Depthwise Separable Convolution based Passive Indoor Localization using CSI Fingerprint

Wenjing Xun[†], Lijuan Sun[†,‡], Chong Han[†,‡], Zhaoxiao Lin[†], Jian Guo[†,‡]
[†]*College of Computer, Nanjing University of Posts and Telecommunications*
[‡]*Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks*
Nanjing, China
nacn315@gmail.com, sunlijuan_nupt@163.com, hc@njupt.edu.cn, linzx@njupt.edu.cn, guoj@njupt.edu.cn

***Abstract*—Wi-Fi-based indoor localization has received extensive attention in the academic community. However, most Wi-Fi-based indoor localization systems have complex models and high localization delays, which limit the universality of these localization methods. To solve these problems, we propose a depthwise separable convolution based passive indoor localization system (DSCP) using Wi-Fi channel state information (CSI). DSCP is a fingerprint-based localization system, which includes an offline training phase and an online localization phase. In the offline training phase, the indoor scenario is first divided into different areas to set training locations for collecting CSI. Then amplitude differences of these CSI subcarriers are extracted for constructing location fingerprints, thereby training the CNN. In the online localization phase, CSI data is first collected at the test locations, then the location fingerprint is extracted and finally fed to the trained network to obtain the predicted location. The experimental results show that DSCP has a short training time and a low localization delay. DSCP achieves a high localization accuracy, upper than 97%, and a small median localization distance error of 0.98 m in open indoor scenarios.**

***Index Terms*—channel state information, indoor localization, location fingerprint, convolutional neural network**

I. Introduction

As mobile smart devices and wireless networks influence all aspects of human production and life, Location Based Service (LBS) has gradually become indispensable in people's lives. The great research significance and practical value of indoor localization attract a large number of researchers, with a large number of indoor localization methods emerged. For now, computer vision, ultrasonic, ultra-wideband, Bluetooth, RFID, etc. [1]–[3] are mainly used in indoor localization. However, most of these localization methods require a large number of hardware devices, which limits the universality of localization. With the widespread adoption of Wi-Fi, Wi-Fi-based indoor localization methods have gradually become mainstream.

Received Signal Strength (RSS), as the main energy characteristic measurement of wireless signals, can be directly obtained from most wireless terminals and is widely used in Wi-Fi-based indoor localization systems. RADAR is the first localization system based on WLAN [4], which uses RSS and fingerprint matching. However, in the indoor environment, wireless signals propagate from the transmitter to the receiver through multiple paths due to the presence of obstacles. As an average of multipath signals, RSS has poor stability indoor. Compared to RSS, channel state information (CSI) is a fine-grained information from PHY layer, which describes multipath propagation. Figuratively speaking, CSI is to RSS what a rainbow is to a sunbeam [5]. By modifying the firmware, a sample version of the channel frequency response (CFR) can be obtained in the form of CSI from some Wi-Fi network interface cards (NIC) such as Intel 5300. Each set of CSI characterizes the amplitude and phase of orthogonal frequency division multiplexing (OFDM) subcarriers. Therefore, CSI has better static stability and dynamic sensitivity [6].

Fingerprints of different locations are generally different. Therefore, fingerprint-based localization can be regarded as a classification task. In recent years, a growing number of researchers have been applying neural networks to fingerprint-based localization. Compared with traditional algorithms, deep learning algorithms often have higher accuracy. However, most CSI fingerprinting-based localization methods based on deep learning require a lot of time in the offline training phase, and the online localization delay is high. In order to solve these problems, we propose a depthwise separable convolution based passive indoor localization system (DSCP) using Wi-Fi CSI. DSCP is based on depthwise separable convolutions, which are lighter than standard convolutions. During the training phase, data from the training locations is collected for training of the DSCP network. In the localization phase, the data of the target location is collected and the feature images are extracted. Then the probability-based method is used to estimate location.

The remainder of this paper is organized as follows. The related work is presented in Section II. The preliminary is presented in Section III. The DSCP system is presented in Section IV. Section V provides the experiment validation while Section VI concludes the paper.

II. Related Work

According to different technical means, CSI-based indoor localization can be mainly divided into indoor localization based on CSI fingerprints and indoor localization based on CSI ranging. Because the signal is affected by the multipath effect indoor, the localization accuracy is low using a single signal-to-distance mapping relationship. Fingerprint-based localization can solve the problem.

Sen et al. propose a fingerprint-based localization system called PinLoc [7] using the distinguishability of CSI at different locations. Xiao et al. use the frequency diversity and

978-1-7281-3106-1/20/$31.00 ©2020 IEEE

spatial diversity of CSI to construct the subchannels power information location fingerprints based on CSI amplitude, and propose the FIFS system [8]. A year later, Xiao et al. propose the first CSI-based passive indoor localization system Pilot [9]. Pilot uses probability estimation to match the obtained CSI with the fingerprint database to achieve position. Chapre et al. use the amplitude differences and phase differences of CSI to construct fingerprints, and propose a Wi-Fi fingerprint system CSI-MIMO [10]. ABDEL-NASSER et al. [11] observe that the CSI probability density distribution of different locations fits a Gaussian distribution mixture nicely, and use the k-means to construct a cluster-based fingerprint database. To improve the localization accuracy, Sabek et al. propose a single AP-MP based localization system MoneStream [12], which models the passive localization as an object recognition problem.

Compared with traditional methods, deep learning-based methods are more accurate. Wang et al. propose deep learning-based CSI fingerprint indoor localization system DeepFi and PhaseFi [13], [14] using the amplitude and phase of CSI, respectively. In the offline training phase, the system uses the weights of the deep learning training as the location fingerprints. In the online localization phase, the probability method based on radial basis function is used to obtain the estimated location. Chen et al. propose the first CNN-based Wi-Fi indoor localization system ConFi [15]. When extracting CSI features, ConFi simulates RGB images, which have three channels. The data of different transmitting-receiving antenna pairs in a multiple input multiple output (MIMO) system is regarded as the different channels of the image. ConFi predicts the location of the target by calculating the probability of the target at different training points. Similarly, Cai et al. construct location fingerprints based on the amplitude of CSI, and propose a passive localization system PILC [16].

Most CSI fingerprint localization systems based on traditional methods have shortcomings such as single feature of location fingerprint information and complex model. Although deep learning methods can mine more features, the training of network models generally takes a lot of time and limits the real-time capability of localization. To this end, we propose a CSI fingerprint localization system based on a lightweight convolutional neural network.

III. Preliminary

A. Channel State Information

CSI represents the channel characteristics of the communication link between the transmitter and the receiver, reflecting the effects of scattering, attenuation, etc. of the signal propagation. According to IEEE 802.11n, a signal can be transmitted through a set of subcarriers with different frequencies and mutually orthogonal by OFDM. In a smooth narrowband channel, the OFDM system can be expressed as:

$$y_i = Hx_i + N_i \tag{1}$$

Where y_i and x_i represent the signal of the receiver and transmitter, respectively. H is the CSI matrix. N_i is the noise.

In order to estimate the CSI matrix, the transmitter transmits a known pilot sequence $x_1, x_2, \ldots, x_n$, and the combined received signal Y can be expressed as:

$$Y = [y_1, y_2, ..., y_n] = HX + N \tag{2}$$

Therefore, the CSI matrix can be approximated as:

$$\hat{H} = \frac{Y}{X} \tag{3}$$

For a system with N subcarriers, H can be expressed as:

$$H = [H_1, H_2, \ldots, H_N] \tag{4}$$

For a MIMO system with m transmitting antennas and n receiving antennas, H_i can be expressed as:

$$H_i = \begin{bmatrix} h_{11} & h_{12} & \ldots & h_{1n} \\ h_{21} & h_{22} & \ldots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{m1} & h_{m2} & \ldots & h_{mn} \end{bmatrix} \tag{5}$$

Where $h_{pq}(p \in [1, m], q \in [1, n])$ corresponds to the complex number of subcarrier amplitude and phase on the stream.

B. Fingerprint based Localization

In an indoor environment, due to the walls and obstacles, wireless signals form multipath effects. Human at different physical locations has different effects on the wireless signal propagation paths. The differences can be expressed as fingerprints that reflect the characteristics of human at different locations [18]. Generally, in the case of small indoor environment changes, a valid location fingerprint needs to meet the following two conditions:

- The fingerprints of different locations have differences at the same time.
- The fingerprints of the same location at different times are stable.

Fingerprint-based localization is usually divided into two phases: offline training and online localization. In the offline training phase, it is necessary to first divide the indoor scenario and set sampling points. The data collected at different sampling points is called location fingerprint. Then the sample data is processed to construct a fingerprint map. In the online localization phase, the data collected at the target location is compared with the fingerprints in the fingerprint map. The location which has the most similar fingerprint to the target fingerprint is the location of the target.

C. Depthwise Separable Convolution

Google proposes a light weight deep neural network MobileNet [17] based on depthwise separable convolutions. Unlike standard convolutions, each depthwise separable convolution is made up of a depthwise convolution and a pointwise convolution. Depthwise convolution does not change the depth of the input image, and each channel of the image corresponds to a convolution kernel with a depth of 1. The pointwise convolution then applies a 1×1 convolution to the image.

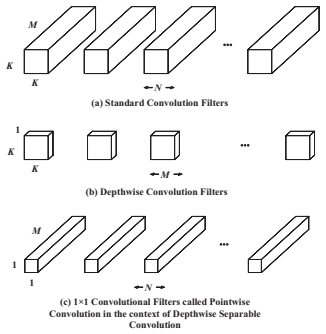

Fig. 1. Depthwise separable convolution [17].

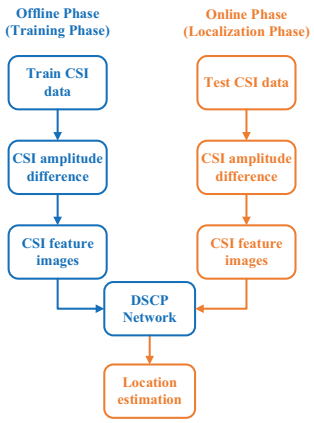

Fig. 2. DSCP architecture.

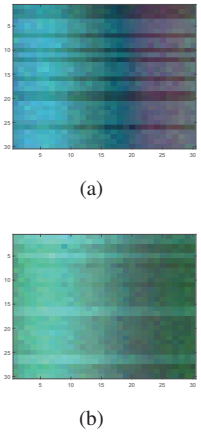

Fig. 3. CSI feature images based on subcarrier amplitudes

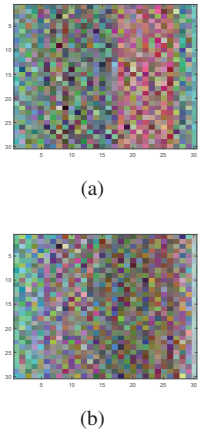

Fig. 4. CSI feature images based on subcarrier amplitudes differences

As shown in Fig. 1, assuming the input image is $H \times W \times M$, the size of the output image that needs to be $H \times W \times N$. If a standard convolution kernel of size $K \times K \times M$ is used, the calculation cost is $H \times W \times M \times N \times K \times K$. Using a depthwise separable convolution, the total calculation cost is $H \times W \times M \times K \times K + H \times W \times M \times N$, which is only $\frac{1}{N} + \frac{1}{K^2}$ of the standard convolution. Depthwise separable convolutions can significantly improve the computational efficiency of the network at only a small reduction in accuracy. Therefore, we use depthwise separable convolutions in the DSCP network.

IV. The DSCP System

A. System Architecture

As shown in Fig. 2, the DSCP system architecture includes an offline phase and an online phase. In the offline phase, the CSI data of each training location is first collected, and then the amplitude of the CSI is extracted. DSCP constructs CSI feature images as fingerprints of each training location for training the network. In the online phase, the CSI data of the target location is collected, the fingerprint of the target location is extracted to be input to the trained network. Finally, the predicted location is output by the DSCP network.

B. Location Fingerprint

30 subcarriers can be read for CSI information via Intel 5300 NIC. In this paper, we use the subcarrier amplitudes when constructing the location fingerprints. Each packet contains a $1 \times 3 \times 30$ CSI matrix, which corresponds to the complex numbers of subcarrier amplitude and phase on the stream. For each transmitting-receiving antenna pair, a sliding window is used to select 30 consecutive data packets in a sequence, and the sliding window size is 30. Multiple sets of 30×30 matrices are constructed. Similar to RGB images, the data of each antenna pair corresponds to one channel, and each location has some CSI feature images of $30 \times 30 \times 3$ [15]. As shown in Fig. 3, the CSI feature images formed by the amplitudes of subcarriers can reflect the difference of adjacent locations.

Different from the ConFi, we consider that the amplitudes of the subcarriers are greatly affected by the physical environment, which may result in the little influence of the human on CSI. In order to amplify the differences between fingerprints of different locations, the amplitude differences between adjacent subcarriers are used in constructing CSI feature images at different locations. The CSI feature images combine with time domain, frequency domain and spatial domain information of CSI data. Therefore, the CSI feature images of each location are used as the location fingerprints, as shown in Fig. 4.

For the transmitting antenna p and the receiving antenna q, the amplitude of the subcarrier i in packet j can be expressed as:

$$h_{amp_{ij}} = |h_{pq}| \tag{6}$$

The amplitude difference between adjacent subcarriers is:

$$h_{d-amp_{ij}} = \begin{cases} h_{amp_{i+1j}} - h_{amp_{ij}} & (i \in [1, 29]) \\ h_{amp_{1j}} - h_{amp_{ij}} & (i = 30) \end{cases} \tag{7}$$

Then a CSI feature image can be expressed as:

$$F = \begin{bmatrix} h_{d\text{-}amp_{1k}} & \cdots & h_{d\text{-}amp_{30k}} \\ \vdots & \ddots & \vdots \\ h_{d\text{-}amp_{1k+29}} & \cdots & h_{d\text{-}amp_{30k+29}} \end{bmatrix} \tag{8}$$

C. Network Structure

TABLE I
Parameter Setting of DSCP Network

Type/Stride	Input Size	Filter Shape	Output Size
conv	$30 \times 30 \times 3$	$3 \times 3 \times 3 \times 64$	$30 \times 30 \times 64$
conv dw/s1	$30 \times 30 \times 64$	$3 \times 3 \times 64$ dw	$30 \times 30 \times 64$
conv/s1	$30 \times 30 \times 64$	$1 \times 1 \times 64 \times 256$	$30 \times 30 \times 256$
conv dw/s2	$30 \times 30 \times 256$	$3 \times 3 \times 256$ dw	$15 \times 15 \times 256$
conv/s1	$15 \times 15 \times 256$	$1 \times 1 \times 256 \times 1024$	$15 \times 15 \times 1024$
conv dw/s2	$15 \times 15 \times 1024$	$3 \times 3 \times 1024$ dw	$8 \times 8 \times 1024$
conv/s1	$8 \times 8 \times 1024$	$1 \times 1 \times 1024 \times 1024$	$8 \times 8 \times 1024$
Avg Pool/s1	$8 \times 8 \times 1024$	Pool 8×8	$1 \times 1 \times 1024$
FC	$1 \times 1 \times 1024$	1024×32	$1 \times 1 \times 32$

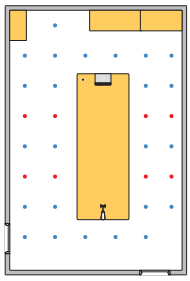

Fig. 5. Scenario 1

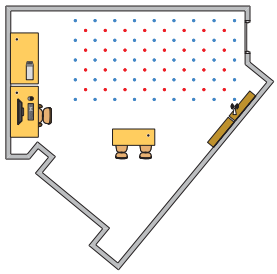

Fig. 6. Scenario 2

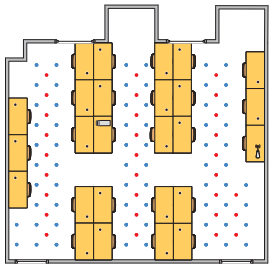

Fig. 7. Scenario 3

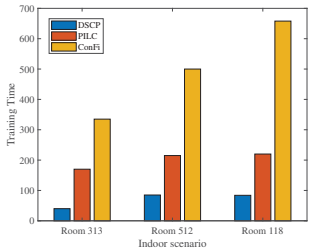

Fig. 8. Training time of models based on CNN

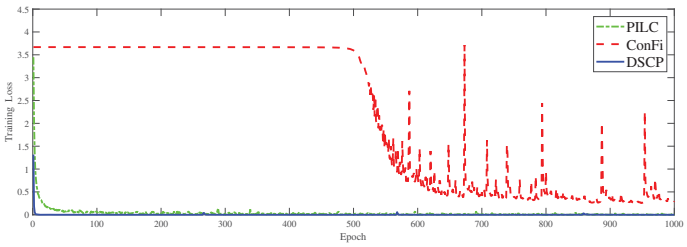

Fig. 9. Training loss of models based on CNN.

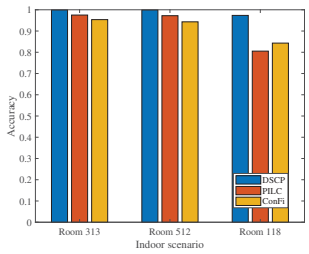

Fig. 10. Localization accuracy of the methods based on CNN.

Inspired by the MobileNet, we design the DSCP network based on depthwise separable convolutions. The DSCP network has seven convolutional layers, the first convolutional layer is a full convolution, and the others are depthwise separable convolutions. Table I shows the network structure. In the network, the standard convolution, pointwise convolutions, and first depthwise convolution stride are set to 1 to keep the size of the image. The second and third depthwise convolution stride are set to 2. There is a batchnorm and ReLU after each depthwise convolution and pointwise convolution. Since only 30 subcarriers can be extracted with the Intel 5300 NIC, the size of the CSI feature images are much smaller than real images. Therefore, the size of each DSCP depthwise convolution is set to 3×3. The output size of the final fully connected layer depends on the number of training locations. Adam optimization [19] is used in network training.

Similar to processing images with the CNN, DSCP considers the amplitude differences between four adjacent subcarriers of the same packet in each convolution. Moreover, for the same pair of adjacent subcarriers, DSCP also considers their amplitude differences at three consecutive time points. Therefore, the DSCP network can learn the correlation of the CSI amplitude differences in the time and frequency domain.

Three channels of the CSI feature images correspond to the data of different transmitting-receiving antenna pairs in the MIMO system. The standard convolution considers different channels of the image simultaneously in the convolution. Different from standard convolutions, the depthwise convolutions of DSCP use different convolution kernels for three channels of the CSI feature images separately. DSCP realizes the complete separation of learning the frequency domain, time domain correlation and learning spatial domain correlation of the CSI subcarriers, reducing the coupling of the convolution kernels.

D. Location Estimation

During the localization phase, test data will be fed to the trained network, which will output the probability of the test data at each of the trained locations. Probability-based methods tend to have higher precision than classification-based methods. So, we use the method of weighted mean based on probability to estimate the final location. Assuming that there are N training locations, the coordinate of the i-th training location is L_i, and the probability of the test data at the i-th training location is P_i, the estimated location can be expressed as:

$$\widehat{L} = \sum_{i=1}^{N} P_i \times L_i \tag{9}$$

V. Experiment Validation

A. Experiments Setup

In the experiments, we use a TP-LINK router as the transmitter. A Dell desktop PC equipped with Intel 5300 NIC and Ubuntu 14.04 is used as the receiver. The tensorflow is accelerated with GPU during the experiments. The sampling rate is set to 50 Hz. When collecting data, the tester stands at each location for 60 seconds, and the receiver collects 3000 data packets. The Intel 5300 NIC has three antennas, and the wireless router has one antenna, which is just a 1×3 MIMO system. The amplitudes and phases of the 30 subcarriers can be acquired from the NIC. In order to simplify the system, we only use the amplitudes of CSI.

In this paper, three indoor scenarios were selected for experiments, as shown in Fig. 5-7. The sampling points are

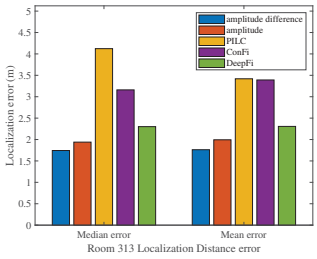

Fig. 11. Localization error in scenario 1.

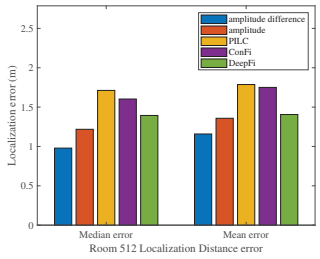

Fig. 12. Localization error in scenario 2.

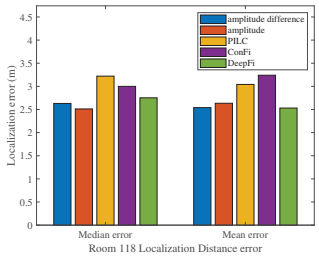

Fig. 13. Localization error in scenario 3.

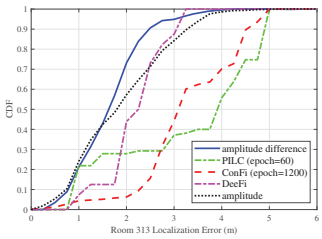

Fig. 14. CDF of localization errors in scenario 1.

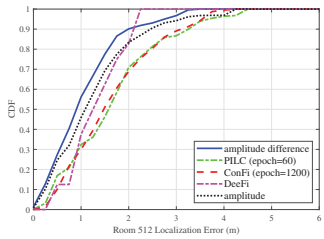

Fig. 15. CDF of localization errors in scenario 2.

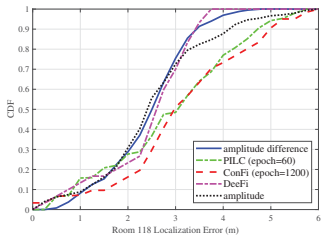

Fig. 16. CDF of localization errors in scenario 3.

divided into training locations and test locations marked as blue and red, respectively. Scenario 1 is a $5.8m \times 8.6m$ laboratory. A total of 32 sampling points are set. The spacing between adjacent sampling points is $1m$. 24 locations are selected for training and 8 for testing. Scenario 2 is a $40m^2$ laboratory. 45 locations are selected for training, 32 for testing. Scenario 3 is an office of $7.5m \times 6.5m$. 66 locations are selected for training and 30 for testing. The spacing between adjacent sampling points is $0.6m$ in both Scenario 2 and 3.

B. Model Evaluation

TABLE II
LOCALIZATION ACCURACY OF THE METHODS BASED ON CNN

Localization Method	Localization Accuracy		
	Scenario 1	*Scenario 2*	*Scenario 3*
DSCP	99.93%	99.94%	97.35%
PILC	97.52%	97.22%	80.54%
ConFi	95.37%	94.33%	84.32%

DSCP is compared with two CNN-based indoor localization methods, PILC and ConFi. As shown in Fig. 9, the training loss of PILC starts to converge when epoch is around 50. The training loss of ConFi converges when epoch is around 1000. The training loss of DSCP converges when epoch is around 20. In subsequent experiments, the values of the epoch were set according to convergence conditions of different networks. The network training time of three localization methods based on CNN is shown in Fig. 8.

In this paper, 75% of all training location data is the training set, and the remaining 25% is the test set. The localization accuracy of DSCP, PILC and ConFi are calculated. The batchsize of DSCP and PILC are the numbers of training locations. The epoch of DSCP, PILC and Confi are 30, 60, 1200 respectively. The batchsize of ConFi is 256. As shown in Fig. 10 and Table II, the localization accuracy of DSCP in scenario 1 and scenario 2 is over 99%, and the localization accuracy in scenario 3 reaches 97%. The localization accuracy of DSCP in each scenario is higher than the other two methods.

C. Localization Performance

The performance of DSCP is evaluated in three scenarios. We compare DSCP with three deep learning-based methods, PILC, ConFi and DeepFi. For DSCP, experiments based on CSI amplitude and amplitude differences are performed respectively. The median distance error and the mean distance error of different localization methods are calculated.

a) Scenario 1: Fig. 11 and Table III show the localization errors of four methods in scenario 1. Scenario 1 is a typical laboratory with some indoor furniture. It can be seen from the experiments that DSCP can obtain a median error of $1.75m$ and a mean error of $1.77m$. Compared with other methods based on deep learning, the localization error of DSCP is smaller. DSCP is 57.71% lower than PILC in median distance error, 44.82% lower than ConFi, and 24.23% lower than DeepFi. Fig. 14 presents the CDF of localization errors of the four methods in scenario 1. With DSCP, about 40% of the test cases have an error under $1.5m$, and about 73% of the test cases have an error under $2m$.

b) Scenario 2: Fig. 12 and Table III show the localization errors in scenario 2. There is no obstacle in the experimental area of scenario 2, and there are line of sight (LOS) paths between the transmitter and the receiver. As the experiments show in scenario 2, DSCP can obtain a median error of $0.98m$ and a mean error of $1.16m$. DSCP is 42.83% lower than PILC in median distance error, 38.93% lower than ConFi, and 29.77% lower than DeepFi. Fig. 15 presents the CDF of localization errors of the four methods in scenario 2. With

TABLE III
LOCALIZATION DISTANCE ERROR

Localization Method	Scenario 1		Scenario 2		Scenario 3	
	Mean Error (m)	*Median Error (m)*	*Mean Error (m)*	*Median Error (m)*	*Mean Error (m)*	*Median Error (m)*
DSCP (amplitude difference)	1.7619	1.7435	1.1584	0.9789	2.5406	2.6307
DSCP (amplitude)	1.9938	1.9394	1.3579	1.2180	2.6348	2.5119
PILC	3.4191	4.1231	1.7869	1.7122	3.0418	3.2211
ConFi	3.3915	3.1597	1.7503	1.6028	3.2408	3.0000
DeepFi	2.3065	2.3011	1.4059	1.3939	2.5314	2.7524

DSCP, 55% of the test cases have an error under $1m$, and about 90% of the test cases have an error under $2m$.

c) Scenario 3: Fig. 13 and Table III show the localization errors in scenario 3. Scenario 3 is a typical office with a large number of office chairs in the room. The experimental result shows that DSCP can obtain a median error of $2.63m$ and a mean error of $2.54m$. DSCP is 18.33% lower than PILC in median distance error, 12.31% lower than ConFi, and 4.42% lower than DeepFi. Fig. 16 depicts the CDF of localization errors of the four methods in scenario 3. With DSCP, 30% of the test cases have an error under $2m$, and about 75% of the test cases have an error under $3m$.

The experiments show that the four methods have better performance in scenario 2 than the others. The reason may be that there is no obstacle between the transmitter and the receiver to block wireless signals. In scenario 3, the multipath propagation of the wireless signal is more complicated, so the localization performance is not so good as the other scenarios. The location fingerprints based on the CSI amplitude differences are slightly better than that based on amplitudes.

VI. CONCLUSION

This paper proposes a passive indoor localization system DSCP based on depthwise separable convolutions using CSI fingerprints. The location fingerprints of DSCP combine data in time domain, frequency domain and spatial domain, and contain rich location information. We propose to use a neural network based on depthwise separable convolutions to solve the location classification problem in fingerprint-based localization. The DSCP network consists of seven convolutional layers. Only one layer uses a standard convolution. The remaining layers use depthwise separable convolutions, which can significantly reduce the computational cost of the network with little decrease in accuracy. We conduct the experiments in three typical indoor scenarios. DSCP is compared with several existing CSI fingerprint-based localization methods. The experimental results show that DSCP is better than other methods in the localization accuracy and localization error.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China under Grant Nos. 61873131, 61702284, 61572261 and Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant No. KYCX19_0972.

REFERENCES

[1] R. Want, A. Hopper, V. Falcão, and J. Gibbons, "The active badge location system," *ACM Trans. Inf. Syst.*, vol. 10, no. 1, pp. 91–102, Jan. 1992.

[2] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *Computer*, vol. 34, no. 8, pp. 57–66, Aug 2001.

[3] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2568–2599, thirdquarter 2019.

[4] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, vol. 2, March 2000, pp. 775–784 vol.2.

[5] Y. Zou, W. Liu, K. Wu, and L. M. Ni, "Wi-Fi Radar: Recognizing human behavior with commodity Wi-Fi," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 105–111, Oct 2017.

[6] H. Zhu, X. Fu, L. Sun, X. Xie, and R. Wang, "CSI-based WiFi environment sensing," *Journal of Nanjing University of Posts and Telecommunications*, 2016.

[7] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "You are facing the Mona Lisa: Spot localization using PHY layer information," in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '12. New York, NY, USA: ACM, 2012, pp. 183–196.

[8] J. Xiao, K. Wu, Y. Yi, and L. M. Ni, "FIFS: Fine-grained indoor fingerprinting system," in *2012 21st ICCCN*, July 2012, pp. 1–7.

[9] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 25:1–25:32, Dec. 2013.

[10] Y. Chapre, A. Ignjatovic, A. Seneviratne, and S. Jha, "CSI-MIMO: An efficient Wi-Fi fingerprinting using channel state information with MIMO," *Pervasive and Mobile Computing*, vol. 23, pp. 89 – 103, 2015.

[11] H. Abdel-Nasser, R. Samir, I. Sabek, and M. Youssef, "MonoPHY: Mono-stream-based device-free WLAN localization via physical layer information," in *2013 IEEE WCNC*, April 2013, pp. 4546–4551.

[12] I. Sabek and M. Youssef, "Monostream: A minimal-hardware high accuracy device-free WLAN localization system," *CoRR*, vol. abs/1308.0768, 2013.

[13] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, Jan 2017.

[14] X. Wang, L. Gao, and S. Mao, "CSI phase fingerprinting for indoor localization with a deep learning approach," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1113–1123, Dec 2016.

[15] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "ConFi: Convolutional neural networks based indoor Wi-Fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18 066–18 074, 2017.

[16] C. Cai, L. Deng, M. Zheng, and S. Li, "PILC: Passive indoor localization based on convolutional neural networks," in *2018 Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS)*, March 2018, pp. 1–6.

[17] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *CoRR*, vol. abs/1704.04861, 2017.

[18] X. Dang, X. Si, Z. Hao, and Y. Huang, "A novel passive indoor localization method by fusion CSI amplitude and phase information," *Sensors*, vol. 19, no. 4, 2019.

[19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.