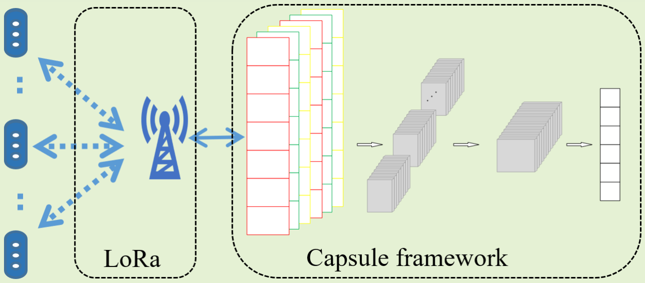
Real-Time Human Activity Recognition System Based on Capsule and LoRa

Leixin Shi, Hongji Xu, *Member, IEEE*, Wei Ji, Beibei Zhang, Xiaojie Sun, and Juan Li



***Abstract*—Human activity recognition (HAR) has become a research hotspot in the field of artificial intelligence and pattern recognition. However, the HAR system still has some deficiencies in the aspects of platform algorithms and wireless access technologies. On the one hand, some state-of-the-art frameworks such as convolutional neural network (CNN) and recurrent neural network (RNN) have been proven successfully in classification tasks of HAR, while those frameworks just identify the feature data of activity but ignore the spatial relationship among features, which may lead to incorrect recognition. On the other hand, some existing transmission modes, such as Bluetooth and 4G, are difficult to realize real-time transmission in the case of a large range and low-power consumption. In this paper, a real-time human activity recognition system based on capsule and "long range" (LoRa) is presented, which pioneers the application of capsule to HAR. The capsule framework encapsulates the multiple convolution layers in parallel to solve the defect that current frameworks cannot identify the spatial relationship among features. Simultaneously, the combination of long-distance transmission and low-power consumption is achieved by using LoRa networking technology instead of other existing transmission modes. The experiments are performed on the dataset WISDM that is collected by the Wireless Sensor Data Mining Lab in Fordham University, and the results demonstrate that the proposed capsule framework achieves a higher classification result than CNN and RNN, and the proposed system makes the real-time HAR based on intelligent sensor devices possible in some special scenarios such as smart prison.**

***Index Terms*— Human activity recognition (HAR), convolutional neural network (CNN), recurrent neural network (RNN), capsule, long range (LoRa).**

I. Introduction

HUMAN activity recognition (HAR) can realize the activity recognition of users by acquiring their activity information and using a reasonable algorithm model. The research of HAR can be traced back to the 1990s [1]. With the development and maturity of advanced technologies, such as internet of things (IoT), artificial intelligence (AI) and cloud computing, more and more scholars have devoted themselves to the research of HAR. There are two main ways of obtaining the activity information in HAR: the vision-based HAR and the wearable-based HAR. Studies have shown that the vision-based HAR has certain shortcomings in terms of privacy and space-time applicability [2], [3]. For example, the activity information will be impossible or unreliable in the blind area of cameras or in the dark environment. Meanwhile, the development of the intelligent terminal provides a good opportunity for the activity recognition based on wearable devices [4], [5]. Nowadays, HAR technologies based on wearable devices have been applied in many fields, such as human motion analysis, smart home, human-computer interaction, medical diagnosis and intelligent monitoring [6], [7].

At present, the activity information from wearable devices is mainly transmitted through Bluetooth and 4G [8], but these transmission technologies cannot be compatible in terms of power consumption and transmission distance. Long-distance transmission can make wearable devices no longer dependent on mobile phones, computers and other terminals, while low-power consumption can improve the stand-by time of wearable devices. Therefore, the transmission technology with low-power consumption and long-distance transmission is very important for the real-time HAR systems [9], [10].

Manuscript received April 22, 2020; revised June 8, 2020; accepted June 9, 2020. Date of publication June 23, 2020; date of current version December 4, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0831001 and Grant 2017YFC0803403, in part by the National Natural Science Foundation of China under Grant 61771292, and in part by the Natural Science Foundation of Shandong Province of China under Grant ZR2016FM29. The associate editor coordinating the review of this article and approving it for publication was Prof. Dongsoo Har. *(Corresponding author: Hongji Xu.)*

The authors are with the School of Information Science and Engineering, Shandong University, Qingdao 266237, China (e-mail: hongjixu@sdu.edu.cn).

Digital Object Identifier 10.1109/JSEN.2020.3004411

After the activity information is acquired, it needs to be identified by the activity recognition framework. The mainstream activity recognition frameworks are traditional data mining, machine learning and deep learning [11]–[14]. Researches show that traditional data mining uses pre-designed features, namely "shallow" features, to represent the activity information of classification tasks and has a better performance in simple activity recognition. Machine learning mainly includes k-nearest neighbor (KNN), support vector machines (SVM), random forests (RF), etc. [15]–[17], and achieves a better performance in complex activity recognition than traditional data mining. Compared with machine learning, deep learning can automatically extract deeper features to achieve a better performance. Typical deep learning models include convolutional neural network (CNN) and recurrent neural network (RNN) [18]–[23]. However, they only consider whether the activity information contains certain features, but cannot identify the spatial relationship among features, which may result in a certain false positive rate.

In order to improve the shortcoming of the current frameworks, the capsule network was first proposed in 2011 by G. Hinton [24] and achieved state-of-the-art accuracy on MNIST in 2017 [25]. The capsule network consists of capsules, which are groups of neurons. The input and output of the capsule network are vectors, whose length represents the existence probability of entities. The orientation of a vector contains information about the instantiation parameters. The capsule network has been applied into license plate recognition, crying recognition, etc. [26], [27].

In this paper, a novel capsule framework is first proposed in the field of HAR to improve the accuracy of the activity recognition. At the same time, "long range" (LoRa) technology is adopted, which realizes low-power consumption and long-distance transmission of signals from wearable devices.

The rest of this paper is organized as follows. The related works on HAR, including traditional data mining, machine learning and deep learning for HAR, are described in Section II. In Section III, we firstly introduce a detailed explanation of the proposed system. Secondly, we give a brief introduction of our proposed framework for HAR. In Section IV, we test the proposed framework on an open human activity dataset WISDM, and make a comparative analysis between our proposed framework and some state-of-the-art HAR frameworks. Additionally, we study the influence of self-parameters on the accuracy rate. Finally, the conclusions and some future works are described in Section V.

II. Related Work

With the rapid development of AI, the frameworks of HAR have been constantly optimized. Traditional data mining manually extracts multiple sets of features from time-series signals and then maps these features to various human activities [28]. This kind of framework is only suitable for some simple applications. In view of the low recognition rate of traditional data mining methods in complex applications, Yin *et al.* first employed a one-class SVM for the activity recognition [29]. Preece *et al.* adopted the KNN algorithm to compare the classification accuracy for each feature set across different combinations of three accelerometer placements [30]. Mannini and Sabatini discussed how to classify HAR using accelerometers and hidden markov models (HMM) [31]. Cheng and Jhan proposed a cascade-AdaBoost SVM classifier to implement a three-axis acceleration-based fall detection [32]. Altun *et al.* [33] and Cleland *et al.* [34] used the principal component analysis (PCA) to extract a feature set from raw sensor data for classification, and also compared some machine learning classification techniques, such as bayesian decision tree (BDT), least squares (LS), KNN and SVM.

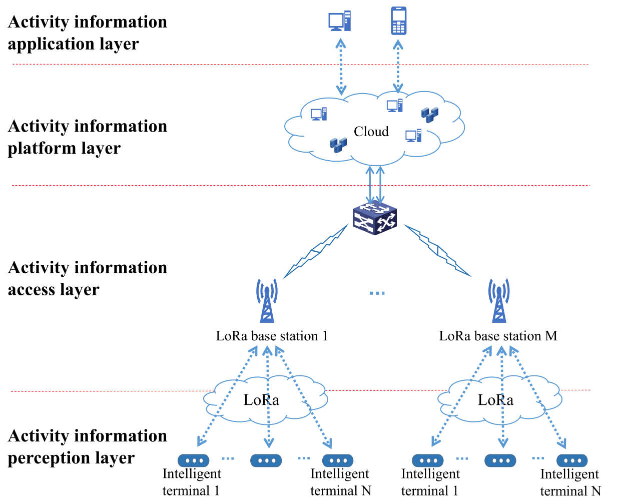

Activity information application layer
Activity information platform layer
Cloud
Activity information access layer
LoRa base station 1
LoRa base station M
LoRa
LoRa
Activity information perception layer
Intelligent terminal 1
Intelligent terminal N
Intelligent terminal 1
Intelligent terminal N

Fig. 1. The system framework of the real-time human activity recognition system based on capsule and LoRa.

In order to optimize the problem, i.e., low accuracy of machine learning methods in complex data classification, deep learning was proposed by Hinton *et al.* in 2006 [35], but it cannot be popularized due to the limitations of hardware. With the improvement of hardware processing ability, deep learning begins to enter our study and life. The typical frameworks in deep learning are CNN and RNN.

Chen and Xue proposed an acceleration-based human activity recognition model, i.e., a deep architecture model of neural network [36]. Yang *et al.* proposed a system feature learning method for HAR, using a deep CNN to automate feature learning from the raw inputs in a systematic way. Through deep architecture, the learning features were considered as the higher-level abstract representations of low-level original time series signals [37]. Zebin *et al.* presented a feature learning method and researched various important hyperparameters such as the number of convolutional layers and the kernel size on the performance of CNN [38]. Zeng *et al.* proposed a partial weight sharing technique for CNN, which had a significant improvement in accuracy [39]. Ha and Choi proposed a partial weight sharing and full weight sharing mechanism for multi-modal data [40]. Ronao and Cho compared the performance of RF and CNN, and the results showed that the recognition accuracy of CNN was higher than that of RF [41]. To sum up, the accuracy of CNN is significantly higher than that of machine learning in the activity recognition based on wearable devices.

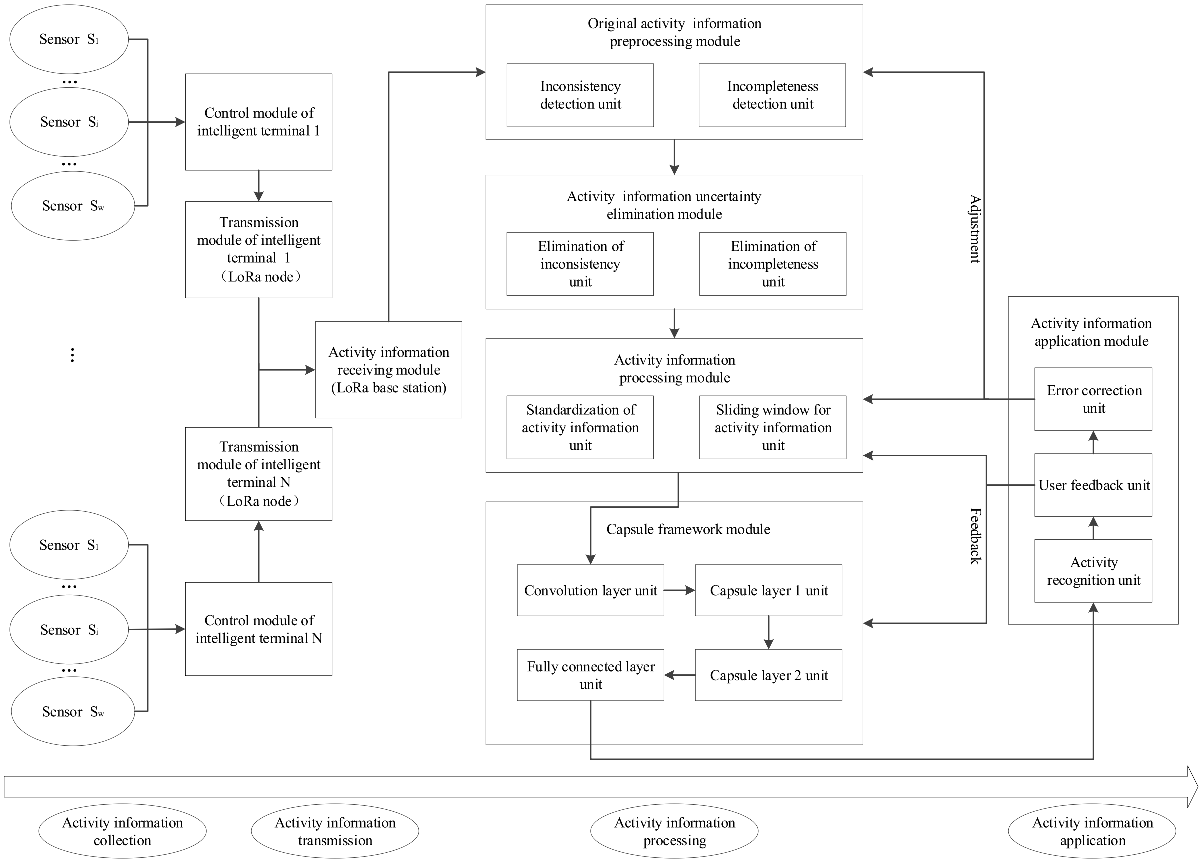

Sensor S_1
Sensor S_i
Sensor S_w
Control module of intelligent terminal 1
Transmission module of intelligent terminal 1 （LoRa node）
Activity information receiving module (LoRa base station)
Transmission module of intelligent terminal N （LoRa node）
Control module of intelligent terminal N
Sensor S_1
Sensor S_i
Sensor S_w
Original activity information preprocessing module
Inconsistency detection unit
Incompleteness detection unit
Activity information uncertainty elimination module
Elimination of inconsistency unit
Elimination of incompleteness unit
Activity information processing module
Standardization of activity information unit
Sliding window for activity information unit
Capsule framework module
Convolution layer unit
Capsule layer 1 unit
Capsule layer 2 unit
Fully connected layer unit
Adjustment
Feedback
Activity information application module
Error correction unit
User feedback unit
Activity recognition unit
Activity information collection
Activity information transmission
Activity information processing
Activity information application

Fig. 2. Module composition and connection diagram of the proposed real-time human activity recognition system based on capsule and LoRa.

The data collected by the wearable device is based on time series. Considering the deficiencies of CNN in dealing with data based on time series, Ordonez and Roggen proposed a general deep activity recognition framework based on convolution and long-term and short-term memory (LSTM) cycle units [42]–[44]. An LSTM-based feature extraction method was proposed by Chen *et al.*, the simulation results on public dataset WISDM showed that the accuracy rate reached 92.1% [45]. Padmaja *et al.* compared the performance of CNN and LSTM models, and the recognition rate of LSTM model was 4% higher than that of CNN [46].

According to the above researches, the accuracy of HAR is effectively improved. However, at present, the inputs and outputs are scalar in mainstream frameworks of HAR, which cannot identify the spatial relationship among features and leads to a certain misjudgment rate.

III. Human Activity Recognition System Based on Capsule and LoRa

In order to make up for the shortcomings of platform algorithms and access technologies of the existing HAR system based on wearable devices, a real-time HAR system based on capsule and LoRa is proposed. This section introduces two aspects. One is the architecture and implementation of the proposed system including system framework, module composition, processing flow and concrete implementation, and the other is the proposed capsule framework.

A. The Architecture and Implementation of the Proposed System

The system framework includes activity information perception layer, activity information access layer, activity information platform layer and activity information application layer. The activity information perception layer mainly collects activity information through intelligent terminals. The activity information access layer mainly transmits activity information through LoRa. The activity information platform layer mainly processes and identifies the activity information through data preprocessing and the proposed capsule framework. The activity information application layer mainly adjusts the stability and generalization of the whole system. The system framework of the real-time human activity recognition system based on capsule and LoRa is shown in Fig. 1. The difference between the proposed framework and the mainstream frameworks is that the LoRa based networking technology can achieve low-power consumption and long-distance transmission, simultaneously.

Fig. 2 shows the module composition and connection diagram of the proposed real-time human activity recognition system based on capsule and LoRa. As shown in Fig. 2, the proposed system consists of the sensor module, the control module of intelligent terminal, the transmission module of intelligent terminal, the activity information receiving module, the original activity information preprocessing module, the activity information uncertainty elimination module,

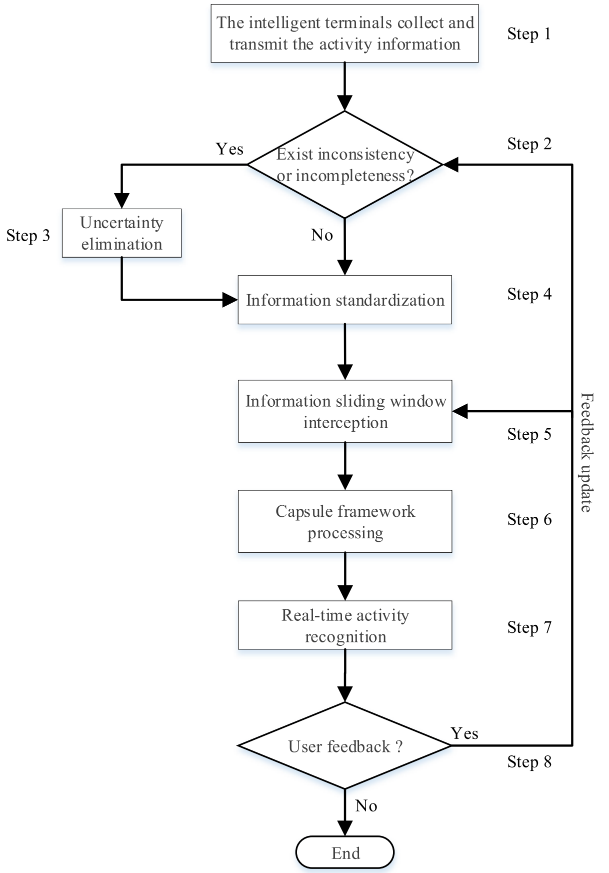

The intelligent terminals collect and transmit the activity information
Step 1
Exist inconsistency or incompleteness?
Step 2
Yes
No
Uncertainty elimination
Step 3
Information standardization
Step 4
Information sliding window interception
Step 5
Capsule framework processing
Step 6
Real-time activity recognition
Step 7
User feedback ?
Yes
No
Step 8
Feedback update
End

Fig. 3. Flowchart of the real-time human activity recognition system based on capsule and LoRa.

the activity information processing module, the capsule framework module and the activity information application module and each module is composed of some specific units. The sensor module with different types of sensors, the microcontroller based control module and the LoRa based transmission module are integrated into the intelligent terminal to collect and transmit the activity information. The original activity information preprocessing module and the activity information uncertainty elimination module deal with the uncertainty of the activity information including inconsistency and incompleteness to improve the credibility of the activity information. The activity information processing module and the network architecture module automatically acquire the spatial relationship among features of the activity information to achieve the activity recognition with higher precision. The activity information application module mainly consists of the activity recognition unit, the error correction unit and the user feedback unit, which improves the stability and generalization of the system.

As shown in Fig. 3, the implement process of the proposed system is as follows:

Step 1: Collect the activity information through intelligent terminals and transmit through the LoRa network.

Step 2: After obtaining the activity information, the system detects whether the information exists uncertainty or not. If yes, the system will execute Step 3, otherwise execute Step 4.

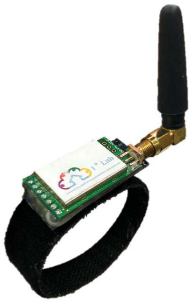


Fig. 4. The wristband to collect the activity information from users.

Step 3: Process the uncertain information according to the degree of uncertainty, such as deletion and context prediction filling, to increase the credibility of the activity information.

Step 4: Standardize various types of information to increase the applicability of the model. The system normalizes numerical data such as acceleration and angular velocity, and uses one-hot encoding for activity types.

Step 5: The time-series interception of the activity information is performed by sliding window, and the activity of users at this time is judged by the change of the activity information in a short time.

Step 6: Build a capsule network model and train the model parameters through a dataset with high credibility and large volume of data.

Step 7: The real-time activity recognition is realized after the activity information is transmitted to the capsule network model.

Step 8: Detect whether there is user feedback in the activity recognition, if yes, the system will adjust and feedback the parameters, such as uncertainty determination threshold (in Step 2), sliding window mechanism (in Step 5), and parameters of the capsule network model (in Step 6), otherwise, this process will go to the end.

The concrete implementation of the proposed system is as follows:

1) Collection of the Activity Information—Wristband: In the proposed system, the wristband is designed to collect the activity information and consists of the sensor module, the microcontroller based control module and the LoRa based transmission module. The sensor module adopts the MPU-6050 to collect the data of 3-axis gyroscope and 3-axis accelerometer. The microcontroller based control module uses STM32 as the controlling core. The LoRa based transmission module used in the wristband is based on SX1278, which is produced by Rejeee, a company focusing on the development of LoRa products. Fig. 4 shows the wristband to get the activity information from users.

2) Transmission of the Activity Information—LoRa: LoRa is a wireless communication technology developed to create the low-power wide-area network (LPWAN) required for different IoT applications. LoRa is pillared on its patented chirp spread spectrum modulation that supports energy-efficient and reliable long-range communication, and it can expand the transmission distance of traditional short-range wireless communication technologies. The transmission of activity

TABLE I
THE PERFORMANCE COMPARISON OF SEVERAL TYPICAL WIRELESS ACCESS TECHNOLOGIES

Wireless Access Technology	Power Consumption	Transmission Distance	Transmission Rate
LoRa	Medium	Medium	Low
Bluetooth	Low	Short	Medium
4G	High	Long	High

information in the proposed system adopts the LoRa technology, which offers a lot of flexibility in the network configuration. For example, the administrator can monitor the activity information of all personnel by using the star topology structure, while if users only want to test or detect their own activity information, the point-to-point network can be adapted. This kind of multi-mode networking strategy improves the generalization of the system. In addition, LoRa shows good immunity to interferences.

Table I shows the performance comparison of several typical wireless access technologies. From the table, we can see that 4G and Bluetooth have a great advantage in transmission rate, but the high-power consumption of 4G will greatly reduce the standby time of wristband, and Bluetooth cannot meet the system requirement due to its limited transmission distance. LoRa achieves a balance between power consumption and transmission distance, which enables the wristband to realize real-time activity recognition with a long distance and a long standby time. In the real-time monitoring scenario with a relatively small population and a large area, such as smart prison, LoRa can be an optional good wireless access method. However, LoRa also faces many challenges. For one thing, it is inherently weak in network deployment due to the lack of operator, and for another, it is limited by the transmission rate and cannot be applied to some high data rate scenarios.

3) Preprocessing of the Activity Information: There could be uncertainty after the activity information is affected by some factors. The uncertainty of the activity information mainly includes incompleteness and inconsistency. If a small amount of data is found to be incomplete, the system would delete the missing data. When a large amount of data is incomplete, the system would fill in the missing data through the context prediction. If there is any inconsistency of the data, the system will use the Dempster-Shafer theory for the inconsistency elimination. After the uncertainty of the original data is eliminated, the activity information will have higher credibility, and then the system standardizes the processed information. There are three main types of data standardization.

(1) Standardization of category-type features uses one-hot encoding.

(2) Standardization of numerical features uses normalization.

(3) Standardization of ordered features uses ordered numerical encoding.

The activity information belongs to the numerical features, so we select the normalization, and the labels in this paper belong to the category-type features, so the one-hot encoding is adopted. The one-hot encoding mainly uses L-bit status registers to encode L states. Each state has its independent register bits, and only one bit is valid at any time. The normalization unifies the scalar to [0, 1], which is convenient for calculation and may help the system to find the global optimal solution. The normalized value of the input is given by

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

where X is the real-time sensor value at a certain moment, X_{max} and X_{min} are the maximum and minimum values for a given range, respectively, and X_{scaled} is the value after standardization.

The next step is time series segmentation for activity information, which uses a sliding window of a fixed length and split each time series into equal segments. In the proposed system, the length of the sliding window is 90, and the sliding mechanism uses 50% data overlap, i.e., the ending point of the next sliding window is the middle point of the current sliding window.

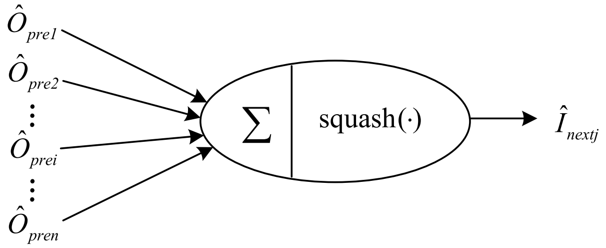


Fig. 5. The working principle diagram of capsule.

B. The Capsule Framework

As we all know, the neural network is composed of many neurons. However, the basic constituent unit of the capsule framework is a capsule rather than a neuron. Each capsule cannot only obtain the features of the object, but also capture the relative position of the object. The direction of the output vector changes accordingly, if the pose of the object changes. The biggest difference from previous frameworks is that the capsule framework is vector-oriented. The vectorization of input and output indicates not only the features of the data but also the spatial relationship among the data features [47]. The working principle diagram of capsule is shown in Fig. 5. The output vector of each capsule in the previous layer is weighted and summed, and then processed using a nonlinear function as the output of the current capsule.

where $\hat{O}_{prei}$ is a prediction vector, which is produced by multiplying the output of a capsule in the layer below by a weight matrix. $\hat{I}_{nextj}$ is the output vector of the current capsule. squash(·) is a new nonlinear function similar to the previously used nonlinear functions, such as tanh(·) and relu(·), and it performs a nonlinear processing for vector information while the other nonlinear functions are mainly for scalar information. The main function of squash(·) is to make

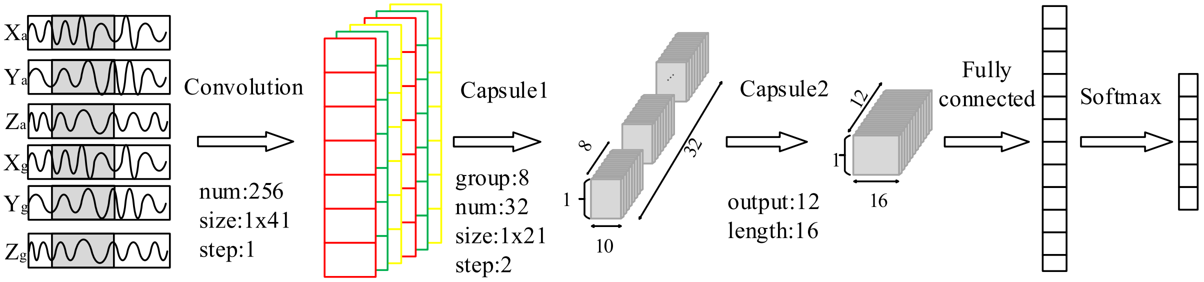


Fig. 6. Block diagram of the proposed capsule network.

the length of the transformed vector no more than 1, and keep the same direction as the original vector, and it is realized by

$$\hat{C}_j = \sum_{i=0}^{n} \hat{O}_{prei} \times w_{ij} \tag{2}$$

$$\hat{I}_{nextj} = \frac{||\hat{C}_j||^2}{1 + ||\hat{C}_j||^2} \times \frac{\hat{C}_j}{||\hat{C}_j||} \tag{3}$$

where $\hat{C}_j$ is a weighted sum over all $\hat{O}_{prei}$ from the capsules in the layer below, w_{ij} refers to a coupling coefficient that is adjusted by the iterative dynamic routing, and it is realized by

$$w_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \tag{4}$$

where b_{ij} is the log prior probability that capsule i should be coupled to capsule j, and its initial value is set by measuring the agreement between the current output $\hat{I}_{nextj}$ of each capsule and the prediction $\hat{O}_{prei}$. The coupling coefficient is iteratively refined by the agreement $\hat{I}_{nextj} \times \hat{O}_{prei}$.

The specific block diagram of the capsule network used in the activity recognition framework in this paper is shown in Fig. 6. The proposed framework is composed of five layers, which are the convolution layer, the first capsule layer, the second capsule layer, the fully connected layer and the classifier layer. The working principle of the convolution layer is to change the input of the current scalar information into the vector information, so that the spatial features can be added among the features of each behavior information. A dynamic routing protocol is used between the capsule layers to update the framework's parameters. The number of updated iterations will have a certain impact on the performance of the framework. The specific process of the proposed network architecture model is as follows.

Step 1: Enter the activity information whose size is $Batch\_size \times 1 \times Window\_size \times 3$, where $Batch\_size$ refers to the number of groups entered per iteration, and $Window\_size$ refers to the length of each input activity information.

Step 2: After the activity information passes through the convolution layer, the input activity information is converted from a scalar to a vector by

$$Y_j = \sum_{i=1}^{n} X_i \times W_{ij} + b_j \tag{5}$$

where X_i refers to the i-th activity information, which is processed through the uncertainty processing, standardization and time series based on sliding window, W_{ij} refers to the j-th weight parameter of the convolutional layer corresponding to X_i, whose initial value is a random number generated by truncating a normal distribution, b_j refers to the offset parameter of the convolution layer, and its initial value is set to be 0, n indicates the number of convolution kernels, and Y_j is the output of the convolution layer, which is a vector that satisfies the input requirements of the capsule network. The size of the output information is $Batch\_Size \times 1 \times \frac{Window\_Size - Nuclear\_Size_1 + 1}{L_1} \times n$, where $Nuclear\_Size_1$ refers to the size of each convolution kernel. It is necessary to ensure that the result of (5) is a positive integer.

Step 3: The variable Con_layer is used to represent $\frac{Window\_Size - Nuclear\_Size_1 + 1}{L_1}$. The system encapsulates the m sets of convolution kernels in the capsule, and then inputs the vector of the activity information into the first capsule layer and converts the input activity information into activity information with spatial features by

$$U_k = squash(\sum_{j=0}^{m} Y_j \times W_{jk} + b_k) \tag{6}$$

where m represents the number of capsules, W_{jk} refers to the k-th weight parameter of the first capsule layer corresponding to Y_j, whose initial value is a random number generated by a truncated normal distribution, b_k refers to the offset parameter of the first capsule layer, whose default initial value is set to 0, and U_k is the output of the capsule layer, which is a vector and identifies the features of activity information and the spatial relationships among different features. The size of the output after the first capsule layer is $Batch\_Size \times \frac{Con\_layer - Nuclear\_Size_2 + 1}{L_2} \times m \times 1$.

Step 4: Input the activity information with spatial features into the second capsule layer. Then the activity information is processed through dynamic routing protocols by

$$S_j = \sum_i \frac{exp(b_{ij})}{\sum_k exp(b_{ik})} \times U_{j|i} \tag{7}$$

where b_{ik} refers to the dynamic routing weight of the i-th neuron in the first capsule layer and the k-th neuron in the second capsule layer, b_{ij} refers to the dynamic routing weight of the i-th neuron in the first capsule layer and the j-th neuron in the second capsule layer, $U_{j|i}$ refers to the output of each capsule in the layer below, and S_j refers to the feature of activity information processed by the second capsule layer through the dynamic routing protocols.

TABLE II
THE NUMBER AND PROPORTION OF EACH ACTIVITY IN WISDM

Type	Number	Proportion
Walking	424, 400	38.6%
Standing	48, 395	4.4%
Going upstairs	122, 869	11.2%
Going downstairs	100, 427	9.1%
Sitting	59, 939	5.5%
Jogging	342, 177	31.2%

The size of the output processed by the second capsule layer is $Batch\_Size \times Num\_Output \times Vec\_Lenv \times 1$.

Step 5: Convert the activity information from a vector to a scalar through the fully connected layer. The size of the output of the fully connected layer is $Batch\_Size \times Output\_Length \times 1$.

Step 6: Use the Softmax classifier to classify and recognize the activity information. By solving the activity probability through the classifier, the system finds the corresponding activity with the largest probability value, which is the final recognition result.

IV. EXPERIMENTS AND EVALUATIONS

The experiments are performed on a PC with Intel Core i7-8700 + 16G RAM + NVIDIA GeForce GTX 1080 Ti, and run on windows system with Pycharm + Anaconda (Python 3.6.5) + CUDA 9.0 + Cudnn + Tensorflow-gpu 1.8. The data collected by the system is consistent with the type of the public dataset WISDM, and in order to quantify the superiority of the proposed framework compared with the existing frameworks in HAR, the experiments are performed on the raw time series of the dataset WISDM. In the experiments, we analyze how some key parameters of the proposed framework affect the classification results, and confirm the optimal configuration of parameters. Finally, we show the performance of the proposed framework by comparing with other state-of-the-art frameworks. In addition, the performance of LoRa transmission scheme is also tested.

A. WISDM

The dataset WISDM is developed by the Wireless Sensor Data Mining Lab at Fordham University. It collects 36 individuals' activities through sensor BMA150 in a smartphone, including the following six activities: walking, standing, going upstairs, going downstairs, sitting and jogging. It has a total of 1, 098, 207 sampling points with a sampling rate of 20Hz. The number and proportion of each activity in the dataset WISDM is shown in Table II. In order to meet the length of time required for each activity, the coverage time of a sliding window is 4.5 seconds. The whole data are divided into 75% training data and 25% testing data.

B. Optimal Configuration of Parameters

In order to obtain the optimal configuration of parameters, i.e. the number of route iterations, the learning rate, the number of capsule's convolution kernels and the number of capsule's convolution nuclear groups, we adjust them in turn and get the best configuration of these parameters.

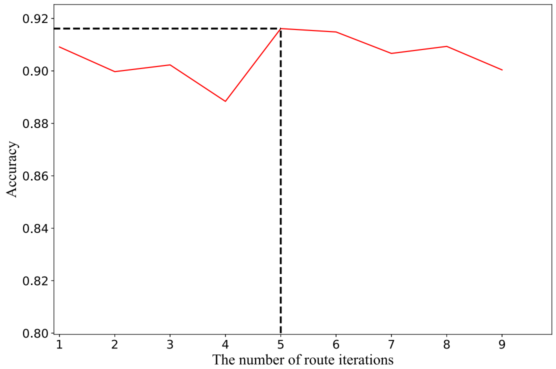


Fig. 7. The effect of the number of route iterations on the accuracy of activity recognition.

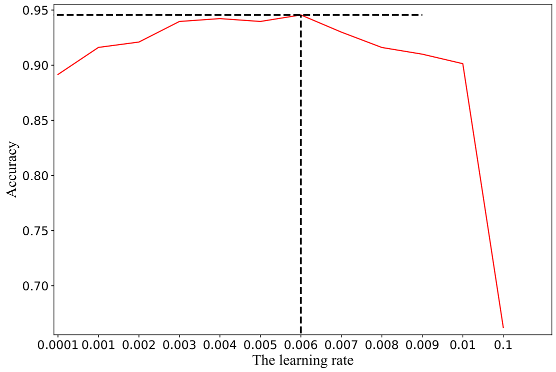


Fig. 8. The effect of the learning rate on the accuracy of activity recognition.

1) The Number of Route Iterations: Firstly, the effect of the number of route iterations on the accuracy of activity recognition is analyzed. Because the number of layers of the framework is small, the number of route iterations is set to be $[1, 9]$ and the step is 1. Fig. 7 shows the effect of the number of route iterations on the accuracy of activity recognition, and it can be seen that the performance of the proposed framework is not obviously affected when the number of route iterations increases. But as we know that the larger the number of route iterations is, the greater complexity the framework will have. After several simulations we find out that the accuracy rate of the framework is optimal (92.3%) when the number of route iterations is 5.

2) The Learning Rate: Fig. 8 shows the effect of the learning rate on the accuracy of activity recognition. The learning rate is firstly set to the different constants i.e., 0.0001, 0.001, 0.01, 0.1. The simulation results show that the optimal parameters may exist between 0.001 and 0.01. Then, 0.001 step is used to find the optimal parameter on $[0.002, 0.009]$. When the learning rate increases to a certain value, the accuracy rate of the proposed framework gets worse. The graph shows that the optimal accuracy rate (94.5%) occurs when the learning rate is 0.006.

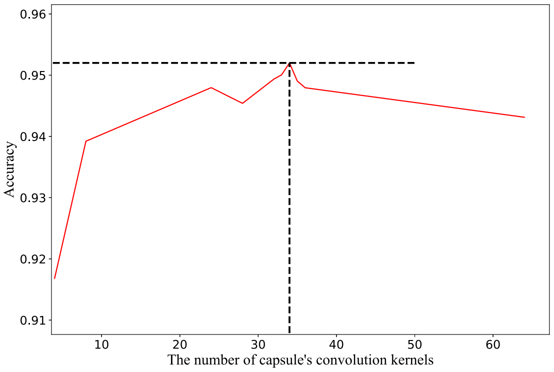


Fig. 9. The effect of the number of capsule's convolution kernels on the accuracy of activity recognition.

TABLE III
THE EFFECT OF THE NUMBER OF CAPSULE'S CONVOLUTION NUCLEAR GROUPS ON THE ACCURACY OF ACTIVITY RECOGNITION

Number	Test Accuracy
4	94.1%
6	94.5%
7	94.6%
8	94.4%
9	94.2%
10	94.8%
11	95.2%
12	94.7%
16	94.1%

3) The Number of Capsule's Convolution Kernels: On the basis of the optimal learning rate and the number of route iterations, we analyze the effect of the number of capsule's convolution kernels on the accuracy of activity recognition. The selection of the number of capsule's convolution kernels mainly adopts the adaptive adjustment method based on dichotomy. The results of tuning the number of capsule's convolution kernels are shown in Fig. 9. From the previous best result configurations, a value 34 improves performance on the test set by 0.6%, and the corresponding optimal accuracy is 95.1%.

4) The Number of Capsule's Convolution Nuclear Groups: Table III shows the effect of the number of capsule's convolution nuclear groups on the accuracy of activity recognition. In this paper, the numbers, i.e., 4, 8 and 16, are selected as the reference number of capsule's convolution nuclear groups, and then this parameter continues to be adjusted with the dichotomy until traversing integers between 4 and 16. As for the number of capsule's convolution nuclear groups, the framework is not very sensitive to this parameter. While the best accuracy rate (95.2%) was obtained for the number of capsule's convolution nuclear groups 11, the accuracy does not drop significantly for the other values.

C. Comparison With the State-of-the-Arts

Here we measure the performance of the proposed capsule framework using the "optimal" configuration of parameters.

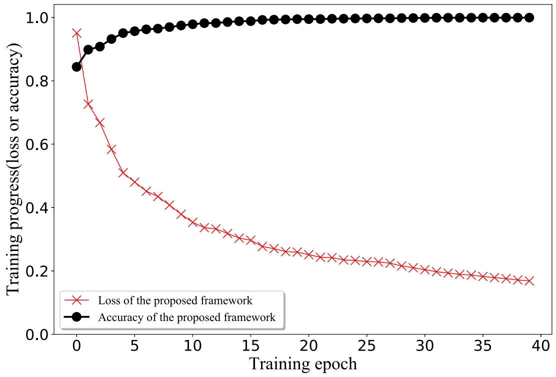


Fig. 10. The loss and accuracy of the proposed capsule framework.

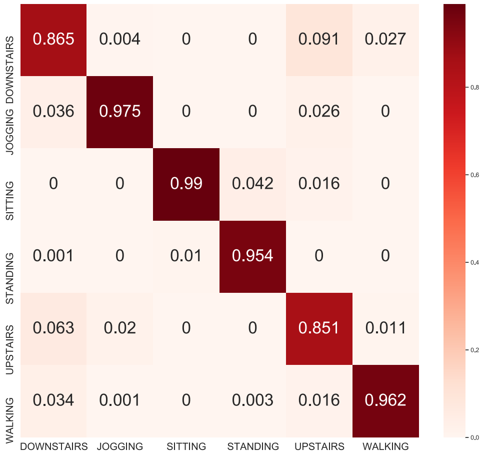


Fig. 11. The confusion matrix of the proposed capsule framework.

Fig. 10 shows the loss and accuracy of the proposed capsule framework. As we can see that the loss value of the proposed framework decreases and the accuracy rate increases continuously when the number of training epoch increases. The training accuracy of the framework reaches 99.9% when the number of training epoch reaches 30.

Fig. 11 shows the confusion matrix of the capsule framework, in which the row represents the activities predicted by the framework and the column represents the actual activities. The diagonal of the confusion matrix indicates that the prediction is the same as the actual one, i.e., the prediction is correct, and the non-diagonal element indicates the error of prediction. It can be seen that the recognition accuracy of jogging, walking, and standing is very high and reaches 97.5%, 99.0%, and 96.2%, respectively. Especially, like most other classifiers, it fails in distinguishing between very similar activities like going upstairs and going downstairs.

In terms of the performance of the capsule framework, we compare the proposed framework with other state-of-the-art frameworks in HAR, such as CNN [20] and LSTM [44]. As can be seen from Fig. 12, i.e., the loss and accuracy of the proposed capsule framework, and Table IV, i.e., the comparison of framework accuracy rate, the training accuracy rate of the capsule framework is always higher than CNN and RNN with increasing training epoch. The testing accuracy of the capsule network is 95.2%, which is 4.5% higher than CNN and 2.9% higher than LSTM. According to the above

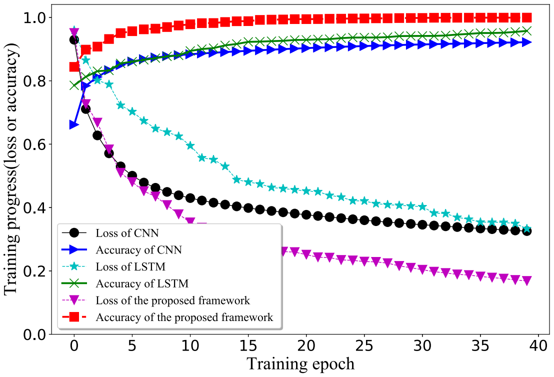


Fig. 12. The loss and accuracy of CNN, LSTM and the proposed framework.

TABLE IV
THE ACCURACY RATE COMPARISON OF DIFFERENT FRAMEWORKS

Framework	Training Accuracy	Testing Accuracy
CNN [20]	92.3%	90.7%
LSTM [44]	95.5%	92.3%
The Proposed Framework	**99.9**%	**95.2**%

results, it is obvious that the capsule framework has a greater improvement in the accuracy rate than other state-of-the-art frameworks in terms of performance.

D. The Performance of LoRa

The wristband is powered by a lithium-ion battery with a voltage of 3.7V. The frequency band of the LoRa module is 433MHz. The tested transmission distance is 200 meters and the transmission rate of collected sensor data is 4.8Kbps on an open field, which can meet the transmission requirement of the collected activity information. The performance may change depending on the different testing environments.

V. CONCLUSION

In this paper, a real-time human activity recognition system based on capsule and LoRa is proposed, which adopts LoRa technology to realize long-distance and low-power transmission of information, providing an effective and feasible method for the real-time monitoring scenario with a small population and a large range of activities, such as smart prison. At the same time, unlike the existing frameworks of machine learning and deep learning that only identify the features of activity information, the proposed capsule framework can identify the features of activity information and the spatial relationship among features. In the experiments, the optimal configuration of parameters is confirmed by analyzing the impact of some key parameters on the proposed framework. To demonstrate the strength of the proposed framework, our framework is compared with two state-of-the-art frameworks, i.e., CNN and LSTM, on the raw time series of the dataset WISDM. The experimental results indicate that the recognition accuracy of the proposed framework is 95.2%, which is 4.5% higher than CNN and 2.9% higher than LSTM. In addition, the LoRa based scheme can meet the transmission requirement of the proposed system. In summary, the proposed system has certain advantages in terms of recognition accuracy, practicability, and adaptability, in despite of some limitations stated above, such as the worse discrimination of some similar behaviors. Future works will consider the scheme with a combination of capsule and LSTM, and a different weighting mechanism as well.

ACKNOWLEDGMENT

The authors thank the editors and anonymous reviewers who gave valuable suggestions that have helped to improve the quality of the article. This article is based on the work that was firstly submitted on November 11, 2019.

REFERENCES

[1] F. Foerster, M. Smeja, and J. Fahrenberg, "Detection of posture and motion by accelerometry: A validation study in ambulatory monitoring," *Comput. Hum. Behav.*, vol. 15, no. 5, pp. 571–583, Sep. 1999.
[2] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognit. Lett.*, vol. 119, no. 1, pp. 3–11, Mar. 2019.
[3] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, 3rd Quart., 2013.
[4] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *Proc. IWVSPETS*, Beijing, China, 2005, pp. 65–72.
[5] A. Mehmood, S. A. Raza, A. Nadeem, and U. Saeed, "Study of multi-classification of advanced daily life activities on shimmer sensor dataset," *Int. J. Commun. Netw. Inf. Secur.*, vol. 8, no. 2, pp. 86–92, Aug. 2016.
[6] Z. Qin *et al.*, "Learning-aided user identification using smartphone sensors for smart homes," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7760–7772, Oct. 2019.
[7] Y. Wang, S. Cang, and H. Yu, "A data fusion-based hybrid sensory system for older people's daily activity and daily routine recognition," *IEEE Sensors J.*, vol. 18, no. 16, pp. 6874–6888, Aug. 2018.
[8] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE Sensors J.*, vol. 15, no. 3, pp. 1321–1330, Mar. 2015.
[9] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 98–104, Oct. 2017.
[10] J. P. Bardyn, T. Melly, O. Seller, and N. Sornin, "IoT: The era of LPWAN is starting now," in *Proc. ESSCIRC*, Lausanne, Switzerland, 2016, pp. 25–30.
[11] T. Sztyler, H. Stuckenschmidt, and W. Petrich, "Position-aware activity recognition with wearable devices," *Pervas. Mobile Comput.*, vol. 38, no. 2, pp. 281–295, Jul. 2017.
[12] T. V. Duong, H. H. Bui, D. Q. Phung, and S. Venkatesh, "Activity recognition and abnormality detection with the switching hidden semi-Markov model," in *Proc. CVPR*, San Diego, CA, USA, 2005, pp. 838–845.
[13] P. Casale, O. Pujol, and P. Radeva, "Human activity recognition from accelerometer data using a wearable device," in *Proc. IbPRIA*, Las Palmas, Spain, 2011, pp. 289–296.
[14] G. M. Sandstrom, N. Lathia, C. Mascolo, and P. J. Rentfrow, "Opportunities for smartphones in clinical care: The future of mobile mood monitoring," *J. Clin. Psychiatry*, vol. 77, no. 2, pp. 135–137, Feb. 2016.
[15] S. Sani, N. Wiratunga, S. Massie, and K. Cooper, "KNN sampling for personalised human activity recognition," in *Proc. ICCBR*, Trondheim, Norway, 2017, pp. 330–344.
[16] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proc. ICML*, Montreal, QC, Canada, 2009, pp. 609–616.
[17] J. Ye, G. Stevenson, and S. Dobson, "KCAR: A knowledge-driven approach for concurrent activity recognition," *Pervasive Mobile Comput.*, vol. 19, no. 12, pp. 47–70, May 2015.
[18] S. M. Lee, S. M. Yoon, and H. Cho, "Human activity recognition from accelerometer data using convolutional neural network," in *Proc. BigComp*, Jeju Island, South Korea, 2017, pp. 131–134.
[19] H. Li and M. Trocan, "Personal health indicators by deep learning of smart phone sensor data," in *Proc. CYBCONF*, Exeter, U.K., 2017, pp. 1–5.
[20] A. Ignatov, "Real-time human activity recognition from accelerometer data using convolutional neural networks," *Appl. Soft Comput.*, vol. 62, no. 15, pp. 915–922, Jan. 2018.

[21] S. Bhattacharya and N. D. Lane, "From smart to deep: Robust activity recognition on smartwatches using deep learning," in *Proc. PerCom Workshops*, Sydney, NSW, Australia, 2016, pp. 1–6.

[22] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," in *Proc. INTERSPEECH*, Singapore, Singapore, 2014, pp. 338–342.

[23] M. Sundermeyer, R. Schlüter, and H. Ney, "LSTM neural networks for language modeling," in *Proc. INTERSPEECH*, Portland, OR, USA, 2012, pp. 194–197.

[24] G. Hinton *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[25] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. NIPS*, Long Beach, CA, USA, 2017, pp. 3856–3866.

[26] M. A. T. Turan and E. Erzin, "Monitoring Infant's emotional cry in domestic environments using the capsule network architecture," in *Proc. INTERSPEECH*, Hyderabad, India, 2018, pp. 132–136.

[27] R. Chen, M. A. Jalal, L. Mihaylova, and R. K. Moore, "Learning capsules for vehicle logo recognition," in *Proc. FUSION*, Shanghai, China, 2018, pp. 565–572.

[28] C. Xu, D. Chai, J. He, X. Zhang, and S. Duan, "InnoHAR: A deep neural network for complex human activity recognition," *IEEE Access*, vol. 7, no. 10, pp. 9893–9902, Jan. 2019.

[29] J. Yin, Q. Yang, and J. J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 8, pp. 1082–1090, Aug. 2008.

[30] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, and D. Howard, "A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 871–879, Mar. 2009.

[31] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors*, vol. 10, no. 2, pp. 1154–1175, Feb. 2010.

[32] W. C. Cheng and D. M. Jhan, "Triaxial accelerometer-based fall detection method using a self-constructing Cascade-AdaBoost-SVM classifier," *IEEE J. Biomed. Health Informa.*, vol. 17, no. 2, pp. 411–419, Mar. 2013.

[33] K. Altun, B. Barshan, and O. Tunçel, "Comparative study on classifying human activities with miniature inertial and magnetic sensors," *Pattern Recognit.*, vol. 43, no. 10, pp. 3605–3620, Oct. 2010.

[34] I. Cleland *et al.*, "Optimal placement of accelerometers for the detection of everyday activities," *Sensors*, vol. 13, no. 7, pp. 9183–9200, Jul. 2013.

[35] R. R. Salakhutdinov and G. E. Hinton, "Using deep belief nets to learn covariance kernels for Gaussian processes," in *Proc. NIPS*, Vancouver, BC, Canada, 2007, pp. 1249–1256.

[36] Y. Chen and Y. Xue, "A deep learning approach to human activity recognition based on single accelerometer," in *Proc. ICSMC*, Kowloon, China, 2015, pp. 1488–1492.

[37] J. Yang, M. N. Nguyen, P. P. San, X. X. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *Proc. IJCAI*, Buenos Aires, Argentina, 2015, pp. 3995–4001.

[38] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach," in *Proc. IEEE Sensors*, Orlando, FL, USA, 2016, pp. 1–3.

[39] M. Zeng *et al.*, "Convolutional neural networks for human activity recognition using mobile sensors," in *Proc. MobiCASE*, Shanghai, China, 2014, pp. 197–205.

[40] S. Ha and S. Choi, "Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors," in *Proc. IJCNN*, Vancouver, BC, Canada, 2016, pp. 381–388.

[41] C. A. Ronao and S. B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, no. 23, pp. 235–244, Oct. 2016.

[42] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Dec. 1997.

[43] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.

[44] F. J. Ordonez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, pp. 115–126, Jan. 2016.

[45] Y. Chen, K. Zhong, J. Zhang, Q. Sun, and X. Zhao, "LSTM networks for mobile human activity recognition," in *Proc. ICAITA*, Dubai, Arabia, 2016, pp. 1035–1042.

[46] B. Padmaja, V. V. R. Prasad, and K. V. N. Sunitha, "Deep RNN based human activity recognition using LSTM architecture on smartphone sensor data," *J. Fundam. Appl.*, vol. 10, no. 5, pp. 1102–1115, Feb. 2018.

[47] E. Xi, S. Bing, and Y. Jin, "Capsule network performance on complex data," Dec. 2017, *arXiv:1712.03480*. [Online]. Available: http://arxiv.org/abs/1712.03480

Leixin Shi received the B.S. degree in the Internet of Things engineering from Mongolia Industrial University, Hohhot, China, in 2017. He is currently pursuing the M.S. degree with the School of Information Science and Engineering, Shandong University, Qingdao, China. His research interests include machine learning, artificial intelligence, activity recognition, ubiquitous computing, data fusion, and quality of context.

Hongji Xu (Member, IEEE) received the B.S. degree in electronic engineering from the Shandong University of Technology, Jinan, China, in 1999, and the M.S. degree in signal and information processing and the Ph.D. degree in communication and information system from Shandong University, Jinan, in 2001 and 2005, respectively. From 2004 to 2005, he was a Visiting Ph.D. candidate with the Telecommunications Technological Center of Catalonia (CTTC) and the Department of Signal Theory and Communication, Polytechnic University of Catalonia (UPC), Barcelona, Spain, and did research in the areas of wireless communication and signal processing. From 2010 to 2015, he was a Postdoctoral Researcher with Tsinghua University—Inspur Group Postdoctoral Scientific Research Station, China, and focused on the research in multimedia information processing for smart home and cloud computing. From December 2014 to December 2015, he was a Visiting Scholar with the Department of Cognitive Science, University of California at San Diego (UCSD), USA, and did research in the ubiquitous computing and human–computer interaction. From January 2018 to April 2018, he was a Visiting Scholar with the Virginia Polytechnic Institute and State University (Virginia Tech), USA, and did research in the interdisciplinary fields related to information science and computer science. He is currently an Associate Professor with the School of Information Science and Engineering, Shandong University. His research interests include wireless communications, ubiquitous computing, blind signal processing, human–computer interaction, and artificial intelligence.

Wei Ji was born in 1978. He received the M.S. degree from the Beijing Institute of Technology in 2003 and the Ph.D. degree from the Beijing University of Posts and Telecommunications in 2006. From September 2012 to September 2013, he worked as a Visiting Scholar with Queen's University, Canada. He is working as a Professor with Shandong University. His research interests include artificial intelligence, data center networking, high-speed optical switching and networking, and optical access networking.

Beibei Zhang received the B.S. degree in communication engineering from Shandong University, Jinan, China, in 2017. She is currently pursuing the M.S. degree with the School of Information Science and Engineering, Shandong University, Qingdao, China. Her research interests include human activity recognition, data fusion, and ubiquitous computing.

Xiaojie Sun received the B.S. degree in communication engineering from Shandong Normal University, Jinan, China, in 2019. She is currently pursuing the M.S. degree with the School of Information Science and Engineering, Shandong University, Qingdao, China. Her research interests include machine learning, artificial intelligence, activity recognition, and context-aware computing.

Juan Li received the B.S. degree in electronic information engineering from Shandong University, Jinan, China, in 2018. She is currently pursuing the M.S. degree with the School of Information Science and Engineering, Shandong University, Qingdao, China. Her research interests include human activity recognition, data fusion, and ubiquitous computing.