

When Learning Meets RFIDs: The Case of Activity Identification

by

Xiaoyi Fan

M.Sc., Simon Fraser University, 2015

B.Eng., Beijing University of Posts and Telecommunications, 2013

Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

in the
School of Computing Science
Faculty of Applied Sciences

© Xiaoyi Fan 2018

SIMON FRASER UNIVERSITY

Spring 2018

All rights reserved.

However, in accordance with the *Copyright Act of Canada*, this work may be reproduced without authorization under the conditions for “Fair Dealing.” Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

Approval

Name: Xiaoyi Fan
Degree: Doctor of Philosophy (Computing Science)
Title: *When Learning Meets RFIDs:
The Case of Activity Identification*
Examining Committee: **Dr. Angelica Lim** (chair)
Assistant Professor

Dr. Jiangchuan Liu
Senior Supervisor
Professor
School of Computing Science

Dr. Jiannan Wang
Supervisor
Assistant Professor
School of Computing Science

Dr. Martin Ester
Internal Examiner
Professor
School of Computing Science

Dr. Mo Li
External Examiner
Associate Professor
School of Computer Science and En-
gineering
Nanyang Technological University

Date Defended: 4 Apr 2018

Abstract

Over the past decades have seen booming interests in human activity identification that is widely used in a range of Internet-of-Things applications, such as healthcare and smart homes. It has attracted significant attention from both academia and industry, with a wide range of solutions based on cameras, radars, and/or various inertial sensors. 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. Radio frequency identification (RFID) is a promising technology that can overcome those difficulties due to its low cost, small form size, and batterylessness, making it widely used in a range of mobile applications.

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 RFID-based activity identification solutions are far from being satisfactory. 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 thesis, we first reviewed the literature and research challenges of multipath effects in activity identification with RFIDs. Then we introduced three advanced RFID learning-based activity identification frameworks, i.e., $i^2\text{tag}$, TagFree and $M^2\text{AI}$, for tag mobility profiling, RFID-based device-free activity identification and tag-attached multi-object activity identification, respectively. Our extensive experiments further demonstrate their superiority on activity identification in the multipath-rich environments.

Keywords: RFID; Internet-of-Things; Deep Learning; Activity Identification

Dedication

To my family.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my senior supervisor Dr. Jiangchuan Liu, for his constant support and insightful directions throughout my studies. His significant enlightenment, guidance and encouragement on my research are invaluable. He is an excellent example as a successful researcher who inspires me to continue pursuing a Ph.D. degree in academia. Without his encouragement and knowledge, this thesis could not have been written.

I also thank Dr. Jiannan Wang for their precious time reviewing my thesis and for advising me on improving this thesis. I am grateful to Dr. Martin Ester, Dr. Mo Li, and Dr. Angelica Lim, for serving in my examining committee.

I thank a number of friends for their assistance in technical matters and for the enjoyable time I spent with them during my stay at Simon Fraser University. In Particular, I thank Dr. Wei Gong, Dr. Feng Wang, and Dr. Xiaoqiang Ma, for helping me with the research as well as many other problems during my study. I also want to thank my friends, Yifei Zhu, Cong Zhang, Lei Zhang, Qiyun He, Si Chen, Fangxin Wang, Yutao Huang and Zhe Yang, for their kind help and support during my PhD study.

Last but not least, I thank my family for their love, care and unconditional support. I sincerely hope that I have made them proud of my achievements today. This thesis is dedicated to you all.

Table of Contents

Approval	ii
Abstract	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Figures	ix
1 Introduction	1
1.1 Overview	1
1.2 Thesis Contributions	2
1.3 Thesis Organization	4
2 Activity Identification with RFIDs	5
2.1 Tag-attached Scenarios	5
2.1.1 Activity Identification Approaches	5
2.1.2 RFID Localization Techniques	5
2.1.3 Phase-based Approaches	6
2.2 Tag-free Scenarios	9
2.3 Summary	10
3 RFID Mobility through Intelligent Profiling	11
3.1 Introduction	11
3.2 Why Single Parameter Doesn't Work	14
3.2.1 Phase Difference as a Single Value	14
3.2.2 Phase Differences as a Vector	18
3.3 System Design	20
3.3.1 Mobility Profile	20
3.3.2 Mobility Detection	25

3.3.3	Understanding the Activities	26
3.4	System Implementation	28
3.5	Evaluation	30
3.5.1	Mobility Accuracy	30
3.5.2	Activities and Mobility	33
3.5.3	Realtime Performance	34
4	TagFree Activity Identification with RFIDs	36
4.1	Introduction	36
4.2	Raw RSSI or Phase? Challenges for Tag-free Activity Identification	37
4.3	TagFree Overview	43
4.4	Data Preprocessing	45
4.4.1	Phase Calibration	47
4.4.2	Multipath De-coupling	47
4.5	Deep Learning for Activity Identification	50
4.5.1	Input Layer	50
4.5.2	Hidden Layer	50
4.5.3	Output Layer	52
4.6	System Implementation	54
4.7	Evaluation	57
4.7.1	Model Training	57
4.7.2	Impact of Phase Calibration	60
4.7.3	Impact of Different System Settings	60
4.7.4	Overall Activity Identification Performance	66
5	Multiple Object Activity Identification using RFIDs	68
5.1	Introduction	68
5.2	M ² AI: Motivation and Overview	71
5.3	Data Preprocessing	73
5.3.1	Phase Calibration	73
5.3.2	Angle of Arrival	74
5.3.3	De-couple Multipath Signals	76
5.4	Deep Learning Design for Activity Identification	79
5.4.1	Input Layer	82
5.4.2	Hidden Layer	82
5.4.3	Output Layer	83
5.5	System Implementation	83
5.6	Performance Evaluation	85
5.6.1	Experiment Setup	85
5.6.2	Activity Identification Performance	86

6 Conclusion and Discussion	93
6.1 Summary of the Thesis	93
6.2 Discussion	94
Bibliography	94

List of Figures

Figure 2.1	Two general scenarios for activity identification with RFIDs	6
Figure 3.1	Our supervised learning framework for mobility detection and activity identification	13
Figure 3.2	Angle of arrival at antenna pair	15
Figure 3.3	Empirical results of existing methods that use the phase difference as a single value	17
Figure 3.4	Our observation that the phase difference vector can be a good indicator for different levels of mobility.	19
Figure 3.5	RSSI of mobility profile	21
Figure 3.6	Rssi variance of mobility profile	21
Figure 3.7	Packet loss rate of mobility profile	23
Figure 3.8	Relative phase fingerprint of mobility profile	24
Figure 3.9	Commercial UHF RFID devices used in experiments	28
Figure 3.10	Accuracy of fine-grained mobility detection with respect to carrier velocities	31
Figure 3.11	Accuracy of mobility detection under various conditions	32
Figure 3.12	Activities and mobility pattern	35
Figure 4.1	The limitation of RSSI and RF phase	38
Figure 4.2	Spectrum for no-activity	39
Figure 4.3	Spectrum for sitting	40
Figure 4.4	Spectrum for walking	40
Figure 4.5	Spectrum for running	41
Figure 4.6	TagFree application examples	42
Figure 4.7	Illustration of the basic idea of TagFree	44
Figure 4.8	TagFree framework	46
Figure 4.9	Details in the data preprocessing stage	49
Figure 4.10	Tag-free architecture	51
Figure 4.11	Details in deep learning stage	53
Figure 4.12	Commercial hardware used to implement TagFree System	54

Figure 4.13	Two typical indoor environments and deployment layouts with the positions of arrays and tags	55
Figure 4.14	Activities in experiments	56
Figure 4.15	Results of phase calibration	58
Figure 4.16	Impact of phase calibration	59
Figure 4.17	Impact of # of antennas	61
Figure 4.18	Impact of # of tags	62
Figure 4.19	Impact of deep learning approaches	63
Figure 4.20	Overall performance of TagFree framework	65
Figure 5.1	M ² AI framework	70
Figure 5.2	The basic idea of M ² AI from a single object to multiple objects of activity identification in the indoor environments.	72
Figure 5.3	Phase jumping caused by frequency hopping	74
Figure 5.4	Illustration of angle of arrival	75
Figure 5.5	Illustration of data preprocessing in our M ² AI framework	78
Figure 5.6	Illustration of spectrum frames in our M ² AI framework	80
Figure 5.7	M ² AI deep learning network architecture	81
Figure 5.8	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.	84
Figure 5.9	Sketches of the tested activities	86
Figure 5.10	Overall activity identification performance	87
Figure 5.11	Impact of phase calibration	87
Figure 5.12	Impact of # of objects	88
Figure 5.13	Impact of different places	89
Figure 5.14	Impact of different distance	89
Figure 5.15	Impact of # of antennas	90
Figure 5.16	Impact of # of tags per person	91
Figure 5.17	Impact of different inputs preprocessing	91
Figure 5.18	Impact of different learning networks	92

Chapter 1

Introduction

1.1 Overview

Over the past decades have seen booming interests in human activity identification that is widely used in a range of Internet-of-Things applications, such as healthcare and smart homes. Yet, sensor or device based solutions [1] require targets carrying sensors/wireless devices that are often not negligible in both size and weight. Radio Frequency Identification (RFID) is a promising technology that can overcome those difficulties, with the advantages of low cost, small form size, and batterylessness, not to mention that RFID tags are able to be read wirelessly and single RFID reader can operate thousands of tags at a time. For example, IKEA Canada has completed a solution that enables shoppers to purchase merchandise with the tap of a spoon that has a built-in tag, freeing shoppers from having to push carts or carry baskets around the store¹. Disney has built an RFID gaming system that can sense when the player is moving objects attached with tags in near real time.

Yet, 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 RFID-based activity identification solutions are far from being satisfactory. It is well known that the RSSI readings almost have no change with small human activities [1], e.g., shaking hands; and the phase angle, though being sensitive to activities, is hardly a reliable indicator. Moreover, real world multipath-rich environments bring more challenges for current RFID based activity identification approaches to be applied in reality. For example, a person is often occluded by furniture and other persons, resulting in that the signals of tags may be deflected and take multiple paths to arrive at the RFID reader. Therefore, the received raw signals are not accurate enough to directly reflect the corresponding activity.

¹IKEA Canada Engages Customers With RFID at Pop-up Store.
[HTTP://www.rfidjournal.com/articles/view?14719](http://www.rfidjournal.com/articles/view?14719)

We however argue that multipath indeed brings rich information that can be explored to identify human activities. Both the backscattered signal power and angle are highly related to human activities, impacting multiple paths with different levels. If we can capture these changing features of the path, the activities could be identified with high sensitivity. Yet, 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, we develop a novel mechanism that jointly considers both pseudospectrum [2] and periodogram [3], so as to take in raw multipath signal mixtures and output carefully de-coupled angles and powers for different paths. We gather massive angle and power information from multiple tags, and preprocess them to extract key features. Yet another challenge is that due to the received signals are a dynamic mixture of multi-path, the features of RFID-based activities are hard to be pre-selected manually, and the rules for making correct estimations are hard to be pre-defined, either.

In this thesis, we note that machine learning, especially deep learning as a new generation technique, can well accommodate this aforementioned challenge, and thus shed new light into the problem for activity identification in a multipath-rich environment. We presented three advanced RFID learning-based activity identification frameworks. In particular, we first introduced *i²tag*, an intelligent mobility-aware activity identification system for RFID tags in multipath-rich environments. Then we presented *TagFree*, the first RFID-based device-free activity identification framework by analyzing the multipath signals. Third, we presented the design of *M²AI* for the first time accommodates both multi-path and multi-object for activity identification. We conducted extensive experiments to evaluate *i²tag*, *TagFree* and *M²AI* in a multipath-rich environments and report significant performance gains over a number of state-of-art feature based approaches. It is also worth noting that our frameworks are 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.

1.2 Thesis Contributions

The contributions of this thesis are summarized as follows:

- Many Radio Frequency Identification (RFID) applications, e.g., virtual shopping-cart and tag-assisted gaming, involve sensing and recognizing tag mobility. Existing RFID localization methods however are mostly designed for static or slowly moving targets (less than 0.3 m/s). More importantly, we observe that prior methods suffer from serious performance degradation for detecting realworld moving tags in typical indoor environments with multipath interference. We present *i²tag*, an intelligent mobility-aware activity identification system for RFID tags in multipath-rich environments, e.g., indoors. *i²tag* employs a supervised learning framework based on

our novel fine-grained *mobility profile*, which can quantify different levels of mobility. Unlike previous methods that mostly rely on phase measurement, $i^2\text{tag}$ takes into account various measurements, including RSSI variance, packet loss rate, and our novel relative-phase-based fingerprint. Additionally, we design a *multiple dimensional dynamic time warping* based algorithm to robustly detect mobility and the associated activities. We show that $i^2\text{tag}$ is readily deployable using off-the-shelf RFID devices. A prototype has been implemented using a Thingmagic reader and standard-compatible tags. Experimental results demonstrate its superiority in mobility detection and activity identification in various indoor environments.

- 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. 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.
- 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 real-world 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. 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 M^2AI , which for the first time accommodates both multi-path and multi-object for activity identification. M^2AI 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 M^2AI is readily deployable using off-the-shelf RFID readers. 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 activity identification accuracy of 97%, a significant gain (27%) over state-of-art solutions.

1.3 Thesis Organization

The remainder of the thesis is organized as follows:

- In Chapter. 2, we presented a literature review on activity identification with RFIDs in both *tag-attached* and *tag-free* activity identification scenarios, where the former directly attaches RFID tags to objects (e.g., human body), and the latter only puts tags on fixed positions in the environment and thus makes objects tag-free.
- In Chapter. 3, we presented the architectural design of i^2tag , an intelligent mobility-aware activity identification system for RFID tags in multipath-rich environments.
- In Chapter. 4, we present **TagFree**, the first RFID-based device-free activity identification system by analyzing the multipath signals.
- In Chapter. 5, we presented M^2AI , the Multipath-aware Multi-object Activity Identification framework, which can identify multiple object activities in typical indoor environments.
- In Chapter. 6, we conclude this thesis and discuss some future works..

Chapter 2

Activity Identification with RFIDs

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 [4], people/object tracking [5] and more complex activity identification [6]. Figure. 2.1 illustrates two general scenarios for activity identification with RFIDs. In tag-attached scenario, an RFID tag is attached to the human body, and the activities are then captured by a tag reader [4] [6]. Tag-free scenario has also been suggested [7]. Instead of attaching tags to human body, which sometimes can be inconvenient and considered intrusive, multiple stationary tags are deployed in the environment 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. We next review literature for both scenarios in Table. 2.1.

2.1 Tag-attached Scenarios

2.1.1 Activity Identification Approaches

Tag-attached scenario is the most straightforward, as RFID tags are suitable to be attached to objects due to its low cost, small for size, and batterylessness. Activity identification solutions exploit the change of wireless signals incurred by the human's actions. RF-IDraw [4] can track human writing by tracking a passive RFID tag attached to his/her pen. Ding *et al.* [6] developed FEMO that uses the frequency shifts of the movements to determine what exercise a user is performing.

2.1.2 RFID Localization Techniques

Accurate RFID localization techniques that have successfully achieved centimeter or even millimeter accuracy are often used to achieve the goal of activity identification. One possibility is to utilize the RSSI of the tag, although RSSI values in backscatter communication

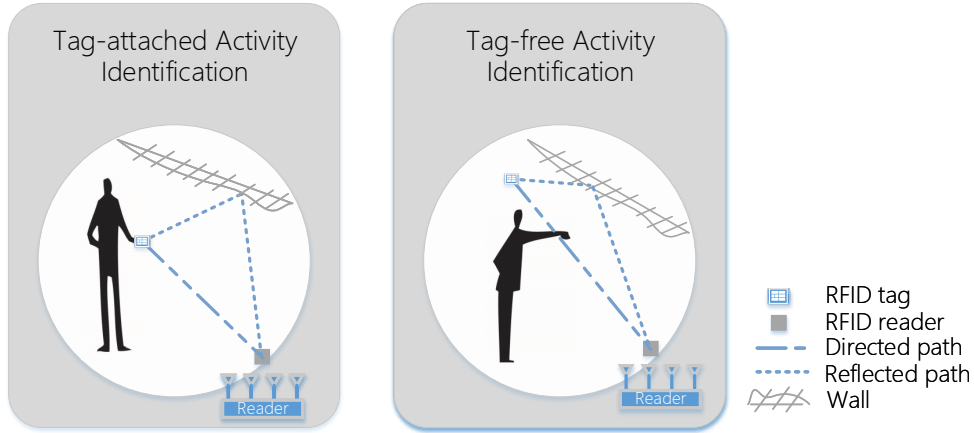


Figure 2.1: Two general scenarios for activity identification with RFIDs

are insensitive to small body movement, and thus difficult to achieve high-precision identification. Making of using phase measurement for wireless localization techniques has successfully achieved centimeter accuracy. We next illustrate the background of the RFID localization techniques and provide a literature review for RSSI-based and Phase-based approaches, as shown in Table. 2.2.

RSSI-based Approaches

Previous work on RF-based positioning primarily relied on RSSI information [1] [8] [9] [10] [11]. The RF fingerprinting, pioneered by Radar [8], employed RSSI based fingerprinting matching against a database to determine the indoor location. LANDMARC [9] introduced the RF fingerprinting technique for localization with fixed location reference tags to help location calibration. Vire [10] used imaginary reference tags, referred to as “virtual tags” to achieve higher accuracy. Later several other improvements over RSSI fingerprinting have been proposed, such as incorporating inertial sensor hints [1]. They typically deployed reference tags on a monitoring region and then used RSSI values to locate a specific tag. The major limitation of RSSI-based approaches is unreliable, since RSSI measured values are highly sensitive to multipath effects, and thus difficult to achieve high-precision localization.

2.1.3 Phase-based Approaches

Phase reflects the distance that a wireless signal traverses in the physical world. There is growing interest in using phase measurement for localization:

Distance ranging: One of the simplest approaches is to calculate the distance between the transmitter and receiver based on received phase measurements. Here, we discuss only some recent and relevant works. Li *et al.* [12] proposed a multi-frequency based ranging method for passive RFID tag localization. Using phase measurement for distance ranging, theoretically, could achieve high localization accuracy. Due to the multipath effects, the

Solutions	Categories	Data	Technical Improvement
RF-IDraw [4]	Gesture Identification	RF phase	employs multiple antennas (8) to eliminate this ambiguity
RFIPad [14]	Handwriting Identification	RSSI & RF phase	transforms a tag plane into a virtual touch screen
FEMO [6]	Activity Identification	RF phase	uses the frequency shifts of the movements
Li <i>et al.</i> [15]	Activity Identification	RSSI	presents a deep learning architecture
TASA [7]	Activity Identification	RSSI	deploys stationary tags as references for location-based activity identification
APID [16]	Activity Identification	Signal energy	uses energy changes of backscatter signals

Table 2.1: Recent research on activity identification with RFIDs

phase measurement is not corresponding to the dominated path, and leads to high ranging error. Liu *et al.* [13] presented an RFID localization scheme by using multiple antennas to receive phase measurements from tags, where the hyperbolic positioning method is employed to correlate phase measurements.

Holography: Holography is the science and practice of making holograms, which is introduced to both the radar and acoustic community for target localization [25]. Miesen *et al.* [17] employ holography to locate a moving tag on a transponder. It achieves an overall accuracy of 7 cm. Parr *et al.* [18] exploited tag mobility and adopted Inverse Synthetic Apertures Radar (ISAR) to generate hologram for tag localization and tracking. Tagoram [5] assumes that the tag movement velocity and its moving track is known in advance, and leverages the tag mobility to construct a virtual antenna array and build a differential augmented hologram using the phase values collected from the antennas. While it fails to address the multipath issue, hence will likely experience practical problems indoors where multipath reflections are prevalent and strong.

Angle-of-Arrival (AoA): Phased-based approaches use antenna arrays or simulated multiple antennas to extract the AoA from RF signals, which can achieve a positioning accuracy on the order of tens of centimeters. AoA information is also employed specifically for RFID localization. Wang *et al.* proposed PinIt [19], which employed a moving antenna to measure the multipath profiles of reference tags at known positions and locates the target tag. PinIt [19] used synthetic aperture radar (SAR) with the moving antenna to extract

Solutions	Categories	Data	Technical Improvement
Radar [8]	Tag-attached	RSSI	employed RSSI based fingerprinting to determine the indoor location
LANDMARC [9]	Tag-attached	RSSI	introduced the RF fingerprinting technique for localization with fixed location reference tags
Vire [10]	Tag-attached	RSSI	used imaginary reference tags
Zee [1]	Tag-attached	RSSI	deployed reference tags on a monitoring region and then used RSSI values to locate a specific tag
Li <i>et al.</i> [12]	Tag-attached	RF phase	proposed a multi-frequency based ranging method
Liu <i>et al.</i> [13]	Tag-attached	RF phase	presented an RFID localization scheme by using multiple antennas
Miesen <i>et al.</i> [17]	Tag-attached	RF phase	employed holography to locate a moving tag
Parr <i>et al.</i> [18]	Tag-attached	RF phase	adopted Inverse Synthetic Apertures Radar (ISAR) to generate hologram for tag localization and tracking
PinIt [19]	Tag-attached	RF phase	employed a moving antenna to measure the multipath profiles of reference tags
STPP [20]	Tag-attached	RF phase	moves the mobile RFID reader with one directional antenna to acquire the spatial order of tags
Yang <i>et al.</i> [21]	Tag-attached	RF phase	proposed a hybrid method for object tracking using noisy passive RFID readings
Tagoram [5]	Tag-attached	RF phase	leveraged the tag mobility to construct a virtual antenna array
D-Watch [22]	Tag-free	RF phase	utilized both the direct path and the reflection paths to track targets
Twins [23]	Tag-free	RF phase	utilized the coupling effect among passive tags
Tadar [24]	Tag-free	RF phase	attached a group of RFID tags on the outer wall and utilized the signals reflected off moving objects

Table 2.2: Related work on RFID localization

the multipath profiles for each tag and leverages the reference tags to locate the target tag. PinIt is not appropriate in our mobile context because the fast-changing environment violates the tag’s multipath profile at every moment, even the movement is very small. STPP [20] was proposed to move the mobile RFID reader with one directional antenna to acquire the spatial order of tags without localizing them. Either the tags or the mobile RFID reader has to move at a constant velocity, while the other kind of device should keep

stationary. Yang *et al.* [21] proposed a hybrid probability model which combines PF with Weighted Centroid Localization (WCL) to achieve high accuracy and low computational cost, but there existed some limitations on the velocity.

2.2 Tag-free Scenarios

The tag-attached approach requires the target to be attached with a tag capable of emitting or reflecting RF signals. This, however, makes the approach not applicable in some scenarios. For example, in intruder detection, the targets will deliberately discard any device that can be tracked. In elderly care, senior people are usually reluctant to hold mobile devices, wear wearables or be attached with RFID tags. These real-life scenarios motivate the needs of tag-free activity identification which does not require any device to be attached to the target. Thus, tag-free activity identification has attracted extensive research interests.

In a tag-free configuration with stationary RFID tags being deployed in the environment as references, e.g., on the walls or on the furniture, the communication link established with fixed readers can be disturbed by human activities in close proximity, hence changing RSSI or phase readings as well. Towards this direction, TASA [7] was proposed to rely on RSSI fingerprints, where reference tags are deployed in a regular way on a monitoring region and training data are generated during the training phase by requiring a person to act in different locations. In the testing phase, resulting RSSI is mapped to the closest fingerprint to identify the status of the person. Such fingerprint-based methods, however, need a large amount of human efforts to acquire and update the fingerprint database. Changes in the environment, such as the movements of furniture, will change the fingerprints, causing mismatches between the database and the new measurements.

Later, Angle of Arrival (AoA) based schemes became popular with the opportunity of multiple antennas attached to a single RFID reader. The Angle of Arrival (AoA) of an RF source is computed by comparing the phases of the received signals at antennas. 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. Tadar [24] presented a system for tracking moving object through walls and behind closed doors using COTS RFID readers and tags. Tadar demands a special RFID reader with a large number of tag arrays, which limits their application in some scenarios. D-Watch [22] efficiently utilized both the direct path and the reflection paths to identify the angle information of the target. Twins [23] leveraged the coupling effect caused by interference among passive tags to detect a single moving subject. APID [16] was proposed to detect arm reaching by analyzing backscatter signals from a passive RFID tag. RFIPad [14] transformed a tag plane into a virtual touch screen by analyzing the induced disturbance of RF signals. However, AoA based schemes may still suffer similar performance degradation when facing

the challenges from the multipath-rich environment [26], as will be further summarized in the next subsection.

2.3 Summary

It is easy to see that although RFID-based activity identification solutions can be categorized into tag-attached and tag-free approaches, where the core technologies and the corresponding challenges are actually similar. For example, in real world indoor scenarios, the received raw RFID signals are the dynamic mixture of many signals from multiple paths, which, as indicated in previous research work [15], may not be accurate enough to be immediately applied to activity identification. Recently, learning-based technique has become a very active research area for general activity understanding. Li *et al.* [15] proposed to directly apply a deep learning approach on collected coarse-grained RSSI readings to range about the tag for activity identification.

Our frameworks well complement these works by demonstrating the necessity and benefits of appropriate data preprocessing on mixed multi-path signals and further proposing a deep learning architecture that can take full advantage of this and maximize the performance gain. In particular, we therefore propose to gather massive angle and power information from multiple tags, and preprocess them for key feature extractions. As the features of RFID-based activities are hard to be pre-selected manually, and the rules for making correct estimations are hard to be pre-defined, we further propose a deep-learning architecture to well handle such dynamics and provide an activity identification framework in a multipath-rich environment.

Chapter 3

RFID Mobility through Intelligent Profiling

3.1 Introduction

The past few years have seen booming interest in human activity identification that provides a range of Internet-of-Things applications, such as healthcare and smart homes [27]. Traditional solutions mainly use radars [28], cameras [29], and various inertial sensors [30]. Yet, sensor or device based radar solutions require targets carrying sensors/wireless devices that are often not negligible in both size and weight. While camera-based and device-free radar-based systems have freed this limitation, they suffer poor performance in accurately identifying multiple objects, especially under Non-Line-of-Sight (NLoS) scenarios. Radio Frequency Identification (RFID) is a promising technology that can overcome those difficulties due to its low cost, small form size, and batterylessness, making it widely used in a range of mobile applications. For example, IKEA Canada has completed a solution that enables shoppers to purchase merchandise with the tap of a spoon that has a built-in tag, freeing shoppers from having to push carts or carry baskets around the store¹. Disney has built an RFID gaming system that can sense when the player is moving or touching objects attached with tags in near real time [31].

The mobility of targets is an essential and important metric to differentiate various human activities [32] [6], e.g., sitting and walking. Yet, the granularity of mobility quantified in existing solutions is not adequate. For example, [32] [33] can only distinguish static and mobile objects, while [6] [4] deal with targets moving at similar speed. Therefore, quantifying the intensity of mobility that is closely related to typical indoor activities is not well addressed yet. One may think of making use of the RFID localization techniques that have successfully achieved centimeter or even millimeter accuracy for mobility detection.

¹IKEA Canada Engages Customers With RFID at Pop-up Store.
[HTTP://www.rfidjournal.com/articles/view?14719](http://www.rfidjournal.com/articles/view?14719)

Unfortunately, while such advanced solutions as RF-IDraw [4] and Tagoram [5] achieve high accuracy through exploring antenna arrays, their performance degrades heavily for indoor environments with multipath. Intuitively, their phase measurement, a core operation, can be remarkably affected by multipath, invalidating the key assumption [34] that the Angle-of-Arrival of the direct path is related to the measurement phase difference between antennas, especially in Non-Line-of-Sight (NLoS) cases. Other localization solutions relying on predeployed reference tags [19] [20] generally require the tagged objects to be static or with limited the moving velocities (i.e., 0.17-0.3 m/s), which is not even applicable for walking (1~1.4 m/s) and running (5~7 m/s). Mobility may also be estimated through the doppler effect [6]. Yet it works with only static communication environments and will again become unstable in fast-changing indoor environments with dynamic multipath, random signal/thermal noise, and varying antenna orientations. Empirically, we show that prior schemes suffer from serious performance degradation for detecting realworld moving tags in typical indoor environments, since using a single parameter for mobility detection is ineffective in multipath scenarios. Our observations motivate us to adopt a profile-based mobility detection that utilizes multiple parameters in tag readings, which is detailed in Section 3.2.

In this chapter, we present i^2 tag, a mobility-aware activity identification system for RFID tags through intelligent profiling, which works robustly in multipath-rich indoor environments. i^2 tag constantly generalizes a huge amount of fine-grained mobility, which further enables us to utilize a supervised learning framework for activity identification as shown in Figure. 3.1. At a high level, it goes through the following major steps:

- **Preprocessing stage.** we employ a novel fine-grained *mobility profile* to quantify different levels of mobility, which seamlessly integrates RSSI variance and packet loss rate, as well as a relative-phase-based fingerprint. The latter is highly effective in distinguishing tag mobility in complicated indoor environments with random signal noise and multipath. By comparing the measured mobility profile against known reference mobility profiles, we detect the tag velocity through a Multiple Dimensional Dynamic Time Warping (MDDTW) [35] algorithm. We classify tag mobility into multiple categories based on the estimated velocity; for instance, *stationary*, *micro-mobility*, and *macro-mobility*.² In this stage, we split tag mobility profile $\mathcal{P}_i = \{p_i^1, p_i^2, \dots\}$ into segments in equal window size τ as $\{\mathbf{p}_i^1, \mathbf{p}_i^2, \dots\}$, which will be transferred into a mobility vector as $\mathbf{v}_i = \{\nu_i^1, \nu_i^2, \dots\}$, where mobility vector as an underpinning unit is applied in a multiclass support vector machine (SVM) [37].
- **Training stage.** Each tag mobility profile \mathcal{P}_i is represented by a corresponding mobility vector \mathbf{v}_i , then we can distinguish different kinds of activities, e.g., sitting, exercising, walking, and running. To be specific, V_{train} in training samples with

²Zhou *et al.* [36] proposed a random mobility model for the different mobile situations, e.g., the user may slowly move the tag although he/she is stationary or his/her movement is confined within a small area.

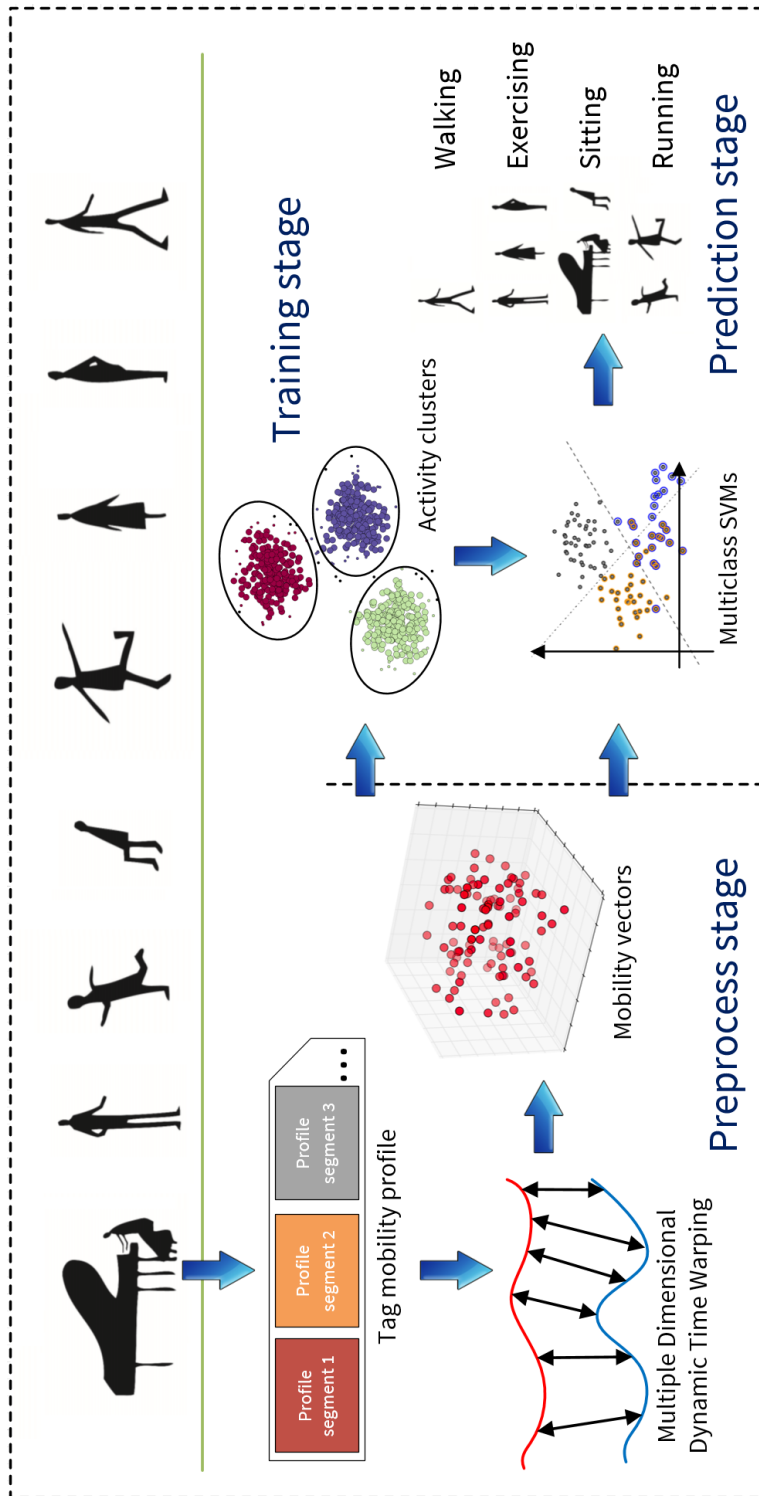


Figure 3.1: Our supervised learning framework for mobility detection and activity identification

corresponding labels will be trained to build the mapping σ from the feature \mathbf{x}_i of mobility vector \mathbf{v}_i to activity label y_i .

- **Prediction stage.** We perform activity recognition in a supervised learning way. For each mobility vector $\mathbf{v}_i \in V_{test}$, we determine whether the feature \mathbf{x}_i of mobility vector is concentrated in certain activities, then label it via σ to achieve corresponding y_i .

i²tag is readily deployable using off-the-shelf RFID readers³ (a single UHF reader with a limited number of antennas) and allows reusing existing RFID readers for indoor activity identification. We have implemented a prototype of **i²tag** using a **Thingmagic** reader and **Impinj** tags, and have conducted extensive experiments in indoor environments. The results demonstrate that, with **i²tag**, a single RFID reader with two connected antennas can accurately distinguish the tag velocity, classify the fine-grained mobility and four categories of activities, with an average detection rate up to 96%.

The rest of the chapter is organized as follows. Section 3.2 illustrates the motivation of our work. Section 3.3 provides our detailed observations on the moving tags properties, then presents activity identification to efficiently solve the problem. Section 3.4 shows the details of implementations. Section 3.5 discusses the performance evaluation results on our approach.

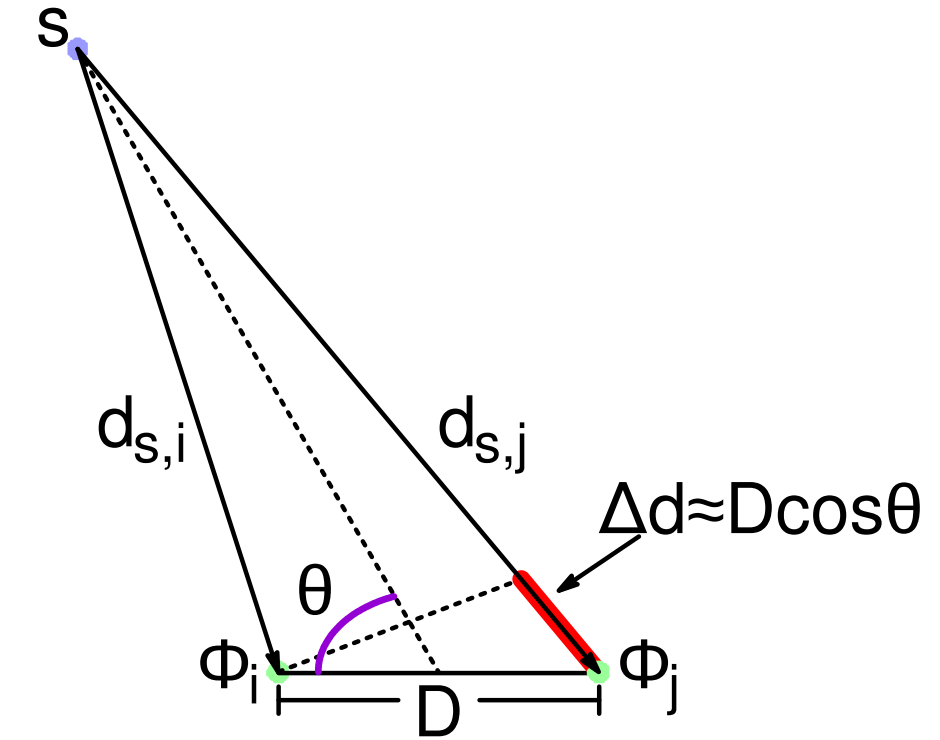
3.2 Why Single Parameter Doesn't Work

3.2.1 Phase Difference as a Single Value

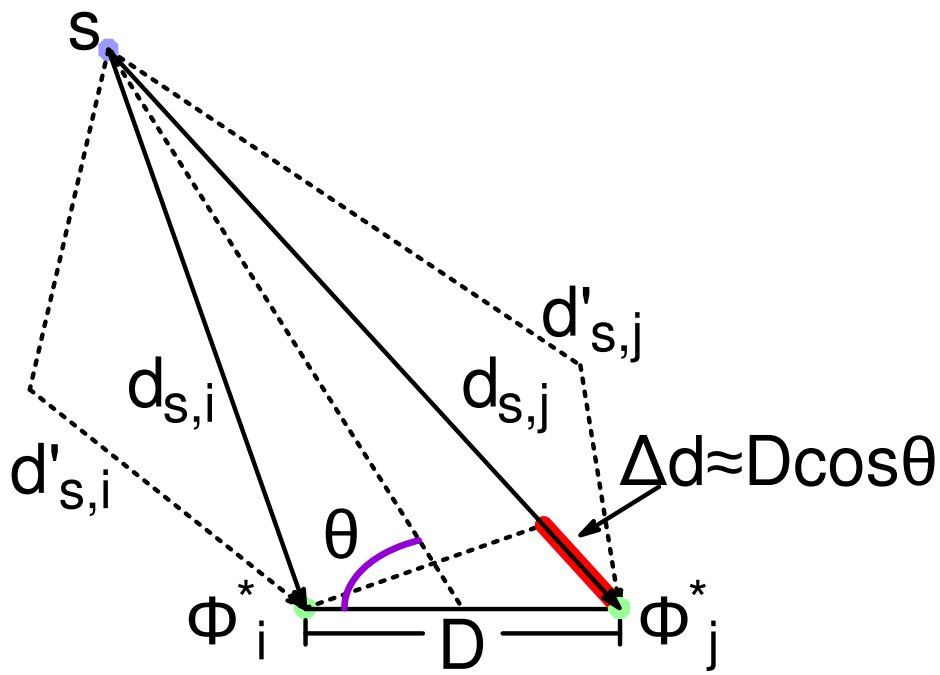
The limited programming interface posed by commercial tag readers⁴ provides only RSSI and phase values. Yet RSSIs are not reliable for location inference, especially for indoor environments, where multipath effects are dominant [38]. On the other hand, the phase value is a relatively reliable choice for deriving statuses of location and mobility. Intuitively, accurate RFID localization can realize tag mobility detection. The tag mobility can be distinguished by the angular velocity, depending on the spatial angle θ as well as the phase difference $\Delta\phi$ [6] [4] [39], as shown in Figure. 3.2(a). In this figure, one 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:

³It is worth noting that the limited programming interface posed by commercial tag readers provides only RSSI and phase values. As such, advanced algorithms for powerful wireless devices are not necessarily applicable here.

⁴ThingMagic M6e RFID reader module. <http://www.thingmagic.com/embedded-rfid-readers>



(a) Single path



(b) Multiple path

Figure 3.2: Angle of arrival at antenna pair

$$\frac{\Delta d_{j,i}}{\lambda} = \frac{\Delta \phi_{j,i}}{2\pi} + k \quad (3.1)$$

where k can be any integer in $[-\frac{D}{\lambda} - \frac{\Delta \phi_{j,i}}{2\pi}, \frac{D}{\lambda} - \frac{\Delta \phi_{j,i}}{2\pi}]$ and D is the distance between two antennas as Figure. 3.2 shows.

However, we find that the above intuition is true only when the multipath effect is negligible. As seen from Figure. 3.2(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 $\alpha_i, \alpha_j, \alpha'_i$ 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:

$$s_i = \alpha \cdot \alpha_i \cdot e^{j\phi_i} \quad (3.2)$$

$$s_j = \alpha \cdot \alpha_j \cdot e^{j\phi_j} \quad (3.3)$$

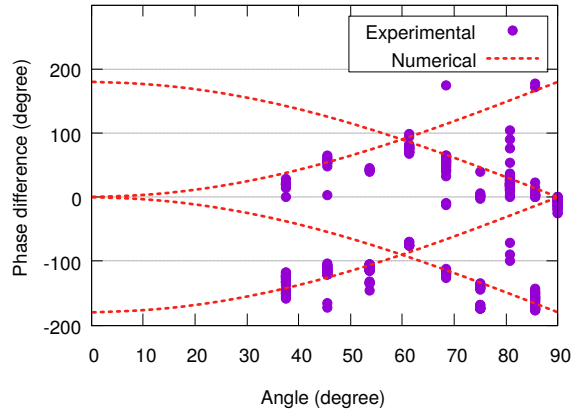
where $\phi_i = \phi_0 + \frac{d}{\lambda} \cdot 2\pi$ and $\phi_j = \phi_0 + (\frac{d}{\lambda} + \frac{D \cos \theta}{\lambda}) \cdot 2\pi$.

$$s_i^* = \alpha \cdot \alpha_i^* \cdot e^{j\phi_i^*} = \alpha \cdot \alpha_i \cdot e^{j\phi_i} + \alpha \cdot \alpha'_i \cdot e^{j\phi'_i} \quad (3.4)$$

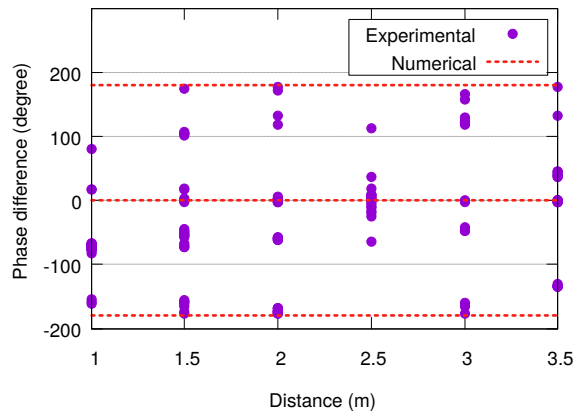
$$s_j^* = \alpha \cdot \alpha_j^* \cdot e^{j\phi_j^*} = \alpha \cdot \alpha_j \cdot e^{j\phi_j} + \alpha \cdot \alpha'_j \cdot e^{j\phi'_j} \quad (3.5)$$

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 $\phi_j^* = 2 \cdot \phi_0 + \frac{d}{\lambda} + \frac{d'_{s,j}}{\lambda} + 2 \cdot \frac{D \cos \theta}{\lambda}$. For instance, we assume $\phi_0 = 0$, $\lambda = 0.33m$, $\alpha = 1$, $\alpha'_i = 0.6$, $\alpha'_j = 0.7$, $\alpha_i = 0.8$, $\alpha_j = 0.9$, $\theta = \frac{\pi}{4}$, $d = 3.3m$, $d'_{s,i} = 3.5m$, $D = 0.165m$ and $d'_{s,j} = 4m$. Then $s_i^* = 0.33 - 0.37i = 0.496e^{-0.8425j}$ and $s_j^* = 0.1325 + 0.5404i = 0.5564e^{1.3304j}$, hence $\phi_i^* = -0.8425$ and $\phi_j^* = 1.3304$. Since $\phi_i = 0$ and $\phi_j = 2.2217$, we have $\phi_i \neq \phi_i^*$ and $\phi_j \neq \phi_j^*$. Obviously, 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.

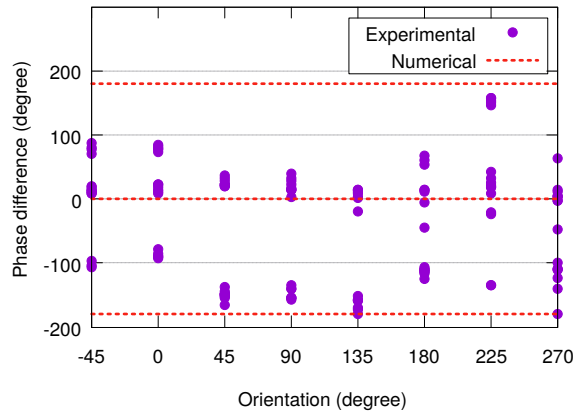
To verify the above hypothesis, we conduct a series of indoor experiments using off-the-shelf tags and the reader by varying positions (spatial angle θ), distances, and tag orientations. The frequency hopping affects phase-angle measurements even for a stationary tag, and thus we fix the channel on the 910 MHz. The results are plotted in Figure. 3.3(a)-(c), which show that the measured phase differences are unreliable, even there exist some experimental results matched with the theoretical benchmark. (i). In Figure. 3.3(a), the red dashed line is the numerical benchmark and the purple dots represent the measured phase difference values $\Delta \phi$ in the experiments. We place the stationary tag at a distance 2m



(a) Spatial angle θ , distance $d = 2m$



(b) Tag orientation, spatial angle $\theta = 90^\circ$, distance $d = 2m$



(c) Tag orientation, spatial angle $\theta = 90^\circ$, distance $d = 2m$

Figure 3.3: Empirical results of existing methods that use the phase difference as a single value

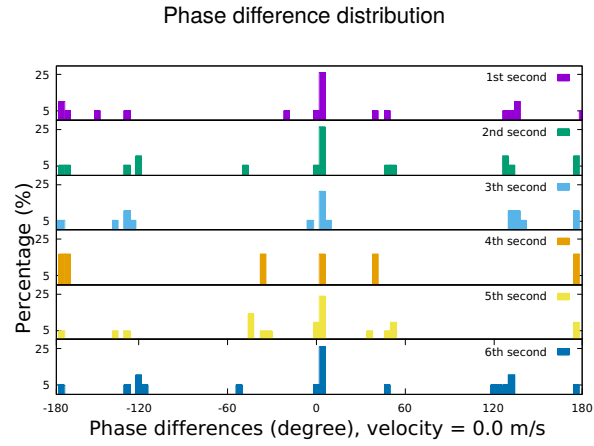
facing to the polarized antennas at different spatial angle θ . It shows the phase differences are not acceptable. We observe that there are significant offsets between the measured phase difference values and the theoretical benchmark; (ii). In Figure. 3.3(b), we put the tag with spatial angle $\theta = 90^\circ$ at different distances. The result demonstrates that the distances have no influences on the phase errors, where the significant phase errors exist at any distances; (iii). In Figure. 3.3(c), the tag orientation is defined as the angle between the reader antenna’s polarization direction and the tag’s antenna. It shows that the tag orientation also introduces the phase errors. Therefore, these measured phase errors as well as random signal noises exist anytime, making phased-based localization ineffective and unreliable. Even worse, a stationary RFID tag can be confused with a tag moving at a high velocity.

3.2.2 Phase Differences as a Vector

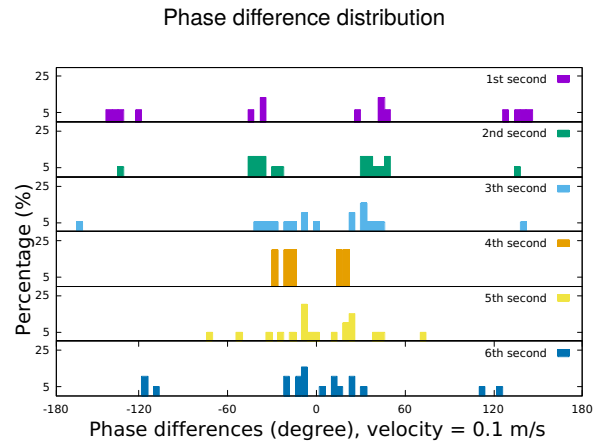
From the above, we know that the phase difference $\Delta\phi_{j,i}$ as the single parameter is ineffective in multipath scenarios. We observe that if we stack the phase differences across a small time interval into a vector, then this vector can be a good indicator of different mobility. To see how this works, we first broadly classify tag mobility into three categories. If the tag is static, it is in the *stationary* status, as Figure. 3.4(a) shows. For the mobile situation, the user may slowly move the tag although he/she is stationary or his/her movement is confined within a small area, e.g., the user may make a telephone call, and a little movement of her head may displace her smartphone. We call that the tag is under *micro-mobility* in Figure. 3.4(b) if it is moving but its location is confined within a small area. Otherwise, tag mobility may also cause the tag to change its location as its user walks from one location to another. In such scenarios as Figure. 3.4(c), we classify the tag to be under *macro-mobility*.

We run three experiments to analyze each kind of mobility in multipath environments. First, we place a stationary tag at a fixed location. Second, for micro-mobility tag, we picked up the tag and moved it around within a meter of its location. Lastly, for evaluating macro-mobility tag, we walked around with the tag in our hand. Figure. 3.4(a) depicts the phase difference for the three categories of tags in 6 seconds. In Figure. 3.4(a), the reader received the signals from a stationary tag, where the phase difference distribution keeps relatively stable. Both the micro-mobility tag with velocity 0.1 m/s in Figure. 3.4(b) and macro-mobility tag with velocity 0.5 m/s in Figure. 3.4(c) return the phase values, where we clearly see that the variance of the phase differences from the macro-mobility tag increases much faster than those of the micro-mobility tag.

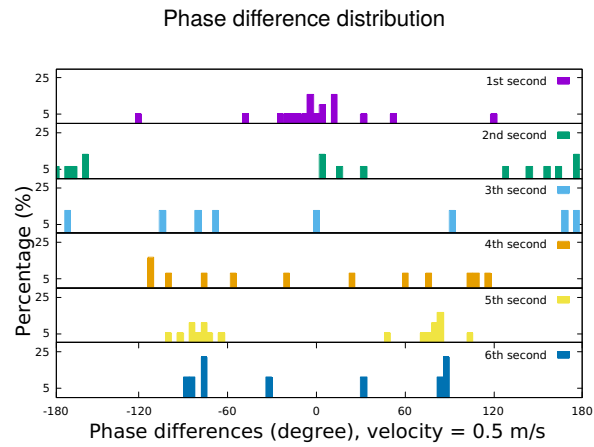
We shall explore more details of the relative phase fingerprint in the next section. To detect the tag mobility, we propose a concept of *relative phase fingerprint*, which denotes the Bhattacharyya distance [40] of the phase difference distribution between two intervals. For illustration, we extract the phase difference distribution at 5 s and 6 s; as can be seen,



(a) Stationary tag



(b) Micro-mobility tag



(c) Macro-mobility tag

Figure 3.4: Our observation that the phase difference vector can be a good indicator for different levels of mobility.

t_i	time i
τ	window size
T	number of periods
$\Delta\phi_i^{t_j}$	phase difference for tag i at time t_j
ψ_i	phase difference density at time i
φ_i	phase difference histogram at time i
r_i^j	RSSI variance of tag i at time j
s_i^j	relative phased-based fingerprint of tag i at time j
e_i^j	packet loss rate during of tag i at time j
p_i^j	measured mobility profile of tag i at time j
\mathbf{p}_i^k	mobility profile segment of tag i
\mathcal{P}_i	mobility profile for tag i
\mathbb{P}	a set of tag profile $\{\mathcal{P}_1, \mathcal{P}_2, \dots\}$
\mathcal{D}	distance matrix
\mathcal{L}	warping path
$C_{\mathcal{L}}$	total cost of warping path
\mathbf{v}_i	mobility vector for tag i
ν_i^j	mobility status
\mathbf{f}_i	mobility feature
y_i	activity label
c_i	activity cluster

Table 3.1: Summary of notations

the stationary tag has much more similarity of phase difference distribution between the two seconds.

3.3 System Design

This section starts from the design of our mobility profile. Then we have shown how to use this profile to effectively detect mobility. Finally, we showcase an accurate indoor activity identification system that builds on our mobility detection scheme.

3.3.1 Mobility Profile

Before we proceed with the detailed solutions for the individual modules of $\mathbf{i}^2\mathbf{tag}$, we first summarize the key notations in Tab. 3.1. The read operation of a commercial UHF RFID reader contains the metadata, namely *measured mobility profile*, about how, where and when the tag was read. The measured mobility profile for each tag read is as follows: antenna ID, read count, timestamp, frequency, phase, and RSSI. We utilize the RSSI, phase and read count for detecting tag mobility.

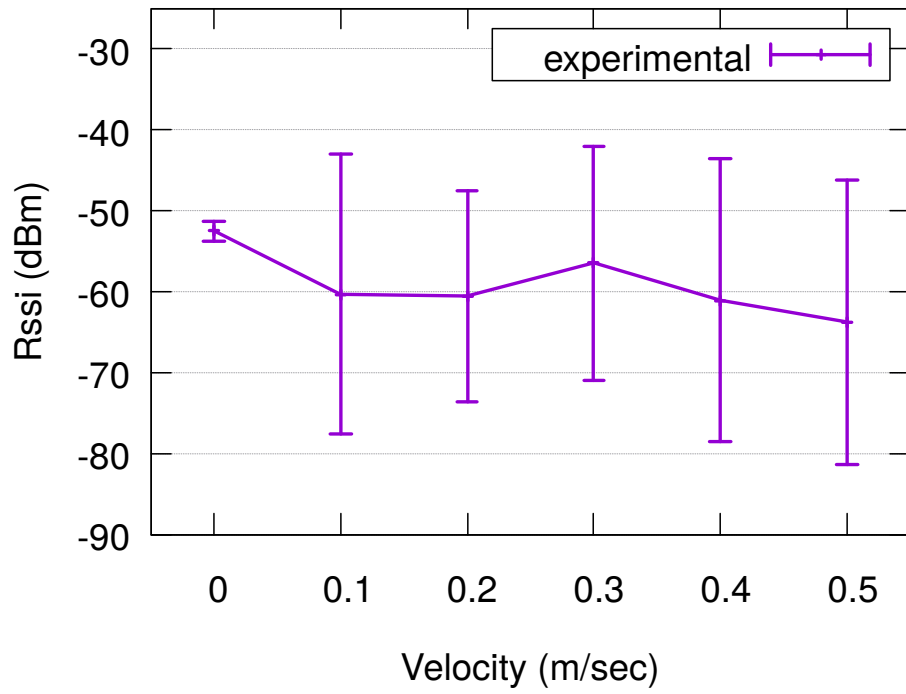


Figure 3.5: RSSI of mobility profile

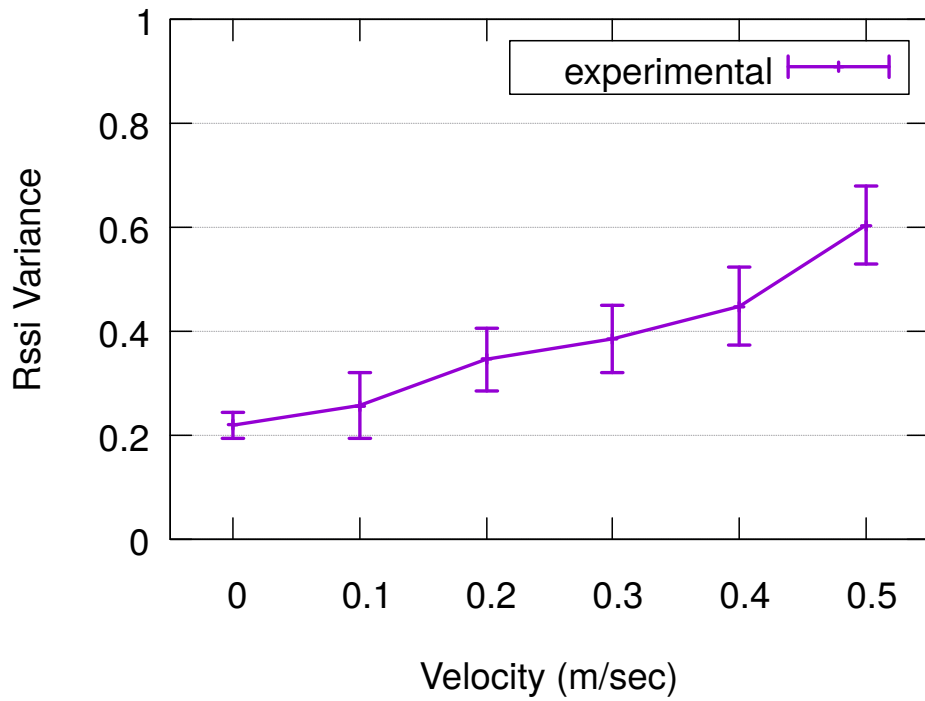


Figure 3.6: Rssi variance of mobility profile

RSSI

One possibility is to utilize the RSSI of the tag, although RSSI values in backscatter communication are not sensitive with the mobility of tags. To provide empirical evidence of the above claim, we measure the RSSI values on ten channels from 910-915 MHz. In our experiments, we found that RSSI is quite stable in stationary scenarios. Yet RSSI is susceptible to any changes in the environment. Often, the RSSI variance under environmental mobility is higher than the observed variation in device mobility.

Figure. 3.5 shows the velocity and corresponding RSSI when the tag is stationary or of other mobility. When the tag is close to the RFID reader, RSSI values are naturally high; yet there are few differences between the stationary tag and moving tags. Therefore, RSSI values cannot be immediately applied in the mobility detection. Fortunately, we observe that the significant differences in RSSI variance between the stationary and moving tags, where we normalize the RSSI variance value between 0 and 1. Although there are multipath in the indoor environment, the RSSI variances of static tags keep relatively stable. There is a significant difference between stationary and moving tags, where the RSSI variances change frequently due to the changing tag position and multipath. Figure. 3.6 shows the RSSI variance at the different tag velocity, where we use normalized standard deviation to represent the RSSI variance, which is range from 0 to 1. Clearly, the RSSI variance can be used to distinguish between stationary and mobility scenarios, where the RSSI variance of velocity 0 m/sec is close to 0.21 and the RSSI variance of velocity 0.2 m/sec jumps to 0.35. The error bars are high for the moving tags; therefore it is difficult to distinguish between micro-mobility and macro-mobility using the RSSI variance.

Packet Loss Rate

Packet loss rate, another important metric in backscatter systems, is the percentage of the maximum number of times that the tag was read during a fixed interval, e.g., one second. Intuitively, mobility and packet loss rate are strongly correlated, since mobility often leads to fast-changing channels.

Hence, the measured packet loss rate can be a dependable indicator of dynamic channel quality. Intuitively, we can use the difference in the loss rates to infer how the tag changes in location or mobility velocities. The experimental results support this hypothesis as shown in Figure. 3.7, which clearly shows that it is straightforward to distinguish between the mobile and stationary case since they have the vastly different packet loss rates. Thus, the packet loss rate is a unique feature of backscatter communication, which involves complementary information about path loss and multipath effects. However, there are still large overlaps between different classes of mobility. Since the loss rate is measured from each of received packets, the moving operation makes the packet loss rate of RFID tag increases rapidly.

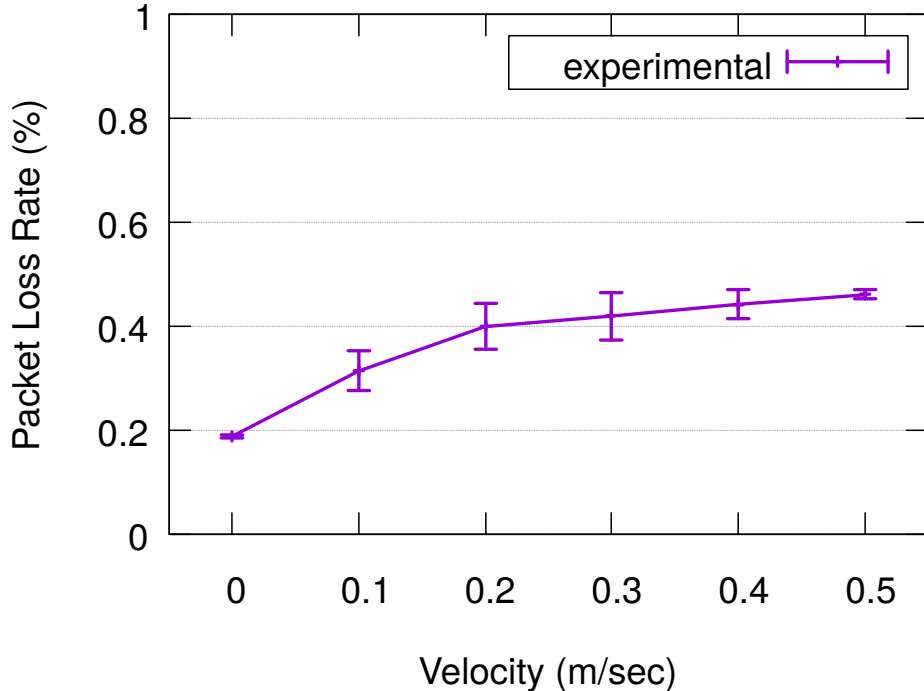


Figure 3.7: Packet loss rate of mobility profile

As such, even if it is possible to use loss rate to distinguish between stationary and macro-mobility, it cannot reliably distinguish between different classes of device mobility.

Relative Phase Fingerprint

We have demonstrated that the measured phase cannot be applied to mobility detection in multipath-free environments. Here, we propose a concept of *relative phase fingerprint* to represent the similarity of phase difference distributions.

Instead of directly using the phase differences, we use histogram formulation to represent the distribution of phase differences at a short interval. The reader antennas receive a set of consecutive signals from the moving tag, where we can capture a set of phase differences $\psi_i = \{\Delta\phi_i^{t_j}, \Delta\phi_i^{t_{j+1}}, \dots\}$ between two antennas. Let ψ_i be the phase difference density of the moving tag, which is discretized into m -bins with the function $\varphi_i = h(\psi_i)$. The histogram φ_i is produced by assigning phase differences $\psi_i = \{\Delta\phi_i^{t_j}, \Delta\phi_i^{t_{j+1}}, \dots\}$ to the corresponding bin.

The estimated state of tag mobility is updated at each time step by incorporating the new observations. Our measurement of the distance between the two phase distributions φ_i and φ_j is based on the Bhattacharyya coefficient [40]. Considering discrete densities such as our phase difference histograms $\varphi_i = \{\varphi_i^1, \varphi_i^2, \dots, \varphi_i^m\}$ and $\varphi_j = \{\varphi_j^1, \varphi_j^2, \dots, \varphi_j^m\}$, the coefficient is defined as

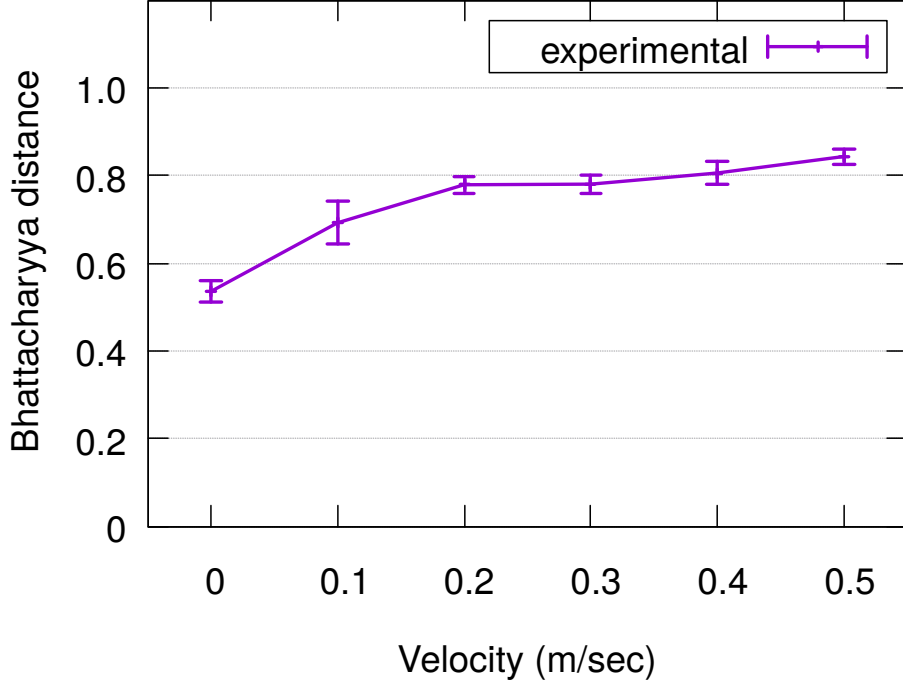


Figure 3.8: Relative phase fingerprint of mobility profile

$$\rho(\varphi_i, \varphi_j) = \sum_{u=1}^m \sqrt{\varphi_i^u \varphi_j^u} \quad (3.6)$$

where m is the number of bins. The larger ρ is, the more similar the distributions are. For two identical normalized histograms we obtain $\rho = 1$, indicating a perfect match. We define the distance between two distributions as

$$d = \sqrt{1 - \rho(\varphi_i, \varphi_j)} \quad (3.7)$$

where d is also called the Bhattacharyya distance [40]. We use this d to quantify the similarity of relative phase fingerprints.

Figure. 3.8 illustrates that the Bhattacharyya distance of the relative phase-based fingerprint can be used to detect the tag velocity. For slow velocity, the Bhattacharyya distance stays low due to the stable environment and slow changes of the phase differences. The Bhattacharyya distance increases once the tag keeps moving. Furthermore, we found that the similarity of fast moving tag (0.4-0.5 m/s) increases faster than slowly moving tag (0.2-0.3 m/s). This happens because a slowly moving tag typically affects only a few multipath components, whereas if the tag itself is moving, all the multipath components will be affected. Therefore, the RFID signal experiences faster variation under macro-mobility than under micro-mobility for the relative phase-based fingerprint.

3.3.2 Mobility Detection

In this section, we introduce an approach to determine mobility statuses of tags. The mobility profile is constantly changing over time with the tag rotation and change of locations. Note that the mobility profile patterns are similar for the same mobility at different rounds but distinctive for different mobility. That said, a particular mobility can be identified by comparing against known profiles.

During the period T of interval length τ , we have the mobility profile set \mathcal{P}_i of RFID tag i with RSSI variance $R_i = \{r_i^1, r_i^2, \dots, r_i^T\}$, relative phase-based fingerprint $S_i = \{s_i^1, s_i^2, \dots, s_i^T\}$ and the packet loss rates $E = \{e_i^1, e_i^2, \dots, e_i^T\}$. We have mobility profile $p_i = \{r_i, s_i, l_i\}$ in interval t_i and mobility profile \mathcal{P}_i as follows:

$$\mathcal{P}_i = \underbrace{\{p_i^1, p_i^2, \dots, p_i^\tau\}}_{1st\ seg}, \underbrace{\{p_i^{\tau+1}, \dots, p_i^{2\tau}\}}_{2nd\ seg}, \dots, \underbrace{\{p_i^{(k-1)\tau+1}, \dots, p_i^{k\tau}\}}_{kth\ seg}, \dots \quad (3.8)$$

where τ is the windows size of segments. Let mobility profile segment \mathbf{p}_i^k represent the k_{th} segment in the mobility profile \mathcal{P}_i .

i²tag detects the tag mobility based on the distance with the multiple dimensional vectors, i.e., the RSSI variance, packet loss rates and relative phase-based fingerprint. To perform multidimensional sequence alignment, **i²tag** employs Multi-Dimensional Dynamic Time Warping [35] to compute the similarity between two mobility profiles. On one hand, MDDTW compares two mobility profiles with different lengths. On the other hand, MD-DTW automatically compresses or stretches a sequence to minimize the distance between two sequences, thus focusing on the shape similarity rather than the absolute values.

We capture the mobility profile based on the tag velocity as the reference $\mathbb{P} = \{\mathcal{P}_1, \mathcal{P}_2, \dots\}$. Then we use the Multiple Dimensional Dynamic Time Warping (MDDTW) technique to match the reference mobility profile against the measured mobility profile. It naturally compensates for the shifts among different mobility profiles caused by varying tag moving velocities. The input of the MDDTW algorithm consists of a reference mobility profile \mathcal{P}_i of length N and a measured mobility profile \mathcal{P}'_j of length M . MDDTW first constructs a distance matrix $\mathcal{D}_{M \times N}$ where each element D_{uv} is defined as the Euclidean distance between element p_i^u and $p'_j{}^v$:

$$\mathcal{D}_{uv} = p_i^u - p'_j{}^v = \|r_i^u - r'_j{}^v\| + \|s_i^u - s'_j{}^v\| + \|e_i^u - e'_j{}^v\| \quad (3.9)$$

where p_i^u and $p'_j{}^v$ are the u_{th} and v_{th} elements of the mobility profiles \mathcal{P}_i and \mathcal{P}'_j , respectively. The MDDTW algorithm find a warping path $\mathcal{L}(\mathcal{P}_i, \mathcal{P}'_j) = \{l_1, l_2, \dots, l_k\}$ such that the total cost $C_L(\mathcal{P}_i, \mathcal{P}'_j)$ of the warping path is minimized:

$$\arg \min_{\mathcal{L}} C_L(\mathcal{P}_i, \mathcal{P}'_j) = \sum_{i=1}^k \mathcal{D}_{l_i} \quad (3.10)$$

ALGORITHM 1: Mobility detection

Input: reference mobility profile set \mathbb{P} , and measured mobility profile set \mathbb{P}'

Output: a set of mobility vector $\{\mathbf{v}_i\}$

```
for each  $\mathcal{P}'_i \in \mathbb{P}'$  do
  Mobility vector  $\mathbf{v}_i = \emptyset$ ;
   $j = 0$ ;
   $\nu_i^j = 0$ ;
  for each  $\mathbf{p}'_i^j \in \mathcal{P}'_i$  do
     $Cost = +\infty$ ;
    for each  $\mathcal{P}_k \in \mathbb{P}$  do
       $C_{tmp} = C_{\mathcal{L}}(\mathbf{p}'_i^j, \mathcal{P}_k)$ ;
      if  $C_{tmp} < Cost$  then
         $Cost = C_{tmp}$ ;
         $\nu_i^j = k$ ;
      end
    end
  end
   $j += 1$ ;
end
  mobility vector  $\mathbf{v}_i = \{\mathbf{v}_i, \nu_i^j\}$ ;
end
Return a set of mobility vector  $\{\mathbf{v}_i\}$ ;
```

where $l_i = (x, y) \in [1 : M] \times [1 : N]$. Then the measured mobility profile \mathcal{P}'_i is classified into different mobility based on the reference mobility profile set \mathbb{P} . Algorithm. 1 shows the workflow to calculate the mobility vector $\mathbf{v}_i = \{\nu_i^1, \nu_i^2, \dots\}$ for each RFID tag mobility profile \mathcal{P}'_i , which contains a sequence of mobility status.

3.3.3 Understanding the Activities

In the preprocessing stage, we first set a window size τ to split the tag mobility profile \mathcal{P}_i into profile segments $\{\mathbf{p}_i^1, \mathbf{p}_i^2, \dots\}$ as illustrated in Figure.3.1. For each profile segment \mathbf{p}_i^j , we treat it as the basic unit and extract its mobility status ν_i^j for labeling. In this section, we will introduce the multiclass SVMs on recognizing and labeling the activities based on the tag mobility vector \mathbf{v}_i . For each mobility vector \mathbf{v}_i , we could simply calculate mobility frequency and denote it as feature \mathbf{f}_i that means the mobility percentage.

When all preparatory steps are done, we can identify the activities in mobility profile \mathcal{P}_i . We assume that there are multiple categories of activities in one specific mobility profile. Specifically, if we have a semantically concentrating profile, for the mobility frequencies, they may have higher variances and lower information entropy. We accordingly formulate the concentrating rating γ as follow:

$$\gamma = \frac{\sum_i^k (f_i - \bar{f})}{\sum_{\mathbf{q}} -q \log(q)} \quad (3.11)$$

in which \mathbf{q} is the normalized form of \mathbf{f} , i.e., $q_i = \frac{f_i}{\sum_j f_j} (f_i \neq 0)$, and $\sum_{\mathbf{q}} -q \log(q)$ here is indeed the entropy of \mathbf{q} . Then, mobility profile segments $\{\mathbf{p}_j^{t_s}, \dots, \mathbf{p}_j^{t_e}\}$ with their concentrating ratings larger than a threshold will be recognized as an activity segment $\mathbf{x}_i = \langle t_s, t_e, \mathbf{f}_i \rangle$, where t_s, t_e indicates the time range of slide, \mathbf{f}_i means the comment frequency on the mobility.

Note that the threshold here is set dynamically in different mobility profiles. We can calculate a series of ratings for the slides and then find the max and min. The threshold is set as $\alpha \times \min + (1 - \alpha) \times \max$, ($0 \leq \alpha \leq 1$), where α is called pass rate and the sensitiveness of α will also be discussed in experimental part.

The set of activity segment $\mathbf{x}_i = \langle t_s, t_e, \mathbf{f}_i \rangle$ are now obtained, and we label each feature \mathbf{x}_i with our preset activity label y_i in a supervised way. We investigate the use of kernel functions to transform the mobility space into a feature space amenable to the Support Vector Machine (SVM) learning methods [37]. SVMs work well in many learning situations since they generalize to unseen data, where the machine is defined by a subset of the training points (i.e., support vectors). In the basic binary classification, SVMs find a hyperplane that provides a maximal separation between two classes. This optimal hyperplane is orthogonal to the shortest line connecting the two classes in their dimensional space, where SVMs maximize the minimal margin. Additional data points, i.e., noises, do not affect the final solution unless they redefine the margin. Therefore, SVMs are amenable to continuous, adaptive on-line learning in activity identification. Multiclass SVMs [37] solves the problem of classifying instances into the more than two classes.

We start with the supervised case. Assume we are given labeled training examples $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$, where each example is assigned a label from a fixed finite set $y_i \in \{1, \dots, k\}$, where k is the total number of categories of activities. Here, we need to extend our feature functions $\sigma(\mathbf{x}, y)$ to include the y -labels explicitly, which provides a separate weight vector \mathbf{w}_k for each class k . Once a complete weight vector has been learned, subsequent test examples \mathbf{x} are classified according to $y^* = \arg \max_y \mathbf{w}^\top \sigma(\mathbf{x}, y)$. The dominant multi-class training procedure for SVMs is formulated as:

$$\begin{aligned}
 w &= \min_{\mathbf{w}, \xi} \frac{\beta}{2} \|\mathbf{w}\|^2 + \xi^\top \mathbf{e} \\
 s.t. \quad &\mathbf{w}^\top (\sigma(\mathbf{x}_i, y_i) - \sigma(\mathbf{x}_i, k)) \geq \delta(y^i, k) - \xi_i, \forall_{i,k}
 \end{aligned} \tag{3.12}$$

where $\delta(y_i, k) = 1_{(y_i \neq k)}$, and w is the multi-class analog of the inverse squared margin. When we get the classifier, every activity segment \mathbf{x}_i can be labeled with a human understandable activity label y_i . In our case, multiclass SVMs [37] are chosen to perform robust and efficient multi-classification.



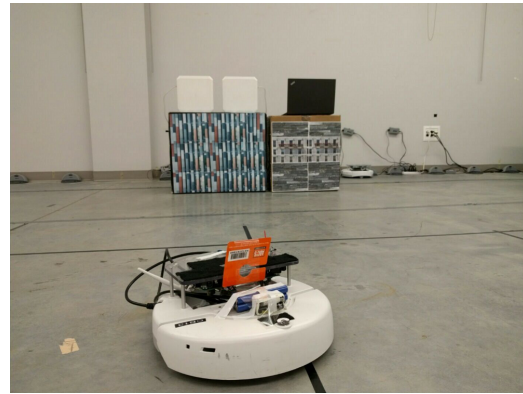
(a) ThingMagic M6e RFID reader



(b) Various RFID tags



(c) Laird Indoor RFID antennas



(d) iRobot Create Programmable Robot

Figure 3.9: Commercial UHF RFID devices used in experiments

3.4 System Implementation

In this section, we describe the key implementation details that are not covered in the previous sections. Our implementation is entirely done based on a commercial reader and requires no modifications to tags. Note that we only highlight the key components here since any real-world activity recognition system requires enormous efforts on implementations [41].

Hardware Settings: Although our system design works with most of the off-the-shelf commercial readers, our prototype implementation uses a Thingmagic reader over other readers (e.g., ImpinJ reader), which have been extensively used in the previous research [19] [20] [6]. The Thingmagic reader works well for mobile applications. For example, the dimensions of a Thingmagic Nano-RFID reader module are $22 \times 26 \times 3.0$ mm⁵, whereas those of an ImpinJ reader R420 are $190.5 \times 175.3 \times 30.5$ mm. Additionally, the ImpinJ reader can report phase readings ranging from to 0° to 360° . In contrast, our ThingMagic M6e 4-port UHF RFID reader ($69 \times 43 \times 7.5$ mm) shown in Figure. 3.9(a), is only able to return phas-

⁵Note that the ThingMagic (<http://www.thingmagic.com>) offers the smallest embedded UHF RFID reader modules.

es ranging from 0° to 180° , which causes ambiguity. Fortunately, our system design does not require accurate phase difference measurement which is necessary for existing methods [19] [20]. As $i^2\text{tag}$ relies on relative phase differences, such ambiguity poses negligible influence.

Another important setting is the distance of two antennas. We connect our ThingMagic M6e reader to the two Laird Indoor RFID Antennas⁶, of which the dimensions are $259 \times 259 \times 33$ mm. Theoretically, the antenna separation D should be spaced by $\lambda/2$, which effectively reduces the ambiguity caused by the high-resolution grating lobes [4]. Due to the available frequencies of RFID, the typical wavelength λ is 0.32786 m (suppose $f = 915$ MHz). Therefore, it is impossible to set D to be smaller than $\lambda/2$ (i.e., 0.16 m). In our implementation, we set D as multiples of $\lambda/2$, $D = 2 \cdot \lambda/2$, which equals to 0.33 m. Although this setting unavoidably introduces ambiguity in phase measurement, unlike RF-IDraw [4] that employs multiple antennas (8) to eliminate this ambiguity, we solve this problem using relative phase differences with only two antennas.

We examine various UHF RFID tags as shown in Figure. 3.9(b), where those different tags have similar performance with phase differences up to 5° . Therefore, we only report the results of a representative type, i.e., Impinj UHF RFID tags, in the rest of this chapter. In our lab, we run the client on a Lenovo laptop (ThinkPad T560), 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 state-of-the-art Dell desktop (OPTIPLEX 7010), each equipped with an Intel Core i7-3770 3.4 GHz quad-core CPU, 8 GB 1333 MHz DDR3 RAM, and a 1 Gbits/sec Network Interface Card (NIC).

Software Settings: The system employs a typical client-server architecture. The processes on clients adopt LLRP protocol [42] to communicate with the reader, and continuously collect the tag readings. The backend module of $i^2\text{tag}$ on the server allows mobile clients to submit the streaming of tag readings, where we store the training results in the MySQL database and execute our algorithms to detect the tag mobility and identify the activity. The Multi-Dimensional Dynamic Time Warping is implemented by C++ language and the multiclass SVMs are implemented based on the Scikit-learn library [43] and LIBSVM [44]. The client is implemented using Java and Mercury API⁷. $i^2\text{tag}$ requires the reader continuously collect tag readings for the further analysis, while using a loop to execute the tag reading operation in a duration leads an excessive delay. Therefore, we utilize the asynchronous reading method *startReading()*, which returns immediately a sequence of RFID reads to the calling thread, then the calling thread uploads the tag readings to the server.

Mobility and Activity Detection: Our experiments include two parts: the mobility detection and activity identification. In the mobility detection, the ground-truth of RFID tag velocity is important for the quality of training set in mobility detection, which also

⁶Laird S9028PCR/S8658PCR (RHCP) INDOOR RFID ANTENNA. [HTTP://rfid.atlasrfidstore.com/](http://rfid.atlasrfidstore.com/)

⁷Mercury API Programmer’s Guide. www.thingmagic.com/

incurs high overhead. Then we employ a carrier attached an Impinj UHF RFID tag as shown in Figure. 3.9(d). The carrier is an iRobot Create programmable robot⁸, of which we can accurately control moving directions and velocities. The robot runs with two powered wheels, while a third passive caster wheel maintains balance. The wheels are controlled independently with a maximum velocity of 500 mm/s. We program the iRobot moving back and forth along a line at a constant velocity. Alternatively, the robot may move along a circle, yet the velocity is uncontrollable and difficult to measure. We examine the effectiveness of tag mobility detection with different forwarding velocities ranging 0.0 m/s to 0.5 m/s. If the velocity is smaller than 0.1 m/s, the tag would be deemed *stationary*. For velocities ranging from 0.1 m/s to 0.4 m/s, we classify them as *micro-mobility*, whereas velocities that are greater than 0.4 m/s are deemed as *macro-mobility*.

In the activity identification, we invite ten volunteers and each volunteer⁹ is attached to an Impinj UHF tag on one’s hand. The volunteers stand 2-5 meters away from the reader antennas in our experiments¹⁰. To conduct a comprehensive evaluation, we test four typical indoor activities, i.e., sitting, exercising, walking and running. In each activity case, the RFID reader continuously queries RFID readings for ten minutes. Then we evaluate the accuracy of activity identification and further explore mobility distributions for the four categories of activities. Furthermore, the volunteers conduct several different kinds of activities in one duration. They are used to evaluate the robustness of *i²tag* for identifying randomly changing activities.

3.5 Evaluation

We conduct experiments in a typical office, which is a multipath-rich environment. We evaluate the performance of *i²tag* in terms of accuracy, effectiveness, and overhead.

3.5.1 Mobility Accuracy

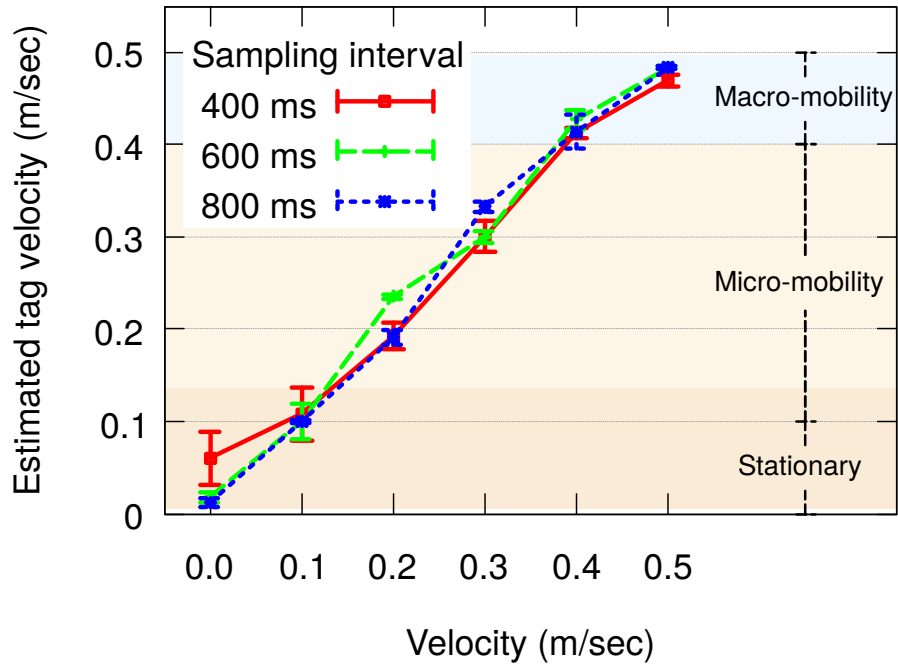
Figure. 3.10 shows the result of fine-grained mobility detection accuracy with respect to carrier velocities varying from 0.0 m/s to 0.5 m/s. In particular, Figure. 3.10(a) plots the performance of *i²tag* with different sample intervals. The accuracy is low for short sampling intervals, because the phase differences may not be stable even under *stationary* status. Figure. 3.10(b) illustrates that the larger the detection window size is, the greater the accuracy achieves. But the large detection windows size will delay the mobility detection. In this work, we identify three broad categories of tag mobility based on velocities.

Then we further evaluate the accuracy of mobility detection. The performance of detecting mobility depends on the sampling period in Figure. 3.11(a). The accuracy is low for

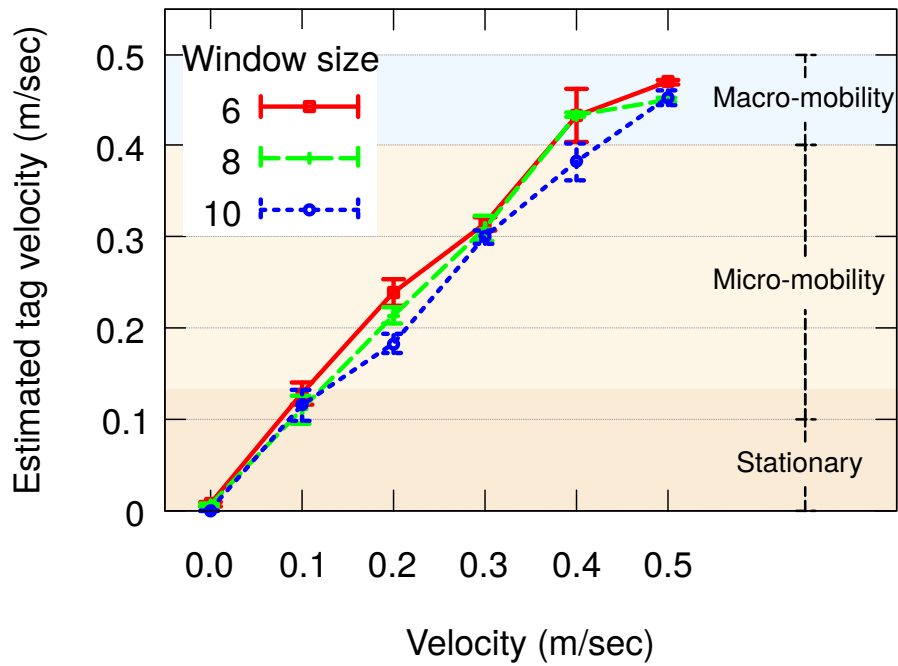
⁸iRobot Create Open Interface (OI) Specification. [HTTP://www.irobot.com/](http://www.irobot.com/)

⁹Note that those volunteers are varied in age, gender, height, and weight.

¹⁰The commercial RFID reader’s range limits the range of our current prototype. Beyond 5 meters, the RFID tag cannot harvest enough energy to wake up.

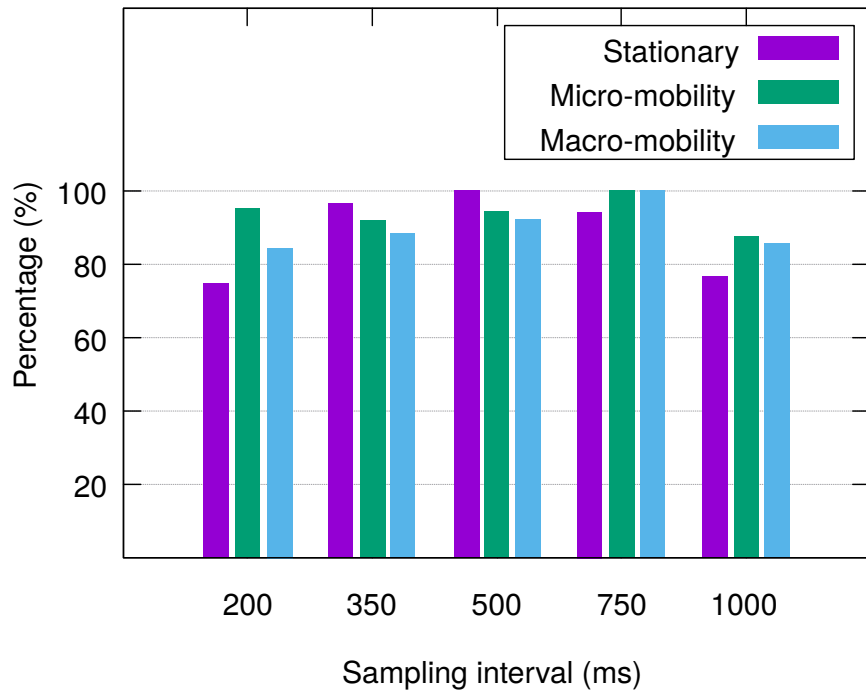


(a) Mobility detection for various sampling intervals

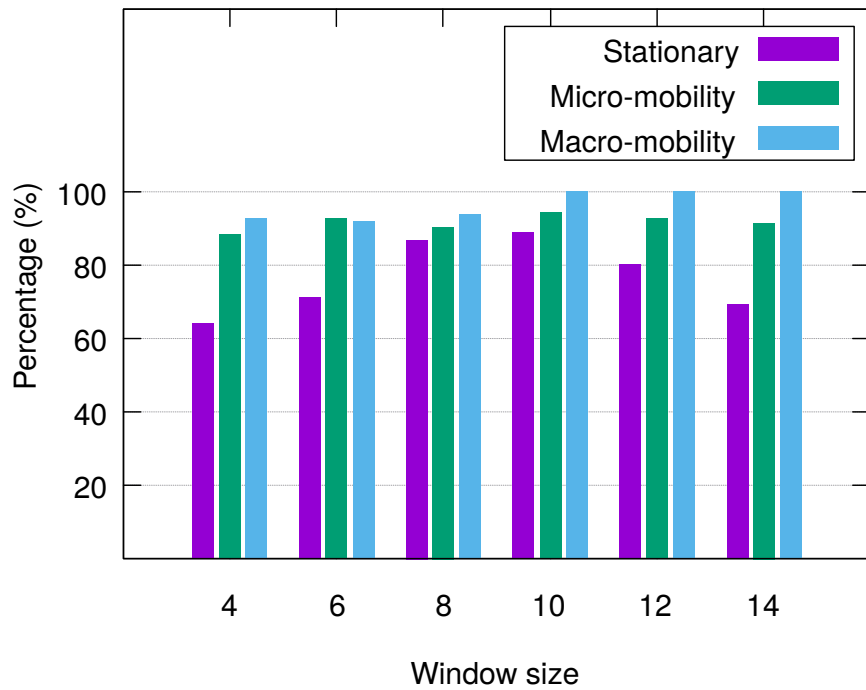


(b) Mobility detection for various windows sizes

Figure 3.10: Accuracy of fine-grained mobility detection with respect to carrier velocities



(a) Mobility detection for various sampling intervals



(b) Mobility detection for various windows size

Figure 3.11: Accuracy of mobility detection under various conditions

short sampling period because the RFID signal of the stationary tag may change very quickly under multipath effects. We use a sampling period of 750 ms in the rest of our evaluation, yielding a median accuracy of 96%. The larger detection window size make the accuracy higher for the moving tag in Figure. 3.11(b). Meanwhile, large detection windows will delay the macro-mobility detection. Nonetheless, we find that a detection window of 8 yields a satisfactory accuracy of 98%, and hence we use this setting in the rest of experiments.

We next evaluate the robustness of $i^2\text{tag}$ with two citations, i.e., tag orientation and distance. **Impact of orientation:** The tag orientation is defined as the angle between the reader antenna’s polarization direction and the tag’s antenna. To understand the effect of tag orientation, we conduct 6 experiments on the fixed frequency, 915 MHz. To measure its influence on the detection accuracy, we adjust the orientation from 0° to 360° . As expected, the result remains at the same level. **Impact of distance:** We evaluate the accuracy with varying distances from 1 m to 3 m. $i^2\text{tag}$ does not exhibit clear correlation with the distance. Therefore, the distance is not a crucial factor affecting the accuracy. Especially, a mean error distance of 5 mm can be obtained, when placing the antenna at a distance of 0.3 m. In fact, it is more reasonable to model the antenna as a point locating at its centroid when it keeps far away from the tag.

3.5.2 Activities and Mobility

Tab. 3.2 shows the results of activity recognition for a single tag. Each row denotes the actual activity performed and each column represents the activity recognized by $i^2\text{tag}$. Each element in the matrix represents the percentage of activities in the row, which is recognized as the activity in the column. As shown in the table, the average accuracy is 83.15% for four activities. This shows that we can extract rich information about the tag mobility and activities. The result clearly shows that $i^2\text{tag}$ achieves a high and stable activity recognition performance, due to its efficient mobility detection and robust activity cluster algorithms. The average accuracy of identifying activities is 83.15%, where the slow activity identification have the accuracy up to 94.62%. The above results show that $i^2\text{tag}$ can distinguish a set of activities with high accuracy.

To understand the effect of multiple activities in one sequence of tag readings, we conduct four experiments, where there are two kinds of activities operating in order. In each experiment, an activity is performed for 5 minutes. $i^2\text{tag}$ can clearly distinguish those activities based on the tag mobility distribution as shown in Tab. 3.3. We further have a detailed look at these experiments as shown in Figure. 3.12. Figure. 3.12(a) shows that walking and running have no stationary status, where activity has approximately 90% macro-mobility and walking only has 25% macro-activity. Figure. 3.12(b) illustrate the high percentage of stationary and micro-mobility represents the activity is in a small area. Figure. 3.12(c) illustrates the same kind of activity that has the similar percentage of stationary and mobility status, where two volunteers just walked and did some daily routines. Figure. 3.12(d) shows

	Identified activities percentage (%)			
	Sitting	Exercising	Walking	Running
Sitting	94.62	5.38	0	0
Exercising	12.50	87.50	0	0
Walking	0	15.00	75.00	10.00
Running	0	0	24.44	75.56

Table 3.2: Activity identification of single RFID tag

	Identified activities percentage (%)			
	Sitting	Exercising	Walking	Running
Walking	0	0	70	30
Running	0	0	10	90
Sitting	96	4	0	0
Exercising	0	80	20	0
Exercising	10	78	12	0
Exercising	8	82	10	0
Exercising	0	85	15	0
Walking	0	0	90	10

Table 3.3: Activity identification of two RFID tags

that micro-mobility and macro-mobility are effective to distinguish different intensities of mobility.

3.5.3 Realtime Performance

Intervals (ms)	Window size					
	4	6	8	10	12	14
200	0.025	0.032	0.036	0.041	0.050	0.054
350	0.022	0.021	0.025	0.029	0.033	0.037
500	0.018	0.018	0.021	0.024	0.031	0.031
750	0.017	0.018	0.021	0.023	0.028	0.031
1000	0.018	0.018	0.021	0.024	0.025	0.031

Table 3.4: Computation Complexity

$i^2\text{tag}$ provides online mobility-detection and activity identification, where the statistical information is displayed in Tab. 3.4. $i^2\text{tag}$ takes an incremental process to generate the relative phase-based fingerprint. After receiving a successful response from the reader, $i^2\text{tag}$ produces intermediate tag features and superimposes them to the Multiple-dimensional Dynamic Timing Warping processing. The read time is the interval during which the reader interrogates two consecutive rounds of reading. It is an upper bound which should be taken for producing an intermediate result. The median read time is 33 ms, and any computation exceeding this bound might affect the real-time performance. Theoretically,

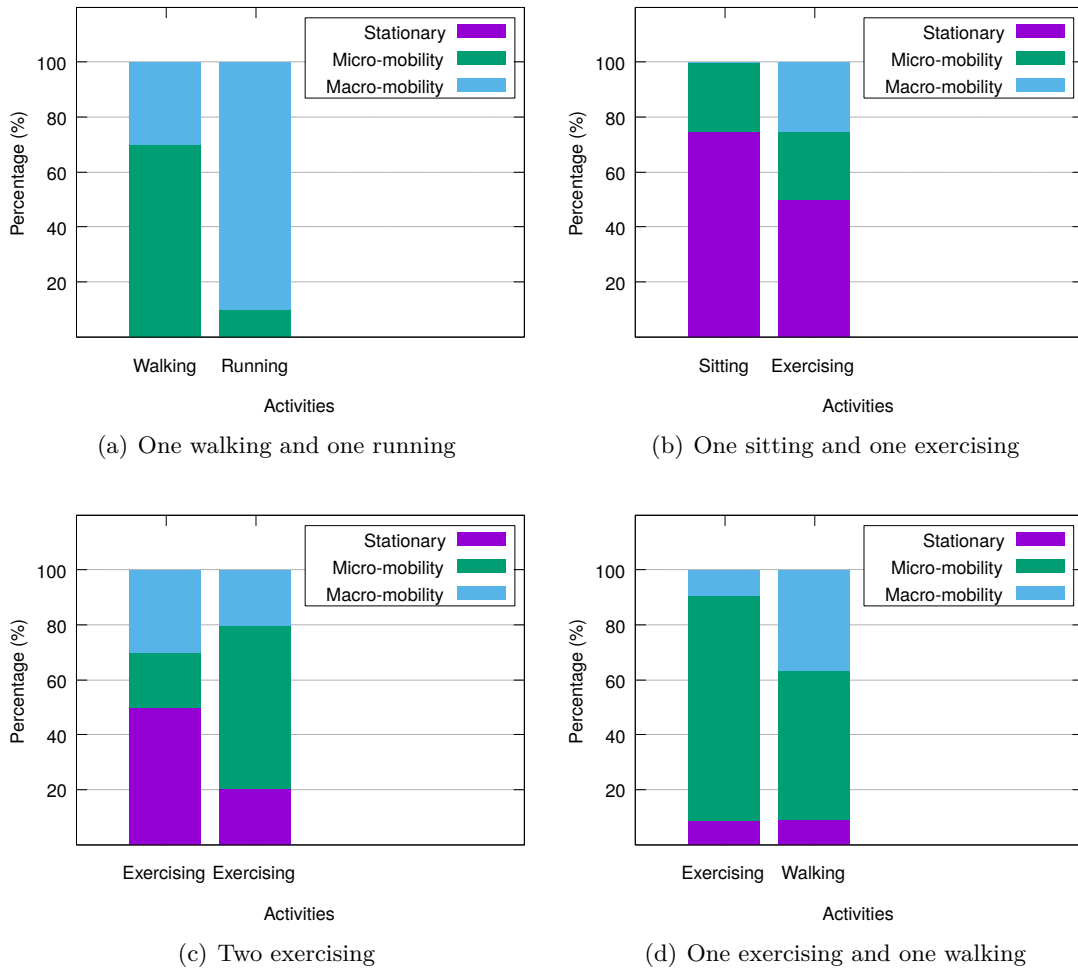


Figure 3.12: Activities and mobility pattern

the fast implementation of Multiple Dimensional Dynamic Timing Warping (MDDTW) provides optimal or near-optimal alignments with an $O(n)$ time and memory complexity. It shows that i^2tag achieves a recognition latency of 30 ms on average. Therefore, we can conclude that i^2tag can provide real-time activity identification results.

Chapter 4

TagFree Activity Identification with RFIDs

4.1 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 [27]. It has received significant attention from both academia and industry, with diverse solutions based on radars [28], cameras [29], inertial sensors [30], 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 [20] [4] [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 [7] [45]. 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) [46] and Long Short Term Memory (LSTM) network [47] 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).

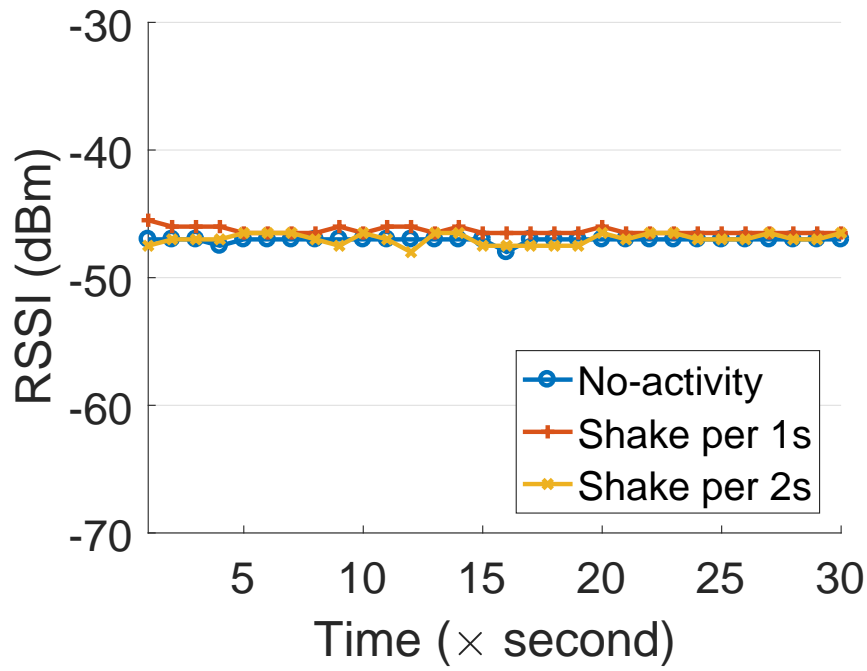
We conduct extensive experiments to evaluate our TagFree in multipath-rich environments and report significant performance gains over different state-of-art feature-based solutions. With an alert mechanism, TagFree can be customized as a monitor for senior patients in healthcare facilities, e.g., identifying 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 4.2 presents the challenges in tag-free activity identification, and Section 4.3 illustrates the basic idea of our work. Section 4.4 provides our data pre-processing scheme on dealing with frequency hopping and de-coupling multipath signals, and then Section 4.5 presents our deep learning approach for activity identification. Section 4.6 discusses the implementation details. The performance evaluation results on our approach are presented in Section 4.7.

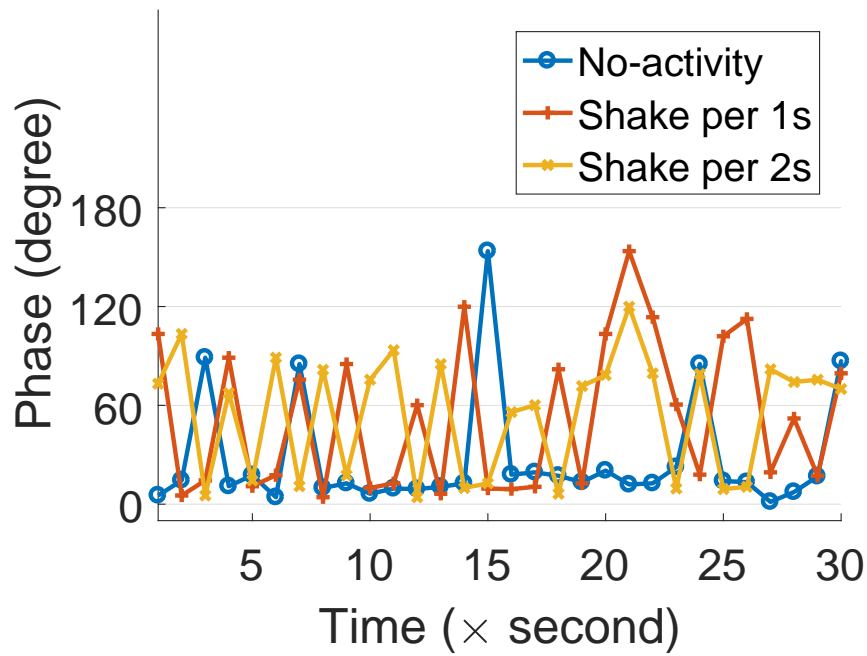
4.2 Raw RSSI or Phase? Challenges for Tag-free Activity Identification

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

¹Thingmagic M6e reader.<http://www.thingmagic.com/>



(a) RSSI is insensitive



(b) Phase is sensitive but inaccurate

Figure 4.1: The limitation of RSSI and RF phase

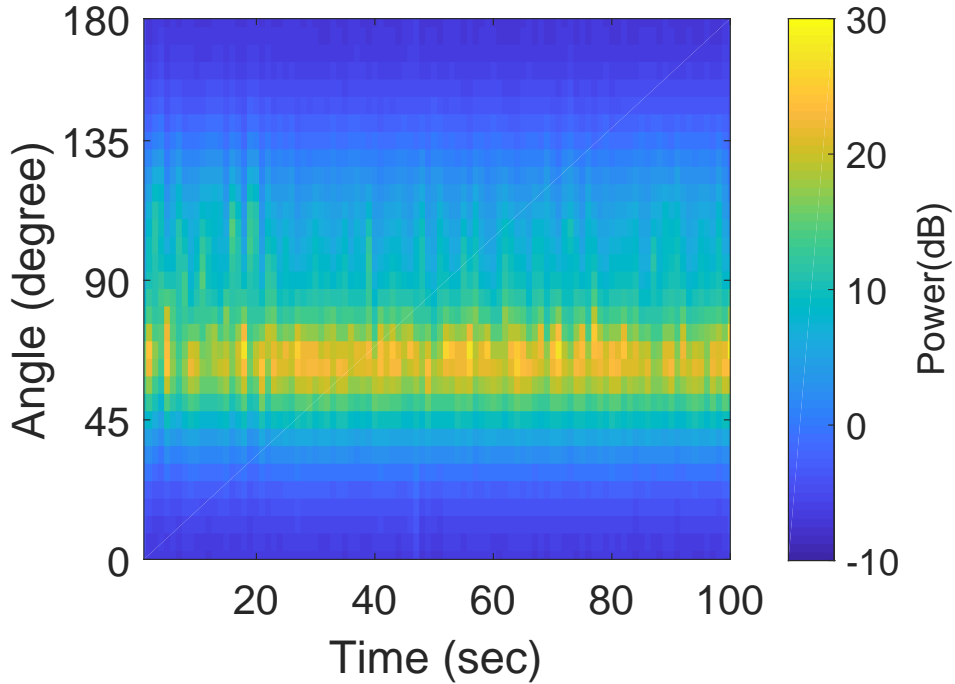


Figure 4.2: Spectrum for no-activity

close proximity, hence changing the RSSI or phase readings as well. For instance, TASA [7] 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 Figure. 4.1 (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 Figure. 4.1 (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 [48] [15]. The phase measurement is affected by multipath, which is inevitable in an indoor environment [4] [5]. As can be seen from Figure. 4.1 (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.,

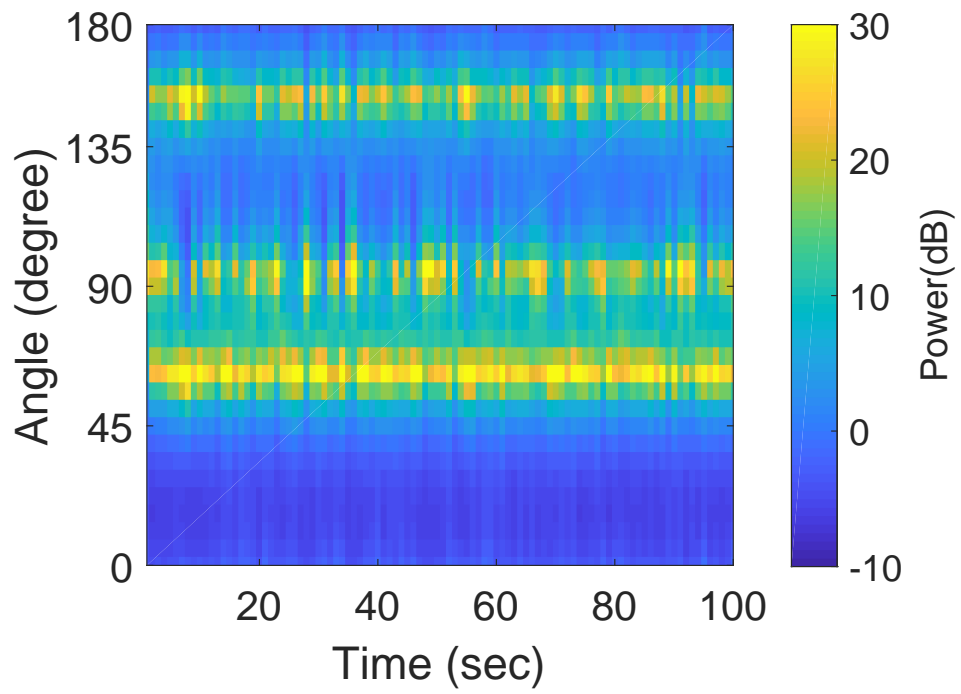


Figure 4.3: Spectrum for sitting

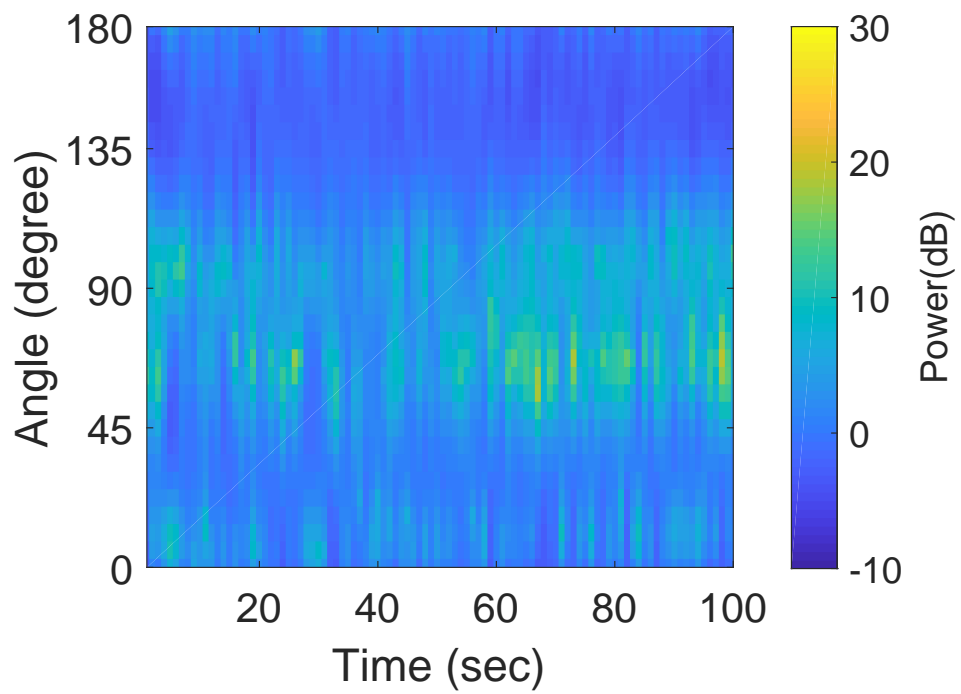


Figure 4.4: Spectrum for walking

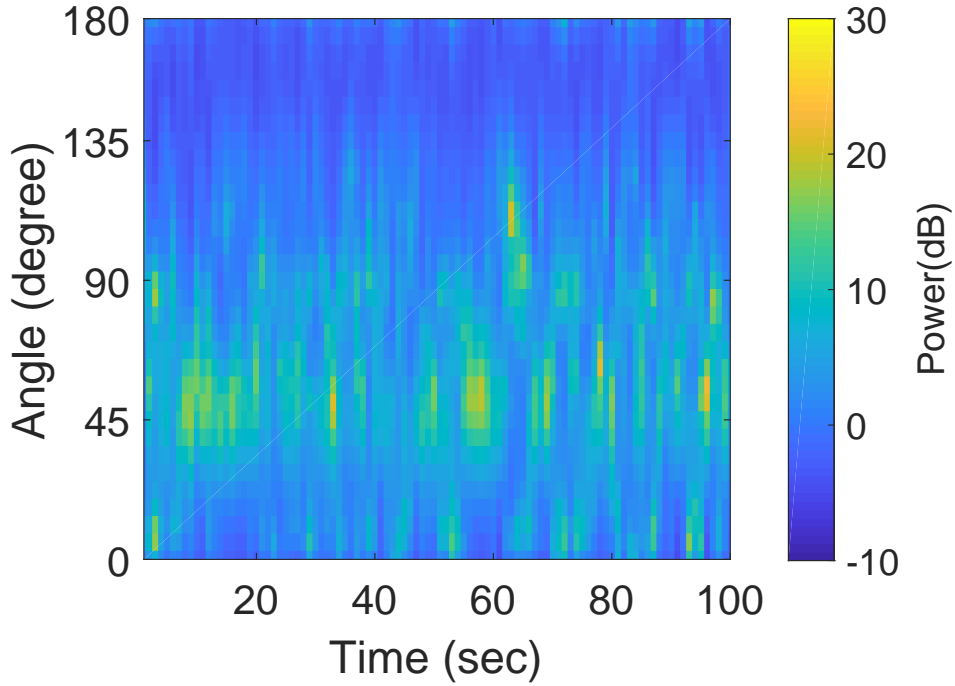
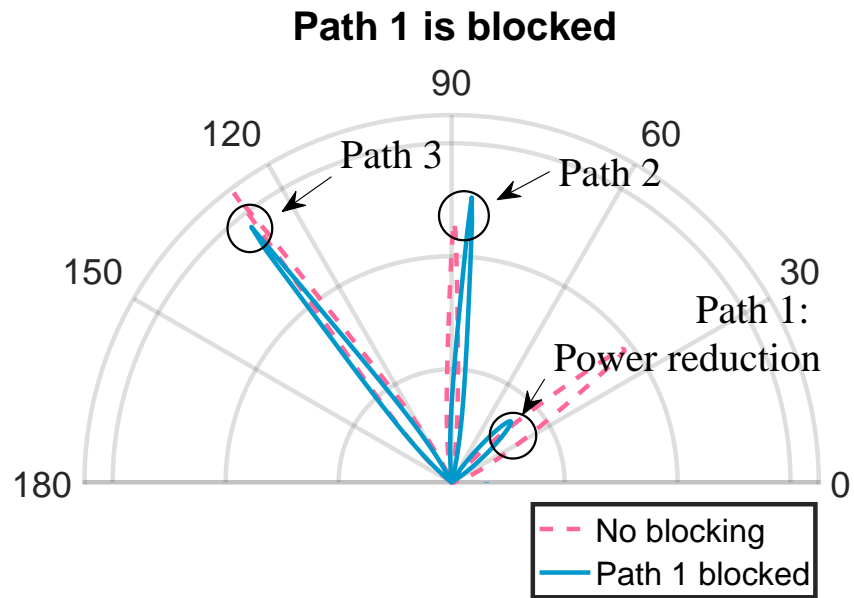


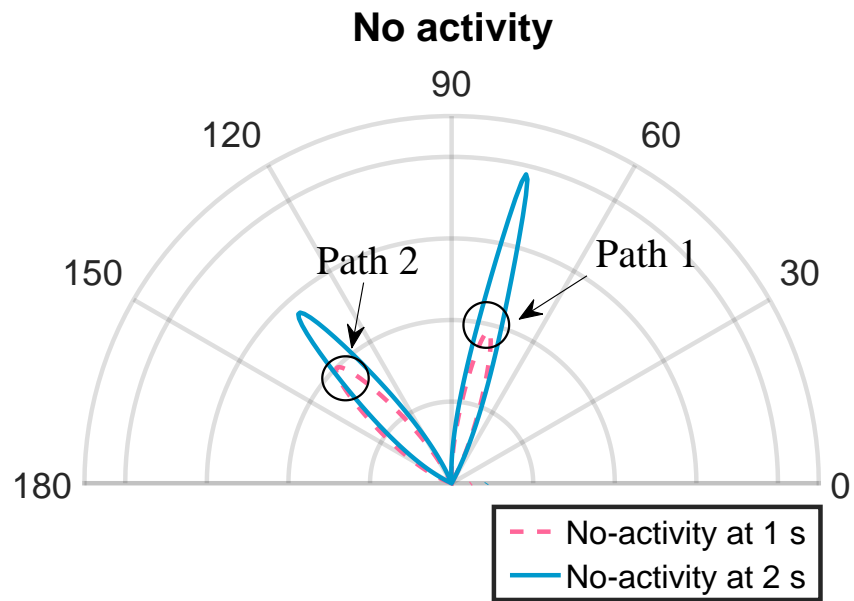
Figure 4.5: Spectrum for running

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 phase through the MUSIC (Multiple Signal Classification) algorithm [2]. This algorithm has been widely used in the multi-path context for calculating the directions of arriving signals. Consider an illustrative example in Figure. 4.2, 4.3, 4.4 and 4.5. 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. Figure. 4.2 shows the no-activity scenario and there is only one significant direct path in the environment. The activity of sitting is shown in Figure. 4.3, 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 Figure. 4.4. Figure. 4.5 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.

We further have a detailed look into the experiments for sitting and no-activity, respectively. Figure. 4.6 (a) demonstrates the spectrums in 2 seconds for the sitting activity,



(a) Spectrum for sitting



(b) Spectrum for no-activity

Figure 4.6: TagFree application examples

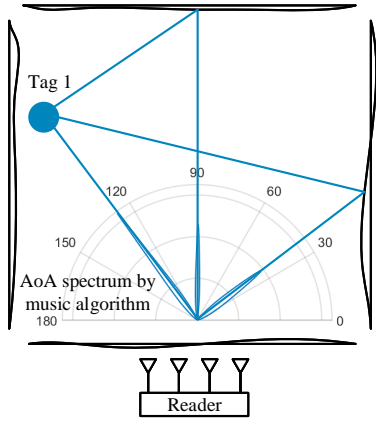
where there exist three signal paths, namely, path 1, 2, and 3. When the person blocks path 1 at 40° , the power of path 1 is decreased, which successfully detects the activity along this directly affected path. Meanwhile, the angles and power of the other two paths are affected as well, though with 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 [2]. 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 Figure. 4.6 (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.

4.3 TagFree Overview

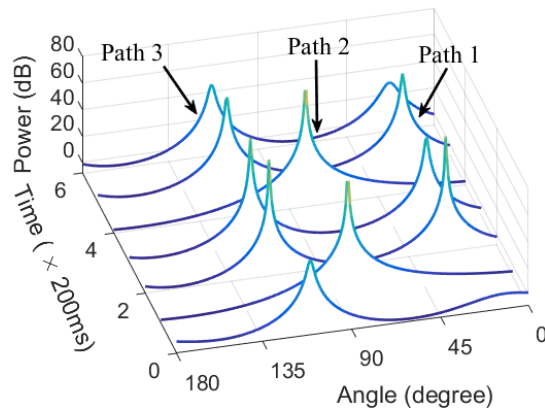
Figure. 4.7 shows the basic idea of our TagFree system. In Figure. 4.7 (a), the spectrum shows that there exist three paths from a stationary tag, where the stationary Tag 1 continuously reflects the signals with the same angle and power from 40° , 90° and 125° , respectively. Figure. 4.7 (b) shows a 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. Figure. 4.7 (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.

To verify the above hypothesis, we conduct a series of field studies to examine the scenarios in Figure. 4.7 (a)-(c).² The results are plotted in Figure. 4.7 (d)-(f), which clearly demonstrate that the angle shift and power reduction are highly related to user activities: (i) In Figure. 4.7 (d), we place a stationary tag with a distance of 3 m facing the uniform linear array at the 125° direction (as shown in Figure. 4.7 (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 Figure. 4.7 (e), we keep the tag at the same place and one person blocks the signal’s propagation along path 1(as shown in Figure. 4.7 (b)). The peak of blocked path 1 experiences a clear drop and the angle of path 2 shifts frequently. (iii) In

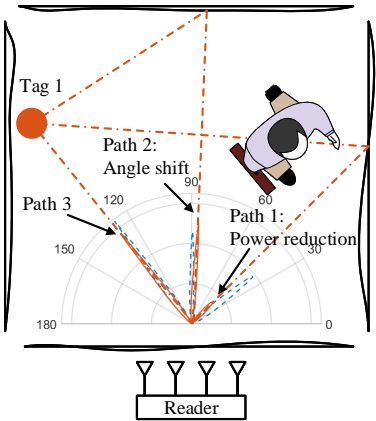
²For ease of exposition, here we disable frequency hopping, which will be further examined in Section 4.4.



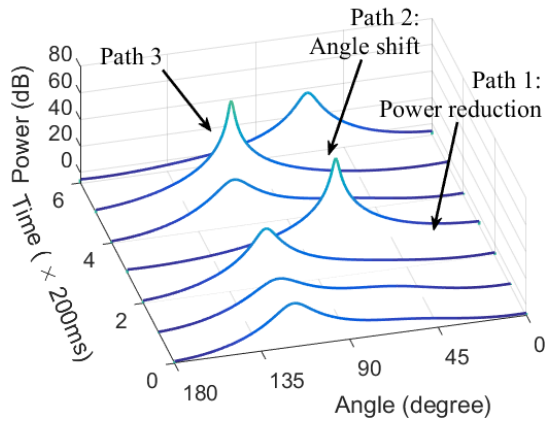
(a) scenario.I



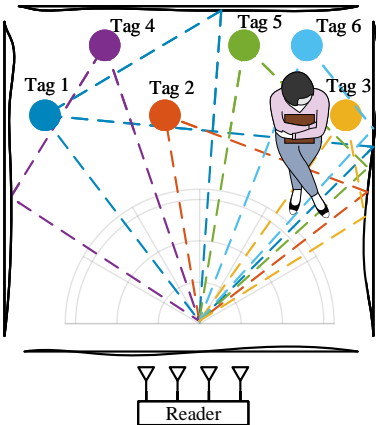
(b) Spectrum stacks in scenario.I



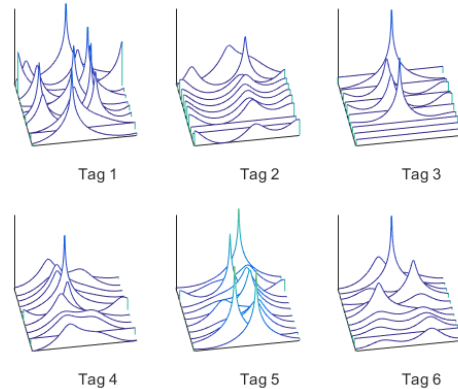
(c) scenario.II



(d) Spectrum stacks in scenario.II



(e) scenario.III



(f) Spectrum stacks in scenario.III

Figure 4.7: Illustration of the basic idea of TagFree

d	distance between two antennas	x_t	input at time t
θ	spatial angle	\mathbf{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
$\mathbf{s}(t)$	source signal vector (M by 1)	γ	activity cluster
$\mathbf{r}(t)$	received signal vector (N by 1)	\mathbf{R}_r	correlation matrix of received signals
$\mathbf{w}(t)$	noise vector (N by 1)	\mathbf{R}_s	correlation matrix of source signals
$\mathbf{a}(\theta)$	steering vector (N by 1)	\mathbf{U}_s	signal subspace
\mathbf{A}	matrix of steering vectors (N by M)	\mathbf{U}_n	noise subspace

Table 4.1: Summary of notations

Figure. 4.7 (f), we place six tags in front of the antenna array (as shown in Figure. 4.7 (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 Figure. 4.6 (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 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) [47] 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 Figure. 4.8. 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.

4.4 Data Preprocessing

Before we proceed with the detailed solutions for the individual modules, we first summarize the key notations in Table. 4.1.

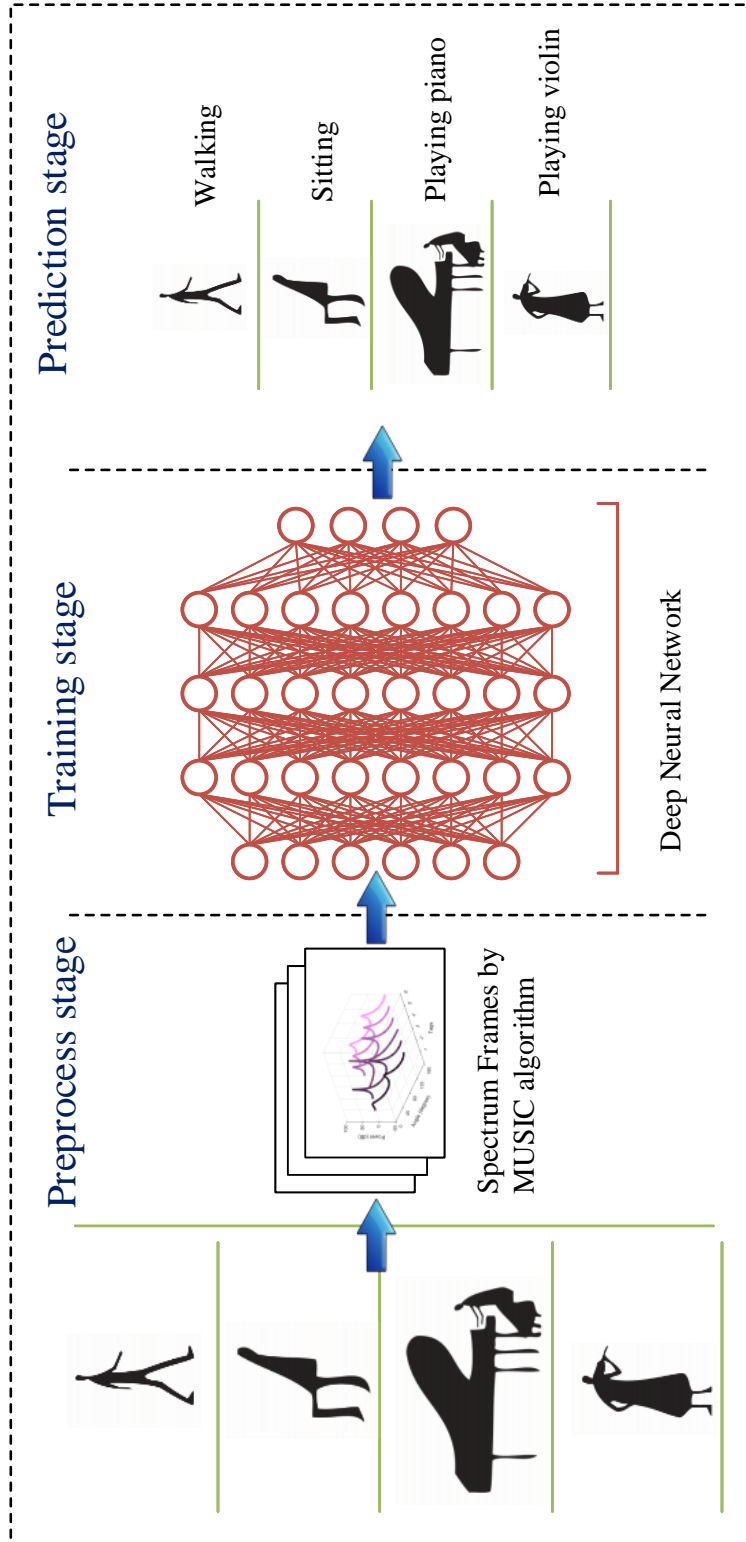


Figure 4.8: TagFree framework

4.4.1 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 [49] 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 Figure. 4.9 (a), where the phase and frequency relation follows a linear model. These experiments imply 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:

$$\phi(t) = \phi^j(t) - \tilde{\phi}^j + \tilde{\phi}^r \quad (4.1)$$

4.4.2 Multipath De-coupling

Intuitively, as shown in Figure. 4.9 (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 ϕ_2 , we can then estimate the AoA θ as:

$$\theta = \arccos\left(\frac{|\phi_1 - \phi_2|}{\pi}\right) \quad (4.2)$$

However, in practice the AoA estimation may not work well because of the multipath effect. To this end, we adopt the MUSIC (MUltiple SIngle Classification) algorithm [2], which is the best known algorithm based on eigenstructure analysis of an $N \times N$ correlation matrix \mathbf{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 $s_1, \dots, s_M(t)$ a from different

angles $\theta_1, \dots, \theta_M$, and into a uniform linear array (ULA) of N antennas. Then we have

$$\mathbf{s}(t) = [s_1(t), \dots, s_M(t)]^\top \quad (4.3)$$

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

$$\mathbf{a}(\theta) = \left[1 \quad e^{-j\frac{2\pi d}{\lambda} \cos(\theta)} \quad e^{-j2\frac{2\pi d}{\lambda} \cos(\theta)} \quad \dots \quad e^{-j(N-1)\frac{2\pi d}{\lambda} \cos(\theta)} \right]^\top \quad (4.4)$$

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

$$\mathbf{r}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{w}(t) \quad (4.5)$$

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

$$\mathbf{R}_r = E\{\mathbf{r}(t)\mathbf{r}^H(t)\} = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \mu^2\mathbf{I} \quad (4.6)$$

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

$$[\mathbf{U}_s \mathbf{U}_n] = \underbrace{[\mathbf{U}_1, \dots, \mathbf{U}_M]}_{\mathbf{U}_s}, \underbrace{[\mathbf{U}_{M+1}, \dots, \mathbf{U}_N]}_{\mathbf{U}_n} \quad (4.7)$$

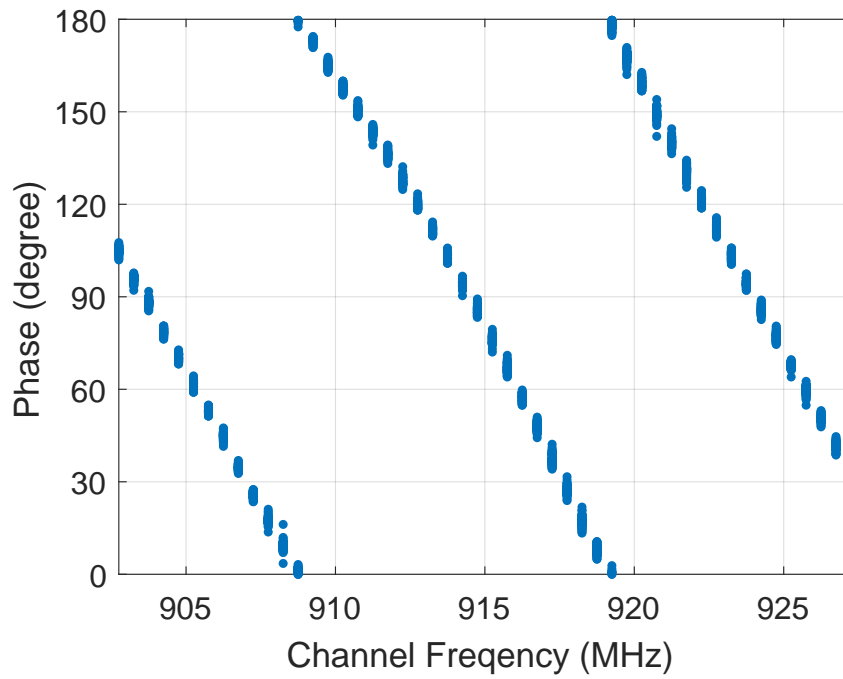
The MUSIC algorithm utilizes the orthogonality relationship between the signal and noise subspaces [2], which implies:

$$\mathbf{a}^H(\theta)\mathbf{U}_n = \mathbf{0} \quad (4.8)$$

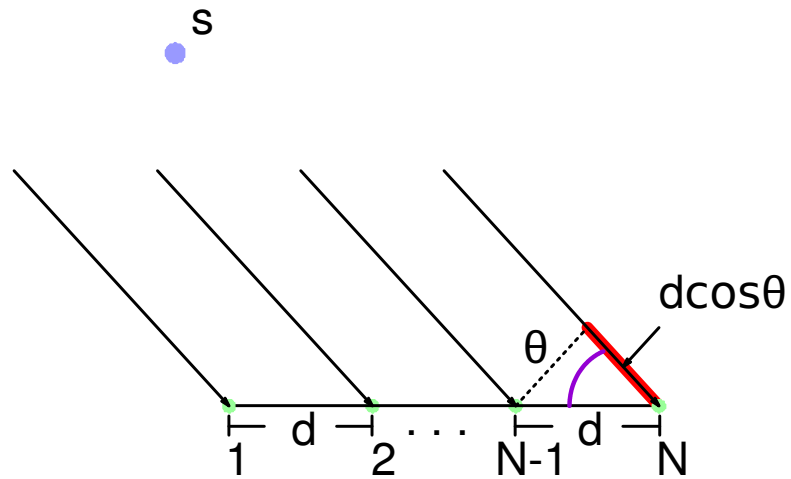
The direction of the arrival angle can be represented in terms of the signal sources and the noise subspaces as shown in the following equation:

$$\theta_{MUSIC} = \arg \min_{\theta} \mathbf{a}^H(\theta)\mathbf{U}_n\mathbf{U}_n^H\mathbf{a}(\theta) \quad (4.9)$$

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



(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.

Figure 4.9: Details in the data preprocessing stage

$$P_{MUSIC} = \frac{1}{\mathbf{a}^H(\theta)\mathbf{U}_n\mathbf{U}_n^H\mathbf{a}(\theta)} \quad (4.10)$$

The M higher peaks are of great power [2] and corresponds to the estimated direction of arrival of the signal source with the angles $\theta_1, \dots, \theta_M$.

4.5 Deep Learning for Activity Identification

This section describes the main components of our TagFree design. As illustrated in Figure. 4.10, 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.

4.5.1 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 Figure. 4.11 (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 $\mathbf{x} = (x_1, \dots, x_T)$, where a series of spectrum frames along the time will further serve as the initial input for the hidden layer.

4.5.2 Hidden Layer

Our hidden layer includes a CNN (Convolutional neural network) structure for effective object classification and detection [46], and an LSTM (Long Short-Term Memory) structure for activity identification [47]. Such layers contain rich *implicit patterns* [50], 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 [51] [52] [53] [54]. 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. 4.7.

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

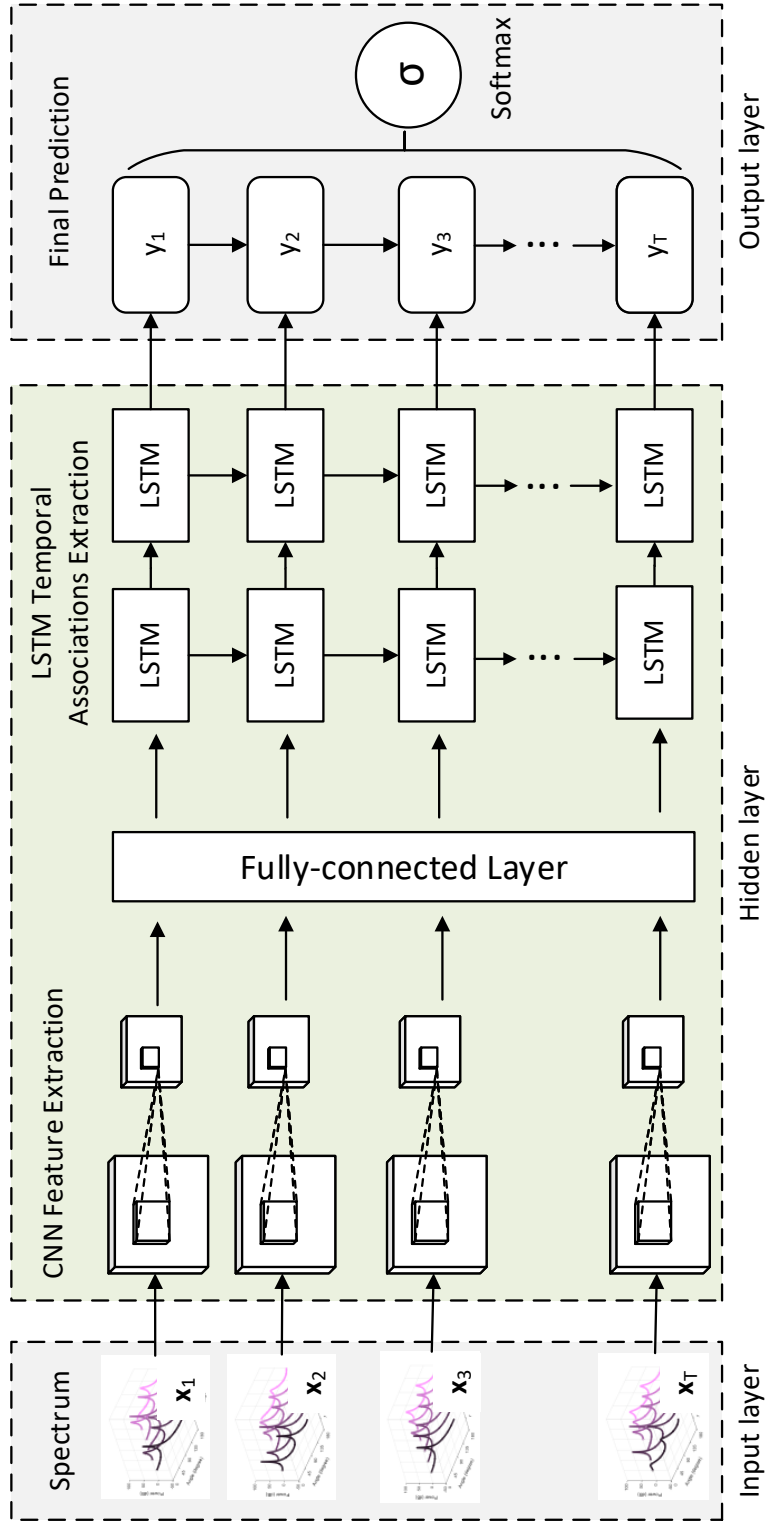


Figure 4.10: Tag-free architecture

these features are outputs of rectified linear units (RELU) [55]. 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 [52], and we applied dropout in all the convolutional layers and the fully connected layer.

LSTM Structure

Figure. 4.11 (b) shows a single LSTM cell [47]. 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.

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

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (4.11)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (4.12)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (4.13)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (4.14)$$

$$h_t = o_t \tanh(c_t) \quad (4.15)$$

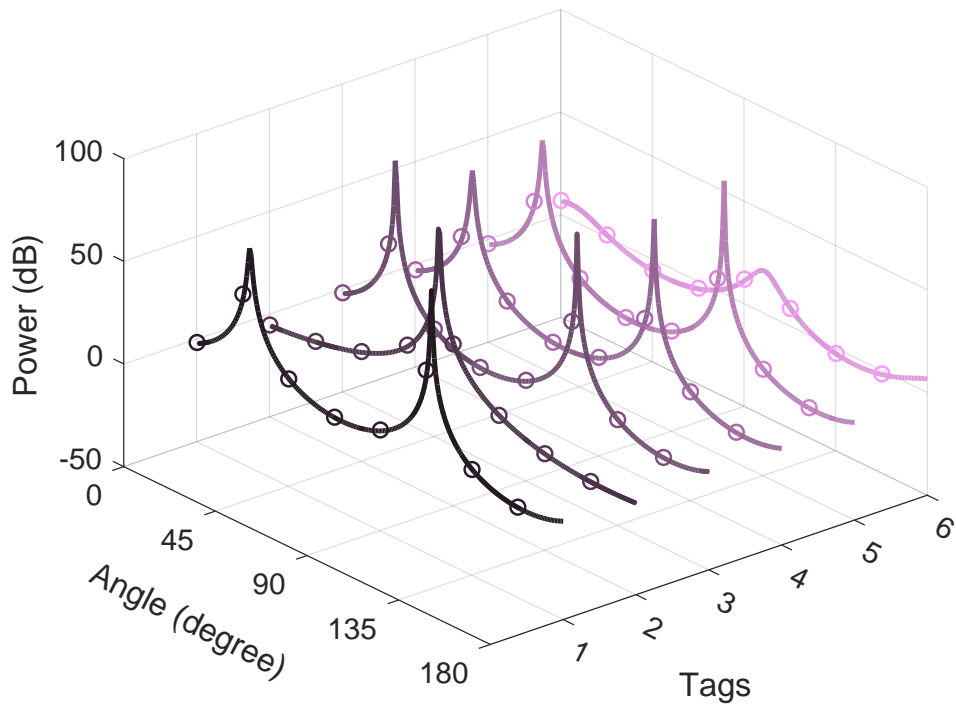
where the W terms denote weight matrices (e.g. W_{hi} is the input-hidden weight matrix), and the b terms denote bias vectors (e.g. b_f is the bias vector of forget gate).

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.

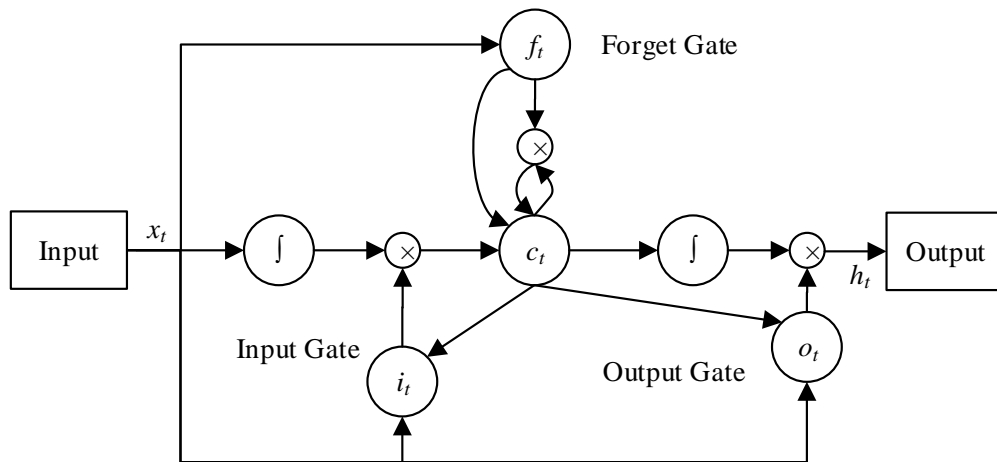
4.5.3 Output Layer

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 :

$$Pr(y|x_i) = \frac{e^{x_i}}{\sum_{y' \in \gamma} e^{x_i}} \quad (4.16)$$



(a) Spectrum Frame



(b) The architecture of LSTM cell

Figure 4.11: Details in deep learning stage

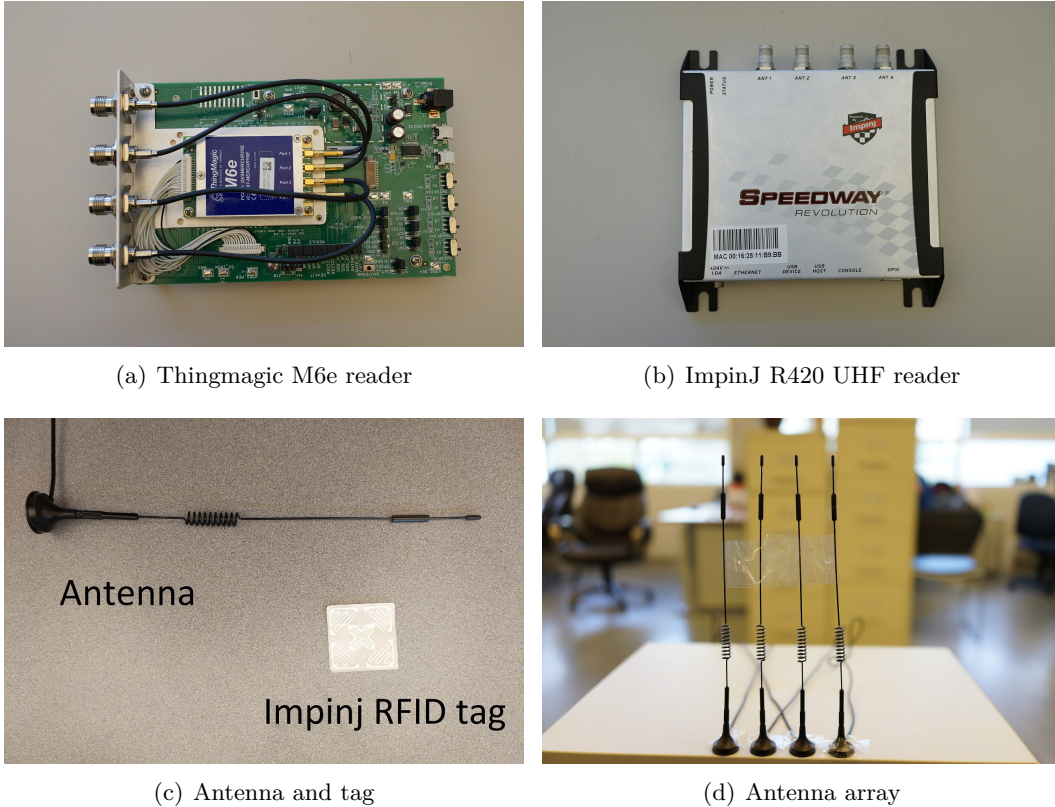


Figure 4.12: Commercial hardware used to implement TagFree System

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.

$$E = - \sum_{\gamma} z_{\gamma} \ln Pr(y|x_i) \quad (4.17)$$

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

4.6 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.

RFID Readers: Our system works with today’s commercial off-the-shelf readers, and our prototype implementation uses a Thingmagic reader³ (Figure. 4.12 (a)), and an Impinj Speedway R420 reader⁴ (Figure. 4.12 (b)) without any hardware or firmware modification. The Thingmagic M6e reader provides the APIs to customize the frequency hop table, thus

³<http://www.thingmagic.com/>

⁴<https://support.impinj.com/>

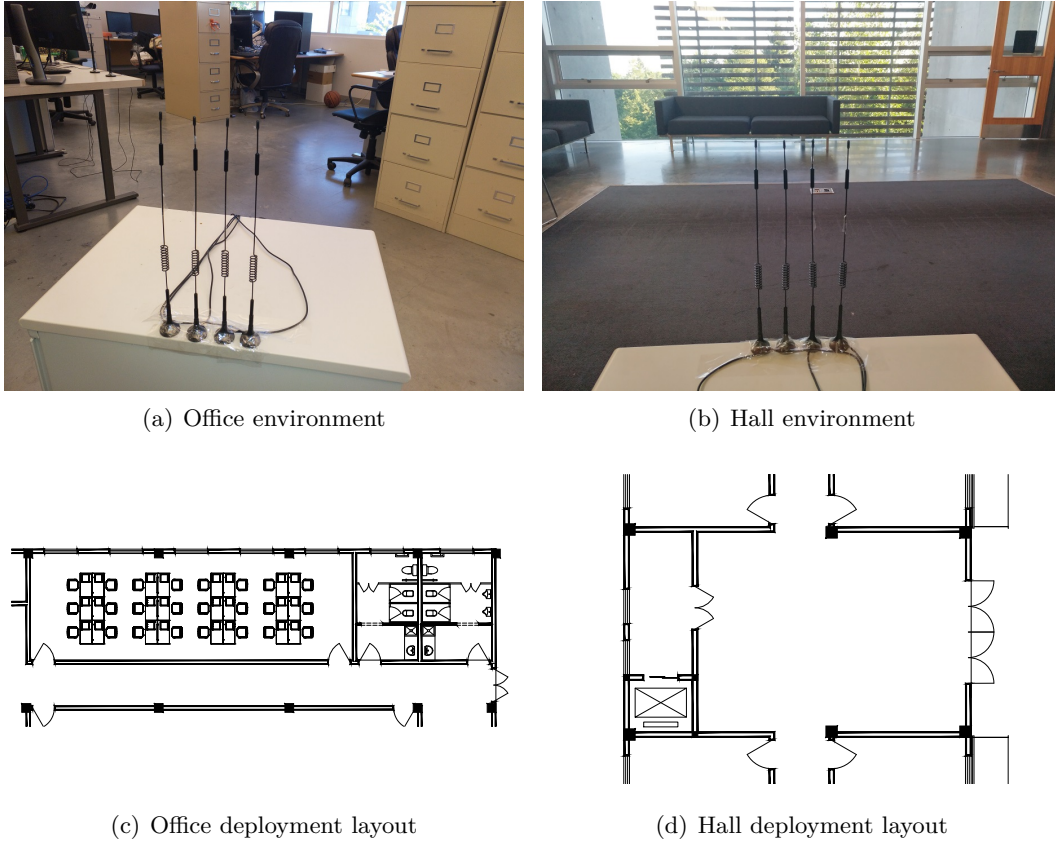


Figure 4.13: Two typical indoor environments and deployment layouts with the positions of arrays and tags

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 Figure. 4.12. 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 [4];

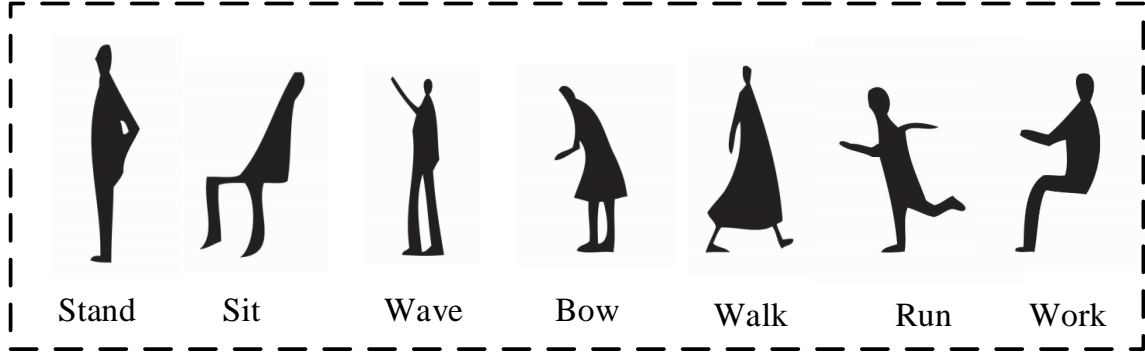


Figure 4.14: Activities in experiments

- 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 $d = \lambda/4$.
- 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 $13.75 \text{ m} \times 10.50 \text{ m}$ has many file cabinets and writing desks, as shown in Figure. 4.13 (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 $8.75 \text{ m} \times 7.50 \text{ m}$ is shown in Fig 4.13 (b). The deployment layouts of the environments are shown in Figure. ?? (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 Figure. 4.12 (c) and place them facing the antenna array with angles $45^\circ, 60^\circ, 75^\circ, 90^\circ, 105^\circ, 120^\circ$, 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 knowing their exact locations. Unless specifically mentioned, we use the default setup for performance evaluation.

Evaluation Metrics: We invite ten volunteers and each volunteer⁵ stands between the tags and the antenna array in our experiments. To conduct a comprehensive evaluation, we test seven scenarios as shown in Figure. 4.14, including standing, sitting, waving, bowing, walking, running, and working. In each activity case, each volunteer is required to repeat-

⁵Note that those volunteers cover a broad range in age, gender, height, and weight.

edly 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 [42] to communicate with the reader, with Mercury API⁶ and Octane SDK Java⁷ 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 Keras⁸ with cuDNN on Dual Nvidia GTX 1080 Ti GPUs, and the multiclass classifiers are implemented based on the Scikit-learn library [43]. 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.

4.7 Evaluation

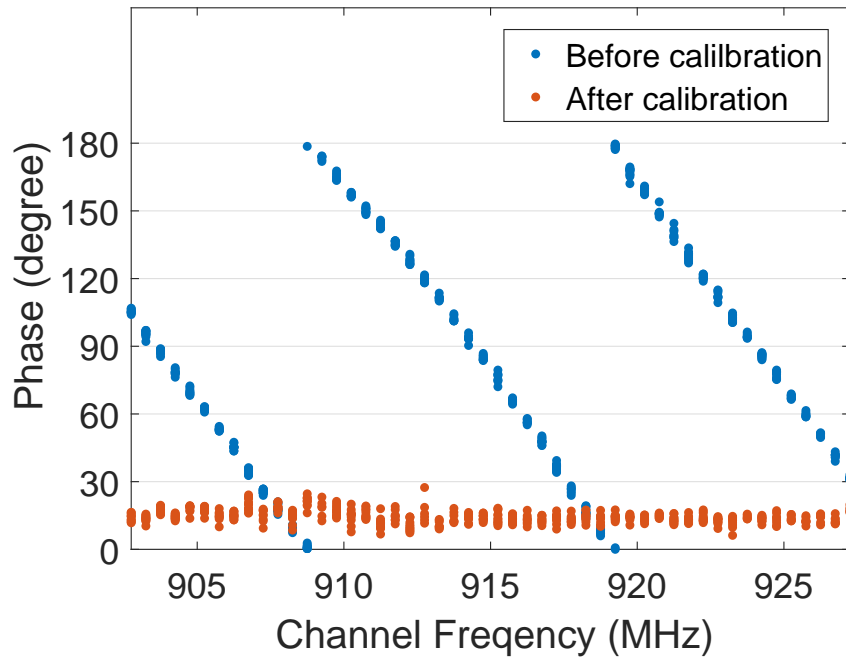
4.7.1 Model Training

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.

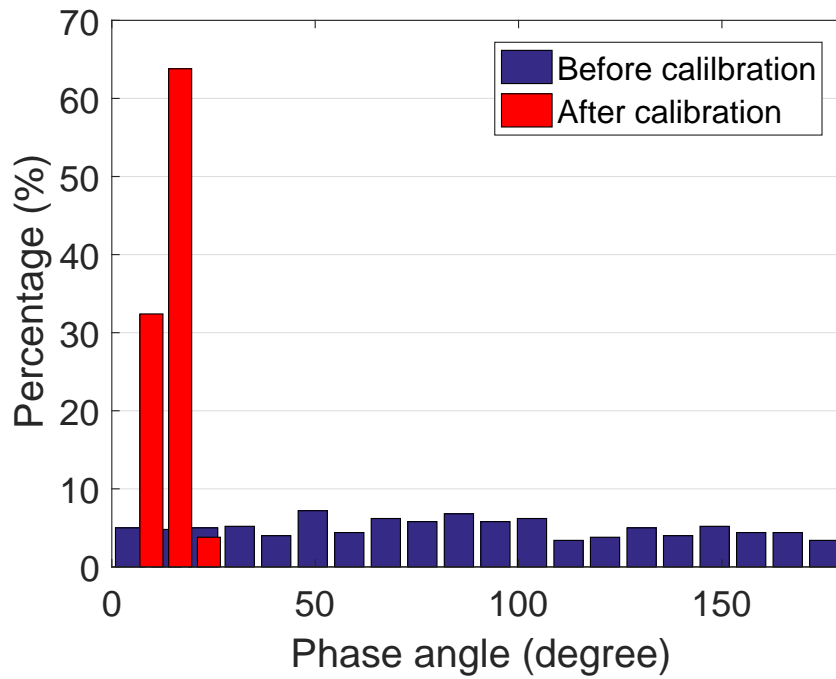
⁶Mercury API Programmer's Guide. <https://www.thingmagic.com/>

⁷Octane SDK Programmer's Guide. <https://support.impinj.com/>

⁸Keras: Deep Learning library for Theano and TensorFlow. <https://keras.io/>

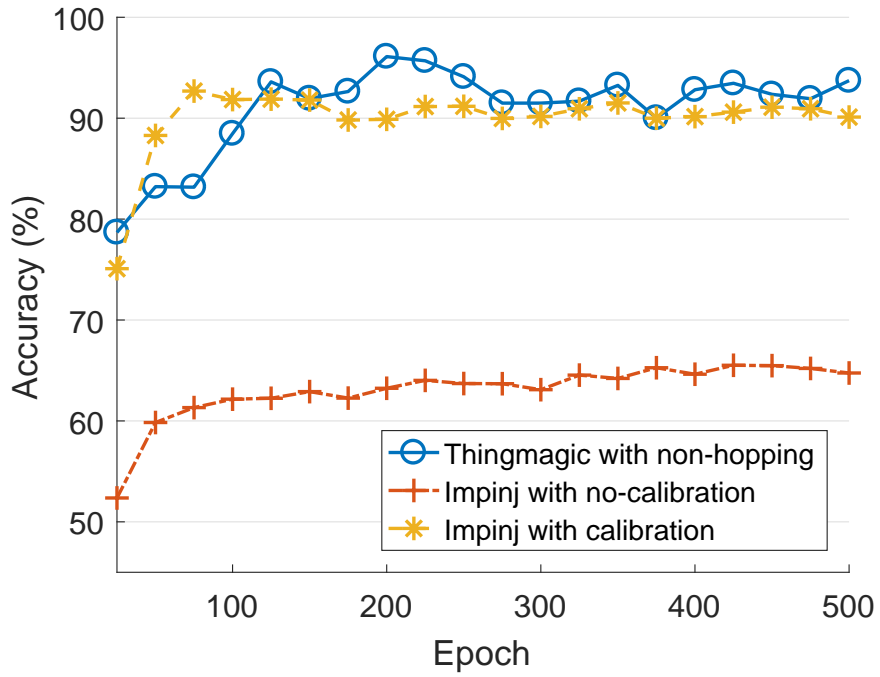


(a) The comparison between the raw phase and the calibrated phase

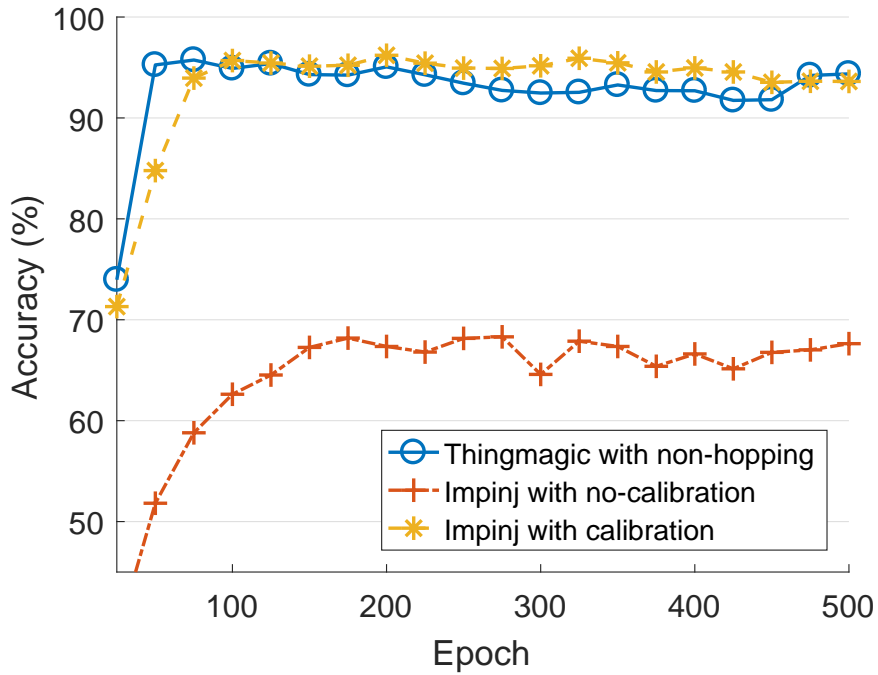


(b) Histogram of phase distribution for frequency hopping calibration

Figure 4.15: Results of phase calibration



(a) In the office with rich-multipath



(b) In the hall with low-multipath

Figure 4.16: Impact of phase calibration

4.7.2 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. Figure. 4.15 (a) shows the result of the calibrated phase data with respect to Figure. 4.9 (a). It illustrates the phase offset causing by frequency jumping can be effectively eliminated by our method. Figure. 4.15 (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° .

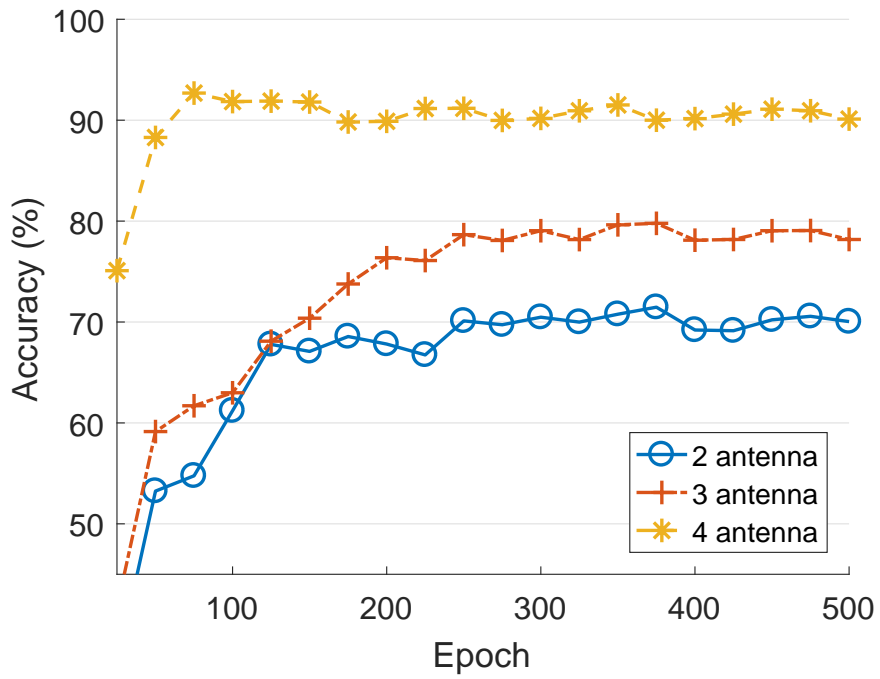
It is worth noting that our phase calibration mechanism contributes toward improving the precision of activity identification, as shown in Figure. 4.16. We evaluate how phase calibration benefits TagFree in Figure. 4.16 (a) and (b), where the x-axis of the figures represents the model training epochs 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.

4.7.3 Impact of Different System Settings

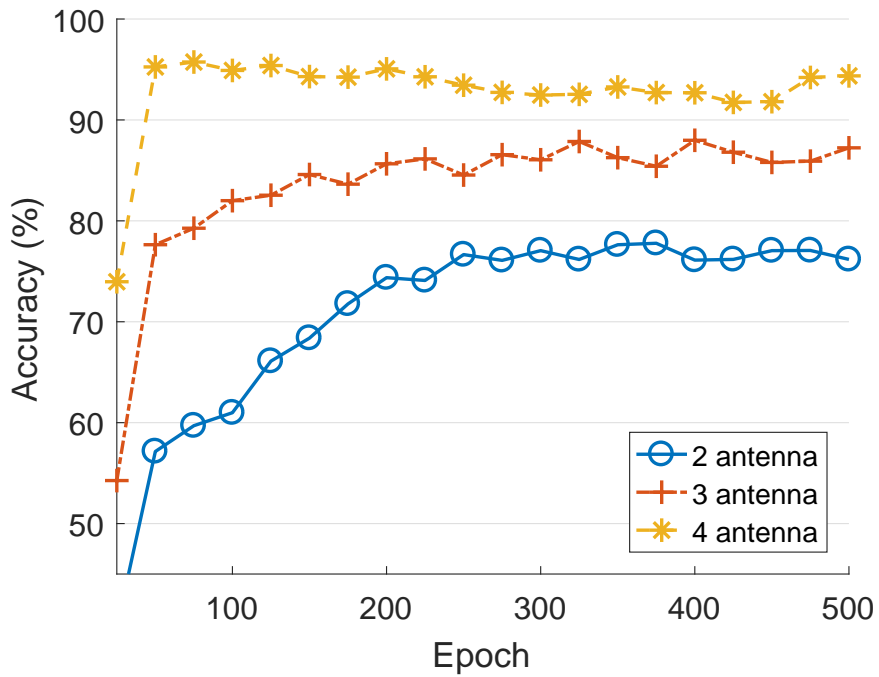
TagFree de-couples the multipaths using an array of antennas. Therefore the number of antennas limits the number of multipaths that can be detected by our pre-processing scheme. With the information of more signal paths, TagFree achieves a higher multipath density in the area and improves the activity identification accuracy. We thus investigate the impact of the number of antennas in two typical environments, as shown in Figure. 4.17 (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 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 Figure. 4.18 (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, more tags only bring marginal increase to the system cost, which is considerably lower than

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

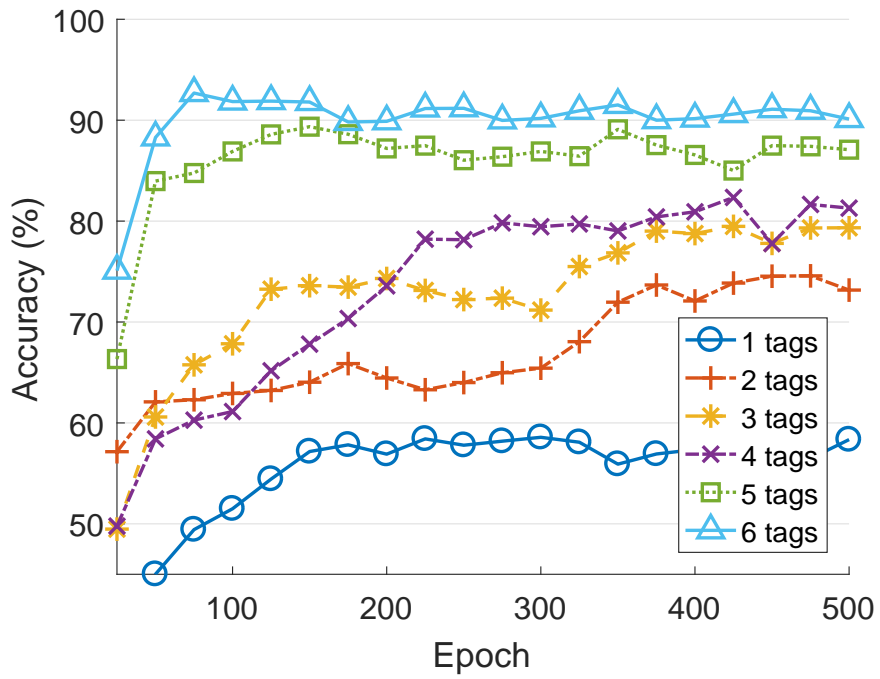


(a) In the office with rich-multipath

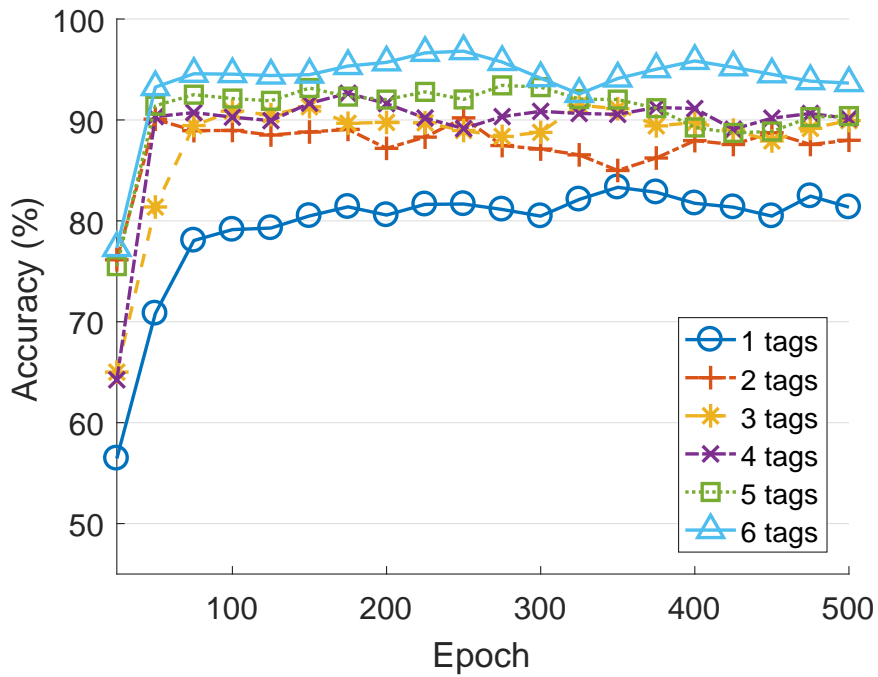


(b) In the hall with low-multipath

Figure 4.17: Impact of # of antennas

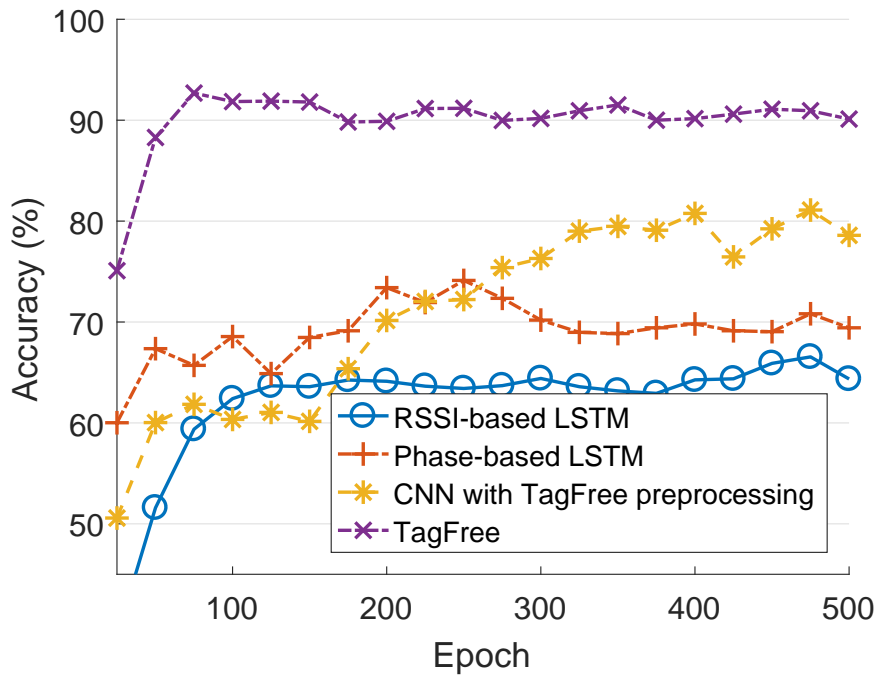


(a) In the office with rich-multipath

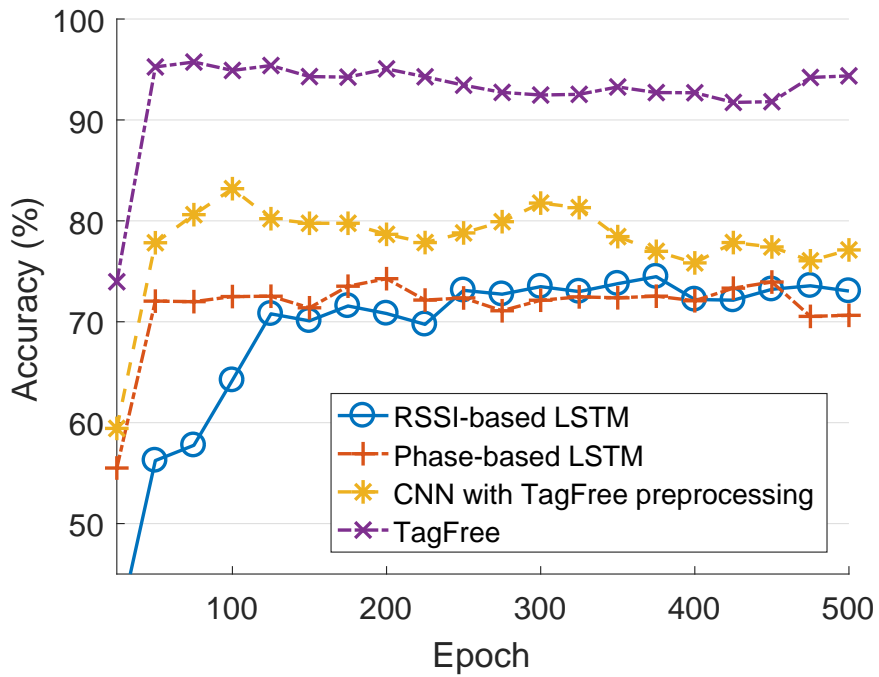


(b) In the hall with low-multipath

Figure 4.18: Impact of # of tags



(a) In the office with rich-multipath



(b) In the hall with low-multipath

Figure 4.19: Impact of deep learning approaches

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 4.2: 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	0.97	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	

Table 4.3: Comparison of performance with different physical configuration in the hall with LOW-multipath

increasing the number of antennas on readers [22]. To this end, we use a tag number of 6 as the default setting of our evaluation.

Tab. 4.2 and Tab. 4.3 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. 4.2 and Tab. 4.3 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 Figure. 4.19 (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

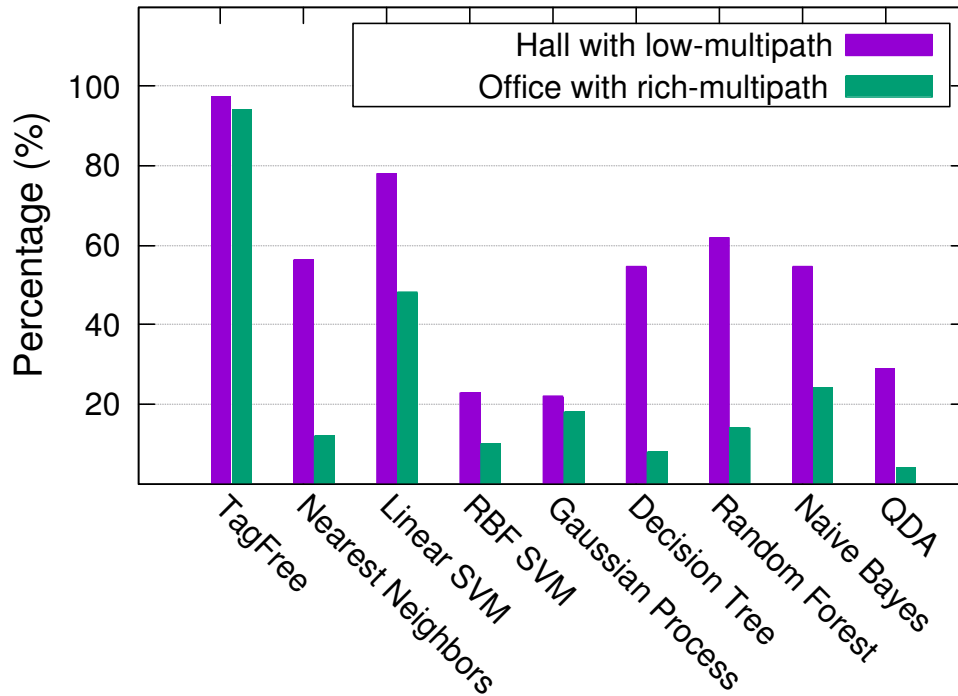


Figure 4.20: Overall performance of TagFree framework

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. 4.4, 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. 4.4 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. 4.5 show that the size of 2×16 has the best performance. In Table. 4.6, 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. 4.7 further suggests that increasing the number of LSTM cell per layer beyond 32 cells achieves only marginal performance improvement, and thus we set the number of LSTM cell to 32.

	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 4.4: Comparison of performance with different number of convolutional layers

	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 4.5: Comparison of performance with different kernel size

	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 4.6: 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 4.7: Comparison of performance with different number of LSTM cells per layer

4.7.4 Overall Activity Identification Performance

Figure. 4.20 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 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 Figure. 4.20 into Tables 4.8 and 4.9. The tables illustrate the details of the experiments in the hall and the office, respectively. Each row

	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 4.8: Low multipath environments in hall

	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

Table 4.9: Rich multipath environments in office

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. 4.8, 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.

Chapter 5

Multiple Object Activity Identification using RFIDs

5.1 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 [29], radars [28], and/or various inertial sensors [30]. 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 batteryless nature [56] [57] [33] [58] [59]. We have seen pioneer studies on RFID-based activity identification as well [4] [6], 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 [31].

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 individual activities. This is further complicated with multi-path 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 [6]. 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.

We however argue that, though often considered harmful, 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 first time accommodates both multi-path and multi-object.¹

As shown in Figure. 5.1, M²AI 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 [3] and pseduospectrum [2] 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) [46] and a Long Short Term Memory (LSTM) network [47]. We show that this integration is computationally effective with high accuracy, beating such conventional classification tools as SVM and Nearest Neighbors.

Our M²AI 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 M²AI 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 chapter is organized as follows. Section 5.2 examines the key challenges for identifying the activities of multiple objects in the indoor environment, and presents our M²AI framework to explore the hidden information in multipath. Section 5.3 introduces our data pre-processing scheme to deal with frequency hopping and de-couple multipath signals. Section 5.4 then presents a deep learning solution for multiple object activity identification

¹An 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.

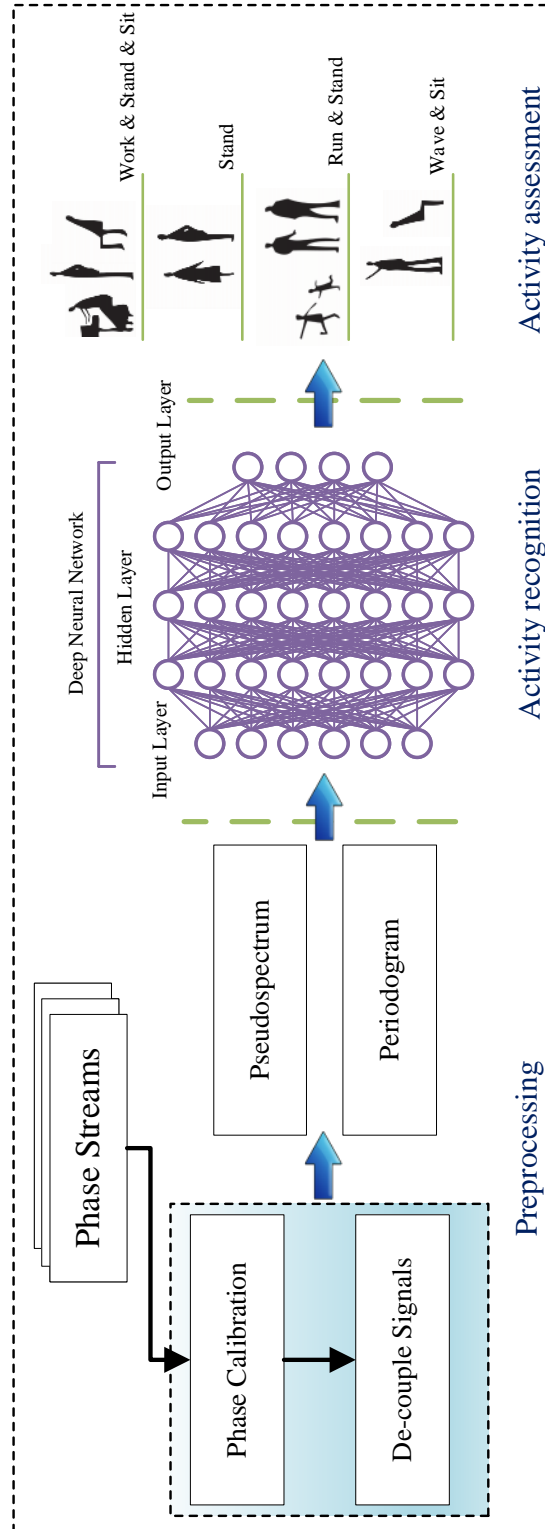


Figure 5.1: M^2 AI framework

from the processed data. Section 5.5 discusses the framework implementation details, with its performance being evaluated in Section 5.6. Section ?? discusses some potential further extensions.

5.2 M²AI: Motivation and Overview

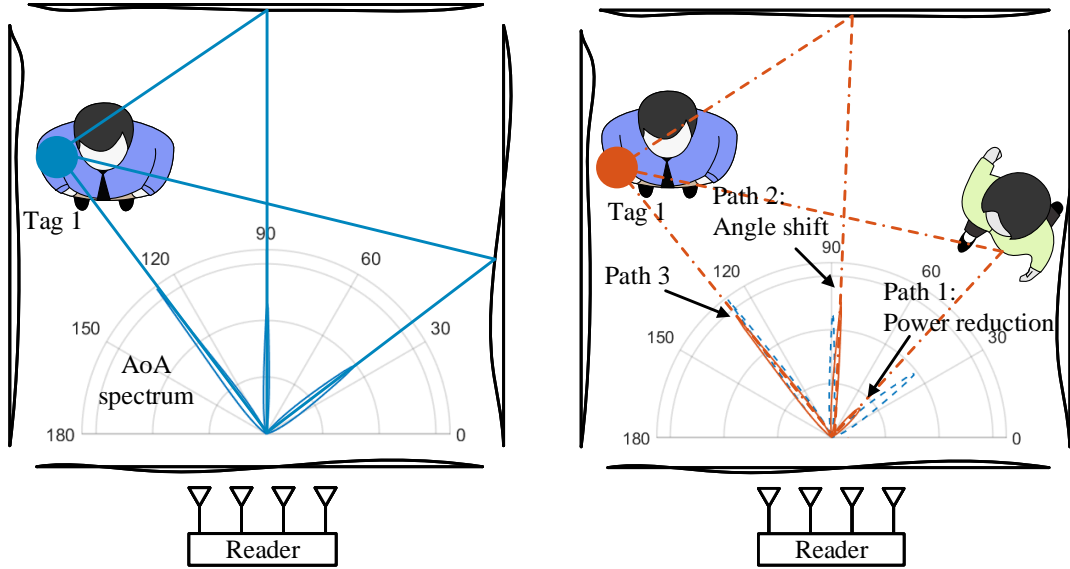
The Angle of Arrival (AoA) of an RF source is computed by comparing the phases of the received signals at antennas [60] [61]. AoA estimation is widely used in RF-based positioning given the different propagation distances to different antennas, and serves as a foundation for activity identification as well [4]. 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 Figure. 5.2. We consider an object of activity identification as a single person equipped with multiple tags. In Figure. 5.2(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, Figure. 5.2(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., Figure. 5.2(c) with five more tags added, we can see that the number of signal paths increases rapidly in multiple objects of activity identification.

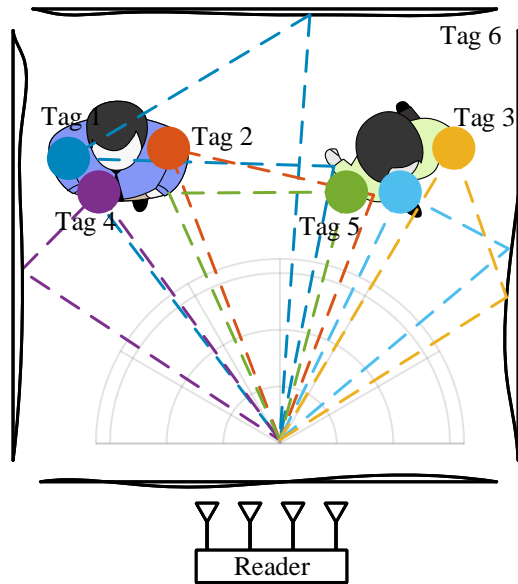
Intuitively, estimating AoA is more difficult in these indoor environments than in an open space. Analyzing the estimated data to derive the corresponding activities can be even more complicated, particularly for the latter two cases with multi-object. On the other hand, the massive multi-path signal 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;
- **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 [49]. The aggregated impact with multi-path and multi-object can be even more dramatic.



(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.

Figure 5.2: The basic idea of M^2AI from a single object to multiple objects of activity identification in the indoor environments.

To deal with the above challenges, our M²AI incorporates a phase calibration scheme to automatically calibrate the phase difference between frequencies without the requirement of human intervention. It then combines the signals' pseduospectrum 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) [46] and a Long Short Term Memory (LSTM) network [47] 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 M²AI 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, M²AI does not need to explicitly know the RFID tags' locations. Hence, the tags can be arbitrarily placed with a high degree of flexibility.

5.3 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 [48] [15], 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 M²AI design to preprocess the measured phase from the reader.

5.3.1 Phase Calibration

To limit the co-channel interference, FCC regulation requires that commercial UHF RFID readers must randomly hop 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 Figure. 5.3, where it is clear to see

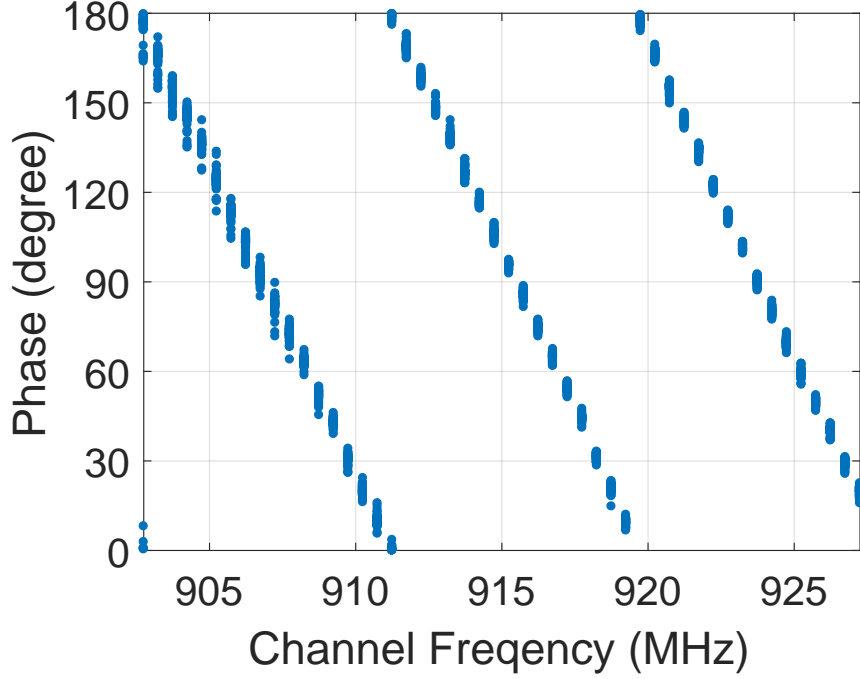


Figure 5.3: Phase jumping caused by frequency hopping

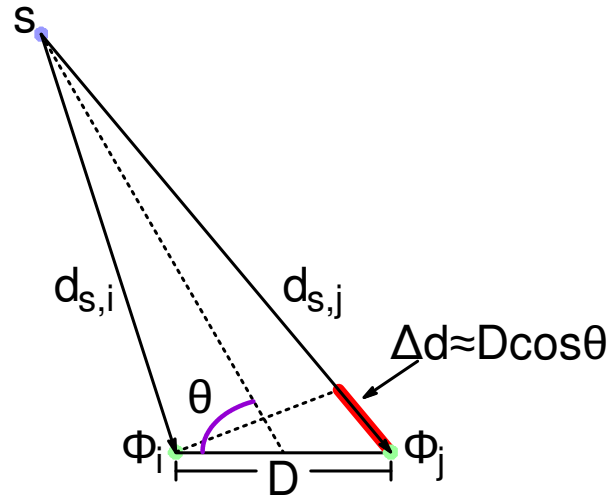
that the phase and frequency relation follows the linear model. These experiments imply that different frequencies induce different initial phase-offsets at the reader. To overcome the issue, we design a mechanism to calibrate the phase difference between frequencies, so that the phase output looks like 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:

$$\phi(t) = \phi^j(t) - \tilde{\phi}^j + \tilde{\phi}^r \quad (5.1)$$

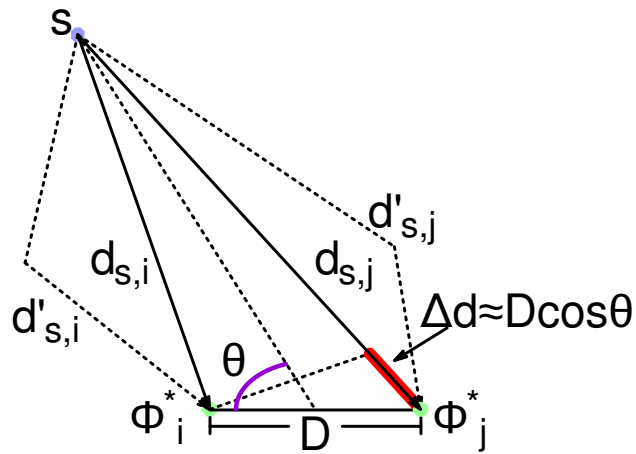
5.3.2 Angle of Arrival

Intuitively, the AoA estimation works as illustrated in Figure. 5.4: A signal source s impinges on the array of N antennas with an angle θ . Let ϕ denote the phases of the received signal that we measure at the antennas. Their relationship can thus be defined by:

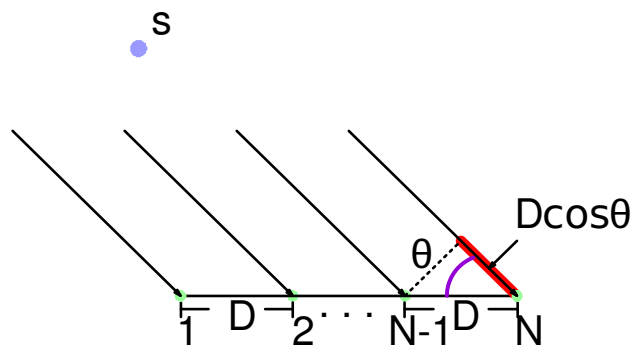
$$\phi = \frac{2\pi}{\lambda} \cdot d \cdot \cos\theta \quad (5.2)$$



(a) Single Path Signal Arrival at Antenna Pair



(b) Multiple Path Signals Arrival at Antenna Pair



(c) Antenna Array and Angle of Arrival

Figure 5.4: Illustration of angle of arrival

where d is the distance between two antennas and λ is the wavelength of the received signal.

However, in practice AoA estimation may not work well because of the multipath effect, which we will address in this subsection. Consider a number of plane waves from M narrow-band sources $s_1(t) = \rho_1 e^{j\phi_1}, \dots, s_M(t) = \rho_M e^{j\phi_M}$ (which may also be multipath signals from the same source) impinging from different angles $\theta_1, \dots, \theta_M$, and into a uniform linear array (ULA) of N antennas. Then we have

$$\mathbf{s}(t) = [s_1(t), \dots, s_M(t)]^\top \quad (5.3)$$

Let $\mathbf{a}(\theta)$ 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:

$$\mathbf{a}(\theta) = [1 \quad e^{-j\phi} \quad \dots \quad e^{-j(N-1)\phi}]^\top \quad (5.4)$$

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

$$\mathbf{r}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{w}(t) \quad (5.5)$$

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

$$\mathbf{R}_r = E\{\mathbf{r}(t)\mathbf{r}^H(t)\} = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \mu^2\mathbf{I} \quad (5.6)$$

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

5.3.3 De-couple Multipath Signals

As shown in Figure. 5.4(c), consider a number of plane waves from M narrow-band sources $s_1(t) = \rho_1 e^{j\phi_1}, \dots, s_M(t) = \rho_M e^{j\phi_M}$ (which may also be multipath signals from the same source) impinging from different angles $\theta_1, \dots, \theta_M$, and into a uniform linear array (ULA) of N antennas. Then we have

$$\mathbf{s}(t) = [s_1(t), \dots, s_M(t)]^\top \quad (5.7)$$

Let $\mathbf{a}(\theta)$ 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:

$$\mathbf{a}(\theta) = [1 \quad e^{-j\phi} \quad \dots \quad e^{-j(N-1)\phi}]^\top \quad (5.8)$$

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

$$\mathbf{r}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{w}(t) \quad (5.9)$$

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

$$\mathbf{R}_r = E\{\mathbf{r}(t)\mathbf{r}^H(t)\} = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \mu^2\mathbf{I} \quad (5.10)$$

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

Pseudospectrum Estimation

Our pseudospectrum estimation design is mainly based on the MUSIC (Multiple Signal Classification) algorithm [2], 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 \mathbf{R}_r has N eigenvalues associated with N eigenvectors $\mathbf{U} = [\mathbf{U}_1, \dots, \mathbf{U}_N]$. The largest M eigenvalues correspond to the M incoming signals while the rest $N - M$ correspond to the noise. The corresponding eigenvectors in \mathbf{U} can be classified into the signal subspace \mathbf{U}_s and noise subspace \mathbf{U}_n :

$$[\mathbf{U}_s \mathbf{U}_n] = [\underbrace{\mathbf{U}_1, \dots, \mathbf{U}_M}_{\mathbf{U}_s}, \underbrace{\mathbf{U}_{M+1}, \dots, \mathbf{U}_N}_{\mathbf{U}_n}] \quad (5.11)$$

Similar to the MUSIC algorithm utilizing the orthogonality relationship between the signal and noise subspaces [2], which implies $\mathbf{a}^H(\theta)\mathbf{U}_n = \mathbf{0}$, the direction of arrival angle can be represented in terms of a spectral estimation plots:

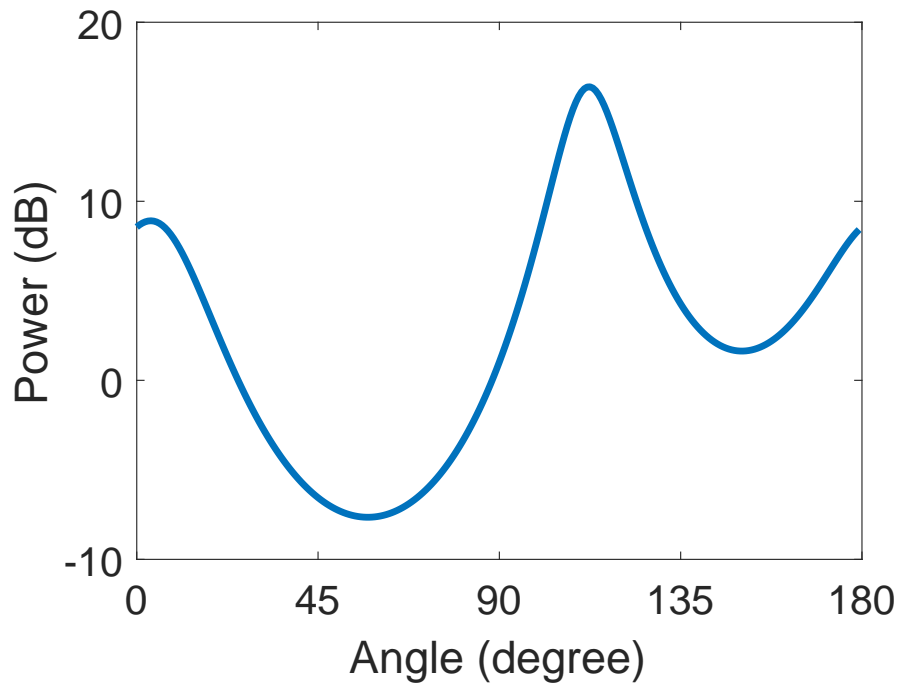
$$P_{MUSIC} = \frac{1}{\mathbf{a}^H(\theta)\mathbf{U}_n\mathbf{U}_n^H\mathbf{a}(\theta)} \quad (5.12)$$

The above equation results in high peaks, when the direction of arrival of the signal source is exactly equal to that of θ . In Figure. 5.5(a), the M higher peaks are of great power, where each corresponds to an estimated arrival angle.

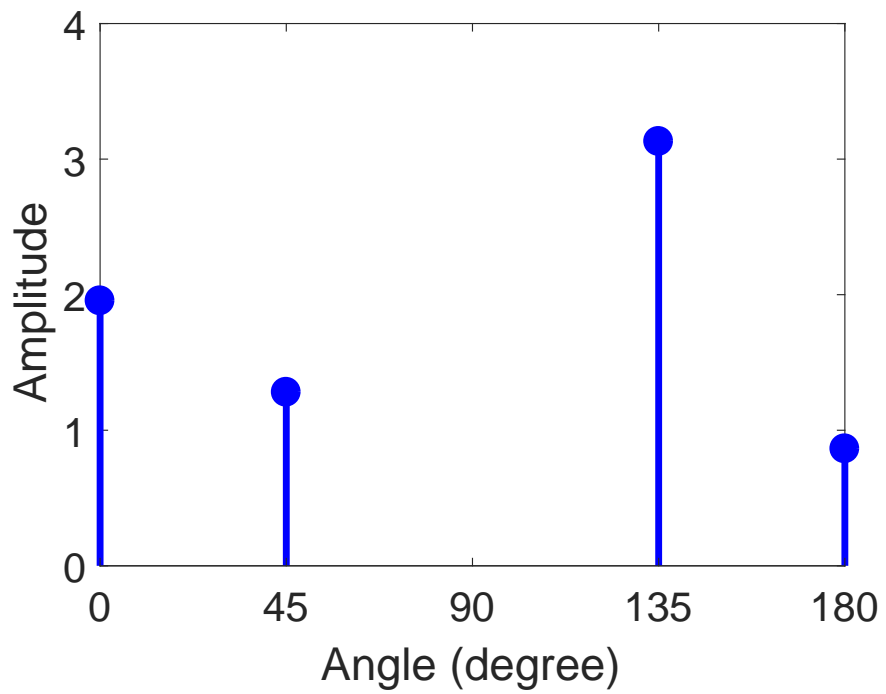
Periodogram Estimation

We introduce the periodogram [3] to strengthen pseudospectrum 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:

$$\phi(\omega) = \sum_{k=-\infty}^{\infty} r(k)e^{-j\omega k} \quad (5.13)$$



(a) Pseudospectrum Estimation



(b) Periodogram Estimation

Figure 5.5: Illustration of data preprocessing in our M^2AI framework

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:

$$\hat{\phi}_p(\omega) = \frac{1}{N} \left| \sum_{t=1}^N y(t) e^{-j\omega t} \right|^2 \quad (5.14)$$

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:

$$\omega = \frac{2\pi}{N} k, k = 0, \dots, N - 1 \quad (5.15)$$

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:

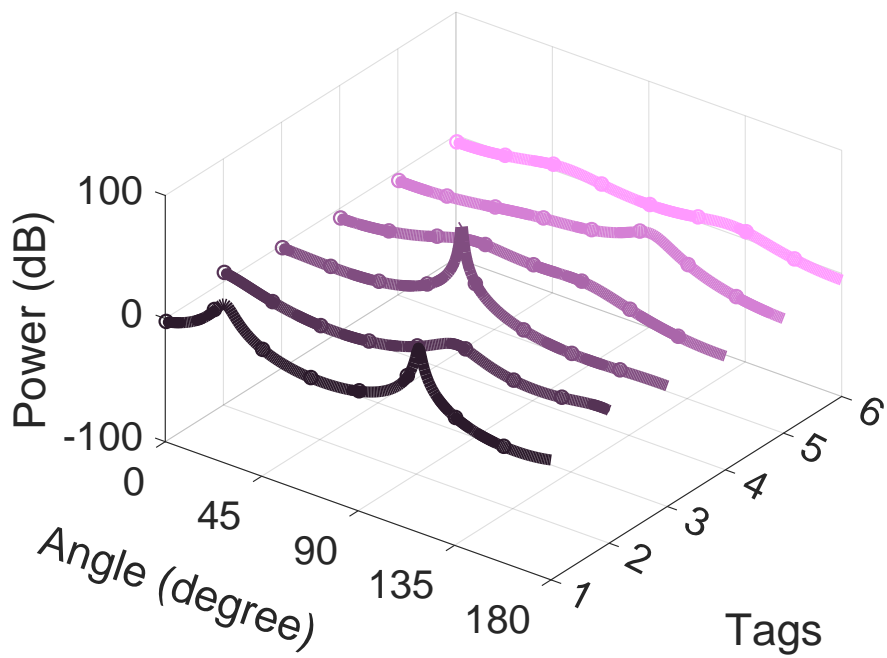
$$Y(k) = \sum_{t=1}^N y(t) W^{tk}, k = 0, \dots, N - 1 \quad (5.16)$$

In our system, we use Fast Fourier Transform [62] to estimate the power distribution. According to the Parseval's theorem [63], 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.

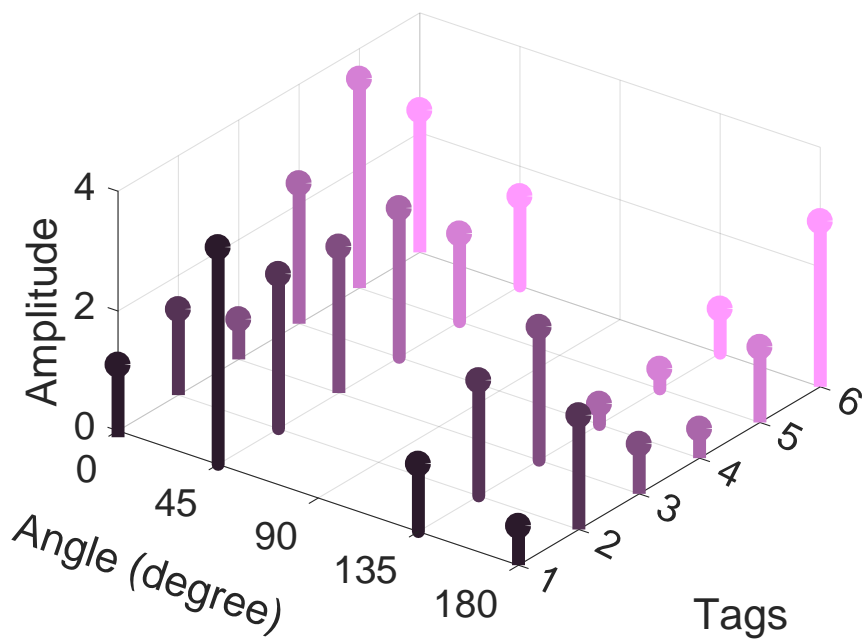
As illustrates in Figure. 5.5(b), we can get four values in the periodogram for the power density distribution. In the implementation, we have four antennas to connect to the Impinj Speedway R420 reader, where the number of RF ports in the reader limits the scale of our antenna array. We can increase the antenna number by Impinj antenna hubs. The effectiveness of the periodogram estimation has also been demonstrated by our experiments in Section. 5.6.

5.4 Deep Learning Design for Activity Identification

This section describes the main components of our M²AI design. Our deep learning design takes the results from data preprocessing into our M²AI deep learning architecture, i.e., periodogram and psuedospectrum frames, as shown in Figure. 5.6(a) and Figure. 5.6(b), respectively. As illustrated in Figure. 5.7, we develop an integrated design of a Convolutional Neural Network (CNN) [46] and a Long Short Term Memory (LSTM) network [47]. 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) Pseudospectrum Frame



(b) Periodogram Frame

Figure 5.6: Illustration of spectrum frames in our M^2AI framework

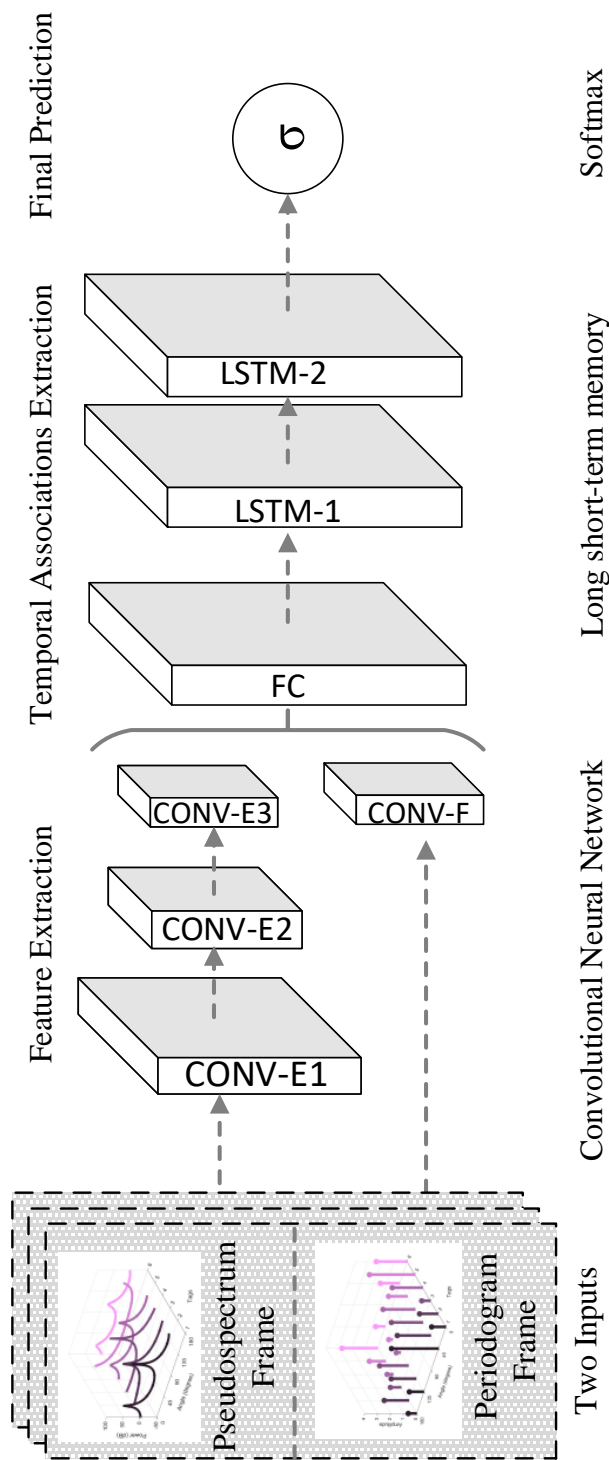


Figure 5.7: M²AI deep learning network architecture

5.4.1 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. Specifically, we provide the following as input to the model: (1) pseudospectrum frames for angle of arrival as illustrated in Figure. 5.6(a), and (2) periodogram frames for power spectral density as Figure. 5.6(b) 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.

5.4.2 Hidden Layer

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

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 $\mathbf{x} = (x_1, \dots, x_T)$ 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 Figure. 5.7, 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.

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.

5.4.3 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

$$E = - \sum_{\gamma} z_y \ln Pr(y|x_i) \quad (5.17)$$

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

5.5 System Implementation

Our M²AI 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.

Readers and Tags: Our prototype implementation uses an Impinj Speedway R420 reader² 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 psuedospectrum and periodogram estimation can be well implemented with Impinj R420 reader, which has been extensively used in the research community [5] [22]. 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 Figure. 5.8(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 Figure. 5.8. One important setting is the distance d between antenna pairs, where we set d as $\lambda/8$ with the following rationale:

²<https://support.impinj.com/>

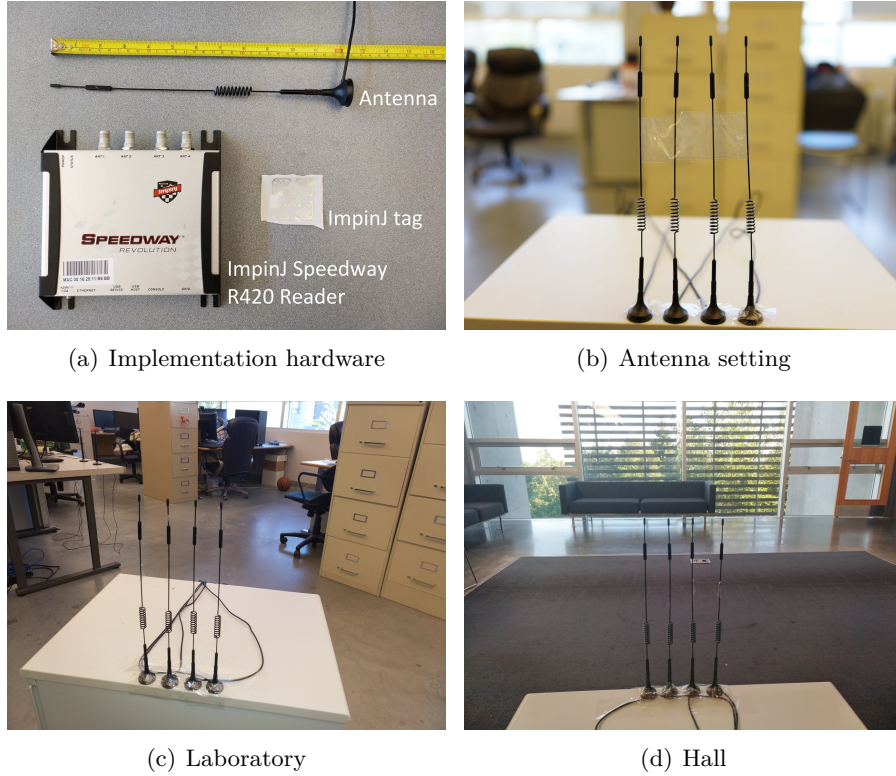


Figure 5.8: Commercial hardware used to implement M^2AI and two typical indoor environments, i.e., laboratory and hall, corresponding to high and low multipath environments, respectively.

- Theoretically, the antenna separation d should be spaced by $\lambda/2$, which effectively reduces the ambiguity caused by the high-resolution grating lobes [4].
- 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 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 Guide³ with LLRP protocol to communicate with the

³<https://support.impinj.com/>

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 5.1: Confusion matrix of activity identification

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.

5.6 Performance Evaluation

5.6.1 Experiment Setup

We conduct extensive experiments in two typical indoor environments, as Figure. 5.8(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 \times 10.50 m has many file cabinets and writing desks, as shown in Figure. 5.8(c). The empty hall with a size of 8.75 m \times 7.50 m is shown in Fig 5.8(d). In each environment, we deploy one 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 volunteers⁴ 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

⁴Note that those volunteers vary in age, gender, height, and weight.













					
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

Figure 5.9: Sketches of the tested activities

12 activity scenarios with two people as shown in Figure. 5.9. 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 on the test set for each epoch.

5.6.2 Activity Identification Performance

Figure. 5.10 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, Neural Net, 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). Tab. 5.1 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-

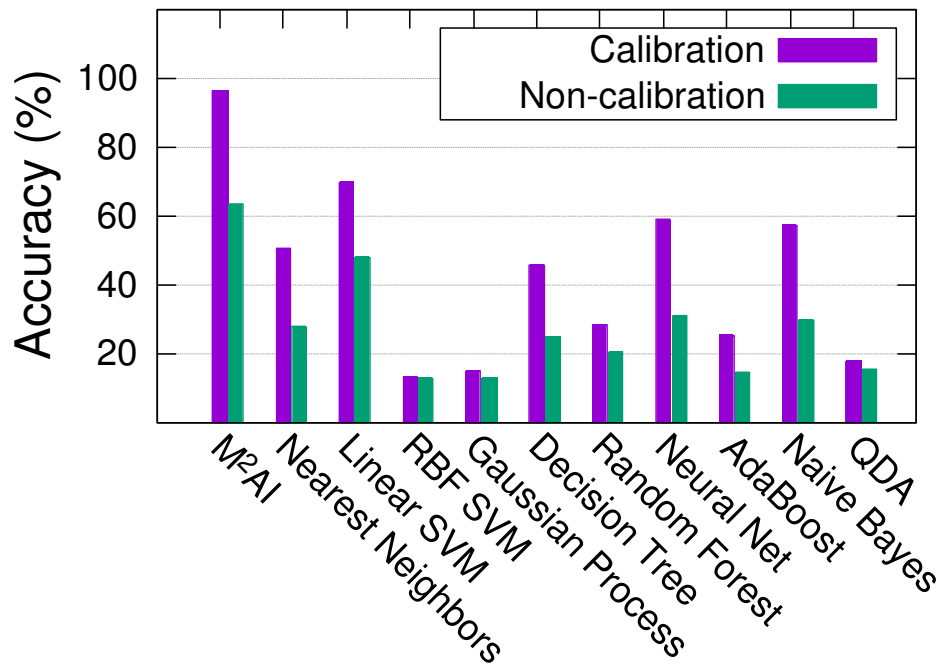


Figure 5.10: Overall activity identification performance

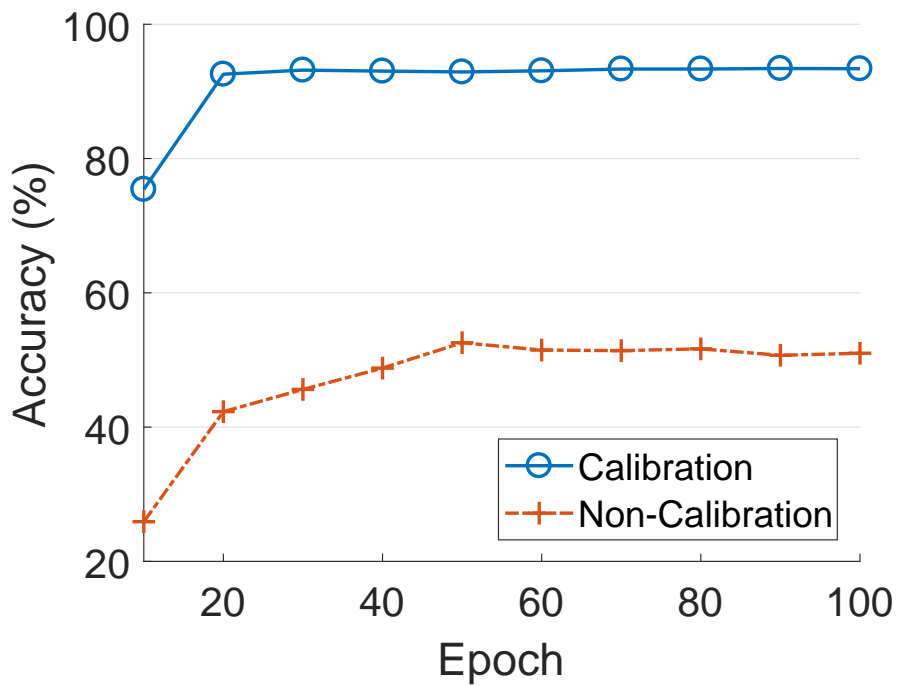


Figure 5.11: Impact of phase calibration

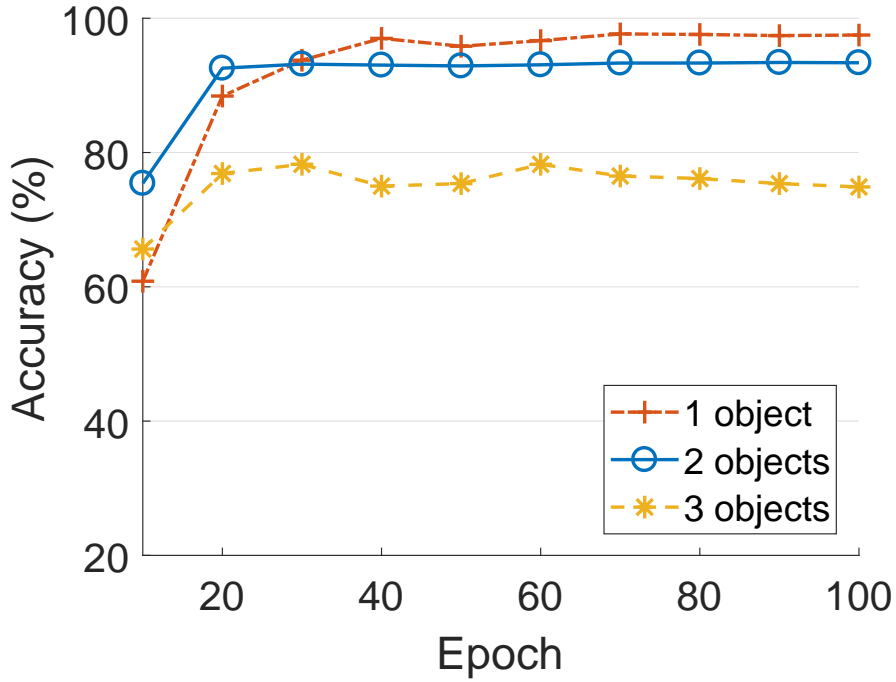


Figure 5.12: Impact of # of objects

calibration. The results are shown in Figure. 5.11, 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 high AoA estimation accuracy. The result also explains the statement in [15] 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 Figure. 5.12. 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. Figure. 5.13 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 Figure. 5.14, 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 multipath diversity in the area and improves the activity identification accuracy. Since the number of antennas may limit the multipath

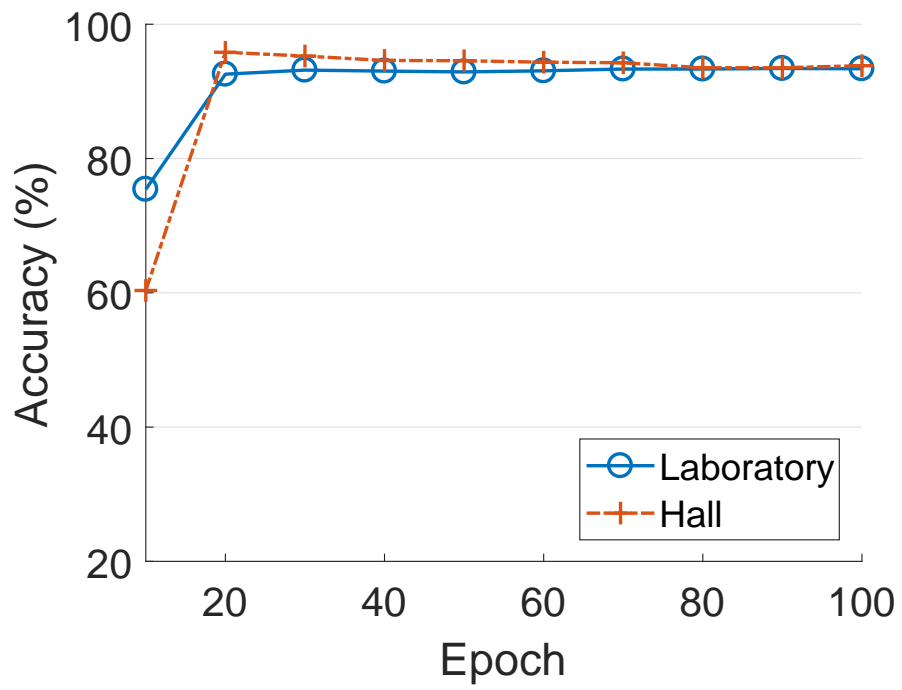


Figure 5.13: Impact of different places

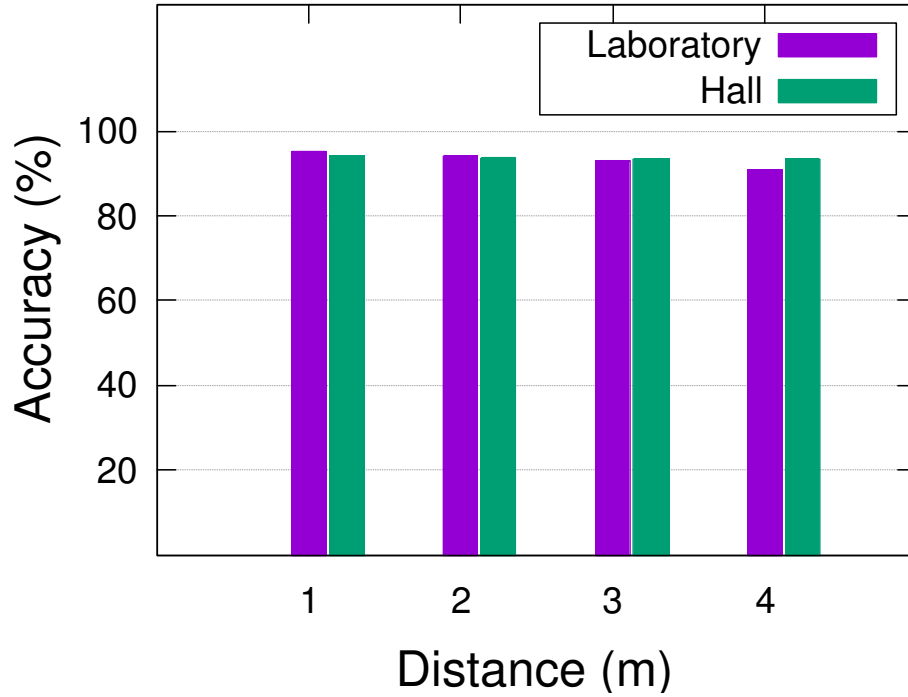


Figure 5.14: Impact of different distance

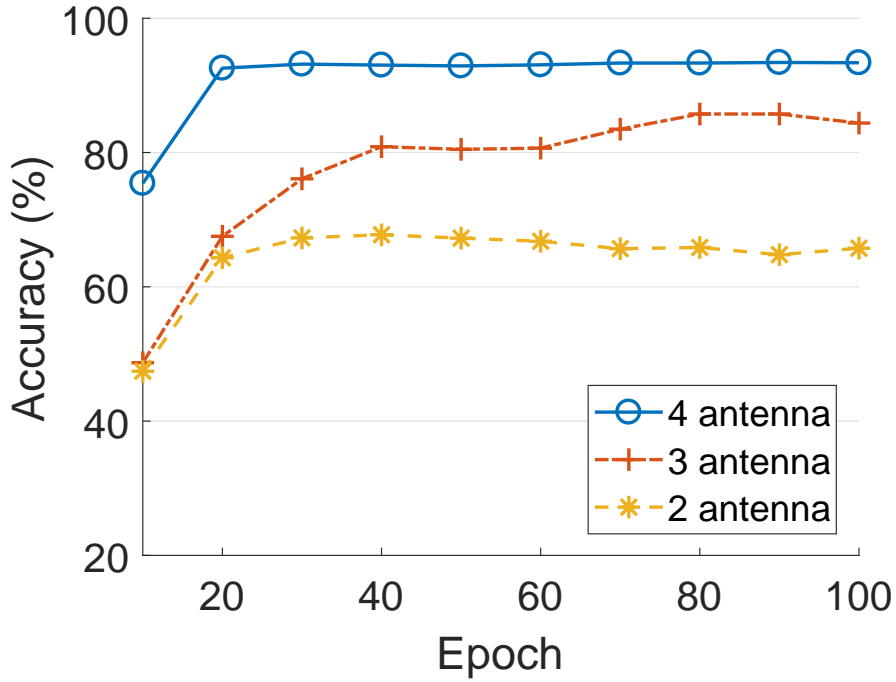


Figure 5.15: Impact of # of antennas

decoupling ability by our pre-processing scheme, we thus investigate the impact of the number of antennas as shown in Figure. 5.15. We can see that when the number of antennas increases from 2 to 4⁵ more angle information of multipath can be detected, and thus M²AI 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 Figure. 5.16. 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 Figure. 5.17, 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 M²AI shows the effectiveness of our preprocessing scheme. In Figure. 5.18, we further compare the results of our M²AI with different deep learning network architectures. It clearly shows that M²AI can achieve 30% higher accuracy than the ar-

⁵Note that Impinj Speedway R420 reader has maximally four ports to connect with antennas.

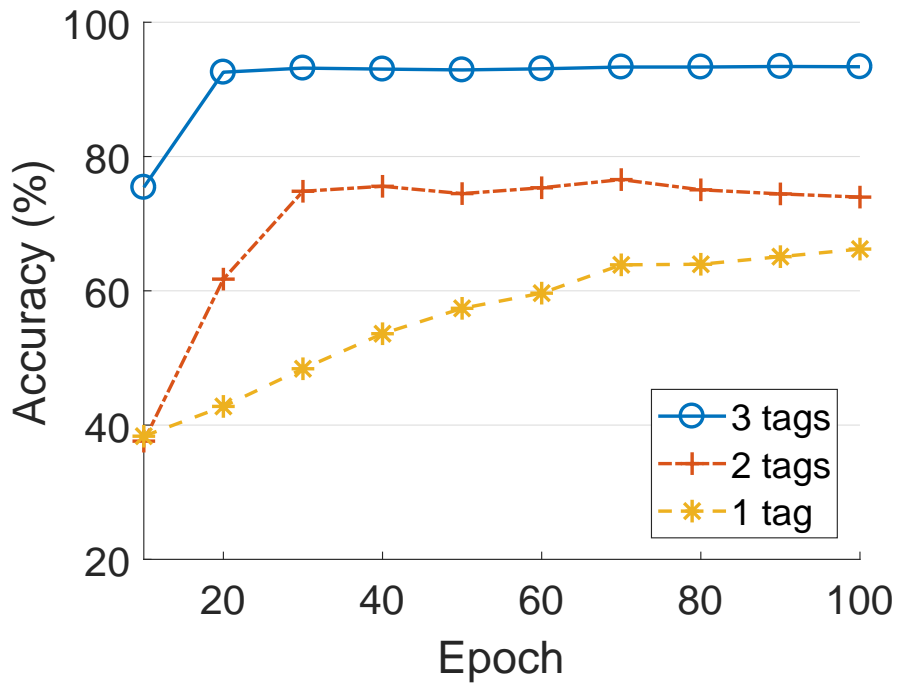


Figure 5.16: Impact of # of tags per person

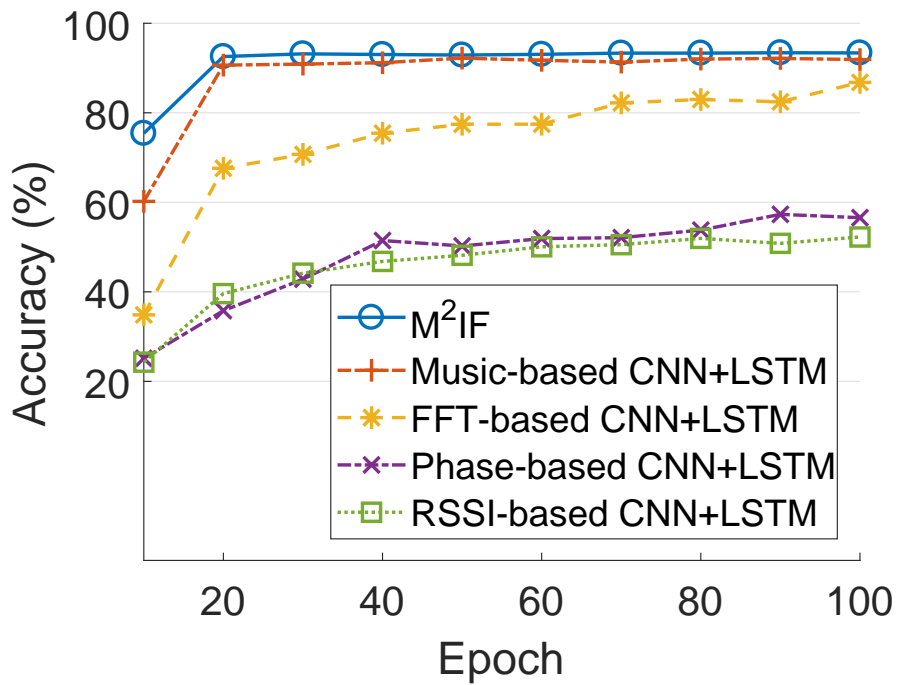


Figure 5.17: Impact of different inputs preprocessing

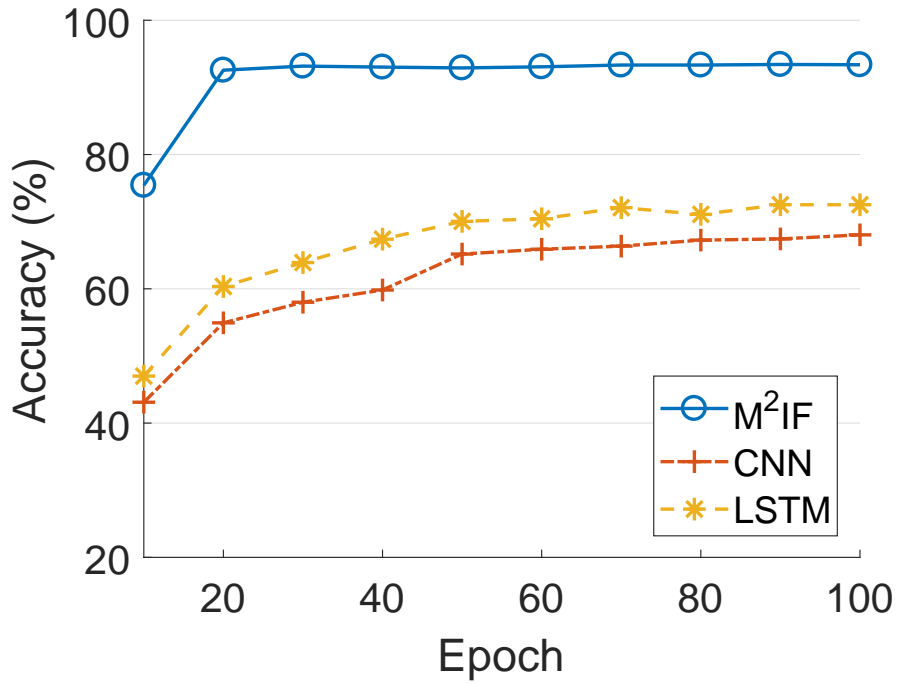


Figure 5.18: Impact of different learning networks

chitecture only using CNN and our preprocessing scheme, which demonstrates that LSTM architectures are necessary for activity identification. On the other hand, M²AI obtains 25% higher accuracy than the architecture only using LSTM, which illustrates that CNN can efficiently extract the features for activity identification. In summary, the benefits of M²AI 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.

Chapter 6

Conclusion and Discussion

6.1 Summary of the Thesis

In this thesis, we covered a broad spectrum of activity identification with RFIDs, from the perspectives of profiling RFID tag mobility, RFID-based device-free activity identification, as well as multi-path and multi-object for activity identification.

- First, we presented the architectural design of $i^2\text{tag}$, that can detect tag mobility and identify activities in typical indoor environments. $i^2\text{tag}$ employs a novel *mobility profile* to quantify different levels of mobility, which seamlessly integrates RSSI variance and packet loss rate, as well as a relative-phase-based fingerprint. We have offered a *multiple dimensional dynamic time warping* algorithms to detect the tag mobility and utilize the multiclass SVMs algorithm to recognize human activities. A prototype has been implemented using a **Thingmagic** reader and **Impinj** tags and has been examined under various indoor environments. Experimental results have demonstrated its superiority in mobility detection and activity identification in various indoor environments.
- Second, we presented **TagFree**, the first RFID-based device-free activity identification system by analyzing the multipath signals, which 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.

- Third, We presented M^2AI , the Multipath-aware Multi-object Activity Identification framework, which can identify multiple object activities in typical indoor environments. M^2AI 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^2AI achieves the activity identification accuracy of 97%, which is 27% better than the state-of-art machine learning approaches.

6.2 Discussion

Our frameworks mark an important step toward enabling accurate indoor activity identification. We envision the basic design 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 works well on accurately identifying predefined scenarios with multiple objects and multiple activities, it is not clear how our deep learning based framework 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 our system 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.

In addition, our current design focuses on identifying multiple activities of multiple objects. Yet, it may also be useful to further identify and differentiate users. Future research may achieve this by extracting robust and representative user features instead of activity features to train a model for user identification, where a solution similar to the one based on WiFi signals [64] may be adopted for this future extension.

Bibliography

- [1] Anshul Rai, Krishna Kant Chintalapudi, Venkata N Padmanabhan, and Rijurekha Sen. Zee: Zero-effort crowdsourcing for indoor localization. In *Proceedings of ACM MobiCom 2012*.
- [2] Ralph Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, 1986.
- [3] Petre Stoica and Randolph L Moses. *Introduction to spectral analysis*. Prentice hall Upper Saddle River, 1997.
- [4] Jue Wang, Deepak Vasisht, and Dina Katabi. Rf-idraw: Virtual touch screen in the air using rf signals. In *Proceedings of ACM SIGCOMM 2015*.
- [5] 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*.
- [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] Daqiang Zhang, Jingyu Zhou, Minyi Guo, Jiannong Cao, and Tianbao Li. Tasa: Tag-free activity sensing using rfid tag arrays. *IEEE Transactions on Parallel and Distributed Systems*, 22(4):558–570, 2011.
- [8] Paramvir Bahl and Venkata N Padmanabhan. Radar: An in-building rf-based user location and tracking system. In *Proceedings of IEEE INFOCOM 2000*.
- [9] Lionel M Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P Patil. Landmarc: indoor location sensing using active rfid. *Wireless networks*, 10(6):701–710, 2004.
- [10] Yiyang Zhao, Yunhao Liu, and Lionel M Ni. Vire: Active rfid-based localization using virtual reference elimination. In *Proceedings of IEEE ICPP 2007*.
- [11] Krishna Chintalapudi, Anand Padmanabha Iyer, and Venkata N Padmanabhan. Indoor localization without the pain. In *Proceedings of ACM MobiCom 2010*.
- [12] Xin Li, Yimin Zhang, and Moeness G Amin. Multifrequency-based range estimation of rfid tags. In *Proceedings of IEEE RFID 2009*.
- [13] Tianci Liu, Lei Yang, Qiongzhen Lin, Yi Guo, and Yunhao Liu. Anchor-free backscatter positioning for rfid tags with high accuracy. In *Proceedings of IEEE INFOCOM 2014*.

- [14] Han Ding, Chen Qian, Jinsong Han, Ge Wang, Wei Xi, Kun Zhao, and Jizhong Zhao. Rfipad: Enabling cost-efficient and device-free in-air handwriting using passive tags. In *Proceedings of IEEE ICDCS 2017*.
- [15] 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*.
- [16] Han Ding, Chen Qian, Jinsong Han, Ge Wang, Zhiping Jiang, Jizhong Zhao, and Wei Xi. Device-free detection of approach and departure behaviors using backscatter communication. In *Proceedings of ACM UbiComp 2016*.
- [17] Robert Miesen, Fabian Kirsch, and Martin Vossiek. Holographic localization of passive uhf rfid transponders. In *Proceedings of IEEE RFID 2011*, pages 32–37. IEEE, 2011.
- [18] Andreas Parr, Robert Miesen, and Martin Vossiek. Inverse sar approach for localization of moving rfid tags. In *Proceedings of IEEE RFID 2013*.
- [19] Jue Wang and Dina Katabi. Dude, where’s my card?: Rfid positioning that works with multipath and non-line of sight. In *Proceedings of ACM SIGCOMM 2013*.
- [20] 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*.
- [21] Lei Yang, Jiannong Cao, Weiping Zhu, and Shaojie Tang. Accurate and efficient object tracking based on passive rfid. *IEEE Transactions on Mobile Computing*, 14(11):2188–2200, 2015.
- [22] 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*.
- [23] Jinsong Han, Chen Qian, Xing Wang, Dan Ma, Jizhong Zhao, Wei Xi, Zhiping Jiang, and Zhi Wang. Twins: Device-free object tracking using passive tags. *IEEE/ACM Transactions on Networking*, 24(3):1605–1617, 2016.
- [24] Lei Yang, Qiongzhen Lin, Xiangyang Li, Tianci Liu, and Yunhao Liu. See through walls with cots rfid system! In *Proceedings of ACM MobiCom 2015*.
- [25] Marwan Younis, Christian Fischer, and Werner Wiesbeck. Digital beamforming in sar systems. *IEEE Transactions on Geoscience and Remote Sensing*, 41(7):1735–1739, 2003.
- [26] Xiaoyi Fan, Wei Gong, and Jiangchuan Liu. i2tag: Rfid mobility and activity identification through intelligent profiling. *ACM Transactions on Intelligent Systems and Technology*, 9(1):5, 2017.
- [27] Muhammad Raisul Alam, Mamun Bin Ibne Reaz, and Mohd Alauddin Mohd Ali. A review of smart homes—past, present, and future. *IEEE Transactions on Systems, Man, and Cybernetics*, 42(6):1190–1203, 2012.
- [28] Jiang Xiao, Zimu Zhou, Youwen Yi, and Lionel M Ni. A survey on wireless indoor localization from the device perspective. *ACM Computing Surveys*, 49(2), 2016.

- [29] Jose M Chaquet, Enrique J Carmona, and Antonio Fernández-Caballero. A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding*, 117(6):633–659, 2013.
- [30] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*, 46(3), 2014.
- [31] Andrew Spielberg, Alanson Sample, Scott E Hudson, Jennifer Mankoff, and James McCann. RapID: A Framework for Fabricating Low-Latency Interactive Objects with RFID Tags. In *Proceedings of ACM CHI 2016*.
- [32] Pengyu Zhang, Jeremy Gummesson, and Deepak Ganesan. Blink: A high throughput link layer for backscatter communication. In *Proceedings of ACM MobiSys 2012*.
- [33] Chuyu Wang, Lei Xie, Wei Wang, Tao Xue, and Sanglu Lu. Moving tag detection via physical layer analysis for large-scale rfid systems. In *Proceedings of IEEE INFOCOM 2016*.
- [34] Jihoon Ryoo and Samir R Das. Phase-based ranging of rfid tags with applications to shopping cart localization. In *Proceedings of ACM MSWiM 2015*.
- [35] Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580, 2007.
- [36] Liang Zhou. Mobile device-to-device video distribution: Theory and application. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 12(3):38, 2016.
- [37] Stefan Knerr, Léon Personnaz, and Gérard Dreyfus. Single-layer learning revisited: a stepwise procedure for building and training a neural network. In *Neurocomputing*, pages 41–50. Springer, 1990.
- [38] Joshua D Griffin and Gregory D Durgin. Complete link budgets for backscatter-radio and rfid systems. *IEEE Antennas and Propagation Magazine*, 51(2):11–25, 2009.
- [39] Cory Hekimian-Williams, Brandon Grant, Xiuwen Liu, Zhenghao Zhang, and Piyush Kumar. Accurate localization of rfid tags using phase difference. In *Proceedings of IEEE RFID 2010*.
- [40] Abdelhamid Djouadi, Oe. Snorrason, and FD Garber. The quality of training sample estimates of the bhattacharyya coefficient. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):92–97, 1990.
- [41] Jake K Aggarwal and Michael S Ryoo. Human activity analysis: A review. *ACM Computing Surveys*, 43(3):16, 2011.
- [42] Low Level Reader Protocol. Version 1.0. 1. *EPCglobal Inc, Aug, 2007*.
- [43] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.

- [44] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27, 2011.
- [45] Benjamin Wagner and Dirk Timmermann. Adaptive clustering for device free user positioning utilizing passive rfid. In *Proceedings of ACM UbiComp 2013*.
- [46] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [47] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
- [48] Jinsong Han, Han Ding, Chen Qian, Wei Xi, Zhi Wang, Zhiping Jiang, Longfei Shang-guan, and Jizhong Zhao. Cbid: A customer behavior identification system using passive tags. *IEEE/ACM Transactions on Networking*, 24(5):2885–2898, 2016.
- [49] Teng Wei and Xinyu Zhang. Gyro in the air: Tracking 3d orientation of batteryless internet-of-things. In *Proceedings of ACM MobiCom 2016*.
- [50] Mathieu Aubry and Bryan C Russell. Understanding deep features with computer-generated imagery. In *Proceedings of IEEE CVPR 2015*.
- [51] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of IEEE CVPR 2016*.
- [52] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of NIPS 2012*.
- [53] 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*.
- [54] 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*.
- [55] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of AISTATS 2011*.
- [56] Jinsong Han, Chen Qian, Xing Wang, Dan Ma, Jizhong Zhao, Pengfeng Zhang, Wei Xi, and Zhiping Jiang. Twins: Device-free object tracking using passive tags. In *Proceedings of IEEE INFOCOM 2014*.
- [57] Han Ding, Jinsong Han, Alex X Liu, Jizhong Zhao, Panlong Yang, Wei Xi, and Zhiping Jiang. Human object estimation via backscattered radio frequency signal. In *Proceedings of IEEE INFOCOM 2015*.
- [58] Jia Liu, Min Chen, Shigang Chen, Qingfeng Pan, and Lijun Chen. Tag-compass: Determining the spatial direction of an object with small dimensions. In *Proceedings of IEEE INFOCOM 2017*.

- [59] Yuxiao Hou, Yanwen Wang, and Yuanqing Zheng. Tagbreathe: Monitor breathing with commodity rfid systems. In *Proceedings of IEEE ICDCS 2017*.
- [60] Jie Xiong and Kyle Jamieson. Arraytrack: A fine-grained indoor location system. In *Proceedings of USENIX NSDI 2013*.
- [61] Wei Gong and Jiangchuan Liu. Robust indoor wireless localization using sparse recovery. In *Proceedings of IEEE ICDCS 2017*.
- [62] GD Bergland. A guided tour of the fast fourier transform. *IEEE spectrum*, 6(7):41–52, 1969.
- [63] Weng Cho Chew and Weng Cho Chew. *Waves and fields in inhomogeneous media*, volume 522. IEEE press New York, 1995.
- [64] Wei Wang, Alex X Liu, and Muhammad Shahzad. Gait recognition using wifi signals. In *Proceedings of ACM UbiComp 2016*.