Retrieve-to-Restore: Efficient All-in-One Image Restoration with a Retrieval-Based Degradation Bank

Supplementary Material

In the supplementary material, we first provide more implementation details of our work in Sec. A. Next, we provide further experiments and ablations in Sec. B. Lastly, we conclude by presenting additional visual comparisons across the all-in-one settings, including single type of degradation, five degradations and composited degradations, as detailed in Sec. C.

A. Further Implementation Details

Throughout all our experiments, we maintained a fixed random seed for reproducibility purposes. We based our implementation on the public PyTorch-based PromptIR [8] and AirNet [5] code for architecture development and training.

Training Time and Details. Training the proposed R2R in the all-in-one setting (dehazing, deraining, denoising) takes 40 hours on a single NVIDIA RTX 5090 GPU; for the five degradation setting, training extends to 48 hours. During inference, the Degradation Amalgamator is removed and restoration relies solely on priors retrieved from the Degradation Bank. In the composited degradations setting (Sec. B.1), a single model is trained to handle both isolated degradations and combinations of up to three types; all other settings follow the three degradation setup except for a batch size of 96, and training completes in 20 hours. Crucially, we replace the ***Argmax*** in Degradation Matching with ***Sigmoid***, enabling composited degradations to retrieve mixed priors.

Degradation Prior Construction and Retrieval. In the early stage of training, the Degradation Amalgamator is first optimized, while the task-level prior of each degradation type is stored to disk only after training reaches the midpoint. During inference, these stored task-level priors are loaded into the network for degradation retrieval and subsequent image restoration. Moreover, because training is performed on degraded inputs with a fixed spatial resolution, the saved task-level priors also have a fixed size. Therefore, for inference on images with arbitrary resolutions, D_K and D_V are first interpolated to the size required for retrieval. Once the corresponding prior is retrieved, it is resized back to the original resolution and then fed into the decoder.

Datasets. For composited degradation evaluation (Sec. B.1), we employ the CDD11 [3] dataset, comprising 11 distinct single and multi degradation combinations that cover diverse mixtures of low-light, haze, rain, and snow, thus providing a stringent benchmark for robustness. For further zero-shot evaluations in Sec. B.3, we use the following datasets: defocus deblurring (DPDD) [1] and motion deblurring (HIDE [9] and GoPro [6]), which contain real-world blur patterns under varying motion and exposure conditions, allowing us to assess the generalization ability of R2R.

Table 1. Comparison to language-guided methods. We evaluate our framework against InstructIR [2] and OneRestore [3] in the AIO-5 and Composited degradation settings, focusing on scenarios where these methods rely solely on visual inputs without additional language priors. Our results demonstrate that R2R performs favorably against these methods. 'LM' indicates the inclusion of language guidance. We report PSNR (dB, ↑) and SSIM (↑) metrics. For the composited degradation setting, we present the average metrics across single-level, double-level, and triple-level degraded inputs, as well as the overall average.

(a) *AIO-5.*

Method	LM	SOTS		Rain100L		BSD68		GoPro		LoLv1		Avg.	
InstructIR	-	25.20	.938	35.58	.967	31.09	.883	26.65	.810	20.70	.820	27.84	.884
	✓	27.10	.956	**36.84**	.973	**31.40**	.873	29.40	.886	**23.00**	.836	29.55	.908
R2R	-	**30.64**	**.974**	36.61	**.975**	31.35	**.891**	**30.93**	**.911**	22.88	**.856**	**30.48**	**.921**

(b) *Composited degradations.*

Method	LM	Single		Double		Triple		Overall	
OneRestore	-	31.68	.938	27.35	.866	24.84	.789	28.47	.878
	✓	31.81	.939	27.65	.871	25.23	.796	28.72	.882
MoCE-IR-S	-	32.50	.940	27.67	.870	25.20	.788	29.11	.880
R2R	-	**32.96**	**.945**	**28.00**	**.879**	**25.34**	**.804**	**29.32**	**.889**

B. Additional Experiments and Ablations

B.1. All-in-One: Composited Degradations

To simulate more realistic degradation scenarios, recent work [3] has trained all-in-one restorers not only on a mix of different degradations but also on more challenging cases where multiple degradation types are combined within a single image. We extend the previous all-in-one settings by including rain, haze, snow, and low illumination as individual degradation types, along with various composite degradation scenarios, resulting in a total of eleven distinct restoration settings. As shown in Table 2, our method achieves average improvements of 0.85dB (vs. OneRestore [3]) and 0.27dB (vs. MoCE-IR-S [10]) over previous all-in-one restorers, establishing a new state-of-the-art. This further validates the effectiveness of our adaptive prior retrieval conditioned on the degraded input.

Table 2. Comparison to state-of-the-art on composited degradations. PSNR (dB, ↑) and SSIM (↑) are reported on the full RGB images with (*) denoting general image restorers, others are specialized all-in-one approaches. Our R2R method consistently outperforms even larger models, achieving new state-of-the-art performance with particularly favorable results in composited degradation scenarios.

Method	MACs.	*CDD11-Single*								*CDD11-Double*										*CDD11-Triple*				Avg.	
		Low (L)		Haze (H)		Rain (R)		Snow (S)		L+H		L+R		L+S		H+R		H+S		L+H+R		L+H+S			
AirNet [5]	238G	24.83	.778	24.21	.951	26.55	.891	26.79	.919	23.23	.779	22.82	.710	23.29	.723	22.21	.868	23.29	.901	21.80	.708	22.24	.725	23.75	.814
PromptIR [8]	132G	26.32	.805	26.10	.969	31.56	.946	31.53	.960	24.49	.789	25.05	.771	24.51	.761	24.54	.924	23.70	.925	23.74	.752	23.33	.747	25.90	.850
WGWSNet [12]	-	24.39	.774	27.90	.982	33.15	.964	34.43	.973	24.27	.800	25.06	.772	24.60	.765	27.23	.955	27.65	.960	23.90	.772	23.97	.771	26.96	.863
WeatherDiff [7]	263G	23.58	.763	21.99	.904	24.85	.885	24.80	.888	21.83	.756	22.69	.730	22.12	.707	21.25	.868	21.99	.868	21.23	.716	21.04	.698	22.49	.799
OneRestore [3]	-	26.48	**.826**	32.52	.990	33.40	.964	34.31	.973	25.79	**.822**	25.58	.799	25.19	.789	**29.99**	.957	**30.21**	.964	24.78	.788	24.90	**.791**	28.47	.878
MoCE-IR-S [10]	37G	**27.26**	.824	**32.66**	**.990**	**34.31**	**.970**	**35.91**	**.980**	**26.24**	.817	**26.25**	**.800**	**26.04**	**.793**	29.93	**.964**	30.19	**.970**	**25.41**	**.789**	**25.39**	.790	**29.05**	**.881**
R2R (*ours*)	**12G**	**27.29**	**.836**	**32.88**	**.992**	**35.12**	**.971**	**36.54**	**.981**	**26.21**	**.833**	**26.70**	**.816**	**26.33**	**.809**	**30.41**	**.966**	**30.37**	**.971**	**25.47**	**.804**	**25.21**	**.803**	**29.32**	**.889**

B.2. Comparison to Language-guided Methods.

Language prompts can also act as priors to boost restoration performance. For example, as shown in Table 1a, InstructIR-3D [2] gains 1.71dB over its non-language baseline, and in the composited-degradation setting OneRestore [3] improves by 0.25dB (Table 1b). However, such language guidance typically incurs additional computation and requires extra text encoders or large-scale language models. By contrast, our Degradation Bank priors deliver similar performance benefits using only image domain information, without any textual side information or auxiliary networks. This design enables R2R to enjoy the advantages of prior based adaptation while retaining a lightweight and efficient architecture.

B.3. Zero-Shot Generalization

Following [10], We evaluate the zero-shot generalization of R2R, AirNet [5], PromptIR [8] and MoCE-IR [10], all trained under the three-degradation setting, by testing their performance on unseen degradations using PSNR, SSIM, and LPIPS metrics. Specifically, we assess various deblurring tasks, including defocus deblurring on DPDD [1] and motion deblurring on GoPro [6] and HIDE [9], with none of the models trained on these degradations. For fair comparison, we use the official model checkpoints for AirNet, PromptIR and MoCE-IR without retraining. As shown in Table 3, R2R outperforms these methods, achieving higher PSNR and SSIM scores and consistently lower LPIPS values across all datasets.

B.4. Visual Analysis of M

We visualize softmax weights between 500 SOTS [4] images and the haze bank (Fig. 1). With M=24, responses are overly uniform and show poor separability; with M=32, several channels stay low, indicating unused priors; with M=128, weights are small and similar, suggesting redundancy. M=64 best balances prior diversity and utilization.

C. Visual Results

Visual Comparison on Single Degradation. We present additional single degradation visual comparisons against PromptIR [8], InstructIR [2] and VLU-Net [11]. In noisy scenes (Fig. 2), our method produces sharper, more detailed denoised results. In foggy conditions (Fig. 3), our approach delivers competitive restoration quality with plausible scene structure and color. Under rainy conditions (Fig. 4), competing methods leave visible rain streaks, whereas our method yields noticeably cleaner outputs. Most strikingly, in blurred scenes (Fig. 5), both InstructIR and VLU-Net struggle to recover fine details and leave noticeable residual blur, whereas our approach produces substantially sharper reconstructions with fewer artifacts. Lowlight scenes (Fig. 6) show a similar advantage. Taken together with the quantitative evaluations, these visual results further demonstrate the effectiveness of our method.

Table 3. Zero-shot generalization. We present results for AirNet [5], PromptIR [8], MoCE-IR [10] and R2R compared for zero-shot generalization to degradations unseen during training. We report PSNR (dB, ↑), SSIM (↑) and LPIPS (↓) on the RGB images.

Method	MACs.	DPDD			GoPro			HIDE		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
AirNet [5]	238G	20.17	.674	.376	21.95	.748	.330	20.68	.731	.350
PromptIR [8]	132G	21.76	.673	.386	22.15	.749	.332	22.78	.740	.333
MoCE-IR [10]	81G	21.83	.678	.372	22.50	.758	.326	22.92	.743	.334
R2R	**12G**	**22.07**	**.678**	**.370**	**25.59**	**.778**	**.316**	**23.89**	**.747**	**.333**

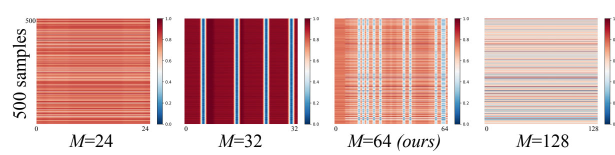

Figure 1. Softmax weight heatmaps for haze prior retrieval under different M.

Visual Comparison on Five Degradations. We present visual comparisons of R2R trained on five degradation types against MoCE-IR-S [10] and VLU-Net [11] in Fig. 7. Similar to the results observed in the single-degradation setting, our model consistently restores image details from hazy, rainy, and noisy inputs. Beyond this, R2R demonstrates strong performance in recovering clearer details from motion-blurred images and achieves more accurate color and detail restoration in low-illumination scenarios compared to both methods. Overall, these qualitative results indicate that R2R delivers visually competitive restorations across all five degradation types, in line with the quantitative performance reported in our five-degradation experiments.

Visual Comparison on Composited Degradations. In Fig. 8-10, we compare the visual quality of restored test samples against MoCE-IR-S [10] and the non-language-based version of OneRestore [3]. The comparisons cover a wide range of degradation scenarios: four single degradations (low-illumination, haze, rain, and snow), five double-composited degradations (low + haze, low + rain, low + snow, haze + rain, and haze + snow), and two triple-composited degradations (low + haze + rain and low + haze + snow).

To better expose reconstruction differences, we enlarge representative regions of each restored image, allowing closer inspection of fine details, residual artifacts, and color fidelity. Across single-degradation cases, our method tends to recover sharper structures and more faithful colors, particularly in challenging regions with dense haze, heavy rain, or severe darkness. In the composited degradation setting, visible remnants of haze and rain, as well as color inaccuracies induced by low illumination, are clearly observed in the baseline results, whereas our framework effectively suppresses these artifacts and produces cleaner, more visually plausible reconstructions.

Visual Comparison on Real-World Scenes. We further present several real-world image restoration examples, including low, haze, snow, and low+haze, to demonstrate the effectiveness of our method in practical scenarios. As shown in Fig. 11, we evaluate different methods using NIQE and PIQE, where lower values indicate better perceptual quality. Compared with MoCE-IR-S [10] and OneRestore [3], our method achieves better perceptual quality, produces more visually pleasing results, and maintains higher efficiency.

Failure Cases. Fig. 12 presents several failure cases of our method under extreme scenarios, including spatially non-uniform degradations and severely degraded conditions. Specifically, in the first row, we construct a spatially non-uniform degradation setting by applying four different degradation types(noise, haze, rain and snow) to four separate regions of a single image. In the second row, we show real-world extreme cases, such as scenes jointly affected by extremely low illumination and snow. The results indicate that both spatially non-uniform degradations and severe compound degradations under extreme weather conditions remain highly challenging for existing restoration methods, which still struggle to produce satisfactory results. In future work, we will further explore patch-based restoration strategies to better handle these complex degradation scenarios.

References

[1] Yuang Ai, Huaibo Huang, Xiaoqiang Zhou, Jiexiang Wang, and Ran He. Multimodal prompt perceiver: Empower adaptiveness generalizability and fidelity for all-in-one image restoration. In *CVPR*, pages 25432–25444, 2024. 1, 2

[2] Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *ECCV*, pages 1–21, 2024. 1, 2

[3] Yu Guo, Yuan Gao, Yuxu Lu, Huilin Zhu, Ryan Wen Liu, and Shengfeng He. Onerestore: A universal restoration framework for composite degradation. In *ECCV*, pages 255–272, 2024. 1, 2, 3, 10, 11, 12, 13

[4] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE TIP*, 28(1):492–505, 2018. 2

[5] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *CVPR*, pages 17452–17462, 2022. 1, 2

[6] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, pages 3883–3891, 2017. 1, 2

[7] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE TPAMI*, 45(8):10346–10357, 2023. 2

[8] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. In *NeurIPS*, pages 71275–71293, 2023. 1, 2

[9] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, pages 17683–17693, 2022. 1, 2

[10] Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yuedong Tan, Danda Pani Paudel, Yulun Zhang, and Radu Timofte. Complexity experts are task-discriminative learners for any image restoration. In *CVPR*, pages 12753–12763, 2025. 1, 2, 3, 9, 10, 11, 12, 13

[11] Haijin Zeng, Xiangming Wang, Yongyong Chen, Jingyong Su, and Jie Liu. Vision-language gradient descent-driven all-in-one deep unfolding networks. In *CVPR*, pages 7524–7533, 2025. 2, 9

[12] 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 *CVPR*, pages 21747–21758, 2023. 2

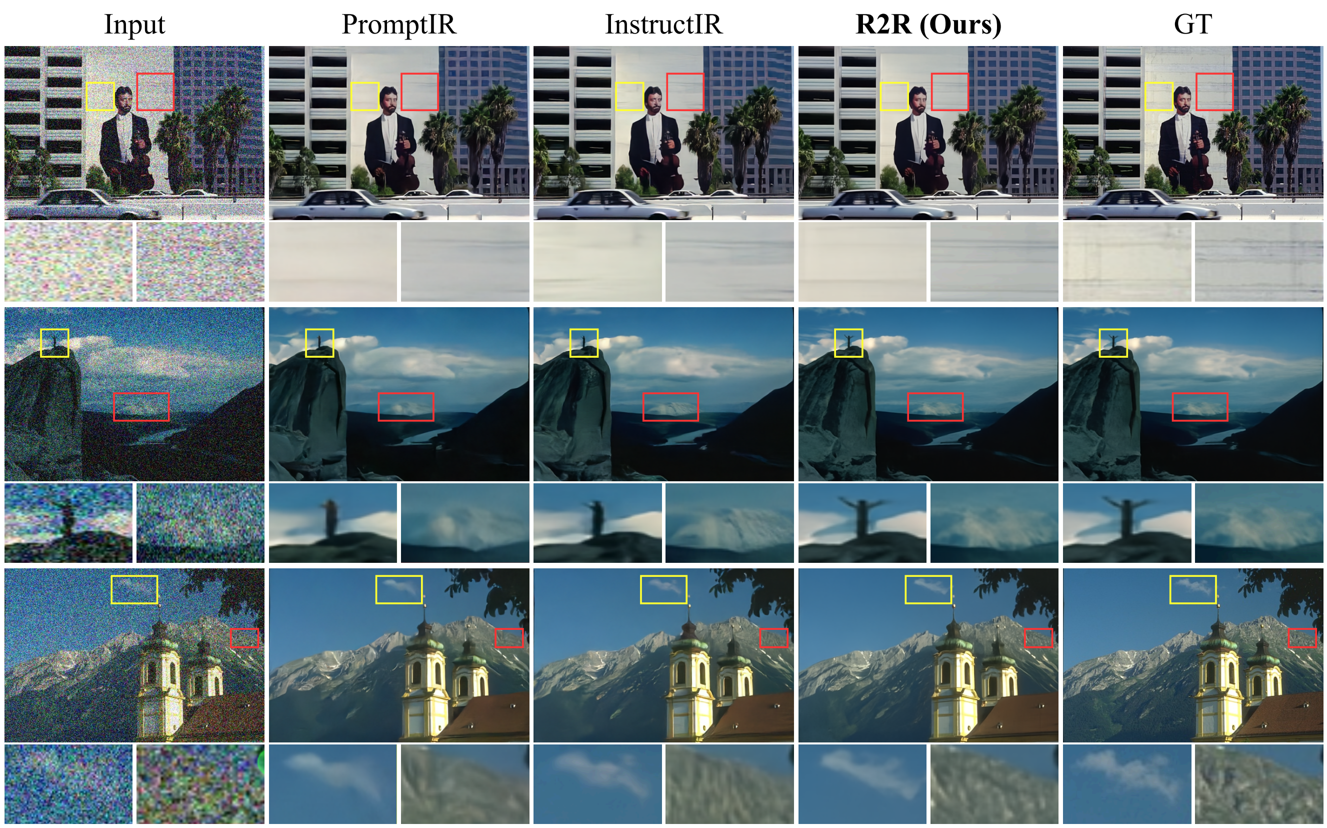

Figure 2. Visual comparison of R2R with state-of-the-art methods on the single noise degradation task. Zoom in for a better view.

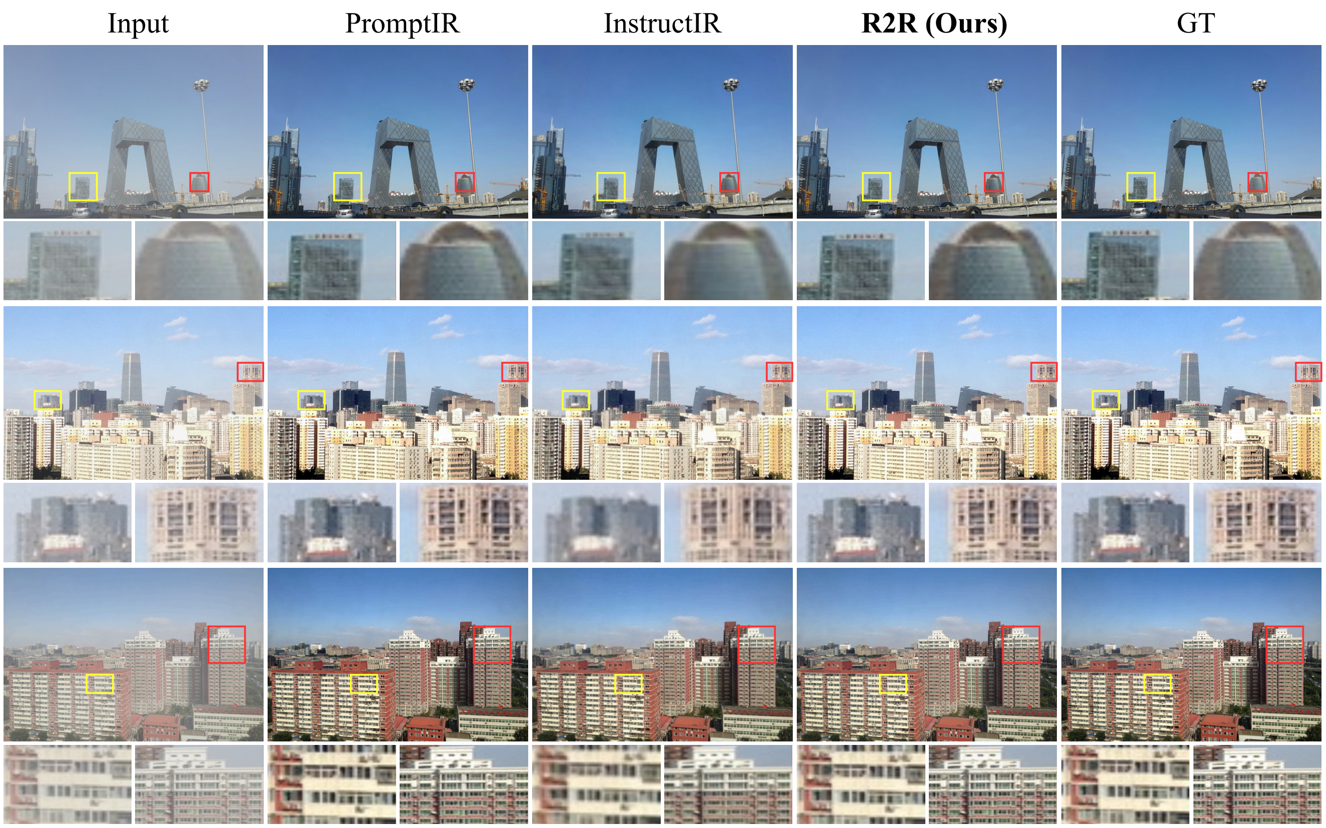

Figure 3. Visual comparison of R2R with state-of-the-art methods on the single haze degradation task.

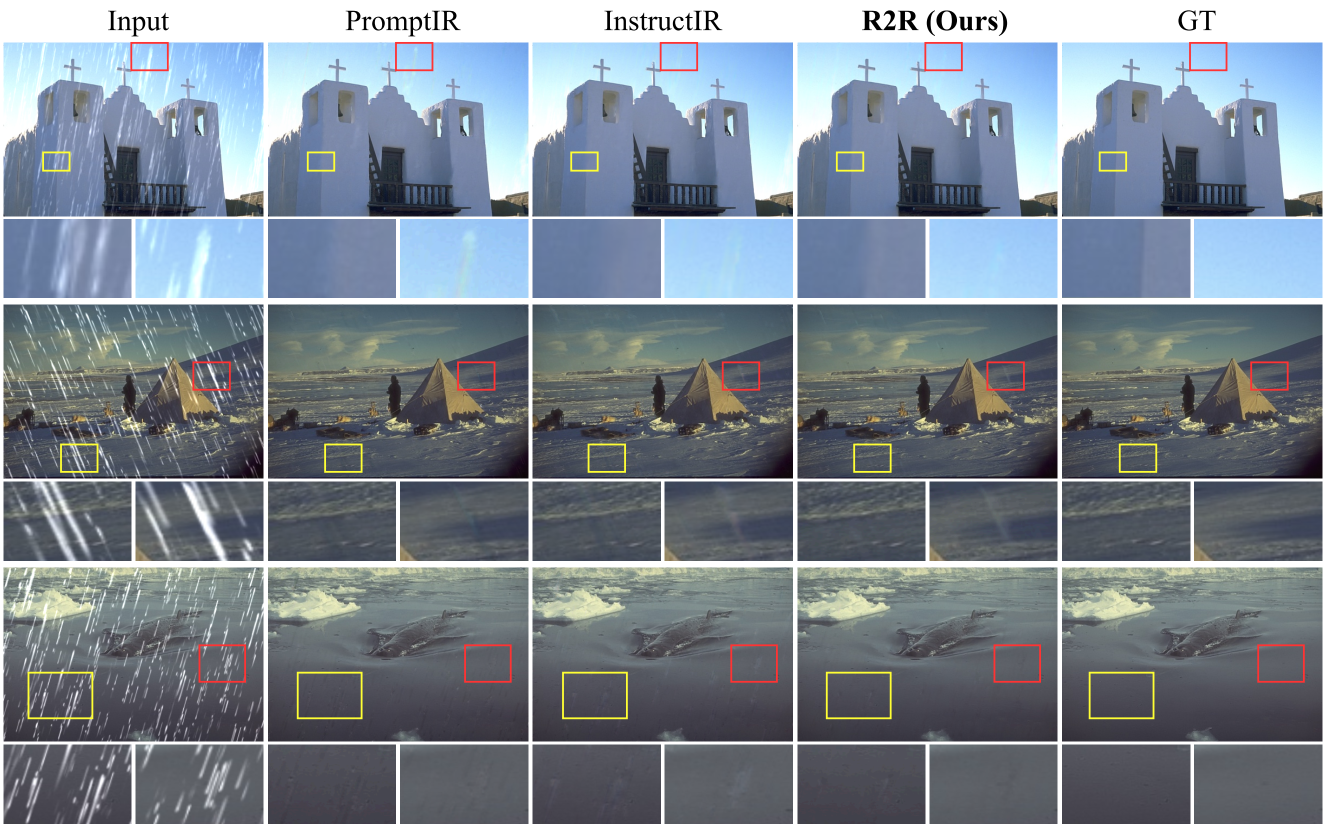

Figure 4. Visual comparison of R2R with state-of-the-art methods on the single rain degradation task.

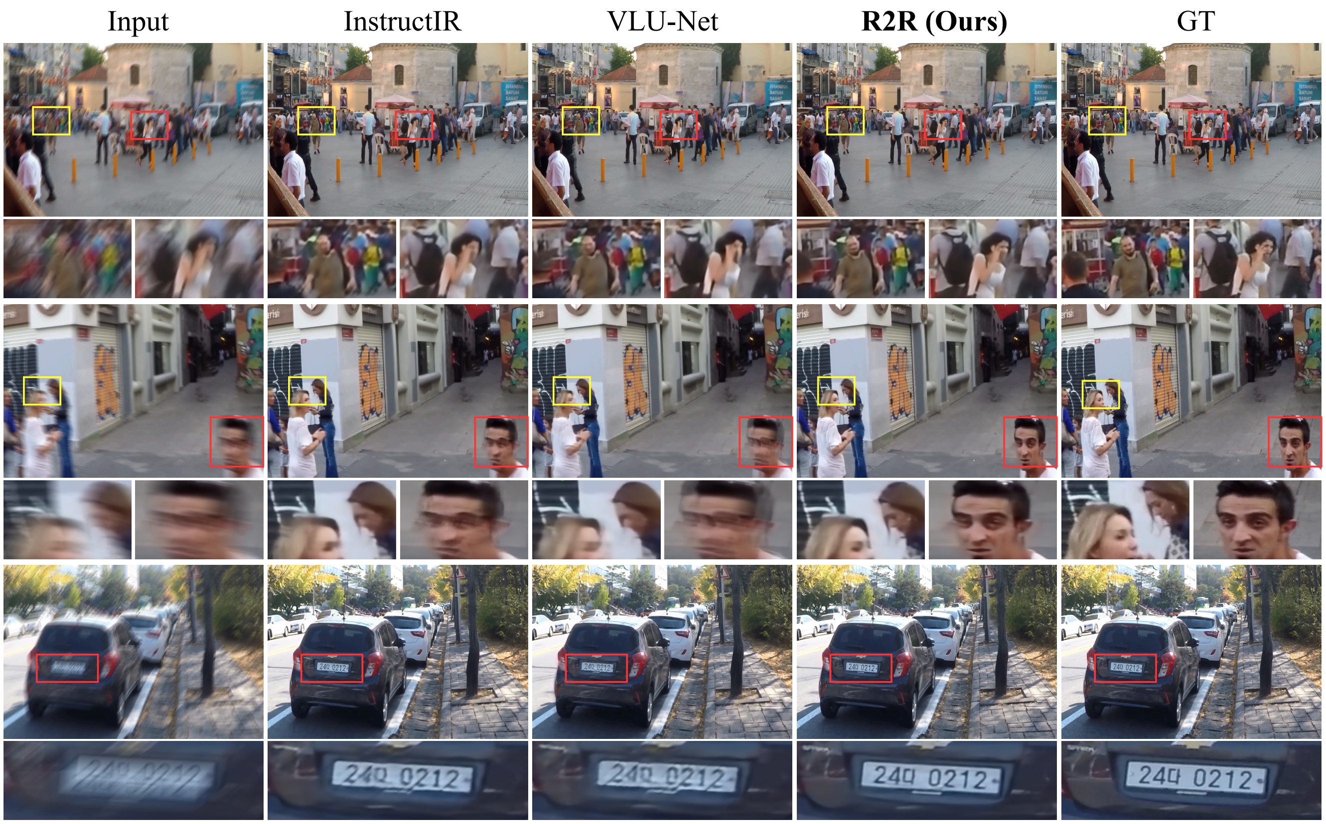

Figure 5. Visual comparison of R2R with state-of-the-art methods on the single blur degradation task.

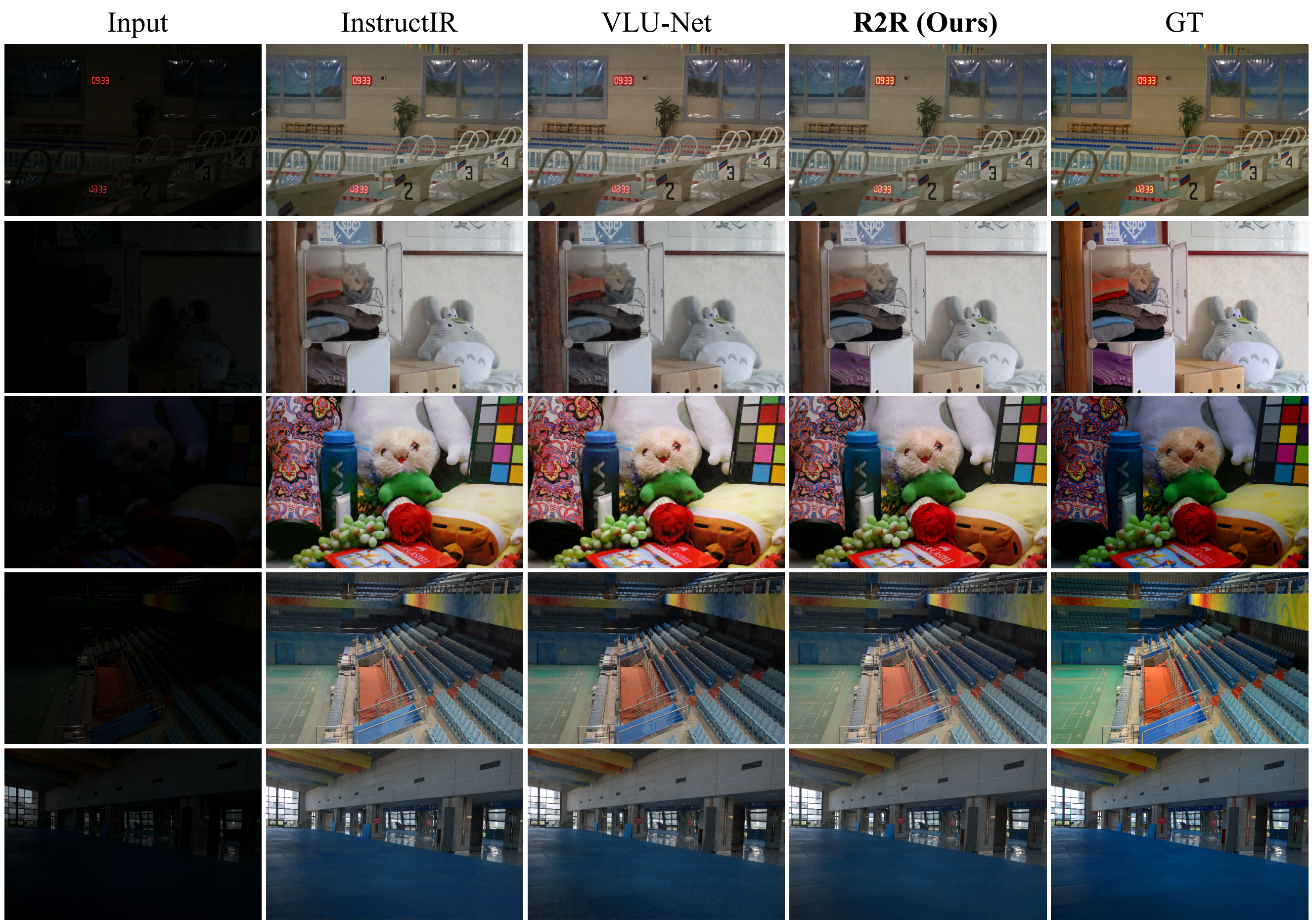

Figure 6. Visual comparison of R2R with state-of-the-art methods on the single lowlight degradation task.

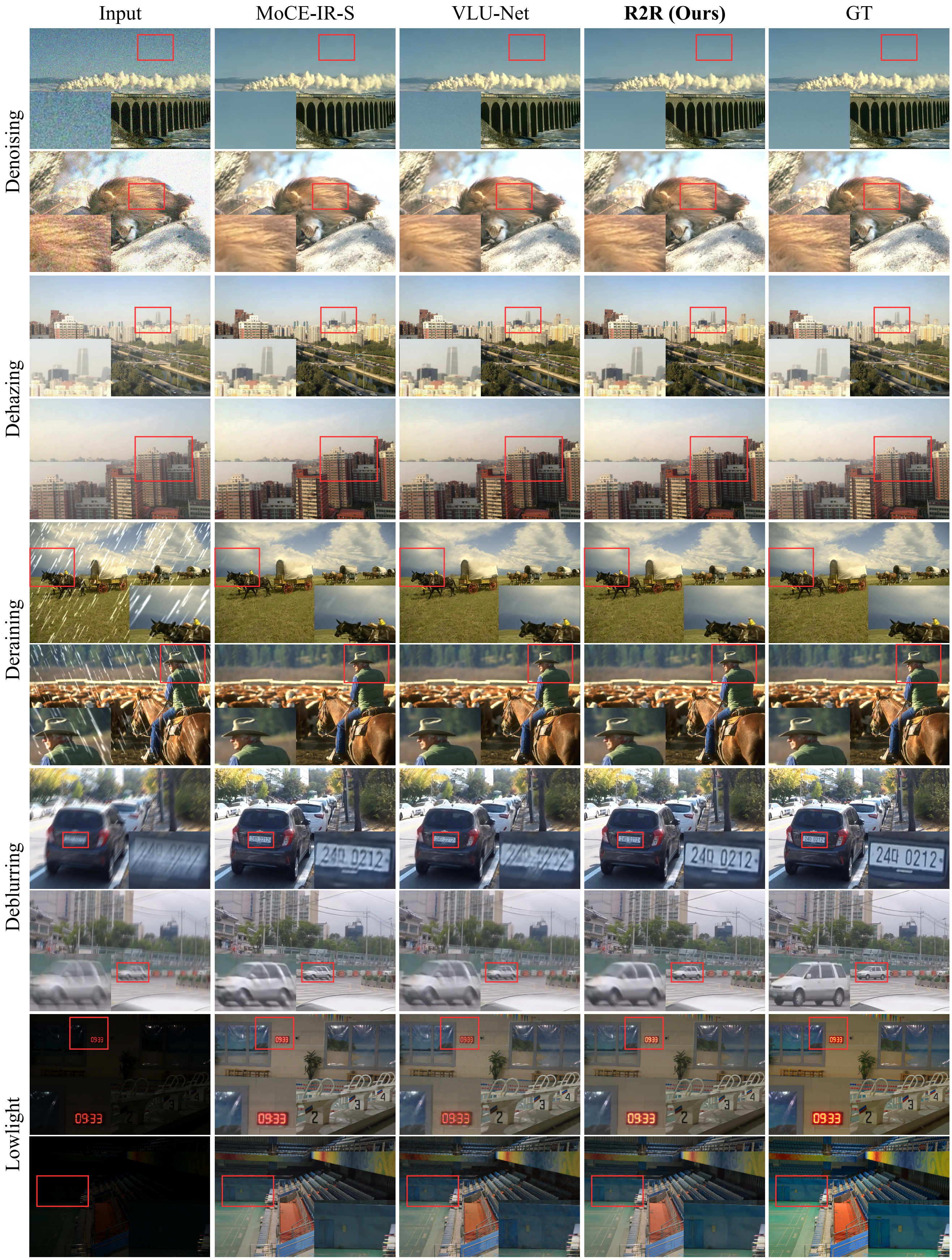

Figure 7. We provide a visual comparison of R2R with MoCE-IR-S [10] and VLU-Net [11] in the all-in-one setting with five degradations. Zoom in for a better view.

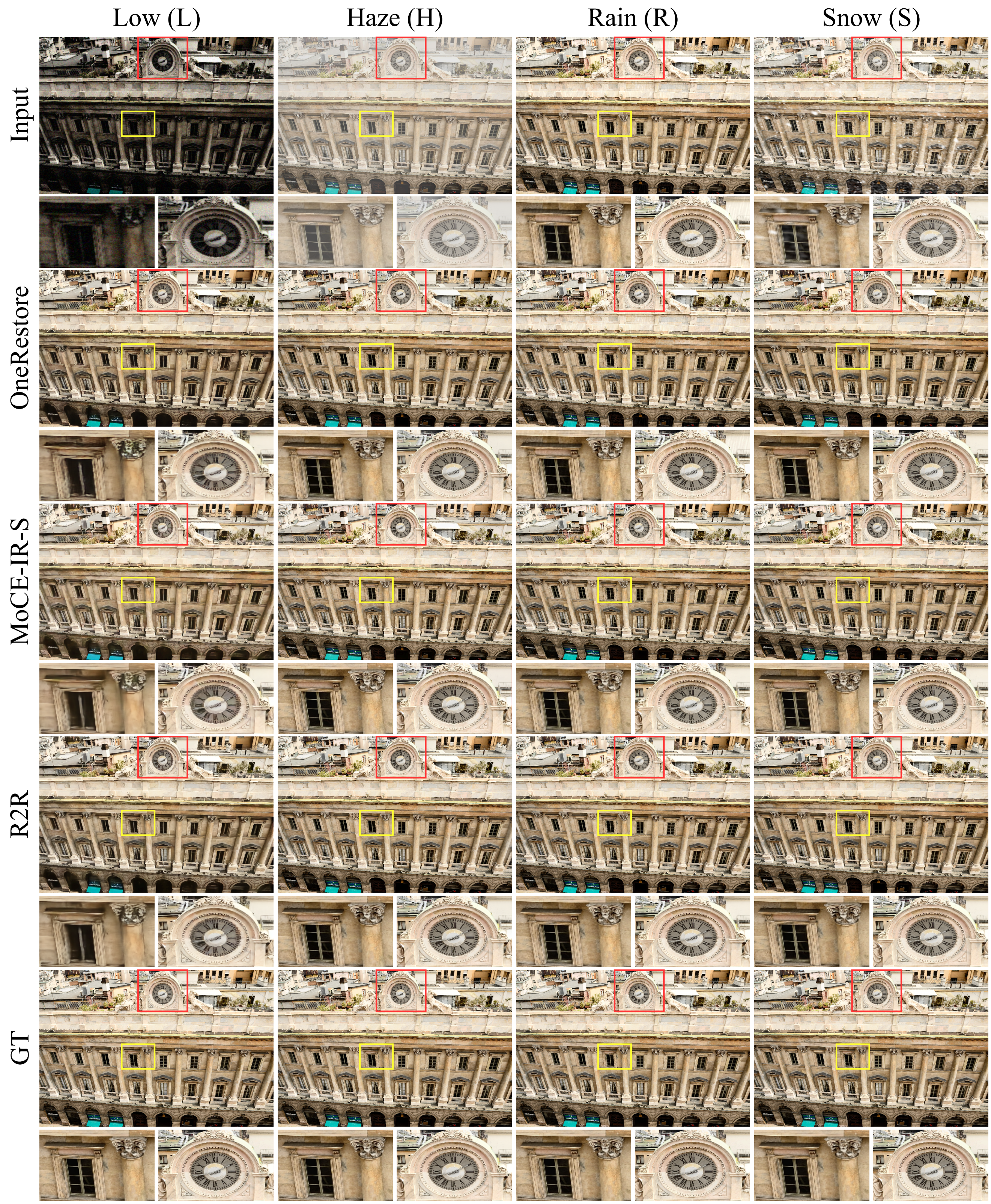

Figure 8. Comparing R2R with MoCE-IR-S [10] and OneRestore [3] on various composited degradations, including low illumination, haze, rain and snow.

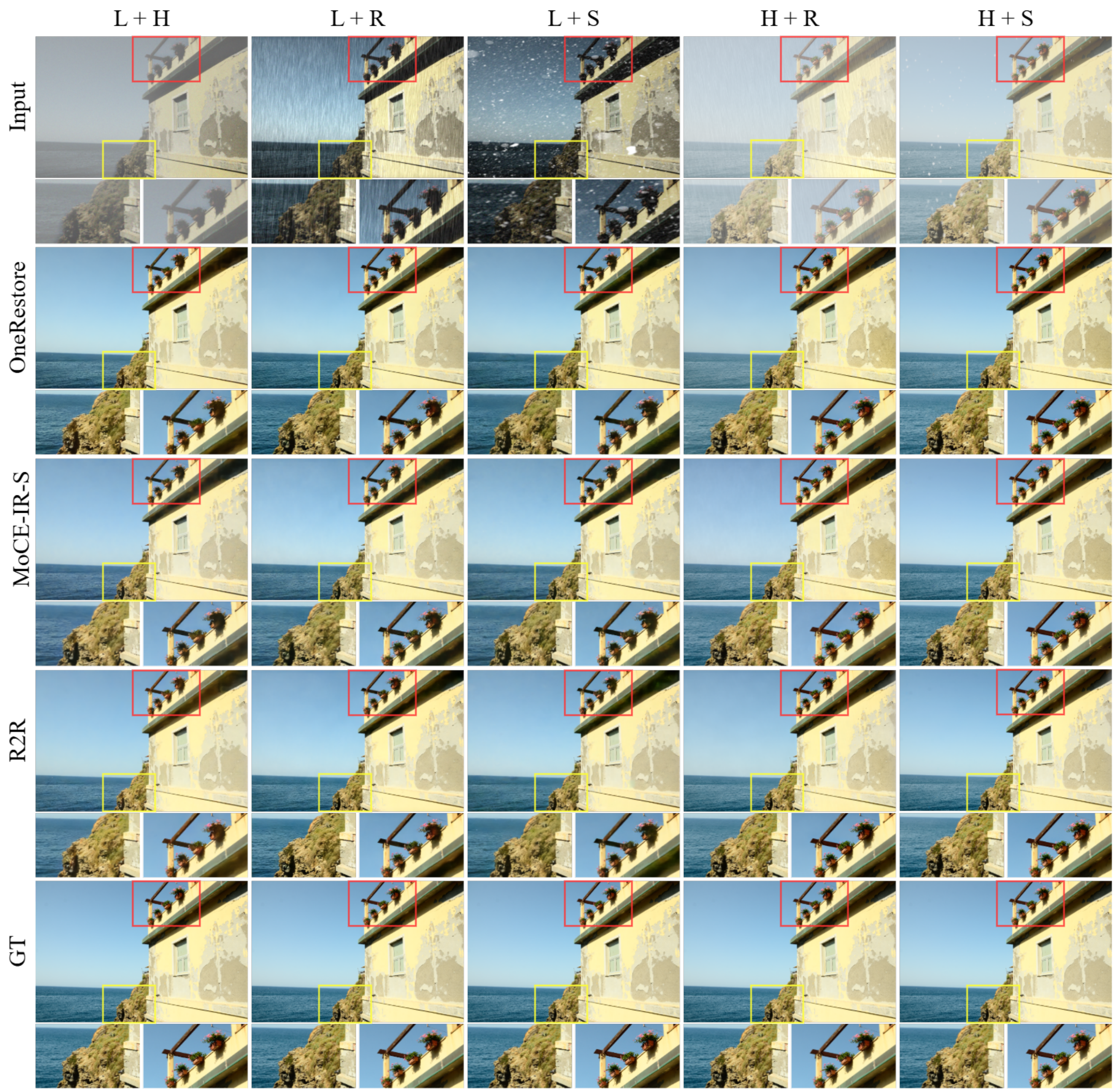

Figure 9. Comparing R2R with MoCE-IR-S [10] and OneRestore [3] on various composited degradations, including low + haze, low + rain, low + snow, haze + rain and haze + snow.

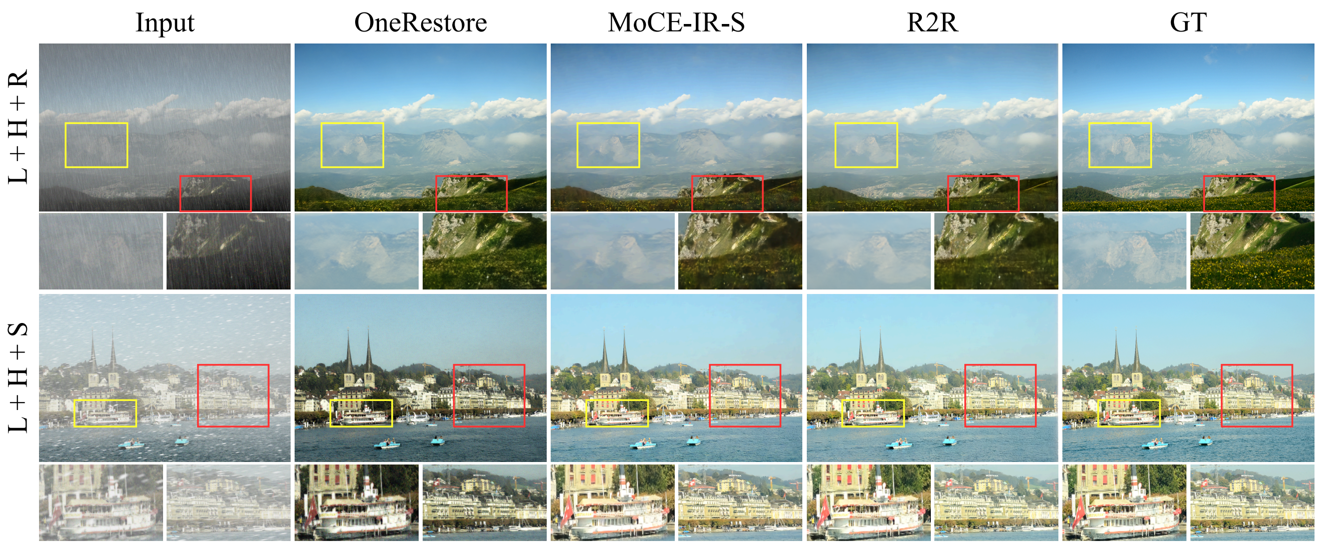

Figure 10. Comparing R2R with MoCE-IR-S [10] and OneRestore [3] on various composited degradations, including low + haze + rain and low + haze + snow.

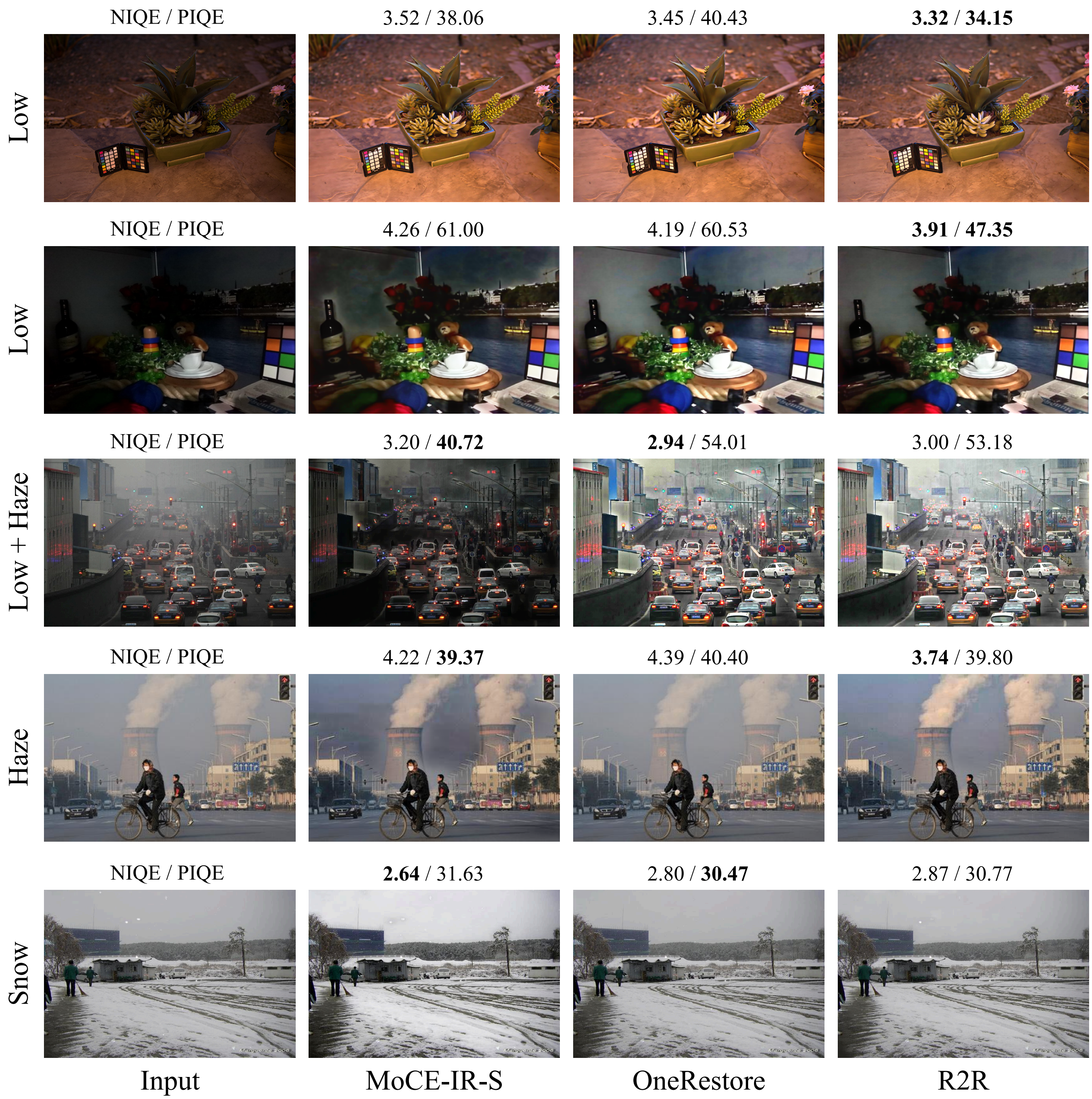

Figure 11. We provide a visual comparison of R2R with MoCE-IR-S [10] and OneRestore [3] in real-world scenes. Zoom in for a better view.

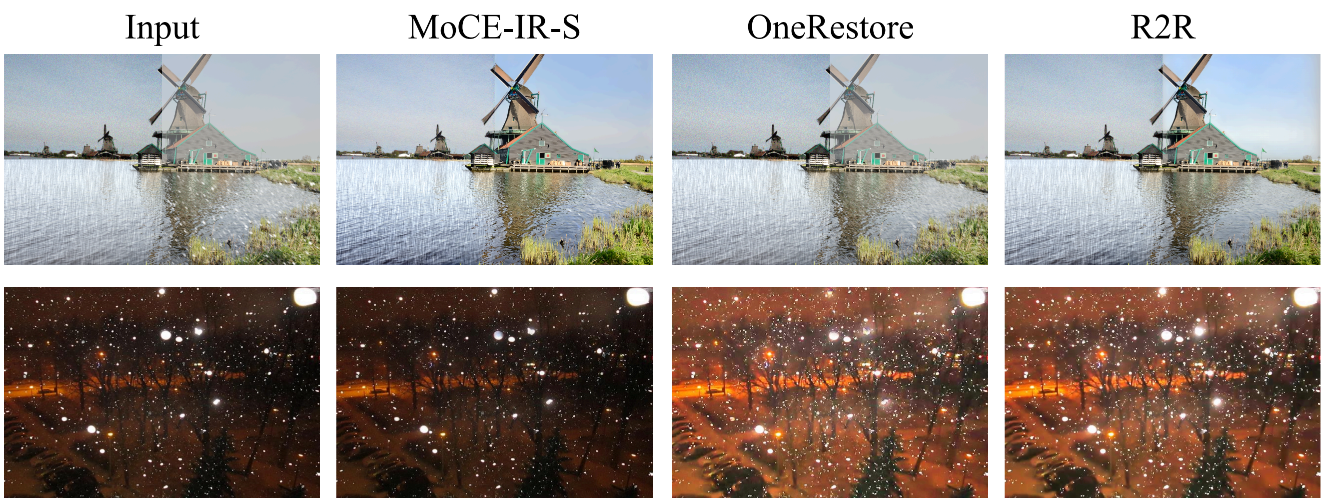

Figure 12. Failure cases of our method under extreme scenarios, including spatially non-uniform degradations and severely degraded conditions.