LEARNING FROM FAILURE: LEVERAGING UNRELIABLE PREDICTIONS IN SEMI-SUPERVISED REAL-WORLD ADVERSE WEATHER REMOVAL

Anonymous authors
Paper under double-blind review

ABSTRACT

Adverse weather image restoration aims to recover images degraded by rain, haze, snow, and other weather-induced artifacts, thereby improving the robustness of outdoor vision systems. Existing unified restoration models exhibit limited generalization to real-world scenes due to their reliance on synthetic supervision and insufficient semantic constraints. In this paper, we propose a novel student–teacher semi-supervised framework that addresses both challenges. Specifically, we introduce an unreliable database that preserves failed teacher predictions as informative negative samples for contrastive learning, while a reliable database stores high-quality teacher predictions as positive samples. By jointly exploiting reliable pseudo-ground truths and unreliable teacher outputs, the proposed framework learns to enhance desirable restoration characteristics while avoiding common failures. We further propose a phase spectrum-based semantic constraint that replaces computationally expensive text-based supervision with an efficient and naturally aligned semantic prior. An adaptive phase consistency loss is also designed to dynamically balance supervision between the degraded input and teacher pseudo-ground truths according to degradation severity. Extensive experiments on real-world benchmarks demonstrate that the proposed method consistently outperforms existing state-of-the-art approaches in restoration quality and perceptual fidelity while exhibiting stronger generalization to real-world adverse weather conditions.

1 INTRODUCTION

Images captured in real-world environments are often degraded by adverse weather conditions such as rain, haze, and snow, which significantly deteriorate the performance of outdoor vision systems. Existing image restoration methods have mainly focused on task-specific models tailored to individual weather phenomena (Wang et al., 2025b). Although these methods achieve strong performance through extensive training in large-scale synthetic datasets, their highly specialized designs limit scalability in diverse weather conditions.

To address this limitation, recent studies have proposed unified restoration frameworks capable of handling multiple weather degradations within a single model (Wang et al., 2025a; Wu et al., 2026). Although these approaches demonstrate promising results on synthetic benchmarks, they exhibit limited generalization to real-world scene. We attribute this to two key factors: their heavy reliance on synthetic supervision and neglect of the semantic context of the scenes, due to insufficient constraints during training.

To alleviate the dependency on synthetic data, unsupervised restoration methods have been proposed that require only unpaired real-world images (Liu et al., 2025a). However, these approaches fail to exploit the abundant labeled synthetic datasets that have become widely available. Semi-supervised learning offers a more attractive compromise by leveraging both labeled synthetic and unlabeled real-world data. For example, WResVLM (Xu et al., 2024) adopts a student–teacher framework in which the teacher generates pseudo-ground truths (pseudo-GTs) for unlabeled images. Nevertheless, the framework performs contrastive learning using only positive samples derived from reliable pseudo-ground truths, overlooking informative negative samples that could explicitly guide the stu-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

dent away from poor restoration results. We observe that the teacher frequently produces low-quality predictions during training. These predictions often contain severe artifacts, color distortions, or incomplete restoration (Fig. 9b), and have previously been simply discarded. However, we note that they are *not just* generic negative samples, but are in fact sampled directly from the model in an *on-policy* manner, and hence provide valuable information (Tang et al., 2024). To address the previous limitation, we propose a mechanism that leverages these informative failures for contrastive semi-supervised learning. More specifically, we retain these cases in an unreliable database and treat them as negative samples for contrastive learning. By simultaneously attracting the student toward reliable restorations while repelling it from unreliable ones, the proposed framework provides stronger supervision and improves generalization to real-world adverse weather conditions.

Besides the supervision strategy, effective semantic guidance is equally important for restoring visually plausible images. Recent methods employ large vision-language models (VLMs) to generate textual descriptions that explicitly constrain image semantics (Xu et al., 2024). Although effective, text supervision suffers from two inherent drawbacks. First, textual descriptions are only weakly aligned with the fine-grained spatial structures required for low-level vision tasks because of language ambiguity and the lossy language-vision representation gap (Wu et al., 2024). Second, generating accurate descriptions with large VLMs introduces substantial computational overhead, slowing the training process.

Instead of relying on language supervision, we exploit the Fourier phase spectrum as an efficient structural prior. It is well established that the Fourier phase spectrum primarily preserves structural and semantic information, whereas the amplitude spectrum mainly represents low-level appearance (Oppenheim et al., 1979). Existing Fourier-based restoration methods (Guo et al., 2022) mainly focus on recovering degraded amplitude information. In contrast, we employ the phase spectrum to constrain structural consistency during semi-supervised learning. Even though the pseudo-GTs are accessible, using them alone is not optimal. (Oppenheim & Lim, 1980) shows that the phase spectrum is inherently robust to illumination variations and color shifts, making it possible to also use the phase spectrum of the input image to constrain the structure of the output image. Nevertheless, severe weather degradations may also corrupt phase information by obscuring structural details (Liu et al., 2025a). Therefore, directly enforcing consistency with the phase spectrum of either the degraded input or pseudo-GT can lead to sub-optimal solutions. To overcome this issue, we propose an *adaptive phase consistency loss* that dynamically balances supervision between the degraded input and the teacher-generated pseudo-GT according to degradation severity. Consequently, the model relies more on the input phase for mild degradations while progressively shifting toward the pseudo-GT phase under severe degradations.

Our contributions are summarized as follows:

- We propose a new mechanism in the family of semi-supervised weather removal methods that identifies and preserves failed teacher predictions in an unreliable database, enabling their use as negative samples for contrastive semi-supervised learning and improving generalization to real-world adverse weather image restoration.
- We introduce the Fourier phase spectrum as an efficient structural prior, eliminating the need for computationally expensive and ambiguous text-based supervision.
- We propose an adaptive phase consistency loss that automatically adjusts structural supervision according to degradation severity, producing more robust restorations across diverse adverse weather conditions.
- The proposed framework establishes new state-of-the-art performance among unified and semi-supervised image restoration methods, outperforming existing approaches on both quantitative and qualitative evaluation.

2 Related Works

2.1 Task-specific Image Restoration

Image restoration is a fundamental problem in computer vision, aiming to recover high-quality images from degraded observations. Early task-specific methods (Jiang et al., 2017; Ancuti & Ancuți, 2013) primarily relied on physical imaging models and handcrafted priors tailored to individual

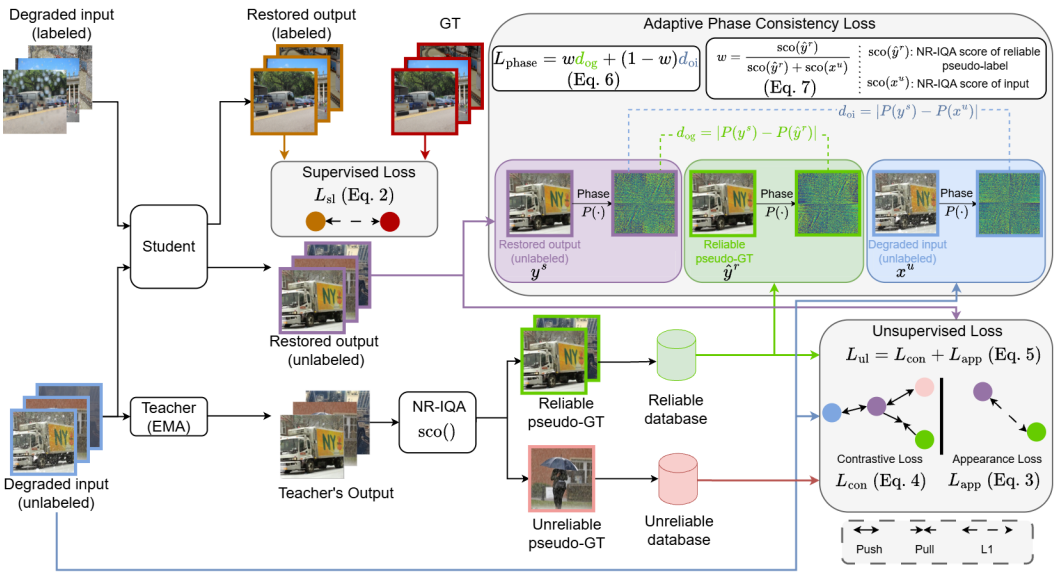

Figure 1: The illustration of the proposed semi-supervised learning framework.

degradation types. With the advent of convolutional neural networks (Yang et al., 2024; 2023) and, more recently, transformer-based architectures have significantly advanced restoration performance across a wide range of tasks. For example, Restormer (Zamir et al., 2021) introduces an efficient multi-scale attention mechanism that achieves an excellent balance between restoration quality and computational efficiency. These transformer-based architectures have been successfully applied to deraining (Chen et al., 2023), desnowing (Zhang et al., 2022), and dehazing (Song et al., 2023). Despite their impressive performance, most existing methods remain task-specific and require separate models for different degradation types, leading to increased training, storage, and deployment costs.

2.2 All-in-One Image Restoration

To mitigate the inefficiency of deploying multiple models for diverse weather degradations, unified image restoration methods have attracted increasing attention (Zhu et al., 2023; Ye et al., 2023). While these approaches achieve impressive performance on synthetic benchmarks, they often struggle in real-world scenarios due to the substantial domain gap between synthetic and real-world degradations. To bridge this gap, (Liu et al., 2025b) proposed a self-evolving framework that combines physics-driven synthetic data generation with a dual-level reinforcement learning strategy. Meanwhile, (Tang et al., 2026) addressed the problem in an unsupervised setting employing global and local discriminators to enhance image realism while preserving structural consistency. WResVLM (Xu et al., 2024) introduced a semi-supervised framework that leverages both labeled synthetic data and unlabeled real-world data through VLM-based semantic supervision. Although effective, the reliance on large VLMs incurs substantial computational overhead and limits scalability and adaptability to resource-constrained applications. we observe that the Fourier phase spectrum inherently preserves the structural and semantic information of an image while being significantly more computationally efficient than VLM-based representations. Motivated by this observation, we propose a phase spectrum-based semantic constraint that provides explicit semantic guidance for semi-supervised learning

3 Methodology

3.1 Overall Framework

Motivated by the observations that existing semi-supervised image restoration methods lack informative negative samples and that VLM-based semantic constraints are computationally expensive,

we propose a novel semi-supervised framework that addresses both limitations. Specifically, we incorporate an unreliable image database into the unsupervised learning process to provide challenging negative samples, while replacing conventional VLM-based semantic supervision with a lightweight phase spectrum-based semantic constraint. This design improves both training effectiveness and computational efficiency.

As illustrated in Fig. 1, our method follows the conventional student–teacher paradigm widely adopted in semi-supervised learning (Sohn et al., 2020). The framework consists of a student network f_{θ_s} and a teacher network f_{θ_t} sharing the same architecture. The teacher parameters are updated using the exponential moving average (EMA) of the student parameters, while the student network is optimized by minimizing the following objective,

$$L_{\text{total}} = L_{\text{sl}} + \alpha L_{\text{ul}} + \beta L_{\text{phase}} \,, \tag{1}$$

where

$$L_{\text{sl}} = \sum_{i=1}^{N} |f_{\theta_s}(x_i^l) - y_i^l| \tag{2}$$

denotes the supervised reconstruction loss on the labeled dataset, L_{ul} denotes the unsupervised loss (Eq. 5), and L_{phase} denotes the proposed adaptive phase consistency loss for structural constraint (Eq. 6). Here, (x_i^l, y_i^l) represents a labeled degraded image and its corresponding ground truth image, while α and β are balancing coefficients. In our paper, both are empirically set to 0.2.

3.2 Contrastive Learning for Unsupervised Learning

Ideally, the teacher network produces higher-quality predictions than the student network and therefore provides reliable supervisory signals for unlabeled data. To improve the quality of pseudo-GTs, other methods (Huang et al., 2023; Xu et al., 2024) retain only reliable teacher outputs, where reliability is assessed using the no-reference image quality assessment (NR-IQA) metric. However, this assumption does not always hold in practice. As shown in Fig. 9, the teacher may still generate unsatisfactory predictions containing severe artifacts, color distortion, or only marginal improvements over the degraded input. Such erroneous outputs are simply discarded in existing methods, leaving their valuable negative information unexplored.

To exploit these failure cases, we introduce an *unreliable database* that explicitly stores the best historical failure for each unlabeled sample. Instead of treating poor predictions as useless, we regard them as informative negative samples that indicate undesirable restoration outcomes. During training, the student network is encouraged to avoid producing outputs similar to those stored in the unreliable database, while the teacher network inherits this behavior through EMA updates.

Specifically, we initialize both the reliable database and the unreliable database as empty. For each unlabeled image $x_i^u \in D_U = \{x_i^u\}_{i=1}^M$, the current teacher prediction is compared with the degraded input, the student prediction, and the historical entries in both databases. If the teacher prediction achieves the highest image quality, it replaces the corresponding entry in the reliable database. Otherwise, if it represents a failed prediction but is of higher quality than the existing entry in the unreliable database, the unreliable database is updated accordingly. The complete updating procedure is summarized in Algo. 1. Consequently, the reliable database $B_R = \{\hat{y}_i^r\}_{i=1}^M$ maintains the best pseudo-GT for each unlabeled image, whereas the unreliable database $B_U = \{\hat{y}_i^u\}_{i=1}^M$ records the strongest failure encountered during training, providing explicit negative supervision that discourages the model from repeating similar restoration errors.

With the reliable pseudo-GTs and unreliable samples available, the unsupervised appearance loss is formulated as follows:

$$L_{\text{app}} = \sum_{i=1}^{M} |y_i^s - \hat{y}_i^r|, y_i^s = f_{\theta_s}(x_i^u) \,, \tag{3}$$

where y_i^s is the restored student network output. To keep the restored output close to the reliable pseudo-GT and separated from the degraded input image and unreliable sample, we implement contrastive loss in the training

$$L_{\text{con}} = \sum_{j=i}^{K} \sum_{i=1}^{M} \gamma_j \frac{e^{|\phi_j(y_i^s) - \phi_j(\hat{y}_i^r)|}}{\sum_{y \in \{\hat{y}_i^r, \hat{y}_i^u, x_i^u\}} e^{|\phi_j(y_i^s) - \phi_j(y)|}} \,, \tag{4}$$

Algorithm 1 Update of Reliable and Unreliable Database

Require: NR-IQA method $\text{sco}(\cdot)$, unlabeled dataset D_U, batch size b, teacher network f_{θ_t}, student network f_{θ_s}, unreliable database B_U, and reliable database B_R
Ensure:
Initialize $B_U = \emptyset, B_R = \emptyset$
for batch $B = \{x_i^u\}_{i=1}^b$ in D_U **do**
  for x_i^u in B **do**
    $y_i^t = f_{\theta_t}(x_i^u)$ *# teacher's output*
    $y_i^s = f_{\theta_s}(x_i^u)$ *# student's output*
    # Compute NR-IQA scores of x_i^u, y_i^t, y_i^s, $\hat{y}_i^r \in B_R$, *and* $\hat{y}_i^u \in B_U$
    $z_i^x = \text{sco}(x_i^u), z_i^t = \text{sco}(y_i^t), z_i^s = \text{sco}(y_i^s)$
    $z_i^r = \text{sco}(\hat{y}_i^r), z_i^u = \text{sco}(\hat{y}_i^u)$
    if $z_i^t > z_i^x$ and $z_i^t > z_i^s$ and $z_i^t > z_i^r$ **then**
      Replace $\hat{y}_i^r$ in B_R by y_i^t
    else if $z_i^t > z_i^u$ **then**
      Replace $\hat{y}_i^u$ in B_U by y_i^t
    end if
  end for
end for

where $\phi_j(\cdot)$ represents the j_{th} hidden layer of the pre-trained VGG-19 (Simonyan & Zisserman, 2014), and γ_j is the weight coefficient. From Eq. 3 and 4, we derive our loss function for unsupervised learning,

$$L_{\text{ul}} = L_{\text{con}} + L_{\text{app}} \tag{5}$$

3.3 Structural Constraint for Unsupervised Learning

Restoring images degraded by adverse weather requires not only recovering visual fidelity but also preserving the underlying structures obscured by weather-induced artifacts. Semantic supervision is particularly beneficial in semi-supervised learning, where reliable labels are unavailable for real-world data. To exploit semantic information, WResVLM (Xu et al., 2024) employs VLMs to construct a contrastive semantic supervision objective. Although effective, such approaches incur substantial computational overhead and rely on language representations that are inherently ambiguous and weakly aligned with the fine-grained spatial structures required for low-level vision tasks.

Instead, we propose a lightweight phase spectrum-based structural constraint. Our design is motivated by two observations. First, previous studies have shown that the Fourier phase spectrum preserves most of the semantic and structural information of an image (Guo et al., 2022). Second, although the phase spectrum is generally robust to illumination changes and color shifts, severe degradations may still corrupt its structural information (Liu et al., 2025a). Consequently, neither the degraded input nor the teacher-generated pseudo-GT should be treated as an absolutely reliable structural reference.

To address this issue, we construct an adaptive phase consistency loss by combining the phase spectrum of the degraded input and the reliable pseudo-GT according to their estimated image qualities. Specifically, the student prediction is encouraged to align its phase spectrum with this adaptive reference through the following phase consistency loss

$$L_{\text{phase}} = \sum_{i=1}^{M} w_i d_{\text{og}} + (1 - w_i) d_{\text{oi}} \,, \tag{6}$$

where $P(\cdot)$ denotes the Fourier phase spectrum, $d_{\text{og}} = |P(y_i^s) - P(\hat{y}_i^r)|$ and $d_{\text{oi}} = |P(y_i^s) - P(x_i^u)|$ are L1 distance.

$$w_i = \frac{z_i^r}{z_i^x + z_i^r} \,, \tag{7}$$

where $z_i^r = \text{sco}(\hat{y}_i^r)$ and $z_i^x = \text{sco}(x_i^u)$ denote the NR-IQA scores of the reliable pseudo-GT and degraded input, respectively. Intuitively, severe degradations result in a significantly lower quality score for the input image than for the pseudo-GT, yielding a larger w_i . In this case, the structural

Models		Rain		Snow		Haze		Average		#Param	FLOP
Name	Year	LIQE↑	CLIP-IQA↑	LIQE↑	CLIP-IQA↑	LIQE↑	CLIP-IQA↑	LIQE↑	CLIP-IQA↑	M	G
Restormer	CVPR22	2.277	0.437	3.172	0.510	1.918	0.366	2.456	0.438	26.12	141.00
TransWeather	CVPR22	1.924	0.358	2.770	0.416	1.502	0.292	2.065	0.355	38.05	12.24
TKL	CVPR22	2.028	0.392	2.830	0.428	1.590	0.318	2.149	0.379	28.47	49.22
WeatherDiffusion	TPAMI23	2.050	0.395	2.950	0.466	1.520	0.326	2.173	0.396	82.96	475.16
WGWS-Net	CVPR23	1.965	0.389	2.619	0.395	1.506	0.310	2.030	0.365	5.19	13.61
MWDT	ICCV23	2.068	0.377	2.903	0.407	1.720	0.315	2.230	0.366	-	-
PromptIR	NeurIPS23	2.250	0.439	3.121	0.517	1.941	0.372	2.437	0.443	35.59	15.81
DA-CLIP	ICLR24	2.250	0.412	3.050	0.499	2.014	0.325	2.438	0.412	174.10	118.50
WResVLM	ECCV24	2.563	0.477	3.293	0.519	2.064	0.371	2.640	0.456	-	-
Histoformer	ECCV24	2.563	0.443	3.273	0.513	1.982	0.386	2.606	0.447	16.62	91.59
MODEM	NeurIPS25	**3.499**	0.582	3.823	0.584	2.832	0.447	3.385	0.538	19.96	463.86
HOGformer	AAAI26	3.396	0.585	3.827	**0.612**	2.659	0.456	3.294	0.551	16.65	91.77
Ours	-	3.466	**0.597**	**3.852**	0.611	**3.263**	**0.507**	**3.527**	**0.572**	19.67	153.94

Table 1: Quantitative comparison in all-in-one setting. **Bold** and Underlining indicate the best and second best results, respectively.

Models	RAIN100L	
	PSNR↑	SSIM↑
UMR	32.39	0.921
SIRR	32.37	0.926
MSPFN	33.50	0.948
LPNet	33.61	0.958
AirNet	34.90	0.977
Restormer	36.74	0.978
PromptIR	37.04	0.979
DFPIR	**39.08**	0.984
Ours	39.03	**0.989**

Models	Snow100K-S		Snow100K-L	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
MPRNet	36.22	0.973	30.63	0.938
SwinIR	33.45	0.960	28.40	0.915
Restormer	36.50	0.973	30.86	0.938
NAFNet	35.94	9.970	29.50	0.924
UniRestorer	36.94	0.974	31.14	0.940
Ours	**37.07**	**0.981**	**31.38**	**0.949**

Models	SOTS	
	PSNR↑	SSIM↑
DehazeNet	22.46	0.851
MSCNN	22.06	0.908
AODNet	20.29	0.877
EPDN	22.57	0.863
FDGAN	23.15	0.921
AirNet	23.18	0.900
Restormer	30.87	0.969
DFPIR	32.00	**0.981**
Ours	**32.03**	0.980

Table 2: Quantitative comparison in single-task setting. From left to right: deraining, desnowing, and dehazing.

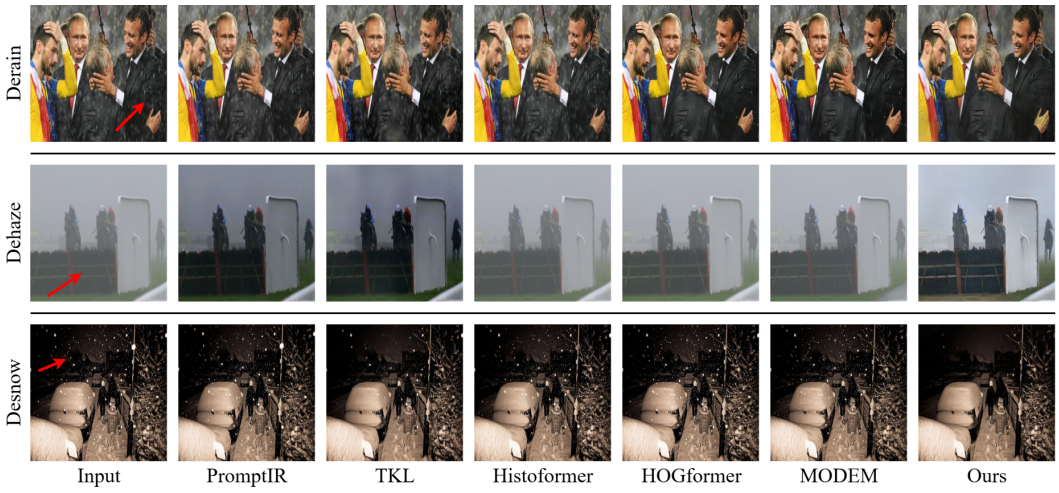

Figure 2: Qualitative comparisons of adverse weather restoration. The red arrows highlight regions where the proposed method produces superior visual results. Please zoom in for better visualization.

constraint relies more heavily on the reliable pseudo-GT. Conversely, when the degradation is mild, the degraded input already retains sufficient structural information, causing w_i to decrease and the supervision to place greater emphasis on the input image. This adaptive strategy enables the model to exploit the most reliable structural information available under varying degradation levels while avoiding excessive dependence on either source alone.

4 EXPERIMENTAL RESULTS

In this section, we present the experimental results of our method. The implementation details are presented in Appendix B.

4.1 DATASETS

Following previous works (Ye et al., 2023; Sun et al., 2024; Wang et al., 2025a), we adopt the all-weather dataset (Ye et al., 2023) as labeled dataset. Meanwhile, we leverage the unlabeled real-world adverse weather images for unsupervised learning. To achieve this, we utilize the real hazy image in URHI (Li et al., 2017b), the real rainy image in the SPA+ test set (Zhu et al., 2023), and the real snowy image from the RealSnow dataset (Zhu et al., 2023). Following (Xu et al., 2024), we employ real adverse weather image datasets for qualitative and quantitative evaluation for the unified model, including RTTS (Li et al., 2017b), DDN-SIRR (Wei et al., 2019), and Real3000 (Liu et al., 2021), and Snow100k Realistic (Liu et al., 2017). As ground truth is not available, we employ widely used non-reference metrics LIQE (Zhang et al., 2023) and CLIP-IQA (Wang et al., 2023).

For the model in the single-task setting, $i.e.$ separate models are trained for different restoration tasks, we use Rain100L (Ye et al., 2023) for rain, Snow100K-S and Snow100K-L (Liu et al., 2017) for snow, and SOTS (Li et al., 2017b) for haze. The results are reported in PSNR and SSIM.

4.2 QUANTITATIVE RESULTS

The results of competing methods are obtained either by directly citing the reported results from the original papers, reproducing the results using the authors' publicly released models, or retraining the models using their official implementations.

We compare the proposed method with recent state-of-the-art all-in-one image restoration models, including Restormer (Zamir et al., 2021), TransWeather (Ye et al., 2023), TKL (Chen et al., 2022b), WeatherDiffusion (Sun et al., 2024), WGWS-Net (Zhu et al., 2023), MWDT (Patil et al., 2023), PromptIR (Potlapalli et al., 2023), DA-CLIP (Luo et al., 2023), WResVLM (Xu et al., 2024), Histoformer (Sun et al., 2024), MODEM (Wang et al., 2025a), and HOGformer (Wu et al., 2026). The quantitative comparisons are presented in Tab. 1. Our method achieves the best overall performance, surpassing the second-best method by 0.142 in LIQE and 0.021 in CLIP-IQA. Moreover, within the family of semi-supervised methods, our approach consistently outperforms WResVLM across all evaluated weather restoration tasks. Despite these improvements, our method remains computationally efficient. Compared with MODEM, it achieves comparable model complexity with the same number of parameters while requiring only approximately one-third of the computational cost (GFLOPs).

To further evaluate the generalization ability of the proposed framework, we also compare it with representative single-task image restoration methods, including UMR (Yasarla & Patel, 2019), SIRR (Wei et al., 2019), MSPFN (Jiang et al., 2020), DehazeNet (Cai et al., 2016), MSCNN (Ren et al., 2016), AODNet (Li et al., 2017a), EPDN (Qu et al., 2019), FDGAN (Dong et al., 2020b), LPNet (Gao et al., 2019), AirNet (Li et al., 2022), Restormer (Zamir et al., 2021), PromptIR (Potlapalli et al., 2023), SwinIR (Liang et al., 2021), NAFNet (Chen et al., 2022a), UniRestorer (Lin et al., 2026), and DFPIR (Tian et al., 2025a). The results are summarized in Tab. 2. Our method achieves the highest PSNR on the desnowing task, outperforming UniRestorer by at least 0.13 dB. On the remaining restoration tasks, our method consistently delivers competitive performance, closely matching or exceeding the recent diffusion-based approach DFPIR. In this setting, the unsupervised branch of our framework is trained using the same synthetic dataset as the supervised branch with the GT stored in a reliable database.More quantitative results are provided in Appendix D.1 and D.2.

4.3 QUALITATIVE RESULTS

We qualitatively compare our method with state-of-the-art image restoration approaches on challenging real-world adverse weather images. The visual comparisons are presented in Fig. 2. As can be observed, most competing methods struggle to simultaneously remove rain, haze, and snow artifacts, often leaving noticeable degradations or introducing undesirable visual distortions. In particular, PromptIR (Potlapalli et al., 2023) and TKL (Chen et al., 2022b) partially suppress haze but

Models	H PSNR↑/SSIM↑	S PSNR↑/SSIM↑	R PSNR↑/SSIM↑	H+S PSNR↑/SSIM↑	H+R PSNR↑/SSIM↑	R*+S PSNR↑/SSIM↑	H+S+R* PSNR↑/SSIM↑
MODEM	19.73/0.8789	27.72/0.8953	23.76/0.9029	**17.87/0.7918**	20.16/0.8801	20.34/0.6734	14.23/0.4327
HOGFormer	18.87/0.8693	28.96/0.9129	23.54/0.9051	16.51/0.7635	20.30/0.8876	20.41/0.6771	13.66/0.4557
Ours	**19.89/0.8805**	**33.46/0.9568**	**24.01/0.9336**	17.27/0.7874	**21.36/0.9161**	**22.84/0.7390**	**14.45/0.5012**

Table 3: Result with composited datasets. * indicates that this degradation is synthetically added by us.

tend to produce texture distortions and over-smoothed details. In contrast, our method consistently restores clearer scene structures while preserving natural textures, resulting in superior visual quality with enhanced clarity, contrast, and overall perceptual realism. Moreover, our approach effectively removes residual haze commonly present in rainy and snowy scenes, leading to significantly improved visibility. Due to the page limit, additional visual results are provided in Appendix D.3.

4.4 Quantitative Results with Composited Dataset

Following (Wu et al., 2026), we also conduct our experiments with the composited CDD11 dataset (Guo et al., 2024), which includes several types of degradation and their combination. As the CDD11 dataset lacks some combinations of degradation that we are interested in testing, we have synthetically added them into existing combinations.

In this experiment, we compare our method with HOGformer and MODEM unified models. We have reported the test in Tab. 3 and Fig. 3, showing that our method consistently outperforms the two models in many combinations of degradation. Note that all models are not trained on the CDD11 dataset.

4.5 Ablation Studies

Impact of components To evaluate the effectiveness of the proposed components, we conduct ablation studies on three benchmark restoration tasks. In these experiments, the baseline model is the semi-supervised framework without L_{con} (Eq. 4) and phase consistency loss (Eq. 6). The average results of three tasks (desnowing, dehazing, and deraining) are reported in Tab. 4. As shown, the proposed contrastive loss consistently improves the baseline performance. Introducing the unreliable database further enhances restoration quality, with an improvement of roughly 0.7 in LIQE and 0.08 in CLIP-IQA. The proposed phase spectrum-based constraint yields an additional improvement compared with the text-based (aligning the textual embedding of the restored output and reliable pseudo-GT with LLaVA (Xu et al., 2024)) and pixel-based (LPIPS of restored output and reliable pseudo-GT) structural constraints.

Choice of NR-IQA Metric Since the proposed framework relies on NR-IQA to assess pseudo-GT quality, the choice of metric is crucial. We compare several NR-IQA methods, and the results are summarized in Tab. 5. Among all candidates, RELI consistently achieves the best performance across all restoration tasks, justifying its use in our framework.

Effectiveness of Adaptive Phase consistency loss To validate the proposed adaptive phase consistency loss, we compare it with alternative unsupervised constraints, including an L_1 loss using the reliable pseudo-GT as the supervision target and a contrastive loss using the reliable pseudo-GT and degraded input as the positive and negative samples, respectively. The results are summarized



Figure 3: Comparisons on multiple degradations (rain-haze-snow).

Baseline	*w/* L_{con}		*w/* unsupervised constraint			LIQE↑	CLIP-IQA↑
	w/o unreliable database	*w/* unreliable database	LPIPS	Text	Phase		
x						2.390	0.402
x	x					2.429	0.445
x		x				3.098	0.525
x		x	x			3.368	0.565
x		x		x		3.407	0.545
x		x			x	**3.527**	**0.572**

Table 4: Ablation study on effectiveness of proposed components. We report the average results of three tasks (desnowing, dehazing, and deraining).

NR-IQA	Rain		Snow		Haze	
	LIQE↑	CLIP-IQA↑	LIQE↑	CLIP-IQA↑	LIQE↑	CLIP-IQA↑
LIQE	1.029	0.418	1.013	0.389	1.032	0.453
MUSIQ	1.774	0.469	2.507	0.532	1.329	0.420
NIMA	2.194	0.516	2.228	<u>0.610</u>	<u>2.995</u>	0.468
CLIP-IQA	<u>2.445</u>	<u>0.553</u>	<u>3.120</u>	0.603	1.696	**0.508**
RALI	**3.466**	**0.597**	**3.852**	**0.611**	**3.263**	<u>0.507</u>

Table 5: Performance comparison of different NR-IQA methods.

Degradation types	Metrics	Choices of L_{sc}		
		L1	Contrastive	Ours
Rain	LIQE↑	3.302	<u>3.380</u>	**3.466**
	CLIP-IQA↑	0.592	<u>0.593</u>	**0.597**
Snow	LIQE↑	3.742	**3.862**	<u>3.852</u>
	CLIP-IQA↑	0.596	**0.612**	<u>0.611</u>
Haze	LIQUE↑	3.136	<u>3.192</u>	**3.263**
	CLIP-IQA↑	<u>0.504</u>	0.489	**0.507**
Average	LIQE↑	3.394	<u>3.478</u>	**3.527**
	CLIP-IQA↑	0.564	<u>0.565</u>	**0.572**

Table 6: Performance comparison with different L_{phase}.

in Tab. 6. As shown, even though the contrastive loss performs better in the desnowing task, the proposed adaptive phase consistency loss consistently achieves the best overall performance.

Further ablation studies presented in the Appendix include phase visualization (C.1), overhead comparison with other SOTAs (C.2), overhead comparison with other constraint methods (C.3), ablation studies on backbone models (C.4) and adaptive weight (C.5), and impact of unreliable data (C.6).

5 Conclusion

This paper presented an unreliable-aware semi-supervised learning framework for real-world adverse weather image restoration. The proposed framework combines an unreliable bank for exploiting failed pseudo-GTs as hard negative samples with an adaptive phase consistency loss that replaces expensive vision-language model-based semantic supervision. Together, these components effectively reduce confirmation bias while providing efficient structural guidance during training. Future work will investigate extending the implementation.

AI USE STATEMENT

In this work, we used generative AI tools for cleaning and reformatting the dataset, grammar checking. We have not used generative AI tools to generate synthetic data sets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, assist with translation, support qualitative and interpret results, assist in the writing of proofs and polish the manuscript. Thematic data analysis are not applicable to this work Additionally, we used generative AI tools to propose a title or keywords for a research paper. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Codruta Orniana Ancuti and Cosmin Ancuți. Single image dehazing by multi-scale fusion. *IEEE Transactions on Image Processing*, 22:3271–3282, 2013. URL https://api.semanticscholar.org/CorpusID:14713713.

Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25:5187–5198, 2016. URL https://api.semanticscholar.org/CorpusID:14092238.

Liangyu Chen, Xiaojie Chu, X. Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision*, 2022a. URL https://api.semanticscholar.org/CorpusID:248085491.

Wei-Ting Chen, Zhi-Kai Huang, Cheng-Che Tsai, Hao-Hsiang Yang, Jianping Ding, and Sy-Yen Kuo. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17632–17641, 2022b. URL https://api.semanticscholar.org/CorpusID:250581068.

Xiang Chen, Hao-Ran Li, Mingqiang Li, and Jin shan Pan. Learning a sparse transformer network for effective image deraining. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5896–5905, 2023. URL https://api.semanticscholar.org/CorpusID:257636509.

Hang Dong, Jin shan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2154–2164, 2020a. URL https://api.semanticscholar.org/CorpusID:216562731.

Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing. *ArXiv*, abs/2001.06968, 2020b. URL https://api.semanticscholar.org/CorpusID:210838848.

Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6):2944–2956, 2017. ISSN 1941-0042. doi: 10.1109/tip.2017.2691802. URL http://dx.doi.org/10.1109/TIP.2017.2691802.

Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia. Dynamic scene deblurring with parameter selective sharing and nested skip connections. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3843–3851, 2019. URL https://api.semanticscholar.org/CorpusID:198185751.

Xin Guo, Xueyang Fu, Man Zhou, Zhen Huang, Jialun Peng, and Zhengjun Zha. Exploring fourier prior for single image rain removal. In *International Joint Conference on Artificial Intelligence*, 2022. URL https://api.semanticscholar.org/CorpusID:250638879.

Yu Guo, Yuan Gao, Yuxu Lu, Ryan Wen Liu, and Shengfeng He. Onerestore: A universal restoration framework for composite degradation. In *European Conference on Computer Vision*, 2024.

Shirui Huang, Keyan Wang, Huan Liu, Jun Chen, and Yunsong Li. Contrastive semi-supervised learning for underwater image restoration via reliable bank. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18145–18155, 2023. URL https://api.semanticscholar.org/CorpusID:257557240.

Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8343–8352, 2020. URL https://api.semanticscholar.org/CorpusID:214623217.

Tai-Xiang Jiang, Tingzhu Huang, Xile Zhao, Liang-Jian Deng, and Yao Wang. A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2827, 2017. URL https://api.semanticscholar.org/CorpusID:4910523.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5128–5137, 2021. URL https://api.semanticscholar.org/CorpusID:237048383.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. URL https://api.semanticscholar.org/CorpusID:6628106.

Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4780–4788, 2017a. URL https://api.semanticscholar.org/CorpusID:30151664.

Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28:492–505, 2017b. URL https://api.semanticscholar.org/CorpusID:39760169.

Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xiaocui Peng. All-in-one image restoration for unknown corruption. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17431–17441, 2022. URL https://api.semanticscholar.org/CorpusID:250551851.

Jingyun Liang, Jie Cao, Guolei Sun, K. Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 1833–1844, 2021. URL https://api.semanticscholar.org/CorpusID:237266491.

Jingbo Lin, Zhilu Zhang, Wenbo Li, Renjing Pei, Hang Xu, Hongzhi Zhang, and Wangmeng Zuo. Unirestorer: Universal image restoration via adaptively estimating image degradation at proper granularity. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=nDrZow7fCF.

Chengxu Liu, Lu Qi, Jinshan Pan, Xueming Qian, and Ming-Hsuan Yang. Frequency domain-based diffusion model for unpaired image dehazing. *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7538–7547, 2025a. URL https://api.semanticscholar.org/CorpusID:280135104.

Fuyang Liu, Jiaqi Xu, and Xiaowei Hu. Real-world adverse weather image restoration via dual-level reinforcement learning with high-quality cold start. *Conference on Neural Information Processing Systems*, abs/2511.05095, 2025b. URL https://api.semanticscholar.org/CorpusID:282890021.

Yang Liu, Ziyu Yue, Jinshan Pan, and Zhixun Su. Unpaired learning for deep image deraining with rain direction regularizer. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4733–4741, 2021. doi: 10.1109/ICCV48922.2021.00471.

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*, 27:3064–3073, 2017. URL https://api.semanticscholar.org/CorpusID:837707.

Ziwei Luo, Fredrik K. Gustafsson, Zheng Zhao, Jens Sjolund, and Thomas B. Schon. Controlling vision-language models for multi-task image restoration. In *International Conference on Learning Representations*, 2023. URL https://api.semanticscholar.org/CorpusID:263605463.

Alan V. Oppenheim and Jae Sung Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69:529–541, 1980. URL https://api.semanticscholar.org/CorpusID:23790114.

Alan V. Oppenheim, Jae Sung Lim, Gary E. Kopec, and Stephen C. Pohlig. Phase in speech and pictures. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1979. URL https://api.semanticscholar.org/CorpusID:23520948.

Prashant W. Patil, Sunil Gupta, Santu Rana, Svetha Venkatesh, and Subrahmanyam Murala. Multi-weather image restoration via domain translation. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 21639–21648, 2023. doi: 10.1109/ICCV51070.2023.01983.

Vaishnav Potlapalli, Syed Waqas Zamir, Salman Siddique Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one blind image restoration. *ArXiv*, abs/2306.13090, 2023. URL https://api.semanticscholar.org/CorpusID:259224666.

Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8152–8160, 2019. URL https://api.semanticscholar.org/CorpusID:195489762.

Wenqi Ren, Sibo Liu, Hua Zhang, Jin shan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*, 2016. URL https://api.semanticscholar.org/CorpusID:17763780.

S M A Sharif, Abdur Rehman, Fayaz Ali, Radu Timofte, and Rizwan Ali Naqvi. Degradation-aware all-in-one image restoration via latent prior encoding, 2026. URL https://openreview.net/forum?id=4fhbxglvYs.

Liang Shen, Zihan Yue, Quan Chen, Fan Feng, and Jie Ma. Deep joint rain and haze removal from single images. *ArXiv*, abs/1801.06769, 2018. URL https://api.semanticscholar.org/CorpusID:32784402.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. URL https://api.semanticscholar.org/CorpusID:14124313.

Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin Dogus Cubuk, Alexey Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Conference on Neural Information Processing Systems*, abs/2001.07685, 2020. URL https://api.semanticscholar.org/CorpusID:210839228.

Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023.

Shangquan Sun, Wenqi Ren, Xinwei Gao, Rui Wang, and Xiaochun Cao. Restoring images in adverse weather conditions via histogram transformer. In *European Conference on Computer Vision*, 2024. URL https://api.semanticscholar.org/CorpusID:271212187.

Yunhao Tang, Daniel Zhaohan Guo, Zeyu Zheng, Daniele Calandriello, Yuan Cao, Eugene Tarassov, Rémi Munos, Bernardo Ávila Pires, Michal Valko, Yong Cheng, and Will Dabney. Understanding the performance gap between online and offline alignment algorithms, 2024. URL https://arxiv.org/abs/2405.08448.

Yunyi Tang, Bowei Fang, Chunyu Zhao, Haoran Liu, Fei Yan, and Tao Deng. Adverse effect removal network via unsupervised weather type transfer. In *ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 10057–10061, 2026. doi: 10.1109/ICASSP55912.2026.11460995.

Xiangpeng Tian, Xiangyu Liao, Xiao Liu, Meng Li, and Chao Ren. Degradation-aware feature perturbation for all-in-one image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 28165–28175, 2025a.

Yunjie Tian, Qixiang Ye, and David Doermann. Yolov12: Attention-centric real-time object detectors, 2025b. URL https://arxiv.org/abs/2502.12524.

Hainuo Wang, Qiming Hu, and Xiaojie Guo. Modem: A morton-order degradation estimation mechanism for adverse weather image recovery. *Conference on Neural Information Processing Systems*, abs/2505.17581, 2025a. URL https://api.semanticscholar.org/CorpusID:278886994.

Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *AAAI*, 2023.

Ruiyi Wang, Yushuo Zheng, Zicheng Zhang, Chunyi Li, Shuaicheng Liu, Guangtao Zhai, and Xiaohong Liu. Learning hazing to dehazing: Towards realistic haze generation for real-world image dehazing. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23091–23100, 2025b. URL https://api.semanticscholar.org/CorpusID:277313707.

Yan Wang and Bo Wu. Improved single image dehazing using dark channel prior. *2010 IEEE International Conference on Intelligent Computing and Intelligent Systems*, 2:789–792, 2010. URL https://api.semanticscholar.org/CorpusID:1294867.

Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *The IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Chunyi Li, Liang Liao, Annan Wang, Erli Zhang, Wenxiu Sun, Qiong Yan, Xiongkuo Min, Guangtai Zhai, and Weisi Lin. Q-align: Teaching lmms for visual scoring via discrete text-defined levels. *arXiv preprint arXiv:2312.17090*, 2023. Equal Contribution by Wu, Haoning and Zhang, Zicheng. Project Lead by Wu, Haoning. Corresponding Authors: Zhai, Guangtai and Lin, Weisi.

Haoning Wu, Zicheng Zhang, Erli Zhang, Chaofeng Chen, Liang Liao, Annan Wang, Chunyi Li, Wenxiu Sun, Qiong Yan, Guangtao Zhai, and Weisi Lin. Q-bench: A benchmark for general-purpose foundation models on low-level vision. *International Conference on Learning Representations*, abs/2309.14181, 2024. URL https://api.semanticscholar.org/CorpusID:262824606.

Jiawei Wu, Zhifei Yang, Zhe Wang, and Zhi Jin. Gradient as conditions: Rethinking hog for all-in-one image restoration. In *AAAI Conference on Artificial Intelligence*, 2026. URL https://api.semanticscholar.org/CorpusID:277781355.

Jiaqi Xu, Mengyang Wu, Xiaowei Hu, Chi-Wing Fu, Qi Dou, and Pheng-Ann Heng. Towards real-world adverse weather image restoration: Enhancing clearness and semantics with vision-language models. In *European Conference on Computer Vision*, 2024. URL https://api.semanticscholar.org/CorpusID:272368249.

Yang Yang, Chaoyue Wang, Xiaojie Guo, and Dacheng Tao. Robust unpaired image dehazing via density and depth decomposition. *International Journal of Computer Vision*, 132:1557 – 1577, 2023. URL https://api.semanticscholar.org/CorpusID:265452111.

Yang Yang, Chun-Le Guo, and Xiaojie Guo. Depth-aware unpaired video dehazing. *IEEE Transactions on Image Processing*, 33:2388–2403, 2024. URL https://api.semanticscholar.org/CorpusID:268635792.

Rajeev Yasarla and Vishal M. Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8397–8406, 2019. URL https://api.semanticscholar.org/CorpusID:195657934.

Tian Ye, Sixiang Chen, Jinbin Bai, Jun Shi, Chenghao Xue, Jingxia Jiang, Junjie Yin, Erkang Chen, and Yun Liu. Adverse weather removal with codebook priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12653–12664, October 2023.

Syed Waqas Zamir, Aditya Arora, Salman Hameed Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5718–5729, 2021. URL https://api.semanticscholar.org/CorpusID:244346144.

Ting Zhang, Nanfeng Jiang, Junhong Lin, Jielian Lin, and Tiesong Zhao. Desnowformer: an effective transformer-based image desnowing network. *2022 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pp. 1–5, 2022. URL https://api.semanticscholar.org/CorpusID:255995928.

Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 14071–14081, 2023.

Shijie Zhao, Xuanyu Zhang, Weiqi Li, Junlin Li, Li Zhang, Tianfan Xue, and Jian Zhang. Reasoning as representation: Rethinking visual reinforcement learning in image quality assessment. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2026.

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 *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

A MOTIVATION FURTHER ANALYSIS

As described in the Introduction of the main paper, the degradation information is manifested by the amplitude spectrum, while the texture structure information is manifested by the phase spectrum, which is highly immune to degradation. That is to say, the semantic information is well-preserved in the phase spectrum of the degraded input image. Therefore, our key motivation is to leverage such semantic information to constrain the unsupervised learning process.

To justify this motivation, we follow the assumption of the classical dark channel prior (Wang & Wu, 2010; Shen et al., 2018), and analyze it by studying the statistical properties of the dark channel before and after the exchange of the phase spectrum. More specifically, as shown in Fig. 5, we replace the phase spectrum of the clean images ($i.e.$, Clean) with the phase spectrum of the degraded image ($i.e.$, Degraded), obtaining a set of synthetic clean images ($i.e.$, SynClean). Then we follow (Li et al., 2017b; Liu et al., 2025a) using the patch with the size of 15×15 to calculate their dark channels.

Fig. 4 depicts multiple example images and the corresponding dark channels. Figs. 4(a), (c), and (e) are the Degraded images, the Clean images, and the SynClean images, respectively. Figs. 4(b), (d), and (f) are their corresponding dark channels. Visually, the dark channels of the SynClean image are closer to the dark channels of the Clean image, despite replacing the phase spectrum of the Clean image with the phase spectrum of the Degraded image. This proves that semantic information is well preserved inside the phase spectrum of the Degraded image.

To further support our motivation, we follow (Liu et al., 2025a) and report the histogram of the pixel intensities in all the dark channels of the Degraded images, the Clean images, and the SynClean images in Figs. 6(a), (b), and (c), respectively.

In sum, this unique frequency domain property provides inspiration for adverse weather removal.

B IMPLEMENTATION DETAILS

We train our method in two settings: all-in-one and single-task, with the same setup. Our semi-supervised learning framework is easily compatible with various image restoration networks. Following (Xu et al., 2024), we opt for MSBDN (Dong et al., 2020a) as our backbone due to its balanced performance and rapid inference speed in our main study. For the NR-IQA method, we incorporate the SOTA RALI score (Zhao et al., 2026). Each batch comprises 8 labeled and 8 unlabeled images, and the number of epochs is set to 200. During training, we used the Adam optimizer (Kingma & Ba, 2014) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and the learning rate $2e-5$. All training was conducted on $2 \times H100$.

C ABLATION STUDIES

C.1 SPECTRAL VISUALIZATIONS

To prove the reliability of our model, we present the spectral visualizations in Fig. 7 on CDD11 dataset. Our model's results are more aligned with the ground truth than results from HOGformer. In addition, we measure the L2 distance ($\downarrow$) of the frequency features with GT to show the superiority and reported in Tab. 7. Compared with HOGFormer, our method substantially reduces both phase and amplitude errors across all three degradation types. The largest improvements are observed for rain and haze, where the phase error decreases from 52.645 to 33.735 and 57.057 to 32.623, respectively. Amplitude error is also consistently reduced, with the average decreasing from 11.184 to 8.111. Overall, our method lowers the average phase error from 57.453 to 41.667, demonstrating better preservation and recovery of both structural phase information and frequency amplitude characteristics across different degradations.

C.2 OVERHEAD COMPARISON WITH OTHER SOTAS

We have conducted an overhead comparison with HOGformer (Wu et al., 2026) and MODEM (Wang et al., 2025a) with the input of 256×256 on a single H100 GPU. The results, shown in

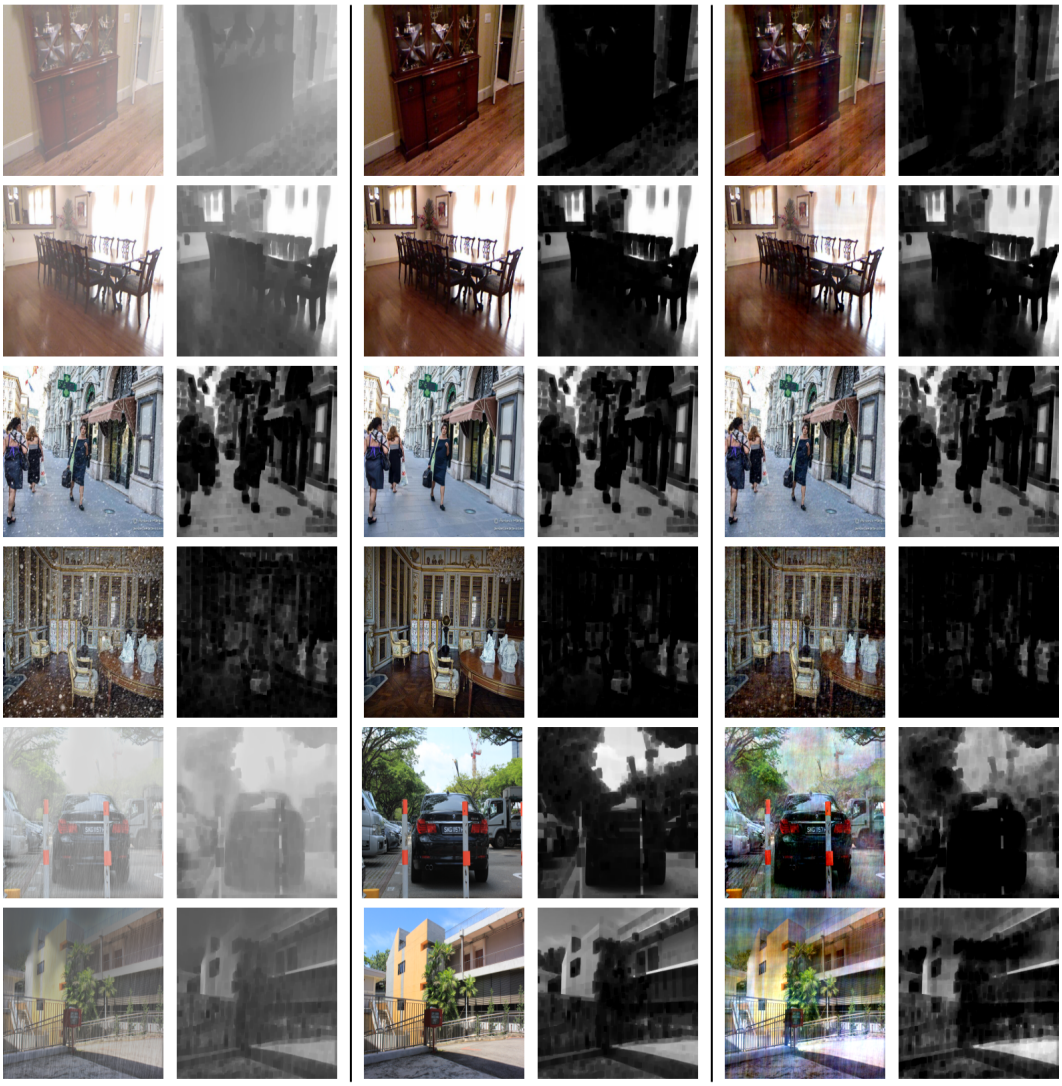

Figure 4: Example images. (a), (c), and (e) are the Degraded images, the Clean images, and the SynClean images, respectively. (b), (d), and (f) are their corresponding dark channels. The dark channel of the SynClean image is more similar to the Clean image indicating that the phase spectrum preserve semantic information well through degradation process.

Degradation type	Methods			
	HOGFormer		Ours	
	Phase	Amplitude	Phase	Amplitude
Rain	52.645	12.331	**33.735**	**8.836**
Snow	62.657	9.616	**58.644**	**8.953**
Haze	57.057	11.604	**32.623**	**6.545**
Avg	57.453	11.184	**41.667**	**8.111**

Table 7: Average L2↓ distance of the frequency features with GT on CDD11 dataset.

Tab. 8, indicate that our model provides a favorable efficiency profile compared with existing approaches. While using a comparable number of trainable parameters to MODEM and HOGFormer,

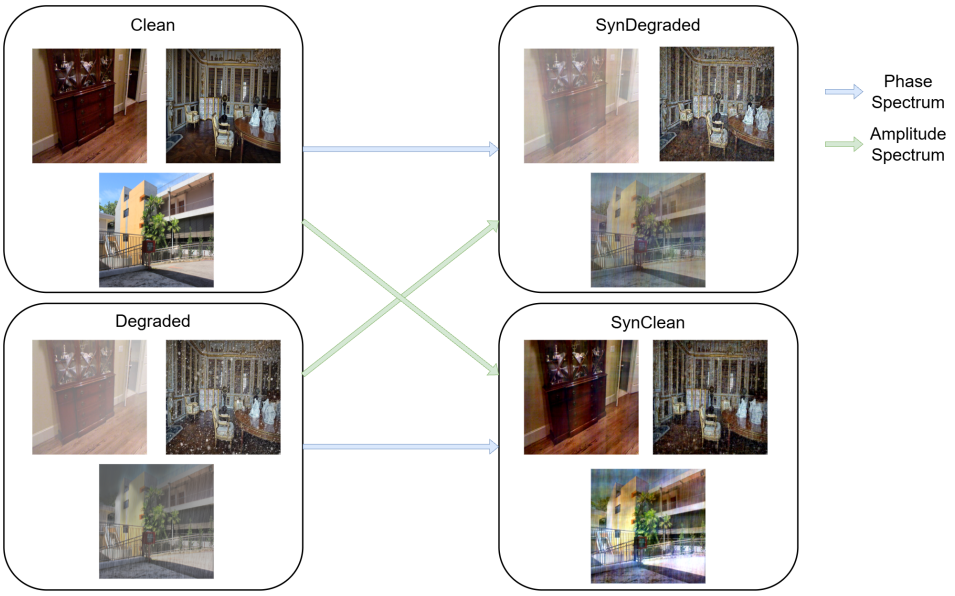

Figure 5: Illustration of the property of adverse weather conditions in the frequency domain.

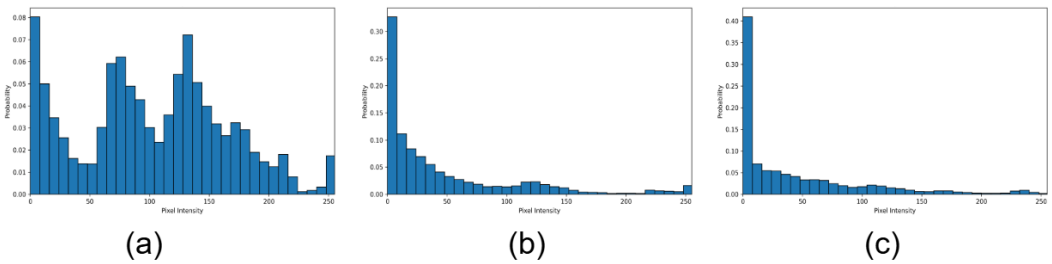

Figure 6: Statistics of the dark channels. (a), (b), and (c) are the histograms of the pixel intensities in all of the dark channels of the Degraded images, the Clean images, and the SynClean images, respectively.

it achieves substantially lower inference latency at only 21.20 ms, compared with 100.83 ms and 218.82 ms, respectively. It also requires the least GPU memory (497.08 MB), despite having a larger overall model size of 225 MB. These results indicate that our method is particularly efficient at inference time, offering faster processing and lower runtime memory consumption while maintaining competitive parameter complexity.

Method	Trainable Param M	Latency ms	GPU Mem MB	Model Size MB
MODEM	19.96	100.83	551.21	77
HOGformer	16.65	218.82	760.43	65
Ours	19.67	21.20	497.08	225

Table 8: Inference overhead comparison with other SOTAs. Input size of 256×256 on a single H100.

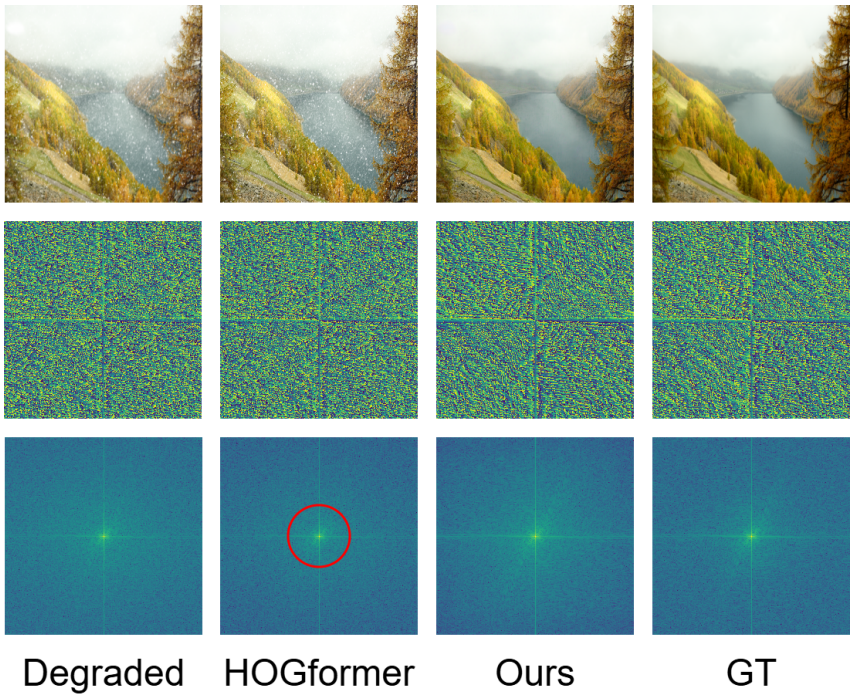

Figure 7: Visual results on CDD11 dataset. From top to bottom: images, phase spectra, and amplitude spectra. Zoom in to see better visualization.

Method	Latency ms	GPU mem MB	Average Performance LIQE↑	CLIP-IQA↑
Using LPIPS	2.14	1774.66	3.368	0.565
Using Text	1198.14	13846.31	3.407	0.545
Using Phase	41.00	2894.09	3.527	0.572

Table 9: Overhead comparison with other constraint methods. Input size of 256×256 on a single A100.

C.3 Overhead comparison with other unsupervised constraint methods

We have conducted an overhead comparison with LPIPS and text (using VLM) with the input of 256×256 on a single A100 GPU. In this experiment, we want to capture the GPU usage and latency that models take to produce a semantic signal from the input image. The results, shown in Tab. 8, indicate that our model achieves the best tradeoff of inference time and performance. The average performance of three tasks (deraining, desnowing, and dehanzing) is also cited from the main paper. Compared with the text-based constraint, the phase-based approach reduces latency from 1198.14 ms to 41.00 ms and GPU memory consumption from 13,846.31 MB to 2,894.09 MB, while achieving better performance on both LIQE (3.527) and CLIP-IQA (0.572). Although LPIPS introduces the lowest computational overhead, its average LIQE score is lower. Overall, the phase-based constraint achieves the strongest perceptual performance while remaining substantially more efficient than the text-based alternative.

C.4 Impact of different backbones

We evaluate our proposed methods with different backbone models and report the results in Tab. 10. The comparison across different backbones shows that the proposed framework can be effectively integrated with multiple restoration architectures. Among them, NAFNet achieves the best perceptual performance, reaching 3.545 LIQE and 0.574 CLIP-IQA, but at the cost of substantially higher computational complexity with 67.89M trainable parameters and 290.84G FLOPs. Uformer is con-

Backbone	Trainable Param M	FLOP G	Average Performance	
			LIQE↑	CLIP-IQA↑
NAFNet	67.89	290.84	**3.545**	**0.574**
Uformer	5.29	42.74	3.268	0.497
MSBDN	19.67	153.94	<u>3.527</u>	<u>0.572</u>

Table 10: Results of different backbone models. We report the average performance on three tasks using the dataset presented in the main paper.

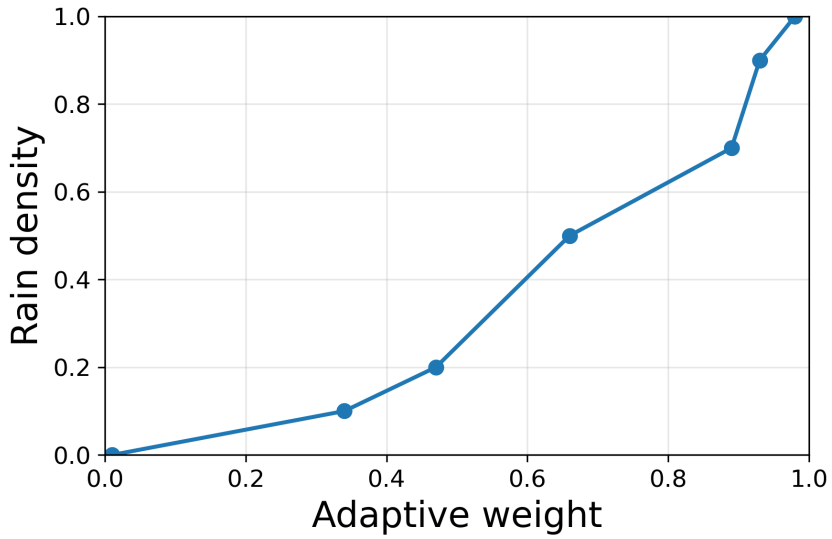

Figure 8: Relationship between adaptive weight and rain density. Higher rain densities are assigned larger adaptive weights, enabling the model to place greater emphasis on pseudo-GT.

siderably lighter but yields lower restoration quality. MSBDN provides a more balanced configuration, achieving 3.527 LIQE and 0.572 CLIP-IQA with fewer parameters and FLOPs than NAFNet. These results indicate that the choice of backbone allows a flexible trade-off between computational cost and perceptual restoration performance.

C.5 Abalation study on adaptive weight w

Figure 8 illustrates how the adaptive weight varies with rain density. The weight increases monotonically as the degradation becomes more severe. For light rain, the weight remains relatively low, while a sharper increase is observed at higher densities, particularly beyond a density of 0.7. This behavior allows the model to adapt its restoration strength according to degradation severity. In this experiment, we use synthetic rain so that we can control the rain density.

C.6 Impact of unreliable data

We further validate the effectiveness of the unreliable database in Fig. 9. Compared with the baseline, the teacher network produces more reliable pseudo-GTs with fewer artifacts and distortions, resulting in improved supervision for the student.

D More Experimental Results

In this section, we first provide results on downstream tasks and more quantitative results. Then, we show more qualitative results. All of these experiments are conducted in the all-in-one setting.

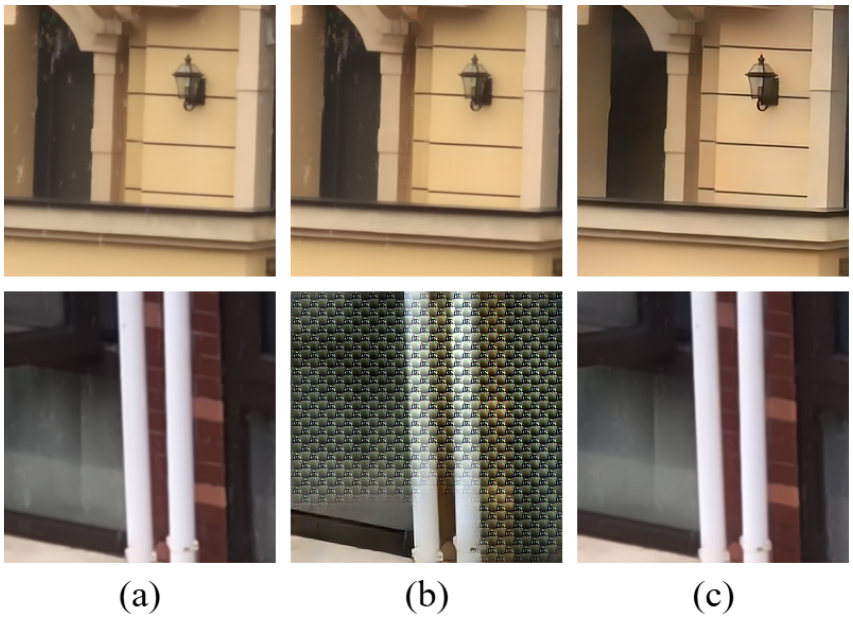

Figure 9: (a) Input image. Output of teacher model: (b) without unreliable database; (c) with unreliable database.

Method	AP			AP by Object Size		
	IoU=0.5:0.95	IoU=0.5	IoU=0.75	Small	Medium	Large
Uformer	29.6	31.3	29.9	37.5	27.3	27.9
AirNet	31.0	32.9	32.1	36.1	26.6	30.7
PromptIR	31.2	33.7	32.3	42.1	**29.9**	31.1
DFPIR	28.8	30.9	29.8	27.4	26.4	30.1
Ours	**35.3**	**38.7**	**36.6**	**44.4**	**29.9**	**35.3**

Table 11: Performance comparison of object detection models. We report average precision (AP) across different IoU thresholds and object sizes.

D.1 Downstream task results - Object Detection (OD)

To evaluate our model on downstream tasks, we follow the setting of (Sharif et al., 2026). We construct a comprehensive evaluation framework using 1,000 test images spanning three common image degradations: deraining (Fu et al., 2017), dehazing (Li et al., 2017b), and desnowing (Liu et al., 2017). Leveraging the paired nature of these datasets, we first annotate the clean ground-truth images and transfer the annotations to their corresponding degraded counterparts, ensuring consistent and reliable labels. We then restore the degraded images using several baseline all-in-one image restoration methods and evaluate their downstream object detection performance using YOLOv12-Large (Tian et al., 2025b).

Tab. 11 shows that our method achieves the strongest overall performance across all evaluation metrics. It improves the overall AP to 35.3, outperforming the best baseline, PromptIR Potlapalli et al. (2023), by 4.1 points. Similar gains are observed at different IoU thresholds, reaching 38.7 AP_{50} and 36.6 AP_{75}, which indicates improvements in both detection accuracy and localization quality. For object-scale evaluation, our method obtains the highest performance on small (44.4) and large (35.3) objects, while matching the best result on medium-sized objects (29.9). These results demonstrate that the proposed method provides consistent improvements across different IoU criteria and object scales.

Degradation types	Metrics	Methods	
		HOGformer	Ours
Rain	Q-Align↑	2.8022	**2.8372**
	MUSIQ↑	53.958	**55.776**
Snow	Q-Align↑	**2.7557**	2.6678
	MUSIQ↑	**54.947**	52.336
Haze	Q-Align↑	2.4597	**2.5241**
	MUSIQ↑	50.438	**51.710**
Average	Q-Align↑	2.6725	**2.6764**
	MUSIQ↑	53.114	**53.274**

Table 12: Quantitative results on different metrics.

D.2 More quantitative results with different metrics

To further prove the effectiveness of our all-in-one model, we have conducted a comparison with HOGformer (Wu et al., 2026) in two different no-reference metrics, Q-Align (Wu et al., 2023) and MUSIQ (Ke et al., 2021). The results, which are reported in Tab. 12, indicate that our method achieves the best average performance across both perceptual quality metrics, improving Q-Align from 2.6725 to 2.6764 and MUSIQ from 53.114 to 53.274. In particular, our method consistently outperforms HOGFormer on both metrics for rain degradation, while also providing clear improvements for haze restoration. For snow degradation, HOGFormer performs better, suggesting that our method may be less effective for this specific corruption type. Nevertheless, the higher overall average indicates that our approach provides more balanced restoration quality across different degradation conditions. We use the same dataset as the main paper.

D.3 More qualitative results

To further verify the effectiveness of our method, we show more comparison results among the proposed model and other advanced methods. Figs. 10, 11, and 12 depict the visual results on test dataset.

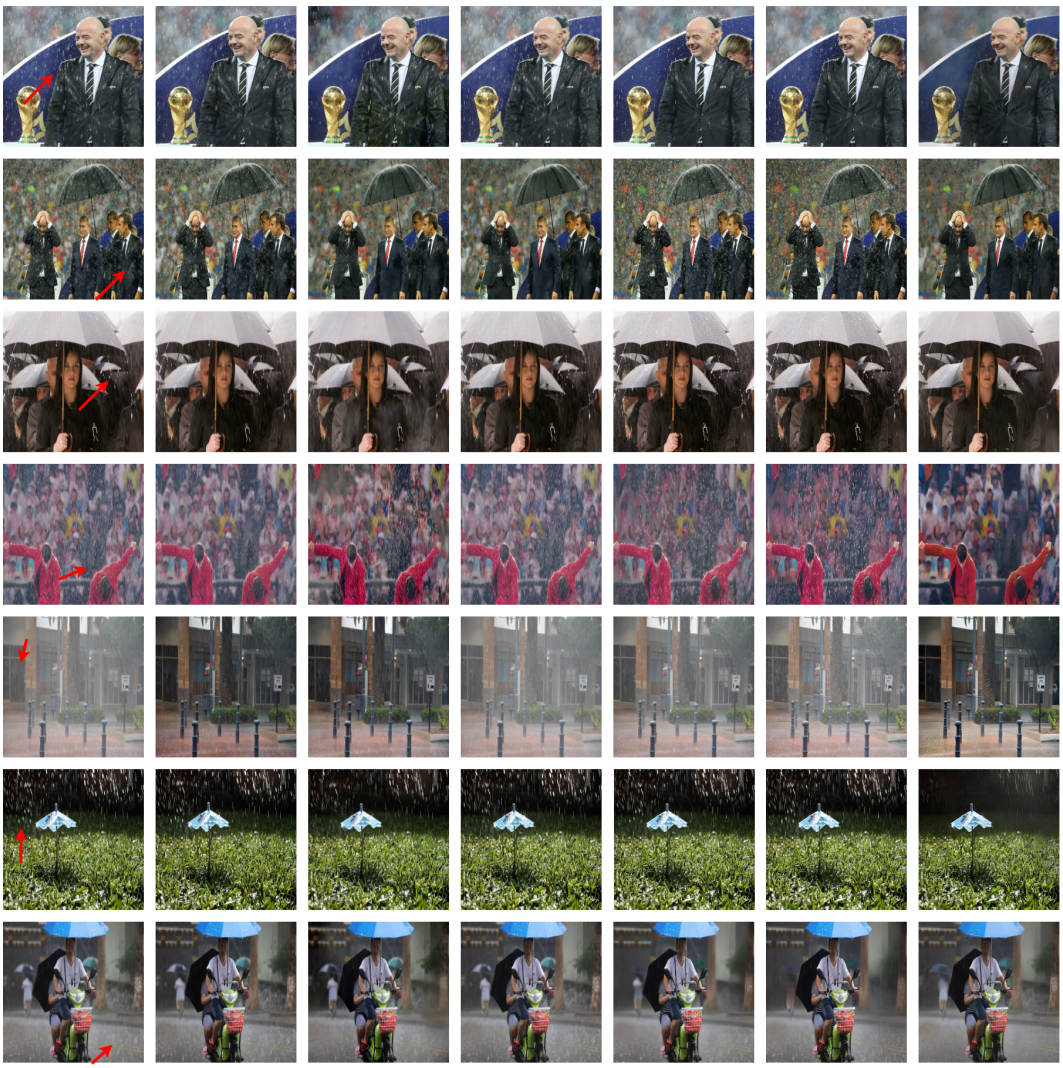

Figure 10: Qualitative comparisons of deraining. Please zoom in for better visualization.

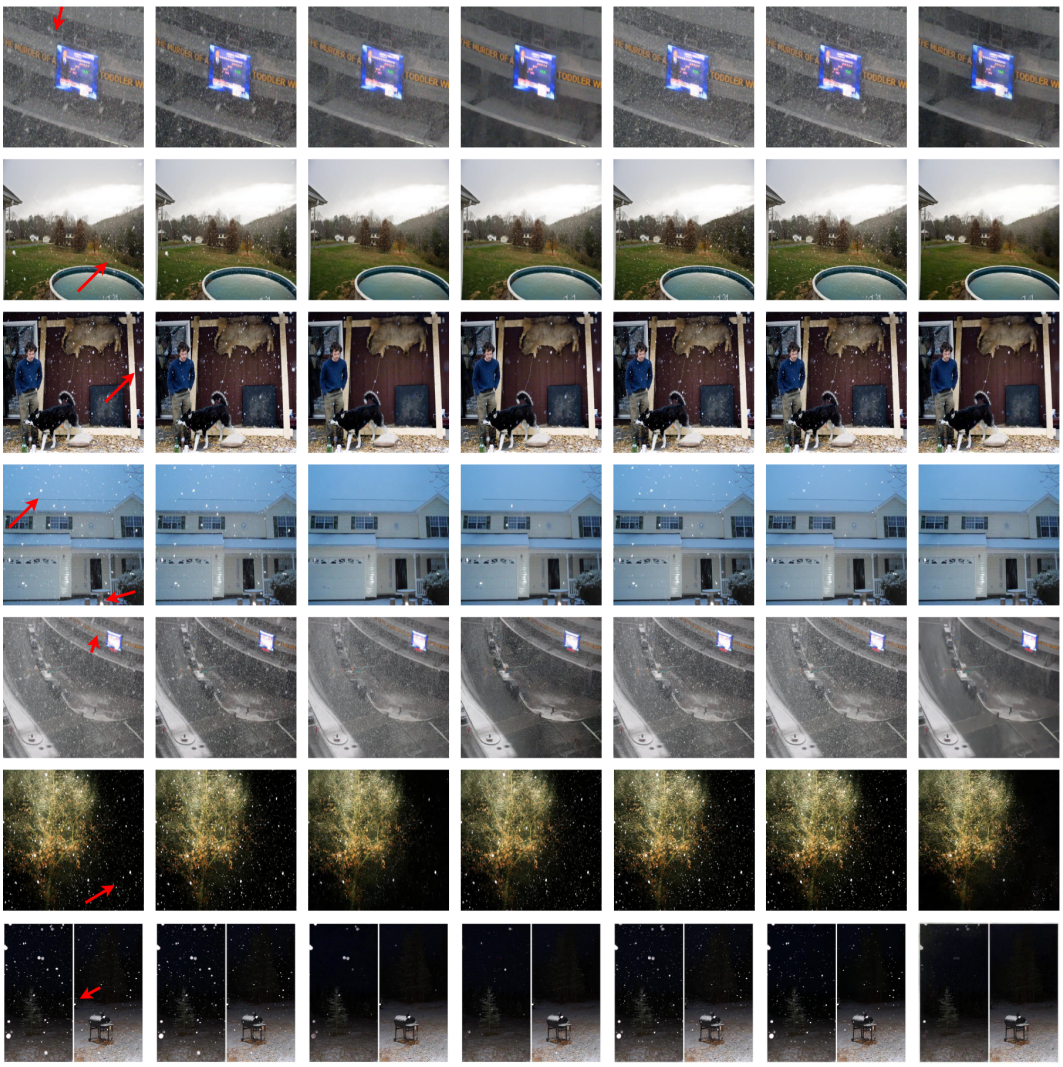

Figure 11: Qualitative comparisons of desnowing. Please zoom in for better visualization.

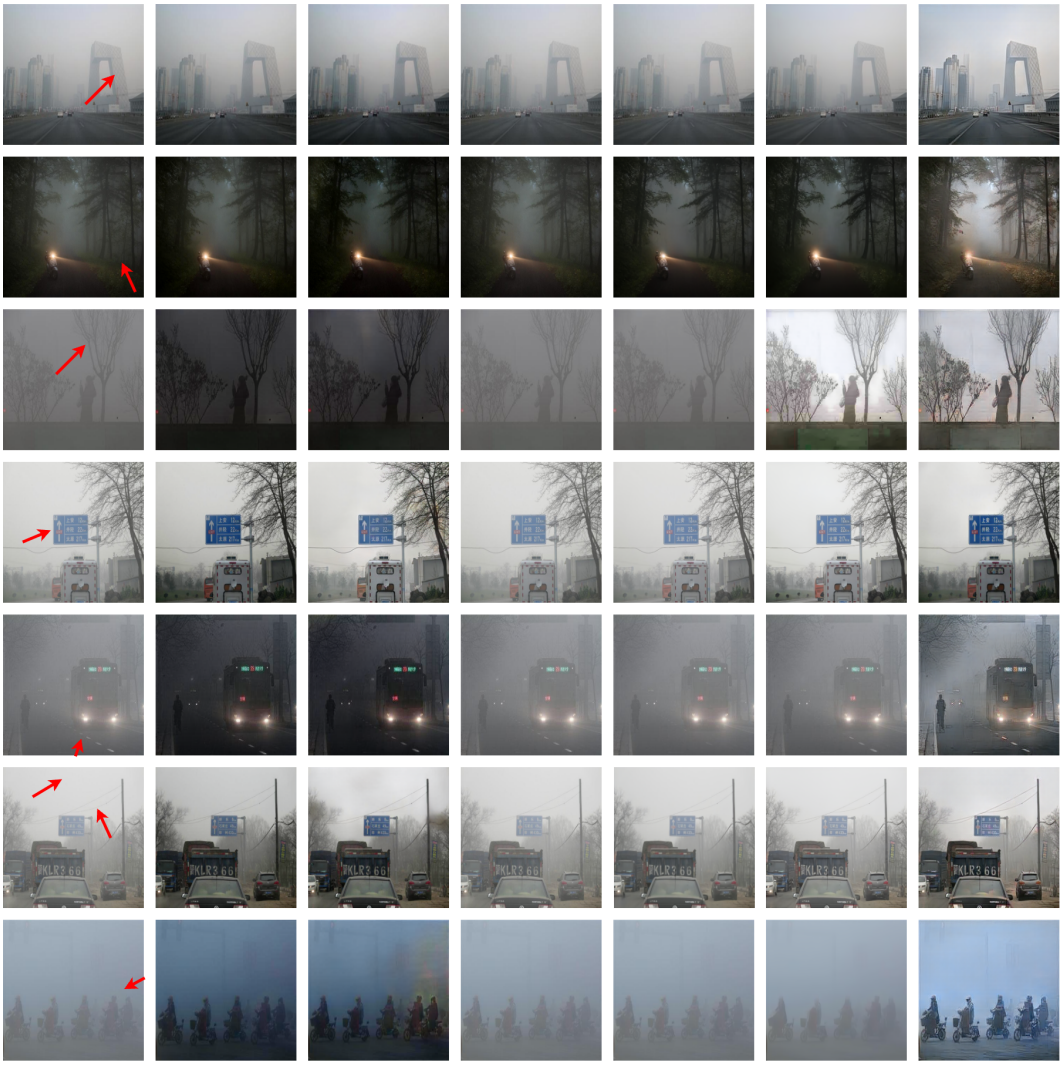

Figure 12: Qualitative comparisons of dehazing. Please zoom in for better visualization.