Important declarations

Please remove this info from manuscript text if it is also present there.

Associated Data

Data supplied by the author:
Data is available at Github (https://github.com/fuqinbing).

Required Statements

Competing Interest statement:
The authors declare that they have no competing interests.

Funding statement:
This research was supported by the National Natural Science Foundation of China under Grant No. 62376063.

A neuronal assembly model with elevated time derivative boosts loom-selectivity

Qinbing Fu [Corresp., 1], **Mengying Wang** [Equal first author, 1], **Renyuan Liu** [Equal first author, 1], **Wanlin He** [1]

[1] School of Mathematics and Information Science, Guangzhou University, Guangzhou, Guangdong, China

Corresponding Author: Qinbing Fu
Email address: qifu@gzhu.edu.cn

In nature, the ability of sighted animals to selectively respond to approaching objects is crucial for survival. The capacity to discriminate approaching (looming) movements from other motion patterns, termed loom-selectivity, is fundamental to collision perception. Driven by biological research, the single neuron computation of lobula giant movement detectors (LGMD) in locust's visual systems has ramified, inspiring numerous artificial vision systems. However, a notable gap remains: existing computational models, typically based on single-neuron implementations, exhibit limited or task-specific looming selectivity compared to biological organisms. To bridge this gap, we proposed a novel computational framework combining LGMD1 and LGMD2 neuronal models. This composite model effectively processes diverse visual movements and demonstrates enhanced selectivity for looming stimuli, robustly distinguishing them from translating and receding motion. The proposed framework was simplified and embedded into a micro-robotic vision system, guiding collision-free navigation within visually dynamic environments. In this paper, we further investigated the internal structure of the neuronal assembly network to clarify the mechanisms underlying this improved selectivity. Through ablation studies on the previously developed LGMD2-LGMD1 composite model and comparative experiments with state-of-the-art methods, we identified that elevating the time derivative in the second-stage neuronal processing module is central to enhanced looming selectivity, rendering subsequent neural computations redundant. Accordingly, we developed a more economical neuronal model emphasizing time derivative computation within visual streams. This optimized model maintains performance comparable to the LGMD2-LGMD1 composite model and consistently outperforms existing bio-inspired methods. Real-world testing through online implementation on micro-mobile robots further confirmed that increasing the sensitivity to temporal changes substantially enhances looming selectivity, reinforcing the value of this approach for practical collision avoidance tasks.

A neuronal assembly model with elevated time derivative boosts loom-selectivity

Mengying Wang[1,2,†], **Renyuan Liu**[1,†], **Wanlin He**[1], **and Qinbing Fu**[1,2,†,*]

[1]**School of Mathematics and Information Science, Guangzhou University, China**
[2]**Machine Life and Intelligence Research Centre, Guangzhou University, China**

Corresponding author:
Qinbing Fu

Email address: qifu@gzhu.edu.cn

ABSTRACT

In nature, the ability of sighted animals to selectively respond to approaching objects is crucial for survival. The capacity to discriminate approaching (looming) movements from other motion patterns, termed loom-selectivity, is fundamental to collision perception. Driven by biological research, the single neuron computation of lobula giant movement detectors (LGMD) in locust's visual systems has ramified, inspiring numerous artificial vision systems. However, a notable gap remains: existing computational models, typically based on single-neuron implementations, exhibit limited or task-specific looming selectivity compared to biological organisms. To bridge this gap, we proposed a novel computational framework combining LGMD1 and LGMD2 neuronal models. This composite model effectively processes diverse visual movements and demonstrates enhanced selectivity for looming stimuli, robustly distinguishing them from translating and receding motion. The proposed framework was simplified and embedded into a micro-robotic vision system, guiding collision-free navigation within visually dynamic environments. In this paper, we further investigated the internal structure of the neuronal assembly network to clarify the mechanisms underlying this improved selectivity. Through ablation studies on the previously developed LGMD2-LGMD1 composite model and comparative experiments with state-of-the-art methods, we identified that elevating the time derivative in the second-stage neuronal processing module is central to enhanced looming selectivity, rendering subsequent neural computations redundant. Accordingly, we developed a more economical neuronal model emphasizing time derivative computation within visual streams. This optimized model maintains performance comparable to the LGMD2-LGMD1 composite model and consistently outperforms existing bio-inspired methods. Real-world testing through online implementation on micro-mobile robots further confirmed that increasing the sensitivity to temporal changes substantially enhances looming selectivity, reinforcing the value of this approach for practical collision avoidance tasks.

1 INTRODUCTION

Collision detection and avoidance are critical capabilities for all living creatures. Through millions of years of evolution, animals have developed sophisticated abilities to perceive collisions within complex and dynamic environments. Specifically, they excel at distinguishing objects that move directly toward them—potential collision threats known as looming stimuli—from other types of motion, such as translation, recession, or rotation caused either by their ego-movements or by nearby moving objects. A fundamental question in neuroscience concerns how animals with normal vision, including mammals and insects, achieve this exceptional sensitivity to looming stimuli—a phenomenon termed "*loom-selectivity*". Gaining insights into the neural mechanisms underlying this ability can guide the development of advanced dynamic vision systems, enabling us to replicate natural visual capabilities and address real-world collision detection challenges more effectively.

Within recent decades, computational models inspired by biological insights—particularly detailed investigations into the neural circuits of insects—have developed rapidly and significantly advanced our understanding of visual processing mechanisms and strategies, as reviewed in (Fu et al., 2019; Fu, 2023). Among these biological models, flies and locusts are particularly noteworthy examples of animals studied extensively for their robust looming perception capabilities (Muijres et al., 2014; Bertrand et al., 2015;

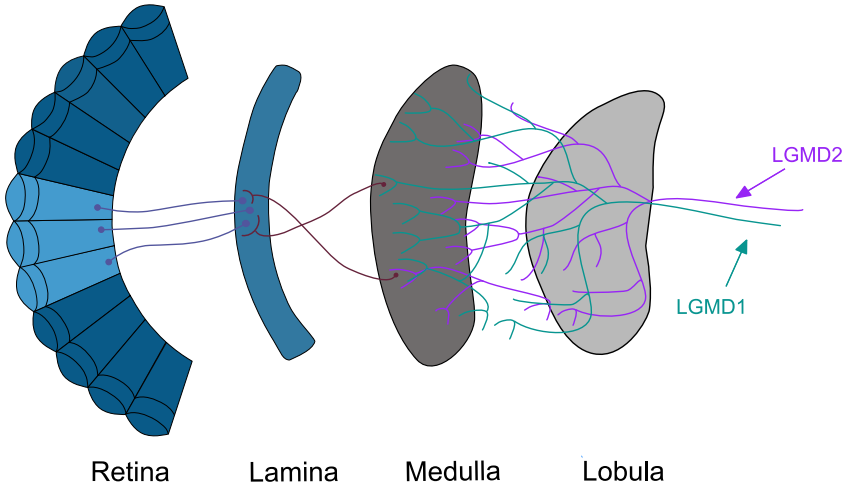

Figure 1. Schematic illustration of presynaptic neuropils of LGMD in either compound eye of locust: each LGMD visual pathway adopts a four-layer structure: retina, lamina, medulla and lobula.

Klapoetke et al., 2017a; Rind and Simmons, 1998; Rind et al., 2016). Their visual system mechanisms have inspired and been successfully integrated into artificial vision systems, including robotic platforms for ground and aerial navigation (Franceschini, 2014; Serres and Ruffier, 2017; Fu et al., 2018a).

Motivated by the visual systems of flies, numerous optical flow (OF)-based perception methods have emerged, particularly drawing on insights from lobula plate tangential cells (LPTCs) (Serres and Ruffier, 2017). These OF-strategies rely on local pixel-wise motion cues to guide diverse navigational behaviors in flies, including tunnel crossing (Wang et al., 2021), terrain following (Wang et al., 2019), collision avoidance (Bertrand et al., 2015), etc., and offer significant advantages in computational simplicity. Consequently, they have been widely adopted in robotic applications, especially in aerial platforms such as micro aerial vehicles (Green et al., 2004; Green and Oh, 2008; Milde et al., 2015). In recent years, research has expanded beyond LPTCs to investigate and model deeper visual projection neurons located in the insect brain, specifically lobula plate/lobula column type-2 (LPLC2) neurons (Klapoetke et al., 2017b; Zhou et al., 2022; Zhao et al., 2023). These neurons exhibit exceptional selectivity, responding preferentially to looming objects approaching from the visual center (Hua et al., 2022). Additionally, models incorporating LPLC2 neurons have been further enhanced by integrating attention mechanisms, enabling effective detection and localization of multiple looming targets in complex, natural environments (Liu and Fu, 2025).

Another type of looming perception models draws on the physiological mechanisms of locusts' visual systems. Specifically, a group of neurons called lobula giant movement detectors (LGMD, LGMD1 and LGMD2) have been identified in locusts' optic lobe (Simmons and Rind, 1997; Rind, 2002; Gabbiani et al., 2002, 2004; Rind and Bramwell, 1996). Such neurons are highly effective at detecting objects that move in depth signaling collision danger, and responsible for triggering escape behavior. Despite LGMD1 and LGMD2 are adjacent to each other and sharing close physiological characteristics, these neurons exhibit distinct loom-selectivity: the LGMD1 is able to detect both darker and brighter approaching targets (OFF and ON contrasts), whereas the LGMD2 specifically responds to darker approaching objects (OFF contrast). In recent two decades, by incorporating different modeling theories to shape the loom-selectivity of LGMD (Rind and Bramwell, 1996; Yue and Rind, 2006; Keil, 2011; Bermúdez i Badia et al., 2010; Fu et al., 2018b; Keil et al., 2004; Salt et al., 2019; Fu et al., 2020a, 2023; Zhao et al., 2024), the LGMD-based single-neuron models have been investigated extensively with many successful application in machine vision and navigation (Fu et al., 2019, 2018a; Fu, 2023). Compared to fly-inspired optical flow (OF)-based approaches, LGMD models exhibit a distinct preference for detecting directly approaching targets, relying on global rather than local motion cues. Furthermore, unlike LPLC2-based models, LGMD systems show greater flexibility in detecting looming motion initiated from various spatial positions within the visual field.

Although existing neural models have shown robustness and effectively addressed some real-world

collision detection problems, achieving consistent loom-selectivity remains a challenge, especially in dynamic physical scenes that involve multiple types of movements, such as translation, rotation, and recession. This limitation highlights the inherent constraints of single-neuron computations.

To overcome these deficiencies, researchers have explored the modularization of neural models or networks to accomplish complex perceptual tasks (Kelkar and Medaglia, 2021; Amer and Maul, 2019; Bertolero et al., 2015; Meunier et al., 2010; Clune et al., 2013; Amer, 2019). Modularity refers to the capacity of a neural or artificial system to be decomposed into relatively independent, reusable, and combinable subsystems (modules). Building upon this principle, Li et al. developed a composite model combining LGMD1 (Fu et al., 2018b) and LGMD2 (Fu et al., 2020a) neural network structures. Their proposed modular approach, presented in two variations—LGMD1-LGMD2 and LGMD2-LGMD1—significantly improved looming selectivity by effectively reducing interference from translating and receding stimuli (Li et al., 2023). Specifically, each neural model employs a four-layer structure to emulate the presynaptic neuropils of either LGMD1 or LGMD2 neural circuit (see Fig. 1). Each model functions as an independent computational unit, or module, which can be combined sequentially, with the output of the preceding module serving as the input to the next.

In our previous work related to this paper (Wang et al., 2024), we optimized both the time and space complexity of this composite LGMD model, successfully implementing it within the embedded vision modules of ground-based micro-robots characterized by limited computational resources. In robotic implementations, the proposed composite model significantly enhances looming selectivity, enabling robots to respond exclusively to approaching targets and achieve collision-free navigation within an arena (Wang et al., 2024). Specifically, the two neuronal modules mutually reinforce each other, with the preliminary looming selectivity generated by the first module refined and strengthened by the second, irrespective of their sequence in the model.

This observation prompts several questions: (1) What core structure within the second module is essential for enhancing loom-selectivity? (2) Is it necessary for this second module to retain the entire four-layer structure, or could it be simplified without compromising performance?

To identify the core structural component responsible for enhancing looming selectivity, we initially conducted offline redundancy experiments focused on the second module of the composite model. The results clearly demonstrated that introducing an additional time-derivative operation is the critical factor behind the improved loom-selectivity. Consequently, we developed a more economical neuronal assembly model that emphasizes the elevated temporal derivative in the cascaded module. We compared the proposed optimized model with state-of-the-art methods and validated its superior performance in loom-selectivity across diverse real-world visual scenarios. Additionally, we conducted online comparative tests using micro-robots, visualizing internal layer representations of the embedded neuronal model and benchmarking its performance against our previously established model (Wang et al., 2024).

The systematic experimental results yield the following primary findings:

- The most significant finding is that simply introducing an additional temporal derivative operation after the first-stage module—specifically, further processing limited to the retinal layer of the second cascade—significantly enhances loom-selectivity. This result confirms that elevating the temporal derivative is crucial for strengthening looming-selective responses.

- We propose a time-derivative-based model that substantially streamlines the neural architecture without compromising the loom-selectivity of the original composite model. This creates an efficient visual perception module well-suited for mobile robots with limited computational resources.

- Our research also provides plausible insights into neuroscience, i.e., the efficacy of locust LGMD neuronal ensembles, suggesting that these loom-selective neurons might achieve optimal looming perception through interconnected neural assemblies rather than operating independently.

Section 2 details the proposed computational model. Experimental results and analysis are presented in Section 3. Section 4 summarizes this paper and points out future work.

2 FORMULATION OF THE COMPUTATIONAL FRAMEWORK

In this section, we present the algorithms of the proposed model in detail, which is essentially a derivative of the LGMD2-LGMD1 model, as shown schematically in Fig. 2.

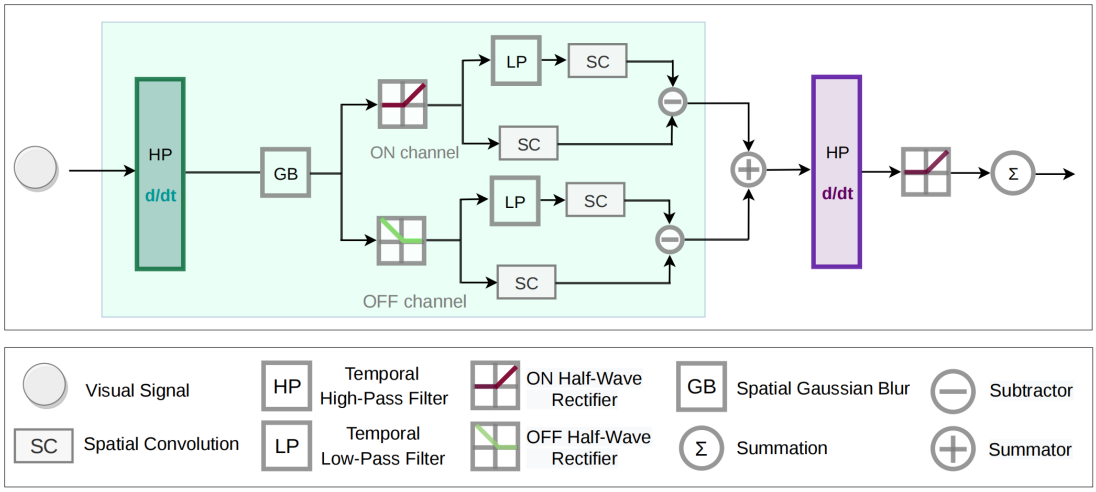

Figure 2. Schematic diagram of the proposed computational framework with raised temporal derivative by feeding the output of first neuronal module (light blue box) as input to the second: for simplicity, only one single neuronal processing pipeline is shown.

2.1 Model Description

This time derivative model, also called LGMD2-Derivative, consists of two modules: the first module includes the retina, lamina and medulla layers in the LGMD2 model, and the second one only consists of the retina layer and the final LGMD1 cell for output, rendering the intermediate layer processing redundant. Specifically, the first retina layer (P layer) is used to extract changes of light intensities with respect to time. The lamina layer comprises ON/OFF-type cells that split visual motion information into two parallel processing pathways. Such mechanism is well established at the preliminary stages of visual signal processing and has been found in the visual systems of many animal species, such as invertebrates like flies (Borst and Euler, 2011) and dragonflies (Geurten et al., 2007), and vertebrates like rabbits (Borst and Euler, 2011) and cats (Troyer et al., 1998). The medulla layer, including excitation units (E layer), inhibition units (I layer), and summation units (S layer), is a key component for achieving loom-selectivity. The second retina layer raises temporal derivative of outputs from the previous LGMD2 module with an additional high-pass filter. The LGMD1 neuron integrates information from its retina layer and activates membrane potentials. Finally, the temporal potential is transformed to frequency domain using a spiking mechanism in order to indicate potential collision threats towards artificial and robotic systems.

It is worth noting that in our preliminary research (Wang et al., 2024), we observed that the LGMD1-LGMD2 model exhibited slightly lower robustness and selectivity compared to the LGMD2-LGMD1 configuration. Therefore, this paper specifically focuses on the LGMD2-LGMD1 model to clearly identify the core structural component responsible for enhancing looming selectivity.

2.2 Formulation

Here we present the formulation of proposed simplified LGMD2-Derivative model with emphasis laid upon the elevated time derivative. Our original composite model (Wang et al., 2024), named LGMD2-Excitation model for comparison, is elucidated in the Appendix.

2.2.1 LGMD2—Retina layer (The first time derivative)

The LGMD2's retina is the first computational layer of the proposed time-derivative model and consists of photoreceptors arranged in a matrix. Each receives stimuli that presented in image streams at gray-scale luminance from the external environment and further extracts changes between two consecutive frames, with respect to time, as the following

$$M(x,y,t) = \alpha_1 \cdot \big(L(x,y,t) - L(x,y,t-1) + M(x,y,t-1)\big), \tag{1}$$

where $L(x,y,t) \in \mathbb{R}^3$ denotes the pixel values of the input image, (x, y) represents spatial position of each receptor and (t) represents temporal position. α_1 is a delay coefficient, and is calculated by

$$\alpha_1 = \frac{\tau_1}{\tau_1 + \tau_{in}}, \tag{2}$$

where τ_1 and τ_{in} are time delay constant and the time interval between two consecutive frames, respectively.

Subsequently, a normalized Gaussian kernel of size 3×3 with a standard deviation at 1 is applied to $M(x,y,t)$ for blurring to suppress irrelevant noise, defined as follows

$$P(x,y,t) = \iint M(u,v,t) G_p(x-u, y-v)\, du\, dv. \tag{3}$$

Furthermore, when a large number of photoreceptor cells in the visual field are activated, such as in scenes of impending collision, or significant changes in the external environment like robot turning, they exerts strong, whole-field inhibition directly upon LGMD neurons. To simulate this, an adaptive inhibition mechanism (AIM) was introduced to adjust the time-varying biases within ON/OFF pathways (Fu et al., 2020a). The mathematical expression is defined as

$$PM(t) = \sum_{x=1}^{R} \sum_{y=1}^{C} |M(x,y,t)| \cdot (C \cdot R)^{-1}, \tag{4}$$

among them, R and C refer rows and columns of the matrix, and $C \cdot R$ simulates the total number of receptors in the visual field. Moreover, the AIM mechanism has a slight delay, that is

$$\hat{PM}(t) = (PM(t), \hat{PM}(t-1), \hat{PM}(t-2)) \cdot \overrightarrow{\theta}, \tag{5}$$

where $\overrightarrow{\theta}$ is the delay coefficient vector $(0.6, 0.3, 0.1)'$. The time-varying biases are then given by

$$\omega_1(t) = \max(\omega_3, \frac{\hat{PM}(t)}{T_{PM}}),\ \omega_2(t) = \max(\omega_4, \frac{\hat{PM}(t)}{T_{PM}}), \tag{6}$$

ω_3 and ω_4 denote the different bias baselines in ON and OFF pathways, respectively.

2.2.2 LGMD2—Lamina layer

The lamina layer receives signals from the retina layer, in which ON-type neurons respond to dark-to-light contrast polarity (ON-contrast), transmitting luminance increases into the ON pathway, and OFF-type neurons respond to light-to-dark contrast polarity (OFF-contrast), transmitting luminance decreases into the OFF pathway. In summary, the whole process can be achieved by half-wave rectify mechanism as

$$P_{on}(x,y,t) = [P(x,y,t)]^+ + \beta \cdot P_{on}(x,y,t-1), \tag{7}$$

$$P_{off}(x,y,t) = -[P(x,y,t)]^- + \beta \cdot P_{off}(x,y,t-1), \tag{8}$$

where $[x]^+$ is $max(0,x)$, and $[x]^-$ is $min(x,0)$. β denotes a delay coefficient that signifies the fraction of the previous signal allowed to pass.

2.2.3 LGMD2—Medulla layer

After that, the medulla layer continues to handle visual information by incorporating competition between excitatory and inhibitory signal flows.

In the ON channels, convolution operation is conducted on P_{on} to mimic the spread of local excitation. On the other hand, the surrounding delayed excitation D_{on} is obtained using first-order low-pass filtering, and then generates the local Inhibition I_{on}. The whole spatiotemporal computation is given by

$$E_{on}(x,y,t) = \iint P_{on}(u,v,t) W_1(x-u, y-v)\, du\, dv, \tag{9}$$

$$D_{on}(x,y,t) = (E_{on}(x,y,t), D_{on}(x,y,t-1), D_{on}(x,y,t-2)) \cdot \overrightarrow{\alpha_{on}}, \tag{10}$$

$$I_{on}(x,y,t) = \iint D_{on}(u,v,t) W_{I_{on}}(x-u, y-v)\, du\, dv, \tag{11}$$

where W_1 stands for a convolution kernel in Eq. (12) as

$$W_1 = \frac{1}{8}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 8 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \tag{12}$$

and $\overrightarrow{\alpha_{on}}$ is the delay coefficient vector $(0.6, 0.2, 0.2)'$.

In the OFF channels, the generation of local excitation (E_{off}), local inhibition(I_{off}) and surrounding delayed excitation (D_{off}) accords well with those in the ON channels, but with different delay coefficient vector $\overrightarrow{\alpha_{off}} = (0.4, 0.3, 0.3)'$. The computation is as the following

$$E_{off}(x,y,t) = \iint P_{off}(u,v,t) W_1(x-u, y-v)\, du\, dv, \tag{13}$$

$$D_{off}(x,y,t) = (E_{off}(x,y,t), D_{off}(x,y,t-1), D_{off}(x,y,t-2)) \cdot \overrightarrow{\alpha_{off}}, \tag{14}$$

$$I_{off}(x,y,t) = \iint D_{off}(u,v,t) W_{I_{off}}(x-u, y-v)\, du\, dv, \tag{15}$$

where $W_{I_{off}}$ and $W_{I_{on}}$ are defined as

$$W_{I_{off}} = \frac{1}{32}\begin{bmatrix} 1 & 2 & 4 & 2 & 1 \\ 2 & 4 & 8 & 4 & 2 \\ 4 & 8 & 16 & 8 & 4 \\ 1 & 2 & 4 & 2 & 1 \\ 2 & 4 & 8 & 4 & 2 \end{bmatrix}, \; W_{I_{on}} = 2 * W_{I_{off}}. \tag{16}$$

Following the above computations, the local summation units in the medulla layer linearly integrate local excitation and inhibition in the ON/OFF pathways, respectively, which reflects the critical race between excitation and inhibition. The formulas are presented as

$$S_{on}(x,y,t) = [E_{on}(x,y,t) - \omega_1(t) \cdot I_{on}(x,y,t)]^+, \tag{17}$$

$$S_{off}(x,y,t) = [E_{off}(x,y,t) - \omega_2(t) \cdot I_{off}(x,y,t)]^+. \tag{18}$$

Finally, there are linear summation between the ON/OFF channels to form the output feature representation of LGMD2 module, that is

$$S(x,y,t) = S_{on}(x,y,t) + S_{off}(x,y,t). \tag{19}$$

2.2.4 LGMD1—Retina layer (The second time derivative)

The second LGMD1's retina layer plays a central role in this computational framework which works as the entry of second cascaded module, elevates the time derivative from the feature representation of first module. This works essentially as another high-pass filter, mathematically defined as

$$TD(x,y,t) = S(x,y,t) - S(x,y,t-1). \tag{20}$$

Subsequently, the output is rectified to only allow positive signals to pass through,

$$\phi(x,y,t) = [TD(x,y,t)]^+ + \beta \cdot \phi(x,y,t-1). \tag{21}$$

2.2.5 LGMD1—Neuronal integration and activation

The LGMD1 neuron integrates its retina's remaining signals to generate membrane potential $k(t)$ which is then activated by a sigmoid function, in order to normalize $k(t)$ to gain $K(t)$ within $[0.5, 1)$. The whole process is mathematically described as

$$k(t) = \sum_{x=1}^{R}\sum_{y=1}^{C} \phi(x,y,t), \; K(t) = \left(1 + e^{-k(t)\cdot(C\cdot R\cdot\alpha_2)^{-1}}\right)^{-1}, \tag{22}$$

the α_2 represents a scale factor to avoid saturation of the activation function.

In order to suppress the effect of irrelevant motion stimuli such as translating and receding objects on the model to further sculpt loom-selectivity, the spike frequency adaptation mechanism (SFA) in a previous modeling study (Fu et al., 2018b) is also applied as

$$\hat{K}(t) = \begin{cases} \alpha_3 \cdot \left(\hat{K}(t-1) + K(t) - K(t-1)\right), & \text{if } \left(K(t) - K(t-1)\right) \leq T_{sfa} \\ \alpha_3 \cdot K(t), & \text{otherwise} \end{cases}, \tag{23}$$

where T_{sfa} is the threshold constant of SFA mechanism, α_3 indicates the adaptation rate, calculated by a time constant τ_s as

$$\alpha_3 = \frac{\tau_s}{\tau_s + \tau_{in}}. \tag{24}$$

2.2.6 Spiking frequency

The temporal membrane potential $\hat{K}(t)$ herein is transformed to frequency domain using an exponential mapping as

$$Spi(t) = \left[e^{(\alpha_4 \cdot (\hat{K}(t) - T_{sp}))}\right], \tag{25}$$

where T_{sp} is the predefined spike threshold, α_4 is a scale parameter that influences the firing rate. Finally, the potential collision threat can be indicated by computing the spike frequency within a specified time window. That is

$$Col(t) = \begin{cases} True, & \text{if } \left(\sum_{i=t-n_t}^{t} Spi(i)\right) \times 1000/(n_t \cdot \tau_{in}) \geq T_c \\ False, & \text{otherwise} \end{cases}, \tag{26}$$

where n_t denotes the time window to update spike frequency, and T_c is collision warning threshold. This is particularly useful for initiating urgent avoidance behaviors in intelligent machines like mobile robots.

2.3 Model Parameters Setting

Table 1. Setting parameters

Parameter	Description	Value
τ_1	time delay constant	100(ms)
τ_{in}	time interval	$15 \sim 50$(ms)
$\{\omega_3, \omega_4\}$	local inhibition bias baseline	{0.6, 0.3}
T_{PM}	AIM threshold	$10 \sim 50$
$\{C, R\}$	columns and rows	adaptable
β	delay coefficient	0.1
τ_s	time delay constant in SFA	$500 \sim 1000$(ms)
α_4	scale parameter of firing rate	4
T_{sp}	spiking threshold	0.7
n_t	time window	10(frames)
T_c	collision warning threshold	$15 \sim 20$(Hz)

Table 1 summarizes the parameter settings used in the proposed time-derivative model. Certain parameters depend directly on the physical characteristics of the input videos: specifically, C and R correspond to the width and height of the image streams, respectively, while τ_{in} is determined by the video frame rate. The remaining parameters were empirically selected based on prior studies (Li et al., 2023; Wang et al., 2024), balancing optimal model performance in both computational simulations and robotic implementations. Given the limited number of adjustable parameters, explicit learning processes were not within the scope of this model-based study but represent a valuable direction for future research.

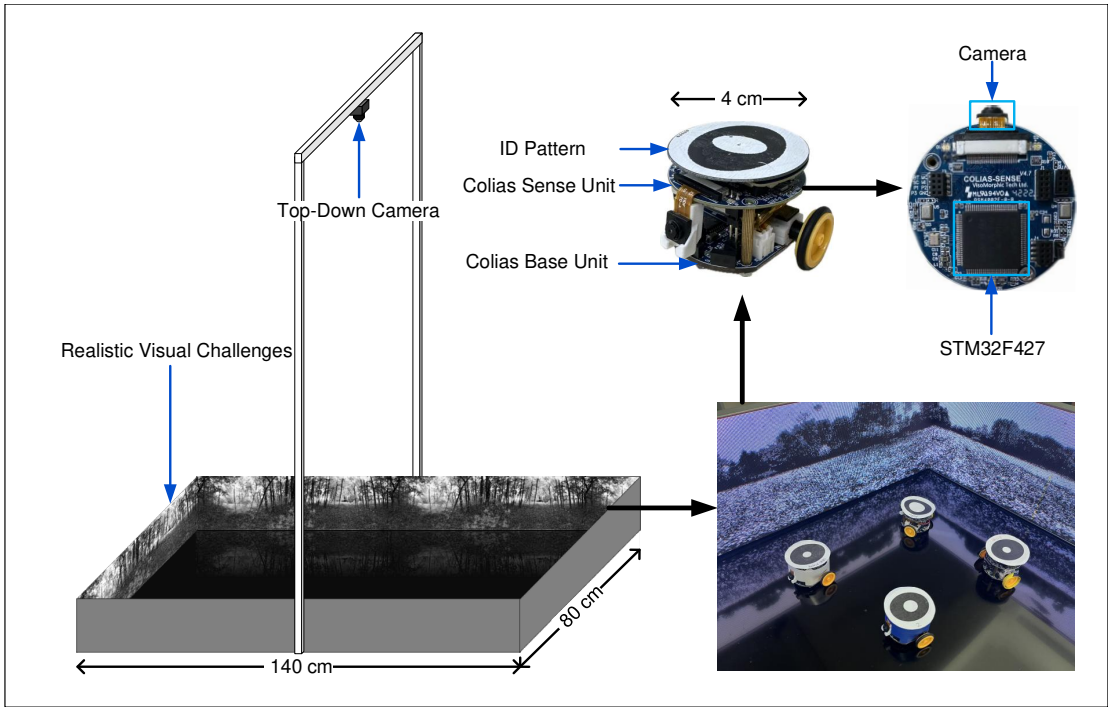

Figure 3. Illustration of the experimental arena and the structure of grounded micro-robot *Colias*. The surrounding LED walls can display different scenes to challenge visual systems. The top-down camera was utilized to localize the robots with ID-specific pattern on each top, with respect to time. The proposed model was used as the only collision sensing modality in arena tests.

3 EXPERIMENTAL VALIDATION

In this section, we present the experiments, including offline and online tests. For offline experiments, we used physical stimuli composed of indoor and outdoor scenes to test and compare the proposed model with state of the arts. For online experiments, we integrated the proposed approach into the vision module of a ground-based micro-robot *Colias* (Hu et al., 2018) for arena tests with layer visualization from robotic visual processing. We compared the proposed model with related state of the arts: the early LGMD2-LGMD1 composite model (Li et al., 2023), the LGMD1 model (Fu et al., 2018b), the LGMD2 model(Fu et al., 2020a), the hybrid LGMD model (Fu et al., 2021), and the LGMD2-Excitation model (algorithms in the Appendix). These experiments aimed to explore the key component in neuronal assemblies which essentially boosts loom-selectivity, and verify its value in robotic real-time implementation.

3.1 Experimental Setting

3.1.1 Offline tests setting

In the offline tests, both the proposed and comparative models were implemented in computational simulations, and their results were visualized for analysis. The visual data used in this study were adapted from two previous works (Fu et al., 2020b; Qin et al., 2024), and were recorded in real-world indoor and outdoor environments. Specifically, the indoor stimuli consisted of a black ball performing approaching, receding, and translating movements, captured at sampling rates of 30 fps, 30 fps, and 60 fps, respectively. The outdoor stimuli included vehicle collisions and UAV approaching and retreating from an obstacle cylinder, recorded at 30 fps, 20 fps, and 60 fps, respectively. To reduce the computational burden and unify the input stimuli dimension, we applied bi-linear interpolation to scale the resolution of all visual sequences to 100×100.

3.1.2 Online tests setting

In the online tests, we implemented the proposed approach as embedded vision module in the ground-based micro-robot *Colias* (Hu et al., 2018). There were two primary reasons for selecting *Colias* as the test platform: (1) it is a lightweight robot, measuring approximately 4 cm in diameter and weighing 50 grams,

capable of responding quickly and accurately to motion commands; and (2) it is widely used in swarm robotics research (Liu et al., 2021a,b), and is among the smallest and most cost-effective micro-robots in the field.

As illustrated in Fig. 3, the *Colias* robot comprises two main components: the Base Unit (CBU) and the Sensing Unit (CSU). The CBU, a circular platform at the robot's base, is responsible for motion and power management. It is equipped with two micro DC motors and 2.2 cm diameter wheels, enabling a maximum speed of 35 cm/s. A 3.7 V, 320 mAh battery supports autonomous operation for up to two hours on a full charge. The CSU, mounted on top of the robot, serves as the vision processing module. It features an ARM Cortex-M4F microcontroller and a compact OV7670 CMOS image sensor. The camera provides a 70-degree field of view, operating at 30 fps with a resolution of 72×99 pixels—balancing computational efficiency and visual fidelity.

As shown in Fig. 3, experiments were conducted in an arena measuring 1.4×0.8 meters, enclosed by LED display walls decorated with static or dynamic grating patterns. A top-down camera was installed above the arena to record the motion and performance of the *Colias* robot. Additionally, specific ID markers were placed on top of each robot, allowing a localization system (Krajník et al., 2014) to track trajectories and calculate the collision avoidance success rate.

3.2 Metrics

To evaluate the proposed and comparative methods, we adopted the confusion matrix and F_1 score to quantify their performance on looming detection. The confusion matrix reflects the prediction results of each model with the basic form as

$$\begin{bmatrix} \mathrm{TN} & \mathrm{FP} \\ \mathrm{FN} & \mathrm{TP} \end{bmatrix}, \tag{27}$$

among them, TP (true positive) refers to the number of instances in which the model correctly generates a collision warning in response to looming stimuli. FP (false positive) denotes the number of instances where the model incorrectly triggers a collision warning for non-looming stimuli, such as translating or receding motion. TN (true negative) is the number of instances where the model correctly withholds a collision warning for non-looming stimuli. FN (false negative) represents the number of instances where the model fails to detect and respond to looming stimuli. F_1 score is the weighted average of precision and recall, which is defined as

$$F_1 = 2 \cdot \frac{\mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}. \tag{28}$$

In this research, we focused on analyzing F_1 score, with which a larger F_1 value indicates a better looming perception ability of the model. The formulas for precision and recall values are respectively defined as

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{29}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{30}$$

3.3 Redundancy Experiments

Firstly, we conducted redundancy experiments to investigate the key mechanism that improves the loom-selectivity. We compared the response and selectivity of the proposed LGMD2-Derivative model with the LGMD2-Excitation, with inhibition processing ablated from our previous model (Wang et al., 2024), and the original composite LGMD2-LGMD1 model (Li et al., 2023). The formulas of the LGMD2-Excitation model are given in the Appendix.

As shown in Fig. 4 and Fig. 5, the proposed model exhibits behavior comparable to the composite LGMD2-LGMD1 model in both indoor and outdoor environments, responding exclusively to approaching objects while showing no activation in response to translating or receding stimuli. In contrast, the LGMD2-Excitation model is highly sensitive to translating motion and vehicle collision stimuli, demonstrating significantly weaker looming selectivity than the other two models. To ensure the reliability of our findings and mitigate the influence of chance, all three models were evaluated using 104 real-world video sequences, 70% of which contained looming stimuli. The statistical outcomes—including confusion matrices, precision, recall, and F_1 scores—are presented in Fig. 4, Fig. 5, Fig. 6, and Table 2. The results

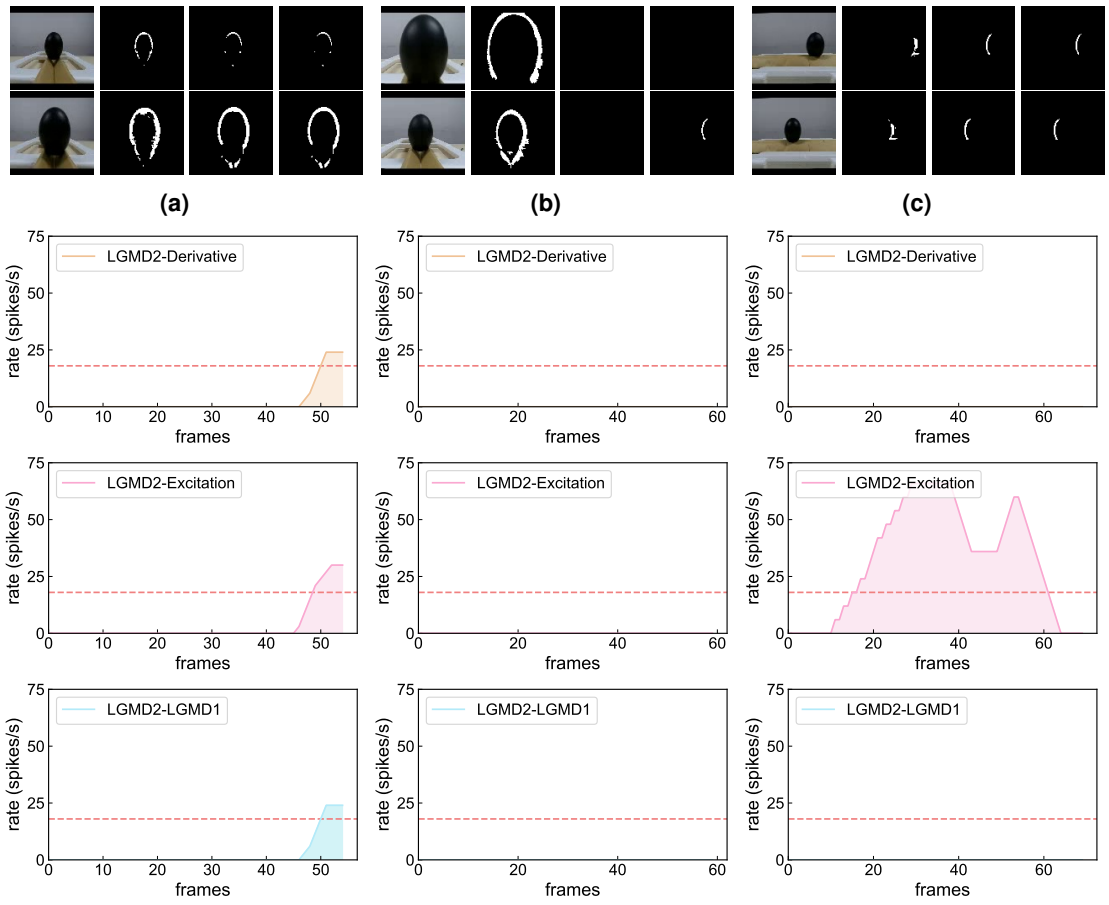

Figure 4. Offline testing results of model selection studies in indoor scenes: LGMD2-Derivative indicates the proposed model with just raised time derivative. LGMD2-Excitation indicates the composite model (Wang et al., 2024) with inhibition ablated in the second LGMD1 module. LGMD2-LGMD1 model is the intact composite model (Wang et al., 2024) from earlier research. Spiking frequency, with respect to time, is represented for each of the three models, stimulated by the dark ball approaching (a), receding (b) and translating (c) within the field of view. For each stimulus, the first column is the input image, the second column is the representation of retina layer in LGMD2 module, the third column is the output layer in LGMD2, and the fourth column is the retina layer in LGMD1 module. Red dotted line refers to collision warning threshold at 18 Hz. **The proposed simplified model and full composite LGMD2-LGMD1 model respond only to looming motion as the excitation retains after raising the time derivative at the second module.**

Table 2. F_1 score (%) of three models to 104 real-world sets in total

Criterion	LGMD2-Derivative	LGMD2-Excitation	LGMD2-LGMD1
Precision	73.77%	27.27%	**90.74%**
Recall	83.33%	**96.00%**	76.56%
F_1 Score	78.26%	42.47%	**83.05%**

show that the proposed model's F_1 score is only 4.79% lower than that of LGMD2-LGMD1, suggesting a marginal decrease in looming detection performance, though the difference is not statistically significant. In contrast, the LGMD2-Excitation model under-performs substantially, with its F_1 score 35.79% and 40.58% lower than those of LGMD2-Derivative and LGMD2-LGMD1, respectively—indicating poor looming recognition capability. Moreover, feature visualizations in Fig. 4 and Fig. 5 reveal that as

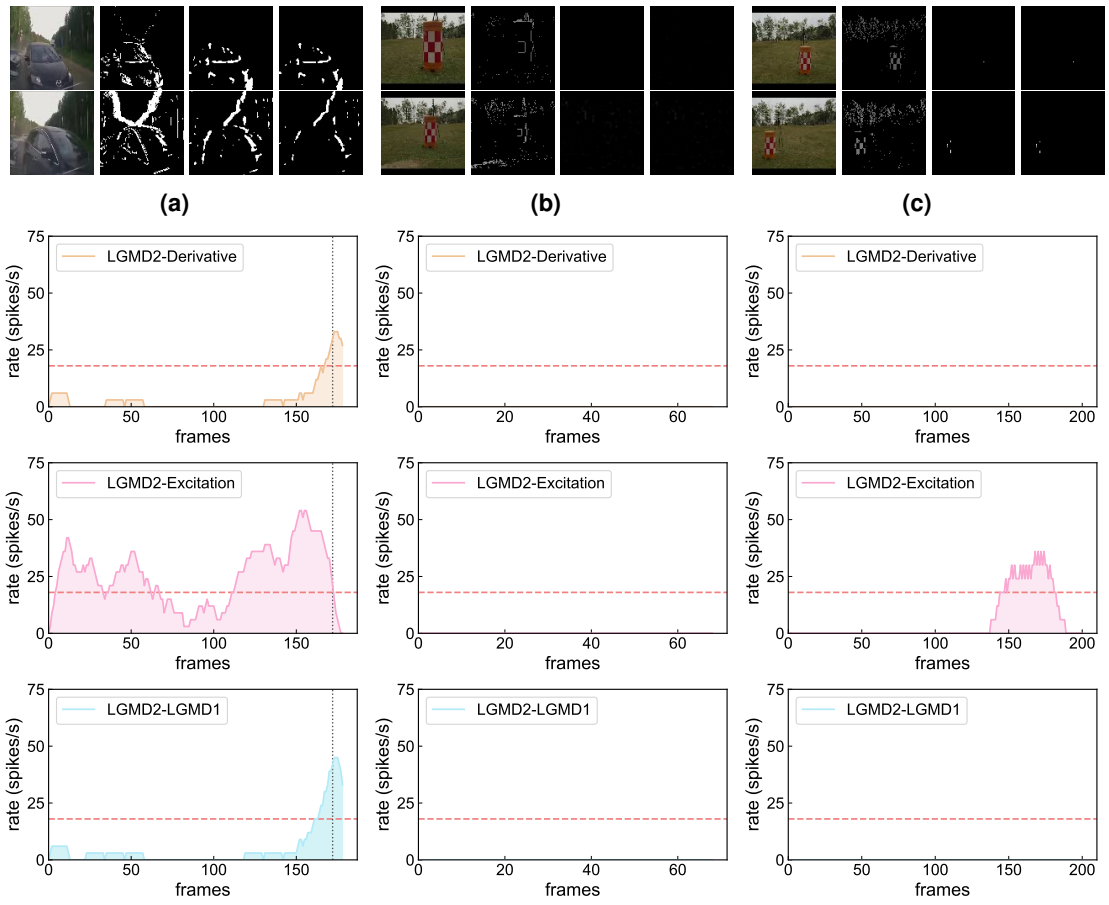

Figure 5. Offline testing results of model selection studies challenged by outdoor recordings from UAV and car (a) crash, (b) recession, (c) shifting views. The vertical dashed line indicates the moment of ground-truth crash at frame 172. Other notations conform to Fig. 4. **The proposed simplified model and full composite LGMD2-LGMD1 model work effectively to extract only looming motion in complex dynamic scenes.**

visual signals propagate deeper into the network, only looming-related features remain after applying the elevated temporal derivative at the retina layer of the second LGMD1 module.

In conclusion, the experiments demonstrate that strong loom-selectivity can be achieved after visual information passes through only the retina layer of the second module, suggesting that a full four-layer structure is redundant. Crucially, the results highlight that elevating the temporal derivative at the output of the first neuronal unit is key to improving loom-selectivity. This approach ensures the model selectively responds to approaching stimuli while significantly simplifying its structural complexity.

3.4 Comparative Experiments

Through the above experiments, we demonstrated that a complete four-layer structure in the cascaded neuronal module is redundant; rather, the incorporation of a temporal derivative computation is central to enhancing loom-selectivity. To further validate the effectiveness of the proposed time-derivative model in looming perception, we conducted comparative tests using real-world physical stimuli. The model's performance was benchmarked against state-of-the-art approaches, including two single-neuron models, LGMD1 (Fu et al., 2018b) and LGMD2 (Fu et al., 2020a), and a hybrid LGMDs neural system in which LGMD1 and LGMD2 operate in parallel with their spiking frequencies integrated (Fu et al., 2021).

3.4.1 Indoor structured scenes

In the first set of comparative experiments, we employed relatively simple indoor stimuli to evaluate model performance. Fig. 7 illustrates the spiking frequency outputs of each model across three test

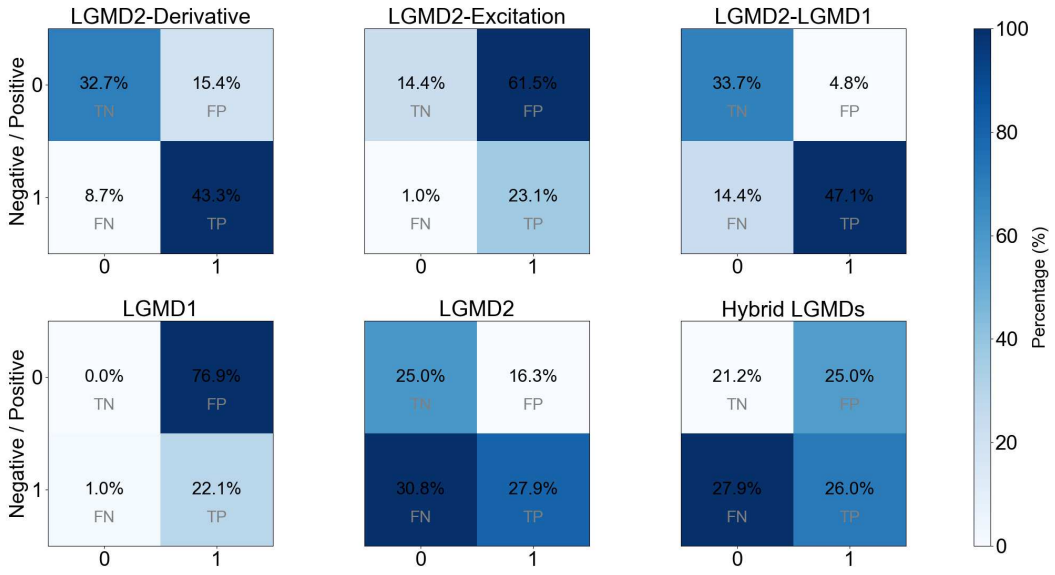

Figure 6. Confusion matrices for six comparative models on real physical datasets. Each matrix shows true positive (TP), false positive (FP), true negative (TN) and false negative (FN). **Compared with the other models, the proposed simplified neuronal assembly model and the full LGMD2-LGMD1 composite model have higher rates of TP and TN challenged by car and UAV scenarios, which demonstrate enhanced loom-selectivity with raised time derivative.**

scenarios. Overall, the results reveal clear differences in activation behavior among the four models. All models successfully detect the approaching black ball; however, the three comparative models exhibit susceptibility to non-looming stimuli. Specifically, LGMD1 shows transient activation to receding motion and strong responses to translation. Although LGMD2 and the hybrid LGMDs do not respond to receding stimuli, both exhibit significant activation to translational motion, misclassifying it as potential collision risk. In contrast, the proposed model robustly suppresses responses to both receding and translational motion, preserving selective activation only for genuine looming stimuli.

These experimental findings underscore a critical insight: compared to existing models, the proposed LGMD2-Derivative model more closely approximates the biological capability of loom-selectivity, thereby advancing the development of artificial visual systems for robust looming perception.

3.4.2 Vehicle and UAV scenes

In the second set of comparative experiments, we evaluated the proposed model in more complex and challenging outdoor scenarios to further assess its robustness and applicability in ground vehicle and UAV platforms. Fig. 8 presents the performance of all models in these scenarios. Consistent with the findings from the first experimental set, the proposed time-derivative model successfully detects imminent collisions while exhibiting no response to non-threatening stimuli, such as a UAV translating laterally or moving away from an obstacle (i.e., receding motion).

In contrast, the comparative models continue to misclassify these irrelevant motions—particularly translation and recession—as potential collision threats. Notably, the LGMD2 and the hybrid LGMDs system demonstrate accurate responses to actual looming stimuli but still exhibit partial activation under non-looming conditions. Additionally, LGMD1 generates a premature response when the approaching vehicle is still at a relatively distant position, indicating a lower selectivity threshold.

These results further confirm that the proposed model maintains strong loom-selectivity and robustness even in complex and naturalistic environments. Moreover, the findings highlight the potential of the proposed model for deployment in both terrestrial and aerial robotic platforms, where reliable and efficient collision perception is crucial under real-world operational constraints.

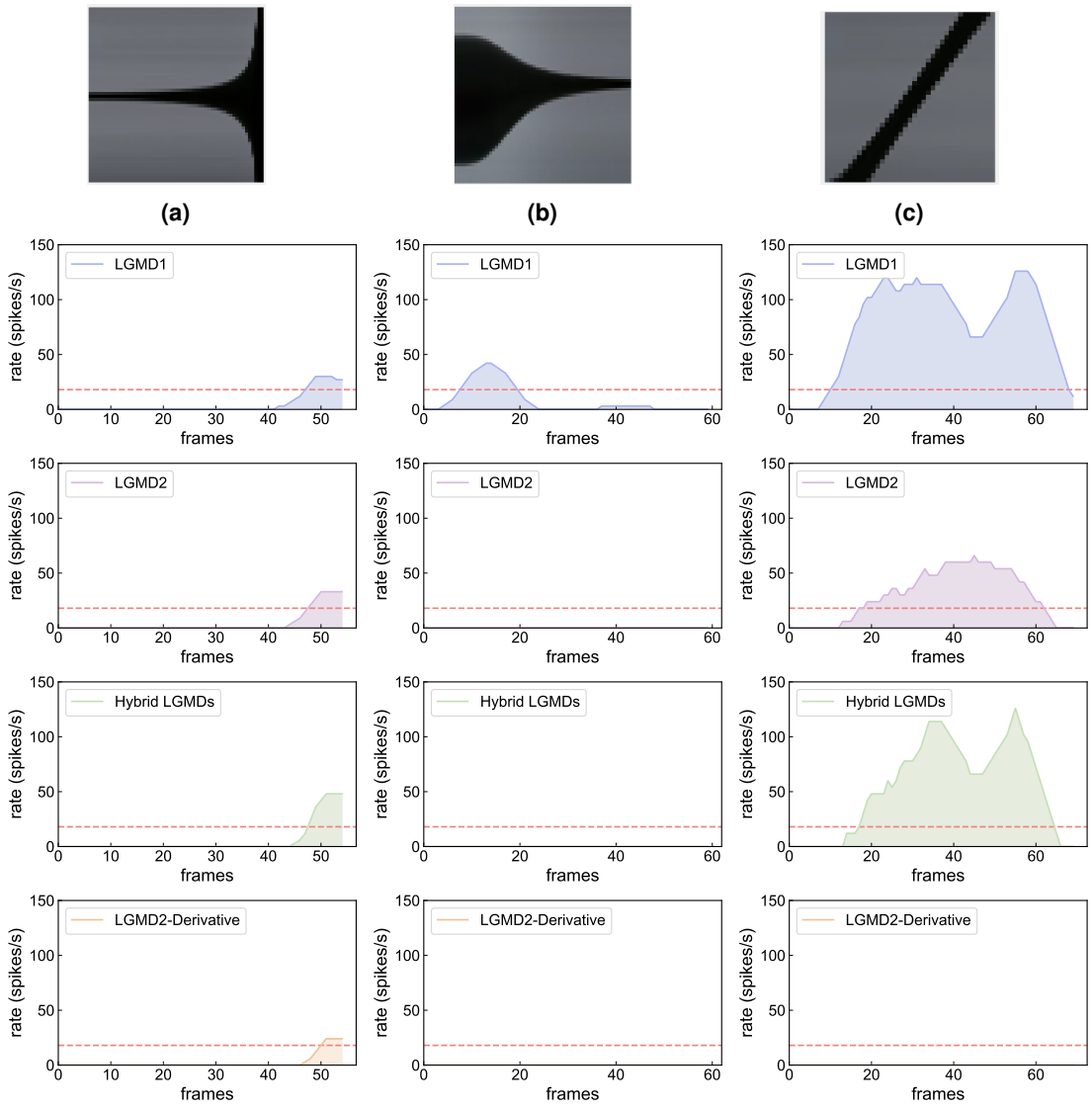

Figure 7. Comparative model responses against indoor visual movements, captured from a fixed camera: The stimuli on top are represented by intensity change of center line with respect to time. (a) dark ball (black pixels) approaching over time, (b) dark ball receding, (c) dark ball translating. Horizontal dashed line denotes threshold for collision warning. **The proposed model only responds to looming motion with enhanced selectivity.**

3.4.3 Statistical results

Finally, Table 3 summarizes the statistical results of precision, recall, and F_1 scores for the four evaluated models across various real-world scenarios, with corresponding confusion matrices illustrated in Fig. 6. Notably, the proposed model achieves the highest F_1 score of 78.26%, reflecting an optimal balance between precision and recall, and demonstrating a strong capability to accurately detect approaching objects. Moreover, the model exhibits both low false positive and false negative rates, indicating a reduced likelihood of false alarms and missed detections. In contrast, although LGMD1 achieves the highest recall (95.83%), its precision is markedly low (22.33%), suggesting that while the model is highly sensitive to looming stimuli, it frequently misclassifies non-looming motions—such as translation or recession—as collision threats. This tendency is clearly reflected in its confusion matrix, which reveals a substantial number of false positives. In summary, the proposed time-derivative model outperforms the comparative LGMD-based models in terms of both selectivity and robustness, offering a more reliable solution for looming perception in real-world visual environments.

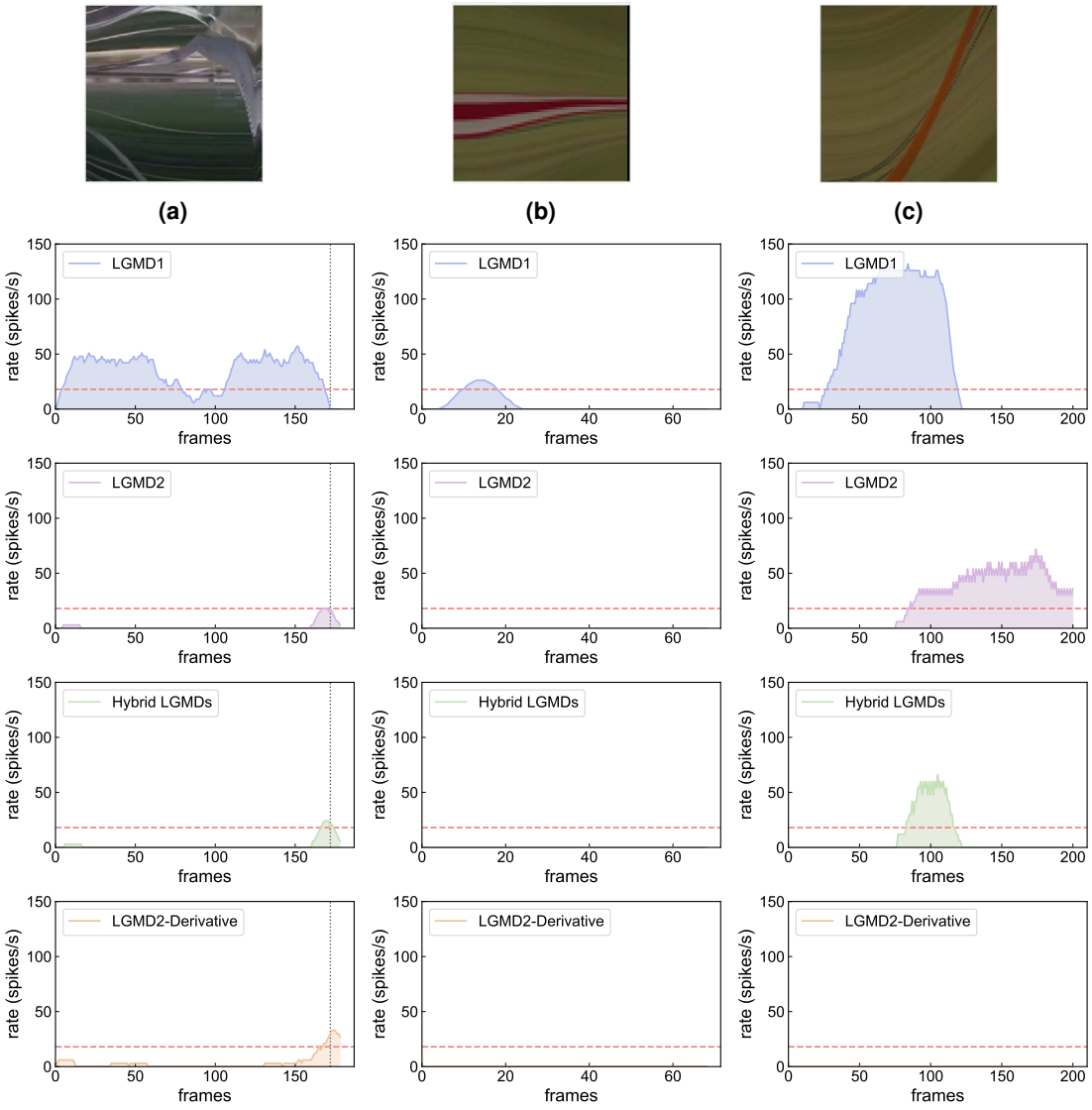

Figure 8. Comparative model responses against complex outdoor visual movements, captured from recordings by ground vehicle and UAV: The stimuli on top are represented by intensity change of center line with respect to time. (a) ground vehicle (gray pixels) approaching over time, (b) UAV recession from a cylindrical obstacle (marked with red-and-white pixels), (c) UAV shifting in front of the cylindrical obstacle. Horizontal dashed line denotes threshold for collision warning, vertical line indicates ground truth crash moment. **The proposed model only responds to looming motion with enhanced selectivity.**

3.5 Embodiment in Micro-Robot Vision

If the temporal derivative is indeed the core factor enhancing loom-selectivity, a critical question arises: does this mechanism retain its efficacy when embodied in a machine vision system? To address this, we implemented the proposed LGMD2-Derivative model on the *Colias* robot and conducted comparative experiments against the previously developed hybrid LGMDs model (Fu et al., 2021), which adopts a different strategy by integrating multiple neuron models in parallel. Specifically, the hybrid model fuses the outputs of LGMD2 and LGMD1 neurons through a parallel architecture, providing a representative benchmark for evaluating alternative combination strategies. This comparison enabled us to rigorously assess whether raising the temporal derivative, as proposed in our model, offers superior performance in robotic looming perception under real-world conditions. The robot and arena configuration are illustrated in Fig. 3.

Table 3. F_1 score (%) of four models to 104 real-world sets in total

Criterion	LGMD2-Derivative	Hybrid LGMDs	LGMD1	LGMD2
Precision	**73.77%**	50.94%	22.33%	63.04%
Recall	83.33%	48.21%	**95.83%**	47.54%
F_1 Score	**78.26%**	49.54%	36.22%	54.20%

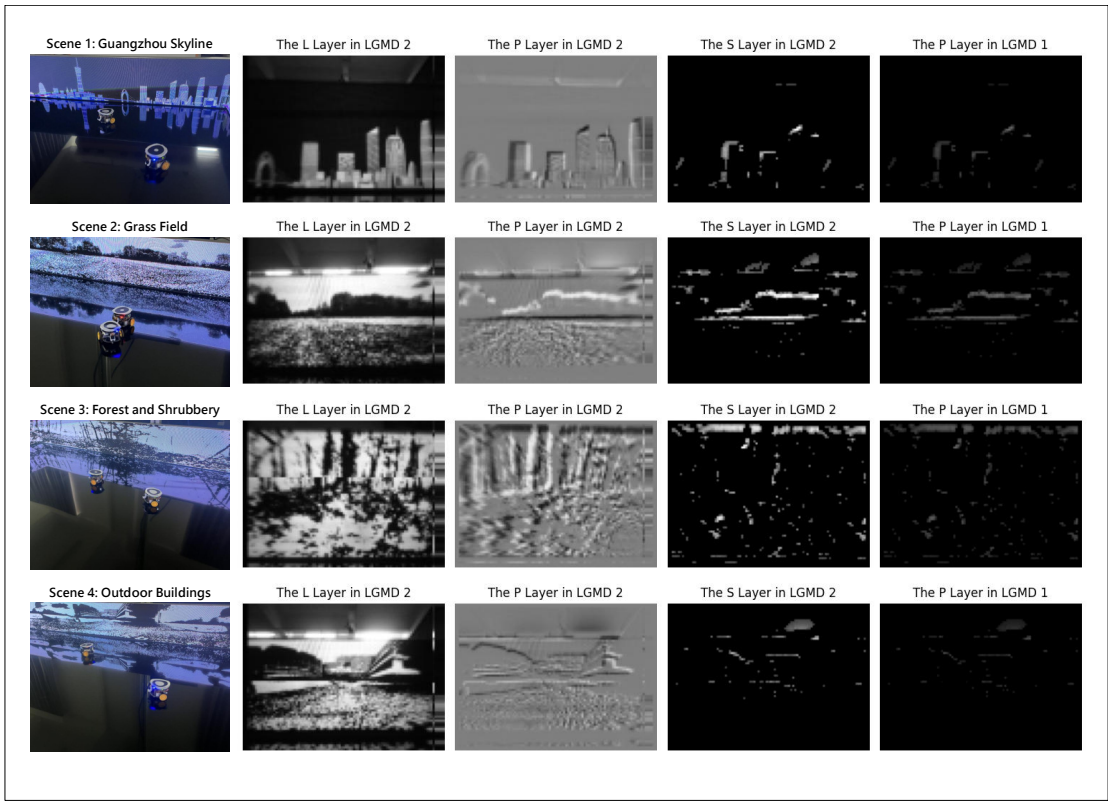

Figure 9. Four experimental scenarios (left column) in the arena setting, and the corresponding layer-wise visualizations of the proposed time-derivative model are shown. Scenes 1∼4 correspond to rows 1∼4, featuring a Guangzhou city skyline and three complex natural environments, respectively.

3.5.1 Robot visualization

What does the elevation of temporal derivative contribute to machine vision? To investigate the role of higher-order temporal derivatives in visual processing, we visualized the robot's perspective during forward motion toward the LED wall as exemplified in Fig. 3. The layer-wise neural responses of the proposed model across Scenes $1 \sim 4$ are presented in Fig. 9.

The visualizations reveal that the application of a second-order temporal derivative at the retina (P) layer significantly enhances the model's capacity to extract behaviorally relevant features. Specifically, only the most salient motion contours—corresponding to looming trajectories—remain after the derivative calculation. These contours are spatially aligned with actual approaching objects, highlighting the model's ability to isolate critical collision-related cues. In Scenes $1 \sim 2$, the final-layer responses clearly delineate the silhouettes of approaching objects, preserving only prominent expanding edges such as the top boundaries of buildings and distant tree lines. Conversely, in Scenes $3 \sim 4$—characterized by low-contrast boundaries and complex, noisy backgrounds—no obvious expansion features persist. However, faint motion traces remain sufficient to elicit a neural response, indicating that the model retains functional sensitivity to subtle dynamic changes.

To sum up, the elevation of temporal derivative thus acts as a selective amplifier for rapid, transient motion cues while suppressing static or slowly varying background features. This mechanism signif-

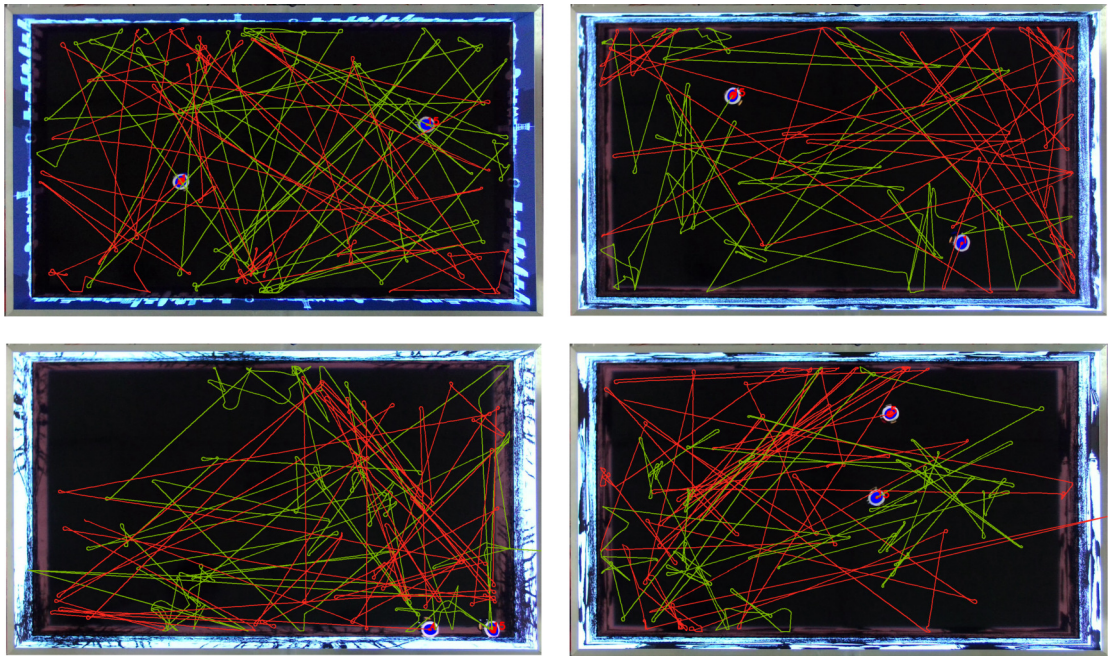

Figure 10. Robot arena tests results: overtime trajectories of two robots are illustrated where green one indicates the proposed LGMD2-Derivative model and red one denotes the comparative hybrid LGMD model. According to Fig. 9, four visual scenes as LED background were tested. Upper-left: Scene-1, Guangzhou skyline; Upper-right: Scene-2, Grass field; Bottom-left: Scene-3, Forest and shrubbery; Bottom-right: Scene-4, Outdoor buildings. For clarity, performance of 5-minute time window is shown for each scene. The linear velocity of each robot was configured constantly ≈ 9.75cm/s, and the turning angular speed was set at $150^\circ \pm 30^\circ$. **Generally, both model systems work effectively and robustly to guide micro-robot looming perception and collision avoidance in all tested scenes. The proposed model performs slightly better regarding success rate in all tested scenes as demonstrated in Table 4.**

Table 4. Success rate of collision avoidance demonstrated by two models in robot arena tests

Robot embodiment	Guangzhou Skyline	Grass Field	Forest/Shrubbery	Outdoor Buildings
Hybrid LGMD model	95.88%	83.21%	88.93%	92.09%
LGMD2-Derivative model	96.25%	90.55%	91.48%	93.16%

icantly enhances the model's robustness and loom-selectivity, reinforcing the quantitative advantages demonstrated in the preceding experimental sections.

3.5.2 Closed-loop arena tests

In the final type of robotic arena experiments, we conducted four sets of comparative online trials. The experimental setup followed the configuration described in Section 3.1.2, as illustrated in Fig. 3, Fig. 9, and Fig. 10. It is important to emphasize three key points: (1) the visual models under evaluation served as the sole collision detection mechanisms for guiding autonomous navigation, (2) no human intervention was involved during the trials, except in cases where the robot became immobilized against the arena boundaries, (3) each visual scene was tested under a 15-minute duration.

The results of the online robotic tests are summarized in Fig. 10, with the corresponding success rates for each of the four visual environments listed in Table 4. Overall, both the proposed LGMD2-Derivative model and the comparative hybrid LGMD model demonstrated effective and robust looming detection and collision avoidance capabilities across all tested scenes. As embodied in micro-robot vision, the LGMD2-Derivative model achieved slightly higher success rates than the hybrid LGMD model. Notably, both models performed best in scenarios featuring the Guangzhou skyline and outdoor buildings, where distinct object edges are more prominent. In contrast, their performance declined in the grass field

scenario, which presented naturally blurred textures and low-contrast boundaries.

To sum up, the most salient distinction between the two compound models lies in the order of the temporal derivative employed in visual information processing. Experimental results strongly suggest that elevating the order of temporal derivative is pivotal for enhancing model robustness, particularly when confronted with complex and dynamic real-world visual scenarios.

4 CONCLUSION

In this paper, we extended our preliminary investigation into LGMD neuronal assembly models to identify the core mechanism responsible for enhancing loom-selectivity in both computational simulations and robotic implementations. Our modeling study revealed that increasing the temporal derivative in continuous visual streams significantly improves the discrimination of looming stimuli against various other types of visual motion. Building upon this insight, we proposed a simplified, time-derivative-based neuronal model. Comparative evaluations demonstrated that this new model performs on par with the previously developed composite model (Wang et al., 2024) and outperforms state-of-the-art methods. When implemented in robotic platforms, the proposed model exhibited reduced computational complexity, enabling real-time processing. This efficiency makes it extremely advantageous for deployment in micro-mobile robots that have constrained computational resources.

Our future work will primarily, though not exclusively, address the following directions. First, as the current study focuses mainly on second-order temporal derivatives, we intend to further explore higher-order derivatives—especially at the retinal processing stage. We aim to determine whether looming detection performance continues to improve with increasing derivative order and identify an optimal derivative level for robust looming detection.

Second, directional selectivity mechanisms, such as elementary motion detectors (EMD), will be incorporated into the proposed LGMD2-Derivative model. The integration of EMD aims to simulate biologically inspired attention mechanisms more effectively. While our current model emphasizes overt attention by detecting centrally approaching objects, incorporating EMD-based modules would enable covert attention capabilities, allowing the detection of dynamic changes in the peripheral visual field.

Lastly, we plan to investigate the use of event-driven sensors as input streams for the proposed approach. Event-driven sensors are well-suited for real-time motion detection tasks due to their advantages in low latency, high dynamic range, low power consumption, and asynchronous signal processing.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China under Grant No. 62376063. †Qinbing Fu, Mengying Wang and Renyuan Liu share first authorship.

APPENDIX

This appendix supplements the paper by elaborating on the algorithms of LGMD2-Excitation model with inhibition ablated from our previous framework (Wang et al., 2024). The key difference is laid upon the neural processing of the second LGMD1 module.

LGMD2—Retina layer

$$M(x,y,t) = \alpha_1 \cdot \big(L(x,y,t) - L(x,y,t-1) + M(x,y,t-1)\big) \tag{a.1}$$

$$\alpha_1 = \frac{\tau_1}{\tau_1 + \tau_{in}} \tag{a.2}$$

$$P(x,y,t) = \iint M(u,v,t) G_p(x-u, y-v)\, du\, dv \tag{a.3}$$

$$PM(t) = \sum_{x=1}^{R} \sum_{y=1}^{C} |M(x,y,t)| \cdot (C \cdot R)^{-1} \tag{a.4}$$

$$\hat{PM}(t) = (PM(t), \hat{PM}(t-1), \hat{PM}(t-2)) \cdot \overrightarrow{\theta} \tag{a.5}$$

$$\omega_1(t) = max(\omega_3, \frac{\hat{PM}(t)}{T_{PM}}),\ \omega_2(t) = max(\omega_4, \frac{\hat{PM}(t)}{T_{PM}}) \tag{a.6}$$

LGMD2—Lamina layer

$$P_{on}(x,y,t) = [P(x,y,t)]^+ + \beta \cdot P_{on}(x,y,t-1) \tag{a.7}$$

$$P_{off}(x,y,t) = -[P(x,y,t)]^- + \beta \cdot P_{off}(x,y,t-1) \tag{a.8}$$

LGMD2—Medulla layer

$$E_{on}(x,y,t) = \iint P_{on}(u,v,t) W_1(x-u, y-v)\, du\, dv \tag{a.9}$$

$$D_{on}(x,y,t) = (E_{on}(x,y,t), D_{on}(x,y,t-1), D_{on}(x,y,t-2)) \cdot \overrightarrow{\alpha_{on}} \tag{a.10}$$

$$I_{on}(x,y,t) = \iint D_{on}(u,v,t) W_{I_{on}}(x-u, y-v)\, du\, dv \tag{a.11}$$

$$W_1 = \frac{1}{8}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 8 & 2 \\ 1 & 2 & 1 \end{bmatrix} \tag{a.12}$$

$$E_{off}(x,y,t) = \iint P_{off}(u,v,t) W_1(x-u, y-v)\, du\, dv \tag{a.13}$$

$$D_{off}(x,y,t) = (E_{off}(x,y,t), D_{off}(x,y,t-1), D_{off}(x,y,t-2)) \cdot \overrightarrow{\alpha_{off}} \tag{a.14}$$

$$I_{off}(x,y,t) = \iint D_{off}(u,v,t) W_{I_{off}}(x-u, y-v)\, du\, dv \tag{a.15}$$

$$W_{I_{off}} = \frac{1}{32}\begin{bmatrix} 1 & 2 & 4 & 2 & 1 \\ 2 & 4 & 8 & 4 & 2 \\ 4 & 8 & 16 & 8 & 4 \\ 1 & 2 & 4 & 2 & 1 \\ 2 & 4 & 8 & 4 & 2 \end{bmatrix},\ W_{I_{on}} = 2 * W_{I_{off}} \tag{a.16}$$

$$S_{on}(x,y,t) = [E_{on}(x,y,t) - \omega_{on}(t) \cdot I_{on}(x,y,t)]^+ \tag{a.17}$$

$$S_{off}(x,y,t) = [E_{off}(x,y,t) - \omega_{off}(t) \cdot I_{off}(x,y,t)]^+ \tag{a.18}$$

$$\omega_{on}(t) = \omega_1(t),\ \omega_{off}(t) = \omega_2(t) \tag{a.19}$$

$$S(x,y,t) = S_{on}(x,y,t) + S_{off}(x,y,t) \tag{a.20}$$

LGMD1—Retina layer

$$TD(x,y,t) = S(x,y,t) - S(x,y,t-1) \tag{a.21}$$

LGMD1—Excitation Layer

$$\hat{P_{on}}(x,y,t) = [TD(x,y,t)]^{+} + \beta \cdot \hat{P_{on}}(x,y,t-1) \tag{a.22}$$

$$\hat{P_{off}}(x,y,t) = -[TD(x,y,t)]^{-} + \beta \cdot \hat{P_{off}}(x,y,t-1) \tag{a.23}$$

$$\hat{E_{on}}(x,y,t) = \iint TD(u,v,t) W_1(x-u, y-v)\, du\, dv \tag{a.24}$$

$$\hat{E_{off}}(x,y,t) = \iint TD(u,v,t) W_1(x-u, y-v)\, du\, dv \tag{a.25}$$

$$\hat{\phi}(x,y,t) = \hat{E_{on}}(x,y,t) + \hat{E_{off}}(x,y,t) \tag{a.26}$$

LGMD1—Neuronal integration and activation

$$k(t) = \sum_{x=1}^{R} \sum_{y=1}^{C} \hat{\phi}(x,y,t),\ K(t) = \left(1 + e^{-k(t)\cdot(C\cdot R\cdot \alpha_2)^{-1}}\right)^{-1} \tag{a.27}$$

$$\hat{K}(t) = \begin{cases} \alpha_3 \left(\hat{K}(t-1) + K(t) - K(t-1)\right), & \text{if } (K(t) - K(t-1)) \le T_{sfa} \\ \alpha_3 K(t), & \text{otherwise} \end{cases} \tag{a.28}$$

$$\alpha_3 = \frac{\tau_s}{\tau_s + \tau_{in}} \tag{a.29}$$

Spiking frequency

$$Spi(t) = \left[e^{(\alpha_4 \cdot (\hat{K}(t) - T_{sp}))}\right] \tag{a.30}$$

$$Col(t) = \begin{cases} True, & \text{if } \left(\sum_{i=t-n_t}^{t} Spi(i)\right) \times 1000/(n_t \cdot \tau_{in}) \ge T_c \\ False, & \text{otherwise} \end{cases} \tag{a.31}$$

REFERENCES

Amer, M. (2019). *Modularity in artificial neural networks*. PhD thesis, The University of Nottingham.

Amer, M. and Maul, T. (2019). A review of modularization techniques in artificial neural networks. *Artificial Intelligence Review*, 52:527–561.

Bermúdez i Badia, S., Bernardet, U., and Verschure, P. F. (2010). Non-linear neuronal responses as an emergent property of afferent networks: a case study of the locust lobula giant movement detector. *PLoS Computational Biology*, 6(3):e1000701.

Bertolero, M. A., Yeo, B. T., and D'Esposito, M. (2015). The modular and integrative functional architecture of the human brain. *Proceedings of the National Academy of Sciences*, 112(49):E6798–E6807.

Bertrand, O. J., Lindemann, J. P., and Egelhaaf, M. (2015). A bio-inspired collision avoidance model based on spatial information derived from motion detectors leads to common routes. *PLoS Computational Biology*, 11(11):e1004339.

Borst, A. and Euler, T. (2011). Seeing things in motion: models, circuits, and mechanisms. *Neuron*, 71(6):974–994.

Clune, J., Mouret, J.-B., and Lipson, H. (2013). The evolutionary origins of modularity. *Proceedings of the Royal Society b: Biological sciences*, 280(1755):20122863.

Franceschini, N. (2014). Small brains, smart machines: from fly vision to robot vision and back again. *Proceedings of the IEEE*, 102(5):751–781.

Fu, Q. (2023). Motion perception based on on/off channels: A survey. *Neural Networks*, 165:1–18.

Fu, Q., Hu, C., Liu, P., and Yue, S. (2018a). Towards computational models of insect motion detectors for robot vision. In *Annual Conference Towards Autonomous Robotic Systems*, pages 465–467. Springer.

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 Transactions on Cybernetics*, 50(12):5074–5088.

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

Fu, Q., Li, Z., and Peng, J. (2023). Harmonizing motion and contrast vision for robust looming detection. *Array*, 17:100272.

Fu, Q., Sun, X., Liu, T., Hu, C., and Yue, S. (2021). Robustness of bio-inspired visual systems for collision prediction in critical robot traffic. *Frontiers in Robotics and AI*, 8:529872.

Fu, Q., Wang, H., Hu, C., and Yue, S. (2019). Towards computational models and applications of insect visual systems for motion perception: A review. *Artificial Life*, 25(3):263–311.

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.

Gabbiani, F., Krapp, H., Koch, C., and Laurent, G. (2002). Multiplicative computation by a looming-sensitive neuron. In *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society][Engineering in Medicine and Biology*, volume 3, pages 1968–1969. IEEE.

Gabbiani, F., Krapp, H. G., Hatsopoulos, N., Mo, C.-H., Koch, C., and Laurent, G. (2004). Multiplication and stimulus invariance in a looming-sensitive neuron. *Journal of Physiology-Paris*, 98(1-3):19–34.

Geurten, B. R., Nordstrom, K., Sprayberry, J. D., Bolzon, D. M., and O'Carroll, D. C. (2007). Neural mechanisms underlying target detection in a dragonfly centrifugal neuron. *Journal of Experimental Biology*, 210(18):3277–3284.

Green, W. E. and Oh, P. Y. (2008). Optic-flow-based collision avoidance. *IEEE Robotics & Automation Magazine*, 15(1):96–103.

Green, W. E., Oh, P. Y., and Barrows, G. (2004). Flying insect inspired vision for autonomous aerial robot maneuvers in near-earth environments. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, volume 3, pages 2347–2352. IEEE.

Hu, C., Fu, Q., and Yue, S. (2018). Colias iv: The affordable micro robot platform with bio-inspired vision. In *Annual Conference Towards Autonomous Robotic Systems*, pages 197–208. Springer.

Hua, M., Fu, Q., Peng, J., Yue, S., and Luan, H. (2022). Shaping the ultra-selectivity of a looming detection neural network from non-linear correlation of radial motion. In *IEEE The International Joint Conference on Neural Networks*.

Keil, M. (2011). Emergence of multiplication in a biophysical model of a wide-field visual neuron for computing object approaches: Dynamics, peaks, & fits. *Advances in Neural Information Processing Systems*, 24.

Keil, M. S., Roca-Moreno, E., and Rodriguez-Vazquez, A. (2004). A neural model of the locust visual system for detection of object approaches with real-world scenes. In *Proceedings of the fourth IASTED international conference on visualization, imaging, and image processing*, pages 340–345. IASTED.

Kelkar, A. and Medaglia, J. D. (2021). Evidence of brain modularity. In *Encyclopedia of evolutionary psychological science*, pages 2432–2441. Springer.

Klapoetke, N. C., Nern, A., Peek, M. Y., Rogers, E. M., Breads, P., Rubin, G. M., Reiser, M. B., and Card, G. M. (2017a). Ultra-selective looming detection from radial motion opponency. *Nature*, 551(7679):237–241.

Klapoetke, N. C., Nern, A., Peek, M. Y., Rogers, E. M., Breads, P., Rubin, G. M., Reiser, M. B., and Card, G. M. (2017b). Ultra-selective looming detection from radial motion opponency. *Nature*, 551:237–241.

Krajník, T., Nitsche, M., Faigl, J., Vaněk, P., Saska, M., Přeučil, L., Duckett, T., and Mejail, M. (2014). A practical multirobot localization system. *Journal of Intelligent & Robotic Systems*, 76:539–562.

Li, J., Sun, X., Li, H., Peng, J., and Fu, Q. (2023). On the ensemble of collision perception neuron models towards ultra-selectivity. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Liu, R. and Fu, Q. (2025). Attention-driven LPLC2 neural ensemble model for multi-target looming detection and localization. In *IEEE The International Joint Conference on Neural Networks*.

Liu, T., Sun, X., Hu, C., Fu, Q., and Yue, S. (2021a). A multiple pheromone communication system for swarm intelligence. *IEEE Access*, 9:148721–148737.

Liu, T., Sun, X., Hu, C., Fu, Q., and Yue, S. (2021b). A versatile vision-pheromone-communication platform for swarm robotics. In *IEEE International Conference on Robotics and Automation (ICRA)*.

Meunier, D., Lambiotte, R., and Bullmore, E. T. (2010). Modular and hierarchically modular organization of brain networks. *Frontiers in neuroscience*, 4:200.

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 *2015 International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP)*, pages 1–7. IEEE.

Muijres, F. T., Elzinga, M. J., Melis, J. M., and Dickinson, M. H. (2014). Flies evade looming targets by executing rapid visually directed banked turns. *Science*, 344(6180):172–177.

Qin, Z., Fu, Q., and Peng, J. (2024). A computational efficient and robust looming perception model based on dynamic neural field. *Neural Networks*, 179:106502.

Rind, F. C. (2002). Motion detectors in the locust visual system: from biology to robot sensors. *Microscopy Research and Technique*, 56(4):256–269.

Rind, F. C. and Bramwell, D. (1996). Neural network based on the input organization of an identified neuron signaling impending collision. *Journal of Neurophysiology*, 75(3):967–985.

Rind, F. C. and Simmons, P. J. (1998). Local circuit for the computation of object approach by an identified visual neuron in the locust. *Journal of Comparative Neurology*, 395(3):405–415.

Rind, F. C., Wernitznig, S., Pölt, P., Zankel, A., Gütl, D., Sztarker, J., and Leitinger, G. (2016). Two identified looming detectors in the locust: ubiquitous lateral connections among their inputs contribute to selective responses to looming objects. *Scientific Reports*, 6(1):35525.

Salt, L., Howard, D., Indiveri, G., and Sandamirskaya, Y. (2019). Parameter optimization and learning in a spiking neural network for uav obstacle avoidance targeting neuromorphic processors. *IEEE Transactions on Neural Networks and Learning Systems*, 31(9):3305–3318.

Serres, J. R. and Ruffier, F. (2017). Optic flow-based collision-free strategies: From insects to robots. *Arthropod Structure & Development*, 46(5):703–717.

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. *Journal of Comparative Physiology A*, 180:203–214.

Troyer, T. W., Krukowski, A. E., Priebe, N. J., and Miller, K. D. (1998). Contrast-invariant orientation tuning in cat visual cortex: thalamocortical input tuning and correlation-based intracortical connectivity. *Journal of Neuroscience*, 18(15):5908–5927.

Wang, H., Fu, Q., Wang, H., Baxter, P., Peng, J., and Yue, S. (2021). A bioinspired angular velocity decoding neural network model for visually guided flights. *Neural Networks*, 136:180–193.

Wang, H., Fu, Q., Wang, H., Peng, J., and Yue, S. (2019). Constant angular velocity regulation for visually guided terrain following. In *Artificial Intelligence Applications and Innovations*, pages 597–608. Springer International Publishing.

Wang, M., Huang, J., Sun, X., Hu, C., Peng, J., and Fu, Q. (2024). A composite neuronal model as miniaturized visual modality for collision perception. In *International Conference on Robot Intelligence Technology and Applications*. Springer.

Yue, S. and Rind, F. C. (2006). Collision detection in complex dynamic scenes using an LGMD-based visual neural network with feature enhancement. *IEEE Transactions on Neural Networks*, 17(3):705–716.

Zhao, J., Xi, S., Li, Y., Guo, A., and Wu, Z. (2023). A fly inspired solution to looming detection for collision avoidance. *iScience*, 26:106337.

Zhao, J., Xie, Q., Shuang, F., and Yue, S. (2024). An angular acceleration based looming detector for moving uavs. *Biomimetics*, 9(1):22.

Zhou, B., Li, Z., Kim, S., Lafferty, J., and Clark, D. A. (2022). Shallow neural networks trained to detect collisions recover features of visual loom-selective neurons. *eLife*, 11:e72067.