SpikeRain: Towards Energy-Efficient Single Image Deraining with Spiking Neural Networks

Md Tanvir Islam[1] Inzamamul Alam[2] Sambit Bakshi[3] Khan Muhammad[2,*]
Javier Del Ser[4] Sangtae Ahn[1,*]
[1]Kyungpook National University, South Korea [2]Sungkyunkwan University, South Korea
[3]National Institute of Technology Rourkela, India [4]University of the Basque Country, Spain
tanvirnwu@knu.ac.kr inzi15@g.skku.edu bakshisambit@ieee.org
khan.muhammad@ieee.org javier.delser@ehu.eus stahn@knu.ac.kr

Abstract

With the rapid deployment of vision systems on edge devices, energy-efficient and temporally aware image deraining models are increasingly needed. We propose SpikeRain, a spiking neural network (SNN) that achieves competitive deraining performance with substantially lower computational cost than conventional artificial neural networks (ANNs). Unlike ANN-based approaches with dense activations and high memory demands, SpikeRain leverages the event-driven sparse-firing nature of spiking neurons for efficient temporal integration and contextual learning. Built on an encoder-decoder framework, SpikeRain incorporates three spiking native modules: a Dense Spiking Residual Block (DSRB) for temporal integration and feature reuse, a Multi-Dimensional Spiking Attention (MDSA) module to model temporal channel spatial dependencies, and an Adaptive Residual Feature Enhancement (ARFE) block with gated attention to refine salient features. Experiments on synthetic and real-world benchmarks show that SpikeRain achieves state-of-the-art PSNR and SSIM while reducing parameters by approximately 40% and FLOPs by approximately 89%, with energy efficiency on par with existing SNN methods. These results highlight the potential of SNNs for real-time low-power image restoration on neuromorphic platforms. SpikeRain code is available on ***GitHub****.*

1. Introduction

Single image deraining aims to recover clean visual content from rain-degraded images, where streaks not only reduce image quality but also hinder downstream tasks such as segmentation [40, 71] and object detection [15, 16, 39]. Therefore, with the growing use of vision-based systems in outdoor environments, robust deraining models are crucial for ensuring safety, reliability, and high-fidelity perception.

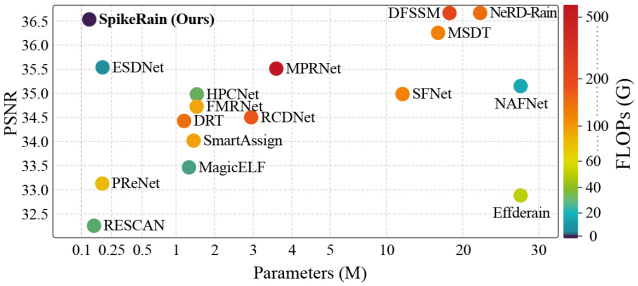


Figure 1. Comparison of model efficiency and performance using PSNR (↑ is better) versus Parameters (↓) with FLOPs (↓) encoded by color. SpikeRain achieves the near-highest PSNR with the fewest parameters and FLOPs.

Early single image deraining methods based on handcrafted priors and optimization techniques [31, 34] assumed simplified rain models, like sparse textures or additive noise. While physically intuitive, these approaches struggled with complex rain patterns, overlapping streaks, and diverse scene structures. With the advent of deep learning, the convolutional neural network (CNN) became the dominant paradigm for image deraining. CNN-based deraining methods [4, 7, 13, 13, 30, 38, 47, 60] leveraged hierarchical feature extraction and large receptive fields to achieve substantial improvements in performance over traditional approaches. However, CNNs inherently lack the capacity to capture long-range dependencies and global context, which are crucial for modeling non-local rain patterns and structures. To overcome these downsides, Transformer-based models [23, 33, 49, 51] have recently been introduced into single image deraining tasks, offering powerful self-attention mechanisms that enable global information modeling. Despite their effectiveness, these models tend to be computationally intensive and demand large memory foot-

*Corresponding authors

prints and extensive training data. Furthermore, most existing architectures overlook temporal dynamics and energy efficiency, factors that are increasingly critical in deployments on resource-constrained devices.

To tackle these limitations, we propose SpikeRain, a novel SNN-based image deraining architecture that is both energy-efficient and temporally dynamic. Unlike conventional CNNs and Transformer-based models, SpikeRain integrates biologically inspired spiking neurons into a deep encoder-decoder framework, allowing the network to model temporal dynamics and event-driven activations across multiple time steps. This enables the network to accumulate contextual information efficiently, while maintaining low computational overhead. Furthermore, we introduce a spiking residual block with dense connectivity (DSRB) to preserve fine-grained spatial features, and a custom-designed Multi-Dimensional Spiking Attention (MDSA) module that captures temporal, channel, and spatial dependencies effectively. Our architecture also incorporates a lightweight Feature Refinement Block with a gated attention fusion mechanism and a Temporal Fusion Module to aggregate temporal features into a consolidated representation that facilitates a clean reconstruction through the decoder. Extensive experiments on standard deraining benchmarks demonstrate that SpikeRain achieves competitive performance while significantly reducing energy consumption, making it highly suitable for deployment on neuromorphic and low-power edge devices. Overall, our research advanced this field with the following key contributions:

- For our proposed SpikeRain method, we design a DSRB with Leaky Integrate-and-Fire (LIF) neurons [65] to enable temporal integration, dense feature reuse, and stable training. Inside DSRB, we further employ an MDSA module that jointly models temporal, channel, and spatial dependencies, allowing the network to suppress complex rain streak patterns.
- We develop an Adaptive Residual Feature Enhancement (ARFE) block with gated attention to refine salient features and enhance perceptual quality in the restored output, which is distinct compared to the residual connection used in the prior SNN-based deraining method [43].
- We conduct extensive experiments on synthetic and real-world benchmarks, showing that SpikeRain achieves competitive PSNR and SSIM while reducing parameters by about 40%, FLOPs by nearly 89%, and energy consumption compared to existing methods, confirming its suitability for low-power neuromorphic deployment.

2. Related Work

2.1. Single Image Deraining

Single image deraining is a challenging task due to the ill-posed nature of recovering clean visual content from rain-degraded images. Conventional deraining techniques [1, 9, 34] heavily rely on handcrafted priors and explicit constraints to guide the rain removal process. While these conventional methods have demonstrated the ability to capture certain low-level characteristics of rain streaks, they often fall short in modeling the underlying semantic structures of the scene, leading to suboptimal visual recovery [9, 34].

With the evolution of learning-based paradigms, deep CNNs have emerged as the dominant approach. These methods approach the deraining task as a supervised image-to-image regression problem, mapping rainy images directly to their clean counterparts. Notable approaches [2, 22, 60] have proposed hierarchical or multi-branch architectures to capture local textures and global context, significantly improving restoration performance. Recent methods [69] decompose the deraining task into multiple subspaces to extract rich multi-scale features that enhance reconstruction accuracy. More recently, vision transformer (ViT) models have gained popularity due to their strong ability to model long-range dependencies. When applied to image deraining, ViT-based architectures such as Restormer [61] and SwinIR [32] have elicited superior performance by leveraging token-level interactions and self-attention mechanisms. However, despite their representational advantages, ViTs typically require higher computational resources and memory, limiting their deployment on edge devices and/or real-time setups.

To address these limitations, researchers have explored lightweight designs using Laplacian pyramids [12], recursive learning [38], and sparse or low-rank representations [42, 49, 54] to reduce computational complexity. While such techniques have improved inference efficiency, they often compromise the model's ability to capture fine-grained spatial details, affecting the overall deraining quality. In contrast to these methods, recent interest has emerged around neuromorphic computing models (e.g., SNNs) for their event-driven nature and low energy consumption.

2.2. Spiking Neural Networks

Deep SNNs have mainly evolved through two approaches. The first is ANN to SNN conversion, where a pretrained ANN is transformed by replacing activations with spiking neurons [8]. Although straightforward, this method often requires many simulation steps to match ANN performance, resulting in high latency and energy costs [59]. Furthermore, the performance of a converted SNN is inherently limited by the source ANN, restricting improvements beyond the original model. In contrast, a more promising approach involves directly training SNNs using surrogate gradient descent [35, 62] or biologically inspired rules such as spike-timing-dependent plasticity (STDP) [53]. This route unfolds the temporal dynamics of the SNN over discrete time steps, enabling effective learning through backpropa-

gation. Unlike converted models, directly trained SNNs are capable of operating with significantly fewer time steps, enhancing their energy efficiency [25]. Recent advances have demonstrated that directly trained SNNs can match ANN performance in various tasks, e.g., object recognition [28], detection [44], and segmentation [27] while maintaining reduced computational costs. For instance, Su et al. [44] showed that an SNN could achieve comparable results to its ANN counterpart using only four time steps. Additionally, techniques such as threshold-dependent batch normalization (tdBN) [65] have been introduced to deepen SNN architectures, improving model expressiveness. SNNs have already demonstrated competitive results in classification tasks [19, 41, 50, 58, 66, 67]. However, their use in pixel-level regression tasks, such as single-image deraining, remains relatively underexplored.

3. Methodology

3.1. Proposed Method: SpikeRain

The overall architecture of our proposed SpikeRain is depicted in Figure 2. It follows an encoder–decoder modeling paradigm to progressively encode the degraded rainy image into a robust representation, which is decoded back to a clean output while preserving structural integrity and reducing computational overhead.

To capture temporal dynamics in SNN-based image restoration, the input rainy image $\mathbf{X}_t$ is replicated over T time steps to form a spike sequence $\mathbf{X} = \{\mathbf{X}_t\}_{t=1}^{T}$. The encoder, beginning with an overlap patch embedding layer, then projects these inputs into a high-dimensional spiking feature space. Subsequently, spiking features are passed through multiple DSRB stages, each comprising cascaded spiking convolutional layers built with LIF neurons [65] and interleaved with tdBN [65]. Each DSRB extracts and fuses rain-specific spike patterns over time using stacked convolutions and the embedded MDSA module. MDSA dynamically recalibrates the spike responses using temporal, channel, and spatial attention, enabling SpikeRain to better focus on the structure and locality of rain streaks.

Downsampling is performed with spiking max pooling and stride convolutions to reduce resolution and capture high-level semantics efficiently, while upsampling uses bilinear interpolation followed by spike-aware convolutions. In contrast, upsampling in the decoder employs bilinear interpolation, followed by spike-aware convolutions to recover spatial resolution. Decoder blocks reuse the DSRB structure to symmetrically refine the decoded features. A Temporal Fusion (TF) module is placed after the decoding layers to aggregate spiking features across time steps using learned soft attention weights.

To further enhance the reconstructed output, SpikeRain introduces an ARFE module. ARFE fuses spatial and channel-specific statistics through gated attention mechanisms, selectively amplifying salient features and suppressing residual artifacts. Finally, a 3×3 convolutional head produces the derained image. The 3×3 kernel size offers a balanced trade-off between computational efficiency and the ability to capture fine spatial details, which is essential for restoring local structures affected by rain streaks. Residual learning is applied not only in the final prediction head, but also throughout the network via residual connections in each DSRB. This strategy stabilizes training and accelerates convergence by preserving high-frequency cues from the input image at multiple stages.

3.2. Dense Spiking Residual Block (DSRB)

Regions in rainy images typically exhibit pronounced local intensity fluctuations compared to non-rainy regions, leading to highly variable spiking neuron activations. The proposed DSRB (Figure 2.a) uses dense connections to integrate spiking neuron outputs across multiple convolutional layers, effectively modeling complex spike interactions caused by rain degradation.

The block begins by processing the spike-based inputs through a series of cascaded *Spiking Control Units* (SCUs). Each SCU comprises LIF neurons [65] that discretize the input into binary spike sequences based on dynamic membrane potential thresholds, effectively extracting temporal spike representations of rain streaks. Formally, at the t-th time step, the spiking convolutional operation for the i-th SCU within DSRB is formulated as follows:

$$\mathbf{X}_{\text{SCU}}^{t,i} = \text{LIF}\left(\text{Conv}\left(\mathbf{X}^{t,i-1}; \boldsymbol{\theta}_i\right)\right), \tag{1}$$

where $\mathbf{X}^{t,i-1}$ denotes the input feature maps from the previous SCU layer, $\text{Conv}(\cdot; \boldsymbol{\theta}_i)$ indicates convolutional operation with learnable parameters $\boldsymbol{\theta}_i$, and $\text{LIF}(\cdot)$ represents the spike activation function of the LIF neuron:

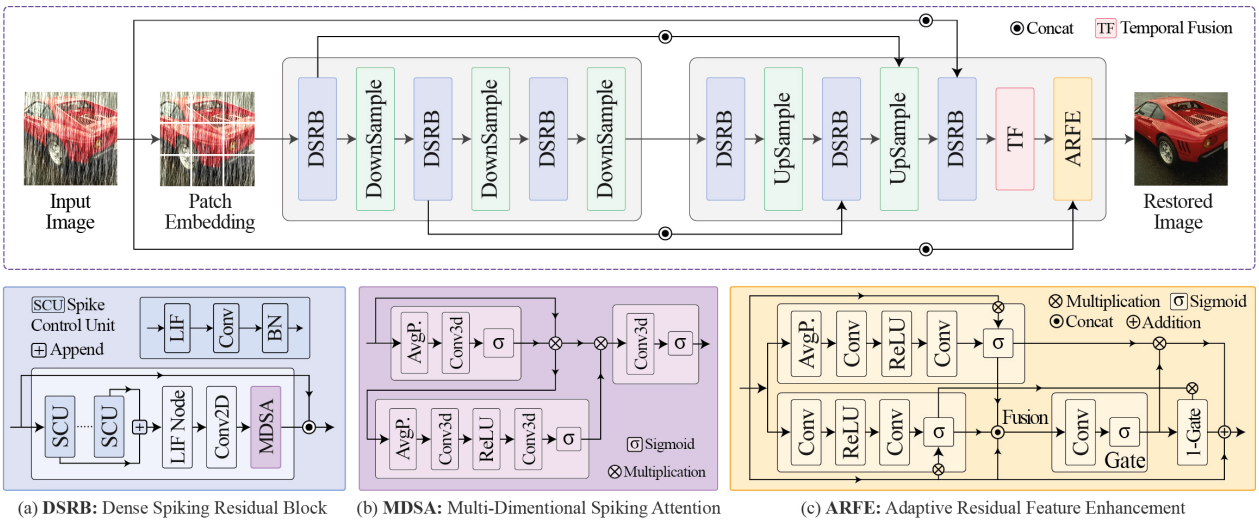
$$\text{LIF}(z) = \begin{cases} 1, & V_{\text{mem}}(z) \geq V_{\text{th}}, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

with membrane potential $V_{\text{mem}}(z)$ dynamically evolving and resetting after each spike event, and V_{th} denoting the neuron firing threshold. To effectively handle the instability from sparse spike information, a parallel convolutional branch followed by tdBN [65] is introduced. This normalization layer maintains both spatial and temporal coherence, mitigating spike-driven activation instabilities and gradient explosions. The normalized output at step t for convolutional layer j can be represented as follows:

$$\mathbf{X}_{\text{tdBN}}^{t,j} = \text{tdBN}\left(\text{Conv}\left(\mathbf{X}_{\text{SCU}}^{t,j-1}; \boldsymbol{\phi}_j\right)\right), \tag{3}$$

where $\boldsymbol{\phi}_j$ denotes convolution parameters, distinct from $\boldsymbol{\theta}_i$.

Subsequently, the spiking features undergo further recalibration through the embedded MDSA module. MDSA



(a) **DSRB:** Dense Spiking Residual Block (b) **MDSA:** Multi-Dimentional Spiking Attention (c) **ARFE:** Adaptive Residual Feature Enhancement

Figure 2. **Overview of the proposed SpikeRain architecture**, which adopts an encoder-decoder structure with Dense Spiking Residual Block (DSRB) and temporal fusion (TF) for efficient deraining. Key modules include the DSRB with Spike Control Unit (SCU), a Multi-Dimensional Spiking Attention (MDSA), and an Adaptive Residual Feature Enhancement (ARFE).

refines spike sequences by adaptively assigning attention weights across temporal, channel, and spatial dimensions, formalized as follows:

$$\mathbf{X}^t_{\text{MDSA}} = \text{MDSA}\left(\mathbf{X}^{t,j}_{\text{tdBN}}\right), \tag{4}$$

where $\text{MDSA}(\cdot)$ is described in detail in Subsection 3.3. Finally, a dense concatenation of all prior outputs is fused through local feature fusion (LFF), compressing spiking representations and forming the residual structure of the block as follows:

$$\mathbf{X}^{t,n} = \text{LFF}\left(\left[\mathbf{X}^t_{\text{MDSA}}, \mathbf{X}^{t,1}_{\text{SCU}}, \ldots, \mathbf{X}^{t,n}_{\text{SCU}}\right]\right) + \mathbf{X}^{t,n-1}, \tag{5}$$

where $[\cdot]$ denotes channel-wise concatenation and LFF is a 1×1 convolutional operation to reduce dimensionality.

3.3. Multi-Dimensional Spiking Attention (MDSA)

The MDSA module addresses the sparsity and variability of spiking representations by selectively emphasizing informative features across temporal, channel, and spatial dimensions (Figure 2.b). Integrated within the DSRB, it adaptively recalibrates spike activities to suppress rain streaks and preserve structural details. Unlike prior attention-based SNNs (e.g., ASNNs [43, 59]), which mainly focus on channel-wise recalibration, MDSA jointly models temporal, channel, and spatial dependencies. This richer formulation is particularly suited for deraining, where temporal spike dynamics and localized streak patterns must be captured together. To remain efficient, MDSA employs bottleneck convolutions with reduction ratios of $r_c = 16$ for the channel branch and $r_t = 4$ for the temporal branch, reducing computation while preserving attention selectivity.

Given the input spike features $\mathbf{X} \in \mathbb{R}^{T\times B\times C\times H\times W}$, where T, B, C, H, and W represent the temporal dimension, batch size, channels, height, and width, respectively, MDSA concurrently computes three attention maps:

- *Temporal Attention* (TA), which captures the significance of spike activity over time. To obtain temporal attention weights, we first conduct adaptive temporal pooling, followed by a sigmoid-based gating operation as follows:

$$\mathbf{A}_{\text{T}} = \sigma\left(\text{Conv3D}\left(\text{AvgPool}_{3D}(\mathbf{X})\right)\right), \tag{6}$$

  where $\mathbf{A}_{\text{T}} \in \mathbb{R}^{B\times 1\times T\times 1\times 1}$ encodes the temporal significance of each frame, and $\sigma(\cdot)$ is the sigmoid activation.
- *Channel Attention* (CA): It emphasizes the channel-wise importance of spiking features. CA is computed by performing channel-wise dimensionality reduction and restoration via two sequential convolutions with nonlinear ReLU activations as follows:

$$\mathbf{A}_{\text{C}} = \sigma\left(\text{Conv3D}\left(\text{ReLU}\left(\text{Conv3D}\left(\text{AvgPool}_{3D}(\mathbf{X})\right)\right)\right)\right), \tag{7}$$

  where $\mathbf{A}_{\text{C}} \in \mathbb{R}^{B\times C\times 1\times 1\times 1}$ characterizes the relative channel importance within the spike sequence.
- *Spatial Attention* (SA), which targets the spatial distribution of spikes to highlight local features most indicative of rain degradation. SA is calculated through a spatial convolution operation on pooled spatial spike features as:

$$\mathbf{A}_{\text{S}} = \sigma\left(\text{Conv3D}\left(\text{AvgPool}_{\text{channel}}(\mathbf{X})\right)\right), \tag{8}$$

  where $\mathbf{A}_{\text{S}} \in \mathbb{R}^{B\times 1\times T\times H\times W}$ assesses the spatial relevance of spike activations.

The three attention maps multiply with spike features to produce recalibrated representations as follows:

$$\mathbf{X}_{\text{MDSA}} = \mathbf{X} \odot \mathbf{A}_{\text{T}} \odot \mathbf{A}_{\text{C}} \odot \mathbf{A}_{\text{S}}, \tag{9}$$

where $\odot$ represents element-wise multiplication. Through this threefold attention scheme, MDSA effectively refines spike responses, ensuring adaptive modulation tailored to temporal dynamics, channel relevance, and spatial prominence, thereby significantly enhancing the capability to model complex rain-degraded patterns.

3.4. Temporal Fusion (TF)

After refining spiking features across spatial, channel, and temporal dimensions through the DSRB and MDSA modules, we employ a temporal fusion mechanism to consolidate the time-distributed feature representations into a single map. This is essential for bridging the temporally encoded dynamics of spiking neural responses with the spatially dense operations that follow. Let the temporally encoded output from the final DSRB be denoted as $\mathbf{X} \in \mathbb{R}^{T\times B\times C\times H\times W}$, where T is the spike time window. TF introduces a set of learnable scalar weights $\boldsymbol{\alpha} \in \mathbb{R}^T$, normalized by a softmax to produce $\boldsymbol{w} = \text{softmax}(\boldsymbol{\alpha})$. The fused representation is then computed as $\bar{\mathbf{X}} = \sum_{t=1}^{T} w_t \cdot \mathbf{X}_t$, where $\mathbf{X}_t$ is the spike feature map at time step t.

This soft attention over time steps enables the model to adaptively emphasize more informative spike responses, while reducing redundancy from temporally correlated patterns. The fused output $\bar{\mathbf{X}}$ is then passed to the ARFE module for final reconstruction.

3.5. Adaptive Residual Feature Enhancement (ARFE)

Unlike MDSA, ARFE operates purely in the spatial-channel domain after temporal fusion, refining continuous feature maps to enhance visual clarity without temporal consideration. Although spike-based representations effectively capture temporal and structural characteristics of rain-induced degradation, residual artifacts and subtle distortions can persist in the decoded images. To address this limitation, the ARFE module (illustrated in Figure 2.c) adaptively refines decoded spiking features through dual-channel and spatial attention mechanisms integrated via a gated fusion strategy. By selectively amplifying essential spiking patterns and attenuating noisy signals, ARFE significantly enhances the perceptual clarity and restoration accuracy of derained.

Given the decoded spike-driven features $\mathbf{F}$, ARFE sequentially applies attention-driven modulation in the channel and spatial domains. On one hand, **Channel Attention** (CA) first aggregates global context via adaptive average pooling, capturing global spike activation statistics. Subsequently, a two-layer gating network dynamically determines the significance of each channel as follows:

$$\mathbf{F}_{\text{CA}} = \sigma\left(\text{Conv}_{1\times1}\left(\text{ReLU}\left(\text{Conv}_{1\times1}\left(\text{AvgPool}_{2D}\left(\mathbf{F}\right)\right)\right)\right)\right) \odot \mathbf{F}, \tag{10}$$

where $\mathbf{F}_{\text{CA}}$ denotes the channel-attended features, and the $\sigma(\cdot)$ function implements sigmoid-based gating.

On the other hand, **Spatial Attention** (SA) computes pixel-level relevance to pinpoint spatially significant spike responses. This spatial attention operation consists of a light convolutional projection followed by non-linear activation and refinement through a larger receptive field convolution:

$$\mathbf{F}_{\text{SA}} = \sigma\left(\text{Conv}_{5\times5}\left(\text{ReLU}\left(\text{Conv}_{1\times1}\left(\mathbf{F}\right)\right)\right)\right) \odot \mathbf{F}, \tag{11}$$

where $\mathbf{F}_{\text{SA}}$ denote spatially refined feature maps.

To optimally integrate the attentional insights produced by CA and SA, ARFE introduces an adaptive gating mechanism. This mechanism learns to dynamically balance the two attention streams by generating a gate map conditioned on the concatenated original features, and the respective channel and spatial attention maps as follows:

$$\mathbf{G} = \sigma\left(\text{Conv}_{1\times1}\left(\left[\mathbf{F}, \mathbf{F}_{\text{CA}}, \mathbf{F}_{\text{SA}}\right]\right)\right), \tag{12}$$

where $[\cdot]$ indicates concatenation along the channel dimension, and $\mathbf{G}$ defines the adaptive gating weights. Finally, the enhanced residual feature representation $\mathbf{F}_{\text{ARFE}}$ is obtained by a gated combination of both attention maps, preserving complementary information from each dimension as:

$$\mathbf{F}_{\text{ARFE}} = \mathbf{G} \odot \mathbf{F}_{\text{CA}} + (1 - \mathbf{G}) \odot \mathbf{F}_{\text{SA}} + \mathbf{F}. \tag{13}$$

Through this adaptive and residual formulation, ARFE enhances feature representation, reduces artifacts, and yields cleaner, more coherent derained outputs.

4. Experiments and Results

4.1. Experimental Setup

Datasets. To maintain consistency across evaluations, all models under comparison were retrain on a common set of four publicly accessible datasets: Rain12 [31], Rain200L [57], Rain200H [56], and Rain1200 [63]. The Rain12 dataset includes 12 artificially generated rainy images, used exclusively for testing. Both Rain200L [57] and Rain200H [56] provide 1,800 synthetic rain images for training and an additional 200 for testing. The Rain1200 [63] dataset offers a larger scale, comprising 12,000 training images and 1,200 test images with diverse rain patterns in terms of direction and intensity. Additionally, we utilize the RW-Data [64] benchmark, which contains 185 real-world rainy images.

Comparison Methods. We compare our proposed method with a wide range of image deraining baselines, including seven CNN-based methods (i.e., RESCAN [30], PReNet [38], RCDNet [47], MPRNet [60], Effderain [13], NAFNet [4], and SFNet [7]), seven Transformer-based networks (i.e., DRT [33], MagicELF [23], HPCNet [49], SmartAssign [51], MSDT [3], NeRD-Rain [6], and FMR-Net [24]), one self-supervised and one spike neural network

model (i.e., DFSSM [55], and ESDNet [43]). For SmartAssign, due to the lack of publicly available code, we re-implement the model based on the details provided in their paper and retrain it under our setting. For other methods, if official pretrained models are available, we use the released code and weights for evaluation. Otherwise, we retrain them under the same experimental setup as our model.

Evaluation Metrics. To ensure practical and consistent evaluation, we adopt two widely used image quality metrics: Peak Signal-to-Noise Ratio (PSNR) [18] and Structural Similarity Index (SSIM) [52]. These metrics are utilized to quantitatively assess the fidelity of the generated derained images. In line with prior works [5, 43], evaluations are carried out on the Y (luminance) channel within the YCbCr color space. For real-world datasets lacking ground truth, we employ no-reference quality assessment methods, including the Perception-based Image Quality Evaluator (PIQE) [46] and the Meta-learning-based Image Quality Assessment (MetaIQA) [68], to measure the visual quality.

Implementation Details. We implement SpikeRain using PyTorch and SpikingJelly's activation-based interface. All spiking components, including LIF neurons [65] and tdBN [65], operate under the multistep mode (`step_mode='m'`) to enable temporal integration over discrete spike time steps. The encoder–decoder follows a 3-stage spatial hierarchy with channel scaling by a factor of 2 at each level. The default model variant (SpikeRain-M) uses an embedding dimension of 48 with [4, 4, 8, 8] DSRBs in the encoder and [2, 2, 2, 2] in the decoder. All DSRBs incorporate LIF-based convolution, dense skip connections, and a built-in MDSA module, with attention reduction ratios set to $r_c = 16$ and $r_t = 4$. The temporal length is fixed at $T = 4$ for all experiments unless otherwise stated. For training, we adopt surrogate gradients [35, 62] to enable backpropagation through the non-differentiable LIF neurons [65], using a sigmoid approximation with slope $\alpha_s = 4.0$ for stable temporal credit assignment across time steps. The network is optimized with the SSIM loss, which better preserves structural fidelity than pixel-wise objectives, and trained using the AdamW optimizer with a batch size of 22 input-output pairs, and an initial learning rate of 1×10^{-3}, annealed to 1×10^{-7} using cosine decay after 5 warm-up epochs. All models are trained for 1,000 epochs, while inputs are cropped to 64×64 during training. Experiments are run on two NVIDIA RTX 3080Ti GPUs, each with 24GB VRAM.

4.2. Quantitative Analysis of the Results

Comparison with Deraining Methods. Table 1 presents a comprehensive comparison between our SpikeRain model and 12 recent state-of-the-art (SOTA) deraining methods across the aforementioned four synthetic benchmarks. SpikeRain achieves the highest average PSNR (36.02 dB) and SSIM (0.9630), while maintaining the lowest computational footprint: only 0.808 GFLOPs and 1.01×10^4 μJ energy. On the Rain200H dataset, which is particularly challenging due to heavy streaking, SpikeRain attains 31.17 dB PSNR, outperforming ESDNet and SFNet by a clear margin. Although some transformer-based and CNN models, such as MPRNet and DFSSM, achieve competitive or higher PSNR values, they require a significantly higher cost, exceeding 500 GFLOPs and 2.6×10^6 μJ in some cases. This highlights the critical performance-efficiency trade-off in deraining tasks. Furthermore, we evaluate the real-world effectiveness of our model on the RW-Data benchmark. As shown in Table 2, our SpikeRain achieves the best performance with a PIQE score of 16.254 (lower is better) and a MetaIQA score of 0.463 (higher is better), outperforming all baselines. These results demonstrate that our model restores clearer content with higher perceptual quality while maintaining an efficient PSNR to energy ratio suitable for low-power real-time vision tasks.

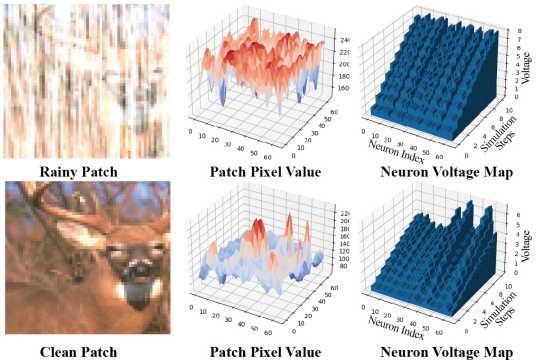


Figure 3. Neuron membrane potential dynamics in SpikeRain. Higher voltage fluctuations are observed in rain-affected regions, reflecting increased spiking activity and the model's adaptive response to local rain patterns.

Comparison with Multi-Weather Methods. Table 3 presents a comprehensive comparison of our SpikeRain model against 10 recent multi-weather restoration approaches, evaluated on the Rain1400 [11] and SPA+ [70] benchmarks. SpikeRain achieves the highest performance on both datasets, attaining 33.05 dB PSNR and 0.94 SSIM on Rain1400, and 39.85 dB PSNR and 0.98 SSIM on SPA+, and outperforming Transformer-based models such as MWFormer [69], WGWS-Net [70], and TransWeather [45]. Compared to traditional CNN methods like AirNet [29] and PatilNet [36], SpikeRain demonstrates substantial improvements in both synthetic and real-world scenarios. These findings validate the generalization capability of our SpikeRain architecture across diverse weather-induced degradations, preserving perceptual quality and structural fidelity.

Model Efficiency. In conventional ANNs, the computational cost is dominated by the FLOPs, typically realized

Type	Model	Venue	Rain12 [31]		Rain200L [57]		Rain200H [56]		Rain1200 [63]		Average		Params↓	FLOPs↓	Energy↓
			PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	(M)	(G)	(uJ)
CNN	RESCAN [30]	ECCV'18	35.48	0.9499	33.82	0.9547	26.22	0.8219	32.60	0.9271	32.03	0.9134	0.149	32.119	4.014×10^5
	PReNet [38]	CVPR'19	36.32	0.9593	36.06	0.9735	27.18	0.8698	33.24	0.9280	33.24	0.9327	0.168	66.249	8.281×10^5
	RCDNet [47]	CVPR'20	37.67	0.9612	38.84	0.9854	28.98	0.8889	34.68	0.9189	34.54	0.9386	2.958	194.501	2.431×10^6
	MPRNet [60]	CVPR'21	37.55	0.9665	39.82	0.9863	29.94	0.8999	34.50	0.9369	35.46	0.9474	3.637	548.651	6.858×10^6
	Effderain [13]	AAAI'21	36.12	0.9588	36.06	0.9731	26.11	0.8341	32.85	0.9147	32.78	0.9202	27.654	52.915	6.614×10^5
	NAFNet [4]	ECCV'22	37.53	0.9639	39.48	0.9847	29.19	0.8880	33.69	0.9367	35.22	0.9433	29.102	16.064	2.008×10^5
	SFNet [7]	ICLR'23	36.11	0.9503	39.50	0.9850	29.75	0.9008	34.51	0.9383	34.97	0.9437	13.234	124.439	1.555×10^6
ViT	DRT [33]	CVPR'22	37.74	0.9674	38.81	0.9830	28.67	0.8796	32.88	0.9283	34.78	0.9395	1.176	166.045	2.075×10^6
	MagicELF [23]	ACM MM'22	35.06	0.9420	38.85	0.9800	28.93	0.8852	34.00	0.9360	33.48	0.9358	1.221	23.667	2.958×10^5
	HPCNet [49]	IEEE ICME'23	36.84	0.9639	39.14	0.9847	29.17	0.8962	33.92	0.9378	34.90	0.9378	1.411	29.540	3.692×10^5
	SmartAssign [51]	CVPR'23	36.87	0.9618	38.41	0.9814	27.71	0.8536	34.03	0.9154	34.03	0.9281	1.359	90.386	1.129×10^6
	MSDT [3]	AAAI'24	36.54	0.9514	41.75	0.9904	32.45	0.9349	34.40	0.9332	36.28	0.9532	16.6	129.9	1.623×10^6
	NeRD-Rain [6]	CVPR'24	37.95	0.9654	41.71	0.9903	32.40	0.9373	34.65	0.9395	**36.67**	0.9581	22.89	156.3	1.953×10^6
	FMRNet [24]	AAAI'24	36.96	0.9440	37.81	0.9740	30.69	0.9000	33.21	0.9240	34.66	0.9355	1.551	91.23	1.140×10^6
SSM	DFSSM [55]	ACCV'24	37.95	0.9655	**41.81**	0.9905	**32.99**	**0.9403**	33.92	0.9254	36.66	0.9554	19.0	212.4	2.655×10^6
SNN	ESDNet [43]	IJCAI'24	37.82	0.9681	39.85	0.9869	30.01	0.9132	34.52	0.9388	35.55	0.9518	0.165	7.320	9.150×10^4
	SpikeRain (ours)	-	**38.12**	**0.9699**	40.01	**0.9948**	31.17	0.9352	**34.80**	**0.9421**	36.02	**0.9630**	**0.099**	**0.808**	$\mathbf{1.01 \times 10^4}$

Table 1. Comparison of quantitative results on four synthetic benchmarks. Here, **bold** and underline indicate the overall best and second-best results, respectively, among the models from all categories. The gray color rows indicate the best-performing models from each category, e.g., CNN, ViT, SSM, and SNN, based on the average PSNR (↑) and SSIM (↑).

Methods	RESCAN	PReNet	RCDNet	MPRNet	Effderain	NAFNet	SFNet	DRT	MagicELF	HPCNet	SmartAssign	ESDNet	**SpikeRain (Ours)**
PIQE (↓)	18.364	18.309	19.692	18.061	18.400	18.106	18.124	18.392	18.037	18.287	18.316	17.959	**16.254**
MetaIQA (↑)	0.419	0.420	0.421	0.426	0.421	0.420	0.423	0.419	0.422	0.426	0.417	0.423	**0.463**

Table 2. Quantitative comparison of proposed SpikeRain with SOTA methods testing on real-world rainy dataset "RW-Data".

Model	Venue	Rain1400 [11]		SPA+ [70]	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑
AirNet [29]	CVPR'22	22.86	0.69	21.15	0.75
TransWeather [45]	CVPR'22	26.30	0.91	30.12	0.90
WeatherDiff. [72]	TPAMI'23	23.45	0.82	20.21	0.78
PatilNet [36]	ICCV'23	24.85	0.89	27.00	0.88
WGWS-Net [70]	CVPR'23	25.00	0.76	28.35	0.89
MetaWeather [26]	ECCV'24	23.65	0.84	23.45	0.81
HazeSpace2M [21]	ACM MM'24	22.10	0.80	19.50	0.68
TANet [48]	ACCV'24	28.90	0.91	31.25	0.91
MWFormer [69]	TIP'24	30.60	0.93	31.90	0.92
AWRaCLe [37]	AAAI'25	30.25	0.91	31.45	0.91
SpikeRain	(Ours)	**33.05**	**0.94**	**39.85**	**0.98**

Table 3. Performance comparison of our SpikeRain with recent multi-weather image restoration methods on Rain1400 (Synthetic) and SPA+ (Real-world) datasets.

Model	Venue	FLOPs (G)↓	Energy (μJ)↓	PSNR↑	SSIM↑
ANN-ResNet [17]	CVPR'16	83.453	1.043×10^6	28.79	0.8942
SEW-ResNet [10]	NeurIPS'21	6.998	8.765×10^4	26.61	0.8613
MS-ResNet [20]	IEEE TNNLS'24	6.970	8.730×10^4	25.20	0.8230
DSRB	(Ours)	**0.752**	$\mathbf{0.937 \times 10^4}$	**30.19**	**0.9327**

Table 4. Ablation study with various residual blocks in SpikeRain.

through multiply-accumulate instructions. Our proposed SpikeRain network primarily builds upon spiking neuron modules, but uses conventional FLOPs in the patch embedding and final output layers. Following prior work [14], we estimate the energy consumed by our SpikeRain based on the cost of a single FLOP as 12.5 pJ. Given the total computational complexity of 0.808 GFLOPs in SpikeRain, the estimated energy consumption is 10,100 μJ for processing a 64×64 input image. Compared to CNN- and transformer-based counterparts, SpikeRain achieves significantly improved energy efficiency while maintaining strong PSNR and SSIM scores (Table 1).

4.3. Qualitative Analysis of the Results

Figure 3 illustrates the spike response dynamics across the spatial domain, where higher spike activity is observed around rainy regions. This confirms that SpikeRain's event-driven architecture adaptively responds to rain-affected patterns with higher spiking frequency. Additionally, we present visual comparisons to assess the perceptual effectiveness of SpikeRain in both synthetic and real-world scenarios. As shown in Figures 4 and 5, SpikeRain produces cleaner and sharper results compared to SOTA methods, effectively removing rain streaks while preserving fine textures and edges. On both synthetic (Rain200H, Rain1400) and real-world (SPA+) datasets, our SpikeRain recovers details closer to the ground truth, demonstrating strong generalization across diverse scenarios.

4.4. Ablation Studies

We conduct ablations on Rain200H to assess the impact of residual blocks, time steps, and key network components.

Effectiveness of Different Residual Blocks. Table 4 compares various residual block designs. Our proposed DSRB achieves the best PSNR (30.19 dB) and SSIM (0.9327) while requiring only 0.752 GFLOPs, the lowest among all the residual blocks. These gains stem from dense skip connections and embedded attention, which enhance spike

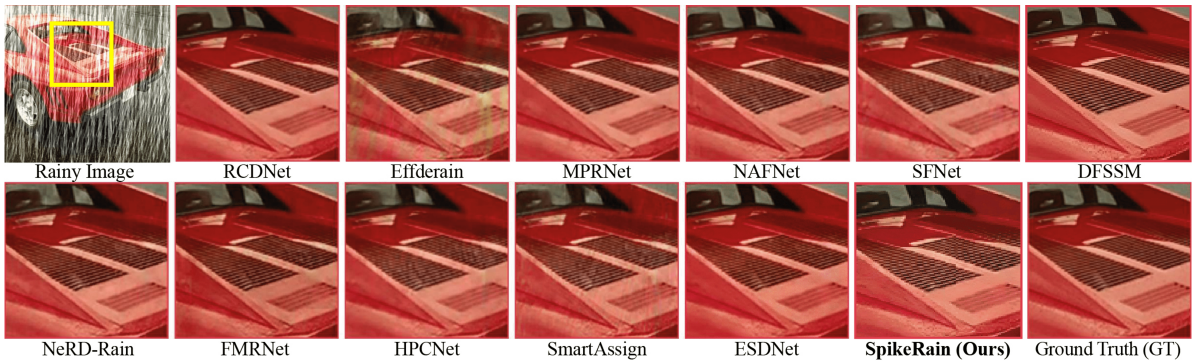


Figure 4. Visual comparison of deraining performance on a sample image from the Rain200H dataset.

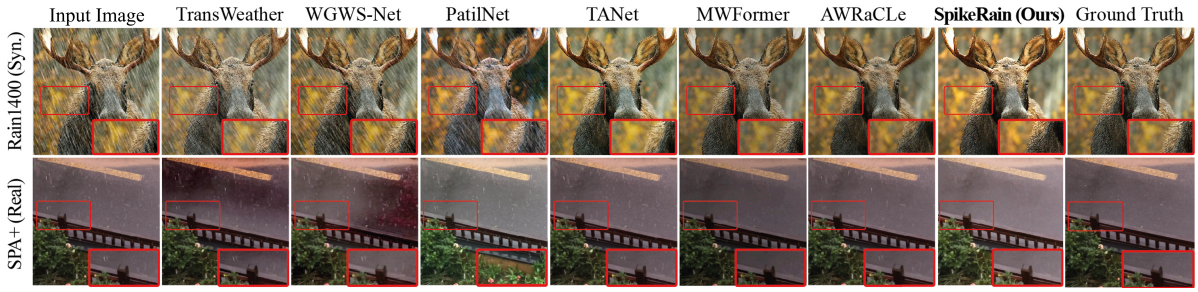


Figure 5. Results for the Rain1400 (synthetic) and SPA+ (real) datasets show that our SpikeRain recovers finer textures and details better than existing multi-weather image registration methods, closely matching the ground truth.

Time steps	FLOPs (G)↓	Energy (uJ)↓	PSNR↑	SSIM↑
$T = 1$	0.047	0.587×10^4	28.81	0.8963
$T = 2$	0.065	0.812×10^4	30.19	0.9327
$T = 4$ (**Baseline**)	0.808	$\mathbf{1.01 \times 10^4}$	31.17	0.9452

Table 5. Performance impact of different T on the Rain200H.

Model	DSRB	MDSA	ARFE	TF	PSNR↑	SSIM↑
(a)	✓				30.19	0.9327
(b)	✓	✓			30.91	0.9393
(c)	✓		✓		30.43	0.9345
(d)	✓	✓	✓		31.05	0.9423
(e)	✓	✓	✓	✓	**31.17**	**0.9452**

Table 6. Normalized ablation results of different combinations of SpikeRain components on the Rain200H dataset.

modulation and gradient propagation. Compared to MS-ResNet and SEW-ResNet, our DSRB reduces energy usage, validating its strength for efficient spike-based deraining.

Effectiveness of Different Time Steps. Table 5 reports the performance of SpikeRain for varying time steps T. As the value of T increases from 1 to 4, PSNR improves from 28.81 to 31.17, demonstrating that temporal integration enhances the extraction of spatiotemporal features. However, a higher T also raises FLOPs and energy consumption. We propose to use $T = 4$ with SpikeRain as it gets better performance while maintaining minimum energy consumption and GFLOPS compared to the existing methods.

Effectiveness of Different Designs. We finally evaluate the relative contribution to performance of MDSA, ARFE, and TF via ablations. The results of this study are summarized in Table 6. Incorporating MDSA into the baseline (a) boosts temporal feature learning (30.91 PSNR), while ARFE refines spatial and channel cues for further improvement. The complete model, combining all modules, achieves the best performance (31.17 PSNR), confirming the contribution of each compounding block of SpikeRain.

5. Conclusion

In this work, we have presented SpikeRain, a novel SNN architecture tailored for energy-efficient single image deraining. Unlike conventional ANN- and Transformer-based approaches, SpikeRain exploits the temporal dynamics and sparse event-driven nature of spiking neurons to achieve low-cost yet high-quality restoration. With the proposed DSRB, MDSA, ARFE, and temporal fusion, SpikeRain achieves competitive PSNR and SSIM while reducing parameters by 40% and FLOPs by 89%. Results on synthetic and real datasets highlight its ability to preserve fine details under strict efficiency constraints, making it well-suited for neuromorphic hardware and low-power edge deployment.

Acknowledgments

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2025-02214941); in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean Government (MSIT) (No. RS-2025-02218631); in part by the BK21 FOUR Project (Bigdata Research and Education Group for Enhancing Social Connectedness Thorough Advanced Data Technology and Interaction Science Research, 5199990913845), funded by the Ministry of Education (MOE, Korea) and the National Research Foundation of Korea (NRF); and in part by the "Regional Innovation System & Education (RISE)" through the Seoul RISE Center, funded by the Ministry of Education (MOE) and the Seoul Metropolitan Government (2025-RISE-01-018-04).

References

[1] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1375–1383. IEEE, 2019. 2

[2] Hongming Chen, Xiang Chen, Jiyang Lu, and Yufeng Li. Rethinking multi-scale representations in deep deraining transformer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1046–1053, 2024. 2

[3] Hongming Chen, Xiang Chen, Jiyang Lu, and Yufeng Li. Rethinking multi-scale representations in deep deraining transformer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1046–1053, 2024. 5, 7

[4] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022. 1, 5, 7

[5] 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. 6

[6] Xiang Chen, Jinshan Pan, and Jiangxin Dong. Bidirectional multi-scale implicit neural representations for image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25627–25636, 2024. 5, 7

[7] Yuning Cui, Yi Tao, Zhenshan Bing, Wenqi Ren, Xinwei Gao, Xiaochun Cao, Kai Huang, and Alois Knoll. Selective frequency network for image restoration. In *The eleventh international conference on learning representations*, 2023. 1, 5, 7

[8] Peter U Diehl, Daniel Neil, Jonathan Binas, Matthew Cook, Shih-Chii Liu, and Michael Pfeiffer. Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. In *2015 International joint conference on neural networks (IJCNN)*, pages 1–8. ieee, 2015. 2

[9] Xinghao Ding, Liqin Chen, Xianhui Zheng, Yue Huang, and Delu Zeng. Single image rain and snow removal via guided l0 smoothing filter. *Multimedia Tools and Applications*, 75: 2697–2712, 2016. 2

[10] Wei Fang, Zhaofei Yu, Yanqi Chen, Tiejun Huang, Timothée Masquelier, and Yonghong Tian. Deep residual learning in spiking neural networks. *Advances in Neural Information Processing Systems*, 34:21056–21069, 2021. 7

[11] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 6, 7

[12] Xueyang Fu, Borong Liang, Yue Huang, Xinghao Ding, and John Paisley. Lightweight pyramid networks for image deraining. *IEEE Transactions on Neural Networks and Learning Systems*, 31(6):1794–1807, 2019. 2

[13] Qing Guo, Jingyang Sun, Felix Juefei-Xu, Lei Ma, Xiaofei Xie, Wei Feng, Yang Liu, and Jianjun Zhao. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1487–1495, 2021. 1, 5, 7

[14] Yufei Guo, Yuanpei Chen, Xiaode Liu, Weihang Peng, Yuhan Zhang, Xuhui Huang, and Zhe Ma. Ternary spike: Learning ternary spikes for spiking neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, pages 12244–12252, 2024. 7

[15] Yazan Hamzeh, Zaid El-Shair, and Samir A Rawashdeh. Effect of adherent rain on vision-based object detection algorithms. *SAE International Journal of Advances and Current Practices in Mobility*, 2(2020-01-0104):3051–3059, 2020. 1

[16] Sinan Hasirlioglu and Andreas Riener. Challenges in object detection under rainy weather conditions. In *First International Conference on Intelligent Transport Systems*, pages 53–65. Springer, 2018. 1

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 7

[18] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369. IEEE, 2010. 6

[19] Yifan Hu, Lei Deng, Yujie Wu, Man Yao, and Guoqi Li. Advancing spiking neural networks toward deep residual learning. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–15, 2024. 3

[20] Yifan Hu, Lei Deng, Yujie Wu, Man Yao, and Guoqi Li. Advancing spiking neural networks toward deep residual learning. *IEEE Transactions on Neural Networks and learning systems*, 36(2):2353–2367, 2024. 7

[21] Md Tanvir Islam, Nasir Rahim, Saeed Anwar, Muhammad Saqib, Sambit Bakshi, and Khan Muhammad. Hazespace2m: A dataset for haze aware single image dehazing. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 9155–9164, 2024. 7

[22] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale

progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8346–8355, 2020. 2

[23] Kui Jiang, Zhongyuan Wang, Chen Chen, Zheng Wang, Laizhong Cui, and Chia-Wen Lin. Magic elf: Image deraining meets association learning and transformer. In *Proceedings of the 30th ACM International Conference on Multimedia*, page 827–836, New York, NY, USA, 2022. Association for Computing Machinery. 1, 5, 7

[24] Kui Jiang, Junjun Jiang, Xianming Liu, Xin Xu, and Xianzheng Ma. Fmrnet: Image deraining via frequency mutual revision. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12892–12900, 2024. 5, 7

[25] Jinseok Kim, Kyungsu Kim, and Jae-Joon Kim. Unifying activation-and timing-based learning rules for spiking neural networks. *Advances in Neural Information Processing Systems*, 33:19534–19544, 2020. 3

[26] Youngrae Kim, Younggeol Cho, Thanh-Tung Nguyen, Seunghoon Hong, and Dongman Lee. Metaweather: Few-shot weather-degraded image restoration. In *European Conference on Computer Vision*, pages 206–222. Springer, 2024. 7

[27] Paul Kirkland, Gaetano Di Caterina, John Soraghan, and George Matich. Spikeseg: Spiking segmentation via stdp saliency mapping. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020. 3

[28] Yuxiang Lan, Yachao Zhang, Xu Ma, Yanyun Qu, and Yun Fu. Efficient converted spiking neural network for 3d and 2d classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9211–9220, 2023. 3

[29] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17452–17462, 2022. 6, 7

[30] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European conference on computer vision (ECCV)*, pages 254–269, 2018. 1, 5, 7

[31] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. Rain streak removal using layer priors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2736–2744, 2016. 1, 5, 7

[32] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 2

[33] Yuanchu Liang, Saeed Anwar, and Yang Liu. Drt: A lightweight single image deraining recursive transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 589–598, 2022. 1, 5, 7

[34] Yu Luo, Yong Xu, and Hui Ji. Removing rain from a single image via discriminative sparse coding. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3397–3405, 2015. 1, 2

[35] Emre O Neftci, Hesham Mostafa, and Friedemann Zenke. Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine*, 36 (6):51–63, 2019. 2, 6

[36] Prashant W Patil, Sunil Gupta, Santu Rana, Svetha Venkatesh, and Subrahmanyam Murala. Multi-weather image restoration via domain translation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21696–21705, 2023. 6, 7

[37] Sudarshan Rajagopalan and Vishal M Patel. Awracle: All-weather image restoration using visual in-context learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025. 7

[38] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3937–3946, 2019. 1, 2, 5, 7

[39] Thomas Rothmeier, Diogo Wachtel, Tetmar von Dem Bussche-Hünnefeld, and Werner Huber. I had a bad day: Challenges of object detection in bad visibility conditions. In *2023 IEEE Intelligent Vehicles Symposium (IV)*, pages 1–6. IEEE, 2023. 1

[40] Suvash Sharma, Chris Goodin, Matthew Doude, Christopher Hudson, Daniel Carruth, Bo Tang, and John Ball. Understanding how rain affects semantic segmentation algorithm performance. Technical report, SAE Technical Paper, 2020. 1

[41] Xinyu Shi, Zecheng Hao, and Zhaofei Yu. Spikingresformer: bridging resnet and vision transformer in spiking neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5610–5619, 2024. 3

[42] Tianyu Song, Pengpeng Li, Guiyue Jin, Jiyu Jin, Shumin Fan, and Xiang Chen. Image deraining transformer with sparsity and frequency guidance. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1889–1894. IEEE, 2023. 2

[43] Tianyu Song, Guiyue Jin, Pengpeng Li, Kui Jiang, Xiang Chen, and Jiyu Jin. Learning a spiking neural network for efficient image deraining. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 2024. 2, 4, 6, 7

[44] Qiaoyi Su, Yuhong Chou, Yifan Hu, Jianing Li, Shijie Mei, Ziyang Zhang, and Guoqi Li. Deep directly-trained spiking neural networks for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6555–6565, 2023. 3

[45] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M. Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2353–2363, 2022. 6, 7

[46] Narasimhan Venkatanath, D Praneeth, Maruthi Chandrasekhar Bh, Sumohana S Channappayya, and Swarup S Medasani. Blind image quality evaluation using perception

based features. In *2015 Twenty First National Conference on Communications (NCC)*, pages 1–6. IEEE, 2015. 6

[47] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3103–3112, 2020. 1, 5, 7

[48] Hsing-Hua Wang, Fu-Jen Tsai, Yen-Yu Lin, and Chia-Wen Lin. Tanet: Triplet attention network for all-in-one adverse weather image restoration. In *Proceedings of the Asian Conference on Computer Vision*, pages 835–851, 2024. 7

[49] Qiong Wang, Kui Jiang, Jinyi Lai, Zheng Wang, and Jianhui Zhang. Hpcnet: A hybrid progressive coupled network for image deraining. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*, pages 2747–2752. IEEE, 2023. 1, 2, 5, 7

[50] Shuai Wang, Malu Zhang, Dehao Zhang, Ammar Belatreche, Yichen Xiao, Yu Liang, Yimeng Shan, Qian Sun, Enqi Zhang, and Yang Yang. Spiking vision transformer with saccadic attention. In *The Thirteenth International Conference on Learning Representations*, 2025. 3

[51] Yinglong Wang, Chao Ma, and Jianzhuang Liu. Smartassign: Learning a smart knowledge assignment strategy for deraining and desnowing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3677–3686, 2023. 1, 5, 7

[52] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 6

[53] Ziqing Wang, Yuetong Fang, Jiahang Cao, Qiang Zhang, Zhongrui Wang, and Renjing Xu. Masked spiking transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1761–1771, 2023. 2

[54] Yunyang Xiong, Zhanpeng Zeng, Rudrasis Chakraborty, Mingxing Tan, Glenn Fung, Yin Li, and Vikas Singh. Nyströmformer: A nyström-based algorithm for approximating self-attention. In *Proceedings of the AAAI conference on artificial intelligence*, pages 14138–14148, 2021. 2

[55] Shugo Yamashita and Masaaki Ikehara. Image deraining with frequency-enhanced state space model. In *Proceedings of the Asian Conference on Computer Vision*, pages 3655–3671, 2024. 6, 7

[56] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1357–1366, 2017. 5, 7

[57] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1357–1366, 2017. 5, 7

[58] Man Yao, Jiakui Hu, Zhaokun Zhou, Li Yuan, Yonghong Tian, Bo Xu, and Guoqi Li. Spike-driven transformer. *Advances in Neural Information Processing Systems*, 36: 64043–64058, 2023. 3

[59] Man Yao, Guangshe Zhao, Hengyu Zhang, Yifan Hu, Lei Deng, Yonghong Tian, Bo Xu, and Guoqi Li. Attention spiking neural networks. *IEEE Transactions on Pattern Analysis and Machine intelligence*, 45(8):9393–9410, 2023. 2, 4

[60] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, 2021. 1, 2, 5, 7

[61] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022. 2

[62] Friedemann Zenke and Tim P Vogels. The remarkable robustness of surrogate gradient learning for instilling complex function in spiking neural networks. *Neural computation*, 33 (4):899–925, 2021. 2, 6

[63] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 695–704, 2018. 5, 7

[64] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *IEEE Tansactions on Circuits and Systems for Video technology*, 30(11):3943–3956, 2019. 5

[65] Hanle Zheng, Yujie Wu, Lei Deng, Yifan Hu, and Guoqi Li. Going deeper with directly-trained larger spiking neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, pages 11062–11070, 2021. 2, 3, 6

[66] Chenlin Zhou, Han Zhang, Zhaokun Zhou, Liutao Yu, Liwei Huang, Xiaopeng Fan, Li Yuan, Zhengyu Ma, Huihui Zhou, and Yonghong Tian. Qkformer: Hierarchical spiking transformer using qk attention. *Advances in Neural Information Processing Systems*, 37:13074–13098, 2024. 3

[67] Zhaokun Zhou, Jun Niu, Yang Zhang, Li Yuan, and Yuesheng Zhu. Spiking transformer with spatial-temporal spiking self-attention. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2025. 3

[68] Hancheng Zhu, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi. Metaiqa: Deep meta-learning for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14143–14152, 2020. 6

[69] Ruoxi Zhu, Zhengzhong Tu, Jiaming Liu, Alan C. Bovik, and Yibo Fan. Mwformer: Multi-weather image restoration using degradation-aware transformers. *IEEE Transactions on Image Processing*, 33:6790–6805, 2024. 2, 6, 7

[70] Yurui Zhu, Tianyu Wang, Xueyang Fu, Xuanyu Yang, Xin Guo, Jifeng Dai, Yu Qiao, and Xiaowei Hu. Learning weather-general and weather-specific features for image restoration under multiple adverse weather conditions. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21747–21758, 2023. 6, 7

[71] Simone Zini and Marco Buzzelli. On the impact of rain over semantic segmentation of street scenes. In *International Conference on Pattern Recognition*, pages 597–610. Springer, 2021. 1

[72] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):10346–10357, 2023. 7