Multi-axis Prompt and Multi-dimension Fusion Network for All-in-one Weather-degraded Image Restoration

Yuanbo Wen[1], **Tao Gao**[2*], **Jing Zhang**[3], **Ziqi Li**[1], **Ting Chen**[1*]

[1]School of Information Engineering, Chang'an University, China
[2]School of Data Science and Artificial Intelligence, Chang'an University, China
[3]School of Computing, Australian National University, Australia
{wyb, gaotao, tchen, lzq}@chd.edu.cn, jing.zhang@anu.edu.au

Abstract

Existing approaches aiming to remove adverse weather degradations compromise the image quality and incur the long processing time. To this end, we introduce a multi-axis prompt and multi-dimension fusion network (MPMF-Net). Specifically, we develop a multi-axis prompts learning block (MPLB), which learns the prompts along three separate axis planes, requiring fewer parameters and achieving superior performance. Moreover, we present a multi-dimension feature interaction block (MFIB), which optimizes intra-scale feature fusion by segregating features along height, width and channel dimensions. This strategy enables more accurate mutual attention and adaptive weight determination. Additionally, we propose the coarse-scale degradation-free implicit neural representations (CDINR) to normalize the degradation levels of different weather conditions. Extensive experiments demonstrate the significant improvements of our model over the recent well-performing approaches in both reconstruction fidelity and inference time.

Code — https://github.com/chdwyb/MPMF-Net

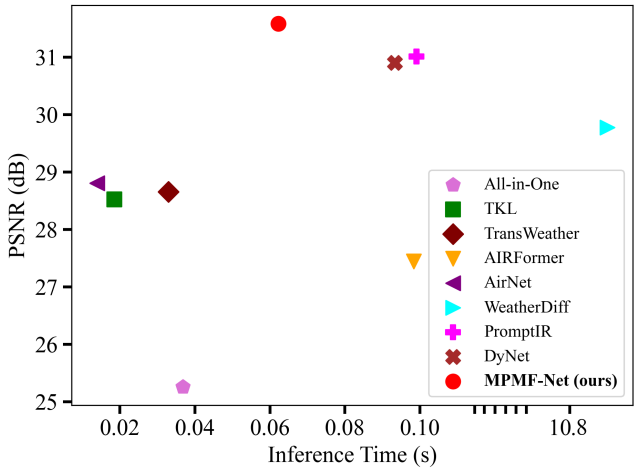

Figure 1: Efficiency comparisons of the involved comparative approaches. Our proposed model achieves the best trade-off between the image quality and inference time.

Introduction

Traffic images often suffer from texture loss and color distortion due to adverse weather conditions (Wen, Gao, and Chen 2024a; Jiang et al. 2024), which significantly reduce the effectiveness of subsequent computer vision algorithms (Gao et al. 2024a,b; Wen, Gao, and Chen 2024b). Currently, weather removal algorithms can be divided into three categories, task-specific, task-aligned and all-in-one methods. Task-specific approaches (Ding et al. 2025a; Zhang et al. 2021) involve creating custom physical models for specific weather conditions, allowing the network to accurately identify the corresponding weather effects. Meanwhile, task-aligned methods (Wang et al. 2022) employ a single network architecture to handle different weather degradations, training the model with various weather-degraded datasets. Recent advancements (Gao et al. 2024c) train an all-in-one model that can simultaneously address multiple weather conditions. However, these research faces issues with reduced image quality and longer processing times.

In this work, we propose a multi-axis prompt and multi-dimension fusion network (MPMF-Net) for all-in-one traffic weather removal. Specifically, existing methods utilize learnable prompts to capture task-related knowledge, but they only focus on either the spatial dimension (Valanarasu, Yasarla, and Patel 2022) or the channel dimension (Potlapalli et al. 2023). This narrow focus leads to inaccurate task representation, affecting the overall task understanding and the quality of reconstructed images. Additionally, these methods require a large number of parameters to learn the prompts, which makes the solution space for the prompts unstable. To this end, we propose a multi-axis prompts learning block (MPLB), which learns the prompts separately along the hw-axis ch-axis, and cw-axis. This approach enables more comprehensive and accurate prompt learning. Moreover, our MPLB uses fewer parameters than existing approaches but achieves the best all-in-one traffic weather-degraded image restoration.

Furthermore, existing methods for cross-stage intra-scale feature fusion (Wen et al. 2023; Gao et al. 2024c; Wang et al. 2022) struggle to adjust the weighting of different features based on input variations, leading to less effective

*Corresponding authors.

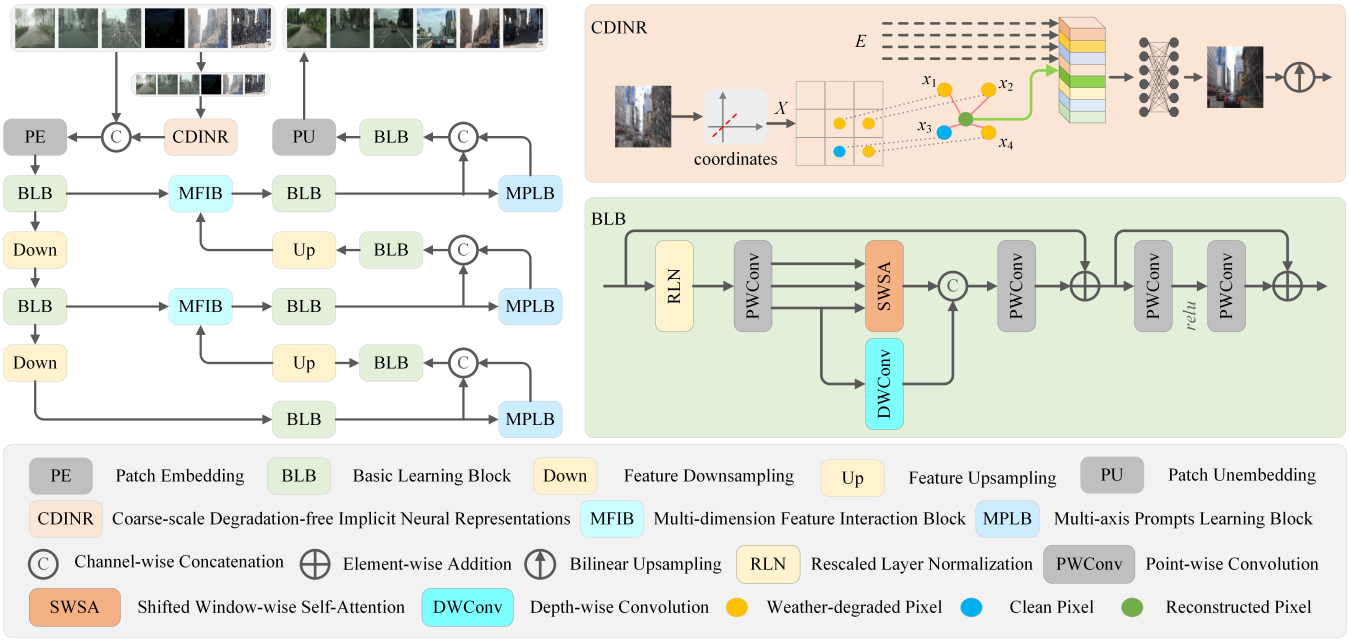

Figure 2: Overview of our proposed multi-axis prompt and multi-dimension fusion network (MPMF-Net) for all-in-one traffic weather removal. Our model contains three key components, multi-dimension feature interaction block (MFIB) achieves the features decomposed fusion in three dimensions and multi-axis prompts learning block (MPLB) learns the prompts from three axis, coarse-scale degradation-free implicit neural representations (CDINR) represent the coarse-scale degradation-free images.

fusion results. To address this issue, we introduce a multi-dimension feature interaction block (MFIB), which breaks down input features into height, width and channel dimensions. Our MFIB applies self-attention separately to each dimension, enabling adaptive feature fusion. Unlike traditional self-attention mechanisms (Zamir et al. 2022; Chen et al. 2023c), our approach uses distinct query and key representations to capture the attention maps from different features. This creates adaptive fusion weights specifically tailored to each group of features, improving the robustness of feature fusion process. Additionally, by decomposing feature fusion across three dimensions, we achieve greater precision in assessing the importance of different features.

Recently, implicit neural representations (INR) (Lee and Jin 2022) have become an effective way to encode images as continuous functions, allowing for the normalization of degradation levels and simplifying the subsequent processes (Yang et al. 2023; Chen, Pan, and Dong 2024). Therefore, we develop the coarse-scale degradation-free implicit neural representations (CDINR) to create the non-degraded neural representations from coarse-scale degraded images, which are then employed as the inputs in fine-scale reconstruction network. Our neural representations approach also aligns with the visual self-prompting paradigm (Wang et al. 2023; Wen et al. 2024a; Ma et al. 2023) for all-in-one image restoration. By utilizing the degradation-free representations at coarse scales to guide fine-scale restoration, our model continuously adapts to the input degraded images, which enhances the model ability in handling with various weather degradations.

Figure 2 illustrates the architectural overview of our proposed multi-axis prompting and multi-dimension fusion network for all-in-one traffic weather removal, and Figure 1 depicts that our model outperforms the other existing all-in-one methods with a better trade-off between the image quality and inference time. Our main contributions can be summarized as follows.

- We introduce a multi-axis prompts learning block aimed at learning more accurate all-in-one task-related representations.
- We propose a multi-dimension feature interaction block to effectively acquire the adaptive mutual attention and discern the significance of distinct features.
- We incorporate the implicit neural representations spanning adjacent branches to facilitate the precise fine-scale reconstruction and engenders the robustness against complex degradations.

Related Work

Adverse Weather Removal In adverse weather removal, Li *et al.* (Li, Tan, and Cheong 2020) introduce a groundbreaking methodology utilizing the task-specific optimized encoders to combat various weather degradations. Subsequently, Valanarasu *et al.* (Valanarasu, Yasarla, and Patel 2022) incorporate a spatially constrained self-attention mechanism alongside task-specific queries to mitigate diverse weather artifacts. In parallel, Chen *et al.* (Chen et al. 2022) employ a dual-phase knowledge amalgamation approach in a multi-contrastive learning framework. Drawing from the diffusion model (Ho, Jain, and Abbeel 2020),

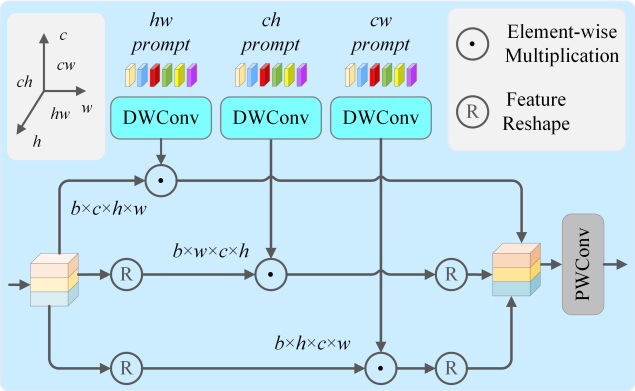

Figure 3: Illustration of our proposed multi-axis prompts learning block (MPLB), which employs three sets of prompts, corresponding to the hw, ch and cw axes, to regulate the flow of information and enhances the all-in-one task-related representations.

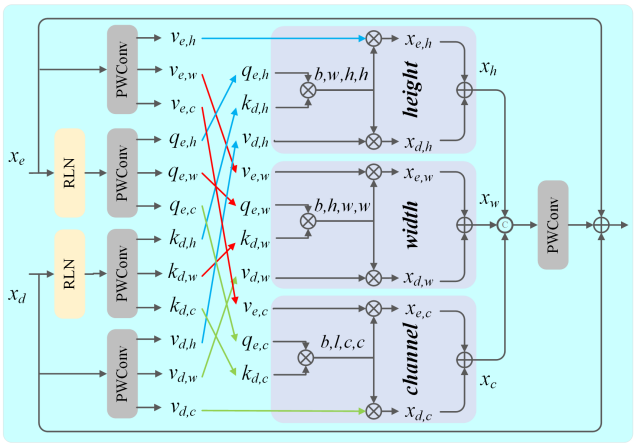

Figure 4: Illustration of our proposed multi-dimension feature interaction block (MFIB), which decomposes the two sets of features into the height, width and channel dimensions to accomplish accurate and adaptive feature fusion.

Özdenizc *et al.* (Özdenizci and Legenstein 2023) present a patch-based diffusion model aimed at restoring high-resolution weather-degraded images. Gao *et al.* (Gao et al. 2024c) explore the frequency variations during weather removal while proposing a frequency-guided and frequency-refined approach. Meanwhile, Ye *et al.* (Ye et al. 2023) develop a unified method leveraging the code-book priors to mitigate adverse weather-induced degradations. Furthermore, Luo *et al.* (Luo et al. 2023) and Wen *et al.* (Wen et al. 2024a) utilize the degradation-aware prompts to guide the image reconstruction process.

Prompt Learning Contemporary studies attempt to employ prompt learning techniques in the context of down-stream visual tasks (Ding et al. 2025b; Li et al. 2024). For instance, MAE-VQGAN (Bar et al. 2022) employs a grid-like input configuration in the course of inference to autonomously perform inpainting operations on regions lacking visual content. TransWeather (Valanarasu, Yasarla, and Patel 2022) and AIRFormer (Gao et al. 2024c) utilize several groups of task queries as the query representation in calculating self-attention maps to perform the latent features decoding. Painter (Wang et al. 2023) employs the paired images as a fixed visual prompts to elicit network alignment with task-specific objectives. Recently, PromptIR (Potlapalli et al. 2023) and DyNet (Dudhane et al. 2024) employ the learnable prompts in each decoding stage to achieve all-in-one task representation and achieve competitive performance in adverse weather removal.

Proposed Method

Multi-axis Prompts Learning Block

Existing prompt-based all-in-one methods (Dudhane et al. 2024; Gao et al. 2024c) primarily focus on either the spatial (Valanarasu, Yasarla, and Patel 2022) or channel (Potlapalli et al. 2023) dimensions of learnable parameters, resulting in the inaccurate task representations. Please note that although the learnable parameters are not technically prompt learning, we also refer to our learnable parameters as prompts following (Potlapalli et al. 2023). As noted in (Chen et al. 2023b; Ruan et al. 2023), applying the same feature transformations across all feature channels leads to an overload of redundant information. Therefore, we introduce a lightweight multi-axis prompts learning block, as shown in Figure 3. Given the input features $\boldsymbol{X}$ with the structure $(B, 3 \times C, H, W)$, we divide them into three feature groups, $\boldsymbol{X}_{hw}$, $\boldsymbol{X}_{ch}$ and $\boldsymbol{X}_{cw}$. Each group follows the structure (B, C, H, W). The learnable axis-wise prompts for $\boldsymbol{X}_{hw}$ are configured as $(1, C, x, y)$, for $\boldsymbol{X}_{ch}$ as $(1, 1, C, x)$ and for $\boldsymbol{X}_{cw}$ as $(1, 1, C, y)$, where x and y correspond to the dimensions of prompts in the height and width dimensions, respectively. All prompts are randomly initialized. We initially transform $\boldsymbol{X}_{hw}$, $\boldsymbol{X}_{ch}$ and $\boldsymbol{X}_{cw}$ into (B, C, H, W), (B, W, C, H) and (B, H, C, W), respectively. Meanwhile, three axis-wise prompts are interpolated to $\boldsymbol{P}_{hw} \in \mathbb{R}^{B\times C\times H\times W}$, $\boldsymbol{P}_{ch} \in \mathbb{R}^{B\times W\times C\times H}$ and $\boldsymbol{P}_{cw} \in \mathbb{R}^{B\times H\times C\times W}$ and optimized through depth-wise convolution. The prompted features can be obtained by

$$\boldsymbol{X}' = Concat(\boldsymbol{X}_{hw} \odot \boldsymbol{P}_{hw}, \boldsymbol{X}_{ch} \odot \boldsymbol{P}_{ch}, \boldsymbol{X}_{cw} \odot \boldsymbol{P}_{cw}), \quad (1)$$

where $Concat$ denotes the channel-wise concatenation, $\odot$ is the element-wise multiplication. Our proposed MPLB transforms the task representations by incorporating prompts from three axes, ensuring the learned prompts capture relevant information from each axis and leading to more accurate all-in-one task representations. Additionally, our prompt learning approach is efficient and stable, requiring fewer parameters than existing methods (Gao et al. 2024c; Potlapalli et al. 2023; Dudhane et al. 2024), resulting in improved computational efficiency and more reliable convergence.

Multi-dimension Feature Interaction Block

Existing methods (Zamir et al. 2022; Gao et al. 2024c; Wang et al. 2022) for fusing cross-stage intra-scale features commonly rely on point-wise convolution following direct addition. However, this approach fails to effectively establish

information interaction between features and determine the relative significance of different features. To this end, we propose the multi-dimension feature interaction block. As shown in Figure 4, with two given sets of input features, we initially employ the point-wise convolution to derive the query representation $\boldsymbol{Q}$ and value representation $\boldsymbol{V}$ of the encoding features $\boldsymbol{X}_e$. Similarly, we acquire the key representation $\boldsymbol{K}$ and value representation $\boldsymbol{V}$ of the decoding features $\boldsymbol{X}_d$ using the same strategy. Each of these four representations are then divided into three components, each corresponding to the dimensions of height, width and channel. In height dimension, we first employ the query representation $\boldsymbol{Q}_{e,h}$ of encoding features and the key representation $\boldsymbol{K}_{d,h}$ of decoding features to compute the interactive self-attention. Subsequently, we utilize the interactive attention map of height dimension to augment both sets of height-dimension value representations. The enhanced features are combined as height-dimensional fused features. This process can be expressed as

$$\begin{aligned} \boldsymbol{A}_h &= (\boldsymbol{Q}_{e,h} \otimes \boldsymbol{K}_{d,h}) \odot \alpha_h, \\ \boldsymbol{X}_h &= \boldsymbol{A}_h \otimes \boldsymbol{V}_{e,h} + \boldsymbol{A}_h^T \otimes \boldsymbol{V}_{d,h}, \end{aligned} \tag{2}$$

where $\boldsymbol{A}_h \in \mathbb{R}^{B\times W\times H\times H}$ indicates the height-dimension mutual attention map, α_h is a learnable parameter, $\boldsymbol{X}_h$ are the fused height-dimension features, $\otimes$ denotes the matrix multiplication. Furthermore, in width dimension, we utilize the query representation $\boldsymbol{Q}_{e,w}$ of encoding features and the key representation $\boldsymbol{K}_{d,w}$ of decoding features to compute the interactive attention. Subsequently, we apply the interactive attention map of width dimension to enhance both two sets of width-dimension value representations separately. This process can be expressed as

$$\begin{aligned} \boldsymbol{A}_w &= (\boldsymbol{Q}_{e,w} \otimes \boldsymbol{K}_{d,w}) \odot \alpha_w, \\ \boldsymbol{X}_w &= \boldsymbol{A}_w \otimes \boldsymbol{V}_{e,w} + \boldsymbol{A}_w^T \otimes \boldsymbol{V}_{d,w}, \end{aligned} \tag{3}$$

where $\boldsymbol{A}_w \in \mathbb{R}^{B\times H\times W\times W}$ indicates the width-dimension mutual attention map, α_w is a learnable parameter, $\boldsymbol{X}_w$ are the fused width-dimension features. In addition, in channel dimension, we employ the query representation $\boldsymbol{Q}_{e,c}$ of encoding features and the key representation $\boldsymbol{K}_{d,c}$ of decoding features to calculate the mutual attention, and then use the interactive attention map of channel dimension to enhance both two sets of channel-dimension value representations, namely

$$\begin{aligned} \boldsymbol{A}_c &= (\boldsymbol{Q}_{e,c} \otimes \boldsymbol{K}_{d,c}) \odot \alpha_c, \\ \boldsymbol{X}_c &= \boldsymbol{A}_c \otimes \boldsymbol{V}_{e,c} + \boldsymbol{A}_c^T \otimes \boldsymbol{V}_{d,c}, \end{aligned} \tag{4}$$

where $\boldsymbol{A}_c \in \mathbb{R}^{B\times C\times C}$ indicates the channel-dimension mutual attention map, α_c is a learnable parameter, $\boldsymbol{X}_c$ are the fused channel-dimension features. This dynamic weighting mechanism enhances the model ability to adapt to various features, resulting in a context-sensitive representation of the input and further boosting the reconstruction performance. Additionally, the multi-dimension decoupling approach facilitates the precise integration of different features.

Coarse-scale Degradation-free Implicit Neural Representations

Although (Chen, Pan, and Dong 2024) have introduced the implicit neural representations in image deraining (Wen et al. 2024b), which fails to fully explore how these representations can normalize various weather degradations. However, images captured in adverse weather conditions often display diverse degradation patterns and intensities, presenting significant challenges for current methods, which struggle particularly with complex scenarios involving multiple stochastic degradations. Recent advancements in implicit neural representations (Lee and Jin 2022) have shown their ability to encode images as continuous functions. This approach aids in normalizing various levels of weather degradations, leading to more uniform degradation distribution and simplifying the subsequent image processing tasks (Yang et al. 2023). To this end, in our proposed coarse-scale degradation-free implicit neural representations, we employ a simple lightweight network to extract the high-dimension features $\boldsymbol{E} \in \mathbb{R}^{H\times W\times D}$ from the coarse-scale degraded images. Meanwhile, we capture the spatial information of weather degradations through the relative coordinate set $\boldsymbol{X} \in \mathbb{R}^{H\times W\times 2}$, where D denotes the dimensions of the learned features and 2 signifies the horizontal and vertical coordinates. As depicted in Figure 2, we concatenate the positional coding $\boldsymbol{X}$ with the features $\boldsymbol{E}$, followed by employing multi-layer perceptron (MLP) to generate the neural representation images

$$\boldsymbol{I}_{inr}(i,j) = MLP(\boldsymbol{E}(i,j), \boldsymbol{X}(i,j)), \tag{5}$$

where (i,j) is the location of a pixel. In our approach, we directly impose a constraint on $\boldsymbol{I}_{inr}$ to mirror the corresponding clean images through the utilization of the $L1$-norm. As recommended in (Lee and Jin 2022), preceding the fusion of coordinates with image features, we employ a high-frequency function $\mathcal{G}(\cdot)$ to transform the original coordinates $\boldsymbol{X}$ from $\mathbb{R}$ into a higher-dimensional space $\mathbb{R}^{2L}$. This transformation can be expressed as

$$\mathcal{G}(x) = [\cdots, \sin(2^i - \pi x), \cos(2^i - \pi x), \cdots], \tag{6}$$

where i ranges from 0 to $L-1$, and L denotes a hyper-parameter governing the dimension. Instead of employing original-scale implicit neural representations for image representation (Yang et al. 2023), we introduce a coarse-scale image representation strategy (Chen, Pan, and Dong 2024) to generate the clean representations based on the given degraded images, which are then utilized to augment the input of fine-scale reconstruction network. This approach enables the fine-scale network to accurately capture the features of degraded images and simplifies the complexity of subsequent processes (Chen, Pan, and Dong 2024). Meanwhile, our neural representations branch is also consistent with the visual self-prompting paradigm (Wang et al. 2023; Ma et al. 2023). Unlike existing algorithms that depend on fixed image pairs or predefined features (Wang et al. 2023; Ma et al. 2023), our visual self-prompting mechanism enables ongoing adaptation to the specific degraded images provided as input. This dynamic adaptability improves the versatility of our model in handling complex degradations.

Method	FoggyCityscapes		
	PSNR	SSIM	LPIPS
All-in-One	28.356	0.9353	0.0719
TransWeather	32.130	0.9754	0.0263
TKL	34.293	0.9824	0.0228
AirNet	31.448	0.9712	0.0314
AIRFormer	34.090	0.9816	0.0235
WeatherDiff	35.165	0.9887	0.0119
PromptIR	37.027	0.9910	0.0110
DyNet	35.949	0.9890	0.0130
MPMF-Net	37.592	0.9910	0.0105

(a) Image dehazing

Method	RainCityscapes		
	PSNR	SSIM	LPIPS
All-in-One	29.741	0.9204	0.1254
TransWeather	33.928	0.9695	0.0283
TKL	34.348	0.9740	0.0298
AirNet	32.873	0.9646	0.0349
AIRFormer	34.282	0.9736	0.0332
WeatherDiff	35.849	0.9864	0.0122
PromptIR	37.040	0.9853	0.0165
DyNet	37.191	0.9853	0.0155
MPMF-Net	37.435	0.9857	0.0131

(b) Rain-by-haze removal

Method	RSCityscapes		
	PSNR	SSIM	LPIPS
All-in-One	27.278	0.8229	0.2561
TransWeather	31.160	0.9235	0.0698
TKL	31.768	0.9339	0.0691
AirNet	30.272	0.9110	0.0879
AIRFormer	31.711	0.9372	0.0669
WeatherDiff	33.350	0.9623	0.0337
PromptIR	34.314	0.9604	0.0426
DyNet	34.286	0.9606	0.0399
MPMF-Net	34.670	0.9620	0.0330

(c) Rain-by-snow removal

Method	SnowTrafficData		
	PSNR	SSIM	LPIPS
All-in-One	25.267	0.8591	0.1955
TransWeather	25.971	0.9075	0.1032
TKL	25.529	0.9015	0.1228
AirNet	23.534	0.8978	0.1205
AIRFormer	26.132	0.9040	0.1294
WeatherDiff	26.428	0.9207	0.0853
PromptIR	26.139	0.9227	0.0862
DyNet	25.949	0.9262	0.0797
MPMF-Net	27.053	0.9280	0.0819

(d) Image desnowing

Method	LowLightTrafficData		
	PSNR	SSIM	LPIPS
All-in-One	17.160	0.5692	0.4965
TransWeather	21.259	0.7371	0.3695
TKL	17.885	0.5810	0.4326
AirNet	20.967	0.7315	0.3779
AIRFormer	19.824	0.7492	0.4124
WeatherDiff	18.148	0.6541	0.3440
PromptIR	21.328	0.7662	0.3758
DyNet	22.088	0.7690	0.3728
MPMF-Net	22.584	0.7658	0.3563

(e) Image enhancement

Method	RainDS-syn		
	PSNR	SSIM	LPIPS
All-in-One	23.739	0.7134	0.1105
TransWeather	26.700	0.8456	0.1844
TKL	28.118	0.8568	0.1801
AirNet	25.559	0.8258	0.2154
AIRFormer	26.801	0.8322	0.2339
WeatherDiff	29.733	0.8971	0.1166
PromptIR	30.220	0.9123	0.1057
DyNet	29.954	0.9075	0.1182
MPMF-Net	30.185	0.9057	0.0958

(f) Rain-by-raindrop removal

Table 1: Quantitative comparisons of the different methods on all-in-one traffic weather removal. The red and blue metrical scores denote the best and second-best quantitative performance. Our proposed method achieves the best metrical scores across the six weather-degraded datasets.

Experimental Results

Implementation Details

Our model is constructed on NVIDIA Tesla A40 GPU. During the 400 training epochs, we utilize a patch size of 256×256 and a batch size of 32. The Adam optimizer (Kingma and Ba 2014) is employed, and we implement a learning rate schedule that decreases from 2×10^{-4} to 1×10^{-7} using the cosine annealing decay (Loshchilov and Hutter 2022). The feature dimensions and basic learning depths are set to $\{24, 48, 96, 48, 24, 24\}$ and $\{8, 16, 17, 9, 9, 8\}$, respectively. Meanwhile, the sizes of learnable prompts are set to 16×16, 32×32 and 64×64, respectively. Additionally, we employ random horizontal and vertical flips to augment the training dataset.

Datasets and Evaluation Protocols

We conduct experiments (following (Wen et al. 2024a)) using FoggyCityscapes (Sakaridis, Dai, and Van Gool 2018), RainCityscapes (Hu et al. 2019), RSCityscapes (Wen et al. 2024c), SnowTrafficData (Chen et al. 2023a), LowLightTrafficData (Li et al. 2018) and RainDS-syn (Quan et al. 2021) datasets, corresponding to haze, rain-by-haze, rain-by-snow, snow, low light and rain-by-raindrop degradations. We calculate the peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and learned perpetual image patch similarity (LPIPS) (Zhang et al. 2018) between the reconstructed and clean images, where decreased LPIPS values correspond to the enhanced image quality, and vice versa for the others.

Comparisons with Existing Methods

We conduct experiments to evaluate the performance of our proposed method with eight existing all-in-one approaches, namely All-in-One (Li, Tan, and Cheong 2020), TransWeather (Valanarasu, Yasarla, and Patel 2022), TKL (Chen et al. 2022), AirNet (Li et al. 2022), AIRFormer (Gao et al. 2024c), WeatherDiff (Özdenizci and Legenstein 2023), PromptIR (Potlapalli et al. 2023) and DyNet (Dudhane et al. 2024). We train our model on the all-in-one dataset (with six degradations simultaneously) and subsequently test the trained model on six individual testing datasets separately. As illustrated in Table 1, compared with several well-performing methods, our proposed model obtains consistent quantitative superiority. We also calculate the average metrical scores of different all-in-one methods in Table 6, where our proposed model advances the existing well-performing PromptIR (Potlapalli et al. 2023) and DyNet (Dudhane et al. 2024) by 0.576 dB and 0.684 dB in PSNR, respectively. Meanwhile, as the visual results and error maps in Figure 5 depicted, our proposed model also achieves the best vision reconstruction.

Ablation Studies

We conduct ablation experiments to evaluate the effectiveness of our proposed multi-axis prompt and multi-dimension

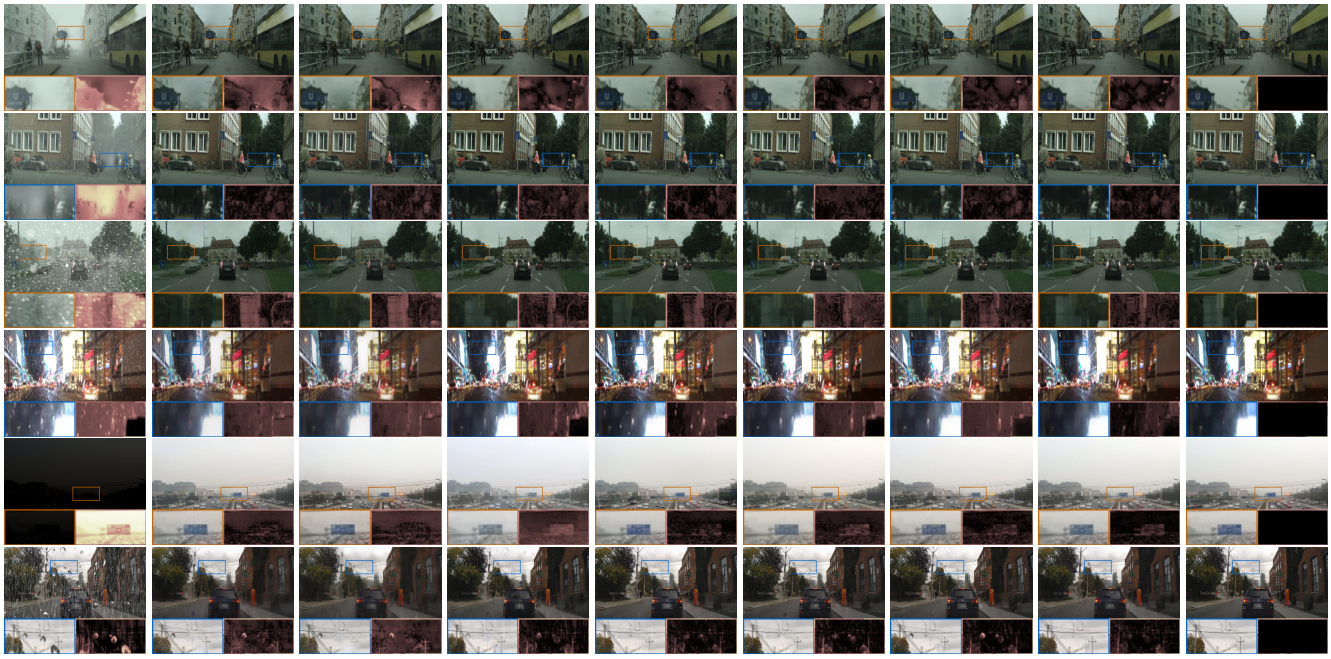

Figure 5: Visual comparisons of different all-in-one methods. Each reconstructed image is accompanied by its respective error map corresponding to the clean image. Our proposed approach effectively mitigates degradation, yielding superior visual outcomes with minimal error maps.

fusion network. All the experimental results are calculated by averaging the metrical scores across the six testing datasets.

Individual Components We start with a baseline model based on a commonly used encoder-decoder architecture with the basic learning block and convolution fusion approach. As shown in Table 2, incorporating coarse-scale degradation-free implicit neural representations into the baseline model results in a performance gain of 0.741 dB. Additionally, replacing the convolution fusion approach with our proposed multi-dimension feature interaction block improves the restoration performance by 0.516 dB. Furthermore, introducing the multi-axis prompts learning block into the decoding stage enhances the overall performance by an additional 0.718 dB.

Model	PSNR	SSIM	LPIPS
Baseline	29.612	0.8949	0.1297
w/ CDINR	30.353	0.9098	0.1209
w/ MFIB	30.869	0.9140	0.1112
w/ MPLB	31.587	0.9230	0.0984

Table 2: Ablation experiments on individual components. Each presents a positive effect on the overall performance.

Feature Interaction As shown in Table 3, convolution fusion approach achieves a performance gain of 0.368 dB over the simple addition. Furthermore, employing SKFusion (Song et al. 2023) results in a performance gain of 0.674 dB over the convolution. However, our proposed MFIB achieves a performance gain of 0.669 dB over SKFusion.

Component	PSNR	SSIM	LPIPS
Addition	29.876	0.8930	0.1249
Convolution	30.244	0.8991	0.1202
SKFusion	30.918	0.9097	0.1126
MFIB	31.587	0.9230	0.0984

Table 3: Ablation experiments on the feature interaction approaches. Our proposed multi-dimension feature interaction block advances the other intra-scale feature fusion strategies.

Prompt Learning Since different prompts learning methods cannot be directly embedded into our network, we directly integrate their modules containing prompts learning into our network for comparative experiments. As Table 4 depicted, our proposed multi-axis prompts learning block achieves performance gains of 2.265 dB, 1.757 dB and 0.833 dB over the prompts learning approaches in TransWeather (Valanarasu, Yasarla, and Patel 2022), AIRFormer (Gao et al. 2024c) and PromptIR (Potlapalli et al. 2023), respectively. Additionally, our method utilizes the fewest learnable prompt parameters.

Implicit Neural Representations To understand the effect of implicit neural representations, we first visualize the outputs of our proposed CDINR and the pixel value distributions of images in Figure 6. Since RSCityscapes dataset

Prompt	Number	PSNR	SSIM	LPIPS
in TransWeather	7.37×10^4	29.322	0.8891	0.1431
in AIRFormer	9.83×10^4	29.830	0.8984	0.1307
in PromptIR	2.38×10^6	30.754	0.9179	0.1103
MPLB	6.04×10^4	31.587	0.9230	0.0984

Table 4: Ablation experiments on prompts learning approaches. Our proposed multi-axis prompts learning block achieves better performance over the other prompts learning strategies.

(Wen et al. 2024c) is generated based on RainCityscapes dataset (Hu et al. 2019), we can easily select images with the same background but different degradation distributions as examples. As depicted, the two inputs of CDINR show significant differences in the pixel value distributions. However, the representation images through our CDINR are degradation-free and the pixel value distributions are normalized to close levels.

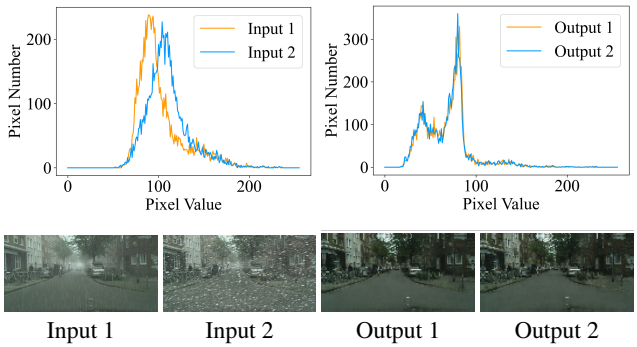

Figure 6: Comparisons of the pixel value distributions between the input and output of our proposed coarse-scale degradation-free implicit neural representations. Our CDINR can normalize the degradation levels of different weather conditions.

Image Scale As Table 5 reported, when utilizing the original degraded images as the input to our CDINR, our model performs the worst. This indicates that directly representing non-degraded images at their original scale is challenging due to the presence of numerous degraded details in fine-scale images. Moreover, when using $\frac{1}{2}$-scale and $\frac{1}{8}$-scale degraded images as inputs to CDINR, our model shows performance reductions of 0.654 dB and 1.062 dB, respectively, compared to using $\frac{1}{4}$-scale degraded images.

Efficiency Analysis As Table 6 reported, our model utilizes the fewest parameters (#Param). Additionally, AIRFormer (Gao et al. 2024c) requires the least inference time. However, compared with the most recent three methods (see the sub-box), our model demonstrates significant improvements in inference time. Specifically, our model utilizes only 62.765% inference time of the recent well-performing PromptIR (Potlapalli et al. 2023), while achieving a performance gain of 0.576 dB. In Figure 1, we demonstrate that our proposed approach strikes an optimal balance between image quality and inference time.

Scale	PSNR	SSIM	LPIPS
Original	29.580	0.8982	0.1320
0.5	30.933	0.9147	0.1064
0.25	31.587	0.9230	0.0984
0.125	30.525	0.9150	0.1109

Table 5: Ablation experiments on representation scales. $\frac{1}{4}$-scale degradation-free representations achieve the best performance.

Method	#Param	Time	PSNR	SSIM	LPIPS
All-in-One	50.226	0.0368	25.257	0.8034	0.2093
TransWeather	37.682	0.0185	28.525	0.8931	0.1303
TKL	28.713	0.0329	28.657	0.8716	0.1429
AirNet	5.7665	0.0984	27.442	0.8837	0.1447
AIRFormer	58.383	0.0139	28.807	0.8963	0.1499
WeatherDiff	1998.7	10.804	29.779	0.9016	0.1006
PromptIR	32.966	0.0991	31.011	0.9230	0.1063
DyNet	14.227	0.0933	30.903	0.9229	0.1065
MPMF-Net	3.2393	0.0622	31.587	0.9230	0.0984

Table 6: Efficiency analysis of the different all-in-one methods. Our proposed method achieves the best trade-off between the quality of generated images and inference time.

Limitations Our model is limited by the narrow scope of training dataset, which covers only six types of weather degradations and restricts the trained model to handle complex weather conditions such as dust and back-light. Furthermore, our research primarily centers on enhancing the quality of traffic weather-degraded images and our experimental focus remains confined to image reconstruction, preventing the utilization of our model as a pre-processing operator. However, as emphasized in (Jiang et al. 2020; Liu et al. 2023), the imperative for achieving superior performance in subsequent high-level vision tasks evidently lies in the accuracy of weather degradation removal.

Conclusion

We present a novel method to enhance visibility in traffic scenes affected by adverse weather conditions. Specifically, we tackle a significant drawback of existing prompts learning approaches by leveraging the prompts from three axes. Furthermore, we explore the dimension-wise decomposition and adaptive weighting strategies for integrating features across different stages, facilitating multi-dimension interaction among height, width and channel dimensions. Additionally, we demonstrate that implicit neural representations effectively normalize the degradation levels of different weather conditions.

Acknowledgments

This research is partially supported by the National Key R&D Program of China (Grant No. 2023YFB2504703), the Shaanxi International S&T Cooperation Program Project (Grant No. 2024GH-YBXM-24), the National Natural Science Foundation of China (Grant No. 52172379), and the Fundamental Research Funds for the Central Universities (Grant No. 300102242901).

References

Bar, A.; Gandelsman, Y.; Darrell, T.; Globerson, A.; and Efros, A. 2022. Visual prompting via image inpainting. *Advances in Neural Information Processing Systems*, 35: 25005–25017.

Chen, H.; Ren, J.; Gu, J.; Wu, H.; Lu, X.; Cai, H.; and Zhu, L. 2023a. Snow Removal in Video: A New Dataset and A Novel Method. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 13165–13176. IEEE.

Chen, J.; Kao, S.-h.; He, H.; Zhuo, W.; Wen, S.; Lee, C.-H.; and Chan, S.-H. G. 2023b. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12021–12031.

Chen, W.-T.; Huang, Z.-K.; Tsai, C.-C.; Yang, H.-H.; Ding, J.-J.; and Kuo, S.-Y. 2022. 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*, 17653–17662.

Chen, X.; Li, H.; Li, M.; and Pan, J. 2023c. Learning a sparse transformer network for effective image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5896–5905.

Chen, X.; Pan, J.; and Dong, J. 2024. Bidirectional Multi-Scale Implicit Neural Representations for Image Deraining.

Ding, J.; Du, Y.; Li, W.; Pei, L.; and Cui, N. 2025a. LG-Diff: Learning to follow local class-regional guidance for nearshore image cross-modality high-quality translation. *Information Fusion*, 117: 102870.

Ding, J.; Li, W.; Yang, M.; Zhao, Y.; Pei, L.; and Tian, A. 2025b. SeaTrack: Rethinking Observation-Centric SORT for Robust Nearshore Multiple Object Tracking. *Pattern Recognition*, 159: 111091.

Dudhane, A.; Thawakar, O.; Zamir, S. W.; Khan, S.; Khan, F. S.; and Yang, M.-H. 2024. Dynamic Pre-training: Towards Efficient and Scalable All-in-One Image Restoration. *arXiv preprint arXiv:2404.02154*.

Gao, T.; Li, Z.; Wen, Y.; Chen, T.; Niu, Q.; and Liu, Z. 2024a. Attention-Free Global Multiscale Fusion Network for Remote Sensing Object Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–14.

Gao, T.; Wen, Y.; Zhang, K.; Zhang, J.; Chen, T.; Liu, L.; and Luo, W. 2024b. Frequency-Oriented Efficient Transformer for All-in-One Weather-Degraded Image Restoration. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(3): 1886–1899.

Gao, T.; Wen, Y.; Zhang, K.; Zhang, J.; Chen, T.; Liu, L.; and Luo, W. 2024c. Frequency-Oriented Efficient Transformer for All-in-One Weather-Degraded Image Restoration. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(3): 1886–1899.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Hu, X.; Fu, C.-W.; Zhu, L.; and Heng, P.-A. 2019. Depth-attentional features for single-image rain removal. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 8022–8031.

Jiang, J.; Zuo, Z.; Wu, G.; Jiang, K.; and Liu, X. 2024. A survey on all-in-one image restoration: Taxonomy, evaluation and future trends. *arXiv preprint arXiv:2410.15067*.

Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Huang, B.; Luo, Y.; Ma, J.; and Jiang, J. 2020. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8346–8355.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lee, J.; and Jin, K. H. 2022. Local texture estimator for implicit representation function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1929–1938.

Li, B.; Liu, X.; Hu, P.; Wu, Z.; Lv, J.; and Peng, X. 2022. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17452–17462.

Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; and Wang, Z. 2018. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1): 492–505.

Li, J.; Zheng, K.; Gao, L.; Ni, L.; Huang, M.; and Chanussot, J. 2024. Model-Informed Multistage Unsupervised Network for Hyperspectral Image Super-Resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–17.

Li, R.; Tan, R. T.; and Cheong, L.-F. 2020. All in one bad weather removal using architectural search. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3175–3185.

Liu, R. W.; Lu, Y.; Guo, Y.; Ren, W.; Zhu, F.; and Lv, Y. 2023. AiOENet: All-in-One Low-Visibility Enhancement to Improve Visual Perception for Intelligent Marine Vehicles Under Severe Weather Conditions. *IEEE Transactions on Intelligent Vehicles*.

Loshchilov, I.; and Hutter, F. 2022. SGDR: Stochastic Gradient Descent with Warm Restarts. In *International Conference on Learning Representations*.

Luo, Z.; Gustafsson, F. K.; Zhao, Z.; Sjölund, J.; and Schön, T. B. 2023. Controlling vision-language models for universal image restoration. *arXiv preprint arXiv:2310.01018*.

Ma, J.; Cheng, T.; Wang, G.; Zhang, Q.; Wang, X.; and Zhang, L. 2023. Prores: Exploring degradation-aware visual prompt for universal image restoration. *arXiv preprint arXiv:2306.13653*.

Özdenizci, O.; and Legenstein, R. 2023. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Potlapalli, V.; Zamir, S. W.; Khan, S.; and Khan, F. S. 2023. PromptIR: Prompting for All-in-One Blind Image Restoration. *arXiv preprint arXiv:2306.13090*.

Quan, R.; Yu, X.; Liang, Y.; and Yang, Y. 2021. Removing raindrops and rain streaks in one go. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9147–9156.

Ruan, J.; Xie, M.; Gao, J.; Liu, T.; and Fu, Y. 2023. Egeunet: an efficient group enhanced unet for skin lesion segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 481–490. Springer.

Sakaridis, C.; Dai, D.; and Van Gool, L. 2018. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126: 973–992.

Song, Y.; He, Z.; Qian, H.; and Du, X. 2023. Vision Transformers for Single Image Dehazing. *IEEE Transactions on Image Processing*, 32: 1927–1941.

Valanarasu, J. M. J.; Yasarla, R.; and Patel, V. M. 2022. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2353–2363.

Wang, X.; Wang, W.; Cao, Y.; Shen, C.; and Huang, T. 2023. Images speak in images: A generalist painter for in-context visual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6830–6839.

Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; and Li, H. 2022. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 17683–17693.

Wen, Y.; Gao, T.; and Chen, T. 2024a. Neural Schrödinger bridge for unpaired real-world image deraining. *Information Sciences*, 682: 121199.

Wen, Y.; Gao, T.; and Chen, T. 2024b. Unpaired Photo-realistic Image Deraining with Energy-informed Diffusion Model. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 360–369.

Wen, Y.; Gao, T.; Li, Z.; Zhang, J.; Zhang, K.; and Chen, T. 2024a. All-in-one Weather-degraded Image Restoration via Adaptive Degradation-aware Self-prompting Model. *arXiv preprint arXiv:2411.07445*.

Wen, Y.; Gao, T.; Zhang, J.; Li, Z.; and Chen, T. 2023. Encoder-Free Multiaxis Physics-Aware Fusion Network for Remote Sensing Image Dehazing. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–15.

Wen, Y.; Gao, T.; Zhang, J.; Zhang, K.; and Chen, T. 2024b. From heavy rain removal to detail restoration: A faster and better network. *Pattern Recognition*, 148: 110205.

Wen, Y.; Gao, T.; Zhang, K.; Cheng, P.; and Chen, T. 2024c. Restoring vision in rain-by-snow weather with simple attention-based sampling cross-hierarchy Transformer. *Pattern Recognition*, 110743.

Yang, S.; Ding, M.; Wu, Y.; Li, Z.; and Zhang, J. 2023. Implicit neural representation for cooperative low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12918–12927.

Ye, T.; Chen, S.; Bai, J.; Shi, J.; Xue, C.; Jiang, J.; Yin, J.; Chen, E.; and Liu, Y. 2023. Adverse Weather Removal with Codebook Priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12653–12664.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5728–5739.

Zhang, K.; Li, R.; Yu, Y.; Luo, W.; and Li, C. 2021. Deep dense multi-scale network for snow removal using semantic and depth priors. *IEEE Transactions on Image Processing*, 30: 7419–7431.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.