Language-driven All-in-one Adverse Weather Removal

Hao Yang[1], Liyuan Pan[1] †, Yan Yang[2,3], and Wei Liang[1]
[1]Beijing Institute of Technology [2] Australian National University [3]A&F, CSIRO
{hao.yang, liyuan.pan, liangwei}@bit.edu.cn, {yan.yang}@anu.edu.au

Abstract

*All-in-one (AiO) frameworks restore various adverse weather degradations with a single set of networks jointly. To handle various weather conditions, an AiO framework is expected to adaptively learn weather-specific knowledge for different degradations and shared knowledge for common patterns. However, existing methods: 1) rely on extra supervision signals, which are usually unknown in real-world applications; 2) employ fixed network structures, which restrict the diversity of weather-specific knowledge. In this paper, we propose a Language-driven Restoration framework (LDR) to alleviate the aforementioned issues. First, we leverage the power of pre-trained vision-language (PVL) models to enrich the diversity of weather-specific knowledge by reasoning about the occurrence, type, and severity of degradation, generating description-based degradation priors. Then, with the guidance of degradation prior, we sparsely select restoration experts from a candidate list dynamically based on a Mixture-of-Experts (MoE) structure. This enables us to adaptively learn the weather-specific and shared knowledge to handle various weather conditions (*e.g.*, unknown or mixed weather). Experiments on extensive restoration scenarios show our superior performance.*

1. Introduction

Imaging under unpleasant degradation conditions poses challenges for vision-based systems, such as self-driving cars [15, 33, 34, 46] and outdoor surveillance systems [9, 12], which require 24/7 service regardless of adverse weather conditions like rain [6, 13, 47], haze [14, 38, 41], and snow [4, 26, 29].. To meet the safety demand, compared to tackling each type of weather condition with independent models, the joint task process, *i.e.*, all-in-one framework, has a broader application scenario [17, 35].

Formally, degraded images can be modeled as masked additive combination of clean images and degradation residuals [35]. Consequently, several works [5, 21, 42] use a single network for all degradation types. Though shared knowledge is learned for restoration, they neglect that different degradations still hold different mathematical formulations, *e.g.*, transmission map produced by scattering effect in haze model [41] is unnecessary for raindrop model [37].

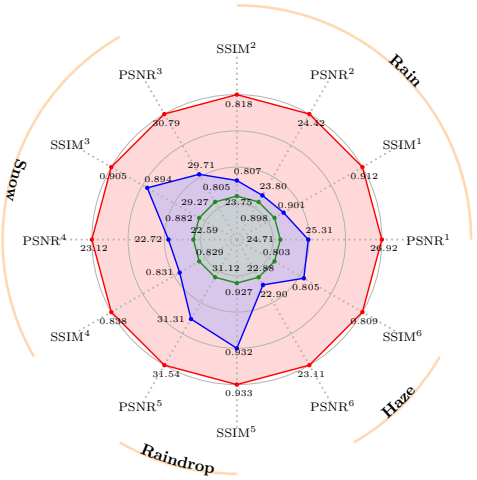
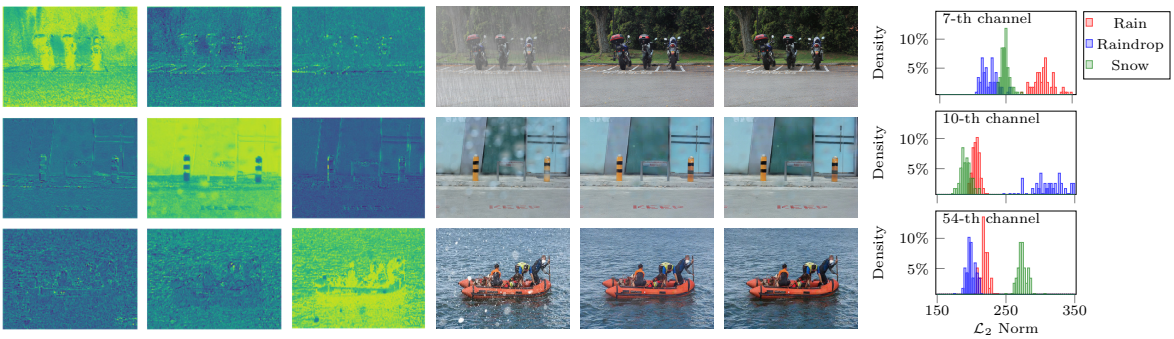


Figure 1. *Score (PSNR and SSIM) comparisons. We compare our model (red) with the top 2 (blue and green) baselines on benchmark datasets with various weather scenarios. Superscripts besides evaluation metrics are used to differentiate datasets.*

Hence, several works [17, 35, 36, 50, 57] use different sub-networks for weather-specific knowledge learning. However, auxiliary supervisions are required to assign the sub-networks, *e.g.*, degradation types [35] or depth maps [57]. Furthermore, the fixed sub-network architecture of existing methods restricts the diversity of learned weather-specific knowledge and their ability to handle images with various weather conditions [43, 55], such as images degraded by weather severity or conditions that have not been encountered before [1, 37], or mixed weather conditions like snow with haze in real-world scenarios.

In this paper, we question – can we restore the image degraded by various weather conditions by adaptive learning of diverse weather-specific knowledge and shared knowledge, without requiring ground truth weather types and severity data? We answer it by our *language-driven all-in-one restoration* (LDR) framework in two aspects.

† Corresponding author.



(a) 7-th (b) 10-th (c) 54-th (d) Input (e) Transweather (f) Zero-out (g) Distribution

Figure 2. *Ineffective computations of Transweather [42]. We find the activity of some parameters in Transweather [42] are degradation dependent, and zero-outing computations for those inactivate parameters barely affect image restoration. (a) 7-th channel, (b) 10-th channel, and (c) 54-th channel of feature maps from Transweather, the brighter, the larger activation value, are activated for degradation of rain, raindrop, and snow. We show the (d) degraded images, (e) restorations from Transweather, and (f) restorations by zero-outing inactive channels. (g) From top to bottom, distribution of $\mathcal{L}_2$ norm for the three channels across all images in the All-weather dataset [21].*

1) The knowledge within the feature space of a pre-trained vision-language (PVL) model can benefit various tasks, while its potential in our task is still under exploration. A straightforward way is to use the PVL model as the image degradation classifier. In contrast, we go one step further by reasoning diverse weather-specific knowledge from the feature space of the PVL model beyond the type of weather conditions.

We start by formulating a question prompt to query the occurrence, type, and severity of degradation in a degraded image. The obtained degradation prior describes *what, where, and severity* of degradations in high-level textual semantics. Then, we translate the high-level degradation prior to a 2D degradation map by aligning the prior with the degraded image. This degradation map provides a pixel-wise representation of the diverse knowledge of image degradation from the PVL model.

2) We then unleash the potential for various weather removal with the guidance of degradation maps. Observing model parameters are weather-specific, *e.g.*, the rain-related parameters are usually inactive for unrelated degradations, and zero-outing the computations of unrelated parameters also barely affects the restoration quality. Take TUM [5] and Transweather [42] as examples, the PSNR and SSIM w/wo zero-outing on the All-weather dataset are '27.93/26.52 (dB) and 0.883/0.864', and '27.98/26.65 (dB) and 0.884/0.866'. This observation, illustrated in Fig. 2, inspired us to bypass computations for parameters unrelated to the specific weather type and severity during image restoration. With the assistance of MoE structure [28, 40], our LDR framework selects experts dynamically for restoration, therefore, ensuring adaptive learning of weather-specific knowledge which is not limited to a fixed network architecture.

Specifically, we maintain a candidate list of restoration experts and utilize the degradation map to sparsely select the most related restoration experts for each degradation. By applying the selected expert pixel-wisely to restore weather-specific features, we create flexible and degradation-adaptive expert combinations/model architectures for restoration.

Though we can have a preliminary restoration from experts restored feature, considering image regions with similar values from the degradation map tend to benefit restoration features of each other, we re-use the degradation map to aggregate the restoration features, and improve the locality of the obtained restoration by a simple convolutional feedforward network. Our main contributions are:

- We present an LDR framework to adaptively remove various adverse weather conditions in an all-in-one solution;
- We propose a degradation map measurement module for extracting diverse weather-specific knowledge from a pre-trained vision-language model;
- We propose a Top-K expert restoration module, sparsely and adaptively computing pixel-wise restoration features.

The overall comparison in Fig. 1 shows the superiority of our framework for handling various degradations compared to state-of-the-art methods.

2. Related Work

Adverse Weather Restoration. The field of adverse weather restoration encompasses two distinct approaches: task-specific methods [2, 3, 23, 51, 52] train the model independently for each individual degradation and all-in-one frameworks [5, 17, 21, 32, 35, 42, 44, 54, 57] develop a single unified model capable of handling various adverse weather degradations. A prevalent strategy for the all-in-one framework is to build upon task-specific models by employing multi-task learning techniques [44, 54, 57] or knowledge distillation [5] that consolidates multiple task-specific networks into a single network. Nevertheless, these methods typically rely on fixed computations, and the computation effectiveness is discounted as model parameters are

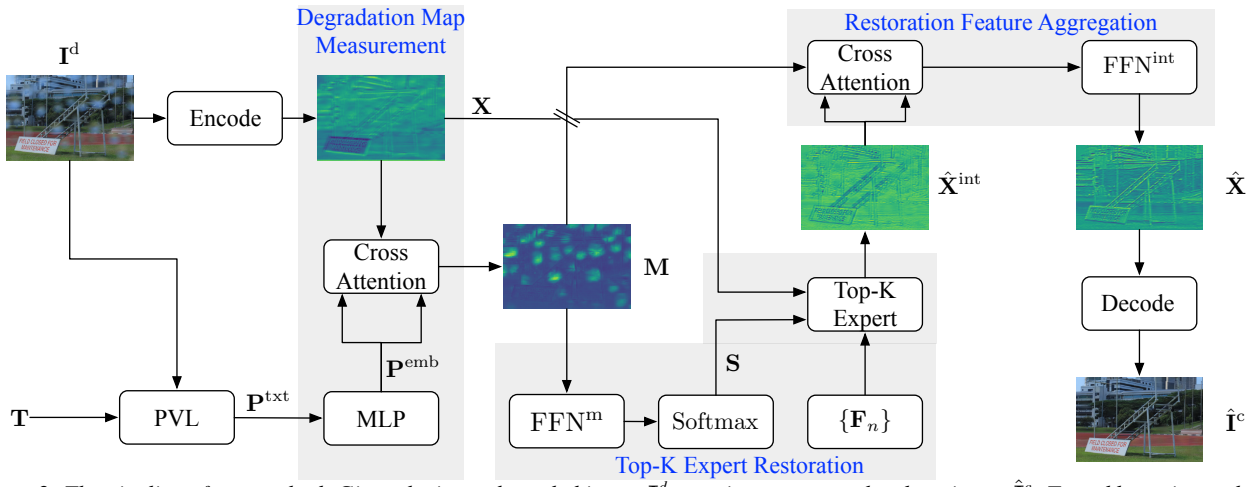


Figure 3. *The pipeline of our method. Given the input degraded image $\mathbf{I}^d$, we aim to recover the clean image $\hat{\mathbf{I}}^c$. To tackle various adverse weather conditions, we format a question prompt $\mathbf{T}$, and query a pre-trained vision-language (PVL) model with $\mathbf{I}^d$ and $\mathbf{T}$, to estimate the degradation type and severity. The generated descriptions $\mathbf{P}^{txt}$ are transformed by a multilayer perceptron (MLP) to get the degradation prior $\mathbf{P}^{emb}$. Meanwhile, we extract a feature map $\mathbf{X}$ from the input degraded image $\mathbf{I}^d$. A degradation map $\mathbf{M}$ is computed by cross-attending $\mathbf{X}$ with the prior $\mathbf{P}^{emb}$, describing pixel-wise degradation pattern. We maintain a trainable candidate list of convolution filters, $\{\mathbf{F}_n|n=1,...,N\}$, and denote them as experts. We first parse the degradation map $\mathbf{M}$ into a score map $\mathbf{S}$ via a feedforward network FFN^m and a softmax layer, and then use $\mathbf{S}$ to find the best K experts describing the degradation of $\mathbf{X}$. Best experts are convolved with $\mathbf{X}$ to generate an intermediate feature map $\hat{\mathbf{X}}^{int}$. Finally, a cross-attention layer is used to aggregate $\hat{\mathbf{X}}^{int}$ with the guidance of the degradation map $\mathbf{M}$. We improve the feature locality with a feedforward network FFN^{int}, and the output $\hat{\mathbf{X}}$ is decoded to the clean image $\hat{\mathbf{I}}^c$.*

separately and degradation-dependently activated.

Conversely, several methods [17, 35, 57] have been developed to overcome these limitations by focusing on learning weather-specific knowledge. They utilize indicators like degradation types [27, 35] or depth maps [57] during training to categorize and direct images to appropriate sub-networks for restoration. Though AirNet [17] learns weather-specific knowledge with contrastive learning. Its fixed sub-network designs can only restore images with certain degradation types, and does not generalize to unseen weather degradations, *e.g.*, rain mixed with haze. Furthermore, these mentioned works overlook the fact that varying levels of degradation severity also warrant adaptively tailored computational processes, for a specific degradation type.

Diverging from previous methods, we harness PVL models to reason both the type and severity of degradations, enabling adaptive restoration with dynamic sub-networks informed by this rich, weather-specific knowledge.

Sparse Mixture of Expert. The concept of expert models [28] is defined as a subset of model parameters/computation, where each expert is specialized in handling distinct aspects of the input data. The sparse mixture of experts usually employs a routing mechanism [40] to dynamically and adaptively forward input to a subset of these experts, bypassing unnecessary and irrelevant computations. Given fixed computation costs, due to the sparsity, this framework can readily scale the number of experts to improve the model capability that has been widely verified in the domain of natural language processing [10, 40, 58] and computer vision [7, 11, 22, 30]. We study a prior derived from a PVL model to dynamically select experts, and adaptively apply expert to restore degraded images.

Vision-language Model. With the release of ChatGPT, remarkable reasoning abilities of large language model (LLM) [16, 18, 25, 31, 39, 56] have been shown. It motivates the vision-language community to transfer the reasoning abilities to visual data. A common pipeline is to project an image to a joint space of LLM [19, 24, 25, 48, 49], and feed the projected image alongside user-provided text for conditional text generation. This paper proposes to leverage the PVL (*e.g.*, [16]) to generate prior of an adverse weather degraded image for adaptive image restoration.

3. Method

Overview. Given an input image $\mathbf{I}^{\mathrm{d}}$ degraded by adverse weather, we aim to recover the corresponding clean image $\hat{\mathbf{I}}^{\mathrm{c}}$. Our method adaptively recovers $\hat{\mathbf{I}}^{\mathrm{c}}$ with degradation prior $\mathbf{P}^{\mathrm{emb}}$ obtained from a pre-trained vision-language (PVL) model. The overall architecture is shown in Fig. 3.

To derive degradation prior $\mathbf{P}^{\mathrm{emb}}$, we query a PVL model with a question prompt $\mathbf{T}$ to reason about the occurrence, type, and severity of degradation in the input image $\mathbf{I}^{\mathrm{d}}$, outputting $\mathbf{P}^{\mathrm{txt}}$. Then a mapping network projects $\mathbf{P}^{\mathrm{txt}}$ to $\mathbf{P}^{\mathrm{emb}}$. Meanwhile, we encode $\mathbf{I}^{\mathrm{d}}$ to $\mathbf{X}$ by using an encoder.

We adaptively restore $\mathbf{X}$ into $\hat{\mathbf{X}}$ in the embedding space, and decode $\hat{\mathbf{X}}$ to a restored image $\hat{\mathbf{I}}^{\mathrm{c}}$, with three steps: i) degradation map measurement, measuring degradation for

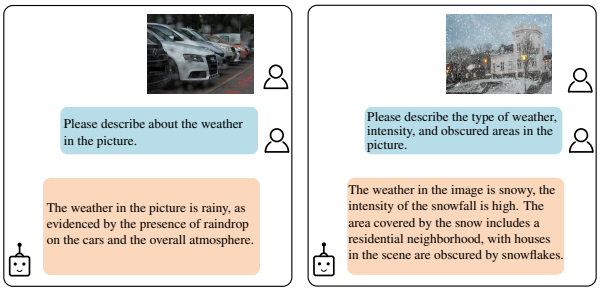


Figure 4. *Examples text descriptions from LISA [16] for reasoning weather degraded images.*

each pixel of $\mathbf{I}^{\mathrm{d}}$ by using the degradation prior $\mathbf{P}^{\mathrm{emb}}$; ii) top-K expert restoration, selecting expert/parameters from a trainable candidate list of convolution filters according to the degradation map $\mathbf{M}$, and convolving with $\mathbf{X}$ to get intermediate restoration features $\hat{\mathbf{X}}^{\mathrm{int}}$; iii) restoration feature aggregation, deriving $\hat{\mathbf{X}}$ by pixel-wisely aggregating $\hat{\mathbf{X}}^{\mathrm{int}}$ with respect to the degradation map $\mathbf{M}$, and improving the feature locality with a feedforward network. Finally, $\hat{\mathbf{X}}$ is decoded to get the restored image $\hat{\mathbf{I}}^{\mathrm{c}}$.

3.1. Degradation Prior

Degradation Prior Generation. We leverage the context learning capability of the PVL model (*e.g.*, [16]) to reason diverse degradation knowledge of the degraded image $\mathbf{I}^{\mathrm{d}}$ with a question prompt $\mathbf{T}$. We format the question prompt $\mathbf{T}$ inspired by the chain-of-thought reasoning that makes the model to identify the occurrence, type, and severity of degradation. In Fig. 4, we provide the text description examples obtained by using different prompts. The generated descriptions $\mathbf{P}^{\mathrm{txt}}$ are defined as

$$\mathbf{P}^{\mathrm{txt}} = \mathrm{VL}(\mathbf{I}^{\mathrm{c}}, \mathbf{T}) , \quad \mathbf{P}^{\mathrm{txt}} \in \mathbb{R}^{L \times C^{\mathrm{vl}}} , \tag{1}$$

where $\mathrm{VL}(\cdot,\cdot)$ is the PVL model, L is the description length, and C^{vl} is the channel dimension. To preserve the representation capabilities of PVL, we prune out the text description output layer of PVL, and use the embedding before the output layer as $\mathbf{P}^{\mathrm{txt}}$.

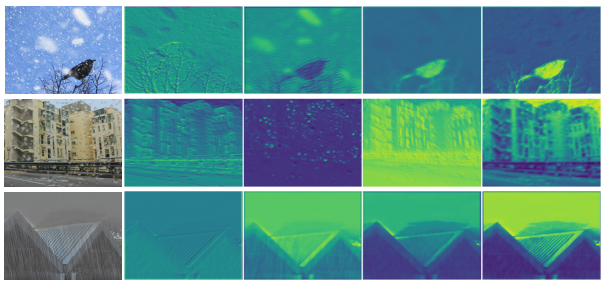
Degradation Prior Embedding. We align $\mathbf{P}^{\mathrm{txt}}$ to the embedding space with size C of our restoration model by using a multilayer perceptron network $\mathrm{MLP}(\cdot)$. We have

$$\mathbf{P}^{\mathrm{emb}} = \mathrm{MLP}(\mathbf{P}^{\mathrm{txt}}) , \quad \mathbf{P}^{\mathrm{emb}} \in \mathbb{R}^{L \times C} . \tag{2}$$

The $\mathbf{P}^{\mathrm{emb}}$ is then used as the degradation prior in our restoration model to adaptively recover $\hat{\mathbf{I}}^{\mathrm{c}}$.

3.2. Language-driven Restoration Model

We use an encoder-decoder architecture [8] as a backbone, and adaptively recover $\hat{\mathbf{I}}^{\mathrm{c}}$ in the embedding space with $\mathbf{P}^{\mathrm{emb}}$ and $\mathbf{X}$ (the encoded embedding of $\mathbf{I}^{\mathrm{d}}$) as inputs.



(a) $\mathbf{I}^{\mathrm{d}}$ (b) $\mathbf{X}$ (c) $\mathbf{M}$ (d) $\hat{\mathbf{X}}^{\mathrm{int}}$ (e) $\hat{\mathbf{X}}$

Figure 5. *Visualization of* $\mathbf{X}$, $\mathbf{M}$, $\hat{\mathbf{X}}^{int}$ *and* $\hat{\mathbf{X}}$. *Without cherry-picking, we show the channel with maximum gradients. From (b)* $\mathbf{X}$ *to (c)* $\mathbf{M}$, *degraded regions are highlighted, and regions with similar degradation severity have similar activation values on the feature map* $\mathbf{M}$. *From (b)* $\mathbf{X}$ *to (d)* $\hat{\mathbf{X}}^{int}$ *and (e)* $\hat{\mathbf{X}}$, *degradations are removed step by step.*

Degradation Map Measurement. The degradation prior $\mathbf{P}^{\mathrm{emb}}$ is text description related with high-level semantics. However, for $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$ with height H, width W, and channel C, the degradation is usually pixel-wise and distinct in severity. To allow pixel-wise computation of adverse weather removal, we measure the 2D degradation map $\mathbf{M}$ for $\mathbf{X}$ by the cross-attention mechanism,

$$\mathbf{Q} = \mathbf{X}\mathbf{W}^{\mathrm{q}_1} , \quad \mathbf{K} = \mathbf{P}^{\mathrm{emb}}\mathbf{W}^{\mathrm{k}_1} , \quad \mathbf{V} = \mathbf{P}^{\mathrm{emb}}\mathbf{W}^{\mathrm{v}_1} , \tag{3}$$

$$\mathbf{M} = \mathrm{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{Softmax}\left(\mathbf{Q}\mathbf{K}^{\top}\right)\mathbf{V} , \tag{4}$$

where $\mathbf{W}^{\mathrm{q}_1}$, $\mathbf{W}^{\mathrm{k}_1}$, and $\mathbf{W}^{\mathrm{v}_1} \in \mathbb{R}^{C \times C}$ are linear projection matrices for obtaining the query, key, and value. Our cross-attention mechanism first computes the semantic alignments between $\mathbf{X}$ and $\mathbf{P}^{\mathrm{emb}}$ via $\mathrm{Softmax}\left(\mathbf{Q}\mathbf{K}^{\top}\right)$, and then transforms $\mathbf{P}^{\mathrm{emb}}$ into a 2D degradation map $\mathbf{M}$ that describes pixel-wise degradation of $\mathbf{X}$. We visualize $\mathbf{X}$ and $\mathbf{M}$ in Fig. 5. The regions of similar degradation severity have been individually highlighted, by applying Eq. (4).

Top-K Expert Restoration. Not all pixels degrade equally, and we adaptively restore each of them by using pixel-wise degradation knowledge within $\mathbf{M}$. We have N candidate experts (convolution filters), $\{\mathbf{F}_n | n = 1, ..., N\}$, for diverse adverse weather conditions, where each candidate $\mathbf{F}_n$ is expertise at generating restoration features for specific degradation types or severities (see 4.2). With the degradation map $\mathbf{M}$, we select experts for each pixel of $\mathbf{X}$.

Specifically, the degradation map $\mathbf{M} \in \mathbb{R}^{H \times W \times C}$ is fed to a feedforward network $\mathrm{FFN}^{\mathrm{m}}$, followed by a Softmax layer along the last dimension, to output a normalized pixel-wise selection score $\mathbf{S} \in \mathbb{R}^{H \times W \times N}$, *i.e.*, $\mathbf{S} = \mathrm{Softmax}(\mathrm{FFN}^{\mathrm{m}}(\mathbf{M}))$. For a pixel with location $(i, j), i \in [1, H], j \in [1, W]$, $\mathbf{S}(i, j) \in \mathbb{R}^{N}$ and $\mathbf{S}(i, j, n)$ measures how likely the n-th expert $\mathbf{F}_n$ can be used to describe the degradation of $\mathbf{X}(i, j)$.

To find the best experts describing the degradation of $\mathbf{X}(i, j)$, we compute the Top-K scores of $\mathbf{S}(i, j)$ and record

Table 1. *Quantitative comparison on the All-weather dataset. We respectively color the best and the second-best methods in red and blue.*

Type	Method	Rain		Snow		Raindrop		Average	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
	BestT + VL	24.33	0.860	28.47	0.872	29.23	0.895	27.34	0.876
	BestT + GT	27.04	0.913	30.61	0.900	31.63	0.936	29.76	0.916
General	MPRNet [51]	23.08	0.839	27.69	0.849	28.75	0.879	26.51	0.856
General	NAFNet [3]	23.21	0.840	27.68	0.847	28.90	0.890	26.60	0.859
General	Uformer [45]	22.93	0.835	27.50	0.838	28.51	0.871	26.31	0.848
General	Restormer [52]	23.37	0.845	27.81	0.850	29.10	0.890	26.76	0.862
General	GRL [23]	23.31	0.842	27.79	0.849	29.05	0.888	26.72	0.860
All-in-One	All-in-One [21]	24.71	0.898	28.33	0.882	31.12	0.927	28.05	0.902
All-in-One	AirNet [17]	23.12	0.837	27.92	0.858	28.23	0.892	26.42	0.862
All-in-One	TUM [5]	23.92	0.855	29.27	0.884	30.75	0.912	27.98	0.884
All-in-One	Transweather [42]	23.18	0.841	27.80	0.854	28.98	0.902	26.65	0.866
All-in-One	WDiff [32]	26.18	0.907	29.69	0.893	29.71	0.911	28.53	0.904
All-in-One	WGWS [57]	25.31	0.901	29.71	0.894	31.31	0.932	28.78	0.909
All-in-One	Ours	26.92	0.912	30.79	0.905	31.54	0.933	29.75	0.916

corresponding indices as the best experts. The best experts restored feature for location (i, j) is given by,

$$\hat{\mathbf{X}}^{\text{int}}(i,j) = \sum_{k=1}^{K} \mathbf{S}(i,j,\rho(k)) \cdot \mathcal{E}(i,j,\rho(k)) \,, \tag{5}$$

$$\mathcal{E}(i,j,\rho(k)) = \sum_{\Delta i, \Delta j} \mathbf{X}(i+\Delta i, j+\Delta j) \cdot \mathbf{F}_{\rho(k)}(\Delta i, \Delta j) \,,$$

where $\rho(k)$ is the index of the selected k-th expert. $\mathcal{E}(i,j,\rho(k))$ is the convolution result of $\mathbf{X}(i,j)$ and the k-th expert $\mathbf{F}_{\rho(k)}$, and $(\Delta i, \Delta j)$ iterates over the convolution kernel size. $\mathbf{S}(i,j,\rho(k))$ are weights for prioritizing different experts. With our pixel-wise Top-K experts, the intermediate restoration feature $\hat{\mathbf{X}}^{\text{int}}$ for each pixel of $\mathbf{X}$ is adaptively generated. Practically, we find the best performance by region-wisely smoothing the expert selection score $\mathbf{S}$, as pixels of a region frequently suffer the same degradation.

Restoration Feature Aggregation. Pixels with similar degradation information in $\mathbf{M}$ potentially benefit each other in deriving the restoration $\hat{\mathbf{X}}$. We compute the compatibility between degradation prior $\mathbf{M}$ and restoration features $\hat{\mathbf{X}}^{\text{int}}$, and aggregate weighted restoration features pixel-wisely by using the cross-attention mechanism,

$$\mathbf{Q} = \mathbf{M}\mathbf{W}^{\text{q}_2} \,, \quad \mathbf{K} = \hat{\mathbf{X}}^{\text{int}}\mathbf{W}^{\text{k}_2} \,, \quad \mathbf{V} = \hat{\mathbf{X}}^{\text{int}}\mathbf{W}^{\text{v}_2} \,, \tag{6}$$

$$\hat{\mathbf{X}} = \text{FFN}^{\text{int}}(\text{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V})) \,, \tag{7}$$

where FFN^{int} is a feedforward convolutional network to improve the feature locality. Finally, $\hat{\mathbf{X}}$ is decoded to the restoration $\hat{\mathbf{I}}^{\text{c}}$. In Fig. 5, we compare $\hat{\mathbf{X}}^{\text{int}}$ and $\hat{\mathbf{X}}$, and find that $\hat{\mathbf{X}}$ is cleaner than $\hat{\mathbf{X}}^{\text{int}}$, as pixels with similar degradation measurements benefit each other in the restorations.

3.3. Loss Function

We train our network with Charbonnier loss $\mathcal{L}_{\text{char}}$ and gradient-level edge loss $\mathcal{L}_{\text{edge}}$, to penalize the deviation of restoration from ground truth clean image, and encourage the consistent image gradients. We have $\mathcal{L}_{\text{char}}$ as

$$\mathcal{L}_{\text{char}} = \sqrt{\|\mathbf{I}^{\text{c}} - \hat{\mathbf{I}}^{\text{c}}\|^2 + \varepsilon^2} \,, \tag{8}$$

where $\varepsilon = 10^{-4}$ is a constant in all experiments. The $\mathcal{L}_{\text{edge}}$ is

$$\mathcal{L}_{\text{edge}} = \sqrt{\|\nabla\mathbf{I}^{\text{c}} - \nabla\hat{\mathbf{I}}^{\text{c}}\|^2 + \varepsilon^2} \,, \tag{9}$$

where ∇ is the laplacian gradient operator. With a balance parameter λ, the total loss is given by

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{char}} + \lambda\mathcal{L}_{\text{edge}} \,. \tag{10}$$

4. Experiments and Analysis

Implementation Details. Our model is implemented using the PyTorch framework and all experiments are conducted on an RTX A6000 GPU. We train our network with batch size 4 and 4×10^6 iterations using the ADAM optimizer. The learning rate is decayed from 2×10^{-4} to 1×10^{-6} by cosine annealing strategy. In training, images are randomly cropped to size 256×256, and $\lambda = 0.05$. Code and more results are in our project page.

Datasets. Our experiments are conducted on both synthetic dataset and real dataset, *i.e.*, All-weather dataset [21] and WeatherStream [53] dataset. All-weather dataset comprises $18,609$ training images and $17,609$ testing images, and is composed of subsets from Outdoor-rain [20], Snow100K-L [26], and Raindrop [37], corresponding to rain, snow, and raindrops weather conditions respectively. WeatherStream dataset is a real-world dataset that has three weather conditions, *i.e.*, rain, snow, and fog, with a total of $176,100$ training images and $11,400$ testing images.

Evaluation Metrics. We use averages of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as evaluation metrics (RGB-channel). The higher, the better.

Table 2. *Quantitative comparison on the WeatherStream dataset. We color the best and the second-best methods in red and blue.*

Type	Method	Rain		Haze		Snow		Average	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
	BestT + VL	21.20	0.781	21.60	0.755	20.32	0.772	21.04	0.769
	BestT + GT	23.95	0.810	22.97	0.804	22.70	0.828	23.21	0.814
General	MPRNet [51]	21.50	0.791	21.73	0.763	20.74	0.801	21.32	0.785
General	NAFNet [3]	23.01	0.803	22.20	0.803	22.11	0.826	22.44	0.811
General	Uformer [45]	22.25	0.791	18.81	0.763	20.94	0.801	20.67	0.785
General	Restormer [52]	23.67	0.804	22.90	0.803	22.51	0.828	22.86	0.812
General	GRL [23]	23.75	0.805	22.88	0.802	22.59	0.829	23.07	0.812
All-in-One	All-in-One [21]	-	-	-	-	-	-	-	-
All-in-One	AirNet [17]	22.52	0.797	21.56	0.770	21.44	0.812	21.84	0.793
All-in-One	TUM [5]	23.22	0.795	22.38	0.805	22.25	0.827	22.62	0.809
All-in-One	Transweather [42]	22.21	0.772	22.55	0.774	21.79	0.792	22.18	0.779
All-in-One	WGWS [57]	23.80	0.807	22.78	0.800	22.72	0.831	23.10	0.813
All-in-One	Ours	24.42	0.818	23.11	0.809	23.12	0.838	23.55	0.822

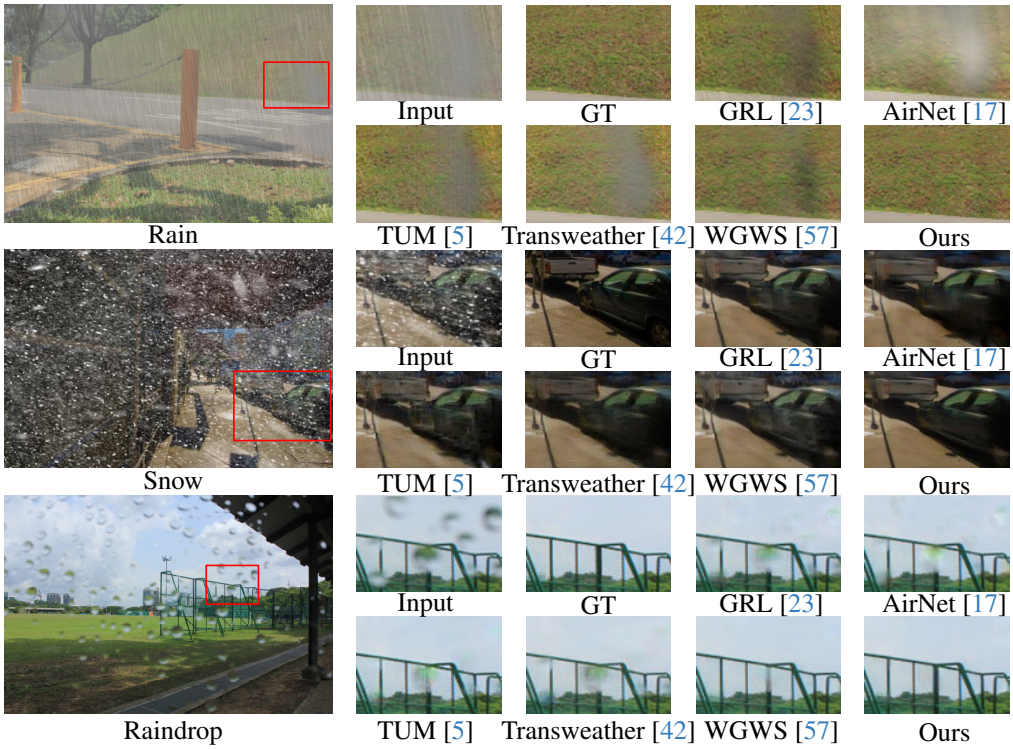


Figure 6. *Qualitative comparison on the All-weather dataset. The first column shows degraded images, while the crops for the bounding box regions of degraded images, ground truth, restoration from SOTA methods and our method are shown in the subsequent columns.*

Baseline Methods. We compare with the state-of-the-art (SOTA) general and all-in-one methods. For general methods, task-specific methods are trained with multi-task learning of different weather conditions, *i.e.*, MPRNet [51], NAFNet [3], Uformer[45], Restormer [52], and GRL [23]. All-in-one methods are All-in-One [21], AirNet[17], TUM [5], Transweather [42], WDiff [32] and WGWS [57].

4.1. Experimental Results

Quantitative Comparison. We compare with the SOTA general and all-in-one methods on the All-weather and WeatherStream datasets in Tab. 1 and Tab. 2, respectively. Our method achieves the best performance.

Furthermore, we create two strong baselines: i) BestT + VL. We use a PVL model for zero-shot adverse weather classification, and select the best task-specific methods for restoring degraded images; ii) BestT + GT. The ground truth adverse weather type is used to select the best task-specific method for image restoration. We find that i) Our method significantly outperforms 'BestT + VL', though it uses the same PVL as our model and uses multiple models for different adverse weather conditions. This shows

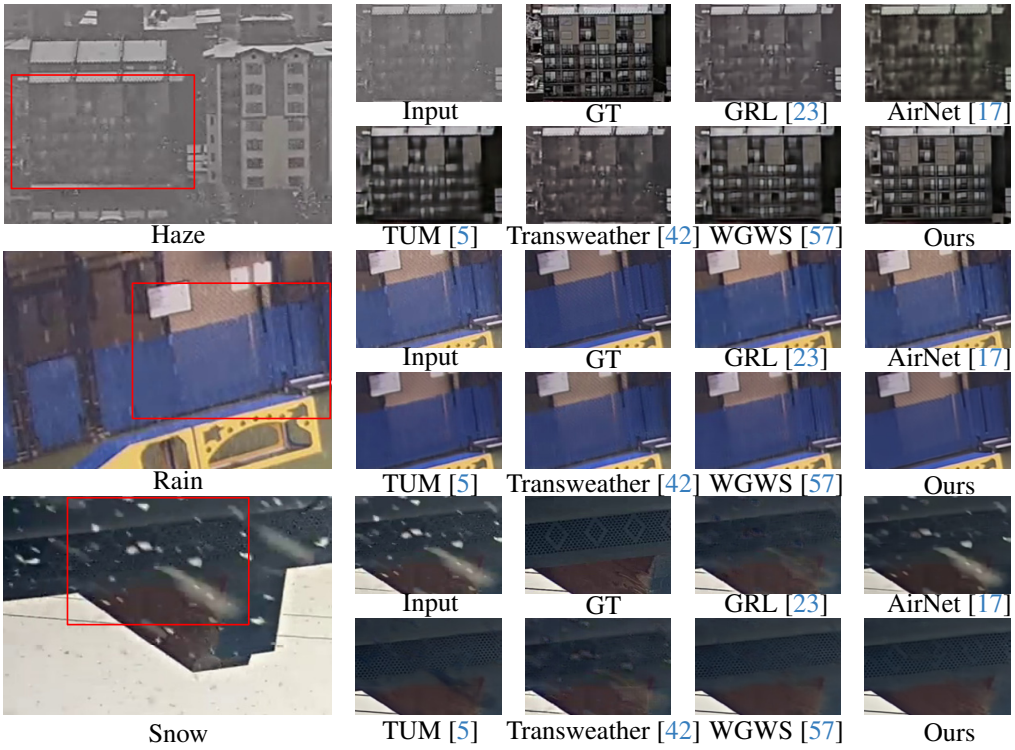


Figure 7. *Qualitative comparison on the WeatherStream dataset. The first column shows degraded images, while the crops for the bounding box regions of degraded images, ground truth, restoration from SOTA methods and our method are shown in the subsequent columns.*

that non-trivial designs are required for effectively leveraging the PVL model in all-in-one adverse weather removal; ii) Although 'BestT + GT' outperforms most SOTA general and all-in-one methods, our method still achieves competitive performance compared to it. This indicates: 1) the benefit of accessing prior knowledge of adverse weather conditions, and 2) our method has learned the shared and weather-specific knowledge adaptively by selecting sparse experts dynamically.

Qualitative Comparison. We compare with the SOTA methods on restoring images degraded by different adverse weather conditions. The results are given in Fig. 6 and Fig. 7. Our method consistently restores clearer images than SOTA methods under different weather conditions.

4.2. Ablation studies and Discussions

We validate the effectiveness and components of our framework on the All-weather dataset.

Model Architecture. We study the effectiveness of the degradation map measurement (DMM), top-K expert restoration (TER), and restoration feature aggregation (RFA) modules. The results are given in Tab. 3. The best performance is achieved by using all modules.

Degradation Severity. We use the PVL model to partition the All-weather dataset [21] into three subsets based on degradation severity: slight, moderate, and heavy. We then compare our results with those of AirNet and WGWS in Tab. 4. Our method shows significant improvement over AirNet [17] and WGWS [57] in the heavily degraded subset. This indicates the effectiveness of reasoning with diverse weather-specific knowledge, such as severity.

Pixel-wise Expert. Different regions of the same degraded image often exhibit varying degrees of degradation severity, and should be adaptively and pixel-wisely restored. We use the degradation prior provided by the PVL model to select the same expert for the degraded image. It shows a

Table 3. *The effectiveness of our model components.*

DMM	TER	RFA	PSNR	SSIM
✗	✓	✗	27.93	0.882
✗	✗	✓	28.37	0.889
✓	✓	✗	28.25	0.890
✓	✗	✓	29.11	0.902
✗	✓	✓	28.55	0.895
✓	✓	✓	29.75	0.916

Table 4. *Comparisons on the All-weather dataset with degradation severity of slight, moderate, and heavy.*

Method	Slight		Moderate		Heavy	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
AirNet	27.59	0.895	26.49	0.865	24.50	0.818
WGWS	29.74	0.921	28.77	0.910	27.14	0.886
Ours	30.46	0.925	29.78	0.918	28.38	0.899

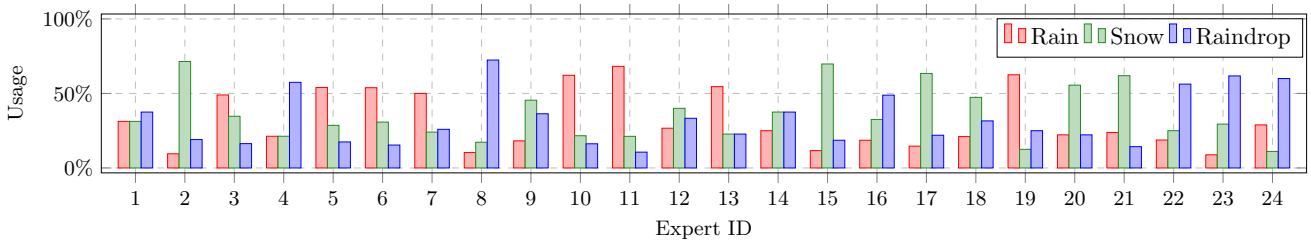
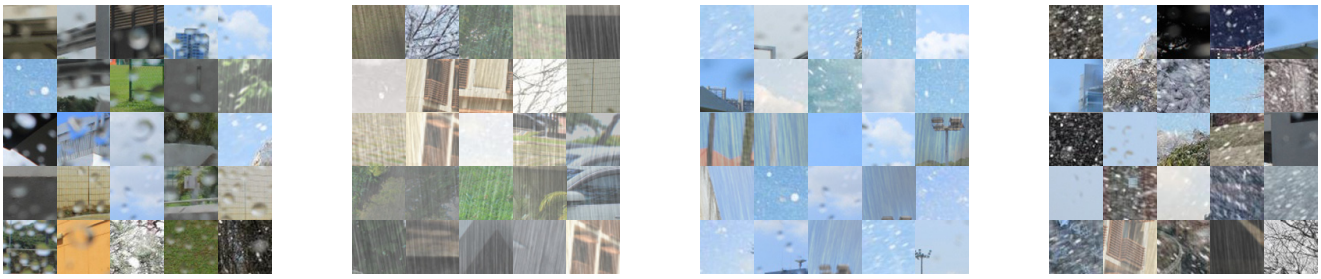
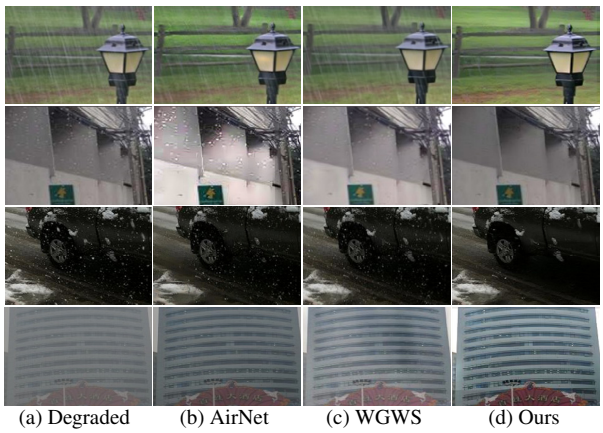


Figure 8. *The usage frequencies of experts for different degradation types. For example, Expert 8, 11, and 15 focuses on raindrop, rain, and snow, respectively. Expert 12 handles all types of degradation. Please refer to Fig. 9 for visualization.*



(a) Expert 8 (b) Expert 11 (c) Expert 12 (d) Expert 15

Figure 9. *Sample image regions, activating different experts. We show regions with highest selection scores for experts 8, 11, 12, and 15 on the All-weather dataset.*



(a) Degraded (b) AirNet (c) WGWS (d) Ours

Figure 10. *From the first to last rows, the degradation is rain, raindrop, snow, and haze, respectively. From left to right, (a) degraded images are restored by (b) AirNet, (c) WGWS, and (d) our method.*

0.33 dB/0.07 decrease in PSNR/SSIM, indicating the necessity of selecting experts pixel-wisely.

Adaptive Adverse Weather Removal. We study the model response to different adverse weather degradation by measuring the usage of our experts. Fig. 8 shows the usage for rain, snow, and raindrop. We find that there are experts for weather-specific knowledge and shared knowledge among degradation, such as Expert 8, Expert 11, Expert 12, and Expert 15. We visualize the image patches selecting the four experts in Fig. 9, where Expert 8, Expert 11, and Expert 15 are selected by image regions with raindrop, rain, and snow degradation, and Expert 12 is selected by image regions with sky in all types of degradation.

Model Ability Under Complex Weather Condition. We test our model trained on the All-weather dataset with real images degraded by rain, raindrop, snow, and haze, as shown in Fig. 10. We compare our results with the two most competitive methods, AirNet and WGWS. Our model successfully disentangles weather-specific knowledge and generalizes to restore images degraded by haze.

5. Conclusion and Broader Impact

We have proposed an LDR framework that adaptively removes various adverse weather conditions in an all-in-one solution. Our key insight is to leverage a pre-trained vision-language model to reason diverse weather-specific knowledge in a degraded image. We then use this knowledge to restore a clean image with three modules: degradation map measurement, Top-K expert restoration, and restoration feature aggregation. Experiments on standard benchmark datasets demonstrate that our method outperforms past works by a large margin.

Broader Impact. Our method is promising to be developed as an image restoration foundation model, prompting by degradation prior generated by a vision-language model.

Acknowledgment. This work was supported in part by the Beijing Institute of Technology Research Fund Program for Young Scholars, BIT Special-Zone, and National Natural Science Foundation of China 62302045.

References

[1] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12299–12310, 2021. 1

[2] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normalization network for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 182–192, 2021. 2

[3] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision*, pages 17–33. Springer, 2022. 2, 5, 6

[4] Wei-Ting Chen, Hao-Yu Fang, Cheng-Lin Hsieh, Cheng-Che Tsai, I Chen, Jian-Jiun Ding, Sy-Yen Kuo, et al. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4196–4205, 2021. 1

[5] Wei-Ting Chen, Zhi-Kai Huang, Cheng-Che Tsai, Hao-Hsiang Yang, Jian-Jiun Ding, and Sy-Yen Kuo. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17653–17662, 2022. 1, 2, 5, 6, 7

[6] 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*, pages 5896–5905, 2023. 1

[7] Zixiang Chen, Yihe Deng, Yue Wu, Quanquan Gu, and Yuanzhi Li. Towards understanding the mixture-of-experts layer in deep learning. *Advances in neural information processing systems*, 35:23049–23062, 2022. 3

[8] Yuning Cui, Yi Tao, Zhenshan Bing, Wenqi Ren, Xinwei Gao, Xiaochun Cao, Kai Huang, and Alois Knoll. Selective frequency network for image restoration. In *The Eleventh International Conference on Learning Representations*, 2023. 4

[9] Keval Doshi and Yasin Yilmaz. Multi-task learning for video surveillance with limited data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3889–3899, 2022. 1

[10] Nan Du, Yanping Huang, Andrew M Dai, Simon Tong, Dmitry Lepikhin, Yuanzhong Xu, Maxim Krikun, Yanqi Zhou, Adams Wei Yu, Orhan Firat, et al. Glam: Efficient scaling of language models with mixture-of-experts. In *International Conference on Machine Learning*, pages 5547–5569. PMLR, 2022. 3

[11] Markus Enzweiler and Dariu M Gavrila. A multilevel mixture-of-experts framework for pedestrian classification. *IEEE Transactions on Image Processing*, 20(10):2967–2979, 2011. 3

[12] Sergio Escalera, Hugo Jair Escalante, Zhen Lei, Hao Fang, Ajian Liu, and Jun Wan. Surveillance face presentation attack detection challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6360–6370, 2023. 1

[13] Xueyang Fu, Jie Xiao, Yurui Zhu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Continual image deraining with hypergraph convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 1

[14] Trung Hoang, Haichuan Zhang, Amirsaeed Yazdani, and Vishal Monga. Transer: Hybrid model and ensemble-based sequential learning for non-homogenous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1670–1679, 2023. 1

[15] Yihan Hu, Jiazhi Yang, Li Chen, Keyu Li, Chonghao Sima, Xizhou Zhu, Siqi Chai, Senyao Du, Tianwei Lin, Wenhai Wang, et al. Planning-oriented autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17853–17862, 2023. 1

[16] Xin Lai, Zhuotao Tian, Yukang Chen, Yanwei Li, Yuhui Yuan, Shu Liu, and Jiaya Jia. Lisa: Reasoning segmentation via large language model. *arXiv preprint arXiv:2308.00692*, 2023. 3, 4

[17] 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*, pages 17452–17462, 2022. 1, 2, 3, 5, 6, 7

[18] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International Conference on Machine Learning*, pages 12888–12900. PMLR, 2022. 3

[19] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *arXiv preprint arXiv:2301.12597*, 2023. 3

[20] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1633–1642, 2019. 5

[21] Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3175–3185, 2020. 1, 2, 5, 6, 7

[22] Xinjie Li and Huijuan Xu. Meid: mixture-of-experts with internal distillation for long-tailed video recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1451–1459, 2023. 3

[23] Yawei Li, Yuchen Fan, Xiaoyu Xiang, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Efficient and explicit modelling of image hierarchies for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18278–18289, 2023. 2, 5, 6, 7

[24] 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*, pages 8094–8103, 2023. 3

[25] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *arXiv preprint arXiv:2304.08485*, 2023. 3

[26] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27(6): 3064–3073, 2018. 1, 5

[27] Yulin Luo, Rui Zhao, Xiaobao Wei, Jinwei Chen, Yijie Lu, Shenghao Xie, Tianyu Wang, Ruiqin Xiong, Ming Lu, and Shanghang Zhang. Mowe: mixture of weather experts for multiple adverse weather removal. *arXiv preprint arXiv:2303.13739*, 2023. 3

[28] Saeed Masoudnia and Reza Ebrahimpour. Mixture of experts: a literature survey. *Artificial Intelligence Review*, 42: 275–293, 2014. 2, 3

[29] Hamam Mokayed, Amirhossein Nayebiastaneh, Kanjar De, Stergios Sozos, Olle Hagner, and Björn Backe. Nordic vehicle dataset (nvd): Performance of vehicle detectors using newly captured nvd from uav in different snowy weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5313–5321, 2023. 1

[30] Lynnette Hui Xian Ng and Kathleen M Carley. Botbuster: Multi-platform bot detection using a mixture of experts. In *Proceedings of the International AAAI Conference on Web and Social Media*, pages 686–697, 2023. 3

[31] Vicente Ordonez, Girish Kulkarni, and Tamara Berg. Im2text: Describing images using 1 million captioned photographs. *Advances in neural information processing systems*, 24, 2011. 3

[32] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 2, 5, 6

[33] Liyuan Pan, Yuchao Dai, Miaomiao Liu, Fatih Porikli, and Quan Pan. Joint stereo video deblurring, scene flow estimation and moving object segmentation. *IEEE Transactions on Image Processing*, 29:1748–1761, 2019. 1

[34] Liyuan Pan, Cedric Scheerlinck, Xin Yu, Richard Hartley, Miaomiao Liu, and Yuchao Dai. Bringing a blurry frame alive at high frame-rate with an event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6820–6829, 2019. 1

[35] Dongwon Park, Byung Hyun Lee, and Se Young Chun. All-in-one image restoration for unknown degradations using adaptive discriminative filters for specific degradations. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5815–5824. IEEE, 2023. 1, 2, 3

[36] Prashant W Patil, Sunil Gupta, Santu Rana, Svetha Venkatesh, and Subrahmanyam Murala. Multi-weather image restoration via domain translation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21696–21705, 2023. 1

[37] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2482–2491, 2018. 1, 5

[38] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI conference on artificial intelligence*, pages 11908–11915, 2020. 1

[39] 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*, pages 8748–8763. PMLR, 2021. 3

[40] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2016. 2, 3

[41] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023. 1

[42] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2353–2363, 2022. 1, 2, 5, 6, 7

[43] Yecong Wan, Yuanshuo Cheng, Mingwen Shao, and Jordi Gonzàlez. Image rain removal and illumination enhancement done in one go. *Knowledge-Based Systems*, 252:109244, 2022. 1

[44] Yinglong Wang, Chao Ma, and Jianzhuang Liu. Smartassign: Learning a smart knowledge assignment strategy for deraining and desnowing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3677–3686, 2023. 2

[45] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022. 5, 6

[46] Xuan Xiong, Yicheng Liu, Tianyuan Yuan, Yue Wang, Yilun Wang, and Hang Zhao. Neural map prior for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17535–17544, 2023. 1

[47] Hao Yang, Dongming Zhou, Jinde Cao, and Qian Zhao. Dpnet: Detail-preserving image deraining via learning frequency domain knowledge. *Digital Signal Processing*, 130: 103740, 2022. 1

[48] Hao Yang, Liyuan Pan, Yan Yang, and Miaomiao Liu. Ldp: Language-driven dual-pixel image defocus deblurring network. *arXiv preprint arXiv:2307.09815*, 2023. 3

[49] Yan Yang, Liyuan Pan, and Liu Liu. Event camera data pre-training. In *Proceedings of the IEEE/CVF International*

Conference on Computer Vision (ICCV), pages 10699–10709, 2023. 3

[50] Tian Ye, Sixiang Chen, Jinbin Bai, Jun Shi, Chenghao Xue, Jingxia Jiang, Junjie Yin, Erkang Chen, and Yun Liu. Adverse weather removal with codebook priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12653–12664, 2023. 1

[51] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021. 2, 5, 6

[52] 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*, pages 5728–5739, 2022. 2, 5, 6

[53] Howard Zhang, Yunhao Ba, Ethan Yang, Varan Mehra, Blake Gella, Akira Suzuki, Arnold Pfahnl, Chethan Chinder Chandrappa, Alex Wong, and Achuta Kadambi. Weatherstream: Light transport automation of single image deweathering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13499–13509, 2023. 5

[54] Jinghao Zhang, Jie Huang, Mingde Yao, Zizheng Yang, Hu Yu, Man Zhou, and Feng Zhao. Ingredient-oriented multi-degradation learning for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5825–5835, 2023. 2

[55] Zhihong Zhang, Yuxiao Cheng, Jinli Suo, Liheng Bian, and Qionghai Dai. Infwide: Image and feature space wiener deconvolution network for non-blind image deblurring in low-light conditions. *IEEE Transactions on Image Processing*, 32:1390–1402, 2023. 1

[56] Deyao Zhu, Jun Chen, Xiaoqian Shen, Xiang Li, and Mohamed Elhoseiny. Minigpt-4: Enhancing vision-language understanding with advanced large language models. *arXiv preprint arXiv:2304.10592*, 2023. 3

[57] Yurui Zhu, Tianyu Wang, Xueyang Fu, Xuanyu Yang, Xin Guo, Jifeng Dai, Yu Qiao, and Xiaowei Hu. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21747–21758, 2023. 1, 2, 3, 5, 6, 7

[58] Simiao Zuo, Qingru Zhang, Chen Liang, Pengcheng He, Tuo Zhao, and Weizhu Chen. Moebert: from bert to mixture-of-experts via importance-guided adaptation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1610–1623, 2022. 3