Machine Intelligence Research
www.mi-research.net

19(5), October 2022, 412-424
DOI: 10.1007/s11633-022-1370-z

Towards a New Paradigm for Brain-inspired Computer Vision

Xiao-Long Zou[1,2] Tie-Jun Huang[1,3,4] Si Wu[1,2,4]

[1] Beijing Academy of Artificial Intelligence, Beijing 100084, China

[2] School of Psychology and Cognitive Sciences, IDG/McGovern Institute for Brain Research, Center for Quantitative Biology, PKU-Tsinghua Center for Life Sciences, Peking University, Beijing 100084, China

[3] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, Beijing 100871, China

[4] Institute for Artificial Intelligence, Peking University, Beijing 100871, China

Abstract: Brain-inspired computer vision aims to learn from biological systems to develop advanced image processing techniques. However, its progress so far is not impressing. We recognize that a main obstacle comes from that the current paradigm for brain-inspired computer vision has not captured the fundamental nature of biological vision, i.e., the biological vision is targeted for processing spatio-temporal patterns. Recently, a new paradigm for developing brain-inspired computer vision is emerging, which emphasizes on the spatio-temporal nature of visual signals and the brain-inspired models for processing this type of data. In this paper, we review some recent primary works towards this new paradigm, including the development of spike cameras which acquire spiking signals directly from visual scenes, and the development of computational models learned from neural systems that are specialized to process spatio-temporal patterns, including models for object detection, tracking, and recognition. We also discuss about the future directions to improve the paradigm.

Keywords: Brain-inspired computer vision, spatio-temporal patterns, object detection, object tracking, object recognition.

Citation: X. L. Zou, T. J. Huang, S. Wu. Towards a new paradigm for brain-inspired computer vision. *Machine Intelligence Research*, vol.19, no.5, pp.412–424, 2022. http://doi.org/10.1007/s11633-022-1370-z

1 Introduction

Nowadays, computer vision or machine vision, represented especially by deep convolutional neural networks (DCNNs), has achieved great success in many vision tasks[1, 2]. Compared to biological vision, however, computer vision is still lagging far behind in both performances and variety of capabilities[3–5]. For instance, DCNNs, which mainly mimic the feedforward and hierarchical structure of the ventral pathway of biological vision, has achieved an extremely high accuracy in image classification[1, 2, 6], but in other tasks, such as video analysis and imaging understanding, they are still far from satisfactory[7]. Thus, learning from biological vision, the so-called brain-inspired computer vision, is still a promising and efficient way to speed up the development of computer vision[4, 8, 9].

Although the importance of developing brain-inspired computer vision has been widely recognized[8, 10], up to now, we have not achieved any really breakthrough in the field that can match the achievement of AlphaGo to GO game[11] or AlphaFold to protein prediction[12]. So, what is the obstacle in the development? We identify that an important issue that is missed in the current practice of brain-inspired computer vision is the ignorance of a key nature of biological vision, i.e., biological vision is targeted on processing spatio-temporal patterns[10, 13]. This is fundamentally different from static images which DCNNs are good at. Shortly speaking, the characteristic of having both spatial and temporal structures is the nature of neural signals in every part of the brain[14, 15]. At the beginning stage of acquiring visual information from the external world, the signals received by retina are in the form of continuous optical flow; these signals are converted into spike trains by retinal ganglion cells, which are subsequently transmitted layer by layer to the visual cortex, where the visual input is integrated with spikes from other cortical regions conveying the prior knowledge or memory[16]; eventually, the visual information is extracted in the form of continuous neuronal responses (see illustration in Fig. 1). The whole process is very complicated with many fine details remaining unknown, nevertheless, the fact that the visual system computes spatio-

Review
Special Issue on Brain-inspired Machine Learning
Manuscript received May 4, 2022; accepted August 19, 2022
Recommended by Associate Editor Gang Pan
Colored figures are available in the online version at https://link.springer.com/journal/11633

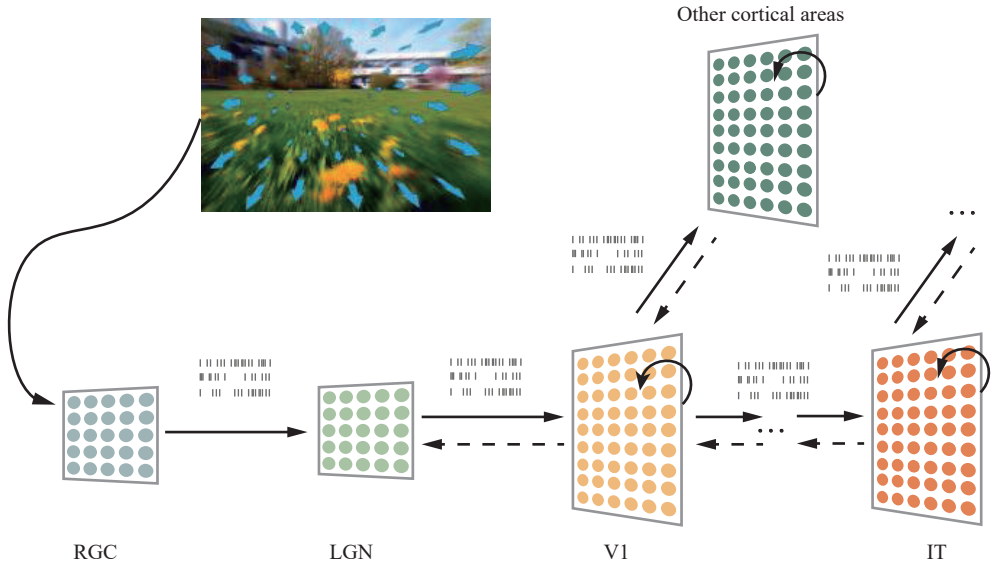


Fig. 1 A schematic of the hierarchical processing of spatio-temporal patterns in the visual system in the form of spike trains. The visual information (in the form of optical flow) of the external world is received by the retina, converted into spike trains by retinal ganglion cells (RGCs), and then processed layer by layer through lateral geniculate nucleus (LGN), V1, V2, V4, IT, etc. There are rich feedback and recurrent interactions within and between visual cortical areas, and interactions with other cortical regions.

temporal patterns in the form of spike trains is fully validated by experiments[17].

In recognition of the aforementioned difference between machine vision and biological vision, a new paradigm which captures this fundamental difference is emerging for developing brain-inspired computer vision. Specifically, in such a paradigm, from beginning the visual information in the external world is expressed in the form of spike trains, which is subsequently processed by computation models inspired by the biological system. Notably, the current popular spiking neural networks (SNNs) in the field do not fulfill this goal[18], and they normally miss many features in biological systems that are crucial for processing spatio-temporal patterns efficiently, such as the differentiation of excitatory and inhibitory neurons, the stochastic neuronal firing[19], the recurrent and feedback interactions between neurons[15], the short-term plasticity of synapses[20], etc.

The organization of the paper is as follows. In Section 2, we will first review the recent develop of spike cameras[21, 22], in particular Vidar[23], which provides a way to represent visual inputs by spike trains. In Section 3, we will review three brain-inspired computational models, which can implement rapid signal detection[24], anticipative object tracking[25], and spatio-temporal pattern recognition[26], respectively based on spike trains. Finally, in Section 4, we will discuss about the future development of brain-inspired computer vision.

2 Brain-inspired cameras for sensing visual information

As described above, to develop brain-inspired computer vision, it is critical to represent visual inputs starting from the sensory level to draw analogies with biological systems. In recent years, brain-like sensing devices have been developed rapidly, which are able to sense visual scenes with high temporal and spatial resolutions, and they convert light signals directly into spike trains. Here, we introduce two of them: one is dynamical vision sensor (DVS)[27, 28], and the other is spike camera[21, 22], e.g., Vidar[23]. Both of them were inspired by the retina system of the biological brain.

DVS is a retina-like sensing device. A traditional camera perceives image information frame by frame at a fixed temporal frequency, such as 20 fs, and each frame contains the integrated luminance information of the image over the frame interval, while the variance of image luminance between frames is ignored. DVS works in a different way. It asynchronously senses the luminance change at each pixel in the image and outputs a stream of spike events. The output spikes are represented in an address-event manner. Each spike signal contains four elements $\langle x, y, t, p\rangle$, which are the horizontal position x, the vertical position y, the spiking moment t, and a variable p of binary values indicating the direction of luminance change. In other words, $\langle x, y, t\rangle$ describe the spatio-temporal position of a spike event, and p describes the way of luminance change. For DVS, a spike is triggered only when the luminance change at the corresponding pixel exceeds a defined threshold, and DVS typically outputs spikes in a spatially sparse and temporally discrete manner. Since DVS only records the relative change of luminance, it is more suitable to sense the motion information in a visual scene, while the detailed texture information of the image is largely ignored. Compared with traditional cameras, DVS has the advantages of broader dynamical range, higher temporal resolution, lower energy

consumption, higher pixel band, etc.[28] As a brain-inspired method, DVS has been used recently to simulate a three-layer retina system and implemented a simplified photo receptor-bipolar cell-ganglion cell pathway[29].

If we roughly regard that DVS is mimicking the peripheral region of the retina, then Vidar is mimicking the fovea of the retina[23]. Vidar builds a pixel processing array, which contains an analog-to-digital converter (ADC) and an accumulator at each pixel position. ADC converts light intensity into voltage, and this voltage signal is sent to the corresponding accumulator, as shown in Fig. 2. The accumulator integrates the input signal and outputs a spike when the accumulated voltage exceeds a defined threshold, and then resets. The whole process can be regarded as a simplified modeling of photo receptor-bipolar cell-ganglion cell pathway. An ADC simulates a photoreceptor, and an accumulator simulates an integrate-and-fire neuron. Vidar perceives the light intensity information, rather than the change of luminance as done by DVS. The larger the light intensity at a pixel position is, the larger the output of the corresponding ADC generates; consequently, the corresponding accumulator can reach the threshold more rapidly, and generate spikes more frequently. In other words, the light intensity at each pixel is represented by the firing frequency of the corresponding neuron. In Vidar, each accumulator outputs spikes and resets in an asynchronous manner. The spike trains generated by Vidar can then be used to reconstruct the texture features in the visual scene. Vidar is designed to sample data in a temporal resolution as high as millisecond[21]. In theory, we can obtain the light intensity information of the image at any time from the recorded spike trains.

3 Computational models for brain-inspired computer vision

In the above, we have introduced brain-inspired sensing devices which transform visual inputs directly into spike trains. In this section, we further discuss models for extracting information efficiently from these spike trains. In AI, many image processing algorithms have been proposed, however, these algorithms mainly focus on processing static images, rather than spatio-temporal patterns. Simply transforming artificial neurons in a neural network into spiking neurons does not help much. To develop efficient computational models for processing spatio-temporal patterns, we should learn from biological vision, as the latter is evolved over millions of years to perform this task efficiently[30].

In essence, computer vision needs to perform three fundamental functions: object detection, object tracking, and object recognition. In the below, we will review three type of models inspired by neural systems that are able to perform these three tasks, respectively.

3.1 A brain-inspired model for fast object detection

3.1.1 Biology background

The ability to respond to external stimuli rapidly is critical for animals to survive in natural environments, e.g., to escape from predators. Over millions of years, the biological brain has evolved to process visual information extremely fast. For examples, the response delay of neurons in macaques' visual cortex is only about tens of milliseconds[31], and human brain can complete complex visual scene analysis in around 150 ms[32]. Computational neuroscience studies have revealed that the capability of fast processing information in the brain is largely attributed to that neural circuits are excitation and inhibition (E-I) balanced, i.e., neurons in a neural circuit on average receive E-I balanced currents. The idea of E-I balanced network was originally proposed by Vreeswijk and Sompolinsky to explain the irregular firings of cortical

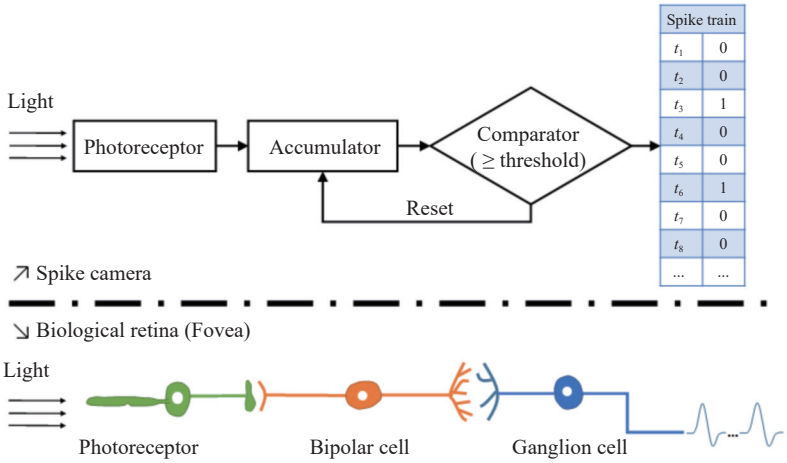


Fig. 2 Workflow of the spike camera Vidar. It contains an analog-to-digital converter (ADC) converting light intensity into voltage, an accumulator integrating inputs, and a comparator generating spike by comparing whether the accumulated voltage exceeds a defined threshold. Because of resetting after spiking, Vidar senses the light intensity change of a visual scene. Vidar can be regarded as modeling the photo receptor-bipolar cell-ganglion cell pathway in the fovea of retina, adapted from [23].

neurons widely observed in the cortex[33, 34]. From the computational point of view, they argued that the E-I balanced network has the advantage of responding to external inputs extremely fast, since neuronal activities in such a network are very noisy. The underlying mechanism can be intuitively understood as follows. Without noise, the reaction time of neurons is restricted by the membrane time constant (see illustration in Fig. 3(b)). In an E-I balanced network, however, the system is at a chaotic state with neurons' membrane potentials widely distributed (Fig. 3(b)). As a result, no matter how small the external input change is, there are always a number of neurons responding fast to detect this signal. Thus, at the neuron ensemble level, the network can detect external signals extremely fast. So far, a large amount of experimental data has confirmed that E-I balance is a general property of neural systems[35–38].

3.1.2 The computational model

Since E-I balanced neural circuits are the substrates of the brain to realize fast computation, it is natural to ask whether such type of model can be used in brain-inspired computer vision. Recently, Tian et al.[24] explored this issue and demonstrated the feasibility. In the below, we briefly introduced their work.

Tian et al.[24] firstly studied a recurrent network with homogeneous connections, as shown in Fig. 3(a). The network has a large size N, with $N_E = q_E N$ and $N_I = q_I N$ being the numbers of excitatory and inhibitory neurons, respectively. For simplicity of theoretical analysis, they consider no-leaky integrate-and-fire neurons. Each neuron receives recurrent inputs from other neurons and an external input, whose dynamics is written as

$$\tau_a \frac{\mathrm{d}v_{a,i}}{\mathrm{d}t} = \sum_{b=E,I} \sum_j J_{ij}^{ab} \sum_k \frac{1}{\tau_{b,s}} \mathrm{e}^{-(t-t_{j,k})/\tau_{b,s}} + f_a u \quad a = E, I \tag{1}$$

where the subscript a denotes the neuron population, with $a = E$ or $a = I$ representing neurons are excitatory or inhibitory, respectively. τ_a is the integration time constant of neurons in population a. $v_{a,i}$ is the membrane potential of neuron i in population a. $J_{i,j}^{a,b}$ denotes the connection between neuron j in population b and neuron i in population a. If a connection exists between them, the connection strength is set to be $J_{i,j}^{a,b} = j_{ab}/\sqrt{N}$; otherwise $J_{i,j}^{a,b} = 0$. $\tau_{b,s}$ is the synaptic time constant of population b, and $t_{j,k}$ is the moment of the kth spike of neuron j. The probability that neuron j in population b connects to neuron i in population a is $p_{a,b}$ for all i, j. $u(t)$ denotes the external forward input, i.e., the signal, whose strength is controlled by f_a. When $v_{a,i}$ reaches a threshold θ, the neuron generates a spike and $v_{a,i}$ is reset to v_0.

By using the mean-field approximation, Tian et al.[24]

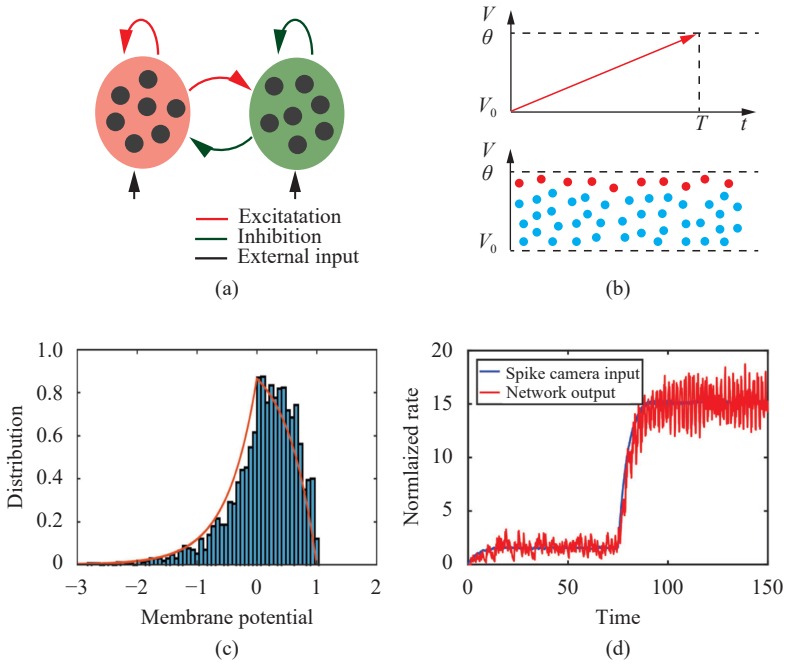


Fig. 3 An E-I balanced neural network for rapid object detection: (a) An example of E-I balanced neural network, in which a neuron receives balanced excitatory and inhibitory recurrent inputs. (b) Illustrating the working mechanism. Upper panel: the reaction time of a non-leaky integrate-and-fire neuron is restricted, which takes time $T = (\theta - v_0)\tau/I$ to reach the threshold θ starting from the resting state v_0, where τ is the membrane time constant and I the input current. Lower panel: in an E-I balanced network, the membrane potentials of neurons are widely distributed, such that neurons whose potentials are close to the firing threshold (red dots) can always respond rapidly to input changes. (c) Stationary distribution of neuronal membrane potentials in an E-I balanced network when the external input is a constant. The red curve is the theoretical prediction and the blue histogram the simulation result. (d) An E-I balanced network can almost instantly track the change of external inputs generated by a spike camera. Figs. 3(b)–3(d) adapted from [24].

derived the condition for E-I balance in the network, which is $f_E/f_I > W_{EI}/W_{II} > W_{EE}/W_{IE}$, where $W_{a,b} = p_{ab}j_{ab}q_{bb}$. Under this condition, the network receiving spatially uniform inputs maintains a stable distribution of neurons' membrane potentials (see Fig. 3(c)). By solving the corresponding Fokker-Planck equation, the stable distribution is obtained as

$$p_a(v) = \begin{cases} \dfrac{1}{\theta}\left[1-\exp(-2\tau_a\beta_a)\right]\exp\left(\dfrac{2\tau_a v}{\beta_a}\right), & \text{if } v < 0 \\ \dfrac{1}{\theta}\left[1-\exp\left(\dfrac{-2\tau_a(\theta-v)}{\beta_a}\right)\right], & \text{if } 0 \le v \le \theta \\ 0, & \text{if } v > \theta \end{cases} \tag{2}$$

where β_a is the variance-to-mean ratio. The derived distribution is confirmed by the simulation (see Fig. 3(c)). This distribution is invariant with respect to the input strength; in other words, the network is always ready to detect the input change.

In practice, however, when the external input is spatially localized, the above condition for E-I balance no longer holds. To solve this problem, Tian et al.[24] consider a localized neuronal connections, in term of that each neuron tends to have a higher connection probability to close neighbors than to distal neighbors, with the value decreasing as a Gaussian function of the distance. Tian et al. derived that the E-I balance condition in such a heterogeneous network becomes $\bar{f}_E/\bar{f}_I > \bar{W}_{EI}/\bar{W}_{II} > \bar{W}_{EE}/\bar{W}_{IE}$, where the bar denotes spatial average. For more details, see the analysis in [24].

Tian et al.[24] applied their model to simulated data of a spike camera and demonstrated that the model can detect the rapid change of the external input (see Fig 3(d)).

3.1.3 Future development

Overall, previous studies have demonstrated that the E-I balanced network originated from neuroscience has potential to be applied in brain-inspired computer vision, serving for fast object detection. To fully validate this application, however, there are still a lot of researches to be done. These include, for instances, 1) the reliability of the model in various visual conditions; 2) the robustness of the model to various noise forms; 3) the acceleration of the model to match the high speed of spike cameras; 4) the comparison of the model with other methods; 5) the integration of the model with other modular functions of computer vision; 6) the test of the model in real-world applications.

3.2 A brain-inspired model for object tracking

3.2.1 Biology background

In navigation tasks, animals need to track their spatial locations and head-directions smoothly[39]. A large volume of modelling studies has revealed that the brain can exploit a type of recurrent neural network called continuous attractor neural networks (CANNs) to achieve this task[39, 40]. In the brain, in addition to tracking the movement of an object continuously, there is an extra imposed requirement of tracking the object movement anticipatively, i.e., to predict the future position of the object. This is because in the brain, the transmission of neural signal is significantly delayed. For example, the propagation of visual signal from the retina to the primary visual cortex takes about 50–80 ms[41]. If this delay is not compensated, our perceived position of the object will lag its true position in the external world considerably, impairing our vision. A strategy to compensate the transmission delay is anticipation, and experimental evidences unveil that the brain does adopt such a strategy. For example, in spatial navigation, the internal head direction encoded by neurons in anterior dorsal thalamic nuclei in a rodent precedes the true head position of the rodent by 25 ms[42], i.e., the internal representation of the neural circuit predicts the external input by 25 ms. In the below, we will review a computational neural model to realize anticipative object tracking and discuss about its potential applications in brain-inspired computer vision. Non-anticipative tracking will be treated as a special case when the amount of anticipative time is less than zero.

3.2.2 Computational model

Motivated by the experimental findings, Mi et al.[25] proposed a computational model for anticipative object tracking and applied it to real-world problems[43, 44]. Specifically, they consider a CANN with adaptive neuronal responses, which is introduced below. Without loss of generality, a one-dimensional CANN is introduced.

A CANN is canonical network model which has been successfully used to explain the representations of continuous variables in neural system, including motion direction, object orientation, head direction, spatial location, etc.[45] Its biological relevance was also supported by recent neuroscience experiments[40, 46]. A 1D CANN is illustrated in Fig. 4(a). Denote x as a 1D continuous stimulus encoded by a CANN, whose value is in the range of $(-\pi, \pi]$ with a periodic boundary. Denote $U(x,t)$ as the synaptic input received by neuron at time t with a preferred stimulus x, and $r(x,t)$ as the corresponding neuronal firing rate. The dynamics of $U(x,t)$ and $r(x,t)$ are given by

$$\tau\frac{\partial U(x,t)}{\partial t} = -U(x,t) + \rho\int_{x'} J(x,x')r(x',t)\mathrm{d}x' - V(x,t) + I_{ext}(x,t) \tag{3}$$

$$r(x,t) = \frac{U(x,t)^2}{1+\kappa\rho\int_{x'}U(x',t)^2\mathrm{d}x'} \tag{4}$$

where τ is the synaptic time constant, ρ is the neuron

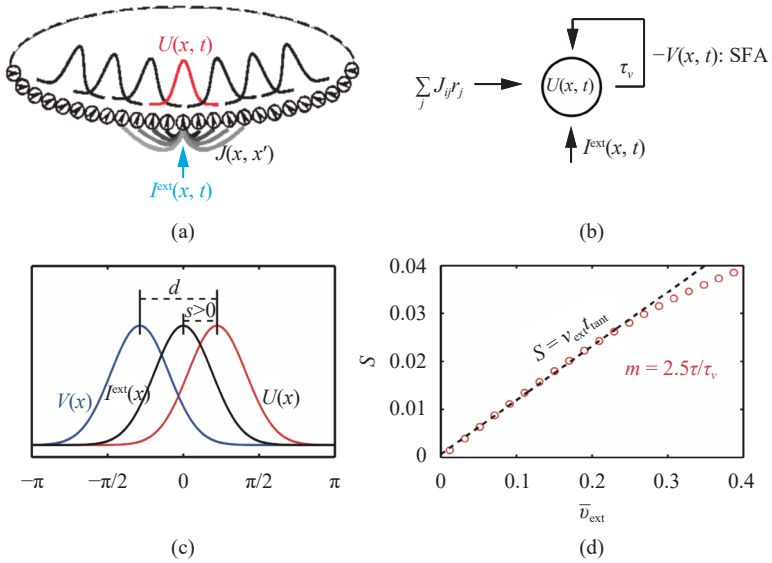


Fig. 4 A CANN with negative feedback for anticipate object tracking: (a) Schematic diagram of 1D CANN encoding the head direction; (b) Diagram of the negative feedback mechanism; (c) An example of anticipative tracking of the moving stimulus (black line). Because of slow spike frequency adaptation (SFA), the bump of $V(x)$ (blue line) is lagging behind that of $U(x)$ (red line). (d) Distance s that the network bump leads the moving object VS. the speed of the moving object v_{ext}. In anticipative tracking condition, the leading distance of the bump are proportional to the speed of the moving object in a wide speed regime. Figures are adapted from [25].

density, and $I_{ext}(x,t)$ is the external input. $J(x,x') = (J_0/\sqrt{2\pi}a)\mathrm{e}^{-(x-x')^2/2a^2}$ represents the recurrent connection from neurons x' to x, with J_0 controlling the connection strength and the Gaussian width a controlling the connection range. The relationship between $r(x,t)$ and $U(x,t)$ satisfies divisive normalization, with κ controls the inhibition strength.

The current $-V(x,t)$ denotes the adaptation effect, whose dynamics is written as

$$\tau_v \frac{\partial V(x,t)}{\partial t} = -V(x,t) + mU(x,t) \tag{5}$$

where τ_v is the time constant of adaptation and m controls the amplitude of adaptation.

The adaptation introduces a negative feedback to suppress neuronal activities, as shown in Fig. 4(b). More active a neuron is, stronger the suppression is. Without adaptation, a CANN is known to hold a continuous family of Gaussian-shape stationary states called bumps. Adaptation destabilizes the bump state to induce travelling wave of the bump, i.e., the bump moves spontaneously in the attractor space without relying on external inputs. This reflects the intrinsic mobility of the CANN caused by adaptation, measured by the speed of travelling wave v_{int}. When the network receives an external moving input, the network tracking behaviour is determined by two competing factors: the intrinsic speed of the network v_{int} and the speed of the external input v_{ext}. Interestingly, Mi et al.[25] found that when $v_{ext} < v_{int}$, the bump position leads that of the external input, achieving anticipative tracking, as shown in Fig. 4(c). Furthermore, they found that for a wide range of speed values, the leading distance is proportional to the speed of the external input, implying that the anticipation time is constant, which agrees with experimental data[47], as shown in Fig. 4(d).

The CANN is essentially a computational model employed by the brain to track moving objects. Compared with current machine learning algorithms, it has a number of advantages, including: 1) the computation is performed by the network dynamics, without the need of feature extraction in image frame by frame; 2) the network tracking is robust to noises, a property coming from attractors of a dynamical system; 3) the anticipation time of the model is approximately a constant, independent of the speed of the moving object over a wide range; 4) the model parameters can be defined theoretically according to the task requirement without training by a large dataset; 5) the network computation can be implemented on hardware for neuromorphic computing.

Because of these appealing properties, efforts have been tried to apply CANNs to real-world problems. In [43], a CANN was successfully implemented on the "Tianjic" chip and integrated with other methods in a self-driving bicycle system to track a running object. Recently, a spiking CANN was developed to track moving objects anticipatively[44].

3.2.3 Future development

Overall, previous studies have demonstrated that the CANN with adaptation originated from neuroscience has potential to be applied in brain-inspired computer vision, serving for anticipative object tracking. To fully validate this application, however, there are still a lot of researches to be done. These include, for instances, 1) the acceleration of the running speed of a CANN to be compatible with spike camera; 2) the integration of the mod-

el with other feature-based methods in computer vision to improve the tracking accuracy; 3) the implementation of a CANN with adaptive neural responses on hardware.

3.3 A brain-inspired model for spatio-temporal pattern recognition

3.3.1 Biology background

A large volume of experimental studies have revealed that the subcortical pathway, which goes from the retina, to superior colliculus (SC), and to higher visual cortex, plays an important role in rapid object recognition[48, 49]. For example, the mice study showed that this pathway mediates the rapid innate responses of the animal[50]. The capability of the subcortical pathway for rapid motion processing relies on its two main features. Firstly, different from the ventral visual pathway that slowly and hierarchically processes visual information, the subcortical pathway provides a shortcut from the retina to higher visual cortex with no explicit feature extraction, where the retina behaves like a reservoir network. Secondly, SC can linearly read out the retina output and perform fast decision making. The experimental study[51] showed that the wide vertical cells in SC, which have large receptive fields and wide-acting inhibition, can realize sampling over a large retina area and implement winner-take-all computation. The monkey experiment also demonstrated that SC plays an causal role in perceptual decision making[52]. In addition to the retina-SC subcortical pathway, the early stages of the auditory pathway also share a similar structure, i.e., a reservoir to decision-making pathway, suggesting that there may serve as a canonical mechanism for fast spatio-temporal pattern recognition.

3.3.2 Computational model

Inspired by the structure and computations of the subcortical pathway, Lin et al.[26] recently proposed a computational model for rapid motion patterns recognition. As shown in Fig. 5(a), they built a reservoir module followed by a decision module to mimic the retina-SC pathway, referred to as reservoir decision-making network (RDMN) hereafter. The details of RDMN are introduced below.

In RDMN, a hierarchical reservoir network was employed to simulate the information processing in the retina, which consists of L feedforwardly connected layers, as shown in Fig. 5(a). Denote x_i^l as the synaptic input current received by neuron i in layer l, for $i = 1, \cdots, N_l$; $l = 1, \cdots, L$, with N_l as the number of neurons in layer l. The recurrent dynamics in the reservoir layer l is given by

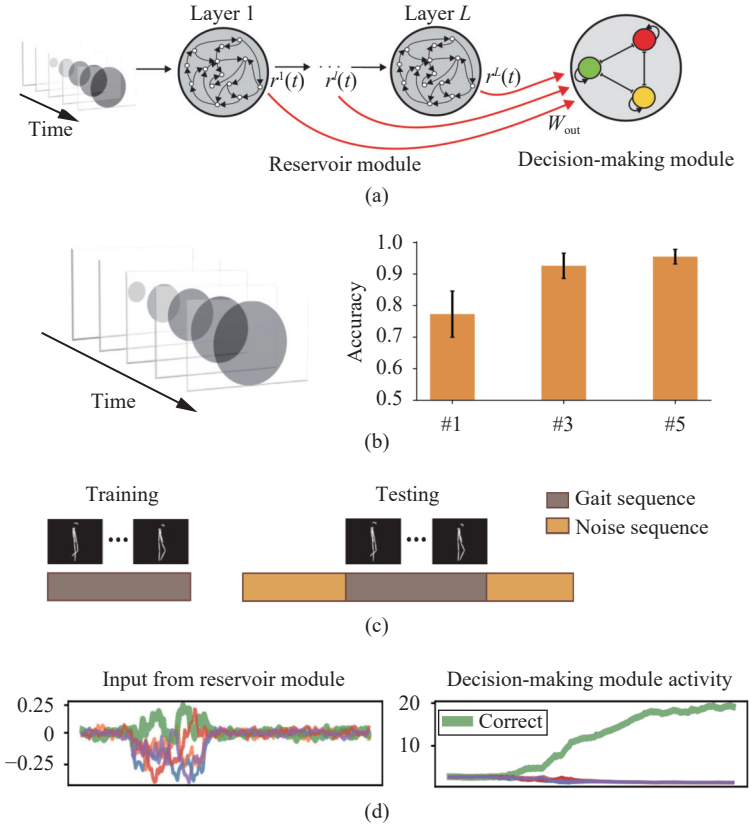


Fig. 5 RDMN performs gait recognition in an event-driven manner. (a) Structure of RDMN, which consists of a reservoir module and a decision-making module. (b) An example of looming pattern (left panel) and the generalization performance in the looming pattern discrimination task (right panel). (c) An example trial of the gait of a subject. (d) Neural dynamics of the decision-making module in the event-based gait recognition task. Figure is adapted from [26].

$$\tau_l \frac{\mathrm{d}x_i^l}{\mathrm{d}t} = -x_i^l + \sum_{j=1}^{N_{l-1}} M_{i,j}^{l,l-1} r_j^{l-1} + \sum_{j\neq i}^{N_l} M_{i,j}^{l,l} r_j^l + \sum_{j=1}^{N_{in}} M_{i,j}^{l,0} I_j^{ext} \delta_{l,1} \tag{6}$$

where $r_i = \tanh(x_i)$ is the activation function of neuron i, $M_{i,j}^{l,l-1}$ is the feedforward connections from neuron j in layer $l-1$ to neuron i in layer l, $M_{i,j}^{l,l}$ denotes the recurrent connections from neuron j to neuron i in layer l. τ_l is the time constant of layer l, $M_{i,j}^{l,0}$ denotes the input connections. I_j^{ext} denotes the external input, with N_{in} is the input dimension. $\delta_{l,1} = 1$ for $l = 1$, and otherwise 0, which indicates that the external input is only sent to layer 1. The connection matrices $M_{i,j}^{l,0}$, $M_{i,j}^{l,l-1}$, $M_{i,j}^{l,l}$ are all random and sparse, which are sampled from Gaussian distributions and not trained. Notably, the largest eigenvalue of recurrent connections $M_{i,j}^{l,l}$ in the reservoir module is set to be slightly larger than 1 to enable the the network dynamics to operate on the edge of chaos[53].

After the external visual input is mapped into a special state of the retina network, SC performs temporal information integration and decision making. Thus, a decision-making model is used to model the information processing in SC[54]. In RDMN, the decision-making model was greatly simplified and extended to multi-class classification tasks[26]. The simplified model is composed of several competing neurons, which receive all the inputs from the neurons in the reservoir module, as shown in Fig. 5(a). And each neuron in the decision-making module represents one category. Denote I_i as the input summation from all neurons in the reservoir module, y_i as the total synaptic input of decision neuron i, r_i as the corresponding activation function, and s_i as the synaptic input due to NMDA receptors. The dynamics of the decision-making network are given as

$$I_i = I_0^{\star} + \sum_{l=1}^{L} \sum_{j=1}^{N^l} M_{i,j}^{dm,l} r_j^l \tag{7}$$

$$y_i(t) = J_E s_i + \sum_{j\neq i}^{N_{dm}} J_M s_j + I_i \tag{8}$$

$$r_i(t) = \frac{\beta}{\gamma} \ln\left[1 + \exp\left(\frac{y_i - \theta}{\alpha}\right)\right] \tag{9}$$

$$\tau_s \frac{\mathrm{d}s_i}{\mathrm{d}t} = -s_i + \gamma(1 - s_i) r_i \tag{10}$$

where $M_{i,j}^{dm,l} r_j^l$ is the feedforward connection from neuron j of layer l to neuron i in the decision-making module, I_0^* is the bias input. The synaptic input y_i is composed of three parts: 1) $J_E s_i$ is the self-excitation input, representing the excitatory interactions between neurons encoding the same category; 2) $\sum_{j\neq i}^{N_{dm}} J_M s_j$ denotes the mutual inhibition between neurons; 3) I_i is the feedforward input from the reservoir module. The parameters β, γ and α control the shape of the nonlinear activation function. Equation (10) models the the slow dynamics of the synaptic current due to NMDA receptors, with the time constant $\tau >> 1$. τ_s controls the time window for integrating input over time by decision-making neurons. The parameters I_0^*, J_E, J_M are both chosen optimally based on a thorough mathematical analysis[26].

The only parameters needed to learn in RDMN are the feedforward connection matrix $M_{l,j}^{dm,i}$ from the reservoir module to the decision-making module, shown by red lines in Fig. 5(a). The loss function is the discrepancy between the actual inputs received by decision-making neurons and the target inputs, $E = (1/2)\sum_{i=1}^{N_{dm}} \sum_{k=1}^{N_k} \int_0^T \mathrm{d}t \times [f_i^k(t) - I_i^k(t)]^2$, where $f_i^k(t)$ is the target input. $M_{i,j}^{dm,l}$ can be optimized by minimizing the loss function using back-propagation through time, or FORCE learning[55], a biologically more plausible method.

When a spatio-temporal input is presented to RDMN, due to the echo properties of the reservoir network, the input is firstly projected from a low-dimensional space to a high-dimensional neural state space, which tends to become linearly separable. Moreover, the hierarchical reservoirs operate at different time and frequency scales, which further enhances the linear separation. With the evidence emerged from the reservoir module, the decision-making module performs a temporal information integration process via its attractor dynamics. Because of the self-excitation and mutual inhibition in the decision-making module, neurons representing different categories integrate information and compete with each other. When a neuron wins the competition, the corresponding classification is made.

Several appealing properties of RDMN are demonstrated through experiments: 1) RDMN successfully reproduces the looming pattern discrimination task as observed in the animal experiment, demonstrating that the model can achieve an excellent generalization performance with only a few trials, see Fig. 5(b); 2) When training data is limited, RDMN outperforms deep learning counterparts, such as LSTM and gated recurrent unit (GRU) on the gait recognition task, and notably, RDMN achieves this by using much fewer numbers of training parameters than long short-term memory (LSTMs) and GRUs. This property is appealing for few-shot learning; 3) Due to the self-excitation and mutual inhibition in the decision-making process, RDMN can perform recognition in an event-based manner, enabling the model to automatically detect and recognize the input pattern, as shown in Figs. 5(c) and 5(d). This property is appealing in real-world applications.

3.3.3 Future direction

Overall, previous studies have demonstrated that RDMN originated from mimicking the subcortical path-

way have potential to be applied in brain-inspired computer vision, serving for rapid moving object recognition. To fully validate this application, however, there are still a lot of researches to be done. These include, for instances, 1) the development of temporal structure extraction in RDMN to further enhance its spatio-temporal pattern recognition capability; 2) the deployment of event-based decision module on neuromorphic computing devices for efficient computing on spike signals; 3) the integration of the ventral and subcortical models for integrated local-global and fast-slow visual information processing; 4) the combination of moving object detection, anticipative tracking, and recognition models to build a unified computational framework for spatio-temporal information processing.

4 Conclusions and discussions

SNNs were regarded as the "third generation of neural network models" as early as in the 1990s[56]. SNNs are biologically more plausible than artificial neural networks, however, up to now, the performances of SNNs are far behind that of artificial neural networks[18]. We need to learn more from biological systems to promote the development of brain-inspired computer vision. A key issue that is missed is the fact that biological vision is targeted on processing spatio-temporal patterns. To capture this missing point, recently, a new paradigm for brain-inspired computer vision is emerging, which takes into account the spatio-temporal nature of signals in every part of a vision task, from signal sensing, to object detection, object tracking, object recognition, etc. In recent years, with the rapid development of spike cameras, we are able to collect spatio-temporal spike data from a large and complex scene[21–23]; with the rapid advance of computational neuroscience, we are able to build more capable computational models to process spatio-temporal spike data efficiently[24–26, 57]; with the fast progress of neuromorphic computing[58–61], we have hope to develop computing platforms that can support the efficient running of brain-inspired neural network models. Combining the above progresses, we are at the edge of entering a new era of practicing brain-inspired computer vision.

In this paper, we have reviewed some recent primary studies on developing spike cameras[21–23] and computational models for object detection[24], tracking[25], and recognition[26]. This is just the beginning and there are still a lot of researches to be done. In the below, we look ahead to some key issues needed to be solved for the success of the new brain-inspired computer vision paradigm.

1) Developing biologically more plausible brain-like sensing devices. The retina in the brain has rich cell types and wiring structures, which enable the retina to perform smart neural computations[62], such as light adaptation[63], image sharpening, etc. However, recent brain-like sensing devices have only simply converted light patterns into spike signals, and generated a pixel-level representation of an image. In the future, we need to develop brain-like sensing devices that can mimic more computational features of the retina. For example, DVS records the motion information, which approximately simulates the peripheral information processing of the retina; while the spike camera records the detailed texture information, which approximately simulates the information processing of the fovea. Integrating the advantages of two devices can better simulate the sensory function of the retina of the brain[21].

2) Designing much smarter brain-inspired computational vision models. This paper has only introduced three rather simple computational models for object detection, tracking, and recognition. The biological vision system has much more complicated architectures and richer computational functions, e.g., visual information is processed in parallel through multiple pathways, visual cognition is from global to local, and prior knowledge affects our perception of an image through feedback. We can learn from these characteristics of the biological visual system to develop improved brain-inspired computational models. Also, we can utilize machine learning methods to trained brain-inspired models from data to improve their performances.

3) Exploring more suitable application scenarios for brain-inspired vision models. As the biological vision system is evolved to adapt to natural environments, its computational advantages should be reflected on the efficient and dynamic interactions with the surrounding environments. Thus, for brain-inspired vision, we should also consider suitable application tasks that can fully utilize its advantages, rather than applying it to the tasks, such as static image classification, which deep neural networks are good at. Recently, some interesting researches along this direction have started and they have already demonstrated promising performances, including, for instances, neuromorphic vision-based action detection[64], neuromorphic vision sensor for surveillance (NeuroAED)[65], and event-based neuromorphic vision for autonomous driving[66].

4) Developing more convenient and efficient programming tools for brain-inspired vision models. The fast development of deep learning technologies has benefited from not only large datasets and computing power such as GPU, but also the convenient softwares, such as Pytorch and Tensorflow. These programming tools lower the entry barrier for new comers and boost the whole field. Brain-inspired computational models rely heavily on neural dynamics, event-based computation, and sparse connectivity between neurons. Up to now, simulating and training large-size brain-inspired models are still challenging. Developing efficient and convenient programming tools, similar to Pytorch[67] and Tensorflow[68] to deep neural networks (DNNs), are urgently needed. Recently, some software platforms towards this goal have been de-

veloped, such as BrainPy[69].

Acknowledgements

This work was supported by National Key R&D Program of China (No. 2020AAA0105200), Science and Technology Innovation 2030-Brain Science and Brain-inspired Intelligence Project (No. 2021ZD0200204), National Key Research and Development Program of China (No. 2020AAA0130401), Huawei Technology Co., Ltd, China (No. YBN2019105137).

References

[1] A. Krizhevsky, I. Sutskever, G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Lake Tahoe, USA, vol. 1, pp. 1097–1105, 2012.

[2] Y. LeCun, Y. Bengio, G. Hinton. Deep learning. *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. DOI: 10.1038/nature14539.

[3] R. Geirhos, C. R. M. Temme, J. Rauber, H. H. Schütt, M. Bethge, F. A. Wichmann. Generalisation in humans and deep neural networks. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Montréal, Canada, pp. 7549–7561, 2018.

[4] R. S. van Bergen, N. Kriegeskorte. Going in circles is the way forward: The role of recurrence in visual inference. *Current Opinion in Neurobiology*, vol. 65, pp. 176–193, 2020. DOI: 10.1016/j.conb.2020.11.009.

[5] I. J. Goodfellow, J. Shlens, C. Szegedy. Explaining and harnessing adversarial examples. In *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, USA, 2015.

[6] K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun. Deep residual learning for image recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp. 770–778, 2016. DOI: 10.1109/CVPR.2016.90.

[7] Y. X. Zhu, T. Gao, L. F. Fan, S. Y. Huang, M. Edmonds, H. X. Liu, F. Gao, C. Zhang, S. Y. Qi, Y. N. Wu, J. B. Tenenbaum, S. C. Zhu. Dark, beyond deep: A paradigm shift to cognitive AI with humanlike common sense. *Engineering*, vol. 6, no. 3, pp. 310–345, 2020. DOI: 10.1016/j.eng.2020.01.011.

[8] D. Hassabis, D. Kumaran, C. Summerfield, M. Botvinick. Neuroscience-inspired artificial intelligence. *Neuron*, vol. 95, no. 2, pp. 245–258, 2017. DOI: 10.1016/j.neuron.2017.06.011.

[9] T. J. Huang. Imitating the brain with neurocomputer a "new" way towards artificial general intelligence. *International Journal of Automation and Computing*, vol. 14, no. 5, pp. 520–531, 2017. DOI: 10.1007/s11633-017-1082-y.

[10] D. D. Cox, T. Dean. Neural networks and neuroscience-inspired computer vision. *Current Biology*, vol. 24, no. 18, pp. R921–R929, 2014. DOI: 10.1016/j.cub.2014.08.026.

[11] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, vol. 529, no. 7587, pp. 484–489, 2016. DOI: 10.1038/nature16961.

[12] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, vol. 596, no. 7873, pp. 583–589, 2021. DOI: 10.1038/s41586-021-03819-2.

[13] W. Wei. Neural mechanisms of motion processing in the mammalian retina. *Annual Review of Vision Science*, vol. 4, pp. 165–192, 2018. DOI: 10.1146/annurev-vision-091517-034048.

[14] N. C. Rust, O. Schwartz, J. A. Movshon, E. P. Simoncelli. Spatiotemporal elements of macaque V1 receptive fields. *Neuron*, vol. 46, no. 6, pp. 945–956, 2005. DOI: 10.1016/j.neuron.2005.05.021.

[15] C. D. Gilbert, W. Li. Top-down influences on visual processing. *Nature Reviews Neuroscience*, vol. 14, no. 5, pp. 350–363, 2013. DOI: 10.1038/nrn3476.

[16] N. C. Rust, M. R. Cohen. Priority coding in the visual system. *Nature Reviews Neuroscience*, vol. 23, no. 6, pp. 376–388, 2022. DOI: 10.1038/s41583-022-00582-9.

[17] M. Humphries. *The Spike: An Epic Journey Through the Brain in 2.1 seconds*, Princeton, USA: Princeton University Press, 2021.

[18] M. Pfeiffer, T. Pfeil. Deep learning with spiking neurons: Opportunities and challenges. *Frontiers in Neuroscience*, vol. 12, Article number 774, 2018. DOI: 10.3389/fnins.2018.00774.

[19] M. N. Shadlen, W. T. Newsome. Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, vol. 4, no. 4, pp. 569–579, 1994. DOI: 10.1016/0959-4388(94)90059-0.

[20] M. Tsodyks, S. Wu. Short-term synaptic plasticity. *Scholarpedia*, vol. 8, no. 10, Article number 3153, 2013. DOI: 10.4249/scholarpedia.3153.

[21] L. Zhu, J. N. Li, X. Wang, T. J. Huang, Y. H. Tian. NeuSpike-Net: High speed video reconstruction via bio-inspired neuromorphic cameras. In *Proceedings of IEEE/CVF International Conference on Computer Vision*, IEEE, Montréal, Canada, pp. 2380–2389, 2021. DOI: 10.1109/ICCV48922.2021.00240.

[22] J. Zhao, R. Q. Xiong, H. F. Liu, J. Zhang, T. J. Huang. Spk2ImgNet: Learning to reconstruct dynamic scene from continuous spike stream. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Nashville, USA, pp. 11991–12000, 2021. DOI: 10.1109/CVPR46437.2021.01182.

[23] L. Zhu, S. W. Dong, T. J. Huang, Y. H. Tian. A retina-in-

spired sampling method for visual texture reconstruction. In *Proceedings of IEEE International Conference on Multimedia and Expo*, Shanghai, China, pp. 1432–1437, 2019. DOI: 10.1109/ICME.2019.00248.

[24] G. S. Tian, S. Y. Li, T. J. Huang, S. Wu. Excitation-inhibition balanced neural networks for fast signal detection. *Frontiers in Computational Neuroscience*, vol. 14, Article number 79, 2020. DOI: 10.3389/fncom.2020.00079.

[25] Y. Y. Mi, C. C. A. Fung, K. Y. M. Wong, S. Wu. Spike frequency adaptation implements anticipative tracking in continuous attractor neural networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Montréal, Canada, vol. 1, pp. 505–513, 2014.

[26] X. H. Lin, X. L. Zou, Z. L. Ji, T. J. Huang, S. Wu, Y. Y. Mi. A brain-inspired computational model for spatio-temporal information processing. *Neural Networks*, vol. 143, pp. 74–87, 2021. DOI: 10.1016/j.neunet.2021.05.015.

[27] Z. C. Bi, S. W. Dong, Y. H. Tian, T. J. Huang. Spike coding for dynamic vision sensors. In *Proceedings of Data Compression Conference*, IEEE, Snowbird, USA, pp. 117–126, 2018. DOI: 10.1109/DCC.2018.00020.

[28] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, D. Scaramuzza. Event-based vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 154–180, 2022. DOI: 10.1109/TPAMI.2020.3008413.

[29] C. Posch, T. Serrano-Gotarredona, B. Linares-Barranco, T. Delbruck. Retinomorphic event-based vision sensors: Bioinspired cameras with spiking output. *Proceedings of IEEE*, vol. 102, no. 10, pp. 1470–1484, 2014. DOI: 10.1109/JPROC.2014.2346153.

[30] A. M. Zador. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature Communications*, vol. 10, no. 1, Article number 3770, 2019. DOI: 10.1038/s41467-019-11786-6.

[31] S. E. Raiguel, D. K. Xiao, V. L. Marcar, G. A. Orban. Response latency of macaque area Mt/V5 neurons and its relationship to stimulus parameters. *Journal of Neurophysiology*, vol. 82, no. 4, pp. 1944–1956, 1999. DOI: 10.1152/jn.1999.82.4.1944.

[32] S. Thorpe, D. Fize, C. Marlot. Speed of processing in the human visual system. *Nature*, vol. 381, no. 6582, pp. 520–522, 1996. DOI: 10.1038/381520a0.

[33] C. Van Vreeswijk, H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, vol. 274, no. 5293, pp. 1724–1726, 1996. DOI: 10.1126/science.274.5293.1724.

[34] W. R. Softky, C. Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, vol. 13, no. 1, pp. 334–350, 1993. DOI: 10.1523/JNEUROSCI.13-01-00334.1993.

[35] M. Zhou, F. X. Liang, X. R. Xiong, L. Li, H. F. Li, Z. J. Xiao, H. W. Tao, L. I. Zhang. Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex. *Nature Neuroscience*, vol. 17, no. 6, pp. 841–850, 2014. DOI: 10.1038/nn.3701.

[36] Y. S. Shu, A. Hasenstaub, M. Badoual, T. Bal, D. A. McCormick. Barrages of synaptic activity control the gain and sensitivity of cortical neurons. *Journal of Neuroscience*, vol. 23, no. 32, pp. 10388–10401, 2003. DOI: 10.1523/JNEUROSCI.23-32-10388.2003.

[37] B. V. Atallah, M. Scanziani. Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. *Neuron*, vol. 62, no. 4, pp. 566–577, 2009. DOI: 10.1016/j.neuron.2009.04.027.

[38] M. Okun, I. Lampl. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neuroscience*, vol. 11, no. 5, pp. 535–537, 2008. DOI: 10.1038/nn.2105.

[39] B. K. Hulse, V. Jayaraman. Mechanisms underlying the neural computation of head direction. *Annual Review of Neuroscience*, vol. 43, pp. 31–54, 2020. DOI: 10.1146/annurev-neuro-072116-031516.

[40] S. S. Kim, H. Rouault, S. Druckmann, V. Jayaraman. Ring attractor dynamics in the Drosophila central brain. *Science*, vol. 356, no. 6340, pp. 849–853, 2017. DOI: 10.1126/science.aal4835.

[41] L. G. Nowak, M. H. J. Munk, P. Girard, J. Bullier. Visual latencies in areas V1 and V2 of the macaque monkey. *Visual Neuroscience*, vol. 12, no. 2, pp. 371–384, 1995. DOI: 10.1017/S095252380000804X.

[42] J. P. Bassett, M. B. Zugaro, G. M. Muir, E. J. Golob, R. U. Muller, J. S. Taube. Passive movements of the head do not abolish anticipatory firing properties of head direction cells. *Journal of Neurophysiology*, vol. 93, no. 3, pp. 1304–1316, 2005. DOI: 10.1152/jn.00490.2004.

[43] J. Pei, L. Deng, S. Song, M. G. Zhao, Y. H. Zhang, S. Wu, G. R. Wang, Z. Zou, Z. Z. Wu, W. He, F. Chen, N. Deng, S. Wu, Y. Wang, Y. J. Wu, Z. Y. Yang, C. Ma, G. Q. Li, W. T. Han, H. L. Li, H. Q. Wu, R. Zhao, Y. Xie, L. P. Shi. Towards artificial general intelligence with hybrid Tianjic chip architecture. *Nature*, vol. 572, no. 7767, pp. 106–111, 2019. DOI: 10.1038/s41586-019-1424-8.

[44] L. T. Yu, T. H. Chu, Z. Zhao, Y. Y. Mi, Y. C. Yang, S. Wu. Spiking continuous attractor neural networks with spike frequency adaptation for anticipative tracking. In *Proceedings of IEEE International Workshop on Future Computing*, Hangzhou, China, 2019. DOI: 10.1109/IWOFC48002.2019.9078445.

[45] S. Wu, K. Y. M. Wong, C. C. A. Fung, Y. Y. Mi, W. H. Zhang. Continuous attractor neural networks: Candidate of a canonical model for neural information representation. *F1000Research*, vol. 5, Article number F1000, 2016. DOI: 10.12688/f1000research.7387.1.

[46] R. J. Gardner, E. Hermansen, M. Pachitariu, Y. Burak, N. A. Baas, B. A. Dunn, M. B. Moser, E. I. Moser. Toroidal topology of population activity in grid cells. *Nature*, vol. 602, no. 7895, pp. 123–128, 2022. DOI: 10.1038/S41586-021-04268-7.

[47] J. P. Goodridge, D. S. Touretzky. Modeling attractor deformation in the rodent head-direction system. *Journal of Neurophysiology*, vol. 83, no. 6, pp. 3402–3410, 2000. DOI: 10.1152/jn.2000.83.6.3402.

[48] P. F. Wei, N. Liu, Z. J. Zhang, X. M. Liu, Y. Q. Tang, X. B. He, B. F. Wu, Z. Zhou, Y. H. Liu, J. Li, Y. Zhang, X. Y. Zhou, L. Xu, L. Chen, G. Q. Bi, X. T. Hu, F. Q. Xu, L. P. Wang. Processing of visually evoked innate fear by a non-canonical thalamic pathway. *Nature Communications*, vol. 6, Article number 6756, 2015. DOI: 10.1038/ncomms7756.

[49] B. De Gelder, M. Tamietto, G. Van Boxtel, R. Goebel, A. Sahraie, J. Van den Stock, B. M. C. Stienen, L. Weiskrantz, A. Pegna. Intact navigation skills after bilateral loss of striate cortex. *Current Biology*, vol. 18, no. 24, pp. R1128–R1129, 2008. DOI: 10.1016/j.cub.2008.11.002.

[50] G. De Franceschi, T. Vivattanasarn, A. B. Saleem, S. G. Solomon. Vision guides selection of freeze or flight defense strategies in mice. *Current Biology*, vol. 26, no. 16, pp. 2150–2154, 2016. DOI: 10.1016/j.cub.2016.06.006.

[51] S. D. Gale, G. J. Murphy. Distinct representation and distribution of visual information by specific cell types in mouse superficial superior colliculus. *Journal of Neuroscience*, vol. 34, no. 40, pp. 13458–13471, 2014. DOI: 10.1523/JNEUROSCI.2768-14.2014.

[52] K. W. Latimer, A. C. Huk. Superior colliculus activates new perspectives on decision-making. *Nature Neuroscience*, vol. 24, no. 8, pp. 1048–1050, 2021. DOI: 10.1038/s41593-021-00885-7.

[53] I. B. Yildiz, H. Jaeger, S. J. Kiebel. Re-visiting the echo state property. *Neural Networks*, vol. 35, pp. 1–9, 2012. DOI: 10.1016/j.neunet.2012.07.005.

[54] K. F. Wong, X. J. Wang. A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, vol. 26, no. 4, pp. 1314–1328, 2006. DOI: 10.1523/JNEUROSCI.3733-05.2006.

[55] D. Sussillo, L. F. Abbott. Generating coherent patterns of activity from chaotic neural networks. *Neuron*, vol. 63, no. 4, pp. 544–557, 2009. DOI: 10.1016/j.neuron.2009.07.018.

[56] W. Maass. Networks of spiking neurons: The third generation of neural network models. *Neural Networks*, vol. 10, no. 9, pp. 1659–1671, 1997. DOI: 10.1016/S0893-6080(97)00011-7.

[57] S. Denève, A. Alemi, R. Bourdoukan. The brain as an efficient and robust adaptive learner. *Neuron*, vol. 94, no. 5, pp. 969–977, 2017. DOI: 10.1016/j.neuron.2017.05.016.

[58] K. Roy, A. Jaiswal, P. Panda. Towards spike-based machine intelligence with neuromorphic computing. *Nature*, vol. 575, no. 7784, pp. 607–617, 2019. DOI: 10.1038/s41586-019-1677-2.

[59] B. Cramer, S. Billaudelle, S. Kanya, A. Leibfried, A. Grübl, V. Karasenko, C. Pehle, K. Schreiber, Y. Stradmann, J. Weis, J. Schemmel, F. Zenke. Surrogate gradients for analog neuromorphic computing. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 119, no. 4, Article number e2109194119, 2022. DOI: 10.1073/pnas.2109194119.

[60] Y. P. Guo, X. L. Zou, Y. F. Hu, Y. F. Yang, X. X. Wang, Y. H. He, R. K. Kong, Y. Z. Guo, G. Q. Li, W. Zhang, S. Wu, H. L. Li. A Marr's three-level analytical framework for neuromorphic electronic systems. *Advanced Intelligent Systems*, vol. 3, no. 11, Article number 2100054, 2021. DOI: 10.1002/aisy.202100054.

[61] F. Zenke, S. M. Bohté, C. Clopath, I. M. Comşa, J. Göltz, W. Maass, T. Masquelier, R. Naud, E. O. Neftci, M. A. Petrovici, F. Scherr, D. F. M. Goodman. Visualizing a joint future of neuroscience and neuromorphic engineering. *Neuron*, vol. 109, no. 4, pp. 571–575, 2021. DOI: 10.1016/j.neuron.2021.01.009.

[62] T. Gollisch, M. Meister. Eye smarter than scientists believed: Neural computations in circuits of the retina. *Neuron*, vol. 65, no. 2, pp. 150–164, 2010. DOI: 10.1016/j.neuron.2009.12.009.

[63] L. Xiao, M. S. Zhang, D. J. Xing, P. J. Liang, S. Wu. Shifted encoding strategy in retinal luminance adaptation: From firing rate to neural correlation. *Journal of Neurophysiology*, vol. 110, no. 8, pp. 1793–1803, 2013. DOI: 10.1152/jn.00221.2013.

[64] G. Chen, S. Q. Qu, Z. J. Li, H. T. Zhu, J. X. Dong, M. Liu, J. Conradt. Neuromorphic vision-based fall localization in event streams with temporal-spatial attention weighted network. *IEEE Transactions on Cybernetics*, vol. 52, no. 9, pp. 9251–9262, 2022. DOI: 10.1109/TCYB.2022.3164882.

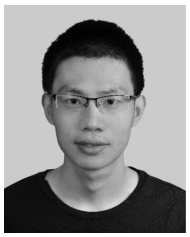
[65] G. Chen, P. G. Liu, Z. F. Liu, H. J. Tang, L. Hong, J. H. Dong, J. Conradt, A. Knoll. NeuroAED: Towards efficient abnormal event detection in visual surveillance with neuromorphic vision sensor. *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 923–936, 2021. DOI: 10.1109/TIFS.2020.3023791.

[66] G. Chen, H. Cao, J. Conradt, H. J. Tang, F. Rohrbein, A. Knoll. Event-based neuromorphic vision for autonomous driving: A paradigm shift for bio-inspired visual sensing and perception. *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 34–49, 2020. DOI: 10.1109/MSP.2020.2985815.

[67] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. M. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. J. Bai, S. Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Vancouver, Canada, Article number 721, 2019.

[68] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. F. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Q. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Q. Zheng. TensorFlow: Large-scale machine learning on heterogeneous distributed systems, [Online], Available: https://arxiv.org/abs/1603.04467, 2016.

[69] C. M. Wang, Y. Q. Jiang, X. Y. Liu, X. H. Lin, X. L. Zou, Z. L. Ji, S. Wu. A just-in-time compilation approach for neural dynamics simulation. In *Proceedings of the 28th International Conference on Neural Information Processing*, Springer, Bali, Indonesia, pp. 15–26, 2021. DOI: 10.1007/978-3-030-92238-2_2.



Xiao-Long Zou received B. Sc. degree in vehicle engineering from Jilin University, China in 2012, and the Ph. D. degree in system theory from Beijing Normal University, China in 2018. Currently, he is a postdoctoral fellow in Beijing Academy of Artificial Intelligence and School of Psychological and Cognitive Sciences, Peking University, China.

His research interests include computational neuroscience and brain-inspired computing.

E-mail: xiaolz@mail.pku.edu.cn
ORCID iD: 0000-0001-9397-6480

Tie-Jun Huang (Senior Member, IEEE) received the Ph. D. degree in pattern recognition and intelligent system from the Huazhong (Central China) University of Science and Technology, China in 1998. He is currently a professor in School of Computer Science, Peking University, China, and the director of Beijing Academy for Artificial Intelligence. He has published two books, more than 200 peer-reviewed papers on leading journals and conferences, held more than 50 granted patents, and is the co-editor of four ISO/IEC standards, five national standards of China, and four IEEE standards.

He is a fellow of CAAI and CCF, the Secretary General of the Artificial Intelligence Industry Technology Innovation Alliance, and the vice chair of the China National General Group on AI Standardization. He received the National Award for Science and Technology of China (Tier-2) for three times. He was awarded the Distinguished Young Scholar by the National Natural Science Foundation of China in 2014 and the Distinguished Professor of the Chang Jiang Scholars Program by the Ministry of Education of China in 2015.

His research interests include visual information processing and neuromorphic computing.

E-mail: tjhuang@pku.edu.cn

Si Wu received the B. Sc. degree in general physics in 1990, the M. Sc degree in general relativity in 1992, and the Ph. D. degree in statistical physics in 1995, all from Beijing Normal University, China. Currently, he is a professor in School of Psychological and Cognitive Sciences, PI in PKU-IDG/McGovern Institute for Brain Research, PI in Center for Quantitative Biology, and PI in PKU-Tsinghua Center for Life Sciences, Peking University, China. He is also a researcher in Beijing Academy of Artificial Intelligence, China, and Co-Editor-in-Chief of *Frontiers in Computational Neuroscience*.

His research interests include computational neuroscience and brain-inspired computing.

E-mail: siwu@pku.edu.cn (Corresponding author)
ORCID iD: 0000-0001-9650-6935

Machine Intelligence Research

Citation: X. L. Zou, T. J. Huang, S. Wu. Towards a new paradigm for brain–inspired computer vision. *Machine Inteligence Research*, vol.19, no.5, pp.412–424, 2022. https://doi.org/10.1007/s11633–022–1370–z

Articles may interest you