Exploring the Potentials of Spiking Neural Networks for Image Deraining

Shuang Chen[1], Tomas Krajnik[2], Farshad Arvin[1], Amir Atapour-Abarghouei[1*],

[1]Computer Science Department of Durham University, UK
[2]Artificial Intelligence Centre of Czech Technical University, Czechia
{shuang.chen, farshad.arvin, amir.atapour-abarghouei}@durham.ac.uk
tomas.krajnik@fel.cvut.cz

Abstract

Biologically plausible and energy-efficient frameworks such as Spiking Neural Networks (SNNs) have not been sufficiently explored in low-level vision tasks. Taking image deraining as an example, this study addresses the representation of the inherent high-pass characteristics of spiking neurons, specifically in image deraining and innovatively proposes the Visual LIF (VLIF) neuron, overcoming the obstacle of lacking spatial contextual understanding present in traditional spiking neurons. To tackle the limitation of frequency-domain saturation inherent in conventional spiking neurons, we leverage the proposed VLIF to introduce the Spiking Decomposition and Enhancement Module and the lightweight Spiking Multi-scale Unit for hierarchical multi-scale representation learning. Extensive experiments across five benchmark deraining datasets demonstrate that our approach significantly outperforms state-of-the-art SNN-based deraining methods, achieving this superior performance with only 13% of their energy consumption. These findings establish a solid foundation for deploying SNNs in high-performance, energy-efficient low-level vision tasks.
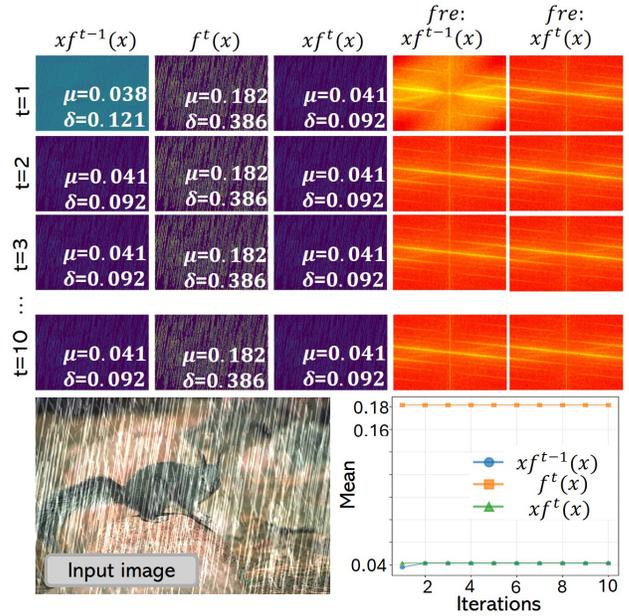
Code — https://github.com/ChrisChen1023/VLIF

Figure 1: LIF $f(\cdot)$ highlights high-frequency rain but exhibits frequency saturation with repeated applications after t=1. x is embedded features, fre is frequency spectrums.

Introduction

Spiking Neural Networks (SNNs) have recently emerged as highly efficient and biologically plausible alternatives to traditional Artificial Neural Networks (ANNs) and Deep Neural Networks (DNNs) with significant improvements in energy efficiency and computational simplicity. Recent studies have demonstrated the promising potential of SNNs across high-level computer vision tasks, including image classification (Deng et al. 2022), object detection (Luo et al. 2024) and semantic segmentation (Patel et al. 2021), with effective sequential information processing and representation.

Given the remarkable efficiency and success in high-level visual tasks, a natural question arises: can SNNs also excel in low-level vision tasks? In this paper, we focus specifically on image deraining, a fundamental low-level vision task. Addressing this challenge requires answering a fundamental **Question** and overcoming a critical **Obstacle**. The

---

*Corresponding Author

Question is: what distinctive features do SNNs represent in the image deraining context? Unlike convolutional layers, which aggregate local information using weighted filters, the conventional spiking neuron, LIF (Leaky Integrate-and-Fire), integrates discrete spikes over time to trigger neuronal firing events (Maass 1997). Such temporal accumulation differs fundamentally from the spatially continuous convolutional operation, potentially leading to distinct feature extraction mechanisms. Although (Song et al. 2024; Su, Wu, and Zheng 2025; Liang et al. 2025; Xiao et al. 2025) introduce SNNs to vision restoration, the existing works have not addressed this differentiation at the feature representation level, a gap particularly crucial for low-level vision tasks.

To answer this fundamental **Question**, we analyse and identify two crucial functional properties inherent in LIF neurons: Intensity-Triggered Activation and High-

Frequency Indication. We conduct extensive experiments and reveal that the discrete, threshold-triggered spikes of LIF neurons inherently emphasise high-frequency components dominated by high-intensity, high-frequency rain artifacts.

Moreover, our investigation reveals another significant representational bottleneck in conventional LIF neurons: repeated applications of the standard LIF mechanism do not further enrich feature representations beyond an initial high-frequency extraction. Motivated by this insight, we propose the Spiking Decomposition and Enhancement Module (SDEM) and a Spiking Multi-scale Unit (SMU). These novel modules are designed to overcome the saturation effect intrinsic to naïve spiking neurons by systematically decomposing input features into complementary high- and low-frequency components. This refines features across multiple spatial scales for a better hierarchical representation of fine-grained rain structures.

The primary **O**bstacle lies in the inherently pixel-level receptive field of the spiking neuron, which severely limits its capability to capture spatial dependencies necessary for representation-based visual tasks. For instance, in image deraining, accurately identifying and removing rain streaks requires contextual understanding beyond pixel-level perception. Previous attempts to enhance spiking neurons' expressivity through architectural modifications achieve limited success, as they failed to provide spiking neurons with true visual perception capabilities. Other efforts introduced convolutional operations to expand the receptive field in SNNs. Unfortunately, this approach fundamentally contradicts the intrinsic efficiency of SNNs by significantly increasing parameters and computational overhead.

To overcome the identified **O**bstacle, we draw inspiration from neurophysiological findings showing that neurons in the inferotemporal (IT) cortex of the ventral visual pathway integrate stimuli across spatially extended receptive fields to recognise complex object features (Kobatake and Tanaka 1994). Motivated by this mechanism, we leverage the cumulative firing behaviour of Leaky Integrate-and-Fire (LIF) neurons to provide visual awareness. However, due to the intrinsic firing-resetting mechanism of LIF neurons, such binary accumulation alone fails to represent rich visual features. To address this, we propose the Visual LIF (VLIF), combining the above insights with Integrate-and-Fire dynamics enhanced by continuous membrane potentials, thus equipping LIF neurons with spatially meaningful receptive fields without introducing additional parameters.

Extensive evaluations across four benchmark deraining datasets confirm the effectiveness and efficiency of our approach. Our proposed method outperforms existing SNN-based models and achieves competitive or superior results compared to state-of-the-art CNN and transformer-based methods, all while significantly reducing computational overhead and energy consumption.

Our primary contributions are summarised as follows:

- We bridge the gap between SNNs and low-level vision tasks by identifying and characterising the distinctive representational capabilities of Spiking neurons in the context of image deraining.

- We provide visual perception capabilities to LIF neurons and explicitly elucidate the feature representation mechanism of conventional LIF neurons in image deraining.

- We propose the novel spiking neuron Visual LIF (VLIF), to incorporate local visual context through spatial aggregation and continuous membrane potentials, overcoming the spatial insensitivity and representational limitations of conventional LIF neurons.

- Leveraging the VLIF neuron, we introduce a novel image deraining framework comprising the proposed Spiking Decomposition and Enhancement Module (SDEM) and the Spiking Multi-scale Unit (SMU). This framework significantly surpasses the performance of the best existing SNN-based deraining models while consuming only 13% of their energy.

Related Work

Image Deraining

Single image deraining has rapidly advanced with the application of deep learning, where initial convolutional neural network (CNN)-based methods typically approached deraining as pixel-wise regression tasks (Jiang et al. 2020; Zamir et al. 2021). More recent methods have employed multi-stage or multi-scale CNN architectures and self-attention mechanisms to improve rain removal performance by capturing intricate patterns and modelling long-range dependencies (Zamir et al. 2022). Despite their effectiveness, these CNN and attention-based approaches often incur heavy computational costs, limiting deployment on resource-constrained devices. Very recently, a different paradigm was explored with spiking neural networks (Song et al. 2024). The SNN-based approach opens a new direction for efficient rain removal on low-power devices.

Spiking Neural Network

Spiking Neural Networks (SNNs) offer energy-efficient computation through sparse, event-driven processing, making them promising alternatives to conventional ANNs. SNNs are typically implemented via ANN-to-SNN conversion or direct training. The former maps a trained ANN to an SNN by replacing activations with spiking units (Cao, Chen, and Khosla 2015; Diehl et al. 2015), but often requires long simulation time and suffers from performance degradation. In contrast, direct training with surrogate gradients enables end-to-end optimisation and achieves competitive accuracy with much fewer timesteps and lower latency (Neftci, Mostafa, and Zenke 2019). In computer vision, SNNs have been applied to high-level tasks such as image classification (Hu et al. 2021), object detection (Luo et al. 2024) and segmentation (Patel et al. 2021). However, their use in pixel-level low-level tasks (such as image deraining) remains rare. We explore this under-investigated area and demonstrate that SNNs can be effectively adapted for dense image regression.

Method

In image deraining, the key challenge lies in accurately identifying and separating high-frequency rain structures from
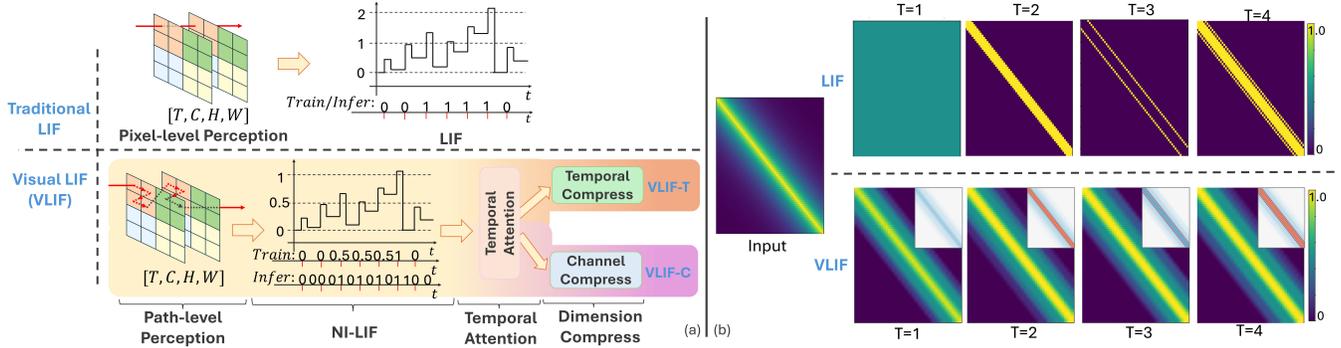
Figure 2: The comparison of LIF and VLIF. (a) shows differences in pipeline, (b) shows differences response areas, in the top-right insets, blue and orange denote the response for LIF and VLIF, respectively.
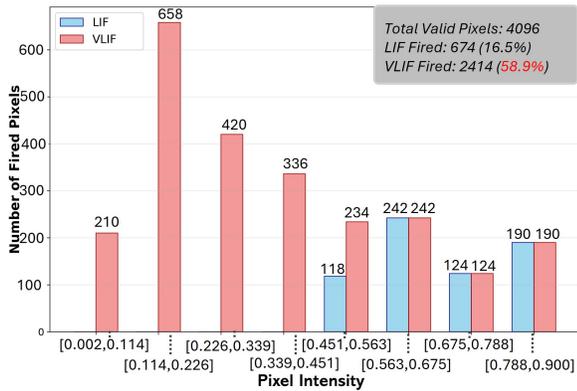

Figure 3: Activation comparison. VLIF activates $3.58\times$ more than LIF by incorporating local spatial context, enabling better feature representation in low-response regions.

degraded images. However, directly applying SNNs to this task suffers from limited visual awareness due to the binary and sequential firing behaviour of standard LIF neurons.

Through empirical analysis, we identify two functional properties of LIF neurons relevant to deraining: a) Intensity-Triggered Activation and b) High-Frequency Indication. Building on this, we derive two fundamental theoretical properties: a) high-frequency selection and b) frequency saturation, which expose a representational bottleneck in standard LIF neurons. These findings motivate our framework, which addresses these limitation through three core components: 1) the Vision-LIF (VLIF) neuron with spatial awareness, (2) the Spiking Decomposition & Enhancement Module (SDEM) for feature disentanglement and refinement and (3) the Spiking Multi-scale Unit (SMU) for efficient multiscale representation learning at high-resolution stages.

Spiking Neurons for Image Deraining?

We begin by introducing our key finding: In image deraining, LIF neurons act as high-pass indicators that emphasise high-frequency rain structures. Notably, LIF neurons do not inherently possess high-pass filtering capabilities. Their ef-

fectiveness arises from the close alignment between their firing behaviour and the core objective of the deraining task, which is to identify rain structures that typically manifest as high-frequency noise superimposed on the image through residual representation learning.

As the rain streaks dominate this discrepancy between clear and rainy images, these regions contribute more significantly to the gradient during backpropagation, leading the network to amplify their intensity in the learned feature representations. The membrane potential dynamics of LIF neurons are inherently sensitive to such localised, high-intensity responses. When the feature intensity at a given spatial location surpasses the neuron's firing threshold θ, a spike is triggered. Consequently, the sparse binary output of the spiking neuron effectively serves as a spatial mask that highlights regions corresponding to high-frequency rain patterns.

Based on this behaviour, we define the spiking neuron as an Intensity-Triggered High-Frequency Indicator in the context of image deraining. This role is characterised by two essential functional properties: 1) **Intensity-Triggered Activation**: The neuron's output depends solely on whether local feature intensities exceed a predefined threshold. 2) **High-Frequency Indication**: The neuron's binary outputs mark spatial locations where intensity surpasses the threshold, which typically aligns with high-frequency artefacts such as rain streaks. We formally define the concept of a high-frequency indicator as follows: Given a feature representation, x, and a LIF neuron $f(\cdot)$ applied t times, denoted as $f^t(x)$, the neuron acts as a high-frequency indicator if it satisfies the following condition:

$$||x \odot f^t(x)||_2 = \begin{cases} ||HC[x]||_2 = C, & \text{if } t = 1, \\ ||x \odot f^{t-1}(x)||_2 = C, & \text{if } t > 1, \end{cases} \quad (1)$$

where $\odot$ denotes the Hadamard product and $HC[x]$ denotes the high-frequency components in the feature space of x.

This formulation reveals two critical theoretical properties: 1) **High-Frequency Highlighting**: At the first application (t=1), the operator $f(\cdot)$ identifies spatial and temporal positions of high-frequency components through intensity thresholding. The resulting sparse binary mask $f(x)$, when applied to x via element-wise multiplication, enhances
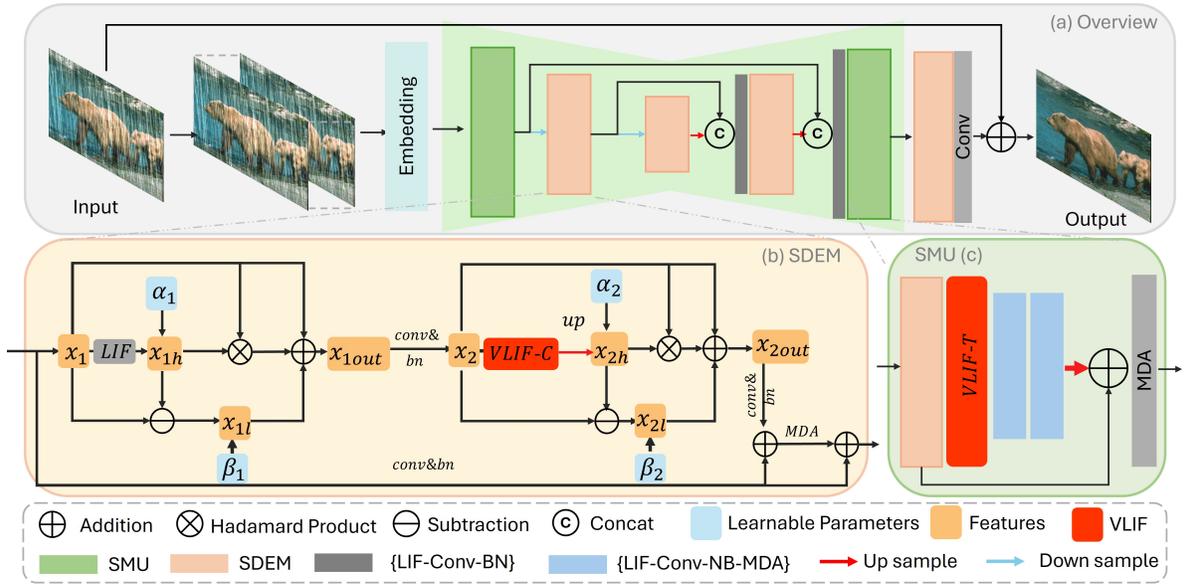
Figure 4: (a) Architecture; (b) Spiking Decomposition & Enhancement Module (SDEM), (c) Spiking Multi-scale Unit (SMU).

the visibility of high-frequency rain artefacts. Validated in Fig. 1, for $t = 1$, compared to $fre : xf^{t-1}(x)$, $fre : xf^{t}(x)$ shows a clear reduction in low-frequency energy. 2) **Frequency Saturation**: For subsequent applications ($t>1$), repeated use of $f(\cdot)$ does not produce a more refined or sparser high-frequency representation. The extracted high-frequency content remains equivalent to that obtained at $t = 1$. In Fig 1, μ is fixed after t=1. This property is mathematically reflected in the invariance for all $t > 1$.

Such properties reveal a fundamental limitation of the standard LIF: repeated applications of the same neuronal mechanism do not enrich the feature representation beyond the initial extraction of high-frequency information. Inspired by this observation, we propose novel modules VLIF, SDEM and SMU, to leverage more diverse feature representations across multiple scales, thereby overcoming the saturation effect and enabling a more effective architecture for image deraining. Next, we introduce the proposed Vision-LIF.

Vision-LIF: Spiking Neuron with Visual Awareness

As discussed above, LIF neurons serve as effective high-frequency indicators, producing sparse binary activations for strong rain artefacts. However, the spatial insensitivity is a drawback. The firing decision at each pixel is made independently and relies solely on the pixel-level intensity. Such a mechanism fails to account for spatially correlated but low-intensity regions (such as smooth peripheries of rain patterns), which are unlikely to be strong enough to trigger the firing threshold. As a result, these semantically meaningful yet low-contrast components are ignored, leading to incomplete structural representations. To solve this drawback, we introduce **Vision-LIF (VLIF)**, a modified spiking neuron with local visual awareness. Unlike conventional LIF neurons, VLIF incorporates spatial context into the spike gener-

ation process by aggregating feature intensities from neighbourhoods of pixels. The comparison is shown in Fig 2.

Patch-to-Time Reorganisation. Given the input feature $X \in \mathbb{R}^{T \times C \times H \times W}$, we first apply a *pixel unshuffle* (Shi et al. 2016) with downsampling factor r, downscale X to $X' \in \mathbb{R}^{T \times C \cdot r^2 \times \frac{H}{r} \times \frac{W}{r}}$, which splits X into r^2 non-overlapping spatial patches. Next, we reshape and permute the tensor by bringing the patch dimension into the temporal axis to have $X'' \in \mathbb{R}^{T \cdot r^2 \times C \times \frac{H}{r} \times \frac{W}{r}}$. This transformation effectively converts local spatial patches into temporally ordered sequences, which allows the neuron to integrate local visual context over the temporal axis. Therefore, low-intensity pixels within high-response regions are more likely to fire, thus enhancing spatial continuity in rain structure encoding:

$$F(h, w) = \sum_{t=1}^{T \cdot r^2} U_t(h, w), \qquad (2)$$

where $F(h, w)$ is the fused output at spatial location (h,w), $U_t(h, w)$ is the membrane potential at timestep t. This accumulation explicitly incorporates local spatial context into the firing dynamics of VLIF, which will allow even low-intensity pixels, especially those within high-response neighbourhoods, to surpass the firing threshold. In contrast to conventional LIF neurons, which fire strictly based on individual pixel intensity, VLIF facilitates more context-aware activation of semantically relevant but low-contrast features.

Spiking and Stability Challenge While VLIF fires based on spatial context, the spatial accumulation carries greater structural variability and finer-grained information. In such cases, binary spikes are insufficient to express continuous intensity differences. To address this, we consider I-LIF (Luo et al. 2024), an integer-valued spiking neuron that generates discrete spike counts for better representation. However,

	Methods	Rain12		Rain200L		Rain200H		Rain1200		Params↓	FLOPs↓	Energy↓
		PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	(M)	(G)	$(\times 10^4 \mu J)$
C	PReNet	36.32	0.9593	36.06	0.9735	27.18	0.8698	33.39	0.9280	0.168	66.249	82.81
	Effderain	36.12	0.9588	36.06	0.9731	26.11	0.8341	32.85	0.9147	27.654	52.915	66.14
T	Restormer	37.33	0.9672	36.62	0.9731	27.93	0.8823	31.56	0.9009	26.10	12.33	11362
	SmartAssign	36.87	0.9618	38.41	0.9814	27.71	0.8536	33.11	0.9154	1.359	90.386	112.9
S	ESDNet	36.38	0.9610	38.34	0.9845	28.39	**0.8944**	**32.78**	0.9126	0.165	7.320	9.150
	Ours	**36.70**	**0.9649**	**38.46**	**0.9847**	**28.63**	0.8935	32.74	**0.9184**	**0.088**	**2.794**	**1.166**

Table 1: Quantitative results on four synthetic benchmarks. **C**, **T**, and **S** represent CNN-based, Transformer-based, and SNN-based methods, respectively. **Bold** denote the best score. Our method achieves consistently superior performance compared to Restormer, a representative and widely adopted Transformer-based baseline, while substantially reducing computational cost.

similar to the observation in Spike2Former (Lei et al. 2025), I-LIF leads to training instability. In our task, this manifests as integer overflow in the output images. We argue that the potential reason is that the large integer values accumulated by I-LIF lead to corrupted or saturated reconstruction results, which fail to represent finer-grained features.

To mitigate this, we adopt virtual timestep normalisation (Lei et al. 2025), to bound the spike values and stabilise gradient flow, named NI-LIF, by scaling spike magnitudes with D and apply clipping to bound x to $[x, min, max]$:

$$U[t] = H[t-1] + X[t], \qquad (3)$$

$$S[t] = \frac{\text{Clip}(\text{round}(U[t]), \ 0, \ D)}{D}, \qquad (4)$$

$$H[t] = \beta \cdot (U[t] - S[t] \cdot D), \qquad (5)$$

where $U[t]$ is the membrane potential, $S[t]$ is the normalised spike output, $H[t]$ is the updated state, D is the virtual timestep scaling constant, and β is the leakage factor.

Following the spiking stage, we apply a temporal attention (Yao et al. 2023) to obtain the attended spike features with the shape $[r^2 T, C, H/r, W/r]$. We denote this output as S_{attn}, which is subsequently used for output compression.

Output Compression. To compact the additional dimension, we introduce temporal- and channel-based compression to fuse global contextual information. For *Temporal Compression (VLIF-T)*, we apply a $r^2 \times 1 \times 1$ 3D convolution to aggregate information across the temporal dimension:

$$S_{\text{TC}} = \text{Conv3D}_{r^2 \times 1 \times 1}(S_{\text{attn}}). \qquad (6)$$

For *Channel Compression (VLIF-C)*, we flatten the temporal axis and apply a lightweight MLP. This enables flexible channel-wise fusion and allows the model to learn channel correlations across the aggregated temporal features:

$$S_{\text{CC}} = \text{MLP}(\text{Flatten}(S_{\text{attn}})). \qquad (7)$$

Both variants produce a compressed representation with reduced temporal dimensionality, which is then passed to subsequent network stages.

Analysis of more Fired Pixels

As shown in Fig. 2b, we compare VLIF with traditional LIF using a synthetic decay matrix, where diagonal elements have the intensity of 0.9 and off-diagonal elements decay exponentially to simulate the rainy structure.

(Ding et al. 2025) finds that the output of the vanilla LIF is poorly informative at the first timestep and only achieves decent performance after integrating subsequent temporal steps, as the LIF with $t=1$ suffers from strong membrane potential suppression due to the zero-initialised hidden state. As a result, input features may fail to exceed the firing threshold at $t=1$ (Fig. 2b), even with relatively high intensity. This behaviour limits LIF to represent fine-grained spatial details, which is critical in pixel-level restoration tasks such as image deraining. In contrast, VLIF normalises spike outputs and encourages early activation, yielding significantly denser responses. As shown in Fig. 3, LIF achieves only 16.46% activation, while VLIF reaches 58.94% activation (3.6× improvement) and maintains 25.4% fire rate even for the lowest input range [0.002, 0.114], where LIF is unable to fire. This showcases VLIF's strength in preserving weak signals and reconstructing fine details through spatio-temporal reorganisation and multi-level quantisation.

Spiking Decomposition & Enhancement Module

To further enhance the model's capacity to detect fine-grained rain structures, we propose the Spiking Residual Block (SDEM), which systematically decomposes input features into high- and low-frequency components via spiking neurons. As shown in Fig. 4(b), SDEM comprises two cascaded stages. Each stage follows an identical processing pipeline but differs in the spiking neuron utilised, LIF for the first stage to detect sharp and strong high-frequency response, while VLIF-C for the second stage for context-aware detection of subtle or blurred rain structures. This progressive design facilitates coarse-to-fine enhancement of rain features. Next, we only detail Stage 2 for simplicity.

Stage-wise Feature Decomposition. Let the input to the second stage be denoted as $X \in \mathbb{R}^{T \times C \times H \times W}$. We begin by applying the VLIF-C on X, after employing a bilinear interpolation for upsampling, we get a high-frequency indicator X_h scaled by a learnable parameter α, then subtract X_h to obtain the complementary low-frequency representation X_l:

$$X_h = \alpha \cdot \text{VLIF-C}(X), \quad X_l = X - X_h, \qquad (8)$$

where X_h serves as a structural high-pass mask and X_l retains the low-frequency component. Additionally, we use X_h to extract the high-frequency feature X'_h. A learnable scalar β_2 is adopted to scale the low-frequency path:

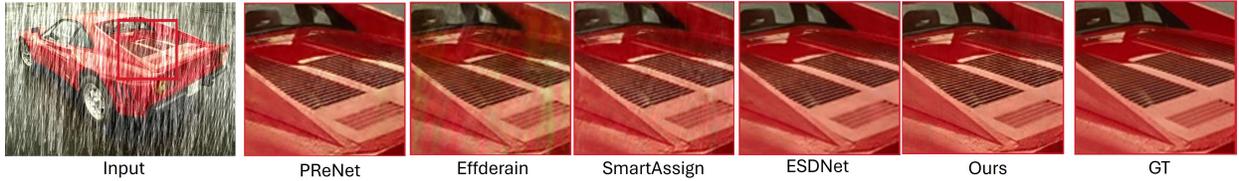$$X'_h = X_h \cdot X, \quad X'_l = \beta_2 X_l. \qquad (9)$$

Figure 5: Qualitative comparison on the Rain200H. Our model generates finer-grained details and clearer textures.

Model	Effderain	Restormer	ESDNet	Ours
PIQE↓	63.56	32.58	26.10	24.26
BRISQUE↓	75.28	28.55	22.75	20.99

(a) Quantitative result for Real-World data.

Model	T-EMB	SMU	SDEM	REFINE	PSNR↑	SSIM↑
(a)					26.51	0.8612
(b)	✓				26.59	0.8639
(c)	✓		✓		26.96	0.8657
(d)	✓	✓	✓		27.11	0.8778
Ours	✓	✓	✓	✓	**27.21**	**0.8799**

(b) Ablation study on architectural components.

Model	LIF	I-LIF	VLIF-C	VLIF-T	Ours
PSNR↑	21.8362	26.19	23.10	26.19	27.21
SSIM↑	0.6722	0.8595	0.7719	0.8755	0.8799

(c) Ablation study on different LIF variants.

Metric	TACL	UIE-WD	SFGNet	UIR-PK	Ours
PSNR↑	20.99	17.80	22.68	**26.42**	25.69
SSIM↑	0.782	0.760	0.585	0.866	**0.8817**
Params (M)↓	28.29	14.46	1.30	1.84	**0.088**
FLOPs (G)↓	240.06	102.76	163.16	27.34	**2.7938**

(d) Quantitative comparison on EUVP.

Table 2: (a) Quantitative real-world results, (b) ablation study, (c) impact of LIF variants, (d) UIE task performance.

We then aggregate all components with a residual connection, followed by a convolution and batch normalisation (Zheng et al. 2021) to refine the features:

$$X_{\text{refined}} = \text{BN}(\text{Conv}(X'_h + X'_l + X)). \quad (10)$$

Residual and Attention Fusion. We include a residual shortcut path to further preserve the input feature distribution and ensure gradient stability:

$$X_{\text{shortcut}} = \text{BN}(\text{Conv}(X)), \quad (11)$$

$$X'_{\text{refined}} = X_{\text{refined}} + X_{\text{shortcut}}. \quad (12)$$

Additionally, a multi-dimensional attention module (MDA) (Yao et al. 2023) enhances spatio-temporal and channel interactions. Then the final SDEM output is given:

$$X_{\text{MDA}} = \text{MDA}(X_{\text{refined}}), \quad (13)$$

$$X_{\text{SDEM}} = X_{\text{refined}'} + X_{\text{shortcut}} + X_{\text{MDA}}. \quad (14)$$

Spiking Multi-scale Unit (SMU)

Inspired by the fact that multi-scale modelling at the shallow level can enhance representational capacity (Cui et al. 2023),

we introduce a lightweight multi-scale module, **SMU**, which leverages the proposed SDEM and a VLIF-T, to incorporate fine-grained rain cues from both local and extended neighbourhoods with minimal computational overhead, offering an efficient multi-scale enhancement mechanism in highest-resolution stages of both encoder and decoder.

As illustrated in Fig. 4(c), SMU first receives input feature $x \in \mathbb{R}^{T \times C \times H \times W}$. x then is processed by the proposed SDEM, followed by a VLIF-T neuron that further downsamples the spatial resolution to $[H/r, W/r]$ via patch-to-time reorganisation. The downsampled feature is then passed through two stacked Spiking Units (SU), each composed of a LIF spiking neuron, batch normalisation, convolution, and a multi-dimensional attention (MDA) module (Yao et al. 2023). Two SUs further extract high-frequency representations at lower spatial resolutions. The features are subsequently upsampled back to $[T, C, H, W]$ using bilinear interpolation. To retain shallow spatial information, a residual connection bypasses intermediate layers, with merged features refined by an MDA block before the next step.

Experiments

This section presents datasets, metrics, our main experimental results on draining, ablation studies, and a generalisation study on another low-level vision task, Underwater Image Enhancement (UIE). The implementation details, training setting, and hyperparameter tuning (e.g. time step, number of blocks) are provided in supplementary materials.

Datasets and Metrics

To comprehensively evaluate the performance, we conduct experiments on five widely used synthetic rain datasets (Rain12 (Li et al. 2016), Rain200L (Yang et al. 2017), Rain200H (Yang et al. 2017), Rain1200 (Zhang and Patel 2018)) and the Real-World rain dataset (Ren et al. 2019). All datasets are used following the original train/test splits (shown in the supplementary material with training iterations). We evaluate model performance for synthetic data on two widely used metrics, Peak Signal to Noise Ratio (PSNR) (Hore and Ziou 2010) and Structural Similarity (SSIM) (Wang et al. 2004), shown in Tab. 1 (reproduced from (Song et al. 2024)). We retrain ESDNet based on their official setting, as their pre-trained models are inaccessible. Perception-based Image Quality Evaluator (PIQE) (Venkatanath et al. 2015) and Blind/Referenceless Image Spatial QUality Evaluator BRISQUE (Mittal, Moorthy, and Bovik 2012) are used for the real-world data, shown in Tab. 2a. Following (Hu, Tang, and Pan 2021; Song et al.

2024), energy is calculated with: 12.5pJ per FLOP, 77fJ per SOP (synaptic operation), and 3.7pJ per spike-based Sign operation (calculated using energy per spike). All ablation studies are conducted on RainH200 for 30K iterations.

Experiment Results

In Tab. 1, our method fully exploits the potential of spiking neurons for image deraining. Compared with the latest SNN-based method ESDNet, our approach achieves superior performance while using only **53%** of its parameters (0.088M vs. 0.165M), **38%** of its FLOPs (2.79G vs. 7.32G), and **13%** of its energy consumption ($1.17 \times 10^4\ \mu$J vs. $9.15 \times 10^4\ \mu$J). Moreover, our model further narrows the performance gap between SNNs and strong ANN-based baselines. Qualitative results in Fig. 5 show that our method generates finer-grained and more structurally coherent outputs compared to ANN models such as Effderain (Guo et al. 2021) and SmartAssign (Wang, Ma, and Liu 2023), while requiring only **0.3% / 6.5%** of their parameter counts (0.088M vs. 27.65M / 1.36M) and **1.8% / 1.0%** of their energy consumption ($1.17 \times 10^4\ \mu$J vs. $6.61 \times 10^5\ \mu$J / $1.13 \times 10^6\ \mu$J). Lower PIQE and BRISQUE from Tab. 2a and comparison from Fig. 6a underscore our robustness in real-world scenarios.

Ablation Study

Effectiveness of Each Components. As shown in Tab. 2b, model (a) is the baseline, where the spiking neuron degrades to a vanilla LIF within a standard U-Net structure. Adding Temporal EMBedding (b) brings slight improvement. Model (c) further introduces SDEM, which results in a significant PSNR improvement. This shows effectiveness to extract and enhance fine-grained rain features. Comparing (c) and (d), the inclusion of the SMU notably improves SSIM, indicating better preservation of spatial structure. Finally, the refinement block composed of four stacked SDEM modules provides further gains to achieve the best overall performance.

Effectiveness of VLIF. We evaluate different spiking variants by replacing the VLIF neurons in our full model with LIF, I-LIF, VLIF-C, and VLIF-T. Table 2c shows that, stacking standard LIF leads to significant degradation (19.7%↓ in PSNR and 23.6%↓ in SSIM) compared to the full model. This is because repeated use of basic LIF provides limited features, as it suffers from *frequency saturation*. The drop observed with VLIF-C indicates that channel-only compression is insufficient for handling extended sequences. While I-LIF, VLIF-T, and our hybrid design achieve comparable quantitative scores, qualitative results in Fig. 6b show that our full model produces cleaner and more detailed outputs, benefiting from joint temporal and channel compression.

Generalisability

To further validate the task-oriented behaviour and generalisability of spiking neurons in image restoration, we train our model on the EUVP (Islam, Xia, and Sattar 2020) Underwater Image Enhancement (UIE) dataset, where the goal is to correct large-scale low-frequency degradation from light scattering and colour shifts. As shown in Tab. 2d, our method surpasses recent UIE approaches (Liu et al. 2022;
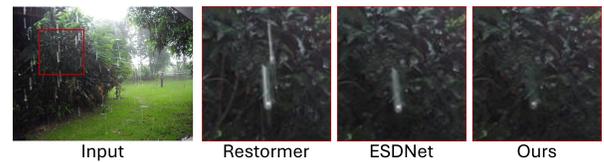
(a) Real-world rain image.
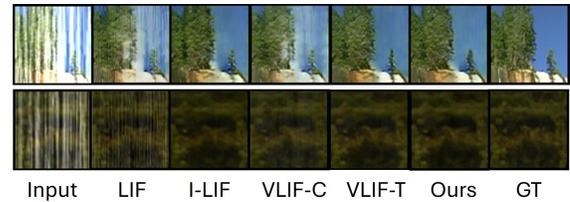
(b) Spiking neuron variant results.
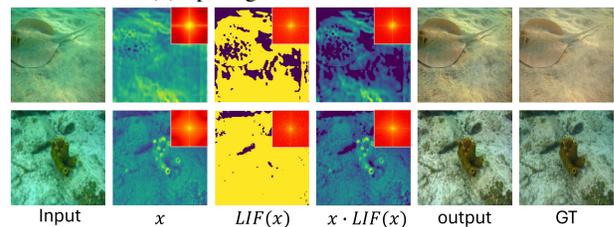
(c) Visualisation of UIE task. Left to right: input image, feature before LIF, binary spike map, LIF-highlighted feature, our result, and ground truth. Top right corners show the frequency spectrums.

Figure 6: (a) Real-world rainy image; (b) Spiking neuron variants; (c) Visualisation on the UIE task.

Ma and Oh 2022; Zhao et al. 2024) while using the least computation. Consistent with our frequency-domain analysis (Fig. 6c), LIF neurons act as low-frequency indicators in UIE, highlighting the task-adaptive nature of spiking neurons in image restoration.

Conclusion

This work demonstrates the potential of Spiking Neural Networks for single-image deraining. We answer the key **Q**uestion of how spiking neurons represent rain artefacts, showing that in deraining they behave as task-driven, intensity-triggered high-frequency indicators. To overcome the **O**bstacle of spatial insensitivity in conventional LIF neurons, we propose Visual LIF (VLIF) to enable spatial awareness. In addition, our Spiking Decomposition & Enhancement Module and Spiking Multi-scale Unit provide hierarchical, multi-scale representations of rain patterns.

Experimentally, we beat existing SNN-based approaches and achieves performance comparable to ESDNet while using only 13% of its energy. Nonetheless, a performance gap between SNNs and ANNs remains, largely due to differences in representational richness and optimisation maturity. Closing this gap is an important direction for future work.

References

Cao, Y.; Chen, Y.; and Khosla, D. 2015. Spiking deep convolutional neural networks for energy-efficient object recognition. *IJCV*, 113: 54–66.

Cui, Y.; Ren, W.; Cao, X.; and Knoll, A. 2023. Image restoration via frequency selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(2): 1093–1108.

Deng, S.; Li, Y.; Zhang, S.; and Gu, S. 2022. Temporal efficient training of spiking neural network via gradient reweighting. *arXiv preprint arXiv:2202.11946*.

Diehl, P. U.; Neil, D.; Binas, J.; Cook, M.; Liu, S.-C.; and Pfeiffer, M. 2015. Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. In *IJCNN*, 1–8.

Ding, Y.; Zuo, L.; Jing, M.; He, P.; and Deng, H. 2025. Rethinking Spiking Neural Networks from an Ensemble Learning Perspective. In *The Thirteenth International Conference on Learning Representations*.

Guo, Q.; Sun, J.; Juefei-Xu, F.; Ma, L.; Xie, X.; Feng, W.; Liu, Y.; and Zhao, J. 2021. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. In *AAAI*, volume 35, 1487–1495.

Hore, A.; and Ziou, D. 2010. Image quality metrics: PSNR vs. SSIM. In *IEEE ICIP*, 2366–2369. IEEE.

Hu, Y.; Tang, H.; and Pan, G. 2021. Spiking deep residual networks. *IEEE TNNLS*.

Hu, Y.; Wu, Y.; Deng, L.; and Li, G. 2021. Advancing residual learning towards powerful deep spiking neural networks. *arXiv e-prints*, arXiv–2112.

Islam, M. J.; Xia, Y.; and Sattar, J. 2020. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2): 3227–3234.

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

Kobatake, E.; and Tanaka, K. 1994. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of neurophysiology*, 71(3): 856–867.

Lei, Z.; Yao, M.; Hu, J.; Luo, X.; Lu, Y.; Xu, B.; and Li, G. 2025. Spike2former: Efficient spiking transformer for high-performance image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 1364–1372.

Li, Y.; Tan, R. T.; Guo, X.; Lu, J.; and Brown, M. S. 2016. Rain streak removal using layer priors. In *ICCV*, 2736–2744.

Liang, H.; Zhong, X.; Liu, W.; Zheng, Y.; Huang, W.; Yu, Z.; and Huang, T. 2025. SpikeDerain: Unveiling Clear Videos from Rainy Sequences Using Color Spike Streams. *arXiv preprint arXiv:2503.20315*.

Liu, R.; Jiang, Z.; Yang, S.; and Fan, X. 2022. Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Transactions on Image Processing*, 31: 4922–4936.

Luo, X.; Yao, M.; Chou, Y.; Xu, B.; and Li, G. 2024. Integer-valued training and spike-driven inference spiking neural network for high-performance and energy-efficient object detection. In *European Conference on Computer Vision*, 253–272. Springer.

Ma, Z.; and Oh, C. 2022. A wavelet-based dual-stream network for underwater image enhancement. In *ICASSP*, 2769–2773.

Maass, W. 1997. Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9): 1659–1671.

Mittal, A.; Moorthy, A. K.; and Bovik, A. C. 2012. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12): 4695–4708.

Neftci, E. O.; Mostafa, H.; and Zenke, F. 2019. 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.

Patel, K.; Hunsberger, E.; Batir, S.; and Eliasmith, C. 2021. A spiking neural network for image segmentation. *arXiv preprint arXiv:2106.08921*.

Ren, D.; Zuo, W.; Hu, Q.; Zhu, P.; and Meng, D. 2019. Progressive image deraining networks: A better and simpler baseline. In *CVPR*, 3937–3946.

Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A. P.; Bishop, R.; Rueckert, D.; and Wang, Z. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1874–1883.

Song, T.; Jin, G.; Li, P.; Jiang, K.; Chen, X.; and Jin, J. 2024. Learning a Spiking Neural Network for Efficient Image Deraining. In *IJCAI*.

Su, X.; Wu, C.; and Zheng, Z. 2025. Bridge the Gap between SNN and ANN for Image Restoration. *arXiv preprint arXiv:2504.01755*.

Venkatanath, N.; Praneeth, D.; Bh, M. C.; Channappayya, S. S.; and Medasani, S. S. 2015. Blind image quality evaluation using perception based features. In *NCC*, 1–6.

Wang, Y.; Ma, C.; and Liu, J. 2023. SmartAssign: Learning a Smart Knowledge Assignment Strategy for Deraining and Desnowing. In *CVPR*, 3677–3686.

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

Xiao, Y.; Yuan, Q.; Jiang, K.; Huang, W.; Zhang, Q.; Zheng, T.; Lin, C.-W.; and Zhang, L. 2025. Spiking Meets Attention: Efficient Remote Sensing Image Super-Resolution with Attention Spiking Neural Networks. *arXiv preprint arXiv:2503.04223*.

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

Yao, M.; Zhao, G.; Zhang, H.; Hu, Y.; Deng, L.; Tian, Y.; Xu, B.; and Li, G. 2023. Attention spiking neural networks. *IEEE TPAMI*.

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

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

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

Zhao, C.; Cai, W.; Dong, C.; and Zeng, Z. 2024. Toward sufficient spatial-frequency interaction for gradient-aware underwater image enhancement. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3220–3224.

Zheng, H.; Wu, Y.; Deng, L.; Hu, Y.; and Li, G. 2021. Going deeper with directly-trained larger spiking neural networks. In *AAAI*, volume 35, 11062–11070.