



Article

A Novel Artificial Visual System for Motion Direction Detection in Grayscale Images

Sichen Tao ¹, Yuki Todo ^{2,*}, Zheng Tang ^{1,*}, Bin Li ², Zhiming Zhang ¹ and Riku Inoue ²¹ Faculty of Engineering, University of Toyama, Toyama-shi 930-8555, Japan² Faculty of Electrical and Computer Engineering, Kanazawa University, Kanazawa-shi 920-1192, Japan

* Correspondence: yktodo@ec.t.kanazawa-u.ac.jp (Y.T.); ztang@eng.u-toyama.ac.jp (Z.T.)

Abstract: How specific features of the environment are represented in the mammalian brain is an important unexplained mystery in neuroscience. Visual information is considered to be captured most preferentially by the brain. As one of the visual information elements, motion direction in the receptive field is thought to be collected already at the retinal direction-selective ganglion cell (DSGC) layer. However, knowledge of direction-selective (DS) mechanisms in the retina has remained only at a cellular level, and there is a lack of complete direction-sensitivity understanding in the visual system. Previous studies of DS models have been limited to the stage of one-dimensional black-and-white (binary) images or still lack biological rationality. In this paper, we innovatively propose a two-dimensional, eight-directional motion direction detection mechanism for grayscale images called the artificial visual system (AVS). The structure and neuronal functions of this mechanism are highly faithful to neuroscientific perceptions of the mammalian retinal DS pathway, and thus highly biologically reasonable. In particular, by introducing the horizontal contact pathway provided by horizontal cells (HCs) in the retinal inner nuclear layer and forming a functional collaboration with bipolar cells (BCs), the limitation that previous DS models can only recognize object motion directions in binary images is overcome; the proposed model can solve the recognizing problem of object motion directions in grayscale images. Through computer simulation experiments, we verified that AVS is effective and has high detection accuracy, and it is not affected by the shape, size, and location of objects in the receptive field. Its excellent noise immunity was also verified by adding multiple types of noise to the experimental data set. Compared to a classical convolutional neural network (CNN), it was verified that AVS is completely significantly better in terms of effectiveness and noise immunity, and has various advantages such as high interpretability, no need for learning, and easy hardware implementation. In addition, activation characteristics of neurons in AVS are highly consistent with those real in the retinal DS pathway, with strong neurofunctional similarity and brain-like superiority. Moreover, AVS will also provide a novel perspective and approach to understanding and analyzing mechanisms as well as principles of mammalian retinal direction-sensitivity in face of a cognitive bottleneck on the DS pathway that has persisted for nearly 60 years.

Keywords: artificial visual system; motion direction detection; direction-selective ganglion cell; retinal direction-selective pathway

MSC: 68T07



Citation: Tao, S.; Todo, Y.; Tang, Z.; Li, B.; Zhang, Z.; Inoue, R. A Novel Artificial Visual System for Motion Direction Detection in Grayscale Images. *Mathematics* **2022**, *10*, 2975. <https://doi.org/10.3390/math10162975>

Academic Editor: Zhaobin Wang

Received: 3 August 2022

Accepted: 16 August 2022

Published: 17 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The human brain optimizes and governs the next activities of the human body by processing inputs from nature and actively accesses specific external information needed through learning, creating a continuous closed loop of positive feedback with significant importance for biological activity. The brain is an excellent complex system, composed of approximately 10^{11} neurons with more than 10^{15} connections between them, continually creating complex patterns. Collective actions of single nerve cells linked by a dense web

of intricate connectivity guide behaviors, shape thoughts, form and retrieve memories, organized into the brain; thus, behaviors, thoughts, memories, and consciousness become possible. Understanding these integrative functions of the brain requires an understanding of neural networks in the brain and the complex dynamic patterns they create [1]. How specific features of the environment are represented in the brain is an important unsolved mystery in neuroscience. The visual system in the brain assists in processing more than 80% of the external information and is an extremely important branch of the brain neural networks. Visual information is thought to be captured and utilized more preferentially by brain activity [2,3]. Therefore, the study of the visual system is widely considered a sure way to reveal the mystery of the brain.

The elements of visual information are light, dark, and change. Color brings light and dark contrast while highlighting shape distinctions; the temporal dimension brings their change. Thus color perception, shape perception, and direction perception are the basic functions of the visual system [4,5]. A striking array of highly organized laminar neural circuits exists in the mammalian retina for the detection, processing, and extraction of these critical visual features. In 1963, Barlow and Hill first identified direction-selective ganglion cells (DSGCs) in the rabbit retina which assist the recognition of the global motion direction, and the circuits dedicated to direction detection were uncovered. The global motion direction is the motion direction in which mostly occurs in the global receptive field. It is important information that can be used to indicate the most obvious changes in the environment. Compared to the color and shape, the direction is a higher dimensional information perception function because of the addition of a temporal dimension. Such higher dimensional visual information is already extracted by the retina at the very beginning of the visual system before it enters the visual cortex, which has led to a long-term study and exploration of the function of the highly organized laminar structure of the retina [6,7].

Various in vivo and in vitro studies have been performed on DSGCs in the mammalian retina. Experiments in mice and rabbits have shown that DSGCs are present in a considerable proportion of the cells in the output layer of the retina, i.e., retina ganglion cells [8–10]. However, as with various other neural systems in the brain, a single nerve cell alone cannot perform any of even slightly more complex visual tasks, including DSGC.

The frontal neural pathway of DSGCs starts from the initial light input layer and runs through the entire retinal layers. This highly distinctive functional pathway is also known as the “DSGC vertical pathway”. It includes photoreceptor cells (PCs) in the outer nuclear layer (ONL), horizontal cells (HCs), bipolar cells (BCs), and amacrine cells (ACs) in the inner nuclear layer, and DSGCs in the ganglion cell layer [11]. In the vertical pathway, layers form interactable connections through synapses, and after completing their respective cellular functions, share the output signals successively to neighboring cells connected through synapses thus reaching intercellular information sharing and cooperation and constituting a functional vertical pathway. Photoreceptors do not respond specifically to motion phenomena in the global receptive field, but DSGCs downstream of the retina exhibit a selection of different directions. What neural computational processes exist between them, and how is this computation implemented in terms of neural circuits and membrane biophysics? This question has become a classic example of neural computation, which has been the focus of intensive research for decades and has attracted a large number of researchers from different fields.

Mammalian retinal circuits that confer unique tuning characteristics to DSGCs have been extensively investigated [5,7,12]. Several related models of motion detectors have been proposed in the past to calculate the direction of motion from the changes in luminance captured by the photoreceptors. Srinivasan proposed a gradient model by using a high-pass filter and calculating the spatial gradient of the luminance information received by a local neighboring receptor and the temporal gradient through the distance between them and dividing the two gradient values. A velocity estimate that excludes the spatial structure interference of motion patterns is obtained [13]. However, although it avoids the interference of luminance differences in adjacent spaces on the detection estimates by

physical concepts and mathematics methods, making it looks attractive, most of the proposed models do not involve the computational process of spatial and temporal gradients in motion images. Instead, these models directly or indirectly invalidate the motion in other directions by exploiting the temporal differences of neighboring receptors in space, thus yielding cleverly correct direction detection results without complex operations. The pioneer of correlation detectors was proposed by Hassenstein et al. based on experimental studies of insects' optokinetic behavior [14–16]. This type of correlation detector is often referred to as a Reichardt detector [17]. It consists of two mirrored subunits that multiply each of the two inputs processed by the temporal low-pass filter with the adjacent input before processing, and then the difference between the two is used as the output. The Reichardt detector has also been used to explain motion detection mechanisms in various vertebrates, including humans [18]. The simplest correlation detector was proposed by Barlow and Levick, and it is the first reasonable explanatory model for this experiment since the discovery of DSGCs in the rabbit retina. The model takes one of two inputs carrying luminance information and, after time delaying it, interacts with the other in an inhibitory manner. The input with time delay in the nonpreferred direction is activated first, and the delayed inhibitory signal arrives just as the adjacent input is activated and denies the excitatory output of the adjacent end; conversely, the input without time delay in the preferred direction is activated first and pre-empts the output before the inhibitory signal is generated at the other end, and the inhibitory effect fails while completing the task of outputting the preferred motion signal. Structurally, the model is equivalent to a subunit of the Reichardt model. In addition, another mechanism known as the motion energy model is often applied to human psychophysics and motion-sensitive neurons in the mammalian cortex [19]. Interestingly, even if the internal structures are different when the Reichardt model is equipped with the same temporal and spatial filters at the input, it presents at the output a functional characteristic equivalent to the motion energy model, and the process is mathematically equivalent.

Directional information is not explicitly encoded at the level of a single photoreceptor. Instead, it must be computed from at least two photoreceptors at the level of spatiotemporal excitation. Thus, although these motion direction detection models are not designed in the same way, there is a necessary consistency in their underlying principles, which can be summarized in three elements:

1. the existence of at least two spatially separated light intensity detectors;
2. the existence of asymmetry in the temporal gradient based on spatial separation;
3. the existence of at least one nonlinear computational unit.

Over the past two decades, thanks to advances in optical and genetic methods, we have made great progress in understanding direction-selective cellular and subcellular mechanisms. This includes the understanding of the presence of dendrites in direction-selective (DS) circuits, asymmetric structures [20–24] and the understanding of the non-negligibility of dendritic internal nonlinear computations exhibited at the ultrastructural level [25,26]. Moreover, it is a surprising fact that researchers have been enthusiastically discussing and working for almost 60 years on how the visual system accomplishes directional selection in a single direction to reach the current level of cognition. This problem is still far from the complex intelligence of humans in terms of computation and information processing. The existing understanding of motion direction selectivity remains at the level of cellular functions of various cellular units in the visual pathway. Meanwhile, existing proposed motion direction detectors based on the mammalian visual system can only perform the directional selection task for a single point in a black-and-white image in a single-dimensional direction. There is no doubt that the neural computational principles of directional selection in the mammalian visual system remain unknown. As with other neural mechanisms in the brain, the goal of revealing the principles through traditional biological experiments alone seems to be out of reach. Therefore, we have taken a slightly different approach to this problem, i.e., if we can use the cellular functions in the neural pathway to construct a biologically sound model of the neural computational pathway

based on the known intercellular structure, so that it can have the same two-dimensional global direction-selectivity as the mammalian visual system, with the best result that its output state is similar to the biological experiments, it will provide a powerful clue to our ultimate understanding of direction-selective computation in the visual system. To this end, several new studies have recently made similar efforts. Han et al. modeled the inhibitory model conjecture proposed by Barlow et al. [27]. In their work, the previous one-dimensional approach was successfully extended to an efficient two-dimensional, eight-way mechanism for direction detection in a black-and-white (binary) environment after exploiting the light-sensitive properties of PC with the time-delayed action of HC. However, in this model, the function of BC, which is a mandatory pathway of the DS circuit, fails to be considered. Therefore, the model is only able to identify the direction of object motion in a black-and-white (binary) environment. Thereafter, Tang et al. modeled a postsynaptic dendritic DSGC model conjecture proposed by Taylor et al. [28]. This model is a more complete development of the proposed circuit based on the Han et al. model. It eliminates the HC from the Barlow model and introduces the BC and the AC at the back end of the BC; this is the first time that the ON/OFF response mechanism of the BC appears in the direction detection model. The model utilizes the structural features of AC to interfere with BC through continuous forward inhibition to achieve the inhibition of null direction motion, thus assisting PC and BC to identify the direction of motion. However, the model lacks the functional properties underlying the sensitivity of BC cells to light changes in biological characteristics in terms of the definition of ON response. At the same time, the lateral regulation function of HC cells at the initial stage is skipped, which may result in the DS pathway losing a part of its performance by loss of a necessary opportunity for information collation, such as the ability to detect the direction of object motion in rich color environments or noise immunity in various complex environments. In addition, despite the new introduction of AC and BC into the model, it still only solves the direction selection problem in black-and-white (binary) environments. Thus, although these recent advances in modeling the DS pathway have been quite novel and productive, and have had a positive impact on the understanding of the retinal DS pathway, these models still fail to approach the real retinal directional selection principles.

In this paper, we address the aforementioned dilemmas by proposing an innovative motion direction detection mechanism as the Artificial Vision System (AVS) that is in close agreement with the biological knowledge of the DS pathway in terms of cellular function, model structure, overall function and performance of the implemented model. Based on the existing understanding of the biology and cellular physiology of the DSGCs vertical pathway in the mammalian retina, AVS uses a dendritic neuron model to implement a two-dimensional, eight-way motion direction detection mechanism that can work efficiently in a grayscale environment. It uses PCs for light signal perception; the concept of local receptive fields to delineate the photoreceptive range of a single DSGC; defines local motion direction detection neurons to collect information from local receptive fields; defines global motion direction detection neurons (GMDNs) to collect the output of local motion direction detection neurons (LMDNs) and infer the global motion direction. Relative to previous studies:

1. We innovatively introduced both BCs and HCs;
2. We faithfully defined their original biological properties and connection structures;
3. Because not necessary to further rely on any other cell function to complete the motion direction detection task, we temporarily eliminated the AC located in the posterior inner plexiform layer (IPL) considering that it can be used in a non-basic more advanced information integration mechanism such as being a post filter to help improve the noise immunity;
4. We achieved a breakthrough in its limitations of black-and-white (binary) detection without adding any non-essential cellular structure and constructed a set of DS pathway model for grayscale images that are faithful, reasonable, and complete in both cellular function and model structure.

We implemented AVS in computer simulations. In grayscale images motion direction detection experiments with no noise and multiple signal-to-noise ratio noises, we used a series of instances of objects with random shapes and different sizes moving in eight directions and compared the activation phenomena with the prior biological experiments. The experimental results show that activating occurring in global motion direction detection neurons is very similar to that in the brain visual cortex [29,30] in both of the activation characteristics and the inference methods. AVS achieves high accuracy with 100% in the noiseless and noisy environment. To demonstrate its high efficiency, we let the conventional convolutional neural network take on the same detection task; the experimental results show that the mechanism maintains an absolute significant superiority over CNN regardless of the presence or absence of noise or the amount of noise, which proves the model's high efficiency and high noise immunity. In addition, the model maintains its irreplaceable advantages over CNN in many aspects, including the interpretability of its principles, biosimilarity, high hardware implementation possibilities, simplicity without learning, the flexibility of the model for further learning, and the availability of excellent initial parameters. We also found that the mechanism can naturally be an edge detector in the condition of motion, which coincides with the edge-sensitive nature of the visual system [31,32]. This study also represents bold conjecture and reverse validation of the principles of neural computation in further understanding the motion direction selection pathway of the visual system. We hope that a comprehensive understanding of this direction-selective neural computation will, while providing a novel computer vision technique, also provide an important stepping stone to our understanding of the more complex functions of the brain neural networks [33].

This work spans two completely different fields of biology and engineering, more specifically, neuroscience and computer science. It makes the following contributions to both.

In terms of biology, contributions as follows.

1. This work gives an advanced quantitative way and mechanism for the DS circuit in the visual system of mammals (note that including human) brain. It offers a reasonable interpretation to solve the important problem that has plagued us for decades.
2. This work can be extended as a framework for understanding a variety of basic visual phenomena, including shape orientation, motion direction, and motion velocity, as well as that in stereo vision.
3. Because of the first success and effectiveness in interpretation of the visual system, the AVS can probably be used to help us understand other mammalian perception systems that also encode in cortical circuits, such as olfaction, taste, and touch.

In terms of engineering, contributions as follows.

1. A very biologically based dendritic neural network algorithm for grayscale motion direction detection, namely AVS, is proposed for the first time.
2. The AVS is verified to be an advanced and very efficient motion direction detector based on the mammalian DS circuit. It first achieves the detection for grayscale images and has a extremely high accuracy.
3. The other superiorities are also verified, such as, high noise immunity in some high complex environments, no need for learning, no parameter, easy to hardware implementation, and high interpretability.

The rest of this paper is organized as follows. Section 2 introduces the fundamental structures and functions of the utilized dendritic neuron model and how used to construct AVS as well as the related biological basis. Section 3 shows the experimental results for verifying the effectiveness and noise immunity with comprison of the convolutional neuron network. Section 4 discusses and draws a conclusion for this work.

2. Material and Method

In this section, we discuss and develop the AVS. In the AVS, the structure and function of eight local motion direction detection neurons (LMDNs) and corresponding global motion direction detection neurons (GMDNs) that are sensitive to different specific motion directions are defined separately using a dendritic neuron model. The LMDN collects visual information from a local area in the global receptive field, which is called the local receptive field, through PCs at the front end of the pathway. PCs, HCs, BCs, and ACs in the DS pathway collect and process visual information from the local receptive field, and output by their corresponding LMDN. Then, the GMDNs collect all the outputs from each of the eight corresponding sensitive LMDNs and sum them up at the membrane layer at the end of dendrites. Finally, the sensitive direction corresponding to the GMDN with the strongest output signal (i.e., the highest activation) was inferred as the global motion direction.

2.1. Dendritic Neuron Model

Considering the existence and necessity of cellular dendritic structures and intercellular nonlinear computations in this circuit, we utilize a set of biologically sound unsupervised learnable nonlinear dendritic models called Dendritic Neuronal Models to implement the relevant operations [34]. Numerous studies have shown that brain computing is similar to computer computing in that electrical signals are transmitted in a large number of simple units that can only perform simple calculations [35]. Koch, Poggio, and Torre found that in the dendrites of retinal nerve cells, excitatory synapses are intercepted if activated inhibitory synapses are closer to the soma than excitatory synapses. They suggested considering the role of synapses with branching turning points from the perspective of logical operations [36,37]. Some experimental examples, whether the subject of this study, direction selectivity of retinal ganglion cells [38], or others such as coincidence detection of the auditory system [39], provide strong circumstantial evidence for a dendritic nonlinear model based on logical operations. Using the dendritic neuron model, we implemented the intercellular interaction architecture of PCs, BCs, and HCs in the retinal vertical pathway of DSGCs, and defined eight LMDNs for DSGCs with different directional sensitivity for local direction recognition within the local receptive field; and also defined eight corresponding GMDNs.

The dendritic neuron model (DNM) was first proposed in 2014 [40]. This is a neuronal model that takes into account the nonlinear interaction between synapses and aims to compensate for the part of dendritic nonlinear computation that is not considered in traditional artificial neural network models. In this model, the synaptic layer receives the output signals from other neurons and processes these signals separately with a sigmoid function; the dendrite layer processes the output signals from the synaptic layer's processing with a multiplication function; the membrane layer processes the output signals from each branch in the dendrite layer with a summation function; the somatic layer processes the output signals from the membrane layer with another sigmoid function, and thus the output signal of the whole dendritic neuron model is obtained. The functions of each part of the model are described in detail below, and the model structure is shown in Figure 1.

2.1.1. Synaptic Layer

A synapse is a structure of mutual contact between one neuron and another neuron that connects the dendrites of the postsynaptic neuron to the axons of the presynaptic neuron and is functionally used to feedforwards transmit output signals from cell bodies in the presynaptic neuron, excitatory or inhibitory to influence the postsynaptic potential and thus the biological response of the postsynaptic neuron. The connection function of the i th ($i = 1, 2, \dots, N$) synapse on the j th ($j = 1, 2, \dots, M$) synaptic layer is given as:

$$S_{i,j} = \frac{1}{1 + e^{-k(\omega_{i,j}X_{i,j} + \theta_{i,j})}} \quad (1)$$

where $S_{i,j}$ is the output of the i th synapse on the j th synaptic layer. k is a positive constant. $x_{i,j}$ reflects the input signal of the i th synapse on the j th synaptic layer. The weights $\omega_{i,j}$ and the threshold $\theta_{i,j}$ are the connection parameters planned for learning. The combination of the parameters $\omega_{i,j}$ and $\theta_{i,j}$ final learning outcomes can be grouped into four connection types, constant 0, constant 1, excitatory and inhibitory. As shown in Figure 2, where the horizontal coordinates indicate the output signal from the neuron at the presynapse and the vertical coordinates indicate the output of the synaptic layer in the neuron at the postsynapse. Since the range of X is $[0, 1]$, only the part that meets the requirement needs to be focused on.

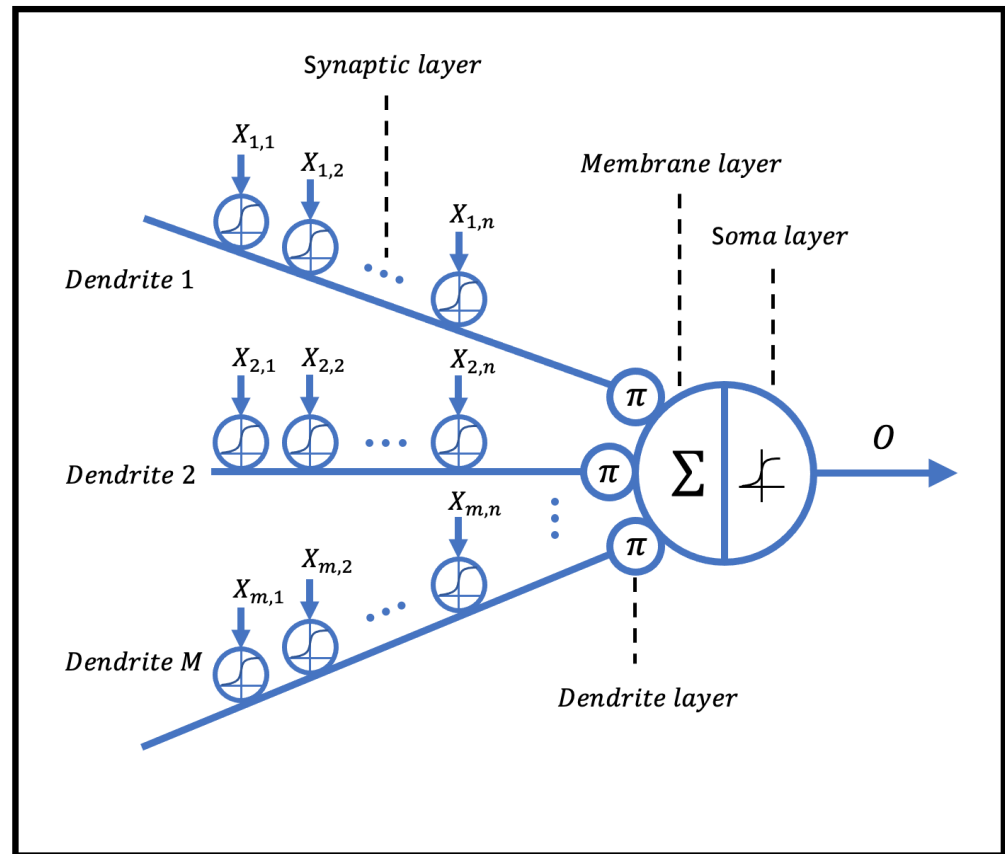


Figure 1. The structure of Dendritic Neuron Model.

The case where the output is always close to 0 regardless of how the input varies between 0 and 1 is called a constant 0 connection (when $\omega_{i,j} < 0 < \theta_{i,j}$ or $0 < \omega_{i,j} < \theta_{i,j}$); the case where the output is always close to 1 regardless of how the input varies between 0 and 1 is called a constant 1 connection (when $\theta_{i,j} < \omega_{i,j} < 0$ or $\theta_{i,j} < 0 < \omega_{i,j}$); the case where the output is always positively correlated with the input regardless of how the input varies between 0 and 1 is called excitatory connection (when $0 < \theta_{i,j} < \omega_{i,j}$); and the case where the output is always negatively correlated with the input regardless of how the input varies between 0 and 1 is called inhibitory connection (when $\omega_{i,j} < \theta_{i,j} < 0$). Notably, these four synaptic connection types are particularly important for inferring the morphology of neurons by specifying the location of dendrites and the type of synapses.

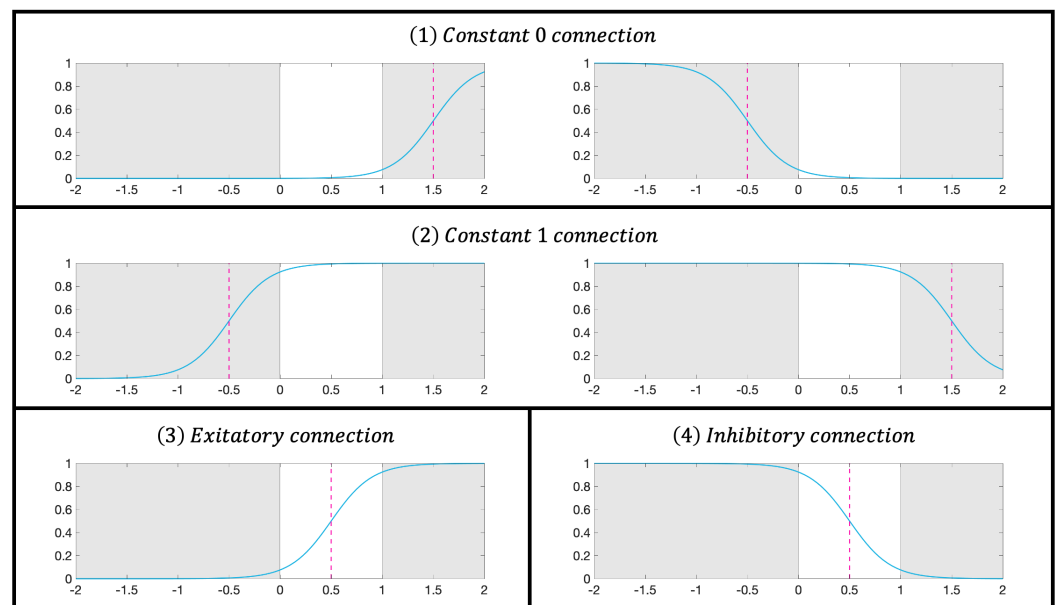


Figure 2. Four types of connection in the synaptic layer.

Furthermore, in the case where the input is binary, the four types of connections in the synaptic layer can be approximated and simplified as follows.

Constant 0 connection:

$$S_{i,j}(X_{i,j}) = 0 \quad (2)$$

Constant 1 connection:

$$S_{i,j}(X_{i,j}) = 1 \quad (3)$$

Excitatory connection:

$$S_{i,j}(X_{i,j}) = \begin{cases} 0, & X_{i,j} = 0 \\ 1, & X_{i,j} = 1 \end{cases} \quad (4)$$

Inhibitory connection:

$$S_{i,j}(X_{i,j}) = \begin{cases} 1, & X_{i,j} = 0 \\ 0, & X_{i,j} = 1 \end{cases} \quad (5)$$

2.1.2. Dendrite Layer

The dendrite layer processes the output from multiple synapses in the synaptic layer through a multiplicative function. The multiplicative equation is chosen for dendritic branches because it reproduces the nonlinear relationship between synapses on the dendrite layer of a neuron. In the case where the inputs and outputs of a dendrite layer correspond to binary, the multiplicative equation is equivalent to a logical AND operation. In Figure 1, the symbol π denotes the multiplication operator. The output of the dendrite layer branch of the j th one can be written as:

$$D_j = \prod_{i=1}^n S_{i,j} \quad (6)$$

where D_j is the output of the j th dendritic branch.

2.1.3. Membrane Layer

The membrane layer processes the output of each branch in the dendrite layer through a summation function, which is similar to the logical OR operation in binary. The output of

the membrane layer will be input and affect the soma layer. The function of the membrane layer can be expressed as follows.

$$M = \sum_{j=1}^m D_j \quad (7)$$

where M is the output of the membrane layer.

2.1.4. Soma Layer

The soma is the last part of the neuronal model. Biologically, it is connected to the axon, and the soma transmits output signals to the axon and through synapses at the end of the axon to the dendritic layer of the next level of the neuron. If the output of the membrane layer exceeds the threshold of the soma, the soma will fire; conversely, the soma outputs an inhibitory signal. The soma function is expressed as a sigmoid function, and its value is the final output signal of such a dendritic neuron model.

$$O = \frac{1}{1 + e^{-k_s(M - \theta_s)}} \quad (8)$$

where O refers to the output of the soma, i.e., that of the dendritic neuron model. k_s is a positive constant. θ_s is the firing threshold of the soma, which ranges from $[0, 1]$.

In addition, the biological functions of soma can also be realized with step functions in cases where the magnitude of neuronal signal strength does not need to be considered. This approach realizes the soma functions in neurons more concisely and can be expressed as follows.

$$O = \begin{cases} 0, & M \leq \theta_s \\ 1, & M > \theta_s \end{cases} \quad (9)$$

where when the output of the membrane layer exceeds the threshold of the soma, the neuron fires, and the firing signal is denoted by 1; conversely, the neuron does not fire and is denoted by 0.

2.1.5. Structure of Dendritic Neuronal Model

Figure 1 completely depicts the structure of DNM, which contains four layers. A series of input signals X from presynaptic neurons are first processed by four types of connections in the synaptic layer to obtain the synaptic layer outputs. The synaptic layer outputs on the same dendritic layer branch are multiplied nonlinearly to obtain a series of dendritic layer outputs. The dendritic layer outputs are summed over the membrane layer to obtain the membrane layer output. Finally, the cell body computes the final output signal of this neuron and transmits the output signal to the next level neuron via axonal and synaptic structure.

2.2. Direction-Selective Pathway and Visual Cortex Responses

The core of the retinal DS pathway is the DSGCs, and each of the different species of DSGCs has a specific sensitivity to motion phenomena occurring in a particular direction in its local sensory field. In the pre-pathway of DSGCs, four types of cells are known to exist, including PCs, which are perceptrons with photoelectric conversion function, HCs and ACs in the lateral information exchange function structure, and BCs, which are vertical information transmitters between PCs and DSGCs; where HCs implement lateral regulation of electrical signals immediately after the completion of photoelectric conversion by PCs, and ACs implement posterior lateral regulation of information after processing by HCs and BCs. The ACs implement the back-end lateral regulation of the information processed by the HCs and BCs. In conclusion, in the perceptual recognition of motion direction in the brain, the light visual signals from the external world rely on the PCs to be converted into electrical signals to be recognized by the retina, and the layers of cells in the DS pathway sort out the recognition signals layer by layer to produce the motion direction selectivity of

the DSGC. Ultimately, the activation state output by the DSGC is transmitted to the cortex and is integrated. Figure 3 depicts the outline process and structure by which a particular motion direction information is extracted as the visual signal is collated from the external world through the retina and integrated into the visual cortex. The eight DSGCs defined in this paper correspond to eight sets of DS model pathways with the same structure but different functions as shown in Figure 3. In the proposed model, the integration link of ACs (represented by fading in Figure 3) is temporarily eliminated in this research stage because it is non-essential in the underlying direction selection function. Since it is responsible for the lateral information exchange at the end of the DS pathway, we believe it will perform other higher-level information integration functions rather than simply participating in or repeating the work already done at the front end of the pathway.

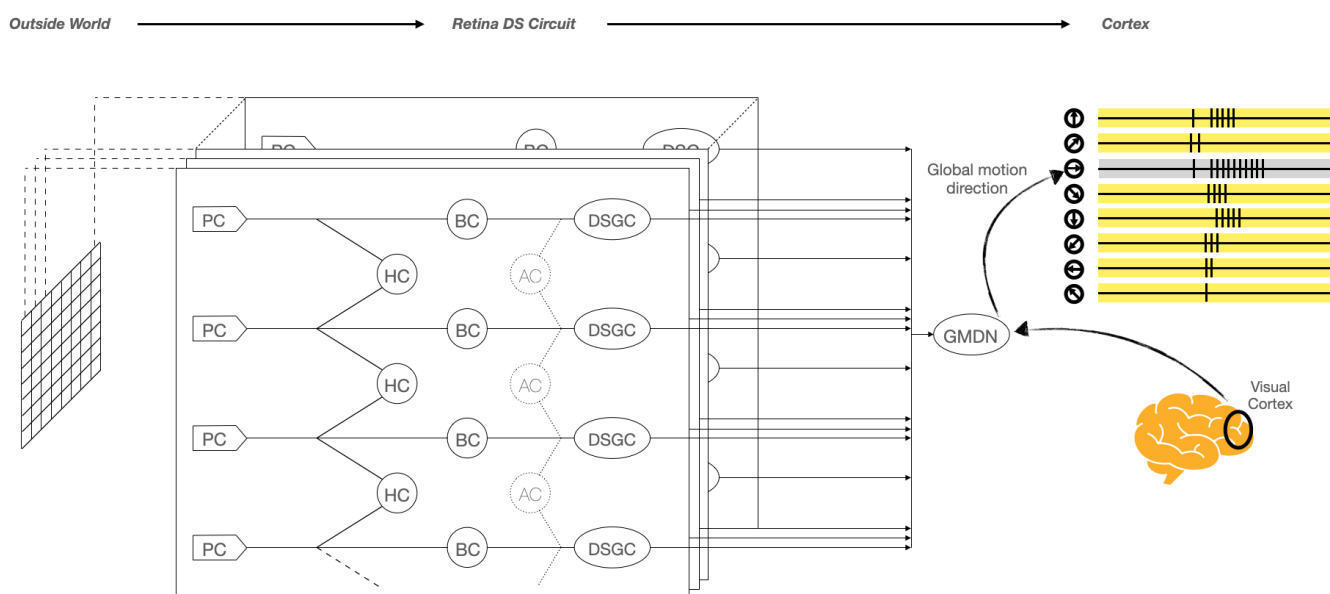


Figure 3. The structure of direction-selective pathway in the visual system.

It was found that for global motion in a certain direction, neurons in the primary visual cortex that are selective for the corresponding direction are strongly activated. At the same time, neurons sensitive to other directions were also activated to varying degrees, but the activation intensity was less than that of the neuron selective for the same direction as the global motion [30]. The one-dimensional statistical graph on the right side in Figure 3 shows a example activation phenomena in the visual cortex triggered by rightward global motion. We expect that our GMDNs in the cortex can activated as a similar way.

2.3. Local Motion Direction Detection Neurons

We propose eight LMDNs and their corresponding local receptive fields to define DSGCs sensitivities to different motion directions; using the cellular distribution structure shown in Figure 3, we scan the light signals in the global receptive field and collect the motion direction information in the local receptive fields.

2.3.1. Local Receptive Fields

The concept of local receptive fields was described by Hubel and Wiesel as follows: individual optic nerve fibers respond only when a specific area of the retina is illuminated, and these areas are referred to as local receptive fields [30]. In 1938, Hartline found that the local receptive field, although small, was still at an observable level and that its location on the retina was fixed [41]. He also noted that a retinal ganglion cell can receive excitatory influences over many convergent pathways; its axon is the final common path for nervous activity originating in many sensory elements [31]. Thus, the front end of each LMDN is connected to a local receptive field consisting of multiple PCs.

In AVS, the local receptive field is defined as a 3×3 square region with nine pixel points as in Figure 4, which includes one central point and eight edge points. The center pixel position of the local receptive field is denoted by (i, j) in the global receptive field, and so on for the edge pixels.

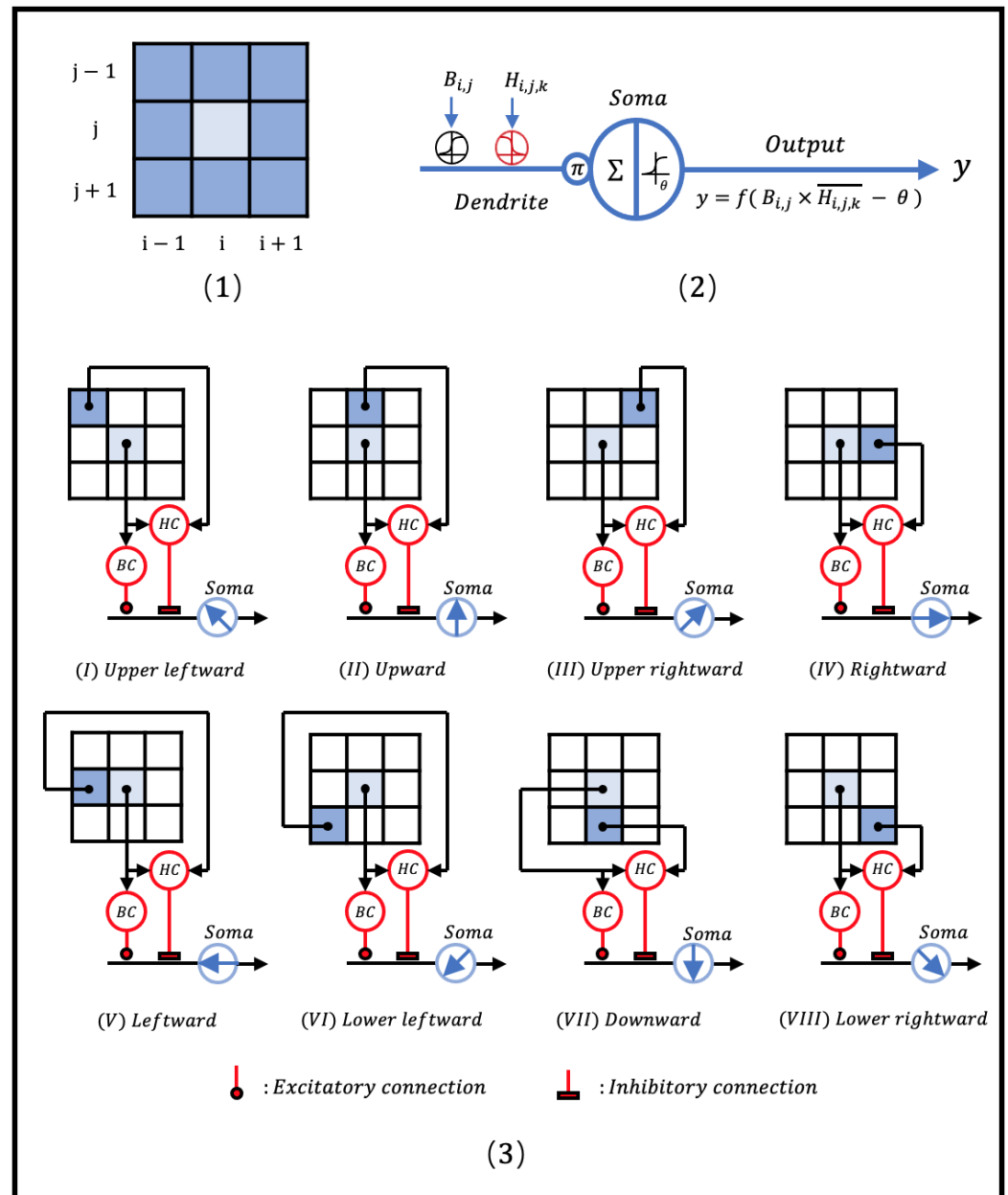


Figure 4. The structure of motion direction-selective pathway in the visual system: (1) the local receptive field and its position coordinate code when scanning; (2) the LMDN model based on DNM (black synapse: excitatory connection; red synapse: inhibitory connection); (3) DS circuits for eight types of LMDNs with different motion direction-sensitivities.

2.3.2. Implement of Dendritic Neuronal Model

Figure 4(2) shows how DNM is applied to the LMDN. X_{BC} is the output signal from the BC axon and \bar{X}_{HC} is the output signal from the HC axon, which are two inputs on a dendritic branch of the LMDN; the synaptic layers are connected in excitatory (black) and inhibitory (red) forms (corresponding to the physiological properties of BC and HC, respectively). X_{BC} and \bar{X}_{HC} are located in the same dendritic branch, and the output of the dendrite layer is obtained directly by the multiplication function; this collaborative

relationship is equivalent to the logic AND operation under the binary condition. Since there is only one output of the dendrite layer, the membrane layer input is equal to the output. The soma determines whether the size of the input value exceeds the firing threshold θ of the neuron. If so, it outputs 1, otherwise 0.

2.3.3. Eight Types of Local Motion Direction Detection Neurons

We defined eight types of LMDNs (corresponding to DSGCs with different direction-sensitivities) and the function of each cell in their forward pathways (including PC, BC, and HC) based on DNM and the real physiological structure and function of each cell for collecting local information on different motion directions. As shown in Figure 4(3), the PCs in LMDN forward pathway receive light grayscale information from the center pixel of its local receptive field, and the edge pixel in a particular direction.

PCs, as the most anterior photoreceptors in the DSGCs pathway, convert light signals into neuroelectrical signals with their unique photoelectric converting function. In AVS, each PC is responsible for detecting grayscale information of one pixel. the function of a PC is defined as follows.

$$P_{i,j,t} = x_{i,j,t}, x \in [0, 255] \quad (10)$$

where $P_{i,j,t}$ is the output of the corresponding position PC at the moment t . $x_{i,j,t}$ is the grayscale value at the moment t , corresponding to the pixel position (i, j) , which is ranged to $[0, 255]$.

In the DS pathway, the BC is only connected to the central PC. It should be noted that although Hartline argued that ON and OFF responses existed in separate regions of the local receptive field [31], i.e., the central region could be only one of the two feedback modes, subsequent experiments by Barlow confirmed that motion in any area within the boundary of the local receptive field leads to a simultaneous response of both [42]. This indicates that the activation of BC depends only on whether the light information detected by its corresponding PC changes, and is not limited to one of dark-to-light or light-to-dark, which is contradicted by the definition of BC function in the model proposed by Tang et al. The BC in such a model is divided into two functions according to ON-response and OFF-response, and they are not fully complementary, which affects the rationality and efficiency of the model. Therefore, BC function should be defined in a more comprehensive but simpler form as follows.

$$B_{i,j} = \begin{cases} 0, & P_{i,j,t} - P_{i,j,t-\Delta t} \leq \theta \\ 1, & P_{i,j,t} - P_{i,j,t-\Delta t} > \theta \end{cases} \quad (11)$$

where $B_{i,j,t}$ is the output produced by the BC connected to (i, j) , which represents the central pixel position in the local receptive field, at moment t . $(t - \Delta t)$ and t denote the two adjacent moments captured during the object motion, respectively. Δt represents the value of the interval between the two moments. θ is the activation threshold, which is a positive number, and the pair of light information is considered to have no difference if their differential value does not exceed this threshold.

HC spans the local receptive field's center and edge. It plays a pivotal role in modulating cone output via reciprocal feedback. In addition, HCs can release GABA (an essential related synaptic transmitter) [43] and GABA receptors have been found on BC dendrites [43–45]. Hence, depending on the local chloride equilibrium potential in BC dendrites, HCs can provide these cells with feedforward inhibition [11,46–48]. HCs are widely thought to be involved in global signal processing, for example in the context of contrast. The basic principle is lateral inhibition and the subtraction is included [49]. The spatiotemporal modulation by inhibitory feedback affects the input of BC [48] and thus indirectly affects the output of BC [50,51]. Therefore, we use the subtraction function of HC to determine whether there is a difference between the outputs of the cells connected at its ends, and thus extract information about the difference in the horizontal dimension. As

HC's inhibitory connection can be presented in the LMDN's synaptic layer later. The HC intracellular function is expressed generally as follows.

$$H_{i,j} = \begin{cases} 0, & P_{edge,t} - P_{i,j,t-\Delta t} \leq \theta \\ 1, & P_{edge,t} - P_{i,j,t-\Delta t} > \theta \end{cases} \quad (12)$$

where $P_{edge,t}$ represents the grayscale value at the moment t on an edge pixel in the LMDN's local receptive field. Corresponding to LMDNs of different direction-sensitivities, the eight types of HCs' spatiotemporal intercellular functions can be defined in detail as follows.

First HC for LMDNs with the sensitivity of upper leftward motion:

$$H_{i,j,1} = \begin{cases} 0, & P_{i-1,j-1,t} - P_{i,j,t-\Delta t} \leq \theta \\ 1, & P_{i-1,j-1,t} - P_{i,j,t-\Delta t} > \theta \end{cases} \quad (13)$$

First HC for LMDNs with the sensitivity of upper leftward motion:

$$H_{i,j,2} = \begin{cases} 0, & P_{i,j-1,t} - P_{i,j,t-\Delta t} \leq \theta \\ 1, & P_{i,j-1,t} - P_{i,j,t-\Delta t} > \theta \end{cases} \quad (14)$$

Third HC for LMDNs with the sensitivity of upper rightward motion:

$$H_{i,j,3} = \begin{cases} 0, & P_{i+1,j-1,t} - P_{i,j,t-\Delta t} \leq \theta \\ 1, & P_{i+1,j-1,t} - P_{i,j,t-\Delta t} > \theta \end{cases} \quad (15)$$

Fourth HC for LMDNs with the sensitivity of rightward motion:

$$H_{i,j,4} = \begin{cases} 0, & P_{i+1,j,t} - P_{i,j,t-\Delta t} \leq \theta \\ 1, & P_{i+1,j,t} - P_{i,j,t-\Delta t} > \theta \end{cases} \quad (16)$$

Fifth HC for LMDNs with the sensitivity of lower rightward motion:

$$H_{i,j,5} = \begin{cases} 0, & P_{i+1,j+1,t} - P_{i,j,t-\Delta t} \leq \theta \\ 1, & P_{i+1,j+1,t} - P_{i,j,t-\Delta t} > \theta \end{cases} \quad (17)$$

Sixth HC for LMDNs with the sensitivity of downward motion:

$$H_{i,j,6} = \begin{cases} 0, & P_{i,j+1,t} - P_{i,j,t-\Delta t} \leq \theta \\ 1, & P_{i,j+1,t} - P_{i,j,t-\Delta t} > \theta \end{cases} \quad (18)$$

Seventh HC for LMDNs with the sensitivity of lower leftward motion:

$$H_{i,j,7} = \begin{cases} 0, & P_{i-1,j+1,t} - P_{i,j,t-\Delta t} \leq \theta \\ 1, & P_{i-1,j+1,t} - P_{i,j,t-\Delta t} > \theta \end{cases} \quad (19)$$

Eighth HC for LMDNs with the sensitivity of leftward motion:

$$H_{i,j,8} = \begin{cases} 0, & P_{i-1,j,t} - P_{i,j,t-\Delta t} \leq \theta \\ 1, & P_{i-1,j,t} - P_{i,j,t-\Delta t} > \theta \end{cases} \quad (20)$$

where $H_{i,j,k}$ ($k \in \{1, 2, 3, 4, 5, 6, 7, 8\}$) denotes the output produced by the corresponding HC at the k th type of LMDN scanning (i, j) position. $(t - \Delta t)$ and t denote the two adjacent moments during the object motion, respectively. θ is also used for BC; the sensitivity of the cells to recognize light changes will peak when θ equals 0. Therefore, in the experiments of this paper, we make θ equal to 0.

In the anterior synapse connection of BC, there are axons of both PC and HC, and it is in this particular synaptic gap that HC can have the opportunity to receive signals from PCs and regulate BC laterally [11]. HC indirectly regulates the output of BC by influencing the input of BC. This process is equated to the nonlinear interaction of HC with BC in the dendrite layer, before which are the excitatory and inhibitory synaptic layer connections of BCs and HCs respectively. The synaptic layer and dendrite layer of the LMDN can be defined together as follows.

$$D_{i,j,k} = B_{i,j} \times \overline{H_{i,j,k}}, k \in \{1, 2, 3, 4, 5, 6, 7, 8\} \quad (21)$$

where $D_{i,j,k}$ is the output of the k th type of LMDN dendrite layer branch, which is subsequently fed into the membrane layer. $H_{i,j,k}$ is the output of the HC corresponding to the k th type of LMDN. The overline indicates the inhibitory connection function that $H_{i,j,k} = 1$ causes $\overline{H_{i,j,k}} = 0$ and $\overline{H_{i,j,k}} = 1$ if $H_{i,j,k} = 0$.

Since there is only one dendrite layer branch of the LMDN, the output of the membrane layer is equivalent to that of the dendrite layer.

$$M_{i,j,k} = D_{i,j,k} \quad (22)$$

The membrane layer output is the input of the soma. The soma compares the input value with the firing threshold, and if the input value exceeds the threshold, the soma fires and outputs 1; conversely, the soma does not fire and outputs 0. The function of the LMDN soma is defined as:

$$Y_{i,j,k} = \begin{cases} 0, & M_{i,j,k} \leq \theta_s \\ 1, & M_{i,j,k} > \theta_s \end{cases} \quad (23)$$

where $Y_{i,j,k}$ is the output of soma at the k th type of LMDN scanning (i, j) position. θ_s is the firing threshold of the soma, due to the membrane layer output of 0 or 1. In the model for the experiments in this paper, the parameter θ_s can take any value of $(0, 1)$ without affecting the results because the simplest neural calculation, i.e., the logic operation, is used.

2.3.4. Scan

We call the entire receptive space in the visual field the global receptive field. Representing the global receptive field as a rectangular two-dimensional space consisting of a series of arranged pixels of size $M \times N$, there are $8 \times M \times N$ LMDNs in the model; LMDNs of the same sensitivity are spread across the global receptive field, corresponding to the presence of $M \times N$. To facilitate understanding, Figure 5 shows a scanning example; the same LMDNs spread across the global receptive field scan the field of size 4×4 . The position of the central pixel in the local receptive field is represented in the form of coordinates.

Generally, during the scanning process, we denote the position of the center of the LMDN local receptive field in the global receptive field by (i, j) ($i = 1, 2, \dots, M$; $j = 1, 2, \dots, N$), as shown in Figure 4(1). Thereby, the positions of the pixel points in the upper-left, up, upper-right, right, lower-right, down, lower-left and left of the local receptive field in global receptive field during the scanning process can be represented by $(i-1, j-1)$, $(i, j-1)$, $(i+1, j-1)$, $(i+1, j)$, $(i+1, j)$, $(i+1, j+1)$, $(i, j+1)$ and $(i-1, j+1)$, respectively.

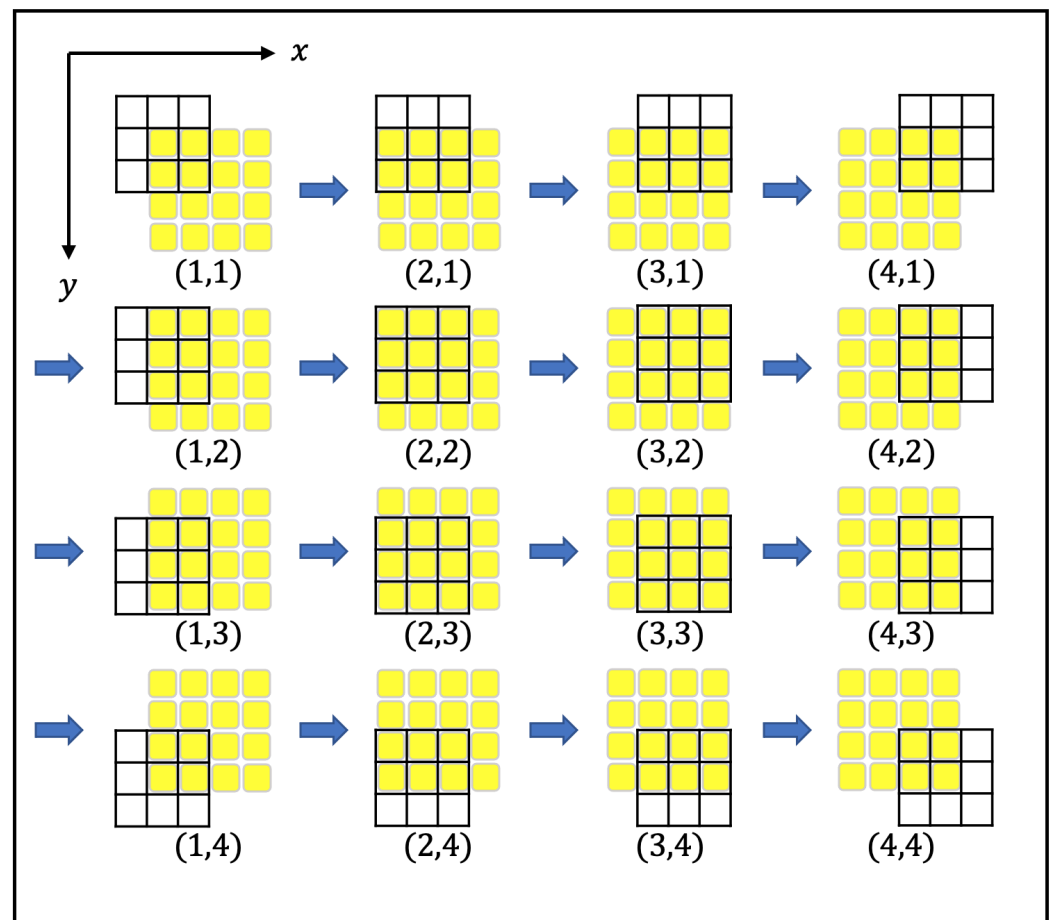


Figure 5. The scanning process of LMDNs. The center pixel of the local receptive field crosses each pixel of the global receptive field in turn. For the situations on scanning the edge pixels, padding of any constance affects few to AVS performance.

2.4. Global Motion Direction Detection Neurons

In a global receptive field of size $M \times N$, the same type of LMDNs will get $(M \times N)$ local motion direction feature values after one scan, which can be represented by a matrix of $M \times N$, i.e., local motion direction feature maps. eight types of LMDNs will get $(8 \times M \times N)$ feature values after scanning a global receptive field of $M \times N$, which can be represented by $8 \times M \times N$ local motion direction feature maps. In the visual cortex, the eight corresponding GMDNs collect all the local motion direction information from each of the eight feature maps to obtain the global feature values for a specific motion direction. The function of the GMDN is defined as an average pooling process.

$$G_k = \frac{\sum_{i=1}^M \sum_{j=1}^N Y_{i,j,k}}{M \times N} \quad (24)$$

where G_k is the output of the k th GMDN, corresponding to the characteristic map of local motion directions output by the same sensitive LMDNs. G_k is used to represent the intensity of the occurrence of the k th local motion direction at the global scale and is therefore subsequently used to calculate the probability of the occurrence of global motion in a particular direction.

$$P_k = \frac{G_k}{\sum_{k=1}^8 G_k} \quad (25)$$

where P_k refers to the probability that the k th motion direction is the global motion direction.

Eventually, GMDNs infer the motion direction corresponding to the largest P_k value as the global motion direction. Eight GMDNs process the eight feature maps scanned by the corresponding group of LMDNs in the global range, calculate the activation intensity by average pooling, and infer the actual motion direction of the object in the global receptive field accordingly, and the process is shown in Figure 6.

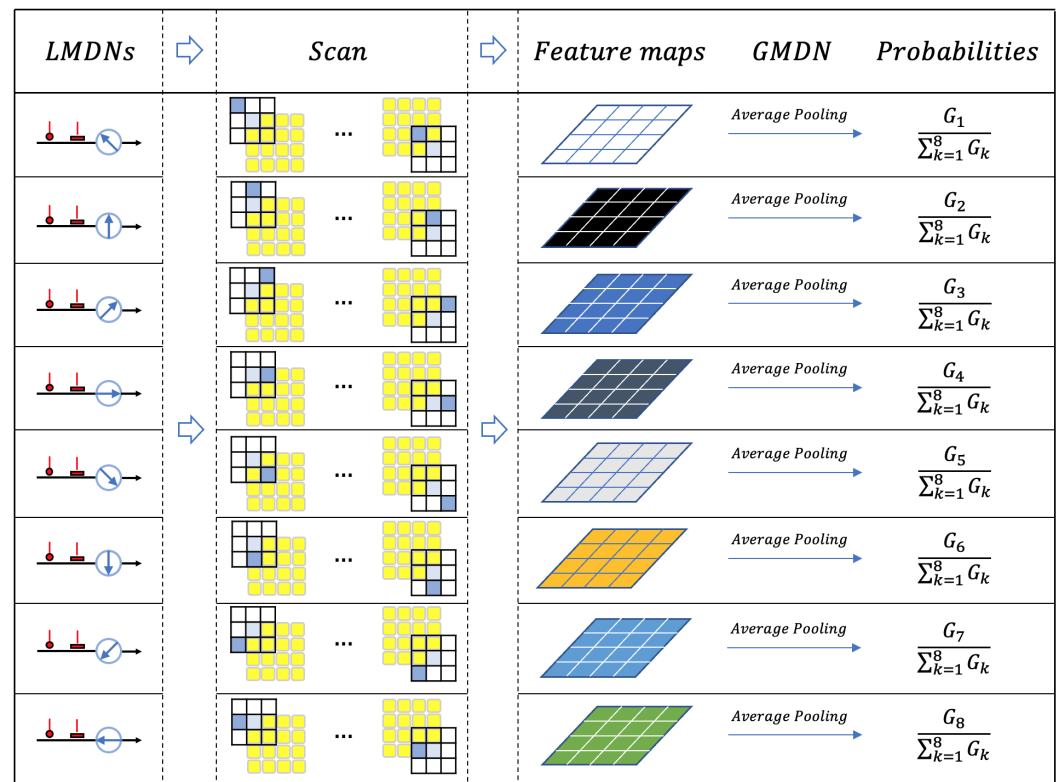


Figure 6. The global motion direction inferring process of GMDNs. Eight feature maps are given by scanning the input images using the eight corresponding LMDNs. GMDNs finally give the probabilities by average pooling to eight feature maps.

For easier understanding to the whole procedure, AVS is summarized in a pseudo-code manner in Algorithm 1. Figure 7 shows a case of detection of an “L” shaped object in a 5×5 global receptive field of 25 pixels, moving lower rightwards. In the two images at $(t - \Delta t)$ and t , the grayscale values of each pixel of the object are the same. The grayscale values of all pixel points in the background are randomly selected. For ease of observation, the pixel points covered by the object in Figure 7 are highlighted in red.

Algorithm 1: AVS.

Input: 2 grayscale images

Output: the global motion direction occurring in the input images

- 1 **Local information collection:** scan the 2 images by the 8 LMDNs, thus obtaining 8 local motion direction feature maps
 - 2 **Global information collection:** average pooling processing by the 8 GMDNs
 - 3 **Recognition result:** infer the corresponding sensitive direction of the GMDN with the highest activation intensity to the present global motion direction.
-

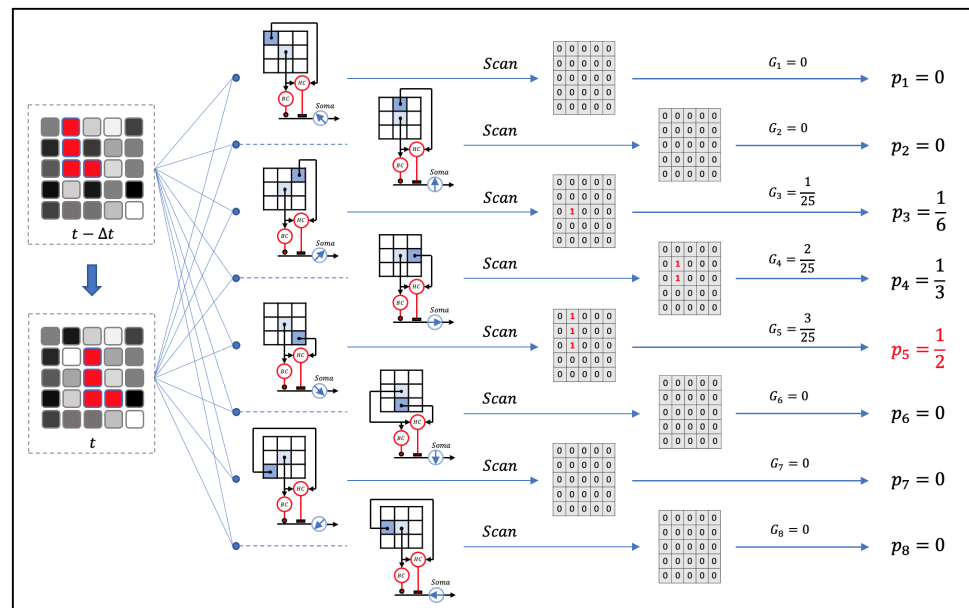


Figure 7. An detection example of AVS for an “L” shaped object. First step, the two grayscale images occurring a lower rightward global motion direction are input into the AVS. Second step, the eight types of LMDNs scan the input images. Third step, eight local motion direction feature maps are generated from LMDNs recording the local motion direction information (the activated local areas are marked by red “1”). Fourth step, eight corresponding GMDNs statistically collect and give a average pooling output. Final step, the direction corresponding to the GMDN with the highest activation intensity is inferred to have the most probability to be the global motion direction.

3. Experiment

3.1. Performance

A grayscale image of size 32×32 with 1024 pixels is used as input; the image content is obtained by sequentially generating the background and the object. Each pixel point in the background has the same grayscale value, which is generated randomly. The object consists of several randomly positioned but continuous pixel points with random and consistent grayscale values. The object is randomly placed at a random position above the background, and then the object is allowed to move in one of eight directions (\uparrow , \nearrow , \rightarrow , \searrow , \downarrow , \swarrow , \leftarrow , \nwarrow), and two successive frames of the motion are extracted and input to AVS. In multiple types of instances, comparing the actual object motion direction with the detection results of AVS, statistical detection accuracy is used to analyze its effectiveness. Considering that different object sizes may affect the detection performance, the motion instances of objects of 10 different sizes (pixel scale of 1, 2, 4, 8, 16, 32, 64, 128, 256, 512) in 8 directions are implemented separately; different sizes of objects are tested 125 times in each motion direction, so a total of 10 sizes of objects in 8 motion directions are 10,000 testing experiments.

The experimental results are reported in Table 1 using AVS to detect the motion direction of objects in different instances. The results show that AVS can achieve a 100% detection accuracy for the motion of 10 different sizes of objects in grayscale images in different directions. This indicates that AVS is able to effectively detect the direction of object motion through successive grayscale image frames in the video, regardless of the size and shape of the objects and regardless of where they are located in the image.

Figure 8a shows an instance of AVS motion direction detection in the theoretical image condition. It shows two successive grayscale images of a randomly shaped object consisting of 16 pixels moving downward in an image field consisting of 1024 pixels in 32×32 . Figure 8b depicts the number of times the eight LMDNs were activated using the one-dimensional statistics used in neuroscience studies. It was shown in mammalian visual cortex that, for a motion direction occurring in the global receptive field, neurons also have direction-sensitivities. The global motion direction corresponds to the GMDN

with the highest activation intensity which is highly similar to phenomena in the primary cortex shown in Figure 3. LMDNs were activated 12, 9, 11, 6, 12, 7, 21, and 7 times, which indicates the activation intensity of the corresponding GMDNs, respectively, in the time of Δt , with the neuron sensitive to the downward motion being activated 21 times and the density of the corresponding activation bar being significantly higher than any of the others, the corresponding direction is consistent with the actual motion direction of the object in the image, and thus a successful detection. Figure 8c counts the sum number of activations of each LMDNs, i.e., the corresponding GMDNs response strength, with a bar chart. Figure 8d marks the pixel locations that caused the LMDNs to activate during the Δt in blue, a region that clearly depicts the general outline of the moving object, due to the particular sensitivity of the LMDNs to the edges of moving objects within the local receptive field. Thus, the model can also perform edge tracking of objects in motion, dynamically displaying the direction in which a moving object is located in the global receptive field in real-time. LMDNs naturally demonstrate the edge-sensitivity in animal vision along with the local motion direction selectivity where the response to non-preferred directions is null. Furthermore, the LMDN has the ability to correctly extract directional information in object motion phenomena at arbitrary greyscale, which is consistent with Barlow's basic experimental conclusion that changes in light do not interfere with directional selectivity. The result also excludes the significant resistance to the interpretation of the motion detection mechanism posed by the opposite functions of the ON and OFF response regions in the same local receptive field. In conclusion, all the properties of AVS correspond highly to the core physiological experimental findings.

Table 1. Experimental results of AVS on motion direction detection with different scales of object.[illegible]

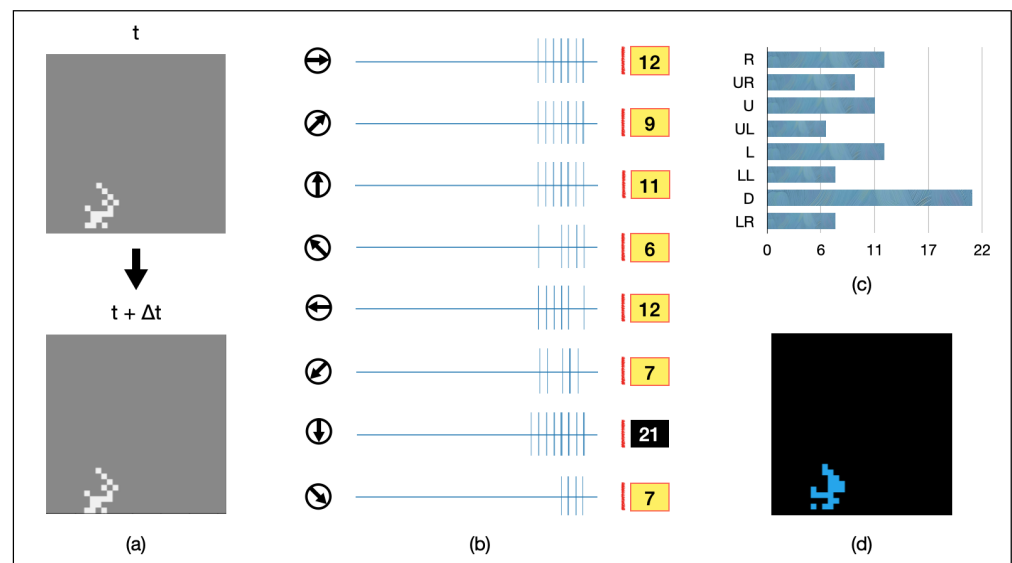


Figure 8. An AVS motion direction detection instance in the theoretical image condition: (a) two successive frames of a downward motion of an 16-pixel object; (b) activation map of the eight LMDNs during scan; (c) LMDNs activations are counted by a bar chart; (d) activated pixels by all the LMDNs during detection process are marked by blue.

3.2. Comparing the Performance of CNN

The traditional CNN is considered to be the most effective algorithm in pattern recognition. To verify the superiority of AVS, we let CNN undertake the same detection task for comparing the performance. We use a classical and representative CNN method called LeNet-5 [52], and the optimizer uses Adam [53]. The network structure is shown in Figure 9, where two grayscale images of the object motion process are input into two channels of the CNN. We generate a total of 10,000 sets of motion instances, 1000 sets for each of the 10 different sizes of objects, which contain another 125 sets for each of the 8 directions. Since there is a model learning optimization process for the CNN, we divide the data of the same category in 10,000 sets into a learning set and a test set by 8:2. The test set was also used to test AVS. to eliminate the randomness in the optimization process for learning as much as possible, the CNN was allowed to learn 30 times independently, and the mean accuracy was used for experimental analysis. In the learning of CNN, epochs and batches value are both 100; the best model in the epochs process is selected for subsequent testing based on the largest accuracy value.

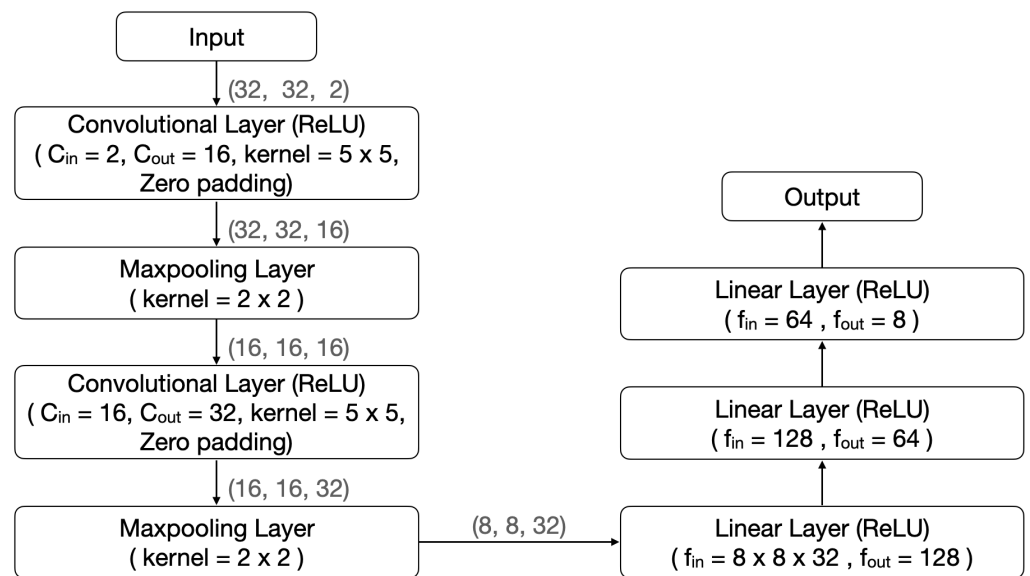


Figure 9. The structure of CNN.

The experimental results are reported in Table 2. CNN can achieve more than 90% correct rate in detecting object motion only at the scale of 256 pixels or more. Meanwhile, AVS shows 100% error-free superior accuracy in detecting object motions of all sizes including the single-point object. In terms of the mean value, ARVS achieves 100% accuracy, while CNN is only 81.98%. Figure 10 records the accuracy trends of the different methods with object size. The smaller the object size, the lower the detection accuracy of CNN; for single-point object motion, the detection accuracy is the lowest, only 65.48%. While for each different object size, AVS is significantly higher than CNN and shows excellent robustness. Therefore, AVS has a surprising superiority over CNN in the motion direction detection task.

Table 2. The experimental results of AVS and CNN in motion direction detection on different scales of objects.

Object Scale	AVS	CNN
1-pixel	100%	65.48%
2-pixel	100%	71.58%
4-pixel	100%	76.73%
8-pixel	100%	77.92%
16-pixel	100%	83.07%
32-pixel	100%	86.22%
64-pixel	100%	89.63%
128-pixel	100%	86.8%
256-pixel	100%	90.4%
512-pixel	100%	91.97%
Mean	100%	81.98%

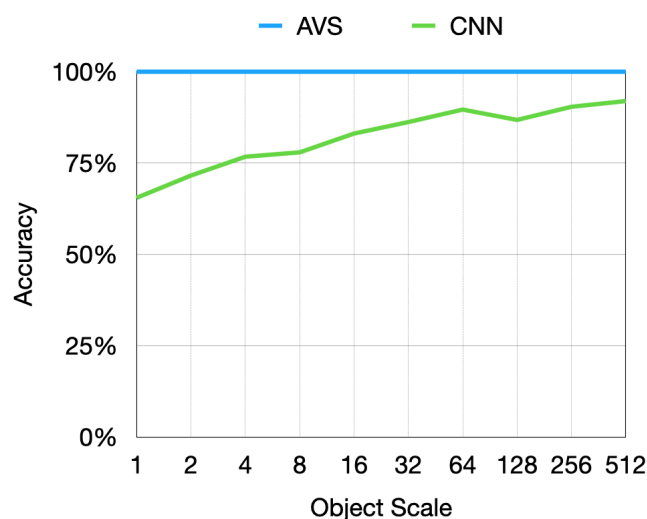


Figure 10. The detection accuracy trends of AVS and CNN to motion direction of different scales of object.

3.3. Performance in Complex Environments

Although AVS is able to achieve error-free in the theoretical conditions, some random and unknown interference information might affect the AVS performance in complex environments. To demonstrate more comprehensively the performance of AVS in complex environments. We experiment with static and dynamic noise separately. In addition, we further classify them as in the background or being on top of the image. Thus, four types of noise are given, namely static background noise, dynamic background noise, static full-image noise, and dynamic full-image noise. In the experiments, the grayscale value of each noise pixel is randomly given. By controlling the number of noise pixels added, we tested 3 progressive levels of complexity for this noise environment. The number of noise pixels was set to 10%, 20%, and 30% of the total number of picture pixels, respectively. The purpose was to examine the noise immunity trend of AVS in each type of complex environment.

The first type of complex environment, static background noise. From the background of each theoretical image, a number of pixels at the same position are selected and they are replaced with an equal number of noise pixels. This noise can make the static background of the moving object more cluttered, thus increasing the complexity of the environment. Figure 11a shows an detection example of AVS under static background noise. It shows two successive images of a randomly shaped object consisting of 128-pixel points moving downward in an image consisting of 1024 pixel points in 32×32 . In these two images, some pixel points in the noiseless background are replaced by new random grayscale pixel points with fixed positions. Figure 11b records the responses of the LMDNs in an intuitive one-dimensional statistical method used in neuroscience studies, the sum of which is the activation intensity of each corresponding GMDN. LMDNs were activated 47, 32, 35, 38, 44, 35, 83, and 40 times, respectively, during Δt , with the LMDN preferring downward motion being activated 83 times and its bar density was significantly higher than any of other neurons and thus was inferred as the global motion direction; the sensitive direction of this class of neurons was consistent with the actual motion direction of the object in the image, and thus a successful detection. The histogram in Figure 11c records the response strength of GMDNs. The pixels activated by either neuron in Δt time are highlighted in blue in Figure 11d.

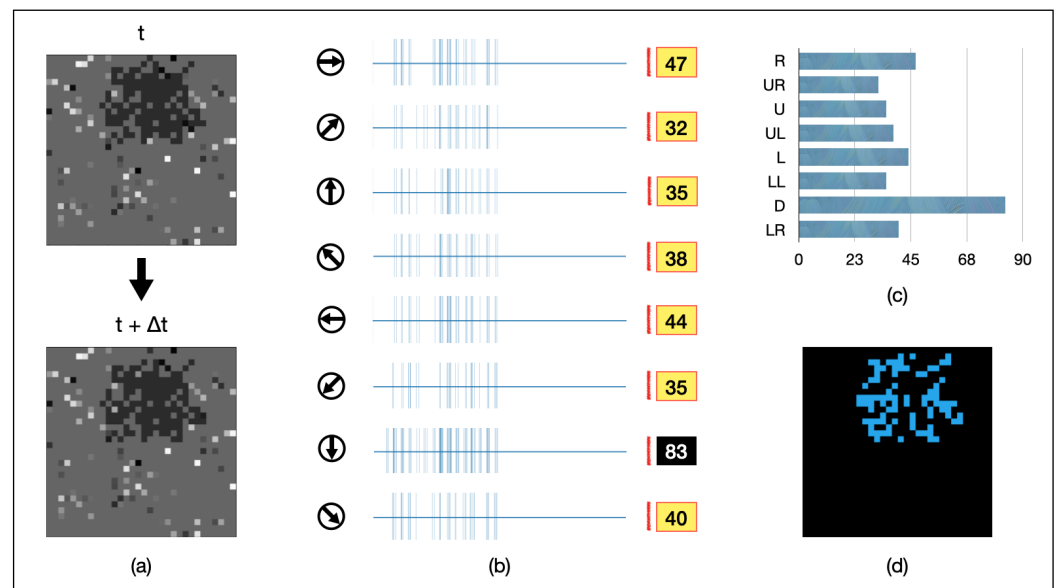


Figure 11. An AVS motion direction detection instance in static background noise: (a) two successive frames of a downward motion of a 128-pixel object; (b) activation map of the eight LMDNs during scan; (c) LMDNs activations are counted by a bar chart; (d) activated pixels by all the LMDNs during detection process are marked by blue.

The second type of complex environment, dynamic background noise. From the background of each theoretical image, a number of pixels at different positions are independently selected and replaced with an equal number of noise pixels. This noise can make the background of a moving object more cluttered while also containing real-time changes, thus further increasing the complexity of the environment. Figure 12a shows an detection example under dynamic background noise. It shows two successive frames of a randomly shaped object consisting of 64-pixel points moving upper leftward in an image consisting of 1024 pixel points in 32×32 . In these two frames, some pixel points at different locations in the background are replaced by random grayscale noise points. It can be seen from the figure that the noise points are present only in the background and change dynamically during the motion of the object. Figure 12b depicts the number of activations of LMDNs, i.e., the activation intensity of each corresponding GMDN. LMDNs were activated 81, 88, 90, 116, 89, 92, 81, and 84 times, respectively, during Δt , with the LMDN preferring upper leftward motion being activated 116 times and the density of activation bars being significantly higher than any of the others. The sensitive direction of this class of neurons is consistent with the actual motion direction of the object in the image, and thus a successful detection. The bar graph in Figure 12c records the response strength of GMDNs. The pixels activated by either neuron in Δt time are highlighted in blue in Figure 12d.

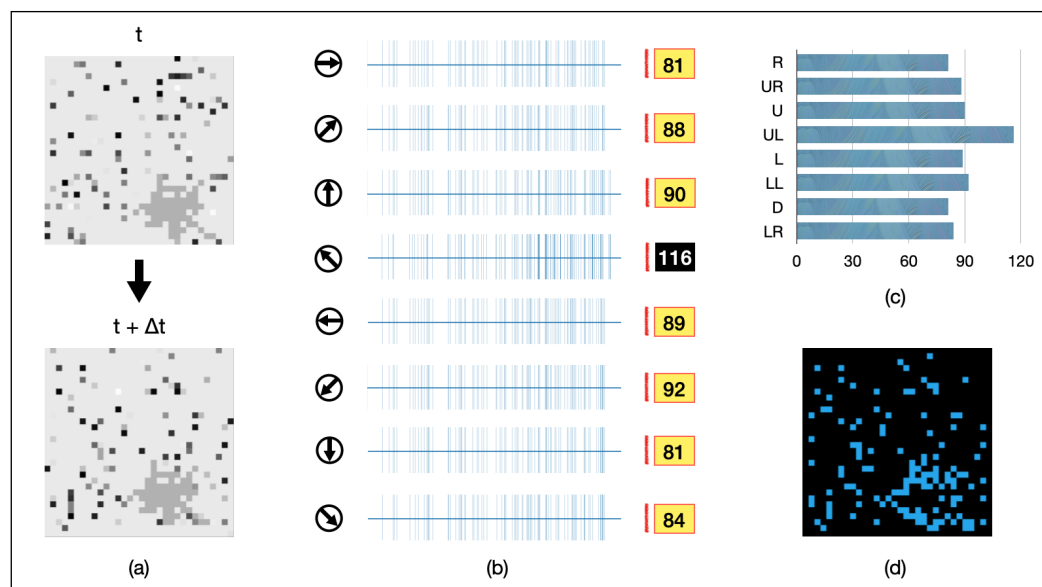


Figure 12. An AVS motion direction detection instance in dynamic background noise: (a) two successive frames of an upper leftward motion of a 64-pixel object; (b) activation map of the eight LMDNs during scan; (c) LMDNs activations are counted by a bar chart; (d) activated pixels by all the LMDNs during detection process are marked by blue.

The third type of complex environment, static full-image noise. From each theoretical image, a number of pixels at the same position are selected and replaced with an equal number of noise pixels. This type of noise not only has the characteristics of static background noise, but also obscures object pixels, causing clutter everywhere in the full image. Figure 13a shows an example of detection under static full-image noise. It shows two successive frames of a randomly shaped object consisting of 32-pixel points moving lower rightward in an image consisting of 1024 pixels in 32×32 . The uppermost pixel points at the same location in both images are covered by random grayscale noise points. It can be seen that the noise acts on the uppermost layer of the image and that the noise point locations do not change during the object's motion. Figure 13b depicts the number of times the LMDNs were activated, i.e., the response strength of each corresponding GMDN. LMDNs were activated 12, 6, 9, 6, 7, 14, 11, and 24 times during the time interval of Δt . The LMDNs preferring lower rightward motion were activated 24 times with a significantly higher activation bar density than any of the others. The sensitive direction of this neuron is consistent with the actual motion direction of the object, and thus a successful detection. The bar graph in Figure 13c records the response intensity of the GMDNs. The pixels activated by either neuron in Δt time are highlighted in blue in Figure 13d.

The fourth complex environment, dynamic full-image noise. From each theoretical image, a number of pixels at different positions are independently selected and replaced with an equal amount of noise pixels. This noise not only has the characteristics of static full-image noise, but also contains real-time changes. This type of noise has the characteristics of the three aforementioned noise types in parallel, and thus is the most complex noise type. Figure 14a shows an example of detection under static full-image noise. It shows two successive frames of a randomly shaped object consisting of 256-pixel points moving lower leftward in an image consisting of 1024 pixel points in 32×32 . Some pixel points at different locations of the two images are covered by random grayscale noise points. We can see that the noise acts on the uppermost layer of the images, and the noise point positions are dynamically changing during the object's motion. Figure 14b depicts the number of activations of LMDNs, i.e., the response strength of each corresponding GMDN. LMDNs were activated 113, 116, 118, 117, 119, 201, 138, and 127 times, respectively, in the time of Δt , where the LMDNs preferring the lower rightward motion were activated 201 times. The density of activation bars is significantly higher than any of the others, the sensitive

direction of this neuron is consistent with the actual motion direction of the object, and the detection is successful. The bar graph in Figure 14c records the response intensity of the GMDNs. The pixels activated by either neuron in Δt time are highlighted in blue in Figure 14d.



Figure 13. An AVS motion direction detection instance in static full-image noise: (a) two successive frames of a lower rightward motion of a 32-pixel object; (b) activation map of the eight LMDNs during scan; (c) LMDNs activations are counted by a bar chart; (d) activated pixels by all the LMDNs during detection process are marked by blue.

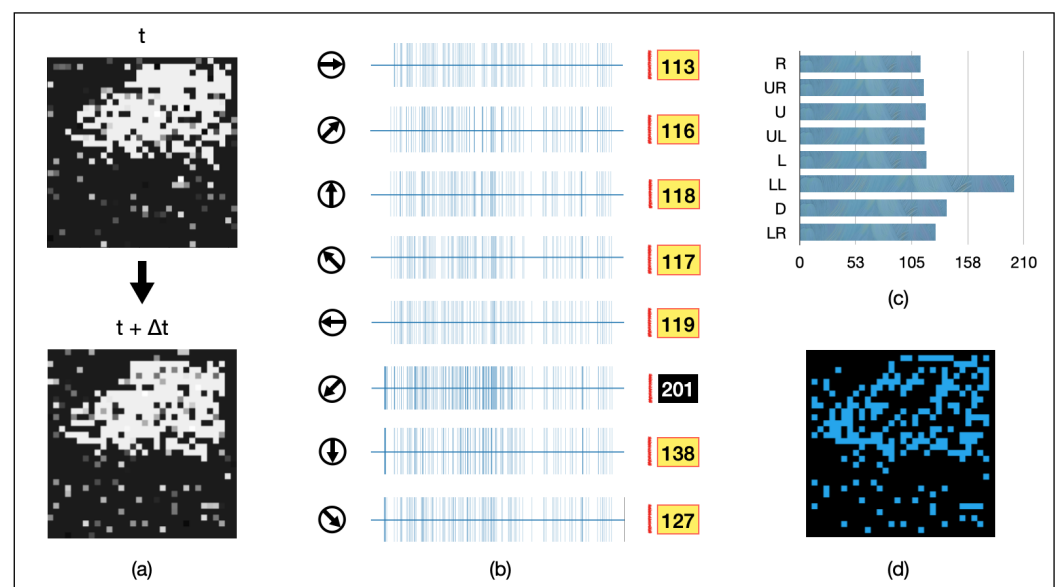


Figure 14. An AVS motion direction detection instance in dynamic full-image noise: (a) two successive frames of a lower leftward motion of a 256-pixel object; (b) activation map of the eight LMDNs during scan; (c) LMDNs activations are counted by a bar chart; (d) activated pixels by all the LMDNs during detection process are marked by blue.

The above four different types of noise are added to the noise-free object motion examples to generate four noise test sets to examine the noise immunity performance of AVS. The percentage of noise points amount to the total number of image pixels is set to 10%, 20%, and 30%, and the detection accuracy is counted. To further verify the

noise immunity performance of AVS, the CNN is also employed to complete the direction detection task in such four noisy environments for comparison.

For each of the above noises, 10,000 sets of motion instances are generated by computer simulation respectively. All other experimental conditions are the same as the CNN comparison experiments without noise. The detection accuracies of objects of different sizes in the four noise environments are calculated separately. The experimental results are shown in Tables 3–6.

Table 3. The experimental results of AVS and CNN under the static background noise.

Noise Type	10%		20%		30%	
Proportion	AVS	CNN	AVS	CNN	AVS	CNN
1-pixel	92.5%	13.18%	89%	13.37%	82%	13.13%
2-pixel	94.5%	13.15%	91.5%	13.08%	86.5%	12.82%
4-pixel	97.5%	16.40%	94%	13.83%	96.5%	13.35%
8-pixel	100%	17.78%	99.5%	15.53%	99%	15.70%
16-pixel	100%	20.65%	100%	18.65%	100%	15.67%
32-pixel	100%	26.03%	100%	21.82%	100%	18.35%
64-pixel	100%	32.68%	100%	27.70%	100%	23.70%
128-pixel	100%	42.70%	100%	33.20%	100%	27.80%
256-pixel	100%	55.20%	100%	42.47%	100%	37.65%
512-pixel	100%	69.70%	100%	55.88%	100%	52.22%
Mean	98.45%	30.75%	97.4%	25.55%	96.4%	23.04%

Table 3 statistically shows the detection accuracies of AVS and CNN under static background noise. In the 10% noise environment, AVS can achieve 100% detection accuracy in all situations when detecting objects of scale larger than 4 pixels; however, a small number of detection errors occur when the object scale is smaller than 8 pixels. This is due to the fact that under a certain amount of noise, the smaller the object size, the greater the relative interference of the noise. the same trend exists for CNN detection, i.e., the accuracy is higher for objects of larger size. However, the CNN achieves only 69.70% for the largest scale object of 512 pixels in the test set, which is just not as good as the AVS for the smallest scale 1-pixel object, i.e., 92.5%. As the object size decreases, the accuracy of CNN reduces rapidly, and the accuracy of 1-pixel object detection is only 13.18%. When the noise ratio increases to 20%, the accuracy of both methods reduce to some extent relative to a 10% noisy environment, with AVS having an all-100% error-free detection capability in tasks for scales larger than 8 pixels, while the accuracy falls from 100% to 99.5% in tasks for 8-pixel objects. The detection accuracy was reduced by 3.5%, 3%, and 3.5% in the tasks of 1, 2, and 4-pixel objects, respectively, while the accuracy of CNN showed a significant reduction in tasks of each object size except for 1 and 2-pixel objects. In the 30% noise condition, AVS still maintains all-100% error-free detection performance for objects larger than 8, while the accuracy reduction is 7%, 5% and 0.5% for 1, 2, and 8-pixel objects, respectively, relative to the 20% noise environment. Although the accuracy in doing the 4-pixel object task becomes lower in the 20% noise environment, this increase should be considered as a singular value due to experimental randomness, because the overall trend shows that the error-free detection range of AVS becomes narrower as the background noise proportion increases, and the accuracy becomes worse overall in tasks with detection errors. CNN, on the other hand, similarly becomes worse on all scales excluding the 8-pixel object. The small accuracy increase of 8-pixel objects should also be considered as a singular value due to experimental randomness. In the experiments with three different noise concentrations of static background noise, AVS maintains high accuracy performance of 100% for all objects larger than 16 pixels. The mean values of AVS are 98.45%, 97.4%, and 96.4%, which are much higher than CNN, i.e., 30.75%, 25.55%, and 23.04%, respectively. It is obvious from

the line graph of the mean values in Figure 15a that the accuracy of both methods reduces as the noise percentage becomes larger, but AVS significantly outperforms CNN in all cases.

Table 4. The experimental results of AVS and CNN under the dynamic background noise.

Noise Type	10%		20%		30%	
Proportion	AVS	CNN	AVS	CNN	AVS	CNN
1-pixel	29.00%	12.68%	18.50%	12.22%	19.00%	12.45%
2-pixel	37.00%	13.58%	27.50%	13.45%	24.00%	12.72%
4-pixel	61.00%	14.42%	30.00%	13.17%	29.50%	12.75%
8-pixel	78.50%	16.13%	46.50%	14.33%	39.00%	14.05%
16-pixel	93.00%	18.95%	78.00%	16.78%	62.00%	14.80%
32-pixel	99.50%	22.72%	96.00%	19.08%	83.50%	17.17%
64-pixel	100.00%	29.33%	100.00%	23.18%	98.50%	18.45%
128-pixel	100.00%	40.70%	100.00%	28.80%	99.50%	25.53%
256-pixel	100.00%	45.83%	100.00%	39.28%	100.00%	33.03%
512-pixel	100.00%	64.13%	100.00%	53.00%	100.00%	46.23%
Mean	79.80%	27.85%	69.65%	23.33%	65.50%	20.72%

Table 4 provides statistics on the detection accuracy of AVS and CNN under dynamic background noise. In the 10% noise environment, AVS can achieve 100% accuracy for all objects larger than 32 pixels; however, detection errors occur for objects smaller than 64 pixels. CNN is also easier to recognize the object's motion with larger object scales. Both the two methods had the lowest accuracy rates for the 1-pixel object, which is 29.00% and 12.68%, respectively; the mean accuracy values for the 10 different sizes of objects were 79.80% and 27.85%, respectively, with a decrease, compared to the static background noise, but AVS still significantly outperformed CNN. When the noise ratio increases to 20%, the detection accuracy of both methods reduce relative to the 10% noise environment, but AVS still maintains all-100% error-free detection capability in tasks for object scales larger than 32 pixels; in tasks for smaller scales, the accuracy decreases naturally. CNN, on the other hand, has decreased for all the object scales. AVS maintains all-100% error-free performance for those larger than 128 pixels in the 30% noise condition, while the accuracy decreases significantly for the smaller object, except for 1-pixel objects (which should be considered as singular values due to experimental randomness), relative to the 20% noise environment. The CNN similarly decreases at all object scales except for 1-pixel objects (which should be considered as singular values due to experimental randomness). In the three experiments with different concentrations of dynamic background noise, AVS maintains high accuracy performance of 100% for all the objects larger than 128 pixels. The mean accuracy of AVS is 79.80%, 69.65%, and 65.50%, respectively, which is much higher than that of CNN at 27.85%, 23.33%, and 20.72%. It is obvious from Figure 15b that the accuracy of both methods decreases as the noise percentage becomes larger, but AVS significantly outperforms CNN in all cases. Relative to static background noise, dynamic background noise produces stronger interference in the motion direction recognition task.

Table 5 statistically shows the detection accuracies of AVS and CNN under static full-image noise. In a 10% noise environment, AVS can achieve 100% detection accuracy for all objects larger than 8 pixels; however, a small number of detection errors occur for objects smaller than 16 pixels. CNN obtains the highest accuracy for the largest scale 512-pixel object in the test set, but it is only 55.15%, which is even far less than the accuracy of AVS when detecting 1-pixel objects, i.e., 77.5%. Both the two methods have the lowest accuracy for the task of 1-pixel objects, which is 77.5% and 13.22%, respectively; the mean accuracy values for 10 different scales of objects are 94.4% and 27.29%, respectively, with AVS significantly outperforming CNN. When the percentage of noise increases to 20%, the accuracy of both methods decreases relative to the 10% noisy environment, but AVS

still maintains an all-100% error-free detection capability in object detection tasks larger than 16 pixels; the accuracy decreases in tasks of smaller sizes. The AVS maintains all-100% error-free performance for objects larger than 64 pixels in the 30% noise condition, while the accuracy decreases in the smaller object tasks compared to the 20% noise environment, and the CNN also decreases in tasks at all object scales. In the three experiments with different concentrations of static full-image noise, AVS maintains a high accuracy performance of 100% for all objects larger than 64 pixels. the mean accuracy of AVS is 94.4%, 88.4%, and 83.35%, respectively, which is much higher than that of CNN at 27.29%, 21.35%, and 17.79%. It is obvious from Figure 15c that the accuracy of both methods decreases as the percentage of noise becomes larger, but AVS is significantly better than CNN in all cases. Static full-image noise is a stronger noise compared to static background noise.

Table 5. The experimental results of AVS and CNN under the static full-image noise.

Noise Type	10%		20%		30%	
Proportion	AVS	CNN	AVS	CNN	AVS	CNN
1-pixel	77.5%	13.22%	51%	12.98%	43%	12.80%
2-pixel	77.5%	13.60%	64%	13.20%	50.5%	12.42%
4-pixel	93%	15.18%	79%	14.53%	67%	12.67%
8-pixel	96%	17.93%	91.5%	15.38%	79.5%	13.88%
16-pixel	100%	20.03%	98.5%	16.97%	95%	14.80%
32-pixel	100%	24.28%	100%	18.83%	99%	17.22%
64-pixel	100%	30.87%	100%	21.57%	99.5%	18.28%
128-pixel	100%	36.35%	100%	27.17%	100%	22.63%
256-pixel	100%	46.25%	100%	32.88%	100%	25.38%
512-pixel	100%	55.15%	100%	39.95%	100%	27.77%
Mean	94.4%	27.29%	88.4%	21.35%	83.35%	17.79%

Table 6. The experimental results of AVS and CNN under the dynamic full-image noise.

Noise Type	10%		20%		30%	
Proportion	AVS	CNN	AVS	CNN	AVS	CNN
1-pixel	23%	12.87%	18.5%	13.23%	10%	12.32%
2-pixel	36%	13.27%	19.5%	12.75%	16%	13.02%
4-pixel	51.5%	13.75%	29%	13.55%	29.5%	13.47%
8-pixel	73.5%	15.40%	43.5%	14.40%	31%	12.87%
16-pixel	93.5%	17.57%	65.5%	15.60%	50%	13.50%
32-pixel	98.5%	22.13%	89.5%	16.63%	65.5%	14.33%
64-pixel	100%	27.53%	100%	19.85%	89.5%	16.53%
128-pixel	100%	31.48%	100%	21.28%	98.5%	17.42%
256-pixel	100%	41.87%	100%	30.53%	100%	22.57%
512-pixel	100%	46.45%	100%	33.53%	100%	28.43%
Mean	77.6%	24.23%	66.55%	19.14%	59%	16.45%

Table 6 statistically shows the detection accuracies of AVS and CNN under dynamic full-image noise. The accuracy of the CNN is also easier to recognize the movement of objects on a larger scale. Both the two methods had the lowest accuracy rates in the 1-pixel object task, at 23% and 12.87%, respectively; the mean accuracy values for the 10 different sizes of objects were 77.6% and 24.23%, respectively, which were lower compared to the three noise types mentioned above, but AVS still significantly outperformed CNN. When the noise ratio increases to 20%, the accuracy of both methods decreases relative to the 10% noise environment, but AVS still maintains all-100% error-free detection capability

for the task of objects larger than 32 pixels; in smaller object tasks, the accuracy decreases, while CNN decreases in all different sizes in experiments. AVS maintains all-100% error-free performance for objects larger than 128 pixels in the 30% noise condition, while the accuracy decreases significantly in the smaller object tasks compared to the 20% noise environment, and the CNN also decreases in tasks of all object sizes except for 2-pixel objects (which should be considered as singular values due to experimental randomness). In three experiments with different concentrations of dynamic full-image noise, AVS maintains high accuracy. In three experiments with different concentrations of dynamic full-image noise, AVS maintains high performance of accuracy of 100% for objects larger than 128 pixels. the mean accuracy values of AVS are 77.6%, 66.55%, and 59%, respectively, which is much higher than that of CNN at 24.23%, 19.14%, and 16.45%. It is obvious from the mean value line graph in Figure 15d that the accuracy of both methods decreases as the noise percentage gets larger, but AVS significantly outperforms CNN in all cases. Relative to the first three types of noise, the dynamic full-image noise produces the strongest interference in the motion direction recognition task. In different concentration environments of the four types of noise, the accuracy of both detection methods decreases gradually with the decrease of object size.

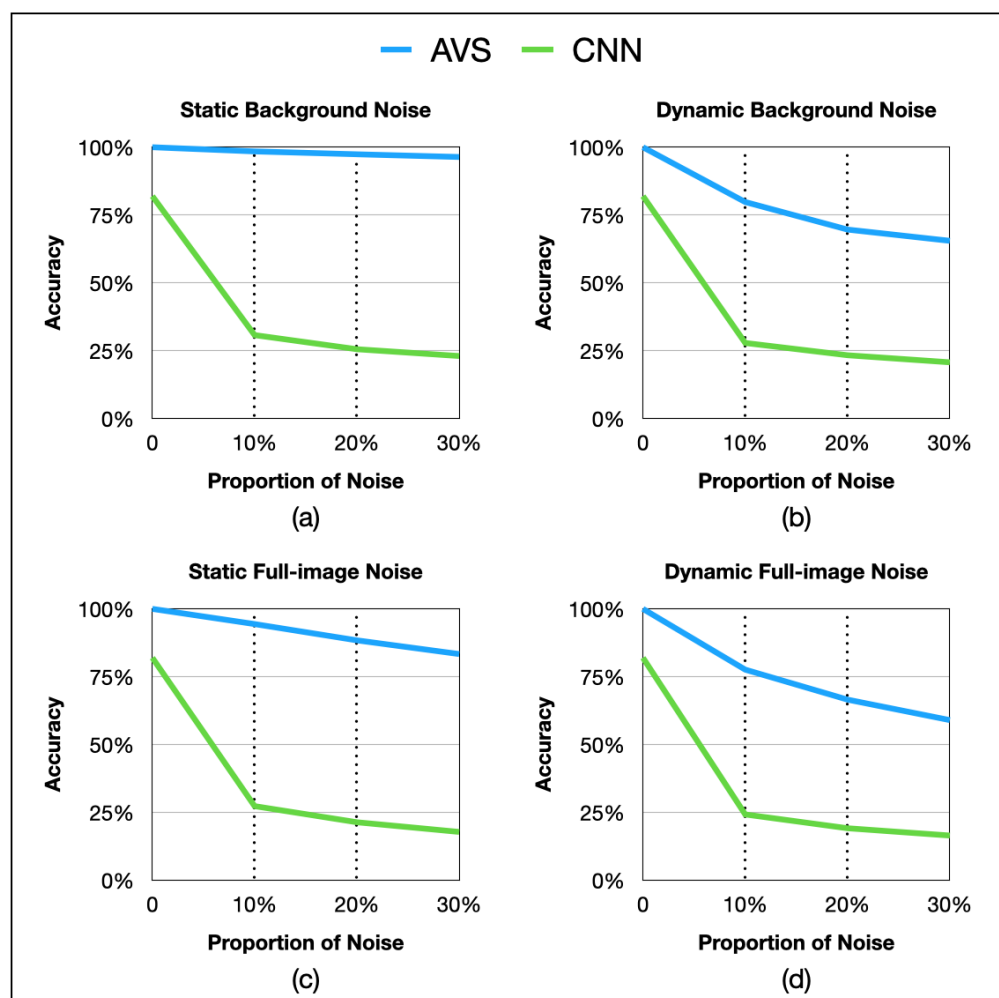


Figure 15. The motion direction detection accuracy trends of AVS and CNN to progressive proportions of different noise types: (a) accuracy trends in static background noise; (b) accuracy trends in dynamic background noise; (c) accuracy trends in static full-image noise; (d) accuracy trends in dynamic full-image noise.

Figure 15 depicts the relationship between different noise ratios and the detection accuracy of the two methods under the four types of noise environments with line graphs, respectively. From the figure, it can be seen that for both methods, the impact of static noise on detection difficulty is smaller than that of dynamic noise under the same other noise conditions, and the impact of background noise is smaller than that of full-image noise. However, when compared to CNN for each type of noise, AVS demonstrates superior noise immunity performance and also a significantly higher noise-free detection capability.

The experimental results show that AVS achieves very high detection accuracy and robustness under both noise-free and four different types of noise compared to the CNN method. At the same time, since the design of AVS is based on the real function of each corresponding cell in the retina to reproduce the physiological collaboration phenomenon of the relevant direction-sensitive pathway, there is no learning optimization process, no teaching data import, and no need to introduce any parameters in AVS. Compared with CNNs, which require a huge model parameter optimization task for a large amount of data during the learning process, AVS has the inherent advantage of being an extremely efficient motion direction detector even if the learning process is skipped, thus saving significant computational resources. This is because AVS has a strong physiological background, which makes it fully interpretable; whereas CNN has been considered a black-box optimizer with a lack of interpretability. Moreover, in terms of hardware implementation, AVS is designed based entirely on the physiological computational properties of relative neurons, requiring only three simple operation modes: comparison, summation, and logic operations, and is therefore extremely easy to implement in hardware, allowing for a very high computational speed while greatly liberating the scope of application scenarios. In summary, AVS has four important characteristics of high effectivity, robustness, efficiency, and rationality at the same time.

4. Discussion and Conclusions

The aim of this study is to construct a novel computer vision technique using the known cellular functional properties of the retinal direction-selective neural pathway. It also provides an innovative explanation for the mechanism underlying the operation of the DS vertical pathway in mammalian visual system to help understand the functional principles of DSGCs and the visual system in general. We propose a novel structure and principle conjecture for the entire retinal direction-selective pathway, construct a new two-dimensional eight-direction motion detection mechanism and verify its effectiveness, efficiency, and robustness using the following key knowledge:

1. BCs have ON-OFF response feedback mechanisms that can respond instantly to the changing phenomenon of local light sources;
2. HCs have asymmetric lateral connection structures and inhibitory signal feedback communication mechanism, which can complete the exchange of light information between the central PC and the surrounding PCs;
3. The dendritic neurons of the DS pathway in the retina have nonlinear computational properties;
4. According to neuroscientific knowledge, neurons can only complete extremely simple computations;
5. The existence of local receptive fields;
6. The superposition of excitation from local to global neurons;
7. A known highly biologically sound dendritic neuron model for modeling the DS pathway;
8. Specific neurons in the brain cortex respond more strongly to specific directions of motion occurring in the receptive field than to other directions.

In this mechanism, the judgment of global motion direction is determined by the activation strength of GMDNs. The response of GMDNs in the experiments is highly consistent with physiological activation phenomena actual in the mammalian retina (LMDN) and brain cortex (GMDN); thus, we have reasons to believe that this mechanism can provide a new cognitive perspective for the mammalian DS pathway as well as the visual system.

Additionally, except for the LMDN and GMDN sensitive to the present global motion direction, the other LMDNs and GMDNs are also activated during detection. LMDNs record the occurrence of different sensitive directions for each pixel and GMDNs record the total intensity of different directions in the global receptive field. Therefore, not only the main motion direction in the whole global receptive field but that in any local area can be derived by the GMDN calculation method. In other words, mammals can supervise motion directions occurring in any area of the visual field, except that the global motion direction is easier to notice because the correlated neurons are most strongly activated.

Experiments have shown that AVS performed excellent resistance ability to interference under a wide range of noise conditions while detecting the motion direction occurs noise-free grayscale images including single, multi-point, and large objects with no error. AVS makes correct motion direction determination regardless of the shape, size, and location of the moving object in the global receptive field. Its structure is a collection of dendritic neuron models that can detect the motion direction of objects in the global receptive field by only three very simple operations: comparison, summation, and logic operations, making AVS extremely easy to be hardwareized. This confirms the fact that the brain can perform complex tasks by performing simple nonlinear operations on a large number of neurons with dendrite, and also brings us, for example, a very low device threshold, very fast detection speed, and a very wide range of application scenarios. Coupled with the structural independence of the eight LMDNs, the mechanism allows for parallel operations to further achieve faster detection speed. In addition, the edge-sensitive property of the mechanism also naturally matches the visual properties of the brain. Therefore, AVS is in high agreement with the relevant physiological experiments in terms of various details of structural functions, as well as in terms of its high accuracy, high anti-interference ability, and various other characteristics (e.g., computational approach, computational efficiency, edge sensitivity, and output intensity pattern) exhibited during the detection process. This suggests that AVS can be a useful forward-looking approach to the current cognitive bottleneck of the DS pathway faced by researchers. It is also a fresh way of thinking to analyzing the mechanisms and principles of DS neural pathways and even the whole brain neural networks. At the same time, AVS also proposes a new and efficient method for motion direction detection.

The AVS used in this study is also compatible to solve the problem of object motion direction detection in two-dimensional eight-directional black-and-white binary conditions. In the future, the mechanism can be extended in the following ways. First, the grayscale motion detection mechanism is an indispensable basis for developing a color motion detection mechanism; designing the color motion direction detection mechanism to fully implement the retinal motion direction detection function and to cope with the widest range of color data in the real world; second, for the other additional motion directions, the current version of AVS can only output a result in the given eight directions. However, by adjusting and expanding the structure and scale of AVS to enable it to detect more types of motion directions (not just the basic eight). The detail method is that expand the 3×3 local receptive field to larger. For example, 5×5 for 16 directions and 7×7 for 24 directions; third, expanding it to direction detection mechanism in the three-dimensional environment in order to further implement the human eye stereoscopic function; fourth, the velocity or acceleration detection mechanism can also be implemented based on the current AVS. Fifth, although the current AVS does not require learning and maintains a high level of interpretability, we do not rule out developing a learnable AVS and trying to apply it to more than just direction-selective tasks. For learnable AVS, the existing AVS can provide an excellent set of initial parameters to greatly improve the learning efficiency. With AVS as the cornerstone, all these future works will ensure high efficiency while being highly physiologically consistent and interpretable.

Author Contributions: Conceptualization, S.T., Z.T. and Y.T.; methodology, S.T., Z.T. and Y.T.; software, S.T., B.L., Z.Z. and R.I.; validation, S.T., Z.T. and Y.T.; writing—original draft preparation, S.T. and R.I.; writing—review and editing, S.T., Z.T. and Y.T.; visualization, S.T. and Z.Z.; supervision, Z.T. and Y.T.; project administration, Z.T. and Y.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JSPS KAKENHI Grant No. 19K12136.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: the source codes of the AVS-grayscale used in experiments can be download from this link: <https://github.com/terrysc/AVS-grayscale>. If there's any problem, please contact the authors.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism

References

1. Sporns, O. *Networks of the Brain*; MIT Press: Cambridge, MA, USA, 2016.
2. Fiske, S.T.; Taylor, S.E. *Social Cognition*; McGraw-Hill Book Company: New York, NY, USA, 1991.
3. Medina, J.J.; Hanlon, D. *12 Principles for Surviving and Thriving at Work, Home and School*; Pear Press: Seattle, WA, USA, 2009.
4. Viviani, P.; Aymoz, C. Colour, form, and movement are not perceived simultaneously. *Vis. Res.* **2001**, *41*, 2909–2918.
5. Mauss, A.S.; Vlasits, A.; Borst, A.; Feller, M. Visual circuits for direction selectivity. *Annu. Rev. Neurosci.* **2017**, *40*, 211.
6. Vaney, D.I.; Sivy, B.; Taylor, W.R. Direction selectivity in the retina: Symmetry and asymmetry in structure and function. *Nat. Rev. Neurosci.* **2012**, *13*, 194–208.
7. Hamilton, N.R.; Scasny, A.J.; Kolodkin, A.L. Development of the vertebrate retinal direction-selective circuit. *Dev. Biol.* **2021**, *477*, 273–283.
8. Ohki, K.; Chung, S.; Ch'ng, Y.H.; Kara, P.; Reid, R.C. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* **2005**, *433*, 597–603.
9. Lipin, M.Y.; Taylor, W.R.; Smith, R.G. Inhibitory input to the direction-selective ganglion cell is saturated at low contrast. *J. Neurophysiol.* **2015**, *114*, 927–941.
10. Sethuramanujam, S.; McLaughlin, A.J.; DeRosenroll, G.; Hoggarth, A.; Schwab, D.J.; Awatramani, G.B. A central role for mixed acetylcholine/GABA transmission in direction coding in the retina. *Neuron* **2016**, *90*, 1243–1256.
11. Chapot, C.A.; Euler, T.; Schubert, T. How do horizontal cells 'talk' to cone photoreceptors? Different levels of complexity at the cone–horizontal cell synapse. *J. Physiol.* **2017**, *595*, 5495–5506.
12. Wei, W.; Feller, M.B. Organization and development of direction-selective circuits in the retina. *Trends Neurosci.* **2011**, *34*, 638–645.
13. Srinivasan, M. Generalized gradient schemes for the measurement of two-dimensional image motion. *Biol. Cybern.* **1990**, *63*, 421–431.
14. Hassenstein, B.; Reichardt, W. Systemtheoretische analyse der zeit-, reihenfolgen-und vorzeichenauswertung bei der bewegungsperzeption des rüsselkäfers chlorophanus. *Z. Naturforschung B* **1956**, *11*, 513–524.
15. Reichardt, W. Autocorrelation, a principle for evaluation of sensory information by the central nervous system. In *Symposium on Principles of Sensory Communication 1959*; MIT Press: Cambridge, MA, USA, 1961; pp. 303–317.
16. Reichardt, W. Evaluation of optical motion information by movement detectors. *J. Comp. Physiol. A* **1987**, *161*, 533–547.
17. Van Santen, J.P.; Sperling, G. Elaborated reichardt detectors. *JOSA A* **1985**, *2*, 300–321.
18. Borst, A.; Egelhaaf, M. Principles of visual motion detection. *Trends Neurosci.* **1989**, *12*, 297–306.
19. Adelson, E.H.; Bergen, J.R. Spatiotemporal energy models for the perception of motion. *Josa a* **1985**, *2*, 284–299.
20. Euler, T.; Detwiler, P.B.; Denk, W. Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature* **2002**, *418*, 845–852.
21. Reiff, D.F.; Plett, J.; Mank, M.; Griesbeck, O.; Borst, A. Visualizing retinotopic half-wave rectified input to the motion detection circuitry of Drosophila. *Nat. Neurosci.* **2010**, *13*, 973–978.

22. Elyada, Y.M.; Haag, J.; Borst, A. Different receptive fields in axons and dendrites underlie robust coding in motion-sensitive neurons. *Nat. Neurosci.* **2009**, *12*, 327–332.
23. Poleg-Polsky, A.; Diamond, J.S. Imperfect space clamp permits electrotonic interactions between inhibitory and excitatory synaptic conductances, distorting voltage clamp recordings. *PLoS ONE* **2011**, *6*, e19463.
24. Briggman, K.L.; Helmstaedter, M.; Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **2011**, *471*, 183–188.
25. Schachter, M.J.; Oesch, N.; Smith, R.G.; Taylor, W.R. Dendritic spikes amplify the synaptic signal to enhance detection of motion in a simulation of the direction-selective ganglion cell. *PLoS Comput. Biol.* **2010**, *6*, e1000899.
26. Hausselt, S.E.; Euler, T.; Detwiler, P.B.; Denk, W. A dendrite-autonomous mechanism for direction selectivity in retinal starburst amacrine cells. *PLoS Biol.* **2007**, *5*, e185.
27. Han, M.; Todo, Y.; Tang, Z. Mechanism of Motion Direction Detection Based on Barlow's Retina Inhibitory Scheme in Direction-Selective Ganglion Cells. *Electronics* **2021**, *10*, 1663.
28. Tang, C.; Todo, Y.; Ji, J.; Tang, Z. A novel motion direction detection mechanism based on dendritic computation of direction-selective ganglion cells. *Knowl.-Based Syst.* **2022**, *241*, 108205.
29. Hubel, D.H. Exploration of the primary visual cortex, 1955–78. *Nature* **1982**, *299*, 515–524.
30. Hubel, D.H.; Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **1962**, *160*, 106.
31. Hartline, H.K. The receptive fields of optic nerve fibers. *Am. J. Physiol.-Leg. Content* **1940**, *130*, 690–699.
32. Barlow, H.; Levick, W.R. The mechanism of directionally selective units in rabbit's retina. *J. Physiol.* **1965**, *178*, 477.
33. Borst, A.; Euler, T. Seeing things in motion: models, circuits, and mechanisms. *Neuron* **2011**, *71*, 974–994.
34. Todo, Y.; Tang, Z.; Todo, H.; Ji, J.; Yamashita, K. Neurons with multiplicative interactions of nonlinear synapses. *Int. J. Neural Syst.* **2019**, *29*, 1950012.
35. Wasserman, P.D. *Neural Computing: Theory and Practice*; Van Nostrand Reinhold Co.: Washington, DC, USA, 1989.
36. Koch, C.; Poggio, T.; Torre, V. Nonlinear interactions in a dendritic tree: localization, timing, and role in information processing. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 2799–2802.
37. Koch, C.; Poggio, T.; Torre, V. Retinal ganglion cells: a functional interpretation of dendritic morphology. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1982**, *298*, 227–263.
38. Taylor, W.R.; He, S.; Levick, W.R.; Vaney, D.I. Dendritic computation of direction selectivity by retinal ganglion cells. *Science* **2000**, *289*, 2347–2350.
39. Segev, I. Sound grounds for computing dendrites. *Nature* **1998**, *393*, 207–208.
40. Todo, Y.; Tamura, H.; Yamashita, K.; Tang, Z. Unsupervised learnable neuron model with nonlinear interaction on dendrites. *Neural Netw.* **2014**, *60*, 96–103.
41. Hartline, H.K. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *Am. J. Physiol.-Leg. Content* **1938**, *121*, 400–415.
42. Barlow, H.B.; Hill, R.M. Selective sensitivity to direction of movement in ganglion cells of the rabbit retina. *Science* **1963**, *139*, 412–414.
43. Liu, X.; Hirano, A.A.; Sun, X.; Brecha, N.C.; Barnes, S. Calcium channels in rat horizontal cells regulate feedback inhibition of photoreceptors through an unconventional GABA-and pH-sensitive mechanism. *J. Physiol.* **2013**, *591*, 3309–3324.
44. Puller, C.; Haverkamp, S.; Neitz, M.; Neitz, J. Synaptic elements for GABAergic feed-forward signaling between HII horizontal cells and blue cone bipolar cells are enriched beneath primate S-cones. *PLoS ONE* **2014**, *9*, e88963.
45. Hoon, M.; Sinha, R.; Okawa, H.; Suzuki, S.C.; Hirano, A.A.; Brecha, N.; Rieke, F.; Wong, R.O. Neurotransmission plays contrasting roles in the maturation of inhibitory synapses on axons and dendrites of retinal bipolar cells. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12840–12845.
46. Vardi, N.; Sterling, P. Subcellular localization of GABAA receptor on bipolar cells in macaque and human retina. *Vis. Res.* **1994**, *34*, 1235–1246.
47. Vardi, N.; Zhang, L.L.; Payne, J.A.; Sterling, P. Evidence that different cation chloride cotransporters in retinal neurons allow opposite responses to GABA. *J. Neurosci.* **2000**, *20*, 7657–7663.
48. Duebel, J.; Haverkamp, S.; Schleich, W.; Feng, G.; Augustine, G.J.; Künner, T.; Euler, T. Two-photon imaging reveals somatodendritic chloride gradient in retinal ON-type bipolar cells expressing the biosensor Clomeleon. *Neuron* **2006**, *49*, 81–94.
49. VanLeeuwen, M.; Fahrenfort, I.; Sjoerdsma, T.; Numan, R.; Kamermans, M. Lateral gain control in the outer retina leads to potentiation of center responses of retinal neurons. *J. Neurosci.* **2009**, *29*, 6358–6366.
50. Van Hateren, H. A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells. *J. Vis.* **2005**, *5*, 5.
51. Van Hateren, J. A model of spatiotemporal signal processing by primate cones and horizontal cells. *J. Vis.* **2007**, *7*, 3.
52. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324.
53. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.