

Crowd Counting Based on WiFi Channel State Information and Transfer Learning

Zhongqiang Wu, Pengyu Ji, Manxi Ma, Wencheng Zhuang, Zhenyu Li, Jihao Cui, Zhengjie Wang*

College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao, China
Email: 1463175858@qq.com, 1925290292@qq.com, 2772692129@qq.com, 2831599729@qq.com, 1245689084@qq.com, 1228054435@qq.com, *cieewangzj@163.com

How to cite this paper: Wu, Z.Q., Ji, P.Y., Ma, M.X., Zhuang, W.C., Li, Z.Y., Cui, J.H. and Wang, Z.J. (2022) Crowd Counting Based on WiFi Channel State Information and Transfer Learning. *Journal of Computer and Communications*, 10, 22-36.
<https://doi.org/10.4236/jcc.2022.106003>

Received: May 4, 2022

Accepted: June 12, 2022

Published: June 15, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

With the popularity and development of indoor WiFi equipment, they have more sensing capability and can be used as a human monitoring device. We can collect the channel state information (CSI) from WiFi device and acquire the human state based on the measurements. These studies have attracted wide attention and become a hot research topic. This paper concentrated on the crowd counting based on CSI and transfer learning. We utilized the CSI signal fluctuations caused by human motion in WiFi coverage to identify the person count because different person counts would lead to unique signal propagation characteristics. First, this paper presented recent studies of crowd counting based on CSI. Then, we introduced the basic concept of CSI, and described the fundamental principle of CSI-based crowd counting. We also presented the system framework, experiment scenario, and neural network structure transferred from the ResNet. Next, we presented the experiment results and compared the accuracy using different neural network models. The system achieved recognition accuracy of this 100 percent for seven participants using the transfer learning technique. Finally, we concluded the paper by discussing the current problems and future work.

Keywords

Crowd Counting, Channel Status Information, Transfer Learning, ResNet

1. Introduction

In recent years, with the popularity of WiFi devices and the development of wireless sensing technology, device-free human activity identification research has achieved some results because WiFi devices provide some environment-sensing capability. For example, we can implement much human behavior recognition,

including daily behavior recognition [1] [2] [3] [4], fall detection [5] [6] [7], hand recognition [8] [9], and human state monitoring [10] [11] [12] [13]. The sensing technique can effectively identify the human state in the region and provide more information for intelligence applications. These CSI applications have many advantages over traditional technologies requiring the subject to wear sensors actively. We focus on crowd counting with WiFi CSI in device-free human sensing techniques. The following sections divide the crowd count into two categories according to whether they use WiFi signals. We present the current crowd count progress based on non-WiFi devices and WiFi signals.

1) Crowd counting based on non-WiFi signals

There are many different kinds of crowd counting methods based on non-WiFi signals. Here, we investigate some typical signals used to implement crowd counting, including video, light, sound, and ultra-wide bandwidth (UWB) signals. The traditional crowd counting methods are image-based and have achieved many research results. An extensive review of video-based demographics and behavioral understanding by Grant [14] *et al.* has investigated the computational methods of population number and density. It summarizes the datasets used for population activity video perception. With the continuous development of technology, light-based population detection has also attracted the attention of researchers. Aubida A. Al-Hameed [15] *et al.* first proposed LiDal, an indoor light detection radar system for personnel counting and positioning. The system uses multiple transmitters and receivers to capture visible light reflected from the target's body and uses light-emitting diodes or lasers to encode the data as light intensity. The experimental results prove that the system can decrease errors in an indoor environment with 15 people. Oliver Shih [16] *et al.* propose an effective sensing scheme that utilizes fluctuation in a room's acoustic features to calculate the number of participants. A dense crowd counting scheme based on the CTF-DBF mixed feature extraction by the IR-UWB radar was proposed by Xiuzhu Yang [17] *et al.* In the case of 20 people, the highest recognition accuracy is above 97%.

2) Crowd counting based on WiFi signals

Several non-WiFi population counting methods have many privacy, cost, deployment, and other issues, limiting their wide application. With the extensive deployment of wireless networks, the crowd counting method based on WiFi signal has low cost, convenient deployment, privacy protection and many other advantages, making it has a wide range of application prospects. At present, two typical wireless signals, including received signal strength (RSS) and channel state information (CSI), are used in WiFi-based population counting applications. This paper mainly focuses on the application of the CSI.

Most crowd counting based on WiFi CSI is mainly divided into pattern-based and deep learning methods. The pattern-based approach establishes the one-to-one correspondence between the signal change mode and the number of human activities by finding the transformation law of CSI signals and constructing fea-

tures to identify the number of populations. Han Zou [18] *et al.*, in 2018, proposed WiFree, a device-free WiFi detection scheme using the commercial Internet of Things devices. It implemented occupancy detection and crowd counting. It proposed a scheme to model human motion using information theory. Experiments show that the proposed system achieves an average population accuracy of 92.8%.

With the development of artificial intelligence, crowd counting based on CSI and deep learning has become a hot research topic. It can extract CSI features from a given data sample and build the relationship between signal variation and the number of the person through neural networks. Compared with the machine learning methods, the deep learning method has the advantages of automatically extracting behavioral features and implementing high accuracy. Typical neural networks include convolutional neural network (CNN), long short-term memory (LSTM). Liu [19] *et al.* first proposed an algorithm for crowd counting in a WiFi environment named DeepCount. The system automatically extracts features using both CNN and LSTM. The model identification accuracy reached 90% in the 5-person experimental scenario. Zhou [20] *et al.* in 2020 proposed a scheme based on WiFi CSI using only a pair of transceivers to achieve crowd counting. This scheme uses a DNN regression model to infer the number of people from the changes in real-time CSI. Experiments show that the scheme achieves a counting accuracy of 100% for six people, an accuracy of 97.7% for 34 people, and an accuracy of 99.3% for two people.

Although the above studies have achieved good system performance, these studies usually employed a specific neural network model. At the same time, many deep learning models have been proved very effective for computer vision applications. Therefore, transfer learning for the CSI application is a hot research topic. This paper focuses on transfer learning for crowd counting using CSI. Specifically, we used the ResNet for crowd counting using the CSI signal.

The contribution of this paper can be summarized as follows. We proposed a crowd counting system using CSI and the transfer learning technique. It achieved 100 percent for seven participants just using the standard ResNet model. This result proves that transfer learning can be utilized to implement crowd counting using the CSI signal. It also indicates that transfer learning has a broad application perspective about human behavior recognition using the traditional transfer learning model.

2. Methodology

2.1. CSI Introduction

Channel state information is mainly designed to measure the channel properties of the propagation link, describing the factors in each propagation path. The CSI can be depicted as a complex value matrix H , and it is affected by many factors, such as signal scattering, environmental weakness (fading, multipath fading or shadowing fading), distance decay (power decay of distance) and other informa-

tion. In the frequency domain, the channel can be described as:

$$Y = H \times X + N \quad (1)$$

In the Formula (1), X represents the transmit signal vector; Y is the received signal vector; H denotes the signal matrix and N describes the additive Gaussian white noise. The CSI can be classified into different subcarrier groups based on the drive of the receiving device. The subcarrier of the channel can be explained as:

$$H_i = \|H_i\| e^{j\sin(\angle H_i)} \quad (2)$$

In the Formula (2), $\|H_i\|$ is the amplitude of the i th subcarrier, indicating the WiFi signal amplitude; the angle is the phase of the i th subcarrier, indicating the periodic change of the signal. Therefore, the amplitude information and phase information are used to describe the channel state and can be utilized for real environmental sensing.

2.2. Framework of the System

The basic principle of crowd counting using WiFi CSI can be explained as follows. When the person passes through the range of WiFi signal coverage, the propagation path of WiFi signal will undergo some complex changes. The crowd count can be recognized by analyzing the relationship between the CSI fluctuations and the number of a crowd [19].

Next, we describe the system framework. The WiFi CSI-based crowd counting system mainly consists of three parts, including experimental data collection, CSI data processing, and population number identification, as shown in **Figure 1**. When the PC and AP can communicate normally, the CSI data can be collected in the notebook with the Intel 5300 network card by modifying the driver

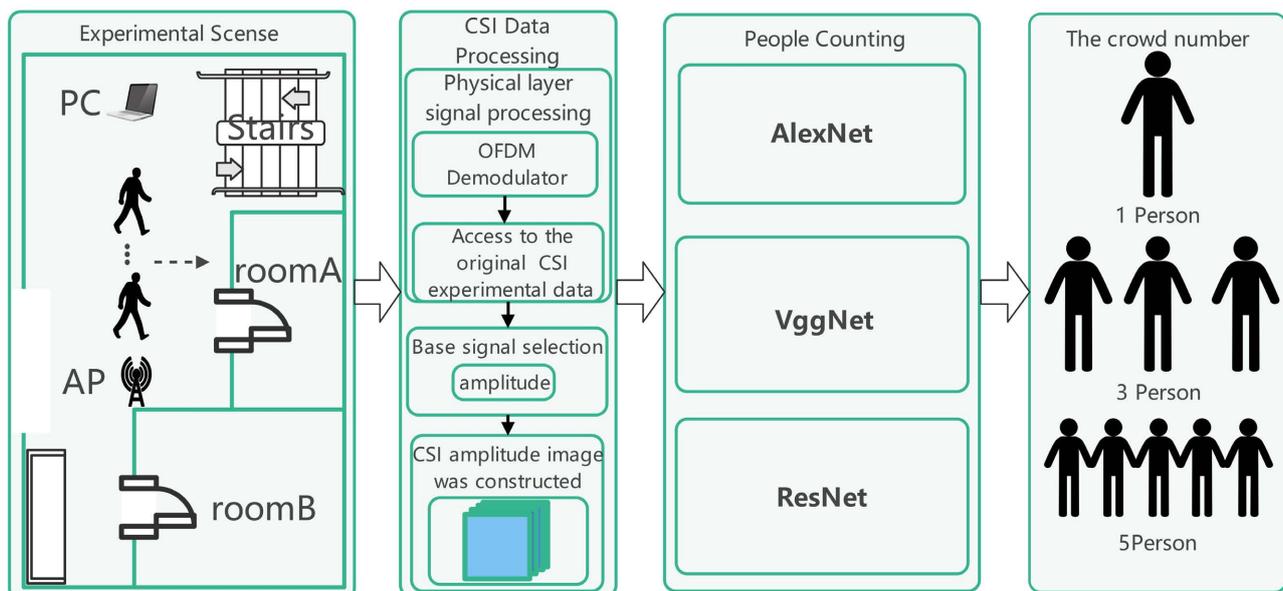


Figure 1. The framework of WiFi CSI-based people counting.

of the Intel 5300 network card. After collecting the raw CSI data, the amplitudes are selected as the base signal and normalized for the analysis. The amplitude information of the CSI is then converted into a spectral feature map. Finally, the feature map containing the population quantity information is input into the three designed neural networks (AlexNet [21], VggNet [22], ResNet [23]). By training, checking and testing the network model, the identification of seven populations (1 - 7 people) is finally realized.

Finally, we describe the experiment devices. We use JCG-134E30 wireless router as the transmitter of WiFi signal and the Lenovo E420 laptop that contains a built-in Intel 5300 network card. After collecting the original CSI data, we choose the amplitude as the base signal because it will generate severe fluctuations when the number of people changes, which is crucial for signal measurements.

2.3. Transfer Learning and Neural Network

The concept of deep learning stems from the research of artificial neural networks. At the same time, deep learning is also recognized as a kind of machine learning. Unlike traditional machine algorithms, this algorithm mainly realizes data classification by training neural network models.

Transfer learning is a technique in which we can transfer the model trained to other datasets that have never been seen. Specifically, we can transfer the learning of the model with its training parameters on a generic dataset to our data sample. For example, we can use the computer vision model trained on millions of images on the CSI data.

This paper utilizes the transfer learning technique to implement human crowd counting. Specifically, we employ ResNet151 to fulfill the crowd counting using CSI. We also compare the model with other typical neural network models, such as AlexNet and VggNet, to validate the effectiveness of the selection. Next, we introduce the three network models (AlexNet, VggNet and ResNet) used in this paper.

The basic structure of AlexNet is shown in **Figure 2**. The AlexNet consists of 8 layers, the first five layers are convolutional, and the last three layers are full-connected. If any convolution layer is removed, the classification performance of the network is greatly reduced.

The name of VggNet is derived from the study group name (Visual Geometry Group), and its structure is shown in **Figure 3**. Most of the whole network use convolution filters of $3 * 3$ and max-pooling of $2 * 2$ to deepen the network depth. The VGG 19 is used in this paper and includes 19 hidden layers of 16 convolutional layers and three fully connected layers, respectively.

ResNet [23] (Residual Net) solves the degradation problem of neural networks by adding the residues blocks. The basic structure of ResNet can be described as the component of some typical network modules [24], as shown in **Figure 4**. The core of ResNet is that each stack layer adopts residual learning. The crucial

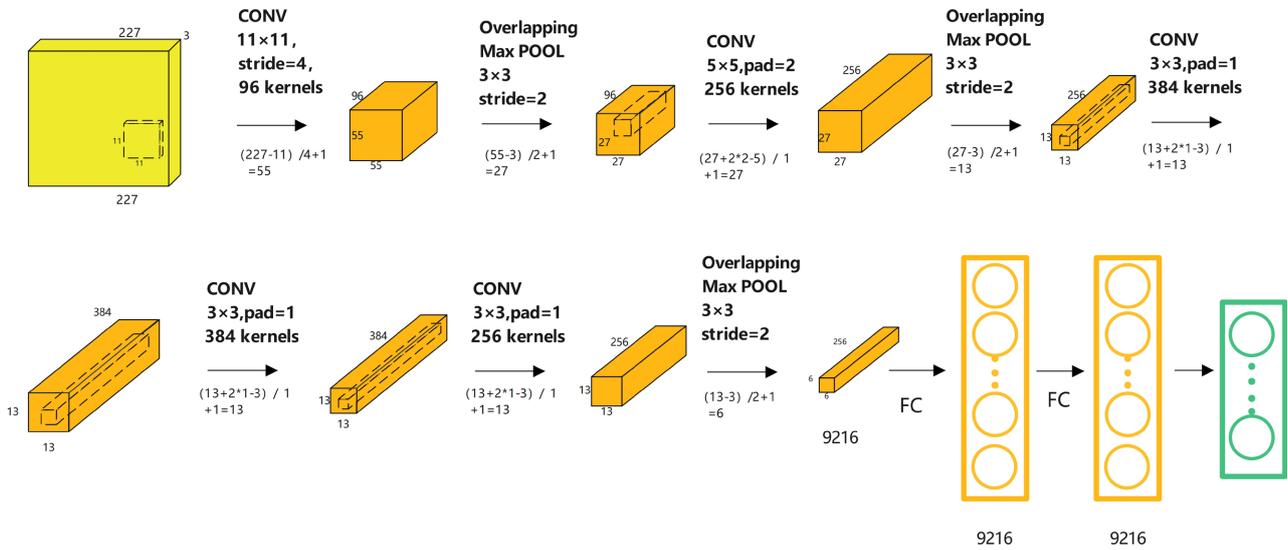


Figure 2. The basic structure of the AlexNet.

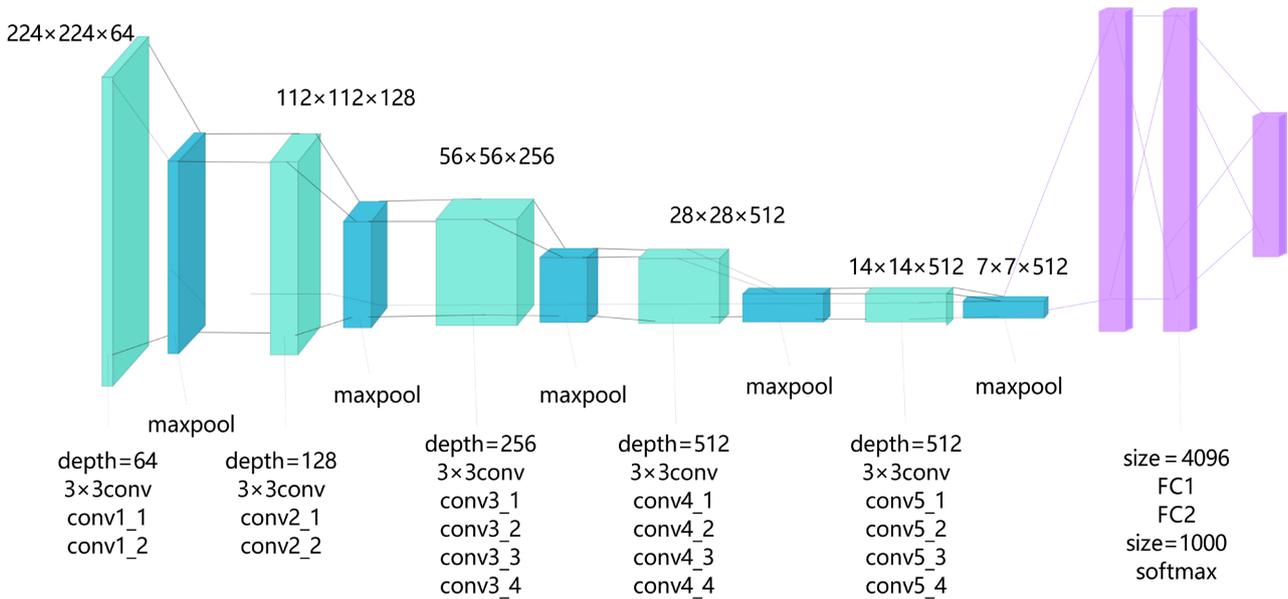


Figure 3. The basic structure of the VggNet.

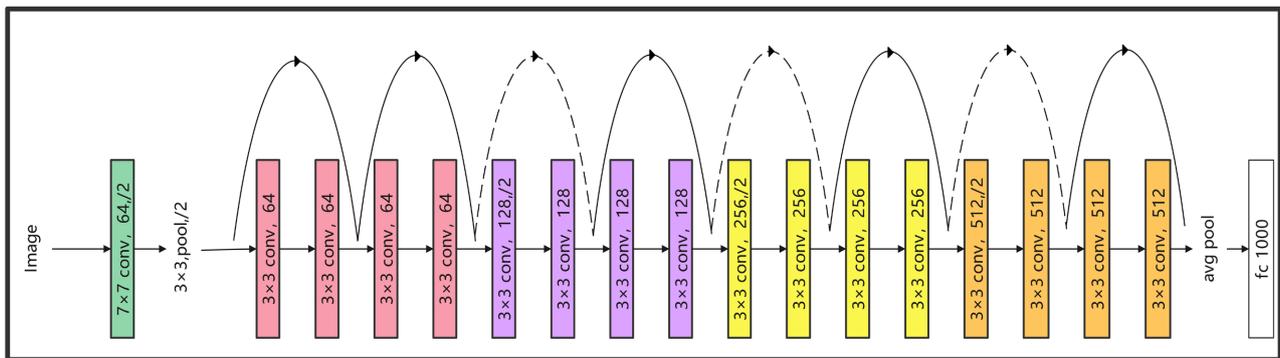


Figure 4. The basic structure of the ResNet.

structure of the network model of ResNet is the block definition and can be expressed as:

$$y = F(x, \{W_i\}) + x \quad (3)$$

Here x and y are the input and output vectors of the layers. The function F represents the residual mapping to be learned. The advantage of this structure is that the network can map the feature of the shallow layer to the deep layer, and the shallow layer and the deep layer can be effectively communicated. In the forward propagation of the network, the shallow features are not easy to be ignored in the deep layer. In the back propagation, the gradient of the deep layer can be directly passed back to the shallow layer, which solves the problem of network degradation well. Therefore, a network model with greater depth can be designed to achieve better network performance.

2.4. Results

We collect 2100 CSI samples for 7 group participants. Therefore, each group of a different number of persons contains 300 samples. We split the total sample into training, validation, and test as 8:1:1. We employ the ResNet 151 neural network and modify the network parameter according to our CSI data.

1) The CSI data representation

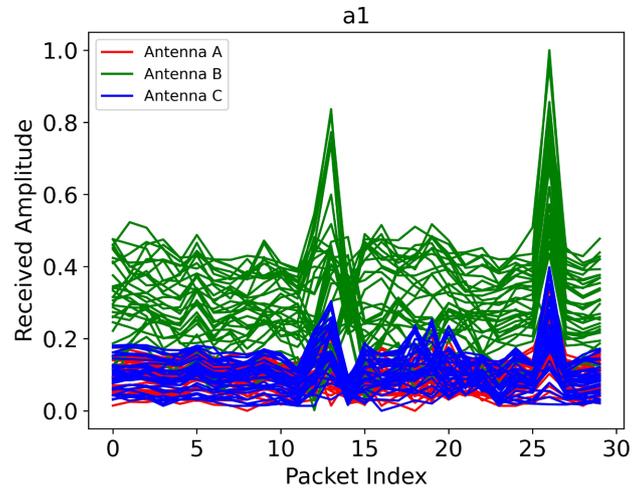
The representation of the CSI data can be described as follows. We collect the CSI data, normalize them, and visualize them using the figure to illustrate the similarity and differences among the number of persons. In **Figure 5**, the picture's color (red, green and blue) represents the received CSI signal corresponding to the three signal receiver antennas. **Figure 5(a)** and **Figure 5(b)** show the amplitude changes in different data packets and the signal subcarrier when one person walks among the WiFi area.

In **Figure 6**, to show the amplitude characteristics of CSI subcarrier more intuitively, we draw the CSI amplitude figure using the subcarrier index and package index as coordinate axis. **Figure 6(a)** is a two-dimensional representation, and **Figure 6(b)** is a three-dimensional representation.

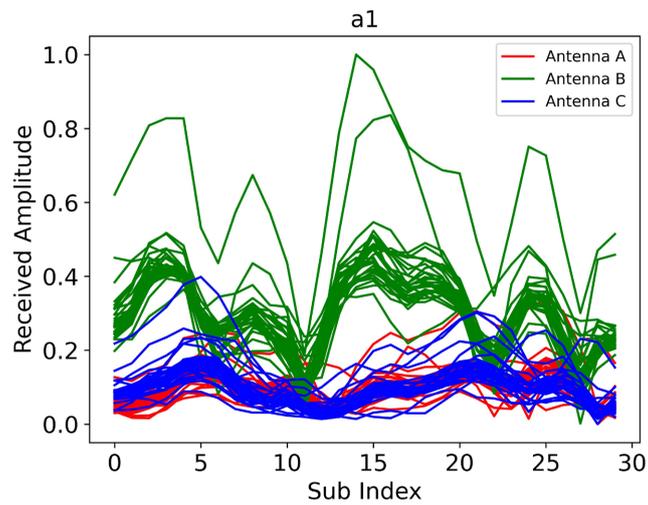
After obtaining the experimental data, we normalized the experimental data and then drew the corresponding three-dimension figure. The intuitive difference among the different numbers of persons can be exhibited in **Figure 7**. The same group has a similar CSI waveform at different times, as shown in **Figure 8**. This figure represents the CSI amplitude when seven persons walk in the WiFi area.

2) The result of the experiment

This paper uses the ResNet151 model to implement crowd counting, showing a transfer learning technique. We exhibit the training and validation accuracy along with the epoch, as shown in **Figure 9**. We find that the training process converges fast after ten epochs because we employ transfer learning and the network parameter initiated when we load the model.



(a)



(b)

Figure 5. Received amplitude along with packet and subcarrier index. (a) Amplitude along packet index; (b) amplitude along subcarrier index.

Besides, we draw the confusion matrix to evaluate the recognition results for the crowd counting from one to seven persons. As shown in **Figure 10**, every number of group persons is always 100 percent on the test data.

2.5. Discussion

This section compares recognition accuracy among different neural network models, including AlexNet and VggNet, because they are the typical image recognition models used in many image applications.

First, we analyze the effect of different epochs among three neural network models. We find that the three models converge very fast, and they reach the ideal training result after ten epochs. Therefore, there is no evident difference among these models. As a result, we can choose each model without considering the training epoch.

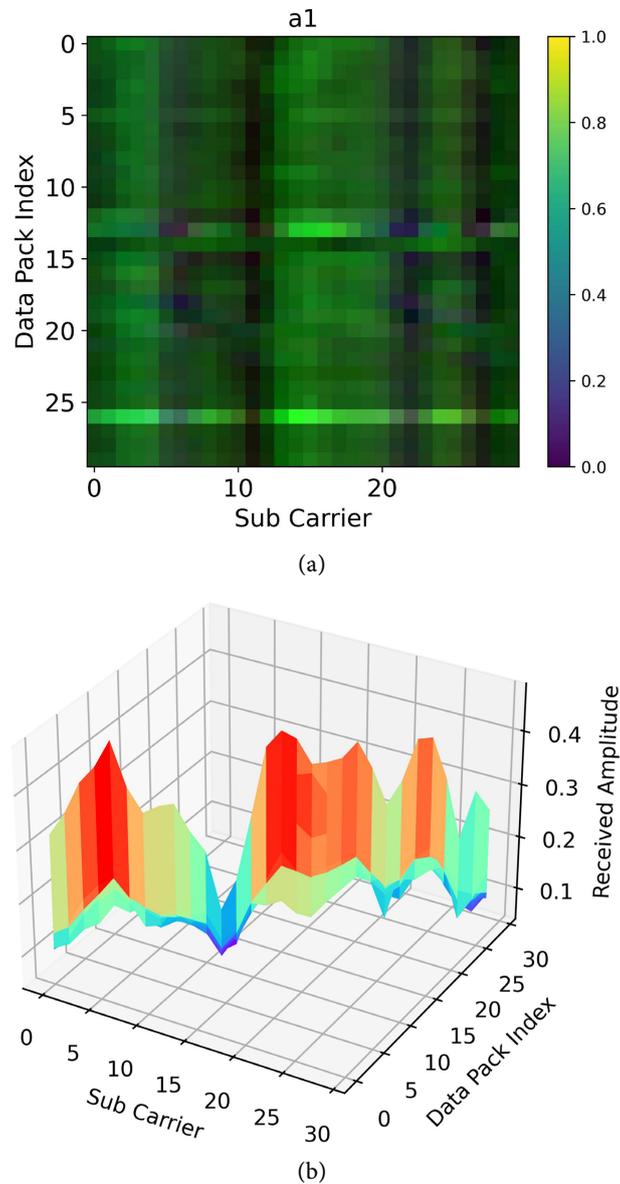


Figure 6. Spectrum graph of CSI amplitude information. (a) 2D spectrum graph; (b) 3D spectrum graph.

Next, we analyze the confusion matrix generated by the other neural networks, as shown in **Figure 11**. As we exhibit earlier, ResNet151 achieves 100 percent recognition accuracy for all test groups and obtains the ideal experiment result. As a comparison, AlexNet achieves an average accuracy of 99%, and the identification accuracy of the two people is the lowest, about 97%. The identification accuracy of all the other crowd counting is 100 percent. At the same time, the average accuracy of VGGNet is also 99 percent, and the lowest identification accuracy for the five persons is 90 percent. The identification accuracy of all the other person groups is 100%. Therefore, we think ResNet151 has an average accuracy of 100% for all seven groups, and it is very suitable for crowd counting using CSI. It holds the best performance among the three networks.

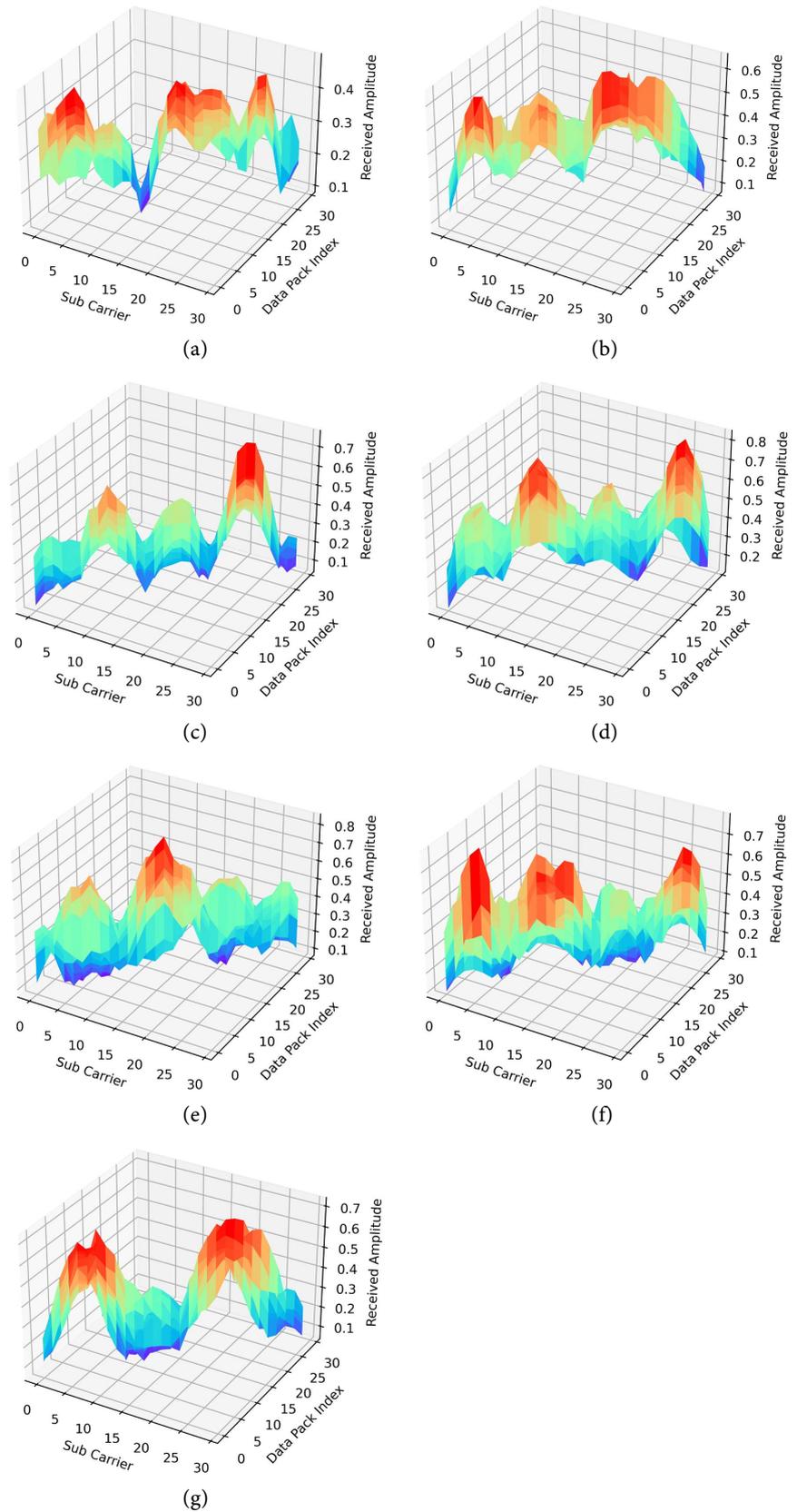


Figure 7. Subcarriers of the different participants. (a) One person; (b) two people; (c) three people; (d) four people; (e) five people; (f) six people; (g) seven people.

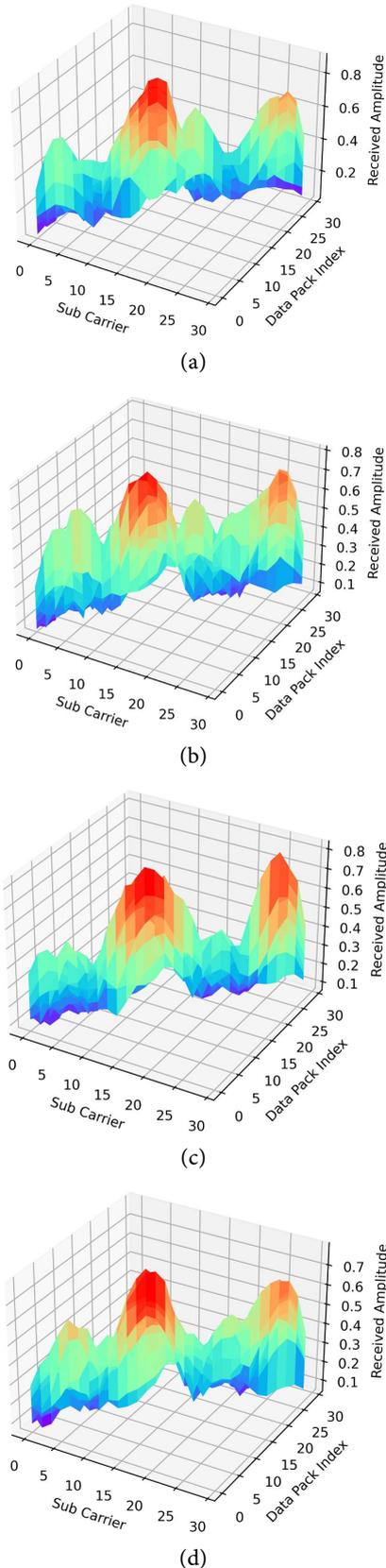


Figure 8. Subcarrier of the same persons (7 people), at different times. (a) First sample; (b) second sample; (c) third sample; (d) fourth sample.

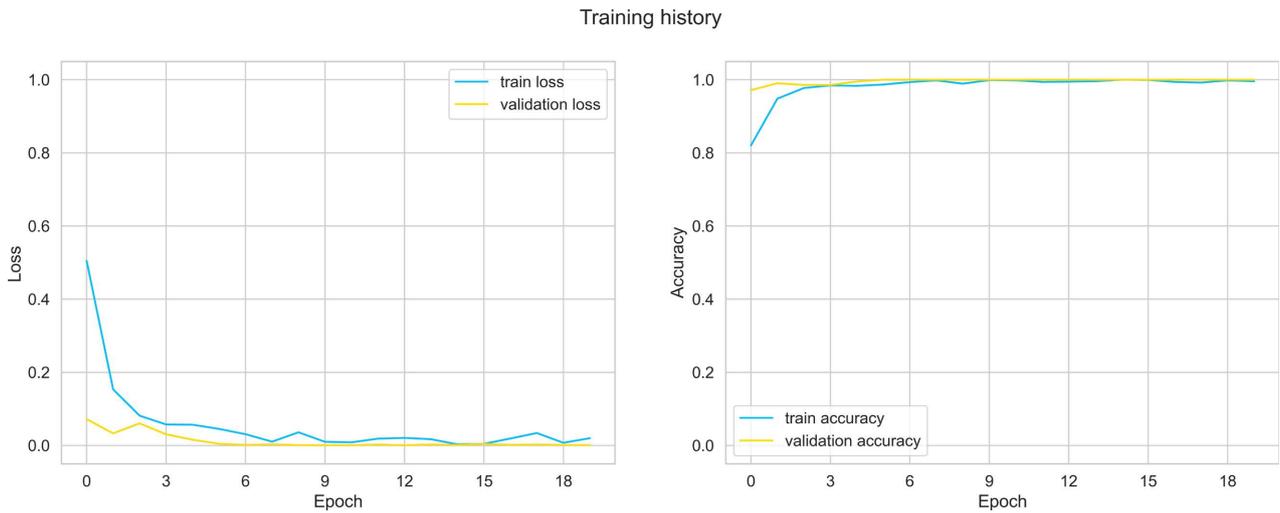


Figure 9. ResNet loss values and training accuracy.

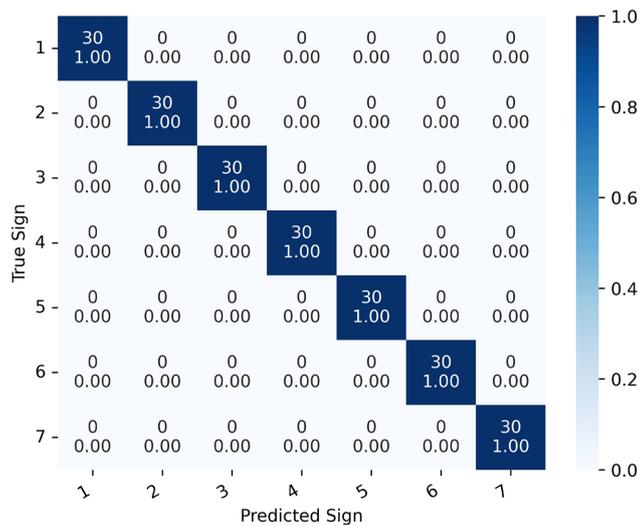


Figure 10. The confusion matrix of ResNet for crowd counting.

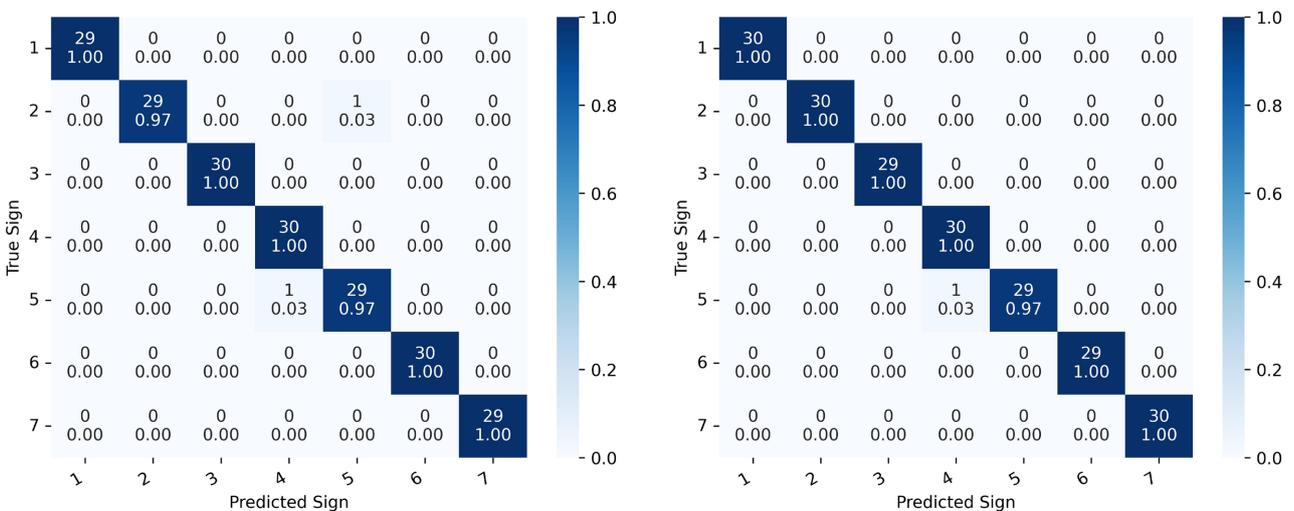


Figure 11. The confusion matrix of AlexNet and VGG for crowd counting.

3. Conclusions

At present, human-machine interaction is an important research topic in the artificial intelligence field. The WiFi-based crowd counting method has attracted more and more attention due to its advantages of device-free pattern and no privacy violations. This paper studies crowd counting from one to seven-person using transfer learning and CSI. Specifically, we transfer ResNet model into our application and implement crowd counting based on the collected CSI amplitude signal. Based on data collection, neural network model choosing, model training, and testing, we achieve the person counting of seven persons and acquire 100 percent test performance. As a comparison, AlexNet and VggNet models also obtain 99 percent recognition accuracy. The result shows that we can transfer the image processing model into crowd counting using CSI.

Although we obtain an ideal recognition result, we have some challenges when considering crowd counting using CSI. Firstly, the experiment may choose more persons to evaluate the model. Secondly, the experiment should be conducted in more different scenarios to validate the model's robustness. To sum up, we may conduct more experiments to prove that transfer learning is a feasible solution to crowd counting using CSI.

Acknowledgements

The work is funded by the foundation of the Innovation and Entrepreneurship Training Program for College Students (S202110424012).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Chowdhury, T.Z., Leung, C. and Miao, C.Y. (2017) WiHACS: Leveraging WiFi for Human Activity Classification Using OFDM Subcarriers' Correlation. *Proceeding of the 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Montreal, 14-16 November 2017, 338-342. <https://doi.org/10.1109/GlobalSIP.2017.8308660>
- [2] Venkatnarayan, R.H., Page, G. and Shahzad, M. (2018) Multi-User Gesture Recognition Using WiFi. *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, Association for Computing Machinery, Munich, 10-15 June 2018, 401-413. <https://doi.org/10.1145/3210240.3210335>
- [3] Wang, J., Zhang, L., Gao, Q., Pan, M. and Wang, H. (2018) Device-Free Wireless Sensing in Complex Scenarios Using Spatial Structural Information. *IEEE Transactions on Wireless Communications*, **17**, 2432-2442. <https://doi.org/10.1109/TWC.2018.2796086>
- [4] Wu, X., Chu, Z., Yang, P., Xiang, C., Zheng, X. and Huang, W. (2019) TW-See: Human Activity Recognition through the Wall with Commodity Wi-Fi Devices. *IEEE Transactions on Vehicular Technology*, **68**, 306-319. <https://doi.org/10.1109/TVT.2018.2878754>

- [5] Han, C., Wu, K., Wang, Y. and Ni, L.M. (2014) WiFall: Device-Free Fall Detection by Wireless Networks. *Proceeding of the IEEE INFOCOM 2014—IEEE Conference on Computer Communications*, Toronto, 27 April-2 May 2014, 271-279. <https://doi.org/10.1109/INFOCOM.2014.6847948>
- [6] Wang, H., Zhang, D., Wang, Y., Ma, J., Wang, Y. and Li, S. (2017) RT-Fall: A Real-Time and Contactless Fall Detection System with Commodity WiFi Devices. *IEEE Transactions on Mobile Computing*, **16**, 511-526. <https://doi.org/10.1109/TMC.2016.2557795>
- [7] Yang, X., Xiong, F., Shao, Y. and Niu, Q. (2018) WmFall: WiFi-Based Multistage Fall Detection with Channel State Information. *International Journal of Distributed Sensor Networks*, **14**, Article ID: 1550147718805718. <https://doi.org/10.1177/1550147718805718>
- [8] Li, H., Yang, W., Wang, J., Xu, Y. and Huang, L. (2016) WiFinger: Talk to Your Smart Devices with Finger-Grained Gesture. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Heidelberg, 12-16 September 2016, 250-261. <https://doi.org/10.1145/2971648.2971738>
- [9] Man, D., Yang, W., Wang, X., Lv, J., Du, X. and Yu, M. (2018) PWiG: A Phase-Based Wireless Gesture Recognition System. *Proceedings of the 2018 International Conference on Computing, Networking and Communications (ICNC)*, Maui, 5-8 March 2018, 837-842. <https://doi.org/10.1109/ICCNC.2018.8390335>
- [10] Lv, J., Man, D., Yang, W., Du, X. and Yu, M. (2018) Robust WLAN-Based Indoor Intrusion Detection Using PHY Layer Information. *IEEE Access*, **6**, 30117-30127. <https://doi.org/10.1109/ACCESS.2017.2785444>
- [11] Qian, K., Wu, C., Yang, Z., Liu, Y., He, F. and Xing, T. (2018) Enabling Contactless Detection of Moving Humans with Dynamic Speeds Using CSI. *ACM Transactions on Embedded Computing Systems*, **17**, Article No. 52. <https://doi.org/10.1145/3157677>
- [12] Soltanaghaei, E., Kalyanaraman, A. and Whitehouse, K. (2017) Peripheral WiFi Vision: Exploiting Multipath Reflections for More Sensitive Human Sensing. *Proceedings of the 4th International on Workshop on Physical Analytics*, New York, 19 June 2017, 13-18. <https://doi.org/10.1145/3092305.3092308>
- [13] Zhu, H., Xiao, F., Sun, L., Wang, R. and Yang, P. (2017) R-TTWD: Robust Device-Free Through-the-Wall Detection of Moving Human with WiFi. *IEEE Journal on Selected Areas in Communications*, **35**, 1090-1103. <https://doi.org/10.1109/JSAC.2017.2679578>
- [14] Grant, J. and Flynn, P. (2017) Crowd Scene Understanding from Video: A Survey. *ACM Transactions on Multimedia Computing, Communications, and Applications*, **13**, Article No. 19. <https://doi.org/10.1145/3052930>
- [15] Al-Hameed, A.A., Younus, S.H., Hussein, A.T., et al. (2019) LiDAL: Light Detection and Localization. *IEEE Access*, **7**, 85645-85687. <https://doi.org/10.1109/ACCESS.2019.2925076>
- [16] Shih, O. and Rowe, A. (2015) Occupancy Estimation Using Ultrasonic Chirps. *Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems*, Seattle, 14-16 April 2015, 149-158. <https://doi.org/10.1145/2735960.2735969>
- [17] Yang, X., Yin, W., Li, L. and Zhang, L. (2019) Dense People Counting Using IR-UWB Radar with a Hybrid Feature Extraction Method. *IEEE Geoscience and Remote Sensing Letters*, **16**, 30-34. <https://doi.org/10.1109/LGRS.2018.2869287>
- [18] Zou, H., Zhou, Y., Yang, J. and Spanos, C.J. (2018) Device-Free Occupancy Detection and Crowd Counting in Smart Buildings with WiFi-Enabled IoT. *Energy and*

- Buildings*, **174**, 309-322. <https://doi.org/10.1016/j.enbuild.2018.06.040>
- [19] Zhao, Y., Liu, S., Xue, F., Chen, B. and Chen, X. (2019) DeepCount: Crowd Counting with Wi-Fi Using Deep Learning. *Journal of Communications and Information Networks*, **4**, 38-52.
- [20] Zhou, R., Lu, X., Fu, Y. and Tang, M. (2020) Device-Free Crowd Counting with WiFi Channel State Information and Deep Neural Networks. *Wireless Networks*, **26**, 3495-3506. <https://doi.org/10.1007/s11276-020-02274-7>
- [21] Krizhevsky, A., Sutskever, I. and Hinton, G. (2017) ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90. <https://doi.org/10.1145/3065386>
- [22] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. *The 3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, 7-9 May 2015, 1-14.
- [23] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [24] Ayyadevara, V.K. and Reddy, Y. (2020) *Modern Computer Vision with PyTorch*. Packt Publishing, Birmingham.