input TransWeather Chen *et al.* WGWSNet WeatherDiff Histoformer Ours GT

Figure 2. Visual comparison on Snow100K-S dataset [3].

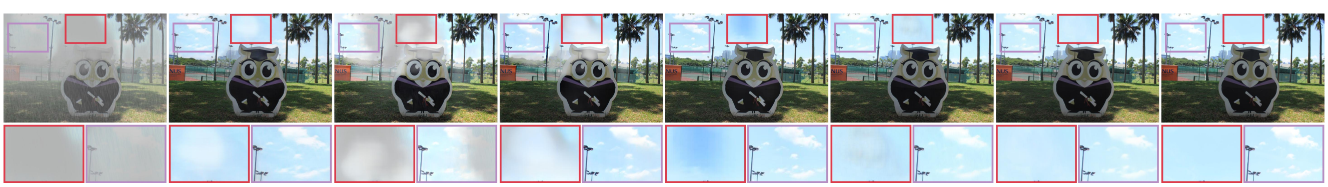
Input TransWeather Chen *et al.* WGWSNet WeatherDiff Histoformer Ours GT

Figure 3. Visual comparison on Snow100K-L dataset [3].



Input TransWeather Chen *et al.* WGWSNet WeatherDiff Histoformer Ours GT

Figure 4. Visual comparison on Snow100K-L dataset [3].



Input TransWeather Chen *et al.* WGWSNet WeatherDiff Histoformer Ours GT

Figure 5. Visual comparison on Outdoor-Rain dataset [2].

Input TransWeather Chen *et al.* WGWSNet WeatherDiff Histoformer Ours GT

Figure 6. Visual comparison on Outdoor-Rain dataset [2].

Input TransWeather Chen *et al.* WGWSNet WeatherDiff Histoformer Ours GT

Figure 7. Visual comparison on RainDrop dataset [5].

Input TransWeather Chen *et al.* WGWSNet WeatherDiff Histoformer Ours GT

Figure 8. Visual comparison on RainDrop dataset [5].

Input TransWeather Chen *et al.* WGWSNet WeatherDiff Histoformer Ours GT

Figure 9. Visual comparison on real-world images from Snow100K dataset [3].

Input TransWeather Chen *et al.* WGWSNet WeatherDiff Histoformer Ours GT

Figure 10. Visual comparison on real-world images from Snow100K dataset [3].