Learning to Remove Coupled Rain and Mist from Single Degradation Priors

Yan Zhang[1] , **Yuxin Feng**[2] , **Zhe Huang**[1] , **Fan Zhou**[1,3] and **Zhuo Su**[1,3,*]
[1]School of Computer Science and Engineering, National Engineering Research Center of Digital Life, Sun Yat-sen University
[2]Hangzhou Institute of Technology, Xidian University
[3]Research Institute of Sun Yat-sen University in Shenzhen
suzhuo3@mail.sysu.edu.cn

Abstract

Existing image deraining models struggle to handle complex real-world scenarios with rain-mist coupled degradation. The core challenge stems from the high visual similarity and spatial coupling between rain streaks and mist, making it difficult to accurately model their joint degradation patterns. Consequently, both specialized deraining models and all-in-one models for multiple degradations are ineffective in rain-mist coexistence scenarios. To address these challenges, we propose a novel **R**ain-**M**ist **R**emoval (**RMR**) framework. It effectively utilizes single degradation priors from existing deraining and dehazing datasets to model the joint degradation, thereby achieving effective rain streak removal while preserving background structures obscured by mist. To enhance the generalization to real-world scenarios, we leverage text prompts trained in the CLIP perceptual space to drive the generated results toward real samples. Extensive experiments demonstrate that the proposed RMR outperforms state-of-the-art methods in rain-mist coexistence scenarios.

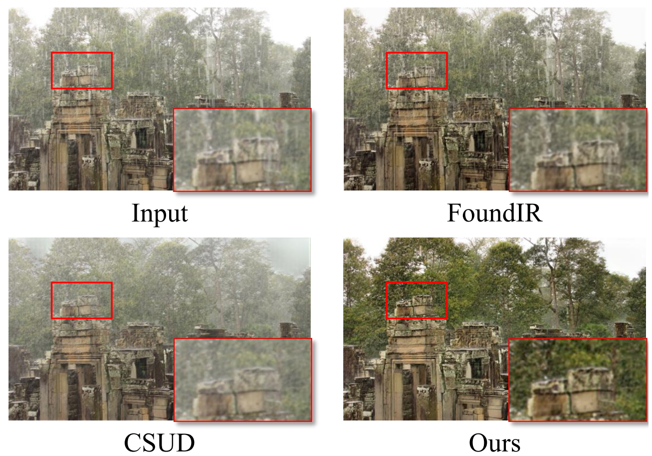


Figure 1: Deraining results in the rain-mist coexistence scenarios from SPA dataset [Wang *et al.*, 2019]. Compared to the unsupervised method CSUD [Dong *et al.*, 2025] and the all-in-one method FoundIR [Li *et al.*, 2025b], our method effectively eliminates the rain-mist coupled degradation and achieves the best visual quality.

1 Introduction

Image deraining is a foundational low-level vision task focused on removing rain streaks to restore clean images, essential for subsequent high-level applications [Chen *et al.*, 2025]. Existing deraining algorithms have demonstrated excellent performance on various synthetic benchmark datasets [Chen *et al.*, 2024a; Wang *et al.*, 2024]. However, most of these models are designed and evaluated under an idealized assumption that rain is the sole degradation factor in images [Yang *et al.*, 2017; Zhang and Patel, 2018].

This simplified assumption significantly diverges from the authentic rainy environments. In real-world scenarios, rain and mist (or haze) are not independent. Instead, they often coexist and form a complex coupled degradation pattern [Li *et al.*, 2024]. This presents a severe challenge for current deraining algorithms, with the core technical limitation stems from the profound confusion and intertwining of rain and haze features [Chen *et al.*, 2023]. Specifically, rain streaks manifest as directional high-frequency patterns, whereas mist presents as spatially low-frequency global blur. Despite their distinct frequency characteristics [Feng *et al.*, 2024; Li *et al.*, 2025a; Feng *et al.*, 2025; Zhang *et al.*, 2026], these phenomena are closely coupled in spatial distribution and exhibit high visual similarity. This inherent correlation makes the accurate modeling of the rain-mist joint degradation pattern difficult.

Existing methods generally exhibit systematic failures when handling rain-mist coupled degradations, as shown in Figure 1. On the one hand, specialized deraining models [Cui *et al.*, 2022; Huang *et al.*, 2023] focus on capturing local high-frequency information, mistakenly identifying mist regions as clean backgrounds or weak rain streaks, thus resulting in blurred images even after deraining. On the other hand, all-in-one models designed for multiple degradations lack prior modeling of the coupling relationship between rain and mist [Potlapalli *et al.*, 2023; Li *et al.*, 2025b; Luo *et al.*, 2025]. This limitation prevents precise decoupling in complex mixed degradation scenarios, resulting in

*Corresponding author: Zhuo Su

incomplete rain streak removal and residual mist. The deeper predicament is that the transient and dynamic characteristics of rain-mist render the creation of precisely paired datasets nearly impossible, posing a fundamental limitation on the application of traditional supervised learning approaches.

The dual challenges of model failure and data scarcity compel us to explore novel learning paradigms: how to tackle the complex problem of rain-mist coupled degradation using only readily available single degradation data. In this paper, we propose a Rain-Mist Removal (RMR) framework based on bilevel coupling learning, as illustrated in Figure 2. Its core innovation lies in leveraging well-established single degradation priors to guide the decoupled modeling of coupled degradation. Specifically, the lower-level optimization process incorporates dehazing priors into the network to achieve core decoupling at the feature level, ensuring precise separation of rain streaks while effectively suppressing residual mist. Building upon this, the upper-level optimization process leverages the powerful vision-language alignment capabilities of CLIP to enhance the model's generalization to real-world scenarios by optimizing a learnable text prompt. By formulating the task as a bilevel problem, our framework establishes a synergistic interaction between low-level structural restoration and high-level semantic guidance. We summarize the main contributions as follows:

- To address the inherent limitations of existing deraining models in real-world rain-mist coexistence scenarios, we propose a novel Rain-Mist Removal (RMR) framework. By leveraging the prior knowledge from single degradation tasks to circumvent the difficulty of end-to-end modeling for rain-mist coupled degradation.
- We introduce a bilevel coupling learning strategy that synergistically integrates low-level feature adaptation with high-level semantic guidance, thereby significantly enhancing the model's generalization capability across diverse real-world scenarios.
- Comprehensive evaluations demonstrate that RMR surpasses state-of-the-art specific deraining and all-in-one methods in both real-world rainy conditions and complex rain-mist scenarios.

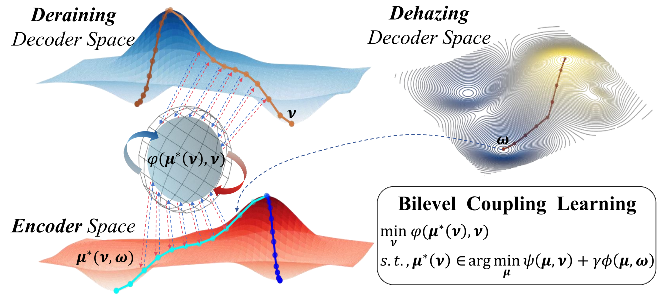


Figure 2: By treating the encoder as a latent lower-level constraint and injecting priors from a pre-trained dehazing model, our bilevel coupling framework enhances model robustness under rain-mist mixed degradations.

2 Related Works

2.1 Single Image Deraining

Single image deraining aims to remove rain streaks and restore obscured background details from degraded images. Early approaches primarily modeled the task based on physical priors [Li *et al.*, 2016]. Driven by the rapid advances in deep learning, various architectures have been proposed to model the physical properties of rain streaks [Li *et al.*, 2025a], while others exploit domain-specific information across different color spaces [Guan *et al.*, 2025]. However, fully-supervised methods heavily rely on large-scale, high-quality paired training data, which significantly limits their generalization capability in real-world scenarios. To overcome the dependency on paired data, researchers have begun exploring unsupervised or semi-supervised deraining methods. Semi-supervised approaches enhance model generalization in practical settings by combining limited paired data with unpaired data during training [Huang *et al.*, 2021; Cui *et al.*, 2022; Huang *et al.*, 2023]. Unsupervised methods leverage the framework of generative adversarial networks to achieve deraining through style transfer between unpaired rainy and rain-free image domains [Dong *et al.*, 2025]. Although these approaches alleviate data dependency to some extent, their performance in real-world scenarios remains constrained. Particularly in complex environments where rain-mist coexist, they often only partially eliminate rain streaks while remaining ineffective against blurred backgrounds and lost details caused by mist.

2.2 Universal Image Restoration

Unified restoration methods based on multi-task learning aim to simultaneously learn to remove multiple degradations within a single network. Some approaches explicitly model distinct degradation characteristics and learn corresponding restoration strategies for them [Li *et al.*, 2022], while others adopt implicit modeling, attempting to map diverse degradation features into a shared latent space before performing unified restoration [Zheng *et al.*, 2024]. However, the performance of such methods heavily relies on the diversity and balance of the training data. When dealing with degradations like rain and mist, which have distinct physical properties and are highly visually coupled, the unified restoration network is difficult to collaboratively optimize and achieve the best performance in both tasks. In recent years, with the rise of large vision models, pre-trained unified restoration methods have demonstrated immense potential [Conde *et al.*, 2024]. These methods learn powerful and generalizable image restoration priors by pre-training on large-scale datasets containing multiple degradation types [Li *et al.*, 2025b]. During the inference stage, they can be guided by text prompts or instructions to execute specific restoration tasks [Potlapalli *et al.*, 2023], giving them unprecedented flexibility and generalization capabilities. Nevertheless, when confronted with coupled degradations not explicitly defined during training, these models also struggle to precisely decoupling and thoroughly remove the different degradation components.

3 Motivation

Through a review of existing work, we find that research on handling the rain-mist coupled degradation in real-world scenarios has reached a dilemma:

Inherent limitations of single-task degradation removal approaches. Existing deraining frameworks are fundamentally constrained by the "clear background + rain streaks" additive assumption, which ignores the presence of low-frequency haze components. On the other hand, dehazing methods target global contrast restoration, leaving them unable to suppress localized high-frequency rain streaks. Such task-specific designs are inherently limited, leading to partial restoration when multiple degradations coexist.

Difficulty of universal restoration models in accurately decoupling rain-mist mixtures. Although universal restoration models for the joint removal of multiple degradations attempt to broaden their applicability, they often treat these degradations as independent or simply linearly superimposable entities. In real-world scenarios where rain-mist coexist, such models may even misinterpret rain streaks as scene details to be preserved, thereby enhancing their visibility during the dehazing process and producing suboptimal results.

4 Methodology

4.1 From Naive Combination to Coupling Learning

To address the challenge of real-world rain-mist removal, we propose a novel paradigm: rather than modeling the complex and intertwined rain-mist coupled degradation, we strategically leverage and integrate existing prior knowledge from single degradation tasks. Recognizing the inherent complementarity between deraining and dehazing at the feature-processing level, we combine an encoder pre-trained on the SOTS-outdoor haze dataset [Li *et al.*, 2019a] with a decoder pre-trained on the DID rain dataset [Zhang and Patel, 2018]. The core insight is to enable the decoder to focus on removing high-frequency rain streaks and reconstructing image details within the feature space where low-frequency mist interference has been eliminated.

As illustrated in Figure 3, this straightforward combination demonstrates surprising potential for removing mixed rain–mist degradation, significantly outperforming single-task models. However, since the encoder and decoder are trained independently on different synthetic datasets, the model exhibits noticeable color distortion while removing the mixed degradation. This result reveals that naive combination cannot yield robust solutions without careful adaptation and collaborative optimization between the two components.

To resolve feature mismatches and improve real-world generalization, we introduce a bilevel coupling learning strategy. This strategy seamlessly integrates the two pre-trained modules into a cohesive system via low-level feature adaptation and high-level semantic guidance, substantially enhancing the model's adaptability to real-world scenarios. Consequently, the model can effectively suppress dense rain streaks while restoring background details and structures typically obscured by haze. The algorithmic details of the bilevel coupling modeling are elaborated below.

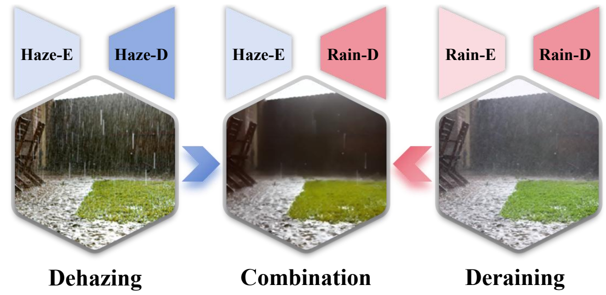


Figure 3: Naive combining can effectively eliminate mixed degradation, but it also introduces color distortion issues.

4.2 Bilevel Coupling Modeling

The bilevel coupling learning strategy is fundamentally designed to leverage and adapt prior knowledge from single degradation tasks for effectively addressing complex rain-mist mixture removal. We formulate the architectural framework as a coupled encoder-decoder system:

$$\begin{cases} \mathbf{z}_{\text{encoder}} = \mathcal{F}_{\boldsymbol{\mu}}(\mathbf{y}), \\ \mathbf{x}_{\text{dehaze}} = \mathcal{F}_{\boldsymbol{\omega}}(\mathbf{z}_{\text{encoder}}), \\ \mathbf{x}_{\text{derain}} = \mathcal{F}_{\boldsymbol{\nu}}(\mathbf{z}_{\text{encoder}}). \end{cases} \tag{1}$$

The proposed pipeline takes an input $\mathbf{y}$, processes it with an encoder $\mathcal{F}_{\boldsymbol{\mu}}$ to extract a latent feature vector $\mathbf{z}_{\text{encoder}}$, and then utilizes this shared feature vector with both $\mathcal{F}_{\boldsymbol{\omega}}$ and $\mathcal{F}_{\boldsymbol{\nu}}$. The terms $\boldsymbol{\mu}$, $\boldsymbol{\omega}$ and $\boldsymbol{\nu}$ denote the parameters for the encoder, dehazing decoder, and deraining decoder, respectively. Then, we formulate the learning objective using a bilevel optimization strategy:

$$\begin{aligned} &\min_{\boldsymbol{\nu}} \varphi(\boldsymbol{\mu}^*(\boldsymbol{\nu}), \boldsymbol{\nu}), \\ &\text{s.t., } \boldsymbol{\mu}^*(\boldsymbol{\nu}) \in \arg\min_{\boldsymbol{\mu}} \psi(\boldsymbol{\mu}, \boldsymbol{\nu}) + \gamma\, \phi(\boldsymbol{\mu}, \boldsymbol{\omega}). \end{aligned} \tag{2}$$

Here, $\varphi(\cdot)$ represents the upper-level objective aimed at real-world adaptation. Concurrently, the lower-level optimization updates the encoder by minimizing a combination of the deraining loss $\psi(\cdot)$ and the dehazing regularization $\phi(\cdot)$, where $\gamma > 0$ serves as the regularization coefficient.

Lower-Level Optimization. The lower-level objective imposes a multi-task constraint on the shared encoder by jointly minimizing reconstruction errors for both deraining and dehazing on synthetic datasets. Formally, the loss functions for the two tasks are defined as:

$$\begin{aligned} \psi(\boldsymbol{\mu}, \boldsymbol{\nu}) &= \left\| \mathcal{F}_{\boldsymbol{\nu}} \circ \mathcal{F}_{\boldsymbol{\mu}}(\mathbf{y}_{\text{rain}}) - \mathbf{x}^{\text{gt}}_{\text{rain}} \right\|_1, \\ \phi(\boldsymbol{\mu}, \boldsymbol{\omega}) &= \left\| \mathcal{F}_{\boldsymbol{\omega}} \circ \mathcal{F}_{\boldsymbol{\mu}}(\mathbf{y}_{\text{haze}}) - \mathbf{x}^{\text{gt}}_{\text{haze}} \right\|_1, \end{aligned} \tag{3}$$

where $\mathbf{y}_{\{\cdot\}}$ and $\mathbf{x}^{\text{gt}}_{\{\cdot\}}$ denote the degraded inputs and their corresponding ground truths. In the lower-level optimization stage, the decoder parameters $\boldsymbol{\nu}$ and $\boldsymbol{\omega}$ are frozen to concentrate the update on the shared encoder $\boldsymbol{\mu}$. We employ the ℓ_1 norm for reconstruction loss, while a hyperparameter γ is introduced to balance the trade-off between deraining performance and dehazing capability. The sensitivity analysis of γ is provided in Section 6.4.

Method	Venue	IVIPC_DQA			MPID			SPA		
		Brisque↓	Entropy↑	CLIPIQA↑	Brisque↓	Entropy↑	CLIPIQA↑	Brisque↓	Entropy↑	CLIPIQA↑
Syn2Real	CVPR'20	25.996	7.3071	0.4777	30.737	7.4635	0.4395	31.754	7.0842	0.4015
MOSS	CVPR'21	28.888	7.3919	0.4804	30.626	7.5239	0.4340	32.641	7.1265	0.3925
SSID-KD	TCSVT'22	26.620	7.4029	0.5186	30.122	7.5332	0.4291	32.961	7.1448	0.4399
MUSS	TPAMI'23	26.389	7.3019	0.4593	29.940	7.4654	0.4171	30.684	7.0839	0.3811
DRSformer	CVPR'23	25.996	7.3071	0.4777	31.730	7.5379	0.4443	33.470	7.0534	0.4477
NeRD-Rain	CVPR'24	29.908	7.4243	0.5289	31.984	7.5557	0.4392	34.481	7.1717	0.4329
MFDNet	TIP'24	31.681	7.4456	0.5357	32.158	7.5545	0.4405	36.829	7.1725	0.4310
CSUD	CVPR'25	24.327	7.3569	0.4850	**24.973**	7.4900	0.3663	28.234	7.1428	0.4360
Restormer	CVPR'22	30.271	7.4360	0.5042	31.786	7.5479	0.4422	35.748	7.1602	0.4258
PromptIR	NeurIPS'23	28.358	7.4532	0.4992	32.038	7.5653	0.4373	33.005	7.2180	0.4407
DiffUIR	CVPR'24	27.985	7.4489	0.5033	31.916	7.5655	0.4420	33.305	7.2329	0.4288
FoundIR	ICCV'25	31.287	7.4523	0.5053	32.209	7.5703	0.4447	37.821	7.1677	0.4461
RMR	Ours	**22.674**	**7.4558**	**0.5420**	25.413	**7.5791**	**0.4639**	**27.588**	**7.2648**	**0.4871**

Table 1: Quantitative comparison on IVIPC_DQA [Wu *et al.*, 2019], MPID [Li *et al.*, 2019b], and SPA [Wang *et al.*, 2019] datasets. The best and second-best performances are highlighted in bold and underlined, respectively.

Upper-Level Optimization. Unlike the lower-level optimization that relies on synthetic paired data, the upper-level objective $\varphi(\cdot)$ is designed to bridge the domain gap and enhance generalization to real-world degradation. Motivated by prior work [Liang *et al.*, 2023; Ma *et al.*, 2025], we employ the CLIP [Radford *et al.*, 2021] latent space to define a contrastive learning objective. To this end, we construct a reference [Guo *et al.*, 2023] set consisting of 500 real rain-degraded images $\mathcal{I}_{\mathcal{D}} = \{I_{\mathcal{D}}^{(i)}\}_{i=1}^{500}$ and 500 real clean images $\mathcal{I}_{\mathcal{C}} = \{I_{\mathcal{C}}^{(i)}\}_{i=1}^{500}$, coupled with learnable text prompts $T_{\mathcal{D}}$ and $T_{\mathcal{C}}$ representing degraded and clean categories, respectively. Let $\Phi_{\text{image}}(\cdot)$ and $\Phi_{\text{text}}(\cdot)$ denote the pre-trained CLIP image and text encoders. The text prompts are optimized via binary cross entropy loss to minimize classification error within the embedding space. The upper-level loss is then formulated as:

$$\mathcal{L}_{\text{upper}} = \frac{e^{\cos\left(\Phi_{\text{image}}(I_{\mathcal{R}}),\Phi_{\text{text}}(T_{\mathcal{D}})\right)}}{\sum_{i\in\{\mathcal{D},\mathcal{C}\}} e^{\cos\left(\Phi_{\text{image}}(I_{\mathcal{R}}),\Phi_{\text{text}}(T_i)\right)}}, \tag{4}$$

in which $I_{\mathcal{R}}$ denotes the restored output of a real rainy image from the DDN-SIRR dataset [Wei *et al.*, 2019]. This objective encourages the restored image to align with the clean distribution in the semantic feature space, thereby improving visual fidelity and perceptual realism. The algorithmic solution for this bilevel optimization problem is detailed in the following section.

Algorithm 1 Bilevel Coupling Learning

Require: Initial parameters $\boldsymbol{\mu}_0$ (pre-trained dehazing encoder), $\boldsymbol{\nu}_0$ (pre-trained deraining decoder), fixed dehazing decoder $\boldsymbol{\omega}$, step-sizes η_μ, η_ν, regularization γ, the maximum iteration K, the lower-loop N_l, the upper-loop N_u, and set $k = 0$.

Ensure: Optimal parameters $\boldsymbol{\mu}^*$, $\boldsymbol{\nu}^*$.

while not converged and $k \leq K$ **do**
  % Update the lower-level parameters
  for $n = 1$ **to** N_l **do**
    $f \leftarrow \psi(\boldsymbol{\mu}_k^n, \boldsymbol{\nu}_k) + \gamma\phi(\boldsymbol{\mu}_k^n, \boldsymbol{\omega})$
    $\boldsymbol{\mu}_k^{n+1} \leftarrow \boldsymbol{\mu}_k^n - \eta_\mu \nabla_{\boldsymbol{\mu}} f(\boldsymbol{\mu}_k^n, \boldsymbol{\nu}_k, \boldsymbol{\omega})$
  end for
  Set $\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k^{N_l}$ to approximate $\boldsymbol{\mu}^*(\boldsymbol{\nu}_k)$
  % Update the upper-level parameters
  for $n = 1$ **to** N_u **do**
    % Estimate hypergradient $\boldsymbol{g}_\nu$ by Eq. (9)
    $\boldsymbol{\nu}_k^{n+1} \leftarrow \boldsymbol{\nu}_k^n - \eta_\nu \boldsymbol{g}_\nu(\boldsymbol{\mu}_{k+1}, \boldsymbol{\nu}_k^n)$
  end for
  $\boldsymbol{\nu}_{k+1} = \boldsymbol{\nu}_k^{N_u}; k = k + 1.$
end while
return $\boldsymbol{\mu}^*, \boldsymbol{\nu}^*$

4.3 Hypergradient Estimation via Approximate Inference

Solving the bilevel optimization problem (2) primarily requires computing the hypergradient of the upper-level objective. With $\boldsymbol{\mu}^*(\boldsymbol{\nu})$ being the lower-level optimal solution, the hypergradient $\boldsymbol{g}_\nu \triangleq \frac{d}{d\boldsymbol{\nu}}\varphi(\boldsymbol{\mu}^*(\boldsymbol{\nu}), \boldsymbol{\nu})$ can be decomposed via the chain rule as follows:

$$\boldsymbol{g}_\nu = \nabla_{\boldsymbol{\nu}}\varphi(\boldsymbol{\mu}^*, \boldsymbol{\nu}) + \nabla_{\boldsymbol{\nu}}\boldsymbol{\mu}^*(\boldsymbol{\nu})^T \nabla_{\boldsymbol{\mu}}\varphi(\boldsymbol{\mu}^*, \boldsymbol{\nu}), \tag{5}$$

where $\nabla_{\boldsymbol{\nu}}\varphi$ and $\nabla_{\boldsymbol{\mu}}\varphi$ denote the partial derivatives of the upper-level objective with respect to $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$, respectively.

Directly computing the Jacobian matrix $\nabla_{\boldsymbol{\nu}}\boldsymbol{\mu}^*(\boldsymbol{\nu})$ is challenging. To circumvent this, we denote the lower-level objective as f and assume it is twice continuously differentiable. By invoking the implicit function theorem [Krantz and Parks, 2002] on the first-order stationary condition $\nabla_{\boldsymbol{\mu}} f(\boldsymbol{\mu}^*, \boldsymbol{\nu}) =$

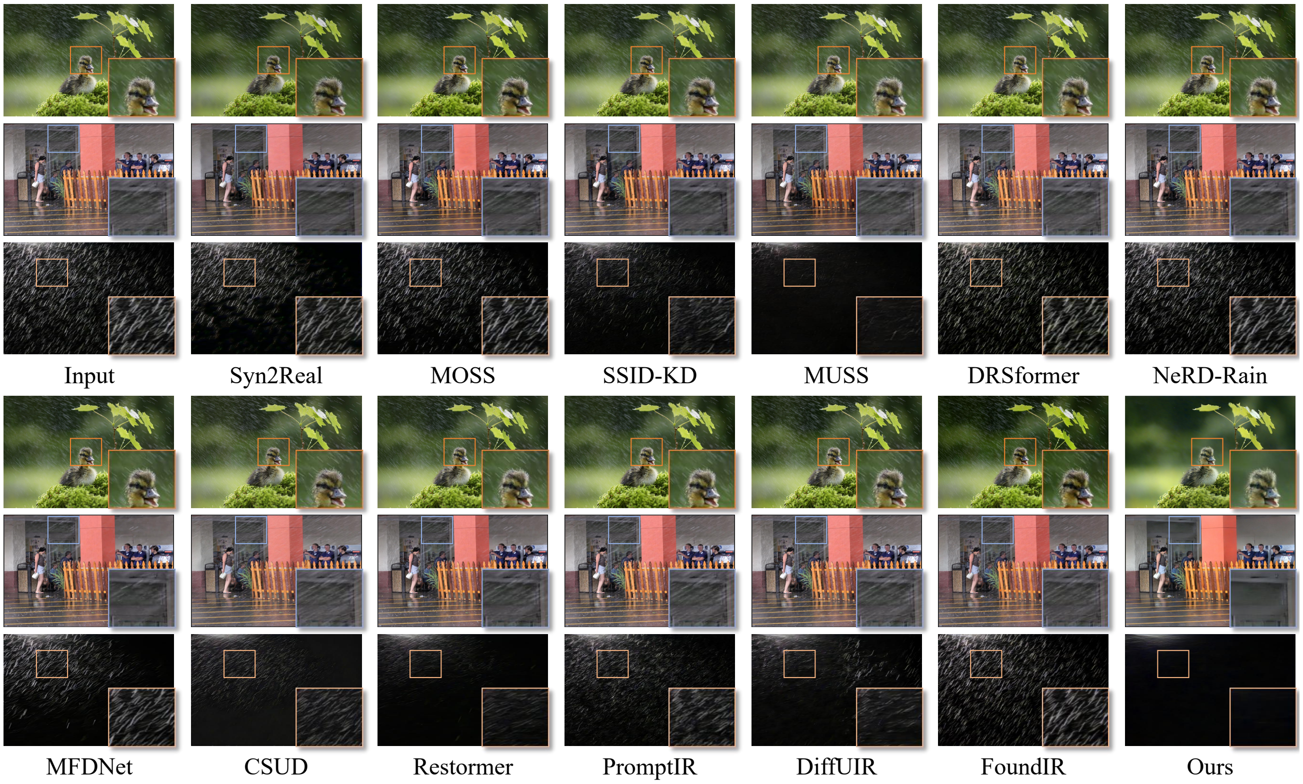


Figure 4: Qualitative comparison of different state-of-the-art methods and our method. The boxes highlight the distinct differences. (These data are sourced from the IVIPC_DQA, MPID, and SPA datasets, respectively.)

$\mathbf{0}$, and differentiating this condition with respect to $\boldsymbol{\nu}$:

$$\nabla^2_{\boldsymbol{\mu}\boldsymbol{\mu}} f(\boldsymbol{\mu}^*, \boldsymbol{\nu}) \nabla_{\boldsymbol{\nu}} \boldsymbol{\mu}^*(\boldsymbol{\nu}) + \nabla^2_{\boldsymbol{\mu}\boldsymbol{\nu}} f(\boldsymbol{\mu}^*, \boldsymbol{\nu}) = \mathbf{0}. \tag{6}$$

Provided that the Hessian $\nabla^2_{\boldsymbol{\mu}\boldsymbol{\mu}} f$ is non-singular, the Jacobian can be explicitly expressed as:

$$\nabla_{\boldsymbol{\nu}} \boldsymbol{\mu}^*(\boldsymbol{\nu}) = -\left[\nabla^2_{\boldsymbol{\mu}\boldsymbol{\mu}} f(\boldsymbol{\mu}^*, \boldsymbol{\nu})\right]^{-1} \nabla^2_{\boldsymbol{\mu}\boldsymbol{\nu}} f(\boldsymbol{\mu}^*, \boldsymbol{\nu}). \tag{7}$$

However, computing and inverting the second-order Hessian $\nabla^2_{\boldsymbol{\mu}\boldsymbol{\mu}} f$ is computationally expensive. To address this, we employ an outer-product approximation [Liu *et al.*, 2024] to represent the second-order curvature:

$$\nabla^2_{\boldsymbol{\mu}\boldsymbol{\mu}} f \approx \nabla_{\boldsymbol{\mu}} f \nabla_{\boldsymbol{\mu}} f^T, \quad \nabla^2_{\boldsymbol{\mu}\boldsymbol{\nu}} f \approx \nabla_{\boldsymbol{\mu}} f \nabla_{\boldsymbol{\nu}} f^T. \tag{8}$$

This corresponds to a rank-one quasi-Newton approximation of the Hessian matrix. By substituting Eq. (8) into Eq. (7) and then into Eq. (5), the hypergradient $\boldsymbol{g}_\nu$ can be efficiently approximated using only first-order information:

$$\boldsymbol{g}_\nu \approx \nabla_{\boldsymbol{\nu}} \varphi - \left(\frac{\nabla_{\boldsymbol{\mu}} \varphi^T \nabla_{\boldsymbol{\mu}} f}{\nabla_{\boldsymbol{\mu}} f^T \nabla_{\boldsymbol{\mu}} f}\right) \nabla_{\boldsymbol{\nu}} f. \tag{9}$$

In addition to the current method, other algorithms are available for solving Eq. (2) [Liu *et al.*, 2022]. Algorithm 1 summarizes the training procedure for the RMR model. The framework alternates between two nested stages: the lower-level loop approximates $\boldsymbol{\mu}^*(\boldsymbol{\nu})$ via N_l gradient steps, and the upper-level loop updates $\boldsymbol{\nu}$ using the hypergradient estimate in Eq. (9) over N_u iterations.

5 Experiments

5.1 Experimental Settings

Training settings. The RMR model was implemented using the PyTorch framework on a single NVIDIA GeForce RTX 3090 GPU. For all experiments, we adopted DEA [Chen *et al.*, 2024b] as the baseline and employed the Adam optimizer to update the parameters, with β_1, β_2, and ε set to 0.9, 0.999, and 1×10^{-8}, respectively. During the pre-training stage, the encoder and decoder were pre-trained on the SOTS-outdoor and DID datasets. Furthermore, the learnable text prompts for the upper-level optimization were optimized on 500 real rain-degraded and clean images [Guo *et al.*, 2023]. The learning rate was initialized at 1×10^{-3} for pre-training and 1×10^{-4} for bilevel coupling learning, decaying to 1×10^{-4} and 1×10^{-5}, respectively. During all training stages, we augmented the training data by cropping random 256×256 patches from the images. Notably, all comparative experiments were performed with the dehazing decoder frozen and the hyperparameter γ set to 1.

Benchmarks and Metrics In this paper, we evaluated the proposed method on three representative real-world datasets: IVIPC_DQA [Wu *et al.*, 2019], MPID [Li *et al.*, 2019b], and SPA [Wang *et al.*, 2019], which consisted of 206, 185, and 103 images, respectively. Objective performance was measured via no-reference metrics such as Brisque [Mittal *et al.*, 2012] and Entropy [Wang and Bovik, 2006], as well as the LLM-driven metric CLIPIQA [Wang *et al.*, 2023].

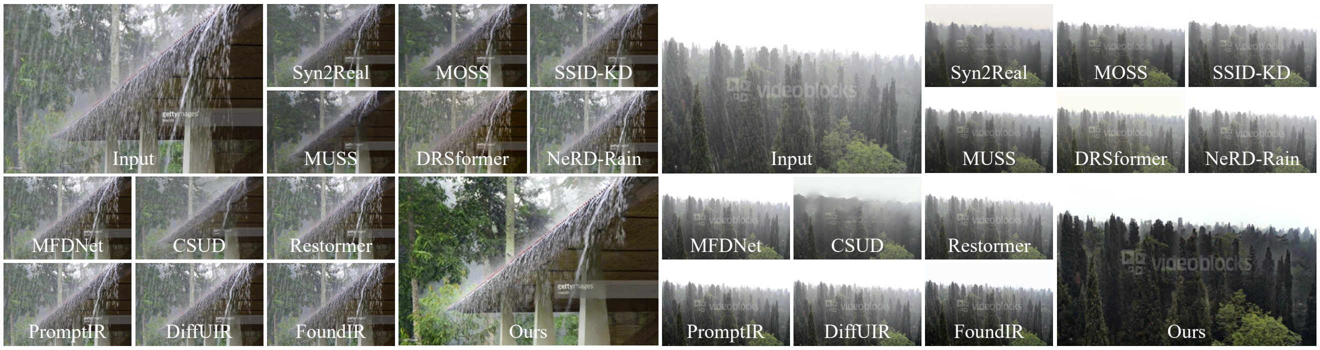


Figure 5: Qualitative comparison of different methods on challenging rain-mist coexisting scenarios in the SPA datasets.

Method	Params (M)	FLOPs (G)	Runtime (ms)	
			256×256	720×1280
DRSformer	33.63	220.4	173.6	2294
NeRD-Rain	22.89	148.0	100.2	1178
CSUD	29.16	16.07	21.21	147.4
PromptIR	32.97	158.1	87.39	1180
DiffUIR	12.40	71.92	117.1	1641
FoundIR	36.23	131.8	181.9	2363
RMR	**3.653**	**34.04**	**6.472**	**83.24**

Table 2: Computational efficiency comparison of different methods. FLOPs are calculated on 256×256 images.

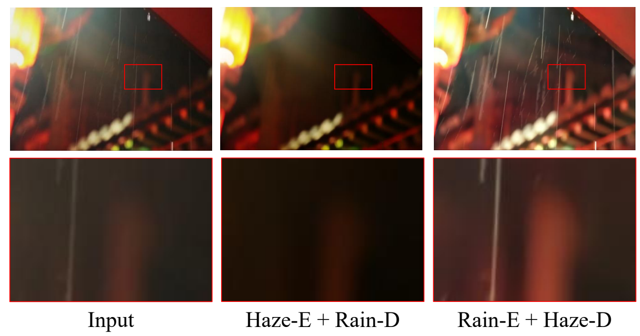


Figure 6: Complementarity of degradation priors. (These data are sourced from the SPA dataset.)

Comparison Methods To thoroughly evaluate the generalization capability of RMR, we conduct quantitative and qualitative comparisons against existing state-of-the-art methods. For specialized deraining algorithms, we selected Syn2Real [Yasarla *et al.*, 2020], MOSS [Huang *et al.*, 2021], SSID-KD [Cui *et al.*, 2022], MUSS [Huang *et al.*, 2023], DRSformer [Chen *et al.*, 2023], NeRD-Rain [Chen *et al.*, 2024a], MFDNet [Wang *et al.*, 2024], and CSUD [Dong *et al.*, 2025]. Additionally, we compared RMR with all-in-one restoration frameworks, including Restormer [Zamir *et al.*, 2022], PromptIR [Potlapalli *et al.*, 2023], DiffUIR [Zheng *et al.*, 2024], and FoundIR [Li *et al.*, 2025b].

5.2 Results and Comparison

Quantitative Evaluation. Table 1 provides a comprehensive quantitative comparison between RMR and several state-of-the-art methods. Given the lack of ground-truth labels in real-world datasets, we employ three authoritative no-reference evaluation metrics to assess the restoration quality objectively. Results demonstrate that our method ranks among the top across nearly all metrics, validating the effectiveness of RMR in diverse real-world scenarios. Furthermore, Table 2 illustrates that RMR achieves a superior trade-off between restoration quality and computational efficiency.

Qualitative Comparison. Figure 4 illustrates that RMR delivers consistently robust performance for rain streak removal across a variety of scenes. To further examine its effectiveness in real-world complex scenarios, we select challenging rain-mist coexisting scenarios from the SPA dataset. As shown in Figure 5, RMR demonstrates superior capability in handling the coupled rain-mist degradation, simultaneously removing both rain streaks and mist while preserving fine details, whereas other methods often suffer from residual visual artifacts and blur.

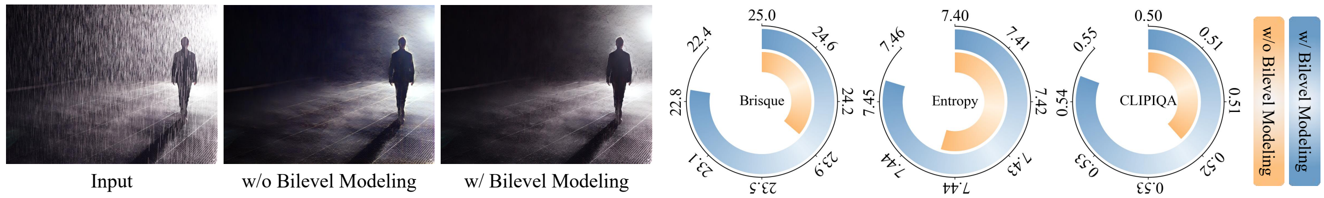
6 Algorithmic Analyses

6.1 Complementarity of Degradation Priors

We investigate the synergistic effects of different prior combinations by comparing our "Dehazing Encoder + Deraining Decoder" (Haze-E + Rain-D) configuration with its inverted counterpart (Rain-E + Haze-D). As illustrated in Figure 6, the inverted architecture yields suboptimal results. This performance degradation stems from the inherent architectural bias where a deraining-oriented encoder lacks the capacity to effectively capture global low-frequency haze distributions, while a dehazing-oriented decoder tends to misinterpret residual rain streaks as high-frequency scene details. In contrast, the Haze-E + Rain-D strategy exhibits superior task complementarity, providing essential prior knowledge that facilitates subsequent bilevel coupling learning.

6.2 Necessity of Bilevel Coupling Modeling

When only lower-level optimization is employed, the inherent domain shift between synthetic and real data often leads



(a) Qualitative Comparison (a) Quantitative Comparison

Figure 7: Necessity of bilevel coupling modeling. (These data are sourced from the IVIPC_DQA dataset.)

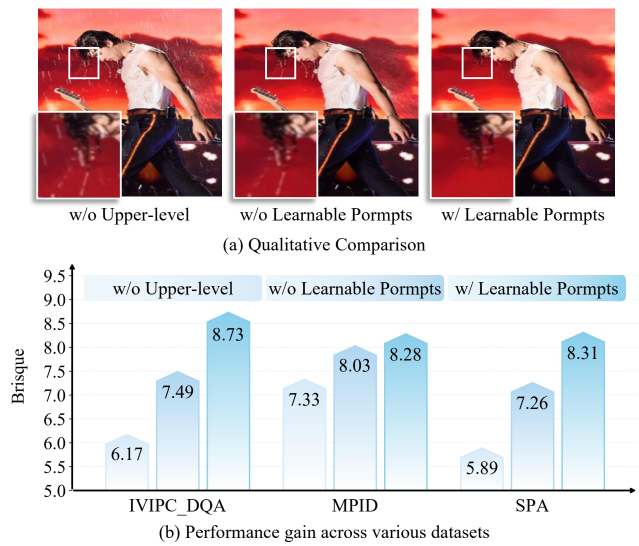


Figure 8: Learnable prompts improve Brisque performance across various datasets.

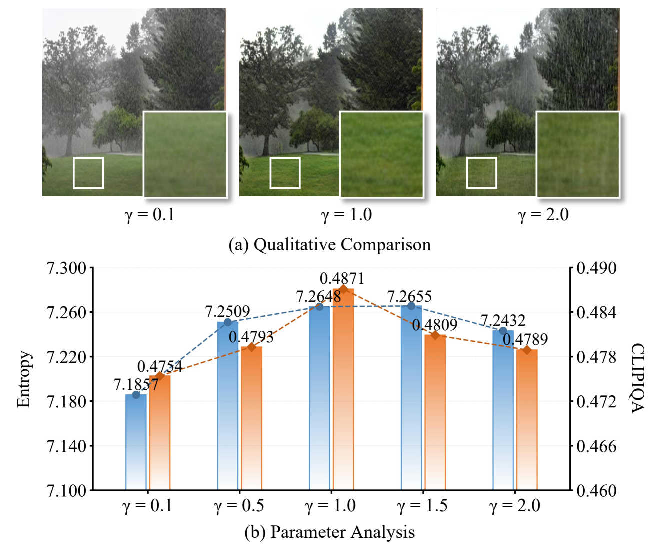


Figure 9: Influence of Prior Strength. (These data are sourced from the SPA dataset.)

to residual rain streaks. To validate the necessity of the bilevel coupling modeling, we compare the performance of fine-tuning versus bi-level optimization strategy in Figure 7. Experimental results demonstrate that incorporating upper-level objectives solely through fine-tuning leads to visual artifacts while removing rain streaks. This is attributed to a lack of lower-level pixel-wise constraints, which causes the model to overfit to semantic objectives. In contrast, bilevel optimization synergizes lower-level and upper-level constraints, effectively addressing complex degradations in real-world environments while preserving structural consistency.

6.3 Importance of Learnable Prompts

During the upper-level optimization stage, learnable prompts substantially enhance adaptation performance. As illustrated in Figure 8, learnable prompts consistently outperform fixed prompts (e.g., “a photo with rain” or “a clear photo”). This performance disparity primarily arises from the static nature of fixed prompts, which restricts their capacity to dynamically capture complex rain-related features. In contrast, text prompts refined through parameter optimization achieve more precise semantic representations, thereby effectively eliminating residual rain streaks.

6.4 Influence of Prior Strength

Figure 9 illustrates the impact of the hyperparameter γ on model performance. As γ increases, the model integrates more robust dehazing priors, which helps maintain stable dehazing performance during the deraining process. This improved balance is evidenced by the consistent improvement in no-reference metrics. However, when γ exceeds a certain threshold, the dehazing prior becomes overly restrictive, eventually suppressing the model’s capacity to effectively process rain streak features.

7 Conclusion

In this paper, we propose a Rain-Mist Removal (RMR) framework to address the limitations of existing image deraining models in rain-mist coexistence scenarios. By employing a bilevel coupling learning strategy which leverages prior knowledge from single degradation tasks, RMR achieves precise decoupling of rain and haze features during lower-level optimization process. Simultaneously, the upper-level decoder incorporates semantic guidance from CLIP to enhance generalization. Experiments demonstrate RMR’s effectiveness on multiple real-world datasets. Future work aims to enhance its applicability in resource-constrained environments.

Acknowledgments

This research was supported by the foundations of Guangdong Basic and Applied Basic Research Foundation (No. 2024A1515011563), Natural Science Foundation of Hangzhou (No. 2025SZRJJ1901), Shenzhen Science and Technology Program (No. SYSPG20241211173951079).

References

[Chen *et al.*, 2023] Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan. Learning a sparse transformer network for effective image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5896–5905, 2023.

[Chen *et al.*, 2024a] Xiang Chen, Jinshan Pan, and Jiangxin Dong. Bidirectional multi-scale implicit neural representations for image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 25627–25636, 2024.

[Chen *et al.*, 2024b] Zixuan Chen, Zewei He, and Zhe-Ming Lu. Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE Transactions on Image Processing*, 33:1002–1015, 2024.

[Chen *et al.*, 2025] Xiang Chen, Jinshan Pan, Jiangxin Dong, and Jinhui Tang. Towards unified deep image deraining: A survey and a new benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7):5414–5433, 2025.

[Conde *et al.*, 2024] Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *European Conference on Computer Vision (ECCV)*, pages 1–21, 2024.

[Cui *et al.*, 2022] Xin Cui, Cong Wang, Dongwei Ren, Yunjin Chen, and Pengfei Zhu. Semi-supervised image deraining using knowledge distillation. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(12):8327–8341, 2022.

[Dong *et al.*, 2025] Guanglu Dong, Tianheng Zheng, Yuanzhouhan Cao, Linbo Qing, and Chao Ren. Channel Consistency Prior and Self-Reconstruction Strategy Based Unsupervised Image Deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7469–7479, 2025.

[Feng *et al.*, 2024] Yuxin Feng, Long Ma, Xiaozhe Meng, Fan Zhou, Risheng Liu, and Zhuo Su. Advancing real-world image dehazing: Perspective, modules, and training. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9303–9320, 2024.

[Feng *et al.*, 2025] Yuxin Feng, Jufeng Li, Tao Huang, Fangfang Wu, Yakun Ju, Chunxu Li, Weisheng Dong, and Alex C. Kot. Cross-frequency attention and color contrast constraint for remote sensing dehazing. *IEEE Transactions on Image Processing*, 34:8552–8567, 2025.

[Guan *et al.*, 2025] Qiyuan Guan, Xiang Chen, Guiyue Jin, Jiyu Jin, Shumin Fan, Tianyu Song, and Jinshan Pan. Rethinking nighttime image deraining via learnable color space transformation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, pages 1–13, 2025.

[Guo *et al.*, 2023] Yun Guo, Xueyao Xiao, Yi Chang, Shumin Deng, and Luxin Yan. From sky to the ground: A large-scale benchmark and simple baseline towards real rain removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12097–12107, 2023.

[Huang *et al.*, 2021] Huaibo Huang, Aijing Yu, and Ran He. Memory Oriented Transfer Learning for Semi-Supervised Image Deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7732–7741, 2021.

[Huang *et al.*, 2023] Huaibo Huang, Mandi Luo, and Ran He. Memory Uncertainty Learning for Real-World Single Image Deraining. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 208(4447):1019–1022, 2023.

[Krantz and Parks, 2002] Steven G. Krantz and Harold R. Parks. *The Implicit Function Theorem: History, Theory, and Applications*. Springer Scence & Business Media, 2002.

[Li *et al.*, 2016] Yu Li, Robby T. Tan, Xiaojie Guo, Jiangbo Lu, and Michael S. Brown. Rain streak removal using layer priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2736–2744, 2016.

[Li *et al.*, 2019a] 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(1):492–505, 2019.

[Li *et al.*, 2019b] Siyuan Li, Iago Breno Araujo, Wenqi Ren, Zhangyang Wang, Eric K. Tokuda, Roberto Hirata Junior, Roberto Cesar-Junior, Jiawan Zhang, Xiaojie Guo, and Xiaochun Cao. Single image deraining: A comprehensive benchmark analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3838–3847, 2019.

[Li *et al.*, 2022] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17452–17462, 2022.

[Li *et al.*, 2024] Xin Li, Yuxin Feng, Fan Zhou, Yun Liang, and Zhuo Su. Revitalizing real image deraining via a generic paradigm towards multiple rainy patterns. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1029–1037, 2024.

[Li *et al.*, 2025a] Dong Li, Yidi Liu, Xueyang Fu, Senyan Xu, and Zhengjun Zha. Fouriermamba: Fourier learning integration with state space models for image deraining. In *International Conference on Machine Learning (ICML)*, volume 267, pages 35342–35365. PMLR, 2025.

[Li *et al.*, 2025b] Hao Li, Xiang Chen, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. Foundir: Unleashing million-scale training data to advance foundation models for image restoration. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 12626–12636, 2025.

[Liang *et al.*, 2023] Zhexin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Iterative prompt learning for unsupervised backlit image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8094–8103, 2023.

[Liu *et al.*, 2022] Risheng Liu, Long Ma, Xiaoming Yuan, Shangzhi Zeng, and Jin Zhang. Task-oriented convex bilevel optimization with latent feasibility. *IEEE Transactions on Image Processing*, 31:1190–1203, 2022.

[Liu *et al.*, 2024] Risheng Liu, Jiaxin Gao, Xuan Liu, and Xin Fan. Learning with constraint learning: New perspective, solution strategy and various applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(7):5026–5043, 2024.

[Luo *et al.*, 2025] Wenyang Luo, Haina Qin, Zewen Chen, Libin Wang, Dandan Zheng, Yuming Li, Yufan Liu, Bing Li, and Weiming Hu. Visual-instructed degradation diffusion for all-in-one image restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 12764–12777, 2025.

[Ma *et al.*, 2025] Long Ma, Yuxin Feng, Yan Zhang, Jinyuan Liu, Weimin Wang, Guangyong Chen, Chengpei Xu, and Zhuo Su. Coa: Towards real image dehazing via compression-and-adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11197–11206, 2025.

[Mittal *et al.*, 2012] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.

[Potlapalli *et al.*, 2023] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. In *Advances in Neural Information Processing Systems*, volume 36, pages 71275–71293, 2023.

[Radford *et al.*, 2021] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning (ICML)*, pages 8748–8763, 2021.

[Wang and Bovik, 2006] Zhou Wang and Alan C. Bovik. No-reference image quality assessment. *Cham: Springer International Publishing*, pages 79–102, 2006.

[Wang *et al.*, 2019] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson W.H. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12270–12279, 2019.

[Wang *et al.*, 2023] Jianyi Wang, Kelvin C.K. Chan, and Chen Change Loy. Exploring CLIP for assessing the look and feel of images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2555–2563, 2023.

[Wang *et al.*, 2024] Qiong Wang, Kui Jiang, Zheng Wang, Wenqi Ren, Jianhui Zhang, and Chia-Wen Lin. Multi-scale fusion and decomposition network for single image deraining. *IEEE Transactions on Image Processing*, 33:191–204, 2024.

[Wei *et al.*, 2019] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3877–3886, 2019.

[Wu *et al.*, 2019] Qingbo Wu, Lei Wang, King N. Ngan, Hongliang Li, and Fanman Meng. Beyond synthetic data: A blind deraining quality assessment metric towards authentic rain image. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 2364–2368, 2019.

[Yang *et al.*, 2017] Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1357–1366, 2017.

[Yasarla *et al.*, 2020] Rajeev Yasarla, Vishwanath A. Sindagi, and Vishal M. Patel. Syn2real transfer learning for image deraining using gaussian processes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2726–2736, 2020.

[Zamir *et al.*, 2022] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5728–5739, 2022.

[Zhang and Patel, 2018] He Zhang and Vishal M. Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 695–704, 2018.

[Zhang *et al.*, 2026] Yan Zhang, Yuxin Feng, Xin Li, Fan Zhou, and Zhuo Su. Pdda: Prompt-driven domain adaptation for real-world image dehazing. *IEEE Transactions on Image Processing*, 35:4280–4294, 2026.

[Zheng *et al.*, 2024] Dian Zheng, Xiaoming Wu, Shuzhou Yang, Jian Zhang, Jianfang Hu, and Weishi Zheng. Selective hourglass mapping for universal image restoration based on diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 25445–25455, 2024.