A Real-Time Cross-Domain Wi-Fi-based Gesture Recognition System For Digital Twins

Jian Su, Qiankun Mao, Zhenlong Liao, Zhengguo Sheng, Chenxi Huang, Xuedong Zhang*

***Abstract*—The rapid development of Internet of Things has led more realization of digital twins (DT), such as healthcare, smart homes, virtual reality, etc, gesture recognition is a fundamental component of DT. Its implementation can provide users with personalized services or improved human-computer interaction, such as smart home control, in-car interaction, etc, most of existing gesture recognition methods are based on vision or wearable device. However, the vision-based methods face the problem of privacy breach, whereas the wearable-based methods may bring inconvenience to users. With the wide deployment of Wi-Fi networks, lots of consumer devices are widely accessible in people's homes. Motivated by the fact that Wi-Fi signal propagation can be affected by human motion, the opportunity to use Wi-Fi signals for gesture recognition can be further explored. However, the challenge is that the received Wi-Fi signal shows great differences when the same person performs the same gesture in different environments or different person performs the same gesture in the same environment. Therefore, the signal alignment across different domain needs to be solved. In this paper, we propose a gesture recognition system named Phase-Attention-based-Conv-CSI (PAC-CSI), which consists of two modules: data processing and gesture recognition. In the data processing module, we eliminate random phase noise in channel state information (CSI) and perform phase calibration. In the gesture recognition module, we feed the processed phase sequence into a lightweight deep neural network for gesture recognition. PAC-CSI can obtain the gesture category in about 200ms, which can meets the real-time requirements of DT. The gesture recognition accuracy of our proposed system in a single domain is 99.46%, and its performance across new locations, orientations, users, and environments is 98.77%, 98.90%, 97.54%, and 96.47%, respectively.**

***Index Terms*—CSI (Channel State Information), Deep Learning, Digital Twin, Gesture Recognition, Wi-Fi, Wireless Sensing.**

I. Introduction

DIGITAL twins (DT) have more opportunities to be realized thanks to the Internet of Things' rapid development [1]–[3]. DT is a digital representation of the real physical world, which has been applied in a number of applications, for example, a DT framework is proposed for smart manufacturing [4], for a given task and application, the framework can evaluate the suitability of different design solutions, Lu et al. [5] proposed a novel DT-based anomaly detection process flow, realizing continuous anomaly detection, Björnsson et al. [6] built a DT for individual patients to finds the optimal drug, DT has also been explored in the field of smart cities [7], [8]. DT system can be integrated with the gesture recognition system to build digital replicas of humans, obtain the gesture categories performed by humans in real-time, and provide personalized services for users. Building a DT requires sensors to collect data and further process the data into usable information for the DT. Traditional gesture recognition systems are mainly based on visual sensors [9]–[11] or wearable sensors [12]–[14]. But the vision-based methods require Line-of-Sight (LoS) and good lighting conditions, which may not be satisfied in weak light situations, besides, it also has the risk of privacy disclosure. The wearable-based methods bring additional device costs, and they are user-unfriendly due to the inconvenience of wearing the device all the time. In recent years, Wi-Fi communication technology and commercialization have achieved rapid development, researchers have started to study the use of Wi-Fi signals for non-intrusive gesture recognition [15]–[17]. In environments full of Wi-Fi signals, the Wi-Fi signal reaches the receiver through the ground, ceiling, and human body, the Wi-Fi signal propagation path changes when a user performs a gesture, and the receiver can capture the signal changes caused by the user's gesture motion and conduct gesture recognition. Non-intrusive Wi-Fi-based gesture recognition system is suitable for integration with DT, and can avoid the above challenges in vision-based methods and wearable-based methods.

However, Wi-Fi-based gesture recognition suffers from the cross-domain problem: the received Wi-Fi signal pattern is not only affected by gesture motion, but also by different environment and habits of users performing gestures. Therefore the received Wi-Fi signal patterns can be different when different users perform gestures or a same user performs gestures in different environment or different location and different orientation in the same environment, which affects the performance of the gesture recognition system. We define these gesture-independent factors as domain, including user, location, orientation, and environment. Most existing Wi-Fi-based gesture recognition systems are trained and tested in a fixed domain, and they cannot accurately recognize gestures in other domains. Some works have been explored to solve the problem of cross-domain Wi-Fi gesture recognition. For ex-

* Xuedong Zhang is the corresponding author.

J. Su is with the School of Software, Nanjing University of Information Science and Technology, Jiangsu 210044, China (e-mail: sj890718@gmail.com).

Q. Mao is with the School of Computer Science, Nanjing University of Information Science and Technology, Jiangsu 210044, China (e-mail: maoqiankun97@outlook.com).

Z. Liao is with the Hwamei College of Life and Health Sciences, Zhejiang Wanli University, Zhejiang 315100, China (e-mail: liaozhenlong1105@163.com).

Z. Sheng is with the Department of Engineering and Design, University of Sussex, Brighton BN1 9RH, U. K. (e-mail: z.sheng@sussex.ac.uk).

C. Huang is with the School of Informatics, Xiamen University, Fujian 361003, China (e-mail: supermonkeyxi@xmu.edu.cn).

X. Zhang is with the School of Management Science and Engineering, Anhui University of Finance & Economics, Bengbu 233000, China (e-mail: zxd_01@163.com).

ample, authors in [18], [19] use generative adversarial network (GAN) [20] to automatically extract domain-independent features, then use these features for gesture recognition, authors in [21], [22] manually extract domain-independent features from channel state information (CSI) through signal processing methods, and then use neural networks for gesture recognition. However, the recognition accuracy of these existing works is insufficient or the time complexity of signal processing is too high to perform real-time gestures.

Cross-domain gesture recognition still faces three challenges. First, a large number of labeled data samples need to be obtained to train the model, which is labor-intensive and requires specialized knowledge. Second, Wi-Fi signals contain significant ambient noise and phase errors caused by transmitter-receiver hardware, these noises affect the recognition accuracy.

Third, complex data processing methods or huge gesture recognition models will bring too long gesture recognition time, which limit the deployment on resource-constrained devices, and in most application scenarios, the DT system requires low latency between the real world and DT, otherwise, it will affect the quality of service and even cause security risks [23], therefore gesture recognition needs to have low time complexity for real-time recognition. Existing research works can achieve high gesture recognition accuracy [22], [24] or perform recognition with low time complexity [16], but existing works cannot perform gesture recognition with high accuracy in real-time.

We noticed that the propagation path of the Wi-Fi signal will change during the execution of a gesture motion, and the change in the length of the propagation path is reflected in the phase change of the Channel State Information (CSI). Based on this, we propose a cross-domain gesture recognition system: Phase-Attention-based-Conv-CSI (PAC-CSI). PAC-CSI processes CSI phase to remove noise and gesture-independent signal patterns, and use a data augmentation method to augment the training samples, the proposed method then performs gesture recognition with a lightweight attention-based deep neural network. Due to the simple data processing method and a lightweight neural network model, PAC-CSI performs gesture recognition in real-time, which makes it suitable for integration with DT. We use Unity to build a DT system integrated with PAC-CSI, when the user gesture is executed, the DT system can obtain the type of gesture performed in real-time.

The main contributions of this paper are summarized as follows:

- We propose a cross-domain gesture recognition system for DT to address the performance degradation problem of cross-domain gesture recognition. Our system can give gesture classification results in real-time.
- We propose a data augmentation method to augment the training samples, and the experiments show that our data augmentation method can improve the accuracy of gesture recognition.
- We use Unity to build a DT system integrated with our gesture recognition system, the gesture state of human DT in the DT system can be synchronized with real human gestures.

The rest of this paper is organized as follows. Related works are reviewed in Section II. Section III introduces PAC-CSI in details before the performance evaluation in Section IV. Section V concludes the paper and discusses future work.

II. Relate work

This section first introduces the related work of DT, and then introduces the research progress of Wi-Fi-based gesture recognition.

A. Digital Twins

With the development of machine learning, deep learning, and Internet of Things, DT has been researched and applied in the fields of intelligent manufacturing, intelligent building, and intelligent medical treatment. Lee et al. [4] proposed a DT-based framework for smart manufacturing. The framework integrates various DT models, such as product DT, process DT, and resource DT, to create a comprehensive digital representation of the manufacturing system. This framework can improve product quality, shorten delivery times, and increase manufacturing flexibility, but it requires accurate and timely data. Lu et al. [5] proposed a novel DT-based anomaly detection process flow, and achieved continuous anomaly detection of the centrifugal pumps in the HVAC system. Zhou et al. [25] proposes an intelligent small object detection system for digital twin in smart manufacturing with industrial cyber-physical systems. The system uses a combination of computer vision and deep learning techniques to detect small objects in the manufaccturing environment, such as screws or nuts, and integrates the detection results into the DT model of the manufacturing system, therefore, the system can help reduce errors ans increase efficiency in the manufacturing process, but the system need for accurate and reliable object detection. Ren et al. [26] embeds machine learning modules into DT, and built a complete life cycle DT for sophisticated equipment, in the application of diesel locomotives maintenance, the abnormal axle temperature can alarm in advance a week or so. In terms of emergency resource scheduling, Hu et al. [27] focuses on the use of digital twin in the scheduling of hospital emergency resources, the authors propose a system that uses real-time data to generate a DT of the hospital emergency department, which can then be used to simulate different scheduling scenarios and optimize resource allocation. The system can help improve patient outcomes by ensuring that the right resources are available at the right time. [6] builds DT for individual patients, and finds the optimal drug for the patient through computationally treated. Barricelli et al. [28] use wearable sensors to collect the athlete's fitness-related data to build the athlete's DT, and analyze it in the DT system to dynamically predict the health status of athlete and give suggestions. At present, there are few studies on DT-based smart homes, however, with the further development of DT related technology, DT-based smart home applications can greatly improve the user's life happiness and home security level, the gesture recognition is a fundamental component for DT-based smart homes.

B. Wi-Fi-based Gesture Recognition

In recent years, there have been many works on Wi-Fi-based gesture recognition. These works can be divided into two categories: non-cross-domain and cross-domain, the non-cross-domain works perform model training and gesture recognition under one domain setting without considering the cross-domain problem, and the cross-domain works study gesture recognition in cross-domain.

1) Non-Cross-Domain: Although most of the non-cross-domain gesture recognition algorithms suffer from severe performance degradation when the domain changes, these algorithms are still an important fundamental for the research of Wi-Fi gesture recognition algorithms. WiFinger [17] calibrates and denoises the collected CSI, then selects subcarriers that are more sensitive to motion, and uses the multi-dimensional DTW algorithm to compare them with the predefined gestures to obtain the gesture recognition results. WiKey [29] uses Butterworth low-pass filter and PCA to remove the noise in CSI, then uses DWT to extract features from the denoised CSI, and uses KNN algorithm to recognize the user's keystrokes. CSI-Time [30] treats CSI data as multi-dimensional time series data and then uses deep learning methods for activity classification, in addition, CSI-Time utilizes two data augmentation strategies to expand the training samples for improving recognition accuracy. SignFi [16] analyzing the changes in signal strength and phase caused by the movement of the hands and arms, and using a convolutional neural network to classify 276 sign language gestures. In the laboratory environment, SignFi achieves a sign language recognition accuracy of 98.01%. Zhang et al. [31] proposed a neural network consisting of DenseNet and LSTM for activity classification using spectrograms derived from CSI, they augmented data by dropout, adding Gaussian noise, frequency filtering, etc, and achieved around 90% of recognition accuracy on a small-size dataset. ABLSTM [32] proposed an attention-based bi-directional LSTM for human activity recognition, and achieved recognition accuracy high than 95%, they believe that handcrafted features inevitably lose implicit gesture information, so they use the amplitude of CSI without additional processing as the model input.

2) Cross-Domain: CARM [15] uses discrete wavelet transform (DWT) to extract features from the denoised CSI to obtain the duration and frequency of activities, and use the hidden markov model (HMM) for gesture classification, the CARM achieves recognition accuracy of 96%, when the environment changes, in the worst case, the accuracy drops to 72%. WiDar3.0 [21] extracts the BVP feature from the denoised CSI data to represent the velocity coordinate system with the human body as the origin, which is independent of the domain, and then uses a deep neural network consisting of CNN and GRU for gesture classification, Widar3.0 is a one-fits-all gesture recognition model that can adapt to different domains after only one-time training, but its cross-domain recognition accuracy can only reach 82.6%, and the time complexity of BVP extraction algorithm is too high to perform real-time gesture recognition. WiHARAN [18] uses GAN to align features in multiple environments to obtain environment-independent features for gesture classification, WiHARAN can achieve an average cross-environment gesture recognition accuracy of 85.71%. MCBAR [33] uses GAN and semi-supervised learning method to extract environment-independent features for gesture classification based on labeled CSI data in the source domain and unlabeled CSI data in the target domain. CsiGAN [19] also uses GAN to extract user-independent features for cross-user behavior recognition. WiHF [22] extracts a domain-independent arm motion change pattern from CSI, and then uses a dual-task deep neural network for simultaneous gesture recognition and user identification. WiGRUNT [24] denoise the random phase error caused by hardware with CSI-Ratio, then visualization the CSI-Ratio phase value as an image, and design a network for gesture recognition: ResNet with the spatial-temporal attention mechanism, in the result, WiGRUNT achieves gesture recognition accuracy of 96.6%, 93.8%, 93.7% in terms of cross-location, cross-orientation, and cross-environment, respectively. WiGRUNT uses a deep neural network to automatically explore key gesture features distributed in CSI-Ratio for cross-domain gesture recognition, and achieves a higher recognition accuracy than WiDar3.0 and WiHF which use handcrafted domain-independent features for gesture recognition, this indicates that handcrafted features will lose implicit gesture information. This inspires our system design, only the necessary data processing is performed to avoid loss of implicit gesture information, and the neural network is used to automatically extract domain-independent features.

III. System Overview

The PAC-CSI and DT sytem is thoroughly introduced in this section. The system architecture overview of the integration of PAC-CSI and DT is shown in Figure 1. PAC-CSI modify the gesture state of the human DT through the open API of the DT system, and then the DT system can analyzes the state of the human DT according to different service scenarios and performs corresponding processing, for example, in a smart home scenario, the DT system performs home appliance control and modifies the state of the corresponding home appliance DT according to the gesture state of the human DT. In the following subsections, we first introduce the basics of CSI and CSI-Ratio, and then introduce the data preprocessing module and gesture classification module of PAC-CSI in detail. Finally, we introduce our DT system.

A. CSI and CSI-Ratio

1) CSI: Wi-Fi signal propagation from transmitter to receiver is affected by reflection, scattering and fading effects, the combination of which is reflected in CSI. Wi-Fi signals are propagated from the transmitter to the receiver through static paths like LoS paths and wall reflection paths, and dynamic paths caused by human gesture motion. We divide the propagation paths into static paths and dynamic paths, the CSI can be expressed as

$$H(f,t) = H_s(f,t) + \sum_{p \in P} A_p(f,t) e^{-j2\pi \frac{d_p(t)}{\lambda}}, \qquad (1)$$

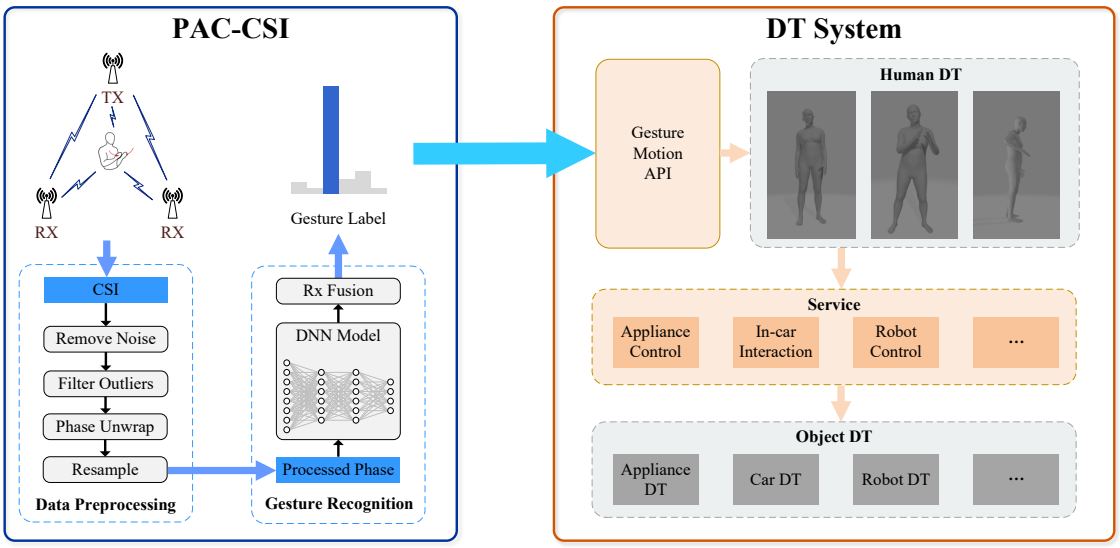


Fig. 1. System overview.

where $H_s(f,t)$ is the CSI component corresponding to the static path, P is the dynamic path set, $A_p(f,t)$ is the CSI amplitude corresponding to the dynamic path p, and $d_p(t)$ is the length of the dynamic path p that changes with time, the meanings of various symbols can be checked in Table IV.

In practice, due to the hardware imperfections of commercial Wi-Fi, the received CSI additionally introduces random phase errors $\epsilon(f,t)$ caused by carrier frequency offset, sampling frequency offset, and timing alignment offset [15]:

$$\bar{H}(f,t) = e^{j\epsilon(f,t)}\left(H_s(f,t) + \sum_{p\in P} A_p(f,t)e^{-j2\pi\frac{d_p(t)}{\lambda}}\right). \quad (2)$$

Due to these time-varying random phase errors, the phase value of the acquired CSI data contains a lot of noise and cannot be used directly. Therefore, we need to eliminate the random hardware phase errors.

2) CSI-Ratio: Since the hardware is shared by multiple antennas connected to the same NIC, the received CSI of different antenna share the same random hardware phase error. We can choose one antenna on the receiver and divide that antenna's CSI using the other antenna's CSI to remove random phase errors. Without loss of generality, considering the case of only one dynamic path [34]:

$$\begin{aligned} R(f,t) &= \frac{\bar{H}_m(f,t)}{\bar{H}_n(f,t)} \\ &= \frac{H_{s,m}(f,t) + A_m(f,t)e^{-j2\pi\frac{d_m(t)}{\lambda}}}{H_{s,n}(f,t) + A_n(f,t)e^{-j2\pi\frac{d_n(t)}{\lambda}}}. \end{aligned} \quad (3)$$

Since the physical distance of different antennas of the same receiver is close, and the human gesture motion in a short time has little change to the propagation path length, the difference of the two reflection path lengths at two close-by antennas can be considered as a constant Δd [35]:

$$R(f,t) = \frac{H_{s,m}(f,t) + A_m(f,t)e^{-j2\pi\frac{d_m(t)}{\lambda}}}{H_{s,n}(f,t) + A_n(f,t)e^{-j2\pi\frac{\Delta d}{\lambda}}e^{-j2\pi\frac{d_m(t)}{\lambda}}}. \quad (4)$$

Let $A = H_{s,m}(f,t)$, $B = A_m(f,t)$, $C = H_{s,n}(f,t)$, $D = A_n(f,t)e^{-j2\pi\Delta d/\lambda}$, $Z = e^{-j2\pi d_m(t)/\lambda}$, then there is

$$R(f,t) = \frac{A + BZ}{C + DZ}, \quad (5)$$

which is in the form of Mobius transformation [34]. Since the Mobius transformation can be decomposed into translation, inversion, similarity, and rotation transformations, these transformations do not change the correlation of the phase with the propagation path. Therefore, we adopt CSI-Ratio to eliminate random phase errors.

B. Data Processing

1) data processing: One CSI packet for Intel 5300 NIC is in form:

$$CSI_t = \begin{bmatrix} H_{1,1} & H_{1,2} & ... & H_{1,30} \\ H_{2,1} & H_{2,2} & ... & H_{2,30} \\ H_{3,1} & H_{3,2} & ... & H_{3,30} \end{bmatrix}, \quad (6)$$

where $H_{i,j}$ represents the CSI value of the j-th subcarrier of the i-th antenna, the Intel 5300 NIC is equipped with three antennas, and each antenna provides CSI data on 30 subcarriers. Generally speaking, gesture perform takes one to several seconds, so the receiver can obtain a CSI sequence related to the gesture:

$$CSI = [CSI_1, CSI_2, ..., CSI_T] \in C^{T\times 3\times 30}. \quad (7)$$

For the three antennas of the receiver, we choose the antenna that is least sensitive to gesture as the reference antenna

[36]. Specially, we calculate the amplitude-variance ratios of all subcarriers of all antennas, choose the antenna with the smallest sum of the amplitude-variance ratios of all subcarriers as the reference antenna, and then divide the CSI data of the other two antennas by the CSI data of the reference antenna to get the CSI-Ratio:

$$CR = [CR_1, CR_2, ..., CR_T] \in C^{T\times 2\times 30}. \tag{8}$$

Due to the different physical locations of different antennas on the same receiver, the data of two antennas can depict the impact of gestures on the signal from different perspectives, we treat the data of the two antennas as two different samples, furthermore, this approach makes our method applicable to receivers with only two antennas.

$$CR^i = [CR_1^i, CR_2^i, ..., CR_T^i] \in C^{T\times 30}, \tag{9}$$

where i indicates index of antennas, then extract the phase P of CSI-Ratio, for brevity, the antenna index superscript are omitted below.

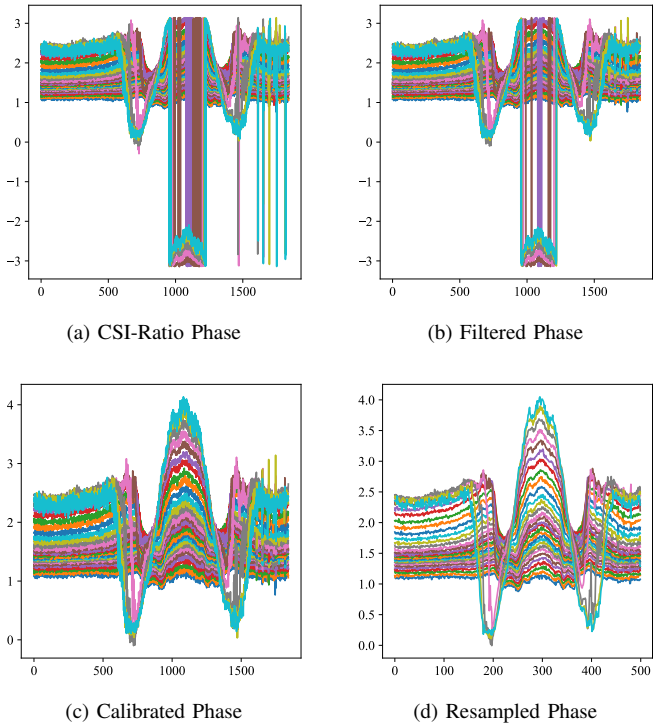
$$\begin{aligned} P &= angle(CR) \\ &= [angle(CR_1), angle(CR_2), ...angle(CR_T)] \\ &= [P_1, P_2, ..., P_T] \in R^{T\times 30}. \end{aligned} \tag{10}$$

The original CSI-Ratio phase sequences are shown in Figure 2a, it can be seen that these phase sequences contain outliers and phase wrapping problems due to measurement limitations. This will affect the recognition accuracy of the neural network, so it needs to be processed. We first use a Hampel filter with a window size of 3 to remove outliers, the filtering results are shown in Figure 2b. The phase calibration is then performed using the phase calibration algorithm, for phase data at each period, if the absolute value of the phase difference with the previous period is greater than $2\pi - \epsilon$, we add or subtract the phase data with 2π until the absolute value of the phase difference is within $2\pi - \epsilon$, where ϵ is empirically set to 0.3, phase calibration resulting in Figure 2c, it should be noted that due to the noise of CSI-Ratio, our phase calibration algorithm cannot be effective for all samples, but it's simple and efficient, and handles most cases, the experimental results also show that our phase calibration algorithm is effective, see section IV for details. At the sampling rate of 1000Hz, the phase sequence length of most gesture samples is more than 1500, which slows down the training and inference time of the deep neural network model and takes up a lot of memory. And the time for the user to perform the gesture is not constant, that is, different data samples have different sequence lengths, which brings inconvenience to the processing of the deep neural network. A common solution for different sequence lengths of data samples is to zero-pad each batch of sequence samples to the same length, however, the zero-pad strategy may affect the recognition ability of the model. We solve the above problem by resample to unify the sequence length of all data samples to 500:

$$\tilde{P} = resample(P) = [\tilde{P}_1, \tilde{P}_2, ..., \tilde{P}_{500}] \in R^{500\times 30}, \tag{11}$$

the results are shown in Figure 2d. This solution reduces the computational complexity of deep neural network, and solves the problem of inconsistent sequences lengths of samples, and further eliminates high-frequency noise. Experiments show that the resample can reduce inference time consumption while ensuring the accuracy of gesture recognition.

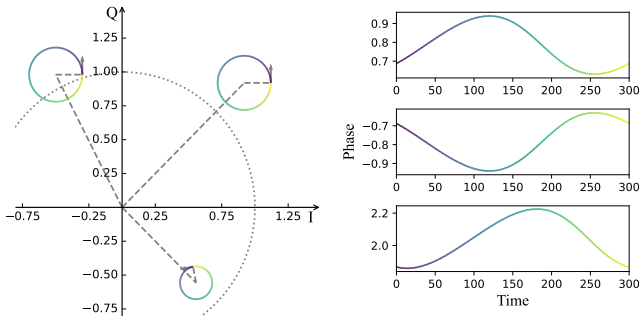
2) data augmentation: As shown in equation 5, CSI-Ratio is in the form of Mobius transformation. Assumed that the CSI-ratio dynamic components caused by gesture motion are fixed, different domain settings or transformations will bring different CSI-Ratio static components, and the inversion transformation will change both the static component of the CSI-Ratio and the trajectory direction of the CSI-Ratio on the complex plane. These factors will change the phase pattern of CSI-Ratio. As shown in Figure 3, Figure 3a represents signal trajectory of the dynamic component $0.2 \cdot e^{j\frac{t}{150}\pi}; t \in [0, 300)$ with different static component components or after inversion transformation, specifically, the CSI-Ratio trajectory in the first quadrant is the dynamic component with static component $1.3 \cdot e^{j\frac{1}{4}\pi}$, in the second quadrant is the dynamic component with static component $1.1 \cdot e^{j\frac{13}{20}\pi}$, in the fourth quadrant is the inversion transformation of the CSI-Ratio trajectory in the first quadrant. Figure 3b upper, middle, lower are the phases of CSI-Ratio trajectory in the first, second, and fourth quadrants of the complex plane, respectively. We noticed that the phase change trend of the same dynamic component is different with different static components or after inversion transformation, but it is roughly the vertical or horizontal flip of the original



(a) CSI-Ratio Phase (b) Filtered Phase

(c) Calibrated Phase (d) Resampled Phase

Fig. 2. Data processing result.



(a) CSI-Ratio Trajectory (b) Phase

Fig. 3. Phase results of a dynamic component with different static components.

phase change trend. Based on this, we augment the training dataset by taking the opposite number or reversing along the time dimension or both of two for the phase sequence of the training samples.

C. Gesture Recognition

PAC-CSI uses a lightweight neural network model to extract features of phase sequences and perform gesture recognition. Specifically, we first normalize the phase sequence samples to remove gesture-uncorrelated phase scale gaps caused by different static components or Mobius transforms. Then we extract features from the normalized phase sequence with an attention-based one-dimensional convolutional neural network, the extracted features are then feed into the FC layer to get gesture recognition results.

1) Phase Sequences Normalization: As shown in Figure 3, the same dynamic path length change has different phase scales in different static components or after different transformations. We normalize the phase sequence, which makes the phase values of all samples scaled between -1 and 1. After normalization, the neural network can better focus on the trend of phase change rather than the scale of phase.

$$P_N^i = 2 \cdot \frac{\tilde{P}^i - min(\tilde{P}^i)}{max(\tilde{P}^i) - min(\tilde{P}^i)} - 1, \tag{12}$$

where i represents the subcarrier indicate, P_N is the normalized phase sequence.

2) Network Architecture: Heavyweight deep neural network models will consume more computing resources, slow down the inference time of the model, and limit the model deployment on resource-constrained devices. The neural network model used in PAC-CSI is consists of one-dimensional convolutional layer and lightweight attention layer, which enables our system to perform real-time gesture recognition. The network structure is shown in Figure 4, network is consisting of Conv1D Blocks and Attention Blocks. As shown in Figure 4a, Conv1D Block consists of a conv1d layer with stride 1 and a conv1d layer with stride 2, all conv1d layers are followed by a BatchNorm layer [37] and a ReLU layer. Our Attention Block design is inspired by CBAM [38], CBAM is a lightweight attention module for Conv2D, the additional computational overhead caused by adding the CBAM attention module is small. Therefore our Attention Block is designed as a 1D convolutional version of CBAM, as shown in Figure 4b.

D. DT System

We use Unity to build a DT system. Specifically, we build a scene and human DT in the DT system, and integrate the DT system with the gesture recognition system so that the gesture motion of the human DT is synchronized with the real-world human gesture motion in real-time.

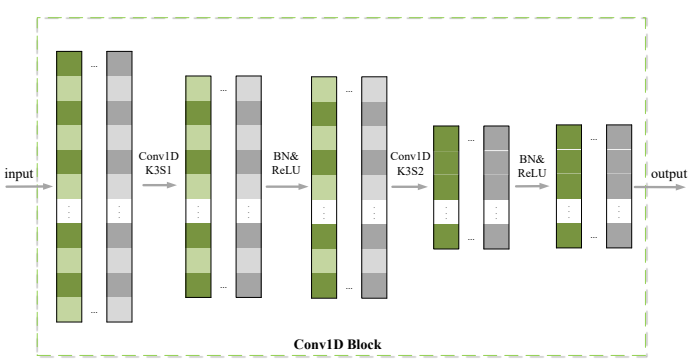
1) Human Gesture Modeling: Thanks to the development of deep learning in recent years, we can automatically model human gesture motions using deep learning methods. We model human gestures using [39], this work can automatically generate 3D models of human motion from text, which can greatly reduce the workload of human gesture modeling. We take gesture *Clap* and *Sweep* as examples, and the modeling results are shown in Figure 5.

2) Scene Modeling: Unity has a wealth of materials available, which can greatly reduce the difficulty of development, and its scalability allows us to easily integrate it with the gesture recognition system, so we use Unity as our DT system platform. We built a simple scene using Unity, and introduced the modeled human DT into the digital scene. By integrating with PAC-CSI, the human DT gesture motion in the scene is synchronized with the real human gesture motion, as shown in Figure 6.

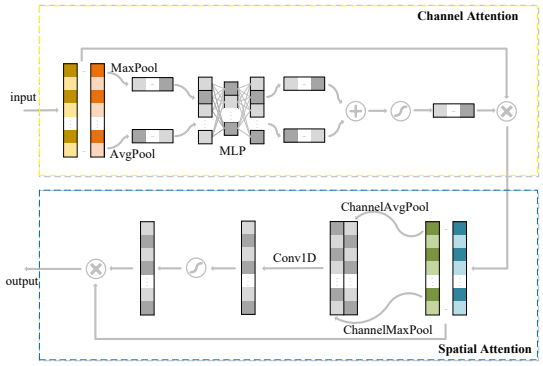
IV. Experiment and Evaluation

This section presents the experiment implementation details and evaluation the experiment results.

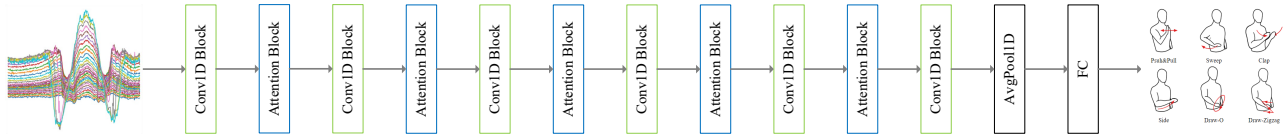
a) Dataset: Widar3 [21] published a dataset for Wi-Fi cross-domain gesture recognition research, it contains data on gestures performed by multiple users in different location, orientation, environment. We choose nine gestures on the Widar3 dataset for experiments, as shown in Figure 7. We designed two datasets: Base and Extend. The Base dataset contains six gesture classes, while the Extend dataset contains nine gesture classes. The selection of data samples is shown in Table I. We conduct experiments on the Extend dataset to verify the performance of PAC-CSI on more gesture classes, while the remaining experiments were conduct on the Base dataset. Specifically, we conduct five experiments, in each experiment, the data of one instance were divided into a verification set and a test set with a ratio of 1:4, and the data of the remaining four instances were used as a training set, the experimental result is averaged over five experiments. We perform cross-location, cross-orientation and cross-user experiments on environment 1 with similar data settings. In the cross-environment experiment, the data of one environment were divided into a verification set and a test set with a ratio of 1:4, and the data of the remaining environment were used as a training set, the average result of the experiments was taken as the final result.



(a) Conv1D Block

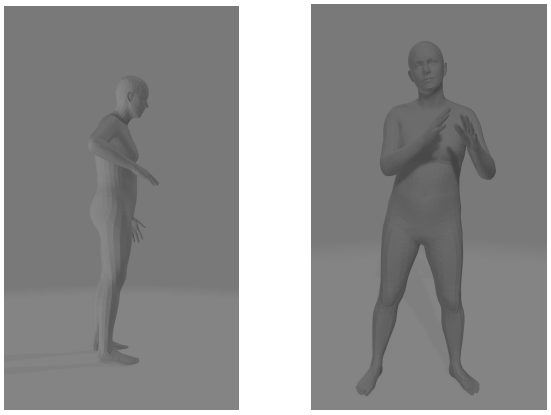


(b) Attention Block



(c) Network Architecture

Fig. 4. Network structure.



(a) Sweep (b) Clap

Fig. 5. Human gesture modeling.

Fig. 6. Scene modeling with Unity.

TABLE I
DATA SAMPLES

Dataset	Env	User	Location	Orientation	Instance
Base	1	5,10,11,12, 13,14,15,16	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5
	2	1,2,3,6	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5
	3	3,7,8,9	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5
Extend	1	13,14,15,16,17	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5
	2	1,2,3	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5

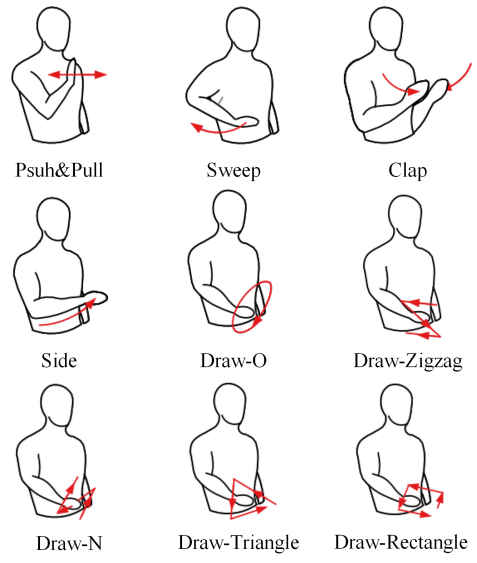


Fig. 7. Gesture labels.

b) Implement detail: We use Adam optimizer with an initial learning rate of 10^{-4} in training, and adjust the learning rate with a Cosine Annealing strategy [40]. We set the batch size to 128, and train for 10000 iterations, and the adjust learning rate every 50 iterations, the learning rate is shown in Figure 8. All experiments are performed on a PC

equipped with Intel 12600K CPU and RTX 3060 GPU. In the scenario of multi-receiver deployment, when the user performs a gesture, multiple data samples are acquired in the corresponding receiver, and multiple receivers provide gesture information from different perspectives, which can improve the accuracy of gesture recognition. Most of the existing work [21], [24] combines the data samples of multiple receivers, and obtains a feature containing the information of multiple receivers to input into the network. Although the information of multiple receivers can also be used in this way, it cannot adapt to the settings of different number of receivers. We use a multi-receiver fusion mechanism to enable our model to adapt quickly to environments with different number of receivers. Specifically, in the training process, we separate multiple data samples from multiple receivers, and in the testing process, we comprehensively consider the classification results of all receivers to obtain the final classification result, specifically, the network model obtains a gesture probability distribution based on the data samples of each receiver, and we take the average of the gesture probability distributions of all receivers, the gesture label with the highest probability in the average probability distribution is used as the final gesture recognition result.

A. Overall Result

The overall result of PAC-CSI is shown in Figure 9. The PAC-CSI achieved recognition accuracy of 99.46% for in-domain, and in the case of cross-domain, PAC-CSI can also get recognition accuracy of 98.77%, 98.90%, 97.54%, 96.47% in terms of cross-location, cross-orientation, cross-user, and cross-environment, respectively, The experiment results demonstrated the cross-domain recognition ability of PAC-CSI. It can be seen from the confusion matrix that *Clap* gestures and *Slide* gestures are easily misclassified when performing cross-domain gesture recognition, since the right hand in the *Clap* gesture follows essentially the same trajectory as the *Slide* gesture.

B. Comparative Study

We compared our PAC-CSI with existing methods [21], [22], [24], the result is shown in Table II. The performance of PAC-CSI is slightly lower than that of WiGRUNT in in-domain experiment, but our PAC-CSI achieves the highest recognition accuracy in all cross-domain experiments. The recognition accuracy of PAC-CSI outperforms the current best solutions by 2.15%, 5.05%, 1.89%, and 2.74% in terms of cross-location, cross-orientation, cross-user, cross-env, respectively. The experimental results on the Extend dataset also demonstrate that PAC-CSI can maintain a high recognition accuracy even with more gesture classes. Since SignFi [16] uses amplitude which is highly correlated with domain factors as gesture features, it can be seen that when the domain changes, the recognition accuracy of SignFi is greatly affected, especially when the environment changes, its recognition accuracy drops severely. WiHF [22] simultaneously recognizes gesture categories and user identification, while PAC-CSI does not have this function. And the superiority of PAC-CSI is not only in the recognition accuracy, but also in the computational efficiency. As shown in Table III, the total time-consuming of PAC-CSI is lower than all comparison methods because of the simple data processing flow and lightweight network structure. Although removing the resampling operation makes the data processing time slightly reduced, the excessively long data length greatly affects the model reasoning time. The overall time consumption has increased by more than 20%. It can be seen that after removing the attention module, the time consumption of gesture recognition is only reduced a little, which shows that the improvement of network complexity by our attention module is negligible. The data processing time consumption of Widar3 is extremely high, and each sample requires 126.523s, which is not enough to support real-time gesture recognition. In the inferring of neural networks, the lightweight network structure makes the inferring of PAC-CSI more efficient than other methods. Overall, PAC-CSI outperforms existing methods in real-time gesture recognition.

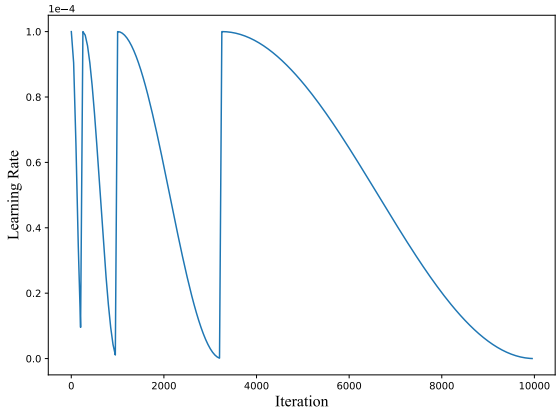


Fig. 8. Learnng rate.

TABLE II
RESULT COMPARE TO EXIST METHODS

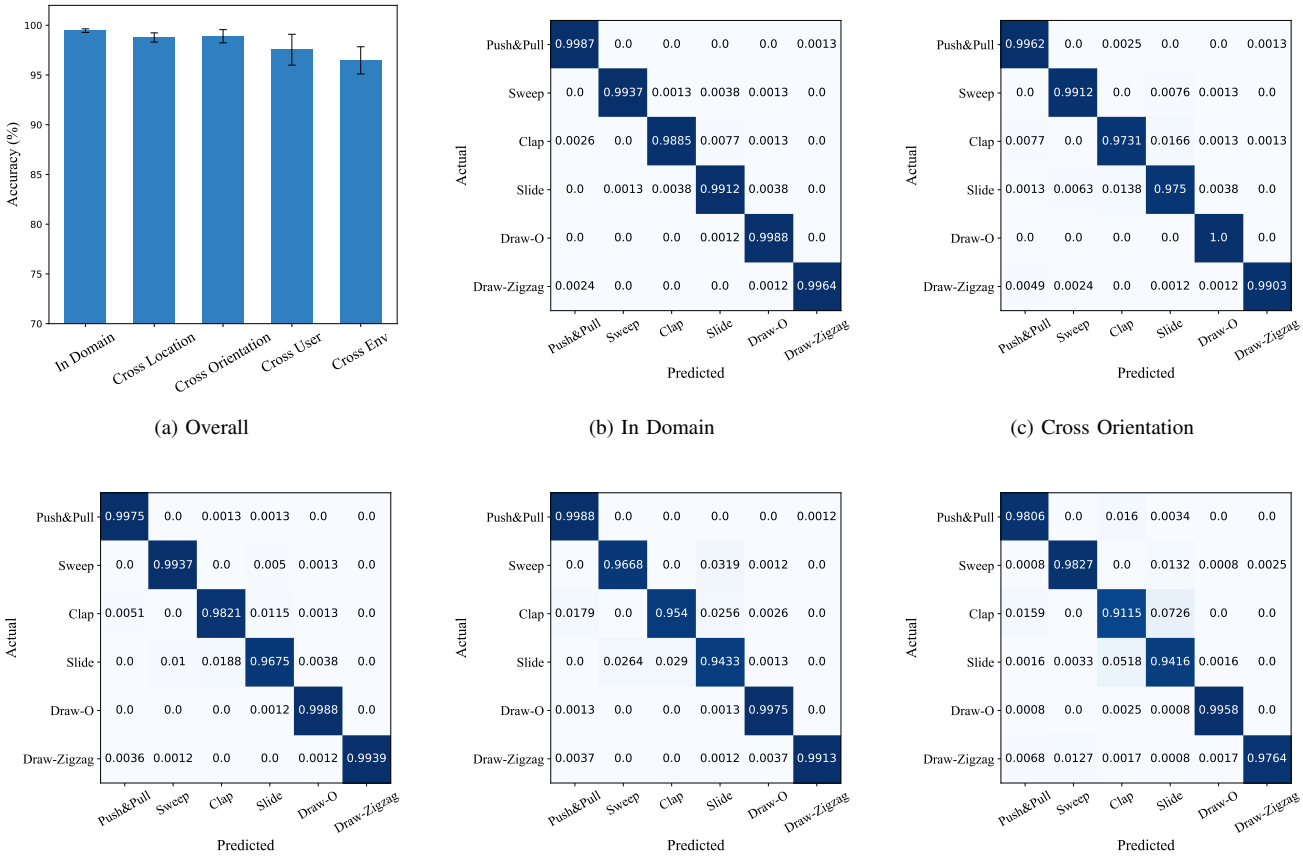
	In Domain	Cross Location	Cross Orientation	Cross User	Cross Env
PAC-CSI	99.46%	**98.77%**	**98.90%**	**97.54%**	**96.47%**
PAC-CSI-Extend	97.40%	96.81%	94.89%	93.15%	92.03%
SignFi	94.22%	81.29%	73.32%	82.66%	54.43%
Widar3	92.7%	89.7%	82.6%	88.9%	92.4%
WiHF	97.65%	92.07%	81.89%	-[1]	91.07%
WiGRUNT	**99.71%**	96.62%	93.85%	95.65%	93.73%

[1] WiHF performs user identification simultaneously, so there are no cross-user result.

TABLE III
THE TIME CONSUMPTION OF DATA PROCESS AND GESTURE RECOGNITION

Method	Data Processing	Gesture Recognition	Total
PAC-CSI	0.173s	0.044s	0.217s
PAC-CSI-No-Att[1]	0.173s	0.029s	0.202s
PAC-CSI-No-Resample[2]	0.147s	0.116s	0.263s
SignFi	0.112s	0.017s	0.129s
Widar3	126.523s	0.02s	126.543s
WiHF	1.128s	0.051s	1.179s
WiGRUNT	0.425s	0.185s	0.61s

[1] PAC-CSI-No-Att refers to the network with the attention block removed.
[2] PAC-CSI-No-Resample refers to the removal of the resample step from the data processing.



(a) Overall (b) In Domain (c) Cross Orientation

(d) Cross Location (e) Cross User (f) Cross Env

Fig. 9. Experiment result.

C. Impact of Data Processing

We conduct experiments to illustrate the impact of data processing on recognition accuracy. As shown in Fig 10, where 'w/o csi ratio', 'w/o calibration', 'w/o resample' means remove CSI-Ratio, phase calibration, and resample in data processing step, respectively, when the CSI-Ratio step is removed, the recognition accuracy drops significantly, this is because gesture information is swamped by non-negligible hardware random phase noise. Phase calibration improves the recognition accuracy from 83.51% to 96.47% in cross-environment experiments, demonstrating the effectiveness of our phase calibration algorithm. And it can be seen from the figure that the resample operation not only maintain the recognition accuracy, but also slightly improves the recognition accuracy because it avoids the impact of too many zero-pads on the model.

We conduct experiments in in-domain scenario to illustrate the effectiveness of the data augmentation method. Specifically, we randomly select 1/4, 1/2, 3/4 and all of the training data for network training, and perform training with and without data augmentation method, respectively. The experimental results are shown in Figure 11. In the experiment of using 1/4 of the data samples without data enhancement, the recognition accuracy is only 74.81%, and after data augmentation, the recognition accuracy improved 21.82%. When using 1/2, 3/4 and all data samples, the improvement in recognition accuracy of data augmentation decreases successively. Our data augmentation method greatly improves the recognition accuracy when the training samples is insufficient. When the training samples are sufficient, the performance improvement brought by the data augmentation method is relatively small.

D. Impact of Data Acquisition

Although different antennas on one receiver provide gesture information from different perspectives due to differences in their physical locations, the multi-perspectives gesture information that only one receiver can provide is limited due to the closing physical distance. Then multiple receivers can capture the influence of gestures on the signal from more perspectives to improve the accuracy of recognition. We conduct experiments to compare the impact of different number of receivers on the accuracy of recognition. The experimental results are shown in Figure 12. In a scenario with only one receiver, the accuracy of in-domain gesture recognition can only reach 86.76%. This is because in the case of only one receiver, when gestures are performed at certain locations or orientations, the impact of the gesture motion on the signal cannot be well captured by the receiver due to the occlusion of the user's body. Multiple receivers in different locations solve this problem and provide different perspectives on the

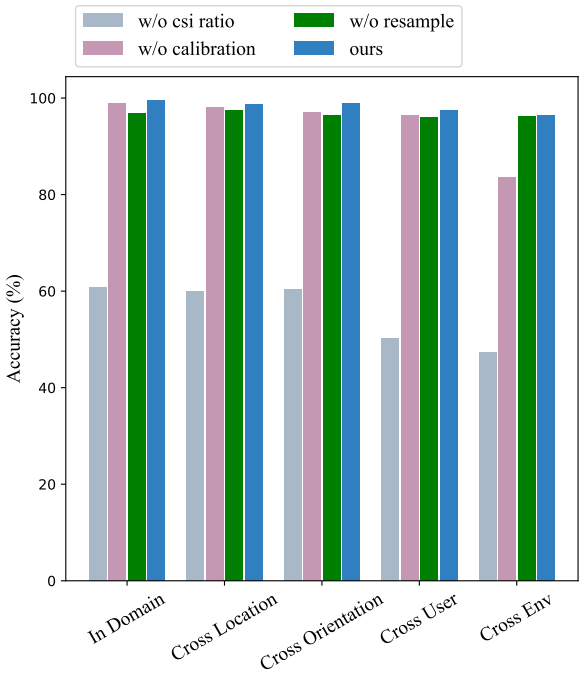


Fig. 10. Impact of data processing.

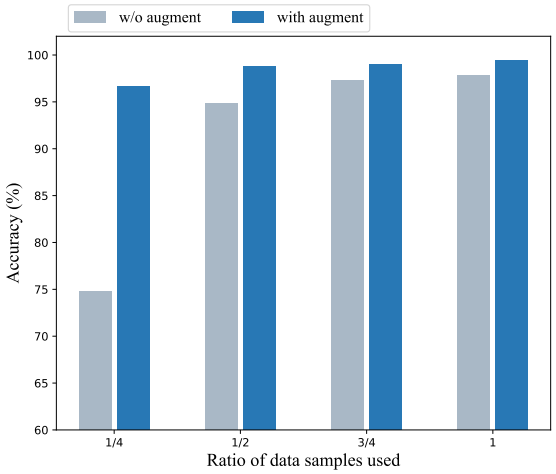


Fig. 11. Impact of data augmentation.

impact of gestures on the signal, resulting in higher recognition accuracy.

The difference in data sample rate will also affect the accuracy of recognition. We conduct experiments to compare the recognition accuracy with different sample rates. For all sample rate settings, we use the same data processing method, which is to unify the sequence length of all data samples to 500. The experimental results are shown in Figure 13. It can be seen that at a sample rate of 50Hz, the insufficient sample rate leads to the loss of gesture details, and ultimately the recognition accuracy is affected, while the sample rate above 100Hz is sufficient to capture the details of daily gestures, and higher sample rate has very little improvement in recognition accuracy.

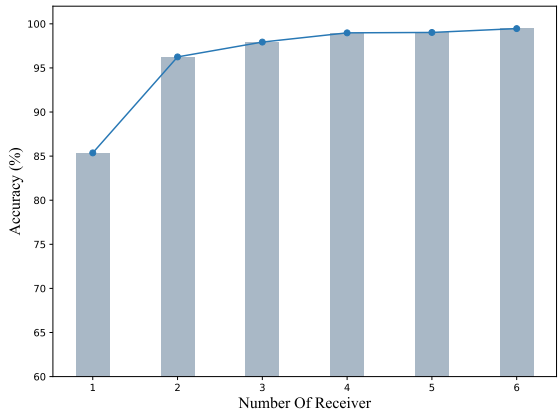


Fig. 12. Impact of receivers number.

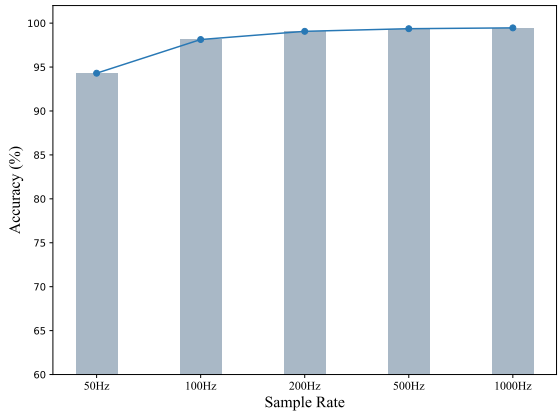


Fig. 13. Impact of sample rate.

E. Impact of Network And Hyperparameter

We conduct experiments to investigate the impact of network structure and hyperparameters on the results. As shown in Figure 14, compared to the network without attention, channel attention and spatial attention brought different degrees of improvement, since the temporal dimension is more relevant to gestures than information from different subcarriers, spatial attention brought greater improvement in accuracy. Our attention block combines two attention mechanisms, and the results were significantly better than those of networks without attention or with only one attention mechanism. The effects of different hyperparameters on the experimental results are shown in Figure 15. The learning rate adjustment strategy adjusts the learning rate to make the model converge better, so we used a learning rate adjustment strategy during training. For different learning rates, when the initial learning rate was 10^{-5}, the convergence was too slow, and when the initial learning rate was 10^{-3} or 10^{-2}, the convergence process was unstable, and when without Cosine Annealing strategy, the model is difficult to converge to the most optimum, therefore, we chose an initial learning rate of 10^{-4}. Regarding batch size, when the batch size was 64 or 32, the model had not yet converged at the end of training, while the results were similar for a batch size of 256 and 128. Considering training

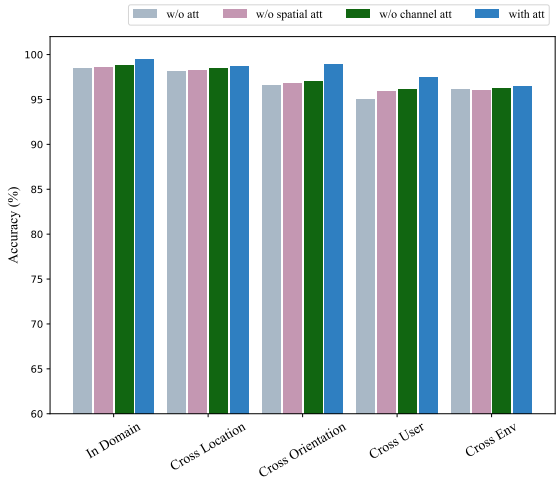


Fig. 14. Impact of attention block.

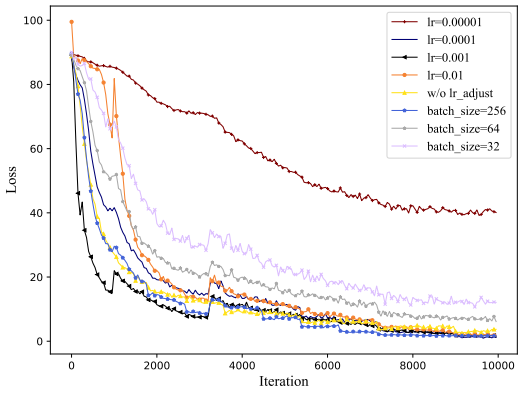


Fig. 15. Impact of hyperparameters.

time and model performance, we selected a batch size of 128.

V. Conclusion

In conclusion, we have introduced a cross-domain gesture recognition system for DT. Our system effectively addresses the issue of random hardware phase errors in CSI by leveraging the CSI-Ratio. By applying the CSI-Ratio, we have successfully eliminated these random phase errors, resulting in improved accuracy of the CSI. Furthermore, we have rectified phase value errors using a phase calibration algorithm, which significantly enhances the precision and reliability of the phase sequences. The integration of an attention-based multi-layer one-dimensional convolutional network in our system has enabled the achievement of robust and accurate gesture recognition. Through experiments conducted on the open dataset widar3, our system achieved an impressive accuracy of 99.46%. Furthermore, our system demonstrated remarkable accuracy in cross-location, cross-orientation, cross-user, and cross-environment scenarios, with gesture recognition accuracies of 98.77%, 98.90%, 97.54%, and 96.47%, respectively. These results underscored the substantial advancements our system brought compared to existing methods, solidifying its effectiveness and practical applicability. Moreover, our gesture recognition system achieved low recognition times, successfully meeting the real-time demands of DT systems. It should be noted that our work is based on manually segmented gesture data sets, and does not have user recognition capabilities. We believe that in the daily application of gesture recognition, gesture data can be automatically segmented to achieve gesture recognition including user authentication is very important. In future work, we will explore the potential for further applications of gesture recognition and user identification through automatic data segmentation in DT systems.

TABLE IV
Symbols

Indices	
f	index of subcarrier ($f \in \{1, ..., 30\}$)
t	index of period ($t \in \{1, ..., T\}$)
Parameters	
λ	subcarrier wavelength
$H(f,t)$	CSI of subcarrier f in period t
$H_{a,f}$	CSI of antenna a and subcarrier f
CSI_t	CSI in period t
CR_t	CSI-Ratio in period t
CR_t^a	CSI-Ratio of antenna a in period t
P_t	phase of CSI-Ratio in period t
$\tilde{P}_t$	resampled phase of CSI-Ratio in period t
$\tilde{P}^f$	resampled phase of subcarrier f

References

[1] A. Ghosh, D. Chakraborty, and A. Law, "Artificial intelligence in internet of things," *CAAI Transactions on Intelligence Technology*, vol. 3, no. 4, pp. 208–218, 2018.
[2] M. A. Jamshed, K. Ali, Q. H. Abbasi, M. A. Imran, and M. Ur-Rehman, "Challenges, applications and future of wireless sensors in internet of things: A review," *IEEE Sensors Journal*, 2022.
[3] A. Ghasempour, "Internet of things in smart grid: Architecture, applications, services, key technologies, and challenges," *Inventions*, vol. 4, no. 1, p. 22, 2019.
[4] J. Lee, M. Azamfar, and B. Bagheri, "A unified digital twin framework for shop floor design in industry 4.0 manufacturing systems," *Manufacturing Letters*, vol. 27, pp. 87–91, 2021.
[5] Q. Lu, X. Xie, A. K. Parlikad, and J. M. Schooling, "Digital twin-enabled anomaly detection for built asset monitoring in operation and maintenance," *Automation in Construction*, vol. 118, p. 103277, 2020.
[6] B. Björnsson, C. Borrebaeck, N. Elander, T. Gasslander, D. R. Gawel, M. Gustafsson, R. Jörnsten, E. J. Lee, X. Li, S. Lilja *et al.*, "Digital twins to personalize medicine," *Genome medicine*, vol. 12, no. 1, pp. 1–4, 2020.
[7] T. Ruohomäki, E. Airaksinen, P. Huuska, O. Kesäniemi, M. Martikka, and J. Suomisto, "Smart city platform enabling digital twin," in *2018 International Conference on Intelligent Systems (IS)*. IEEE, 2018, pp. 155–161.
[8] Y.-C. Lin and W.-F. Cheung, "Developing wsn/bim-based environmental monitoring management system for parking garages in smart cities," *Journal of Management in Engineering*, vol. 36, no. 3, p. 04020012, 2020.
[9] T. Li, Q. Liu, and X. Zhou, "Practical human sensing in the light," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, 2016, pp. 71–84.
[10] M. Wang, B. Ni, and X. Yang, "Recurrent modeling of interaction context for collective activity recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3048–3056.
[11] G. Gkioxari, R. Girshick, P. Dollár, and K. He, "Detecting and recognizing human-object interactions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8359–8367.

[12] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, pp. 1–33, 2014.
[13] S. Shen, H. Wang, and R. Roy Choudhury, "I am a smartwatch and i can track my user's arm," in *Proceedings of the 14th annual international conference on Mobile systems, applications, and services*, 2016, pp. 85–96.
[14] Y. Guan and T. Plötz, "Ensembles of deep lstm learners for activity recognition using wearables," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, pp. 1–28, 2017.
[15] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial wifi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, 2017.
[16] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "Signfi: Sign language recognition using wifi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–21, 2018.
[17] S. Tan and J. Yang, "Wifinger: Leveraging commodity wifi for fine-grained finger gesture recognition," in *Proceedings of the 17th ACM international symposium on mobile ad hoc networking and computing*, 2016, pp. 201–210.
[18] Z. Wang, S. Chen, W. Yang, and Y. Xu, "Environment-independent wi-fi human activity recognition with adversarial network," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 3330–3334.
[19] C. Xiao, D. Han, Y. Ma, and Z. Qin, "Csigan: Robust channel state information-based activity recognition with gans," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 191–10 204, 2019.
[20] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
[21] Y. Zhang, Y. Zheng, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Widar3. 0: Zero-effort cross-domain gesture recognition with wi-fi," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
[22] C. L. Li, M. Liu, and Z. Cao, "Wihf: Gesture and user recognition with wifi," *IEEE Transactions on Mobile Computing*, 2020.
[23] S. Mihai, M. Yaqoob, D. V. Hung, W. Davis, P. Towakel, M. Raza, M. Karamanoglu, B. Barn, D. Shetve, R. V. Prasad *et al.*, "Digital twins: a survey on enabling technologies, challenges, trends and future prospects," *IEEE Communications Surveys & Tutorials*, 2022.
[24] Y. Gu, X. Zhang, Y. Wang, M. Wang, H. Yan, Y. Ji, Z. Liu, J. Li, and M. Dong, "Wigrunt: Wifi-enabled gesture recognition using dual-attention network," *IEEE Transactions on Human-Machine Systems*, 2022.
[25] X. Zhou, X. Xu, W. Liang, Z. Zeng, S. Shimizu, L. T. Yang, and Q. Jin, "Intelligent small object detection for digital twin in smart manufacturing with industrial cyber-physical systems," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 1377–1386, 2021.
[26] Z. Ren, J. Wan, and P. Deng, "Machine-learning-driven digital twin for lifecycle management of complex equipment," *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 1, pp. 9–22, 2022.
[27] X. Hu, H. Cao, J. Shi, Y. Dai, and W. Dai, "Study of hospital emergency resource scheduling based on digital twin technology," in *2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, vol. 2. IEEE, 2021, pp. 1059–1063.
[28] B. R. Barricelli, E. Casiraghi, J. Gliozzo, A. Petrini, and S. Valtolina, "Human digital twin for fitness management," *Ieee Access*, vol. 8, pp. 26 637–26 664, 2020.
[29] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Recognizing keystrokes using wifi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1175–1190, 2017.
[30] S. K. Yadav, S. Sai, A. Gundewar, H. Rathore, K. Tiwari, H. M. Pandey, and M. Mathur, "Csitime: Privacy-preserving human activity recognition using wifi channel state information," *Neural Networks*, vol. 146, pp. 11–21, 2022.
[31] J. Zhang, F. Wu, B. Wei, Q. Zhang, H. Huang, S. W. Shah, and J. Cheng, "Data augmentation and dense-lstm for human activity recognition using wifi signal," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4628–4641, 2020.
[32] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "Wifi csi based passive human activity recognition using attention based blstm," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2714–2724, 2018.
[33] D. Wang, J. Yang, W. Cui, L. Xie, and S. Sun, "Multimodal csi-based human activity recognition using gans," *IEEE Internet of Things Journal*, vol. 8, no. 24, pp. 17 345–17 355, 2021.
[34] 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," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–26, 2019.
[35] Y. Zeng, D. Wu, R. Gao, T. Gu, and D. Zhang, "Fullbreathe: Full human respiration detection exploiting complementarity of csi phase and amplitude of wifi signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–19, 2018.
[36] K. Qian, C. Wu, Z. Zhou, Y. Zheng, Z. Yang, and Y. Liu, "Inferring motion direction using commodity wi-fi for interactive exergames," in *Proceedings of the 2017 CHI conference on human factors in computing systems*, 2017, pp. 1961–1972.
[37] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.
[38] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.
[39] G. Tevet, S. Raab, B. Gordon, Y. Shafir, A. H. Bermano, and D. Cohen-Or, "Human motion diffusion model," *arXiv preprint arXiv:2209.14916*, 2022.
[40] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.

Jian Su has been an associate professor in the School of Software at the Nanjing University of Information Science and Technology since 2017. He received his PhD with distinction in communication and information systems at University of Electronic Science and Technology of China (UESTC) in 2016. He holds a B.S. in Electronic and information engineering from Hankou university and an M.S. in electronic circuit and system from Central China Normal University. His current research interests cover Internet of Things, RFID, and Wireless sensors networking. He is a member of IEEE and a member of ACM.

Qiankun Mao received the B.S. degree in School of Software, East China Jiao Tong University. Since September 2021, he has been working toward the M.S. degree with the School of Computer Science, Nanjing University of Information Science and Technology. His research interests include wireless sensing, human activity recognition, and deep leaning.

Zhenlong Liao received his B.S. degree in Internet of Things Engineering from Binjiang College of Nanjing University of Information Engineering in 2020, M.S. degree in Software Engineering at the School of Software, Nanjing University of Information Science and Technology in 2023. His research interests include WiFi-based human activity sensing, Internet of Things. He is current with the Hwamei College of Life and Health Sciences, Zhejiang Wanli University.

Zhengguo Sheng has been a senior lecturer in the Department of Engineering and Design at the University of Sussex since 2015. He received his Ph.D. and M.S. with distinction at Imperial College London in 2011 and 2007, respectively, and his B.Sc. from the University of Electronic Science and Technology of China (UESTC) in 2006. His current research interests cover the Internet of Things (IoT), connected vehicles, and cloud/ edge computing.

Chenxi Huang s currently an Assistant Professor with the School of Informatics, Xiamen University. His research interests include image processing, image reconstruction, data fusion, 3D visualization, and machine learning. He serves as an Associate Editor for Journal of Medical Imaging and Health Informatics (SCIE) and Frontiers in Medical Technology. He is also a reviewer for IEEE Access, Neurocomputing, Peerj, Journal of Grid computing, IEEE journal of biomedical imaging and health informatics, IEEE Transactions on Emerging Topics in Computational Intelligence, Journal of Medical Imaging and Health Informatics, and other SCI journals. He has published many high-level academic papers in related research fields. Recently, he has published 10+ SCI journal papers as the first author or corresponding author in ACM Transactions on Multimedia Computing, Communications, and Applications, IEEE Transactions on Instrumentation and Measurement, Complexity and Frontiers in Neuroscience.

Xuedong Zhang received his M.S. degree in Computer Applications from Suzhou University in 2004, and has been working in the School of Management Science and Engineering of Anhui University of Finance & Economics since 2004. His research interests are intelligent algorithms, Internet of Things.