# UIESNN: A Scale-Aware Spiking Network for Underwater Image Enhancement

Shuang Chen*, Ruochen Li*, Zihan Zhu‡, Ronald Thenius†, Farshad Arvin*, Amir Atapour-Abarghouei*
*Computer Science Department, Durham University, UK
{shuang.chen, ruochen.li, farshad.arvin, amir.atapour-abarghouei}@durham.ac.uk
†Institute of Biology, University of Graz, Graz, Austria. ronald.thenius@gmail.com
‡University of Cambridge zz566@cam.ac.uk

***Abstract*—Underwater image enhancement (UIE) is a practically important yet underexplored application of spiking neural networks (SNNs), where the dominant degradations are large-scale and low-frequency, such as wavelength-dependent colour casts and scattering-induced veiling. Existing SNN restoration designs rely on locally bounded spiking perception, which can limit global correction and lead to saturated or inconsistent representations. To address these challenges, we propose a scale-aware SNN framework for UIE named UIESNN. At its core is a Multi-scale Pooling LIF Block (MPLB) that injects hierarchical multi-scale pooling responses into membrane dynamics, thereby enlarging the effective receptive field while preserving fine-grained details and inducing heterogeneous scale-dependent activations. Building on MPLB, we design a spiking residual architecture that integrates frequency decomposition and attention-based refinement in a fully spike-driven pipeline. Extensive experiments on the EUVP and LSUI benchmarks demonstrate that UIESNN achieves state-of-the-art performance among SNN-based methods, delivering improved colour fidelity and spatial coherence with competitive energy cost.**

***Index Terms*—Underwater Image Enhancement, Spiking Neural Network.**

## I. Introduction

Spiking Neural Networks (SNNs) are widely regarded as energy-efficient and biologically plausible alternatives to conventional deep neural networks. In particular, the Leaky Integrate-and-Fire (LIF) spiking neuron [1], [2] integrates input evidence over time and emits discrete spikes, which leads to distinctive spatio-temporal processing behaviours that differ fundamentally from convolutional layers that aggregate spatial context via learned filters. Although a performance gap between SNNs and Artificial Neural Networks (ANNs) often still exists, SNNs have achieved impressive results in high-level vision tasks [3], [4]. More recently, they have also been explored for low-level image restoration, such as static image deraining, which is attractive for extreme scenarios such as autonomous driving where cameras must operate under rain [5]. However, in another extreme setting, deep-sea exploration, underwater image enhancement (UIE) remains largely underexplored in the SNN literature.

In single-image deraining, the primary challenge is to remove thin, high-frequency rain streaks, whereas UIE focuses on correcting large-scale, low-frequency degradations such as colour casts and haze-like blur caused by light scattering (shown in Fig. 1). This spectral discrepancy implies that

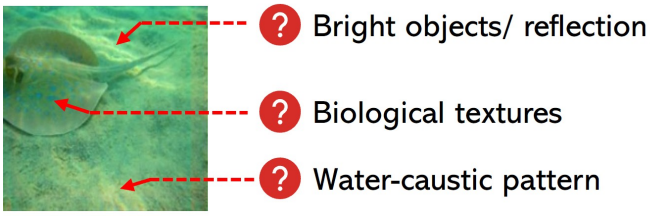


Fig. 1. Visualisation of complex degradation in the underwater scenario.

neuron and architecture designs effective for deraining may not generalise to UIE. As shown in [6], LIF neurons exhibit task-driven frequency selectivity and behave as high-frequency indicators in deraining. In contrast, UIE requires modelling spatially extensive, slowly varying distortions whose cues are distributed over broad regions. Consequently, a limited, locally bounded receptive field can degrade the ability of spiking neurons to capture the global low-frequency structure underlying underwater degradations. This motivates us to explicitly expand the receptive field in a scale-aware manner.

Another major obstacle lies in the limited spatial perception of conventional spiking neurons. Standard LIF units make firing decisions in a largely point-wise manner, integrating inputs over time but with no explicit mechanism to aggregate wider spatial context. Such spatial insensitivity is particularly restrictive for dense enhancement tasks like UIE, where the degradation cues (e.g., veiling light and global color bias) are distributed over neighborhoods and even image-wide regions rather than isolated pixels. Moreover, as observed in [6], naïvely stacking LIF neurons may lead to frequency-domain saturation, where repeated applications of the same spiking mechanism fail to further enrich the underlying representation. For underwater enhancement, this poses an additional challenge: correcting spatially extensive, low-frequency distortions requires both long-range contextual awareness and the ability to refine representations progressively across scales. These considerations raise a key question: how can we endow spiking neurons with a larger receptive field while preserving fine-grained details, and simultaneously encourage heterogeneous representations that can capture underwater degradations across different spatial extents and frequency bands?

To address this challenge, we propose **Multi-scale Pooling LIF Block (MPLB)** to expand the receptive field through *hierarchical multi-scale pooling*. Instead of relying on a single

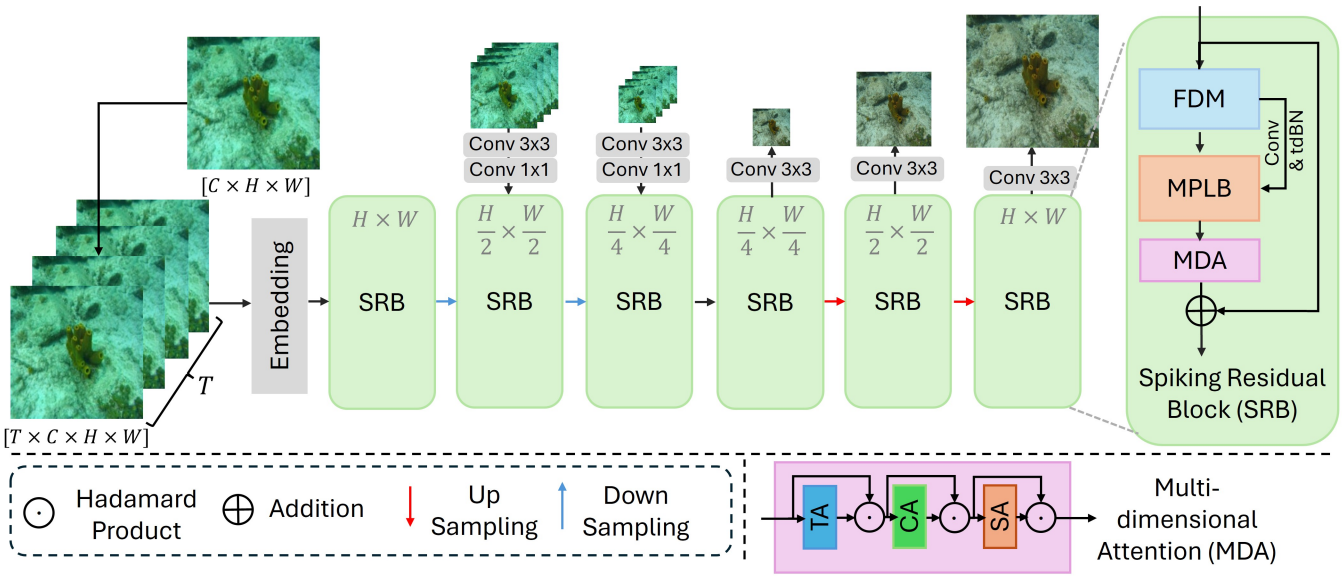


Fig. 2. Overview of the proposed framework. The top panel presents the overall pipeline. The right panel illustrates the proposed Spiking Residual Network (SRB) for feature refinement with spiking dynamics.

fixed neighbourhood, MPLB integrates pooled responses from multiple spatial scales into the membrane potential dynamics, allowing each neuron to jointly encode local details and region-level context. Leveraging the thresholded firing mechanism and task-driven frequency selectivity of spiking neurons, MPLB produces scale-dependent activations: coarse-scale inputs emphasise spatially extensive low-frequency degradations (e.g., veiling light and colour bias), while fine-scale inputs preserve texture-sensitive cues. Therefore, MPLB enlarges spatial awareness without indiscriminately smoothing details, and encourages heterogeneous representations across scales.

Built upon MPLB, we develop **UIESNN**, an end-to-end SNN framework tailored for underwater image enhancement. UIESNN captures fine-grained structures and global colour/contrast corrections in a spike-driven manner, making it suitable for power-constrained underwater vision systems. Experiments on **EUVP** and **LSUI** show UIESNN consistently outperforms prior SNN-based restoration baselines [5], [6], which shows importance of scale-aware spiking representations for low-frequency degradations. Code is available: https://github.com/ChrisChen1023/UIESNN. Our contributions are:

- We identify a core challenge in applying SNNs to UIE: dominant multi-scale low-frequency degradations are insufficiently captured by conventional LIF neurons with limited receptive fields and saturated representations.
- We propose **MPLB**, a multi-scale spike-driven pooling block to enlarge receptive fields while preserving details and inducing heterogeneous feature representations.
- We develop **UIESNN**. Extensive experiments on EUVP and LSUI demonstrate that UIESNN achieves state-of-the-art performance among SNN-based methods.

## II. Related Work

### A. Underwater Image Enhancement

Underwater image enhancement has progressed from prior-based, handcrafted pipelines to learning-based methods. Early methods relied on imaging priors such as attenuation and scattering, which often generalise poorly across diverse underwater conditions [8]. Deep learning has improved robustness by combining data-driven representations with physical insights, including model-based designs [9], prior-guided CNNs [10], and GAN-based restoration [11], though GAN training can be unstable and CNN receptive fields remain limited.

To better handle large-scale degradations, recent UIE models adopt Transformers for long-range dependency modeling [12] and phase-aware attention for improved structural fidelity [13]. Meanwhile, frequency-domain modelling is increasingly popular, leveraging wavelet or Fourier priors [14], shallow-layer frequency features [15], and spatial–frequency interaction with FFT-based refinement [16] to better capture the low-frequency nature of underwater degradations.

More recently, energy-efficient SNNs have been explored for UIE. [17] proposed a convolutional spiking encoder–decoder, which shows that competitive restoration can be achieved with $T = 5$ timesteps while reducing computation and energy consumption. [18] further explored SNNs for UIE by directly adopting two existing SNN architectures. However, existing SNN-based UIE efforts have not explicitly addressed the key challenge that arises when spiking neurons are applied to underwater imagery: standard LIF-style dynamics are inherently local and receptive-field limited, and thus struggle to capture multi-level degradations (e.g., region-wide colour casts and haze-like scattering) that dominate underwater scenes.

### B. Spiking Neural Network

Spiking Neural Networks are increasingly studied as biologically plausible and energy-efficient alternatives to ANNs [19], leveraging sparse event-driven spikes for low-power computation [7], [20], [20]. Most modern SNNs are built on Leaky Integrate-and-Fire neurons and are trained either by converting pre-trained ANNs into spiking counterparts [21] or by direct training with surrogate gradients [22]. With these advances,

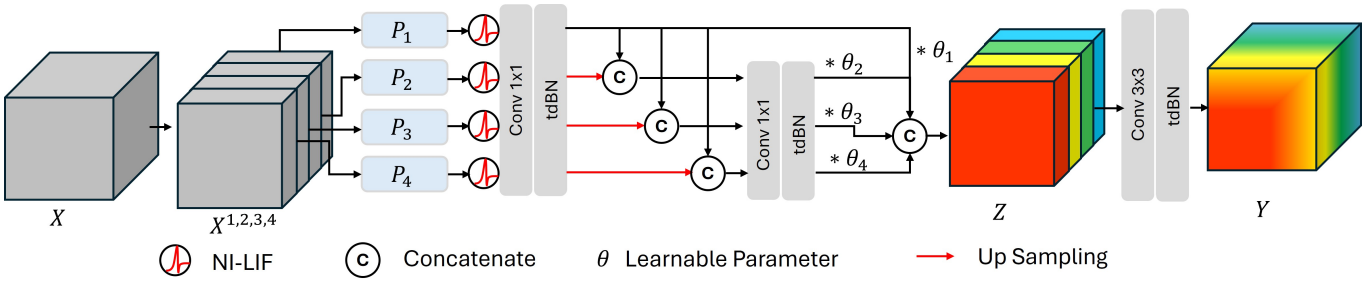


Fig. 3. The illustration of the proposed Multi-scale Pooling LIF Block. NI-LIF is illustrated in Fig. 4.

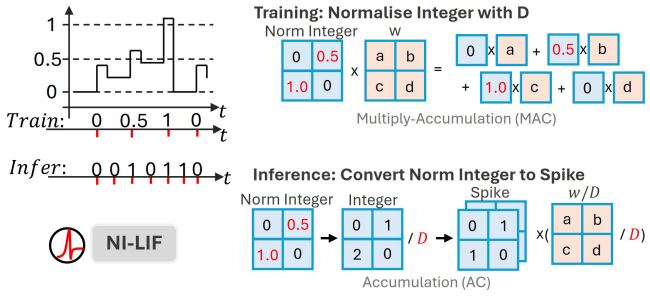


Fig. 4. The illustration of the NI-LIF [7].

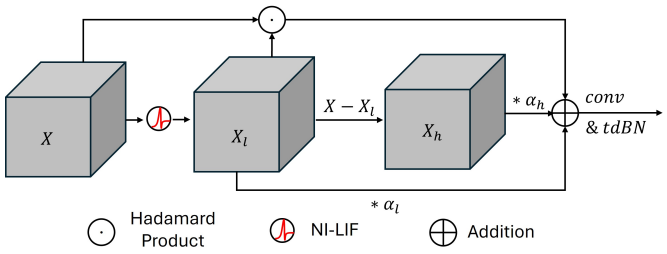


Fig. 5. The illustration of the frequency decomposition module.

SNNs have achieved strong performance on high-level vision tasks such as image classification.

For low-level vision, only a limited number of works have explored SNNs for image restoration [23], and most focus on high-frequency degradations such as single-image deraining [5], [6]. In contrast, our work targets underwater image enhancement, where degradations are predominantly low-frequency and region-level, and proposes a spiking solution tailored to this distinct challenge.

## III. Methods

This section presents **UIESNN** for static underwater image enhancement. We first discuss why spiking neurons require a UIE-specific design, then introduce a **Multi-Scale Pooling LIF Block** (III-B) that expands the effective receptive field. Next, we build a **Spiking Residual Block** (III-C) by integrating frequency decomposition, multi-scale spiking perception, and attention. Finally, we describe the overall encoder–decoder architecture and the multi-scale training objective.

### A. Why Spiking Neuron Receptive Fields Matter?

Underwater image enhancement is dominated by low-frequency degradations such as wavelength-dependent colour casts and scattering-induced veiling. [6] finds that spiking

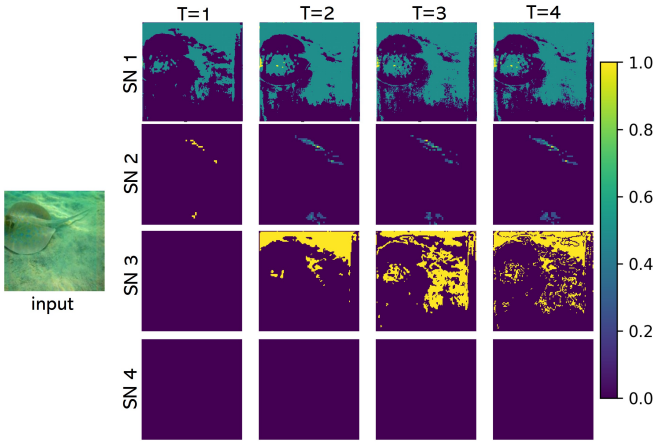


Fig. 6. Temporal spiking feature visualisation in the Multi-scale Pooling LIF Block. SN1–SN4 denote the four spike neurons in the four scale branches of MPLB after the corresponding $P_1$–$P_4$ in Sec. III-B

neurons can behave as frequency-domain indicators in a task-driven manner. When the degradation is primarily low-frequency, the spiking responses tend to emphasise low-frequency components. However, a conventional LIF neuron is inherently local because its membrane integrates signals at each spatial location independently, which results in a unit receptive field in the spatial domain. This locality becomes a bottleneck for UIE because underwater degradations are often global yet spatially heterogeneous. As shown in Fig. 1, a pollution-induced green tint can be global in distribution, but its manifestation becomes non-uniform around bright objects, saturated colors, biological textures, and water-caustic patterns. These heterogeneous low-frequency artifacts frequently appear as irregular regions that are locally continuous, and they are difficult to be reliably captured when the spiking decision is made from a single pixel or a single feature location. This motivates us to design the enlarged receptive fields for spiking neurons to solve this problem.

### B. Multi-scale Pooling LIF Block

To address the receptive field limitation without sacrificing efficiency, we propose the **Multi-scale Pooling LIF Block** (MPLB). Instead of enlarging receptive fields with large-kernel convolutions, we use parameter-free average pooling at multiple spatial scales and let spiking dynamics decide which scale to activate, as illustrated in Fig. 3. Specifically, given an input tensor $X \in \mathbb{R}^{T\times B\times C\times H\times W}$, we split it evenly along

the channel dimension into four groups and feed them to four scale branches:

$$\left(X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}\right) = \mathrm{Split}_c(X), \tag{1}$$

$$X^{(i)} \in \mathbb{R}^{T\times B\times \frac{C}{4}\times H\times W}, \tag{2}$$

so that $X = [X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}]_c$ (channel-wise concatenation). The four branches correspond to: (i) pixel-level perception ($\mathcal{P}_1$), (ii) coarse local perception with $4\times4$ average pooling, (iii) intermediate local perception with $2\times2$ average pooling, and (iv) global perception with adaptive global average pooling:

$$\mathcal{P}_1 = \mathrm{Identity}, \quad \mathcal{P}_2 = \mathrm{AvgPool}_{4\times4}, \tag{3}$$

$$\mathcal{P}_3 = \mathrm{AvgPool}_{2\times2}, \quad \mathcal{P}_4 = \mathrm{AdaGAP}. \tag{4}$$

Each branch applies a spiking neuron, followed by a lightweight $1\times1$ convolution and threshold-dependent batch normalisation (tdBN), then upsamples back to the original spatial resolution when pooling is used. We implement the NI-LIF [7] spiking neuron using a normalised multi-spike variant (shown in Fig. 4). For each branch $i \in \{1,2,3,4\}$ and time step $t$, we first obtain the input current via the corresponding pooling operator:

$$U_t^{(i)} = \mathcal{P}_i\left(X_t^{(i)}\right), \tag{5}$$

then update the membrane potential as:

$$\mu_t^{(i)} = \gamma\left(\mu_{t-1}^{(i)} - S_{t-1}^{(i)}\right) + U_t^{(i)}, \tag{6}$$

where $\gamma$ is the decay constant. Inspired by [7], the spike output is produced by a quantised surrogate activation:

$$S_t^{(i)} = \mathcal{Q}\left(M_t^{(i)}\right), \tag{7}$$

which clips and rounds membrane values into a small set of discrete spike levels and normalises them to improve stability. After spiking, we apply a projection and normalisation, and upsample if needed:

$$\hat{F}_t^{(i)} = \mathrm{tdBN}\left(\mathrm{Conv}_{1\times1}\left(S_t^{(i)}\right)\right), \qquad F_t^{(i)} = \mathcal{U}_i\left(\hat{F}_t^{(i)}\right), \tag{8}$$

where $\mathcal{U}_1 = \mathrm{Id}$ and $\mathcal{U}_i$ (for $i = 2,3,4$) upsamples features back to $H\times W$. Collecting all time steps yields

$$F_i \in \mathbb{R}^{T\times B\times \frac{C}{4}\times H\times W}, \qquad i = 1,2,3,4. \tag{9}$$

We then fuse multi-scale features by forming three pairwise mixtures with the pixel-level branch to enlarge the receptive field while preserving fine-grained details:

$$M_1 = \phi\big([F_1, F_2]_c\big), \quad M_2 = \phi\big([F_1, F_3]_c\big), \tag{10}$$

$$M_3 = \phi\big([F_1, F_4]_c\big), \tag{11}$$

where $[\cdot,\cdot]_c$ denotes channel-wise concatenation and $\phi(\cdot)$ denotes a $1\times1$ convolution followed by threshold-dependent batch normalisation (optionally projecting back to $\frac{C}{4}$ channels for efficiency). In this way, MPLB can respond to low-frequency degradations at multiple spatial supports, from regional bias to global cast, without losing fine-grained detail features. To enable input-adaptive scale selection with negligible overhead, we introduce a set of learnable fusion coefficients $\theta_i, i = 1,2,3,4$ to reweight the pixel-level feature and the three mixed features before the final aggregation:

$$\tilde{F}_1 = \theta_1 \odot F_1, \quad \tilde{M}_1 = \theta_2 \odot M_1, \tag{12}$$

$$\tilde{M}_2 = \theta_3 \odot M_2, \quad \tilde{M}_3 = \theta_4 \odot M_3, \tag{13}$$

where $\odot$ denotes element-wise multiplication with $\theta_i$ to introduce an explicit mechanism to resolve cross-scale conflicts: for regions dominated by low-frequency degradations, the model can prioritise the global or coarse pooling. For textures and edges, it can prioritise the pixel-level pathway. After reweighting, we aggregate all branches via concatenation and a lightweight $3\times3$ fusion:

$$Z = [\tilde{F}_1, \tilde{M}_1, \tilde{M}_2, \tilde{M}_3], \tag{14}$$

$$Y = \mathrm{tdBN}\left(\mathrm{Conv}3\times3(Z)\right), \tag{15}$$

where $Y$ is the MPLB output. Notably, the fusion remains efficient: multi-scale context is from parameter-free pooling and small-kernel convolutions, avoiding the expensive cost in large kernel size associated with receptive field expansion.

**Analysis of Multi-scale Pooling LIF Block** Fig. 6 visualises the temporal responses of the four NI-LIF branches in MPLB. The multi-scale pooling operators provide different spatial ranges, therefore each branch produces a distinct spiking pattern. The finest branch (SN 1) shows more spatially selective activations that are helpful for preserving local structures and texture cues, while the pooled branches produce more region-wise responses that better capture the degradations from specific regions, such as global colour cast and veiling light (SN 2). Across timesteps, the post-spike features generally become stronger and more complete because membrane integration accumulates evidence before firing, which helps aggregate weak but globally consistent cues. Moreover, different branches respond with different strengths over time, suggesting that MPLB can emphasise the most informative scales for a given input, and this scale-aware fusion improves colour fidelity and spatial coherence in the restored results.

### *C. Spiking Residual Block*

Based on MPLB, we design a **Spiking Residual Block** (SRB). The block contains two consecutive groups. Inspired by the task-driven frequency indication behaviour [6], as shown in Fig. 5, we build the first group as the Frequency Decomposition Module (FDM), which uses a spiking neuron to perform frequency-like decomposition. Given input $X$, we compute spiking responses:

$$X_l = \mathrm{LIF}(X), \quad X_h = X - X_l, \tag{16}$$

where $X_l$ acts as a low-frequency feature spiked by the indicator LIF [6] and $X_h$ retains the residual component. We apply learnable scalars $\alpha_l$ and $\alpha_h$ to optimise the two

components, and we further introduce an element-wise gated enhancement term $X \odot X_l$. The fused representation becomes

$$\widetilde{X} = \alpha_l X_l + \alpha_h X_h + (X \odot X_l). \tag{17}$$

The $\widetilde{X}$ is fed into a combo of $3\times3$ convolution and tdBN. The second group applies MPLB to enlarge receptive fields and capture multi-level low-frequency degradations. Specifically, we compute:

$$Z = \mathrm{MPLB}(\mathrm{tdBN}(\mathrm{Conv}(\widetilde{X}))). \tag{18}$$

Then we apply another $3 \times 3$ projection and normalisation. Finally, we combine residual learning and attention. We use a projected shortcut path to match distributions and a Multi-Dimensional Attention [24] (MDA) module to refine the output. The block output is

$$Y = \mathrm{MDA}(Z + \mathrm{Shortcut}(X)) + X. \tag{19}$$

By design such a Spiking Residual Block, we preserve stable information flow while allowing the spiking pathway to focus on region-level underwater degradations.

### D. Overall Architecture and Loss Function

**Overall architecture.** UIESNN adopts a three-level encoder-decoder architecture composed of stacked Spiking Residual Blocks. For a static input image $I \in \mathbb{R}^{B\times3\times H\times W}$, we replicate it along the temporal axis to obtain $I^{(t)}$ for $t = 1, \ldots, T$, following the multi-step spiking pipeline. A shallow overlap patch embedding layer maps the input to the feature space using a $3 \times 3$ convolution in multi-step mode. The encoder contains Level 1, Level 2, and Level 3 stages with progressive downsampling.

To improve robustness to diverse underwater degradations, we inject multi-scale inputs into deeper encoder stages. Specifically, at Level 2 and Level 3, we downsample the original image to the target resolution, embed it using lightweight convolutions, concatenate it with encoder features, and then compress channels with a linear projection. This provides explicit low-level references at multiple scales and helps colour and illumination correction.

The decoder follows a symmetric hierarchy with fewer spiking residual blocks per level. Each level upsamples features per timestep, then fuses the corresponding encoder features via skip connections and refines them with stacked Spiking Residual Blocks. UIESNN outputs predictions at Level 3 and Level 2, and a final full-resolution result at Level 1. For each head, we temporally average features using $\mathrm{mean}$ before a final $3\times3$ reconstruction convolution, and add the input image as a global residual to obtain the restored output.

**Training objective.** We supervise three outputs using a multi-scale loss. $\hat{I}^{(1)}$ is the final output, $\hat{I}^{(1/2)}$ and $\hat{I}^{(1/4)}$ are the Level 2 and Level 3 outputs after resizing to the full resolution and adding the residual input in the forward pass. During training, we compare each prediction with the ground truth at the corresponding scale by downsampling the target and the prediction with bilinear interpolation.

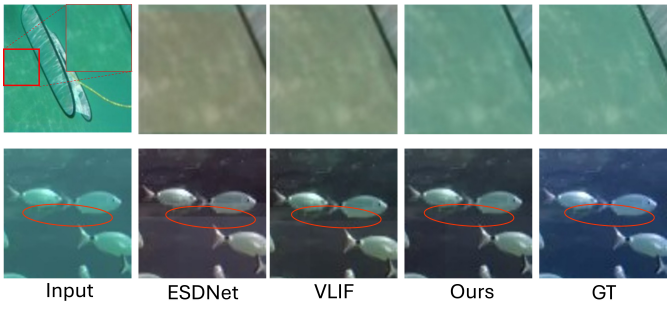


Fig. 7. Qualitative comparison on underwater image enhancement. The top and bottom rows are results on EUVP and LSUI, respectively. The red boxes (top row) highlight that our method restores more accurate colours and finer texture details, while the red ellipses (lower rows) show that our method produces more spatially coherent structures with fewer discontinuities compared with prior SNN-based methods.

We define the content loss as an $L_1$ loss summed over scales

$$\mathcal{L}_{\text{pix}} = \sum_{s\in\{1,\,1/2,\,1/4\}} \left\| D_s(\hat{I}^{(s)}) - D_s(I_{\text{gt}}) \right\|_1, \tag{20}$$

where $D_s(\cdot)$ denotes bilinear downsampling by scale factor $s$ and $I_{\text{gt}}$ is the ground truth image. We also apply an SSIM loss at each scale

$$\mathcal{L}_{\text{ssim}} = \sum_{s\in\{1,\,1/2,\,1/4\}} \left(1 - \mathrm{SSIM}\left(D_s(\hat{I}^{(s)}), D_s(I_{\text{gt}})\right)\right). \tag{21}$$

To explicitly regularise frequency content, we further include a Fourier-domain loss. Let $\mathcal{F}(\cdot)$ denote the 2D FFT. We compute the frequency discrepancy by measuring the $L_1$ distance between the Fourier spectra of the prediction and the ground truth at multiple scales:

$$\mathcal{L}_{\text{fft}} = \sum_{s\in\{1,\,1/2,\,1/4\}} \left\| \mathcal{F}\big(D_s(\hat{I}^{(s)})\big) - \mathcal{F}\big(D_s(I_{\text{gt}})\big) \right\|_1. \tag{22}$$

The final loss is:

$$\mathcal{L} = \lambda_{\text{pix}}\mathcal{L}_{\text{pix}} + \lambda_{\text{ssim}}\mathcal{L}_{\text{ssim}} + \lambda_{\text{fft}}\mathcal{L}_{\text{fft}}, \tag{23}$$

where we set $\lambda_{\text{fft}} = 0.1$ and $\lambda_{\text{ssim}} = 1$, and $\lambda_{\text{pix}} = 0.5$. This objective jointly enforces spatial fidelity, structural consistency, and frequency alignment, which is particularly important for underwater restoration where low-frequency color and illumination shifts coexist with fine texture details.

## IV. Experiments

### A. Setting and Energy Calculation

**Setting** Unless noted otherwise, we run all comparative experiments under unified training protocols. Our UIESNN is built by stacking Spiking Residual Blocks in a stage-wise layout of $[4, 4, 8, 2, 2, 2]$. Both downsampling and upsampling are implemented with spike-driven convolutional operators to maintain a fully event-based feature transformation throughout the network. We set the virtual timestep scaling constant to $D$=4 and use $T$=4 timesteps in all experiments. During training, we crop patches of $64 \times 64$ with a batch size of 12. Following [5], we adopt the Sigmoid surrogate function

TABLE I

QUANTITATIVE COMPARISON ON EUVP AND LSUI IN TERMS OF PSNR, SSIM, PARAMS (M), AND ENERGY (MJ). **BOLD** INDICATES THE BEST PERFORMANCE IN EACH COLUMN, AND UNDERLINE INDICATES THE SECOND-BEST. ESDNET AND VLIF ARE FULLY TRAINED ON EUVP AND LSUI FOLLOWING THEIR OFFICIAL SETTINGS.

| Dataset | EUVP | | | | | LSUI | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | ANN | | SNN | | | ANN | | SNN | | |
| Metric | TACL [25] | UIE-WD [26] | ESDNet [5] | VLIF [6] | Ours | TACL [25] | UIE-WD [26] | ESDNet [5] | VLIF [6] | Ours |
| PSNR↑ | 20.99 | 17.80 | 25.36 | <u>25.69</u> | **26.97** | 22.97 | 19.23 | 23.8747 | <u>24.1731</u> | **24.7346** |
| SSIM↑ | 0.782 | 0.760 | 0.8645 | <u>0.8817</u> | **0.8936** | 0.8280 | 0.8036 | 0.8725 | <u>0.8744</u> | **0.8754** |
| Params(M)↓ | 28.29 | 14.46 | **12.81** | <u>15.72</u> | 16.72 | 28.29 | 14.46 | **12.81** | <u>15.72</u> | 16.72 |
| Energy(mJ)↓ | 1104.3 | 472.7 | **174.63** | 697.63 | <u>199.33</u> | 1104.3 | 472.7 | **174.63** | 697.63 | <u>199.33</u> |

for gradient backpropagation. All models are implemented in PyTorch and trained on a single NVIDIA A6000 GPU. We evaluate UIESNN on the widely used EUVP [27] and LSUI [12] benchmarks. Since official checkpoints are not publicly available for the compared SNN baselines, we strictly follow the original experimental configurations of ESDNet and VLIF to reproduce their performance on UIE. Unless specified otherwise, ablation studies are conducted on EUVP.

**Energy Calculation** We estimate inference energy using an operation-count proxy that is widely adopted in the SNN literature [4], [7], [24], where the dominant cost is attributed to multiply-accumulate (MAC) operations in ANNs and spike-triggered accumulation (AC) operations in SNNs. For a convolution layer with output spatial size $O \times O$, input and output channels $C_{in}$ and $C_{out}$, and kernel size $k$, the ANN energy is proportional to the dense MAC count:

$$E_{\text{ANN}} = O^2 \cdot C_{in} \cdot C_{out} \cdot k^2 \cdot E_{\text{MAC}}. \tag{24}$$

In contrast, spike-driven inference replaces dense MACs with sparse accumulations that occur only when spikes arrive. We therefore scale the operation count by the average firing rate $fr$ measured during inference, and by the effective number of inference steps. When integer-valued training is used and converted to binary spikes by expanding virtual steps, we denote the effective steps as $T \times D$. The resulting SNN energy proxy becomes:

$$E_{\text{SNN}} = (T \times D) \cdot fr \cdot O^2 \cdot C_{in} \cdot C_{out} \cdot k^2 \cdot E_{\text{AC}}. \tag{25}$$

We use widely recognised per-operation costs $E_{\text{MAC}} = 4.9pJ$ and $E_{\text{AC}} = 0.9pJ$ for MAC and AC [28], respectively, and report total network energy by summing $E$ over all layers.

### B. Experimental Results

Tab. I shows the quantitative results on the EUVP and LSUI datasets. Compared with the previous SNN state-of-the-art method VLIF, UIESNN leverages multi-scale information to better capture low-frequency underwater degradations, enabling **UIESNN** to achieve state-of-the-art performance among SNN methods. Specifically, UIESNN reaches 26.97 dB PSNR and 0.8936 SSIM on EUVP, exceeding VLIF by 1.28 dB and 0.0119. On LSUI, UIESNN attains 24.7346 dB PSNR and 0.8754 SSIM, improving PSNR by 0.56 dB over VLIF and 0.86 dB over ESDNet. We additionally report two ANN-based methods (TACL and UIE-WD) for reference. Fully closing the performance gap between ANNs and SNNs is not the goal of this work. Instead, we focus on advancing SNN-based UIE with strong restoration quality and improved efficiency. Notably, UIESNN outperforms these ANN baselines on both datasets while using only 18.05% and 42.17% of their energy consumption (199.33 mJ vs. 1104.3 mJ and 472.7 mJ), and it requires 59.1% of TACL's parameters (16.72M vs. 28.29M).

TABLE II

ABLATION STUDY ON ARCHITECTURAL COMPONENTS.

| Model | FDB | MDA | MPLB | PSNR↑ | SSIM↑ |
|---|---|---|---|---|---|
| (a) | | | | 25.39 | 0.8643 |
| (b) | ✓ | | | 25.45 | 0.8677 |
| (c) | ✓ | ✓ | | 25.88 | 0.8718 |
| Ours | ✓ | ✓ | ✓ | **26.97** | **0.8936** |

Fig. 7 presents qualitative results on two benchmark datasets EUVP and LSUI. On EUVP (top row), our method exhibits more accurate global colour restoration and finer-grained appearance details. In the highlighted regions, competing methods tend to produce noticeable colour shifts or over-smoothed textures, which weaken material cues and local contrast. In contrast, our results better preserve subtle texture variations and natural colour transitions, indicating stronger capability in capturing global illumination and colour statistics while maintaining fine details. On LSUI (bottom row), we observe that ESDNet and VLIF suffer from different degrees of spatial inconsistency, especially around object boundaries and elongated structures, as marked by the red ellipses. A plausible reason is that local perception in spiking networks introduces an inductive bias, so when feature integration relies on local windows or limited receptive fields, small response mismatches between neighbouring regions can accumulate and cause discontinuous structures and unstable geometry. Our method alleviates this issue by strengthening the cross-region feature coupling and reducing the reliance on purely local evidence aggregation, which leads to more coherent contours, smoother structural transitions, and fewer block-like artefacts.

### C. Ablation Studies

*1) Effectiveness of Each Components:* Table II evaluates the contribution of each component using PSNR and SSIM on EUVP. Compared to the baseline (a), which involves simple spike-driven convolution and tdBN in SRB modules, adding FDB (b) yields a modest improvement, suggesting that frequency decomposition promotes more diverse temporal features. Adding MDA on top of FDB (c) further improves performance by strengthening spatio-temporal and channel interactions. With MPLB enabled, the full model achieves the best results, yielding significant gains of 6.22%↑ / 3.39%↑ (PSNR / SSIM) over (a) and 4.21%↑ / 2.50%↑ over (c).

TABLE III
IMPACT OF DIFFERENT TIME STEPS (T) AND QUANTISATION STEPS (D) ON EUVP DATASET.

| T × D | Energy $(mJ)\downarrow$ | PSNR↑ | SSIM↑ |
|---|---|---|---|
| (1 × 1) | 18.84 | 25.66 | 0.8591 |
| (1 × 4) | 43.42 | 25.84 | 0.8634 |
| (4 × 1) | 52.20 | 26.57 | 0.8902 |
| (4 × 4) **Ours** | 199.33 | **26.97** | **0.8936** |

This indicates that the scale-aware representations induced by MPLB are crucial for underwater image enhancement.

*2) Effectiveness of Different Time Steps and Quantisation Steps:* Tab. III evaluates the effects of time steps ($T$) and quantisation steps ($D$) on EUVP. Extending the temporal horizon from $T$=1 to $T$=4 yields clear gains of 3.55%↑ / 3.62%↑ (PSNR / SSIM), showing that longer time steps facilitate temporal aggregation and better capture complex degradation patterns. Increasing $D$ from 1 to 4 provides smaller gains (0.70%↑ / 0.50%↑) with lower energy cost (43.42 mJ) than increasing $T$ (52.20 mJ). The full setting ($T$=4, $D$=4) performs best, improving over (4×1) by 1.51%↑ / 0.38%↑, but at much higher energy (199.33 mJ vs. 52.20 mJ). Overall, temporal integration contributes most of the gains for modeling underwater degradations in SNNs, while finer quantisation offers limited improvements at a high energy cost.

## V. CONCLUSION

This work studies spiking neural networks for underwater image enhancement (UIE). We explored that, UIE requires correcting spatially extensive, low-frequency degradations such as colour casts and haze-like blur, which are difficult for limited-receptive-field spiking neurons. To solve this, we propose the **Multi-scale Pooling LIF Block (MPLB)** to expand the receptive field with hierarchical pooling and inject scale-aware context into membrane dynamics, capturing global cues while preserving local details. Building on MPLB, we develop **UIESNN**, an end-to-end SNN tailored for UIE. Experiments on large-scale real-world underwater datasets show that UIESNN consistently outperforms prior SNN baselines with low energy cost. Future work will extend UIESNN to video enhancement and improve the performance-energy trade-off with more adaptive spiking dynamics.

## REFERENCES


[1] W. Maass, "Networks of spiking neurons: the third generation of neural network models," *Neural networks*, vol. 10, no. 9, pp. 1659–1671, 1997.

[2] J. Wang, L. Yu, L. Huang, C. Zhou, H. Zhang, Z. Song, H. Liu, M. Zhang, Z. Ma, and Z. Zhang, "Efficient speech command recognition leveraging spiking neural networks and progressive time-scaled curriculum distillation," *Neural Networks*, vol. 195, p. 108253, 2026.

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

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

[5] T. Song, G. Jin, P. Li, K. Jiang, X. Chen, and J. Jin, "Learning a spiking neural network for efficient image deraining," in *IJCAI*, 2024.

[6] S. Chen, T. Krajnik, F. Arvin, and A. Atapour-Abarghouei, "Exploring the potentials of spiking neural networks for image deraining," *arXiv preprint arXiv:2512.02258*, 2025.

[7] Z. Lei, M. Yao, J. Hu, X. Luo, Y. Lu, B. Xu, and G. Li, "Spike2former: Efficient spiking transformer for high-performance image segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 2, 2025, pp. 1364–1372.

[8] J. Y. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1756–1769, 2011.

[9] K. Wang, Y. Hu, J. Chen, X. Wu, X. Zhao, and Y. Li, "Underwater image restoration based on a parallel convolutional neural network," *Remote sensing*, vol. 11, no. 13, p. 1591, 2019.

[10] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognition*, vol. 98, p. 107038, 2020.

[11] R. Cong, W. Yang, W. Zhang, C. Li, C.-L. Guo, Q. Huang, and S. Kwong, "Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators," *IEEE Transactions on Image Processing*, 2023.

[12] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," *IEEE Transactions on Image Processing*, 2023.

[13] M. Khan, A. Negi, A. Kulkarni, S. S. Phutke, S. K. Vipparthi, and S. Murala, "Phaseformer: Phase-based attention mechanism for underwater image restoration and beyond," *arXiv preprint arXiv:2412.01456*, 2024.

[14] C. Zhao, W. Cai, C. Dong, and C. Hu, "Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 8281–8291.

[15] X. Guo, Y. Dong, X. Chen, W. Chen, Z. Li, F. Zheng, and C.-M. Pun, "Underwater image restoration via polymorphic large kernel cnns," *arXiv preprint arXiv:2412.18459*, 2024.

[16] S. Chen, R. Thenius, F. Arvin, and A. Atapour-Abarghouei, "Deep-sea: Deep-learning enhancement for environmental perception in submerged aquatics," *arXiv preprint arXiv:2508.12824*, 2025.

[17] V. Sudevan, F. Zayer, R. Kausar, S. Javed, H. Karki, G. De Masi, and J. Dias, "Underwater image enhancement by convolutional spiking neural networks," *arXiv preprint arXiv:2503.20485*, 2025.

[18] J. Shao, H. Zhang, and J. Miao, "Lamsnn: Learnable adaptive modulation for artifact suppression in spiking underwater image enhancement networks," *Neural Networks*, p. 108210, 2025.

[19] J. Wang, Z. Ma, X. Shen, C. Zhou, L. Zhao, H. Zhang, Y. Zhong, S. Cai, Z. Song, and Z. Zhang, "S$^2$m-former: Spiking symmetric mixing branchformer for brain auditory attention detection," in *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. [Online]. Available: https://openreview.net/forum?id=WtMuGdHvh6

[20] J. Wang, L. Yu, X. Shen, S. Guo, C. Zhou, L. Zhao, Y. Zhong, Z. Zhang, and Z. Ma, "Spikcommander: A high-performance spiking transformer with multi-view learning for efficient speech command recognition," *arXiv preprint arXiv:2511.07883*, 2025.

[21] Y. Cao, Y. Chen, and D. Khosla, "Spiking deep convolutional neural networks for energy-efficient object recognition," *IJCV*, vol. 113, pp. 54–66, 2015.

[22] E. O. Neftci, H. Mostafa, and F. Zenke, "Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks," *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 51–63, 2019.

[23] R. Xu, J. Xie, J. Nie, J. Cao, and Y. Pang, "Snnsir: A simple spiking neural network for stereo image restoration," *arXiv preprint arXiv:2508.12271*, 2025.

[24] M. Yao, G. Zhao, H. Zhang, Y. Hu, L. Deng, Y. Tian, B. Xu, and G. Li, "Attention spiking neural networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 8, pp. 9393–9410, 2023.

[25] R. Liu, Z. Jiang, S. Yang, and X. Fan, "Twin adversarial contrastive learning for underwater image enhancement and beyond," *IEEE Transactions on Image Processing*, vol. 31, pp. 4922–4936, 2022.

[26] Z. Ma and C. Oh, "A wavelet-based dual-stream network for underwater image enhancement," in *ICASSP*, 2022, pp. 2769–2773.

[27] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3227–3234, 2020.

[28] M. Horowitz, "1.1 computing's energy problem (and what we can do about it)," in *2014 IEEE international solid-state circuits conference digest of technical papers (ISSCC)*. IEEE, 2014, pp. 10–14.