RF-CGR: Enable Chinese Character Gesture Recognition With RFID

Zhixiong Yang, Ziyi Zhen, Zijian Li, Xu Liu, Bo Yuan, and Yajun Zhang

***Abstract*— Gesture recognition is the basis of human–computer interaction (HCI), and radio frequency identification (RFID) has attracted wide attention for its advantages of lightweight, low cost, and universality. We propose a Chinese gesture recognition system based on RFID (RF-CGR), which implements cross-domain Chinese character gesture recognition intuitively and effectively. The fundamental concept of RF-CGR is to extract hand motion patterns from RFID phase information and display distinctive gesture features. Firstly, we build a signal sensing model to capture gesture information effectively. Secondly, we innovatively convert the Chinese character phase signal into an intuitive picture representation and use the gesture movement change pattern as input through data preprocessing operations. Finally, modify-visual geometry group network (M-VGG) model is used for cross-domain gesture recognition. The M-VGG network learns high-level semantic features of gestures through spatiotemporal convolutions to extract cross-domain gesture features. Compared with other models, M-VGG has fewer parameters, and shorter training time, while still achieving high-accuracy gesture recognition. We evaluate its performance in three scenarios. The accuracy is 98.75% and 98% for new users and new scenarios, respectively, significantly outperforming existing wireless signal-based gesture recognition methods.**

***Index Terms*— Gesture recognition, radio frequency identification (RFID), tag.**

I. Introduction

GESTURE recognition has received a lot of attention in HCI and perception recognition, where interaction between the user and the device is achieved by the user performing different body movements. Among them, gesture recognition based on Chinese characters can not only facilitate communication between hearing impaired and person without disability [1], but can also be applied to the interaction of gaming devices. By drawing different gesture movements instead of tapping the keyboard or clicking the mouse as input to the computer, it can make HCI more intelligent and natural [2]. In addition, in 5G-based smart homes, people interact with smart assistants by performing specific gestures, which will further facilitate the development of smart homes and improve the convenience of everyday life [3].

In the past decades, researchers have used wearable sensor-based, computer vision (CV), and wireless signal-based approaches for human activity recognition. For example, IMU [4], [5], EMG [6], [7], WiFi [8], [9], [10], [12], radio frequency identification (RFID) [13], [14], acoustic [15], [16], millimeter wave [17], [18], ultrasonic [19], [20], and radar [21], [22], [23] technologies are used for gesture recognition. The wearable sensor-based approach has the advantage of low cost and high sensitivity, but it is less portable and more invasive. The vision-based approach is highly dependent on line of sight (LOS) and lighting conditions and has the potential to violate user privacy. The measurement granularity using WiFi is too coarse to recognize tiny gestures with fixed wavelengths, and multiuser recognition is a fundamental challenge for WiFi sensing. Dedicated devices such as USRP and Radar are not cost-effective. These limit their widespread use in real-world environments.

Fortunately, RFID-based gesture recognition has noninvasive and low-cost advantages, enabling its widespread application across scenarios such as smart homes, museums, and art galleries. However, existing systems still have many limitations. For example, GRfid [13] achieves basic recognition of predefined gestures with a relatively large range of hand motion (20–30 cm). RF-finger [24] captures fine-grained finger motions by deploying dense tag arrays and reduces the impact of coupling effects on gesture signals by orthogonal placement of adjacent tags. However, these systems either cannot achieve fine-grained gesture recognition or cannot eliminate the coupling effect among tags.

To address the aforementioned issues, we propose a noninvasive and fine-grained RFID-based Chinese gesture recognition system RF-CGR, which is robust to user and environment domain variation. The fundamental principle behind RF-CGR's recognition of different Chinese characters is that each character corresponds to a unique signal pattern. The same character drawn by different users yields similar signal patterns, whereas different characters lead to differentiated signal patterns. By analyzing the signal pattern variations, RF-CGR can identify distinct Chinese characters. Although the underlying rationale seems straightforward, several key challenges require to be addressed.

Challenge 1: The writing process of Chinese characters is complex and dynamic, and it's noteworthy that different Chinese characters may share similar stroke sequences (e.g., the drawing process of "土" and "工"), resulting in highly similar signal patterns. However, commercial RFID readers have limited spatial resolution, which makes it

Manuscript received 28 April 2023; revised 4 September 2023; accepted 21 September 2023. Date of publication 6 October 2023; date of current version 18 October 2023. This work was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region and the Major Science and Technology Special Project of Xinjiang Uygur Autonomous Region under Grant 2022D01C54 and Grant 2022A03016-4. The Associate Editor coordinating the review process was Dr. Alice Buffi. *(Corresponding author: Yajun Zhang.)*

Zhixiong Yang and Ziyi Zhen are with the School of Future Technology, Xinjiang University, Urumqi 830046, China (e-mail: 1075519485@qq.com; 75615689@qq.com).

Zijian Li, Xu Liu, Bo Yuan, and Yajun Zhang are with the School of Software, Xinjiang University, Urumqi 830046, China (e-mail: lzjgit123@163.com; 982880865@qq.com; 1950384084@qq.com; yajunzhang369@163.com).

Digital Object Identifier 10.1109/TIM.2023.3322505

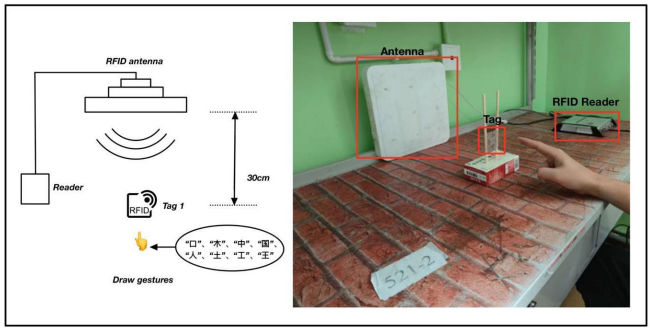


Fig. 1. Experimental scenario deployment.

difficult for them to distinguish the subtle phase differences between Chinese characters. Obtaining complex and fine-grained Chinese character gesture signals under limited spatial resolution poses a challenge. To address this challenge, we started from the principle of signal perception and established a signal sense model to analyze the impact of gesture positions on phase signals. As shown in Fig. 1, we leveraged this model to optimize the antenna-tag layout to improve the granularity of gesture perception, enabling the capture of intricate variations in complex gestures.

Challenge 2: Compared to only using RSSI for gesture recognition, the phase signal contains richer information with higher spatiotemporal resolution. However, the raw phase signals are abstract and unintuitive, which cannot directly reflect gesture motions. In addition, phase signals are sensitive to noise and can be easily influenced by environmental noise and multipath effects. Extracting unique and intuitive features to characterize the intricate gestures of Chinese characters poses a challenge. To address this challenge, our insight is to convert the abstract signal processing problem into an intuitive image recognition problem through problem transformation. Specifically, a series of preprocessing operations are performed on the collected data to improve the quality and reliability of gesture data. Meanwhile, the abstract raw phase data is converted into intuitive 224 × 224 pixel images. Such conversion makes the data more interpretable and intuitive.

Challenge 3: Gestures are often highly dependent on the gesture performers and environments, which we refer to as domain in this article. When the same user performs the same gesture in different environments or different users perform the same gesture in the same environment, domain shifts lead to large deviations in gesture signals, making models trained on specific domains difficult to transfer to new domains. Cross-domain gesture recognition is a challenge. To address this challenge, our insight is to learn high-level semantic features of gestures through spatiotemporal convolutions of the modify-visual geometry group network (M-VGG) network instead of relying on specific spatial environmental information for extracting cross-domain gesture features. Meanwhile, we collect data from different users and environments and use data augmentation to generate more diverse training data so that the model learns to distinguish between semantic and domain features. The resulting trained model will have cross-user and cross-environment gesture recognition capabilities.

To evaluate our system, we conducted experiments in three different scenarios (classroom, conference room, and dormitory). Fourteen volunteers (seven males and seven females) are invited to collect 25 200 sample sets. We conduct comprehensive experiments in various scenarios to evaluate the performance of RF-CGR in identifying eight Chinese characters, namely "口," "木," "中," "国," "人," "土," "工" and "王". We chose Chinese characters as the representation for several reasons: Chinese characters often involve multiple variations in gestures, resulting in richer and more diverse gesture signal patterns. Recognizing Chinese characters opens up a broader range of recognition possibilities and applications. However, it also brings higher complexity and challenges. On the other hand, the shape variations of digits 0–9 or individual letters are relatively limited, conveying less information and having narrower applicability. Considering these factors, we chose to use Chinese characters as the representation in our system.

Our contributions are summarized as follows.

Contribution 1: It presents an RFID gesture recognition system that exploits the signal sensing model to optimize the antenna-tag layout to improve the granularity of gesture perception. As a result, the design achieves fine-grained gesture recognition even with a single antenna-single tag layout.

Contribution 2: RF-CGR converts complex RFID signal processing problems into image processing problems, realizes the conversion of abstract phase signals into intuitive image forms, and successfully uses it to recognize gestures. Although our design and results are proposed in the context of RFID, the fundamental ideas can be extended to other wireless systems.

Contribution 3: We implemented the RF-CGR system on commercial RFID devices and conducted extensive experiments that validated the cross-domain performance of RF-CGR.

The rest of the article is organized as follows, with a discussion of related work in Section II, an elaboration of preliminary work in Section III, a description of the detailed design of the RF-CGR identification system in Section IV, implementation and evaluation of our experimental approach in Section V, a discussion of the experiments and future perspectives in Section VI, and a conclusion of the article in Section VII.

II. Related Work

A. Gesture Recognition Technology

Gesture recognition is crucial in the field of HCI. Traditional gesture recognition uses wearable sensors [25], [26], [27], [28] and CV-based [29], [30], [31], [32], [33], [34], [35], [36], [37] methods for recognition. Although these techniques have excellent gesture recognition results, these methods suffer from privacy issues and cost concerns. Gesture recognition based on wireless technologies [38], [39], [40] can solve the above-mentioned troubles, among which RFID, as an emerging technology, has excellent performance in the field of gesture recognition. On the one hand, RFID-based gesture recognition has the advantage of being low-cost and easy to deploy. On the other hand, it is nonintrusive and can effectively protect the user's privacy in the process of gesture recognition.

Grfid [13] used phase information of RFID for gesture recognition for the first time, overcoming the trouble of requiring tags on the user. Zhang et al. [14] extracted gesture recognition by combining coarse-grained statistical attributes and fine-grained wavelet decomposition attributes and realized low latency and accurate noncontact gesture recognition. RFnet [58] employs a multibranch 1D-CNN framework to enable human gesture detection and user identification. We design the RF-CGR as an RFID-based system that can recognize fine gestures without requiring the user to carry any additional sensor equipment.

B. RFID-Based Technology Sensing

RFID has been used to identify objects, and the literature [41], [42] uses RFID signals for target material identification, especially liquids. Recent studies have shown that RFID signals are rich in information and can be used for localization [43], [44], [45], activity recognition [46], [47], [48], gesture recognition [13], [24], [49], and vital sign detection [50], [51], [52]. The literature [53] achieves high accuracy in locating vehicle positions, breaking through the high degree of distortion caused by the performance limitations of traditional localization systems in specific application scenarios such as tunnels, under multistory bridges, and densely shaded areas with vegetation. TagFocus [54] converts the localization problem to a matching problem through the combination of CV and RFID, and fine-grained multiobject recognition and tracking can be performed through visual assistance. Duan et al. [55] proposed Taggo to design an efficient self-checkout model using the lightweight RFID, freeing the cashier's labor and shortening the traditional checkout queue. The RF-ECG proposed in reference [14] monitors the target person's heart rate by deploying a tag array in the chest area of clothes, but it cannot detect the heart rate of multiple users at the same time. Tagbreath [56] separates the backscattered signal of RFID to realize multiuser heart rate monitoring, improves the heart rate monitoring performance, and reduces the detection error by fusing the displacement of two consecutive phase readings of the multitag array at the same frequency and separating the displacement value dominated by human respiration. ReActor [57] realizes gesture recognition with low delay and high accuracy by combining machine learning to quickly calculate the statistical data of signal summary representing coarse-grained features and the wavelet (transform) coefficient of signal summary representing fine-grained local features. The difference between RF-CGR and the above work is that RF-CGR only adopts a single antenna-single tag layout, which can realize gesture recognition of Chinese characters in a cross-domain environment.

III. Preliminaries

In this section, some basic principles of RFID technology are introduced, and from them, our RFID sensing model is constructed.

A. Signal Sensing Model

RFID-based gesture recognition systems are usually composed of a reader, antenna, and tag. The reader is connected to the antenna to emit a continuous electromagnetic wave to activate the surrounding tag, which reflects the backscattered signal to the antenna, which is demodulated and decoded by the reader. The user draws a gesture that affects the transmission channel between the antenna and the tag, which in turn affects the tag's backscattered signal. The captured backscatter signal $S(t)$ is processed to identify the gesture performed by the user

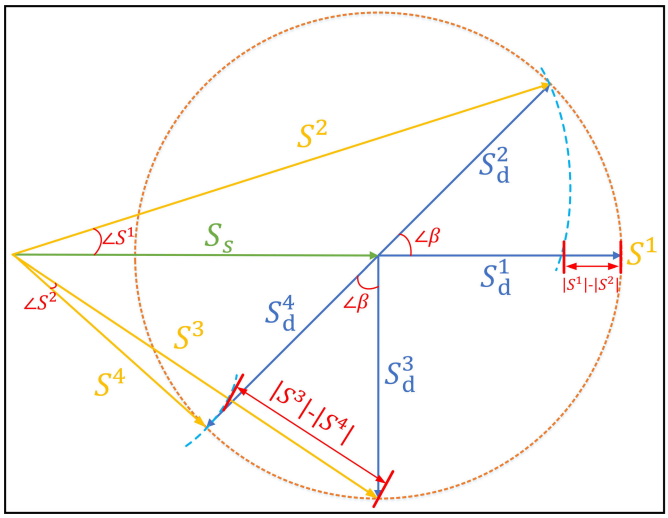


Fig. 2. Signal vector variation in the complex plane. The impact of S_d rotation at different angles on the amplitude and phase of $S(t)$.

$$S(t) = S_s + S_d = \alpha_s(t)e^{-i\theta_s(t)} + \alpha_d(t)e^{-i\theta_d(t)} \tag{1}$$

$$\theta(t) = \left(\frac{2d}{\lambda} \times 2\pi + \theta_0\right) \tag{2}$$

where S_s denotes the combined complex signal of the LoS path and the ambient multipath, i.e., the static signal, and S_d denotes the dynamic signal of the human gesture movement. $\alpha(t)$ and $\theta(t)$ denote the amplitude and phase of the backscattered signal, respectively. θ_0 is the initial offset of the phase, i is an imaginary unit, λ denotes the signal wavelength, d denotes the distance between the antenna and the tag, and the signal propagation distance is $2d$ in the actual transmission process.

To visualize the gesture signal changes, we describe the RF signal on a two-dimensional plane. Within a short time, the static component S_s of the backscattered signal remains constant and the signal amplitude $\alpha_d(t)$ of S_d is essentially stable, but the phase changes dramatically. As shown in Fig. 2, The change in the phase of S_d causes a rotation of S_d relative to S_s. To represent the impact of different rotation angles on the signal, we assume that S_d rotates at the same angle $\angle\beta$ relative to S_s. S_d^1, S_d^2, S_d^3, and S_d^4 represent the rotated phases of S_d, with the same amplitude but at different angles. $\angle S$ represents the phase change of the composite signal, and $|S|$ represents the amplitude change of the composite signal. When S_d rotates the angle $\angle\beta$ relative to S_s at 0°, i.e., S_d^1 rotates to S_d^2. At this time, the composite signal $S(t)$ rotates from S^1 to S^2, resulting in a phase change of $\angle S^1$ for the composite signal, and an amplitude change of $|S^1 - S^2|$. Similarly, When S_d rotates the angle $\angle\beta$ relative to S_s at 90°, i.e., S_d^3 rotates to S_d^4. At this time, the composite signal $S(t)$ rotates from S^3 to S^4,

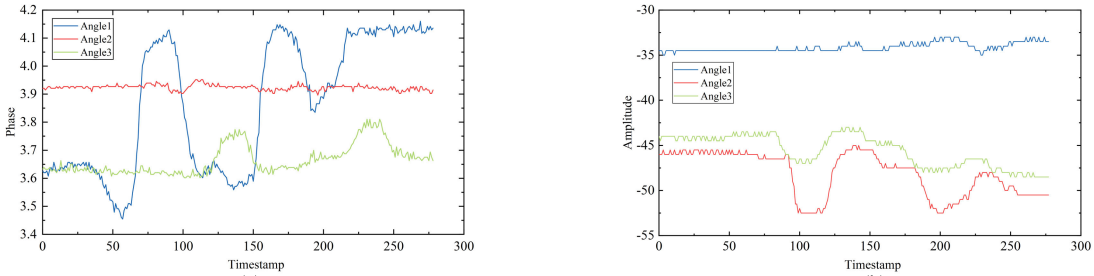


Fig. 3. Impact of different angles on the gesture of the Chinese character "口". (a) Phase of Chinese characters "口" from different angles. (b) Amplitude of Chinese characters "口" from different angles.

resulting in a phase change of $\angle S^2$ for the composite signal and an amplitude change of $|S^3 - S^4|$. It can be clearly from the figure that $\angle S^1 > \angle S^2$, and $|S^1 - S^2| < |S^3 - S^4|$. Therefore, the phase change of the signal is more significant when the phase difference with near 0°. A significant phase change is the key to accurately detecting changes in user gestures. In addition, we can see that when the phase difference with near 0°, the change in signal amplitude is much smaller than the change at 90°. Based on the modeling above, we can conclude that when the user draws gestures around 0°, the recognition performance of the gestures will be better.

To verify the validity of the signal sensing model, we ask a volunteer to draw the character "口". The angle between the volunteer's gesture and the tag 1, 2, and 3 are 90°, 0°, and 45°, respectively. As shown in Fig. 3, it can be seen that when the angle between the user and the tag is 90°, the phase hardly changes, while the amplitude changes significantly. In addition, when the user's angle between S_s and S_d is around 0°, the phase changes significantly, while the amplitude remains relatively constant. Hence, a suitable gesture acquisition direction can reduce the gesture perception blind spot.

B. Preliminary Experiments

Before human gesture recognition experiments, there are two basic problems: 1) How to select and set up tags to capture motion gesture information and how to deploy an effective antenna-tag layout? 2) How to control the distance between antennas and tags? Based on previous research experience, we choose the H47UHF tag as shown in Fig. 4.

The antenna-tag layout is shown in Fig. 5, and the experimenter performs the acquisition of gestures behind the antenna-tag. It is known from the literature [61] that there is a perceptual blind area due to multipath attenuation, and it is known from Section III-A that when phase information is used as the feature, its recognition effect is better near 0°, and our unique single antenna-single tag deployment method can increase the perceptual sensitivity. In our system, volunteers only need to move their fingers slightly to draw the corresponding Chinese characters. As a result, it is unnecessary to substantially increase the gap between the tag and the antenna. We control the antenna-tag distance to 30 cm in RF-CGR to conduct our experiments and we can also avoid the coupling effects [51].

Fig. 4. H47UHF tag.

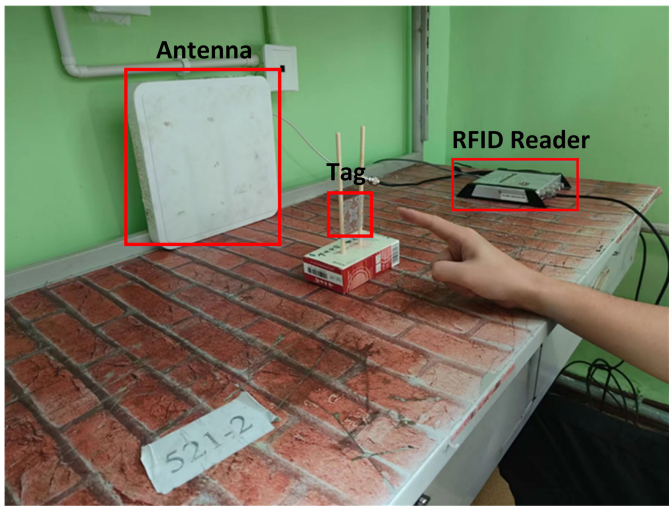


Fig. 5. Experimental setup.

C. Gesture Feature Selection

To verify the accuracy of phase as a gesture recognition feature, three experimenters are asked to draw the Chinese character "木". The RSS and phase graphs for the identical dynamic gesture used by three users with the same tag are displayed in Figs. 6 and 7, respectively. The RSS signals of the gestures collected by different users are significantly different, while for the phase, the waveform plots of the signals collected by different users are relatively similar. This is because the phase received by the antenna is influenced by the reflection of the human gesture movement, and for the same gesture, everyone writes similarly. Although the different writing speed of each person leads to the up-and-down translation or the different width of the phase waveform (user1 and user3), it can still be seen that the same gesture of different users has

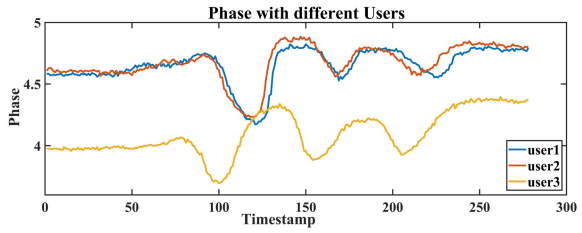


Fig. 6. Phase of different users drawing the Chinese character "木".

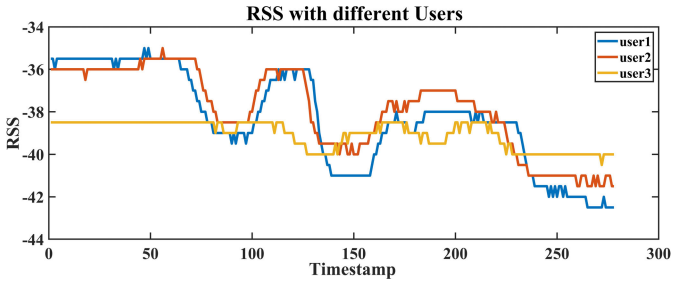


Fig. 7. RSS of different users drawing the Chinese character "木".

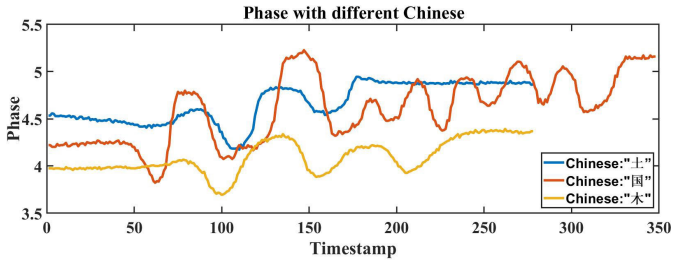


Fig. 8. Phase of different Chinese characters.

similar characteristics in the phase diagram. Therefore, in the experiment, we chose the phase to analyze human gesture recognition.

To verify that the phases behave differently for different gesture signals, we invite a volunteer to draw the Chinese characters "土," "国" and "木", respectively. Fig. 8 shows the phase diagram for the three gestures written by the volunteer under the same tag, and it can be seen that the different gestures have different characteristics. This indicates that the phase contains distinct human gestures and that the gestures made by the human body can be recognized by analyzing the phases.

Finally, to verify the generality of different tags for human gesture recognition, we invite a volunteer to draw the "国" under different tags, and Fig. 9 represents the phase diagram when the user performs the same dynamic gesture writing under five tag cases, respectively. From Fig. 9, we can see that the different tags show high similarity in the features of the same user, following the same time-varying features, and therefore, the different tags are universal for human gesture recognition.

IV. System Design

In this section, we detail the system design of the RF-CGR, including data acquisition, signal preprocessing, feature extraction, and gesture recognition.

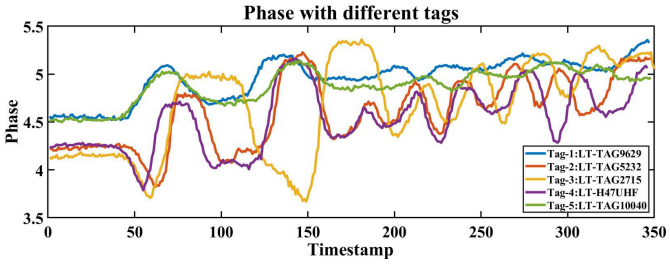


Fig. 9. Phase of Chinese characters under different tags.

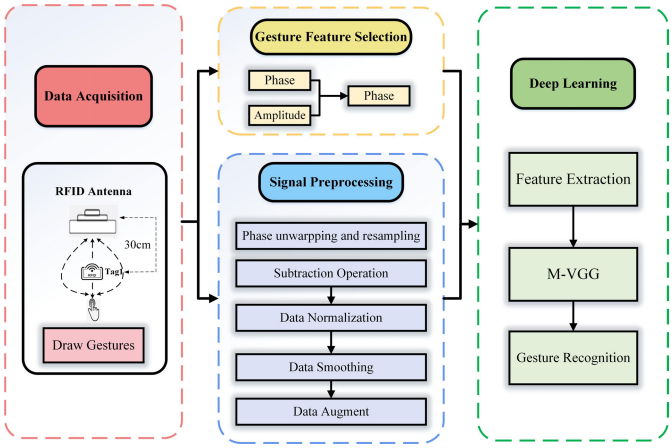


Fig. 10. Overview of the RF-CGR system.

A. Overview of the System

RF-CGR is a Chinese character recognition system based on an RFID device, which is capable of capturing complex Chinese characters using only a single antenna-single tag layout, and the basic idea of RF-CGR is to extract the characteristics of different Chinese characters to distinguish between different Chinese character gestures when users appear in the sensing area. The system architecture is shown in Fig. 10 and consists of the following modules.

Signal Acquisition Module: We use the signal sensing model in Section III-A to construct a single antenna-single tag layout to obtain the gesture signal and select the phase signal as the gesture signal feature.

Signal Preprocessing Module: We first perform phase unwrapping and signal resampling on the original signal and then highlight the differences between the different gesture signals by subtraction operations. The processed gesture signals are normalized and the interference caused by ambient noise is reduced using the Savitzky–Golay filter (SG Filter) [62]. In addition, we further increase the diversity of the Chinese character gesture data through data expansion operations, converting the gesture data into an intuitive image format.

Gesture Recognition Module: M-VGG deep learning module is used to recognize gestures. M-VGG improves the VGG16 network, which greatly reduces the parameters of the network and shortens the training time. It also achieves highly accurate recognition of Chinese characters.

B. Signal Preprocessing Module

Due to the various environmental noises and interferences in the raw RFID gesture signals, the system performs a set

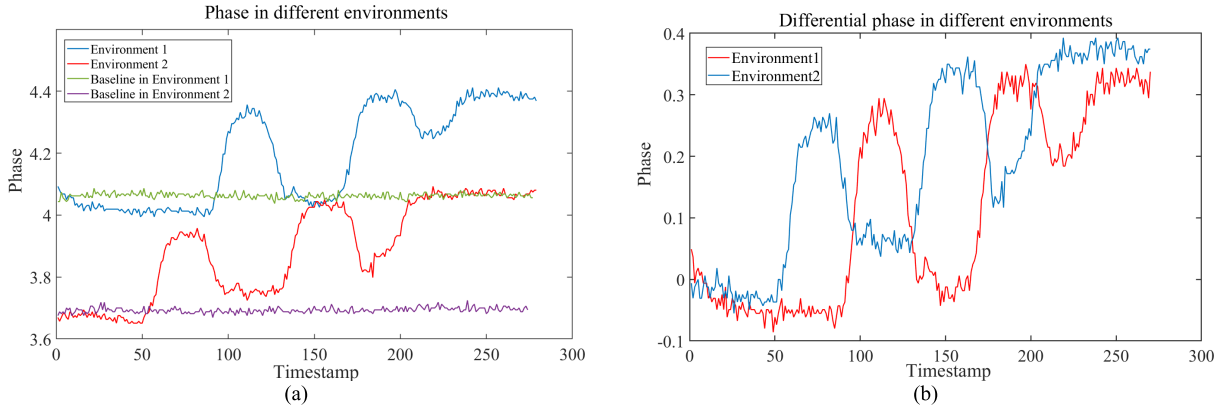


Fig. 11. Remove static interference from different environments. (a) Phase of gestures for different environments. (b) Differential phase of gestures in different environments.

of preprocessing steps to obtain clear phase and amplitude readings in this section. Through problem transformation, the RF signal is converted to the form of a picture as input to the M-VGG network for training.

1) Phase Unwrapping and Resampling: As the RF phase received by the reader is in the range $[0, 2\pi)$, the RF phase at this point is known as the wrapped phase. The user draws the gesture as a continuous process, but the wrapped phase is discontinuous and causes a jump when the phase is at a boundary value $[0, 2\pi)$. To accurately represent the process of continuous gestures, RF-CGR corrects the phase values by unfolding the wrapped phase and we use the method in [43] where the difference between the absolute values of the consecutive phase values is less than the absolute difference between two adjacent read phase values less than π. The inconsistent polling interval of passive tags in the time domain during the reader-antenna transmission signal results in the tags being sampled out of sync. To obtain phase data at the same sampling frequency, we use linear interpolation to process the phase stream, and the linear interpolation process is as follows. We resample the data stream by interspacing the linear interpolation between successive values by 20 ms

$$\frac{y - y_0}{x - x_0} = \frac{y_1 - y_0}{x_1 - x_0} \tag{3}$$

$$y = y_0 + \frac{(x - x_0)y_1 - (x - x_0)y_0}{x_1 - x_0} \tag{4}$$

where the known points (x_0, y_0), (x_1, y_1), (x, y) are the newly inserted values.

2) Subtraction Operation: A superposition of the LOS signal and the non-LOS (NLOS) signal makes up the phase signal that RFID readers acquire. NLOS signal is mainly generated by static reflection (e.g., table, chair), dynamic reflection (e.g., human gesture movement), and within the RFID hardware itself. We collect two sets of data, the first set is the baseline data from two environments, that is, without dynamic reflection signals. The second set is the acquisition of gestures in both environments. From Fig. 11(a), it can be observed that the phase values of the same gesture are different in these two environments due to varying multipath interferences (Baseline in environments 1 and 2). However, their phase waveforms are similar. By subtracting the corresponding baseline signals of each environment from the gesture phase signal, the differenced phase contains only the dynamic signal of the user's gesture, effectively eliminating static multipath interferences. Fig. 11(b) shows that the differenced phase values start around 0, and the differences in phase between different environmental signals have been reduced (from (3.6–4.4) rad to $(-0.1 - 0.4)$ rad). This indicates that subtraction operation [46] successfully removed the impact of static multipath interferences, resulting in more stable and reliable gesture phase signals.

3) Data Normalization: The size of the gesture drawn and the relative position of the gesture movement differ from one user to another when performing the gesture, resulting in different scales of the phase information we capture and low comparability between the data, which affects the performance of gesture recognition. To solve the problem of different user gesture motion scales, we use a linear function normalization (Min-Max scaling) method to linearly transform the output of the gesture data so that they are in the same order of magnitude

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{5}$$

where X is the gesture data after removing the outliers, $X_{\max}$ is the maximum value of the current gesture sample data, and $X_{\min}$ is the minimum value of the current gesture sample data.

4) Data Smoothing: Due to the multipath effect, there is multipath noise in the collected gesture data, resulting in the need to denoise the gesture data. Due to the multipath effect, the collected gesture data contains multipath noise. As shown in Fig. 12, the original phase waveform of the Chinese character "木" has excessive jitter. This redundant waveform jitter is the noise introduced by the multipath effect, which severely affects the quality of gesture data and thus degrades gesture recognition performance. Therefore, we need to perform data smoothing on the gesture data.

In this study, the data are smoothed using the SG Filter. By fitting the low-frequency components in the gesture signal, the high-frequency components are smoothed. Specifically, SG Filter sets a window of data as $x[i]$, $i = -m, \ldots, 0, \ldots, m$, i take $2m + 1$ consecutive integer values, and fits this set of

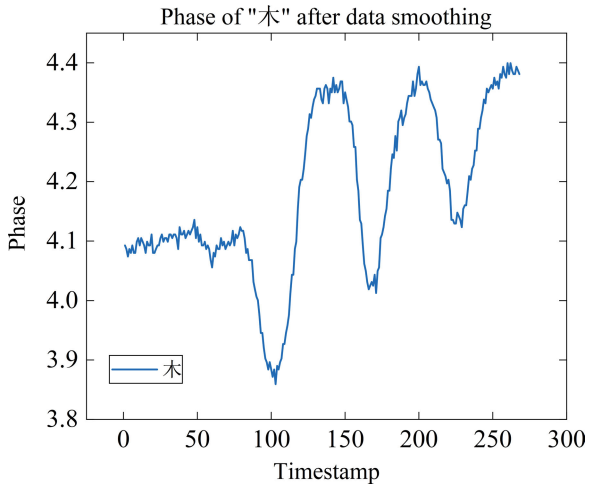


Fig. 12. Original phase of "木".

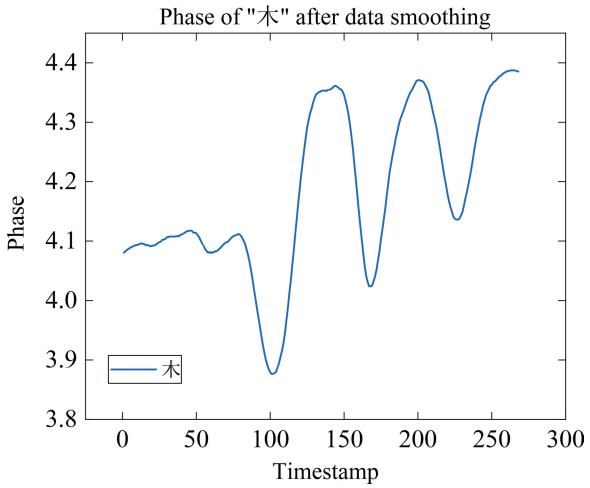


Fig. 13. Phase of "木" after data smoothing.

data by constructing an n-order polynomial

$$f(i) = \sum_{k=0}^{n} b_{nk} i^k = b_{n0} + b_{n1} i + b_{n2} i^2 + \cdots + b_{nn} i^n. \quad (6)$$

To reduce the sum E of squares of residuals between the fit data points and the original data points, the least squares approach is utilized

$$E = \sum_{i=-m}^{m} (f(i) - x[i])^2 = \sum_{i=-m}^{m} \left(\sum_{k=0}^{n} b_{nk} i^k - x[i] \right)^2. \quad (7)$$

When the fitting effect is optimal, a polynomial is produced that may be used to determine the approximate value of the window's center point.

By further solving it can be found that

$$\begin{cases} f[i]|_{i=0} = 0! b_{n0} = a_{n0} \\ \left. \dfrac{df(i)}{di} \right|_{i=0} = 1! b_{n1} = a_{n1} \\ \cdots \\ \dfrac{d^n f(i)}{di^n} = n! b_{nn} = a_{nn}. \end{cases} \quad (8)$$

It can be seen that calculating is equivalent to performing a refine impulse response (FIR) filtering on the original data, that is, the unit impulse of the input and output filters is convolved accordingly to get the output

$$a_{nn} = \sum_{i=-m}^{m} h[i] x[n-i]. \quad (9)$$

This feature is used to create the convolution coefficient table, and by employing discrete convolution of the coefficients, it is possible to quickly acquire the smooth value of the window's center point. The final Savitzky-Golay smoothing formula is

$$x_{k,\text{smooth}} = \overline{x_k} = \frac{1}{H} \sum_{i=-w}^{+w} x_{k+i} h_i \quad (10)$$

where (h_i/H) is the smoothing factor, which is obtained from the least squares polynomial. Fig. 13 illustrates the smoothed gesture waveform after signal processing. It is evident that the waveform appears more steady, indicating that the jitter induced by the multi-path effect has been removed.

5) Data Augment: Neural networks require a large number of datasets to reduce the loss values to obtain high accuracy. When using the convolutional blocks of VGG for feature extraction, a small number of datasets tend to be overfitting. To prevent overfitting and increase the robustness of the model, the dataset needs to undergo data expansion and enhancement operations to achieve high accuracy. Generative adversarial networks (GANs) [58] have been widely used in image generation and image enhancement tasks, using this feature, RF-CGR enhances the diversity of the preprocessed data by adding Gaussian noise and altering the image brightness, providing more variations and diversity for the data generated by GAN. The data generated by GAN, along with the data augmented by adding Gaussian noise and altering image brightness, together form an augmented dataset, offering a richer and more comprehensive set of training samples.

C. Deep Learning Module

M-VGG Motivation: Multibranch structures are used by ResNet, Inception, and other networks to enhance network performance but increase memory use. Because the results of each branch need to be saved until the last step of fusion (e.g., add) to free up the video memory of each branch. VGG has the characteristics of fast running speed, memory saving, and flexibility. We design the M-VGG network with convolutional operations instead of full connectivity, which not only can adapt to images of various sizes but also the network parameters and training time are significantly reduced and the model has high robustness. After the preprocessing operation of the collected gesture data, it is used as the input of the deep learning module for gesture feature extraction and gesture recognition. As shown in Fig. 14, the traditional VGG16 network consists of one input layer and six different blocks. The first five blocks perform convolutional operations, the sixth block performs three fully connected operations, and finally, classification is done via softmax. Although the VGG

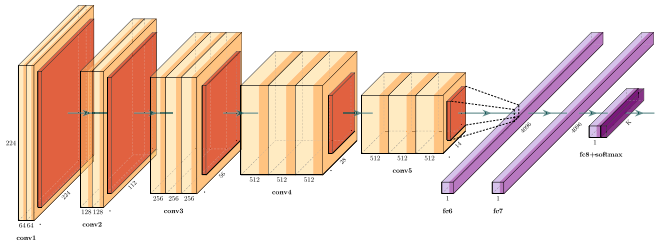


Fig. 14. Structure of VGG16 network.

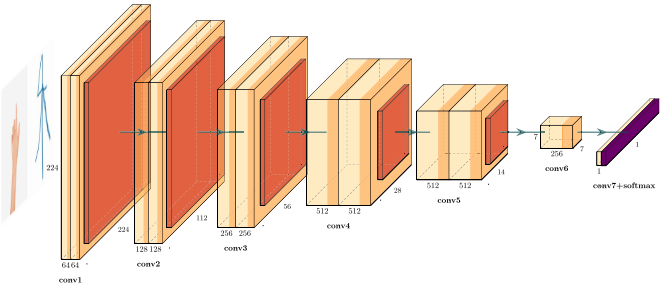


Fig. 15. M-VGG network model.

network structure is very simple, with 3 × 3 convolution kernels used throughout the whole network to increase depth and improve model performance, the three fully connected operations lead to higher memory usage. To mitigate this limitation, we designed the M-VGG network.

1) M-VGG Module: The core idea of the M-VGG module is to replace fully connected layers with convolutional layers and leverage 7 × 7 sized convolution kernels to downscale the image to 1 × 1 dimensions. This design not only achieves equivalent functionality to fully connected layers but also greatly reduces the number of network parameters and decreases training time.

Specifically, as shown in Fig. 15, the first five blocks of the M-VGG network have the same settings as the first five blocks of VGG16c. The main difference is that two convolutional layers are used to replace the three fully connected layers. The input channels of the first convolutional layer are 512 (consistent with the output channels of the previous block), the output channels are 256, the kernel size is 3 × 3, padding = 1, and stride = 1. After convolution, the image size remains unchanged, but the number of channels is halved, to reduce the number of parameters. In addition, Batch Normalization and ReLU operations are performed after convolution this time to increase nonlinear expressive ability. The input channels of the second convolutional layer are 256, the output channels are the number of gesture categories, the kernel size is 7 × 7, padding = 0, and stride = 1. Since the kernel size is the same as the image input size, the output size of 1 × 1 can be obtained, achieving the equivalent effect as the original three fully connected layers in VGG16c. After convolutional operations, softmax is used for model classification.

2) Chinese Character Feature Extractor: For different Chinese character gestures that will produce different waveform pictures, we use the convolution operation to automatically extract the gesture information of different domain environments and the gesture features of different users after converting the abstract phase data into an intuitive picture form. The core operation of the convolution layer can be simplified as

$$X = \text{POOL}(\text{Relu}(WX_0 + B)) \tag{11}$$

where W denotes the kernel, b denotes the bias term, and X is the input to the convolution layer.

After the image has been convolved, a nonlinearity is introduced by ReLU [58] to pool the activated gesture image to capture the change in image and gradient, bringing greater local information variability. The phase features extracted by convolution are represented as follows:

$$\text{RM} = \text{FE}(X, \theta_{\text{fe}}) \tag{12}$$

where X denotes the image of the preprocessing data, and θ_{fe} denotes all parameters after convolution extraction.

3) Chinese Character Gesture Recognizer: The Chinese character features pass through the Chinese character feature extractor to obtain the advanced gesture features, and we use the Softmax layer when classifying the advanced gesture features

$$P(y|x) = \frac{e^{h(x,y_i)}}{\sum_{j=1}^{n} e^{h(x,y_i)}}. \tag{13}$$

Notably, x and y denote gesture samples and gesture categories, respectively.

TABLE I
EXPERIMENTAL PARAMETERS TABLE

Parameters	Value
Distance between tag and antenna	30cm
Angle between antenna and ground	90°C
Reader acquisition frequency	920.875MHz

V. IMPLEMENTATION AND EVALUATION

The implementation and specific performance of the RF-CGR are covered in this section.

A. Evaluation Methodology

1) Experimental Environment: The experimental setup uses a single antenna-single tag layout. The experimental parameters are shown in Table I, and the distance between antennas-tags is 30 cm. Three scenarios are created. Scenario 1: As seen on the left side of Fig. 16, it is set up in a classroom with a 10 × 10 m space, and various tables and chairs are put in the classroom. Scenario 2: As seen in the middle side of Fig. 16, it is set up in a conference room with a 7 × 8 m space and has several electronic gadgets. Scenario 3: As seen on the right side of Fig. 16, it is set up in a dormitory with a 5 × 8 m space and has four iron bunk beds.

2) Hardware: Four components make up the hardware: an ImpinjR420RFID reader (Fig. 17) operating at 920.875 MHz, an RFID UHF circularly polarized antenna (9 dbi) (Fig. 18), a 5 × 5 cm tag, and a Lenovo R7000p computer.

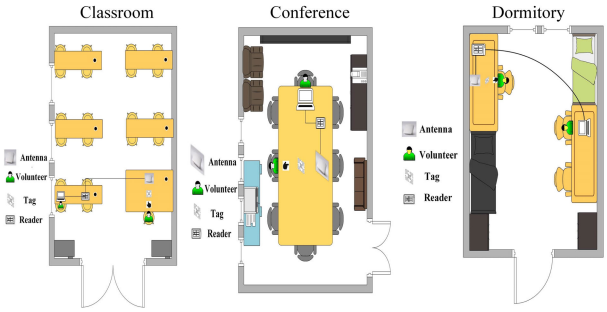


Fig. 16. Experimental scene.

Fig. 17. Impinj R420 RFID reader.

Fig. 18. UHF circular polarization antenna.

3) Measurement Metrics: We use accuracy (ACC) and true positive rate (TPR) as evaluation metrics for model performance. ACC is a measure of the probability of correctly recognizing a user's gesture and is calculated as

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}} \tag{14}$$

where true-positive (TP): Positive cases where the number of positive cases is judged to be positive, true-negative (TN): Samples where negative cases are judged to be negative. False-positive (FP): Samples where negative cases are judged to be positive. False-negative (FN): Samples where positive cases are judged to be negative. TPR represents the recall rate, i.e., how many of all positive samples are successfully identified by the model, and is expressed as follows:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{15}$$

4) Experimental Dataset: We invite 14 volunteers (seven males, seven females) to collect our data. Ten (five male, five female) of these are chosen at random to serve as the training set for the gesture data and the remaining four (two males, two females) as the test set. Each volunteer draw eight different Chinese characters in three different scenes, and each Chinese character is drawn 25 times. During data collection, we require each user to draw complete Chinese characters following the stroke order of the characters. This is because different stroke orders and partial omissions of the same Chinese character will produce different signal patterns, leading to multiple signal patterns for the same character and thus affecting the experimental results. The preprocessing gesture data three times is to embed noise to expand the dataset. We collect a total of 25 200 samples [25 (time) × 14 (people) × 8 (character) × 3 (times augment) × 3 (scenes)], including 18 000 samples in the training set (25 × 10 × 8 × 3 × 3) and 7200 samples in the test set (25 × 4 × 8 × 3 × 3).

B. Micro-Benchmark

1) Impact of Antenna-Tag Distance: The distance between the antenna and the tag affects the acquisition of gesture signals. To explore the effect of antenna-tag distance, five distances (25, 30, 35, 40, 45 cm) are set up for evaluation. As follows, we invite a volunteer to perform gesture drawing at the above five distances, and each character is captured 25 times. As can be seen from Fig. 19(a), the best performance is achieved at the antenna-tag distance of 30 cm, with a slight drop in performance when the distance is on either side of the optimum, and a significant drop in gesture recognition performance when the distance is greater than 45 cm. Therefore, the antenna tag distance needs to be set to within 40 cm, which is sufficient for most real-life applications.

2) Impact of Finger-Tag Distance: To evaluate the effect of finger-tag distance on gesture recognition. We set five different distances between the finger and the tag: 3, 6, 9, 12, and 15 cm. We invite a volunteer to perform gesture drawing at the above five distances, and each character is captured 25 times. From Fig. 19(b), it can be observed that when the distance between the finger and the tag is less than 12 cm, RF-CGR can effectively recognize different gestures with an accuracy exceeding 93%. However, when the distance is increased to 15 cm, the gesture recognition accuracy decreases. This is because as the distance between the finger and the tag increases, the signal strength received by the tag decreases, making the demodulation and recognition process more challenging, thus affecting the system's accuracy.

C. Overall Performance of RF-CGR

We divided the collected dataset, where 70% was used as the training set and the remaining 30% as the test set. The test set consists of data that has not been used for training. For the training set, we employed fivefold cross-validation to train the model, which effectively prevents overfitting. We evaluated the gesture recognition performance on the test dataset, as shown in Fig. 20. The overall accuracy of RF-CGR reached 98.375%. However, the accuracy for the characters "工" and "土" was relatively lower. This is mainly because

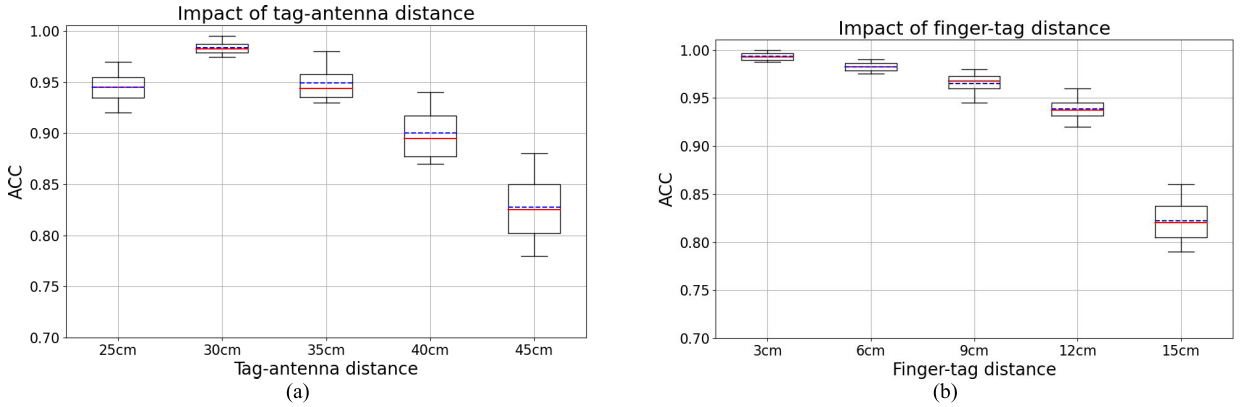


Fig. 19. Micro-benchmark. (a) Effect of different Tag-antenna distances on gesture recognition performance. (b) Effect of different Finger-tag distances on gesture recognition performance.

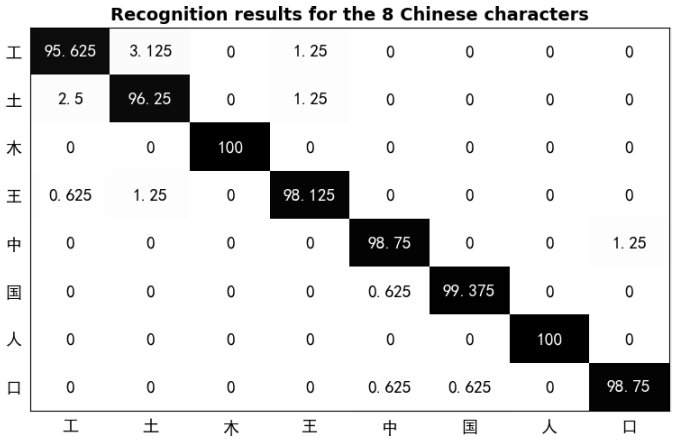


Fig. 20. RF-CGR overall performance confusion matrix.

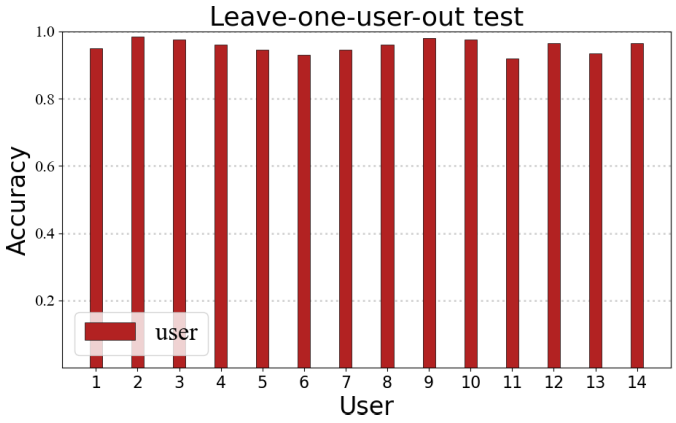


Fig. 21. Leave-one-user-out-test.

these two characters have the same stroke order, leading to a slight decrease in recognition rate. The results demonstrate that RF-CGR is capable of effectively recognizing different Chinese characters with high accuracy.

D. Leave-One-User-Out Test

Leave-one-user-out test means selecting one user's data from the dataset as the test set (with a different user chosen each time) and using the data from other users as the training set. This process is repeated until all users have been tested. Through this method, we can evaluate the model's adaptation to different users and identify issues such as overfitting or underfitting. As shown in Fig. 21, we assessed the testing performance for all users. The overall accuracy for all users (sum of individual user accuracies divided by the number of users) is 95%. However, Individual users (user 6, user 11) draw smaller or slower characters, resulting in reduced gesture recognition.

E. Performance Under Different Environments

To validate the effectiveness of RF-CGR under different environmental conditions, we conducted tests in a classroom, meeting room, and dormitory, corresponding to low, medium, and high multipath interference, respectively. It is important to note that, to ensure cross-domain robustness, the experimental locations for data collection during testing were different from those used for training the dataset. Each user performed 200 gestures (25 times for each Chinese character) in each environment. We used the datasets from these different environments as the test sets and evaluated the system's Accuracy and TPR using a pretrained RF-CGR model, as shown in Fig. 22. Overall, RF-CGR achieved an accuracy rate of 98%. The average matching accuracies in the classroom, meeting room, and dormitory were 100%, 98.3%, and 95.7%, respectively. However, due to the similarity in phase characteristics between the Chinese characters "工" and "土" the TPR experienced a slight decrease (see Fig. 23). Furthermore, the presence of multiple electronic devices' interference in the meeting room and the influence of the dormitory's iron beds on gesture signals caused a slight reduction in the accuracy of Chinese character recognition, but it remained above 95%.

F. Performance Under Different Users

Different users' drawing speeds and variations in the size of Chinese characters can affect the phase signal. To evaluate the performance of RF-CGR in the user domain, we invited five volunteers (three males and two females) to participate in the experiment. In scenario 1 (classroom), each volunteer drew eight Chinese characters, repeating each character 25 times.

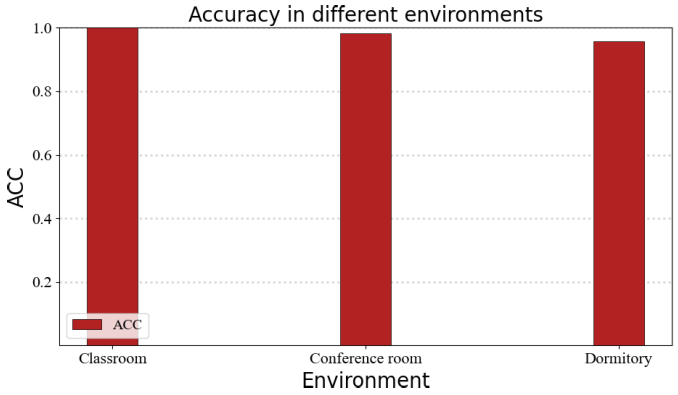


Fig. 22. ACC in different environments.

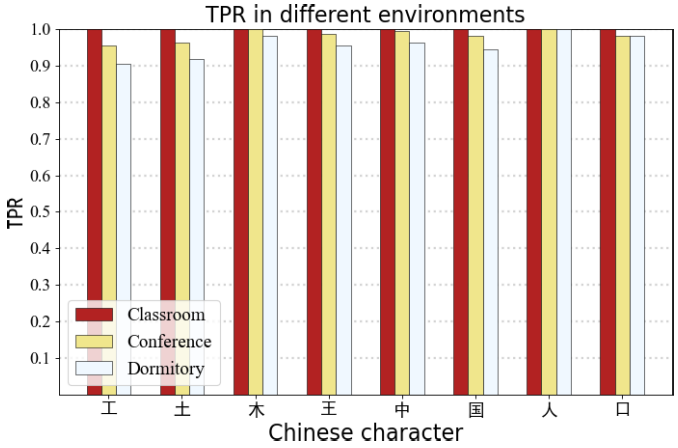


Fig. 23. TPR in different environments.

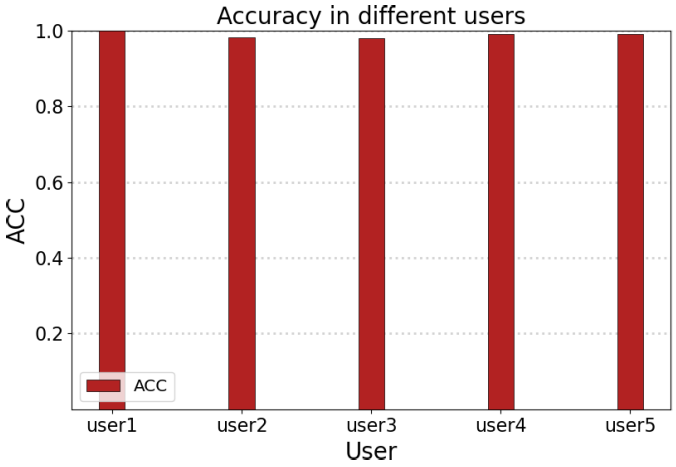


Fig. 24. ACC in different users.

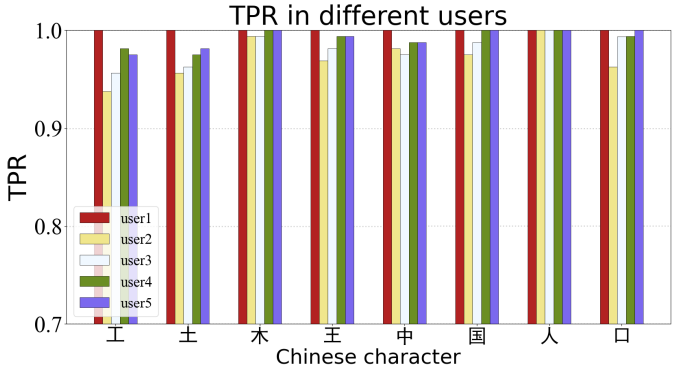


Fig. 25. TPR in different users.

As shown in Fig. 24, the overall accuracy of RF-CGR was 98.75%. The average accuracies for the five volunteers were 100%, 97.25%, 98.15%, 99.125%, and 99.225%, respectively. The lower accuracy of User 2 was due to the similarity between the small size of the characters "土" and "工" that she drew. The TPR performance results in Fig. 25 can be divided into two groups. The first group includes gestures for "木," "人," "国," "口," "王" and "中", with an average TPR rate of 99%. The second group involves gestures for "工" and "土", with average TPR rates of 96.875% and 97.5%, respectively. The second group of gestures has the same stroke order during the drawing process, leading to a slight decrease in gesture recognition accuracy, yet the system still achieves over 97% accuracy. This demonstrates the high robustness of RF-CGR across different users.

G. Impact of Sample Numbers

The accuracy of RF-CGR recognition is affected by the sample count. We plot the accuracy of RF-CGR concerning the number of samples in Fig. 26. It can be seen that when there are only 80 training samples for each Chinese character, the correct recognition rate of Chinese characters is close to 80%. As the number of samples increases, the recognition accuracy improves and stabilizes after the number of samples is greater than 200, at which time the correct recognition rate of Chinese characters can reach more than 98%.

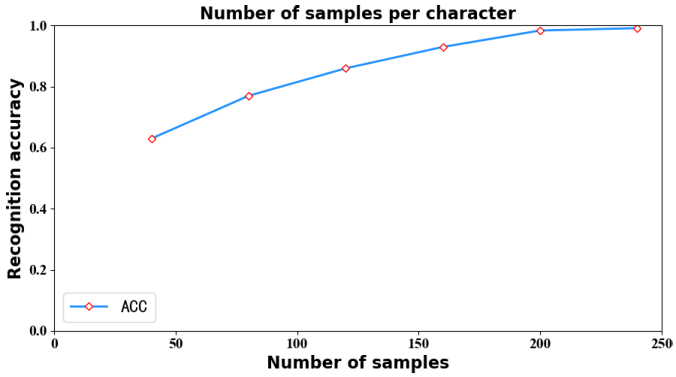


Fig. 26. Effect of sample size on the recognition rate of Chinese characters.

H. Performance Under Different Disturbances

Different interference factors refer to the presence of different objects around the user when performing a gesture, and RFID tags are susceptible to the effects of the metal. When metal objects are present, eddy currents are generated during gesture capture, which in turn reduces gesture recognition performance. To explore the performance of RF-CGR against carrying interference factors, three scenarios are set up for comparison. For scenario 1, the user performs a gesture in a classroom without interference. For scenario 2, a metal water cup is placed near the tag and reader. The user performs the gesture and the metal cup provides interference to the user.

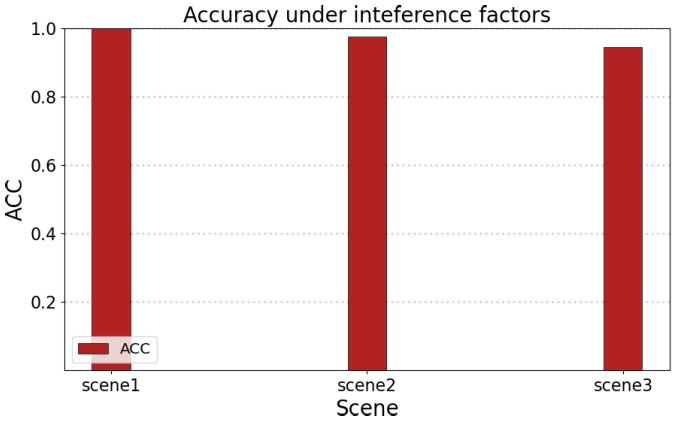


Fig. 27. ACC under different interference factors.

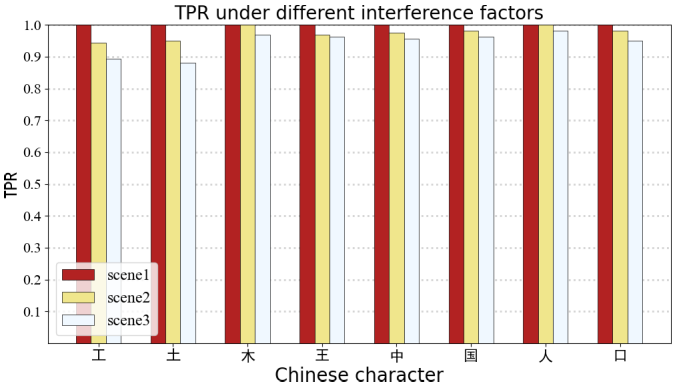


Fig. 28. TPR under different interference factors.

For scenario 3, the user draws a gesture in the classroom with multiple users walking simultaneously 1 m behind them while drawing the gesture action. The user executes each scene 200 times (25 times for each Chinese character). The experimental results are plotted in Figs. 27 and 28. For scenarios 2 and 3, the average accuracy of the method is reduced compared to the interference-free environment due to the effects of metal devices and multiuse interference, still providing an average accuracy of 97.5% and 94.5%. This indicates that RF-CGR is highly resistant to environmental interference.

I. Recognition Performance of Similar Chinese Characters

For Chinese characters that have the same stroke order but different shapes, this article refers to them as similar characters. The collected signal waveforms for these similar characters tend to be quite similar, confusing recognition results. To verify the recognition performance of RF-CGR on similar characters, we set up five groups of similar characters: {“日” and “曰,” “未”, and “末,” “天” and “夫”, “刀” and “力”, “土” and “工”}. These five groups can be divided into two types. The first type has the same stroke order but different lengths. The second type is the same stroke order, head and not head. We invited five volunteers to collect data by drawing each character 40 times. 70% of the dataset was used as the training set and 30% as the test set, with no overlap between the two sets. Fivefold cross-validation was used during training to prevent overfitting.

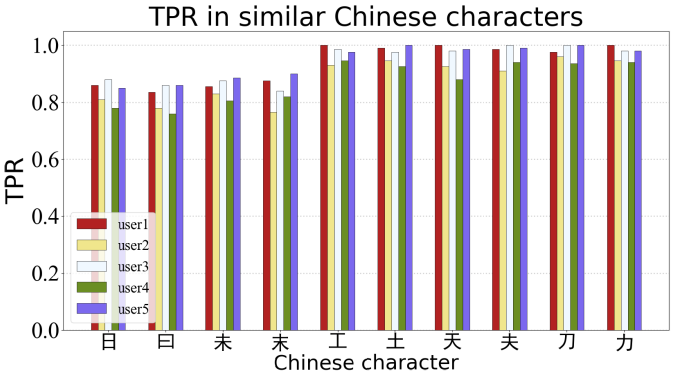


Fig. 29. TPR in similar Chinese characters.

As shown in Fig. 29, the average recognition rate of RF-CGR for the five groups of “similar” characters reached 91%. Among them, the TPR of the first type of similar characters was lower (83.6%). This is because different users drew characters in different sizes and distances, and the gesture features of the first type of similar characters were mainly distinguished by different drawing distances, resulting in lower recognition rates. Fortunately, such similar character groups are very few in the Chinese character set. The TPR of the second character group was higher (96.6%). This shows that the signal sensing model of RF-CGR can capture fine-grained gesture variations. Although the recognition rate of some similar characters is slightly lower, RF-CGR overall performs steadily and has a certain robustness for similar characters.

J. Complex and Commonly Used Chinese Character Recognition Performance

RF-CGR recognizes Chinese characters based on their unique signal patterns. To validate its recognition capability on complex and commonly used new characters, we tested eight newly introduced complex characters: {“破,” “智,” “你,” “我,” “他,” “她,” “其,” “能”}. Five volunteers were invited to draw each character 40 times. 70% of the collected dataset was allocated as the training set and the remaining 30% as the test set, with no overlap between the two sets. Fivefold cross-validation was leveraged during training to prevent overfitting. As shown in Fig. 30, the system achieved an average TPR of 97%, demonstrating that RF-CGR can accurately recognize complex and commonly used characters when new ones are introduced.

K. Performance Comparison of M-VGG With VGG16, ResNet34, and Inception-ResNet-v2

The deep learning module of RF-CGR is an improvement of the traditional VGG16 network. The traditional VGG16 network requires three fully connected operations after the convolution of blocks, which will make the number of parameters of the model large and thus greatly increase the training time of the model. For extracting image features, the image

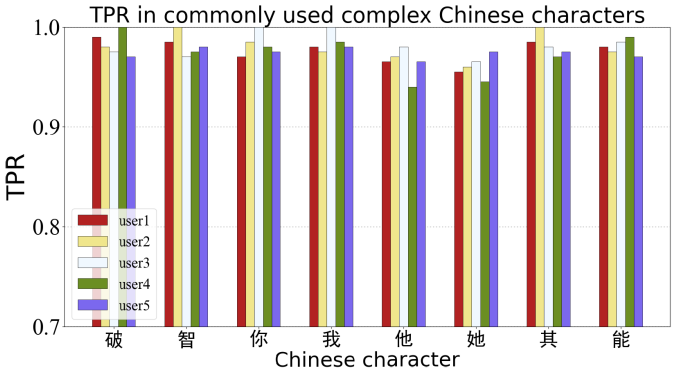


Fig. 30. TPR in commonly used complex Chinese characters.

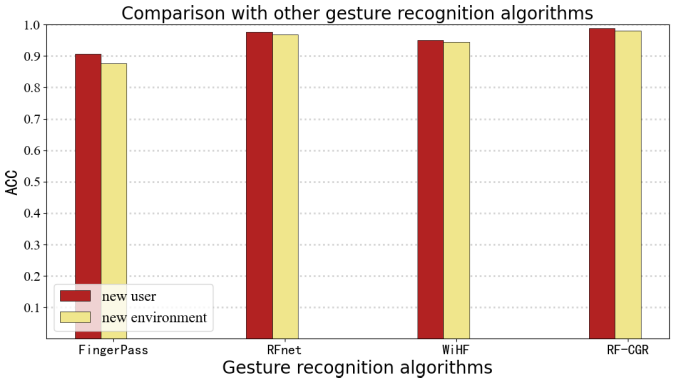


Fig. 31. Comparison with other gesture recognition algorithms.

TABLE II
COMPARISON WITH OTHER GESTURE RECOGNITION ALGORITHMS

	M-VGG	VGG16	ResNet34	Inception-ResNet-v2
Params Size(MB)	**43.05**	512.32	81.18	207.21
Estimated Total Size(MB)	**365.67**	834.84	1208.76	554.74
Epoch	10	10	10	10
Train Time(s)	**3545**	14836	18124	11103
TestAcc(%)	**98.75**	97.5	98.625	**99**

features can be well extracted by a convolution operation. We reduce the number of channels in the first convolutional layer of the sixth module by half it, replacing the three fully connected layers with two convolutional processes, and using a convolutional kernel of size 7×7 and padding $= 0$ in the second convolutional operation, which can down-sample the image to 1×1 and achieve the same effect as fully connected. Moreover, the number of parameters of the model decreases after this operation, which in turn shortens the training time. Table II shows the performance of M-VGG compared with VGG16, ResNet34, and Inception-ResNet-v2 networks. Inception- ResNet-v2 combines ResNet and GoogLeNet, and the "larger network size" allows the model to Inception-ResNet-v2 combines ResNet and GoogLeNet, and the "larger network size" allows the model to reach 99% correct rate, but the model takes longer to train because to its deeper network. In contrast, the M-VGG network has fewer parameters, shorter training time, and model accuracy up to 98.75%.

L. Comparison With Other Gesture Recognition Algorithms

The RF-CGR senses complex Chinese character gestures with a single antenna-single tag layout designed by a signal sensing model, converting abstract phase data into intuitive gesture signal waveforms and converting the signal recognition problem into an image recognition problem. The waveform pictures of different Chinese character gestures are used as input to a training model of the M-VGG network for gesture recognition. We compare RF-CGR with several other gesture recognition systems, such as FingerPass [60], RFnet [58], and WiHF [3], as shown in Fig. 31, the average accuracy of FingerPass using WiFi signal for intradomain and cross-domain scenes is 90.6% and 87.6%, respectively. WiHF's precision of gesture recognition and user recognition in the domain based on WiFi signal was 97.65% and 96.74%, respectively. RFnet improved the average accuracy to 95.1% and 94.4% for new users and environments, respectively, using the 7×7 tag array. In contrast, RF-CGR showed the best performance in both tests, where fewer tags were involved and accuracy was higher, at 98.75% and 98% for different users and different environments, respectively.

VI. DISCUSSION

RF-CGR has some limitations in practical applications.

First, to meet the needs of practical applications, more gesture data needs to be collected to train the network. However, the cost of data acquisition based on RFID is much higher than image-based recognition. In the future, we plan to adopt the idea of RF-CM [64] transfer learning, using models well-trained on large image-based gesture datasets to guide models with less RFID gesture data, to reduce the reliance on large-scale RFID gesture data, optimize the data collection process, and lower labor costs.

Secondly, RF-CGR achieves gesture recognition for individual characters, but not for continuous characters. For continuous Chinese gesture recognition, the performance of gesture segmentation will directly affect continuous Chinese gesture recognition. We consider combining the DeepSeg [65] dynamic segmentation framework for segmenting continuous Chinese character gesture gestures and also the long short term memory (LSTM) network for sequential gesture recognition, and extend it to the RF-CGR system.

Third, the signal sensing model maximizes the gesture signal sensing range, but the limitation of the single tag is that the upper limit of the number of characters recognized is bounded. In the future we intend to combine the multitag array approach of [61] to build a small tag array while still being able to increase the upper limit of the Chinese character recognition. The deployment position continues to follow the signal sensing model keeping the antenna array placement near 0° to further improve the sensing range of the multitag array. The second is to increase the upper limit of Chinese character recognition by extending the perception of gesture signals through multiangle-multi tag array coverage.

Finally, continuously retraining the model on new characters is necessary to recognize more Chinese characters. In this

regard, transfer learning can help reduce the training cost for new characters. However, given the vast number of Chinese characters, which can be tens of thousands, this method is impractical for large-scale character sets. Another solution is decomposing Chinese characters into distinct radicals and components. The number of radicals (around 120) and components (around 1000) in a commonly used Chinese character set is limited. We can first construct fundamental datasets of radicals and components, and train the model to recognize these basic parts. When recognizing handwritten Chinese, we can reconstruct characters by sequentially combining components to identify them. This method breaks down many characters into a limited set of component combinations, making it better suited for large-scale Chinese character datasets.

VII. Conclusion

The RF-CGR that is proposed in this study is an RFID-based, device-free, domain-independent system for recognizing Chinese character gestures. The antenna-tag layout of RF-CGR can effectively capture the spatiotemporal feature transformation of gestures. The signal can be extracted from the original RFID gesture signal with low differentiation by preprocessing operations such as subtraction operation, data normalization, data smoothing, and data expansion. We effectively minimize the number of parameters and the training time of the network by converting the RFID signals into understandable images for feature capture by the M-VGG network. Through several experiments, we show that the average correct recognition rate of RF-CGR can reach 98% and 98.75% in different domains (different experimental environments, different users). In our opinion, our RF-CGR system, which is based on Chinese character gestures, may significantly advance HCI.

References

[1] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "SignFi: Sign language recognition using WiFi," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–21, Mar. 2018.

[2] K. Ling, H. Dai, Y. Liu, A. X. Liu, W. Wang, and Q. Gu, "UltraGesture: Fine-grained gesture sensing and recognition," *IEEE Trans. Mobile Comput.*, vol. 21, no. 7, pp. 2620–2636, Jul. 2022.

[3] C. Li, M. Liu, and Z. Cao, "WiHF: Enable user identified gesture recognition with WiFi," in *Proc. IEEE Conf. Comput. Commun.*, Jul. 2020, pp. 586–595.

[4] C. Xu, P. H. Pathak, and P. Mohapatra, "Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch," in *Proc. 16th Int. Workshop Mobile Comput. Syst. Appl.*, 2015, pp. 9–14.

[5] H. Wen, J. Ramos Rojas, and A. K. Dey, "Serendipity: Finger gesture recognition using an off-the-shelf smartwatch," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2016, pp. 3847–3851.

[6] X. Jiang et al., "Measuring neuromuscular electrophysiological activities to decode HD-sEMG biometrics for cross-application discrepant personal identification with unknown identities," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–15, 2022.

[7] Q. Li, Z. Luo, and J. Zheng, "A new deep anomaly detection-based method for user authentication using multichannel surface EMG signals of hand gestures," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[8] R. H. Venkatnarayan, G. Page, and M. Shahzad, "Multi-user gesture recognition using WiFi," in *Proc. 16th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2018, pp. 401–413.

[9] Y. Zheng et al., "Zero-effort cross-domain gesture recognition with Wi-Fi," in *Proc. 17th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2019, pp. 313–325.

[10] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using WiFi," in *Proc. 15th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2017, pp. 252–264.

[11] Y. Bai and X. Wang, "CARIN: Wireless CSI-based driver activity recognition under the interference of passengers," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–28, Mar. 2020.

[12] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of WiFi signal based human activity recognition," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 65–76.

[13] Y. Zou, J. Xiao, J. Han, K. Wu, Y. Li, and L. M. Ni, "GRfid: A device-free RFID-based gesture recognition system," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 381–393, Feb. 2017.

[14] S. Zhang, C. Yang, X. Kui, J. Wang, X. Liu, and S. Guo, "ReActor: Real-time and accurate contactless gesture recognition with RFID," in *Proc. 16th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Jun. 2019, pp. 1–9.

[15] T. Amesaka, H. Watanabe, M. Sugimoto, and B. Shizuki, "Gesture recognition method using acoustic sensing on usual garment," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 6, no. 2, pp. 1–27, Jul. 2022.

[16] K. Sun, T. Zhao, W. Wang, and L. Xie, "VSkin: Sensing touch gestures on surfaces of mobile devices using acoustic signals," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2018, pp. 591–605.

[17] H. P. Liu et al., "Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 4, pp. 1–28, 2020.

[18] P. S. Santhalingam, A. A. Hosain, D. Zhang, P. Pathak, H. Rangwala, and R. Kushalnagar, "MmASL: Environment-independent ASL gesture recognition using 60 GHz millimeter-wave signals," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–30, Mar. 2020.

[19] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2016, pp. 82–94.

[20] Y. Iravantchi, M. Goel, and C. Harrison, "BeamBand: Hand gesture sensing with ultrasonic beamforming," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2019, pp. 1–10.

[21] J. Lien et al., "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–19, 2016.

[22] B. Li, J. Yang, Y. Yang, C. Li, and Y. Zhang, "Sign language/gesture recognition based on cumulative distribution density features using UWB radar," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.

[23] E. Hayashi et al., "RadarNet: Efficient gesture recognition technique utilizing a miniature radar sensor," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2021, pp. 1–14.

[24] C. Wang et al., "Multi–touch in the air: Device-free finger tracking and gesture recognition via COTS RFID," in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2018, pp. 1691–1699.

[25] R. Mardiyanto, M. F. R. Utomo, D. Purwanto, and H. Suryoatmojo, "Development of hand gesture recognition sensor based on accelerometer and gyroscope for controlling arm of underwater remotely operated robot," in *Proc. Int. Seminar Intell. Technol. Its Appl. (ISITIA)*, Aug. 2017, pp. 329–333.

[26] W. Guo, Y. Fang, X. Sheng, and X. Zhu, "Measuring motor unit discharge, myofiber vibration, and haemodynamics for enhanced myoelectric gesture recognition," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023.

[27] X. Chu, J. Liu, and S. Shimamoto, "A sensor-based hand gesture recognition system for Japanese sign language," in *Proc. IEEE 3rd Global Conf. Life Sci. Technol. (LifeTech)*, Mar. 2021, pp. 311–312.

[28] L. Meng et al., "User-tailored hand gesture recognition system for wearable prosthesis and armband based on surface electromyogram," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–16, 2022.

[29] H. Gao, C. Liu, Y. Li, and X. Yang, "V2VR: Reliable hybrid-network-oriented V2V data transmission and routing considering RSUs and connectivity probability," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3533–3546, Jun. 2021.

[30] C. Zhang, J. Tabor, J. Zhang, and X. Zhang, "Extending mobile interaction through near-field visible light sensing," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 345–357.

[31] J. Taylor et al., "Efficient and precise interactive hand tracking through joint, continuous optimization of pose and correspondences," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–12, Jul. 2016.

[32] Z. Zhang, B. Wu, and Y. Jiang, "Gesture recognition system based on improved YOLO v3," in *Proc. 7th Int. Conf. Intell. Comput. Signal Process. (ICSP)*, Apr. 2022, pp. 1540–1543.

[33] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree, and J. Kautz, "Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4207–4215.

[34] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, Feb. 2012.

[35] B. Fang, J. Co, and M. Zhang, "DeepASL: Enabling ubiquitous and non-intrusive word and sentence-level sign language translation," in *Proc. 15th ACM Conf. Embedded Netw. Sensor Syst.*, Nov. 2017, pp. 1–13.

[36] L. E. Potter, J. Araullo, and L. Carter, "The leap motion controller: A view on sign language," in *Proc. 25th Austral. Computer-Human Interact. Conference: Augmentation, Appl., Innov., Collaboration*, Nov. 2013, pp. 175–178.

[37] F. Jiang, S. Zhang, S. Wu, Y. Gao, and D. Zhao, "Multi-layered gesture recognition with Kinect," *J. Mach. Learn. Res.*, vol. 16, pp. 227–254, Feb. 2015.

[38] Z. Tang, Q. Liu, M. Wu, W. Chen, and J. Huang, "WiFi CSI gesture recognition based on parallel LSTM-FCN deep space-time neural network," *China Commun.*, vol. 18, no. 3, pp. 205–215, Mar. 2021.

[39] J. Song et al., "In-air gestures around unmodified mobile devices," in *Proc. 27th Annu. ACM Symp. User interface Softw. Technol.*, Oct. 2014, pp. 319–329.

[40] P. Wang, R. Jiang, and C. Liu, "Amaging: Acoustic hand imaging for self-adaptive gesture recognition," in *Proc. IEEE Conf. Comput. Commun.*, May 2022, pp. 80–89.

[41] B. Xie et al., "Tagtag: Material sensing with commodity RFID," in *Proc. 17th Conf. Embedded Networked Sensor Syst.*, Nov. 2019, pp. 338–350.

[42] J. Wang, J. Xiong, X. Chen, H. Jiang, R. K. Balan, and D. Fang, "TagScan: Simultaneous target imaging and material identification with commodity RFID devices," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2017, pp. 288–300.

[43] H. Xu, D. Wang, R. Zhao, and Q. Zhang, "AdaRF: Adaptive RFID-based indoor localization using deep learning enhanced holography," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–22, Sep. 2019.

[44] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2014, pp. 237–248.

[45] L. Shangguan and K. Jamieson, "Leveraging electromagnetic polarization in a two-antenna whiteboard in the air," in *Proc. 12th Int. Conf. Emerg. Netw. EXperiments Technol.*, Dec. 2016, pp. 443–456.

[46] H. Ding et al., "FEMO: A platform for free-weight exercise monitoring with RFIDs," in *Proc. 13th ACM Conf. Embedded Networked Sensor Syst.*, Nov. 2015, pp. 141–154.

[47] S. Amendola, L. Bianchi, and G. Marrocco, "Movement detection of human body segments: Passive radio-frequency identification and machine-learning technologies," *IEEE Antennas Propag. Mag.*, vol. 57, no. 3, pp. 23–37, Jun. 2015.

[48] J. Han et al., "CBID: A customer behavior identification system using passive tags," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2885–2898, Oct. 2016.

[49] P. Asadzadeh, L. Kulik, and E. Tanin, "Gesture recognition using RFID technology," *Pers. Ubiquitous Comput.*, vol. 16, no. 3, pp. 225–234, Mar. 2012.

[50] R. Zhao, D. Wang, Q. Zhang, H. Chen, and A. Huang, "CRH: A contactless respiration and heartbeat monitoring system with COTS RFID tags," in *Proc. 15th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Jun. 2018, pp. 1–9.

[51] C. Liu, J. Xiong, L. Cai, L. Feng, X. Chen, and D. Fang, "Beyond respiration: Contactless sleep sound-activity recognition using RF signals," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–22, Sep. 2019.

[52] L. Chen et al., "LungTrack: Towards contactless and zero dead-zone respiration monitoring with commodity RFIDs," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–22, Sep. 2019.

[53] R. Chen, X. Huang, Y. Zhou, Y. Hui, and N. Cheng, "UHF-RFID-based real-time vehicle localization in GPS-less environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 9286–9293, Jul. 2022.

[54] J. Yin, S. Liao, C. Duan, X. Ding, Z. Yang, and Z. Yin, "Robust RFID-based multi-object identification and tracking with visual aids," in *Proc. 18th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Jul. 2021, pp. 1–9.

[55] C. Duan, J. Liu, X. Ding, Z. Li, and Y. Liu, "Full-dimension relative positioning for RFID-enabled self-checkout services," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 1, pp. 1–23, Mar. 2021.

[56] Y. Hou, Y. Wang, and Y. Zheng, "TagBreathe: Monitor breathing with commodity RFID systems," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2017, pp. 404–413.

[57] C. Wang, L. Xie, W. Wang, Y. Chen, Y. Bu, and S. Lu, "RF-ECG: Heart rate variability assessment based on COTS RFID tag array," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 2, pp. 1–26, Jul. 2018.

[58] H. Ding et al., "RFNet: Automatic gesture recognition and human identification using time series RFID signals," *Mobile Netw. Appl.*, vol. 25, no. 6, pp. 2240–2253, Dec. 2020.

[59] H. Ding et al., "A platform for free-weight exercise monitoring with passive tags," *IEEE Trans. Mobile Comput.*, vol. 16, no. 12, pp. 3279–3293, Dec. 2017.

[60] H. Kong, L. Lu, J. Yu, Y. Chen, L. Kong, and M. Li, "FingerPass: Finger gesture-based continuous user authentication for smart homes using commodity WiFi," in *Proc. 20th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jul. 2019, pp. 201–210.

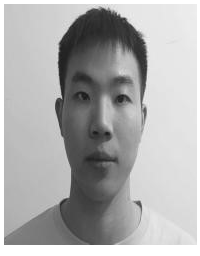
[61] C. Dian, D. Wang, Q. Zhang, R. Zhao, and Y. Yu, "Towards domain-independent complex and fine-grained gesture recognition with RFID," *Proc. ACM Human-Computer Interact.*, vol. 4, pp. 1–22, Nov. 2020.

[62] W. H. Press and S. A. Teukolsky, "Savitzky-Golay smoothing filters," *Comput. Phys.*, vol. 4, no. 6, pp. 669–672, Nov. 1990.

[63] C. Xia, A. Saito, and Y. Sugiura, "Using the virtual data-driven measurement to support the prototyping of hand gesture recognition interface with distance sensor," *Sens. Actuators A, Phys.*, vol. 338, May 2022, Art. no. 113463.

[64] X. Wang, T. Liu, C. Feng, D. Fang, and X. Chen, "RF-CM: Cross-modal framework for RF-enabled few-shot human activity recognition," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 7, no. 1, pp. 1–28, Mar. 2022.

[65] C. Xiao, Y. Lei, Y. Ma, F. Zhou, and Z. Qin, "DeepSeg: Deep-learning-based activity segmentation framework for activity recognition using WiFi," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5669–5681, Apr. 2021.



Zhixiong Yang was born in Ningde, Fujian, China, in 1998. He received the B.S. degree in software engineering from the Changchun University of Technology, Changchun, China, in 2019. He is currently pursuing the Ph.D. degree in computer science and technology with Xinjiang University, Ürümqi, China.

His research interests include indoor localization technology, pattern recognition, and signal processing.

Ziyi Zhen was born in Xinzhou, Shanxi, China, in 1998. She received the bachelor's degree in mathematics and applied mathematics from Shanxi Normal University, Linfen, China, in 2020. She is currently pursuing the Ph.D. degree majoring in mathematics with Xinjiang University, Ürümqi, China.

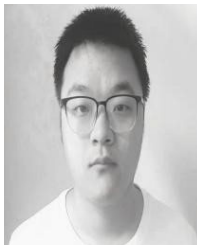
She is mainly involved in the study of fluid mechanics.

Zijian Li was born in Xuzhou, Jiangsu, China, in 1998. He received the bachelor's degree in computer science and technology from Changzhou University, Changzhou, China, in 2020. He is currently pursuing the master's degree in software engineering with Xinjiang University, Ürümqi, China.

His research interests include image processing, big data technology, and signal processing.

Xu Liu was born in Zhoukou, Henan, China, in 1998. He received the B.S. degree in software engineering from Tianjin Polytechnic University, Tianjin, China, in 2020. He is currently pursuing the master's degree in software engineering with Xinjiang University, Ürümqi, China.

His research interests include indoor localization technology, pattern recognition, and signal processing.



Bo Yuan was born in Linfen, Shanxi, China, in 1995. He received the bachelor's degree in computer science and technology from Changzhi University, Changzhi, China, in 2019. He is currently pursuing the master's degree in software engineering with Xinjiang University, Ürümqi, China.

Yajun Zhang was born in Nanyang, Henan, China, in 1983. He received the bachelor's and master's degrees from Xinjiang University, Ürümqi, China, in 2007 and 2010, respectively, and the Ph.D. degree from Shandong University, Jinan, China, in 2021.

He is currently working as an Associate Professor with Xinjiang University. His research interests include indoor localization technology, pattern recognition, and signal processing.