IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 1, 2017, accepted March 20, 2017, date of publication March 28, 2017, date of current version April 24, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2688362

BiLoc: Bi-Modal Deep Learning for Indoor Localization With Commodity 5GHz WiFi

XUYU WANG[1], (Student Member, IEEE), LINGJUN GAO[2], (Student Member, IEEE), AND SHIWEN MAO[1], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201, USA
[2]DataYes, Inc., Shanghai 200120, China

Corresponding author: S. Mao (smao@ieee.org)

This work is supported in part by the U.S. National Science Foundation under Grant CNS-1320664 and Grant CNS-1702957, and in part by the Wireless Engineering Research and Education Center, Auburn University.

ABSTRACT In this paper, we study fingerprinting-based indoor localization in commodity 5-GHz WiFi networks. We first theoretically and experimentally validate three hypotheses on the channel state information (CSI) data of 5-GHz OFDM channels. We then propose a system termed BiLoc, which uses bi-modality deep learning for localization in the indoor environment using off-the-shelf WiFi devices. We develop a deep learning-based algorithm to exploit bi-modal data, i.e., estimated angle of arrivings and average amplitudes (which are calibrated CSI data using several proposed techniques), for both the off-line and online stages of indoor fingerprinting. The proposed BiLoc system is implemented using commodity WiFi devices. Its superior performance is validated with extensive experiments under three typical indoor environments and through comparison with three benchmark schemes.

INDEX TERMS Indoor localization, fingerprinting, deep learning, 5GHz commodity WiFi, channel state information, bi-modality fingerprinting.

I. INTRODUCTION

The wide use of mobile devices has fostered great interest in indoor location-based services, such as indoor navigation [1]–[5], robot tracking in factories [6], locating workers on construction sites [7], and activity recognition [8], all requiring accurate identification of locations of mobile devices indoors. The indoor environment results in a complex radio propagation channel, including multipath propagation, blockage, and shadow fading, and has stimulated considerable research efforts involving 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. The position of a mobile user can then be estimated by matching the newly received data with the information stored 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) for indoor localization, because of its low requirements for hardware [10], [11]. The Radar system is the first RSS-based scheme that incorporates a deterministic approach for indoor localization [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 performance of RSS-based schemes is limited by two inherent shortcomings. 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, an RSS value only reflects the coarse channel information, which is the total power for all received signals.

Unlike RSS, *channel state information* (CSI) represents accurate channel information, which can currently be extracted from several commodity wireless network interface cards (NIC), e.g., Intel WiFi Link 5300 MIMO 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 and provides 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] leverages the weighted average of CSI amplitude data from three antennas for fingerprinting based localization. To fully exploit the diversity of multiple antennas and subcarriers, DeepFi [16] incorporates a

deep auto-encoder network for training subcarriers' data over multiple antennas to extract the CSI features for indoor localization. However, these CSI-based techniques only use CSI amplitude data. The raw phase data is often highly random and not directly usable [17].

Recently, two effective methods have been presented to remove the randomness of raw CSI phase data extracted from the Intel 5300 NIC in the 2.4GHz band. In [18], the measured phases from 30 subcarriers are processed with a linear transformation to mitigate the high randomness in the CSI phase; the cleansed phase is then employed for passive human movement detection. In [17], in addition to 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 the phase will be zero (i.e., lost) after such processing. This is caused by the firmware design of the Intel 5300 NIC when it is operating on the 2.4GHz band [19]. To address this issue, Phaser [19] is the first to exploit CSI phase in the 5GHz WiFi. Phaser constructs an angle of arriving (AOA) pseudospectrum for phase calibration in a single Intel 5300 NIC. Motivated by the above studies, we explore the use of effectively cleansed phase data for indoor localization with commodity 5GHz WiFi.

In this paper, we study the problem of fingerprinting-based indoor localization using commodity 5GHz WiFi. We first present three hypotheses on CSI amplitude and phase information for 5GHz OFDM channels. The first hypothesis states that the average amplitude of two antennas is more stable than that of a single antenna, as well as, RSS. The second hypothesis asserts that CSI phase difference values from two antennas in 5GHz are 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 the measured phase differences unusable. However, we find this phenomenon does not exit in the 5GHz band, where all the phase differences are concentrated around one value. We then design simple multi-radio hardware for phase calibration which is very different from the technique [19] that uses AOA pseudospectrum searching with high computation complexity to calibrate the phase of a single Intel 5300 NIC. As a result, the randomness from the frequency and time difference between the receiver and transmitter, and the unknown phase offset can all be removed, and stable phase information can be obtained. The third hypothesis is that 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.

Next, we design BiLoc, **Bi**-modal deep learning for indoor **loc**alization using commodity WiFi devices, to incorporate the three hypotheses while building an indoor fingerprinting system. In BiLoc, we first extract raw amplitude and phase data from three antennas, each with 30 subcarriers, utilizing a modified firmware. We then obtain bi-modal data, including average amplitudes over pairs of antennas and estimated AOAs, with the calibration procedure previously discussed. In the training phase, we adopt a deep autoencoder network to extract the unique channel features hidden in the bi-modal data, and leverage the weights of the deep autoencoder network as the extracted features (i.e., fingerprints). To reduce the computational complexity, a greedy learning scheme is leveraged to train the deep autoencoder network using a Restricted Boltzmann Machine (RBM) model. In the test phase, bi-modal test data is first collected from a mobile device. Based on the radial basis function (RBF), a Bayesian probability model is employed to estimate position.

The main contributions of this paper are summarized below.

- We theoretically and experimentally validate the feasibility of using bi-modal CSI data for indoor localization. In particular, we analyze the measured phase errors and design a multi-radio hardware for calibrating the unknown phase offset difference in a single Intel 5300 NIC. To the best of our knowledge, we are the first to employ both average amplitudes and estimated AOAs for indoor fingerprinting with commodity 5GHz WiFi networks.
- We propose a deep learning approach for indoor fingerprinting. In particular, we design a deep autoencoder network to extract OFDM channel features hidden in the rich CSI bi-modal data, and use weights to build the bi-modal fingerprint database. Furthermore, we propose a probability fusion technique for accurately estimating position with bi-modal test data.
- The BiLoc system is implemented using off-the-shelf 5GHz WiFi and is shown, through extensive experimentation to achieve superior performance in three typical indoor scenarios. Our test results demonstrate that BiLoc outperforms three representative existing schemes regarding localization accuracy.

In the remainder of this paper, the preliminaries and hypotheses are presented in Section II. We introduce the BiLoc system in Section III and validate its performance in Section IV. Related work is discussed in Section V. Section VI summarizes this paper.

II. PRELIMINARIES AND HYPOTHESES

A. CHANNEL STATE INFORMATION PRELIMINARIES

The OFDM technique is widely employed in the physical layer (PHY) of Wi-Fi and LTE systems, where the total spectrum is partitioned into multiple orthogonal subcarriers. Moreover, the subcarriers can carry wireless data based on the same modulation and coding scheme (MCS), which mitigates frequency selective fading. Leveraging the device driver for the Intel 5300 NIC, we can extract CSI, which can reveal the channel characteristics experienced by the received signal such as the multipath effect, shadow fading, and distortion.

With OFDM technology, every WiFi subcarrier becomes a narrowband flat fading channel. In the frequency domain,

the channel can be modeled by

$$\vec{R} = \text{CSI} \cdot \vec{T} + \vec{N}, \tag{1}$$

where $\vec{R}$ and $\vec{T}$ are the received and transmitted signals, respectively, and $\vec{N}$ is the noise.

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 subcarrier i CSI, CSI_i, is as follows:

$$\text{CSI}_i = \mathcal{I}_i + j\mathcal{Q}_i = |\text{CSI}_i| \exp(j\angle \text{CSI}_i), \tag{2}$$

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

B. 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 [20]. The probability distribution function (PDF) of the amplitude response is defined by

$$f(|\text{CSI}_i|) = \frac{|\text{CSI}_i|}{\sigma^2} \exp\left(-\frac{|\text{CSI}_i|^2 + |\text{CSI}_0|^2}{2\sigma^2}\right) \times I_0\left(\frac{|\text{CSI}_i| \cdot |\text{CSI}_0|}{\sigma^2}\right), \tag{3}$$

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 $\mathcal{N}(\sqrt{|\text{CSI}_0|^2 + \sigma^2}, \sigma^2)$ [20].

The phase response of subcarrier i is computed by $\angle\text{CSI}_i = \arctan(\mathcal{Q}_i/\mathcal{I}_i)$ [20]. The phase PDF is expressed as:

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

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 $\mathcal{N}\left(0, (\sigma/|\text{CSI}_0|)^2\right)$ [20]. The distribution of amplitude and phase of the subcarriers would be useful to guide the design of localization algorithms.

C. HYPOTHESES

We have the following three important hypotheses about the 5GHz CSI data, which are demonstrated and tested with our measurement study and theoretical analysis.

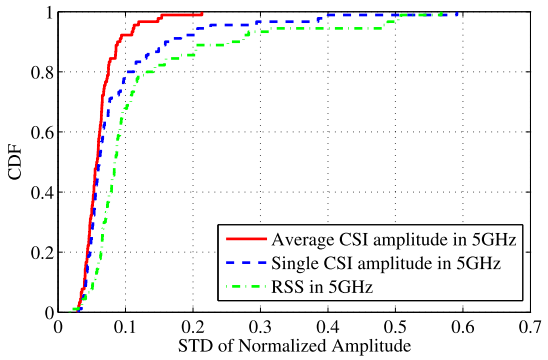


FIGURE 1. The standard deviations CDF of normalized average CSI amplitude, a single CSI amplitude, and a single RSS in the 5GHz OFDM channel for 90 positions.

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 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 in 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% in the case of a single antenna CSI and 70% in the case of a single antenna RSS. Thus, averaging over two adjacent antennas can make CSI amplitude highly stable for a fixed location with 5GHz OFDM channels. We conducted the measurements over a long period of time, during both business hours and overnight 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 two adjacent antennas are separated by more than a half wavelength [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 [21](as observed in Fig. 1). 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 bi-model data, such as estimated AOAs and average amplitudes. This requires that we use the same amount of nodes as input to a deep autoencoder network.

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 have been used to remove the randomness in the CSI phase. The first approach is to make a linear transformation of the phase values measured from the 30 subcarriers [18], [22]. 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 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.

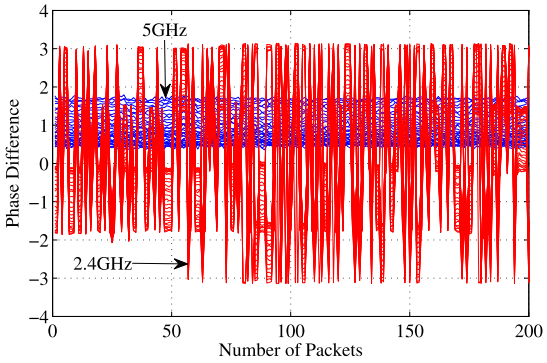


FIGURE 2. The measured phase differences of the 30 subcarriers between two antennas for 200 consecutively received packets in the 5GHz (blue) and 2.4GHz (red) bands.

Switching to the 5GHz band, we find the phase difference becomes highly stable. Fig. 2 shows 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 between [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. 4. We find that all the 5GHz measurements concentrate around 30°, while the 2.4GHz measurements form four clusters around 0°, 90°, 180°, and 270°. The latter occurs 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]. Compared 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 observations made from our experiments. Let $\angle\widehat{CSI}_i$ denote the measured phase of subcarrier i, which is provided by [14], [23]

$$\angle\widehat{CSI}_i = \angle\text{CSI}_i + (\lambda_p + \lambda_s)m_i + \lambda_c + \beta + Z, \tag{4}$$

where $\angle\text{CSI}_i$ 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, and 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 [23], [24], which are expressed by

$$\begin{cases} \lambda_p = 2\pi \dfrac{\Delta t}{N} \\ \lambda_s = 2\pi \left(\dfrac{T' - T}{T}\right) \dfrac{T_s}{T_u} n \\ \lambda_c = 2\pi \, \Delta f T_s n, \end{cases} \tag{5}$$

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_s is the length of the data symbol plus the guard interval, T_u is the data symbol length, n is the sampling time offset for the current packet, and Δf is the difference between the center frequencies at the transmitter and receiver. We cannot obtain the exact values of Δt, $\frac{T'-T}{T}$, n, Δf, and β in (4) and (5). Moreover, λ_p, λ_s and λ_c vary for different packets with different Δt's and n's. 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], [24]. Thus the measured phase difference on subcarrier i between two antennas can be approximated as

$$\Delta\angle\widehat{CSI}_i = \Delta\angle\text{CSI}_i + \Delta\beta + \Delta Z, \tag{6}$$

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 find that $\Delta\angle\widehat{CSI}_i$ is stable for different packets because of the above equation (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 compared 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. (4)).

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 of subscriber i can be translated into an estimation of angle of arrival (AOA), as

$$\theta = \arcsin\left(\frac{\Delta\angle\widehat{CSI}_i\lambda}{2\pi d}\right), \tag{7}$$

where λ is the wavelength of subcarrier i and d is the separation between the two antennas (set to $d = 0.5\lambda$ in our experiments). Although the measured phase difference $\Delta\angle\widehat{CSI}_i$ 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 existing approach for one NIC is to search for $\Delta\beta$ within an AOA pseudospectrum in the range of $[-\pi, \pi]$, which, however, has a high time complexity [19].

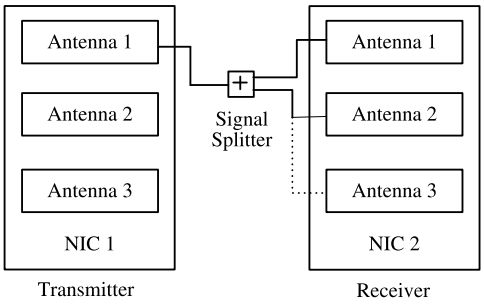


FIGURE 3. The multi-radio hardware design for calibrating the unknown phase offset difference $\Delta\beta$.

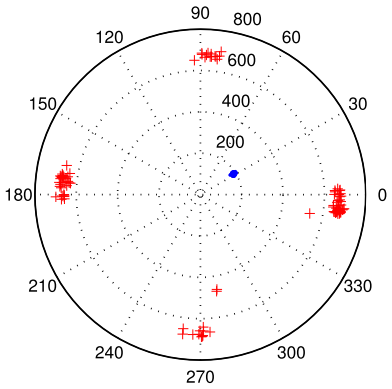


FIGURE 4. The measured phase differences of the 5th subcarrier between two antennas for 200 consecutively received packets in the 5GHz (blue dots) and 2.4GHz (red crosses) bands.

For the proposed Biloc system, we design a novel method to remove the unknown phase offset difference $\Delta\beta$ using two Intel 5300 NICs. As in Fig. 3, we use one Intel 5300 NIC as the transmitter and the other as the receiver, while a *signal splitter* is used to route the 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 and to obtain the unknown phase offset difference between them as well. The unknown phase offset difference is found to be relatively stable over time.

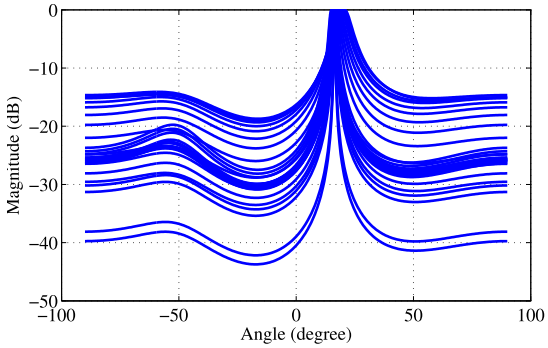


FIGURE 5. The estimated AOAs from the 30 subcarriers using the MUSIC algorithm, while the real AOA is 14°.

After calibrating the unknown phase offset differences for the three antennas, we then use the MUSIC algorithm for AOA estimation [25]. In Fig. 5, 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 in Fig. 5 that the peak occurs at around 20°, indicating an AOA estimation error of about 6°. In fact, there are multiple paths indoor environments. Thus, using only three antennas still cannot obtain accurate angle estimation. Moreover, we implemented more experiments for angles estimation by using MUSIC indoor environments; we find that the estimated angles change for different locations because of NLOS paths. Although the calibrated phase differences are not available for estimating angles when using three antennas, we believe that the new phase calibrated method can be used for future WiFi systems, such as IEEE 802.11ac, that have more than three antennas.

We can obtain the true incoming angle with MUSIC when the LOS component is strong. To deal with the case of strong NLOS paths (typical in indoor environments), we adopt a deep autoencoder network to learn the estimated AOAs 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.

$$\theta = \arcsin\left(\left(\Delta\angle\widehat{CSI}_i - \Delta\beta\right)\frac{\lambda}{2\pi d}\right) + \frac{\pi}{2}, \tag{8}$$

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

III. THE BiLoc DESIGN

A. BiLoc SYSTEM FRAMEWORK

The overall framework of BiLoc is illustrated in Fig. 6. The BiLoc design uses one mobile device and one access point, each equipped with an Intel 5300 NIC, serving as receiver and transmitter, respectively [26], [27]. All the communications are on the 5GHz band. We can collect 90 CSI data for every received packet from the three antennas of the NIC. We then calibrate the phase information from the received

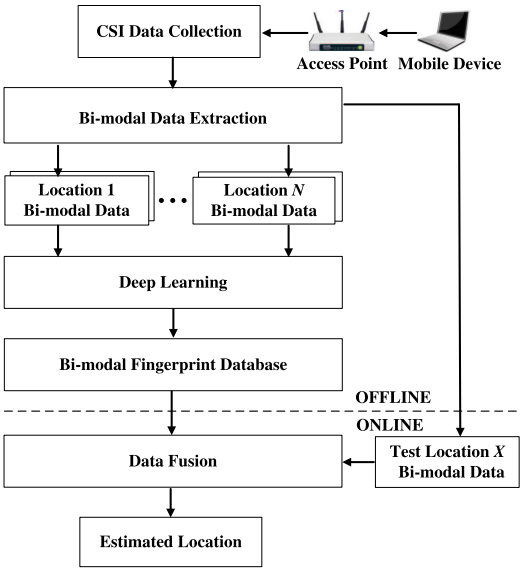


FIGURE 6. The BiLoc system architecture.

CSI data using our multi-radio hardware design (see Fig. 3). 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) those 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 will be significantly reduced. However, the estimated AOA will be less affected. 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, the weights in the deep autoencoder network will be used as features-based fingerprints for every position. By obtaining the optimal weights with the bi-modal data on estimated AOAs 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.

B. 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 network that involves three phases: pretraining, unrolling, and fine-tuning [16], [28]. In the pretraining phase, a deep autoencoder 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 each layer i, for $i = 1, 2, 3$, 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 the adjacent layers.

To reduce the computational complexity of obtaining training weights, a greedy learning algorithm is developed to learn the weights *layer after layer* based on an RBM stack [29]. We consider an RBM as a bipartite undirected graphical model [29] with joint distribution $\Pr(h^{i-1}, h^i)$, as

$$\Pr(h^{i-1}, h^i) = \frac{\exp(-\mathbb{E}(h^{i-1}, h^i))}{\sum_{h^{i-1}} \sum_{h^i} \exp(-\mathbb{E}(h^{i-1}, h^i))}, \tag{9}$$

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

$$\mathbb{E}(h^{i-1}, h^i) = -b^{i-1}h^{i-1} - b^i h^i - h^{i-1}W_i h^i, \tag{10}$$

where b^{i-1} and b^i are the biases for layer $(i-1)$ and layer i units, respectively. To obtain the joint distribution $\Pr(h^{i-1}, h^i)$, we use a contrastive divergence with a one step iteration (CD-1) algorithm to solve it as [29]

$$\begin{cases} \Pr(h^{i-1}|h^i) = \prod_{j=1}^{K_{i-1}} \Pr(h_j^{i-1}|h^i) \\ \Pr(h^i|h^{i-1}) = \prod_{j=1}^{K_i} \Pr(h_j^i|h^{i-1}), \end{cases} \tag{11}$$

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:

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

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 the conditional probability $\Pr(h^i|\hat{h}^{i-1})$, we can obtain an estimate $\hat{h}^i$ [30]. We thus have the following equations for updating the parameters.

$$\begin{cases} \Delta W_i = \epsilon(h^{i-1}h^i - \hat{h}^{i-1}\hat{h}^i) \\ \Delta b^i = \epsilon(h^i - \hat{h}^i) \\ \Delta b^{i-1} = \epsilon(h^{i-1} - \hat{h}^{i-1}), \end{cases} \tag{12}$$

where ϵ is the step size.

When the pretraining is over, we then unroll the deep autoencoder network using *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 autoencoder network according to the error between the reconstructed and input data. The optimal weights are obtained by minimizing the error.

In BiLoc, we use estimated AOAs and average amplitudes as input data, and obtain two sets of optimal weights for the bi-modal fingerprint database.

C. ONLINE DATA FUSION FOR POSITION ESTIMATION

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

$$\Pr(x_i|v^1, v^2) = \frac{\Pr(x_i)\Pr(v^1, v^2|x_i)}{\sum_{k=1}^{K}\Pr(x_k)\Pr(v^1, v^2|x_k)}, \tag{13}$$

where x_i is the ith training position, K is the total number of training positions, and $\Pr(x_i)$ is the prior probability, which is uniformly distributed for any training position x_i [30]. The posteriori probability $\Pr(x_i|v^1, v^2)$ becomes

$$\Pr(x_i|v^1, v^2) = \frac{\Pr(v^1, v^2|x_i)}{\sum_{k=1}^{K}\Pr(v^1, v^2|x_k)}. \tag{14}$$

In BiLoc, we approximate $\Pr(v^1, v^2|x_i)$ with an RBF to model the degree of similarity between the reconstructed bi-modal data and the test bi-modal data, given by

$$\Pr(v^1, v^2|x_i) = \exp\left(-(1-\rho)\frac{\|v^1 - \hat{v}^1\|}{\eta_1\sigma_1} - \rho\frac{\|v^2 - \hat{v}^2\|}{\eta_2\sigma_2}\right), \tag{15}$$

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 (15), the average amplitudes $\hat{v}^1$ and the estimated AOAs $\hat{v}^2$ are the input to the deep autoencoder network. Then, by employing 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 are used to compute the likelihood function $\Pr(v^1, v^2|x_i)$.

The location of the device is finally determined as a weighted average of all the reference positions, that is:

$$\hat{x} = \sum_{i=1}^{K}\Pr(x_i|v^1, v^2)\cdot x_i. \tag{16}$$

IV. EXPERIMENTAL STUDY

A. TEST CONFIGURATION

We will present our experimental study with BiLoc in the 5GHz band in this section. In the experiments, a Dell laptop serves as the mobile device and a desktop computer serves as the access point, both of which are furnished with an Intel 5300 NIC. Our implementation of BiLoc is executed on the Ubuntu desktop 14.04 LTS OS at both the access point and mobile device. We use QPSK modulation and a 1/2 coding rate for the OFDM system. The access point is set in the monitor mode. The two adjacent antennas are separated by a distance of $d = 2.68$ cm, which is half of a wavelength for the 5GHz band. The mobile device transmits at 100 packets per second using only one antenna in injection mode. 5GHz CSI data can be obtained by using the packet injection technique based on LORCON version 1. Then, we extract bi-modal data for the training and test stages as described in Section III-B.

We built three representative schemes from the literature, i.e., Horus [11], FIFS [15], and DeepFi [16], which are discussed in Section I. Moreover, the same dataset captured in the 5GHz band is used by all the schemes to ensure a fair comparison. We conduct extensive experiments with the schemes in the following three representative indoor environments.

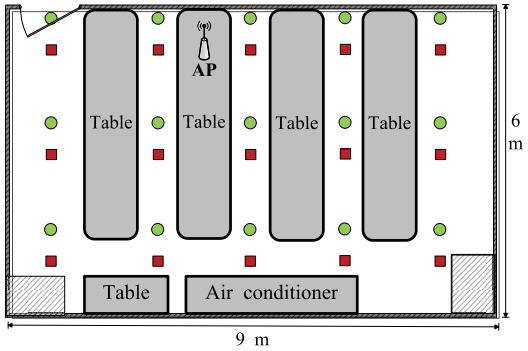


FIGURE 7. The floor plan of the computer laboratory: training locations are represented as red squares and testing locations are represented as green dots.

1) COMPUTER LABORATORY

We utilize a 6×9 m^2 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. 7, 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. The bi-modal CSI data from 1000 received packets is collected for each training location, and that from 25 received packets is collected for each test location. The deep autoencoder 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.

2) CORRIDOR

The 2.4×24 m^2 corridor is illustrated in Fig. 8. In this scenario, the AP is installed at the left 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 1.8 m as before. We again 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 $\{K_1 = 150, K_2 = 100, K_3 = 50\}$, and the ratio ρ for the bi-modal data is set as 0.1.

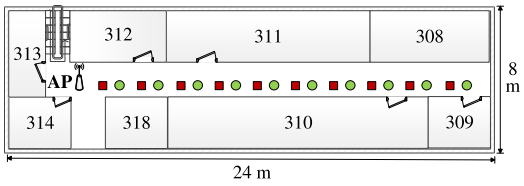


FIGURE 8. The floor plan of the corridor: training locations are represented as red squares and testing locations are represented as green dots.

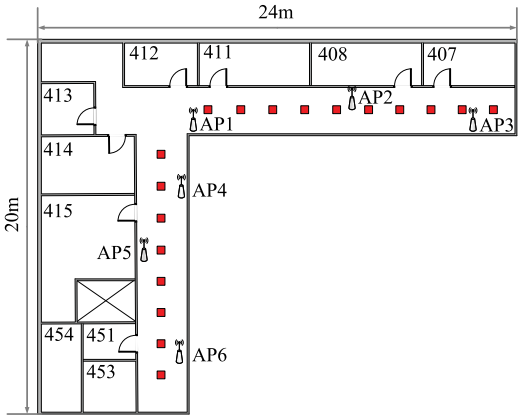


FIGURE 9. The floor plan of the two corridors: training locations are represented as red squares.

3) TWO CORRIDORS

Employed are a 2.4 × 24 m^2 corridor and a 2.4 × 20 m^2 corridor, as shown in Fig. 9. In this scenario, six APs are placed in the two corridors and there are enormous LOS paths. Eighteen training positions (red squares) are arranged along the two corridors, and a number of test positions, which are not shown in Fig. 9, are randomly chosen in the two corridors. The separation between two adjacent training locations is again 1.8 m. As previously, we measure bi-modal data from 1000 packets for each training position and from 25 packets for each test position. Other parameters in the deep network are set in the same way as in the above, one corridor case. We use this deployment scenario to study the effect of the number of access points on indoor localization accuracy for different schemes.

B. ACCURACY OF LOCATION ESTIMATION

We present the mean and STD of localization errors and the execution time of the four schemes for the two scenarios, respectively, in Tables 1 and 2. In the laboratory environment, the mean error of BiLoc is 1.5743 m and the STD error is 0.8312 m for the 15 test points. In the corridor experiment, because only one access point is used for this larger space, the mean error of BiLoc is 2.1501 m and the STD error is 1.5420 m for 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., it is 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. 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

TABLE 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

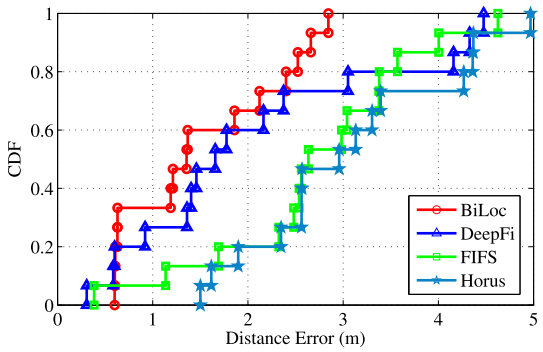


FIGURE 10. CDF of localization errors in 5GHz for the laboratory experiment.

In Fig. 10, we plot the CDF of location errors of the four methods in the laboratory scenario. In this complex propagation environment, with BiLoc, 100% of the test positions have an error smaller than 2.8 m, while DeepFi, FIFS, and Horus achieved an error of less than 2.8 m in approximately 72%, 52%, and 45% of the test locations, 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 AOAs). 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 methods based solely on CSI or RSS amplitudes will be affected.

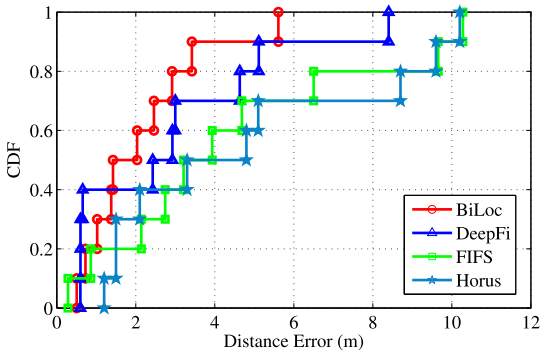


FIGURE 11. CDF of localization errors in 5GHz for the corridor experiment.

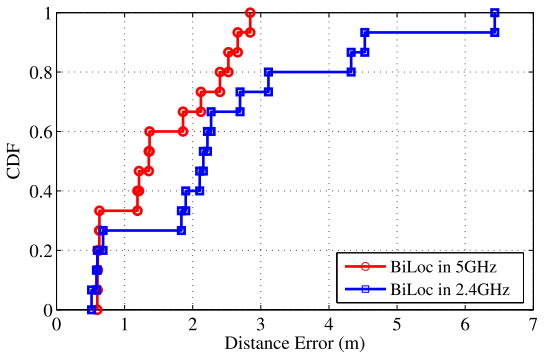


FIGURE 12. CDF of localization errors in 5GHz and 2.4GHz for the laboratory experiment.

In Fig. 11, we plot the error CDFs achieved by the four techniques in the corridor case. Only one access point is used at one end for this 24 m long corridor; thus, it is hard to determine the device's position. With BiLoc, more than 90% of the test positions have an error smaller than 4 m, while with DeepFi, FIFS, and Horus, near to 70%, 60%, and 50% of the test positions, respectively, have an error smaller than 4 m. 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. Regarding the corridor scenario, BiLoc primarily 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.

C. 2.4GHz VERSUS 5GHZ

We also compare the 2.4GHz channel and 5GHz channel using 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. 12 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 smaller than 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 difference between two antennas in 2.4GHz exhibits great variations, which lead to lower localization accuracy. This experiment also validates our Hypothesis 2.

D. IMPACT OF PARAMETER ρ

Recall that the parameter ρ is used to trade off the impacts of average amplitudes and estimated AOAs in location estimation as in (15). We study the impact of ρ on localization accuracy under the two environments. With BiLoc, we use bi-modal data for online testing, and ρ directly influences the likelihood probability $\Pr(v^1, v^2|x_i)$ (15), which in turn influences the localization accuracy.

Table 3 illustrates the mean localization errors for increasing ρ in the laboratory and corridor experiments. In the laboratory experiment, when ρ is increased from 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 AOAs 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 AOAs 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 AOAs provide similar characteristics for deep learning and are not useful for distinguishing the positions. Therefore BiLoc should mostly 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.

TABLE 3. Mean localization error versus parameter ρ.

ρ	0.0	0.01	0.1	0.3	0.5	0.7	0.9	0.99	1.0
Laboratory	2.6799	2.6222	2.0435	1.0567	1.5529	1.5432	1.8120	1.9602	1.9756
Corridor	2.1305	2.1311	2.1501	2.4356	3.2569	4.1484	4.3485	4.2574	4.2817

E. EFFECT OF THE NUMBER OF PACKETS

We study the effect of the number of packets used in the online test stage of BiLoc. In this experiment, we estimate the location of the mobile device using different numbers of packets for the two indoor environments. Although 1000 test packets are received for each position, we only use 25, 50, 100, and 300 of them in the online test for location estimation. We also randomly select the parameter ρ value to guarantee the consistency of localization results obtained with different number of packets.

TABLE 4. Mean localization error versus the number of packets used in the online test stage.

# of Packets	25	50	100	300
Laboratory	1.6829	1.6341	1.5966	1.5479
Corridor	2.3030	2.2906	2.2819	2.0691

In Table 4, we plot the average errors for different numbers of packets in the corridor and laboratory experiments. We can see that the mean distance error in the laboratory experiment is lower than that in the corridor experiment for different amounts of packets. Moreover, with the increase of packets, the mean distance error for both experiments is decreased. Also, we can find that the maximum distance errors for the laboratory and corridor experiments are 1.7 m and 2.3 m, respectively, while the minimum distance errors for the laboratory and corridor experiments are 1.58 m and 2.1 m, respectively. In fact, with the increase in packets, the decrease of mean distance error is small for both experiments. Therefore, we choose 25 packets for the test phase in the proposed BiLoc system, which can obtain a lower computational complexity and a satisfactory localization performance.

F. EFFECT OF THE NUMBER OF NODES IN THE DEEP NETWORK

We study the impact of the number of nodes in deep network on localization results of our BiLoc system. Although there are lots of values we can set for the parameters K_1, K_2 and K_3, the number of all nodes $(K_1 + K_2 + K_3)$ in deep network is selected. In addition, we choose the number of nodes as 200, 300 and 400, respectively, for the two indoor cases.

TABLE 5. Mean localization error versus the number of nodes in the deep network.

# of Nodes	200	300	400
Laboratory	1.7906	1.5743	1.5201
Corridor	2.1710	2.1501	2.0937

In Table 5, we plot the mean errors for the increased number of nodes in the deep network for the laboratory and corridor experiments. We can see that the mean distance error decreases with the increasing number of nodes in the deep network in both experiments. It is noticed that the difference of the mean errors is small for different numbers of nodes, where the mean error is from 1.8m to 1.5m in the laboratory experiment, and that is from 2.2m to 2.1m in the corridor experiment. This demonstrates that our BiLoc is robust for different number of nodes in the deep network. Thus, we select the proper number of nodes with 300 that is $K_1 = 150$, $K_2 = 100$ and $K_3 = 50$, thus obtaining the lower cost.

G. IMPACT OF THE NUMBER OF APs

Finally, we study the impact of the number of APs on localization results for different methods, where we consider the two corridors deployment with six APs. For multiple APs, we evaluate the online localization scheme by multiplying equation (14) for all APs to obtain the fusion likelihood function for every location [15]. Moreover, we implement the proposed BiLoc system without calibration for indoor localization. Also, we take into account the transmitter device can access the maximum three APs.

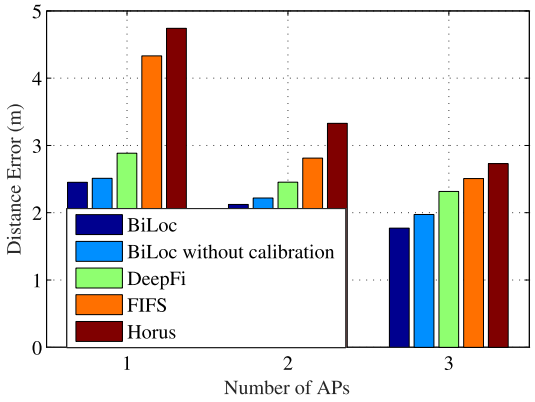


FIGURE 13. Mean errors versus the number of APs for the two corridors experiment.

Fig. 13 presents the mean distance errors for the increased number of APs based on five different schemes in the two corridors environment. It is noticed that, the mean error is decreased with more APs for all the schemes. Moreover, for the BiLoc system, we can see that with the increase of the number of APs from 1 to 3, the distance error is decreased from approximately 2.4 m to 1.8 m, which can improve the localization accuracy. However, we can see that, compared with other traditional methods such as Horus and FIFS, the improvement of localization accuracy is minor. Moreover, we can find that the proposed method with a single AP can obtain even higher accuracy than other methods. Furthermore, we also find that the BiLoc system has a better localization performance than BiLoc system without calibration. Thus, we consider using one AP for the BiLoc system based on the proposed method, which achieves the lower localization error and device cost.

V. RELATED WORK

Indoor localization has been widely studied in the last two decades [31]. This work is closely related to two types of indoor localization schemes, i.e., fingerprinting-based and AOA-based schemes, which are discussed in the following.

Fingerprinting is a mainstream approach for indoor localization [32], [33]. RADAR [10] is the one of the first works for RSS-based indoor localization and uses K-nearest-neighbor for location estimation. Another RSS-based scheme, Horus [11] uses a probabilistic method for indoor fingerprinting, thus improving localization accuracy. Furthermore, CSI data are exploited for indoor fingerprinting to improve the localization accuracy. FIFS [15], PinLoc [34], PhaseFi [22], [30] and DeepFi [16], [35] are all CSI-based schemes that leverage CSI data to achieve reliable fingerprints. Although these systems achieve good localization performances, they require offline calibration to create the fingerprint database via war-driving. An effective approach to reducing the burden of war-driving is crowdsourcing, where the load of fingerprinting is shared by multiple users [36]. A example is the Zee system [37], which leverages the inertial sensors and WiFi data to estimate a user's moving trajectory and to build fingerprints during crowdsourcing.

AOA-based indoor localization leverages array antennas to estimate the arriving angle and then uses geometric relationships to obtain the user location. Recently, the CUPID system [13] achieved a mean error of about 20° using the MUSIC algorithm [25]. CUPID uses the Atheros chipset and three antennas, from which CSI can be obtained for AOA estimation. The relatively large angle error is due to the low resolution of the antenna array. To achieve high angle accuracy, the Array-Track system [38] is implemented with two WARP FPGA-based software radios equipped with 16 antennas for AOA estimation. It can then obtain the direct path by using spatial smoothing for indoor localization. Array-Track requires a large number of antennas, which may not be practical for commodity mobile devices.

The proposed BiLoc system is motivated by these interesting prior works. To the best of our knowledge, it is the first to utilize bi-model data, i.e., both estimated AOAs and average amplitudes, for indoor localization with commodity 5 GHz WiFi. It exploits both the spatial diversity from multiple antennas and the spectral diversity of OFDM channels based on a deep autoencoder network, thus overcoming the limitation of the low resolution of antenna array in off-the-shelf wireless devices.

VI. CONCLUSIONS

We proposed BiLoc, a bi-modal deep learning system for fingerprinting-based indoor localization with 5GHz commodity WiFi NICs. In BiLoc, we first 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. In the training phase, we leveraged a deep autoencoder network to train the bi-modal data, and the weights were used to represent the bi-modal fingerprints. In the test phase, a Bayesian approach based probability model was employed for estimating position with bi-model test data. We evaluated the performance of BiLoc with extensive experiments under three representative indoor environments. The experimental results validated the superior performance of BiLoc over several benchmark schemes.

REFERENCES

[1] D. Zhang, S. Zhao, L. T. Yang, M. Chen, Y. Wang, and H. Liu, "NextMe: Localization using cellular traces in Internet of Things," *IEEE Trans Ind. Informat.*, vol. 11, no. 2, pp. 302–312, Apr. 2015.

[2] K. Derr and M. Manic, "Wireless sensor networks—Node localization for various industry problems," *IEEE Trans. Ind. Informat.*, vol. 11, no. 3, pp. 752–762, Jun. 2015.

[3] A. Abu-Mahfouz and G. P. Hancke, "Distance bounding: A practical security solution for real-time location systems," *IEEE Trans Ind. Informat.*, vol. 9, no. 1, pp. 16–27, Feb. 2013.

[4] S. Ivanov and E. Nett, "Localization-based radio model calibration for fault-tolerant wireless mesh networks," *IEEE Trans Ind. Informat.*, vol. 9, no. 1, pp. 246–253, Feb. 2013.

[5] J. M. Pak, C. K. Ahn, Y. S. Shmaliy, and M. T. Lim, "Improving reliability of particle filter-based localization in wireless sensor networks via hybrid particle/FIR filtering," *IEEE Trans. Ind. Informat.*, vol. 11, no. 5, pp. 1089–1098, Oct. 2015.

[6] B.-F. Wu and C.-L. Jen, "Particle-filter-based radio localization for mobile robots in the environments with low-density WLAN APs," *IEEE Trans. Ind. Electron.*, vol. 61, no. 12, pp. 6860–6870, Dec. 2014.

[7] S. Lee, B. Kim, H. Kim, R. Ha, and H. Cha, "Inertial sensor-based indoor pedestrian localization with minimum 802.15.4a configuration," *IEEE Trans. Ind. Informat.*, vol. 7, no. 3, pp. 455–466, Aug. 2011.

[8] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. ACM Mobicom*, Maui, HI, USA, Sep. 2014, pp. 617–628.

[9] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 6, pp. 1067–1080, Nov. 2007.

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

[11] M. Youssef and A. Agrawala, "The Horus WLAN location determination system," in *Proc. ACM MobiSys*, Seattle, WA, USA, Jun. 2005, pp. 205–218.

[12] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," in *Proc. ACM SIGCOMM*, New Delhi, India, Sep. 2010, pp. 159–170.

[13] S. Sen, J. Lee, K.-H. Kim, and P. Congdon, "Avoiding multipath to revive inbuilding WiFi localization," in *Proc. ACM MobiSys*, Taipei, Taiwan, Jun. 2013, pp. 249–262.

[14] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity WiFi," in *Proc. ACM Mobicom*, Paris, France, Sep. 2015, pp. 53–64.

[15] J. Xiao, K. Wu, Y. Yi, and L. M. Ni, "FIFS: Fine-grained indoor fingerprinting system," in *Proc. IEEE ICCCN*, Munich, Germany, Jul. 2012, pp. 1–7.

[16] X. Wang, L. Gao, S. Mao, and S. Pandey, "DeepFi: Deep learning for indoor fingerprinting using channel state information," in *Proc. WCNC*, New Orleans, LA, USA, Mar. 2015, pp. 1666–1671.

[17] C. Wu, Z. Yang, Z. Zhou, K. Qian, Y. Liu, and M. Liu, "PhaseU: Real-time LOS identification with WiFi," in *Proc. IEEE INFOCOM*, Hong Kong, Apr./May 2015, pp. 2038–2046.

[18] K. Qian, C. Wu, Z. Yang, Y. Liu, and Z. Zhou, "PADS: Passive detection of moving targets with dynamic speed using PHY layer information," in *Proc. IEEE ICPADS*, Hsinchu, Taiwan, Dec. 2014, pp. 1–8.

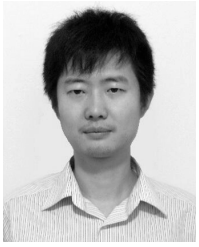
[19] J. Gjengset, J. Xiong, G. McPhillips, and K. Jamieson, "Phaser: Enabling phased array signal processing on commodity WiFi access points," in *Proc. ACM Mobicom*, Maui, HI, USA, Sep. 2014, pp. 153–164.

[20] M. B. Akbar, D. G. Taylor, and G. D. Durgin, "Amplitude and phase difference estimation bounds for multisensor based tracking of RFID tags," in *Proc. IEEE RFID*, San Diego, CA, USA, Apr. 2015, pp. 105–112.

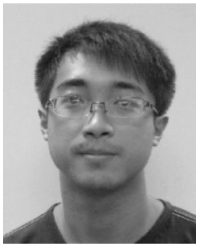
[21] K. Kleisouris, Y. Chen, J. Yang, and R. P. Martin, "The impact of using multiple antennas on wireless localization," in *Proc. IEEE SECON*, San Francisco, CA, USA, Jun. 2008, pp. 55–63.

[22] X. Wang, L. Gao, and S. Mao, "PhaseFi: Phase fingerprinting for indoor localization with a deep learning approach," in *Proc. GLOBECOM*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[23] M. Speth, S. A. Fechtel, G. Fock, and H. Meyr, "Optimum receiver design for wireless broad-band systems using OFDM—Part I," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1668–1677, Nov. 1999.
[24] X. Wang, C. Yang, and S. Mao, "TensorBeat: Tensor decomposition for monitoring multi-person breathing beats with commodity WiFi," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 4, Aug. 2017.
[25] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
[26] A. T. Mariakakis, S. Sen, J. Lee, and K.-H. Kim, "SAIL: Single access point-based indoor localization," in *Proc. ACM MobiSys*, Bretton Woods, NH, USA, Jun. 2014, pp. 315–328.
[27] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single WiFi access point," in *Proc. ACM NSDI*, Santa Clara, CA, USA, Mar. 2016, pp. 165–178.
[28] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
[29] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19. Vancouver, BC, Canada, Dec. 2007, pp. 153–160.
[30] X. Wang, L. Gao, and S. Mao, "CSI phase fingerprinting for indoor localization with a deep learning approach," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1113–1123, Dec. 2016.
[31] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, p. 25, Nov. 2013.
[32] Y. Wen, X. Tian, X. Wang, and S. Lu, "Fundamental limits of RSS fingerprinting based indoor localization," in *Proc. IEEE INFOCOM*, Hong Kong, Apr. 2015, pp. 2479–2487.
[33] S. He, S.-H. G. Chan, L. Yu, and N. Liu, "Fusing noisy fingerprints with distance bounds for indoor localization," in *Proc. IEEE INFOCOM*, Hong Kong, Apr. 2015, pp. 2506–2514.
[34] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka, "You are facing the Mona Lisa: Spot localization using PHY layer information," in *Proc. ACM MobiSys*, Low Wood Bay, U.K., Jun. 2012, pp. 183–196.
[35] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan. 2017.
[36] X. Zhang, Z. Yang, C. Wu, W. Sun, Y. Liu, and K. Liu, "Robust trajectory estimation for crowdsourcing-based mobile applications," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 7, pp. 1876–1885, Jul. 2014.
[37] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proc. ACM Mobicom*, Istanbul, Turkey, Aug. 2012, pp. 293–304.
[38] J. Xiong and K. Jamieson, "ArrayTrack: A fine-grained indoor location system," in *Proc. ACM NSDI*, Lombard, IL, USA, Apr. 2013, pp. 71–84.



XUYU WANG (S'13) received the B.S. degree in electronic information engineering and the M.S. degree in signal and information processing from Xidian University, Xi'an, China, in 2009 and 2012, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA. His research interests include indoor localization, deep learning, wireless communications, software defined radio, and big data. He was a recipient of a Woltolsz Fellowship at Auburn University and a co-recipient of the Second Prize of Natural Scientific Award of Ministry of Education, China, in 2013.



LINGJUN GAO (S'14) received the B.E. degree in electrical engineering from the Civil Aviation University of China, Tianjing, China, in 2013, and the M.S. degree in electrical and computer engineering from Auburn University, Auburn, AL, USA, in 2015. He is currently a Data Engineer with DataYes, Inc., Shanghai, China. His research interests include machine learning, indoor localization, and testbed implementation.

SHIWEN MAO (S'99–M'04–SM'09) received the Ph.D. degree in electrical and computer engineering from Polytechnic University, Brooklyn, NY, in 2004. He is currently the Samuel Ginn Distinguished Professor and the Director of the Wireless Engineering Research and Education Center, Auburn University, Auburn, AL, USA. His research interests include wireless networks and multimedia communications. He is a Distinguished Lecturer of the IEEE Vehicular Technology Society. He received the 2015 IEEE ComSoc TC-CSR Distinguished Service Award, the 2013 IEEE ComSoc MMTC Outstanding Leadership Award, and the NSF CAREER Award in 2010. He was a co-recipient of Best Paper Awards from the IEEE GLOBECOM 2016, the IEEE GLOBECOM 2015, the IEEE WCNC 2015, and the IEEE ICC 2013, and the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems. He serves as a TPC Co-Chair of the IEEE INFOCOM 2018, the Area TPC Chair of the IEEE INFOCOM 2017 and 2016, the Technical Program Vice Chair for Information Systems of the IEEE INFOCOM 2015, a Symposium/Track Co-Chair for many conferences, including the IEEE ICC, the IEEE GLOBECOM, ICCCN, among others, Steering Committee of the IEEE ICME and AdhocNets, and in various roles in the organizing committees of many conferences. He is the Chair of the IEEE ComSoc Multimedia Communications Technical Committee. He is on the Steering Committees of the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING and the IEEE TRANSACTIONS ON MULTIMEDIA, and on the Editorial Boards of the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE INTERNET OF THINGS JOURNAL, the IEEE MULTIMEDIA, and the *ACM GetMobile*, among others.