Activity Recognition and Classification via Deep Neural Networks

Zhi Wang[1], Liangliang Lin[2,3], Ruimeng Wang[4], Boyang Wei[5], Yueshen Xu[6], Zhiping Jiang[6], and Rui Li[6(✉)]

[1] School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China
zhiwang@xjtu.edu.cn

[2] School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China
lin_ll@126.com

[3] Informatization Office, Xi'an Conservatory of Music, Xi'an 710061, China

[4] School of Photovoltaic and Renewable Energy Engineering, The University of New South Wales, Sydney, NSW 2052, Australia
brianrwangmsecs@gmail.com

[5] Geogetown University, Washington DC 20007, USA
bw558@georgetown.edu

[6] School of Computer Science and Technology, Xidian University, Xi'an, China
rli@xidian.edu.cn

Abstract. Based on the Wi-Fi widely separated in the world, Wi-Fi-based wireless activity recognition has attracted more and more research efforts. Now, device-based activity awareness is being used for commercial purpose as the most important solution. Such devices based on various acceleration sensors and direction sensor are very mature at present. With more and more profound understanding of wireless signals, commercial wireless routers are used to obtain signal information of the physical layer: channel state information (CSI) more granular than the RSSI signal information provides a theoretical basis for wireless signal perception. Through research on activity recognition techniques based on CSI of wireless signal and deep learning, the authors proposed a system for learning classification using deep learning, mainly including a data preprocessing stage, an activity detection stage, a learning stage and a classification stage. During the activity detection model stage, a correlation-based model was used to detect the time of the activity occurrence and the activity time interval, thus solving the problem that the waveform changes due to variable environment at stable time. During the activity recognition stage, the network was studied by innovative deep learning to conduct training for activity learning. By replacing the fingerprint way, which is used broadly today, with learning the CSI signal information of activities, we classified the activities through trained network.

Keywords: Channel state information · Pearson correlation coefficient · Deep convolutional neural networks · AlexNet network

H. Gao et al. (Eds.): TridentCom 2019, LNICST 309, pp. 213–228, 2020.
https://doi.org/10.1007/978-3-030-43215-7_15

1 Introduction

The rapid development and wide application of the Internet of Things (IoT) around the world stimulate the study on wireless activity awareness. At present, the two main study directions in the field of wireless activity awareness are device-based activity awareness systems and device-free activity awareness systems. The device-based activity awareness systems collect data mainly through built-in sensors of mobile phones, smart bracelets, smart watches or other sensing devices to obtain the activity status of people or other targets. And the device-free activity awareness systems obtain people's activity status mainly through the impact of people's activities on wireless networks. However, the device-based systems are limited by many conditions, for example, the senior and the infant are typically loath to carry a device, and those devices are not suitable for bringing into private occasions. Therefore, the device-free wireless-based activity awareness systems are needed in more and more scenarios and have ever-increasing research value. At the same time, recognizing activities with wireless networks also faces a lot of challenges.

As the number of mobile smart device increases rapidly around the globe, and the wireless network devices are popularized and widely used in people's life, the methods for activity awareness are diversified (including RFID [9], Wi-Fi [1] and radar), and the characteristic that the wireless signal can spread over the space without any impact on persons lays a solid foundation for the device-free activity awareness system.

At present, many institutions and companies at home and abroad have conducted researches in the field of activity awareness and achieved diversified solutions. The WiTrack [8] system, published by the Massachusetts Institute of Technology at the NSDI International Conference in 2014, performs activity awareness for targets by applying wireless technologies of frequency modulated continuous wave and radar reflection in the Wi-Fi field. The WiTrack 2.0 [15] system, presented at the NSDI International Conference in 2015, allows more than one person to present in the scenario to conduct location awareness. Similarly, WiSee [2] system and WiHear [14] use wireless technologies to perform activity awareness on human body in whole or in part. Currently, in the field of activity awareness, limited by technologies and scenarios, most of the device-free activity awareness systems are still being developed, with strong demands in commercial application.

Despite relatively sluggish progress on commercial application development, the activity awareness system has broad prospect and will become a huge innovative point in commercial application. The wireless activity awareness system can be widely applied in many scenarios, such as a nursing home, where the senior can be protected from dangers by activity monitoring, and safety precaution can be made when the senior suffer from sudden illness. In the private spaces such as washroom and bedroom, the wireless-based activity awareness system can not only monitor the target activity in real time, but also detect and alarm abnormities. From this, as an emerging product, the wireless activity awareness system has a promising commercial prospect thanks to its low hardware cost, high accuracy, convenient installation and other advantages.

The research in this paper contributes to the following aspects: (1) using extensive commercial routers to differentiate more detailed activities; and (2) using deep learning methods to perform activity awareness and recognition with good fault tolerance and stability. Meanwhile, many challenges still remain in the research. The work direction will be: (1) supervised learning, as the people's activities need to be learned thoroughly; and (2) the system is unable to recognize and distinguish activities of several persons at once.

2 Related Work

There are many researches in the field of activity awareness [1–10]. Related researches mainly include video-based, mobile device-based, radar-based and wireless fingerprint-based activity awareness systems.

The products of video-based activity awareness system are relatively mature. Key of this system lies in using video images to recognize and identify the target [11–13]. The user's behaviors are obtained by analyzing activities of human bodies. Among these systems, Kinect [4] and Leap Motion [5] are widely used in commercial application. Kinect collects human images mainly through video and infrared, and it simulates people's activities by constructing collected images into a 32-node model. The moving activities in the depth direction can not be detected by plane images. The video-based activity recognition uses a depth camera to detect the activities in the depth direction other than the plane motion. To recognize the detailed local motion, a camera and a depth camera are used to detect the edge of human body and construct a denser dot matrix into a network model to obtain the local motion activity. Leap Motion uses facial activity recognition based on binocular vision. It simulates binocular vision effects via dual cameras to capture two pictures simultaneously to realize 3D modeling and sense the 3D activities out of 2D planes. Applying different technologies, the above two widely-used commercial products ultimately aim at recognizing activities via video. More importantly, they overcome the key problems in detection of 3D activities through image recognition under the relatively mature conditions. Application of these two products, however, is constrained by strict environmental requirements, such as sufficient light and no obstruction.

The mobile device-based activity awareness system mainly relies on the sensor module built into the mobile device [16–25]. Especially during the current period when the mobile phone is commonly used, this system typically detects the people's activities by the compass, gyroscope, the acceleration sensor and the sensors in the telephone which cooperate with each other. With the direction indicated by the compass, the device status in 3D space indicated by the gyroscope, the velocity indicated by the acceleration sensor, and detailed displacement distance calculated by time, this system can obtain the user's specific motions, but the activity details cannot be detected. This is the most widely-used method at this stage, with inability to directly detect fine activities as its chief drawback. Individual's ongoing activities can only be inferred by these built-in sensors of the mobile phones. For example, the system can detect that a person is exercising, climbing the stairs or sleeping, but it is unable to sense the specific activities such as standing up, sitting down or raising hands. In addition to mobile phones, today's most popular activity awareness devices are smart bracelets which are designed for monitoring. The bracelets are equipped with various sensors, such as the heartbeat sensor to detect the heart rate and the temperature sensor to detect the body temperature. Representative products include Apple Watch, Huawei Sports Watch and Xiaomi Sports Bracelet. These wearable devices are sufficient as a simple activity awareness device, but they are unacceptable as a part that must be carried by human body, especially for children, the new born and the senior, because it is undesirable for theses persons, both physically and mentally, to wear such a special device.

The radar-based activity awareness system senses the surrounding environment through high-frequency radar signals. When the target is active in the area, it causes reflection of radar signals. Through the collection of high-frequency signal, people's motion at every moment can be obtained. By combining these motions together, a continuous activity is formed and thus the activity of the target is obtained. As a sophisticated technical means, radar is highly sensitive to people's activities. But it is very expensive with strict application requirements. What's more, the high-precision radar system is gigantic and difficult to deploy, restricting its application only to the military field. Despite its high precision, its huge size and high cost make this system not suitable for the current commercial environment.

The technical solution we studied is mainly based on 2.4 GHz wireless RF signals, similar to detecting target activity with RF fingerprints. Earlier, researches of wireless-based activity awareness were mainly through wireless RSSI information. For example, the SigComm conference published an article about using RSSI to detect people's heartbeat [9] in 2011, but the RSSI had rigorous requirements on environment and the content of information was insufficient, so it was difficult to get more valuable information from it. There are many methods for activity recognition, such as application of sensor-based activity recognition systems [1, 2], but the sensors either require to be carried by the target, or deployed around the activity area of the target. The systems applying acceleration sensors to detect the motion speed of human body are easily disturbed, for example, the GrandCare will be disturbed by the motion of doors. There are other wearable device systems that use BodyScop as a sound sensor to classify people's activities to identify eating, coughing and other people's activities.

3 Wireless Activity Recognition System

To design a relatively universal system and enable the system to recognize the activities of the target in a low-power Wi-Fi environment, we proposed a better method to record changes of Wi-Fi signals. Besides, we carried out training based on collection of a large amount of data for the purpose of being suitable for more extensive scenarios.

3.1 Architectural Design of Wireless Activity Recognition System

This system uses wireless signals to recognize the people's activities. Based on existing wireless router and CSI signal monitoring means, there are many insuperable difficulties for the system to rely only on the channel of the Wi-Fi signal to monitor the people's activities. First of all, when recognizing people's activities, data needs to be learned, thus requiring a large amount of stable data. When collecting data, factors including environment and equipment must be changed singly, so that we can determine that the signal is affected by people's activities rather than change of environment. In addition, the stability of the data must be ensured during data filtering. Second, it is critical to determine when the motion changes. After receiving the data of the target motion, it is necessary to recognize the time at which the motion occurs and the time at which the motion ends. For example, within a video surveillance area, when a quiescent person suddenly stands up from different positions at different speeds, the start and end time of the motion and the type of motion are required to be accurately recognized by the algorithm. The environment has changed due to people's motion, regardless of whether they become quiescent again (Fig. 1).

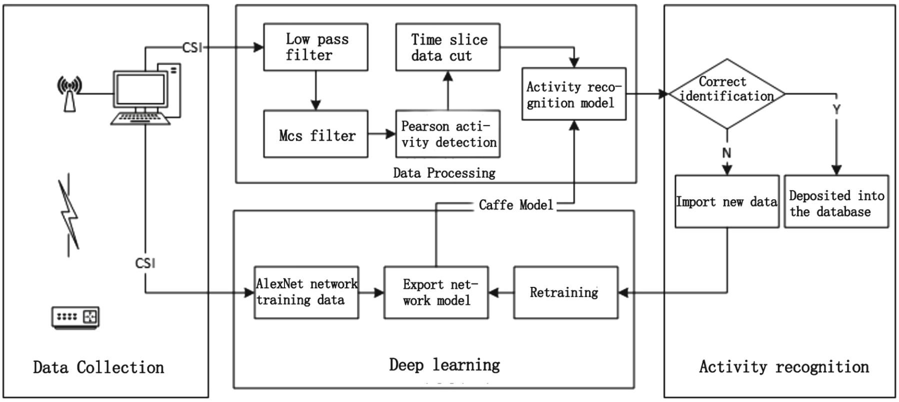

Fig. 1. System architecture diagram

We mainly studied two aspects of the wireless activity recognition system. On the one hand, as home wireless routers become more and more popular, watching TV, playing computer games, surfing the internet and other daily applications can be supported by the Wi-Fi signals provided by devices. With more stable Wi-Fi signal and enhanced network bandwidth, finer granularity of data can be achieved through a better way under 802.11n MIMO system. With the standard 20 MHz signal and 40 MHz signal, the 802.11 system provides 52 and 128 OFDM subcarriers respectively. In this system, we are able to collect 30 subcarriers, and the CSI impact on people's activities is shown in Fig. 2. On the other hand, the far-reaching application of the system is to recognize and differentiate target groups through activity recognition, that is, distinguishing targets through recognition of target activities, a series of meaningful activities and statistics. By tracking one's activities, a time series-based portrait of the target person is created, through which the characteristics of the target groups in the current scenario are identified.

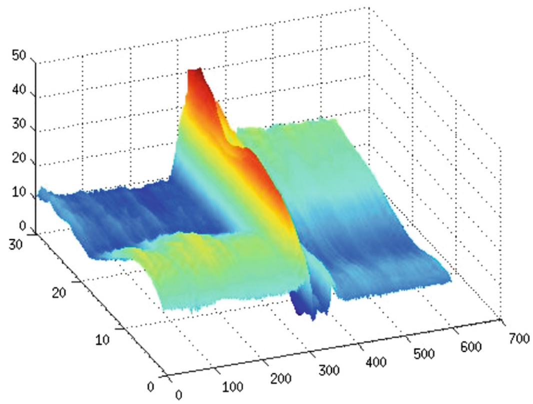

Fig. 2. Chart of CSI fluctuation caused by people's activities

We collected data by using a commercial wireless router based on the 802.11n MIMO system at a frequency of 2.3 GHz to 2.4 GHz. At this frequency, although it was not easy to obtain subtle motions, the impact of major motion of human body on wireless signals was very obvious. Besides, the difficulty in feature extraction was reduced through deep learning, thus compensating the poor recognition accuracy. The basic conception of the system is to recognize the human body motion based on signal characteristics and sequences by deeply learning the characteristics of the CSI link signal. We collected data of 30 channels through the network card, and after data preprocessing, we deeply learned the convolutional network. By modifying the parameters, the data was trained for several times to obtain an activity recognition model, through which the activity can be accurately recognized.

In the actual operation of the system, the following problems need to be solved. First one is data collection. Because the wireless signals, which can be received, are invisible, and can only be described by human imagination. In scientific research, the most common method for data collection is control variable method, by which other factors are controlled as much as possible during data collection to make the wireless signal fluctuate only by people's activities. To this end, we strictly controlled the environmental change during data collection. At the same time, the sampling rate is also essential for data collection. The sampling rate concerns the accuracy of the system's recognition of people's activity. Of course, the higher sampling rate causes a more complicated calculation, so the sampling rate shall be determined in line with the time required by the motion. Therefore, we determined a proper sampling rate by investigating the time of each motion made by target groups. The core of the system lies in learning the CSI through deep learning method, which has more advantages for target recognition, including the capability to judge a more detailed behaviors of the users and distinguish different target users. The core part of the system is activity recognition, which includes the learning and recognition of the activity. Activity recognition is divided into two parts: one is the recognition of the activity start & end time of the system to distinguish two different motions. Generally speaking, when the activity occurs, the CSI link will change drastically. And the end time of the activity is usually not easy to determine.

3.2 Design of Deep Learning Network Model

The model for activity recognition in the system performs image learning recognition by converting data into grayscale images. Therefore, the network model AlexNet is a more advantageous network structure model, which is the classic model of CNN in image classification.

The network model is shown in Fig. 3. You can see that this network structure uses two GPU servers (of course one GPU is also acceptable), so there are two processes. From the perspective of network structure, the entire AlexNet network model has a total of 8 layers, of which 5 layers are located in the convolutional layer, and the remaining 3 layers are fully connected layers. Each convolutional layer includes the excitation function and normalized function as partial response, and then pooling processing is conducted to simplify the calculation.

The input images are 224 × 224 × 3 pixel pictures, and change to 227 pixels after preprocessing to facilitate the convolution kernel selection and calculation of the network, and then the pooling layer is used to simplify the calculation. The network uses a 5 × 5 convolution kernel for feature extraction. The 48 in the figure means that each GPU uses 48 convolution kernels because two GPU are used for processing. In addition, although the size of the convolution kernel is 5 × 5, since the general training image uses an RGB three-channel image, the corresponding convolution kernel also has three layers in the third dimension.

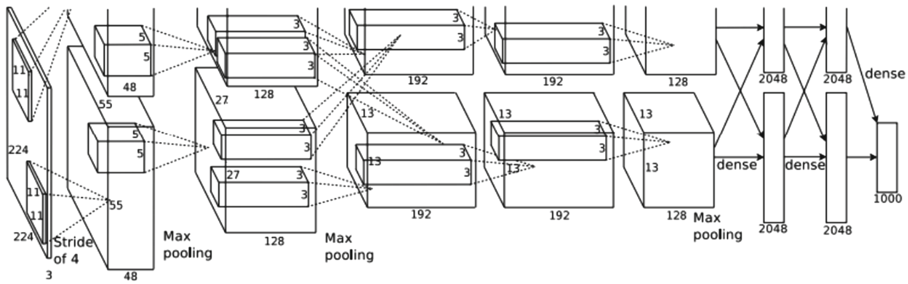

Fig. 3. AlexNet network structure

3.3 Design of Activity Recognition Scheme

In CSI, the number of subcarriers available in the system was 30, so the data used was a matrix of n × 30. The feature for activity recognition was 30 subcarriers.

$$h = s_0 + s_1 \times x_1 + s_2 \times x_2 \cdots + s_{30} \times x_{30}$$

where x_i is a vector representing the ith data. Appropriate parameters can be obtained by the gradient descent method.

Within the context of CSI, the impact of motions performed by targets within the region on CSI is a relatively stable distribution of CSI, and the impact can represent the activity and the activity time of this target. At the same location, CSI waveform distributions caused by different activities vary. CSI waveforms caused by the repetitive motion of the target and the different postures of the user are also different. First, the CSI waveforms of the same motion are compared to see if the CSI fluctuations caused by the same motion are related or similar. Then the CSI fluctuations caused by different motions are compared to observe whether it is possible to distinguish the two motions. Through real experimental comparison, it can be observed that the waveform similarity between the same motions is much higher than that between different motions, so it is theoretically possible to complete activity recognition by wireless means.

For known activities, only the known raw data are needed during data collection. After filtering and noise reduction, the occurrence time of the activity is obtained through an activity detection model, and then the current target activity is obtained through two schemes, one of which only compares the CSI fingerprint information from fingerprint database by a proper algorithm, and obtains the target motion from the nearest CSI waveform distribution. Comparing the waveform distribution or the ratio of wave peaks to waveforms are favorable fingerprint match algorithms. The second is to conduct supervised learning in advance by learning algorithm. When training known activity data by learning algorithms, namely, using data as training samples, the samples of the activities themselves are very similar during training and learning. The valid recognition can even be realized after data conversion by the traditional comparison method using fingerprint matching, so it can be very easy to recognize the known activities after using the CNN. During the learning of the model, complete matching will not be conducted to avoid overfitting, which is problematic for a highly accurate recognition with tiny error. However, it is very suitable here, because the impacts of the same activity on CSI are different, including the error of the hardware and effect of noise in the scenario. What more, when different persons do the same motion, the local details of their bodies vary obviously. This is conducive to the recognition accuracy of the learning algorithm, which can further recognize the very detailed local features to obtain the target within the current region.

The recognition of unknown target activities is the research priority of this system. For the fingerprint matching system, the fingerprint database stores all the active fingerprints. Although there are many targets and different active fingerprints, it is impossible to include fingerprints of all persons. In this case, the signal fingerprints formed by the same activity of different targets established by the fingerprint matching system are eventually similar despite distinction. A problem raises therefore. Although the fingerprints on images of different targets may be similar, errors are still remaining. Through comparison of fingerprint database, the risk of error will be relatively high. Such risks may reduce as the fingerprint library increases, but as the types of activity increase, similar activities have increasingly higher requirements for fingerprint match algorithm. The approach of deep learning that we proposed also focuses on this point. Deep convolutional neural networks can extract local features well by convolution operations of convolution kernels.

4 Implementation of Wireless Activity Recognition System

The wireless activity recognition system first collects data, then preprocesses the original data collected to obtain smoother data before learning the activity data of targets, including the same activity of different targets and different activities of the same target, and finally improves the accuracy of the system with higher frequency.

4.1 Data Collection

The initially received CSI data are $30 \times N \times M$ and each data is plural, N is the number of antennae and M is the number of data packets received. The scenario for data collection represents an important factor, and the representations of CSI signals are different in different scenarios. To recognize activities of targets in different scenarios, a large amount of data need to be collected in these scenarios to ensure the learning model used can better improve its accuracy. When collecting data, stable scenario shall be ensured, and data of different persons in the same scenario and data of the same person in different scenarios shall be collected to ensure the comparability of the data.

4.2 Data Preprocessing

Given the instability of hardware itself and the impact of ambient noise contained in the original CSI data collected, the original CSI data need to be preprocessed. Observation shows that wireless signal is vulnerable to the environment, for example, changes in location of items in the environment and sway of clothes and other items affected by wind will cause a great deal of noise to the original CSI data, resulting in the inability to obtain more effective information. For data training, dimension conversion of data is the main task. After receiving the original CSI data, the wireless signal data are converted and the signal information of 30 subcarriers is extracted. For a sequence, the whole data are the data set of 2D data $30 \times n$, and when training data, the data set needs to be converted into a data set of 3D data to facilitate the data training. The training sample and Label file are input to perform training. Data preprocessing is designed to improve the reliability of CSI data, and the data representing noise in the original CSI data are removed to allow CSI data to better reflect the activity of the target. The noise mostly comes from the Wi-Fi device transmitting in the indoor environment.

(1) Low-pass filtering
Low-pass filtering is a method to filter data, and its principle features that low frequency signals are allowed to pass while high frequency signals reaching the threshold will be blocked or weakened, and the amplitudes of blocking and weakening varies according to frequency and filtering process. High-pass filtering is a filter relative to low-pass filtering. The calculation formula of its critical frequency is as follows.

$$f_c = \frac{1}{2R_2C}$$

The low-pass filtering is used to remove the high frequency noise not caused by people's activities. Relatively, people's activities generally affect low frequency, so high frequency ambient noise needs to be removed. We used an exponential smoothing filter (DESF) which can smooth data changes of samples by exponential smoothing based on changes in original samples, and can remove high frequency noise and protect CSI data changes caused by people's activities from being removed.

(2) Filtering of modulation and coding index
CSI data are not only influenced by people's activities and ambient noise, but also modulation and coding index. We found in our experiment that modulation and coding index may also cause fluctuation of CSI signal and change fluctuation of CSI data due to signal fluctuation resulting from inherent unstable wireless signals, so CSI needs to be filtered when minimum impact was imposed on CSI by people's activities. When this part of impact not caused by human factors is removed, more accurate data will be obtained. The activity behavior data are only caused by people's activities, so it is impossible to remove all noise, but greatest efforts can be made to highlight the CSI signal data affected by people's activities.

(3) Trigger of detection activities
In this system, activities occur and stop randomly, so it is an ongoing monitoring process. It keeps collecting data during operation and people's activity itself is a random event, so the time when the activity is going to happen or stop and when the next activity is to happen shall be monitored in order to recognize people's activities via wireless signals. By observing the fluctuation of CSI signals, we can see that every time one activity happens, the wireless signal produces a violent fluctuation, so we can know the start and end time of people's activities by only detecting the time when the wireless signal begins to fluctuate and when the fluctuation turns to be relatively quiet (not totally still). There are two ways for detecting fluctuation, and one is threshold method. This method is very simple: an initial threshold is set and during receiving CSI signals, activities are considered to happen when one or more signals exceeding the threshold are detected, while activities are considered to have stopped when stable one or more signals below the threshold are detected.

The Pearson product-moment correlation coefficient is mainly used to measure the correlation between two variables, and the correlation coefficient between two variables is defined as the quotient of covariance and standard deviation between two variables. The calculation formula is as follows.

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

The formula above defines the overall correlation of X and Y, and covariance and standard deviation of the samples is represented by r, and the calculation formula is as follows.

$$r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$$

The Pearson correlation coefficient does very well at the beginning. After data preprocessing, we got the image shown in Fig. 4 by detection of Pearson correlation coefficient, and we can see from the image that Pearson correlation coefficient does a very good job in detecting the time when the fluctuation starts and ends. However, things are not always the same. Pearson correlation coefficient is outstanding when sound data are available and the time when the activity happens and stops can be easily judged. However, data are not always that sound, and environmental changes and fluctuation of hardware itself or minor irregular movement caused by wind blowing curtains and other items can also cause undesired data. As shown in Fig. 5, the judgment of the time when the activity starts and ends is very important for activity recognition, and when such case of Fig. 5 happens, it is difficult to know the occurrence time of the data. Although Pearson correlation coefficient serves well in judging the occurrence of activities when sound data are available, some errors may happen due to the instability of the signal itself and presence of noise. To reduce such misjudgment caused by the instability of the signal itself, we modified the data content of the algorithm, namely the sum of three data packets were used to judge the fluctuation, as shown in Fig. 6.

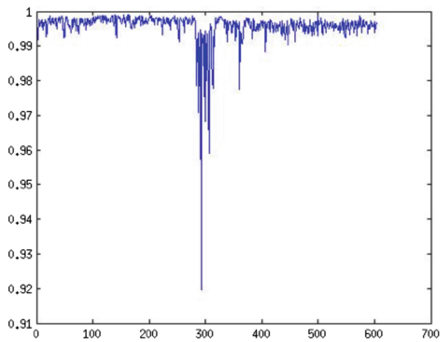

Fig. 4. Pearson correlation coefficient

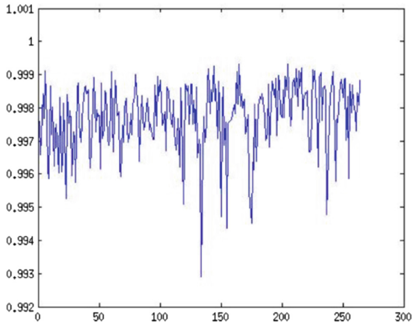

Fig. 5. CSI under Pearson correlation coefficient

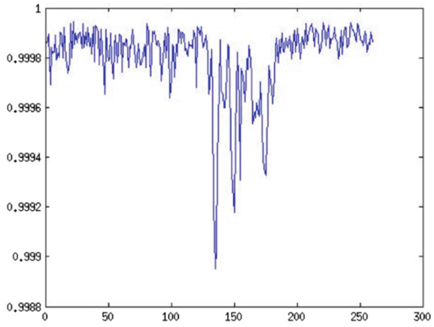

Fig. 6. Performance of Pearson coefficient

5 Performance Evaluation

Since we used a deep learning-based convolutional neural network to recognize activities and the accuracy of the time when activities happen is the key indicator of the whole system, it is necessary to evaluate the performance of activity detection of the system.

(1) Performance of Pearson correlation coefficient model
As the time in the activity detection model was obtained by comparing the lowest coefficient points within the most recent time window, the model had nothing to do with the environment, and the overall waveform of CSI would not be impacted by environmental changes, To get more detailed performance evaluation, we chose two test sites to conduct the test and obtained the performance evaluation analysis of the activity detection model. We carried out two groups of experiments in each of the two sites, and collected adequate data and established an activity detection model for each group of experiments. The true activities and the time when the activities happened were compared to obtain the accuracy rate of the activity detection models and the time error of detection activities. For detection activity trigger, accuracy rate of 95% was acceptable, and the time error of its activity detection was about 0.1 s, as shown in Fig. 7.

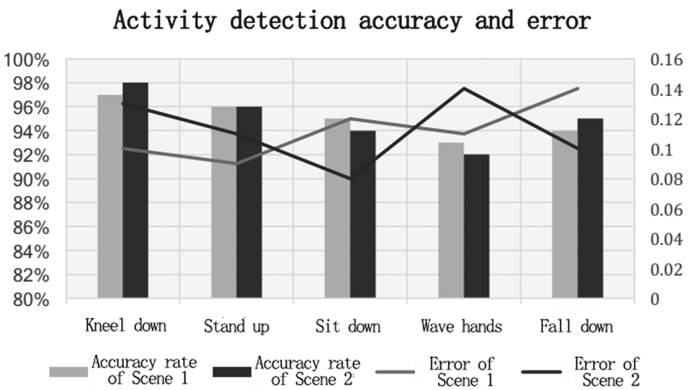

Fig. 7. Accuracy rate and time error of different activity detections

During the model detection, the system was randomly tested when the test target squatted down, stood up, sat down and waved hands, and the time when activities happened was detected by testing the models. After filtering the data collected, the detection model calculated the time error of activity occurrence, which was compared with the actual time when environmental activities happened. Each activity was sampled for 200 times with 1000 experiments conducted in total. It can be seen that the fluctuation time point was detected in a relatively accurate way. The figure above shows that the average error of this detection model is about 0.1 s.

(2) Performance of activity recognition model
We analyzed five deep learning network models including AlexNet and VGG, and conducted a multi-dimensional and detailed comparison on their similarities and their own advantages and disadvantages. Figure 8 shows the training comparison between the two deep convolutional network models namely AlexNet and VGG. We found through comparison that under the same data situation, AlexNet can achieve relatively high accuracy when much less learning data were needed in the network learning while VGG network showed high depth, so when less single data are available, it is very like to cause overfitting. That is to say, it works very well when learning, but when it comes to actual classification, the accuracy rate may fall.

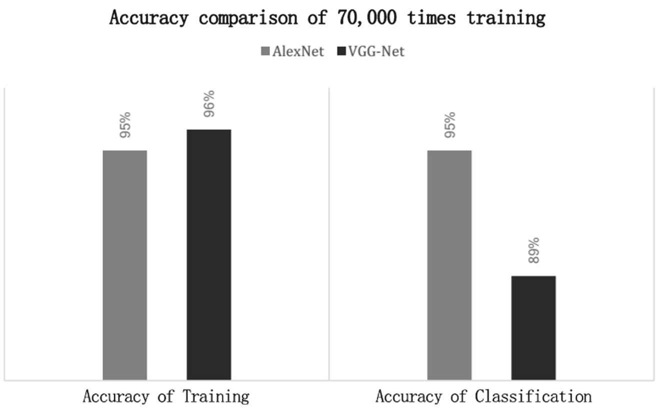

Fig. 8. Comparison of accuracy between AlexNet and VGG

After the system was deployed in Scenario 1, we respectively sampled data samples with 50 actions as a group by systematic data collection and training, and each sampling lasted 10 s and data transmission rate was 50 kpt/s. If the classification was right, the results were considered to be right, otherwise the results were considered to be wrong regardless of any other classification. Experiments showed that among the total 250 matching calculations, the accuracy rate was

98%, which mainly resulted from unstable hardware. Expensive device failed to conform to the design purpose of the experimental system, and the inexpensive commercial routers with an accuracy rate of 92% was acceptable. In Scenario 2, we conducted the same experiment with a consequent accuracy rate of 95% and a false positive rate of 3%, which were mainly due to the unstable fluctuation of the hardware itself. Compared to other systems like WiSee [1], an accuracy of 94% was obtained by very expensive device (USRP) (Fig. 9).

Fig. 9. Comparison of accuracy among WiSensor, WiSee and WiFall

6 Conclusion

By comparing and analyzing the existing activity recognition technologies, we summarized and analyzed the advantages and disadvantages of the current scheme and proposed to carry out activity recognition by training wireless CSI signals by means of the deep learning methods, and designed an activity recognition system. The comparison of various schemes in terms of their advantages and disadvantages showed that wireless activity recognition had better suitability and stability. And we chosen a network model most suitable for the current system by comparing various deep learning network models.

The authors used the means of converting wireless signals into pictures in an innovative manner. The activity detection model was used to obtain the time slice of activity occurrence which was clipped before being converting into pictures, and then the pictures were classified by means of the deep learning methods, thus the results and the classification of activities were obtained simultaneously. This means compensated for the inconsistency of waveforms of wireless signals at different time points. All waveforms, high or low, can be removed during training as backgrounds of pictures. Such scheme largely resolves the inconsistency of waveforms when signals are stable.

References

1. Wang, Y., Liu, J., Chen, Y., et al.: E-eyes: device-free location-oriented activity recognition using fine-grained WiFi signatures. In: Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, pp. 617–628. ACM (2014)
2. Pu, Q., Gupta, S., Gollakota, S., et al.: Whole-home gesture recognition using wireless signals. In: Proceedings of the 19th Annual International Conference on Mobile Computing & Networking, pp. 27–38. ACM (2013)
3. Zheng, X., Wang, J., Shangguan, L., et al.: Smokey: ubiquitous smoking detection with commercial WiFi infrastructures. In: IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications, pp. 1–9. IEEE (2016)
4. Ren, Z., Meng, J., Yuan, J., et al.: Robust hand gesture recognition with kinect sensor. In: Proceedings of the 19th ACM International Conference on Multimedia, pp. 759–760. ACM (2011)
5. Weichert, F., Bachmann, D., Rudak, B., et al.: Analysis of the accuracy and robustness of the leap motion controller. Sensors **13**(5), 6380–6393 (2013)
6. IEEE Std. 802.11n-2009: Enhancements for higher throughput (2009). http://www.ieee802.org
7. Silver, D., Veness, J.: Monte-Carlo planning in large POMDPs. In: Advances in Neural Information Processing Systems, pp. 2164–2172 (2010)
8. Adib, F., Kabelac, Z., Katabi, D., et al.: 3D tracking via body radio reflections. In: NSDI, vol. 14, pp. 317–329 (2014)
9. Gollakota, S., Hassanieh, H., Ransford, B., et al.: They can hear your heartbeats: noninvasive security for implantable medical devices. ACM SIGCOMM Comput. Commun. Rev. **41**(4), 2–13 (2011)
10. Asadzadeh, P., Kulik, L., Tanin, E.: Gesture recognition using RFID technology. Pers. Ubiquit. Comput. **16**(3), 225–234 (2012)
11. Tongrod, N., Lokavee, S., Kerdcharoen, T., et al.: Gestural system based on multifunctional sensors and ZigBee networks for squad communication. In: 2011 Defense Science Research Conference and Expo (DSR), pp. 1–4. IEEE, 2011
12. Wang, Y., Wu, K., Ni, L.M.: Wifall: device-free fall detection by wireless networks. IEEE Trans. Mob. Comput. **16**, 581–594 (2016)
13. Molchanov, P., Gupta, S., Kim, K., et al.: Short-range FMCW monopulse radar for hand-gesture sensing. In: 2015 IEEE Radar Conference (RadarCon), pp. 1491–1496. IEEE (2015)
14. Wang, G., Zou, Y., Zhou, Z., et al.: We can hear you with Wi-Fi! IEEE Trans. Mob. Comput. **15**(11), 2907–2920 (2016)
15. Adib, F., Kabelac, Z., Katabi, D.: Multi-person localization via RF body reflections. In: NSDI, pp. 279–292 (2015)
16. Xie, Y., Li, Z., Li, M.: Precise power delay profiling with commodity Wi-Fi. In: Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, pp. 53–64. ACM (2015)
17. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
18. Kleisouris, K., Firner, B., Howard, R., Zhang, Y., Martin, R.P.: Detecting intra-room mobility with signal strength descriptors. In: ACM MobiHoc (2010)
19. Lei, J., Ren, X., Fox, D.: Fine-grained kitchen activity recognition using RGB-D. In: ACM UbiComp (2012)

20. Keally, M., et al.: PBN: towards practical activity recognition using smartphone based body sensor networks. In: ACM SenSys (2011)
21. Adib, F., Katabi, D.: See through walls with WiFi! In: ACM SIGCOMM (2013)
22. Yatani, K., Truong, K.N.: BodyScope: a wearable acoustic sensor for activity recognition. In: Proceedings of the ACM UbiComp (2012)
23. Halperin, D., et al.: Tool release: gathering 802.11n traces with channel state information. ACM SIGCOMM CCR **41**(1), 1 (2011)
24. Xia, P., Zhou, S., Giannakis, G.B.: Adaptive MIMO-OFDM based on partial channel state information. IEEE Trans. Signal Process. **52**(1), 202–213 (2004)
25. Hong, J., Ohtsuki, T.: Ambient intelligence sensing using array sensor: device-free radio based approach. In: CoSDEO Workshop (2013)