Adverse Weather Image Restoration with Multi-domain Collaborative Guidance

Kai Xiong Siyu Wu Xiyu Pan Jianjun Li ✉
Central South University of Forestry and Technology
Changsha, China
xkbbk123@163.com, wusiyu19980913@163.com, 20231100392@csuft.edu.cn, T20010539@csuft.edu.cn

Abstract

Images captured under adverse weather often suffer severe distortions, thereby degrading the performance of downstream vision tasks. Existing methods rely on limited weather descriptors to guide multi-weather restoration, overlooking the intrinsic properties of degradation and yielding subpar results. Motivated by the observation that different weather degradations present distinguishable frequency characteristics, we are inspired to combine frequency and spatial cues to capture intrinsic degradation attributes. Building on this, we propose MCGNet, a multi-domain collaborative guidance based image restoration framework that enables the model to dynamically determine a more appropriate restoration strategy from intrinsic degradation cues in the input. Specifically, we design a collaboration-driven shared domain that cross-fuses multi-granularity frequency features with spatial features to form cross-domain shared guidance, which in turn dynamically refine the outputs of the frequency and spatial domains to adjust restoration strategies. Furthermore, we introduce multi-scale learning of shared representations to ensure consistent restoration of global structures and local details. Extensive experiments on 15 datasets demonstrate that MCGNet outperforms state-of-the-art methods and enhances the visual quality of real-world images captured in adverse weather.

***Keywords:** Image restoration, Adverse weather, Frequency feature, Multi-domain learning*

1. Introduction

Adverse weather, such as rain, haze, and snow, significantly degrades image quality. Removing these weather artifacts is a critical preprocessing step for numerous practical applications, notably autonomous driving [39]. Most existing methods are tailored for specific tasks, such as dehazing [17, 34], deraining [4, 51], or desnowing [2, 11]. However, adverse weather often exhibits multiple interacting degradations, causing these methods to perform poorly in such scenes [19], which motivates unified models for restoration under composite multi-weather conditions.

Several approaches have been proposed for multi-weather image restoration. A straightforward strategy involves training separate branches for each weather [25], yet this approach results in linear growth of model complexity with increasing degradation types. Another strategy uses a two-stage restoration process [50]: the first stage extracts general features with frozen parameters, followed by the second stage refining weather-specific features. This approach complicates the training process and heavily relies on weather labels (one-hot codes such as [1,0,0] for snow, [0,1,0] for rain, [0,0,1] for haze), leading to performance degradation due to misclassification under composite degradations. To simplify training and unify the representation of different degradations, Transformer-based methods have been proposed [19, 22], utilizing cross-attention to introduce weather descriptors to guide the model. Nevertheless, these approaches introduce expensive computations, limiting their adaptability to large-scale degradations, and similarly depend on explicit weather descriptors. We list the challenges in current methods as follows.

Figure 1. Weather descriptors fail to capture fine-grained variations in degradation; even with the same description, weather degradations can differ greatly in morphology and intensity.

- **Limited weather descriptors**: Weather descriptors are ineffective in capturing the intrinsic characteristics of weather. For example, as shown in Fig 1, three images described as "rain" can differ substantially in morphology and intensity. In multi-weather scenarios where degradation patterns are more complex and variable, they are even less representative and inadequate for driving the optimal restoration strategy.
- **Limited dynamic strategy adjustment**: The lack of dynamic adjustment of restoration strategies reduces robustness under multi-weather scenarios. Previous meth-

ods [19, 50] lack flow and feedback of restoration cues, leaving inference to static choices within a restricted strategy space, which in turn suppresses overall restoration quality and generalization ability.

We focus on dynamically inferring the optimal restoration strategy from the intrinsic characteristics of degradation patterns, addressing the aforementioned challenges.

Our proposed. To tackle these limitations, we propose MCGNet, an efficient framework for multi-weather restoration. MCGNet addresses the above challenges by making the following contributions.

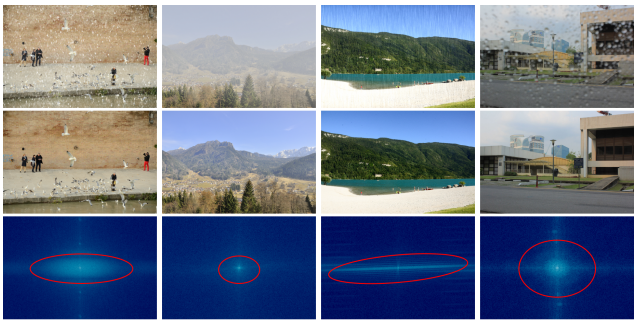

Figure 2. From top to bottom: weather-degraded images, the corresponding ground-truth images, and the frequency spectra obtained by applying a fast Fourier transform to their residuals.

Frequency features. We incorporate frequency domain knowledge into MCGNet and design a multi-granularity Frequency-Aware Prompter (FAP) to enhance the perception of intrinsic weather characteristics. Fig. 2 shows that frequency features can capture the properties of weather—rain streaks and snow grains elevate high-frequency energy, whereas haze induces global low-frequency attenuation [15]. FAP captures intrinsic characteristics from the frequency domain that weather descriptors cannot provide.

Multi-domain collaborative guidance. To better fuse multi-domain knowledge into reliable restoration guidance and enable dynamic strategy adjustment, we design a collaboration-driven shared domain together with FAP, forming the Multi-domain Collaborative Guidance Block (MCGB). Through a two-stage cross-fusion, the shared domain bridges the semantic gap between frequency and spatial domains and produces unified cross-domain shared representations, which align and refine multi-domain knowledge. Moreover, the shared domain modulates the frequency and spatial outputs, enabling MCGNet to adaptively select an appropriate domain combination for the current degradation and thus infer the restoration strategy. This dynamic inference balances domain contributions instead of relying on a fixed domain preference, accommodating continuous variations in multi-weather degradation.

Enhanced feature extraction and fusion. We design two modules for enhancing feature extraction and fusion: (1) Dual-path refinement module (DRM). DRM enhances domain-specific features along both the channel and spatial dimensions, yielding more discriminative shared representations. (2) Multi-scale Progressive Fusion Block (MPFB). MPFB aggregates multi-scale shared representations from coarse to fine, achieving consistency between global structures and local details.

Evaluations. We evaluate MCGNet on 15 datasets spanning five weather restoration tasks. The experimental results show that MCGNet achieves state-of-the-art performance on each task with fewer parameters. In composite multi-weather benchmarks *CDD11* [19], MCGNet outperforms the two descriptor-guided baselines [19, 50] by 0.35dB and 1.86dB PSNR. Its cross-domain shared guidance encodes richer degradation features than weather descriptors, enabling better separation of degradations from images. Further evaluations on real-world benchmarks yield competitive quantitative and visual results, demonstrating the practical effectiveness of our multi-domain collaborative strategy for multi-weather restoration. We find that the collaboration-driven shared domain effectively models intrinsic degradation patterns; visualizations of final decoder outputs (see Fig 12) show that MCGNet effectively captures residual frequency patterns. Overall, MCGNet effectively enhances degradation modeling by leveraging multi-domain knowledge and dynamic strategy adjustments, achieving robustness in multi-weather scenarios and providing practical utility in real-world settings.

2. Related Work

2.1. Specific Weather Restoration

Image restoration under adverse weather has been extensively studied [2, 4, 11, 17, 34, 51]. Early methods were based on physical priors [20]. Later, deep learning approaches achieved better performance by end-to-end dehazing [31], overcoming reliance on specific priors. Recent studies have proposed various advanced architectures and strategies, such as hybrid architectures combining CNN and Transformer [8]. Rainy degradation is mainly characterized by rain streaks and raindrops [30]. Recent studies have turned to attention [51] and Transformer-based methods [4]. The latter advances deraining by leveraging global modeling capability. For raindrop removal, techniques such as branched architectures [51] have improved the performance of raindrop removal. Snowy images often suffer from snow particles and severe blur [28]. Liu *et al.* [28] combine chromatic-aberration estimation and residual compensation to remove the veiling snow layer, while the diffusion-based method [52] reconstructs clear details by progressively inverting the degradation of snow particles.

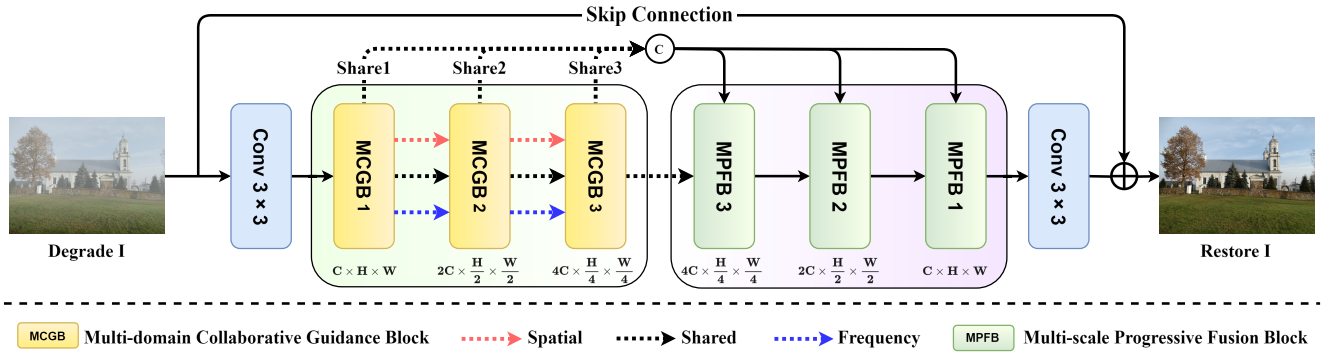

Figure 3. The architecture of MCGNet. The **red** dashed lines represent spatial feature flow, the **black** dashed lines denote shared feature flow, and the **blue** dashed lines depict frequency feature flow.

Although effective in addressing specific weather conditions, these approaches struggle to restore images suffering from multi-weather effects [50].

2.2. Multi-Weather Restoration

Recent works have explored unified networks for image restoration under multiple weather conditions. Li *et al.* [25] use dedicated encoders for each degradation, leading to increased model complexity. Jose Valanarasu *et al.* [22] utilize a Transformer-based encoder-decoder architecture with learnable weather-query embeddings to address multiple degradations, while Chen *et al.* [3] integrate knowledge from multiple task-specific teacher models via two-stage knowledge distillation and multi-contrastive regularization. Özdenizci *et al.* [52] model restoration as a conditional diffusion process. Potlapalli *et al.* [29] introduce learnable prompt vectors for various types of degradation. Guo *et al.* [19] use cross-attention with external scene descriptors to support user-guided restoration. Despite these advances, most methods rely on weather descriptors, which fail to capture intrinsic degradation features and hinder adaptation to complex multi-weather scenarios. Therefore, we adopt cross-domain shared representations that fuse multi-granularity frequency and spatial features as guidance signals for restoration. Compared with weather descriptors, they encode richer degradation information and provide more comprehensive support.

2.3. Multi-domain Learning

In the spectrum, rain streaks and snow particles manifest as dense local occlusions, increasing high-frequency energy, whereas haze suppresses low-frequency components [12]. This frequency distinction serves as a natural prior for identifying degradations [10, 34]. Recent works propose merging the spatial and frequency domains for image restoration. Zhou *et al.* [49] injects frequency priors into the decoder via filtering and a dual-branch prompting module, which modulates them to suppress frequency-domain artifacts. Existing methods face several limitations: (1) they lack fine-grained frequency modulation and dynamic adaptation [10, 49], hindering their ability to differentiate diverse degradation patterns; (2) they treat spatial and frequency features independently, leading to limited cross-domain interaction and sharing [9, 34].

Our work addresses these limitations by introducing a multi-granular frequency modulation to adaptively extract essential features for different degradations, while enabling deep cross-domain interaction among spatial, frequency, and shared features within a unified network.

3. Methodology

3.1. Motivation

Multi-weather is often composed of coupled degradations at multiple scales. Existing methods [19, 50] typically use weather descriptors to guide restoration. However, since these descriptors overlook the fine-grained characteristics, models cannot dynamically adjust strategies from the degradations themselves and performance is limited. To address this, we propose MCGNet, which consists of three components as shown in Fig. 4-6: (1) Multi-domain Collaborative Guidance Block (MCGB), (2) Dual-path Refinement Module (DRM), and (3) Multi-scale Progressive Fusion Block (MPFB). Among them, MCGB serves as the core by aggregating multi-granularity frequency features into a unified shared representation. They serve as guidance maps to modulate domain outputs, enabling dynamical strategy inference from degradations themselves. DRM amplifies degradation features by enhancing attention to channel and spatial dimensions, thereby reducing feature redundancy during cross-domain fusion. MPFB strengthens multi-scale adaptability by progressively fusing shared representations from coarse to fine, ensuring consistency between global structures and local details. Thus, incorporating three components, MCGNet shows phenomenal restoration performance under multiple weather conditions.

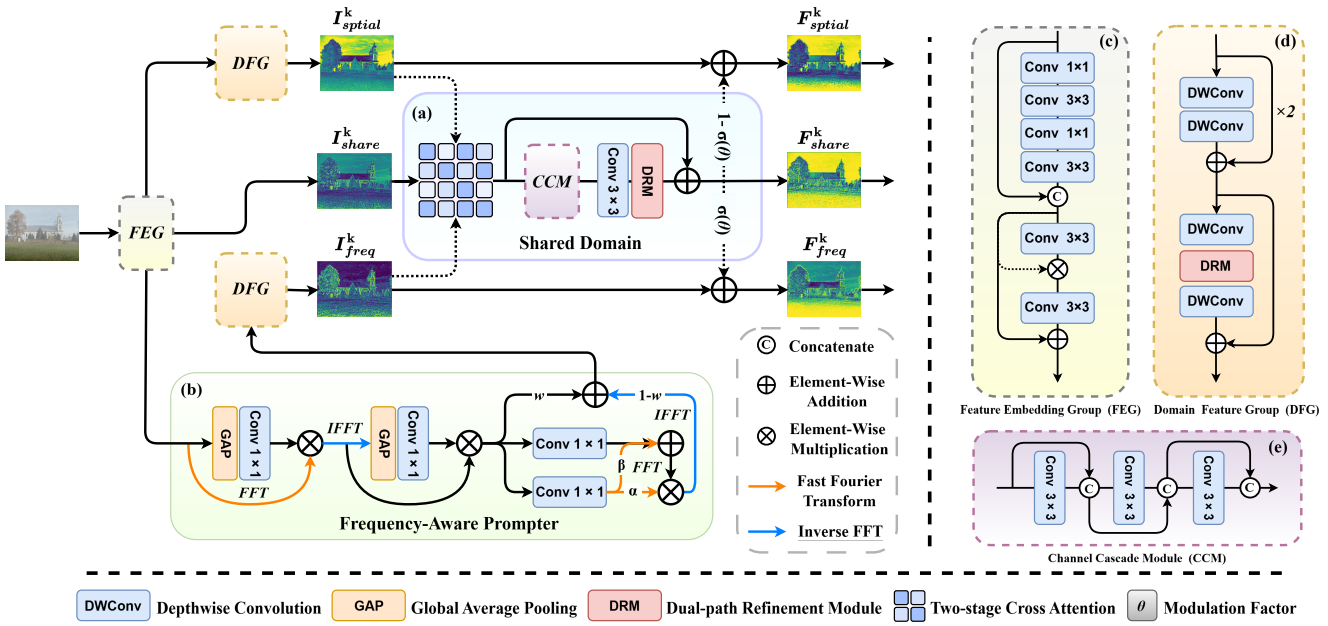

Figure 4. Overview of MCGB. (a) Shared Domain; (b) Frequency-Aware Prompter; (c) Feature Embedding Group; (d) Domain Feature Group; (e) Channel Cascade Module.

3.2. Overall Architecture

As shown in Fig. 3, given a degraded image, MCGNet applies a 3×3 convolution to extract shallow features $F_s \in \mathbb{R}^{C\times H\times W}$, where C, H, and W denote the channel, height, width, respectively. F_s are then processed by three MCGBs, which decompose them into three domain feature flow $\{F^k_{spatial}, F^k_{freq}, F^k_{share}, k = 1, 2, 3\} \in \{\mathbb{R}^{nC\times \frac{H}{n}\times \frac{W}{n}}, n = 1, 2, 4\}$. As illustrated in Fig. 4, MCGB is centered on the shared domain and additionally includes the Frequency-Aware Prompter (FAP) and multiple DRM-embedded residual blocks. During this process, the spatial resolution is halved while the channel width is doubled. The multi-scale F^k_{share} are subsequently concatenated and fed into the MPFBs. As shown in Fig. 6, the MPFB uses the DRM to assign fusion weights and progressively reconstruct a high-resolution restored image. For simplicity, we omit multi-scale inputs and outputs, and only display the top-level skip connection in Fig. 3. Then, we elaborate on each module in the following sections.

3.3. Multi-domain Collaborative Guidance Block (MCGB)

The structure of MCGB is shown in Fig. 4. The objective is to overcome the limitations of weather descriptors in modeling degradations and to equip the model with a dynamic strategy for adjustment. To this end, MCGB comprises two key components (Fig. 4(a)–(b)): the shared domain, which fuses frequency and spatial features into a unified representation and, via a learnable modulation factor, produces guidance maps that guide the selection of the most suitable domain combination for degradations; the Frequency-Aware Prompter (FAP), which performs coarse-to-fine modulation to provide fine-grained frequency cues for the shared domain. Section 3.3.1 details the FAP.

Specifically, given the input features $F_s \in \mathbb{R}^{C\times H\times W}$, after being processed by a series of Feature Embedding Groups (Fig. 4(c)) and Domain Feature Groups (Fig. 4(d)), the initial domain features $I^k_{spatial}$, I^k_{share}, and I^k_{freq} are obtained. Notably, the frequency branch first passes through the FAP to extract multi-granularity frequency features before entering Domain Feature Group.

Subsequently, the three initial domain features are fused in the shared domain as described in Algorithm 1. After projecting them into a common interaction space, the shared domain undergoes a two-stage cross-domain information exchange via cross-attention [43]. The first stage focuses on spatial-frequency interaction to establish correspondence between structural and spectral cues, producing intermediate representations F_{s1} where spatial and frequency information is aligned. The second stage integrates F_{s1} with the shared stream to yield the three-domain aligned representations F_{s2}. Unlike direct merging, this staged fusion reduces cross-domain mismatch by first establishing clean spatial-frequency correspondence and then using the shared stream to calibrate and consolidate the aligned cues. We then refine F_{s2} using the Channel Cascade Module (CCM, Fig. 4(e)) and the Dual-path Refinement Module (DRM) to enhance degradation-relevant responses and suppress redundancy. This results in the final shared guidance features:

Algorithm 1: The Shared Domain

Input: Initial features
$\{I^k_{spatial}, I^k_{share}, I^k_{freq} \in \mathbb{R}^{C_k \times H_k \times W_k}\}$

Output: Refined features
$\{F^k_{spatial}, F^k_{share}, F^k_{freq} \in \mathbb{R}^{C_k \times H_k \times W_k}\}$

// ***Step 1: Feature Projection***

$I^k_{spatial} \leftarrow (Q_{spatial}, K_{spatial}, V_{spatial})$;

$I^k_{share} \leftarrow (Q_{share}, K_{share}, V_{share})$;

$I^k_{freq} \leftarrow (Q_{freq}, K_{freq}, V_{freq})$;

// ***Step 2: Cross-domain Fusion***

$F_{freq \to spatial} = Softmax(\frac{Q_{freq} K^T_{spatial}}{\sqrt{d}}) V_{spatial}$;

$F_{spatial \to freq} = Softmax(\frac{Q_{spatial} K^T_{freq}}{\sqrt{d}}) V_{freq}$;

$F_{s1} = \mathcal{C}(\mathrm{Conv}_{1\times1}(F_{freq \to spatial}),\ \mathrm{Conv}_{1\times1}(F_{spatial \to freq}))$;

$F_{s2} = \mathcal{C}(\mathrm{Conv}_{1\times1}(F_{s1 \to share}),\ \mathrm{Conv}_{1\times1}(F_{share \to s1}))$;

// ***Step 3: Feature Refinement***

$F^k_{share} = F_{s2} + \mathcal{CCM}\big(\mathrm{Conv}_{3\times3}(\mathcal{DRM}(F_{s2}))\big)$;

// ***Step 4: Cross-domain Modulation***

$F^k_{spatial} = I^k_{spatial} + [1 - \sigma(\theta)] F^k_{share}$;

$F^k_{freq} = I^k_{freq} + [\sigma(\theta)] F^k_{share}$;

$F^k_{share} = F_{s2} + \mathcal{CCM}(\mathrm{Conv}_{3\times3}(\mathcal{DRM}(F_{s2})))$. F^k_{share} deeply integrate multi-domain knowledge, capturing fine-grained features superior to limited weather descriptors.

With a learnable modulation factor θ and sigmoid function $\sigma(\cdot)$, F^k_{share} serve as guidance maps injected into the spatial and frequency domains. Here, $\sigma(\theta)$ controls the injection ratio of shared guidance, while the dynamic adjustment of restoration strategy is mainly driven by the input-adaptive F^k_{share}. When degradation patterns vary subtly, such changes first perturb the underlying frequency and spatial cues in the input, which are subsequently amplified by FAP through multi-granularity frequency modulation and reflected as distinct activation patterns in F^k_{share}. These variations ultimately lead to dynamic updates of both domain outputs. In Fig 4(a), we visualize the initial features and outputs of the three domains, where brighter regions indicate stronger responses. After injecting F^k_{share}, the spatial and frequency branches exhibit enhanced responses in degradation-relevant regions, and the shared domain outputs show richer activations. These indicate that the shared domain aggregates cross-domain features into shared representations and adaptively distributes them to both domains, dynamically coordinating domain composition to accommodate variations in degradation morphology and severity.

3.3.1 Frequency-Aware Prompter (FAP)

Discriminative frequency features are key to revealing intrinsic properties of the degradations. Many methods [34, 49] lack fine-grained frequency control, limiting learning of morphology and intensity. FAP performs multi-granularity modulation across the full spectrum. Using global channel statistics, it highlights degradation-relevant responses over multiple frequency bands and preserves co-existing, heterogeneous frequency cues via coarse-to-fine refinement and gated fusion. As a result, FAP can simultaneously model low-frequency haze attenuation and mid-to-high-frequency rain/snow components in composite degradations, avoiding oversimplification to a single dominant band; the final features can capture fine spectral variations of degradations and assist in producing reliable guidance.

As shown in Fig. 4(b), given embedded features $F_e \in \mathbb{R}^{C\times H\times W}$, FAP forms a global channel descriptor using global average pooling (GAP) and $\mathrm{Conv}_{1\times1}$, and uses it to re-weight the fast Fourier transform (FFT) of the features, producing coarse frequency features:

$$F_{coarse} = \mathrm{GAP}(\mathrm{Conv}_{1\times1}(F_e)) \otimes \mathcal{F}(F_e) \tag{1}$$

where $\otimes$ denotes element-wise multiplication and $\mathcal{F}(\cdot)$ is FFT. F_{coarse} is processed by the inverse FFT (IFFT), followed by another round of GAP and $\mathrm{Conv}_{1\times1}$ to extract a fine-grained channel modulator $F^{'}_{fine}$:

$$F'_{fine} = \mathcal{F}^{-1}(F_{coarse}) \otimes \mathrm{GAP}\big(\mathrm{Conv}_{1\times1}(\mathcal{F}^{-1}(F_{coarse}))\big) \tag{2}$$

where $\mathcal{F}^{-1}(\cdot)$ denotes IFFT. Two $\mathrm{Conv}_{1\times1}$ branches (spatial view and frequency view) are then fused with learnable factors α and β to obtain a multi-level fused feature:

$$F_{mf} = \alpha \cdot \mathrm{Conv}^1_{1\times1}(F'_{fine}) \otimes \Big(\beta \cdot \mathrm{Conv}^1_{1\times1}(F'_{fine}) + \mathrm{Conv}^2_{1\times1}(F'_{fine})\Big) \tag{3}$$

Finally, a learnable gate w activated by the sigmoid $\sigma(\cdot)$ mixes $F^{'}_{fine}$ and the IFFT of the fused frequency feature:

$$F_{fine} = \sigma(w) \cdot F^{'}_{fine} + \sigma(1-w) \cdot \mathcal{F}^{-1}(F_{mf}) \tag{4}$$

3.4. Dual-path Refinement Module (DRM)

The shared domain seeks comprehensive degradation modeling via multi-domain learning, which demands more discriminative domain-specific features for effective complementarity and less redundancy. We therefore propose the Dual-path Refinement Module (DRM) in Fig. 5, which strengthens degradation-relevant features along channel and spatial dimensions and is embedded in each Domain Feature Group of MCGB. DRM has two branches: a channel branch that generates complementary channel maps to amplify degradation-relevant channels, and a spatial branch that applies multi-directional depthwise convolutions to enrich edges and textures. Their outputs are fused with the

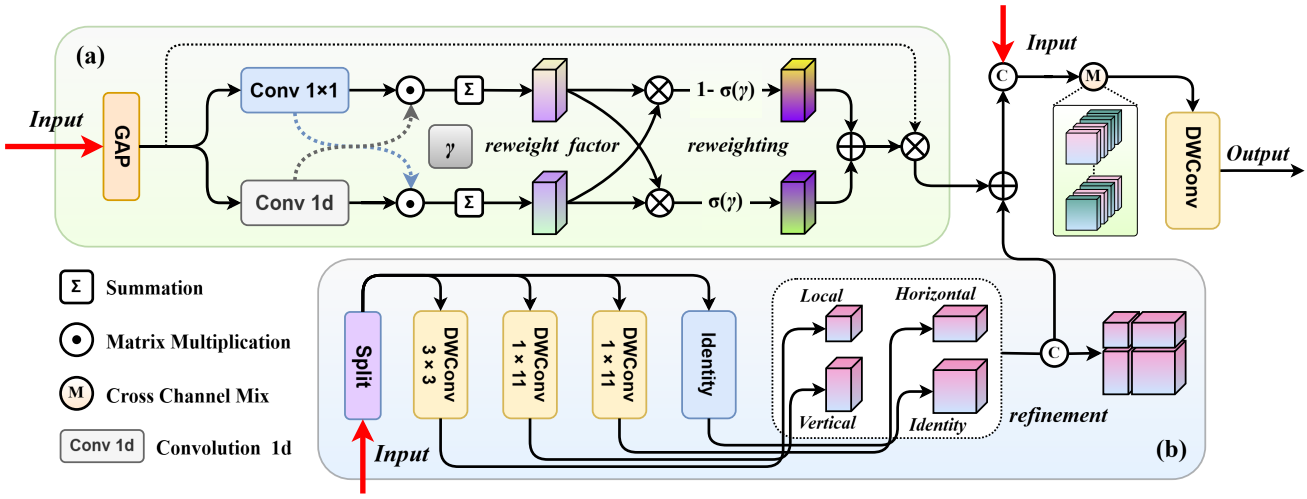

Figure 5. The structure of DRM. DRM includes two parts: (a) the channel branch that adaptively amplifies degradation-relevant channel features; (b) the spatial branch that captures multi-directional context and fine details.

input to yield degradation-sensitive and less redundant representations. Next, we detail the two branches separately.

As shown in Fig. 5(a), the channel branch operates as follows: Given an input feature map $X \in \mathbb{R}^{C\times H\times W}$, GAP is applied to obtain a global vector $V \in \mathbb{R}^{C\times 1\times 1}$. Two paths produce different channel descriptors $X_1 = \mathrm{Conv}_{1\times 1}(V)$ and $X_2 = \mathrm{Conv}_{1d}(V)$. Complementary channel maps X_{12S} and X_{21S} are formed by matrix products and channel-wise aggregation, highlighting more discriminative channels for degradation. A re-weight factor γ is then introduced to dynamically adjust the fusion between X_{12S} and X_{21S}, generating the comprehensive attention map M:

$$\begin{aligned} X_{12S} &= \mathcal{S}_c\left[X_1 X_2^\top\right] \otimes \mathcal{S}_c\left[X_1^\top X_2\right], \\ X_{21S} &= \mathcal{S}_c\left[X_2 X_1^\top\right] \otimes \mathcal{S}_c\left[X_2^\top X_1\right] \end{aligned} \tag{5}$$

$$M = \sigma(1-\gamma)\cdot X_{12S} + \sigma(\gamma)\cdot X_{21S} \tag{6}$$

where T and $\mathcal{S}_c[\cdot]$ denote transpose and summation over the channel dimension. Finally, M is multiplied element-wise with the input feature map X to obtain the final output of the channel branch: $F_{CR} = M \otimes X$.

Meanwhile, X is fed into the spatial branch (Fig. 5(b)), where it splits along channels into $S_1 \sim S_4$ and applies depthwise convolutions (DWC) that have different size kernels to capture multi-directional context, while keeping one identity path for global information. Then these features are concatenated $\mathcal{C}(\cdot)$ along the channel dimension to generate F_{SR} with a refined perception of the spatial structure:

$$\begin{aligned} Y_1 &= \mathrm{DWC}_{3\times 3}(S_1), \quad Y_2 = \mathrm{DWC}_{1\times 11}(S_2), \\ Y_3 &= \mathrm{DWC}_{1\times 11}(S_3), \quad Y_4 = S_4 \end{aligned} \tag{7}$$

$$F_{SR} = \mathcal{C}\left(Y_1,\ Y_2,\ Y_3,\ Y_4\right) \tag{8}$$

Let $\oplus$ denote element-wise addition. We fuse the two branches and refine with a $\mathrm{DWC}_{7\times 7}$, generate the final refined feature representation F_{DRM}:

$$\begin{aligned} F_{DRM} &= \mathcal{DRM}(X) \\ &= \mathrm{DWC}_{7\times 7}\left(\mathcal{C}\left(X,\ (F_{CR} + F_{SR})\right)\right) \end{aligned} \tag{9}$$

3.5. Multi-scale Progressive Fusion Block (MPFB)

MCGB provides cross-domain shared representations $\{F^k_{share},\ k = 1,2,3\}$ that encode restoration cues by aligning frequency and spatial features at three encoder scales. To fully exploit these representations, we propose the MPFB in Fig 6, which progressively fuses three F^k_{share} with high-level details, so that the shared representations from MCGB guide coarse-to-fine reconstruction of both global structures and fine textures. Let $k \in \{1,2,3\}$ denote the current stage (k=3 coarsest, k=1 finest). F^{k+1}_{MPFB} is the output from the previous (coarser) stage. In MPFB, the progressive fusion proceeds in the following three steps:

Multi-scale Aggregation. MPFB aggregates F^1_{share}, F^2_{share}, F^3_{share} at the current stage scale and obtain the multi-scale aggregated features F_{ma}:

$$\begin{aligned} F_{ma} = \mathrm{Conv}_{3\times 3}\Big(\mathrm{Conv}_{1\times 1}\big(\mathcal{C}(F^1_{share}, \\ F^2_{share},\ F^3_{share})\big)\Big) \end{aligned} \tag{10}$$

Progressive Fusion. F^{k+1}_{MPFB} and F_{ma} are processed through channel operations and addition to obtain the channel-adapted high-level feature F_H, as well as the guidance feature F_G that coarsely fuses high-level details:

$$\begin{aligned} F_H &= \mathrm{Conv}_{1\times 1}\left(F^{k+1}_{MPFB}\right), \\ F_G &= \mathrm{Conv}_{1\times 1}\left(\mathcal{C}(F_{ma},\ F_H) + F_H\right) \end{aligned} \tag{11}$$

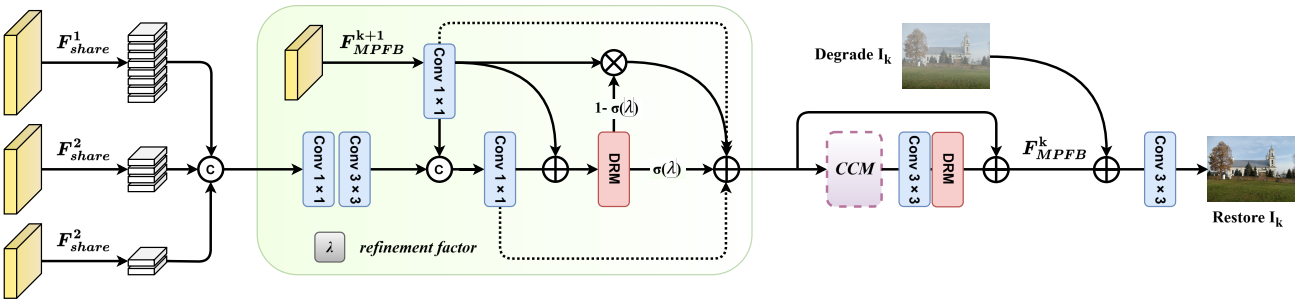

Figure 6. The structure of MPFB. By progressively fusing multi-scale shared representations, MPFB enhances consistency between global structures and local details.

To reduce feature redundancy during fusion, we further refine F_G by the refinement factor λ and DRM. F_G is processed by DRM to generate refined guidance with enhanced spatial and channel features, after which λ and DRM produce dynamic fusion weights. Finally, a skip connection combines F_G and F_H, alleviating gradient vanishing and yielding the progressively fused feature F_{PF}:

$$\begin{aligned} F_{PF} = F_G &+ \sigma(\lambda)\, \mathcal{DRM}(F_G) \\ &+ (1-\sigma(\lambda))\, \mathcal{DRM}(F_G) \cdot F_H \end{aligned} \tag{12}$$

Feature Refinement. F_{PF} are further refined through Channel Cascade Module (CCM) in Fig. 4(e) and DRM, thereby obtaining the output feature F^k_{MPFB} that integrates global structure and local details.

$$F^k_{MPFB} = \mathcal{DRM}(\mathcal{CCM}(F_{PF})) + F_{PF} \tag{13}$$

Finally, a 3×3 convolution derives the residual s from F^k_{MPFB}, and the final image is obtained as $\hat{x} = x + s$.

Through the above steps, MPFB effectively realizes progressive multi-scale feature fusion, gradually enhances the consistency between global structures and local details.

3.6. Loss Function

We introduce a dual-domain loss in the optimization flow, which consists of the spatial and frequency L_1 loss. Since each MPFB produces restored images at different scales, we extend this dual-domain loss to multiple scales.

$$\begin{cases} L_{\text{spatial}} = \sum_{r=1}^{3} \frac{1}{S_r} \left\| \hat{I}_r - Y_r \right\|_1, \\ L_{\text{frequency}} = \sum_{r=1}^{3} \frac{1}{S_r} \left\| \mathcal{F}(\hat{I}_r) - \mathcal{F}(Y_r) \right\|_1 \end{cases} \tag{14}$$

where r denotes the index of multiple inputs/outputs; $\hat{I}_r$ and Y_r are the restored image and ground truth, respectively; S denotes the number of elements for normalization; $\mathcal{F}(\cdot)$ is FFT. The final loss function is given by:

$$L = L_{\text{spatial}} + \lambda\, L_{\text{frequency}} \tag{15}$$

where λ is set to 0.1 for balancing multi-domain training.

4. Experimental Setup

4.1. Investigated Datasets.

We perform experiments on 15 benchmarks in Tab 1 to evaluate adverse weather image restoration. These benchmarks are grouped into three types: (1)Multi-Weather, (2)Specific-Weather, and (3)Real-World Weather.

4.2. Experimental Settings

Evaluation Metrics. We evaluate performance on multi-weather and specific-weather benchmarks using Peak Signal-to-Noise Ratio (PSNR) and Structure Similarity Index Measure (SSIM) following [10]. For real-world benchmarks, we adopt the non-reference metrics Q-Align and Neural Image Assessment (NIMA), consistent with [40].
Implementation. We use the following training settings: Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, initial learning rate of 2×10^{-4}, and cosine annealing strategy as in [50]. The model was trained on 256×256 patches. To accommodate different task complexities, we introduced two variants by adjusting channels: MCGNet and MCGNet-L (Large), with $c = 20$ and $c = 32$, respectively. All models were trained on four NVIDIA GeForce RTX 3090 Ti GPUs.

5. Experimental Evaluations

We evaluate MCGNet's quantitative and visual results on multi-weather, specific-weather, and real-world weather benchmarks, demonstrating its effectiveness for adverse-weather image restoration compared to baselines.

5.1. Evaluation on multi-weather benchmarks

We evaluate the performance of MCGNet under multi-weather using the *CDD11* [19] and *WeatherStream* [45].

Quantitative results on *CDD11* are reported in Tab. 2. With the fewest parameters, MCGNet achieves the best performance on all SSIM and almost all PSNR. It is slightly lower than the baselines only in the two composite scenarios involving snow. In particular, compared with weather descriptor-guided methods OneRestore [19] and WGWS-Net [50], MCGNet improves average PSNR by 0.35 dB and

Benchmark Type	Benchmark Name	Degradation Type	# Training Pairs	# Testing Pairs	Venue	Synthetic/Real
Multi-Weather	*WeatherStream* [45]	Rain, Haze, Snow	188,000	13,500	CVPR 2023	Real (different time)
	CDD11 [19]	Rain, Haze, Snow, Low-light	13,013	2,200	ECCV 2024	Synthetic
Specific-Weather	*RESIDE-OTS* [24]	Haze	313,950	550	TIP 2018	Synthetic
	SateHaze1K [21]	Haze	960	240	WACV 2020	Synthetic
	LMHAZE [37]	Haze	3,925	1,115	MMAsia 2024	Real
	Rain100L [42]	Rain	1,800	100	CVPR 2017	Synthetic
	Rain100H [42]	Rain	1,800	100	CVPR 2017	Synthetic
	Test2800 [14]	Rain	11,200	2,800	CVPR 2017	Synthetic
	Test1200 [46]	Rain	12,000	1,200	CVPR 2018	Synthetic
	RainDrop [30]	Raindrop	861	58	CVPR 2018	Synthetic
	*Snow100K** [28]	Snow	2,500	2,000	TIP 2018	Synthetic
	*SRRS** [1]	Snow	2,500	2,000	ECCV 2020	Synthetic
	*CSD** [2]	Snow	2,500	2,000	ICCV 2021	Synthetic
Real-World Weather	*RESIDE-RTTS* [24]	Haze	-	4,322	TIP 2018	Real
	WReal [40]	Rain, Snow	-	4,451	ECCV 2024	Real

Table 1. Summary of benchmarks used this paper. They are grouped into Multi-Weather, Specific-Weather, and Real-World Weather for evaluation. * denotes that we adopt the same dataset configuration as Cui *et al.* [10] to ensure a fair comparison.

Method	*CDD11-Single*								*CDD11-Double*										*CDD11-Triple*				*Average*		Params
	Low (L)		Haze (H)		Rain (R)		Snow (S)		L+H		L+R		L+S		H+R		H+S		L+H+R		L+H+S				
AirNet [23]	24.83	.778	24.21	.951	26.55	.891	26.79	.919	23.23	.779	22.82	.710	23.29	.723	22.21	.868	23.29	.901	21.80	.708	22.24	.725	23.75	.814	8.63M
TransWeather [22]	23.39	.725	23.95	.924	26.69	.899	25.74	.890	22.24	.721	22.62	.694	21.80	.661	23.10	.876	22.34	.868	21.55	.678	21.01	.655	23.13	.781	21.90M
PromptIR [29]	26.32	.805	26.10	.969	31.56	.946	31.53	.960	24.49	.789	25.05	.771	24.51	.761	24.54	.924	23.70	.925	23.74	.752	23.33	.747	25.90	.850	38.45M
WGWSNet [50]	24.39	.774	27.90	.982	33.15	.964	34.43	.973	24.27	.800	25.06	.772	24.60	.765	27.23	.955	27.65	.960	23.90	.772	23.97	.771	26.96	.863	25.76M
WeatherDiff [52]	23.58	.763	21.99	.904	24.85	.885	24.80	.888	21.83	.756	22.69	.730	22.12	.707	21.25	.868	21.99	.868	21.23	.716	21.04	.698	22.49	.799	82.96M
InstructIR [7]	26.70	.809	32.61	.978	33.51	.940	34.45	.948	24.36	.800	25.41	.782	**25.63**	.778	28.80	.921	29.64	.959	24.84	.777	24.32	.760	28.21	.859	15.80M
OneRestore [19]	26.48	.826	32.52	.990	33.40	.964	34.31	.973	25.79	.822	25.58	.799	25.19	.789	29.99	.957	30.21	.964	24.78	.788	**24.90**	.791	28.47	.878	5.98M
MCGNet	**27.13**	**.841**	**32.94**	**.992**	**33.96**	**.970**	**35.48**	**.975**	**26.03**	**.847**	**26.17**	**.824**	24.72	**.799**	**30.06**	**.968**	**30.70**	**.967**	**25.36**	**.819**	24.50	**.804**	**28.82**	**.891**	**5.75M**

Table 2. Comparison on *CDD11* [19] under Single/Double/Triple settings. Each cell shows PSNR (left) and SSIM (right).

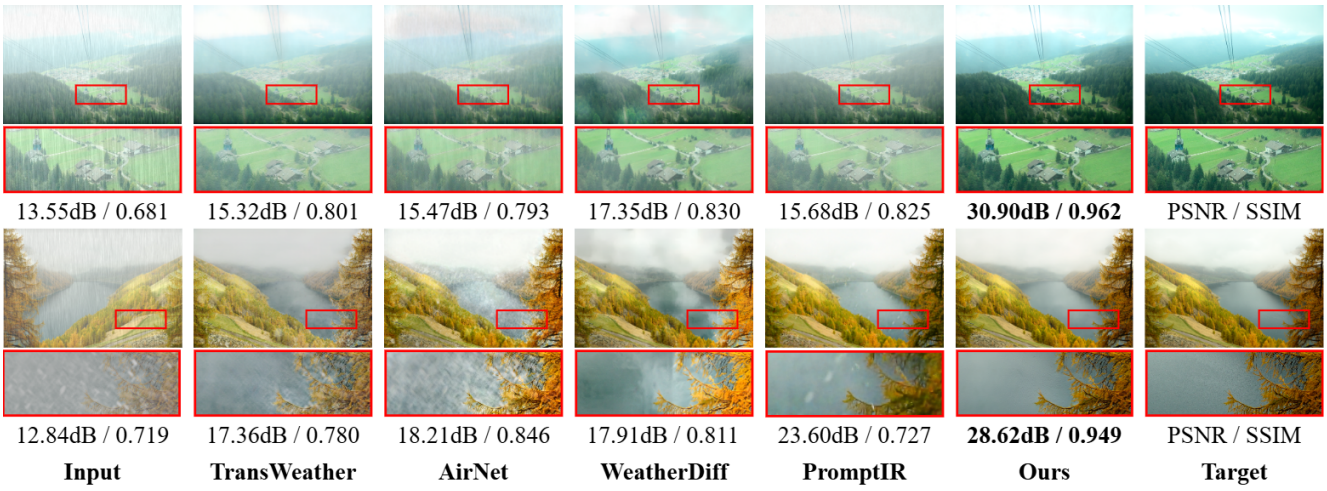

Figure 7. Visual comparison on *CDD11 [19]*: H+R (haze+rain, top) and L+H+S (low+haze+snow, bottom) degradations.

1.86 dB, respectively, highlighting the advantage of cross-domain shared representations. MCGNet further improves the average SSIM over OneRestore [19] by 0.013, indicating that multi-domain collaboration and multi-scale learning better capture structural variations in degraded images without causing structural distortions. Visual comparisons on *CDD11* (Fig. 7) show that MCGNet removes overlapping artifacts while preserving fine details, whereas other methods often exhibit color shifts and residual noise. These results demonstrate that MCGNet more effectively models

Method	Rain		Haze		Snow		#Params
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	
TKL [3]	23.22	0.795	22.38	0.805	22.25	0.827	28.71M
Uformer [36]	22.25	0.791	18.81	0.763	20.94	0.801	20.63M
GRL [26]	23.75	0.805	22.88	0.802	22.59	0.829	20.20M
WGWSNet [50]	23.80	0.807	22.78	0.800	22.72	**0.831**	25.76M
Restormer* [27]	23.64	0.805	23.01	**0.806**	22.66	0.821	26.13M
RainRobust* [27]	23.68	0.799	22.71	0.797	22.54	0.816	12.90M
SFSNiD [8]	24.25	0.842	22.63	0.751	22.33	0.786	8.35M
MCGNet	**24.82**	**0.868**	**23.82**	0.773	**22.84**	**0.831**	**5.75M**

Table 3. Comparison on *WeatherStream* [45]. * indicates the model optimized with the CLC and IAS Liu *et al.* [27] proposed.

intrinsic degradation and produces more natural results.

On *WeatherStream*, the results in Tab. 3 show that MCGNet attains the best performance on almost all metrics, with the lowest parameters; only on haze does the SSIM fall slightly below [27]. Relative to [8], MCGNet improves PSNR by 0.57, 1.19, and 0.51dB on rain, haze, and snow, further validating its effectiveness under complex multi-weather conditions.

5.2. Evaluation on specific-weather benchmarks

This section provides controlled validation of MCGNet's capability on each weather, showing attributable gains per task rather than merely averaging better in multi-weather scenes. We evaluate dehazing, deraining, raindrop removal, and desnowing to verify performance.

Image desnowing. We conduct desnowing experiments on three benchmarks: *CSD* [2], *SRRS* [1], and *Snow100K* [28] (referring to the data setting of the practical implementation in Cui *et al.* [10] proposed). Quantitative results on these benchmarks are reported in Tab. 4. Compared with the recent MHNet [16], we improve PSNR by 1.81, 0.39, and 0.82 dB on *CSD*, *Snow100K*, and *SRRS*, respectively, with 11.15M fewer parameters. Visual results in Fig. 8 present that MCGNet cleaner removes stubborn snow particles and distant haze, recovering occluded details to a greater extent.

Image deraining and Raindrop Removal. We conduct deraining experiments on five benchmarks. As shown in Tabs. 5–6, we achieve the best performance with fewer parameters. Tab 5 shows comparison on *Rain100L/H* [42] and *RainDrop* [30], and Tab 6 shows comparison on *Test1200* [46] and *Test2800* [46]. Specifically, on *Rain100L/H*, MCGNet attains PSNR gains of 1.44 dB and 1.18 dB over MHNet [16] with 14% fewer parameters. On *RainDrop*, MCGNet surpasses GridFormer [35] by 0.37 dB PSNR with 15.63M fewer parameters. Fig. 9 illustrates that MCGNet removes large raindrops more thoroughly while restoring local details. On the *Test1200* and *Test2800*, MCGNet provides a PSNR improvement of 1.94 dB over the DMSR [51] with 20.98M fewer parameters.

Method	*Snow100K*		*CSD*		*SRRS*		#Params
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	
DesnowNet [28]	30.50	0.940	20.13	0.810	20.38	0.840	15.60M
TKL [3]	30.31	0.896	30.11	0.934	29.37	0.934	28.71M
TransWeather [22]	31.82	0.930	31.76	0.930	28.29	0.920	21.90M
WeatherDiff [52]	30.43	0.915	35.23	0.976	-	-	82.96M
ConvIR-B [10]	33.92	**0.960**	39.10	**0.990**	32.29	**0.980**	8.63M
FSNet [9]	33.76	0.950	38.37	**0.990**	32.33	**0.980**	13.28M
MHNet [16]	33.97	**0.960**	37.44	0.980	32.18	0.970	16.90M
MCGNet	**34.36**	0.952	**39.25**	**0.990**	**33.00**	0.971	**5.75M**

Table 4. Comparison of image desnowing on the *Snow100K* [28], *CSD* [2], and *SRRS* [1], using the same dataset setting as [10]

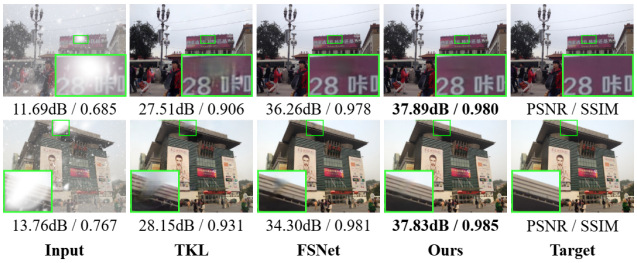

Figure 8. Visual comparison of image desnowing on *CSD* [2].

Method	*Rain100L*		*Rain100H*		*RainDrop*		#Params
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	
MAXIM-2S [33]	38.06	0.977	30.81	0.903	31.87	0.935	14.10M
Restormer [44]	38.99	0.978	31.46	0.904	30.85	0.928	26.13M
WeatherDiff [52]	36.73	0.969	31.08	0.902	29.66	0.923	82.96M
PromptIR [29]	36.59	0.968	30.83	0.894	32.03	0.938	38.45M
GridFormer [35]	34.24	0.965	-	-	32.39	0.936	30.12M
FSNet [9]	38.00	0.972	31.77	0.906	30.85	0.925	13.28M
MHNet [16]	40.04	0.985	31.08	0.899	-	-	16.90M
MCGNet-L	**41.48**	**0.989**	**32.26**	**0.924**	**32.76**	**0.939**	**14.49M**

Table 5. Comparison of image deraining on *Rain100L* [42], *Rain100H* [42] and *RainDrop* [30].

Method	*Test1200*		*Test2800*		*Average*		#Params
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	
MAXIM-2S [33]	32.37	0.922	33.80	0.943	33.08	0.933	14.10M
SFNet [12]	32.55	0.911	33.69	0.937	33.12	0.924	13.27M
DRSformer [4]	30.99	0.906	32.80	0.931	31.90	0.919	33.79M
FSNet [9]	33.08	0.916	33.64	0.936	33.36	0.926	13.28M
ConvIR-L [10]	-	-	33.73	0.937	-	-	14.83M
MambaIR [18]	32.56	0.923	33.58	0.927	33.07	0.925	31.51M
DMSR [51]	31.35	0.910	32.50	0.931	31.93	0.921	35.47M
MCGNet-L	**33.85**	**0.931**	**33.89**	**0.944**	**33.87**	**0.941**	**14.49M**

Table 6. Comparison of image deraining on *Test1200* [46] and *Test2800* [14].

Image dehazing. We conduct dehazing experiments on three benchmarks: *RESIDE-OTS* [24], *SateHaze1K* [21], and *LMHAZE* [37]. As shown in Tab. 7, MCGNet achieves the best performance. Specifically, on *OTS*, it surpasses the DCMPNet [47] by 1.77dB PSNR, with 11.61M fewer pa-

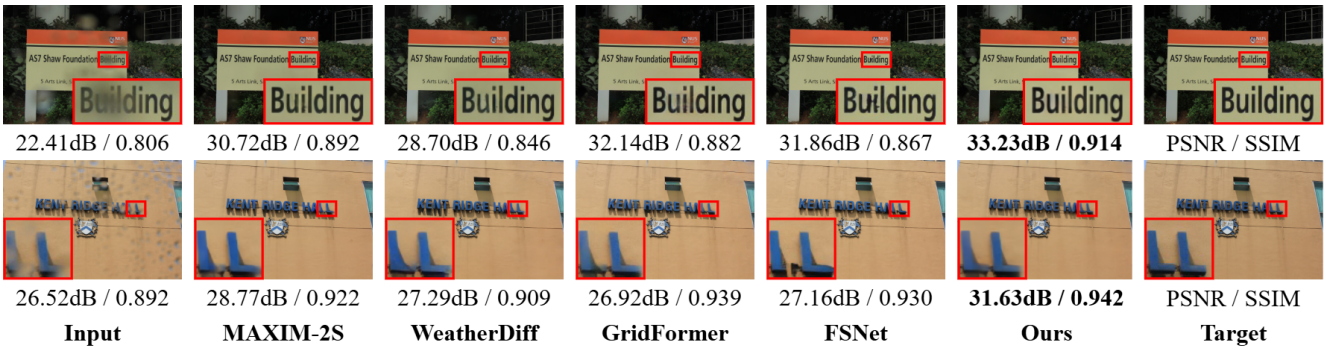

Figure 9. Visual comparison of raindrop removal on *RainDrop* [30].

Method	*RESIDE-OTS*		Method	*SateHaze1K*		Method	*LMHAZE*	
	PSNR/SSIM ↑	#Params		PSNR/SSIM ↑	#Params		PSNR/SSIM ↑	#Params
DeHamer [17]	35.18 / 0.986	132.45M	Uformer [36]	22.55 / 0.887	20.63M	AECRNet [38]	15.80 / 0.466	2.61M
MAXIM-2S [33]	34.19 / 0.985	14.10M	Restormer [44]	22.13 / 0.869	26.13M	DeHamer [17]	15.76 / 0.578	132.45M
C2PNet [48]	36.68 / 0.990	7.17M	C2PNet [48]	20.41 / 0.870	7.17M	DHFormer [32]	17.70 / 0.762	25.44M
DCMPNet [47]	36.56 / **0.995**	17.36M	Trinity-Net [6]	21.96 / 0.879	20.24M	MB-Taylor [31]	17.73 / 0.729	7.43M
FPro [49]	32.85 / 0.983	22.30M	FCDM [34]	17.76 / 0.826	155.33M	DEA-Net [5]	17.96 / 0.745	3.65M
OneRestore [19]	35.58 / 0.981	5.98M	OKNet [11]	22.09 / 0.872	4.72M	LDR [41]	17.97 / 0.746	9.30M
FDTANet [15]	34.73 / 0.989	34.10M	AU-Net [13]	23.48 / 0.893	7.78M	MambaIR [18]	17.94 / 0.736	26.78M
MCGNet	**38.33** / 0.991	**5.75M**	**MCGNet**	**25.39 / 0.902**	**5.75M**	**MCGNet**	**18.69 / 0.788**	**5.75M**

Table 7. Comparison of image dehazing on *RESIDE-OTS* [24], *SateHaze1K* [21], and *LMHAZE* [37].

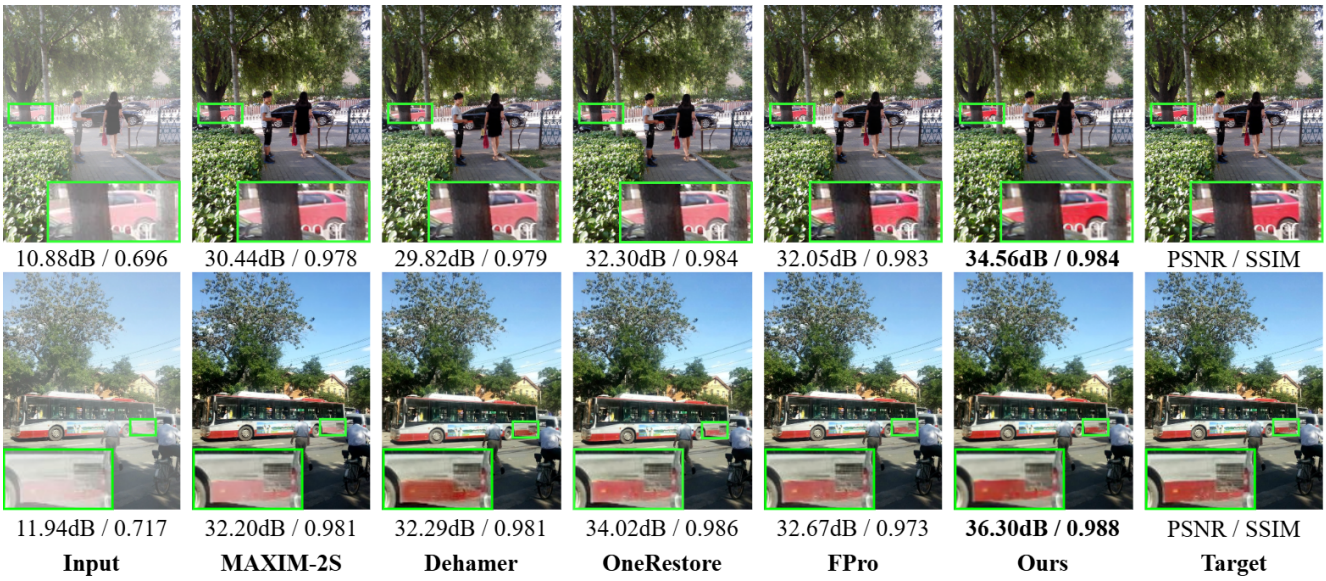

Figure 10. Visual comparison of image dehazing on *RESIDE-OTS* [24].

rameters. Fig. 10 shows that MCGNet delivers higher color fidelity and sharper local textures. On the remote-sensing *SateHaze1K* [21] and real-world *LMHAZE* [37], MCGNet also attains the best results, with PSNR improvements of 1.91dB and 0.72dB over the best baselines [13, 41].

These results show MCGNet's adaptivity to each weather and attributable gains per task rather than merely averaging better in multi-weather scenes. MCGNet perceives the frequency properties of various weather via FAP and further amplifies feature discriminability through DRM. The shared domain converts multi-domain knowledge into reliable guidance, enabling effective handling of degradations with diverse scales and intensities through multi-domain collaboration and multi-scale fusion.

Figure 11. Visual comparison on real-world benchmarks *RTTS* [24] and *WReal* [40].

Method	Rain		Haze		Snow		#Params
	Q-Align↑	NIMA↑	Q-Align↑	NIMA↑	Q-Align↑	NIMA↑	
PromptIR [29]	3.548	5.424	3.093	4.798	3.698	5.404	38.45M
WGWSNet [50]	3.553	5.459	2.915	4.821	3.736	5.414	25.76M
ConvIR [10]	3.618	5.427	3.084	**4.826**	3.694	5.437	8.63M
OneRestore [19]	3.492	5.341	2.979	4.758	3.628	5.353	5.98M
MCGNet	**3.660**	**5.468**	**3.244**	4.814	**3.817**	**5.476**	**5.75M**

Table 8. Comparison on real-world benchmarks *RTTS* [24] and *WReal* [40].

5.3. Evaluation on real-world benchmarks

To verify the generalization of MCGNet under real-world adverse weather conditions, we evaluated models trained on a composite dataset comprising multiple public datasets [2, 14, 24, 28, 39, 45, 50] and assessed its performance on real-world benchmarks *RTTS* [24] and WReal [40]. Whenever available, we used public checkpoints for comparison. As shown in Tab. 8, our method achieves the best performance on almost all metrics compared to existing state-of-the-art all-in-one image restoration methods, only on haze does the NIMA fall slightly below the baselines. Compared with the descriptor-guided baseline OneRestore [19], MCGNet surpasses it on Q-Align by 0.168, 0.265, and 0.189 for rain, haze, and snow, respectively. It presents the effectiveness of our multi-domain collaborative guidance strategies.

In addition, Fig. 11 provides the visual comparison, where our method effectively removes haze, rain, and snow artifacts in real-world images while better preserving object edges and texture details. Furthermore, our approach shows robust performance in composite degradation (such as haze and snow in Fig. 11(4)), handling both distant haze and nearby snow simultaneously. Visually, our method achieves more accurate color and luminance restoration. As shown in Fig. 11(3), our restored results avoid the false colors caused by excessive enhancement, yielding outputs that are closer.

6. Ablation Study

In this section, we conduct ablation studies to demonstrate the effectiveness of the proposed components. All models are configured with $c = 20$ and trained on the *CDD11* [19] for 200 epochs. The reported FLOPs and Latency are measured on a 256×256 image.

6.1. Multi-domain Configurations in MCGB

Tab. 9 evaluates domain configurations in MCGB by selectively activating the spatial, frequency, and shared domains. The full configuration achieves a 1.47 dB improvement in PSNR over the spatial-only baseline. Turning on only the frequency branch adds 0.34 dB over spatial-only, showing that frequency cues capture degradation more effectively than purely spatial features. The shared domain is crucial for cross-domain alignment, as it provides a further 0.98 dB PSNR gain relative to the simple two-branch setup (spatial + frequency). Although it adds a small amount of complexity, the improvement is well justified.

We visualize in Fig. 12 the spectra of the learning residuals between degraded images and restored images across four weather types. From the highlighted red circles, we observe: for deraining, slanted bright streaks, indicating strong high-frequency energy from many aligned rain streaks; for raindrop removal, a clear central cross with oblique blobs, reflecting local circular edges caused by drop occlusions; for desnowing, a near-circular granular halo around the center, revealing isotropic high-frequency energy from dense small particles; dehazing concentrates

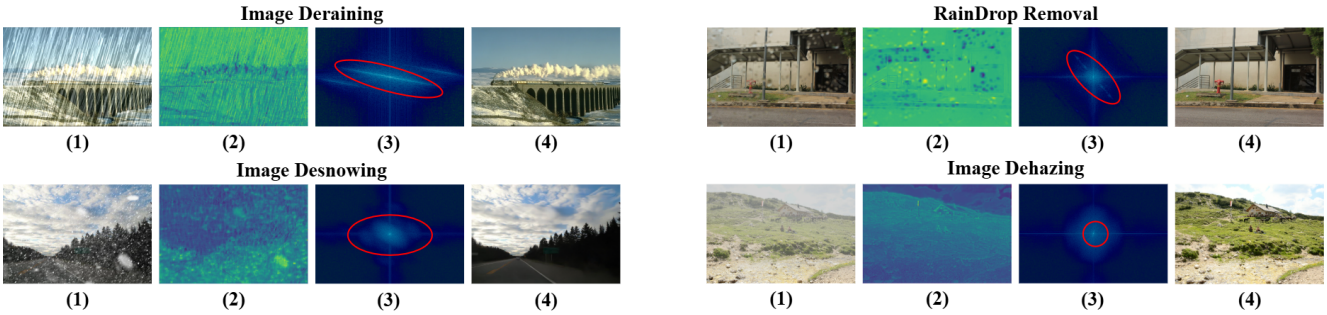

Figure 12. The residual features and corresponding spectra for four weather degradations. The results are sampled from the last MPFB. The numbers in parentheses represent: (1) degraded image; (2) learned residual features; (3) frequency spectra of (2); (4) restored image.

Spatial	Freq	Shared	PSNR↑	SSIM↑	#Params	FLOPs	Latency
✓	✗	✗	27.35	0.878	3.55M	42.90G	25.24ms
✗	✓	✗	27.69	0.880	3.58M	43.01G	26.14ms
✓	✓	✗	27.84	0.882	3.76M	44.23G	28.92ms
✗	✗	✓	28.08	0.884	5.35M	54.64G	28.93ms
✓	✗	✓	28.35	0.885	5.54M	56.03G	34.04ms
✗	✓	✓	28.51	0.886	5.57M	56.14G	35.27ms
✓	✓	✓	**28.82**	**0.891**	**5.75M**	**57.53G**	**38.92ms**

Table 9. Ablation study of multi-domain configurations in MCGB.

CF	CM	PSNR↑	SSIM↑	#Params	FLOPs	Latency
✗	✗	28.12	0.884	5.69M	56.80G	37.07ms
✗	✓	28.26	0.884	5.69M	56.81G	37.49ms
✓	✗	28.63	0.887	5.75M	57.52G	38.27ms
✓	✓	**28.82**	**0.891**	**5.75M**	**57.53G**	**38.92ms**

Table 10. Ablation study of Cross-domain Fusion (CF) and Cross-domain Modulation (CM) in the shared domain.

In MCGB	In MPFB	PSNR↑	SSIM↑	#Params	FLOPs	Latency
✗	✗	28.11	0.884	5.48M	54.71G	33.26ms
✗	✓	28.54	0.886	5.61M	55.86G	36.13ms
✓	✗	28.61	0.887	5.62M	56.12G	36.19ms
✓	✓	**28.82**	**0.891**	**5.75M**	**57.53G**	**38.92ms**

Table 11. Ablation study of DRM embedding position

energy near the center that rapidly decays outward, evidencing large-scale low-frequency attenuation. These results confirm that our approach effectively captures distinct frequency patterns of various weather degradations. Such spectral differences guide dynamic strategy adjustment: when low-frequency attenuation dominates, the strategy should emphasize global structural recovery; when mid-to-high-frequency components dominate, it should focus on suppressing streak/particle artifacts while preserving details. Accordingly, our model leverages shared guidance to modulate the spatial and frequency branches, thereby dynamically balancing cross-domain collaboration to accommodate variations in degradation morphology and severity.

6.2. The Shared Domain Configurations

The core of the shared domain is the Cross-domain Fusion (CF, Algorithm 1 Step 2) and the Cross-domain Modulation (CM, Algorithm 1 Step 4). We analyze their effects via ablations in Tab 10. Relative to the baseline, introducing CF or CM alone increases PSNR by 0.51 dB and 0.14 dB, respectively, and enabling both yields a 0.70 dB gain with negligible added complexity. These results validate our design: the Cross-domain Fusion(CF) reduces the semantic gap across domains, producing more consistent features. The Cross-domain Modulation(CM) uses the shared guidance to modulate the frequency and spatial outputs, allowing MCGNet to select the most suitable domain combination and thus adjust the restoration strategy dynamically.

6.3. DRM Embedding Position

We evaluate the effect of embedding DRM in different components (see Tab 11). Within MCGB, DRM strengthens domain-specific features and reduces redundancy during cross-domain fusion. In MPFB, DRM enables dynamic allocation of fusion weights. Compared to the baseline without DRM, our full configuration yields a 0.71 dB PSNR improvement. Introducing DRM only in MCGB or only in MPFB brings 0.50 dB and 0.41 dB gains, respectively, striking a good balance between parameters and performance. These results highlight the positive impact of DRM on multi-domain collaboration and multi-scale fusion.

6.4. Multi-scale Fusion Configurations in MPFB

MPFB reconstructs images from coarse to fine through progressive fusion guided by MCGB's cross-domain shared representations. Its core comprises Multi-scale Aggregation (MA, Step 1 in Section 3.5) and Progressive Fusion (PF, Step 2 in Section 3.5). MA aggregates the shared representations from multiple MCGB stages at the current resolution; PF describes the process of progressively fusing the output of MA and high-level MPFB. We therefore ablate MA and PF to assess their contributions at both the MPFB

with MA	*with* PF	PSNR↑	SSIM↑	#Params	FLOPs	Latency
✗	✗	28.19	0.883	5.65M	56.24G	36.18ms
✓	✗	28.34	0.885	5.66M	56.50G	37.06ms
✗	✓	28.54	0.886	5.74M	57.27G	38.06ms
✓	✓	**28.82**	**0.891**	**5.75M**	**57.53G**	**38.92ms**

Table 12. Ablation study of fusion configurations in MPFB

and overall levels. As shown in Tab 12, the full MPFB configuration yields a 0.63 dB PSNR gain over the baseline. Examined in isolation, only PF improves PSNR by 0.15 dB, while only MA adds 0.35 dB, indicating that the multi-scale shared representations encode rich degradation information that boosts decoder reconstruction quality. In addition, PF adaptively fuses high-level guidance into lower-level features, stabilizing fusion. This combination of MA and PF achieves the best performance, strengthening the consistency of global structure and local detail.

7. Conclusion

We propose MCGNet, an efficient framework for adverse-weather image restoration that dynamically infers the optimal restoration strategy from the intrinsic characteristics of degradations. MCGNet consists of three components: MCGB, DRM, and MPFB. MCGB forms unified cross-domain shared representations via the shared domain. These representations not only encode fine-grained weather-degradation information but also serve as adaptive guidance to modulate the outputs of the two domains, enabling dynamic adjustment of the restoration strategy. DRM amplifies fine-grained degradation features along the channel and spatial dimensions, thereby reducing feature redundancy during fusion. MPFB improves multi-scale adaptability through progressive fusion, ensuring consistency between global structures and local details. Extensive experiments on 15 datasets demonstrate that MCGNet surpasses state-of-the-art methods with fewer parameters and significantly improves the visual quality of images captured under real-world adverse weather conditions. Although MCGNet is parameter-efficient, its multi-domain feature pipeline and cross-domain fusion introduce additional computational overhead and runtime latency. We consider pruning or distilling the frequency and spatial branches into a lighter student while retaining the shared-domain guidance, aiming to balance computational cost and performance.

References

[1] W.-T. Chen, H.-Y. Fang, J.-J. Ding, C.-C. Tsai, and S.-Y. Kuo. Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In *ECCV*, page 754–770, 2020. 8, 9

[2] W.-T. Chen, H.-Y. Fang, C.-L. Hsieh, C.-C. Tsai, I.-H. Chen, J.-J. Ding, and S.-Y. Kuo. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In *ICCV*, pages 4176–4185, 2021. 1, 2, 8, 9, 11

[3] W.-T. Chen, Z.-K. Huang, C.-C. Tsai, H.-H. Yang, J.-J. Ding, and S.-Y. Kuo. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In *CVPR*, pages 17632–17641, 2022. 3, 9

[4] X. Chen, H. Li, M. Li, and J. Pan. Learning a sparse transformer network for effective image deraining. In *CVPR*, pages 5896–5905, 2023. 1, 2, 9

[5] Z. Chen, Z. He, and Z.-M. Lu. Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE Trans. Image Process.*, 33:1002–1015, 2024. 10

[6] K. Chi, Y. Yuan, and Q. Wang. Trinity-net: Gradient-guided swin transformer-based remote sensing image dehazing and beyond. *IEEE Trans. Geosci. Remote Sens.*, 61:1–14, 2023. 10

[7] M. V. Conde, G. Geigle, and R. Timofte. Instructir: High-quality image restoration following human instructions. In *ECCV*, pages 1–21, 2024. 8

[8] X. Cong, J. Gui, J. Zhang, J. Hou, and H. Shen. A semi-supervised nighttime dehazing baseline with spatial-frequency aware and realistic brightness constraint. In *CVPR*, pages 2631–2640, 2024. 2, 9

[9] Y. Cui, W. Ren, X. Cao, and A. Knoll. Image restoration via frequency selection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(2):1093–1108, 2024. 3, 9

[10] Y. Cui, W. Ren, X. Cao, and A. Knoll. Revitalizing convolutional network for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(12):9423–9438, 2024. 3, 7, 8, 9, 11

[11] Y. Cui, W. Ren, and A. Knoll. Omni-kernel network for image restoration. In *AAAI*, pages 1426–1434, 2024. 1, 2, 10

[12] Y. Cui, Y. Tao, Z. Bing, W. Ren, X. Gao, X. Cao, K. Huang, and A. Knoll. Selective frequency network for image restoration. In *ICLR*, 2023. 3, 9

[13] Y. Du, J. Li, Q. Sheng, Y. Zhu, B. Wang, and X. Ling. Dehazing network: Asymmetric unet based on physical model. *IEEE Trans. Geosci. Remote Sens.*, 62:1–12, 2024. 10

[14] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley. Removing rain from single images via a deep detail network. In *CVPR*, pages 1715–1723, 2017. 8, 9, 11

[15] H. Gao, B. Ma, Y. Zhang, J. Yang, and D. Dang. Frequency domain task-adaptive network for restoring images with combined degradations. *Pattern Recognit.*, 158:111057, 2025. 2, 10

[16] H. Gao, Y. Zhang, J. Yang, and D. Dang. Mixed hierarchy network for image restoration. *Pattern Recognit.*, 161:111313, 2025. 9

[17] C. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren, and C. Li. Image dehazing transformer with transmission-aware 3d position embedding. In *CVPR*, pages 5802–5810, 2022. 1, 2, 10

[18] H. Guo, J. Li, T. Dai, Z. Ouyang, X. Ren, and S. Xia. Mambair: A simple baseline for image restoration with state-space model. In *ECCV*, pages 222–241, 2024. 9, 10

[19] Y. Guo, Y. Gao, Y. Lu, R. W. Liu, and S. He. Onerestore: A universal restoration framework for composite degradation.

In *ECCV*, pages 255–272, 2024. 1, 2, 3, 7, 8, 10, 11

[20] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. In *CVPR*, pages 1956–1963, 2009. 2

[21] B. Huang, Z. Li, C. Yang, F. Sun, and Y. Song. Single satellite optical imagery dehazing using sar image prior based on conditional generative adversarial networks. In *WACV*, pages 1806–1813, 2020. 8, 9, 10

[22] J. M. Jose Valanarasu, R. Yasarla, and V. M. Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *CVPR*, pages 2343–2353, 2022. 1, 3, 8, 9

[23] B. Li, X. Liu, P. Hu, Z. Wu, J. Lv, and X. Peng. All-in-one image restoration for unknown corruption. In *CVPR*, pages 17431–17441, 2022. 8

[24] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang. Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.*, 28(1):492–505, 2019. 8, 9, 10, 11

[25] R. Li, R. T. Tan, and L.-F. Cheong. All in one bad weather removal using architectural search. In *CVPR*, pages 3172–3182, 2020. 1, 3

[26] Y. Li, Y. Fan, X. Xiang, D. Demandolx, R. Ranjan, R. Timofte, and L. Van Gool. Efficient and explicit modelling of image hierarchies for image restoration. In *CVPR*, pages 18278–18289, 2023. 9

[27] X. Liu, Z. Zhang, X. Wu, Y. Wu, and Y. Liu. Learning real-world image de-weathering with imperfect supervision. In *AAAI*, pages 3738–3746, 2024. 9

[28] Y. Liu, D. Jaw, S. Huang, and J. Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Trans. Image Process.*, 27(6):3064–3073, 2018. 2, 8, 9, 11

[29] V. Potlapalli, S. W. Zamir, S. Khan, and F. S. Khan. Promptir: Prompting for all-in-one image restoration. In *NIPS*, pages 71275–71293, 2023. 3, 8, 9, 11

[30] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu. Attentive generative adversarial network for raindrop removal from a single image. In *CVPR*, pages 2482–2491, 2018. 2, 8, 9, 10

[31] Y. Qiu, K. Zhang, C. Wang, W. Luo, H. Li, and Z. Jin. Mb-taylorformer: Multi-branch efficient transformer expanded by taylor formula for image dehazing. In *ICCV*, pages 12756–12767, 2023. 2, 10

[32] Y. Song, Z. He, H. Qian, and X. Du. Vision transformers for single image dehazing. *IEEE Trans. Image Process.*, 32:1927–1941, 2023. 10

[33] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li. Maxim: Multi-axis mlp for image processing. In *CVPR*, pages 5759–5770, 2022. 9, 10

[34] J. Wang, S. Wu, Z. Yuan, Q. Tong, and K. Xu. Frequency compensated diffusion model for real-scene dehazing. *Neural Networks.*, 175:106281, 2024. 1, 2, 3, 5, 10

[35] T. Wang, K. Zhang, Z. Shao, W. Luo, B. Stenger, T. Lu, T.-K. Kim, W. Liu, and H. Li. Gridformer: Residual dense transformer with grid structure for image restoration in adverse weather conditions. *Int. J. Comput. Vis.*, 132(10):4541–4563, 2024. 9

[36] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, pages 17662–17672, 2022. 9, 10

[37] Z. Wang, H. Wu, Y. Yang, Y. Qu, L. Ma, and S. Lin. Lmhaze: Intensity-aware image dehazing with a large-scale multi-intensity real haze dataset. In *Proc. ACM Int. Conf. Multimedia in Asia*, 2024. 8, 9, 10

[38] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma. Contrastive learning for compact single image dehazing. In *CVPR*, pages 10546–10555, 2021. 10

[39] J. Xu, X. Hu, L. Zhu, and P.-A. Heng. Unifying physically-informed weather priors in a single model for image restoration across multiple adverse weather conditions. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(10):9575–9591, 2025. 1, 11

[40] J. Xu, M. Wu, X. Hu, C.-W. Fu, Q. Dou, and P.-A. Heng. Towards real-world adverse weather image restoration: Enhancing clearness and semantics with vision-language models. In *ECCV*, pages 147–164, 2024. 7, 8, 11

[41] H. Yang, L. Pan, Y. Yang, and W. Liang. Language-driven all-in-one adverse weather removal. In *CVPR*, pages 24902–24912, 2024. 10

[42] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan. Deep joint rain detection and removal from a single image. In *CVPR*, pages 1685–1694, 2017. 8, 9

[43] Y. Yang, G. Yuan, and J. Li. Sffnet: A wavelet-based spatial and frequency domain fusion network for remote sensing segmentation. *IEEE Trans. Geosci. Remote Sens.*, 62:1–17, 2024. 4

[44] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, pages 5718–5729, 2022. 9, 10

[45] H. Zhang, Y. Ba, E. Yang, V. Mehra, B. Gella, A. Suzuki, A. Pfahnl, C. Chandrappa, A. Wong, and A. Kadambi. Weatherstream: Light transport automation of single image deweathering. In *CVPR*, pages 13499–13509, 2023. 7, 8, 9, 11

[46] H. Zhang and V. M. Patel. Density-aware single image deraining using a multi-stream dense network. In *CVPR*, pages 695–704, 2018. 8, 9

[47] Y. Zhang, S. Zhou, and H. Li. Depth information assisted collaborative mutual promotion network for single image dehazing. In *CVPR*, pages 2846–2855, 2024. 9, 10

[48] Y. Zheng, J. Zhan, S. He, J. Dong, and Y. Du. Curricular contrastive regularization for physics-aware single image dehazing. In *CVPR*, pages 5785–5794, 2023. 10

[49] S. Zhou, J. Pan, J. Shi, D. Chen, L. Qu, and J. Yang. Seeing the unseen: A frequency prompt guided transformer for image restoration. In *ECCV*, pages 246–264, 2024. 3, 5, 10

[50] Y. Zhu, T. Wang, X. Fu, X. Yang, X. Guo, J. Dai, Y. Qiao, and X. Hu. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *CVPR*, pages 21747–21758, 2023. 1, 2, 3, 7, 8, 9, 11

[51] S. Zou, Y. Zou, M. Zhang, S. Luo, G. Gao, and G. Qi. Learning dual-domain multi-scale representations for single image deraining. In *ICME*, pages 1–6, 2025. 1, 2, 9

[52] O. Özdenizci and R. Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(8):10346–10357, 2023. 2, 3, 8, 9