MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Fingerprinting-Based Indoor Localization with Commercial MMWave WiFi: A Deep Learning Approach

Koike-Akino, Toshiaki; Wang, Pu; Pajovic, Milutin; Sun, Haijian; Orlik, Philip V.

TR2020-054 May 02, 2020

Abstract

Existing fingerprint-based indoor localization uses either fine-grained channel state information (CSI) from the physical layer or coarse-grained received signal strength indicator (RSSI) measurements. In this paper, we propose to use a mid grained intermediate-level channel measurement — spatial beam signal-to-noise ratios (SNRs) that are inherently available and defined in the IEEE 802.11ad/ay standards — to construct the fingerprinting database. These intermediate channel measurements are further utilized by a deep learning approach for multiple purposes: 1) location-only classification; 2) simultaneous locationand orientation classification; and 3) direct coordinate estimation. Furthermore, the effectiveness of the framework is thoroughly validated by an in-house experimental platform consisting of 3 access points using commercial-off-the-shelf millimeter-wave WiFi routers. The results show a 100% accuracy if the location is only interested, about 99% for simultaneous location-and orientations classification, and an averaged root mean-square error (RMSE) of 11.1 cm and an average median error of 9.5 cm for direct coordinate estimate, greater than 2-fold improvements over the RMSE of 28.7 cm and median error of 23.6 cm for RSSI-like single SNR-based localization

IEEE Access

Date of publication xxxx 00, 2020, date of current version xxxx 00, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.DOI

Fingerprinting-Based Indoor Localization with Commercial MMWave WiFi: A Deep Learning Approach

TOSHIAKI KOIKE-AKINO[1], (Senior Member, IEEE), PU WANG[1], (Senior Member, IEEE), MILUTIN PAJOVIC[1], (Member, IEEE), HAIJIAN SUN[2], (MEMBER, IEEE), AND PHILIP V. ORLIK[1], (Senior Member, IEEE)

[1]Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA
[2]University of Wisconsin-Whitewater, Whitewater, WI 53190, USA

Corresponding author: Toshiaki Koike-Akino (e-mail: koike@merl.com).

ABSTRACT Existing fingerprint-based indoor localization uses either fine-grained channel state information (CSI) from the physical layer or coarse-grained received signal strength indicator (RSSI) measurements. In this paper, we propose to use a mid-grained intermediate-level channel measurement — spatial beam signal-to-noise ratios (SNRs) that are inherently available and defined in the IEEE 802.11ad/ay standards — to construct the fingerprinting database. These intermediate channel measurements are further utilized by a deep learning approach for multiple purposes: 1) location-only classification; 2) simultaneous location-and-orientation classification; and 3) direct coordinate estimation. Furthermore, the effectiveness of the framework is thoroughly validated by an in-house experimental platform consisting of 3 access points using commercial-off-the-shelf millimeter-wave WiFi routers. The results show a 100% accuracy if the location is only interested, about 99% for simultaneous location-and-orientations classification, and an averaged root mean-square error (RMSE) of 11.1 cm and an average median error of 9.5 cm for direct coordinate estimate, greater than 2-fold improvements over the RMSE of 28.7 cm and median error of 23.6 cm for RSSI-like single SNR-based localization.

INDEX TERMS Indoor localization, WiFi, millimeter wave, fingerprinting, machine learning, deep neural networks, location, orientation, coordinate estimation.

I. INTRODUCTION

LOCALIZATION of people, objects and devices in indoor environments has received tremendous attention over the past few decades. Although the global positioning system (GPS) is a prevailing technology for outdoor localization, its use for indoor localization has been prevented due to its large attenuation when penetrating buildings.

Radio frequency (RF) technologies, e.g., WiFi, infrared, RF identification, ultra wide-band (UWB), Zigbee, Bluetooth, digital television, cellular and frequency-modulation (FM) radio, have been proposed for indoor localization with varying degree of implementation complexity and resulting accuracy. Most of them have been built upon information/estimation either on i) time, e.g., time of arrival (ToA), time of flight (ToF), time difference of arrival (TDoA), ii) angles, e.g., angles of arrival (AoA) and departure (AoD), iii) phases, e.g., phase of arrival and phase difference, and iv) power, e.g., received signal strength indicator (RSSI) or signal-to-noise ratio (SNR) [1], [2].

Compared with technologies requiring dedicated hardware, such as anchors in UWB localization systems, indoor localization systems using existing infrastructure are more cost-effective solutions. Given its ubiquitous presence, WiFi stands out as a technology for infrastructure-free indoor localization. Most WiFi-based indoor localization frameworks use either fine-grained channel state information (CSI) from the physical layer [3]–[12] or coarse-grained RSSI measurements from the MAC layer [13]–[29] for fingerprinting or direct localization; see more detailed literature review in the next section.

The conventional RSSI measurement suffers from the measurement instability and coarse granularity of the channel information, leading to limited accuracy for localization. The CSI measurement is more fine-grained but requires access

to physical-layer interfaces and high computational power to process a large amount of sub-carrier data. These limitations motivate us to use mid-grained intermediate channel measurements which are more informative (e.g., in the spatial domain) than the RSSI measurement and easier to access than the lower-level CSI measurement. Specifically, this paper proposes to use a new type of intermediate channel measurement — spatial beam SNRs — that are inherently available (with zero overhead) for beam training for the fifth-generation (5G) and IEEE 802.11ad/ay standards operating at millimeter-wave (mmWave) bands, to construct the fingerprinting database.

Using commercial-off-the-shelf (COTS) 802.11ad routers, we conduct proof-of-concept experiments to collect the beam SNR measurements at several location-of-interests for constructing a fingerprinting dataset at regular office environments. For the in-house measurement dataset, both classification and coordinate estimation are considered using a deep neural network architecture inspired by residual network (ResNet) [30] for location/orientation identification and coordinate estimation. To verify the advantage of proposed beam SNRs fingerprinting and neural networks, the location accuracy and estimation error are analyzed through the comparison to various machine learning methods. Our contributions and results are summarized as follows:

- We propose to fingerprint beam SNR measurements for location and orientation for indoor localization as they provide relatively rich information on spatial propagation paths of mmWave signals used during beam training phase in IEEE 802.11ad standards, and are accessible from COTS 802.11ad chipsets.
- We introduce a ResNet-inspired deep neural network (DNN) by fusing feedforward fully-connected layers and shortcut connections for one-dimensional beam SNR vectors from multiple access points (APs).
- We implement a mmWave fingerprint-based indoor localization system consisting of 4 COTS 802.11ad-compliant WiFi routers and collect real-world measurements in an office space during regular business hours.
- We conduct comprehensive performance analysis by evaluating performance as a function of the number of APs, training data size, sliding-window size, orientation mismatch, and off-grid locations.
- High-accuracy localization performance is achieved by using beam SNRs, which is greater than 2-fold improvements over the conventional RSSI-like single SNR-based fingerprinting localization.

It is noted that this paper takes one step further from our preliminary work in [31] and [32] by introducing the customized deep learning (DL) neural network and achieving significant improvements, especially for the coordinate estimation.

It is worth noting that our work is inspired by earlier efforts in [33]–[35] which enabled easy access to beam SNR measurements from COTS 802.11ad WiFi routers. However, rather than formulating it to a direct localization as a constrained optimization and requiring dedicated chamber measurements of beam patterns, we propose to direct fingerprint beam SNR measurements as features for location and orientation. This is motivated by the conventional wisdom that fingerprinting yields better performance than direct localization by registering locations-of-interest directly with WiFi propagation features without the need for an accurate propagation model.

The remainder of the paper is organized as follows. Section II reviews the existing literature of using the coarse-grained RSSI measurements and fine-grained CSI measurements for indoor localization. In Section III, we introduce the principle of a multi-AP data collection system. Section IV details the offline fingerprinting phase to build the labeled training dataset and the deep learning-based online localization phase. Section V describes the in-house experiment setup, the classification performance, and the accuracy of direct coordinate estimation. Finally, conclusions are drawn in Section VI.

II. LITERATURE REVIEW

In the following, we provide a literature review on WiFi-based indoor localizations using RSSI and CSI measurements and related applications.

A. RSSI FINGERPRINTING

Early WiFi-based indoor localization systems used RSSI measurements to estimate indoor location in a direct localization fashion [13]–[16]. For fingerprinting-based methods, RSSI was used directly as fingerprinting data in systems such as Radar [17], Compass [18], and Horus [19] due to easy access to 802.11ac- and 802.11n-compliant devices.

Classical machine learning methods such as the k-nearest neighbor (kNN) and support vector machine (SVM) were applied to RSSI fingerprinting measurements [17], [20]–[23]. In [19], a probabilistic Bayesian method was proposed to measure the similarity between the test and fingerprinted RSSI measurements. Instead of using parametric statistical distributions such as the Gaussian and lognormal distributions, non-parametric kernel methods were applied to the RSSI measurements to extract statistical distribution of RSSI measurements to infer the likelihood of test measurements [24], [25]. Leveraging modern machine learning frameworks such as discriminant-adaptive neural network [26], robust extreme learning machines [27], and multi-layer neural networks [28], RSSI fingerprinting-based indoor localization methods showed improved localization performance over classical machine learning approaches. More recently, [29] proposed to apply recurrent neural networks (RNNs) to RSSI measurements for utilizing trajectory information.

Nevertheless, RSSI measurements have limitations such as 1) instability of RSSI measurements at a given location and 2) coarse-grained channel information.

B. CSI FINGERPRINTING

At low frequency bands, CSI measurements can be accessed from COTS 802.11n, 802.11ac and 802.11h devices. These data are complex-valued channel measurements over multiple subcarriers at 2.4 and 5 GHz bands [3]–[8]. With richer channel information, a larger amount of CSI measurements from fingerprinted locations can be trained by more advanced deep learning architectures to learn the mapping from CSI to locations. For instance, ConFi [9] used convolutional neural networks (CNNs) to train CSI measurements from three antennas, for classifying the location, and estimating location coordinates with weights equal to the classified category posteriors. [10] fingerprinted full CSI over multiple time instants, calibrated their phases and fitted one autoencoder for one location. An unknown location was estimated as centroid of fingerprinted locations with weights computed from autoencoders' reconstruction errors. Besides the above classification-first localization methods, CSI measurements were trained directly to provide the coordinate estimation by formulating a regression problem in [11], [12].

At mmWave bands (e..g, 28-GHz for 5G communication and 60 GHz for IEEE 802.11ad [36] and 802.15.3c [37]), the use of CSI measurements for fingerprinting was much less reported in the literature due to the cost of a dedicated mmWave platform or no access to CSI measurements from COTS mmWave WiFi devices. RSSI and AoA from multiple APs were fingerprinted and then used to estimate location using the weighted nearest neighbor algorithm [38]. A two-dimensional power delay profile (PDP) over multiple beampatterns was used as fingerprints at 28 GHz band for outdoor localization [39]. It exploited the fact that clients' locations can be registered by multipath delays due to surrounding obstructions (e.g., buildings and trees). To obtain high-resolution PDP, it assumed that base stations can transmit short pulses with a sequence of directive beamforming patterns and a high sample rate was required at the client to separate closely-spaced delays. However, this concept was verified only using ray-tracing simulated datasets.

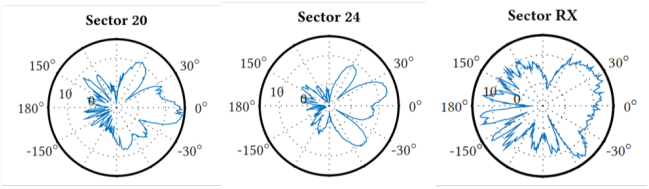
C. RELATED LITERATURE

In the following, we provide a brief overview of direct localization and related sensing applications.

1) mmWave Direct Localization

With no requirements of offline fingerprinting, direct localization methods using mm-Wave channel features were proposed. Examples include a three-stage location and orientation estimation method in [40], direct localization for massive multi-input multi-output (MIMO) based on AoA and ToA in [41], and three-dimensional (3D) localization using a large-scale mmWave uniform cylindrical array [42]. Similarly, [43]–[45] estimated location from knowledge of mmWave channel in the angular domain with one or more APs. Nonetheless, hardware constrains limit the number of RF chains that can be employed in a mmWave device due to cost and power consumption, rendering the above referenced direct localization methods impractical.

(a) Talon AD7200 Router



(b) Beampatterns for beam training

FIGURE 1: (a) Commercial off-the-shelf 802.11ad device: Talon AD7200 router; (b) Two directional transmitting beampatterns and a quasi-omnidirectional receiving beampattern used for beam training are shown. The beampatterns were measured in a chamber at the TU Darmstadt [33].

2) Human Sensing

Beyond indoor localization, WiFi-band and mmWave frequency-modulated continuous-wave (FMCW) signals from dedicated devices and commercial sensing evaluation boards (e.g., TI AWR/IWR chipsets) were utilized to take advantage of their high-resolution range and angle information to track persons behind the wall, determine personal identity, estimate pose and gestures, and track 2D/3D skeleton movements [46]–[51].

With success of mmWave FMCW signals for human sensing, commercial WiFi signals, especially CSI measurements from commercial 802.11n chipsets at low frequency (2.4 GHz) bands, were trained via supervised learning or cross-modal deep learning for human sensing tasks such as device-free localization, activity recognition, fall detection, personal identification, emotion sensing, and skeleton tracking [52]–[62]. Most recently, [60] used annotations from camera images to train fine-grained CSI measurements over 30 subcarriers and 5 frames from 3 transmitting and 3 receiving antennas. The cross-modal deep learning approach showed the great potential of commercial WiFi signals for sensing applications. Nevertheless, explicit utilization of beam features from commercial mmWave communication (5G and WiFi) signals was not yet reported in the literature.

III. DATA COLLECTION SYSTEM

A. HARDWARE

We use TP-Link Talon AD7200 routers to build our in-house data collection system. Complying with IEEE 802.11ad standards, this router implements Qualcomm QCA9500 transceiver that supports a single stream communication in

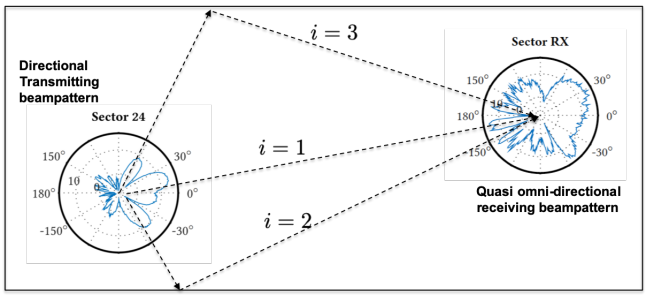


FIGURE 2: Illustration of beam SNR measurements as a function of transmitting and receiving beampatterns.

60 GHz range using analog beamforming over 32-element planar array, as shown in Fig. 1(a).

To search for desired directions, a series of pre-defined beampatterns or sectors are used by APs to send beacon messages to potential clients which are in a listening mode with a quasi-omnidirectional beampattern. These beampatterns were measured in a chamber at the TU Darmstadt [33], [34] and three selected beampatterns (two for transmitting and one for receiving) are shown in Fig. 1(b). Then, clients send a series of beampatterns while the APs are in a listening mode. After beam training, the link can be established by choosing the pair of beampatterns between the AP and clients. Such beam training is periodically repeated and the beam sectors are updated to adapt to the environmental changes. It is noted that the resulting beampatterns depart from the theoretical ones and exhibit fairly irregular shapes due to hardware imperfections and housing at 60 GHz.

B. BEAM SNR

When directional beampatterns are used, beam SNRs are collected by 802.11ad devices as a measure of beam quality. For a given pair of transmitting and receiving beampatterns, corresponding beam SNR can be defined as

$$h_m = \mathsf{BeamSNR}_m = \frac{1}{\sigma^2} \sum_{i=1}^{I} \gamma_m(\theta_i)\zeta_m(\psi_i)P_i, \quad (1)$$

where m is the index of beampattern, I is the total number of paths, θ_i and ψ_i are the transmitting and receiving azimuth angles for the ith path, respectively, P_i is the signal power at the ith path, $\gamma_m(\theta_i)$ and $\zeta_m(\psi_i)$ are the transmitting and receiving beampattern gains at the ith path for the mth beampattern, respectively, and σ^2 is the noise variance. Fig. 2 shows an example of $I = 3$ paths between the transmitting side that probes the spatial domain using the $(m = 24)$th directional beampattern and the receiving side which is in a listening mode. For Talon AD7200 routers, the beam SNR measurements are further quantized in a stepsize of 0.25 dB. Overall, from one beam training, one AP can collect M beam SNRs for M transmitting beampatterns.

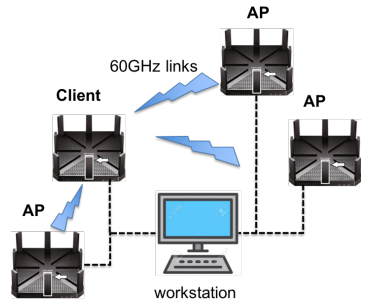
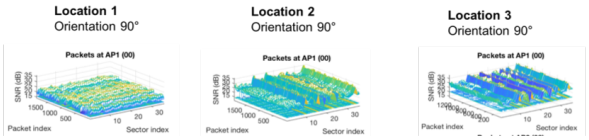
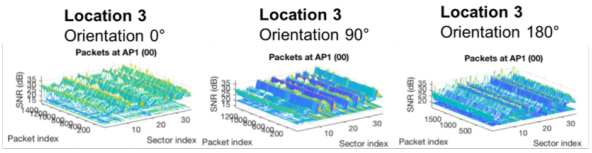


FIGURE 3: The data collection system uses multiple commercial 802.11ad devices as APs and one 802.11ad device as client for fingerprinting. The client sequentially performs beam training over multiple APs. During the beam training phase, beam SNR measurements are collected from each AP to a workstation via Ethernet cables.



(a) Beam SNR measurements at different locations for a given orientation



(b) Beam SNR measurements at different orientations for a given location

FIGURE 4: Beam SNR measurements when the client is located at (a) three locations with the same orientations; and (b) the same location but with different orientations.

C. CONFIGURATION

To access the raw beam SNR measurements at Talon AD7200 routers, we followed the work in [33]–[35] and used the open-source software package in [63]. Particularly, we used the Nexmon firmware patching framework [64], which enables the development of binary firmware extensions in C. By matching the patterns of IEEE 802.11ad beam training frames with the memory inside the chip, one can identify parts of the firmware handling the beam training frames and extract beam SNR measurements from these memory addresses.

The data collection system consists of multiple Talon AD7200 routers, three serving as APs and one as the client, in a configuration shown in Fig. 3. The client sequentially performs beam training over multiple APs. During the beam training phase, beam SNR measurements are collected from each AP to a workstation via Ethernet cables.

D. DATA VISUALIZATION

Fig. 4 shows collected beam SNRs over the time (packet index) and spatial beam (sector index) domains from one AP

to a client. The top row shows the beam SNR measurements when the client is located at three different locations (i.e., Locations 1, 2, 3) with the same orientation (Orientation 90°), while the bottom row shows the beam SNR measurements when the client is located at the same location (Location 3) but with different orientations (i.e., Orientations $0°, 90°, 180°$). Overall, beam SNR measurements are stable over time (packet index) with only a few fluctuations, possibly due to people moving. On the other hand, the measurements are sensitive to the client's location and orientation as beam SNRs change more rapidly over sector index.

IV. INDOOR LOCALIZATION BY MMWAVE BEAM FINGERPRINTING

In the following, we utilize the beam SNRs to build fingerprinting dataset at reference locations and orientations, and then introduce a ResNet-based deep learning approach for classification and coordinate estimation. Compared with our earlier classical machine learning approaches such as the kNN, SVM and Gaussian process (GP) [32], the deep learning approach shows significant improvements on localization errors, as verified in Section. V.

A. OFFLINE TRAINING DATASET

To construct the fingerprinting dataset, we follow the standard procedure by stacking all SNR measurements from all beam sectors as a fingerprinting vector, e.g., $\mathbf{h} = [h_1, h_2, \ldots, h_M]^T$ where M is the number of beampatterns used for beam training and $[\cdot]^T$ denoting the transpose. When multiple APs are used, we combine beam SNR measurements from each AP to form one long fingerprinting snapshot, i.e., $\tilde{\mathbf{h}} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \ldots, \mathbf{h}_P^T]^T \in \mathbb{R}^{MP\times 1}$, where P is the number of APs. For a given location and orientation, R fingerprinting snapshots, $\tilde{\mathbf{h}}_1(l,o), \ldots, \tilde{\mathbf{h}}_R(l,o)$, are collected to construct the offline training dataset, where l and o are the indices for the location and orientation, respectively. By collecting many realizations of beam SNR measurements at multiple APs over L locations-of-interest and O orientations, we will have LO sets of $MP \times R$ beam SNR measurements in the training dataset.

Albeit simple, the offline fingerprinting phase is time- and manpower-consuming. This issue becomes worse when one sets the resolution of fingerprinting positions and orientations to a finer granularity. In our experiment, it is not uncommon to collect the fingerprinting datasets in days. To alleviate this issue, one can borrow the concept of crowdsourcing [65], [66] which exploits pervasive (mmWave) WiFi devices to collect training samples and labels with unconscious cooperation among volunteering users [67], and adaptive sampling which exploits adaptivity to identify highly informative fingerprinting positions and, hence, reduces the amount of labeled samples.

B. ONLINE LOCALIZATION

When new fingerprinting measurements from an unknown location are available, the problem of interest is to identify its location and/or orientation and estimate its coordinate. To this end, we propose a deep learning architecture by fusing feedforward fully connected (FC) layers and shortcut connections (SC) of the ResNet for both classification and coordinate estimation.

1) Proposed Network Architecture

The proposed deep neural network architecture for indoor localization is shown in Fig. 5. It first feeds beam SNRs from multiple APs to an input layer with a dimension of N_w, where N_w refers to the layer width. In the case of three APs, a total of 108 beam SNRs by cascading measurements from APs is used as an input. The input layer is implemented by using a fully-connected linear layer, i.e., $\mathbf{y}_0 = \mathbf{W}_{\text{input}}\tilde{\mathbf{h}} + \mathbf{b}_{\text{input}}$, for a weight of $\mathbf{W}_{\text{input}} \in \mathbb{R}^{N_w\times 108}$ and a bias of $\mathbf{b}_{\text{input}} \in \mathbb{R}^{N_w\times 1}$.

Then, the output $\mathbf{y}_0$ is fed into N_d consecutive residual blocks, where a shortcut connection is used to jump from the input to the output of each residual block in order to learn residual gradient for improved training stability,

$$\mathbf{y}_\ell = f_\ell(\mathbf{y}_{\ell-1}, \boldsymbol{\theta}_\ell) + \mathbf{y}_{\ell-1}, \quad \ell = 1, 2, \ldots, N_d, \qquad (2)$$

where f_ℓ is the nonlinear mapping with weights $\boldsymbol{\theta}_\ell$ to be learned, $\mathbf{y}_\ell$ is the output of the ℓth residual block and input for the next residual block, and N_d is the number of residual blocks.

For the residual block architecture, the form of f_ℓ can be flexible in terms of the number of hidden layers, the use of bottleneck layers for dimension reduction and computational reduction, activation functions, and regularization formats. In Fig. 5, we consider the batch normalization (BN) and rectified linear unit (ReLU) activation function followed by hidden layers implemented by two fully-connected layers of the same dimension of N_w. The use of the same dimension through the residual block allows an identity-mapping shortcut connection which introduces neither additional parameters nor computation complexity, but allows for more efficient gradient backpropagation to mitigate gradient exploding or vanishing. More specifically, the output of the previous residual block $\mathbf{y}_{\ell-1}$ first goes through a batch normalization layer and a ReLU activation layer. Then a fully-connected layer of $N_w \times N_w$ is used for linear combination. This process is repeated again to generate the output of the nonlinear mapping $f_\ell(\mathbf{y}_{\ell-1}, \boldsymbol{\theta}_\ell)$ which is added to the input $\mathbf{y}_{\ell-1}$ which passes through the shortcut connection path. In other words, for the particular architecture, the weights $\boldsymbol{\theta}_\ell$ in (2) includes the linear weights of two hidden layers and associated bias vectors. Finally, dropout operations are used to silence a proportion of nodes of hidden layers to prevent overfitting.

It is easy to see that the proposed deep neural network is inspired by the ResNet [30] which stacks two-dimensional convolution layers and uses shortcut connections for two-dimensional image recognition. By comparing the original ResNet with the proposed architecture, one can note a number of subtle differences here: First, we replace the two-dimensional convolution layers with simple fully-connected

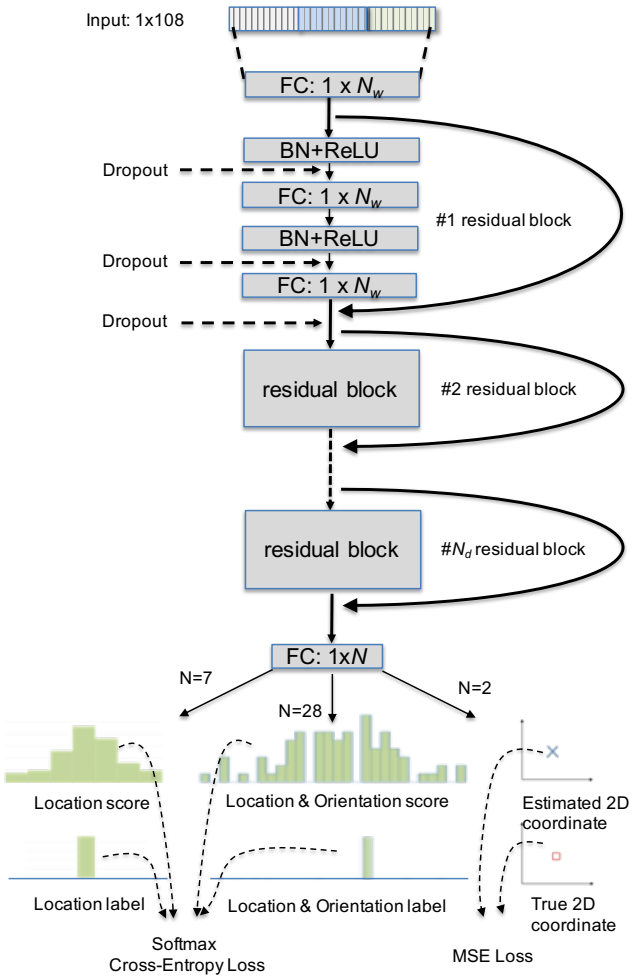


FIGURE 5: Proposed deep learning architecture by fusing feedforward fully connected (FC) layers and shortcut connections (SC) of ResNet along with batch normalization (BN) and dropout regularization operations for multi-purpose indoor localization: 1) location-only classification; 2) simultaneous location-and-orientation classification; and 3) direct coordinate estimation.

layers since we deal with one-dimension vectors of beam SNRs and linear combinations of input are sufficient to capture the interaction among them. Second, as a consequence of the fully connected layers, the shortcut connection is operated over the same dimension (i.e., N_w) as opposed to the skip links in the original ResNet have to bridge over different dimensions by zero-padding identity mapping or projection if a stride of 2 or larger is used. Third, with simple fully-connect layers, dropout operations are more meaningful to randomly silence nodes in hidden layers and prevent overfitting.

Finally, for the output layer, we use a fully-connected layer to generate an output vector $\mathbf{u} = \mathbf{W}_{\text{output}}\mathbf{y}_{N_d} + \mathbf{b}_{\text{output}}$ with a dimension of N, where N is determined by the objective: 1) $N = 7$ for the location-only classification; 2) $N = 28$ for the simultaneous location-and-orientation classification; and

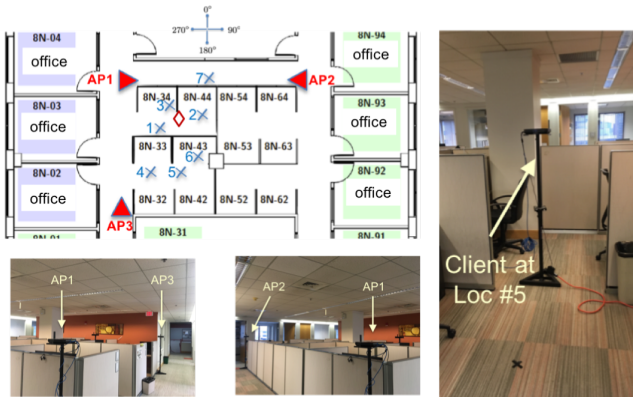


FIGURE 6: Experimental setup with 3 APs (denoted by triangles) in 7 locations-of-interest (denoted by crosses) and 4 orientations in an office environment during regular hours.

3) $N = 2$ for the two-dimensional coordinate estimation. In the following, we further elaborate the three cases.

2) Classification: Location-Only and Simultaneous Location-and-Orientation

With the above network architecture, one can attach a classification output layer to assign new beam SNRs into one of fingerprinted locations and orientations. This is achieved by formulating it as a classification problem. If only the location is interested, the dimension of the last fully-connected output layer is $N = 7$ for our experiments, while $N = 28$ if 7 locations and 4 orientations are simultaneously identified. For a training input with a label, the corresponding output of the last layer $\mathbf{u}$ is first normalized with the softmax operation as

$$s_n = \exp(u_n) \Big/ \sum_{i=1}^{N} \exp(u_i), \quad n \in \{1, 2, \ldots, N\}. \tag{3}$$

where s_n is the nth element of the normalized output u_n that is referred to as the location or location-orientation score vector in Fig. 5. Then, the cross-entropy loss function is computed over the score vector $\mathbf{s} = [s_1, s_2, \ldots, s_N]$ and the corresponding one-hot label vector $\mathbf{c} = [c_1, c_2, \ldots, c_N]$ as

$$L_{\text{classification}} = -\sum_{n} c_n \log(s_n). \tag{4}$$

The average probability of successful classification (or accuracy) is calculated by the ratio between the number of correct estimations and total samples, i.e., $\Pr(\arg\max_i s_i = \arg\max_i c_i)$ where $\Pr(\cdot)$ denotes the sample probability that the argument event is true.

3) Regression: Coordinate Estimation

One can also estimate the coordinates of new measurements by formulating it as a regression problem. For the fingerprinting training dataset, the label is changed from the pair of location and orientation to the coordinate values of the fingerprinted location. Therefore, we set $N = 2$ in the

TABLE 1: Number of training (test) samples for each location and orientation

Loc. / Ori.	0°	90°	180°	270°
1	417 (480)	594 (549)	562 (361)	560 (326)
2	572 (140)	546 (302)	582 (176)	402 (224)
3	207 (319)	267 (253)	565 (328)	428 (299)
4	520 (204)	510 (192)	453 (167)	129 (223)
5	511 (287)	498 (322)	396 (307)	118 (303)
6	507 (427)	419 (220)	300 (190)	281 (156)
7	530 (199)	210 (72)	413 (139)	510 (196)

output layer $\mathbf{u}$ for the two coordinate values in the Cartesian coordinate system. Then, the mean-square error (MSE) of the coordinate estimation is used as a loss function:

$$L_{\text{regression}} = |x - u_1|^2 + |y - u_2|^2, \tag{5}$$

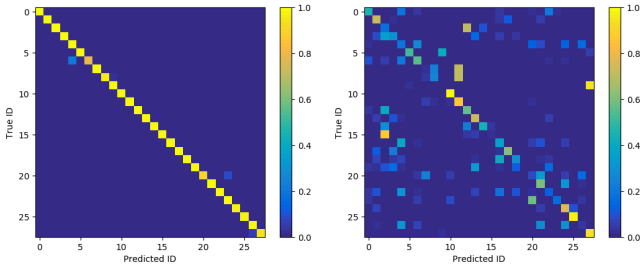
where (x, y) is the Cartesian coordinate of the true fingerprinting location for the training sample.

C. IMPLEMENTATION

The proposed neural network is implemented in Chainer 7 with python 3.7. A MacBook Pro 2016 with 2.9 GHz i7-6920HQ processor and 16 GB memory is used for data analysis. For optimization, adaptive momentum (Adam) stochastic gradient descent method is used with a learning rate of 0.001 and a mini-batch size of 100. The maximum number of epochs is 500 while early stopping with a patience of 10 is used. Training the DNN architecture takes about 1.03 seconds per epoch on the laptop computer.

D. COMPUTATIONAL COMPLEXITY

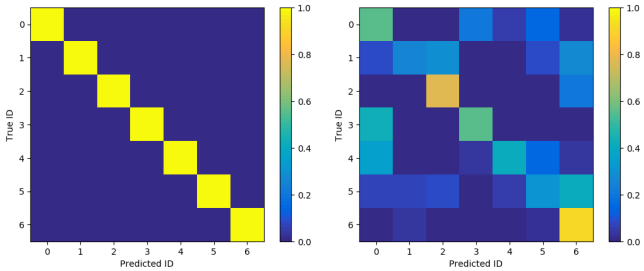
Now we analyze the computational complexity of the proposed neural network during the test phase. As seen from Fig. 5, the main building block is the FC layer. For each FC layer with an input dimension of N_{in} and an output dimension of N_{out}, the forward procedure mainly consists of two components: the matrix multiplication between the input vector and the $N_{\text{out}} \times N_{\text{in}}$ weight matrix and the addition of the bias vector of $N_{\text{out}} \times 1$, which gives a total of about $2N_{\text{in}}N_{\text{out}}$ floating-point operations (FLOPs). For each residual block, it consists of two FC layers, ReLU layers and a skip connection. In general, the ReLU activation layer can be implemented with bit-wise AND operations per dimension, while the skip connection introduces single addition per dimension. Therefore, each residual block has about $(4N_w + 3)N_w$ FLOPs. Adding the FC-based input and output layers, the overall computational complexity is about $N_d(4N_w + 3)N_w + 2MPN_w + 2NN_w$ FLOPs, where MP is the input dimension, and N_d and N_w denote the number of residual blocks and the number of nodes per hidden layer, respectively. In the case of simultaneous location and orientation classification of $N = 28$ with $N_w = 100$, we have about 67.5 thousand FLOPs for $N_d = 1$ residual block and, respectively, 148.1 thousand FLOPs for $N_d = 3$ residual blocks.



(a) Beam SNR: 98.9% Acc. $F_1 = 0.99$

(b) Conv. SNR: 44.5% Acc. $F_1 = 0.31$

FIGURE 7: Confusion matrices for simultaneous location-and-orientation classification using the proposed DL approach.



(a) Beam SNR: 100% Acc. $F_1 = 1.00$

(b) Conv. SNR: 56.6% Acc. $F_1 = 0.54$

FIGURE 8: Confusion matrices for location-only classification using the proposed DL approach.

V. PERFORMANCE EVALUATION

A. EXPERIMENT SETUP

The data collection system is deployed in an office environment during office hours, as shown in Fig. 6. There are 6 offices on both sides and 8 cubicles in the middle. All 6 offices and 4 cubicles on the right are occupied by staff members. Furniture including chairs, tables, and desktops are present in the cubicles.

These 3 APs, denoted as red triangles, are fixed in the aisle with fixed orientations. Specifically, AP1, AP2 and AP3 point to $90°$, $180°$ and $0°$, respectively, where the orientation reference is marked out in Fig. 6. To collect fingerprinting training data, we location a client at one of 7 locations-of-interest marked by crosses in Fig. 6. At each of the 7 locations, we collect beam SNR measurements by rotating the client to 4 orientations at $[0°, 90°, 180°, 270°]$. Overall, the offline training dataset consists of beam SNR measurements from $L = 7$ locations and $O = 4$ orientations[1]. The number of labeled training data for each location and orientation is listed in Table 1.

[1]The in-house mmWave Beam SNR Fingerprinting (mmBSF) dataset will be released at https://www.merl.com/demos/mmBSF.

TABLE 2: Average probability of successful classification for location and orientation identification with different methods

	Location & Orientation		Location-Only	
	Single SNR	Beam SNRs	Single SNR	Beam SNRs
LDA	41.1%	96.2%	47.7%	**100%**
QDA	41.0%	92.3%	54.2%	**100%**
SVM	39.9%	96.7%	54.1%	99.7%
DT	31.3%	72.1%	46.1%	91.7%
1NN	33.2%	98.5%	46.7%	99.9%
3NN	35.6%	98.6%	48.1%	99.8%
DL	**44.5%**	**98.9%**	**56.6%**	**100%**

TABLE 3: Average probability of successful classification as a function of training size R for the proposed DL method

No. of Training (R)	Loc. & Orn.	Loc. Only
10	94.1%	99.8%
20	96.8%	99.9%
50	98.4%	100%
100	97.9%	100%
original size	98.9%	100%

B. PERFORMANCE OF CLASSIFICATION

We first present our results on the location and orientation classification for our mmWave beam SNR fingerprinting-based localization system. For this purpose, we use the confusion matrix C as a performance visualization:

$$C(i,j) = \frac{1}{T_j} \sum_{t=1}^{T_j} \mathbb{1}[\hat{l}(\tilde{\mathbf{h}}_t(j)) = i], \tag{6}$$

where i and j are indices, respectively, for the estimated and true locations/orientations, and T_j is the number of sample in the test dataset for the index j. In addition, $\hat{l}(\tilde{\mathbf{h}}_t(j))$ is the location/orientation estimate by using the tth sample batch from the test data collected at jth location/orientation.

We first evaluate the localization performance of the proposed DL approach with $N_w = 100$ and $N_d = 1$, i.e., one residual block, for both location and orientation determination. Fig. 7(a) shows the confusion matrix using the proposed approach using the beam SNR measurements. The indices are arranged as $\ell = (l-1) \times 4 + (o-1)$ where $l \in \{1, \ldots, 7\}$ is the location index and $o \in \{1, \ldots, 4\}$ is the orientation index. It is seen from Fig. 7 that the proposed DL approach is able to localize both location and orientation with high probability. The probability of successful classification is 98.96% on average. The averaged F_1 score (harmonic mean of precision and recall) is also present in the figure captions for reference. When only the location is interested, corresponding confusion matrix is shown in Fig. 8(a). The results show that the DL approach with the beam SNRs can achieve an accuracy of 100%.

1) Beam SNRs versus Conventional SNR

To illustrate advantages using beam SNRs, we compare the performance with the traditional fingerprint-based approach with only one SNR measurement (or RSSI) available at each AP. For this purpose, we extract only one SNR measurement (from the highest average SNR) from all $M = 36$ beam SNRs at each AP and, therefore, the fingerprinting training data are R realizations of the RSSI-like single SNR values at each location and orientation. We apply the DL approach with the same architecture except that the input dimension is now 3. Corresponding confusion matrices are shown in Fig. 7(b) for the simultaneous location-and-orientation classification and Fig. 8(b) for location-only classification. This comparison clearly shows significant performance gains from conventional RSSI-like measurements to beam SNRs that carry richer spatial channel information.

2) Impact of Classification Methods

Next, we confirm that the proposed DL approach yields better performance over several classical machine learning methods, such as linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), SVM, decision tree (DT), and kNN. The results are shown in Table 5. Overall, we have the following observations:

- Classification using beam SNRs significantly improves the accuracy compared to the cases using single SNR.
- With beam SNRs, all classification methods except the DT show excellent performance with a nearly 100% accuracy.
- Our DL method consistently outperforms other considered machine learning methods.

We further remark a comparison in terms of computational complexity between the DL method and the simple kNN method. As analyzed in Section IV-D, once the training is done, the computation complexity of the proposed DL method depends on the dimension but not on the number of training samples. In contrast, the complexity of the kNN is a function of the number of labeled training samples in the fingerprinting dataset (i.e., $R \cdot L \cdot O$), the dimension of fingerprinting samples (i.e., MP), and the value of k. In the most naïve implementation, e.g., calculating each Euclidean distance and identifying the labeled samples closest to the test sample, the kNN method has a computational complexity of $\mathcal{O}(RLO(k+MP))$ although a further complexity reduction can be achieved [68] by using partial distance, editing, and prototype pruning. In our case, the total number of training samples is $RLO = 12{,}007$ in Table 1, and the dimension of the sample is $MP = 108$. Hence, the DNN method can be simpler than the naïve kNN implementation.

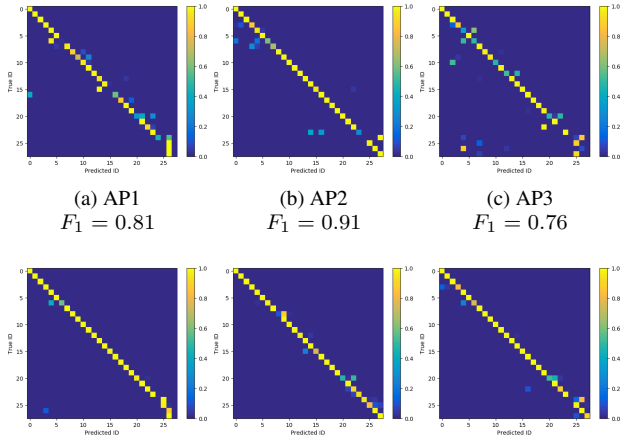
In the following, we focus on evaluating its performance as a function of training data size, sliding-window size, the number of APs, and orientation mismatch, as the DL method achieves the best performance.

3) Impact of Training Data Size

In the above performance evaluations, all training data listed in Table 1 were used for training the proposed neural network. To evaluate the impact of the number of training data on the localization performance, we truncate the original training dataset to smaller datasets with $R = \{10, 20, 50, 100\}$ beam SNR snapshots in each location and

TABLE 4: Average probability of successful classification as a function of window size Q

Window Size (Q)	Loc. & Orn.	Loc. Only
1	98.9%	100%
2	98.7%	100%
5	99.6%	100%
10	99.2%	100%



(a) AP1 $F_1 = 0.81$ (b) AP2 $F_1 = 0.91$ (c) AP3 $F_1 = 0.76$

(d) AP1+AP2 $F_1 = 0.92$ (e) AP1+AP3 $F_1 = 0.90$ (f) AP2+AP3 $F_1 = 0.93$

FIGURE 9: Impact of the number of APs on the performance of the simultaneous location-and-orientation classification accuracy.

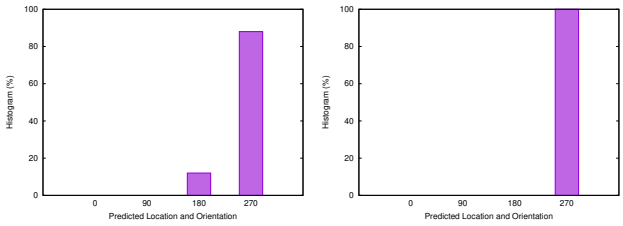
orientation. The average probabilities of successful classification are listed in Table 3. It is not surprising to see that the performance degrades as the number of training data reduces. Nevertheless, even in the case of only $R = 10$ fingerprinting beam SNRs, the average success probabilities are greater than 94% for the simultaneous location-and-orientation classification and maintain a nearly 100% for the location-only classification.

4) Impact of Window Size

Then we evaluate the localization performance as a function of window size Q, where Q denotes the number of consecutive packets used for the location and orientation classification. For each location and orientation, the beam SNRs fingerprint is now expanded to a $Q \times 108$ matrix. To keep the dimensionality constant regardless of the window size Q, we employed principal component analysis (PCA) before the feed to the proposed neural network. As listed in Table 4, the results reveal that the window size Q has a minor effect on the classification performance for both location-only and simultaneous location-and-orientation classification. In turn, this further confirms that the spatial feature (beam SNRs) is more dominant than the temporal feature (snapshots) in the fingerprinting training dataset.

TABLE 5: Average probability of successful localization for various combinations of APs

APs		Loc. & Orn.	Loc. Only
1 AP	AP1	84.3%	85.5%
	AP2	90.9%	99.6%
	AP3	80.2%	89.4%
2 APs	AP1 + AP2	94.9%	100%
	AP1 + AP3	94.5%	100%
	AP2 + AP3	93.8%	99.7%
All 3 APs	AP1 + AP2 + AP3	98.9%	100%



(a) The test orientation of 225° (b) The test orientation of 315°

FIGURE 10: Histogram of predicted orientations from simultaneous location-and-orientation classification on two test datasets with a $45°$ orientation mismatch at Location 5. The predicted location from the proposed DL approach is always Location 5 with a 100% accuracy.

5) Impact of APs

We now evaluate the impact of classification performance by changing the combinations of multiple APs. When only one AP is available, the result of confusion matrices are shown in the top row of Fig. 9. It shows that each AP has its own ambiguity region in terms of locations and orientations. For example, AP1 is hard to distinguish some orientations in Locations 2, 4 and 7, i.e., the 7th, 16th, 26th and 28th diagonal elements are missing, while AP2 shows several misclassifications at Location 7, i.e., the 25th diagonal element is missing. The average probabilities of successful classification are shown in Table 5, where the success probabilities of simultaneous location-and-orientation classification can still reach at 84.3%, 90.9% and 80.2% for AP1, AP2 and AP3, respectively.

With one more AP available (i.e., 2 APs), the ambiguity region is significantly reduced as seen from the bottom row of Fig. 9. This is particularly true for the combination of AP1 and AP2, where the average probability of successful classification jumps to 94.9%. When all three APs are available, the accuracy improves to 98.9% for the simultaneous location-and-orientation classification.

6) Sensitivity to Orientation Mismatch

Finally, we evaluate the sensitivity of the classification performance with respect to the orientation mismatch. To this end, we collect another independent test dataset at Location 5 with two additional orientations at $225°$ and $315°$ with a $45°$ orientation mismatch to their nearest fingerprinted orientations in the training dataset.

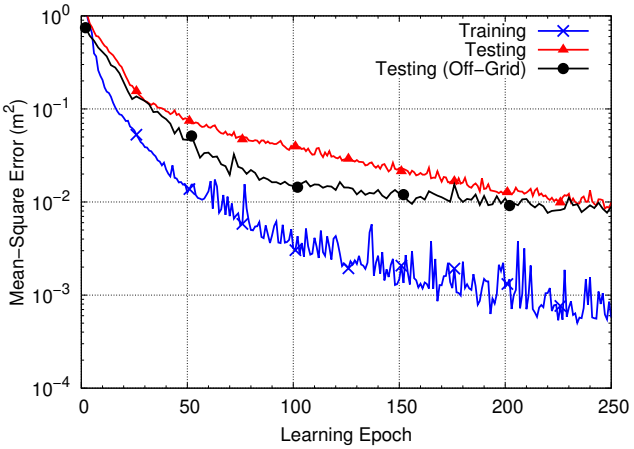


FIGURE 11: Learning trajectory in localization MSEs of the proposed DL method over epochs.

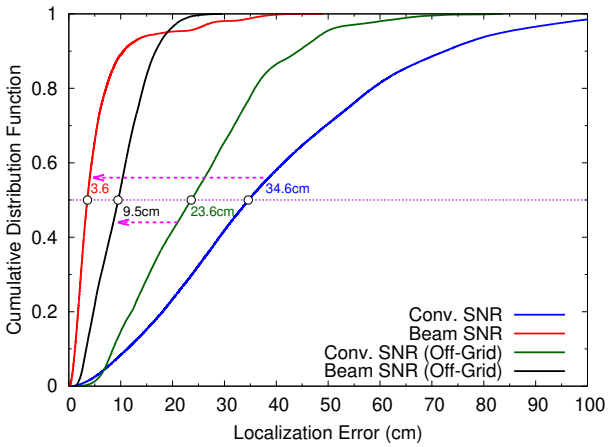


FIGURE 12: CDF curves of localization error for the proposed DL approach using beam SNRs and RSSI-like single SNR for 7 on-grid and 4 off-grid testing locations. The results were averaged over 20 time with different initializations.

TABLE 6: Configuration of APs, on-grid training and testing locations, and off-grid testing locations

AP	1	2	3				
x (cm)	11	306	0				
y (cm)	226	221	0				
On-grid location	1	2	3	4	5	6	7
x (cm)	53	122	67	48	88	110	142
y (cm)	133	167	188	62	62	99	233
Off-grid location	A	B	C	D			
x (cm)	83	99	67.5	103			
y (cm)	155	133.5	94.25	192			

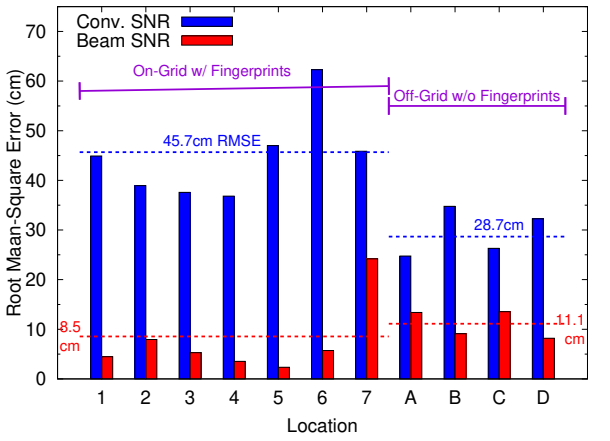


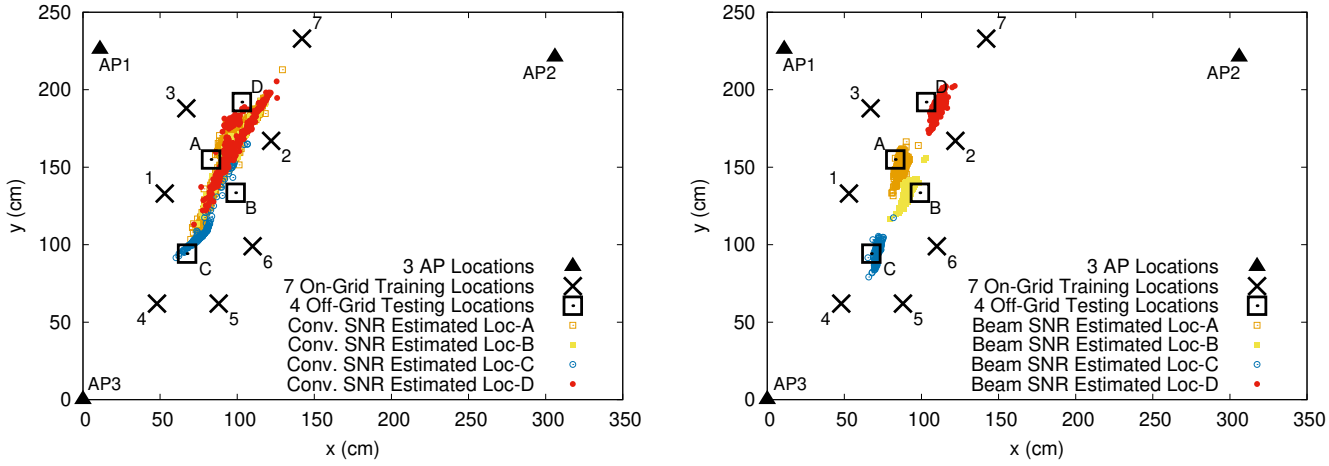
FIGURE 13: Location-wise RMSEs of coordinate estimation for the proposed DL approach using beam SNRs and RSSI-like single SNR at 7 on-grid and 4 off-grid testing locations.

For both mismatch cases, there is no compromise on the performance for the location classification. In other words, it maintains 100% accuracy to classify the location even if there is an orientation mismatch between the training and test datasets. Taking closer look at the test case of orientation 225° in Fig. 10(a), 88.0% out of the test samples are classified to the orientation 270° and the remaining 12.0% to the orientation 180°, two closest orientations included in the training dataset. Similarly, for the test case of orientation 315° at Location 5, the histogram of orientation classification Fig. 10(b) shows that all test samples are classified to 270°, again, one of two closest orientations in the training dataset. Overall, the results of Fig. 10 imply that 4 orientations for constructing the training fingerprint data may be sufficient to localize the client location even for the case when the true orientation of the test data is not included in the 4 orientations.

In a short summary, the above results on the classification performance confirm that the beam SNR measurements are able to register distinctive fingerprinting signatures for localization and orientation. Both classical classification (except the DT) and DL methods are able to a nearly 100% accuracy. In terms of computational complexity, the DL method can be faster than the naïve kNN implementation for testing, although it takes additional training time.

C. PERFORMANCE OF COORDINATE ESTIMATION

In this section, we directly predict the 2D coordinate of the client location by formulating it as a regression problem. Particularly, we consider a more practical scenario where an independent test dataset at 4 off-grid locations (denoted as A, B, C, and D in Figs. 14) was collected on a different date (four months later than the date of training data collection) during regular business hours. As shown in Table 6, these off-grid locations are not the same as the 7 fingerprinted locations (denoted as on-grid locations with labels $1, 2, \cdots, 7$ in Fig. 14) in the training dataset and the distance from each off-grid test location to its closest on-grid fingerprinted location is about the same and less than 70 cm to test the capability of sub-meter localization accuracy.



(a) Conv. SNR: 28.7 cm RMSE, 23.6 cm Median

(b) Beam SNR: 11.1 cm RMSE, 9.5 cm Median

FIGURE 14: Coordinate estimates at 4 off-grid testing locations (referred to as A,B,C,D in black squares).

1) Learning Trajectory

To predict the 2D coordinate, we attach an output layer of dimension 2 to the proposed neural network architecture of $N_w = 100$ and $N_d = 1$ in Fig. 5 with the MSE loss function. To achieve better generalizability for off-grid coordinate estimation, the dropout rate is increased from 0.1 to 0.8, and reduced the learning rate of Adam to $\alpha = 0.0001$. We applied a data augmentation technique based on a pairwise superposition with Gaussian noise injection to both beam SNR values and fingerprinted location coordinates with variances of 0.5 dB^2 and 0.02 m^2, respectively.

Fig. 11 presents the MSE trajectories of 2D coordinate estimation as a function of epochs for both training and testing. One can see that the training MSE (blue curves) at 7 on-grid locations rapidly decreases from 1m^2 to 0.001m^2, while the testing MSE (red curves) at the same 7 on-grid locations but in a different date exhibits a slower convergence and finally reaches to a level slightly below 0.01m^2 over 250 epochs. More importantly, the proposed DL approach can achieve the testing MSE at 4 off-grid locations at a level of 0.01m^2. It is worth noting that the testing MSE at the 4 off-grid locations is smaller than that at the 7 on-grid locations because the average distance among the off-grid locations is smaller than that of the fingerprinting locations and all the off-grid locations are inside the regions encompassed by the 7 on-grid locations.

2) Average Localization Performance

To evaluate the average localization performance, we trained the proposed neural network for 20 times starting with different initialization setups. Fig. 12 shows the averaged cumulative distribution function (CDF) of coordinate estimation error over the 4 off-grid locations. Compared with the RSSI-like single SNR fingerprinting, the proposed beam SNR fingerprinting along with the deep learning approach achieves significant improvements. Specifically, the averaged median root mean-square error (RMSE) is improved by an order of magnitude from 34.6 cm to 3.6 cm for the 7 on-grid testing locations. For the 4 off-grid testing locations, the averaged median RMSE of 9.5 cm by using the beam SNR is considerably better than that by using the RSSI-like single SNR with a median RMSE of 23.6 cm. The proposed DL-based approach also outperforms the classical machine learning method (i.e., GP) with a median RMSE of about 18 cm as reported in [32].

3) Location-Wise Localization Performance

Fig. 13 shows location-wise RMSE at the 7 on-grid locations and 4 off-grid locations. For the conventional RSSI-like single SNR fingerprinting at on-grid locations, the proposed DL approach achieves an RMSE of about 45.7 cm, where the best performance is obtained at Location 4 which is closest to AP3, whereas the worst performance is obtained at Location 6 possibly because it is relatively far from any APs. By using the beam SNR, one can achieve an RMSE of 3.6 cm, which is nearly 10-fold better than the single SNR-based fingerprinting. It is noted that the RMSE at Location 7 was exceptionally higher than those at the other 6 locations. This may be due to a few scattering paths for Location 7 to exploit spatial beam patterns as it is at the line-of-sight propagation between AP1 and AP2 and on the edge of fingerprinted coverage.

For the 4 off-grid testing locations, the single SNR fingerprinting shows an RMSE of 28.7 cm, while the beam SNR fingerprinting gives an RMSE of 11.1 cm. The sample distributions of coordinate estimates at the off-grid locations are shown in Fig. 14. It is clear to see that the single SNR fingerprinting-based coordinate estimates are scattered around the middle regions of fingerprinted locations and the beam SNR-based counterpart shows well-clustered coordinate estimates around corresponding true locations.

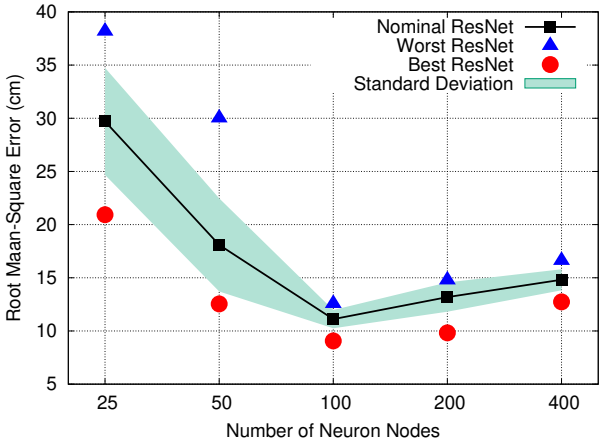


FIGURE 15: Nominal, worst and best RMSEs of coordinate estimation as a function of the number of nodes N_w when $N_d = 1$.

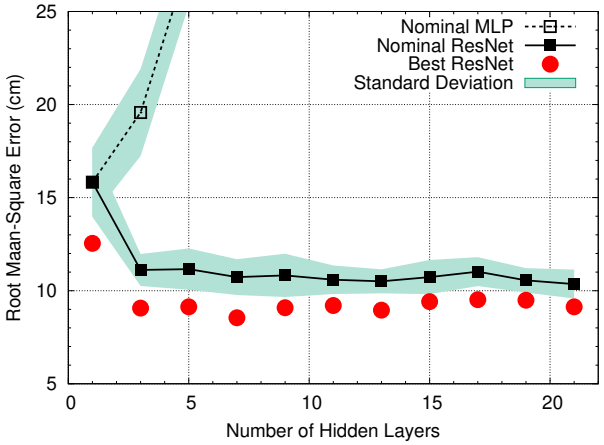


FIGURE 16: Nominal, worst and best RMSEs of coordinate estimation as a function of the network depth N_d when $N_w = 100$. The total number of hidden layers is given by $2N_d + 1$.

4) Impact of Neural Network Architecture

Finally, we show the impact of the neural network architecture in terms of the number of neuron nodes N_w of the hidden layers and the network depth N_d. Fig. 15 shows the nominal, best and worst RMSEs from 20 independently trained neural networks as a function of N_w when $N_d = 1$, i.e., there is only one residual block and three hidden layers (one input layer and two FC layers in the residual block) in total. It is seen that the RMSEs rapidly reduce when the number of node increases from $N_w = 25$ to $N_w = 100$ and then increase again when $N_w > 100$. When $N_w = 100$, the nominal RMSE is about 10 cm with the best performance can break into the centimeter-level accuracy, i.e., 8.6 cm.

Fig. 16 shows the nominal, best and worst RMSEs as a function of the network depth N_d when $N_w = 100$, i.e., the number of nodes is fixed to 100 for each layer. Given the structure of the residual block in Fig. 5, the total number of hidden layers is given by $2N_d + 1$ as each residual block contains two hidden layers plus the input layer. We also include the performance of a plain multilayer perceptron (MLP) that is identical to the proposed architecture in Fig. 5 but without shortcut connections. First, it can be verified from Fig. 16 that deeper networks with shortcut connections give slightly improved performance in terms of the nominal MSE. Second, concerning the best RMSE, the proposed architecture with at least one residual block, i.e., at least three hidden layers, can give a centimeter-level localization accuracy. Finally, the proposed architecture with shortcut connections can maintain the robustness against the network depth, while the RMSE of the plain MLP quickly explodes over the network depth.

In a short summary, it is noticeable that the proposed DL approach can achieve higher accuracy than the conventional machine learning methods for the direct coordinate estimation. For instance, the median RMSE is improved from 18 cm of the GP method to 9.5 cm of the proposed DL approach. The use of the beam SNR measurement over the RSSI-like single SNR measurement is also justified with about 2-fold improvements on the median RMSE.

VI. CONCLUSION

This paper has demonstrated that, by fingerprinting real-world beam SNRs from multiple COTS mmWave WiFi devices in our office environment, the proposed deep learning approach can identify the location and orientation of a client with high accuracy (100% accuracy if the location is only interested and about 99% for simultaneous location-and-orientations classification) and directly estimate the coordinate of a client with localization performance of 9.5 cm and 11.1 cm in term of the median and mean RMSEs, respectively. The localization performance was further evaluated as a function of various factors such as training data size, window size, the number of access points, orientation mismatch, and network width and depth.

REFERENCES

[1] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 37, no. 6, pp. 1067–1080, Nov 2007.

[2] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," IEEE Communications Surveys & Tutorials, vol. 11, no. 1, pp. 13–32, 2009.

[3] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single WiFi access point," in 13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16), Santa Clara, CA, Mar. 2016, pp. 165–178.

[4] C. Chen, Y. Chen, Y. Han, H. Lai, F. Zhang, and K. J. R. Liu, "Achieving centimeter-accuracy indoor localization on WiFi platforms: A multi-antenna approach," IEEE Internet of Things Journal, vol. 4, no. 1, pp. 122–134, Feb 2017.

[5] C. Chen, Y. Chen, Y. Han, H. Lai, and K. J. R. Liu, "Achieving centimeter-accuracy indoor localization on wifi platforms: A frequency hopping approach," IEEE Internet of Things Journal, vol. 4, no. 1, pp. 111–121, Feb 2017.

[6] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," IEEE Transactions on Vehicular Technology, vol. 66, no. 1, pp. 763–776, Jan 2017.

[7] X. Wang, L. Gao, and S. Mao, "Biloc: Bi-modal deep learning for indoor localization with commodity 5GHz WiFi," IEEE Access, vol. 5, pp. 4209–4220, 2017.

[8] C. Hsieh, J. Chen, and B. Nien, "Deep learning-based indoor localization using received signal strength and channel state information," IEEE Access, vol. 7, pp. 33256–33267, 2019.

[9] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "ConFi: Convolutional neural networks based indoor Wi-Fi localization using channel state information," IEEE Access, vol. 5, pp. 18066–18074, 2017.

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

[11] C. Xiang, Z. Zhang, S. Zhang, S. Xu, S. Cao, and V. Lau, "Robust sub-meter level indoor localization - a logistic regression approach," in ICC 2019 - 2019 IEEE International Conference on Communications (ICC), May 2019, pp. 1–6.

[12] C. Xiang, S. Zhang, S. Xu, X. Chen, S. Cao, G. C. Alexandropoulos, and V. K. N. Lau, "Robust sub-meter level indoor localization with a single WiFi access point?regression versus classification," IEEE Access, vol. 7, pp. 146309–146321, 2019.

[13] X. Li, "RSS-based location estimation with unknown pathloss model," IEEE Transactions on Wireless Communications, vol. 5, no. 12, pp. 3626–3633, December 2006.

[14] S. Mazuelas, A. Bahillo, R. M. Lorenzo, P. Fernandez, F. A. Lago, E. Garcia, J. Blas, and E. J. Abril, "Robust indoor positioning provided by real-time RSSI values in unmodified WLAN networks," IEEE Journal of Selected Topics in Signal Processing, vol. 3, no. 5, pp. 821–831, Oct 2009.

[15] M. Pajovic, P. Orlik, T. Koike-Akino, K. J. Kim, H. Aikawa, and T. Hori, "An unsupervised indoor localization method based on received signal strength (RSS) measurements," in 2015 IEEE Global Communications Conference (GLOBECOM), Dec 2015, pp. 1–6.

[16] C. Liu, D. Fang, Z. Yang, H. Jiang, X. Chen, W. Wang, T. Xing, and L. Cai, "RSS distribution-based passive localization and its application in sensor networks," IEEE Transactions on Wireless Communications, vol. 15, no. 4, pp. 2883–2895, April 2016.

[17] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in Proceedings IEEE INFOCOM 2000, March 2000, vol. 2, pp. 775–784.

[18] T. King, S. Kopf, T. Haenselmann, C. Lubberger, and W. Effelsberg, "COMPASS: A probabilistic indoor positioning system based on 802.11 and digital compasses," in Proc. First ACM Intl Workshop on Wireless Network Testbeds, Experimental evaluation and Characterization (WiNTECH), Sep. 2006.

[19] M. Youssef and A. Agrawala, "The horus location determination system," Wirel. Netw., vol. 14, no. 3, pp. 357–374, June 2008.

[20] M. Brunato and R. Battiti, "Statistical learning theory for location fingerprinting in wireless LANs," Computer Networks, vol. 47, no. 6, pp. 825 – 845, 2005.

[21] Z. Wu et. al., "Location estimation via support vector regression," IEEE Trans. on Mobile Computing, vol. 6, no. 3, pp. 311–321, March 2007.

[22] D. Li, B. Zhang, Z. Yao, and C. Li, "A feature scaling based k-nearest neighbor algorithm for indoor positioning system," in 2014 IEEE Global Communications Conference, Dec 2014, pp. 436–441.

[23] S. He and S. . G. Chan, "Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons," IEEE Communications Surveys Tutorials, vol. 18, no. 1, pp. 466–490, 2016.

[24] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, "Kernel-based positioning in wireless local area networks," IEEE Transactions on Mobile Computing, vol. 6, no. 6, pp. 689–705, June 2007.

[25] C. Figuera, I. Mora-Jimenez, A. Guerrero-Curieses, J. Rojo-Alvarez, E. Everss, M. Wilby, and J. Ramos-Lopez, "Nonparametric model comparison and uncertainty evaluation for signal strength indoor location," IEEE Transactions on Mobile Computing, vol. 8, no. 9, pp. 1250–1264, Sep. 2009.

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

[27] X. Lu, H. Zou, H. Zhou, L. Xie, and G. Huang, "Robust extreme learning machine with its application to indoor positioning," IEEE Transactions on Cybernetics, vol. 46, no. 1, pp. 194–205, Jan 2016.

[28] H. Dai, W. Ying, and J. Xu, "Multi-layer neural network for received signal strength-based indoor localisation," IET Communications, vol. 10, no. 6, pp. 717–723, 2016.

[29] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Reddy, "Recurrent neural networks for accurate RSSI indoor localization," IEEE Internet of Things Journal, vol. 6, no. 6, pp. 10639–10651, Dec 2019.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016, pp. 770–778.

[31] M. Pajovic, P. Wang, T. Koike-Akino, H. Sun, and P. V. Orlik, "Fingerprinting-based indoor localization with commercial MMWave WiFi–Part I: RSS and Beam Indices," in 2019 IEEE Global Communications Conference (GLOBECOM), Dec. 2019.

[32] P. Wang, M. Pajovic, T. Koike-Akino, H. Sun, and P.V. Orlik, "Fingerprinting-based indoor localization with commercial mmwave WiFi–Part II: Spatial beam SNRs," in 2019 IEEE Global Communications Conference (GLOBECOM), Dec 2019.

[33] D. Steinmetzer, D. Wegemer, M. Schulz, J. Widmer, and M. Hollick, "Compressive millimeter-wave sector selection in off-the-shelf IEEE 802.11ad devices," in CoNEXT 2017, December 2017.

[34] G. Bielsa et. al., "Indoor localization using commercial off-the-shelf 60 GHz access points," in IEEE INFOCOM 2018, April 2018, pp. 2384–2392.

[35] S. K. Saha et. al., "Fast and infuriating: Performance and pitfalls of 60 GHz WLANs based on consumer-grade hardware," in SECON'18, June 2018.

[36] "IEEE standards 802.11ad-2012: Enhancements for very high throughput in the 60 GHz band," in IEEE Standards Association, 2012.

[37] "IEEE standards 802.15.3c-2009: Millimeter-wave-based alternate physical layer extension," in IEEE Standards Association, 2009.

[38] Z. Wei, Y. Zhao, X. Liu, and Z. Feng, "DoA-LF: A location fingerprint positioning algorithm with millimeter-wave," IEEE Access, vol. 5, pp. 22678–22688, 2017.

[39] J. Gante, G. Falcão, and L. Sousa, "Beamformed fingerprint learning for accurate millimeter wave positioning," in 2018 VTC Fall, Aug 2018.

[40] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "Position and orientation estimation through millimeter-wave MIMO in 5G systems," IEEE Transactions on Wireless Communications, vol. 17, no. 3, pp. 1822–1835, March 2018.

[41] N. Garcia, H. Wymeersch, E. G. Larsson, A. M. Haimovich, and M. Coulon, "Direct localization for massive MIMO," IEEE Transactions on Signal Processing, vol. 65, no. 10, pp. 2475–2487, May 2017.

[42] Z. Lin, T. Lv, and P. T. Mathiopoulos, "3-D indoor positioning for millimeter-wave massive MIMO systems," IEEE Transactions on Communications, vol. 66, no. 6, pp. 2472–2486, June 2018.

[43] F. Lemic, J. Martin, C. Yarp, D. Chan, V. Handziski, R. Brodersen, G. Fettweis, A. Wolisz, and J. Wawrzynek, "Localization as a feature of mmwave communication," in 2016 International Wireless Communications and Mobile Computing Conference (IWCMC), Sep. 2016, pp. 1033–1038.

[44] J. Chen, D. Steinmetzer, J. Classen, E. Knightly, and M. Hollick, "Pseudo lateration: Millimeter-wave localization using a single RF chain," in 2017 IEEE Wireless Communications and Networking Conference (WCNC), March 2017, pp. 1–6.

[45] J. Palacios, P. Casari, and J. Widmer, "JADE: Zero-knowledge device localization and environment mapping for millimeter wave systems," in IEEE INFOCOM 2017 - IEEE Conference on Computer Communications, May 2017, pp. 1–9.

[46] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, "Capturing the human figure through a wall," ACM Trans. Graph., vol. 34, no. 6, Oct. 2015.

[47] C.-Y. Hsu, Y. Liu, Z. Kabelac, R. Hristov, D. Katabi, and C. Liu, "Extracting gait velocity and stride length from surrounding radio signals," in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2017, pp. 2116–2126.

[48] M. Zhao, T. Li, M. A. Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018, pp. 7356–7365.

[49] M. Zhao, Y. Tian, H. Zhao, M. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, "RF-based 3D skeletons," in Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication, 2018, pp. 267–281.

[50] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, N. Trigoni, and A. Markham, "mID: Tracking and identifying people with millimeter wave radar," in 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS), May 2019, pp. 33–40.

[51] A. Singh, S. Sandha, L. Garcia, and M. Srivastava, "RadHAR: Human activity recognition from point clouds generated through a millimeter-wave radar," in Proceedings of the 3rd ACM Workshop on Millimeter-Wave Networks and Sensing Systems, 2019, pp. 51–56.
[52] X. Lu, H. Wen, H. Zou, H. Jiang, L. Xie, and N. Trigoni, "Robust occupancy inference with commodity WiFi," in 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Oct 2016, pp. 1–8.
[53] Y. Zeng, P. H. Pathak, and P. Mohapatra, "WiWho: WiFi-based person identification in smart spaces," in 2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), April 2016, pp. 1–12.
[54] D. Wu, D. Zhang, C. Xu, H. Wang, and X. Li, "Device-free WiFi human sensing: From pattern-based to model-based approaches," IEEE Communications Magazine, vol. 55, no. 10, pp. 91–97, Oct 2017.
[55] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "WiFi-based human identification via convex tensor shapelet learning," in Proceedings of the 2018 AAAI Conference on Artificial Intelligence, 2018, pp. 1711–1718.
[56] J. Yang, H. Zou, H. Jiang, and L. Xie, "Device-free occupant activity sensing using WiFi-enabled iot devices for smart homes," IEEE Internet of Things Journal, vol. 5, no. 5, pp. 3991–4002, Oct 2018.
[57] Y. Gu, T. Liu, J. Li, F. Ren, Z. Liu, X. Wang, and P. Li, "Emosense: Data-driven emotion sensing via off-the-shelf WiFi devices," in 2018 IEEE International Conference on Communications (ICC), May 2018, pp. 1–6.
[58] Y. Gu, X. Zhang, Z. Liu, and F. Ren, "Besense: Leveraging WiFi channel data and computational intelligence for behavior analysis," IEEE Computational Intelligence Magazine, vol. 14, no. 4, pp. 31–41, Nov 2019.
[59] Z. Wang, K. Jiang, Y. Hou, W. Dou, C. Zhang, Z. Huang, and Y. Guo, "A survey on human behavior recognition using channel state information," IEEE Access, vol. 7, pp. 155986–156024, 2019.
[60] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang, and J. Han, "Joint activity recognition and indoor localization with WiFi fingerprints," IEEE Access, vol. 7, pp. 80058–80068, 2019.
[61] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang, "Person-in-wifi: Fine-grained person perception using WiFi," in Proceedings of (ICCV) International Conference on Computer Vision, July 2019.
[62] L. Zhang, X. Ruan, and J. Wang, "WiVi: A ubiquitous violence detection system with commercial WiFi devices," IEEE Access, vol. 8, pp. 6662–6672, 2020.
[63] D. Steinmetzer, D. Wegemer, and M. Hollick, "Talon tools: The framework for practical IEEE 802.11ad research," in Available: https://seemoo.de/talon-tools/, 2018.
[64] M. Schulz, D. Wegemer, and M. Hollick, "Nexmon: The c-based firmware patching framework," in Available: https://nexmon.org, 2017.
[65] W. Zhao, S. Han, R. Q. Hu, W. Meng, and Z. Jia, "Crowdsourcing and multisource fusion-based fingerprint sensing in smartphone localization," IEEE Sensors Journal, vol. 18, no. 8, pp. 3236–3247, April 2018.
[66] Y. Jiang, B. Liu, Z. Wang, and X. Yi, "Start from scratch: A crowdsourcing-based data fusion approach to support location-aware applications," Sensors, vol. 19, no. 20, 2019.
[67] X. Liu, S. Aeron, V. Aggarwal, X. Wang, and M. Wu, "Adaptive sampling of RF fingerprints for fine-grained indoor localization," IEEE Transactions on Mobile Computing, vol. 15, no. 10, pp. 2411–2423, Oct 2016.
[68] R .O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, Wiley, New York, 2001.