
MODEM: A Morton-Order Degradation Estimation Mechanism for Adverse Weather Image Recovery

Hainuo Wang, Qiming Hu, Xiaojie Guo*

College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
hainuo@tju.edu.cn huqiming@tju.edu.cn xj.max.guo@gmail.com

Abstract

Restoring images degraded by adverse weather remains a significant challenge due to the highly non-uniform and spatially heterogeneous nature of weather-induced artifacts, *e.g.*, fine-grained rain streaks versus widespread haze. Accurately estimating the underlying degradation can intuitively provide restoration models with more targeted and effective guidance, enabling adaptive processing strategies. To this end, we propose a Morton-Order Degradation Estimation Mechanism (MODEM) for adverse weather image restoration. Central to MODEM is the Morton-Order 2D-Selective-Scan Module (MOS2D), which integrates Morton-coded spatial ordering with selective state-space models to capture long-range dependencies while preserving local structural coherence. Complementing MOS2D, we introduce a Dual Degradation Estimation Module (DDEM) that disentangles and estimates both global and local degradation priors. These priors dynamically condition the MOS2D modules, facilitating adaptive and context-aware restoration. Extensive experiments and ablation studies demonstrate that MODEM achieves state-of-the-art results across multiple benchmarks and weather types, highlighting its effectiveness in modeling complex degradation dynamics. Our code will be released at here.

1 Introduction

Computer vision systems are increasingly integral to critical applications, such as autonomous driving [51, 1] and intelligent surveillance [49, 70], demanding reliable performance in diverse environments. However, their effectiveness deteriorates significantly under adverse weather conditions, such as rain [16, 42, 57, 75, 78, 92], haze [76, 77, 21, 62, 71, 83, 85], and snow [38, 45, 59, 87, 13, 12], which introduce complex visual artifacts and obscure critical scene information. Thus, effectively restoring clear images from such weather-degraded inputs is essential to boost the robustness and safety of modern computer vision technologies in real-world deployments.

Early task-specific methods [2, 4, 23, 60, 61, 30, 32, 90] largely rely on physical models or hand-crafted statistical priors tailored to individual weather phenomena. Due to the limited feature representation capabilities, these schemes are brittle in the face of complex scenes or deviations from assumed degradation patterns. With the advent of deep learning, numerous deep networks have achieved impressive performance by training on large-scale datasets for specific tasks such as image deraining [16, 42, 57, 75, 78, 5, 41, 69, 83], dehazing [76, 77, 21, 62, 71, 83, 85], or desnowing [38, 45, 59, 87]. These models excel at implicitly learning the inverse mapping through extensive supervision. But their highly specialized nature necessitates separate models for each weather condition, severely limiting scalability and practical deployment in unconstrained environments.

Recently, much attention has been directed toward unified or multi-task frameworks designed to cope with various weather conditions within a single model [95, 67, 11, 40, 55, 65, 79]. These

*Corresponding Author

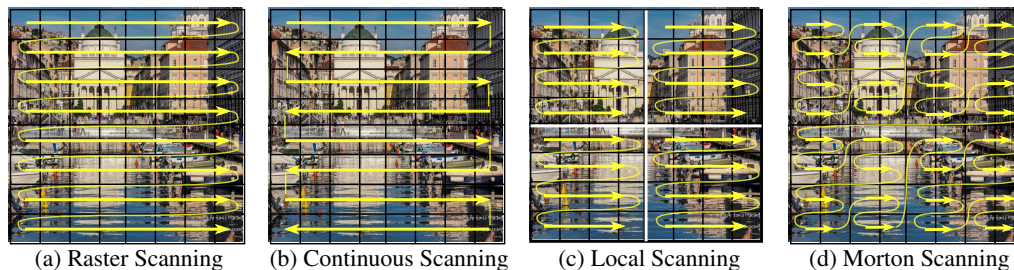


Figure 1: Comparison of various image scanning methods. (a) Raster Scanning. (b) Continuous Scanning. (c) Local Scanning. (d) Morton Scanning, which can better preserve spatial locality of neighboring pixels in the resulting 1D sequence, beneficial for capturing contextual information.

approaches adopt a variety of sophisticated strategies, including architecture search for optimal generalization [40], disentanglement of weather-shared and weather-specific representations [95], knowledge distillation and regularization [11], generative diffusion modeling [55], learned priors or codebooks [79], and transformer-based architectures that offer strong global modeling capabilities [67, 65]. Although these models mark notable progress and offer broader applicability, they still struggle with the challenge of modeling the inherently distinct and often highly spatially heterogeneous characteristics of different weather degradations. For example, haze tends to cause smooth intensity attenuation, whereas rain streaks and snowflakes introduce sharp, local occlusions with specific structural patterns. Most existing architectures lack dedicated degradation estimation mechanisms to explicitly model and leverage these fine-grained, spatially variant degradation patterns, which undermines their performance in real-world deployment. Therefore, *effective estimation of spatially variant degradation characteristics, encompassing both global trends and local structures, as guidance is critical for adaptive and context-aware restoration under dynamic weather conditions.*

For an element (*e.g.*, pixel) of a degraded image, its degradation characteristic can be regarded as a latent “state” within a degradation space, encapsulating the influence of adverse weather on its visual appearance. These states evolve spatially, governed by both local (*e.g.*, rain streaks) and non-local patterns (*e.g.*, drifting fog). This perspective naturally lends itself to State Space Modeling (SSM) [17]. The connection between the image degradation modeling and SSM is formally given in Sec. 3. By treating degradation as a structured sequence of evolving states, SSM can capture long-range contextual dependencies while maintaining sensitivity to local features. When coupled with degradation estimation mechanisms, SSM empowers the restoration process to be contextually aware and spatially adaptive, effectively aligning restoration strategies with the heterogeneous and scene-specific nature of weather-induced artifacts.

Motivated by the above insight, we propose a novel Morton-Order Degradation Estimation Mechanism (MODEM) that **estimates** degradation state via SSM and **guides** adaptive restoration in degraded images. At the heart of MODEM lies the Morton-Order 2D-Selective-Scan Module (MOS2D), which integrates SSM with Morton-coded spatial traversal. Unlike conventional raster scanning, the Morton-order scan follows a hierarchical, locality-preserving sequence [50, 34, 53, 31, 7, 54] that enables structured and efficient modeling of both local and long-range dependencies, as shown in Fig. 1(d). To complement MOS2D, we further introduce a Dual Degradation Estimation Module (DDEM), designed to extract two complementary forms of degradation information from the input: (i) a global degradation representation that encapsulates high-level weather characteristics such as type and severity, and (ii) a set of spatially adaptive kernels that encode local degradation structures and variations. These dual degradation representations are then utilized to dynamically modulate the restoration process. Specifically, the global representation adaptively influences feature transformations within MODEM, while the spatially adaptive kernels guide the matrix construction within MOS2D, refining the spatial dependencies captured along the Morton-order sequence. This dual-modulation strategy equips the restoration network with both global awareness and local sensitivity, thereby delivering more precise and effective restoration. Our primary contributions can be summarized as follows:

- We propose a novel Morton-Order Degradation Estimation Mechanism that introduces the MOS2D module, which integrates Morton-coded spatial ordering with selective state space modeling to effectively capture spatially heterogeneous weather degradation dynamics.
- We design a Dual Degradation Estimation Module that jointly estimates global degradation descriptors and spatially adaptive kernels. These two complementary representations are used to dynamically modulate MOS2D, allowing contextually aware and spatially adaptive restoration.

- Extensive experiments and ablation studies are conducted to demonstrate the effectiveness of the proposed MODEM, and its superiority over other state-of-the-art competitors in restoring images under diverse and complex adverse weather conditions.

2 Related Work

This section briefly reviews representative approaches to single adverse weather restoration including rain streak removal, raindrop removal, haze removal and snow removal, as well as unified all-in-one models. Additionally, recent advances in SSM-based image restoration techniques are also discussed.

Rain Streak Removal. Traditional rain streak removal methods typically relied on image decomposition [32] or tensor-based priors [30]. With the advent of deep learning, various models have emerged, like the deep detail network for splitting rain details [16], recurrent networks for context aggregation in single images [41], and spatio-temporal aggregation in videos [42]. Further developments include uncertainty-guided multi-scale designs [78], density-aware architectures [83], spatial attention [69], joint detection-removal frameworks [75], and NAS-based attentive schemes [5].

Raindrop Removal. Early efforts addressed specific scenarios [15] or adherent raindrops in videos [80]. Deep learning subsequently provided more generalizable solutions, including learning from synthetic photorealistic data [22], using attention-guided GANs for realistic inpainting [57], and dedicated networks for visibility through raindrop-covered windows [58]. General restoration architectures like Dual Residual Networks [43] and Adaptive Sparse Transformers [92] are also applicable, with advanced dual attention-in-attention models [86] tackling joint rain streak and raindrop removal.

Haze Removal. Traditional image dehazing methods often utilized statistical priors the Dark Channel Prior (DCP) [23], non-local similarity techniques [4], or multi-scale fusion [2]. But these approaches faced robustness issues in diverse hazy scenes due to strong underlying assumptions. Deep learning has since become dominant, with methods focusing on estimating transmission and atmospheric light, sometimes using depth awareness [76] or unpaired learning via decomposition [77]. Other approaches tackle haze density variations [83, 85], domain adaptation [62], and contrastive learning [71].

Snow Removal. While specific traditional priors for snow removal are less common compared to other weather conditions, principles from general image restoration were often adapted. Deep learning solutions have gained traction, including powerful general-purpose backbones [13, 12] and more specialized networks. Desnowing-specific models often involve context-aware designs [45] or leverage semantic and depth priors [87]. Techniques inspired by classical matrix decomposition [59] and methods for online video processing [38] have also been integrated into deep frameworks.

While these specialized models excel under specific weather conditions, their limited generalization requires separate models for each degradation type, hindering real-world practicality.

Unified Adverse Weather Restoration. To mitigate the inefficiency of deploying multiple models for complex weather conditions, unified models have been devised [95, 67, 11, 40, 55, 65, 79]. However, these unified ones face a fundamental challenge that *the degradation space covered by unified models is exponentially larger than that of weather-specific counterparts*, substantially increasing the complexity of accurate modeling and generalization. To tackle this, existing approaches employ different strategies to learn more generalizable representations. Early efforts included using NAS to find a unified structure [40] and leveraging Transformers like TransWeather [67] for their global context modeling. Others focused on training strategies, such as two-stage knowledge learning with multi-contrastive regularization [11]. More recent works have explored disentangling weather-general and weather-specific features [95]), adapting powerful generative models like diffusion models [55]), incorporating learned codebook priors [79]), and enhancing Transformers with global image statistics like Histoformer [65]. These methods showcase a trend towards more sophisticated and adaptive unified restoration. Despite progress, the unified models still struggle to effectively capture the diverse and spatially heterogeneous characteristics of different weather phenomena.

SSM-based Image Restoration. Recent State Space Models (SSMs) [18, 17], notably Mamba [17] with its selective scan mechanism, offer efficient long-sequence modeling and have rapidly permeated computer vision [52, 94, 44, 48, 64]. Inspired by these advancements, SSMs are increasingly applied to image restoration. General frameworks like MambaIR [20], VMambaIR [63], and CU-Mamba [14] demonstrate the potential of Mamba-based models as strong baselines. Specific low-level vision applications have also seen tailored solutions. For example, WaterMamba [19] has been developed

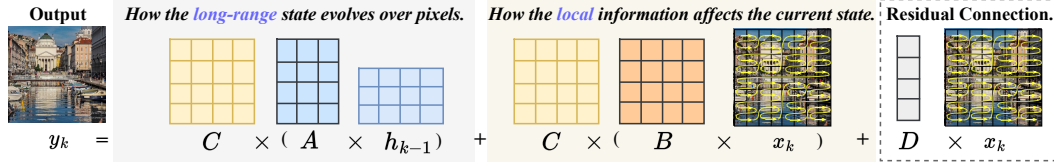


Figure 2: Connection between MODEM and SSM.

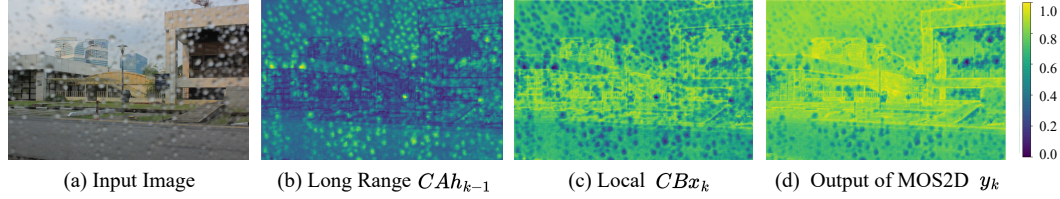


Figure 3: With respect to a sample (a), (b)-(d) visualize the long-range CAh_{k-1} , (c) local CBx_k , and (d) output of MOS2D y_k , respectively. More cases can be found in Appendix A.4

for underwater image enhancement, and IRSRMamba [26] addresses infrared super-resolution. Low-light image enhancement has been tackled by models like RetinexMamba [3] and LLEMamba [89], while efficient SSM designs have been proposed for image deblurring [35]. In image deraining, notable works include FourierMamba [37] and FreqMamba [91]. These efforts underscore the growing success of SSMs in addressing image restoration challenges. Our work builds upon this trend, leveraging SSMs for adaptive modeling of spatially heterogeneous weather degradations.

3 Bridging Image Degradation Estimation and State Space Modeling

Restoring images corrupted by adverse weather is inherently a highly ill-posed problem. In general, the relationship between an observed degraded image x and its clean version y can be modeled as:

$$x = \text{Degrade}(y, \theta := \{\theta_G, \theta_L\}) \Leftrightarrow y = \text{Restore}(x, \theta := \{\theta_G, \theta_L\}), \quad (1)$$

where $\text{Degrade}(\cdot, \theta)$ designates a complex, often spatially varying weather degradation function parameterized by θ . By considering the spatial influence of degradation, the parameter set (θ) can be partitioned into long-range/global degradation (θ_G , *e.g.* atmospheric haze) and local one (θ_L , *e.g.* rain streak). The objective is to recover the latent clean y from the degraded observation x . The core challenge lies in the fact that both the degradation process and the artifacts are unknown and can vary significantly across scenes and weather types. Thus, estimating the degradation characteristic θ and constructing the function $\text{Restore}(\cdot, \theta)$ in Eq. (1), is central to solving the problem.

We propose that State Space Models (SSMs) [17] offer a suitable and robust framework for this degradation estimation task. Recall the core recurrence of a discrete-time SSM:

$$h_k = Ah_{k-1} + Bx_k, \quad y_k = Ch_k + Dx_k, \quad (2)$$

where x_k is the input Morton-order sequence at step k , h_k is the latent hidden state summarizing historical context, and y_k is the output. The matrices A , B , and C are learnable parameters that define the SSM's dynamics and output mapping. Please notice that the term Dx_k is a practical (non-theoretical) addition for training ease, which is analogous to residual connection. To better understand how the state dynamics contribute to the output y_k , we can analyze the system by omitting the residual Dx_k and substitute h_k into the $y_k = Ch_k$ part, as follows:

$$y_k = CAh_{k-1} + CBx_k. \quad (3)$$

By comparing Eq. (1) with Eq. (3), we interpret SSM components as degradation estimators, as illustrated in Fig. 2. The term CAh_{k-1} conceptually models the long-range, spatially propagated degradation context, as shown in Fig. 3(b). The previous hidden state h_{k-1} summarizes accumulated degradation cues (*e.g.*, overall haze level, rain intensity) along the Morton-order scan path. The state transition matrix A then models how these summarized characteristics evolve or persist as the scan progresses spatially (*e.g.*, the smooth attenuation of visibility due to widespread haze or the consistent statistical properties of rain streaks over a larger area.). It essentially dictates “how the long-range state evolves over pixels” as noted in Fig. 2. The output matrix C then translates this evolved context Ah_{k-1} into a contribution to the current output y_k , reflecting broader, non-local degradation patterns.

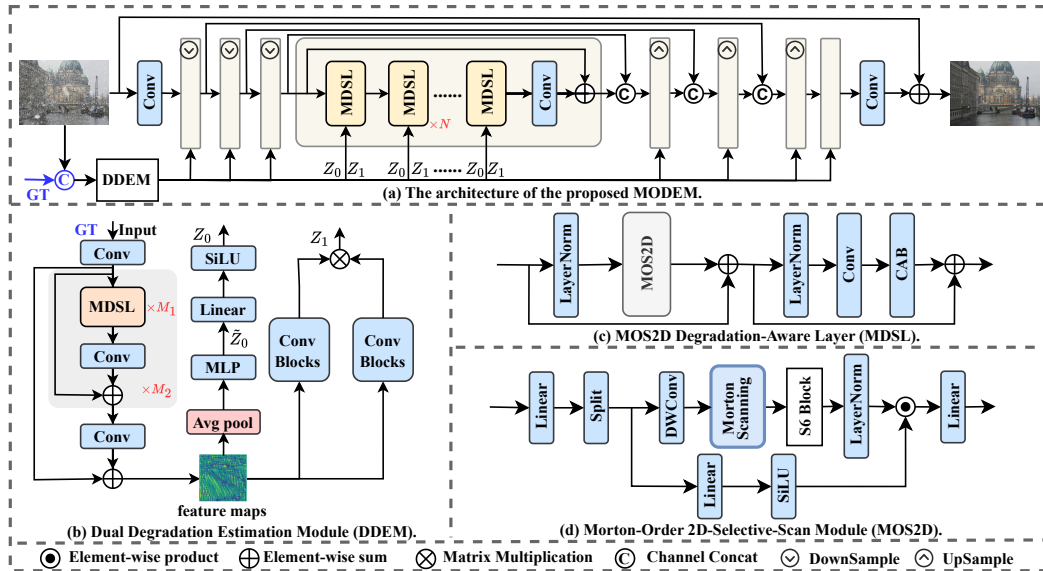


Figure 4: (a) Overall architecture of MODEM. (b) The DDEM for extracting global descriptor Z_0 and adaptive degradation kernel Z_1 degradation priors. (c) The MDSL incorporating the core MOS2D module (d) within a residual block. The blue-colored components indicate elements exclusive to the first training stage. N , M_1 , M_2 denote the number of the corresponding module, respectively.

Concurrently, the $C B x_k$ term accounts for the impact of the immediate, local input features x_k , as visualized in Fig. 3(c). The input matrix B determines “how the local information affects the current state” (Fig. 2), allowing features from x_k (e.g., local degradation artifacts like dense fog patches, or crucial local image content like fine textures/sharp edges) to directly influence the current hidden state h_k . This enables the SSM to react to fine-grained local evidence, with C projecting this locally-informed component into y_k for specific adjustments based on immediate observations.

4 Morton-Order Degradation Estimation Mechanism

The Morton-Order Degradation Estimation Mechanism (MODEM), depicted in Fig. 4(a), employs a two-stage training strategy designed to accurately learn degradation characteristics.

Stage 1: In the first stage, the Dual Degradation Estimation Module (DDEM), shown in Fig. 4(b), is provided with both the degraded image I_{LQ} and its corresponding ground-truth I_{GT} from the training set. This allows the DDEM to explicitly learn the mapping from an image pair to its underlying degradation patterns. It outputs two degradation priors: a global descriptor Z_0 and a spatially adaptive degradation kernel Z_1 . These priors, representing the degradation information, are then injected into MODEM’s main restoration backbone. The backbone itself, comprised of N stacked MOS2D Degradation-Aware Layers (MDSLs) detailed in Fig. 4(c), only receives the degraded image I_{LQ} and uses these priors to perform adaptive restoration. Each MDSL leverages the Morton-Order 2D-Selective-Scan module (MOS2D), shown in Fig. 4(d) and Fig. 5, to adaptively modulate features based on the degradation priors Z_0 and Z_1 . The degradation representation learned by the DDEM in this stage serves as the supervisory target for the Stage 2.

Stage 2: In the second stage, the inputs to both the DDEM and the main backbone consist of only the degraded image I_{LQ} , without any GT. The Stage 2 model inherits its parameters from Stage 1, and the DDEM in this stage receives only the degraded image I_{LQ} as input. Simultaneously, there is a frozen DDEM receiving both the GT and the degraded image. The degradation information from this frozen DDEM is then used to supervise the trainable DDEM.

4.1 Dual Degradation Estimation Module (DDEM)

The Dual Degradation Estimation Module (DDEM), shown in Fig. 4(b), extracts global degradation descriptor Z_0 and adaptive degradation kernel Z_1 . In the first stage, DDEM processes the degraded image I_{LQ} and ground-truth I_{GT} . Otherwise, it uses only I_{LQ} . The input undergoes several

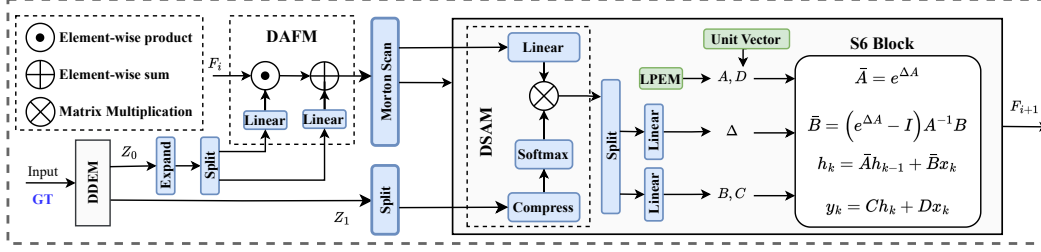


Figure 5: Detailed illustration of the degradation modulation mechanism within a MOS2D module in the main restoration backbone, which employs the Degradation-Adaptive Feature Modulation (DAFM) and Degradation-Selective Attention Modulation (DSAM) to dynamically adjust feature representations based on the degradation priors Z_0 and Z_1 .

MOS2D Degradation-Aware Layers (MDSLs) to extract degradation feature map $F \in \mathbb{R}^{H \times W \times C_d}$. Subsequently, the global descriptor Z_0 and adaptive kernels Z_1 are derived from F as follows:

$$\tilde{Z} = \text{MLP}(\text{AvgPool}(F)), \quad Z_0 = \sigma(\text{Linear}(\tilde{Z})), \quad Z_1 = \text{Conv}(F) \times \text{Conv}(F)^T, \quad (4)$$

where $\sigma(\cdot)$ denotes the SiLU activation function, $\tilde{Z} \in \mathbb{R}^{4C_d}$ will be used for degradation supervision. These priors $Z_0 \in \mathbb{R}^{C_d}$ and $Z_1 \in \mathbb{R}^{C_{d1} \times C_{d2}}$ then guide the main restoration network.

MOS2D Degradation-Aware Layers (MDSL). The iterative application of MDSL enhances sensitivity to degradation patterns, yielding enriched feature maps. For the feature map F_i at the i -th MDSL, this process of MDSL can be formulated as:

$$\tilde{F}_i = \text{MOS2D}(\text{LN}(F_i)) + F_i, \quad F_{i+1} = \text{CAB}(\text{Conv}(\text{LN}(\tilde{F}_i))) + \tilde{F}_i, \quad (5)$$

where $\text{CAB}(\cdot)$ denotes the Channel Attention Block. $\text{LN}(\cdot)$ represents Layer Normalization.

4.2 Morton-Order 2D-Selective-Scan Module (MOS2D)

To effectively model spatially heterogeneous degradations in 2D images, our MOS2D employs Morton-Order scan, as illustrated in Fig. 1(d). This converts 2D spatial features into locality-preserving 1D sequence, facilitating structured feature interaction and aggregation by SSM.

Specifically, the Morton encoding z maps 2D pixel coordinates (i, j) , where $0 \leq i < H, 0 \leq j < W$, to a 1D index by interleaving the bits of their binary representations. For $i = (i_n, \dots, i_0)_2$ and $j = (j_n, \dots, j_0)_2$, with $n = \lceil \log_2(\max(H, W)) \rceil - 1$, the encoding z is:

$$z = \text{interleave}(i, j) = (j_n, i_n, j_{n-1}, i_{n-1}, \dots, j_1, i_1, j_0, i_0)_2. \quad (6)$$

In the DDEM, Morton-order coding is followed by standard SSM operations to effectively extract global degradation descriptor Z_0 and adaptive degradation kernel Z_1 . In contrast, in our main restoration backbone, the MOS2D module employs degradation-aware modulations using Z_0 and Z_1 through DAFM and DSAM, detailed in Fig. 5. This dual-modulation ensures that the MOS2D is dynamically conditioned on both long range contextually aware and spatially adaptive restoration.

Degradation-Adaptive Feature Modulation (DAFM). The i -th layer's input feature map F_i is first modulated by the global degradation descriptor Z_0 . As shown in Fig. 5, Z_0 is expanded and split to produce channel-wise adaptive weights Z_0^w and biases Z_0^b , which applied to F_i using a feature-wise linear modulation operation, thus incorporates global degradation characteristics:

$$F_{\text{DAFM}} = (Z_0^w \odot F_i) + Z_0^b, \quad Z_0^w, Z_0^b = \text{Split}(\text{Linear}(Z_0)), \quad (7)$$

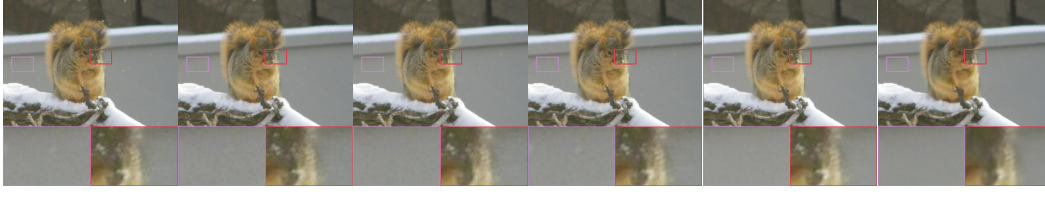
where $\odot$ denotes element-wise multiplication.

Degradation-Selective Attention Modulation (DSAM). To further guide the S6 Block with local degradation information, the spatially adaptive kernel Z_1 is utilized as follows:

$$F_{\text{DSAM}} = W_F F_{\text{DAFM}} \times \text{Softmax}(W_Z Z_1), \quad (8)$$

where W_F and W_Z are learnable linear projection matrices.

Degradation-Guided S6 Block. The core selective scan operation is performed by our Degradation-Guided S6 Block. Its parameters are dynamically adapted based on the degradation-sensitive features



(a) Input (b) WGWSNet (c) WeatherDiff (d) Histoformer (e) Histoformer† (f) Ours

Figure 6: Visual comparison on the real world snow dataset [45]. Compared to the prior methods [95, 55, 65], where “†” denotes the Histoformer [65] for real snow from their official repository, our MODEM achieves superior results without additional training.

Table 1: Quantitative comparison with recent state-of-the-art unified methods [40, 67, 11, 95, 55, 79, 65] across various datasets. The best and second-best results are in **bold** and underlined, respectively.

Methods	Snow100K-S		Snow100K-L		Outdoor		RainDrop		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
All-in-One [40]	-	-	28.33	0.8820	24.71	0.8980	31.12	0.9268	28.05	0.9023
TransWeather [67]	32.51	0.9341	29.31	0.8879	28.83	0.9000	30.17	0.9157	30.21	0.9094
Chen <i>et al.</i> [11]	34.42	0.9469	30.22	0.9071	29.27	0.9147	31.81	0.9309	31.43	0.9249
WGWSNet [95]	34.31	0.9460	30.16	0.9007	29.32	0.9207	32.38	0.9378	31.54	0.9263
WeatherDiff ₆₄ [55]	35.83	0.9566	30.09	0.9041	29.64	0.9312	30.71	0.9312	31.57	0.9308
WeatherDiff ₁₂₈ [55]	35.02	0.9516	29.58	0.8941	29.72	0.9216	29.66	0.9225	31.00	0.9225
AWRCP [79]	36.92	0.9652	31.92	0.9341	31.39	0.9329	31.93	0.9314	33.04	0.9409
Histoformer [65]	<u>37.41</u>	<u>0.9656</u>	<u>32.16</u>	0.9261	<u>32.08</u>	<u>0.9389</u>	33.06	0.9441	<u>33.68</u>	<u>0.9437</u>
MODEM (Ours)	38.08	0.9673	32.52	<u>0.9292</u>	33.10	0.9410	<u>33.01</u>	<u>0.9434</u>	34.18	0.9452

F_{DSAM} derived from the DSAM stage. F_{DSAM} is split into three components, which are then linearly transformed to generate the SSM parameters B , C , and Δ :

$$F_{\Delta}, F_B, F_C = \text{Split}(F_{\text{DSAM}}), \quad \Delta = W_{\Delta} F_{\Delta}, \quad B = W_B F_B, \quad C = W_C F_C, \quad (9)$$

where W_{Δ} , W_B , and W_C are learnable linear mappings. We use zero-order hold (ZOH) discretization with timescale Δ to obtain the discrete matrices $A = e^{\Delta A}$ and $\bar{B} = (\exp(\Delta A) - I)A^{-1}B$, where I is the identity matrix. The S6 Block then processes the Morton-Ordered sequence x_k , which is the Morton-ordered sequence derived from the DAFM features F_{DAFM} , as follows:

$$y_k = Ch_k + Dx_k, \quad h_k = \bar{A}h_{k-1} + \bar{B}x_k, \quad x_k = F_{\text{DAFM}}[z], \quad (10)$$

where h_k is the hidden state at step k . The dynamically generated B and C (and Δ influencing $\bar{A}$, $\bar{B}$) ensure that the state evolution and output generation are adaptive to the degradation characteristics.

4.3 Loss Function

Our model is trained in two stages. In the first stage, the loss $\mathcal{L}_{\text{stage1}}$ combines a $\mathcal{L}_1$ loss and a correlation loss $\mathcal{L}_{\text{cor}}$ [65] to ensure accuracy and structural fidelity between the output of MODEM I_{HQ} and ground-truth I_{GT} :

$$\mathcal{L}_{\text{stage1}} = \mathcal{L}_1(I_{\text{HQ}}, I_{\text{GT}}) + \mathcal{L}_{\text{cor}}(I_{\text{HQ}}, I_{\text{GT}}), \quad (11)$$

The correlation loss $\mathcal{L}_{\text{cor}}(I_{\text{HQ}}, I_{\text{GT}})$ [65] is based on the Pearson correlation coefficient $\rho(I_{\text{HQ}}, I_{\text{GT}})$:

$$\mathcal{L}_{\text{cor}}(I_{\text{HQ}}, I_{\text{GT}}) = \frac{1}{2} (1 - \rho(I_{\text{HQ}}, I_{\text{GT}})), \quad \rho(I_{\text{HQ}}, I_{\text{GT}}) = \frac{\sum_{i=1}^N (I_{i,\text{HQ}} - \bar{I}_{\text{HQ}})(I_{i,\text{GT}} - \bar{I}_{\text{GT}})}{N \cdot \sigma(I_{\text{HQ}}) \cdot \sigma(I_{\text{GT}})}. \quad (12)$$

where N is the total number of pixels, $\bar{I}$ denotes mean, and $\sigma(I)$ denotes standard deviation.

In the second stage, we introduce a KL divergence loss $\mathcal{L}_{\text{KL}}$ to maintain consistency of the degradation representation $\tilde{Z}$ across stages. Let $\tilde{Z}_{0,\text{st1}}$ and $\tilde{Z}_{0,\text{st2}}$ be the $\tilde{Z}$ vectors from the first (fixed) and second stages, respectively. The KL divergence is computed between their softmax distributions $\phi(\cdot)$:

$$\mathcal{L}_{\text{KL}} = D_{\text{KL}} \left(\phi(\tilde{Z}_{0,\text{st1}}) \parallel \phi(\tilde{Z}_{0,\text{st2}}) \right) = \sum_{j=1}^{4C_d} \phi(\tilde{Z}_{0,\text{st1}}(j)) \log \left(\frac{\phi(\tilde{Z}_{0,\text{st1}}(j))}{\phi(\tilde{Z}_{0,\text{st2}}(j))} \right). \quad (13)$$

Table 2: Desnowing Task

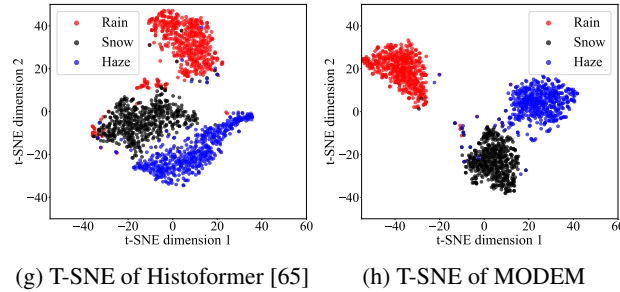
Methods	Snow100K-S		Snow100K-L	
	PSNR	SSIM	PSNR	SSIM
SPANet [69]	29.92	0.8260	23.70	0.7930
JSTASR [10]	31.40	0.9012	25.32	0.8076
RESCAN [41]	31.51	0.9032	26.08	0.8108
DesnowNet [45]	32.33	0.9500	27.17	0.8983
DDMSNet [87]	34.34	0.9445	28.85	0.8772
Restormer [82]	36.01	0.9579	30.36	0.9068
ConvIR [13]	<u>37.98</u>	<u>0.9686</u>	<u>32.11</u>	0.9300
FSNet [12]	37.42	0.9654	31.62	0.9246
MODEM (Ours)	38.08	0.9673	32.52	<u>0.9292</u>

Table 3: Raindrop Removal Task

Methods	RainDrop	
	PSNR	SSIM
pix2pix [27]	28.02	0.8547
DuRN [43]	31.24	0.9259
RaindropAttn [58]	31.44	0.9263
AttentiveGAN [57]	31.59	0.9170
IDT [73]	31.87	0.9313
MAXIM [66]	31.87	0.9352
Restormer [82]	<u>32.18</u>	<u>0.9408</u>
AST [92]	30.57	0.9333
MODEM (Ours)	33.01	0.9434

Table 4: Deraining & Dehazing Task

Methods	Outdoor-Rain	
	PSNR	SSIM
CycleGAN [93]	17.62	0.6560
pix2pix [27]	19.09	0.7100
HRGAN [39]	21.56	0.8550
PCNet [29]	26.19	0.9015
MPRNet [81]	28.03	0.9192
NAFNet [9]	29.59	0.9027
Restormer [82]	<u>30.03</u>	<u>0.9215</u>
MODEM (Ours)	33.10	0.9410



(g) T-SNE of Histoformer [65] (h) T-SNE of MODEM

Figure 7: T-SNE results of Histoformer [65] and MODEM, which reflect that MODEM exhibits better clustering of features corresponding to different weather types than Histoformer [65].

The total loss for the second stage, $\mathcal{L}_{\text{stage2}}$, can be formulated as:

$$\mathcal{L}_{\text{stage2}} = \mathcal{L}_1(I_{\text{HQ}}, I_{\text{GT}}) + \mathcal{L}_{\text{cor}}(I_{\text{HQ}}, I_{\text{GT}}) + \mathcal{L}_{\text{KL}}(\tilde{Z}_{0,\text{st1}}, \tilde{Z}_{0,\text{st2}}). \quad (14)$$

5 Experimental Validation

Our MODEM is implemented in PyTorch [56] and trained on 4 NVIDIA RTX 3090 GPUs in two stages. We employ the AdamW optimizer [47] with a Cosine Annealing Restart Cyclic LR scheduler [46]. It is trained on an all-weather dataset [67, 65], consistent with prior works [40, 67, 55, 65, 95, 79]. For evaluation, MODEM is tested on established benchmarks including Test1 [39, 40], RainDrop [57], Snow100K-L/S [45], and a real-world snow test set [45]. Due to the page limit, more details are given in Appendix A.1 and A.2.

5.1 Comparisons with State-of-the-arts

Quantitative Comparison. Our quantitative evaluation assesses MODEM against both unified models [40, 67, 11, 95, 55, 79, 65] and task-specific methods [69, 10, 41, 45, 87, 9, 82, 43, 27, 58, 57, 73, 66, 93, 39, 29, 81]. As reported in Table 1, MODEM achieves SOTA performance, with an average PSNR improvement of 0.5dB across benchmarks [39, 40, 57, 45], and a notable 1.02dB gain on Outdoor-Rain [39, 40]. While slightly below Histoformer on RainDrop [57] by 0.05dB PSNR, MODEM yields compelling visual results, a point further elaborated in our qualitative comparisons. Furthermore, comparisons with leading task-specific methods in Tables 2 to 4 confirm MODEM consistently attains SOTA results. These evaluations underscore MODEM’s capability to effectively estimate and adapt to diverse, spatially heterogeneous weather degradations. Additionally, the comparisons of complexity can be found in Appendix A.3.

Qualitative comparison. We provide qualitative comparisons across diverse scenarios in Figs. 6 and 8. For image desnowing on Snow100K [45], MODEM effectively removes heavy snowflakes and visual artifacts that other models [95, 55, 65] struggle to address. For joint deraining and dehazing on Outdoor-Rain [39, 40], MODEM excels in restoring richer texture details and produces images with noticeably fewer artifacts. For raindrop removal on the Raindrop [57], MODEM again

Table 5: Comparison of perceptual metrics, including referenced (LPIPS $\downarrow$) and non-referenced (Q-Align $\uparrow$, MUSIQ $\uparrow$) scores. Best results are **bolded**; second-best are underlined.

Method		Snow100K-L	Snow100K-S	Outdoor	Raindrop	Snow100K-Real
LPIPS	Histoformer [65]	<u>0.0919</u>	<u>0.0445</u>	<u>0.0778</u>	0.0672	–
	WeatherDiff [55]	0.0982	0.0541	0.0887	0.0615	–
	MODEM (Ours)	0.0880	0.0407	0.0699	<u>0.0650</u>	–
Q-Align	Histoformer [65]	<u>3.7207</u>	<u>3.7598</u>	<u>4.1445</u>	<u>4.0156</u>	<u>3.5449</u>
	WeatherDiff [55]	3.4531	3.5293	3.8691	4.0000	3.4512
	MODEM (Ours)	3.7324	3.7695	4.1875	4.0664	3.5586
MUSIQ	Histoformer [65]	64.2526	<u>64.2581</u>	<u>67.7461</u>	68.4852	59.4040
	WeatherDiff [55]	62.6267	<u>63.1278</u>	<u>67.4814</u>	69.3608	59.4493
	MODEM (Ours)	<u>64.2438</u>	64.2853	68.2926	69.7925	59.6042

Table 6: Comparison of different methods on various real-world datasets using the Q-Align metric.

Method	Snow100K-Real	RainDrop	NTURain	RESIDE	WeatherStream
WeatherDiff [55]	3.4531	4.0000	3.2031	3.4219	<u>1.9561</u>
Histoformer [65]	<u>3.7207</u>	<u>4.0156</u>	<u>3.2266</u>	3.2891	1.9434
MODEM (Ours)	3.7324	4.0664	3.2891	<u>3.3164</u>	1.9863

demonstrates superior detail preservation and artifact reduction. Furthermore, on real-world snowy images [45], MODEM achieves superior results as shown in Fig. 6 even without any additional fine-tuning, showcasing excellent generalization and real-world applicability, which can be attributed to its profound understanding and adaptive modeling of degradation characteristics. Due to the page limit, more visual comparisons can be found in Appendix A.8.

5.2 Perceptual Quality and Real-World Performance

Comparison of perceptual metrics. We report referenced (LPIPS [88]) and non-referenced (Q-Align [72], MUSIQ [33]) scores, all computed using the pyiqa library [6], with the same settings applied for all methods. As shown in Table 5, our method achieves SOTA perceptual quality.

Comparison of real-world datasets. To provide quantitative evidence of our model’s performance on the challenging real-world scenarios, we evaluated our model on the real-world data from the Snow100K-Real [45], RainDrop [57], RESIDE [36], NTURain [8] and WeatherStream [84] datasets using the Q-Align [72]. The results are presented in Table 6. As the results show, our method achieves state-of-the-art performance on real-world datasets when compared to the previous state-of-the-art method, Histoformer [65], and the diffusion-based approach, WeatherDiff [55].

5.3 Ablation Studies

We evaluate the impact of the Morton-Order scan, DDEM, DAFM and DSAM. Morton-Order scan facilitates structured spatial feature processing within the SSM, enabling the model to better capture contextual dependencies and preserve local structural coherence. As shown in Table 7, configurations incorporating the Morton scan generally yield improved performance, underscoring its benefit in organizing spatial information for sequential modeling. DDEM provides the degradation priors Z_0 and Z_1 that guide the DAFM and DSAM. As indicated in Table 7, configurations lacking DDEM exhibit a significant performance drop. DAFM, which utilizes the global degradation prior Z_0 , plays a crucial role in adaptively modulating features based on the overall estimated weather type and severity. DSAM, guided by the spatially adaptive kernel Z_1 , allows the MOS2D to selectively focus on and modulate features pertinent to local degradation characteristics or specific image regions. While removing DSAM shows marginal gains on certain snow metrics, it impairs performance on tasks like raindrop removal, which demands strong local adaptation due to the highly local nature of raindrops, and on mixed rain&haze scenarios where intricate local texture recovery is challenging. Thus, DSAM, guided by the spatially adaptive kernel Z_1 , is crucial for achieving robust, balanced performance across diverse weather conditions by guiding the network to selectively address these local characteristics. The combination of all components employed in full MODEM, consistently



Figure 8: Visual comparisons of MODEM against state-of-the-art unified methods across diverse adverse weather restoration tasks. **Top Row**: Image desnowing. **Middle Row**: Joint deraining and dehazing. **Bottom Row**: Raindrop removal.

Table 7: Ablation study results across different datasets and factor combinations.

Factors				Snow100K-S		Snow100K-L		Outdoor		RainDrop	
Morton	DDEM	DAFM	DSAM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
✗	✓	✓	✓	38.03	0.9672	32.33	0.9283	32.89	0.9399	32.69	0.9423
✓	✗	N/A	N/A	37.61	0.9650	32.12	0.9249	32.37	0.9224	32.38	0.9390
✓	✓	✓	✗	38.15	0.9675	32.59	0.9298	32.77	0.9404	32.72	0.9435
✓	✓	✗	✓	37.43	0.9643	32.07	0.9240	32.19	0.9308	32.62	0.9387
✓	✓	✓	✓	38.08	0.9673	32.52	0.9292	33.10	0.9410	33.01	0.9434

yields the best results. This result is further supported by t-SNE [68] in Fig. 7. Compared to Histoformer [65], MODEM exhibits significantly improved clustering of features corresponding to different weather types. This indicates that the combined action of the components enables MODEM to learn more discriminative and well-separated feature representations, which directly contributes to enhance restoration capabilities. More ablation studies can be found in Appendix A.6.

6 Conclusion

In this paper, we introduced the Morton-Order Degradation Estimation Mechanism (MODEM) to address the challenge of restoring images degraded by diverse and spatially heterogeneous adverse weather. MODEM integrates a Dual Degradation Estimation Module (DDEM) for extracting global and local degradation priors, with a Morton-Order 2D-Selective-Scan Module (MOS2D) that employs Morton-coded spatial ordering and selective state-space models for adaptive, context-aware restoration. Extensive experiments demonstrate MODEM’s state-of-the-art performance across multiple benchmarks and weather types. This superiority is attributed to its effective modeling of complex degradation dynamics via explicit degradation estimation guiding the restoration process, leading to more discriminative feature representations and strong generalization to real-world scenarios.

Acknowledgement

We would like to thank Mingjia Li for the insightful discussions and feedback. This work was partially supported by the National Natural Science Foundation of China under Grant no. 62372251. The computational resources of this work was partially supported by TPU Research Cloud (TRC).

References

- [1] Yasin Almalioglu, Mehmet Turan, Niki Trigoni, and Andrew Markham. Deep learning-based robust positioning for all-weather autonomous driving. *Nature machine intelligence*, 4(9): 749–760, 2022.
- [2] Codruta O. Ancuti and Cosmin Ancuti. Single image dehazing by multi-scale fusion. *IEEE Trans. Image Process.*, 22(8):3271–3282, 2013.
- [3] Jiesong Bai, Yuhao Yin, and Qiyuan He. Retinexmamba: Retinex-based mamba for low-light image enhancement. *arXiv preprint arXiv:2405.03349*, 2024.
- [4] Dana Berman, Tali Treibitz, and Shai Avidan. Non-local image dehazing. In *CVPR*, pages 1674–1682. IEEE Computer Society, 2016.
- [5] Lei Cai, Yuli Fu, Wanliang Huo, Youjun Xiang, Tao Zhu, Ying Zhang, Huanqiang Zeng, and Delu Zeng. Multiscale attentive image de-raining networks via neural architecture search. *IEEE Trans. Circuits Syst. Video Technol.*, 33(2):618–633, 2023. doi: 10.1109/TCSVT.2022.3207516. URL <https://doi.org/10.1109/TCSVT.2022.3207516>.
- [6] Chaofeng Chen and Jiadi Mo. IQA-PyTorch: Pytorch toolbox for image quality assessment. [Online]. Available: <https://github.com/chaofengc/IQA-PyTorch>, 2022.
- [7] Guangyan Chen, Meiling Wang, Yi Yang, Kai Yu, Li Yuan, and Yufeng Yue. Pointgpt: Auto-regressively generative pre-training from point clouds. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.
- [8] Jie Chen, Cheen-Hau Tan, Junhui Hou, Lap-Pui Chau, and He Li. Robust video content alignment and compensation for rain removal in a cnn framework. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6286–6295, 2018.
- [9] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022.
- [10] Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, Cheng-Che Tsai, and Sy-Yen Kuo. Jstasr: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*, pages 754–770. Springer, 2020.
- [11] Wei-Ting Chen, Zhi-Kai Huang, Cheng-Che Tsai, Hao-Hsiang Yang, Jian-Jiun Ding, and Sy-Yen Kuo. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In *CVPR*, pages 17632–17641. IEEE, 2022.
- [12] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Image restoration via frequency selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(2):1093–1108, 2023.
- [13] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Revitalizing convolutional network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12): 9423–9438, 2024.
- [14] Rui Deng and Tianpei Gu. Cu-mamba: Selective state space models with channel learning for image restoration. *arXiv preprint arXiv:2404.11778*, 2024.
- [15] David Eigen, Dilip Krishnan, and Rob Fergus. Restoring an image taken through a window covered with dirt or rain. In *Proceedings of the IEEE international conference on computer vision*, pages 633–640, 2013.
- [16] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John W. Paisley. Removing rain from single images via a deep detail network. In *CVPR*, pages 1715–1723. IEEE Computer Society, 2017.

- [17] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.
- [18] Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021.
- [19] Meisheng Guan, Haiyong Xu, Gangyi Jiang, Mei Yu, Yeyao Chen, Ting Luo, and Yang Song. Watermamba: Visual state space model for underwater image enhancement. *arXiv preprint arXiv:2405.08419*, 2024.
- [20] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. *arXiv preprint arXiv:2402.15648*, 2024.
- [21] Xiaojie Guo, Yang Yang, Chaoyue Wang, and Jiayi Ma. Image dehazing via enhancement, restoration, and fusion: A survey. *Inf. Fusion*, 86-87:146–170, 2022.
- [22] Zhixiang Hao, Shaodi You, Yu Li, Kunming Li, and Feng Lu. Learning from synthetic photorealistic raindrop for single image raindrop removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019.
- [23] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *CVPR*, pages 1956–1963. IEEE Computer Society, 2009.
- [24] Vincent Tao Hu, Stefan Andreas Baumann, Ming Gui, Olga Grebenkova, Pingchuan Ma, Johannes Fischer, and Björn Ommert. Zigma: A dit-style zigzag mamba diffusion model. In *European conference on computer vision*, pages 148–166. Springer, 2024.
- [25] Tao Huang, Xiaohuan Pei, Shan You, Fei Wang, Chen Qian, and Chang Xu. Localmamba: Visual state space model with windowed selective scan. *arXiv preprint arXiv:2403.09338*, 2024.
- [26] Yongsong Huang, Tomo Miyazaki, Xiaofeng Liu, and Shinichiro Omachi. Irsrmamba: Infrared image super-resolution via mamba-based wavelet transform feature modulation model. *arXiv preprint arXiv:2405.09873*, 2024.
- [27] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
- [28] Min Wu Jeong and Chae Eun Rhee. Lc-mamba: Local and continuous mamba with shifted windows for frame interpolation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 17671–17681, 2025.
- [29] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Zheng Wang, Xiao Wang, Junjun Jiang, and Chia-Wen Lin. Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining. *IEEE Transactions on Image Processing*, 30:7404–7418, 2021.
- [30] Tai-Xiang Jiang, Ting-Zhu Huang, Xi-Le Zhao, Liang-Jian Deng, and Yao Wang. A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors. In *CVPR*, pages 2818–2827. IEEE Computer Society, 2017.
- [31] Xin Jin, Haisheng Su, Kai Liu, Cong Ma, Wei Wu, Fei Hui, and Junchi Yan. Unimamba: Unified spatial-channel representation learning with group-efficient mamba for lidar-based 3d object detection. *arXiv preprint arXiv:2503.12009*, 2025.
- [32] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Trans. Image Process.*, 21(4):1742–1755, 2012. doi: 10.1109/TIP.2011.2179057. URL <https://doi.org/10.1109/TIP.2011.2179057>.
- [33] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5148–5157, 2021.
- [34] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

- [35] Lingshun Kong, Jiangxin Dong, Ming-Hsuan Yang, and Jinshan Pan. Efficient visual state space model for image deblurring. *arXiv preprint arXiv:2405.14343*, 2024.
- [36] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE transactions on image processing*, 28(1):492–505, 2018.
- [37] Dong Li, Yidi Liu, Xueyang Fu, Senyan Xu, and Zheng-Jun Zha. Fourierrmamba: Fourier learning integration with state space models for image deraining. *arXiv preprint arXiv:2405.19450*, 2024.
- [38] Minghan Li, Xiangyong Cao, Qian Zhao, Lei Zhang, and Deyu Meng. Online rain/snow removal from surveillance videos. *IEEE Trans. Image Process.*, 30:2029–2044, 2021.
- [39] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1633–1642, 2019.
- [40] Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3175–3185, 2020.
- [41] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European conference on computer vision (ECCV)*, pages 254–269, 2018.
- [42] Jiaying Liu, Wenhan Yang, Shuai Yang, and Zongming Guo. Erase or fill? deep joint recurrent rain removal and reconstruction in videos. In *CVPR*, pages 3233–3242. Computer Vision Foundation / IEEE Computer Society, 2018.
- [43] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7007–7016, 2019.
- [44] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, and Yunfan Liu. Vmamba: Visual state space model, 2024. URL <https://arxiv.org/abs/2401.10166>.
- [45] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Trans. Image Process.*, 27(6):3064–3073, 2018.
- [46] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.
- [47] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- [48] Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024.
- [49] Tim Meinhardt, Alexander Kirillov, Laura Leal-Taixe, and Christoph Feichtenhofer. Trackformer: Multi-object tracking with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8844–8854, 2022.
- [50] Guy M Morton. A computer oriented geodetic data base and a new technique in file sequencing. *IBM, Ottawa, Canada*, 1966.
- [51] Valentina Musat, Ivan Fursa, Paul Newman, Fabio Cuzzolin, and Andrew Bradley. Multi-weather city: Adverse weather stacking for autonomous driving. In *ICCVW*, pages 2906–2915. IEEE, 2021.
- [52] Eric Nguyen, Karan Goel, Albert Gu, Gordon Downs, Preey Shah, Tri Dao, Stephen Baccus, and Christopher Ré. S4nd: Modeling images and videos as multidimensional signals with state spaces. *Advances in neural information processing systems*, 35:2846–2861, 2022.

- [53] Jack A Orenstein and Tim H Merrett. A class of data structures for associative searching. In *Proceedings of the 3rd ACM SIGACT-SIGMOD Symposium on Principles of Database Systems*, pages 181–190, 1984.
- [54] Zongzhi Ouyang and Wenhui Li. Mmamba: Enhancing image deraining with morton curve-driven locality learning. *Neurocomputing*, 638:130161, 2025.
- [55] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(8):10346–10357, 2023. doi: 10.1109/TPAMI.2023.3238179. URL <https://doi.org/10.1109/TPAMI.2023.3238179>.
- [56] A Paszke. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019.
- [57] Rui Qian, Robby T. Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *CVPR*, pages 2482–2491. Computer Vision Foundation / IEEE Computer Society, 2018.
- [58] Yuhui Quan, Shijie Deng, Yixin Chen, and Hui Ji. Deep learning for seeing through window with raindrops. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2463–2471, 2019.
- [59] Weihong Ren, Jiandong Tian, Zhi Han, Antoni B. Chan, and Yandong Tang. Video desnowing and deraining based on matrix decomposition. In *CVPR*, pages 2838–2847. IEEE Computer Society, 2017.
- [60] Stefan Roth and Michael J. Black. Fields of experts: A framework for learning image priors. In *CVPR (2)*, pages 860–867. IEEE Computer Society, 2005.
- [61] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.
- [62] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *CVPR*, pages 2805–2814. Computer Vision Foundation / IEEE, 2020.
- [63] Yuan Shi, Bin Xia, Xiaoyu Jin, Xing Wang, Tianyu Zhao, Xin Xia, Xuefeng Xiao, and Wenming Yang. Vmambair: Visual state space model for image restoration. *arXiv preprint arXiv:2403.11423*, 2024.
- [64] Yuheng Shi, Mingjing Dong, Mingjia Li, and Chang Xu. Vssd: Vision mamba with non-causal state space duality. *arXiv preprint arXiv:2407.18559*, 2024.
- [65] Shangquan Sun, Wenqi Ren, Xinwei Gao, Rui Wang, and Xiaochun Cao. Restoring images in adverse weather conditions via histogram transformer. In *European Conference on Computer Vision*, pages 111–129. Springer, 2024.
- [66] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5769–5780, 2022.
- [67] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2353–2363, 2022.
- [68] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.
- [69] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson WH Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12270–12279, 2019.
- [70] Zhengwei Wang, Qi She, and Aljosa Smolic. Action-net: Multipath excitation for action recognition. In *CVPR*, pages 13214–13223. Computer Vision Foundation / IEEE, 2021.

- [71] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *CVPR*, pages 10551–10560. Computer Vision Foundation / IEEE, 2021.
- [72] Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Chunyi Li, Liang Liao, Annan Wang, Erli Zhang, Wenxiu Sun, Qiong Yan, Xiongkuo Min, Guangtai Zhai, and Weisi Lin. Q-align: Teaching Imms for visual scoring via discrete text-defined levels. *International Conference on Machine Learning (ICML)*, 2024. Equal Contribution by Wu, Haoning and Zhang, Zicheng. Project Lead by Wu, Haoning. Corresponding Authors: Zhai, Guangtai and Lin, Weisi.
- [73] Jie Xiao, Xueyang Fu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Image de-raining transformer. *IEEE transactions on pattern analysis and machine intelligence*, 45(11):12978–12995, 2022.
- [74] Chenhongyi Yang, Zehui Chen, Miguel Espinosa, Linus Ericsson, Zhenyu Wang, Jiaming Liu, and Elliot J Crowley. Plainmamba: Improving non-hierarchical mamba in visual recognition. *arXiv preprint arXiv:2403.17695*, 2024.
- [75] Wenhan Yang, Robby T. Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(6):1377–1393, 2020.
- [76] Yang Yang, Chun-Le Guo, and Xiaojie Guo. Depth-aware unpaired video dehazing. *IEEE Trans. Image Process.*, 33:2388–2403, 2024.
- [77] Yang Yang, Chaoyue Wang, Xiaojie Guo, and Dacheng Tao. Robust unpaired image dehazing via density and depth decomposition. *Int. J. Comput. Vis.*, 132(5):1557–1577, 2024.
- [78] Rajeev Yasarla and Vishal M. Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning CNN for single image de-raining. In *CVPR*, pages 8405–8414. Computer Vision Foundation / IEEE, 2019.
- [79] Tian Ye, Sixiang Chen, Jinbin Bai, Jun Shi, Chenghao Xue, Jingxia Jiang, Junjie Yin, Erkang Chen, and Yun Liu. Adverse weather removal with codebook priors. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12653–12664, 2023.
- [80] Shaodi You, Robby T Tan, Rei Kawakami, Yasuhiro Mukaigawa, and Katsushi Ikeuchi. Adherent raindrop modeling, detection and removal in video. *IEEE transactions on pattern analysis and machine intelligence*, 38(9):1721–1733, 2015.
- [81] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14821–14831, 2021.
- [82] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022.
- [83] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 695–704, 2018.
- [84] Howard Zhang, Yunhao Ba, Ethan Yang, Varan Mehra, Blake Gella, Akira Suzuki, Arnold Pfahnl, Chethan Chinder Chandrappa, Alex Wong, and Achuta Kadambi. Weatherstream: Light transport automation of single image deweathering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13499–13509, 2023.
- [85] Jingang Zhang, Wenqi Ren, Shengdong Zhang, He Zhang, Yunfeng Nie, Zhe Xue, and Xiaochun Cao. Hierarchical density-aware dehazing network. *IEEE Trans. Cybern.*, 52(10):11187–11199, 2022.

- [86] Kaihao Zhang, Dongxu Li, Wenhan Luo, and Wenqi Ren. Dual attention-in-attention model for joint rain streak and raindrop removal. *IEEE Transactions on Image Processing*, 30:7608–7619, 2021.
- [87] Kaihao Zhang, Rongqing Li, Yanjiang Yu, Wenhan Luo, and Changsheng Li. Deep dense multi-scale network for snow removal using semantic and depth priors. *IEEE Trans. Image Process.*, 30:7419–7431, 2021.
- [88] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.
- [89] Xuanqi Zhang, Haijin Zeng, Jinwang Pan, Qiangqiang Shen, and Yongyong Chen. Llemamba: Low-light enhancement via relighting-guided mamba with deep unfolding network. *arXiv preprint arXiv:2406.01028*, 2024.
- [90] Haiyu Zhao, Yuanbiao Gou, Boyun Li, Dezhong Peng, Jiancheng Lv, and Xi Peng. Comprehensive and delicate: An efficient transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14122–14132, 2023.
- [91] Zou Zhen, Yu Hu, and Zhao Feng. Freqmamba: Viewing mamba from a frequency perspective for image deraining. *arXiv preprint arXiv:2404.09476*, 2024.
- [92] Shihao Zhou, Duosheng Chen, Jinshan Pan, Jinglei Shi, and Jufeng Yang. Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2952–2963, 2024.
- [93] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.
- [94] Lianghai Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*, 2024.
- [95] Yurui Zhu, Tianyu Wang, Xueyang Fu, Xuanyu Yang, Xin Guo, Jifeng Dai, Yu Qiao, and Xiaowei Hu. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *CVPR*, pages 21747–21758. IEEE, 2023.
- [96] Zhen Zou, Hu Yu, Jie Huang, and Feng Zhao. Freqmamba: Viewing mamba from a frequency perspective for image deraining. In *Proceedings of the 32nd ACM international conference on multimedia*, pages 1905–1914, 2024.

A Appendix

A.1 Implement Details

MODEM is implemented in PyTorch [56] and trained on 4 NVIDIA RTX 3090 GPUs. The main backbone contains residual groups with [4, 4, 6, 8, 6, 4, 4] MDSLs sequentially, and a final refinement group with 4 MDSLs. Downsampling and upsampling are performed using PixelUnshuffle and PixelShuffle, respectively. The initial channel size is 36. The DDEM employs residual groups with 2 MDSLs each, operating with 96 channels. We train MODEM in two stages. In Stage 1, DDEM takes the channel-concatenated degraded image I_{LQ} and ground-truth I_{GT} as input. We use the AdamW optimizer [47] with a learning rate of 3×10^{-4} , a weight decay of 1×10^{-4} , and betas set to (0.9, 0.999). The learning rate is scheduled by a Cosine Annealing Restart Cyclic LR [46] scheduler with periods of [92k, 208k] iterations and restart weights of [1, 1]. The minimum learning rates, $\eta_{\min}$, for these cycles are $[3 \times 10^{-4}, 1 \times 10^{-6}]$. Progressive training is adopted with patch sizes [128, 160, 256, 320, 384] and corresponding per-GPU mini-batch sizes [6, 4, 2, 1, 1] for [92k, 84k, 56k, 36k, 32k] iterations each, totaling 300k iterations. In Stage 2, DDEM processes only I_{LQ} , initializing parameters from Stage 1. The same AdamW optimizer settings and Cosine Annealing Restart Cyclic LR scheduler parameters are employed. Progressive training uses patch sizes [128, 160, 256, 320, 376] with per-GPU mini-batch sizes [8, 5, 2, 1, 1].

A.2 Datasets and Metrics

To ensure a fair and comprehensive evaluation, MODEM is trained and tested consistent with those utilized in prior works [40, 67, 55, 65, 95, 79]. Our composite training data is aggregated from multiple sources, including 9,000 synthetic images featuring snow degradation from Snow100K [45], 1,069 real-world images affected by adherent raindrops from the Raindrop [57], and an additional 9,000 synthetic images from Outdoor-Rain [39] which are degraded by a combination of both fog and rain streaks. For performance evaluation, we utilize several distinct test sets: the Test1 [39, 40], the designated test split from the RainDrop [57], both the Snow100K-L and Snow100K-S subsets [45], and a challenging real-world test set from Snow100K comprising 1,329 images captured under various adverse weather conditions. We report PSNR and SSIM on these test datasets.

A.3 Complexity Analysis

We report the parameters and inference time. The inference time is performed on a single Nvidia RTX 3090, with single 256×256 input image, detailed in Table 8.

Table 8: Comparison of parameters and inference time, along with average PSNR.

Methods	WGWSNet [95]	WeatherDiff [55]	Histoformer [65]	MODEM (Ours)
Time (ms)	24.83	1.67×10^6	109.07	92.86
Parameters (M)	2.65	82.96	16.62	19.96
Average PSNR	31.54	31.57	33.68	34.18

Table 9: Comparison of inference time (ms) for different input sizes and average PSNR.

Input Size	WGWSNet [95]	WeatherDiff [55]	Histoformer [65]	MODEM (Ours)
256×256	24.83	1.67×10^6	109.07	92.86
512×512	110.34	5.37×10^6	576.15	443.02
1024×1024	439.13	1.35×10^7	3056.29	1946.34
Average PSNR	31.54	31.57	33.68	34.18

To be more comprehensive, we evaluate the inference speed on larger resolutions in Table 9. As can be seen, compared to Transformer-based architectures like Histoformer [65], as the number of pixels increases, MODEM’s complexity scales in a near-linear fashion, demonstrating significantly better scalability. This stands in contrast to the quadratic complexity often associated with Transformers, granting MODEM a distinct computational advantage.

Table 10: Comparison of parameters and inference time for other Mamba-style methods.

Method	MambaIR [20]	FreqMamba [91]	MODEM (Ours)
Parameters (M)	15.78	8.93	19.96
Time (ms)	790.61	233.41	92.86

Table 11: Comparison of inference time (ms) between Hilbert scan and Morton scan.

Input Size	256×256	512×512	1024×1024
Hilbert scan [28]	604.00	10134.91	88288.46
Morton scan (Ours)	92.86	443.02	1946.34

To further contextualize our model’s performance, we compared MODEM with other Mamba-style methods, MambaIR [20] and FreqMamba [91]. As shown in Table 10, although MODEM has a larger parameter count compared to both MambaIR [20] and FreqMamba [91], it achieves a significantly faster inference time, highlighting the superior efficiency of our architectural design.

For scanning methods, Hilbert scan [28] is another locality-preserving alternative. However, it comes with a significantly higher computational cost, whereas the Morton-order can be calculated efficiently with simple bitwise operations as shown in our Eq. (6). To be clear, we compared the inference speed of the Morton and Hilbert scans. For the Hilbert scan, we used the official implementation from LC-Mamba [28]. The speed (ms) of different resolutions are shown in the Table 11, with all tests performed on a single RTX 3090 GPU. Our method is significantly faster than Hilbert.

A.4 More Visualizations of Features

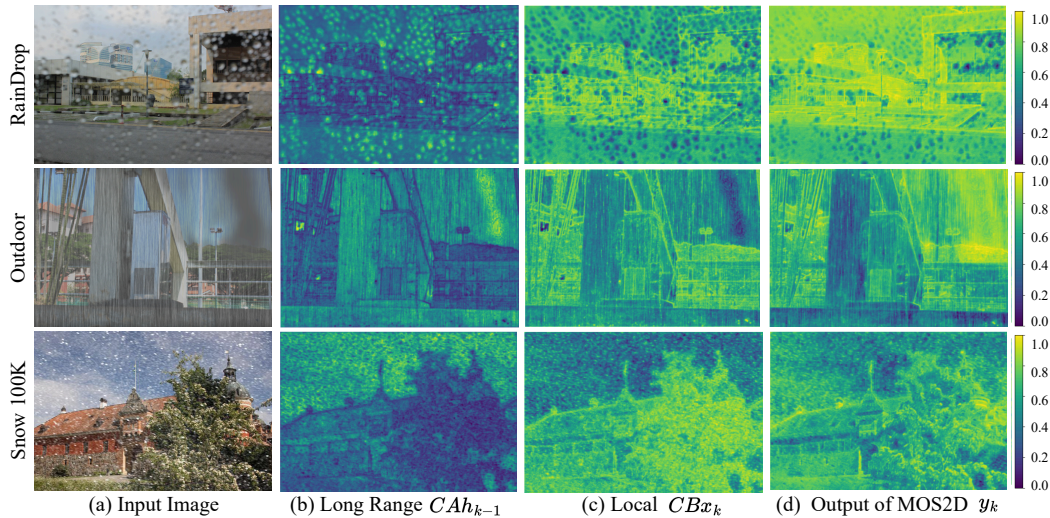


Figure 9: With respect to a sample (a), (b)-(d) visualize the long-range CAh_{k-1} , (c) local CBx_k , and (d) output of MOS2D y_k , respectively.

A.5 Limitations

While our proposed MODEM demonstrates state-of-the-art performance across a variety of adverse weather conditions, we acknowledge certain limitations. As illustrated in Fig. 10 of challenging real-world snow scenarios, MODEM, like other contemporary methods, can encounter difficulties in achieving perfect restoration when faced with images containing extremely large, high-contrast snowflakes. Such scenarios are particularly challenging if these specific visual patterns of snow, differing significantly in scale, density, or opacity from typical training examples, are underrepresented or entirely absent in the training data distribution. Nevertheless, empowered by its robust degradation estimation capabilities, MODEM still achieves comparatively better results in these extreme cases, effectively suppressing artifacts and preserving some structural detail. This highlights an ongoing

challenge in achieving perfect generalization to all unseen severe degradations, but also underscores the significant benefit of incorporating explicit and adaptive degradation estimation.

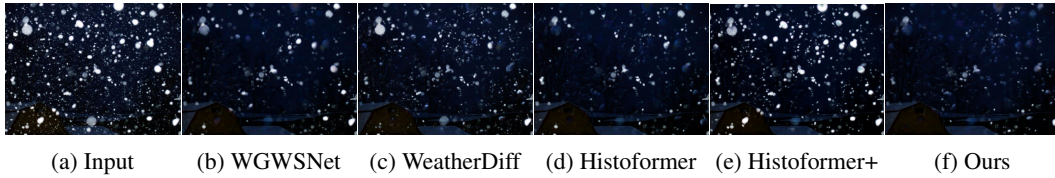


Figure 10: Visual comparison on a challenging real-world snow case from the dataset in [45], illustrating performance under severe degradation unseen during training. We compare MODEM to the prior methods [95, 55, 65], where “+” denotes additional training of Histoformer.

A.6 More Ablation Studies

The Morton-order scans an image by processing it in small, contiguous block-like regions, moving from one adjacent block to the next (refer to the scale-space theory). This ensures that pixels that are close in the 2D image stay close in the 1D sequence. To further investigate the impact of the scanning scheme itself, we offer an ablation study within our MODEM, comparing the Morton scan against others [20, 74, 63, 25, 28, 24]. As shown in Table 12, our Morton achieves the best performance.

Table 12: Ablation study of different scanning scheme.

Methods	Snow100K-L		Snow100K-S		Outdoor		RainDrop	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Raster (MambaIR [20])	32.33	0.9283	38.03	<u>0.9672</u>	32.89	0.9399	32.69	0.9423
Continuous (Zigzag [74])	32.17	0.9266	37.74	0.9662	32.35	0.9385	32.59	0.9422
OSS (VmambaIR [63])	32.14	0.9262	37.39	0.9651	32.61	0.9387	32.11	0.9413
Local (LocalMamba [25])	32.23	0.9266	37.75	0.9661	32.60	0.9382	32.53	0.9418
Hilbert (LC-Mamba [28])	<u>32.46</u>	<u>0.9287</u>	<u>37.96</u>	0.9671	<u>32.99</u>	0.9414	<u>32.82</u>	<u>0.9433</u>
Morton (Ours)	32.52	0.9292	38.08	0.9673	33.10	<u>0.9410</u>	33.01	0.9434

We conduct an ablation study on the contribution of the loss terms. We have also performed an ablation study on the placement (before/after the average pooling) of the KL divergence loss. The results are presented in the table below. They confirm that each component contributes positively.

Table 13: Ablation study of the loss function.

Methods	Snow100K-L		Snow100K-S		Outdoor		RainDrop	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
w/o Correlation Loss	32.30	0.9278	37.79	0.9668	32.91	0.9403	32.69	0.9427
w/o KL Loss	32.12	0.9249	37.61	0.9650	32.37	0.9224	32.38	0.9390
Replace KL Loss	31.85	0.9237	37.16	0.9638	32.77	0.9389	32.57	0.9387
MODEM	32.52	0.9292	38.08	0.9673	33.10	0.9410	33.01	0.9434

A.7 More Quantitative Comparisons

We retrain three prominent Mamba-style restoration networks [20, 63, 96] on the all-weather setting. These experiments were conducted carefully following the settings used by other recent methods like ConvIR [13], FSNet [92], and Histoformer [65], while also respecting their original training configurations. The results are presented in Table 14. Our MODEM achieves the best performance.

Table 14: Comparison with Mamba-like Methods.

Methods	Snow100K-L		Snow100K-S		Outdoor		RainDrop	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
MambaIR [20]	28.59	0.8729	33.34	0.9330	<u>24.73</u>	0.8808	32.16	<u>0.9393</u>
FreqMamba [96]	27.09	0.8624	33.52	0.9331	19.89	0.7519	30.60	0.9198
VmambaIR [63]	<u>31.07</u>	<u>0.9168</u>	<u>36.35</u>	<u>0.9605</u>	24.23	<u>0.8558</u>	<u>32.18</u>	0.9392
MODEM	32.52	0.9292	38.08	0.9673	33.10	0.9410	33.01	0.9434

A.8 More Visual Comparisons

Further visual comparisons are presented in Figs. 11 to 14.

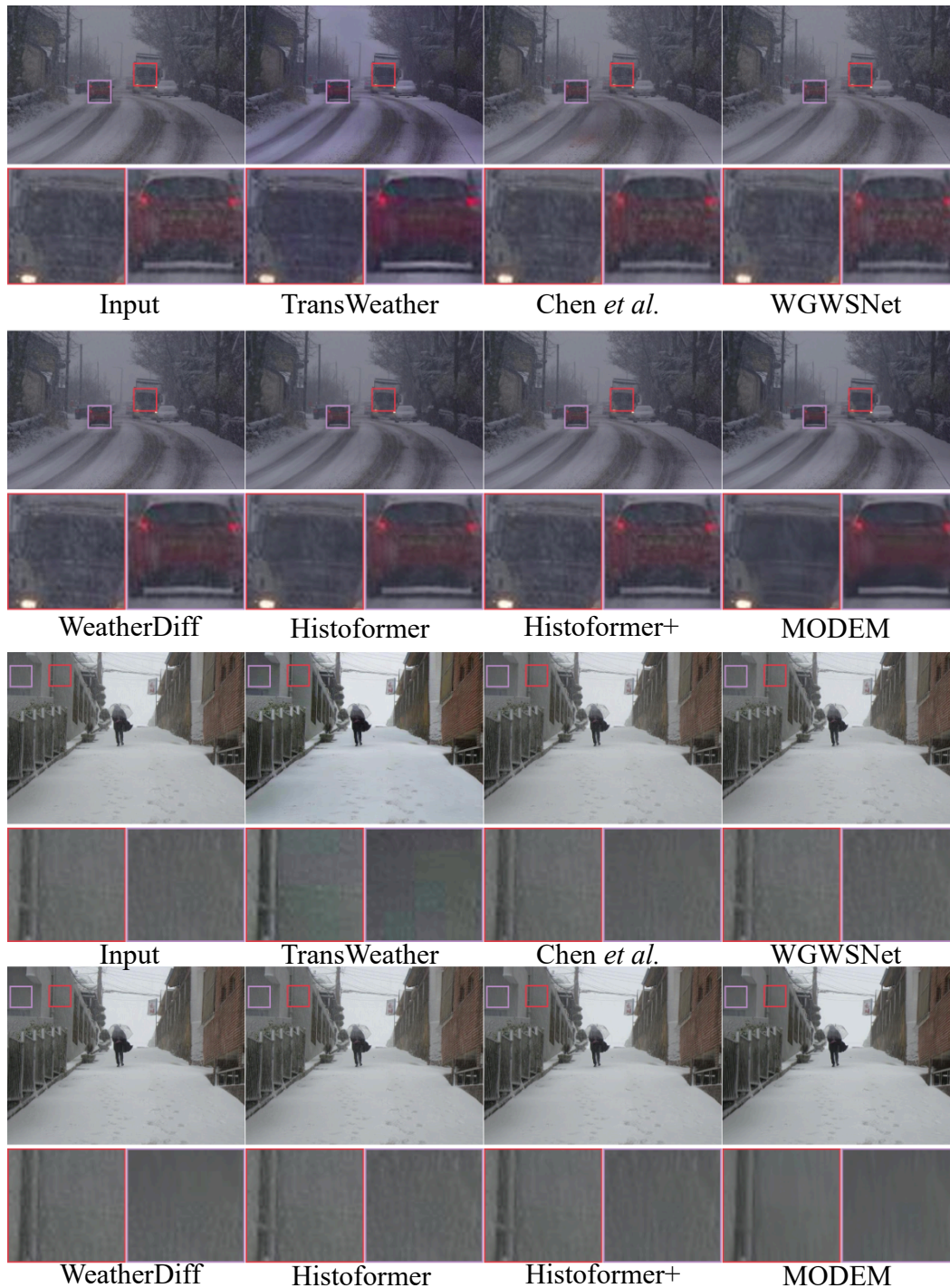


Figure 11: More visual results for desnowing on real-world snowy images [45].

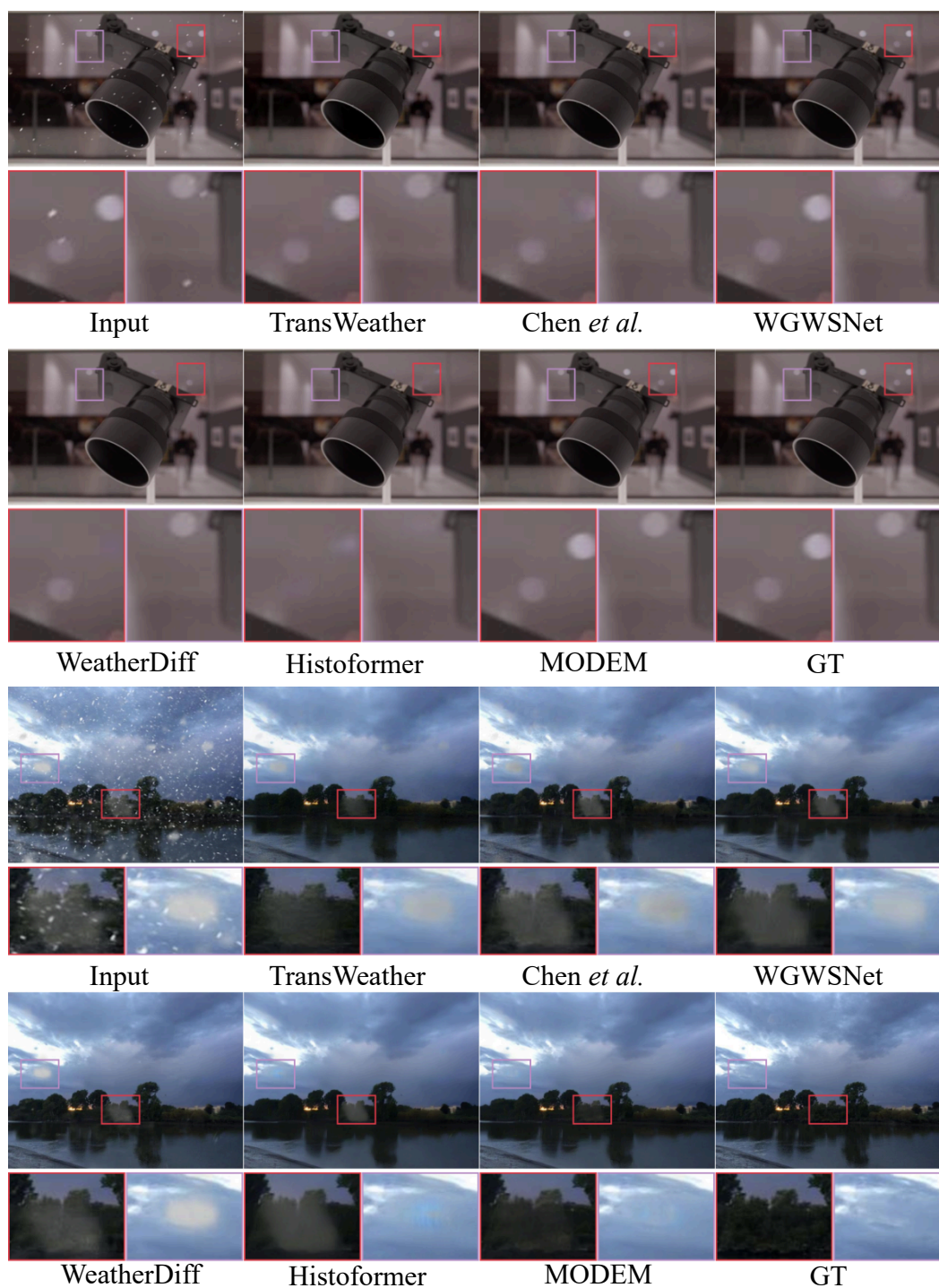


Figure 12: More visual results for image desnowing on the Snow100K [45].

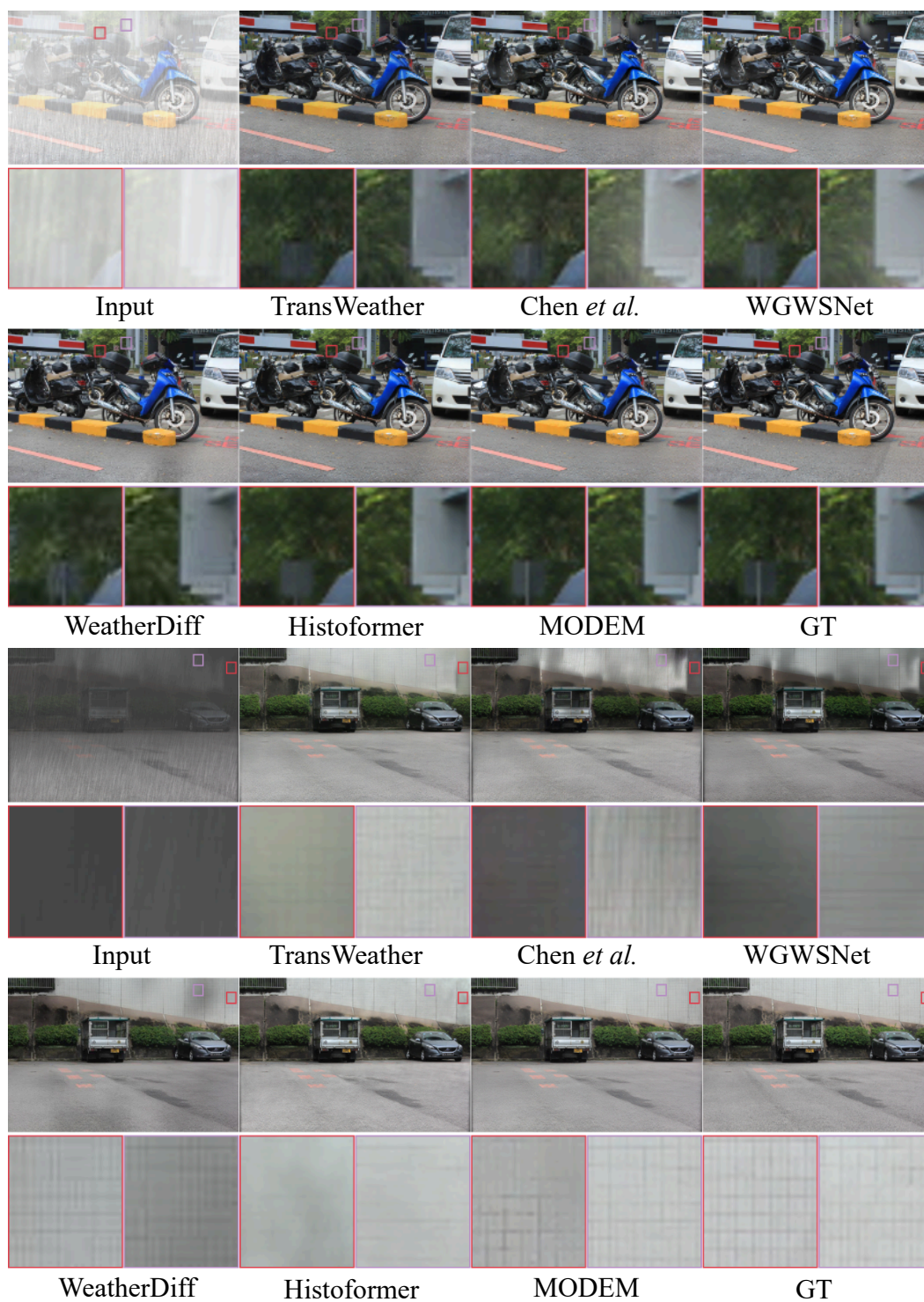


Figure 13: More visual results for deraining&dehazing on the Test1 dataset [39, 40].



Figure 14: More visual results for raindrop removal on the Raindrop dataset [57].

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: In the abstract, we summarize our main claims and the scope of our research. In the introduction, we clearly state our contributions.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: We have discussed our limitations in Section A.5.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: The paper does not include theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper provides detailed descriptions of the proposed MODEM architecture in Section 4, training settings, and datasets used for training and evaluation, along with the evaluation metrics in Sections A.1, A.2 and 5.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The code will be made open-source.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The paper specifies key training and test details, including the datasets utilized, data splits for training/validation/testing where applicable, and the primary hyperparameters for training the MODEM framework (e.g., learning rate, batch size, optimizer type) in Sections A.2, A.5 and 5.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: Consistent with common practice in this specific field of image restoration, error bars or detailed statistical significance tests are not explicitly reported for the main quantitative results. We follow established evaluation protocols and reporting standards used by prior state-of-the-art methods to ensure fair and direct comparisons on benchmark datasets.

Guidelines:

- The answer NA means that the paper does not include experiments.

- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Section 5 of the paper provides detailed information regarding the computer resources.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: The research conducted and presented in this paper conforms, in every respect, with the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: We discuss the potential positive societal impacts of our work in both the Introduction and Conclusion sections. The nature of this research, focused on improving image quality under adverse weather, does not inherently present negative societal impacts.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: This paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The assets used in this work (e.g., code, data, models) are properly credited, and the corresponding licenses and terms of use are respected.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.

- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: This paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: This paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: This paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this paper does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.