Memristor-Based Attention Network for Online Real-Time Object Tracking

Zekun Deng, Chunhua Wang, Hairong Lin, *Member, IEEE,* Quanli Deng, and Yichuang Sun, *Senior Member, IEEE,*

***Abstract*—Most existing visual object tracking approaches are implemented based on von Neumann computation systems, which inevitably have the problems of high latency. Additionally, remote server processing of video resources requires a large amount of data transmission over the Internet, which limits real-time tracking performance. The integration of visual object tracking technology into electronic devices has become a new trend. However, current visual object tracking approaches have high algorithm complexity, making it difficult to design circuits to implement the corresponding functions. In this paper, a memristor-based attention network and its corresponding algorithm are proposed to achieve online real-time tracking under parallel computing. Memristors are used to construct attention encoding circuits to record changes of the target in historical frames, and adjust attention signals to the target online and in real-time during the tracking process, avoiding the latency problem of the von Neumann architecture. Inspired by the working process of γ-GABAergic interneuron and tripartite synapse, we propose an attention allocation module to selectively allocate attention values. Combining the Winner-Take-All principle, we design a target localization circuit and an optimal attention zone selection circuit for parallel computation to track the location of the target. Finally, experiments and analyses on OTB-100, NFS, and VOT-RTb2022 benchmark datasets verify that the proposed memristor-based attention network has promising tracking performance and achieves a tracking speed of 1000 FPS, demonstrating superior real-time performance.**

***Index Terms*—Attention network, memristor, object tracking, allocation module, online, real-time, winner-take-all.**

I. Introduction

VISUAL Object Tracking (VOT) is one of the key problems in computer vision, which aims to track any given object in a sequence of video frames based on an initial state. VOT has wide applications in the fields of autonomous driving, visual surveillance, visual navigation, and robotics [1], [2].

Currently popular trackers are mostly designed based on Correlation Filter (CF) or Siamese networks [3]. The CF-based approaches convert computation from the time domain to the frequency domain, training filters for distinguishing between background and target. In 2017, Valmadre *et al.* [4] proposed a tracking model called CFNet, which interprets the correlation filter learner as a differentiable layer in deep neural networks, enabling the learning of deep features tightly coupled with the correlation filters. In 2024, Yang *et al.* [5] proposed a new correlation filter through saliency-driven localization and cascaded scale estimation (SDCS-CF), which improves the confidence and robustness of target localization. Nevertheless, the CF-based approaches usually require additional computational components and have a large computational cost. The Siamese network-based approaches takes Siamese networks as the main body, and traditional Siamese networks perform feature matching by directly calculating the distance between two features. In 2016, Bertinetto *et al.* [6] proposed a novel fully-convolutional Siamese network called SiamFC for end-to-end training of a convolutional neural network (CNN) for tracking objects. In 2018, Li *et al.* [7] proposed a Siamese region proposal network (SiamRPN). SiamRPN consists of Siamese subnetwork for feature extraction and region proposal subnetwork including the classification branch and regression branch, and is trained end-to-end using a large-scale dataset of image pairs. In 2020, Chen *et al.* [8] views the visual tracking problem as a parallel classification and regression problem, and proposed a Siamese Box Adaptive Network (SiamBAN) to directly classifies objects and regresses their bounding boxes in a unified fully convolutional network. In 2023, Wei *et al.* [9] proposed an online updatable Siamese tracker called SiamSTC, which exploits spatiotemporal context to improve the accuracy and robustness of tracking. However, matching operations and complex network structures make Siamese network-based approaches time-consuming.

Recently, Transformer that introduces the attention mechanism to enhance and integrate the features of the tracked object has achieved good tracking performance. In 2021, Chen *et al.* [10] proposed a Transformer Tracking (TransT) method, which effectively combines the template and search region features solely using attention, including an ego-context augment module based on self-attention and a cross-feature augment module based on cross-attention. In 2022, Mayer *et al.* [11] proposed a Transformer-based target model prediction network for tracking called ToMP, where Transformer is used to capture global relationships. In 2023, Nardo *et al.* [12] proposed a novel visual object tracking model called ViTCRT based on Siamese network and vision Transformer, which utilizes the learning and memory capabilities of the visual Transformer to achieve tracking. Although the Transformer based approaches achieve higher tracking performance, the lack of sufficient training data to train network structures and

This work is supported by the National Natural Science Foundation of China (Grant Nos. 62271197, 62201204), Guangdong Basic and Applied Basic Research Foundation (Grant No. 2024A1515011910), China Postdoctoral Science Foundation (Grant No.2022M71104).(*Corresponding author: Chunhua Wang, email: wch1227164@hnu.edu.cn.*)

Zekun Deng, Chunhua Wang, and Quanli Deng are with College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China.

Hairong Lin is with School of Electronic Information, Central South University, Changsha 410083, China.

Yichuang Sun is with the School of Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, U.K.

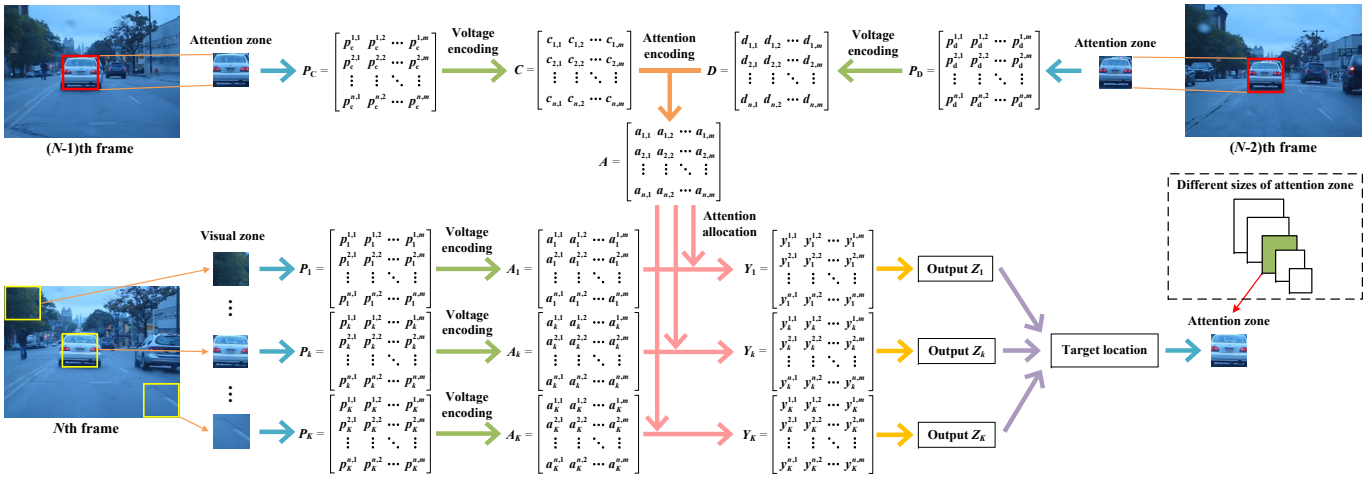


Fig. 1. Framework of the memristor-based attention network for object tracking.

the complexity of the models limit their applications.

Current VOT approaches are almost all run on the computing systems based on the von Neumann architecture such as CPUs and GPUs. Due to the physical separation of the storage unit and computing unit of the von Neumann architecture, the frequent data transmission between the storage unit and the computing unit, as well as the performance difference between them, result in varying degrees of latency in trackers used for VOT, limiting the real-time performance of the tracker at the hardware level [13]. In addition, the processing of video resources by remote servers requires a large amount of data transmission over the Internet, and the time-consuming data transmission also limits the real-time tracking performance of trackers [14], [15]. Therefore, integrating object tracking technology into electronic devices is a promising trend. However, the existing VOT approaches contain a large number of network models and complex functions, making it difficult to design corresponding circuits to implement related algorithms [16]. As a new type of non-volatile nanodevice, memristors have advantages such as in-memory computing, low power consumption, and ease of integration [17], providing a novel approach for hardware design of VOT. The memristor as the fourth basic circuit element to represent the relationship between charge and flux was defined by Chua [18], and was first physically produced by Hewlett Packard Laboratories in 2008 [19]. Since then, memristors have been widely used in circuit designs such as network circuits [20]–[23], computing circuits [24]–[27], memristor crossbar [28]–[30], and brain-inspired circuit [31], [33]. Taking advantage of the in-memory computing characteristics of memristors, VOT circuit design based on memristors can greatly simplify the circuit structure, achieve online real-time computation, and greatly improve tracking speed [32]. However, the attention network based on memristors for online real-time object tracking has yet to be proposed.

In this paper, a memristor-based attention network (MAN) and its corresponding algorithm are proposed for online real-time object tracking. The MAN records changes of the target in historical frames to adjust the attention encoding of the attention zone online, while allocating attention values to the current frame according to the encoded attention signals, and performs parallel computation based on the Winner-Take-All (WTA) principle to locate the target. The contributions of this paper are as follows:

1) Attention encoding circuits are designed based on memristors to record target information from historical frames and dynamically encode attention signals for the attention zone online and in real-time, avoiding the latency problem of the von Neumann architecture.
2) Inspired by the working process of γ-GABAergic (GABA) interneuron and tripartite synapse, an attention allocation module is designed to selectively allocate attention values. Combined with the WTA principle, a target localization circuit and an optimal attention zone selection circuit are designed for parallel computing attention allocation results and tracking the target location.
3) The proposed MAN is designed based on the corresponding algorithm and realizes online real-time tracking under parallel computing, with a tracking time of only 1 ms per frame and a tracking speed of 1000 frames per second (FPS), which is higher than the other trackers. Experiments on the tracking benchmarks demonstrate that the MAN has promising tracking performance and superior real-time performance.

The rest of the paper is organized as follows: Section II introduces the network framework and algorithms. Section III provides the circuit design and analysis of the MAN. Section IV performs experiments and performance analysis on the MAN. Section V concludes this work.

II. Network Framework and Algorithm

A. Network framework

The framework of the memristor-based attention network is shown in Fig. 1. The framework mainly consists of five parts: voltage encoding, attention encoding, attention allocation, target localization, and optimal attention zone selection. Set the current frame be the Nth frame, and the attention

zones that successfully locate the target in the $(N-1)$th and $(N-2)$th frames are set as P_{C} and P_{D}, respectively. The image size of each frame is set to $H \times W$. The size of the attention zone is set to $n \times m$, where the number of pixels is $L = m \cdot n$. The size of each visual zone is the same as the attention zone, and the number of visual zones can be calculated as $K = \left(\frac{H-n}{s}+1\right)\left(\frac{W-m}{s}+1\right)$, where s is the pixel intervals of each visual zone. We perform voltage encoding on the pixel matrixes P_{C} and P_{D} of the attention zones, obtaining voltage-encoded matrixes C and D, respectively. At the same time, attention encoding is performed on the voltage values within C and D to obtain the attention signal matrix A, which is used to allocate attention values to different visual zones of the current frame. The historical target information is stored in memristors to adjust the attention encoding process online, thereby eliminating tracking drift or even tracking loss caused by tracking errors. The attention allocation is calculated from the encoded voltages in matrices $A_1, A_2, \cdots, A_K$ of visual zones and the attention signals in matrix A, and the attention allocation rule is inspired by the working process of GABA interneuron and tripartite synapse. Based on the WTA principle, we perform parallel computation on the attention allocation results and locate the target by the maximum attention output values. Furthermore, we set the location of the attention zone in the first frame, and set Q attention zones of different sizes to work independently to output their respective maximum attention results, and then utilize the WTA principle to select the optimal attentional zone, thereby eliminating the impact of target scale changes.

B. Algorithm design

The algorithms of the MAN are shown in Algorithm 1 and Algorithm 2. In Algorithm 1, $c_r(t)$ is the voltage-encoded signal of the pixel p_{c}^r, and $d_r(t)$ is the voltage-encoded signal of the pixel p_{d}^r, where α is an encoding coefficient and $r = 1, 2, \ldots, L$. The weight adjustment signal $e_r(t)$ is used to adjust the rth attention weight $w_r(t)$, and $e_r(t)$ is calculated from $c_r(t)$ and $d_r(t)$. The constants β_1 and β_2 are the gain factors of $c_r(t)$ and $d_r(t)$, respectively. V_{th1} and V_{th2} are the thresholds for weight adjustment. If $e_r(t) > V_{th1} > 0$, the change of $w_r(t)$ varies with $\eta_1(t)e_r(t)$. If $V_{th1} \geq e_r(t) \geq V_{th2}$, $w_r(t)$ remains unchanged. If $0 > V_{th2} > e_r(t)$, the change of $w_r(t)$ varies with $\eta_2(t)e_r(t)$. Thereby the weight change is affected by $e_r(t)$, where $\eta_1(t)$ and $\eta_2(t)$ are weight adjustment variables. In this paper, the conductance of the memristor is used to represent the weight, thus the weight adjustment parameters are consistent with the memristor model. The signals $c_r(t)$, $e_r(t)$ and $w_r(t)$ jointly participate in the generation of the attention signal $a_r(t)$, where λ_1 and λ_2 are the gain coefficients. The signal $a_i^r(t)$ is the voltage-encoded signal of the pixel p_i^r in the ith visual zone. The allocation control signal $f_i^r(t)$ is calculated from $a_i^r(t)$ and the attention signal $a_r(t)$, and $f_i^r(t) = \xi\left(a_i^r(t) + a_r(t)\right)$, where ξ is a gain factor. The allocation control signal $f_i^r(t)$ is used to control the attention value allocated to $y_i^r(t)$. If $f_i^r(t) \leq V_{th3}$, then the attention value allocated to $y_i^r(t)$ is 0. If $V_{th4} > f_i^r(t) > V_{th3}$, then the attention value allocated to $y_i^r(t)$ is $\varphi(f_i^r(t))$, where $\varphi(f_i^r(t))$ is a function related to the value of $f_i^r(t)$. If $f_i^r(t) \geq V_{th4}$, then the attention value allocated to $y_i^r(t)$ is 0. V_{th3} and V_{th4} are the thresholds of the attention allocation, and ξ is a gain coefficient. The attention value allocation rule for $y_i^r(t)$ is inspired by the working process of GABA interneuron and tripartite synapse, which is elaborated in Section III-C. $z_i(t)$ is the attention allocation result of the ith visual zone, where ρ is a gain factor. $v_{\mathrm{z}}(t)$ is the maximum value from the attention allocation results $z_1(t), z_2(t), \cdots, z_K(t)$ of K visual zones. If $z_i(t) \geq v_{\mathrm{z}}(t)$,

Algorithm 1 Attention encoding, attention allocation, and target localization algorithm

$K \leftarrow$ Number of visual zones
$L \leftarrow$ Number of pixels in an attention zone
$t \leftarrow 1$
for $i = 1$ **to** K **do**
  for $r = 1$ **to** L **do**
    $c_r(t) \leftarrow \alpha p_{\mathrm{c}}^r$ //Voltage encoding of pixel p_{c}^r
    $d_r(t) \leftarrow \alpha p_{\mathrm{d}}^r$ //Voltage encoding of pixel p_{d}^r
  // Subalgorithm 1. Attention encoding
    $e_r(t) = \beta_1 c_r(t) - \beta_2 d_r(t)$ //Weight adjustment signal
    if $e_r(t) > V_{th1} > 0$ **then**
      $\Delta w_r = \eta_1(t) e_r(t)$ //Change of weight
    end if
    if $V_{th1} \geq e_r(t) \geq V_{th2}$ **then**
      $\Delta w_r = 0$
    end if
    if $0 > V_{th2} > e_r(t)$ **then**
      $\Delta w_r = \eta_2(t) e_r(t)$
    end if
    $a_r(t) = -\left(\lambda_1 c_r(t) + \lambda_2 w_r(t) e_r(t)\right)$ //Attention signal
  // Subalgorithm 2. Attention allocation
    $a_i^r(t) \leftarrow \alpha p_i^r$ //Voltage encoding of pixel p_i^r
    $f_i^r(t) = \xi\left(a_i^r(t) + a_r(t)\right)$ //Allocation control signal
    if $f_i^r(t) \leq V_{th3}$ **then**
      $y_i^r(t) = 0$ //Attention allocation value
    end if
    if $V_{th4} > f_i^r(t) > V_{th3}$ **then**
      $y_i^r(t) = \varphi(f_i^r(t))$
    end if
    if $f_i^r(t) \geq V_{th4}$ **then**
      $y_i^r(t) = 0$
    end if
  end for
  // Subalgorithm 3. Target localization
  $z_i(t) = \rho(y_i^1(t) + y_i^2(t) + \ldots + y_i^L(t))$ //Allocation result
end for
$v_{\mathrm{z}}(t) = \max(z_1(t), z_2(t), \cdots, z_K(t))$ //Maximum value
for $i = 1$ **to** K **do**
  if $z_i(t) \geq v_{\mathrm{z}}(t)$ **then**
    $Z_i(t) = \phi$ //Localization result
  end if
  if $z_i(t) < v_{\mathrm{z}}(t)$ **then**
    $Z_i(t) = 0$
  end if
end for
$t \leftarrow t + 1$

Algorithm 2 Optimal attention zone selection algorithm

$Q \leftarrow$ Number of attention zones
$t \leftarrow 1$
for $j = 1$ **to** Q **do**
  $x_{\mathrm{o}}(t) = \max(\varepsilon_1 v_{\mathrm{z}1}(t), \varepsilon_2 v_{\mathrm{z}2}(t), \cdots, \varepsilon_Q v_{\mathrm{z}Q}(t))$
  if $\varepsilon_j v_{\mathrm{z}j}(t) \geq x_o(t)$ **then**
    $O_j(t) = \psi$ //Selection result
  end if
  if $\varepsilon_j v_{\mathrm{z}j}(t) < x_{\mathrm{o}}(t)$ **then**
    $O_j(t) = 0$
  end if
end for
$t \leftarrow t + 1$

then $Z_i(t) = \phi$, and the ith visual zone is the target location corresponding to the maximum attention allocation result.

In Algorithm 2, $x_{\rm o}(t)$ is the maximum value in $v_{\rm z1}(t)$, $v_{\rm z2}(t), \ldots, v_{\rm zQ}(t)$ obtained from Q attention zones of different sizes, where $v_{{\rm z}j}(t)$ is the maximum attention allocation results from the jth attention zone, and $j = 1, 2, \ldots, Q$. The coefficient ε_j is the normalization factor of $v_{{\rm z}j}(t)$, and its value is the reciprocal of the number of pixels in the j-th attentional zone. The optimal attention zone is selected by comparing the amplitudes of $\varepsilon_j v_{{\rm z}j}(t)$ and $v_{\rm o}(t)$. If $\varepsilon_j v_{{\rm z}j}(t) \geq x_{\rm o}(t)$, then $O_j(t) = \psi$, and the jth attention zone is the optimal attention zone. The processing time for one frame is set as $T_{\rm f}$, and all calculations are performed in parallel.

Existing VOT methods involve a large number of network models and computational functions, leading to high algorithm complexity [3], [16]. It is difficult to design circuits to implement corresponding functions, so these methods need to run on CPUs or GPUs with von Neumann architecture, which causes the latency problem and cannot achieve higher real-time performance [13]. Different from VOT methods running on CPUs and GPUs, the proposed algorithm can design a corresponding memristor-based VOT circuit with non-von Neumann architecture to achieve online real-time tracking of the target. The algorithm parameters correspond to the circuit parameters, and the algorithm parameter settings are as follows: $\alpha = 0.1$, $\beta_1 = 0.1$, $\beta_2 = 0.1$, $V_{th1} = 0.5$, $V_{th2} = -0.5$, $V_{th3} = -5.5$, $V_{th4} = 2.5$, $\lambda_1 = 1$, $\lambda_2 = -3000$, $\xi = 0.5$, $\rho = 0.01$, $\phi = 5$, $\psi = 5$.

III. Circuit Design

A. Memristor Model

In recent years, various memristor models have been proposed for circuit simulation, such as the TiO_2 memristor model [34] and the VTEAM model [35]. In this work, we utilize a voltage-controlled threshold memristor model [36] for PSPICE simulation, which is designed based on the experimental data of Ag/AgInSbTe/Ta (AIST) memristors. Compared with other memristor models, the AIST-based memristor model can achieve stable regulation of memristance when applied pulse voltages. The derivative of state variable in the AIST-based memristor model is expressed as:

$$\frac{{\rm d}w(t)}{{\rm d}t} = \begin{cases} \mu_{\rm v}\frac{R_{\rm ON}}{D}\frac{i_{\rm off}}{i(t)-i_0}F\left(w(t)\right), & v(t) > V_{\rm T+} > 0 \\ 0, & V_{\rm T+} \geq v(t) \geq V_{\rm T-} \\ \mu_{\rm v}\frac{R_{\rm ON}}{D}\frac{i(t)}{i_{\rm on}}F\left(w(t)\right), & 0 > V_{\rm T-} > v(t), \end{cases} \tag{1}$$

$$F\left(w(t)\right) = 1 - \left(\frac{2w(t)}{D} - 1\right)^{2p}, \tag{2}$$

$$M(t) = R_{\rm ON}\frac{w(t)}{D} + R_{\rm OFF}\left(1 - \frac{w(t)}{D}\right), \tag{3}$$

where i_0, $i_{\rm off}$ and $i_{\rm on}$ are constants, $w(t)$ represents the state variable of the memristor, $\mu_{\rm v}$ is the average ion mobility and D is the semiconductor film thickness. $M(t)$ is the memristance. $R_{\rm OFF}$ and $R_{\rm ON}$ are the internal high and low memristances, respectively. p is a positive integer. $V_{\rm T+}$ and $V_{\rm T-}$ are positive and negative threshold voltages, respectively. $F(w(t))$ is a window function. The simulation parameter settings of the memristors are listed in Table I.

TABLE I
Simulation Parameters of Memristors

Parameters	Setting
D (nm)	10
$\mu_{\rm v}$ ($\rm m^2 s^{-1} \Omega^{-1}$)	3e-7
$R_{\rm ON}$ (kΩ)	1
$R_{\rm OFF}$ (kΩ)	10
$V_{\rm T+}$ (V)	0.5
$V_{\rm T-}$ (V)	-0.5
$i_{\rm off}$ (A)	8e-8
$i_{\rm on}$ (A)	1
i_0 (A)	1e-7
p	10

The configuration circuit for a memristor and the PSPICE simulation results of the memristor are shown in Fig. 2. When the voltage applied to the memristor does not exceed the thresholds, the memristance remains unchanged. When the voltage applied to the memristor exceeds $V_{\rm T+}$, the memristance first decreases rapidly and then slowly. When the voltage applied to the memristor exceeds $V_{\rm T-}$, the memristance increases rapidly and then slowly.

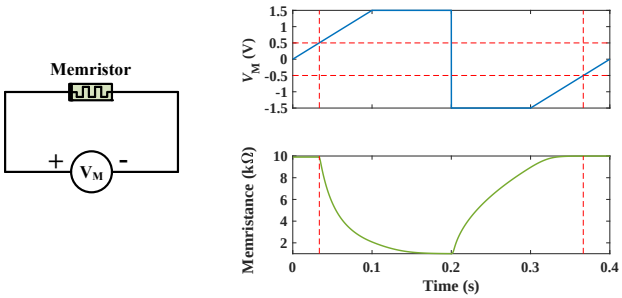


Fig. 2. Configuration circuit for a memristor and PSPICE simulation results of the memristor under different applied voltages.

B. Voltage encoding and attention encoding

Images are captured using CMOS image sensors, which output the image information in digital signal form by Analog-to-Digital Converter (ADC). For example, for an 8-bit digital signal, there are a total of 256 (2^8) gray values ranging from 0 (black) to 255 (white). The voltage encoding is used to converting image information into voltage signals. In the voltage encoding circuit as shown in Fig. 3, a Digital-to-Analog Converter (DAC) is used to encode the u-bit digital signal of a gray value into an analog voltage signal proportional to the gray value. Each input of the DAC receives one bit of the digital signal. The voltage encoding of the DAC for the gray value can be represented as:

$$V_{out} = \frac{V_{ref}}{2^u} p_{grad} \frac{R_{\rm d3}}{R_{\rm d2}}, \tag{4}$$

where p_{grad} is the gray value corresponding to the u-bit digital signal, V_{out} is the output analog voltage. V_{ref} is the reference voltage of the DAC, which is generally the maximum output

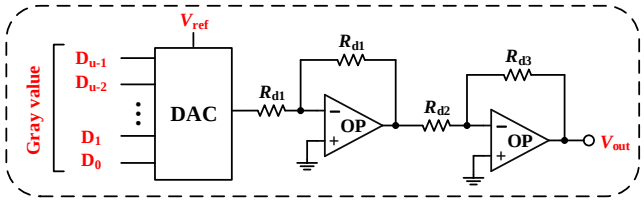


Fig. 3. Voltage encoding circuit.

voltage. In this work, u is set to 8, V_{ref} is set to +2.56 V, $R_{\mathrm{d1}} = R_{\mathrm{d2}} = 1$ MΩ, $R_{\mathrm{d3}} = 10$ MΩ. The voltage encoding time is set to 1 ms, which is also the parallel computing time (the processing time) of our tracker for one frame, i.e. $T_{\mathrm{f}} = 1$ ms.

Research indicates that the selection history of attention plays an important role in the tracking performance of target objects [37]. We combine the voltage-encoded signals of the attention zones from the previous two frames with the historical information stored in the memristor to perform attention encoding. The attention encoding fully utilize the historical information related to the target and pays attention to the dynamic change of the scene, which can adjust the tracking status online. Fig. 4 shows the attention encoding circuit for encoding an attention signal $a_{1,1}(t)$. The signals $a_{1,1}(t)$ and $e_{1,1}(t)$ can be calculated as:

$$e_{1,1}(t) = \left(\frac{R_{\mathrm{a2}}}{R_{\mathrm{a1}}} + 1\right) \frac{R_{\mathrm{a4}}}{R_{\mathrm{a3}} + R_{\mathrm{a4}}} c_{1,1}(t) - \frac{R_{\mathrm{a2}}}{R_{\mathrm{a1}}} d_{1,1}(t), \quad (5)$$

$$a_{1,1}(t) = -\left(c_{1,1}(t) - \frac{R_{\mathrm{a5}}}{M_{1,1}} e_{1,1}(t)\right), \quad (6)$$

where $c_{1,1}(t)$ and $d_{1,1}(t)$ are the voltage-encoded signals of the $(N-1)$th frame and the $(N-2)$th frame, respectively. The conductance of the memristor represents the attention weight, i.e. $w_{1,1} = \frac{1}{M_{1,1}}$.

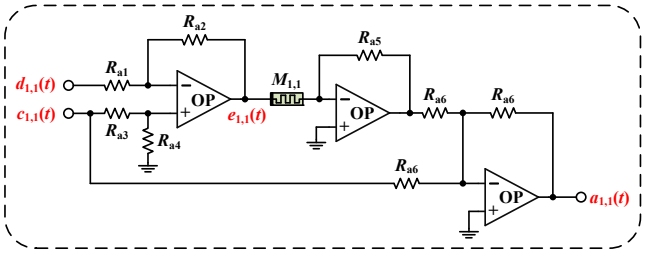


Fig. 4. Attention encoding circuit for encoding an attention signal $a_{1,1}(t)$.

Fig. 5 shows the encoding process of $a_{1,1}(t)$. The signals $c_{1,1}(t)$ and $d_{1,1}(t)$ both contain ten pulses with different voltage amplitudes. When the voltage of $e_{1,1}(t)$ is greater than the positive threshold 0.5 V, the memristance of $M_{1,1}$ decreases. When the voltage of $e_{1,1}(t)$ is less than the negative threshold -0.5 V, the memristance of $M_{1,1}$ increases. The simulation results indicates that the smaller the memristance, the smaller the voltage amplitude of $a_{1,1}(t)$. The memristance stores the voltage changes of the historical attention zones, which contains change information of the target. The more

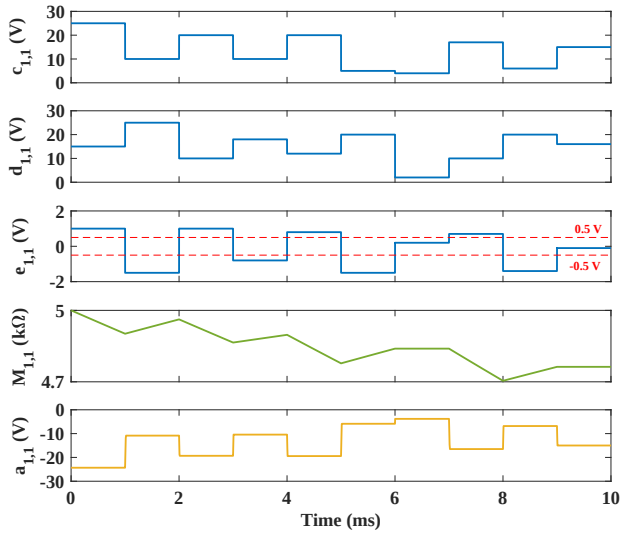


Fig. 5. Encoding process of an attention signal $a_{1,1}(t)$.

significant the voltage changes, the more the memristance adjusts, while the voltage change is less than the thresholds of the memristor, the memristance remains unchanged. The historical information recorded by the memristor is used to modify the attention encoding process of the current frame, allowing the attention encoding considers historical information such as the color, position, and shape changes of the target during the tracking process. In simulations, the circuit parameters are set as $R_{\mathrm{a1}} = R_{\mathrm{a3}} = R_{\mathrm{a6}} = 10$ MΩ, $R_{\mathrm{a2}} = R_{\mathrm{a4}} = 1$ MΩ, $R_{\mathrm{a5}} = 3$ kΩ.

C. Attention allocation and target localization

The traditional tripartite synapse structure includes presynaptic axons, postsynaptic dendrites, and astrocytes. Recent studies have shown that the GABA interneuron is involved in the activity of the tripartite synapse [38], [39], and the signaling pathway between the GABA neuron and the tripartite synapse is shown in Fig. 6. The working process of the GABA neuron and the tripartite synapse has the following three situations, and relevant works [39], [40] indicate that the regulation of IP_3 and synaptic efficacy in the working process can be described as shown in Fig. 7.

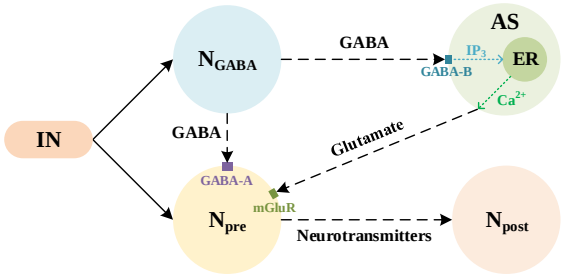


Fig. 6. Signaling pathway between the GABA neuron and the tripartite synapse. $\mathrm{N_{GABA}}$ represents the GABA interneuron, $\mathrm{N_{pre}}$ represents the presynaptic neuron, $\mathrm{N_{post}}$ represents the postsynaptic neuron, AS represents the astrocyte, and IN represents the input signal. ER is the endoplasmic reticulum of the astrocyte.

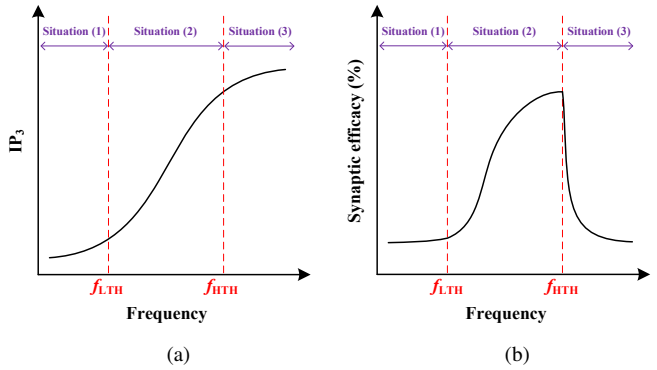


Fig. 7. Regulation process of the GABA neuron and the tripartite synapse with changes in input frequency. (a) IP_3. (b) Synaptic efficacy.

(1) When both the presynaptic neuron and the GABA neuron receive low-frequency pulse signals less than a threshold frequency f_{LTH}, the GABA neuron release GABA to both the presynaptic neuron and the astrocyte. However, the amount of inositol-1,4,5-trisphosphate (IP_3) released by GABA binding to GABA-B receptors on the astrocyte is not sufficient to induce the release of Ca^{2+} from the endoplasmic reticulum (ER) of the astrocyte. Thereby Ca^{2+} induced glutamate cannot be released from the astrocyte to enable the presynaptic neuron to release neurotransmitters. At the same time, GABA also binds to GABA-A receptors at the presynaptic neuron, resulting in inhibitory effects. Therefore, the presynaptic neuron is unable to release neurotransmitters, and synaptic efficacy cannot be enhanced.

(2) When both the presynaptic neuron and the GABA neuron receive high-frequency pulse signals with frequencies greater than f_{LTH} and less than f_{HTH}, GABA binds to GABA-B receptors on the astrocyte and produces sufficient IP_3, leading to the release of Ca^{2+} from the ER. The release of Ca^{2+} causes glutamate to be released outside of the astrocyte. Subsequently, glutamate binds to metabotropic glutamate receptors (mGluR) at the presynaptic neuron, causing the release of excitatory neurotransmitters from the presynaptic neuron. This process overcomes the inhibitory effects of GABA-A on the presynaptic neuron. As glutamate increases, the release of excitatory neurotransmitters also increases, leading to a continuous enhancement of synaptic efficacy.

(3) If the input frequency continues to increase, IP_3 in the astrocyte will also increase. When the input frequency is greater than the threshold frequency f_{HTH}, the release of Ca^{2+} stops instantly, and glutamate in the astrocyte is no longer released, resulting in the cessation of the release of excitatory neurotransmitters at the presynaptic neuron. The inhibitory effects of GABA-A once again dominate, and the presynaptic neuron is unable to release neurotransmitters, leading to a rapid decrease in synaptic efficacy.

Therefore, the interaction between the GABA neuron and the tripartite synapse exhibits selective behavior towards frequency information. Researches have shown that in the human visual system, when attention is focused on target information, the response of neurons in specific visual cortex regions significantly enhances [41], while irrelevant information are ignored by attention and cannot elicit neural responses. Thereby, the attention mechanism demonstrates selective processing of visual information. Furthermore, the accurate tracking of targets depends on the allocation of attentional resources in the human brain. Allocating more attentional resources can improve the tracking performance of target objects [42]. Inspired by the selective behavior in the working processes of the GABA neuron and the tripartite synapse, we design an attention allocation module to selectively allocate attention to different visual zone.

Fig. 8 shows the circuit structure and schematic diagram of attention allocation module (AAM). $a_i^{1,1}(t)$ is the voltage-encoded signal of the (1,1) pixel in the ith visual zone. $y_i^{1,1}(t)$ is the attention allocation signal of the (1,1) pixel in the ith visual zone. The allocation control siganl $f_i^{1,1}(t)$ can be calculated as:

$$f_i^{1,1}(t) = \frac{a_{1,1}(t) + a_i^{1,1}(t)}{2}, \tag{7}$$

where $i = 1, 2, \ldots, K$. The PSPICE simulation results of AAM are shown in Fig. 9, where the parameters are set to $V_{\mathrm{A}} = 2$ V, $R_{\mathrm{y1}} = 10$ MΩ, and $R_{\mathrm{y2}} = 3$ MΩ. The results indicate that there are three cases for the attention allocation value of signal $y_i^{1,1}(t)$, and these three cases correspond to the working process of GABA neuron and tripartite synapse:

(1) When $f_i^{1,1}(t) \leq -5.5$ V, PMOS $\mathrm{P_y}$ is fully turned on. Since the output voltage of NOT gate is 5 V, NMOS $\mathrm{N_y}$ is turned on. So the voltage of $y_i^{1,1}(t) = \frac{-V_{\mathrm{A}}+V_{\mathrm{A}}}{2} = 0$ V. Thereby the low voltage of $f_i^{1,1}(t)$ causes the attention allocation module to be inhibited, and the attention allocation value is 0 V.

(2) When -5.5 V$< f_i^{1,1}(t) < 2.5$ V, the output voltage of NOT gate is 5 V, NMOS $\mathrm{N_y}$ is turned on. As the voltage of $f_i^{1,1}(t)$ gradually increases from -5.5 V, PMOS $\mathrm{P_y}$ is

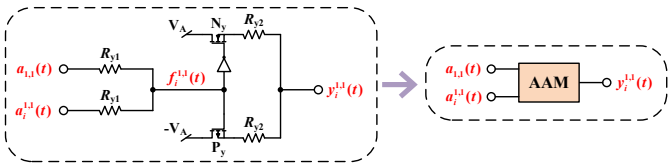


Fig. 8. Circuit structure and schematic diagram of the attention allocation module, where $i = 1, 2, \ldots, K$.

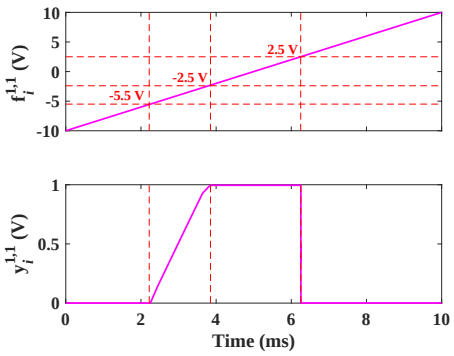


Fig. 9. PSPICE simulation results of the attention allocation module.

gradually turned off, and the voltage of $y_i^{1,1}(t)$ gradually increases. When $f_i^{1,1}(t) \geq -2.5$ V, $\mathrm{P_y}$ is completely turned off, and $y_i^{1,1}(t)$ reaches the maximum voltage $\frac{V_\mathrm{A}}{2}$. Therefore, the high voltage of $f_i^{1,1}(t)$ can overcomes the inhibitory effects and enable the attention allocation module to allocate positive attention values.

(3) When $f_i^{1,1}(t) \geq 2.5$ V, PMOS $\mathrm{P_y}$ is turned off. Since the output voltage of NOT gate is 0 V, NMOS $\mathrm{N_y}$ is also turned off. So the voltage of $y_i^{1,1}(t)$ is 0 V. Therefore, the excessive voltage of $f_i^{1,1}(t)$ exceeding the threshold causes the attention allocation module to stop working, resulting in no attention value being generated and the attention allocation value is 0 V.

The above attention allocation rule shows that when the difference in pixel information between the visual zone and the attention zone is small, the pixel information is focused by attention as important information and is allocated more attention value. When the difference in pixel information between the visual zone and the attention zone is large, the pixel information is regarded as irrelevant information and ignored by the attention, and the attention value cannot be assigned. Therefore, the designed attention allocation module demonstrates the selective processing of visual information by attention, which can allocate more attention to important information related to the target, thereby achieving target tracking.

For color images, we separately perform attention encoding and attention allocation on the RGB color components, and design an RGB-attention allocation module (RGB-AAM), as shown in Fig. 10. In RGB-AAM, $r_{1,1}(t)$, $g_{1,1}(t)$ and $b_{1,1}(t)$ are the attention signals of the red, green and blue color components, respectively. $r_i^{1,1}(t)$, $g_i^{1,1}(t)$ and $b_i^{1,1}(t)$ are the voltage-encoded signals of the red, green and blue color components, respectively. $y_{\mathrm{r}i}^{1,1}(t)$, $y_{\mathrm{g}i}^{1,1}(t)$ and $y_{\mathrm{b}i}^{1,1}(t)$ are the attention allocation values of the red, green and blue color components, respectively. $y_{\mathrm{rgb}i}^{1,1}(t)$ is the attention allocation signal of the (1,1) pixel in the ith visual zone, which can be calculated as:

$$y_{\mathrm{rgb}i}^{1,1}(t) = -\frac{R_{\mathrm{y}4}}{R_{\mathrm{y}3}}\left(y_{\mathrm{r}i}^{1,1}(t) + y_{\mathrm{g}i}^{1,1}(t) + y_{\mathrm{b}i}^{1,1}(t)\right). \quad (8)$$

In simulations, $R_{\mathrm{y}3} = 3$ MΩ and $R_{\mathrm{y}4} = 0.1$ MΩ.

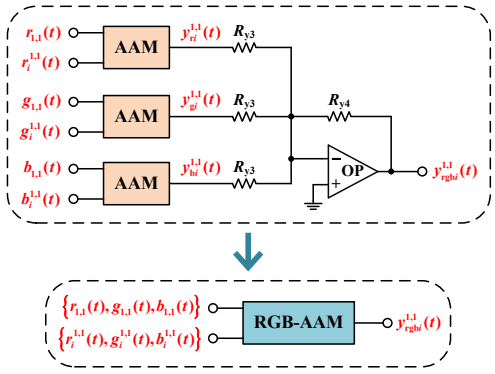


Fig. 10. Circuit structure and schematic diagram of RGB-AAM.

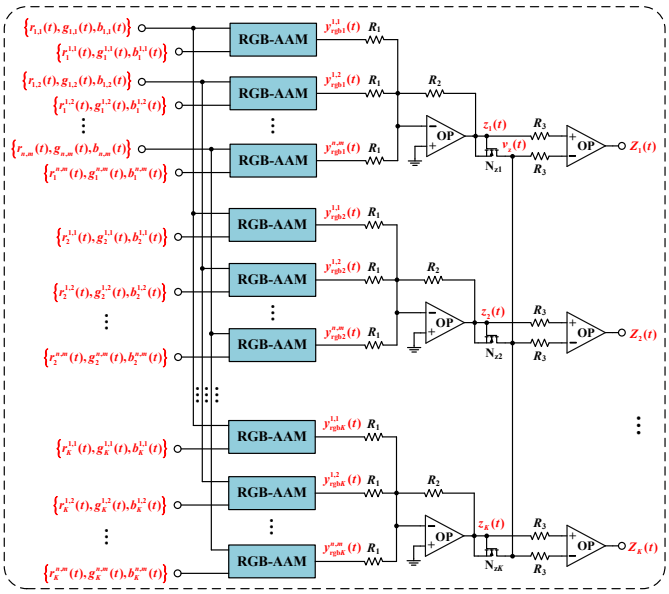


Fig. 11. Target localization circuit.

We combined RGB-AAMs with WTA principles to design a target localization circuit as shown in Fig. 11. The attention allocation result $z_i(t)$ can be calculated as:

$$z_i(t) = \frac{R_2}{R_1}\left(y_{\mathrm{rgb}i}^{1,1}(t) + y_{\mathrm{rgb}i}^{1,2}(t) + \cdots + y_{\mathrm{rgb}i}^{n,m}(t)\right). \quad (9)$$

To locate the target from K attention allocation results, the target localization circuit perform parallel computations on the attention allocation results. According to the WTA principle, all attention allocation results compete with each other, the result with the maximum value wins and is output, while the other attention allocation results are suppressed. In Fig.11, the signal $v_{\mathrm{z}}(t)$ is obtained by attention allocation results $z_1(t), z_2(t), \ldots, z_K(t)$ competing with each other, and the voltage of $v_{\mathrm{z}}(t)$ is the maximum value of all the attention allocation results. The attention allocation result that equals $v_{\mathrm{z}}(t)$ wins and outputs a pulse at the corresponding output terminal, which can be expressed as:

$$v_{\mathrm{z}}(t) = max(z_1(t), z_2(t), \ldots, z_K(t)), \quad (10)$$

$$Z_i(t) = \begin{cases} 5 \text{ V}, & z_i(t) \geq v_{\mathrm{z}}(t) \\ 0 \text{ V}, & z_i(t) < v_{\mathrm{z}}(t). \end{cases} \quad (11)$$

When $Z_i(t) = 5$ V, it means that the ith visual zone is the target location, and the attention zone of this frame is the i visual zone. In simulations, $R_1 = 1$ MΩ, $R_2 = 10$ kΩ, and $R_3 = 1$ MΩ.

D. Optimal attention zone selection

To address the negative impact of target scale changes on tracking, we set up Q attention zones of different sizes to work simultaneously in corresponding circuits, and obtain their respective maximum attention allocation results $v_{\mathrm{z}1}(t)$, $v_{\mathrm{z}2}(t)$, …, $v_{\mathrm{z}Q}(t)$.

Fig. 12 shows an optimal attention zone selection circuit for selecting the optimal attention zone. Since the different

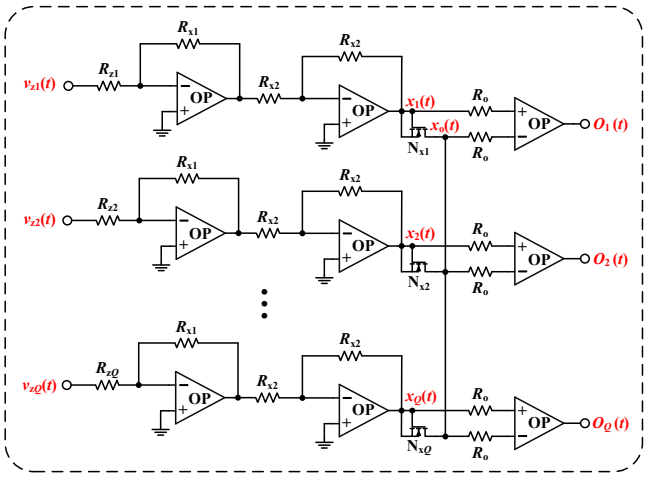


Fig. 12. Optimal attention zone selection circuit.

Fig. 13. Experimental results of the MAN on object tracking of video sequences from the OTB-100, NFS, and VOT-RTb2022 datasets. The first two rows of video sequences are from the OTB-100 dataset, the middle two rows of video sequences are from the NFS dataset, and the last two rows of video sequences are from the VOT-RTb2022 dataset. The frame number is displayed in the upper-right corner of each frame. The first column is the first frame, with four attention zones of different sizes. The yellow box represents the target bounding box tracked by the MAN, and the red box indicates the ground-truth bounding box of the datasets.

number of pixels contained in different sizes of attention zones, the normalization of attention values is achieved by resistance settings of $R_{z1}, R_{z2}, \ldots, R_{zQ}$. For example, if the number of pixels in the attention zone corresponding to $v_{z1}(t)$ is $L = m \cdot n$, then the resistance of R_{z1} is set to $R_{z1} = LR_{x1}$. In simulations, $R_{x1}, R_{x2}, \ldots, R_{xQ}$ are set to 1 kΩ. In addition to $R_{z1}, R_{z2}, \ldots, R_{zQ}$, the other resistors are set to 1 MΩ. The output signal $x_o(t)$ can be calculated as:

$$x_o(t) = max(x_1(t), x_2(t), \ldots, x_Q(t)), \tag{12}$$

$$O_j(t) = \begin{cases} 5 \text{ V}, & x_j(t) \geq x_o(t) \\ 0 \text{ V}, & x_j(t) < x_o(t), \end{cases} \tag{13}$$

where $j = 1, 2, \ldots, Q$. If O_j outputs a pulse voltage, it indicates that the attention allocation value of the jth attention zone is the largest, and the jth attention zone is the optimal attention zone.

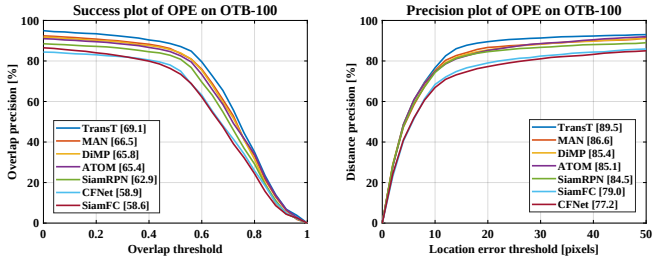
IV. Experiment and Analysis

In the experiments, we use OTB-100 [43], NFS [44] and VOT-RTb2022 [45] benchmark datasets for performance evaluation. The OTB-100 dataset is a popular tracking benchmark contains 100 videos with the frame rate of 30 FPS. The NFS dataset contains 100 videos with the frame rate of 240 FPS. The VOT-RTb2022 focused on real-time short-term tracking by bounding boxes, which contains 62 videos with the frame rate of 30 FPS. The frame image size of OTB-100 dataset is set to 240×180, and the frame image size of NFS and VOT-RTb2022 datasets are set to 240×135. The proposed tracker processes the tracking problem based on one-shot detection, which obtains the target object in the initial frame and locates the target in subsequent frames. By manually providing the initial positions of the attention zones containing the target in the first frame, the proposed tracker dynamically fine-tunes the attention to detect the target.

A. Experimental results

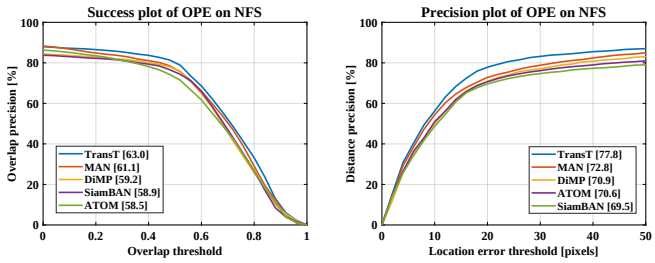
Fig. 13 shows six tracking examples from the OTB-100, NFS, and VOT-RTb2022 datasets performed by the proposed MAN. The number of attention zones is set to $Q = 4$, and their sizes are set as: 12×12, 20×20, 28×28, and 36×36. The pixel interval of visual zones is set to $s = 1$. The processing time for one frame is set to $T_f = 1$ ms. The tracking results demonstrate that the MAN can accurately track the target, and the attention zone can be adaptively selected according to the scale change of the target.

We comprehensively evaluate the performance of the proposed MAN on OTB-100 and NFS datasets on Success plots and Precision plots. In the Success plot, Overlap Score (OS) is defined as the Intersection over Union (IoU) ratios of the predicted target bounding box and the ground-truth box. When the OS of a frame is greater than a set threshold, the frame is considered successful for tracking, and the percentage of total success frames out of all frames is defined as Success Rate (SR). The success plot shows the SR from threshold 0 to 1, and the Area Under Curve (AUC) of the Success plot is used to evaluate trackers. The Precision plot is the percentage of video frames in which the Euclidean distance between the center



(a) (b)

Fig. 14. Performance on the OTB-100 dataset. (a) The AUC in the success plot. (b) The distance precision at the threshold of 20 pixels (DP20) in the precision plot.



(a) (b)

Fig. 15. Performance on the NFS dataset. (a) The AUC in the success plot. (b) The DP20 in the precision plot.

point of the predicted target bounding box and the center point of the ground-truth bounding box is less than a given threshold.

Fig. 14 shows the experimental evaluation results of all 100 videos in the OTB-100 dataset, scores are obtained by using One-Pass Evaluation (OPE) protocol. We compared the performance of our proposed MAN with trackers CFNet [4], SiamFC [6], SiamRPN [7], ATOM [46], DiMP [47], SiamBAN [8], TransT [10]. In the OPE Success plot of Fig. 14(a), the AUC of the MAN is 66.5%, which is 7.9% and 7.6% higher than the baseline trackers SiamFC and CFNet, respectively. In the OPE Precision plot of Fig. 14(b), the precision score of the MAN is 86.6%, which is higher than the trackers CFNet, SiamFC, SiamRPN, ATOM, and DiMP. Additionally, Fig. 15 shows the experimental evaluation results of all 100 videos in the NFS dataset. In the OPE Success plot of Fig. 15(a), the AUC of the MAN is 61.1%. In the OPE Precision plot of Fig. 15(b), the precision score of the MAN is 72.8%, which is higher than the trackers ATOM, SiamBAN and DiMP.

In addition, the overall performance of the trackers on VOT-RTb2022 dataset is evaluated by Expected Average Overlap (EAO), accuracy and robustness. In the short-term performance evaluation protocol of VOT2022 [45], accuracy is defined as the average overlap on frames before tracking failure, averaged over all sub-sequences. Robustness is defined as the percentage of successfully tracked subsequence frames, averaged over all sub-sequences. Tracking failure is defined as the frame at which the overlap between the predicted target bounding box and the ground-truth bounding box dropped below 0.1 and did not increase above this during the next 10 frames. EAO is the main evaluation index to evaluate the overall performance of trackers, which is obtained by combining accuracy and robustness. The larger the EAO, the better the overall performance of the tracker. Table II lists the evaluation performance of trackers MAN, ATOM, DiMP, ViTCRT, ToMP, SBT [48], and TransT on VOT-RTb2022. The comparison results show that the proposed MAN has higher EAO compared to ATOM, DiMP, ViTCRT, and ToMP.

TABLE II
PERFORMANCE COMPARISON ON VOT-RTB2022

Trackers	EAO	Accuracy	Robustness
MAN	0.480	0.736	0.785
ATOM	0.391	0.672	0.728
DiMP	0.434	0.689	0.761
ViTCRT	0.434	0.774	0.711
ToMP	0.478	0.728	0.796
TransT	0.513	0.781	0.800
SBT	0.523	0.791	0.814

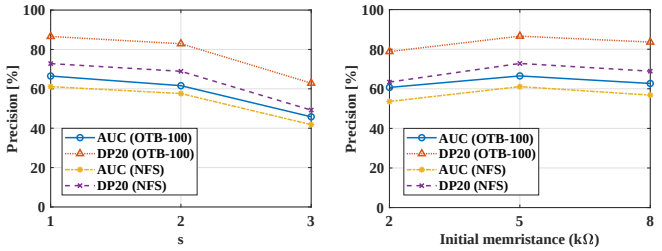
Furthermore, tracking speed is also an important evaluation metric. The speed of the tracker reflects the degree of real-time tracking, and the faster the speed, the better the real-time performance. Table III shows the comparison results in terms of tracking speed for all trackers. Because the processing time for the MAN to process each frame is 1 ms, the tracking speed of the MAN is 1000 FPS, which is higher than the other trackers. Meanwhile, the trackers CFNet, SiamFC, SiamPRN and SiamBAN are implemented on Intel CPU and NVIDIA GPU, and the trackers DiMP, ATOM, ViTCRT, TransT, ToMP, and SBT are implemented on NVIDIA GPU. Since CPUs and GPUs adopt the von Neumann architecture, where the storage unit and the computing unit are physically separated. The frequent data transmission between the storage unit and the computing unit, as well as the performance difference between them, lead to varying degrees of latency in trackers running on CPUs and GPUs. The proposed circuit integrates storage and computation using memristors, avoiding the latency problem present in trackers running on CPUs and GPUs, thus enabling tracking speeds consistent with the processing time. Therefore, the MAN achieves higher real-time tracking performance compared to the other trackers.

TABLE III
THE TRACKING SPEED OF TRACKERS

Trackers	MAN	CFNet	SiamFC	SiamRPN
Speed (FPS)	1000	75	86	160
Trackers	ATOM	DiMP	SiamBAN	TransT
Speed (FPS)	30	43	40	50
Trackers	ToMP	ViTCRT	SBT	
Speed (FPS)	26	83	62	

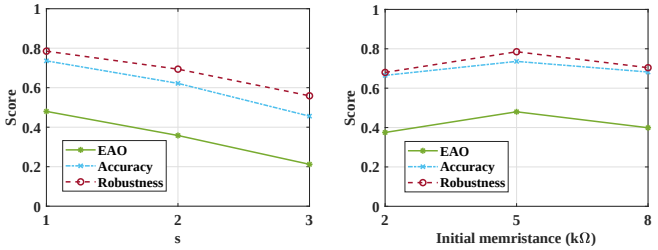
B. Parameter influence analysis

Since the irregular change of the target location in the video sequence, setting different pixel intervals for the visual zones will produce different tracking effects. Fig. 16(a) and Fig. 17(a) shows the evaluation scores of the MAN on OTB-100, NFS, and VOT-RTb2022 datasets with the pixel interval s set to 1, 2, and 3. The results indicate that as s increases, the



(a) (b)

Fig. 16. Influence of different parameters on tracking performance of NFS and OTB datasets. (a) Different pixel interval s of the visual zones. (b) Different initial memristances.



(a) (b)

Fig. 17. Influence of different parameters on tracking performance of VOT-RTb2022 dataset. (a) Different pixel interval s of the visual zones. (b) Different initial memristances.

tracking performance gradually decrease, which we analyze is mainly caused by the accumulation of errors in the attention zone.

In addition, considering that different initial memristances will affect the experimental results, we set different initial memristances for experiments. The initial memristances are set to 2 kΩ, 5 kΩ and 8 kΩ respectively, and the evaluation scores of the OTB-100, NFS and VOT-RTb2022 datasets are shown in Fig. 16(b) and Fig. 17(b). The results illustrate that the tracking performance at both 2 kΩ and 8 kΩ is lower than that of 5 kΩ. Therefore, excessively large or small initial memristances will have a negative impact on tracking process, setting the initial memristance to the intermediate value can generally achieve better tracking performance.

Furthermore, considering that the duration of the voltage applied to the memristor affects the change in memristance, different processing times will result in different tracking performance. Therefore, we analyze the impact of processing time on tracking performance by setting different processing times. As listed in Table IV, the processing times are set to 0.1 ms, 1 ms and 10 ms respectively. Combining the analysis of memristance change and tracking performance under different processing times, the results indicate that a longer processing time can lead to an excessively large change in memristance, resulting in an overly significant adjustment of the attention signal, making it difficult for our tracker to accurately track the target. Conversely, a shorter processing time can cause too small change in the memristance, leading to an insufficient adjustment of the attention signal, making it challenging for our tracker to track rapidly changing targets and potentially causing target loss. Therefore, setting an appropriate processing time can achieve better tracking performance.

TABLE IV
THE IMPACT OF PROCESSING TIME ON TRACKING PERFORMANCE

Processing times	0.1 ms	1 ms	10 ms
AUC (OTB-100)	31.2%	66.5%	24.8%
DP20 (OTB-100)	45.3%	86.6%	30.5%
AUC (NFS)	27.9%	61.1%	22.7%
DP20 (NFS)	36.4%	72.8%	25.6%
EAO (VOT-RTb2022)	0.072	0.480	0.041
Accuracy (VOT-RTb2022)	0.276	0.736	0.205
Robustness (VOT-RTb2022)	0.316	0.785	0.240

C. Power Consumption and Area Overhead

The components of MAN include the attention encoding circuit, the attention allocation module, the RGB-attention allocation module, the target localization circuit, and the optimal attention zone selection circuit. The circuit elements in MAN include memristors, NMOSs, PMOSs, resistors, operational amplifiers, 8-bit DAC, and NOT gates. The average power consumption of MAN is obtained from PSPICE simulation report, as shown in Table V. The parameters are set to $K = 1, n = 12, m = 12$. The size of the AIST memristor is 100 nm × 25 nm × 100 nm [49]. The area of an NMOS or a PMOS is approximately 0.9 $\mu\mathrm{m}^2$. Using the sheet resistance of 25 Ω/square whose poly width is 0.6 μm, the area of a 1 kΩ resistor can be calculated as $(1\ \mathrm{k\Omega} \div 25\ \Omega) \times 0.6\ \mu\mathrm{m} \times 0.6\ \mu\mathrm{m} = 14.4\ \mu\mathrm{m}^2$. The area of an operational amplifier is approximately 600 $\mu\mathrm{m}^2$, the area of a 8-bit DAC is approximately 9.0 mm^2, and the area of a NOT gate is about 1.8 $\mu\mathrm{m}^2$. Based on the above analysis, the area overhead of the proposed circuit can be estimated as shown in Table V.

TABLE V
POWER CONSUMPTION AND AREA OVERHEAD

Components	Power consumption	Area overhead
Voltage encoding circuit	0.15 mW	9.188 mm^2
Attention encoding circuit	0.58 mW	7.506 $\times 10^5$ μm^2
Attention allocation module	0.01 mW	3.744 $\times 10^5$ μm^2
RGB-attention allocation module	0.04 mW	1.255 mm^2
Target localization circuit	5.86 mW	1.828 $\times 10^2$ mm^2
Optimal attention zone selection circuit	0.05 mW	6.149 $\times 10^4 \mu\mathrm{m}^2$

V. CONCLUSION

In this paper, we propose a memristor-based attention network and its corresponding algorithm for online real-time object tracking. By constructing attention encoding circuits based on memristors, the changes of the target in the historical frame are recorded in memristors. At the same time, the memristors adjust the attention signals of the attention zone online and in real-time, allowing the MAN to track the target in dynamically changing scenes. The tracking of the target relies on the attention values selectively allocated by the attention allocation module, and the attention allocation rule is inspired by the working process of GABA interneuron and tripartite synapse. The optimal target location is determined by the target localization circuit and the optimal attention zone selection circuit based on the WTA principle. Experimental results on OTB-100, NFS, and VOT-RTb2022 benchmark datasets show that the proposed MAN achieves competitive tracking performance and has higher real-time tracking performance than the other trackers. Therefore, our work can provide support for high-real-time tracking applications of electronic devices. In the future, we will further optimize the algorithm and circuit structure of the attention network to achieve higher tracking performance.

REFERENCES

[1] J. Deng, Z. Shi, and C. Zhuo, "Energy-efficient real-time UAV object detection on embedded platforms," *IEEE Trans. Comput-Aided Des. Integr. Circuits Syst.*, vol. 39, no. 10, pp. 3123–3127, 2020.

[2] A. H. A. Zargari, M. Dautta, M. Ashrafiamiri, M. Seo, P. Tseng, and F. Kurdahi, "Newertrack: ML-based accurate tracking of in-mouth nutrient sensors position using spectrum-wide information," *IEEE Trans. Comput-Aided Des. Integr. Circuits Syst.*, vol. 39, no. 11, pp. 3833–3841, 2020.

[3] S. M. Marvasti-Zadeh, L. Cheng, H. Ghanei-Yakhdan and S. Kasaei, "Deep learning for visual tracking: A comprehensive survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 5, pp. 3943-3968, 2022.

[4] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, and P. H. S. Torr, "End-to-end representation learning for correlation filter based tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 5000–5008.

[5] P. Yang, Q. Wang, J. Dou, and L. Dou, "SDCS-CF: Saliency-driven localization and cascade scale estimation for visual tracking," *J. Vis. Commun. Image Represent.*, vol. 98, no. 4, p. 104040, 2024.

[6] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "Fully-convolutional siamese networks for object tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2016, pp. 850–865.

[7] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu, "High performance visual tracking with siamese region proposal network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 8971–8980.

[8] Z. Chen, B. Zhong, G. Li, S. Zhang, and R. Ji, "Siamese box adaptive network for visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 6667–6676.

[9] B. Wei, H. Chen, Q. Ding, and H. Luo, "SiamSTC: Updatable siamese tracking network via spatio-temporal context," *Knowledge-Based Syst.*, vol. 263, no. 5, p. 110286, 2023.

[10] X. Chen, B. Yan, J. Zhu, D. Wang, X. Yang, and H. Lu, "Transformer tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 8122–8131.

[11] C. Mayer *et al.*, "Transforming model prediction for tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 8721–8730.

[12] E. D. Nardo and A. Ciaramella, "Tracking vision transformer with class and regression tokens," *Inf. Sci.*, 2023, vol. 619, pp. 276–287.

[13] A. Ammar, H. B. Fredj, S. Sghaier, and C. Souani, "Algorithmic optimization effects on energy consumption for real time object tracking," in *Proc. International Conference on Control, Automation and Diagnosis (ICCAD)*, 2021, pp. 1–4.

[14] R. Liu, L. Zhang, J. Wang, H. Yang, and Y. Liu, "Petri: Reducing bandwidth requirement in smart surveillance by edge-cloud collaborative adaptive frame clustering and pipelined bidirectional tracking," in *Proc. 58th ACM/IEEE Design Automation Conference (DAC)*, 2021, pp. 421–426.

[15] A. Fjodorov, S. Ulp, M. M. Alam, and A. Kuusik, "Inertial and positioning sensors fusion for indirect location tracking in warehouse inventory management," in *Proc. International Conference on Control, Automation and Diagnosis (ICCAD)*, 2023, pp. 1–7.

[16] L. Jiao, D. Wang, Y. Bai, P. Chen, and F. Liu, "Deep learning in visual tracking: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 9, pp. 5497–5516, 2023.

[17] X. Yang, B. Taylor, A. Wu, Y. Chen, and L. O. Chua, "Research progress on memristor: From synapses to computing systems," *IEEE Trans. Circuits Syst. I-Regul. Pap.*, vol. 69, no. 5, pp. 1845–1857, 2022.

[18] L. Chua, "Memristor-The missing circuit element," *IEEE Trans. Circuit Theory*, vol. 18, no. 5, pp. 507–519, 1971.

[19] D. B. Strukov, G. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *Nature*, vol. 453, pp. 80–83, 2008.

[20] R. Yan, Q. Hong, C. Wang, J. Sun, and Y. Li, "Multilayer memristive neural network circuit based on online learning for license plate detection," *IEEE Trans. Comput-Aided Des. Integr. Circuits Syst.*, vol. 41, no. 9, pp. 3000–3011, 2022.

[21] Z. Deng, C. Wang, H. Lin, and Y. Sun, "A memristive spiking neural network circuit with selective supervised attention algorithm," *IEEE Trans. Comput-Aided Des. Integr. Circuits Syst.*, vol. 42, no. 8, pp. 2604–2617, 2023.

[22] C. Wang, D. Tang, H. Lin, F. Yu, and Y. Sun, "High-dimensional memristive neural network and its application in commercial data encryption communication," *Expert Syst. Appl.*, vol. 242, p. 122513, 2024.

[23] Q. Deng, C. Wang, and H. Lin, "Memristive Hopfield neural network dynamics with heterogeneous activation functions and its application," *Chaos Solitons Fractals*, vol. 178, p. 114387, 2024.

[24] C. Yang, X. Wang, and Z. Zeng, "Full-circuit implementation of transformer network based on memristor," *IEEE Trans. Circuits Syst. I-Regul. Pap.*, vol. 69, no. 4, pp. 1395–1407, 2022.

[25] C. Xu, C. Wang, J. Jiang, J. Sun, and H. Lin, "Memristive circuit implementation of context-dependent emotional learning network and its application in multitask," *IEEE Trans. Comput-Aided Des. Integr. Circuits Syst.*, vol. 41, no. 9, pp. 3052–3065, 2022.

[26] H. Bao, Z. Hua, H. Li, M. Chen, and B. Bao, "Memristor-based hyperchaotic maps and application in auxiliary classifier generative adversarial nets," *IEEE Trans. Ind. Inform.*, vol. 18, no. 8, pp. 5297–5306, 2022.

[27] Q. Deng, C. Wang, J. Sun, Y. Sun, J. Jiang, H. Lin, and Z. Deng, "Nonvolatile CMOS memristor, reconfigurable array, and its application in power load forecasting," *IEEE Trans. Ind. Inform.*, vol. 20, no. 4, pp. 6130–6141, 2024.

[28] Y. Zhang, G. He, G. Wang, and Y. Li, "XBarNet: Computationally efficient memristor crossbar model using convolutional autoencoder," *IEEE Trans. Comput-Aided Des. Integr. Circuits Syst.*, vol. 41, no. 12, pp. 5489–5500, 2022.

[29] A. Bhattacharjee, L. Bhatnagar, Y. Kim, and P. Panda, "Neat: Nonlinearity aware training for accurate, energy-efficient, and robust implementation of neural networks on 1T-1R crossbars," *IEEE Trans. Comput-Aided Des. Integr. Circuits Syst.*, vol. 41, no. 8, pp. 2625–2637, 2022.

[30] B. Li, Q. Xie, and G. Shi, "A memristor crossbar based lyapunov equation solver," *IEEE Trans. Comput-Aided Des. Integr. Circuits Syst.*, vol. 42, no. 11, pp. 4324–4328, 2023.

[31] H. Lin, C. Wang, L. Cui, Y. Sun, C. Xu, and F. Yu, "Brain-like initial-boosted hyperchaos and application in biomedical image encryption," *IEEE Trans. Ind. Inform.*, vol. 18, no. 12, pp. 8839–8850, 2022.

[32] Y. Xiao *et al.*, "A review of memristor: Material and structure design, device performance, applications and prospects," *Sci. Technol. Adv. Mater.*, vol. 24, no. 1, p. 2162323, 2023.

[33] S. Zhang, C. Li, J. Zheng, X. Wang, Z. Zeng, and G. Chen, "Memristive autapse-coupled neuron model with external electromagnetic radiation effects," *IEEE Trans. Ind. Electron.*, vol. 70, no. 11, pp. 11 618–11 627, 2023.

[34] D. Biolek, V. Biolková, and Z. Biolek, "SPICE model of memristor with nonlinear dopant drift," *Radioengineering*, vol. 18, no. 2, pp. 210–214, 2009.

[35] S. Kvatinsky, M. Ramadan, E. G. Friedman, and A. Kolodny, "VTEAM: A general model for voltage-controlled memristors," *IEEE Trans. Circuits Syst. II-Express Briefs*, vol. 62, no. 8, pp. 786–790, 2015.

[36] Y. Zhang, X. Wang, Y. Li, and E. G. Friedman, "Memristive model for synaptic circuits," *IEEE Trans. Circuits Syst. II-Express Briefs*, vol. 64, no. 7, pp. 767–771, 2017.

[37] J Theeuwes, "Goal-driven, stimulus-driven, and history-driven selection," *Curr. Opin. Psychol.*, vol. 29, pp. 97–101, 2019.

[38] G. Perea *et al.*, "Activity-dependent switch of GABAergic inhibition into glutamatergic excitation in astrocyte-neuron networks," *eLife*, vol. 5, p. e20362, 2016.

[39] J. Liu *et al.*, "Exploring self-repair in a coupled spiking astrocyte neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 865–875, 2019.

[40] J. Liu *et al.*, "GABA regulation of burst firing in hippocampal astrocyte neural circuit: A biophysical model," *Front. Cell. Neurosci.*, vol. 13, no. 335, pp. 1–14, 2019.

[41] T. Moore and M. Zirnsak, "Neural mechanisms of selective visual attention," *Annu. Rev. Psychol.*, vol. 68, no. 5, pp. 47–72, 2017.

[42] S. E. Donohue, J. M. Hopf, M. V. Bartsch, M. A. Schoenfeld, H. Heinze, M. G. Woldorff, "The rapid capture of attention by rewarded objects," *J. Cogn. Neurosci.*, vol. 28, no. 4, pp. 529–541, 2016.

[43] Y. Wu, J. Lim, and M.-H. Yang, "Object tracking benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1834–1848, 2015.

[44] H. K. Galoogahi, A. Fagg, C. Huang, D. Ramanan, and S. Lucey, "Need for speed: A benchmark for higher frame rate object tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 1134–1143.

[45] M. Kristan *et al.*, "The tenth visual object tracking VOT2022 challenge results," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2023, pp. 431–460.

[46] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "ATOM: Accurate tracking by overlap maximization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 4655–4664.

[47] G. Bhat, M. Danelljan, L. Gool, and R. Timofte, "Learning discriminative model prediction for tracking," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 6181–6190.

[48] F. Xie, C. Wang, G. Wang, Y. Cao, W. Yang, and W. Zeng, "Correlation-aware deep tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 8741–8750.
[49] Y. Zhang and Z. Zeng, "A multi-functional memristive pavlov associative memory circuit based on neural mechanisms," *IEEE Trans. Biomed. Circuits Syst.*, vol. 15, no. 5, pp. 978–993, 2021.

Zekun Deng received the B.S. degree in communications engineering from Northeastern University, Shenyang, China, in 2018. He is currently pursuing the Ph.D. degree in College of Computer Science and Electronic Engineering, Hunan University, Changsha, China. His research interests include memristive neural network circuits, algorithm design of neural networks, and analog implementation of neuromorphic systems.

Chunhua Wang received the M.S. degree from Zhengzhou University, Zhengzhou, China, in 1994, and the Ph.D. degree from Beijing University of Technology, Beijing, China, in 2003. He is currently Professor of College of Information Science and Engineering, Hunan University, Changsha, China. He is Doctor tutor, director of advanced communication technology key laboratory of Hunan universities, member of academic committee of Hunan university, director of chaos and nonlinear circuit professional committee of circuit and system branch of China electronic society. Now, his research interests include memristor circuit, complex networks, chaotic circuit, chaos secure communication, current-mode circuit and neural networks based on memristor. He has presided over 8 national and provincial projects, and published more than 200 papers, among which more than 180 were retrieved by SCI.

Hairong Lin received M.S. and Ph.D. degrees in information and communication engineering and computer science and technology from Hunan University, Changsha, China, in 2015 and 2021, respectively. From 2022 to 2023, he was a Postdoctoral Fellow with the School of Computer Science and Electronic Engineering, Hunan University, China. He is currently an Associate Professor at the School of Electronic Information, Central South University, Changsha, China. He is a member of the Chaos and Nonlinear Circuit Professional Committee of Circuit and System Branch of China Electronic Society. He has presided over three national and provincial projects, and published more than 50 papers in related international journals, such as IEEE-TIE, IEEE-TII, IEEE-TCAD, etc. His research interests include memristive neural networks, chaotic systems and circuits, information and network security, and Internet of Things

Quanli Deng received the B.S. degree in Xiangtan University School of Physics and Optoelectronics, Xiangtan, China, in 2016, and the M.S. degree in College of Computer Science and Electronic Engineering, Hunan University, Changsha, China, in 2020. He is currently pursuing the Ph.D. degree in College of Computer Science and Electronic Engineering, Hunan University, China. His research interests include modeling and analysis of neural systems, fundamental theory of nonlinear systems and circuits, and analog implementation of neuromorphic systems.

Yichuang Sun (M'90–SM'99) received the B.Sc. and M.Sc. degrees from Dalian Maritime University, Dalian, China, in 1982 and 1985, respectively, and the Ph.D. degree from the University of York, York, U.K., in 1996, all in communications and electronics engineering.

Dr. Sun is currently Professor of Communications and Electronics, Head of Communications and Intelligent Systems Research Group, and Head of Electronic, Communication and Electrical Engineering Division in the School of Engineering and Computer Science of the University of Hertfordshire, UK. He has published over 330 papers and contributed 10 chapters in edited books. He has also published four text and research books: Continuous-Time Active Filter Design (CRC Press, USA, 1999), Design of High Frequency Integrated Analogue Filters (IEE Press, UK, 2002), Wireless Communication Circuits and Systems (IET Press, 2004), and Test and Diagnosis of Analogue, Mixed-signal and RF Integrated Circuits - the Systems on Chip Approach (IET Press, 2008). His research interests are in the areas of wireless and mobile communications, RF and analogue circuits, microelectronic devices and systems, and machine learning and deep learning.

Professor Sun was a Series Editor of IEE Circuits, Devices and Systems Book Series (2003-2008). He has been Associate Editor of IEEE Transactions on Circuits and Systems I: Regular Papers (2010-2011, 2016-2017, 2018-2019). He is also Editor of ETRI Journal, Journal of Semiconductors, and Journal of Sensor and Actuator Networks. He was Guest Editor of eight IEEE and IEE/IET journal special issues: High-frequency Integrated Analogue Filters in IEE Proc. Circuits, Devices and Systems (2000), RF Circuits and Systems for Wireless Communications in IEE Proc. Circuits, Devices and Systems (2002), Analogue and Mixed-Signal Test for Systems on Chip in IEE Proc. Circuits, Devices and Systems (2004), MIMO Wireless and Mobile Communications in IEE Proc. Communications (2006), Advanced Signal Processing for Wireless and Mobile Communications in IET Signal Processing (2009), Cooperative Wireless and Mobile Communications in IET Communications (2013), Software-Defined Radio Transceivers and Circuits for 5G Wireless Communications in IEEE Transactions on Circuits and Systems-II (2016), and the 2016 IEEE International Symposium on Circuits and Systems in IEEE Transactions on Circuits and Systems-I (2016). He has also been widely involved in various IEEE technical committee and international conference activities.