Towards Universal Rainy Image Restoration: Benchmark and Baseline

Hujie Yan

California Institute of Technology
1200 E California Blvd, Pasadena
California, 91125 USA
hyan3@caltech.edu

Abstract

Despite significant progress has been made in image deraining, most existing methods are limited to handling only a single type of rain degradation or a specific pattern of rain. However, real-world rain scenarios tend to contain diverse rainy patterns due to variations in the rainfall process and lighting conditions. To address this dilemma and advance this field, we introduce a new task: Universal Rainy Image Restoration (URIR), which aims to handle multiple types of rain degradation on a single model. To benchmark this task, we construct a high-quality dataset called URIR-8K, which contains four patterns: rain streak, raindrop, rain accumulation and nighttime rain. Building upon this dataset, we present a comprehensive study on existing approaches by evaluating their universal deraining capabilities and their effect on downstream object detection task. In addition, we design a multi-scale vision Mamba as a baseline model, leveraging the benefits of multi-scale learning for its robustness to diverse rain appearances. Unlike existing methods that use fixed-scale scanning for feature extraction, we employ a multi-scale 2D scanning technique to better help image restoration in the richer scale space. Extensive experimental analysis shows the potential of our proposed task and the effectiveness of our model.

Introduction

Rainy image restoration aims to enhance the quality of images captured in rainy conditions, thereby improving their visual clarity and the accuracy of perception systems (Chen et al. 2021). To solve this problem, recent years have witnessed the emergence of diverse datasets and the proposal of numerous deep learning-based methods. As this field has developed, researchers have focused addressing different rain degradation patterns, such as rain streaks, raindrops, rain accumulation, and so on. These well-defined settings enable researchers to design specific models tailored to the unique characteristics of each rainy scenario (Chen et al. 2023b).

However, these models designed to focus solely on specific rain patterns, often suffer from significant performance declines when applied to other types of rain. In real complex rainy scenarios, multiple types of rain degradation frequently change throughout the rainfall process. For instance, autonomous vehicles may simultaneously experience interference from both rain streaks and raindrops. In addition, the characteristics of rain can shift from day to night. As a result, intelligent systems would have to adapt to and switch between different rain degradation types. In other words, using separate models for different types of rain may lead to inefficiencies in deployment and maintenance. Besides practical limitations, focusing solely on single-degradation rain models is still constrained by their task-specific nature, hindering the field's progress from specialization toward more general intelligence (Li et al. 2022). To advance this field, we introduce a new task: Universal Rainy Image Restoration (URIR). The goal of this new task is to handle multiple types of rain degradation using a single universal model. To better achieve this goal, we need to consider two main factors related to the URIR problem: benchmark and baseline.

Datasets	Year	Rain Categories				Annotation
		RS	RD	RA	NR	
Rain200L/H	2017	✓	×	×	×	×
RainDrop	2018	×	✓	×	×	×
RID/RIS	2019	✓	✓	×	×	✓
SPA-Data	2019	✓	×	×	×	×
RainCityscapes	2019	✓	×	✓	×	✓
Rain13k	2020	✓	×	×	×	×
RainDS	2021	✓	✓	×	×	×
GTAV-NightRain	2022	×	×	×	✓	×
URIR-8K (Ours)	2024	✓	✓	✓	✓	✓

Table 1: Overview of recent datasets for rain removal tasks. The abbreviations "RS", "RD", "RA" and "NR" stand for rain streak, raindrop, rain accumulation, and nighttime rain, respectively. The inclusion of annotations indicates that the dataset contains labels for object detection tasks.

To build a dataset for the URIR task, the most straightforward approach is to simply combine existing datasets of different specific types of rain, as shown in Table 1. However, significant differences exist among these datasets in terms of background quality, resolution, and synthesis methods. For instance, Rain200L/H (Yang et al. 2017) consists of regular rain streaks synthesized using Photoshop software, RainCityscapes (Hu et al. 2019) is designed for driving scenarios by incorporating scene depth, and GTAV-NightRain (Zhang et al. 2023) utilizes a game engine to render nighttime rain degradation. Directly combining these datasets could introduce distribution inconsistencies, potentially confusing the

models and reducing their performance and generalization capabilities. Thus, it is necessary to build a unified benchmark to comprehensively evaluate the URIR capability.

One more thing, how to develop a robust deep model for the URID task is worth exploring. In fact, multi-scale learning has been demonstrated to be an effective strategy in rainy image restoration because it naturally captures rain appearances of various sizes (Fu et al. 2019; Chen, Pan, and Dong 2024; Jiang et al. 2020; Chen et al. 2024). Despite the numerous multi-scale CNN-based and Transformer-based approaches that have been proposed, with the recent popularity of state space models (SSMs), multi-scale representations in SSMs to facilitate the rain removal are still unexplored. This motivates us to design a multi-scale Mamba to explore richer scale-space information for better image deraining.

In this paper, we first construct a new benchmark dataset URIR-8K for the URIR task. To facilitate real-world applications, the proposed URIR-8K is designed for autonomous driving scenarios, with each image containing corresponding object detection labels. Note that we provide a unified pipeline for synthesizing datasets of four common rain patterns, including rain streak, raindrop, rain accumulation, and nighttime rain. We would open-source this pipeline's code, offering researchers a convenient reference for URIR data generation in other scenarios. Inspired by the recent popularity of SSMs, we develop an effective multi-scale Mamba as a new baseline, incorporating a multi-scale 2D scanning mechanism to better help image restoration. Figure 1 shows that our model achieves state-of-the-art performance in various rainy image restoration and object detection tasks compared to existing approaches, which implies the potential of our method to become a pentagonal warrior.

This paper makes the following contributions to the field:

- We introduce a new task setting for universal rainy image restoration, which aims to address different types of rain degradation using a single model.
- We propose a new benchmark dataset for the URIR task to pave the way for future research in this field. We conduct benchmark experiments on existing methods to report their comprehensive deraining capabilities.
- We design an effective multi-scale Mamba to remove rain effects while maintaining a low model complexity. Extensive experiments show that our baseline performs favorable performance against state-of-the-art ones.

Related Work

In this section, we briefly review the progress in image deraining, universal image restoration and state space models.

Rainy Image Restoration. Upon revisiting the field of image deraining, it is evident that numerous deraining methods and datasets have been introduced in recent years, achieving notable success (Chen et al. 2023b). Most existing methods and datasets focus on removing rain streaks, taking into account variations in streak length, density, and direction (Yang et al. 2017; Zhang and Patel 2018). To address the problem of raindrop removal, Qian *et al.* (Qian et al. 2018) create the first dataset specifically for raindrop removal. Hu

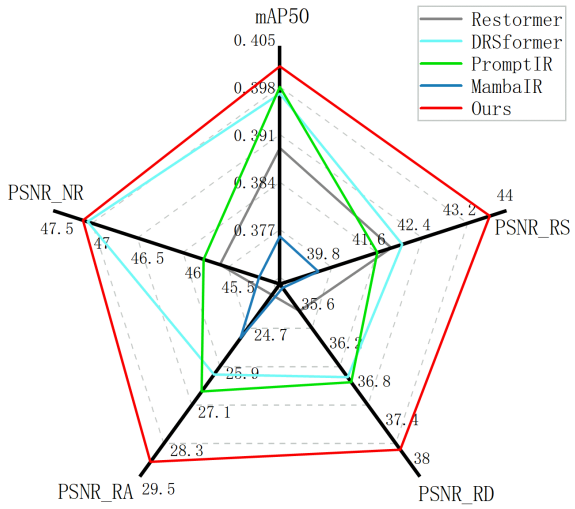

Figure 1: A five-dimensional radar chart compares the comprehensive capacity of state-of-the-art models on the URIR-8K dataset. The dimensions include PSNR across four types of rainy image restoration and mAP50 for object detection.

et al. (Hu et al. 2019) observe that heavy rain is often accompanied by rain fog effects and proposed a new dataset specifically for rain fog. Recently, nighttime deraining has begun to attract researchers' attention (Zhang et al. 2023), as it exhibits distinct differences from daytime rain patterns. Although various datasets and solutions for different types of rain have been proposed, they often overlook the fact that in real-world applications, there is a preference for a single model capable of addressing all types of rain degradation. To fill the gap in this research, we explore universal rainy image restoration to facilitate real-world applications.

Universal Image Restoration. Universal image restoration refers to the comprehensive process of improving the quality of images by addressing various types of degradations. In recent year, several pioneers have conducted studies on universal image restoration models and have made significant progress. For example, Li *et al.* (Li et al. 2022) present a unified approach for dehazing, deraining and denoising, utilizing an image encoder trained via contrastive learning to effectively model latent representations of degradations. Potlapalli *et al.* (Potlapalli et al. 2023) introduce PromptIR, a prompt-based learning approach that implicitly generates degradation-conditioned prompts to direct the restoration of input images with unknown degradations. Inspired by these popular trends, our goal is to explore universal rainy image restoration to establish a robust foundation model for image deraining. In addition, we explore utilizing the recent Mamba architecture to achieve universal image restoration.

Visual State Space Models. Recently, state space models (SSMs) (Gu et al. 2020, 2021; Gu, Goel, and Ré 2021) have interested researchers due to their linearly scalable computational complexity and global awareness capabilities. A recent advancement in this field is Mamba (Gu and Dao 2023), a novel architecture based on SSMs. Mamba uses adaptable

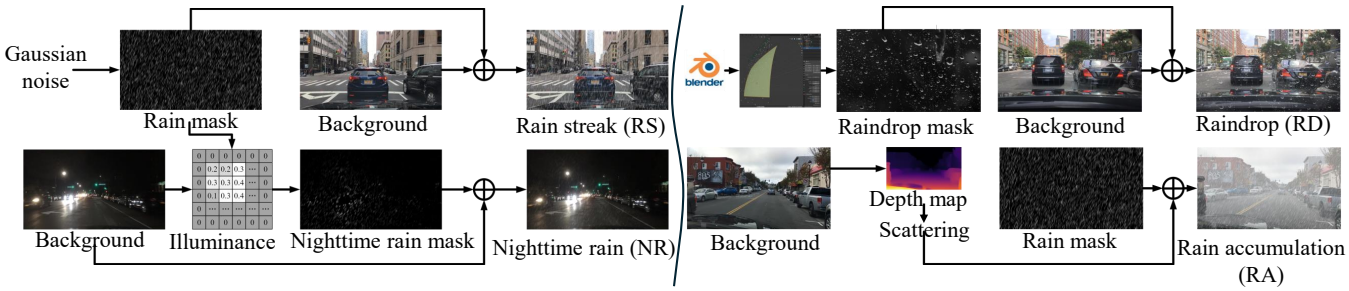

Figure 2: Illustration of the URIR-8K dataset generation pipeline (rain streak, raindrop, rain accumulation, and nighttime rain).

parameters to capture non-local dependencies and hardware-level techniques to balance memory efficiency with performance. Its selective scanning mechanism focuses on extracting essential semantics from long sequences, reducing semantic redundancy. Mamba has been demonstrated to perform well on large-scale datasets. For example, ViM (Zhu et al. 2024) employs multiple scanning directions to handle the non-causal nature of image data, achieving competitive results. Motivated by their success, numerous Mamba-based studies have emerged in low-level vision tasks (Zhou et al. 2024; Guo et al. 2024), obtaining further advancements compared to CNNs and ViTs. However, these methods are limited to exploring feature representations at a fixed image scale. In this work, we investigate multi-scale representations in SSMs for better boosting image restoration.

Dataset Construction

To evaluate the performance of existing approaches for universal rainy image restoration, we first create a high-quality benchmark dataset named URIR-8K. We note that existing image deraining datasets often focus on particular types of rain, lacking the diversity needed for a comprehensive evaluation. Simply combining these existing datasets may result in background inconsistencies and unknown gaps, failing to represent the wide range of rainy conditions adequately. To this end, we propose a new pipeline for synthesizing data of different rain types, which will be described below.

Background Collection and Object Detection Labels. To facilitate the real-world application of universal rainy image restoration, we collect numerous rain-free backgrounds from the BDD100K dataset (Seita 2018). Our ground-truth data includes a diverse range of typical daytime and nighttime first-person driving scenes in urban environments. Specifically, these scenes are characterized by elements such as buildings, streets, and the sky. It is worth noting that current rain datasets focus solely on rain synthesis, overlooking the downstream task integration. Our dataset also contains object detection labels, which are equally crucial for investigating the effect of image deraining for downstream vision-based tasks, such as object detection. Here, we emphasize traffic-relevant objects, including cars, buses, pedestrians, bicycles, trucks, motorcycles, and traffic lights. These objects are labeled in an accompanying file for each image.

Rain Streak Generation. Diversity and fidelity are the two main considerations in rain streak generation. For diversity, factors such as rain density, thickness, length, and direction play crucial roles. To control these factors, we apply a motion blur process (Garg and Nayar 2007; Wang et al. 2020b) to generate the rain layer, transforming Gaussian noise into directed streaks with specific thickness by tuning the motion blur kernels. To enhance fidelity, we use an alpha blending technique (Porter and Duff 1984) to organically mix the rain layer with the background layer. This produces the desired visual effect, where rain streaks are less visible in lighter areas (such as the sky), which is common in driving scenarios.

Raindrop Generation. To synthesize raindrops with realistic morphology, we employ open-source computer graphics software, Blender [1], to simulate the process of rain hitting a window, and capture raindrop morphology. This software engine is capable of rendering random raindrops with a physical model, facilitated by the Rain Generator plugin in Blender. A random cropping approach with image augmentation is applied to ensure each raindrop mask is unique. For better realism, we tune the transparency of raindrop masks and blend them with background images in the alpha channel to produce the occluded effect (Qian et al. 2018).

Rain Accumulation Generation. In the case of heavy rain, the rain streaks and water particles in the atmosphere create a rain veiling effect similar to fog, which blurs background objects (Yang et al. 2017). This phenomenon is known as rain accumulation. As a result, objects at different distances have varying levels of visibility: distant objects are blurred more, while closer objects are less blurred. To authentically model this effect, we adopt an atmospheric scattering model (McCartney 1976) where object visibility decays exponentially with depth and the resulting background is superimposed onto a white fog blurring mask. To better obtain the depth of the objects, we use Depth Anything (Yang et al. 2024) to generate a depth mask. We randomly sample a coefficient A to represent the intensity of rain accumulation and a coefficient β to determine the intensity of atmospheric scattering.

Nighttime Rain Generation. During the day, illumination comes from the sun reflection, which can be considered uniform, making rain streaks clearly visible. At nighttime, light

[1] https://blendermarket.com/products/rain-generator

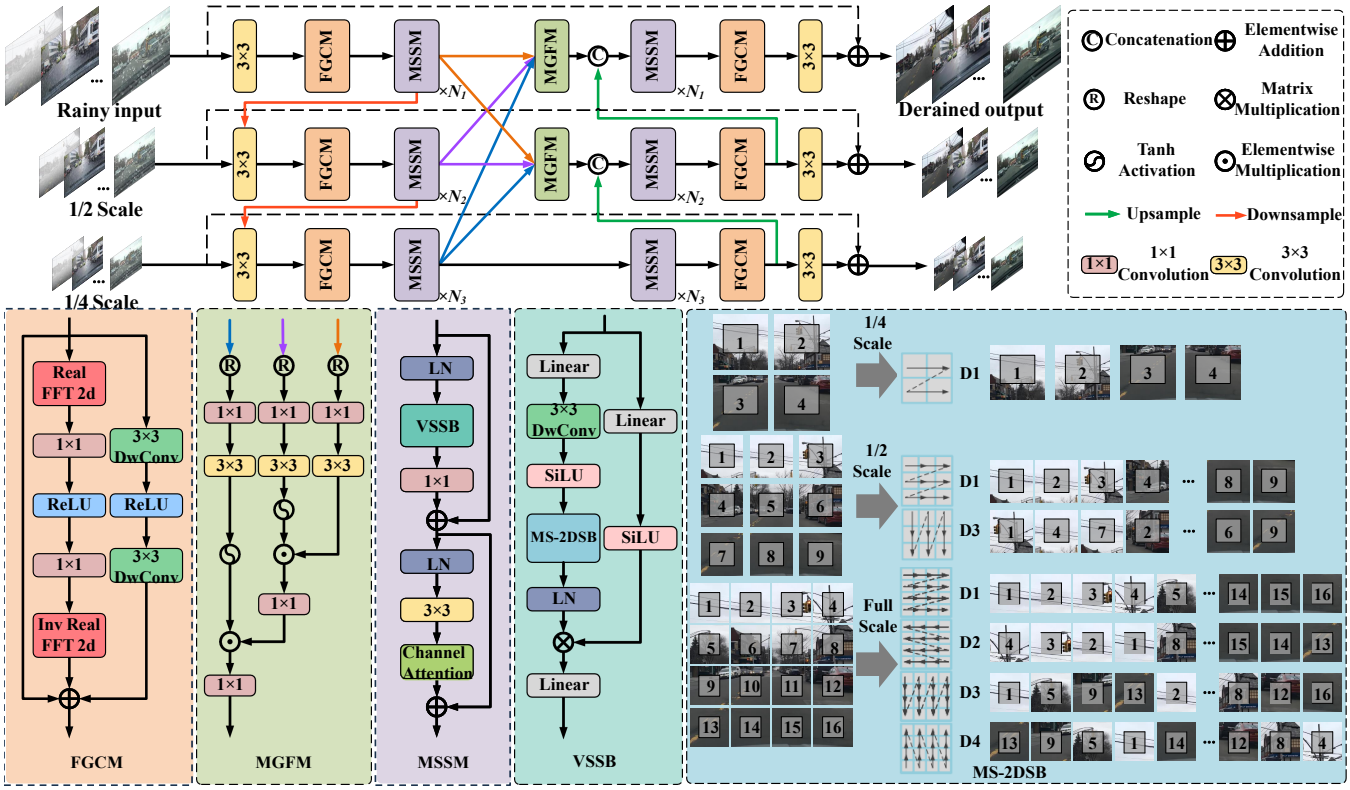

Figure 3: Overall architecture of the proposed multi-scale Mamba (MSDM) for rainy image restoration.

mainly comes from some artificial light sources, such as traffic lights and vehicle headlights, resulting in uneven illumination. As the results, rain streaks are not visible in areas with low illumination (Cheng et al. 2023). In addition, areas with extremely high illumination (such as the center of a headlight) also obscure the visibility of rain streaks. To authentically depict nighttime rain, we transform the background image from RGB to HSV format, where the V channel reflects the distribution of image brightness. We use the normalized V channel to create a visibility mask: areas with V values either below 0.1 or above 0.8 are set to low values, while other areas retain their original V values. Then, we convolute the visibility mask with the rain streak mask generated from Gaussian motion blur. Finally, we use an alpha blending technique to mix the rain mask with the background images to generate rain streaks in visible areas.

Benchmark Statistics. Figure 2 presents the overall data synthesis pipeline. In total, our URIR-8K dataset contains 7,200 training pairs and 800 test images, which is divided into four subsets: rain streaks (RS), raindrops (RD), rain accumulation (RA), and nighttime rain (NR). These scenes and data for the training and test sets are completely separate, ensuring no overlap. The average resolution is 1280×720.

Baseline Method

In this section, we first summarize the overall pipeline of our method, followed by an analysis of the main components.

Overall Framework

The overall framework of our proposed multi-scale deraining Mamba (MSDM) is illustrated in Figure 3. Given an input rainy image $I \in \mathbb{R}^{H\times W\times 3}$, our method down-samples it into two coarser scales, *i.e.* $I_{mid} = \frac{1}{2}, I_{min} = \frac{1}{4}$. The network uses pyramid rainy images as multiple inputs and employs 3×3 convolutional layers to extract shallow features. To better enhance the locality of the network in the frequency domain, we introduce the frequency-guided convolutional module (FGCM) (Mao et al. 2023), leveraging the initial features from each image scale. Then, we develop the multi-scale state-space module (MSSM) to model deep global information with linear complexity. Finally, we leverage the multi-scale gated fusion module (MGFM) to aggregate scale-specific representations, which are then fed into the network decoder to reconstruct high-quality outputs. The detailed description of these components is provided below.

To supervise the network training procedure, we employ a weighted sum of three loss functions, defined as follows:

$$L_{total} = \lambda_1 L_c + \lambda_2 L_f + \lambda_3 L_e,$$

where L_c is the Charbonnier loss (Charbonnier et al. 1994), L_f is the frequency reconstruction loss (Cho et al. 2021), L_e is the edge loss (Zamir et al. 2021). Each component loss is calculated across all image scales. The empirical values for λ_1, λ_2, and λ_3 are set to 1, 0.01, and 0.05, respectively.

Multi-scale State Space Module

Preliminaries. The SSM serves as a fundamental mathemat-

ical framework, drawing inspiration from continuous linear time-invariant (LTI) systems. These models transform one-dimensional functions or sequences $x_t \in \mathbb{R}$ into outputs $y_t \in \mathbb{R}$ through implicit latent states $h_t \in \mathbb{R}^N$. Mathematically, SSMs are typically represented by a set of first-order differential equations. In the context of deep learning, these continuous equations are discretized using the zero-order hold (ZOH) method, resulting in the discrete-time equations:

$$h_t = Ah_{t-1} + Bx_t, \quad y_t = Ch_t + Dx_t,$$

where A, B, C, and D denote learnable weight matrices.

A recent advancement in the application of SSMs is the Mamba architecture (Gu and Dao 2023), which leverages a selective scanning mechanism. This mechanism allows the weights to dynamically adapt to changes in the input while maintaining the advantage of linear complexity. The way to apply the scanning mechanism becomes a key factor in the performance of Mamba-based image restoration tasks.

Multi-scale 2D Scanning Block. For recent Mamba-based method, a variety of image scanning strategies flourish. For example, Vim (Zhu et al. 2024) and VMamba (Liu et al. 2024) have demonstrated that utilizing different scanning orders, including both row-wise and column-wise scans in multiple directions, can effectively enhance model performance. However, these approaches primarily focus on feature representations of a fixed image scale, neglecting potentially useful information from other scales. Therefore, exploring multi-scale scanning to boost image deraining performance is crucial. Based on the above multi-scale architecture using a multi-input encoder and a multi-output decoder, the most straightforward approach is to apply the existing multi-directional scanning mechanism to each image scale. In fact, the effect of rain degradation varies across different image scales (Chen, Pan, and Dong 2024). In other words, the information contained at larger scales differs from that at smaller scales. As a result, applying the same scanning operation to different scales may result in information redundancy and additional computational burden. To this end, we develop an efficient multi-scale 2D scanning block that achieves data scanning by applying different numbers of geometric transformations at each scale. Compared to smaller scales, we adopt a greater number of scans at larger scales to comprehensively model the information. Mathematically, this step is represented as a piecewise function:

$$\hat{\boldsymbol{B}} = \begin{cases} \boldsymbol{B} & \text{if } k = 1 \\ \boldsymbol{B} \bigoplus \boldsymbol{B}^F & \text{if } k = 2 \\ \boldsymbol{B} \bigoplus \boldsymbol{B}^F \bigoplus \boldsymbol{B}^T \bigoplus (\boldsymbol{B}^T)^F & \text{if } k = 4 \end{cases}$$

where k represents the number of 2D scanning operations. Here, $k = 1$ is used for the small scale, $k = 2$ for the medium scale, and $k = 4$ for the large scale. The transpose operation ($\boldsymbol{B}^T$) swaps the image's horizontal and vertical axes, and the flip operation ($\boldsymbol{B}^F$) reverses pixel arrangement along both axes. The direct sum symbol $\bigoplus$ represents the stacking operation on the images. We will show the effectiveness of these design choices in the experimental section.

Multi-scale Gated Fusion Module

As is well known, multi-scale feature fusion integrates features extracted from different scales of an image, allowing the network to capture both global context and fine details. Different from simply concatenating outputs from different scales of the MSSM (Mao et al. 2023), we employ a dual-input gating unit that dynamically adapts to inputs of varying scales. The formula for this gating unit can be expressed as:

$$\text{Gate}(\mathbf{X}, \mathbf{Y}) = F_{3\times3}(F_{1\times1}(\mathbf{X})) \odot \sigma(F_{3\times3}(F_{1\times1}(\mathbf{Y}))),$$

where $F_{3\times3}$ represents 3×3 convolution layer, σ represents the tanh activation function, and $\odot$ denotes element-wise multiplication. This dynamic gating unit can be used for pairwise integration of different scales. To integrate features from three different scales, we combine the gating units into the MGFM, which can be mathematically expressed as:

$$F_{GFM} = F_{1\times1}(\text{Gate}(F_{1\times1}(\text{Gate}(\mathbf{B}_S, \mathbf{B}_M)), \mathbf{B}_L)),$$

where $\mathbf{B}_L$, $\mathbf{B}_M$, $\mathbf{B}_S$ denotes three different image scales.

Experiments

In this section, we conduct extensive experiments on the proposed URIR-8K dataset and several pubilc benchmarks.

Experimental Setup

Datasets and Metrics. We conduct benchmark experiments on the proposed URIR-8K dataset. We compare it with both classical and recent methods, including four CNN-based approaches: LPNet (Fu et al. 2019), JORDER-E (Yang et al. 2019), RCDNet (Wang et al. 2020a), and SPDNet (Yi et al. 2021); four Transformer-based networks: IDT (Xiao et al. 2022), Restormer (Zamir et al. 2022), DRSformer (Chen et al. 2023a), and PromptIR (Potlapalli et al. 2023); and one Mamba-based architecture, MambaIR (Guo et al. 2024).

Furthermore, we evaluate the effectiveness of our method on existing benchmarks, including two rain streak datasets, Rain200L (Yang et al. 2017) and Rain200H (Yang et al. 2017); a raindrop dataset, UAV-Rain1K (Chang et al. 2024); and a real-world rain dataset, SPA-Data (Wang et al. 2019). To assess the quality of the restored images, we employ two metrics: PSNR (Huynh-Thu and Ghanbari 2008) and SSIM (Wang et al. 2004). To evaluate the performance in object detection tasks, we employ the mAP50 metric (Redmon et al. 2016) to measure accuracy in traffic scenes.

Implementation Details. In our network, the stack size of MSSMs is set to 6. The models are trained using the PyTorch framework with the Adam optimizer. The initial learning rate is 1×10^{-4}, which is gradually decreased to 1×10^{-6} following a cosine annealing strategy (Loshchilov and Hutter 2016). An exception is made for the training on Rain200H, where the initial learning rate is set to 2×10^{-4}. The models are trained on the URIR-8K, Rain200L, Rain200H, and UAV-Rain1K datasets for 300 epochs, and on SPA-Data for 5 epochs. To enhance the training procedure, images in these datasets are randomly flipped for data augmentation. All experiments are conducted with a batch size of 1 and a patch size of 256, utilizing one NVIDIA GeForce RTX 4090 GPU.

Model	Venue	RS		RD		RA		NR		Average		
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	mAP50
LPNet	TNNLS'2019	16.09	0.7254	16.64	0.6688	17.76	0.8027	21.38	0.7630	17.97	0.7400	0.200
JORDER-E	TPAMI'2019	36.72	0.9674	32.11	0.9549	23.76	0.8985	42.80	0.9929	33.85	0.9534	0.366
RCDNet	CVPR'2020	32.00	0.9317	27.01	0.8973	21.47	0.8610	38.52	0.9847	29.75	0.9187	0.343
SPDNet	ICCV'2021	32.52	0.9399	27.97	0.9064	23.55	0.8737	39.17	0.9861	30.80	0.9265	0.341
IDT	TPAMI'2022	40.89	0.9889	35.48	0.9802	**29.34**	**0.9711**	46.11	0.9964	37.96	0.9842	0.401
Restormer	CVPR'2022	41.46	0.9864	35.42	0.9671	23.84	0.9140	45.66	0.9955	36.59	0.9657	0.390
DRSformer	CVPR'2023	41.70	0.9874	36.45	0.9752	26.32	0.9491	47.12	0.9966	37.90	0.9771	0.398
PromptIR	NeurIPS'2023	41.15	0.9864	36.53	0.9758	26.85	0.9467	45.84	0.9956	37.59	0.9761	0.399
MambaIR	ECCV'2024	39.85	0.9825	35.06	0.9681	25.18	0.9187	45.23	0.9953	36.33	0.9661	0.377
Ours	-	**43.63**	**0.9918**	**37.58**	**0.9811**	29.04	0.9679	**47.17**	**0.9967**	**39.35**	**0.9844**	**0.402**

Table 2: Quantitative comparison on the URIR-8K dataset. Bold and underline highlights the best and second-best results. RS, RD, RA, and NR refer to rain streak, raindrop, rain accumulation, and nighttime rain, respectively.

(a) Input (b) Restormer (c) DRSformer (d) PromptIR (e) MambaIR (f) Ours (g) GT

Figure 4: Visual deraining results on the URIR-8K dataset. Best zoom in the figures for better visual comparison.

Experimental Results

Evaluations on the proposed URIR-8K dataset. Table 2 presents the quantitative results of different approaches on the URIR-8K dataset. It is evident that our proposed baseline achieves the highest average PSNR and SSIM values. Notably, our method surpasses the second-highest performing model IDT, by 1.4dB in PSNR. In Figure 4, we further show the visual recovery results of different methods under various rain degradations. Although DRSformer is competitive in removing rain streaks, it still has limitations in restoring rain accumulation. Compared to the recent MambaIR, our method introduces multi-scale scanning to better capture rain degradations of varying scales. In contrast, our method produces clearer restoration results with better detail repair.

Evaluations on public benchmarks. We validate the effectiveness of our method on public benchmark datasets with specific rain degradations. The quantitative results in Table 4 show that our method still achieves the best performance, indicating that our multi-scale architecture has good generalization capabilities for different types of rain degradation.

Model	Restormer	DRSformer	PromptIR	MambaIR	Ours
#FLOPs	174.7	242.9	172.7	172.4	**99.2**
#Params	26.1	33.7	35.6	30.8	**11.8**

Table 3: Comparison of model complexity for 256×256 pixel images. "#FLOPs" and "#Params" represent FLOPs (in G) and the number of trainable parameters (in M).

Evaluations on model complexity. We evaluate the complexity of our model and recent methods using FLOPs and model parameters. As shown in Table 3, our model offers a lower FLOPs count and fewer parameters, yet it maintains excellent performance as evidenced in Table 2.

Evaluations on downstream tasks. To evaluate the impact of the image deraining process on downstream vision-based applications such as object detection, we utilize mainstream object detection pre-trained models, specifically YOLOv5,

Model	RCDNet		SPDNet		Uformer		Restormer		IDT		DRSformer		Ours	
Metrics	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Rain200L	39.17	0.9885	40.50	0.9875	40.20	0.9860	40.99	0.9890	40.74	0.9884	<u>41.23</u>	<u>0.9894</u>	**41.47**	**0.9899**
Rain200H	30.24	0.9048	31.28	0.9207	30.80	0.9105	32.00	0.9329	32.10	<u>0.9344</u>	32.17	0.9326	**32.32**	**0.9361**
UAV-Rain1K	22.48	0.8753	22.54	0.8594	-	-	24.78	0.9054	22.47	0.8957	<u>24.93</u>	<u>0.9155</u>	**25.04**	**0.9174**
SPA-Data	43.36	0.9831	43.20	0.9871	46.13	0.9913	47.98	0.9921	47.35	<u>0.9930</u>	<u>48.54</u>	0.9924	**48.95**	**0.9931**

Table 4: Quantitative comparison on public benchmark datasets. Bold and underline highlights the best and second-best results.

(a) Input (b) RCDNet (c) SPDNet (d) DRSformer (e) MambaIR (f) Ours

Figure 5: Comparison results of traffic object detection on images degraded by rain and restored using different methods.

Models	(a)	(b)	(c)	(d)	(e)	(f)	(g)
S_1	×	×	✓	✓	✓	✓	✓
S_2	×	✓	✓	✓	✓	✓	✓
S_3	✓	✓	✓	✓	✓	✓	✓
2DSB	✓	✓	×	×	×	×	×
MS-2DSB	×	×	✓	✓	✓	✓	✓
FGCM	✓	✓	×	✓	✓	✓	✓
MGFM	×	✓	✓	×	✓	✓	✓
L_c	✓	✓	✓	✓	✓	✓	✓
L_f	✓	✓	✓	✓	×	✓	✓
L_e	✓	✓	✓	✓	×	×	✓
PSNR	39.85	39.99	40.68	41.16	41.09	41.27	41.37

Table 5: Ablation studies on different variants of our method. Here, S_1, S_2, and S_3 represent $1/4$, $1/2$ and full image scale. The term 2DSB denotes the 2D scanning block, while MS-2DSB refers to the multi-scale 2D scanning block.

to evaluate the outcomes. As illustrated in Figure 5, our approach not only reconstructs clear images but also enhances target recognition accuracy, particularly in correctly identifying the "traffic light" category and reducing misidentification of building surface as "traffic sign".

Ablation Studies

Effectiveness of main components. We first analyze the effectiveness of each component in the proposed framework, including the multi-scale configuration, MS-2DSB, FGCM, MGFM, and loss functions. Here, we train these models using the same number of epochs for fairness. Table 5 presents the quantitative results of different variants. Compared to the fixed-scale 2DSB, our proposed MS-2DSB achieves superior results. This advantage is due to our model, where exploring multi-scale representations facilitates rain removal. The comparison results of other variants reveals that each component of MSDM contributes to the final performance.

Effect of the scanning operation in the MS-2DSB. We further analyze the impact of the scanning operation in the MS-2DSB. Table 6 presents the quantitative results of applying different scanning operations across various scales.

Models	(a)			(b)			(c)			(d, Ours)		
	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3
D1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
D2	×	×	×	✓	✓	✓	✓	✓	✓	×	✓	✓
D3	×	×	×	×	×	×	✓	✓	✓	×	×	✓
D4	×	×	×	×	×	×	✓	✓	✓	×	×	✓
PSNR	40.58			40.59			40.89			40.86		
#Params	11.7			11.8			12.1			11.8		
#FLOPs	96.9			98.3			100.9			99.2		

Table 6: Ablation studies on the MS-2DSB. D_1, D_2, D_3, and D_4 denotes different scanning directions in Figure 3.

Compared to model (c), which uses four-directional scanning at each scale, our MS-2DSB achieves comparable performance with unequal scanning operations across different scales, thus reducing the model's computational complexity. This finding offers new insights into efficient data scanning techniques for multi-scale Mamba-based methods.

Concluding Remarks

This paper explores the task of universal rainy image restoration (URIR) for the first time and introduces a high-quality dataset URIR-8K that contains four distinct rain degradation patterns. Based on this benchmark, we conduct extensive experiments to evaluate the comprehensive capabilities of different approaches in the URIR task. Furthermore, we also propose an effective Mamba-based baseline model that utilizes multi-scale 2D scanning mechanisms to jointly model rain distribution. Experimental results demonstrate the effectiveness of our framework, surpassing existing models on both our URIR-8K dataset and several public benchmarks.

References

Chang, W.; Chen, H.; He, X.; Chen, X.; and Shen, L. 2024. UAV-Rain1k: A Benchmark for Raindrop Removal from UAV Aerial Imagery. In *CVPR Workshops*, 15–22.

Charbonnier, P.; Blanc-Feraud, L.; Aubert, G.; and Barlaud, M. 1994. Two deterministic half-quadratic regularization

algorithms for computed imaging. In *ICIP*, volume 2, 168–172.

Chen, H.; Chen, X.; Lu, J.; and Li, Y. 2024. Rethinking Multi-Scale Representations in Deep Deraining Transformer. In *AAAI*, volume 38, 1046–1053.

Chen, H.; Wang, Y.; Guo, T.; Xu, C.; Deng, Y.; Liu, Z.; Ma, S.; Xu, C.; Xu, C.; and Gao, W. 2021. Pre-trained image processing transformer. In *CVPR*, 12299–12310.

Chen, X.; Li, H.; Li, M.; and Pan, J. 2023a. Learning a sparse transformer network for effective image deraining. In *CVPR*, 5896–5905.

Chen, X.; Pan, J.; and Dong, J. 2024. Bidirectional multi-scale implicit neural representations for image deraining. In *CVPR*, 25627–25636.

Chen, X.; Pan, J.; Dong, J.; and Tang, J. 2023b. Towards unified deep image deraining: A survey and a new benchmark. *arXiv preprint arXiv:2310.03535*.

Cheng, Y.; Wu, Z.; Li, J.; and Xu, J. 2023. Retinex Meets Transformer: Bridging Illumination and Reflectance Maps for Low-Light Image Enhancement. In *International Conference on Neural Information Processing*, 388–402.

Cho, S.-J.; Ji, S.-W.; Hong, J.-P.; Jung, S.-W.; and Ko, S.-J. 2021. Rethinking coarse-to-fine approach in single image deblurring. In *ICCV*, 4641–4650.

Fu, X.; Liang, B.; Huang, Y.; Ding, X.; and Paisley, J. 2019. Lightweight pyramid networks for image deraining. *IEEE TNNLS*, 31(6): 1794–1807.

Garg, K.; and Nayar, S. K. 2007. Vision and rain. *IJCV*, 75: 3–27.

Gu, A.; and Dao, T. 2023. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.

Gu, A.; Dao, T.; Ermon, S.; Rudra, A.; and Ré, C. 2020. Hippo: Recurrent memory with optimal polynomial projections. *NeurIPS*, 33: 1474–1487.

Gu, A.; Goel, K.; and Ré, C. 2021. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*.

Gu, A.; Johnson, I.; Goel, K.; Saab, K.; Dao, T.; Rudra, A.; and Ré, C. 2021. Combining recurrent, convolutional, and continuous-time models with linear state space layers. *NeurIPS*, 34: 572–585.

Guo, H.; Li, J.; Dai, T.; Ouyang, Z.; Ren, X.; and Xia, S.-T. 2024. Mambair: A simple baseline for image restoration with state-space model. In *ECCV*.

Hu, X.; Fu, C.-W.; Zhu, L.; and Heng, P.-A. 2019. Depth-attentional features for single-image rain removal. In *CVPR*, 8022–8031.

Huynh-Thu, Q.; and Ghanbari, M. 2008. Scope of validity of PSNR in image/video quality assessment. *Electronics letters*, 44(13): 800–801.

Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Huang, B.; Luo, Y.; Ma, J.; and Jiang, J. 2020. Multi-scale progressive fusion network for single image deraining. In *CVPR*, 8346–8355.

Li, B.; Liu, X.; Hu, P.; Wu, Z.; Lv, J.; and Peng, X. 2022. All-in-one image restoration for unknown corruption. In *CVPR*, 17452–17462.

Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; and Liu, Y. 2024. VMamba: Visual State Space Model. arXiv:2401.10166.

Loshchilov, I.; and Hutter, F. 2016. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.

Mao, X.; Liu, Y.; Liu, F.; Li, Q.; Shen, W.; and Wang, Y. 2023. Intriguing findings of frequency selection for image deblurring. In *AAAI*, volume 37, 1905–1913.

McCartney, E. J. 1976. Optics of the atmosphere: scattering by molecules and particles. *New York*.

Porter, T.; and Duff, T. 1984. Compositing digital images. In *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, 253–259.

Potlapalli, V.; Zamir, S. W.; Khan, S. H.; and Shahbaz Khan, F. 2023. Promptir: Prompting for all-in-one image restoration. *NeurIPS*, 36.

Qian, R.; Tan, R. T.; Yang, W.; Su, J.; and Liu, J. 2018. Attentive generative adversarial network for raindrop removal from a single image. In *CVPR*, 2482–2491.

Redmon, J.; Divvala, S.; Girshick, R.; and Farhadi, A. 2016. You only look once: Unified, real-time object detection. In *CVPR*, 779–788.

Seita, D. 2018. BDD100k: A large-scale diverse driving video database. *The Berkeley Artificial Intelligence Research Blog. Version*, 511: 41.

Wang, H.; Xie, Q.; Zhao, Q.; and Meng, D. 2020a. A model-driven deep neural network for single image rain removal. In *CVPR*, 3103–3112.

Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; and Lau, R. W. 2019. Spatial attentive single-image deraining with a high quality real rain dataset. In *CVPR*, 12270–12279.

Wang, Y.-T.; Zhao, X.-L.; Jiang, T.-X.; Deng, L.-J.; Chang, Y.; and Huang, T.-Z. 2020b. Rain streaks removal for single image via kernel-guided convolutional neural network. *IEEE TNNLS*, 32(8): 3664–3676.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4): 600–612.

Xiao, J.; Fu, X.; Liu, A.; Wu, F.; and Zha, Z.-J. 2022. Image de-raining transformer. *IEEE TPAMI*, 45(11): 12978–12995.

Yang, L.; Kang, B.; Huang, Z.; Xu, X.; Feng, J.; and Zhao, H. 2024. Depth anything: Unleashing the power of large-scale unlabeled data. In *CVPR*, 10371–10381.

Yang, W.; Tan, R. T.; Feng, J.; Guo, Z.; Yan, S.; and Liu, J. 2019. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE TPAMI*, 42(6): 1377–1393.

Yang, W.; Tan, R. T.; Feng, J.; Liu, J.; Guo, Z.; and Yan, S. 2017. Deep joint rain detection and removal from a single image. In *CVPR*, 1357–1366.

Yi, Q.; Li, J.; Dai, Q.; Fang, F.; Zhang, G.; and Zeng, T. 2021. Structure-preserving deraining with residue channel prior guidance. In *ICCV*, 4238–4247.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 5728–5739.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.-H.; and Shao, L. 2021. Multi-stage progressive image restoration. In *CVPR*, 14821–14831.

Zhang, F.; You, S.; Li, Y.; and Fu, Y. 2023. Learning rain location prior for nighttime deraining. In *ICCV*, 13148–13157.

Zhang, H.; and Patel, V. M. 2018. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*, 695–704.

Zhou, H.; Wu, X.; Chen, H.; Chen, X.; and He, X. 2024. RS-Dehamba: Lightweight Vision Mamba for Remote Sensing Satellite Image Dehazing. *arXiv preprint arXiv:2405.10030*.

Zhu, L.; Liao, B.; Zhang, Q.; Wang, X.; Liu, W.; and Wang, X. 2024. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*.