ADVANCED INTELLIGENT SYSTEMS
Open Access
www.advintellsyst.com

Looming Detection in Complex Dynamic Visual Scenes by Interneuronal Coordination of Motion and Feature Pathways

*Bo Gu, Jianfeng Feng, and Zhuoyi Song**

Detecting looming signals for collision avoidance encounters challenges in real-world scenarios, where moving backgrounds can interfere as an agent navigates through complex natural environments. Remarkably, even insects with limited neural systems adeptly respond to looming stimuli while in motion at high speeds. Existing insect-inspired looming detection models typically rely on either motion-pathway or feature-pathway signals, yet both are susceptible to dynamic visual scene interference. Coordinating interneuron signals from both pathways can enhance the looming detection performance under dynamic conditions. An artificial neural network is employed to construct a combined-pathway model based on *Drosophila* anatomy. The model outperforms state-of-the-art bio-inspired looming-detection models in tasks involving dynamic backgrounds, simulated by animated 2D-moving natural scenes or recorded in reality when an unmanned aerial vehicle performs obstacle collision avoidance tasks. Notably, by combining neural anatomy architecture and appropriate multiobjective tasks, the model exhibits convergent neural dynamics with biological counterparts post-training, offering network explanations and mechanistic insights. Specifically, a multiplicative interneuron operation enhances looming signal patterns and reduces background interferences, generalizing to more complex scenarios, such as AirSim 3D environments and real-world situations. The work introduces testable biological hypotheses and a promising bioinspired solution for looming detection in dynamic visual environments.

1. Introduction

Detecting looming signals, the expanding moving patterns generated by incoming objects, is at the heart of collision avoidance. Nevertheless, real-world challenges arise as the agent's motion generates dynamic background movements, bewildering the foreground object features and corrupting those vital looming signals (**Figure 1**). Remarkably, animals, even insects with simple brains, excel at looming detection while moving swiftly. Behavioral experiments show that even flying flies, the simplest insect with only 100 000 neurons in the brain, can react to looming stimuli with ultrafast visually directed banked turns.[1]

In the spirit of this extraordinary natural talent, the realm of neuroinspired computations has found innovative applications to address challenges of collision avoidance in artificial systems,[2,3] such as automobiles and unmanned aerial vehicles (UAVs).[4,5] This integration has led to significant reductions in computation time while showcasing impressive achievements in detecting imminent obstacles.[6] However, traditional solutions excel only in uncomplicated settings or static natural scenes,[7] and they stumble when confronted with complexities—especially scenarios where foreground objects and backgrounds dynamically intertwine.[8] In this particular investigation, we introduce an innovative neuroinspired computational framework driven by recent biological insights. This framework is poised to enhance the detection of looming signals in visually dynamic environments.

The core of a looming signal lies in how its subtended angle on the observer (θ) changes over time.[9] Thus, there are two theoretical schemes for looming detection: one based on calculating changing image features and the other on detecting motion patterns. Similarly, insect-inspired looming detection models fall into motion- and feature-pathway models, depending on whether they integrate local motion signals or contrast edge changes to form a looming selective response.

B. Gu, J. Feng, Z. Song
Institute of Science and Technology for Brain-Inspired Intelligence
Fudan University
Shanghai 200433, China
E-mail: zhuoyi.song@fudan.edu.cn

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aisy.202400198.

DOI: 10.1002/aisy.202400198

B. Gu, J. Feng, Z. Song
Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Ministry of Education
Fudan University
Shanghai 200433, China

J. Feng, Z. Song
MOE Frontiers Center for Brain Science
Fudan University
Shanghai 200433, China

J. Feng, Z. Song
Zhangjiang Fudan International Innovation Center
Shanghai 201203, China

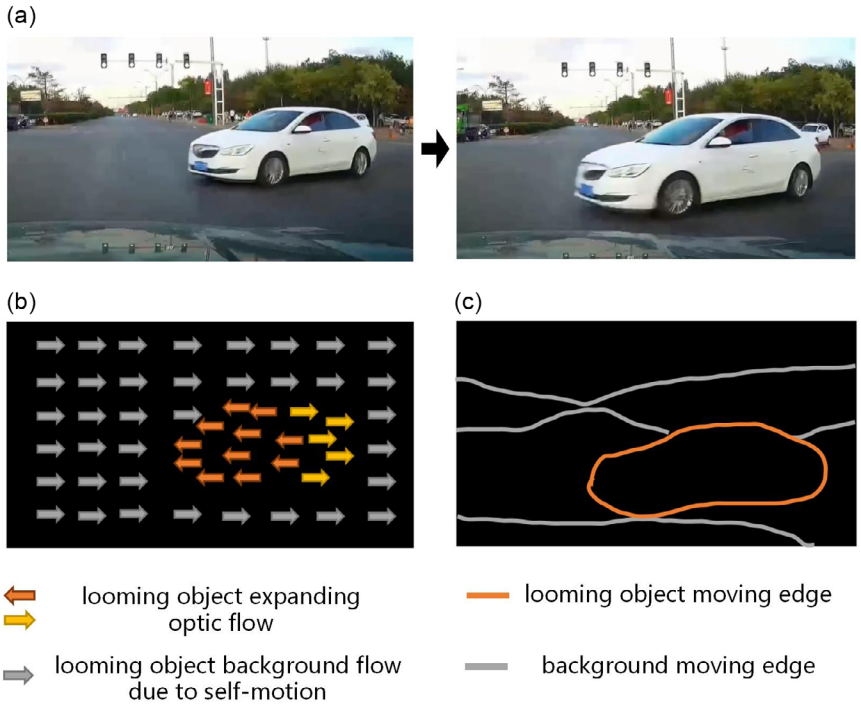

Figure 1. Looming estimation encounters interference from dynamic backgrounds. a) Illustration of a car approaching within a natural scene. b) The optic flow displayed in sequential images reveals a blending of the car's optic flow with the background optic flow, posing a challenge in distinguishing the car's looming pattern. c) The moving edges within the visual space appear indistinct, as the edges of the car assimilate with those of the surrounding environment.

The feature-pathway model, represented by the lobula giant movement detector (LGMD) model,[10,11] draws inspiration from the LGMD neuron in the locust visual system. With around four to five layers, this model operates as a shallow neural network using local illumination changes as inputs rather than explicitly encoding directional motion signals.[12] Achieving a nuanced balance between excitation and delayed inhibition, the model selectively responds to the feature of an expanding edge while disregarding receding ones. However, a challenge with this edge detection approach lies in its susceptibility to irrelevant moving features, resulting in false alarms for nearby or accelerating translating stimuli.[3] Researchers have attempted various enhancements to improve the accuracy and robustness of LGMD models in complex visual scenes. These enhancements include grouping nearby excitations,[2] exploring ON/OFF pathway competitions,[13,14] incorporating a separate contrast pathway to scale looming responses,[15] and introducing competitions between local excitation and inhibition signals.[16] Despite these efforts, persistent issues with background feature interference remain in these models.

An alternative network structure was proposed two decades ago, integrating local directional motion signals for looming detection,[17] even the biological evidence only came recently for LPLC2 (lobula plate/lobula columnar, type II) neurons in *Drosophila*.[18,19] The LPLC2 neuron integrates local motion signals and selectively responds to the diverging expanding motion pattern for effective looming detection.[15,20] However, this model structure prioritizes ultraselective responses to the radial motion expansion when an object approaches head-on, potentially compromising robustness in detecting various looming patterns from different directions.

Leveraging insights from recent connectomic data in *Drosophila*,[21] our investigation explores looming detection circuits that receive synaptic inputs from both motion- and feature-detection pathways. We delve into the coordination between the interneuronal signals from the feature and the motion pathways. Additionally, we inquire about the potential functional benefits of the coordination for looming detection in dynamic backgrounds. Our approach involves constructing an artificial neural network (ANN) model constrained by *Drosophila* neuroanatomy.

The ANN offers flexibility, serving as a combined pathway model or reducing to either a motion-pathway or a feature-pathway model, based on whether looming detection exclusively utilizes motion or feature detection inputs. Through training, we aim to predict the looming size (θ) and the expansion speed (ρ) with high accuracy in various dynamic visual scenes, including 2D animated dynamic backgrounds and those encountered in UAV obstacle detection & avoidance (ODA) tasks. Across all tested conditions, the combined-pathway model consistently outperforms single-pathway models, including traditional LGMD and LPLC2 models. An intriguing finding emerges: while integrating two pathways offers marginal benefits and may seem unnecessary in simple or static backgrounds, it becomes a significant catalyst in dynamic environments, seeing a 15–30% performance increase in the combined-pathway model.

Remarkably, a multiobjective optimization during training, emphasizing the detection of various motion patterns, leads the ANN to converge to solutions with neural dynamics akin to real biological counterparts. Mechanistically, the performance boost primarily stems from a novel multiplication between interneuron signals from motion and feature pathways. This interneuronal coordination empowers the model to amplify the looming signal pattern while adeptly suppressing interference noise generated by unpredictable background movements. These conclusions extend to other testing environments, including AirSim-simulated 3D environments and real-world scenarios.

In summary, our study shows that the interneuronal coordination in motion and feature pathways can be a potent catalyst for enhancing the looming detection efficacy of bioinspired models in intricate dynamic scenes. Furthermore, our multiobjective optimization highlights the importance of designing relevant training tasks to unveil the biological insights of neuroanatomy-constrained ANN models.

2. Related Work

Insect's visual system is a refined biological structure that enables daily survival tasks. *Drosophila* is a traditional model organism with a tiny brain but complicated behaviors, contributing to the science community with excellent experimental tools in genetics, anatomy, and physiology. More importantly, stereotyped neural circuits come in handy for mechanistic modeling. Here, we briefly review related looming detection models and neural circuits, focusing on the implementations with complementary pathways.

2.1. Complementary Motion and Feature Pathways in *Drosophila* Neural Circuit

In *Drosophila*, visual information is processed through four layers of neuropils in its optic lobe before being sent to the central brain and integrated with information from other sensory modalities.[22–24] These four optic lobe neuropils are the retina, lamina, medulla, and lobula complex. The lobula complex is divided into two distinct structural loci, the lobula plate and the lobula, providing parallel motion and feature processing pathways that coordinate behaviors.

The neurons in the lobula plate specifically code motion signals, such as optical flow information in the visual field.[25] Two parallel motion pathways for ON and OFF motion emerge early in the lamina but show motion-sensitive responses in T4 and T5 neurons within the lobula plate. T4 and T5 cells integrate excitation or inhibition inputs from adjacent column signals to form direction selectivity, coding local motion signals.[26] Furthermore, they send their dendrites to four separate layers in the lobula plate, coding moving signals in four coordinated directions, respectively.[12] Thousands of T4 and T5 cells tile the visual space, forming an optic flow map, like ganglion cells in vertebrates. The lobula plate tangential cells (LPTC) then integrate these local motion signals to create a sense of the global pattern induced by self-motion.

On the other hand, the lobula is a feature-detecting structure responsible for recognizing behaviorally relevant visual objects, such as conspecifics or predators.[27] Visual projection neurons in the lobula, like lobula columnar (LC) and lobula tangential cells,[28] send signals to specific brain regions coding various visual features.[29] For example, LC18, LC21, and LC11 can detect small moving objects that extend 2–4° and even code the object's moving speed with linear speed tuning curves.[30]

Although the lobula and the lobula plate are densely connected, indicating some interdependence between parallel processing of motion and feature, complementary signals may assist in certain behavioral functions. For example, looming detection circuitry involving LPLC1 and LPLC2 may show this cooperation.[28] LPLC2 selectively integrates local motion signals from T4 and T5 neurons to detect radial expanding motion patterns, indicating an approaching object on a collision course.[19] LPLC1, however, receives input from both motion pathways and feature detection interneurons (T2 and T3), triggering slowing behaviors upon stimuli of back-to-front motion, mimicking a frontal or parallel approaching object.[31] Unlike well-studied T4 and T5 neurons, T2 and T3 cells are only gaining more research attention these years.

T2 and T3 neurons have tight size tuning curves and provide excitatory presynaptic inputs to object-selective visual projection neurons.[32,33] They protrude the outline of a small object within a small target moving detector (STMD) circuit.[29,32] Their integration with T4 and T5 cells in the looming detection circuit and the extra functional benefits they provide still need to be fully understood. We hypothesized in this article that combining T2&T3 and T4&T5 signals may enhance the looming detection performance in dynamic visual scenes.

2.2. Insect Biology-Inspired Movement Detector Models

Researchers have created numerous bioinspired mechanistic models for movement detection in insects. However, these models perform well only in simple laboratory settings due to the difficulty of distinguishing object features from complex backgrounds. Continuous efforts have been made to enhance the performance of the models by introducing new specific mechanisms or pathways. Here, we briefly highlight the improvements made in three classes of related models: wide-field movement detectors, looming detectors, and STMDs.

2.2.1. Wide-Field Movement Detectors

Lobula plate tangential cells (LPTCs) process wide-field motion signals in the fly's visual system. They integrate local motion inputs coded by T4 and T5 cells, which can be modeled as elementary motion detectors (EMDs). Research suggests that LPTCs have receptive fields similar to matched filters for various optical flow patterns.[34] Although theories on LPTC computations have not progressed much since the early 2000s, discoveries in neurological implementations of EMD circuits have advanced rapidly. These include the ON–OFF pathway split,[35] separate four layers of T4 and T5 cells for different coordinates,[12] and hybrid motion computations involving preferred direction enhancement and nonpreferred direction

inhibition.[36,37] These new models improve direction selectivity and reduce output ambiguities, leading to improved optical flow estimations.[38]

2.2.2. Looming Detectors

Insect looming detectors were established in the mid-1990s and have seen significant advancements in recent years. There are two types of looming detectors,[10,11] LGMD models for the locust visual system and LPLC2 models for *Drosophila*.[15,20] The critical difference is whether local motion signals are explicitly coded in the network.

The LGMD and LPLC2 models face challenges from background interference in dynamic visual scenes. Researchers have proposed combining feature- and motion-based models to improve the performance of various motion patterns. Yue proposed two decades ago to integrate a translating-sensitive neural network to extract whole-field visual motion cues,[39] even though they concluded that such a redundant scheme does not add much to the LGMD performance.[40] Fu integrated an LPTC model and an LGMD model to enhance detection selectivities,[41] reducing false alarms for nearby translating stimuli.[42] More recently, Shuang et al. combined the two types of looming detection models to achieve both image velocity selectivities of LGMD and the radial motion opponency characteristics of LPLC2.[8]

However, most previous attempts integrated different models as independent modules, introducing competition mechanisms only in the output. Instead, a more effective approach is to introduce interaction among the pathways at the level of interneurons, creating a more reliable and efficient neural circuit. Moreover, the looming detector mostly outputs zero-one alarming signals instead of estimating the continuous looming size and speed. The latter are more valuable variables for subsequence collision avoidance control and are potentially coded in the brain. Looming size (θ) and looming speed (ρ) neurons may exist in neural circuits of various species,[43,44] such as locusts, pigeons, fish, flies, and mice.

2.2.3. STMDs

Our model also takes inspiration from the STMD model, a crucial set of object detection models.[45,46] Although not typically discussed with LGMD or LPLC models, recent studies in *Drosophila* suggest correlations between them. Neurons in the lobula, like LC18 and LC11, have been identified as STMDs. LC11 and LPLC1 may even share interneurons that encode the outline features of an object.[21,32] The classic STMD model detects fast-moving objects by correlating the leading and trailing edge signals with opposite directional contrast changes.[47] Modifications accounted for new experimental findings, such as LC18's smaller tuning size for small targets, by introducing a bounded crossover inhibition mechanism.[29] On the other hand, the selectivities for small targets of LC11 neurons were successfully replicated by introducing spatial–temporal pooling.[21] These novel STMD models rely on T2 and T3 to provide the moving edge signals before the correlation step.[32] We speculate that moving edges may also enhance looming object detection in dynamic backgrounds in *Drosophila*.

2.3. Biological Anatomy-Constrained Neural Networks

The mechanistic models discussed earlier are carefully designed to replicate the responses of specific neurons, but they may only work well under specific conditions. On the other hand, ANNs trained on datasets can perform efficiently with numerous parameters but lack interpretability. Recently, neural scientists have started constraining ANNs with real neural circuit anatomy,[38,48,49] showing that such models can converge to biological solutions and be more efficient and robust.[50] This convergence of biological neural circuits and ANNs can mutually validate their rationality under realistic conditions.

In the context of movement detectors in *Drosophila* neural circuits, several anatomy-constrained ANNs have been developed. For example, Mano et al. created a shallow convolutional ANN that models the processing steps of the elementary motion detection circuit.[38] Their model was trained in moving scenes to estimate speed using T4 and T5 signals. Drews et al. developed a similar ANN based on T4 and T5 neurons with divisive normalization, showing that spatial contrast normalization improves motion estimation on different natural scenes.[51] Zhou et al. established a two-layer convolutional ANN for the looming-detection neuron LPLC2, successfully distinguishing between looming objects that will hit or miss.[31] However, there has yet to be an ANN modeling the elementary motion detection circuits, looming, and translating object detection circuits to investigate their interaction effects.

3. Results

3.1. Anatomically Constrained ANN Model for Looming Detection in Dynamic Visual Scenes

There are various motion patterns in dynamic scenes of the real world. The *Drosophila* neural circuits have evolved parallel branches to handle them. Our model also incorporates parallel branches for wide-field background motion, translating object motion, and looming motion patterns (**Figure 2**a). These branches share the same interneurons (T2, T3, T4, and T5), making the network compact.

The background branch estimates background motion direction using local motion signals from T4 and T5 neurons.[38] We used a basic EMD model with static compression for this branch (Figure 2b).[37] The translation branch takes inputs from T2 and T3 neurons (Figure 2c) and uses an STMD structure to detect small translation objects (Figure 2d). The STMD model adopts ideas from the inhibition-of-inhibition mechanism from neural models for LC18 in *Drosophila*.[29] We then calculated the translating object's speed using a two-layer convolutional neural network (CNN). The looming detection branch combines inputs from all four types of interneurons to acquire a looming map that shapes the expanding pattern of the looming object. Based on this map, we computed the expanding size and speed of the looming object using two-layered CNNs (Figure 2e).[43] The two-layered CNNs mimic the unknown nonlinear mapping of some neurons that can convert the interneuron signals from T2/T3 and T4/T5 to continuous size or speed estimations. This network structure essentially allowed us to investigate

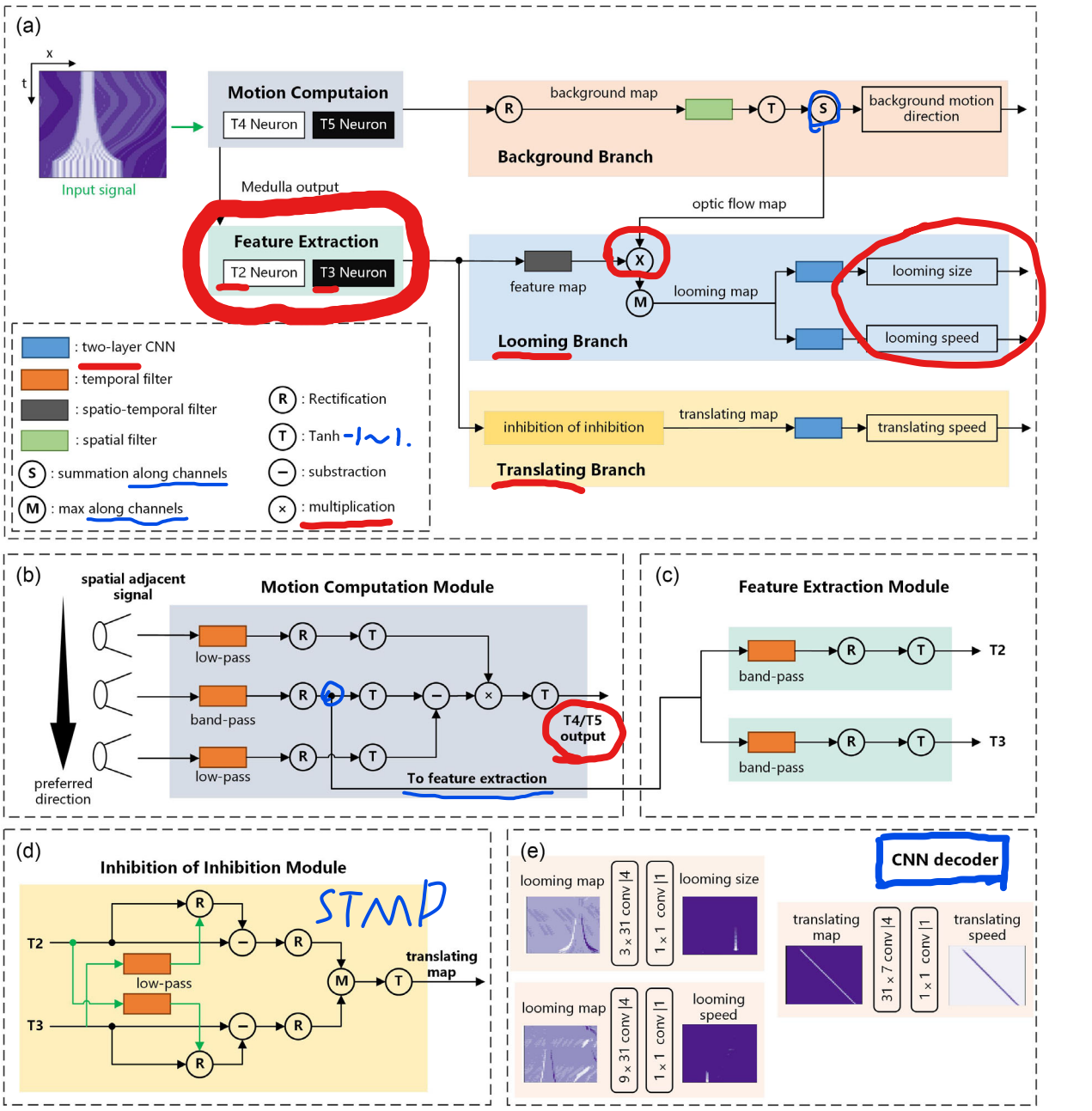

Figure 2. Model framework. a) The model comprises three branches dedicated to background motion direction, translating object speed, and detection of looming objects' size and speed, respectively. The looming branch synergizes signals from motion-detection interneurons (T4 and T5 cells) and feature-detection interneurons (T2 and T3 cells). b) The elementary motion detection model computes the local motion direction by integrating signals from three adjacent columns. c) The feature extraction module calculates the T2 and T3 signals based on cascaded filtering and nonlinear operations. d) The inhibition of inhibition module is grounded in the LC18 neuron model, calculating the translating map using the object's ON and OFF moving edge signals. e) A two-layer CNN was used as a decoder for each branch.

how the neural circuit creates more useful interneuron signals for dynamic looming detection.

The interactions between T2/T3 and T4/T5 in *Drosophila* are not fully understood. We speculated that they sharpen each other's signals and filter out noise. Thus, we introduced a multiplication operation to combine the visual information from T2/T3 and T4/T5. Multiplication makes the ANN a combined-pathway model but reduced to a motion-pathway or feature-pathway model if only T4/T5 or T2/T3 signals are used for looming detection. We trained the spatial–temporal receptive fields of all interneurons using a loss function that weights errors from all three branches. This multiobjective optimization

alleviates the fine-tuning problem when each branch is trained separately and allows the network to converge to biologically plausible solutions that otherwise would not be possible (data not shown). This emphasizes the significance of designing training tasks for bioinspired neural networks targeted to solve real-world problems.

3.2. Visual Stimuli Dataset for Training the ANN Model

As looming detection is at the core of collision avoidance, the ideal datasets for training the ANN models should meet three conditions: 1) Source of videos: The datasets should consist of videos recorded onboard by an agent, such as a UAV. 2) Realistic obstacle Interaction: The agent in the videos should navigate through natural and complex outdoor environments, approaching and avoiding obstacles dynamically. 3) Labeling of videos: The recorded videos must be labeled with information about the size and expansion speed of the encountered obstacles.

However, obtaining datasets that meet all these conditions is challenging. Some datasets are recorded in controlled lab settings with simple backgrounds,[52] while others captured in the wild lack accurate labels.[53] To address this limitation, we conducted training and testing on artificial and publicly available real datasets.

Both datasets are highly dynamic videos that contain moving foreground objects and moving background scenes. In the artificial datasets, we used panoramic natural images as the backgrounds (**Figure 3**a left) and manually added nonlinear expanding looming blocks and translating moving objects to simulate the foreground (Figure 3b). The natural images shifted horizontally with varied speeds to mimic moving backgrounds (Figure 3a middle), allowing us to assess the models' performance with naturalistic image dynamics. The public real datasets, known as the TUDelft ODA datasets,[52] were recorded indoor by Micro Air Vehicle Laboratory (MAVLab) for UAV ODA tasks. This real-world dataset allowed us to evaluate our models using data associated with practical applications.

We focused on a simplified 1D scenario, using a line of pixels as raw data (Figure 3a left). The motion of this pixel line created a 2D spatial–temporal pattern labeled with background motion directions, looming size, looming speed, and translating speed (Figure 3a right). While public ODA datasets lacked these labels, the TUDelft ODA datasets included videos, obstacle positions, and the UAV's 6D pose recordings, enabling us to calculate these labels directly (Figure 3c).

To test the robustness of our looming detection models across diverse dynamic scenes, we varied the statistical properties of both background and foreground objects in the artificial datasets. This involved manipulating static and dynamic backgrounds with different speeds, sizes of foreground translating objects, uniform or sine-wave pattern contrast for looming objects, and controlled object contrast to mimic natural predators. To enhance background complexities in the ODA dataset, we introduced low-amplitude Gaussian white noise to simulate the noisy environment in nature. Additionally, we added miniature white blocks and made them move homogeneously or divergently to mimic dynamic background interferences from complex environments (Figure 3d). Detailed information on the generation of artificial datasets and the labeling process for the TUDelft ODA datasets can be found in Section 5.

3.3. The Combined-Pathway Model Demonstrates Significant Looming Estimation Enhancement in Dynamic Backgrounds

We first compared the looming estimation performances of five models. The models included three of our own (combined-pathway, motion-pathway, and feature-pathway models) and two state-of-the-art insect-inspired looming detection models, CLGMD and LPLC2. The latter two represent traditional looming detection routes based on moving feature detection or motion detection, respectively. The CLGMD model integrates a spatial contrast pathway into the classical locust-inspired LGMD model for enhanced illumination invariance.[7] The LPLC2 model combines local motion signals from T4/T5 using a large divergent receptive field.[20] We reproduced the CLGMD and LPLC2 models with slight modifications to facilitate comparability with our models. We connected them with a looming decoder identical to our models and allowed parameter training in these two models.

The presence of dynamic backgrounds posed challenges for various motion detection tasks. The accuracy of background direction estimation significantly dropped when background moved stochastically with random speeds instead of homogeneously with a uniform speed (see **Figure 4**a). All models, excluding the LPLC2 model, performed well in static background (static columns in Figure 4b–d). However, their performance notably decreased for dynamic backgrounds (uniform and stochastic columns in Figure 4b–d). The CLGMD model experienced the most significant decrease in performance, with the R^2 of looming size estimation dropping from 0.85 on static backgrounds to below 0.4 on dynamic backgrounds (Figure 4c), indicating limited robustness. In contrast, the LPLC2 model exhibited greater robustness, maintaining performances at similar lower levels, with R^2 reaching around 0.4 for looming size estimation and 0.6 for looming speed estimation (Figure 4c,d).

Encouragingly, our three models generally outperformed the CLGMD and LPLC2 models in looming detection, irrespective of whether the evaluation used artificial or ODA datasets (Figure 4c–f). In particular, in all conditions tested with dynamic backgrounds, the combined-pathway model consistently outperformed the other two of our own models, achieving a 10–25% increase in the estimation accuracy (Figure 4c,d). This suggests that coordination between the feature and motion pathways provided better interneuronal signals for estimating looming size and speed. Furthermore, our three models exhibited greater robustness across different datasets than the CLGMD and LPLC2 models. The CLGMD and LPLC2 models experienced a dramatic performance drop in the ODA dataset, reaching a value of R^2 of 0.1, while the performance drop for our three models remained marginal.

Notably, intriguing results emerged when comparing looming estimation performances for artificial and ODA datasets. Logically, looming estimation tasks should be more challenging in nature than in laboratory conditions. Surprisingly, looming estimation performances decreased in the ODA dataset compared to the artificial dataset. Although the artificial dataset more

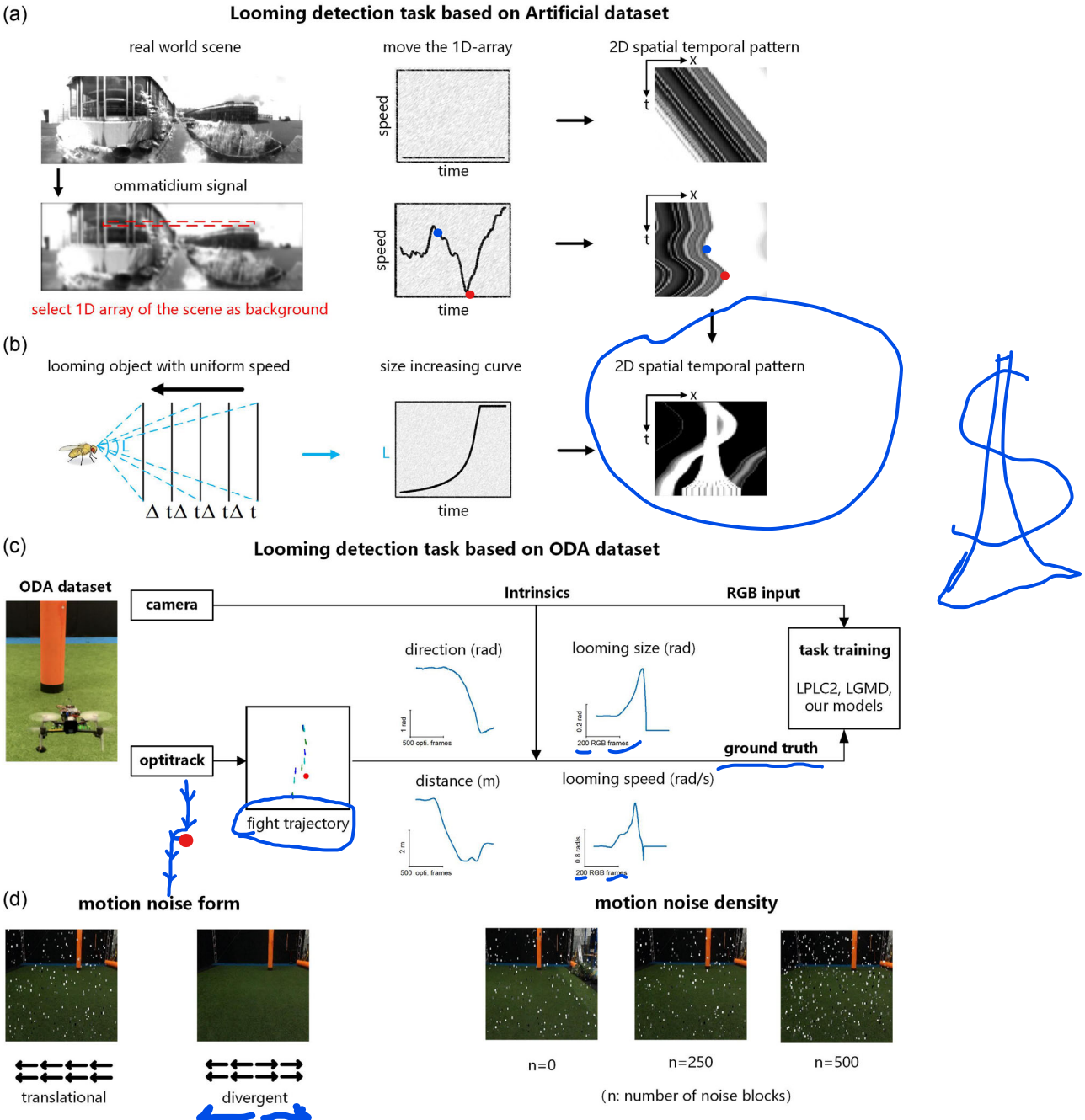

Figure 3. Dataset generation procedure: a) The process for generating the artificial dataset involves selecting a 1D array of signals from a panoramic image of a natural scene. The 1D image moves either with a uniform speed or a low-pass filtered speed sequence, and the movement of the 1D image over time generates a 2D spatiotemporal signal pattern. b) An object moves constantly toward collision, producing an exponentially expanding looming signal on the retina. c) The procedure for generating the ODA dataset involves utilizing the publicly available DelftODA dataset,[52] recorded during a flying UAV's laboratory ODA tasks. Looming size and speed are computed by aligning the camera-recorded video with the Optitrack-recorded pose data. d) Moving blocks are introduced into images in the ODA dataset to replicate background interference (c,d are adapted with permission.[52]).

closely resembles complex naturalistic background statistics, two factors may contribute to the diminished performance of the ODA dataset. First, the addition of low-amplitude Gaussian noise to the ODA dataset may perturb the models' robustness. Second, and perhaps more importantly, the ODA dataset reflects changes in an embodied environment as an agent moves within it. This embodied looming detection encompasses data with approaching, receding, side-way passing, or even void looming objects, resembling the varied looming cases encountered in real-world collision avoidance scenarios.

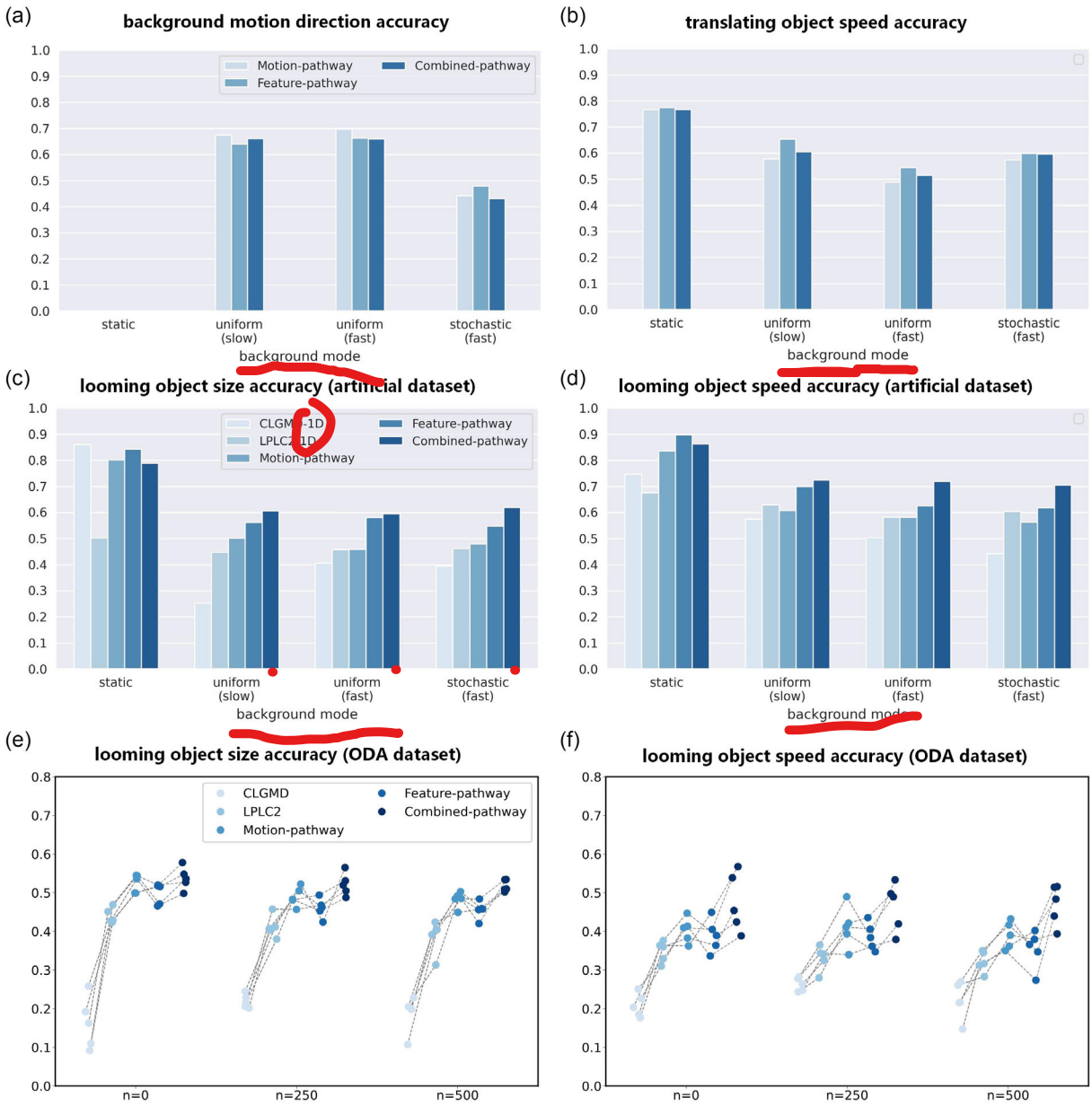

Figure 4. The combined-pathway model exhibited superior and more resilient performances on both a–d) artificial datasets and e,f) ODA datasets. We compared the performances of five looming detection models in static or dynamic backgrounds, featuring exponentially expanding objects with backgrounds moving stochastically or at various uniform speeds. The five models include two traditional bioinspired looming detection models, namely the CLGMD and LPLC2 models, and three of our ANN models—the motion-pathway, the feature-pathway, and the combined-pathway models. All models exhibited a noticeable decline in performance when the background was in motion. However, our three models outperformed and demonstrated greater robustness than the other two traditional models. Notably, the combined-pathway model achieved the best looming detection performance in all tested dynamic conditions. (a,b) Our three models performed similarly in estimating background motion direction and translating object speed, suggesting that the complementary mechanism did not adversely affect the functions of these two branches. The two traditional looming detection models did not perform such tasks. (c,d) The combined-pathway model improved looming estimation accuracy in all tested dynamic conditions but not in static backgrounds. (e,f) Our three models exhibited superior and more robust performance than the other two traditional models on the ODA dataset.

3.4. Testing with AirSim Simulated 3D Scenes and Real-World Looming Scenarios

Given that all three of our models were trained on 2D synthetic images or in controlled laboratory settings, we sought to assess the extent to which our results could be extrapolated to more realistic scenarios. To address this, we conducted additional tests on AirSim simulated 3D scenes and real-world looming data captured using a cell phone.[54]

In the AirSim simulated data, a flying agent maneuvered within a selected 3D environment, and the installed camera recorded the scene from its own perspective. To simulate a looming object

approaching, we introduced an expanding contrast block with a sine wave pattern. The agent's movement in the 3D scene varied, either moving forward with a looming object on a direct collision course or moving laterally with a looming object approaching from the side (see **Figure 5**a). In real-world scenes, an individual recorded the surroundings with a cell phone while walking or running toward a specific object, with an apparent looming object simulating the trajectory toward a collision in daylight. The recordings were intentionally left uncontrolled to capture a diverse range of looming scenarios in natural settings.

Notably, the AirSim simulated data differed from the earlier synthetic training data due to more realistic background

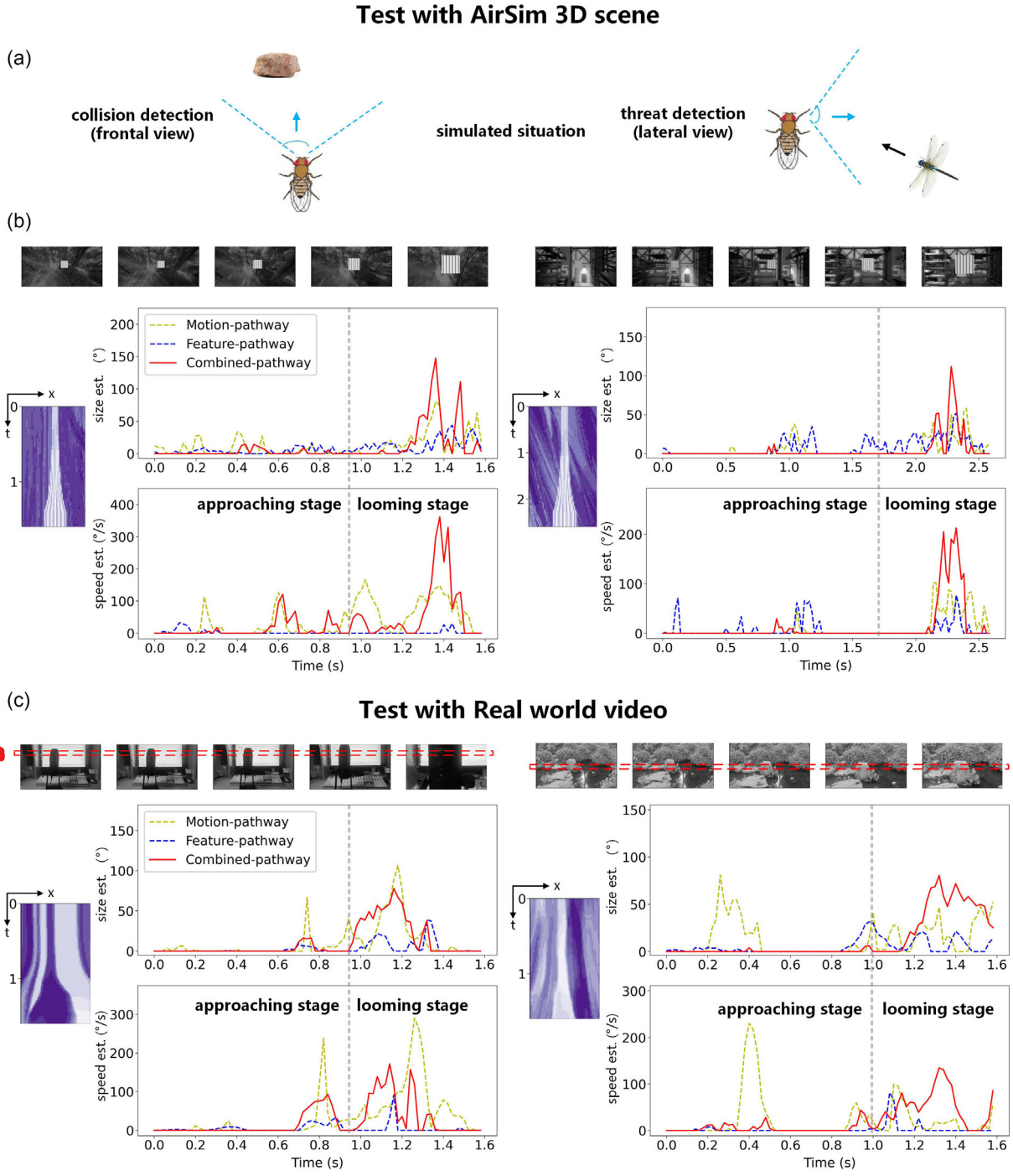

Figure 5. The combined-pathway model excels in looming estimation across both AirSim simulated 3D dynamic visual scenes and real-world scenarios. a) Looming detection during frontal approaches in AirSim simulated scenes. b) Looming detection during sideways approaches in AirSim simulated scenes. c) Videos recorded with a cell phone in the real world as a person runs toward an apparent looming object.

movements and diversified velocities of moving objects influenced by various depths. Consequently, the looming responses exhibited more noise (see Figure 5b) compared to those tested with 2D image movements, where looming estimation curves displayed clear exponential rising trends with minimal noise (ground truth in **Figure 6**). This contrast underscores how movements within a 3D scene can introduce more background interference to looming signals than pure 2D image movements. Interestingly, the impact of background interference varied for frontal versus sideways looming approaches (see Figure 5b). The motion-pathway model experienced more noise during frontal approaches (Figure 5b left), while the feature-pathway model exhibited more noise during sideways approaches (Figure 5b right). This observation highlights the importance of testing in an embodied 3D environment.

Despite the challenges, the combined-pathway model consistently outperformed the others, demonstrating enhanced responses and reduced noise in predicting looming size and speed (red lines in Figure 5b). For both frontal and sideways looming signals, the motion- and feature-pathway models often produced extremely noisy responses in the approaching stage (blue and yellow lines in Figure 5b), potentially leading to false alarms or faint speed estimations that may fail to detect looming (blue lines in Figure 5b–bottom). In contrast, the combined-pathway model improved the looming response, amplifying faint signals in speed estimations (blue lines in Figure 5b bottom), and suppressed noise, compensating for the limitations of each pathway (red lines in Figure 5b).

Real-world scenes introduced additional complexities such as uneven contrast edges and off-center looming objects with edges

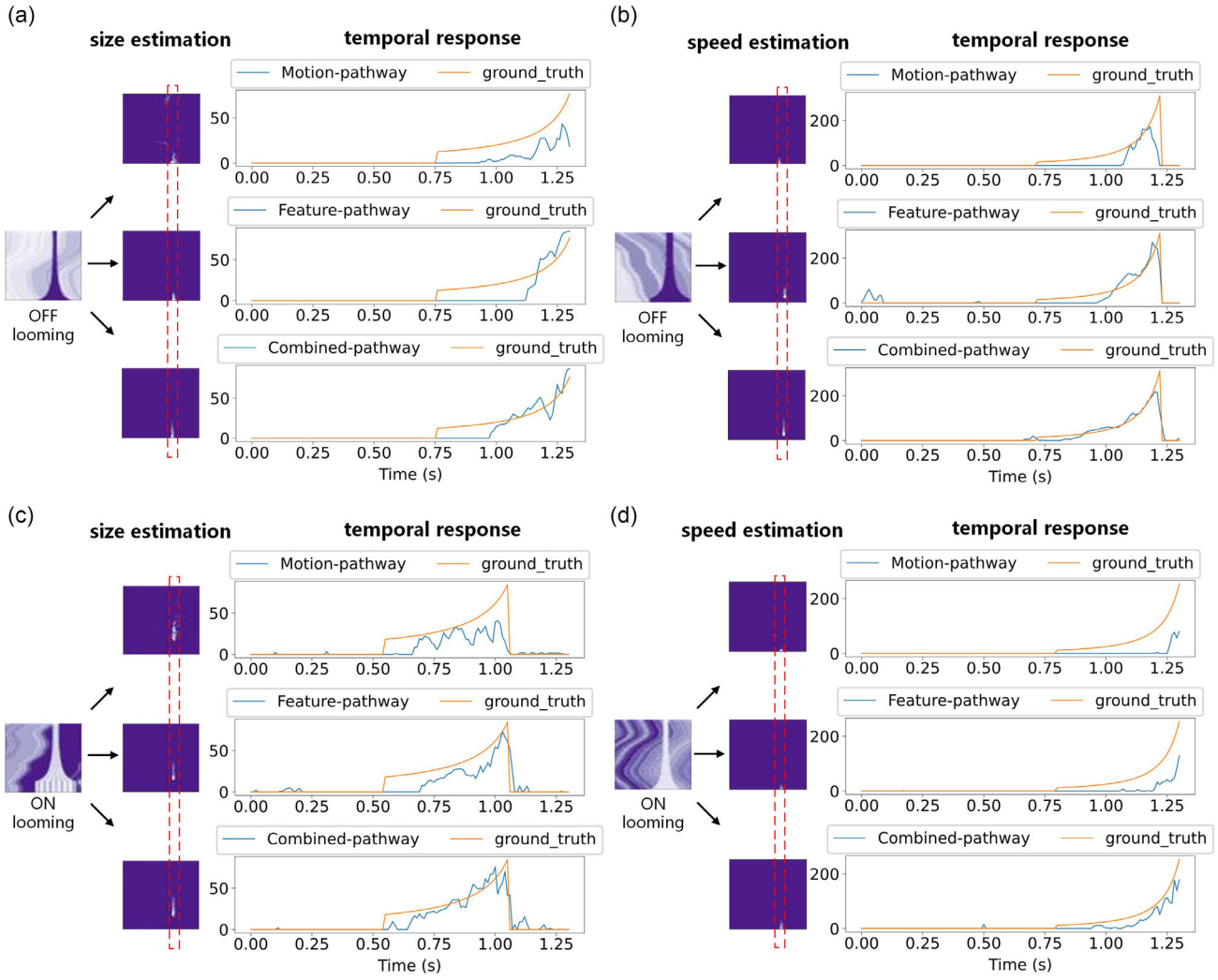

Figure 6. On the artificial dataset, when stimulated with an ON or OFF looming object approaching with a uniform speed, the ground truth looming size and speed expand exponentially on the retina (yellow lines). The combined-pathway model produced temporal responses that matched the ground truth more closely in all cases. Here, the combined-pathway model excels in promptly estimating the size and speed of looming objects. Early detection of collisions or predators is crucial for survival or flight safety. The combined-pathway model identified looming objects 50–200 ms earlier than the other two models. However, similar results are less evident in the ODA dataset (data not shown). a) Size estimation map and temporal responses for an ON-looming object. b) Speed estimation map and temporal responses for an ON-looming object. c) Size estimation map and temporal responses for an OFF-looming object. d) Speed estimation map and temporal responses for an OFF-looming object.

expanding at inhomogeneous speeds, posing even greater challenges for looming estimation. Despite this, the combined-pathway model consistently outperformed the others, providing clearer speed estimates, reducing noise and false alarms (red lines in Figure 5c). The performance of our model in both AirSim simulated data and real-world looming recordings showcased its adaptability to complex scenarios. It is important to note that the models were not retrained on new data; their performance could potentially improve with training on more realistic datasets in the future, although this was beyond the scope of our current research. Overall, the tests certify the general benefits of looming estimation by coordinating signals from both the motion and feature pathways.

3.5. Interneuron Responses to Looming and Translating Objects Show Expected Neural Dynamics

We sought to understand why our combined-pathway model exhibited superior performance and whether it could offer biologically plausible mechanistic insights. Initially, we examined the responses of interneurons, namely, T2\T3 and T4\T5 cells, in our ANN models to validate their resemblance to biological counterparts (**Figure 7**). We subjected the models to four types of moving objects in the visual field: ON and OFF looming objects (Figure 7a,b) and ON and OFF translating objects (Figure 7c,d). Importantly, all four types of interneurons encoded distinct visual features, as expected from biology.

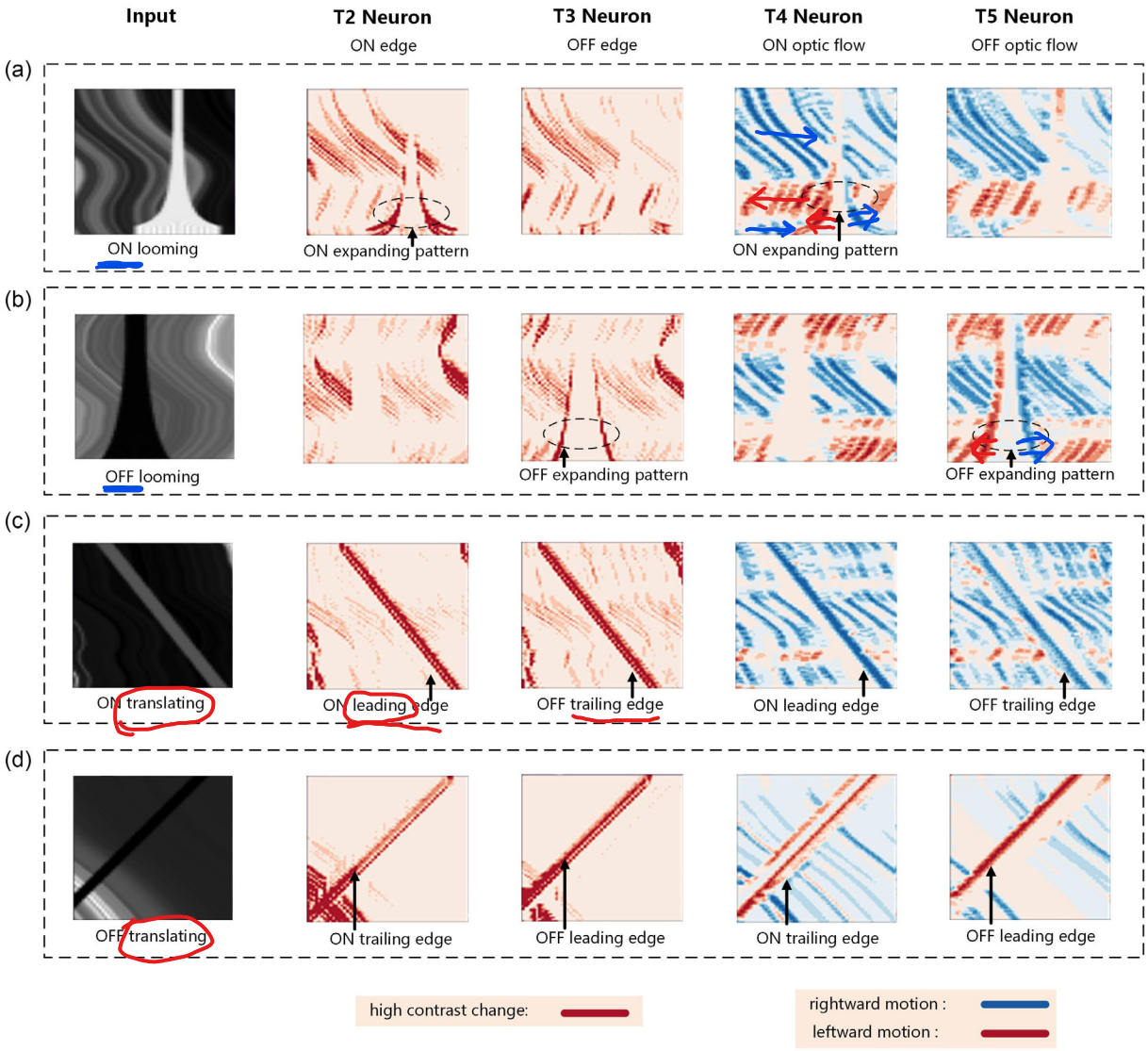

Figure 7. The interneurons perform as anticipated, resembling their biological counterparts. T2 and T3 segregate into ON/OFF pathways akin to T4 and T5 cells. T2 and T4 cells exhibit robust responses to ON-looming objects, while T3 and T5 cells selectively respond to OFF-looming objects. T2 and T3 focus on local high-contrast change signals with less background noise, whereas T4 and T5 extract local motion signals while maintaining motion direction opponency. a) Interneuron response map for an ON-looming object with a sine-wave pattern; T2 and T4 neuron responses depict a distinct expanding pattern. T4 shows clear opponent responses for opposite motion directions. b) Interneuron response map for an OFF-looming object with a sine-wave pattern. c) Interneuron response map for an ON-translating object. d) Interneuron response map for an OFF-translating object.

T2 and T3 primarily focused on detecting the moving edge of objects, responding minimally to background motion (see T2 and T3 in Figure 7d), except in the presence of a high-contrast object (see T2 and T3 in Figure 7c). Similar to T4 and T5, T2 and T3 segregated into ON/OFF pathways, indicating responsiveness to specific contrasted edges. For instance, T2 detected the ON leading edge of an ON translating object (see T2 in Figure 7c) or the ON trailing edge of an OFF translating object (see T2 in Figure 7d). Simultaneously, T3 responded to the OFF leading edge of an OFF translating object (see T3 in Figure 7d) or the OFF trailing edge of an ON translating object (see T3 in Figure 7c). In the presence of a looming object, T2 and T3 detected the expanding pattern for ON and OFF looming objects (see T2 and T3 in **Figure 8**a,b), respectively.

In contrast, T4 and T5 responded to local motion signals and were sensitive to ON or OFF looming objects, showing minimal response to objects with nonpreferred contrast. T4 responded to an ON looming object with a clear looming pattern (see T4 in

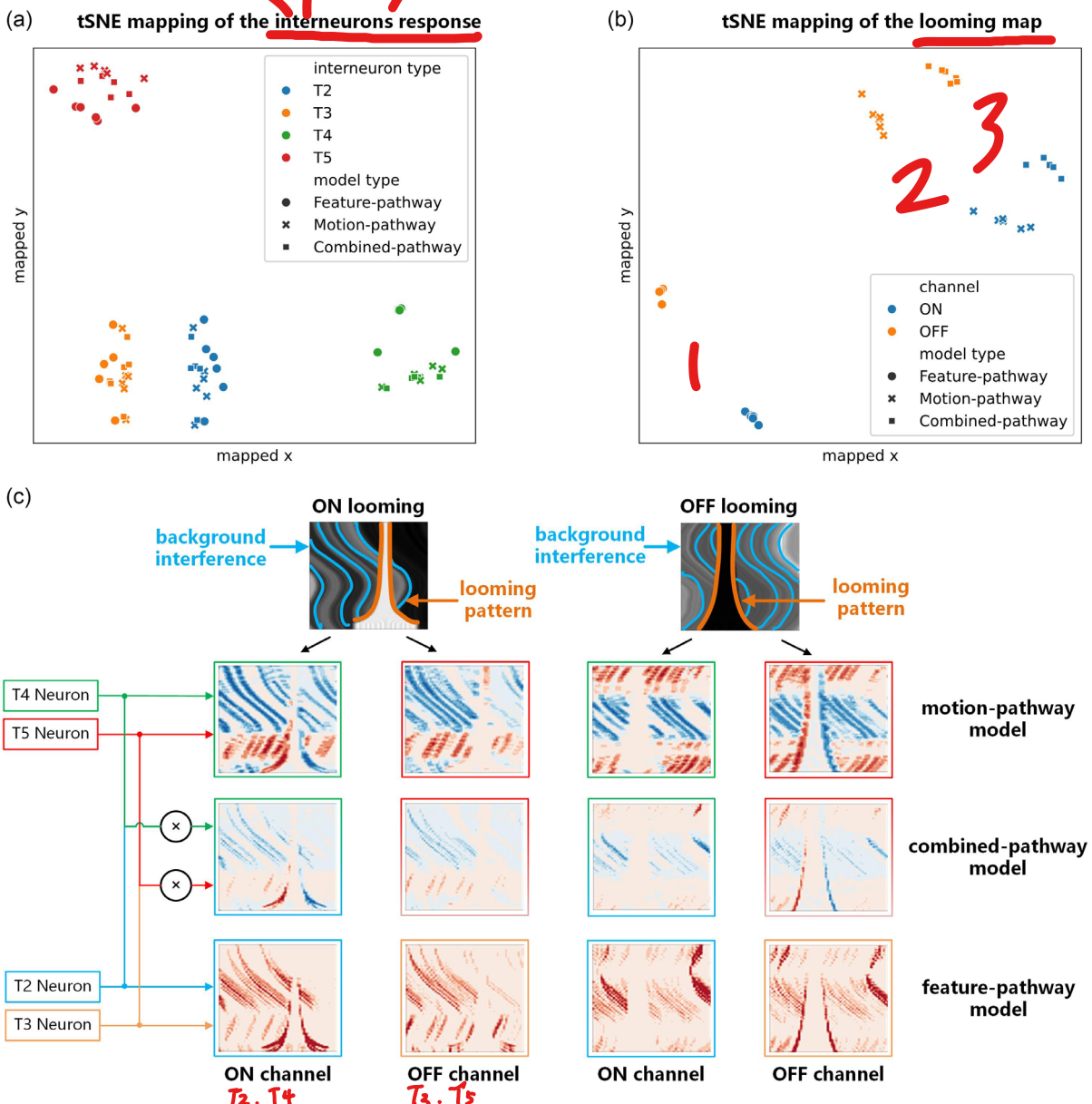

Figure 8. The multiplication is pivotal to the enhanced looming estimation performance in the combined-pathway model. a) In the tSNE mapping of interneuron signals from our three model types, four clusters corresponding to the four types of interneurons are evident, while data points for different model types overlap; each model type consists of six independently trained models. b) The tSNE mapping of the looming feature map reveals three clusters for the three model types. Together, (a, b) signify that multiplication is the reason behind the enhanced looming detection performance in the combined-pathway model. c) The multiplication elevates the looming signal pattern while mitigating interference noise from background movements. The looming object feature map of the motion-pathway model appears cluttered with background optic flows. The looming object feature map of the feature-pathway model is cleaner, but lacks direction opponency. In contrast, the combined-pathway model amalgamates the advantages of the other two models, resulting in a less noisy and direction-opponent looming object feature map.

Figure 7a), while T5 coded for OFF looming objects (see T5 in Figure 7b). Both exhibited motion direction opponency, characterized by positive and negative responses for two edges moving in opposite directions, as illustrated by the red and blue strips for leftward and rightward moving edges in their feature map (see T4 and T5 in Figure 7a,b). For translating objects, T4 responded to the ON leading edge of an ON-contrast object or the ON trailing edge of an OFF-contrast object (see T4 in Figure 7c,d). Conversely, T5 responded reversely (see T5 in Figure 7c,d). Compared to T2 and T3, the responses of T4 and T5 were more crowded due to background optical flow, indicating more severe background interference from the motion pathway.

3.6. The Multiplication of Interneuronal Signals Is the Key for Enhanced Looming Estimation Performance

Finally, we delved into the mechanism that led to the superior performance of the combined-pathway model in looming estimation. Examining the results in Figure 4, it became apparent that the model's enhancement lay in the postprocessing of interneuron signals—specifically, T4, T5, T2, and T3 signals. All three models demonstrated similar performances in estimating background motion direction and translating object speed (Figure 4a,b). This indicated that the complementary mechanism did not adversely affect the training of the background and translating object branches. Consequently, the combined-pathway model should have converged on similar interneuron functions compared to those trained independently for separate tasks.

To validate this, we compared the interneuron responses and the looming object feature map using t-distributed stochastic neighbor embedding (tSNE) mapping. Visualization of interneuron responses from six independently trained models for each model type revealed four distinct clusters corresponding to the four interneuron types, with data points for different model types intermingled (Figure 8a). In contrast, the tSNE mapping of the looming object feature map demonstrated different model types clustering together (Figure 8b). This inverse clustering feature illustrated that the multiplication operation, the postprocessing step between the interneuron signals of the feature map and the optical flow map, underscores the performance boost in the combined-pathway model.

The multiplication operation proved beneficial as it enhanced selectivity for the looming object while mitigating interference noise from background movements (Figure 8b). The motion-pathway model, relying solely on T4 and T5 signals, faced challenges with a cluttered looming object map (LOM) due to background flows, despite the improvement from direction opponency. The feature-based model, dependent on T2 and T3 signals, presented a cleaner LOM but lacked motion direction selectivity, resulting in confusion with high-contrast background objects and reduced robustness against dynamic backgrounds. In contrast, the multiplication allowed the combined-pathway model to inherit advantages and circumvent drawbacks from the other two models. The looming signal pattern in the optic flow and feature maps exhibited significant correlation, and the multiplication enhanced the correlated looming signal while preserving motion direction opponency. Conversely, background motions and features produced sparse complementary interferences, which were mitigated through multiplication.

4. Conclusion

In the pursuit of detecting imminent collisions within dynamic real-world scenarios, the challenges are compounded by the intricate dynamics of moving backgrounds that obscure looming objects. Nevertheless, nature's wonders are evident as even tiny-brained insects respond swiftly to looming threats during high-speed flights. Despite advancements in bioinspired solutions, existing looming detection models, rooted in either motion or feature pathways, grapple with interference in dynamic scenes.

This study unveils a promising approach to elevate looming detection in dynamic backgrounds by harmoniously coordinating feature and motion detection pathways. Drawing inspiration from cutting-edge *Drosophila* neuroanatomy, we crafted an ANN model and honed its skills through various motion detection tasks. By creatively accepting synaptic inputs from motion and feature detection neurons, the looming detection branch dynamically coordinates their interneuronal signals through an electrifying multiplication process. This signal coordination empowers the combined-pathway model with superior and more robust looming detection performances across various training and testing datasets.

In summary, three key contributions emerge from our study: 1) Multiplication of interneuron signals: the seemingly simple yet potent mathematical operation—the multiplication of interneuron signals—elegantly equips the model to counter background disturbances. This synergy reduces interference from background noise, enhancing accuracy in looming estimations for realistic scenes featuring looming objects. The direct link between this mechanistic implementation and task performance underscores the strength of developing explainable artificial intelligence solutions inspired by biology. 2) Decoding continuous looming variables: unlike traditional bioinspired models that provide zero-one alarm signals, our ANN models decode continuous looming variables, such as looming size and speed, from local interneuron signals. Computing such continuous variables proves advantageous for subsequent collision avoidance control. Importantly, the decoding process relies on simple two-layer CNNs, implying the potential for real neural circuits to decode such variables, even if these decoding neurons have not been identified. 3) Parallel structure for diverse motion patterns: the model's success is further enriched by a parallel structure proficiently handling diverse motion patterns, including wide-field background motion, translating object motion, and looming motion patterns. Propelled by multiple objective functions, the optimization process considers various motion patterns simultaneously. This multiobjective optimization is crucial for the network to converge to biologically plausible solutions, emphasizing the importance of designing training tasks for bioinspired neural networks tailored to address real-world problems.

Finally, as an inherently neural network model reliant on training for performance enhancement, the efficacy and applicability of our models in real-world scenarios depend on the provided training data. While the model trained on the artificial

dataset exhibits some robustness when generalized to the AirSim simulated dataset with similar naturalistic background distributions, retraining is necessary for new data types, such as the ODA dataset. Nevertheless, the implications of the model, including pathway coordination, interneuronal multiplication, and the importance of multiobjective task training, hold general significance. Overall, this proposed looming detection model certifies the importance of pathway coordination for collision detection in the real world, affirming the benefits of bioinspiration from intricate neural circuitries. Concurrently, it also suggests testable hypotheses for pathway coordination in biological experiments.

5. Experimental Section

The Background Branch and the EMD: The background branch (blue branch in Figure 2a) processes local motion signals from T4 and T5 cells using neuroinspired mathematical operations to estimate the whole-field motion direction, which has a ground truth label. The difference between the estimation and the ground truth contributes to a multiobjective loss function that tunes the parameters for interneuron filtering profiles in the motion computation module.

The input is a preprocessed high-resolution image contrast signal $I_c(x,t)$. To mimic the low resolution of the *Drosophila* compound eye, $I_c(x,t)$ was convolved with a Gaussian-shaped receptive field whose full width half maximum and intervals are both 3°[22]

$$P(x,t) = \int I_c(x,t) \times G(x)dx \tag{1}$$

where $P(x,t)$ denotes the optic signals received by each ommatidium in the compound eye. We integrated this step in the process of generating the dataset rather than in the runtime, saving much redundant computation during training.

Following the columnized structure in *Drosophila*'s optic lobe, each pixel corresponds to a "column" structure in the laminar, where computations are mostly independent across columns (Figure 2b). Here, signals are filtered and rectified into light increments and decrements,[55] representing the well-known ON/OFF pathway split in the lamina. The computations in the ON and OFF pathways are independent. The mechanisms may be implemented differently in biology for these two pathways; we used identical mathematical operations for simplicities.

The signals from adjacent laminar columns are then filtered in the medulla and integrated to estimate the local motion directions in the lobula plate. We used an EMD model with three input arms: leading, central, and trailing edges,[25,37] based on recent neuroscience findings. These arms have varied temporal dynamics,[37] which delay and filter the signals to enhance or separate spatial–temporal correlations caused by motion. Typically, the leading- and trailing-edge signals are much delayed compared to the central ones

$$M^i_{ON}(x,t) = \left[\int P(x,s) \times A^i_{ON}(t-s)ds\right]^+$$
$$M^i_{OFF}(x,t) = \left[\int P(x,s) * A^i_{OFF}(t-s)ds\right]^+, i \in 1,2,3 \tag{2}$$

where M_{ON} and M_{OFF} are the ON/OFF pathways output signals. $A^i_{ON/OFF}, i \in 1,2,3$ are the medulla temporal filters for the three adjacent arms in ON/OFF pathways. $[\ \ ]^+$ represents the rectification process. For each of these arms, we also added a tanh nonlinearity to account for the contrast adaptation in medulla.

In our approach, we used the structure of the EMD model, but instead of manually setting up the medulla filters, we trained them to optimize the task. Previous studies have shown that training the medulla filters can lead to biologically meaningful results, even when guided by neural signals from only one T4 or T5 neuron.[38] In the final step of EMD, the signals from the three arms were combined to enhance the preferred direction and inhibit the nonpreferred direction along the local movement direction. Traditionally, this is achieved using a correlation or multiplication between the trailing and central edges, similar to the Hassenstein–Reichardt correlator model, followed by a division between the central and leading edges, similar to the Barlow–Levick model. However, recent research suggests that other nonlinear signal combinations can also work well for the three-input model.[56] To avoid training difficulties associated with division operations, we used subtraction to model the inhibition instead.

Figure 2b shows the basic idea for 1D motion detection. The leftward and rightward motion outputs can be formally expressed as follows

$$\begin{aligned} L_{ON}(x,t) &= F\left(M^1_{ON}(x+1,t), M^2_{ON}(x,t), M^3_{ON}(x-1,t)\right) \\ L_{OFF}(x,t) &= F\left(M^1_{OFF}(x+1,t), M^2_{OFF}(x,t), M^3_{OFF}(x-1,t)\right) \\ R_{ON}(x,t) &= F\left(M^1_{ON}(x-1,t), M^2_{ON}(x,t), M^3_{ON}(x+1,t)\right) \\ R_{OFF}(x,t) &= F\left(M^1_{OFF}(x-1,t), M^2_{OFF}(x,t), M^3_{OFF}(x+11,t)\right) \\ &\text{and} \quad F(x_1,x_2,x_3) = [x_1(x_2 - x_3)]^+ \end{aligned} \tag{3}$$

where F is the spatial correlation function, and $R_{ON/OFF}$, $L_{ON/OFF}$ is the rightward or leftward motion output from ON/OFF pathways.

After spatial correlation, the signals undergo another tanh nonlinearity to obtain a normalized direction estimation. To filter out noise, a spatial filter that covers three columns is applied. Finally, we subtracted the leftward and rightward outputs to obtain a signed value for direction estimation

$$\begin{aligned} T_4 &= \mathrm{Tanh}\left(\int B(x) \times (L_{ON}(x) - R_{ON}(x))dx\right) \\ T_5 &= \mathrm{Tanh}\left(\int B(x) \times (L_{OFF}(x) - L_{OFF}(x)dx)\right) \end{aligned} \tag{4}$$

where B stands for the spatial filter, and tanh stands for the tanh-function nonlinearity.

The Translating Object Detection Branch and the Moving Edge Sensitive Neurons: We added a new branch called the translating object detection branch, which helps train the feature-detection interneurons' filtering profiles and assists in looming object detection (yellow branch in Figure 2a). To estimate the moving speed of a translating object, we used a two-layer CNN that takes the translating object map as input. The translating object map is a feature map that marks potential objects.

To calculate the translating object map, we hypothesized that T2 and T3 neurons in the Locula play a role (Figure 2c). These neurons are sensitive to local high-contrast changes caused by moving objects, assisting small object feature detection. Moreover, they are not direction-selective, suggesting they are feature-detection neurons. Additionally, T2 and T3 neurons provide inputs to looming detection neurons like LPLC1.

We took inspiration from the STMD model for *Drosophila* LC18 neuron, a small moving object detection neuron that receives inputs from T2 and T3 neurons.[29] The most prominent feature of a translating object is a leading moving edge followed by a trailing moving edge with opposite contrast changes. However, these moving edges can be confused by background moving edges. To increase the model's selectivity, instead of directly calculating the correlation between the delayed leading edge and trailing edge signals as in the classical STMD model, we implemented an inhibition of inhibition mechanism, which proved particularly useful for selecting a translating object.[29]

The inhibition of inhibition means that T2 and T3 are both inhibited by their own and each other's signals (Figure 2d). T2 and T3 first inhibit themselves through a self-inhibition branch, which receives inhibition from the delayed counterpart's signal (denoted by $T_2^{inhibit}$, $T_3^{inhibit}$). When a translating object is present, the leading edge's signal is delayed to inhibit the inhibition branch of the trailing edge signal, resulting in a prominent signal. On the other hand, when no trailing edge follows, T2 and T3 signals are completely inhibited by themselves, preventing the model's responses to single moving edges in the background. The outputs from the inhibition of the inhibition mechanism are then compressed using a tanh nonlinearity to form a small object map (SOM), representing the traces of the object's moving edges:

$$\begin{aligned}\mathrm{SOM}_{\mathrm{ON}}(x,t) &= \mathrm{Tanh}\left(\left[T_2 - \left[T_2 - T_3^{\mathrm{inhibit}}\right]^+\right]^+\right)\\ \mathrm{SOM}_{\mathrm{OFF}}(x,t) &= \mathrm{Tanh}\left(\left[T_2 - \left[T_2 - T_3^{\mathrm{inhibit}}\right]^+\right]^+\right)\end{aligned} \tag{5}$$

$$\begin{aligned}T_2^{\mathrm{inhibit}} &= \int C(t-s) \times T_2(s)ds\\ T_3^{\mathrm{inhibit}} &= \int C(t-s) \times T_3(s)ds\end{aligned} \tag{6}$$

where SOM_i stands for SOM. The parameters x and t for T_2 and T_3 are omitted for simplification. C is the temporal filter to delay T_2 and T_3 signals.

Note that T2 and T3 pathways have two SOMs for the ON/OFF moving target traces, respectively. A simple competition mechanism is adopted to select the maximum value between the two SOMs

$$\mathrm{SOM}(x,t) = \max(\mathrm{SOM}_{\mathrm{ON}}(x,t), \mathrm{SOM}_{\mathrm{OFF}}(x,t)) \tag{7}$$

Upper in the pathways, T2 and T3 neurons receive presynaptic signals from fast-responding neurons Mi3, etc., in the medulla. Thus, we model T2 and T3's function by temporal filtering medulla neuron outputs

$$\begin{aligned}T_2(x,t) &= \left[\int D_{\mathrm{ON}}(t-s) \times M_{\mathrm{ON}}^2(s)ds\right]^+\\ T_3(x,t) &= \left[\int D_{\mathrm{OFF}}(t-s) \times M_{\mathrm{OFF}}^2(s)ds\right]^+\end{aligned} \tag{8}$$

where $D_{\mathrm{ON/OFF}}$ is the temporal filter ahead of T2 and T3 cells.

The Looming Object Detection Branch: The looming detection branch is a key aspect of our article (orange branch in Figure 2a). We created a LOM, marking potential looming objects based on inputs from the motion detection and/or feature detection pathways. This leads to three different looming detection models: one using only motion detection inputs, another with only edge feature detection inputs, and a third model with inputs from both pathways, called the combined-pathway model. In the combined-pathway model, we calculated the LOM by multiplying the interneuron's signals, which are the optic flow map of T4/T5 signals and the edge feature map composed of spatial–temporal filtered T2/T3 signals. In the motion-pathway model, the LOM is the same as the optic flow map, while in the feature-pathway model, it reduces to the edge feature map. The multiplication could be reinforcing and exclusive depending on whether the signals correlate. The multiplication can reinforce the looming signals, which are the dominant feature in both the optic flow map and the edge feature map and are highly correlated. However, the unwanted signals from the background interference can be represented differently by the motion or the feature pathway, composing uncorrelated noise features in the optic flow map and the feature map, which the multiplication can suppress. In this way, the multiplication enhanced the model's looming object detection capability while suppressing the disturbances in a dynamic background.

$$\begin{aligned}\mathrm{LOM}_{\mathrm{ON}}(x,t) &= \mathrm{Tanh}\left(T_4 \int E(x-u,t-v) \times T_2(u,v)dudv\right)\\ \mathrm{LOM}_{\mathrm{OFF}}(x,t) &= \mathrm{Tanh}\left(T_5 \int E(x-u,t-v) \times T_3(u,v)dudv\right)\end{aligned} \tag{9}$$

where LOM_i stands for LOM. The two channels LOM are also compressed into one channel by competition

$$\mathrm{LOM}(x,t) = \max(\mathrm{LOM}_{\mathrm{ON}}, \mathrm{LOM}_{\mathrm{OFF}}) \tag{10}$$

Unlike standard looming detection models that only give binary outputs (zero or one) to signal the presence of a looming object, we utilized two two-layer convolution networks to map the LOM into the size and speed of the looming object, respectively. There are two reasons for this approach: 1) An alarming model is limited in its predictive capabilities. Existing evidence suggests that individual neurons can encode information about looming size or speed, indicating that animals might perform more complex computations than just alarming. 2) It is relatively straightforward to design or train a network to achieve high accuracy as a binary classifier (zero or one). However, in such cases, it becomes challenging to demonstrate the additional benefits of combining motion and feature detection pathways for looming detection.

However, we must emphasize that apart from implementing four separate channels to account for the various object sizes, the convolution network was intentionally set simple with only a hidden layer to avoid taking over the roles of pattern recognition from the bioinspired models (Figure 2e).

Finally, all filters for interneurons, including *A*, *B*, *C*, *D*, and *E*, as well as filters and CNNs in three separate branches, have free parameters to be optimized. It is essential to note that they are not initialized to function according to their names. Our objective was to demonstrate that by combining optimization tasks and an anatomy-constrained network, we can achieve a network that exhibits neural dynamics similar to their biological counterparts, specifically, T2, T3, T4, and T5 neurons. We also aimed to show that our network can converge to a biological solution by hypothetically multiplying the motion and feature detection signals, validating our approach to model the coordination between the two pathways.

Artificial Dataset: We chose 241 high-quality panoramic high dynamic range images from the original set of 421 images to serve as backgrounds in our video sequences.[57] In simulating ego motion, we assumed the insect to remain stationary while the entire background moves, mimicking the motion as if the insect is turning. The background's movement speed is determined either by a uniform value from a slow speed distribution ($33 \pm 5\ \mathrm{s}^{-1}$) or a fast speed distribution ($75 \pm 5\ \mathrm{s}^{-1}$), or a stochastic speed time sequence generated from various speed distributions. The process to generate the speed sequence follows the approach outlined in ref. [38]: sampling from a Gaussian distribution over time and then low-pass filtering the sequence to achieve a reasonable autocorrelation half-life of 200 ms. Our speed distributions have a mean of $0\ \mathrm{s}^{-1}$ and a standard deviation of either 40 or $80\ \mathrm{s}^{-1}$, representing slow or fast stochastic movements.

We manually introduced moving objects in the foreground, encompassing both translating and looming objects, with the largest object size covering a visual field of 90°. To conserve computational resources, we downsampled the high-resolution images using spatial filters to match the resolution of a fly compound eye. Focusing on 1D scenarios, we stimulated the model with the movement of a line of pixels, lasting for 2 s and sampled at 100 Hz, resulting in a 2D spatial–temporal pattern with 72×200 data points (see Figure 3a right).

To mimic the wide range of contrast distributions seen in natural objects, the contrast distribution of the added objects is more diverse than that of the background (**Figure 9**). Then the Weber contrast of the normalized background scene is slightly biased with a zero mean and a standard variation.[58]

To assess the robustness of looming detection models across various dynamic scenes, we varied the statistical properties of both background and foreground objects in the artificial datasets, generating 2886 training samples and 972 testing samples. We considered: 1) Static backgrounds

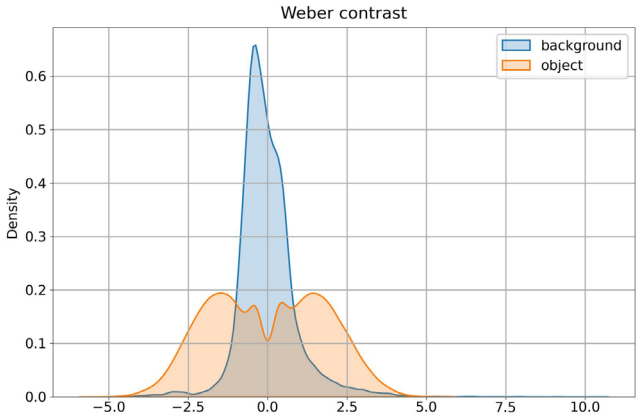

Figure 9. Weber contrast distribution of the background scenes and manually added object.

as a control and dynamic backgrounds with different speeds, including stochastic and uniform speeds at slow or fast rates. 2) Foreground translating objects with randomly generated sizes (10°–20°) moving leftward or rightward against the background. 3) Looming objects approaching with a uniform speed, resulting in a highly nonlinear expanding extended angle on the retina (Figure 3b). 4) Looming objects with either uniform or sinewave pattern contrast, where patterned looming objects presented a greater challenge for detection. 5) Carefully controlled object contrast to mimic natural predators, where predators from bright skies or dark caves may have distinct contrasts against the background. The contrast distribution of added objects is more diverse than that of the background, with the object's Weber contrast exhibiting about three times larger variation (see Figure 9). The Weber contrast of the normalized background scene is slightly biased with a zero mean and a standard variation.[58]

The Weber contrast is defined as

$$I_c = \frac{I - I_{\text{mean}}}{I_{\text{mean}}} \tag{11}$$

ODA Dataset: The TUDelft ODA dataset is a publicly accessible collection of data created by the MAVLab group. This dataset was recorded in an indoor laboratory for tasks related to detecting and avoiding obstacles using UAVs. It consists of 1369 samples, capturing information from sensors on the UAV and the OptiTrack motion capture system. As the original dataset lacks ground truth labels for looming size and speed, we generated these labels by analyzing video data from an onboard camera, obstacle positions, and UAV pose measurements from the OptiTrack system. Here is a simplified breakdown of the calibration process: 1) Data selection: Removed samples with undesired conditions (e.g., dim lighting, multiple obstacles, OptiTrack system failures). After exclusion, 370 samples remained. 2) Coordinate transformation and calibration: Transformed obstacle location from world coordinates to UAV camera coordinates, calibrated looming size using azimuth direction of obstacle edges, and calculated looming speed from temporal differences in looming size after low-pass filtering to minimize measurement noise. 3) Handling object disappearance: Adjusted looming size and speed to zero when the orange obstacle moved out of the UAV's view. We determined obstacle disappearance time by monitoring color changes (from orange to other colors) in the Red, Green and Blue (RGB) frame's left or right patch. 4) Temporal alignment: Addressed synchronization issues between video and OptiTrack data, aligned looming labels and RGB frames at the obstacle disappearance time, and interpolated looming labels at RGB timestamps. 5) Standardization of samples: Unified all training samples to the same length by dividing them into segments. Each segment consists of 80 frames with a 30-frame overlap with adjacent segments.

Moreover, to augment background complexities within the ODA dataset, we incorporated low-amplitude Gaussian white noise to replicate the inherent noise found in natural environments. Additionally, we introduced miniature white blocks, making them move uniformly or divergently to emulate dynamic background interferences encountered in intricate surroundings (see Figure 3c).

Optimization and Evaluation: Here, we detailed the optimization and evaluation procedures for training on the artificial dataset. The procedure for training on the ODA dataset follows similarly.

Loss Function: We trained our model in a supervised way to optimize a multiobjective loss function that considers the background moving direction, the looming object size, the looming object speed, and the translating object speed. To ensure unbiased training, we sampled the same number of datasets with looming objects and translating objects. We used mean square error (MSE) as the training loss in optimization and *R*-square to evaluate models' performances. The loss function is defined as

$$L = \frac{\alpha_1}{N}\sum \left(\hat{Y}_{\text{background}} - Y_{\text{background}}\right)^2 + \tag{12}$$

$$\frac{\alpha_2}{N}\sum \left(\hat{Y}_{\text{Lsize}} - Y_{\text{Lsize}}\right)^2 + \tag{13}$$

$$\frac{\alpha_3}{N}\sum \left(\hat{Y}_{\text{Lspeed}} - Y_{\text{Lspeed}}\right)^2 + \tag{14}$$

$$\frac{\alpha_4}{N}\sum \left(\hat{Y}_{\text{Tspeed}} - Y_{\text{Tspeed}}\right)^2 \tag{15}$$

where $\alpha_j, j \in 1, 2, 3, 4$ are the weights for the four tasks; $Y_{\text{background}}$ is the background motion direction formulated by values between −1 and 1; Y_{Lsize} and Y_{Lspeed} are the size and speed of the looming object; Y_{Tspeed} is the speed of the translating object; and the hat above each Y stands for the model estimations.

Evaluation: The evaluation *R*-square is defined by

$$E\left(Y, \hat{Y}\right) = 1 - \frac{\sum_{i=0}^{N}\left(\hat{Y}_i - \overline{Y}_i\right)^2}{\sum_{i=0}^{N}\left(Y_i - \overline{Y}_i\right)^2} \tag{16}$$

where E is the evaluation for the looming size estimation; Y_i is the label and $\hat{Y}_i$ is the model estimation; and $\overline{Y}_i$ is the mean value of Y_i. The E value is upbounded by 1, and the larger its value, the better the model in estimating the labeled values.

Optimization: Our training process comprises two stages for experiments on an artificial dataset. Initially, we focused on training the background and the translating object detection branch. In the subsequent stage, we trained all three branches. During the first stage, the training involved two branches that integrated the T4/T5 and T2/T3 signals separately, thereby reinforcing the interneuron function to emulate their biological counterparts. In this stage, we set the parameters α_j, where j ranges from 1 to 4, as [50, 0, 0, 1].

Moving to the second stage of training, we adjusted the parameters α_j differently for each model. For the combined-pathway model, they were set as [100, 1, 1, 1], for the feature-pathway model as [50, 1, 1, 3], and for the motion-pathway model as [50, 1, 1, 1]. These parameter variations account for the sensitivities of interneuron signals in each model. The three models converged after ≈4000 iterations in the first training stage and another 4000 in the second training stage.

Specifically designed only for the looming object detection task, the CLGMD and LPLC2 models exhibited faster convergence, achieving it after only about 1000 iterations. Throughout all training stages, the models received input data with a batch size of 64 and were optimized using the Adam optimizer with a learning rate of 1e−3.

MSE loss and *R*-square evaluation were also employed for experiments on the ODA dataset. In this case, all five models received input data with a batch size of 4 and were optimized by Adam with a learning rate of 1e−3, converging after ≈20 000 iterations.

Acknowledgements

The authors thank Prof. Marion Sillies, Prof. Shigang Yue, and Dr. Yu Zhou for valuable feedback on the initial drafts of the manuscripts. Project was supported by the Young Scientists Fund of the National Natural Science Foundation of China (grant no. 12001111), Fund from the National Natural Science Foundation of China (grant no. 11925103), Shanghai Municipal Science and Technology Major Project (grant no. 2018SHZDZX01), ZJ Lab, and Shanghai Center for Brain Science and Brain-Inspired Technology, the 111 Project (grant no. B18015), and the 2021 STCSM (grant no. 2021SHZDZX0103).

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

B.G. and Z.S. designed the study. B.G. programmed the model and ran the simulations, with supervision from Z.S. Z.S. wrote the paper with an initial draft from B.G. and editing from everybody. B.G. and Z.S. analyzed the results and designed the figures. J.F.F. advised the project.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

Drosophila neural circuits, insect inspired neural computations, looming detection

Received: March 13, 2024
Revised: May 11, 2024
Published online:

[1] F. T. Muijres, M. J. Elzinga, J. M. Melis, M. H. Dickinson, *Science* **2014**, *344*, 172.
[2] S. Yue, F. C. Rind, *IEEE Trans. Neural Netw.* **2006**, *17*, 705.
[3] Q. Fu, C. Hu, J. Peng, F. C. Rind, S. Yue, *IEEE Trans. Cybern.* **2019**, *50*, 5074.
[4] R. Strydom, A. Denuelle, M. V. Srinivasan, *Aerospace* **2016**, *3*, 21.
[5] J. R. Serres, F. Ruffier, *Development* **2017**, *46*, 703.
[6] L. Salt, G. Indiveri, Y. Sandamirskaya, in *2017 IEEE Int. Symp. Circ. Syst. (ISCAS)*, IEEE, Baltimore, MD **2017**, pp. 1–4.
[7] Q. Fu, Z. Li, J. Peng, *Array* **2023**, *17*, 100272.
[8] F. Shuang, Y. Zhu, Y. Xie, L. Zhao, Q. Xie, J. Zhao, S. Yue, *arXiv:2302.10284* **2023**.
[9] N. Hatsopoulos, F. Gabbiani, G. Laurent, *Science* **1995**, *270*, 1000.
[10] F. C. Rind, D. Bramwell, *J. Neurophysiol.* **1996**, *75*, 967.
[11] F. C. Rind, S. Wernitznig, P. Pölt, A. Zankel, D. Gütl, J. Sztarker, G. Leitinger, *Sci. Rep.* **2016**, *6*, 35525.
[12] M. S. Maisak, J. Haag, G. Ammer, E. Serbe, M. Meier, A. Leonhardt, T. Schilling, A. Bahl, G. M. Rubin, A. Nern, B. J. Dickson, D. F. Reiff, E. Hopp, A. Borst, *Nature* **2013**, *500*, 212.
[13] Q. Fu, C. Hu, J. Peng, S. Yue, *Neural Netw.* **2018**, *106*, 127.
[14] F. Lei, Z. Peng, M. Liu, J. Peng, V. Cutsuridis, S. Yue, *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 8362.
[15] M. Hua, Q. Fu, J. Peng, S. Yue, H. Luan, in *2022 Int. Joint Conf. Neural Netw. (IJCNN)*. IEEE, Padua, Italy **2022**, pp. 1–8.
[16] J. Zhao, H. Wang, N. Bellotto, C. Hu, J. Peng, S. Yue, *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *34*, 2539.
[17] S. B. i Badia, P. F. Verschure, in *2004 IEEE Int. Joint Conf. Neural Netw. (IEEE Cat. No. 04CH37541)*. IEEE, Budapest, Hungary **2004**, pp. 1757–1761.
[18] T. Schilling, A. Borst, *Biol. Open* **2015**, *4*, 1105.
[19] N. C. Klapoetke, A. Nern, M. Y. Peek, E. M. Rogers, P. Breads, G. M. Rubin, M. B. Reiser, G. M. Card, *Nature* **2017**, *551*, 237.
[20] J. Zhao, S. Xi, Y. Li, A. Guo, Z. Wu, *iScience* **2023**, *26*, 106337.
[21] R. Tanaka, D. A. Clark, *Curr. Biol.* **2022**, *32*, 2357.
[22] A. Borst, J. Haag, D. F. Reiff, *Annu. Rev. Neurosci.* **2010**, *33*, 49.
[23] A. I. Meinertzhagen, *Prog. Retinal Res.* **1993**, *12*, 13.
[24] K. F. Fischbach, A. Dittrich, *Cell Tissue Res.* **1989**, *258*, 441.
[25] A. Borst, J. Haag, A. S. Mauss, *J. Comp. Physiol. A* **2020**, *206*, 109.
[26] E. Gruntman, S. Romani, M. B. Reiser, *Elife* **2019**, *8*, e50706.
[27] M. F. Keles, M. A. Frye, *Curr. Biol.* **2017**, *27*, 680.
[28] M. Wu, A. Nern, W. R. Williamson, M. M. Morimoto, M. B. Reiser, G. M. Card, G. M. Rubin, *Elife* **2016**, *5*, e21022.
[29] N. C. Klapoetke, A. Nern, E. M. Rogers, G. M. Rubin, M. B. Reiser, G. M. Card, *Neuron* **2022**, *110*, 1700.
[30] K. Nordström, P. D. Barnett, D. C. O'Carroll, *PLoS Biol.* **2006**, *4*, e54.
[31] B. Zhou, Z. Li, S. Kim, J. Lafferty, D. A. Clark, *Elife* **2022**, *11*, e72067.
[32] M. F. Keleş, B. J. Hardcastle, C. Städele, Q. Xiao, M. A. Frye, *Cell Rep.* **2020**, *30*, 2115.
[33] R. Tanaka, D. A. Clark, *Curr. Biol.* **2020**, *30*, 2532.
[34] H. J. Dahmen, M. O. Franz, H. G. Krapp, in *Extracting Egomotion from Optic Flow: Limits of Accuracy and Neural Matched Filters*, Springer, Berlin, Heidelberg **2001**.
[35] R. Behnia, D. A. Clark, A. G. Carter, T. R. Clandinin, C. Desplan, *Nature* **2014**, *512*, 427.
[36] J. C. S. Leong, J. J. Esch, B. Poole, S. Ganguli, T. R. Clandinin, *J. Neurosci.* **2016**, *36*, 8078.
[37] A. Arenz, M. S. Drews, F. G. Richter, G. Ammer, A. Borst, *Curr. Biol.* **2017**, *27*, 929.
[38] O. Mano, M. S. Creamer, B. A. Badwan, D. A. Clark, *Curr. Biol.* **2021**, *31*, 4062.
[39] S. Yue, F. C. Rind, *Comput. Vision Image Understanding* **2006**, *104*, 48.
[40] S. Yue, F. C. Rind, *IEEE Trans. Auton. Mental Dev.* **2013**, *5*, 173.
[41] Q. Fu, S. Yue, *Biol. Cybern.* **2020**, *114*, 443.
[42] Q. Fu, S. Yue, in *2020 5th Int. Conf. Adv. Robot. Mechatron.* IEEE, Shenzhen, China **2020**, pp. 609–615.
[43] F. Gabbiani, H. G. Krapp, G. Laurent, *J. Neurosci.* **1999**, *19*, 1122.
[44] M. Y. Peek, G. M. Card, *Curr. Opin. Neurobiol.* **2016**, *41*, 167.
[45] K. Nordström, *Curr. Opin. Neurobiol.* **2012**, *22*, 272.
[46] R. Palavalli-Nettimi, J. Theobald, *Curr. Biol.* **2020**, *30*, R761.
[47] S. D. Wiederman, P. A. Shoemaker, D. C. O'Carroll, *J. Neurosci.* **2013**, *33*, 13225.
[48] B. R. Cowley, A. J. Calhoun, N. Rangarajan, E. Ireland, M. H. Turner, J. W. Pillow, M. Murthy, *Nature* **2024**, *629*, 1100–1108.
[49] V. Lobato-Rios, S. T. Ramalingasetty, P. G. Özdil, J. Arreguit, A. J. Ijspeert, P. Ramdya, *Nat. Methods* **2022**, *19*, 620.
[50] A. Zador, S. Escola, B. Richards, B. Ölveczky, Y. Bengio, K. Boahen, M. Botvinick, D. Chklovskii, A. Churchland, C. Clopath, J. DiCarlo, S. Ganguli, J. Hawkins, K. Körding, A. Koulakov, Y. LeCun, T. Lillicrap, A. Marblestone, B. Olshausen, A. Pouget, C. Savin, T. Sejnowski, E. Simoncelli, S. Solla, D. Sussillo, A. S. Tolias, D. Tsao, *Nat. Commun.* **2023**, *14*, 1597.
[51] M. S. Drews, A. Leonhardt, N. Pirogova, F. G. Richter, A. Schuetzenberger, L. Braun, E. Serbe, A. Borst, *Curr. Biol.* **2020**, *30*, 209.
[52] J. Dupeyroux, R. Dinauxd, N. Wessendorp, G. De Croon, in *System Engineering for Constrained Embedded Systems, DroneSE and RAPIDO*, Association for Computing Machinery, New York, NY **2022**, pp. 8–13.
[53] A. Loquercio, A. I. Maqueda, C. R. Del-Blanco, D. Scaramuzza, *IEEE Robot. Autom. Lett.* **2018**, *3*, 1088.
[54] S. Shah, D. Dey, C. Lovett, A. Kapoor, in *Field and Service Robotics: Results of the 11th Inter. Conf.*, Springer, Zurich, Switzerland **2018**, pp. 621–635.
[55] H. H. Yang, F. St-Pierre, X. Sun, X. Ding, M. Z. Lin, T. R. Clandinin, *Cell* **2016**, *166*, 245.
[56] J. E. Fitzgerald, D. A. Clark, *Elife* **2015**, *4*, e09123.
[57] H. G. Meyer, M. Egelhaaf, A. K. Warzecha, *Front. Physiol.* **2013**, *4*, 155.
[58] E. Peli, *J. Opt. Soc. Am. A* **1990**, *7*, 2032.