WeaFU: Weather-Informed Image Blind Restoration via Multi-Weather Distribution Diffusion

Bodong Cheng, Juncheng Li*, Jun Shi, Yingying Fang, Guixu Zhang, Yin Chen, Tieyong Zeng, Zhi Li*

***Abstract*—The extraction of distribution from images with diverse weather conditions is crucial for enhancing the robustness of visual algorithms. When addressing image degradation caused by different weather, accurately perceiving the data distribution of weather-informed degradation becomes a fundamental challenge. However, given the highly stochastic nature, modelling weather distribution poses a formidable task. In this paper, we propose a novel multi-Weather distribution difFUsion blind restoration model, named WeaFU. Firstly, the model employs representation learning to map image distribution into a latent space. Subsequently, WeaFU utilizes a diffusion-based approach, with the assistance of Diffusion Distribution Generator (DDG), to perceive and extract corresponding weather distribution. This strategy ingeniously injects data distribution into the recovery process, significantly enhancing the robustness of the model in diverse weather scenarios. Finally, a Conditional Distribution-Aware Transformer (CDAT) is constructed to align the distribution information with pixels, thereby obtaining clear images. Extensive experiments on real and synthetic datasets demonstrate that WeaFU achieves superior performance.**

***Index Terms*—Image restoration, diffusion model, distribution learning.**

I. Introduction

In various visual tasks, the impact of extreme weather information on images is significantly amplified. This is particularly evident in tasks such as object detection [1]–[5], image segmentation [6]–[9], and facial recognition [10], [11]. Confronted with such ill-posed problems, image restoration for various weather often poses even more intricate challenges. Due to the inability to determine the weather type of the current image, restoration models usually incur higher costs when learning this unstable degradation form.

Numerous methods [12]–[19] have been proposed to address tasks associated with the specific weather information, including image deraining, dehazing, snow removal, and smoke removal. However, these methods are limited to individual tasks, presenting difficulties in extending their applicability to diverse scenarios. In the broader framework of unified restoration models, the intelligent integration of diverse methods seems ostensibly uncomplicated. Nevertheless, this strategy comes with a higher storage and computational costs. More importantly, it is unable to perform restoration for multi-weather scenarios.

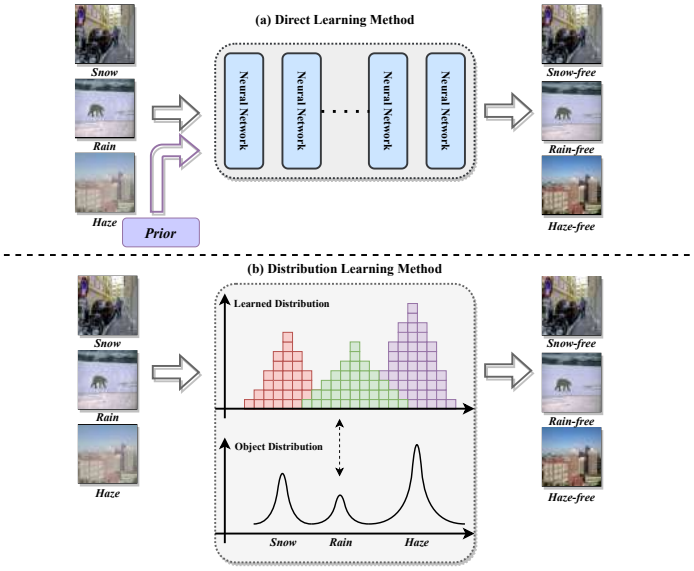

Fig. 1: The proposed method differs from traditional direct learning method. The direct learning method uses neural networks to narrow the gap between LQ and HQ images. This strategy limits the ability of the model to extract distribution information. In contrast, our method focuses primarily on the distribution. By leveraging diffusion-based representations, we effectively learn the distribution of images. Subsequently, a pixel reconstructor is employed to obtain clean images.

Currently, an increasing number of studies are focusing on developing general multi-weather restoration models. However, a significant portion of research [20]–[23] still emphasizes directly shaping the unified representation of images through learning, as shown in Fig. 1. These methods use supervised learning to fit various weather degradation by employing different strategies, thereby achieving the goal of handling multi-weather image degradation. Typically, they learn the explicit features of different degraded images, overlooking the implicit distribution behind the images. Moreover, the random distribution of weather information can lead to inaccuracies in model predictions. Given the randomness and uncertainty of weather data, these models require optimization over a

*Corresponding author.

Bodong Cheng, Guixu Zhang and Zhi Li are with the School of Computer Science and Technology, East China Normal University, Shanghai, China. (E-mail:52275901029@stu.ecnu.edu.cn, gxzhang@cs.ecnu.edu.cn, zli@cs.ecnu.edu.cn)

Juncheng Li and Jun Shi is with the School of Communication and Information Engineering, Shanghai University, Shanghai, China. (E-mail: junchengli@shu.edu.cn, junshi@shu.edu.cn)

Yingying Fang and Tieyong Zeng are with the Department of Mathematics, The Chinese University of Hong Kong, New Territories, Hong Kong. (E-mail: yyfang@math.cuhk.edu.hk, zeng@math.cuhk.edu.hk)

Yin Chen is with the Department of Cyberspace Security, Beijing Electronic Science and Technology Institute, Beijing, China. (E-mail:ychen@besti.edu.cn)

more extensive search space, which limits their generalization performance. Consequently, their performance is constrained when facing real weather degradation.

To capture the weather distribution within images, some methods [24], [25] introduced generated models into image restoration. Initially, Generative Adversarial Networks (GANs), recognized as efficient end-to-end generators, were employed for tasks such as image denoising and deblurring, yielding impressive results. With the advent of Denoising Diffusion Probabilistic Models (DDPMs), a considerable body of current research has begun integrating DDPMs into image restoration [26]–[29]. Undoubtedly, these endeavors have significantly advanced the field. However, these methods often confine themselves to a singular form of degradation, thereby limiting the outstanding distribution learning capabilities of DDPMs. To fully leverage the capabilities of DDPMs, we integrate them into a multitask framework, utilizing it as a data distribution generator to assist the recovery model in obtaining more universally applicable distribution information.

The crux of multi-task image blind restoration lies in precisely modelling and extracting the distribution of the images, and embedding distribution information into the network to guide the restoration process. The inherent challenge of this concept is finding a suitable approach capable of accurately mapping complex weather in images to corresponding distributions. To achieve this, we have developed a multi-**Wea**ther distribution dif**FU**sion blind restoration model (WeaFU) that extracts distribution representations from which abstract mappings are reconstructed into a new pixel space (Fig. 1(b)). Our method offers two significant advantages. Firstly, the diffusion process expands the search space of distribution representations, mitigating unnecessary disturbances in the original pixel space. Additionally, learned representations will yield dense data mappings, enabling the utilization of distribution information to avoid computations involving sparse tensors and enhance overall efficiency. Therefore, compared to direct pixel-to-pixel restoration methods, distribution information endows the model with a high level of robustness.

- We design a Diffusion Distribution Generator (DDG) to perceive the data distribution in diverse weather. This efficient generator compresses sparse pixel information from the image space into a denser high-dimensional space. This process yields more generalizable features, thereby enhancing the robustness of the multi-weather model.
- We construct a Conditional Distribution-Aware Transformer (CDAT). This transformer utilizes the data distribution as a condition to guide the model towards more accurate results.
- We propose a novel multi-**Wea**ther distribution dif**Fu**sion blind restoration model (WeaFU), which creatively employs diffusion model to perceive corresponding weather distributions to address various weather recovery scenarios. Extensive experiments demonstrate that WeaFU achieves optimal results in multi-weather blind restoration, surpassing the performance of most specialized methods.

II. Related Works

A. Single Weather Degradation Removal

Single degradation removal aims to reconstruct clean images from specific weather. For rain removal, [30] introduced a Progressive Recurrent Network (PReNet), leveraging recurrent layers for learning feature dependency. [31] proposed a Multi-Stage Progressive Reconstruction Network (MPRNet), using a competitive objective to co-design spatial details and high-level contextual information. This architecture progressively learns and decomposes the restoration process, achieving significant improvements across tasks with per-pixel adaptive designs and information exchange mechanisms at each stage. [32] introduced a compact learning rain removal transformer, emphasizing task invariance and knowledge transfer. For snow removal, DesnowNet [33] is a multi-level network addressing translucent and opaque snow particle removal, while SM-GARN [16] uses a snow mask-guided adaptive residual network. For haze removal, FFA-Net [34] is an end-to-end feature fusion attention network designed for haze removal. It utilizes feature attention modules and attention-based fusion structures at different levels. AECR-Net [35] employs contrastive learning for efficient fog removal, balancing performance, and memory usage through a compact class autoencoder (AE) framework.

These methods perform well for individual tasks, but they have not yet achieved in unified restoration in different weather. Thus, in constructing a unified model, the constrained generalization of these methods impedes their performance in multi-weather scenarios.

B. Multi Degradation Removal

Various weather phenomena introduce distinct forms of interference to images, each exhibiting unique visual characteristics. Despite these apparent differences, these phenomena share commonalities. For instance, the impact of rain streaks, snow stripes, and snowflakes on images shows similarities. Moreover, in rainy and snowy weather, the inevitable occurrence of haze also affects images. Building upon these observations, several studies [20]–[23], [36] have conducted in-depth research in the field of image restoration under various weather.

[20] proposed a model named All-in-One, capable of addressing image degradation caused by various severe weather conditions. Through neural architecture search optimization of the generator and employing a novel adversarial learning scheme, it achieves competitiveness in the restoration of images under different types of severe weather conditions. [22] tackles the ill-posedness of the problem of removing multiple severe weather effects by introducing a knowledge distillation learning mechanism. [21] introduced an end-to-end model called TransWeather, based on the Transformer architecture, to efficiently handle image restoration under different weather conditions through an encoder and a decoder. [23] designed an efficient unified framework by discovering that distorted images under different weather conditions contain general and specific features.

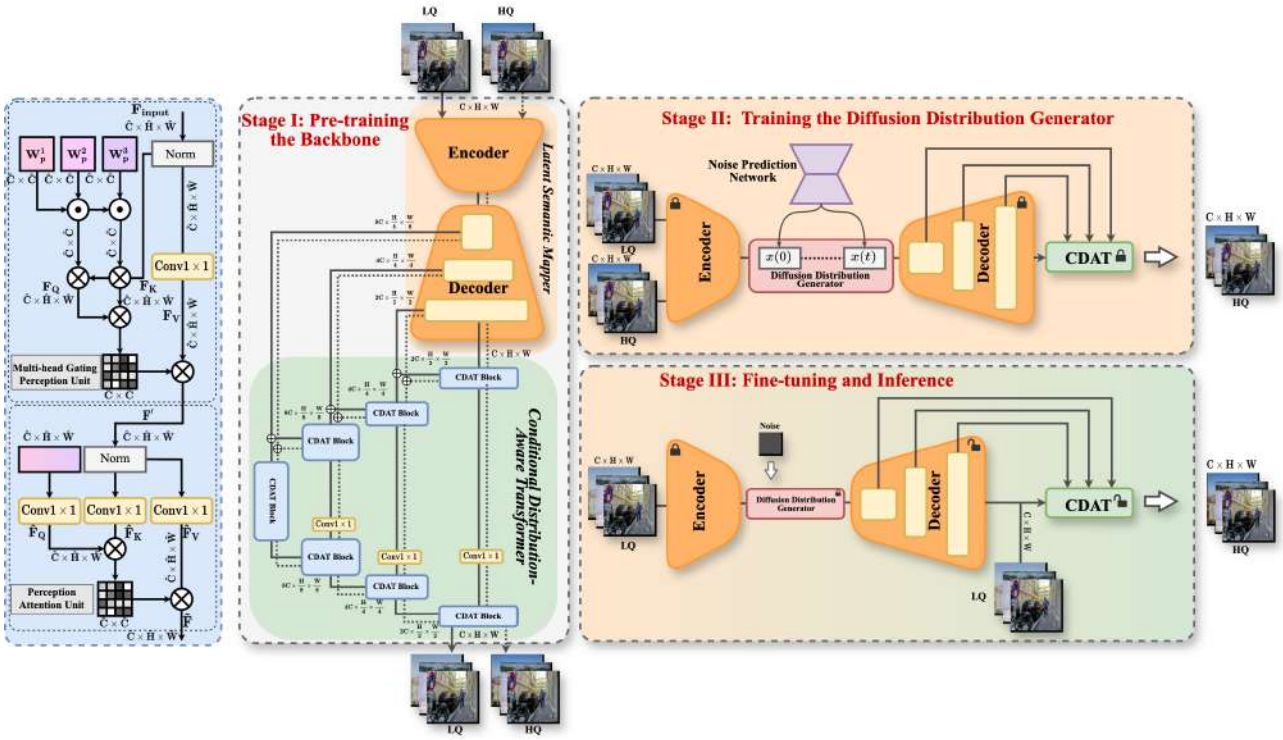

Fig. 2: WeaFU is composed of three main components: the Latent Semantic Mapper (LSM), the Diffusion Distribution Generator (DDG), and the Conditional Distribution-Aware Transformer (CDAT). Training is carried out in three stages. Stage I: Low-quality (LQ) and high-quality (HQ) images are simultaneously input to train the LSM, mapping the images into the feature space. During this stage, the CDAT is also pre-trained. Stage II: The Noise Prediction Network within the DDG is trained to effectively sample distribution representations across multi-weather scenarios. Stage III: The decoder part of the LSM and the entire CDAT are fine-tuned, ensuring accurate reconstruction of the generic features sampled by the DDG into the new pixel space for clean image generation.

These methods focus on pixel-space restoration and overlook deeper weather-informed distributions, limiting the search space of the model. Although [22] considers distributions across different weather, it only integrates multiple methods through knowledge distillation, neglecting further exploring distributions among distinct weather. Moreover, this integration introduces model complexity.

C. Diffusion Model for Restoration

Significant advancements in diffusion models for AI-Generated Content (AIGC) have led to their integration into image restoration tasks, outperforming previous GAN-based methods. For instance, [26] introduced SR3, an image super-resolution method based on DDPMs, employing stochastic iterative denoising and U-Net architecture. Another approach, WeatherDiff [27] utilizes a patch-based diffusion model for image restoration under adverse weather conditions, independent of image size. Despite the advanced performance of diffusion models, their inefficiency in estimating entire images during image restoration (IR) has been addressed by [28] with the introduction of the efficient DiffIR. Luo et al. [29]proposed a Stochastic Differential Equation (SDE) method applicable to general image restoration, degrading HQ images and recovering LQ images through inverse SDE.

However, these methods only apply DDPMs to single-task scenarios and do not consider multi-task settings. Therefore, we believe this limits the exceptional distribution modelling capabilities of Diffusion. To fully unleash the potential of the diffusion model, we aim to construct a multi-weather restoration model with distribution-aware capabilities.

III. Method

A. Overview

As shown in Fig. 2, WeaFU primarily consists of the Latent Semantic Mapper (LSM), the Diffusion Distribution Generator (DDG), and the Conditional Distribution-Aware Transformer (CDAT). Additionally, we divide the training process into three stages: Stage I involves representation learning training of the LSM and pixel reconstructor (CDAT), mapping HQ and LQ images to the feature space and enabling the network to acquire fundamental reconstruction capabilities; Stage II focuses on training the DDG; Stage III fine-tunes the decoder of the LSM and CDAT to adapt to the distribution of the generator. Define $X = [X_{rain}, X_{snow}, X_{haze}]$ as the LQ image, $Y = [Y_{rain}, Y_{snow}, Y_{haze}]$ represents the HQ image, and we use $rain$, $snow$, and $haze$ as subscripts to distinguish different weather conditions.

B. Stage I: Pre-training the Backbone

The training in Stage I has two objectives: training the Latent Semantic Mapper (LSM) and the pixel reconstructor (CDAT), aimed at compressing the image space and capturing distribution information.

Latent Semantic Mapper. WeaFU maps pixels from the image space to the feature space through the previously trained LSM. This form of feature carries denser information, allowing the model to understand the data distribution better. Furthermore, it contains rich semantic information to guide pixel-level reconstruction. We pre-train this mapper using a representation-learning approach that leverages the architecture of U-Net to compress the images. As illustrated in Fig. 2, this process involves parallel operations on LQ images of different weather and their corresponding HQ images. In order to more effectively compress and restore the image space, we conducted a cross-operation between the representations and underlying features of LQ and HQ at the deepest layer of the network:

$$X_l, X_h = LSM_{encoder}(X), \tag{1}$$

$$Y_l, Y_h = LSM_{encoder}(Y), \tag{2}$$

$$X' = LSM_{decoder}(X_l, X_h), \tag{3}$$

$$Y' = LSM_{decoder}(Y_l, X_h), \tag{4}$$

where the latent and hidden representations of the LQ image output by the encoder are denoted as X_l and X_h, respectively, while Y_l and Y_h represent the latent and hidden representations of the HQ image. It is important to emphasize that while our primary goal is to obtain the encoder ($LSM_{encoder}$) within the LSM, the pre-trained decoder ($LSM_{decoder}$) will perform feature alignment in the subsequent pixel reconstruction.

Conditional Distribution-Aware Transformer. Precisely reconstructing abstract distributions into specific pixels is a pivotal concern for WeaFU. To address this, we introduce a lightweight Conditional Distribution-Aware Transformer (CDAT) to map distribution information into the image space.

Given the image distribution generated by DDG as x', to reconstruct HQ images, it is necessary to satisfy:

$$p(\hat{x}) = p_\phi(\hat{x}|x, x'), \tag{5}$$

where p_ϕ represents the pixel reconstruction pipeline, x is the LQ image, and $\hat{x}$ is the reconstructed clean image. Previous research has mainly focused on x or derived features of x (such as color, texture, luminance channels, etc., as prior information). In contrast, this work embeds weather distribution information as the conditional control in the pixel reconstruction process. Specifically, CDAT consists of multiple CDAT blocks, each containing a Multi-Head Gated Perception Unit (MGPU) and a Perception Attention Unit (PAU).

Multi-head Gating Perception Unit: In the Transformer, Q, K, and V represent abstract concepts of the query mechanism in database retrieval. In the design of the pixel reconstructor, we extend this idea and propose a Multi-Head Gating Perception Unit (MGPU). MGPU introduces three proposed matrices, W_p^1, W_p^2, and $W_p^3 \in \mathcal{R}^{\hat{C}\times\hat{C}}$. Through these three matrices, we can obtain the Query $F_Q \in \mathcal{R}^{\hat{C}\times\hat{H}\times\hat{W}}$ and the Key $F_K \in \mathcal{R}^{\hat{C}\times\hat{H}\times\hat{W}}$:

$$F_Q = (W_p^1 \odot W_p^2) \otimes Norm(F_{input}), \tag{6}$$

$$F_K = (W_p^1 \odot W_p^2 \odot W_p^3) \otimes Norm(F_{input}), \tag{7}$$

where $\odot$ indicates element-wise multiplication and $\otimes$ indicates matrix multiplication. Through the mainstream, we can obtain the Value $F_V \in \mathcal{R}^{\hat{C}\times\hat{H}\times\hat{W}}$:

$$F_V = Conv1(F_{input}). \tag{8}$$

The overall process of MGPU can be described as follows:

$$F' = Softmax\left(F_Q \otimes F_K^T\right) F_V. \tag{9}$$

Perception Attention Unit: We use a unified learnable query $\hat{F}_Q \in \mathcal{R}^{\hat{C}\times\hat{H}\times\hat{W}}$ as the embedding for attention, and the additional Key $\hat{F}_K$ and Value $\hat{F}_V$ can be obtained through the mapping of F':

$$\hat{F}_K = Conv1_k(Norm(F')), \tag{10}$$

$$\hat{F}_V = Conv1_v(Norm(F')). \tag{11}$$

Subsequently, we employ an attention mechanism across the channel dimensions to process the features. The process of Perception Attention Unit (PGU) can be described as:

$$\tilde{F} = Softmax\left(\hat{F}_Q \otimes \hat{F}_K^T\right) \hat{F}_V. \tag{12}$$

In this stage, we adopt the $\mathcal{L}_1$ loss function for pre-training, which is used to guide LSM to learn the representation of images:

$$\mathcal{L}_{rec} = \|X' - X\|_1, \tag{13}$$

$$\mathcal{L}_{res} = \|Y' - Y\|_1, \tag{14}$$

where $\mathcal{L}_{rec}$ and $\mathcal{L}_{res}$ represent the reconstruction loss and restoration loss, respectively. The purpose of $\mathcal{L}_{rec}$ is to guide the encoder to generate representations of LQ images, while $\mathcal{L}_{res}$ aims to help the decoder learn the degradation information within LQ images to reconstruct images close to HQ images. Additionally, we impose constraints at the distribution level of the representations, where $E()$ represents the mean and $Std()$ represents the standard deviation.

$$\mathcal{L}_{dis} = \|E(X') - E(X)\|_1 + \|Std(X') - Std(X)\|_1, \tag{15}$$

Finally, we constrain the output of CDAT with $\mathcal{L}_{acc}$ to ensure the accuracy of image reconstruction:

$$\mathcal{L}_{acc} = \|CDAT(Y', X) - Y\|_1. \tag{16}$$

In summary, the objective function to be optimized in Stage I is:

$$\mathcal{L} = \mathcal{L}_{rec} + \mathcal{L}_{res} + \mathcal{L}_{dis} + \mathcal{L}_{acc}. \tag{17}$$

C. Stage II: Training the Diffusion Distribution Generator

WeaFU differs from previous methods by introducing the idea of generating distribution representations in multi-weather scenarios. We propose a Diffusion Distribution Generator (DDG), which perceives the latent feature extracted by the mapper using a diffusion process. To further enhance its learning capacity for distribution representations, we employ the simpler IR-SED [29]. The method can achieve high-quality

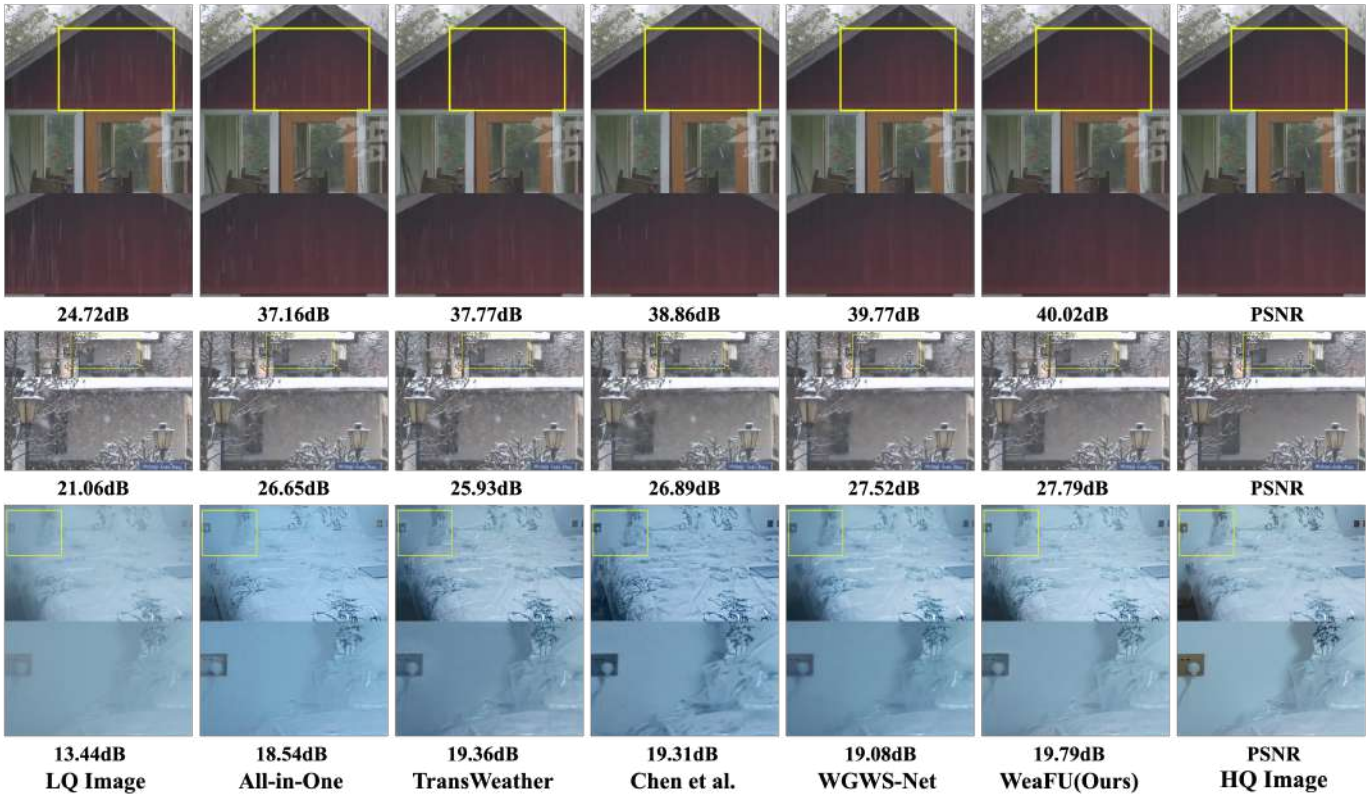

Fig. 3: Visual comparisons with other SOTA methods in real-world scenarios.

Algorithm 1 Inference Processing

Input: LQ Image X, implicit sampling steps T.

Parameter: Latent Semantic Mapper $LSM()_\theta$, Noise prediction Network $\epsilon()_\theta$, Conditional Distribution-Aware Transformer $CDAT()_\theta$.

Output: Weather-free image $\hat{X}$.

$X_l, X_h = LSM_{encoder}(X)_\theta$
for $t = 1...T$ **do**
  $X_{dist}(t) = X_l + \sigma_{max}\epsilon \sim N(0, I)$
  $dX_l = \left[\theta_t(\mu - X_l) - \sigma_t^2 \frac{\epsilon(X_{dist}(t), X_l)_\theta}{v_t}\right] dt + \sigma_t d\hat{w}$
  $X_{dist}(t-1) = X_{dist}(t) - dX_l$
end for
$X' = LSM_{decoder}(X_{dist}(0), X_h)_\theta$
$\hat{X} = CDAT(X', X)_\theta$
return $\hat{X}$

distributions using shorter processes. The forward process of IR-SDE is defined as:

$$dX = \theta_t(\mu - X)dt + \sigma_t dw, \tag{18}$$

where θ_t and σ_t represent the mean-reversion speed and stochastic volatility, respectively. And μ is latent feature of LQ image. According to IR-SDE, when $\frac{\sigma_t^2}{\theta_t^2} = 2\lambda^2$ holds, the score in the reverse process can be solved:

$$dX = \left[\theta_t(\mu - X) - \sigma_t^2 \bigtriangledown_X log_{p_t}(X)\right] dt + \sigma_t d\hat{w}, \tag{19}$$

$$\bigtriangledown_X log_{p_t}(X) = -\frac{X(t) - m_t}{v_t}, \tag{20}$$

here, m_t and v_t are derived from the marginal distribution $p_t(X) \sim N(X(t)|m_t, v_t)$:

$$m_t := \mu + (X(0) - \mu)e^{\bar{\theta}_t}, \tag{21}$$

$$v_t := \lambda^2(1 - e^{-2\bar{\theta}_t}), \tag{22}$$

as $X(t)$ can be obtained through m_t and v_t, the $\bigtriangledown_X log_{p_t}(X)$ can be expressed as $-\frac{\epsilon_t}{v_t}$.

D. Stage III: Fine-tuning and Inference

The LSM and CDAT use the U-Net architecture to facilitate the pre-training. By summing the features at each level of these U-Nets, information from both is integrated, ensuring that CDAT can fully perceive the distribution information during reconstruction. However, since the DDG, which generates distribution information, is trained separately in Stage II, a divergence occurs between the original feature space and the pre-trained network. Therefore, in the third phase, fine-tuning focuses on adjusting the parameters of the decoders and CDAT. This process aims to realign the distribution information output by DDG and ensure the accurate reconstruction of image pixels.

As illustrated in Stage III of Fig. 2, we froze the LSM's encoder and the parameters within DDG. Simultaneously, we employ the $\mathcal{L}_1$ loss function to measure the difference between the reconstructed images and the originals, which guides updating the parameters in LSM's decoder and CDAT. This approach allows us to fine-tune the model to closely align with the output distribution of DDG, ensuring that the pixel

TABLE I: Quantitative comparisons in real-world scenarios on the SPA+, RealSnow, and REVIDE datasets. The best results are **highlighted**, and the second best results are underlined.

Methods	SPA+				RealSnow				REVIDE			
	PSNR↑	LPIPS↓	SSIM↑	IL-NIQE↓	PSNR↑	LPIPS↓	SSIM↑	IL-NIQE↓	PSNR↑	LPIPS↓	SSIM↑	IL-NIQE↓
Specific Methods												
DeSnowNet [33]	-	-	-	-	28.52	0.258	0.8952	27.16	-	-	-	-
SMGARN [16]	-	-	-	-	29.27	0.221	0.8994	26.45	-	-	-	-
PReNet [30]	37.30	0.124	0.9631	24.67	-	-	-	-	-	-	-	-
MPRNet [31]	39.02	0.086	0.9803	23.89	-	-	-	-	20.81	0.513	0.8453	43.91
ESTINet [37]	38.73	0.093	0.9823	23.50	-	-	-	-	-	-	-	-
GridDehazeNet [38]	-	-	-	-	-	-	-	-	19.40	0.682	0.8421	45.34
FFA-Net [34]	-	-	-	-	-	-	-	-	**21.19**	**0.411**	0.8504	43.78
Mb-TaylorFormer [39]	-	-	-	-	-	-	-	-	20.78	0.510	0.8501	43.90
SwinIR [40]	38.64	0.089	0.9776	24.12	28.93	0.210	0.9067	26.12	19.67	0.692	0.8502	45.01
Uformer [41]	37.58	0.120	0.9759	25.23	28.77	0.237	0.9073	26.45	18.60	0.808	0.8142	46.45
Restormer [42]	38.87	0.084	0.9820	24.89	29.41	0.216	0.8787	25.45	19.86	0.599	0.8316	44.82
MambaIR [43]	38.42	0.083	0.9835	23.75	29.36	0.243	0.8930	25.91	20.61	0.527	0.8461	44.31
GridFormer [44]	38.92	0.089	0.9814	22.85	29.47	0.223	0.8986	25.23	20.73	0.503	0.8482	43.81
Diffusion-Based Methods												
Refusion [45]	36.40	0.208	0.9613	25.45	28.04	0.496	0.8834	28.65	19.70	0.608	0.8210	45.12
WeatherDiff [27]	38.57	0.088	0.9824	23.78	29.44	0.275	0.9078	25.23	19.65	0.629	0.8289	44.23
DiffIR [28]	**39.08**	0.082	**0.9873**	22.45	**29.63**	**0.172**	0.8907	24.89	20.39	0.547	0.8476	44.67
Multi-Weather Methods												
All-in-One [20]	34.78	0.161	0.9436	26.12	29.28	0.253	0.9048	25.89	20.24	0.570	0.8256	44.56
TransWeather [21]	33.64	0.168	0.9258	27.45	29.16	0.248	0.8962	26.23	20.31	0.532	0.8297	44.40
Chen et al. [22]	38.15	0.092	0.9744	24.78	29.35	0.221	0.8983	25.56	20.47	0.516	0.8372	44.32
WGWS-Net [23]	38.94	0.086	0.9823	24.45	29.46	0.216	0.9083	25.12	20.56	0.507	0.8311	44.16
WeaFU (Ours)	39.06	**0.071**	0.9844	**22.12**	29.58	0.208	**0.9094**	**24.51**	20.82	0.498	**0.8592**	**43.56**

values generated by the decoder closely resemble those of the original image. Additionally, we introduce the LQ image as a conditional factor for the pixel reconstructor. This enables the network to preserve the fundamental pixel distribution of the image, thereby enhancing the quality of the reconstructed image. Upon completing Stage III, the parameters of these two components remain frozen for inference. The entire inference process of the model is detailed in Algorithm 1.

IV. Experiment

A. Implementation

WeaFU consists of three main components: the Latent Semantic Mapper, the Diffusion Distribution Generator, and the Conditional Distribution-Aware Transformer. The Latent Semantic Mapper adopts a 4-level down-sampling strategy, with the number of residual units per layer set as $[4, 8, 8, 16]$. The Conditional Distribution-Aware Transformer follows a similar 4-level design, with the number of units per layer specified as $[1, 1, 2, 4]$ and an embedding dimension of 48.

For the Diffusion Distribution Generator, the batch size is 8, and the training block size is 128×128 pixels. We utilize the AdamW optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The initial learning rate is set to 3×10^{-5} and is decayed to $1e^{-7}$ using a cosine scheduler. The noise level is fixed at 50 for all tasks, and the diffusion denoising steps are set to 100.

B. Datasets

We evaluated different datasets to validate the effectiveness of WeaFU on multi-weather images. The training data is divided into two groups: a composite dataset from real-world images (SPA+ [23], RealSnow [23], and REVIDE [46]) and a composite dataset from synthetic images (Rain1400 [47], Rain800 [48], Rain200 [49], Snow100K [33], and RESIDE [50]).

C. Compare with SOTA Methods

For evaluating the performance of WeaFU, we employ several metrics, including PSNR (Peak Signal-to-Noise Ratio), LPIPS (Learned Perceptual Image Patch Similarity), SSIM (Structural Similarity Index Measure), and IL-NIQE (Integrated Local Natural Image Quality Evaluator).

TABLE I compares WeaFU with other state-of-the-art (SOTA) models on the real-world datasets. As can be observed, WeaFU achieves competitive results on all datasets. While DiffIR demonstrates slightly higher PSNR among diffusion methods, WeaFU excels in SSIM and IL-NIQE, highlighting its robust performance in multi-weather restoration tasks through unified training. In addition, compared with the multi-weather methods, WeaFU similarly yields competitive scores, indicating the higher robustness of our proposal towards different scenarios.

To validate the robustness of WeaFU, we created a test dataset (**MultiWeather**) by randomly sampling subsets of degraded images containing rain, snow, and haze from various datasets to evaluate the ability of general models to handle unknown weather types. According to Fig. 6, we can observe that our WeaFU outperforms other methods significantly under multi-weather conditions. Despite being a recently proposed general method, the performance of WGWS-Net [23] notably degrades in unknown weather, further highlighting the robustness of WeaFU in multi-weather tasks.

D. Comparison on Synthetic Datasets

We further evaluate the reconstruction performance of WeaFU on synthetic datasets to emphasize its robustness in diverse weather scenarios.

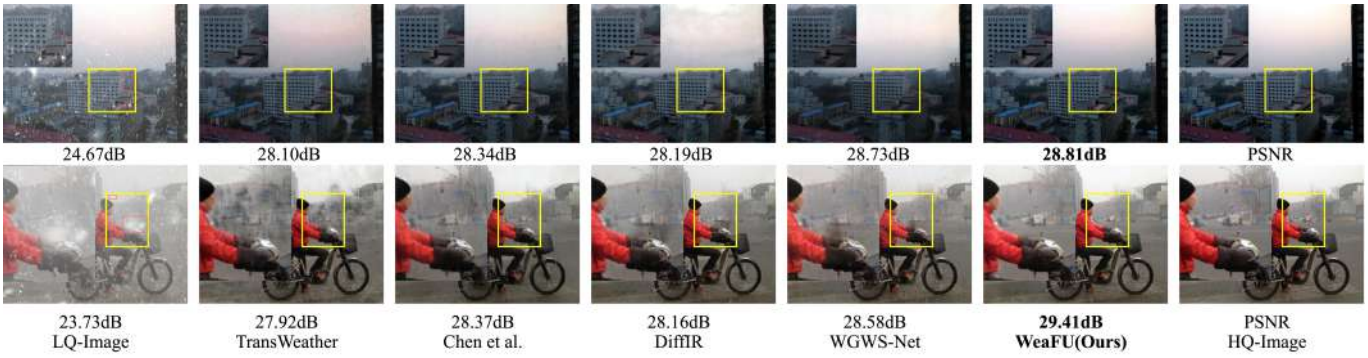

Fig. 4: Visual comparisons of SOTA models on the CSD dataset. The red boxes of LQ image depict varying snowflake sizes and dense rain streaks. From a global perspective, the image also exhibits vailing effects.

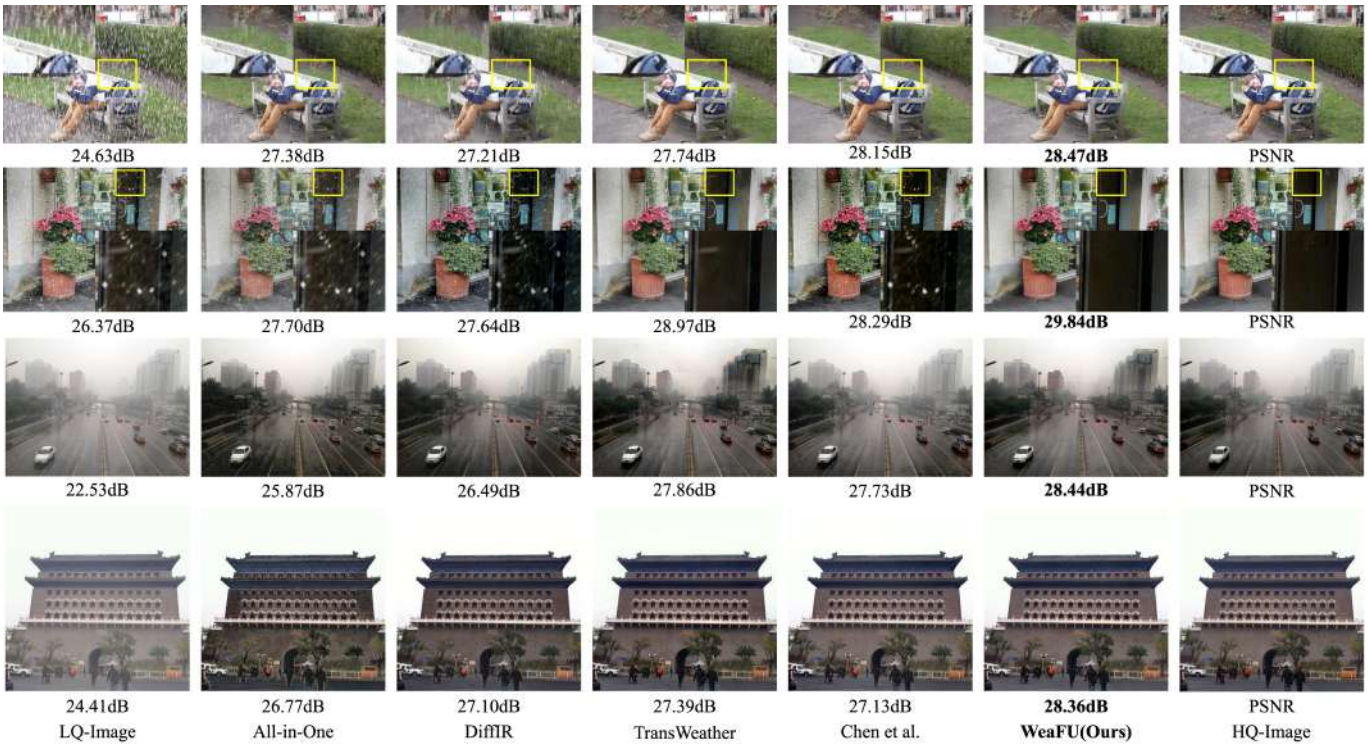

Fig. 5: Visualization assessment of the model in unknown weather restoration on the MultiWeather dataset.

TABLE II compares the performance of our method with existing methods on the RESIDE dataset. WeaFU achieves 29.37 PSNR, 0.236 LPIPS and 22.80 NIQE, which outperforms previous state-of-the-art methods, indicating its exceptional ability to recover clear images from the haze. Additionally, WeaFU achieves 0.8962 SSIM, which is still a decent result (0.8975 SSIM achieved by Restormer [42]).

In TABLE III, we compare WeaFU with state-of-the-art restoration models on Rain200 (L and H), Rain800, and Rain1400. In contrast to the previous methods, our method achieves competitive results on all rainy datasets, demonstrating its robust performance in deraining tasks under different rainfall. Specifically, WeaFU achieves 37.65 PSNR on Rain200-L, the highest among all methods, indicating its exceptional ability to restore clear images from rainy scenes. It can be observed that MPRNet achieves excellent results for SSIM and IL-NIQE on the Rain200 dataset. However, they usually generalize poorly to other datasets.

TABLE IV demonstrates the quantitative results of our proposed method compared with specific methods and multi-weather restoration methods. One can observe that our network, while being trained for multiple bad weather types, achieves competitive performance compared with specific methods. WeaFU achieves the best PSNR, demonstrating its ability to restore clear images from snowy scenes. WeaFU also achieves the best LPIPS and SSIM, indicating its effectiveness in preserving image quality. Furthermore, WeaFU attains 18.12, 20.34 and 23.23 IL-NIQE. Such results demonstrate exceptional performance under heavy snow.

We also provide a comparative analysis of this experiment from a visual perspective. Fig. 4 showcases the results of different methods on the CSD dataset. It is worth emphasizing that the CSD dataset integrates diverse weather degradation, providing a robust assessment of the comprehensive performance of the methods. To further evaluate the overall performance of WeaFU, we also include the single-task method

TABLE II: Quantitative comparisons of image dehazing on the RESIDE dataset. The best results are **highlighted**, and the second best results are underlined.

Methods	RESIDE			
	PSNR↑	LPIPS↓	SSIM↑	IL-NIQE↓
Specific Methods				
AOD-Net [51]	27.50	0.653	0.8569	26.78
GridDehazeNet [38]	29.01	0.315	0.8801	23.96
FFA-Net [34]	29.27	0.281	0.8736	23.84
MPRNet [31]	28.69	0.411	0.8822	24.12
SwinIR [40]	28.92	0.353	0.8795	23.89
Restormer [42]	29.36	0.240	**0.8975**	23.13
Uformer [41]	28.77	0.376	0.8705	24.34
MambaIR [43]	28.73	0.381	0.8752	24.39
Mb-TaylorFormer [39]	28.92	0.346	0.8771	24.05
GridFormer [44]	29.11	0.293	0.8864	23.92
Multi-Weather Methods				
All-in-One [20]	28.29	0.483	0.8823	23.45
TransWeather [21]	28.43	0.452	0.8750	24.12
Chen et al. [22]	28.35	0.457	0.8904	23.67
WSWG-Net [23]	28.82	0.384	0.8815	23.34
WeaFU (Ours)	**29.37**	**0.236**	0.8962	**22.80**

DiffIR [28], based on the diffusion model, for comparison. Upon closer inspection of the region inside the yellow box after local magnification, it can be observed that WeaFU exhibits fewer rain/snow artifacts. While TransWeather [21] can eliminate more rain and snow textures, it fails to remove haze in the image, resulting in uneven, patchy artifacts. In contrast, the other three methods can eliminate these artifacts but at the cost of generating more residual raindrops or snow streaks. Additionally, results in Fig. 5 reflect the outcomes on the MultiWeather dataset, which randomly samples images with different weather degradation to assess the capability of the model in handling unknown weather. Firstly, compared to the single-task image restoration model DiffIR [28], WeaFU demonstrates more robust general restoration capabilities. The reconstructed weather-free images exhibit less residual weather degradation. Compared to other general methods, WeaFU produces more accurate image details and shows no residues in the edge regions.

Overall, these results highlight the outstanding performance of WeaFU in deraining (Rain200, Rain800, and Rain1400 datasets), desnowing (Snow100K dataset), and dehazing (RESIDE dataset). They further validate the exceptional performance of WeaFU in diverse weather scenarios. Notably, WeaFU consistently outperforms state-of-the-art unified and specific methods, emphasizing its robustness in addressing various weather-related challenges.

E. Effectiveness in Multi-Weather and Mixed-Weather Conditions

In addition, we integrated methods designed for specific weather conditions into a comprehensive model to address mixed-weather degradation in CSD [12]. In other words, in CSD, each image contains multiple different weather conditions simultaneously, and they are randomly distributed on the image. This greatly tests the robustness of the

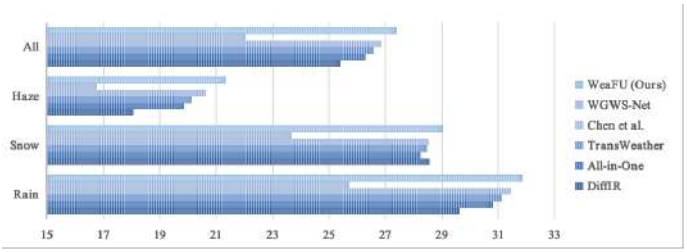

Fig. 6: PSNR comparisons on the MultiWeather dataset.

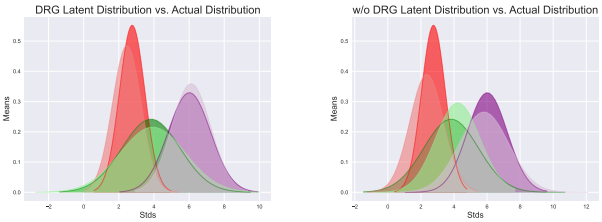

Fig. 7: Quantitative analysis of the influence of DDG on representation distribution. In this context, the light-colored region signifies the distribution generated by the model, whereas the dark-colored region represents the original data distribution. The distribution in Fig. 1 is obtained by computing the mean and variance of feature maps.

model. As shown in Table VI, the performance of the integrated and multi-weather models significantly deteriorated in mixed-weather tasks, while our WeaFU still performed excellently. Table VII includes the evaluation results for the RainKITTI [55] and JRSRD [56] datasets. It is evident that our method has achieved the best results when compared with other multi-weather methods. This further demonstrates the excellence of WeaFU.

F. Ablation Study

In this part, we validate the effectiveness of each component in WeaFU. For detailed information, please refer to TABLE V, where $WeaFU_{base}$ denotes our WeaFU, serving as a control reference.

Study of the Diffusion Distribution Generator. As the core of our method, DDG utilizes the diffusion mechanism to perceive the distribution information in the image space, thereby enhancing restoration performance across diverse weather. We remove DDG from the model, resulting in $WeaFU_1$. Clearly, the comprehensive performance of $WeaFU_{base}$ is significantly superior to that of $WeaFU_1$ across various weather conditions. This indicates that DDG contributes to enhancing the overall robustness of the model across diverse weather.

We also conduct a quantitative analysis of the features generated by DDG to validate the accuracy of generating distribution representations when matching sample distributions, as illustrated in Fig. 7. Clearly, under the guidance of DDG, the model can better fit data distributions under different weather conditions. This further confirms the rationality and feasibility of WeaFU in leveraging distribution information.

Study of the Conditional Distribution-Aware Transformer. The role of CDAT in WeaFU is to remap the obtained distribution information back to the pixel space without being affected by weather. We remove CDAT from the model,

TABLE III: Quantitative comparisons of image deraining on Rain200 (-L and -H), Rain800, and Rain1400 datasets. The best results are **highlighted**, and the second best results are underlined.

Methods	Rain200-L				Rain200-H				Rain800				Rain1400			
	PSNR↑	LPIPS↓	SSIM↑	IL-NIQE↓	PSNR↑	LPIPS↓	SSIM↑	IL-NIQE↓	PSNR↑	LPIPS↓	SSIM↑	IL-NIQE↓	PSNR↑	LPIPS↓	SSIM↑	IL-NIQE↓
Specific Methods																
RESCAN [52]	28.79	0.331	0.8720	27.67	26.54	0.594	0.8545	28.13	25.81	0.358	0.8634	30.72	29.81	0.254	0.8898	25.58
PReNet [30]	33.66	0.175	0.8831	23.45	29.04	0.320	0.8703	26.89	28.14	0.244	0.8687	28.73	31.76	0.176	0.8976	24.90
SwinIR [40]	34.52	0.166	0.9130	22.56	29.23	0.306	0.8799	25.68	28.03	0.263	0.8729	28.15	32.88	0.155	0.9010	24.45
Restormer [42]	34.87	0.143	0.9566	22.34	29.82	0.265	0.8921	24.89	28.11	0.277	0.8773	27.98	32.43	0.159	0.8991	24.67
Uftormer [41]	33.10	0.184	0.8895	23.78	28.56	0.278	0.8697	27.23	27.47	0.306	0.8643	29.11	31.89	0.183	0.8964	25.23
MPRNet [31]	37.41	0.116	**0.9835**	21.89	30.27	0.176	0.8942	**24.43**	28.57	0.228	0.8791	27.67	32.96	0.146	0.9034	23.98
MAXIM [53]	36.28	0.127	0.9724	22.23	29.79	0.254	0.8909	25.01	28.33	0.265	0.8785	28.56	32.81	0.151	0.9020	24.46
MambaIR [43]	36.93	0.125	0.9753	22.12	30.08	0.205	0.8930	24.71	28.45	0.235	0.8788	27.84	32.91	0.150	0.9018	24.41
ESTINet [37]	37.34	0.120	0.9772	22.01	30.22	0.182	0.8925	24.86	28.49	0.230	0.8726	27.72	32.90	0.150	0.9013	24.35
GridFormer [44]	37.39	0.118	0.9736	22.04	30.25	0.181	0.8905	24.92	28.52	0.236	0.8762	27.79	32.81	0.148	0.9032	24.05
Multi-Weather Methods																
All-in-One [20]	33.13	0.192	0.9157	23.87	29.11	0.331	0.8672	25.78	27.59	0.282	0.8661	28.99	32.22	0.162	0.8982	24.89
TransWeather [21]	33.30	0.189	0.9420	23.56	28.92	0.348	0.8609	26.34	27.40	0.295	0.8635	29.12	32.08	0.166	0.8970	25.04
Chen et al. [22]	37.06	0.124	0.9621	22.27	29.35	0.295	0.8823	25.78	27.45	0.286	0.8679	29.34	32.34	0.157	0.8998	24.54
WGWS-Net [23]	35.46	0.138	0.9456	22.14	29.74	0.220	0.8754	25.45	27.97	0.254	0.8702	28.78	32.67	0.164	0.9002	24.58
WeaFU (Ours)	**37.65**	**0.108**	0.9792	**21.56**	**30.58**	**0.156**	**0.8958**	24.67	**28.84**	**0.227**	**0.8803**	**27.34**	**33.16**	**0.122**	**0.9045**	**23.78**

TABLE IV: Quantitative comparisons of image desnowing on Snow100K (-S, -M, and -L) datasets. The best results are **highlighted**, and the second best results are underlined.

Methods	Snow100K-S				Snow100K-M				Snow100K-L			
	PSNR↑	LPIPS↓	SSIM↑	IL-NIQE↓	PSNR↑	LPIPS↓	SSIM↑	IL-NIQE↓	PSNR↑	LPIPS↓	SSIM↑	IL-NIQE↓
Specific Methods												
DesnowNet [33]	32.23	0.277	0.9500	20.48	30.68	0.407	0.9035	22.95	28.79	0.528	0.8823	25.34
SwinIR [40]	33.27	0.214	0.9521	19.56	31.44	0.366	0.9409	21.77	29.28	0.473	0.8947	24.12
Restormer [42]	33.89	0.163	0.9534	18.73	32.42	0.265	0.9301	21.10	29.36	0.456	0.8989	23.48
Uformer [41]	31.04	0.335	0.9356	20.52	31.07	0.381	0.9134	22.89	28.76	0.501	0.8901	25.67
SMGARN [16]	34.16	0.153	0.9610	18.56	32.47	0.246	0.9434	20.67	29.45	0.438	0.9178	23.45
DDMSNET [54]	31.46	0.318	0.9276	20.34	31.35	0.370	0.9268	22.61	28.85	0.498	0.8834	25.11
MambaIR [43]	34.62	0.148	0.9638	18.42	32.50	0.241	0.9455	20.55	29.47	0.435	0.9185	23.41
GridFormer [44]	34.05	0.178	0.9550	19.20	32.11	0.280	0.9350	21.40	29.27	0.456	0.9060	23.89
Multi-Weather Methods												
All-in-One [20]	32.73	0.243	0.9512	19.01	31.82	0.339	0.9198	21.74	28.99	0.476	0.8843	24.12
TransWeather [21]	32.60	0.247	0.9598	19.45	31.22	0.378	0.9073	21.88	29.14	0.460	0.8865	24.01
Chen et al. [22]	32.86	0.210	0.9445	19.23	31.68	0.354	0.9101	21.55	29.33	0.443	0.9089	23.56
WSWG-Net [23]	33.56	0.236	0.9528	19.45	31.48	0.360	0.9289	21.68	29.06	0.461	0.8954	24.12
WeaFU (Ours)	**34.77**	**0.139**	**0.9723**	**18.12**	**32.63**	**0.239**	**0.9456**	**20.34**	**29.49**	**0.431**	**0.9198**	**23.23**

resulting in $WeaFU_2$. Obviously, in the absence of CDAT, WeaFU struggles to correspond distribution information with weather-free pixels. To further validate the effectiveness of the Multi-Head Gating Perception Unit (MGPU) and Perception Attention Unit (PAU) in CDAT, we replace these two units with CNN and the original attention structure, resulting in $WeaFU_{w/oPAG}$ and $WeaFU_{w/oMGPU}$. TABLE V shows that both sets of experimental configurations fail to surpass $WeaFU_{base}$. All of these studies demonstrate the effectiveness of the proposed model.

Study of Training Strategy In order to explore a more effective training strategy, we compare end-to-end optimization with the optimization employed by WeaFU. End-to-end optimization allows the entire model to be trained in a single pass, optimizing from input to output. This facilitates the direct learning of complex mappings for the model, eliminating the need for manually designing intermediate representations or features. However, this proves detrimental for WeaFU, which possesses diffusion generation capabilities. Due to the indirect learning approach of WeaFU , which involves sampling the distribution of images rather than directly learning from the images, the training process occurs independently of the mainstream model. As a result, end-to-end optimization methods are not applicable. Additionally, the staged training proves effective in controlling the information flow through the intermediate layers of the model, providing valuable assistance in guiding model learning.

In the TABLE VIII, $WeaFU_{base}$ and $WeaFU_{e2e}$ contrast the final performance of the two training strategies, where $WeaFU_{base}$ represents the segmented optimization strategy employed by our WeaFU, and $WeaFU_{e2e}$ represents the end-to-end optimization strategy. Clearly, the performance of $WeaFU_{e2e}$ is significantly inferior to $WeaFU_{base}$, demonstrating the applicability of segmented optimization for dif-

TABLE V: Quantitative comparisons of different components within WeaFU across multi-weather conditions (**PSNR** and **LPIPS**).

Method	DDG	CDAT	CDAT-MGPU	CDAT-PAG	Rain		Snow		Haze	
					PSNR↑	LPIPS↓	PSNR↑	LPIPS↓	PSNR↑	LPIPS↓
$WeaFU_{base}$	✔	✔	✔	✔	**39.06**	**0.071**	**29.58**	**0.208**	**20.82**	**0.498**
$WeaFU_1$	✘	✔	✔	✔	38.82	0.092	29.14	0.274	20.36	0.520
$WeaFU_2$	✔	✘	✘	✘	37.66	0.288	28.38	0.325	19.52	0.652
$WeaFU_{w/oPAG}$	✔	✔	✔	✘	38.77	0.106	29.30	0.259	20.55	0.501
$WeaFU_{w/oMGPU}$	✔	✔	✘	✔	38.69	0.112	29.27	0.261	20.60	0.506

TABLE VI: Quantitative comparisons of multi-task integration models and unified models on CSD dataset (**PSNR**). R, S, and H represent single-task methods for rain, snow, and haze removal, respectively. Among them, the best results are **highlighted** and the second best results are underline.

Type	Methods	CSD
Integrad Methods		
R+S	PReNet+SMGARN	29.25
	MPRNet+Restormer	29.11
S+H	Restormer+FFA-Net	30.19
	SMGARN+GridDehazeNet	30.38
H+R	FFA-Net+PReNet	28.54
	GridDehaze+MPRNet	28.26
R+S+H	PReNet+SMGARN+FFA-Net	30.45
	MPRNet+Restormer+GridDehazeNet	31.12
Multi-Weather Methods		
(R, S, H)	All-in-One [20]	29.83
	TransWeather [21]	30.15
	Chen et al. [22]	31.29
	WGWS-Net [23]	31.34
	WeaFU (Ours)	**32.87**

TABLE VII: Quantitative comparisons on the RainKITTI and JRSRD datasets. The best results are **highlighted**.

Methods	RainKITTI	JRSRD
All-in-One [20]	32.78	27.42
TransWeather [21]	33.71	27.83
Chen et al. [22]	34.62	28.54
WGWS-Net [23]	34.55	28.62
WeaFU (Ours)	**35.13**	**28.97**

TABLE VIII: Quantitative comparisons of different training strategies within WeaFU across multi-weather conditions (**PSNR**).

Models	Rain	Snow	Haze
$WeaFU_{e2e}$	35.42	27.53	18.76
$WeaFU_{S_1S_2}$	36.58	28.03	19.20
$WeaFU_{S_2S_3}$	37.16	28.39	19.25
$WeaFU_{base}$	**39.06**	**29.58**	**20.82**

fusion models. Additionally, we conducted combined experiments with Stage I, Stage II, and Stage III, where $WeaFU_{S_1S_2}$ involves joint optimization of Stage I and Stage II, and $WeaFU_{S_2S_3}$ involves joint optimization of Stage II and Stage III. As reflected in the results listed in TABLE VIII, end-to-end optimization, including diffusion models, severely impacts the performance of the model. Hence, we employ a segmented optimization strategy to train the proposed generated distribution mapping model.

Study of Hyperparameters In this part, we ablate image patch size, embedding dimension, and network depth, as shown in Table IX and X. Larger image patch sizes and embedding dimensions can continuously improve PSNR. Based on this, we set the patch size to 128×128, with an embedding dimension of 48. Additionally, as shown in Table X, PSNR also increases with the depth of the network. Therefore, we follow the setup of previous work and set the network depth for LSM and CDAT to 4 to achieve the best PSNR.

V. Conclusion

In this work, we addressed the fundamental challenge of image restoration under unknown weather conditions by emphasizing the precise modelling of image representation distributions. Specifically, we proposed a multi-**Wea**ther distribution dif**FU**sion blind restoration model (WeaFU), which introduces a novel idea creatively utilizing the diffusion-based weather information distribution for image restoration, enhancing robustness under various weather conditions. WeaFU employed a Latent Semantic Mapper (LSM) and Diffusion Distribution Generator (DDG) to perceive universal distributions by constraining sparse image features into a denser feature space. Additionally, the designed lightweight Conditional Distribution-Aware Transformer (CDAT) further enhanced the accuracy of restoration results. Extensive experiments demonstrated that WeaFU excels in multi-weather image restoration under various weather conditions, validating its outstanding robustness.

Acknowledgments

This work was supported in part by Grant NSFC/RGC N_CUHK 415/19, the National Natural Science Foundation of China (62301306, 62271203, 62202173, 62371190, 61971234), Grant ITF ITS/173/22FP, Grant RGC 14300219, 14302920, 14301121, CUHK Direct Grant for Research, in part by the Science and Technology Commission of Shanghai Municipality under Grants 23ZR1422200 and 23YF1412800.

References

[1] G.-P. Ji, D.-P. Fan, Y.-C. Chou, D. Dai, A. Liniger, and L. Van Gool, "Deep gradient learning for efficient camouflaged object detection," *Machine Intelligence Research*, vol. 20, no. 1, pp. 92–108, 2023.

[2] S. Chen, P. Sun, Y. Song, and P. Luo, "DiffusionDet: Diffusion model for object detection," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2023, pp. 19 830–19 843.

[3] Z. Chen, J. Li, T. Li, T. Fan, C. Meng, C. Li, J. Kang, L. Chai, Y. Hao, Y. Tang *et al.*, "A CRISPR/Cas12a-empowered surface plasmon resonance platform for rapid and specific diagnosis of the omicron variant of SARS-CoV-2," *National Science Review*, vol. 9, no. 8, p. 2200016, 2022.

[4] Z. Chen, C. Meng, X. Wang, J. Chen, J. Deng, T. Fan, L. Wang, H. Lin, H. Huang, S. Li *et al.*, "Ultrasensitive DNA origami plasmon sensor for accurate detection in circulating tumor DNAs," *Laser & Photonics Reviews*, p. 2400035, 2024.

TABLE IX: Quantitative comparisons of different patch sizes and embedding dimensions across multi-weather conditions (**PSNR**).

Patch Size				**Embedding Dimension**			
Patch size	Rain	Snow	Haze	Dimension	Rain	Snow	Haze
32	38.61	29.25	20.50	12	37.26	27.06	19.20
64	38.76	29.41	20.68	24	38.11	28.49	20.14
128	**39.06**	**29.58**	**20.82**	48	**39.06**	**29.58**	**20.82**

TABLE X: Quantitative comparisons of different network depths within LSM and CDAT across multi-weather conditions (**PSNR**).

LSM				**CDAT**			
Depth	Rain	Snow	Haze	Depth	Rain	Snow	Haze
1	37.23	27.32	19.08	1	36.97	27.32	18.08
2	38.14	28.69	19.73	2	37.42	28.25	19.23
3	38.62	29.11	20.36	3	38.63	28.81	20.37
4	**39.06**	**29.58**	**20.82**	4	**39.06**	**29.58**	**20.82**

[5] F. Zheng, Z. Chen, J. Li, R. Wu, B. Zhang, G. Nie, Z. Xie, and H. Zhang, "A highly sensitive CRISPR-empowered surface plasmon resonance sensor for diagnosis of inherited diseases with femtomolar-level real-time quantification," *Advanced Science*, vol. 9, no. 14, p. 2105231, 2022.

[6] D. Zou, Q. Zhu, and P. Yan, "Unsupervised domain adaptation with dual-scheme fusion network for medical image segmentation." in *Proc. Int. Joint Conf. Artificial Intell.*, 2020, pp. 3291–3298.

[7] G. Gao, G. Xu, J. Li, Y. Yu, H. Lu, and J. Yang, "FBSNet: A fast bilateral symmetrical network for real-time semantic segmentation," *IEEE Trans. Multimedia*, vol. 25, pp. 3273–3283, 2022.

[8] S. Ouyang, H. Wang, S. Xie, Z. Niu, R. Tong, Y.-W. Chen, and L. Lin, "SLViT: scale-wise language-guided vision transformer for referring image segmentation," in *Proc. Int. Joint Conf. Artificial Intell.*, 2023, pp. 1294–1302.

[9] G. Xu, J. Li, G. Gao, H. Lu, J. Yang, and D. Yue, "Lightweight real-time semantic segmentation network with efficient transformer and CNN," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 12, pp. 15 897–15 906, 2023.

[10] P. Terhörst, M. Ihlefeld, M. Huber, N. Damer, F. Kirchbuchner, K. Raja, and A. Kuijper, "QMagFace: Simple and accurate quality-aware face recognition," in *Proc. Winter Conf. Applications of Comp. Vis.*, 2023, pp. 3484–3494.

[11] J. C. Pérez, M. Alfarra, A. Thabet, P. Arbeláez, and B. Ghanem, "Towards characterizing the semantic robustness of face recognition," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2023, pp. 315–325.

[12] W.-T. Chen, H.-Y. Fang, C.-L. Hsieh, C.-C. Tsai, I. Chen, J.-J. Ding, S.-Y. Kuo *et al.*, "ALL Snow Removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2021, pp. 4196–4205.

[13] R. Yasarla and V. M. Patel, "Uncertainty guided multi-scale residual learning-using a cycle spinning CNN for single image de-raining," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2019, pp. 8405–8414.

[14] K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma, and J. Jiang, "Multi-scale progressive fusion network for single image deraining," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2020, pp. 8346–8355.

[15] H. Liu, Z. Wu, L. Li, S. Salehkalaibar, J. Chen, and K. Wang, "Towards multi-domain single image dehazing via test-time training," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2022, pp. 5831–5840.

[16] B. Cheng, J. Li, Y. Chen, and T. Zeng, "Snow mask guided adaptive residual network for image snow removal," *Computer Vision and Image Understanding*, vol. 236, p. 103819, 2023.

[17] Y. Qiao, M. Shao, L. Wang, and W. Zuo, "Learning depth-density priors for fourier-based unpaired image restoration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2604 – 2618, 2023.

[18] K. Jiang, Z. Wang, P. Yi, C. Chen, Z. Han, T. Lu, B. Huang, and J. Jiang, "Decomposition makes better rain removal: An improved attention-guided deraining network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3981–3995, 2020.

[19] J.-L. Yin, Y.-C. Huang, B.-H. Chen, and S.-Z. Ye, "Color transferred convolutional neural networks for image dehazing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3957–3967, 2019.

[20] R. Li, R. T. Tan, and L.-F. Cheong, "All in one bad weather removal using architectural search," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2020, pp. 3175–3185.

[21] J. M. J. Valanarasu, R. Yasarla, and V. M. Patel, "TransWeather: Transformer-based restoration of images degraded by adverse weather conditions," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2022, pp. 2353–2363.

[22] 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 *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2022, pp. 17 653–17 662.

[23] 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 *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2023, pp. 21 747–21 758.

[24] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2019, pp. 8878–8887.

[25] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2018, pp. 8183–8192.

[26] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4713–4726, 2022.

[27] O. Özdenizci and R. Legenstein, "Restoring vision in adverse weather conditions with patch-based denoising diffusion models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 10 346–10 357, 2023.

[28] B. Xia, Y. Zhang, S. Wang, Y. Wang, X. Wu, Y. Tian, W. Yang, and L. Van Gool, "DiffIR: Efficient diffusion model for image restoration," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2023, pp. 13 095–13 105.

[29] Z. Luo, F. K. Gustafsson, Z. Zhao, J. Sjölund, and T. B. Schön, "Image restoration with mean-reverting stochastic differential equations," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 23 045–23 066.

[30] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng, "Progressive image deraining networks: A better and simpler baseline," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2019, pp. 3937–3946.

[31] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2021, pp. 14 821–14 831.

[32] M. Liu, W. Yang, Y. Hu, and J. Liu, "Dual prompt learning for continual rain removal from single images," in *Proc. Int. Joint Conf. Artificial Intell.*, 2023, pp. 7215–7223.

[33] Y.-F. Liu, D.-W. Jaw, S.-C. Huang, and J.-N. Hwang, "DesnowNet: Context-aware deep network for snow removal," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 3064–3073, 2018.

[34] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," in *Proc. Conf. AAAI*, vol. 34, 2020, pp. 11 908–11 915.

[35] 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 *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2021, pp. 10 551–10 560.

[36] C. Wang, Z. Zheng, R. Quan, Y. Sun, and Y. Yang, "Context-aware pretraining for efficient blind image decomposition," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2023, pp. 18 186–18 195.

[37] K. Zhang, D. Li, W. Luo, W. Ren, and W. Liu, "Enhanced spatio-temporal interaction learning for video deraining: faster and better,"

IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 1, pp. 1287–1293, 2022.

[38] X. Liu, Y. Ma, Z. Shi, and J. Chen, "GridDehazeNet: Attention-based multi-scale network for image dehazing," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2019, pp. 7314–7323.

[39] 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 *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2023, pp. 12 802–12 813.

[40] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2021, pp. 1833–1844.

[41] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general u-shaped transformer for image restoration," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2022, pp. 17 683–17 693.

[42] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2022, pp. 5728–5739.

[43] H. Guo, J. Li, T. Dai, Z. Ouyang, X. Ren, and S.-T. Xia, "Mambair: A simple baseline for image restoration with state-space model," *arXiv preprint arXiv:2402.15648*, 2024.

[44] 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. Vision*, pp. 1–23, 2024.

[45] Z. Luo, F. K. Gustafsson, Z. Zhao, J. Sjölund, and T. B. Schön, "Refusion: Enabling large-size realistic image restoration with latent-space diffusion models," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2023, pp. 1680–1691.

[46] X. Zhang, H. Dong, J. Pan, C. Zhu, Y. Tai, C. Wang, J. Li, F. Huang, and F. Wang, "Learning to Restore Hazy Video: A new real-world dataset and a new method," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2021, pp. 9239–9248.

[47] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2017, pp. 3855–3863.

[48] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3943–3956, 2019.

[49] 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 *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2017, pp. 1357–1366.

[50] 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.*, vol. 28, no. 1, pp. 492–505, 2018.

[51] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2017, pp. 4770–4778.

[52] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha, "Recurrent squeeze-and-excitation context aggregation net for single image deraining," in *Proc. Eur. Conf. Comp. Vis.*, 2018, pp. 254–269.

[53] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, "MAXIM: Multi-axis mlp for image processing," in *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2022, pp. 5769–5780.

[54] K. Zhang, R. Li, Y. Yu, W. Luo, and C. Li, "Deep dense multi-scale network for snow removal using semantic and depth priors," *IEEE Trans. Image Process.*, vol. 30, pp. 7419–7431, 2021.

[55] K. Zhang, W. Luo, Y. Yu, W. Ren, F. Zhao, C. Li, L. Ma, W. Liu, and H. Li, "Beyond monocular deraining: Parallel stereo deraining network via semantic prior," *Int. J. Comput. Vision*, vol. 130, no. 7, pp. 1754–1769, 2022.

[56] K. Zhang, D. Li, W. Luo, and W. Ren, "Dual attention-in-attention model for joint rain streak and raindrop removal," *IEEE Trans. Image Process.*, vol. 30, pp. 7608–7619, 2021.

VI. Biography Section

Bodong Cheng He received the master's degree in Computer Science and Technology from Xidian University, Shaanxi, China, in 2023. He is currently pursuing his Ph.D. at the School of Computer Science and Technology, East China Normal University. His main research interests lie in image processing and medical imaging.

Juncheng Li He received the Ph.D. degree in computer science from East China Normal University, Shanghai, China, in 2021. He has been a Postdoctoral Fellow in the Center for Mathematical Artificial Intelligence, The Chinese University of Hong Kong, China. He is currently an Assistant Professor with the School of Communication and Information Engineering, Shanghai University. He has published more than 40-articles in international journal/conference. His main research interests include artificial intelligence and its applications to computer vision (e.g., image segmentation) and image processing (e.g., image super resolution, image denoising, image dehazing, and medical image processing).

Jun Shi He received the B.S. degree and the Ph.D. degree from the Department of Electronic Engineering and Information Science, University of Science and Technology of China in 2000 and 2005, respectively. In 2005, he joined the School of Communication and Information Engineering, Shanghai University, Shanghai, China, where he has been a professor since 2015. From 2011 to 2012, he was a visiting scholar with the University of North Carolina at Chapel Hill, Chapel Hill, USA. His current research interests include machine learning in medical imaging.

Yingying Fang She obtained the Ph.D. degree in mathematics from Hong Kong Baptist University in September 2020. Her research interests mainly include image restoration and medical image processing. She became a postdoctoral research associate at National Heart Lung Institute in Imperial College London from October 2021. Her current research is focused on the medical imaging.

Guixu Zhang He received the Ph.D. degree from the Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou, China, in 1998. He is currently a Professor with the Department of Computer Science and Technology, East China Normal University, Shanghai, China. His research interests include hyperspectral remote sensing, image processing, and artificial intelligence.

Yin Chen He received the Ph.D. in Applied Mathematics from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences in 2005. He is currently a professor at Beijing Electronic Science and Technology Institute, specializing in research areas such as data mining, artificial intelligence, and image processing.

Tieyong Zeng He received the B.S. degree from Peking University, Beijing, China, in 2000, the M.S. degree from École Polytechnique, Palaiseau, France, in 2004, and the Ph.D. degree from the Université of Paris XIII, Paris, France, in 2007. He is currently a Professor with the Department of Mathematics, The Chinese University of Hong Kong. He worked as a Postdoctoral Researcher with ENS de Cachan from 2007 to 2008 and an Assistant/Associate Professor with Hong Kong Baptist University from 2008 to 2018. His research interests are image processing, machine learning, and scientific computing.

Zhi Li He received the B.S. degree and the M.S. degree from China University of Petroleum, in 2007 and 2010, respectively. He also received the M.S. degree in applied science from Saint Mary's University, Canada, in 2012. After being awarded the HK Ph.D. Fellowship, he went to Hong Kong Baptist University, where he received the Ph.D. degree in 2016. Then he worked as a Postdoctoral Researcher at Michigan State University, USA, from 2016 to 2019. He is currently a professor with the Department of Computer Science and Technology, East China Normal University, Shanghai, China. His main research interests include image processing, medical imaging and object detection.