Check for updates

SLoN: a spiking looming perception network exploiting neural encoding and processing in ON/OFF channels

Zhifeng Dai†, Qinbing Fu*†, Jigen Peng and Haiyang Li

Machine Life and Intelligence Research Centre, School of Mathematics and Information Science, Guangzhou University, Guangzhou, China

Looming perception, the ability to sense approaching objects, is crucial for the survival of humans and animals. After hundreds of millions of years of evolutionary development, biological entities have evolved efficient and robust looming perception visual systems. However, current artificial vision systems fall short of such capabilities. In this study, we propose a novel spiking neural network for looming perception that mimics biological vision to communicate motion information through action potentials or spikes, providing a more realistic approach than previous artificial neural networks based on sum-then-activate operations. The proposed spiking looming perception network (SLoN) comprises three core components. Neural encoding, known as phase coding, transforms video signals into spike trains, introducing the concept of phase delay to depict the spatial-temporal competition between phasic excitatory and inhibitory signals shaping looming selectivity. To align with biological substrates where visual signals are bifurcated into parallel ON/OFF channels encoding brightness increments and decrements separately to achieve specific selectivity to ON/OFF-contrast stimuli, we implement eccentric down-sampling at the entrance of ON/OFF channels, mimicking the foveal region of the mammalian receptive field with higher acuity to motion, computationally modeled with a leaky integrate-and-fire (LIF) neuronal network. The SLoN model is deliberately tested under various visual collision scenarios, ranging from synthetic to real-world stimuli. A notable achievement is that the SLoN selectively spikes for looming features concealed in visual streams against other categories of movements, including translating, receding, grating, and near misses, demonstrating robust selectivity in line with biological principles. Additionally, the efficacy of the ON/OFF channels, the phase coding with delay, and the eccentric visual processing are further investigated to demonstrate their effectiveness in looming perception. The cornerstone of this study rests upon showcasing a new paradigm for looming perception that is more biologically plausible in light of biological motion perception.

1 Introduction

Looming perception is an essential ability of sighted animals that detects objects moving in depth, accordingly plays critical roles in their daily movements including escaping from predators, preying, and so forth. With the development of robotic technologies, the vast majority of mobile robots nowadays are capable of detecting and avoiding collisions through obtaining sensor information from their surroundings, for example, infrared (Benet et al., 2002), ladar (Manduchi et al., 2005), and ultrasonic (Nelson and MacIver, 2006). Meanwhile, vision-based sensing modalities are benefiting from their economy, energy saving, and high-dimension feature acquisition, thus gradually prevailing over other collision sensing techniques in mobile robotics (Fu et al., 2019) and ground vehicles (Mukhtar et al., 2015). However, in terms of reliability and robustness in complex and dynamic environments, current approaches are far from acceptable for serving human society.

On the other hand, humans and animals possess highly robust and efficient dynamic vision systems to handle looming perception in fast-changing visual environments, which has always been inspiring researchers to explore the biological visual systems in order to develop robust artificial vision systems for addressing real world challenges. In this regard, early studies of mammalian vision began at the last century of 1930s, in which Hartline pointed out that the function of brightness increase (ON) and decrease (OFF) in early visual motion is separated into parallel processing pathways (Hartline, 1938). This means that the visual neural systems evolved to split motion information in parallel for better adapting to dynamic environments, with probably more energy consumption. After that, neuro-scientists studied the receptive fields (RF) of cats and rabbits and, in the 1960s, discovered direction-selective neurons that are sensitive to ON/OFF moving stimuli along specific directions (Hubel and Wiesel, 1962; Barlow and Levick, 1965). In the 1980s, Schiller et al. summarized the function of ON and OFF channels in the mammalian visual system (Schiller et al., 1986); he pointed out that ON/OFF retinal ganglion cells constitute the two polarity pathways of early visual processing.

With advances in biotechnology, the mammalian visual system has been better understood in recent decades. Many studies have demonstrated how the mammalian visual systems detect looming objects by recording the rats' reactions to the sight of an approaching disc (Yilmaz and Meister, 2013; Busse, 2018; Lee et al., 2020). In addition, scientists found that photoreceptors in the mammalian visual systems are not evenly distributed in the retina, with rods and cones having a high density in the foveal region and decreasing toward the marginal regions (Harvey and Dumoulin, 2011; Wurbs et al., 2013). At the same time, the size of individual RF increases toward the periphery. This results in a biological vision that images the foveal region clearly, while the peripheral region is relatively blurry. These visual characteristics of mammals have also inspired modeling works accounting for motion perception (Borst and Helmstaedter, 2015; Fu, 2023). However, the underlying circuits and mechanisms of biological vision remain largely unknown.

Although the biological substrates are elusive, computational modeling is particularly useful for testing hypotheses on biological signal processing. Regarding computational modeling of looming perception visual systems, there are many methods inspired by insects' visual systems (Fu et al., 2019). Fu et al. introduced a series of visual neural networks based on the locust's lobula giant movement detector-2 (LGMD2) neuronal mechanism, mimicking its distinctive looming selectivity for darker objects approaching relative to the background through bio-plausible ON/OFF pathways (Fu and Yue, 2015; Fu et al., 2018). Owing to the efficiency in hardware implementation, the model was applied to the vision systems of micro-robots for quick collision detection in navigation (Fu et al., 2016). More specifically, the object movement will elicit the image brightness to change with respect to time. When the brightness increases, the stimulus will enter the ON pathway, while the brightness decreases will enter the OFF pathway. The significance of ON/OFF channels has been systematically investigated through a recent research upon implementing different selectivity in motion perception (Fu, 2023). Moreover, there are also many other models for motion detection that employ the ON/OFF channels, such as the elementary motion detectors (EMD) (Franceschini, 2014). Franceschini et al., for the first time, proposed the splitting of EMD inputs into ON-EMD and OFF-EMD structures encoding light and dark moving edges separately for micro-simulation of photoreceptors in RF (Franceschini et al., 1989). Subsequently, following advancements in physiological research on the fruit fly *Drosophila*, various computational models employing different combinations of ON/OFF-EMD were proposed to simulate the motion vision of the fly (Eichner et al., 2011; Joesch et al., 2013; Fu and Yue, 2020). Additionally, models for small target motion detectors (STMDs) were developed, drawing inspiration from insects such as dragonflies and hoverflies. In these models, the ON/OFF channels play an indispensable role in discriminating between small moving targets and cluttered, dynamic backgrounds (Halupka et al., 2011; Wiederman et al., 2013).

However, these bio-inspired motion perception models essentially belong to the second generation of artificial neural networks (ANN), where neurons within the network transmit and calculate real numbers, which are weighted, summed, and then activated for delivery. Such approaches are not as realistic as biological visual systems that encode external stimuli as action potentials or spikes for signal communication between neurons. Specifically, when the pre-synaptic neuron receives a stimulus large enough, it is charged to transmit a spike to communicate with the post-synaptic neuron. Encoding spikes with neuronal dynamics concerning time gives rise to the third generation of spiking neural networks (SNNs) (Tan et al., 2020). In terms of energy efficiency, the SNN undoubtedly prevails over the previous generations of ANN as the SNN transmits binary information without multiplication before summation. In addition, sparse and asynchronous signal processing can also be allowed in SNN (Lobo et al., 2020). The vast majority of SNN modeling works have been proposed for addressing real-world challenges including object detection (Cordone et al., 2022), pattern recognition (Kheradpisheh et al., 2018; Tavanaei et al., 2019), and perceptual system (Masuta and Kubota, 2010; Tan et al., 2020). Recently, Yang et al. (2022, 2023) and Yang and Chen (2023a,b) offered new insights into enhancing spike-based machine learning performance

using advanced information-theoretic learning methods. These studies have introduced novel frameworks to enhance SNN performance in specific tasks, allowing the model to achieve high-level intelligence, accuracy, robustness, and low power consumption compared with state-of-the-art artificial intelligence works.

In terms of motion perception, to the best of our knowledge, there are very few modeling works based on the SNN because decoding video signals to extract motion cues with temporal coherence is challenging. Salt (2016) and Salt et al. (2017, 2020) introduced a collision perception SNN implemented in an aerial robot. However, this study lacked systematic testing in a closed-loop flight scenario. Recently, the development of dynamic vision sensors (DVS) has significantly bolstered computer vision and vision-based robotic applications, e.g., Posch et al. (2014), Milde et al. (2015), Milde et al. (2017), Vasco et al. (2017), and Gallego et al. (2022). Compared with traditional, frame-based sensors, the DVS, such as, event-based camera features low latency, high speed, and high dynamic range. Most importantly, it can report ON (onset)/OFF (offset) motion events, which aligns much closer to the revealed principles of biological visual processing. Utilizing input from DVS, a novel spiking-EMD (sEMD) model was proposed for encoding ON/OFF motion events and decoding the direction of motion via the "time-of-travel" concept (Milde et al., 2018; D'Angelo et al., 2020). However, the cost of DVS is high, limiting its flexibility in mobile robotics due to its larger size and higher power requirements. Here, we aim to develop a more general method for SNN in motion perception that can be used with either frame-based or event-driven sensors.

Accordingly, this study introduces a novel SNN model, the Spiking Looming perception Network (SLoN), which offers two-fold advantages over previous ANN models. On one hand, the SLoN is more biologically realistic as it encodes visual streams into spikes for communication within the network. On the other hand, the SLoN mimics the receptive field (RF) of mammalian vision for looming perception with new bio-plausible concepts and mechanisms to achieve specific looming selectivity. The cornerstone of this study is based on the following aspects:

- In consonance with mammalian motion vision (Fu, 2023), the SLoN introduces ON/OFF channels that split the input signals into parallel pathways for neural encoding and down-sampling. This structure enables the SLoN to achieve varying looming selectivity to ON/OFF-contrast stimuli, improving robustness in complex and dynamic visual scenarios.
- Various neural coding models transform continuous signals into spatiotemporal spike trains, generally categorized as rate-based and temporal coding methods (Guo et al., 2021). Currently, neural coding is predominantly used to convert images into spike trains over a time window, with limited application in processing sequences of images for extracting motion information with temporal coherence. In the proposed SLoN model, phase coding is employed to convert video signals into spike trains as input. The phase coding method encodes video signals into spike trains using a binary representation, dividing each frame into eight periods with weighted spikes and emphasizing the significance of early

TABLE 1 Nomenclature of this study.

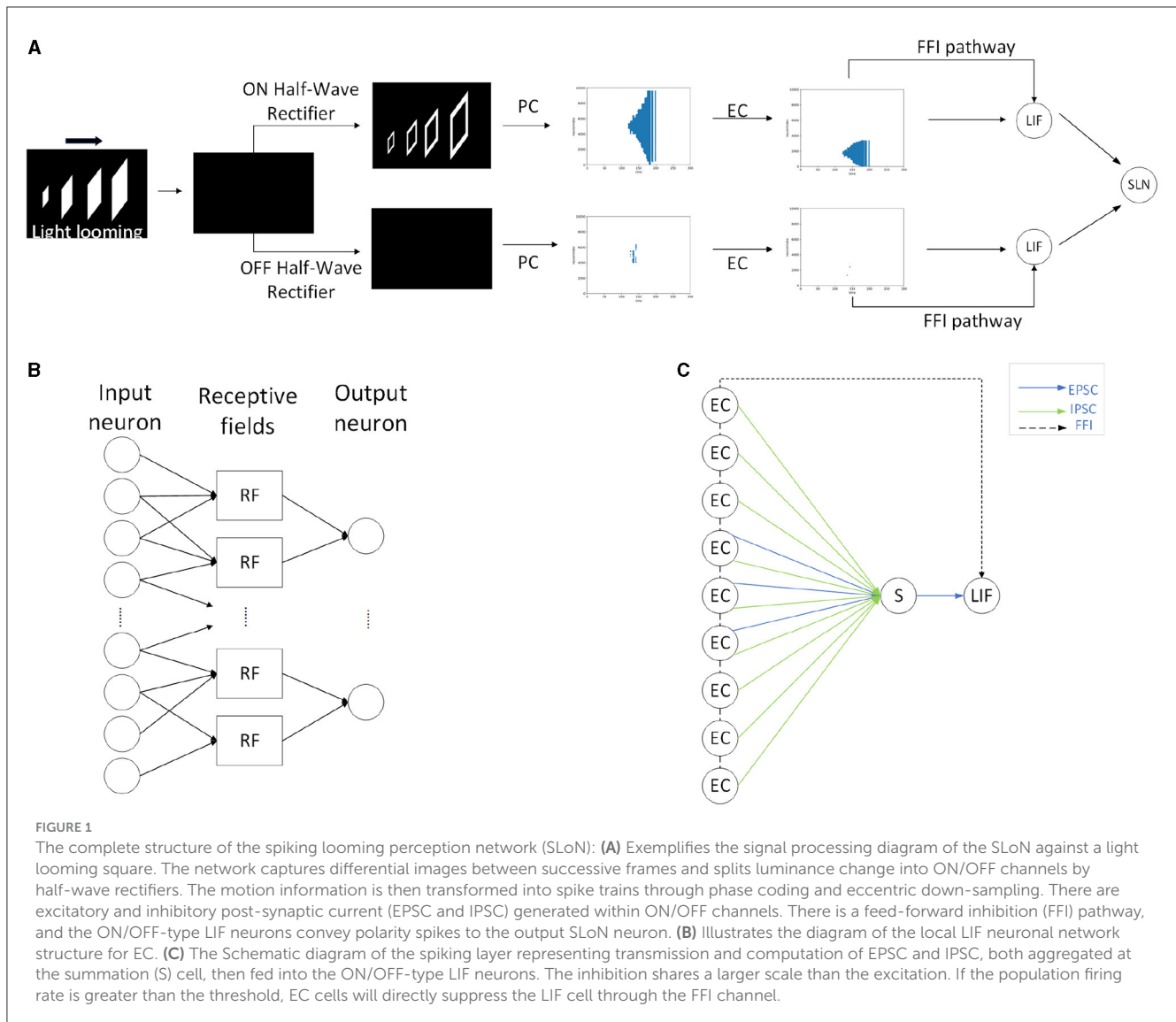|	
ANN	Artificial neural network
DVS	Dynamic vision sensor
EC	Eccentric down-sampling
EPSC	Excitatory post-synaptic current
FFI	Feed-forward inhibition
IPSC	Inhibitory post-synaptic current
LGMD	Lobula giant movement detector
LIF	Leaky integrate-and-fire
PC	Phase coding
PD	Phase delay
RF	Receptive field
sEMD	Spiking elementary motion detector
SLoN	Spiking looming perception network
SNN	Spiking neural network

spikes (Kim et al., 2018). Importantly, a new concept of phase delay (PD) is introduced to capture the spatiotemporal competition between excitatory and inhibitory currents within the ON/OFF channels, generating specific looming selectivity. The phase coding with delay mechanism not only combats noise but also portrays the fundamental characteristics of interactions between excitatory and inhibitory signals at a small time scale. This enables the proposed SNN model to process data from frame-based cameras.

- The mammalian receptive field (RF) does not sample as uniformly as traditional cameras (D'Angelo et al., 2020). To align with this, a strategy of eccentric down-sampling (EC) is adopted at the entry of ON/OFF channels of the proposed SLoN. The EC uses square regions to approximate the round retina of the animal, maintaining a quadrilateral camera resolution. Each RF spatially integrates information within its sensitive region through leaky integrate-and-fire (LIF) neuron models. The EC thus implements the foveal region with higher acuity to motion. Compared with previous study, this study delineates the EC for a better understanding of its computer implementation.

The rest of this study is structured as follows. Section 2 introduces the framework of SLoN with elaborated algorithms and illustrations, the setting of network parameters and experiments. Section 3 presents the systematic experimental results with analysis. Section 4 provides further discussions and concludes this research. The nomenclature used in this study is shown in Table 1.

2 Materials and methods

Within this section, we propose the methodology, including the framework of SLoN, the parameter setups, and the experimental settings. As the underlying mechanisms of biological looming

FIGURE 1
The complete structure of the spiking looming perception network (SLoN): **(A)** Exemplifies the signal processing diagram of the SLoN against a light looming square. The network captures differential images between successive frames and splits luminance change into ON/OFF channels by half-wave rectifiers. The motion information is then transformed into spike trains through phase coding and eccentric down-sampling. There are excitatory and inhibitory post-synaptic current (EPSC and IPSC) generated within ON/OFF channels. There is a feed-forward inhibition (FFI) pathway, and the ON/OFF-type LIF neurons convey polarity spikes to the output SLoN neuron. **(B)** Illustrates the diagram of the local LIF neuronal network structure for EC. **(C)** The Schematic diagram of the spiking layer representing transmission and computation of EPSC and IPSC, both aggregated at the summation (S) cell, then fed into the ON/OFF-type LIF neurons. The inhibition shares a larger scale than the excitation. If the population firing rate is greater than the threshold, EC cells will directly suppress the LIF cell through the FFI channel.

perception visual systems remain largely elusive, this SNN model is inspired by a few relevant studies. These include the modeling of ON/OFF channels in the locust's LGMD model to separate different looming selectivities to ON/OFF-contrast stimuli (Fu et al., 2020a), eccentric down-sampling methods mimicking the mammalian receptive field (D'Angelo et al., 2020), and phase coding (Kim et al., 2018) with the inaugural concept of phase delay to process and correlate video signals. The complete structure of SLoN is shown in Figure 1. The EC and PD are shown in Figures 2, 3, respectively.

2.1 Framework of SLoN

In general, the proposed SLoN consists of three layers: the neural encoding layer, the spiking interaction layer, and the output layer. Specifically, the main difference from the vast majority of SNNs is that the SLoN prioritizes motion retrieval from videos to respond to looming motion features rather than handling

singular images. Accordingly, within the first layer, the differential images between successive frames are split into ON/OFF channels encoding luminance increments/decrements, respectively. Phase coding is adopted to transform pixel intensity features into spatial-temporal spike trains, where each period is represented by eight phases. Importantly, the most significant information is retained in the first phases of each period, with the significance decaying by phase. The ON/OFF spike trains are then fed into an eccentric down-sampling mechanism, a partial neural network with LIF neurons, to simulate mammalian vision with higher acuity in the foveal region.

Within the second, spiking interaction layer, both EPSC and IPSC are generated within the ON/OFF channels. The inhibitory signals are temporally phase-delayed and spatially influenced by a larger field relative to the excitatory ones. This competition between them effectively responds most strongly to the expanding edges of objects over other types of motion stimuli. The ON/OFF-type LIF neurons integrate the corresponding EPSC and IPSC to generate polarity spikes. In addition, there are feed-forward inhibition (FFI)

pathways that can directly suppress the ON/OFF-type LIF neurons if either the population rate exceeds a threshold. At the last layer, the output SLoN neuron integrates spikes from the ON/OFF channels, which inherently is another LIF neuron.

2.1.1 ON/OFF motion retrieving

At the initial step, the SLoN does not handle every single image from a visual stream. Instead, motion information between two adjacent frames is retrieved. Assume $L(x, y, f) \in R^3$ is the fth frame image brightness value, where x, y denote spatial positions, we can obtain the motion information as

$$P(x, y, f) = L(x, y, f) - L(x, y, f - 1) \quad (1)$$

As movement inevitably induces brightness increment and decrement within receptive field, the motion information is spit into parallel ON/OFF channels. This is achieved by half-wave rectifiers where the brightness increments/decrements flow into ON/OFF channels, respectively, mathematically expressed as

$$P_{on}(x, y, f) = [P(x, y, f)]^+ + \alpha_1 P_{on}(x, y, f - 1) \quad (2)$$

$$P_{off}(x, y, f) = [P(x, y, f)]^- + \alpha_1 P_{off}(x, y, f - 1) \quad (3)$$

where α_1 is a coefficient, which stands for the residual information in time. $[x]^+$, $[x]^-$ are denoted as $max(x, 0)$ and $max(-x, 0)$, respectively.

2.1.2 Phase coding with weighted spikes

There are many neural coding methods as reviewed in the study by Guo et al. (2021), such as rate coding (Heeger, 2000), time-to-first-spike coding (Park et al., 2020), and burst coding (Eyherabide et al., 2009). Among these, phase coding was initiated in relation to the oscillatory firing pattern of neurons (O'Keefe and Recce, 1993; Laurent, 1996). Laurent conducted biological experiments on locusts, where a specific odor was directed toward one antenna of a locust. Laurent observed an oscillatory response in the mushroom body calyx of the locust. Building on this foundation, Kim et al. (2018) introduced an enhanced phase coding method with weighted spikes, emphasizing timing information. In this approach, the processing time of the stimulus is divided into k phases, and the stimulus is converted into binary values. Each phase is assigned a 0 or 1 (1 for a spike and 0 for no spike), with different weights assigned to each phase highlighting the importance of early information. The weight decays significantly concerning the phase time in each period, indicating that the most critical information is conveyed in the earlier phases. The proposed SLoN aligns with this coding strategy but introduces a novel concept of phase delay to accommodate spatial-temporal interactions between excitatory and inhibitory signals, aiming to shape the looming selectivity. Taking a few examples in Figure 2, the phase coding divides motion information, i.e., intensities of luminance change in ON/OFF channels into eight phases that constitute a time unit of computer processing. The weight of each phase is $2^{-1}, 2^{-2}, \ldots, 2^{-8}$, and then

the algorithm assigns 0, 1 to each phase with different weights. The computation can be defined as

$$C_{on}(x, y, 8(f - 1) + i) = mod(\frac{P_{on}(x, y, f)}{2^i}, 2), \ i = 0, 1, \ldots, 7 \quad (4)$$

$$C_{off}(x, y, 8(f - 1) + i) = mod(\frac{P_{off}(x, y, f)}{2^i}, 2), \ i = 0, 1, \ldots, 7 \quad (5)$$

where $mod(x, y)$ is the modulus function that divides x by y. We herein use time phase instead of video frame to represent time t. $C_{on/off}(x, y, 8(f - 1) + i)$ thus can be expressed as $C_{on/off}(x, y, t)$ in the following equations defined in this section. The ON/OFF-type spike trains can be obtained by assembling phasic spikes, spatially. Then, the weight of each phase in every period is computed as

$$\omega(t) = 2^{-(1+mod(t,8))}. \quad (6)$$

2.1.3 Eccentric down-sampling at the entrance of ON/OFF channels

Eccentric down-sampling uses a square RF to approximate a circular retina, as shown in Figure 3. The RF decreases linearly from the edge to the foveal area, where each RF corresponds to a single image pixel. First, we define the size and center of the RFs. The same sized RFs form a square ring. We can define only the position of upper left corner of each square ring, and then the foveal region will be defined. Accordingly, the size and center of the upper left corner of RFs can be computed as

$$R^c(i) = R^c(i - 1) + \frac{1}{2}R^s(i - 1), \ i = 1, 2, \ldots, l \quad (7)$$

$$R^s(i) = max(R^s) - \frac{max(R^s)}{d_{fovea}}R^c(i), \ i = 1, 2, \ldots, l \quad (8)$$

where i is the number of layers of the square ring, $R^s(i)$ represents the size of the RF of the upper left corner of the ith layer, and $R^c(i)$ represents the center of the RF of the upper left corner of the ith layer. $max(R^s)$ indicates the size of the outermost edge of the RF, and d_{fovea} is the total distance from the periphery to the edge of the foveal region whose dimension is 10% of the size of the image. l is the last layer number of edge square ring and $R^s(l + 1) < 2$. The square ring starts at layer-0. The length of the upper left corner of the 0th layer square ring is $max(R^s)$, and its center is $(\frac{max(R^s)}{2}, \frac{max(R^s)}{2})$. For ease of description, we arrange all RFs in a column as a set A: $A = \{RF_1, RF_2, \ldots, RF_m\}$, where the side of RF_k is R_k^s and its center position is (R_k^x, R_k^y).

Basically, the EC herein is equivalent to a local SNN, and each RF is modeled as an LIF neuron (Gerstner et al., 2014), which integrates the information of the pixels it contains. When the membrane potential of the RF exceeds the threshold, it fires a spike to the next layer of its connected neurons. Taking the ON channel for example, the entire process can be defined mathematically as

$$M_{on}(k, t) = M_{on}(k, t_1)e^{-\frac{dt}{\tau}} + \sum_{i=\left\lfloor -\frac{R_k^s}{2} \right\rfloor}^{\left\lfloor \frac{R_k^s}{2} \right\rfloor} \sum_{j=\left\lfloor -\frac{R_k^s}{2} \right\rfloor}^{\left\lfloor \frac{R_k^s}{2} \right\rfloor} \frac{C_{on}(R_k^x + i, R_k^y + j, t)}{R_k}$$
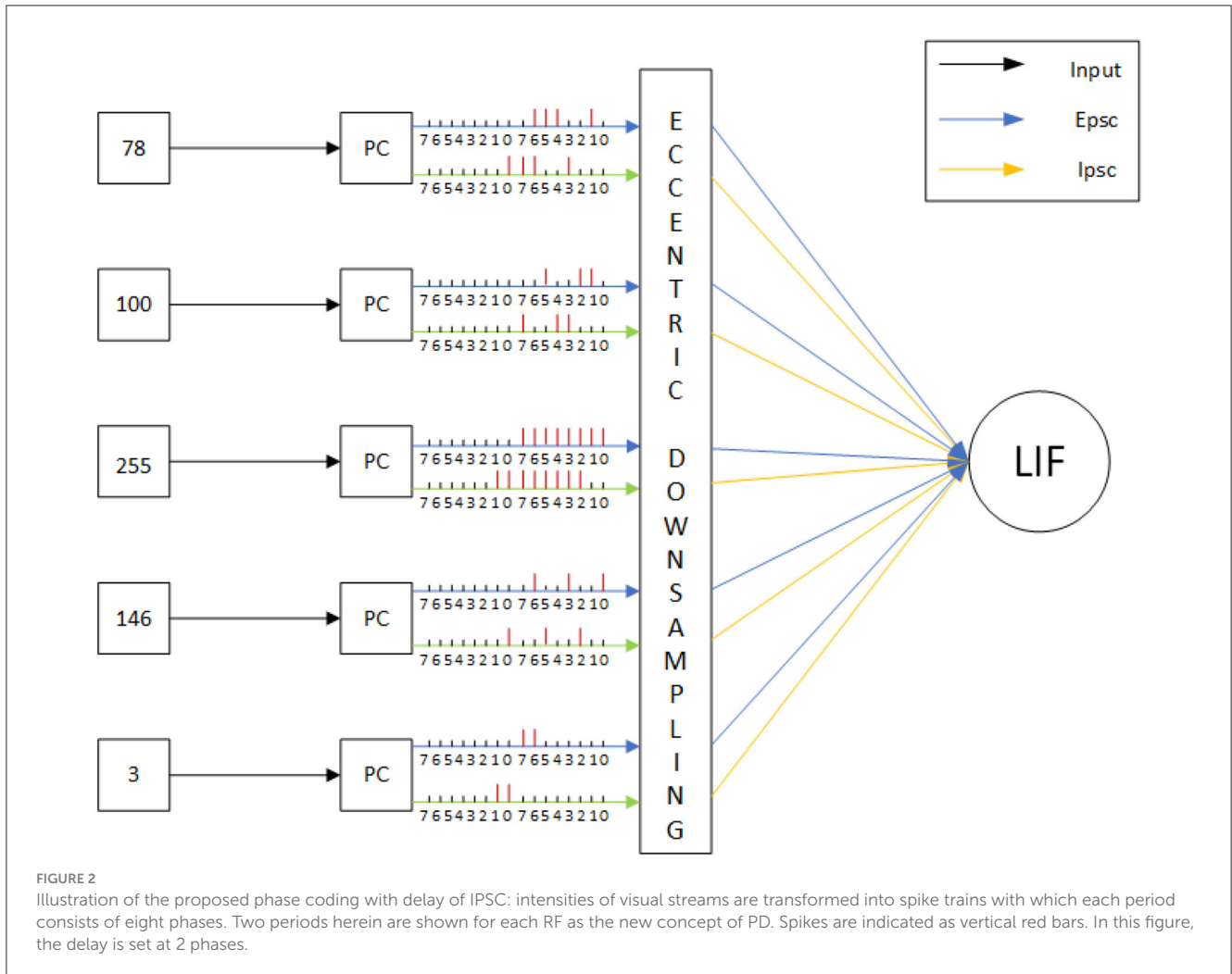
$$(9)$$

FIGURE 2
Illustration of the proposed phase coding with delay of IPSC: intensities of visual streams are transformed into spike trains with which each period consists of eight phases. Two periods herein are shown for each RF as the new concept of PD. Spikes are indicated as vertical red bars. In this figure, the delay is set at 2 phases.

$$S_{on}(k,t) = \begin{cases} 1 & \text{if } M_{on}(k,t) \geq \rho \\ 0 & \text{if } M_{on}(k,t) < \rho \end{cases} \quad (10)$$

where $M_{on}(k,t)$ represents the membrane potential of RF_k, t_1 indicates the moment of last emitted spike, and R_k is a percentage area of the RF. R_k affects the amount of current received by the pixel which RF_k contains. $\lfloor \cdot \rfloor$ denotes the floor function, and $S_{on}(k,t)$ indicates whether the RF_k emits spike at moment t. ρ is a spiking threshold for each RF_k, and τ is a time constant ($\tau = \frac{1,000}{fps}$, fps represents frames per second, i.e., the sampling frequency).

Next, we discuss which RFs the output neurons are connected to. We rezone the image, for which we need to take out a series of points. Denote the set of all upper left corner points of the square ring at the edge as a set B: $B = \{(x_0,y_0), (x_1,y_1),\ldots,(x_l,y_l)\}$, the set of diagonal points in the foveal region as a set C: $C = \{(x_{l+1},y_{l+1}),(x_{l+2},y_{l+2}),\ldots,(x_{l+n},y_{l+n})\}$, and the set of all bottom right corner points of the square ring at the edge as a set D: $D = \{(x_{l+n+1},y_{l+n+1}),(x_{l+n+2},y_{l+n+2}),\ldots,(x_{2l+n+1},y_{2l+n+1})\}$, and the entire process can be defined mathematically as

$$x_i = y_i = R^c(i) - \frac{1}{2}R^s(i); \quad i = 1,\ldots,l \quad (11)$$

$$x_{l+1} = y_{l+1} = R^c(l), x_{l+i} = y_{l+i} = x_{l+i-1} + 1; \quad i = 2,3,\ldots,n \quad (12)$$

$$x_{l+n+i} = y_{l+n+i} = L - x_{l-i+1}; \quad i = 1,\ldots,l+1 \quad (13)$$

where L denotes the side length of the image, n is the dimension of the foveal region which starts at $R^c(l)$, and (x_0,y_0) is $(0,0)$. Now, we re-divide the image into regions and the delineated areas are
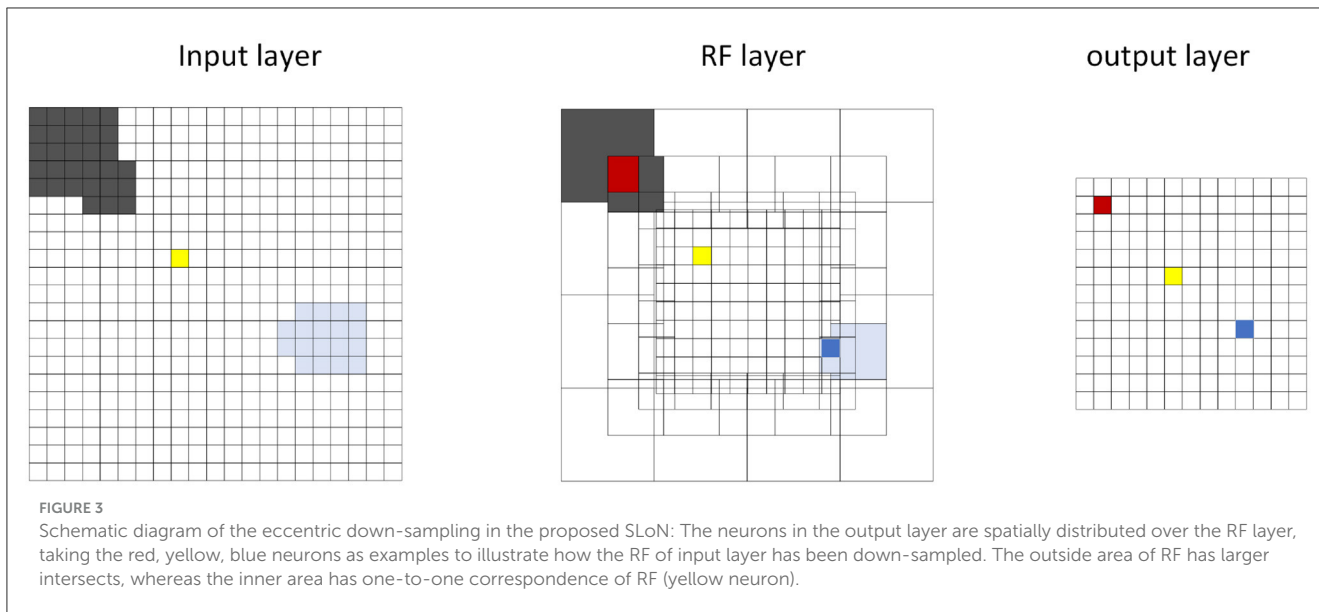
$$\triangle_n^{(x)} = [x_n, x_{n+1}], \triangle_n^{(y)} = [y_n, y_{n+1}]; \quad n = 0,1,\ldots,2l+n \quad (14)$$

$\triangle_n^{(x)}, \triangle_n^{(y)}$ denote the nth region of x-axis and y-axis in Cartesian coordinate system, respectively. From this, we can establish the connection between the RF and the output neurons as

$$R(\tilde{x},\tilde{y}) = \left\{ RF_k \mid (x,y) \cap (\triangle_{\tilde{x}}^{(x)} \times \triangle_{\tilde{y}}^{(x)}) \neq \varnothing, \ (x,y) \in RF_k, \ k = 1,2,\ldots,m \right\} \quad (15)$$

$R(\tilde{x},\tilde{y})$ represents the RF which the output neuron $(\tilde{x},\tilde{y})$ fastened. Furthermore, the output of neuron $(\tilde{x},\tilde{y})$ can be calculated as

$$\tilde{M}_{on}(\tilde{x},\tilde{y},t) = \tilde{M}_{on}(\tilde{x},\tilde{y},t_2)e^{-\frac{dt}{\tau}} + \sum_{RF_k \in R(\tilde{x},\tilde{y})} S_{on}(k,t) \quad (16)$$

FIGURE 3
Schematic diagram of the eccentric down-sampling in the proposed SLoN: The neurons in the output layer are spatially distributed over the RF layer, taking the red, yellow, blue neurons as examples to illustrate how the RF of input layer has been down-sampled. The outside area of RF has larger intersects, whereas the inner area has one-to-one correspondence of RF (yellow neuron).

$$\tilde{S}_{on}(\tilde{x}, \tilde{y}, t) = \begin{cases} 1 & \text{if } \tilde{M}(\tilde{x}, \tilde{y}, t) \geq \rho \\ 0 & \text{if } \tilde{M}(\tilde{x}, \tilde{y}, t) < \rho \end{cases} \quad (17)$$

where t_2 indicates the moment of last emitted spike, $\tilde{M}_{on}(\tilde{x}, \tilde{y}, t)$, $\tilde{S}_{on}(\tilde{x}, \tilde{y}, t)$ represents the membrane potential and spike of the neuron at $(\tilde{x}, \tilde{y})$. The eccentric down-sampling computations of OFF channels align with ON channels, which is not explicitly reiterated.

2.1.4 Spike-based interaction with phase delay

After neural encoding and eccentric down-sampling, the information within the ON/OFF pathways is transmitted to the second, spiking interaction layer, further extracting looming motion features. In the ON pathway, the excitatory input will be directly conveyed to the next summation sub-layer without time delay by convolving surrounding spikes as EPSC. Meanwhile, the inhibitory input is phase-delayed by convolving surrounding delayed spikes as IPSC. The computations in the OFF pathway can be obtained in the same way. The entire process can be defined as

$$E_{on}(\tilde{x}, \tilde{y}, t) = \omega(t)(\sum_{i=-1}^{1} \sum_{j=-1}^{1} \tilde{S}_{on}(\tilde{x}+i, \tilde{y}+j, t) W_1(\tilde{x}+i, \tilde{y}+j))\rho \quad (18)$$

$$I_{on}(\tilde{x}, \tilde{y}, t) = \omega(t-\epsilon)(\sum_{i=-4}^{4} \sum_{j=-4}^{4} \tilde{S}_{on}(\tilde{x}+i, \tilde{y}+j, t-\epsilon) W_2(\tilde{x}+i, \tilde{y}+j))\rho \quad (19)$$

$$E_{off}(\tilde{x}, \tilde{y}, t) = \omega(t)(\sum_{i=-1}^{1} \sum_{j=-1}^{1} \tilde{S}_{off}(\tilde{x}+i, \tilde{y}+j, t) W_1(\tilde{x}+i, \tilde{y}+j))\rho \quad (20)$$

$$I_{off}(\tilde{x}, \tilde{y}, t) = \omega(t-\epsilon)(\sum_{i=-4}^{4} \sum_{j=-4}^{4} \tilde{S}_{off}(\tilde{x}+i, \tilde{y}+j, t-\epsilon) W_2(\tilde{x}+i, \tilde{y}+j))\rho \quad (21)$$

where $E_{on/off}(\tilde{x}, \tilde{y}, t)$ represents the EPSC and $I_{on/off}(\tilde{x}, \tilde{y}, t)$ denotes the IPSC. ϵ indicates the delayed phase time, and W_1, W_2 are spatial

convolution matrices, obeying Gaussian distribution to adjust the connection weights of intermediate neurons.

Although the previous ANN methods apply similar strategy to shape the looming selectivity, e.g., Fu et al. (2018, 2020a), the proposed SLoN differs mainly in the following aspects:

- The SLoN is more biologically plausible as transmitting and processing action potentials between connected neurons.
- The looming selectivity can be implemented through spatiotemporal competition between excitation and inhibition, as demonstrated in previous modeling studies (Fu et al., 2019; Fu, 2023). Here, we introduce the new concept of phase delay to reflect the spatiotemporal competition between non-delayed EPSC and delayed IPSC.
- The EPSC is generated by the activities of a relatively smaller area of excitatory input spikes; on the other hand, the IPSC has a larger impact despite being delayed.

Therefore, the competitive interaction between EPSC and IPSC within the ON/OFF channels happened at the summation sub-layer as

$$G_{on}(\tilde{x}, \tilde{y}, t) = E_{on}(\tilde{x}, \tilde{y}, t) - I_{on}(\tilde{x}, \tilde{y}, t) \quad (22)$$

$$G_{off}(\tilde{x}, \tilde{y}, t) = E_{off}(\tilde{x}, \tilde{y}, t) - I_{off}(\tilde{x}, \tilde{y}, t) \quad (23)$$

The summation LIF neuron receives remaining current injection, causing an increase in neuronal membrane potential. If there is no input spikes, the neuron's membrane potential decays to the resting potential level. On the other hand, when the membrane potential exceeds the threshold, the neuron fires a spike, and the potential immediately declines to the resting level. Such neuronal dynamics can be described as

$$\widehat{S}_{on}(\tilde{x}, \tilde{y}, t) = U(V_{on}(\tilde{x}, \tilde{y}, t_3)e^{-\frac{dt}{\tau}} + G_{on}(\tilde{x}, \tilde{y}, t) - \omega(t)\rho)) \quad (24)$$

$$V_{on}(\tilde{x}, \tilde{y}, t) = V_{on}(\tilde{x}, \tilde{y}, t_3)e^{-\frac{dt}{\tau}} + G_{on}(\tilde{x}, \tilde{y}, t) - \omega(t)\widehat{S_{on}}(\tilde{x}, \tilde{y}, t)\rho \tag{25}$$

$$\widehat{S_{off}}(\tilde{x}, \tilde{y}, t) = U(V_{off}(\tilde{x}, \tilde{y}, t_4)e^{-\frac{dt}{\tau}} + G_{off}(\tilde{x}, \tilde{y}, t) - \omega(t)\rho)) \tag{26}$$

$$V_{off}(\tilde{x}, \tilde{y}, t) = V_{off}(\tilde{x}, \tilde{y}, t_4)e^{-\frac{dt}{\tau}} + G_{off}(\tilde{x}, \tilde{y}, t) - \omega(t)\widehat{S_{off}}(\tilde{x}, \tilde{y}, t)\rho \tag{27}$$

where $U(\cdot)$ is the Heaviside step function, and t_3, t_4 indicate the moment when last spike was emitted from corresponding neurons within ON/OFF channels, respectively. $V_{on/off}(\tilde{x}, \tilde{y}, t)$ represents the membrane potential, and $\widehat{S_{on/off}}(\tilde{x}, \tilde{y}, t)$ represents whether the neuron at $(\tilde{x}, \tilde{y})$ releases a spike at t phase time. Notably, the spike weight $\omega(t)$ and the threshold ρ are associated with the calculation of neuronal membrane potential and spike.

2.1.5 ON/OFF-type LIF neurons and feed-forward inhibition

Subsequently, there are ON/OFF-type LIF neurons, pooling spikes from ON/OFF channels. Notably, the summation sub-layer and LIF neurons are fully connected with a global Gaussian distributed matrix W_3 for either ON/OFF pathways. Inspired by the locust's LGMD computational models (Fu et al., 2019), here we introduce ON/OFF-type feed-forward inhibition (FFI) mechanism which works effectively to suppress directly the ON/OFF-type LIF neurons if large area of RF is highly activated inside a tiny time window. The previous research has demonstrated the effectiveness of FFI mechanism to prohibit the model from responding to the non-collision stimuli including rapid view shifting. Differently to the previous computations of FFI, we nevertheless compute the population firing rate of ON/OFF-type events to determine its efficacy. The computation can be defined as

$$F_{on}(t) = \omega(t - \epsilon)\left(\frac{\sum_{i=1}^{DL}\sum_{j=1}^{DL}\tilde{S}_{on}(\tilde{x}, \tilde{y}, t - \epsilon)}{DL^2}\right) \tag{28}$$

$$F_{off}(t) = \omega(t - \epsilon)\left(\frac{\sum_{i=1}^{DL}\sum_{j=1}^{DL}\tilde{S}_{off}(\tilde{x}, \tilde{y}, t - \epsilon)}{DL^2}\right) \tag{29}$$

where DL represents the length of down-sampled image after EC. Notably, the phase weight ω is also coupled with the calculation yet with the phase delay ϵ. The FFI output decides whether the ON/OFF-type LIF neurons receive spikes from the ON/OFF channels. Precisely, when the population rate exceeds a threshold, the LIF neuron will be directly suppressed, not receiving any input spikes. The computation can be expressed as

$$\tilde{I}_{on}(t) = \begin{cases} \omega(t)(\sum_{i=1}^{DL}\sum_{j=1}^{DL}\widehat{S_{on}}(i,j,t)W_3(i,j))\rho & \text{if } F_{on}(t) < S_{th} \\ 0 & \text{if } F_{on}(t) \geq S_{th} \end{cases} \tag{30}$$

$$\tilde{I}_{off}(t) = \begin{cases} \omega(t)(\sum_{i=1}^{DL}\sum_{j=1}^{DL}\widehat{S_{off}}(i,j,t)W_3(i,j))\rho & \text{if } F_{off}(t) < S_{th} \\ 0 & \text{if } F_{off}(t) \geq S_{th} \end{cases} \tag{31}$$

where S_{th} is the threshold. Otherwise, the ON/OFF-type LIF neurons are charged by input spikes working as the following dynamics:

$$\bar{S_{on}}(t) = U(\tilde{V}_{on}(t_5)e^{-\frac{dt}{\tau}} + \tilde{I}_{on}(t) - \omega(t)\rho)) \tag{32}$$

$$\bar{S_{off}}(t) = U(\tilde{V}_{off}(t_6)e^{-\frac{dt}{\tau}} + \tilde{I}_{off}(t) - \omega(t)\rho)) \tag{33}$$

$$\tilde{V}_{on}(t) = \tilde{V}_{on}(t_5)e^{-\frac{dt}{\tau}} + \tilde{I}_{on}(t) - \omega(t)\bar{S_{on}}(t)\rho \tag{34}$$

$$\tilde{V}_{off}(t) = \tilde{V}_{off}(t_6)e^{-\frac{dt}{\tau}} + \tilde{I}_{off}(t) - \omega(t)\bar{S_{off}}(t)\rho \tag{35}$$

Similarly, t_5, t_6 indicate the moment of last emitted spike from ON/OFF-type LIF neurons, respectively. $S_{on/off}(t)$ represents the generated spike in current phase time, and $\tilde{V}_{on/off}(t)$ is the neuronal membrane potential.

2.1.6 SLoN output neuron

As shown in Figure 1, the SLoN has only one output neuron to indicate whether there is any potential looming motion identified by the network. The outputs of ON/OFF-type neurons are linearly combined at the SLoN output neuron, which is

$$I(t) = \omega(t)(\theta_1 \bar{S_{on}}(t) + \theta_2 \bar{S_{off}}(t)) \tag{36}$$

where $\{\theta_1, \theta_2\}$ are term coefficients to adjust the different selectivity to ON/OFF-contrast looming stimuli. We will demonstrate the selection of such parameters in the experiments. Finally, the output of the SLoN is calculated as

$$S(t) = U(V(t_7)e^{-\frac{dt}{\tau}} + I(t) - \omega(t)\rho)) \tag{37}$$

$$V(t) = V(t_7)e^{-\frac{dt}{\tau}} + I(t) - \omega(t)S(t)\rho \tag{38}$$

where $U(\cdot)$ is the Heaviside step function, t_7 indicates the moment of last emitted spike, and $S(t), V(t)$ represent the spike and the membrane potential of the SLoN output neuron, respectively. The detailed online signal processing algorithms are also shown in Algorithm 1.

2.2 Setting the parameters

The parameters and their settings are shown in Table 2. It is essential to emphasize that the SLoN processes visual signals in a feed-forward structure without any feedback connections. Furthermore, learning methods were not employed in this modeling endeavor. The primary focus of this study was to propose an SNN framework that emulates biological visual neural systems for looming perception and integrates ON/OFF channels for eccentric neural encoding. The parameters were determined with considerations of mainly two aspects as follows:

- Certain parameters, such as the LIF spiking threshold and the standard deviations in connection matrices, were adjusted during experimental validation to enhance looming perception performance. The paramount consideration was achieving robust looming selectivity tested across diverse visual scenarios.

- On the other hand, the SLoN drew modeling inspiration from several notable studies: (1) The parameters of PC with weighted spikes were adapted from the modeling study by Kim et al. (2018), and the novel concept of PD was further investigated in the experiments. (2) The parameters of EC were derived from the modeling study by D'Angelo et al. (2020), and we provided a more detailed description of the EC algorithms. (3) The fundamental structure of ON/OFF channels was based on recent biological and computational advancements in visual motion perception (Fu, 2023).

2.3 Setting the experiments

To validate our proposed SLoN and investigate its key components, we conducted experiments across four categories, encompassing computer-simulated and real-physical stimuli. The goals were to demonstrate: (1) the basic functionality of the proposed model for looming perception with robust selectivity, (2) the efficacy of the ON/OFF channels in implementing diverse selectivity to ON/OFF-contrast looming stimuli, (3) the impact of the inaugurated concept of phase delay on looming perception, and (4) the results of an ablation study with comparison on down-sampling methods in the SLoN model.

Specifically, the experimental videos were divided into two parts. The first part included computer-simulated movements such as approaching, receding, translating, and grating stimuli, aligning with related physiological research. The second part involved more challenging real-world vehicle crash videos from a dashboard camera recording, which was sourced from related modeling research (Fu et al., 2020b). All input videos were of size 100×100 pixels, sampled at 30Hz. The source code of SLoN and visual stimuli used in the experiments can be accessed as open source at Github.

3 Results and analysis

Within this section, we present the results of systematic experiments with analysis. This includes four categories of experiments to verify the effectiveness of the proposed SLoN model. The main objectives of these tests fall within the following scope of

---

Input: $L(x, y, t)$

Output: $S(t)$

1 Receive video input : $L(x, y, t)$
2 **1. ON/OFF motion retrieving:**
3 The motion information is spit into parallel ON/OFF channels by Equations 1-3.
4 **2. Neural encoding and eccentric down-sampling:**
5 The signal is encoded and down-sampled by Equations 4-17.
6 **3. Spike-based interaction:**
7 The competitive interaction between EPSC and IPSC via Equations 18-27
8 **4. ON/OFF-type LIF neurons and FFI:**
9 Calculation of population firing rate $F_{on/off}(t)$ by Equations 28, 29.
10 **if** $F_{on/off}(t) \geq S_{th}$ **then**
11   ON/OFF-type LIF neurons do not receive spike input via Equations 30-35.
12 **else**
13   ON/OFF-type LIF neurons are charged via Equations 30-35.
14 **5. Output neuron:**
15 SLoN neuron integrates ON/OFF input via Equations 36-38.

Algorithm 1. Online algorithm of the SLoN.

TABLE 2 Parameters of the SLoN.

| Parameter	Description	Value
α_1	Coefficient at signal bifurcation of the neural encoding layer (Equations 2, 3)	0.1
$max(R^s)$	Largest size of RF in Equation 8	10
τ	Time constant in Equations 9, 16, 24, 25, 26, 27, 32–35, 37, 38	$\frac{1,000}{fps}$
R_k	Percentage area of the RF in Equation 9	0.6
ρ	LIF neuron firing threshold in Equations 10, 17–21, 24–27, 30–35, 37, 38	0.9
σ_1	Standard deviation of Gaussian kernel in Equations 18, 20	1
σ_2	Standard deviation of Gaussian kernel in Equations 19, 21	0.5
σ_3	Standard deviation of Gaussian kernel in Equations 30, 31	1
ϵ	Phase delay time in Equations 19, 21, 28, 29	2
L	Side length of original input visual stimuli	100
DL	Down-sampled side length in Equations 28–31	58
S_{th}	Threshold of FFI population rate in Equations 30, 31	0.1
$\{\theta_1, \theta_2\}$	Term coefficients in Equation 36	{0.5, 0.5}
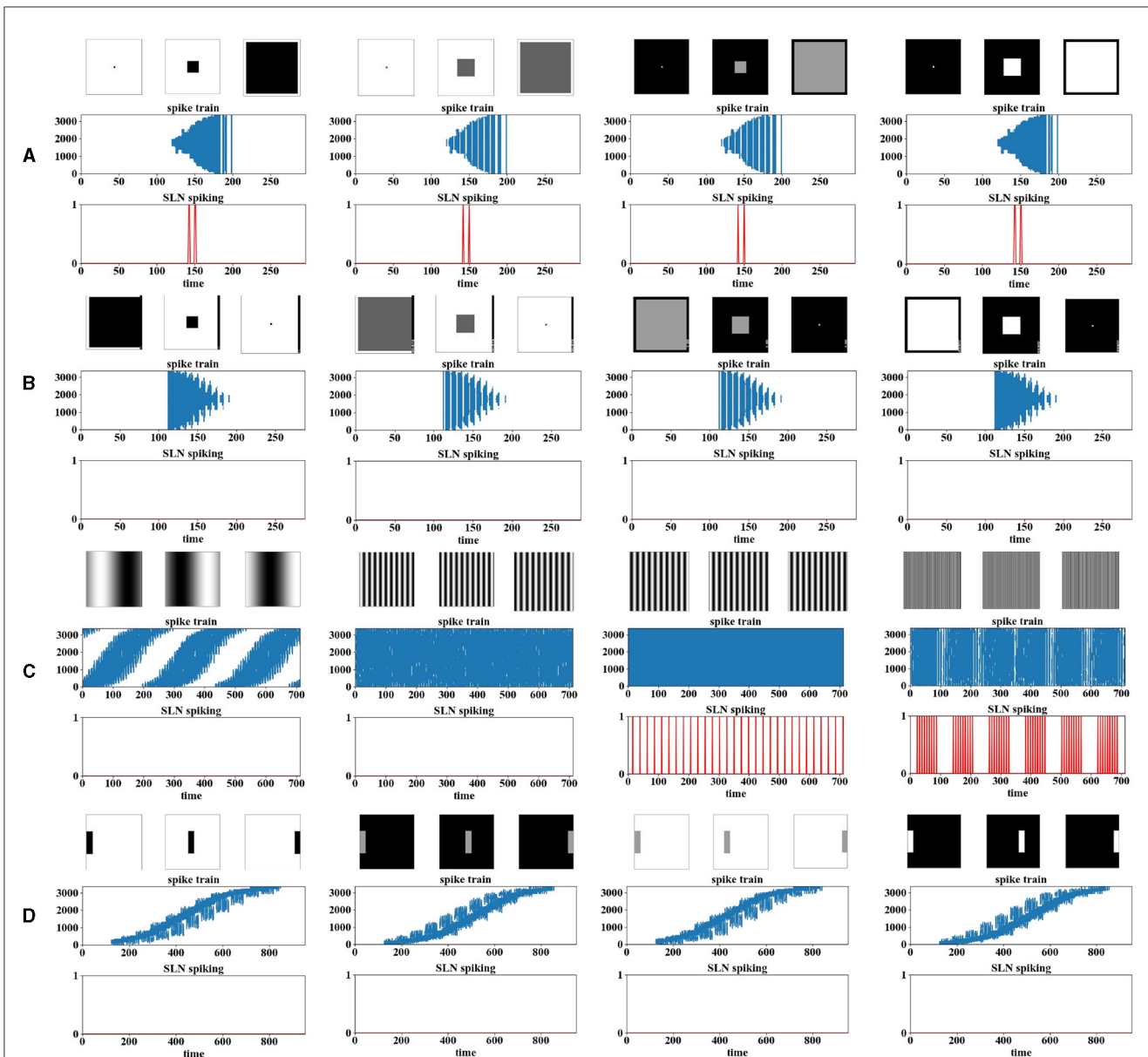
FIGURE 4
The results of the computer-simulated movements are presented, including **(A)** looming, **(B)** receding, **(C)** grating, and **(D)** translating. The images above the experimental results are snapshots of the input stimuli. The results are organized into two sub-graphs each, with one illustrating the position of the activated neurons over time as raster plots (i.e., the spike trains), and the other depicting the final spike output of the model. In the spike trains, the X-axis represents the phase time, while the Y-axis indicates ON/OFF-events in a vector after phase coding. In the model response, the X-axis indicates the phase time, and the Y-axis indicates the binary output of spikes.

- Verifying the effectiveness of SLoN for looming perception against a variety of visual collision challenges,
- Implementing the different looming selectivity to ON/OFF-contrast aligned with biological principles,
- Demonstrating the efficacy of phase delay in spike-based interaction shaping looming selectivity,
- Highlighting the robustness of EC mechanism increasing the fidelity of looming perception across various complicated scenarios.

3.1 Basic functionality

For collision perception models, the most important property to achieve is whether the model can respond most strongly to moving object signaling approaching rather than other types of movements. This is also naturally one criterion of "looming selectivity" for computer models. First, we carried out experiments on the approaching of squares, at four contrasts, as shown in Figure 4A. We can observe that the SLoN spikes for either white or dark objects, and the contrast matters. The encoded spike trains
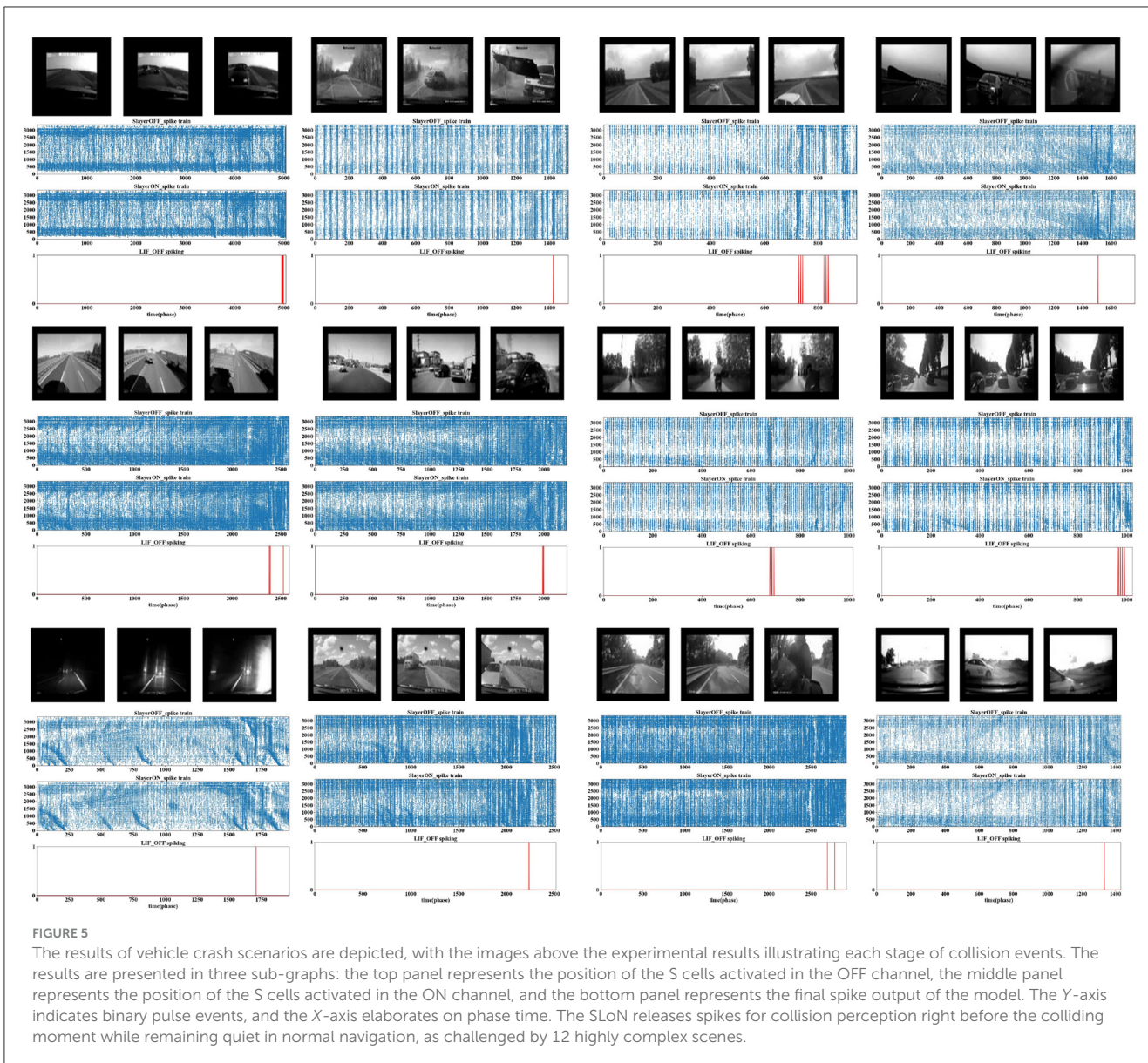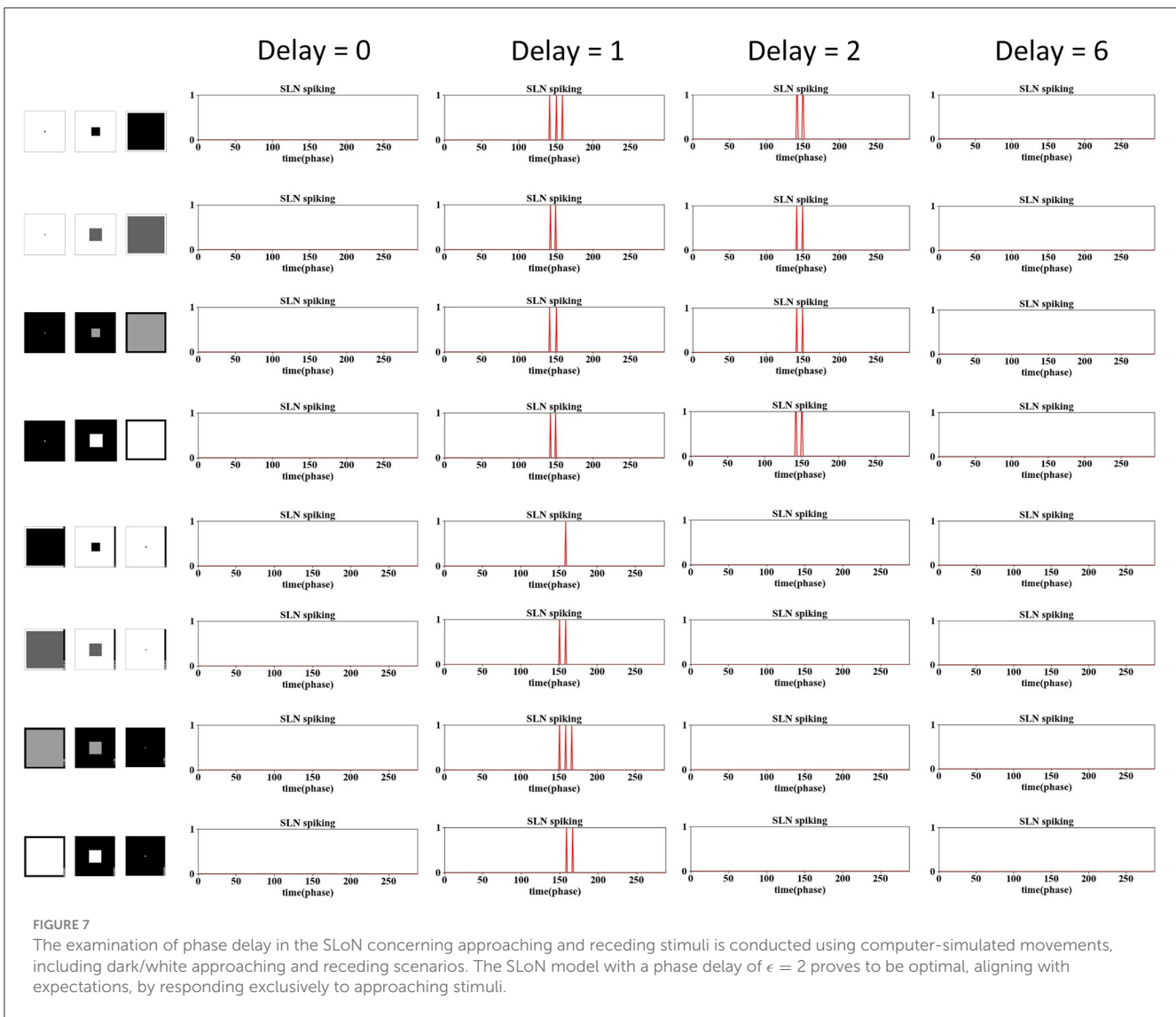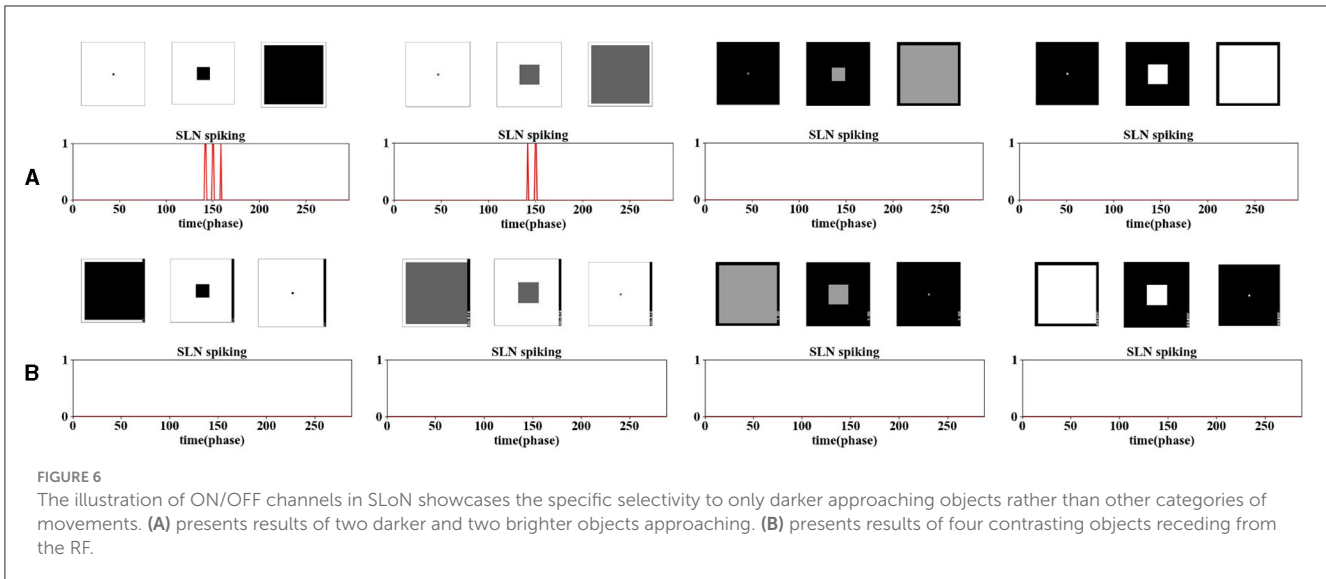
FIGURE 5

The results of vehicle crash scenarios are depicted, with the images above the experimental results illustrating each stage of collision events. The results are presented in three sub-graphs: the top panel represents the position of the S cells activated in the OFF channel, the middle panel represents the position of the S cells activated in the ON channel, and the bottom panel represents the final spike output of the model. The Y-axis indicates binary pulse events, and the X-axis elaborates on phase time. The SLoN releases spikes for collision perception right before the colliding moment while remaining quiet in normal navigation, as challenged by 12 highly complex scenes.

show more denser events within ON/OFF channels delivered to the spiking interaction layer, and the events of action potentials spread out spatially with respect to the approaching of squares. At all the tested contrasts, the SLoN performs stably on perception of looming motion.

On the other hand, we also did experiments with objects receding to observe if the model would generate spike to this opposite stimulus. Figure 4B shows, conversely, the spatiotemporal distributions of spike trains challenged by receding stimuli. Interestingly, the SLoN does not generate any spikes in this case, marking the emergence of looming selectivity.

Additionally, we tested the SLoN with stimuli involving translating movements, including sinusoidal grating at different frequencies and single-bar crossing at four contrasts. Figure 4C demonstrates that the SLoN does not consistently respond to grating at certain frequencies but is regularly activated in some cases. We observed that the PD and the threshold of FFI population

rate S_{th} can influence the SLoN's response to grating. Increasing the threshold could suppress the SLoN's response to grating stimuli but might impact the perception of looming motion if set too high. The subsequent experiments in Figure 8 will illustrate the influence of PD on grating stimuli. While adjusting the delay can suppress the response to grating stimuli, it can also compromise the model's ability to detect collisions. Attempts to make the model resolve both grating and approaching stimuli by adjusting the activation threshold within the neuron model were not entirely successful in distinguishing between approaching and receding stimuli. Moreover, the proposed EC facilitates the SLoN model in capturing motion features near the center of view. Zooming in on the grating pattern reveals motion features that were not discernible before the implementation of EC. Raising the threshold could suppress the SLoN's response to grating stimuli but may affect the perception of looming motion. Additionally, the spike trains, as shown in Figure 4D, display action potential events as the bar

FIGURE 6
The illustration of ON/OFF channels in SLoN showcases the specific selectivity to only darker approaching objects rather than other categories of movements. **(A)** presents results of two darker and two brighter objects approaching. **(B)** presents results of four contrasting objects receding from the RF.

FIGURE 7
The examination of phase delay in the SLoN concerning approaching and receding stimuli is conducted using computer-simulated movements, including dark/white approaching and receding scenarios. The SLoN model with a phase delay of $\epsilon = 2$ proves to be optimal, aligning with expectations, by responding exclusively to approaching stimuli.

moves rightward across the RF, and the SLoN remains inactive against such motion patterns.

To increase the challenge, we conducted experiments using real-world car crashes recorded by dashboard cameras. The SLoN model exhibits robust performance against various crash scenarios, generating brief spikes just before the moment of collision (Figure 5). Due to the high complexity of the vehicle scenes and the redundancy of spike trains after down-sampling, we present the activation location of ON/OFF polarity S-cells instead of EC cells. It could be difficult to find collision events directly from such spike train patterns. Nevertheless, the proposed SLoN model works

effectively to filter out irrelevant information, accurately extracting looming motion cues.

3.2 Potential of ON/OFF channels

The preceding bio-inspired collision perception visual systems demonstrate the capability to achieve diverse looming selectivity to ON/OFF-contrast, representing darker or brighter objects moving against a background. Specifically, the neural network model
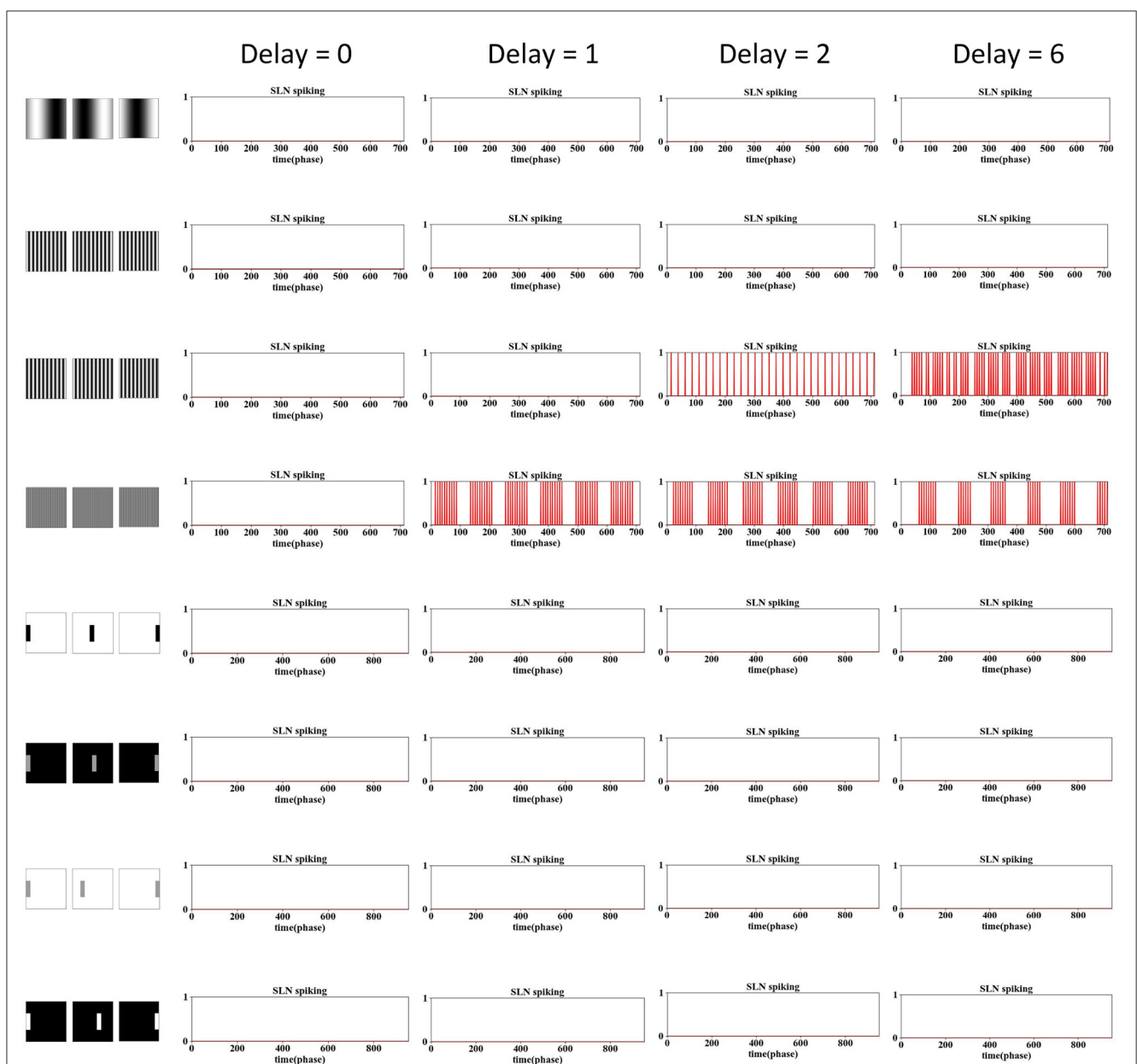

FIGURE 8
The examination of phase delay in the SLoN concerning grating and translating stimuli is conducted using computer-simulated movements. Four different frequencies of grating movements are tested. The SLoN model with a phase delay of $\epsilon = 0$ outperforms other comparative tests in this case, indicating that the SLoN should not respond to any grating or translating stimuli.

proposed by Fu et al. (2018) achieves the selectivity of the locust's LGMD1 neuron, responding to both dark and white looming objects. With the introduction of biased-ON/OFF channels, the neural model presented by Fu et al. (2020a) realizes the unique selectivity of the locust's LGMD2 neuron to only darker looming objects. In a recent study (Chang et al., 2023), feedback neural computation is incorporated into the original model based on ON/OFF channels to achieve diverse selectivity to either ON/OFF-contrast looming stimuli.

To align with the functionality of ON/OFF channels revealed in neuroscience (Fu, 2023), we conducted experiments to achieve a similar selectivity to the aforementioned LGMD2 neuron model, which is only sensitive to darker approaching objects, signaling OFF-contrast. The experiments so far have demonstrated that the SLoN model can achieve the selectivity of the locust's LGMD1, responding to the approaching of both white and dark objects. Figure 6 illustrates situations involving white and dark objects approaching and receding. As our proposed SLoN combines ON/OFF-type LIF neurons at the final output layer, we have the flexibility to adjust the contribution of either channel to the final LIF neuron. In previous experiments, ON-type and OFF-type spikes contributed equally. However, in these specific tests, we adjusted the coefficients in Equation 36, with the coefficient for the OFF pathway increased to 0.7 and that of the ON pathway decreased to 0.3. This adjustment indicates that OFF-type spikes contribute relatively more to the network. Consequently, as shown in Figure 6, the SLoN only generates spikes for dark looming patterns against a white background, at two contrasts, over any other movements. The model also exhibits a higher frequency of spikes in response to darker looming stimuli.

It is noteworthy that, unlike the model of the locust's LGMD2 proposed in the study by Fu et al. (2020a), which also responds briefly to a white object receding within a dark background (OFF-contrast), the proposed SLoN can better distinguish approaching from receding courses. Therefore, the ON/OFF channels in the SLoN contribute potentially to enriching the selectivity to align with the requirements of looming detection in more complex, real-world scenarios.

3.3 Investigation of phase delay

The phase time delay is a novel concept introduced in this modeling work. This subsection delves into the selection of the delay window in looming perception. As shown in Section 3.1, the delay is a crucial parameter that empowers our model to address the grating problem and refine the looming selectivity. The default setting for the delay is two phases, as shown in Table 2. In this investigation, we explored a range between 0 and 8, where 0 implies no delay in the IPSC, while 8 indicates that the IPSC is delayed to the same phase as the previous video frame.

The results are shown in Figures 7, 8, encompassing various motion patterns, with only representative results showcased. When the delay is set to 0, the SLoN model does not respond to any stimulus, including dark/white approaching, receding, translating, and grating movements, rendering it devoid of motion sensitivity. This underscores the crucial role of phase delay in constructing a motion perception SNN when utilizing phase coding to transform motion features into ON/OFF-type spike trains. The spatiotemporal competition between excitation and inhibition

TABLE 3 Results of different phase delay in the SLoN against real-world collision challenges.

Dataset	Delay 0	Delay 1	Delay 2	Delay 3	Delay 4	Delay 5	Delay 6	Delay 7	Delay 8
Car collision1	✗	✓	✓	✓	✓	✓	✓	✓	✗
Car collision2	✗	✗	✓	✓	✗	✓	✗	✓	✗
Car collision3	✗	✓	✓	✓	✓	✓	✓	✓	✗
Car collision4	✗	✓	✓	✓	✗	✓	✗	✓	✗
Car collision5	✗	✓	✓	✓	✗	✓	✓	✓	✗
Car collision6	✗	✓	✓	✓	✗	✓	✓	✓	✗
Car collision7	✗	✓	✓	✓	✗	✓	✗	✓	✗
Car collision8	✗	✓	✓	✓	✓	✓	✓	✓	✗
Car collision9	✗	✗	✓	✗	✗	✗	✓	✗	✗
Car collision10	✗	✓	✓	✓	✗	✓	✓	✓	✗
Car collision11	✗	✗	✓	✓	✗	✓	✓	✗	✗
Car collision12	✗	✗	✓	✗	✗	✗	✗	✗	✗
Ball collision1	✗	✓	✓	✓	✗	✓	✓	✓	✗
Ball collision2	✗	✗	✓	✗	✗	✗	✗	✗	✗
Ball collision3	✗	✗	✓	✗	✗	✗	✗	✗	✗
Ball collision4	✗	✗	✓	✗	✗	✗	✗	✗	✗

The notation "delayX" indicates that X-phases are delayed in the SLoN model.
✗ indicates the model fails to looming perception while ✓ indicates true positives.

consistently proves to be pivotal in shaping the specific looming selectivity as affirmed by numerous prior studies (Fu, 2023).

Obviously, when the PD increases to 1, the SLoN works effectively to detect approaching stimuli, however, also responds to receding stimuli. After further increasing the PD to 2, the SLoN can distinguish between approach, receding very well which is more in line with the expected looming selectivity. A limitation of the SLoN model is its inability to effectively handle grating movements, as it continues to be activated by certain frequencies, resulting in regular clustering of spikes. Addressing this issue requires the introduction of new bio-plausible mechanisms and algorithms in future research. Additionally, the PD cannot be set too large, as shown in Figures 7, 8, since the SLoN loses its ability for looming perception. We further investigated real-world collision challenges with different delays, and the results are shown in Table 3.

In summary, our investigation indicates that a PD of two-phases is an optimal temporal delay for the interaction between EPSC and IPSC in the network. This choice corresponds to the concentration of ON/OFF-type spike events, which are more heavily weighted within the initial phases through phase coding. This observation suggests that the most pertinent information is effectively conveyed to the ON/OFF-type LIF neurons during the early time window of each video frame. However, the current analyses are solely based on experiments where the mathematical analysis of PD should be involved in the future work.

3.4 Ablation study on eccentric down-sampling

Photoreceptors density in the mammalian retina is high at the fovea and decreases toward the periphery area. A recent study (D'Angelo et al., 2020) combined event-driven visual processing
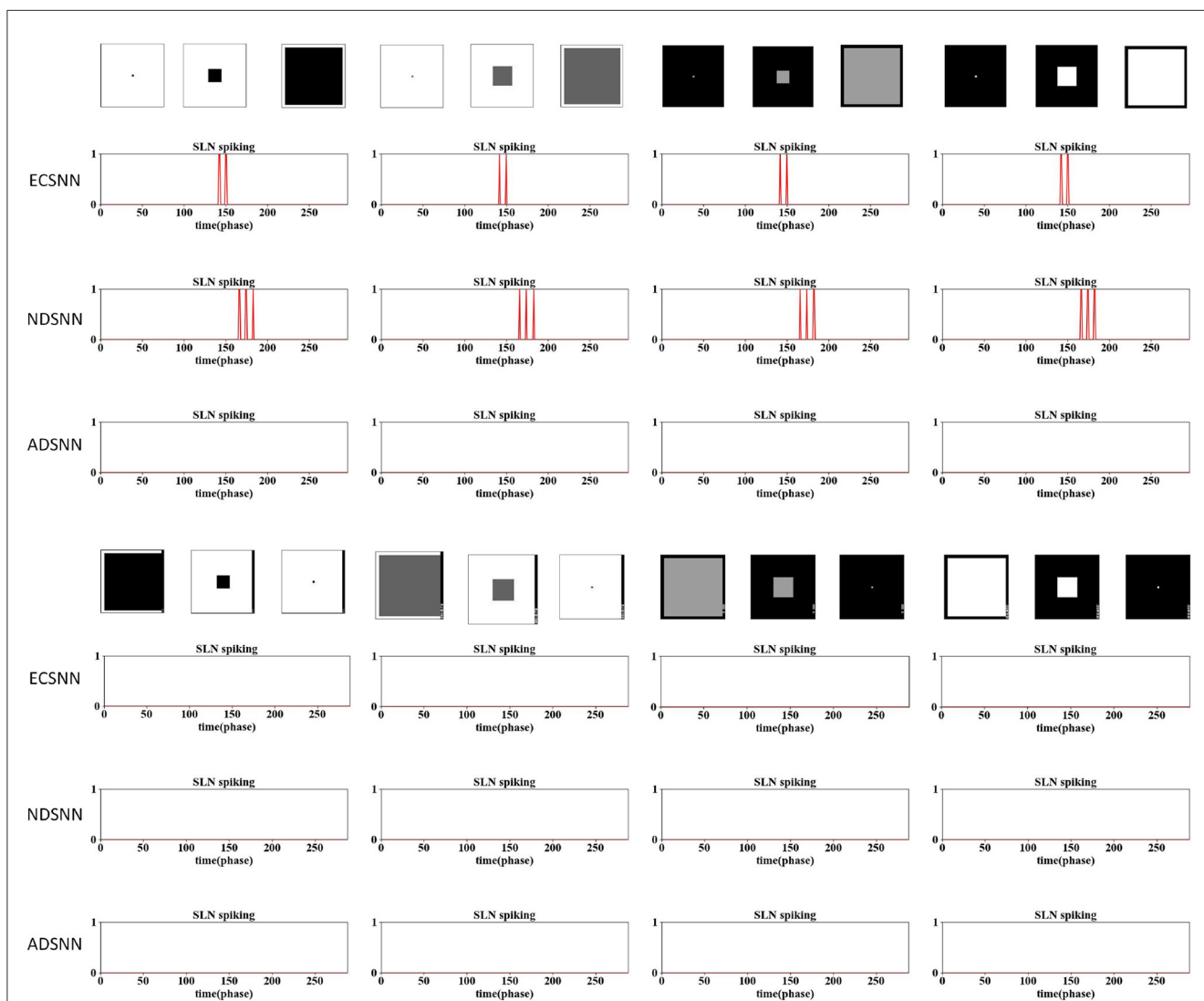

FIGURE 9
In the comparative results of the ablation study on down-sampling methods against dark/white looming and receding stimuli, ECSNN represents the SLoN with EC down-sampling method, NDSNN represents the SLoN without any down-sampling method, and ADSNN represents the SLoN with the traditional average down-sampling method. Both ECSNN and NDSNN can distinguish between approaching and receding movements, with NDSNN demonstrating a higher spiking rate. However, ADSNN cannot respond to looming stimuli.

FIGURE 10
Comparative results of ablation study on down-sampling methods against grating and dark/white translating stimuli following Figure 9.
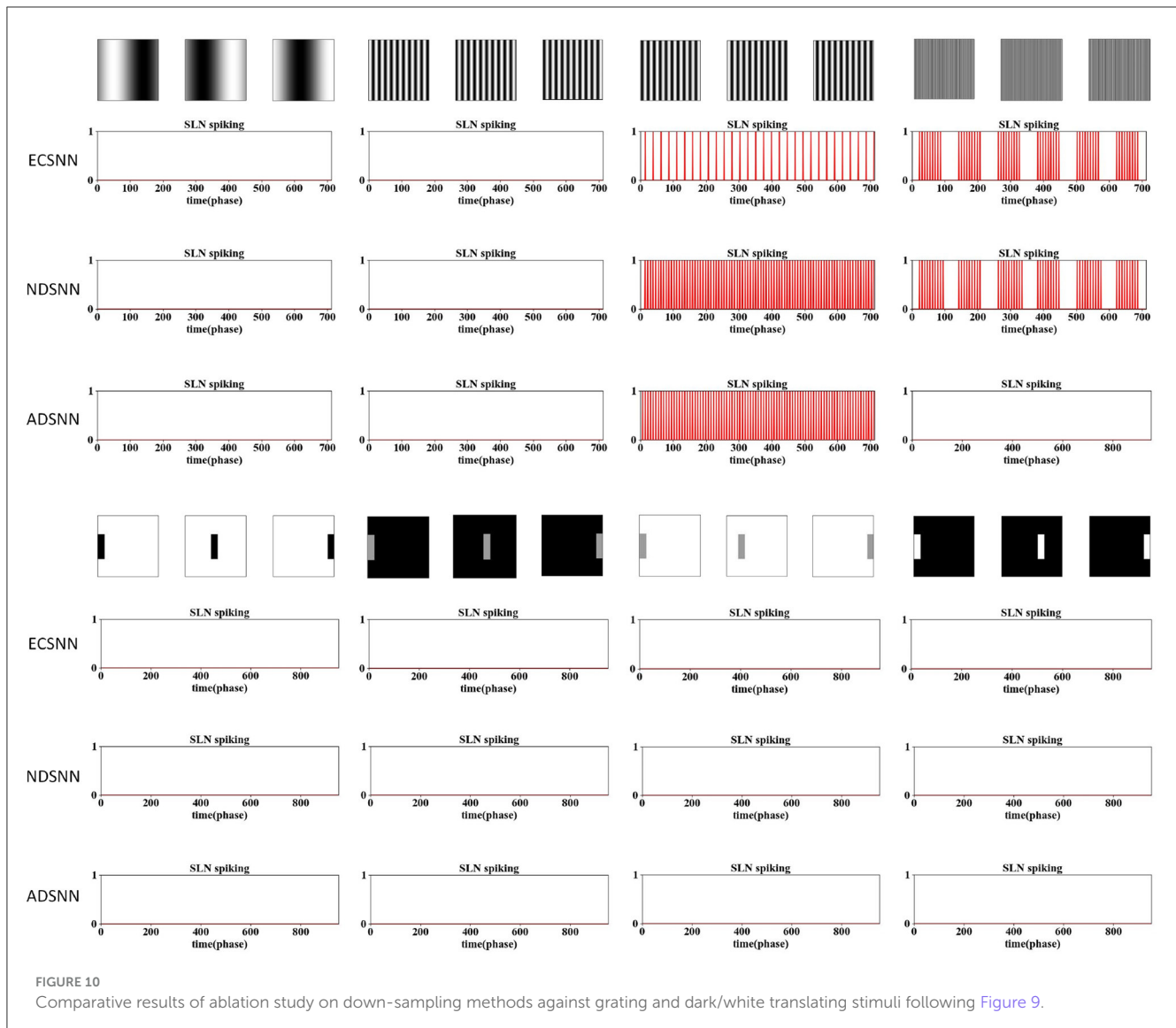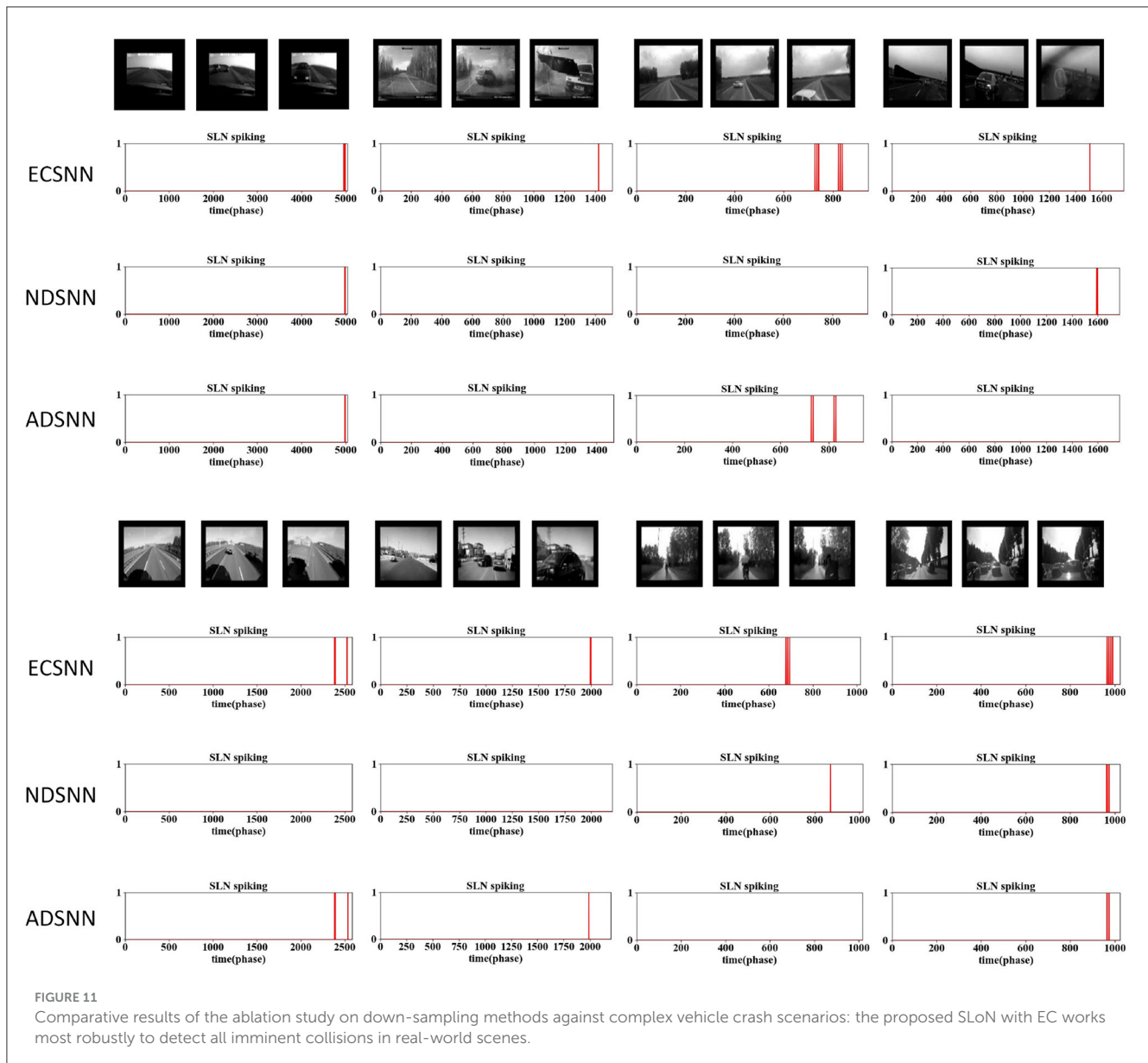
to model sEMD with the non-uniform retina model as a down-sampling of the visual field. We agreed with this strategy mimicking mammalian motion vision and incorporated the EC mechanism in the neural encoding of ON/OFF channels. Other motivations for us to model EC were (1) further sharpening up the looming selectivity as the object normally expands from the center area of visual field and (2) enhancing the robustness across complex visual scenes as the EC could potentially reduce noisy optical flows resident in the periphery. Accordingly, this subsection carries on ablation study on the down-sampling method and compares the EC with typical uniform down-sampling in the framework by SLoN's.

The results are shown in Figures 9–12. Three sets of experiments were conducted for each investigated model, including the EC down-sampling SLoN (ECSNN), the SNN with no down-sampling (NDSNN), and the average down-sampling SLoN (ADSNN). The synthetic stimuli, as shown in Figures 9, 10, revealed similar behavior between ECSNN and NDSNN. Both models responded to white/dark objects approaching but not to

objects receding and translating. However, they were not effective in addressing the grating problem at certain frequencies.

Furthermore, ECSNN exhibits earlier spikes than NDSNN in response to approaching objects (Figure 9), indicating that the EC mechanism concentrates motion-induced excitations in the central RF. Additionally, NDSNN is more responsive to grating stimuli than ECSNN. Conversely, ADSNN proves ineffective at detecting looming objects and other types of motion stimuli. This study underscores the unsuitability of uniform down-sampling for the proposed algorithm.

To assess the capability of EC down-sampling in improving the model's performance in real-world complex scenarios, experiments with vehicle collision stimuli were conducted. Twelve different scenarios were used to test the three comparative models. The results presented in Figures 11, 12 demonstrate the superior performance of ECSNN compared with the other two models, exhibiting robustness against all collision perception scenarios. Specifically, ECSNN recognizes all collisions, NDSNN detects only

FIGURE 11
Comparative results of the ablation study on down-sampling methods against complex vehicle crash scenarios: the proposed SLoN with EC works most robustly to detect all imminent collisions in real-world scenes.
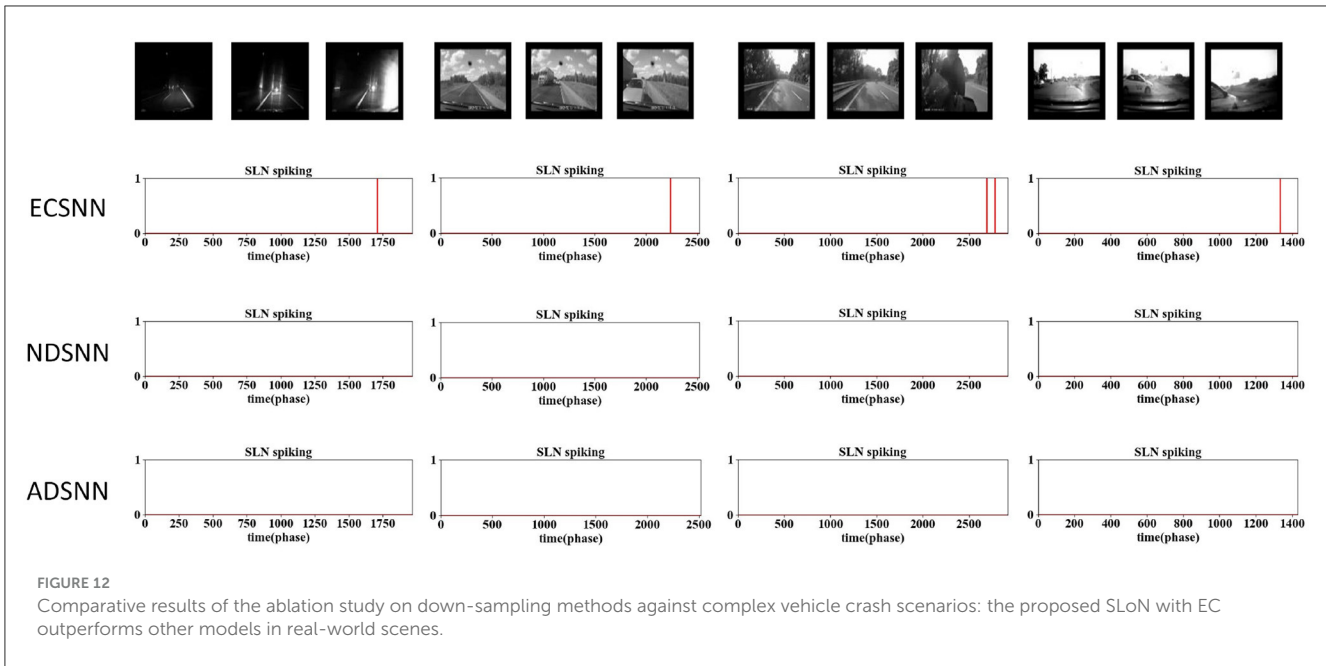
four collision events, and ADSNN detects five events. The small size of the RF in the EC-based SLoN model in the foveal region and its larger size in the edge region contribute to its heightened sensitivity to motion at the fovea, aligning with mammalian vision characteristics. The EC mechanism significantly enhances the SLoN model's response to approaching objects. The summarized results of real-world collisions are presented in Table 4, and a couple of additional experiments on ball collisions were included.

4 Discussion

In this study, we introduce a spiking looming perception network (SLoN) inspired by biological motion vision. Our systematic experiments demonstrate the efficacy of the SLoN model in detecting approaching objects while remaining unresponsive

to receding and translating ones. Additionally, we showcase the model's robustness in addressing complex challenges posed by real-world scenarios. The investigations into the intrinsic structures of SLoN, encompassing the ON/OFF channels, eccentric down-sampling mechanism, and phase delay, affirm the following key achievements in this modeling research:

- The incorporation of ON/OFF channels in motion-sensitive spiking neural networks enables the implementation of diverse looming selectivity to ON/OFF-contrast stimuli. This capability enhances the stability of model performance in complex and dynamic scenes.
- The newly introduced concept of phase delay extends the theory of phase coding, aligning well with the spatiotemporal interaction between excitatory and inhibitory signals to establish specific looming selectivity. As a result, the proposed

FIGURE 12
Comparative results of the ablation study on down-sampling methods against complex vehicle crash scenarios: the proposed SLoN with EC outperforms other models in real-world scenes.

SLoN distinguishes itself from typical SNN models that are designed for pattern recognition or object detection.

- The eccentric down-sampling mechanism not only mimics mammalian vision but also aligns with the characteristic of looming perception, focusing more on the expanding edges of objects from the foveal region of the visual field. This feature proves particularly beneficial in complex visual scenes, reducing the model's sensitivity to strong optical flows at the periphery.

Moreover, the SLoN model can achieve the selectivity of locust's LGMD2 (Simmons and Rind, 1997). By adjusting the weighted ON/OFF-type LIF neurons, the OFF channels contribute more than ON channels to the final output neuron. Consequently, the SLoN model achieves specific selectivity for darker objects approaching.

The final ablation experiment also verified the effectiveness of EC down-sampling. The SLoN responds to objects approaching rather than receding and translating. Such fidelity in looming perception is maintained in complex vehicle crash scenarios, which would be practical for utilization in mobile machines. This non-uniform sampling implies that the peripheral RF demands more stimulation to emit spikes. Essentially, the EC mechanism also reduces the dimension of input to the subsequent network with a decreased number of intermediate neurons.

We also note some shortcomings of the SLoN model. First, it cannot address cases of low-contrast motion, i.e., when there is a small luminance difference between the moving target and the background. In such instances, the SLoN model may struggle to detect approaching objects, as the LIF neurons in the initial encoding layer rely on visual contrast for spiking. Additionally, the SLoN model responds to grating movements at certain frequencies, which is not in line with expectations from either biological looming perception circuits or their computational models. Second, the feed-forward inhibition mechanism, learned from the modeling of locust's LGMD neuronal models, works

TABLE 4 Results of the ablations study on down-sampling methods in the SLoN model against real-world collision challenges.

Dataset	ECSNN	ADSNN	NDSNN
Car collision1	✓	✓	✓
Car collision2	✓	✕	✕
Car collision3	✓	✓	✕
Car collision4	✓	✕	✓
Car collision5	✓	✓	✕
Car collision6	✓	✓	✕
Car collision7	✓	✕	✓
Car collision8	✓	✓	✓
Car collision9	✓	✕	✕
Car collision10	✓	✕	✕
Car collision11	✓	✕	✕
Car collision12	✓	✕	✕
Ball collision1	✓	✓	✓
Ball collision2	✓	✕	✕
Ball collision3	✓	✓	✕
Ball collision4	✓	✕	✕

✕ indicates the model fails to looming perception while ✓ indicates true positives.

effectively to suppress such stimuli, but to some extent, it influences the responsive preference to approaching stimuli. There is a trade-off in tuning the FFI mechanism, as the threshold is not adaptive at present. Finally, it is essential to note that the SLoN model, like other vision-based neural models, is highly dependent on realistic lighting conditions. Under extreme conditions, such as excessively bright or dark scenes, our model may face challenges.

The future direction of this study encompasses three key aspects. First, we aim to incorporate event-based sensing modalities into the SLoN, providing an ideal match at the neural encoding layer. This approach not only enhances energy efficiency but also offers a high dynamic range, making it adaptable to extreme lighting environments (Gallego et al., 2022). In the SLoN model, the sensor could potentially replace the neural coding layer, directly generating ON/OFF-type spike trains as input for the EC down-sampling, as demonstrated in previous studies (D'Angelo et al., 2020).

Second, in the investigation of phase delay, we have set the maximum value at 8, considering that one video frame is composed of eight phases. However, further research is needed to explore the optimal choice of delay for different types of stimuli. In a previous LGMD2 modeling study (Fu et al., 2020a), inhibitory currents were found to be related not only to the excitatory current of the current frame but also to the excitatory current of the previous frame. Adjusting the delay can partially address specific types of stimuli. The exploration of different delay scopes and whether smaller delays are preferable requires systematic investigation, representing a focus for future work.

Third, our future research will delve into learning methods or time-varying, adaptive mechanisms to effectively handle grating stimuli. This represents a significant area for improvement in the SLoN model.

In conclusion, this study presents a computational model of a spiking looming perception network (SLoN) inspired by biological vision. Our objective is to capture looming perception in a manner consistent with neural information processing in the brain. The proposed model features an eccentric down-sampling mechanism that connects ON/OFF channels for neural encoding and transmission. Notably, we employ phase coding to transform video signals into spike trains for input to the SLoN, enabling the model to process data from frame-based cameras. Additionally, we introduce phase delay to represent the spatiotemporal interaction between excitation and inhibition for achieving looming selectivity. Systematic experiments validate the effectiveness and robustness of the SLoN model across a variety of collision challenges. The foundational structure of SLoN positions it well for integration with event-driven cameras, a focus for future investigations.

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 author.

Author contributions

ZD: Data curation, Investigation, Software, Visualization, Writing—original draft. QF: Conceptualization, Methodology, Project administration, Supervision, Validation, Writing—original draft, Writing—review & editing. JP: Formal analysis, Funding acquisition, Supervision, Validation, Writing—review & editing. HL: Formal analysis, Funding acquisition, Supervision, Validation, Writing—review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Barlow, H. B., and Levick, W. R. (1965). The mechanism of directionally selective units in rabbit's retina. *J. Physiol.* 178, 477–504. doi: 10.1113/jphysiol.1965.sp007638

Benet, G., Blanes, F., Simó, J., and Pérez, P. (2002). Using infrared sensors for distance measurement in mobile robots. *Rob. Auton. Syst.* 40, 255–266. doi: 10.1016/S0921-8890(02)00271-3

Borst, A., and Helmstaedter, M. (2015). Common circuit design in fly and mammalian motion vision. *Nat. Neurosci.* 18, 1067–1076. doi: 10.1038/nn.4050

Busse, L. (2018). "Chapter 4 - The mouse visual system and visual perception," in *Handbook of Object Novelty Recognition, Volume 27 of Handbook of Behavioral Neuroscience*, eds A. Ennaceur, and M. A. de Souza Silva (Amsterdam: Elsevier), 53–68. doi: 10.1016/B978-0-12-812012-5.00004-5

Chang, Z., Fu, Q., Chen, H., Li, H., and Peng, J. (2023). A look into feedback neural computation upon collision selectivity. *Neural Netw.* 166, 22–37. doi: 10.1016/j.neunet.2023.06.039

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)* (Padua), 1–8. doi: 10.1109/IJCNN55064.2022.9892618

D'Angelo, G., Janotte, E., Schoepe, T., O'Keeffe, J., Milde, M. B., Chicca, E., et al. (2020). Event-based eccentric motion detection exploiting time difference encoding. *Front. Neurosci.* 14:451. doi: 10.3389/fnins.2020.00451

Eichner, H., Joesch, M., Schnell, B., Reiff, D., and Borst, A. (2011). Internal structure of the fly elementary motion detector. *Neuron* 70, 1155–1164. doi: 10.1016/j.neuron.2011.03.028

Eyherabide, H., Rokem, A., Herz, A., and Samengo, I. (2009). Bursts generate a non-reducible spike-pattern code. *Front. Neurosci.* 3:1. doi: 10.3389/neuro.01.002.2009

Franceschini, N. (2014). Small brains, smart machines: from fly vision to robot vision and back again. *Proc. IEEE* 102, 751–781. doi: 10.1109/JPROC.2014.2312916

Franceschini, N., Riehle, A., and Le Nestour, A. (1989). "Directionally selective motion detection by insect neurons," in *Facets of Vision*, eds D. G. Stavenga, and R. C. Hardie (Berlin: Springer Berlin Heidelberg), 360–390. doi: 10.1007/978-3-642-74082-4_17

Fu, Q. (2023). Motion perception based on on/off channels: a survey. *Neural Netw.* 165, 1–18. doi: 10.1016/j.neunet.2023.05.031

Fu, Q., Hu, C., Peng, J., Rind, F. C., and Yue, S. (2020a). A robust collision perception visual neural network with specific selectivity to darker objects. *IEEE Trans. Cybern.* 50, 5074–5088. doi: 10.1109/TCYB.2019.2946090

Fu, Q., Hu, C., Peng, J., and Yue, S. (2018). Shaping the collision selectivity in a looming sensitive neuron model with parallel on and off pathways and spike frequency adaptation. *Neural Netw.* 106, 127–143. doi: 10.1016/j.neunet.2018.04.001

Fu, Q., Wang, H., Hu, C., and Yue, S. (2019). Towards computational models and applications of insect visual systems for motion perception: a review. *Artif. Life* 25, 263–311. doi: 10.1162/artl_a_00297

Fu, Q., Wang, H., Peng, J., and Yue, S. (2020b). Improved collision perception neuronal system model with adaptive inhibition mechanism and evolutionary learning. *IEEE Access* 8, 108896–108912. doi: 10.1109/ACCESS.2020.3001396

Fu, Q., and Yue, S. (2020). Modelling *Drosophila* motion vision pathways for decoding the direction of translating objects against cluttered moving backgrounds. *Biol. Cybern.* 114, 443–460. doi: 10.1007/s00422-020-00841-x

Fu, Q., Yue, S., and Hu, C. (2016). "Bio-inspired collision detector with enhanced selectivity for ground robotic vision system," in *British Machine Vision Conference* (Durham: BMVA Press), 1–13. doi: 10.5244/C.30.6

Fu, Q., and Yue, S. (2015). "Modelling lgmd2 visual neuron system," in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)* (Boston, MA), 1–6. doi: 10.1109/MLSP.2015.7324313

Gallego, G., Delbrck, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., et al. (2022). Event-based vision: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 154–180. doi: 10.1109/TPAMI.2020.3008413

Gerstner, W., Kistler, W. M., Naud, R., and Paninski, L. (2014). *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. Cambridge: Cambridge University Press.

Guo, W., Fouda, M. E., Eltawil, A. M., and Salama, K. N. (2021). Neural coding in spiking neural networks: a comparative study for robust neuromorphic systems. *Front. Neurosci.* 15:638474. doi: 10.3389/fnins.2021.638474

Halupka, K. J., Wiederman, S. D., Cazzolato, B. S., and O'Carroll, D. C. (2011). "Discrete implementation of biologically inspired image processing for target detection," in *2011 Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing* (Adelaide, SA: IEEE), 143–148. doi: 10.1109/ISSNIP.2011.6146617

Hartline, H. K. (1938). The response of single optic-nerve fibers of the vertebrate eye to illumination of the retina. *Am. J. Physiol.-Legacy Content* 121, 400–415. doi: 10.1152/ajplegacy.1938.121.2.400

Harvey, B. M., and Dumoulin, S. O. (2011). The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. *J. Neurosci.* 31, 13604–13612. doi: 10.1523/JNEUROSCI.2572-11.2011

Heeger, D. (2000). *Poisson Model of Spike Generation*, Volume 5. Handout: University of Standford, 76.

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154. doi: 10.1113/jphysiol.1962.sp006837

Joesch, M., Weber, F., Eichner, H., and Borst, A. (2013). Functional specialization of parallel motion detection circuits in the fly. *J. Neurosci.* 33, 902–905. doi: 10.1523/JNEUROSCI.3374-12.2013

Kheradpisheh, S. R., Ganjtabesh, M., Thorpe, S. J., and Masquelier, T. (2018). STDP-based spiking deep convolutional neural networks for object recognition. *Neural Netw.* 99, 56–67. doi: 10.1016/j.neunet.2017.12.005

Kim, J., Kim, H., Huh, S., Lee, J., and Choi, K. (2018). Deep neural networks with weighted spikes. *Neurocomputing* 311, 373–386. doi: 10.1016/j.neucom.2018.05.087

Laurent, G. (1996). Dynamical representation of odors by oscillating and evolving neural assemblies. *Trends Neurosci.* 19, 489–496. doi: 10.1016/S0166-2236(96)10054-0

Lee, K. H., Tran, A., Turan, Z., and Meister, M. (2020). The sifting of visual information in the superior colliculus. *Elife* 9, e50678. doi: 10.7554/eLife.50678

Lobo, J. L., Del Ser, J., Bifet, A., and Kasabov, N. (2020). Spiking neural networks and online learning: an overview and perspectives. *Neural Netw.* 121, 88–100. doi: 10.1016/j.neunet.2019.09.004

Manduchi, R., Castano, A., Talukder, A., and Matthies, L. (2005). Obstacle detection and terrain classification for autonomous off-road navigation. *Auton. Robots* 18, 81–102. doi: 10.1023/B:AURO.0000047286.62481.1d

Masuta, H., and Kubota, N. (2010). "Perceptual system using spiking neural network for an intelligent robot," in *2010 IEEE International Conference on Systems, Man and Cybernetics* (Istanbul), 3405–3412. doi: 10.1109/ICSMC.2010.5642471

Milde, M. B., Bertrand, O. J., Benosmanz, R., Egelhaaf, M., and Chicca, E. (2015). "Bioinspired event-driven collision avoidance algorithm based on optic flow," in *Event-Based Control, Communication, and Signal Processing* (Krakow: IEEE), 1–7. doi: 10.1109/EBCCSP.2015.7300673

Milde, M. B., Bertrand, O. J. N., Ramachandran, H., Egelhaaf, M., and Chicca, E. (2018). Spiking elementary motion detector in neuromorphic systems. *Neural Comput.* 30, 2384–2417. doi: 10.1162/neco_a_01112

Milde, M. B., Blum, H., Dietmüller, A., Sumislawska, D., Conradt, J., Indiveri, G., et al. (2017). Obstacle avoidance and target acquisition for robot navigation using a mixed signal analog/digital neuromorphic processing system. *Front. Neurorobot.* 11:28. doi: 10.3389/fnbot.2017.00028

Mukhtar, A., Xia, L., and Tang, T. B. (2015). Vehicle detection techniques for collision avoidance systems: a review. *IEEE Trans. Intell. Transp. Syst.* 16, 2318–2338. doi: 10.1109/TITS.2015.2409109

Nelson, M. E., and MacIver, M. A. (2006). Sensory acquisition in active sensing systems. *J. Comp. Physiol. A* 192, 573–586. doi: 10.1007/s00359-006-0099-4

O'Keefe, J. and Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3, 317–330. doi: 10.1002/hipo.450030307

Park, S., Kim, S., Na, B., and Yoon, S. (2020). "T2FSNN: deep spiking neural networks with time-to-first-spike coding," in *2020 57th ACM/IEEE Design Automation Conference (DAC)* (San Francisco, CA), 1–6. doi: 10.1109/DAC18072.2020.9218689

Posch, C., Serrano-Gotarredona, T., Linares-Barranco, B., and Delbruck, T. (2014). Retinomorphic event-based vision sensors: bioinspired cameras with spiking output. *Proc. IEEE* 102, 1470–1484. doi: 10.1109/JPROC.2014.2346153

Salt, L. (2016). *Optimising a Neuromorphic Locust Looming Detector for UAV Obstacle Avoidance* [PhD thesis, Master's thesis]. St Lucia, QLD: The University of Queensland.

Salt, L., Howard, D., Indiveri, G., and Sandamirskaya, Y. (2020). Parameter optimization and learning in a spiking neural network for uav obstacle avoidance targeting neuromorphic processors. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 3305–3318. doi: 10.1109/TNNLS.2019.2941506

Salt, L., Indiveri, G., and Sandamirskaya, Y. (2017). "Obstacle avoidance with LGMD neuron: towards a neuromorphic UAV implementation," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)* (Baltimore, MD), 1–4. doi: 10.1109/ISCAS.2017.8050976

Schiller, P. H., Sandell, J. H., and Maunsell, J. H. R. (1986). Functions of the on and off channels of the visual system. *Nature* 322, 824–825. doi: 10.1038/322824a0

Simmons, P. J., and Rind, F. C. (1997). Responses to object approach by a wide field visual neurone, the lgmd2 of the locust: characterization and image cues. *J. Comp. Physiol. A* 180, 203–214. doi: 10.1007/s003590050041

Tan, C., Šarlija, M., and Kasabov, N. (2020). Spiking neural networks: background, recent development and the neucube architecture. *Neural Process. Lett.* 52, 1675–1701. doi: 10.1007/s11063-020-10322-8

Tavanaei, A., Ghodrati, M., Kheradpisheh, S. R., Masquelier, T., and Maida, A. (2019). Deep learning in spiking neural networks. *Neural Netw.* 111, 47–63. doi: 10.1016/j.neunet.2018.12.002

Vasco, V., Glover, A., Mueggler, E., Scaramuzza, D., Natale, L., Bartolozzi, C., et al. (2017). "Independent motion detection with event-driven cameras," in *International Conference on Advanced Robotics* (Hong Kong), 530–536. doi: 10.1109/ICAR.2017.8023661

Wiederman, S. D., Shoemaker, P. A., and O'Carroll, D. C. (2013). Correlation between off and on channels underlies dark target selectivity in an insect visual system. *J. Neurosci.* 33, 13225–13232. doi: 10.1523/JNEUROSCI.1277-13.2013

Wurbs, J., Mingolla, E., and Yazdanbakhsh, A. (2013). Modeling a space-variant cortical representation for apparent motion. *J. Vis.* 13, 2–2. doi: 10.1167/13.10.2

Yang, S., Wang, H., and Chen, B. (2023). Sibols: robust and energy-efficient learning for spike-based machine intelligence in information bottleneck framework. *IEEE Trans. Cogn. Dev. Syst.* 1–13. doi: 10.1109/TCDS.2023.3329532

Yang, S., and Chen, B. (2023b). SNIB: improving spike-based machine learning using nonlinear information bottleneck. *IEEE Trans. Syst. Man. Cybern.* 53, 7852–7863. doi: 10.1109/TSMC.2023.3300318

Yang, S., Tan, J., and Chen, B. (2022). Robust spike-based continual meta-learning improved by restricted minimum error entropy criterion. *Entropy* 24:455. doi: 10.3390/e24040455

Yang, S., and Chen, B. (2023a). Effective surrogate gradient learning with high-order information bottleneck for spike-based machine intelligence.

IEEE Trans. Neural Netw. Learn. Syst. 1–15. doi: 10.1109/TNNLS.2023.332 9525

Yilmaz, M., and Meister, M. (2013). Rapid innate defensive responses of mice to looming visual stimuli. *Curr. Biol.* 23, 2011–2015. doi: 10.1016/j.cub.2013.0 8.015