IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 23, 2019, accepted October 7, 2019, date of publication October 10, 2019, date of current version October 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2946870

Learning Spatiotemporal Features of CSI for Indoor Localization With Dual-Stream 3D Convolutional Neural Networks

YUAN JING, JINSHAN HAO, AND PENG LI
School of Information, Liaoning University, Shenyang 110036, China

Corresponding authors: Yuan Jing (yjing@lnu.edu.cn) and Peng Li (35780780@qq.com)

This work was supported in part by the Scientific Research Project of Liaoning Provincial Department of Education of China under Grant LYB201616, in part by the Public Sentiment and Network Security Big Data System Engineering Laboratory of Liaoning Province of China under Grant 2016(294), and in part by the China Scholarship Council.

ABSTRACT Recently, the research of WiFi-based indoor localization combing with deep-learning techniques has earned wide attention due to its potential applications in smart cities. In this paper, a novel fingerprinting system is proposed to achieve indoor localization via learning spatiotemporal features from channel state information (CSI) of multiple-input multiple-output wireless channels (CSI-MIMO) by a dual-stream three-dimensional (3D) convolutional neural network (DS-3DCNN). In the proposed system, the gathered raw CSI-MIMO data are firstly preprocessed through amplitude outliers elimination and phase sanitization for constructing a pair of 3D CSI-MIMO matrices including a 3D amplitude matrix and a 3D phase matrix. Next, the 3D matrices will be input to the DS-3DCNN deep neural network which consists of two parallel subnetworks with specific architecture of several convolution, batch normalization, max-pooling, and fully connected layers. Through this DS-3DCNN network, learning spatiotemporal features of CSI-MIMO is carried out simultaneously from 3D amplitude and phase matrices. And then, probabilistic classification results of two subnetworks are fused in the final output layer of the proposed DS-3DCNN based on Bayes' theorem. Moreover, in the offline training stage, a dual-stream joint optimization method is presented for efficiently optimizing network parameters. After offline training of the DS-3DCNN, in the online locating stage, current CSI-MIMO data are firstly collected from the mobile device to be located. Then probabilistic classification results are obtained from the output layer of the DS-3DCNN, and further used to approximate the posterior distribution of the mobile device's location with a Gaussian mixture model. Finally, a novel location estimation algorithm is deduced based on the minimum mean square error (MMSE) criterion. Since unique features of wireless MIMO channels are jointly learnt in spatial, temporal, and frequency domains, the proposed DS-3DCNN based fingerprinting system is reasonable to provide accurate localization results in indoor environments, which is verified in corresponding experiments.

INDEX TERMS Indoor localization, deep learning, convolutional neural network (CNN), channel state information (CSI).

I. INTRODUCTION

In recent years, with the rapid development of artificial intelligence, deep learning (DL) techniques [1], [2] have been widely and successfully applied to physical layer wireless communications to improve system performance, such as nonorthogonal multiple access scheme [3], automatic modulation recognition [4], wideband RF power amplifiers [5], channel estimation [6], channel prediction [7], and wireless localization [8]. For the indoor wireless localization, it is one of the most attracting research hot-spots due to increasing demands of location-based applications such as human activity recognition, mobile robot navigation, and health surveillance [9]–[11]. Differing from Bluetooth beacons, infrared sensors, and other equipments, WiFi devices have wide-spread coverage and adopt IEEE 802.11 series standards, which activate considerable research efforts on the WiFi-based indoor localization [12], [13]. Moreover,

The associate editor coordinating the review of this manuscript and approving it for publication was Guan Gui.

due to the convenience of gathering without extra infrastructures, received signal strength (RSS) and channel state information (CSI) have been commonly used in most existing WiFi-based indoor localization systems [14]–[16] compared with other measurements such as time difference of arrival (TDoA), time difference of flight (TDoF), and angle of arrival (AoA).

In most RSS/CSI-based indoor localization systems, fingerprinting becomes a popular technique due to some advances such as simple principle and easy implementation [14]–[16]. In fingerprinting systems, there are two stages in the localization progress: the offline training stage and the online locating stage [13]. During the former stage, a set of RSS/CSI data are firstly collected through arranging a mobile device on different reference points (RPs) with known locations. Then, a fingerprint database is constructed to store identifiable features of RSS/CSI data corresponding to different RPs; When going to the online locating stage, the online acquired RSS/CSI data corresponding to the mobile device with unknown location will be sent to fingerprinting systems for extracting or learning features and making a matching with stored versions in the fingerprint database. And then, we can determine which RP the mobile device is closer to. Therefore, localization performances of WiFi-based fingerprinting systems mainly depend on the effective learning of identifiable features from RSS/CSI data.

However, due to complex indoor wireless environments such as multi-path propagation, obstacle occlusion, and possible shadow fading, it is still a challenge to accurately learn detailed features and construct an identifiable fingerprint database based on WiFi signals.

Many traditional research efforts have been devoted to directly store RSS data or those transformed versions in fingerprint database and achieve indoor localization. In [17], the RADAR localization system was proposed to pioneer the combination of empirical RSS data with signal propagation modelling. Then, the location of the mobile device can be estimated by minimize the Euclidean distance between the online-gathered RSS and the stored version. To further improve the localization performance, Horus system was presented in [18] through investigating a probabilistic model to describe RSS data and determining the location of mobile device based on the K-nearest neighbor (KNN) algorithm. Subsequently, many other RSS-based fingerprinting methods were proposed to improve indoor localization accuracy by using various machine learning methods [19].

Unfortunately, in practical applications, RSS values sometimes change significantly over time in complex indoor wireless environments due to some characters of wireless channels such as the multi-path effect and shadow fading, which may result in the performance degradation of RSS-based fingerprinting systems [20]. Therefore, RSS may be not the best choice for achieving accurate indoor localization.

In WiFi networks adopting IEEE 802.11 series standards [21], Orthogonal Frequency Division Multiplexing (OFDM) technology is used to obtain the frequency diversity for the wireless communication link by utilizing multiple subcarriers simultaneously to complete the data transmission. Accordingly, CSI data on each subcarrier could be easily collected through some network interface cards (NICs) such as Intel 5300 NIC [22] and Atheros NICs [23] on either APs or mobile devices. Furthermore, compared with RSS, CSI could provide both amplitude and phase information on each subcarrier between any transmit-receive antenna pair. Most importantly, CSI offers more stable characters under complex indoor wireless environments. Consequently, it is possible to obtain more identifiable and relatively stable channel features from CSI which can be used to determine the geographical location of the transmitter or receiver [15]. Especially, when utilizing multiple antennas on both transmitters and receivers, CSI of the multi-input and multi-output wireless channels (CSI-MIMO) could also be conveniently collected to construct a fingerprint database and obtain more accurate localization performance [24].

In addition, the occurrence of DL techniques makes the fingerprinting to be more efficient and effective [1], [2]. In DL-based fingerprinting approaches, deep neural networks (DNNs) are generally explored to learn core features of WiFi signals for achieving indoor localization [25]. Accordingly, after offline training stage, network parameters are stored in a database, which could save much more memory space than storing the raw WiFi signals. In the online locating stage, currently gathered raw WiFi signals would be input into DNNs to learn features and determine which RP is closest to the mobile device's location. Since DL-based fingerprinting systems are able to automatically learn the intrinsic features of raw WiFi signals, it is reasonable to obtain more accurate localization results compared with other traditional machine-learning methods, such as KNN and Support Vector Machine (SVM) [25].

With the great success of DL, in recent years, many pioneer and valuable works have been done on CSI-based fingerprinting systems for indoor localization [8], [25]–[29]. In [26], DeepFi system exhibited the strong ability in indoor localization by utilizing a DNN with four hidden layers to learn CSI features. However, too many network parameters needed to be trained and stored, which limited its application. Furthermore, in [27], a deep residual sharing learning method was proposed to learn detailed features from dual-channel CSI tensor data. The corresponding ResLoc system provided more accurate localization performance and required less storage space than the DeepFi system. Differing from DeepFi and ResLoc systems, [28] tried to use a five-layers convolutional neural network (CNN) for learning features of wireless channels from CSI amplitude feature images. Then, indoor localization was formulated as a classification problem and solved by a ConFi fingerprinting system. Based on the convolution calculations in the ConFi system, localization accuracy was effectively increased for the typical indoor scenario. To further improve the localization performance, in [29], CiFi system was designed by constructing AoA fingerprint maps

based on CSI phase information, and then to learn the unique features of AoA fingerprint maps for indoor localization. Experiments show that more accurate locating results can be obtained by using the CiFi system. Additionally, [8] explored a one-dimensional CNN to learn features from both RSS and CSI data for indoor locating, which earned outstanding performance in runtime efficiency.

As introduced above, most of existing CSI-based fingerprinting systems tried to explore CSI features in one or two dimensions for indoor localization. To our knowledge, however, no method till now tries to learn unique 3D spatiotemporal features simultaneously from the amplitude and phase information of CSI-MIMO data for improving the localization performance. Therefore, how to design a 3D DNN [30] to learn spatiotemporal features from CSI-MIMO data would be valuable for improving the localization performance of fingerprinting systems.

In this paper, a novel dual-stream 3D convolutional neural network (DC-3DCNN) based fingerprinting system is proposed to simultaneously learn spatiotemporal features from the amplitude and phase information of CSI-MIMO data for accurate indoor localization. In this fingerprinting system, amplitude outliers elimination and phase sanitization are firstly applied to gather CSI-MIMO data for constructing 3D amplitude and phase matrices. Then, these 3D CSI-MIMO matrices are input to the proposed DS-3DCNN which consists of two parallel subnetworks with several convolution, batch normalization (BN) [32], max-pooling, and fully-connected layers to implement dual-stream features learning. Through this parallel deep network architecture, identifiable space-time-frequency features of wireless MIMO channels are simultaneously learnt from amplitude and phase information of CSI-MIMO data. Then, in the last fully-connected layer, probabilistic classification results of the dual-stream subnetworks are fused based on Bayes' theorem [33]. Correspondingly, network parameters will be efficiently optimized in the offline training stage and stored in the database. When it comes to the online locating stage, newly gathered CSI-MIMO data are processed to construct the 3D CSI-MIMO matrices which will be further input to the proposed DS-3DCNN. At the final output layer, the fused probabilistic classification results are obtained from spatiotemporal feature maps. Since the outputs of DS-3DCNN can be regarded as the probabilities of a mobile device's location matching some known RPs, the Gaussian mixture model (GMM) [34] is subsequently used to approximate the posterior distribution of the mobile device's location. Finally, a novel estimation algorithm is deduced to estimate the mobile device's location based on the minimum mean square error (MMSE) criterion [35].

In brief, the main contributions of this paper can be summarized as follows:

1) A DS-3DCNN network is proposed to learn spatiotemporal features simultaneously from both amplitude and phase information of CSI-MIMO with a dual-stream network architecture including two parallel subnetworks. Each subnetwork includes several composite-network-units (CNUs), max-pooling layers, and fully-connected layers. And in each CNU, a convolution layer, a BN layer, and an activation function are connected end-to-end. Then, the space-time-frequency features of wireless MIMO channels can be learned from 3D CSI-MIMO matrices through 3D convolutions in two parallel subnetworks, respectively.

2) For the proposed DS-3DCNN network, based on the Bayes' theorem, a fusion strategy is deduced to fuse probabilistic classification results delivered by subnetworks. Consequently, classification results of the output layer in the proposed DS-3DCNN can be regarded as the probabilities of a mobile device's location matching those of the known RPs.

3) To efficiently optimize the network parameters, a novel parallel optimization method is designed for the proposed DS-3DCNN based on the maximum likelihood criterion. Through this method, the whole network optimization problem can be decomposed into two independent optimization sub-problems corresponding to dual-stream subnetworks architecture. Then, the parameters of two subnetworks are available to be efficiently optimized by using stochastic gradient descent (SGD) algorithm [36], separatively.

4) In the online locating stage, the GMM is applied to approximate the posterior distribution of the mobile device's location using the output probabilistic classification results of the proposed DS-3DCNN. And then, utilizing several pairs of input 3D CSI-MIMO matrices gathered from multiple packets, the optimal MMSE estimation can be approximately calculated from the GMM-based posterior distribution according to the Bayes' theorem.

The rest of this paper is organized as follows. In Section II, a preliminary on the basics of CSI-MIMO and fingerprint-based indoor localization system is introduced. In Section III, the proposed CSI-based fingerprinting system is presented in detail; We first describe how to construct 3D CSI-MIMO matrices based on the amplitude outliers elimination and phase sanitization performed on the gathered raw CSI-MIMO data; Then, a DS-3DCNN is designed to realize the supervised learning of spatiotemporal features of CSI-MIMO, and whose network architecture is also introduced in detail; Next, based on the probabilistic classification results of DS-3DCNN, the GMM is applied to approximate the posterior distribution of the mobile devices's location. After that, the optimal estimation of the mobile devices's location is deduced based on the MMSE criterion in the proposed fingerprinting system. Moreover, experimental results and discussion are given in Section IV, including the comparison with some traditional methods. At last, the conclusion is drawn in Section V.

II. PRELIMINARY

A. CSI-MIMO

In WiFi networks with IEEE 802.11 series standards, MIMO and OFDM techniques are utilized to obtain high channel

capacity and transform frequency-selective fading channels into a set of flat-fading channels in parallel [21]. In these networks, CSI-MIMO is generally used to reflect channel properties on subcarriers between every transmit-receive antenna pair [24]. Essentially, CSI is the channel frequency response of wireless channels. Therefore, spectrum features of wireless channels could be also represented by the CSI. Fortunately, several CSI gathering tools have been released to collect raw CSI data by utilizing some NICs such as the Intel 5300 [22] and Qualcomm Atheros products [23].

According to the principle of WiFi systems, the received signal $y_{i,j,n}$ on the nth subcarrier between the ith transmit-antenna and the jth receiver-antenna can be described by [24], [29]

$$y_{i,j,n} = \mathrm{CSI}_{i,j,n} x_{i,j,n} + z_{i,j,n} \tag{1}$$

where $\mathrm{CSI}_{i,j,n}$, and $x_{i,j,n}$ denote the CSI and the transmitted signal, respectively. And $z_{i,j,n}$ represents the corresponding channel noise. Based on [37], $\mathrm{CSI}_{i,j,n}$ is further defined as follows

$$\mathrm{CSI}_{i,j,n} = \sum_{l=1}^{L} \alpha_l \exp\{-\jmath 2\pi (f_0 + n\triangle f)\tau_l\} \tag{2}$$

where $\jmath = \sqrt{-1}$, f_0 is the central frequency, $\triangle f$ represents the frequency interval of adjacent subcarriers, α_l and τ_l are the signal magnitude and time-of-flight (ToF) of the lth signal propagation path, respectively.

Equivalently, $\mathrm{CSI}_{i,j,n}$ could be also rewritten as the following simplified format [24], [37]

$$\mathrm{CSI}_{i,j,n} = |\mathrm{CSI}_{i,j,n}| \exp\{-\jmath \varphi_{i,j,n}\} \tag{3}$$

where $|\mathrm{CSI}_{i,j,n}|$ and $\varphi_{i,j,n}$ denote the amplitude and phase information of the CSI, respectively.

B. FINGERPRINT-BASED INDOOR LOCALIZATION PROBLEM DESCRIPTION

In fingerprinting systems [13], [16], the procedure of indoor locating is generally divided into two stages: the offline training stage and the online locating stage.

In the former stage, first of all, CSI corresponding to several RPs with known locations are collected to construct fingerprint data directly or after some preprocessing. Next, these fingerprint data are used to analyze or learn CSI features which can be applied to efficiently and identifiably represent different patterns corresponding to the locations of RPs.

Furthermore, in the online locating stage, current CSI between the WiFi access point (AP) and the mobile device with unknown location will be gathered and processed for extracting the corresponding features. And then, machine-learning methods can be applied to examine the matching between features of current CSI measurements and the previous versions stored in the database. After that, matching results (in the form of probabilities) are output. At last, the mobile device's location can be estimated through Bayes methods or roughly recognized as the RP's location with highest matching probability.

Consequently, as we described above, it is reasonable to model the fingerprint-based indoor localization as a classification problem which can be solved by machine-leaning techniques such as DL methods [25].

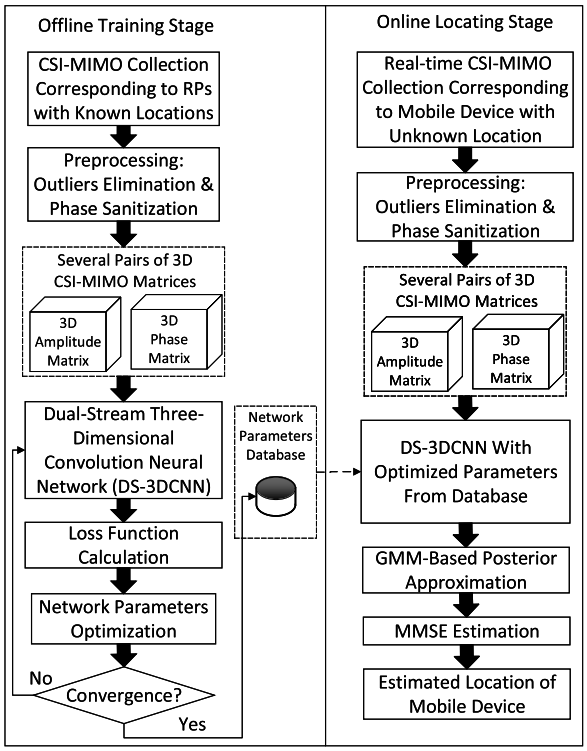


FIGURE 1. The proposed DS-3DCNN fingerprinting localization system.

III. THE PROPOSED CSI-BASED FINGERPRINTING SYSTEM FOR INDOOR LOCALIZATION

In this paper, we design a CSI-based fingerprinting system as shown in Figure 1. In the proposed localization system, WiFi wireless devices are working on the 5GHz frequency band which can provide more stable characters specially in the phase information of CSI-MIMO data. There are two stages included in the procedure of this fingerprinting system: offline training stage and online locating stage.

In the offline training stage, after collecting the raw CSI-MIMO, a data preprocessing including amplitude outliers elimination and phase sanitization is operated for the construction of the pair of 3D CSI-MIMO matrices including a 3D amplitude matrix and a 3D phase matrix. Then, both 3D amplitude and phase matrices are input to the DS-3DCNN for spatiotemporal features learning of CSI-MIMO. Through many rounds of forward supervised feature learning and backward parameter optimizing, the whole DS-3DCNN achieves stable state and obtains optimal classification performance. Then, network parameters are stored in a database.

When it comes to the online locating stage, real-time CSI-MIMO corresponding to a mobile device with unknown location are continuously gathered from several packets and preprocessed to construct 3D amplitude and phase matrices. Since the optimal network parameters are available from the database, spatiotemporal features of CSI-MIMO can be directly learnt utilizing the DS-3DCNN. Completing feature learning, a set of classification results will be output in probabilities. Based on these probabilistic results, the posterior distribution of the mobile device's location is approximated by a GMM. Furthermore, a MMSE estimation algorithm is carried out to locate the mobile device.

In the following sections, we will introduce each component of the proposed fingerprinting system in detail.

A. PREPROCESSING OF RAW CSI-MIMO

As described and proven in [37], due to complex indoor wireless environments, the collected raw CSI-MIMO data are unavoidably disturbed by the channel noise. It makes amplitudes of CSI-MIMO containing a few outliers which are unfavorable to learn detailed features from amplitudes of CSI-MIMO in our system. In addition, the phase information of CSI-MIMO data is also required to be calibrated to cancel the disturbances such as the carrier frequency offset (CFO), the sampling frequency offset (SFO), and the phase noises. Therefore, the preprocessing including amplitude outliers elimination and phase sanitization is indispensable for constructing 3D CSI-MIMO matrices.

1) AMPLITUDE OUTLIERS ELIMINATION

As mentioned in [39], many statistical methods (such as the Chauvenet's Criterion, Pauta criterion, and median absolute deviation (MAD)) have been investigated to detect and eliminate abrupt change points occurring in different kinds of data. Compared with other methods, Pauta criterion has the lower computational complexity and stable detection performance without carefully setting the threshold. Consequently, in this paper, we adopt the Pauta criterion to detect amplitude outliers of raw CSI-MIMO data and further remove them.

let $|\mathrm{CSI}_{i,j,n}(k)|$ denotes the amplitude value of the CSI on the nth subcarrier between the ith transmit-antenna and the jth receive-antenna gathered from the kth packet. Then, based on the Pauta criterion, $|\mathrm{CSI}_{i,j,n}(k)|$ will be identified as an outlier when

$$|d(k)| > 3\sigma \tag{4}$$

where $d(k)$ and σ are calculated as follows, respectively

$$d(k) = |\mathrm{CSI}_{i,j,n}(k)| - \mu_c \tag{5}$$

$$\sigma = \sqrt{\frac{\sum_k d(k)}{K-1}} \tag{6}$$

where μ_c and K are the mean and the number of CSI amplitude values, respectively. After identifying outliers, the outliers of CSI-MIMO data will be replaced by newly gathered samples matching the condition of (4).

2) PHASE SANITIZATION

Besides amplitudes, phase values of gathered CSI-MIMO data may also suffer disturbances from CFO, SFO, and phase noises due to the hardware imperfection on the non-ideal timing and frequency synchronization in OFDM systems. From [29], [37], [40], gathering from the kth packet, the CSI phase value on the nth subcarrier between the ith transmit-antenna and the jth receive-antenna could be written as

$$\varphi_{i,j,n} = \phi_{i,j,n} + \frac{2\pi \nu_n \Delta t}{N} + \theta_C + w \tag{7}$$

where $\phi_{i,j,n}$ is the genuine value of the phase on the the nth subcarrier, ν_n is the subcarrier index of the nth subcarrier, Δt indicates the SFO, θ_C and w represents the CFO and the measurement error, respectively.

As shown in (7), the phase value of the gathered CSI generally contains not only the genuine phase value but also disturbances (including CFO, SFO, measurement error, and the noise), which inevitably destroy the stable and identifiable features of CSI-MIMO data. Referring to [29], [37], [40], a common method for phase sanitization is the linear transformation which can effectively cancel the disturbances with low computational complexity.

Based on [40], the phase slope $\Delta\varphi$ is defined by

$$\Delta\varphi = \frac{\varphi_{i,j,N} - \varphi_{i,j,1}}{\nu_N - \nu_1} = \frac{\phi_{i,j,N} - \phi_{i,j,1}}{\nu_N - \nu_1} - \frac{2\pi(\Delta t)}{N} \tag{8}$$

and the average of all gathered CSI phase values of N subcarriers is

$$\mu_\varphi = \frac{1}{N}\sum_{n=1}^{N} \varphi_{i,j,n} = \mu_\phi + \frac{2\pi\,\Delta t}{N^2}\sum_{n=1}^{N} \nu_n + \theta_C + w \tag{9}$$

where $\mu_\phi = 1/N \sum_{n=1}^{N} \phi_{i,j,n}$ is the average of genuine phase values.

Suppose B_o, f_C, and N denote the OFDM channel bandwidth, the center frequency, and the number of subcarriers respectively. Based on the description in IEEE 802.11n standard [41], the frequency of nth subcarrier can be given by

$$f_n = f_c + \frac{B_o}{N}\nu_n \tag{10}$$

where ν_n is the subcarrier index of the nth subcarrier as in equations (7) and (9). Referring to the Table 7-25f on page 50 of the IEEE 802.11n standard [41], when the WiFi devices are working on the 5GHz center frequency with 40MHz bandwidth, the values of 30 subcarrier indexes which can be used to collect CSIs are $\{-58, -54, -50, -46, -42, -38, -34, -30, -26, -22, -18, -14, -10, -6, -2, 2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58\}$. Therefore, it can be obviously observed that the subcarrier indexes $\{\nu_n, n = 1, \ldots, N\}$ have symmetric properties which lead to

$$\sum_{n=1}^{N} \nu_n = 0 \tag{11}$$

Then, subtracting the linear term $\Delta\varphi\nu_n + \mu_\varphi$ from the phase $\varphi_{i,j,n}$ in (7), the calibrated phase is given by

$$\bar{\varphi}_{i,j,n} = \phi_{i,j,n} - \frac{\phi_{i,j,N} - \phi_{i,j,1}}{\nu_N - \nu_1}\nu_n - \mu_\phi \tag{12}$$

Comparing the equations (7) and (12), after phase sanitization, the disturbances (Δ, θ_C, and w) are cancelled from the calibrated phase $\bar{\varphi}_{i,j,n}$ which presents more stable and identifiable features of CSI.

B. 3D CSI-MIMO MATRICES CONSTRUCTION

As described in [28], [29], CSI amplitudes and phases are widely utilized in fingerprinting systems mostly in the form of two-dimensional (2D) images or matrices. To our knowledge, however, there is still no effort made on jointly learning the fine features of wireless MIMO channels in spatial, temporal, and frequency domains through 3D convolution calculations to improve localization performance of fingerprinting systems. In this paper, in order to realize 3D learning of spatiotemporal features of CSI-MIMO in our fingerprinting system, it is necessary to construct corresponding 3D CSI-MIMO matrices as input data of the follow-up DNN.

Let N_p be the number of packets transmitted over the link of the ith transmit-antenna and the jth receive-antenna, after the preprocessing including the amplitude outliers elimination and phase sanitization, the 3D amplitude matrix $|\mathbf{CSI}|_{i,j}$ and phase matrix $\Phi_{i,j}$ could be constructed as follows

$$|\mathbf{CSI}|_{i,j} = \begin{bmatrix} |\mathrm{CSI}_{i,j,1}(1)| & \dots & |\mathrm{CSI}_{i,j,30}(1)| \\ \vdots & \dots & \vdots \\ |\mathrm{CSI}_{i,j,1}(N_p)| & \dots & |\mathrm{CSI}_{i,j,30}(N_p)| \end{bmatrix} \tag{13}$$

$$\Phi_{i,j} = \begin{bmatrix} |\bar{\varphi}_{i,j,1}(1)| & \dots & |\bar{\varphi}_{i,j,30}(1)| \\ \vdots & \dots & \vdots \\ |\bar{\varphi}_{i,j,1}(N_p)| & \dots & |\bar{\varphi}_{i,j,30}(N_p)| \end{bmatrix} \tag{14}$$

where $|\mathrm{CSI}_{i,j,n}(k)|$ denotes the CSI amplitude value on the nth subcarrier between the ith transmit-antenna and the jth receive-antenna gathered from the kth packet, and $\bar{\varphi}_{i,j,n}(k)$ is the corresponding calibrated phase. Additionally, there are 30 elements in each row of 2D matrices $|\mathbf{CSI}|_{i,j}$ and $\Phi_{i,j}$ because it is feasible to gather CSI data from 30 subcarriers between each transmit-receive antenna pair [24]. Assume that there are three antennas equipped on each transmitter/reveicer in a WiFi network, the 3D amplitude matrix $\mathbf{M}_A$ and phase matrix $\mathbf{M}_P$ of CSI-MIMO are defined as follows (as shown in Figure 2)

$$\begin{aligned} \mathbf{M}_A(\ ,\ ,1) &= \left[\ |\mathbf{CSI}|_{1,1}\ \right], \\ \mathbf{M}_A(\ ,\ ,2) &= \left[\ |\mathbf{CSI}|_{2,1}\ \right], \\ &\vdots \\ \mathbf{M}_A(\ ,\ ,9) &= \left[\ |\mathbf{CSI}|_{3,3}\ \right] \end{aligned} \tag{15}$$

where the matrices $\left\{|\mathbf{CSI}|_{i,j}, i,j = 1,2,3\right\}$ are given by equation (13), and

$$\begin{aligned} \mathbf{M}_P(\ ,\ ,1) &= \left[\ \Phi_{1,1}\ \right], \\ \mathbf{M}_P(\ ,\ ,2) &= \left[\ \Phi_{2,1}\ \right], \\ &\vdots \\ \mathbf{M}_P(\ ,\ ,9) &= \left[\ \Phi_{3,3}\ \right] \end{aligned} \tag{16}$$

where the matrices $\left\{\Phi_{i,j}, i,j = 1,2,3\right\}$ are defined in equation (14).

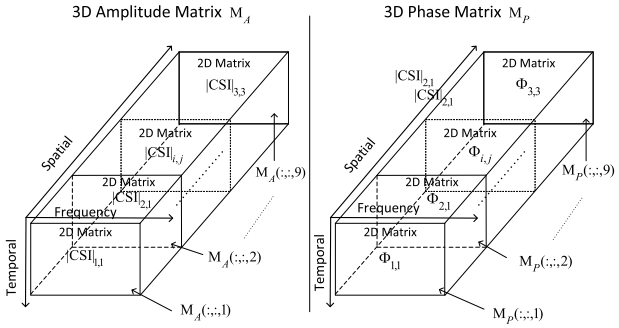


FIGURE 2. Construction of 3D amplitude and phase fingerprint matrices.

It is obviously shown in (15-16) and Figure 2 that the unique features of wireless MIMO channels in spatial, frequency, and temporal domains are centrally embodied in 3D amplitude and phase matrices.

C. DUAL-STREAM 3D CONVOLUTIONAL NEURAL NETWORK

The constructed 3D amplitude and phase matrices in Fig. 2 will be further input to the proposed DS-3DCNN for learning space-time-frequency features of MIMO wireless channels. Differing from 2D-CNNs, this operation is completed by performing a set of 3D convolutions on 3D CSI-MIMO matrices simultaneously in spatial, temporal, and frequency domains, which is more favorable for the subsequent classification.

1) DS-3DCNN ARCHITECTURE

The architecture of the proposed DS-3DCNN is given in Fig. 3. To simultaneously learn features from amplitude and phase information of CSI-MIMO, a dual-stream structure is adopted in the proposed DS-3DCNN which contains two 3D-CNN subnetworks whose inputs are 3D amplitude and phase matrices, respectively.

For the upper amplitude subnetwork (in Fig. 3), following the input layer, we continuously arrange three CNUs with the same internal composition and structure. Each CNU contains a convolution layer followed by a BN layer and a nonlinear activation function such as the rectified linear unit (ReLU) function ($ReLU(x) = \max(0, x)$) [1], [2], in which the convolution layer is responsible for learning features through 3D convolution kernels and obtaining several feature maps. The 3D convolution calculation can be defined as follows [30], [31]

$$v_{l_i,l_j}^{x,y,z} = \sum_{m} \sum_{r=0}^{R_{l_i}-1} \sum_{p=0}^{P_{l_i}-1} \sum_{q=0}^{Q_{l_i}-1} w_{l_i,l_j,m}^{r,p,q} v_{l_i-1,m}^{x+r,y+p,z+q} + b_{l_i,l_j} \tag{17}$$

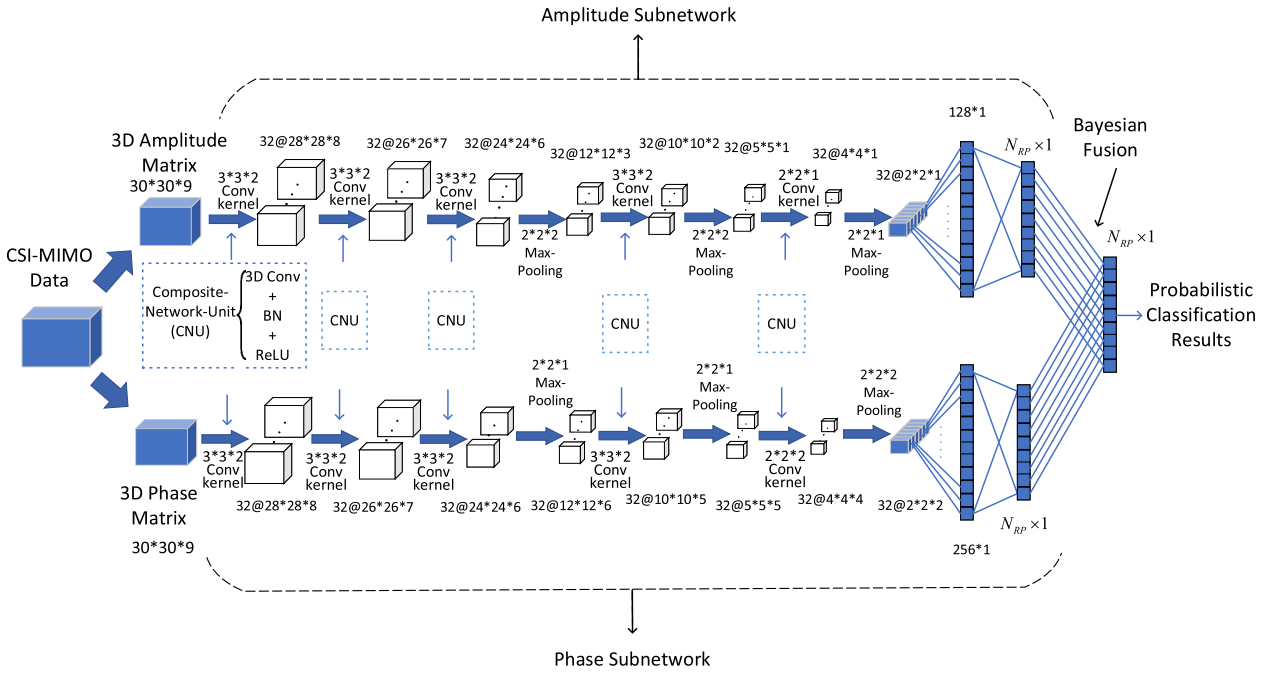


FIGURE 3. The proposed DS-3DCNN deep neural network.

where R_{l_i}, P_{l_i}, and Q_{l_i} are the sizes of the kernel along temporal, frequency, and spatial dimensions respectively. And $v_{l_i,l_j}^{x,y,z}$ is the value of the (x, y, z)th element of the l_jth feature map in the l_i layer, $w_{l_i,l_j,m}^{r,p,q}$ denotes the value of the (r, p, q)th element of the 3D convolution kernel connected to the mth feature map, and b_{l_i,l_j} represents the bias.

Next, feature maps would be completed the de-correlation and normalization operations in the BN layer, besides that the distribution of feature maps can be also modified to accelerate the network training. The BN operation is first proposed in [32] to effectively avoid the over-fitting and accelerate the training speed of DNNs. Following the BN layer, the ReLU activation function is applied on the feature maps to realize the nonlinear mapping with fast training speed. It should be noted that, in first three consecutive CNUs of the amplitude subnetwork, there is no pooling layer for decreasing network parameters. Since the dimensions of input 3D CSI-MIMO matrices are only $30 \times 30 \times 9$, we want to perform more convolutions to learn finer features from the amplitude information of CSI-MIMO. After these three CNUs, feature maps are downsampled through a max-pooling layer. After that, feature maps will continue to go through two CNUs plus max-pooling layers to complete the whole feature learning process. It is obviously shown that, there are 32 convolution kernels with dimensions of $3 \times 3 \times 2$ used in each convolution layer. But in the last convolution layer, dimensions of the convolution kernel is decreased to $2 \times 2 \times 1$ for learning features in detail from smaller feature maps. Next, a fully-connected layer is arranged to generate the 128×1 feature vector which is finally sent to the output layer of the amplitude subnetwork. Then, the following Softmax activation function [1], [2] is used to offer probabilistic classification results whose number is the same as that of RPs

$$\zeta_{l_o} = \frac{e^{\mathbf{w}_{l_o}^T v_{l_o}}}{\sum_{l_o=1}^{N_o} e^{\mathbf{w}_{l_o}^T v_{l_o}}} \tag{18}$$

where ζ_{l_o} denotes the l_oth neuron output in the output layer of the amplitude subnetwork, N_o is the number of neurons, $\mathbf{w}_{l_o}$ is the weight parameter vector of the connected neurons of the link between the second last layer and the output layer of the amplitude subnetwork, and v_{l_o} corresponds to the output of the second last layer.

For the phase subnetwork in the proposed DS-3DCNN, it has the same architecture as the amplitude subnetwork except the dimensions of max-pooling layers and the convolution kernels in the following convolution layers as shown in Fig. 3, which are more suitable for features learning of 3D phase matrices proven by experiments in Section IV. Furthermore, after learning features, the outputs of amplitude and phase subnetworks are fused in the final output layer of the whole DS-3DCNN network based on the Bayes' theorem.

2) BAYESIAN FUSION OF SUBNETWORKS

Given the probabilistic classification results of two-stream subnetworks, the global output results could be obtained by Bayesian fusing in the fusion layer of the proposed DS-3DCNN.

let W_A and W_P denote the parameter sets corresponding to the amplitude and phase subnetworks in the proposed DS-3DCNN, respectively. Given the 3D amplitude and phase matrices $\mathbf{M}_A$ and $\mathbf{M}_P$ defined in (15-16) and

let $\mathbf{X}_A = \{\mathbf{M}_A, \mathrm{W}_A\}$ and $\mathbf{X}_P = \{\mathbf{M}_P, \mathrm{W}_P\}$, based on the Bayes' theorem, the probabilistic classification result corresponding to the l_oth RP's location $\mathbf{r}_{l_o}$ can be expressed in the conditional probability form shown below

$$\mathrm{P}(\mathbf{r}_{l_o}|\mathbf{X}_A, \mathbf{X}_P) = \frac{\mathrm{P}(\mathbf{r}_{l_o}, \mathbf{X}_A, \mathbf{X}_P)}{\mathrm{P}(\mathbf{X}_A, \mathbf{X}_P)} \tag{19}$$

Considering $\mathbf{X}_A$ and $\mathbf{X}_P$ are uncorrelated and hence $\mathrm{P}(\mathbf{X}_A, \mathbf{X}_P|\mathbf{r}_{l_o}) = \mathrm{P}(\mathbf{X}_A|\mathbf{r}_{l_o})\mathrm{P}(\mathbf{X}_P|\mathbf{r}_{l_o})$, the above equation can be further written as

$$\begin{aligned}\mathrm{P}(\mathbf{r}_{l_o}|\mathbf{X}_A, \mathbf{X}_P) &= \frac{\mathrm{P}(\mathbf{X}_A, \mathbf{X}_P|\mathbf{r}_{l_o})\mathrm{P}(\mathbf{r}_{l_o})}{p(\mathbf{X}_A, \mathbf{X}_P)} \\ &\propto \frac{\mathrm{P}(\mathbf{X}_A|\mathbf{r}_{l_o})\mathrm{P}(\mathbf{X}_P|\mathbf{r}_{l_o})}{\mathrm{P}(\mathbf{X}_A, \mathbf{X}_P)} \\ &= \frac{\mathrm{P}(\mathbf{r}_{l_o}|\mathbf{X}_A)\mathrm{P}(\mathbf{X}_A)p(\mathbf{r}_{l_o}|\mathbf{X}_P)\mathrm{P}(\mathbf{X}_P)}{\mathrm{P}(\mathbf{X}_A, \mathbf{X}_P)}\end{aligned} \tag{20}$$

where $\mathrm{P}(\mathbf{r}_{l_o}|\mathbf{X}_A)$ and $\mathrm{P}(\mathbf{r}_{l_o}|\mathbf{X}_P)$ correspond to probabilistic classification results of the amplitude and phase subnetworks, respectively. Since $\mathbf{X}_A$ and $\mathbf{X}_P$ are uncorrelated, $\mathrm{P}(\mathbf{X}_A, \mathbf{X}_P) = \mathrm{P}(\mathbf{X}_A)\mathrm{P}(\mathbf{X}_P)$ is valid, which means that the fused classification result (20) can be given by

$$\mathrm{P}(\mathbf{r}_{l_o}|\mathbf{X}_A, \mathbf{X}_P) \propto \mathrm{P}(\mathbf{r}_{l_o}|\mathbf{X}_A)\mathrm{P}(\mathbf{r}_{l_o}|\mathbf{X}_P) \tag{21}$$

After Bayesian fusing in (21), the following normalization calculation is performed on $\mathrm{P}(\mathbf{r}_{l_o}|\mathbf{X}_A, \mathbf{X}_P)$ to make the sum to be one

$$\mathrm{P}(\mathbf{r}_{l_o}|\mathbf{X}_A, \mathbf{X}_P) = \frac{\mathrm{P}(\mathbf{r}_{l_o}|\mathbf{X}_A, \mathbf{X}_P)}{\sum_{i_o=1}^{N_{RP}} \mathrm{P}(\mathbf{r}_{i_o}|\mathbf{X}_A, \mathbf{X}_P)} \tag{22}$$

Based on the above derivations, it is reasonably concluded that the dual-stream amplitude and phase subnetworks in the proposed DS-3DCNN are independent and complementary, which makes the Bayesian fusion results more accurate than those of any single amplitude/phase subnetwork.

D. OFFLINE TRAINING STAGE

As with other fingerprinting systems, the proposed DS-3DCNN contains two stages in the locating process: offline training stage and online locating stage. In the offline training stage, network parameters are optimized by minimizing a loss function. Next, we will present the derivation of the loss function used in the network optimization.

Let $\mathrm{X}_{in,l_o} = \{\mathbf{M}_A, \mathbf{M}_P\}$ be the set of inputs where $\mathbf{M}_A$ and $\mathbf{M}_P$ denote 3D amplitude and phase matrices corresponding to the location of the l_oth RP, respectively. Additionally, let $\mathrm{W} = \{\mathrm{W}_A, \mathrm{W}_P\}$ denote the whole network parameters where W_A and W_P are the subnetwork parameters. Then, the proposed DS-3DCNN can be formulated as the following mathematic model

$$\tilde{\zeta}_{l_o} = g_{l_o} + z_{l_o} = f(\mathrm{X}_{in,l_o}, \mathrm{W}) + z_{l_o}, \quad l_o = 1, \ldots, N_{RP} \tag{23}$$

where $g_{l_o} = f(\mathrm{X}_{in,l_o}, \mathrm{W})$ describes the nonlinear mapping from the network input X_{in,l_o} and the parameter set W to the target result g_{l_o} through the supervised learning. Additionally, $\tilde{\zeta}_{l_o}$ is the output result of the proposed DS-3DCNN corresponding the l_oth RP, and z_{l_o} denotes the error following Gaussian distribution.

From (23), let $\mathrm{X}_{in} = \left\{\mathrm{X}_{in,l_o}, l_o = 1, \ldots, N_{RP}\right\}$ and $\tilde{\zeta} = \left\{\tilde{\zeta}_{l_o}, l_o = 1, \ldots, N_{RP}\right\}$, then the optimal network parameters $\hat{\mathrm{W}}$ can be obtained as follows based on the maximum likelihood criterion [35]

$$\hat{\mathrm{W}} = \arg\max_{\mathrm{W}} p(\tilde{\zeta}|\mathrm{X}_{in}, \mathrm{W}) = \arg\min_{\mathrm{W}} \log p(\tilde{\zeta}|\mathrm{X}_{in}, \mathrm{W}) \tag{24}$$

Considering that $\mathbf{M}_A$ and W_A are uncorrelated to $\mathbf{M}_P$ and W_P, the equation (24) is further written as follows referring to the derivations in (20-22)

$$\begin{aligned}\hat{\mathrm{W}} &= \arg\min_{\mathrm{W}} \log\left(p(\tilde{\zeta}|\mathbf{M}_A, \mathrm{W}_A)p(\tilde{\zeta}|\mathbf{M}_P, \mathrm{W}_P)\right) \\ &= \arg\min_{\mathrm{W}} \left(\underbrace{\log p(\tilde{\zeta}|\mathbf{M}_A, \mathrm{W}_A)}_{\mathcal{L}_A} + \underbrace{\log p(\tilde{\zeta}|\mathbf{M}_P, \mathrm{W}_P)}_{\mathcal{L}_P}\right)\end{aligned} \tag{25}$$

Since the final output label $\tilde{\zeta}$ corresponds to the fused results of the amplitude and phase subnetworks, it is difficult to calculate the partial derivatives: $\partial\mathcal{L}_A/\mathrm{W}_A$, and $\partial\mathcal{L}_P/\mathrm{W}_P$. Consequently, we make the following assumptions for the convenience of optimizing network parameters

$$\mathcal{L}_A \approx \log p(\tilde{\zeta}^A|\mathbf{M}_A, \mathrm{W}_A) = \sum_{i_o=1}^{N_{RP}} (\tilde{\zeta}_{i_o}^A - g_{i_o})^2 \tag{26}$$

$$\mathcal{L}_P \approx \log p(\tilde{\zeta}^P|\mathbf{M}_P, \mathrm{W}_P) = \sum_{i_o=1}^{N_{RP}} (\tilde{\zeta}_{i_o}^P - g_{i_o})^2 \tag{27}$$

where $\tilde{\zeta}_{i_o}^A$ and $\tilde{\zeta}_{i_o}^P$ are the results before fusing of amplitude and phase subnetworks, respectively. Then, the global parameters optimization is approximately equivalent to the following two separately local optimizations

$$\hat{\mathrm{W}}_A = \arg\min_{\mathrm{W}_A} \mathcal{L}_A \tag{28}$$

$$\hat{\mathrm{W}}_P = \arg\min_{\mathrm{W}_P} \mathcal{L}_P \tag{29}$$

where (28) and (29) represent the optimizations of subnetworks in the proposed DS-3DCNN, respectively. Then, the SGD algorithm [36] is utilized to solve the above optimization problems until the loss functions $\mathcal{L}_A$ and $\mathcal{L}_P$ converge to the values under the given threshold.

Based on the above description, the whole training process of the proposed DS-3DCNN can be roughly summarized in the following three phases: (1) Constructing 3D CSI-MIMO matrices as the training data set; (2) Inputting the 3D CSI-MIMO matrices into the proposed DS-3DCNN network and calculating the subnetworks's loss functions $\mathcal{L}_A$ and $\mathcal{L}_P$; (3) Updating the subnetwork parameters W_A and W_P of DS-3DCNN through SGD algorithm, till $\mathcal{L}_A$ and $\mathcal{L}_P$ are under a given threshold. Next, we give the detailed steps of the training algorithm of the DS-3DCNN in Algorithm 1.

Algorithm 1 Training Algorithm of the Proposed DS-3DCNN

Input: Gathered raw CSI values at all RPs

Output: Optimized parameters of the DS-3DCNN

1: CSI preprocessing: amplitude outliers elimination and phase sanitization
2: Construct 3D CSI-MIMO matrices
3: Set the initial values of network parameters W_A and W_P to be zeros. Furthermore, let the error threshold $\gamma = 10^{-6}$
4: **while** $(\mathrm{W}_A < \gamma)$ and $(\mathrm{W}_P < \gamma)$, **do**
5: Randomly select a mini-batch from all pairs of 3D CSI-MIMO matrices corresponding all RPs
6: Input the mini-batch to the DS-3DCNN
7: Calculate the loss functions $\mathcal{L}_A$ and $\mathcal{L}_P$ in (26-27)
8: Update the amplitude subnetwork parameters W_A through SGD algorithm based on $\mathcal{L}_A$
9: Update the phase subnetwork parameters W_P through SGD algorithm based on $\mathcal{L}_P$
10: **end while**

E. ONLINE LOCATING STAGE

1) TRADITIONAL METHOD

After offline training of the proposed DS-3DCNN network, network parameters are stored in a database. In the online locating stage, when there is a mobile device required to be located, raw CSI-MIMO data are real-time collected and preprocessed to construct a pair of CSI-MIMO matrices including a 3D amplitude matrix and a 3D phase matrix. Next, these 3D matrices are input to the proposed DS-3DCNN adopting optimized parameters stored in the database in advance. In the output layer, using the Softmax activation function, the output of the l_oth neuron can be viewed as the probability that the mobile device locates at l_oth RP. Then, the probability weighted centroid method is commonly used to calculate the estimated location $\hat{\mathbf{r}}_M$ of the mobile device as follows

$$\hat{\mathbf{r}}_M = \sum_{l_o \in \Theta} \mathbf{r}_{l_o} \frac{\zeta_{l_o}}{\sum_{l_o \in \Theta} \zeta_{l_o}} \tag{30}$$

where $\mathbf{r}_{l_o}$ is the location of the l_oth RP, and Θ denotes the set of selected RPs.

Analyzing the equation (30), it can be regarded as a discrete approximation of the following MMSE estimation of the mobile device's location

$$\hat{\mathbf{r}}_M = \mathrm{E}[\mathbf{r}_M|\mathrm{CSI}] = \int_{x \in \Theta} x p(x|\mathrm{CSI}) dx \tag{31}$$

where $\mathrm{E}[\cdot]$ denotes mathematical expectation, Θ is the same set as that in (30), and $p(x|\mathrm{CSI})$ denotes the posterior distribution of the mobile device's location. In this situation, the l_oth neuron's output of the proposed DS-3DCNN in the online locating stage is actually the value of $p(x|\mathrm{CSI})|_{x=\mathbf{r}_{l_o}}$.

However, using (30) to be as the approximate estimation of equation (31) is not accurate enough due to the following two reasons: (1) There are usually not enough candidate RPs in the set Θ; (2) More importantly, the estimation method in (30) is only based on one pair of 3D CSI-MIMO matrices, but several pairs of 3D CSI-MIMO matrices from multiple packets may be favorable to obtain more accurate estimate of the mobile device's location.

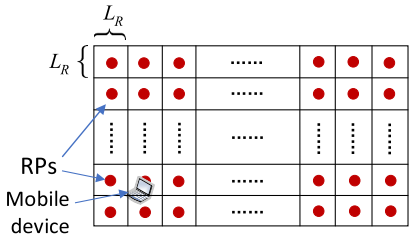


FIGURE 4. Layout of RPs in the monitoring region of the university lab.

In this paper, we try to give a novel MMSE estimation of the mobile device's location based on CSI measurements from multiple packets, and hence provide more accurate localization results of the proposed DS-3DCNN based fingerprinting system. Firstly, we will give the approximate description of the posterior distribution of the mobile device's location.

2) GMM FOR APPROXIMATING POSTERIOR DISTRIBUTION

Let $\mathrm{CSI}(i_t)$ denote the i_tth pair of input 3D CSI matrices (including a 3D fingerprint matrix and a 3D phase fingerprint matrix) of the proposed DS-3DCNN in the online locating stage, and there are totally N_{RP} probabilistic outputs $\{\zeta_{l_o}(i_t), l_o = 1, \ldots, N_{RP}\}$ from the proposed DS-3DCNN.

Since the output $\zeta_{l_o}(i_t)$ could be viewed as the probability of the mobile device locating at the l_oth RP, as shown in following equation

$$\zeta_{l_o}(i_t) = \mathrm{P}(\mathbf{r}_{l_o}|\mathrm{CSI}(i_t)) \tag{32}$$

then the following GMM model is defined to approximate the posterior distribution $p(\mathbf{r}_M|\mathrm{CSI}(i_t))$ of the mobile device's location $\mathbf{r}_M$

$$\begin{aligned} p(\mathbf{r}_M|\mathrm{CSI}(i_t)) &= \sum_{l_o=1}^{N_G} \zeta_{l_o}(i_t) \mathcal{N}(\mathbf{r}_M; \mathbf{r}_{l_o}, \Sigma_{l_o}) \\ &= \sum_{l_o=1}^{N_G} \mathrm{P}(\mathbf{r}_{l_o}|\mathrm{CSI}(i_t)) \mathcal{N}(\mathbf{r}_M; \mathbf{r}_{l_o}, \Sigma_{l_o}) \end{aligned} \tag{33}$$

where N_G is the number of candidate RPs with relatively higher values in $\{\zeta_{l_o}(i_t), l_o = 1, \ldots, N_{RP}\}$, and $\mathcal{N}(\mathbf{r}_M; \mathbf{r}_{l_o}, \Sigma_{l_o})$ denotes the Gaussian distribution with mean $\mathbf{r}_{l_o}$ (which is the location of the l_oth RP) and covariance matrix Σ_{l_o}. Since the RPs are generally designed as the centers of a few of square regions as shown in Fig. 4, it is reasonable to set the covariance matrix $\Sigma = (L_R/2)^2 \mathrm{I}_2$, where L_R is the side length of the square region of the l_oth RP's location, and I_2 is the 2×2 identification matrix.

3) MMSE ESTIMATION OF MOBILE DEVICE'S LOCATION

Without loss of generality, it is assumed that the location of the mobile device does not change within the time of con-

structing and learning N_T pairs of input 3D CSI-MIMO matrices. Let $\mathrm{CSI}_{1:N_T} = \{\mathrm{CSI}(1), \ldots, \mathrm{CSI}(N_T)\}$ denote all the pairs of input 3D CSI-MIMO matrices, the global posterior distribution of the mobile device's location based on $\mathrm{CSI}_{1:N_T}$ can be further given as follows based on the equation (33)

$$p(\mathbf{r}_M|\mathrm{CSI}_{1:N_T}) = \sum_{l_o=1}^{N_G} \mathrm{P}(\mathbf{r}_{l_o}|\mathrm{CSI}_{1:N_T})\mathcal{N}(\mathbf{r}_M; \mathbf{r}_{l_o}, \Sigma_{l_o}) \quad (34)$$

where $\mathrm{P}(\mathbf{r}_{l_o}|\mathrm{CSI}_{1:N_T})$ is the probability of the mobile device locating at l_oth RP. Based on the Bayes' theorem, $\mathrm{P}(\mathbf{r}_{l_o}|\mathrm{CSI}_{1:N_T})$ can be written as

$$\begin{aligned}\mathrm{P}(\mathbf{r}_{l_o}|\mathrm{CSI}_{1:N_T}) &= \frac{\mathrm{P}(\mathrm{CSI}_{1:N_T}|\mathbf{r}_{l_o})\mathrm{P}(\mathbf{r}_{l_o})}{\mathrm{P}(\mathrm{CSI}_{1:N_T})} \\ &= \frac{\left[\prod_{i_t=1}^{N_T} \mathrm{P}(\mathrm{CSI}(i_t)|\mathbf{r}_{l_o})\right]\mathrm{P}(\mathbf{r}_{l_o})}{p(\mathrm{CSI}_{1:N_T})}\end{aligned} \quad (35)$$

where $\mathrm{P}(\mathrm{CSI}(i_t)|\mathbf{r}_{l_o}) = \mathrm{P}(\mathbf{r}_{l_o}|\mathrm{CSI}(i_t))\mathrm{P}(\mathrm{CSI}(i_t))/p(\mathbf{r}_{l_o})$, then the equation (35) can be rewritten as

$$\mathrm{P}(\mathbf{r}_{l_o}|\mathrm{CSI}_{1:N_T}) = \frac{\left[\prod_{i_t=1}^{N_T} \mathrm{P}(\mathbf{r}_{l_o}|\mathrm{CSI}(i_t))\right]\left[\prod_{i_t=1}^{N_T} \mathrm{P}(\mathrm{CSI}(i_t))\right]}{\mathrm{P}(\mathrm{CSI}_{1:N_T})\left[\mathrm{P}(\mathbf{r}_{l_o})\right]^{N_T-1}} \quad (36)$$

where $\mathrm{P}(\mathrm{CSI}(i_t))$ is obviously a constant because $\mathrm{CSI}_{1:N_T}$ are the CSI-MIMO measures, and $\mathrm{P}(\mathbf{r}_{l_o}) = 1/N_{RP}$ is the probability of randomly selecting the l_oth RP from N_{RP} RPs. Consequently, considering that $\{\mathrm{CSI}(1), \ldots, \mathrm{CSI}(N_T)\}$ are separately collected from different packets and uncorrelated with each other, $p(\mathbf{r}_{l_o}|\mathrm{CSI}_{1:N_T})$ in the global posterior distribution (34) is further decided by

$$\mathrm{P}(\mathbf{r}_{l_o}|\mathrm{CSI}_{1:N_T}) \propto \prod_{i_t=1}^{N_T} \mathrm{P}(\mathbf{r}_{l_o}|\mathrm{CSI}(i_t)) \quad (37)$$

Consequently, given N_T pairs of 3D CSI-MIMO matrices gathered from multiple packets and the corresponding probabilistic output results of the proposed DS-3DCNN, we can further approximate the posterior distribution $p(\mathbf{r}_M|\mathrm{CSI}_{1:N_T})$ of the mobile device's location $\mathbf{r}_M$ based on equations (37) and (34). Then, we can calculate the MMSE estimate of the mobile device's location as follows

$$\hat{\mathbf{r}}_M = \mathrm{E}[\mathbf{r}_M|\mathrm{CSI}_{1:T}] = \int x p(x|\mathrm{CSI}_{1:T})dx \approx \sum_{l_o=1}^{\bar{N}_{RP}} \bar{\bar{w}}_{l_o}\mathbf{r}_{l_o} \quad (38)$$

where the parameter $\bar{\bar{w}}_{l_o}$ is given by

$$\bar{\bar{w}}_{l_o} = \frac{\bar{w}_{l_o}}{\sum_{l_o}\bar{w}_{l_o}}, \quad l_o = 1, \ldots, \bar{N}_{RP} \quad (39)$$

and $\{\bar{w}_{l_o}\}_{l_o=1}^{\bar{N}_{RP}}$ are the first $\bar{N}_{RP}$ largest values in $\{\bar{w}_{l_o}\}_{l_o=1}^{N_{RP}}$ which are calculated as follows using the probabilistic output $\{\zeta_{l_o}(i_t), l_o = 1, \ldots, N_{RP}\}$ of the proposed DS-3DCNN based on the equation (37)

$$\bar{w}_{l_o} = \prod_{i_t=1}^{N_T} \zeta_{l_o}(i_t), \quad l_o = 1, \ldots, N_{RP} \quad (40)$$

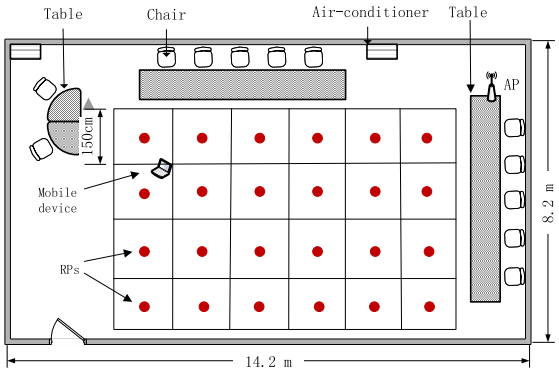


FIGURE 5. Layout of the university laboratory.

In brief, all the steps of the proposed DS-3DCNN based locating/testing algorithm are summarized in Algorithm 2.

Algorithm 2 DS-3DCNN Based Locating/Testing Algorithm

Input: Gathered raw CSI values from the mobile device at an unknown location

Output: Estimated location

1: CSI preprocessing: amplitude outliers elimination and phase sanitization
2: Construct N_T pairs of 3D CSI-MIMO matrices as the inputs of DS-3DCNN
3: **for** $i_t = 1 : N_T$, **do**
4: Obtain outputs $\{\zeta_{l_o}(i_t)\}_{l_o=1}^{N_{RP}}$ of the DS-3DCNN
5: **end for**
6: // MMSE estimation:
7: Calculate the weights $\{\bar{w}_{l_o}\}_{l_o=1}^{N_{RP}}$ using (40)
8: Select the first $\bar{N}_{RP}$ largest values in $\{\bar{w}_{l_o}\}_{l_o=1}^{N_{RP}}$
9: Normalize the selected $\bar{N}_{RP}$ values based on (39)
10: Obtain the estimated location using (38)

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. EXPERIMENTAL ENVIRONMENTS AND CONFIGURATION

To verify the effectiveness of the proposed localization system, two general indoor environments are considered in our experiments: an university laboratory (lab) with dimensions of $14.2 \times 8.2\mathrm{m}^2$ and a long corridor with dimensions of $35 \times 2.2\mathrm{m}^2$.

In the university lab as shown in Figure 5, there are some office furniture (such as tables and chairs), two air-conditioners, and several computers. For the corridor as shown in Figure 6, it is a long and narrow space without any objects in it. Both in the lab and the corridor, there are a desktop computer and a laptop computer used as the AP and the mobile device, respectively. The Intel 5300 NIC with 3 antennas is equipped on both the AP and the mobile device for the transmission of packets and the collection of CSI data on the 5G frequency-band. In the locating region, we set a RP every 1.5 meters for the CSI gathering in the offline training

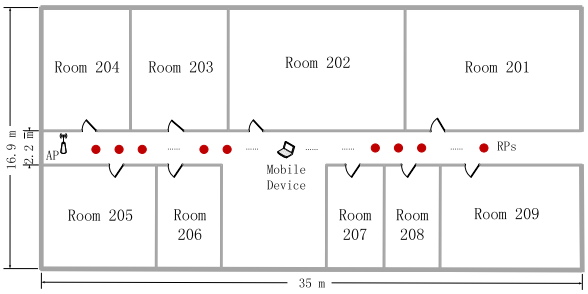


FIGURE 6. Layout of the indoor corrider.

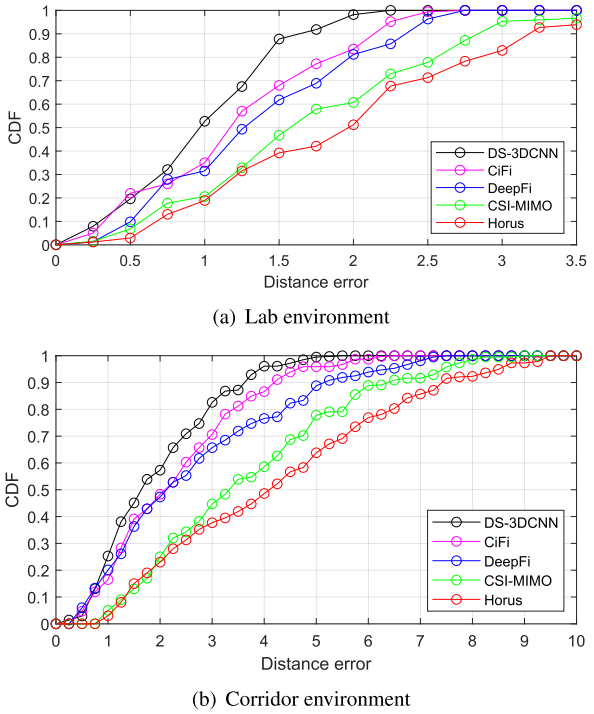
stage, and there are totally 26 and 16 RPs set in the locating regions of the lab and the corridor, respectively. Moreover, there is also a desktop computer with the Geforce 1080Ti graph-processor-unit (GPU) acting as the central server for training the deep neural network.

Before our testing experiments, we need to perform the training of the proposed DS-3DCNN. In the lab environment, at each RP, the laptop receives 12000 packets at each receive-antenna. Based on the constructing method of 3D CSI-MIMO matrices presented in subsection III.A, the laptop can once construct 400 pairs of 3D CSI-MIMO matrices at each RP. Considering all the RPs, there are 10400 pairs of 3D CSI-MIMO matrices are used to construct the whole training data set in the lab experimental environments. Then, the data set is saved in the central server, and the Algorithm 1 is performed to complete the training of the proposed DS-3DCNN. Similarly, in the corridor experimental environment, 800 pairs of 3D CSI-MIMO matrices are constructed at each RP. Correspondingly, considering all the RPs, there are totally 12800 pairs of 3D CSI-MIMO matrices in the training data set for the training of the proposed DS-3DCNN network in the corridor environment.

After obtaining the optimized parameters of the proposed DS-3DCNN, in the online locating/testing experiments, we randomly select 10 testing locations whose coordinates are more or less different from those of RPs in the surveillance regions of the lab and corridor environments, respectively. Then, at each testing location, we use the laptop to receive $(30 * N_T)$ packets at each receive-antenna. Subsequently, N_T pairs of CSI-MIMO matrices can be constructed at each testing location. At last, we perform the Algorithm 2 to achieve the MMSE estimation of the testing location. For each testing location, we do 1000 Monte Carlo locating/testing experiments to obtain the stable performance of our method. It is should be noted that, we can obtain the best locating performance when the value of N_T is set to 5 as shown and proven in our subsequent experiments.

B. COMPARISON OF LOCALIZATION PERFORMANCE

To make a comparison of the localization performance, several state-of-the-art methods are realized in our experiments including Horus [18], CSI-MIMO [24], DeepFi [26], and CiFi [29]. Comparison results in both lab and corridor environments are shown in Figure 7.



(a) Lab environment

(b) Corridor environment

FIGURE 7. Cumulative distribution functions (CDFs) of different methods' localization errors.

1) UNIVERSITY LABORATORY

For the lab environment, localization performance comparison of the five methods are given in Figure 7(a) where the horizontal axis is the localization distance error and the vertical axis is the cumulative distribution functions (CDFs) of distance errors. As shown in Figure 7(a), for the Horus, there are almost 40% test cases whose localization errors are below 1.5 meters, while percentages of other methods are obviously higher than 45%. It seems that the CSI-based locating methods (including CSI-MIMO, DeepFi, CiFi, and the proposed DS-3DCNN) significantly outperform the traditional RSS-based method (Horus). The main reason is that CSI owns more stable characters than RSS whose values exhibit more significant changes in each measurement.

Similarly, compared with Horus and CSI-MIMO, the DL-based methods (including DeepFi, CiFi, and the proposed DS-3DCNN) provide higher percentage of testing cases whose localization errors are below 1.5 meters. It is concluded that DNNs are more effective in learning CSI features for achieving more accurate localization than other methods. Different DNNs are designed and applied to learn different features from raw CSI measurements in DeepFi, CiFi, and the proposed DS-3DCNN systems, respectively. But only the proposed DS-3DCNN system is able to simultaneously learn space-time-fequency features of MIMO channels from both CSI amplitude and phase information. Consequently, more detailed and identifiable features can be used to obtain better classification results. Moreover, due to utilizing the GMM in (34), the posterior distribution of the mobile device's location can be accurately approximated. Hence, the MMSE estimate

TABLE 1. Performance comparison of localization systems for the lab environment.

Methods	Mean error (m)	Std. dev. (m)	Mean running time (s)
Horus	1.794	0.841	0.068
CSI-MIMO	1.631	0.793	0.073
DeepFi	1.340	0.679	0.095
CiFi	1.202	0.647	0.124
DS-3DCNN	0.984	0.502	0.198

of the mobile device's coordinate can be calculated by using the proposed Algorithm 2. Correspondingly, as shown in Figure 7(a), the localization results of the proposed DS-3DCNN method is better than other DL-based methods such as DeepFi and CiFi systems.

In Table 1, we also give the comparison of mean locating error, standard deviation, and mean running time of these five methods. The mean error is defined as follows

$$M_{err} = \frac{\sum_{i_c}^{N_c}\sum_{i_e}^{N_e}\sqrt{(x_{i_e,i_c}-\hat{x}_{i_e,i_c})^2+(y_{i_e,i_c}-\hat{y}_{i_e,i_c})^2}}{N_eN_c} \quad (41)$$

where $(x_{i_e,i_c}, y_{i_e,i_c})$ is the genuine coordinate of the mobile device located on the i_eth testing point, and $(\hat{x}_{i_e,i_c}, \hat{y}_{i_e,i_c})$ denotes the estimated coordinate of the i_eth testing point in the i_cth testing case.

It is obviously shown in Table 1 that, the proposed DS-3DCNN system outperforms other methods in both mean error and standard deviation. Unfortunately, however, this superior locating performance is achieved at the cost of more running time as shown in Table 1. The main reason is 3D convolution calculations in our DS-3DCNN own higher computational complexity than other methods. Therefore, how to further decrease the computational burden of the proposed DS-3DCNN while holding the same lower mean locating error is our next work. Our preliminary research shows that 3D residual networks [42] and distributed DNNs [43] may be a good choice to reach the aim.

2) CORRIDOR

For the corridor environment, comparison results are given in Figure 7(b). Better localization performance are provided by DL-based methods, in which the proposed DS-3DCNN method also provides best localization results. It should be also noted that, however, the localization performance of all methods degrade more or less in the corridor environment compared with those in the lab as shown in Fig. 7(a). The main reason is that there certainly appears more effect of multiple pathes of the wireless channel in the corridor due to its narrow space. Accordingly, both CSI and RSS appear significantly unstable features in the corridor environment which makes the localization more difficult.

Besides CDFs results, other performances comparison of those methods are also shown in the Table 2 such as mean localization error, standard deviation of the localization error, and the mean running time. As obviously shown in the Table 2, the proposed DS-3DCNN reaches the smallest localization error in these methods. In the meanwhile, however, our method also need more running time to implement the smallest locating error mainly because of more 3D convolution calculations required for learning spatiotemporal features of CSI-MIMO.

TABLE 2. Performance comparison of localization systems for the corridor environment.

Methods	Mean error (m)	Std. dev. (m)	Mean running time (s)
Horus	4.216	2.325	0.069
CSI-MIMO	3.640	1.931	0.088
DeepFi	2.614	1.806	0.102
CiFi	2.279	1.362	0.131
DS-3DCNN	1.909	1.125	0.204

TABLE 3. Localization errors of the proposed DS-3DCNN with different dimensions of input 3D CSI-MIMO matrices.

Dimension	Mean error (m)	Std. dev. (m)
$15 \times 30 \times 9$	1.187	0.781
$30 \times 30 \times 9$	0.984	0.502
$45 \times 30 \times 9$	1.027	0.628
$60 \times 30 \times 9$	1.237	0.834

C. IMPACT OF SYSTEM PARAMETERS SETTING

1) DIMENSIONS OF 3D CSI-MIMO MATRICES

As described in [22], [24], We can collect CSI on 30 subcarriers from a single packet between each transmit-receive antenna pair. Therefore, for a 3×3 MIMO-OFDM WiFi system, a pair of input 3D CSI-MIMO matrices with the dimensions of $30 \times 30 \times 9$ can be constructed as shown in Figure 2, through utilizing 30 packets between each transmit-receive antenna pair. In this experiment, we try to use different numbers of packets to construct 3D amplitude and phase matrices with different dimensions for examining the performance of the proposed DS-3DCNN localization system.

The localization errors corresponding to different dimensions of 3D amplitude and phase matrices are given in Table 3. As obviously shown that, the proposed DS-3DCNN achieves the smallest localization error when the dimensions of amplitude and phase matrices are $30 \times 30 \times 9$. It means that the best localization performance will be obtained when using 30 packets between each transmitter-receiver antenna pair to construct 3D CSI-MIMO matrices as the input of the proposed DS-3DCNN.

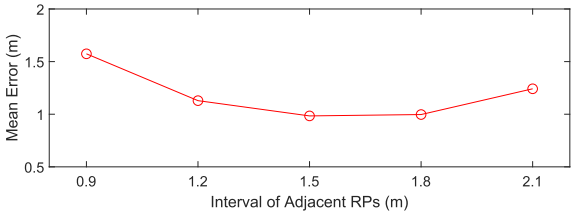
2) NUMBER OF 3D CONVOLUTION KERNELS

For both 2D-CNNs and 3D-CNNs, using more convolution kernels is beneficial for constructing more detailed and identifiable feature maps and hence obtain more reliable recognition or classification results. However, too many convolution operations and corresponding parameters will inevitably increase the computational burden.

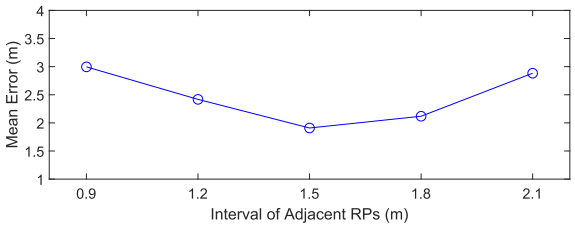
In Table 4, for the lab environment, we give the localization results corresponding to different numbers of 3D convolution kernels in each convolution layer of the proposed DS-3DCNN. From the comparison we can find that the most accurate localization performance is obtained by using

TABLE 4. Localization error vs. number of convolution kernels.

Number of kernels	Mean error (m)	Std. dev. (m)
16	1.154	0.673
32	0.984	0.502
48	0.919	0.482



(a) Lab environment



(b) Corridor environment

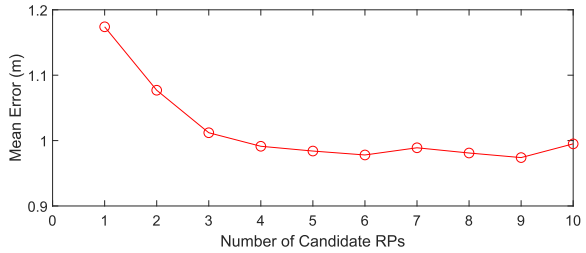
FIGURE 8. Localization errors of the proposed DS-3DCNN versus different intervals of adjacent RPs.

48 3D convolution kernels in our network. When the number of kernels reducing to 32, the localization error increases by 14.3%. When the number of kernels changes from 32 to 16, the localization error rises by more than 17%. Therefore, we decide to use 32 kernels in the convolution layers of our DS-3DCNN to reach a balance between performance and computation cost.

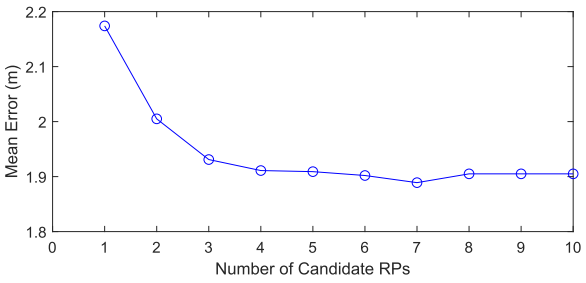
3) INTERVAL OF RPS

As discussed in previous literatures, different numbers of RPs used in the network training lead to various localization errors of fingerprint-based systems, because the changing of the interval between two adjacent RPs will result in different classification results of DNNs. In fig. 8, we select five different intervals between two adjacent RPs to observe the performance changes of the proposed DS-3DCNN system.

As Shown in Fig. 8(a), when the interval is set to 1.5 and 1.8 meters, the optimal localization performance can be obtained by the proposed DS-3DCNN system in the lab environment. However, when we enlarge the interval to 2.1 m, the mean distance error will obviously increase. Similarly, reducing the interval from 1.5 to 0.9 meters, the mean error also increases more or less. Additionally, localization errors with various intervals in the corridor environment are also presented in Fig. 8(b). We can observe that the lowest localization error occurs when the interval of adjacent RPs is 1.5 meter. Consequently, concluding from the results in Fig. 8, the localization performance of the proposed DS-3DCNN system suffers a significant impact of the interval between two adjacent RPs, just liking other fingerprinting



(a) Lab environment



(b) Corridor environment

FIGURE 9. Localization errors of the proposed DS-3DCNN versus the number of candidate RPs.

systems. Without loss of generality, it is proven that bigger intervals make RPs easier to be distinguished, however, it will also bring more built-in errors in the estimation of the mobile device's location. On the other hand, built-in errors are decreased when the interval becomes smaller, but the CSI features corresponding to adjacent RPs are less identifiable and unreliable for matching the testing data with the correct training data.

4) NUMBER OF CANDIDATE RPS

As shown in Algorithm 2, several candidate RPs are selected to calculate the estimate of the mobile device's location. In Figure 9, we give the impact of the number of candidate RPs on final estimation performance in the proposed DS-3DCNN system.

As shown in Figure 9, the localization error of our system is bigger when there are not enough candidate RPs used in GMM to approxiamte the posterior and estimate the mobile device's location in (34,38-40). It is mainly because that the accurate approximation of any distribution by the GMM generally depends on the number of Gaussian components. When there are five RPs used to approximate the posterior, the proposed MMSE estimation reaches relatively stable mean localization error.

However, it should be also noted that more candidate RPs in the GMM will definitely increase the computational complexity of the proposed DS-3DCNN system, but not significantly contribute to the improvement of localization performance.

5) NUMBER OF THE PAIRS OF 3D CSI-MIMO MATRICES

In the proposed DS-3DCNN system, a MMSE estimation method is investigated to estimate the mobile device's

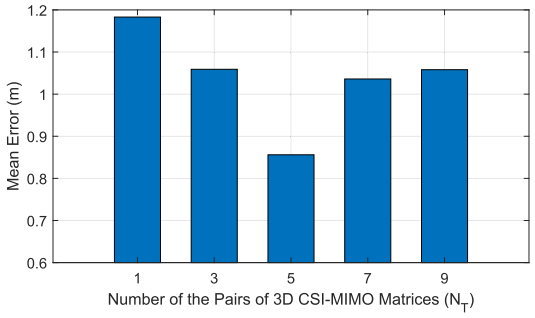


FIGURE 10. Localization errors of the proposed DS-3DCNN system versus different values of N_T for the corridor environment.

location using the outputs of the proposed DS-3DCNN, which is based on the assumption on that the mobile device's location does not change during the constructing and learning of N_T pairs of 3D CSI-MIMO matrices in the proposed DS-3DCNN localization system. The posterior distribution of the mobile device's location is deduced by means of Bayes' theorem in equations (34) and (37). Viewing from the perspective of Bayesian filtering, this step is actually to perform the smoothing on the proposed DS-3DCNN's probabilistic results which are corresponding to N_T pairs of 3D CSI-MIMO matrices. Therefore, it is reasonably expected that the proposed MMSE method is able to bring more accurate localization result than the case of only using the CSI-MIMO from one packet.

Taking the lab environment as an example, the localization errors versus different values of N_T are given in Figure 10. It is obviously shown that, the localization error of our system descents to the lowest value when $N_T = 5$ in (40). We also observe that the localization error becomes higher when increasing the value of N_T to be bigger than 5. The main reason seems that, after many times of product calculation of equation (40), the difference between the adjacent elements $\bar{w}_{l_o}$ and $\bar{w}_{l_o-1}$ becomes larger, which means that elements with larger values in $\{\bar{w}_{l_o}, \quad l_o = 1, \ldots, N_{RP}\}$ become larger, but relatively small elements tend to smaller. Then, after many times of product calculations in (40), the largest value of $\{\bar{\bar{w}}_{l_o}, \quad l_o = 1, \ldots, \bar{N}_{RP}\}$ will approximately equal to one, which means that the estimation of mobile devices's location is only based on one RP. From Fig. 10, when $N_T = 5$, the MMSE estimation method is proven to be feasible to improve localization accuracy of the proposed DS-3DCNN.

V. CONCLUSION

In this paper, we propose a DS-3DCNN to learn space-time-frequency features of MIMO channels from CSI-MIMO data for the fingerprint-based indoor localization. Firstly, CSI data gathered from MIMO-OFDM WiFi systems are preprocessed and used to construct 3D amplitude and phase matrices which are further used as the inputs of the DS-3DCNN network. Several 3D convolution layers, BN layers, max-pooling layers, and fully-connected layers are utilized to learn identifiable features simultaneously from the amplitude and phase information of CSI measurements in the offline training stage. Then, based on the newly collected CSI-MIMO data in the online locating stage, probabilistic classification results are obtained from the output layer and utilized to construct the posterior distribution of the mobile device's location using the GMM. At last, based on the MMSE criterion, a novel algorithm is deduced to achieve the accurate estimation of the mobile device's location. Extensive experiments are performed to examine the localization performance of the proposed DS-3DCNN system. Experimental results show that the proposed DS-3DCNN system outperforms several existing methods, which obviously verify that 3D deep CNNs are feasible and powerful to learn features of MIMO wireless channels jointly in spatial, temporal, and frequency domains and achieve accurate localization in indoor environments.

ACKNOWLEDGMENT

The authors would like to thank the Editors and all the anonymous reviewers for their valuable suggestions and comments. The quality improvement of this article is inseparable from their earnest work.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[2] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[3] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.
[4] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019.
[5] J. Sun, W. Shi, Z. Yang, J. Yang, and G. Gui, "Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems," *IEEE Trans. Veh. Technol.*, to be published.
[6] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and doa estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.
[7] J. Wang, Y. Ding, S. Bian, Y. Peng, M. Liu, and G. Gui, "UL-CSI data driven deep learning for predicting DL-CSI in cellular FDD systems," *IEEE Access*, vol. 7, pp. 96105–96112, 2019.
[8] C.-H. Hsieh, J.-Y. Chen, and B.-H. Nien, "Deep learning-based indoor localization using received signal strength and channel state information," *IEEE Access*, vol. 7, pp. 33256–33267, 2019.
[9] L. Chen, X. Chen, L. Ni, Y. Peng, and D. Fang, "Human behavior recognition using Wi-Fi CSI: Challenges and opportunities," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 112–117, Oct. 2017.
[10] J. Biswas and M. Veloso, "WiFi localization and navigation for autonomous indoor mobile robots," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Anchorage, AK, USA, 2010, pp. 4379–4383.
[11] M. S. Hossain, "Cloud-supported cyber–physical localization framework for patients monitoring," *IEEE Syst. J.*, vol. 11, no. 1, pp. 118–127, Mar. 2017.
[12] C. Yang and H.-R. Shao, "WiFi-based indoor positioning," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 150–157, Mar. 2015.
[13] S. He and S.-H. G. Chan, "Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 466–490, 1st Quart., 2015.
[14] F. Zampella, A. R. J. Ruiz, and F. S. Granja, "Indoor positioning using efficient map matching, RSS measurements, and an improved motion model," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1304–1317, Apr. 2015.
[15] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, Dec. 2013, Art. no. 25.

[16] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. M. Ni, "CSI-based Indoor Localization," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 7, pp. 1300–1309, Jul. 2013.

[17] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. IEEE INFOCOM*, Tel Aviv, Israel, vol. 2, Mar. 2000, pp. 775–784.

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

[19] Y.-K. Cheng, H.-J. Chou, and R. Y. Chang, "Machine-learning indoor localization with access point selection and signal strength reconstruction," in *Proc. IEEE Int. Conf. Veh. Technol. (VTC Spring)*, Nanjing, China, May 2016, pp. 1–5.

[20] Y. Wen, X. Tian, X. Wang, and S. Lu, "Fundamental limits of RSS fingerprinting based indoor localization," in *Proc. IEEE INFOCOM*, Hong Kong, Apr./May 2015, pp. 2479–2487.

[21] L. Hanzo, J. Akhtman, L. Wang, and M. Jiang, *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent Versus Non-Coherent and Cooperative Turbo Transceivers*. Hoboken, NJ, USA: Wiley, 2010.

[22] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, 2011.

[23] S. Sen, J. Lee, K.-H. Kim, and P. Congdon, "Avoiding multipath to revive inbuilding WiFi localization," in *Proc. 11th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, Taipei, Taiwan, Jun. 2013, pp. 249–262.

[24] Y. Chapre, A. Ignjatovic, A. Seneviratne, and S. Jha, "CSI-MIMO: Indoor Wi-Fi fingerprinting system," in *Proc. 39th Annu. IEEE Conf. Local Comput. Netw.*, Edmonton, AB, Canada, Sep. 2014, pp. 202–209.

[25] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2224–2287, 3rd Quart., 2019.

[26] X. Wang, L. Gao, S. Mao, and S. Pandey, "DeepFi: Deep learning for indoor fingerprinting using channel state information," in *Proc. IEEE WCNC*, stanbul, Turkey, Mar. 2015, pp. 1666–1671.

[27] X. Wang, X. Wang, and S. Mao, "ResLoc: Deep residual sharing learning for indoor localization with CSI tensors," in *Proc. IEEE Int. Symp. Pers., Indoor, Mobile Radio Commu. (PIMRC)*, Montreal, QC, Canada, Oct. 2017, pp. 1–6.

[28] 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.

[29] X. Wang, X. Wang, and S. Mao, "Deep convolutional neural networks for indoor localization with CSI images," *IEEE Trans. Netw. Sci. Eng.*, to be published. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8468057/

[30] E. Ahmed, A. Saint, K. Cherenkova, R. Das, G. Gusev, D. Aouada, B. Ottersten, and A. E. R. Shabayek, "A survey on deep learning advances on different 3D data representations," 2018, *arXiv:1808.01462*. [Online]. Available: https://arxiv.org/abs/1808.01462

[31] K. Hara, H. Kataoka, and Y. Satoh, "Can spatiotemporal 3D CNNs retrace the history of 2D CNNs and ImageNet?" in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, Jun. 2018, pp. 6546–6555.

[32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167?context=cs

[33] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics: Distribution Theory*, vol. 1, 6th ed. London, U.K.: Holder Arnold, 1994.

[34] K. N. Plataniotis and D. Hatzinakos, "Gaussian mixtures and their applications to signal processing," in *Advanced Signal Processing Handbook*. Boca Raton, FL, USA: CRC Press, 2001.

[35] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[36] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, Sep. 1951.

[37] L. Zhang, E. Ding, Y. Hu, and Y. Liu, "A novel CSI-based fingerprinting for localization with a single AP," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, Feb. 2019, Art. no. 51.

[38] A. H. Salamah, M. Tamazin, M. A. Sharkas, and M. Khedr, "An enhanced WiFi indoor localization system based on machine learning," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat.*, Alcala de Henares, Spain, Oct. 2016, pp. 1–8.

[39] C. Wang, J. Caja, and E. Gómez, "Comparison of methods for outlier identification in surface characterization," *Measurement*, vol. 117, pp. 312–325, Mar. 2018.

[40] S. Sen, B. Radunovic, T. Minka, and R. R. Choudhury, "You are facing the Mona Lisa: Spot localization using PHY layer information," in *Proc. ACM Int. Conf. Mobile Syst., Appl., Services*, Lake District, U.K., Jun. 2012, pp. 183–196.

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

[42] Z. Qiu, T. Yao, and T. Mei, "Learning spatio-temporal representation with pseudo-3D residual networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 5533–5541.

[43] E.-J. Lim, S.-Y. Ahn, Y.-M. Park, and W. Choi, "Distributed deep learning framework based on shared memory for fast deep neural network training," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Jeju, South Korea, Oct. 2018, pp. 1239–1242.

YUAN JING was born in Liaoyang, China, in 1980. He received the B.S. degree in communication engineering from the Liaoning University of Technology, Jinzhou, China, in 2002, the M.S. degree in communication and information systems from Yanshan University, Qinhuangdao, China, in 2005, and the Ph.D. degree in signal and information processing from the Dalian University of Technology, Dalian, China, in 2009. He joined the Department of Communication Engineering, Liaoning University, Shenyang, China, as a Lecturer, in 2010, and became an Associate Professor, in 2012. He is currently the Dean with the Department of Communication Engineering, Liaoning University. His research interests include statistical signal processing, broadband wireless communication, speech processing, and image processing.

JINSHAN HAO received the B.S. degree in computer science and technology from the Shenyang University of Chemical Technology, Shenyang, China, in 2017. He is currently pursuing the M.S. degree in computer application with Liaoning University, Shenyang. His research interests include signal and information processing, speech signal processing, and computer networks.

PENG LI received the B.S. degree in communication engineering from Liaoning University, Shenyang, China, in 1997, the M.S. and Ph.D. degrees in computer science and technology from Northeast University, Shenyang, in 2005 and 2012, respectively. He is currently the Lecturer of communication engineering, Liaoning University. His research interests include broadband wireless communication, software networks, and computer networks.