



OPEN ACCESS

EDITED BY

Lei Deng,
Tsinghua University, China

REVIEWED BY

Chenglong Zou,
Peking University, China
Fan Yimeng,
Tianjin University, China

*CORRESPONDENCE

Gang Wang
✉ g_wang@foxmail.com
Yong Song
✉ yongsong@bit.edu.cn

RECEIVED 08 July 2025

ACCEPTED 03 September 2025

PUBLISHED 22 September 2025

CITATION

Wu S, Wang G, Song Y, Zhao Y, Zhou Y,
Meng Q and Liao Y (2025) Bidirectional
dynamic threshold SNN for enhanced object
detection with rich spike information.
Front. Neurosci. 19:1661916.
doi: 10.3389/fnins.2025.1661916

COPYRIGHT

© 2025 Wu, Wang, Song, Zhao, Zhou, Meng
and Liao. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Bidirectional dynamic threshold SNN for enhanced object detection with rich spike information

Shaoxing Wu¹, Gang Wang^{2*}, Yong Song^{1*}, Yufei Zhao³,
Ya Zhou¹, Qingzhu Meng¹ and Yizhao Liao¹

¹School of Optics and Photonics, Beijing Institute of Technology, Beijing, China, ²Center of Brain Sciences, Beijing Institute of Basic Medical Sciences, Beijing, China, ³Beijing Institute of Automation and Control Equipment, Beijing, China

Spiking Neural Networks (SNNs), inspired by neuroscience principles, have gained attention for their energy efficiency. However, directly trained SNNs lag behind Artificial Neural Networks (ANNs) in accuracy for complex tasks like object detection due to the limited information capacity of binary spike feature maps. To address this, we propose BD-SNN, a new directly trained SNN equipped with Bidirectional Dynamic Threshold neurons (BD-LIF). BD-LIF neurons emit +1 and -1 spikes and dynamically adjust their thresholds, enhancing the network's information encoding capacity and activation efficiency. Our BD-SNN incorporates two new all-spike residual blocks, BD-Block1 and BD-Block2, for efficient information extraction and multi-scale feature fusion, respectively. Experiments on the COCO and Gen1 datasets demonstrate that BD-SNN improves accuracy by 3.1% and 2.8% compared to the state-of-the-art EMS-YOLO method, respectively, validating BD-SNN's superior performance across diverse input scenarios. Project will be available at <https://github.com/Ganpei576/BD-SNN>.

KEYWORDS

RGB and event, spiking neural networks, neuromorphic computing, neuron model, object detection

1 Introduction

Object detection is a challenging task in the field of computer vision, which aims to determine the location and category of each object in images or videos, forming the foundation for further analysis and processing. In recent years, ANNs have achieved remarkable results in many fields including object detection. However, as model complexity grows, these methods entail significant computational and memory demands, posing challenges for deployment in real-time applications and resource-constrained environments (Howard, 2017). Therefore, there is an urgent need to explore new methods capable of delivering comparable object detection performance to existing ANN methods, while significantly reducing computational costs. A promising approach is to train SNNs directly with surrogate gradient (Srinivasan et al., 2020), which can achieve high performance with few time steps and process both static images and event data efficiently.

With the development of artificial intelligence, SNN, known as the next generation of neural networks (Maass, 1997), has attracted widespread attention due to its unique advantages such as asynchronous discrete event drive, spike activation that is more in line with physiological characteristics, and no floating-point multiplication operations during

the operation process. SNNs employ spiking neurons as computational units (Zador, 1997) and convey information through binary spike signals. Information is transmitted via spikes only when the neurons membrane potential, which denotes the internal state variable of the spiking neuron and corresponds to the membrane potential in biological neurons, reaches the excitation threshold (Andrew, 2003). Therefore, during network inference, the floating-point multiplication involved in weight computation and neuron activation in SNNs can be substituted with addition, enabling more efficient and faster computations compared to ANNs (Merolla et al., 2014; Poon and Zhou, 2011). In addition, event-driven computing methods can also show higher energy efficiency on neuromorphic hardware (Davies et al., 2018; Akopyan et al., 2015; Liu et al., 2019). Today, SNN has been widely used in many fields, including object classification (Hu et al., 2024; Zhu et al., 2024; Shan et al., 2024; Wu et al., 2023), detection (Kim et al., 2020b,a; Li et al., 2022), and tracking (Liu et al., 2024; Zhang J. et al., 2023), etc.

However, when dealing with complex regression tasks such as object detection, the accuracy of directly trained SNNs is inferior to that of ANNs (Zhang H. et al., 2023; Hopkins et al., 2018). In object detection tasks, the network must not only identify objects but also precisely delineate their boundaries, necessitating robust feature representation capabilities (He et al., 2019). SNNs encode information via spike emissions, but their inherently discrete signal transmission constrains the networks feature representation capabilities. Compared to the activation feature maps of ANNs, the binary nature of spikes hinders the smooth capture and representation of complex feature variations, impairing the networks ability to discern subtle features and potentially causing information loss (Guo et al., 2024). Recent efforts to directly train deep SNNs for object detection using the surrogate gradient method have achieved higher accuracy than ANN-to-SNN conversion methods, requiring only four time steps compared to hundreds (Su et al., 2023). However, a performance gap persists when compared to ANNs with equivalent network architectures. Some studies have sought to enhance network performance by mitigating information loss in SNNs (Kim et al., 2023; Guo et al., 2023, 2022; He et al., 2024). However, a systematic analysis of this challenge in object detection tasks remains absent.

To address this issue, we employed information entropy theory (Paninski, 2003) into the object detection task and showed that the binary spike activation maps of SNNs carry substantially less information than ANN activations, leading to reduced accuracy. To tackle this challenge, we propose BD-SNN network, designed to comprehensively express the information embedded in input data for efficient object detection. The BD-SNN network integrates two new full-spike residual blocks, BD-Block 1 and BD-Block 2, alongside the new designed bidirectional dynamic threshold neuron model (BD-LIF) employed as the activation unit. Unlike LIF neurons used in traditional SNNs, BD-LIF neurons emit spikes in two distinct modes, -1 and 1 , enabling the transmission of diverse information. Additionally, this neuron model dynamically adjusts its spike threshold, with the threshold varying in response to the depolarization rate of the membrane potential (Azouz and Gray, 2000), mirroring the biological principle of inverse proportionality between spike threshold and depolarization rate. Experimental results indicate that the proposed

network outperforms those networks through existing ANN-to-SNN conversion methods and direct SNN training methods in terms of accuracy.

In summary, our main contributions are fourfold:

- Through theoretical analysis and experimental validation, we identified the issue of limited information capacity in the activation process of conventional binary spike neurons. The information capacity of the LIF neuron activation map is found to be 3 times lower than that of the ReLU activation map.
- We propose the BD-SNN network, designed to enable SNN-based object detection with enhanced information representation. The network incorporates two new full-spike residual blocks, BD-Block1 and BD-Block2, which are designed to efficiently extract information and fuse multi-scale features respectively.
- We developed a BD-LIF neuron model capable of emitting spikes in two distinct forms, 1 and -1 , while dynamically adjusting its threshold. This model is integrated into the BD-SNN network to enhance the information capacity of the spike feature map.
- Experimental results on the COCO and Gen1 datasets demonstrate that BD-SNN, trained with BD-LIF neurons, enhances both information representation and reasoning efficiency. On the COCO dataset, BD-SNN achieves approximately 3.1% higher accuracy than other state-of-the-art methods while requiring only three time steps instead of four.

2 Related work

2.1 Learning strategies of spiking neural networks

There are two main methods to obtain high-performance deep SNNs. The first method involves converting a pre-trained ANN into an SNN with an identical structure, which is called ANN-to-SNN conversion (Bu et al., 2023; Rueckauer et al., 2017). However, this method has some inherent defects that are difficult to solve. First, models trained using this method require extended time steps to approximate the accuracy of the activation values in the original ANN (Bu et al., 2023). This prolongs model inference time and escalates energy consumption, counteracting the low-energy design principles of the SNN. Secondly, the limitations of the rate coding scheme (Al-Hamid and Kim, 2020) cause this method to forgo the rich temporal dynamics of the SNN (Deng et al., 2020; Van Rullen and Thorpe, 2001), rendering it unsuitable for dynamic datasets captured by event cameras. These limitations restrict both the practical application and research potential of this method.

The second method is to train SNN using direct training method (Zhou et al., 2024). This method optimizes the network model by backpropagating simultaneously in the time and space dimensions (Wu et al., 2018). The network trained with this method significantly reduces the time steps required for inference (Wu et al., 2019), thereby reducing energy consumption during

inference. Furthermore, when processing dynamic datasets, the direct training method's ability to simultaneously optimize both the time and space dimensions allows the trained model to better handle time-varying inputs (Deng et al., 2020) and adapt to diverse real-time application scenarios. These advantages have recently drawn increasing attention to direct training method. In this study, we chose the direct training method to optimize the network model, aiming to fully leverage its advantages in inference efficiency and dynamic data processing.

2.2 Energy efficient object detection methods

Currently, mainstream deep learning-based object detection frameworks can be broadly categorized into two types: two-stage frameworks and one-stage frameworks. The two-stage framework is represented by the RCNN series (Girshick et al., 2014). These methods first generate region proposal boxes and subsequently classify and regress them. One-stage frameworks, including the YOLO series (Redmon and Farhadi, 2018), SSD (Wei et al., 2016), and Transformer-based models (Dosovitskiy et al., 2021), adopt a more direct approach by performing object detection and classification within a single network. Although existing ANN-based methods can achieve good target detection results, they all have high energy consumption problems. In practical applications, this high energy consumption limits its application in some scenarios with high energy consumption requirements. Therefore, more and more studies are trying to use SNN to provide an energy-efficient target detection solution with high efficiency and low energy consumption.

The initial approach to object detection using SNN adopted ANN-SNN based conversion method (Kim et al., 2020b,a; Li et al., 2022). This makes it take a long time to infer and is unsuitable for dynamic datasets captured by event cameras. Additionally, there are some hybrid architectures (Johansson, 2021; Lien and Chang, 2022) try to use directly trained SNN backbones and ANN detection heads for object detection. However, these detection heads introduce additional floating point multiplication operations, which destroy the spike-driven nature of SNNs. Moreover, such architectures are incompatible with certain neuromorphic hardware that exclusively supports spike-based operations (Davies et al., 2018; Akopyan et al., 2015; Liu et al., 2019). It was only in the past two years that fully directly trained deep SNN object detection networks have been successfully developed. For instance, EMS-YOLO (Su et al., 2023) represents one of the pioneering successes in this domain. However, it still faces limitations in fully extracting information from data, particularly in complex scenarios, leaving room for further optimization of the models performance.

2.3 Information loss in spiking neural networks

In order to improve the performance of SNN in different tasks, extensive research efforts have focused on mitigating information loss in SNNs, leading to the development of numerous methods

and strategies. One specific work (Kim et al., 2023) has analyzed the distribution of temporal dynamic information in SNNs during training by estimating the Fisher information of weights, uncovering the impact of temporal information concentration on SNN performance. In InfLoR-SNN (Guo et al., 2023), it is posited that the reset mechanism of SNN membrane potentials overlooks differences between potentials, leading to information loss. To address this, a Soft Reset mechanism and a Membrane Potential Rectifier are proposed to reduce errors. IM-Loss (Guo et al., 2022) suggests that the spike quantization process in SNNs causes information loss and diminished accuracy. To counteract this, it introduces an information maximization loss function designed to optimize information flow within SNNs. MSAT (He et al., 2024) posits that the uniform response of a constant threshold to varying inputs may result in information loss. To mitigate this, it introduces a multi-stage adaptive threshold mechanism to dynamically adjust membrane potential and input thresholds, thereby reducing information loss.

While the aforementioned studies have advanced efforts to mitigate information loss in SNNs, they still quantize membrane potentials into binary spikes. Some works have attempted to leverage non-binary, bidirectional spikes to enhance SNN performance. For example, Spiking-YOLO (Kim et al., 2020b) recognizes that the negative activation regions in leaky-ReLU-based ANN networks occupy a substantial portion of the network, and proposes neurons capable of emitting both positive and negative activations to compensate for information loss during ANN-to-SNN conversion. Similarly, Ternary Spike (Guo et al., 2024) introduces ternary neurons that emit +1 and -1 spikes, carrying richer information. It further proposes trainable spike amplitudes, allowing the network to represent different information during training and convert to standard ternary SNNs during inference. Although these approaches employ bidirectional spiking neurons, they do not provide a theoretical quantitative analysis of the information-carrying capabilities of spiking neurons versus ANN activations, nor do they quantitatively assess the contribution of bidirectional spikes in reducing information loss during inference, particularly for complex tasks such as object detection that require rich information representation.

3 Methodology

3.1 Information loss in spiking neural networks for object detection

While directly trained SNNs have demonstrated comparable performance to ANNs with reduced power consumption in tasks like object classification, their performance in object detection remains suboptimal when compared to ANNs. We posit that a key limitation lies in the binary spike feature map's insufficient capacity to convey the requisite information for object detection and complex regression tasks, leading to information loss and reduced accuracy. To verify our hypothesis, we employed information entropy theory, integrating it with the membrane potential dynamics at each layer during binary spike neuron inference for theoretical analysis and experimental verification.

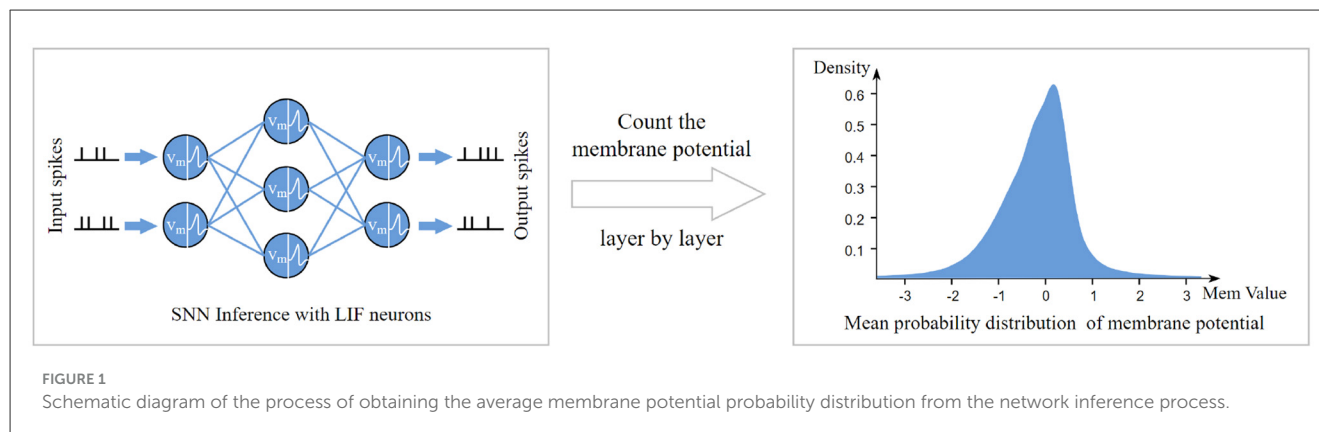


TABLE 1 Information entropy of LIF and ReLU activation function in different calculation ranges.

| Calculation range | $H(FB)$ | $H(FR)$ | $H(FR)/H(FB)$ |
|-------------------|--------------|--------------|---------------|
| Shallow layers | 0.715 | 2.207 | 3.087 |
| Middle layers | 0.695 | 2.152 | 3.096 |
| Deep layers | 0.833 | 2.486 | 2.984 |
| Global average | 0.746 | 2.281 | 3.056 |

$H(FB)$ denotes the information entropy of LIF neurons, while $H(FR)$ denotes the information entropy of the ReLU activation function.

Bold values highlight our findings regarding the information entropy under different activation methods.

The information expression capabilities $R(X)$ and $R(Y)$ of discrete random variables X and continuous random variables Y can be quantified by the information entropy $H(X)$ and $H(Y)$ of X and Y , respectively, which are expressed by the following Equations 1, 2.

$$R(X) = H(X) = - \sum_i P(x_i) \log_b P(x_i) \quad (1)$$

$$R(Y) = H(Y) = - \int_{-\infty}^{+\infty} P(y) \log_b P(y) \quad (2)$$

Here, $P(x_i)$ represents the probability of the random variable X taking the value x_i , $P(y)$ denotes the probability density function of Y , and b is the logarithmic base, typically set to 2, indicating that the information is measured in bits. Building upon the aforementioned formula, we investigated the disparity in information capacity between binary spike neurons and ReLU activation functions in activating feature maps during inference, employing both qualitative analysis and quantitative evaluation.

From a qualitative analysis perspective, the binary spike output is restricted to two states, 0 and 1, which limits its information entropy. Let FB denote the binary spike feature map, where $FB \in B^{C \times H \times W}$. For each pixel, the output of the spike feature map is binary (0 or 1), and thus the entropy of each pixel, $H(FB)$, can be expressed as Equation 3.

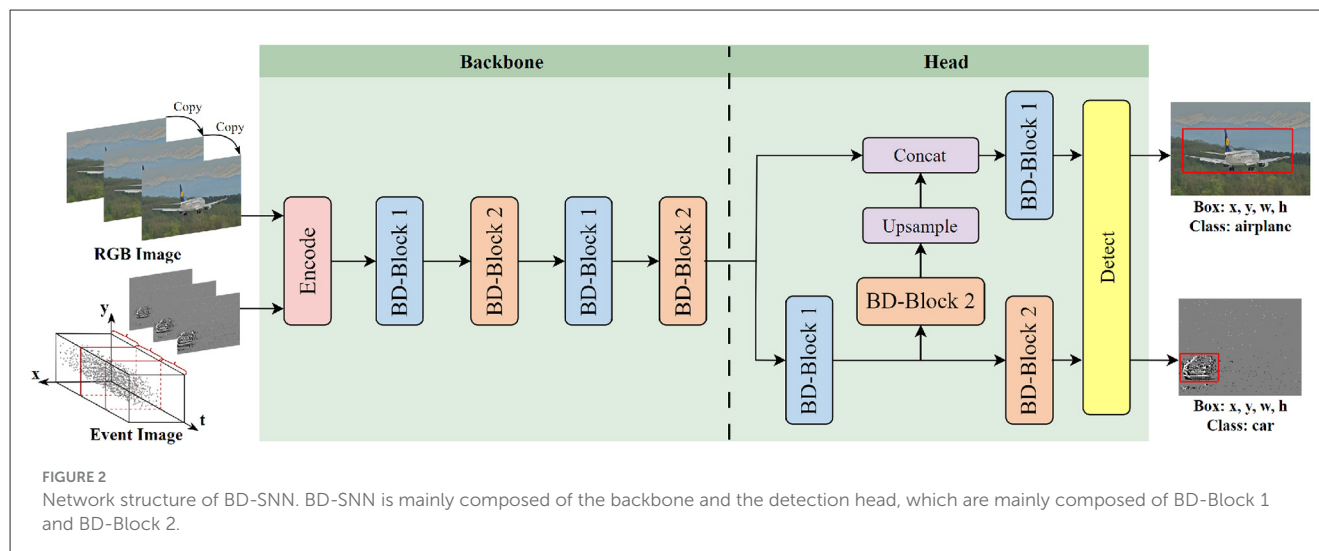
$$H(FB) = -P(0) \log_2 P(0) - P(1) \log_2 P(1) \quad (3)$$

The output of the spike feature map is limited to two discrete states, hindering its capacity to represent complex information. Consequently, it is unable to effectively represent multi-class data or dynamically changing information, leading to information loss. In contrast, the output of the ReLU activation function is a continuous, non-negative real value, allowing it to convey richer information. Let FR represent the ReLU activation feature map, where $FR \in B^{C \times H \times W}$. Since each pixel in the membrane potential map can assume any real value, the entropy of the feature map activated by the ReLU function is continuous within the non-negative range, necessitating the use of a probability density function for its definition. The entropy $H(FR)$ of each pixel can thus be expressed as Equation 4.

$$H(FR) = -P(0) \log_2 P(0) - \int_0^{+\infty} P(x) \log_2 P(x) \quad (4)$$

The probability density function $P(x)$ for positive values must be determined by statistically analyzing their distribution. While the specific information content of the ReLU activation feature map depends on the probability distribution of the membrane potential, it can theoretically represent more information than the binary spike feature map, as it outputs non-negative continuous values.

When performing quantitative calculations, in order to avoid the sampling error caused by relying solely on the distribution of values from a single layer's membrane potential, we recorded the membrane potential distributions layer by layer during network inference, as illustrated in Figure 1. Subsequently, we divided the membrane potential distributions into three groups, corresponding to the shallow, middle, and deep layers of the network, and averaged them separately to obtain average membrane potential probability distributions for each depth. These averaged distributions were then substituted into Equations 3, 4 to compute the information entropy of the feature maps obtained using LIF neurons and ReLU activation functions. The results are presented in Table 1, indicate that when LIF neurons were used for activation, the average information capacity per pixel was only 0.746 bits, whereas with ReLU activation, it increased to 2.281 bits per pixel. Comparatively, the information capacity of the ReLU-activated feature map was 3.056 times that of the binary spike feature map. This finding underscores the limitations of binary spike feature maps in information representation compared to ReLU-activated feature



maps, as quantizing real-valued membrane potentials into binary spikes results in approximately threefold information loss.

3.2 Design of BD-SNN network

To effectively represent the information in the input features and perform efficient object detection tasks, we propose the BD-SNN network. The BD-SNN network can process both frame-based RGB image data and event image data captured by event cameras for inference, utilizing 1 or -1 spikes to enhance information representation. This feature is mainly based on the design and implementation of the bidirectional variable threshold spike neuron (BD-LIF), the details of which are presented in Section 3.3.

3.2.1 Input representation

Static image inputs: Given the spatiotemporal nature of SNNs, when the network input consists of frame-based RGB images, we replicate the image across the time dimension, ensuring that it serves as input at each time step. This method fully leverages the spatiotemporal information processing capabilities of SNNs.

Event-based inputs: Event cameras operate fundamentally differently from frame cameras. Each pixel in an event camera can independently respond to changes in light and output these changes in the form of an asynchronous event stream, representing variations in the logarithmic brightness of the pixel. This enables an exceptionally wide dynamic range and superior temporal resolution in the microsecond range. When the logarithmic light level of a pixel exceeds the threshold V_{th} , an event $e_n = (x_n, y_n, t_n, p_n)$ is generated. Here, x_n and y_n represent the pixel coordinates, t_n is the timestamp of the event, and the polarity $p_n \in \{-1, 1\}$ represents an increase or decrease in light intensity.

Given a time window ζ , the asynchronous event stream $E = \{e_n \in \zeta : n = 1, \dots, N\}$ represents sparse event points in 3D space. In this work, we partition E into segments with a fixed time window dt and map the event points to a 2D representation resembling

an image within each segment. The event image generated in each segment is fed into the network as a time step during inference.

3.2.2 Network structure

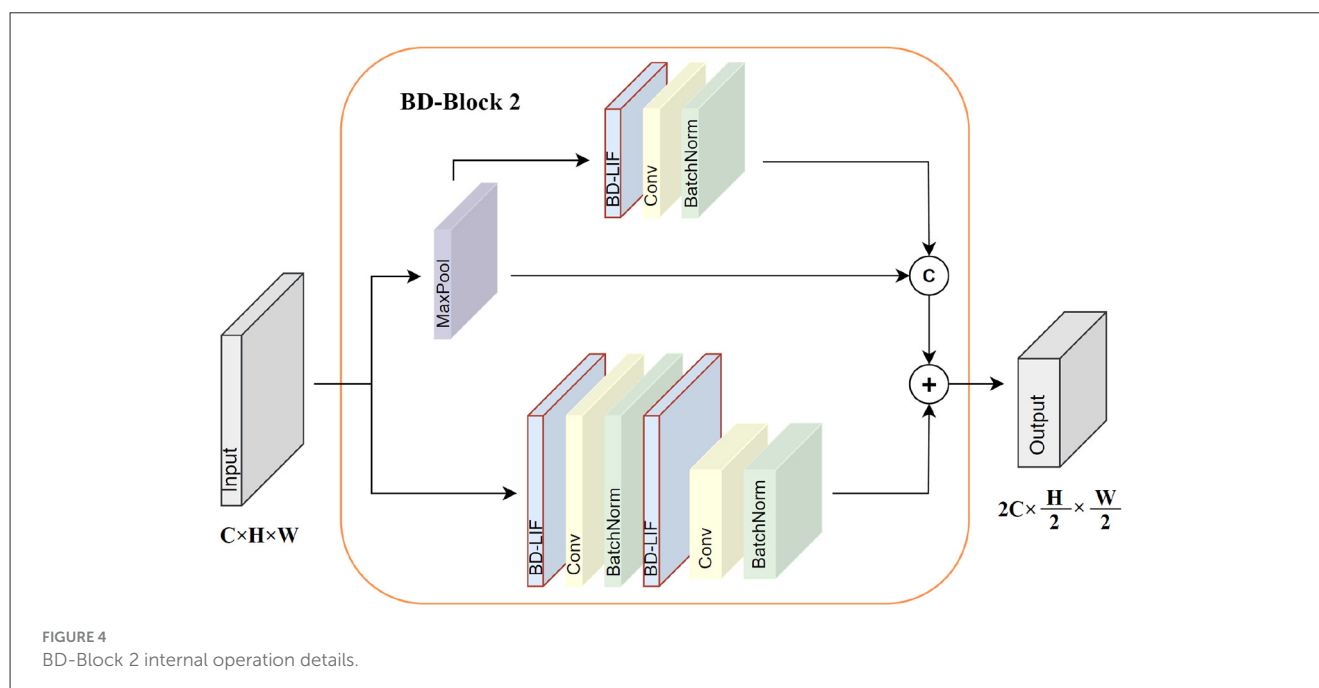
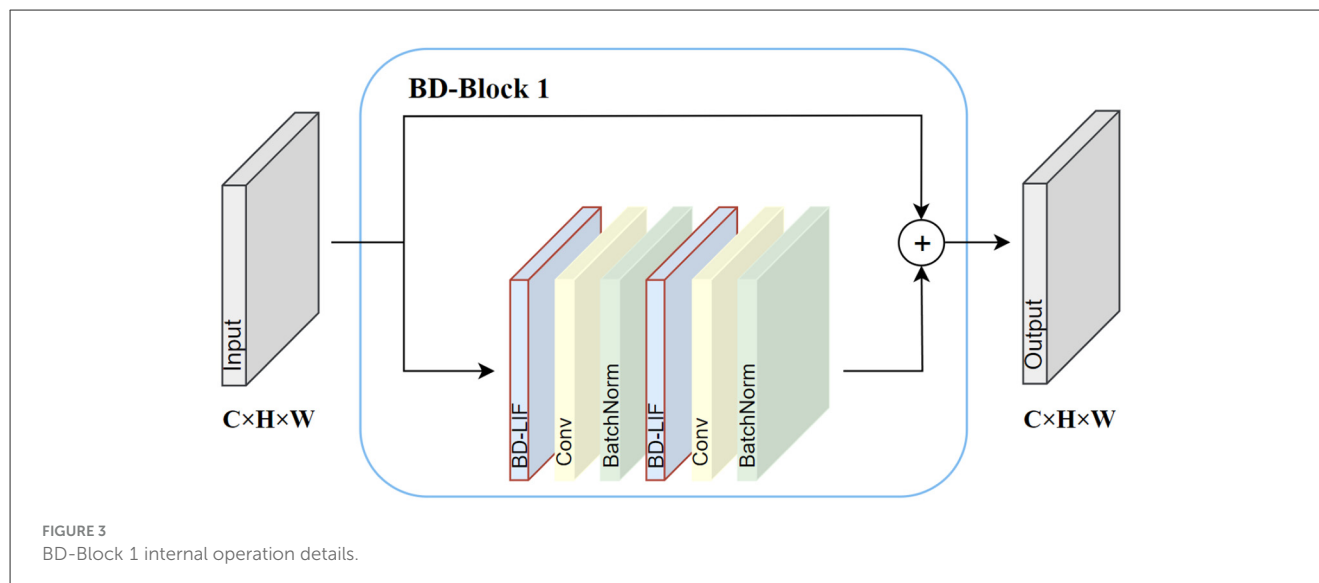
As shown in Figure 2, BD-SNN consists of two main components: the backbone for feature extraction and the detection head for object detection. Upon input of the RGB or event image into the network, it is first processed by an encoding layer, which includes a convolutional layer and a normalization layer, to convert the input into spikes. Subsequently, a series of BD-Blocks perform information extraction and feature fusion on the resulting spike activation map. Specifically, the BD-LIF neuron learns and integrates weighted inputs from different layers, emitting positive or negative spikes once the threshold is reached, based on the accumulation of membrane potential. We used two different BD-Blocks to extract and fuse features across different dimensions and channels, thereby enhancing the network's robustness. To address the issue of non-differentiable spikes during backpropagation and enable direct training of the network, we employ an alternative gradient (Wu et al., 2018), expressed as Equation 5.

$$\frac{\partial X_i^{t,n}}{\partial V_i^{t,n}} = \text{sign} \left(\|v_i^{t,n} - v_{th}\| < a \right) \quad (5)$$

where a is used to limit the range over which the gradient can propagate.

For the object detection task of the SNN model, the main challenge lies in accurately mapping the features extracted from the spike sequence to continuous bounding box coordinate representations. In this study, we feed the final membrane potential of the neuron into the detector to generate anchor boxes of varying scales. After applying Non-Maximum Suppression (NMS), we obtain the category and bounding box coordinates for each object.

BD-SNN consists primarily of two core modules: BD-Block 1 and BD-Block 2. The operational details of BD-Block 1 are depicted in Figure 3. BD-Block 1 is primarily used to extract features from the spike map, maintaining the input size by performing



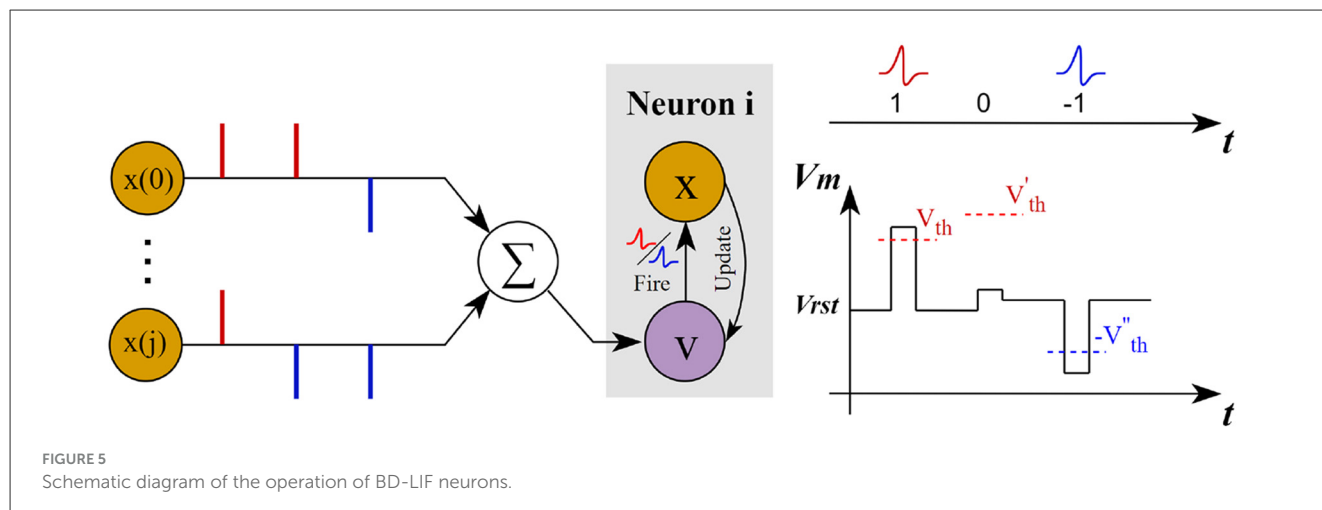
two consecutive spike activations, convolutions, and normalization operations. Residual connections are also introduced to enhance gradient propagation, enabling more effective training of deeper networks. In addition to the feature extraction function, BD-Block 2 is also responsible for fusing multi-scale features and downsampling the spike map. The specific operation details are shown in Figure 4. BD-Block 2 adopts a CSP-inspired dual-branch structure: one doubles the number of channels in the input spike map and reduces its size by half through convolution operations, while the other branch combines multi-scale features via pooling, convolution, and channel concatenation, resulting in an output with the same shape as the first branch. Finally, the outputs from both branches are combined to produce the final output, integrating multi-scale information.

Unlike EMS-Block (Su et al., 2023), which employs conventional LIF neurons generating unidirectional binary

spikes with fixed thresholds, both BD-Block 1 and BD-Block 2 are built upon our proposed BD-LIF neurons. BD-LIF introduces a bidirectional spiking mechanism (outputting $-1, +1$ instead of $0, 1$) and a learnable dynamic threshold adaptation strategy. These neuron-level innovations significantly increase the information capacity of the spike feature maps and alleviate activation inefficiency in deeper layers.

3.3 Bidirectional dynamic threshold neuron model

As analyzed in Section 3.1, the conversion of membrane potentials into binary spikes introduces substantial quantization errors, significantly limiting the network model's expressive



As shown in [Figure 5](#), BD-LIF neurons update the membrane potential by integrating both positive and negative spike information across different time steps, while dynamically adjusting their activation threshold. Once the membrane potential surpasses the activation threshold, the neuron will emit either positive or negative spikes accordingly. The detailed neuron model is described by the following [Equation 6](#).

$$V_i^{t+1,n+1} = \tau V_i^{t,n+1} \left(1 - X_i^{t,n+1}\right) + \sum_j W_{ij}^n X_j^{t+1,n} \quad (6)$$

$$V'_{th} = V_{th} - \alpha \cdot \tanh(\psi - b) \quad (7)$$

$$\psi = \frac{|\Delta V_i|}{\Delta t} = \frac{|V_i^t - V_i^{t-1}|}{\Delta t} \quad (8)$$

$$X_i^{t+1,n+1} = \begin{cases} 1, & \text{if } V_i^{t+1,n+1} \geq V'_{th} \\ -1, & \text{if } V_i^{t+1,n+1} \leq -V'_{th} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

As shown in Figure 6, BD-LIF neurons can not only emit both positive and negative spikes but also adaptively adjust their thresholds according to input variations. For instance, in example (a), when the membrane potential increases significantly at the second time step, the neuron’s activation threshold decreases accordingly. Conversely, when the membrane potential increase is

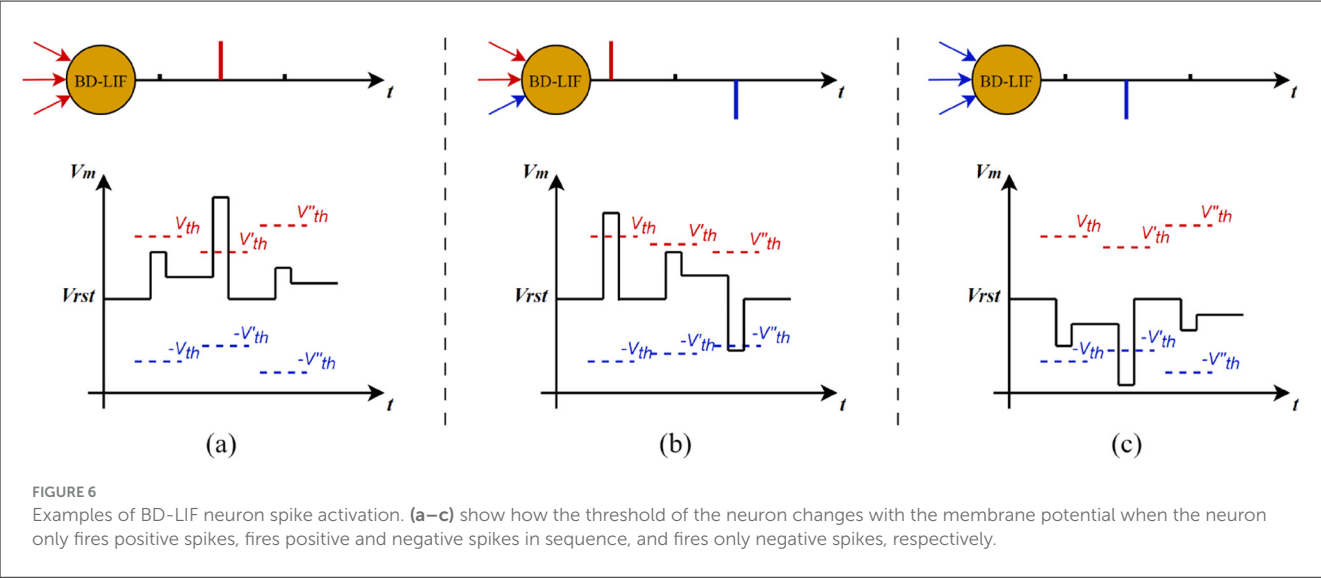


TABLE 2 Information entropy of LIF and BD-LIF activation function in different calculation ranges.

| Calculation range | $H(FB)$ | $H(FL)$ | $H(FL)/H(FB)$ |
|-------------------|--------------|--------------|---------------|
| Shallow layers | 0.715 | 1.524 | 2.131 |
| Middle layers | 0.695 | 1.519 | 2.186 |
| Deep layers | 0.833 | 1.657 | 1.989 |
| Global average | 0.746 | 1.566 | 2.099 |

$H(FB)$ denotes the information entropy of LIF neurons, while $H(FL)$ denotes the information entropy of the BD-LIF neurons. Bold values highlight our findings regarding the information entropy under different activation methods.

smaller at the third time step, the activation threshold of the neuron increases.

We hypothesize that BD-LIF neurons enhance the information capacity of spike-activated feature maps. Using the information entropy theory-based method introduced in Section 3.1, we further compare the information capacity of networks employing BD-LIF neurons and conventional LIF neurons across shallow, intermediate, and deep layers. As shown in Table 2, BD-LIF neurons achieve an average information capacity of 1.57 bits/pixel, which is 2.1 times higher than that of traditional LIF neurons. This layer-wise analysis provides quantitative evidence that BD-LIF neurons substantially enhance the feature expressiveness of SNNs during inference.

4 Experiment

4.1 Implementation details

In order to thoroughly evaluate the effectiveness of the proposed network, we performed experiments on both the static COCO2017 dataset (Lin et al., 2014) and the dynamic Gen1 dataset (De Tournemire et al., 2020) using the BD-SNN network, and compared its performance with several other state-of-the-art methods.

In all of our experiments, we set the number of detection heads to 2 in order to ensure a fair comparison with prior work. For the BD-LIF neurons, we set the reset potential $V_{res} = 0$, the initial threshold $V_{th} = 0.5$ (Su et al., 2023), and the membrane potential decay factor $\tau = 0.25$. The model was trained on two NVIDIA RTX2080Ti GPUs using Python 3.8 and the PyTorch 1.11 deep learning framework, with the SGD optimizer and the cross-entropy loss function. The learning rate was set to 1×10^{-2} . The network was trained on the COCO2017 dataset for 200 epochs with a batch size of 8, while the model was trained on the Gen1 dataset for 100 epochs with a batch size of 16.

4.2 Quantitative evaluation

We conduct a comparative analysis of our network against previous works, evaluating its performance on both the frame-based COCO2017 dataset and the event-based Gen1 dataset.

As shown in Table 3, on the COCO dataset, the previous best method EMS-YOLO, which uses EMS-ResNet34 as the backbone, achieves the highest mAP@0.5 of 0.501 and mAP@0.5:0.95 of 0.301. BD-SNN improves the performance to 0.532 and 0.327 with fewer time steps, representing an effective improvement of 3.1%. As shown in Table 4, on the Gen1 dataset, we trained our model using the same number of time steps as the comparison method, ultimately surpassing it by 2.8%. These experimental results show the advantages of BD-SNN in terms of speed and accuracy.

4.3 Ablation studies

We performed ablation experiments to assess the impact of the proposed BD-LIF components on both static and dynamic datasets. Specifically, we conducted experiments on the COCO and Gen1 datasets to evaluate the effects of BD-SNN using traditional LIF neurons, neurons with bidirectional spike features

TABLE 3 Results on the COCO dataset.

| Architecture | Model | Param(M) | Time step | mAP@0.5 | mAP@0.5:0.95 |
|----------------------|------------------------------------|----------|-----------|--------------|--------------|
| ANN | Tiny-Yolo | 4.3 | / | 0.258 | - |
| | ResNet18 | 11.2 | / | 0.478 | 0.299 |
| | ResNet34 | 21.8 | / | 0.497 | 0.319 |
| | SSD513 (Wei et al., 2016) | 36.0 | / | 0.504 | 0.312 |
| ANN-SNN Conversion | SNN-VGG16 | 20.0 | 64 | 0.359 | 0.192 |
| | SNN-InceptionV3 | 23.9 | 64 | 0.413 | 0.237 |
| | Spiking-Yolo (Kim et al., 2020b) | 10.2 | 3500 | - | 0.257 |
| | Bayesian Optim (Kim et al., 2020a) | 10.2 | 5000 | - | 0.259 |
| | Spike Calib (Li et al., 2022) | 17.1 | 512 | 0.454 | 0.259 |
| Directly-trained SNN | SpikeDet (Fan et al., 2025) | 22.0 | 4 | 0.626 | 0.462 |
| | SG ResNet (Zhang H. et al., 2023) | 22.5 | 4 | 0.484 | 0.296 |
| | EMS-YOLO (Su et al., 2023) | 26.8 | 4 | 0.501 | 0.301 |
| | M-SpikeFormer (Yao et al., 2024) | 75.0 | 4 | 0.503 | - |
| | BD-SNN (Ours) | 27.0 | 3 | 0.532 | 0.327 |

“/” denotes that the attribute is not applicable to the corresponding method, while “-” indicates that the value is not reported in the reference. Bold values emphasize the results obtained by our proposed method.

TABLE 4 Results on the Gen1 dataset.

| Architecture | Model | Param(M) | Time step | mAP@0.5 | mAP@0.5:0.95 |
|--------------|---|----------|-----------|--------------|--------------|
| ANN | YOLOv3-tiny (Redmon, 2018) | 10.2 | / | 0.312 | - |
| | Asynet (Messikommer et al., 2020) | 11.4 | / | - | 0.145 |
| | Aegnn (Schaefer et al., 2022) | 20.0 | / | - | 0.163 |
| | Inception+SSD (Iacono et al., 2018) | - | / | - | 0.301 |
| | MatrixLSTM (Cannici et al., 2020) | 61.5 | / | - | 0.310 |
| SNN | SpikeDet (Fan et al., 2025) | 22.0 | 5 | 0.692 | 0.465 |
| | MobileNet-64+SSD | 24.3 | 5 | - | 0.147 |
| | VGG-11+SSD | 12.6 | 5 | - | 0.174 |
| | DenseNet121-24+SSD (Cordone et al., 2022) | 8.2 | 5 | - | 0.189 |
| | Spiking-Yolo (Kim et al., 2020b) | 7.9 | 500 | 0.442 | - |
| | Tr-Spiking-Yolo (Yuan et al., 2024) | 7.9 | 5 | 0.453 | - |
| | EMS-YOLO (Su et al., 2023) | 14.4 | 5 | 0.541 | 0.311 |
| | SFOD (Fan et al., 2024) | 11.9 | 5 | - | 0.321 |
| | BD-SNN(Ours) | 14.5 | 5 | 0.569 | 0.326 |

“/” denotes that the attribute is not applicable to the corresponding method, while “-” indicates that the value is not reported in the reference. Bold values emphasize the results obtained by our proposed method.

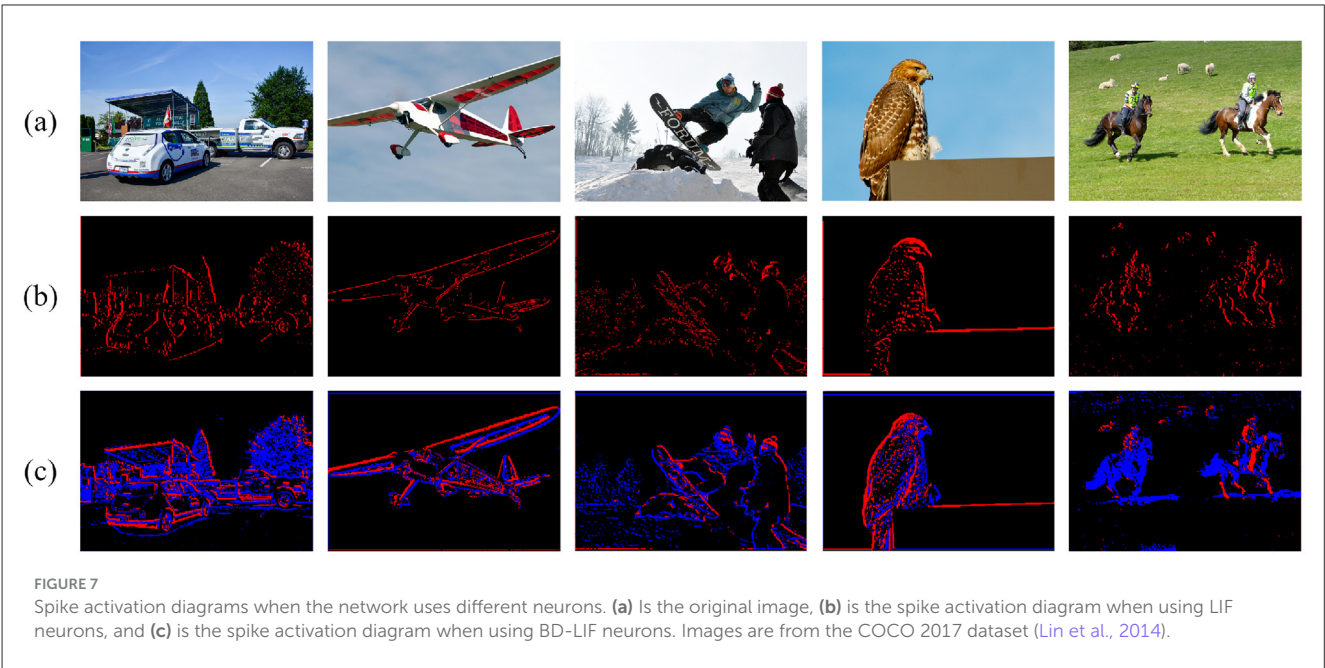
only, neurons with dynamic threshold features only, and neurons incorporating both features simultaneously. The experimental results are presented in Table 5. The results show that each component of the BD-LIF neuron model contributes to the enhancement of network performance. This further verifies that the SNN network using the proposed BD-LIF neurons is more competitive when processing traditional frame image data and event data.

4.4 Qualitative analysis

To intuitively illustrate that BD-LIF enhances the information-carrying capacity compared to traditional LIF neurons during the network inference process, we analyzed the inference results of networks utilizing different types of neurons on static datasets. Specifically, we extracted several activation maps from the shallow layers of the BD-SNN network and visualized these feature maps,

TABLE 5 Ablation study of BD-LIF on each characteristic.

| Dataset | Bidirectional spike | Dynamic threshold | Results | |
|---------|---------------------|-------------------|---------|--------------|
| | | | mAP@0.5 | mAP@0.5:0.95 |
| COCO | | | 0.498 | 0.302 |
| | ✓ | | 0.525 | 0.323 |
| | | ✓ | 0.506 | 0.303 |
| | ✓ | ✓ | 0.532 | 0.327 |
| Gen1 | | | 0.537 | 0.310 |
| | ✓ | | 0.560 | 0.321 |
| | | ✓ | 0.548 | 0.315 |
| | ✓ | ✓ | 0.569 | 0.326 |



as shown in Figure 7. It can be seen that, compared to traditional LIF neurons, BD-LIF neurons can flexibly utilize two distinct spike modes to represent key image details with greater precision. For instance, in the third column, the network using LIF neurons failed to accurately identify the body contours of the two individuals in the image, leading to the loss of critical details. The network employing BD-LIF neurons not only clearly delineated the body contours of the two individuals through two distinct spike forms, but also precisely captured the ski boundaries. These results demonstrate that BD-LIF neurons enhance the network's sensitivity to fine details and complex features, thereby facilitating superior performance in object detection tasks.

5 Conclusion

In this paper, we propose a Bidirectional Dynamic threshold SNN named by BD-SNN, which is capable of emitting both 1 and -1 spikes to convey richer information during inference. This addresses the challenge of limited feature information extraction from neuronal membrane potentials in traditional SNN-based

object detection networks, thereby effectively enhancing detection accuracy. Specifically, we designed two new all-spike residual blocks, termed BD-Block, to efficiently extract information and fuse features, and incorporated a new type of spiking neuron, BD-LIF, within these blocks. This neuron can emit both 1 and -1 spikes to enhance the information capacity of the spike activation feature map, while also adaptively adjusting its threshold in a biologically plausible manner to optimize activation efficiency. Experimental results show that BD-SNN improves the accuracy of state-of-the-art method (EMS-YOLO) by 3.1% and 2.8% on the COCO2017 and Gen1 datasets, respectively. We believe that this work provides new insights into enhancing SNN performance in object detection and further expands the potential of spiking neurons for spatiotemporal information processing.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

SW: Conceptualization, Methodology, Software, Visualization, Writing – original draft. GW: Project administration, Supervision, Writing – review & editing. YS: Funding acquisition, Resources, Writing – review & editing. YZha: Investigation, Writing – review & editing. YZho: Visualization, Writing – review & editing. QM: Writing – review & editing. YL: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work is sponsored by the Beijing Nova Program (20220484097, 20240484703), the National Natural Science Foundation of China General Program (82272130), the National Natural Science Foundation of China Key Program (U22A20103), and the Aeronautical Science Foundation (2023Z019072001, 2024Z074072001).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships

References

- Akopyan, F., Sawada, J., Cassidy, A., Alvarez-Icaza, R., Arthur, J., Merolla, P., et al. (2015). Truenorth: design and tool flow of a 65 mw 1 million neuron programmable neurosynaptic chip. *IEEE Trans. Comput.-Aided Des. Integr. Circ. Syst.* 34, 1537–1557. doi: 10.1109/TCAD.2015.2474396
- Al-Hamid, A. A., and Kim, H. (2020). Optimization of spiking neural networks based on binary streamed rate coding. *Electronics* 9:1599. doi: 10.3390/electronics9101599
- Andrew, A. M. (2003). Spiking neuron models: single neurons, populations, plasticity. *Kybernetes* 32:3. doi: 10.1108/k.2003.06732gae.003
- Azouz, R., and Gray, C. M. (2000). Dynamic spike threshold reveals a mechanism for synaptic coincidence detection in cortical neurons in vivo. *Proc. Nat. Acad. Sci.* 97, 8110–8115. doi: 10.1073/pnas.130200797
- Bu, T., Fang, W., Ding, J., Dai, P., Yu, Z., and Huang, T. (2023). Optimal ann-snn conversion for high-accuracy and ultra-low-latency spiking neural networks. *arXiv preprint arXiv:2303.04347*.
- Cannici, M., Ciccone, M., Romanoni, A., and Matteucci, M. (2020). “A differentiable recurrent surface for asynchronous event-based data,” in *European Conference on Computer Vision* (Springer), 136–152. doi: 10.1007/978-3-030-58565-5_9
- Cordone, L., Miramond, B., and Thierion, P. (2022). “Object detection with spiking neural networks on automotive event data,” in *2022 International Joint Conference on Neural Networks (IJCNN)* (IEEE), 1–8. doi: 10.1109/IJCNN55064.2022.9892618
- Davies, M., Srinivasa, N., Lin, T.-H., Chinya, G., Cao, Y., Choday, S. H., et al. (2018). Loihi: a neuromorphic manycore processor with on-chip learning. *IEEE Micro* 38, 82–99. doi: 10.1109/MM.2018.112130359
- De Tournemire, P., Nitti, D., Perot, E., Migliore, D., and Sironi, A. (2020). A large scale event-based detection dataset for automotive. *arXiv preprint arXiv:2001.08499*.
- Deng, L., Wu, Y., Hu, X., Liang, L., Ding, Y., Li, G., et al. (2020). Rethinking the performance comparison between snns and anns. *Neural Netw.* 121, 294–307. doi: 10.1016/j.neunet.2019.09.005
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2021). “An image is worth 16x16 words: Transformers for image recognition at scale,” in *International Conference on Learning Representations*, 1–15.
- Fan, Y., Liu, C., Li, M., Liu, D., Liu, Y., and Zhang, W. (2025). “SpikeDet: better firing patterns for accurate and energy-efficient object detection with spiking neuron networks. *arXiv e-prints*, page arXiv:2501.15151.
- Fan, Y., Zhang, W., Liu, C., Li, M., and Lu, W. (2024). “SFOD: spiking fusion object detector,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17191–17200. doi: 10.1109/CVPR52733.2024.01627
- Fontaine, B., Peña, J. L., and Brette, R. (2014). Spike-threshold adaptation predicted by membrane potential dynamics in vivo. *PLoS Comput. Biol.* 10:e1003560. doi: 10.1371/journal.pcbi.1003560
- Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 580–587. doi: 10.1109/CVPR.2014.81
- Guo, Y., Chen, Y., Liu, X., Peng, W., Zhang, Y., Huang, X., et al. (2024). Ternary spike: learning ternary spikes for spiking neural networks,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 12244–12252. doi: 10.1609/aaai.v38i11.29114
- Guo, Y., Chen, Y., Zhang, L., Liu, X., Wang, Y., Huang, X., et al. (2022). “Im-loss: information maximization loss for spiking neural networks,” in *Advances in Neural Information Processing Systems*, 156–166.
- Guo, Y., Chen, Y., Zhang, L., Wang, Y., Liu, X., Tong, X., et al. (2023). INFLO-SNN: Reducing information loss for spiking neural networks. *arXiv preprint arXiv:2307.04356*.
- He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., and Li, M. (2019). “Bag of tricks for image classification with convolutional neural networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 558–567. doi: 10.1109/CVPR.2019.00065
- He, X., Li, Y., Zhao, D., Kong, Q., and Zeng, Y. (2024). Msat: biologically inspired multistage adaptive threshold for conversion of spiking neural networks. *Neural Comput. Applic.* 36, 8531–8547. doi: 10.1007/s00521-024-09529-w
- Hopkins, M., Pineda-García, G., Bogdan, P. A., and Furber, S. B. (2018). Spiking neural networks for computer vision. *Interface Focus* 8:20180007. doi: 10.1098/rsfs.2018.0007
- Howard, A. G. (2017). Mobilenets: efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- Hu, Y., Deng, L., Wu, Y., Yao, M., and Li, G. (2024). Advancing spiking neural networks toward deep residual learning. *IEEE Trans. Neural Netw. Learn. Syst.* 36, 2353–2367. doi: 10.1109/TNNLS.2024.3355393
- Iacono, M., Weber, S., Glover, A., and Bartolozzi, C. (2018). “Towards event-driven object detection with off-the-shelf deep learning,” in *2018 IEEE*

that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- RSJ International Conference on Intelligent Robots and Systems (IROS), 1–9. doi: 10.1109/IROS.2018.8594119
- Johansson, O. (2021). *Training of object detection spiking neural networks for event-based vision*. Independent thesis Advanced level.
- Kim, S., Park, S., Na, B., Kim, J., and Yoon, S. (2020a). Towards fast and accurate object detection in bio-inspired spiking neural networks through bayesian optimization. *IEEE Access* 9, 2633–2643. doi: 10.1109/ACCESS.2020.3047071
- Kim, S., Park, S., Na, B., and Yoon, S. (2020b). Spiking-yolo: spiking neural network for energy-efficient object detection,” in *Proceedings of the AAAI conference on artificial intelligence*, 11270–11277. doi: 10.1609/aaai.v34i07.6787
- Kim, Y., Li, Y., Park, H., Venkatesha, Y., Hambitzer, A., and Panda, P. (2023). “Exploring temporal information dynamics in spiking neural networks,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 8308–8316. doi: 10.1609/aaai.v37i7.26002
- Li, Y., He, X., Dong, Y., Kong, Q., and Zeng, Y. (2022). Spike calibration: Fast and accurate conversion of spiking neural network for object detection and segmentation. *arXiv preprint arXiv:2207.02702*.
- Lien, H.-H., and Chang, T.-S. (2022). Sparse compressed spiking neural network accelerator for object detection. *IEEE Trans. Circ. Syst. I* 69, 2060–2069. doi: 10.1109/TCSI.2022.3149006
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). “Microsoft coco: common objects in context,” in *Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13* (Springer), 740–755. doi: 10.1007/978-3-319-10602-1_48
- Liu, Q., Richter, O., Nielsen, C., Sheik, S., Indiveri, G., and Qiao, N. (2019). “Live demonstration: face recognition on an ultra-low power event-driven convolutional neural network asic,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*.
- Liu, S., Wang, G., Song, Y., Huang, J., Huang, Y., Zhou, Y., et al. (2024). Siameft: adaptive-time feature extraction hybrid network for rgbe multi-domain object tracking. *Front. Neurosci.* 18:1453419. doi: 10.3389/fnins.2024.1453419
- Maass, W. (1997). Networks of spiking neurons: the third generation of neural network models. *Neural Netw.* 10, 1659–1671. doi: 10.1016/S0893-6080(97)00011-7
- Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., et al. (2014). A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* 345, 668–673. doi: 10.1126/science.1254642
- Messikommer, N., Gehrig, D., Loquercio, A., and Scaramuzza, D. (2020). “Event-based asynchronous sparse convolutional networks,” in *European Conference on Computer Vision* (Springer), 415–431. doi: 10.1007/978-3-030-58598-3_25
- Paninski, L. (2003). Estimation of entropy and mutual information. *Neural Comput.* 15, 1191–1253. doi: 10.1162/089976603321780272
- Poon, C.-S., and Zhou, K. (2011). Neuromorphic silicon neurons and large-scale neural networks: challenges and opportunities. *Front. Neurosci.* 5:108. doi: 10.3389/fnins.2011.00108
- Redmon, J., and Farhadi, A. (2018). Yolov3: an incremental improvement. *arXiv preprint arXiv:1804.02767*.
- Redmon, J., and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv e-prints*.
- Rueckauer, B., Lungu, I.-A., Hu, Y., Pfeiffer, M., and Liu, S.-C. (2017). Conversion of continuous-valued deep networks to efficient event-driven networks for image classification. *Front. Neurosci.* 11:682. doi: 10.3389/fnins.2017.00682
- Schaefer, S., Gehrig, D., and Scaramuzza, D. (2022). “AEGNN: asynchronous event-based graph neural networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12371–12381. doi: 10.1109/CVPR52688.2022.01205
- Shan, Y., Zhang, M., Zhu, R.-J., Qiu, X., Eshraghian, J. K., and Qu, H. (2024). Advancing spiking neural networks towards multiscale spatiotemporal interaction learning. *arXiv preprint arXiv:2405.13672*.
- Srinivasan, G., Lee, C., Sengupta, A., Panda, P., Sarwar, S. S., and Roy, K. (2020). “Training deep spiking neural networks for energy-efficient neuromorphic computing,” in *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE), 8549–8553. doi: 10.1109/ICASSP40776.2020.9053914
- Su, Q., Chou, Y., Hu, Y., Li, J., Mei, S., Zhang, Z., et al. (2023). “Deep directly-trained spiking neural networks for object detection,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6555–6565.
- Van Rullen, R., and Thorpe, S. J. (2001). Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Comput.* 13, 1255–1283. doi: 10.1162/08997660152002852
- Wei, L., Dragomir, A., Dumitru, E., Christian, S., Scott, R., Cheng-Yang, F., et al. (2016). *Ssd: Single Shot Multibox Detector*. Springer, Cham.
- Wu, X., Zhao, Y., Song, Y., Jiang, Y., Bai, Y., Li, X., et al. (2023). Dynamic threshold integrate and fire neuron model for low latency spiking neural networks. *Neurocomputing* 544:126247. doi: 10.1016/j.neucom.2023.126247
- Wu, Y., Deng, L., Li, G., Zhu, J., and Shi, L. (2018). Spatio-temporal backpropagation for training high-performance spiking neural networks. *Front. Neurosci.* 12, 331. doi: 10.3389/fnins.2018.00331
- Wu, Y., Deng, L., Li, G., Zhu, J., Xie, Y., and Shi, L. (2019). “Direct training for spiking neural networks: faster, larger, better,” in *Proceedings of the AAAI conference on artificial intelligence*, 1311–1318. doi: 10.1609/aaai.v33i01.33011311
- Yao, M., Hu, J., Hu, T., Xu, Y., Zhou, Z., Tian, Y., et al. (2024). Spike-driven transformer v2: Meta spiking neural network architecture inspiring the design of next-generation neuromorphic chips. *arXiv preprint arXiv:2404.03663*.
- Yuan, M., Zhang, C., Wang, Z., Liu, H., Pan, G., and Tang, H. (2024). Trainable spiking-yolo for low-latency and high-performance object detection. *Neural Networks* 172:106092. doi: 10.1016/j.neunet.2023.106092
- Zador, A. (1997). Spikes: Exploring the neural code. *Science* 277, 772–773. doi: 10.1126/science.277.5327.772b
- Zhang, H., Li, Y., He, B., Fan, X., Wang, Y., and Zhang, Y. (2023). *Direct training high-performance spiking neural networks for object recognition and detection*. Frontiers in Neuroscience. doi: 10.3389/fnins.2023.1229951
- Zhang, J., Wang, Y., Liu, W., Li, M., Bai, J., Yin, B., et al. (2023). “Frame-event alignment and fusion network for high frame rate tracking,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9781–9790. doi: 10.1109/CVPR52729.2023.00943
- Zhou, C., Zhang, H., Yu, L., Ye, Y., Zhou, Z., Huang, L., et al. (2024). Direct training high-performance deep spiking neural networks: a review of theories and methods. *Front. Neurosci.* 18:1383844. doi: 10.3389/fnins.2024.1383844
- Zhu, R.-J., Zhang, M., Zhao, Q., Deng, H., Duan, Y., and Deng, L.-J. (2024). Tcja-SNN: Temporal-channel joint attention for spiking neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* 36, 5112–5125. doi: 10.1109/TNNLS.2024.3377717