WiMPP: An Indoor Multi-person Positioning Method Based on Wi-Fi Signal

Pengsong Duan, Biao Ye, Chenfei Jiao, Weixing Zhang(✉), and Chao Wang

School of Software, Zhengzhou University, Zhengzhou 450002, China
{duanps,rjwxzhang}@zzu.edu.cn

Abstract. In the era of Internet of things, convenient and high-precision location service is of great importance for the connection among things. In recent years, the indoor positioning technology based on Wi-Fi devices has developed rapidly, but there is still space for the improvement of accuracy in multi-target positioning. In this paper, a multi person positioning method named WiMPP based on Wi-Fi signal is proposed for the high-precision positioning in indoor scenes. WiMPP first collects the Wi-Fi sensing signals in environment with only one pair of transmit and receive antennas, and then estimates AOA, TOF and other parameters using two-dimensional MUSIC algorithm; Then, the estimated parameters are constructed as a heat map which is then inputted into a two-dimensional convolution neural network for training and classification such that the positioning of targets can be obtained. The experimental results show that WiMPP can achieve high precision positioning accuracy (average error distance is 6 cm, median error distance is 8 cm) under the condition that two persons are in the indoor scene. Compared with other location methods based on Wi-Fi signal, WiMPP not only can position multiple persons, but also improves the location accuracy to a certain extent.

Keywords: Indoor multi-person positioning · Wi-Fi sensing · MUSIC algorithm · CNN

1 Introduction

With the rapid development of computer technology and mobile communication technology, human society is experiencing the change of information technology. Smart life and smart city have become hot topics. More convenient and easier interaction with the surrounding environment and equipment has gradually become people's primary need. With the huge growth in the number of wirelessly connected devices, wireless signals are everywhere. These signals interact with people and objects in the environment and imply information about various properties of the surroundings Therefore, using wireless signals to sense surrounding environment has aroused great interest among researchers. For example, many attempts have been made to use wireless signals for positioning [1, 2], health detection [3, 4], motion recognition [5, 6], identity recognition [7], and gesture recognition [8–10]. In particular, passively locate multiple people walking in an

C. T. Calafate et al. (Eds.): MONAMI 2021, LNICST 418, pp. 115–126, 2022.
https://doi.org/10.1007/978-3-030-94763-7_9

area, without relying on any equipment carried by participants, this is a very challenging problem, and has important significance in many applications, such as elderly life monitoring, intrusion detection and retail analysis.

At present, indoor positioning technologies include Radio Frequency Identification (RFID) positioning technology [11], Ultra-Wideband (UWB) positioning technology [12], Bluetooth positioning technology [13], Ultrasonic positioning technology [14], Infrared positioning technology [15], ZigBee technology [16] and various other positioning technologies. After years of research and exploration, significant progress has been made. However, these positioning technologies usually require special hardware equipment or special deployment, which leads to high application cost and popularization difficulty. Wi-Fi network has been widely used in indoor positioning system research because of its high penetration rate and low cost. Therefore, Wi-Fi signal-based indoor positioning technology has become the mainstream of current indoor positioning technology. In this paper, the main problem we want to deal with is to achieve multiple targets passive localization in a specific environment.

In this paper, we propose a localization method that uses only few Wi-Fi resources to achieve passive in-region multi-person localization.

- We use two-dimensional MUSIC algorithm to estimate parameters such as AoA and ToF contained in received Wi-Fi signals and generate a heat map. Then, a two-dimensional convolutional neural network is constructed to classify the parametric heat map obtained after signal processing for target localization.
- We conduct experiments to validate our proposed multi-person localization method in a test environment, and test the localization results for the 2-person case. The results show that high-precision localization for two people can be achieved in the target environment with an average error distance of 6 cm and a median error distance of 8 cm.

2 Related Work

2.1 Related Research Review

In this section, we review the latest developments in passive target location using wireless signals. We first discuss the work that only focuses on single-target positioning. Then summarized the work of realizing passive multi-target positioning. A detailed comparison of the different proposed methods (including ours) for single and multiple target tracking is shown in Table 1.

Single-Target Positioning. In [17], an indoor fingerprint recognition system based on CSI was proposed, which uses the ownership value of the deep neural network as the weight of fingerprint training layer by layer. The average distance error tested in the living room is about 0.95 m.

The Widar system measures the relationship between CSI information and the user's position and speed without using statistical learning techniques, and can reach an error distance of 38 cm [18].

Yin Zhendong proposed a positioning scheme UWB-IP based on Wavelength Division Multiple Access (WDMA), which achieves centimeter-level positioning accuracy with a second-level response, and can also achieve an error within 10 cm in a noisy environment [19]. In [20], a UWB positioning algorithm based on the attention mechanism was proposed. This method can reduce positioning errors caused by multipath effects and non-line-of-sight in a dynamic environment. Experimental results show that the error distance can reach 0.0012 m [20].

Multi-target Positioning. In [21], a two-person positioning method based on channel state information (CSI) is proposed. Two kinds of resolution fingerprint libraries are constructed in the offline phase. In the online matching stage, it matches with the fingerprint database to obtain accurate positioning results. Through experimental verification, the minimum positioning error of this method is 1.12 m.

In [22], a multi-parameter indoor passive positioning technology is proposed. First, the relationship between user motion and channel state information (CSI) is quantified through the parameter model of the angle of arrival (AoA), time of flight (ToF) and Doppler frequency shift (DFS), and the generalized expectation maximization algorithm (Alternating Generalized Expectation Maximization algorithm (SAGE) refines the wrong original parameters to a certain precise range, and finally outputs the target position.

Table 1. Comparison with the technologies in target positioning using RF signals

Paper	Number of targets	Bandwidth	Number of devices used	Positioning error
DeepFi,	Single	Narrowband	2 WiFi NICs	0.95 m
Widar	Single	Narrowband	2 WiFi NICs	0.38 m
[19]	Single	UWB radar	1 UWB radar	0.1 m
[20]	Single	UWB radar	1 UWB radar	0.01 m
[21]	Multiple (2)	Narrowband	2 WiFi NICs	1.12 m
[22]	Multiple (5)	Narrowband	2 WiFi NICs	0.8 m−1.3 m
This paper	Multiple (2)	Narrowband	2 WiFi NICs	0.08 m

3 System Design

3.1 2D MUSIC Algorithm

In this section, we use the two-dimensional multi-signal classification algorithm to estimate the value of the parameters (AoA, ToF) related to the target position, because it has the advantage of high resolution.

Consider a scenario where the receiving array contains M and the distance between antennas A is *dant*. The antenna of the array samples the received signal for *Twin* length at a rate of *1/T* (sample per second). Therefore, the number of samples in space and time is M_A and $M_T = Twin / Ts$, respectively. Use C to represent the $M_A \times M_T$ matrix of the square of the magnitude in the space-time window:

$$C = \begin{bmatrix} |c_{1,1}|^2 & |c_{1,2}|^2 & \dots & |c_{1,M_T}|^2 \\ |c_{2,1}|^2 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ |c_{M_A,1}|^2 & \dots & \dots & |c_{M_A,M_T}|^2 \end{bmatrix} \tag{1}$$

$c_{i,j}$ are channel state information. We express C as the following mathematical model:

$$C = AS + N \tag{2}$$

Where N is the noise, A is the steering vector of the signal, S is the incident signal:

$$S = [S_1(t), S_2(t), \dots S_n(t)]^T \tag{3}$$

Rc is the correlation matrix of vector C:

$$R_c = E\left[CC^H\right] = AR_S A^H + R_N \tag{4}$$

Where $R_S = E[SS^H]$, $R_N = \sigma^2 I$, E is the expectation, σ^2 is the noise power, and I is the identity matrix.

The feature vector of Rc is divided into a signal subspace (its dimension is equal to the rank of Rs) and a noise subspace (which is orthogonal to all direction vectors corresponding to the N signal paths to the receiving array). Therefore, we can define the frequency spectrum as:

$$P\left(\psi^M, \psi^A\right) = \frac{1}{a^H(\psi^M, \psi^A) E_n {E_n}^H a(\psi^M, \psi^A)} \tag{5}$$

Where E_N is a matrix composed of various noise eigenvectors. a is the guiding vector.

The denominator in this formula is the inner product of the signal vector and the noise matrix, because the signal subspace is orthogonal to the noise subspace. Therefore, when ψ^M and ψ^A are equal to the position corresponding to the target, the denominator gets 0, where the spectrum takes a maximum value. Generally, we can obtain the correlation value (θ and τ) of the target's position parameters (AoA and ToF) by searching for the peak of the spectrum peak.

However, in the actual environment, because of the coherence of the signals of different targets, the correct position cannot be obtained through spectral search. Although the frequency spectrum does not directly reflect the final result, it contains sufficient multi-object spatial features. Therefore, our method uses deep learning techniques to train a model, extract features from the spectrum and predict the location of the target. After calculation, a set of (AoA, ToF) parameter estimates can be obtained, which are displayed on a high-resolution heatmap, as shown in Fig. 1, so that we can sort them in the network model in next section.

The performance and estimation accuracy of the MUSIC algorithm mentioned above are affected by the number of sources, so at present this article can only estimate the motion parameters of two targets when only a pair of transceiver devices are used.

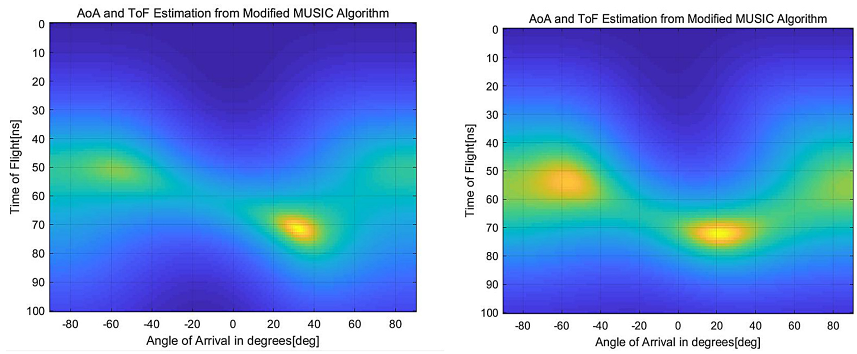

Fig. 1. Examples of heat maps processed by MUSIC algorithm. The highlighted part in the figure is the estimation of the parameter value pair containing the target position information

3.2 Convolutional Neural Network

In this section, we show how to input the heat map obtained by processing the original CSI data through the MUSIC algorithm into a convolutional neural network we proposed to locate the target. Its overall structure is shown in Fig. 2.

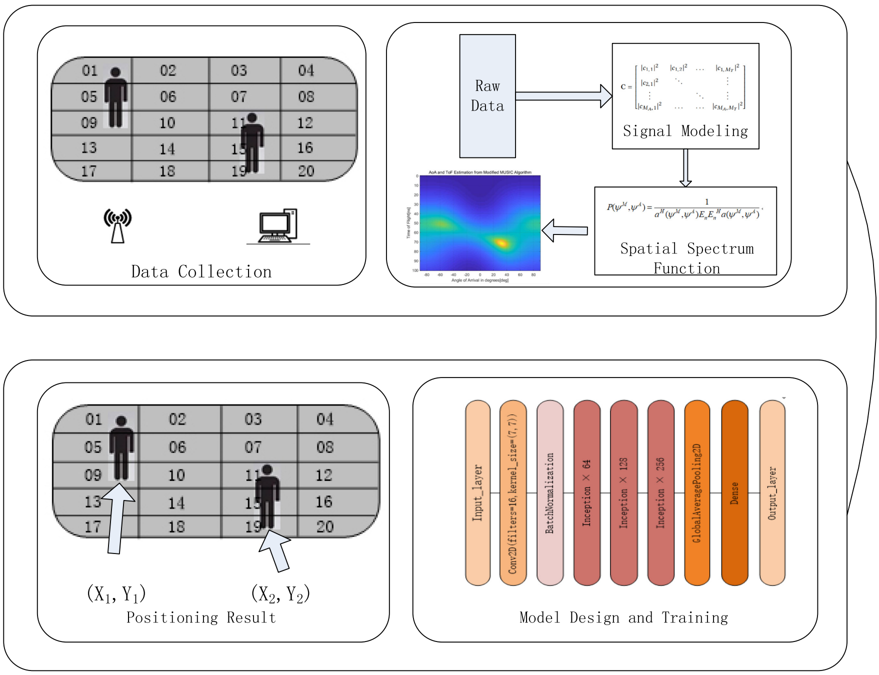

Fig. 2. The overall structure diagram of the neural network model. Including four parts, data acquisition, signal modeling, construction of spatial spectrum function, target positioning.

3.3 Neural Network Structure

The model structure of this article is built with Keras, and the parameters of the model, the size of the convolution kernel, the number of model layers, etc. are reasonably set according to the experimental needs. The detailed model structure is shown in Fig. 3.

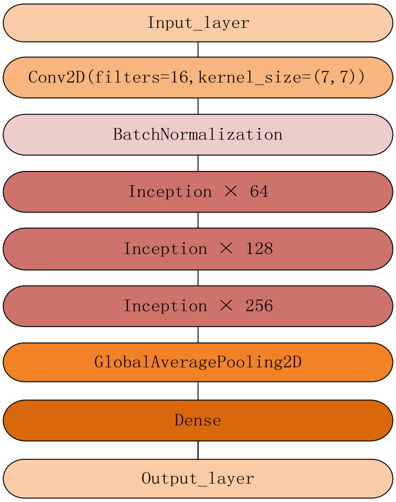

Fig. 3. Neural network model structure diagram, which includes the input layer, two-dimensional convolutional layer, improved Inception module, dense layer and output layer.

3.4 Module Introduction

Improvements in the Inception Module
In the model, the convolution operation is the core operation. Inception V3 solves the GoogleNet spatial convolution into an asymmetric convolution, and accelerating the calculation also increases the nonlinearity of the network [24]. Therefore, this article improves the Inception module on its basis, not only introduces the residual link [25] on its basis, adds the underlying features to the subsequent operations; and uses 1×1 convolution instead of the Concatenate layer to make features of different scales Perform a linear combination. The improved module is shown in Fig. 4:

The convolution module is to perform preliminary feature extraction on the data of the input model, and extract some overall data features, and use them as input to enter the next module.

The improved Inception module effectively solves the problem of feature degradation, so that the neural network can reach a deeper level and better deal with features of different scales.

Classifier
According to the definition of the problem, the input of the model is a heat map, denoted

by X, and the output is the coordinates of the positions of two people, denoted by Y. The function of the model is to predict the position category Y of the two persons through the input X. The number of monitored locations in the location problem studied in the paper is greater than 2, so the problem is determined to be a multi-classification problem. The Softmax function can be used for multiple classifications, so this article chooses the Softmax function to distinguish the action categories. The expression formula is as follows:

$$P(y|X) = \frac{\exp(z_y)}{\sum_{y=1}^{r} \exp(z_y)}, \text{ y} \in [1, \text{r}] \tag{6}$$

In the formula, r is the number of categories of y, which represents the result of the global average pooling operation, and represents the posterior probability of the model input X predicted belonging to the y category.

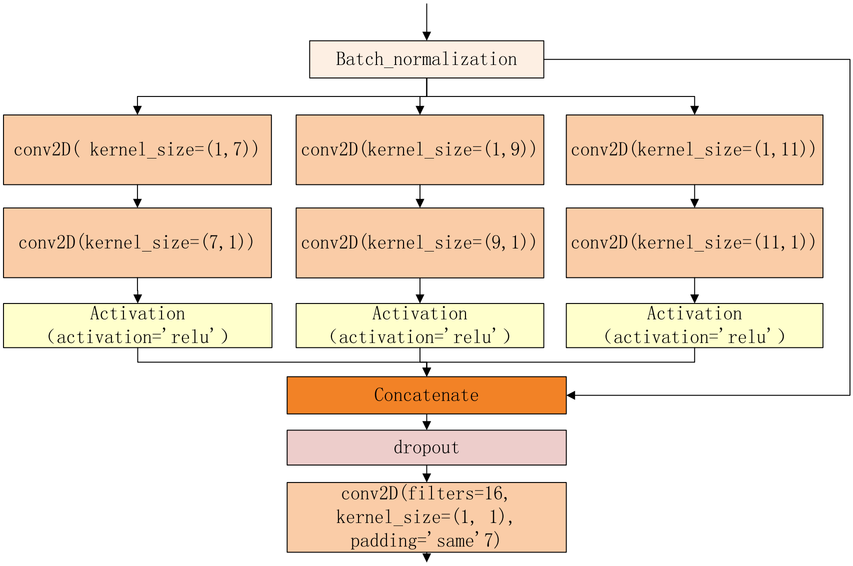

Fig. 4. The improved Inception module structure diagram, which combines residual linking and asymmetric convolution.

4 Experiment and Result Analysis

4.1 Evaluation Index

This paper uses Average Distance Error, Median Distance Error, and CDF (Cumulative Distribution Function) curves to evaluate the classification effect of the model. Among them, the average distance error is defined as:

$$\text{Average Distance Error} = \frac{\sum_{i=1}^{m} \sqrt{(x_i - p_i)^2 + (y_i - q_i)^2}}{m} \tag{7}$$

Where m is the number of test points, (x_i,y_i) is the predicted coordinates of the i-th test point, and (p_i,q_i) is the true coordinate of the i-th test point.

The Median Distance Error is the median obtained from the sum of the error distances.

CDF (Cumulative distribution function) curve refers to the sum of the probability of occurrence of all values less than or equal to a for a continuous function. That is, F(a) = P(x< = a).

4.2 Experimental Setup

We choose a relatively empty laboratory for multi-person positioning experiments. In each experiment, two targets are invited to be located in the detection area on the side of the Wi-Fi device. A total of 2000 pieces of CSI data in a two-person environment were collected as the training set and the test set.

The experimental equipment is composed of a TP_LINK AC1750 wireless router as a transmitter and a ThinkPad X201 portable computer terminal equipped with an Intel 5300 802.11n Wi-Fi NIC as a receiver.

Among them, the open source CSI Tools is installed on the receiving end device, which can realize real-time collection of CSI information. The transmitter and receiver of the Wi-Fi device have 1 and 3 antennas respectively, that is, there are (1 × 3) 3 antenna pairs, and each antenna pair contains 30 sub-carriers. Therefore, the experiment collected (1 × 3 × 30) CSI data of 90 sub-carriers. The placement of the experimental device is shown in Fig. 5.

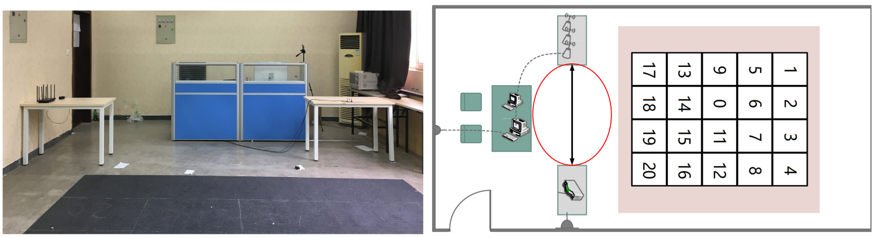

Fig. 5. The experimental environment map is composed of the receiving device area (left) and the positioning area (right). The positioning area is divided into 20 sub-areas of the same size, numbered starting from 1. When positioning data is collected, the inspected person is stationary in the 20 sub-areas.

There are 90 CSI data streams collected in this experiment, and all experiments are done at 5 GHz frequency to ensure the integrity and accuracy of the data. The sampling rate of the CSI data packet is set to 1000 data packets per second at the receiving end. In order to extract more feature information, the sliding step length d = 200 and the sliding window T = 600 ms are set. The sliding step length and the overlap between the windows ensure the continuity of the data after segmentation. Based on this, the sample data was constructed and different heat maps were generated.

In the selection of deep learning tools, the deep learning framework of this article uses Keras, and the model construction is implemented in Python. The collected CSI data is labeled with the positions of two persons, and the proposed neural network is used to supervise and learn the heat map of the localized target. The data of the data set is divided into two parts: training data and test data. Among them, training data accounts for 90% of all collected CSI data, and the remaining 10% is used as test data. The test set is used to evaluate the positioning effect after training. At present, the verification and evaluation of the model are based on the data set collected by our team.

4.3 Experimental Results

Analysis of Experimental Results

In order to verify the positioning effect of our model, this article designed a comparative experiment using different models in this scenario. In terms of model selection, we chose the classic LSTM and VGG16 [26] to compare with this experimental model, using the above The detailed comparison results of the proposed evaluation methods are as shown in Table 2 (Fig. 6).

Table 2. Comparison table of experimental results of different models.

Index model	ADE (m)	MDE (m)	Parameters	GFLOPs	Duration (ms)
LSTM	0.76	0.76	572,244	4.4	7571.22
VGG	0.33	0.39	138,357,544	15.5	5962.71
This paper	0.08	0.06	2,541,974	7.6	2771.56

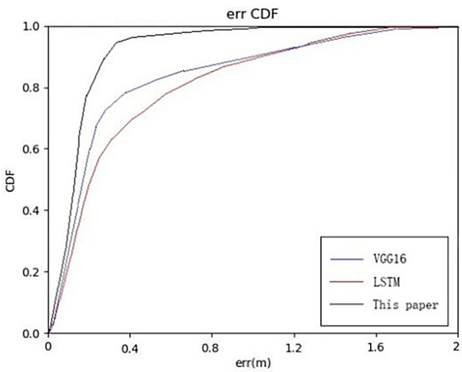

Fig. 6. The cumulative distribution function graph of the error in the experimental results of different neural network models.

LSTM is an excellent variant model of RNN. It inherits most of the characteristics of RNN models and at the same time solves the problem of gradient disappearance caused by the gradual reduction of the gradient backpropagation process. Therefore, LSTM is suitable for dealing with problems that are highly related to time series. It performs poorly in the image classification problem of this experiment. The VGG16 model is composed of several convolutional layers and pooling layers to form a deeper network structure. The depth of the layer makes the feature map wider and more suitable for large data sets, but the 16-layer network structure now looks It is not very deep, the fitting effect is not as good as our 23-layer model, and VGG has a very large number of parameters, resulting in too long training time. After updating the Inception module, the network model we proposed adds residual links, which can overcome the problem of feature degradation, thereby making the model more in-depth and improving the processing of data with different scale features. Therefore, the performance is the best in the comparison test, and the parameter quantity is moderate. This not only ensures the model fitting ability, but also enables the training time to meet actual needs.

The Impact of Movement on Positioning Results

In order to verify the influence of the action of the tested person on the positioning result in the experiment, we collected different actions in the same indoor environment and the same deployment. We additionally collected four different actions of the target which including the up, down, left, and right actions of the target in the same position grid, as shown in Fig. 7.

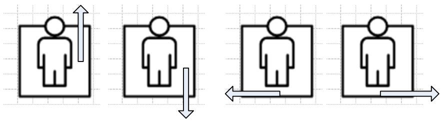

Fig. 7. Four different actions of the target in the same position grid, which include the up, down, left, and right actions of the target.

In this experiment, two targets stand in different positions. One target is stationary, and the other one can choose one of the four movement actions. The results of locating two targets at the same time are shown in Table 3 (Fig. 8):

Table 3. Comparison table of the influence of different actions on the experimental results

Index prerequisite	ADE (m)	MDE (m)
Static & Static	0.06	0.08
Static & Up	1.74	0.09
Static & Down	1.58	0.08
Static & Left	2.19	0.08
Static & Right	2.38	0.09

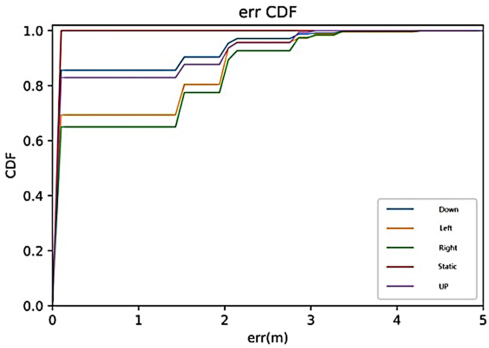

Fig. 8. The cumulative distribution function graph of the error of different action experiment results, the influence of different target actions on the experiment.

5 Conclusion

This paper proposes a new positioning method. First, we use the MUSIC method to process the signal, estimate the position parameter (AoA, ToF) of the positioning target and generate heatmaps. Then building a neural network to classify the images to achieve the purpose of positioning. In view of the feature degradation and poor effect of traditional deep neural networks for deep network structures, our neural network model effectively solves the feature degradation problem, enabling the neural network to reach deeper levels and better handle features of different scales. Experimental results show that the average error distance is only 8 cm and the median error distance is only 6 cm for simultaneous positioning of two persons in an indoor environment. In the future, we will study multi-person positioning for more than two persons. And to realize the trajectory tracking of the targets, so as to further improved the practical value of positioning under the Wi-Fi signal.

References

1. Li, Q.Y., et al.: AF-DCGAN: amplitude-feature deep convolutional GAN for fingerprint construction in indoor localization system. IEEE Trans. Emer. Top. Comput. Intell. **5**(3), 468–480 (2021)
2. Ding, Y.L., Sun, D.G., Yang, S.J.: Research on positioning technology of wireless sensor network based on ZigBee. In: Information Technology and Informatization, pp.187–188 (2021)
3. Lin, N., et al.: Contactless body movement recognition during sleeping via WiFi signal. IEEE Int. Things J. **7**(3), 2028–2037 (2020)
4. Gu, Y., Wang, Y.T., Liu, Z., Liu, J., Li, J.: SleepGuardian: an RF-based healthcare system guarding your sleep from Afar. IEEE Network **34**(2), 164–171 (2020)
5. Duan, P., Li, H., Zhang, B.: APFNet: Amplitude-Phase Fusion Network for CSI-Based Action Recognition. Mobile Netw. Appl. **26**, 2024–2034 (2021)
6. Gu, Y., Zhang, X., Liu, Z., Ren, F.J.: BeSense: leveraging WiFi channel data and computational intelligence for behavior analysis. IEEE Comput. Intell. Mag. **14**(4), 31–41 (2019)

7. Gu, Y., et al.: WiONE: one-shot learning for environment-robust device-free user authentication via commodity WiFi in man-machine system. IEEE Trans. Comput. Soc. Syst. **8**(3), 630–642 (2021)
8. Wang, C.: Research and implementation of Wi-Fi signal gesture recognition technology based on multi-modality. Beijing University of Posts and Telecommunications (2020)
9. Wang, X.: Research on Gesture Recognition Based on Improved EMA in Wi-Fi Environment. Beijing University of Posts and Telecommunications (2020)
10. Zhang, K.Q., Huang, Q.: Context-Aware Wireless Based Cross Domain Gesture Recognition. IEEE Int. Things J. (2021)
11. Cui, D., Zhang, Q.: The RFID data clustering algorithm for improving indoor network positioning based on LANDMARC technology. Clust. Comput. **22**(3), 5731–5738 (2017). https://doi.org/10.1007/s10586-017-1485-0
12. Zhang, K., Shen, C., Zhou, Q., Wang, H., Gao, Q., Chen, Y.: A combined GPS UWB and MARG locationing algorithm for indoor and outdoor mixed scenario. Clust. Comput. **22**(3), 5965–5974 (2018). https://doi.org/10.1007/s10586-018-1735-9
13. Kalbandhe, A.A., Patil, S.C.: Indoor positioning system using Bluetooth low energy. In: Proceedings of the 2016 International Conference on Computing, Analytics and Security Trends, pp. 451–455 (2016)
14. Paredes, J.A., Alvarez, F.J., Aguilera, T.: 3D indoor positioning of UAVs with spread spectrum ultrasound and time-of-flight cameras. Sensors **18**(1), 89 (2017)
15. Martin-gorostiza, E., Garcia-Garrido, M.A., Pizarro, D.: An indoor positioning approach based on fusion of cameras and infrared sensors. Sensors **19**(11), 2519 (2019)
16. Hahnloser, R.H.R., Sarpeshkar, R., Mahowald, M.A.: Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature **405**(6789), 947–951 (2020)
17. Wang, X., Gao, L., Mao, S., Pandey, S.: DeepFi: deep learning for indoor fingerprinting using channel state information. In: IEEE Wireless Communications and Networking Conference (WCNC), pp. 1666–1671 (2015)
18. Qian, K., Wu, C.S., Yang, Z., Liu, Y.H., Jamieson, K.: Widar: decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi. In: Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing. Association for Computing Machinery (2017)
19. Yin, Z., Jiang, Z., Yang, Z., Zhao, N., Chen, Y.: WUB-IP: a high-precision UWB positioning scheme for indoor multiuser applications. IEEE Syst. J. **13**(1), 279–288 (2019)
20. Ye, X.T., Zhang, Y., Song, J.D.: UWB indoor positioning algorithm based on attention mechanism. Computer Applications and Software, pp. 198–201,(2021)
21. Dang, X.C., Cao, Y, Hao, Z.J., Duan, Y.: A two-person positioning method based on CSI. J. Sens. Technol. (2019)
22. Wang, Y.Y., Chang, J., Wu, H.: Research on multi-parameter optimization of indoor WiFi positioning technology. Comput. Eng., 128–135 (2021)
23. Karanam, C.R., Korany, B., Mostofi, Y.: Tracking from One Side - Multi-Person Passive Tracking with WiFi Magnitude Measurements. In: 18th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), pp. 181–192 (2019)
24. Szegedy, C., Vanhoucke, V., Ioffe, S.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
25. He, K., Zhang, X., Ren, S.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations (2015)