PRE-Mamba: A 4D State Space Model for Ultra-High-Frequent Event Camera Deraining

Ciyu Ruan[1,*], Ruishan Guo[1,*], Zihang Gong[2], Jingao Xu[3], Wenhan Yang[4], Xinlei Chen[1,†]
[1]Shenzhen International Graduate School, Tsinghua University, [2]Harbin Institute of Technology,
[3]Carnegie Mellon University, [4]Pengcheng Laboratory,
{rcy23, grs24}@mails.tsinghua.edu.cn, gongzihang0201@gmail.com,
jingaox@andrew.cmu.edu, yangwh@pcl.ac.cn, chen.xinlei@sz.tsinghua.edu.cn

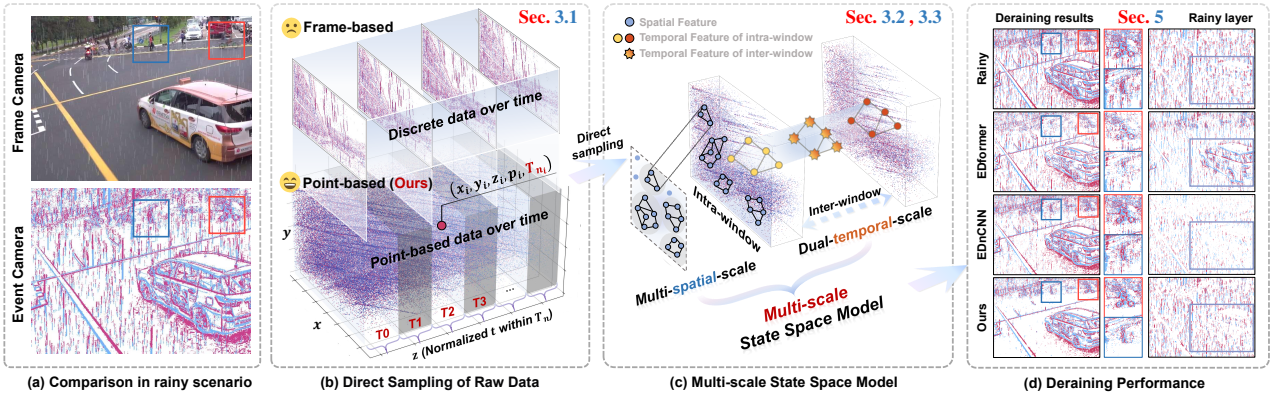

Figure 1. (a) Rain severely degrades event camera data compared to conventional cameras. (b) Unlike frame-based aggregation methods, our approach directly samples raw data, preserving temporal resolution. (c) The multi-scale state space model integrates dual-temporal scales for intra- and inter-window information and multi-spatial scales for rain dynamics. (d) Deraining results of our method.

Abstract

Event cameras excel in high temporal resolution and dynamic range but suffer from dense noise in rainy conditions. Existing event deraining methods face trade-offs between temporal precision, deraining effectiveness, and computational efficiency. In this paper, we propose PRE-Mamba, a novel point-based event camera deraining framework that fully exploits the spatiotemporal characteristics of raw event and rain. Our framework introduces a 4D event cloud representation that integrates dual temporal scales to preserve high temporal precision, a Spatio-Temporal Decoupling and Fusion module (STDF) that enhances deraining capability by enabling shallow decoupling and interaction of temporal and spatial information, and a Multi-Scale State Space Model (MS3M) that captures deeper rain dynamics across dual-temporal and multi-spatial scales with linear computational complexity. Enhanced by frequency-domain regularization, PRE-Mamba achieves superior performance (0.95 SR, 0.91 NR, and 0.4s/M events) with only 0.26M parameters on EventRain-27K, a comprehensive dataset with labeled synthetic and real-world sequences. Moreover, our method generalizes well across varying rain intensities, viewpoints, and even snowy conditions. Code and dataset: https://github.com/softword-tt/PRE-Mamba.

1. Introduction

Event cameras have revolutionized mobile applications with their high temporal resolution (microsecond level), high dynamic range (>120 dB), and low power consumption (<10 mW) [1, 2], driving breakthroughs in high-speed tracking [3, 4], SLAM [5–7], localization [8] and obstacle avoidance [9, 10]. Unlike frame cameras, they asynchronously capture pixel-level intensity changes without global exposure, eliminating motion blur even at extreme speeds. However, despite excelling in dynamic scenes, event cameras suffer significant performance degradation under adverse weather, particularly in rain [11, 12]. The rapid motion of rain streaks triggers excessive intensity changes, generating dense noise that overwhelms valid scene information and exacerbates sampling rates. This severely impacts accuracy and latency in downstream tasks on drones [13–15] and autonomous vehicles [16–18], limiting real-world reliability.

Despite significant advances in deep learning for frame

*These authors contributed equally.
†Corresponding author.

camera deraining [19–22], event camera deraining remains largely unexplored. Existing solutions predominantly follow a frame-based paradigm, converting event streams into grayscale frames [12] and applying frame-based deraining techniques [11], which sacrifice the fine-grained temporal resolution and sparsity of event data. Point-based methods align naturally with event streams' asynchronous, sparse nature, but remain unapplied to deraining due to notable limitations. Point Networks [23, 24], GNNs [25, 26], and SNNs [27–29] still underperform compared to frame-based methods, while Point Transformers face computational bottlenecks under high event rates (>10M events/s) due to quadratic attention complexity [30, 31]. As a result, balancing temporal precision, deraining effectiveness, and computational efficiency remains an open challenge.

Recently, the State Space Model (SSM) [32], particularly Mamba, provides a promising solution with linear complexity and long-range context modeling. However, introducing Mamba to event camera deraining is non-trivial. The original design of the Mamba is aimed at solving causal sequential language tasks, which differ from the asynchronous, sparse, and high-temporal-resolution nature of event data. Rain streaks further complicate this by introducing diverse spatiotemporal patterns, requiring the model to capture complex dependencies across both time and space.

To address this, we propose **PRE-Mamba**, the first **P**oint-based de**R**aining framework for **E**vent camera. As a point-based method, it processes raw events directly instead of quantized frames, fully preserving the sensor's native microsecond (μs) temporal resolution and facilitating downstream high-speed vision tasks. Our framework first introduces a 4D event cloud representation (§3.1), integrating inter- and intra-temporal windows to preserve high temporal precision. To handle the spatial irregularity and motion characteristics of rain, we propose a Spatio-Temporal Decoupling and Fusion module (STDF, §3.2), enabling shallow decoupling and interaction of temporal and spatial information. For deeper modeling, we propose a Multi-Scale State Space Model (MS3M, §3.3) that efficiently captures complex rain dynamics across dual temporal and multi-spatial scales with linear computational complexity. To improve feature discrimination, we introduce a frequency regularization term (§3.4) in the loss function, guiding the model to learn rain distribution patterns in the frequency domain. To evaluate our approach, we construct EventRain-27K (§4), the first comprehensive point-based event deraining dataset, comprising 18K labeled synthetic and 9K unlabeled real-world rain sequences. Our model achieves state-of-the-art performance (0.95 SR, 0.91 NR, and 0.4s/M events) with only 0.26M parameters, validated through extensive experiments (§5). Moreover, it generalizes well across diverse rain intensities and snowy conditions. Our contributions are summarized as follows:

- To the best of our knowledge, we propose the first point-based event deraining framework.
- To effectively extend Mamba for event deraining, we address the asynchronous, sparse nature of event data and the diverse spatiotemporal patterns of rain streaks by introducing a 4D event cloud, an STDF module, and an MS3M enhanced with frequency regularization loss.
- We introduce EventRain-27K, the first point-based event deraining with labeled synthetic and real-world sequences, providing a benchmark for future research.
- We achieve superior performance on synthetic and real datasets with high efficiency and a lightweight design.

2. Related Work

Existing deraining approaches for frame camera. Deraining methods for frame cameras are categorized into traditional and deep learning-based approaches. Traditional methods, such as dictionary learning and low-rank matrix decomposition [33–36], rely on handcrafted assumptions about the directional nature or sparsity of rain streaks, often failing to generalize to complex real-world scenarios. In contrast, deep learning-based methods, leveraging CNNs and Transformer [37, 38] architectures, have shown significant advances. CNNs effectively model local spatial patterns, capturing rain stripe textures and background structures [19, 39], while Transformer-based models [20–22] enhance performance by capturing long-range dependencies and global context, offering superior deraining capabilities.

Existing deraining approaches involving event camera. Research on event-based deraining remains limited despite its importance for outdoor applications. Existing methods, such as EGVD [40], RainVID&SS [41], and NESD [42], leverage the high temporal resolution and dynamic range of event cameras to assist frame camera by providing prior knowledge of rain location and density. Other methods designed for event-based deraining typically convert asynchronous event streams into synchronous frame-like representations, aligning with image deraining but sacrificing temporal resolution and sparsity. For instance, DistillNet [11] partially preserves temporal resolution through voxel representations but still loses their asynchronous nature. Moreover, the unique motion and spatiotemporal distribution of rain create a domain barrier, limiting the applicability of traditional event denoising methods like filtering [43–45] and clustering [46]. In addition, existing deraining datasets primarily focus on frame camera [47–50] or represent event data in image-like formats [51], further hinder the event camera deraining research.

Event Data Processing Architectures. Event data processing approaches are broadly categorized into frame-based and point-based frameworks. Frame-based methods [45, 52–54] not only face the aforementioned limitations but also incur redundant computational overhead by

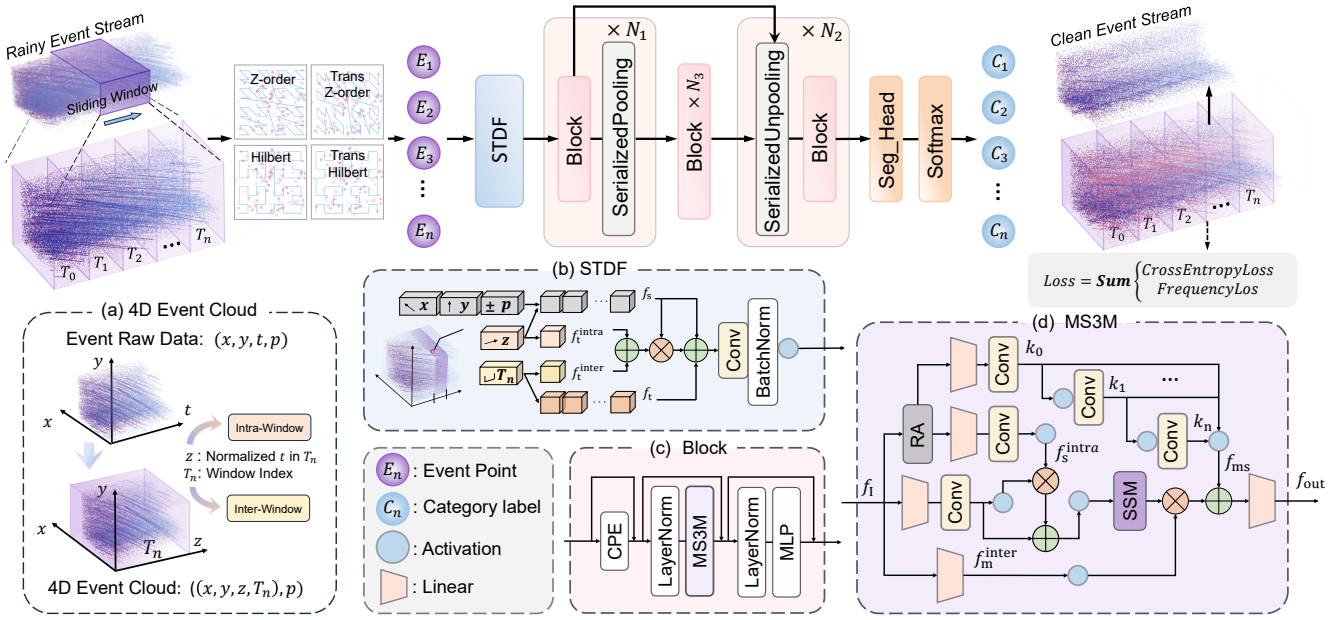

Figure 2. **Overview of Our Proposed PRE-Mamba.** The model takes n temporal windows as input to generate a 4D event cloud, serialized via four scanning modes. The STDF module decouples and enhances spatiotemporal features across dual temporal scales, which flow through a U-Net style encoder-decoder with multiple blocks. Each block employs MS3M as its core, integrating intra- and inter-window branches to capture appearance and motion features, alongside a multi-spatial-scale pathway for handling diverse rain streak scales while preserving local dynamics. Finally, the decoder outputs are fed into a linear layer followed by softmax for per-event predictions.

processing event-free regions [55], making them inefficient for resource-constrained environments [53, 56]. In contrast, point-based methods process events in their native form, preserving temporal resolution and sparsity. These include Point Networks [57, 58], Graph Neural Networks (GNNs) [59], and Spiking Neural Networks (SNNs) [29, 60], excel in capturing rapid motion and fine-grained temporal dynamics. However, GNNs and SNNs often require specialized hardware or face performance issues, while Point Cloud networks, based on assumptions of permutation and transformation invariance, are poorly suited for temporal event data and require time-aware feature extraction.

To address these limitations, this paper proposes a novel approach based on selective State Space Models (SSMs) [61–63]. Renowned for its linear computational complexity and parameter efficiency [32, 64], Mamba is well-suited for processing long event sequences exceeding 1M length [30]. It provides a promising direction to bridge performance gap between point-based and frame-based methods [30, 65] while preserving the inherent advantages of event data, facilitating the development of lightweight, real-time deraining systems.

3. Method

To perform event deraining effectively and efficiently while preserving temporal precision, we propose a novel framework, PRE-Mamba. We formulate event deraining as an event-by-event classification task, akin to event denoising, reducing computational complexity and avoiding artifacts compared to reconstruction-based methods. Figure 2 illustrates the architecture of our system. At the input level, a 4D Event Cloud Representation (§3.1) integrates inter- and intra-temporal windows to preserve high temporal precision. For feature extraction, a Spatio-Temporal Decoupling and Fusion module (§3.2) enables effective decoupling and interaction of temporal and spatial information to capture complex rain dynamics. For deep modeling, a Multi-Scale State Space Model (§3.3) captures rain dynamics across dual temporal and multi-spatial scales with linear computational complexity. Finally, a Frequency-Regularized Optimization (§3.4) enhances feature discrimination by learning rain distribution patterns in the frequency domain, ensuring robust deraining accuracy. Together, these components form a unified framework that achieves both computational efficiency and deraining performance.

3.1. 4D Event Cloud Representation

To preserve fine-grained spatiotemporal information while enabling efficient processing, we transform raw event streams into a structured 4D event cloud representation. The raw event data is represented as a sequence $\mathcal{E} = \{e_i\}_{i=1}^{N}$, where each event $e_i = (x_i, y_i, t_i, p_i)$ consists of spatial coordinates (x_i, y_i), timestamp t_i, and polarity p_i.

3D Pseudo-Point Cloud Construction: The event sequence is divided into fixed-duration temporal windows $\{\mathcal{W}_k\}_{k=1}^{L}$, each spanning a time interval T. Within each window $\mathcal{W}_k$, events are normalized into a 3D pseudo-point cloud representation:

$$p_i = (x_i, y_i, z_i) = \left(x_i, y_i, \frac{t_i - t_0}{t_e - t_0} \right), \tag{1}$$

where t_0 and t_e denote the window's start and end timestamps, respectively. The normalized coordinate z_i encodes the relative temporal position of the event within the window, mapping event timestamps to a unified temporal scale.

4D Cross-Window Temporal Modeling: To model long-term dependencies across windows, we introduce a fourth dimension, T_n, which represents the global temporal index of the window. This extends the 3D representation to a 4D event cloud $E_n = \{(x_i, y_i, z_i, T_n)\}_{i=1}^{M_n}$, where M_n is the number of events in window n. The complete 4D event cloud is then $E = \bigcup_{n=1}^{L} E_n$.

This hierarchical dual-scale framework captures local intra-window dynamics through normalized timestamp z and models global inter-window evolution via window index T_{n}. Unlike fixed event-count methods that fragment or regroup events [31, 66], compromising temporal integrity, our approach dynamically adapts to event sparsity and irregularity by processing variable-length sequences within fixed time windows. We also serialize 4D event clouds into SSM-compatible sequences using z-order and Hilbert curves, aligning with point cloud processing paradigms [67, 68].

3.2. Spatio-Temporal Decoupling and Fusion

To address the complex spatiotemporal noise in rain events while maximizing the temporal discrimination power of event camera, we propose a Spatio-Temporal Decoupling and Fusion Module (STDF). This module explicitly decouples spatial and temporal features while modeling their interactions across dual scales in the 4D event cloud.

We extract spatial features f_{s} via depthwise 1D convolutions on (x, y, z, p) coordinates and temporal features f_{t} through embeddings of window indices T_{n}. To exploit rain events' temporal characteristics and event cameras' high resolution, we prioritize temporal information over spatial information through cross-domain feature modulation. Specifically, spatial features f_{s} are refined through Hadamard products with complementary temporal representations: intra-window dynamics f_t^{intra} from 1D convolutions along z, encoding microsecond correlations within windows, and inter-window trends f_t^{inter} from 1D convolutions on T_{n}, capturing long-term temporal trends across consecutive windows. This modulation is formulated as:

$$f_{\text{s}}^{*} = f_{\text{s}} \otimes \left(f_{\text{t}}^{\text{intra}} + f_t^{\text{inter}} \right). \tag{2}$$

Finally, the temporally-modulated features are combined with the original spatial and temporal representations via residual addition ($\mathcal{R}$). The integrated features are sequentially processed by 1D convolution (Φ_{1d}), batch normalization (BN), and activation (σ) to ensure training stability:

$$\begin{aligned} \mathcal{R}(f_{\text{s}}, f_{\text{t}}) &= f_{\text{s}} + f_{\text{t}} + f_{\text{s}}^{*}, \\ f_{\text{STDF}} &= \sigma\left(\text{BN}\left(\Phi_{1d}\left(\mathcal{R}(f_{\text{s}}, f_{\text{t}})\right)\right)\right). \end{aligned} \tag{3}$$

3.3. Multi-Scale State Space Model

Building upon the hierarchical input representation and spatiotemporal feature embedding, we propose the Multi-Scale State Space Model (MS3M) to further model deeper rain dynamics from multi-temporal and spatial scales. This framework incorporates parallel intra- and inter-window branches to simultaneously capture appearance features and motion patterns, coupled with a multi-scale spatial pathway that preserves structural details and local rain dynamics.

Dual-Temporal-Scale Global Modeling: While Mamba excels in sequence modeling, it struggles to leverage temporal features [69] for event deraining. Inspired by the Motion-aware State Space Model (MSSM), which excels in LiDAR-based moving object segmentation [70], we integrate MSSM into our dual-temporal-scale architecture. This enables explicit modeling of rain events' appearance and motion, improving rain–background separation.

The architecture mainly consists of two complementary branches: an intra-window branch processes spatial appearance features $f_{\text{s}}^{\text{intra}}$ through Reversed Aggregation (RA) and 1D convolution, capturing rain patterns and structural details for rain-background separation; an inter-window branch extracts motion-aware features $f_{\text{m}}^{\text{inter}}$ by analyzing temporal dependencies, enabling separation of dynamic rain streaks from moving objects or scene changes. An adaptive gating mechanism [71] generates weight f_{G} to emphasize rain-relevant features. Outputs from both branches are fused via cross-product attention with inter-branch interaction, then passed to SSM for global feature learning via scanning [72], followed by gating signal modulation.

$$\begin{aligned} f_{\text{s}}^{\text{intra}} &= \sigma(\Phi_{1d}(\text{RA}(f_{\text{I}}))) \\ f_{\text{fuse}} &= \sigma(f_{\text{m}}^{\text{inter}}) \otimes f_{\text{s}}^{\text{intra}} + f_{\text{m}}^{\text{inter}} \\ f_{\text{dual}} &= \text{SSM}(\sigma(f_{\text{fuse}})) \otimes f_{\text{G}}. \end{aligned} \tag{4}$$

Multi-Spatial-Scale Local Pathway: To address rain scale and appearance diversity, we introduce a Multi-Scale Local Spatial Pathway. This pathway extracts multi-scale local features from the intra-window branch using convolutional layers with varying kernel sizes. Small kernel sizes (k_s) capture fine details of thin rain streaks, while larger kernels model broader and more diffuse rain patterns. This design enhances the model's ability to handle diverse rain scenarios while preserving local structural details.

$$\begin{aligned} &f_i = \sigma(\Phi_{1d}^{k_i}(f_{i-1})), \quad i = 1, 2, \ldots, n \\ &f_0 = \Phi_{1d}^{k_0}(\mathbb{L}(\text{RA}(f_{\text{I}}))) \quad f_{\text{ms}} = \sigma(\sum\nolimits_{i=1}^{n} f_i) \end{aligned} \tag{5}$$

The multi-scale features f_{ms} are aggregated and fused with the global outputs f_{dual} via residual addition, seamlessly integrating local details and global context. Finally, a linear layer ($\mathbb{L}$) refines the fused features, formulated as $f_{\text{out}} = \mathbb{L}(f_{\text{ms}} + f_{\text{dual}})$, enhancing spatial precision while maintaining structural integrity.

3.4. Training Loss

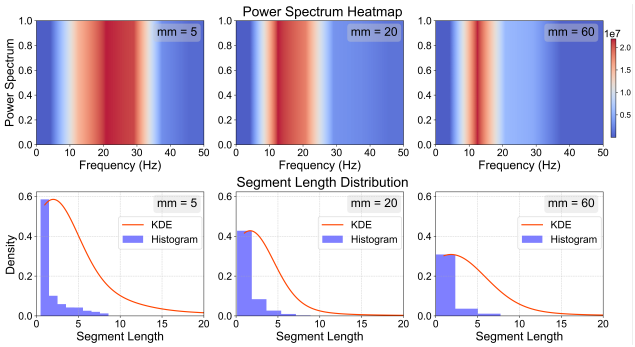

Figure 3. **Power Spectrum Heatmap and Segment Length Distribution for Rainfall Intensities (5, 20, 60 mm).** Top: Heatmaps show higher frequencies at peak power for lower densities, reflecting sparse rain. Bottom: Distributions of continuous signals reveal longer segments at higher densities, indicating sustained rain.

Our training objective combines a standard binary cross-entropy loss ($\mathcal{L}_{\text{ce}}$) with a frequency-domain regularization term ($\mathcal{L}_{\text{fft}}$) to guide the model in learning both point-wise accuracy and global frequency patterns, forming the joint loss as $\mathcal{L} = \mathcal{L}_{\text{ce}} + \lambda \mathcal{L}_{\text{fft}}$.

Frequency-Domain Regularization: Existing image deraining methods integrate frequency layers to capture frequency-domain features [61, 73, 74]. However, for event camera with millions of event data, embedding Fast Fourier Transform (FFT) layer directly introduces significant computational overhead. Alternatively, we propose incorporating $\mathcal{L}_{\text{fft}}$ as a regularization term during model optimization, maintaining efficiency while mitigating overfitting.

While $\mathcal{L}_{\text{ce}}$ focuses on per-event accuracy, it may overlook global distribution patterns, resulting in unrealistic predictions like isolated rain events. In contrast, $\mathcal{L}_{\text{fft}}$ enforces statistical consistency by aligning frequency-domain distributions via amplitude and phase constraints:

$$\mathcal{L}_{\text{fft}} = \frac{1}{L} \sum_{i=1}^{L} \left(\frac{|\mathcal{F}(P)_i - \mathcal{F}(Y)_i|}{\max(|\mathcal{F}(P) - \mathcal{F}(Y)|, \epsilon)} + \epsilon' \right)^2, \quad (6)$$

where $\mathcal{F}(\cdot)$ denotes the 1D FFT, L is the sequence length, and ϵ, ϵ' ensure numerical stability, and Y and P denote the ground truth labels and predicted outputs, respectively. Unlike traditional $\mathcal{L}_{\text{fft}}$ used in image reconstruction [61, 75], our formulation operates directly on event labels, leveraging event cameras' microsecond-level resolution to capture rain dynamics. Rainfall intensity directly correlated to event density and label 1's density (rain events labeled as 1). As illustrated in Figure 3, high-density rainfall manifests as low-frequency continuous patterns in the label sequence, while low-density rain exhibits high-frequency sparse patterns. This spectral distinction enables $\mathcal{L}_{\text{fft}}$ to align predictions with physically consistent global patterns [76].

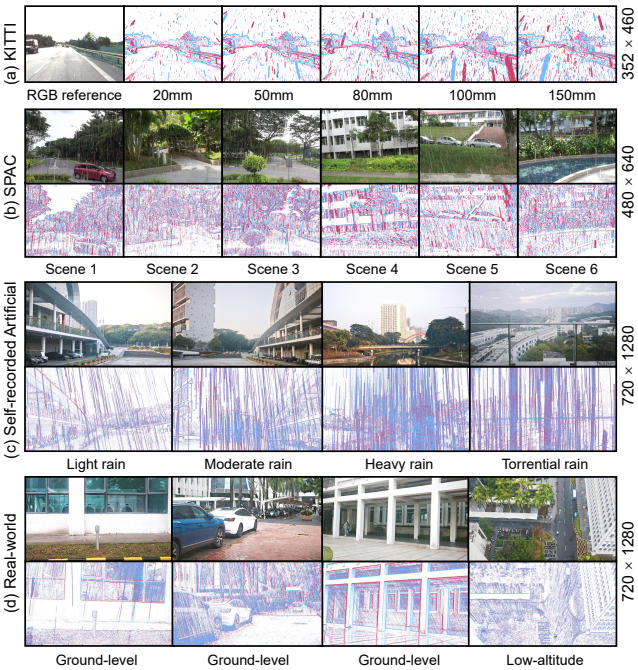

Figure 4. **EventRain-27K: The First Point-based Event Deraining Dataset.** Additional dataset visualizations with varying rainfall intensities and scenarios are included in the Appendix.

4. Dataset

To support model training and testing, we published EventRain-27K, the first point-based benchmark dataset for event-based deraining. The dataset comprises over 7K self-generated synthetic rain samples, over 7K self-recorded artificial rain samples, and over 9K real rain samples.

Synthetic Dataset (labeled). Supervised event deraining relies on labeled data; however, annotating discrete, sparse, and highly dynamic event points is non-trivial. To address this limitation, we use a two-step synthetic data generation approach. We first utilize the KITTI dataset [77] and the SPAC dataset [78] as clean video sources, then synthesize rain effects using state-of-the-art rendering simulators [79, 80]. The KITTI rainy dataset provides videos with varying camera motion speeds and rain intensities (4(a)), covering rainfall rates ranging from light rain (5 mm/hour) to severe storms (200 mm/hour). The SPAC rainy dataset includes 16 diverse scenes, such as cityscapes and natural environments, each augmented with 3 to 4 distinct rain patterns (4(b)). Finally, we generate event sequences from these rainy videos using the Vid2E simulator [81].

Self-recorded Artificial Dataset (labeled). To bridge the domain gap between synthetic and real-world event data for model training, we introduce a self-recorded artificial dataset, capturing real event camera data under controlled artificial rainfall. A showerhead was employed to simulate raindrop free-fall motion, captured by a Prophesee EVK4 event camera (1280 × 720 resolution, 10k fps, 120dB dy-

Metrics	5mm			20mm			50mm			80mm			125mm			150mm		
	SR ↑	NR ↑	DA ↑	SR ↑	NR ↑	DA ↑	SR ↑	NR ↑	DA ↑	SR ↑	NR ↑	DA ↑	SR ↑	NR ↑	DA ↑	SR ↑	NR ↑	DA ↑
TS[82]	0.888	0.265	0.576	0.887	0.305	0.596	0.883	0.231	0.557	0.881	0.271	0.576	0.877	0.237	0.556	0.872	0.243	0.557
DWF[83]	0.703	0.394	0.549	0.734	0.464	0.599	0.755	0.352	0.553	0.777	0.432	0.604	0.786	0.371	0.578	0.782	0.375	0.578
Knoise[84]	0.860	0.375	0.618	0.870	0.377	0.624	0.884	0.241	0.563	0.890	0.257	0.574	0.901	0.211	0.556	0.896	0.214	0.555
Ynoise[85]	0.679	0.533	0.606	0.677	0.568	0.623	0.663	0.481	0.572	0.658	0.525	0.591	0.647	0.482	0.564	0.634	0.487	0.561
RED[86]	0.862	0.172	0.517	0.853	0.251	0.552	0.833	0.208	0.520	0.822	0.185	0.503	0.808	0.207	0.507	0.789	0.208	0.499
EDnCNN[86]	0.968	0.905	0.937	0.954	0.904	0.929	0.948	0.888	0.918	0.935	0.870	0.903	**0.933**	0.867	0.900	**0.929**	0.843	0.886
AEDNet[87]	0.941	0.850	0.891	0.938	0.876	0.907	0.928	0.732	0.830	0.925	0.681	0.803	0.922	0.624	0.773	0.923	0.547	0.735
EDformer[66]	0.981	0.818	0.899	0.962	0.832	0.897	0.924	0.844	0.884	0.894	0.834	0.864	0.876	0.840	0.858	0.839	0.834	0.836
Ours	**0.994**	**0.914**	**0.954**	**0.978**	**0.915**	**0.947**	**0.955**	**0.911**	**0.933**	**0.940**	**0.903**	**0.922**	0.918	**0.898**	**0.908**	0.908	**0.895**	**0.902**

Table 1. **Quantitative Comparisons on Synthetic Dataset with Varying Rainfall Intensities.** We mark the **best** and second best results.

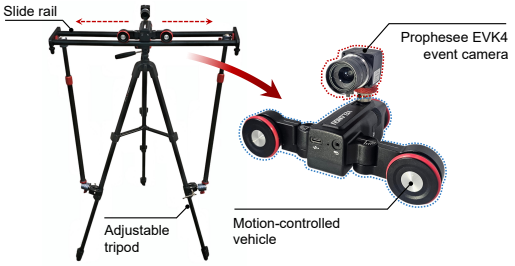

Figure 5. Our fixed-point & speed data acquisition system.

namic range) for both rainy and rain-free data. Slight camera vibrations were applied during captures to obtain complete scene details. Utilizing the K-Nearest Neighbors (KNN) algorithm, background events in rainy data were identified via spatiotemporal alignment with rain-free data, enabling background labeling and isolating rain events. The dataset focuses on urban outdoor scenes (buildings, trees, cars, lakes) and covers diverse rainfall intensities across four scenarios (Figure 4(c)).

Real-world Dataset (label-free). To validate event deraining effectiveness and generalization, we collect a real-world event dataset. As illustrated in Figure 5, we designed a data acquisition system with fixed-point and fixed-speed (10 cm/s) cruising. We mounted a Prophesee EVK4 event camera on a motion-controlled vehicle, which traversed a slide rail to capture background and rain data simultaneously under controlled dynamic conditions. The dataset includes diverse viewpoints and rain intensities, covering a wide range of real-world scenarios (Figure 4(d)). Additionally, it incorporates 4K samples of real snow data captured to evaluate the model's generalization capability.

5. Experiments

5.1. Experiments Setups

Evaluation Metrics. We use event Denoising Accuracy (DA), Signal Retention (SR) and Noise Removal (NR) to measure the event rain removal performance on PRE-Mamba. There is $\mathrm{DA} = \frac{1}{2}(\mathrm{SR} + \mathrm{NR}) = (\frac{\mathrm{PB}}{\mathrm{TB}} + \frac{\mathrm{PR}}{\mathrm{TR}})$, where $\mathrm{PB}, \mathrm{TB}, \mathrm{PR}, \mathrm{TR}$ are the predicted background, the predicted rain, the real background, and the number of real rain, respectively. DA measures rain removal while preserving real motion (e.g., vehicles, pedestrians) and structural details (e.g., roads, building profiles). SR and NR analyze signal retention and noise suppression, respectively.

Implementation Details. Our network is implemented in PyTorch and trained using the AdamW optimizer with an initial learning rate of 4.8×10^{-4} and a weight decay coefficient of 5×10^{-3}. The entire training is conducted on six NVIDIA RTX A6000 GPUs for 50 epochs with a batch size of 6. The model processes a sequence of five temporal windows as input, with each window spanning 0.1 seconds. In the multi-spatial-scale module of the MS3M, we employ kernel sizes of (1, 3, 5) to capture features at different spatial resolutions. Network details can be found in the Appendix.

5.2. Comparisons with State-of-The-Art Methods

Baselines. As this work represents the first attempt at point-based event deraining, there are no directly comparable baselines in the literature. Alternatively, we adapt state-of-the-art event denoising methods, as deraining can be viewed as a specialized case of noise removal. These include filter-based methods TS [82], DWF [83], Knoise [84], Ynoise [85], RED [86] and learning-based algorithms EDnCNN [86], AEDNet [87], EDformer [66]. Additionally, we qualitatively compare with DistillNet [11], a voxel-based GAN deraining method, but its voxelized processing is incompatible with event point denoising metrics, preventing quantitative evaluation.

Results on Synthetic and Artificial Datasets. In Table 1, we present SR, NR, and DA results on the synthetic dataset with six rain intensities, ranging from light rain to heavy rain. Our method achieves state-of-the-art performance, with an average SR/NR/DA of 0.95/0.91/0.93. It is notable that learning-based algorithms outperform filtering methods due to their ability to leverage supervised training. When training EDnCNN [86], we replaced its original event probability mask (EPM) labeling framework with our ground truth labels, which improved its performance. In addition, we provide qualitative results in Figure 6. Filtering methods are ineffective for rain removal, as rain streaks exhibit distinct spatiotemporal patterns that deviate from typical noise. Other learning-based methods yield incomplete deraining results, often retaining residual rain streaks. In contrast, our method effectively removes rain streaks and restores details through multi-scale temporal and spatial modeling

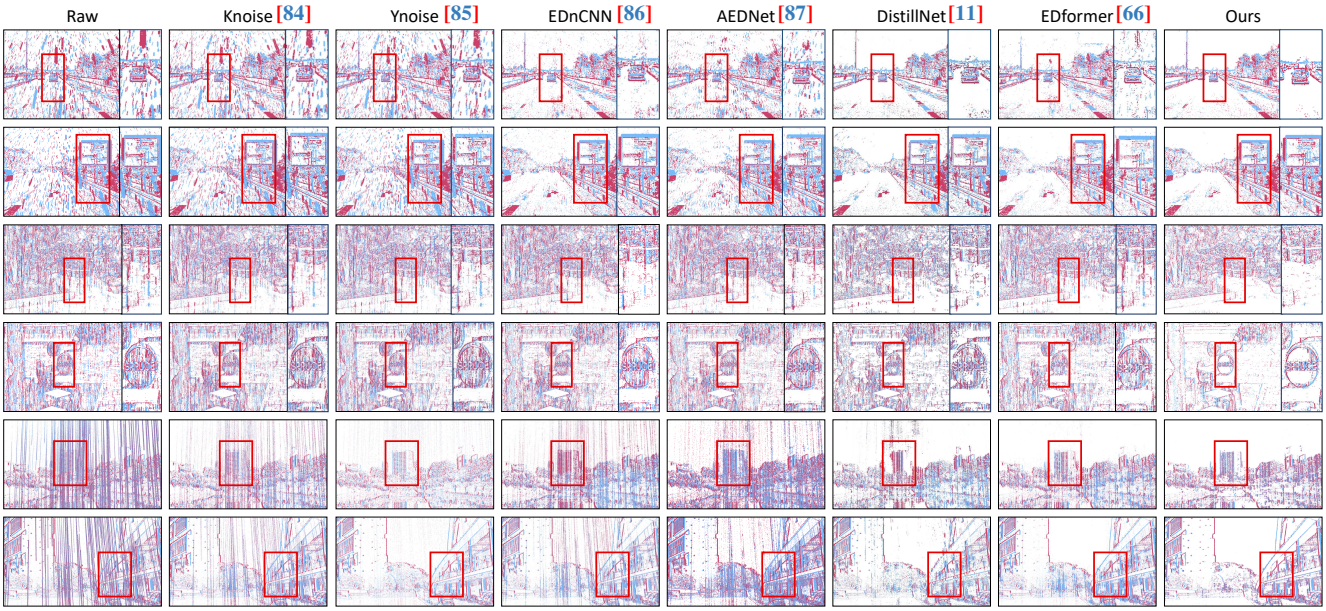

Figure 6. **Qualitative Comparison on Synthetic and Artificial Datasets.** PRE-Mamba demonstrates superior deraining capabilities across diverse structural contours and complex textures with enhanced detail preservation. Please zoom in for details.

	GFLOPs	Parameters	Inference time (s)	Relative speed
TS[82]	N/A	N/A	0.1296	1.0 ×
DWF[83]	N/A	N/A	0.0954	1.36×
KNoise[84]	N/A	N/A	**0.0198**	**6.55**×
YNoise[85]	N/A	N/A	0.0513	2.53×
RED[86]	N/A	N/A	2.2716	0.06×
EDnCNN[86]	234.51	614.55K	20.1885	1.0×
AEDNet[87]	4400.46	45.87M	43.4250	0.46×
DistillNet[11]	255.17	18.96M	0.2029	99.50×
EDformer[66]	8.41	**49.80K**	2.4943	8.09×
Ours	**6.23**	264.63K	**0.0987**	**204.54**×

Table 2. Model complexity comparisons with previous model.

combined with motion and appearance perception.

Generalization on Real-World Dataset. To validate practical applicability, we compare our method against baselines on the real-world data of EventRain-27K. Figure 7 (top four rows) isualizes results for four scenarios, including three ground-level and one low-altitude viewpoints. Our method achieves superior performance compared to competitors. In the low-altitude scenario, our method effectively eliminates rain streaks while maintaining the structural integrity of static objects (e.g., buildings) and preserving the motion patterns of dynamic objects (e.g., vehicles). Filter-based methods treat both background and rain equally, resulting in effects similar to downsampling. While EDnCNN [86], AEDNet [87], and EDformer [66] are trained on synthetic data, their denoising-focused designs fail to generalize well to real rain, further hindered by the synthetic-to-real domain gap. DistillNet [11] removes large rain streaks via voxel-based GANs but loses background details and introduces artifacts. In contrast, our method, designed for point-based event deraining, effectively removes rain while preserving both background details and dynamic motion content, demonstrating robust real-world performance.

Model Complexity and Efficiency Comparison. In Table 2, we compare the computational efficiency of our method and baselines in terms of FLOPs, parameters, and inference time for processing 100K events on an NVIDIA RTX A6000 GPU. Filtering-based methods enable fast inference but suffer from poor deraining performance due to lack of task-specific optimization. EDnCNN [86] and AEDNet [87] incur high computational overhead from encoding or sampling spatial-temporal neighborhoods, resulting in high FLOPs and slow inference. Transformer-based EDformer [66] reduces FLOPs and parameters but suffers from high latency due to the quadratic complexity of its attention mechanism. DistillNet [11] achieves a lower inference time by voxelizing events but at the cost of high computational complexity, large model size, and loss of fine-grained temporal precision. In contrast, our method achieves SOTA performance with only 6.23G FLOPs and 0.26M parameters, processing 100K events in 0.0987s and scaling to 1M events in 0.398s. Compared to EDnCNN and AEDNet, our method reduces FLOPs to 2.66% and 0.14% of theirs, respectively, while achieving 204.5× and 440.0× faster inference. Compared to the lightweight EDformer, our method achieves a 25.3× speedup with comparable model size, excelling in both accuracy and computational efficiency.

Model	STDF	MS3M	$\mathcal{L}_{\text{fft}}$	SR↑	NR↑	DA↑
M1				0.9123	0.8394	0.8759
M2	✓			0.8983	0.8747	0.8865
M3		✓		0.8941	0.8906	0.8923
M4	✓	✓		0.9046	0.8954	0.9000
Ours	✓	✓	✓	0.9080	0.8949	0.9015

Table 3. Ablation about each module in PRE-Mamba on the EventRain-27K validation set.

5.3. Ablation Studies

Baseline Design. We conduct systematic ablation studies to validate the efficacy of three proposed components:

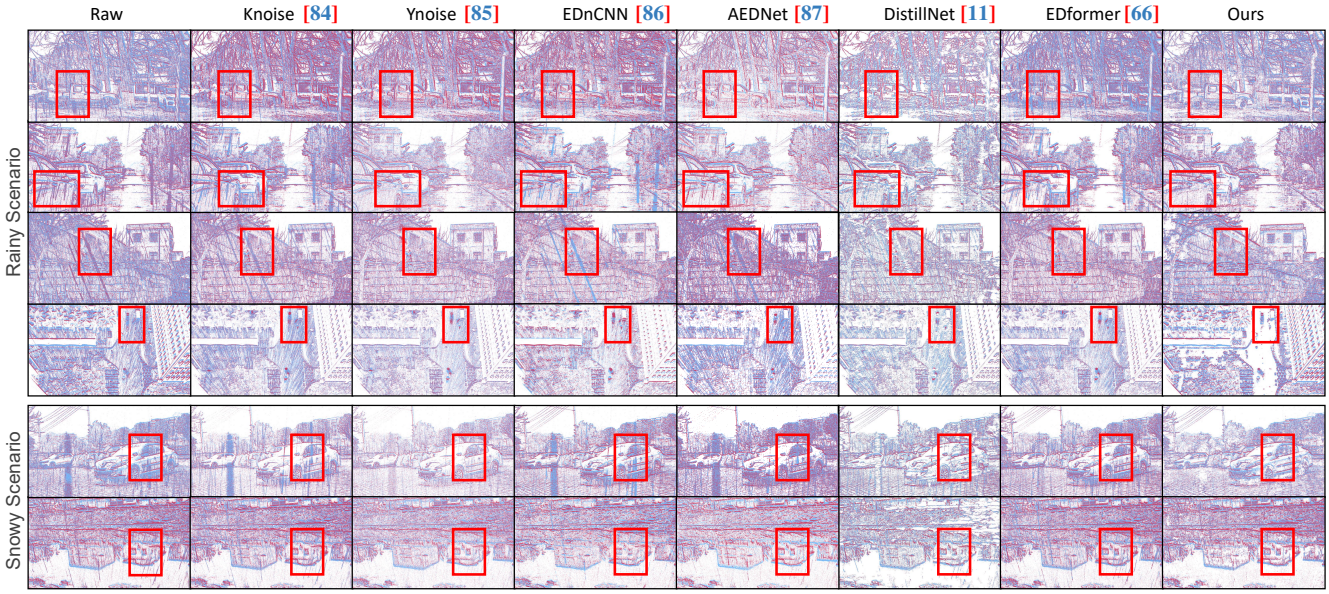

Figure 7. **Qualitative Comparison on Real-World Datasets.** The top four rows highlight our method's superior deraining performance on real rain datasets, while the last two rows showcase its generalization to snow scenes. Please zoom in for details.

Window Number	Num = 3	Num = 5	Num = 8
DA	0.8268	0.9015	0.9330
Improve (%)	-	9.03	12.84
Inference time (ms)	274.02	394.16	525.68
Relative speed	-	0.70 ×	0.52 ×

Table 4. Ablation about time window number on validation set.

the Spatio-Temporal Decoupling and Fusion (STDF) module, Multi-Scale State Space Model (MS3M), and Frequency Loss ($\mathcal{L}_{\text{fft}}$). Our baseline model ("M1") removes all three components, employing naive summation for spatio-temporal feature fusion, reverting to the original MSSM block [70] for temporal modeling, and using only cross-entropy loss. Building upon "M1", "M2" introduces the STDF module to embed structured spatio-temporal features, while "M3" integrates the MS3M architecture for enhanced multi-scale modeling. And "M4" combines both STDF and MS3M components. Finally, we incorporate the Frequency Loss into "M4" to complete our full model.

Quantitative Comparison. From Table 3, we observe that both STDF ("M2") and MS3M ("M3") improve the baseline performance by 1.21% and 1.87% in DA, respectively, validating their individual contributions. "M3" achieves greater improvement, demonstrating MS3M's ability to capture deep spatio-temporal features and adapt to diverse rain patterns. By integrating STDF with MS3M, "M4" further enhances performance, demonstrating the complementary nature of these components in modeling multi-scale rain and event characteristics. Finally, our full network outperforms "M4", which verifies the importance of the frequency-domain regularization term in leveraging both amplitude and phase constraints for improved learning.

Effects of Time Windows Number. In Table 4, we evaluate model performance with time windows of 3, 5, and 8 on validation set. Increasing the number of time windows improves the model's ability to capture temporal dependencies, with a 9.0% improvement in DA from 3 to 5 windows and an additional 3.8% improvement from 5 to 8. However, the inference speed decreases by 30% when the window size is increased from 3 to 5, and by 48% when increased from 3 to 8. Therefore, we set the number of time windows to 5, as further increases in time windows yield diminishing returns, emphasizing the need to balance temporal depth and computational efficiency.

5.4. Generalization to Snowy Datasets

While primarily designed for event-based deraining, our method also generalizes well to snow removal, as demonstrated in Figure 7 (last two rows). Without architectural changes, it delivers promising results under snowy conditions, showcasing robustness to diverse weather patterns. Additional visual results are included in the Appendix.

6. Conclusion

To the best of our knowledge, PRE-Mamba is the first point-based framework for event camera deraining. It adopts a 4D event cloud to preserve native temporal resolution via inter- and intra-temporal windows, a spatiotemporal decoupling and fusion module for efficient feature extraction, and a multi-scale state space model that captures rain dynamics across spatial and temporal scales with linear complexity. A frequency regularization term and cross-entropy loss guide learning, and we contribute a large-scale dataset with diverse rain scenarios. Extensive experiments demonstrate SOTA performance on synthetic and real-world data. The method is efficient, lightweight, and real-time capable, with potential for robust multi-agent perception [88, 89].

Acknowledgments This paper was supported by the Natural Science Foundation of China under Grant 62371269, Guangdong Innovative and Entrepreneurial Research Team Program (2021ZT09L197), and Meituan Academy of Robotics Shenzhen.

References

[1] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(1):154–180, 2020. 1

[2] Haoyang Wang, Ruishan Guo, Pengtao Ma, Ciyu Ruan, Xinyu Luo, Wenhua Ding, Tianyang Zhong, Jingao Xu, Yunhao Liu, and Xinlei Chen. Towards mobile sensing with event cameras on high-agility resource-constrained devices: A survey. *arXiv preprint arXiv:2503.22943*, 2025. 1

[3] Arren Glover and Chiara Bartolozzi. Robust visual tracking with a freely-moving event camera. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3769–3776. IEEE, 2017. 1

[4] Xinyu Luo, Haoyang Wang, Ciyu Ruan, Chenxin Liang, Jingao Xu, and Xinlei Chen. Eventtracker: 3d localization and tracking of high-speed object with event and depth fusion. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 1974–1979, 2024. 1

[5] Jianhao Jiao, Huaiyang Huang, Liang Li, Zhijian He, Yilong Zhu, and Ming Liu. Comparing representations in tracking for event camera-based slam. In *Proceedings of the IEEE/cvf conference on computer vision and pattern recognition*, pages 1369–1376, 2021. 1

[6] Antoni Rosinol Vidal, Henri Rebecq, Timo Horstschaefer, and Davide Scaramuzza. Ultimate slam? combining events, images, and imu for robust visual slam in hdr and high-speed scenarios. *IEEE Robotics and Automation Letters*, 3(2):994–1001, 2018.

[7] Daniel Gehrig and Davide Scaramuzza. Low-latency automotive vision with event cameras. *Nature*, 629(8014):1034–1040, 2024. 1

[8] Haoyang Wang, Jingao Xu, Xinyu Luo, Xuecheng Chen, Ting Zhang, Ruiyang Duan, Yunhao Liu, and Xinlei Chen. Ultra-high-frequency harmony: mmwave radar and event camera orchestrate accurate drone landing. In *Proceedings of the 23rd ACM Conference on Embedded Networked Sensor Systems*, pages 15–29, 2025. 1

[9] Davide Falanga, Kevin Kleber, and Davide Scaramuzza. Dynamic obstacle avoidance for quadrotors with event cameras. *Science Robotics*, 5(40):eaaz9712, 2020. 1

[10] Nitin J Sanket, Chethan M Parameshwara, Chahat Deep Singh, Ashwin V Kuruttukulam, Cornelia Fermüller, Davide Scaramuzza, and Yiannis Aloimonos. Evdodgenet: Deep dynamic obstacle dodging with event cameras. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10651–10657. IEEE, 2020. 1

[11] Ciyu Ruan, Chenyu Zhao, Chenxin Liang, Xinyu Luo, Jingao Xu, and Xinlei Chen. Distill drops into data: Event-based rain-background decomposition network. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 2072–2077, 2024. 1, 2, 6, 7, 8

[12] Long Cheng, Ni Liu, Xusen Guo, Yuhao Shen, Zijun Meng, Kai Huang, and Xiaoqin Zhang. A novel rain removal approach for outdoor dynamic vision sensor event videos. *Frontiers in Neurorobotics*, 16:928707, 2022. 1, 2

[13] Xinlei Chen, Aveek Purohit, Carlos Ruiz Dominguez, Stefano Carpin, and Pei Zhang. Drunkwalk: Collaborative and adaptive planning for navigation of micro-aerial sensor swarms. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pages 295–308, 2015. 1

[14] Shuhang Zhang, Hongliang Zhang, Boya Di, and Lingyang Song. Cellular uav-to-x communications: Design and optimization for multi-uav networks. *IEEE Transactions on Wireless Communications*, 18(2):1346–1359, 2019.

[15] Haoyang Wang, Xuecheng Chen, Yuhan Cheng, Chenye Wu, Fan Dang, and Xinlei Chen. H-swarmloc: Efficient scheduling for localization of heterogeneous mav swarm with deep reinforcement learning. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 1148–1154, 2022. 1

[16] Zhuozhu Jian, Zejia Liu, Haoyu Shao, Xueqian Wang, Xinlei Chen, and Bin Liang. Path generation for wheeled robots autonomous navigation on vegetated terrain. *IEEE Robotics and Automation Letters*, 9(2):1764–1771, 2023. 1

[17] Kaddour Messaoudi, Abdullah Baz, Omar Sami Oubbati, Abderrezak Rachedi, Tahar Bendouma, and Mohammed Atiquzzaman. Ugv charging stations for uav-assisted aoi-aware data collection. *IEEE Transactions on Cognitive Communications and Networking*, 10(6):2325–2343, 2024.

[18] jianzhuozhu, Xuran Pu, Jianjie Fang, Zhiyuan Deng, Xueqian Wang, and Xinlei Chen. A large language model-driven heterogeneous air-ground rescue swarm. In *ICLR 2025 Workshop on Embodied Intelligence with Large Language Models In Open City Environment*, 2025. 1

[19] Lei Yu, Bishan Wang, Jingwei He, Gui-Song Xia, and Wen Yang. Single image deraining with continuous rain density estimation. *IEEE Transactions on Multimedia*, 25:443–456, 2021. 2

[20] Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan. Learning a sparse transformer network for effective image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5896–5905, 2023. 2

[21] Sotiris Karavarsamis, Ioanna Gkika, Vasileios Gkitsas, Konstantinos Konstantoudakis, and Dimitrios Zarpalas. A survey of deep learning-based image restoration methods for enhancing situational awareness at disaster sites: the cases of rain, snow and haze. *Sensors*, 22(13):4707, 2022.

[22] Jiehui Huang, Zhenchao Tang, Xuedong He, Jun Zhou, Defeng Zhou, and Calvin Yu-Chian Chen. Progressive network based on detail scaling and texture extraction: A more general framework for image deraining. *Neurocomputing*, 568: 127066, 2024. 2

[23] Qinyi Wang, Yexin Zhang, Junsong Yuan, and Yilong Lu. Space-time event clouds for gesture recognition: From rgb

cameras to event cameras. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1826–1835. IEEE, 2019. 2

[24] Yusuke Sekikawa, Kosuke Hara, and Hideo Saito. Eventnet: Asynchronous recursive event processing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3887–3896, 2019. 2

[25] Yin Bi, Aaron Chadha, Alhabib Abbas, Eirina Bourtsoulatze, and Yiannis Andreopoulos. Graph-based spatio-temporal feature learning for neuromorphic vision sensing. *IEEE Transactions on Image Processing*, 29:9084–9098, 2020. 2

[26] Simon Schaefer, Daniel Gehrig, and Davide Scaramuzza. Aegnn: Asynchronous event-based graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12371–12381, 2022. 2

[27] Alexander Kugele, Thomas Pfeil, Michael Pfeiffer, and Elisabetta Chicca. Hybrid snn-ann: Energy-efficient classification and object detection for event-based vision. In *DAGM German Conference on Pattern Recognition*, pages 297–312. Springer, 2021. 2

[28] Yannick Schnider, Stanisław Woźniak, Mathias Gehrig, Jules Lecomte, Axel Von Arnim, Luca Benini, Davide Scaramuzza, and Angeliki Pantazi. Neuromorphic optical flow and real-time implementation with event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4129–4138, 2023.

[29] Wei Fang, Zhaofei Yu, Yanqi Chen, Timothée Masquelier, Tiejun Huang, and Yonghong Tian. Incorporating learnable membrane time constant to enhance learning of spiking neural networks. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2641–2651, 2021. 2, 3

[30] Alejandro Hernández Díaz, Rebecca Davidson, Steve Eckersley, Christopher Paul Bridges, and Simon J Hadfield. E-mamba: Using state-space-models for direct event processing in space situational awareness. In *IAA Conference on AI in and for Space (SPAICE)*, 2024. 2, 3

[31] Hongwei Ren, Yue Zhou, Jiadong Zhu, Haotian Fu, Yulong Huang, Xiaopeng Lin, Yuetong Fang, Fei Ma, Hao Yu, and Bojun Cheng. Rethinking efficient and effective point-based networks for event camera classification and regression: Eventmamba. *arXiv preprint arXiv:2405.06116*, 2024. 2, 4

[32] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023. 2, 3

[33] Hongzhong Tang, Ling Zhu, Dongbo Zhang, and Xiang Wang. Single image rain removal model using pure rain dictionary learning. *IET Image Processing*, 13(10):1797–1804, 2019. 2

[34] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3100–3109, 2020.

[35] Jin-Hwan Kim, Jae-Young Sim, and Chang-Su Kim. Video deraining and desnowing using temporal correlation and low-rank matrix completion. *IEEE Transactions on Image Processing*, 24(9):2658–2670, 2015.

[36] Yi Chang, Luxin Yan, and Sheng Zhong. Transformed low-rank model for line pattern noise removal. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. 2

[37] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. *arXiv preprint arXiv:2108.10257*, 2021. 2

[38] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17683–17693, 2022. 2

[39] 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 (CVPR)*, pages 14821–14831, 2021. 2

[40] Yueyi Zhang, Jin Wang, Wenming Weng, Xiaoyan Sun, and Zhiwei Xiong. Egvd: Event-guided video deraining. *ArXiv*, abs/2309.17239, 2023. 2

[41] Shangquan Sun, Wenqi Ren, Jingzhi Li, Kaihao Zhang, Meiyu Liang, and Xiaochun Cao. Event-aware video deraining via multi-patch progressive learning. *IEEE Transactions on Image Processing*, 32:3040–3053, 2023. 2

[42] Chengjie Ge, Xueyang Fu, Peng He, Kunyu Wang, Chengzhi Cao, and Zheng-Jun Zha. Neuromorphic event signal-driven network for video de-raining. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1878–1886, 2024. 2

[43] Peiqi Duan, Zihao W Wang, Boxin Shi, Oliver Cossairt, Tiejun Huang, and Aggelos K Katsaggelos. Guided event filtering: Synergy between intensity images and neuromorphic events for high performance imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11): 8261–8275, 2021. 2

[44] Yang Feng, Hengyi Lv, Hailong Liu, Yisa Zhang, Yuyao Xiao, and Chengshan Han. Event density based denoising method for dynamic vision sensor. *Applied Sciences*, 10(6), 2020.

[45] Yaoyi Chen, Yujie Huang, Feiqiang Li, Xiaoyang Zeng, Wenhong Li, and Mingyu Wang. Denoising method for dynamic vision sensor based on two-dimensional event density. In *2023 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–4, 2023. 2

[46] Pei Zhang, Zhou Ge, Li Song, and Edmund Y. Lam. Neuromorphic imaging with density-based spatiotemporal denoising. *IEEE Transactions on Computational Imaging*, 9:530–541, 2023. 2

[47] 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. 2

[48] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. Depth-attentional features for single-image rain removal. In

Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, pages 8022–8031, 2019.

[49] Wenhan Yang, Robby T Tan, Shiqi Wang, Yuming Fang, and Jiaying Liu. Single image deraining: From model-based to data-driven and beyond. *IEEE Transactions on pattern analysis and machine intelligence*, 43(11):4059–4077, 2020.

[50] Hong Wang, Zongsheng Yue, Qi Xie, Qian Zhao, Yefeng Zheng, and Deyu Meng. From rain generation to rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14791–14801, 2021. 2

[51] Jin Wang, Wenming Weng, Yueyi Zhang, and Zhiwei Xiong. Unsupervised video deraining with an event camera. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10797–10806, 2023. 2

[52] Alexander Andreopoulos, Hirak J. Kashyap, Tapan K. Nayak, Arnon Amir, and Myron D. Flickner. A low power, high throughput, fully event-based stereo system. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7532–7542, 2018. 2

[53] Daniel Gehrig, Antonio Loquercio, Konstantinos G Derpanis, and Davide Scaramuzza. End-to-end learning of representations for asynchronous event-based data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5633–5643, 2019. 3

[54] Fernando Cladera, Anthony Bisulco, Daniel Kepple, Volkan Isler, and Daniel D. Lee. On-device event filtering with binary neural networks for pedestrian detection using neuromorphic vision sensors. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 3084–3088, 2020. 2

[55] Nikola Zubic, Mathias Gehrig, and Davide Scaramuzza. State space models for event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5819–5828, 2024. 3

[56] Simone Undri Innocenti, Federico Becattini, Federico Pernici, and Alberto Del Bimbo. Temporal binary representation for event-based action recognition. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 10426–10432, 2021. 3

[57] Dayong Ren, Zhe Ma, Yuanpei Chen, Weihang Peng, Xiaode Liu, Yuhan Zhang, and Yufei Guo. Spiking pointnet: Spiking neural networks for point clouds. In *Advances in Neural Information Processing Systems*, pages 41797–41808. Curran Associates, Inc., 2023. 3

[58] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 3

[59] Daniel Gehrig and Davide Scaramuzza. Low latency automotive vision with event cameras. 2024. 3

[60] Jiangrong Shen, Qi Xu, Jian K Liu, Yueming Wang, Gang Pan, and Huajin Tang. Esl-snns: An evolutionary structure learning strategy for spiking neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 86–93, 2023. 3

[61] Zou Zhen, Yu Hu, and Zhao Feng. Freqmamba: Viewing mamba from a frequency perspective for image deraining. *arXiv preprint arXiv:2404.09476*, 2024. 3, 5

[62] 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. In *European conference on computer vision*, pages 222–241. Springer, 2024.

[63] Lianghui 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. 3

[64] Shangquan Sun, Wenqi Ren, Juxiang Zhou, Jianhou Gan, Rui Wang, and Xiaochun Cao. A hybrid transformer-mamba network for single image deraining. *arXiv preprint arXiv:2409.00410*, 2024. 3

[65] Yufeng Li, Chuanlong Xie, and Hongming Chen. Multi-scale representation for image deraining with state space model. *Signal, Image and Video Processing*, 19(1):183, 2025. 3

[66] Bin Jiang, Bo Xiong, Bohan Qu, M Salman Asif, You Zhou, and Zhan Ma. Edformer: Transformer-based event denoising across varied noise levels. In *European Conference on Computer Vision*, pages 200–216. Springer, 2024. 4, 6, 7, 8

[67] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 16259–16268, 2021. 4

[68] Xiaoyang Wu, Yixing Lao, Li Jiang, Xihui Liu, and Hengshuang Zhao. Point transformer v2: Grouped vector attention and partition-based pooling. *Advances in Neural Information Processing Systems*, 35:33330–33342, 2022. 4

[69] Jiahao Qin and Feng Liu. Mamba-spike: Enhancing the mamba architecture with a spiking front-end for efficient temporal data processing. *arXiv preprint arXiv:2408.11823*, 2024. 4

[70] Kang Zeng, Hao Shi, Jiacheng Lin, Siyu Li, Jintao Cheng, Kaiwei Wang, Zhiyong Li, and Kailun Yang. Mambamos: Lidar-based 3d moving object segmentation with motion-aware state space model. In *ACM International Conference on Multimedia (MM)*, 2024. 4, 8

[71] Weizhe Hua, Zihang Dai, Hanxiao Liu, and Quoc Le. Transformer quality in linear time. In *International conference on machine learning*, pages 9099–9117. PMLR, 2022. 4

[72] Haotian Zhang, Keyan Chen, Chenyang Liu, Hao Chen, Zhengxia Zou, and Zhenwei Shi. Cdmamba: Remote sensing image change detection with mamba. *arXiv preprint arXiv:2406.04207*, 2024. 4

[73] Lin Liu, Jianzhuang Liu, Shanxin Yuan, Gregory Slabaugh, Aleš Leonardis, Wengang Zhou, and Qi Tian. Wavelet-based dual-branch network for image demoiréing. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*, pages 86–102. Springer, 2020. 5

[74] Dong Li, Yidi Liu, Xueyang Fu, Senyan Xu, and Zheng-Jun Zha. Fouriermamba: Fourier learning integration with state space models for image deraining. *arXiv preprint arXiv:2405.19450*, 2024. 5

[75] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Focal frequency loss for image reconstruction and synthesis. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 13919–13929, 2021. 5

[76] Ji Luo, Zijian Xiao, Zuxin Li, Xuecheng Chen, Chaopeng Hong, Xiao-Ping Zhang, and Xinlei Chen. Smartspr: A physics-informed mobile sprinkler scheduling system for reducing urban particulate matter pollution. *IEEE Transactions on Mobile Computing*, 2025. 5

[77] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361. IEEE, 2012. 5

[78] 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. 5

[79] Maxime Tremblay, Shirsendu Sukanta Halder, Raoul De Charette, and Jean-François Lalonde. Rain rendering for evaluating and improving robustness to bad weather. *International Journal of Computer Vision*, 129:341–360, 2021. 5

[80] Adobe Inc. Adobe after effects software, 2023. 5

[81] Daniel Gehrig, Mathias Gehrig, Javier Hidalgo-Carrió, and Davide Scaramuzza. Video to events: Recycling video datasets for event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3586–3595, 2020. 5

[82] Teresa Serrano-Gotarredona and Bernabé Linares-Barranco. A 128 $\times$ 128 1.5*IEEE Journal of Solid-State Circuits*, 48(3): 827–838, 2013. 6, 7

[83] Shasha Guo and Tobi Delbruck. Low cost and latency event camera background activity denoising. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):785–795, 2023. 6, 7

[84] Alireza Khodamoradi and Ryan Kastner. $o(n)$o(n)-space spatiotemporal filter for reducing noise in neuromorphic vision sensors. *IEEE Transactions on Emerging Topics in Computing*, 9(1):15–23, 2021. 6, 7, 8

[85] Yang Feng, Hengyi Lv, Hailong Liu, Yisa Zhang, Yuyao Xiao, and Chengshan Han. Event density based denoising method for dynamic vision sensor. *Applied Sciences*, 2020. 6, 7, 8

[86] R Baldwin, Mohammed Almatrafi, Vijayan Asari, and Keigo Hirakawa. Event probability mask (epm) and event denoising convolutional neural network (edncnn) for neuromorphic cameras. *arXiv preprint arXiv:2003.08282*, 2020. 6, 7, 8

[87] Huachen Fang, Jinjian Wu, Leida Li, Junhui Hou, Weisheng Dong, and Guangming Shi. Aednet: Asynchronous event denoising with spatial-temporal correlation among irregular data. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1427–1435, 2022. 6, 7, 8

[88] Xuecheng Chen, Haoyang Wang, Yuhan Cheng, Haohao Fu, Yuxuan Liu, Fan Dang, Yunhao Liu, Jinqiang Cui, and Xinlei Chen. Ddl: Empowering delivery drones with large-scale urban sensing capability. *IEEE Journal of Selected Topics in Signal Processing*, 2024. 8

[89] Xuecheng Chen, Zijian Xiao, Yuhan Cheng, Chen-Chun Hsia, Haoyang Wang, Jingao Xu, Susu Xu, Fan Dang, Xiao-Ping Zhang, Yunhao Liu, et al. Soscheduler: Toward proactive and adaptive wildfire suppression via multi-uav collaborative scheduling. *IEEE Internet of Things Journal*, 11(14): 24858–24871, 2024. 8