Complexity Experts are Task-Discriminative Learners for Any Image Restoration

Supplementary Material

In the supplementary material, we first provide more implementation details of our work in Sec. A. Next, we provide further ablations and analyses in Sec. B. Lastly, we conclude by presenting additional visual comparisons across the all-in-one settings, including five types of 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 [41] and AirNet [28] code for architecture development and training. We use fvcore [1] Python package for computing GFLOPS and parameter counts.

The MoE routing implementation is based on the publicly available re-implementation of sparse MoE [51] in PyTorch [2], as well as the official VMoE [48] JAX code [3] , which we have transcribed into PyTorch.

Training Time. Training proposed MoCE-IR in the all-in-one setting, which includes dehazing, deraining, and denoising, requires approximately two days using 4 NVIDIA RTX 4090 GPUs. For the more comprehensive all-in-one task involving five degradations, the training time extends to roughly four days. Optimizing the dispatching and merging of sub-batches for each expert during training could reduce the overall training time, which currently represents the primary computational bottleneck in our approach.

Datasets. For further zero-shot evaluations in Sec. B.2, we use the following datasets: defocus deblurring (DPDD) [1], motion deblurring (HIDE [52] and GoPro [39]), snow removal (RealSnow [81]), underwater restoration (UIEB [29]), and shadow removal (SRD [11]).

B. Additional Ablations

B.1. Model Design

High-Frequency Context. In addition to the current input feature, each routing function receives an additional input: a global context vector. This context vector is derived from the bottleneck output of the UNet by applying a frozen convolution initialized by a Sobel filter to extract high-frequency information.

Table 6. *More ablations on the architecture design.* We provide further insights in the design decisions of our MoCE-IR framework and the optimization function. We present the average PSNR (dB, ↑) and SSIM (↑) across the AIO-3 setting including SOTS, Rain100L and BSD68 benchmarks.

Method	L1	$\mathcal{F}$-L1	HF-Context	Experts	PSNR	SSIM
Baseline	✓	-	-	FFTFormer [25]	31.85	.910
	✓	✓	-	FFTFormer [25]	32.50	.915
	✓	✓	✓	ConvFormer [20]	32.51	.916
MoCE-IR-S	✓	✓	✓	FFTFormer [25]	**32.57**	**.916**

The goal is to prime the routing function to focus on high-frequency components, which are crucial for restoring fine image details. As shown in Tab. 6, incorporating the high-frequency (HF) context vector results in an overall improvement of 0.07 dB. When the full high-frequency awareness is introduced through both the optimization function and the HF context vector, a total improvement of 0.72 dB is observed.

Expert Block Architecture. Our complexity expert framework offers versatility, supporting diverse expert architectures and establishing a foundation for future research. As shown in Tab. 6, our proposed expert design outperforms convolutional-based experts [20], where increased receptive fields are achieved through larger kernel sizes.

Moreover, the complexity expert concept demonstrates potential for generalization across multiple domains, such as segmentation and classification tasks, suggesting its applicability in broader MoE architectures.

Optimization Function. In comparison to using only the traditional L1 loss in RGB space presented in Tab. 6, incorporating the combined losses results in an average performance improvement of 0.72 dB across dehazing, deraining, and denoising tasks. While only a few prior methods explicitly employ the FFT loss [53, 54, 73], other all-in-one approaches often rely on more complex training schedules with multiple stages [28, 76] or leverage large-scale language models for multimodal training [10, 14, 19, 35].

Primarily, incorporating language guidance into these methods typically results in more significant improvements than adding the FFT loss. For example, as shown in Tab. 7a, InstructIR-3D [10] achieves a gain of 1.71 dB over its non-language-based baseline, while OneRestore [19] demonstrates a 0.25 dB improvement in the composited degradation setting, see Tab. 7b. However, while the language guidance demands additional computational resources, it is important to note that the FFT loss delivers these benefits without

[1] github.com/facebookresearch/fvcore
[2] github.com/davidmrau/mixture-of-experts
[3] github.com/google-research/vmoe

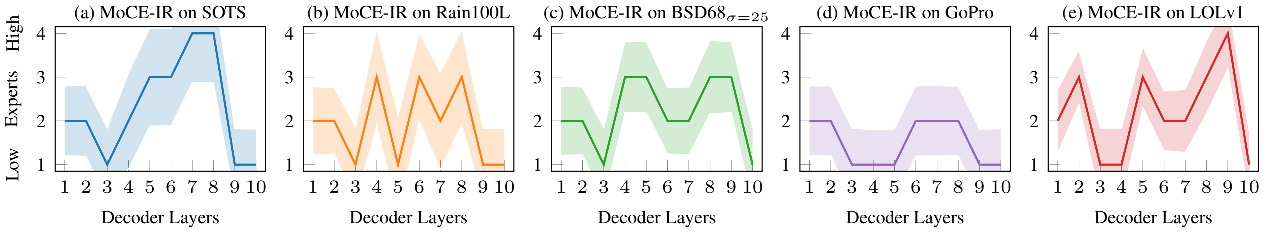


Figure 6. *Routing visualization for the AIO-5 setting.* Complexity-aware routing facilitates task-specific learning by assigning experts to input samples based on the task requirements. Our MoCE framework effectively extends this approach to even more diverse scenarios than the traditional all-in-one with three degradations. To illustrate, we visualize the average routing decisions for tasks such as dehazing, deraining, denoising, motion deblurring, and low-light enhancement, with the y-axis representing increasing expert complexity.

Table 7. *Comparison to language-guided methods.* We evaluate our framework against InstructIR [10] and OneRestore [19] 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 MoCE-IR 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
MoCE-IR	-	**30.48**	**.974**	**38.04**	**.982**	31.34	**.887**	**30.05**	**.899**	**23.00**	**.852**	**30.58**	**.919**

(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

Table 8. *Number of Complexity Experts.* We investigate the model performance of MoCE-IR-S when reducing or increasing the total numbers n of expert blocks. We present the average PSNR (dB, ↑) and SSIM (↑) across the AIO-3 setting including SOTS, Rain100L and BSD68 benchmarks.

# Experts	SOTS		Rain100L		BSD68		Avg.	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
2	30.56	.978	37.74	.980	31.15	.871	32.35	.914
4 *(ours)*	**30.94**	**.979**	**38.22**	**.983**	**31.22**	**.873**	**32.57**	**.916**
6	30.47	.977	37.82	.981	31.14	.870	32.34	.913

Table 9. *Zero-shot generalization.* We present results for AirNet [28], PromptIR [41] and MoCE-IR compared for zero-shot generalization to degradations unseen during training. We report PSNR (db, ↑), SSIM (↑) and LPIPS (↓) on the RGB images. We evaluate MoCE-IR on real-world data using IQA metrics: MANIQA/CLIPIQA/MUSIQ.

(a) *Blur degradations.*

Method	Params.	DPDD			GoPro			HIDE		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
AirNet [28]	9M	20.17	.674	.376	21.95	.748	.330	20.68	.731	.350
PromptIR [41]	36M	21.76	.673	.386	22.15	.749	.332	22.78	.740	.333
MoCE-IR	25M	**21.83**	**.678**	**.372**	**22.50**	**.758**	**.326**	**22.92**	**.743**	**.334**

(b) *Real-world degradations.*

Method	RealSnow (↑)			UIEB (↑)			SRD (↑)		
AirNet	.2040	.2832	38.08	.3148	.4645	51.65	.3869	.5619	58.54
PromptIR	.2133	.3037	38.35	.3164	.4734	51.82	.3925	.5991	**58.75**
MoCE-IR (ours)	**.2136**	**.3234**	**38.68**	**.3193**	**.4984**	**52.09**	**.3966**	**.6381**	58.60

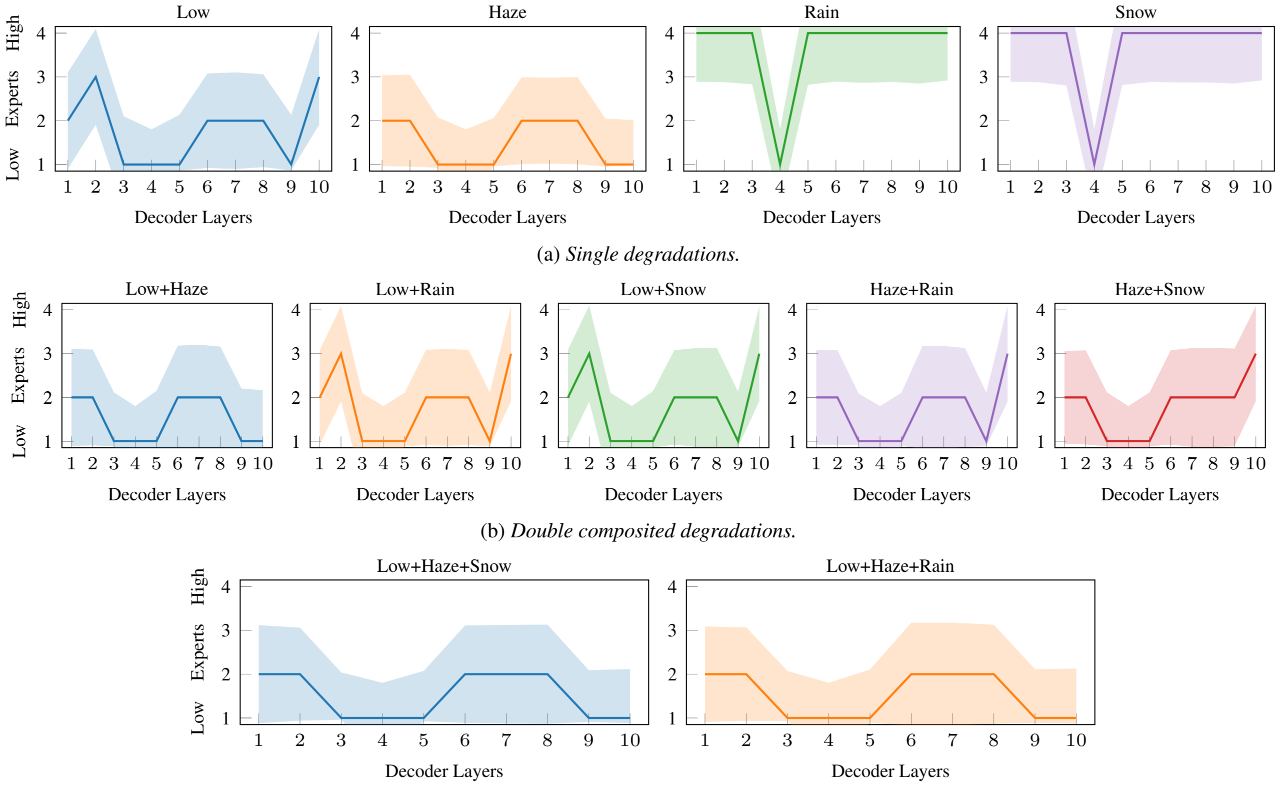
compromising model efficiency.

Number of Complexity Experts. In Tab. 8, we analyze the number of complexity experts within our MoCE layer. The total number of parameters remains constant, adhering to the nested parameter scaling rule, which distributes the number of channels across the experts. Additionally, the window partition sizes grow exponentially. Our observations indicate that neither reducing nor increasing the number of experts yields any significant benefit.

Notably, increasing the number of experts leads to a substantial rise in overall training time without corresponding improvements. Moreover, the expert with the highest complexity is rarely utilized across the considered degradations when the $n = 6$. We attribute this to the increasing homogeneity among experts by design, as more experts share similar parameter counts. Combined with the analysis in the main text, we conclude that experts need to exhibit distinct characteristics to ensure that the optimization process derives measurable benefits from higher-complexity experts.

B.2. Zero-Shot Generalization

We evaluate the zero-shot generalization of MoCE-IR, AirNet [28], and PromptIR [41], all trained under the AIO-3 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 [39] and HIDE [52], with none of the models trained on these degradations. For fair comparison, we use the official model checkpoints for AirNet and PromptIR without retraining. As shown in Tab. 9a, MoCE-IR outperforms both methods, achieving higher PSNR and SSIM scores and consistently lower LPIPS values across all datasets.



(a) *Single degradations.*

(b) *Double composited degradations.*

(c) *Triple composited degradations.*

Figure 7. *Routing visualization for the composited degradations setting.* Using CDD11 [19], we observe a more distinct separation in expert utilization compared to the AIO-3 and AIO-5 settings. Unlike these prior configurations, CDD11 provides uniformly high-quality images with varying combinations of degradations, effectively mitigating imbalances caused by differences in image size, perceptual image quality, the proportion of degraded samples, and content diversity.

Further, Tab. 9b presents real-world results, demonstrating MoCE-IR's effectiveness on unseen degradations such as snow, underwater scenes, and shadows. MoCE-IR maintains strong performance on popular NR-IQA metrics, including MANIQA [67], CLIPIQA [59], and MUSIQ [23].

B.3. Further Routing Analyses

We expand on the analysis presented in the main paper by including visualizations for the AIO-5 setting (see Fig. 6) and the composited setting (see Fig. 7). Furthermore, we investigate the effect of the $\mathcal{F}$-L1 loss on the expert utilization.

Expert Utilization On Complex Degradations. Even when faced with multiple simultaneous degradations or more complex scenarios where a single image exhibits two or three degradations, our MoCE framework demonstrates the desired task-discriminative behavior. Building on the trends discussed in the main text, where processing choices are aligned with the inherent requirements of each degradation, the visualizations for the composited degradation setting reveal additional insights. Specifically, certain degradations tend to dominate the routing decisions, and a general preference for experts with smaller receptive fields becomes apparent.

In all-in-one restoration, the goal is to effectively integrate task-specific learning with task-invariant learning, creating a model that fully leverages the unique aspects of each task while incorporating cross-task knowledge into the restoration process. As shown in Fig. 7, we observe a shift from task individuality toward a preference for shared feature learning. This transition is intuitive, given the nature of combined degradations, which naturally encourage more collaborative processing across tasks.

$\mathcal{F}$-Loss Complements Task-Discrimination. Additionally, we examine the impact of frequency awareness on expert utilization, as illustrated in Fig. 8. The left column depicts the routing decisions of MoCE-IR without incorporating the Fourier loss, while the right column shows the decisions with the Fourier loss applied.

In case of dehazing, visualized in Fig. 8 (a-b), there is a notable difference in expert selection during decoding with

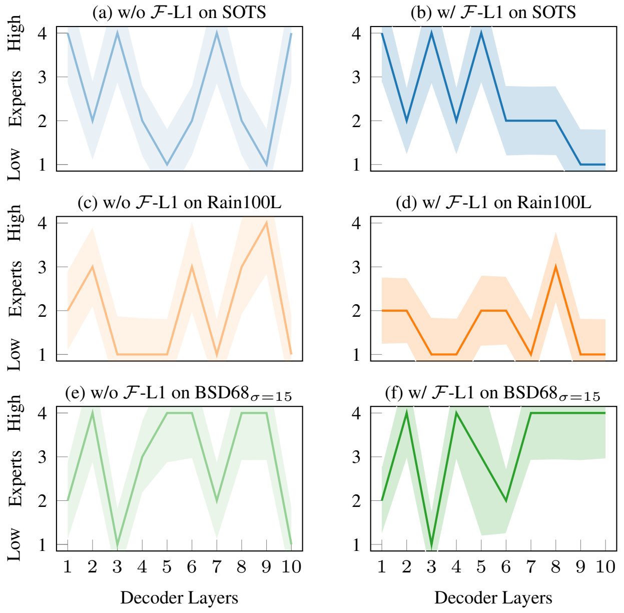


Figure 8. *Effect of $\mathcal{F}$-L1.* Placing greater emphasis on high frequencies enhances task-discriminative behavior, as visualized by comparing expert utilization in MoCE-IR models trained with and without the FFT loss. Without the FFT loss, the routing decisions in the later decoder layers become increasingly similar, resulting in suboptimal capacity utilization and ultimately poorer reconstruction performance.

and without the FFT loss. Without the FFT loss, the choice of experts in the final decoder layers shifts predominantly toward those with larger receptive fields, while with the FFT loss, the model exhibits a smoother and more stable routing, favoring smaller receptive field experts in later layers.

This difference may arises because the FFT loss enforces consistency in the frequency domain, compelling the network to balance global structural information (low frequencies) and local high-frequency details like edges and textures. Potentially, this difference arises because the FFT loss enforces consistency in the frequency domain, seamingly guiding the network to balance global structural information (low frequencies) and local high-frequency details like edges and textures. By doing so, the FFT loss reduces the reliance on large receptive field experts for correcting global inconsistencies in the later stages of the decoder. This dynamic routing demonstrates how the FFT loss guides the model toward more task-aligned expert utilization, optimizing computational complexity and promoting expert specialization to enhance restoration performance and fidelity.

C. Visual Results

Visual Comparison on Three Degradations. Fig. 9, we present additional visual examples under the three degradation settings, comparing our approach with AirNet [28] and PromptIR [41]. In challenging foggy scenarios, both AirNet and PromptIR face difficulties in fully restoring scene clarity, leaving residual fog and some color inconsistencies, while our method achieves more accurate color reconstruction. Similarly, in rainy conditions, the competing methods leave visible rain streaks, whereas our approach produces cleaner outputs.

In noisy scenes, our method delivers sharper and more detailed denoised results. Error heatmaps further illustrate the pixel-level discrepancies, showing fewer errors in our outputs compared to others. These results, in conjunction with quantitative evaluations, highlight the effectiveness of our approach.

Visual Comparison on Five Degradations. We present visual comparisons of MoCE-IR trained on five degradation types against InstructIR [10]. Similar to the results observed in the three-degradation setting, our model consistently restores image details from hazy, rainy, and noisy inputs. Beyond this, MoCE-IR 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 InstructIR [10].

Visual Comparison on Composited Degradations. In Fig. 11 and Fig. 12, we compare the visual quality of restored test samples with the non-language-based version of OneRestore [19]. The comparison contains the same image subjected to various degradation scenarios: two single degradations (low illumination and haze), three double-composited degradations (low + haze, low + rain, haze + rain), and one triple-composited degradation (low + haze + rain). To highlight the reconstruction differences, we include zoomed-in views of the restored outputs alongside corresponding error maps, which effectively visualize discrepancies, particularly those related to accurate color reconstruction.

In Fig. 11, we illustrate various restored regions across the input image, highlighting how different degradations impact specific areas and how our approach effectively restores the original image quality. Meanwhile, in Fig. 12, we examine the same image regions and their individual degradation patterns, demonstrating the strong performance of MoCE-IR compared to OneRestore. In particular, in the composited degradation setting, visible remnants of haze, rain, or color inaccuracies caused by low illumination are evident in the baseline approach, whereas our framework successfully addresses these issues.

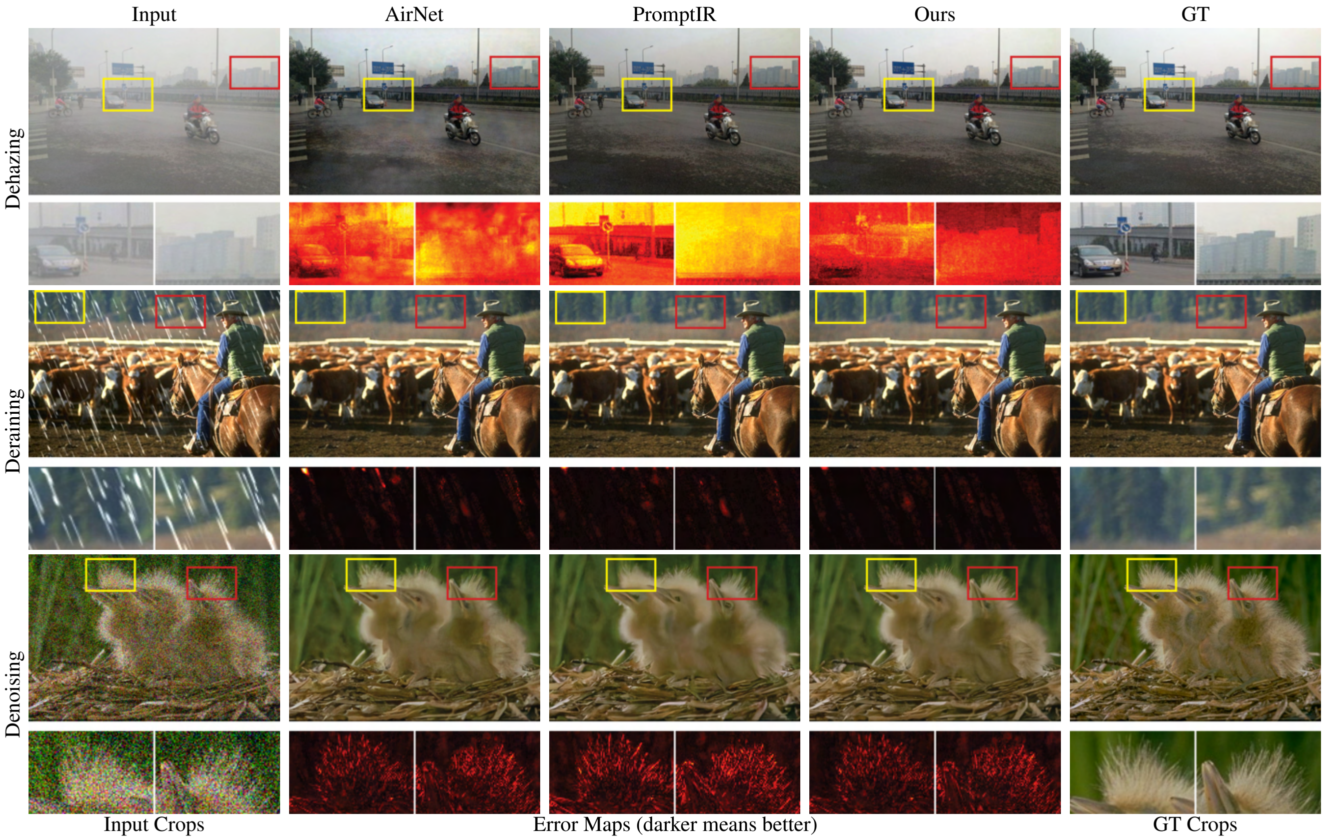


Figure 9. We provide a more detailed visual comparison of MoCE-IR-S with AirNet [28] and PromptIR [41] in the all-in-one setting with three degradations. To illustrate the differences, we include error heatmaps where the color transition from black to white indicates increasing pixel-wise error.

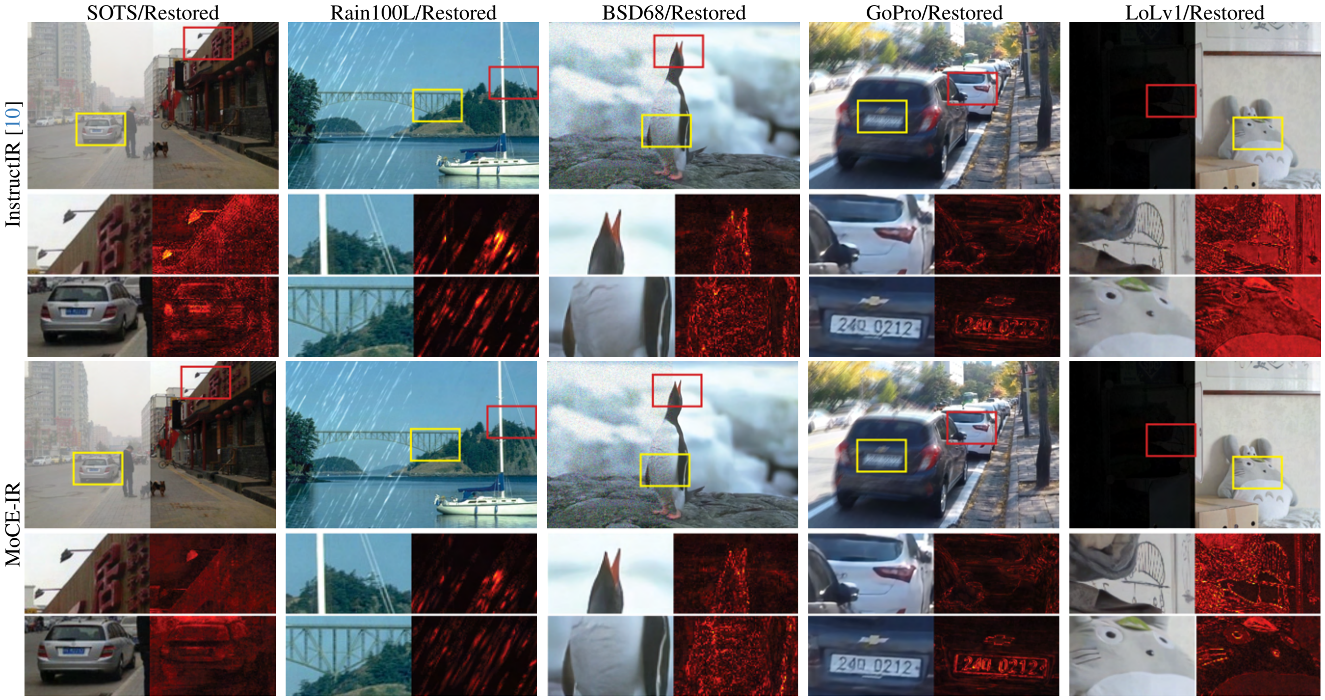


Figure 10. We provide a visual comparison of MoCE-IR with InstructIR [10] in the all-in-one setting with five degradations. To illustrate the differences, we include error heatmaps where the color transition from black to white indicates increasing pixel-wise error.

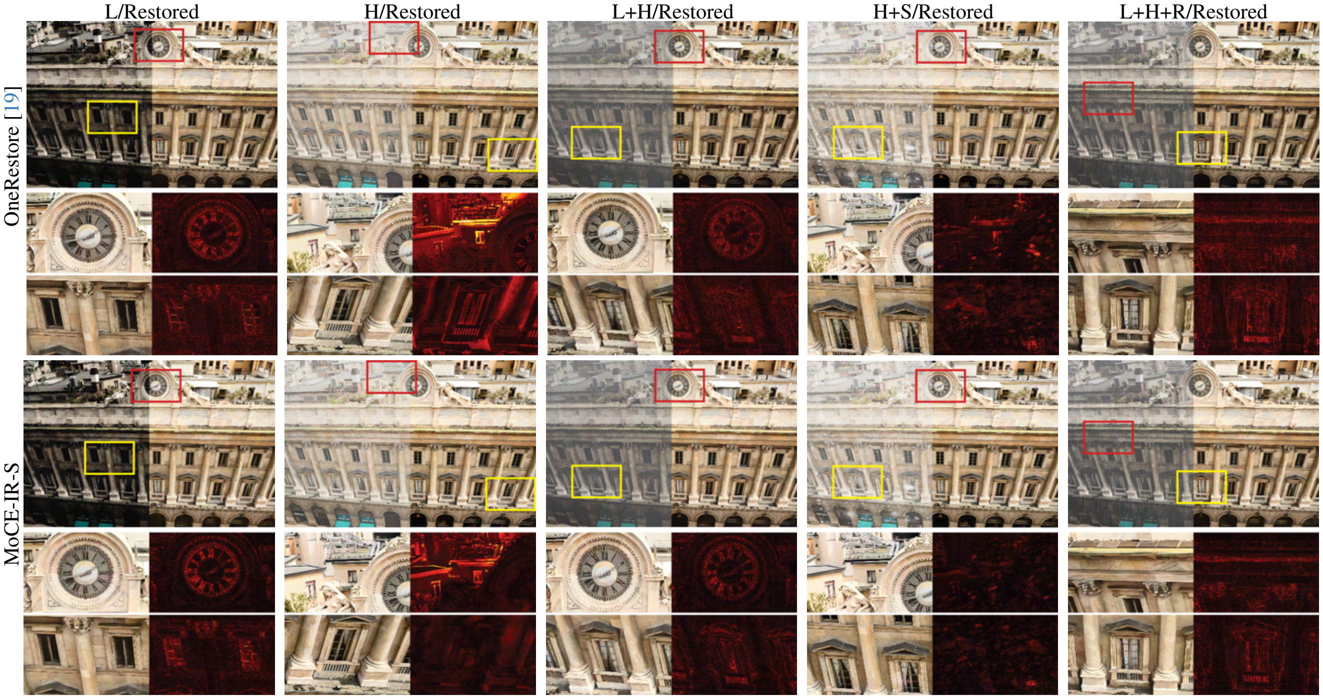


Figure 11. Comparing MoCE-IR-S with OneRestore [19] on various composited degradations, including low illumination, haze, snow, rain, as well as combinations of these in double and triple composited scenarios. We include the error heatmap with color transition from black to white denotes increasing pixel-wise erroneous.

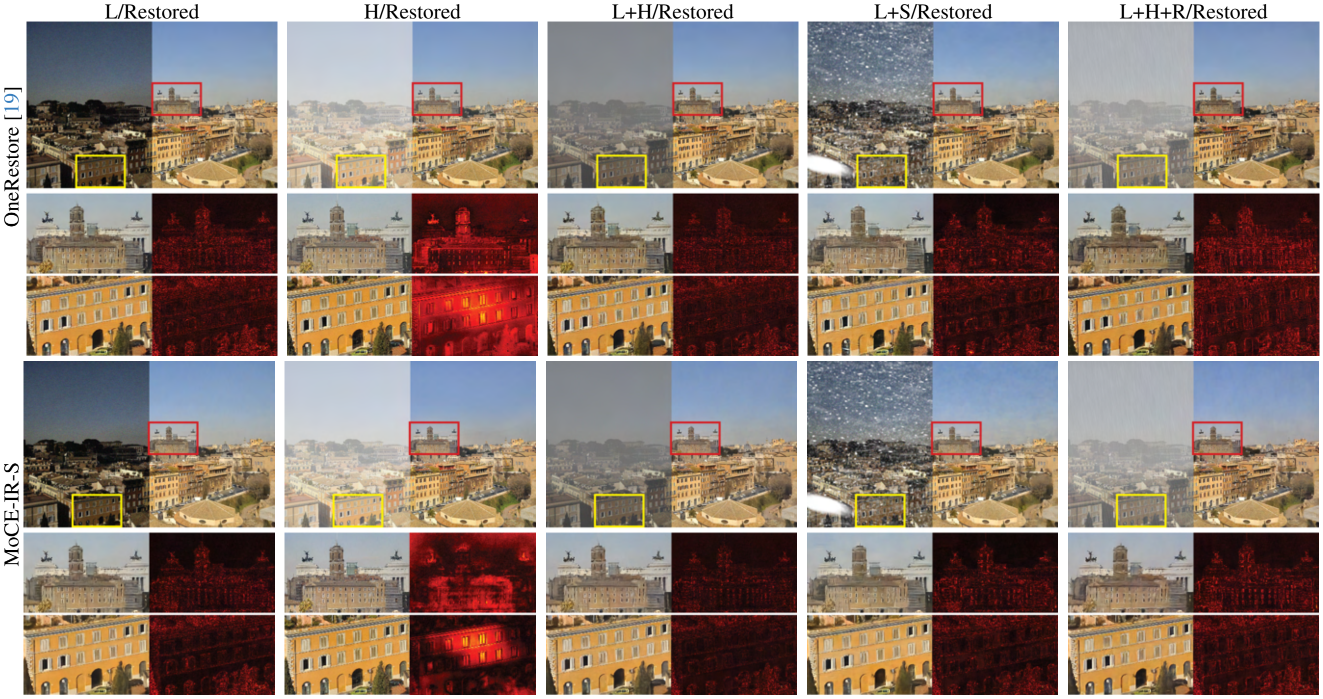


Figure 12. Further visual results comparing MoCE-IR-S with OneRestore [19] on various composited degradations, including low illumination, haze, snow, rain, as well as combinations of these in double and triple composited scenarios from CDD11 [19]. We include the error heatmap with color transition from black to white denotes increasing pixel-wise erroneous.