This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648,

An Efficient Human Activity Recognition In-Memory Computing Architecture Development for Healthcare Monitoring

Xiaoyue Ji, *Member, IEEE*, Zhekang Dong, *Senior Member, IEEE*, Liyan Zhu, Chenhao Hu, Chun Sing Lai, *Senior Member, IEEE*

***Abstract*—Human activity recognition has played a crucial role in healthcare information systems due to the fast adoption of artificial intelligence (AI) and the internet of thing (IoT). Most of the existing methods are still limited by computational energy, transmission latency, and computing speed. To address these challenges, we develop an efficient human activity recognition in-memory computing architecture for healthcare monitoring. Specifically, a mechanism-oriented model of Ag/a-Carbon/Ag memristor is designed, serving as the core circuit component of the proposed in-memory computing system. Then, one-transistor-two-memristor (1T2M) crossbar array is proposed to perform high-efficiency multiply-accumulate (MAC) operation and high-density memory in the proposed scheme. To facilitate understanding of the proposed efficient human activity recognition in-memory computing design, self-attention ConvLSTM module, multi-head convolutional attention module, and recognition module are proposed. Furthermore, the proposed system is applied to perform human activity recognition, which contains eleven different human activities, including five different postural falls, and six basic daily activities. The experimental results show that the proposed system has advantages in recognition performance (⩾ 0.20% accuracy, ⩾ 1.10% F1-score) and time consumption (approximately 8∼10 times speed up) compared to existing methods, indicating an advancement in smart healthcare applications.**

***Index Terms*—Human activity recognition, in-memory computing, memristor, healthcare monitoring**

I. Introduction

With the rapid development of big data, artificial intelligence (AI), and the internet of thing (IoT), the increasing number of wearable sensors integrated with intelligent processors are wildly used to health monitoring in home environment [1]. Human activity recognition can understand human interaction with the living environment and automatically discover behavioural patterns within multimodal information [2]. Human activity recognition has been an active research area in smart healthcare, which can effectively enhance the level of patient rehabilitation and medical decision systems [3]. Considering the privacy, comfort, and portability, some researchers used wearable sensors for human activity recognition [4-6]. However, wearable sensor can only capture the local movement information, which may lead to low precision for complex movements. To solve this problem, machine learning and deep learning techniques have achieved tremendous successes in human activity recognition in smart home applications, which can extract necessary features from multimodal information and perform fusion analysis [7-14]. To handle time-domain data, machine learning-based framework was proposed to perform human activity recognition [7]. Convolutional neural networks (CNNs) with different convolutional kernels were used to capture hidden patterns from visual data in [8-11], showing good performance in human activity recognition. Compared with single modal data, multimodal data fusion can provide more comprehensive information for human activity recognition. Based on this, multimodal information processing architectures were proposed in [12-14], which achieved competitive recognition accuracy and exceeded previous results in human activity recognition. However, the above-mentioned methods based on conventional computing architectures (e.g., CPU, GPU, etc.) have still suffered from computational energy, transmission latency, and computing speed issues.

Consequently, some researchers are turning to a promising computing architecture to overcome these limitations, which is called ‘in-memory computing’ [15]. In-memory computing using non-volatile memory devices executes approximating matrix-vector multiplication computation in a fast, highly parallel, and energy-efficient manner [16]. A multimodal neuromorphic sensory-processing system using memristive circuit was designed in [17], which can realize indoor human behavior recognition with relative high accuracy (about 90.37%). A heterogeneous tightly coupled clustered in-memory computing architecture was proposed in [18], which can perform end-to-end inference of a full mobile-grade DNN (MobileNetV2). In [19], researches used duplex two-dimensional material to construct an in-memory

Manuscript received December 07, 2023.

This work was supported in part by the National Postdoctoral Researcher Support Program under Grant GZB20230356, the Shuimu Tsinghua Scholar program under Grant 2023SM035, the National Natural Science Foundation of China under Grant 62206062, and the Fundamental Research Funds for the Provincial University of Zhejiang under Grant GK229909299001-06. *(Corresponding authors: Zhekang Dong and Chunsing Lai).*

X. Ji is with the Center for Brain-Inspired Computing Research (CBICR), Beijing Innovation Center for Future Chip, Optical Memory National Engineering Research Center, Department of Precision Instrument, Tsinghua University, Beijing 100084, China. (e-mail: jixiaoyue@mail.tsinghua.edu.cn).

Z. Dong, L. Zhu, and C. Hu, are with the School of Electronics and Information, Hangzhou Dianzi University, Hangzhou 310018, China, and also with the Zhejiang Provincial Key Lab of Equipment Electronics, Hangzhou 310018, China (e-mail: englishp@126.com; 232040154@hdu.edu.cn; chenhaohu@hdu.edu.cn;).

C. S. Lai is with the Department of Electronic and Electrical Engineering, Brunel University London, London, UB8 3PH, UK and also with the School of Automation, Guangdong University of Technology, Guangzhou, China 510006 (email: chunsing.lai@brunel.ac.uk).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648,

computing system, achieving 99.86% accuracy in a nonlinear localization task. Based on human brain mechanisms, a hierarchical interactive in-memory computing system was proposed in [20], aiming to solve balancing problem between computing accuracy and energy consumption for conventional computing architecture. A memristor-based hardware implementation of generative complex networks was designed in [21], achieving good performance in speech recognition tasks. Although existing in-memory computing systems show remarkable advantages in computational energy, transmission latency, and computing speed, several fundamental challenges are still open.

Firstly, limited by device performance issues, the existing in-memory computing systems are difficult to popularize and apply. An essential way is the computationally efficient and accurate modelling of the memristor device. secondly, current studies are always rarely considered the long-range spatial dependencies when they process single nature data (time-series) collected from wearable sensors and context-aware devices. Thirdly, the research gaps exist in developing multimodal fusion strategy for time-series data and visual data in smart home environment.

To fully exploit the potential of in-memory computing system in home scenarios, this work aims to investigate an efficient human activity recognition in-memory computing architecture for healthcare monitoring. The main contributions of this work are concluded below:

1) As the core circuit component, a mechanism-oriented model of Ag/a-Carbon/Ag memristor is proposed, which provides the possibility to explore the dynamics of electronic resistive switching memory (ERSM) behavior and helps realize the parallel computing in the proposed system.

2) The circuit design of self-attention convolutional long short-term memory (ConvLSTM) module is developed, which can effectively handle time-series data and retrieve the high-level spatial and temporal features with local-global dependences.

3) Compared with the existing in-memory computing systems, the proposed human activity recognition in-memory computing architecture can fuse the multimodal information simultaneously, showing respectable performance (i.e., recognition accuracy and computational efficiency) in smart healthcare applications.

The remainder of this work is structured as follows. Section II describes the overall architecture of the proposed system. In Section III, the memristor crossbar array is introduced based on the proposed mechanism-oriented model. Section IV describes the specific circuit design of the self-attention ConvLSTM module, the multi-head convolutional attention module, and the recognition module. In Section V, the proposed system is applied to perform human activity recognition. Finally, the entire work is summarized in Section VI.

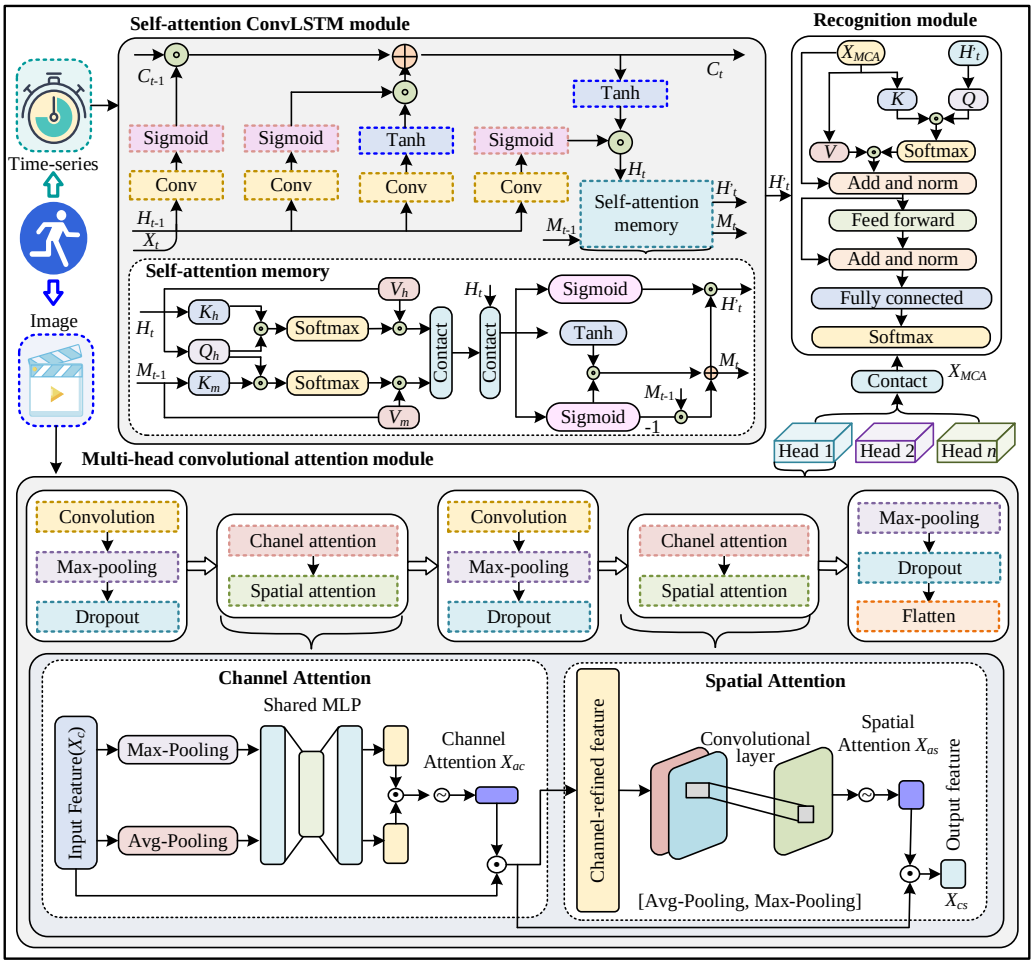


Fig. 1. Schematic diagram of the proposed human activity recognition in-memory computing system in smart healthcare applications.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648,

II. An Efficient Human Activity Recognition In-Memory Computing Architecture

In this paper, we propose an efficient human activity recognition in-memory computing architecture for healthcare monitoring. Notably, the parameters of the proposed system are computed and stored in the crossbar array to facilitate smart healthcare applications. The proposed system consists of three modules to recognize human activities from the multimodal data, as shown in Fig. 1.

Self-attention ConvLSTM module: The time-series data X_t collected from several wearable sensors are injected to the self-attention ConvLSTM module (SA-ConvLSTMM) to capture the high-level spatial and temporal features H'_t with local-global dependences. The SA-ConvLSTMM with cascaded configuration contains two units, i.e., the ConvLSTM unit and the self-attention memory unit. Specifically, the ConvLSTM unit is used to extract features with spatial and temporal dependencies form the time-series data [22]. Then, the self-attention memory unit constructed based on self-attention mechanism is utilized, which can realize features memory with long-range spatial and temporal dependencies through feature aggregation and memory updating [23]. Here, the input, hidden state and cell state of the ConvLSTM unit are denoted as X_t, H_t, and C_t, respectively. The input, memory, and output of self-attention memory unit are expressed by H_t, M_t, and H'_t, respectively.

Multi-head convolutional attention module: The image data X_i collected from multiple cameras are sent to the multi-head convolutional attention module (MHCAM) to capture the relevant patterns H_i from the channel and spatial dimensions. In the MHCAM, each head contains two convolutional operations, three max-pooling operations, three dropout operations, and two channel-spatial attention operations. Firstly, the feature X_c is obtained by a serial operation (containing the convolutional, max-pooling, and dropout). Then, the channel-spatial attention unit captures channel and spatial features from the input feature X_c. Specifically, the channel attention unit allows the MHCAM pay more attention to the channel information X_{ac}, while suppresses the unnecessary information. The spatial attention unit enables the MHCAM to focus on important spatial information X_{sc} by performing the feature filtering operation at different locations in the same spatial dimension. Notably, the channel attention unit and spatial attention unit are cascaded structure, and the output X_{cs} of these two units is entered into the next convolutional unit.

Recognition module: The output features from SA-ConvLSTMM H'_t and MHCAM H_i are sent into the recognition module to recognize human activities through the cross-modal transformer unit. Specifically, the crossmodal transformer mechanism is employed [24], taking H'_t as queries and H_i as keys and values, and then followed by fully connected layer to prevent over-fitting. Furthermore, the softmax layer is utilized to calculate the probability distribution of each human activities, and the final output are determined by the highest probability score.

III. Memristor Technology

Before the circuit design of the proposed efficient human activity recognition in-memory computing system, it is necessary to introduce memristor technology.

A. Memristor Characteristics and Modelling

In this work, the Ag/a-Carbon/Ag memristor device is fabricated on quartz plate and the schematic representation of Ag/a-Carbon/Ag memristor is shown in Fig. 2(a).

Firstly, the quartz plate is cleaned in deionized water, ethyl alcohol, and acetone using ultrasonic cleaner. Secondly, the a-Carbon function layer is synthesized on the dried quartz plate using flame method. Specifically, the quartz plate is transferred on top of the alcohol burner, so that its surface temperature reaches 800℃. Then, the quartz plate is annealed at 350℃ for 180 minutes and naturally cooled to room temperature. Furthermore, the quartz plate is placed to plasma cleaner and treated with air plasma. This leads to the formation of an a-Carbon function layer with a thickness of 185nm. Finally, the top/bottom Ag electrodes with a thickness of 20nm and 120nm are deposited on the a-Carbon function layer using magnetron sputtering method.

For electrical characterization, the voltage-current (V-I) curve is measured using an electrochemical workstation (CHI-600D). In the static measurement, scan rate and scan amplitude are fixed at 0.5V/s and ± 3V. The Ag/a-Carbon/Ag memristor exhibits typical electronic resistive switching memory (ERSM) behavior, as illustrated in Fig. 2(a). To demonstrate its stability, over 1000 cycles of I–V curves are measured on the same memristor, as depicted in Fig. 2(b). The ERMS behavior is well maintained, indicating high cycle-to-cycle (C2C) stability of the fabricated Ag/a-Carbon/Ag memristor. Extensive overlap of I–V curves measured by 160 randomly chosen memristors is shown in Fig. 2(c), demonstrating good device-to-device (D2D) stability of the fabricated Ag/a-Carbon/Ag memristors. Furthermore, a stable resistance ratio (about 10^4) between the high resistance state (HRS) and low resistance state (LRS) is maintained for 60 hours at 0.5 V read voltage, as depicted in Fig. 2(d).

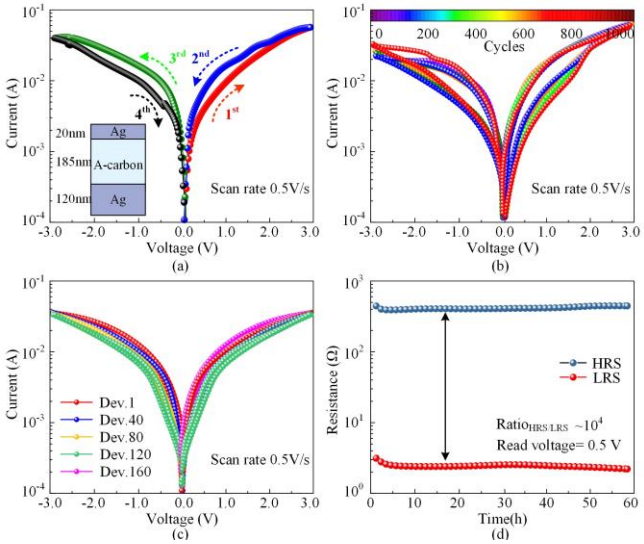


Fig. 2. Electrical characterization of Ag/a-Carbon/Ag memristor. (a) V-I curve; (b) C2C stability; (c) D2D stability; (d) The measurement of resistance.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648,

To investigate the physical mechanism of Ag/a-Carbon/Ag memristor, the electrons trapping and de-trapping processes are further analysed in this paper, as shown in Fig. 3.

At beginning, the Ag/a-Carbon/Ag memristor is in a high-resistance state (HRS). Ag atoms at top electrode have not been ionized, and unfilled traps randomly distribute in the a-Carbon function layer, as shown in Fig. 3(a). When the scanning voltage gradually increases, some traps are filled with electrons, resulting in the current builds steadily, as shown in Fig. 3(b). When the traps are fully filled with electrons, the conduction filament is formed, mean that the Ag/a-Carbon/Ag memristor switching from the HRS to the low-resistance state (LRS) and the "SET" process is completed, shown in Fig. 3(c). When a reverse scanning voltage is applied, the trapped electrons escape the traps, and the conduction filaments are broken, mean that the Ag/a-Carbon/Ag memristor switching from the LRS to the HRS and the "RESET" process is completed, shown in Fig. 3(d). According to above analysis, the ERSM behavior in Ag/a-Carbon/Ag memristor is governed by the electrons trapping and de-trapping processes, satisfying space-charge limited current (SCLC) mechanism in high electric field [25] and the field-assisted Frenkel-Poole mechanism in low electric field [26].

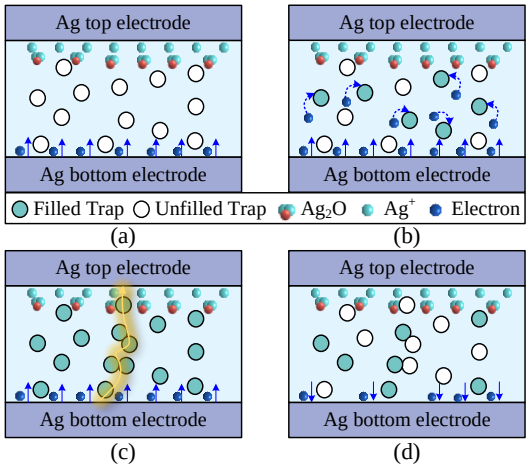


Fig. 3. Physical mechanism of Ag/a-Carbon/Ag memristor. (a) Unfilled Traps; (b) Traps gradually filled; (c) the conduction filament is formed; (d) The conduction filament is broken.

Based on this, a mechanism-oriented model of Ag/a-Carbon/Ag memristor is proposed to explore the dynamics of the ERMS behavior in simulation. The V-I relationship can be mathematically expressed by:

$$i(t)=\lambda x(t)^{n}\frac{v(t)^{m}}{D^{2m+1}}+\gamma\left(1-x(t)\right)^{\tau}\left(1-\exp\left(-\frac{q_0 v(t)}{K_B T}\right)\right) \quad (1)$$

where $i(t)$ is the current in the mechanism-oriented memristor model, $v(t)$ is the applied voltage of the memristor model, λ, n, m, τ and γ are the fitting parameters, D is the thickness of the a-Carbon function layer, q_0 denotes the electric charge in vacuum, K_B is the Boltzmann constant, T is the temperature. $x(t)$ denotes the state variable, which is expressed as follows:

$$\frac{dx}{dt}=\alpha\sinh\beta v(t)\,f(x)-\eta x \quad (2)$$

where α, β, and η are the fitting parameters, $f(x)$ is the window function [27]. Specifically, $f(x)$ can be written as follows:

$$f(x)=0.25\begin{bmatrix}\left(sign(v)+1\right)\left(sign(1-x)+1\right)\\+\left(sign(-v)+1\right)\left(sign(x)+1\right)\end{bmatrix} \quad (3)$$

The fitting results are illustrated in Fig. 4, where the solid spheres represent the experimental data obtained from the Ag/a-Carbon/Ag memristor, and the solid lines represent the V-I curves of the proposed model.

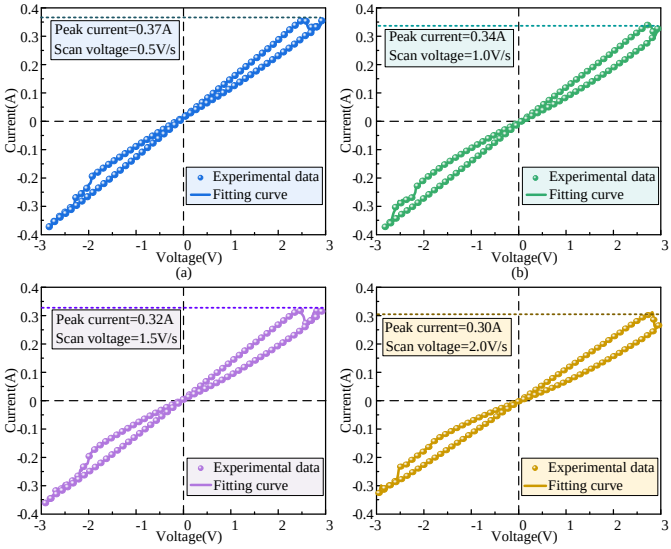


Fig. 4. The fitting results of the proposed mechanism-based memristor model.

To rigorously evaluate the accuracy of mechanism-oriented memristor model, both the human visual system (HVS) for subjective validation and the relative root mean squared error (RRMSE) as an objective metric are employed [28]. Notably, the optimal fitting parameters are obtained using the whale optimization algorithm (WOA) [29]. The specific parameters setting is provided as follows: λ=2.73×10^{-3}, n=2.00, m=1.43, τ=2.52 × 10^{-5}, γ=0.11, α=0.81, β=0.11, η=0.048, D=1.85×10^{-7}, q_0=1.60×10^{-7}, K_B=1.38×10^{-23}, T=25.00. When the voltage scan rate increases from 0.5V/s to 2.0V/s, the preserved asymmetry of I–V curve and the continuity of current variation can be both well maintained, albeit with a marginal decrement in peak current from 0.37A to 0.30A, as depicted in Fig. 4. Based on HVS, the experimental results demonstrated that the fitting curves coincides exactly with the experimental data. From an objective perspective, the mechanism-oriented memristor model accuracy is further verified by RRMSE. The average RRMSE can be obtained with the value of 0.081%, indicating that the constructed mechanism-oriented memristor model is capable to characterize the performance of the Ag/a-Carbon/Ag memristor.

B. Memristor-based Crossbar Array

Considering memristors have the unique characteristics, such as low power, nonvolatility, high density, great scalability, and compatibility with complementary metal oxide semiconductor (CMOS), which provides a subversive way for in-memory computing system [16-21]. Thus, one-transistor-two-memristor (1T2M) crossbar array is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648,

designed to realize the high-efficiency multiply-accumulate (MAC) operation and high-density memory while avoid sneak path issue, as shown in Fig. 5.

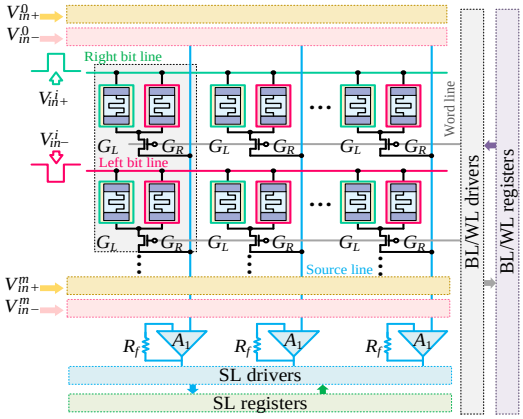


Fig. 5. 1T2M crossbar array

Compared with traditional one-transistor-one-memristor (1T1M) crossbar array, each 1T2M cell in crossbar array are connected to the same word line (WL) and source line (SL), and subtraction can be performed directly in the current domain. In Fig. 5, each 1T2M cell in the crossbar array can be implemented by two Ag/a-Carbon/Ag memristors with the reverse connection structure, which are connected to a common transistor. The word line (WL) and source line (SL) are connected to the transistor gate and transistor source, respectively. The right bit line (BL) and left BL are connected to the top electrode of the right memristor and the left memristor, respectively.

The basic operations of the 1T2M crossbar array are consisted of the crossbar array programming and the MAC operation, as shown in Fig. 6.

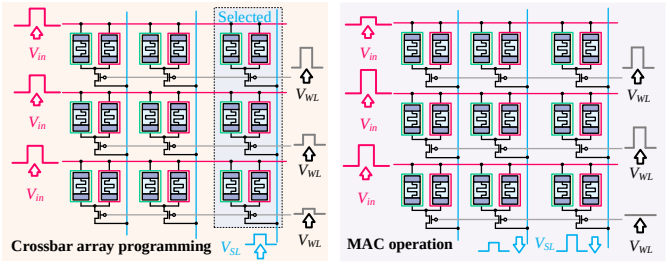


Fig. 6. The basic operations of the 1T2M crossbar array. (a) Crossbar array programming; (b) MAC operation.

During the crossbar array programming, the 1T2M crossbar array is programmed column-by-column using one-shot blind update method [31], as shown in Fig. 6(a). When the selected column is programmed, all SLs are floating, except the selected SL. Each WL is assigned a different voltage based on the targeted conductance. All 1T2M cells have the same input voltage. During the MAC operation, the input voltage $V_{in,j}$ is applied across the BLs, and the corresponding output voltage falling on the memristor is read out through the SLs, as shown in Fig. 6(b). According to the Kirchhoff's law and Ohm's law, the output voltage V can be mathematically expressed by:

$$V = R_f \sum_{i,j}^{row,col} \left(G_{R,i} - G_{L,i}\right) V_{in,j} \tag{4}$$

where $G_{R,i}$ and $G_{L,i}$ are the conductance of the right memristor and the left memristor, respectively. R_f is the feedback resistor in the 1T2M crossbar array. *col* and *row* denote the number of columns and rows in the 1T2M crossbar array, respectively.

The positive, zero, and negative weights can be achieved, as shown in Fig. 7.

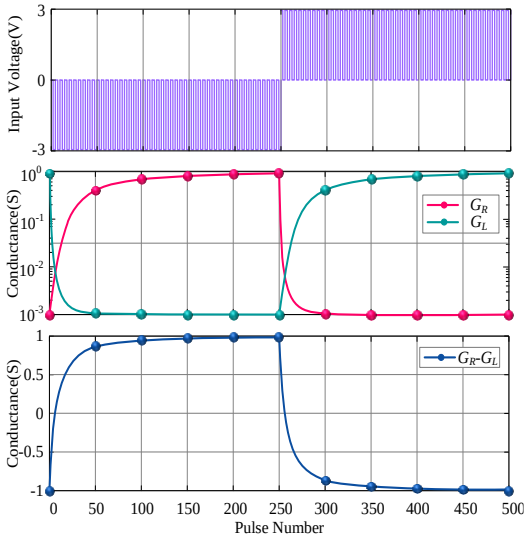


Fig. 7. Circuit results of weight modulation.

Specifically, when the negative input voltage is applied, the conductance of the right memristor $G_{R,i}$ increases and the conductance of the right memristor $G_{L,i}$ decreases, the corresponding weight ($G_{R,i}$- $G_{L,i}$) changes from negative to positive. Conversely, when the positive input voltage is applied, the conductance of the right memristor $G_{R,i}$ decreases and the conductance of the right memristor $G_{L,i}$ increases, the corresponding weight ($G_{R,i}$- $G_{L,i}$) changes form positive to negative. Therefore, the desired weight can be effectively obtained through the weight modulation operation.

IV. Circuit Design of Human Activity Recognition In-Memory Computing System

In-memory computing is the potential candidate to break von Neumann bottleneck and provide a new way towards artificial general intelligence (AGI) [15]. Our motivation is to design a novel in-memory computing system for human activity recognition in smart healthcare applications.

A. Circuit Design of Self-Attention ConvLSTM Module

In this work, the SA-ConvLSTMM is proposed to capture high-level spatial and temporal feature $\hat{V}_H(t)$ from the time-series inputs $V_t(t)$, which consists of two units: the ConvLSTM unit and self-attention memory unit.

(1) Circuit design of ConvLSTM unit

ConvLSTM network has been proposed in [22], through replacing all linear multiplication operations in long short-term memory (LSTM) network [22] with convolution operation with parameter sharing and sparse connectivity. The circuit design of ConvLSTM unit is proposed in this work to capture spatial and temporal dependencies in the time-series inputs $V_t(t)$, as shown in Fig. 8.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648,

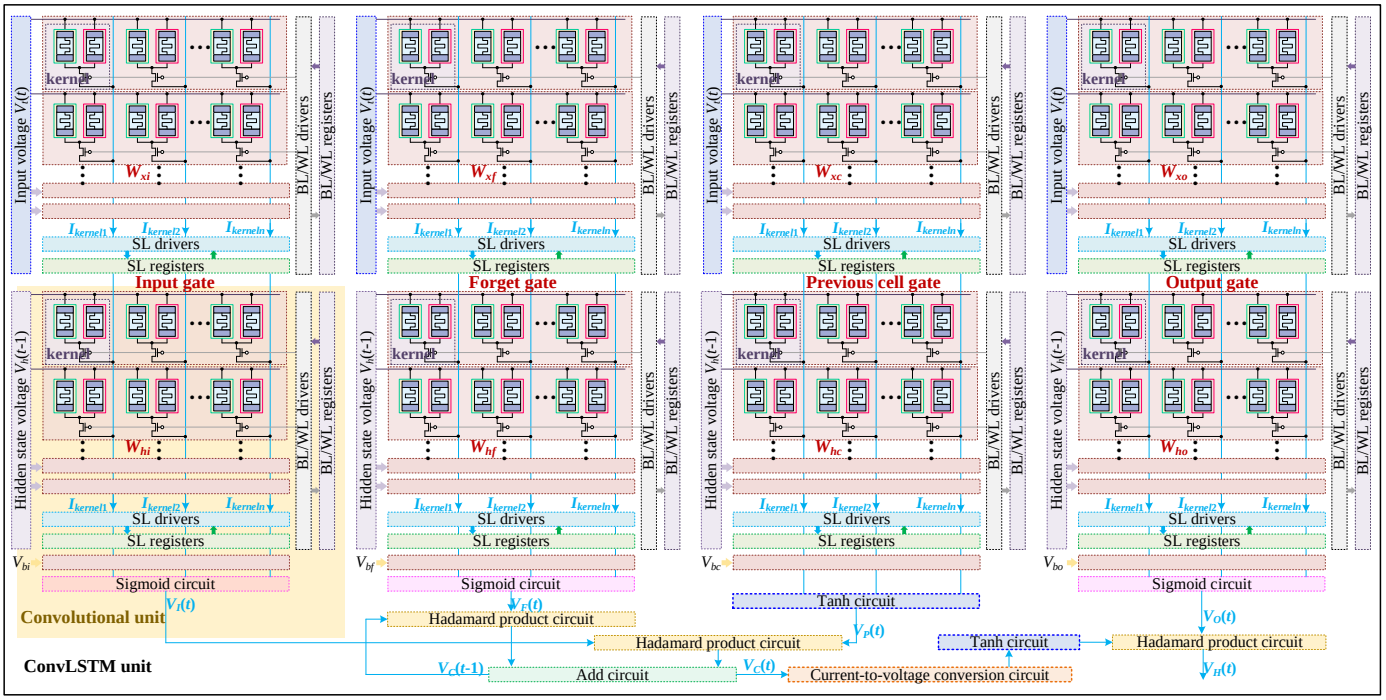


Fig. 8. Circuit design of ConvLSTM unit.

From Fig. 8, the proposed ConvLSTM unit is mainly composed of the 1T2M crossbar arrays, activation circuits (i.e., sigmoid circuit and tanh circuit), Hadamard product circuit, add circuit, and current-to-voltage conversion circuit. Specifically, the 1T2M crossbar array is used to perform convolutional operation, the convolutional kernel weight is encoded to the conductance difference between one pair of memristors. The output currents $I_{knernel1}$, $I_{knernel12}$, …, $I_{knerneln}$ collected for kernels are injected into sigmoid circuit and tanh circuit, respectively. Then, the output voltages of input gate $V_I(t)$, forget gate $V_F(t)$, previous cell gate $V_P(t)$, and output gate $V_O(t)$ can be obtained. Finally, the output voltages of cell gate $V_C(t)$ and ConvLSTM unit $V_H(t)$ can be produced after several operations (i.e., Hadamard dot product, add, and activation). Notably, the above-mentioned sub-circuits have been designed in our previous work [17]. Thus, the input and output of the ConvLSTM unit can be mathematically expressed by:

$$V_I(t)=\sigma\left(\begin{array}{l}\sum_{n=1}^{N}\left(G_{xi,n}^{R}-G_{xi,n}^{L}\right)*V_t(t)+\\ \sum_{m=1}^{M}\left(G_{hi,n}^{R}-G_{hi,n}^{L}\right)*V_h(t-1)+I_{bi}\end{array}\right) \tag{5}$$

$$V_F(t)=\sigma\left(\begin{array}{l}\sum_{n=1}^{N}\left(G_{xf,n}^{R}-G_{xf,n}^{L}\right)*V_t(t)+\\ \sum_{m=1}^{M}\left(G_{hf,n}^{R}-G_{hf,n}^{L}\right)*V_h(t-1)+I_{bf}\end{array}\right) \tag{6}$$

$$V_P(t)=\tanh\left(\begin{array}{l}\sum_{n=1}^{N}\left(G_{xc,n}^{R}-G_{xc,n}^{L}\right)*V_t(t)+\\ \sum_{m=1}^{M}\left(G_{hc,n}^{R}-G_{hc,n}^{L}\right)*V_h(t-1)+I_{bc}\end{array}\right) \tag{7}$$

$$V_O(t)=\sigma\left(\begin{array}{l}\sum_{n=1}^{N}\left(G_{xo,n}^{R}-G_{xo,n}^{L}\right)*V_t(t)+\\ \sum_{m=1}^{M}\left(G_{ho,n}^{R}-G_{ho,n}^{L}\right)*V_h(t-1)+I_{bo}\end{array}\right) \tag{8}$$

$$V_C(t)=V_F(t)\cdot V_C(t-1)+V_I(t)\cdot V_P(t) \tag{9}$$

$$V_H(t)=V_O(t)\cdot\tanh\left(V_C(t)\right) \tag{10}$$

where symbol σ, tanh, *, and · denote the sigmoid activation, tanh activation, convolutional, and Hadamard dot product operations, respectively. N and M denote the row of the 1T2M crossbar arrays. W_{xi}, W_{hi}, W_{xf}, W_{hf}, W_{xc}, W_{hc}, W_{xo}, and W_{ho} are the weight matrixes of the input gate, forget gate, previous cell gate, and output gate, respectively. G_{xi}, G_{hi}, G_{xf}, G_{hf}, G_{xc}, G_{hc}, G_{xo}, and G_{ho} are the conductance of memristor in the 1T2M crossbar arrays. I_{bi}, I_{bf}, I_{bc}, and I_{bo} are the bias current of the input gate, forget gate, previous cell gate, and output gate, respectively.

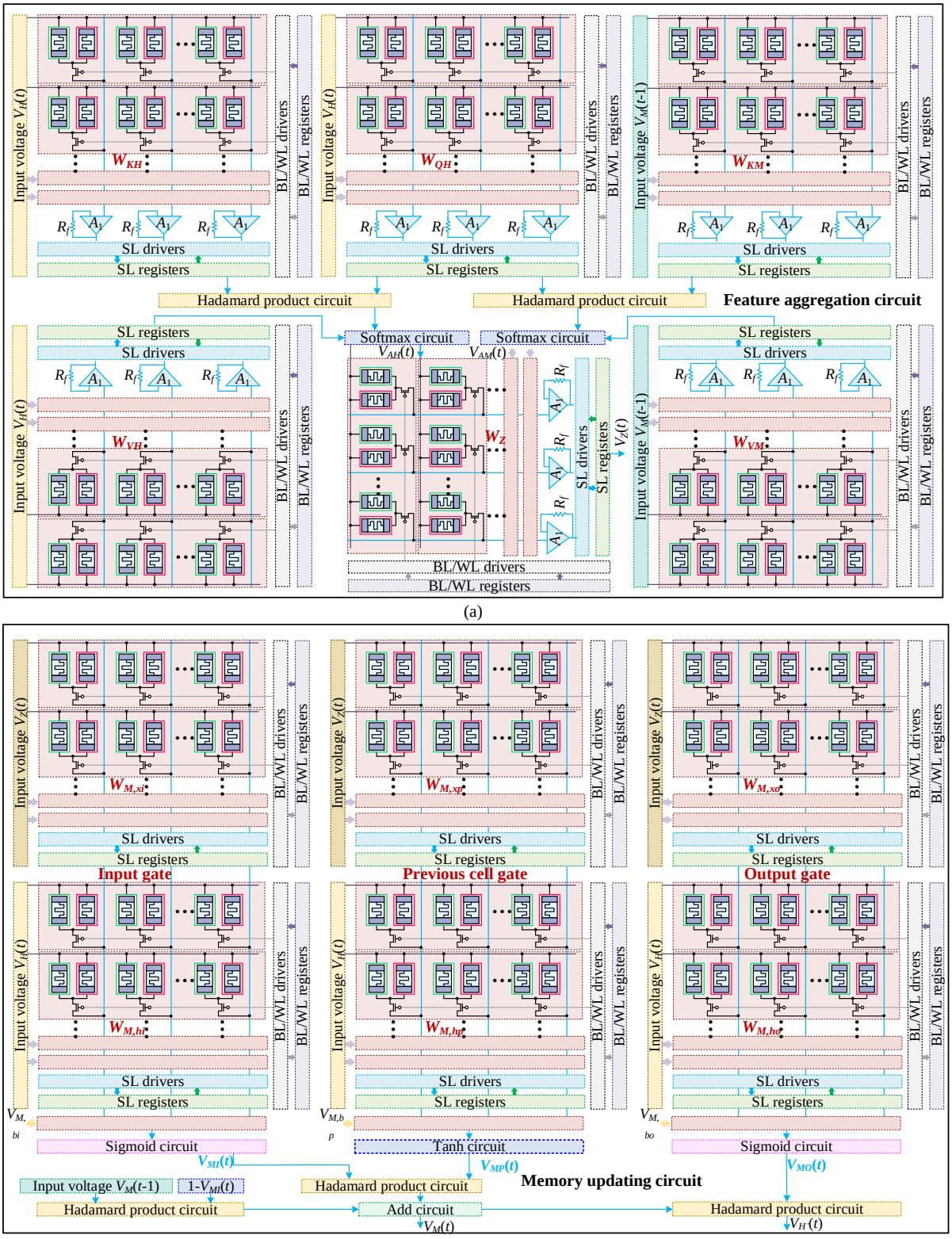
(2) Circuit design of self-attention memory unit

The self-attention memory unit is designed to capture long-range spatial and temporal dependencies from the feature voltage $V_H(t)$ generated by ConvLSTM unit. The proposed self-attention memory unit is mainly composed by the feature aggregation circuit and memory updating circuit.

The self-attention mechanism is used to feature aggregation, and the specific circuit architecture of feature aggregation circuit is shown in Fig. 9 (a).

In Fig. 9 (a), the feature aggregation circuit receives two input voltages, in which the feature voltage $V_H(t)$ belongs to time step t, and the memory voltage $V_M(t\text{-}1)$ belongs to time step t-1. The 1T2M crossbar arrays are employed to store and compute the weight matrixes $\boldsymbol{W}$ (W_{KH}, W_{QH}, W_{VH}, W_{KM}, and W_{VM}) in the feature aggregation circuit. Following the weight matrixes $\boldsymbol{W_H}$ (W_{KH}, W_{QH}, and W_{VH}), the feature voltage $V_H(t)$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648,



(a)

(b)

Fig. 9. Circuit design of self-attention memory unit. (a) feature aggregation circuit; (b) memory updating circuit.

can be transferred into current vectors, representing the attention key K_H, the attention query Q_H, and the attention value V_H, respectively. Following the weight matrixes $\boldsymbol{W_M}$ (W_{KM}, W_{VM}), the memory voltage $V_M(t\text{-}1)$ can be converted to the attention key K_M, and the attention value V_M, respectively. The output voltages of feature aggregation circuit can be obtained by applying softmax operation, which can be described as:

$$V_{AH} = soft\max\left(\frac{\left(W_{KH}\cdot V_H\right)^T\cdot\left(W_{QH}\cdot V_H\right)^T}{\sqrt{d}}\right)\cdot\left(W_{VH}\cdot V_H\right)^T \quad (11)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648,

$$V_{AM} = soft\max\left(\frac{\left(W_{KM}\cdot V_M\right)^T\cdot\left(W_{QH}\cdot V_H\right)^T}{\sqrt{d}}\right)\cdot\left(W_{VM}\cdot V_M\right)^T \quad (12)$$

$$V_Z = W_z\left[V_{AH}, V_{AM}\right] \quad (13)$$

where V_{AH} is the output voltage aggregated by querying on the feature voltage $V_H(t)$ at time step t, V_{AM} is the output voltage aggregated by querying on the memory voltage $V_M(t\text{-}1)$ at time step t-1. d and T are the dimension of the feature aggregation circuit and transpose operation, respectively. V_Z is the output voltage of feature aggregation circuit, which is the fusion of V_{AH} and V_{AM}. W_z is the weight matrix realized by the 1T2M crossbar array.

Then, the gating mechanism is employed to update the memory voltage V_M, and generate the finial output. The specific circuit architecture of memory updating circuit is shown in Fig. 9(b). From Fig. 9(b), the aggregated feature voltage $V_Z(t)$ and the input feature voltage $V_H(t)$ are utilized to generate the output voltages of input gate $V_{MI}(t)$, previous cell gate $V_{MP}(t)$, and output gate $V_{MO}(t)$. To reduce parameters, the output voltage of the forget gate is replaced as (1- $V_{MI}(t)$). The specific memory updating progress and final output of the self-attention memory unit can be formulated as:

$$V_{MI}(t) = \sigma\left(\begin{array}{l}\sum_{n=1}^{N}\left(G^R_{M,xi,n} - G^L_{M,xi,n}\right)\cdot V_Z(t) + \\ \sum_{m=1}^{M}\left(G^R_{M,hi,n} - G^L_{M,hi,n}\right)\cdot V_H(t) + I_{M,bi}\end{array}\right) \quad (14)$$

$$V_{MP}(t) = \tanh\left(\begin{array}{l}\sum_{n=1}^{N}\left(G^R_{M,xp,n} - G^L_{M,xp,n}\right)\cdot V_Z(t) + \\ \sum_{m=1}^{M}\left(G^R_{M,hp,n} - G^L_{M,hp,n}\right)\cdot V_H(t) + I_{M,bp}\end{array}\right) \quad (15)$$

$$V_{MO}(t) = \sigma\left(\begin{array}{l}\sum_{n=1}^{N}\left(G^R_{M,xo,n} - G^L_{M,xo,n}\right)\cdot V_Z(t) + \\ \sum_{m=1}^{M}\left(G^R_{M,ho,n} - G^L_{M,ho,n}\right)\cdot V_H(t) + I_{M,bo}\end{array}\right) \quad (16)$$

$$V_M(t) = \left(1 - V_{MI}(t)\right)\cdot V_M(t-1) + V_{MI}(t)\cdot V_{MP}(t) \quad (17)$$

$$V_{H'}(t) = V_{MO}(t)\cdot V_M(t) \quad (18)$$

where $\boldsymbol{W}$ ($W_{M,xi}$, $W_{M,hi}$, $W_{M,xp}$, $W_{M,hp}$, $W_{M,xo}$, and $W_{M,ho}$) are the weight matrixes realized by the 1T2M crossbar array. $G_{M,xi}$, $G_{M,hi}$, $G_{M,xp}$, $G_{M,hp}$, $G_{M,xo}$, and $G_{M,ho}$ are the conductance of memristor in the 1T2M crossbar arrays. $I_{M,bi}$, $I_{M,bp}$, and $I_{M,bo}$ are the bias current of the input gate, previous cell gate, and output gate, respectively.

B. *Circuit Design of Multi-Head Convolutional Attention Module*

In this work, the MHCAM is proposed to extract channel and spatial feature $V_{MC}(t)$ from the image inputs $V_i(t)$. The convolutional layer is used to produce the feature map $V_C(t)$ through convolutional operation, which is mainly consisted of convolutional filter, max-pooling operation, and dropout operation. In the MHCAM, each head consists two channel-spatial attention unit with the cascaded configuration that can enhance the channel and spatial features. Considering there are existing circuit design of the convolutional unit, max-pooling unit, and dropout unit, this work mainly focuses on the investigation of the circuit design of channel attention unit and spatial attention unit.

(1) Circuit design of channel attention unit

The channel attention unit is employed to measure the significance of each channel information, and the specific circuit architecture of channel attention unit is shown in Fig. 10.

From Fig. 10, the proposed channel attention unit is mainly composed of the 1T2M crossbar arrays, average-pooling circuit, max-pooling circuit, ReLU circuit, sigmoid circuit, add circuit, and current-to-voltage conversion circuit. Specifically, the input voltage $V_C(t)$ generated by the convolutional unit is injected to the average-pooling circuit and max-pooling circuit, respectively, generating two different spatial context representations. Then, the output voltages of average-pooling circuit and max-pooling circuit are further entered to shared network. Notably, the shared network is mainly composed by two 1T2M crossbar arrays with a ReLU circuit in the between. The 1T2M crossbar array is mainly used to calculate and store the learnable parameters W_A and W_B. The output voltage $V_{AC}(t)$ of channel attention unit can be obtained by applying element-wise summation operation, which can be described as:

$$V_{AC}(t) = \sigma\left(\begin{array}{l}W_B\left(W_A\cdot AvgPool\left(V_C\right)\right) + \\ W_B\left(W_A\cdot MaxPool\left(V_C\right)\right)\end{array}\right) \quad (19)$$

where $AvgPool(\cdot)$ and $MaxPool(\cdot)$ denote average-pooling and max-pooling operations, respectively.

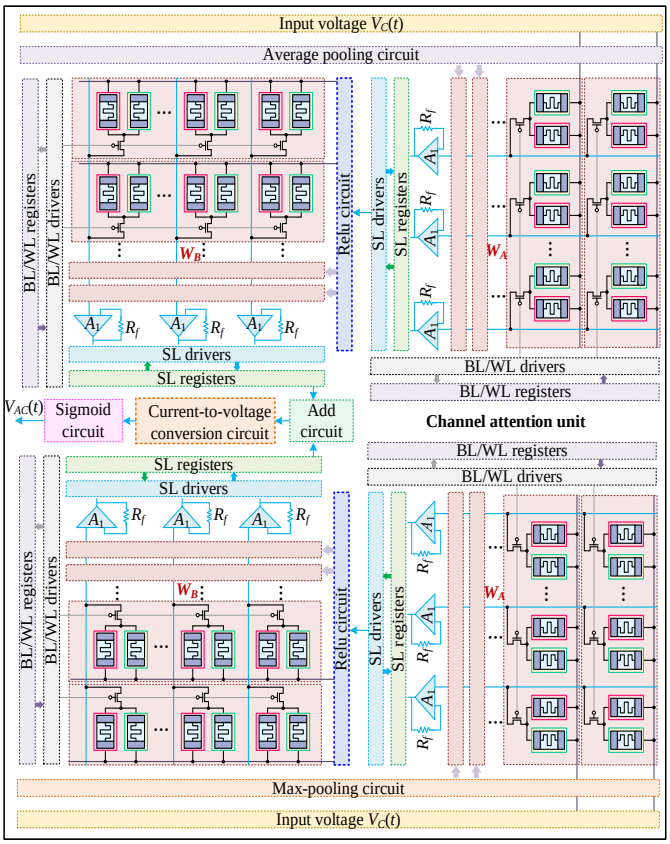


Fig. 10. Circuit design of channel attention unit.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648,

(2) Circuit design of spatial attention unit

The spatial attention unit is complementary the channel attention unit, which focuses on the spatial information processing, as shown in Fig. 11.

From Fig. 11, the proposed spatial attention unit is mainly composed by the 1T2M crossbar array, average-pooling circuit, max-pooling circuit, and sigmoid circuit. Firstly, the efficient features are generated by performing average-pooling and max-pooling operations on the input voltage (V_{AC}·V_C). Then, the efficient features are concatenated and convolved by the convolutional unit to generate the output voltage $V_{AS}(t)$ of spatial attention unit. Notably, the convolutional unit is implemented by the 1T2M crossbar array. Thus, the input and output of the spatial attention unit can be mathematically expressed by:

$$V_{AS}(t)=\sigma\left(conv\left[AvgPool\left(V_{AC}\cdot V_C\right),MaxPool\left(V_{AC}\cdot V_C\right)\right]\right) \quad (20)$$

where $conv(\cdot)$ denotes convolutional operation, $[\cdot]$ is concatenated operation.

On this basis, the overall channel-spatial attention mechanism can be given as:

$$V_{CS}(t)=V_{AS}(t)\cdot\left(V_{AC}(t)\cdot V_C(t)\right) \quad (21)$$

where $V_{CS}(t)$ denotes the output voltage of the channel-spatial attention unit.

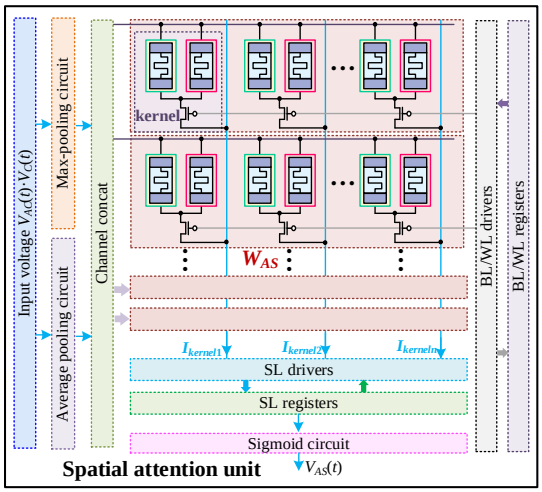


Fig. 11. Circuit design of spatial attention unit.

C. Circuit Design of Recognition Module

Considering the semantic gap between the temporal feature $V_H(t)$ and image feature $V_{MC}(t)$, the crossmodal attention mechanism is employed in the recognition module to effectively capture cross-modal interaction and generate final output for human activity recognition. The specific circuit scheme of recognition module is shown in Fig. 12.

From Fig. 12, the proposed recognition module consists of the 1T2M crossbar arrays, Hadamard product circuit, layer normalization circuit, ReLU circuit, softmax circuit, and fully connected circuit. The weight matrixes $\boldsymbol{W}$ (W_K, W_Q, W_V, W_a, and W_b) are learnable parameters, which are implemented by the 1T2M crossbar arrays. Following weight matrixes $\boldsymbol{W}$ (W_K, W_Q, and W_V), the attention key is defined as $V_K(t)=W_K\cdot V_{MCA}(t)$, the attention query is defined as $V_Q(t)=W_Q\cdot V_H(t)$, and the attention value is defined as $V_V(t)=W_V\cdot V_{MCA}(t)$. The output voltage $V_{CMA}(t)$ of crossmodal attention circuit can be obtained by applying softmax operation, which can be mathematically expressed by:

$$V_{CMA}=soft\max\left(\frac{\left(W_K\cdot V_{MCA}\right)^T\cdot\left(W_Q\cdot V_H\right)^T}{\sqrt{d}}\right)\cdot\left(W_V\cdot V_{MCA}\right)^T \quad (22)$$

Then, the feed-forward circuit is used to process the output voltage $V_{CMA}(t)$ of crossmodal attention circuit. The output voltage $V_F(t)$ of feed-forward circuit is symbolized by:

$$V_F=LN\left(LN\left(V_{MCA},V_{CMA}\right),\mathrm{Re}LU\left(W_a\cdot LN\left(V_{MCA},V_{CMA}\right)\right)\cdot W_b\right) \quad (23)$$

where $LN(\cdot)$ and $ReLU(\cdot)$ denote the layer normalization and ReLU activation operations, respectively.

Finally, the output voltage $V_F(t)$ of feed-forward circuit is applied to the fully connect circuit and softmax circuit in sequence for human activity recognition. The output $V_{out}(t)$ of the recognition module is mathematically expressed by:

$$V_{out}(t)=soft\max\left(\sum V_F(t)\right) \quad (24)$$

V. Application in Human Activity Recognition

In this section, the effectiveness and validity of the proposed human activity recognition in-memory computing system are demonstrated by a series of experiments (including quantitative results and analysis, ablation study, computational efficiency analysis, and robustness analysis). The electrical parameters and neural network parameters used for the proposed human activity recognition in-memory computing system are provided in revised Table I.

TABLE I
LIST OF THE PARAMETERS USED FOR PROPOSED SYSTEM

	Device	Parameter	
Electrical parameters	1T2M memristor crossbar array	R_{on}	~$10^1\Omega$
		R_{off}	~1kΩ
		V_{read}	0.5V
		V_{scan}	0.5V/s
		V_{SET}	2.0V
		V_{RESET}	1.5V
		$[V_{min},V_{max}]$	[0V, 3V]
	Transistor	Gate width/length ratio	4.3
		Gate voltage	1.1V
		Access resistance	15 KΩ
	ADC	Precision	6 bits
Neural network parameters		Learning rate	10^{-2}
		Momentum	0
		Decay	0.9
		Maximum error	10^{-4}

From Table I, the electrical parameters mainly rely on the fabricated Ag/a-Carbon/Ag memristor and selected devices. Inspired by [12-14], the neural network parameters can be roughly determined, which can greatly simplify the process of parameter setting.

A. Dataset Description

The large-scale and publicly available UP-Fall dataset collected by three types devices (i.e., cameras, wearable sensors, and context-aware devices) is employed for experimental evaluation. The data collected from those three types devices are in the shape of time-series and image

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648,

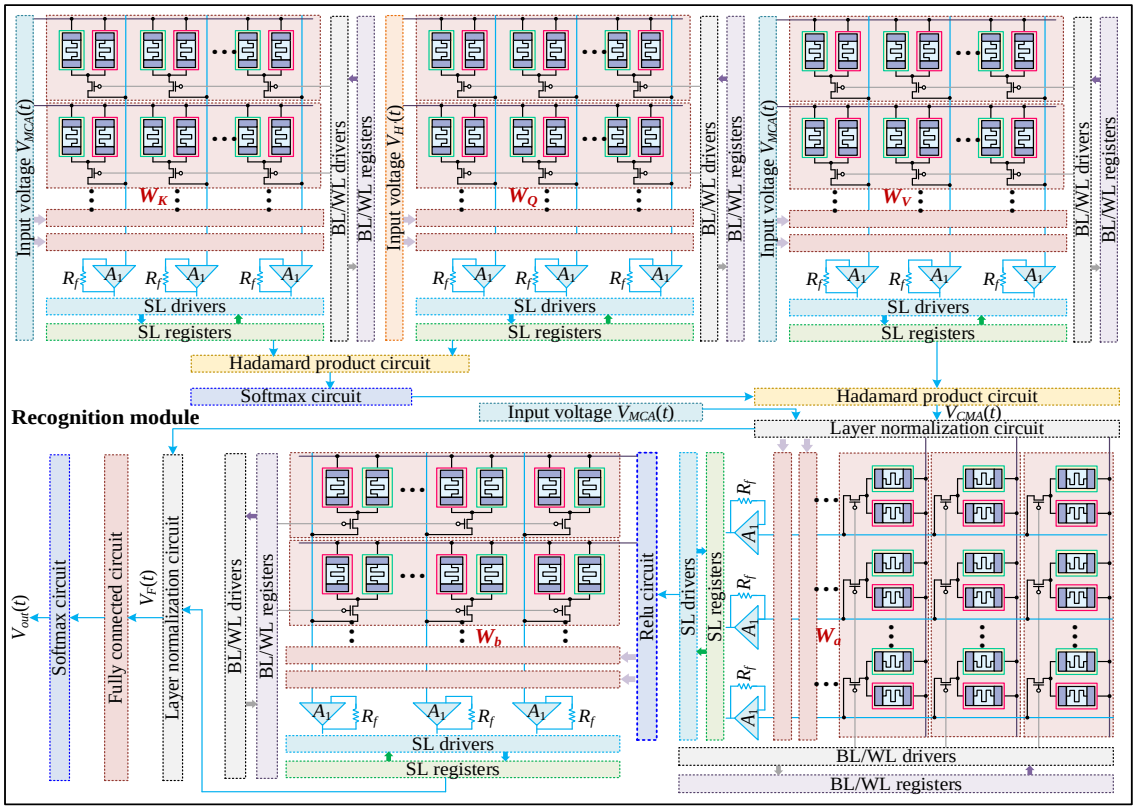


Fig. 12. Circuit design of classification module.

respectively. The UP-Fall dataset was published by Martinez et al. [12] in 2019, which contains eleven different human activities, including five different postural falls, and six basic daily activities. All activities are repeatedly performed three trials by seventeen healthy volunteers aging from 18 to 24, with a mean height and mean weight of 1.66m and 66.8kg, respectively. In this work, the data from trial 1 and trail 2 are used as training dataset, and the data from trail 3 is used as testing dataset. Specifically, the description and the time duration of each activity in the UP-Fall dataset is recorded in Table II.

TABLE II
STATISTICS FOR THE THREE PUBLIC STANDARD DATASETS

Activity ID	Description	Duration (s)
ID 1	Falling forward using hands	10
ID 2	Falling forward using knees	10
ID 3	Falling backward	10
ID 4	Falling sideways	10
ID 5	Falling sitting in empty chair	10
ID 6	Walking	60
ID 7	Standing	60
ID 8	Sitting	60
ID 9	Picking up an object	10
ID 10	Jumping	30
ID 11	Laying	60

To evaluate the influence of multimodal data on human activity recognition, different data sources are used in the proposed system. Specifically, the data combinations can be divided into seven categories: 1) C1: camera 1 only; 2) C2: camera 2 only; 3) TS: time-series only; 4) C1+TS: fusion data of camera 1 and time-series; 5) C2+TS: fusion data of camera 2 and time-series; 6) C1+C2: fusion data of camera 1 and camera 2; 7) C1+C2+TS: fusion data of camera 1, camera 2 and time-series.

B. *Neural Network Training and Inference*

We trained the proposed human activity recognition in-memory computing system to recognize eleven different human activities, which includes ex-situ training, crossbar array programming, and in-situ inference.

Ex-situ training: the ex-situ training is implemented in PyTorch platform (Intel® CoreTM i7-12700KF CPU, graphics card (Nvidia RTX 3070 Ti), and Windows 10 operating system, Python version 3.7). Specifically, the image data in the training dataset are down sampled from 640×480 pixels to 64×64 pixels and then fed to the MHCAM. The time-series data with the total number of 42 are collected from wearable sensors and context-aware devices, which are injected to SA-ConvLSTMM. The desired weights can be obtained when the ex-situ training is completed.

Crossbar array programming: To modulate the weights, the weight difference is mapped to the conductance difference of 1T2M cell. Then, the one-shot blind-update method [30] is used to update the conductance by tuning the gate of the corresponding 1T2M cell in each crossbar array.

In-situ inference: when the 1T2M crossbar arrays have been programmed, the proposed human activity recognition in-memory computing system with the testing dataset is validated using in-situ inference. Specifically, the multimodal dates are converted to voltage inputs by digital-to-analog converter. Then, the input voltages are injected to the BLs of each crossbar arrays. Meanwhile, the recognition module will output a set of voltages with eleven states assigned to eleven

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648,

different human activities. The recognition results are determined by the largest output voltage in each period, and the specific in-situ inference results are illustrated in Fig. 13.

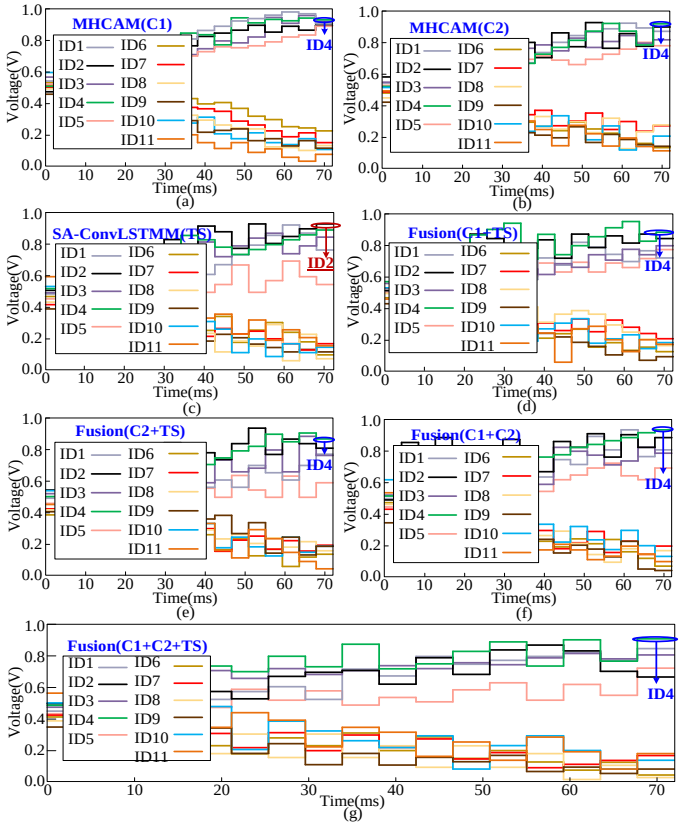


Fig. 13. The in-situ inference results. (a) MHCAM (C1); (b) MHCAM (C2); (c) SA-ConvLSTM (TS); (d) Fusion (C1+TS); (e) Fusion (C2+TS); (f) Fusion (C1+C2); (g) Fusion (C1+C2+TS)

C. Quantitative Results and Analysis

To evaluate the overall performance of the proposed system, the experiment for 100 epochs is conducted. The relationship between accuracy and the number of epochs during training phase is illustrated in Fig. 14(a). The experimental result demonstrates that the proposed system containing camera 1, camera 2 and time-series data sources achieves best performance on human activity recognition. The relationship between loss and the number of epochs during training phase is illustrated in Fig. 14(b). It is can be observed from experimental result that the SA-ConvLSTMM takes more time to generalize since it only contains time-series data. Similar results can also be observed in the accuracy and loss curves during testing phase, as shown in Fig. 14 (c) and Fig. 14 (d).

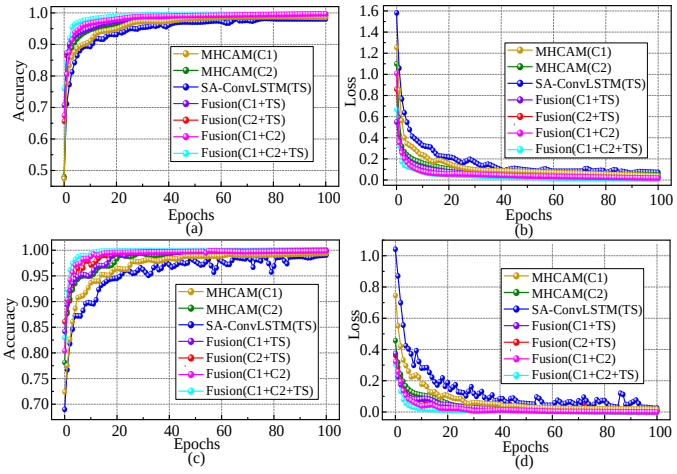


Fig. 14. Performance measures. (a) Accuracy during training process; (b) Loss during training process; (c) Accuracy during testing process; (b) Loss during testing process.

Two widely used common metrics, i.e., accuracy, and F1-score [14] are record for each human activity on the UP-Fall dataset, as illustrated in Table III.

Table III demonstrates that the proposed system achieves close to 100% recognition performance of the most human activities (i.e., ID 6 to ID 11) for fusing the camera 1, camera 2, and time-series data. The lowest recognition performance (53.6% Acc., 55.9 F1) is obtained for the activity ID 3 from SA-ConvLSTMM (TS). The results demonstrate that when fusion data of camera 1, camera 2 and time-series is used as input to the proposed system, the system performance outperform those of the single modality and the bi-modalities.

Furthermore, the proposed human activity recognition in-memory computing system is compared with the state-of-the-art (SOTA) methods on the UP-Fall dataset, as shown in Table IV.

TABLE IV
COMPARISON OF DIFFERENT STATE-OF-THE-ART METHODS ON UP-FALL DATASET

Ref.	Data	Data source	Acc.	F1
[7]	Time-series	Wearable sensors	/	67.0
[8]	Image	Camera 1	95.1	71.2
[9]	Image	Camera 1	96.0	93.0
[10]	Image	Camera 2	96.7	97.4
[11]	Image	Camera 1, 2	95.6	97.4
[12]	Time-series + Image	Wearable sensors + Camera 1	96.4	82.3
[13]	Time-series + Image	Wearable sensors + Camera 1, 2	99.8	98.9
[14]	Time-series + Image	Wearable sensors + Camera 1, 2	97.9	97.9
This work	Time-series + Image	Wearable sensors + Camera 1, 2	100.0	100.0

From Table IV, the proposed system using fusion data of camera 1, camera 2 and time-series obtains the best accuracy and F1-score on the UP-Fall dataset. The experimental results

TABLE III
PERFORMANCE OF THE PROPOSED SYSTEM FOR EACH ACTIVITY RECOGNITION

Activity	MHCAM(C1)		MHCAM(C2)		SA-ConvLSTMM(TS)		Fusion(C1+TS)		Fusion(C2+TS)		Fusion(C1+C2)		Fusion(C1+C2+TS)	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
ID1	90.7	87.2	93.3	89.2	82.7	75.6	94.1	89.7	94.8	91.5	96.2	93.5	97.1	94.3
ID2	76.2	79.2	77.3	81.1	69.3	68.9	78.8	82.5	80.4	83.9	82.3	85.2	84.2	87.5
ID3	84.3	82.4	86.7	86.1	53.6	55.9	85.8	85.3	88.3	87.4	90.1	88.7	93.5	91.5
ID4	80.6	81.6	83.5	85.1	76.0	72.2	84.2	84.0	85.8	85.9	86.3	86.5	90.2	88.7
ID5	81.8	85.3	84.0	88.7	60.7	67.7	84.9	88.7	86.1	89.5	87.8	89.5	88.5	91.7
ID6	98.7	96.1	100.0	97.4	98.2	90.1	100.0	96.8	100.0	95.5	100.0	95.5	100.0	98.7
ID7	98.9	98.7	100.0	99.3	93.4	95.3	100.0	98.7	100.0	98.7	100.0	99.3	100.0	100.0
ID8	99.4	98.0	100.0	98.7	96.3	90.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
ID9	94.7	94.6	97.3	97.8	69.3	78.8	99.4	99.3	98.7	99.3	100.0	100.0	100.0	100.0
ID10	93.4	94.6	96.0	97.3	90.2	93.1	98.6	98.7	99.2	98.7	100.0	100.0	100.0	100.0
ID11	98.6	95.8	100.0	96.6	97.1	92.5	100.0	97.9	100.0	97.9	100.0	97.9	100.0	100.0

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/JBHI.2024.3392648,

are concluded below: 1) Compared to single modality approaches [7-11], the reliability and accuracy of the proposed system are improved by using multimodal information; 2) the SA-ConvLSTMM can capture high-level spatial and temporal feature form time-series more effectively compared with LSTM network used in [12], Gramian angular field (GAF) method used in [13], and ConvLSTM network used in [14]; 3) the MHCAM can extract the relevant patterns from the channel and spatial dimensions compared with CNN network used in [12, 13]; 4) Different from the direct fusion strategy used in [14], the cross-modal transformer unit is adopted in the proposed system, which can realize the cross-modal interactions more sufficiently.

D. Computational Efficiency Analysis

In this work, the computational efficiency of the developed system for human activity recognition in smart home environment are measured in terms of time consumption, energy consumption, and latency.

Fig. 15 presents the time consumption of the proposed system by comparing with SOTA methods on the UP-Fall dataset. On average, the time consumption of the proposed system is about 1.08ms, which is approximately 8~10 times faster than other competitors, satisfying the real-time requirement of smart healthcare applications.

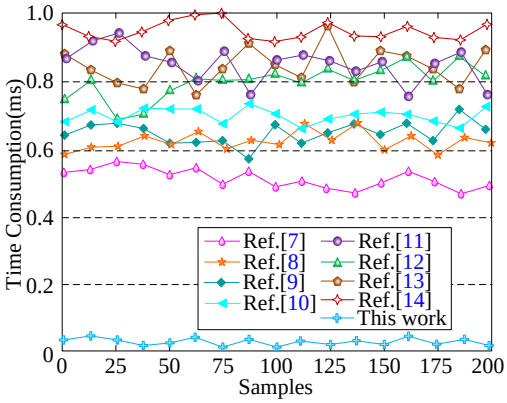


Fig. 15. The time consumption of in-situ inference.

The energy consumption of the in-memory computing system mainly depends on the input/output size, the weight precision, the crossbar array size, and the design scheme. To estimate the inference energy consumption of the proposed system, we rely on numbers obtained by existing in-memory computing platform [31], as shown in Fig. 16(a).

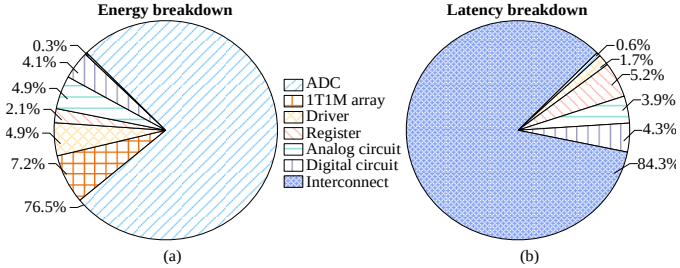


Fig. 16. The breakdown of proposed system. (a) energy breakdown; (b) latency breakdown.

We found the energy consumption is estimated about 2337pJ for 1-bit input/output, 1-bit weight with 0.5V, 50ns read voltage. The ADCs consume almost 76.5% of the total energy, which is the most of the energy consumption rather than the 1T2M array. Fig. 16(b) illustrates the latency breakdown of the proposed system. The latency of the proposed system is estimated about 1.82μs, and the interconnect module account almost 84.3% of the total latency.

E. Robustness Analysis

To evaluate the robustness of the proposed human activity recognition in-memory computing system, the anti-noise analysis and device failure analysis are carried out, as shown in Fig. 17. Firstly, the random noises are added to the multimodal inputs (i.e., time-series input and image inputs), and the accuracy and the F1-score of the proposed system on the UP-fall dataset are demonstrated in Fig. 17(a). In Fig. 17(a), when the standard deviation of random noise is over 0.2, the recognition accuracy and the F1-score can remain higher than 93% on the UP-fall dataset. The experimental results demonstrate that the impact of the random noise on the proposed system is negligible. Then, in order to examine the effect of the device failure, we set the failure ratio of the memristors in LRS/HRS with 0% to 30%, and the system performance on the UP-fall dataset is illustrated in Fig. 17(b). In Fig. 17(b), the accuracy loss and the F1-score loss of nearly 3% and 5% are observed in the proposed system, when the failure ratio of the memristors in LRS/HRS exceeds 20%. The experimental results demonstrate that the proposed system has a good tolerance to device failure.

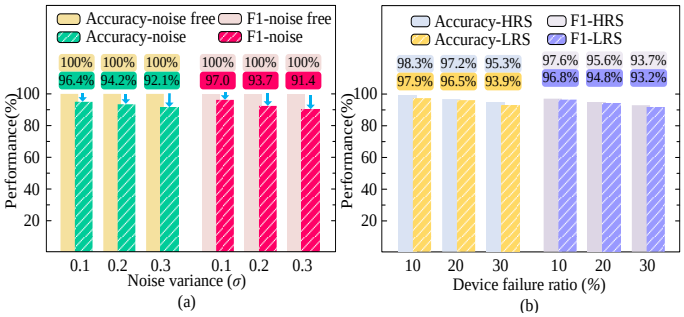


Fig. 17 The robustness analysis of proposed system. (a) anti-noise analysis; (b) device failure analysis.

VI. Discussion

Although proposed human activity recognition in-memory computing system has a capability to process multimodal information in smart healthcare applications, which balances recognition performance and computational efficiency to promote versatility. The scalability of the proposed system for real-world large-scale applications is still complex and challengeable.

At the device level: the perceptual information of different modalities can be simultaneously sensed and processed in human brain. An eco-friendly and intelligent in-sensor or near-sensor computing device should be developed for fusion of different perceptual information in a real-time manner, including tactile, auditory, olfactory, visual, and so on.

At the circuit level: the development of existing in-memory computing systems is limited to specific scenarios, and the versatility requirements of real-world large-scale applications are hard to achieve. Considering the requirements of real-world

large-scale application have become more diversified, reconfigurable functional circuits need to be further developed in the future.

At the algorithm level: the brain-inspired learning algorithms is still in its infancy stage. With better understanding of the structure and function of human brain, general online learning algorithms should be studied to promote the development of in-memory computing system for real-world large-scale applications.

VII. Conclusion

This paper investigates an efficient human activity recognition in-memory computing architecture for healthcare monitoring. Firstly, a mechanism-oriented model is constructed after fabrication of Ag/a-Carbon/Ag memristor. Then, 1T2M crossbar array is designed, which can realize the high-density connection and perform parallel computing. Furthermore, an efficient human activity recognition in-memory computing mainly consisted of SA-ConvLSTMM, MHCAM, and recognition module is designed. Through SA-ConvLSTMM, the high-level spatial and temporal features with local-global dependences from time-series data can be adequately captured. Through MHCAM, relevant patterns from the channel and spatial dimensions in visual data can extract sufficiently. Through recognition module, the crossmodal interaction can be exchanged and the reliable output can be obtained effectively. For verification, the proposed system is applied to perform human activity recognition on UP-Fall dataset, and the experimental results demonstrate that the proposed system outperforms SOTA methods in terms in recognition performance and time consumption (approximately 8∼10 times speed up). In addition, the necessary computational efficiency analysis and robustness analysis are carried out, indicating the high computational efficiency and reliability of proposed system in smart healthcare applications.

References

[1] G. Yang, L. Zhang, C. Bu, S. Wang, H. Wu, and A. Song, "FreqSense: Adaptive sampling rates for sensor-based human activity recognition under tunable computational budgets," *IEEE J. Biomed. Heal. Informatics.* vol. 27, no. 12, pp. 5791-5802, Dec. 2023.

[2] Y. Tang, L. Zhang, H. Wu, J. He, and A. Song, "Dual-branch interactive networks on multichannel time series for human activity recognition," *IEEE J. Biomed. Heal. Informatics.,* vol. 26, no. 10, pp. 5223-5234, Nov. 2022.

[3] D. Ahn, S. Kim, H. Hong, and B. C. Ko, "STAR-transformer: A spatio-temporal cross attention transformer for human action recognition," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, 2-7 Jan., 2023, pp. 3319-3328.

[4] N. Dai, I. M. Lei, Z. Li, Y. Li, P. Fang, and J. Zhong, "Recent advances in wearable electromechanical sensors—Moving towards machine learning-assisted wearable sensing systems," *Nano Energy,* vol. 105, pp. 108041, Jan. 2023.

[5] G. Lee, O. Hossain, S. Jamalzadegan, Y. Liu, H. Wang, A. C. Saville, T. Shymanovich, R. Paul, D. Rotenberg, A. E. Whitfield, J. B. Ristaino, Y. Zhu, and Q. Wei, "Abaxial leaf surface-mounted multimodal wearable sensor for continuous plant physiology monitoring," *Science Advances,* vol. 9, no. 15, pp. eade2232, Apr. 2023.

[6] X. Xiong, Y. Chen, Z. Wang, H. Liu, M. Le, C. Lin, G. Wu, L. Wang, X. Shi, Y. G. Jia, and Y. Zhao, "Polymerizable rotaxane hydrogels for three-dimensional printing fabrication of wearable sensors," *Nat. Commun.,* vol. 14, no. 1, pp. 1331, Dec. 2023.

[7] L. Martínez-Villaseñor, H. Ponce, and R. A. Espinosa-Loera, "Multimodal database for human activity recognition and fall detection," in *12th International Conference on Ubiquitous Computing and Ambient Intelligence*, Punta Cana, 4–7 Dec., 2018, pp. 1237-1237.

[8] L. Martínez-Villaseñor, H. Ponce, J. Brieva, E. Moya-Albor, J. Núñez-Martínez, and C. Peñafort-Asturiano, "Up-fall detection dataset: A multimodal approach," *Sensors,* vol. 19, no. 9, pp. 1988, May 2019.

[9] Y. M. Galvao, L. Portela, J. Ferreira, P. Barros, O. A. De Araujo Fagundes, and B. J. T. Fernandes, "A framework for anomaly identification applied on fall detection," *IEEE Access,* vol. 9, pp. 77264-77274, May 2021.

[10] S. K. Yadav, A. Luthra, K. Tiwari, H. M. Pandey, and S. A. Akbar, "ARFDNet: An efficient activity recognition & fall detection system using latent feature pooling," *Knowledge-Based Syst.,* vol. 239, pp. 107948, Mar. 2022.

[11] R. Espinosa, H. Ponce, S. Gutiérrez, L. Martínez-Villaseñor, J. Brieva, and E. Moya-Albor, "A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the UP-Fall detection dataset," *Comput. Biol. Med.,* vol. 115, pp. 103520, Dec. 2019.

[12] L. Martínez-Villaseñor, H. Ponce, and K. Perez-Daniel, "Deep learning for multimodal fall detection," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Bari, 6-9 Oct., 2019, pp. 3422-3429.

[13] P. Qi, D. Chiaro, and F. Piccialli, "FL-FD: Federated learning-based fall detection with multimodal data fusion," *Inf. Fusion.,* vol. 99, pp. 101890, Nov. 2023.

[14] M. M. Islam, S. Nooruddin, F. Karray, and G. Muhammad, "Multi-level feature fusion for multimodal human activity recognition in Internet of Healthcare Things," *Inf. Fusion,* vol. 94, pp. 17-31, Jun. 2023.

[15] X. Ji, Z. Dong, C. S. Lai, and D. Qi, "A brain-inspired in-memory computing system for neuronal communication via memristive circuits," *IEEE Commun. Mag.,* vol. 60, no. 1, pp. 100-106, Jan. 2022.

[16] X. Huang, C. Liu, Z. Tang, S. Zeng, S. Wang, and P. Zhou, "An ultrafast bipolar flash memory for self-activated in-memory computing," *Nat. Nanotechnol.,* vol. 18, no. 5, pp. 486-492, May 2023.

[17] Z. Dong, X. Ji, G. Zhou, M. Gao, and D. Qi, "Multimodal neuromorphic sensory-processing system with memristor circuits for smart home applications," *IEEE Trans. Ind. Appl.,* vol. 59, no. 1, pp. 47-58, Jan. 2023.

[18] A. Garofalo, G. Ottavi, F. Conti, G. Karunaratne, I. Boybat, L. Benini, and D. Rossi, "A heterogeneous in-memory computing cluster for flexible end-to-end inference of real-world deep neural networks," *IEEE J. Emerg. Sel. Top. Circuits Syst.,* vol. 12, no. 2, pp. 422-435, Jun. 2022.

[19] H. Ning, Z. Yu, Q. Zhang, H. Wen, B. Gao, Y. Mao, Y. Li, Y. Zhou, Y. Zhou, J. Chen, L. Liu, W. Wang, T. Li, Y. Li, W. Meng, W. Li, Y. Li, H. Qiu, Y. Shi, Y. Chai, H. Wu, and X. Wang, "An in-memory computing architecture based on a duplex two-dimensional material structure for in situ machine learning," *Nat. Nanotechnol.,* vol. 18, no. 5, pp. 493-500, May 2023.

[20] X. Ji, Z. Dong, Y. Han, C. S. Lai, and D. Qi, "A brain-inspired hierarchical interactive in-memory computing system and its application in video sentiment analysis," *IEEE Trans. Circuits Syst. Video Technol.* vol. 33, no. 12, pp. 7928 – 7942, Dec. 2023..

[21] Y. Guo, W. Duan, X. Liu, X. Wang, L. Wang, S. Duan, C. Ma, and H. Li, "Generative complex networks within a dynamic memristor with intrinsic variability," *Nat. Commun.,* vol. 14, no. 1, pp. 6134, Dec. 2023.

[22] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional ConvLSTM U-net with densley connected convolutions," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, 27-28 Oct. 2019, pp. 406-415.

[23] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International Cnference on Machine Learning*, Long Beach, CA, 9-15 Jun., 2019, pp. 7354-7363.

[24] Y.-H. H. Tsai, S. Bai, P. P. Liang, J. Z. Kolter, L.-P. Morency, and R. Salakhutdinov, "Multimodal transformer for unaligned multimodal language sequences," in *Proceedings of the conference. Association for Computational Linguistics. Meeting*, Florence, Jul., 2019, pp. 6558-6869.

[25] A. A. Grinberg, S. Luryi, M. R. Pinto, and N. L. Schryer, "Space-charge-limited current in a film," *IEEE Trans. Electron Devices,* vol. 36, no. 6, pp. 1162-1170, Jun. 1989.

[26] E. Rosencher, V. Mosser, and G. Vincent, "Transient-current study of field-assisted emission from shallow levels in silicon," *Phys. Rev. B,* vol. 29, no. 3, pp. 1135-1147, Feb. 1984.

[27] A. Ascoli, F. Corinto, V. Senger, and R. Tetzlaff, "Memristor model comparison," *IEEE Circuits Syst. Mag.,* vol. 13, no. 2, pp. 89-105, May 2013.

[28] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error

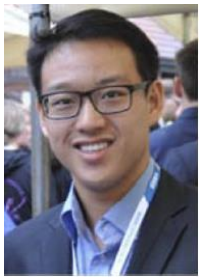
(MAE): When to use them or not," *Geosci. Model Dev.,* vol. 15, no. 14, pp. 5481-5487, Jul. 2022.
[29] S. Mirjalili, and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.,* vol. 95, pp. 51-67, May 2016.
[30] Z. Wang, C. Li, P. Lin, M. Rao, Y. Nie, W. Song, Q. Qiu, Y. Li, P. Yan, J. P. Strachan, N. Ge, N. McDonald, Q. Wu, M. Hu, H. Wu, R. S. Williams, Q. Xia, and J. J. Yang, "In situ training of feed-forward and recurrent convolutional memristor networks," *Nat. Mach. Intell.,* vol. 1, no. 9, pp. 434-442, Sep. 2019.
[31] M. J. Rasch, C. Mackin, M. Le Gallo, A. Chen, A. Fasoli, F. Odermatt, N. Li, S. R. Nandakumar, P. Narayanan, H. Tsai, G. W. Burr, A. Sebastian, and V. Narayanan, "Hardware-aware training for large-scale and diverse deep learning inference workloads using in-memory computing-based accelerators," *Nat. Commun.,* vol. 14, no. 1, pp. 5282, Dec. 2023.

Xiaoyue Ji (Member, IEEE) received the B.E. degree in electronics and information engineering from the Harbin Engineering University, China, in 2016, the M.S. degree in control science and engineering from the National University of Defense Technology, China, in 2019, and the Ph.D. degree in control science and engineering from the Zhejiang University, China, in 2023. She is currently working toward postdoctoral research at Tsinghua University, China. Her research interests cover memristor and memristive system, artificial neural network, neuromorphic computing.

Zhekang Dong (Senior Member, IEEE) received the B.E. and M.E. degrees in electronics and information engineering in 2012 and 2015, respectively, from Southwest University, Chongqing, China. He received the Ph.D. degree from the School of Electrical Engineering, Zhejiang University, China, in 2019. Currently, he is an associate professor in Hangzhou Dianzi University, Hangzhou, China. He is also a Research Assistant (Joint-Supervision) at The Hong Kong Polytechnic University. His research interests cover memristor and memristive system, artificial neural network, the design and analysis of nonlinear systems based on memristor and computer simulation.

Liyan Zhu received the B.E. degree in electronics and information engineering from Hangzhou Dianzi University, Hangzhou, China, in 2023, where he is currently pursuing the M.E. degree in electronic science and technology. His research interests cover artificial neural network and computer vision.

Chenhao Hu received the B.E. degree in electronic information science and technology from Wenzhou University, Wenzhou, China, in 2021. He is currently pursuing the M.E. degree in electronic science and technology with Hangzhou Dainzi University, Hangzhou, China. His research interests cover artificial neural network and computer vision.



Chun Sing Lai (Senior Member, IEEE) received the BEng in electronic and electrical engineering from Brunel University London, UK, and DPhil in engineering science from the University of Oxford, UK in 2013 and 2019, respectively. Dr Lai is currently a Lecturer at the Department of Electronic and Electrical Engineering, Brunel University London, UK. His current interests are in data analytics, power system optimization, energy system modelling, and energy economics for low carbon energy networks and energy storage systems.