Enhancing Collision-Selectivity in Autonomous Micro-Robots by Elevated Temporal Derivatives in Neuronal Assembly Framework

Journal:	*IEEE Transactions on Cognitive and Developmental Systems*
Manuscript ID	TCDS-2025-0806
Manuscript Type:	Regular
Keywords:	biologically inspired feature extraction, dynamics in neural systems, neuronal plasticity, using robots to study development and learning

SCHOLARONE™
Manuscripts

Enhancing Collision-Selectivity in Autonomous Micro-Robots by Elevated Temporal Derivatives in Neuronal Assembly Framework

Anonymous Authors

***Abstract*—Collision selectivity, the ability to distinguish approaching motion from other movement patterns, is fundamental for safe navigation in autonomous micro-robots operating under strict computational and energy constraints. Inspired by biological research, the single-neuron computation of the lobula giant movement detectors (LGMD) in the locust visual system has led to a range of bio-inspired vision models, many of which are particularly attractive for micro-robots due to their low computational cost and rapid response. However, a key limitation persists: most existing robotic LGMD models rely on isolated single-neuron implementations and exhibit limited or task-specific approaching selectivity compared with their biological counterparts. To address this challenge, this paper introduces a neuronal assembly framework in which LGMD1 and LGMD2 modules are cascaded to improve collision selectivity in micro-robotic vision systems. Through systematic and comparative analysis, we identify elevated temporal derivative processing in a second-stage module as the key mechanism underlying improved discrimination of approaching stimuli, rendering additional neural layers largely redundant. Leveraging this insight, we derive a simplified and computationally economical model emphasizing temporal derivative computation within visual streams. The proposed approach consistently outperforms existing bio-inspired methods in distinguishing approaching motion from translating and receding patterns. Real-world micro-robot experiments demonstrate reliable collision avoidance at the success rate over 96% in dynamic environments. These results highlight the effectiveness of temporal-derivative-based neural processing for developing low-power, biologically grounded visual cognition systems in autonomous micro-robots.**

***Index Terms*—Collision selectivity, Micro-robot vision, LGMD, Neuronal assembly framework, Elevated temporal derivatives**

I. 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 approaching motion—a phenomenon termed "*collision 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 [1]–[4]. Among these biological models, flies and locusts are particularly noteworthy examples of animals studied extensively for their robust looming perception capabilities [5]–[9]. Their visual system mechanisms have inspired and been successfully integrated into artificial vision systems, including robotic platforms for ground and aerial navigation [1], [2], [10].

More specifically, numerous optical flow (OF)-based perception methods, motivated by the visual systems of flies, have ramified, particularly drawing on insights from lobula plate tangential cells (LPTCs) [2]. These OF-strategies rely on local pixel-wise motion cues to guide diverse navigational behaviors in flies, including tunnel crossing [11], terrain following [12], collision avoidance [6], 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 [13]–[15]. 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 [16]–[18]. These neurons exhibit exceptional selectivity, responding preferentially to visual centroid-emanated centrifugal motion patterns [19]. 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 [20].

Another type of collision 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 [21]–[25]. 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 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 LGMD's collision selectivity [25]–[34], the LGMD-based single-neuron models have been investigated extensively with many successful application in robotic vision and navigation [3], [4], [10], [35]. Compared to fly-inspired optical flow (OF)-based approaches, LGMD models exhibit a distinct preference for detecting directly approaching targets, relying on global

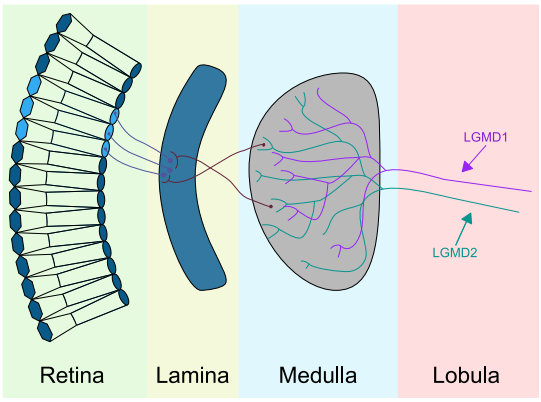

Fig. 1: Schematic illustration of presynaptic neuropils of LGMD in either compound eye of locust: each LGMD visual pathway adopts a four-layer structure: retina, lamina, medulla and lobula.

rather than local motion cues. Furthermore, unlike LPLC2-based models, LGMD systems show greater flexibility in detecting approaching motion initiated from various spatial positions within the visual field.

Although existing neural models have demonstrated robustness and successfully addressed certain real-world collision detection tasks, achieving consistent collision selectivity remains challenging–particularly for robotic vision systems operating in dynamic physical environments under strict computational constraints. This limitation underscores the inherent restrictions of single-neuron-based computational architectures.

To overcome these deficiencies, researchers have explored the modularization of neural models or networks to accomplish complex perceptual tasks [36]–[41]. 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 [29] and LGMD2 [32] neural network structures. Their proposed modular approach, presented in two variations, i.e., LGMD1-LGMD2 and LGMD2-LGMD1, significantly improved selectivity by effectively reducing interference from translating and receding stimuli [42]. 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.

To enable deployment of the composite model in embedded robotic vision systems, Wang et al. optimized both the temporal and spatial complexity of the LGMD composite framework and successfully implemented it on STM32-based ground micro-robots with severely constrained computational resources [43]. In robotic implementations, the proposed composite model significantly enhances collision selectivity, enabling robots to respond exclusively to approaching targets and achieve collision-free navigation within an arena. Specifically, the two neuronal modules mutually reinforce each other, with the preliminary 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 selectivity to approaching motion (2) Is it necessary for this second module to retain the entire four-layer network structure, or could it be simplified without compromising performance?

To identify the core structural component responsible for enhanced collision selectivity, we conducted offline redundancy analyses focusing on the second module of the LGMD2–LGMD1 neuronal assembly. The results demonstrate that the introduction of an additional temporal-derivative operation is the key mechanism underlying the observed improvement in selectivity. Based on this insight, we derived a more economical neuronal assembly model that emphasizes elevated temporal-derivative processing within the cascaded module. The proposed framework was systematically compared with state-of-the-art approaches and exhibited superior collision selectivity across a range of real-world visual scenarios. Furthermore, online comparative experiments were performed using micro-robots, enabling visualization of internal layer representations in the embedded neuronal model and benchmarking its performance against related bio-inspired visual cognition systems.

The rest contents are structured as follows. Section II formulates the proposed computational framework. Experimental evaluations are presented in Section III. Section IV offers further discussions. Section V summarizes this research.

II. Formulation of the Computational Framework

This time derivative model, also called LGMD2-Derivative, consists of two modules: the first module includes the retina, lamina and medulla computational layers in the LGMD2 module, and the second one only consists of the retina computational layer and the final LGMD1 cell for output, rendering the intermediate layer processing redundant. The schematic illustration of the proposed neuronal assembly framework is illustrated in Fig. 2. Specifically, the first retina layer is used to extract changes of light intensities with respect to time using a differential operator. 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 [44] and dragonflies [45], and vertebrates like rabbits [44] and cats [46]. The medulla layer, including excitation units, inhibition units, and summation units, is a key component for extracting edge motion information. 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 presynaptic field and activates membrane potentials which are eventually transformed to frequency domain using a spiking mechanism in order to indicate potential collision threats towards robotic motor systems.

A. The First Temporal Differential Operator

The LGMD2's retina is the first computational layer of the proposed time-derivative model and consists of visual recep-

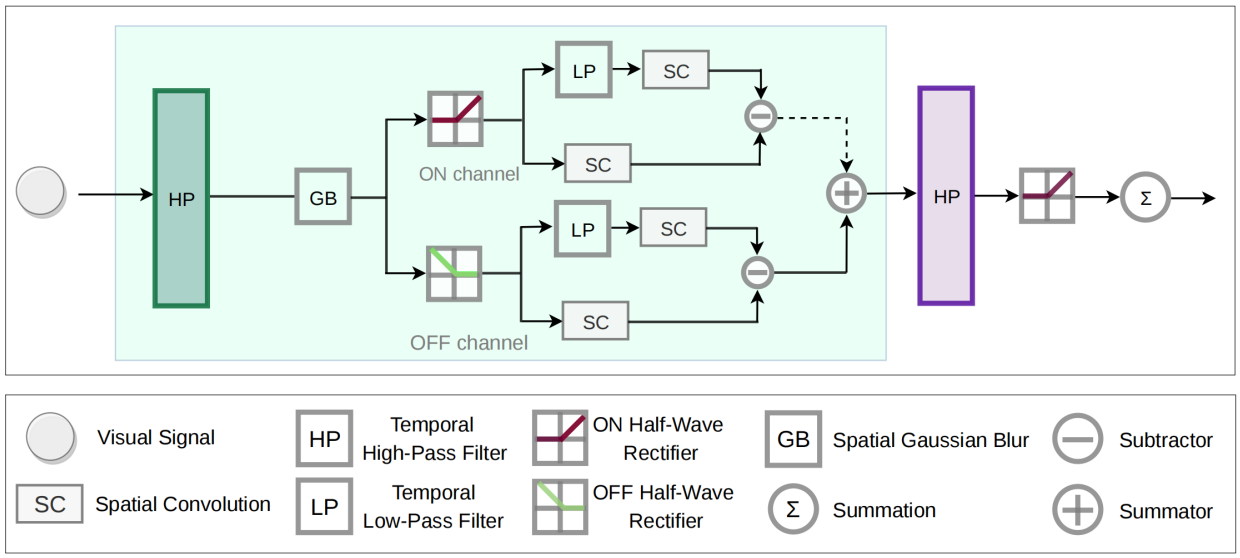

Fig. 2: Schematic diagram of the proposed computational framework. The first neuronal module (light blue box) represents the LGMD2 neuronal system. Input visual signals first pass through a Temporal High-Pass Filter (HP) to capture brightness changes in the visual field. A subsequent Spatial Gaussian Blur (GB) suppresses irrelevant noise. The signals are then transmitted to the ON and OFF channels, with the dashed arrow indicating bias of the ON channel of LGMD2. The output of this module feeds into the retina layer of the second module (LGMD1), where a further temporal derivative is applied (functionally another HP), allowing only positive signals to pass. Finally, the Summation stage integrates the remaining signals to generate the model's output. For simplicity, only one single neuronal processing pipeline is shown.

tors arranged in a matrix. Mathematically, it applies a temporal differential operator to extract changes of light intensities between consecutive image frames, as the following:

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

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

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

τ_1 and Δ_t are time delay constant and the time interval between two discrete frames, respectively.

Subsequently, a Gaussian blur with kernel $[G_p]$ is applied for blurring and mimicking insect compound eyes, 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 receptors in the visual field Ω are activated, such as in scenes of robot turning around, or at the end of colliding, the layer would exert strong inhibition across wide field, directly upon LGMD neurons. To simulate this, a photoreceptor mediation (PM) is introduced to adjust the time-varying biases within ON/OFF pathways. The mathematical expression is defined as

$$PM(t) = \Big(\iint_{\Omega} |M(u,v,t)|\, du\, dv \Big) \cdot \Omega^{-1}. \tag{4}$$

A temporal convolution is applied as the low-pass filter to delay the PM signal, that is

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

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

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

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

B. Signal Bifurcation and Competition

The computational 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 mechanisms as

$$\begin{aligned} P_{on}(x,y,t) &= \big[P(x,y,t)\big]^+ + \beta \cdot P_{on}(x,y,t-\Delta_t), \\ P_{off}(x,y,t) &= -\big[P(x,y,t)\big]^- + \beta \cdot P_{off}(x,y,t-\Delta_t), \end{aligned} \tag{6}$$

where the functions $[x]^+$ and $[x]^-$ denote $\max(0,x)$ and $\min(x,0)$, respectively.

After that, the computational 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

$$\begin{aligned} E_{on}(x,y,t) &= \iint P_{on}(u,v,t) W_e(x-u, y-v)\, du\, dv, \\ D_{on}(x,y,t) &= (E_{on}(x,y,t), D_{on}(x,y,t-\Delta_t), \ldots \\ &\quad D_{on}(x,y,t-2\cdot\Delta_t)) \cdot \overrightarrow{\alpha_{on}}, \\ I_{on}(x,y,t) &= \iint D_{on}(u,v,t) W_{I_{on}}(x-u, y-v)\, du\, dv, \end{aligned} \tag{7}$$

where $[W_e]$ stands for a convolution kernel of excitatory signals, in the following discrete form:

$$W_e = \frac{1}{8}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 8 & 2 \\ 1 & 2 & 1 \end{bmatrix},$$

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 neural computations are as the following:

$$\begin{aligned} E_{off}(x,y,t) &= \iint P_{off}(u,v,t) W_e(x-u, y-v)\, du\, dv, \\ D_{off}(x,y,t) &= (E_{off}(x,y,t), D_{off}(x,y,t-\Delta_t), ... \\ & D_{off}(x,y,t-2\cdot\Delta_t)) \cdot \overrightarrow{\alpha_{off}}, \\ I_{off}(x,y,t) &= \iint D_{off}(u,v,t) W_{I_{off}}(x-u, y-v)\, du\, dv. \end{aligned} \tag{8}$$

$W_{I_{off}}$ and $W_{I_{on}}$ can be implemented as the following discrete forms:

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

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

$$\begin{aligned} S_{on}(x,y,t) &= \left[E_{on}(x,y,t) - \omega_1(t)\cdot I_{on}(x,y,t)\right]^+, \\ S_{off}(x,y,t) &= \left[E_{off}(x,y,t) - \omega_2(t)\cdot I_{off}(x,y,t)\right]^+. \end{aligned} \tag{9}$$

Linear summation between the ON/OFF channels is conducted to form the output feature representation of the first LGMD2 module, that is

$$S(x,y,t) = S_{on}(x,y,t) + S_{off}(x,y,t). \tag{10}$$

C. The Second Temporal Differential Operator

The second LGMD1's computational retina layer plays a central role in this neuronal assembly framework which works as the entry of second cascaded module, elevates the time derivatives from the feature representation of first module, i.e., $[S]$. This works essentially as another differential operator as

$$TD(x,y,t) = S(x,y,t) - S(x,y,t-\Delta_t). \tag{11}$$

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

$$\Phi(x,y,t) = \left[TD(x,y,t)\right]^+ + \beta\cdot\Phi(x,y,t-\Delta_t). \tag{12}$$

D. Neuronal Integration and Activation

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

$$K(t) = \iint_{\Omega} \Phi(u,v,t)\, du\, dv, \;\; \xi(t) = \left(1 + e^{-K(t)\cdot(|\Omega|\cdot\alpha_2)^{-1}}\right)^{-1}, \tag{13}$$

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

TABLE I: Setting the Parameters

Eq.	Parameters	Description
(2)	$\tau_1 = 100$(ms)	time delay constant
(2)	$\Delta_t \in 15 \sim 50$(ms)	time interval between frames
(6)	$\{\omega_3 = 0.6, \omega_4 = 0.3\}$	local inhibition bias baseline
(6)	$T_{pm} \in 10 \sim 50$	Divisive threshold
(12)	$\beta = 0.1$	decay coefficient
(15)	$\tau_s \in 500 \sim 1000$(ms)	time delay constant in SFA
(15)	$\alpha_4 = 4$	scale parameter of firing rate
(15)	$T_{sp} = 0.7$	spiking threshold
(16)	$n_t = 10$(frames)	time window
(16)	$T_c \in 15 \sim 20$(Hz)	collision warning threshold

In order to suppress the effect of irrelevant motion stimuli such as translating and receding objects on the model to further sculpt its selectivity, the spike frequency adaptation mechanism (SFA) [29] is incorporated to inspect the first-order temporal derivative of potentials, that is,

$$\hat{\xi}(t) = \begin{cases} \alpha_3\cdot\xi(t), \text{if } \left(\xi(t) - \xi(t-\Delta_t)\right) > 0 \\ \alpha_3\cdot\left(\hat{\xi}(t-\Delta_t) + \xi(t) - \xi(t-\Delta_t)\right), \text{otherwise} \end{cases}, \tag{14}$$

where α_3 indicates the adaptation rate, calculated by a time constant τ_s as

$$\alpha_3 = \frac{\tau_s}{\tau_s + \Delta_t}.$$

E. Frequency Domain Output

The potential $\hat{\xi}(t)$ herein is transformed to frequency domain using an exponential functions as

$$S_{pi}(t) = \left[\mathrm{e}^{(\alpha_4\cdot(\hat{\xi}(t) - T_{sp}))}\right], \tag{15}$$

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

$$C_{ol}(t) = \begin{cases} \text{True}, & \text{if } \left(\sum_{i=t-n_t}^{t} S_{pi}(i)\right) \times 1000/\left(n_t\cdot\Delta_t\right) \geq T_c \\ \text{False}, & \text{otherwise} \end{cases}. \tag{16}$$

This discrete form of computation indicates the implementation in micro-robots for initiating urgent avoidance behaviors versus collision danger. n_t denotes the specified time window to update spike frequency, and T_c is the warning threshold of collision danger.

F. Setting the Parameters

Table I summarizes the parameter settings used in the proposed time-derivative model. Certain parameters depend directly on the physical characteristics of the input videos, specifically, the spatial sampling of visual field Ω correspond to the physical resolution of image streams, while Δ_t is determined by the video frame rate. The remaining parameters were empirically selected based on prior related studies [42], [43], balancing optimal model performance in both computational simulations and micro-robotic implementations.

Driven by biological principles, visual neural processing of the proposed neuronal assembly framework is sensible and explainable. Given the limited number of adjustable parameters, explicit learning or adaptive mechanisms were not

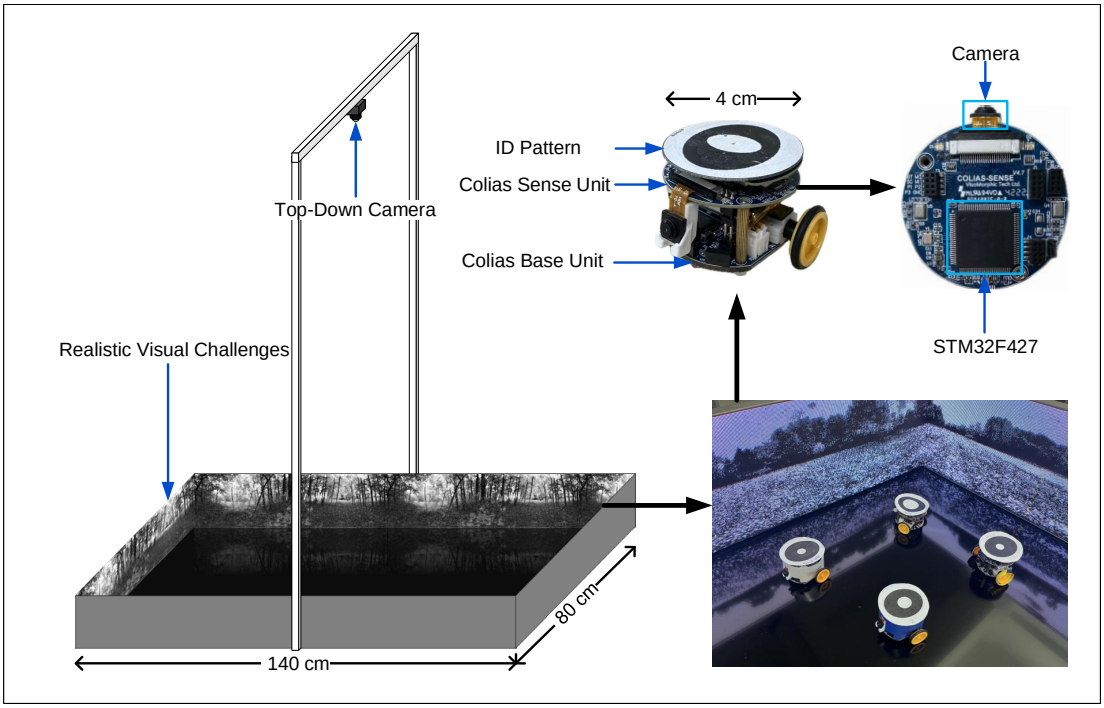

Fig. 3: Illustration of the experimental robot arena and the structure of grounded micro-robot *Colias*. The surrounding LED walls can display different scenes to challenge visual systems. The top-down camera was utilized to localize the robots with ID-specific pattern on each top, with respect to time.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

incorporated into this model-based approach. However, future work could explore integrating automatic parameter searching techniques to further enhance the model's adaptability and robustness across varying environments or sensor configurations.

III. Experimental Evaluation

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* [47] 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 [42], the LGMD1 model [29], the LGMD2 model [32], the hybrid LGMD model [48], and the LGMD2-Excitation model with inhibition of second module ablated from [43] (algorithms in the Appendix). These experiments aimed to explore the key component in neuronal assemblies which essentially boosts approaching motion selectivity, and verify its value in robotic real-time implementation.

A. Experimental Setting

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 [49], [50], 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

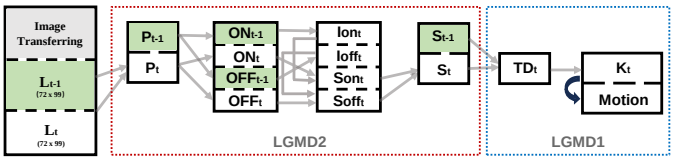

Fig. 4: Illustration of the onboard image-processing strategy implemented on the micro-robot with 256 KB SRAM. Three images are stored in memory at any time: one serves as a buffer, while the remaining two are used for visual neural processing within the proposed neuronal assembly framework. The resulting output signal is then transmitted to the motor control system to guide robot behavior.

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.

2) Online tests setting: In the online experiments, the proposed approach was implemented as an embedded vision module on the ground-based micro-robot *Colias* [47]. *Colias* was selected as the experimental platform because it is a lightweight and cost-effective micro-robot, measuring approximately 4 cm in diameter and weighing about 50 g, while being equipped with an onboard vision module. Its small form factor, rapid response, and precise motion control make it a well-suited testbed for insect-inspired visual intelligence research, particularly under stringent computational and energy constraints.

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,

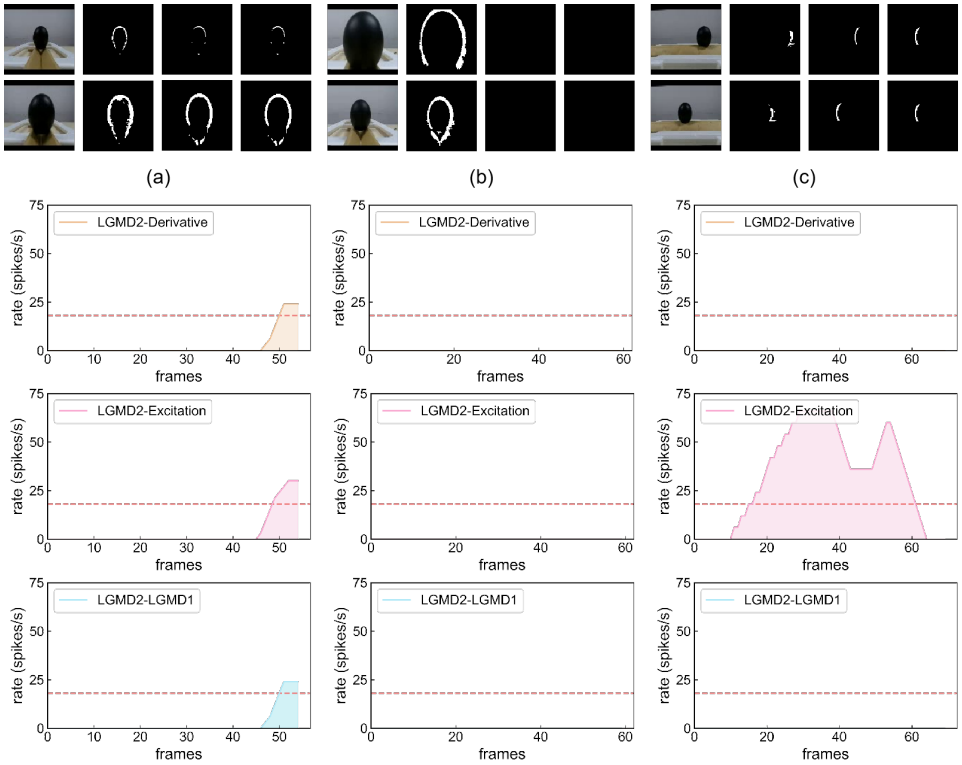

Fig. 5: Offline testing results of ablation studies in indoor scenes: LGMD2-Derivative indicates the proposed model with just raised time derivative. LGMD2-Excitation indicates the composite model [43] with inhibition ablated in the second LGMD1 module. LGMD2-LGMD1 model is the intact composite model [42]. 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 LGMD2-Derivative model and full composite LGMD2-LGMD1 model respond only to approaching motion as the excitation retains after raising the time derivative at the second module.

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. Owing to the robot's limited memory capacity, we designed a circular queue to efficiently transfer and stream image data to the proposed visual processing framework in real time, as illustrated in Fig. 4.

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/natural 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 [51] to track trajectories and calculate the collision avoidance success rate.

B. Metrics

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

$$\begin{bmatrix} \mathrm{TN} & \mathrm{FP} \\ \mathrm{FN} & \mathrm{TP} \end{bmatrix}, \tag{17}$$

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

$$F_1 = 2 \cdot \frac{\mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}. \tag{18}$$

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

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{19}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{20}$$

C. Ablation Studies on Module Redundancy

Firstly, we conducted redundancy experiments to investigate the key component or mechanism that improves the collision selectivity of LGMD-based visual systems. We compared the response and selectivity of the proposed LGMD2-Derivative

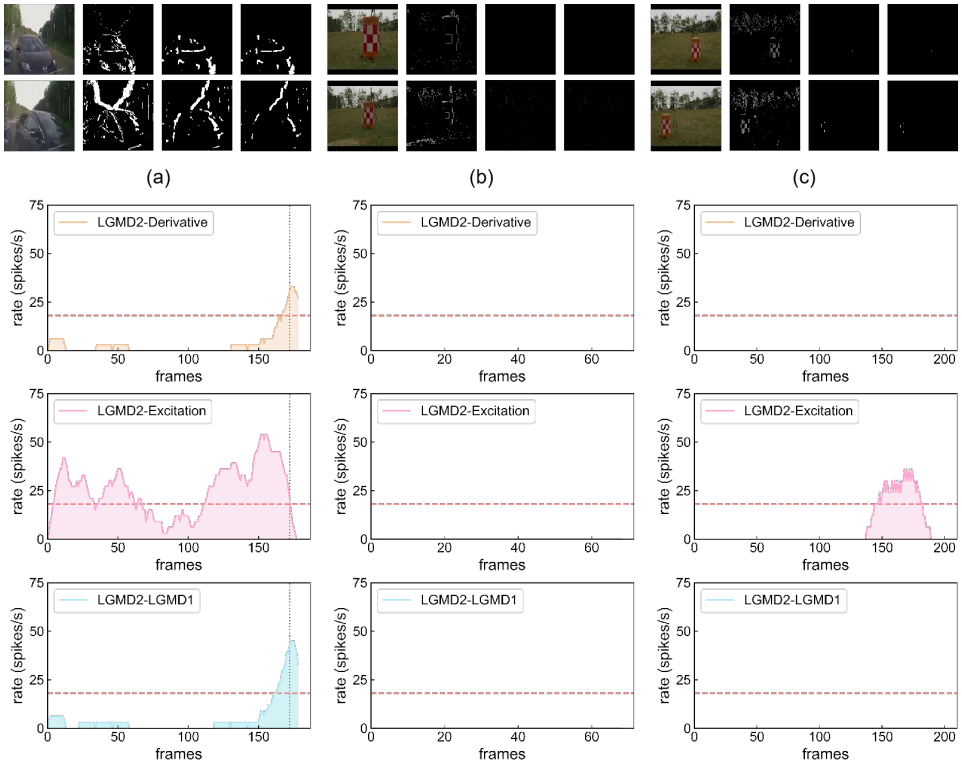

Fig. 6: Offline testing results of ablation studies challenged by outdoor recordings from UAV and car (a) crash, (b) recession, (c) shifting views. The vertical dashed line indicates the moment of ground-truth crash at frame 172. Other notations conform to Fig. 5. The proposed LGMD2-Derivative model and full composite LGMD2-LGMD1 model work effectively to extract only approaching motion in complex dynamic scenes.

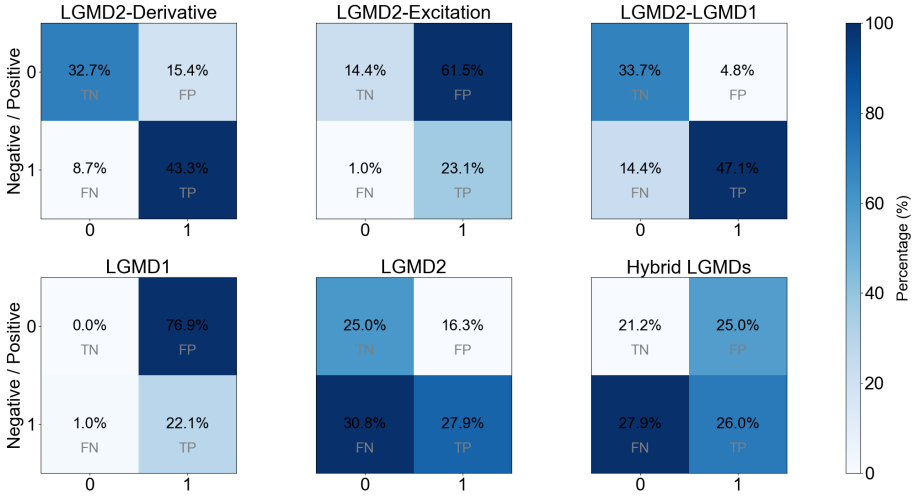

Fig. 7: Confusion matrices for six comparative models on real physical datasets. Each matrix shows TP, FP, TN and FN. The proposed LGMD2-Derivative model and the full LGMD2-LGMD1 composite model have higher rates of TP and TN challenged by car and UAV scenarios, which demonstrate enhanced collision selectivity with raised time derivative.

TABLE II: F_1 score (%) of models to 104 real-world sets

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

model with the LGMD2-Excitation, with inhibition processing ablated from a robotic composite model [43], and the original LGMD2-LGMD1 model [42].

As shown in Fig. 5 and Fig. 6, 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 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 approaching or colliding stimuli. The statistical outcomes–

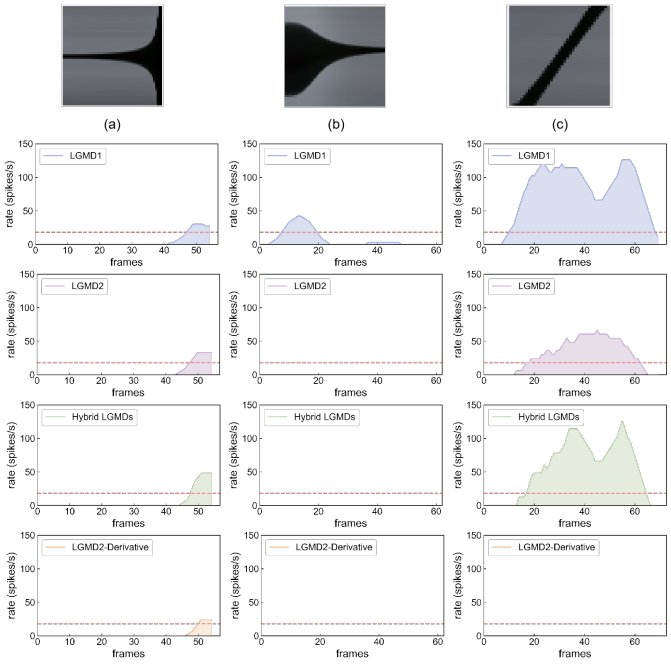

Fig. 8: Comparative model responses against indoor visual movements, captured from a fixed camera: The stimuli on top are represented by intensity change of image center with respect to time. (a) dark ball (black pixels) approaching over time, (b) dark ball receding, (c) dark ball translating. Horizontal dashed line denotes threshold for collision warning. The proposed LGMD2-Derivative model only responds to approaching motion with enhanced selectivity.

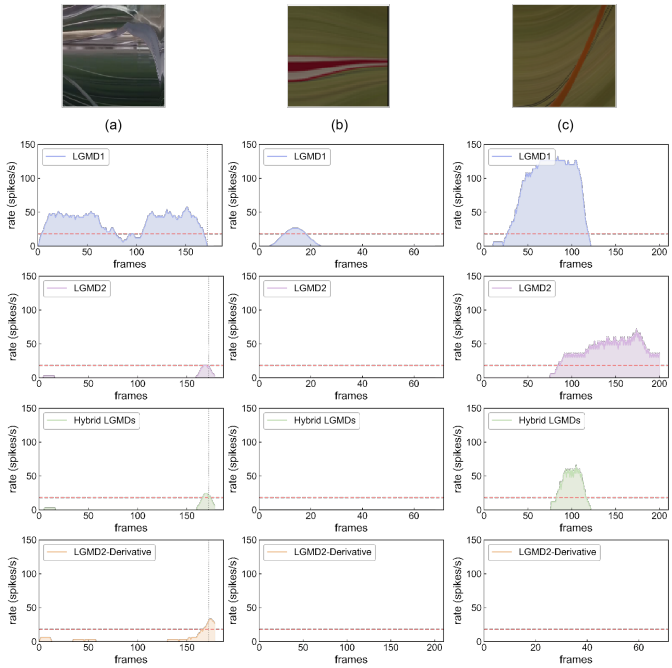

Fig. 9: Comparative model responses against complex outdoor visual movements, captured from recordings by ground vehicle and UAV: (a) ground vehicle (gray pixels) approaching over time, (b) UAV recession from a cylindrical obstacle (marked with red-and-white pixels), (c) UAV shifting in front of the cylindrical obstacle. Vertical line indicates ground truth crash moment. The proposed LGMD2-Derivative model only responds to approaching motion.

including confusion matrices, precision, recall, and F_1 scores–are presented in Fig. 7 and Table II. The results show that the proposed model's F_1 score is only 4.79% lower than that of LGMD2-LGMD1, suggesting a marginal decrease in collision detection performance, though the difference is not statistically significant. In contrast, the LGMD2-Excitation model underperforms substantially, with its F_1 score 35.79% and 40.58% lower than those of LGMD2-Derivative and LGMD2-LGMD1, respectively–indicating poor collision recognition capability. Moreover, feature visualizations in Fig. 5 and Fig. 6 reveal that as 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 ablation studies demonstrate that strong approaching motion 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. The results highlight that elevating the temporal derivative at the output of the first neuronal unit is key to improving the selectivity of LGMD-based visual systems. This approach ensures the model selectively responds to approaching stimuli while significantly simplifying its structural complexity.

D. Comparisons with State of the Arts

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 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 [29] and LGMD2 [32], and a hybrid LGMDs neural system in which LGMD1 and LGMD2 operate in parallel with their spiking frequencies integrated [48].

1) Indoor structured scenes: In the first set of comparative experiments, we employed relatively simple indoor stimuli to evaluate model performance. Fig. 8 illustrates the spiking frequency outputs of each model across three test 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.

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. 9 presents the performance of all models in

TABLE III: F_1 score (%) of models to 104 real-world sets

	The proposed	[48]	[29]	[32]
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%

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 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.

3) Statistical results: Finally, Table III 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. 7. 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 collision perception in real-world visual environments.

Meanwhile, Fig. 10 articulates the F_1 score distribution of the six models across 104 datasets. The results show that the LGMD2-Derivative model achieves near-optimal performance on single-category datasets (i.e., ball, UAV, and crash) as well as on the overall dataset (overall), performing comparably to the LGMD2-LGMD1 model but with more simplified architecture.

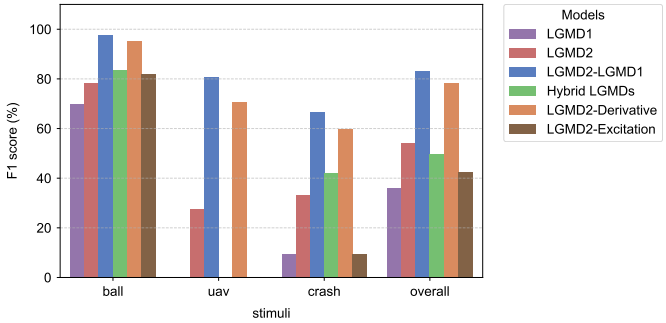

Fig. 10: The statistical results of F_1 scores of six comparative models tested upon 104 real-world datasets. The proposed model, i.e., LGMD2-Derivative, has comparable performance with the best LGMD2-LGMD1 composite model, yet with more simplified architecture well-suited for micro-robot onboard implementation.

E. Embodiment in Micro-Robot Vision

If elevating temporal derivatives is indeed the core factor enhancing approaching motion selectivity, a critical question arises: does this mechanism retain its efficacy and efficiency when embodied in micro-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 [48], 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 derivatives, as proposed in our model, offers superior performance in robotic looming perception under real-world conditions.

1) Robot visualization: What does the elevation of temporal derivative contribute to robot 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 walls. The layer-wise feature representations of the proposed model across Scenes $1 \sim 4$ are presented in Fig. 11.

The visualizations reveal that the application of an additional 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 stimuli, 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 derivatives thus acts as a selective amplifier for rapid, transient motion cues while

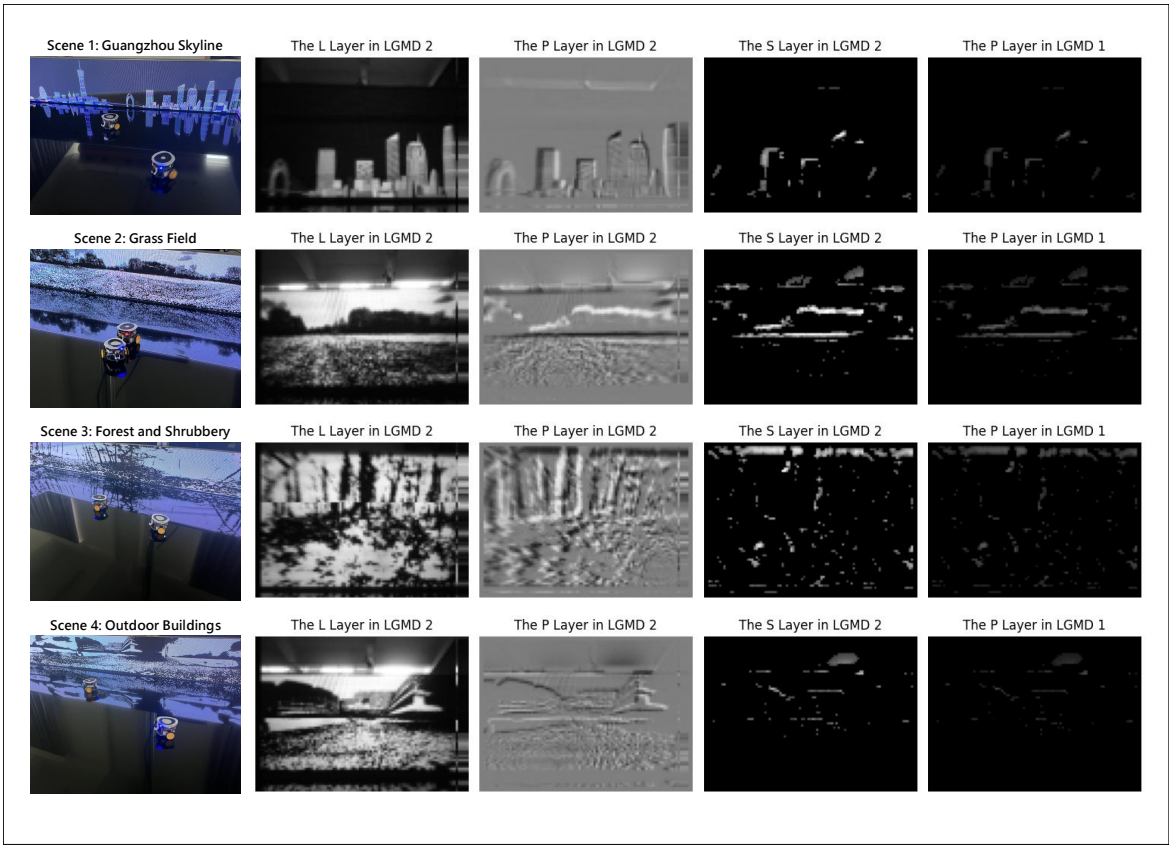

Fig. 11: Four experimental visual scenarios (left column) in the arena setting, and the corresponding layer-wise visualizations of the proposed time-derivative model are shown. Scenes 1∼4 correspond to rows 1∼4, featuring a city skyline and three complex natural environments, respectively.

suppressing static or slowly varying background features. This mechanism significantly enhances the LGMD model's robustness and selectivity, reinforcing the quantitative advantages demonstrated in the preceding experimental sections.

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 III-A2, as illustrated in Fig. 3, Fig. 11, and Fig. 12. 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. 12, with the corresponding success rates for each of the four visual environments listed in Table IV. 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 city 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 derivatives employed in visual information processing. Experimental results strongly suggest that elevating the order of temporal derivatives is pivotal for enhancing model robustness, particularly when confronted with complex and dynamic real-world visual scenarios.

TABLE IV: Success rate (%) of collision avoidance guided by two comparative micro-robot visual systems

Scene/Robot embodiment	[48]	The proposed
City skyline	95.88%	**96.25%**
Grass field	83.21%	**90.55%**
Forest/Shrubbery	88.93%	**91.48%**
Outdoor buildings	92.09%	**93.16%**

IV. Discussion

Through a series of systematic experiments, from offline simulations to real-world online tests, we have validated the robustness and effectiveness of the proposed LGMD2-Derivative neuronal assembly model which is ideally suited to micro-robot vision. By incorporating elevated temporal derivative processing, this model exhibits strong collision selectivity in real-world tasks.

Two types of looming-sensitive neurons–LGMD1 and LGMD2–have been identified in the locust's optic lobe. These neurons are not only spatially adjacent but also share similar morphologies and functions, both specialized in detecting approaching objects. In this study, the LGMD2-LGMD1 model was simplified to preserve only the retina layer of the

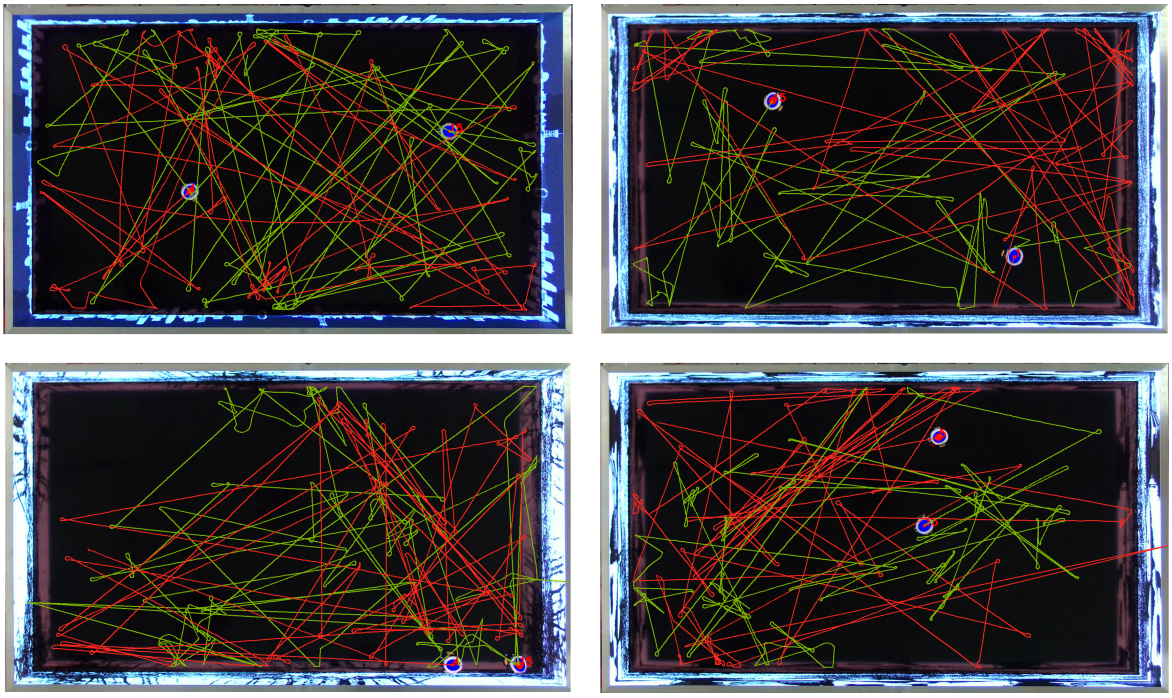

Fig. 12: Robot arena tests results: overtime trajectories of two robots are illustrated where green one indicates the proposed LGMD2-Derivative model and red one denotes the comparative hybrid LGMD model [48]. According to Fig. 11, four visual scenes as LED background were tested. Upper-left: scene-1, city 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 collision perception and avoidance in all tested scenes. The proposed model performs slightly better regarding success rate in all tested scenes as demonstrated in Table IV.

second module. This design emphasizes temporal derivative enhancement at the output of the first module, significantly boosting collision-selectivity and aligning more closely with physiological neuronal characteristics.

Although current anatomical and physiological evidence remains insufficient to confirm the existence of such a compound mechanism in biological systems, our modeling results strongly suggest that introducing neuronal assembly in cascade is effective for improving motion selectivity. These findings offer compelling support for future investigations into the underlying neural mechanisms of collision-sensitive visual systems.

The closed-loop arena tests confirmed the effectiveness and robustness of the proposed LGMD2-Derivative model in guiding robotic collision avoidance (Fig. 12). Notably, the model outperformed the comparative approach, achieving a higher success rate in avoiding collisions across diverse visual environments (Table IV). Although the model reached an F_1 score of 78.26% on datasets containing both looming and non-looming stimuli–demonstrating strong collision detection performance–false positives and missed detections still occurred.

In scenes with cluttered backgrounds (e.g., UAV and crash scenarios), local edge fluctuations in the complex environment were sometimes misinterpreted as looming cues, resulting in false alarms. Conversely, in low-contrast scenes, the model struggled to extract clear edge expansion features, leading to missed detections. These results highlight the need for further enhancement in background suppression and contrast sensitivity for more reliable performance under diverse real-world conditions.

Furthermore, inspired by biological systems, our study proposes a novel approach for constructing artificial collision detection systems. We draw on the motion detection mechanism of locusts, which, despite their simple neural architecture and limited number of neurons, exhibit highly efficient visual processing and complex behavioral control. In contrast, the human brain possesses a far more intricate structure, relying on spiking neural activity for cognitive functions.

Future work will primarily, though not exclusively, address the following questions. First, as the current neuronal module focuses mainly on first-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.

Furthermore, we plan to investigate the use of event-driven sensors as input streams for the proposed approach. Event-driven sensors are ideally suited for real-time, fast motion detection tasks due to their advantages in low latency, high dynamic range, low power consumption, and asynchronous signal processing. However, the current framework operates under a synchronous processing paradigm, making it unsuitable for directly handling asynchronous event streams. To

address this limitation, we plan to adapt the architecture to better accommodate the temporal characteristics of such inputs. One promising direction is to convert the model into a spiking neural network and integrate it with event-based cameras, enabling efficient and low-latency collision detection in real-time dynamic environments.

Lastly, our proposed approach is inspired by the visual system of locusts and is entirely data-free. In contrast, modern event-driven or deep learning-based motion detection models heavily rely on large-scale datasets, making direct comparisons under the current research framework impractical.

V. Concluding Remarks

In this paper, we proposed a time-derivative–based model that substantially streamlines the neural architecture without compromising collision selectivity. Through both computational simulations and robotic implementations, we identified the core mechanism underlying enhanced selectivity to approaching motion, demonstrating that elevating the temporal derivative of continuous visual streams markedly improves discrimination of looming stimuli from other forms of visual motion. Building on this insight, we developed a simplified neuronal assembly framework centered on temporal-derivative processing. Comparative evaluations showed that the proposed model consistently outperforms state-of-the-art bio-inspired approaches. When deployed on micro-robotic platforms, the model achieves performance comparable to previously developed composite frameworks [43], [48], while significantly reducing computational complexity. This efficiency makes the approach particularly well suited for micro-mobile robots with constrained computational resources. Beyond its engineering implications, this work also offers plausible insights for neuroscience, suggesting that the high efficacy of locust LGMD neurons may arise from interconnected neuronal assemblies, rather than isolated single-neuron computation, in achieving optimal looming perception.

Appendix

This appendix supplements the paper by elaborating on the algorithms of comparative LGMD2-Excitation model with inhibition of the second LGMD1 module ablated from a previous composite LGMD model [43]. The neural processing of first LGMD2 module conforms to the algorithms in Section II. The key difference laid upon the second cascaded LGMD1 module as the following equations.

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

$$\hat{P_{on}}(x,y,t) = \left[TD(x,y,t)\right]^{+} + \beta \cdot \hat{P_{on}}(x,y,t-\Delta_t) \tag{a.2}$$

$$\hat{P_{off}}(x,y,t) = -\left[TD(x,y,t)\right]^{-} + \beta \cdot \hat{P_{off}}(x,y,t-\Delta_t) \tag{a.3}$$

$$\hat{E_{on}}(x,y,t) = \iint \hat{P_{on}}(u,v,t) W_e(x-u, y-v)\, du\, dv \tag{a.4}$$

$$\hat{E_{off}}(x,y,t) = \iint \hat{P_{off}}(u,v,t) W_e(x-u, y-v)\, du\, dv \tag{a.5}$$

$$\hat{\Phi}(x,y,t) = \hat{E_{on}}(x,y,t) + \hat{E_{off}}(x,y,t) \tag{a.6}$$

$$K(t) = \iint_{\Omega} \hat{\Phi}(u,v,t)\, du\, dv,\ \ \xi(t) = \left(1 + e^{-K(t)\cdot(|\Omega|\cdot\alpha_2)^{-1}}\right)^{-1} \tag{a.7}$$

$$\hat{\xi}(t) = \begin{cases} \alpha_3 \cdot \xi(t), \text{if } \left(\xi(t) - \xi(t-\Delta_t)\right) > 0 \\ \alpha_3 \cdot \left(\hat{\xi}(t-\Delta_t) + \xi(t) - \xi(t-\Delta_t)\right), \text{otherwise} \end{cases} \tag{a.8}$$

$$\alpha_3 = \frac{\tau_s}{\tau_s + \Delta_t} \tag{a.9}$$

$$S_{pi}(t) = \left[\mathrm{e}^{(\alpha_4 \cdot (\hat{\xi}(t) - T_{sp}))}\right] \tag{a.10}$$

$$C_{ol}(t) = \begin{cases} \text{True}, & \text{if } \left(\sum_{i=t-n_t}^{t} S_{pi}(i)\right) \times 1000/\left(n_t \cdot \Delta_t\right) \geq T_c \\ \text{False}, & \text{otherwise} \end{cases} \tag{a.11}$$

References

[1] N. Franceschini, "Small brains, smart machines: from fly vision to robot vision and back again," *Proceedings of the IEEE*, vol. 102, no. 5, pp. 751–781, 2014.

[2] J. R. Serres and F. Ruffier, "Optic flow-based collision-free strategies: From insects to robots," *Arthropod Structure & Development*, vol. 46, no. 5, pp. 703–717, 2017.

[3] Q. Fu, H. Wang, C. Hu, and S. Yue, "Towards computational models and applications of insect visual systems for motion perception: A review," *Artificial Life*, vol. 25, no. 3, pp. 263–311, 2019.

[4] Q. Fu, "Motion perception based on on/off channels: A survey," *Neural Networks*, vol. 165, pp. 1–18, 2023.

[5] F. T. Muijres, M. J. Elzinga, J. M. Melis, and M. H. Dickinson, "Flies evade looming targets by executing rapid visually directed banked turns," *Science*, vol. 344, no. 6180, pp. 172–177, 2014.

[6] O. J. Bertrand, J. P. Lindemann, and M. Egelhaaf, "A bio-inspired collision avoidance model based on spatial information derived from motion detectors leads to common routes," *PLoS Computational Biology*, vol. 11, no. 11, p. e1004339, 2015.

[7] N. C. Klapoetke, A. Nern, M. Y. Peek, E. M. Rogers, P. Breads, G. M. Rubin, M. B. Reiser, and G. M. Card, "Ultra-selective looming detection from radial motion opponency," *Nature*, vol. 551, no. 7679, pp. 237–241, 2017.

[8] F. C. Rind and P. J. Simmons, "Local circuit for the computation of object approach by an identified visual neuron in the locust," *Journal of Comparative Neurology*, vol. 395, no. 3, pp. 405–415, 1998.

[9] F. C. Rind, S. Wernitznig, P. Pölt, A. Zankel, D. Gütl, J. Sztarker, and G. Leitinger, "Two identified looming detectors in the locust: ubiquitous lateral connections among their inputs contribute to selective responses to looming objects," *Scientific Reports*, vol. 6, no. 1, p. 35525, 2016.

[10] Q. Fu, C. Hu, P. Liu, and S. Yue, "Towards computational models of insect motion detectors for robot vision," in *Annual Conference Towards Autonomous Robotic Systems*. Springer, 2018, pp. 465–467.

[11] H. Wang, Q. Fu, H. Wang, P. Baxter, J. Peng, and S. Yue, "A bioinspired angular velocity decoding neural network model for visually guided flights," *Neural Networks*, vol. 136, pp. 180–193, 2021.

[12] H. Wang, Q. Fu, H. Wang, J. Peng, and S. Yue, "Constant angular velocity regulation for visually guided terrain following," in *Artificial Intelligence Applications and Innovations*. Springer International Publishing, 2019, Conference Proceedings, pp. 597–608.

[13] W. E. Green, P. Y. Oh, and G. Barrows, "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*, vol. 3. IEEE, 2004, pp. 2347–2352.

[14] W. E. Green and P. Y. Oh, "Optic-flow-based collision avoidance," *IEEE Robotics & Automation Magazine*, vol. 15, no. 1, pp. 96–103, 2008.

[15] M. B. Milde, O. J. Bertrand, R. Benosmanz, M. Egelhaaf, and E. Chicca, "Bioinspired event-driven collision avoidance algorithm based on optic flow," in *2015 International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP)*. IEEE, 2015, pp. 1–7.

[16] N. C. Klapoetke, A. Nern, M. Y. Peek, E. M. Rogers, P. Breads, G. M. Rubin, M. B. Reiser, and G. M. Card, "Ultra-selective looming detection from radial motion opponency," *Nature*, vol. 551, pp. 237–241, 2017.

[17] B. Zhou, Z. Li, S. Kim, J. Lafferty, and D. A. Clark, "Shallow neural networks trained to detect collisions recover features of visual loom-selective neurons," *eLife*, vol. 11, p. e72067, jan 2022.

[18] J. Zhao, S. Xi, Y. Li, A. Guo, and Z. Wu, "A fly inspired solution to looming detection for collision avoidance," *iScience*, vol. 26, p. 106337, 2023.

[19] M. Hua, Q. Fu, J. Peng, S. Yue, and H. Luan, "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*, 2022, Conference Proceedings.

[20] R. Liu and Q. Fu, "Attention-driven LPLC2 neural ensemble model for multi-target looming detection and localization," in *IEEE The International Joint Conference on Neural Networks*, 2025, Conference Proceedings.

[21] P. J. Simmons and F. C. Rind, "Responses to object approach by a wide field visual neurone, the LGMD2 of the locust: characterization and image cues," *Journal of Comparative Physiology A*, vol. 180, pp. 203–214, 1997.

[22] F. C. Rind, "Motion detectors in the locust visual system: from biology to robot sensors," *Microscopy Research and Technique*, vol. 56, no. 4, pp. 256–269, 2002.

[23] F. Gabbiani, H. Krapp, C. Koch, and G. Laurent, "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*, vol. 3. IEEE, 2002, pp. 1968–1969.

[24] F. Gabbiani, H. G. Krapp, N. Hatsopoulos, C.-H. Mo, C. Koch, and G. Laurent, "Multiplication and stimulus invariance in a looming-sensitive neuron," *Journal of Physiology-Paris*, vol. 98, no. 1-3, pp. 19–34, 2004.

[25] F. C. Rind and D. Bramwell, "Neural network based on the input organization of an identified neuron signaling impending collision," *Journal of Neurophysiology*, vol. 75, no. 3, pp. 967–985, 1996.

[26] S. Yue and F. C. Rind, "Collision detection in complex dynamic scenes using an LGMD-based visual neural network with feature enhancement," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 705–716, 2006.

[27] M. Keil, "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*, vol. 24, 2011.

[28] S. Bermúdez i Badia, U. Bernardet, and P. F. Verschure, "Non-linear neuronal responses as an emergent property of afferent networks: a case study of the locust lobula giant movement detector," *PLoS Computational Biology*, vol. 6, no. 3, p. e1000701, 2010.

[29] Q. Fu, C. Hu, J. Peng, and S. Yue, "Shaping the collision selectivity in a looming sensitive neuron model with parallel on and off pathways and spike frequency adaptation," *Neural Networks*, vol. 106, pp. 127–143, 2018.

[30] M. S. Keil, E. Roca-Moreno, and A. Rodriguez-Vazquez, "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*. IASTED, 2004, Conference Proceedings, pp. 340–345.

[31] L. Salt, D. Howard, G. Indiveri, and Y. Sandamirskaya, "Parameter optimization and learning in a spiking neural network for uav obstacle avoidance targeting neuromorphic processors," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3305–3318, 2019.

[32] Q. Fu, C. Hu, J. Peng, F. C. Rind, and S. Yue, "A robust collision perception visual neural network with specific selectivity to darker objects," *IEEE Transactions on Cybernetics*, vol. 50, no. 12, pp. 5074–5088, 2020.

[33] Q. Fu, Z. Li, and J. Peng, "Harmonizing motion and contrast vision for robust looming detection," *Array*, vol. 17, p. 100272, 2023.

[34] J. Zhao, Q. Xie, F. Shuang, and S. Yue, "An angular acceleration based looming detector for moving uavs," *Biomimetics*, vol. 9, no. 1, p. 22, 2024.

[35] X. Sun, Q. Fu, J. Peng, and S. Yue, "An insect-inspired model facilitating autonomous navigation by incorporating goal approaching and collision avoidance," *Neural Networks*, vol. 165, pp. 106–118, 2023.

[36] A. Kelkar and J. D. Medaglia, "Evidence of brain modularity," in *Encyclopedia of evolutionary psychological science*. Springer, 2021, pp. 2432–2441.

[37] M. Amer and T. Maul, "A review of modularization techniques in artificial neural networks," *Artificial Intelligence Review*, vol. 52, pp. 527–561, 2019.

[38] M. A. Bertolero, B. T. Yeo, and M. D'Esposito, "The modular and integrative functional architecture of the human brain," *Proceedings of the National Academy of Sciences*, vol. 112, no. 49, pp. E6798–E6807, 2015.

[39] D. Meunier, R. Lambiotte, and E. T. Bullmore, "Modular and hierarchically modular organization of brain networks," *Frontiers in neuroscience*, vol. 4, p. 200, 2010.

[40] J. Clune, J.-B. Mouret, and H. Lipson, "The evolutionary origins of modularity," *Proceedings of the Royal Society b: Biological sciences*, vol. 280, no. 1755, p. 20122863, 2013.

[41] M. Amer, "Modularity in artificial neural networks," Ph.D. dissertation, The University of Nottingham, 2019.

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

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

[44] A. Borst and T. Euler, "Seeing things in motion: models, circuits, and mechanisms," *Neuron*, vol. 71, no. 6, pp. 974–994, 2011.

[45] B. R. Geurten, K. Nordstrom, J. D. Sprayberry, D. M. Bolzon, and D. C. O'Carroll, "Neural mechanisms underlying target detection in a dragonfly centrifugal neuron," *Journal of Experimental Biology*, vol. 210, no. 18, pp. 3277–3284, 2007.

[46] T. W. Troyer, A. E. Krukowski, N. J. Priebe, and K. D. Miller, "Contrast-invariant orientation tuning in cat visual cortex: thalamocortical input tuning and correlation-based intracortical connectivity," *Journal of Neuroscience*, vol. 18, no. 15, pp. 5908–5927, 1998.

[47] C. Hu, Q. Fu, and S. Yue, "Colias iv: The affordable micro robot platform with bio-inspired vision," in *Annual Conference Towards Autonomous Robotic Systems*. Springer, 2018, pp. 197–208.

[48] Q. Fu, X. Sun, T. Liu, C. Hu, and S. Yue, "Robustness of bio-inspired visual systems for collision prediction in critical robot traffic," *Frontiers in Robotics and AI*, vol. 8, p. 529872, 2021.

[49] Q. Fu, H. Wang, J. Peng, and S. Yue, "Improved collision perception neuronal system model with adaptive inhibition mechanism and evolutionary learning," *IEEE Access*, vol. 8, pp. 108 896–108 912, 2020.

[50] Z. Qin, Q. Fu, and J. Peng, "A computational efficient and robust looming perception model based on dynamic neural field," *Neural Networks*, vol. 179, p. 106502, 2024.

[51] T. Krajník, M. Nitsche, J. Faigl, P. Vaněk, M. Saska, L. Přeučil, T. Duckett, and M. Mejail, "A practical multirobot localization system," *Journal of Intelligent & Robotic Systems*, vol. 76, pp. 539–562, 2014.