IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 14, 2017, accepted August 30, 2017, date of publication September 7, 2017, date of current version September 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2749516

ConFi: Convolutional Neural Networks Based Indoor Wi-Fi Localization Using Channel State Information

HAO CHEN[1], YIFAN ZHANG[1], WEI LI[2], XIAOFENG TAO[3], AND PING ZHANG[1]

[1]State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Department of Electrical Engineering, Northern Illinois University, DeKalb, IL 60115, USA
[3]National Engineering Laboratory for Mobile Network Technologies, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Hao Chen (chenhao@bupt.edu.cn)

This work was supported in part by the National Natural Science Foundation for Distinguished Young Scholar of China under Grant 61325006, in part by the National Natural Science Foundation of China under Grant 61231009, and in part by the Shenzhen Science and Technology Project under Grant JCYJ20160531173517680.

ABSTRACT As the technique that determines the position of a target device based on wireless measurements, Wi-Fi localization is attracting increasing attention due to its numerous applications and the widespread deployment of Wi-Fi infrastructure. In this paper, we propose ConFi, the first convolutional neural network (CNN)-based Wi-Fi localization algorithm. Channel state information (CSI), which contains more position related information than traditional received signal strength, is organized into a time-frequency matrix that resembles image and utilized as the feature for localization. The ConFi models localization as a classification problem and addresses it with a five layer CNN that consists of three convolutional layers and two fully connected layers. The ConFi has a training stage and a localization stage. In the training stage, the CSI is collected at a number of reference points (RPs) and used to train the CNN via stochastic gradient descent algorithm. In the localization stage, the CSI of the target device is fed to the CNN and the localization result is calculated as the weighted centroid of the RPs with high output value. Extensive experiments are conducted to select appropriate parameters for the CNN and demonstrate the superior performance of the ConFi over existing methods.

INDEX TERMS Wi-Fi localization, channel state information, convolutional neural network, pattern recognition.

I. INTRODUCTION

As the task of positioning a target device in indoor environment, indoor localization has a wide range of applications such as indoor navigation and people flow monitoring. A number of emerging technologies, including visible light, infrared ray and radio frequency identification (RFID), have been applied in this field. Among them, Wi-Fi based indoor localization stands out due to the widespread deployment of Wi-Fi infrastructures and its potential of being deployed in a transparent manner to users. Various Wi-Fi localization methods are proposed, including angle of arrival based method [1], time of arrival based method [2], and signal propagation model based method [3]. However, fingerprint based localization methods produce the best performance [4] and become the focus of research.

First proposed by RADAR [5], fingerprint based localization methods use certain measurement of Wi-Fi signal as feature and try to capture the difference in the feature across different positions. These methods generally consist of two stages, i.e., a training stage and a localization stage. In the training stage, features are collected at a set of reference points (RPs) and used to train or fit a localization model. In the localization stage, the position of the target device is decided by feeding its feature to the localization model. Therefore, feature utilization and the design of localization model are the core of fingerprint based localization.

The received signal strength (RSS) was widely utilized as a feature in localization [5]–[8], as RSS can be obtained easily at the PHY service access point of Wi-Fi receiver. In RADAR [5], RSS is measured at a number of RPs and localization is conducted by measuring the similarity between the RSS of the target device and the RPs using Euclidean distance. However, RSS has two drawbacks. Firstly, it is sensitive to time varying multipath fading, which results in confusion in

localization results. More importantly, RSS finds it hard to cope with device heterogeneity, which is the phenomenon that different devices such as cell phone and laptop have different transmission parameters such as maximum power and antenna characteristics. As device heterogeneity usually results in difference in RSS even for the same position, model trained using one device may not perform well for another device. To deal with these difficulties, various methods are proposed. Instead of using raw RSS directly, an alternative is to preprocess RSS by normalization and centralization, and calculate statistics such as maximum value, average value, difference between the measurements at different access points (AP). Moreover, dimensionality reduction methods such as PCA [6], LDA [7], and LFDA [8] are also proposed to extract more robust feature from RSS.

Recently, some researchers propose to use channel state information (CSI) as feature [9]–[14]. According to IEEE 802.11n, when APs and client devices work in high throughput (HT) mode, CSI will be included in the CSI field of management frames, which means obtaining CSI is also an easy task. As a complex number indicating the channel condition on one specific subcarrier for an antenna, CSI contains richer information than RSS and provides the possibility for improving localization accuracy. FILA [9] uses the CSI of multiple subcarriers for localization, and achieves a 40% improvement in accuracy compared to RSS based Horus system [10]. As CSIs are complex numbers, various methods are proposed to extract features from it. Xiao *et al.* use only the amplitude of CSI [11] while Wang *et al.* utilize only the phase of CSI [12]. Sen *et al.* utilize the CSI of a single antenna as fingerprints [13] but Chapre *et al.* adopt the CSI from multiple antennas and multiple subcarriers to construct a CSI matrix [14].

From the perspective of the model design, most works formulate fingerprint based localization as a classification problem [15]–[18]. The position of the target device is usually decided as the RP with the most similar feature or the combination of a group of RPs with similar feature. Xie *et al.* adopt KNN [15], in which the Euclidean distance between the feature of the target device and the RPs are calculated, and the resultant position is calculated as the weighted average of RPs with weights inversely proportional to distance. Probability based model treats feature as a random variable and fits the feature at every RP to a distribution. Given the feature of the target device, the probability that the target device resides on a RP can be calculated and localization results is given following the maximum likelihood principle [10]. To estimate the probability distribution of the feature accurately, methods including kernel density estimation [16] and Gaussian process regression [17] are adopted. Decision tree model is also used for localization and achieves higher accuracy than pattern matching [18].

All the methods mentioned above need professional experiences to tune and the selection of the feature is subjective. Neural networks (NN) imitates the signal transition process of neurons and can approximate arbitrary math function. NN can also extract features from input implicitly thus manual feature selection can be avoided. Recently, there is a trend of using NN for fingerprint based localization. Fang and Lin propose DANN [19], which uses a NN with a single hidden layer to extract feature from RSS and improves the probability of the localization error below 2.5m by 17% over RADAR. A three-layer NN is adopted to process the phase of CSI and the weights of the NN is utilized as feature for localization in [12]. DeepFi is proposed in [20] with a four-layer NN and greedy learning algorithm is used for training the model. According to the authors, DeepFi improves accuracy by 20% over FIFS which adopts a probability based model. Note that all existing NN based methods use fully connected (FC) NN and the complexity is positively correlated with the depth of the NN. So the performance of the model is restricted.

In this paper, we propose ConFi, a convolutional NN (CNN) based indoor Wi-Fi localization method that uses CSI as feature. By introducing CNN, the depth of the NN can be increased while keeping the complexity in a proper level [23]. We organize the CSI into what we call CSI feature image. To be more specific, CSIs for different subcarriers at different time are arranged into a matrix, which is similar to one of the RGB channels of an image while CSI matrixes on different antennas are treated as different channels. The CNN consists of three convolutional layers and two FC layers including a softmax output layer. The network is trained using the CSI feature images collected at a number of RPs. The localization results is the weighted centroid of RPs with high output value. Moreover, extensive simulation is conducted to select appropriate parameters for ConFi and compare against existing methods in a real indoor scenario.

The contribution of the paper can be summarized as follows. Firstly, we propose a novel representation of CSI as CSI feature image. With CSI feature image, manual subjective feature selection and preprocessing are avoided while the information contained in CSI is utilized comprehensively. Secondly, to the best of our knowledge, ConFi is the first method that utilizes CNN for Wi-Fi localization, which captures the correlation among time, frequency and antenna domain in CSI. Lastly, ConFi extends the depth of the NN and improves the localization accuracy. Extensive experiments are conducted to compare the performance of ConFi with existing methods and explore the influence of various model parameters.

The remainder of this paper is organized as follows. CSI measurement in Wi-Fi and the construction of CSI feature image are introduced in Section II. Section III presents the structure and the training method of the CNN. Section IV provides the experiment results while Section V concludes the paper.

II. CSI FEATURE IMAGE

In this section, we introduce CSI related background in Wi-Fi and illustrate how to organize CSIs for multiple subcarriers, time slots and antennas into CSI feature image.

A. CSI MEASUREMENT IN WI-FI

Estimating CSI is a fundamental functionality in wireless communication system, which provides support for functionalities such as power control and handover. Wi-Fi uses training sequence for CSI estimation. According to IEEE 802.11n [21], sounding PPDUs (physical layer convergence procedure protocol data unit) are sent from the beamformee to the beamformer to estimate the CSI during transmit beamforming procedure. In the time domain, the received signal can be written as

$$r(t) = s(t) * h(t) + n(t), \tag{1}$$

where $s(t)$ is the transmitted signal made up of known training sequence and $n(t)$ is the random noise. $h(t)$ is the channel impulse response modeling the comprehensive effects of large scale fading, multi-path fading and shadowing. The channel response in frequency domain can be calculated as

$$\hat{H} = R/S, \tag{2}$$

where S is the Fast Fourier Transform (FFT) of the training sequence and R is the FFT of the received sequence. $\hat{H}$ is the CSI between the transmitter and the receiver. The CSI is used for beamforming procedure and can be obtained from the CSI field of MAC management frame. For 20 MHz bandwidth, there are 56 subcarriers in total and three grouping configurations of subcarriers, which are listed in Table 1 [21]. N_S is the number of the subcarriers used for training sequence transmission. The exact no. of the subcarriers used for CSI extraction are shown in the last column. The Wi-Fi vendor should choose at least one configuration to support beamforming.

TABLE 1. CSI Grouping Configuration in 802.11n.

BW	Ns	Carrier no. for CSI extraction
20MHz	56	All
	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
	16	-28,-24,-20,-16,-12,-8,-4,-1,1,5,9,13,17,21,25,28

Fig. 1 shows the amplitude of the CSI on 30 subcarriers from 5000 measurements at a single location. The CSIs for the three antennas are plotted with different colors. The following observations can be made. Firstly, The CSIs on different antennas show different patterns, which means using multiple antennas may better capture location dependent CSI pattern and yield better performance. Secondly, CSIs on adjacent subcarriers are similar (we find CSIs measured at adjacent time slots are also similar), resembling an image in which adjacent pixels usually takes similar values. This similarity is what motivates us to propose CSI feature image. Thirdly, the CSIs on the same antenna show different patterns over the measurement period. For example, RX Antenna A experiences roughly four patterns and the maximum difference in amplitude reaches 35 dB. This suggests there is the need to take the time domain changes in CSI into consideration, which existing works fail to do.

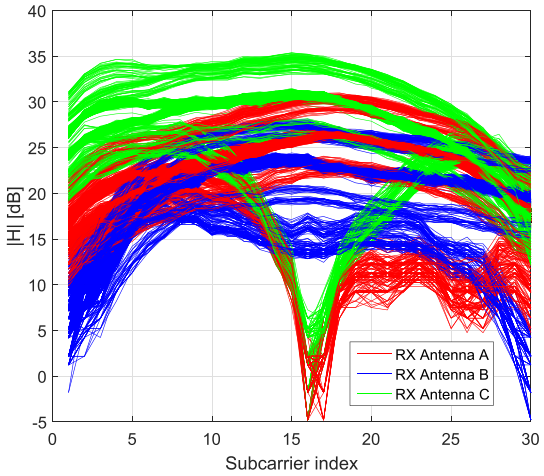


FIGURE 1. CSI amplitude of multiple antennas for a single location.

B. CSI FEATURE IMAGE

As analyzed in [20], the phase of CSI is prone to noise and random fading thus complicated preprocessing is needed before using it as feature. To avoid preprocessing, we only use the amplitude of CSI. For one antenna, we group T CSI measurements for N subcarriers at the same RP to construct a $N * T$ matrix which we call CSI feature sub-image as follows.

$$|\mathbf{H}|_i = \begin{pmatrix} |\mathbf{H_{11}}| & \cdots & |\mathbf{H}_{1T}| \\ \vdots & \ddots & \vdots \\ |\mathbf{H}_{N1}| & \cdots & |\mathbf{H}_{NT}| \end{pmatrix}_i \tag{3}$$

where N is the number of subcarriers, T is the number of CSI measurements in one sub-image and i is the index of antenna. Nowadays, advanced Wi-Fi APs are usually equipped with multiple antennas and as shown in the previous subsection, different antennas usually have quite different CSI patterns. Therefore, we can organize the CSI from different antennas into separate CSI feature sub-images. This means the CSI feature sub-image of an antenna acts like one of the RGB channels of an actual image. The set of CSI feature sub-images on all antennas is called CSI feature image. However, as opposed to images, which usually have three channels, the number of channels in ConFi is decided by the number of antennas. CSI feature images collected at the same RP are treated as samples from the same category when training the CNN.

Some examples of CSI feature images are illustrated in Fig. 2. They are collected at 4 different RPs. Three antennas are used and we map the CSI feature sub-images from the antennas into the RGB channels of the image. The pixel in a column corresponds to CSI amplitude of a subcarrier from three antennas. The elements in the row are composed by the time samples. We can make several assertions from the images. Firstly, the images from different RPs have different

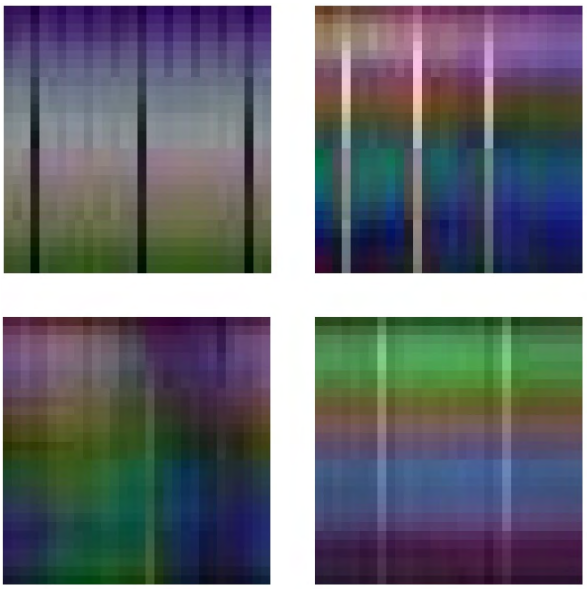


FIGURE 2. Examples of CSI feature images from 4 RPs.

patterns, which suggests CSI feature image is good feature for localization. Secondly, One can tell there are several vertical lines in the images. It implies that some features present on every subcarriers but appear sporadically. Theses features are quite difficult to be captured by a single-shot. Thirdly, the colour of the image is quite different, which indicates different features need to be extracted at different RPs.

As CNN has many parameters to tune, a large number of training samples are needed to prevent overfitting. However, getting training sample can be expensive and translation and horizontal reflection are usually applied to the original images to expand the training set in computer vision. Since the pixels in CSI feature image are actual CSIs, applying translation and horizontal reflection to them may corrupt the information contained in CSI. Instead, we use a sliding window strategy to expand the training set. When the number of CSI measurements in a CSI feature image is T, we generate a CSI feature image every $T/2$, which means adjacent CSI feature images are allowed to overlap in the time domain. The detailed performance comparison with other expanding techniques are shown in section 4.3.

III. CNN BASED LOCALIZATION

CNN is proved as an effective technique in image classification. By using convolutional kernels, CNN is robust to noise and can construct increasingly high level representation of the input images at latter layers. Please refer to [22] for detailed introduction of the CNN in image classification. Therefore, we apply the CNN as our model and formulate the localization as a classification problem. The proposed CNN based localization method consists of two stages, i.e., a training stage and a localization stage. In the training stage, multiple CSI feature images are collected at every RP and the CNN is trained using the CSI feature images as in

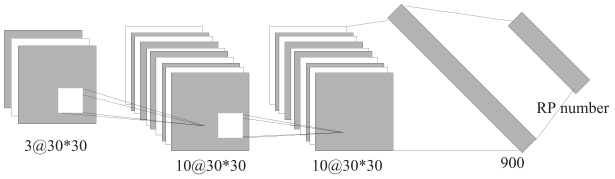


FIGURE 3. Structure of the CNN.

a classical multiclass image classification problem. In the localization stage, the CSI feature image of the target device is fed to the trained neural network and the position is estimated as the weighted centroid of RPs with a high value at the output layer. In this section, we first introduce the structure of the CNN and then present the loss function and training method.

A. STRUCTURE OF THE CNN

The structure of the CNN used in ConFi is inspired by LeNet [23] and Alexnet [24] which produce remarkable performance in image recognition. As shown in Fig. 3, the network has five layers, which consists of three convolutional layers and two FC layers. As CSI feature image is different from actual image, our CNN is also different from conventional CNN in several aspects. The first difference is that we pad the feature image and set the stride step to one so that the size of the input image will not be reduced by the convolutional layers. This is because the size of the feature image is already small and we want the FC layers to have enough number of input features. The second difference is that we do not use the pooling layers, which conducts sampling essentially and reduces the size of the image. We believe there are fine descriptions of location features in CSI feature image, while the pooling process will confuse these information.

TABLE 2. Example parameter settings of CNN.

Layer Type	Input Size	Parameters	Activation Function
Convolutional layer	30*30*3 pad=2 stride=1	5*5 filter kernel 10 feature images	ReLUs
Convolutional layer	30*30*10 pad=2 stride=1	5*5 filter kernel 10 feature images	ReLUs
Convolutional layer	30*30*10 pad=2 stride=1	5*5 filter kernel 10 feature images	ReLUs
FC layer	9000	900 neurons Dropout 50%	ReLUs
FC layer	900	RP number neurons	Softmax

As an example, we give the specific parameters of the CNN for 30 by 30 CSI feature image from 3 transmitting antennas in Table 2. The inputs of the CNN are three 30 by 30 pictures. For the convolutional layers, we set the number of the convolutional kernel to be 10. So the outputs of the convolutional layer are 10 feature images. For the reasons described above, we choose 5 by 5 filter size as convolutional kernels and use padding to keep the image size unchanged.

The stride of the convolutional filter is set to 1 so as to extract the time-frequency information precisely. For the senond last FC layer, we use 50% dropout [25] to avoid overfitting.

The activation function introduces nonlinearity into NN and is an important factor for performance. We choose Rectified Linear Units (ReLUs) as the activation function. It is more plausible biologically than the sigmoid function, and the resultant NN enjoys good sparsity which translates into high computation speed. ReLU can be expressed as follow:

$$f(x) = max(0, x) \tag{4}$$

The number of neurons at the output layer is equal to the number of RPs, therefore each output neuron corresponds to a RP. As the target device may appear at any of the RPs, we use softmax as the activation function of output layer, which means the outputs of all neurons in the output layer sum to one. Therefore the output of a neuron can be interpreted as the probability that the target device is at the corresponding RP. The definition of the softmax function is as follows:

$$y^{(j)} = \frac{e^{\mathbf{w}_j^{\mathbf{T}} x^{(i)}}}{\sum_{j=1}^{K} e^{\mathbf{w}_j^{\mathbf{T}} x^{(i)}}} \tag{5}$$

where $y^{(j)}$ is output of jth neuron in the output layer. j is the index of output neurons while K is total number of output neurons which is equal to the number of RPs. $x^{(i)}$ is the output of second last layer and $\mathbf{w}_j$ is the weight vector connecting the neurons in the second last layer to the output layer. $\mathbf{T}$ means transformation of a vector. Note that softmax function maps the output in the range of [0, 1].

To train the network, we use cross-entropy [26] plus a regularization term as the loss function.

$$J(\mathbf{w}) = -\frac{1}{M}\left[\sum_{i=1}^{M}\sum_{j=1}^{K} 1\{z^{(i)} = j\} log \frac{e^{\mathbf{w}_j^{\mathbf{T}} x^{(i)}}}{\sum_{l=1}^{K} e^{\mathbf{w}_l^{\mathbf{T}} x^{(i)}}}\right] + \frac{\lambda}{2}\sum_{i=1}^{P}\sum_{j=1}^{K} w_{ij}^2 \tag{6}$$

where $1\{\}$ is the indicator function, $\lambda > 0$ is the weight of the regularizer. P is the dimension of the $\mathbf{w}_j$ which corresponds to the number of neurons in the second last layer. M is the size of the training set. $z^{(i)}$ is the index of the RP at which the CSI feature image is collected. The cross-entropy in loss function enforces that if the input CSI feature image is collected at the jth RP, the output of the jth neuron should be close to one. The regularization term can prevent the network weights from taking extremely large value thus helps to avoid overfitting. We train the network to minimize Eq. 7 and its derivative is:

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}_j} = -\frac{1}{M}\sum_{i=1}^{M}\left[x^{(i)}\left(1\{z^{(i)} = j\} - \frac{e^{\mathbf{w}_j^{\mathbf{T}} x^{(i)}}}{\sum_{j=1}^{K} e^{\mathbf{w}_j^{\mathbf{T}} x^{(i)}}}\right)\right] + \lambda \mathbf{w}_j \tag{7}$$

We utilize stochastic gradient descent and backpropagation algorithm to train the network until the decease of the loss function between adjacent iterations falls below a threshold.

B. LOCALIZATION

In the localization stage, the CSI feature image of the target device is fed into the model. The model outputs $y^{(j)}$, which can be interpreted as the probability that the target device is located at the jth RP. For the target device may appear in any position of interested area, we use the probability weighted centroid method to estimate the final location, which is calculated as follows:

$$\hat{L} = \frac{\sum_{j\in\Omega} y^{(j)} R_j}{\sum_{j\in\Omega} y^{(j)}} \tag{8}$$

where R_j is the coordinate of the jth RP. Ω is the set of considered RPs. In our experiments, We typically use 3 RPs with the largest output value to calculate the centroid.

IV. EXPERIMENTS VALIDATION

A. EXPERIMENTS SETUP

We use a ThinkPad E430 laptop equipped with Intel 5300 wireless network card as the target device. TP-link TL-WR885N wireless router which has 3 antennas is used as the AP. A desktop PC with NIVIDA GTX1080 Graphic Card acts as the model training server (based on the Caffe framework [27] and CUDA Tool kit 7.5).

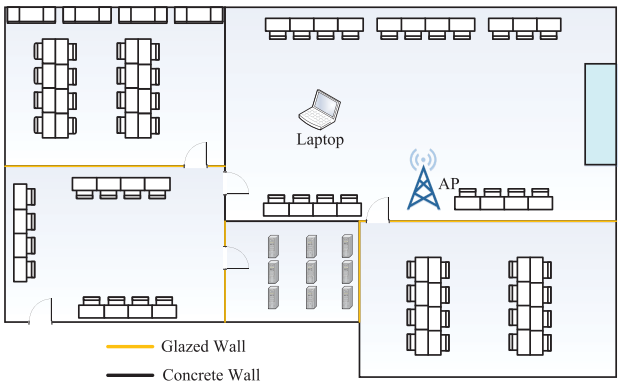


FIGURE 4. The schematic of indoor scenario.

We verify our model in a typical indoor scenario. As shown in Fig. 4, the whole experiment area is about 16.3 m by 17.3 m with five rooms. The walls include both concrete wall and glazed wall. There are also reflectors such as furniture and rack servers. The AP and the target device are set on the desk and a cart with the height of 150 cm, respectively. In this paper, we only focus the localization in 2D space, which means the height of the target device is kept constant.

We choose 64 RPs with a spacing from 1.5m to 2m in between. Therefore, the output layer of the CNN has 64 neurons. In the training stage, the laptop is positioned at the RPs and ICMP packets are collected from AP. The interval of the packets is 0.01s and we record for 2 minutes at every

RP. We conduct 10 independent measurements on different days to take into account the time domain variation of CSI. At every RP, we get 120000 time domain samples. On the training server, these time domain samples from the same RP are grouped into CSI feature images. We partition the entire dataset into training sets, validation sets and test sets using a ratio of 7:2:1.

B. ANALYSIS OF PARAMETER SETTING

In this subsection, we analyze the effect of various parameters on performance by experiment and identify a good set of parameter settings for comparison with existing methods. In our experiments, we find that high classification accuracy usually translates into low localization error. Therefore, we use classification accuracy as the metric for parameter selection. As described above, the validation set is used to determine when to stop training. After training, the test set is used to test the performance of the trained model. Since the test set is not used in the training process, classification accuracy on it should be a good approximation of the generalization error of the model. The learning rate is set as 0.001. The training sets batch size is 256.

1) THE SIZE OF FEATURE MAP

In the experiments, we use 30 subcarriers, so the number of rows of the feature image is 30. We compare the performance of different number of columns using the same amount of CSI samples. Note that a larger number of columns means each CSI feature image spans a longer time but the total number of CSI feature images will be less. The configuration of four CSI feature image sizes and their performance are summarized in Table 3.

TABLE 3. Comparison of feature map size.

Size	Accuracy	
	Train	Test
30*15	91.46%	87.17%
30*30	95.83%	89.58%
30*60	92.86%	88.02%
30*90	90.97%	88.63%

It is obvious that 30*30 CSI feature image gets the highest accuracy. The time span of 30*15 CSI feature image size is too short and fails to capture the time domain correlation between the CSI samples. 30*60 and 30*90 CSI feature image sizes are too long, resulting in an insufficient number of training samples.

2) DATA AUGMENTATION

We also compare different methods to expand the training set. The baseline is the case that no training set expansion method is used, the three considered methods are mirror, random and sliding window.

Mirror is widely used in image classification and it reflects an image in a left-right manner. That is, the right side half of the image is just a copy of the left hand half but the order is reserved. Randomly choosing samples to construct the CSI feature image from the set of samples means the samples in the same CSI feature image may not be adjacent in time. Sliding window has been explained in Section II. The results are shown in Table 4.

TABLE 4. Comparison of data augmentation methods.

Methods	Accuracy	
	Train	Test
Baseline	95.83%	89.58%
Mirror	97.27%	85.03%
Random	92.90%	88.00%
Sliding Windows	96.17%	91.78%

In the table, we can see that mirror and random performs worse than the baseline, and sliding window provides the best performance. Random fails to capture the correlation of CSI over time, which is a common problem of existing works, as they do not consider time domain information in CSI by using only one snap shot of CSI. Note that mirror gets the highest training accuracy but the worst test accuracy, which is a sign of overfitting.

3) SIZE OF CONVOLUTIONAL KERNEL

Convolutional kernel is also called receptive field, which decides how many pixels will contribute to a feature in the succeeding layer and can also be regarded as the window for information acquisition. We compare different sizes of convolutional kernel without data augmentation. In Table 5, we can find 5*5 kernel size is the best choice. 3*3 kernel is too small to capture time domain feature, while 7*7 kernel is too large and introduces noise.

4) THE NUMBER OF CONVOLUTIONAL KERNELS

In CNN, different kernels extract different features from the input and construct individual feature maps. We compare the performance with different number of kernels in Table 6. We can find 10 kernels work best in our model. When reducing to 5 kernels, the accuracy reduces by 4%, suggesting the number of feature maps is insufficient. While doubling the kernels to 20, accuracy only increases by only 0.1%. To balance between performance and computation cost, we use 10 kernels for convolutional layers.

C. COMPARISON WITH EXISTING ALGORITHMS

Different from the parameter selection part, we compare the performance of the algorithms using 32 randomly selected test points (TP) that are not necessarily coincident with the RPs used for training. At each TP, we collect samples for 1 minute in 5 independent trials, which result in 960000 CSI samples for all the TPs. Note that although the samples at one TP are sufficient to construct many CSI feature images, we only use one CSI feature image for localization as

TABLE 5. Comparison of convolutional kernel size.

3*3		5*5		7*7	
Train	Test	Train	Test	Train	Test
95.31%	87.11%	95.83%	89.58%	95.70%	86.98%

TABLE 6. Comparison of convolutional kernel number.

5		10		20	
Train	Test	Train	Test	Train	Test
94.67%	85.71%	95.83%	89.58%	95.18%	89.67%

practical localization usually has a delay requirement. In fact, we turn data collected at one TP into multiple test cases by partitioning its CSI samples into multiple feature images. Localization accuracy is measured by the distance between the output of an algorithm to the ground truth. The parameters of the compared algorithms are all tuned to give the best performance.

1) COMPARISON WITH RSS BASED METHODS

RADAR [5] and Horus [10] are RSS fingerprint localization methods based on KNN and probability theory, respectively. In Table 7, we can observe that ConFi outperforms them by a large margin, i.e., a 42.8% improvement over Horus and a 66.9% improvement over RADAR in mean localization error.

TABLE 7. The comparison of statistic error with RSS based methods.

Algorithm	ConFi	Horus	RADAR
Mean(m)	1.3654	1.9503	2.2775
Std. dev.(m)	0.9005	1.2808	1.4774

The cumulative distribution functions (CDF) of localization error of the three algorithms are plotted in Fig. 5. For ConFi, 70% of the test cases have a localization error under 1.5 meters while less than 50% of the test cases have a localization error below 1.5 meters for RADAR and Horus. This proves again the fact made clear by existing CSI based methods-CSI contains richer information than RSS and yields superior performance.

2) COMPARISON WITH CSI BASED METHODS

We compare ConFi with two CSI based methods, FILA [9] and CSI-MIMO [14]. CSI-MIMO uses complex CSI from multiple antennas instead of only the amplitude and adopts a probability theory based formulation while FILA utilizes the summation of the amplitude of CSI from multiple APs. In Table 8, FILA has the worst mean accuracy performance as it is designed to work for the scenario where multiple APs have line of sight measurement of the target device while we only use a single AP. Note that ConFi can also work with multiple APs by changing the number of input CSI feature images. In conclusion, ConFI has a 17.8% and 31.3%

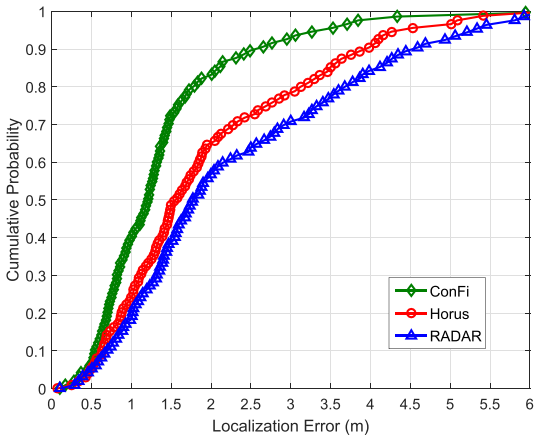


FIGURE 5. The comparison of CDF with RSS based methods.

TABLE 8. The comparison of localization error with CSI based methods.

Algorithm	ConFi	CSI-MIMO	FILA
Mean(m)	1.3654	1.608	1.793
Std. dev.(m)	0.9005	1.077	1.154

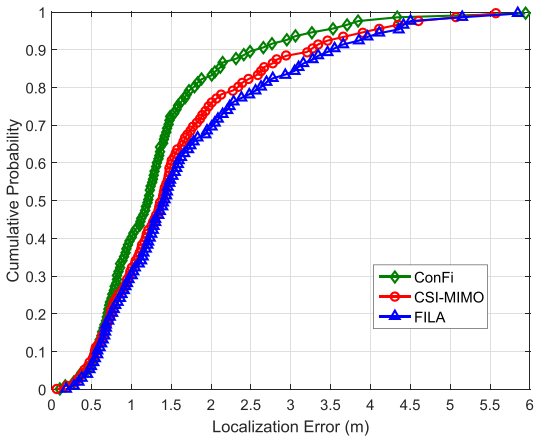


FIGURE 6. The comparison of CDF with CSI based methods.

improvement in mean accuracy over CSI-MIMO and FILA, respectively.

Fig. 6 shows the CDFs of localization error for ConFi, CSI-MIMO and FILA. It can be observed that ConFi increases the percentage of test cases having an error under two meters by 8% and 10% over CSI-MIMO and FILA respectively. This demonstrates CNN is the more effective model for CSI based localization. Moreover, the superior performance of CSI-MIMO over FILA suggests the correlation between the CSI of multiple antennas helps in localization.

3) COMPARISON WITH OTHER NN BASED METHODS

We compare the mean and standard deviation error of ConFi with DeepFi [20] and DANN [19] in Table 9. ConFi improves the mean error by 9.2% and 21.64% over the two algorithms, respectively. Note that DANN performs even worse

TABLE 9. The comparison of localization error with neural network based methods.

Algorithm	ConFi	DeepFi	DANN
Mean(m)	1.3654	1.491	1.6609
Std. dev.(m)	0.9005	0.9798	1.1182

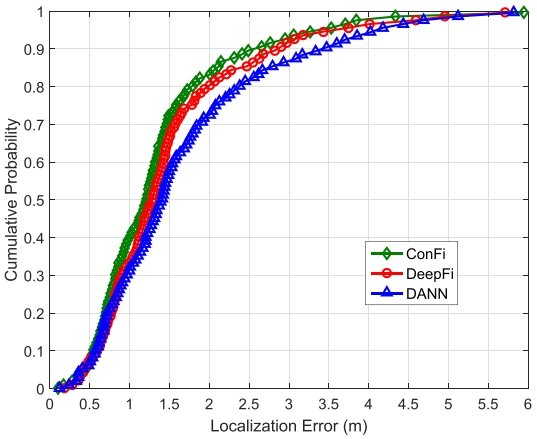


FIGURE 7. The comparison of CDF with NN based methods.

than CSI-MIMO, which can be explained by the fact DANN uses RSS. The performance advantage of ConFi over DeepFi and DANN indicates CNN is more suitable for localization than fully connected NN.

In the error CDF plot for ConFi, DANN and DeepFi in Fig. 7, we can observe that ConFi improves the percentage of test cases having an error below 1.5 meters by 5.6% and 16% over DANN and DeepFi, respectively. Therefore we can conclude CNN can extract feature from multi antennas more effectively than fully connected NN.

V. CONCLUSION

In this paper, we proposed ConFi the first convolutional neural network based indoor Wi-Fi localization system. The CSI from multiple antennas were organized into multiple matrixes indicating CSI over time and frequency domain and used as the input of the convolutional neural network. A five-layer neural network with three convolutional layers and two fully connected layers was utilized to process the CSI feature images. With extensive experiment, we select appropriate parameters for the convolutional neural network and verify that ConFi outperforms most existing methods. Our result suggests that CNN is a powerful tool for capturing the information encoded in CSI for localization, its superior performance demonstrates the power of CNN in pattern recognition, which may also work for problems such as automatic modulation classification.

REFERENCES

[1] A. Cidronali, S. Maddio, G. Giorgetti, and G. Manes, "Analysis and performance of a smart antenna for 2.45-GHz single-anchor indoor positioning," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 1, pp. 21–31, Jan. 2010.

[2] Y. Wang, S. Ma, and C. L. P. Chen, "TOA-based passive localization in quasi-synchronous networks," *IEEE Commun. Lett.*, vol. 18, no. 4, pp. 592–595, Apr. 2014.

[3] J. K.-Y. Ng, K.-Y. Lam, Q. J. Cheng, and K. C. Y. Shum, "An effective signal strength-based wireless location estimation system for tracking indoor mobile users," *J. Comput. Syst. Sci.*, vol. 79, no. 7, pp. 1005–1016, Nov. 2013.

[4] A. Jaffe and M. Wax, "Single-site localization via maximum discrimination multipath fingerprinting," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1718–1728, Apr. 2014.

[5] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Tel Aviv, Israel, Mar. 2000, pp. 775–784.

[6] S.-H. Fang and C.-H. Wang, "A novel fused positioning feature for handling heterogeneous hardware problem," *IEEE Trans. Commun.*, vol. 63, no. 7, pp. 2713–2723, Jul. 2015.

[7] S. H. Fang and T. N. Lin, "Projection-based location system via multiple discriminant analysis in wireless local area networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 5009–5019, Nov. 2009.

[8] Z.-A. Deng, Y. Xu, and L. Chen, "Localized local fisher discriminant analysis for indoor positioning in wireless local area network," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, Apr. 2013, pp. 4795–4799.

[9] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "FILA: Fine-grained indoor localization," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Orlando, FL, USA, Mar. 2012, pp. 2210–2218.

[10] M. Youssef and A. Agrawala, "The Horus location determination system," *Wireless Netw.*, vol. 14, no. 3, pp. 357–374, Jun. 2008.

[11] J. Xiao, K. Wu, Y. Yi, and L. M. Ni, "FIFS: Fine-grained indoor fingerprinting system," in *Proc. 21st Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul./Aug. 2012, pp. 1–7.

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

[13] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "You are facing the Mona Lisa: Spot localization using PHY layer information," in *Proc. 10th Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, New York, NY, USA, 2012, pp. 183–196.

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

[15] Y. Xie, Y. Wang, A. Nallanathan, and L. Wang, "An improved K-nearest-neighbor indoor localization method based on spearman distance," *IEEE Signal Process. Lett.*, vol. 23, no. 3, pp. 351–355, Mar. 2016.

[16] Y. Cao, H. He, and H. Man, "SOMKE: Kernel density estimation over data streams by sequences of self-organizing maps," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 8, pp. 1254–1268, Aug. 2012.

[17] M. M. Atia, A. Noureldin, and M. J. Korenberg, "Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 9, pp. 1774–1787, Sep. 2013.

[18] D. Liang, Z. Zhang, and M. Peng, "Access point reselection and adaptive cluster splitting-based indoor localization in wireless local area networks," *IEEE Internet Things J.*, vol. 2, no. 6, pp. 573–585, Dec. 2015.

[19] S. H. Fang and T. N. Lin, "Indoor location system based on discriminant-adaptive neural network in IEEE 802.11 environments," *IEEE Trans. Neural Netw.*, vol. 19, no. 11, pp. 1973–1978, Nov. 2008.

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

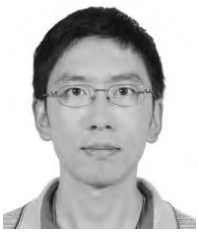
[21] *IEEE Standard for Information Technology—Local and Metropolitan Area Networks—Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput*, IEEE Standard 802.11n-2009, Oct. 2009.

[22] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. 7th Int. Conf. Document Anal. Recognit.*, 2003, pp. 958–963.

[23] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.

[25] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. (Jul. 2012). "Improving neural networks by preventing co-adaptation of feature detectors." [Online]. Available: https://arxiv.org/abs/1207.0580
[26] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 19–67, 2005.
[27] Y. Jia *et al.* (Jun. 2014). "Caffe: Convolutional architecture for fast feature embedding." [Online]. Available: https://arxiv.org/abs/1408.5093



HAO CHEN received the M.S. degree in telecommunication and information systems from the Beijing University of Posts and Telecommunications in 2007, where he is currently pursuing the Ph.D. degree in intelligent communication systems. He was Senior Research Engineer with the Key Laboratory of Universal Wireless Communications, Ministry of Education, from 2007 to 2013. His research interests are localization, pattern recognition, convolutional neural network, and machine learning and applications.

YIFAN ZHANG received the Ph.D. degree in 2007 from the Beijing University of Posts and Telecommunications (BUPT). He is currently an Associate Professor with the School of Information and Communication Engineering, BUPT. His current research interests include compressed sensing, optimization algorithms in wireless networks, and machine learning and applications.

WEI LI received the Ph.D. degree in electrical and computer engineering from the University of Victoria, Canada, in 2004. He is currently an Assistant Professor with the Northern Illinois University, USA. His research interests are wireless networks and applications, Internet of Thing, machine learning and artificial intelligence algorithms, and big data analytics.

XIAOFENG TAO received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 1993, and the M.S.E.E. and Ph.D. degrees in telecommunication engineering from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1999 and 2002, respectively. He is currently a Professor with BUPT and a fellow of the Institution of Engineering and Technology. He is currently focusing on the 5G networking technology and mobile network technology.

PING ZHANG is currently the Chair Professor with the Beijing University of Posts and Telecommunications and the Director of the State Key Laboratory of Networking and Switching Technology, China. His research interests include cognitive wireless networks, fifth generation mobile networks, universal wireless signal detection instrument, and mobile Internet. He was a recipient of the First and Second Prizes of the National Technology Invention and Technological Progress Awards and the First Prize of the Outstanding Achievement Award of Scientific Research in College. He is currently the Executive Associate Editor-in-Chief on *Information Sciences of the Chinese Science Bulletin*, a Guest Editor of the IEEE *Wireless Communications Magazine*, and an Editor of the *China Communications*.