Teaching Tailored to Talent: Adverse Weather Restoration via Prompt Pool and Depth-Anything Constraint

Sixiang Chen[1], Tian Ye[1], Kai Zhang[1],
Zhaohu Xing[1], Yunlong Lin[2], and Lei Zhu[1,3]✉

[1] The Hong Kong University of Science and Technology (Guangzhou), China
[2] Xiamen University, China
[3] The Hong Kong University of Science and Technology, Hong Kong SAR, China
leizhu@ust.hk
Project page: https://ephemeral182.github.io/T3-DiffWeather

This is the supplementary material for Teaching Tailored to Talent: Adverse Weather Restoration via Prompt Pool and Depth-Anything Constraint.

The provided materials are listed as follows:

- Sec.1 Preminliaries of diffusion models.
- Sec.2 Detailed information about dataset configurations.
- Sec.3 Architecture of the T^3-DiffWeather pipeline.
- Sec.4 Additional ablation studies.
- Sec.5 Performance on more datasets.
- Sec.6 Performance on real-world datasets.
- Sec.7 More visual comparisons.
- Sec.8 Limitation discussion.
- Sec.9 Broad impact.
- Sec.10 Future works.

1 Preliminaries

Diffusion Models. Diffusion models (DMs) [5, 10] infuse training data with Gaussian noise and then recover the original data through the inversion of this noise. Initially, DMs implement a diffusion algorithm that incrementally converts a starting image $\boldsymbol{x}_0$ into a noise distribution $\boldsymbol{x}_T \sim \mathcal{N}(0,1)$ across T steps. Each step of this transformation is described by the equation:

$$q\left(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}\right)=\mathcal{N}\left(\boldsymbol{x}_t;\sqrt{1-\beta_t}\boldsymbol{x}_{t-1},\beta_t\boldsymbol{I}\right), \tag{1}$$

with $\boldsymbol{x}_t$ representing the image with noise at timestep t, β_t as a predetermined scale parameter, and $\mathcal{N}$ signifying the Gaussian distribution. Defining $\alpha_t = 1-\beta_t$ and $\bar{\alpha}_t = \prod_{i=0}^{t}\alpha_i$ allows us to simplify Eq.1 as:

$$q\left(\boldsymbol{x}_t|\boldsymbol{x}_0\right)=\mathcal{N}\left(\boldsymbol{x}_t;\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0,(1-\bar{\alpha}_t)\boldsymbol{I}\right). \tag{2}$$

During the inference phase, DMs commence by generating a Gaussian noise map $\boldsymbol{x}_T$ and then progressively apply a denoising process until reaching a high-fidelity result $\boldsymbol{x}_0$:

$$p\left(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0\right)=\mathcal{N}\left(\boldsymbol{x}_{t-1};\boldsymbol{\mu}_t(\boldsymbol{x}_t,\boldsymbol{x}_0),\sigma_t^2\boldsymbol{I}\right), \tag{3}$$

where the mean value $\boldsymbol{\mu}_t(\boldsymbol{x}_t,\boldsymbol{x}_0)=\frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t-\boldsymbol{\epsilon}\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\right)$ and the variance $\sigma_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$. The noise estimate $\boldsymbol{\epsilon}$ is optimized $\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t,t)$ through training as follows: with a clean image $\boldsymbol{x}_0$, DMs select a random timestep t and noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0,\boldsymbol{I})$ to produce the noisy images $\boldsymbol{x}_t$ as per Eq.2. DMs then optimize the model parameters θ of $\boldsymbol{\epsilon}_\theta$ in accordance with [5]:

$$\mathcal{L}_{diff}=\mathbb{E}\left\|\boldsymbol{\epsilon}-\boldsymbol{\epsilon}_{\boldsymbol{\theta}}\left(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0+\boldsymbol{\epsilon}\sqrt{1-\bar{\alpha}_t},t\right)\right\|_2^2. \tag{4}$$

2 Dataset Configurations

Our experiments incorporate three primary benchmark datasets, including various snow conditions, dense rain with haze, real and simulated raindrops and streaks.

Snow100K: The Snow100K [8] dataset is specifically created to assess the performance of image snow removal algorithms. This benchmark is structured into two primary sections: a training set comprising 50,000 images, and a testing set. The testing set is divided into Snow100K-S, Snow100K-M, and Snow100K-L, containing 16,611, 16,588, and 16,801 images, respectively. These subsets are classified based on the intensity of synthetic snow coverage, categorized by the size of the snowflakes as light, medium, and heavy. Furthermore, Snow100K encompasses an additional set of 1,329 real-world snowy images, labeled as Snow100K-Real, to evaluate the effectiveness in adapting to realistic snow scenarios.

Outdoor-Rain: The Outdoor-Rain dataset [6] serves as an integral benchmark for addressing the complex challenge of removing both rain and haze from images, which presents dense rain patterns and integrates a realistic atmospheric perspective. The training set of Outdoor-Rain comprises 9,000 images. For quantitative analysis, we utilize a subset of 750 high-definition images, as Test1 in [6],

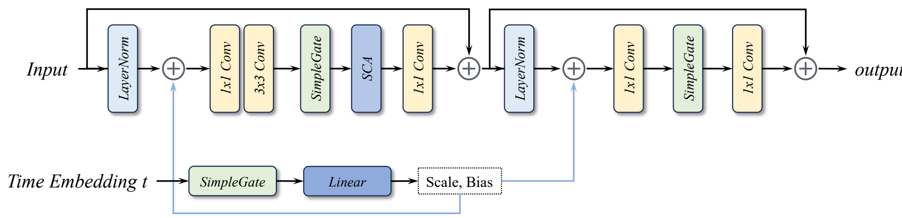

Fig. 1: Structure of the basic block.

which tests the effectiveness of image restoration algorithms under conditions of rain and dense haze.

RainDrop: The RainDrop [12] contains a set of images that simulate the visual obstruction and artifacts caused by raindrops on the camera lens in real scenes. The set contains 861 training images and a test subset of 58 images. This subset is called RainDrop-A.

3 Architecture Implementation

For the diffusion process architecture, we simply adopt the framework from the prior work Refusion [9] as our backbone (refer to Fig.1). To emphasize the superiority of our paradigm and the efficacy of our richly informative conditions, we have significantly reduced the parameter amount from the original 131M [9] to 69M. Specifically, we set the channel width to 64 and 18 blocks for the third stage of the encoder, all blocks across other stages are set to 1. Compared to the latest WeatherDiffusion [11], our intermediate architecture achieves remarkably improved performance in our experiments, while using almost 50M fewer parameters. This not only showcases the excellence of our pipeline but also the ample capability of our conditions.

4 Additional Ablation Studies

To rigorously evaluate the performance and effectiveness of our proposed T^3-DiffWeather framework, we have further carried out a series of comprehensive ablation studies. The more details of these studies are elaborated in the subsequent sections.

Discussion of prompt pool size and topk. Our results show that as the prompt pool size increases, the model has sufficient diversity in prompts to effectively capture and represent various degraded properties. Conversely, a tip pool that is too large may introduce redundancy or noise, thereby compromising the model's resilience. Additionally, for top-k, too fine-grained selection may lead to overfitting of poorly represented degradation attributes or ignoring valuable global information. Smaller top-k values ensure that the most relevant tips are employed, resulting in more robust and general performance across various degraded images.

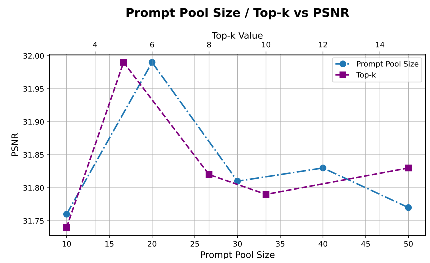

Fig. 2: Abl. of Prompt pool size and k value of top-k.

Gains of different Depth-Anything Architecture. The ablation study shown in Tab.1 compares memory cost with image recovery quality of different Depth-Anything [16] architectures. We found that the ViT-S-14 model stands out for its low memory consumption of only 115.22 MB, while still achieving competitive PSNR and SSIM metrics for our final results. In comparison, ViT-B-14 and ViT-L-14 require significantly more memory with little corresponding gain in PSNR or SSIM.

Table 1: Com. of memory cost, PSNR, and SSIM across the different Depth-Anything [16] architectures.

Method	#Memory Cost	PSNR ↑	SSIM ↑
ViT-S-14 (Ours)	115.22 MB	31.99	0.9365
ViT-B-14	402.29 MB	31.96	0.9365
ViT-L-14	1314.38 MB	**32.06**	**0.9366**

Benefits of Contrast Prompt Loss (CPL). As shown in Tab.2, the CPL distinguishes designed prompts to guide the diffusion process. The table shows that the pull of explicit constraints for prompts and the push of two prompts based on different motivations play an important role in improving the learning effect. They improve the guidance performance of conditions on diffusion by promoting enhanced representation learning. More ablation experiments can be found in the supplementary material.

Table 2: Abl. of Contrastive Prompt Loss (CPL).

Method	PSNR ↑	SSIM ↑
w/o. CPL	31.71	0.9350
w/o. Negative γ	31.81	0.9359
w/o. Positive γ	31.77	0.9358
w. CPL (Ours)	**31.99**	**0.9365**

4.1 Effectiveness of Degradation Residual

In contrast to the state-of-the-art method WeatherDiffusion [11], which focuses on reconstructing clean images, we have adapted the training objective of our diffusion model to reconstruct degradation residual. Such a paradigm greatly reduces the difficulty of diffusion reconstruction, and under the guidance of sufficient conditions, it

Table 3: Ablation studies on degradation residual(§4.1).

Settings	PSNR↑	SSIM↑
w/o. degradation residual (clean image)	31.13	0.9317
degradation residual (Ours)	31.99	0.9365

can improve the quality of the image fidelity. As shown in Tab.3, our pipeline can greatly improve restoration performance without requiring many inference steps.

Table 4: Ablation studies for Loss Function(§4.2).

Settings	PSNR↑	SSIM↑
L2 (sampling)	31.79	0.9358
PSNR Loss (sampling)	31.99	0.9365

4.2 Gains for the Loss Function

We further explore loss function and present the results in Tab.4. It is noticeable that using PSNR loss [2] in the sampling process has better performance than L2 loss.

5 Performance on More Datasets

Table 5: Comparisons results on the datasets of setting in [18].

Datasets	Method	PSNR ↑	SSIM ↑
SPA+	TransWeather [15]	33.64	0.93
	Chen et al. [4]	37.32	0.97
	WGWS-Net [18]	38.94	0.98
	T^3-DiffWeather (Ours)	39.99	0.99
RealSnow	TransWeather [15]	29.16	0.82
	Chen et al. [4]	29.37	0.88
	WGWS-Net [18]	29.46	0.85
	T^3-DiffWeather (Ours)	29.81	0.90
REVIDE	TransWeather [15]	17.33	0.82
	Chen et al. [4]	20.10	0.85
	WGWS-Net [18]	20.44	0.87
	T^3-DiffWeather (Ours)	20.89	0.89

In our extended quantitative comparison, we compare a novel dataset configuration, the same as WGWS-Net [18]. As illustrated in Tab.5, our approach continues to demonstrate state-of-the-art performance. Notably, unlike the WGWS-Net paradigm which requires a two-stage training process, our method adopts a "teaching tailored to talent" pipeline that is more versatile and adaptable to weather degradations across various scenarios. This one-stage, adaptive framework ensures high-quality restoration without needing phase-specific training, providing a universally applicable solution for diverse weather conditions.

6 Evaluating Real-World Performance

To further demonstrate the superiority of our paradigm in the real world, we conduct a quantitative comparison of real-world datasets. Tab.**??** presents an in-depth evaluation of our T^3-DiffWeather approach on datasets captured under real-world meteorological conditions [1, 3, 13]. Our T^3-DiffWeather method

Table 6: Com. on more datasets.

Method	RainDS-Real(RDS)		GT-RAIN	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
WGWS-Net [18]	20.79	0.603	20.65	0.608
WeatherDiff$_{128}$ [11]	21.09	0.605	20.83	0.613
AWRCP [17]	21.31	0.607	20.97	0.615
Ours	**21.96**	**0.612**	**21.27**	**0.619**

achieves impressive scores, emphasizing its excellent adaptability to different environmental conditions. This shows that our "teaching tailored to talent" paradigm can allow the network to adaptively utilize the required attributes, which is effective for diverse real-life scenarios. It shows that our method is widely applicable to actual environments.

7 More Visual Comparison

We provide more visual comparisons to illustrate the superior visual performance of T^3-DiffWeather in adverse weather restoration. The comparison results are shown in the Fig.3-10. As shown in the figure, our method is closer to the reference images in various tasks on synthetic datasets, significantly surpassing previous methods in terms of overall image detail and color recovery. In addition, we have also shown more real-world samples. It can be seen that our overall paradigm of "teaching tailored to talent" is more capable of dealing with complicated scenarios in the real world and has good adaptability.

8 Limitation Discussion

Despite the outstanding performance of our pipeline, the diffusion architecture still faces the challenge of reducing parameter amounts compared to traditional regression models. Moreover, the pursuit of more efficient guidance methods for the diffusion process is essential to enhance the embedding of conditions within the framework. Although the cross-attention [14] approach we currently utilize has achieved notably beneficial performance, there is continuing potential for further advancements in this area.

9 Broad Impacts

Advances in adverse weather image restoration extend across various domains, offering lots of benefits:

1. Improved Visibility and Safety for Transportation: In transportation systems, especially for autonomous vehicles and advanced driver-assistance systems, restoring images from adverse weather conditions like haze, rain, or snow is crucial. Enhanced clarity in visual inputs leads to better decision-making and increased road safety.

2. Reinforced Scientific and Environmental Analysis: Adverse weather conditions can obscure critical details in environmental monitoring and climate research imagery. Effective restoration techniques can unveil obscured features, aiding in more accurate data collection and analysis for climate studies and resource management.

3. Facilitated Geographic and Outdoor Activities: For outdoor enthusiasts and professionals in geography or archaeology, adverse weather restoration can enhance the usability of images captured in suboptimal conditions, providing clearer documentation and analysis of natural sites or historical landmarks.

The application of sophisticated adverse weather restoration methods promises substantial positive societal impacts, particularly in enhancing safety and information accuracy. The technology also has commercial applications in fields that require high-fidelity visual information regardless of weather conditions.

10 Future Works

Moving forward, we aspire to further mitigate the issue of parameter amounts while refining the concepts presented in our paper. Our goal is to extend our approach to other all-in-one tasks, aiming to create a unified framework for image restoration. Additionally, we are committed to ensuring that the overall pipeline design remains free from any redundancy.

References

1. 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: European Conference on Computer Vision. pp. 723–740. Springer (2022)
2. Chen, L., Lu, X., Zhang, J., Chu, X., Chen, C.: Hinet: Half instance normalization network for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 182–192 (2021)
3. 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: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13106–13117 (2023)
4. 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: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17653–17662 (2022)
5. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
6. Li, R., Cheong, L.F., Tan, R.T.: Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)

7. Li, R., Cheong, L.F., Tan, R.T.: Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1633–1642 (2019)
8. Liu: Desnownet: Context-aware deep network for snow removal. IEEE TIP (2018)
9. Luo, Z., Gustafsson, F.K., Zhao, Z., Sjölund, J., Schön, T.B.: Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1680–1691 (2023)
10. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: International Conference on Machine Learning. pp. 8162–8171. PMLR (2021)
11. Özdenizci, O., Legenstein, R.: Restoring vision in adverse weather conditions with patch-based denoising diffusion models. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
12. Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J.: Attentive generative adversarial network for raindrop removal from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2482–2491 (2018)
13. Quan, R., Yu, X., Liang, Y., Yang, Y.: Removing raindrops and rain streaks in one go. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9147–9156 (2021)
14. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
15. Valanarasu, J.M.J., Yasarla, R., Patel, V.M.: Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2353–2363 (2022)
16. Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H.: Depth anything: Unleashing the power of large-scale unlabeled data. arXiv preprint arXiv:2401.10891 (2024)
17. Ye, T., Chen, S., Bai, J., Shi, J., Xue, C., Jiang, J., Yin, J., Chen, E., Liu, Y.: Adverse weather removal with codebook priors. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12653–12664 (2023)
18. 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: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21747–21758 (2023)

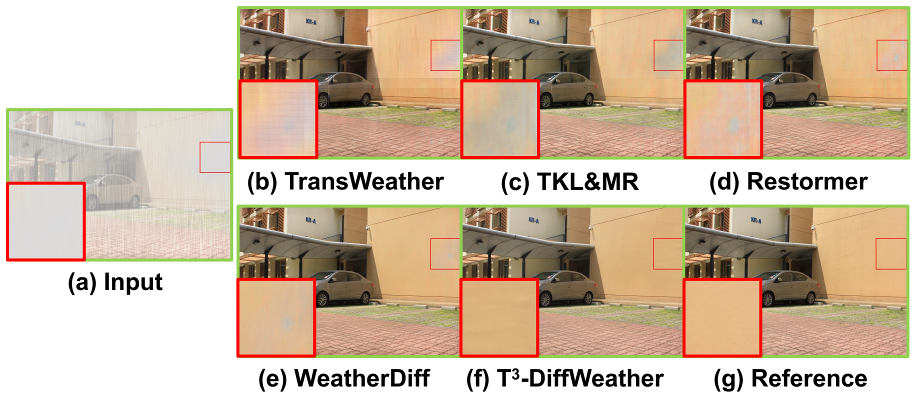

Fig. 3: Visual comparisons of rain & haze removal on Test1 dataset [7].

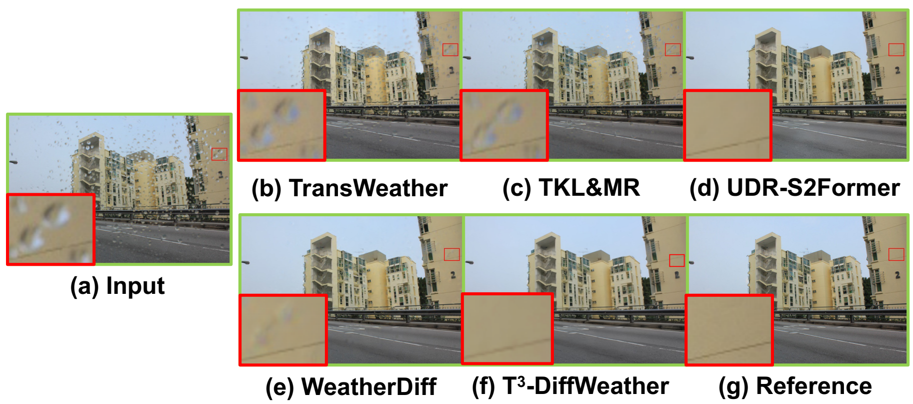

Fig. 4: Visual comparisons of raindrop removal on RainDrop dataset [12].

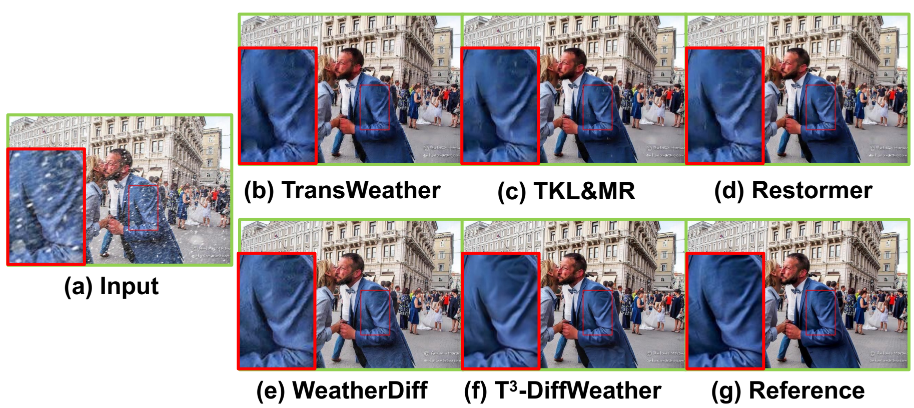

Fig. 5: Visual comparisons of snow removal on Snow100K dataset [8].

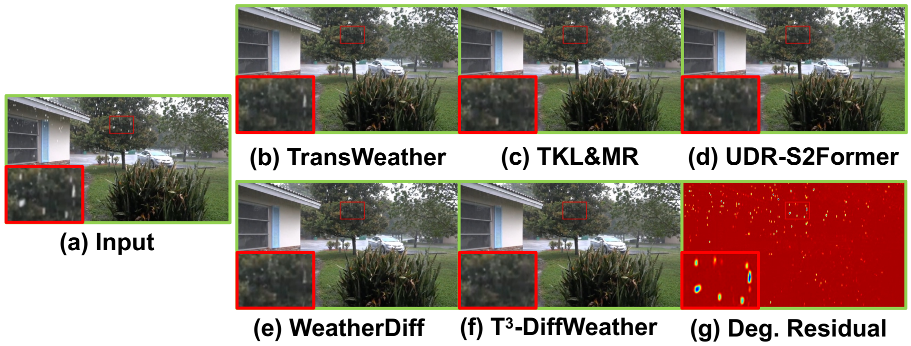

Fig. 6: Visual comparisons of rain removal on real-world sample.

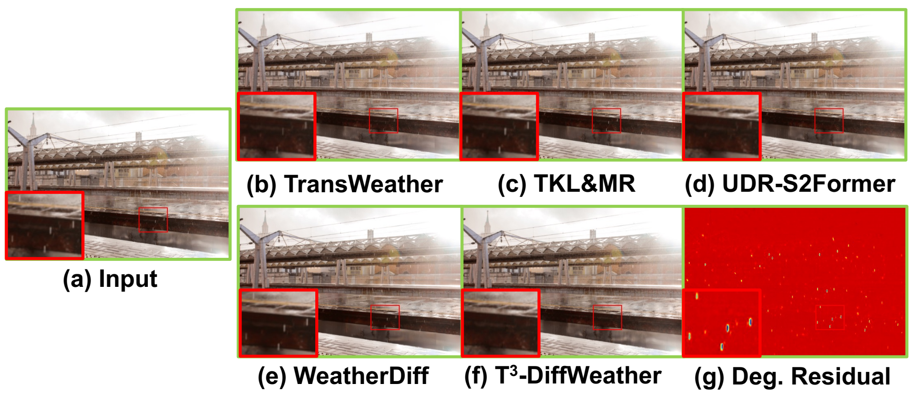

Fig. 7: Visual comparisons of rain removal on real-world sample.

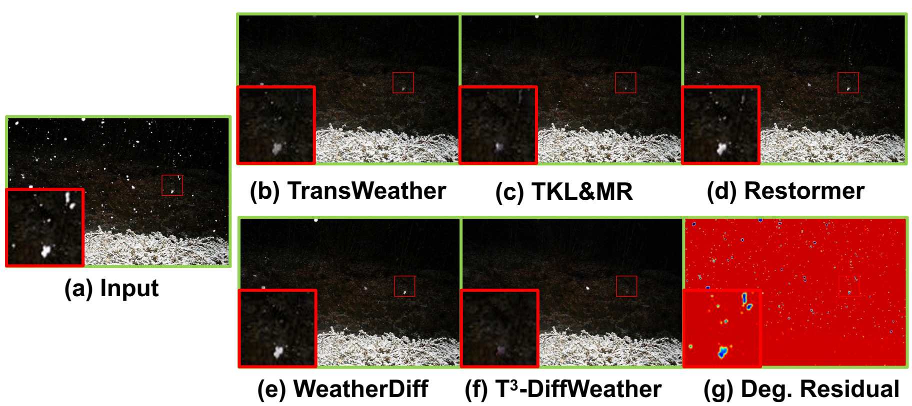

Fig. 8: Visual comparisons of snow removal on real-world sample.

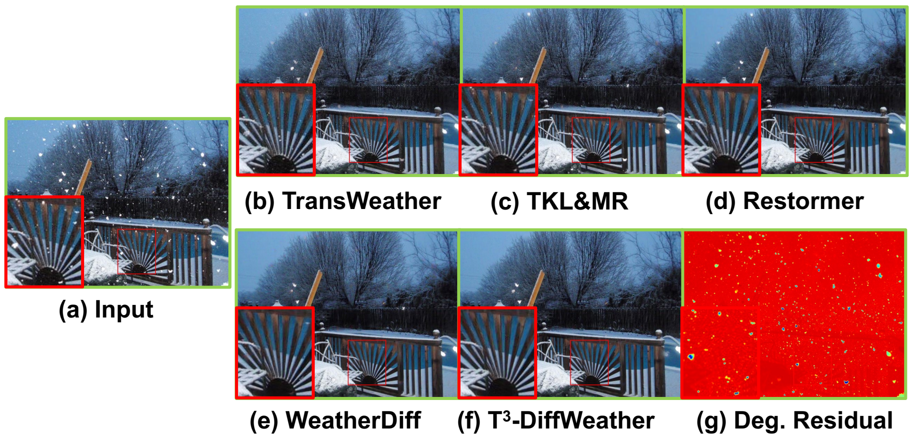

Fig. 9: Visual comparisons of snow removal on real-world sample.

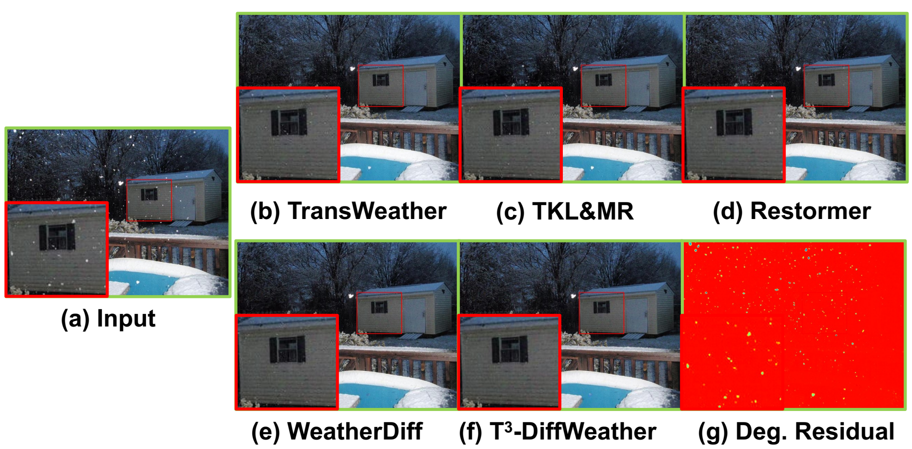

Fig. 10: Visual comparisons of snow removal on real-world sample.