TagFree Activity Identification with RFIDs

XIAOYI FAN, Simon Fraser University, Canada
WEI GONG, Simon Fraser University, Canada
JIANGCHUAN LIU*, Simon Fraser University, Canada

Human activity identification plays a critical role in many Internet-of-Things applications, which is typically achieved through attaching tracking devices, e.g., RFID tags, to human bodies. The attachment can be inconvenient and considered intrusive. A tag-free solution instead deploys stationary tags as references, and analyzes the backscattered signals that could be affected by human activities in close proximity. The information offered by today's RFID tags however are quite limited, and the typical raw data (RSSI and phase angles) are not necessarily good indicators of human activities (being either insensitive or unreliable as revealed by our realworld experiments). As such, existing tag-based activity identification solutions are far from being satisfactory, not to mention tag-free. It is also well known that the accuracy of the readings can be noticeably affected by multipath, which unfortunately is inevitable in an indoor environment and is complicated with multiple reference tags.

In this paper, we however argue that multipath indeed brings rich information that can be explored to identify fine-grained human activities. Our experiments suggest that both the backscattered signal power and angle are correlated with human activities, impacting multiple paths with different levels. We present `TagFree`, the first RFID-based device-free activity identification system by analyzing the multipath signals. Different from conventional solutions that directly rely on the unreliable raw data, TagFree gathers massive angle information as spectrum frames from multiple tags, and preprocesses them to extract key features. It then analyzes their patterns through a deep learning framework. Our TagFree is readily deployable using off-the-shelf RFID devices and a prototype has been implemented using a commercial `Impinj` reader. Our extensive experiments demonstrate the superiority of our TagFree on activity identification in multipath-rich environments.

CCS Concepts: • **Networks** → **Network mobility**; **Sensor networks**; • **Computer systems organization** → **Sensor networks**;

Additional Key Words and Phrases: RFID, Backscatter, Activity Identification, Deep Learning

ACM Reference Format:
Xiaoyi Fan, Wei Gong, and Jiangchuan Liu. 2018. TagFree Activity Identification with RFIDs. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 7 (March 2018), 23 pages. https://doi.org/10.1145/3191739

1 INTRODUCTION

With the widespread deployment of Internet-of-Things, human activity identification has become a key service in many IoT applications, such as healthcare and smart homes [1]. It has received significant

*This is the corresponding author

Authors' addresses: Xiaoyi Fan, Simon Fraser University, 8888 University Dr, Burnaby, BC, V5A 1S6, Canada, xiaoyif@sfu.ca; Wei Gong, Simon Fraser University, 8888 University Dr, Burnaby, BC, V5A 1S6, Canada, gongweig@sfu.ca; Jiangchuan Liu, Simon Fraser University, 8888 University Dr, Burnaby, BC, V5A 1S6, Canada, jcliu@cs.sfu.ca.

2474-9567/2018/3-ART7 $15.00
https://doi.org/10.1145/3191739

attention from both academia and industry, with diverse solutions based on radars [30], cameras [4], inertial sensors [3], etc. Among them, RFID (Radio Frequency Identification) is of particular interest given its low cost, light weight, small footprint, and batteryless-operation. There have been pioneer studies on *tag-based* solutions for human activity identification [22][25][6]. That is, an RFID tag is attached to the human body, and the activities are then captured by a tag reader. Recently, *tag-free* solutions have also been suggested [34][23]. Instead of attaching tags to human bodies, which sometimes can be inconvenient and considered intrusive, multiple stationary tags are deployed as references, whose readings are expected to be affected by human activities in close proximity. Through analyzing the backscattered signals from the reference tags, the activities can then be identified.

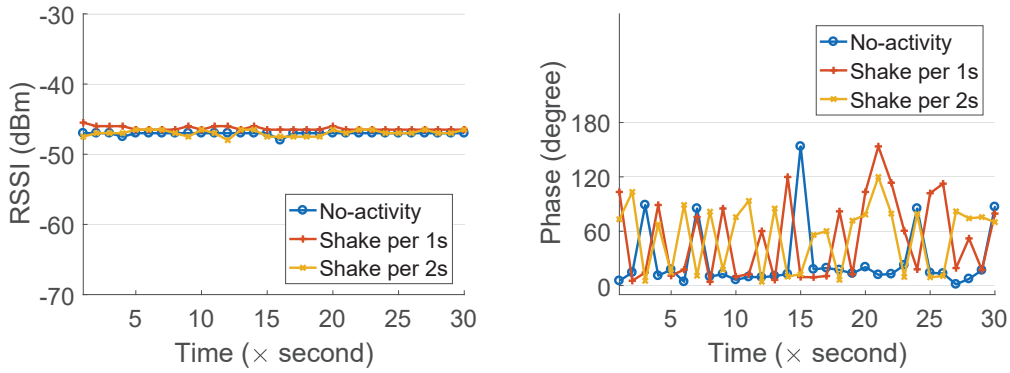
The information offered by today's RFID tags are quite limited, and the typical raw data, namely, *received signal strength indicator* (RSSI) and the *phase angle*, mostly target stationary reading scenarios. As such, existing tag-based activity identification solutions are far from being satisfactory, not to mention tag-free. Our realworld experiments have shown that the RSSI readings almost have no change with small human activities, e.g., shaking hands; The phase angle, though being sensitive to activities, is hardly a reliable indicator. It is also well known that the accuracy of the readings can be noticeably affected by multipath, which unfortunately is inevitable in an indoor environment.

In this paper, we however argue that mutipath indeed brings rich information that can be explored to identify human activities. Our experiments suggest that both the backscattered signal power and angle are highly related to human activities, impacting multiple paths with different levels. In a tag-free configuration with multiple reference tags, if we can capture these changing features of the paths, the activities could be identified with high sensitivity.

Inspired by these observations, we present `TagFree`, the first RFID-based device-free activity identification system by analyzing the multipath signals. Identifying the relevant features can be very time-consuming and complicated, so for defining the rules for accurate classification of activities. Different from conventional solutions that directly rely on the unreliable raw data, TagFree gathers massive angle information as spectrum frames from multiple tags, and preprocesses them to extract key features. It then analyzes their patterns through supervised learning. In particular, a deep learning framework with both Convolutional Neural Network (CNN) [15] and Long Short Term Memory (LSTM) network [13] is applied for common activity identification, which also scales up well to identify complex high-level activities (e.g., hour-long, day-long or more).

We conduct extensive experiments to evaluate our TagFree in multipath-rich environments and report significant performance gains over different state-of-art feature-based solutions. With an alert mechanism, TagFree can be customized as a monitor for senior patients in healthcare facilities, e.g., identifying dangerous events such as falling down on floors. Also, we envision that our TagFree can further facilitate various smart home applications, e.g., activity-based temperature adjustment in homes or exercise assistant equipment in gyms. It is also worth noting that TagFree is readily deployable using off-the-shelf RFID readers (a single UHF reader with a limited number of antennas) and allows reuse of existing RFID readers for indoor activity identification.

The rest of the paper is organized as follows. Section 2 presents the challenges in tag-free activity identification, and Section 3 illustrates the basic idea of our work. Section 4 provides our data pre-processing scheme on dealing with frequency hopping and de-coupling multipath signals, and then Section 5 presents our deep learning approach for activity identification. Section 6 discusses the implementation details. The performance evaluation results on our approach are presented in Section 7. Section 8 illustrates the background of the research area and provides a literature review. We then conclude this paper in Section 9.



(a) RSSI is insensitive

(b) Phase is sensitive but inaccurate

Fig. 1. The limitation of RSSI and Phase

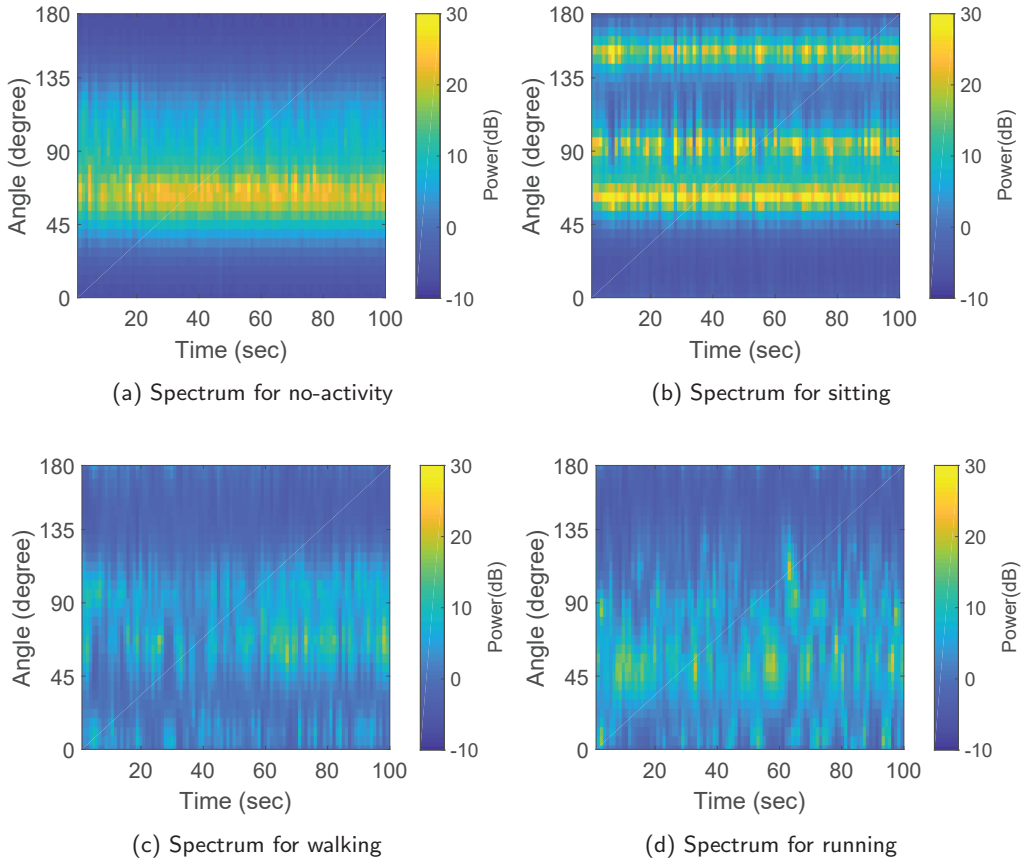
2 RAW RSSI OR PHASE? CHALLENGES FOR TAG-FREE ACTIVITY IDENTIFICATION

Today's commercial tag readers have very limited programming interfaces, which, through the standard Low Level Reader Protocol (LLRP)[1], report such low-level raw data as the *received signal strength indicator* (RSSI) and the *phase angle* only. The raw RSSI and phase data have been widely used for tag-based RFID applications. In a tag-free configuration with stationary RFID tags being deployed as references, e.g., on walls or furniture, the communication link established with fixed readers can be disturbed by human activities in close proximity, hence changing the RSSI or phase readings as well. For instance, TASA [34] measures the RSSI changes of signals received by the readers to infer human movement. Unfortunately, the activity information inferred from the raw RSSI can be quite unreliable and inaccurate for small movement. We have conducted an indoor experiment using off-the-shelf tags placed at a distance of 1 m facing polarized antennas. When shaking hands in front of the tag with different speeds (zero or no-activity, once per 1 sec, and once per 2 sec), we would expect the RSSI or phase readings be affected. Yet as shown in Fig. 1 (a), we observe an almost constant RSSI value, i.e., RSSI is insensitive to such small activities as handshaking. In contrast, the phase readings in Fig. 1 (b) do change when hands shake, which has the potentials to be explored.

It is worth noting that our experiments have fixed the channel on the 908.25 MHz. The U.S. government regulation requires frequency hopping to be enabled for RFID readers, with which the effectiveness of phase measurement can be noticeably affected [11][16]. The phase measurement is affected by multipath, which is inevitable in an indoor environment [25] [31]. As can be seen from Fig. 1 (b), with multipath, the changes in phase are often arbitrary, and the readings for "no-activity" in still swings from 1.05° to 157.8°, which can falsely identify activities.

Interestingly, we find that the dense multipath patterns indeed carry rich information about human activities. Specifically, in a tag-free configuration with multiple reference tags, if we can obtain a description of the paths along which the tag signals propagate, i.e., both the direction and the power of the backscattered signals, the human activity could be identified with high sensitivity. Our TagFree system deploys multiple reference tags to increase the density of multipath, and preprocesses the raw

[1]Thingmagic M6e reader.http://www.thingmagic.com/



(a) Spectrum for no-activity

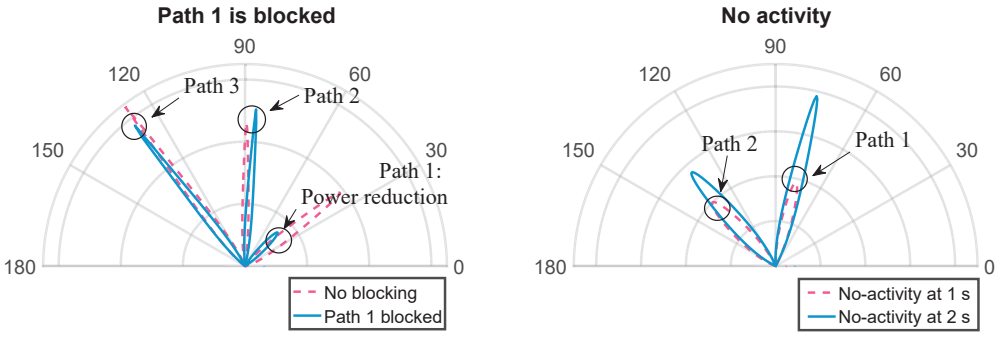
(b) Spectrum for sitting

(c) Spectrum for walking

(d) Spectrum for running

Fig. 2. Spectrum for different activities

phase through the MUSIC (Multiple Signal Classification) algorithm [20]. This algorithm has been widely used in the multi-path context for calculating the directions of arriving signals. Consider an illustrative example in Fig. 2. We place a stationary tag in front of a 4-antenna array with 2-meter distance and continuously collect tag readings with a Thingmagic reader, giving the corresponding spectrums for four human activities: no-activity, sitting, walking and running. The spectrum shows how the directions of arriving signals evolve over time, where high-amplitude angles are colored in yellow. Fig. 2 (a) shows the no-activity scenario and there is only one significant direct path in the environment. The activity of sitting is shown in Fig. 2 (b), where two more signal paths are produced with the sitting person, which are relatively stable. In the walking activity, a volunteer walks back and forth between the antenna array and the tag, where the multipath patterns in the spectrum become explicitly different with sitting as shown in Fig. 2 (c). Fig. 2 (d) shows the spectrum of running, which the signal peaks shift frequently. Using the signal multipath spectrum of the reference tags, we can quantify the correlation between the multipath signal patterns and a specific human activity.



(a) Spectrum for sitting

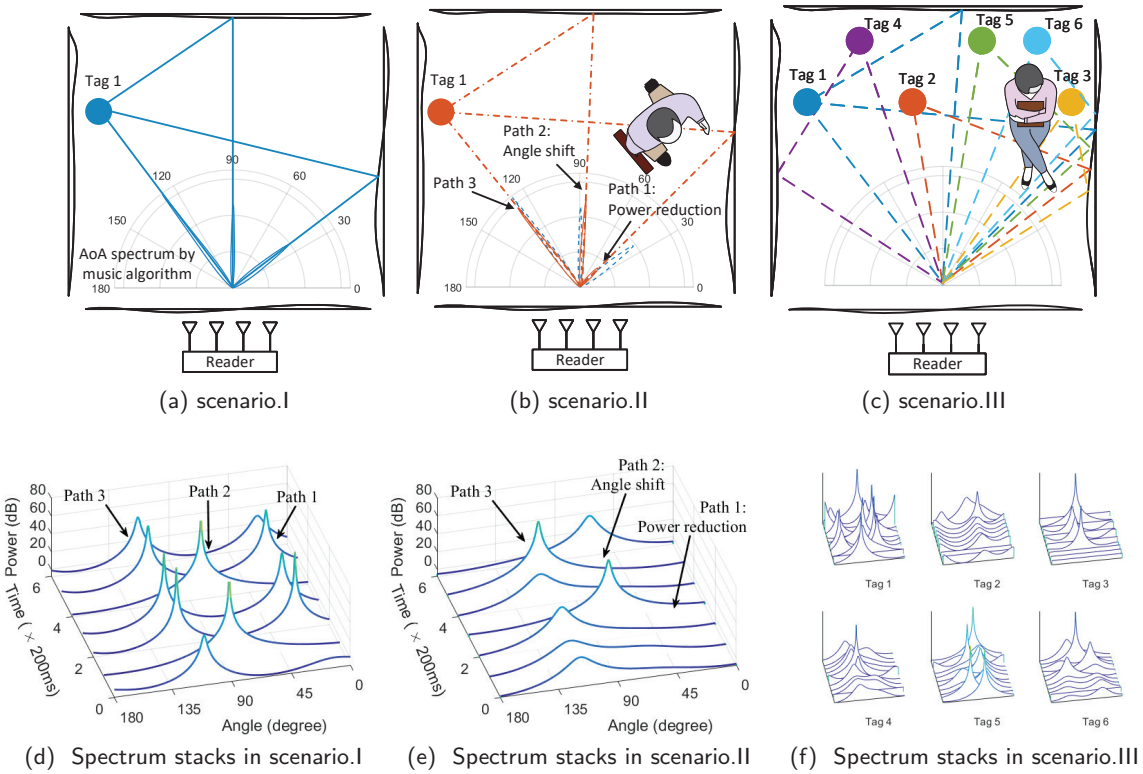
(b) Spectrum for no-activity

Fig. 3. TagFree Application Examples

We further have a detailed look into the experiments for sitting and no-activity, respectively. Fig. 3 (a) demonstrates the spectrums in 2 seconds for the sitting activity, where there exist three signal paths, namely, path 1, 2, and 3. When the person blocks path 1 at 40°, the power of path 1 is decreased, which successfully detects the activity along this directly affected path. Meanwhile, the angles and power of the other two paths are affected as well, though with different levels: path 2 at 90° is shifted to 85° with increased peak amplitude, and path 3 has little change. In short, both the signal power and angle are highly related to human activities, impacting multiple paths with different levels. However, the peak amplitude on an AoA spectrum estimated by MUSIC is a probability function [20]. As such, the peak amplitudes in the spectrum data do not necessarily represent the true signal power. In particular, with "no-activity", the power of a path should remain stable. Yet as Fig. 3 (b) shows, although the tag continuously reflects the signals with the same angle from 80° and 130°, the peak amplitudes of paths 1 and 2 have dramatically changed between the 1st and the 2nd second. The multiple signals also twist with each other and sometimes hide behind noises, so that the relationships to human activities cannot be easily identified. All these call for solutions to dynamically identify and extract intrinsic features from the massive spectrum data with high accuracy. We accordingly introduce a deep learning solution, which is not only effective in uncovering features for the common activities, but also scales up to identify more complex activities.

3 TAGFREE OVERVIEW

Fig. 4 shows the basic idea of our TagFree system. In Fig. 4 (a), the spectrum shows that there exist three paths from a stationary tag, where the stationary Tag 1 continuously reflects the signals with the same angle and power from 40°, 90° and 125°, respectively. Fig. 4 (b) shows a simple case that when one person blocks path 1 at 40°, where not just the peak of the blocked path is decreased, the other peak amplitudes and angle of other paths change as well. Fig. 4 (c) illustrates another case when there are many tags in the area. We can see that the number of signal paths increases rapidly with five more tags. Therefore, such rich and massive multi-path signal information provides an opportunity for activity identification with RFID tags.



(a) scenario.I (b) scenario.II (c) scenario.III

(d) Spectrum stacks in scenario.I (e) Spectrum stacks in scenario.II (f) Spectrum stacks in scenario.III

Fig. 4. Illustration of the basic idea of TagFree

To verify the above hypothesis, we conduct a series of field studies to examine the scenarios in Fig. 4 (a)-(c).[2] The results are plotted in Fig. 4 (d)-(f), which clearly demonstrate that the angle shift and power reduction are highly related to user activities: (i) In Fig. 4 (d), we place a stationary tag with a distance of 3 m facing the uniform linear array at the 125° direction (as shown in Fig. 4 (a)). We observe that there are three amplitude peaks in 1200 ms, which matches the "no-activity" situation, although the peaks are not stable. (ii) In Fig. 4 (e), we keep the tag at the same place and one person blocks the signal's propagation along path 1(as shown in Fig. 4 (b)). The peak of blocked path 1 experiences a clear drop and the angle of path 2 shifts frequently. (iii) In Fig. 4 (f), we place six tags in front of the antenna array (as shown in Fig. 4 (c)); these signals twist with each other and sometimes hide behind noises, making the patterns of the relationships between them and human activities hard to observe.

Recall the challenge we discussed in Fig. 3 (b), the peak amplitudes may dramatically change in a short time, which could be filtered out as noises for activity identification. Each individual spectrum frame forms only a small part of human activity. The traditional machine learning methods, e.g., support vector machine (SVM), would have to deal with incomplete information and suffer from much confusion in activity identification. Learning an activity description in the temporal spectrum data is important for activity

[2]For ease of exposition, here we disable frequency hopping, which will be further examined in Section 4.

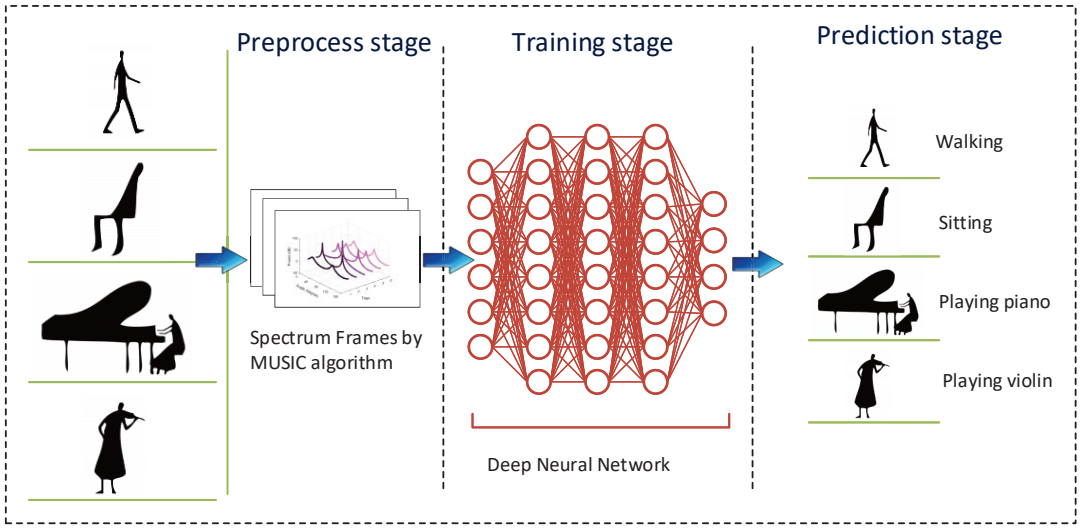


Fig. 5. TagFree framework

identification, since the temporal spectrum data provide more information to the activity identification. Therefore, we employ the recurrent neural network with Long Short Term Memory (LSTM) [13] units to discover long-range temporal relationships, which uses memory cells to store, modify, and access internal state. By sharing parameters through time, LSTM networks can learn how to integrate information and maintain a constant number of parameters while capturing an accurate activity description in the massive spectrum data.

Based on this basic idea, we propose the TagFree framework to identify the activities shown in Fig. 5. Our framework consists of three stages: preprocess stage, training stage, and activity identification stage. We first preprocess a large amount of tag phase readings based on the MUSIC algorithm and stack the spectrum frames in time series, serving as input for training and learning. It is followed by a deep learning architecture consisting of CNN and LSTM layer for model training and then activity identification.

4 DATA PREPROCESSING

Before we proceed with the detailed solutions for the individual modules, we first summarize the key notations in Table. 1.

4.1 Phase Calibration

Accurate AoA estimation is the key part of our activity identification system, which highly depends on the accurate phase measurements. To limit co-channel interference, FCC regulation requires that commercial UHF RFID readers must randomly hop across 50 channels within the 902-928 MHz band for every 400 ms. Such hopping inevitably causes phase offset due to the phase difference of oscillator and non-uniform frequency response of the tags' antennas. Our experiment and Tagyro [29] both verify that the frequency hopping introduces phase offset into the phase measurement.

We measure the phase of a stationary tag for 60 seconds, and plot the phase values versus frequencies in Fig. 6 (a), where the phase and frequency relation follows a linear model. These experiments imply

Table 1. Summary of Notations

d	distance between two antennas	x_t	input at time t
θ	spatial angle	$\boldsymbol{y}$	a set of labels
N	number of received signals	i_t	input gate at time t
f_i	frequency at channel i	f_t	forget gate at time t
$\phi^j(t)$	measured phase at frequency f_i	o_t	output gate at time t
$\phi(t)$	calibrated phase at time t	c_t	memory cell at time t
$s_i(t)$	a signal source i	h_t	hidden state at time t
M	number of source signals	y_t	output at time t
$\boldsymbol{s}(t)$	source signal vector (M by 1)	γ	activity cluster
$\boldsymbol{r}(t)$	received signal vector (N by 1)	$\boldsymbol{R_r}$	correlation matrix of received signals
$\boldsymbol{w}(t)$	noise vector (N by 1)	$\boldsymbol{R_s}$	correlation matrix of source signals
$\boldsymbol{a}(\theta)$	steering vector (N by 1)	$\boldsymbol{U_s}$	signal subspace
$\boldsymbol{A}$	matrix of steering vectors (N by M)	$\boldsymbol{U_n}$	noise subspace

that different frequencies induce different initial phase-offsets at the reader. We accordingly design a mechanism to calibrate the phase difference between frequencies, so that the phase output looks like coming from a fixed frequency. The calibration is done by collecting an initial phase measurement that takes about 10 seconds for the tag in stationary. We have frequency $f_j, \forall j \in [1, 50]$ and set a common frequency at a common frequency f_r (default to 908.25 MHz). Let $\phi^j(t)$ denote the measured phase at frequency f_j at time t, and $\tilde{\phi^j}$ represent the median value of measured phase at frequency j in recent 10 seconds. We map the measured phase ϕ^j at frequency f_j to the calibrated phase ϕ_i as follows:

$$\phi(t) = \phi^j(t) - \tilde{\phi^j} + \tilde{\phi^r} \tag{1}$$

4.2 Multipath De-coupling

Intuitively, as shown in Fig. 6 (b), the AoA estimation works as follows: A signal source $s = e^{j\phi}$ impinges on the array of N antennas with an angle θ, where ϕ denotes the phases of the received signal that we measure at the antennas. Let d be the distance between two antennas and λ be the wavelength of the received signal. Their relationship can be calculated as $\phi = \frac{2\pi}{\lambda} \cdot d \cdot cos\theta$. Assume the phase measurements at the first two antennas are ϕ_1 and $\phi 2$, we can then estimate the AoA θ as:

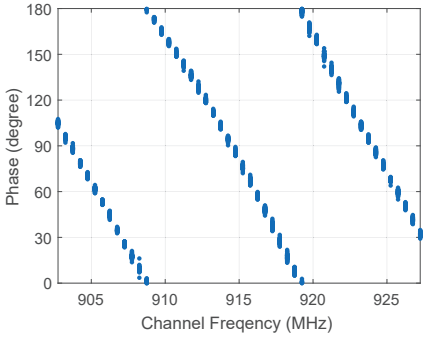
$$\theta = \arccos(\frac{|\phi_1 - \phi_2|}{\pi}) \tag{2}$$

However, in practice the AoA estimation may not work well because of the multipath effect. To this end, we adopt the MUSIC (MUltiple SIgnal Classification) algorithm[20], which is the best known algorithm based on eigenstructure analysis of an $N \times N$ correlation matrix $\boldsymbol{R_r}$. The entry at the l_{th} column and m_{th} row is the mean correlation between the l_{th} column and m_{th} antennas' signals. Consider M signals $s_1, \ldots, s_M(t)$ a from different angles $\theta_1, \ldots, \theta_M$, and into a uniform linear array (ULA) of N antennas. Then we have

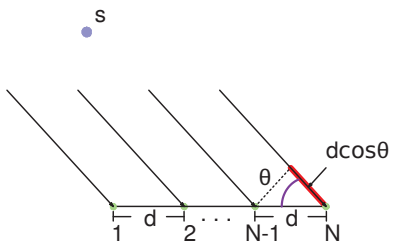
$$\boldsymbol{s}(t) = \left[s_1(t), \ldots, s_M(t)\right]^\top \tag{3}$$

Let $\boldsymbol{a}(\theta)$ be an $N \times 1$ vector which is the array steering vector for that direction. It is given by:

$$\boldsymbol{a}(\theta) = \left[1 \quad e^{-j\frac{2\pi d}{\lambda}\cos(\theta)} \quad e^{-j2\frac{2\pi d}{\lambda}\cos(\theta)} \quad \ldots \quad e^{-j(N-1)\frac{2\pi d}{\lambda}\cos(\theta)}\right]^\top \tag{4}$$



(a) Phase jumping caused by frequency hopping



(b) Based on the signal phase difference measured between a pair of antennas, we can estimate the spatial direction of the source's signal.

Fig. 6. Details in the data preprocessing stage

The $N \times 1$ received signal vector $\boldsymbol{r}(t)$ can be expressed as:

$$\boldsymbol{r}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{w}(t) \tag{5}$$

where $\boldsymbol{A} = \big[\boldsymbol{a}(\theta_1), \ldots, \boldsymbol{a}(\theta_M)\big]$ is an $N \times M$ matrix of the steering vectors, and $\boldsymbol{w}(t)$ is a noise term. Hence the array output consists of the signal plus noise components, where $\boldsymbol{r}(t)$ and $\boldsymbol{w}(t)$ are assumed to be uncorrelated and $\boldsymbol{w}(t)$ is modeled as temporally white and zero-mean complex Gaussian process. The spatial correlation matrix R of the observed signal vector $\boldsymbol{r}(t)$ can be defined as:

$$\boldsymbol{R_r} = E\{\boldsymbol{r}(t)\boldsymbol{r}^H(t)\} = \boldsymbol{A}\boldsymbol{R_s}\boldsymbol{A}^H + \mu^2\boldsymbol{I} \tag{6}$$

where $\boldsymbol{R_s} = E\{\boldsymbol{s}(t)\boldsymbol{s}^H(t)\}$, μ^2 is the noise covariance matrix, and $\boldsymbol{I}$ is an $N \times N$ matrix. The correlation matrix $\boldsymbol{R_r}$ has N eigenvalues associated with N eigenvectors $\boldsymbol{U} = [\boldsymbol{U}_1, \ldots, \boldsymbol{U}_N]$. The largest M eigenvalues correspond to the M incoming signals while the rest $N - M$ correspond to the noise. The corresponding eigenvectors in $\boldsymbol{U}$ can be classified into the signal subspace $\boldsymbol{U_s}$ and noise subspace $\boldsymbol{U_n}$:

$$[\boldsymbol{U_s}\boldsymbol{U_n}] = [\underbrace{\boldsymbol{U}_1, \ldots, \boldsymbol{U}_M}_{\boldsymbol{U_s}}, \underbrace{\boldsymbol{U}_{M+1}, \ldots, \boldsymbol{U}_N}_{\boldsymbol{U_n}}] \tag{7}$$

The MUSIC algorithm utilizes the orthogonality relationship between the signal and noise subspaces [20], which implies:

$$\boldsymbol{a}^H(\theta)\boldsymbol{U_n} = \boldsymbol{0} \tag{8}$$

The direction of the arrival angle can be represented in terms of the signal sources and the noise subspaces as shown in the following equation:

$$\theta_{MUSIC} = \arg\min_{\theta} \boldsymbol{a}^H(\theta)\boldsymbol{U_n}\boldsymbol{U_n}^H\boldsymbol{a}(\theta) \tag{9}$$

The above equation can be represented in terms of its reciprocal to obtain the peaks in a spectral estimations:

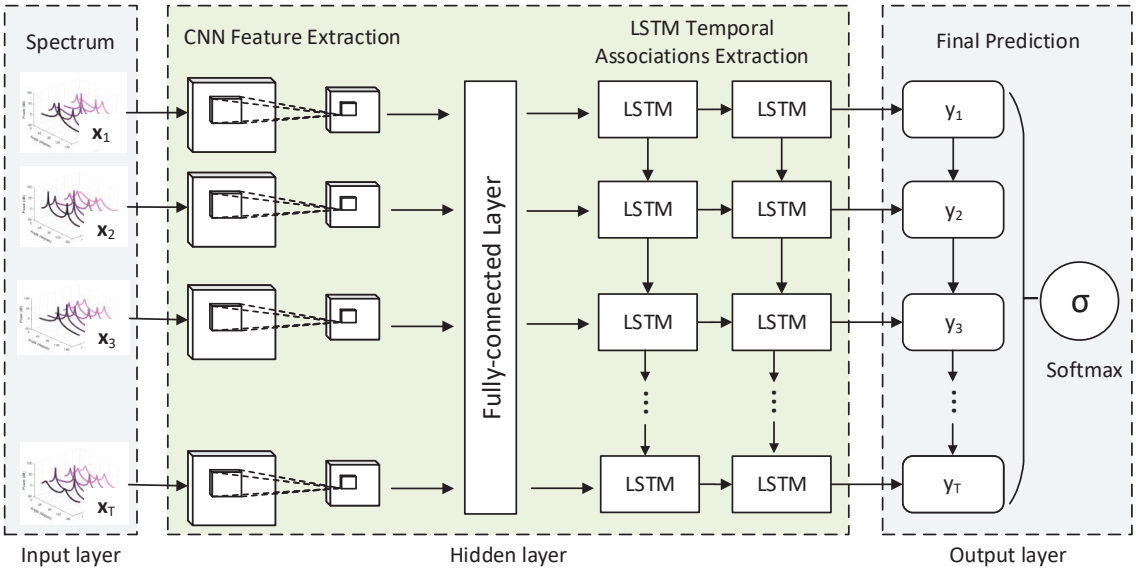


Fig. 7. Tag-free architecture

$$P_{MUSIC} = \frac{1}{\boldsymbol{a}^H(\theta)\boldsymbol{U_n}\boldsymbol{U_n}^H\boldsymbol{a}(\theta)} \tag{10}$$

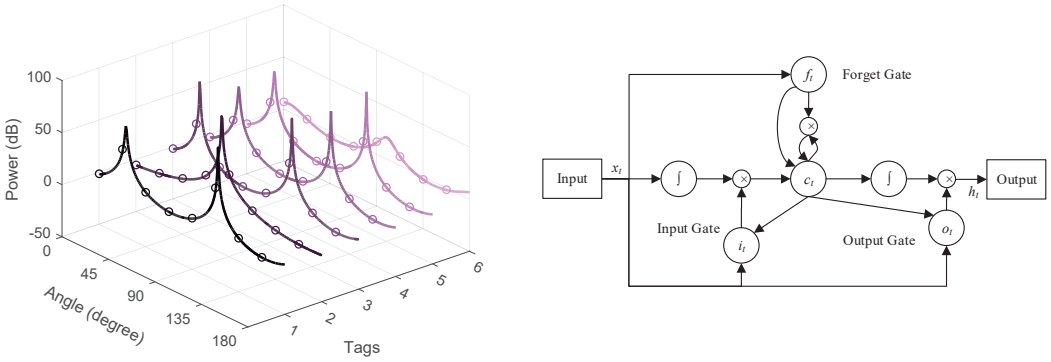
The M higher peaks are of great power [20] and corresponds to the estimated direction of arrival of the signal source with the angles $\theta_1, \ldots, \theta_M$.

5 DEEP LEARNING FOR ACTIVITY IDENTIFICATION

This section describes the main components of our TagFree design. As illustrated in Fig. 7, we construct a deep learning architecture, which is divided into the three main layers: input layer, hidden layer, and output layer. The hidden layer is further divided into a CNN (Convolutional Neural Network) sublayer and an LSTM (Long Short-Term Memory) sublayer, and is stacked as deep networks. We discuss each layer one by one in the following subsections.

5.1 Input Layer

This part starts from the design of our spectrum frames. The preprocessing stage outputs the spectrum for each tag, where we utilize the spectrum of all tags to build the spectrum frame, as Fig. 8 (a) shows. The size of a spectrum frame is $180 \times n$, where n is the number of tags and 180 is the number of angles. The input layer then takes all the spectrum outputs from our preprocessing stage and build the corresponding spectrum frames $\boldsymbol{x} = (x_1, \ldots, x_T)$, where a series of spectrum frames along the time will further serve as the initial input for the hidden layer.



(a) Spectrum Frame

(b) The architecture of LSTM cell

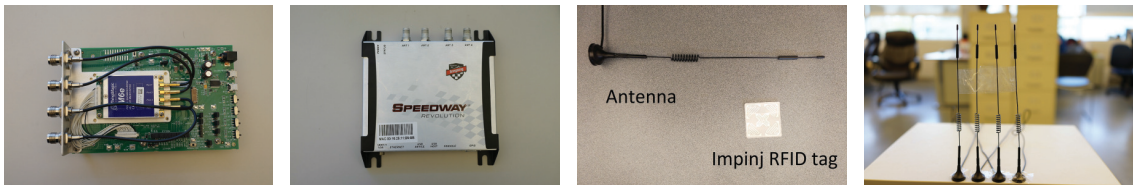
Fig. 8. Details in Deep Learning Stage

5.2 Hidden Layer

Our hidden layer includes a CNN (Convolutional neural network) structure for effective object classification and detection [15], and an LSTM (Long Short-Term Memory) structure for activity identification [13]. Such layers contain rich *implicit patterns* [2], e.g., object patterns and textures, and thus have been extensive used in Deep Learning networks where computer vision researchers have achieved tremendous successes in activity identification [12][14][7][33]. In our framework, we adopt a similar approach that extracts features at every spectrum frame, like image frame in videos, and stacks those features across time into a vector as the input of the LSTM structure. The parameters and settings of our deep learning network will be detailedly examined in Section. 7.

5.2.1 CNN Structure. The CNN takes the spectrum frames as input and provides the output to the LSTM structure. In this work, we report results using the output of the fully-connected layer, where these features are outputs of rectified linear units (RELUs) [8]. In particular, each input spectrum frame is $180 \times n$, where the features are extracted by the CNN. The extracted lower dimension features are then fed to a fully-connected layer and form the inputs to the The fully-connected layer has commonly been used to avoid overfitting [14], and we applied dropout in all the convolutional layers and the fully connected layer.

5.2.2 LSTM Structure. Fig. 8 (b) shows a single LSTM cell [13]. A stacked LSTM first encodes the frames one by one, taking the output of a fully-connected layer. An LSTM cell allows to easily memorize the context information for long periods of time in sequence data, which includes three gates: the input gate i_t, the forget gate f_t, and the output gate o_t, which have the controls to overwrite, keep, or retrieve the memory cell c_t, respectively. Each LSTM cell remembers a single floating point value c_t. This value may be diminished or erased through a multiplicative interaction with forget gate f_t or additively modified by the current input x_t multiplied by the activation of input gate i_t. The output gate o_t controls the emission of the memory value from the LSTM cell.



(a) Thingmagic M6e reader (b) ImpinJ R420 UHF reader (c) Antenna and tag (d) Antenna array

Fig. 9. Commercial hardware used to implement TagFree System

Let $\sigma(x) = (1 + e^{-x})^{-1}$ be the sigmoid function, which controls the inputs to a [0,1] range. We then have

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{11}$$
$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \tag{12}$$
$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \tag{13}$$
$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{14}$$
$$h_t = o_t \tanh(c_t) \tag{15}$$

where the W terms denote weight matrices (e.g. W_{hi} is the input-hidden weight matrix), and the b terms denote bias vectors (e.g. b_f is the bias vector of forget gate).

The LSTM cells are then grouped and organized into a deep LSTM architecture. Inside the architecture, the output from one LSTM layer will be the input for the next LSTM layer. We fine-tune the LSTM architecture with varying numbers of layers and memory cells, and chose to use two stacked LSTM layers, each with 32 memory cells. Following the LSTM layers, a softmax classifier is used to make a prediction at every spectrum frame.

5.3 Output Layer

The outputs from the last hidden layer is normalized with a softmax function, obtaining the probability distribution over the activity label y in the activity cluster γ:

$$Pr(y|x_i) = \frac{e^{x_i}}{\sum_{y' \in \gamma} e^{x_i}} \tag{16}$$

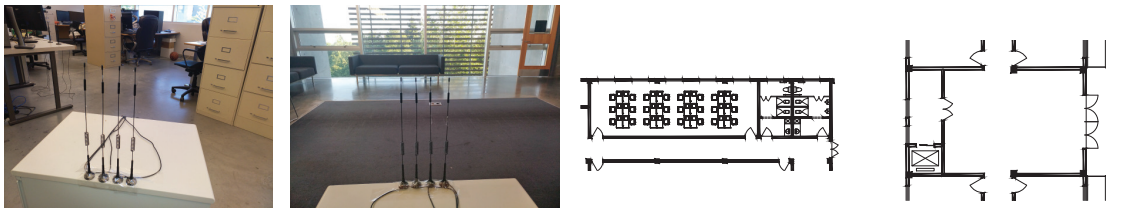
Our goal is to find the maximum likelihood of all training samples. We use the negative log probability, i.e., cross entropy error function, as the objective.

$$E = -\sum_{\gamma} z_y ln Pr(y|x_i) \tag{17}$$

where $z_y \in \{0, 1\}$. $Pr(y|x_i)$ is the predicted probability of class y.

6 SYSTEM IMPLEMENTATION

In this section, we describe the key implementation details that are not covered in the previous sections. Our implementation is entirely done with a commercial reader and requires no modifications to tags.



(a) Office environment (b) Hall environment (c) Office deployment layout (d) Hall deployment layout

Fig. 10. Two typical indoor environments and deployment layouts with the positions of arrays and tags

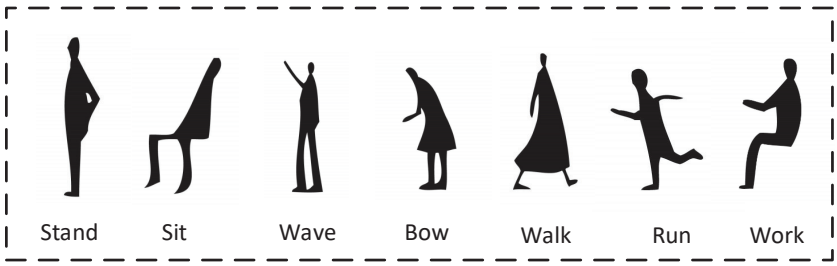


Fig. 11. Activities in experiments

RFID Readers: Our system works with today's commercial off-the-shelf readers, and our prototype implementation uses a Thingmagic reader[3] (Fig. 9 (a)), and an Impinj Speedway R420 reader[4] (Fig. 9 (b)) without any hardware or firmware modification. The Thingmagic M6e reader provides the APIs to customize the frequency hop table, thus we fix the channel on the common frequency 908.25 MHz. We use the fixed channel configuration of Thingmagic M6e reader to compare with Impinj R420 reader with frequency calibration, so as to test the performance of our phase calibration. The Impinj Speedway R420 reader has four antenna ports only. The reader is compatible with EPC Gen2 standard and the antennas work in a time division multiplexing mode. The FCC regulation requires that RFID readers perform frequency hopping in the range of 902-928 MHz, making phase measurements not accurate enough for activity classification. Due to the available frequencies of RFID reader, we set the common frequency $f = 908.25$ MHz, and the typical wavelength λ is 0.33 m.

Antennas Settings: We connect our Impinj Speedway R420 reader to four omni-directional antennas as shown in Fig. 9. An important setting is the distance between antennas, where we set d as $\lambda/8$ in our experiments for the following reasons:

- Theoretically, the antenna separation d should be spaced by $\lambda/2$, which effectively reduces the ambiguity caused by the high-resolution grating lobes [25];
- Since RFIDs communicate by backscattering the reader signal, the signal phase reading returned by the reader reflects the round trip distance instead of the one-way distance. Hence, d should be reduced to half, giving $d = \lambda/4$.

[3]http://www.thingmagic.com/
[4]https://support.impinj.com/

- Most readers can only measure phases within $[0, \pi]$, including the ThingMagic and ImpinJ readers. Although the ImpinJ reader can report phase readings ranging from to 0° to 360°, but it has π radians ambiguity, i.e., the reported phase can be the true phase (θ) or the true phase plus π radians ($\theta + \pi$). To account for such period π instead of 2π, the separation d is further reduced by half, i.e., $d = \lambda/8$.

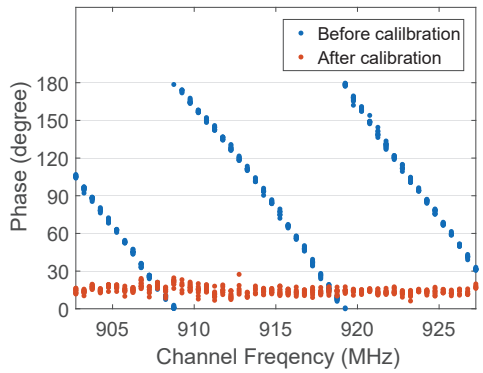
Deployment: One critical deployment issue is to determine the tag placement. We run experiments in two typical indoor environments: a multipath-rich office and an empty hall with little multipath effect. The office with a size of 13.75 m × 10.50 m has many file cabinets and writing desks, as shown in Fig. 10 (a). The file cabinet has a height of 1.32 m and is made of metal, resulting in rich multipaths and strong NLoS. The empty hall with a size of 8.75 m × 7.50 m is shown in Fig 10 (b). The deployment layouts of the environments are shown in Fig. 10 (c) and (d), respectively. In each environment, we deploy one readers and 6 tags. We place the antenna array at a height of 1.25 m. We use the basic ImpinJ tags as shown in Fig. 9 (c) and place them facing the antenna array with angles $45°, 60°, 75°, 90°, 105°, 120°$, respectively. The tags are usually placed on the furniture, so their heights are between 1 to 1.5 m above the ground and their distance to the reader is around 3 to 8 m. Indeed, the tags can be randomly placed without explicitly knowning their exact locations. Unless specifically mentioned, we use the default setup for performance evaluation.

Evaluation Metrics: We invite ten volunteers and each volunteer[5] stands between the tags and the antenna array in our experiments. To conduct a comprehensive evaluation, we test seven scenarios as shown in Fig. 11, including standing, sitting, waving, bowing, walking, running, and working. In each activity case, each volunteer is required to repeatedly perform the activity and the RFID reader continuously queries the RFID tags for 10 minutes. Each activity example contains all RFID readings in one minute period.

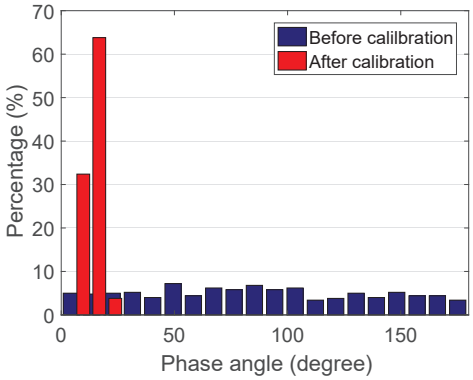
Server and algorithm implementation: The system employs a typical client-server architecture. We run the client on a Lenovo laptop (ThinkPad T560) connecting to the Impinj Speedway R420 reader, equipped with an Intel Core i5-6200U Dual 2.3/2.8GHz CPU and 8 GB 1333 MHz DDR3 RAM. The server runs on a customized PC with an Intel Core i7-6850K 3.60 GHz 12-core CPU and Dual Nvidia GTX 1080 Ti GPUs. The processes on clients adopt the LLRP protocol [18] to communicate with the reader, with Mercury API[6] and Octane SDK Java[7] for Impinj R420 reader. It requires the reader to continuously execute the tag reading operation, and upload tag readings to the backend modules for further analysis. Therefore, we utilize multi-threading, which returns immediately a sequence of RFID reads to the calling thread, which then uploads the tag readings to the server. The backend module on the server accepts and stores the streaming of tag readings as the training data in a database, and executes our algorithms to identify the activity.

CNN and LSTM classifiers are implemented in Keras[8] with cuDNN on Dual Nvidia GTX 1080 Ti GPUs, and the multiclass classifiers are implemented based on the Scikit-learn library [17]. We compared the performance of our deep learning system with ten mainstream classifiers: k-Nearest Neighbors, one-vs-all Linear SVM, one-vs-all RBF SVM, Gaussian Process, Decision Tree, Random Forest, Neural Net, Adaptive Boosting, Bayesian Net and Quadratic Discriminant Analysis. We treated activity identification as a multi-class classification problem and considered the detection of each activity as a binary classification problem.

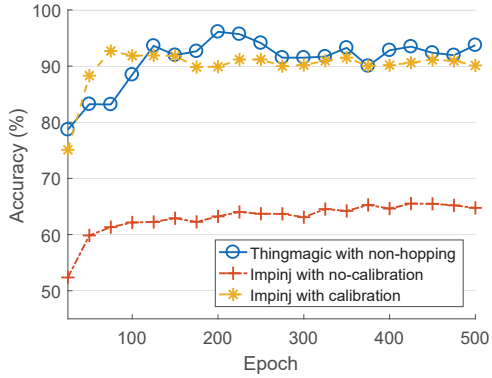
[5]Note that those volunteers cover a broad range in age, gender, height, and weight.
[6]Mercury API Programmer's Guide. https://www.thingmagic.com/
[7]Octane SDK Programmer's Guide. https://support.impinj.com/
[8]Keras: Deep Learning library for Theano and TensorFlow. https://keras.io/



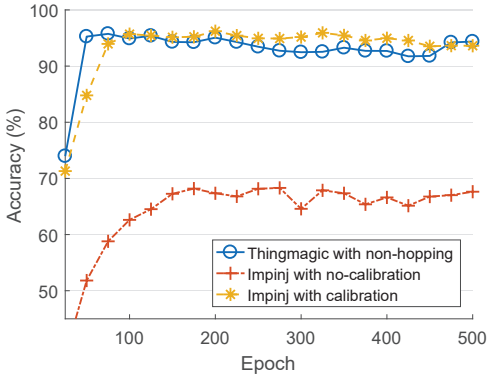
(a) The comparison between the raw phase and the calibrated phase



(b) Histogram of phase distribution for frequency hopping calibration



(c) In the office with rich-multipath



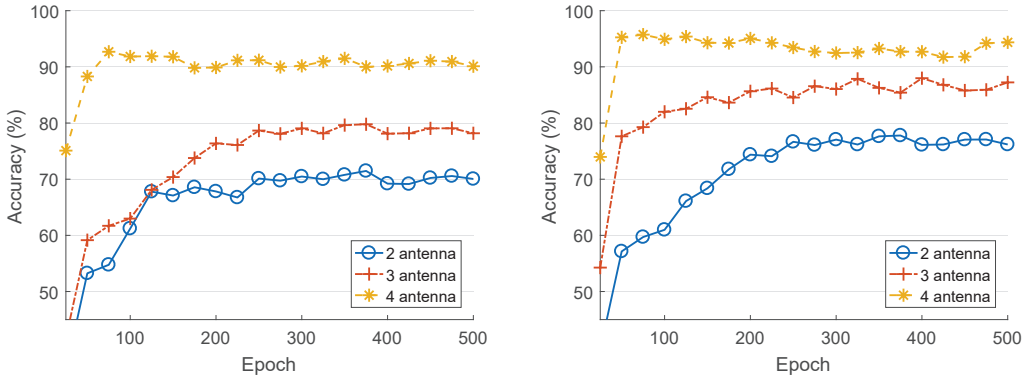
(d) In the hall with low-multipath

Fig. 12. Impact of Phase Calibration

7 EVALUATION AND DISCUSSION

7.1 Model Training

To evaluate the prediction quality, we run an experiment on the real-world data with 5690 activity examples, where 2845 examples come from multipath-rich environments and 2845 examples from environments with little multipath. We train the models for the two different scenarios with cross validation to mitigate overfitting, where 80% of the data is used as a training set and the remaining 20% is used as a test set. The training includes 500 epochs using stochastic gradient descent (SGD). To combat exploding gradients, we scale the norm of the gradient, and both hyperparameters are chosen using the training set. We implement the CNN networks with two convolutional layers with a dropout of 0.5, followed by one fully-connected layer. Our LSTM networks use 32 memory cells per layer. Throughout training, we save the model and compute prediction accuracy on the test set for each epoch.



(a) In the office with rich-multipath (b) In the hall with low-multipath

Fig. 13. Impact of Number of Antennas

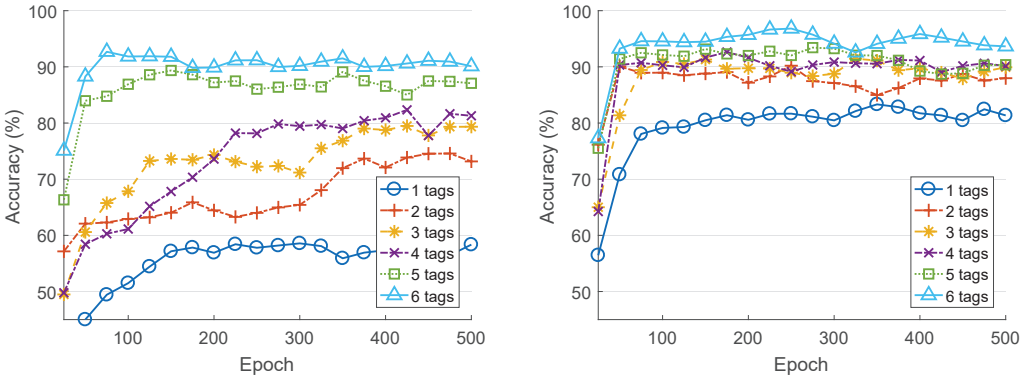
7.2 Impact of Phase Calibration

We first evaluate the impact of the reader's channel hopping on the phase value by measuring the phase of a stationary tag for 2 minutes. Fig. 12 (a) shows the result of the calibrated phase data with respect to Fig. 6 (a). It illustrates the phase offset causing by frequency jumping can be effectively eliminated by our method. Fig. 12 (b) plots the histogram of the phase values. Without calibration, the phases are almost evenly spread over all possible phase values, where our phase calibration method can reduce the standard deviation from 49.69° down to 2.91°.

It is worth noting that our phase calibration mechanism contributes toward improving the precision of activity identification, as shown in Fig. 12. We evaluate how phase calibration benefits TagFree in Fig. 12 (c) and (d), where the x-axis of the figures represents the model training epochs and y-axis represents the prediction accuracy. The results show that our calibration method achieves a high accuracy and minimizes the negative influence of frequency hopping. Specifically, the TagFree system in the multipath-rich office with phase calibration achieves an activity identification accuracy of 91%, against the accuracy of 64% without calibration, for our calibration mechanism achieves a high AoA estimation accuracy. In the low-multipath hall environment, TagFree with phase calibration can reach a 97% accuracy, against the accuracy of 68% without calibration.

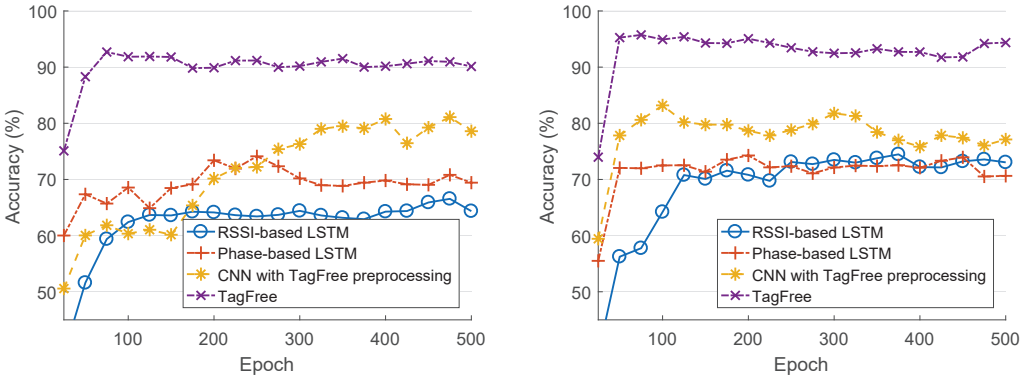
7.3 Impact of Different System Settings

TagFree de-couples the multipaths using an array of antennas. Therefore the number of antennas limits the number of multipaths that can be detected by our pre-processing scheme. With the information of more signal paths, TagFree achieves a higher multipath density in the area and improves the activity identification accuracy. We thus investigate the impact of the number of antennas in two typical environments, as shown in Fig. 13 (a) and (b), respectively. We can see that when the number of antennas increases from 2 to 4, more angle information of multipath can be detected, and thus TagFree can



(a) In the office with rich-multipath (b) In the hall with low-multipath

Fig. 14. Impact of Number of Tags



(a) In the office with rich-multipath (b) In the hall with low-multipath

Fig. 15. Impact of Deep Learning Approaches

achieve even higher activity identification accuracy.[9] The default setting of our remaining experiments is 4 antennas connecting to the reader.

With more tags, more signals will be reflected, creating more paths to cover the monitoring area and providing more information for activity identification. In the next experiments, we vary the number of tags from 1 to 6. Both Fig. 14 (a) and (b) show that more tags are helpful to provide more information and improve the activity identification accuracy. Since the number of multipath that our data pre-processing scheme can detect for each tag is limited by the number of antennas on the reader, using more tags indeed is a better alternative to increase the path diversity in the environment. In an indoor environment,

[9]Note that Impinj Speedway R420 reader has maximally four ports to connect with antennas.

Table 2. Comparison of performance with different physical configuration in the office with RICH-multipath

Height (m)	The distance from tags to antennas (m)					
	3	4	5	6	7	8
1.0	0.92	0.90	0.88	0.86	0.82	0.80
1.25	0.93	0.91	0.90	0.89	0.83	0.81
1.5	0.94	0.92	0.91	0.91	0.86	0.83

Table 3. Comparison of performance with different physical configuration in the hall with LOW-multipath

Height (m)	The distance from tags to antennas (m)					
	3	4	5	6	7	8
1.0	0.97	0.97	0.97	0.96	0.90	0.89
1.25	0.98	0.98	0.97	0.97	0.91	0.90
1.5	0.98	0.98	0.97	0.97	0.93	0.90

more tags only bring marginal increase to the system cost, which is considerably lower than increasing the number of antennas on readers [26]. To this end, we use a tag number of 6 as the default setting of our evaluation.

Tab. 2 and Tab. 3 show the details of activity identification accuracy with different physical configuration of our TagFree approach, where each row represents the tag heights and each column denotes the distance from tags to antennas. Also, the experiments are conducted in high and low multipath environments, as shown in Tab. 2 and Tab. 3 respectively. In the experiments, we evaluate the accuracy with varying tag heights ranging from 1.0 m to 1.5 m, where the results do not exhibit clear correlation with the tag heights. Thus, the tag height around 1.0 m to 1.5 m is not a crucial factor affecting the activity identification accuracy. We use 1.5 meter as the default tag height for the remained experiments. Then we further examine the activity identification accuracy when applying different distance (3 to 8 m) from tags to antennas. The results illustrate that the smaller the distance is, the greater the accuracy achieves. Beyond 6 m, the reduced reading rate of RFID tag makes the activity identification accuracy decrease, due to many effects, e.g., too weak signals. We use the distance from tags to antennas of 6 meter as the default setting of our evaluation, since smaller distances may make our design impractical. The default distance setting of 6 meter yields an average accuracy of 91% and 97% in high and low multipath environments, respectively.

We compare the results of our TagFree with various combinations of deep learning architectures as shown in Fig. 15 (a) and (b). First, we compare the performance TagFree with CNN networks, and both of them integrate the preprocessing scheme of TagFree. TagFree can achieve a 14% higher accuracy on average than the CNN networks, which demonstrates that the LSTM architecture is necessary for activity identification. Then, we evaluate TagFree against the phase-based LSTM and the RSSI-based LSTM networks. It clearly shows that TagFree can achieve a 22% higher accuracy than the RSSI-based LSTM networks, and a 20% higher accuracy than the phase-based LSTM networks. The results illustrates that our preprocessing scheme and CNN can efficiently extract the features for activity identification. In summary, the benefits of TagFree comes from both the preprocessing scheme and the deep learning, which work jointly to harvest the rich phase information for activity identification in a multipath environment.

As illustrated in Table. 4, using more convolutional layers generally achieves better performance, but there exists tradeoff between the performance gain and the computation resources. We have tested the CNN with 1 to 5 convolutional layers. The results in Table. 4 show little gain in precision, recall and F-Score when using more than two convolutional layers, and hence we use two layers in our implementation. With this default setting, we then vary the convolution kernel size, and the results in Table. 5 show that the size of 2×16 has the best performance. In Table. 6, we summarize the results for various number of LSTM layers. Again, though a single layer LSTM network is not effective, the 2-layer LSTM achieves sufficiently good performance. Table. 7 further suggests that increasing the number of LSTM cell per

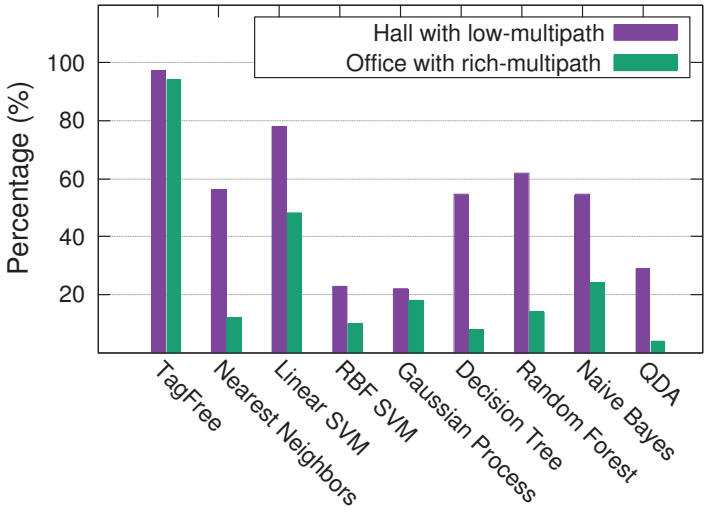


Fig. 16. Overall Performance of TagFree System

Table 4. Comparison of performance with different number of convolutional layers

	1	2	3	4	5
Precision	0.69	0.94	0.95	0.96	0.97
Recall	0.74	0.93	0.93	0.97	0.94
F-score	0.69	0.94	0.87	0.95	0.91

Table 5. Comparison of performance with different kernel size

	2×2	2×8	2×16	2×32	4×64
Precision	0.89	0.93	0.94	0.97	0.93
Recall	0.87	0.88	0.95	0.97	0.92
F-score	0.87	0.90	0.94	0.97	0.92

Table 6. Comparison of performance with different number of LSTM layers

	1	2	3	4	5
Precision	0.84	0.94	0.94	0.92	0.90
Recall	0.77	0.93	0.97	0.93	0.97
F-score	0.73	0.94	0.95	0.81	0.89

Table 7. Comparison of performance with different number of LSTM cells per layer

	1	2	3	4	5
Precision	0.84	0.94	0.94	0.92	0.90
Recall	0.77	0.93	0.97	0.93	0.97
F-score	0.73	0.94	0.95	0.81	0.89

layer beyond 32 cells achieves only marginal performance improvement, and thus we set the number of LSTM cell to 32.

7.4 Overall Activity Identification Performance

Fig. 16 shows the performance of our TagFree compared with 8 mainstream classifiers, including k-Nearest Neighbors, one-vs-all Linear SVM, one-vs-all RBF SVM, Gaussian Process, Decision Tree, Random Forest, Bayesian Net and Quadratic Discriminant Analysis, where all the classification methods take the same spectrum data as input. TagFree achieves a 94% accuracy of activity identification on average. In the low-multipath hall environment, we can see that our TagFree performs the best among all approaches

Table 8. Low multipath environments in hall

	Predicted activities						
	Stand	Sit	Wave	Bow	Walk	Run	Work
Stand	1.00	0	0	0	0	0	0
Sit	0	0.96	0.04	0	0	0	0
Wave	0	0.04	0.96	0	0	0	0
Bow	0	0	0	1.00	0	0	0
Walk	0	0	0	0	0.98	0	0.03
Run	0	0	0	0	0.02	0.98	0
Work	0	0	0	0	0	0.02	0.97

Table 9. Rich multipath environments in office

	Predicted activities						
	Stand	Sit	Wave	Bow	Walk	Run	Work
Stand	0.95	0.03	0	0	0	0	0
Sit	0.05	0.94	0.04	0	0	0	0
Wave	0	0	0.90	0	0	0	0
Bow	0	0.03	0.03	0.91	0.17	0.03	0
Walk	0	0	0.07	0.09	0.83	0.08	0.05
Run	0	0	0	0	0	0.89	0
Work	0	0	0	0	0	0	0.95

with an accuracy 97% on average, which is 20% better than the runner-up (SVM). The 4-antenna array maximally decouples the three signals paths for each tag, which should be easy to cope with the challenge of sparse multipath in the hall. Unfortunately, the classical machine learning methods, e.g., linear SVM, only have an accuracy 77% or lower, reaffirming the effectiveness of our deep learning scheme. The experiments in a multipath-rich office show that the performance of other classifiers degrades heavily and TagFree still maintains a high accuracy of 91% on average.

We break down the results of TagFree in Fig. 16 into Tables 8 and 9. The tables illustrate the details of the experiments in the hall and the office, respectively. Each row denotes the actual activity performed and each column represents the activity identified by TagFree. Each element in the matrix represents the percentage of activities in the row that is recognized as the activity in the column. As shown in the Tab. 8, the average accuracy is 97% in the hall with low multipath for all scenarios, where the identification of slow activities (stand, sit, wave, bow and work) has achieved an accuracy at least 94%. In the multipath-rich environments, TagFree can achieve a 91% accuracy on average, where the identification accuracy for fast moving activities still keeps above 83% . This indicates that TagFree can distinguish activities at different speeds with high accuracy.

7.5 Discussion and Limitations

TagFree marks an important step toward enabling accurate indoor activity identification, which does not require users hold or wear any RFID tags. There are however many possible future works to enhance the basic TagFree design.

Scalability: So far TagFree focus on accurately identifying a single person's activity. We expect to scale the system to identify multiple persons' activities. Given the complex interaction of the persons, data preprocessing and analysis will be much more challenging. Yet our preliminary results have shown that deep learning has great potentials in decoupling the individual activities and is far more effective than conventional learning tools. Also the coverage of TagFree with a single antenna array is limited to 12 m, which is the reading range of the Impinj reader. To cover larger areas, one may deploy multiple antenna arrays with a Impinj antenna hub and place more RFID tags as references.
Model Extension: Our deep learning model trained cannot be directly used in a different environments with non-identical antenna settings or tag placement. As a result, the model has to be re-trained for different environments. A possible extension is to tuning the prediction results based on domain-expert knowledge. For example, suppose there exist several activities that follow a certain sequential order, the final decision can be made based on both the softmax score and the constraints from expert knowledge. Training the model with expert knowledge will be an important future work.
User Identification: Our current design can identify only several activities, but cannot differentiate users. Future research may address this issue by extracting robust and representative user features instead of activity features to train a model for user identification. We may adopt the solutions based on WiFi signals [27] in our future work.

8 RELATED WORK

Radio Frequency Identification (RFID) is a promising technology due to its low cost, small form size, and batterylessness, making it widely used in a range of mobile applications, including detection of human-object interaction [25], people/object tracking [31] and more complex activity identification [6].

In activity identification, previous solutions exploited the changing of RSSI (received signal strength indicator) [35][5][19] incurred by human actions. Yet RSSI is insensitive to small body movement, and thus difficult to achieve high-precision identification. Other works for tag-free localization [32][21] and body movement [34] rely on RSSI fingerprints; they typically deployed reference tags on a monitoring region and generated in the training phase by requiring a person to act in different locations. In the testing phase, they map the resulting RSSI to the closest fingerprint to identify the status of the person. In our work, we have shown that RSSI is insensitive to such small activities such as handshaking, which is not suitable to our goal for tag-free activity identification.

Wireless localization techniques using phase measurement have successfully achieved centimeter accuracy. RF-compass [24] presented a WiFi-based approach to classify a predefined set of nine gestures; E-eyes [28] proposed a location-oriented activity identification system, which utilized WiFi signals to recognize in-home human activities; Tagoram [31] assumed that the tag movement velocity and its moving track are known in advance, and built a differential augmented hologram using the phase values collected from the antennas; RF-IDraw [25] achieved good tracking accuracy with eight antennas connected to two RFID readers. Ding *et al.* [6] further developed FEMO that uses the frequency shifts of the movements to determine what exercise a user is performing. Although the advanced solutions, e.g., RF-IDraw [25] and Tagoram [31], achieved high accuracy through exploring antenna arrays, their performance may degrade heavily in indoor environments with multipath. TagFree tackles the activity identification problem in the indoor multipath-rich environment. We carefully handle multipath signals and develop a deep learning approach to solve the activity identification problem.

Deep learning has become a very active research area for general activity understanding. We employ Long Short Term Memory (LSTM) [13] units to build a recurrent neural network, which discovers long-range temporal relationships by using memory cells to store, modify, and access internal state. LSTM

networks have been successfully applied to many tasks such as handwriting [9] and speech recognition [10]. Our research well complements these works by exploring the potential of deep learning to tag-free activity identification. We demonstrate the necessity and benefits of appropriate data preprocessing to maximize the performance gain from deep learning approaches.

9 CONCLUSION

In this paper, we have shown that TagFree can identify activities without attaching tags on the targets in typical indoor environments. TagFree employs a data preprocessing scheme to handle frequency hopping and de-couple multipath signals, which potentially offers the rich information for activity identification. We then utilize a Convolutional Neural Network and a Long Short Term Memory network to effectively solve the tag-free activity identification problem. A prototype has been implemented using a commercial `Impinj` reader and our extensive experimental results have demonstrated that TagFree achieves an activity identification accuracy of 94% on average in multipath-rich environments, which is better than the state-of-art solutions.

REFERENCES

[1] Muhammad Raisul Alam, Mamun Bin Ibne Reaz, and Mohd Alauddin Mohd Ali. 2012. A review of smart homespast, present, and future. *IEEE Transactions on Systems, Man, and Cybernetics* 42, 6 (2012), 1190–1203.

[2] Mathieu Aubry and Bryan C Russell. Understanding deep features with computer-generated imagery. In *Proceedings of IEEE CVPR 2015*.

[3] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *Comput. Surveys* 46, 3 (2014).

[4] Jose M Chaquet, Enrique J Carmona, and Antonio Fernández-Caballero. 2013. A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding* 117, 6 (2013), 633–659.

[5] Krishna Chintalapudi, Anand Padmanabha Iyer, and Venkata N Padmanabhan. Indoor localization without the pain. In *Proceedings of ACM MobiCom 2010*.

[6] Han Ding, Longfei Shangguan, Zheng Yang, Jinsong Han, Zimu Zhou, Panlong Yang, Wei Xi, and Jizhong Zhao. Femo: A platform for free-weight exercise monitoring with rfids. In *Proceedings of ACM SenSys 2015*.

[7] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of IEEE CVPR 2015*.

[8] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of AISTATS 2011*.

[9] Alex Graves, Marcus Liwicki, Santiago Fernández, Roman Bertolami, Horst Bunke, and Jürgen Schmidhuber. 2009. A novel connectionist system for unconstrained handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 5 (2009), 855–868.

[10] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *Proceedings of IEEE ICASSP 2013*.

[11] Jinsong Han, Han Ding, Chen Qian, Wei Xi, Zhi Wang, Zhiping Jiang, Longfei Shangguan, and Jizhong Zhao. 2016. Cbid: A customer behavior identification system using passive tags. *IEEE/ACM Transactions on Networking* 24, 5 (2016), 2885–2898.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of IEEE CVPR 2016*.

[13] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of NIPS 2012*.

[15] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.

[16] Xinyu Li, Yanyi Zhang, Ivan Marsic, Aleksandra Sarcevic, and Randall S Burd. Deep Learning for RFID-Based Activity Recognition. In *Proceedings of ACM SenSys 2016*.

[17] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, and others. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, Oct (2011), 2825–2830.
[18] Low Level Reader Protocol. 2007. Version 1.0. 1. *EPCglobal Inc, Aug* (2007).
[19] Anshul Rai, Krishna Kant Chintalapudi, Venkata N Padmanabhan, and Rijurekha Sen. Zee: zero-effort crowdsourcing for indoor localization. In *Proceedings of ACM MobiCom 2012*.
[20] Ralph Schmidt. 1986. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation* 34, 3 (1986), 276–280.
[21] Moustafa Seifeldin, Ahmed Saeed, Ahmed E Kosba, Amr El-Keyi, and Moustafa Youssef. 2013. Nuzzer: A large-scale device-free passive localization system for wireless environments. *IEEE Transactions on Mobile Computing* 12, 7 (2013), 1321–1334.
[22] Longfei Shangguan, Zheng Yang, Alex X Liu, Zimu Zhou, and Yunhao Liu. Relative Localization of RFID Tags using Spatial-Temporal Phase Profiling.. In *Proceedings of USENIX NSDI 2015*.
[23] Benjamin Wagner and Dirk Timmermann. Adaptive clustering for device free user positioning utilizing passive RFID. In *Proceedings of ACM UbiComp 2013*.
[24] Jue Wang, Fadel Adib, Ross Knepper, Dina Katabi, and Daniela Rus. RF-compass: robot object manipulation using RFIDs. In *Proceedings of ACM MobiCom 2013*.
[25] Jue Wang, Deepak Vasisht, and Dina Katabi. RF-IDraw: virtual touch screen in the air using RF signals. In *Proceedings of ACM SIGCOMM 2015*.
[26] Ju Wang, Jie Xiong, Hongbo Jiang, Xiaojiang Chen, and Dingyi Fang. D-watch: Embracing" bad" multipaths for device-free localization with COTS RFID devices. In *Proceedings of ACM CoNEXT 2016*.
[27] Wei Wang, Alex X Liu, and Muhammad Shahzad. Gait recognition using wifi signals. In *Proceedings of ACM UbiComp 2016*.
[28] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. E-eyes: device-free location-oriented activity identification using fine-grained WiFi signatures. In *Proceedings of ACM MobiCom 2014*.
[29] Teng Wei and Xinyu Zhang. Gyro in the air: tracking 3D orientation of batteryless internet-of-things. In *Proceedings of ACM MobiCom 2016*.
[30] Jiang Xiao, Zimu Zhou, Youwen Yi, and Lionel M Ni. 2016. A Survey on Wireless Indoor Localization from the Device Perspective. *Comput. Surveys* 49, 2 (2016).
[31] Lei Yang, Yekui Chen, Xiang-Yang Li, Chaowei Xiao, Mo Li, and Yunhao Liu. Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices. In *Proceedings of ACM MobiCom 2014*.
[32] Moustafa Youssef, Matthew Mah, and Ashok Agrawala. Challenges: device-free passive localization for wireless environments. In *Proceedings of ACM MobiCom 2007*.
[33] Joe Yue-Hei Ng, Matthew Hausknecht, Sudheendra Vijayanarasimhan, Oriol Vinyals, Rajat Monga, and George Toderici. Beyond short snippets: Deep networks for video classification. In *Proceedings of IEEE CVPR 2015*.
[34] Daqiang Zhang, Jingyu Zhou, Minyi Guo, Jiannong Cao, and Tianbao Li. 2011. TASA: Tag-free activity sensing using RFID tag arrays. *IEEE Transactions on Parallel and Distributed Systems* 22, 4 (2011), 558–570.
[35] Yiyang Zhao, Yunhao Liu, and Lionel M Ni. VIRE: Active RFID-based localization using virtual reference elimination. In *Proceedings of IEEE ICPP 2007*.

Received May 2017; revised November 2017; accepted January 2018