

Research Article

DSCP: Depthwise Separable Convolution-Based Passive Indoor Localization Using CSI Fingerprint

Chong Han ^{1,2}, Wenjing Xun,¹ Lijuan Sun,^{1,2} Zhaoxiao Lin,¹ and Jian Guo^{1,2}

¹College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

²Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Correspondence should be addressed to Chong Han; hc@njupt.edu.cn

Received 12 August 2020; Revised 7 December 2020; Accepted 22 December 2020; Published 4 January 2021

Academic Editor: Miguel Garcia-Pineda

Copyright © 2021 Chong Han et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Wi-Fi-based indoor localization has received extensive attention in wireless sensing. 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, a depthwise separable convolution-based passive indoor localization system (DSCP) is proposed. DSCP is a lightweight fingerprint-based localization system that 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, the amplitude differences of these CSI subcarriers are extracted to construct location fingerprints, thereby training the convolutional neural network (CNN). In the online localization phase, CSI data are first collected at the test locations, and 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, above 97%, and a small median localization distance error of 0.69 m in typical indoor scenarios.

1. Introduction

1.1. Background. As mobile smart devices and wireless networks influence all aspects of human production and life, location-based services (LBSs) have gradually become indispensable in people's lives [1]. Although global positioning systems (GPSs) have become a leader in outdoor navigation and localization, the severe decline in GPS signals caused by reinforced concrete makes it difficult to navigate and localize indoors. The great research significance and practical value of indoor localization have attracted a large number of researchers, with a large number of indoor localization [2–4] methods emerging. Currently, wireless sensor networks, computer vision, ultrasonic, ultrawideband, Bluetooth, and RFID [5, 6] 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, an increasing number of researchers have applied Wi-Fi to wireless sens-

ing. Wi-Fi-based indoor localization methods [7–9] have gradually become mainstream.

The 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 [10]. RADAR, proposed by Microsoft Research, was the first system to implement indoor localization using WLAN [11]. RADAR uses RSS and fingerprint matching and proposed a fingerprint-based localization method for the first time. However, in the indoor environment, wireless signals propagate from the transmitter to the receiver through multiple paths due to the presence of obstacles such as walls and furniture. The signals received by the receiver are the superposition of signals from multiple paths. As an average of multipath signals, the RSS is unstable in indoor environments, which limits the reliability and localization accuracy of RSS-based indoor localization systems. Compared to RSS, channel state information (CSI) is fine-grained

TABLE 1: CSI grouping of 20 MHz bandwidth in 802.11n.

BW	Grouping	N_g	N_s	Carriers for which matrices are sent
		1	56	All carriers: -28, -27, ..., -2, -1, 1, 2, ..., 27, 28
20 MHz	2	30		-28, -26, -24, -22, -20, -18, -16, -14, -12, -10, -8, -6, -4, -2, -1, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 28
	4	16		-28, -24, -20, -16, -12, -8, -4, -1, 1, 5, 9, 13, 17, 21, 5, 28

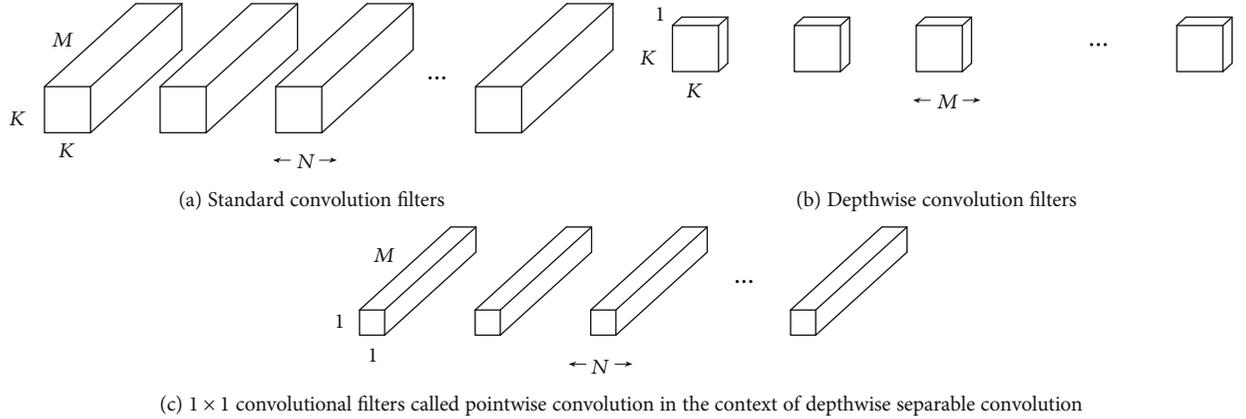


FIGURE 1: Depthwise separable convolution [30].

information from the PHY layer, which describes multipath propagation. Figuratively speaking, CSI is to RSS what a rainbow is to a sunbeam [12]. 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 (NICs), such as Intel 5300. Each set of CSIs characterizes the amplitude and phase of orthogonal frequency division multiplexing (OFDM) subcarriers. Therefore, CSI has better static stability and dynamic sensitivity [13]. Many researchers have begun to use CSI in indoor localization [14–16].

Fingerprints of different locations are generally different. Therefore, many researchers use classification algorithms to implement fingerprint-based localization. As traditional classification algorithms, kNN and Bayesian classification are widely used in fingerprint-based localization. There are many fingerprint-based localizations in Wi-Fi wireless sensing. Sen et al. proposed a fingerprint-based localization system called PinLoc [17] using the distinguishability of CSI at different locations. Xiao et al. used the frequency diversity and spatial diversity of CSI to construct subchannel power information location fingerprints based on CSI amplitude and proposed the FIFS system [18]. Xiao et al. subsequently proposed the first CSI-based passive indoor localization system Pilot [19]. Pilot uses probability estimation to match the obtained CSI with the fingerprint database to achieve position. Chapre et al. used the amplitude differences and phase differences of CSI to construct fingerprints and proposed a Wi-Fi fingerprint system CSI-MIMO [20]. Abdel-Nasser et al. [21] observed that the CSI probability density distribution of different locations nicely fits a Gaussian distribution mixture and used k -means to construct a cluster-based fingerprint database. To improve the localization accuracy, Sabek et al.

propose a single access point-monitoring point- (AP-MP-) based localization system MonoStream [22], which models the passive localization as an object recognition problem.

Compared with traditional methods, deep learning-based methods are more accurate. Wang et al. proposed deep learning-based CSI fingerprint indoor localization systems DeepFi and PhaseFi [23, 24] 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 the radial basis function is used to obtain the estimated location. Chen et al. proposed the first CNN-based Wi-Fi indoor localization system ConFi [25]. 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 are 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. constructed location fingerprints based on the amplitude of CSI and proposed a passive localization system PILC [26].

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

1.2. Our Contributions. In our method, we extract amplitudes of CSI subcarriers from data of different transmitting-

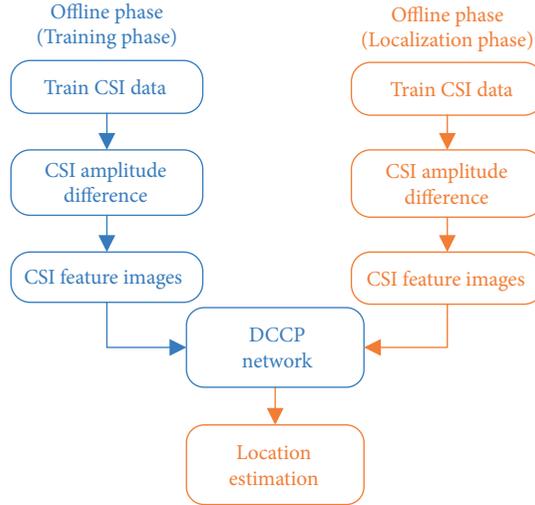


FIGURE 2: DSCP architecture.

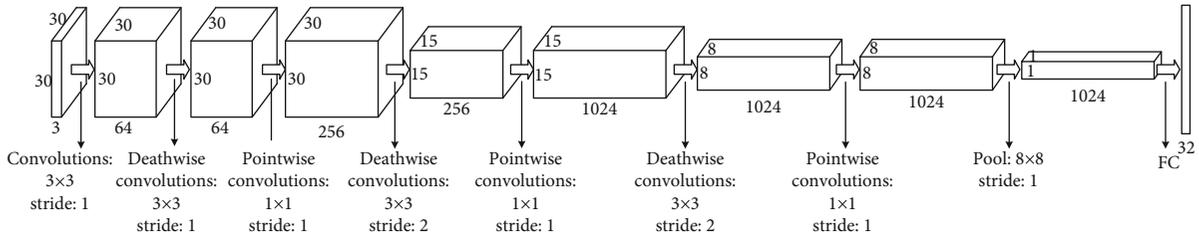


FIGURE 3: DSCP network.

receiving antenna pairs and calculate the differences. Then, we use amplitude differences to construct CSI feature images. Each channel of the CSI feature image corresponds to data of one transmitting-receiving antenna pair. We design a depthwise separable convolution-based network consisting of seven convolutional layers. In the offline training phase, we collect CSI indoors and construct CSI feature images for network training. In the online localization phase, CSI feature images of test data are sent to the trained network to obtain the predicted location.

The main contributions of this paper are as follows:

- (i) This paper proposes to use depthwise separable convolutions in the network to speed up network training and reduce localization delay in an indoor localization system based on a CSI fingerprint
- (ii) The proposed indoor localization system uses CSI feature images constructed from the amplitude difference of CSI subcarriers as location fingerprints. Similar to CSI feature images based on the amplitude of subcarriers, CSI feature images based on the amplitude difference of subcarriers combine time, frequency, and spatial domain information
- (iii) The proposed indoor localization system uses CSI feature images based on the amplitude difference of subcarriers, which can reflect the difference in the signal attenuation of different subcarriers in multiple

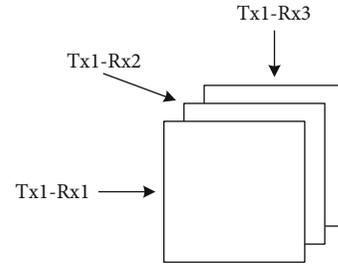


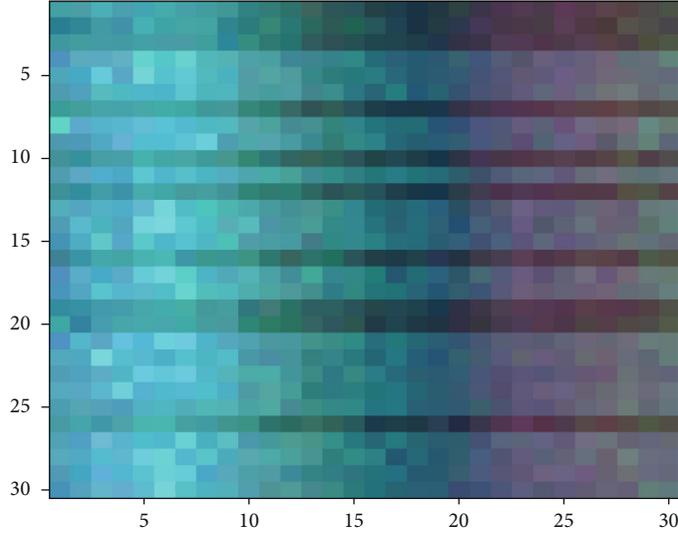
FIGURE 4: Structure of CSI feature image formed by data of 3 antenna pairs.

paths and amplify the differences between different location fingerprints

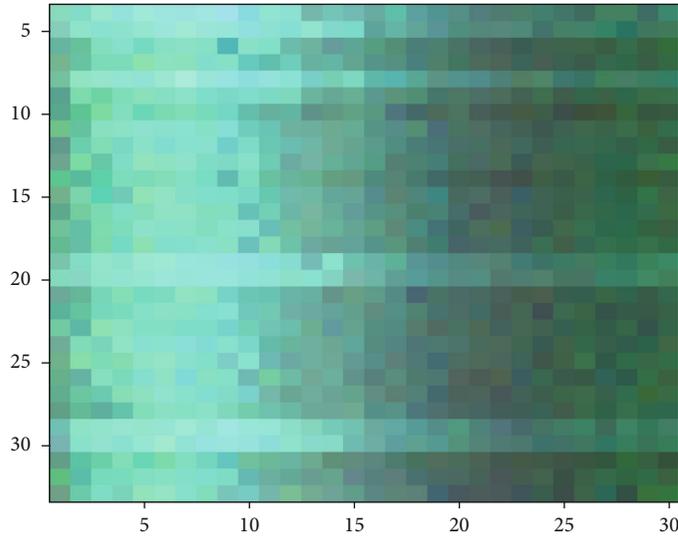
The remainder of this paper is organized as follows. The related work and the preliminary are presented in Section 2. The DSCP system is presented in Section 3. Section 4 provides the experimental validation, while Section 5 concludes the paper.

2. Preliminary

This section introduces the concepts needed for the subsequent analysis, including CSI background information,



(a) CSI fingerprint of location 1



(b) CSI fingerprint of location 2

FIGURE 5: CSI feature images based on subcarrier amplitudes.

fingerprint-based localization, and depthwise separable convolution.

2.1. 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 [27], a signal can be transmitted through a set of subcarriers with different frequencies and mutually orthogonal by OFDM. To estimate the CSI matrix, the transmitter transmits a known pilot sequence x_1, x_2, \dots, x_n , and the combined received signal Y can be expressed as

$$Y = [y_1, y_2, \dots, y_n] = HX + N, \quad (1)$$

where H is the CSI matrix and N is the noise.

Therefore, the CSI matrix can be approximated as

$$\hat{H} = \frac{Y}{X}. \quad (2)$$

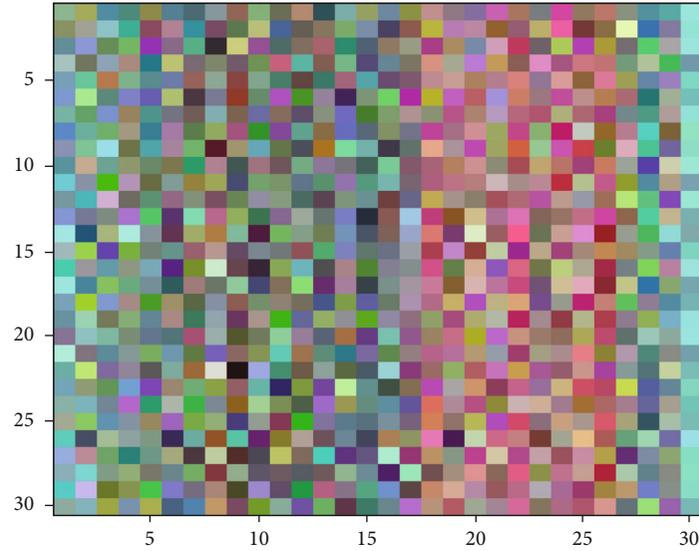
For a system with N subcarriers, H can be expressed as

$$H = [H_1, H_2, \dots, H_N], \quad (3)$$

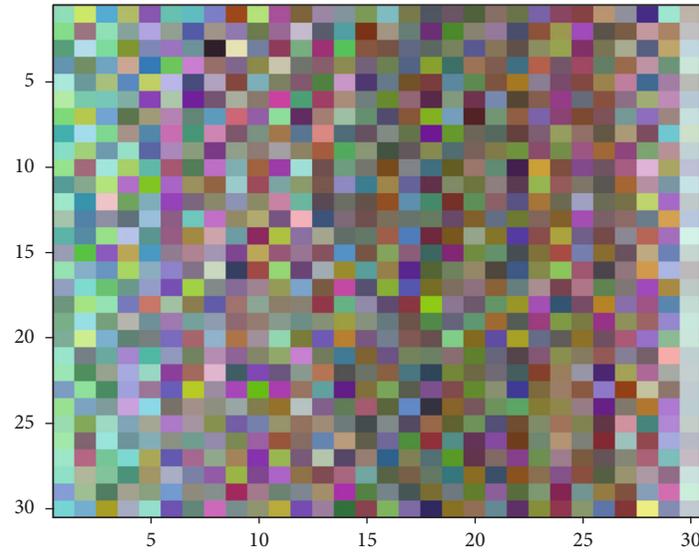
where

$$H_i = |H_i| e^{j \sin^{-1}(\angle H_i)}. \quad (4)$$

For a MIMO system with m transmitting antennas and n



(a) CSI fingerprint of location 1



(b) CSI fingerprint of location 2

FIGURE 6: CSI feature images based on subcarrier amplitudes differences.

receiving antennas, H_i can be expressed as

$$H_i = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{m1} & h_{m2} & \cdots & h_{mn} \end{bmatrix}, \quad (5)$$

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

There are 56 subcarriers in total in the 20 bandwidth and 114 subcarriers in the 40 bandwidth. Grouping is a method

that reduces the size of the CSI Report field by reporting a single value for each group of N_g adjacent subcarriers [27]. The subcarrier frequency spacing is 312.5 in the 20 channel [28]. As shown in Table 1, for a 20 bandwidth, there are three grouping modes defined in IEEE 802.11n. N_s is the number of subcarriers sent. This paper is based on the Wi-Fi CSI in a 20 channel.

2.2. Fingerprint-Based Localization. In an indoor environment, due to walls and obstacles, wireless signals form multipath effects. Humans at different physical locations have different effects on wireless signal propagation paths. The differences can be expressed as fingerprints that reflect the characteristics of humans at different locations [29]. Generally, in the case of small indoor environment changes, a valid location fingerprint needs to meet the following two conditions:

```

Input:  $M$  packets with  $1 \times 3 \times 30$  CSI matrix for each of the  $n$  locations, sliding window size  $s$ .
Output:  $N$  CSI feature images of each location.
for  $i = 1 : 30$  do
  for  $j = 1 : M$  do
    for  $l = 1 : 3$  do
      For antenna pair  $l$ , extract the CSI amplitude  $h_{\text{amp}_{ij}}^{(l)}$ 
      Compute the amplitude difference  $h_{d-\text{amp}_{ij}}^{(l)}$ 
    end
  end
end
 $k = 1$ ;
for  $j = 1 : N$  do
  Construct the CSI feature images  $F_j$  base on  $h_{d-\text{amp}_{1:30k:k+29}}^{(1:3)}$ ;
   $k = k + s$ 
end

```

ALGORITHM 1: Construction of CSI feature images.

TABLE 2: 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	$1 \times 1 \times 32$

- (i) The fingerprints of different locations also have differences
- (ii) The fingerprints of the same location at different times are stable

Fingerprint-based localization is usually divided into two phases: the offline training phase and the online localization phase. In the paper, in the offline training phase of DSCP, we first divide the indoor scenario and set sampling points. The feature of CSI data collected at different sampling points are called location fingerprints. We construct CSI feature images as location fingerprints using amplitude differences of CSI subcarriers. Then, the location fingerprints are processed to construct a fingerprint map. We design DSCP network based on depthwise separable convolutions. The DSCP network is trained by location fingerprints of the fingerprint map during the offline training phase. In the online localization phase of DSCP, the target fingerprint is constructed from the CSI data of target location. The target fingerprint is input into the trained DSCP network. Finally, the predicted location is output by the network.

2.3. Depthwise Separable Convolution. Howard et al. of Google Inc. proposed a lightweight deep neural network, MobileNet [30], 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. As shown in Figure 1, assuming the input image is $H \times W \times M$, where H is the height of the image, W is the width of the image, and M is the number of channels, the size of the output image 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 $(1/N) + (1/K^2)$ of the standard convolution.

Although the calculation cost of a depthwise separable convolution is much lower than that of a standard convolution, the experiments of MobileNet show that the model using depthwise separable convolutions compared to standard convolutions only reduces accuracy by 1% on ImageNet. Depthwise separable convolution 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 to speed up training and improve localization efficiency.

3. The DSCP System

This section introduces the DSCP system, including the architecture of the system, location fingerprints, the DSCP network structure, time complexity of the network, and location estimation.

3.1. System Architecture. As shown in Figure 2, the DSCP system architecture includes an offline phase and an online phase. In the offline phase, the CSI data of each training

```

Input: CSI feature image of the training data  $F_1, F_2, \dots, F_m$  and CSI feature image of the target  $F_{\text{target}}$ , location of  $n$  trained locations  $L_1, L_2, \dots, L_n$ 
Output: Location of the target.
For  $i = 1 : m$  do
  Train DSCP network using CSI feature image of the training data  $F_1, F_2, \dots, F_m$ ;
end
 $L_{\text{target}} = (0, 0)$ ;
for  $i = 1 : n$  do
  Compute the probability of the target at location  $iP_i$  by the trained DSCP network;
end
for  $i = 1 : n$  do
   $L_{\text{target}} = L_{\text{target}} + P_i \times L_i$ ;
end

```

ALGORITHM 2: DSCP network training and location estimation.

TABLE 3: Parameter setting of Conv_DSCP network.

Type/stride	Input size	Filter shape	Output size
conv_1	$30 \times 30 \times 3$	$3 \times 3 \times 3 \times 64$	$30 \times 30 \times 64$
conv_2/s1	$30 \times 30 \times 64$	$3 \times 3 \times 64 \times 256$	$30 \times 30 \times 256$
conv_3/s2	$30 \times 30 \times 256$	$3 \times 3 \times 256 \times 1024$	$15 \times 15 \times 1024$
conv_4/s2	$15 \times 15 \times 1024$	$3 \times 3 \times 1024 \times 1024$	$8 \times 8 \times 1024$
Avg pool/s1	$8 \times 8 \times 1024$	Pool 8×8	$1 \times 1 \times 1024$
FC	$1 \times 1 \times 1024$	1024×32	$1 \times 1 \times 32$

location are 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 are collected, and the fingerprint of the target location is extracted to be input into the trained network. Finally, the predicted location is output by the DSCP network. The DSCP network has seven convolutional layers, as shown in Figure 3.

3.2. Location Fingerprint. Thirty 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 amplitudes and phases on the stream. For each transmitting-receiving antenna pair, a sliding window of size M is used to select M consecutive data packets in a sequence. To match the 30 subcarriers, we make the height and width of CSI feature images the same. So we set M to 30, which means subcarriers of 30 consecutive data packets are chosen to construct one 30×30 matrix. Multiple sets of 30×30 matrices are constructed. The structure of CSI feature images is shown in Figure 4. Similar to RGB images, the data of each antenna pair correspond to one channel, and each location has some CSI feature images of $30 \times 30 \times 3$ [25]. As shown in Figure 5, the CSI feature images formed by the amplitudes of subcarriers can reflect the difference of adjacent locations.

The amplitudes of the subcarriers characterize the power fading of the wireless signals between the transmitter and

receiver [31]. The signal received by the receiver also contains measurement noise. The frequency difference between adjacent subcarriers is 625 kHz in our experiments. Different frequency subcarriers are affected differently by the multipath effect, and the difference can be reflected in the subcarrier amplitude. ConFi [25] is the first CNN-based Wi-Fi indoor localization system, which proposes to use CSI feature images as location fingerprints. ConFi uses amplitudes of subcarriers from consecutive data packets to construct CSI feature images. ConFi maps the CSI feature subimages from the different antenna pairs into the RGB channels of the image. The elements in the row are amplitudes of 30 subcarriers in one packet. The experiments of ConFi show that CSI feature image of size 30×30 gets the highest accuracy, which is about 90% on test set. Therefore, CSI feature images based on amplitude of CSI subcarriers are valid location fingerprints. In indoor environments, wireless signals are affected by multipath effects. We consider that the amplitudes of the subcarriers are greatly affected by the physical environment, which may result in little influence of humans on CSI. Different from ConFi, to amplify the differences between fingerprints of different locations, we use the amplitude differences between adjacent subcarriers to construct CSI feature images at different locations. The CSI feature images combine the 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 Figure 6).

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

$$h_{\text{amp}_{ij}} = |h_{pq}|_{ij}. \quad (6)$$

The amplitude difference between adjacent subcarriers is

$$h_{d-\text{amp}_{ij}} = \begin{cases} h_{\text{amp}_{i+1j}} - h_{\text{amp}_{ij}} & (i \in [1, 29]), \\ h_{\text{amp}_{1j}} - h_{\text{amp}_{30j}} & (i = 30). \end{cases} \quad (7)$$

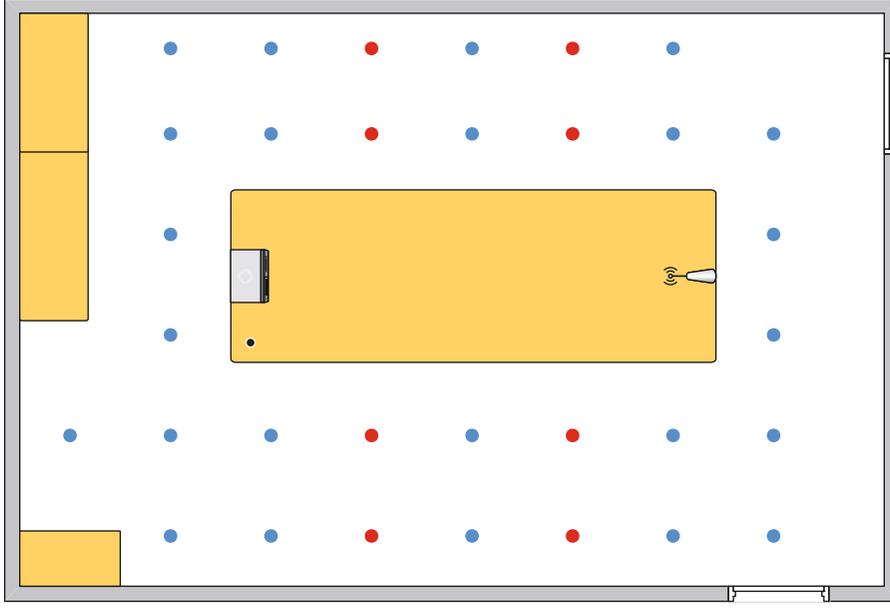


FIGURE 7: Scenario 1 is a laboratory.

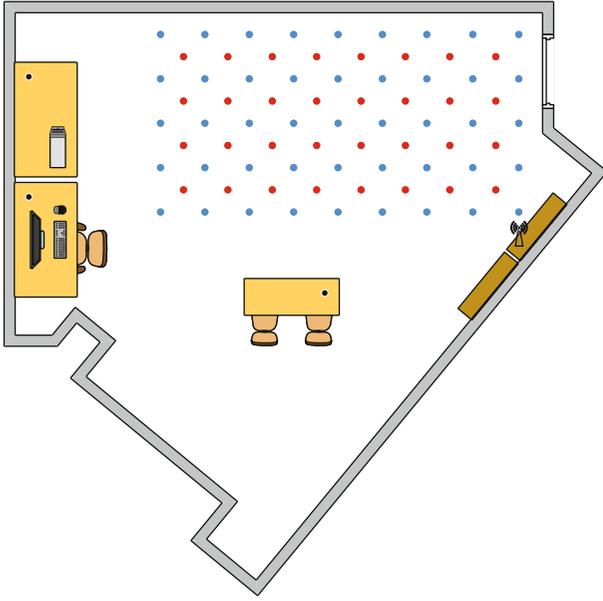


FIGURE 8: Scenario 2 is a laboratory.

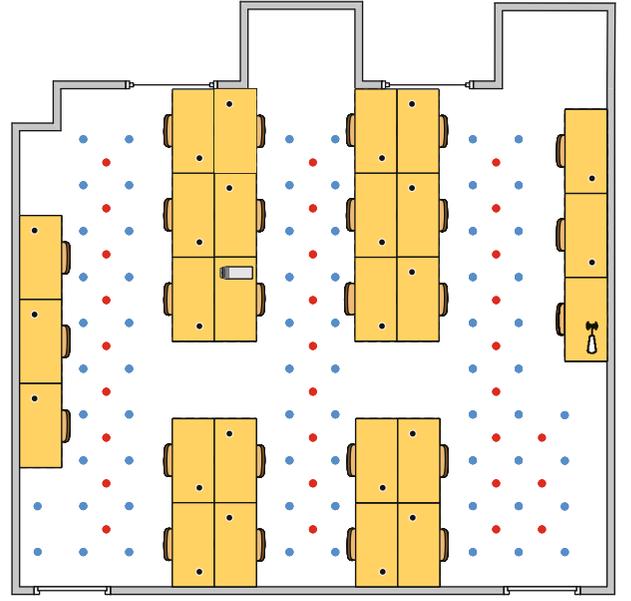


FIGURE 9: Scenario 3 is an office.

Then, a CSI feature image can be expressed as

$$F = \begin{bmatrix} h_{d\text{-amp}_{1k}} & h_{d\text{-amp}_{2k}} & \cdots & h_{d\text{-amp}_{Nk}} \\ h_{d\text{-amp}_{1k+1}} & h_{d\text{-amp}_{2k+1}} & \cdots & h_{d\text{-amp}_{Nk+1}} \\ \vdots & \vdots & \ddots & \vdots \\ h_{d\text{-amp}_{1k+29}} & h_{d\text{-amp}_{2k+29}} & \cdots & h_{d\text{-amp}_{Nk+29}} \end{bmatrix}. \quad (8)$$

The pseudocode for CSI feature image construction is presented in Algorithm 1. The input of the algorithm

includes M packets with $1 \times 3 \times 30$ CSI matrix for each of the n locations and sliding window size s . First, for data of each transmitting-receiving antenna pair, we extract the amplitude of CSI subcarriers from each packet (line 4). Then, we compute the amplitude difference (line 5). Finally, we use a sliding window to group the amplitude difference of 30 packets to construct n CSI feature images of each location (lines 10-13).

3.3. Network Structure. Inspired by MobileNet, we design the DSCP network. The DSCP network is based on depthwise separable convolutions, which have seven convolutional

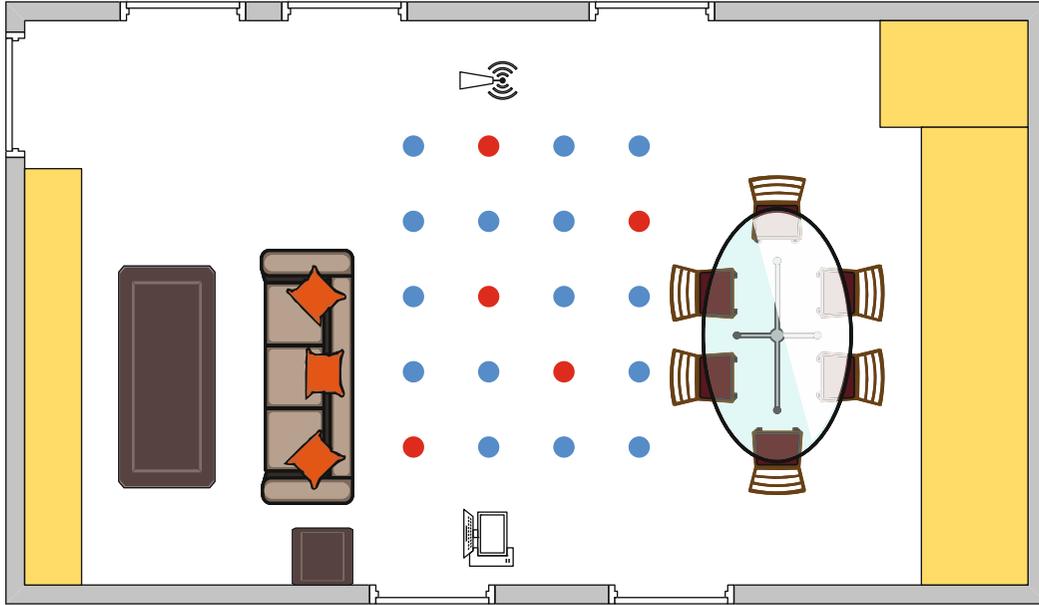


FIGURE 10: Scenario 4 is a living room.

layers. The first convolutional layer of the DSCP network is a full convolution, and the others are depthwise separable convolutions. Table 2 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 strides 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 sizes 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 [32] is an effective random optimization algorithm with high computational efficiency and low memory requirements. We use Adam optimization in the 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 domains.

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.

3.4. 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. Therefore, we use the weighted mean method 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

$$\hat{L} = \sum_{i=1}^N \mathbb{P}P_i \times L_i. \quad (9)$$

The pseudocode for network training and location estimation is presented in Algorithm 2. The input of the algorithm are the CSI feature images for training and CSI feature image of the target. First, CSI feature images are used to train the DSCP network (lines 1-3). Then, the probability of the target at each training location will be computed by sending the CSI feature image of the target to the trained DSCP network (line 6). Then, we use the method of weighted mean based on probability to estimate the location of the target (lines 8-10).

3.5. Time Complexity of the Network. The total time complexity of all convolutional layers [33] is

$$\mathcal{O} \left(\sum_{l=1}^d \mathbb{N}n_{l-1} \cdot s_l^2 \cdot n_l \cdot m_l^2 \right), \quad (10)$$

where l is the index of a convolutional layer, d is the number of convolutional layers, n_l is the output channels of the l th layer, m_l is the length of the output feature

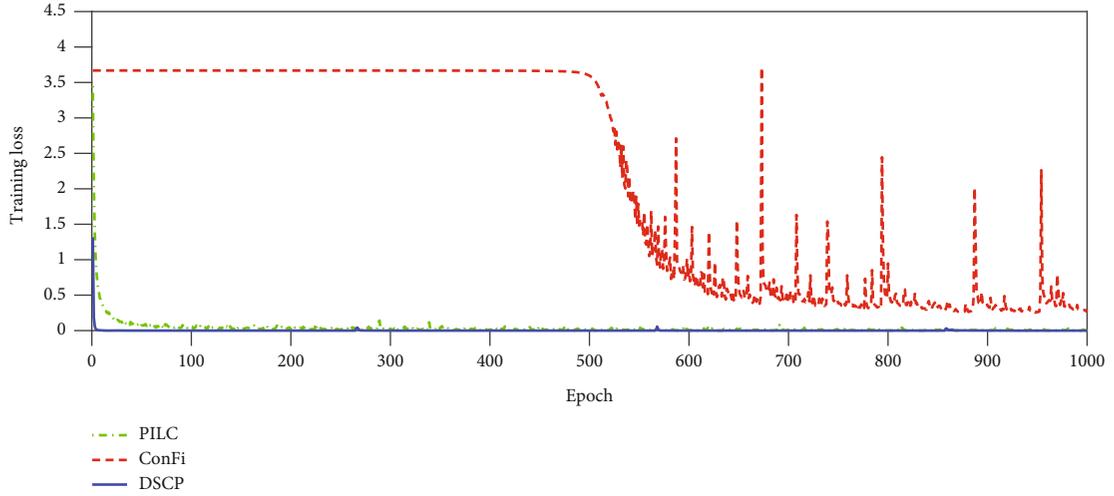


FIGURE 11: Training loss of methods based on CNN.

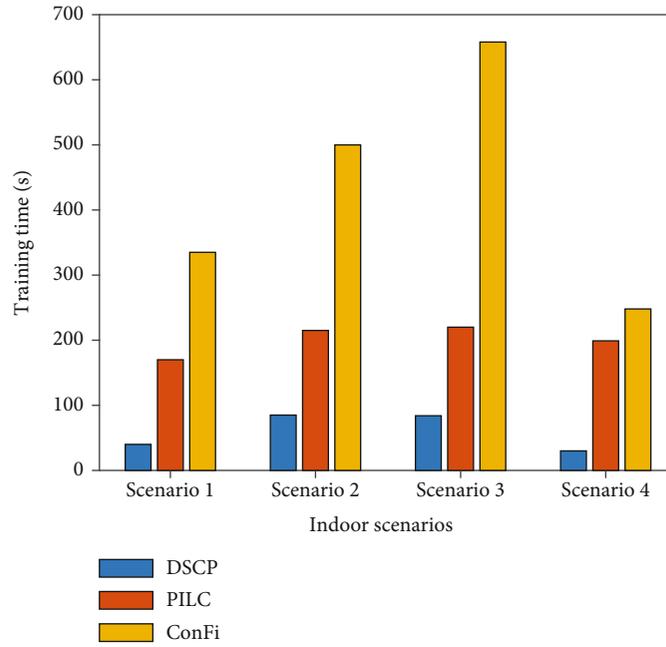


FIGURE 12: Training time of models based on CNN.

TABLE 4: Training time of models based on CNN.

Localization method	Training time			
	Scenario 1	Scenario 2	Scenario 3	Scenario 4
DSCP	40 s	85 s	84 s	30 s
PILC	170 s	215 s	220 s	199 s
ConFi	335 s	500 s	658 s	248 s

image, and s_f is the length of the filter. The time complexity of the network is the sum of the time complexity of each layer.

If only using standard convolution kernels, the DSCP network will have 4 convolutional layers, and the structure is shown in Table 3. We estimate the time complexity of the DSCP and Conv_DSCP networks. For simplicity, we only consider the convolutional layers. Assume the size of the input image is $S \times S$, and the calculation cost of Conv_DSCP is approximately $4,867,776 S^2$. The DSCP is approximately $163,397.5 S^2$, which is only 3.36% of the Conv_DSCP. Thus, the DSCP can achieve a short low network training time and low localization delay.

4. Experiment Validation

To illustrate the performance of the proposed DSCP method, a set of comparisons are performed with the other three

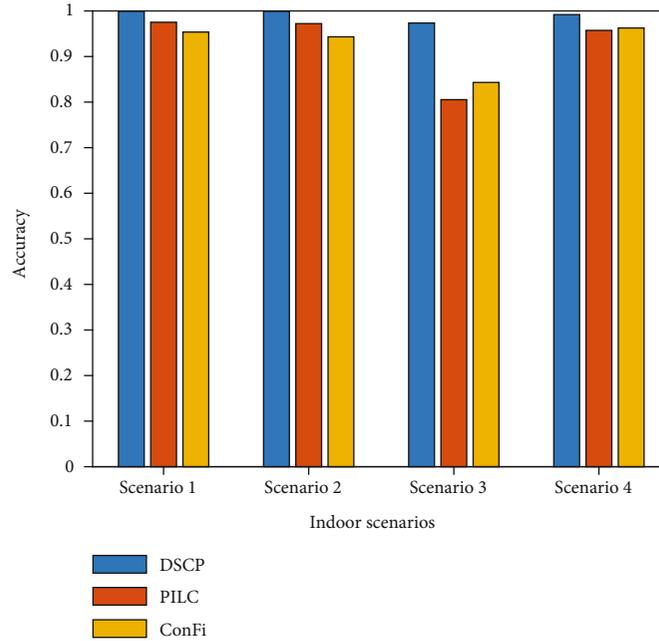


FIGURE 13: Accuracy of methods based on CNN.

TABLE 5: Localization accuracy of the methods based on CNN.

Localization method	Localization accuracy			
	Scenario 1	Scenario 2	Scenario 3	Scenario 4
DSCP	99.93%	99.94%	97.35%	99.20%
PILC	97.52%	97.22%	80.54%	95.73%
ConFi	95.37%	94.33%	84.32%	96.27%

methods, including the two methods ConFi [25] and PILC [26] based on CNN and the other method DeepFi [23] based on deep learning. The comparison analysis is based on the same illustrative examples.

4.1. Experiments Setup. In the experiments, we use a TP-LINK router as the transmitter. A Dell commercial desktop PC equipped with Intel 5300 NIC and Ubuntu 14.04 is used as the receiver. The TensorFlow is accelerated with a NVIDIA RTX2080 graphics card during the experiments. The sampling rate is set to 50 Hz. DSCP is a single-target localization system. When collecting data, there is only one person in the room. We wrote a script to perform data collection. The tester stands at each location for 60 seconds, and the receiver can automatically collect 3,000 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. Due to the hardware, there are errors in the phase information acquired by the common commercial network card, and the original phase information is often not directly used. It is common to calculate the exact phase shift by means of multiple linear regression and then experi-

ment with the phase shift. To simplify the system, we only use the amplitude information of CSI.

In this paper, four indoor scenarios were selected for experiments (as shown in Figures 7–10). The sampling points were divided into training locations and test locations marked as blue and red, respectively. Scenario 1 was a $5.8 \text{ m} \times 8.6 \text{ m}$ laboratory. A total of 32 sampling points were set. The spacing between adjacent sampling points was 1 m. Twenty-four locations were selected for training and 8 for testing. Scenario 2 was a 40 m^2 laboratory. Forty-five locations were selected for training, and 32 locations were selected for testing. Scenario 3 was an office of $7.5 \text{ m} \times 6.5 \text{ m}$. Sixty-six locations were selected for training and 30 for testing. The spacing between adjacent sampling points was 0.6 m in both Scenarios 2 and 3. Scenario 4 was a typical living room of $4 \text{ m} \times 8 \text{ m}$. Fifteen locations are selected for training and 5 for testing. The spacing between adjacent sampling points was 0.5 m.

4.2. The Comparison Works. The performance of DSCP was evaluated in four scenarios. We compared DSCP with three deep learning-based methods, PILC [26], ConFi [25], and DeepFi [23]. ConFi was the first CNN-based Wi-Fi indoor localization system. PILC and ConFi both use CSI amplitudes to construct CSI feature images, which have three channels. The data of different transmitting-receiving antenna pairs are used in a MIMO system. PILC is a passive localization system based on CNN. The PILC network has six layers. DeepFi uses CSI information for all the subcarriers from three antennas and proposes to use the weights in the deep network to represent fingerprints. DeepFi uses a probabilistic data fusion method based on the radial basis for online location estimation.

For DSCP, experiments based on CSI amplitude and amplitude differences were performed. The median distance

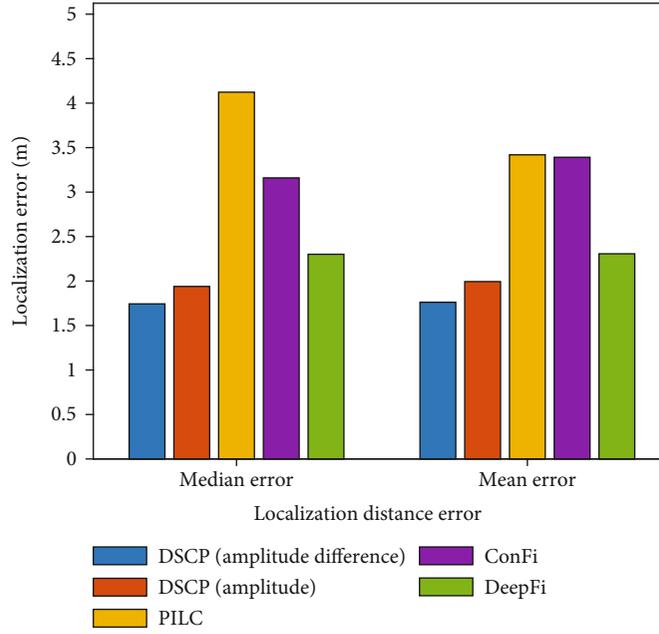


FIGURE 14: Localization error in Scenario 1.

TABLE 6: Localization distance error.

Localization method	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	Mean error	Median error						
DSCP (amplitude-diff)	1.7619 m	1.7435 m	1.1584 m	0.9789 m	2.5406 m	2.6307 m	0.9052 m	0.6939 m
DSCP (amplitude)	1.9938 m	1.9394 m	1.3579 m	1.2180 m	2.6348 m	2.5119 m	1.1243 m	0.9721 m
PILC	3.4191 m	4.1231 m	1.7869 m	1.7122 m	3.0418 m	3.2211 m	1.1070 m	1.0000 m
ConFi	3.3915 m	3.1597 m	1.7503 m	1.6028 m	3.2408 m	3.0000 m	1.3269 m	1.4133 m
DeepFi	2.3065 m	2.3011 m	1.4059 m	1.3939 m	2.5314 m	2.7524 m	0.8259 m	0.8479 m

error and the mean distance error of different localization methods were calculated. The median distance error was the median of all the test case localization errors. The mean distance error is the average of all the test case localization errors.

4.3. Model Evaluation. DSCP was compared with two CNN-based indoor localization methods, PILC and ConFi. As shown in Figure 11, the training loss of PILC started to converge when the number of epochs was approximately 50. The training loss of ConFi converged when the number of epochs was approximately 1,000. The training loss of the DSCP converged when the number of epochs is approximately 20. In subsequent experiments, the values of the epoch were set according to convergence conditions of different networks. The network training time of the three localization methods based on the CNN is shown in Figure 12 and Table 4. The time unit is second.

In this experiment, 75% of all training location data were the training set, and the remaining 25% were the test set. The localization accuracies of DSCP, PILC, and ConFi were calculated. The batch sizes of DSCP and PILC were the numbers of training locations. The epochs of DSCP, PILC, and ConFi were 30, 60, and 1,200, respectively. The batchsize of ConFi

was 256. As shown in Figure 13 and Table 5, the localization accuracies of the DSCP in Scenarios 1 and 2 were both high, above 99%, and the localization accuracy in Scenario 3 reaches 97%. The localization accuracy of the DSCP in each scenario was higher than those of the other two methods.

4.4. Localization Performance

Scenario 1. Figure 14 and Table 6 show the localization errors of the four methods in Scenario 1. Scenario 1 was a typical laboratory with some indoor furniture. It can be seen from the experiments that DSCP obtained a median error of 1.75 m and a mean error of 1.77 m. Compared with other methods based on deep learning, the localization error of DSCP was smaller. DSCP was 57.71% lower than PILC in median distance error, 44.82% lower than ConFi, and 24.23% lower than DeepFi. Figure 15 presents the cumulative distribution function (CDF) of localization errors of the four methods in Scenario 1. With DSCP, approximately 40% of the test cases had an error under 1.5 m, and approximately 73% of the test cases had an error under 2 m.

Scenario 2. Figure 16 and Table 6 show the localization errors in Scenario 2. There was no obstacle in the experimental area

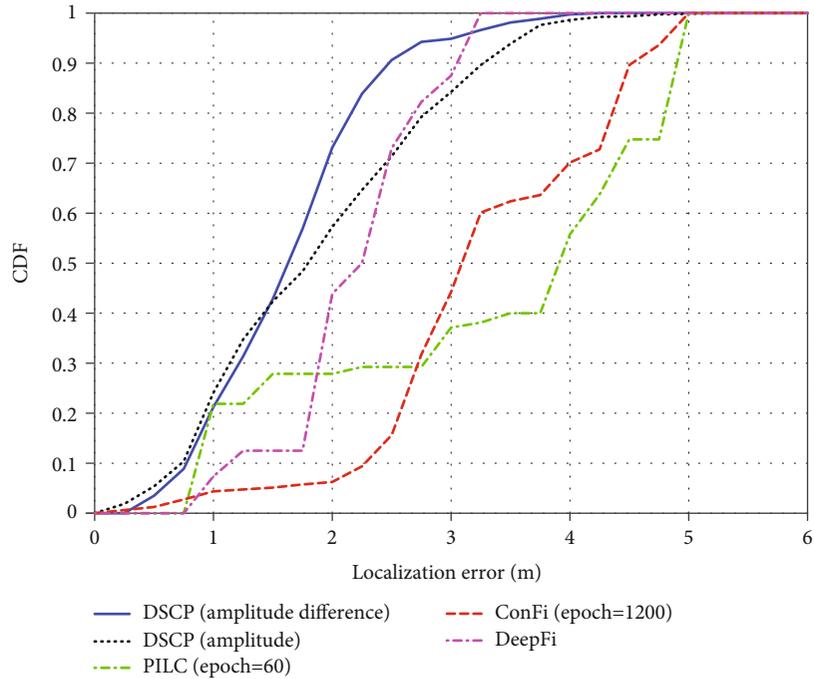


FIGURE 15: CDF of localization errors in Scenario 1.

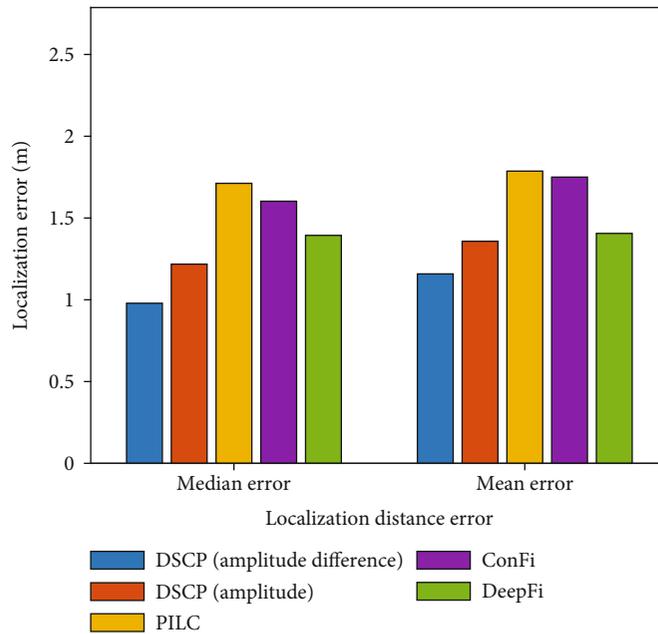


FIGURE 16: Localization error in Scenario 2.

of Scenario 2, and there were line of sight (LOS) paths between the transmitter and the receiver. As the experiments show in Scenario 2, the DSCP obtained a median error of 0.98 m and a mean error of 1.16 m. DSCP was 42.83% lower than PILC in median distance error, 38.93% lower than ConFi, and 29.77% lower than DeepFi. Figure 17 presents the CDF of localization errors of the four methods in Scenario 2. With DSCP, 55% of the test cases had an error under

1 m, and approximately 90% of the test cases have an error under 2 m.

Scenario 3. Figure 18 and Table 6 show the localization errors in Scenario 3. Scenario 3 was a typical office with a large number of office chairs in the room. The experimental results show that DSCP obtained a median error of 2.63 m and a mean error of 2.54 m. DSCP was 18.33% lower than PILC

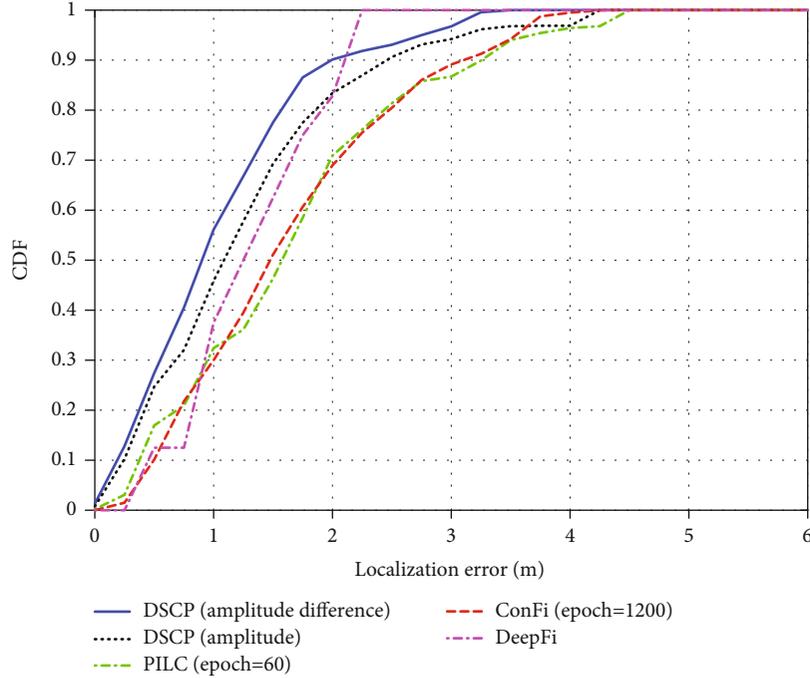


FIGURE 17: CDF of localization errors in Scenario 2.

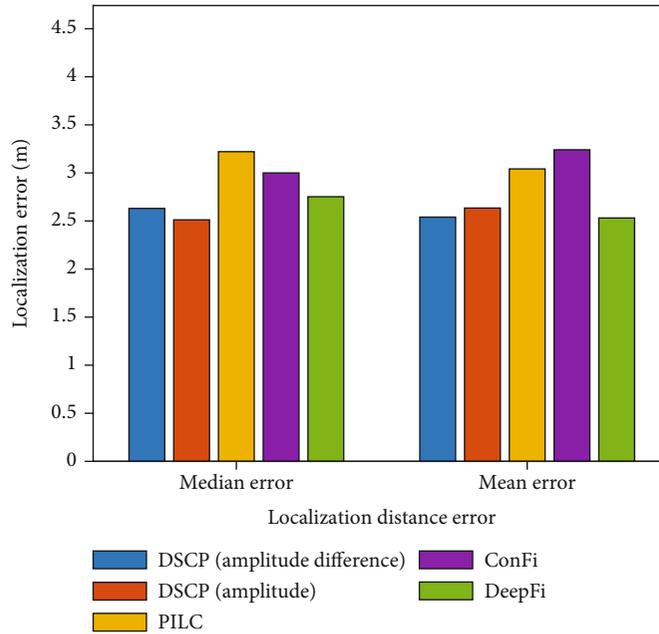


FIGURE 18: Localization error in Scenario 3.

in median distance error, 12.31% lower than ConFi, and 4.42% lower than DeepFi. Figure 19 depicts the CDF of localization errors of the four methods in Scenario 3. With DSCP, 30% of the test cases had an error under 2 m, and approximately 75% of the test cases have an error under 3 m.

Scenario 4. Figure 20 and Table 6 show the localization errors in Scenario 4. Scenario 4 was a living room with ordinary furniture in the room. The experimental results show that DSCP

can obtain a median error of 0.69 m and a mean error of 0.91 m. DSCP was 30.61% lower than PILC in median distance error and 50.90% lower than ConFi. Figure 21 depicts the CDF of localization errors of the four methods in Scenario 4. With DSCP, 45% of the test cases had an error under 0.5 m, and approximately 75% of the test cases have an error under 1 m. The experiments show that the four methods achieved better performance in Scenarios 2 and 4 than the other two scenarios. The reason may be that there was no

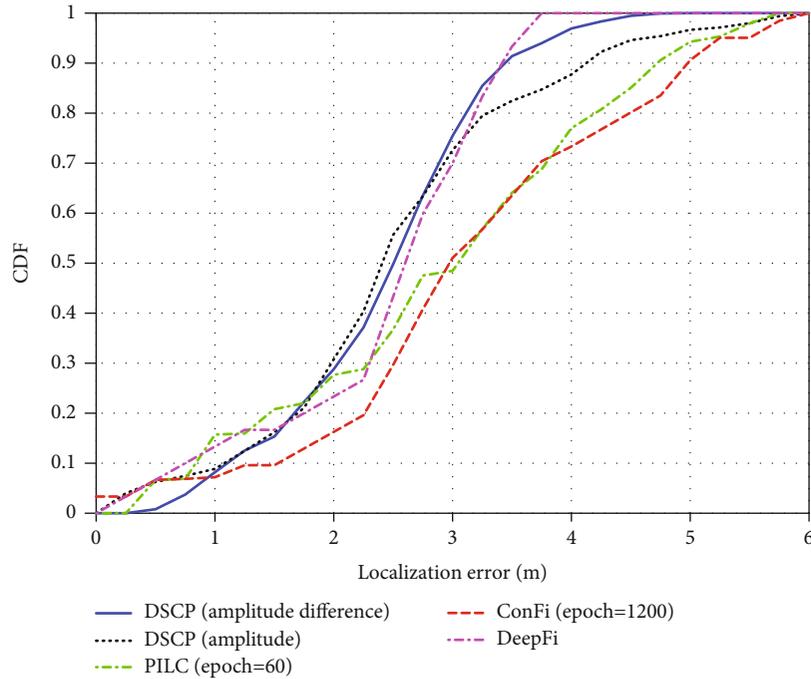


FIGURE 19: CDF of localization errors in Scenario 3.

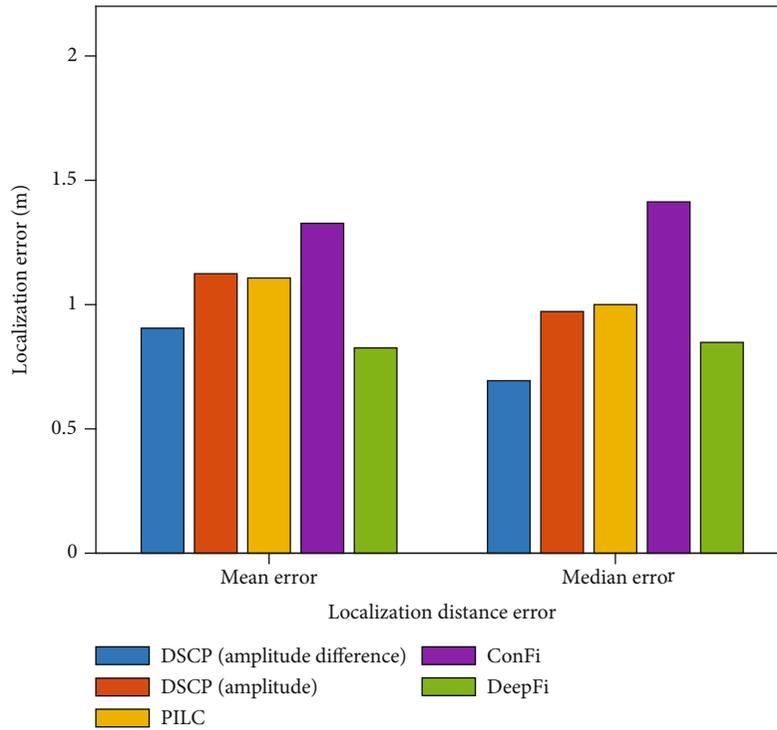


FIGURE 20: Localization error in Scenario 4.

obstacle between the transmitter and the receiver to block wireless signals in Scenarios 2 and 4. In Scenario 3, the multipath propagation of the wireless signal was more complicated, so the localization performance was not as good as that in the other scenarios. The location fingerprints based

on the CSI amplitude differences were slightly better than those based on amplitudes. The experiments show that the four methods achieve better performance in Scenarios 2 and 4 than the other two scenarios. The reason may be that there is no obstacle between the

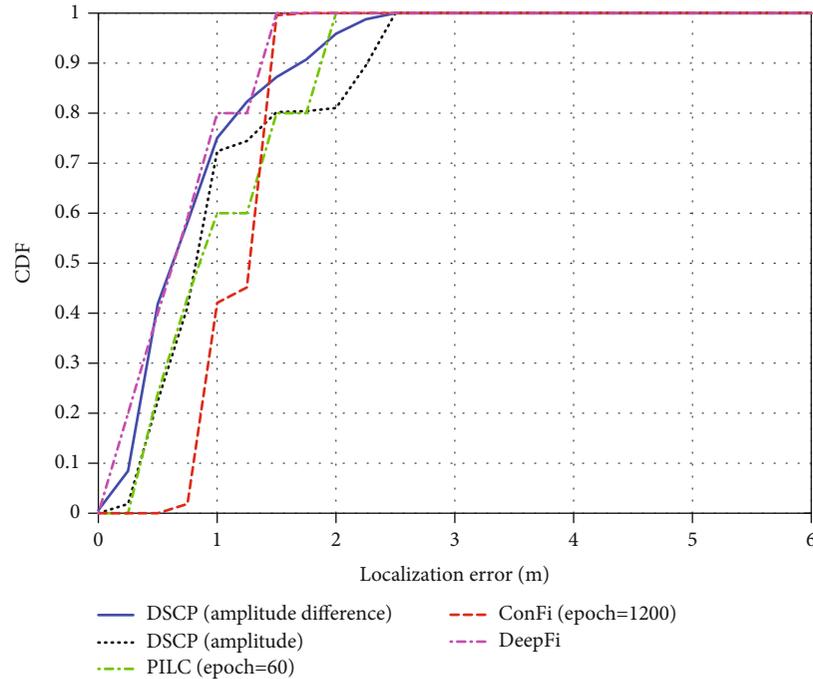


FIGURE 21: CDF of localization errors in Scenario 4.

transmitter and the receiver to block wireless signals in Scenarios 2 and 4. 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.

5. Conclusion

This paper proposed a passive CSI indoor localization system DSCP by Wi-Fi wireless sensing. A neural network based on depthwise separable convolutions in fingerprint-based localization was adopted to speed up network training and reduce localization delay. We conducted the experiments in four typical indoor scenarios. DSCP was compared with several existing CSI fingerprint-based indoor localization methods. However, in different indoor scenarios, the experimental results will be affected by different environments. The experimental results showed that DSCP achieved better performance than other methods in terms of the localization accuracy and localization error. DSCP was also more efficient than other CNN-based localization methods in network training and localization. A possible future extension is using both the amplitude and phase of CSI subcarriers when constructing CSI feature images and then using CSI feature images as location fingerprints in a depthwise separable convolution-based passive indoor localization system.

Data Availability

The CSI fingerprint data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

An earlier and preliminary version of this paper was presented as conference in 2020 IEEE Wireless Communications and Networking Conference (WCNC), <https://ieeexplore.ieee.org/document-/9120638>.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work is partly supported by the National Natural Science Foundation of China under grant nos. 61702284, 61873131, and 61872194 and Postgraduate Research & Practice Innovation Program of Jiangsu Province under grant no. KYCX19_0972.

References

- [1] F. Xiao, L. Chen, H. Zhu, R. Hong, and R. Wang, "Anomaly-tolerant network traffic estimation via noise-immune temporal matrix completion model," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1192–1204, 2019.
- [2] 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, 2019.
- [3] M. Zhou, Y. Wang, Y. Liu, and Z. Tian, "An information-theoretic view of WLAN localization error bound in GPS-denied environment," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 4089–4093, 2019.
- [4] Y. Liu, W. Xiong, Z. Zhu, and S. Li, "Csi based high accuracy device free passivelocalization system," in *2018 IEEE 88th*

- Vehicular Technology Conference (VTC-Fall)*, pp. 1–5, Chicago, IL, USA, August 2018.
- [5] R. Want, A. Hopper, V. Falcão, and J. Gibbons, “The active badge location system,” *ACM Transactions on Information Systems*, vol. 10, no. 1, pp. 91–102, 1992.
 - [6] F. Xiao, Z. Wang, N. Ye, R. Wang, and X. Li, “One more tag enables fine-grained RFID localization and tracking,” *IEEE/ACM Transactions on Networking*, vol. 26, no. 1, pp. 161–174, 2018.
 - [7] P. Dai, Y. Yang, M. Wang, and R. Yan, “Combination of DNN and improved KNN for indoor location fingerprinting,” *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 4283857, 9 pages, 2019.
 - [8] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, “Spotfi: decimeter level localization using wifi,” in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, ser. SIGCOMM’15*. New York, NY, USA: Association for Computing Machinery, pp. 269–282, New York, NY, USA, 2015.
 - [9] X. Yalong, Z. Shigeng, W. Jianxin, and Z. Chengzhang, “A novel indoor localization algorithm for efficient mobility management in wireless networks,” *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 9517942, 12 pages, 2018.
 - [10] A. Achroufene, Y. Amirat, and A. Chibani, “RSS-based indoor localization using belief function theory,” *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 3, pp. 1163–1180, 2019.
 - [11] 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, pp. 775–784, Tel Aviv, Israel, 2000.
 - [12] 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, 2017.
 - [13] C. Han, Q. Tan, L. Sun, H. Zhu, and J. Guo, “Csi frequency domain fingerprint-based passive indoor human detection,” *Information*, vol. 9, no. 4, p. 95, 2018.
 - [14] S. Shi, S. Sigg, and Y. Ji, “Probabilistic fingerprinting based passive device-free localization from channel state information,” in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Nanjing, China, May 2016.
 - [15] B. Berruet, O. Baala, A. Caminada, and V. Guillet, “An evaluation method of channel state information fingerprinting for single gateway indoor localization,” *Journal of Network and Computer Applications*, vol. 159, p. 102591, 2020.
 - [16] X. Wang, X. Wang, and S. Mao, “Cifi: deep convolutional neural networks for indoor localization with 5 GHz Wi-Fi,” in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, Paris, France, May 2017.
 - [17] 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, pp. 183–196, New York, NY, USA, 2012.
 - [18] J. Xiao, K. Wu, Y. Yi, and L. M. Ni, “FIFS: fine-grained indoor fingerprinting system,” in *2012 21st International Conference on Computer Communications and Networks (ICCCN)*, pp. 1–7, Munich, Germany, July 2012.
 - [19] J. Xiao, K. Wu, Y. Yi, L. Wang, and L. M. Ni, “Pilot: passive device-free indoor localization using channel state information,” in *2013 IEEE 33rd International Conference on Distributed Computing Systems (ICDCS)*, pp. 236–245, Philadelphia, PA, USA, July 2013.
 - [20] 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.
 - [21] H. Abdel-Nasser, R. Samir, I. Sabek, and M. Youssef, “Mono-PHY: mono-stream-based device-free WLAN localization via physical layer information,” in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 4546–4551, Shanghai, China, April 2013.
 - [22] I. Sabek and M. Youssef, “Monostream: A minimal-hardware high accuracy device-free WLAN localization system,” *CoRR*, p. 11, 2013, <https://arxiv.org/abs/1308.0768>.
 - [23] 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, 2017.
 - [24] 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, 2016.
 - [25] 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. 18066–18074, 2017.
 - [26] 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)*, pp. 1–6, Wuhan, China, March 2018.
 - [27] IEEE, “IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,” in *IEEE Std 802.11-2012 (Revision of IEEE Std 802.11-2007)*, pp. 1–2793, New York, NY, USA, March 2012.
 - [28] I. Lu and K. Tsai, “Channel estimation in a proposed IEEE802.11n OFDM MIMO WLAN system,” in *2007 IEEE Sarnoff Symposium*, pp. 1–5, Princeton, NJ, USA, April 2007.
 - [29] 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.
 - [30] A. G. Howard, M. Zhu, B. Chen et al., “MobileNets: efficient convolutional neural networks for mobile vision applications,” *CoRR*, p. 9, 2017, <https://arxiv.org/abs/1704.04861>.
 - [31] J. Wang, H. Jiang, J. Xiong et al., “Lifs: low human effort, device-free localization with fine-grained subcarrier information,” in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, ser. MobiCom’16*. New York, USA: Association for Computing Machinery, pp. 243–256, New York, NY, USA, 2016.
 - [32] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization,” 2014, <https://arxiv.org/abs/1412.6980>.
 - [33] K. He and J. Sun, “Convolutional neural networks at constrained time cost,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5353–5360, Boston, MA, USA, June 2015.