Supplementary Material – Adverse Weather Removal with Codebook Priors

Tian Ye[1,3*] Sixiang Chen[1,3*] Jinbin Bai[2 *] Jun Shi[4] Chenghao Xue[3]
Jingxia Jiang[3] Junjie Yin[3] Erkang Chen[3] Yun Liu[5†]

[1]The Hong Kong University of Science and Technology (Guangzhou) [2]National University of Singapore
[3]School of Ocean Information Engineering, Jimei University
[4]Xinjiang University [5]College of Artificial Intelligence, Southwest University
{*owentianye*, *sixiangchen*} @hkust-gz.edu.cn jinbin.bai@u.nus.edu, junshi2022@gmail.com,
{202021115097, 202021114006, 202021112006, *ekchen*} @jmu.edu.cn, yunliu@swu.edu.cn

This is a supplementary material for Adverse Weather Removal with Codebook Priors.

We provide the following materials in this manuscript:

- Sec.1 the detailed architecture of the pre-trained VQ-GAN network and our AWRCP.
- Sec.2 additional ablation studies.
- Sec.3 Comparison with existing codebook-based related works.
- Sec.4 more visual comparisons.
- Sec.5 future works, limitations and broader impacts.

1. The Detailed Architecture

1.1. Efficient Encoder Block

Our approach, the AWRCP, utilizes convolutional blocks to efficiently extract features from degraded images in the Encoder's Feature Extraction stage. To construct more distinctive features, we initially double the channel dimension relative to the input channels, as shown in Fig.1, inspired by previous research[2]. Next, we employ a design comprising DWConv-LN-GELU-DWConv to capture sufficient degradation information. Before reducing the number of channels, we add a channel attention mechanism [6] to improve channel interaction. Our encoder design is both simple and effective, as demonstrated by our model's performance, which ensures exceptional feature extraction capabilities while maintaining low inference time and computational complexity.

1.2. Latent Transformer T

In the latent layer of AWRCP, we perform deep-level global modeling using the vanilla transformer block, including multi-head self-attention [3] and multi-scale feedforward network [4]. For the obtained feature $\mathcal{F}_E \in \mathbb{R}^{H\times W\times C}$, we reshape it to $\mathcal{F}_E \in \mathbb{R}^{N\times C}$ ($N = H \times W$) and adopt learnable $\mathcal{W}_{\mathcal{Q}}^{C\times C}$, $\mathcal{W}_{\mathcal{K}}^{C\times C}$ and $\mathcal{W}_{\mathcal{V}}^{C\times C}$ to project the $\mathcal{F}_E$ into $\mathcal{Q}$ ($\mathcal{F}_E\mathcal{W}_{\mathcal{Q}}$), $\mathcal{K}$ ($\mathcal{F}_E\mathcal{W}_{\mathcal{K}}$) and $\mathcal{V}$ ($\mathcal{F}_E\mathcal{W}_{\mathcal{V}}$). To improve the local extraction ability, we also add a depth-wise convolution with the kernel size of 3×3:

$$\textbf{Softmax}\left(\frac{\mathcal{Q}\mathcal{K}^{\mathrm{T}}}{\sqrt{C}}\right)\times\mathcal{V}+\textbf{DWConv}(\mathcal{F}_E), \qquad (1)$$

where H, W, and C represent the height, width, and number of dimensions. The Softmax operation is employed with these dimensions. To perform these operations, we apply multi-head self-attention using the approach described in reference [3].

To handle the feedforward stage, we use the multi-scale feedforward network introduced in reference [4]. This network helps us model complex and diverse degradations in

*Equal contributions.
†Yun Liu is the corresponding author.

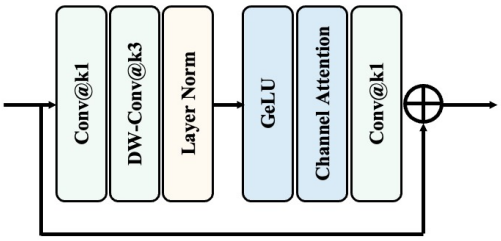


Figure 1. The overview of the proposed Efficient Encoder Block.

Figure 2. Ablation results on Codebook Priors

Table 1. Ablation studies on Transformer Block (§1.2). We conducted a performance comparison using the PSNR metric on the Outdoor-Rain [9] dataset to ensure a convinced comparison.

Setting	Model	PSNR	SSIM
i	w/o DWConv	36.81	0.965
ii	MLP [3]	36.66	0.964
iii	ConvFFN [10]	36.64	0.964
iv	LeFF [11]	36.79	0.965
v	Ours	36.91	0.966

rain scenes. Specifically, we express the transformer block as:

$$\begin{aligned} \mathcal{F}'_E &= \mathcal{F}_E + \textbf{MSA}(\textbf{LN}(\mathcal{F}_E)), \\ \hat{\mathcal{F}}_E &= \mathcal{F}'_E + \textbf{MFFN}(\textbf{LN}(\mathcal{F}'_E)), \end{aligned} \tag{2}$$

where $\hat{\mathcal{F}}_E$ refers to the feature processed by the global modeling via transformer block. **LN** is the LayerNorm, and the residual connection is employed followed by vanilla ViT [3]. In the latent part of AWRCP, we only employ 4 transformer block to balance model performance and complexity.

2. Additional Ablation Studies

In order to fully research the proposed AWRCP, we present a more exhaustive set of ablation studies.

2.1. Improvements of Transformer Block

Our study aims to evaluate how different components within the transformer block affect performance. We specifically investigate the effects of excluding the DWConv operation (w/o DWConv), using a traditional MLP-based feed-forward network [3] instead of the presented MFFN (MLP), employing the ConvFFN [10] instead of the proposed MFFN (ConvFFN), comparing the performance of LeFF [11] and MFFN (LeFF), and demonstrating the superiority of our approach using the MFFN (Ours). Our experimental results, presented in Table 1, demonstrate that the DWConv operation improves performance by complementing self-attention. Additionally, the MFFN has the most significant impact on the PSNR metric compared to other notable designs.

2.2. Effectiveness of Codebook Priors on Real-world Degradations

We analyze the effectiveness of codebook priors for the proposed AWRCP framework. In Fig. 2, we can observe that codebook priors can help our model produce clean results with less degradations.

3. Comparison with existing codebook-based related works

VQFR [5] is a codebook-based face restoration approach that deeply investigates the impact of compression patch size on both fidelity and restored quality. To address this issue, VQFR introduces a parallel decoder to balance both. In contrast, our work focuses on leveraging codebook priors to restore vivid texture details and accurate background structures. FeMaSR [1] utilizes pre-trained codebooks for blind SR and deals with potential matching errors through a Residual Shortcut Module. However, FeMaSR disregards the benefits of modulated high/low-quality features, which, as demonstrated in our work, are essential for adverse weather removal.

4. More Visual Comparisons

We conduct more visual comparisons with other state-of-the-art methods on synthetic and real datasets to demonstrate the excellent generalization performance of AWRCP on adverse weather removal.

4.1. Synthetic Datasets

AWRCP can effectively eliminate complex and challenging adverse weather degradations as demonstrated in Fig. 3, 4, 5, 6, 7, 8, 9, 10 and 11. Due to its robust high-quality codebook prior, AWRCP performs well adverse weather removal across multiple adverse weather degradations, including rain, haze, snow and raindrops. Additionally, the ability to restore realistic texture details is also noticeable compared with previous state-of-the-art methods.

4.2. Real-world Dataset

To demonstrate the impressive performance of our model in real-world scenarios, in Fig.12,13, 14, 15, 17 and 18, we

present extensive visual comparisons for real-world adverse weather removal. Our observations indicate that images processed by AWRCP exhibit improved quality and effectively remove complex degradations, while other algorithms often struggle with complex rain degradations. Moreover, AWRCP surpasses previous state-of-the-art methods in effectively handling fine-grained degradations.

5. Future Works, Limitations and Broader Impacts

Moving forward, we plan to explore the potential of codebook priors for various tasks while enhancing our overall architecture to make it more efficient and less computationally complex. Our AWRCP shows promising results in handling various challenging weather conditions. Nevertheless, the high complexity of our approach currently limit its deployment on edge devices and real-time processing applications.

AWRCP has demonstrated strong performance on real-world adverse weather scenarios, indicating potential applications in various industrial tasks and applications, such as surveillance video, autonomous driving and computational photography. Consequently, our work has the potential to contribute positively to both academia and industry.

References

[1] Chaofeng Chen, Xinyu Shi, Yipeng Qin, Xiaoming Li, Xiaoguang Han, Tao Yang, and Shihui Guo. Real-world blind super-resolution via feature matching with implicit high-resolution priors. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1329–1338, 2022.

[2] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. *arXiv preprint arXiv:2204.04676*, 2022.

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[4] Chen et al. Dual-former: Hybrid self-attention transformer for efficient image restoration. *arXiv*, 2022.

[5] Yuchao Gu, Xintao Wang, Liangbin Xie, Chao Dong, Gen Li, Ying Shan, and Ming-Ming Cheng. Vqfr: Blind face restoration with vector-quantized dictionary and parallel decoder. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVIII*, pages 126–143. Springer, 2022.

[6] 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.

[7] Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3175–3185, 2020.

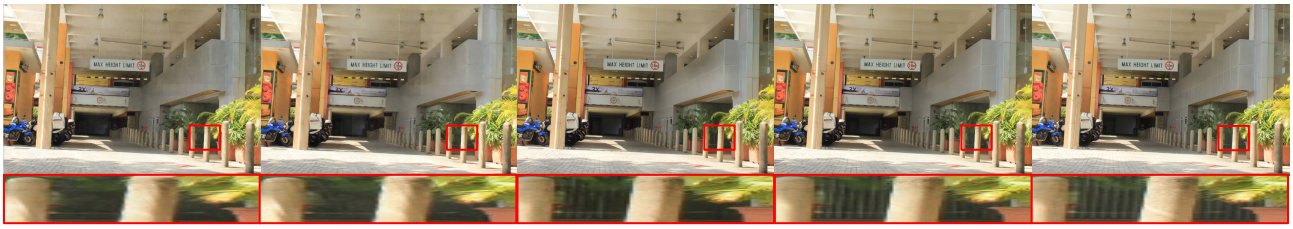
[8] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27(6):3064–3073, 2018.

[9] 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.

[10] Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 568–578, 2021.

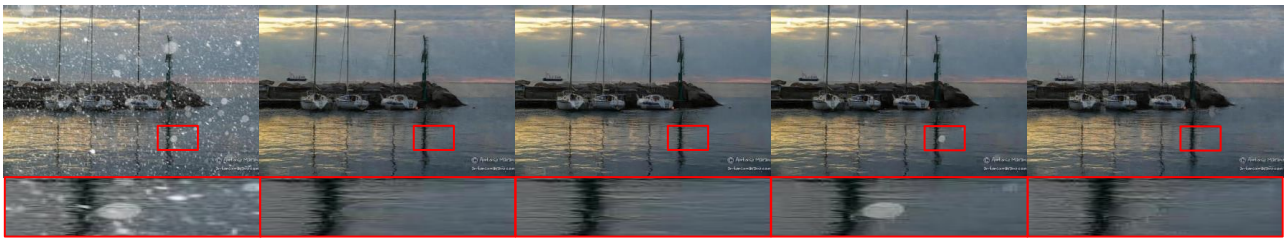
[11] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17683–17693, 2022.

(a) input (b) All-in-One (c) TKL&MR (d) MPRNet (e) TransWeather

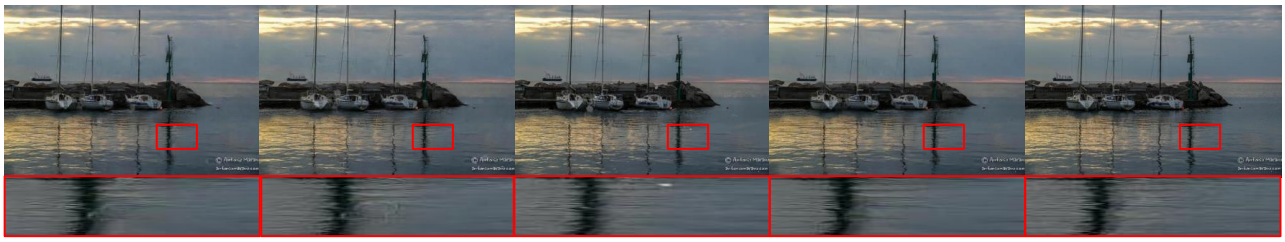


(f) Restormer (g) NAFNet (h) WeatherDiffusion (i) Ours (j) GT

Figure 3. Visual comparisons of synthetic rain and fog image from the Outdoor-Rain [7] testing dataset. The image can be zoomed in for improved visualization.



(a) input (b) All-in-One (c) TKL&MR (d) MPRNet (e) TransWeather



(f) Restormer (g) NAFNet (h) WeatherDiffusion (i) Ours (j) GT

Figure 4. Visual comparisons on a typical degraded image with serious snowy degradations.The proposed method outperforms other state-of-the-art approaches for removing adverse weather effects by producing cleaner results with realistic texture details specifically for lake ripples.

(a) input (b) All-in-One (c) TKL&MR (d) MPRNet (e) TransWeather

(f) Restormer (g) NAFNet (h) WeatherDiffusion (i) Ours (j) GT

Figure 5. Visual comparisons on a typical degraded image with dense fog and heavy rain.The proposed method generates cleaner results with better structural quality than other state-of-the-art adverse weather removal approaches.

(a) input (b) All-in-One (c) TKL&MR (d) MPRNet (e) TransWeather

(f) Restormer (g) NAFNet (h) WeatherDiffusion (i) Ours (j) GT

Figure 6. Visual comparisons of synthetic rain and fog image from the Outdoor-Rain [7] testing dataset. The image can be zoomed in for improved visualization.

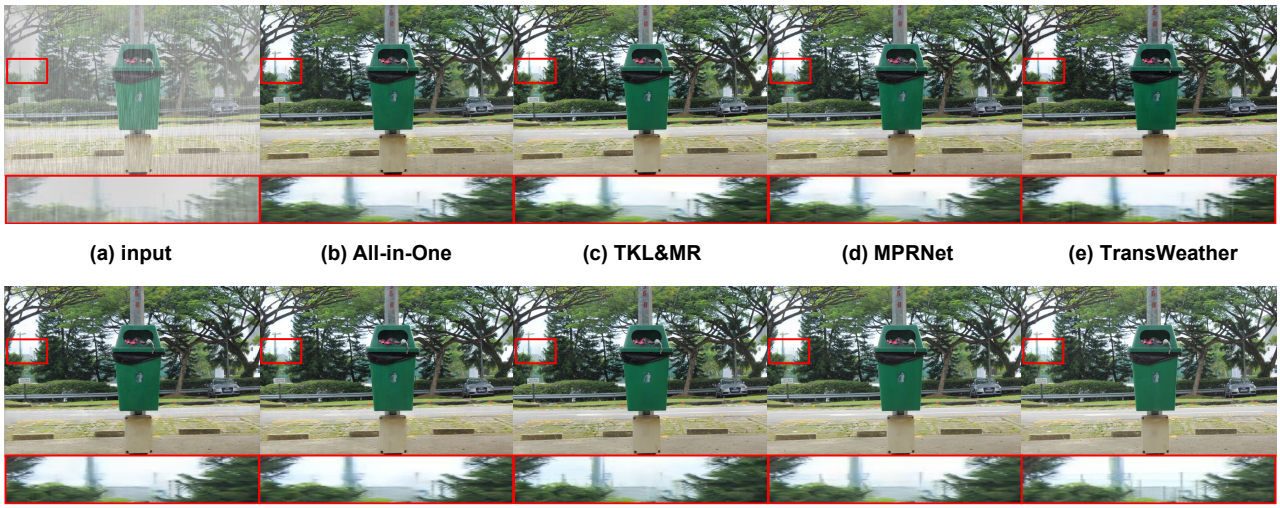


Figure 7. Visual comparisons of the synthetic rain and fog image from the Outdoor-Rain [7] testing dataset. The image can be zoomed in for improved visualization.

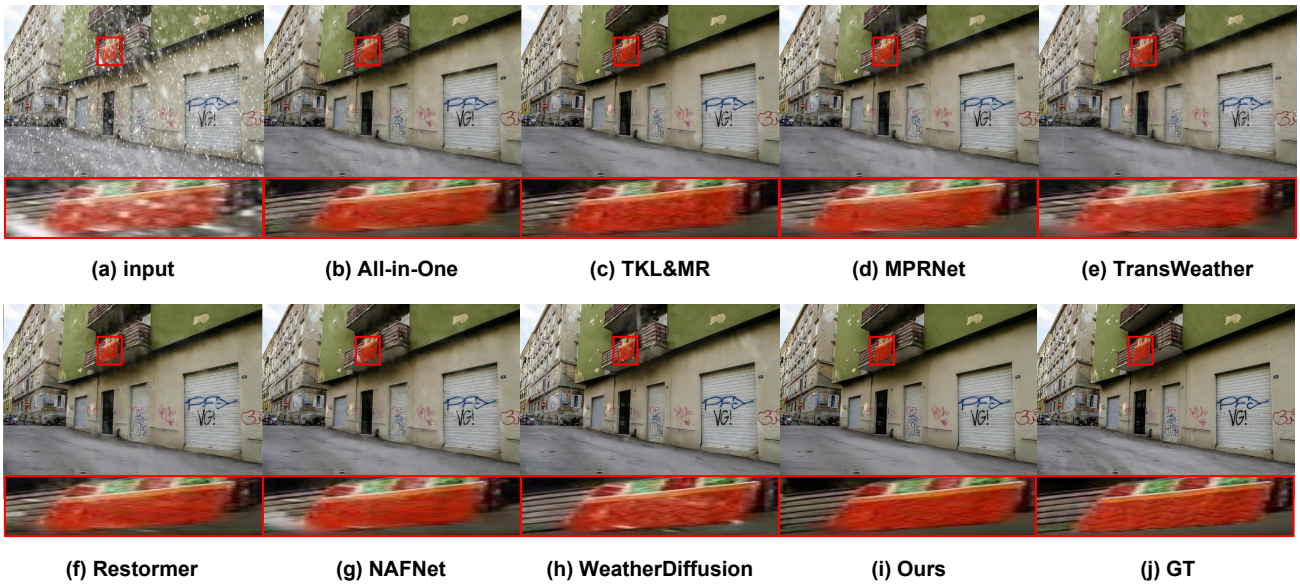


Figure 8. Visual comparisons of the synthetic snow image from the Snow100k [8] testing dataset. The image can be zoomed in for improved visualization.

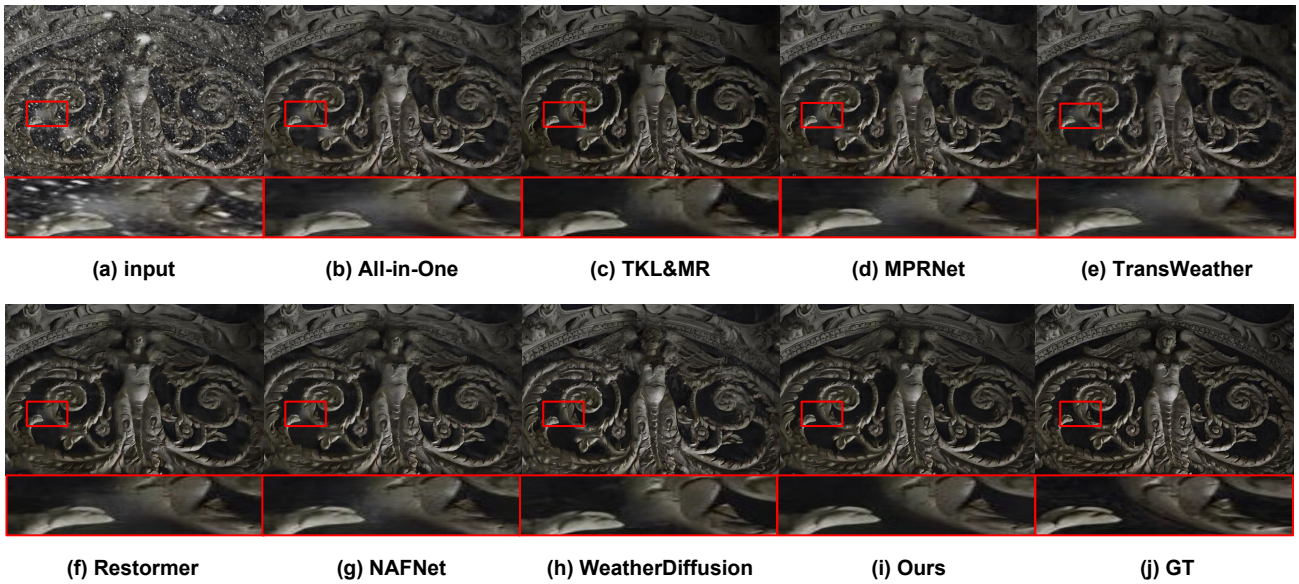


Figure 9. Visual comparisons of the synthetic snow image from the Snow100k [8] testing dataset. The image can be zoomed in for improved visualization.

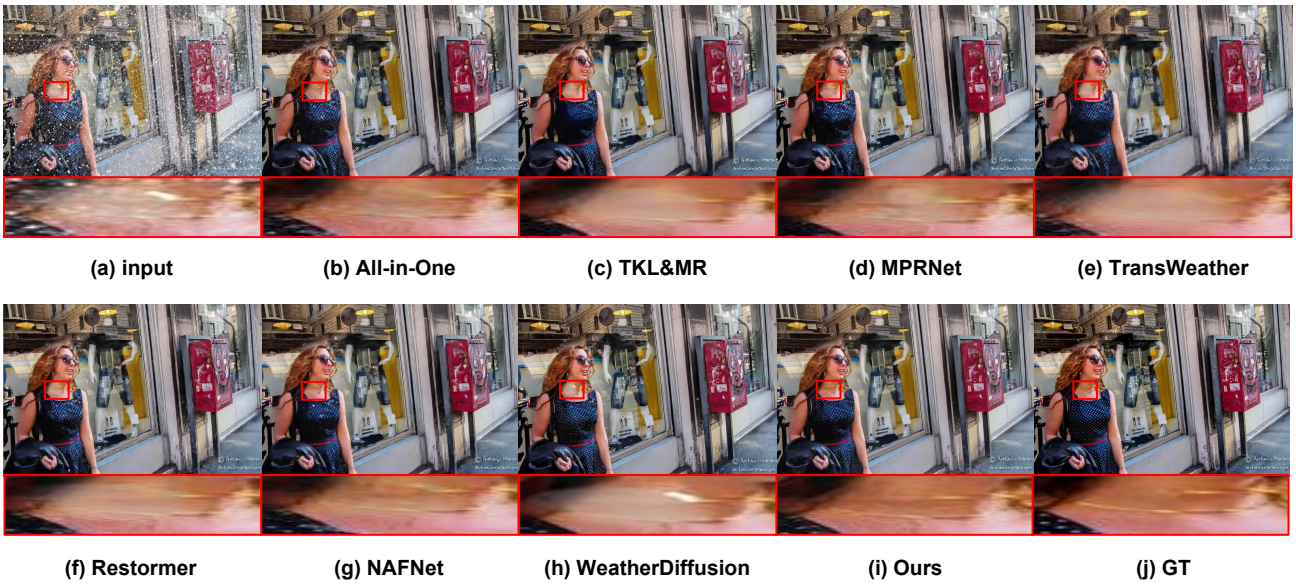


Figure 10. Visual comparisons of the synthetic snow image from the Snow100k [8] testing dataset. The image can be zoomed in for improved visualization.

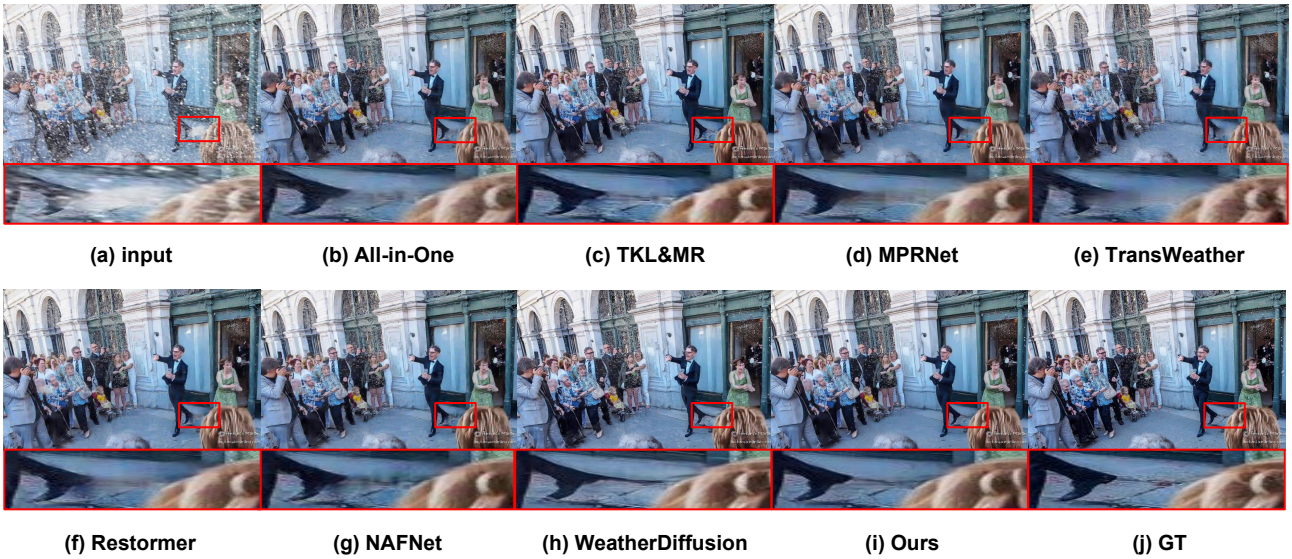


Figure 11. Visual comparisons of synthetic snow image from the Snow100k [8] testing dataset. The image can be zoomed in for improved visualization.

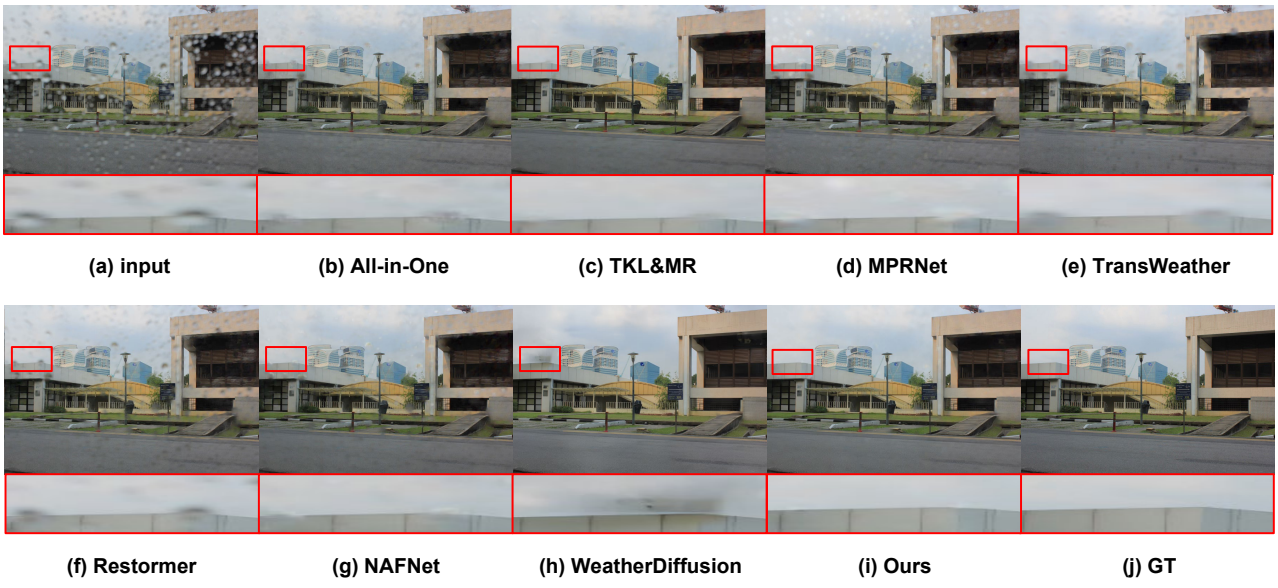


Figure 12. Visual comparisons of the real-world raindrop image from the RainDS [9] testing dataset. The image can be zoomed in for improved visualization.

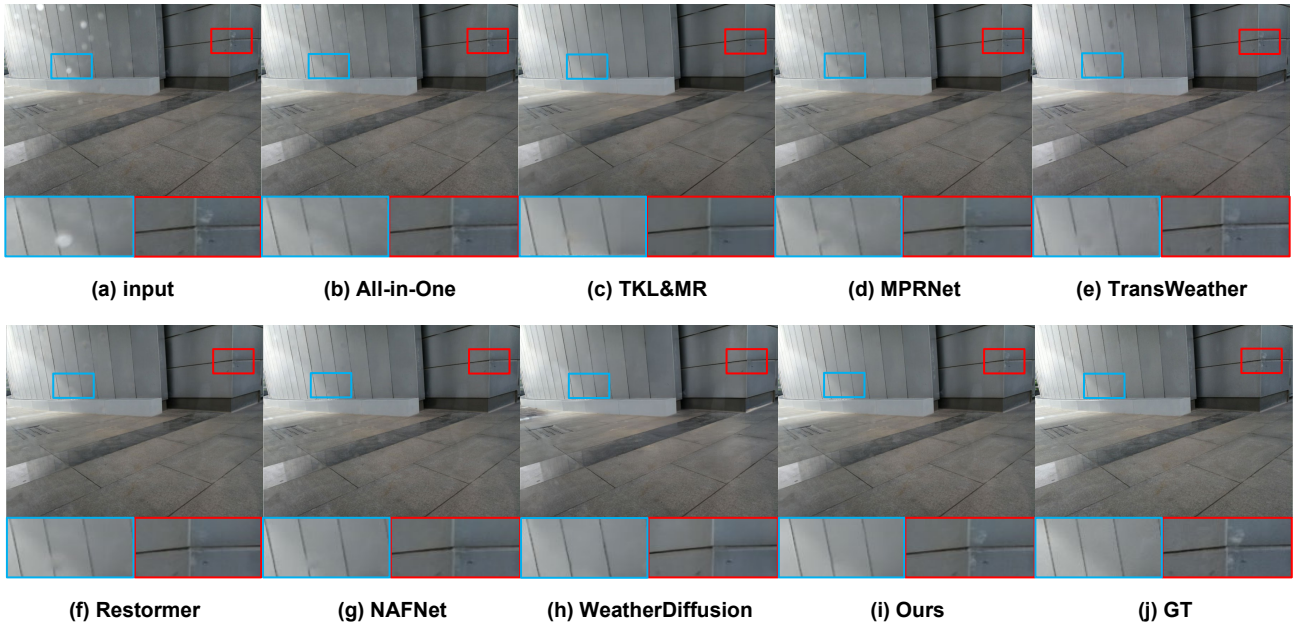


Figure 13. Visual comparisons of the real-world raindrop image from the RainDS [9] testing dataset. The image can be zoomed in for improved visualization.

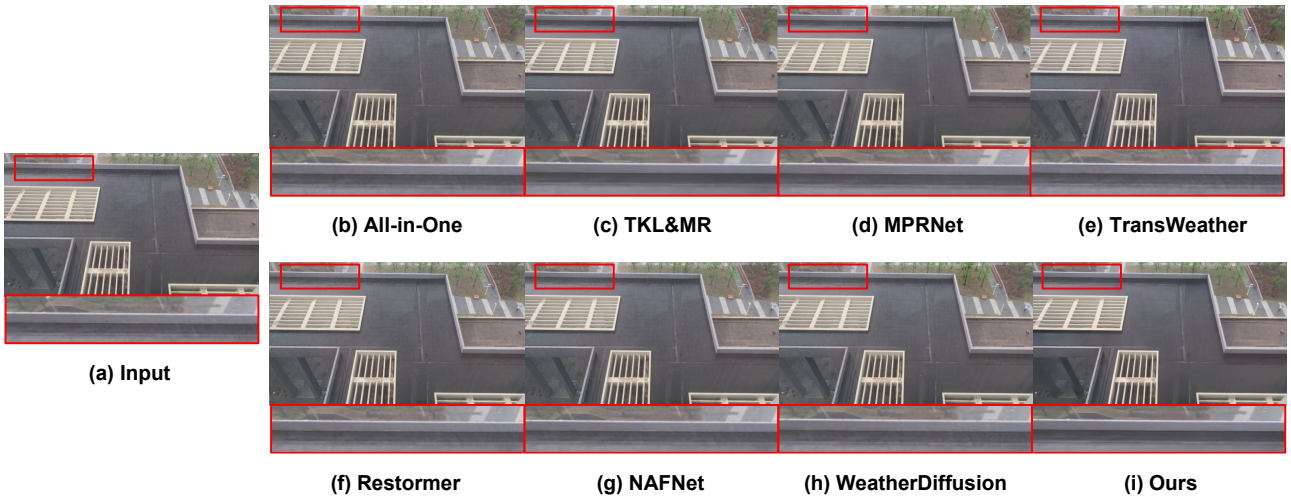


Figure 14. Visual comparisons of the real-world rainy image from Internet. The image can be zoomed in for improved visualization.

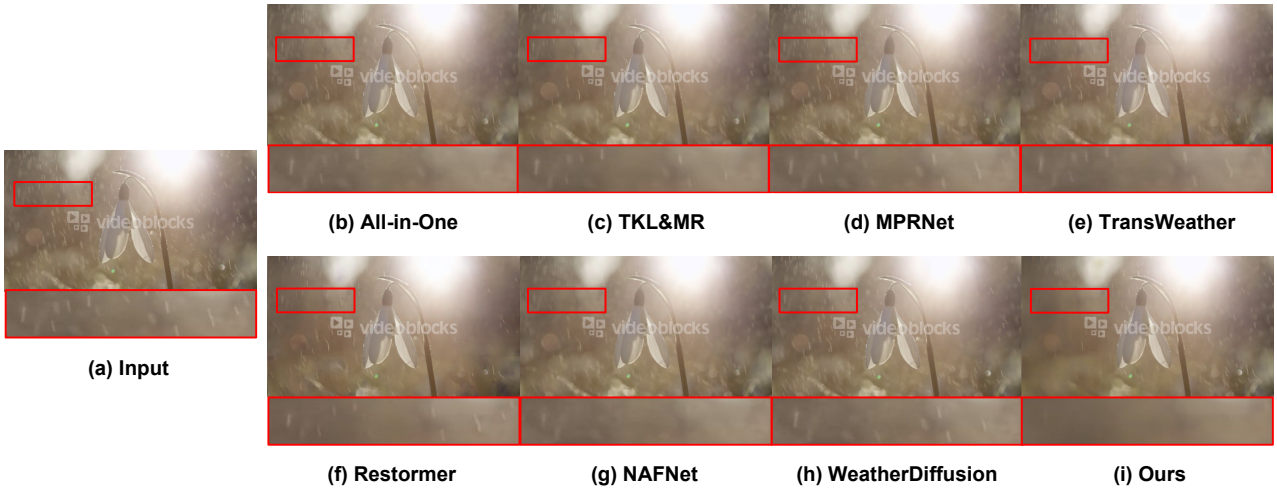


Figure 15. Visual comparisons of the real-world rainy image from Internet. The image can be zoomed in for improved visualization.

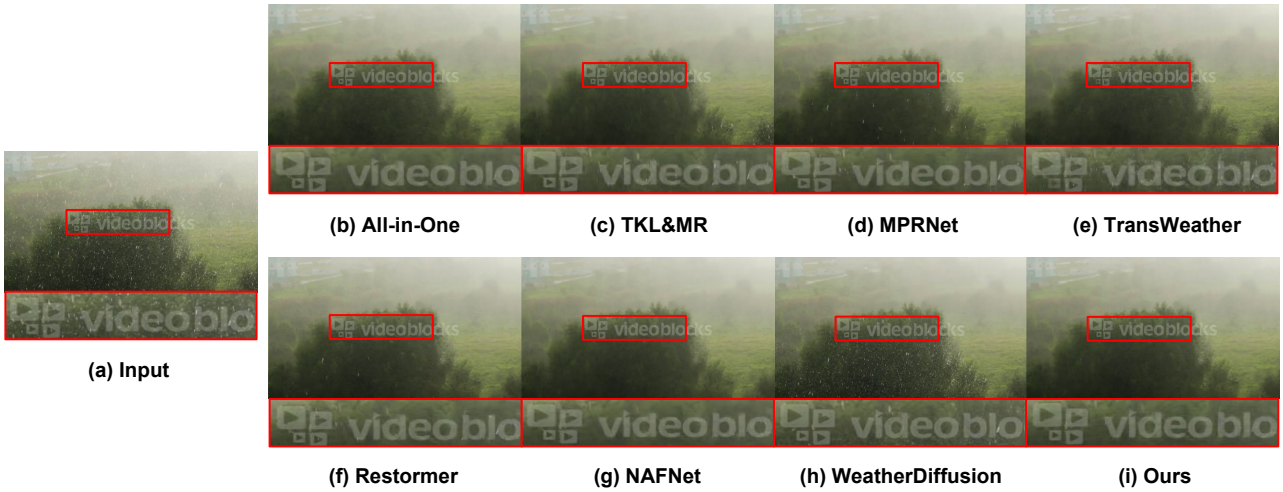


Figure 16. Visual comparisons of the real-world rainy image from Internet. The image can be zoomed in for improved visualization.

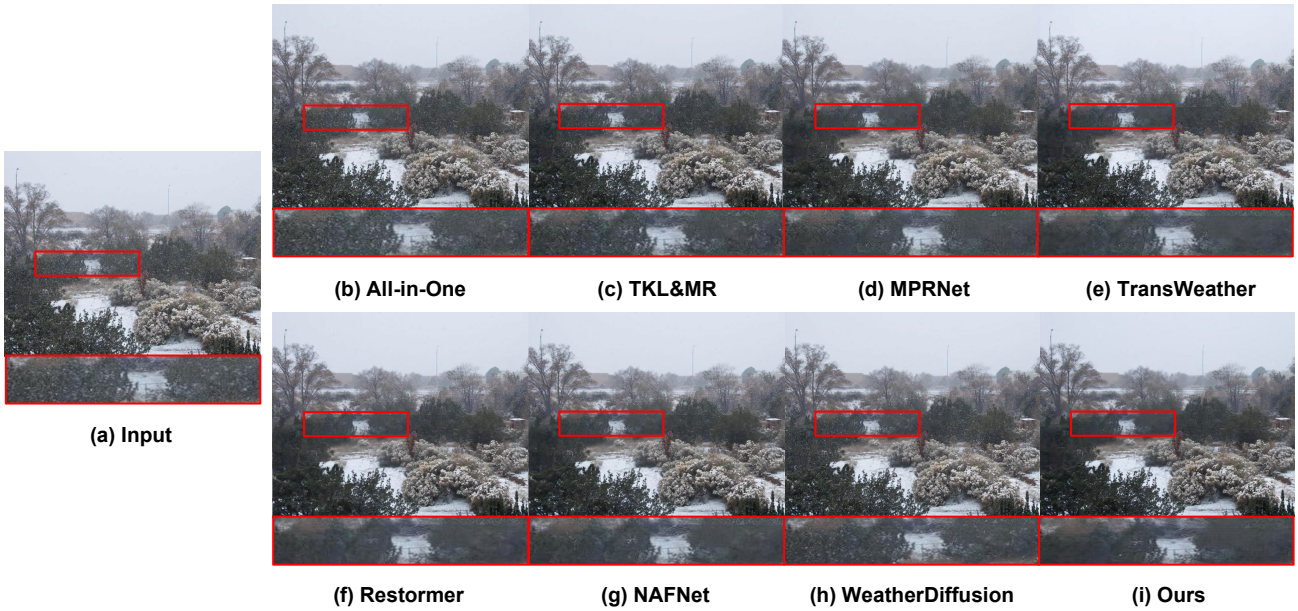


Figure 17. Visual comparisons of the real-world snow image from Internet. The image can be zoomed in for improved visualization.

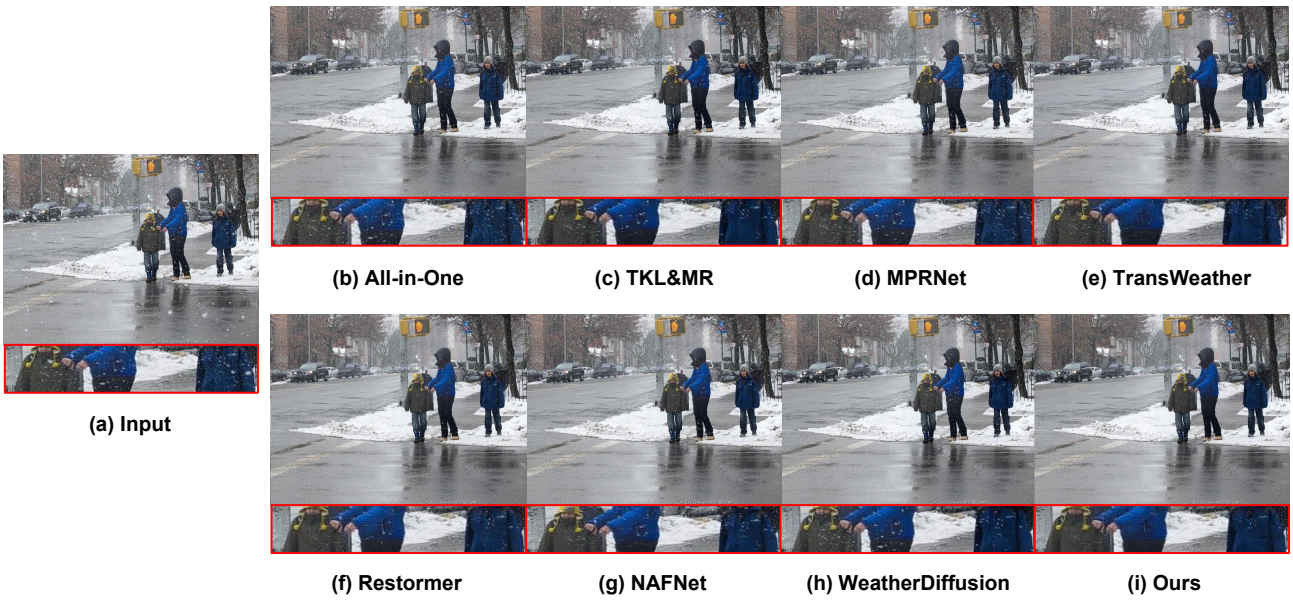


Figure 18. Visual comparisons of the real-world snow image from Internet. The image can be zoomed in for improved visualization.