Matched Averaging Federated Learning Gesture Recognition with WiFi Signals

Zexin Wang, Weidong Zhang, Xuangou Wu and Xiujun Wang
School of Computer Science and Technology, Anhui University of Technology
Anhui Engineering Laboratory for Intelligent Applications and Security of Industrial Internet
Maanshan, China

2021 7th International Conference on Big Data Computing and Communications (BigCom) | 978-1-6654-4252-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/BigCom53800.2021.00018

***Abstract*—WiFi signals based gesture recognition plays an important role in smart homes, such as human-computer interaction, identity authentication and so on. Most of existing approaches need a large number of channel state information dataset for model training to recognized human gestures. However, the robustness of the model to new users will be seriously affected, when the number of users participating in the training dataset is lacking. To address this problem, we propose a gesture recognition system based on matched averaging federated learning framework (WiMA). WiMA exploits parameter matching federated learning to training gesture prediction model, instead of traditional parameter aggregation. Experimental results show that the average accuracy of WiMA can be up to 90.4%, when 7 users participate in model training in 2 different rooms.**

***Index Terms*—Gesture recognition, channel state information, federated learning**

I. Introduction

With the advent of the Internet of Things era, Wi-Fi access points are widely deployed in home environment. Human gesture recognition based on Wi-Fi signals is attracting more and more attentions which can be applied to human-computer interaction applications, identity authentication and so on. Wi-Fi based sensing has been widely concerned because of its low cost and easy deployment. Meanwhile, the subcarrier-level sensing information can be obtained from channel state information (CSI). Through CSI information, many Wi-Fi signal based applications have been proposed, such as human activity recognition (e.g. [1]–[6]), user authentication (e.g. [7]–[10]), indoor localization(e.g. [11]–[13]) and intrusion detection (e.g. [14]–[16]).

Most of the existing approaches exploit the collected CSI training dataset to build a human gesture classification model. In [17], Y Zheng *et.al* proposed Body-coordinate Velocity Profile (BVP) feature to realize cross-domain gesture recognition, which can improve the robustness of the model to the environment. Wi-SL [18] established the correlation mapping between the amplitude and phase difference information in the wireless signal and the sign language action. WiGAN [19] used Generative Adversarial Network to extract and generate gesture features, which shows the robustness under different experimental environments and different users. Since the robustness of model depends on the user diversity of the training dataset, these approaches require a large number of different users participating in training dataset to improve the model robustness for new users. However, user's diversity is influenced by the number of users in the smart homes scenario. When the number of users participating in the training dataset is lacking, the prediction accuracy of model will be seriously affected for new users.

To address the above problem, we propose a Wi-Fi gesture recognition system (WiMA) based on matched averaging federated learning, which exploits parameter matching fusion algorithm instead of the traditional Fedavg federated learning algorithm. Fedavg algorithm has an adverse impact on local user diversity. From the experiment results, when the number of users increases, the accuracy of local model recognition does not improve with increasing the number of users. WiMA is based on the invariance of neuron displacement in the neural network, match local model parameters to improve the local user diversity.

The main contributions of this paper can be summarized as follows:

- We propose cross-local gesture recognition based on federated migration learning. As far as we know, this is the first user robust robustness gesture recognition system with limited users in the training dataset.
- To realize user-robust gesture recognition, we build a deep learning local model, and then fuse the parameters between the local models by matched averaging algorithm.
- Our experimental results show that the accuracy of global model based on the matching average algorithm is better than the traditional federated learning algorithm. When the total number of users in the two local training datasets reaches 7, the gesture recognition average accuracy of the two local users can reach 90.4%.

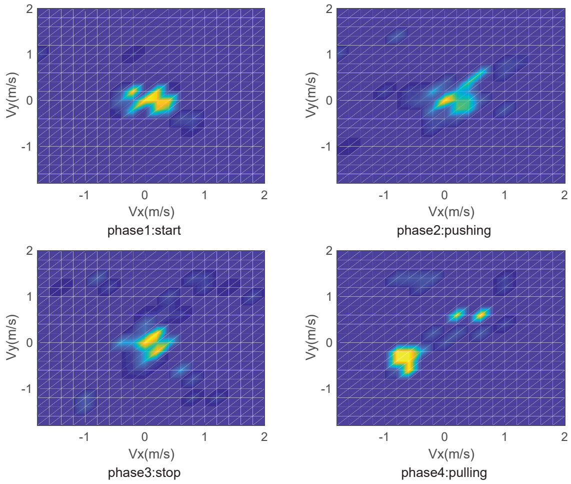
The rest of this paper is organized as follows: In Section II, the background of wireless and motivation is described. The system overview is given in Section III. The local model design and parameter fusion are presented in Section IV in detail. We evaluate the performance of WiMA in Section V. Finally, we summarize the related work and make a conclusion in VI and VII, respectively.

II. Background and Motivation

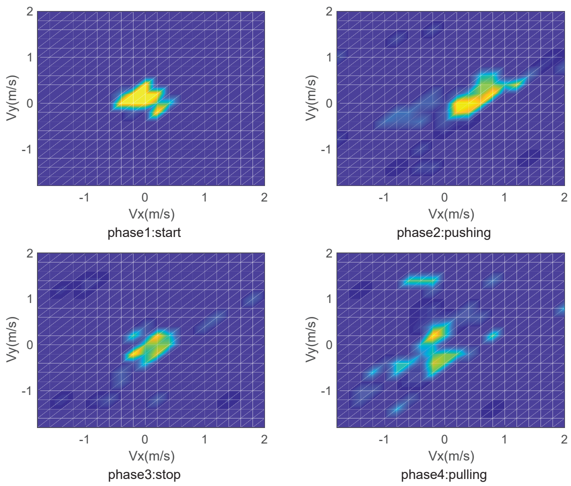
A. CSI and BVP

In frequency-division multiplexing (OFDM) systems, S subcarriers represented by complex values can be collected from

978-1-6654-4252-7/21/$31.00 ©2021 IEEE
DOI 10.1109/BigCom53800.2021.00018



(a) BVP series of user1



(b) BVP series of user2

Fig. 1: BVP series for different user.,

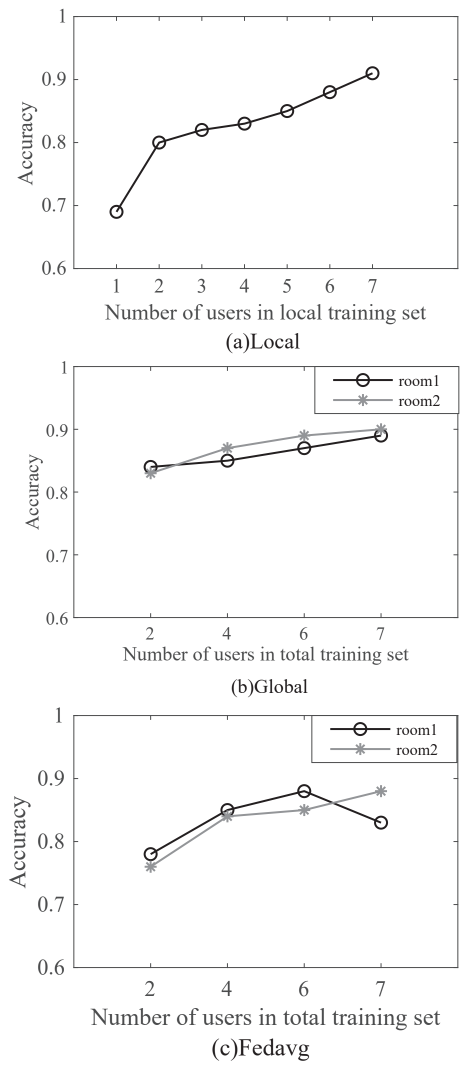


Fig. 2: The accuracy of model with different number of users.

each data packet by using current commercial Wi-Fi devices. CSI can be defined as:

$$H(f_k,t) = |H(f_k,t)|e^{j\angle H(f_k,t)}, k \in [1,S] \tag{1}$$

where $|H(f_k,t)|$ and $\angle H(f_k,t)$ represent the subcarrier f_k as the center frequency, the tth timestamp CSI values of the amplitude and phase respectively.

The relative motion between the transmitter and the receiver causes Doppler Frequency Shift (DFS) [20]. According to CARM, the root reason that leads to DFS is the change of signal propagation path. The frequency shift which results from the reflected signal generated can be written as:

$$f_D(t) = -\frac{1}{\lambda}\frac{\mathrm{d}}{\mathrm{d}t}d(t) = -f\frac{\mathrm{d}}{\mathrm{d}t}\tau(t), \tag{2}$$

where λ,f,$\tau(t)$ corresponds to the wavelength of the signal, the subcarrier frequency, and the flight time of the signal, and $d(t)$ is the distance of the NLOS path.

When users perform gestures, in addition to the body parts of the body movement can produce different speed, these mobiles can cause the relative nonnegligible motion of the DFS. Assuming accumulation caused by doppler frequency shift of the velocity vector for $\vec{v}$, in each timestamp note k transceiver link corresponding to the doppler frequency shift as $F_D^k(\vec{v})$:

$$F_D^k(\vec{v}) = c_x^k \vec{v_x} + c_y^k \vec{v_y}, \tag{3}$$

where c_x^k and c_y^k are determined by the location of the corresponding transceiver link and the user. Derived from $\vec{v}$, $\vec{v_x}$ refers to the user's face orientation and $\vec{v_y}$ refers to the vertical direction [21]. Therefore,c_x^k and c_y^k can be used to solve the possible values of $\vec{v_x}$ and $\vec{v_y}$, calculate $F_D^k(\vec{v})$, and solve the optimal solution of $\vec{v_x}$ and $\vec{v_y}$ with the measurement DFS isolated from the CSI measurement [17]. Body-coordinate velocity profile (BVP) can be represented by $\vec{v_x}$ and $\vec{v_y}$. Different users perform the same action with different patterns. Take push and pull as an example, as shown in Fig. 1, different users at different phases of the same gesture have

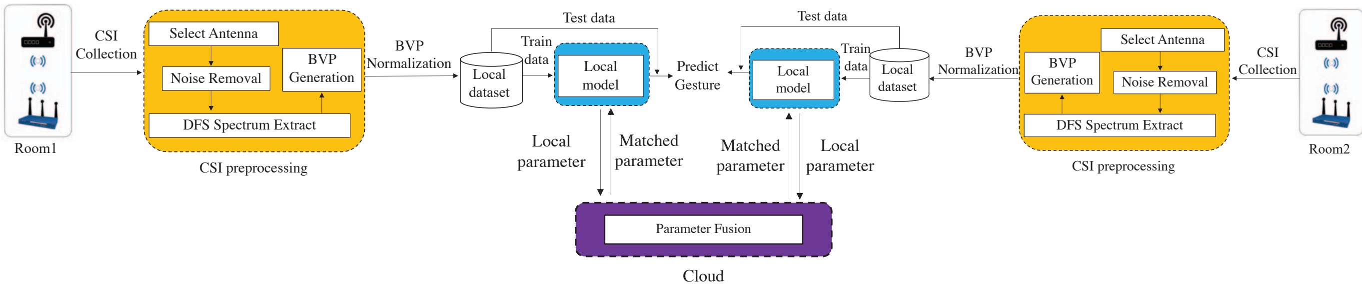


Fig. 3: WiMA System Architecture.

different power distribution of speed components and different execution duration.

B. Motivation

By using BVP of local users to predict new users, with the increase the number of users, as shown in Fig. 2(a), when there are more than 4 users, the test average accuracy of new users can be guaranteed to be above 85%. However, when few local users are involved in building the training dataset of the model, the recognition accuracy for new users is not high. The reason is that .local datasets' heterogeneity is affected when the number of local users is limited. In other words, physiological characteristics such as height and weight of users have diverse influences on Wi-Fi signals. When local data resources are limited, as shown in Fig. 2(b), the models established by global aggregation of different local sides are used to predict new users in two rooms respectively. When the number of users in the global training set reaches 6, the average accuracy of both rooms can reach 87%. Using the traditional federated learning Fedavg algorithm under federated learning, as shown in Fig. 2(c), the average accuracy of one room is steadily improving, while the accuracy of the other room is declining. This phenomenon indicates that the average weight obtained by FedAvg may have an adverse effect on the performance of the global model, which is due to the permutation invariance of neural network parameters. For any given neural network, there are many parameters in its neurons. The order of parameters is different, so Fedavg ignores this point and averages directly according to the inherent positions of neurons in the neural network.

Aiming at the limited gesture recognition of local users, when there is a deviation between the data of new users and users in the training dataset, aggregating all local data can improve the accuracy of local testing for new users. For the need of privacy protection, users tend to keep their data out of the local area and use different local data to build model parameter fusion to realize data fusion. The traditional federated learning Fedavg algorithm will lead to a decrease in recognition accuracy in some local sides. To solve this problem, we need to use the permutation invariance of parameters in the neural network to overcome the global model's adverse impact on local areas under Fedavg.

III. Overview

To train an efficient client federated learning model, we divide the WiMA system into four blocks as shown in Fig. 3, CSI preprocessing block, BVP normalization block, model building block, parameter fusion block, and gesture recognition block. CSI preprocessing block extracts DFS from collected raw CSI measurements and generates BVP from DFS spectrum. BVP normalization block is designed to standardize the BVP series data to generate local datasets. Model building block is to construct local models with train data from local datasets. Parameter fusion block is to match and fuse the local models' parameters and return match parameters to local models. The gesture recognition block is responsible for distinguishing different user gestures with local models with matched parameters.

IV. Design

A. CSI preprocessing

According to IndoTrack [20], the transmitting antenna and the two receiving antennas of the CSI amplitude conjugate multiplication are used to eliminate the quasi-static offset. The band-pass filter is used to filter out-of-band noise, which can remove the random offset. Therefore, To preserve non-zero DFS with gaining multipath components, two receiving antennas need to be selected. Widance [22] studied the influence of different antennas on the dynamic pathes. The influence is characterized by the variance of different antenna pairs in CSI amplitude, which can be used to extract DFS spectrum and generate BVP.

B. BVP normalization

For the obtained BVP series, we need to normalize the BVP series. Durations of different BVP series samples are not uniform, it is necessary to up-sampling and fix the duration of all samples and normalize all elements in the BVP series. In this way, it can be ensured that the BVP series is only related to user gestures.

C. Model building

The local side has the same model structure. Different local data sets are used to train model parameters, and the cloud side uses local parameters for parameter fusion. Every BVP series data can be seen as a picture series composed of pictures

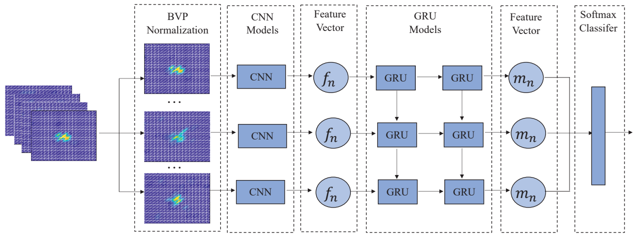


Fig. 4: Local model structure.

depicting velocity components distribution. Each BVP profile depicts how the energy of the user performing a certain gesture is distributed. Using a convolutional neural network (CNN) as the spatial feature extractor, CNN is an effective method that can automatically learn parameters and features for complex image problems. Therefore, CNN is very suitable for feature extraction for each BVP profile.

Furthermore, since the BVP series is timing, it is necessary to introduce a recurrent neural network (RNN) to describe the dynamic characteristics of the moment. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are common RNN types of RNN. Compared with LSTM, we use GRU because GRU can use fewer parameters and obtain quite accurate compared with LSTM.

As shown in Fig. 4, the complete network structure of the local model is two 3x3 convolutional blocks, 2x2 maximum pooling layer, two fully connected layers for each BVP profile. Then input the GRU block and expand the GRU output and the dropout layer is introduced to prevent the model from overfitting. Then expand the input to the fully connected layer classifier and finally, the softmax classifier based on the cross-entropy loss function obtains the prediction result.

D. Parameter fusion and gesture recognition

For the parameters trained by local models, we propose a federated matching algorithm, whose core is to introduce a permutation matrix to realize the permutation invariance of neurons in the neural network. The simplest single-layer fully connected network can be formulated as $y = \sum_{i=1}^{L} W_{2,i}\sigma(< x, W_{1,i} >)$, and L is the number of neurons in the hidden layer. Therefore, there are total L! parameter arrangements for W_1, W_2. Further,

$$Y = \sigma(xW_1)W_2 = \sigma(xW_1\Omega)\Omega^{\mathrm{T}}W_2, \tag{4}$$

where Ω is any $L \times L$ permutation matrix.For two of the same size datasets $X_j, X_{j'}$,weight is obtained by training for $W_1\Omega_j, {\Omega_j}^{\mathrm{T}}W_2$ and $W_1\Omega_{j'}, {\Omega_{j'}}^{\mathrm{T}}W_2$, obviously, most likely $W_1\Omega_j \neq W_1\Omega_{j'}$ and $(W_1\Omega_j + W_1\Omega_{j'})/2 \neq W_1\Omega_j$ for any Ω. Therefore, the first thing to restore replacement $(W_1\Omega_j{\Omega_{j'}}^{\mathrm{T}} + W_1\Omega_{j'}{\Omega_{j'}}^{\mathrm{T}})/2 \neq W_1$. Suppose Wjl is the lth neuron learned on dataset X_j, θ_i represents ith neuron in the global model, and $c(\cdot)$ is defined as an appropriate similarity function between a pair of neurons. The permutation optimization problem can be defined as follows:

$$\min_{\tau^j_{li}} \sum_{i=1}^{L} \sum_{j,l} \min_{\theta i} \tau^j_{li} \cdot c(W_{jl}, \theta_i)$$
$$s.t. \sum_{i} \tau^j_{li} = 1; \sum_{i} \tau^j_{li} = 1 \tag{5}$$

$\Omega^{\mathrm{T}}_{jli} = \tau^j_{li}$ and the weight of a specific provide j th local $\{W_{j,1}, W_{j,2}\}_{j=1}^{J}$ provided by J local sides, we calculate the federated neural network weights $W_1 = \frac{1}{J}\sum_j W_{j,1}{\Omega_j}^{\mathrm{T}}$ and $W_2 = \frac{1}{J}\sum_j {\Omega_{j'}}^{\mathrm{T}} W_{j,2}$. To solve (5) ,we apply Hungarian matching algorithm [23] and BBP_MAP algorithm [24]. For the solving process in detail, please refer to FEDMA [25].

V. Evaluation

A. Experiment Setup

We use the public dataset Widar3.0 [17], which contains 9 gestures by 16 users. The Wi-Fi receiver has three antennas and records the original CSI measurement at a sampling rate of 1000 Hz. The size of the CSI measurement included in the package is $1 \times 3 \times 30$, where 1 represents one transmit antenna, and 3 represents 3 receive antennas. We select 6 gestures performed by 12 users in 2 rooms as a dataset. In WiMA, we assume two rooms as two local sides. We randomly select one user locally as the test user, and then randomly select a specified number of users from the remaining users as the local training dataset. We implement WiMA in MATLAB and Keras. In these experiments, training and testing are performed by a Linux desktop with GPU GeForce GTX 1080.

B. Experimental Analysis

Accuracy: From the experimental results, it can be seen that when the number of users in the two local training datasets reaches 7, as shown in Fig. 5 and Fig. 6 , the recognition average accuracy of the two local sides for new users can both reach 90%, and the local federated model has no adverse effect on the recognition effect of the local side. Furthermore, we can find with the recognition accuracy using the MA algorithm is equivalent to the model established by data fusion. With the increase of the number of users in the training dataset, the recognition accuracy tends to rise steadily.

The number of users: There are 3 users in the training dataset of each local side, the local federated model can identify new users with an average accuracy of 87%, which is better than the 85% accuracy achieved by modeling with at least 4 users for local training. Therefore, the matching average federated learning model can have a more relaxed requirement on the number of users participating in the construction of local datasets and achieve reliable accuracy.

VI. Related Work

In this section, we introduce the related work on human gesture recognition. Gesture recognition is an emerging research topic with various applications. We summarize existing work

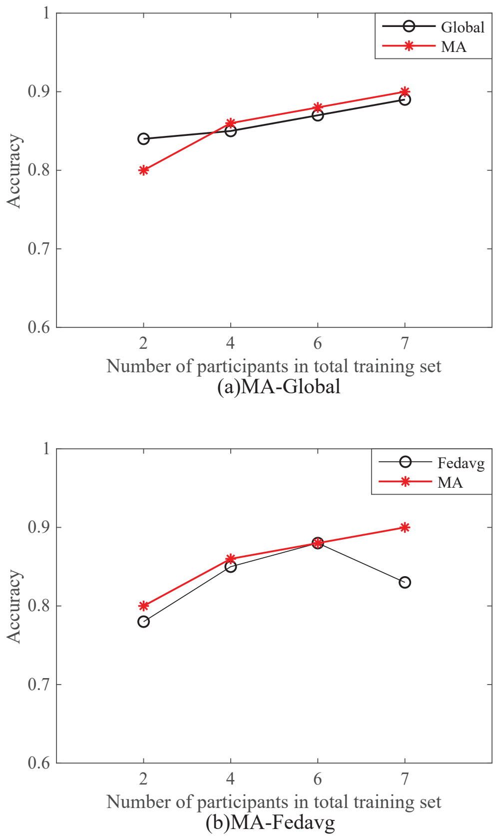


Fig. 5: Comparison results with Room1.

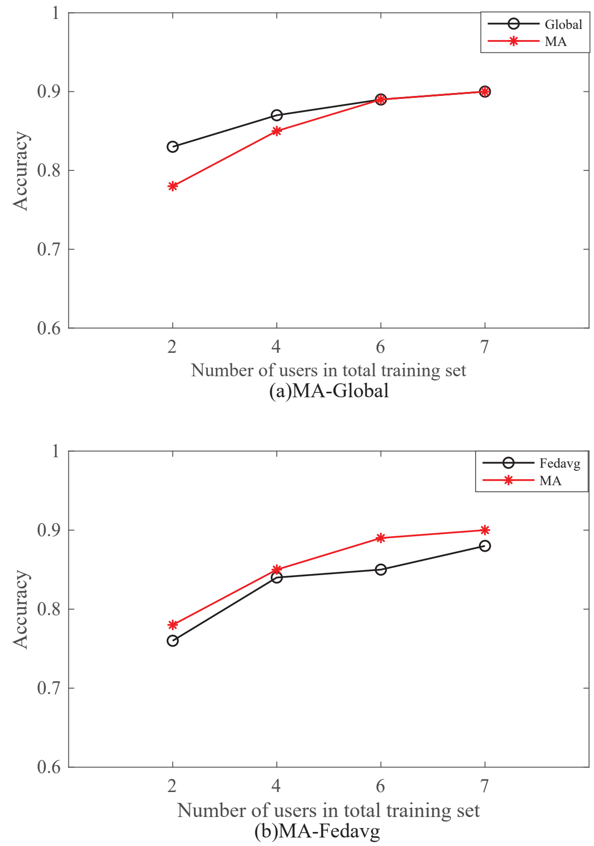


Fig. 6: Comparison results with Room2.

into two categories. (i) Human gesture recognition and (ii) CSI-based activity recognition.

Human activity recognition. Human gesture recognition methods adopt wearable sensors or body contact devices [26]–[29] and computer vision based systems [30]–[32]. WiSee [26] was first of its kind experiment done with Doppler shift for identifying and recognizing the human gestures using Universal Software Radio Peripheral (USRP) device. Chin-Shyurng *et. al* [28] utilized Kinect perform accurate tracking and recognition by capturing the target image from different angles. H Foroughi et al. [32] proposed a novel method to detect various posture-based events, which uses video surveillance to monitor the elderly at home.

CSI-based activity recognition. In recent years, many CSI- based systems have been proposed to recognize human gestures(e.g. [2], [18], [19], [33]–[36]). These existing human gesture recognition systems require sufficient users participate help construct local training dataset. For example, CARM [2] introduced a model of multipath velocity and CSI to realize the recognition of human activity. WiGeR [33] realized gesture recognition when the transceiver LOS link is blocked by walls, WiSign [35] can effectively reduce the false alarm rate of gesture recognition through a transmitter and two receivers, and TW-see [34] realized motion recognition through walls by separating environmental noise and motion components. Wi-SL [18] established the correlation mapping between the amplitude and phase difference information in the wireless signal and the sign language action. WiGAN [19] used Generative Adversarial Network to extract and generate gesture features, which shows the robustness under different experimental environments and different users.

Although these systems have the advantages of passive detection and easy deployment, enough samples are required to be collected in their datasets and they all follow the traditional data planning division method, dividing the whole dataset randomly according to the proportion. In other words, the data bias between the training dataset and the test dataset will not affect the results.

VII. Conclusion

In this paper, we proposed WiMA, a gesture recognition system based on the framework of federated learning. WiMA can solve gesture recognition with limited local data. To solve the adverse effect of the traditional federated learning Fedavg algorithm, WiMA designed a matching average algorithm based on the permutation invariance of neurons in the neural network to improve the robustness of the model and maximizes the use of parameters to learn the characteristics of local datasets. Our results show that when the number of local training users reaches 3, the average accuracy of gesture recognition for new local users can reach 87.2% and when the total number of users in the two local training datasets reaches 7, the gesture recognition average accuracy of the two local users can reach 90.4%.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grants (61672038), the Key Research and Development Project of Anhui Province (Grant No. 202004a05020009).

REFERENCES

[1] R. H. Venkatnarayan, G. Page, and M. Shahzad, "Multi-user gesture recognition using wifi," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, 2018, pp. 401–413.

[2] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of wifi signal based human activity recognition," in *Proceedings of the 21st annual international conference on mobile computing and networking*, 2015, pp. 65–76.

[3] N. Golestani and M. Moghaddam, "Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks," *Nature communications*, vol. 11, no. 1, pp. 1–11, 2020.

[4] J. Ding and Y. Wang, "A wifi-based smart home fall detection system using recurrent neural network," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 4, pp. 308–317, 2020.

[5] S. Palipana, D. Rojas, P. Agrawal, and D. Pesch, "venkatnarayan2018multivenkatnarayan2018multi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–25, 2018.

[6] T. Nakamura, M. Bouazizi, K. Yamamoto, and T. Ohtsuki, "Wi-fi-csi-based fall detection by spectrogram analysis with cnn," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

[7] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 363–373.

[8] C. Shi, J. Liu, H. Liu, and Y. Chen, "Smart user authentication through actuation of daily activities leveraging wifi-enabled iot," in *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2017, pp. 1–10.

[9] H. Kong, L. Lu, J. Yu, Y. Chen, L. Kong, and M. Li, "Fingerpass: Finger gesture-based continuous user authentication for smart homes using commodity wifi," in *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2019, pp. 201–210.

[10] C. Shi, J. Liu, N. Borodinov, B. Leao, and Y. Chen, "Towards environment-independent behavior-based user authentication using wifi," in *2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*. IEEE, 2020, pp. 666–674.

[11] A. Niitsoo, T. Edelhäβer, and C. Mutschler, "Convolutional neural networks for position estimation in tdoa-based locating systems," in *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2018, pp. 1–8.

[12] W. Cui, Q. Liu, L. Zhang, H. Wang, X. Lu, and J. Li, "A robust mobile robot indoor positioning system based on wi-fi," *International Journal of Advanced Robotic Systems*, vol. 17, no. 1, p. 1729881419896660, 2020.

[13] L. Chen, "Enhancing indoor location fingerprinting using channel state information," Ph.D. dissertation, Conservatoire national des arts et metiers-CNAM, 2020.

[14] X. Li, D. Zhang, J. Xiong, Y. Zhang, S. Li, Y. Wang, and H. Mei, "Training-free human vitality monitoring using commodity wi-fi devices," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–25, 2018.

[15] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-invasive detection of moving and stationary human with wifi," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2329–2342, 2015.

[16] F. Li, M. Valero, H. Shahriar, R. A. Khan, and S. I. Ahamed, "Wi-covid: A covid-19 symptom detection and patient monitoring framework using wifi," *Smart Health*, vol. 19, p. 100147, 2021.

[17] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Zero-effort cross-domain gesture recognition with wi-fi," in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, 2019, pp. 313–325.

[18] Z. Hao, Y. Duan, X. Dang, Y. Liu, and D. Zhang, "Wi-sl: Contactless fine-grained gesture recognition uses channel state information," *Sensors*, vol. 20, no. 14, p. 4025, 2020.

[19] D. Jiang, M. Li, and C. Xu, "Wigan: A wifi based gesture recognition system with gans," *Sensors*, vol. 20, no. 17, p. 4757, 2020.

[20] X. Li, S. Li, D. Zhang, J. Xiong, Y. Wang, and H. Mei, "Dynamic-music: accurate device-free indoor localization," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 196–207.

[21] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, "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*, 2017, pp. 1–10.

[22] K. Qian, C. Wu, Z. Zhou, Y. Zheng, Z. Yang, and Y. Liu, "Inferring motion direction using commodity wi-fi for interactive exergames," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 1961–1972.

[23] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[24] R. Thibaux and M. I. Jordan, "Hierarchical beta processes and the indian buffet process," in *Artificial Intelligence and Statistics*. PMLR, 2007, pp. 564–571.

[25] H. Wang, M. Yurochkin, Y. Sun, D. Papailiopoulos, and Y. Khazaeni, "Federated learning with matched averaging," *arXiv preprint arXiv:2002.06440*, 2020.

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

[27] M. L. Gavrilova, Y. Wang, F. Ahmed, and P. P. Paul, "Kinect sensor gesture and activity recognition: New applications for consumer cognitive systems," *IEEE Consumer Electronics Magazine*, vol. 7, no. 1, pp. 88–94, 2017.

[28] F. Chin-Shyurng, S.-E. Lee, and M.-L. Wu, "Real-time musical conducting gesture recognition based on a dynamic time warping classifier using a single-depth camera," *Applied Sciences*, vol. 9, no. 3, p. 528, 2019.

[29] G. Bhat, R. Deb, V. V. Chaurasia, H. Shill, and U. Y. Ogras, "Online human activity recognition using low-power wearable devices," in *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. IEEE, 2018, pp. 1–8.

[30] L. Ding, W. Fang, H. Luo, P. E. Love, B. Zhong, and X. Ouyang, "A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory," *Automation in construction*, vol. 86, pp. 118–124, 2018.

[31] K. Kim, J. Kim, J. Choi, J. Kim, and S. Lee, "Depth camera-based 3d hand gesture controls with immersive tactile feedback for natural mid-air gesture interactions," *Sensors*, vol. 15, no. 1, pp. 1022–1046, 2015.

[32] H. Foroughi, B. S. Aski, and H. Pourreza, "Intelligent video surveillance for monitoring fall detection of elderly in home environments," in *Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on*, 2009.

[33] M. A. A. Al-qaness and F. Li, "Wiger: Wifi-based gesture recognition system," *ISPRS International Journal of Geo-Information*, vol. 5, no. 6, p. 92, 2016.

[34] X. Wu, Z. Chu, P. Yang, C. Xiang, X. Zheng, and W. Huang, "Tw-see: Human activity recognition through the wall with commodity wi-fi devices," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 306–319, 2018.

[35] J. Shang and J. Wu, "A robust sign language recognition system with multiple wi-fi devices," in *Proceedings of the Workshop on Mobility in the Evolving Internet Architecture*, 2017, pp. 19–24.

[36] S. Tan, J. Yang, and Y. Chen, "Enabling fine-grained finger gesture recognition on commodity wifi devices," *IEEE Transactions on Mobile Computing*, 2020.