Pixel Ignores, Superpixel Sees: Adverse Weather Image Restoration via Semantic-Center SSM

Dayu Li[3], Shihao Zhou[1,3], Leizhi Shu[4], Jin Wu[4], Chi Man Vong[5], and Jufeng Yang[1,2,3]*

[1] Nankai International Advanced Research Institute (SHENZHEN·FUTIAN)
[2] Peng Cheng Laboratory
[3] College of Computer Science, Nankai University
[4] University of Science and Technology Beijing
[5] Department of Computer and Information Science, University of Macau

Abstract. Adverse weather image restoration aims to recover clear visibility from degraded images in complex weather conditions. Existing works attempt to address this problem by modeling relationships between pixels, however, this paradigm defies the spatially non-uniformity fact of degradations and learns non-discriminative features from semantic-conflict regions. In this paper, we propose SSR, a **S**emantic-center guilded **S**tate space model for image **R**estoration. The key idea of SSR is to shift the conventional scanning strategy of pixel-serial to semantic-guilded one. Specifically, we introduce a Superpixel-guided Selective Scan Mechanism (S^3M), which first partitions the image into perceptually coherent regions via superpixel clustering and then performs relations modeling within the semantic-related regions. Moreover, a Region-level Gating Mechanism (RGM) is developed to perform intra-region calibration by modulating degradation outliers within each semantic superpixel unit along the channel dimension. Extensive experiments on **6** well-established benchmarks demonstrate that SSR performs favorably against state-of-the-art models with competitive computational cost. The source code is publicly available at `https://github.com/LIDAYU-DayuLi/SSR`.

Keywords: Image restoration · Adverse weather · Superpixel clustering

1 Introduction

Adverse weather conditions, such as rain, haze, and snow, pose challenges to vision systems deployed in real-world environments. Serious degradations distort scene structures and corrupt visual cues in a spatially non-uniform way, leading to performance degradation in downstream applications [8,35,58]. Intuitively, developing a unified image restoration model for handling different adverse weather conditions provides a plausible solution.

* Corresponding author.

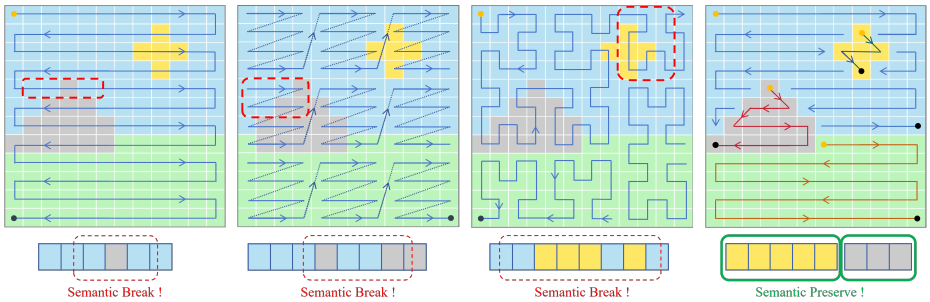
Previous restoration methods [12, 34] primarily rely on hand-crafted priors to approximate the physical process of degradation. Unfortunately, these simplified assumptions often fall short in capturing the diverse and non-linear patterns observed in real-world weather. To address this, deep learning-based approaches [7, 24] have shifted the paradigm toward data-driven one, achieving superior performance by learning complex mappings from large-scale paired datasets. Nevertheless, most existing models are constrained to task-specific settings (e.g., only deraining or dehazing), which leads to limited generalization capability when encountering different or hybrid degradations under in-the-wild environments.

To address the diverse and complex degradations in adverse weather, All-in-one Image Restoration (AIR) frameworks aim to restore images using a single unified model. Early CNN-based methods [17,20] are restricted within limited receptive fields, while Transformer-based architectures [38,65] achieve global modeling but incur quadratic computational overhead. More recently, State Space Models (SSMs) [9, 25] have emerged as a promising alternative, offering linear-complexity long-range dependency modeling. However, directly applying SSMs to the AIR task presents a critical challenge. For the vision tasks, the SSMs rely on flattening 2D images into 1D sequences for recursive state propagation. The vanilla SSMs adopt predefined and geometry-based scanning trajectories, such as standard raster scans [25] (Fig. 1a), window partitioning [13] (Fig. 1b), or hilbert curves [45] (Fig. 1c). These attempts achieve advanced performance boosts with limited computational cost, but they are hindered from generating better results by ignoring image semantic content. SSMs rely on sequential state accumulation to model representation, while these predefined trajectories inevitably include semantically conflicting regions as continuous sequences (e.g., patches cross object boundaries). These fields are not ideally suited for the SSMs to learn informative features, and heavy mixed degradations further impede the models from achieving satisfactory restoration.

In light of this, we propose the **S**emantic-center guided **S**tate space model for **R**estoration (SSR) to advance the content-agnostic scanning manner in current methods. Instead of relying on rigid geometric paths, SSR utilizes superpixels (Fig. 1d) as perceptually coherent units to dynamically construct structure-aware scanning trajectories. This adaptive routing strictly adheres to semantic boundaries, effectively preventing the leakage of degradation noise into clean regions. Furthermore, we introduce a Region-level Gating Mechanism (RGM) to perform intra-region calibration. By adaptively modulating degradation outliers within each semantic superpixel unit along the channel dimension, our framework provides a more interpretable, efficient, and content-adaptive SSM-based model tailored for universal weather restoration.

Our main contributions are summarized as follows.

- We propose SSR, a **S**emantic-center guided **S**tate space model for image **R**estoration, which shifts the conventional scanning paradigm of SSMs from pixel serial to superpixel-based semantic-guided one, effectively preventing learning noisy features from the semantic conflict regions.



(a) Raster [25] (b) Local Window [13] (c) Hilbert [45] (d) S^3M(ours)

Fig. 1: Comparison of scanning mechanisms. (a-c) Existing predefined scan strategies fall into a content-agnostic paradigm, and may learn non-discriminative features from semantic-conflict regions. (d) Our Superpixel-guided Selective Scan Mechanism (S^3M) Scan constrains the trajectory within perceptually coherent regions, effectively preventing information leakage and ensuring structure-aware state evolution.

- We introduce the Superpixel-guided Selective Scan Mechanism (S^3M) to model relations within perceptually coherent regions, alongside a Region-level Gating Mechanism (RGM) that performs intra-region calibration by modulating degradation outliers along the channel dimension.
- Extensive experiments on **6** well-established benchmarks demonstrate that SSR performs favorably against state-of-the-art models with competitive computational cost.

2 Related Work

Task-specific image restoration. Early work in this field primarily relied on image decomposition or statistical priors (e.g., sparse representation for rain streak removal [14, 15] and the Dark Channel Prior for haze removal [12]). With the widespread adoption of deep learning, the focus shifted towards data-driven modeling, leading to the development of various specialized network architectures. such as polarization-aided transformers [6] and detail separation networks [7], recurrent aggregation networks [22], and uncertainty-guided multi-scale designs [49] for rain streak removal; attention-guided generative adversarial networks [32] and dual attention models [56] for raindrop removal; transmission map estimation models [48] along with contrastive learning frameworks [44] for dehazing; and periodicity-aware frameworks for burst flicker removal [33]. On the other hand, researchers have also developed effective desnowing models by incorporating context-aware designs [24] or leveraging semantic and depth priors [57]. More recent solutions are often built upon general CNN backbones [4,5] or Transformers [61–63]. However, this specialization in a single degradation type impedes generalization across diverse weather conditions, thereby complicating practical deployment with multiple dedicated models and increased overhead.

All-in-one weather removal. In recent years, unified all-weather image restoration models have emerged as a significant research direction, aiming to handle multiple weather degradations with a single model. Early efforts primarily focused on exploring generic network architectures [20] or leveraging the global modeling capability of Transformers to construct foundational models [38]. As research has progressed, more diverse approaches have been introduced to enhance model generalization and adaptability, including disentangling weather-general and weather-specific features [66], adopting generative frameworks such as diffusion models [28], integrating learned codebook priors [50], and incorporating global image statistics as priors within Transformer-based models like Histoformer [30]. Moreover, some works have explored strategies such as instruction-based natural language guidance [3] and prompt-encoded degradation conditions [31]. At the same time, efficient long-sequence modeling architectures, represented by State Space Models (SSMs), have been introduced to this task, offering a new approach for modeling long-range degradation dependencies through their linear complexity and selective mechanisms [39]. Nevertheless, accurately modeling diverse and spatially heterogeneous weather degradations remains a core challenge in the field.

Mamba-based image restoration. Recent advances in SSMs have introduced an efficient approach for modeling long-range dependencies in image restoration, leveraging linear complexity and selective scanning mechanisms [9]. General frameworks such as MambaIR [11] and VMambaIR [36] have validated the potential of this architecture as a foundational model for image restoration. Concurrently, research on task-specific designs has proliferated, spanning areas including deraining [18, 47, 68], dehazing [59, 60], deblurring [16], and low-light enhancement [1, 42]. For instance, RainMamba [45] incorporates a Hilbert-order scanning mechanism to enhance the preservation of local structures, while Wave-Mamba [67] integrates wavelet transforms to optimize feature representation in the frequency domain. Furthermore, studies such as Vision Mamba [64] and Mamba-ND [21] have extended SSMs into pure visual architectures and multi-dimensional data processing frameworks,improving model capabilities in dense prediction and multi-dimensional signal modeling. These developments collectively underscore the ongoing evolution of scanning strategies within SSM frameworks, highlighting their critical role in capturing spatially varying degradations and establishing a foundation for further advances in sequential modeling for image restoration.

3 Methodology

3.1 Overall Architecture

The overall architecture of SSR follows a hierarchical encoder-decoder framework (Fig. 2). Given a degraded image $I_{in} \in \mathbb{R}^{H\times W\times 3}$, a 3×3 convolutional layer is first employed to extract shallow features $F_s \in \mathbb{R}^{H\times W\times C}$, where C denotes the number of feature channels. These features are then fed into a 3-level symmetric encoder-decoder network. Each level consists of several S^3M Blocks

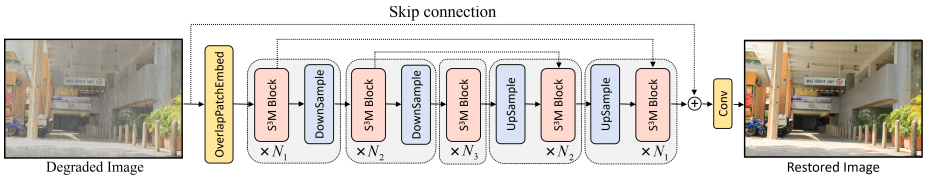


Fig. 2: Overall architecture of the proposed Semantic-center Guided State Space Model **(SSR)**. Our model adopts a hierarchical U-shaped encoder-decoder structure to capture multi-scale features for unified weather restoration. The backbone consists of multiple S^{3}M Blocks, which integrate the Superpixel-guided Selective Scan Mechanism (S^{3}M) and a Region-level Gating Mechanism (**RGM**) to perform intra-region calibration by modulating degradation outliers along the channel dimension. Skip connections are employed to fuse multi-scale encoder features with upsampled decoder features.

(see Sec. 3.2). For the level-l encoder/decoder ($l \in \{1, 2, 3\}$), the intermediate features $F^l_{enc}/F^l_{dec} \in \mathbb{R}^{\frac{H}{2^{l-1}} \times \frac{W}{2^{l-1}} \times 2^{l-1}C}$ are progressively processed. We utilize bilinear interpolation and 1×1 convolutions for downsampling and upsampling, complemented by skip connections between corresponding levels to fuse multi-scale features. Finally, a 3×3 convolution is applied to generate the restoration residual $R \in \mathbb{R}^{H \times W \times 3}$. The final restored image I_{rec} is obtained by:

$$I_{rec} = I_{in} + R = I_{in} + \mathcal{F}_{SSR}(I_{in}), \tag{1}$$

where $\mathcal{F}_{SSR}(\cdot)$ denotes the proposed SSR network.

3.2 S^{3}M Block

The core building block of our encoder and decoder is the S^{3}M Block (Fig. 3). Given an input feature $X \in \mathbb{R}^{H \times W \times C}$, the block consists of a spatial token mixer and a channel-wise feed-forward network:

$$X' = \mathrm{S^3M}(\mathrm{LN}(X)) + X, \tag{2}$$

$$X_{out} = \mathrm{FFN}(\mathrm{LN}(X')) + X', \tag{3}$$

where $\mathrm{LN}(\cdot)$ denotes Layer Normalization, and the proposed Superpixel-guided Selective Scan Mechanism (S^{3}M) module serves as the primary spatial mixer.

The S^{3}M module integrates two primary components: *Superpixel-guided Scanning* and the *Region-level Gating Module*. Specifically, the scanning mechanism partitions the input into perceptually coherent regions via superpixel clustering and performs relations modeling within these semantic-related regions. This process constrains state propagation to prevent learning non-discriminative features from semantic-conflict areas. Simultaneously, the Region-level Gating Module is developed to perform intra-region calibration by modulating degradation outliers within each semantic superpixel unit along the channel dimension. The channel-wise details are subsequently processed by the Feed-Forward Network (FFN) to produce the final output X_{out}.

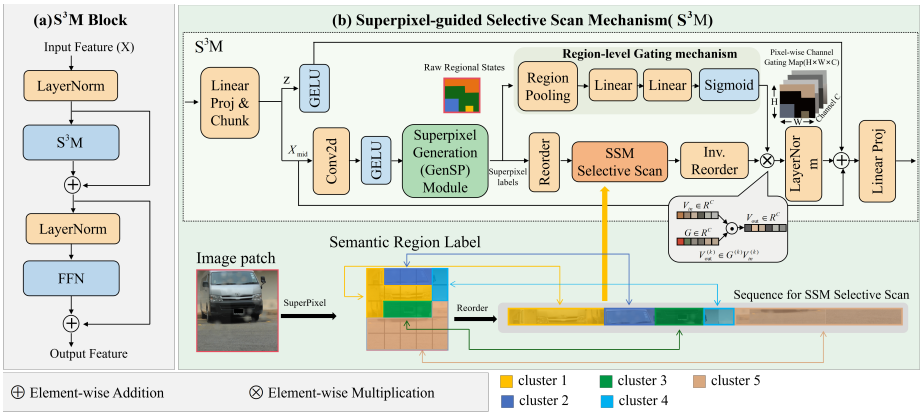


Fig. 3: Structure of the S^{3}M Block. (a) The S^{3}M block enables structure-aware feature transitions by integrating the Superpixel-guided Selective Scan Mechanism (S^{3}M) and a Feed-Forward Network (FFN). (b) The S^{3}M incorporates two synergistic innovations: Superpixel-guided Scanning, which constrains state propagation within perceptually coherent regions to prevent learning non-discriminative features from semantic-conflict areas, and Region-level Gating, which performs intra-region calibration by modulating degradation outliers within each semantic unit along the channel dimension.

Superpixel-guided Selective Scan Mechanism The Superpixel-guided Scanning mechanism partitions the spatial domain of the intermediate feature $X_{mid} \in \mathbb{R}^{H\times W\times D}$ into R perceptually coherent regions $\{\mathcal{S}_r\}_{r=1}^{R}$ using a superpixel generator [55], where each pixel (i,j) is assigned a region label $\ell_{i,j} \in \{1,\dots,R\}$. This partitioning ensures that the subsequent state propagation is confined within semantic-related regions to prevent learning non-discriminative features from semantic-conflict areas.

To align the state evolution with the semantic regions, we define a permutation operator $\mathcal{P}_\ell$ that reorders the flattened spatial tokens such that those belonging to the same superpixel are grouped contiguously. The superpixel-guided sequence modeling is formulated as:

$$F_{seq}^{sp} = \mathcal{P}_\ell(\text{reshape}(X_{mid})), \quad Z_{seq}^{sp} = \text{SSM}(F_{seq}^{sp}), \tag{4}$$

where SSM$(\cdot)$ denotes the selective scan operator. The output sequence is kemudian restored to the original spatial arrangement via an inverse permutation $Y = \text{reshape}(\mathcal{P}_\ell^{-1}(Z_{seq}^{sp}))$. This mechanism ensures that the state propagation is confined within perceptually coherent regions, thereby preventing the aggregation of non-discriminative features from semantic-conflict areas.

Region-level Gating Mechanism The Region-level Gating Mechanism (RGM) is developed to perform intra-region calibration by modulating degradation outliers along the channel dimension (Fig. 4). For each perceptually coherent region

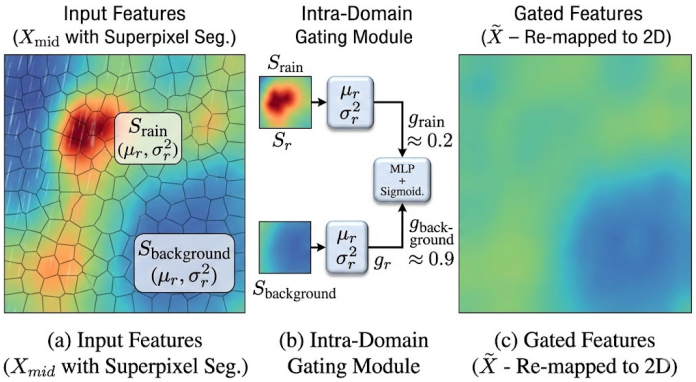


Fig. 4: Illustration of the Region-level Gating Mechanism (RGM). (a) Input features X_{mid} with superpixel labels partitioning the spatial domain into perceptually coherent regions. (b) The RGM computes regional statistics (μ_r, σ_r^2) and employs a lightweight MLP to generate adaptive gating factors g_r for channel-wise modulation. (c) Visualized gated features $\tilde{X}$. The mechanism adaptively modulates degradation outliers within different semantic units to perform intra-region calibration.

$\mathcal{S}_r$, we first compute the regional mean $\mu_r \in \mathbb{R}^C$ and variance $\sigma_r^2 \in \mathbb{R}^C$ from the intermediate feature X_{mid} as follows:

$$\mu_r = \frac{1}{|\mathcal{S}_r|} \sum_{(i,j)\in\mathcal{S}_r} X_{mid}^{(i,j)}, \quad \sigma_r^2 = \frac{1}{|\mathcal{S}_r|} \sum_{(i,j)\in\mathcal{S}_r} (X_{mid}^{(i,j)} - \mu_r)^2. \tag{5}$$

To perceive the degradation characteristics within each semantic unit, these statistics are concatenated as $[\mu_r; \sigma_r^2] \in \mathbb{R}^{2C}$ and processed by a lightweight MLP. The MLP, comprising two fully-connected layers with a non-linear activation, generates a region-specific gating factor $g_r \in \mathbb{R}^C$. This factor adaptively modulates the channel-wise feature responses to perform intra-region calibration. For the k-th token $x_k \in F_{seq}^{sp}$ belonging to region $\mathcal{S}_r$, the gated update and subsequent state evolution are formulated as:

$$\tilde{x}_k = \sigma(g_r) \odot x_k, \tag{6}$$

$$h_k = \bar{\mathbf{A}} h_{k-1} + \bar{\mathbf{B}} \tilde{x}_k, \tag{7}$$

where $\sigma(\cdot)$ denotes the Sigmoid function, $\odot$ represents the element-wise product, and h_k is the hidden state. By modulating degradation outliers within each semantic superpixel unit, the RGM suppresses non-discriminative features from semantic-conflict regions and provides a content-adaptive state transition.

3.3 Loss Function

To train the proposed SSR, we optimize the network using a combination of a pixel-wise reconstruction loss and a sequence-level structural correlation loss.

Given the restored image $\hat{I}$ and the corresponding ground truth I_{gt}, we first employ the L_1 loss to constrain the fundamental pixel-level fidelity:

$$\mathcal{L}_{pixel} = \|\hat{I} - I_{gt}\|_1. \tag{8}$$

While the L_1 loss ensures global color and luminance accuracy, relying solely on it can sometimes lead to overly smoothed results or geometric distortions in heavily degraded regions. To further enhance the structural coherence and linear correlation of the restored features, we introduce a Pearson correlation loss $\mathcal{L}_{seq}$. Let x and y denote the flattened 1D sequences of $\hat{I}$ and I_{gt}, respectively. The sequence-level Pearson loss is defined as:

$$\mathcal{L}_{seq} = 1 - \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}}, \tag{9}$$

where $\bar{x}$ and $\bar{y}$ are the mean values of the respective sequences. This term explicitly penalizes structural discrepancies and encourages the network to recover fine-grained details that are strictly correlated with the underlying clear image structure.

The overall objective function is the unweighted sum of these two terms:

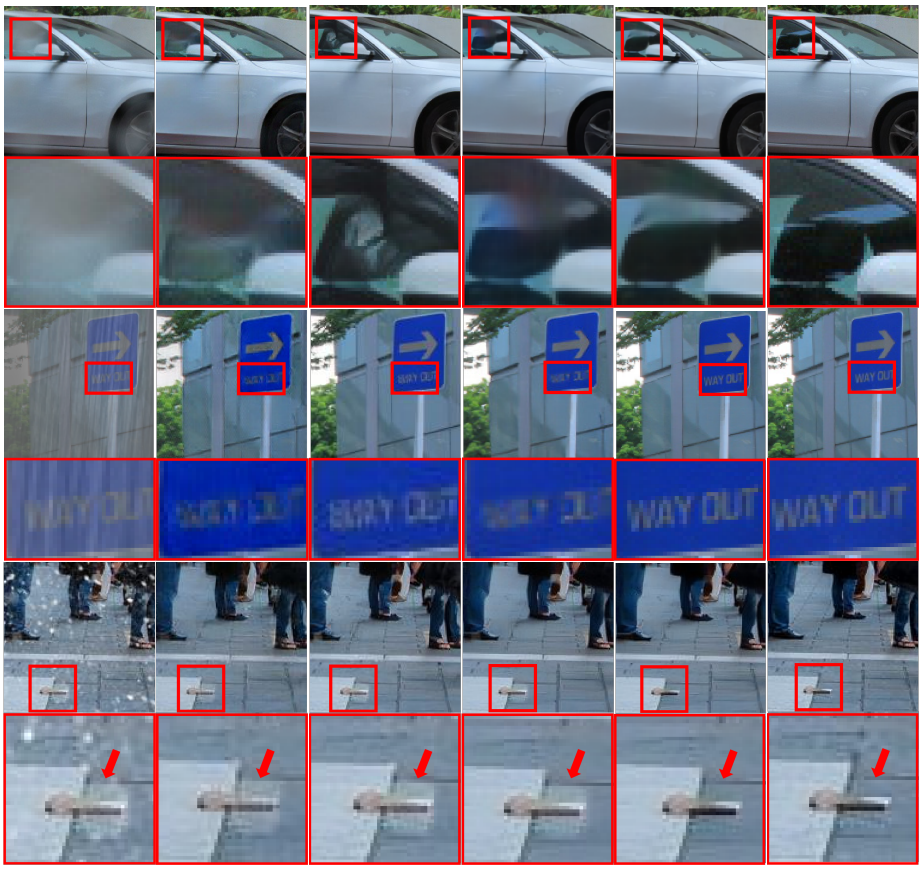
$$\mathcal{L}_{total} = \mathcal{L}_{pixel} + \mathcal{L}_{seq}. \tag{10}$$

4 Experiments

4.1 Experiment Settings

Datasets and Evaluation Metrics. To ensure fair comparison, we follow the standard experimental protocol adopted by previous all-in-one adverse weather restoration methods [20, 28, 37, 38]. The training set consists of 9,000 images from Snow100K [24], 1,069 images from Raindrop [32], and 9,000 images from Outdoor-Rain [19]. Snow100K and Outdoor-Rain are synthetic datasets containing snow, rain, and fog degradations, while Raindrop comprises real-world images degraded by raindrops. For evaluation, we adopt the Test1 set [19, 20], the Raindrop test set [32], and the Snow100K-S and Snow100K-L test sets [24]. In addition, Snow100K provides a real-world test set containing 1,329 images with real adverse weather degradations. We report Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as quantitative evaluation metrics, following common practice in adverse weather image restoration [20].

Implementation Details. Our model is implemented in PyTorch [29] and trained on a single NVIDIA A40 GPU. The network is trained for 300,000 iterations with an initial batch size of 8. Following the progressive learning strategy in Restormer [53], we use an initial patch size of 128. We employ the AdamW optimizer [27] with an initial learning rate of 3×10^{-4} for the first 92,000 iterations, which is then gradually decayed to 1×10^{-6} using a cosine annealing



(a)Input (b)TransW. (c)WeatherDiff (d)Histoformer (e)Ours (f)GT

Fig. 5: Visual comparisons of SSR against state-of-the-art unified methods. **Top Row**: Raindrop removal [28]. **Middle Row**: Joint deraining and dehazing [19, 20]. **Bottom Row**: Image desnowing [24].

schedule [26] over the remaining iterations. The number of blocks at each stage is set to $N_{1,2,3} = \{3, 3, 6\}$, the base channel dimension is 48, and the channel expansion ratio is set to 3. Standard data augmentation including random horizontal and vertical flipping is applied during training.

4.2 Comparison with State-of-the-Art Methods

Quantitative Comparison. We evaluate the performance of SSR against a diverse set of state-of-the-art (SOTA) image restoration methods. These include Transformer-based models such as Restormer [53], AWRCP [50], and Histoformer [30], as well as contemporary SSM-based frameworks like MambaIRv2 [10] and EVSSM [16]. As reported in Tab. 1, SSR achieves SOTA performance across most benchmarks. Notably, on the challenging Outdoor dataset, which consists

Table 1: Quantitative comparisons of adverse weather restoration. The **best** and second-best results are highlighted.

Methods	Snow100K-S		Snow100K-L		Outdoor		RainDrop		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
All-in-One [20] (CVPR'20)	-	-	28.33	0.8820	24.71	0.8980	31.12	0.9268	-	-
SwinIR [23] (ICCV'21)	-	-	28.18	0.880	23.23	0.869	30.82	0.904	-	-
MPRNet [54] (CVPR'21)	-	-	28.66	0.869	30.25	0.914	30.99	0.916	-	-
TKMANet [40] (ICIP'21)	-	-	30.24	0.902	29.92	0.917	30.99	0.927	-	-
AirNet [17] (CVPR'22)	-	-	27.92	0.858	23.12	0.837	28.23	0.892	-	-
TransWeather [38] (CVPR'22)	32.51	0.9341	29.31	0.8879	28.83	0.9000	30.17	0.9157	30.21	0.9094
Chen et al. [2] (CVPR'22)	34.42	0.9469	30.22	0.9071	29.27	0.9147	31.81	0.9309	31.43	0.9249
Restormer [53] (CVPR'22)	36.01	0.9579	30.36	0.9068	30.03	0.9215	32.18	0.9408	32.15	0.9318
WeatherDiff$_{128}$ [28] (TPAMI'23)	35.02	0.9652	29.58	0.8941	29.72	0.9216	29.66	0.9225	31.00	0.9225
WGWSNet [66] (CVPR'23)	34.31	0.9460	30.16	0.9007	29.32	0.9207	32.38	0.9378	31.54	0.9263
WeatherDiff$_{64}$ [28] (TPAMI'23)	35.83	0.9566	30.09	0.9041	29.64	0.9312	30.71	0.9312	31.57	0.9308
AWRCP [50] (ICCV'23)	36.92	0.9652	31.92	**0.9341**	31.39	0.9329	31.93	0.9314	33.04	0.9409
GridFormer [41] (IJCV'24)	37.46	0.9640	31.71	0.9231	31.87	0.9335	32.39	0.9362	33.36	0.9392
Histoformer [30] (ECCV'24)	37.41	0.9656	32.16	0.9261	32.08	0.9389	**33.06**	**0.9441**	33.68	0.9437
Mambairv2 [10] (CVPR'25)	37.13	0.9655	31.75	0.9234	31.45	0.9332	32.22	0.9373	33.14	0.9398
EVSSM [16] (CVPR'25)	37.45	0.9655	32.15	0.9250	32.06	0.9380	32.40	0.9400	33.52	0.9421
MoCE-IR [52] (CVPR'25)	37.10	0.9654	31.88	0.9234	31.42	0.9334	32.23	0.9386	33.16	0.9400
CPL$_{PromptIR}$ [43] (TPAMI'26)	37.82	**0.9667**	32.27	0.9280	32.16	0.9420	32.73	0.9430	33.74	0.9438
SSR(Ours)	**37.87**	0.9656	**32.70**	0.9297	**33.22**	**0.9421**	32.82	0.9440	**34.15**	**0.9442**

of complex joint rain and haze degradations, our method outperforms the second-best competitor (CPL$_{PromptIR}$ [43]) by a significant margin of 1.06 dB in PSNR. Furthermore, SSR attains the best results in terms of average PSNR (34.15 dB) and SSIM (0.9442) across all test sets, demonstrating its superior robustness and generalization capability in handling spatially non-uniform degradations. While Histoformer [30] and CPL$_{PromptIR}$ [43] achieve slightly better results on specific synthetic benchmarks (e.g., Snow100K-S and RainDrop), SSR consistently ranks as the best or second-best method across these tasks. This confirms that our proposed superpixel-guided strategy effectively handles localized artifacts (like raindrops) without sacrificing the modeling of large-scale structural context.

Qualitative Comparison We provide visual comparisons of SSR against leading unified methods across three diverse adverse weather scenarios in Fig. 5 On the raindrop removal task (Top Row), while methods such as TransWeather [38] (b) and WeatherDiff [28] (c) struggle to restore the car's interior details obscured by heavy drops, SSR effectively eliminates the artifacts while preserving sharp structural content. For the challenging joint deraining and dehazing task (Middle Row), SSR excels in restoring fine texture details; notably, the text "WAY OUT" on the blue sign remains clearly legible in our result, validating the efficacy of our semantic-guided scanning mechanism. Finally, in the image desnowing task (Bottom Row), SSR demonstrates superior detail preservation by eliminating dense snowflakes without introducing blurring artifacts, especially on the sharp edges of small objects (indicated by red arrows). Overall, these visual results underscore SSR's remarkable capability to adaptively model spatially non-uniform weather degradations while maintaining high perceptual fidelity.

Efficiency and Model Complexity Analysis Based on the data presented in Tab. 2, we comprehensively evaluate the proposed SSR from the perspectives of performance, parameter efficiency, and computational complexity. Compared to a broad range of state-of-the-art unified restoration models, SSR achieves superior performance on the Outdoor dataset (33.22 dB) while maintaining remarkable parameter efficiency. As shown in Tab. 2, our method utilizes only 7.05M parameters, which is the lowest among all compared approaches.

Compared to recent high-performing methods such as CPL$_{PromptIR}$ [43], SSR achieves a significant performance gain of +1.06 dB while reducing the model size by approximately 80.4%. Even against the second most parameter-efficient model, AirNet [17], our method demonstrates a clear advantage in both restoration quality (+10.10 dB) and parameter economy (-1.88M).

Table 2: Performance and Complexity Comparison on the Outdoor Dataset.

Method	Param./M	FLOPs/G	PSNR/dB
AirNet [17]	<u>8.93</u>	**30.13**	23.12
Restormer [53]	26.12	141.00	30.03
WeatherDiff [28]	82.96	–	29.64
Histoformer [30]	16.61	91.65	32.08
MambairV2 [10]	15.42	76.50	31.45
EVSSM [16]	17.13	122.23	32.06
CPL$_{PromptIR}$ [43]	36.00	–	32.16
SSR (Ours)	**7.05**	<u>54.27</u>	**33.22**

In terms of computational complexity measured by FLOPs, SSR strikes a highly favorable balance between restoration capability and execution efficiency. With 54.27G FLOPs, SSR requires less than half the computation of previous Transformer-based heavyweights like Restormer [53] (141.00G) and is significantly more economical than recent unified models such as EVSSM [16] (122.23G) and Histoformer [30] (91.65G). It also notably undercuts the baseline MambaIRv2 [10] (76.50G). While AirNet [17] achieves lower FLOPs (30.13G), it suffers from a severe performance degradation (23.12 dB) compared to SSR. Overall, these quantitative results rigorously validate that our proposed SSR is a highly competitive unified weather restoration model, achieving SOTA performance while maintaining high efficiency in both model footprint and computational cost.

Table 3: Quantitative comparison on real-world datasets. Left: Evaluation using Q-Align (↑). Right: Evaluation using NIQE and IL-NIQE (↓) along with parameter comparisons.

Method	RainDrop-real		NTURain-real	
	Quality	Aesthetic	Quality	Aesthetic
WeatherDiff [28]	0.6550	0.3231	0.6582	0.2537
Histoformer [30]	0.6551	0.3234	0.6584	0.2541
SSR (Ours)	**0.6591**	**0.3246**	**0.6587**	**0.2561**

Method	Snow100K-Real		RainDS		Param
	NIQE↓	IL-NIQE↓	NIQE↓	IL-NIQE↓	
TransWeather [38]	3.161	22.207	4.005	22.512	38.05M
WeatherDiff_64 [28]	2.985	22.121	3.050	19.800	82.96M
WeatherDiff_128 [28]	2.964	21.976	3.642	19.972	85.61M
SSR (Ours)	**2.879**	**21.909**	**2.924**	**19.215**	**7.05M**

Real-World Evaluation To provide quantitative evidence of our model's performance in challenging real-world scenarios, we evaluate our SSR framework on real-world adverse weather images. Since real-world degradations lack ground-truth clear images for standard reference-based evaluation (e.g., PSNR/SSIM),

we employ no-reference metrics to evaluate the restored results. Specifically, we use the multi-modal no-reference metric, Q-Align [46], to assess both perceptual *quality* and *aesthetic* scores on the real-world subsets of the RainDrop [28] and NTURain [51] datasets. In addition, we further evaluate our model on real-world datasets by adopting NIQE and IL-NIQE on Snow100K-Real and RainDS benchmarks. The results are presented in Tab. 3.

As the results show, our method achieves state-of-the-art performance across all datasets. In the Q-Align dimensions, our method outperforms the recent Transformer-based method, Histoformer [30], and the diffusion-based approach, WeatherDiff [28], in both quality and aesthetic scores. Meanwhile, on the Snow100K-Real and RainDS benchmarks, SSR outperforms existing methods by a clear margin while maintaining a significantly lower parameter budget.

4.3 Ablation Study

We conduct comprehensive ablation studies to analyze the impact of key design choices in our framework. Unless otherwise specified, all ablation experiments are performed under the same SSM-based architecture, and only the corresponding scanning or grouping strategy is modified.

Table 4: Ablation study on the number of superpixels (K) across different datasets.

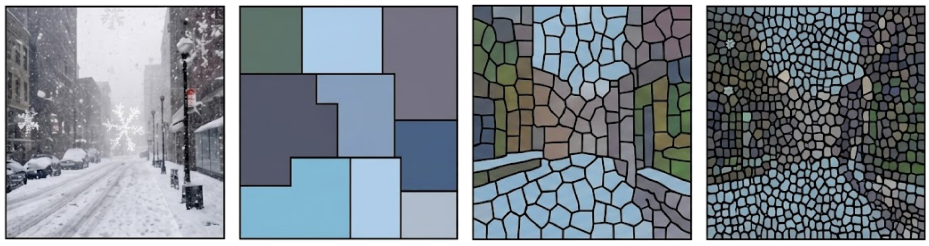
K	FLOPs	Snow100K-S		Snow100K-L		Outdoor		RainDrop		Average	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
8	**4.21G**	37.34	0.9715	31.62	0.9213	31.68	0.9218	32.25	0.9391	33.22	0.9384
16	4.30G	**37.42**	**0.9717**	**31.90**	**0.9227**	**31.91**	**0.9239**	**32.37**	**0.9396**	**33.40**	**0.9395**
32	4.71G	37.26	0.9714	31.88	0.9224	31.57	0.9212	32.28	0.9394	33.25	0.9386

Effect of the Number of Superpixels K. Tab. 4 investigates the influence of the number of superpixels K across different datasets, while the visual impact of region partitioning for different values of K is illustrated in Fig. 6. When K is set to a smaller value ($K = 8$), the image is partitioned into overly coarse regions, which limits the network's ability to accurately model fine-grained and spatially non-uniform degradation patterns. As K increases to 16, the performance peaks across all benchmark datasets, achieving the highest average PSNR of 33.40 dB and SSIM of 0.9395. This indicates that an appropriate region partitioning granularity enables optimal structure-aware state propagation. However, when K is further increased to 32, the performance slightly degrades. Excessively large K results in fragmented regions, diminishing the benefits of regional coherence and causing the scanning behavior to regress toward a pixel-wise traversal. Furthermore, as shown in Tab. 4, a larger K inevitably leads to higher overhead (e.g., FLOPs increases from 4.30G to 4.71G). Consequently, we adopt $K = 16$

Table 5: Comprehensive ablation studies of different scanning and gating strategies across four datasets. Default settings used in our final framework are highlighted in **bold**.

	Method	Outdoor	Snow100K-S	Snow100K-L	RainDrop
Scanning Strategy	Raster Scan	31.77 / 0.9130	36.53 / 0.9548	31.52 / 0.8823	31.82 / 0.9034
	Fixed Window Scan	31.92 / 0.9215	36.78 / 0.9602	31.75 / 0.8904	31.94 / 0.9058
	Hilbert Scan	32.03 / 0.9246	36.72 / 0.9557	31.79 / 0.8923	32.01 / 0.9067
	Superpixel (Ours)	**32.13 / 0.9288**	**36.93 / 0.9613**	**31.86 / 0.8946**	**32.16 / 0.9102**
Gating Strategy	w/o Gating	32.13 / 0.9288	36.93 / 0.9613	31.86 / 0.8946	32.16 / 0.9102
	Global Gating	32.91 / 0.9398	37.23 / 0.9608	32.19 / 0.9237	32.62 / 0.9396
	Region Gating (Ours)	**33.22 / 0.9401**	**37.87 / 0.9656**	**32.70 / 0.9297**	**32.82 / 0.9440**

as the default setting for all experiments to achieve the best balance between restoration quality, regional structural integrity, and computational efficiency.



(a) Input (b) $K = 8$ (c) $K = 16$ (d) $K = 32$

Fig. 6: Visualization of the image partitioning with different values of K. As shown, $K = 8$ leads to overly coarse and mixed semantic features, while $K = 32$ results in fragmented regions. $K = 16$ provides the most continuous and coherent semantic representation.

Effect of the Scanning Strategies. We investigate the impact of various scanning patterns to validate the proposed Superpixel-guided Selective Scan. Quantitative results across four benchmarks and visual partitioning are provided in Tab. 5 and Fig. 6, respectively. Compared to conventional content-agnostic trajectories on the Outdoor dataset, such as Raster Scan (31.77 dB), Fixed Window Scan (31.92 dB), and Hilbert Scan (32.03 dB), our structure-aware scanning strategy achieves superior performance with 32.13 dB PSNR. While the superpixel clustering and permutation operations introduce a slight increase in computational cost (4.28G vs. 4.17G for Raster Scan), the mechanism successfully shifts the scanning paradigm from pixel-serial to a semantic-guided one. By partitioning the image into perceptually coherent regions and performing relations modeling within these semantic-related areas, our approach effectively prevents the model from learning non-discriminative features from semantic-conflict regions. This ensures that the state evolution adheres to semantic boundaries, lead-

ing to more effective restoration compared to rigid, predefined trajectories. This performance advantage consistently holds true across the other three weather datasets in Tab. 5, where our superpixel scanning steadily delivers gains, outperforming the advanced Hilbert scanning by over 0.2 dB on the Snow100K-S benchmark.

Effect of Region-Level Gating. We evaluate the impact of the proposed Region-level Gating Mechanism (RGM) to verify its ability in performing intra-region semantic calibration. As shown in Tab. 5, the "w/o Gating" baseline, which relies solely on the superpixel-guided scanning, yields 32.13 dB PSNR on the Outdoor dataset. While "Global Gating" reaches 32.91 dB, its image-wide statistics fail to capture the spatially non-uniform distribution of degradations. By transitioning to our proposed RGM (Region Gating), the performance is further boosted to the peak of 33.22 dB PSNR and 0.9401 SSIM. This improvement demonstrates that by adaptively modulating channel-wise feature responses based on regional semantic statistics rather than global ones, the model can more effectively emphasize restoration-critical information within each semantic unit while suppressing non-discriminative features from semantic-conflict regions. Notably, although RGM introduces a marginal increase in computational complexity (with an overhead of only 0.04G), this negligible cost leads to a significant performance gain of 1.09 dB PSNR over the baseline, confirming that regional calibration is essential for handling complex weather scenarios. As validated across the Snow100K-S, Snow100K-L, and RainDrop benchmarks in Tab. 5, this regional strategy maintains a similar upward trend, where Region Gating improves upon Global Gating by over 0.5 dB on both the Snow100K-S and Snow100K-L datasets.

4.4 Discussion and Limitations

Although SSR achieves SOTA restoration performance with significantly reduced theoretical computational complexity, its practical inference speed warrants further discussion. As shown in Tab. 6, while SSR is the fastest among Mamba-based models, it remains slightly slower than highly optimized Transformer counterparts. Unlike Transformers, which benefit from mature CUDA ecosystems and highly-fused kernels like FlashAttention, hardware-level acceleration for Mamba-based selective scan mechanisms remains in its infancy. Furthermore, our dynamic superpixel clustering and region-guided selective scanning require frequent memory permutations and irregular spatial indexing, leading to increased synchronization overheads on modern GPUs. Exploring customized, hardware-efficient CUDA kernels to mitigate these bottlenecks and fully unleash the linear-complexity potential of SSMs will be a key focus of our future work.

Furthermore, the robustness of SSR inherently depends on the accuracy of the superpixel clustering. Since the segmentation process is content-dependent, extreme weather conditions (e.g., dense snow or torrential rain) can severely corrupt the underlying image structures. In such failure cases, the clustering

Table 6: Inference latency (s) comparison among representative Mamba-based and Transformer-based architectures.

Transformer-based	**Venue**	**Time**	**Mamba-based**	**Venue**	**Time**
GridFormer	IJCV'24	1.110	MambaIR	ECCV'24	2.599
Restormer	CVPR'22	0.853	MambaIRv2	CVPR'25	1.918
SwinIR	ICCVW'21	0.823	EVSSM	CVPR'25	1.267
Histoformer	ECCV'24	**0.716**	**SSR (Ours)**	–	**0.895**

algorithm may erroneously over-segment heavy degradations into independent semantic regions. This causes our semantic-guided trajectories to degrade into localized noise-fitting paths, thereby limiting the model's ability to leverage long-range clean contexts. Future research will explore degradation-robust clustering mechanisms and multi-scale semantic guidance to enhance the model's stability in severely corrupted scenarios.

5 Conclusions

In this paper, we propose **SSR**, a **S**emantic-center guilded **S**tate space model for image **R**estoration. Unlike conventional SSMs using rigid, content-agnostic scanning, SSR shifts the paradigm from pixel-serial to superpixel-based semantic-guided one. By incorporating superpixel-guided structural cues via the proposed $\mathbf{S}^3\mathbf{M}$, our method ensures recursive state propagation is strictly confined within perceptually coherent regions, preventing the aggregation of semantically conflicting features across boundaries. Furthermore, the **RGM** is developed to perform intra-region calibration by modulating degradation outliers within each semantic superpixel unit along the channel dimension. Extensive experiments on six benchmarks demonstrate that SSR achieves SOTA performance with superior parameter efficiency. This work highlights semantic-aware SSMs as an efficient and interpretable alternative to complex attention-based designs for universal image restoration.

Acknowledgements

This work was supported by the Shenzhen Science and Technology Program (No. JCYJ20240813114229039), PCL Major Key Project of PCL2025A17-2, the Natural Science Foundation of Tianjin, China (No. 24JCZXJC00040), the National Natural Science Foundation of China (No. 624B2072), the Doctoral Student Program of the Young S&T Talents Cultivation Project, CAST, the Fundamental Research Funds for the Central Universities (No. 63263253), and the Supercomputing Center of Nankai University (NKSC).

References

1. Bai, J., Yin, Y., He, Q., Li, Y., Zhang, X.: Retinexmamba: Retinex-based mamba for low-light image enhancement. In: ICONIP (2024)
2. Chen, W.T., Huang, Z.K., Tsai, C.C., Yang, H.H., Ding, J.J., Kuo, S.Y.: Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In: CVPR (2022)
3. Conde, M.V., Geigle, G., Timofte, R.: Instructir: High-quality image restoration following human instructions. In: ECCV (2024)
4. Cui, Y., Ren, W., Cao, X., Knoll, A.: Image restoration via frequency selection. TPAMI (2023)
5. Cui, Y., Ren, W., Cao, X., Knoll, A.: Revitalizing convolutional network for image restoration. TPAMI (2024)
6. Duosheng, C., Shihao, Z., Jinshan, P., Jinglei, S., Lishen, Q., Jufeng, Y.: A polarization-aided transformer for image deblurring via motion vector decomposition. In: CVPR (2025)
7. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: CVPR (2017)
8. Gasperini, S., Morbitzer, N., Jung, H.J., Navab, N., Tombari, F.: Robust monocular depth estimation under challenging conditions. In: ICCV (2023)
9. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. In: CoLM (2024)
10. Guo, H., Guo, Y., Zha, Y., Zhang, Y., Li, W., Dai, T., Xia, S.T., Li, Y.: Mambairv2: Attentive state space restoration. In: CVPR (2025)
11. Guo, H., Li, J., Dai, T., Ouyang, Z., Ren, X., Xia, S.T.: Mambair: A simple baseline for image restoration with state-space model. In: ECCV (2024)
12. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. TPAMI (2010)
13. Huang, T., Pei, X., You, S., Wang, F., Qian, C., Xu, C.: Localmamba: Visual state space model with windowed selective scan. In: ECCV (2024)
14. Jiang, T.X., Huang, T.Z., Zhao, X.L., Deng, L.J., Wang, Y.: A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors. In: CVPR (2017)
15. Kang, L.W., Lin, C.W., Fu, Y.H.: Automatic single-image-based rain streaks removal via image decomposition. TIP (2011)
16. Kong, L., Dong, J., Tang, J., Yang, M.H., Pan, J.: Efficient visual state space model for image deblurring. In: CVPR (2025)
17. Li, B., Liu, X., Hu, P., Wu, Z., Lv, J., Peng, X.: All-in-one image restoration for unknown corruption. In: CVPR (2022)
18. Li, D., Liu, Y., Fu, X., Xu, S., Zha, Z.J.: Fouriermamba: Fourier learning integration with state space models for image deraining. In: ICML (2025)
19. Li, R., Cheong, L.F., Tan, R.T.: Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In: CVPR (2019)
20. Li, R., Tan, R.T., Cheong, L.F.: All in one bad weather removal using architectural search. In: CVPR (2020)
21. Li, S., Singh, H., Grover, A.: Mamba-nd: Selective state space modeling for multi-dimensional data. In: ECCV (2024)
22. Li, X., Wu, J., Lin, Z., Liu, H., Zha, H.B.: Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: ECCV (2018)

23. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: ICCV (2021)
24. Liu, Y.F., Jaw, D.W., Huang, S.C., Hwang, J.N.: Desnownet: Context-aware deep network for snow removal. TIP (2018)
25. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., Liu, Y.: Vmamba: Visual state space model. NeurIPS (2024)
26. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. In: ICLR (2017)
27. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: ICLR (2019)
28. Özdenizci, O., Legenstein, R.: Restoring vision in adverse weather conditions with patch-based denoising diffusion models. TPAMI (2023)
29. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. NeurIPS (2019)
30. Peng, Y.T., Chen, Y.R., Chen, G.R., Liao, C.J.: Histoformer: Histogram-based transformer for efficient underwater image enhancement. JOE (2024)
31. Potlapalli, V., Zamir, S.W., Khan, S.H., Khan, F.S.: Promptir: Prompting for all-in-one image restoration. NeurIPS (2023)
32. Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J.: Attentive generative adversarial network for raindrop removal from a single image. In: CVPR (2018)
33. Qu, L., Zhou, S., Liang, J., Zeng, H., Zhang, L., Yang, J.: It takes two: A duet of periodicity and directionality for burst flicker removal. In: CVPR (2026)
34. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D (1992)
35. Sakaridis, C., Dai, D., Van Gool, L.: Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In: ICCV (2021)
36. Shi, Y., Xia, B., Jin, X., Wang, X., Zhao, T., Xia, X., Xiao, X., Yang, W.: Vmambair: Visual state space model for image restoration. TCSVT (2025)
37. Sun, S., Ren, W., Gao, X., Wang, R., Cao, X.: Restoring images in adverse weather conditions via histogram transformer. In: ECCV (2024)
38. Valanarasu, J.M.J., Yasarla, R., Patel, V.M.: Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In: CVPR (2022)
39. Wang, H., Hu, Q., Guo, X.: Modem: A morton-order degradation estimation mechanism for adverse weather image recovery. NeurIPS (2025)
40. Wang, H., Wang, W., Liu, J.: Temporal memory attention for video semantic segmentation. In: ICIP (2021)
41. Wang, T., Zhang, K., Shao, Z., Luo, W., Stenger, B., Lu, T., Kim, T.K., Liu, W., Li, H.: Gridformer: Residual dense transformer with grid structure for image restoration in adverse weather conditions. IJCV (2024)
42. Weng, J., Yan, Z., Tai, Y., Qian, J., Yang, J., Li, J.: Mamballie: Implicit retinex-aware low light enhancement with global-then-local state space. NeurIPS (2024)
43. Wu, G., Jiang, J., Jiang, K., Liu, X., Nie, L.: Beyond degradation redundancy: Contrastive prompt learning for all-in-one image restoration. TPAMI (2025)
44. Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., Ma, L.: Contrastive learning for compact single image dehazing. In: CVPR (2021)
45. Wu, H., Yang, Y., Xu, H., Wang, W., Zhou, J., Zhu, L.: Rainmamba: Enhanced locality learning with state space models for video deraining. In: ACM MM (2024)
46. Wu, H., Zhang, Z., Zhang, W., Chen, C., Liao, L., Li, C., Gao, Y., Wang, A., Zhang, E., Sun, W., et al.: Q-align: Teaching lmms for visual scoring via discrete text-defined levels. ICML (2024)

47. Yamashita, S., Ikehara, M.: Image deraining with frequency-enhanced state space model. In: ACCV (2024)
48. Yang, Y., Guo, C.L., Guo, X.: Depth-aware unpaired video dehazing. TIP (2024)
49. Yasarla, R., Patel, V.M.: Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In: CVPR (2019)
50. Ye, T., Chen, S., Bai, J., Shi, J., Xue, C., Jiang, J., Yin, J., Chen, E., Liu, Y.: Adverse weather removal with codebook priors. In: ICCV (2023)
51. Yue, Z., Xie, J., Zhao, Q., Meng, D.: Semi-supervised video deraining with dynamical rain generator. In: CVPR (2021)
52. Zamfir, E., Wu, Z., Mehta, N., Tan, Y., Paudel, D.P., Zhang, Y., Timofte, R.: Complexity experts are task-discriminative learners for any image restoration. In: CVPR (2025)
53. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: CVPR (2022)
54. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: CVPR (2021)
55. Zhang, A., Ren, W., Liu, Y., Cao, X.: Lightweight image super-resolution with superpixel token interaction. In: ICCV (2023)
56. Zhang, K., Li, D., Luo, W., Ren, W.: Dual attention-in-attention model for joint rain streak and raindrop removal. TIP (2021)
57. Zhang, K., Li, R., Yu, Y., Luo, W., Li, C.: Deep dense multi-scale network for snow removal using semantic and depth priors. TIP (2021)
58. Zhao, R., Yan, H., Wang, S.: Revisiting domain-adaptive object detection in adverse weather by the generation and composition of high-quality pseudo-labels. In: ECCV (2024)
59. Zheng, Z., Wu, C.: U-shaped vision mamba for single image dehazing. arXiv preprint arXiv:2402.04139 (2024)
60. Zhou, H., Wu, X., Chen, H., Chen, X., He, X.: Rsdehamba: Lightweight vision mamba for remote sensing satellite image dehazing. arXiv preprint arXiv:2405.10030 (2024)
61. Zhou, S., Li, D., Pan, J., Zhou, J., Shi, J., Yang, J.: Devil is in the uniformity: Exploring diverse learners within transformer for image restoration. In: ICCV (2025)
62. Zhou, S., Pan, J., Chen, D., Dong, Y., Yang, J.: Rethinking the importance of high-frequency components in transformers for image restoration. TIP (2026)
63. Zhou, S., Pan, J., Yang, J.: Learning an adaptive sparse transformer for efficient image restoration. TPAMI (2025)
64. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. In: ICML (2024)
65. Zhu, R., Tu, Z., Liu, J., Bovik, A.C., Fan, Y.B.: Mwformer: Multi-weather image restoration using degradation-aware transformers. TIP (2024)
66. Zhu, Y., Wang, T., Fu, X., Yang, X., Guo, X., Dai, J., Qiao, Y., Hu, X.w.: Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In: CVPR (2023)
67. Zou, W., Gao, H., Yang, W., Liu, T.: Wave-mamba: Wavelet state space model for ultra-high-definition low-light image enhancement. In: ACM MM (2024)
68. Zou, Z., Yu, H., Huang, J., Zhao, F.: Freqmamba: Viewing mamba from a frequency perspective for image deraining. In: ACM MM (2024)