
Chapter 1

Deep Learning for Indoor Localization based on Bi-modal CSI Data

Xuyu Wang¹ and Shiwen Mao²

In this chapter, we incorporate deep learning for indoor localization based on channel state information with commodity 5GHz WiFi. We first introduce the state-of-the-art deep learning techniques including deep autoencoder network, convolutional neural network, and recurrent neural network. We then present a deep learning based algorithm to leverage bi-modal CSI data, i.e., average amplitudes and estimated angle of arrivals (AOA), in both offline and online stages of fingerprinting. The proposed scheme is validated with extensive experiments. Finally, we examine several open research problems for indoor localization based on deep learning techniques.

1.1 Introduction

The proliferation of mobile devices has fostered great interest in indoor location-based services, such as indoor navigation, robot tracking in the factories, locating workers on construction sites, and activity recognition [1, 2, 3, 4, 5, 6, 7, 8], all requiring accurately identifying locations of mobile devices indoors. The indoor environment poses a complex radio propagation channel, including multipath propagation, blockage, and shadow fading, and stimulates great research efforts on indoor localization theory and systems [9]. Among various indoor localization schemes, *WiFi-based fingerprinting* is probably one of the most widely used. With fingerprinting, a database is first built with data collected from a thorough measurement of the field in the offline training stage. Then, the position of a mobile user can be estimated by comparing the newly received test data with that in the database. A unique advantage of this approach is that no extra infrastructure needs to be deployed.

Many existing fingerprinting-based indoor localization systems use *received signal strength* (RSS) as fingerprints, due to its simplicity and low hardware requirement [10, 11]. For example, Radar is one of the first RSS-based fingerprinting systems that incorporates a deterministic method for location estimation [10]. For higher accuracy, Horus, another RSS-based fingerprinting scheme, adopts a probabilistic method based on *K*-nearest-neighbor [9] for location estimation [11]. The

¹California State University, Sacramento, CA

²Auburn University, Auburn, AL

performance of RSS-based schemes is limited by two inherent shortcomings of RSS. First, due to the multipath effect and shadow fading, the RSS values are highly diverse, even for consecutively received packets at the same position. Second, RSS value only reflects the coarse channel information, since it is the sum of the powers of all received signals.

Unlike RSS, *channel state information* (CSI) represents fine-grained channel information, which can now be extracted from several commodity WiFi network interface cards (NIC), e.g., Intel WiFi Link 5300 NIC [12], the Atheros AR9390 chipset [13] and the Atheros AR9580 chipset [14]. CSI consists of subcarrier-level measurements of orthogonal frequency division multiplexing (OFDM) channels. It is a more stable representation of channel characteristics than RSS. Several CSI-based fingerprinting systems have been proposed and shown to achieve high localization accuracy [15, 16]. FIFS [15] uses the weighted average of CSI values over multiple antennas. To fully exploit the diversity among the multiple antennas and subcarriers, DeepFi [16] learns a large amount of CSI data from the three antennas and 30 subcarriers with a deep network. However, these CSI-based schemes only use the amplitude information of CSI. The raw phase information is extremely random and not directly usable [17].

Recently, for the Intel 5300 NIC in 2.4GHz, two effective methods are proposed to remove the randomness in raw CSI phase data. In [18], the measured phases from 30 subcarriers are processed with a linear transformation to mitigate the random phase offsets, which is then employed for passive human movement detection. In [17], in addition to the linear transformation, the difference of the sanitized phases of two antennas is obtained and used for line-of-sight (LOS) identification. Although both approaches can stabilize the phase information, the mean value of phase will be zero (i.e., lost) after such processing. This is actually caused by the firmware design of the Intel 5300 NIC when operating on the 2.4GHz band [19]. To address this issue, Phaser [19] is the first to exploit CSI phase in 5GHz WiFi. Phaser constructs an angle of arriving (AOA) pseudospectrum for phase calibration in single Intel 5300 NIC. These interesting works greatly motivate us to explore effectively cleansed phase information for indoor fingerprinting with commodity 5GHz WiFi.

In this chapter, we investigate the problem of fingerprinting-based indoor localization with commodity 5GHz WiFi. We first present three hypotheses on CSI amplitude and phase information for 5GHz OFDM channels. *First*, the average amplitude over two antennas is more stable than that from a single antenna as well as RSS. *Second*, the phase difference of CSI values from two antennas in 5GHz is highly stable. Due to the firmware design of Intel 5300 NIC, the phase differences of consecutively received packets form four clusters when operating in 2.4GHz. Such ambiguity makes measured phase difference unusable. However, we find this phenomenon does not exist in the 5GHz band, where all the phase differences concentrate around one value. We further design a simple multi-radio hardware for phase calibration which is greatly different from the technique [19] that uses AOA pseudospectrum searching with high computation complexity to calibrate phase in single Intel 5300 NIC. As a result, the randomness from the time and frequency difference between the transmitter and receiver, and the unknown phase offset can all be re-

moved; and stable phase information can be obtained. *Third*, the calibrated phase difference in 5GHz can be translated into AOA with considerable accuracy when there is a strong LOS component. We validate these hypotheses with both extensive experiments and simple analysis.

We then design BiLoc, bi-modal deep learning for indoor localization with commodity 5GHz WiFi, to incorporate the three hypotheses in an indoor fingerprinting system [20]. In BiLoc, we first extract raw amplitude and phase data from the three antennas, each with 30 subcarriers, with a modified firmware. We then obtain bi-modal data, including average amplitudes over pairs of antennas and estimated AOAs, with the calibration procedure discussed above. In the offline training stage, we adopt a deep network with three hidden layers to extract the unique channel features hidden in the bi-modal data, and propose to use the weights of the deep network to store the extracted features (i.e., fingerprints). To reduce the computational complexity, we propose a greedy learning algorithm to train the deep network in a layer-by-layer manner with a Restricted Boltzmann Machine (RBM) model. In the online test stage, bi-modal test data is first collected for a mobile device. Then a Bayesian probability model based on the radial basis function (RBF) is leveraged for accurate online position estimation.

In the rest of this chapter, deep learning for indoor localization is introduced 1.2. Then, the preliminaries and hypotheses are given in Section 1.3. We present the BiLoc system in Section 1.4 and validate its performance in Section 1.5. Section 1.6 discusses future research problems for indoor localization and Section 1.7 concludes this paper.

1.2 Deep Learning for Indoor Localization

With the rapid growth of computation platforms like Tensorflow, Caffe and Torch [21], Deep learning has been widely applied in a variety of areas such as object recognition, natural language processing, computer vision, Robotics, automated vehicles and AI games [22]. Compared with shadowing machine learning algorithms such support vector machine (SVM) and K-nearest neighbor (KNN), deep learning is a branch of machine learning, which implements non-linear transformations with multiple hidden layers and has high-level data abstractions. In addition, Deep learning can train the weights and bias of the network with a huge quantity of data for improving classification performance and data representation capacity, which mainly includes unsupervised and supervised learning with different deep learning models [23]. In the chapter, three different deep learning frameworks are discussed for indoor localization problems as the following.

1.2.1 Autoencoder Neural Network

A deep autoencoder neural network is an unsupervised learning, which can produce the output data that is a de-noised input data. Moreover, it is also used to extract data features or reduce the size of data, which is more powerful than principal component analysis (PCA) based methods because of its non-linear transformations with mul-

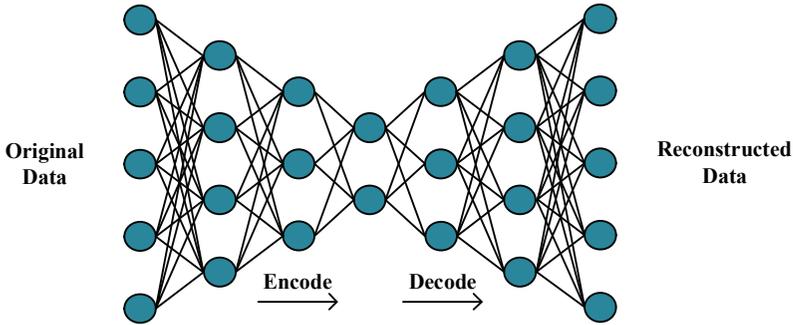


Figure 1.1 Autoencoder.

multiple hidden layers. Fig. 1.1 shows the architecture of the deep autoencoder neural network. For training phase, a deep autoencoder neural network has three stages with pretraining, unrolling, and fine-tuning [24]. In the pretraining stage, each neighboring set of two layer is considered as a restricted boltzmann machines(RBM), which is denoted as a bipartite undirected graphical model. Then, a greedy algorithm is used to train the weights and biases for a stack of RBMs. In the unrolling stage, the deep autoencoder network is unrolled to obtain the reconstructed input data. Finally, the fine-tuning phase employs the *backpropagation* algorithm for training the weights in the deep autoencoder network by minimizing the error.

Based on deep autoencoder networks, the first work for indoor localization is DeepFi [16, 25] system, which is a deep autoencoder network based indoor fingerprinting method with CSI amplitudes. For every training location, the deep autoencoder network is trained to get a set of weights and biases, which are used as fingerprints for corresponding locations. For online test, the final location is estimated based on Bayesian scheme. The experimental results show that the mean distance error in the living room and the laboratory is about 1.2 and 2.3 meters, respectively. In addition, PhaseFi [26, 27] is proposed using CSI calibrated phase, which still use deep autoencoder networks for indoor localization. Moreover, deep autoencoder networks are used for device-free indoor localization problems [28, 29]. The denoising autoencoder-based indoor localization with BLE is also used to provide 3-D localization [30]. In this chapter, we still consider deep autoencoder networks for indoor localization using bimodal CSI data.

1.2.2 Convolutional Neural Network

Convolutional neural network (CNN) is also a useful deep learning architecture, which has been successfully used in computer vision and activity recognition [31, 23]. In 1998, LeCun proposed LeNet-5 [32], which is the first architecture of CNN that is called LeNet-5. Fig. 1.1 shows CNN framework, which includes convolutional layers, subsampling layers, and fully connected layers.

The convolutional layer can obtain feature maps within local regions in the previous layer's feature maps with linear convolutional filters, which is followed by

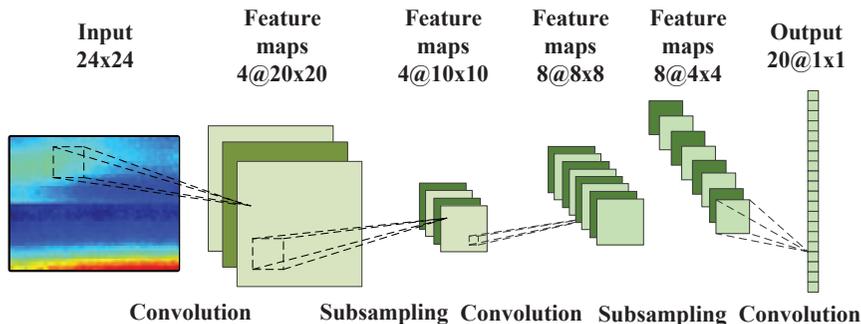


Figure 1.2 CNN.

nonlinear activation functions. For the subsampling layer, it can decrease the resolution of the feature maps by downsampling over a local neighborhood in the feature maps of the previous layer, which is invariant to distortions on the input data [33]. The feature maps in the previous layer are pooled over a local temporal neighborhood using the mean pooling function. Other operations such as the sum or max pooling function can be also exploited in the subsampling layer.

After the convolutional and subsampling layers, the fully-connected layer can consider a basic neural network with one hidden layer to train the output data. Moreover, the loss function is used to measure the difference between the true location label and the output data of CNN, where the squared error or cross-entropy is used as loss function for training these weights. Currently, increasing number of CNN models are proposed, such as AlexNet [31] and ResNet [34]. AlexNet is a bigger and more complex model, where Max pooling and ReLU nonlinear activation function are used in the model [35]. Moreover, the dropout regularization is used to handle overfitting problem. ResNet was proposed by Microsoft, where residual block includes a direct path between the input and output, and adds batch normalization technique to avoid diminishing or exploding of the gradient. ResNet is a 152 layers residual learning framework, which won the ILSVRC 2015 classification task [31].

For indoor localization problems, CiFi [33, 36] system leverages generated images with estimated AOA values with commodity 5GHz WiFi as the input of CNN for indoor localization, which can be trained by backpropagation (BP) algorithm. This system demonstrates that the performance of the localization has outperformed existing schemes, like FIFS and Horus. Motivated by ResNet, ResLoc [37] system uses bimodal CSI tensor data as input to train deep network with a *deep residual sharing learning*, which can achieve the best performance in deep learning based localization methods using CSI. CSI amplitude is also used to obtain CSI images for indoor localization [38]. In addition, input images by using RSSI of Wi-Fi signals are leveraged to train a CNN model [39, 40]. Moreover, CNN is used for TDoA-based localization systems, which can estimate non-linearities in the signal propagation space but also predict the signal for multipath effects [41].

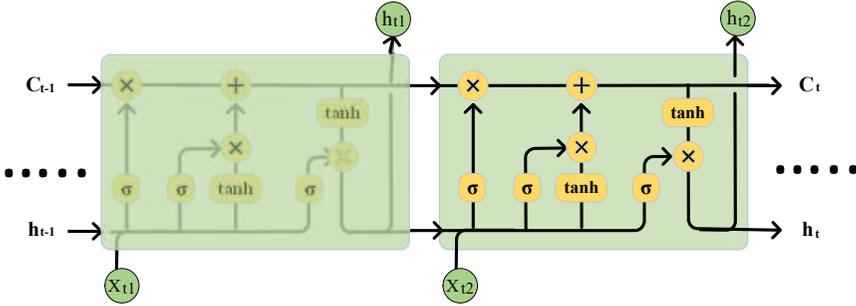


Figure 1.3 LSTM.

1.2.3 Long Short-term Memory

To process variable-length sequence inputs, recurrent neural networks (RNN) are proposed, where long-term dependencies can be solved using the feedback loop in the recurrent layer. However, the dependencies also lead to that RNN cannot be successfully trained, because of diminishing or exploding of the gradient of the loss function. Long short-term memory (LSTM) architecture is proposed to handle the above problem, which has widely applications for sequence data processing [42].

For the LSTM algorithm in Fig. 1.3, the input gate i decides how much new information will be exploited in the current memory cell, the forget gate f controls how much information will be removed from the old memory cell, and the output gate o determines how much data will be output based on the current memory cell c . In addition, the sigmoid function σ can control how much information can be updates and hyperbolic tangent function \tanh can create new candidate values g . Thus, unlike RNN, LSTM scheme can handle long-term dependency and has better data representation ability, which has been employed for speech recognition, machine translation, and time-series problems.

Currently, the DeepML system uses two layers LSTM network for obtaining a higher learning and representation ability with magnetic and light sensor data for indoor localization, which can achieve sub-meter level localization accuracy [43]. LSTM framework can be used for sequence based localization problems with other signals. We have also applied LSTM to wheat moisture level detection [44] and forecasting of renewable energy generation [45].

1.3 Preliminaries and Hypotheses

1.3.1 Channel State Information Preliminaries

OFDM is widely used in wireless network standards, such as WiFi (i.e., IEEE 802.11a/g/n), where the total spectrum is partitioned into multiple orthogonal subcarriers, and wireless data is transmitted over the subcarriers using the same modulation and coding scheme (MCS) to mitigate frequency selective fading. Leveraging the device driver for off-the-shelf NICs, e.g., the Intel 5300 NIC, we can extract CSI that is a

fine-grained physical layer (PHY) information. CSI reveals the channel characteristics experienced by the received signal such as the multipath effect, shadow fading, and distortion.

With OFDM, the WiFi channel at the 5GHz band can be considered as a narrowband flat fading channel. In the frequency domain, the channel model can be expressed as

$$\vec{Y} = \text{CSI} \cdot \vec{X} + \vec{N}, \quad (1.1)$$

where \vec{Y} and \vec{X} denote the received and transmitted signal vectors, respectively, \vec{N} is the additive white Gaussian noise and CSI represents the channel's frequency response, which can be computed from \vec{Y} and \vec{X} .

Although a WiFi receiver uses an OFDM system with 56 subcarriers for a 20 MHz channel, the Intel 5300 NIC can report 30 out of 56 subcarriers. The channel frequency response of subcarrier i , CSI_i , is a complex value, that is

$$\text{CSI}_i = \mathcal{I}_i + j\mathcal{Q}_i = |\text{CSI}_i| \exp(j\angle \text{CSI}_i). \quad (1.2)$$

where \mathcal{I}_i and \mathcal{Q}_i are the in-phase component and quadrature component, respectively; $|\text{CSI}_i|$ and $\angle \text{CSI}_i$ are the amplitude response and phase response of subcarrier i , respectively.

1.3.2 Distribution of Amplitude and Phase

In general both \mathcal{I}_i and \mathcal{Q}_i can be modeled as i.i.d. AWGN of variance σ^2 . The amplitude response is $|\text{CSI}_i| = \sqrt{\mathcal{I}_i^2 + \mathcal{Q}_i^2}$, which follows a Rician distribution when there is a strong LOS component [46]. The probability distribution function (PDF) of the amplitude response is given by

$$f(|\text{CSI}_i|) = \frac{|\text{CSI}_i|}{\sigma^2} \times \exp\left(-\frac{|\text{CSI}_i|^2 + |\text{CSI}_0|^2}{2\sigma^2}\right) \times I_0\left(\frac{|\text{CSI}_i| \cdot |\text{CSI}_0|}{\sigma^2}\right), \quad (1.3)$$

where $|\text{CSI}_0|$ is the amplitude response without noise, $I_0(\cdot)$ is the zeroth order modified Bessel function of the first kind. When the signal to noise ratio (SNR) is high, the PDF $f(|\text{CSI}_i|)$ will converge to the Gaussian distribution as $\mathcal{N}(\sqrt{|\text{CSI}_0|^2 + \sigma^2}, \sigma^2)$ [46].

The phase response of subcarrier i is computed by $\angle \text{CSI}_i = \arctan(\mathcal{Q}_i/\mathcal{I}_i)$ [46]. The phase PDF is given by

$$f(\angle \text{CSI}_i) = \frac{1}{2\pi} \exp\left(-\frac{|\text{CSI}_0|^2}{2\sigma^2}\right) \left(1 + \frac{|\text{CSI}_0|}{\sigma} \sqrt{2\pi} \cos(\angle \text{CSI}_i) \times \exp\left(\frac{|\text{CSI}_0|^2 \cos^2(\angle \text{CSI}_i)}{2\sigma^2}\right) \left(1 - Q\left(\frac{|\text{CSI}_0| \cos(\angle \text{CSI}_i)}{\sigma}\right)\right)\right),$$

where $Q(\cdot)$ is the Q-function. In the high SNR regime, the PDF $f(\angle \text{CSI}_i)$ also converges to a Gaussian distribution as $\mathcal{N}(0, (\sigma/|\text{CSI}_0|)^2)$ [46]. The distribution of amplitude and phase of the subcarriers would be useful to guide the design of localization algorithms.

1.3.3 Hypotheses

We next present three important hypotheses about the CSI data for 5GHz OFDM channels, which are demonstrated and tested with our measurement study and theoretical analysis.

1.3.3.1 Hypothesis 1

The average CSI amplitude value over two adjacent antennas for the 5GHz OFDM channel is highly stable at a fixed location.

We find CSI amplitude values exhibit great stability for continuously received packets at a given location. Fig. 1.4 presents the cumulative distribution functions (CDF) of the standard deviations (STD) of (i) the normalized CSI amplitude averaged over two adjacent antennas, (ii) the normalized CSI amplitude from a single antenna, and (iii) the normalized RSS amplitude from a single antenna, for 90 positions. At each position, 50 consecutive packets are received by the Intel 5300 NIC operating on the 5GHz band. It can be seen that 90% of the testing positions are below 10% of the STD in the case of averaged CSI amplitudes, while the percentage is 80% for the case of single antenna CSI and 70% for the case of single antenna RSS. Thus, averaging over two adjacent antennas can make CSI amplitude highly stable for a fixed location with 5GHz OFDM channels. We conduct the measurements over a long period of time, including midnight hours and business hours. No obvious difference in the stability of CSI is observed over different times, while RSS values exhibit large variations even for the same position. This finding motivates us to use average CSI amplitudes of two adjacent antennas as one of the features of deep learning in the BiLoc design.

Recall that the PDF of the amplitude response of a single antenna is Gaussian in the high SNR regime. Assuming that the CSI values of the two antennas are i.i.d. (true when the antennas are more than a half wavelength apart [17]), the average CSI amplitudes also follow the Gaussian distribution, as $\mathcal{N}(\sqrt{|\text{CSI}_0|^2 + \sigma^2}, \sigma^2/2)$, but with a smaller variance. This proves that stability can be improved by averaging CSI amplitudes over two antennas [47](as observed in Fig. 1.4). On the other hand, we consider the average CSI amplitudes over two antennas instead of three antennas or only one antenna, because BiLoc system employs a bi-model data, such as estimated AOAs and average amplitudes. This requires that we use the same number of nodes as the input for deep network.

1.3.3.2 Hypothesis 2

The difference of CSI phase values between two antennas of the 5GHz OFDM channel is highly stable, compared to that of the 2.4GHz OFDM channel.

Although the CSI phase information is also available from the Intel 5300 NIC, it is highly random and cannot be directly used for localization, due to noise and the unsynchronized time and frequency of the transmitter and receiver. Recently, two useful algorithms are used to remove the randomness in CSI phase. The first approach is to make a linear transform of the phase values measured from the 30 subcarriers [18]. The other one is to exploit the phase difference between two antennas in 2.4GHz and then remove the measured average [17]. Although both methods

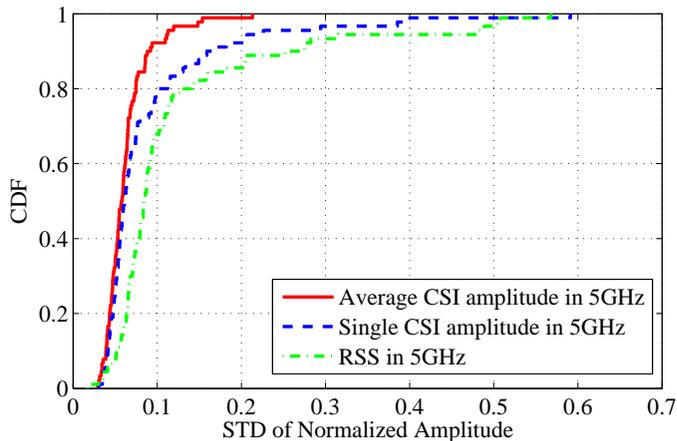


Figure 1.4 CDF of the standard deviations of normalized average CSI amplitude, a single CSI amplitude, and a single RSS in the 5GHz OFDM channel for 90 positions.

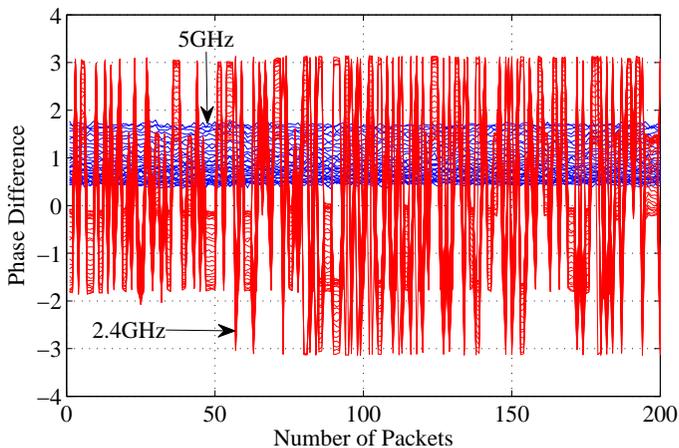


Figure 1.5 CDF of the standard deviations of normalized average CSI amplitude, a single CSI amplitude, and a single RSS in the 5GHz OFDM channel for 90 positions.

can stabilize the CSI phase in consecutive packets, the average phase value they produce is always near zero, which is different from the real phase value of the received signal.

Switching to the 5GHz band, we find the phase difference becomes highly stable. In Fig. 1.5, we plot the measured phase differences of the 30 subcarriers between two antennas for 200 consecutively received packets in the 5GHz (in blue) and 2.4GHz (in red) bands. The phase difference of the 5GHz channel varies be-

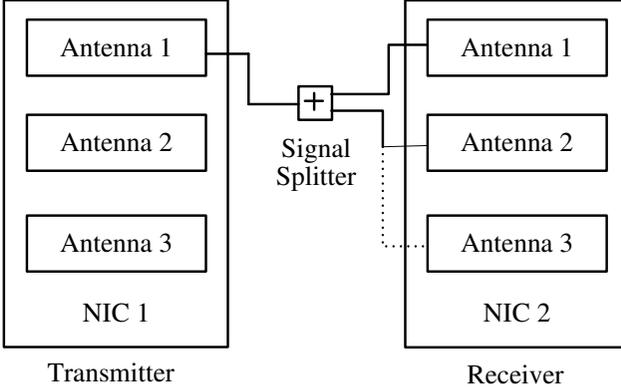


Figure 1.6 The multi-radio hardware design for calibrating the unknown phase offset difference $\Delta\beta$.

tween $[0.5, 1.8]$, which is considerably more stable than that of the 2.4GHz channel (varies between $[-\pi, \pi]$). To further illustrate this finding, we plot the measured phase differences between the 5th subcarrier of two antennas using polar coordinates in Fig. 1.7. We find that all the 5GHz measurements concentrate around 30° , while the 2.4GHz measurements form four clusters around 0° , 90° , 180° , and 270° . It is because of the firmware design of the Intel 5300 NIC when operating on the 2.4GHz band, which reports the phase of channel modulo $\pi/2$ rather than 2π on the 5GHz band [19]. Comparing to the ambiguity in the 2.4GHz band, the highly stable phase difference in the 5GHz band could be very useful for indoor localization.

As in Hypothesis 1, we also provide an analysis to validate the observation from experiments. Let $\angle\widehat{CSI}_i$ denote the measured phase of subcarrier i , which is given by [48, 14]

$$\angle\widehat{CSI}_i = \angle CSI_i + (\lambda_p + \lambda_s)m_i + \lambda_c + \beta + Z, \quad (1.4)$$

where $\angle CSI_i$ is the true phase from wireless propagation, Z is the measurement noise, β is the initial phase offset because of the phase-locked loop (PLL), m_i is the subcarrier index of subcarrier i , λ_p , λ_s and λ_c are phase errors from the packet boundary detection (PBD), the sampling frequency offset (SFO) and central frequency offset (CFO), respectively [48], which are expressed by

$$\begin{cases} \lambda_p = 2\pi\frac{\Delta t}{N} \\ \lambda_s = 2\pi\left(\frac{T'-T}{T}\right)\frac{T_s}{T_u}n \\ \lambda_c = 2\pi\Delta f T_s n, \end{cases} \quad (1.5)$$

where Δt is the packet boundary detection delay, N is the FFT size, T' and T are the sampling periods from the receiver and the transmitter, respectively, T_u is the length of the data symbol, T_s is the total length of the data symbol and the guard interval, n is the sampling time offset for current packet, Δf is the center frequency difference between the transmitter and receiver. It is noticed that we cannot obtain the exact values about Δt , $\frac{T'-T}{T}$, n , Δf , and β in (1.4) and (1.5). Moreover, λ_p , λ_s and λ_c vary

for different packets with different Δt and n . Thus, the true phase $\angle \text{CSI}_i$ cannot be derived from the measured phase value.

However, note that the three antennas of the Intel 5300 NIC use the same clock and the same down-converter frequency. Consequently, the measured phases of subcarrier i from two antennas have identical packet detection delay, sampling periods and frequency differences (and the same m_i) [19]. Thus the measured phase difference on subcarrier i between two antennas can be approximated as

$$\Delta \widehat{\angle \text{CSI}}_i = \Delta \angle \text{CSI}_i + \Delta \beta + \Delta Z, \quad (1.6)$$

where $\Delta \angle \text{CSI}_i$ is the true phase difference of subcarrier i , $\Delta \beta$ is the unknown difference in phase offsets, which is in fact a constant [19], and ΔZ is the noise difference. We can find that $\Delta \widehat{\angle \text{CSI}}_i$ is stable for different packets because of the above equation (1.6) without Δt and n .

In the high SNR regime, the PDF of the phase response of subcarrier i for each of the antennas is $\mathcal{N}(0, (\sigma/|\text{CSI}_0|)^2)$. Due to the independent phase responses, the measured phase difference of subcarrier i is also Gaussian with $\mathcal{N}(\Delta \beta, 2\sigma^2(1 + 1/|\text{CSI}_0|^2))$. Note that although the variance is higher comparing to the true phase response, the uncertainty from the time and frequency differences is removed, leading to much more stable measurements (as shown in Fig. (1.7)).

1.3.3.3 Hypothesis 3

The calibrated phase difference in 5GHz can be translated into the angle of arriving (AOA) with considerable accuracy when there is a strong LOS component.

The measured phase difference on subscriber i can be translated into an estimation of angle of arrival (AOA), as

$$\theta = \arcsin \left(\frac{\Delta \widehat{\angle \text{CSI}}_i \lambda}{2\pi d} \right), \quad (1.7)$$

where λ is the wavelength and d is the distance between the two antennas (set to $d = 0.5\lambda$ in our experiments). Although the measured phase difference $\Delta \widehat{\angle \text{CSI}}_i$ is highly stable, we still wish to remove the unknown phase offset difference $\Delta \beta$ to further reduce the error of AOA estimation. For commodity WiFi devices, the only existing approach for a single NIC, to the best of our knowledge, is to search for $\Delta \beta$ within an AOA pseudospectrum in the range of $[-\pi, \pi]$, which, however, has a high time complexity [19].

In this chapter, we design a simple method to remove the unknown phase offset difference $\Delta \beta$ using two Intel 5300 NICs. As in Fig. 1.6, we use one Intel 5300 NIC as transmitter and the other as receiver, while a *signal splitter* is used to route signal from antenna 1 of the transmitter to antennas 1 and 2 of the receiver through cables of the same length. Since the two antennas receive the same signal, the true phase difference $\Delta \angle \text{CSI}_i$ of subcarrier i is zero. We can thus obtain $\Delta \beta$ as the measured phase offset difference between antennas 1 and 2 of the receiver. We also use the same method to calibrate antennas 2 and 3 of the receiver, to obtain the unknown phase offset difference between them as well. We find that the unknown phase offset difference is relatively stable over time.

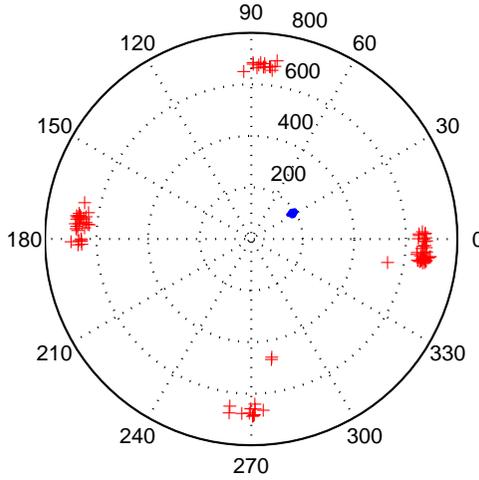


Figure 1.7 The measured phase differences of the 5th subcarrier between two antennas for 200 consecutively received packets in the 5GHz (blue dots) and 2.4GHz (red crosses) bands.

Having calibrated the unknown phase offset differences for the three antennas, we then use the MUSIC algorithm for AOA estimation [49]. In Fig. 1.8, the AOA estimation using MUSIC with the calibrated phase information for the 30 subcarriers is plotted for a high SNR signal with a known incoming direction of 14° . We can see that the peak occurs at around 20° in Fig. 1.8, indicating an AOA estimation error of about 6° .

We can obtain the true incoming angle with MUSIC when the LOS component is strong. To deal with the case with strong NLOS paths (typical in indoor environments), we adopt a deep network with three hidden layers to learn the estimated AOA and the average amplitudes of adjacent antenna pairs as fingerprints for indoor localization. As input to the deep network, the estimated AOA is obtained as follows.

$$\theta = \arcsin \left(\left(\Delta \angle \widehat{CSI}_i - \Delta \beta \right) \frac{\lambda}{2\pi d} \right) + \frac{\pi}{2}, \quad (1.8)$$

where $\Delta \beta$ is measured with the proposed multi-radio hardware experiment. The estimated AOA is in the range of $[0, \pi]$.

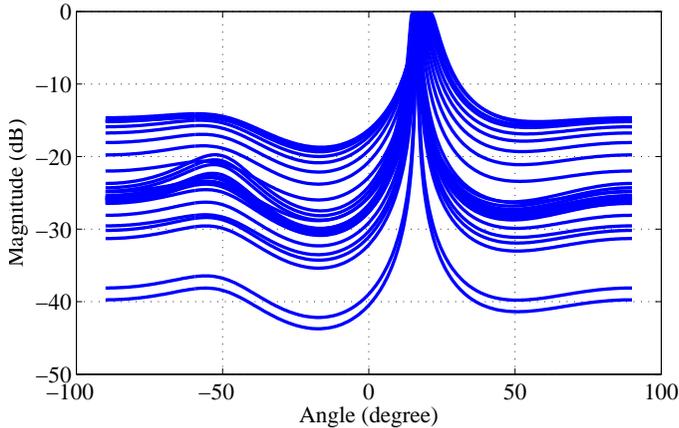


Figure 1.8 The estimated AOAs from the 30 subcarriers using the MUSIC algorithm, while the real AOA is 14° .

1.4 The BiLoc System

1.4.1 BiLoc System Architecture

The overall architecture of BiLoc is illustrated in Fig. 1.9. The BiLoc design uses only one access point and one mobile device, each equipped with an Intel 5300 NIC, servicing as receiver and transmitter, respectively. All the communications are on the 5GHz band. The Intel 5300 NIC has three antennas; at each antenna, we can read CSI data from 30 subcarriers. Thus we can collect 90 CSI data for every received packet. We then calibrate the phase information of the received CSI data using our multi-radio hardware design (see Fig. 1.6). Both the estimated AOAs and average amplitudes of two adjacent antennas are used as location feature for building the fingerprint database.

A unique feature of BiLoc is its bi-modal design. With the three receiving antennas, we can obtain two groups of data: (i) 30 estimated AOAs and 30 average amplitudes from antennas 1 and 2, and (ii) that from antennas 2 and 3. BiLoc utilizes estimated AOAs and average amplitudes for indoor fingerprinting for two main reasons. First, these two types of CSI data are highly stable for any given position. Second, they are usually complementary to each other under some indoor circumstances. For example, when a signal is blocked, the average amplitude of the signal will be significantly weakened; but the estimated AOA becomes more effective. On the other hand, when the NLOS components are stronger than the LOS component, the average amplitude will help to improve the localization accuracy.

Another unique characteristic of BiLoc is the use of deep learning to produce feature-based fingerprints from the bi-modal data in the offline training stage, which is quite different from the traditional approach of storing the measured raw data as fingerprints. Specifically, we use the weights in the deep network to represent the features-based fingerprints for every position. By obtaining the optimal weights

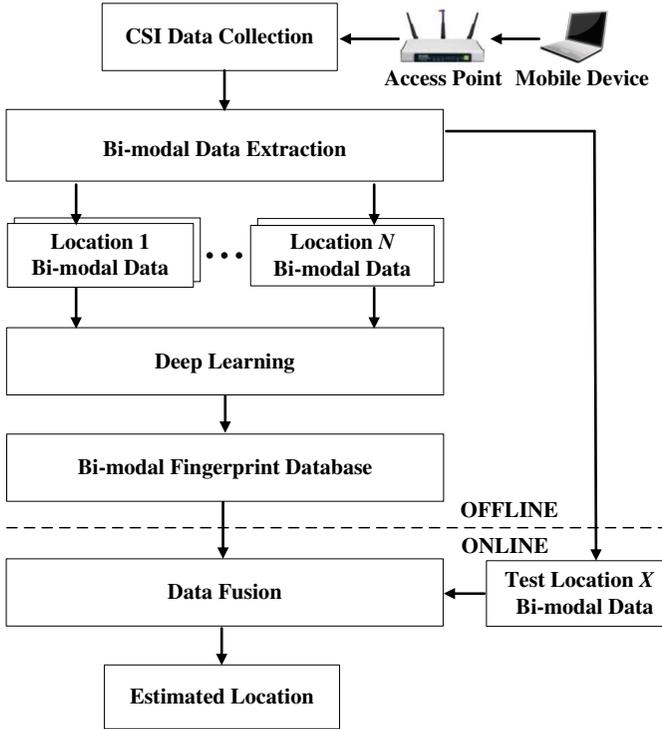


Figure 1.9 The BiLoc system architecture.

with the bi-modal data on estimated AOA and average amplitudes, we can establish a bi-modal fingerprint database for the training positions. The third feature of BiLoc is the probabilistic data fusion approach for location estimation based on received bi-modal data in the online test stage.

1.4.2 Offline Training for Bi-Modal Fingerprint Database

In the offline stage, BiLoc leverages deep learning to train and store the weights to build a bi-modal fingerprint database, which is a deep autoencoder that involves three phases: pretraining, unrolling, and fine-tuning [50]. In the pretraining phase, a deep network with three hidden layers and one input layer is used to learn the bi-modal data. We denote h^i as the hidden variable with K_i nodes at layer i , $i = 1, 2, 3$, and h^0 as the input data with K_0 nodes at the input layer. Let the average amplitude data be v^1 and the estimated AOA data be v^2 . To build the bi-modal fingerprint database, we set $h^0 = v^1$ and $h^0 = v^2$ for database 1 and 2, respectively, each of which is a set of optimal weights. We denote W_1 , W_2 and W_3 as the weights between input data and the first hidden layer, the first and second hidden layer, and the second and third hidden layer, respectively.

We define $\Pr(h^0, h^1, h^2, h^3)$ as the probabilistic generative model for the deep network. To derive the optimal weights, we maximize the marginal distribution of the input data for the deep network, which is given by

$$\max_{\{W_1, W_2, W_3\}} \sum_{h^1} \sum_{h^2} \sum_{h^3} \Pr(h^0, h^1, h^2, h^3). \quad (1.9)$$

Because of the large number of nodes and the complex model structure, it is difficult to find the optimal weights for the input data with the maximum likelihood method. To reduce the computational complexity, BiLoc utilizes a greedy learning algorithm to train the weights *layer by layer* based on a stack of RBMs [51]. We consider an RBM as a bipartite undirected graphical model [51] with joint distribution $\Pr(h^{i-1}, h^i)$, as

$$\Pr(h^{i-1}, h^i) = \frac{\exp(-\mathbb{E}(h^{i-1}, h^i))}{\sum_{h^{i-1}} \sum_{h^i} \exp(-\mathbb{E}(h^{i-1}, h^i))}, \quad (1.10)$$

where $\mathbb{E}(h^{i-1}, h^i)$ denotes the free energy between layer $(i-1)$ and layer i , which is given by

$$\mathbb{E}(h^{i-1}, h^i) = -b^{i-1}h^{i-1} - b^i h^i - h^{i-1}W_i h^i, \quad (1.11)$$

where b^{i-1} and b^i are the biases for the units of layer $(i-1)$ and that of layer i , respectively. To obtain the joint distribution $\Pr(h^{i-1}, h^i)$, the *CD-1 algorithm* is used to approximate it as [51]

$$\begin{cases} \Pr(h^{i-1}|h^i) = \prod_{j=1}^{K_{i-1}} \Pr(h_j^{i-1}|h^i) \\ \Pr(h^i|h^{i-1}) = \prod_{j=1}^{K_i} \Pr(h_j^i|h^{i-1}), \end{cases} \quad (1.12)$$

where $\Pr(h_j^{i-1}|h^i)$, and $\Pr(h_j^i|h^{i-1})$ are given by the sigmoid belief network as follows.

$$\begin{cases} \Pr(h_j^{i-1}|h^i) = \left(1 + \exp(-b_j^{i-1} - \sum_{t=1}^{K_i} W_i^{j,t} h_t^i)\right)^{-1} \\ \Pr(h_j^i|h^{i-1}) = \left(1 + \exp(-b_j^i - \sum_{t=1}^{K_{i-1}} W_i^{j,t} h_t^{i-1})\right)^{-1}. \end{cases} \quad (1.13)$$

We propose a greedy algorithm to train the weights and biases for a stack of RBMs. First, with the CD-1 method, we use the input data to train the parameters $\{b^0, b^1, W_1\}$ of the first layer RBM. Then, the parameters $\{b^0, W_1\}$ are frozen and we sample from the conditional probability $\Pr(h^1|h^0)$ to train the parameters $\{b^1, b^2, W_2\}$ of the second layer RBM. Next, we freeze the parameters $\{b^0, b^1, W_1, W_2\}$ of the first and second layers and then sample from the conditional probability $\Pr(h^2|h^1)$ to train the parameters $\{b^2, b^3, W_3\}$ of the third layer RBM. In order to train the weights and biases of each RBM, we use the CD-1 method to approximate them. For the layer i RBM model, we estimate \hat{h}^{i-1} by sampling from the conditional probability $\Pr(h^{i-1}|h^i)$; by sampling from the conditional probability $\Pr(h^i|\hat{h}^{i-1})$, we can estimate \hat{h}^i . Thus, the parameters are updated as follows.

$$\begin{cases} \Delta W_i = \varepsilon(h^{i-1}h^i - \hat{h}^{i-1}\hat{h}^i) \\ \Delta b^i = \varepsilon(h^i - \hat{h}^i) \\ \Delta b^{i-1} = \varepsilon(h^{i-1} - \hat{h}^{i-1}), \end{cases} \quad (1.14)$$

where ε is the step size.

After the pretraining phase, we obtain the near-optimal weights for the deep network. We then unroll the deep network with *forward propagation* to obtain the reconstructed input data in the unrolling phase. Finally, in the fine-tuning phase, the *backpropagation* algorithm is used to train the weights in the deep network according to the error between the input data and the reconstructed input data. The optimal weights are obtained by minimizing the error. In BiLoc, we use estimated AOA and average amplitudes as input data, and obtain two sets of optimal weights for the bi-modal fingerprint database.

1.4.3 Online Data Fusion for Position Estimation

In the online phase, we adopt a probabilistic approach to location estimation based on the bi-modal fingerprint database and the bi-modal test data. We derive the posteriori probability $\Pr(l_i|v^1, v^2)$ using Bayes' law as

$$\Pr(l_i|v^1, v^2) = \frac{\Pr(l_i) \Pr(v^1, v^2|l_i)}{\sum_{i=1}^N \Pr(l_i) \Pr(v^1, v^2|l_i)}, \quad (1.15)$$

where N is the number of reference locations, l_i is the i th reference location in the bi-modal fingerprint database, and $\Pr(l_i)$ is the prior probability that the mobile device is considered to be at the reference location l_i . Without loss of generality, we assume that $\Pr(l_i)$ is uniformly distributed. The posteriori probability $\Pr(l_i|v^1, v^2)$ becomes

$$\Pr(l_i|v^1, v^2) = \frac{\Pr(v^1, v^2|l_i)}{\sum_{i=1}^N \Pr(v^1, v^2|l_i)}. \quad (1.16)$$

In BiLoc, we approximate $\Pr(v^1, v^2|l_i)$ with an RBF in the similar form of a Gaussian function, to measure the degree of similarity between the reconstructed bi-modal data and the test bi-modal data, given by

$$\Pr(v^1, v^2|l_i) = \exp\left(-\frac{(1-\rho)\|v^1 - \hat{v}^1\|}{\eta_1\sigma_1} - \rho\frac{\|v^2 - \hat{v}^2\|}{\eta_2\sigma_2}\right), \quad (1.17)$$

where \hat{v}^1 and \hat{v}^2 are the reconstructed average amplitude and reconstructed AOA, respectively; σ_1 and σ_2 are the variance of the average amplitude and estimated AOA, respectively; η_1 and η_2 are the parameters of the variance of the average amplitude and estimated AOA, respectively; and ρ is the ratio for the bi-modal data.

For the Eq. (1.17), the average amplitudes \hat{v}^1 and the estimated AOAs \hat{v}^2 are as the input of deep network, where the different nodes of the input can express the different CSI channels. Then, by employing the test data \hat{v}^1 and \hat{v}^2 , we compute the reconstructed average amplitude \hat{v}^1 and reconstructed AOA \hat{v}^2 based on database 1 and database 2, respectively, which is used to compute the likelihood function $\Pr(v^1, v^2|l_i)$.

The location of the mobile device can be finally estimated as a weighted average of all the reference locations, which is given by

$$\hat{l} = \sum_{i=1}^N \Pr(l_i|v^1, v^2) \cdot l_i. \quad (1.18)$$

1.5 Experimental Study

1.5.1 Test Configuration

We present our experimental study with BiLoc in the 5GHz band in this section. In the experiments, we use a desktop computer as an access point and a Dell laptop as a mobile device, both equipped with an Intel 5300 NIC. In fact, we use the desktop computer instead of the commodity routers that are not equipped with the Intel 5300 NIC nowadays. Our implementation of BiLoc is executed on the Ubuntu desktop 14.04 LTS OS for both the access point and mobile device. We use QPSK modulation and a $1/2$ coding rate for the OFDM system. For the access point, it is set in monitor model and the distance between two adjacent antennas is $d = 2.68$ cm, which is half of a wavelength for the 5GHz band. For the mobile device, it transmits packets at 100 packets per second using only one antenna in injection mode. 5GHz CSI data can be obtained by using packet injection technique based on LORCON version 1. Then, we extract bi-modal data for training and test stages as described in Section 1.4.2.

We also implement three representative schemes from the literature, i.e., Horus [11], FIFS [15], and DeepFi [16]. For a fair comparison, all the schemes use the same measured dataset captured in the 5GHz band to estimate the location of the mobile device. We conduct extensive experiments with the schemes in the following two representative indoor environments.

Computer Laboratory: This is a 6×9 m² computer laboratory, a cluttered environment with metal tables, chairs, and desktop computers, blocking most of the LOS paths. The floor plan is shown in Fig. 1.10, with 15 chosen training positions (marked as red squares) and 15 chosen test positions (marked as green dots). The distance between two adjacent training positions is 1.8 m. The single access point is put close to the center of the room. We collect bi-modal data from 1000 packet receptions for each training position, and from 25 packet receptions for each test position. The deep network used for this scenario is configured as $\{K_1 = 150, K_2 = 100, K_3 = 50\}$. Also, the ratio ρ for the bi-modal data is set as 0.5.

Corridor: This is a 2.4×24 m² corridor, as shown in Fig. 1.11. In this scenario, the AP is placed at one end of the corridor and there are plenty of LOS paths. Ten training positions (red squares) and 10 test positions (green dots) are arranged along a straight line. The distance between two adjacent training positions is also 1.8 m. We also collect bi-modal data from 1000 packets for each training position and from 25 packets for each test position. The deep network used for this scenario is configured as $\{K_1 = 150, K_2 = 100, K_3 = 50\}$. Also, the ratio ρ for the bi-modal data is set as 0.1.

1.5.2 Accuracy of Location Estimation

Tables 1.1 and 1.2 present the mean and STD of localization errors, and the execution time of the four schemes for the two scenarios, respectively. In the laboratory environment, BiLoc achieves a mean error of 1.5743 m and an STD error of 0.8312 m across the 15 test points. In the corridor experiment, because only one access point is used for this larger space, BiLoc achieves a mean error of 2.1501 m and an STD

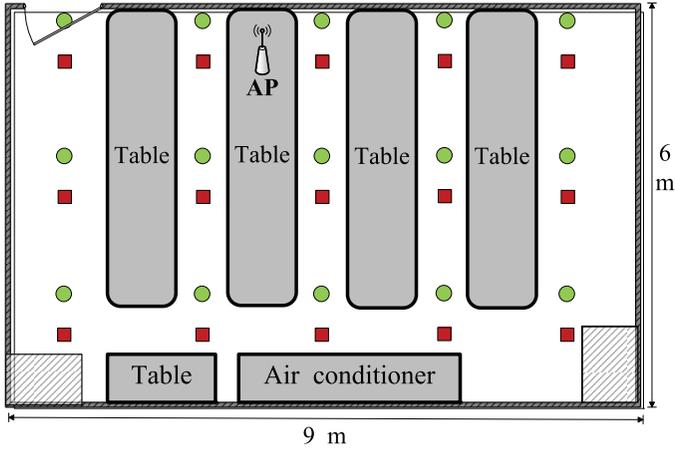


Figure 1.10 Layout of the computer laboratory: training positions are marked as red squares and testing positions are marked as green dots.

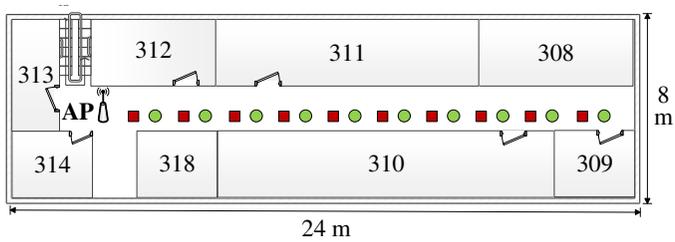


Figure 1.11 Layout of the corridor: training positions are marked as red squares and testing positions are marked as green dots.

Table 1.1 Mean/STD error and execution time of the laboratory experiment

Algorithm	Mean error (m)	Std. dev. (m)	Mean execution time (s)
BiLoc	1.5743	0.8312	0.6653
DeepFi	2.0411	1.3804	0.3340
FIFS	2.7151	1.0805	0.2918
Horus	3.0537	1.0623	0.2849

error of 1.5420 m across the 10 test points. BiLoc outperforms the other three benchmark schemes with the smallest mean error, as well as with the smallest STD error, i.e., being the most stable scheme in both scenarios. We also compare the online test time of all the schemes. Due to the use of bi-modal data and the deep network, the mean executing time of BiLoc is the highest among the four schemes. However, the mean execution time is 0.6653 s for the laboratory case and 0.5440 s for the corridor case, which are sufficient for most indoor localization applications.

Table 1.2 Mean/STD errors and execution time of the corridor experiment

Algorithm	Mean error (m)	Std. dev. (m)	Mean execution time (s)
BiLoc	2.1501	1.5420	0.5440
DeepFi	2.8953	2.5665	0.3707
FIFS	4.4296	3.4256	0.2535
Horus	4.8000	3.5242	0.2505

Fig. 1.12 presents the CDF of distance errors of the four schemes in the laboratory environment. In this complex propagation environment, BiLoc has 100% of the test positions with an error under 2.8 m, while DeepFi, FIFS, and Horus have about 72%, 52%, and 45% of the test positions with an error under 2.8 m, respectively. For a much smaller error of 1.5 m, the percentage of test positions having a smaller error are 60%, 45%, 15%, and 5% for BiLoc, DeepFi, FIFS, and Horus, respectively. BiLoc achieves the highest precision among the four schemes, due to the use of bi-modal CSI data (i.e., average amplitudes and estimated AOA). In fact, when the amplitude of a signal is strongly influenced in the laboratory environment, the estimated AOA can be utilized to mitigate this effect by BiLoc. However, the other schemes based solely on CSI or RSS amplitudes will be affected.

Fig. 1.13 presents the CDF of distance errors of the four schemes for the corridor scenario. Only one access point is used at one end for this 24 m long corridor, making it hard to estimate the location of the mobile device. For BiLoc, more than 90% of the test positions have an error under 4 m, while DeepFi, FIFS, and Horus have about 70%, 60%, and 50% of the test positions with an error under 4 m, respectively. For a tighter 2 m error threshold, BiLoc has 60% of the test positions with an error below this threshold, while it is 40% for the other three schemes. For the corridor scenario, BiLoc mainly utilizes the average amplitudes of CSI data, because the estimated AOAs are similar for all the training/test positions (recall that they are aligned along a straight line with the access point at one end). This is a challenging scenario for differentiating different test points and the BiLoc mean error is 0.5758 m higher than that of the laboratory scenario.

1.5.3 2.4GHz versus 5GHz

We also compare the 2.4GHz channel and 5GHz channel with the BiLoc scheme. For a fair comparison, we conduct the experiments at night, because the 2.4GHz band is much more crowded than the 5GHz band during the day.

Fig. 1.14 presents the CDF of localization errors in the 2.4GHz and 5GHz band in the laboratory environment, where both average amplitudes and estimated AOAs are effectively used by BiLoc for indoor localization. We can see that for BiLoc, about 70% of the test positions have an error under 2 m in 5GHz, while 50% of the test positions have an error under 2 m in 2.4GHz. In addition, the maximum errors in 2.4GHz and 5GHz are 6.4 m and 2.8 m, respectively. Therefore, the proposed BiLoc scheme achieves much better performance in 5GHz than 2.4GHz. In fact, the phase

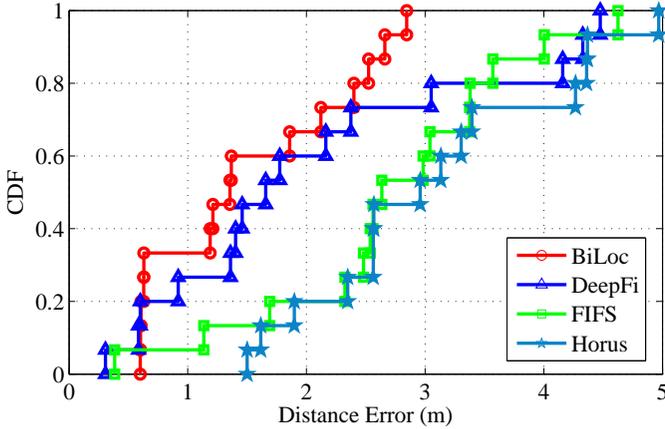


Figure 1.12 CDF of localization errors in 5GHz for the laboratory experiment.

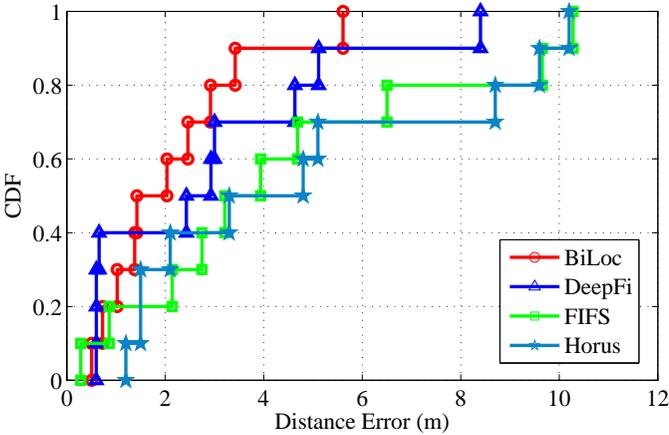


Figure 1.13 CDF of localization errors in 5GHz for the corridor experiment.

difference between two antennas in 2.4GHz exhibits great variations, which lead to lower localization accuracy. This experiment also validates our Hypothesis 2.

1.5.4 Impact of Parameter ρ

Recall that the parameter ρ is used to trade off the impacts of average amplitudes and estimated AOA in location estimation as in (1.17). We examine the impact of ρ on localization accuracy under the two environments. With BiLoc, we use bimodal data for online testing, and ρ directly influences the likelihood probability $\Pr(v^1, v^2 | l_i)$ (1.17), which in turn influences the localization accuracy.

Fig. 1.15 presents the mean localization errors for increasing ρ for the laboratory and corridor experiments. In the laboratory experiment, when ρ is increased from

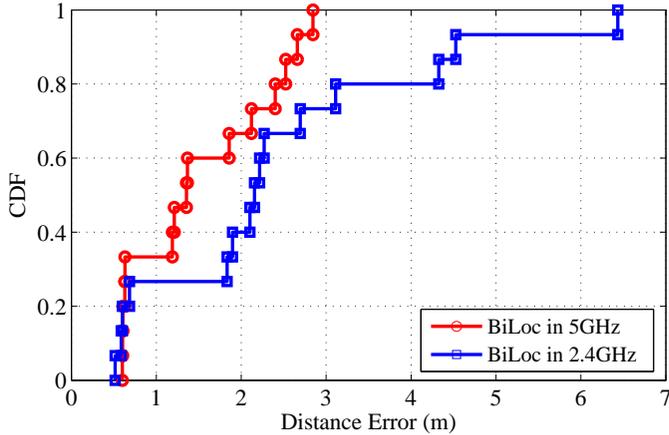


Figure 1.14 CDF of localization errors in 5GHz and 2.4GHz for the laboratory experiment.

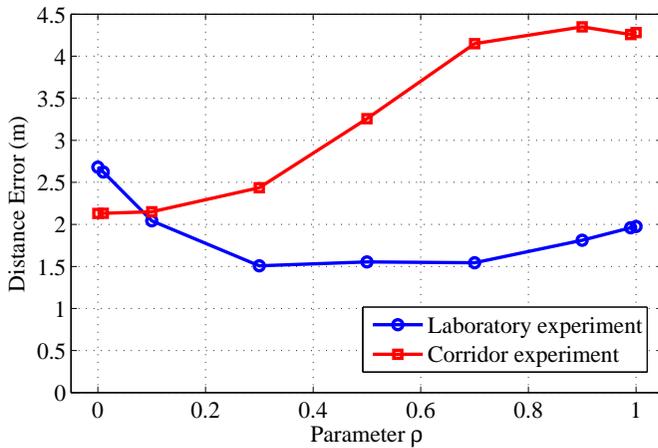


Figure 1.15 Mean localization errors versus parameter ρ for the laboratory and corridor experiments.

0 to 0.3, the mean error decreases from 2.6 m to 1.5 m. Furthermore, the mean error remains around 1.5 m for $\rho \in [0.3, 0.7]$, and then increases from 1.5m to 2 m when ρ is increased from 0.6 to 1. Therefore, BiLoc achieves its minimum mean error for $\rho \in [0.3, 0.7]$, indicating that both average amplitudes and estimated AOA are useful for accurate location estimation. Moreover, BiLoc has higher localization accuracy with the mean error of 1.5m, compared with individual modality such as the estimated AOA with that of 2.6m or the average amplitudes with that of 2.0m.

In the corridor experiment, we can see that the mean error remains around 2.1 m when ρ is increased from 0 to 0.1. When ρ is further increased from 0.1 to 1, the

mean error keeps on increasing from 2.1 m to about 4.3 m. Clearly, in the corridor experiment, the estimated AOA provide similar characteristics for deep learning, and are not useful for distinguishing the positions. Therefore BiLoc should mainly use the average amplitudes of CSI data for better accuracy. These experiments provide some useful guidelines on setting the ρ value for different indoor environments.

1.6 Future Directions and Challenges

1.6.1 *New Deep Learning Methods for Indoor Localization*

This chapter has discussed three deep learning technologies including autoencoder, CNN, LSTM for fingerprinting based indoor localization. With the rapid growth in the artificial intelligence (AI) field, new deep learning approaches are proposed for mainly handling computer vision problems, such as robust object recognition and detection, data generation, as well as the Go game. For example, generative adversarial network (GAN) can be used for generating new data samples; deep reinforcement learning has been leveraged for AlphaGo; deep Gaussian process can be utilized for improving the robustness of object detection. In fact, these new deep learning methods can be also used for solving basic indoor localization problems such as radio map constructions, environment change, and devices calibration. For example, deep reinforcement learning [52] can be used for improving localization performance and reduce cost. Moreover, Bayesian deep learning such as deep Gaussian process [53, 54] has high robustness for environment noise, which can be exploited for radio map construction, and environment change and devices calibration. Moreover, generative adversarial network (GAN) can be incorporated for building radio map and increasing the number of training data samples. In addition, compressed deep learning [55] by using pruning and quantization can be considered for resource limited mobile devices. Thus, we can implement deep learning models on smartphones rather than servers for indoor localization.

1.6.2 *Sensor Fusion for Indoor Localization Using Deep Learning*

In this chapter, we have proposed bi-modal CSI data for indoor localization. In fact, multiple sensor data sources can be fused for improving indoor localization performance. Traditionally, sequence models such as Kalman filter, particle filter, hidden markov model (HMM), and conditional random field (CRF) can fuse WiFi and inertial sensor data on smartphones for indoor localization, which requires for obtaining the sequence data from continuing smartphone movement. Deep learning techniques can improve the performance of indoor localization using multimodal sequence data. For example, LSTM method can be leveraged for indoor localization using sequence RSS or CSI data, which also fuse multimodal data for improving the localization accuracy. Considering WiFi and magnetic sensor data from smartphone, we can integrate them into a large data matrix as input to LSTM for indoor localization. In fact, WiFi and magnetic sensor data are complementary to each other. For example, because of lower resolution of WiFi signals, only using WiFi RSS values cannot obtain better performance at close locations, while magnetic sensor data at

such positions is greatly different. LSTM can effectively fuse them for indoor localization [43]. In addition, an integrated CNN and LSTM model can be used for WiFi RSS or CSI images data, which can be easily created from different access points or different subcarriers. In fact, the LSTM model can be combined as other deep learning models such as autoencoder, GAN, deep reinforcement learning, Bayesian model for different localization problems such as radio map construction, device calibration, and environment change. For sensor data fusion for indoor localization, different sensor data sources should be normalized and aligned [23].

1.6.3 Secure Indoor Localization Using Deep learning

For wireless fingerprinting based indoor localization, security becomes increasingly important, where wireless signals are susceptible to eavesdropping, distributed denial-of-service (DDoS) attacks, and bad data injection [56]. Specially, for crowd-sourcing based indoor localization, fingerprints are from different devices at different times, which greatly exposes the security problem. For attacker models, there are three general scenarios for RSS fingerprinting based localization [57]. First, the attacker does not know the true RSS fingerprints and injects fake RSS data at random. Second, the attacker knows legitimate RSS fingerprints and add noise to them. Third, the attacker can change the mapping between RSS fingerprints and positions. For defense models, they can consider the temporal correlation and spatial correlation within RSS Traces against different attackers. In fact, deep learning can study the feature of Localization signals to address the above security problems. Deep learning can consider different data features from multiple paths of wireless signals to classify eavesdropping, DDoS attack or bad data injection for fingerprinting based indoor localization.

On the other hand, deep learning security problems become important, which mainly focuses on how to recognize adversarial data and clear RSS data. Deep Learning will have bad performance with adversarial data, which is only obtained by adding small noise into clear RSS data. Thus, adversarial data should be recognized before implementing indoor localization systems based on deep learning, thus guaranteeing good localization performance. In addition, privacy persevering deep learning can be used for indoor localization problems, which can protect user location privacy information.

1.7 Conclusions

In this chapter, we proposed a bi-modal deep learning system for fingerprinting-based indoor localization with 5GHz commodity WiFi NICs. First, the state-of-the-art deep learning techniques including deep autoencoder network, CNN and LSTM were introduced. We then extracted and calibrated CSI data to obtain bi-modal CSI data, including average amplitudes and estimated AOAs, which were used in both the offline and online stages. The proposed scheme was validated with extensive experiments. We concluded this chapter with a discussion of future directions and challenges for indoor localization problems using deep learning.

Acknowledgment

This work is supported in part by the US NSF under Grant ACI-1642133, and by the Wireless Engineering Research and Education Center (WEREC) at Auburn University.

References

- [1] Wang Y, Liu J, Chen Y, et al. E-eyes: Device-free Location-oriented Activity Identification Using Fine-grained WiFi Signatures. In: Proc. ACM MobiCom'14. Maui, HI; 2014. p. 617–628.
- [2] Zhang D, Zhao S, Yang LT, et al. NextMe: Localization using cellular traces in internet of things. *IEEE Trans Ind Informat.* 2015 April;11(2):302–312.
- [3] Derr K, Manic M. Wireless sensor networks node localization for various industry problems. *IEEE Trans Ind Informat.* 2015 Jun;11(3):752–762.
- [4] Abu-Mahfouz A, Hancke GP. Distance bounding: a practical security solution for real-time location systems. *IEEE Trans Ind Informat.* 2013 Feb;9(1):16–27.
- [5] Pak J, Ahn C, Shmaliy Y, et al. Improving Reliability of Particle Filter-Based Localization in Wireless Sensor Networks via Hybrid Particle/FIR Filtering. *IEEE Trans Ind Informat.* 2015 Oct;11(5):1089–1098.
- [6] Ivanov S, Nett E. Localization-Based Radio Model Calibration for Fault-Tolerant Wireless Mesh Networks. *IEEE Trans Ind Informat.* 2013 Feb;9(1):246–253.
- [7] Lee S, Kim B, Kim H, et al. Inertial sensor-based indoor pedestrian localization with minimum 802.15.4a configuration. *IEEE Trans Ind Informat.* 2011 Aug;7(3):455–466.
- [8] Wu B, Jen C. Particle filter based radio localization for mobile robots in the environments with low-density WLAN APs. *IEEE Trans Ind Electron.* 2014 Sep;61(12):6860–6870.
- [9] Liu H, Darabi H, Banerjee P, et al. Survey of Wireless Indoor Positioning Techniques and Systems. *IEEE Trans Syst, Man, Cybern C.* 2007 Nov;37(6):1067–1080.
- [10] Bahl P, Padmanabhan VN. Radar: An in-building RF-based User Location and Tracking System. In: Proc. IEEE INFOCOM'00. Tel Aviv, Israel; 2000. p. 775–784.
- [11] Youssef M, Agrawala A. The Horus WLAN Location Determination System. In: Proc. ACM MobiSys'05. Seattle, WA; 2005. p. 205–218.
- [12] Halperin D, Hu WJ, Sheth A, et al. Predictable 802.11 Packet Delivery from Wireless Channel Measurements. In: Proc. ACM SIGCOMM'10. New Delhi, India; 2010. p. 159–170.
- [13] Sen S, J Lee KK, Congdon P. Avoiding Multipath to Revive Inbuilding WiFi Localization. In: Proc. ACM MobiSys'13. Taipei, Taiwan; 2013. p. 249–262.
- [14] Xie Y, Li Z, Li M. Precise power delay profiling with commodity wifi. In: Proc. ACM Mobicom'15. Paris, France; 2015. p. 53–64.

- [15] Xiao J, Wu K, Yi Y, et al. FIFS: Fine-Grained Indoor Fingerprinting System. In: Proc. IEEE ICCCN'12. Munich, Germany; 2012. p. 1–7.
- [16] Wang X, Gao L, Mao S, et al. DeepFi: Deep Learning for Indoor Fingerprinting Using Channel State Information. In: Proc. WCNC'15. New Orleans, LA; 2015. p. 1666–1671.
- [17] Wu C, Yang Z, Zhou Z, et al. PhaseU: Real-time LOS Identification with WiFi. In: Proc. IEEE INFOCOM'15. Hong Kong, China; 2015. p. 2038–2046.
- [18] Qian K, Wu C, Yang Z, et al. PADS: Passive Detection of Moving Targets with Dynamic Speed using PHY Layer Information. In: Proc. IEEE ICPADS'14. Hsinchu, Taiwan; 2014. p. 1–8.
- [19] Gjengset J, Xiong J, McPhillips G, et al. Phaser: Enabling phased array signal processing on commodity WiFi access points. In: Proc. ACM Mobicom'14. Maui, HI; 2014. p. 153–164.
- [20] Wang X, Gao L, Mao S. BiLoc: Bi-Modal Deep Learning for Indoor Localization With Commodity 5GHz WiFi. *IEEE Access*. 2017;5:4209–4220.
- [21] Abadi M, Barham P, Chen J, et al. Tensorflow: a system for large-scale machine learning. In: OSDI. vol. 16; 2016. p. 265–283.
- [22] Mohammadi M, Al-Fuqaha A, Sorour S, et al. Deep Learning for IoT Big Data and Streaming Analytics: A Survey. *IEEE Communications Surveys & Tutorials*. 2018;.
- [23] Wang X, Wang X, Mao S. RF Sensing in the Internet of Things: A General Deep Learning Framework. *IEEE Communications Magazine*. 2018;56(9):62–67.
- [24] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *science*. 2006;313(5786):504–507.
- [25] Wang X, Gao L, Mao S, et al. CSI-based fingerprinting for indoor localization: A deep learning approach. *IEEE Transactions on Vehicular Technology*. 2017;66(1):763–776.
- [26] Wang X, Gao L, Mao S. PhaseFi: Phase fingerprinting for indoor localization with a deep learning approach. In: Proc. GLOBECOM'15. San Diego, CA; 2015. .
- [27] Wang X, Gao L, Mao S. CSI phase fingerprinting for indoor localization with a deep learning approach. *IEEE Internet of Things Journal*. 2016;3(6):1113–1123.
- [28] Chen X, Ma C, Allegue M, et al. Taming the inconsistency of Wi-Fi fingerprints for device-free passive indoor localization. In: INFOCOM 2017-IEEE Conference on Computer Communications, IEEE. IEEE; 2017. p. 1–9.
- [29] Wang J, Zhang X, Gao Q, et al. Device-free wireless localization and activity recognition: A deep learning approach. *IEEE Transactions on Vehicular Technology*. 2017;66(7):6258–6267.
- [30] Xiao C, Yang D, Chen Z, et al. 3-d ble indoor localization based on denoising autoencoder. *IEEE Access*. 2017;5:12751–12760.

- [31] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems; 2012. p. 1097–1105.
- [32] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86(11):2278–2324.
- [33] Wang X, Wang X, Mao S. CiFi: Deep convolutional neural networks for indoor localization with 5GHz Wi-Fi. In: Proc. IEEE ICC 2017. Paris, France; 2017. p. 1–6.
- [34] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–778.
- [35] Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10); 2010. p. 807–814.
- [36] Wang W, Wang X, Mao S. Deep convolutional neural networks for indoor localization with CSI images. IEEE Transactions on Network Science and Engineering;5.
- [37] Wang X, Wang X, Mao S. ResLoc: Deep residual sharing learning for indoor localization with CSI tensors. In: Proc. IEEE PIMRC 2017. Montreal, Canada; 2017. .
- [38] Chen H, Zhang Y, Li W, et al. ConFi: Convolutional neural networks based indoor wi-fi localization using channel state information. IEEE Access Journal. 2017;5:18066–18074.
- [39] Mittal A, Tiku S, Pasricha S. Adapting Convolutional Neural Networks for Indoor Localization with Smart Mobile Devices. In: Proceedings of the 2018 on Great Lakes Symposium on VLSI. ACM; 2018. p. 117–122.
- [40] ZHANG T, Yi M. The Enhancement of WiFi Fingerprint Positioning Using Convolutional Neural Network. DEStech Transactions on Computer Science and Engineering. 2018;(CCNT).
- [41] Niitsoo A, Edelhäuser T, Mutschler C. Convolutional neural networks for position estimation in tdoa-based locating systems. In: Proc. 9th Intl. Conf. Indoor Positioning and Indoor Navigation,(Nantes, France); 2018. p. 1–8.
- [42] Gers FA, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM. Neural Computation. 2000 Oct;12(10).
- [43] Wang X, Yu Z, Mao S. DeepML: Deep LSTM for indoor localization with smartphone magnetic and light sensors. In: Proc. IEEE ICC 2017. Kansas City, MO; 2018. .
- [44] Yang W, Wang X, Cao S, et al. Multi-class wheat moisture detection with 5GHz Wi-Fi: A deep LSTM approach. In: Proc. ICCCN 2018. Hangzhou, China; 2018. .
- [45] Wang Y, Shen Y, Mao S, et al. LASSO & LSTM integrated temporal model for short-term solar intensity forecasting. IEEE Internet of Things Journal. 2018;Revised.

- [46] Akbar MB, Taylor DG, Durgin GD. Amplitude and phase difference estimation bounds for multisensor based tracking of RFID Tags. In: Proc. IEEE RFID'15. San Diego, CA; 2015. p. 105–112.
- [47] Kleisouris K, Chen Y, Yang J, et al. The Impact of Using Multiple Antennas on Wireless Localization. In: Proc. IEEE SECON'08. San Francisco, CA; 2008. p. 55–63.
- [48] I Speth M, Fechtel S, Fock G, et al. Optimum receiver design for wireless broad-band systems using OFDM–Part I. *IEEE Trans Commun.* 1999 Nov;47(11):1668–1677.
- [49] Schmidt R. Multiple emitter location and signal parameter estimation. *IEEE Trans Antennas Propag.* 1986 Mar;34(3):276–280.
- [50] Hinton GE, Salakhutdinov R. Reducing the dimensionality of data with neural networks. *Science.* 2006 July;313(5786):504–507.
- [51] Bengio Y, Lamblin P, Popovici D, et al. Greedy layer-wise training of deep networks. In: Proc. Adv. Neural Inform. Proc. Syst. 19. Vancouver, Canada; 2007. p. 153–160.
- [52] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. *Nature.* 2015;518(7540):529.
- [53] Shi J, Chen J, Zhu J, et al. ZhuSuan: A library for Bayesian deep learning. arXiv preprint arXiv:170905870. 2017;.
- [54] Wang X, Wang X, Mao S, et al. DeepMap: Deep Gaussian Process for indoor radio map construction and location estimation. In: Proc. IEEE GLOBECOM 2018. Abu Dhabi, United Arab Emirates; 2018. .
- [55] Han S, Liu X, Mao H, et al. EIE: Efficient Inference Engine on Compressed Deep Neural Network. *International Conference on Computer Architecture (ISCA).* 2016;.
- [56] Shokri R, Theodorakopoulos G, Troncoso C, et al. Protecting location privacy: optimal strategy against localization attacks. In: *Proceedings of the 2012 ACM conference on Computer and communications security.* ACM; 2012. p. 617–627.
- [57] Li T, Chen Y, Zhang R, et al. Secure crowd-sourced indoor positioning systems. In: *IEEE INFOCOM18;* 2018. .