Multiple Object Activity Identification using RFIDs: A Multipath-Aware Deep Learning Solution

Xiaoyi Fan*, Feng Wang**, Wei Gong†*, Lei Zhang*, Jiangchuan Liu*,

*School of Computing Science, Simon Fraser University, Canada
**Department of Computer and Information Science, The University of Mississippi, USA
†School of Computer Science and Technology, University of Science and Technology of China, China
Email: xiaoyif@sfu.ca, fwang@cs.olemiss.edu, weigong@ustc.edu.cn, lza70@sfu.ca, jcliu@cs.sfu.ca

Abstract—RFID-based human activity identification has become a key component in today's Internet-of-Things applications. State-of-the-art solutions mostly focus on the simple scenario with a single person in the open space. Extension to the more realistic realworld scenarios with multiple persons however is non-trivial. Given the much richer interactions among them, the backscattered signals will inevitably mixed, obscuring the information of individual activities. This is further complicated with multi-path in a common indoor environment.

In this paper, we however argue that, though often considered harmful, the rich interactions combined with multi-path indeed offer more observable data. After careful processing the raw signals, critical information about the activities can be unveiled through modern learning tools. We present M²AI, which for the first time accommodates both multi-path and multi-object for activity identification. M²AI incorporates a phase calibration mechanism to automatically eliminate the frequency hopping offsets, and a novel decoupling mechanism for the periodogram and pseduospectrum in the raw signal mixture. The refined data are then fed into an advanced deep-learning engine that integrates a Convolutional Neural Network and a Long Short Term Memory network, which examines both spatial and temporal information in realtime for activity identification.

Our M²AI is readily deployable using off-the-shelf RFID readers. We have implemented an M²AI prototype with Impinj UHF passive tags and a Speedway R420 reader. Experiments with multiple objects in a multipath-rich indoor environments report an activity identification accuracy of 97%, a significant gain (27%) over state-of-art solutions.

Index Terms—RFID; Backscatter; Activity Identification; Deep Learning;

I. Introduction

Human activity identification has become a key component in such critical Internet-of-Things applications as healthcare and smart homes. It has attracted significant attention from both academia and industry, with a wide range of solutions based on cameras [1], radars [2], and/or various inertial sensors [3]. They generally require the object of identification to carry sensors/wireless transceivers, which are not negligible in both size and weight, not to mention the constraints from the battery. Recently, *Radio Frequency Identification* (RFID) is experiencing an explosion in many application contexts given its low cost, small form size, and batterylessness nature [4] [5] [6] [7] [8]. We have seen pioneer studies on RFID-based activity identification as well [9] [10], together with preliminary adoptions in industry. For instance, Disney has built an RFID gaming system that can sense when a player attached with an RFID tag is moving or touching objects in real time [11].

The state-of-the-art solutions mostly focus on the simple scenario, i.e., a single person in the open space. A reader tracks the wireless signal backscattered from the person attached with RFID tags and, from changes of the signal, identifies the activities of the person. Extension to the more realistic realworld scenarios with multiple persons (each with distinct RFID tags) however is non-trivial. Given the much richer interactions among them, the backscattered signals will inevitably mixed, obscuring the information of multiple on-ongoing individual activities. This is further complicated with multipath in a common indoor environment. Besides reflection from surrounding walls, a person can also be occluded by furniture and other persons, resulting in the signals to be deflected and take multiple paths to arrive at the RFID reader.

There have been efforts towards better understanding and distinguishing signals from different paths caused by mutual interferences among multiple persons' activities, and the Hidden Markov Model (HMM) has been suggested as a useful tool [10]. Unfortunately, the features of RFID-based activities are hard to be pre-selected manually, for the received signals are a mixture of multi-path and multi-object, and the rules for making correct estimations are hard to be pre-defined, either. Without good *a priori* knowledge, the effectiveness of HMM can be quite limited in this context.

In this paper, we however argue that, though often considered harmful in multi-object scenarios, the rich interactions combined with multi-path indeed offer more observable data that carry abundant information about the activities. The raw RFID signals, which is a mixture from multi-objects, can be too noisy to be directly processed and understood; to limit co-channel interference, commercial RFID systems also support frequency hopping that further scrambles the raw signals. Yet after careful processing, critical information about the activities can be unveiled through modern learning tools. We present M²AI (Multipath-aware Multi-object Activity Identification), an RFID-based activity identification framework that for the
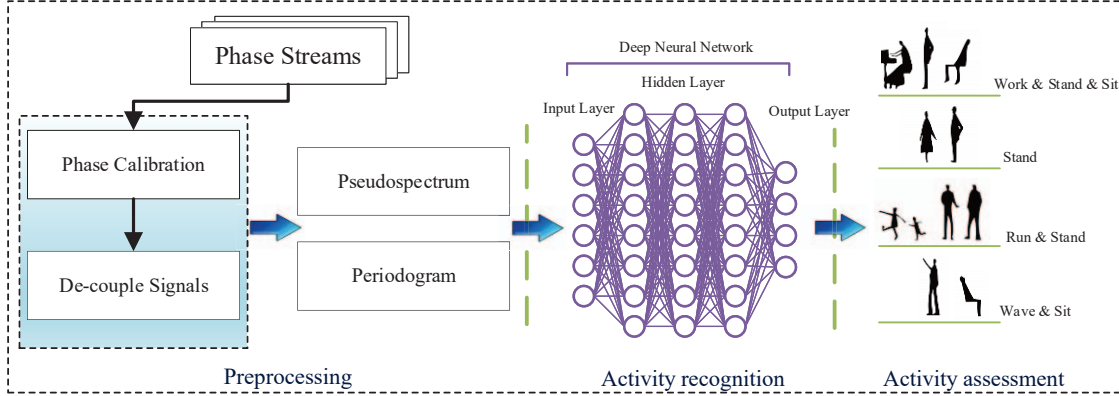
Fig. 1: M^2AI framework

first time accommodates both multi-path and multi-object. [1]

As shown in Fig. 1, M^2AI involves four steps: preprocessing, activity identification, and activity assessment. Our data preprocessing incorporates (1) a phase calibration mechanism to automatically eliminate the frequency hopping offsets; and (2) a novel decoupling mechanism that jointly considers both periodogram [12] and pseduospectrum [13] in the raw signal mixture. The carefully separated angles and powers of different paths are then fed into an advanced deep-learning engine for activity identification. Our engine integrates a Convolutional Neural Network (CNN) [14] and a Long Short Term Memory (LSTM) network [15]. We show that this integration is computationally effective with high accuracy, beating such conventional classification tools as SVM and Nearest Neighbors.

Our M^2AI is readily deployable using off-the-shelf RFID readers (e.g., a single UHF reader with a limited number of antennas), and allows reusing existing RFID readers for indoor activity identification. We have implemented an M^2AI prototype with Impinj UHF passive tags and a Speedway R420 reader. Experiments with multiple objects in a multipath-rich indoor environments report an identification accuracy of 97%, a significant gain (27%) over state-of-art solutions.

The rest of the paper is organized as follows. Section II examines the key challenges for identifying the activities of multiple objects in the indoor environment, and presents our M^2AI framework to explore the hidden information in multipath. Section III introduces our data pre-processing scheme to deal with frequency hopping and de-couple multipath signals. Section IV then presents a deep learning solution for multiple object activity identification from the processed data. Section V discusses the framework implementation details, with its performance being evaluated in Section VI. Section VII discusses some potential further extensions. We provide a literature review on related works in Section VIII, and conclude the paper in Section IX.

[1]In this paper, an object of activity identification refers to a person equipped with multiple RFID tags, where the number of tags is set to 3 by default. Note however our solution works with non-human moving objects as well.

II. M^2AI: MOTIVATION AND OVERVIEW

The Angle of Arrival (AoA) of an RF source is computed by comparing the phases of the received signals at antennas [16] [17]. AoA estimation is widely used in RF-based positioning given the different propagation distances to different antennas, and serves as a foundation for activity identification as well [9]. It is known that the estimation is quite challenging in a multi-path indoor environment, not to mention with multiple objects.

Consider an illustrative example in Fig. 2. In this paper, we consider an object of activity identification as a single person equipped with multiple tags. In Fig. 2(a), Tag 1 is attached to one person, and there exist three paths in the indoor environment as the AoA spectrum shows. In particular, the stationary Tag 1 continuously reflects the signals with the same angle and power from $40°$, $90°$ and $125°$, respectively. Further, Fig. 2(b) illustrates the scenario for a single object of activity identification with another moving object, where the moving person blocks path 1 at $40°$. As such, not only the peak of the blocked path is decreased, but the peak amplitudes and angles of the other paths are also changed. Specifically, the path at $90°$ is shifted to $85°$ with power decrease. When there are more tags in the area, e.g., Fig. 2(c) with five more tags added, we can see that the number of signal paths increases rapidly in multiple objects of activity identification.

Intuitively, estimating AoA is more difficult in these indoor environments than in an open space. Analyzing the estimated data to derive the corresponding activities can be even more complicated, particularly for the latter two cases with multiobject. On the other hand, the massive multi-path signal information in this context also provides opportunities, as they indeed reflect the activities from different observation angles.

It is however non-trivial to explore the rich but hidden information therein.

- **Mixture of signals from multiple paths and multiple objects**: First, the signals may twist with each other and sometimes hide behind noises, the patterns of the relationships between them and human activities are hardly determined with a simple set of predefined rules;
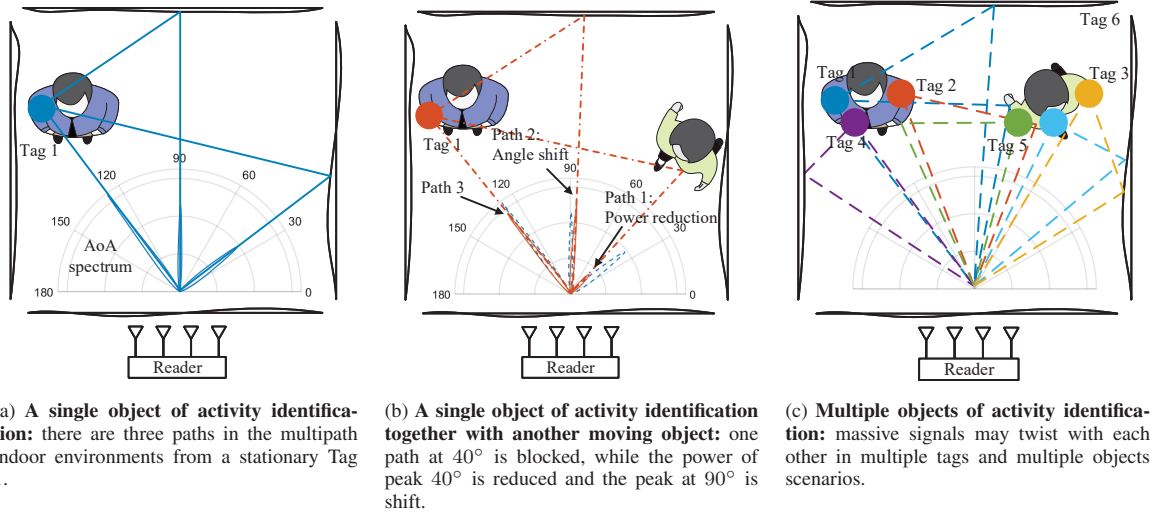
(a) **A single object of activity identification:** there are three paths in the multipath indoor environments from a stationary Tag 1.

(b) **A single object of activity identification together with another moving object:** one path at $40°$ is blocked, while the power of peak $40°$ is reduced and the peak at $90°$ is shift.

(c) **Multiple objects of activity identification:** massive signals may twist with each other in multiple tags and multiple objects scenarios.

Fig. 2: The basic idea of M^2AI from a single object to multiple objects of activity identification in the indoor environments.

- **Frequency hopping:** Second, to limit the co-channel interference, FCC regulation requires that commercial UHF RFID readers must randomly hop to one of the 50 center frequencies within the 902-928 MHz band every 200 ms. The recent study shows that frequency hopping will cause noticeable phase offset due to the phase difference of the oscillator and the non-uniform frequency responses of the tags' antennas [18]. The aggregated impact with multipath and multi-object can be even more dramatic.

To deal with the above challenges, our M^2AI incorporates a phase calibration scheme to automatically calibrate the phase difference between frequencies without the requirement of human intervention. It then combines the signals' psedu-ospectrum and periodogram frames to decouple signals from multiple paths while reserving the accurate signal power and direction estimation at the same time. Such pre-processing provides refined data with the essential information for identifying the activities through automated learning.

Our learning network takes two inputs, i.e., periodogram and pseduospectrum. Since each individual spectrum frame forms only a small part of the human activities, such conventional machine learning methods as *support vector machine* (SVM) and *decision tree* based on the incomplete information can easily be confused. They do not well utilize the temporal spectrum, either, which contain important information about the activity. Hence, we explore the latest deep learning tools; in particular, we use an integrated design of a Convolutional Neural Network (CNN) [14] and a Long Short Term Memory (LSTM) network [15] units, which uses memory cells to store, modify, and access internal state, so as to discover long-range temporal relationships. Our networks integrate information and maintain a constant number of parameters while capturing an accurate activity description in the massive spectrum data.

It is worth noting that although working with rich multi-paths in indoor environments, our M^2AI design requires only one reader with four antennas and can increase the path diversity by adding tags in the area, so the deployment cost can be well reduced. Moreover, M^2AI does not need to explicitly know the RFID tags' locations. Hence, the tags can be arbitrarily placed with a high degree of flexibility.
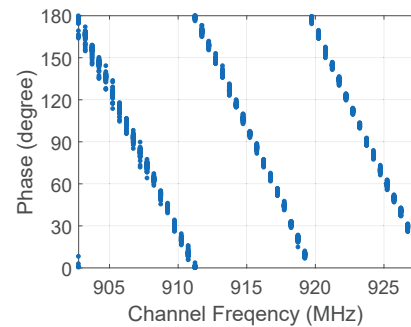

Fig. 3: Phase jumping caused by frequency hopping

III. DATA PREPROCESSING

The state-of-the-art RFID hardware with the standard Low Level Reader Protocol (LLRP) can provide low-level data reports, such as the phase and Doppler shift. Yet, as indicated in previous research work [19] [20], the reported data may not be accurate enough to be immediately applied to activity identification, due to such factors as the multipath effect and frequency hopping. Targeting on these factors, we propose our M^2AI design to preprocess the measured phase from the reader.

A. Phase Calibration

To limit the co-channel interference, FCC regulation requires that commercial UHF RFID readers must randomly hop
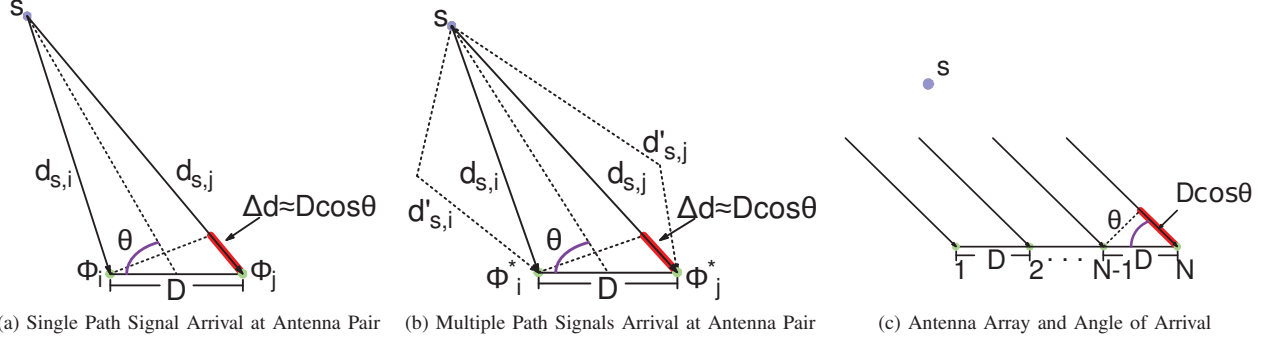
Fig. 4: Illustration of Angle of Arrival

(a) Single Path Signal Arrival at Antenna Pair   (b) Multiple Path Signals Arrival at Antenna Pair   (c) Antenna Array and Angle of Arrival

to one of 50 center frequencies within the 902-928 MHz band every 400 ms, which will cause phase offset due to the phase difference of the oscillator and the non-uniform frequency responses of the tags antennas. To this end, we measure the phase of a stationary tag for 60 seconds, which supposed to keep consistent on phase values. We plot phase values against their frequency in Fig. 3, where it is clear to see that the phase and frequency relation follows the linear model. These experiments imply that different frequencies induce different initial phase-offsets at the reader. To overcome the issue, we 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. In particular, we have frequency $f_j, \forall j \in [1, 50]$ and set a common frequency as f_r (default to 910.25 MHz). Let $\phi^j(t)$ denote the measured phase at frequency f_j at time t. Let $\tilde{\phi}^j$ and $\tilde{\phi}^r$ represent the median value of measured phase in recent 10 seconds at frequency f_j and common frequency f_r, respectively. We map the measured phase ϕ^j at frequency f_j to the calibrated phase $\phi(t)$ as follows:

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

B. Angle of Arrival

Intuitively, the AoA estimation works as illustrated in Fig. 4(a): A signal source s impinges on the antenna pair with an angle θ, and D is the distance between two antennas. We can calculate the spatial angle θ by comparing the phases of the received signals at multiple antennas. The phase ϕ of an RF signal rotates by 2π for every λ (wavelength) distance the signal travels. Let $d_{s,i}$ and $d_{s,j}$ denote the distances from the source s, to the two antennas respectively, and ϕ_i and ϕ_j are the phases of the received signal that we measure at the two antennas. The phase difference between the received signals at the two antennas, $\Delta\phi_{j,i} = \phi_j - \phi_i$, relates to the difference in their distances from the source, $\Delta d_{j,i} = d_{s,j} - d_{s,i}$, as follows:

$$\frac{\Delta d_{j,i}}{\lambda} = \frac{\Delta\phi_{j,i}}{2\pi} + k \qquad (2)$$

where k can be any integer in $[-\frac{D}{\lambda} - \frac{\Delta\phi_{j,i}}{2\pi}, \frac{D}{\lambda} - \frac{\Delta\phi_{j,i}}{2\pi}]$.

However, we find that the above intuition is true only when the multipath effect is negligible. As seen from Fig. 4(b), if the signal arrives at each antenna via two paths, the overall phases received at the two antennas become ϕ_i^* and ϕ_j^*. Let s_i and s_j denote the signals along direct path from source s to antenna i and j. Let s_i' and s_j' denote the signals along second path from source s to antenna i and j. Let α denote the amplitude of s. Let α_i, α_j, α_i' and α_j' represent the propagation attenuation at the path $d_{s,i}, d_{s,j}, d_{s,i}'$ and $d_{s,j}'$. The overall amplitude received at the two antennas become α_i^* and α_j^*. We assume the source s is far from antennas, therefore $d_{s,i} = d_{s,j} = d$, and have the following equations:

$$s_i = \alpha \cdot \alpha_i \cdot e^{j\phi_i} \qquad (3)$$
$$s_j = \alpha \cdot \alpha_j \cdot e^{j\phi_j} \qquad (4)$$

where $\phi_i = \phi_0 + \frac{d}{\lambda} \cdot 2\pi$ and $\phi_j = \phi_0 + (\frac{d}{\lambda} + \frac{D\cos\theta}{\lambda}) \cdot 2\pi$.

$$s_i^* = \alpha \cdot \alpha_i^* \cdot e^{j\phi_i^*} = \alpha \cdot \alpha_i \cdot e^{j\phi_i} + \alpha \cdot \alpha_i' \cdot e^{j\phi_i'} \qquad (5)$$
$$s_j^* = \alpha \cdot \alpha_j^* \cdot e^{j\phi_j^*} = \alpha \cdot \alpha_j \cdot e^{j\phi_j} + \alpha \cdot \alpha_j' \cdot e^{j\phi_j'} \qquad (6)$$

where $\phi_i' = \phi_0 + \frac{d_{s,i}'}{\lambda} \cdot 2\pi$, $\phi_j' = \phi_0 + (\frac{d_{s,j}'}{\lambda} + \frac{D\cos\theta}{\lambda}) \cdot 2\pi$, $\phi_i^* = 2 \cdot \phi_0 + \frac{d}{\lambda} + \frac{d_{s,i}'}{\lambda}$ and $\phi_j^* = 2 \cdot \phi_0 + \frac{d}{\lambda} + \frac{d_{s,j}'}{\lambda} + 2 \cdot \frac{D\cos\theta}{\lambda}$. Thus, the new phase difference under this simple multipath scenario is not equal (nor a good approximation) to the original phase difference, i.e., $\Delta\phi_{j,i} \neq \Delta\phi_{j,i}^*$. Hence, these approaches are ineffective in multipath-rich indoor environments.

C. De-couple Multipath Signals

As shown in Fig. 4 (c), consider a number of plane waves from M narrow-band sources $s_1(t) = \rho_1 e^{j\phi_1}, \ldots, s_M(t) = \rho_M e^{j\phi_M}$ (which may also be multipath signals from the same source) impinging from different angles $\theta_1, \ldots, \theta_M$, and into a uniform linear array (ULA) of N antennas. Then we have

$$\boldsymbol{s}(t) = [s_1(t), \ldots, s_M(t)]^\top \qquad (7)$$

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

$$\boldsymbol{a}(\theta) = \begin{bmatrix} 1 & e^{-j\phi} & \ldots & e^{-j(N-1)\phi} \end{bmatrix}^\top \qquad (8)$$

(a) Pseduospectrum Estimation   (b) Periodogram Estimation   (c) Pseduospectrum Frame   (d) Periodogram Frame
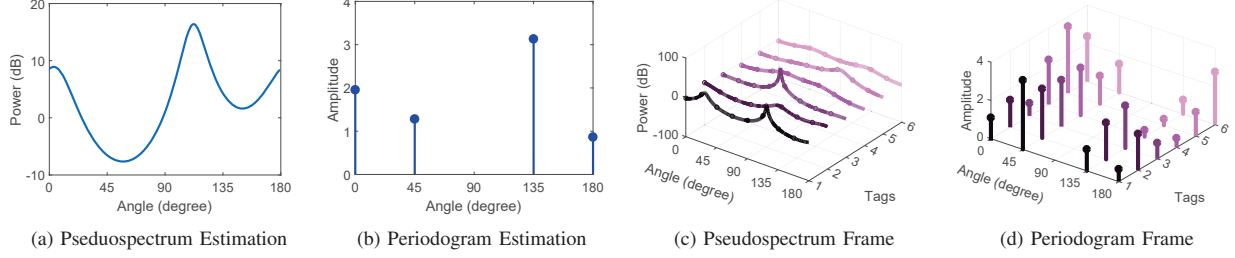
Fig. 5: Illustration of data preprocessing in our M^{2}AI system framework.

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

$$\boldsymbol{r}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{w}(t) \tag{9}$$

where $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}(\theta_1), \ldots, \boldsymbol{a}(\theta_M) \end{bmatrix}$ is an $N \times M$ matrix of steering vectors, and $\boldsymbol{w}(t)$ is a noise term. The spatial correlation matrix R of the observed signal vector $\boldsymbol{r}(t)$ can be defined as:

$$\boldsymbol{R_r} = E\{\boldsymbol{r}(t)\boldsymbol{r}^H(t)\} = \boldsymbol{A}\boldsymbol{R_s}\boldsymbol{A}^H + \mu^2 \boldsymbol{I} \tag{10}$$

where $\boldsymbol{R_s} = E\{\boldsymbol{s}(t)\boldsymbol{s}^H(t)\}$, μ^2 is the noise covariance matrix, and $\boldsymbol{I}$ represents an $N \times N$ matrix.

1) Pseduospectrum Estimation: Our pseduospectrum estimation design is mainly based on the MUSIC (MUltiple SIgnal Classification) algorithm [13], which is one of the high resolution subspace AOA (Arrival of Angle) algorithms and is originally used to estimate the number of received signals from their directions of arrival. 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 $\boldsymbol{U_n}$:

$$[\boldsymbol{U_s}\boldsymbol{U_n}] = [\underbrace{\boldsymbol{U}_1, \ldots, \boldsymbol{U}_M}_{\boldsymbol{U_s}}, \underbrace{\boldsymbol{U}_{M+1}, \ldots, \boldsymbol{U}_N}_{\boldsymbol{U_n}}] \tag{11}$$

Similar to the MUSIC algorithm utilizing the orthogonality relationship between the signal and noise subspaces [13], which implies $\boldsymbol{a}^H(\theta)\boldsymbol{U_n} = \boldsymbol{0}$, the direction of arrival angle can be represented in terms of a spectral estimation plots:

$$P_{MUSIC} = \frac{1}{\boldsymbol{a}^H(\theta)\boldsymbol{U_n}\boldsymbol{U_n}^H\boldsymbol{a}(\theta)} \tag{12}$$

The above equation results in high peaks, when the direction of arrival of the signal source is exactly equal to that of θ. In Fig. 5(a), the M higher peaks are of great power, where each corresponds to an estimated arrival angle.

2) Periodogram Estimation: We introduce the periodogram [12] to strengthen pseduospectrum estimates by taking the accurate power information into consideration. The Power Spectral Density (PSD) is defined as the discrete-time Fourier transform (DTFT) of the covariance sequence:

$$\phi(\omega) = \sum_{k=-\infty}^{\infty} r(k)e^{-jwk} \tag{13}$$

where $r(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi(\omega)e^{j\omega k}d\omega$. We use the periodogram spectral estimator to compute the power density distribution as follows:

$$\hat{\phi}_p(\omega) = \frac{1}{N}|\sum_{t=1}^{N} y(t)e^{-jwt}|^2 \tag{14}$$

where $\{y(t)\}$ denote a deterministic discrete-time data sequence. In practice, it is not possible to evaluate $\hat{\phi}_p(\omega)$ over a continuum of frequencies. Hence, the frequency variable must be sampled for the purpose of computing $\hat{\phi}_p(\omega)$. The following frequency sampling scheme is most commonly used:

$$\omega = \frac{2\pi}{N}k, k = 0, \ldots, N - 1 \tag{15}$$

Let W be $e^{-i\frac{2\pi}{N}}$. Then, the evaluation of $\hat{\phi}_p(\omega)$ at the frequency samples reduces to the computation of the following Discrete Fourier Transform:

$$Y(k) = \sum_{t=1}^{N} y(t)W^{tk}, k = 0, \ldots, N - 1 \tag{16}$$

In our system, we use Fast Fourier Transform [21] to estimate the power distribution. According to the Parseval's theorem [22], the Fourier transform is unitary, i.e., the sum (or integral) of the square of a function is equal to the sum (or integral) of the square of its transform.

As illustrates in Fig. 5(b), we can get four values in the periodogram for the power density distribution. In the implementation, we have four antennas to connect to the Impinj Speedway R420 reader, where the number of RF ports in the reader limits the scale of our antenna array. We can increase the antenna number by Impinj antenna hubs. The effectiveness of the periodogram estimation has also been demonstrated by our experiments in Section. VI.

IV. DEEP LEARNING DESIGN FOR ACTIVITY IDENTIFICATION

This section describes the main components of our M^{2}AI design. Our deep learning design takes the results from data preprocessing into our M^{2}AI deep learning architecure, i.e., periodogram and pseduospectrum frames, as shown in Fig. 5(c) and Fig. 5(d), respectively. As illustrated in Fig. 6, we develop an integrated design of a Convolutional Neural Netwok (CNN) [14] and a Long Short Term Memory (LSTM)
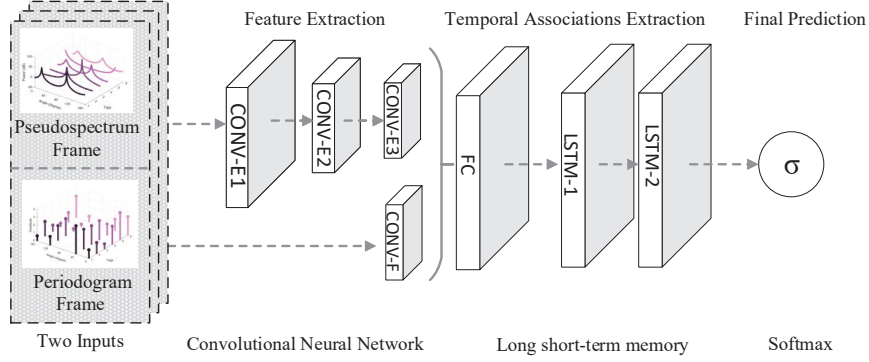
Fig. 6: M^{2}AI deep learning network architecture

network [15]. CNN networks can extract spatial relationships in a single spectrum frame; while LSTM networks can learn dynamic temporal relationships from a sequential spectrum frames. The output is the classification of object activities using a softmax layer. We discuss the design of each layer one by one in the following subsections.

A. 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. Specifically, we provide the following as input to the model: (1) pseudospectrum frames for angle of arrival as illustrated in Fig. 5(c), and (2) periodogram frames for power spectral density as Fig. 5(d) shows. By combining these two types of information, the model can take into account both the angle and power information of signals. The size of pseudospectrum frame is $n \times 180$, and the size of periodogram frame is $n \times N$, where n is the number of tags, N is the number of antennas 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 frame, where a series of spectrum frames along the time will further serve as the initial input for the hidden layer.

B. Hidden Layer

The hidden layer integrates a CNN structure and a LSTM structure.

1) CNN Structure: We construct a Convolutional Neural Network (CNN) to take the spectrum frames as input and provide the output to be fed into the LSTM structure. The extracted lower dimension features then form the input as a sequence of spectrum frames $\boldsymbol{x} = (x_1, \ldots, x_T)$ to the LSTM structure. In this work, we use the fully-connected layer to merge two inputs, where these features are outputs of rectified linear units. As illustrated in Fig. 6, CONV represents convolutional layers (with filter size of kernels: CONV-E1: $n \times 180$, CONV-E2: $n \times 32$, CONV-E3: $n \times 4$, CONV-F: $n \times N$, while FC represents fully-connected layers and LSTM represents LSTM layers.

2) LSTM Structure: In our design, a stacked LSTM first encodes the frames one by one from the output of the CNN. LSTM is a subnet that allows to easily memorize the context information for long periods of time in sequence data. The subnet 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 the forget gate f_t or additively modified by the current input x_t multiplied by the activation of the input gate i_t. The output gate o_t controls the emission of the memory value from the LSTM cell. 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 use two stacked LSTM layers, and each with 32 memory cells. Following the LSTM layers, a softmax classifier at the output layer is used to make a prediction at every spectrum frame.

C. Output Layer

The outputs from the last hidden layer are normalized with the softmax function. We use a softmax function to get the probability distribution over the activity label y in the activity cluster γ. Our goal is to find the maximum likelihood of all training samples. We apply the negative log probability as an objective function, i.e., cross entropy error function

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

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

V. SYSTEM IMPLEMENTATION

Our M^{2}AI can be fully implemented based on a commercial reader and requires no modifications to tags. In this section, we further describe the key implementation details that are not covered in the previous sections.
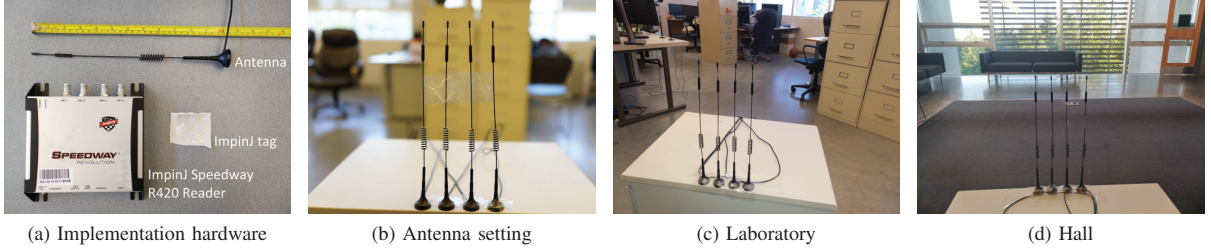
(a) Implementation hardware   (b) Antenna setting   (c) Laboratory   (d) Hall

Fig. 7: Commercial hardware used to implement M^{2}AI and two typical indoor environments, i.e., laboratory and hall, corresponding to high and low multipath environments, respectively.
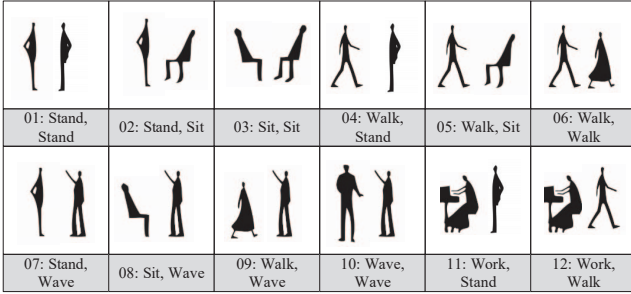

01: Stand, Stand	02: Stand, Sit	03: Sit, Sit
04: Walk, Stand	05: Walk, Sit	06: Walk, Walk
07: Stand, Wave	08: Sit, Wave	09: Walk, Wave
10: Wave, Wave	11: Work, Stand	12: Work, Walk

Fig. 8: Sketches of the tested activities

Readers and Tags: Our prototype implementation uses an Impinj Speedway R420 reader[2] and an antenna array without any hardware or firmware modification. The Impinj Speedway R420 reader is compatible with EPC Gen2 standard. The reader has four antenna ports to construct an antenna array, where the antennas work in a time division multiplexing mode with the inventory duration for an antenna port as 25 ms. The reader perform frequency hopping across the UHF RFID band, 902-928 MHz with hopping occurring between 902.75-927.25 MHz in 500 KHz steps, and the dwell time on each channel is set to 400 ms in a 20-second interval. Note that since the inventory duration for an antenna port is 25 ms, one round of ports switching in a 4-antenna array is 100 ms, which is far less than one channel dwell time, i.e., 400 ms. Thus, the pseduospectrum and periodogram estimation can be well implemented with Impinj R420 reader, which has been extensively used in the research community [23] [24]. According to the available frequencies of RFID reader, we set the common frequency $f_r = 910.25$ MHz, and the typical wavelength λ is 0.32 m. We use Impinj tags as shown in Fig. 7(a), which are one of the cheapest tags available on the market and cost 5 cents USD.

Antennas Settings: We connect our Impinj Speedway R420 reader to four omni-directional antennas as shown in Fig. 7. One important setting is the distance d between antenna pairs, where we set d as $\lambda/8$ with the following rationale:

- Theoretically, the antenna separation d should be spaced by $\lambda/2$, which effectively reduces the ambiguity caused

by the high-resolution grating lobes [9].
- 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, each tightly spaced antenna pair has a separation d of $\lambda/4$.
- Although the Impinj reader can report phase readings ranging from to $0°$ to $360°$, the signal processing component in the Impinj Speedway R420 reader introduces π radians of ambiguity such that the reported phase can be the true phase (ϕ) or the true phase plus π radians ($\phi + \pi$), which leads that the separation d is $\lambda/8$.

Given that λ is 0.32m, we set the separation d to $\lambda/8$ as well as the distance between antenna pair equal to 0.04 m.

Server and Algorithm Implementation:

The system employs a typical client-server architecture. The processes adopt Octane SDK JavaOctane SDK Programmer's Guide[3] with LLRP protocol to communicate with the reader, collect the readings and upload them to backend module. We utilize the multiple threads method, where a loop is used to execute the tag reading operation and returns immediately a sequence of RFID readings to the calling thread. The calling thread then uploads the tag readings to the server. The backend module on the server accepts the streaming of tag readings, where the server also stores the training data in the database and execute our algorithms to identify the activity. CNN and LSTM classifiers are implemented in Keras with Tensorflow backend on Dual NVIDIA GeForce GTX 1080 Ti GPUs and the multiclass classifiers based on machine learning tools are implemented based on the Scikit-learn library.

VI. PERFORMANCE EVALUATION

A. Experiment Setup

We conduct extensive experiments in two typical indoor environments, as Fig. 7(c) and (d) show: a laboratory and an empty hall corresponding to high and low multipath environments, respectively. The laboratory with a size of 13.75 m × 10.50 m has many file cabinets and writing desks, as shown in Fig. 7(c). The empty hall with a size of 8.75 m × 7.50 m is shown in Fig 7(d). In each environment, we deploy one
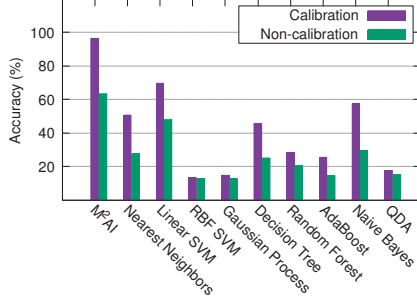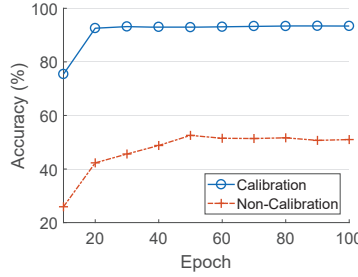
Fig. 9: Overall activity identification performance
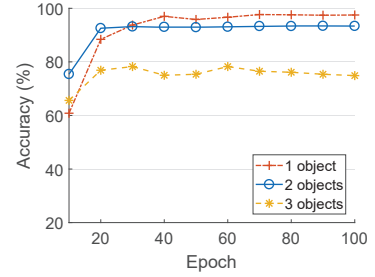
Fig. 10: Impact of phase calibration

Fig. 11: Impact of # of objects

Predicted lable	Actual label											
	A 01	A 02	A 03	A 04	A 05	A 06	A 07	A 08	A 09	A 10	A 11	A 12
A 01	97%	0	0	0	0	0	0	0	0	0	0	0
A 02	0	95%	0	0	0	0	0	0	0	0	0	0
A 03	0	5%	100%	0	0	2%	0	0	0	0	0	0
A 04	0	0	0	97%	0	0	0	0	0	0	0	0
A 05	0	0	0	0	100%	0	3%	0	0	0	0	0
A 06	0	0	0	0	0	98%	0	0	0	0	0	0
A 07	3%	0	0	0	0	0	95%	0	0	7%	0	0
A 08	0	0	0	0	0	0	0	95%	5%	0	7%	2%
A 09	0	0	0	0	0	0	0	0	95%	0	0	0
A 10	0	0	0	3%	0	0	2%	0	0	93%	0	0
A 11	0	0	0	0	0	0	0	5%	0	0	93%	0
A 12	0	0	0	0	0	0	0	0	0	0	0	98%

TABLE I: Confusion matrix of activity identification

readers and one 4-antenna array, where we place the antenna array at a height of 1.25 m. The tags are attached to volunteers, so that their heights are between 1 to 1.5 m above the ground and their distance to the reader is around 3 to 6 m.

We invite ten volunteers[4] and by default each volunteer is attached with three Impinj UHF passive tags on their hand, arm and shoulder, respectively. The volunteers do various activities about 3-6 meters away from the reader antennas in our experiments. We test 12 activity scenarios with two people as shown in Fig. 8. The default setting of our experiments is 4 antennas connecting to the reader.

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 100 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. Throughout training, we save the model and compute prediction accuracy[5] on the test set for each epoch.

B. Activity Identification Performance

Fig. 9 shows the performance of our M^2AI compared with ten commonly used classifiers: k-Nearest Neighbors, one-vs-all Linear SVM, one-vs-all RBF SVM, Gaussian Process, Decision Tree, Random Forest, Adaptive Boosting, Bayesian Net and Quadratic Discriminant Analysis. We can see that our M^2AI performs the best among all approaches with the highest accuracy up to 97%, which is 27% better than the runner-up approach (SVM). It is worth noting that the signal patterns have linear features, which makes the linear SVMs have better performance than the decision tree algorithm. Tab. I shows the detailed results of our M^2AI approach, where each row denotes the actual activity performed and each column represents the activity recognized by M^2AI. 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 table, the identification accuracy is at least 93% for all scenarios. This indicates that M^2AI can distinguish various mixed activities from different objects with high accuracy by efficiently extracting rich information about the activities.

Our phase calibration mechanism also contributes to the improvement of the precision to identify the object activity. We further compare M^2AI with phase calibration and non-calibration. The results are shown in Fig. 10, where M^2AI with phase calibration is able to achieve the activity identification accuracy of 97% against the accuracy of 52% with no calibration, since our calibration mechanism achieves a
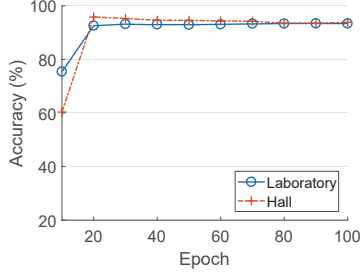
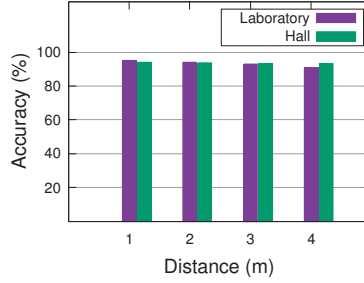Fig. 12: Impact of different places

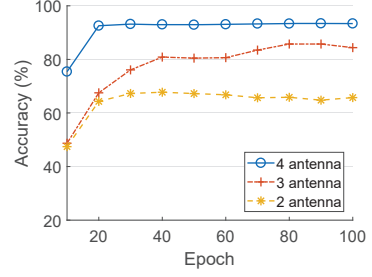Fig. 13: Impact of different distance
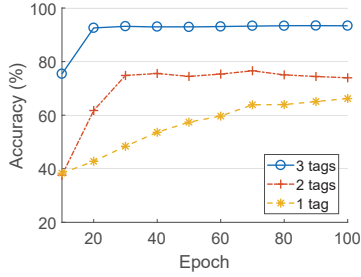

Fig. 14: Impact of # of antennas
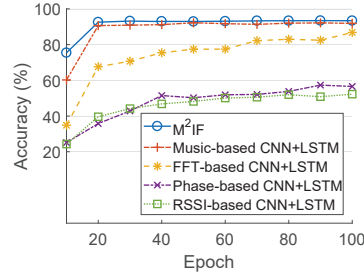

Fig. 15: Impact of # of tags per person

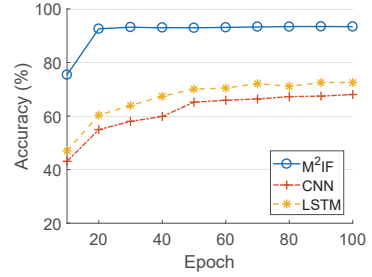Fig. 16: Impact of different inputs preprocessing

Fig. 17: Impact of different learning networks

high AoA estimation accuracy. The result also explains the statement in [20] that directly using the measured phase by Impinj R420 reader API is not accurate enough for activity identification and further demonstrates the effectiveness of our phase calibration mechanism.

We further examine the performance of M^2AI for multiple object activity identification. The results are shown in Fig. 11. Intuitively, with more people acting in the environment, the activity identification accuracy will drop dramatically. Yet M^2AI can achieve a high and relatively stable activity identification performance, where the average accuracy is still close to 80% even when three people are simultaneously acting in the environment. Fig. 12 shows the activity identification accuracy in two different environments, where M^2AI achieves the best performance in the hall (low multipath) environment with the accuracy of 95%, and its accuracy is close to the laboratory (high multipath) environments. In Fig. 13, we further evaluate the accuracy with varying distances from 1 m to 4 m, where the results do not exhibit clear correlation with the distance. Therefore, the distance is not a crucial factor affecting the activity identification accuracy.

M^2AI de-couples the multipaths using the array of antennas, which makes that with more multipaths, M^2AI achieves a higher mulitpath diversity in the area and improves the activity identification accuracy. Since the number of antennas may limit the multipath decoupling ability by our pre-processing scheme, we thus investigate the impact of the number of antennas as shown in Fig. 14. We can see that when the number

of antennas increases from 2 to 4^6 more angle information of multipath can be detected, and thus M^2AI can achieve even higher activity identification accuracy.

With more tags, more signals will be reflected, creating more paths to cover the monitoring area and provide the information of target's activity. In the laboratory environment, we vary the number of tags from 1 to 3 per person and the results are shown in Fig. 15. It is easy to see 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, the number of tags actually is the most effective and cheapest method to increase the path diversity in the environment. Thus, in an indoor environment with more tags, the cost of the equipment can then be dramatically reduced by simply using more tags for better accuracy.

In Fig. 16, we compare the results of our deep learning design with inputs from various preprocessing options. The comparison among the MUSIC-based, FFT-based, Phase-based, RSSI-based and M^2AI shows the effectiveness of our preprocessing scheme. In Fig. 17, we further compare the results of our M^2AI with different deep learning network architectures. It clearly shows that M^2AI can achieve 30% higher accuracy than the architecture only using CNN and our preprocessing scheme, which demonstrates that LSTM architectures are necessary for activity identification. On the other hand, M^2AI obtains 25% higher accuracy than the

---

^{6}Note that Impinj Speedway R420 reader has maximally four ports to connect with antennas.

architecture only using LSTM, which illustrates that CNN can efficiently extract the features for activity identification. In summary, the benefits of M^{2}AI comes from both our preprocessing scheme and deep learning approach that work jointly to enable harvesting the rich phase information for multiple object activity identification in a multipath-rich environment.

VII. Further Discussions

M^{2}AI marks an important step toward enabling accurate indoor activity identification. We envision the basic M^{2}AI can be further extended and explored in the following aspects.

Current deep learning model is trained under the same environment with identical antenna settings and tag placements. As a result, the model may need to be re-trained for different settings and environments. This is because the pseduospectrum and periodogram estimation are sensitive to the activity directions and tag orientations. Therefore, although the current implementation of M^{2}AI works well on accurately identifying predefined scenarios with multiple objects and multiple activities, it is not clear how M^{2}AI performs beyond predefined environments and activities. We expect one extension is to identify multiple activities in different scenarios without frequently re-training, which requires incorporating complex models of human activities. Yet our preliminary results have shown that deep learning has great potentials in de-coupling the individual activities and is thus far more effective than conventional learning tools towards this goal.

Another extension is on the coverage scale. Current coverage of M^{2}AI with a single antenna array is limited to 12 m, which is the reading range of the Impinj reader. Yet, beyond 6 meters, the RFID tag may not harvest enough energy to achieve a desired read rate. To cover larger areas, one may exploit Impinj antenna hubs to deploy multiple RFID antenna arrays. It is thus interesting to explore how the system performs along this extension.

VIII. 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 [9], people/object tracking [23] and more complex problems such as activity identification [10].

In activity identification, most solutions exploited the change of wireless signals incurred by human actions, where accurate wireless localization techniques are often used to achieve effective activity identification. Previous work on RF-based positioning primarily relied on RSSI information [25]. Ding *et al.* [5] estimated the number of human objects based on the received RF signal. Yet the major limitation of RSSI-based approaches is unreliable, since RSSI is insensitive to small body movement, and thus difficult to achieve high-precision localization. There is growing interest in using phase measurement for localization. Han *et al.* [4] leveraged the coupling effect caused by interference among passive tags to detect a single moving subject. Wang *et al.* [6] detected

the moving tags using physical-layer features of RFIDs, i.e., the phase profile and the backscatter link frequency. Liu *et al.* [7] exploited the polarization properties of the RF waves to identify the spatial direction of the tag. Tagoram [23] built a differential augmented hologram based the measured phase values. RF-IDraw [9] achieved good tracking accuracy with eight antennas connected to two RFID readers. Ding *et al.* [10] developed FEMO that uses the frequency shifts of the movements to determine what exercise a user is performing. RFIPad [26] transformed a tag plane into a virtual touch screen by anylzing the induced disturbance of RF signals. CARM [27] presents a state-of-the-art WiFi-based interface, yet it only supports the detection and classification of a predefined set of nine gestures. Different from these aforementioned approaches, M^{2}AI tackles the multiple objects activity identification problem in the indoor multipath-rich environment, where we design a series of solutions to carefully handle multipath signals and develop a deep learning approach to cope with the dynamics caused by multiple objects and multiple tags.

Deep learning has become a very active research area for general activity understanding, which has demonstrated great potentials to achieve high levels of performance and has revolutionized image classification [28] and speech recognition [29]. Recently, Li *et al.* [20] proposed to collect the coarsely grained RSSI readings to range about the tag, and apply a deep learning approach for activity identification. The recurrent neural network architecture we employ is derived from Long Short Term Memory (LSTM) units [15], and uses memory cells to store, modify, and access internal state, allowing it to discover long-range temporal relationships. Our research well complements these works by exploring the potential of applying deep learning approaches to multipath multiple object activity identification, where we also demonstrate the benefits of appropriate data preprocessing to maximize the performance gain from deep learning approaches.

IX. Conclusion

In this paper, we presented the Multipath-aware Multi-object Activity Identification framework (M^{2}AI) that can identify multiple object activities in typical indoor environments. M^{2}AI 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 build a deep learning architecture that can effectively solve the multiple object activity identification problem. Our extensive experimental results have demonstrated that M^{2}AI achieves the activity identification accuracy of 97%, which is 27% better than the state-of-art machine learning approaches.

REFERENCES

[1] J. M. Chaquet, E. J. Carmona, and A. Fernández-Caballero, "A survey of video datasets for human action and activity recognition," *Computer Vision and Image Understanding*, vol. 117, no. 6, pp. 633–659, 2013.

[2] J. Xiao, Z. Zhou, Y. Yi, and L. M. Ni, "A survey on wireless indoor localization from the device perspective," *ACM Computing Surveys*, vol. 49, no. 2, 2016.

[3] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Computing Surveys*, vol. 46, no. 3, 2014.

[4] J. Han, C. Qian, X. Wang, D. Ma, J. Zhao, P. Zhang, W. Xi, and Z. Jiang, "Twins: Device-free object tracking using passive tags," in *Proceedings of IEEE INFOCOM 2014*.

[5] H. Ding, J. Han, A. X. Liu, J. Zhao, P. Yang, W. Xi, and Z. Jiang, "Human object estimation via backscattered radio frequency signal," in *Proceedings of IEEE INFOCOM 2015*.

[6] C. Wang, L. Xie, W. Wang, T. Xue, and S. Lu, "Moving tag detection via physical layer analysis for large-scale rfid systems," in *Proceedings of IEEE INFOCOM 2016*.

[7] J. Liu, M. Chen, S. Chen, Q. Pan, and L. Chen, "Tag-compass: Determining the spatial direction of an object with small dimensions," in *Proceedings of IEEE INFOCOM 2017*.

[8] Y. Hou, Y. Wang, and Y. Zheng, "Tagbreathe: Monitor breathing with commodity rfid systems," in *Proceedings of IEEE ICDCS 2017*.

[9] J. Wang, D. Vasisht, and D. Katabi, "Rf-idraw: Virtual touch screen in the air using rf signals," in *Proceedings of ACM SIGCOMM 2015*.

[10] H. Ding, L. Shangguan, Z. Yang, J. Han, Z. Zhou, P. Yang, W. Xi, and J. Zhao, "Femo: A platform for free-weight exercise monitoring with rfids," in *Proceedings of ACM SenSys 2015*.

[11] A. Spielberg, A. Sample, S. E. Hudson, J. Mankoff, and J. McCann, "RapID: A Framework for Fabricating Low-Latency Interactive Objects with RFID Tags," in *Proceedings of ACM CHI 2016*.

[12] P. Stoica and R. L. Moses, *Introduction to spectral analysis*. Prentice hall Upper Saddle River, 1997.

[13] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] J. Xiong and K. Jamieson, "Arraytrack: A fine-grained indoor location system," in *Proceedings of USENIX NSDI 2013*.

[17] W. Gong and J. Liu, "Robust indoor wireless localization using sparse recovery," in *Proceedings of IEEE ICDCS 2017*.

[18] T. Wei and X. Zhang, "Gyro in the air: Tracking 3d orientation of batteryless internet-of-things," in *Proceedings of ACM MobiCom 2016*.

[19] J. Han, H. Ding, C. Qian, W. Xi, Z. Wang, Z. Jiang, L. Shangguan, and J. Zhao, "Cbid: A customer behavior identification system using passive tags," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2885–2898, 2016.

[20] X. Li, Y. Zhang, I. Marsic, A. Sarcevic, and R. S. Burd, "Deep learning for rfid-based activity recognition," in *Proceedings of ACM SenSys 2016*.

[21] G. Bergland, "A guided tour of the fast fourier transform," *IEEE spectrum*, vol. 6, no. 7, pp. 41–52, 1969.

[22] W. C. Chew and W. C. Chew, *Waves and fields in inhomogeneous media*, vol. 522. IEEE press New York, 1995.

[23] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile rfid tags to high precision using cots devices," in *Proceedings of ACM MobiCom 2014*.

[24] J. Wang, J. Xiong, H. Jiang, X. Chen, and D. Fang, "D-watch: Embracing "bad" multipaths for device-free localization with cots rfid devices," in *Proceedings of ACM CoNEXT 2016*.

[25] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proceedings of ACM MobiCom 2012*.

[26] H. Ding, C. Qian, J. Han, G. Wang, W. Xi, K. Zhao, and J. Zhao, "Rfipad: Enabling cost-efficient and device-free in-air handwriting using passive tags," in *Proceedings of IEEE ICDCS 2017*.

[27] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of wifi signal based human activity recognition," in *Proceedings of ACM MobiCom 2015*.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of NIPS 2012*.

[29] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proceedings of IEEE ICASSP 2013*.