A Deep Learning Based Lightweight Human Activity Recognition System Using Reconstructed WiFi CSI

Xingcan Chen, Yi Zou, Chenglin Li, and Wendong Xiao, *Senior Member, IEEE*

***Abstract*—Human activity recognition (HAR) is a key technology in the field of human–computer interaction. Unlike systems using sensors or special devices, the WiFi channel state information (CSI)-based HAR systems are noncontact and low cost, but they are limited by high computational complexity and poor cross-domain generalization performance. In order to address the above problems, a reconstructed WiFi CSI tensor and deep learning based lightweight HAR system (Wisor-DL) is proposed, which firstly reconstructs WiFi CSI signals with a sparse signal representation algorithm, and a CSI tensor construction and decomposition algorithm. Then, gated temporal convolutional network with residual connections is designed to enhance and fuse the features of the reconstructed WiFi CSI signals. Finally, dendrite network makes the final decision of activity instead of the traditional dense layer. Experimental results show that Wisor-DL is a lightweight HAR system with high recognition accuracy and satisfactory cross-domain generalization ability.**

***Index Terms*—Deep learning, human activity recognition (HAR), signal processing, WiFi channel state information (CSI).**

I. Introduction

With the development of human–computer interaction, more and more attention has been paid to human activity recognition (HAR). Most common systems of recognizing different human activities rely on wearable sensors [1], high-definition cameras [2] or special radio-frequency devices, such as ultrawideband (UWB) [3], radar [4], and WiFi [5].

The activity recognition systems based on wearable sensors will be inconvenient, because users need to wear corresponding devices all the time in the detection process, and it is difficult to achieve long time detection due to the limitation of sensor battery power [6]. High-definition cameras can capture the changes of various parts of human bodies so as to realize activity recognition, which has attracted wide attention. However, the high-definition cameras based systems are subject to the lighting conditions, i.e., the accuracy of activity recognition will be seriously reduced in the environment of severe exposure caused by excessive light or in the dark environment [7]. In addition, the systems based on high-definition cameras may also cause some privacy issues as it involves the shooting and storage of users' photos [8]. Radio-frequency based systems are becoming more popular due to human activity can be recognized in a noncontact manner without interference from lighting conditions. However, HAR systems based on UWBs and radars are still limited by the additional costs of installing devices in the test environment that are not commonly used in daily lives [9].

Due to the development of Internet of things technology, WiFi devices have been widely distributed in various indoor environments. Moreover, different human activities between the transmitter (Tx) and the receiver (Rx) will cause different changes in the WiFi sensing signal [10]. Therefore, WiFi-based HAR systems are noncontact and low cost. WiFi received signal strength indicator (RSSI) and WiFi channel state information (CSI) are two commonly used WiFi wireless sensing signals. WiFi RSSI has been widely used in indoor localization of human [11], [12], whereas WiFi CSI is often used to human intelligent sensing, such as activity recognition [13], gesture recognition [9], and breath detection [14]. As RSSI only acquires coarse-grained signals from the WiFi media access control layer, and is prone to interference from multipath effect, it is not suitable for gesture recognition, breath detection, and other fine-grained HAR [15]. Differently, WiFi CSI is a fine-grained signal from the physical layer, which can reflect the signal scattering, distance attenuation, environmental attenuation, and other information of wireless signals on different propagation paths [16], and can change differently with different human movements, therefore, compared with WiFi RSSI, WiFi CSI is more suitable for activity recognition [17].

WiFi CSI signals can be picked up using a laptop equipped with a CSI acquisition tool, such as the CSI acquisition tool based on an Intel 5300 wireless card [18] or AX-CSI [19]. Due to the presence of environmental noise and system noise in CSI measurements, it is difficult for CSI measurements to be directly used to distinguish different human activities [20]. In order to address this problem, some researchers try to utilize machine learning algorithms to explore discriminating features in CSI measurements [21], [22]. However, machine learning-based

Manuscript received 9 June 2023; revised 16 September 2023; accepted 27 December 2023. Date of current version 26 January 2024. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62173032, in part by the Foshan Science and Technology Innovation Special Project under Grant BK22BF005, in part by the Regional Joint Fund of the Guangdong Basic and Applied Basic Research Fund under Grant 2022A1515140109, and in part by the Natural Science Foundation of Shandong Province under Grant ZR202212040125. This article was recommended by Associate Editor Z. Wang. *(Xingcan Chen and Yi Zou are co-first authors.) (Corresponding author: Wendong Xiao.)*

The authors are with the School of Automation & Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China, also with the Beijing Engineering Research Center of Industrial Spectrum Imaging, Beijing 100083, China, and also with the Shunde Innovation School, University of Science and Technology Beijing, Foshan 528399, China (e-mail: d202210337@xs.ustb.edu.cn; g20208669@xs.ustb.edu.cn; m202210526@xs.ustb.edu.cn; wdxiao@ustb.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/THMS.2023.3348694.

Digital Object Identifier 10.1109/THMS.2023.3348694

systems mostly require expert knowledge and hand-crafted features, which may inevitably lose some features hidden in CSI data streams that are crucial for classifying different activities, leading to unsatisfactory activity recognition accuracy [15]. Recent studies have shown that deep learning-based systems can automatically discover and learn discriminating features and automatically record time features in CSI measurement, and are successfully applied to HAR with significantly higher accuracy than classic machine learning-based systems [23].

Although existing systems based on deep learning have achieved good performance in HAR, there are still some problems that need to be addressed. First, the systems based on deep learning usually require a large amount of CSI data to train the model, and the large scale of original CSI data is easy to lead to the overfitting problem with noisy CSI measurements. Furthermore, in order to achieve better recognition accuracy, usually large-scale representations of complex networks are used, resulting in high model complexity and long training time. Second, most activity recognition systems based on CSI are trained in a well-controlled environment, which are usually inefficient in a new environment. Therefore, the cross-domain generalization ability of HAR systems based on CSI needs to be studied.

To address these problems, a reconstructed WiFi CSI tensor and deep learning based lightweight HAR system (Wisor-DL) is proposed. Wisor-DL consists of a commercial WiFi device as the transmitter (Tx), a laptop with an Intel 5300 NIC as the receiver (Rx), and a lightweight HAR algorithm and model. Specifically, Wisor-DL first filters out most irrelevant noises in original CSI data through principal component analysis (PCA) and low-pass filtering, the signal-to-noise ratio (SNR) of CSI is greatly improved. In order to appropriately reduce the dimensions of CSI measurements and enhance the SNR of CSI measurements, a sparse signal representation algorithm and a CSI tensor construction and decomposition algorithm are, respectively, used to reconstruct CSI signals. After CSI signal reconstruction, Wisor-DL fuse CSI data reconstructed by two algorithms through a gated temporal convolutional network with residual connections (GTCN-RC) rather than temporal convolutional network (TCN). Finally, dendrite network (DD) [24] replaces the traditional dense layer to recognize different human activities. By comparing with some state-of-the-art models in real environments, the experimental results show that Wisor-DL is a system with high accuracy, lightweight, and good cross-domain generalization ability.

The main contributions of this work are summarized as follows.

1) In order to significantly reduce the data dimension while enhancing the SNR of the original CSI measurements, the original CSI measurements are reconstructed by a sparse signal representation algorithm and CSI tensor construction and decomposition algorithm, respectively.
2) In order to fuse the features of reconstructed CSI data, GTCN-RC is designed, which is lightweight compared with the state-of-the-art models, has good recognition accuracy, and cross-domain generalization ability. Moreover, DD replaces the dense layer to make activity decisions to improve the training efficiency.
3) A large number of experiments in real environments show that the proposed Wisor-DL performs well in recognition accuracy, recognition efficiency, and cross-domain generalization ability, and its overall performance is better than that of other state-of-the-art models.

The rest of this article is organized as follows. Related works are introduced in Section II, followed by the preliminary works addressed in Section III. The detailed methodology is presented in Section IV. Experimental results are presented in Sections V. Finally, Section VI concludes this article.

II. Related Works

HAR based on WiFi CSI has been widely studied due to the advantages that it is noncontact and low cost. Zhang et al. [25] theoretically analyzed the perception ability of WiFi signals and proposed a Fresnel zone model through WiFi CSI signals. He et al. [21] proposed WiG, which can capture the characteristics of different human activity changes from CSI data sequences through the support vector machine, so as to realize HAR. Virmani and Shahzad [22] proposed WiAG, a WiFi CSI-based activity recognition system, where discrete wavelet transform is used to extract CSI temporal features from human activity samples, and then the K-nearest neighbor algorithm is used to distinguish different human activities. Based on the CSI-speed model and the CSI-activity model, Wang et al. [26] proposed CARM, which extracts temporal features from WiFi CSI through hidden Markov model to recognize different human activities.

The development of the above efforts is limited by the time and effort required to produce artificial features and the inevitable loss of some implicit but crucial features. These problems can be addressed by automatically capturing key features through deep learning. Wang et al. [27] employed sparse auto-encoder to capture the temporal identification features of WiFi CSI signals to realize HAR. Yousefi et al. [15] employed long short-term memory (LSTM) to automatically extract sequential temporal features from WiFi CSI measurements to recognize different human activities. For better HAR performance, Chen et al. [8] proposed ABLSTM, and Meng et al. [9] proposed ABGRU. Both ABLSTM and ABGRU make use of sequential and reverse CSI sequences and assign higher weights to more important features through the attention mechanism, thus, achieving better activity recognition performance than traditional deep learning.

Deep learning-based HAR system has achieved good recognition accuracy, but it is also accompanied by a large amount of computation, especially for double-layer network structure models, such as attention based bi-directional long short-term memory network (ABLSTM) and modified attention based bi-directional gate recurrent unit network (ABGRU). In order to reduce the computational complexity as much as possible, while maintaining high recognition accuracy, some systems are proposed to integrate different CSI features. Yang et al. [28] employed convolutional neural network (CNN) and recurrent neural network to extract distinguishing features from CSI data streams, respectively, and the features of different networks are combined to realize activity recognition. Similarly, Wang

et al. [29] proposed WiSDAR, which integrates the characteristics of CSI through CNN and LSTM to realize HAR based on WiFi CSI. Li et al. [10] proposed THAT model, which captures representative features from CSI channel streams and CSI temporal streams through the built-in two-tower structure.

In addition to recognition accuracy, recognition efficiency, and computational complexity, cross-domain generalization ability or one-shot learning ability are also the key issues of activity recognition systems based on WiFi CSI. Gu et al. [30] proposed WiGRUNT, a WiFi-enabled activity recognition system using a dual-attention network. WiGRUNT roots in a deep residual network backbone to evaluate the importance of spatial–temporal clues and exploit their built-in sequential correlations for fine-grained activity recognition, achieving good cross-domain generalization. Ding et al. [31] proposed RF-net, which achieves good cross-domain generalization performance by using an original signal adaptive CNN architecture rather than manual features, and utilizing point-group convolution and depth-separable convolution to limit model size and speed up reasoning execution time.

III. Preliminary

In this section, the basics of CSI and tensor are introduced.

A. Channel State Information

CSI is the frequency response of the wireless channel. Let f, t, $X(f,t)$, and $Y(f,t)$ denote the carrier frequency, the time node, the frequency domain representations of the transmitted signals, and the frequency domain representations of the received signals, respectively. The relationship between them is expressed as

$$Y(f,t) = H(f,t) \times X(f,t) + n(f,t) \tag{1}$$

where $n(f,t)$ denotes the additive white Gaussian noises. Equation (1) indicates that the CSI measurements $H(f,t)$ can be estimated by $X(f,t)$ and $Y(f,t)$.

The number of transmitting and receiving antennas are denoted as N_T and N_R. For a CSI-based HAR system using Intel 5300 NIC, there are $30 \times N_T \times N_R$ CSI streams in a time-series CSI measurement, as every CSI value includes 30 matrixes with dimensions $N_T \times N_R$.

The CSI measurements of subcarrier i is $H_i(f,t)$, which can be abbreviated as H_i, and denoted as

$$H_i = I_i + jK_i = |H_i|\exp(j\angle H_i) \tag{2}$$

where $I_i, K_i, |H_i|$, and $\angle H_i$ are the in-phase component, quadrature component, amplitude, and phase response of subcarrier i, respectively. Take the multipath components into account, H_i can also be denoted as

$$H_i = \sum_{q=1}^{N} R_q \cdot e^{-j2\pi f \tau_q} \cdot e^{-j2\pi \Delta f t} \tag{3}$$

where N is the number of multipath components, the initial phase offset and attenuation of the qth path form a complex value, represented as R_q, τ_q is the propagation delay on the qth path, and the phase shift caused by the difference Δf between the transmitting carrier frequency and the receiving carrier frequency is represented as $e^{-j2\pi\Delta f t}$.

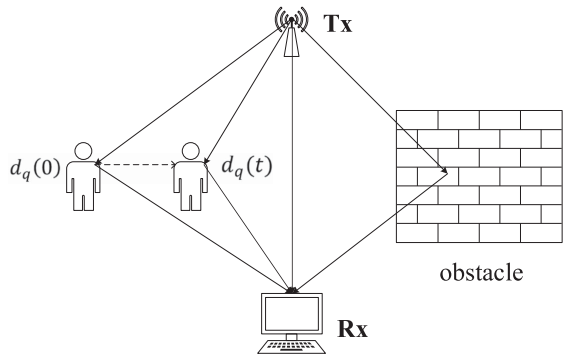

Fig. 1. Multipaths caused by human movements.

B. Impact of Activity on CSI

H_i will vary with the change of multipath and the length of multipath will change with the body motion of a human. Therefore, it is necessary to research the relationship between variation of CSI and body movement [32]. Considering the scenario in Fig. 1, the CSI signals are reflected by human body through the qth path, the length of the qth path varies from $d_q(0)$ to $d_q(t)$ when the human moves a short distance from time 0 to time t. Let λ and v_q represent the wavelength and the constant move speed in a short time period, respectively. The CSI signals spread at the speed of light c, it can be found that

$$\begin{aligned} \lambda &= c/f \\ d_q(t) &= d_q(0) \pm v_q t. \\ \tau_q &= d_q(t)/c. \end{aligned} \tag{4}$$

Therefore, $f \cdot \tau_q = d_q(t)/\lambda$. Then, (3) can be rewritten as

$$H_i = \sum_{q=1}^{N} R_q \cdot e^{-j2\pi d_q(t)/\lambda} \cdot e^{-j2\pi \Delta f t}. \tag{5}$$

It implies that the received subcarrier phase will transform 2π if the path length changes by a wavelength λ.

C. Tensor

A tensor can be viewed as a multidimensional array, the dimensions of a tensor are called ways or modes, the number of the modes is equal to the order of a tensor, an Mth-order or an M-way tensor is considered as a member of a tensor product of M vector spaces with its own coordinate system [33].

Furthermore, a vector is a first-order tensor, a matrix is a second-order tensor, and a cubic structure is a third-order tensor. The higher order tensors have an order of three or higher, which has been widely used in computer vision, wireless communications, healthcare and medical applications, data mining, and brain data analyses [34]. As the exponential increase in space and time complexity with the increase of tensors orders, the higher order tensors face various computing challenges, and may cause the curse of dimensionality [35]. Given the amount of computation and the impact on accuracy, a third-order tensor

is constructed in this article. Tensor decomposition is a useful method to decompose higher order tensors into a finite number of elements. Some necessary equations and definitions of tensor decomposition are afforded, that will be used in subsequent works.

Definition 3.1: For a tensor $\eta \in \mathbb{R}^{G_1 \times G_2 \times \cdots \times G_M}$, the square root of the square sum of all the elements is its Frobenius norm, defined by

$$\| \eta \|_F = \sqrt{\sum_{g_1=1}^{G_1} \sum_{g_2=1}^{G_2} \cdots \sum_{g_M=1}^{G_M} h_{g_1, g_2 \ldots g_M}^2}. \quad (6)$$

Definition 3.2: There are matrixes $U \in \mathbb{R}^{A \times B}$ and $V \in \mathbb{R}^{A \times B}$, $U * V$ represents the Hadamard product of U and V, given as

$$U * V = \begin{bmatrix} u_{11}v_{11} & u_{12}v_{12} & \cdots & u_{1B}v_{1B} \\ u_{21}v_{21} & u_{22}v_{22} & \cdots & u_{2B}v_{2B} \\ \vdots & \vdots & \vdots & \vdots \\ u_{A1}v_{A1} & u_{A2}v_{A2} & \cdots & u_{AB}v_{AB} \end{bmatrix}. \quad (7)$$

Definition 3.3: The Kronecker product of a matrix $U \in \mathbb{R}^{A \times B}$ and another matrix $V \in \mathbb{R}^{C \times D}$ is expressed as $U \otimes V$, which is a matrix with dimensions $AC \times BD$, denoted as

$$U \otimes V = \begin{bmatrix} u_{11}V & u_{12}V & \cdots & u_{1B}V \\ u_{21}V & u_{22}V & \cdots & u_{2B}V \\ \vdots & \vdots & \vdots & \vdots \\ u_{A1}V & u_{A2}V & \cdots & u_{AB}V \end{bmatrix}. \quad (8)$$

Definition 3.4: $U \odot V$ denotes the Khatri–Rao product of $U \in \mathbb{R}^{A \times B}$ and $V \in \mathbb{R}^{C \times B}$, which has a size of $(AC \times B)$ and is defined by

$$U \odot V = \left[u_1 \otimes v_1, u_2 \otimes v_2, \ldots, u_B \otimes v_B \right]. \quad (9)$$

Definition 3.5: An M-way tensor $\eta \in \mathbb{R}^{A_1 \times A_1 \times \cdots \times A_M}$ is rank-1 if it can be denoted as an outer product of M vectors, i.e.,

$$\eta = x_1 \circ x_2 \circ \cdots \circ x_M \quad (10)$$

where the symbol $\circ$ denotes the outer product of vectors, and the rank of the tensor η is the smallest number of rank-1 tensors that make up η.

IV. Methodology

In this section, the details of Wisor-DL, including its overall model architecture, sparse signal representation algorithm, CSI tensor construction and decomposition algorithm, GTCN-RC, and DD will be described.

A. Basic Structure of Wisor-DL

The basic structure of Wisor-DL is shown in Fig. 2, where $\oplus$ indicates the connection operation. PCA and low-pass filtering are firstly used to filter out the noises of raw CSI data. Then, the sparse signal representation algorithm is used to screen out subcarriers that are more relevant to changes in human movements to reduce the data dimension, and the phase difference of the screened subcarriers is used as the inputs of GTCN-RC. In addition, raw CSI is reconstructed into a new CSI tensor, and the peaks of its components are also used as inputs of GTCN-RC. Finally, DD fuses the features of two GTCN-RCs and makes the final activity decision.

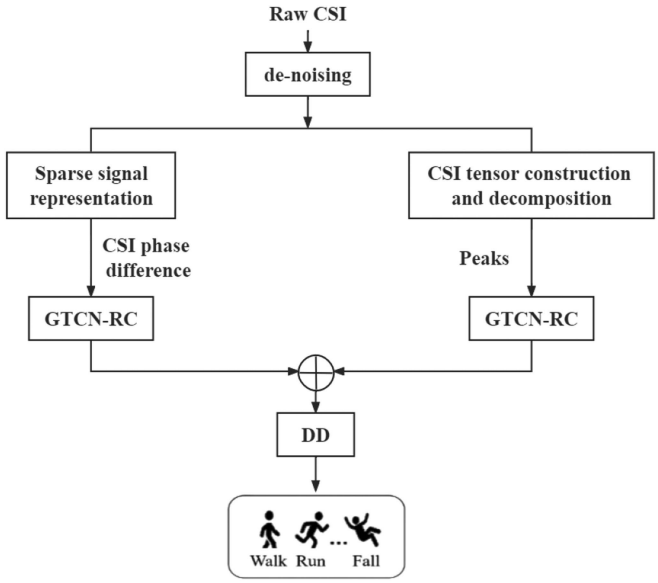

Fig. 2. Overall model architecture of Wisor-DL.

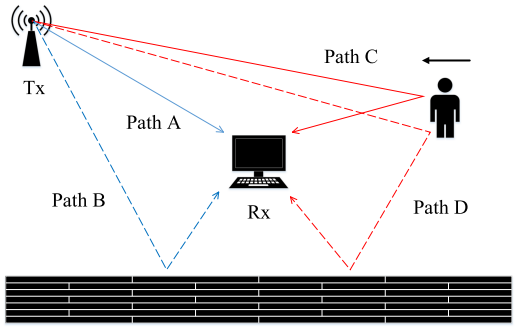

Fig. 3. WiFi signals travel through multiple paths.

B. Sparse Signal Representation

1) Sparse Representation of CSI: Normally, a whole CSI consists of dynamic CSI and static CSI [36], denoted as $H_{\text{dynamic}}(f,t)$ and $H_{\text{static}}(f,t)$, respectively, aliased as H_d and H_s, respectively. The static paths, such as path A and B in Fig. 3, constitute the static CSI, which is often treated as a constant. Dynamic CSI is the sum of paths that can be changed by human movements, such as path C and D in Fig. 3, and can be given as

$$H_d = \sum_{q \in P_d} R_q \cdot e^{-j2\pi d_q(t)/\lambda} \quad (11)$$

where the P_d are the dynamic paths affected by human movements. H_i can be expressed as

$$\begin{aligned} H_i &= H_d + H_s \\ &= \left(\sum_{q \in P_d} R_q \cdot e^{-j2\pi d_q(t)/\lambda} + H_s \right) \cdot e^{-j2\pi \Delta f t}. \end{aligned} \quad (12)$$

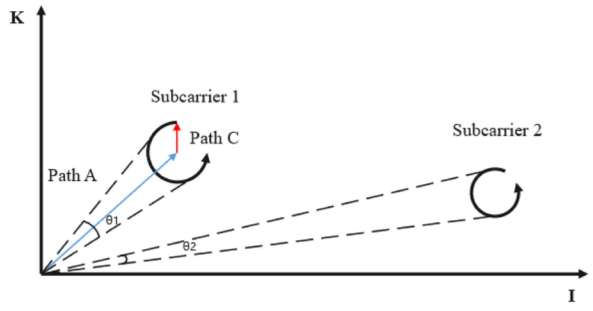

Fig. 4. Subcarrier phase representation.

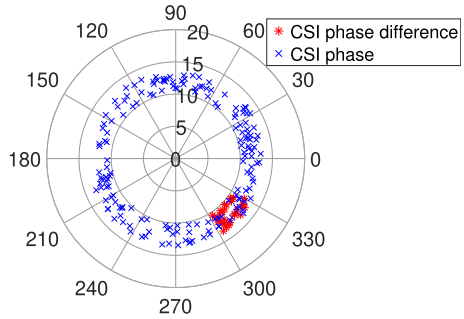

Fig. 5. Raw CSI phases and CSI phase differences of the second subcarrier for 200 packets.

Based on (2) and (12), we plot the Fig. 4. As shown in Fig. 4, it can be observed that the phase changes caused by human activity are different on different subcarriers, the phase change of subcarrier 1 is obviously greater than that of subcarrier 2, because subcarrier 1 has a higher proportion of dynamic components.

In our previous work [9], a sparse recovery algorithm is used to extract five subcarriers related to gesture change from 30 original subcarriers as inputs, which greatly reduces the complexity of the model and improves the recognition accuracy and cross-domain generalization ability. Therefore, in this article, the subcarriers are reordered from 1 to 30 according to the phase change of each subcarrier, and some subcarriers (such as ten subcarriers) are selected to recognize human activities, so that the dimension of CSI data is reduced from $30 \times N_T \times N_R$ to $10 \times N_T \times N_R$.

2) CSI Phase Difference: Generally, the phase of CSI is more related to the human activity change than the amplitude [37]. After subcarrier selection is completed, the CSI phase difference between two adjacent receiving antennas is calculated to classify different human activities. Let $\angle H_{mi}$ denote the phase measurement of subcarrier i, i.e.,

$$\angle H_{mi} = \angle H_i + (n_s + n_p)D_i + n_c + P + E \tag{13}$$

where $\angle H_i$ is the true CSI phase, D_i represents the index of subcarrier i, P is the initial phase offset caused by the phase-locked loop, E is the environment noises, and n_s, n_p, and n_c are the phase shifts due to the sampling frequency offset, the packet boundary detection, and the central frequency offset, respectively [38], given by

$$\begin{cases} n_s = 2\pi(\frac{T_r - T_s}{T_s})\frac{T_w}{T_l}m \\ n_p = 2\pi\frac{\Delta\tau}{S} \\ n_c = 2\pi\Delta f_d m \end{cases} \tag{14}$$

where T_r and T_s are the sampling periods of the receiver and sender, respectively, T_w is the whole length of the guard internal and data symbol, T_l denotes the data symbol length, m represents the time offset of sampling for current packet, S is the fast Fourier transform size, and the difference of center frequency between the sender and receiver is Δf_d. Based on (13), it is difficult to extract the true phase directly from the phase measurements of CSI.

Since the antennas of the Intel 5300 NIC have the same down-converter frequency and system lock, the CSI phases measured by different antennas have the same n_s, n_p, n_c, and m. The CSI phase difference on subcarrier i can be computed as

$$\Delta\angle H_{mi} = \Delta\angle H_i + \Delta P + \Delta E \tag{15}$$

where $\Delta\angle H_i$ denotes the true difference of phases on subcarrier i, ΔE represents the noise difference, and ΔP is an unknown difference of phase offsets, which turns out to be a constant [39].

In (15), we find that the random values, i.e., m, $\Delta\tau$, and Δf_d are all removed, and the difference of phases is more stable than the phase. Fig. 5 is plotted based on 200 continuously received packets from the second subcarrier, in which the phases of a single antenna marked as blue crosses, while the differences of the phases between two adjacent antennas shown as red asterisks. It can be found that the raw CSI phases scatter randomly over all feasible angles, whereas the CSI phase differences concentrate between 300° and 330°. Thus, the use of the CSI phase difference between two adjacent receiving antennas does remove the phase offset.

C. CSI Tensor Construction and Decomposition

1) CSI Tensor: The denoised CSI signals can not be used for tensor decomposition directly, a Hankelization way is used to transform the collected CSI matrixes to a CSI tensor [40]. The 2-D Hankel matrixes rearranged by CSI stream signals, are constructed to a 3-D tensor.

For subcarrier i, let L_i represent the constructed Hankel matrix, which has a size of $O \times Q$ and is created by mapping K packets, where $K = O + Q - 1$. Considering the case that $O = Q = K + 1/2$, the L_i is given by

$$L_i = \begin{bmatrix} l_i(0) & l_i(1) & \cdots & l_i(\frac{K-1}{2}) \\ l_i(1) & l_i(2) & \cdots & l_i(\frac{K+1}{2}) \\ \vdots & \vdots & \vdots & \vdots \\ l_i(\frac{K-1}{2}) & l_i(\frac{K+1}{2}) & \cdots & l_i(K+1) \end{bmatrix} \tag{16}$$

where $l_i(k)$ denotes the denoised CSI phase difference data for packet k. In this work, $O = Q = 300$ and $K = 599$ are set.

Assume that R represents the order of CSI tensor, the rank of L_i is $2R$. Let $S_r(t) = M_r \cos(\omega_r t + \varphi_r)$ denotes the rth activity signal, the whole signal is a sum of individual component signal,

represented by [41]

$$W(t) = \sum_{r=1}^{R} A_r S_r(t) = \sum_{r=1}^{R} \widehat{A}_r \cos(\omega_r t + \varphi_r) \tag{17}$$

where A_r denotes the coefficient of the rth activity, and $\widehat{A}_r = A_r M_r$. Based on the Euler's formula, the decomposition of the rth component $\widehat{A}_r \cos(\omega_r t + \varphi_r)$ can be expressed as

$$\widehat{A}_r \cos(\omega_r t + \varphi_r) = \frac{\widehat{A}_r}{2} e^{j\varphi_r} e^{j\omega_r t} + \frac{\widehat{A}_r}{2} e^{-j\varphi_r} e^{-j\omega_r t}. \tag{18}$$

Therefore, the whole signal can be obtained by

$$W(t) = \sum_{r=1}^{R} \widehat{A}_r \cos(\omega_r t + \varphi_r) = \sum_{r=1}^{2R} \widetilde{A}_r Z_r^t \tag{19}$$

where the new coefficient $\widetilde{A}_r = \widehat{A}_r 2e^{\pm j\varphi_r}$ and $Z_r^t = e^{\pm j\omega_r t}$. For K packets that received at discrete times, $W(t)$ will be converted to $W(k) = \sum_{r=1}^{2R} \widetilde{A}_r Z_r^k, k = 1, 2, \ldots, K$. Accordingly, the Hankel matrix L_i turns to

$$L_i = \begin{bmatrix} \sum_{r=1}^{2R} \widetilde{A}_r Z_r^0 & \sum_{r=1}^{2R} \widetilde{A}_r Z_r^1 & \cdots & \sum_{r=1}^{2R} \widetilde{A}_r Z_r^{\frac{K-1}{2}} \\ \sum_{r=1}^{2R} \widetilde{A}_r Z_r^1 & \sum_{r=1}^{2R} \widetilde{A}_r Z_r^2 & \cdots & \sum_{r=1}^{2R} \widetilde{A}_r Z_r^{\frac{K+1}{2}} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{r=1}^{2R} \widetilde{A}_r Z_r^{\frac{K-1}{2}} & \sum_{r=1}^{2R} \widetilde{A}_r Z_r^{\frac{K+1}{2}} & \cdots & \sum_{r=1}^{2R} \widetilde{A}_r Z_r^{K+1} \end{bmatrix}. \tag{20}$$

L_i can be decomposed as follows by Vandermonde decomposition [40]:

$$L_i = V_r \cdot \mathrm{diag}(\widetilde{A}_1, \widetilde{A}_2, \ldots, \widetilde{A}_{2R}) \cdot \widetilde{V}_r^T. \tag{21}$$

The Vandermonde matrixes V_r and $\widetilde{V}_r$ are denoted as

$$V_r = \widetilde{V}_r = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ Z_1 & Z_2 & \cdots & Z_{2R} \\ \vdots & \vdots & \vdots & \vdots \\ Z_1^{\frac{K-1}{2}} & Z_2^{\frac{K-1}{2}} & \cdots & Z_{2R}^{\frac{K-1}{2}} \end{bmatrix}. \tag{22}$$

As the Vandermonde matrix is full rank, the Hankel matrix has a rank of $2R$.

2) Canonical Polyadic Decomposition: After the Hankel matrix construction is completed, the canonical polyadic (CP) decomposition is used to further process the CSI data. A CSI tensor can be viewed as a sum of $2R$ rank-1 tensors with CP decomposition. Let $\eta \in \mathbb{R}^{A \times B \times C}$ represent the third-order CSI tensor, which can be approximated as a sum of 3-way outer products [34], denoted as

$$\eta \approx \sum_{r=1}^{2R} x_r \circ y_r \circ z_r \tag{23}$$

where $x_r \in \mathbb{A}$, $y_r \in \mathbb{B}$, and $z_r \in \mathbb{C}$ are the vectors for the first, second, and third dimensions at the rth position, respectively. $2R$ is the approximation rank of η and the number of CP decomposition components. For all $a = 1, 2, \ldots, A$, $b = 1, 2, \ldots, B$, and $c = 1, 2, \ldots, C$, the outer product is given by [42]

$$x_r \circ y_r \circ z_r(a, b, c) = x_r(a) y_r(b) z_r(c). \tag{24}$$

Let matrixes $\mathbf{X} = [x_1, x_2, \ldots, x_{2R}] \in \mathbb{R}^{A \times 2R}$, $\mathbf{Y} = [y_1, y_2, \ldots, y_{2R}] \in \mathbb{R}^{B \times 2R}$, and $\mathbf{Z} = [z_1, z_2, \ldots, z_{2R}] \in \mathbb{R}^{C \times 2R}$ represent the vectors combined by rank-1 components. Assuming that $\eta_{(1)} \in \mathbb{R}^{A \times BC}$, $\eta_{(2)} \in \mathbb{R}^{B \times AC}$, and $\eta_{(3)} \in \mathbb{R}^{C \times AB}$ denote the 1-mode, 2-mode, and 3-mode matrixing of the CSI tensor $\eta \in \mathbb{R}^{A \times B \times C}$ respectively [33]. The three matrixing forms can be approximated as

$$\begin{cases} \eta_{(1)} \approx \mathbf{X}(\mathbf{Z} \odot \mathbf{Y})^T \\ \eta_{(2)} \approx \mathbf{Y}(\mathbf{Z} \odot \mathbf{X})^T \\ \eta_{(3)} \approx \mathbf{Z}(\mathbf{Y} \odot \mathbf{X})^T \end{cases}. \tag{25}$$

The smallest square sum of difference between the constructed CSI tensor and the estimated CSI tensor is given by

$$\min_{\mathbf{X},\mathbf{Y},\mathbf{Z}} \left\| \eta - \sum_{r=1}^{2R} x_r \circ y_r \circ z_r \right\|^2 \tag{26}$$

which is nonconvex. The alternating least squares algorithm can fix two factor matrixes, which can simplify (26) to a linear least square minimization problem with the other one factor matrix as variable. Fixing $\mathbf{Y}$ and $\mathbf{Z}$, the problem (26) can be rewritten as

$$\min_{\mathbf{X}} \|\eta_{(1)} - \mathbf{X}(\mathbf{Z} \odot \mathbf{Y})^T\|^2. \tag{27}$$

The optimal solution of (27) is

$$\mathbf{X} = \eta_{(1)}[(\mathbf{Z} \odot \mathbf{Y})^T]^\dagger \tag{28}$$

where the symbol † represents the Moore–Penrose pseudoinverse of a matrix. According to the property of the Khatri–Rao product pseudoinverse, we can rewrite (28) as

$$\mathbf{X} = \eta_{(1)}(\mathbf{Z} \odot \mathbf{Y})(\mathbf{Z}^T\mathbf{Z} * \mathbf{Y}^T\mathbf{Y})^\dagger. \tag{29}$$

It only needs to calculate the pseudoinverse of the matrix with a size of $2R \times 2R$ rather than a $BC \times 2R$ matrix. Note that B and C are much bigger than $2R$, the computational complexity can be reduced effectively. Similarly, the optimal solutions of Y and Z can be obtained as

$$\mathbf{Y} = \eta_{(2)}(\mathbf{Z} \odot \mathbf{X})(\mathbf{Z}^T\mathbf{Z} * \mathbf{X}^T\mathbf{X})^\dagger \tag{30}$$

$$\mathbf{Z} = \eta_{(3)}(\mathbf{Y} \odot \mathbf{X})(\mathbf{Y}^T\mathbf{Y} * \mathbf{X}^T\mathbf{X})^\dagger. \tag{31}$$

The basic theorem about the uniqueness of CP decomposition is given in Theorem 4.1

Theorem 4.1: For a tensor η with rank R, if $p_{\mathbf{X}} + p_{\mathbf{Y}} + p_{\mathbf{Z}} \geq 2R + 2$, then, the CP decomposition of η is unique, where $p_{\mathbf{X}}$, $p_{\mathbf{Y}}$, and $p_{\mathbf{Z}}$ denote the p-rank of matrixes $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, respectively. Here p-rank means the maximum value p, such that any p columns are linearly independent.

The CSI tensor η is created by 30 Hankel matrixes with a rank of $2R$. Based on (22), there are $p_{\mathbf{Y}} = 2R$ and $p_{\mathbf{Z}} = 2R$ for p-rank matrixes $\mathbf{Y}$ and $\mathbf{Z}$. Moreover, the CSI phase differences between Antennas 1 and 2, Antennas 1 and 3, and Antennas 2 and 3 are independent, the p-rank of matrix $\mathbf{X}$ has $p_{\mathbf{X}} \geq 3$. Thus, $p_{\mathbf{X}} + p_{\mathbf{Y}} + p_{\mathbf{Z}} \geq 2R + 2R + 3 > 2(2R) + 2$, i.e., the CP decomposition is unique of the created CSI tensor η.

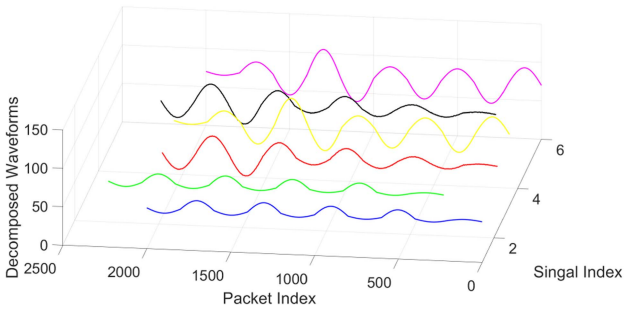

Fig. 6. CP decomposed waveforms.

For a CSI tensor with order 3, the decomposed waveforms are shown in Fig. 6. There are 2×3 waveforms, two of which represent the same components needed to be matched.

3) Waveforms Matching: After CP decomposition, the CSI tensor will be decomposed to $2R$ decomposition waveforms, which include both discrepant waveforms and similar waveforms. Moreover, the decomposed waveforms have a random order and pairwise similarity. Therefore, the waveform matching is necessary. The data length of a decomposed signal may increase due to its autocorrelation, that can improve the performance of the peak detection. There are nonalignment and phase shifts in the the decomposed signals, the phase shifts can be reduced and the periodicity will be enhanced with the autocorrelation of decomposed signals. Two same length autocorrelation signals are different if either of them has a micro shift by the Euclidean distance algorithm. Dynamic time warping (DTW) can automatically identify phase shifts and provide the similar distance measurement between two autocorrelation signals by aligning the corresponding time series, thus, overcoming the limitation of the Euclidean distance algorithm [43]. Therefore, the DTW algorithm is used to measure the distance between two curves with the autocorrelation signals. Downsampling is used in autocorrelation signals to reduce the packets number to K', and the computational complexity of DTW is reduced. For two downsampled autocorrelation signals $F_o = [F_o(0), F_o(1), \ldots, F_o(K'-1)]$ and $F_q = [F_q(0), F_q(1), \ldots, F_q(K'-1)]$, the warping path is $\mathcal{B} = [\beta_1, \beta_2, \ldots, \beta_{\mathcal{L}}]$, where $\mathcal{L}$ is the path length, and $\omega_\ell = (\vartheta_\ell, \gamma_\ell)$ denotes the ℓth element of $\mathcal{B}$, where ϑ and γ are the indexes of packets for a pair of downsampled autocorrelation signals. The minimum total cost function of F_o and F_q is given by

$$\begin{aligned} \min \ & \sum_{\ell=1}^{\mathcal{L}} \|F_o(\vartheta_\ell) - F_q(\gamma_\ell)\| \\ \text{s.t.} \ & (\vartheta_1, \gamma_1) = (0,0). \\ & (\vartheta_{\mathcal{L}}, \gamma_{\mathcal{L}}) = (K'-1, K'-1) \\ & \vartheta_\ell \le \vartheta_{\ell+1} \le \vartheta_\ell + 1 \\ & \gamma_\ell \le \gamma_{\ell+1} \le \gamma_\ell + 1 \end{aligned} \tag{32}$$

Considering the 2-D cost matrix $\mathcal{C}$, which has a size of $K' \times K'$, and whose element $\mathcal{C}(\vartheta_\ell, \gamma_\ell)$ represents the shortest warping path of $F_o = [F_o(0), F_o(1), \ldots, F_o(\vartheta_\ell)]$ and $F_q = [F_q(0), F_q(1), \ldots, F_q(\gamma_\ell)]$. The $\mathcal{C}(\vartheta_\ell, \gamma_\ell)$ can be obtained by

$$\begin{aligned} \mathcal{C}(\vartheta_\ell, \gamma_\ell) &= \|F_o(\vartheta_\ell) - F_q(\gamma_\ell)\| \\ &+ \min[\mathcal{C}(\vartheta_\ell - 1, \gamma_\ell), \mathcal{C}(\vartheta_\ell, \gamma_\ell - 1), \mathcal{C}(\vartheta_\ell - 1, \gamma_\ell - 1)]. \end{aligned} \tag{33}$$

The value of $\mathcal{C}(K'-1, K'-1)$, which can be viewed as the DTW value of a couple of downsampled autocorrelation signals, and can be computed by filling all the elements of the 2-D cost matrix $\mathcal{C}$.

To facilitate the follow-up work, once waveforms matching is completed, a couple of activity signals are fused to a single activity signal by averaging the signal pair to reduce the variance of the decomposed signals while maintain the same period [35]. The fused waveforms are then reordered according to an evaluation function defined as

$$e_{\text{fun}} = \xi(v_{\max} - v_{\min}) + (1-\xi)\text{average}\left(\sum_{g=1}^{g_{\max}} v_g\right) \tag{34}$$

where $\xi \in [0,1]$ is an adjustable parameter, $v_{\max}$ and $v_{\min}$ are the maximum and minimum peaks corresponding to the waveform, and $g_{\max}$ and v_g are the maximum number and the gth peak value of the corresponding waveform, respectively.

The peaks of the reordered waveforms are also used as the input features. In particular, the data length of each waveform is fixed as the number of peaks of the waveform with the largest number of peaks. In order to ensure the consistency of input signal length, the data length is supplemented by complementing 0 on the left side, which will not destroy the temporal features.

D. Gated Temporal Convolutional Network With Residual Connections

GTCN-RC is designed to learn the features of reconstructed CSI signals to recognize different activities. GTCN-RC does not destroy the temporal features of the reconstructed CSI signals because its internal convolution is right aligned, i.e., GTCN-RC strictly adheres to the order of data modeling by right-aligned convolution, where the estimation $e(h^t|h^1, h^2, \ldots, h^{t-1})$ at time step t in the GTCN-RC cannot depend on any future time steps $h^{t+1}, h^{t+2}, \ldots, h^T$.

Fig. 7 shows the convolutional blocks in GTCN-RC, which includes single-layer causal dilated convolution, weight normalization layer, rectified linear unit, dropout layer, and the gated unit. Hyperbolic tangent activation function, sigmoid activation function and Hadamard product are, respectively, represented by $\tanh$, σ, and $\odot$. The gated unit composed of the hyperbolic tangent and sigmoid activation functions, the intuition behind our gated unit is that the hyperbolic tangent activation function acts as a common activation function to constrain the output values within the range of $[-1, 1]$. On the other hand, the sigmoid function serves as a gating mechanism to dynamically activate or deactivate unnecessary time nodes, i.e., time nodes that are not important for cross-domain recognition will be assigned 0, which will not participate in the subsequent calculation, so the computational cost of GTCN-RC is reduced and the computational efficiency of GTCN-RC is improved. Conversely,

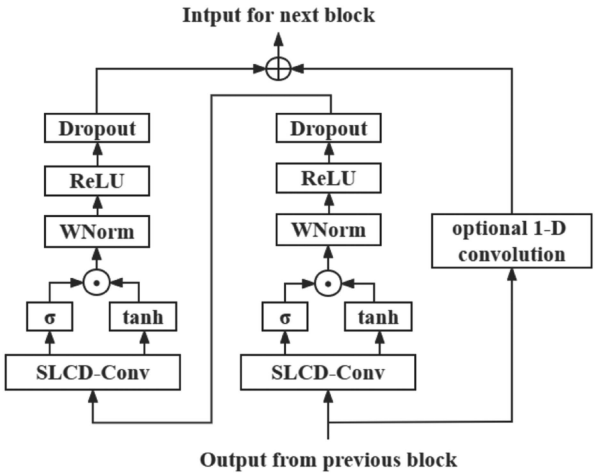

Fig. 7. Convolutional blocks in GTCN-RC.

time nodes that are important for cross-domain recognition are retained. Moreover, the gated unit will assign larger values to the more important time nodes. Therefore, the cross-domain recognition accuracy of GTCN-RC is improved.

E. Dendrite Network

DD is used to make the final activity decision, because it has controllable accuracy for better generalization capability, stronger expression ability, and lower computational complexity than the residual connected dense layer, which only represents the additive relation [24]. DD includes DD modules and linear modules, and the DD module is denoted as

$$\mathbf{Z}_k = \mathbf{U}_{k,k-1}\mathbf{Z}_{k-1} \odot \mathbf{V} \tag{35}$$

where $\mathbf{Z}_{k-1}$ and $\mathbf{Z}_k$ are the input and output of the kth DD module, respectively. $\mathbf{U}_{k,k-1}$ is the weight matrix from the $(k-1)$th module to the kth module, $\mathbf{V}$ denotes the inputs of DD, and $\mathbf{Z}_0 = \mathbf{V}$.

The final output of DD is given by

$$\mathbf{o}_{\text{dd}} = \mathbf{U}_{k,k-1}(\cdots \mathbf{U}_{2,1}(\mathbf{U}_{1,0}\mathbf{V} \odot \mathbf{V}) \odot \mathbf{V} \cdots). \tag{36}$$

DD has significantly lower computational complexity than residual connected dense layer, because it performs matrix multiplication and Hadamard product to achieve high-power representation instead of nonlinear mapping.

V. Experiments

In this section, the details of the experiments are firstly elaborated. Then, the accuracy, efficiency, and cross-domain generalization ability of the proposed Wisor-DL are verified through experiments compared with some existing state-of-the-art models. Finally, an ablation study is conducted to analyze the role of each component of Wisor-DL.

A. Experimental Setup

In all of our experiments, three datasets collected by Intel 5300 NICs are used to evaluate the performance of Wisor-DL. The first dataset is derived from [15], which can be found in https: //github.com/ermongroup/Wifi_Activity_Recognition. WiFi CSI is collected by an Intel 5300 NIC with the sampling frequency is set as 1 kHz. There are six activities in the first dataset, including Lie down, Fall, Walk, Run, Sit down, and Stand up. The first dataset is collected by six volunteers individually at different time. For each volunteer, each activity is collected 20 times with a window size of 2 s.

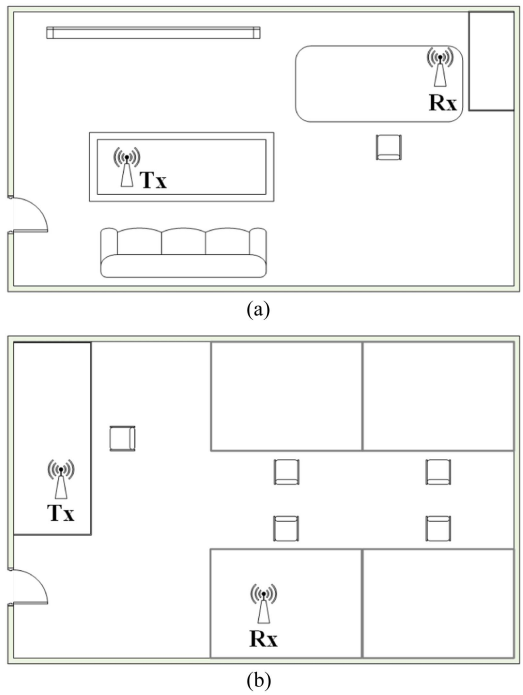

Fig. 8. Layouts of experimental environments. (a) Office room. (b) Laboratory room.

TABLE I
Statistics of the Three Evaluation Datasets

Datasets	Scenes	Activ.	Freq.	Time	Dist.
First dataset	Office room	6	1 kHz	2s	3 m
Second dataset	Office room	6	500 Hz	4s	3.5 m
Third dataset	Laborotary	6	500 Hz	4s	2.5 m

An office room with a size of 4400 mm × 2650 mm and a laboratory room with a size of 4400 mm × 3600 mm are used to collect the second dataset and third dataset, and their layouts are shown in Fig. 8. Tx and Rx are set 3.5 m apart in the office room and 2.5 m apart in the laboratory room under the line-of-sight condition. The Tx is a commercial WiFi router, and the Rx is a laptop equipped with an Intel 5300 NIC. Sampling frequency is set as 500 Hz. At different time, eight volunteers are recruited to individually collect each activity in 15 s with each activity of each volunteer is sampled 100 times, and the data will be split by a sliding window of 4 s. There are six activities that are not exactly the same with the first dataset in both the second dataset and third dataset, i.e., Jump, Stoop, Wave hand, Fall, Sit down, and Stand up. Table I shows the statistical information of the three datasets that are used in all of our experiments.

A workstation with NVIDIA GeForce RTX3090 GPU and Intel i9-10900 K 3.70 GHz CPU is used for all experiments.

TABLE II
CONFUSION MATRIXES FOR ALL THE MODELS ON THE FIRST DATASET

	Lie down	Fall	Walk	Run	Sit down	Stand up	Average
CNN [7]	78.32%	80.69%	82.53%	81.84%	76.14%	82.06%	80.2633%
LSTM [15]	80.15%	85.33%	87.76%	87.42%	82.51%	87.60%	85.1283%
ABLSTM [8]	95.47%	96.28%	97.24%	98.96%	97.59%	99.74%	97.5467%
THAT [10]	96.83%	98.25%	98.14%	98.06%	97.67%	99.51%	98.0767%
Siamese [28]	96.96%	99.04%	97.28%	97.95%	98.43%	99.86%	98.2533%
HAR-SAnet [3]	96.89%	98.43%	98.27%	98.14%	98.64%	99.79%	98.3600%
Wisor-DL	98.52%	98.04%	98.31%	97.69%	98.38%	99.72%	98.4433%

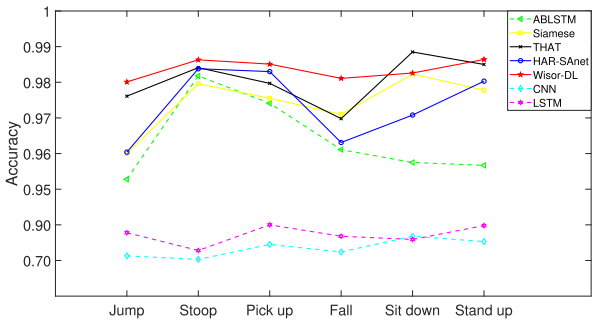

Fig. 9. Average accuracy of all models on the second dataset and third dataset.

TABLE III
TRAINING TIME AND TESTING TIME OF ALL THE MODELS ON THE SECOND DATASET

	Training time (s) of all training samples	Testing time (ms) of each testing sample
CNN [7]	1528.32	2.67
LSTM [15]	5412.68	10.46
ABLSTM [8]	12316.52	16.34
THAT [10]	3372.72	3.69
Siamese [28]	14532.14	17.12
HAR-SAnet [3]	2707.96	3.28
Wisor-DL	1857.44	2.81

Xavier [44] with a gain of 1 is used to initialize all weight parameters. In order to reduce the risk of overfitting, the ADAM [45] optimizer is used to calculate the learning rate for each parameter with the initial learning rate and weight decay set to 0.0001 and 0.001, respectively. The same training settings are applied to all experiments. For all datasets, the ten-fold cross validation is used for evaluation, and the average result of all 10 runs is considered as the final recognition accuracy. In addition, all testing models are limited to run with a maximum of 50 epochs.

In order to verify the performance of Wisor-DL, some state-of-the-art HAR models based on CSI are compared with it. The compared models include CNN [7], LSTM [15] ABLSTM [8], THAT using two-stream CNN [10], Siamese combining CNN and BiLSTM [28], and HAR-SAnet using two-stream signal adapted CNN [3].

B. Recognition Accuracy

To verify the accuracy of Wisor-DL in recognizing different human activities, Wisor-DL is compared with other current state-of-the-art models on the first dataset, and the confusion matrixes are shown in Table II.

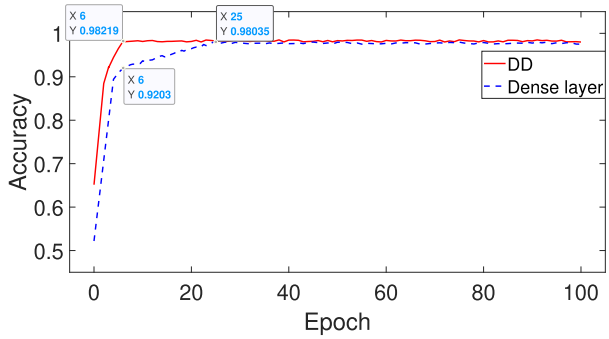

Fig. 10. Accuracy curves of different classification layers.

TABLE IV
MODEL COMPLEXITY OF ALL THE COMPARED MODELS ON THE FIRST DATASET

	Computational complexity	Number of parameters
CNN [7]	0.26 GMac	5.32 M
LSTM [15]	0.52 GMac	11.28 M
ABLSTM [8]	2.24 GMac	32.44 M
THAT [10]	1.68 GMac	27.14 M
Siamese [28]	2.83 GMac	37.56 M
HAR-SAnet [3]	1.15 GMac	20.67 M
Wisor-DL	0.83 GMac	16.43 M

TABLE V
CROSS-DOMAIN RECOGNITION ACCURACIES OF ALL THE MODELS ON THE SECOND DATASET AND THIRD DATASET

	Average accuracy on dataset2 and dataset3	Average accuracy of cross-domain
CNN [7]	75.54%	60.38%
LSTM [15]	82.57%	67.43%
ABLSTM [8]	95.63%	87.31%
THAT [10]	96.93%	93.82%
Siamese [28]	96.24%	92.54%
HAR-SAnet [3]	97.76%	96.24%
Wisor-DL	98.00%	97.57%

It is found that the traditional deep learning models, i.e., CNN and LSTM, are not ideal in the overall recognition accuracy, and are easy to confuse CSI features with similar changes. Particularly, it is more difficult for CNN to distinguish the activities with opposite time-domain characteristics, such as Stand up and Sit down, because CNN is difficult to retain the time sequence of features. Due to the improvement of the traditional network, ABLSTM, THAT, Siamese, HAR-SAnet, and Wisor-DL all perform well on the first dataset, with the recognition accuracy reaching 97.55%, 98.08%, 98.25%, 98.36%, and 98.44%, respectively.

In order to ensure the accuracy of the results, experiments on the second dataset and third dataset are also conducted, and the average accuracy of each activity are shown in Fig. 9 (Due to the large difference in data, equal-spacing coordinate distance representation of unequal spacing data values is applied for more intuitive presentation). As can be seen from Table II and Fig. 9, Wisor-DL performs well in terms of recognition accuracy, which is better than other models used for comparisons.

TABLE VI
ABLATION STUDY RESULTS ON THE SECOND DATASET AND THIRD DATASET

	Average accucy	Δ	Average cross-domain accuracy	Δ
Wisor-DL	98.00%	–	97.57%	–
with CSI phase difference replaced by CSI amplitude	96.33%	–1.67%	94.42%	–3.15%
with CSI phase difference replaced by CSI phase	96.74%	–1.26%	94.87%	–2.70%
with CSI phase difference replaced by complete CSI	97.12%	–0.88%	95.34%	–2.23%
without sparse signal representation algorithm	97.67%	–0.33%	90.53%	–7.04%
without CSI tensor construction and decomposition algorithm	97.43%	–0.57%	92.18%	–5.39%
with GTCN-RC replaced by TCN	84.82%	–13.18%	72.26%	–25.31%

C. Recognition Efficiency

Experiments on the first dataset to evaluate the efficiency of all the systems and experiments on the second dataset to evaluate the computation cost of all the systems are also performed, and the results are shown in Tables III and IV. It can be found that CNN, THAT, HAR-SAne, and Wisor-DL perform well in terms of model training time and testing time because the convolutional network can be computed in parallel. Considering the model training time, model testing time, computational burden, and number of parameters, Wisor-DL has the best performance in recognition efficiency. In addition, since all models can be calculated offline, the time cost of the training phase will not be a major problem. The average testing time for Wisor-DL to recognize a human activity is 2.81 ms. For each human activity with a sliding window size of 4 s, Wisor-DL could recognize a human activity in real time.

D. Cross-Domain Generalization Ability

Cross-domain generalization ability is one of the key properties of WiFi based HAR systems. Experiments on the second and third datasets are performed to verify the cross-domain generalization ability of all models, and the experimental results are shown in the Table V.

It can be found that traditional deep learning networks, such as CNN and LSTM, perform poorly in cross-domain generalization, and their recognition accuracy is reduced by about 15%. The cross-domain recognition accuracies of ABLSTM, THAT, Siamese, and HAR-SAnet decrease by about 8%, 3%, 4%, and 2%, respectively, whereas the average cross-domain recognition accuracy of Wisor-DL only decreases by about 0.5%. Therefore, it can be claimed that Wisor-DL has good cross-domain generalization ability.

E. Ablation Study

Ablation experiments are performed on the second and third datasets to test the role of each module in Wisor-DL, and the average accuracies are shown in Table VI. It can be found that using CSI phase difference instead of CSI amplitude, phase or original CSI can slightly improve the recognition accuracy and cross-domain recognition accuracy, because the CSI phase difference between two adjacent receiving antennas is more stable. The application of the two CSI signal reconstruction algorithms, i.e., the sparse signal representation algorithm and CSI tensor construction and decomposition algorithm, hardly improves the recognition accuracy of Wisor-DL, but significantly improves the cross-domain generalization ability of Wisor-DL. Specifically, the sparse signal representation algorithm extracts part of the subcarriers related to the human activity from the original subcarriers to reduce the dimension of the original CSI data. Therefore, the computational burden of Wisor-DL is reduced, and the SNR of CSI data is enhanced. Then, the CSI tensor construction and decomposition algorithm constructs original CSI to novel CSI tensors and then decomposes them with uniqueness, so the features of CSI are highlighted and enhanced. Its application further improves the accuracy of activity recognition and the cross-domain generalization performance. Moreover, GTCN-RC maintains the temporal features of CSI, and its built-in gated unit reinforces discriminating features and skips unnecessary features. Based on the above reasons, the recognition accuracy, recognition efficiency, and cross-domain generalization performance of Wisor-DL are improved.

In addition, the contribution of DD to Wisor-DL with 100 epochs is tested on the third dataset, and the results are shown in Fig. 10. It can be found that the training accuracy of Wisor-DL with DD reaches 98.22% in the 6th epoch, whereas Wisor-DL with dense layer reaches 92.03% and 98.04% in the 6th and 25th epoch, respectively. Therefore, Wisor-DL with DD outperforms that with dense layer.

VI. CONCLUSION

This article proposes Wisor-DL, a lightweight HAR system through WiFi CSI. Wisor-DL firstly uses two CSI signal reconstruction algorithms, i.e., CSI sparse signal representation algorithm and CSI tensor construction and decomposition algorithm, to reconstruct original CSI data, which reduces the computational complexity and significantly improves the cross-domain generalization ability. GTCN-RC is designed to capture the features of reconstructed CSI data. In addition, DD replaces the traditional dense layer to make activity decisions. Experimental results show that Wisor-DL is a lightweight WiFi CSI based HAR system with better performance compared with existing systems.

Although the proposed system performs well in terms of lightweight and cross-domain generalization, it is still constrained by some problems, such as it is difficult to recognize activities of multiple people simultaneously in the same scene, and it does not support background traffic data for devices when activity recognition is taking place. These problems will be studied in the future work.

REFERENCES

[1] P.-G. Jung, G. Lim, S. Kim, and K. Kong, "A wearable gesture recognition device for detecting muscular activities based on air-pressure sensors," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 485–494, Apr. 2015.

[2] G. Chen et al., "A novel illumination-robust hand gesture recognition system with event-based neuromorphic vision sensor," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 508–520, Apr. 2021.

[3] Z. Chen, C. Cai, T. Zheng, J. Luo, J. Xiong, and X. Wang, "RF-Based human activity recognition using signal adapted convolutional neural network," *IEEE Trans. Mobile Comput.*, vol. 22, no. 1, pp. 487–499, Jan. 2023.

[4] S. Skaria, A. A.-Hourani, M. Lech, and R. J. Evans, "Hand-gesture recognition using two-antenna doppler radar with deep convolutional neural networks," *IEEE Sensors J.*, vol. 19, no. 8, pp. 3041–3048, Apr. 2019.

[5] Y. He, Y. Chen, Y. Hu, and B. Zeng, "WiFi vision: Sensing, recognition, and detection with commodity MIMO-OFDM WiFi," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8296–8317, Sep. 2020.

[6] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1629–1645, Mar. 2020.

[7] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "SignFi: Sign language recognition using WiFi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, 2018, Art. no. 23.

[8] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "WiFi CSI based passive human activity recognition using attention based BLSTM," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2714–2724, Nov. 2019.

[9] W. Meng, X. Chen, W. Cui, and J. Guo, "WiHGR: A robust WiFi-Based human gesture recognition system via sparse recovery and modified attention-based BGRU," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 10272–10282, Dec. 2022.

[10] B. Li, W. Cui, W. Wang, L. Zhang, Z. Chen, and M. Wu, "Two-stream convolution augmented transformer for human activity recognition," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 1, 2021, pp. 286–293.

[11] W. Cui et al., "Received signal strength based indoor positioning using a random vector functional link network," *IEEE Trans. Ind. Informat.*, vol. 14, no. 5, pp. 1846–1855, May 2018.

[12] L. Zhang et al., "WiFi-based indoor robot positioning using deep fuzzy forests," *IEEE Internet Things J.*, vol. 7, no. 11, pp. 10773–10781, Nov. 2020.

[13] J. Huang et al., "PhaseAnti: An anti-interference WiFi-based activity recognition system using interference-independent phase component," *IEEE Trans. Mobile Comput.*, vol. 22, no. 5, pp. 2938–2954, May 2023.

[14] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang, "FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 3, 2019, Art. no. 121.

[15] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.

[16] Y. Tian, C. Chen, Q. Zhang, Y. Li, S. Li, and X. Ding, "Multidimensional information recognition algorithm based on CSI decomposition," *IEEE Internet Things J.*, vol. 10, no. 10, pp. 9234–9248, May 2023.

[17] Y. Zhang, Q. Liu, Y. Wang, and G. Yu, "CSI-based location-independent human activity recognition using feature fusion," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.

[18] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, 2011.

[19] F. Gringoli, M. Cominelli, A. B. Pizarro, and J. Widmer, "AX-CSI: Enabling CSI extraction on commercial 802.11ax Wi-Fi platforms," in *Proc. 15th ACM Workshop Wireless Netw. Testbeds, Exp. Eval. & CHaracterization*, 2021, pp. 46–53.

[20] Z. Shi, Q. Cheng, J. A. Zhang, and R. Yi Da Xu, "Environment-robust WiFi-based human activity recognition using enhanced CSI and deep learning," *IEEE Internet Things J.*, vol. 9, no. 24, pp. 24643–24654, Dec. 2022.

[21] W. He, K. Wu, Y. Zou, and Z. Ming, "WiG: WiFi-based gesture recognition system," in *Proc. 24th Int. Conf. Comput. Commun. Netw.*, 2015, pp. 1–7.

[22] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using WiFi," in *Proc. 15th Annu. Int. Conf. Mobile Syst., Appl., Serv.*, T. Choudhury, S. Y. Ko, A. Campbell, and D. Ganesan, Eds. ACM, 2017, pp. 252–264.

[23] L. Zhang, W. Cui, B. Li, Z. Chen, M. Wu, and T. S. Gee, "Privacy-preserving cross-environment human activity recognition," *IEEE Trans. Cybern.*, vol. 53, no. 3, pp. 1765–1775, Mar. 2023.

[24] G. Liu and J. Wang, "Dendrite net: A white-box module for classification, regression, and system identification," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13774–13787, Dec. 2022.

[25] D. Zhang, H. Wang, and D. Wu, "Toward centimeter-scale human activity sensing with Wi-Fi signals," *Computer*, vol. 50, no. 1, pp. 48–57, 2017. [Online]. Available: https://doi.org/10.1109/MC.2017.7

[26] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, May 2017.

[27] J. Wang, X. Zhang, Q. Gao, H. Yue, and H. Wang, "Device-free wireless localization and activity recognition: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6258–6267, Jul. 2017.

[28] J. Yang, H. Zou, Y. Zhou, and L. Xie, "Learning gestures from WiFi: A Siamese recurrent convolutional architecture," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10763–10772, Jun. 2019.

[29] F. Wang, W. Gong, and J. Liu, "On spatial diversity in WiFi-based human activity recognition: A deep learning-based approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2035–2047, Feb. 2019.

[30] Y. Gu et al., "WiGRUNT: WiFi-Enabled gesture recognition using dual-attention network," *IEEE Trans. Human-Mach. Syst.*, vol. 52, no. 4, pp. 736–746, Apr. 2022.

[31] S. Ding, Z. Chen, T. Zheng, and J. Luo, "RF-net: A unified meta-learning framework for RF-enabled one-shot human activity recognition," in *Proc. 18th Conf. Embedded Netw. Sensor Syst.*, New York, NY, USA, Nov. 2020, pp. 517–530.

[32] K. Xu, J. Wang, L. Zhang, H. Zhu, and D. Zheng, "Dual-stream contrastive learning for channel state information based human activity recognition," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 1, pp. 329–338, Jan. 2023.

[33] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[34] E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos, "Tensors for data mining and data fusion: Models, applications, and scalable algorithms," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 2, 2017, Art. no. 16.

[35] X. Wang, C. Yang, and S. Mao, "TensorBeat: Tensor decomposition for monitoring multi-person breathing beats with commodity WiFi," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 1, pp. 8.1–8.27, 2018.

[36] N. Yu, W. Wang, A. X. Liu, and L. Kong, "QGesture: Quantifying gesture distance and direction with WiFi signals," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, 2018, Art. no. 51.

[37] W. Cui, B. Li, L. Zhang, and Z. Chen, "Device-free single-user activity recognition using diversified deep ensemble learning," *Appl. Soft Comput.*, vol. 102, 2021, Art. no. 107066.

[38] X. Wang, L. Gao, and S. Mao, "CSI phase fingerprinting for indoor localization with a deep learning approach," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1113–1123, Jun. 2016.

[39] J. Gjengset, J. Xiong, G. McPhillips, and K. Jamieson, "Phaser: Enabling phased array signal processing on commodity WiFi access points," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, S. Lee, A. Sabharwal, and P. Sinha, Eds. ACM, 2014, pp. 153–164.

[40] L. D. Lathauwer, "Blind separation of exponential polynomials and the decomposition of a tensor in Rank-$(l_r, l_r, 1)$ terms," *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 4, pp. 1451–1474, 2011.

[41] A. Cichocki et al., "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Feb. 2015.

[42] Y. Sun and M. Kumar, "A numerical solver for high dimensional transient Fokker-Planck equation in modeling polymeric fluids," *J. Comput. Phys.*, vol. 289, pp. 149–168, 2015.

[43] J. Wang and D. Katabi, "Dude, where's my card?: RFID positioning that works with multipath and non-line of sight," in *Proc. ACM SIGCOMM Conf.*, D. M. Chiu, J. Wang, P. Barford, and S. Seshan, Eds. ACM, 2013, pp. 51–62.

[44] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Thirteenth Int. Conf. Artif. Intell. Statist.*, ser. JMLR Proceedings, Y. W. Teh and D. M. Titterington, Eds., vol. 9. JMLR.org, 2010, pp. 249–256.

[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, Y. Bengio and Y. LeCun, Eds., 2015, doi: 10.48550/arXiv.1412.6980.