IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 20, 2021, accepted November 12, 2021, date of publication November 15, 2021,
date of current version November 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3128293

RF-Eletter: A Cross-Domain English Letter Recognition System Based on RFID

ZHIXIONG YANG, XU LIU, ZIJIAN LI, BO YUAN, AND YAJUN ZHANG
School of Software Engineering, Xinjiang University, Ürümqi, Xinjiang 830046, China

Corresponding author: Yajun Zhang (yajunzhang369@163.com)

This work was supported in part by the National Key Research and Development Program under Grant 2018YFF01014304; and in part by the Xinjiang Uygur Autonomous Region University Joint Fund under Grant XJEDU2019Y049 and Grant XJEDU2019Y057.

ABSTRACT As the basis of human-computer interaction (HCI), gesture recognition interprets user-performed gestures as commands, followed by the content execution expressed by users' gestures. Gesture recognition through wireless signals denotes a novel branch of human perception. Despite the recent popularity of Radio Frequency Identification (RFID) following specific advantages (lightweight, low-cost, and universality), several intricacies remain unresolved in RFID sensing research. First, most studies performed simplified body movements assessments instead of identifying complex and fine-grained or subtle gestures. Second, users require extensive training in a novel discipline to collect training data in a specific pattern. Given the paucity of an intuitive and effective means of identifying user gestures, the RF-E-letter proposed in this study denotes an RFID recognition system for complex, fine-grained, and domain-independent gestures. A multi-label array was utilized to gather gesture signals. Fine-grained gesture data could be obtained pre-processing with a novel data-processing method. Seemingly irregular RFID phase data could be converted into intuitive images for the deep learning module input as convolutional neural networks (CNNs) encompass automatic extraction characteristics for complex space-time features. The average accuracy of new environments for novel users is 95.6% and 96.6%, respectively (significantly better than current RFID-based solutions), thus demonstrating effectiveness and versatility.

INDEX TERMS Gesture recognition, RFID, tags.

I. INTRODUCTION

Given the rapid development of wireless technologies to locate and track people [33], researchers have been exploring different means of performing fine-grained human perception. Human gesture recognition implies one of the emerging branches of human perception that significantly influences multiple applications [32]: smart homes, virtual reality (VR), sign language recognition, and smart cities [33]–[36]. Gesture recognition facilitates improved user experience compared to conventional techniques. Gesture recognition in public areas, such as hospitals, libraries, supermarkets, and museums could improve user experience without physical contact due to the current health crisis. The non-contact approach could also prevent the spread of germs and bacterial infections. People could also execute specific gestures to interact with smart assistants in 5G smart homes and further promote smart home development.

Traditional solutions involving gesture recognition typically utilize wearable sensors [3]–[9] and cameras [1]–[4], [37] for gesture recognition. Nevertheless, the approaches remain limited despite the high recognition accuracies of such alternatives. Although computerized vision-based gesture recognition through cameras could achieve high gesture recognition accuracy, the camera-based approach is inefficient in dark environments and potentially violates user privacy [10], [11]. Wearable sensor-based approaches utilize inertial sensors, accelerometers, smartphones, tablets, and smartwatches for gesture recognition. In [61], the Myo armband was utilized for electromyographic (EMG) signal acquisition. A simple network structure of a fully connected neural network was employed to achieve seven specific gesture recognition. Although the method could be effective and accurate for hand gesture recognition, wearing such sensing devices might be

The associate editor coordinating the review of this manuscript and approving it for publication was Antonio Piccinno.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

practically inconvenient. For example, some older adults and children might refuse to wear the devices or forget them at home.

This study explored a flexible and deployable gesture recognition mechanism. Generally, wireless signal-based solutions recognize gestures as a particular user gesture that would affect the wireless channel and cause RF signal changes (amplitude or phase). Through specific gesture data acquisition, data pre-processing, and gesture data feature extraction [19], [20], the system could perform gesture recognition with the extracted features: dynamic time warping (DTW)-based template matching or distance-oriented classifiers. Regardless, both components require extensive and labor-intensive data pre-processing and feature extraction. Gesture recognition performance is also highly dependent on the selection of feature extraction algorithm.

The recent development of deep learning has rapidly advanced the CNNs in natural language processing (NLP), image recognition, and other disciplines [48], [49]. As CNNs could automatically learn and extract complex features, the possibility of automatically extracting the feature representation of gesture signals from complex time-series wireless signals is questioned. In this vein, a device-free, fine-grained, and domain-independent (different environments and users) gesture recognition system named RF-Eletter was designed. Following Figure 1, the tag array and antenna were placed facing one another while users drew specific letters in between. The RF-Eletter subsequently mapped the captured signal to the performed gesture. In the experiment, volunteers freely drew the given letters between the antennas and tag arrays in three different places: dormitory, conference room, and classroom.

FIGURE 1. RF-Eletter's lab scenario deployment.

Three barriers were encountered in designing the RF-Eletter. Although the volunteers were not statically placed (immobile) between the tag array and antenna in the letter-drawing process, the users dynamically performed letter-drawing that involved complex, diverse, and fine-grained gesture transformations as many letters encompassed similar hand movements (c vs. l). Nevertheless, current commercial RFID readers provide signal (received signal strength (RSS) and phase) indicators with limited spatial resolution. As hand gesture recognition involves complex spatio-temporal transformations, some subtle finger movements might be difficult to identify. Due to the multipath effects within the experimental environment, the additional noise produced in the collected data caused complexities in determining the letters drawn by users.

Compared to the conventional target recognition activity using Received Signal Strength Indication (RSSI), the phase reflected higher fine-grained resolution. Given that RSSI is highly influenced by multi-path effects, the spectrograms were compared through RSS and phase. The phase was utilized for letter recognition works based on significant gesture recognition features. The final barrier involved domain-independent feature extraction works. Typically, RF signals carry much gesture-independent information with high dependence on users and their environment (referred to as domain in this study). As a gesture recognition model trained by users in one environment would be significantly less accurate in a novel domain, domain-independent feature extraction denotes a highly challenging issue. The RF signal attributes also instigate complexities in implementing domain-specific feature extraction.

Several solutions were identified in the raw signal and deep learning network contexts to address the aforementioned intricacies. As current commercial RFID readers could only provide limited information, sensing capabilities could be optimized with two-dimensional multi-tag arrays. In this study, a 2*3 two-dimensional tag array was deployed where the spatial distribution could capture complex and fine-grained gestures. Given the raw noise in the collected data that is not easily recognized, smoothing, subtraction and normalization operations, and data expansion were performed on the gathered data to further optimize RF signal sensing capability. A novel idea was proposed to convert the raw phase data into a 100*100 pixel picture as the conversion could intuitively represent the executed gesture as the deep learning module input. A multimodal CNN was designed to extract features and complex temporal and spatial features in each tag. Spatio-temporal warping was employed to extract modal features across time and space by extracting phase features at higher levels. Accurate gesture recognition is thus achieved by performing feature analysis.

This study experimented with three different scenarios (conference rooms, classrooms, and dormitories) and invited three volunteers (two men and one woman) to collect approximately 6,000 sample sets for system evaluation. Notably, the distance between the label array and antenna denotes the distance set by the actual experiment.

The current study contributions could be summarized as follows: (1) the RF-Eletter is device-free as the system employs a (i) multi-tag array (2*3) and a (ii) multi-modal CNN for feature collection and extraction to sense complex and fine-grained gesture data; (2) RF-Eletter adopts a novel approach to convert the original low-resolution RFID phase

data into a more intuitive and high-resolution image as the deep learning module input, thus facilitating the conversion of complex and low-resolution gesture data while retaining gesture-specific information; (3) dynamic, complex, fine-grained, and device-free continuous gesture recognition is attained on commercial RFID devices. Based on the results, RF-Eletter is deemed to be flexible, deployable, and highly accurate in recognizing different alphabetic gestures with an overall accuracy of 96.1%.

The remaining sections in this article are organized as follows: Section 2 presents the research overview, Section 3 describes the preliminary work, Section 4 thoroughly explains the posture recognition system design, Section 5 discusses the experimental method implementation and evaluation, Section 6 elaborates on the study experiment and future research prospects, and Section 7 concludes with future studies and the study summary.

II. RELATED WORK

A. GESTURE RECOGNITION TECHNOLOGY

Current gesture recognition is divided into three categories: wearable-based gesture recognition, computerized vision-based gesture recognition, and wireless technology-based gesture recognition. Wearable sensor-based approaches utilize sensors that are embedded in sensing devices to capture hand and finger movements. For example, [16], [50] employed inertial sensors that were embedded into a bracelet to recognize eating and smoking gestures while [15] integrated gyroscopes and accelerometers with a glove and utilized the glove to track subtle finger movements. Other scholars employed bracelets for fine-grained gesture recognition [5], [18], [61]. In [61], the Myo armband was placed on the forearm for gesture signal acquisition and duly processed and utilized as input for a fully connected network to recognize gestures. Despite the prevalence of real-life wearable sensing devices, people inevitably forget to wear the devices or experience discomfort.

Computerized vision-based gesture or human recognition systems employ cameras or light sensors [21]–[24] to recognize gesture movements or humans. In [22], the RGB camera in a mobile device was utilized to recognize gestures. In the deep learning environment context, computerized vision-based methods were significantly optimized and incorporated into activity and gesture recognition accuracies [1], [2]. Some researchers also utilized Kinect [26] and Leap Motion [27], [28] to further improve gesture recognition performance. Nevertheless, the systems are susceptible to variations in lighting conditions that do not apply to occlusion-oriented situations. Although computerized vision-based approaches encounter specific intricacies, such as the invasion of user privacy, RF-Eletter is independent of illumination-based conditions due to its lightness, scalability, and pervasiveness.

Wireless infrastructure could provide device-free gesture recognition given the prevalence of gesture recognition with wireless technology. The Wi-Fi [12]–[14], RFID [38]–[40], ultrasound [41], [42], radar [43]–[45], and other wireless technologies have been employed for gesture recognition, such as the channel state information (CSI) of WiFi utilized in reference [12] to whole-body activity and coarse-grained gesture recognition. For example, [14] and [22] employed WiFi signals to recognize common sign language and gesture recognition. In [41], LLAP utilized a microphone and speaker to precisely recognize hand-tracking in the millimeter range. Regarding wireless-based positioning and tracking technologies, some researchers tracked object movements with RF signal analysis [29], [30]. Witrack [29] employed the frequency-modulated continuous-wave (FMCW) technology to track target personnel with unique and complex equipment. Although target gesture tracking was achieved by eliminating the reflection from surrounding objects in [30], fine-grained gesture recognition could not be attained as such methods typically require specific equipment for gesture recognition. Consequently, IoT devices were incorporated into various applications [31]. Overall, RF-Eletter facilitates device-free fine-grained gesture recognition as the system is incorporated into commercial RFID devices.

B. RFID-BASED PERCEPTION

The RFID is generally utilized for object identification. For example, [46], [47] employed RFID signals to determine target materials, specifically liquid. Notwithstanding, recent studies have revealed RFID signals to be information-rich for localization [51]–[53], activity identification [54], [55], human identification [56], [57], and vital sign detection [58]–[60]. Specifically, [60] proposed the LunkTrack system that performs breath detection on a commercial RFID device without any equipment placed on the target. Detection could be accomplished through receivers' signal fluctuations following chest movements while breathing. Meanwhile, optimization techniques serve to locate multiple RFID tags and facilitate the system to monitor the breathing of two target individuals.

The TagSleep system designed in [59] proposed a two-layer sensing concept with breathing information as the first layer to obtain rich second-layer sensing information, including sleep activities (coughing, snoring, and sleep talking) and detect the target individual's breathing status through subtle changes. Meanwhile, RFIDraw [55] was utilized for a high-precision tracking of subtle hand movements. Some researchers explored tagless sensing to avoid the inconvenience of physical body tagging. For example, TagFree [54] extracted signal angle-of-arrival (AOA) information from a multi-antenna array to achieve device-free activity recognition albeit with AOA information processing. Although current device-free RFID sensing solutions emphasize the recognition of coarse-grained or simple gestures rather than fine-grained and complex counterparts, accuracy tends to decrease dramatically in novel environments or with new users despite being domain-specific.

C. RFID-BASED GESTURE RECOGNITION

Gesture recognition with wireless signals, such as RFID is an emerging non-touch user interface technology that has garnered much attention following its lightness, affordability, and prevalence. CAO DIAN *et al.* [25] proposed RFree-GR, a domain-independent RFID system that utilizes a 3*4 array of tags to capture users' gesture signals through the designed multi-modal convolutional neural network (MCNN) to aggregate information between signals, abstract complex spatio-temporal patterns, and facilitate complex and fine-grained gesture recognition. Despite an average accuracy of 90% for novel users and environments, the system failed to recognize dynamic gestures. Meanwhile, Ding *et al.* [17] recommended RFnet to recognize static or dynamic gestures with time-series RFID signals through extensive experiments in three environments. Resultantly, RFnet achieved an average accuracy of 94.8% in dynamic gesture recognition. The system reflected high recognition accuracy for dynamic gestures albeit with reliance on a sensing plane that encompassed an array of 7*7 tags.

Multiple tags might instigate coupling effects between tags and cause the sensing plane to occupy a large space. In alleviating such effects and minimizing the occupied area, a 2*3 tag array was adopted as the sensing plane. The RF-Eletter, a framework based on multi-branch CNN, was also proposed as a novel approach to convert the raw low-resolution RFID phase data into a more intuitive and high-resolution image as the deep learning module input. The input subsequently enabled the conversion of complex low-resolution gesture data while retaining gesture-specific information. The model was validated with extensive experiments. Based on the study outcomes, the RF-Eletter system was found to be flexible, deployable, and highly accurate in dynamic gesture recognition.

III. PRELIMINARIES

This section presents some fundamental RFID technology principles for an optimal RFID sensing model.

A. RF TECHNOLOGY PRINCIPLES

The RFID tags are utilized to perceive target environments in RFID technology. When the tag accesses the antenna receiving range, the antenna subsequently senses the RF tag product information in the environment through the RF signal emitted by the reader. The read and decoded information is then conveyed to the central information system for relevant data processing. As a natural reflecting and receiving medium, the human body encompasses a wide range of applications in RFID. In the simulation environment of the RFID system, the signal emitted from the antenna would be disrupted by various media in the environment and induce different waveforms. The RF-Eletter system parallels the RFID technology principle and deploys a tag array to capture human finger movements and complete the gesture recognition spectrum.

1) RFID-AWARE TECHNOLOGY PRINCIPLES

In RFID systems, the reader transmits an RF signal from the antenna for target label access, derives energy from the RF signal to respond, and returns the backward-scattering signal $S(t)$ to the reader as follows:

$$\mathrm{S}(\mathrm{t}) = \alpha(\mathrm{t})\,\mathrm{e}^{-\mathrm{i}\theta(\mathrm{t})} \tag{1}$$

$$\theta(t) = \theta_0 + \frac{4\pi d}{\lambda} \tag{2}$$

Specifically, $\alpha(t)$ and $\theta(t)$ denote the receiving signal amplitude and phase, respectively, θ_0 implies the initial offset, i is an imaginary unit, d reflects the propagation path length, and λ demonstrates the wavelength. A dynamic signal impacts the signal change when the user gestures between antennas and labels. Static signals beyond user influence are also identified. Thus, $S(t)$ can be computed as follows:

$$S(t) = S_s(t) + S_d(t) = \leftrightarrows \alpha_s e^{-i\theta_s} + \alpha_d(t)\, e^{-i\theta_d(t)} \tag{3}$$

The S_s and S_d represent static and reflective dynamic signals, respectively. Human reflection causes shifts in phase and receiving signal. As amplitude is not a primary factor in this experiment, dynamic amplitude(α_d) and static amplitude (α_s) are perceived as constants and (3) could be computed as follows:

$$S(t) = \alpha\left(e^{-i\theta_s} + e^{-i\theta_d(t)}\right) \tag{4}$$

The final $S(t)$ result is conclusively obtained.

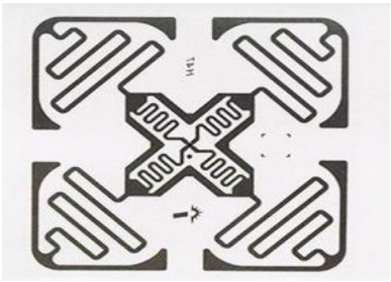
2) PRELIMINARY EXPERIMENTS

A set of 2 × 3 tag arrays were utilized to study the effect of human motion on the back-scattered signal while 4 × 3 tag arrays were employed for data acquisition following [25]. Resultantly, 2 × 3 tag arrays enabled the acquisition of complex and fine-grained gesture data. Dynamic gestures were also performed between the tag arrays and antenna to collect RSS and phase values. A high coupling effect was identified when the tags were placed closer to one another. As the spacing between the tags deployed in this study exceeded 15 cm [59], higher accuracy was attained with fewer tags.

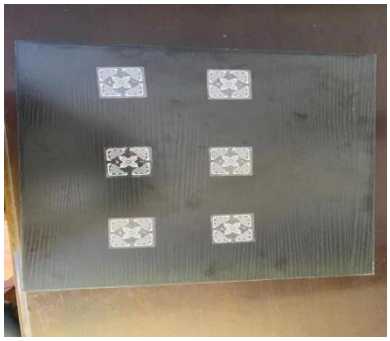
B. SETTINGS OF THE EXPERIMENTAL DEVICE

Two fundamental questions are posed before conducting a human gesture recognition experiment: (1) How to select and set labels to capture motion gesture information? and (2) How can the distance between the antenna and label be controlled? Based on past research, the H47UHF label was selected and formed into a 2 × 3 label array to precisely cover human gestures and produce more accurate signal changes. The label array of 2 × 3 could also better eliminate coupling effects with higher accuracy and fewer labels. Figures 1a and b demonstrate the H47UHF labels and label array settings employed by the RF-Eletter system, respectively.

Specifically, volunteers were only required to move their fingers slightly in drawing the corresponding letters without overstretching the label to the antenna (direct distance) and controlling the label-antenna distance in RF-Eletter at 35 cm.



(a)



(b)

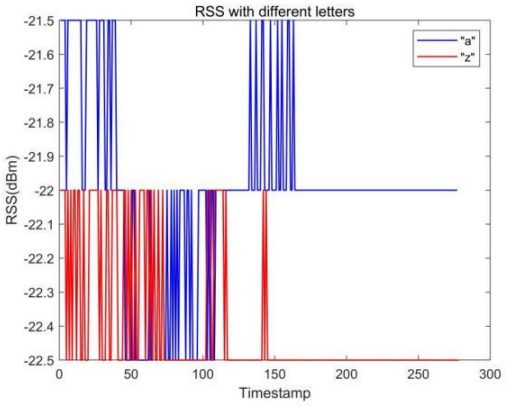
FIGURE 2. a) H47UHF label. b) Experimental label array.

C. GESTURE AWARENESS ANALYSIS

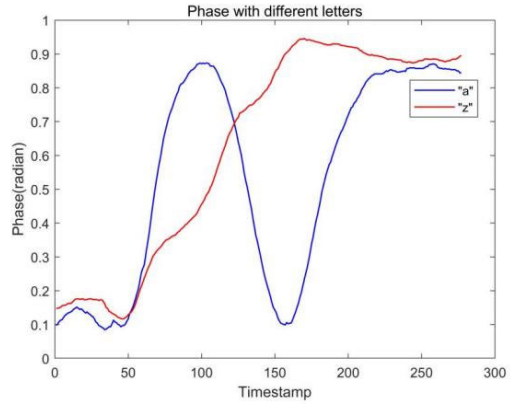
This study analyzed the perceptions of human gesture recognition. Figures 3a and b illustrate the RSS and phases under various dynamic letters, respectively. Different gestures exhibited distinct waveforms with higher recognition compared to RSS. Generally, the signal received by the antenna from the label contained human gesture information that could be analyzed by docking the signal to deduce human body gestures. In Figure 3a, the RSS in human gesture perception demonstrated a less obvious trend of change. Thus, a phase was chosen in this experiment to perform the corresponding experimental analysis of human gesture recognition [25].

Different users' human gesture recognition capabilities were also studied. Figures 4a and b represent the RSS and phase maps, respectively when various users wrote the same dynamic gesture under the same label. Specifically, different users demonstrated distinct characteristics for the same gesture. In this experiment, as many gestures as possible were gathered from different users to increase dataset diversity. In comparing the RSS and phase spectrums, phase was found to contain palpable and recognizable human gesture characteristics.

The ability of different labels to recognize human gestures was also examined. Figure 5 represents a phase map where the same user wrote the same dynamic gesture under different labels. Resultantly, different labels reflected different variations in the user's characteristics. Based on the user's time-varying characteristics, a 2 × 3 label array was implemented to minimize the perceived blind spot of the RFID signal.



(a)



(b)

FIGURE 3. a) RSS information under different letters. b) Phase information under different letters.

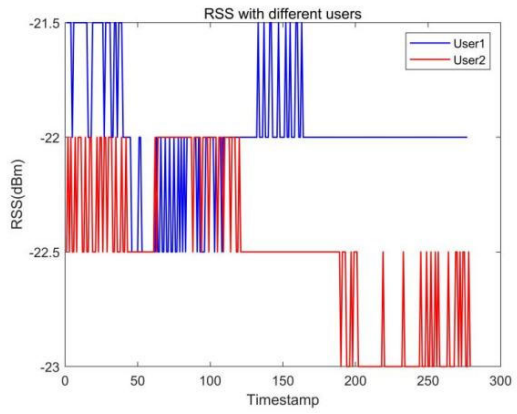
Based on the study experiment, label 5 was selected as the experimental analysis label given that tag 5 was more recognizable to the user's gestures. For example, the volunteers habitually utilized five labels as references when writing gestures. In comparing the RSS and phase spectrums, phase was found to contain palpable and recognizable human gesture characteristics.

IV. SYSTEM DESIGN

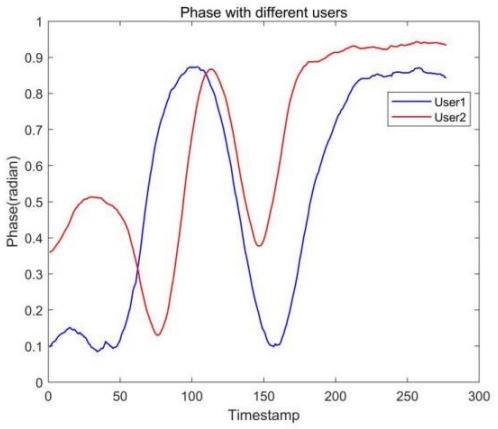
This section provides an overview of the study system (see Section 4.1) and thoroughly elaborates on the core system modules.

A. OVERVIEW OF THE SYSTEM

Following Figure 6, RF-Eletter primarily encompasses three modules: signal acquisition, signal pre-processing, and deep learning. In the signal acquisition module, a 2 × 3 RFID tag array was employed to capture the original RSS and phase values. Based on the original signal analysis, the phase was finally selected as the characteristic signal input follow-up module. In the signal pre-processing module, the original phase was subtracted to highlight the differences between different gestures and normalize the processed signal data



(a)



(b)

FIGURE 4. a) RSS information for different users. b) Phase information for different users.

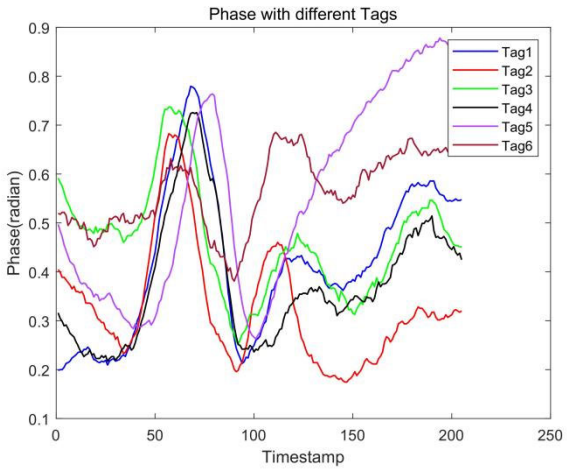


FIGURE 5. Phase information under different labels.

into the data-smoothing operation for low noise interference. Additionally, the data was expanded to further improve data diversity and output for conversion into the image format. Lastly, the pre-processed signal data were incorporated into the deep learning module for gesture recognition.

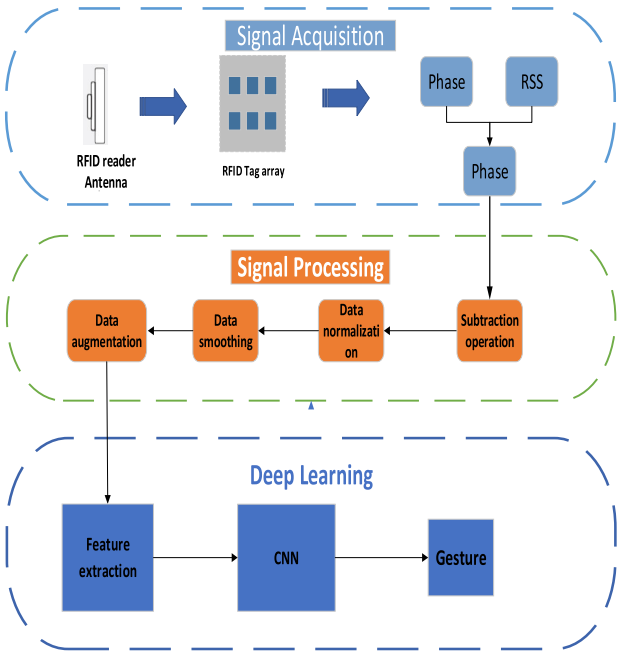


FIGURE 6. Overview of RF-Eletter systems.

B. SIGNAL PRETREATMENT MODULE

Owing to environmental noise and interference, it was deemed inappropriate to directly integrate the original signal with the neural network for training. Thus, the signal was initially incorporated into the pre-processing module for optimal signal recognition.

1) SUBTRACTION OPERATIO

Based on the antenna-tag array distance, reader, and tag (physical characteristics of the impact), the original signal was frequently accompanied by environmental noise in the output. In this vein, the signal output significantly differed from the anticipated counterpart. The RFID acquisition of phase denoted the super-position of LOS and NLOS signals. As the NLOS signal primarily occurred through static (tables and chairs) and dynamic (human gesture motions) reflection within the RFID hardware, the signal was initially gathered in an empty environment without the dynamic reflection signal. Gesture motion was then performed between the RFID and tag for signal collection by subtracting both signals [17]. The dynamic reflection signal could be extracted following human gesture motions and the reflected dynamic signal caused by human gesture movements. The signal was then utilized as the characteristic signal of human gesture movements. The noise interference from physical characteristics, such as the environment could be effectively omitted for a distinct feature curve.

2) DATA NORMALIZATION

In RFID systems, the original phases gathered might demonstrate different scale units following the nature of the label and position of human gesture movements to alleviate data

comparability and gesture recognition accuracy. Data normalization operation aimed to normalize the original data to the same order of magnitude, resolve the comparability between data indicators, and improve the model convergence rate and accuracy. Specifically, the minimum-maximum standardized method was implemented for a linear transformation of the original data (X_1) output as follows:

$$X_1 = \frac{X - min}{max - min} \tag{5}$$

Specifically, X denotes the subtraction processed data, max implies the maximum value of the current sample data, and min reflects the minimum value of the present sample data.

3) DATA SMOOTHING

Excessive noise from raw data is a common challenge in gesture recognition experiments, such as the original data "a" output from tag 1 (see Figure 7). Specifically, output "a" was merely smoothed due to the jittery signal and complexities in distinguishing accuracy.

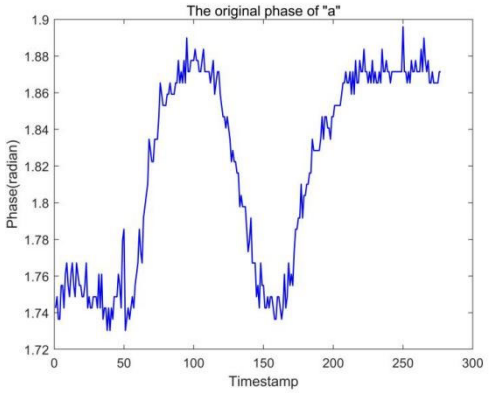


FIGURE 7. Label 1 outputs the original phase information for the letter "a."

The fundamental notion of data smoothing could be summarized as "hijacking the rich and the poor". In other words, the probability distribution was inclined to the actual level as much as possible for increased zero probability (low probability) and reduced high probability. In this study, the data were smoothed with a Savitzky-Golay filter that proved more appropriate for RFID following data change dominance. The width of the filter window was set to $m = 2n + 1$, the prediction point was $x = (t - n, t - 1, \cdots, t, t + 1, t + n)$, and the data in the window was fitted with a $k-1$ polynomial. The fitting formula is presented in (6) as follows:

$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_{k-1} x^{k-1} \tag{6}$$

In obtaining a similar solution and $2n + 1 \geq k$, the least squares were obtained to determine the fitted parameter A as presented in (7):

$$\begin{pmatrix} y_{t-n} \\ \vdots \\ y_{t-1} \\ y_t \\ y_{t+1} \\ \vdots \\ y_{t+n} \end{pmatrix} = \begin{pmatrix} 1 & t-n & (t-n)^2 & \cdots & (t-n)^{k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t-1 & (t-1)^2 & \cdots & (t-1)^{k-1} \\ 1 & t & t^2 & \cdots & t^{k-1} \\ 1 & t+1 & (t+1)^2 & \cdots & (t+1)^{k-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t+n & (t+n)^2 & \cdots & (t+n)^{k-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{k-1} \end{pmatrix} + \begin{pmatrix} e_{t-n} \\ \vdots \\ e_{t-1} \\ e_t \\ e_{t+1} \\ \vdots \\ e_{t+n} \end{pmatrix} \tag{7}$$

The matrix above is simplified as follows:

$$Y_{(2n+1)\times 1} = X_{(2n+1)\times k} \cdot A_{k\times 1} + E_{(2n+1)\times 1} \tag{8}$$

In the aforementioned formula, Y, X, A, and reflect the matrix representations of the formulas in (7) while the subscripts represent their respective dimensions. For example, $A_{k\times 1}$ represents a parameter with k rows and 1 column. The solution of $A_{k\times 1}$ could be derived from the least square method as follows:

$$A = \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot Y \tag{9}$$

The aforementioned mark, T, indicates transposition. The predicted or filtered model Y value is computed as follows:

$$P = X \cdot A = X \cdot \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot Y = B \cdot Y \tag{10}$$

Lastly, a matrix of the filter value-observation relationship is formulated as follows:

$$B = X \cdot \left(X^T \cdot X\right)^{-1} \cdot X^T \tag{11}$$

The observations could be quickly converted into filter values for data smoothing upon obtaining matrix B.

4) DATA AUGMENTATION

Neural network training requires large data to mitigate value loss and obtain high accuracy. Convolutional networks tend to overfit in managing small datasets. To prevent overfitting and attain a better training effect, the study dataset required expansion for high accuracy. In the study system, the data deformation and adjustment method for increased dataset optimized data diversity and robustness. Specifically, the data output was smoothed as a 100 × 100 pixel image. The literature [17] uses GAN for data expansion. Regardless, the expanded data image resulted in biased training data. As the effect of incorporating the biased data into the deep learning module could not attain the anticipated effect, noise was embedded into the converted image to expand the training set

and convey the data to the deep learning module as network input.

C. DEEP LEARNING MODULE

The collected gesture signals were pre-processed and utilized as the deep learning module input to implement feature extraction and gesture recognition. This section introduces the deep learning module in detail (see Figure 8). The framework diagram of this designed network encompasses two parts: feature extractor and gesture recognizer.

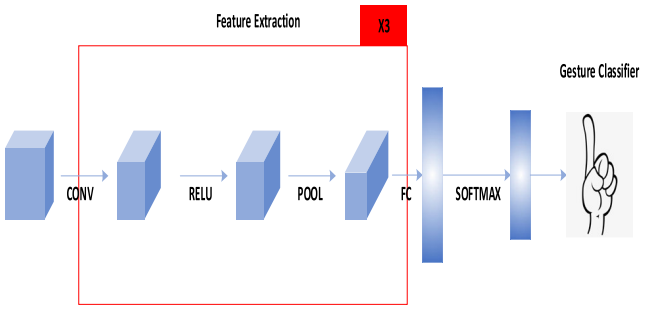


FIGURE 8. CNN frame.

1) FEATURE EXTRACTION PHASE

A three-tier CNN network structure was employed for feature extraction. The data were pre-processed for input to the deep learning module. The CNN was utilized following its suitability in analyzing the time-series of sensing data and optimizing feature extraction from the fixed-length segment of the whole sequence. The convolution layer output characteristics are defined as follows:

$$X = POOL\left(ReLu\left(WX0 + b\right)\right) \tag{12}$$

This study utilized a three-layer convolution. The convolution core size was 3 × 3, the number of data input channels was three, the number of channels in the first layer convolution was eight, the number of channels in the second-layer convolution was 16, and the number of channels in the third convolution was 24. Through the three-layer convolution, different dimensional phase features could be extracted from the pre-processed image.

After the image passed through the convolution layer, the ReLU activation layer [19] was employed to increase the nonlinear segmentation ability of the image and apply the average pooling operation to the activated image along the time dimension. As such, the network parameters could be reduced to prevent overfitting through pooling operation. A high-level phase representation was extracted using CNN as follows:

$$RM = FE\left(X, \theta fe\right) \tag{13}$$

Specifically, x denotes a picture of the pre-processed data while θfe represents all the feature extraction parameters.

2) GESTURE RECOGNITION STAGE

As the advanced feature representation of gesture data was obtained post-feature extraction, the gesture recognizer could simply employ the complete connection and softmax layers for gesture classification.

$$GP = GR\left(RM, \theta gr\right) \tag{14}$$

The θgr implies all the gesture recognizer (GR) parameters. Additionally, the cross-entropy loss function could be implemented to compute the LG loss between the predicted GP and ground truth (GT):

$$L_g\left(\theta_{fe}, \theta_{gr}\right) = -\frac{1}{N}\sum_{1}^{N}\sum_{1}^{J} G_P^{ij}\, log\left(GT_P^{ij}\right) \tag{15}$$

Notably, N and J denote the number of gesture samples and categories, respectively.

V. IMPLEMENTATION

This section presents the hardware and software employed in this experiment with RFID equipment and multi-label arrays to test and verify the model accuracy and performance. Based on the experimental parameters (see Table 1), the study model was designed based on different novel domains and various users with average accuracies of 96.6% and 95.6%, respectively.

A. EXPERIMENTAL ENVIRONMENT

Three experimental sites were designed in this study. Experimental site 1 was arranged in a classroom area of approximately 7m × 10m with several tables and chairs (see Figure 9). Experimental site 2 was arranged in a conference room area of approximately 7.2m × 8.5m with much electronic equipment (see Figure 10). Experimental site 3 was arranged in a dormitory area of approximately 4.5m × 8.5m with four iron bunk beds (see Figure 11). The label array was designed as a 2*3 two-dimensional multi-label array with a target in the antenna and label array at a 35 cm distance. Regarding gesture recognition, the user sat in front of the desk and performed specific gestures between the antenna and tag array for users' gesture movements, data extraction, and identification by transmitting Ethernet to the PC side.

B. HARDWARE FACILITIES

The hardware encompassed four parts: an Impinj R420RFID reader (see Figure 12) operating at 920.875MHz, an RFID UHF circular polarization antenna (see Figure 13), six 4 cm × 4 cm labels, and a Lenovo R7000p computer.

C. SOFTWARE FACILITIES

The study model was operated on a Lenovo computer equipped with 2.5GHz AMDR7 and 16G memory (for data acquisition and pre-processing), RFID card readers connected to laptops through Ethernet cables, and low-level card reader protocols (LLRPs) for communication. The method

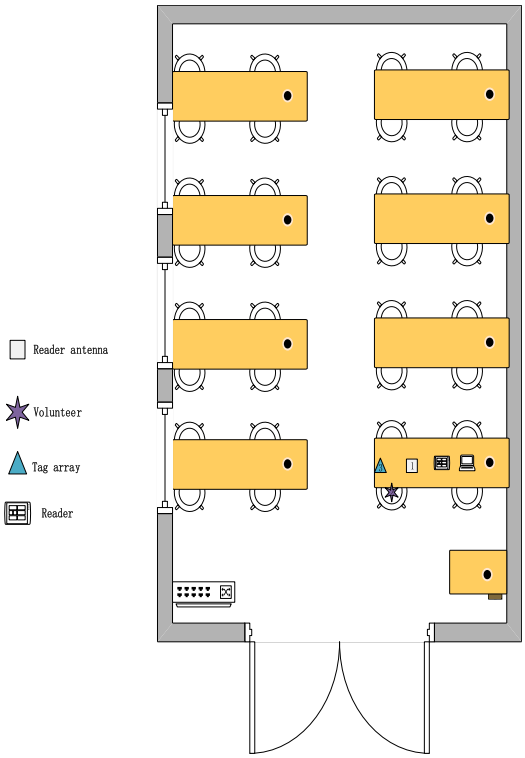


FIGURE 9. Classroom experiment scene diagram.

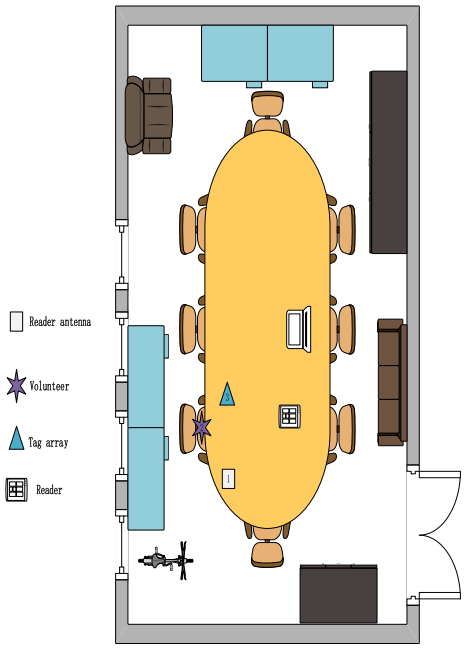


FIGURE 10. A picture of the lab scene in a conference room.

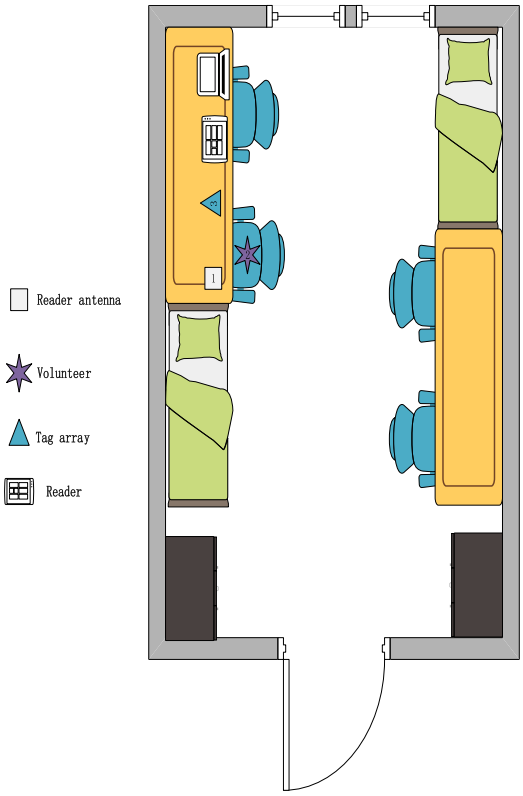


FIGURE 11. Dormitory experiment scene map.

FIGURE 12. Impinj R420 RFID reader.

FIGURE 13. UHF circular polarization antenna.

was implemented by c while the designed neural model was implemented with Python.

D. EXPERIMENTAL DATA SET

A total of 15 volunteers (eight males and seven females) were invited for data collection. Specifically, 10 of the individuals were randomly employed for dataset training while the remaining five encompassed the test set (three males and two females). Each volunteer drew seven different letters in three distinct scenes. Drawings c and l produced more similar images while i and w produced images that significantly differed. All seven letters were utilized for drawing as the letters proved sufficient to cover the 26-letter differences

TABLE 1. Experimental parameter table.

Parameter	Value
Spacing between adjacent labels	5cm
Spacing between tag and antenna	35cm
Angle between antenna and ground	90℃
Reader frequency	920.875MHz

between them. Each letter was drawn 25 times. Lastly, the collected dataset was expanded by embedding noise into it thrice. Out of the 23, 625 gathered data (25*7*15*3*3), 15, 750 (25*7*10*3*3) were training set samples while 2625 (25*7*5*3) were test set counterparts.

E. METRICS

Two evaluation metrics accuracy and correct recognition (*FRR*) were employed to describe the model performance. The *ACC* measures the likelihood of precisely recognizing users' gesture movements as follows:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}} \tag{16}$$

The *TP* represents the number of accurately predicted positive examples while *TN* reflects the number of accurately predicted negative examples. The *FP* represents the number of inaccurately predicted positive examples while *FN* represents the number of negative inaccurately predicted examples. The *FRR* implies the model likelihood of precisely identifying a word as 'yes' and is computed as follows:

$$FRR = \frac{TP}{TP + FN} \tag{17}$$

F. PERFORMANCE IN DIFFERENT ENVIRONMENTS

Comparative experiments were performed to authenticate the cross-domain performance of the study model. Three environmental types were established without loss of generality. Environment 1 involved a classroom with four tables and chairs in an area of approximately 7m × 10m. Environment 2 encompassed a conference room with an area of approximately 7.2m × 8.5m and multiple electronic devices interfering in the experimental environment. Environment 3 involved a dormitory encompassing an area of approximately 4.5m × 8.5m with four iron bunk beds. Each user drew 175 gestures per environment (25 per letter). The FRR rate of the RF-Eletter system was evaluated using the model (see Figures 14 and 15). The matching model accuracies in conference rooms, dormitories, and classrooms were 93.14%, 97.14%, and 100%, respectively. Following the interference of multiple electronic devices in the conference room, the

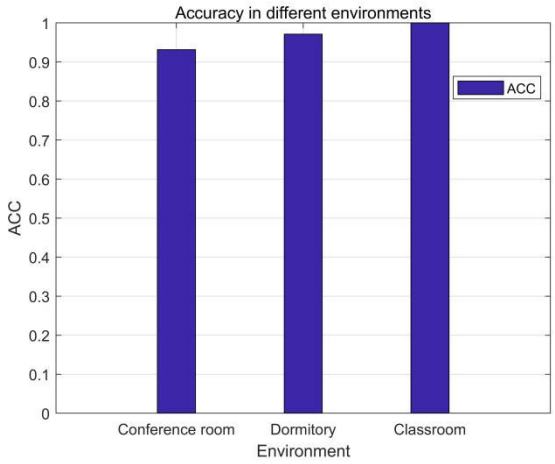


FIGURE 14. ACC in different environments.

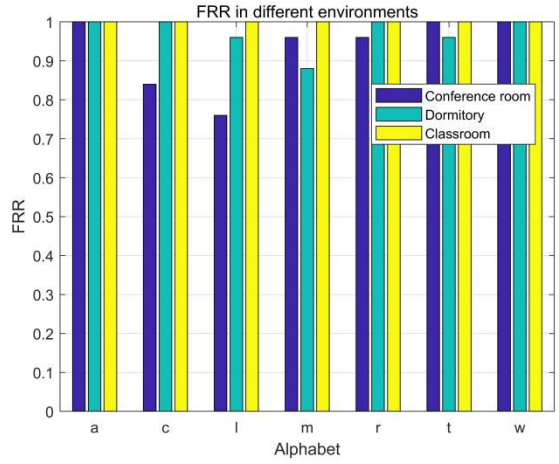


FIGURE 15. FRR in different environments.

phase characteristics of the words c and l were similar while the FRR rate was under 90%. Regardless, the average accurate recognition rate of other words exceeded 95%.

G. PERFORMANCE UNDER DIFFERENT USERS

The impact of different users' performance regarding the same gesture on the channel was explored despite performing the same action. Regardless, different users demonstrated various signal impacts following distinct letter-drawing speed, size, and other relevant factors. In authenticating the RF-Eletter model accuracy for different users' gesture recognition, five volunteers were invited to participate in this experiment. The participants executed seven letters in Environment 2 by drawing each letter 25 times. The FRR rates of the RF-Eletter system for different users' gesture recognition are presented in Figures 16 and 17.

The average matching accuracy of recognition among the five volunteers with the RF-Eletter model was 100%, 98.28%, 89.14%, 96%, and 96.57%, respectively. The accuracy of user 3 was under 90% as the phase maps of c and l that were

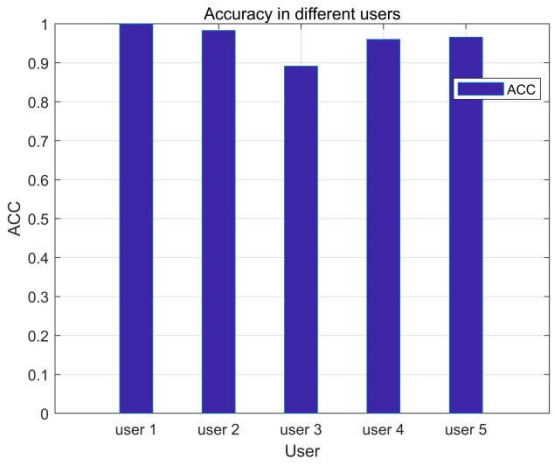


FIGURE 16. ACC for different users.

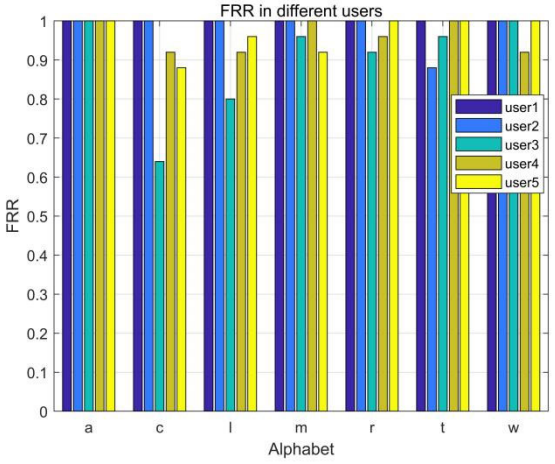


FIGURE 17. FRR for different users.

drawn by the user were more similar and resulted in lower model accuracy. The FRR rate performance results could be divided into two groups. The first group encompassed gesture activities a, m, r, t, and w with an average FRR rate exceeding 97%. The second group involved gesture activities c and l with an average FRR rate of 88.8% and 93.6%, respectively. Compared to the first group of activities, the measurement accuracy of gesture recognition was mitigated as the c and t gestures were highly similar during the drawing process. Although the phase graphs were not significantly different, the system still provided a correct rate exceeding 88.8%.

H. PERFORMANCE UNDER DIFFERENT INTERFERENCE FACTORS

The system design and production are influenced by metal and electronic products as RFID tags and other electronic products are susceptible to interference from metals and other wireless signals. Metal would cause eddying currents around the over-clocked RFID tags and readers and reduce the overall effectiveness of the RFID electromagnetic field. The surrounding objects would also reflect RFID signals and cause interference. In verifying the model performance under the interference factors, three scenarios were established for comparison. Regarding Scenario 1, the user performed a gesture action in Environment 3 without interference factors. Regarding Scenario 2, a metal water cup was placed close to the tag and reader. The user performed the gesture action while the metal water cup caused interference.

Concerning Scenario 3, the user performed a gesture with multiple people walking around the house and causing interference. The user performed such gestures 175 times (25 times for each letter) in every scenario. The experimental results are presented in Figures 18 and 19. For each different scenario, the average accuracy reflected 100%, 98.28%, and 97.14%, respectively. The accuracies measured in Scenarios 2 and 3 proved similar to the counterpart without interference factors (Scenario 1), thus indicating the high resistance of RF-Eletter against environmental interference.

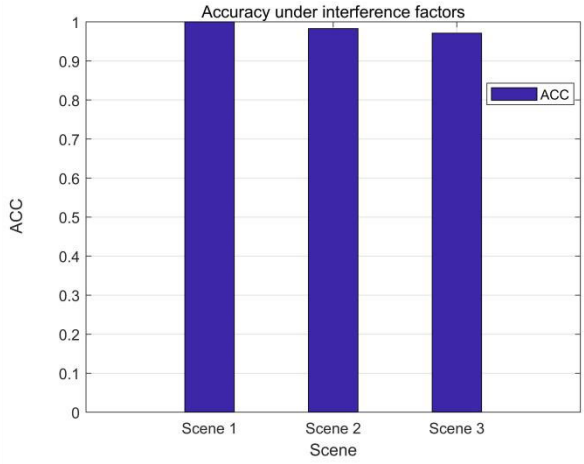


FIGURE 18. ACC under interference factors.

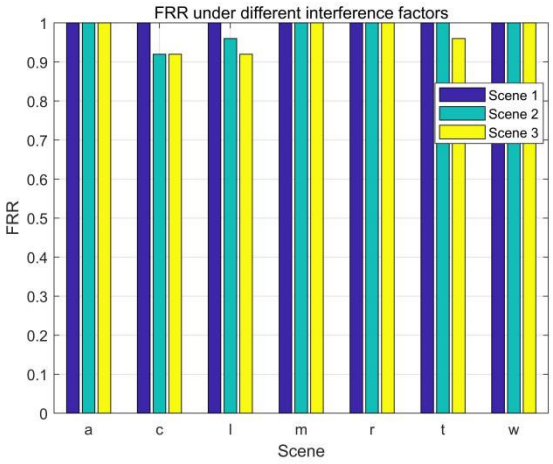


FIGURE 19. FRR under different interference factors.

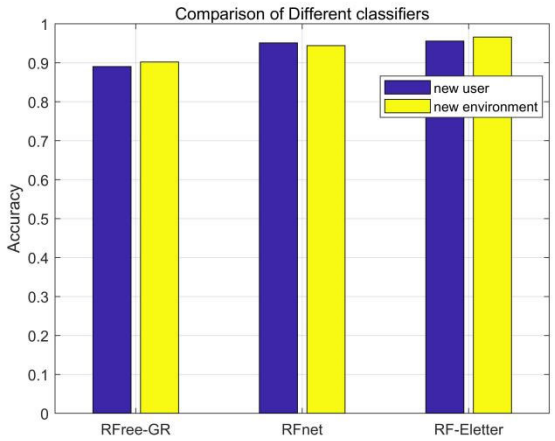


FIGURE 20. Comparison of different classification algorithms.

I. COMPARISON WITH OTHER GESTURE RECOGNITION ALGORITHMS

A novel method was adopted in RF-Eletter to convert the initial low-resolution phase data into more intuitive and high-resolution pictures. The pictures were utilized as the training model input of the CNN network for gesture recognition. A series of comparative experiments were constructed to verify the gesture recognition algorithm effectiveness. Several advanced gesture recognition systems were compared based on wireless signals, such as RFree-GR [25] and RFnet [17] (see Figure 20). The average accuracy rates of RFree-GR for novel users and environments were 89.03% and 90.21%, respectively (lower than the two preceding counterparts). The RFnet utilized a 7*7 tag array to increase the average accuracy of novel users and environments to 95.1% and 94.4%, respectively. Contrarily, RF-Eletter demonstrated the most optimal performance in both tests at 95.6% and 96.6%, respectively with fewer tags for higher accuracy.

VI. DISCUSSION

The RF-Eletter system encountered some limitations in real-life applications. First, more gesture data were required to practically meet specific application prerequisites. Additionally, the designed neural network required additional parameters for real-time gesture recognition. This study aimed to employ the simple fully connected neural network in [61] for real-time gesture recognition and improve the gesture recognition algorithm in the future. In this vein, the MUSIC (multiple signal classification) algorithm was deemed appropriate. The researchers intend to design a neural network with better robustness and generalizability for real-time detection and gesture recognition to complement realistic future applications.

Although the designed RF-Eletter could achieve individual letter recognition in different domains, word recognition proved challenging following the highly complex and fine-grained gesture recognition of words (an intriguing complexity to be examined in future research). As such, data segmentation and subsequent techniques [62] could be considered to split the words into different parts and splice the content for word recognition. Different signal waveforms could also be generated for different words through relevant word recognition analysis.

Lastly, RF-Eletter required more gesture data to accurately identify gesture performance for different users. The researchers aim to implement the DCGAN adversarial neural network in future research to retain specific gesture-related information and omit domain-specific knowledge using the adversarial learning process for improved model generalization capabilities. In this vein, the DCGAN network facilitates data expansion and user-environmental domain distinctions for accurate user gesture recognition.

VII. CONCLUSION

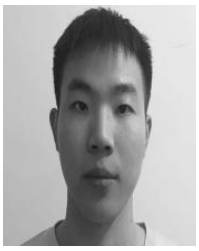
This study proposed RF-Eletter, a device-free, RFID-based, and domain-independent system that facilitated complex and fine-grained gesture recognition. The RFID tag array was implemented to efficiently capture gesture space-time transformations. Subtraction operation and the data normalization, smoothing, and expansion of some pre-processing columns were performed to ensure that the original variance was not particularly high. The RFID gesture signal could be utilized for CNN network feature extraction towards accurate and robust gesture recognition. Multiple experiments demonstrated RF-Eletter to reflect an average accuracy of 100% in classrooms and 95.6% and 96.6% in other counterparts (different users in different scenarios). As the rate proved higher than the current RFID-based gesture recognition scheme, RF-Eletter designs could significantly facilitate gesture-based HCIs.

REFERENCES

[1] 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 networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4207–4215.
[2] M. Wang, B. Ni, and X. Yang, "Recurrent modeling of interaction context for collective activity recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7408–7416.
[3] (2017). *Leap Motion*. [Online]. Available: https://www.vicon.com
[4] (2017). *X-Box Kinect*. [Online]. Available: https://www.xbox.com
[5] G. Laput and C. Harrison, "Sensing fine-grained hand activity with smartwatches," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 1–13.
[6] S. Shen, H. Wang, and R. Roy Choudhury, "I am a smartwatch and i can track my user's arm," in *Proc. 14th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2016, pp. 85–96.
[7] J. Hou, X.-Y. Li, P. Zhu, Z. Wang, Y. Wang, J. Qian, and P. Yang, "SignSpeaker: A real-time, high-precision SmartWatch-based sign language translator," in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, Aug. 2019, pp. 1–15.
[8] H. Wen, J. Ramos Rojas, and A. K. Dey, "Serendipity: Finger gesture recognition using an off-the-shelf smartwatch," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2016, pp. 3837–3851.
[9] H. Gao, W. Huang, and Y. Duan, "The Cloud-edge-based dynamic reconfiguration to service workflow for mobile ecommerce environments: A QoS prediction perspective," *ACM Trans. Internet Technol.*, vol. 21, no. 1, pp. 1–23, Feb. 2021.

[10] 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.
[11] J. Yang, H. Zou, Y. Zhou, and L. Xie, "Learning gestures from WiFi: A Siamese recurrent convolutional architecture," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10763–10772, Dec. 2019.
[12] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2014, pp. 617–628.
[13] 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.
[14] 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.
[15] K. Li, Z. Zhou, and C.-H. Lee, "Sign transition modeling and a scalable solution to continuous sign language recognition for real-world applications," *ACM Trans. Accessible Comput.*, vol. 8, no. 2, pp. 1–23, Jan. 2016.
[16] A. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan, and E. Kalogerakis, "RisQ: Recognizing smoking gestures with inertial sensors on a wristband," in *Proc. 12nd Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2014, pp. 149–161.
[17] H. Ding, L. Guo, C. Zhao, F. Wang, G. Wang, Z. Jiang, W. Xi, and J. Zhao, "RFnet: Automatic gesture recognition and human identification using time series RFID signals," *Mobile Netw. Appl.*, vol. 25, no. 6, pp. 2240–2253, Dec. 2020.
[18] G. Laput, R. Xiao, and C. Harrison, "ViBand: high-fidelity bio-acoustic sensing using commodity smartwatch accelerometers," in *Proc. 29th Annu. Symp. User Interface Softw. Technol.*, Oct. 2016, pp. 321–333.
[19] H. Ding, J. Han, L. Shangguan, W. Xi, Z. Jiang, Z. Yang, Z. Zhou, P. Yang, and J. Zhao, "A platform for free-weight exercise monitoring with passive tags," *IEEE Trans. Mobile Comput.*, vol. 16, no. 12, pp. 3279–3293, Dec. 2017.
[20] H. Ding, C. Qian, J. Han, J. Xiao, X. Zhang, G. Wang, W. Xi, and J. Zhao, "Close-proximity detection for hand approaching using backscatter communication," *IEEE Trans. Mobile Comput.*, vol. 18, no. 10, pp. 2285–2297, Oct. 2019.
[21] H. Gao, C. Liu, Y. Li, and X. Yang, "V2 VR: Reliable hybrid-network-oriented V2 V data transmission and routing considering RSUs and connectivity probability," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3533–3546, Jun. 2021.
[22] J. Song, F. Pece, S. R. Fanello, C. Keskin, and O. Hilliges, "In-air gestures around unmodified mobile devices," in *Proc. ACM UIST*, 2014, pp. 319–329.
[23] J. Taylor, L. Bordeaux, T. Cashman, B. Corish, C. Keskin, E. Soto, D. Sweeney, J. Valentin, B. Luff, A. Topalian, E. Wood, S. Khamis, P. Kohli, T. Sharp, S. Izadi, R. Banks, A. Fitzgibbon, and J. Shotton, "Efficient and precise interactive hand tracking through joint, continuous optimization of pose and correspondences," *ACM Trans. Graph.*, vol. 35, no. 143, pp. 1–12, Jul. 2016.
[24] 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.
[25] C. Dian, D. Wang, Q. Zhang, R. Zhao, and Y. Yu, "Towards domain-independent complex and fine-grained gesture recognition with RFID," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, pp. 1–22, Nov. 2020.
[26] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE MultiMedia*, vol. 19, no. 2, pp. 4–10, Feb. 2012.
[27] 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.
[28] L. E. Potter, J. Araullo, and L. Carter, "The leap motion controller: A view on sign language," in *Proc. 25th Austral. Comput.-Hum. Interact. Conf. Augmentation, Appl., Innov., Collaboration*, 2013, pp. 175–178.
[29] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *Proc. USENIX NSDI*, 2014, pp. 317–329.
[30] L. Yang, Q. Lin, X. Li, T. Liu, and Y. Liu, "See through walls with COTS RFID system!" in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 487–499.
[31] H. Gao, Y. Duan, L. Shao, and X. Sun, "Transformation-based processing of typed resources for multimedia sources in the IoT environment," *Wireless Netw.*, vol. 27, no. 5, pp. 3377–3393, Jul. 2021.
[32] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.
[33] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFi: Decimeter level localization using WiFi," in *Proc. ACM Conf. Special Interest Group Data Commun.*, Aug. 2015, pp. 269–282.
[34] T. Westeyn, H. Brashear, A. Atrash, and T. Starner, "Georgia tech gesture toolkit: Supporting experiments in gesture recognition," in *Proc. 5th Int. Conf. Multimodal Interface*, 2003, pp. 85–92.
[35] M. Bouet and G. Pujolle, "3-D localization schemes of RFID tags with static and mobile readers," in *Proc. Netw.*, 2008, pp. 112–123.
[36] X. Yang, X. Chen, X. Cao, S. Wei, and X. Zhang, "Chinese sign language recognition based on an optimized tree-structure framework," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 4, pp. 994–1004, Jul. 2017.
[37] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 3, pp. 311–324, May 2007.
[38] H. Ding, L. Shangguan, Z. Yang, J. Han, Z. Zhou, P. Yang, W. Xi, and J. Zhao, "FEMO: A platform for free-weight exercise monitoring with RFIDs," in *Proc. 13rd ACM Conf. Embedded Netw. Sensor Syst.*, Nov. 2015, pp. 141–154.
[39] L. Shangguan, Z. Zhou, and K. Jamieson, "Enabling gesture-based interactions with objects," in *Proc. 15th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2017, pp. 239–251.
[40] 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.
[41] 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.
[42] Y. Iravantchi, M. Goel, and C. Harrison, "BeamBand: Hand gesture sensing with ultrasonic beamforming," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 1–10.
[43] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–19, 2016.
[44] A. Patra, P. Geuer, A. Munari, and P. Mähönen, "Mm-wave radar based gesture recognition: Development and evaluation of a low-power, low-complexity system," in *Proc. 2nd ACM Workshop Millim. Wave Netw. Sens. Syst.*, Oct. 2018, pp. 51–56.
[45] E. Hayashi, J. Lien, N. Gillian, L. Giusti, D. Weber, J. Yamanaka, L. Bedal, and I. Poupyrev, "RadarNet: Efficient gesture recognition technique utilizing a miniature radar sensor," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2021, pp. 1–14.
[46] B. Xie, J. Xiong, X. Chen, E. Chai, L. Li, Z. Tang, and D. Fang, "Tagtag: Material sensing with commodity RFID," in *Proc. 17th Conf. Embedded Networked Sensor Syst.*, Nov. 2019, pp. 338–350.
[47] 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.
[48] Z. Huang, M. Dong, Q. Mao, and Y. Zhan, "Speech emotion recognition using CNN," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 801–804.
[49] X. Xing, M. Dong, C. Bi, and L. Yang, "Self-quotient image based CNN: A basic image processing assisting convolutional neural network," in *Proc. 3rd Int. Conf. Digit. Signal Process.*, 2019, pp. 17–21.
[50] R. I. Ramos-Garcia, E. R. Muth, J. N. Gowdy, and A. W. Hoover, "Improving the recognition of eating gestures using intergesture sequential dependencies," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 3, pp. 825–831, May 2015.
[51] K. Chawla, G. Robins, and L. Zhang, "Object localization using RFID," in *Proc. IEEE 5th Int. Symp. Wireless Pervasive Comput.*, Dec. 2010, pp. 301–306.
[52] 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.
[53] 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.

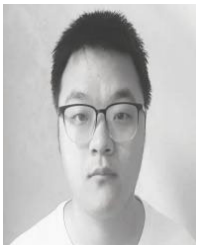
[54] X. Fan, W. Gong, and J. Liu, "TagFree activity identification with RFIDs," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–23, Mar. 2018.
[55] J. Wang, D. Vasisht, and D. Katabi, "RF-IDraw: Virtual touch screen in the air using RF signals," in *Proc. SIGCOMM*, vol. 44, no. 4, 2014, pp. 235–246.
[56] Q. Zhang, D. Li, R. Zhao, D. Wang, Y. Deng, and B. Chen, "RFree-ID: An unobtrusive human identification system irrespective of walking cofactors using COTS RFID," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mar. 2018, pp. 1–10.
[57] A. Huang, D. Wang, R. Zhao, and Q. Zhang, "Au-id: Automatic user identification and authentication through the motions captured from sequential human activities using RFID," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 2, pp. 1–26, Jun. 2019.
[58] 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.
[59] C. Liu, J. Xiong, L. Cai, L. Feng, X. Chen, and D. Fang, "Beyond respiration: Contactless sleep sound-activity recognition using RF signals," *ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–22, Sep. 2019.
[60] L. Chen, J. Xiong, X. Chen, S. I. Lee, D. Zhang, T. Yan, and D. Fang, "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.
[61] A. A. Neacsu, G. Cioroiu, A. Radoi, and C. Burileanu, "Automatic EMG-based hand gesture recognition system using time-domain descriptors and fully-connected neural networks," in *Proc. 42nd Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2019, pp. 232–235.
[62] 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, China, in 2019. He is currently pursuing the master's degree in software engineering with Xinjiang University, China. His research interests include indoor localization technology, pattern recognition, 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, China, in 2020. He is currently pursuing the master's degree in software engineering with Xinjiang University, China. His research interests include indoor localization technology, pattern recognition, and signal processing.

ZIJIAN LI was born in Xuzhou, Jiangsu, China, in 1998. He received the bachelor's degree in computer science and technology from Changzhou University, in 2020. He is currently pursuing the master's degree in software engineering with Xinjiang University, China. His research interests include image processing, big data technology, 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, in 2019. He is currently pursuing the master's degree in software engineering with Xinjiang University.

YAJUN ZHANG was born in Nanyang, Henan, China, in 1983. He received the B.S. and master's degrees from Xinjiang University, China, in 2007 and 2010, respectively. He is currently working as an Associate Professor with Xinjiang University. His research interests include indoor localization technology, pattern recognition, and signal processing.