Robust Adverse Weather Removal via Spectral-based Spatial Grouping

Yuhwan Jeong*, Yunseo Yang*, Youngho Yoon*, and Kuk-Jin Yoon
Visual Intelligence Lab., KAIST, Korea
{jeongyh98,acorn,dudgh1732,kjyoon}@kaist.ac.kr

Abstract

Adverse weather conditions cause diverse and complex degradation patterns, driving the development of All-in-One (AiO) models. However, recent AiO solutions still struggle to capture diverse degradations, since global filtering methods like direct operations on the frequency domain fail to handle highly variable and localized distortions. To address these issue, we propose Spectral-based Spatial Grouping Transformer (SSGformer), a novel approach that leverages spectral decomposition and group-wise attention for multi-weather image restoration. SSGformer decomposes images into high-frequency edge features using conventional edge detection and low-frequency information via Singular Value Decomposition. We utilize multi-head linear attention to effectively model the relationship between these features. The fused features are integrated with the input to generate a grouping-mask that clusters regions based on the spatial similarity and image texture. To fully leverage this mask, we introduce a group-wise attention mechanism, enabling robust adverse weather removal and ensuring consistent performance across diverse weather conditions. We also propose a Spatial Grouping Transformer Block that uses both channel attention and spatial attention, effectively balancing feature-wise relationships and spatial dependencies. Extensive experiments show the superiority of our approach, validating its effectiveness in handling the varied and intricate adverse weather degradations.

1. Introduction

Adverse weather conditions, such as snow, rain, raindrop, and haze, severely degrade visual information, making image restoration an essential task for various applications [20, 24, 58]. To address this issue, researchers have explored methods to remove weather-induced degradation from images [5, 27, 31, 37, 38, 40, 59, 62, 68]. In particular, significant efforts have been made to leverage image-based priors, such as edges and fourier patterns, as a foundation for restoring degraded images [22, 43, 64, 75, 76].

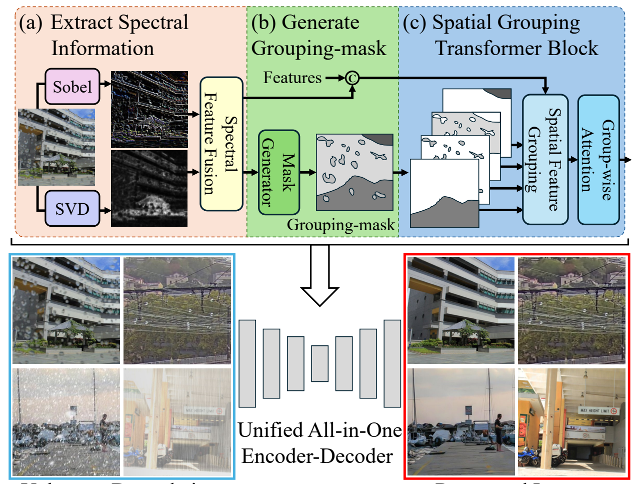


Figure 1. **Overview.** We propose SSGformer, an All-in-One adverse weather removal model with three key components. (a) Sobel operator and SVD extract spectral information from degraded image, producing degradation-aware features through spectral feature fusion. (b) The fused features identify degradation patterns and generate a "grouping-mask" for spatial grouping. (c) Features are grouped by the mask for group-wise spatial/channel attention.

Early research on adverse weather removal primarily focused on handling single weather conditions [4, 8, 46, 60]. However, as the demand for comprehensive models to address multiple types of degradation grew, the focus shifted toward All-in-One (AiO) approaches for adverse weather removal [7, 12, 26, 30, 33–36, 49, 61, 63, 70, 72, 77, 78].

Despite these advancements, developing a unified model capable of handling diverse weather conditions remains a major challenge due to the highly variable nature of weather-induced degradations. The complex interactions between different atmospheric conditions lead to unpredictable visual degradations, making it difficult to create a single framework that generalizes well across all scenarios.

To address the challenges posed by adverse weather conditions, several studies have explored image restoration techniques in the frequency domain. AIRFormer [18], a general-purpose AiO model, constructs a frequency-

*Denotes equal contribution.
The code is available at github.com/jeongyh98/SSGformer.

guided transformer encoder by incorporating wavelet-based prior information to enhance feature extraction, leveraging structured frequency priors to improve restoration quality. Fourmer [73], similarly, employs the Fourier transform to disentangle image degradation from content components, utilizing the global nature of the Fourier domain to process degradation patterns comprehensively. AdaIR [13] is designed to identify and mitigate degradation patterns through frequency domain analysis. These methods show the potential of frequency-based restoration, which transforms an image into the frequency domain and manipulates its frequency components. However, operations in the frequency domain that affect the entire region modify the frequency content of the entire image. This global filtering method may be suitable for addressing repetitive patterns, such as blur or noise. However, it is not effective for handling random and localized degradations that occur in adverse weather conditions. Additionally, while the wavelet transform processes images by dividing them into frequency bands, it reduces the image resolution by half. Consequently, to use this frequency information, the image needs to be resized back to its original dimensions, which may introduce distortions. Therefore, an effective solution is required that extracts valuable spectral information for frequency-based image restoration, while preserving spatial details to prevent the loss of crucial contextual information.

To conquer this problem, we propose SSGformer, which not only extracts essential features using image-based priors but also ensures their effective integration within a comprehensive framework. As shown in Fig. 1 (a), the first step extracts edge and low-rank information from the input image while preserving spatial details. Conventional edge detection captures high-frequency edge features, while Singular Value Decomposition (SVD) analyzes degradation textures in the low-frequency domain. To model the relationships between spectral information, we use multi-head linear attention to capture the interdependencies within the spectral features. Building on this, we expand the information exchange beyond local spectral regions by identifying and comparing spatially similar areas across the image, facilitating broader degradation-aware feature sharing and more efficient grouping. To achieve the grouping, we employ a mask generator to capture spatial similarity from the spectral information and generate a grouping-mask that clusters relevant features based on their values (Fig. 1 (b)).

To effectively utilize these grouped features and facilitate global information exchange, we design a specialized architecture tailored to optimize their integration. In this architecture, we propose attention mechanisms that operate on mask-based grouped features exhibiting similar spatial characteristics as shown in Fig. 1 (c). These group-wise attention is applied both spatially and across channels, allowing the model to focus on the most relevant features. Consequently, our proposed architecture enables robust degradation removal across diverse adverse weather conditions.

Our contributions can be summarized as:

- We propose SSGformer, an innovative transformer architecture that enhances restoration effectively handling adverse weather conditions.
- We reveal that directly extracting image information via edge detection and SVD is effective for robust adverse weather removal. Moreover, efficient grouping of this information is key to performance gains.
- Our extensive experiments showcase the robust restoration capability of SSGformer, achieving state-of-the-art performance across a complex weather degradation.

2. Related works

2.1. Task-specific adverse weather removal

Attempts to restore images degraded by adverse weather conditions to clean images initially began as single-task approaches. Common restoration tasks involves deraining, desnowing, dehazing, and raindrop removal.

Deraining aims to remove rain streaks from images. In the early stages, the task was solved using hand-crafted or rule-based methods [10, 22]. The development of single image deraining has entered the deep learning era, leading to the emergence of numerous deep learning-based approaches [9, 16, 17, 52]. Some of works utilized a conditional GAN [27, 68], edge loss [59], and a multi-scale residual network [62] for deraining. Since transformers emerged [50], studies [8, 54] have leveraged ViT's image analysis capabilities [14].

Dehazing has been widely studied across various fields, thereby aiding daily life. Early methods [1, 42] utilized CNNs for dehazing. Proposed learning methods include considering both atmospheric light and transmission map [25], GAN [38, 60], and contrastive learning [53].

Desnowing [2, 4, 31, 41] tasks remove snow particles from an image and fill in the empty areas behind them. The research has been proposed using a method that is aware of size and transparency [5], semantic and geometric priors [71], and dual-tree wavelet transform [6].

Rain drop removal has been studied using temporal information from videos [65] in the early era. During the machine learning era, CNNs [15] were initially used to clean images, followed by enhancements through GAN-based methods [37, 55]. More research [19, 57] has been widely proposed using edge information [40], multi-scale attention [44], and transformer [46].

2.2. All-in-One model for adverse weather removal

Early restoration methods were constrained by the need for separate, task-specific models for each adverse weather condition. This limitation has led to efforts to develop unified frameworks. Li *et al.* [28] introduced task-specific en-

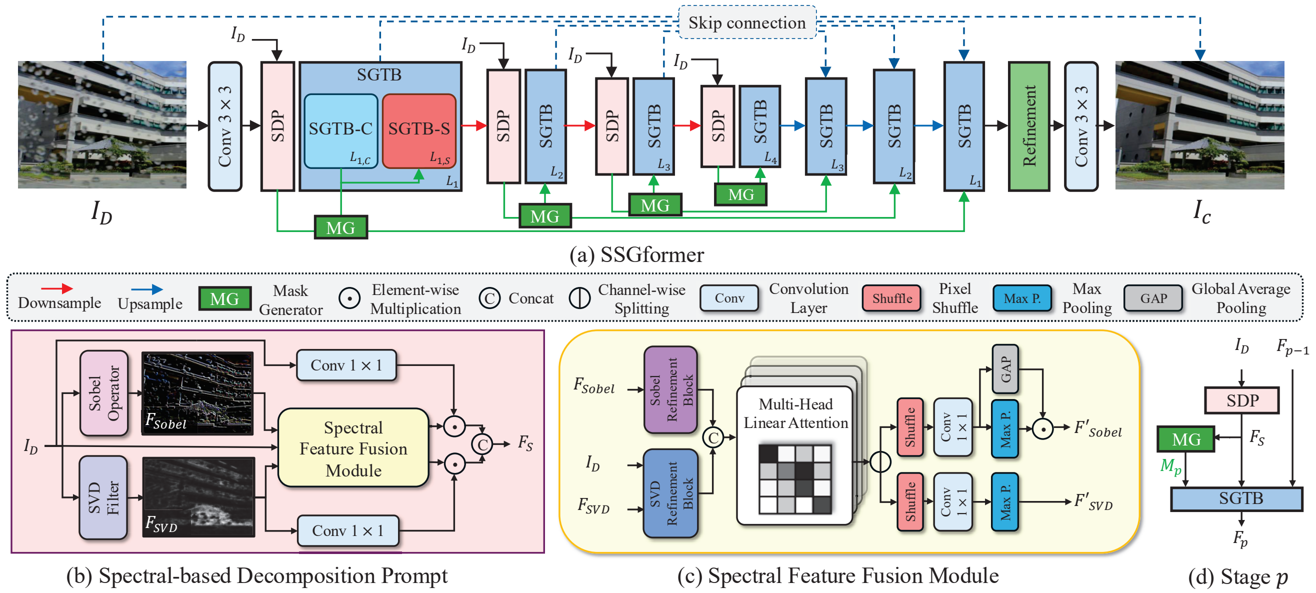


Figure 2. **Overall framework.** (a) SSGformer is a transformer-based encoder-decoder network that restores a weather-degraded image I_D to a clean image I_C. The encoder consists of Spectral-based Decomposition Prompt (SDP), Mask Generator (MG), and Spatial Grouping Transformer Block (SGTB), which include two types of blocks: SGTB-C and SGTB-S. At stage p, there exist as many blocks as the number indicated by L_p. The stage p increases from 1 to 4 and then decreases back to 1 in the decoder. (b) SDP takes the input image I_D, passes it through Sobel operator and SVD filter to generate F_{Sobel} and F_{SVD}, which are processed by the Spectral Feature Fusion Module to produce F_S. (c) In the Spectral Feature Fusion Module, F_{Sobel}, F_{SVD}, and I_D, are passed through multi-head linear attention to obtain the fused features F'_{Sobel} and F'_{SVD}. (d) Feature flow at p-th stage in the encoder. F_S from SDP is sent to both MG and SGTB, while the previous stage output F_{p-1} and the "grouping-mask" M_p from MG are fed into SGTB to produce F_p.

coders, which were later expanded upon with contrastive learning [26]. Subsequent advancements include two-stage knowledge distillation [7], single encoder-decoder architectures [49], and the decoupling of general and specific weather features [78]. Ye *et al.* [63] utilized quantized codebook prior to remove degradation, while Sun *et al.* [47] further improved restoration quality by grouping similar features using pixel-sorting. However, in some challenging scenarios, intensity may not sufficiently distinguish semantically different regions, possibly affecting attention behavior. Although recent methods [32, 56] incorporate external knowledge through LLMs or VLMs, our approach is fundamentally grounded in intra-image information without resorting to any external source.

3. Methods

3.1. Overview

Our goal is to restore images damaged by random degradations and highly diverse weather conditions within a unified framework. Many existing AiO models address severe weather degradation through global filtering but struggle with highly variable conditions. To solve this object, one of our key idea is that adding comparisons between similar features through grouping can lead to more robust adverse weather removal. Moreover, grouping features based on spectral characteristics provides valuable information for image restoration. With these ideas, we propose a novel architecture, **SSGformer**, that captures inherent variability and randomness through spectral analysis and effectively integrates them within a comprehensive framework.

SSGformer is a transformer-based unified 4-stage encoder-decoder network designed to restore a weather-degraded image I_D to a clean image I_C. The encoder consists of a Spectral-based Decomposition Prompt (SDP), a Mask Generator (MG), and a Spatial Grouping Transformer Block (SGTB). The decoder incorporates SGTB and a refinement block. For more details, please refer to Fig. 2 (a).

To consider the pixel similarity, it is necessary to group the features, which requires a criterion for dividing the groups. Therefore, we begin the design process selecting the information to determine the rule. To extract information for the criterion, SDP is designed to perform spectral analysis using two types of filters, the Sobel operator and SVD filter, and fuse the features with multi-head linear attention. Then, the MG generates a grouping-mask from degradation-aware feature F_S, the output of the SDP, enabling the spatial grouping of features with similar characteristics. Using this mask, SGTB applies group-wise attention to enhance feature interactions. Depending on the type of attention used, SGTB is categorized into SGTB-C for channel attention, and SGTB-S for spatial attention. Detailed descriptions of SDP, MG, and SGTB are provided in Sec. 3.2, Sec. 3.3, and Sec. 3.4, respectively.

3.2. Spectral-based Decomposition Prompt

Spectral-based Decomposition Prompt (SDP) (Fig. 2 (b)) is divided into two stages. In the first stage, spectral information of the input image I_D is decomposed using multiple filters. Subsequently, we integrate these features through our Spectral Feature Fusion Module and convolution layers, yielding the degradation-aware feature F_S.

Spectral analysis. To extract essential priors related to spectral information, we employ two decomposition filters. One is the Sobel operator which detects high frequency information by highlighting intensity changes. The other, Singular Value Decomposition (SVD), finds low frequency knowledge of the degraded image. The features obtained from the two filter operations, F_{Sobel} and F_{SVD}, both in $\mathbb{R}^{H\times W\times 1}$, capture different information from the same input I_D. H and W denote the spatial dimensions at current stage. This complementary spectral analysis enables early capture of various types of degradation, which can provide valuable spectral features for effective restoration.

Spectral Feature Fusion Module. To find the relationship between each spectral feature, we adopt multi-head linear attention [23], which improves feature representation by sharing spectral information with different characteristics. Prior to this, we introduce the Sobel and SVD refinement blocks to enhance each spectral feature (see Fig. 2 (c)). The Sobel refinement block uses a convolution layer and feature reorganization, which refine information by incorporating surrounding context. The SVD refinement block consists of convolution layers, feature reorganization, and a deformable convolution layer, which refine the low-frequency components with I_D. The refined features are then passed through multi-head linear attention, which is followed by channel splitting before further processing. Afterward, we apply pixel shuffle [45] to upscale the feature, expanding the receptive field to better understand the context. The upscaled features are processed using max pooling or global average pooling to emphasize salient features and capture global context information, yielding F'_{Sobel} and F'_{SVD}.

The manipulated spectral features are combined to form the degradation-aware feature F_S. Specially, F'_{Sobel} is element-wise multiplied with the feature extracted from I_D, while the F'_{SVD} is multiplied with the F_{SVD} after a convolution layer. These two components are then integrated as:

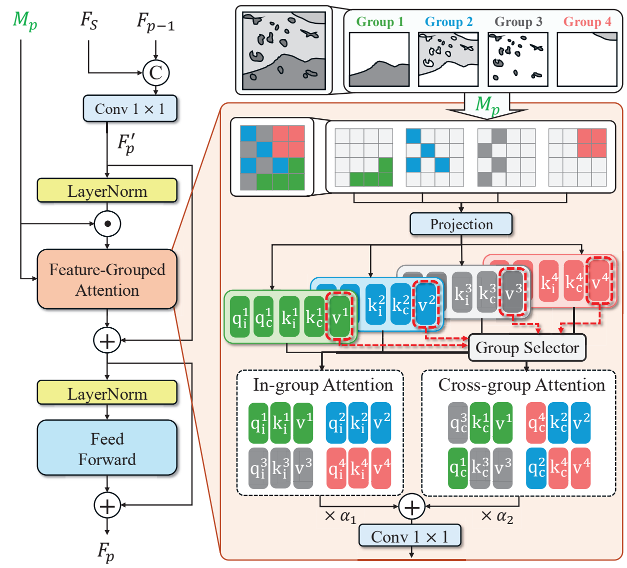
$$F_S = [F'_{\text{Sobel}} \odot \text{Conv}_1(I_D),\ F'_{\text{SVD}} \odot \text{Conv}_1(F_{SVD})] \quad (1)$$

where $\odot$ denotes element-wise multiplication. $[\bullet, \bullet]$ is the concatenation operation.

3.3. Mask Generator

Previous methods [13, 29, 47] typically rely on the feature derived from the prompt without any additional processing, often simply concatenating it and feeding it directly into the network. However, these approaches do not fully utilize the extracted information. To overcome this, we introduce a spatial grouping method that first generates a grouping-mask from the degradation-aware feature F_S to enhance feature utilization. This mask identifies regions in the feature map with similar characteristics, enabling effective spatial grouping. It plays a crucial role in focusing on regions that share similar degradation patterns, thereby improving the efficiency and effectiveness of the restoration process. At stage p, the mask generator consists of a simple convolution layer that synthesizes information across multiple channels to produce a single-channel mask, $M_p = \text{Conv}_7(F_S)$, where $M_p \in \mathbb{R}^{H\times W\times 1}$. Notably, the mask is used in both the encoder and decoder at the same stage, ensuring coherent and effective feature manipulation.

With this grouping-mask M_p, feature elements are divided into g_p number of groups, structuring them for further processing. The mask utilization is described in Sec. 3.4.



(a) SGTB (b) Feature-Grouped Attention

Figure 3. (a) **Spatial Grouping Transformer Block (SGTB)** starts with a convolution layer that fuses the previous stage features F_{p-1} and the degradation-aware features F_S. The grouping-mask M_p is element-wise multiplied with F'_p and is also fed into Feature-Grouped Attention. (b) **Illustration of Feature-Grouped Attention.** Features are grouped using M_p, which encodes spatial similarity. In-group and cross-group attention are applied simultaneously, with a group selector handling query exchanges before cross-group attention. The same color represents the same group. This is an example when the number of groups, g_p is 4.

3.4. Spatial Grouping Transformer Block

One of our key idea is analyzing spatially correlated features and considering their contextual relationships can be helpful to handle the unpredictability of weather-induced distortions. To this end, we design the Spatial Grouping

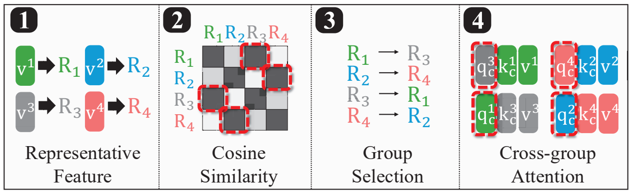


Figure 4. **Illustration of the group selector process.** (1) Each group's value is pooled to obtain a representative feature ($R_{1\sim4}$). (2) Cosine similarity is computed using the representative features to measure inter-group similarity. The darker the matrix, the higher the similarity score. (3) The most similar group is selected. (4) Attention is computed using the query from the selected group. For clarity, the example is illustrated with four groups, where the same color represents the same group.

Transformer Block (SGTB), which performs group-wise attention by leveraging spatial similarity information from generated masks. Furthermore, unlike existing transformer-based AiO models that rely solely on channel attention, SGTB incorporates not only channel attention but also spatial attention, enhancing restoration performance by facilitating information exchange along the spatial axis.

SGTB follows a standard transformer structure, comprising a normalization layer, attention, and feed-forward network, as shown in Fig. 3 (a). At the beginning of each SGTB, the features F_{p-1} from the previous stage (for the first block, the input degraded image) and the degradation-aware features F_S extracted from SDP are combined through a convolution layer as follows:

$$F'_p = \text{Conv}_1[F_{p-1}, F_S]. \tag{2}$$

The resulting feature F'_p goes through a normalization layer, Feature-Grouped Attention (FGA), and a feed-forward network (FFN), and is expressed as follows:

$$\hat{F}_p = F'_p + \text{FGA}(\text{LayerNorm}(F'_p) \odot M_p,\ M_p), \tag{3}$$

$$F_p = \hat{F}_p + \text{FFN}(\text{LayerNorm}(\hat{F}_p)). \tag{4}$$

In FGA, we employ two attention mechanisms: channel (SGTB-C) and spatial attention (SGTB-S). The grouped features undergo attention operations, either channel-wise, by emphasizing the most relevant channels, or spatially, by focusing on key regions within the feature map. These attention operations are structurally identical, with the only distinction being the dimension of the attention matrix.

Feature-Grouped Attention, which is the key component, utilizes the grouping-mask M_p to perform spatial grouping of features. We group features based on mask values, dividing them evenly into g_p groups (*e.g.*, $g_p = 4$ in Fig. 3 (b)). Each grouped feature has the same spatial dimensions, *i.e.*, the same number of elements. This grouping mechanism enables restoration by focusing on regions with similar characteristics, effectively handling spatial variability.

While attention within individual groups allows the model to capture local feature relationships, it is limited to learning only within those specific groups. To overcome this limitation, we extend the attention mechanism to capture cross-group relationships, enabling the model to learn broader feature dependencies and expand its receptive field.

Accordingly, we implement two types of group attention mechanisms, in-group and cross-group. In-group attention is applied within each feature group, while cross-group attention identifies the relationship between itself and another group. As shown in Fig. 3 (b), to perform both types of attention simultaneously, the grouped features are projected to generate two queries, two keys, and one value:

$$q_i^m, q_c^m, k_i^m, k_c^m, v^m = \text{Projection}(G_p^m), \tag{5}$$

where $G_p^m \in \mathbb{R}^{\frac{H\times W}{g_p}\times C}$ represents the feature of the m-th group at stage p, with $m \in \{1, 2, ..., g_p\}$. The subscript i indicates in-group attention, and c indicates a feature used in cross-group. The dimension of G_p^m is derived by spatially grouping the original feature map, which has dimensions $\mathbb{R}^{H\times W\times C}$, where C represents the channel dimension. Each projected feature has the same dimensions as G_p^m. Next, the in-group attention is calculated within the group as follows:

$$A_{\text{in}}^m = \text{In-Group}(q_i^m, k_i^m, v^m). \tag{6}$$

For cross-group attention, rather than performing comparisons across all groups, which would be a computationally heavy task, we aim to efficiently target a specific group. Therefore, to identify the most relevant group, we propose the group selector which computes the similarity between all v^m and selects the most relevant group index for group m while excluding itself, as follows:

$$a_m = \text{GroupSelector}(\{v^n\}_{n=1}^{g_p}), \quad s.t.\ \ n \neq m \tag{7}$$

and a_m denotes the selected group index (Details are described in the next section.) The attention is performed using the query from the selected group, $q_c^{a_m}$:

$$A_{\text{cross}}^m = \text{Cross-Group}(q_c^{a_m}, k_c^m, v^m). \tag{8}$$

After calculating A_{in}^m and A_{cross}^m for each group, they are aggregated to obtain the final attention outputs, A_{in} and A_{cross}. The output for FGA is combined with skip-connection as follows:

$$\hat{F}_p = F'_p + \text{Conv}_1(\alpha_1 \odot A_{\text{in}} + \alpha_2 \odot A_{\text{cross}}). \tag{9}$$

Here, α_1 and α_2 are learnable weights that balance the contributions of A_{in} and A_{cross}.

Group selector. To identify the most relevant group, each group generates a representative feature R_m by channel-wise pooling its values v^m, which aggregates the features across the channel dimension. The cosine similarity is then

Table 1. Quantitative comparisons on the All-weather dataset [28]. Methods with a * mark are adopted from All-in-One image restoration and trained with the All-weather dataset. The best values are marked in **bold**, while the second-best values are underlined.

	Rain			Snow			Raindrop			Average	
	Method	PSNR	SSIM	Method	PSNR	SSIM	Method	PSNR	SSIM	PSNR	SSIM
Single	CycleGAN [74]	17.62	0.6560	SPANet [51]	23.70	0.7930	pix2pix [21]	28.02	0.8547	-	-
	pix2pix [21]	19.09	0.7100	JSTASR [5]	25.32	0.8076	RaindropAttn [40]	31.44	0.9263	-	-
	HRGAN [27]	21.56	0.8550	DesnowNet [31]	27.17	0.8983	AttentiveGAN [37]	31.59	0.9170	-	-
	MPRNet [66]	28.03	0.9192	DDMSNet [71]	28.85	0.8772	MAXIM [48]	31.87	0.9352	-	-
	Fourmer [73]	26.58	0.8740	Fourmer [73]	27.38	0.8555	Fourmer [73]	25.79	0.8500	26.58	0.8508
All-in-One	Restormer* [67]	30.03	0.9215	Restormer* [67]	30.36	0.9068	Restormer* [67]	32.18	0.9408	30.86	0.923
	PromptIR* [36]	30.49	0.9263	PromptIR* [36]	30.91	0.9148	PromptIR* [36]	32.56	0.9426	31.32	0.928
	AdaIR* [13]	30.85	0.9286	AdaIR* [13]	31.01	0.9160	AdaIR* [13]	32.87	0.9432	31.58	0.929
	All-in-One [28]	24.71	0.8980	All-in-One [28]	28.33	0.8820	All-in-One [28]	31.12	0.9268	28.05	0.902
	TransWeather [49]	28.83	0.9000	TransWeather [49]	29.31	0.8879	TransWeather [49]	30.17	0.9157	29.44	0.901
	TUM [7]	29.27	0.9147	TUM [7]	30.22	0.9071	TUM [7]	31.81	0.9309	30.43	0.918
	WGWS [78]	29.32	0.9207	WGWS [78]	30.16	0.9007	WGWS [78]	32.38	0.9378	30.62	0.920
	AWRCP [63]	31.39	0.9329	AWRCP [63]	31.92	**0.9341**	AWRCP [63]	31.93	0.9314	31.75	0.933
	Histoformer [47]	<u>32.08</u>	<u>0.9389</u>	Histoformer [47]	<u>32.16</u>	0.9261	Histoformer [47]	<u>33.06</u>	<u>0.9441</u>	<u>32.43</u>	<u>0.936</u>
	Ours	**32.43**	**0.9410**	Ours	**32.22**	<u>0.9272</u>	Ours	**33.24**	**0.9489**	**32.63**	**0.939**

Table 2. Experiments on WeatherStream dataset. Trans. and Histo. denote TransWeather [49] and Histoformer [47], respectively.

	Rain		Fog		Snow		Avg.	
Model	PNSR	SSIM	PNSR	SSIM	PNSR	SSIM	PNSR	SSIM
Trans. [49]	25.72	0.7911	24.30	0.6979	23.66	0.7386	24.56	0.742
WGWS [78]	25.75	0.7928	23.88	0.6752	23.66	0.7383	24.43	0.735
Histo. [47]	<u>25.85</u>	<u>0.8215</u>	<u>24.36</u>	**0.7348**	<u>23.72</u>	<u>0.7631</u>	<u>24.64</u>	<u>0.773</u>
Ours	**26.01**	**0.8218**	**24.39**	<u>0.7335</u>	**23.91**	**0.7652**	**24.77**	**0.774**

computed between these representative features, and the most similar group is selected based on the highest similarity score. This allows the model to adaptively establish cross-group relationships, even if the groups were not initially grouped together. The process is illustrated step by step in Fig. 4. The insight is that interacting with the most similar group allows the model to achieve more robust restoration by effectively leveraging relationships beyond the initial grouping.

Attention configuration. Existing AiO models primarily employ channel attention, which performs attention mechanism along the channel dimension. While channel attention focuses on feature-wise relationships, it does not explicitly account for spatial dependencies. In adverse weather conditions, spatial attention is crucial as it directly captures relationships in space, making it more effective for restoration tasks where degraded or occluded areas need to be reconstructed using surrounding context. However, as image resolution increases, the computational cost of spatial attention grows quadratically. To balance the benefits of spatial attention and computational efficiency, half of the SGTB blocks at each stage use channel attention (SGTB-C), while the other half use spatial attention (SGTB-S). In each stage, SGTB-C is applied first and followed by SGTB-S.

3.5. Objective function

We train our model with the combination of $L1$ loss, $\mathcal{L}_{rec}$, and Pearson correlation loss [11, 47], $\mathcal{L}_{cor}$ to consider pixel-level and patch-level correlations together. The total objective function we used is as:

$$\mathcal{L} = \mathcal{L}_{rec} + \beta \mathcal{L}_{cor}, \tag{10}$$

where β is the weight of the correlation loss. More about objective function are described in the supplementary.

4. Experiments

4.1. Experimental settings

Datasets. We use two dataset to evaluate our model. One is a synthetic dataset, All-weather dataset [28], which contains Outdoor-rain [27], Snow100K-L [31], and Raindrop [37] dataset as previous researches [26, 47, 49] use. The other is a real dataset, WeatherStream [69], which contains the scenes of rain, snow, and fog. Each sub-dataset has A, B, and C degraded images. For fair comparisons, we follow the dataset utilization protocol of previous works [28, 47, 78].

Baseline methods. To contextualize our results, we compare with single-task methods for raindrop removal (pix2pix [21], RaindropAttn [40], AttentiveGAN [37], MAXIM [48], Fourmer [73]), snow removal (SPANet [51], JSTASR [5], DesnowNet [31], DDMSNet [71], Fourmer [73]), and rain removal (CycleGAN [74], pix2pix [21], HRGAN [27], MPRNet [66], Fourmer [73]). We also evaluate against AiO restoration models (Restormer [67], PromptIR [36], AdaIR [13]) and AiO adverse weather removal models (All-in-One [28], TransWeather [49], TUM [7], WGWS [78], AWRCP[63], Histoformer [47]) that use only images as input without any external knowledge, such as language models, trained with previous strategies [28, 47] for All-weather dataset [28]. Experiments on WeatherStream dataset [69], we compare our methods with TransWeather, WGWS, and Histoformer, which are recent AiO models.

4.2. Comparisons with the state-of-the-arts

Quantitative evaluation. In Tab. 1, we quantitatively compare our method with other single weather removal mod-

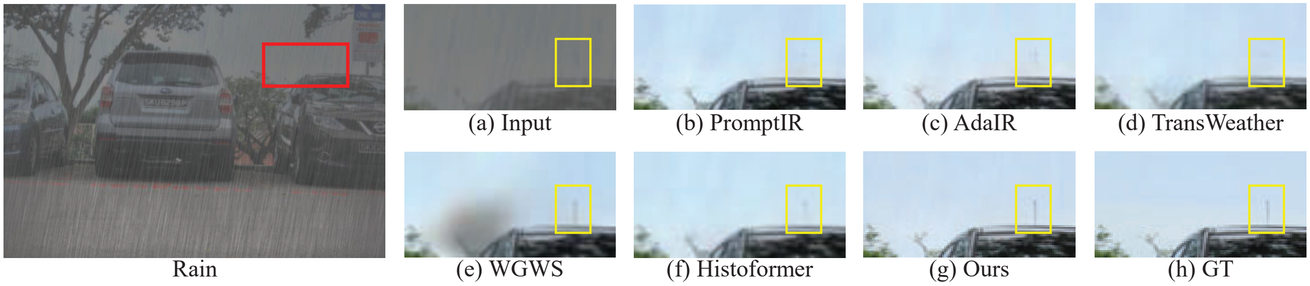


Figure 5. Qualitative results of deraining on Outdoor-rain [27] dataset.

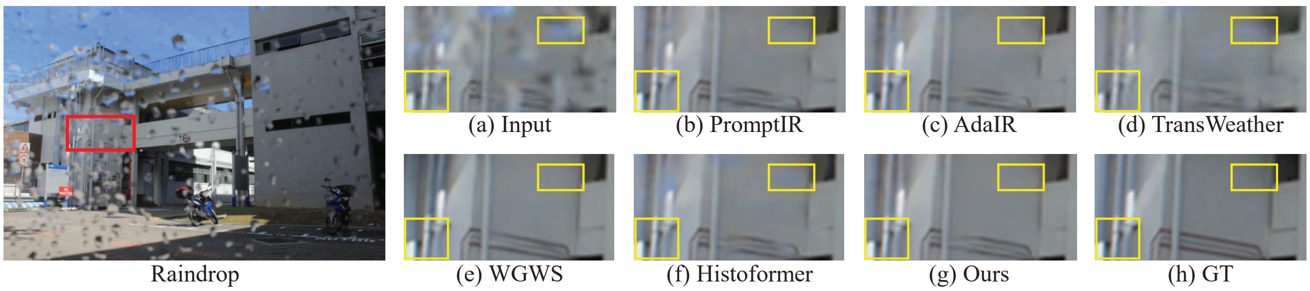


Figure 6. Qualitative results of raindrop removal on RainDrop [37] dataset.

els, general AiO models, and AiO adverse weather removal models on the All-Weather dataset. Our method achieves the best performance on rain and raindrop, outperforming the second-best method by +0.35dB and +0.18dB in PSNR, respectively, and the highest PSNR with the second-best SSIM in snow. On average, it demonstrates the best overall performance on the All-Weather dataset.

Moreover, Tab. 2 presents the comparison on the WeatherStream dataset [69]. Our SSGformer consistently achieved top performance across most cases, ranking second only in SSIM for fog degradation. Our method performs the best on average in the real-world dataset.

Qualitative evaluation. We compare visualizations with recent methods across three tasks in Fig. 5, 6, and 7. Our method restores the original objects more sharply than others and achieves the closest color condition to the clean image. In particular, as shown in Fig. 5, the yellow box highlights how our method accurately restores fine objects that are obscured by rain degradation, preventing misclassification as rain artifacts. We can also see that our method fills in the empty areas left after removing degradations rather than others. The visualization for WeatherStream is provided in the supplementary.

Real-world scenario. Additionally, we conduct a visual comparison to evaluate the model trained on the All-Weather dataset under real-world conditions. The results in Fig. 8 show that our method successfully removes real-world degradations better than others.

4.3. Ablation studies and discussions

To reveal the relationship between our proposed methods and the degradation removal performance of each weather type, we demonstrate ablation studies with the evaluation in all weather, not specific weather type.

Table 3. Ablation study for the proposed Spectral-based Decompotision Prompt (SDP).

		Rain		Snow		Raindrop		Avg.	
Sobel	SVD	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
✓		32.32	0.9401	32.13	0.9259	**33.30**	0.9488	32.58	0.9383
	✓	32.19	0.9400	32.08	0.9257	33.07	0.9486	32.45	0.9381
✓	✓	**32.43**	**0.9410**	**32.22**	**0.9272**	33.24	**0.9489**	**32.63**	**0.9390**

Table 4. SGTB Attention configuration. Ch. and Spa. denote channel-wise attention and spatial-wise attention, respectively.

		Rain		Snow		Raindrop		Avg.	
Ch.	Spa.	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
✓		32.16	0.9399	31.94	0.9239	**33.26**	0.9485	32.45	0.9374
✓	✓	**32.43**	**0.9410**	**32.22**	**0.9272**	33.24	**0.9489**	**32.63**	**0.9390**

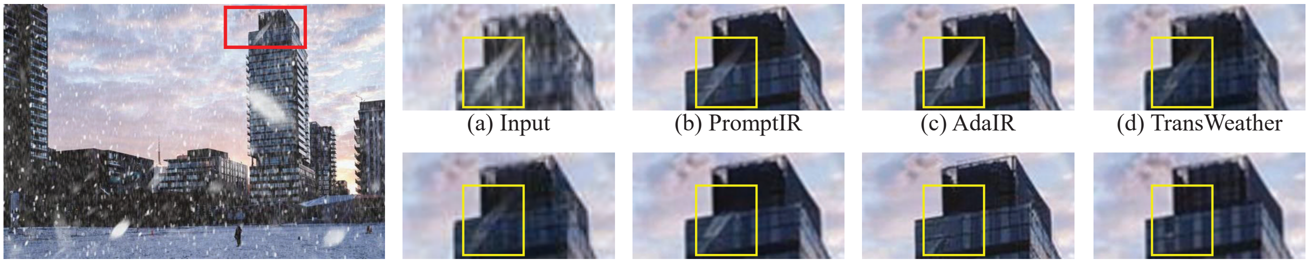
Table 5. Group attention configuration. In and Cross denote ingroup attention and cross-group attention, respectively.

		Rain		Snow		Raindrop		Avg.	
In	Cross	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
✓		32.28	0.9403	32.00	0.9246	33.10	0.9485	32.46	0.9378
✓	✓	**32.43**	**0.9410**	**32.22**	**0.9272**	**33.24**	**0.9489**	**32.63**	**0.9390**

Table 6. Ablation Study on Prompt and Transformer Block Choices. Feature-Grouped Attention (FGA) and Spectral-based Decomposition Prompt (SDP) are added progressively.

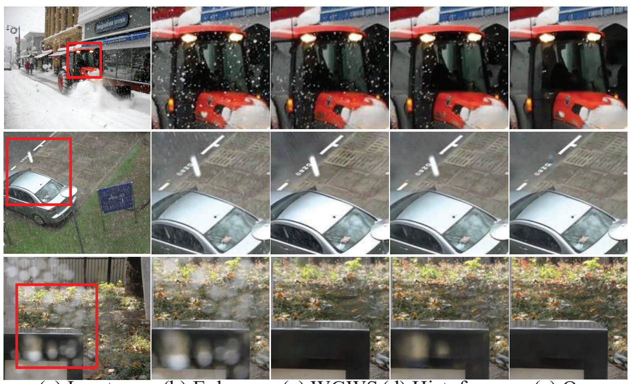
		Rain		Snow		Raindrop		Avg.	
FGA	SDP	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
		31.93	0.9386	31.92	0.9238	33.10	0.9485	32.32	0.9369
	✓	32.27	0.9399	32.01	0.9249	33.19	0.9486	32.49	0.9378
✓		32.18	0.9400	32.07	0.9256	33.14	0.9486	32.46	0.9380
✓	✓	**32.43**	**0.9410**	**32.22**	**0.9272**	**33.24**	**0.9489**	**32.63**	**0.9390**

Design of prompts. To demonstrate the effectiveness of our proposed SDP, we conduct independent experiments about the components in the SDP. As shown in Tab. 3, using the Sobel operator and SVD together shows the best performance in the degradation type of rain and snow. In raindrop, the SSIM is the highest, and the PSNR is confirmed



Snow (a) Input (b) PromptIR (c) AdaIR (d) TransWeather (e) WGWS (f) Histoformer (g) Ours (h) GT

Figure 7. Qualitative results of desnowing Snow100K-L [31] dataset.



(a) Input (b) Enlarge (c) WGWS (d) Histoformer (e) Ours

Figure 8. Real-world degradation removal examples. From the first to last rows, the adverse weather type is snow (RealSnow [78]), rain (NTURain [3]), and, raindrop (RainDS [39]), respectively. Zoom for better view.

to be the second best. This experiment demonstrates that the combination of the two decomposition filters can efficiently extract information from the image.

Spatial Grouping Transformer Block configuration. To demonstrate the effectiveness of the proposed SGTB, we conduct experiments by varying its configuration. In Tab. 4, the proposed configuration is compared with two variants. The first uses SGTB-C for all blocks per stage, where each stage p consists of L_p blocks. The second adopts a hybrid setup, in which half of the blocks per stage are replaced with SGTB-S (*i.e.*, $L_p/2$ SGTB-C and $L_p/2$ SGTB-S). It is observed that the mixed configuration with SGTB-C and SGTB-S performs better on all metrics. Notably, this benefit comes without any increase in model capacity.

Group attention configuration. We analyzed about how the formation of attention mechanism is executed within each transformer block in Tab. 5. The results showed that incorporating cross-group attention, which enables information exchange between groups, yielded better performance than using in-group attention.

The synergy between FGA and SDP. We conducted an ablation study to evaluate the effects of FGA and SDP, as shown in Tab. 6. Four models were tested for comparison: the baseline model using the standard transformer block, the model using SDP as a prompt, the model using FGA as a transformer block, and the model incorporating both FGA and SDP. The results demonstrate that the inclusion of SDP significantly improves performance, with the model using SDP outperforming the baseline. This highlights the effectiveness of SDP in extracting degradation information and enhancing restoration. Further improvements are observed with FGA, which enables the model to focus on relevant areas of the scene. For FGA, a simple visual prompt is used to generate a mask, which serves as input. Unlike the baseline model, which lacks a grouping mechanism, this model leverages the mask to group similar areas, leading to better performance, particularly in complex degradation scenarios. Finally, we conducted an experiment combining both FGA and SDP to assess their synergy. As shown in the last column of Tab. 6, SDP improves restoration by handling degradations, while FGA enhances this process by adding an effective grouping mechanism. The combination of both results in the best overall performance.

5. Conclusion

We propose SSGformer, a novel restoration architecture that incorporates Spatial Grouping Transformer Block with Spectral-based Decomposition Prompt. By leveraging the Sobel operator to extract high-frequency edge information and Singular Value Decomposition for low-frequency components, our method effectively captures valuable spectral features from degraded images. With the manipulated features, we generate the grouping-mask. This mask guides the group-wise attention mechanism across both the channel and spatial domains, enabling the model to comprehensively learn degradation-related information across diverse frequency ranges. Through comprehensive experiments, we showcase the effectiveness and advantages of our approach.

Limitations and future works. Our method leverages information from classical filters, which makes it effective in scenarios where classical filters are applicable. This design reflects an internal strategy for AiO, relying on intrinsic visual structures to guide restoration. However, in scenes with high background complexity—especially when weather-specific cues are absent—such internal cues may be insufficient. To overcome these limitations, we plan to explore external strategies that incorporate knowledge from large-scale pretrained models, such as LLM/VLMs [56].

Acknowledgment

This research was supported by the Challengeable Future Defense Technology Research and Development Program through the Agency For Defense Development(ADD) funded by the Defense Acquisition Program Administration(DAPA) in 2025(No.915102201)

References

[1] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE transactions on image processing*, 25(11):5187–5198, 2016. 2

[2] Haoyu Chen, Jingjing Ren, Jinjin Gu, Hongtao Wu, Xuequan Lu, Haoming Cai, and Lei Zhu. Snow removal in video: A new dataset and a novel method. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13165–13176. IEEE, 2023. 2

[3] Jie Chen, Cheen-Hau Tan, Junhui Hou, Lap-Pui Chau, and He Li. Robust video content alignment and compensation for rain removal in a cnn framework. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6286–6295, 2018. 8

[4] Sixiang Chen, Tian Ye, Yun Liu, Taodong Liao, Jingxia Jiang, Erkang Chen, and Peng Chen. Msp-former: Multi-scale projection transformer for single image desnowing. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023. 1, 2

[5] Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, Cheng-Che Tsai, and Sy-Yen Kuo. Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*, pages 754–770. Springer, 2020. 1, 2, 6

[6] 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. 2

[7] 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, 3, 6

[8] 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, 2

[9] Xiang Chen, Jinshan Pan, Jiyang Lu, Zhentao Fan, and Hao Li. Hybrid cnn-transformer feature fusion for single image deraining. In *Proceedings of the AAAI conference on artificial intelligence*, pages 378–386, 2023. 2

[10] Yi-Lei Chen and Chiou-Ting Hsu. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *Proceedings of the IEEE international conference on computer vision*, pages 1968–1975, 2013. 2

[11] Israel Cohen, Yiteng Huang, Jingdong Chen, Jacob Benesty, Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. *Noise reduction in speech processing*, pages 1–4, 2009. 6

[12] 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. 1

[13] Yuning Cui, Syed Waqas Zamir, Salman Khan, Alois Knoll, Mubarak Shah, and Fahad Shahbaz Khan. Adair: Adaptive all-in-one image restoration via frequency mining and modulation. *arXiv preprint arXiv:2403.14614*, 2024. 2, 4, 6

[14] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 2

[15] David Eigen, Dilip Krishnan, and Rob Fergus. Restoring an image taken through a window covered with dirt or rain. In *Proceedings of the IEEE international conference on computer vision*, pages 633–640, 2013. 2

[16] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3855–3863, 2017. 2

[17] 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*, 45(8):9534–9551, 2023. 2

[18] Tao Gao, Yuanbo Wen, Kaihao Zhang, Jing Zhang, Ting Chen, Lidong Liu, and Wenhan Luo. 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, 2023. 1

[19] Yina Guo, Jianguo Chen, Xiaowen Ren, Anhong Wang, and Wenwu Wang. Joint raindrop and haze removal from a single image. *IEEE Transactions on Image Processing*, 29:9508–9519, 2020. 2

[20] Himanshu Gupta, Oleksandr Kotlyar, Henrik Andreasson, and Achim J Lilienthal. Robust object detection in challenging weather conditions. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7523–7532, 2024. 1

[21] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 6

[22] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE transactions on image processing*, 21(4):1742–1755, 2011. 1, 2

[23] Angelos Katharopoulos, Apoorv Vyas, Nikolaos Pappas, and François Fleuret. Transformers are rnns: Fast autoregressive transformers with linear attention. In *International conference on machine learning*, pages 5156–5165. PMLR, 2020. 4

[24] Younkwan Lee, Yechan Kim, Jongmin Yu, and Moongu Jeon. Learning to remove bad weather: towards robust visual perception for self-driving. *IEEE Robotics and Automation Letters*, 2022. 1

[25] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision*, pages 4770–4778, 2017. 2

[26] 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, 3, 6

[27] 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. 1, 2, 6, 7

[28] 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. 2, 6

[29] Zilong Li, Yiming Lei, Chenglong Ma, Junping Zhang, and Hongming Shan. Prompt-in-prompt learning for universal image restoration. *arXiv preprint arXiv:2312.05038*, 2023. 4

[30] Jingbo Lin, Zhilu Zhang, Yuxiang Wei, Dongwei Ren, Dongsheng Jiang, Qi Tian, and Wangmeng Zuo. Improving image restoration through removing degradations in textual representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2866–2878, 2024. 1

[31] 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, 2, 6, 8

[32] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for universal image restoration. *arXiv preprint arXiv:2310.01018*, 3(8), 2023. 3

[33] 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*, 45(8): 10346–10357, 2023. 1

[34] 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.

[35] 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.

[36] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36, 2024. 1, 6

[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, 2, 6, 7

[38] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8160–8168, 2019. 1, 2

[39] Ruijie Quan, Xin Yu, Yuanzhi Liang, and Yi Yang. Removing raindrops and rain streaks in one go. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9147–9156, 2021. 8

[40] Yuhui Quan, Shijie Deng, Yixin Chen, and Hui Ji. Deep learning for seeing through window with raindrops. In *Proceedings of the IEEE/CVF International*

Conference on Computer Vision, pages 2463–2471, 2019. 1, 2, 6

[41] Yuhui Quan, Xiaoheng Tan, Yan Huang, Yong Xu, and Hui Ji. Image desnowing via deep invertible separation. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(7):3133–3144, 2023. 2

[42] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 154–169. Springer, 2016. 2

[43] Yoav Y Schechner, Srinivasa G Narasimhan, and Shree K Nayar. Instant dehazing of images using polarization. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, pages I–I. IEEE, 2001. 1

[44] Ming-Wen Shao, Le Li, De-Yu Meng, and Wang-Meng Zuo. Uncertainty guided multi-scale attention network for raindrop removal from a single image. *IEEE Transactions on Image Processing*, 30:4828–4839, 2021. 2

[45] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 4

[46] 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, 2

[47] Shangquan Sun, Wenqi Ren, Xinwei Gao, Rui Wang, and Xiaochun Cao. Restoring images in adverse weather conditions via histogram transformer. In *European Conference on Computer Vision*, pages 111–129. Springer, 2025. 3, 4, 6

[48] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5769–5780, 2022. 6

[49] 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, 3, 6

[50] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017. 2

[51] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson WH Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12270–12279, 2019. 6

[52] 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

[53] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10551–10560, 2021. 2

[54] Jie Xiao, Xueyang Fu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Image de-raining transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11):12978–12995, 2022. 2

[55] Xu Yan and Yuan Ren Loke. Raingan: Unsupervised raindrop removal via decomposition and composition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 14–23, 2022. 2

[56] Hao Yang, Liyuan Pan, Yan Yang, and Wei Liang. Language-driven all-in-one adverse weather removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24902–24912, 2024. 3, 8

[57] Tao Yang, Xiaofei Chang, Hang Su, Nathan Crombez, Yassine Ruichek, Tomas Krajnik, and Zhi Yan. Raindrop removal with light field image using image inpainting. *IEEE Access*, 8:58416–58426, 2020. 2

[58] Xin Yang, Wending Yan, Yuan Yuan, Michael Bi Mi, and Robby T Tan. Semantic segmentation in multiple adverse weather conditions with domain knowledge retention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 6558–6566, 2024. 1

[59] Youzhao Yang and Hong Lu. Single image deraining using a recurrent multi-scale aggregation and enhancement network. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1378–1383. IEEE, 2019. 1, 2

[60] Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, and Dacheng Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2037–2046, 2022. 1, 2

[61] Yijun Yang, Hongtao Wu, Angelica I Aviles-Rivero, Yulun Zhang, Jing Qin, and Lei Zhu. Genuine knowl-

edge from practice: Diffusion test-time adaptation for video adverse weather removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25606–25616, 2024. 1

[62] Rajeev Yasarla and Vishal M Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8405–8414, 2019. 1, 2

[63] 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, 3, 6

[64] Chia-Hung Yeh, Chih-Yang Lin, Kahlil Muchtar, and Pin-Hsian Liu. Rain streak removal based on non-negative matrix factorization. *Multimedia Tools and Applications*, 77:20001–20020, 2018. 1

[65] Shaodi You, Robby T Tan, Rei Kawakami, Yasuhiro Mukaigawa, and Katsushi Ikeuchi. Adherent raindrop modeling, detection and removal in video. *IEEE transactions on pattern analysis and machine intelligence*, 38(9):1721–1733, 2015. 2

[66] 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. 6

[67] 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. 6

[68] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *IEEE transactions on circuits and systems for video technology*, 30(11):3943–3956, 2019. 1, 2

[69] 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. 6, 7

[70] 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. 1

[71] Kaihao Zhang, Rongqing Li, Yanjiang Yu, Wenhan Luo, and Changsheng Li. Deep dense multi-scale network for snow removal using semantic and depth priors. *IEEE Transactions on Image Processing*, 30: 7419–7431, 2021. 2, 6

[72] Rongyu Zhang, Yulin Luo, Jiaming Liu, Huanrui Yang, Zhen Dong, Denis Gudovskiy, Tomoyuki Okuno, Yohei Nakata, Kurt Keutzer, Yuan Du, et al. Efficient deweahter mixture-of-experts with uncertainty-aware feature-wise linear modulation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 16812–16820, 2024. 1

[73] Man Zhou, Jie Huang, Chun-Le Guo, and Chongyi Li. Fourmer: An efficient global modeling paradigm for image restoration. In *International conference on machine learning*, pages 42589–42601. PMLR, 2023. 2, 6

[74] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 6

[75] Lei Zhu, Chi-Wing Fu, Dani Lischinski, and Pheng-Ann Heng. Joint bi-layer optimization for single-image rain streak removal. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. 1

[76] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE transactions on image processing*, 24(11):3522–3533, 2015. 1

[77] Ruoxi Zhu, Zhengzhong Tu, Jiaming Liu, Alan C Bovik, and Yibo Fan. Mwformer: Multi-weather image restoration using degradation-aware transformers. *IEEE Transactions on Image Processing*, 2024. 1

[78] 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, 3, 6, 8