Robust Recognition of Indoor Educational Activity Based on Wi-Fi Signals with SSGAN

Jian Zhao[1], Xin Li[1], Liang Huang[1], Shangwu Chong[1], Jian Jia[2]

[1] School of Information Science and Technology, Northwest University, China
[2] School of Mathematics, Northwest University, China
zjctec@nwu.edu.cn, lixin1072@stumail.nwu.edu.cn, huangliang@stumail.nwu.edu.cn,
201932067@stumail.nwu.edu.cn, jiajian@nwu.edu.cn

Abstract

Wi-Fi-based activity recognition has been widely used in many fields. However, there is no research dedicated to educational activity recognition, and a lot of works using Wi-Fi signals for human activity recognition does not have generalization performance for left-out users whose CSI data are not training in model. In order to solve the above problems, we propose an educational activity recognition model which can effectively target at left-out user. In addition, we considered the generalization performance of the model when facing the activities with different directions and different scenes. The experimental results show that our model has excellent performance in the above situations, and also has higher recognition accuracy in the face of left-out user.

Keywords: CSI, Activity recognition, Wi-Fi, SSGAN, Robustness

1 Introduction

Human activity recognition technology has a wide range of applications in many aspects such as smart homes and the Internet of Things [1], and it is an important means of human-computer interaction. The current human activity recognition technology can be roughly divided into contact recognition and non-contact recognition. Contact recognition usually requires users to actively wear a specific device [2], which can quantitatively analyze the movement of the human body to achieve detailed Granular perception. However, this recognition method is costly and inconvenient to install and adopt, so that is only suitable for professional scenes with few recognition objects. Traditional non-contact recognition has strict requirements on the surrounding environment. It cannot work under weak light conditions and obstructions, and often infringes the user's privacy [3]. This leads the equipment is difficult to apply to scenes with high privacy protection requirements (Such as in bedroom).

In recent years, Wi-Fi signals have been gradually used in the field of human recognition. Researchers have realized a series of activity recognition technologies such as sleep monitoring [4], fall detection [5], gesture recognition [6], and driving activity recognition [7] by extracting the CSI of Wi-Fi devices. Benefiting from the large-scale popularization of commercial Wi-Fi, Wi-Fi-based activity recognition technology can achieve universal promotion at a very low cost.

At present, many CSI-based activity recognition technologies do not have generalization performance for left-out user. Even if the model trained with the deep network still shows high enough accuracy on the test set, when facing new users, due to the different physical characteristics and movement habits of the individual device users, the recognition accuracy will often decrease a lot of. This is reflected in many works. In a real scene, the device user needs to obtain a high recognition effect in a short period of time, so it is not feasible to collect a large number of labeled samples from left-out user. To train the model with a small number of labeled samples and a large number of unlabeled samples. By extracting the characteristics of left-out user, the model can obtain a higher generalization performance. In view of this situation, we proposed the use of SSGAN [8] to improve the accuracy of the recognition model with left-out user.

In the case of fewer labeled samples, in order to use unlabeled samples for training, we add a generator to the original model and modify the output of the original classifier to k+1 class. Therefore, the classifier gets a total of three types of input: labeled samples, unlabeled samples, and fake samples. The fake samples are generated by GAN [9]. The output of the k items of the classifier is the sample classification of the data, and the output of the k+1 item is the label of unlabeled samples or fake samples. The classifier accepts the optimization of the supervised model and the semi-supervised model at the same time. After the training is

*Corresponding Author: Jian Jia; E-mail: jiajian@nwu.edu.cn
DOI: 10.53106/160792642021122207008

completed, the classifier can be used to identify the educational activity of left-out user.

The experimental results prove that our model can achieve fine-grained perception of educational activity, and it is also robust when faced with different directions and different scenes. Even in the face of left-out user with only a small number of labeled samples, the semi-supervised model can achieve good recognition results.

In summary, the main contributions of our work are as follows:

(1) As far as we know, this is the first time that commercial Wi-Fi human activity recognition technology has been applied to the educational field.

(2) The model proposed in this paper can accurately identify six common educational activities, and proves the robustness of the model when facing activities in different directions and different scenes.

(3) This paper is based on the semi-supervised model implemented by GAN, which has good generalization ability in the face of left-out user, shows that it has practical value.

2 Materials and Methods

2.1 Data Collection

CSI is the information used to estimate the channel state in the commercial Wi-Fi OFDM technology, represents the transmission characteristics of the signal in the communication link (Figure 1). When encountering objects with a size smaller than the wavelength, scattering will occur. CSI can be captured on almost all commercial Wi-Fi devices, which provides us with low-cost and universal research conditions.

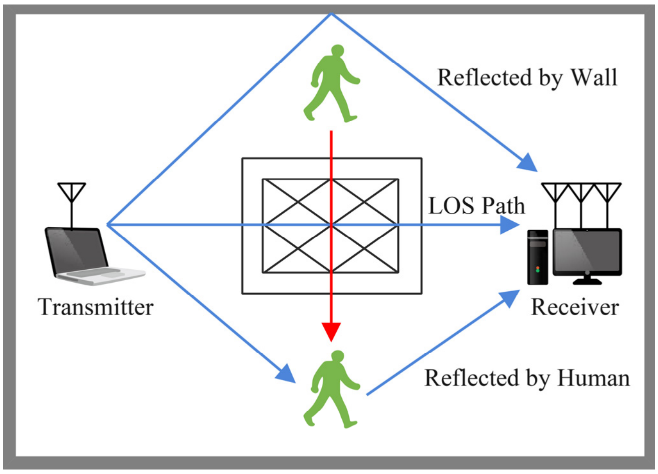


Figure 1. The transmission path of indoor wireless signals

CSI Tool released by Halperin et al. [10] in 2011 greatly promoted the research status based on CSI. The tool modifies the driver of the Intel 5300 network card and provides a CSI read interface. At present, most CSI-based work is done with the tool. The Intel 5300 network card uniformly extracts 30 sub-carriers according to the 802.11n standard. Therefore, the collected CSI matrix is as follows:

$$H_{CSI} = N_{sc} \times N_{tx} \times N_{rx} \quad \textbf{(1)}$$

Where $N_{sc} = 30$ is the number of subcarriers, N_{tx} is the number of transmitting antennas, and N_{rx} is the number of receiving antennas.

Our experimental equipment uses two computers equipped with Intel 5300 network card as the transmitter (TX) and receiver (RX) of the CSI respectively. We extend the 3 antennas of the receiver to the outside to obtain better signal transmission Quality. Our experimental data is collected on the 165 channel of the 5G frequency band. Compared to the 2.4G frequency band, there is no overlap between Wi-Fi channels on this frequency band, and there are currently fewer electronic devices working in this frequency band [11]. Therefore, collecting data on this frequency band can reduce irrelevant electromagnetic interference and obtain reliable signal transmission. The center frequency of the channel is 5.825 GHz, and the wavelength is 5.15 cm. In our experiment, the distance between the three receiving antennas is greater than $\lambda/2$ (λ is the wavelength), so that the fading signals received by different antennas are approximately independent, so as to obtain better spatial diversity gain and improve the reliability of signal transmission.

We collected 6 types of common educational activity from 11 volunteers, including 6 males and 5 females. The volunteers were 22-26 years old. These 6 types of educational activity include hand up, attend lecture, bend down, hand clap, Walk, and sit down (Figure 2). We assume that the duration of each type of activity is 3 seconds. Considering that the CSI signal fluctuation frequency caused by human activities in the indoor environment is not greater than 300Hz, while the frequency classification caused by human walking is even lower than 60Hz, we set the sampling frequency to 1000Hz to meet the Nyquist sampling theorem, the number of sent packets is counted as 3000.

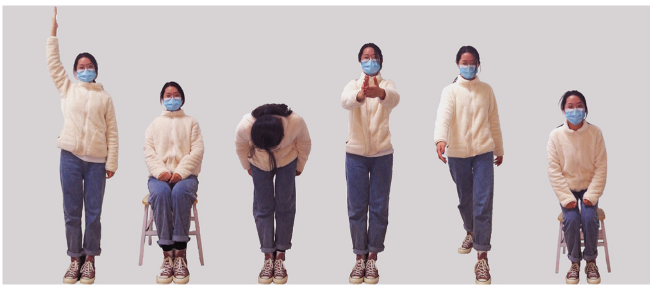
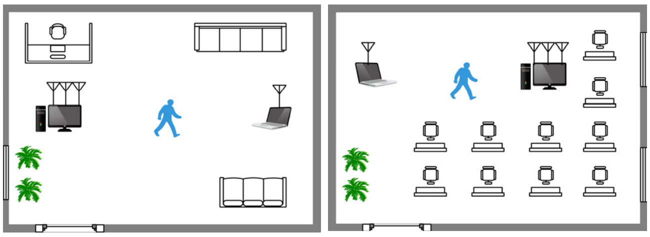


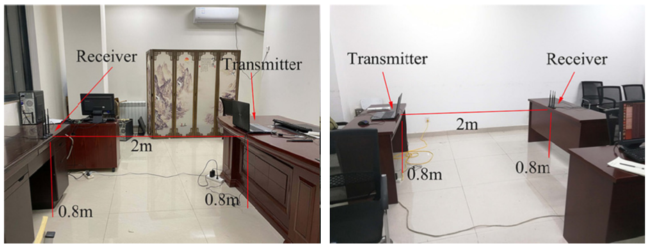
Figure 2. Educational activities (hand up, attend lecture, bend down, hand clap, walk, sit down)

We hope that the recognition model can work well in any scenes, so it is very important to obtain a robust model. In order to verify the robustness of the model in different scenes, our data was collected in the office and the conference room (Figure 3 and Figure 4). Different from the office scene, there is more furniture interference in the conference room, and the signal transmitter/receiver is located in the corner, which leads to more complicated multi-path transmission than scene 1.



(a) Office (b) Conference room

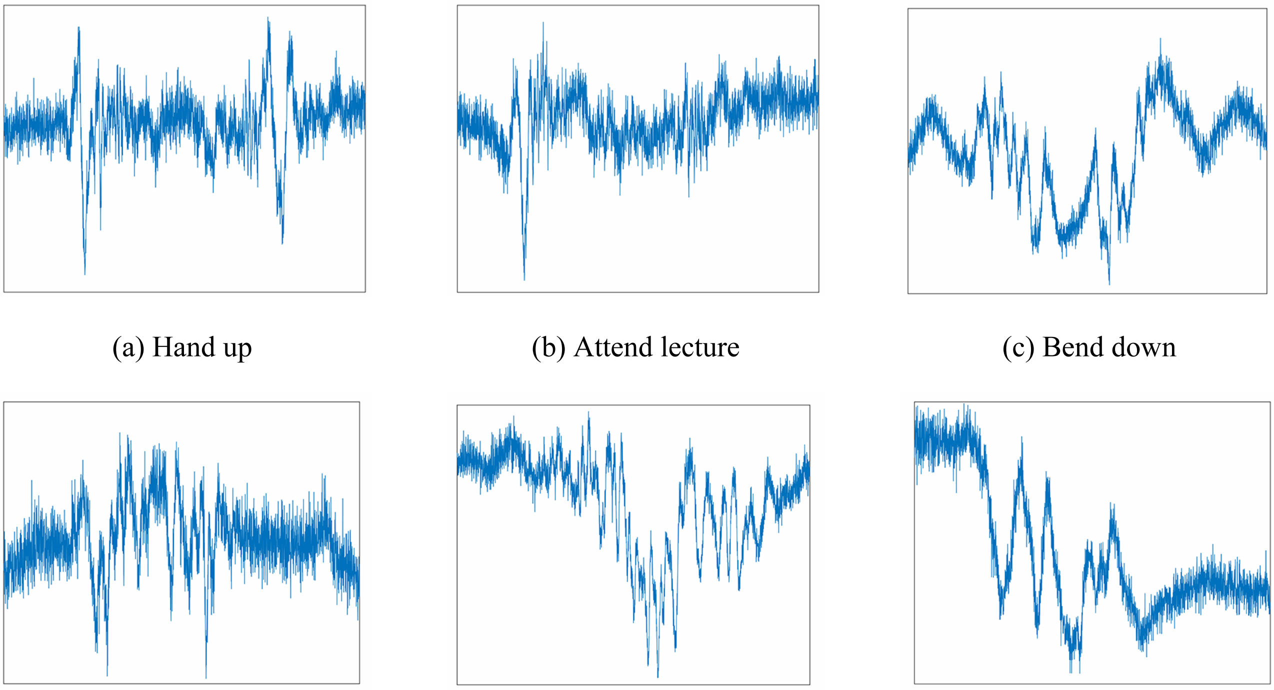
Figure 3. Plan structure diagrams of experimental scenes



(a) Office (b) Conference room

Figure 4. Real experimental scenes

Figure 5 shows the CSI sequence of different educational activities in the office. It can be seen from the figure that the CSI sequence changes caused by different activities are varied, but the characteristics that can distinguish different activity sequences cannot be intuitively given based on the figure.



(a) Hand up (b) Attend lecture (c) Bend down

(d) Hand clap (e) Walk (f) Sit down

Figure 5. CSI sequence caused by different educational activities

2.2 Data Preprocessing

CSI receives varying degrees of noise interference on the transmitter/receiver and transmission paths. In order to improve the recognition accuracy, it is necessary to preprocess the dataset first to filter out the noise that may affect the recognition effect. Different from RSS [12], CSI contains detailed amplitude information and phase information of the wireless signal in the transmission path. Referring to the processing methods of other documents [13], we consider only using amplitude information. The experimental results show that only using amplitude information can train a sufficiently good model, which is shown in the third part of this paper. The flow chart of the preprocessing is shown in Figure 6.

We use Hampel detection to check and eliminate outliers. The specific implementation method is to traverse the CSI sequence sub-carrier [14], and generate an observation window around each sample of the sequence, and use the absolute value of the median to estimate the standard for the median of each sample

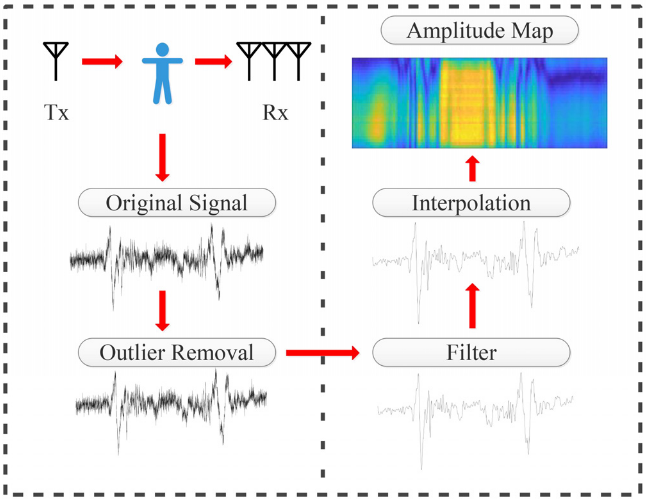


Figure 6. The flow chart of the preprocessing

pair difference. If a sample differs from the median by more than three standard deviations, the value is determined to be an outlier and the sample is replaced with the median.

CSI noise is mainly caused by environmental noise and changes in the internal transmission power and transmission rate of the transmitter/receiver, including some high-frequency components and DC components [15]. Considering that the CSI amplitude fluctuation caused by human activity is mainly in the low frequency range, we use a low-pass Butterworth filter to filter out the noise. The advantage of Butterworth filter lies in its maximum flat frequency response in the passband, so it will not cause excessive distortion of signal amplitude [16].

In the signal acquisition process, even if the data packets are transmitted at a fixed transmission rate, the received packets are not continuous due to electromagnetic interference, or some data packets are lost due to the delay in the decoding and sampling process, causing the CSI sequence to be non-uniform in time distributed. In order to ensure the uniform time interval of the packets, we use the time stamp linear interpolation method [17].

Figure 7 shows the preprocessed CSI subcarrier sequence. It can be observed that in each subcarrier sequence of CSI, outliers and high-frequency noise that affect the recognition performance have been filtered out, which is beneficial to the recognition of our model.

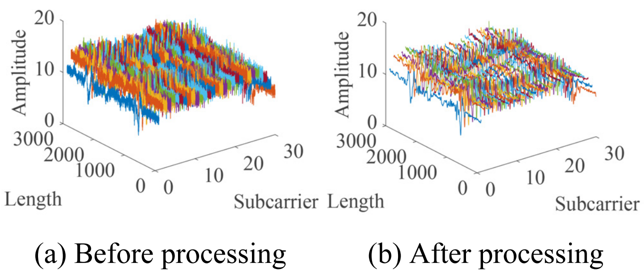


Figure 7. CSI sequence preprocessing

Our work converts the CSI sequence into amplitude map. Different colors in the amplitude map represent different amplitudes. We consider using 30 sub-carriers at the same time. As shown in Figure 8, different color block sequences in the amplitude map correspond to different educational activities. The size of the amplitude map generated by the CSI sequence is relatively large, so we compress it to a size of 64×64. The target tensor size is set to adapt to the output framework of GAN.

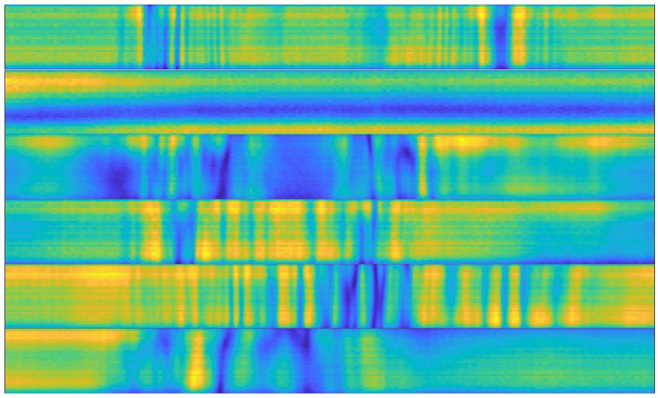


Figure 8. CSI amplitude map (hand up, attend lecture, bend down, hand clap, walk, sit down)

2.3 Model Framework

Semi-supervised learning can provide a better solution. In the semi-supervised field, the SSGAN model has achieved good results in image classification [18]. Therefore, our model is based on SSGAN design, and the model's framework is shown in Figure 9.

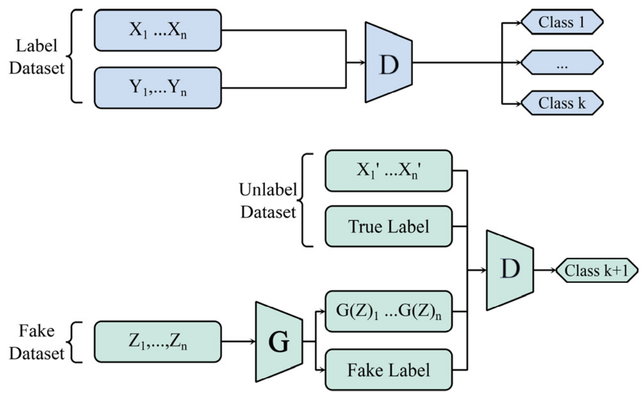


Figure 9. Framework of the educational activity recognition model

In supervised learning, labeled samples only passes through the classifier of model, and k-class labels are output. In the original DCGAN model [19], the discriminator is a 5-layer CNN network. The reason why we choose ResNet as the classifier is that it usually has better performance than CNN, and has fewer network parameters than VGG [20], DenseNet, GoogleNet [21], etc. The specific results are shown in the third part.

GAN is a kind of generative model, which is widely used in image generation, image conversion, image style transfer, data enhancement and other fields [22]. Vanilla GAN model obtains the data distribution of real samples, and the generator evolves random noise into fake samples that the discriminator cannot distinguish.

SSGAN [23] is an extend model of Vanilla GAN. By modifying the output of the discriminator in Vanilla GAN, it is transformed from a discriminator that can only distinguish true and false to a classifier that can classify k+1 categories. Where is the category of educational activity, and is to distinguish the unlabeled samples and fake samples, that is the work done by the discriminator in Vanilla GAN. The purpose of semi-supervised GAN is to train a classifier to have the ability to distinguish labeled samples and unlabeled samples at the same time. In the case that the generated samples are fake labels, the unlabeled samples get its labels as “true”, so the semi-supervised GAN classifier accepts 3 inputs, (1) Labeled sample is used to train the supervised model. (2) Unlabeled samples, (3) Generate samples, to train semi-supervised models.

The generator is a contrary structure of the classifier in SSGAN, in which Dense Layer expands the dimension of random noise, and Reshape Layer reshapes the expanded vector into a size suitable for the input of the Residual module. Convolutional Transpose Layer up samples the output of the Residual module, and finally obtains fake sample images with a size of 64×64×3, which is labeled as False. We use the fake samples in the model to stimulate the classifier to draw the class boundary in the low-density area to improve the performance of the classifier.

Our model is implemented based on keras 2.4.3 and trained on an Ubuntu server equipped with 2 RTX 2080 GPUs.

3 Evaluation and Results

3.1 Antenna Analysis

Our experimental samples are collected by 1 sending antenna and 3 receiving antennas. The purpose of using multiple antennas is to collect independent CSI sequences of different antennas to obtain better spatial diversity gain and improve the reliability of signal transmission. In addition, using the data of 3 receiving antennas at the same time can effectively expand the CSI dataset and avoid overfitting in the deep network. The effect of different combinations of receiving antennas on the accuracy of educational activity recognition is shown in Table 1.

Table 1. The effect of different antenna combinations on accuracy

Antenna combinations	1	2	3	12	13	23	123
Accuracy	95.3	95.5	96.4	97.1	97.0	97.0	97.4

Table 1 shows the impact of the dataset combined by different antennas on the results in the supervised mode, where the accuracy is obtained by stratified 10-fold cross-validation. The results show that only one single antenna has been able to achieve quite good accuracy, especially the third antenna. With the increase in the number of antennas, its results on the test set gradually increase, which shows that increasing the training data can prevent the model from overfitting and significantly improve our network prediction performance. This is also reflected in other work [23]. Unless otherwise specified, the following experimental results use all three antennas receiving samples for training, and the results are obtained by stratified cross-validation.

3.2 Semi-supervised Mode

3.2.1 Performance Analysis

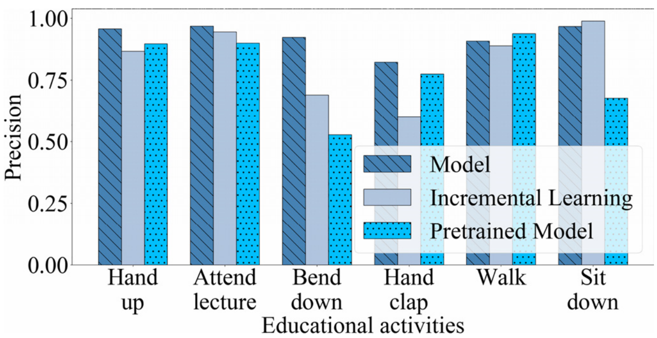
For left-out user, only a small number of labeled samples can be collected, while terminal user often needs to obtain better performance in a short time. In order to test the generalization performance of the model in the face of left-out user, we select labeled samples of a volunteer in the dataset, and divide each category into labeled samples and unlabeled samples in proportion. The experimental results are shown in Table 2.

Table 2. The influence of the number of labeled samples on semi-supervised learning

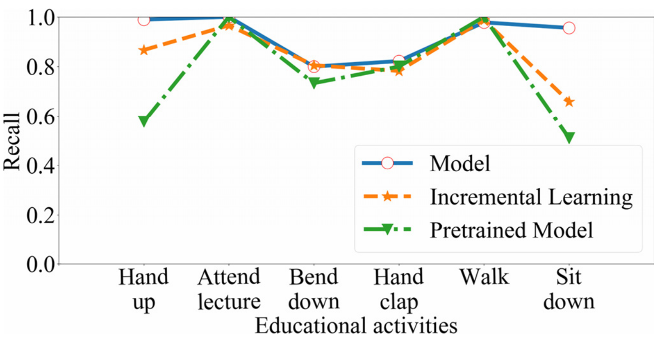
Accuracy / epoch / Samples	20	40	60	80	100
5	69.3	74.1	75.4	75.4	79.8
10	71.1	80.6	80.7	85.2	86.1
15	**78.9**	**87.6**	**89.1**	**89.8**	**92.4**
20	83.1	87.2	91.5	91.5	94.4

It can be seen from Table 2 that when there are only 5 labeled samples for each type of educational activity, the accuracy of the semi-supervised model reaches 80%. As the number of labeled samples of each type increases to 10, the accuracy of the semi-supervised model on left-out user has increased to about 86%. When the number of labeled samples of each type is increased to 15, the accuracy of the model increases to 92%. The result shows that with only a small number of labeled samples, semi-supervised model has been able to achieve a high accuracy. When the number of labeled samples for each type is 20, the accuracy of the model is 94%, which is close to the accuracy of the supervised model. The corresponding reduction in the number of labeled samples will result in a significant decrease in the performance of the semi-supervised mode. Considering that the increase in the number of labeled samples brings about the problem of rising labeling costs, we select 15 labeled samples for each type to illustrate the performance of our model in the semi-supervised mode.

In order to verify the effectiveness of the SSGAN in the face of left-out user, we compare the model with an incremental semi-supervised model and call it Incremental learning. The model is pretrained by labeled samples, and unlabeled samples are iteratively fed into the pretrained model. Use unlabeled samples and fake labels to enhance the trainset until the entire unlabeled samples set is traversed. In addition, we compare the performance of the above two methods with the pretrained model test on left-out dataset to obtain the performance improvement results of the semi-supervised model. The experimental results are shown in Figure 10.



(a) Precision



(b) Recall

Figure 10. Semi-supervised mode performance evaluation Precision and Recall

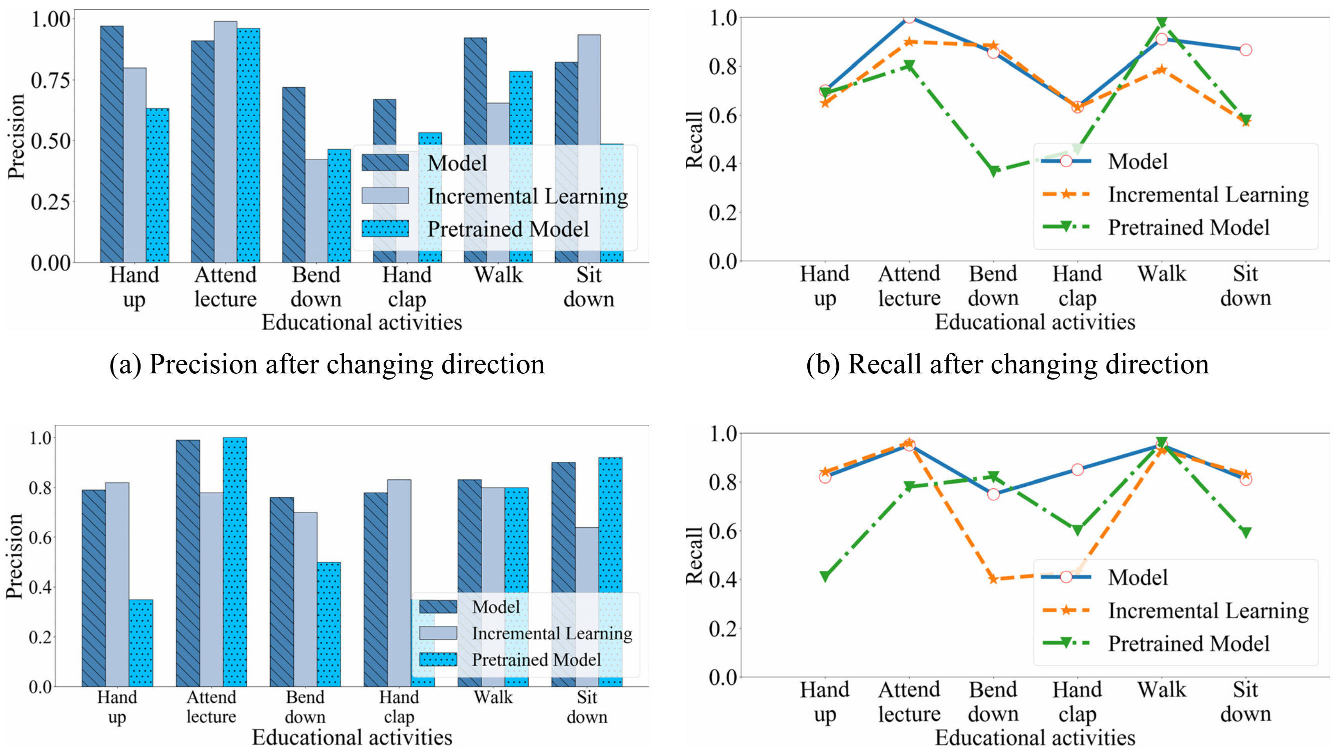
It can be seen from Figure 10 that the performance of pretrained model in bend down activity has been reduced to 50%. It also performed poorly on hand clap (77%) and sit down (68%). In addition, the recall for hand up and sit-down activities are about 58% and 51%, respectively, which means that it is difficult to make correct predictions about hand up and sit-down activities in real scene. Incremental learning can improve the generalization ability of the model to a certain extent, and shows better recognition ability in actions such as sit down. The precision in bend down and hand clap is still only about 65%. In addition, incremental learning is sensitive to outliers, and the model needs to maintain considerable recognition accuracy on the initial sample and subsequent samples in order to improve the performance of the model through iteration. If the model makes too many mistakes on a certain sample, it will cause errors to accumulate and affect subsequent judgments. In contrast, our model still achieved good recognition ability in the above 6 educational activities. The accuracy of bend down and sit down also reached 92% and 91%. It also has 82% precision in hand clap, even if it has more difference in personal habit. In summary, the accuracy of our model on the standard data set is about 92%. This is significantly higher than incremental learning (83%) and pretrained models (77%).

3.2.2 Robustness Analysis

We also verify the robustness of our model in the face of activities from different directions and scenes. Figure 11 shows the experimental results of our model when facing direction dataset.

It can be seen from Figure 11 that without changing the scene and the position and direction of the transmitter/receiver, only the direction at which the educational activity starts is changed. Comparing the results of standard data, the performance of the three methods are all degraded. The precision and recall of the pretrained model in each educational activity have dropped significantly, especially in the activity of bend down and sit down, the precision is less than 50%, and the recall of bend down and hand clap is about 40%, which shows the pretrained model is difficult to accurately recognize the above activities in practice. In contrast, incremental learning achieves performance improvements in hand up, attend lecture, and sit down. The precision in bend down and hand clap even declined. This shows that incremental learning has accumulated errors in the recognition of these two activities. Our model still shows higher performance than the pretrained model and incremental learning. The precision of hand up, attend lecture and walk are all higher than 90%, and it is also showing the high precision in bend down and hand clap. The accuracy of our model is about 83%, which shows that our model has better robustness when faced with activities that change directions.

It can be seen from Figure 11 that the performance of the three methods has all declined in the more complex environment scene 2. The precision of the pretrained model in hand up and hand clap is lower than 40%, and the recall in hand up and sit down are 41% and 60% respectively. The result means that more complex environments bring more wireless signal



(a) Precision after changing direction

(b) Recall after changing direction

(c) Precision after changing scene

(d) Recall after changing scene

Figure 11. Semi-supervised mode performance evaluation Precision and Recall

transmission path makes it more difficult for the model to generalize and identify left-out user. In contrast, incremental learning has improved accuracy in activities such as hand up and hand clap, but the recall in bend down and hand clap is less than 50%, which means that incremental learning is difficult to accurately predict the above two in real scenes. In contrast, the performance of our method in each educational activity is relatively stable, especially in the recall. The accuracy of our model in scene 2 is about 86%, which is also higher than the incremental learning (77%) and pretrained model (70%), proving that our model is robust in the face of more complex environments. Table 3 shows the accuracy in different conditions of the semi-supervised mode. With the change of volunteer's direction and the complex scene, the accuracy of our work decreases significantly less than that of other models.

Table 3. Accuracy in semi-supervised mode

Model / Dataset	**Our work**	Incremental learning	Pretrained model
Normal	**92**	83	77
Direction	**83**	71	64
Scene 2	**86**	77	71

4 Conclusion

In this paper we propose a method of using commercial Wi-Fi to recognize educational activities, and verify the robustness of the method when facing different directions of activity and different scenes. Meanwhile we considered the general model's difficulty in generalizing the left-out user in real scenes, and proposed a method to improve the generalization performance of the model by using a small number of labeled samples and some unlabeled samples. The experimental results show that our model can effectively solve the above problems. Nevertheless, our model has certain limitations. First of all, our experimental data can only be collected one by one, which greatly reduces our data collection efficiency. There are already some methods that can automatically detect the moment when activity starts in a long sequence and tailor the desired sequence. Secondly, the performance of our model is limited by the size of the dataset used, and we will collect more data to train a deeper network model in order to achieve better recognition performance. The above limitations can be further research directions.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Project 12071369) and Key Research and Development Program of Shaanxi (Program No. 2019ZDLSF02-09-02).

References

[1] V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M.

Mordonini, I. D. Munari, IoT Wearable Sensor and Deep Learning: An Integrated Approach for Personalized Human Activity Recognition In A Smart Home Environment, *IEEE Internet of Things Journal*, Vol. 6, No. 5, pp. 8553-8562, October, 2019.
[2] A. Mannini, S. Intille, Classifier Personalization for Activity Recognition Using Wrist Accelerometers, *IEEE Journal of Biomedical and Health Informatics*, Vol. 23, No. 4, pp. 1585-1594, July, 2019.
[3] U. M. Kamthe, C. G. Patil, Suspicious Activity Recognition In Video Surveillance System, *2018 4th International Conference on Computing Communication Control and Automation (ICCCA)*, Pune, India, 2018, pp. 1-6.
[4] X. Liu, J. Cao, S. Tang, J. Wen, Wi-Sleep: Contactless Sleep Monitoring via WiFi Signals, *2014 35th IEEE Real-Time Systems Symposium (RTSS)*, Rome, Italy, 2014, pp. 346-355.
[5] R. M. Keenan, L. N. Tran, Fall Detection Using Wi-Fi Signals and Threshold-Based Activity Segmentation, *2020 31st IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, London, United Kingdom, 2020, pp. 1-6.
[6] J. Yang, H. Zou, Y. Zhou, L. Xie, Learning Gestures From WiFi: A Siamese Recurrent Convolutional Architecture, *IEEE Internet of Things Journal*, Vol. 6, No. 6, pp. 10763-10772, December, 2019.
[7] F. Wang, J. Liu, W. Gong, WiCAR: WiFi-based in-Car Activity Recognition with Multi-Adversarial Domain Adaptation, *2019 27th IEEE/ACM International Symposium on Quality of Service (IWQoS)*, Phoenix, AZ, USA, 2019, pp. 1-10.
[8] N. Souly, C. Spampinato, M. Shah, Semi Supervised Semantic Segmentation Using Generative Adversarial Network, *2017 11th IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 5689-5697.
[9] H. Ko, D. Y. Lee, S. Cho, A. C. Bovik, Quality Prediction on Deep Generative Images, *IEEE Transactions on Image Processing*, Vol. 29, pp. 5964-5979, April, 2020.
[10] D. Halperin, W. Hu, A. Sheth, D. Wetherall, Tool release: Gathering 802.11n Traces with Channel State Information, *ACM SIGCOMM Computer Communication Review*, Vol. 41, No. 1, pp. 53-53, January, 2011.
[11] Z. Han, L. Guo, Z. Lu, X. Wen, W. Zheng, Deep Adaptation Networks Based Gesture Recognition Using Commodity WiFi, *2020 22th IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, South Korea, 2020, pp. 1-7.
[12] X. Guo, N. R. Elikplim, N. Ansari, L. Li, L. Wang, Robust WiFi Localization by Fusing Derivative Fingerprints of RSS and Multiple Classifiers, *IEEE Transactions on Industrial Informatics*, Vol. 16, No. 5, pp. 3177-3186, May, 2020.
[13] Z. Wang, B. Guo, Z. Yu, X. Zhou, Wi-Fi CSI-Based Behavior Recognition: From Signals and Actions to Activities, *IEEE Communications Magazine*, Vol. 56, No. 5, pp. 109-115, May, 2018.
[14] M. T. Rahman, S. Han, N. Tadayon, S. Valaee, Ising Model Formulation of Outlier Rejection, with Application in WiFi based Positioning, *2019 44th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, 2019, pp. 4405-4409.
[15] J. Zhang, F. Li, J. Chen, Detection for OFDM-AF Relay Systems Under Gaussian Mixture Noise with Imperfect CSI, *IEEE Transactions on Vehicular Technology*, Vol. 69, No. 5, pp. 5666-5671, May, 2020.
[16] L. Jia, Y. Gu, K. Cheng, H. Yan, F. Ren, BeAware: Convolutional Neural Network (CNN) Based User Behavior Understanding Through WiFi Channel State Information, *Neurocomputing*, Vol. 397, pp. 457-463, July, 2020.
[17] Y. Lu, F. Wu, S. Tang, L. Kong, C. Chen, Pushing the Limit of CSI-based Activity Recognition: An Enhanced Approach via Packet Reconstruction, *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, Boston, MA, USA, 2019, pp. 1-9.
[18] G. J. Qi, L. Zhang, H. Hu, M. Edraki, J. Wang, X. S. Hua, Global Versus Localized Generative Adversarial Nets, *2018 31th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 1517-1525.
[19] T. Hinz, M. Fisher, O. Wang, S. Wermter, Improved Techniques for Training Single-Image GANs, *2021 26th IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 2021, pp. 1299-1308.
[20] S. Gu, L. Ding, A Complex-Valued VGG Network Based Deep Learing Algorithm for Image Recognition, *2018 9th International Conference on Intelligent Control and Information Processing (ICICIP)*, Wanzhou, China, 2018, pp. 340-343.
[21] B. K. Kim, H. Kang, S. Park, Drone Classification Using Convolutional Neural Networks With Merged Doppler Images, *IEEE Geoscience and Remote Sensing Letters*, Vol. 14, No. 1, pp. 38-42, January, 2017.
[22] H. Tang, H. Liu, N. Sebe, Unified Generative Adversarial Networks for Controllable Image-to-Image Translation, *IEEE Transactions on Image Processing*, Vol. 29, pp. 8916-8929, September, 2020.
[23] C. Min, Y. Li, L. Fang, P. Chen, Conditional Generative Adversarial Network on Semi-supervised Learning Task, *2019 5th IEEE International Conference on Computer and Communications (ICCC)*, Chengdu, China, 2019, pp. 1448-1452.

Biographies

Jian Zhao received his Ph.D. degree in Signal and Information Processing from Northwestern Polytechnical University in 2003. He is currently the Professor in the School of Information Science and Technology, Northwest University. His currently research interest includes signal and information processing, artificial intelligence and cognitive psychology.

Xin Li received the B.S. degree from the Northwest University in 2019. He is currently pursuing the M.S. degree in Northwest University. He current research interests include information processing and artifical intelligence.

Liang Huang received the B.S. degree from the Anhui University of Science and Technology in 2019. He is currently pursuing the M.S. degree in Northwest University. His current research interests are wireless sensor networks.

Shangwu Chong received the B.S. degree from the Northwest University in 2019. He is currently pursuing the M.S. degree in Northwest University. He current research interests include information processing and artifical intelligence.

Jian Jia received his Ph.D. degree in pattern recognition and intelligent systems from Xidian University in 2008. He is currently the professor in the School of Mathematics Science, Northwest University. His current research interests include image processing and artificial intelligence.