



Article

Towards Adaptive Adverse Weather Removal via Semantic and Low-Level Visual Perceptual Priors

Wei Dong , Han Zhou *, Terry Ji and Jun Chen *

Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4L8, Canada; dongw22@mcmaster.ca (W.D.); jit7@mcmaster.ca (T.J.)

* Correspondence: zhouh115@mcmaster.ca (H.Z.); chenjun@mcmaster.ca (J.C.)

Abstract

Adverse weather removal aims to restore images degraded by haze, rain, or snow. However, existing unified models often rely on implicit degradation cues, making them vulnerable to inaccurate weather perception and insufficient semantic guidance, which leads to over-smoothing or residual artifacts in real scenes. In this work, we propose **AWR-VIP**, a prior-guided adverse weather removal framework that explicitly extracts semantic and perceptual priors using a frozen vision–language model (VLM). Given a degraded input, we first employ a degradation-aware prompt extractor to produce a compact set of semantic tags describing key objects and regions, and simultaneously perform **weather-type perception** by prompting the VLM with explicit weather definitions. Conditioned on the predicted weather type and selected tags, the VLM further generates two levels of restoration guidance: a **global instruction** that summarizes image-level enhancement goals (e.g., visibility/contrast) and **local instructions** that specify tag-aware refinement cues (e.g., recover textures for specific regions). These textual outputs are encoded by a text encoder into a pair of priors ($\mathbf{P}_{global}$ and $\mathbf{P}_{local}$), which are injected into a UNet-based restorer through global-prior-modulated normalization and instruction-guided attention, enabling weather-adaptive and content-aware restoration. Extensive experiments on a combined benchmark show that AWR-VIP consistently outperforms state-of-the-art methods. Moreover, the VLM-derived priors are plug-and-play and can be integrated into other restoration backbones to further improve performance.

Keywords: adverse weather removal; semantic prior; low-level visual prior

1. Introduction

Image restoration has been a fundamental part of computer vision, aimed at restoring degraded images to their high-quality counterpart. Specifically, image restoration in adverse weather is essential for emerging sectors such as autonomous driving [1] where high-quality images are needed for accurate decision-making. These weather degradations include, but are not limited to, different levels of rain [2,3], snow [4,5], and haze [6–9].

To address these weather degradations, early weather restoration methods involve traditional methods using Convolutional Neural Network (CNN) [10] to restore images affected by degradations, leveraging convolutional layers to learn spatial patterns and feature details. Transformers [11] have also been extensively studied, often outperforming traditional CNNs with their superior ability in capturing global dependencies and generalization at the expense of larger computational resources. Denoising Diffusion Probabilistic Models (DDPMs) [12] have also shown to be highly effective in image restoration through



Received: 27 December 2025

Revised: 26 January 2026

Accepted: 10 February 2026

Published: 12 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

their iterative denoising process, ideal for learning weather degradations. More importantly, many existing works focus on single-weather degradation network [1–3,6–9,13,14], while few propose non-blind general purpose models designed to restore all weather conditions [15]. However, these models often rely on an initial weather prior given or task-specific models for each weather degradation type, which is suboptimal for blind real-life situations. For improved efficiency, recent studies pivoted their approach to an all-in-one adverse weather restoration (AWR) model [16–19] designed to adaptively learn various weather conditions in a single network. Despite these advances, two challenges remain, hindering generalization in real-world scenarios and the quality of restored images.

First, a fundamental unresolved issue in all-in-one AWR is the reliable disentanglement of weather factors from corrupted observations, i.e., inferring the latent weather type and severity directly from degraded inputs. This capability is consequential because it enables the restoration pipeline to condition its computation—via mechanism selection or architectural modulation—on the inferred degradation, thereby producing more weather-aligned reconstructions. Existing efforts either embed weather characteristics implicitly through learnable restoration backbones [17,20] or explicitly estimate the condition using auxiliary predictors, ranging from conventional classifiers [21,22] to CLIP-driven formulations [23–25]. However, such condition-awareness is often brittle under distribution shifts, leading to pronounced performance deterioration on out-of-distribution (OOD) data and revealing limited extrapolative generalization.

Second, incorporating semantic knowledge is essential for AWR because restoration decisions are inherently content-dependent: preserving object contours, maintaining material-consistent textures, and avoiding over-smoothing all require understanding what the scene depicts rather than only how it is corrupted. Semantic priors can guide the model toward structure-preserving, context-aware reconstructions by constraining the solution space toward perceptually plausible and object-consistent outputs. Despite this importance, how to reliably extract semantic priors from degraded observations and integrate them into AWR models in a principled manner remains underexplored.

To tackle these issues, we propose to leverage semantic and low-level Visual Perceptual priors extracted from pre-trained vision language models for adaptive Adverse Weather Removal (AWR-VIP). Specifically, we first introduce an innovative pipeline that leverages comprehensive priors from pre-trained vision language models, capturing both semantic and attribute-related information such as degradation types and restoration instructions. To utilize these priors for removing adverse weather conditions effectively, we deploy a UNet architecture that integrates these priors to enhance layer normalization and cross-channel attention within its backbone. Our main contributions are summarized below:

- We propose a VLM-guided prior extraction pipeline that explicitly produces (i) **weather-type perception**, (ii) **compact semantic tags**, and (iii) **global/local restoration instructions**, which are encoded as **semantic and perceptual priors**.
- We develop **AWR-VIP**, a unified adverse weather removal framework that performs restoration by conditioning a UNet-based backbone on the extracted **global and local priors**, enabling **weather-adaptive** and **content-aware** enhancement.
- We design two complementary prior injection mechanisms: the **global prior** modulates the affine parameters of **layer normalization** for image-level adaptation, while the **local prior** guides a **cross-attention** module to refine key semantic regions.
- Extensive experiments on combined hazy/rainy/snowy benchmarks demonstrate that AWR-VIP achieves state-of-the-art performance. Moreover, the extracted priors are **plug-and-play** and can be integrated into existing restoration backbones to further improve their performance.

The remainder of this article is organized as follows. Section 2 reviews the related work on adverse weather removal and VLM-guided restoration. Section 3 introduces the motivation of employing guidance derived from vision language models (VLMs) for adverse weather removal, proposes an effective pipeline that guides VLMs to perceive semantic contents—the degradation type in weather-affected images—and then outputs global and tag-specific restoration. Moreover, Section 3.3 also presents a UNet architecture with several special modules to incorporate these priors to guide the adverse weather restoration process. Section 4 reports the experimental findings and analysis. Finally, Section 5 summarizes the main contributions and outlines directions for future research.

2. Related Work

2.1. Adverse Weather Deep Learning

Although single-weather restoration methods [26–29] often deliver strong results for a specific degradation, their specialization limits applicability in real-world imagery where adverse conditions are diverse, overlapping, and difficult to pre-identify. This limitation has motivated the development of blind and non-blind all-in-one AWR models. As an early non-blind attempt, all-in-one [15] employs a CNN whose task-dependent encoders are discovered via neural architecture search. To avoid explicit weather labels, Zhu et al. [16] propose a two-stage formulation that disentangles weather-agnostic and weather-specific priors within a UNet backbone. With the rise of Transformers [11], TransWeather [17] casts restoration as query–key/value matching across weather types, while AWRCP [18] performs latent-space restoration guided by codebook priors. PromptIR [20] further introduces plug-in prompt blocks that encode degradation-dependent cues and a prompt interaction mechanism that adaptively steers a Transformer-based restorer.

Building on prompt-based conditioning, PIP [30] factorizes prompting into a degradation-aware prompt and a general restoration prompt, coupling them through prompt-to-prompt and selective prompt-to-feature interactions to enhance robustness under mixed weather. MiOIR [31] adopts sequential task learning with prompts to reduce inter-task interference and stabilize optimization across heterogeneous conditions. U-WADN [21] incorporates a nested, unified-width backbone with an automatic width selector to emphasize salient regions while improving efficiency. UtilityIR [32] explicitly models degradation type and severity, combining a marginal quality ranking loss with adaptive normalization/attention to calibrate restoration strength and generalize to unseen mixtures. OneRestore [33] addresses composite degradations by fusing scene descriptors with image features, enabling controllable restoration without assuming a single weather category. Finally, AWRaCLe [34] treats degraded–clean context pairs as visual prompts, jointly extracting weather/type semantics and appearance cues so that the model leverages before–after contrast for targeted correction.

2.2. Language-Driven Image Restoration

Language-driven restoration models treat natural language as a high-level control signal for removing degradation and producing clean images, thereby extending image restoration from purely signal-level correction to instruction-following generation. Recent breakthroughs in VLM architectures have made this direction increasingly viable by injecting semantic alignment and instruction understanding into restoration pipelines. TextIR [35] uses CLIP [36] to enforce text–image feature consistency, guiding the restoration output toward the content implied by textual descriptions. DACLIP [24] advances this idea by training a degradation-aware VLM from text annotations and using it to guide an SDE-based restoration model toward stronger feature quality. InstructPix2Pix [37] provides a diffusion-based instruction-following editor that performs targeted image modifications

from natural-language commands. InstructIR [38] adapts instruction guidance to an all-in-one restoration setting via feature masking under human-written prompts. Co-Instruct [39] develops a large multimodal instruction-tuned model that can reason explicitly and answer open-ended questions, offering richer forms of guidance. VLU-Net [25] proposes an interpretable unfolding framework where a VLM steers gradient steps to automatically choose degradation transformations; combined with hierarchical feature unfolding, it enables all-in-one restoration. LDR [19] queries a VLM to estimate pixel-level degradation maps and uses them to route MoE experts, achieving adaptive restoration without explicit weather supervision. InstructRestore [40] further enables instruction-guided, region-customized restoration by releasing a new dataset and adopting a ControlNet-like design that converts region descriptions into masks and region-wise feature modulation. These studies demonstrate state-of-the-art performance on various image restoration benchmarks and suggest that the use of text-based instructions are effective in understanding and restoring degraded images. Different from existing works, we propose to leverage both certain types of text descriptions and instructions to better capture global and local dependencies, enhancing image restoration.

3. AWR-VIP

As illustrated in Figure 1, our framework follows a two-stage pipeline that first extracts semantic and low-level priors via VLMs and then injects them into AWR-VIP to perform adverse weather removal. The main contributions of this work are twofold: (1) we extract priors that capture both content and degradation in weather-affected images and generate effective global and local restoration instructions; (2) we develop an adverse weather removal network guided by these priors to produce satisfactory restoration results. The overall framework is illustrated in Figure 2. In this section, we first discuss the motivation of employing guidance derived from vision language models (VLMs) for adverse weather removal (Section 3.1). Then, we propose an effective pipeline that guides VLMs to perceive the semantic content and degradation type of weather-affected images and to generate global and tag-specific restoration instructions (Section 3.2). Moreover, we develop a UNet-based architecture with dedicated modules to incorporate these priors, thereby guiding the adverse weather restoration process (Section 3.3).

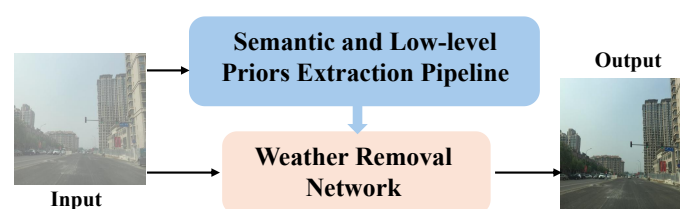


Figure 1. The flowchart of our proposed AWR-VIP. The VLM-based **Semantic and Low-level Priors Generation Pipeline** is introduced to guide the weather removal network.

3.1. Motivation

An ideal all-in-one method, capable of restoring clear images across various weather conditions using a single set of pre-trained weights, should incorporate dedicated modules to detect weather conditions and adapt feature representations accordingly. An intuitive approach involves developing a classifier to categorize weather types. However, the classifier trained on existing adverse weather datasets may not perform effectively in real-world scenarios due to the diversity of weather conditions encountered. Recent works (DACLIP [24]) have developed an image controller based on pre-trained CLIP model to identify degradation types. While this controlling strategy somewhat mitigates over-fitting compared to direct fine-tuning of VLMs, the training dataset remains considerably smaller than the

expansive datasets employed in VLMs. Thus, the DACLIP struggles to retain the impressive zero-shot capabilities of the original VLMs. Inspired by the recent success of VLMs in mastering low-level visual perception and understanding (Q-instruct [41], GPP-LLIE [42], and VAR-LIDE [43]), this paper aims to explore the potential of these VLMs to facilitate the removal of adverse weather conditions.

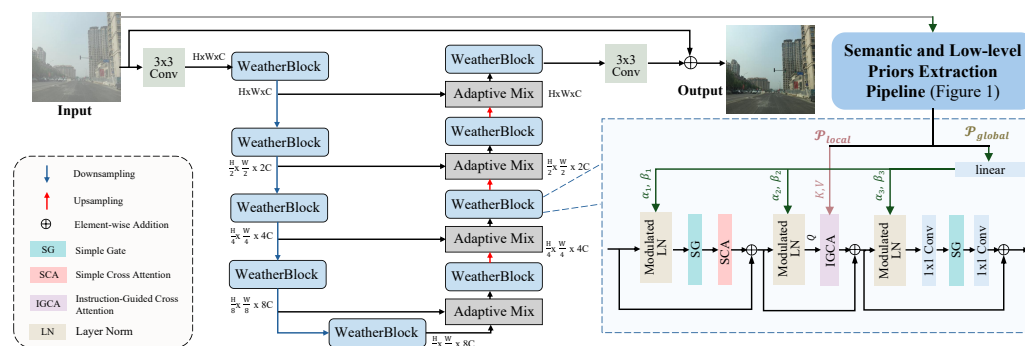


Figure 2. The framework of our proposed adverse weather network. A weather-degraded input is firstly fed into the VLM-based Semantic and Low-level Priors Perception Pipeline to obtain **global and local priors** (P_{global} and P_{local}), which are further injected into AWR-VIP to condition the restoration. $\{\alpha_1, \beta_1\}$, $\{\alpha_2, \beta_2\}$, and $\{\alpha_3, \beta_3\}$ are modulation parameters estimated from P_{global} . Q , K , and V are computed from different sources: the query Q is derived from the input feature, whereas the key K and value V are generated under the guidance of the local prior P_{local} . [Key: 3×3 Conv: 3×3 Convolution].

3.2. Semantic and Low-Level Prior Extraction Pipeline

Pre-trained on extensive text–image pairs, VLMs typically exhibit exceptional zero-shot capability in aligning textual and visual information. Therefore, leveraging such pre-trained knowledge is promising for enabling the perception of diverse weather conditions in unified adverse weather removal models. However, most VLMs primarily concentrate on discerning the semantic content within images, yet they do not adequately capture the specific weather conditions depicted in those images. Even with fine-tuning or a controlling strategy, these models still exhibit limited capability in representing weather conditions, especially on unseen data. In contrast, the VLMs (LLaVA [44]) employed in this paper are fine-tuned with 200 K instruction–response pairs related to low-level visual aspects. In this paper, we introduce a new pipeline to employ LLaVA for all-in-one adverse weather removal: we design text prompts to guide LLaVA to perceive the adverse weather in low-quality images and to output the global restoration instruction. In addition, we introduce the tag selection strategy based on DAPE to acquire primary tag information, and then local restoration instructions are generated by tag-specific question answering session. Our pipeline for low-level visual priors extraction is shown in Figure 3.

3.2.1. VLM Configuration

Notably, we employ a frozen LLaVA-1.5 model for prior extraction using the fixed prompt templates shown in Figure 3. Weather-type perception is implemented as a forced-choice query over {Hazy, Rainy, Snowy}. We adopt deterministic inference settings to ensure stable and reproducible outputs.

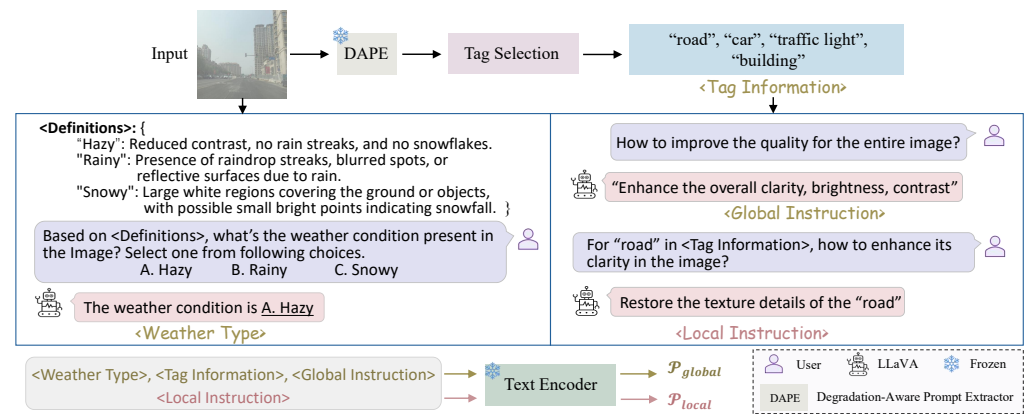


Figure 3. Overview of our VLM-guided prior extraction pipeline. The input image is processed by DAPE [45] to obtain candidate tags, followed by primary tag selection (see Algorithm 1). Conditioned on the weather definitions, the VLM performs weather-type perception and generates global and tag-specific restoration instructions. The resulting texts are encoded by a text encoder into P_{global} and P_{local} , which guide the downstream restoration network.

Algorithm 1 Primary Tag Selection

Require: Low-quality image I_{lq} , pre-trained DAPE θ and the vocabulary $\mathcal{V}$.

```

1: Employ  $\theta$  on  $I_{lq}$ .  $T = \{T_i\}, P = \{P_i\} \leftarrow \theta(I_{lq})$ 
2: Sort  $T$  using  $P$  and select top 8 tags ( $T_s$ )
3: Define  $Index = [], PrimaryTag = []$ 
4: for  $j = 0 : \text{length}(T_s)$  do
5:   Get index  $I_j$  for  $T_s[j]$  from  $\mathcal{V}$ 
6:   Set  $Not\_keep = \text{False}$ 
7:   for  $k = 0 : \text{length}(Index)$  do
8:     if  $\|I_j - I_k\| < 50$  do
9:        $Not\_keep = \text{True}$ 
10:      break
11:   end
12: end
13: if  $Not\_keep = \text{False}$  do
14:    $Index.append(I_j)$ 
15:    $PrimaryTag.append(T_s[j])$ 
16: Return  $PrimaryTag$ 

```

3.2.2. Accurate and Fine-Grained Weather Perceiving

To achieve accurate weather perceiving, we offer definitions **<Definition>** for three common adverse weather conditions to help VLMs better understand the task. Specifically, given an input image, VLMs are prompted to select the option that most likely represents the depicted weather condition based on our developed definitions. In particular, these definitions are crucial, as VLMs generally offer a universal understanding for images affected by weather conditions. These definitions enable VLMs to better distinguish these weather conditions. Without these definitions, there is a high likelihood that images with rainy or snowy scenes are misidentified as hazy because both rain and snow significantly lower visibility and contrast, essential attributes also observed in hazy images. Table 1 summarizes the accuracy of perceived weather conditions, illustrating that the introduction of **<Definition>** significantly improves the performance of VLMs in recognizing weather, particularly in rainy and snowy images. Although DACLIP achieves relatively high accuracy in predicting weather conditions via training, its performance on out-of-distribution data reveals limited generalizability, as reported in Table 2. In contrast, our pipeline excels in offering consistently precise and robust predictions across diverse weather scenarios.

Table 1. The accuracy comparison of weather condition prediction. With our developed <Definition>, our pipeline demonstrates markedly improved performance. All accuracies are computed over the full test set and reported as mean $\pm$ standard deviation over three runs, with one-decimal precision for readability.

Methods	<i>Hazy</i>	<i>Rainy</i>	<i>Snowy</i>	<i>Average</i>
VLMs w/o <Definition>	99.0%	77.5%	73.5%	88.1%
DACLIP (requires training)	100.0 $\pm$ 0.0%	97.5 $\pm$ 1.3%	98.7 $\pm$ 1.5%	99.3 $\pm$ 0.5%
Our pipeline	99.2%	82.0%	88.7%	93.8%

Table 2. The weather condition prediction on out-of-distribution dataset. Compared to DACLIP, our weather condition perceiving strategy exhibits higher accuracy and enhanced generalizability.

Methods	<i>Hazy</i>	<i>Rainy</i>	<i>Snowy</i>	<i>Average</i>
Pre-trained DACLIP	100.0%	100.0%	30.8%	80.9%
Our pipeline	100.0%	100.0%	84.6%	95.7%

3.2.3. Primary Tag Selection

Motivated by the success of semantic prompts in other image restoration tasks, we further explore the potential of semantic priors for adverse weather removal. Specifically, our exploration starts with DAPE [45], a robust degradation-aware extractor whose tag information has been successfully integrated into state-of-the-art image restoration methods. However, when DAPE is directly adopted to weather-affected images, its performance is not perfect as expected: (1) **Excessive tag length**: The predicted tags in DAPE are determined by the input image and a pre-defined threshold. For the image in Figure 3, DAPE tends to generate a large number of tags: “building”, “car”, “city”, “city street”, “drive”, “intersection”, “pole”, “road”, “stop light”, “traffic light”, “street scene”, “street sign”, and “urban”. However, since our framework relies on VLMs to generate tag-specific restoration instructions, such an extensive set of tags can introduce substantial computational overhead without proportionate gains in restoration performance. (2) **Redundant Tags**. The DAPE model frequently generates redundant tags due to its vocabulary containing many semantically similar terms. To enhance efficiency, it is crucial to maintain only one tag per semantic group to reduce unnecessary computational overhead.

To this end, we design a primary tag selection strategy as shown in Algorithm 1. First, instead adopting the fixed threshold, we rank all tags based on their probability and select top 8 tags for further processing. Subsequently, for the remaining tags, we ascertain their index within the vocabulary and we remove tags whose indices are within 50 positions relative to a tag of higher probability to reduce the redundancy.

As demonstrated in Figure 3, our proposed primary tag selection mechanism is capable of producing accurate and succinct tag information. Moreover, based on these tags, VLMs are prompted to provide restoration instructions globally and locally. Finally, the answers from VLMs are transformed into global and local priors ($\mathbf{P}_{global}$ and $\mathbf{P}_{local}$), which help achieve adaptive adverse weather removal in this paper. Overall, our introduced semantic and low-level priors extraction pipeline is summarized in Algorithm 2, where Prompt templates follow Figure 3 and are kept fixed for all experiments; weather-type perception is implemented as a forced-choice query over hazy, rainy, and snowy. Note that the VLM is only used to generate weather-aware textual priors (weather type and global/local instructions) in a training-free manner; the restoration is performed exclusively by AWR-VIP conditioned on these priors.

3.2.4. Flexible Semantic Elements and Controllable Inference

Although the proposed pipeline runs automatically by default, the semantic elements used for guidance are not restricted to a fixed set. Specifically, users can optionally override the predicted weather type and provide customized semantic tags as well as global/local instructions at inference time, enabling controllable restoration without retraining. This flexibility is particularly useful in challenging cases where the automatic tag extraction or weather perception may miss small but important objects or when user preferences favor different restoration styles (e.g., sharper details versus smoother appearance). In our framework, these user-provided texts are encoded by the same text encoder into priors ($\mathbf{P}_{global}, \mathbf{P}_{local}$) and seamlessly injected into AWR-VIP, following the same conditioning mechanism as the automatic mode.

Algorithm 2 VLM-guided Prior Extraction Pipeline

Require: Low-quality image $\mathbf{I}_{lq}$, frozen VLM $\mathcal{M}$ (LLaVA-1.5), pre-trained DAPE θ , CLIP text encoder $\mathcal{E}_{txt}$, and primary tag selection $\mathcal{S}(\cdot)$ (Algorithm 1).

```

1: Tag selection:  $\mathbf{K} \leftarrow \mathcal{S}(\theta(\mathbf{I}_{lq}))$  ▷ Top-K compact tags.
2: Weather-type perception:
3:   Construct prompt  $\mathcal{Q}_w$  with weather definitions and multiple-choice options.
4:    $\hat{w} \leftarrow \mathcal{M}(\mathbf{I}_{lq}, \mathcal{Q}_w)$  ▷  $\hat{w} \in \{\text{Hazy, Rainy, Snowy}\}$ .
5: Global instruction generation:
6:   Construct prompt  $\mathcal{Q}_g$ .
7:    $\mathbf{s}_g \leftarrow \mathcal{M}(\mathbf{I}_{lq}, \mathcal{Q}_g)$ 
8: Local instruction generation:
9:   Define LocalText = []
10:  for each tag  $k$  in  $\mathbf{K}$  do
11:    Construct prompt  $\mathcal{Q}_l^k$  and  $k$ .
12:     $\mathbf{s}_l^k \leftarrow \mathcal{M}(\mathbf{I}_{lq}, \mathcal{Q}_l^k)$ 
13:    LocalText.append( $\mathbf{s}_l^k$ )
14:  end
15: Prior encoding:
16:   $\mathbf{P}_{global} \leftarrow \mathcal{E}_{txt}(\hat{w}, \mathbf{K}, \mathbf{s}_g)$ 
17:   $\mathbf{P}_{local} \leftarrow \mathcal{E}_{txt}(\mathbf{LocalText})$ 
18: return  $\mathbf{P}_{global}, \mathbf{P}_{local}$ 

```

3.3. Priors Guided Adverse Weather Removal Network

To fully exploit the priors extracted by our pipeline, we also design an all-in-one network with several special designs.

3.3.1. Overview

The overall framework of our adverse weather removal network (AWR-VIP) is shown as Figure 2. Notably, AWR-VIP serves as the main restoration backbone that produces the final clean output, while the VLM branch provides auxiliary conditioning priors to adapt the restoration behavior under different adverse weather degradations. Specifically, the input image is first processed by a 3×3 convolution to capture its feature embedding. Then, this shadow feature is fed into 4-level encoder-decoder architecture to obtain the deep representation. Finally, another 3×3 convolution is applied, and the residual term is added to obtain the restored image. In addition, each level of encoder-decoder contains multiple weather blocks, and adaptive mix-up modules are adopted to replace the skip connection operation. Moreover, we introduce several special designs in each weather block to effectively incorporate the priors extracted by our proposed pipeline.

3.3.2. Layer Norm (LN) Modulated by Global Prior

To effectively integrate the global prior $\mathbf{P}_{global}$ derived from VLMs into our weather block, we modulate the layer normalization process. This modulation is driven by scale and shift parameters (α and β) conditioned on $\mathbf{P}_{global}$, enabling the normalization process to better reflect the semantic and perceptual information encoded in the global prior. Given an input feature $\mathbf{z}_{in}$, the output of modulated LN is calculated by $\mathbf{z}_{out} = \alpha \cdot \text{LN}(\mathbf{z}_{in}) + \beta$, where $\alpha, \beta = \text{Linear}(S)$. Notably, each WeatherBlock contains three modulated-layer normalization layers; therefore, the linear projection outputs three pairs of modulation parameters, denoted as $\{\alpha_1, \beta_1\}$, $\{\alpha_2, \beta_2\}$, and $\{\alpha_3, \beta_3\}$, respectively.

3.3.3. Instruction-Guided Cross-Attention (IGCA)

To reduce the huge computational cost caused by spatial self-attention mechanism, we adopt simplified cross-attention (SCA [46]) to calculate the attention map along the channel dimension in our weather block. Moreover, we develop the channel attention mechanism guided by the local prior $\mathbf{P}_{local}$. Specifically, *query* element $\mathbf{Q}$ is calculated upon the input feature, while the calculation of *key* and *value* elements ($\mathbf{K}$, $\mathbf{V}$) (see Figure 2) are guided by the local prior $\mathbf{P}_{local}$.

4. Experiments

4.1. Experiment Settings

To train our proposed adverse weather network, we first build a combined dataset that includes hazy, rainy, and snowy images and their corresponding clean counterparts. The source datasets are Reside-6K [47], Rain100H [48], and Snow100K-L [49], respectively. In our combined dataset, we have 6000 hazy-clean, 1800 rainy-clean, and 1872 snowy-clean image pairs for training, and the other 1000 hazy-clean, 200 rainy-clean, and 601 snowy-clean image pairs for evaluation. Our adverse weather network is trained using the Adam optimizer with the total training iterations of 200 K with the batch size of 8. The initial learning rate is set to 10^{-3} and it is reduced to 10^{-7} by the end of training. Each training input is cropped into 256×256 , and we use horizontal flips and rotations for data augmentation. Two metrics (PSNR and SSIM) are calculated for quantitative evaluation. Besides the test samples in this combined dataset, we also evaluate the generalization of our method on multiple real-world images without ground-truth correspondences (BeDDE [50], SPA+ [16], and RealSnow [16]).

Software and Hardware Environment: All experiments are conducted in a conda environment with Python 3.8, PyTorch 1.13.0, pytorch-lightning 1.9.0, and CUDA 11.7. All experiments were run on two NVIDIA RTX 2080 Ti GPUs.

4.2. Performances and Comparisons

We compare our method with several approaches: Restormer [51], WGWS-Net [16], TransWeather [17], NAFNet [46], HistoFormer [52], HOGFormer [53], and DACLIP [24]. Restormer, HistoFormer, and HOGFormer are advanced transformer-based networks. WGWS-Net proposes to learn weather-general and weather-specific features, whereas DACLIP presents a degradation-aware VLM that guides the SDE-based restoration model.

4.2.1. Quantitative Results

Table 3 summarizes the statistical results of our method and above models. Our method achieves the superior performance on all metrics, highlighting its advanced and perfect generalization. Notably, the PSNR of our method surpasses the best of SOTA method (HistoFormer) by 1.31 dB. Moreover, we observe enhanced performance by integrating our

extracted priors into existing methods. These numbers demonstrate the effectiveness and generalization of semantic and visual perceptual priors in our method.

Table 3. Quantitative comparisons on the combined dataset; our method achieves superior performance compared to current SOTA methods. Moreover, our extracted priors can be integrated into current SOTA methods for performance enhancement. [Key: $\uparrow$: the larger the better; **Best**, **Second Best**; *: trained with our extracted priors; † : replacing the degradation and content embedding with our extracted priors. Results for AWR-VIP are reported as mean $\pm$ standard deviation over three runs].

Methods	Hazy		Rainy		Snowy		Average	
	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$
Restormer	24.38	0.911	22.45	0.749	24.29	0.806	24.14	0.858
Restormer *	26.10	0.929	24.36	0.810	26.36	0.838	25.99	0.885
WGWS-Net	25.24	0.920	25.44	0.806	26.28	0.833	25.61	0.878
WGWS-Net *	26.58	0.943	26.85	0.833	27.15	0.875	26.80	0.908
TransWeather	28.87	0.945	24.30	0.815	26.95	0.877	27.72	0.908
TransWeather *	29.47	0.951	26.17	0.830	27.53	0.889	28.46	0.917
NAFNet	29.28	0.948	26.03	0.849	28.19	0.895	28.56	0.920
NAFNet *	29.97	0.954	27.52	0.861	28.37	0.891	29.16	0.923
HOGFormer	26.23	0.945	25.90	0.823	26.26	0.879	26.20	0.909
HOGFormer *	27.16	0.954	26.84	0.837	27.21	0.888	27.14	0.919
HistoFormer	28.88	0.953	27.26	0.855	27.39	0.887	28.20	0.920
HistoFormer *	29.95	0.965	27.87	0.867	28.06	0.892	29.09	0.930
DACLIP	29.12	0.937	26.80	0.850	27.03	0.870	28.16	0.905
DACLIP †	29.70	0.954	27.43	0.877	27.59	0.890	28.74	0.924
AWR-VIP (Ours)	30.55 $\pm$ 0.05	0.957 $\pm$ 0.002	28.15 $\pm$ 0.06	0.883 $\pm$ 0.003	28.22 $\pm$ 0.05	0.895 $\pm$ 0.005	29.51 $\pm$ 0.05	0.928 $\pm$ 0.003

4.2.2. Qualitative Results

As depicted in Figures 4 and 5, our method demonstrates a marked improvement in image clarity and color fidelity over competing methods such as DACLIP and HistoFormer. Notably, our AWR-VIP effectively mitigates the over-saturation observed in HistoFormer's outputs and achieves a harmonious balance of sharpness and color correctness, closely approximating the ground truth. These visualizations illustrate the superiority of our method and validate the effectiveness of our weather removal network and our extracted priors.

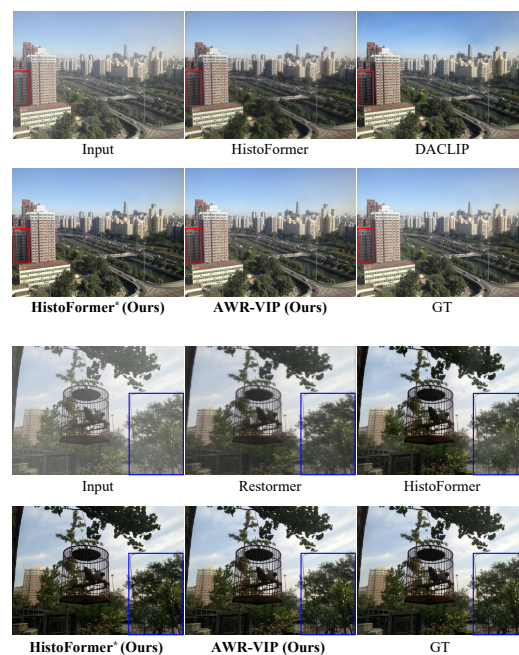


Figure 4. Visual comparisons under hazy scenarios. Our VAR-LIDE is capable of preserving fine details and maintaining color fidelity (see regions highlighted using red or blue boxes). In addition, our extracted semantic and low-level prior significantly enhances SOTA method (HistoFormer).

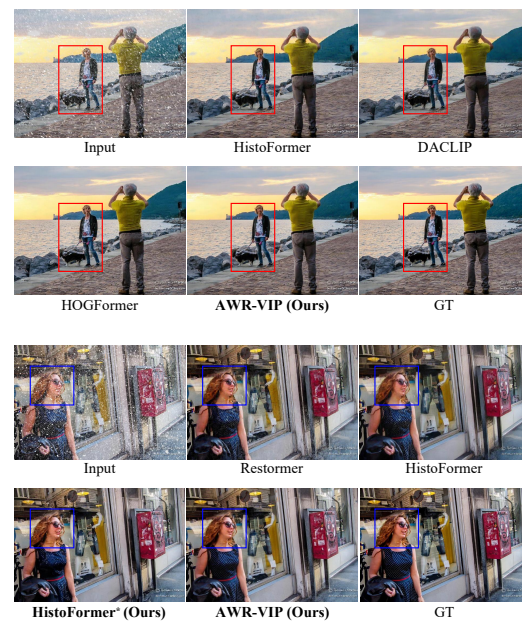


Figure 5. Visual comparisons on snowy scenarios, where our AWR-VIP excels in structure restoration. These two scenarios are from Snow100K-L test set (link: <https://drive.google.com/drive/folders/1Ox7Fj4WVophj5YBmKfUVH8YdIp7NQfer> accessed on 10 November 2025).

4.3. Ablation Study

To evaluate the importance of each component, we conduct ablation experiments by removing each module from our method. As presented in Table 4, though removing the local instruction can still achieve competitive result, its performance is significantly inferior to our full model, demonstrating that local instruction plays an important role in our method. Similarly, we also observe distinct performance decrease when the global instruction or the weather type are removed. All these numbers in Table 4 demonstrate the positive contributions of each component to our final satisfactory outcomes.

Table 4. Ablation results. Removing one or multiple components from our method leads to poor performance.

Configurations	PSNR	SSIM
Full AWR-VIP	29.51	0.928
w/o <Local Instruction> (P_{local})	28.62	0.914
w/o <Global Instruction>	29.16	0.919
w/o <Weather Type>	28.85	0.916
Baseline (w/o P_{local} and P_{global})	28.20	0.902

4.4. Computational Efficiency

The efficiency comparisons between AWR-VIP and other methods are reported in Table 5. Notably, in our AWR-VIP, 90% of the parameters come from language model with extremely high efficiency. Therefore, employing the VLM only results in 0.23 s inference delay compared to the Baseline, whereas DACLIP suffers from 18.4 s inference time due to its iterative diffusion process. Additionally, we emphasize that our superior performance primarily arises from methodological advancements, as scaling the Baseline's hidden dimension to match AWR-VIP's runtime (denoted as "Baseline *" in Table 5) yields significantly inferior performance (PSNR: 28.57; SSIM: 0.910) compared to our complete AWR-VIP.

Table 5. Computational efficiency for 512×512 input on one RTX 2080Ti GPU. Simply increasing the Baseline’s hidden dimension to match AWR-VIP’s runtime (denoted as Baseline *) yields only 28.57 dB in PSNR and 0.910 in SSIM, far below AWR-VIP’s result in Table 3.

Methods	HistoFormer	DACLIP	Baseline *	AWR-VIP (Baseline + VLM)
Params (M)	16.6	174	67.8	18.2 + 467
FLOPs (G)	366	474	291	80.5 + 950
Runtime (s)	0.80	18.4	0.32	0.08 + 0.23

4.5. Evaluations on Real-World Data and Generalization to More Restoration Tasks

4.5.1. Visual Comparisons with DACLIP on Real Hazy Images

Figure 6 provides visual comparisons between our AWR-VIP and DACLIP on real hazy images. Notably, DACLIP is capable of producing visually compelling results under hazy conditions [50]. This outcome can be analyzed using two factors: (1) its diffusion-based nature, which enhances visual quality despite lower PSNR and SSIM; (2) haze is DACLIP’s best-performing condition, aligning most closely with our perceived weather type. However, in other scenarios, DACLIP struggles with accurate weather prediction and is prone to such misclassification, leading to limited performance. More importantly, even in such a DACLIP-preferred hazy scenario, our AWR-VIP, guided by global and local priors, outperforms DACLIP and achieves enhanced clarity and contrast.

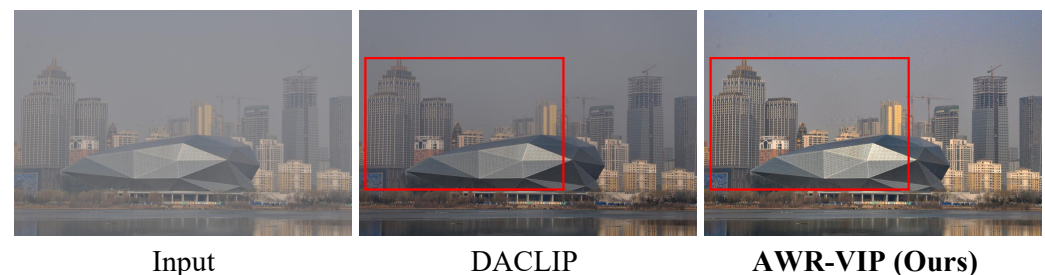


Figure 6. Visual comparisons on real-world data. Thanks to our extracted semantic and low-level priors, our AWR-VIP significantly outperforms DACLIP (see regions highlighted using red boxes).

4.5.2. Generalization to Broader Restoration Tasks

Besides common weather-related restoration tasks (dehazing, deraining, and desnowing), we add one more degradation (low-light, LOL dataset [54]) by re-designing the definition and prompt in our extraction pipeline. With this modification, our AWR-VIP retains over 90% degradation prediction accuracy and delivers satisfactory enhancements as presented in Figure 7, highlighting its potential in handling diverse degradations similarly to DACLIP while remaining training-free. Notably, AWR-VIP achieves comparable or even better visual effects compared to LLIE-specific methods (ECMamba [55] and GALRE [56]), highlighting the generalization and effectiveness of our proposed prior extraction pipeline in Figure 3.

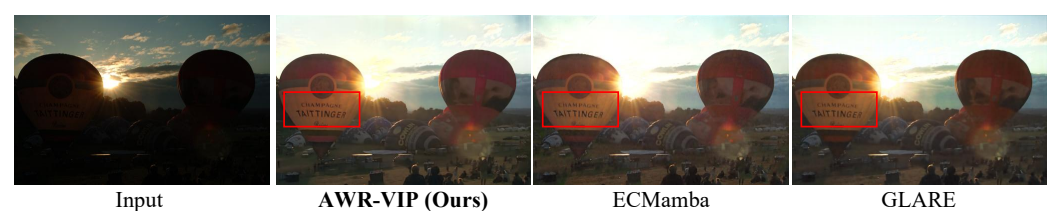


Figure 7. Visual comparisons on unpaired real-world low-light images (MEF dataset). Our AWR-VIP effectively enhance the visibility and contrast, achieving comparable or even better performance than LLIE-specific methods (see red boxes).

4.5.3. Performance on Extreme Weather Conditions

Beyond the standard evaluation setting, we further investigate the robustness of AWR-VIP under out-of-distribution (OOD) *extreme* adverse weather conditions, where degradations are significantly stronger and may be accompanied by haze patterns [57]/challenging illumination [58]. Such scenarios are well-known to be difficult for conventional restoration pipelines, which often suffer from inaccurate degradation perception and tend to produce over-smoothed results or residual artifacts.

Figure 8 presents representative qualitative comparisons between DACLIP and our AWR-VIP. As shown in the highlighted regions, DACLIP may introduce noticeable artifacts or suppress fine structures when the degradation becomes extreme, while our method yields clearer structures and a more natural visual appearance. We attribute the improvement to the explicit global/local restoration guidance derived from the VLM, which provides content-aware priors to anchor the restoration behavior even under OOD conditions. Overall, these results indicate that AWR-VIP generalizes favorably to extreme weather conditions, demonstrating improved robustness in challenging real-world cases.

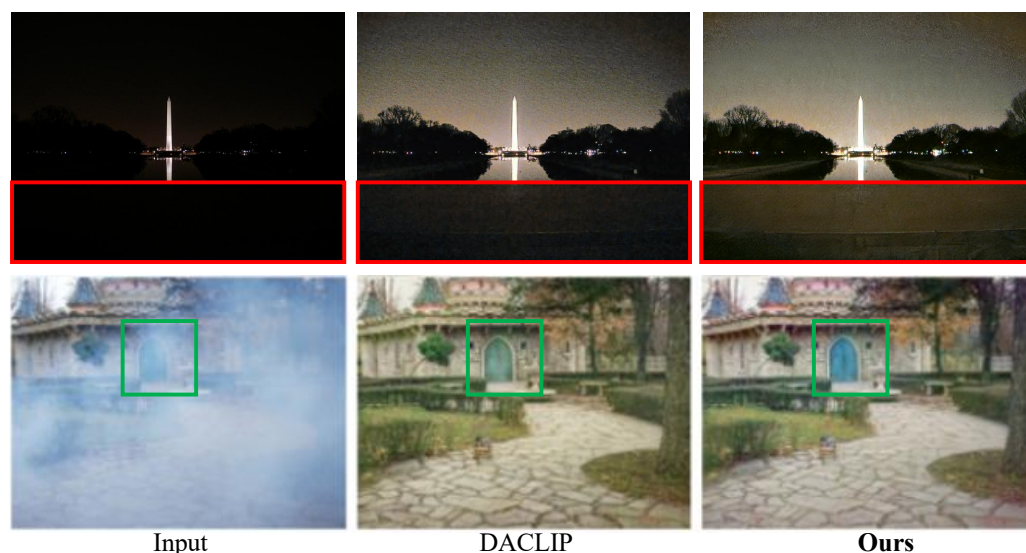


Figure 8. Visual comparisons on out-of-distribution *extreme* adverse weather conditions. The highlighted boxes highlight show that our method better preserves semantic structures and improve local clarity in extreme scenarios in highlighted red and green boxes, compared to DACLIP.

4.6. Incorporating Conventional Low-Level Priors: Blur Scalar Guidance

While our framework mainly leverages VLM-derived semantic and perceptual priors, we note that classical image processing techniques can provide complementary low-level cues, which are particularly useful under complex degradations. As a concrete example, we incorporate a lightweight *blur scalar* as an additional low-level prior.

4.6.1. Blur Scalar Estimation

Given an input image I_{lq} , we first convert it to grayscale and compute its Laplacian response ΔI_{lq} . We then estimate the blur level via the inverse variance of the Laplacian:

$$b = \frac{1}{\text{Var}(\Delta I_{lq}) + \epsilon}, \quad (1)$$

where ϵ is a small constant for numerical stability. Intuitively, a sharper image yields stronger high-frequency responses and thus a larger Laplacian variance, resulting in a smaller b ; conversely, more defocus blur leads to a larger b .

4.6.2. Integration into AWR-VIP

We treat b as a global low-level prior and inject it into AWR-VIP through the same global conditioning interface. Specifically, we concatenate b with the extracted global prior $\mathbf{P}_{global}$ and use an MLP to produce the modulation parameters for the modulated layer normalization in each WeatherBlock:

$$\{\alpha_i, \beta_i\}_{i=1}^3 = f_{MLP}\left(\left[\mathbf{P}_{global}; b\right]\right). \quad (2)$$

In this way, the restoration backbone can adapt its global normalization behavior according to both semantic guidance and the estimated blur severity.

By incorporating this blur scalar prior, we observe a slight improvement (PSNR: 29.62 dB) in restoration quality, suggesting that conventional low-level cues can complement VLM-derived guidance. This result indicates the potential of enriching AWR-VIP with additional classical priors (e.g., noise/exposure-related cues) to further enhance robustness; however, identifying the most effective priors and designing reliable estimators remain important directions to explore.

5. Conclusions

To achieve adaptive adverse weather removal in real-world scenarios, we propose **AWR-VIP**, a unified framework guided by semantic and visual perceptual priors. Specifically, we develop a VLM-based prior extraction pipeline by designing weather definitions and prompt-driven querying to perceive the weather type and generate restoration instructions at both global and local levels. To obtain accurate yet concise semantic descriptions, we further introduce a primary tag selection strategy that reduces redundancy and computational overhead. Moreover, we design a UNet-based restoration backbone equipped with a weather-aware module, where global priors modulate layer normalization for image-level adaptation and local priors guide cross-attention to refine key semantic regions. Extensive experiments on combined benchmarks validate the effectiveness of AWR-VIP and demonstrate that the extracted priors are transferable to other restoration backbones.

Overall, the main contributions of this work can be summarized as follows: (1) a VLM-guided prior extraction pipeline that produces weather-type perception, compact semantic tags, and global/local restoration instructions as semantic and perceptual priors; (2) a unified adverse weather removal framework, AWR-VIP, that performs restoration by conditioning a UNet-based backbone on the extracted priors; (3) two complementary prior injection mechanisms via global-prior-modulated normalization and local-prior-guided cross-attention for content-aware refinement; (4) a user-controllable inference mode that allows overriding the automatically generated prompts (weather type and instructions) to achieve preference-driven restoration without retraining; (5) extensive experimental results showing state-of-the-art performance and strong plug-and-play applicability of the extracted priors across different restoration architectures.

Limitations and Future Work: A potential limitation of our method is its reliance on the performance of VLMs. In some extreme scenarios, the efficacy of VLMs may diminish, thereby adversely affecting the overall performance of our method. In addition, differing from the hard weather-type perception proposed in this paper, we are currently investigating a soft perception mechanism and we plan to present it in the future work, which allows for flexible restoration for images with multiple degradations (e.g., “rainy + snowy”).

Author Contributions: Conceptualization, W.D., H.Z., and J.C.; methodology, W.D. and H.Z.; software, W.D.; validation, H.Z.; formal analysis, W.D., H.Z., and T.J.; investigation, W.D., H.Z., and J.C.; resources, J.C.; data curation, W.D.; writing—original draft preparation, W.D. and T.J.; writing—review and editing, H.Z.; visualization, W.D., H.Z., and T.J.; supervision, J.C. and H.Z.;

project administration, H.Z.; funding acquisition, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets used in this study are publicly available. The combined version for Reside-6K, Rain100H, and Snow100K-L is available at https://drive.google.com/file/d/1k7QYg215cGpS8xvRu0k9wO8YKv7_vzgx/view?usp=sharing accessed on 10 November 2025.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
DDPM	Denoising Diffusion Probabilistic Models
AWR	Adverse Weather Restoration
OOD	Out-Of-Distribution
SOTA	State-Of-The-Art
VLM	Vision Language Models
LN	Layer Norm
IGCA	Instruction-Guided Cross-Attention
SCA	Simplified Cross Attention
PSNR	Peak Signal-To-Noise Ratio
SSIM	Structural Similarity Index Measure

References

1. Zang, S.; Ding, M.; Smith, D.; Tyler, P.; Rakotoarivelo, T.; Kaafar, M.A. The impact of adverse weather conditions on autonomous vehicles: How rain, snow, fog, and hail affect the performance of a self-driving car. *IEEE Veh. Technol. Mag.* **2019**, *14*, 103–111. [\[CrossRef\]](#)
2. Li, R.; Cheong, L.-F.; Tan, R.T. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2019; pp. 1633–1642.
3. Zhang, R.; Yu, J.; Chen, J.; Li, G.; Lin, L.; Wang, D. A Prior Guided Wavelet-Spatial Dual Attention Transformer Framework for Heavy Rain Image Restoration. *IEEE Trans. Multimed.* **2024**, *26*, 7043–7057. [\[CrossRef\]](#)
4. Chen, W.T.; Fang, H.Y.; Hsieh, C.L.; Tsai, C.C.; Chen, I.; Ding, J.J.; Kuo, S.Y. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: Piscataway, NJ, USA, 2021; pp. 4196–4205.
5. Zhang, K.; Li, R.; Yu, Y.; Luo, W.; Li, C. Deep dense multi-scale network for snow removal using semantic and depth priors. *IEEE Trans. Image Process.* **2021**, *30*, 7419–7431. [\[CrossRef\]](#)
6. Li, B.; Peng, X.; Wang, Z.; Xu, J.; Feng, D. AOD-Net: All-in-one dehazing network. In *2017 IEEE International Conference on Computer Vision (ICCV)*; IEEE: Piscataway, NJ, USA, 2017; pp. 4770–4778.
7. Cai, B.; Xu, X.; Jia, K.; Qing, C.; Tao, D. DehazeNet: An end-to-end system for single image haze removal. *IEEE Trans. Image Process.* **2016**, *25*, 5187–5198. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Li, B.; Gou, Y.; Liu, J.Z.; Zhu, H.; Zhou, J.T.; Peng, X. Zero-shot image dehazing. *IEEE Trans. Image Process.* **2020**, *29*, 8457–8466. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Song, Y.; He, Z.; Qian, H.; Du, X. Vision transformers for single image dehazing. *IEEE Trans. Image Process.* **2023**, *32*, 1927–1941. [\[CrossRef\]](#)
10. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*; Curran Associates, Inc.: Red Hook, NY, USA, 2012; Volume 25, pp. 1097–1105.
11. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30, pp. 5998–6008.
12. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 6840–6851.
13. Fu, X.; Huang, J.; Ding, X.; Liao, Y.; Paisley, J. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Trans. Image Process.* **2017**, *26*, 2944–2956. [\[CrossRef\]](#)

14. Zhou, H.; Dong, W.; Chen, J. LITA-GS: Illumination-Agnostic Novel View Synthesis via Reference-Free 3D Gaussian Splatting and Physical Priors. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2025; pp. 21580–21589.
15. Li, R.; Tan, R.T.; Cheong, L.-F. All in one bad weather removal using architectural search. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2020; pp. 3175–3185.
16. Zhu, Y.; Wang, T.; Fu, X.; Yang, X.; Guo, X.; Dai, J.; Qiao, Y.; Hu, X. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2023; pp. 21747–21758.
17. Valanarasu, J.M.J.; Yasarla, R.; Patel, V.M. TransWeather: Transformer-based restoration of images degraded by adverse weather conditions. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2022; pp. 2353–2363.
18. Ye, T.; Chen, S.; Bai, J.; Shi, J.; Xue, C.; Jiang, J.; Yin, J.; Chen, E.; Liu, Y. Adverse weather removal with codebook priors. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: Piscataway, NJ, USA, 2023; pp. 12653–12664.
19. Yang, H.; Pan, L.; Yang, Y.; Liang, W. Language-driven All-in-one Adverse Weather Removal. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2024; pp. 24902–24912.
20. Potlapalli, V.; Zamir, S.W.; Khan, S.; Khan, F.S. PromptIR: Prompting for All-in-One Blind Image Restoration. In *Advances in Neural Information Processing Systems 36 NeurIPS 2023*; Curran Associates, Inc.: Red Hook, NY, USA, 2023; Volume 36, pp. 71275–71293.
21. Xu, Y.; Gao, N.; Zhong, Y.; Chao, F.; Ji, R. Unified-Width Adaptive Dynamic Network for All-In-One Image Restoration. *arXiv* **2024**, arXiv:2401.13221. [[CrossRef](#)]
22. Hu, J.; Jin, L.; Yao, Z.; Lu, Y. Universal Image Restoration Pre-training via Degradation Classification. *arXiv* **2025**, arXiv:2501.15510. [[CrossRef](#)]
23. Jiang, Y.; Zhang, Z.; Xue, T.; Gu, J. AutoDir: Automatic all-in-one image restoration with latent diffusion. In *Computer Vision–ECCV 2024*; Springer: Cham, Switzerland, 2024; pp. 340–359.
24. Luo, Z.; Gustafsson, F.K.; Zhao, Z.; Sjölund, J.; Schön, T.B. Controlling vision-language models for universal image restoration. *arXiv* **2023**, arXiv:2310.01018.
25. Zeng, H.; Wang, X.; Chen, Y.; Su, J.; Liu, J. Vision-Language Gradient Descent-driven All-in-One Deep Unfolding Networks. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2025; pp. 7524–7533.
26. Dong, W.; Zhou, H.; Wang, R.; Liu, X.; Zhai, G.; Chen, J. DehazeDCT: Towards Effective Non-Homogeneous Dehazing via Deformable Convolutional Transformer. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; IEEE: Piscataway, NJ, USA, 2024; pp. 6405–6414.
27. Zhou, H.; Dong, W.; Liu, Y.; Chen, J. Breaking Through the Haze: An Advanced Non-Homogeneous Dehazing Method Based on Fast Fourier Convolution and ConvNeXt. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; IEEE: Piscataway, NJ, USA, 2023; pp. 1895–1904.
28. Ancuti, C.O.; Ancuti, C.; Vasluianu, F.-A.; Timofte, R.; Zhou, H.; Dong, W.; Liu, Y.; Chen, J.; Liu, H.; Li, L.; et al. NTIRE 2023 HR Nonhomogeneous Dehazing Challenge Report. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; IEEE: Piscataway, NJ, USA, 2023; pp. 1808–1825.
29. Ancuti, C.O.; Ancuti, C.; Vasluianu, F.-A.; Timofte, R.; Liu, Y.; Wang, X.; Zhu, Y.; Shi, G.; Lu, X.; Fu, X.; et al. NTIRE 2024 Dense and Non-Homogeneous Dehazing Challenge Report. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; IEEE: Piscataway, NJ, USA, 2024; pp. 6453–6468.
30. Li, Z.; Lei, Y.; Ma, C.; Zhang, J.; Shan, H. Prompt-In-Prompt Learning for Universal Image Restoration. *arXiv* **2023**, arXiv:2312.05038.
31. Kong, X.; Dong, C.; Zhang, L. Towards Effective Multiple-in-One Image Restoration: A Sequential and Prompt Learning Strategy. *arXiv* **2024**, arXiv:2401.03379. [[CrossRef](#)]
32. Chen, Y.-W.; Pei, S.-C. Always Clear Days: Degradation Type and Severity Aware All-In-One Adverse Weather Removal. *IEEE Access* **2025**, *13*, 7650–7662. [[CrossRef](#)]
33. Guo, Y.; Gao, Y.; Lu, Y.; Zhu, H.; Liu, R.W.; He, S. OneRestore: A Universal Restoration Framework for Composite Degradation. In *Computer Vision–ECCV 2024*; Springer: Cham, Switzerland, 2024; pp. 255–272.
34. Rajagopalan, S.; Patel, V.M. AWRaCLe: All-Weather Image Restoration Using Visual In-Context Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Philadelphia, PA, USA, 25 February–4 March 2025; Volume 39, pp. 6675–6683.
35. Bai, Y.; Wang, C.; Xie, S.; Dong, C.; Yuan, C.; Wang, Z. TextIR: A Simple Framework for Text-Based Editable Image Restoration. *arXiv* **2023**, arXiv:2302.14736. [[CrossRef](#)]
36. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*; PMLR: Cambridge MA, USA, 2021; Volume 139, pp. 8748–8763.
37. Brooks, T.; Holynski, A.; Efros, A.A. InstructPix2Pix: Learning to Follow Image Editing Instructions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; PMLR: Cambridge MA, USA, 2023; pp. 18392–18402.

38. Conde, M.V.; Geigle, G.; Timofte, R. InstructIR: High-Quality Image Restoration Following Human Instructions. In *Computer Vision–ECCV 2024*; Springer: Cham, Switzerland, 2024; pp. 1–21.
39. Wu, H.; Zhu, H.; Zhang, Z.; Zhang, E.; Chen, C.; Liao, L.; Li, C.; Wang, A.; Sun, W.; Yan, Q.; et al. Towards Open-Ended Visual Quality Comparison. In *Computer Vision–ECCV 2024*; Springer: Cham, Switzerland, 2024; pp. 360–377.
40. Liu, S.; Ma, J.; Sun, L.; Kong, X.; Zhang, L. InstructRestore: Region-Customized Image Restoration with Human Instructions. *arXiv* **2025**, arXiv:2503.24357.
41. Wu, H.; Zhang, Z.; Zhang, E.; Chen, C.; Liao, L.; Wang, A.; Xu, K.; Li, C.; Hou, J.; Zhai, G.; et al. Q-Instruct: Improving Low-level Visual Abilities for Multi-modality Foundation Models. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2024; pp. 25490–25500.
42. Zhou, H.; Dong, W.; Liu, X.; Zhang, Y.; Zhai, G.; Chen, J. Low-light Image Enhancement via Generative Perceptual Priors. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, Philadelphia, PA, USA, 25 February–4 March 2025; pp. 10752–10760.
43. Dong, W.; Zhou, H.; Lin, J.; Chen, J. Zero-Reference Joint Low-Light Enhancement and Deblurring via Visual Autoregressive Modeling with VLM-Derived Modulation. *arXiv* **2025**, arXiv:2511.18591.
44. Liu, H.; Li, C.; Wu, Q.; Lee, Y.J. Visual Instruction Tuning. In *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*; Curran Associates, Inc.: Red Hook, NY, USA, 2023; Volume 36, pp. 34892–34916.
45. Wu, R.; Yang, T.; Sun, L.; Zhang, Z.; Li, S.; Zhang, L. SeeSR: Towards Semantics-Aware Real-World Image Super-Resolution. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE: Piscataway, NJ, USA, 2024; pp. 25456–25467.
46. Chen, L.; Chu, X.; Zhang, X.; Sun, J. Simple Baselines for Image Restoration. In *Computer Vision–ECCV 2022*; Springer: Cham, Switzerland, 2022; pp. 17–33.
47. Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; Wang, Z. Benchmarking Single-Image Dehazing and Beyond. *IEEE Trans. Image Process.* **2018**, *28*, 492–505. [[CrossRef](#)] [[PubMed](#)]
48. Yang, W.; Tan, R.T.; Feng, J.; Liu, J.; Guo, Z.; Yan, S. Deep Joint Rain Detection and Removal From a Single Image. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2017; pp. 1357–1366.
49. Liu, Y.-F.; Jaw, D.-W.; Huang, S.-C.; Hwang, J.-N. DesnowNet: Context-Aware Deep Network for Snow Removal. *IEEE Trans. Image Process.* **2018**, *27*, 3064–3073. [[CrossRef](#)]
50. Zhao, S.; Zhang, L.; Huang, S.; Shen, Y.; Zhao, S. Dehazing Evaluation: Real-World Benchmark Datasets, Criteria, and Baselines. *IEEE Trans. Image Process.* **2020**, *29*, 6947–6962. [[CrossRef](#)]
51. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.-H. Restormer: Efficient Transformer for High-Resolution Image Restoration. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2022; pp. 5728–5739.
52. Sun, S.; Ren, W.; Gao, X.; Wang, R.; Cao, X. Restoring Images in Adverse Weather Conditions via Histogram Transformer. In *Computer Vision–ECCV 2024*; Springer: Cham, Switzerland, 2024; pp. 111–129.
53. Wu, J.; Yang, Z.; Wang, Z.; Jin, Z. Beyond Degradation Conditions: All-in-One Image Restoration via HOG Transformers. *arXiv* **2025**, arXiv:2504.09377. [[CrossRef](#)]
54. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep Retinex Decomposition for Low-Light Enhancement. *arXiv* **2018**, arXiv:1808.04560. [[CrossRef](#)]
55. Dong, W.; Zhou, H.; Zhang, Y.; Liu, X.; Chen, J. ECMamba: Consolidating Selective State Space Model with Retinex Guidance for Efficient Multiple Exposure Correction. In *Advances in Neural Information Processing Systems (NeurIPS)*; Curran Associates, Inc.: Red Hook, NY, USA, 2024; pp. 53438–53457.
56. Zhou, H.; Dong, W.; Liu, X.; Liu, S.; Min, X.; Zhai, G.; Chen, J. Glare: Low Light Image Enhancement via Generative Latent Feature Based Codebook Retrieval. In *Computer Vision–ECCV 2024*; Springer: Cham, Switzerland, 2024; pp. 36–54.
57. Ancuti, C.O.; Ancuti, C.; Timofte, R. NH-HAZE: An Image Dehazing Benchmark with Non-Homogeneous Hazy and Haze-Free Images. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; IEEE: Piscataway, NJ, USA, 2020; pp. 444–445.
58. Lee, C.; Lee, C.; Kim, C. Contrast Enhancement Based on Layered Difference Representation of 2D Histograms. *IEEE Trans. Image Process.* **2013**, *22*, 5372–5384. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.