A Trajectory-based Gesture Recognition in Smart Homes based on the Ultra-Wideband Communication System

Anna Li, Eliane Bodanese, *Member, IEEE*, Stefan Poslad, Tianwei Hou, *Member, IEEE*, Kaishun Wu, *Member, IEEE*, and Fei Luo

***Abstract*—In this paper, a cost-effective ultra-wideband (UWB) communication system for gesture recognition in a smart home environment is proposed, which uses gesture trajectories and a deep learning model. Most previous studies of gesture recognition using UWB technology used electromagnetic signals directly, which may bring problems like radar clutter, signal coupling, multi-path, fading, and interference. However, instead of using UWB's high-frequency pulse signals, the proposed method only uses gesture trajectories by data positioning. To this end, firstly, a dataset of four gesture activities was created. Then, this dataset was trained using a convolutional neural network (CNN) integrated with a squeeze-and-excitation (SE) block, namely the SE-Conv1D model. Finally, the system was prototyped to interact with appliances in practical smart homes. The experimental data was used to demonstrate the superiority of the SE-Conv1D model in comparison with four baselines: support vector machines, K-nearest neighbor, random forest, and binarized neural networks. Experimental results show that all collected gesture activities are correctly recognized with an overall accuracy of over 95%, among which the proposed SE-Conv1D model achieves the best accuracy of 99.48%. The proposed system is a complete end-to-end sensing system specifically designed for tracking and recognizing human gestures, which is robust against interference and changes in distance or direction. In addition, the proposed system can tackle the device selection problems for smart homes, which means it is reliable for real-world applications.**

***Index Terms*—Gesture recognition, Smart home, Squeeze-and-excitation, Trajectory, Ultra-Wideband**

I. Introduction

Gesture recognition is one of the most critical sub-topics in human activity recognition (HAR), and plays a key role in the development of multiple applications, including smart homes [1], health care [2], and virtual reality [3]. Gesture recognition is able to control devices remotely, which is convenient and efficient for users [4]. Recent developments in gesture recognition applied to smart homes, e.g., SeleCon [5], have attracted great attention in using gestures to control smart devices.

This work is supported by the Fundamental Research Funds for the Central Universities 2021RC260 and the Queen Mary–China Scholarship Council Co-Funded Scholarships. (*Corresponding author: Fei Luo.*)

Anna Li, Eliane Bodanese and Stefan Poslad are with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: anna.li@qmul.ac.uk, eliane.bodanese@qmul.ac.uk, stefan.poslad@qmul.ac.uk).

Fei Luo and Kaishun Wu are with the School of Computer Science, Shenzhen University, Shenzhen 518000, China. (e-mail: luofei2018@outlook.com, wu@szu.edu.cn).

Tianwei Hou is with the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China. (e-mail: twhou@bjtu.edu.cn).

A. Limitation of State of the Art

Previous attempts for gesture recognition used different sensing modalities like cameras, audio-based approaches, Wi-Fi, radio-frequency identification (RFID), and Bluetooth techniques. However, these approaches suffer from inherent drawbacks, including privacy leakage, inconvenience, as well as limited sensing range and interference. For example, vision-based approaches have to deal with well-known environmental challenges, where the line-of-sight (LoS) is strictly required between the camera and users [6]. In addition, the feasibility of the vision-based recognition is impacted by the variability in brightness, contrast, and exposure [7]. It is worth noting that both vision-based and speech-based recognition approaches may violate users' privacy because the recorded video and audio data are released to remote cloud servers. To overcome privacy issues, Wi-Fi, RFID, and Bluetooth techniques are more applicable. However, Wi-Fi is limited by the low spatial resolution, signal strength, multi-path reflections, as well as electromagnetic interference [8]. Bluetooth and RFID are greatly restricted to the short-range sensing capability [9].

Recently, radar-based gesture recognition [10] has been considered as alternatives for overcoming the problems mentioned above. Radar-based recognition has no privacy and LoS issues, which stands as a potential solution for gesture recognition [11]. However, radar-based recognition also suffers from interference from other devices and multi-path effects, which dramatically decrease the signal-to-interference-noise-ratio of the wireless signals [12]. Most of the solutions utilize digital signal processing techniques for mitigating co-channel interference. However, it is still hard to distinguish the interference for gesture recognition in practice.

Another challenge is that previous studies of gesture recognition have not thoroughly tackled the device selection problem, especially for smart home applications. Only a few existing gesture recognition solutions can select a device and control it without increasing the number of tags needed in the different objects [5]. In most studies, different gestures are assigned to various devices in a smart home. However, assigning semantic tags for each device, such as 'light 1' or 'light 2', results in an increased burden to users. With the increasing number of devices in smart homes, this process

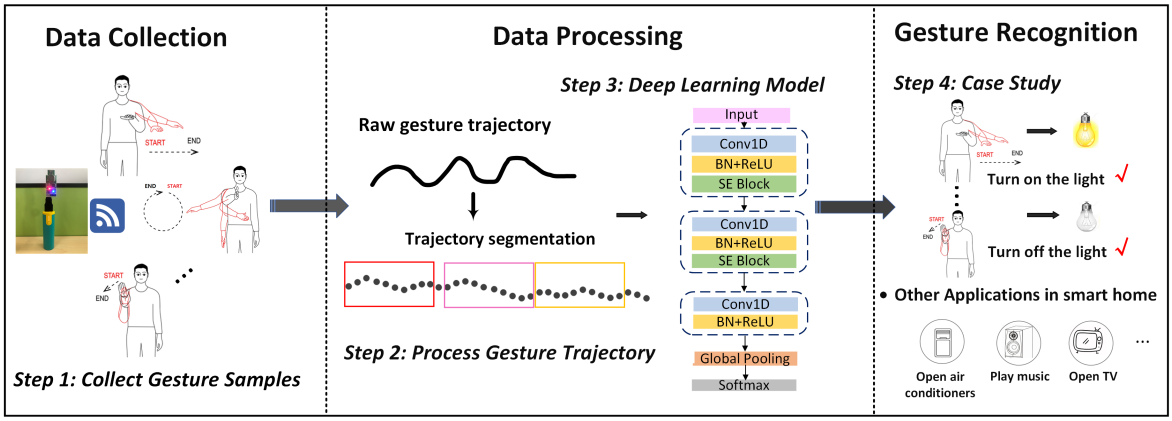

Fig. 1: Architecture of the proposed UWB communication system.

becomes cumbersome. It is natural to ask if it is possible to simultaneously control and select a specific device without defining a massive number of gestures.

B. Motivations

To solve the mentioned problems, a practical solution that can be applied to smart homes is strongly needed. Therefore, this paper presents a cost-effective communication system based on Ultra-Wideband (UWB) technology with a wide bandwidth ($\geq$500MHz). Previous research mainly focused on using electromagnetic signals directly to perform gesture recognition, e.g., UWB radars [13]. However, these existing studies using UWB radars do not completely address problems like signal coupling, multi-path, fading, and interference. In addition, if gesture verification is performed at a distance or direction that is not used for training, the accuracy may be reduced.

Different from the previous studies, this paper focuses on the high localization accuracy of UWB, which is regarded as one of the most accurate and promising technologies. A novel solution that uses gesture trajectories instead of using electromagnetic signals directly to recognize human gestures is proposed based on UWB technology. Additionally, a deep learning model is developed to recognize human gestures. In this way, the proposed UWB communication system can solve the problem of interference and device selection for smart homes.

C. Contributions

A trajectory-based solution for gesture recognition using the UWB communication system is proposed, which aims to overcome the constraints of the existing works. A deep learning model is presented for recognizing dynamic human gestures. It is worth noting that the proposed system is not just a sensor chip or a new learning algorithm. As shown in Fig. 1, it is a complete end-to-end sensing system specifically designed for tracking and recognizing human gestures. The main contributions of this paper are summarized as follows:

- A novel solution of recognizing human activities only using trajectories of human activities based on a cost-effective UWB communication system is proposed. Then, a dataset of four pre-defined gestures, e.g., turning on/off the LED light, was collected in a laboratory that simulates a room of a smart home. Finally, the trajectories of pre-defined gestures from different directions and distances were produced.
- A novel framework for gesture recognition is proposed, namely the SE-Conv1D model, which achieves excellent gesture recognition performance of 99.48% overall accuracy. All of these gestures are also correctly classified with an accuracy of over 95% by the four baseline models. It proves that the proposed solution is robust against interference, and it works robustly against changes in distance or direction, which means it is more reliable for real-world applications.
- The proposed system was prototyped to interact with appliances in smart homes. The proposed system provides a practical method of device selection for smart homes only using gestures. It proves that the proposed system is a complete end-to-end sensing system specifically designed for tracking and recognizing human gestures.
- The archived UWB gesture datasets and code[1] have been already published, which may provide the basis for the comparison of techniques and improvements.

D. Organization

The rest of this paper is organized as follows. Section II presents the related work. Section III presents the system overview. Section IV presents a detailed review of the UWB technology, double-sided two-way ranging, and the proposed learning model. Section V shows the evaluation setup. The experimental results and analysis are illustrated in section VI. Section VII presents the limitations and future work. Finally, Section VIII concludes this paper.

II. Related Work

Human gesture recognition is a widely concerned research field. Given the enormous volume of research conducted on this topic and space limitation, the three most relevant aspects

[1]Data and code available at https://github.com/Annaliisme/UWB-based-gesture-recognition.

to this paper are introduced: UWB-based gesture recognition, trajectory-based gesture recognition, and squeeze and excitation (SE) blocks.

A. Gesture Recognition based on Ultra-Wideband Technology

Recently, there has been renewed interest in UWB technology. UWB technology is regarded as one of the most accurate technologies for providing high localization accuracy, high immunity against multi-paths, and low output power [14]. Most previous work [15] focused on indoor localization based on UWB technology, but few focused on gesture recognition.

An early study for gesture recognition using UWB technology was published in 2016, where Park *et al.* [16] recognized human gestures by using an impulse radio UWB radar. With the help of the principal component analysis (PCA) method, the features were successfully extracted from the received signals. Then, support vector machines (SVMs) were used for training and classifying six gestures using the extracted features. Lately, Khan *et al.* [17] improved this system with a particular focus on user-friendliness and flexibility. They proposed a new hand-based gesture recognition algorithm that can be used to control different electronic devices inside a car. Based on this work, Ahmed *et al.* [18] developed a convolutional neural network (CNN) model for counting fingers based on gestures to control electronic devices in cars, which achieved approximately 97% overall accuracy. Recently, several research efforts have started to focus on gesture-driven applications for smart homes. Alanwar *et al.* [5] proposed a pointing approach to interact with different devices in smart homes based on a UWB equipped smart-watch. Their results demonstrated that it achieves 84.5% accuracy for device selection and 97% overall accuracy for gesture recognition.

The above methods based on UWB technology showed exciting gesture recognition accuracy in the corresponding dataset. However, radar-based gesture recognition suffers from the interference of nearby moving objects and multi-path effects (shown in Fig. 2). In addition, most of the studies mentioned above are not sensitive to users' orientation. For example, if users perform gestures differently from any of the directions where the training samples are collected, the systems' accuracy would suffer. It is essential to note that few works have investigated the problem of equipment selection without increasing the tags for each device.

B. Trajectory-based Gesture Recognition

Trajectory-based methods provide a means of solving the problem of interference [19]. Most of the previous research on trajectory-based gesture recognition used cameras. The Time of Flight (ToF) sensors (e.g., Kinect 2.0) were widely used in the fields of computer vision. ToF sensors have been reported to be quite efficient and robust to the hand's orientation, size, and distance measurement task, which can be utilized to recognize complex hand trajectory gestures [20]. ChaLearn [21] is a well-known trajectory gesture dataset that consists of 50,000 hand and arm gestures video recorded with a Microsoft Kinect^{TM} camera that can provide both RGB and depth images. However, trajectory-based approaches using cameras have failed to address the problem of small visual scope because the trajectories extracted from a video heavily depend on the azimuth and inclination of the camera. In addition, the privacy issue is still a much-debated question.

	UWB radar	UWB communication
Method	Use electromagnetic signals directly	Use data positioning
Pros	• No privacy issues • Contactless • Hardly affected by lighting and noise	• High precise positioning • Interference-free • No multi-path effects • No privacy issues • Hardly affected by lighting and noise
Cons	• Affected by radar clutter, signal coupling, multi-path, fading, and interference	• Wearable

Fig. 2: The difference between UWB radar and the proposed UWB communication system.

Lately, real-time wearable hand gesture recognition can be used for gesture detection, e.g., Chen *et al.* [22] proposed a novel wearable wrist camera-worn sensor, the WristCam, for recognizing hand trajectory gestures. By dividing the video into gesture segments, they finally achieved an accuracy of 97.6%. Vu *et al.* [23] explored the possibility of using wrist-worn devices based on the trajectory tracking solution to track the hand movement and gestures for gesture-driven applications accurately. They built a Hidden Markov Model (HMM)-based framework, which achieved 95% stroke recognition accuracy. It should be noted that few studies have focused on trajectory-based gesture recognition using the UWB communication system.

C. Squeeze and Excitation (SE) Blocks

In the last decade, machine learning techniques have been widely used for gesture recognition, which have laid a good foundation for gesture recognition and met the requirements of typical applications for recognition accuracy. Due to the complexity and diversity of gesture movement, classification models have gradually developed from manual feature extraction to model training of the deep network.

Since it was reported in 2018, SE blocks have attracted a lot of interest and some achievements have been made. Rundo *et al.* [24] incorporated SE blocks into U-Net for prostate zonal segmentation of multi-institutional Magnetic Resonance Imaging (MRI) datasets. Their research revealed that SE blocks could provide superior intra-dataset generalization in multi-institutional scenarios. In [25], an end-to-end intelligent recognition of epileptic electro-encephalogram (EEG) seizure detection framework was proposed by using a channel-embedding spectral-temporal SE network. Experimental results demonstrated the effectiveness of their proposed framework in recognizing epileptic EEGs, indicating its powerful capability in automatic seizure detection. In [26], a faster region-based CNN by using SE mechanisms was proposed for ship detection. The authors claimed that their proposed method outperforms the state-of-the-art methods.

TABLE I: The parameter setting of the proposed UWB communication system.

Parameters	Values
Centre Frequency	3 GHz
Bandwidth	500 MHz
Bit Rate	11520 bps (UART)
Range	0-9 meter

The SE blocks have not been used in gesture recognition systems even though they have achieved some signs of success in many fields. This paper incorporated the SE blocks in the one-dimensional CNN to improve gesture recognition accuracy for smart home applications.

III. System Overview

Throughout the paper, this paper assumes that human gestures may be accurately recognized by the trajectories of different hand movements obtained by the UWB communication system. With the help of deep learning models, trajectories of human gestures can be used directly to classify meaningful gestures. A deep learning-based UWB communication system is designed to verify the hypothesis in the rest of this paper. The framework used in this paper has three stages (shown in Fig. 1): data collection, data processing, and gesture recognition.

Data collection: A UWB communication system is employed, which enables the integration of sensing and wireless data transfer into a single step. The user wears the tag, and the anchors can be placed in the room. The human motions are spatially translated and oriented during physical activities, which change the relative location. Collecting the received ToF signals transmitted from the UWB communication system can model the relative location of human motions to represent different human gestures, which can be expressed as a sequence of coordinates containing spatial and temporal information.

Data processing: The collected point sets are grouped into different clusters that represent various human activities in a scene. This paper uses points-based segmentation to process trajectories obtained from the UWB communication system, splitting trajectories into several segments with the same number of points.

Gesture recognition: A deep learning model is applied to classify geometric features extracted upon these point clusters for detecting human gestures. It consequently enhances the accuracy of the proposed UWB communication system in recognizing human gestures.

IV. Methodology

This section gives an overall review of the UWB technology-based on DWM1000[2], the theory of double-sided two-way ranging (DS-TWR), SE blocks, and the proposed SE-Conv1D model.

[2]The DWM1000 module is based on DecaWave's DW1000 Ultra-Wideband (UWB) transceiver IC. For further information on this, please refer to www.decawave.com.

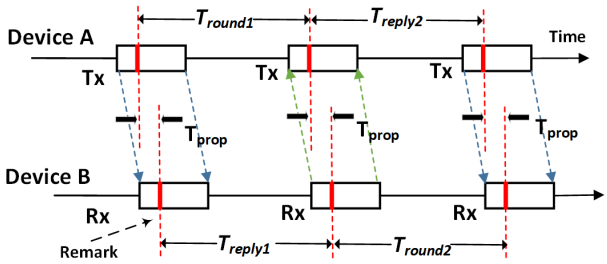

Fig. 3: The theorem of double-sided two-way ranging system.

A. The Ultra-Wideband Technology based on DWM1000

UWB technology is one of the most potent choices for critical positioning applications that require highly accurate results [27]. In this research, one type of sensor (DWM1000 module) was utilized, which is compliant with the IEEE 802.15.4-2011 UWB standard. The module size is 54 mm × 20 mm × 2.9 mm, and it integrates antennas, all RF circuits, power management, and clock circuitry in one module. As shown in TABLE I, the centre frequency is 3 GHz, bandwidth is 500 MHz, and the bit rate is 11520 bps by universal asynchronous receiver-transmitter (UART).

B. Double-sided Two-way Ranging (DS-TWR)

UWB ranging is suitable for real-time locating systems (RTLS) [28]. In addition, the wide frequency bandwidth allows for high-resolution channel impulse response estimation, along with accurate ToF measurements in a dense multi-path environment [29]. The system used in this research is based on the double-sided two-way ranging (DS-TWR) [30], in which two round trip time measurements are used and combined to give a ToF result. The core of a DS-TWR exchange is shown in Fig. 3.

In this paper, three DWM1000 modules were configured as anchors (anchor A, anchor B, anchor C), while another was configured as a tag. The trilateration solver gives two solutions equidistant from each side of the plane of the anchors, which are assumed to be all horizontal. Each device precisely time-stamps the transmission and reception times of the messages. The remark is the part of the frame that is assumed to be time-stamped at the device antennas. The final message communicates the tag's T_{round} and T_{reply} times to the anchors. Assuming that any two anchors can communicate between them. The resultant ToF estimate, $\hat{T}_{prop}$, which is the propagation time of the message between tag and anchors, is calculated by:

$$\hat{T}_{prop} = \frac{(T_{round1} \cdot T_{round2} - T_{\text{reply}1} \cdot T_{\text{reply}2})}{(T_{round1} + T_{round2} + T_{\text{reply}1} + T_{\text{reply}2})}, \quad (1)$$

where T_{round1} represents the time from sending the polling signal to Anchor A to receiving the response signal of the tag; T_{round2} represents the time from sending the response signal to the Rx to receiving the ranging information sent by anchor A; $T_{\text{reply}1}$ represents the time from the tag receiving the polling signal to the time when the response signal is sent; $T_{\text{reply}2}$ represents the time from the Anchor A receiving the

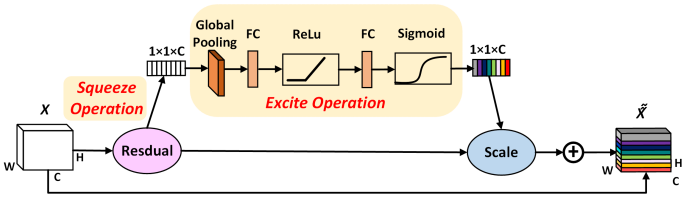

Fig. 4: The scheme of SE module.

response signal to the time sending the ranging information. Finally, the tag's position can be estimated by TOF from anchors. This paper assumes the speed of the radio waves through the air is equal to the speed of light c, then the distance between the tag and anchors can be expressed by:

$$Distance = c \cdot \hat{T}_{prop}. \tag{2}$$

C. The Proposed Model Architecture

Recently, *squeeze-and-excitation* (SE) blocks [31] have been proposed to become the integral parts of models, which are used to rescale the input feature map to highlight valuable channels. These blocks are lightweight to decrease the model complexity and computation time and ease the training of the network by improving gradient flow. The general scheme of the SE module is shown in Fig. 4. For an input feature map $\mathbf{U}^{W\times H\times C}$, a *squeeze* operation is utilized to aggregate the feature map across dimensions $W \times H$ to produce a channel descriptor, which embeds the global distribution of channel-wise feature responses [32]. This is achieved by using global average pooling to generate channel-wise statistics as:

$$\mathbf{z}_c = \frac{1}{\mathrm{W}\times H}\sum_{i=1}^{W}\sum_{j=1}^{H} u_c(i,j). \tag{3}$$

where H and W are the height and width of the features from the previous layer, respectively. The $u(i,j)$ is the data element. The width can be regarded as W = 1 for the one-dimensional convolution.

For temporal sequence data, the channel-wise statistics is generated by shrinking $\mathbf{U}$ through the temporal dimension T, where the c-th element of $\mathbf{z}$ is calculated by:

$$\mathbf{z}_c = \frac{1}{T}\sum_{t=1}^{T} u_c(t). \tag{4}$$

The aggregated information obtained from the *squeeze* operation is followed by the *excitation* operation, which aims to capture the channel-wise dependencies. To meet these criteria, a simple gating mechanism is employed with a sigmoid activation:

$$\mathbf{s}_c = u(i,j)\sigma(g(\mathbf{z}_c,\mathbf{W}) = u(i,j)\sigma(\mathbf{W_1}\delta(\mathbf{W_2 z_c})), \tag{5}$$

where σ refers to the Sigmoid activation function, δ refers to the ReLU activation function. $\mathbf{W}_1$ and $\mathbf{W}_2$ are weights to limit model complexity and aid generalisation.

The SE blocks are integrated into the one-dimensional CNN model to enhance recognition accuracy, for convenience,

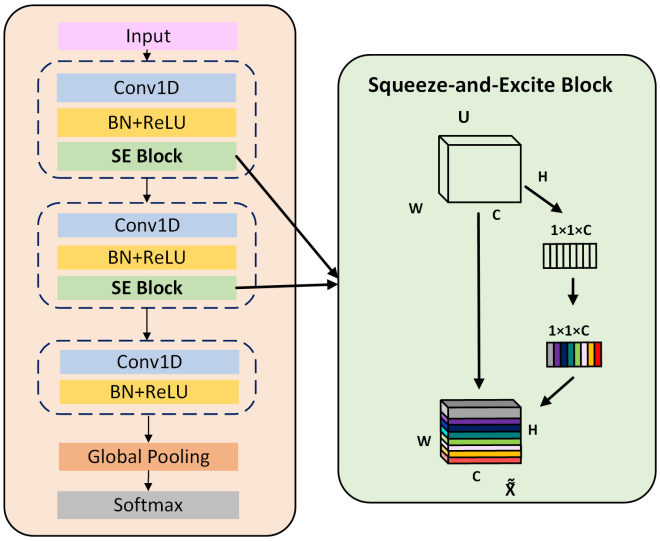

Fig. 5: The proposed SE-Conv1D model.

namely the SE-Conv1D model, as illustrated in Fig. 5. One-dimensional CNN takes one-dimensional data as input. It is a variant of traditional two-dimensional CNN, which includes an input layer, an output layer, and multiple hidden layers. Compared with two-dimensional CNN, one-dimensional CNN uses one-dimensional arrays instead of two-dimensional matrixes to represent the kernels and feature maps [33]. In the SE-Conv1D model, the segmented trajectories are taken as inputs. The fully convolutional block contains three temporal convolutional blocks, which are used as feature extractors. The SE-Conv1D model is comprised of 2 blocks of (128, 256, 128) filters for all models, with kernel sizes of 16, 3, and 5, respectively. Each convolutional layer is succeeded by the batch normalization layer and the ReLU activation function. A global average pooling layer follows the final temporal convolutional block.

V. Evaluation Setup

This section presents a comprehensive experiment on the collected gesture dataset. The research questions of this experiment are raised through three aspects: robustness, effectiveness, and usefulness. The data acquisition and processing procedure, baselines, and evaluation metrics are introduced. Finally, the hypotheses are set up to answer research questions.

A. Research Questions

Given the fact that the deep learning model fundamentally determines the effectiveness of the proposed system, different parameter settings are analyzed to present readers an extensive understanding of this paper. $RQ1$ is designed as follows:

> **[RQ1. Robustness]** How do the experimental parameters influence the performance of the proposed model?

Secondly, $RQ2$ is designed to investigate whether the proposed system is effective for recognizing different human gestures. $RQ2$ is formulated as follows:

TABLE II: The description of pre-defined gestures.

ID Number	Gesture Activity	Descripition	Case study
0	Swipe right	The gesture of swiping the right hand to the right	Turn on the LED light
1	Up and down	The gesture of lifting the right hand from bottom to top and then put it down	Change the color of an LED light
2	Circle clockwise	The gesture of rotating the right hand for a circle clockwise	Turn on all LED lights
3	Push	The gesture of pushing the right hand outwards perpendicular to the ground	Turn off all LED lights
4	Others	Daily behaviors, e.g., walking and sitting	/

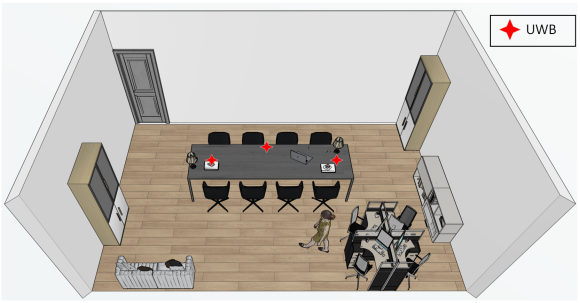

Fig. 6: The experimental scenario is 6.5m × 6m, which mainly contains one sofa, two tables, and three bookcases. The UWB devices were configured on the table.

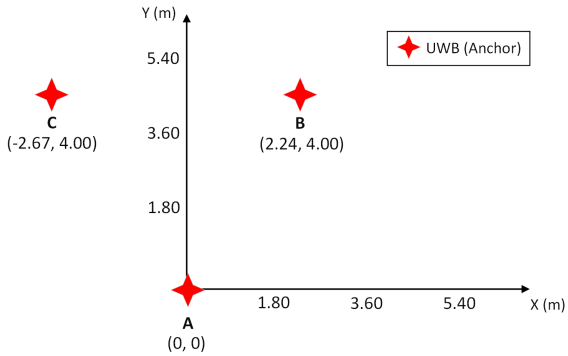

Fig. 7: The location of device configuration.

[RQ2. Effectiveness] To what extent can the proposed system accurately recognize different human gestures?

Finally, it is essential to note that even though we investigate the effectiveness of the proposed system in $RQ2$, the usefulness of that in practice needs to be further validated. The proposed system was prototyped to interact with appliances in smart homes and compared with the state-of-the-art approaches. $RQ3$ is designed as follows:

[RQ3. Usefulness] To what extent can the proposed system accurately recognize different human gestures in uncontrolled environments?

B. Data Measurement and Processing

In this section, the experimental setup is presented, including device configuration, the experimental environments, gesture sets, participants, and data processing.

1) Data Measurement: As shown in Fig. 6, the data collection was conducted in an office, where we simulated the smart home scenarios. In this experiment, three DWM1000 modules were configured as anchors while another was configured as a tag. The tag was connected to the human body while the data was being collected in real-time. As shown in Fig. 7, three anchors were deployed, namely anchors A, B, and C. Anchors B and C were located 89cm above the ground, and anchor A was located 33cm above the ground. This position was chosen because most of the electric outlets are at this height. The idea is to have a plug-and-play solution. Anchor B and anchor C were kept fixed at a separation distance of 4.91m from each other.

A set of gestures were predefined to represent different user commands in a smart home, including four continuous gestures, 'swipe right', 'up and down', 'circle clockwise' and 'push', which are shown in Fig. 8 (a). To explore how the proposed system would be incorrectly triggered by daily behaviors, e.g., 'walking' and 'sitting', volunteers were asked to perform the daily behaviors for 10 minutes. These activities were defined as 'Others'. The standard for each class is shown in TABLE II. The data were collected from three volunteers: two males and one female. They were unpaid volunteers recruited from different departments of Shenzhen university. It is worth noting that all activities were performed in naturally different orientations, as would be the case in a real-world scenario. When the instruction 'start' was given, a volunteer, facing different directions at different positions, performed a pre-defined gesture around 100 times.

2) Data Processing: Finally, trajectories from different directions and distances were produced by the proposed UWB communication system (shown in Fig. 8 (b)). A trajectory is the result of the displacement of a human gesture obtained by the UWB communication system, which can be expressed as a sequence of coordinates containing spatial and temporal information. This paper uses points-based segmentation of sliding window to process trajectories obtained from the UWB communication system, which split trajectories into several segments with the same number of points [19]. A sliding window is a well-established method of feature extraction for data pre-processing. The information could be extracted over a 'sliding window' that contains a fixed number of samples.

As shown in Fig. 9, a sliding window with a length of d was utilized to perform segmentation at the interval of v. The choice of window size is essential for gesture recognition. If the size is too small, the signals of a human gesture cannot be entirely captured by the window. On the contrary, if the window size is too large, signals of two or more

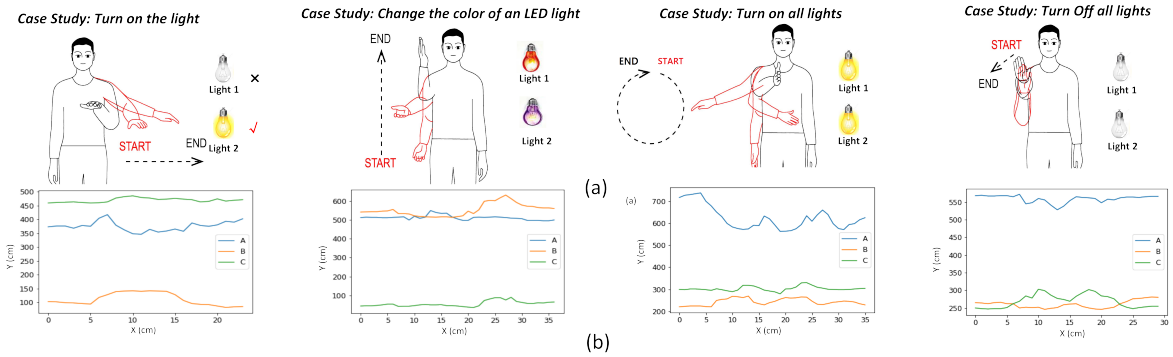

Fig. 8: (a) The design of four potential human gesture activities in homes. From left to right: 'swipe right', 'up and down', 'circle clockwise', and 'push'. (b) Trajectories are randomly chosen from different directions and distances produced by UWB. 'A', 'B', and 'C' refer to Anchor A, B, and C.

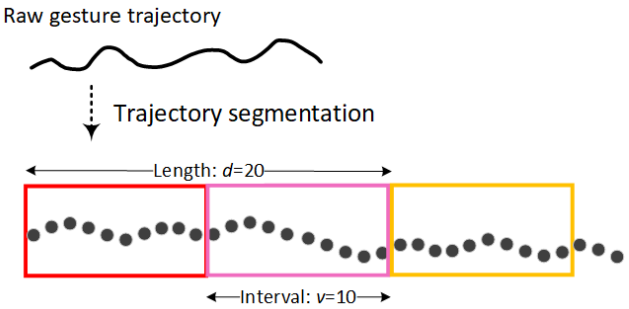

Fig. 9: An example of trajectory segmentation.

human gestures can be included. Considering different users' different shapes and speeds, we can observe that each sliding window contains 8-30 location points extracted from gesture trajectories. A sliding window was finally chosen according to the average completion time of gestures. Although there is no guarantee that a sliding window only contains one motion cycle for different speeds of gestures, it at least contains the whole gesture, which is enough for gesture classification. After testing, gesture trajectories were then segmented using a sliding window with an average length of 20 (location points). The final dataset comprises 19121 samples, including 3145 samples of 'swipe right', 3291 samples of 'up and down', 4737 samples of 'circle clockwise', 3948 samples of 'push forward', 4000 samples of 'others', respectively.

C. Evaluation Metrics

The metrics used to evaluate recognition performance from different perspectives are: 1) Overall Accuracy (OA); 2) Recall; 3) F_1 score and 4) Normalized confusion matrix. The metrics used to evaluate performance of the system are defined as follows:

$$\mathrm{OA} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \tag{6}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{7}$$

$$F_1 = 2 \cdot \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{8}$$

where TP, FP, TN, and FN denote true positive, false positive, true negative, and false negative, respectively. During the training process, the trajectory samples were divided into three segments: 80% for the training, 10% for the validation, and 10% for testing.

D. Baselines

The machine learning techniques have been proven to be effective for gesture recognition, including support vector machines (SVM), K-nearest neighbor (KNN), and random forest (RF). In addition, a recent deep learning method based on a binarized neural network is presented. Details of the comparative approaches are listed in the following:

- SVM: SVM supports to select the hyperplanes that maximize the distance between the nearest training samples and the hyperplanes [34]. In the case of SVM, the radial basis function (RBF) kernel function [35] is used. Here, two hyperparameters (C, γ) in SVM-RBF are specified manually. Given the hyperparameter space C: [0, 20], γ: [0, 1.0E - 5], a different pair of parameters will be selected from the hyperparameter space by using the cross-validation in each training and validation epoch. In this paper, the gesture recognition performance is best when C is set to 10, and γ is set to 1.0E - 4.3.
- KNN: KNN follows an assumption that similar things are closed to each other [36]. Adjusting the k value is critical in assuring that the model is robust to noise, and is able to discriminate the gesture classes. In the case of KNN, different values for k are examined, and the gesture recognition performance is best when k is set to 2.
- RF: RF is a holistic learning method of classification and regression by constructing a collection of decision trees [37]. In the case of RF, an RF built with 110 decision trees is utilized.
- BinaryDilatedDenseNet [38]: BinaryDilatedDenseNet is a recent model that designed for sensor-based human

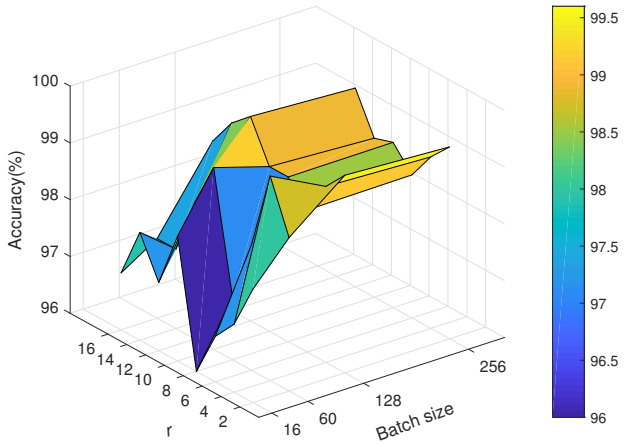

Fig. 10: The accuracy under the different reduction ratios r and batch size of the SE-Conv1D model.

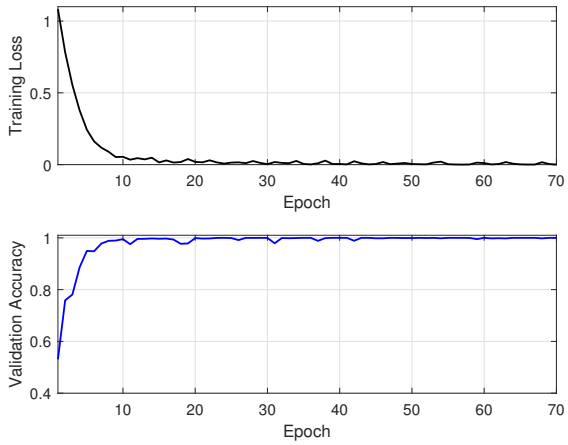

Fig. 11: Validation accuracy and loss of the proposed SE-Conv1D model.

activity recognition. The same structure used in [38] is adopted. In training, categorical cross-entropy loss is used to optimize the network. Adam is chosen as the optimization function, with the initialized learning rate of 0.0001, and the batch size is set to 32.

E. Hypotheses Setup

To answer **RQ1**, the following null hypothesis $H1_0$ and the alternative hypothesis $H1_A$ are set up:

- *$H1_0$: The reduction ratio r cannot significantly influence the performance.*
- *$H1_A$: The reduction ratio r can significantly influence the performance.*

To answer questions in **RQ2**, the null hypothesis $H2_0$ and the alternative hypothesis $H2_A$ are set up:

- *$H2_0$: The overall gesture accuracy of the proposed SE-Conv1D model is not significantly higher than the baselines (SVM, RF, KNN, and BinaryDilatedDenseNet).*
- *$H2_A$: The overall gesture accuracy of proposed SE-Conv1D model is significantly higher than the baselines (SVM, RF, KNN, and BinaryDilatedDenseNet).*

To comprehensively answer **RQ3**, the null hypotheses $H3_0$ and the alternative hypothesis $H3_A$ are set up:

- *$H3_0$: The environmental changes cannot significantly influence the performance of gesture recognition.*
- *$H3_A$: The environmental changes can significantly influence the performance of gesture recognition.*

VI. RESULT AND ANALYSIS

This section presents the experimental results to answer three research questions through three aspects: robustness, effectiveness, and usefulness. Firstly, different parameter settings are shown. Then, the proposed SE-Conv1D model results and different baseline approaches are given. Also, the case study is presented. Finally, a detailed discussion of cost-effectiveness, computational complexity, time consumption, and applications of the proposed system is provided.

A. Answering Research Question 1: Robustness

[RQ1. Robustness]: How do the experimental parameters influence the performance of the the proposed model?

This paper combines debugging and grid search methods to select the appropriate hyperparameters. During the training phase, the ADAM optimizer [39] is utilized with a learning rate of 0.001 for 70 epochs. In addition, all the networks are trained using sparse categorical cross-entropy [40].

The reduction ratio r is an important hyperparameter that allows us to vary the capacity and computational cost of the SE blocks in the proposed model. To find the optimal value of reduction ratio r in the proposed method, we thus analyze the performance of the SE-Conv1D model under different r values (from 2 to 16), along with batch size. Meanwhile, to answer the hypothesis of RQ1, Friedman's rank tests over the accuracy of different r on each gesture type (batch size is set to 60) are conducted. The null hypothesis $H1_0$ is that the reduction ratio r cannot significantly influence the performance. The accuracy under different r is presented in Fig 10. The following remarks are made:

- Fig. 10 shows that the SE block always improves performance at different reduction ratios. with the increase of r, the accuracy goes down. The proposed SE-Conv1D model reaches the best classification accuracy with the reduction ratio 16.
- The test results (p<0.01) show that $H1_0$ is rejected, which means the parameter r influences the recognition accuracy in the proposed system.

B. Answering Research Question 2: Effectiveness

[RQ2. Effectiveness]: To what extent can the proposed system accurately recognize different human gestures?

1) Results of the Proposed SE-Conv1D Model: The training loss and validation accuracy of the SE-Conv1D model are shown in Fig. 11. By using the proposed SE-Conv1D model, the achieved overall accuracy is 99.48%, and the loss is less than 0.02. The validation accuracy curves level off after 15

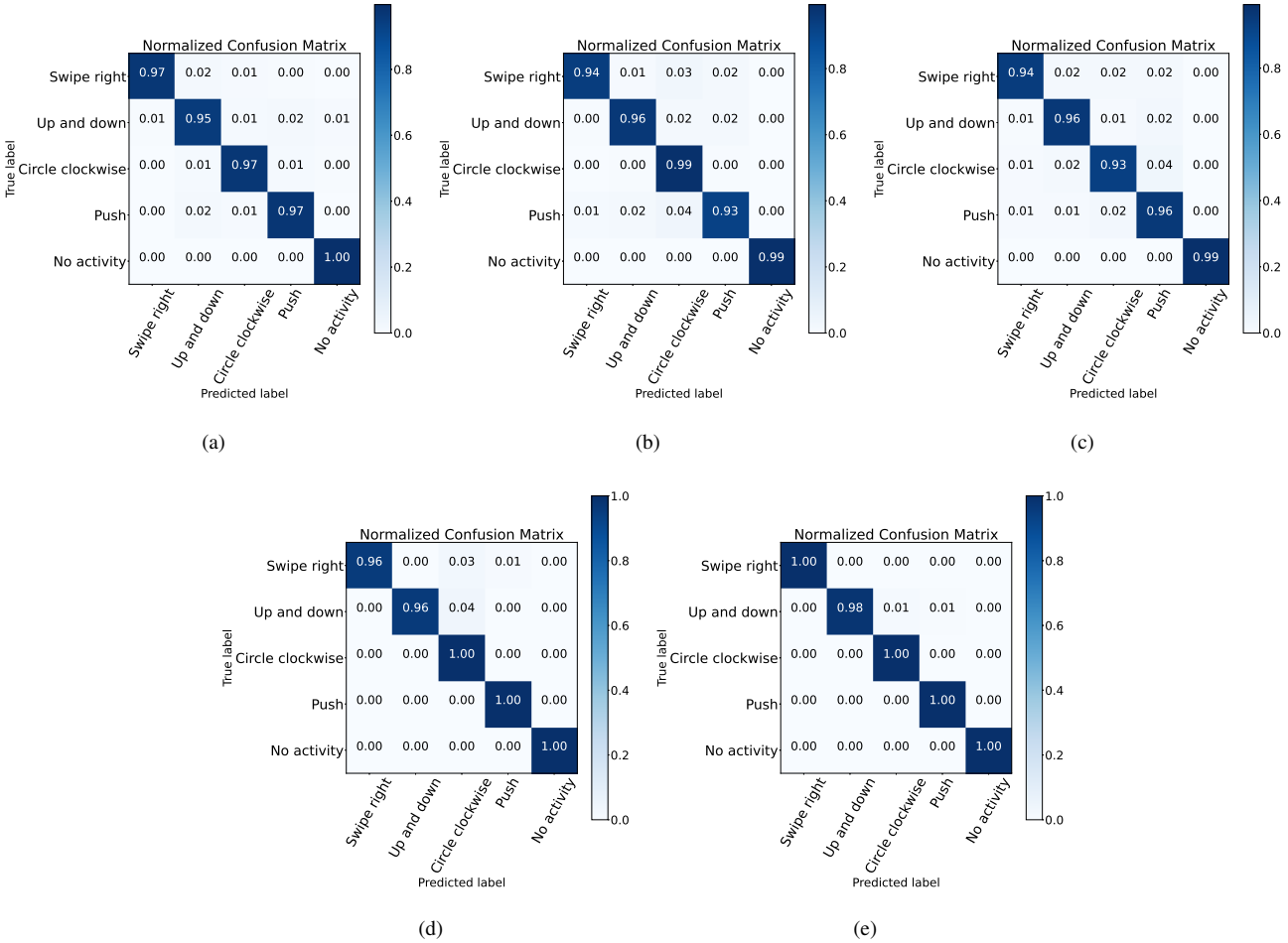

Fig. 12: Normalized confusion matrix. (a) Normalized confusion matrix using SVM-RBF. (b) Normalized confusion matrix using RF. (c) Normalized confusion matrix using KNN. (d) Normalized confusion matrix using BinaryDilatedDenseNet. (e) Normalized confusion matrix using SE-Conv1D model.

epochs and remain roughly constant thereafter, confirming that there is no overfitting in any of the deep architectures implemented.

A confusion matrix is commonly utilized for analyzing recognition performance. As shown in the normalized confusion matrix of Fig. 12, all gestures are correctly classified with nearly 100%, except the gesture 'up and down'. 1% of 'up and down' samples were misclassified into 'push', while the rest of those were misclassified into 'circle clockwise'. Although these motions have similar kinematic structures, the proposed system still achieves excellent performance. This is a significant result as it shows the clear benefits of SE-Conv1D model, especially in cases when the classes being considered are highly similar.

2) Gesture Recognition Results Comparison: The proposed SE-Conv1D model is also compared with four baselines. As shown in TABLE III, the achieved overall accuracies by SVM, RF, KNN, and BinaryDilatedDenseNet are 98.12%, 96.51%, 95.47%, and 98.57%, respectively. BinaryDilatedDenseNet performs best among four baselines. As shown in Fig. 12, SVM achieves the highest performance in 'swipe right' (97%), 'push' (97%), and 'no activity (100%). SVM achieves poor performance for the activity 'up and down', while 1% of samples have been misclassified into 'swipe right', 'circle clockwise' and 'no activity, respectively, and 2% of samples have been misclassified into 'push'. Some trajectory samples from these classes have similar patterns compared with other activities. RF shows better performance on activities 'up and down' and 'circle clockwise'. Especially for activity 'circle clockwise', the OA is nearly 100%. However, for activity 'push', the OA is only 93%. BinaryDilatedDenseNet achieves the highest performance in 'circle clockwise', 'push' and 'no activity'. All collected gesture activities are correctly

TABLE III: Comparison of the classification models.

Model	OA (%)	Recall (%)	F1 (%)
SVM-RBF	98.12	98.05	98.04
RF	96.51	96.22	96.42
KNN	95.47	95.54	95.48
BinaryDilatedDenseNet [38]	98.82	98.23	98.21
SE-Conv1D	**99.48**	**99.45**	**99.49**

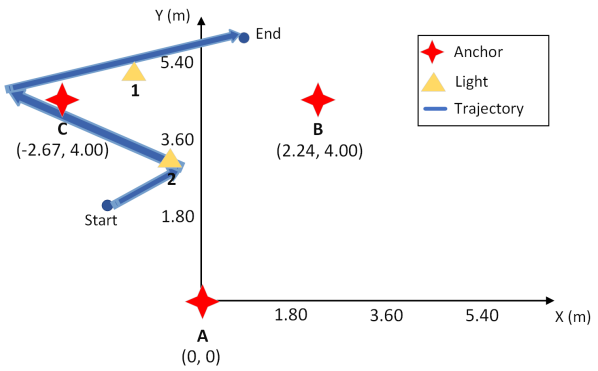

Fig. 13: The tested experiment with dynamic trajectory.

recognized with an overall accuracy of over 95%, which demonstrates that the proposed system yields stable performance among all baselines.

To assess the statistical significance of the results, Wilcoxon signed-rank tests [41] are conducted between the proposed SE-Conv1D model and four baselines. In this assessment, the level of significance used is α. Note that if the test reports a large p-value than the significant level on each of the models, then this means that there is no significant difference. Otherwise, one model is performing significantly better than the other. A hypothesis is carried out to prove beyond a shadow of a doubt that the proposed SE-Conv1D model performs better than the four baseline approaches and it does not work due to chance. In this case study, the null hypothesis $H2_0$ is that the overall gesture accuracy of the proposed SE-Conv1D model is not significantly higher than the baselines (SVM, RF, KNN, and BinaryDilatedDenseNet). The obtained p-value from the Wilcoxon test is smaller than the significance level (0.01), which suggests rejecting the null hypothesis $H2_0$. Thus, saying that the observed differences are statistically significant. In conclusion, the proposed SE-Conv1D model performs better than the four baseline approaches.

C. Answering Research Question 3: Usefulness

[RQ3. Usefulness:] To what extent can the proposed system accurately recognize different human gestures in uncontrolled environments?

In the previous experiments, it was assumed that the proposed SE-Conv1D model was trained in a controlled environment. Hence, an interesting question is how the trained SE-Conv1D model would effectively recognize hand gestures in uncontrolled environments. Twelve unseen users are invited to interact with LED lights, whose heights are from 157cm to 182 cm. It should be noted that uncontrolled environments mean i) Unseen people; ii) A new testing place; iii) Randomly combinations of distance and direction.

1) Case Study Setup: The proposed system was prototyped to interact with appliances in practical smart homes. In detail, this paper realized its functions responding to the following pre-defined gestures, which is also shown in TABLE II:

- Swipe right: Turn on the LED light.
- Up and down: Change the color of an LED light.
- Circle clockwise: Turn on all LED lights.
- Push: Turn off all LED lights.

The detailed operation is formulated in Alg.1, and the dynamic trajectory collection process for each person is shown in Fig. 13. The selection of LED lights is controlled by the user's distance to the closest light instead of adding more pre-defined gestures, which may bring convenience to real-world applications. The demonstration for this experiment can be found in the link[3]. It is worth noting that only two LED lights were used in the test experiment. In the future, more electric devices could be added.

Algorithm 1: The proposed gesture control algorithm

```
Input: Location results; Learning parameter
Output: Turning on/off light1 and/or light2; changing
        light1 and/or light2 to different colors
Initialize the locations of all anchors and lights
foreach A volunteer has been detected do
    while Gesture detection: Gesture0 do
        Turn on the nearest light;
    end
    while Gesture detection: Gesture1 do
        Change the color of the nearest light;
    end
    while Gesture detection: Gesture2 do
        Turn on both light1 and light2;
    end
    while Gesture detection: Gesture3 do
        Turn off both light1 and light2;
    end
end
```

2) Impact of User Distance and Direction: Most previous gesture recognition systems required that the user orientations be always the same since this could achieve the best recognition performance. In addition, distance is also an essential factor that could influence accuracy in most previous gesture recognition systems. This paper considers the impact of different user positions and orientations on the proposed system's performance. Eight compass directions (i.e., 0=South, 1=South-East, 2=East, 3=North-East, 4=North, 5=North-West, 6=West, 7=South-West) are applied here, with different distances between the volunteer and the LED light 1. The overall accuracy of performing gestures in an uncontrolled environment was analyzed with an independent two-sample t-test.

TABLE IV presents some results with different combinations of distances and directions. The following remarks are made:

- The average accuracy of 99.69% (p=0.34) in this experiment is very close to the average accuracy of 99.84% that we obtained in Section VI-B. The result reveals that the proposed system is robust to the changes of users' position and orientation.
- The result (p=0.34) shows there is not enough evidence to reject $H3_0$ in Section V-E. This again demonstrates that

[3]The Demo available at https://www.youtube.com/watch?v=cAG665SbTpk.

TABLE IV: The selected results of the random combinations in terms of distances and directions.

Distance (m)	Direction	Accuracy (%)
1	0	100.00
2	1	99.55
1	7	99.82
3	1	98.53
2	2	99.48
4	3	99.86
5	5	98.83
3	4	99.26

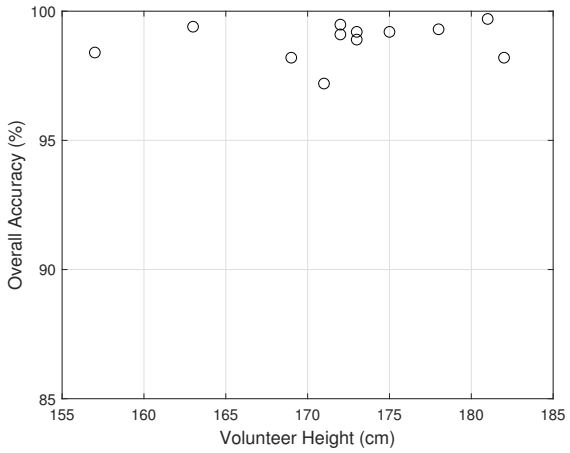

Fig. 14: Accuracy on unseen users.

this paper does not find enough evidence to support that the environmental changes can significantly influence the overall accuracy of performing gestures.

3) Impact of Unseen User: To study the impact of unseen users on the accuracy of the proposed system, we present the performance of the proposed system on 12 unseen users. Fig. 14 shows that the accuracies of unseen volunteers are always above 95% and have only small deviations across volunteers. Further, there is no trend in recognition accuracies with changing heights, demonstrating that the proposed method is unaffected by users' physiology. Most importantly, it proves that the proposed system is robust to the changes of unseen users.

4) Surveys for User Satisfaction: To quantify the user experience for the future improvement of the proposed system, a user evaluation was conducted. For this, the designed questionnaires were based on the following aspects: convenience, flexibility, accuracy, robustness, and interactivity. For each item, users were asked to grade their perception into the following levels, i.e., 'A' refers to excellent, 'B' refers to very good, 'C' refers to good, 'D' fair, or 'E' does poorly. All users found that the proposed UWB-based system is more convenient than the traditional method, e.g., using smartphones, because it does not ask the users to open the app to choose which device they want to open. More than 80% of users believed that the proposed system is more difficult to be triggered by mistake in daily usage than the voice-control solution. Compared with the current popular solutions for smart homes, the proposed system has a more convenient user experience.

TABLE V: Comparison with related works.

Study	Device	Algorithm	Accuracy
[43]	IR-UWB Radar	K-means	97%
[42]	IR-UWB Radar	CNN	97%
[13]	UWB Radar	Stacked-LSTM	97%
[5]	UWB equipped smart-watch	HMM	97%
Proposed	**UWB communication**	**SE-Conv1D**	**99.48%**

5) Comparison with Related Works: TABLE V presents the performance comparison of the proposed approach with the existing state-of-the-art methods based on UWB technology. Ahmed *et al.* [42] performed gesture recognition within cars using impulse UWB radar with CNN, which achieved an overall accuracy of approximately 97%. Khan *et al.* [43] proposed a novel gesture recognition algorithm, which achieved the best performance of 97%. Maitre *et al.* [13] recognized activities of daily living from UWB radars by using a stacked long short-term memory (LSTM). Alanwar *et al.* [5] used a UWB equipped smart-watch to interact with different devices in smart homes, which achieved 97% overall accuracy for gesture recognition based on the Hidden Markov model (HMM). It should be noted that it is impossible to directly compare previous works and ours because none of them use a trajectory-based UWB communication system. However, we still can argue that the proposed solution is robust against changes in distance or direction and gives a better accuracy rate in trajectory-based human gesture recognition than other standard gesture-based learning algorithms. It proves that trajectory-based activity recognition is a potentially promising method for real-world applications.

D. Discussion

A detailed discussion of the cost-effectiveness, computational complexity, time consumption, and applications of the proposed system is provided as follows:

1) Cost Effectiveness: The cost of a UWB module is approximately around 10 US dollars, which is lightweight, portable, inexpensive to be worn as accessories such as belts and wristbands. The proposed UWB communication system contains four UWB modules, approximately 40 US dollars.

2) Computational Complexity: For the proposed SE block to be viable in practice, it must provide an effective trade-off between model complexity and performance, which is essential for scalability. In this paper, the proposed SE-Conv1D model has 538,245 total parameters, with 537,221 trainable parameters. The SE block automatically recalibrates the incoming feature maps with the reduction ratio r of 16. The number of parameters required to learn these maps is reduced such that the overall model size increases by just 3-10%. More precisely, the number of additional parameters introduced is given by:

$$P = \frac{2}{r} \sum_{s=1}^{S} N_s \cdot C_s^2, \tag{9}$$

where P is the total number of additional parameters, r denotes the reduction ratio, S refers to the number of stages (where each stage refers to the collection of blocks operating on feature maps of a common spatial dimension), C_s denotes the number of output feature maps for stage s and N_s denotes the repeated block number for stage s. Thus, given r = 16, the number of parameters incurred by SE blocks is $\frac{2}{16} \cdot (128^2 + 256^2) = 10240$. This provides evidence of the effectiveness of the SE blocks in improving the classification performance in gesture recognition tasks with the proposed SE-Conv1D model exhibiting fewer parameters that make it efficient and effective.

3) Analysis about Time Consumption: The hardware platform is a laptop with an Intel(R) Core(TM) i7-10510U CPU. The CPU clock frequency and the memory size are 2.3 GHz and 16.0 GB, respectively. The software platform is python with TensorFlow, and the operating system is Windows 10. The SE-Conv1D model consumes around 7min30s for training and testing all gestures. In real-world applications, gesture recognition will require real-time processing. Since the running time for classifying one hand gesture by the proposed SE-Conv1D model is only 0.023s, which means the gesture recognition system can respond efficiently.

4) Applications and Impacts: The overall idea of the proposed system is to showcase the potential of UWB-sensing in realizing a vital application. Such a cost-effective UWB-based system is lightweight and portable to be worn as accessories such as belts and wristbands, which could be easily adopted by the public. More functions of the proposed system will be researched in the coming future, such as opening air conditioners, playing music, etc.

VII. Limitations and Future Work

Our experimental results show that the proposed system outperforms competitive baselines by a significant gesture recognition accuracy. However, there are several additional issues that we must consider in the following works.

Limitation of fine-grained gestures: The current version of the proposed method is limited to monitoring relatively fine-grained gestures, e.g., finger movements. The reason is that the core of the proposed solution depends on the changes in gesture trajectory. However, fine-grained gestures are usually detected at close range, where the millimeter level is required.

Limitation of multi-hand gestures: The current approach is designed for tracking the gesture of a single arm. In future work, multi-object tracking technology should be considered to solve the multi-hand gestures recognition problem.

VIII. Conclusion

In this paper, a novel solution was proposed to perform gesture recognition by using a low-cost UWB communication system. With the aid of SE-Conv1D model, trajectories of different human gestures could be used directly to recognize different human gestures. This paper presented the experimental results on the collected datasets of four different human gesture activities, and evaluated the results from the standpoints of robustness, effectiveness, and usefulness. The proposed SE-Conv1D model achieved an excellent result of 99.48% OA ($p<0.01$), which is superior to the results achieved by the baseline models. In addition, the proposed system was prototyped to interact with appliances in smart homes. The experiment results showed that the proposed system system generally outperforms the state-of-the-art approach based on UWB technology. It proves that the proposed system is a complete end-to-end sensing system specifically designed to track and recognize human gestures. Note that the archived UWB gesture datasets and code have been published. The future work will focus on monitoring relatively fine-grained gestures and multi-hand gestures recognition problems.

References

[1] H. Liu, Y. Wang, A. Zhou, H. He, W. Wang, K. Wang, P. Pan, Y. Lu, L. Liu, and H. Ma, "Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 4, pp. 1–28, 2020.

[2] X. Chen, L. Gong, L. Wei, S.-C. Yeh, L. Da Xu, L. Zheng, and Z. Zou, "A wearable hand rehabilitation system with soft gloves," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 943–952, 2021.

[3] K. M. Sagayam and D. J. Hemanth, "Hand posture and gesture recognition techniques for virtual reality applications: a survey," *Virtual Reality*, vol. 21, no. 2, pp. 91–107, 2017.

[4] G. Du, B. Zhang, C. Li, B. Gao, and P. X. Liu, "Natural human–machine interface with gesture tracking and cartesian platform for contactless electromagnetic force feedback," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 11, pp. 6868–6879, 2020.

[5] A. Alanwar, M. Alzantot, B.-J. Ho, P. Martin, and M. Srivastava, "Selecon: Scalable IoT device selection and control using hand gestures," in *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation*, 2017, pp. 47–58.

[6] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 3, pp. 478–500, 2009.

[7] Y. Wu and T. S. Huang, "Vision-based gesture recognition: A review," in *International gesture workshop*. Springer, 1999, pp. 103–115.

[8] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th annual international conference on Mobile computing & networking*, 2013, pp. 27–38.

[9] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "Landmarc: Indoor location sensing using active RFID," in *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.(PerCom 2003).* IEEE, 2003, pp. 407–415.

[10] Y. Kim and B. Toomajian, "Hand gesture recognition using micro-Doppler signatures with convolutional neural network," *IEEE Access*, vol. 4, pp. 7125–7130, 2016.

[11] C. Ding, H. Hong, Y. Zou, H. Chu, X. Zhu, F. Fioranelli, J. Le Kernec, and C. Li, "Continuous human motion recognition with a dynamic range-Doppler trajectory method based on FMCW radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6821–6831, 2019.

[12] K. A. Hamdi, "On the statistics of signal-to-interference plus noise ratio in wireless communications," *IEEE transactions on communications*, vol. 57, no. 11, pp. 3199–3204, 2009.

[13] J. Maitre, K. Bouchard, C. Bertuglia, and S. Gaboury, "Recognizing activities of daily living from UWB radars and deep learning," *Expert Systems with Applications*, vol. 164, p. 113994, 2021.

[14] M. R. Mahfouz, C. Zhang, B. C. Merkl, M. J. Kuhn, and A. E. Fathy, "Investigation of high-accuracy indoor 3-D positioning using UWB technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 6, pp. 1316–1330, 2008.

[15] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.

[16] J. Park and S. H. Cho, "IR-UWB radar sensor for human gesture recognition by using machine learning," in *2016 IEEE 18th International Conference on High Performance Computing and Communications*, 2016, pp. 1246–1249.

[17] F. Khan, S. K. Leem, and S. H. Cho, "Hand-based gesture recognition for vehicular applications using IR-UWB radar," *Sensors*, vol. 17, no. 4, p. 833, 2017.

[18] S. Ahmed, F. Khan, A. Ghaffar, F. Hussain, and S. H. Cho, "Finger-counting-based gesture recognition within cars using impulse radar with convolutional neural network," *Sensors*, vol. 19, no. 6, 2019.

[19] F. Luo, S. Poslad, and E. Bodanese, "Temporal convolutional networks for multiperson activity recognition using a 2-D lidar," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7432–7442, 2020.

[20] H. Cheng, L. Yang, and Z. Liu, "Survey on 3D hand gesture recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1659–1673, 2016.

[21] I. Guyon, V. Athitsos, P. Jangyodsuk, B. Hamner, and H. J. Escalante, "Chalearn gesture challenge: Design and first results," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2012, pp. 1–6.

[22] F. Chen, H. Lv, Z. Pang, J. Zhang, Y. Hou, Y. Gu, H. Yang, and G. Yang, "Wristcam: A wearable sensor for hand trajectory gesture recognition and intelligent human–robot interaction," *IEEE Sensors Journal*, vol. 19, no. 19, pp. 8441–8451, 2019.

[23] T. H. Vu, A. Misra, Q. Roy, K. C. T. Wei, and Y. Lee, "Smartwatch-based early gesture detection 8 trajectory tracking for interactive gesture-driven applications," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, Mar. 2018.

[24] L. Rundo, C. Han, Y. Nagano, J. Zhang, R. Hataya, C. Militello, A. Tangherloni, M. S. Nobile, C. Ferretti, D. Besozzi *et al.*, "USE-Net: Incorporating squeeze-and-excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets," *Neurocomputing*, vol. 365, pp. 31–43, 2019.

[25] Y. Li, Y. Liu, W.-G. Cui, Y.-Z. Guo, H. Huang, and Z.-Y. Hu, "Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 4, pp. 782–794, 2020.

[26] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 5, pp. 751–755, 2019.

[27] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067–1080, 2007.

[28] D. Zhang, L. T. Yang, M. Chen, S. Zhao, M. Guo, and Y. Zhang, "Real-time locating systems using active RFID for internet of things," *IEEE Systems Journal*, vol. 10, no. 3, pp. 1226–1235, 2014.

[29] D. Dardari, A. Conti, U. Ferner, A. Giorgetti, and M. Z. Win, "Ranging with ultrawide bandwidth signals in multipath environments," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 404–426, 2009.

[30] P. Corbalán, G. P. Picco, and S. Palipana, "Chorus: UWB concurrent transmissions for GPS-like passive localization of countless targets," in *2019 18th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2019, pp. 133–144.

[31] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[32] H. Zhang, X. Zhao, Z. Wu, B. Sun, and T. Li, "Motor imagery recognition with automatic EEG channel selection and deep learning," *Journal of Neural Engineering*, vol. 18, no. 1, p. 016004, 2021.

[33] L. Su, L. Ma, N. Qin, D. Huang, and A. H. Kemp, "Fault diagnosis of high-speed train bogie by residual-squeeze net," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 3856–3863, 2019.

[34] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: a local SVM approach," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 3. IEEE, 2004, pp. 32–36.

[35] Z. Ramedani, M. Omid, A. Keyhani, S. Shamshirband, and B. Khoshnevisan, "Potential of radial basis function based support vector regression for global solar radiation prediction," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 1005–1011, 2014.

[36] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "Knn model-based approach in classification," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer, 2003, pp. 986–996.

[37] M. Pal, "Random forest classifier for remote sensing classification," *International journal of remote sensing*, vol. 26, no. 1, pp. 217–222, 2005.

[38] F. Luo, S. Khan, Y. Huang, and K. Wu, "Binarized neural network for edge intelligence of sensor-based human activity recognition," *IEEE Transactions on Mobile Computing*, 2021.

[39] Z. Zhang, "Improved adam optimizer for deep neural networks," in *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, 2018, pp. 1–2.

[40] P. Covington, J. Adams, and E. Sargin, "Deep neural networks for youtube recommendations," in *Proceedings of the 10th ACM conference on recommender systems*, 2016, pp. 191–198.

[41] F. Wilcoxon, "Individual comparisons by ranking methods," in *Breakthroughs in statistics*. Springer, 1992, pp. 196–202.

[42] S. Ahmed, F. Khan, A. Ghaffar, F. Hussain, and S. H. Cho, "Finger-counting-based gesture recognition within cars using impulse radar with convolutional neural network," *Sensors*, vol. 19, no. 6, 2019.

[43] F. Khan, S. K. Leem, and S. H. Cho, "Hand-based gesture recognition for vehicular applications using IR-UWB radar," *Sensors*, vol. 17, no. 4, 2017.