Study on Efficient Crowdsensing-based Indoor Localization and Their Applications

高, 路路

Study on Efficient Crowdsensing-based Indoor Localization and Their Applications

Lulu Gao

August 2023

Department of Advanced Information Technology

Graduate School of Information Science and Electrical Engineering

Kyushu University

Contents

List of Figures

List of Tables

Abstract

With the increasing user demands for the ubiquitous availability of location-based services, and the acknowledgement of their substantial business prospects in many fields, including healthcare, security and entertainment, researchers have extensively studied indoor localization techniques that do not rely on the Global Positioning System (GPS) or other localization technologies that do not work well in indoor environments. The explosive growth and wide proliferation of mobile devices (e.g., smartphones, smartwatches, in-vehicle sensors, etc.), the majority of which are smartphones, led to a huge boost in smartphone-based indoor localization approaches due to its integration with rich advanced sensors, including Wi-Fi and inertial sensors. Consequently, Wi-Fi based localization is becoming a popular approach for providing location based services in indoor environment and pedestrian dead reckoning (PDR) based on inertial sensor readings has become one of the most practical methods for indoor location inference. However, different technologies come with their own challenges and shortcomings when precise positioning is concerned with the indoor environment.

Due to the labor-intensive and time-consuming tasks of radio signature collection in Wi-Fi based indoor localization systems, it is hard to build a comprehensive radio map constructed with received signal strength (RSS) for location prediction. Although crowdsensing could solve the problem of radio signature collection, there are various uncertainties about the location annotations contributed by the crowd, which would affect the performance of the localization model. To address such is-

sues and realize efficient indoor localization systems based on Wi-Fi fingerprint, we propose a crowdsensing-based indoor localization framework, ALCIL, which utilizes the active learning technique to collect the informative data to improve the performance of the model under a certain cost. We then employ global and local optimization strategies considering the multiple attributes of locations to improve the accuracy of location prediction in different dimensions. In addition, we propose a sample selection method based on stream-based active learning so as to improve the quality of radio maps and enhance the performance of the indoor localization model without penalizing the location-annotation process. The effectiveness of the proposed framework is verified through the experiments in the context of practical multi-story buildings. Our experiment shows that the proposed method can localize users' mobile devices accurately at the given fixed budget.

Regarding the PDR, which suffers from the severe position error accumulation with time due to the sensor drift. Activity landmark, which is available in many scenarios without additional deployment cost, is employed to calibrate the cumulative error. We propose an indoor navigation based on landmark without extra deployment cost, in which landmark is recognized by an unsupervised feature learning method to automatically extracts and selects the features to reduce the effort of data processing. The proposed method jointly trains denoising autoencoder implemented by convolutional neural network (CNN) and long short-term memory (LSTM) neural networks producing a compact feature representation of the data to identify landmarks. Besides, the relative distance between different landmarks is estimated by PDR to generate the indoor landmark map with the help of the multidimensional scaling technique. The effectiveness of the proposed framework is verified through experiments in the context of practical buildings. Furthermore, it was applied to the proposed indoor spatiotemporal contact awareness framework (iSTCA) due to the spread of COVID-19, which explicitly considers the self-containing quantitative contact analytics approach with spatiotemporal infor-

mation to provide accurate awareness of the virus quanta concentration in different origins at various times. Smartphone-based PDR is employed to precisely detect the locations and trajectories for distance estimation and time assessment without the need to deploy extra infrastructure. Another custom deep learning model is designed, composing of bidirectional long short-term memory (Bi-LSTM) and multi-head CNNs for extracting the local correlation and long-term dependency to recognize landmarks. By considering the spatial distance and time difference in an integrated manner, we can quantify the virus quanta concentration of the entire indoor environment at any time with all contributed virus particles. We conducted an extensive experiment based on practical scenarios to evaluate the performance of the proposed system, showing that the average positioning error is reduced to less than $0.7m$ with high confidence and demonstrating the validity of our system for the virus quanta concentration quantification involving virus movement in a complex indoor environment.

To make the proposed activity landmark-based indoor localization system more efficient, we conduct an extensive study on crowdsensing-based activity landmark recognition. Federated learning (FL) is a distributed framework that enables multiple parties to enable the collaborative learning without uploading the data of each participant. However, the assumption of FL is to rely on the annotated data on client, which is difficult to acquire the annotations for sensor-based activity recognition on all uploaders due to the lack of expertise or resource. Moreover, a general model is not suitable for each person because of the data heterogeneity, resulting from the different physical characteristics and various contextual information. To this end, in this work, we propose a semi-supervised learning method for personalized federated human activity recognition (HAR), in which clients have completely unlabeled data, while the server has a small amount of labeled data contributed by volunteers. Clients conduct unsupervised learning on autoencoders with locally unlabeled data to collaboratively learn a general rep-

resentation model. Server conducts supervised learning on an activity classifier with labeled data stored on the server. After that, the shared global model is personalized using individually pseudo-labeled data on each client side, wherein both confidence and uncertainty are taken into account concurrently, with the aim of achieving a balanced selection for assigning pseudo labels to samples. We conduct extensive experiments using two different real-world HAR datasets, demonstrating the effectiveness of the proposed methods.

Acknowledgments

Until undertaking this Ph.D., I received warm help from so many people and had a valuable experience. I would like to thank all of them.

First, I would like to thank my supervisor Professor Shin'ichi Konomi for his continuous support, advice, and patience throughout my graduate studies.

Also, I would like to express my gratitude to my parents and family in China for teaching me independence and instilling a sense of wonder and adventure in me. They always believing me and supporting me, helping me face every challenge during my Ph.D. studies. It would be impossible for me to complete my study without their understanding and encouragement.

Finally, thanks to all my friends for all the memorable moments. I am very glad to be a part of our laboratory at Kyushu University and very grateful to all the members of our laboratory. The research experience during my Ph.D. study has deepened my understanding of research. My horizons broadened and life changed. The experience during my Ph.D. study is really valuable and will encourage me to face all the challenges and adventures in the future!

Chapter 1

Introduction

In recent years, various Location-Based Services (LBS) have significantly penetrated into many aspects of people's daily life with the explosive growth and large-scale proliferation of smart devices and other wireless mobile devices. The primary objectives of LBS are providing users with location identification, navigation assistance in unfamiliar places such as shopping centers and airports and other services to meet the positioning demand of different groups under varied scenarios [106]. Particularly, the contact tracing service has witnessed a tremendous increase in interest as a non-pharmaceutical and practical measure for disease prevention, primarily due to its effectiveness in reducing the transmission of contagious diseases, such as COVID-19 [54]. Consequently, these types of services have become increasingly integrated into our societal activities, which indicates the importance of precise location information for offering better services and a higher quality of satisfaction for users. The widely accepted and popular localization system, the global positioning system (GPS) does not work well in a complicated indoor environment due to the attenuation of satellite signals by obstacles like walls, furniture, and human beings, failing to meet the needs for accurate indoor localization.

Nowadays, advancements in hardware technologies have led to novel sensing technologies, which provide compact yet powerful sensors with the ability to cap-

ture a wide variety of readings. Hence, various solutions have been studied to address the constraints of GPS and other global navigation satellite systems (GNSS), such as the wireless signal-based methods using Wi-Fi signals, ultra-wideband (UWB), Bluetooth, radio frequency identification (RFID), Zigbee, visible light and acoustic-based technologies [161]. Within this framework, the unique wireless signal characteristics, usually represented by the received signal strength indicator (RSSI), at a particular spatial location serve as the fingerprint for that position, thereby allowing the estimation of a user's location through fingerprint matching with a pre-established position–fingerprint relationship database. Among all these radio communication technologies, Wi-Fi signal is the ideal candidate for indoor localization and becomes one of the extensively studied localization technologies due to the popularity of smartphones with embedded Wi-Fi chips [124]. Moreover, the ubiquity of sensor-rich smartphones with wireless communication capabilities not only eases the physical measurements of ambient signals but also presents an opportunity to investigate the fused sensor data for positioning, further promoting the development of indoor localization. An algorithm, namely, pedestrian dead reckoning (PDR), can utilize the estimated motion dynamic (such as speed, heading, orientation, or motion states) of smartphone carrier's via fused inertial sensor data, including accelerometer, magnetometer, and gyroscope, to locate a pedestrian in GNSS challenging environments. Since there is no need for extra infrastructure and no coverage limitation, PDR has become one of the mainstream indoor localization methods [44].

Different localization techniques have different advantages and limitations in terms of coverage, accuracy, requirement for infrastructure and cost of deployment. Therefore, the merits and demerits of an indoor localization system should be comprehensively considered first before the positioning system is designed. Among the factors, the deployment cost plays a paramount role in the large-scale practical application to provide the indoor LBS and the high expenses would significantly

hinder the adoption of a localization system. The deployment cost essentially comprises *pre-configured infrastructure, data acquisition and data processing and analysis.*

The prevalent Wi-Fi fingerprinting-based indoor localization techniques can be deployed without the need for additional infrastructure due to existing Wi-Fi access points. However, the Wi-Fi signals would not be accessed because of the out-of-coverage and signal shielding, which are common in developing countries with poor information and communication technology infrastructure. Moreover, since existing Wi-Fi networks are typically configured for communication instead of localization and the RSSI is sensitive to complex indoor environments, great effort in terms of time and labor should be expended on the collection and updating of the fine-grained fingerprint database to achieve high localization accuracy [107]. The investment in data engineering, including data labeling, data preprocessing and data feature computing, is worthwhile with raw datasets, which may not yet be labeled and cleaned, before performing data analysis effectively for location inference. Thus, we should devote additional effort to accurately annotate and process the data for subsequent feature extraction, thereby enabling efficient data analysis to infer the location. The significant PDR technique is self-contained, which continuously provides relative location estimation without considering wireless signal coverage or infrastructure, and appears to require no deployment cost. However, it suffers from accumulated errors due to sensor drift, resulting in large deviations over time. Wi-Fi fingerprinting technique is usually integrated with PDR techniques to reduce errors, in which both the drift problem of PDR and the failure of wireless localization methods for continuous localization (tracking) [161, 98].

Consequently, in this thesis, we focus on the investigation of various techniques to reduce deployment costs, including pre-established infrastructure cost, data collection cost and data processing and learning cost, thus achieving efficient indoor

localization systems and their practical applications.

1.1 Major Research Contributions

In order to attain the overall objective, the major research contributions of this dissertation are briefly described below:

First of all, the crowdsensing technique is applied to reduce the cost of labor-intensive and time-consuming fingerprint collection for radio map construction in popular Wi-Fi fingerprinting-based indoor localization systems with Wi-Fi signal covered. Although crowdsensing could solve the problem of radio signature collection, there are various uncertainties about the location annotations contributed by the crowd, which would affect the performance of the localization model. To address such issues and realize efficient indoor localization systems based on Wi-Fi fingerprints, we propose a crowdsensing-based indoor localization framework, ALCIL, which utilizes the active learning technique to collect the informative data to improve the performance of the localization model under a certain cost and the machine learning approaches to learn the strong patterns heuristically for accurate location estimation. We conducted extensive experiments to demonstrate the effectiveness of the proposed framework, locating users' mobile devices efficiently at the given fixed budget.

Secondly, we develop an activity landmark-based PDR to realize efficient indoor navigation with high precision in various environments regardless of the configuration status of wireless access points. The activity landmark, which stands for a location point that imposes a certain pattern on the motion sensor readings, is properly recognized with the sensors inside smartphones and applied to the calibration of the accumulated errors in PDR at no extra cost. To further reduce the deployment cost of data processing, the feature of activity landmarks is extracted by an unsupervised feature learning method without manual calculation,

producing a compact representation for landmark identification. The validity of the proposed approach to provide efficient indoor navigation was verified with 9,271 samples collected using a crowdsensing technique for 27 landmarks in the context of practical buildings.

Thirdly, the above-mentioned activity landmark-based PDR was utilized to develop an indoor spatiotemporal contact awareness framework (iSTCA) to enable precise contact tracing of Covid-19 in various indoor environments. The iSTCA explicitly considers the self-containing quantitative contact analytics approach with spatiotemporal information to provide accurate awareness of the virus quanta concentration in different areas at various times. PDR technique is employed to precisely detect the locations and trajectories for distance estimation and time assessment with recognized activity landmarks using a designed deep learning model. Thus, the contact-tracing feature within this framework allows for a more detailed and quantitative understanding of indoor exposure to virus with virus lifespan considered, which is difficult to discern using conventional techniques based on relative distances between devices. Furthermore, the integration of activity landmark-based PDR positioning methods with a spatiotemporal model of virus concentration variations has led to the development of a cost-effective solution that can be employed in diverse indoor environments, including those lacking ICT. We performed an evaluation using actual movement history data collected in a building and confirmed the effectiveness of the developed system.

Lastly, we employ federated learning -based activity landmark recognition to make the activity landmark-based indoor localization system more efficient and useful. We proposed a method based on semi-supervised learning and federated learning (FL) that can perform personalized human activity recognition (HAR) while considering user privacy and cost reduction on data annotation, in which FL clients have completely unlabeled data, while the FL server has a small amount of labeled data contributed by volunteers. This approach is characterized by collab-

orative semi-supervised learning conducted between client devices and a server for the general model training, using pseudo-labels that take into account trustworthiness and uncertainties for model personalization. This method can be applied to efficiently identify activity landmarks in indoor localization techniques. Through evaluation experiments with two different real-world activity recognition datasets, the persuasive accuracy improvement is confirmed compared to conventional techniques.

1.2 Thesis Organization

This thesis is composed of the following chapters.

Chapter 1 introduces the overview of research motivations of this thesis and our approaches and contributions.

Chapter 2 presents the background and gives an overview of related work, including research in crowdsensing-based indoor localization and the applications of indoor localization, especially the contact tracing.

Chapter 3 describes an efficient crowdsensing-based indoor localization using active learning techniques to reduce the deployment cost. We present the detail of the methods and evaluation results.

Chapter 4 provides an activity landmark based PDR to realize efficient indoor navigation in various environments regardless of the configuration status of wireless access points.

Chapter 5 introduces the iSTCA framework for precise contact tracing of the Covid-19 virus in various indoor environments, in which a more detailed and quantitative understanding of indoor exposure to virus can be achieved by considering virus lifespan. We provide the detail of iSTCA and the evaluation.

Chapter 6 presents our study on efficient activity recognition using semi-supervised learning and FL, which involve clients with totally unlabeled data and a server with

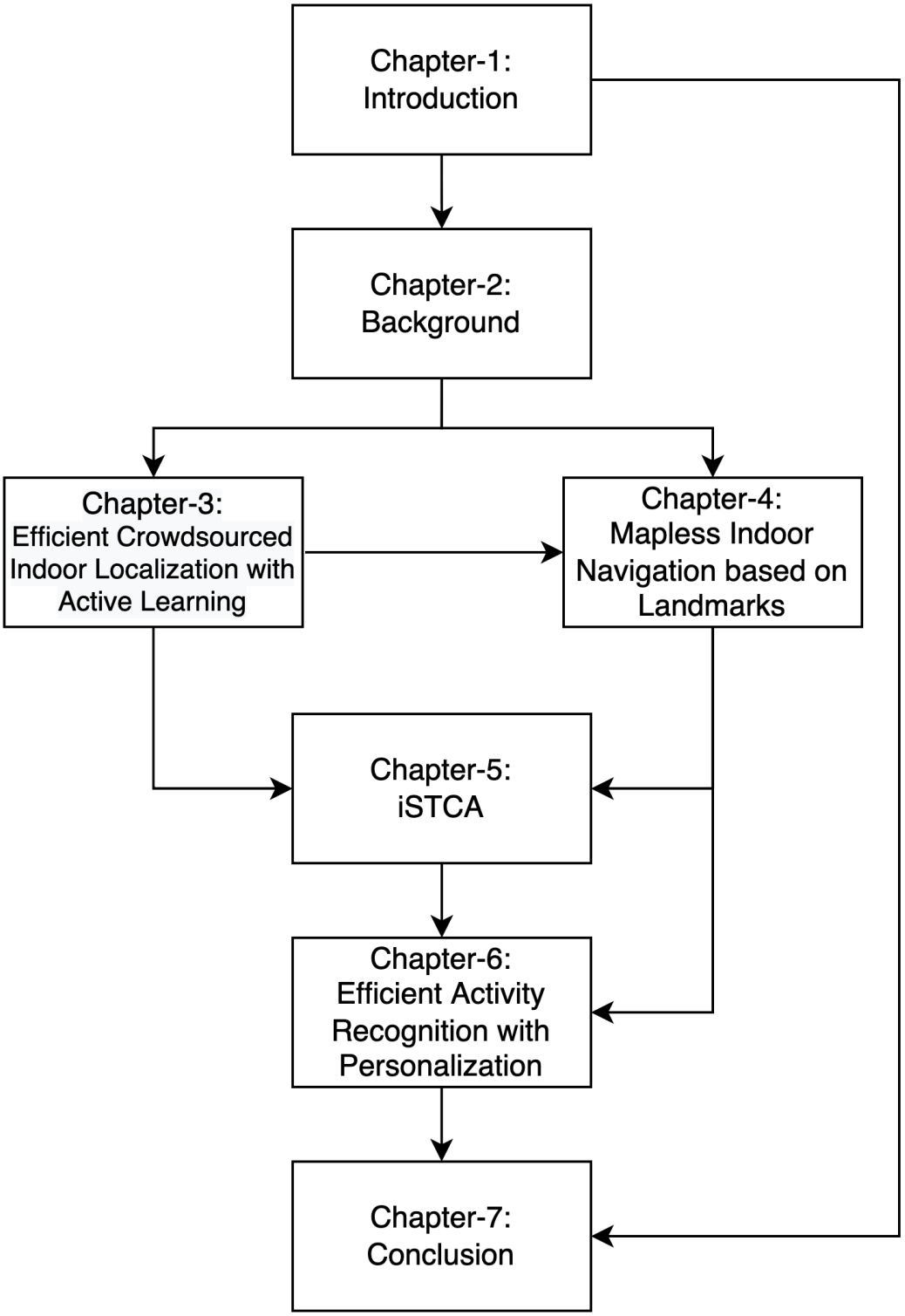


Figure 1-1: Structure of the thesis

a small amount of labeled data and utilize pseudo-labels with trustworthiness and uncertainties considered for model personalization.

Finally, chapter 7 concludes our contributions of this thesis and discusses possible future research.

Chapter 2

Background

2.1 Indoor Localization Systems

Indoor localization systems hold significant importance in expanding LBS into indoor settings where GNSS is unreliable. Such systems are typically integrated and deployed within homes and buildings, facilitating interaction with smart devices to provide them with spatial context. Therefore, over the past decades, numerous innovations have emerged to address the complexity within indoor environments and to strive for higher accuracy in localization through the utilization of various techniques and diverse technologies. Typically, the following three categories of techniques are often used: triangulation-based, fingerprinting-based and dead reckoning techniques [161, 124].

Triangulation-based positioning systems mainly use radio technology, such as Wi-Fi, Bluetooth, UWB and RFID. The detected physical measurements of the signals are firstly converted to some geometric parameters (e.g., distance and angle), and then determine the actual location with trilateration or triangulation approaches, which include Time of Arrival (ToA), Time Difference of Arrival (TDoA), Time of Flight (ToF), Return Time of Flight (RToF), Angle of Arrival (AoA). However, they usually require dedicated devices to achieve stable measurements,

which greatly limit their applications in reality. Furthermore, the deployment cost of those dedicated devices also includes careful calibration of indoor environments with experts.

Fingerprinting technique is widely adopted as the basic scheme of location determination due to the already installed wireless local area network infrastructure in some buildings. The main idea is to calculate the location of the object by matching a set of measurements called fingerprints with a set of fingerprints that are collected and stored in a pre-built database. A fingerprint is the specifications or measurements from various signals at a certain location, such as the received signal strength (RSS) of Wi-Fi access points. Fingerprinting comprises an offline training phase and an online location estimation phase. During the training phase, a fingerprint database is created within the specified area of interest. In the subsequent online localization phase, the target's location is estimated by comparing the measured fingerprints against the pre-established database. Apart from the cost of previously configured infrastructure, the notable expenses associated with the deployment of fingerprinting techniques are the collection, updating and processing of fine-grained fingerprints, which are time-consuming and labor-intensive.

The dead reckoning (DR) technique relies on inertial measurement unit (IMU) sensors to approximate relative location with little or no infrastructure to be deployed, in which the IMU sensors can track target movement by the equipped accelerometers, gyroscopes, and magnetometers. An inertial navigation system that provides accurate directional information uses DR and is applied. In this study, we focus on DR for pedestrians called PDR. The latest location of the user will be calculated based on the previously determined position and displacement derived from the achieved measurements of moving directions, velocity, and interval sampling using IMU sensors. However, it demands a precise initial position to avoid errors and suffers from cumulative inaccuracy due to sensor drift and contextual interference, leading to deviation in position estimation. Although PDR

can be deployed without additional equipment, it is essential to establish reference points, such as the Wi-Fi fingerprints, for the purpose of error correction and accurate localization.

2.1.1 Crowdsensing-based Indoor Localization

In recent years, there has been a rapid increase in smartphones and wearable devices outfitted with various built-in sensors, causing crowdsensing-based mobile applications to become increasingly widespread. Crowdsensing is an extension of crowdsourcing that refers to leveraging the intelligence of distributed participants to accomplish one complex task. Crowdsensing presents a unique opportunity for utilizing inertial sensors in mobile devices, as human mobility can offer unparalleled sensing coverage and data transmission without incurring the prohibitive costs associated with traditional sensor networks. Generally, within this approach, individuals acquire data utilizing sensors embedded within their smartphones, navigate the environment, and extract information for phenomena investigation that yields benefits for themselves [65, 101, 169].

As crowdsensing can take advantage of human mobility and the mobile devices they carry allows for extensive environmental sensing and data transmission without the need for deploying additional infrastructure, the possibility of building large-scale sensing applications with efficiencies is greatly enhanced. Thus, the crowdsensing paradigm is perfectly suitable for fingerprinting-based indoor localization, in which the construction of a timely fingerprint database, also known as a radio map in Wi-Fi fingerprinting-based localization, needs regular site surveys with an expert in the target area. Many researchers used crowdsensing technology to propose efficient indoor localization systems in order to reduce the deployment costs, time-consuming and labor-intensive, such as LiFS, Walkie-Markie [156, 118]. Besides, the extensive engagement of anonymous users can also benefit indoor localization by facilitating the automated development of indoor floor plans and

establishing navigation spaces that include all walkable paths within particular environments, thereby further contributing to the advancement of intelligent indoor LBS.

2.1.2 Applications of Indoor Localization

The application of indoor localization systems has seen a drastic increase in use around the world. They vary from tracking and navigation to asset management and autonomous vehicle navigation to facilitate our daily life. The following provides a brief overview of these applications.

The primary aim of indoor localization systems is to locate the position of an object, track the position sequence for any moving object, and facilitate navigation in indoor environments. In construction management, this technology has been employed to monitor laborers, materials, and machinery. Doctors or nurses can track the location of their patients and track their mobility to ensure patient safety in the hospital using a localization system [161, 124]. Indoor localization can be employed for user location-based authentication purposes, such as implementing location-based access control for sensitive business information and allocating hardware resources based on the user's position. Surveillance can be another paradigm where a suspect may be tracked indoors.

Asset management can fundamentally benefit from tracking, as it would allow different businesses to track the location of their assets, allowing for better inventory management. In shopping malls or markets, personalized marketing and advertisement can be achieved based on the customers' location. In a warehouse, there is a high demand for tracking assets such as electronic equipment and manufacturing objects. Key requirements for indoor asset tracking include monitoring goods throughout the supply chain, implementing autonomous tracking systems, and addressing theft protection and other security-related concerns [106].

Autonomous vehicle navigation is another significant application of indoor lo-

calization. In indoor environments, unmanned aerial vehicles (UAVs) and unmanned ground vehicles (UGVs) have obtained significant interest for their extensive capabilities in aerial surveillance support, emergency evacuation routes determination, search and rescue operations conduction, and many more. Additionally, autonomous navigation is crucial for various equipment, including service robots, self-driving cars, and smart wheelchairs. Since the working environment is likely to be an unstructured and unknown area for these vehicles, indoor localization and navigation are particularly indispensable to ensure safe navigation without any collision in an indoor area [124].

The aforementioned applications illustrate that localization can offer efficient and effective services across various application areas, with the underlying objective of assisting users and customers. In the future, we anticipate a broader range of services and applications that will be facilitated by indoor localization.

2.2 Contact Tracing

Infectious diseases, which can be spread, directly or indirectly, from one person to another, pose a serious threat to human health, national economy and societal development, especially COVID-19 that has struck a devastating blow to the global economy. In response to the spread of infectious viruses, contact tracing is a well-established part of the management of disease outbreaks, which has become one of the most critical measures to effectively curb the spread of the virus. Contact tracing involves identifying, notifying, and quarantining people who have had close contact with new cases in order to prevent further transmission within the community [57, 16]. When systematically applied, contact tracing will break the chains of transmission of infectious disease and is thus an essential public health tool for controlling infectious disease outbreaks [54].

In history, contact tracing has been widely used in the control of infectious

diseases and has become a pillar of communicable disease control in public health with a lot of human and material resources for decades. Thus, the traditional contact tracing methods suffer from some major limitations when dealing with large-scale outbreaks of infectious diseases, such as the pandemic of COVID-19. Relying mainly on manual interviews and investigations, the traditional tracing methods are time-consuming and labor-intensive. The more recent digital contact tracing methods are far more efficient and have helped greatly reduce the consumption of human labor and other material resources [61]. Given their great advantages, various digital information technologies in contact tracing have been applied in many countries as effective means of COVID-19 inhibition [5].

Usually, digital contact tracing applications are installed on portal devices, typically smartphones, to conveniently and intelligently realize tracing with the help of existing sensors based on various technologies, such as GNSS, Bluetooth, and Wi-Fi [5, 26]. There are typically two approaches for contact determinations, peer-to-peer proximity detection-based and geolocation-based. Peer-to-peer proximity can be estimated by the received signal strength (RSS) of wireless signals, such as Bluetooth and UWB. Many contact tracing applications rely on Bluetooth-based systems that can directly detect whether users came in proximity of each other. The proximity can be approximated by the strength of the signal, which is very short and can be obstructed by buildings and walls. Therefore, in a high-risk environment for close contact like buildings or public transit, it can reflect functional proximity more effectively and accurately. However, with applications that evaluate exposure risk based on Bluetooth, proximity exchange is in essence not sufficient because of the fact that apart from the human-to-human interaction, some infectious diseases like coronavirus can also transmit through common environments or commonly touched surfaces[26]. This kind of contact can only be determined by geolocation-based approaches. The distance between two devices in geolocation-based approaches is precisely derived from the cross-examination

after obtaining the accurate location and trajectory with the help of localization techniques using various technologies, such as GNSS, Wi-Fi and PDR.

Contact tracing in indoor environments can complement the ones used in outdoor environments to enable comprehensive digital contact tracing. However, indoor contact tracing imposes unique technical challenges due to virus concentrations and unreliable GNSS signals in indoor environments. The virus concentrations, which play a critical role in calculating the amount of viruses we are exposed to and further assessing the infection risk, should be explicitly considered in indoor contact tracing applications [119, 140]. Although the virus concentration will gradually decrease due to inactivation, deposition, and air purification after the virus-laden droplets are exhaled, the poor air exchange rate, superspreaders, and more virulent variants will keep it at a relatively high concentration for a long time in an indoor environment. Li et al. utilized active Wi-Fi sensing to collect the data and leveraged signal processing approaches and similarity metrics to align and detect virus exposure with location-dependent virus concentration [72]. However, to accurately estimate the concentration, investigating the airborne transmission of these ejected particles is, thus, of fundamental importance in a closed environment because of the assemblage, in which human movement is implicitly involved to achieve the initial motion state of droplets. Tu et al. an epidemic contact tracing with Wi-Fi network and smartphone-based PDR, involving not only coarse-grained duration but also the fine-grained distance between students [134]. Although the motion state of humans can be acquired in the contact awareness system with PDR and Wi-Fi, the Wi-Fi network would not be accessed sometimes because of the out-of-coverage and signal shielding.

Chapter 3

A Cost-Effective and Quality-Ensured Framework for Crowdsourced Indoor Localization

3.1 Introduction

With the rapid development of wireless technology and pervasive computing, LBSs, such as social networks, tracking, navigation, recommendation and social distancing, have shown tremendous value [58, 96]. Particularly, contact tracing services have garnered considerable attention as a non-pharmaceutical and practical measure for disease prevention, primarily due to their effectiveness in reducing the transmission of infectious diseases, such as COVID-19. The essence of LBSs is to locate the user and then provide useful information at the appropriate time and right location. Therefore, the performance of the LBSs is greatly influenced by the accuracy of localization measures [34].

Localization is a mechanism for determining the spatial relationship based on the physical position or logical position of different entities [108]. Depending on the target environment, it can be divided into outdoor localization and indoor

localization. GNSS, such as GPS, Galileo Satellite Navigation (Galileo), BeiDou Navigation Satellite System (BDS) and other satellite systems can locate us precisely and reliably in an outdoor open environment, which is widely exploited in our everyday lives [6]. However, GNSS does not perform well in urban canyons, underground environments and indoor environments in which we spend most of the time because of the lack of a unified infrastructure and the weak signal strength of satellites due to the absence of line of sight, the attenuation of satellite signals as they cross through physical objects, especially walls, and noise interference, resulting in inaccurate localization of us or devices [27, 17]. What's more, indoor localization has played an important role in tracking and navigation, especially in large buildings such as shopping malls and underground parking lots. Therefore, to fill the gap, indoor localization technology has been extensively researched for many years.

Indoor localization is the process of obtaining the location of a user or device in an indoor setting or environment, which has been well-developed with the joint effort of researchers and engineers in the past few decades. For different scenarios, researchers investigate lots of technologies to build indoor localization systems: RFID, Bluetooth, Zigbee, UWB, wireless local area network (WLAN), infrared ray (IR), ultrasound, magnetic field and visible light. Among the aforementioned technologies, as an infrastructure-free technology, WLAN (or Wi-Fi) is widely used because of the ubiquity in deployment and accessibility on the device [161, 132]. Many indoor localization methods based on Wi-Fi using different techniques that mainly cluster into trilateration (including TOA, TDOA and AOA) and fingerprinting have been proposed. In the trilateration method, the distance to three points that could obtain the relative location of a user via basic geometry and trigonometry is estimated according to the RSSI and the path-loss propagation model [4]. In the fingerprinting approach, a number of fingerprints from different grids within the target area are collected to obtain the training dataset to train

a machine-learning algorithm that could predict the coordinates online given the measured fingerprints at the location of the user. Fingerprinting techniques apply machine learning algorithms to find the spatial pattern behind the sensed RSS data in the target area to reduce the effects of RSS fluctuation in trilateration caused by the complicated indoor environment, plentiful multipath fading and various non-line-of-sight (NLOS) conditions that result in inaccurate propagation mode even though calibrating considerable samples, which is widely researched for accurately locating the user now [56].

The RSSs from detected access points (APs) in a different position can be recognized as the fingerprint in fingerprinting indoor localization system, which includes an offline training phase and an online location prediction phase. Since machine learning algorithms require plentiful training data to achieve great performance, the fingerprinting method also demands a large number of fingerprints in the offline training phase to improve the accuracy of location prediction. Numerous fingerprints should be obtained from different positions of interested regions to construct the fingerprinting dataset with annotation by researchers to construct an elaborated radio map in the training phase. Subsequently, in the online estimation phase, machine-learning methods trained with the collected dataset could properly return the estimated user's position after conducting a location query using the stored radio map. However, owing to the RSS variance caused by environmental changes between the two phases, along with the limited availability of experts, the collection and annotation of fingerprints can be quite laborious and time-consuming [147, 166].

Crowdsourcing is a potential solution to solve the site survey problem because everyone could become a contributor. With the development of sensor technology and the popularity of wireless mobile terminal devices, such as laptops, smartphones and so forth, mobile devices integrate more and more sensors, including Wi-Fi and camera, leading to more and more powerful abilities of computing and

sensing, which make it possible to encourage ordinary mobile devices users to contribute their effort and large quantities and scalabilities of fingerprint data could be achieved, thus so-called crowdsourcing based indoor localization, reducing the burden of site survey. As the low deployment cost and efficient approach to constructing a radio map, much different crowdsourcing-based indoor localization systems have been proposed [166, 155, 43]. Crowdsourcing-based indoor localization techniques do reduce the burden on researchers, while some new variances like unsure annotation because of the lack of expert knowledge about positioning or Geographic Information System (GIS) are introduced. In fact, without the assistance of GNSS, specific measurements are demanded to calculate the location, which requires enough specialized knowledge and equipment. In a crowdsourcing scenario, we do not know the information about the task performer that may work in different fields with different capacities and are unable to obtain his/her precise location in detail, including height, latitude and longitude. Therefore, many researchers now bypass the process of data annotation, trying to automatically generate the annotation by some auxiliary sensors and find other data patterns from different perspectives to assist in positioning users. While the annotation is indispensable for achieving better performance of fingerprinting approach in indoor scenarios.

In this chapter, for crowdsourcing-based indoor localization, a framework, named ALCIL, Active Learning-based Crowdsourced Indoor Localization, which could reduce the efforts of site survey and ensure the performance of the positioning methods has been proposed. ALCIL can reduce the number of fingerprint data that constitute the radio map without affecting the performance of location prediction, using active learning. Active learning is a modern method in machine learning, aiming to reduce the sample size, and complexity by selecting the informative data according to informativeness measures and increase the accuracy of data tasks with minimal costs. Moreover, ALCIL can ensure the accuracy of

the fingerprint label via participants' relative annotation of all data and experts' re-verification of reduced instances with high informativeness. Location can be labeled by absolute position with precise height, latitude and longitude and logical position. The relative position associated with an indoor environment that is available for everyone even those without relevant expertise. To achieve ALCIL, participants only need to describe the uncomplicated relative position and the complex and specialized labeling work with informative data can be calculated by experts, reducing the annotation variances and cost.

To demonstrate the effectiveness of ALCIL, we have conducted extensive experiments over the dataset collected in West Zone of Kyushu University's Ito Campus to evaluate the proposed methods. We have developed an application to collect data and annotate data with relative information and physical values of different attributes of location. Experiments indicate that ALCIL can obtain the better dataset with high quality and accurate performance of indoor localization with the constraint of cost.

The remainder of this chapter is organized as follows. Related work about crowdsourced indoor localization and active learning is reviewed in Section 3.2. Section 3.3 described the theoretical methodology and the architecture of the proposed ALCIL system. Section 3.4 presents the experimental methodology and results in two datasets. Finally, the conclusion and future work are presented in Section 3.5.

3.2 Related Work

In this section, we discuss related works in fingerprinting-based indoor localization, crowdsourced indoor localization and active learning.

3.2.1 Crowdsourced Fingerprinting-based Indoor Localization

Ambient radio signals can be conveniently obtained to identify different locations, serving as fingerprints for location interference in indoor environment [110]. A plethora of indoor localization methods that adopt the radio-based fingerprint to predict a user's location has been proposed. The core idea is to build up a fine-grained radio map consisting of the fingerprint of each interested location, which the position can be achieved using matching algorithms. RADAR is the first attempt to apply the fingerprinting-based technique with a KNN matching algorithm in indoor localization, which utilizes the Wi-Fi fingerprint [169, 9]. Though the wide GSM, TV signals and FM radio signals are researched as fingerprints for positioning, Wi-Fi fingerprints can be recognized as the most representative fingerprinting-based in indoor localization due to the rapid development of wireless communication and the extensive deployment of WLAN infrastructure [138, 104, 102, 120].

Wi-Fi fingerprinting-based indoor localization can measure the RSSs from detected APs at target area to construct the detailed radio map on training phase, which will be used to determine the location via deterministic and probabilistic algorithms with different similarity metrics like Euclidean distance, Kullback-Leibler (KL) divergence and Jensen-Shannon (JS) divergence [159, 2, 92, 93, 3]. The RSS of target location can be calculated in deterministic approaches, whose accuracy is greatly affected by noise and variation which are tackled in probabilistic methods [33]. All these approaches will distinguish the location within the previously surveyed fingerprints database, whose accuracy and coverage are critical attributes to achieve better performance, which means labor-intensive and time-consuming [65].

Crowdsourcing is the most suitable approach to collect large-scale fingerprints because everyone can be the potential contributor using the terminal devices with

plenty of sensors in-built, including smartphones, PDA, and human mobility. The term crowdsourcing was first presented in 2006 by Jeff Howe [50]. Crowdsourcing was traditionally used to be a distributed problem-solving and production model, but now, it can be seen a promising approach to address some of the growing challenges associated with data collection and data processing, as demonstrated by Amazon Turk, Netflix, and the ESP game, which expand human computation [15, 165]. As a low-cost and efficient method to collaborate the intelligence of different people, crowdsourcing has been widely employed in the acquisition of Wi-Fi fingerprints [149, 167]. We can't achieve the complete training data just with data collection because annotation is the important part to train the prediction model. However, due to anyone can be a contributor, including people who do not have the expertise about location technique, and the label may be inaccurate. Although there are some solutions without site survey processing to realize indoor localization, label is inevitable to return accurate location prediction in indoor scenario [141, 1, 123].

3.2.2 Active Learning

Active learning is a modern method in machine learning, aiming to reduce the sample size, complexity, and increase the accuracy of the data tasks as much as possible with fewer data. The key hypothesis of active learning is that the learning mechanisms will be more intelligent if the learning algorithm can actively choose the most significant unlabeled data. An active learner will query only a small number of valuable unlabeled instances to be labeled by an oracle or annotator to automatically enlarge the labeled dataset in an intelligent manner [111].

There are three main scenarios that have been studied of active learning, membership query synthesis, stream-based selective sampling, and pool-based sampling [111, 73, 154].

- In query synthesis, any unlabeled instance can be queried by an active

learner, including the model-generated, even though it may have no practical meaning and cannot be labeled by a human annotator. While the other scenarios do not have this problem that cannot be labeled, because the learner must query the instances of what it thinks is important from the actual input pool.

- In stream-based selective sampling, the unlabeled instances will be queried sequentially by the learner [81]. And the learner will decide whether the instance be annotated or not.

- In pool-based sampling, many unlabeled instances are assumed to be available. In this kind of scenario, the learner should rank the entire unlabeled instances according to an informativeness measure, that is the pool of unlabeled instances, then, query the most informative one [81]. The main difference between pool-based sampling and stream-based selective sampling is that the former should evaluate all unlabeled data before selecting a query, while the latter just query the instance in sequence [112].

Selective sampling is the most relevant scenario to crowdsourcing-based indoor localization, including stream-based active learning and pool-based active learning. In the stream-based scenario, the decision of annotating the current piece of data can be determined by the informativeness measures, which is also applicable to data ranking for data labeling in pool-based active learning after we collect a certain amount of data. The measure of informativeness evaluation is vital in all active scenarios and can be parted into uncertainty selection, query by committee, expected objective change, and data-centered method [73, 12, 49].

- Uncertainty selection, which would query the most uncertain instance on the prediction of the current model.

- Query by committees (QBC), which queries the most disagreeing instance

of the committees' prediction. Each committee member is a different model based on the current training set.

- Expected objective change, which queries the instance that could make the maximizing impact on the objective. For example, maximizing model change, maximizing the generalization error reduction, and maximizing the output variance reduction.

- Data-centered method, which queries the most representative of the most informative instance.

It can be seen that active learning and crowdsourcing are critical technologies for optimizing data collection and processing, and many researchers have conducted a lot of studies on the ways to integrate them [42]. Lease suggests that crowdsourcing with active learning may provide new insights for better focusing annotation effort on the examples that will be most informative to the learner to accelerate model training, as well as reduce the cost of data annotation [67]. Costa, et al. propose methods of combining crowdsourcing and active learning. The methods were tested with the Jester data set, a text humor classification benchmark, and the result shows promising improvements [24].

ALSense is a novel active learning framework under crowdsourced scenarios that can be used for indoor localization, which considers the cost of data annotation, and its main goal is minimizing the prediction error of crowdsourced tasks within a fixed annotation cost. Crowdsourcers calculate the informativeness of current data based on their own simple model trained by the initial dataset and upload the data with higher informativeness. After the server collects a certain amount of data, pool-based sampling is utilized to determine the instances that need to be labeled from the collected data and requests the participant to annotate it [154]. Although ALSense can control the cost of annotation, it is necessary to ensure the label's accuracy, which is difficult to achieve in practice due to different

crowdsourced workers with diverse expertise, and just only target for the classification tasks. Hence, we propose ALCIL to minimize the predictions of indoor localization tasks, including classification and regression, subject to collection and annotation cost constraints under different crowdsourced workers.

3.3 Methodology

In this section, we analyze the annotation cost of Wi-Fi fingerprints and present the detailed framework of ALCIL. Besides, the optimization strategies are explained for indoor localization with multi-labels.

3.3.1 Problem Construction and Overview

To achieve Wi-Fi fingerprinting-based indoor localization, it is necessary that one or more people carries a terminal device with WLAN access and RSSs measurement from different APs to collect the fingerprints on various location of the target indoor region [85]. The fingerprints in different positions should be labeled with some identification, like the building number, floor number, room number, and physical coordinate to form the training dataset, called radio map, which is time-consuming and labor-intensive and requires some expertise or specific equipment.

Given a set of fingerprints of X, $\{x_1,\ x_2, \ldots, x_n\}$, where x_k is the RSSs vector of the k-th sample we collect and the value of it should belong to the value space R^x. A set of labels of Y, $\{y_1,\ y_2, \ldots, y_n\}$, where y_k is the labels of the k-th RSSs sample x_k, which may include more than one attribute, like height, latitude, and longitude, and the value of it should in the label space R^y. (X, Y) is trained in the offline phase to discover the reflection of $f : X \to Y$ to obtain the location in the online phase. Actually, f can be each one of hypothesis spaces F and the prediction error or interference error E of f can be equated by $E = \sum_{i=0}^{n}\ L(y_i, f\,(x_i))$ that we can measure the performance of f and it should be minimized, where L is

a specific measurement. No matter what the selection and the measurement of f, we all need accurate label y. However, the cost of obtaining Y is higher than collecting X, hence improving the quality of Y at a certain cost to make f perform better is essential. Under the restriction of cost, M, which we can obtain from the incentive method, only selected instances set X' of X can be annotated, and the label of X' we can get is Y'.

$$\begin{aligned} \min_{x_i' \in X',\ \ y_i' \in Y'} E &= \sum_{i=0}^{n} L\left(y_i', f\left(x_i'\right)\right) \\ s.t.\ \ \left|Y'\right| &\leq M \end{aligned} \tag{3.1}$$

3.3.2 Architecture

Figure 3-1 shows the proposed indoor localization system architecture, which can be divided into two parts: the mobile device and the positioning server. On the mobile device side, there is a local active learner which can detect the informativeness of collected fingerprints for the localization method based on utility measure, then decides whether to upload a data item with relative annotation, like the building number, floor number, room number and the distance to some obvious indoor objects. Moreover, the local active learner can be updated by the server in a certain period. The global active learner can select informative samples from the uploaded data summited by crowdsourcers based on informativeness measurement to remind us to re-calibrate the fingerprints, including the absolute height, latitude and longitude. Therefore, the uploaded data item with accurate annotations is added to the database of labeled fingerprints to train a better model. The indoor localization model and global learner are updated when a new re-calibrated fingerprint appears in the labeled database to make the system perform better.

For this system, a few calibrated fingerprints and the informativeness measure need to be settled down according to the localization approaches. A small amount of initial data with labels should be collected to accelerate the deployment, providing the localization service and incentive the crowdsourcer to participate, and

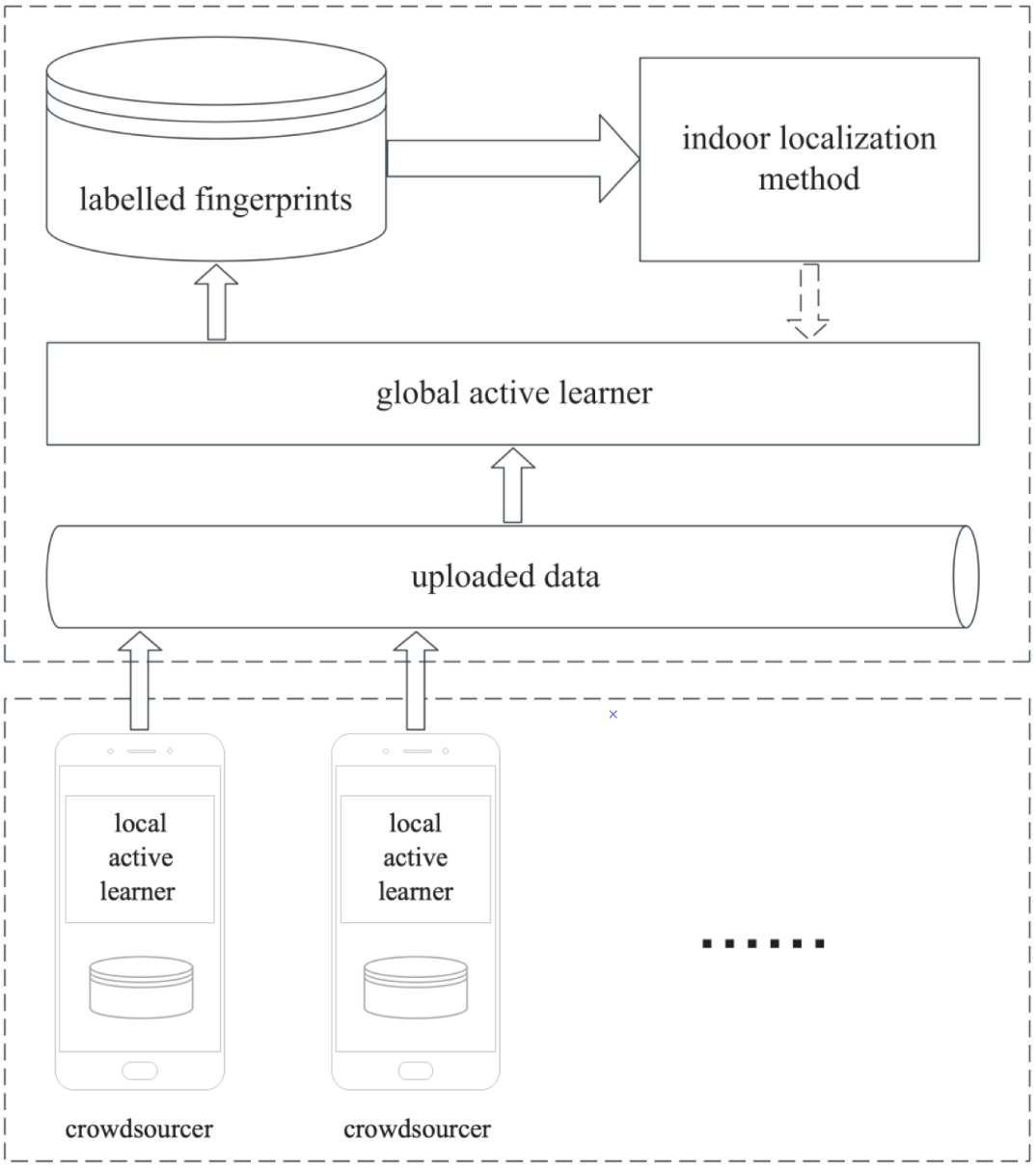


Figure 3-1: The architecture of ALCIL

thus, the performance of the initial indoor positioning model should achieve a certain level. As we mentioned above, there are many informativeness measures, which can be selected based on our target. The purpose of this system is to obtain a precise model that predicts the position stably according to the data we calibrated, therefore, we need to continuously reduce the variance of the model

derived. QBC is a model-driven selection approach, which involves maintaining a committee $\mathcal{C} = \{\theta^1, \theta^2, \ldots, \theta^c\}$ of models trained on the labeled dataset and select the most disagree one to constrain the hypothesis space, reducing the model variance. To detect the level of disagreement, various distance forms are utilized, among which KL divergence can measure the difference between different probability distributions [113, 89]. KL divergence is characterized by,

$$x^*_{KL} = argmax_x \frac{\sum_{c=1}^{C} D(P_{\theta^c}||P_c)}{C} \tag{3.2}$$

where

$$D(P_{\theta^c}||P_c) = \sum_i P_{\theta^c}(y_i|x) \log \frac{P_{\theta^c}(y_i|x)}{P_c(y_i|x)} \tag{3.3}$$

Here θ^c represents a particular model in the committee, and $\mathcal{C}$ represents the committee as a whole, hence $P_{\mathcal{C}}(y_i|x) = \frac{\sum_{c=1}^{C} P_{\theta^c}(y_i|x)}{C}$ is the agreement that y_i is the correct label. Thus, this disagreement measure considers the most informative instance to be the one with the largest average difference between the label distributions of any one committee member and the consensus. As the variant of KL divergence with some useful improvement, including that it is symmetric between two distributions, and it always has a finite value [146], JS divergence is also used to measure the disagreement. Therefore, QBC is appropriate as the informativeness measure in indoor localization systems to improve the prediction accuracy.

3.3.3 Optimization Strategies

As mentioned above, location can be expressed in different terms, including the relative address and absolute address, each of them has different attributes. Relative location is the description of how a place is related to other places like the floor number is the relative height to the ground, and the building number is the relative location from other buildings of a region, directly showing the connection

with others. Geographic coordinates of longitude and latitude help us pinpoint the absolute location which is a kind of relative address relative to the equator (latitude) and prime meridian (longitude) and help us understand each other with the fixed standard. Depending on the different purposes, the importance of various attributes is also different. For the significant application of measuring the distance between people, which is widely applied to detect the social distance under COVID-19, building number and floor number are more straightforward than a string of digits because we can directly obtain the conclusion we desired when we have no knowledge about geography.

However, accurate coordinates do help us calculate the precise information we need, like the social distance, and if we focus on certain attributes, we cannot achieve it. Thus, the importance of various attributes is different at diverse phases. The prediction accuracy of the building number and floor number is more important at first, as the project carried on, and when there is no significant improvement about it, we should focus on the coordinates' prediction. For indoor localization with multiple labels, we design two optimization strategies in different stages, namely the local optimization strategy and the global optimization strategy. The former focuses on the optimization of a subset of all features, and the other one optimizes all attributes at the same time. When the current strategy cannot have a significant impact on the location prediction, the accuracy reaches a critical point, and it is a waste to take more effort to continuously improve, and it is time to change the strategies or attributes we focused on.

3.4 Experiments

In this section, we evaluate the performance of ALCIL on a dataset we collected in the West Zone of Kyushu University. The dataset is randomly divided into training data and test data, and a small collection of training data is selected at

random.

3.4.1 Data Collection

To collect the necessary data for our experiments, we developed a data collection application on Android smartphone, integrated the Wi-Fi fingerprint sensing and data annotation via text or image, where the labels can be relation position or absolute position. It records the (Service Set Identifier (SSID), MAC address, RSSI) for each detected AP. Several participants with knowledge of localization contribute more than 4000 fingerprints with accurate labels of building ID, floor ID and geographic coordinate, using the same Android phone, LG Nexus 5, over different days. Among them, building ID and floor ID are categorical labels and latitude and longitude are quantitative values. The dataset was divided into training (70%), validation (10%) and testing (20%) sets, randomly, without overlapping, and then 200 samples from the training sets are randomly selected to achieve an initial model for bootstrapping the system.

3.4.2 Localization Methods and Evaluation Metrics

Various machine learning methods are widely researched in many different fields, among which random forest (RF) has been deeply studied as the critical feature of convenient behavior both in terms of accuracy and efficiency in prediction tasks [20]. RF is an ensemble learning approach that behaved on many independent decision trees, and the result is calculated as,

$$f(x) = \sum_{n=1}^{N} \frac{1}{N} f_n(x) \tag{3.4}$$

where f_n is the $n-th$ decision tree and N is the total number which is not always the same, depending on the specific tasks and dataset [137]. Whether it is a classification mission or regression analysis, RF has achieved tremendous success,

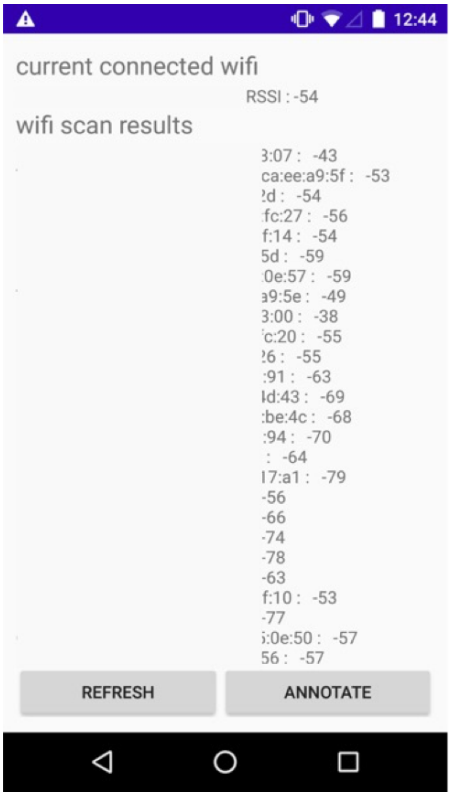


Figure 3-2: Data collection system

which is the main reason for being chosen because there are quantitative values of latitude and longitude and categorical labels of building and floor ID [48]. Some separate RF models are performed for the regression of continuous attributes and the classification of discrete labels, respectively, to return the different descriptions of location.

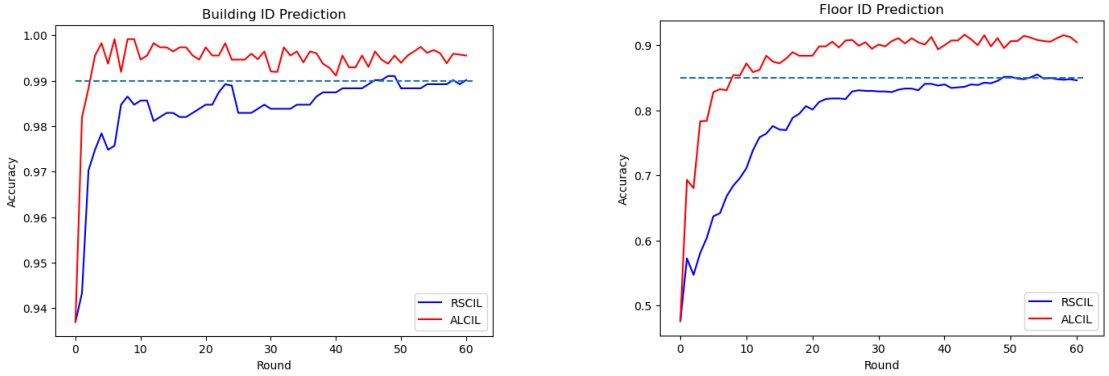
The prediction accuracy of two categorical location attributes (building ID and floor ID) is applied to evaluate the performance of the corresponding model. We utilize root-mean-square error to measure the regression results for latitude inference and longitude inference, which is defined as:

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i)^2} \tag{3.5}$$

where $\hat{y}_i$ represent the estimated value of $i-th$ sample and y_i is the corresponding true value.

3.4.3 Results and Discussion

In the experiments, we evaluate how the proposed system performs with the increase of annotation acquisition cost M. Because there are multiple labels of the location, to show its efficiency clearly, we can consider it as multiple independent problems, including two classification and two regression tasks using a collected dataset.

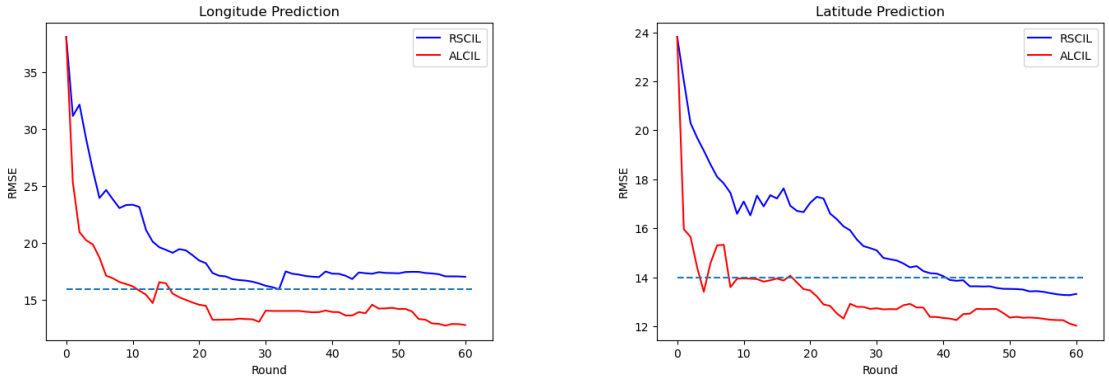


(a) building ID prediction (b) floor ID prediction

Figure 3-3: Accuracy of the proposed framework on building ID prediction and floor ID prediction. The dashed lines in these figures represent the communication rounds required to achieve the corresponding accuracy.

There are two methods involved in this experiment, the proposed ALCIL and RSCIL. RSCIL represents random selection-based crowdsourced indoor localization, in which the crowdsourcers would select the data to upload to the server at random and the server randomly selects the data to re-calibrate from the uploaded dataset. As the system runs, assuming we select λ samples from Λ the data set each time, the cost of recalibration for each data is fixed at η and then, each round, all the accurate annotation we obtained will spend $\eta\lambda$ which should lower than M. Here we set Λ to 50, λ to 10, and we perform 60 rounds to construct the complete

labeled dataset. It means that the number of uploaded samples should be greater than 50 to be used as the data pool of the server.



(a) longitude prediction (b) latitude prediction

Figure 3-4: RMSE of the proposed framework on longitude prediction and latitude prediction. The dashed lines in these figures represent the communication rounds required to achieve the corresponding prediction errors.

As we mentioned before, more queries usually imply better performance for supervised learning applications. However, except for the quantity, the quality of queries matters as well. Therefore, we consider the four attributes of location as four separate tasks, indicating that the system possesses a singular objective, rather than varying objectives as the experiment progresses. Consequently, to test the quality of query samples in different tasks, the selection of 60 rounds of data would focus on one of the four specific objectives: building prediction, floor prediction, longitude estimation, and latitude estimation. The results of building and floor predictions are presented in Figure 3-3. As can be observed, with the help of active learning, a higher accuracy performance is available at the same annotation cost, thus increasing efficiency in both building and floor predictions and realizing efficient localization. Furthermore, it can be discerned that the performance gain for floor predictions is superior to that of building prediction. This outcome can be attributed to the relatively high precision acquired by the initial model in building ID prediction, resulting in a comparatively smaller actual gain. The evaluation

result of longitude prediction and latitude prediction is presented in Figure 3-4. The continuously decreasing errors of active learning based models not only suggest that we approach the accurate location, but also indicate we are converging towards it at a relatively faster pace.

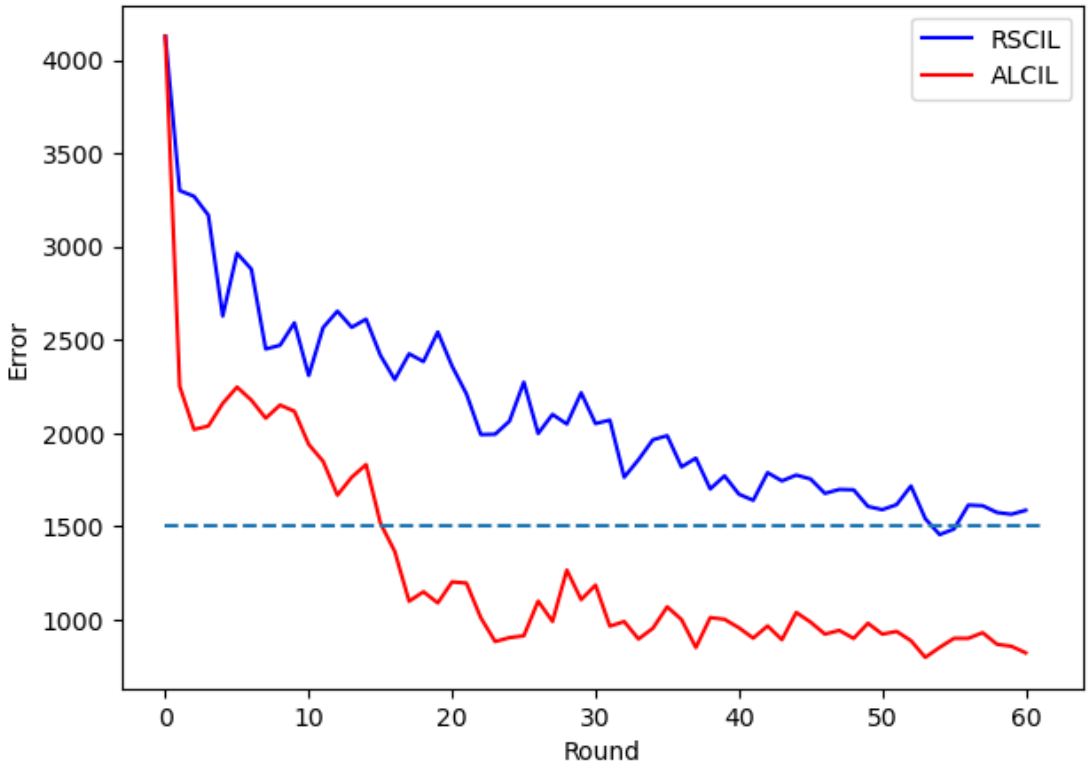


Figure 3-5: The holistic evaluation of the proposed framework. The dashed line in the figure represents the communication rounds required to achieve the corresponding error.

We assemble the prediction results of the four models and try to evaluate the system using a holistic outcome. Regarding the classification tasks, we count the number of incorrect predictions made by the classification models and generate results with corresponding penalty factors to simulate regression tasks. Therefore, the comprehensive metric is proposed as:

$$F = EN_{building} * PF_{building} + EN_{floor} * PF_{floor} + RMSE_{latitude} + RMSE_{longitude} \quad (3.6)$$

where $EN_{building}$ and EN_{floor} are the number of inaccurate building IDs and floor IDs, respectively. Considering that incorrectly predicted building ID and floor ID can have a significant impact on final location estimation results, we assign penalty coefficients to 50 and 4. We perform a comprehensive evaluation of the system based on the new criterion, and the results are illustrated in Figure 3-5. As observed from Figure 3-5, our proposed ALCIL is capable of achieving better performance compared to RSCIL in general. In comparison to the gradual improvement of RSCIL, the distinct characteristics of ALCIL can be clearly observed. Although the model's performance would improve with the increase in labeled samples, the impact of the four different subtasks on the global results varies. It is evident that in the first two rounds, the system's performance significantly improved, followed by a period of steady decline and then experienced a substantial drop afterward. This phenomenon occurs because the accuracy of the initial building prediction model dramatically increases, and subsequently, as the overall performance begins to stabilize, the optimization turns towards refining the floor prediction model. After the acquired performance gain of the floor prediction model in several rounds, the overall performance of the system gradually improves as the number of labeled samples increases. Consequently, our proposed system would swiftly attain a better result and higher efficiency with lower labeling costs compared to RSCIL.

3.5 Conclusion

We proposed a cost-effective and quality-ensured framework for crowdsourced indoor localization, which collects informative data to further improve the performance of the model at a certain cost. In ALCIL, the researcher needs to obtain a collection of Wi-Fi fingerprints with precise labels as the initial data to train a primary model, which can be comprehensive when some users are motivated by

incentive measures to contribute more samples in actual deployment. While considering the participants' ability, the annotation can be the relative location. Then the informativeness of the sample is detected based on QBC to determine whether this sample uploads to the server. After the server obtains a certain number of samples, considering the multiple attributes of location, the data that are valuable for the current target will be selected to recalibrate. Avoiding the annotation of all data can reduce the cost of obtaining the complete dataset. The effectiveness of this framework is demonstrated by the collected data. Possible extensions to this framework include the integration of an appropriate incentive mechanism to motivate more users to collect Wi-Fi fingerprints within the target area, and mechanisms for considering the device heterogeneity.

Chapter 4

Mapless Indoor Navigation based on Landmarks

4.1 Introduction

Navigation problems about how to reach the destination from the origin is often encountered in human daily life and work, especially in unfamiliar area. Depending on the target environment, it can be divided into outdoor navigation and indoor navigation. GNSS, such as GPS, Galileo, BDS and other satellite systems can navigate us precisely and reliably in an outdoor open environment, which are widely exploited and broadly applied in our everyday lives [64]. Although outdoor navigation is well satisfied, indoor navigation is more challenging. Since GNSS does not perform well in urban canyons, underground environments, and indoor environments in which we spend most of the time and tend to lose position more easily because of the lack of a unified infrastructure and the weak signal strength of satellites due to the absence of line of sight, the attenuation of satellite signals as they cross through physical objects, especially walls, and noise interference, resulting in inaccurate navigation of us or devices [27, 17]. Furthermore, indoor navigation has played an important role in our routine, especially in a large building like a

shopping mall or underground parking lot. Thus, in an effort to find an alternative technique that can provide indoor navigation service with high precision, a variety of approaches have been extensively researched in decades.

Many indoor navigation systems have been researched with the joint effort of researcher and engineer in the past few decades. For different scenarios, researchers investigate lots of technologies and techniques to build indoor navigation systems: RFID, Bluetooth, Zigbee, UWB, WLAN, IR and ultrasound. Among these radio frequency-based methods, Wi-Fi is the most commonly used and extensively applied for indoor navigation systems using received signal strength and channel state information as the Wi-Fi settings are considered broadly distributed. While the Wi-Fi signals would not be accessed when the user is out of Wi-Fi coverage range, which are common in developing country due to poor ICT infrastructure [63], or the radio coverage has to be blocked sometimes. Therefore, the pre-deployment and special hardware hinder the wide adoption and practical application.

Meanwhile, mobile devices of all kinds are rapidly involving, and our daily life is significantly changed. The range of application, including indoor navigation, are efficiently growing with more and more advanced built-in sensors equipped. With the help of portable devices people usually carry around, PDR based on built-in inertial sensors, which is self-contained, has been obtaining growing attention in different scenarios [80]. Nevertheless, PDR is prone to suffer from the accumulated error because of sensor drift, resulting in accuracy degradation. The combination of PDR and Wi-Fi fingerprinting which is extensively adopted, Bluetooth, and some other localization methods to calibrate the trajectory are common remedies. In addition, a recent trend is to remove the requirements of infrastructure support.

Spatial context, such as maps and landmarks, which is available in many scenarios, can be another choice to calibrate the indoor navigation system based on PDR without additional deployment cost. Fusing spatial information which is easy to understand for successful wayfinding is an effective way to achieve indoor navi-

gation with little or mostly with no need for complementary infrastructure [36]. As a piece of important spatial information, the landmark is a salient point in sensor readings when people pass the location, which can be detected by built-in sensors. There are many different proposed methods that integrated landmark in indoor navigation and positioning systems, in which landmark detection and matching are both involved to realize indoor navigation in works [36, 141, 47]. Currently, the manual designed features and threshold are extracted to detect different landmarks [44]. When numerous identified landmarks are in proximity, the adoption of additional Wi-Fi fingerprints is broadly implemented for recognition [167].

In this paper, to reduce additional infrastructure demand in indoor navigation systems, we propose *mapless indoor navigation system based on landmark identification*. The motivation of this work is to provide navigation service to people with a typical smartphone equipped coming to a new area and without recourse to any additional deployment and spatial information in advance. The proposed methods can be divided into two phases: offline training and online navigating. In the former phase, the records produced by the embedded sensors of mobile devices are collected to automatically extract features and train a landmark recognition model after being preprocessed, including filtering and segmenting. The semantic map is subsequently constructed with the help of detected landmarks and PDR. In the later phase, online navigation is achieved based on the location estimated by PDR and detected landmarks with sensor reading stream. The major contribution of this work consists of two aspects: the first aspect involves a novel extra infrastructure-free landmark-based indoor navigation systems without radio coverage to characterize landmarks without a detailed floor plan ahead, which reduces the deployment without extra cost requirements and the cognitive load of people with a human-friendly navigation experience compared with physical coordinates to locate and navigate users. For the second part, a custom deep learning model is designed and implemented for landmark identification with the ability to auto-

matically extract features through the raw signals to create an indoor landmark map.

The remainder of this chapter is organized as follows. Related work about indoor navigation, especially the landmark-based navigation, is reviewed in Section 4.2. Section 4.3 introduce the theoretical methodology and the architecture of proposed mapless landmark-based indoor navigation system. Section 4.4 presents the experimental methodology and results in datasets collected. Finally, conclusion and future work are presented in Section 4.5.

4.2 Related Work

There are three different modes to display the guidance in a navigation system: geographical coordinates, symbolic modes, and hybrid information [153]. Geographical coordinate is a detailed representation of location, a machine-friendly way, which is widely applied to navigation system in many professional devices. For indoor navigation, since the GNSS signal reception is degraded accuracy of the coordinates will be affected, therefore, representation via the earth coordinate system will be changed to a relative coordinate system, which is relative to a pre-defined location in the indoor environment. Symbolic modes denote the environment with the logical relationship of different areas or locations, which can generally navigate people both indoors and outdoors in a human-friendly way. These two representations are combined in the hybrid mode navigation systems.

This section discusses a brief overview of the relevant literature, which can be divided into indoor navigation based on geographical coordinates and landmark-aided navigation.

4.2.1 Indoor Navigation

With the help of indoor localization technologies and techniques promoted by the rapid development of wireless technology and pervasive computing, the theoretical research and practical application of indoor navigation systems has been efficiently broadened. Multiple deterministic models are designed in combination of physical laws and mathematical, such as: AoA, ToA, and TDoA [153]. Among the various methods, acquiring the distance traveled through the Log-Distance Path Loss (LDPL) model from the RSS of sensed environmental radio signals constitutes a significant approach for location estimation within buildings [114]. Since the accuracy of the statistical models is greatly affected by noise and variation, the probabilistic methods are investigated [33]. Additionally, a plethora of indoor positioning methods adopt the RSS as fingerprint to realize localization has been proposed. The core idea is to build up a fine-grained radio map consisting of the fingerprint of each interested location, which the position can be achieved using matching algorithms. RADAR is the first attempt to apply the fingerprinting-based technique with k-nearest neighbors (KNN) matching algorithm in indoor localization, which utilizes the Wi-Fi fingerprint [9]. Wi-Fi fingerprinting-based indoor localization can measure the RSSs from detected access points (APs) at target area to construct the detailed radio map on training phase, which will be used to determine the location via deterministic and probabilistic algorithms with different similarity metrics like Euclidean distance, Kullback-Leibler divergence and Jensen-Shannon divergence [94, 2, 93, 3]. Therefore, the quality and quantity of radio map, which depend on the number of wireless devices installed and annotation accuracy, makes a significant impact on the performance of Wi-Fi fingerprinting based positioning systems [83].

Simultaneous Localization and Mapping (SLAM) is a well-known technique that dominates the robotic field, which usually rely on the laser range sensor or cameras to build the map based on landmarks and simultaneously infer robot

location in an unknown area. SmartSLAM, as a modified algorithm based on smartphone, is proposed to gradually construct the indoor floor plan for anonymous buildings, employing inertial sensors to track users and using Wi-Fi signals as an indicator to find anchor point [121]. SematicSLAM uses the estimated landmark in the environment as reference point and combine the inertial sensors as an odometer to keep track of users [1]. Wi-Fi related fingerprint is also used in WalkSLAM and Wi-Fi-RTT-SLAM [86, 41].

DR, as a radio signal-free localization method based on inertial sensors such as accelerometers, gyroscope, magnetometer, and barometer, are presented in many works. Given a starting point, locations can be continuously calculated by combing last position and the displacement inferred by the motion information which is provided by three parts: step detection, estimated strides length and heading changes. Because of no additional requirements and no coverage limitation of DR, it is popular in wireless blocked or denied area and emergencies [22, 117]. While, due to environmental contamination, vibration and temperature fluctuations, inevitable sensor drift causes the positioning error accumulated along movement, resulting in degraded localization performance. Wi-Fi fingerprint is commonly integrated to enhance the accuracy. Predefined Wi-Fi fingerprint as the reference point to When the Wi-Fi fingerprint whose corresponding location is known as the reference point is passed, the accumulated error can be eliminated by the absolute landmark position. Besides, with the ubiquitous mobile sensor rich device, DR is gradually extended to PDR.

In addition to PDR, SLAM and Wi-Fi based localization approaches, plenty of research based on light Intensity, RFID, Bluetooth, Zigbee and other technologies realize the indoor positioning. However, extra infrastructure deployment or specific devices are required in the sensor-based approaches, even for the self-contained inertial sensor-based PDR methods, in which the absolute position of anchor point is needed to calibrate the positioning error.

4.2.2 Landmark-based Navigation

Landmark is defined as a spatial point with salient features and semantic characteristics from its near environment in indoor navigation systems, which can be used to calibrate the localization error based on the inherent spatial information. Fuqiang presents the concept of sensory landmark with distinctiveness to distinguish, stability patterns to detect and identifiability to identify, categorizing the different landmarks based on the type of built-in sensor within the smartphone to assist indoor localization [45]. The location of these landmarks, presented by geographical coordinates or the relationship with other locations/areas, where people perform specific and predictable movements can be detected and correspondingly identifiable change displayed on the changes of the readings of at least one type of sensor, as an anchor point to correct the position we calculated.

To identify the landmark, plenty of features are manually calculated, and the special thresholds of different sensors within various kinds of landmark recognition are analyzed. For instance, the threshold of angular velocity produced by a gyroscope is usually used to detect the corner landmark, and the acceleration changes can recognize the stairs. The combination of different thresholds of various sensors forms the decision tree, which can reveal the standing motion state to further distinguish common landmarks, such as corners, stairs, and elevators [87]. However, the calculation, extraction, and selection of features of different sensors for various landmarks are heuristic with professional knowledge of the domain and time-consuming, and laborious. To simplify feature engineering and improve performance, deep neural networks are applied in a variety of works. A convolutional neural network (CNN) is trained on the one-dimensional sensor data to learn the proper features automatically and landmark identification in [168]. The long short-term memory (LSTM) based deep RNNs (DRNNs) to classify the location mapped from variable-length input sequences of sensor data for landmark classification [13]. Wang Y, et.al. improve the LSTM neural network to recognize

different kinds of spatial structure-related sensory landmarks [145].

To navigate people based on the detected landmarks, in addition to the location, how to organize the landmarks precisely and effectively is also necessary. Multiple landmarks that the logical relationships are well displayed in an easy-understanding construction of an indoor floor plan for indoor navigation. Although almost all buildings provide the indoor map at some conspicuous locations, it is still not easy for people, especially the spatial cognitive disability, children, to understand the map and reach the destination. ALIMC estimates the relative distance of all the landmarks detected and then generates the indoor floor plan using multidimensional scaling algorithms [167]. IndoorWaze calculates the movement trajectory with the help of PDR between different landmarks to construct the landmark graph to guide people, and the same fashion of the detected landmarks is also presented in many works [47, 75, 55]

4.3 Methods

In this section, the proposed indoor navigation system is presented, which only needs a smartphone, precisely the sensors in it. Both hardware sensors and virtual sensors are in a smart device. Hardware-based sensors derive their data by directly measuring specific environmental properties and physical attributes, such as the barometer, accelerometer, and gyroscope. Readings of virtual sensors are calculated by one or more hardware sensors, such as the gravity sensor, linear acceleration, and rotation sensor. Here, the accelerometer, gyroscope, barometer, and rotation sensor.

4.3.1 Architecture

Figure 4-1 shows the architecture of proposed indoor navigation systems, which can be divided into two phases: offline training and online indoor navigation. In

the offline phase, various sensor data flows for analyzing are firstly collected from a handheld smartphone's built-in sensors which record the changes of environment and body motion. The collected signals need to be preprocessed, which includes data filtering and data scaling, to reduce the noise for better motion state estimation. Next, with the help of PDR techniques, the locations and the trajectories are estimated to build the indoor landmark map, which is to further assist the navigation on online phase. Besides, the preprocessed data are also used to extract the features for landmark identification model training to aid online navigation. The same data collection module and PDR algorithms are utilized in online phase. On online data processing, for timely navigation, the fixed sliding window with a degree of overlap is applied to generate the same inputs as the model trained. Therefore, the landmark identification model can distinguish the landmark and calibrate the location due to sensor drift and cumulative error simultaneously. What's more, in combination of the indoor semantic map constructed by landmarks, users are navigated with accessible instructions based on their connectivity relationships.

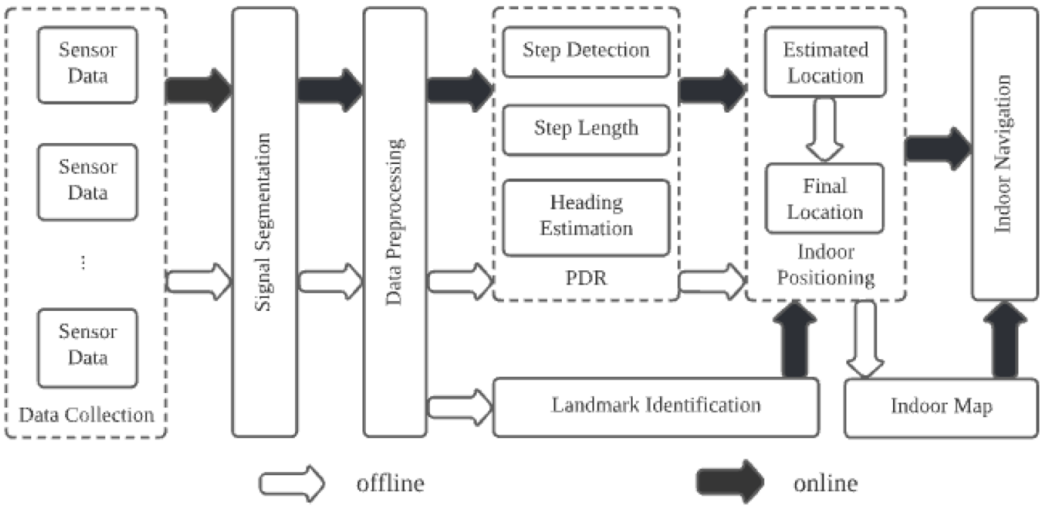


Figure 4-1: Architecture of proposed approach

4.3.2 Data Preprocessing

To combine the data from different sensors for landmark identification, the specific processing process is described below, including data alignment, data interpolation, data filtering, data scaling and data segmentation.

Data alignment The same sampling rate for data collection is set to 50 Hz due to the low frequency of human movements [131]. Although the constant rate is defined, the time interval between the recorded adjacent readings of each sensor is not always the same because of the observational error and random error, and it oscillates within a certain range in practice. To acquire the same number of samples for conveniently performing the subsequent procedures, we take the timestamp of the first data collected as the starting time to align the sensor readings at the same time interval with the help of data interpolation.

Data interpolation During the practical data collection using smartphone sensors, some data points in the acquired dataset are lost due to malfunctioning; such data points are typically replaced by 0, NaN, or none [59]. To fill in the missing values, the data interpolation technique was developed, in which the new data point is estimated based on the known information. Linear interpolation, as the prevalent type of interpolation approach, was adopted in this paper, using linear polynomials to construct new data points [152]. Generally, the strategy for linear interpolation is to use a straight line to connect the known data points on either side of the unknown point and, thus, it is defined as the concatenation of linear interpolation between each pair of data points on a set of samples.

Data filtering Due to the environmental noise and interference caused by the unconscious jittering of the human body, there are many undesirable components in the obtained signals that need to be dealt with [152]. This usually means removing some frequencies to suppress interfering signals and reduce the back-

ground noise. A low-pass filter is a type of electronic filter that attempts to pass low-frequency signals through the filter unchanged while reducing the amplitude of signals with a frequency above what is known as the cutoff frequency. A Butterworth low-pass filter with a cutoff frequency of 3 Hz is applied to denoise and smooth the raw signals.

Data scaling The difference in the scale of each input variable increases the difficulty of the problem being modeled. If one of the features has a broad range of values, the objective functions of THE established model will be highly probably governed by the particular feature without normalization, suffering from poor performance during learning and sensitivity to input values and further resulting in a higher generalization error [29]. Therefore, the range of all data should be normalized so that each feature contributes approximately proportionately to the final result. Standardization makes the values of each feature in the data have zero means by subtracting from the mean in the numerator and unit variance, as shown in Equation 4.1:

$$X_i^{'} = \frac{X_i - \mu}{\sigma} \ (i = 1,\ 2,\ 3 \cdots,\ n) \tag{4.1}$$

where the $X_i^{'}$ is the standardized data, n represents the number of data channels, and μ and σ are the mean and standard deviations of the i-th channel of the samples [29]. This method is widely used for normalization in many machine learning algorithms and is also adopted in this work to normalize the range of data we obtained.

Data segmentation A sensor-based landmark recognition model is typically fed with a short sequence of continuously recorded sensor readings, since only a single data point cannot reflect the characteristics of landmarks. The sequence consists of all the channels of selected sensors. To preserve the temporal relationship between

the acquired data points with the aligned times, we partition the multivariate time-series sensor signals into sequences or segments leveraging the sliding operation, which consists of 128 samples (corresponding to 2.56 s for the sampling frequency at 50 Hz) [152, 46]. It is noteworthy that the length of the window is picked empirically to achieve the segments for all considered landmarks, in which the features of the landmarks can be precisely captured to promote the landmark identification model training [131, 162].

4.3.3 PDR

Generally, PDR consists of three main components: step detection, stride length estimation and Heading estimation.

Step detection As the most popular method for accurate step detection, peak detection is employed in this work, which relies on the repeating fluctuation patterns during human movement. Using the smartphone's accelerometer to determine whether the pedestrian is stationary, or walking is straightforward as it directly reflects the moving acceleration. The magnitude of acceleration in three dimensions (a_x, a_y, a_z) instead of the vertical part is employed as the input for peak findings to improve the accuracy, which can be expressed as:

$$a = \sqrt{a_x^2 + a_y^2 + a_z^2} \tag{4.2}$$

where a_x, a_y, a_z denotes the three-axis accelerometer values in the smartphone [157]. A peak is detected when a is greater than the given threshold. To further enhance the performance, the low-pass filter is further applied to the magnitude to reduce the signal noise. Due to the acceleration jitter, the incumbently detected peak points need to be eliminated. Hence, an adaptive threshold technique of the maximum and minimum acceleration is adopted to fit different motion states with a time interval limitation between adjacent detected steps.

Stride length estimation Various linear and nonlinear methods are proposed to estimate the step length, which varies from person to person because of different walking postures determined by various factors, including height, weight, and step frequency. Therefore, it is not easy to precisely construct the same step-length estimation model. Some researchers assume that the step length is a static value affected by the individual characteristics of different users. On the contrary, the empirical Weinberg model estimates the stride length according to the dynamic movement state, which is closer to reality [148]. The model is given by:

$$SL = k\sqrt[4]{a_{\max} - a_{\min}} \tag{4.3}$$

where k is the dynamic value concerned with the acceleration of each step and $a_{\max}$, $a_{\min}$ are the maximum and minimum accelerations for each step [28].

Heading estimation Heading information is a critical component for the entire PDR implementation, which seriously affects localization accuracy. To avoid the cumulative error in the direction estimation based on the gyroscope, and short-term direction disturbances based on the magnetometer, the combination of the gyroscope and magnetometer is typically adopted for heading estimation [157]. The current magnetometer heading signals, current gyroscope readings, and previously fused headings are weight-averaged to form the fused heading. The weighting factor is adaptive and is based on the magnetometer's stability as well as the correlation between the magnetometer and the gyroscope [28]. As they are already fused in the rotation vector achieved from the rotation sensor in the smartphone, the heading change can be calculated by a rotation matrix transformed from the rotation vector [158]. The rotation vector is defined as: $[x, y, z, w]$, and the matrix is defined as $M, M \in R^{3\times3}$. The heading direction on three dimensions can be

evaluated by:

$$M = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} = \begin{bmatrix} 1-2y^2-2z^2 & 2xy-2zw & 2xz+2yw \\ 2xy+2zw & 1-2x^2-2z^2 & 2yz-2xw \\ 2xz-2yw & 2yz+2xw & 1-2x^2-2y^2 \end{bmatrix} \tag{4.4}$$

$$\theta = \begin{bmatrix} arctan2(M_{12},\ M_{22}) \\ arcsin\,(-M_{32}) \\ arctan2(-M_{31}, M_{33}) \end{bmatrix} = \begin{bmatrix} arctan2(2xy-2zw,\ 1-2x^2-2z^2) \\ arcsin(-2yz-2xw) \\ arctan2(2yw-2xz,\ 1-2x^2-2y^2) \end{bmatrix} \tag{4.5}$$

4.3.4 Landmark Identification

Autoencoder is a type of unsupervised neural networks that can be used to learn feature representation of data. It learns the feature representation by training the network to reconstruct the data at the output layer. Each autoencoder consists of three parts: encoder, compressed features and output, whereby the compressed features extracted by encoder are sent to the decoder part to reconstruct the input. Many variants are proposed to solve different problems, such as the sparse autoencoder, variational autoencoder and denoising autoencoder [139]. Denoising autoencoder is a useful variant of vanilla autoencoder to learn a robust feature representation by introducing stochastic noise to the input data and the critical part will be reconstructed from the corrupted data [53, 40].

Usually, the denoising autoencoder are built by fully connected layers. Since the sensor data flow has a strong 1D structure that the previous state and the next state connect tightly. 1D convolution operation can efficiently capture the local correlation features by limiting the hidden units' receptive field to be local. CNN considers each frame of sensor data as independent and extracts the feature for these isolated portions of data without considering the temporal context beyond the boundaries of the frame. Due to the continuity of sensor data flow produced by user behavior, local correlation and long-term connections are both important to

identify the landmark [162]. While LSTM with learnable gates, which modulate the flow of information and control when to forget previous hidden states, as a variant of vanilla Recurrent Neural Networks (RNN), allows a neural network to effectively extract the long-range dependencies of time-series sensor data.

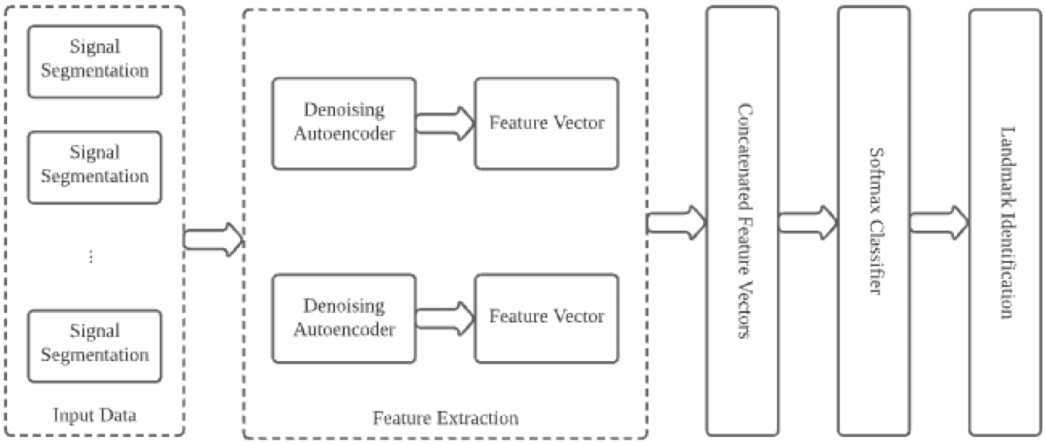


Figure 4-2: The block diagram of landmark identification

The structure of the landmark identification module is shown in Figure 4-2. When preprocessed data segmentation of multiple sensors comes, features in different domains are automatically extracted by unsupervised featuring learning pipelines based on denoising autoencoders implemented by alternatively stacked 1D convolutional layer and 1D max-pooling layer and multiple LSTM layers, respectively. Denoising autoencoder extracts the robust feature, both short-time correlation and long-term dependencies from the sensor data corrupted by environmental noise. Finally, features extracted by encoder parts will be concatenated to train a classification model used in online predicting phase to identify landmarks to improve the localization accuracy and the performance of navigation based on the landmark map. The indoor landmark map can be generated based on the relative distance of different landmarks calculated by PDR, with the help of multidimensional scaling techniques which are commonly used to find the spatial relationship based on dissimilarity information, such as the relative distance in

this work, between objects.

4.4 Performance Evaluation

In this section, we evaluate the performance of proposed methods on the dataset we collected on the fourth floor of the Building 1, at the Center Zone of Kyushu University's Ito campus.

4.4.1 Data Collection

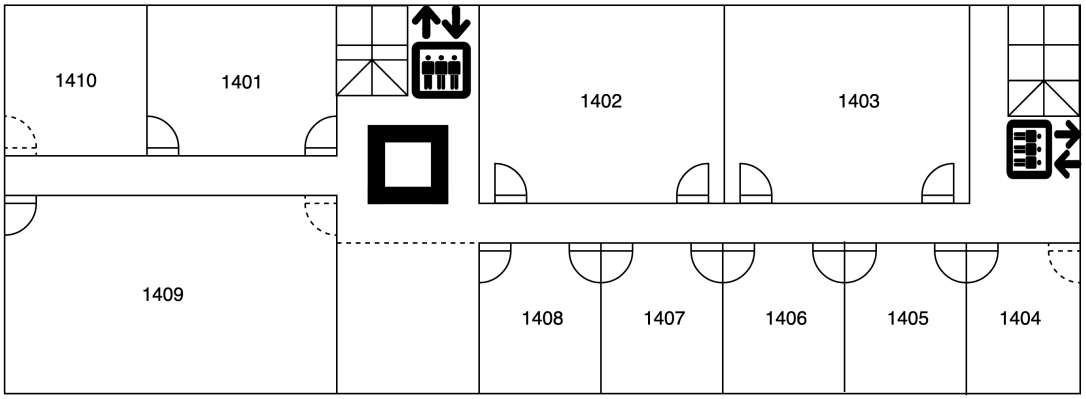


Figure 4-3: The floor plan in our experiment

For recording sensor data flow, we used the built-in sensors of a smartphone (Pixel 4a) and developed an Android application that periodically read and stored the readings. To collect the sensor dataset when the landmark passed and build the landmark map, five participants were invited to collect the data. During each collection, participants were required to keep their phones in hand and walk with them held to their chest level. They were also required to record the timestamp of passing by the landmark and the identification of the landmark. Landmark information they passed are the type of scenario, the type of landmark, and the moment of passing by the landmark.

After data preprocessing, some statics with a brief description of the clean dataset are shown in Table 4.1. There are 9,271 samples in total for all 27 landmarks, including almost all (toilet are not included, only one door selected of room 1401 and 1409, and no landmark for room 1410) the corners, stairs, elevators and two doors, which is assumed to be open during our experiment of each room of the corridor, shown on Figure 4-3.

Table 4.1: Landmark identification performances of different models in the collected dataset

Users	**S. Rate**	**# Landmark**	**# Samples**	**Sensors**
5	50Hz	27	9271	A,G,R,B

4.4.2 Hyperparameter Settings

The landmark identification model is trained on the collected data with different hyperparameters, which are listed in Table 4.2. Keras framework with TensorFlow backend is used for the implementation of the classifier to minimize the cross-entropy loss. Both LSTM and 1D CNN share the same parameters, if not specifically summarized. The number of hidden layers represents the number of encoder layers, the same number of layers on decoder parts.

4.4.3 Results and Discussion

Visual results Figure 4-6 depicts the landmark graph generated by our system based on all the data we collected. The map presented in both Figure 4-5 and Figure 4-4 are very similar to the ground truth in Figure 4-3. The difference between these two landmark maps is whether the door landmark was selected or not. Doors are usually not selected as a landmark in existing systems, while the building where we conduct our experiments is a teaching building, and many people

Table 4.2: Training hyperparameters

Hyperparameters	**Value**
Optimizer	Adam
Batch size	64
Dropout rate	0.5
Filter size	5
Learning rate	0.002
Input vector size	128
Input channels	12
Number of epochs	800
Number of hidden layers	2
Units of hidden layers	64

come to study or take classes in the classroom. Therefore, we spend triple time to obtain the landmark map with door in, as a frequently visited location, we still choose the doors of the classroom as an important kind of landmark, which makes the map more comprehensive and guides people to the greatest extent possible.

Landmark Identification. The landmark identifications are performed using the collected data, which were divided into training set and test set randomly without overlap between these two datasets. There are a total of 27 landmarks, and most of them are similar, which can be divided into four categories in total, including half-occupied doors, corners, stairs, and elevators.

We evaluated the classification performance of proposed approaches for four

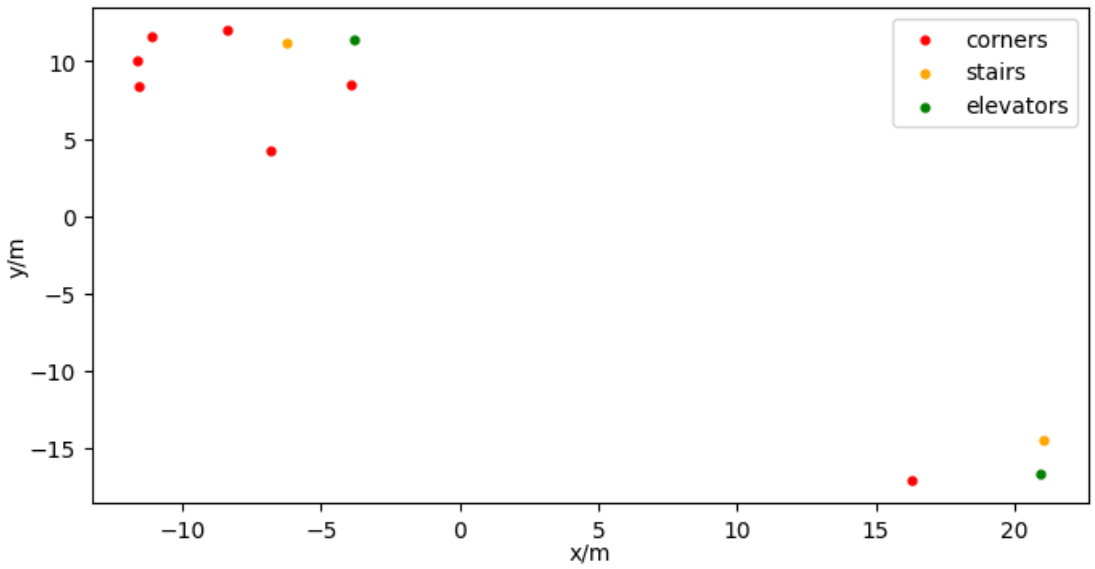


Figure 4-4: The landmark-based indoor map without door

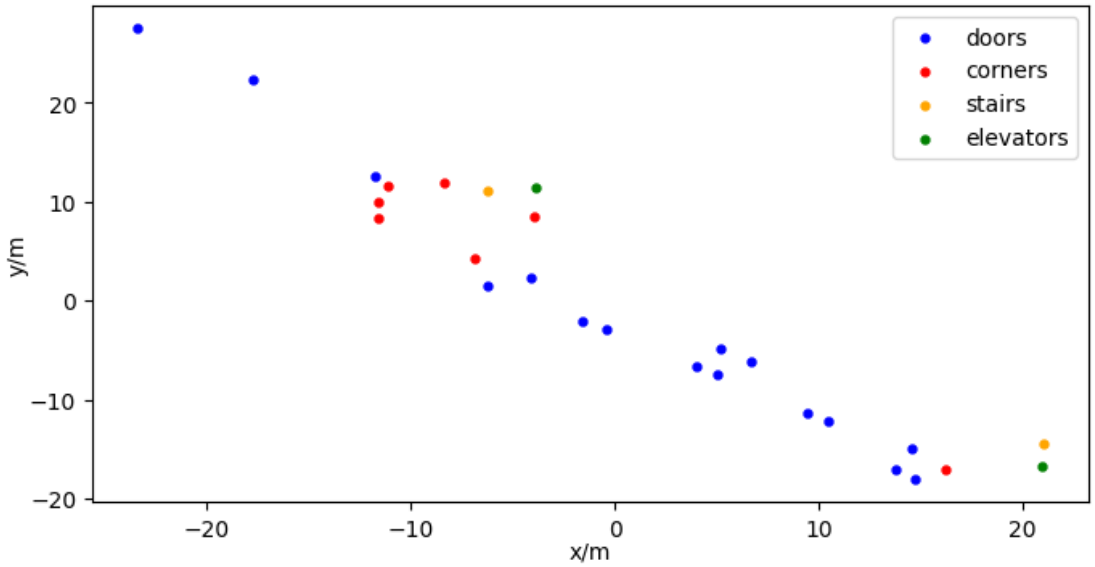


Figure 4-5: The landmark-based indoor map with door

kinds of landmark. The recognition result is presented by the confusion matrix in Table 4.3. The confusion between the individual labels is small. Many other landmarks are misclassified as doors, which mainly because people have to make the same movement when entering and exiting the classroom as they do around

Table 4.3: Confusion matrix for the kinds of landmark classification

Prediction/Truth	**Doors**	**Corners**	**Elevators**	**Stairs**
Doors	1578	10	1	0
Corners	20	652	2	0
Elevators	13	18	280	0
Stairs	2	2	4	140

corners, turning a direction and continuing to walk. The same situation occurs in elevator detection, mainly when entering and exiting elevators. Figure 4-6 is the classification confusion matrix on 27 landmarks of the experiment conducted floor.

Indoor navigation We test the navigation service based on the generated landmark graph. In this experiment, four different routes are randomly selected. We reached each destination by following the landmark based indoor relative map with different wrong steps. Most of the wrong steps are at the beginning to identify the landmark, in which people have to move more than three steps to start the landmark identification model and plenty of space near the landmark. Additionally, if it is the incorrect landmark detected, users will be misguided, which happens in the fourth trail. There are three different directions near Room 1401 for users to start the walking, and the wrong direction is chosen without landmark identification at the first several steps. While it can be corrected during movement in a large area. Nevertheless, the results show that the proposed approach is accurate as an infrastructure-free indoor navigation system.

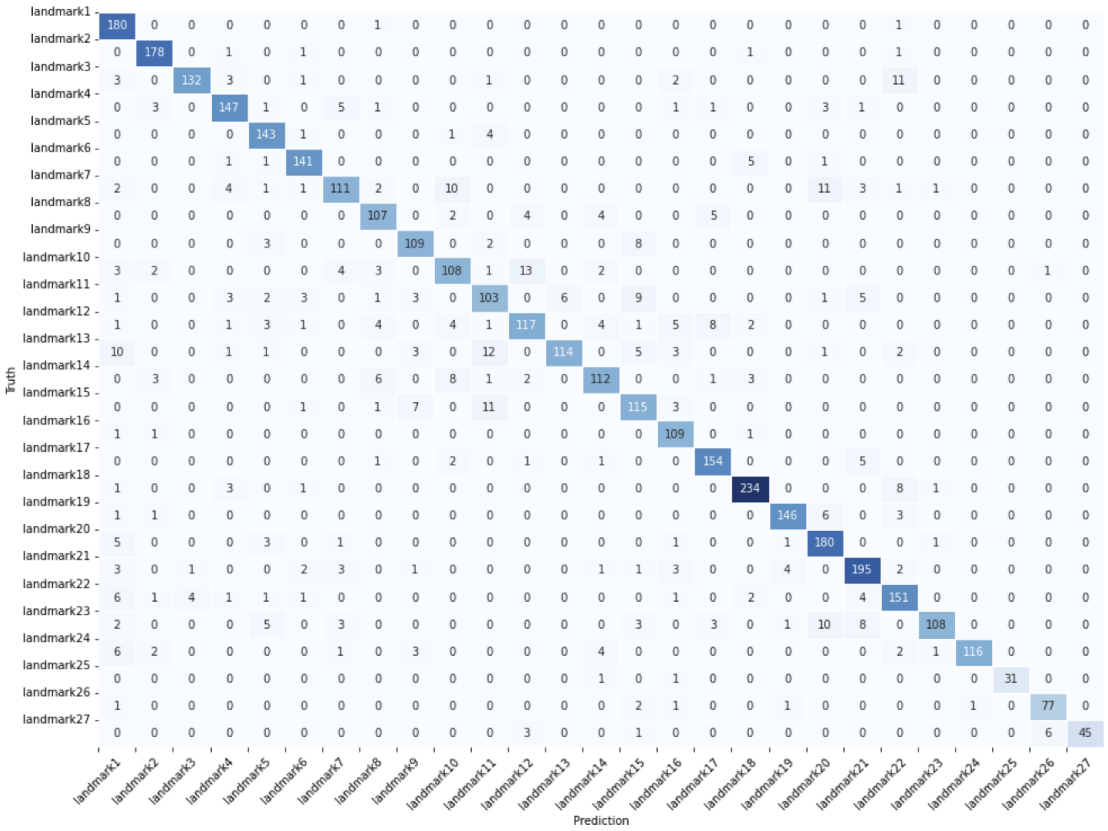


Figure 4-6: Confusion matrix for landmark identification

Table 4.4: Navigation result

	Start Point	**End Point**	**Total steps**	**Wrong Steps**
#1	Front door of Room 1404	Front Door of Room 1401	77	2
#2	Elevator near Room 1404	Front door of Room 1409	96	3
#3	Stairs near Room 1401	Stairs near Room 1403	69	5
#4	Corner near Room 1401	Front door of Room 1404	41	12

4.5 Conclusion

In this paper, based on the sensor data flow from smartphone built-in sensors, including an accelerometer, a gyroscope, a barometer, and a virtual rotation sensor fused by inertial sensors, indoor navigation with the help of spatial landmark identification and PDR to generate the indoor map is achieved. What's more, the landmarks, such as stairs, elevators, corners, and doors are distinguished by denoising autoencoder for automatically extracting the features without the requirement of extra infrastructure to significantly eliminate the laborious manual feature design. The effectiveness of this framework is demonstrated by extensive experiments based on the collected data.

Chapter 5

Indoor Spatiotemporal Contact Analytics Using Landmark-Aided Pedestrian Dead Reckoning on Smartphones

5.1 Introduction

The worldwide COVID-19 pandemic has brought about many changes in our daily lives and struck a devastating blow to the global economy. It is widely recognized that airborne transmission serves as the primary pathway for the spread of COVID-19 via expiratory droplets, especially in indoor environments [115]. During the viral outbreak, many people were infected due to exposure to virus droplets generated by human ex-halation activities [88, 140, 77]. Some infected patients spread the virus unknowingly without properly being examined because there is an incubation period that varies for different mutations and asymptomatic patients who never experience apparent symptoms [38]. Reliable and efficient tracing and quarantining have become more important than ever to alert individuals to take

actions to interrupt the transmission between people and further curb the spread of the disease. Contact tracing involves identifying, assessing, and managing people who are at risk of the infection, and tracking subsequent victims as recorded by the public health department [7]; contact tracing can be performed via manual or digital methods. Since the manual contact tracing is labor-intensive and time-consuming and may be incomplete and inaccurate due to forgetfulness; automatic digital contact tracing has been widely researched in recent years [97]. Usually, digital contact tracing applications are installed on portal devices, typically smartphones, to conveniently and intelligently realize tracing with the help of existing sensors based on various technologies, such as a GNSS, Bluetooth, and Wi-Fi.

Contact tracing in indoor environments can complement the ones used in outdoor environments to enable comprehensive digital contact tracing. However, indoor contact tracing imposes unique technical challenges due to virus concentrations and unreliable GNSS signals in indoor environments [72]. The virus concentration, which plays a critical role in calculating the amount of a virus we are exposed to and further assesses the infection risk, should be explicitly considered in indoor contact tracing applications [140, 35]. The quantitative infection risk for a susceptible person is significantly associated with the quantity of the pathogen inhaled in the surrounding ambient air, from the respiratory droplets exhaled by infected individuals [11]. Thus, inhaling a large amount of the virus in a short period, i.e., under the 15 min time mark, can greatly increase the infection risk, especially for so-called "superspreading events", which invariably occur indoors [21]. Moreover, the majority of time has to be spent by people in indoor contexts with plenty of daily activities performed. However, GNSS-based approaches do not work well in indoor environments due to signal attenuation. Phone-to-phone pairing-based methods using Bluetooth low energy (BLE) work only for direct face-to-face contact tracing scenarios and are inapplicable to indirect virus exposure in ambient aerosols. The expelled pathogen-containing particles can remain active in the air

for hours without sufficient sanitization, especially in indoor environments, constructing a significant fraction of the virus concentration [140]. Recently, vContact was proposed as a means to detect exposure to the virus with the consideration of asynchronous contacts by leveraging Wi-Fi networks, while the spatiotemporal dynamism in the virus concentration is not fully being considered [72].

Although the virus concentration will gradually decrease due to inactivation, deposition, and air purification after the virus-laden droplets are exhaled, the poor air ex-change rate, superspreaders, and more virulent variants will keep it at a relatively high concentration for a long time in an indoor environment [35, 18]. The viral particles are continuously ejected by infected people at different locations, relying on human movement. Moreover, due to the initial motion state and environmental airflow, these droplets maintain a ceaseless transmission before they are removed and meet somewhere (at some time), which leads to constant changes in the virus concentration within the control volume [142]. To accurately estimate the concentration, investigating the airborne transmission of these ejected particles is, thus, of fundamental importance in a closed environment because of the assemblage, in which human movement is implicitly involved to achieve the initial motion state of droplets [142]. The qualitatively location-specific assessment of the viral concentration is proposed with the dual use of computational fluid dynamic simulations and surrogate aerosol measurements for different real-world settings [100]. Moreover, the transmission of the virus brings about changes in the viral concentration of a specific location in an overall space, as well as the movements of people. Z. Li et al. analyzed the dispersion of cough-generated droplets in the wake of a walking person [77].

To be precisely aware of the amount of the virus one is exposed to and to detect both direct and indirect contacts, an indoor spatiotemporal contact awareness (iSTCA) framework is proposed. Since the virus concentration (at different times in the same area) is not the same because of the dispersion and diffusion of

the virus and human movements, we employed a self-contained PDR technique to calculate the human trajectory with accuracy and further achieve the location and time of the expelled virus droplets for the quantitative measurement of the concentration at any time in different spots. Moreover, based on the acquired changing virus concentration and reliable trajectories, the exposure time and distance of both direct and indirect contacts can be derived via cross-examination to realize quantitative spatiotemporal contact awareness.

Our main contributions are as follows:

1. To accurately present the virus concentrations at different times, we established quantitative virus concentration changes in various areas of indoor environments at different times by infected individuals. The viral-laden droplets were continuously released during the expiratory activities, moving forward. During the movements of viral-loaded droplets exhaled by infectious individuals at different locations and times, the virus instances met in certain spots at certain times and contributed to the calculation of the concentration. Finally, the concentration of each virus instance was integrated.

2. We employed PDR for the acquisition of the trajectory to conduct contact awareness without requiring extra infrastructure or being affected by coverage limitations compared with other indoor positioning techniques.

3. We considered various landmarks to calibrate the cumulative error for trajectory achievement by using PDR. A custom deep neural network using bidirectional long short-term memory (Bi-LSTM) and multi-head CNNs with residual concatenations was designed and implemented to extract temporal information in forward and backward directions and spatial features at various resolutions from built-in sensor readings for landmark identification.

4. Additionally, we demonstrate the effectiveness of the proposed Bi-LSTM-CNN classification model for landmark identification through empirical ex-

periments, as well as the performance of our proposed iSTCA system for quantitative spatiotemporal contact analytics.

The remainder of this part is organized as follows. The related work about contact awareness and indoor localization techniques, including PDR, is reviewed in Section 5.2. Definitions and preliminaries about virus concentrations and different contact types are introduced in Section 5.3. Section 5.4 introduces the theoretical methodology and the architecture of the proposed iSTCA. The experimental methodology and results based on the collected datasets are presented in Section 5.5. Section 5.6 reveals the limitations of this work. Finally, we present the conclusion and future work in Section 5.7.

5.2 Related Work

Contact tracing is used to identify and track people who may have been exposed to a virus, due to the prevalence of many infectious diseases in our society. To conduct contact tracing, it is necessary for the infected individuals to provide their visited locations and people whom they encountered based on the specific definitions of meetups for different diseases. Instead of interviews and questionnaires via traditional manual tracing, technology-aided contact tracing can track people at risk conveniently and intelligently. To reduce the spread of COVID-19 effectively, digital contact tracing, which generally depends on applications installed on smartphones, has been developed in both academia and industry, using various technologies, such as GNSS, Bluetooth, and Wi-Fi.

There are typically two approaches for encounter determinations, *peer-to-peer proximity detection*-based and *geolocation*-based. Peer-to-peer proximity can be estimated by the RSS of wireless signals, such as Bluetooth and UWB, and the distance between two devices in geolocation-based approaches can be precisely derived from the cross-examination after obtaining the accurate location and tra-

jectory with the help of localization techniques using various technologies, such as GPS, Wi-Fi, and PDR.

Some systems based on peer-to-peer proximity using Bluetooth or BLE have been implemented, and part of them are deployed by the governments of various countries, such as Australia (COVIDSafe), Singapore (Trace together), and the United Kingdom (NHS COVID-19 App) due to their ubiquitous embedding in mobile phones [116]. Among these systems, the most representative protocols are Blue Trace and ROBERT [21, 10]. The data from Bluetooth device-to-device communications are stored and checked against the data uploaded by the infector. In Blue Trace, the health authority contacts individuals who had a high probability of virus exposure, whereas ROBERT users need to periodically probe the server for their infection risk scores. In addition, Google and Apple provide a broadly used toolkit based on Bluetooth, named Google and Apple Exposure Notification (GAEN), to facilitate a contact tracing system in Android and iOS and curb the spread of COVID-19 [69]. Despite some minor differences in implementation and efficiency, these schemes are all independently designed and very similar. When exposure is detected, the RSS in the communication data frame is utilized to estimate the distance between two devices and notify the user. However, it has been demonstrated that the signal strengths can only provide very rough estimations of the actual distances between devices, as they are affected by device orientation, shadowing, shading effects, and multipath losses in different environments [68, 70]. Although it is difficult to measure the distances among users accurately by using Bluetooth and other technologies, the UWB radio technology has the capacity to measure distances at the accuracy level of a few centimeters, which is significantly bettering than Bluetooth [52]. The use of UWB, however, has some significant drawbacks, including the fact that UWB is not widely supported by mobile devices, requires extra infrastructure, and is not energy efficient, which makes UWB less useful in practice [14]. All of the above works that are based on calculated

proximity using RSS do not consider the user's specific physical location, resulting in unsatisfactory tracing results. Moreover, these approaches cannot be applied to the detection of temporal contact due to the dispersion and lifespan of the virus.

To achieve accurate geolocation in contact tracing, plenty of localization systems have been researched with the joint efforts of researchers and engineers in the past based on GNSS, cellular technology, RFID, and quick response (QR) code [97]. GNSS can be used for contact tracing as the exact position of a person can be located and it is available globally. Many countries, including Israel (HaMagen 2.0) and Cyprus (CovTracer), use GPS-based contact tracing approaches [116] as well. GNSS signals are usually weak in indoor environments due to the absence of the line of sight and the attenuation of satellite signals, as well as the noisiness of the environment. Many people may spend most of their time in indoor environments, which can result in limited contact coverage. It is difficult to detect contact based on cellular data due to the large coverage of cell towers and high location errors [72]. RFID was used to reveal the spread of infectious diseases and detect face-to-face contact in [109, 51]. QR codes for contact tracing require users to check in at various venues by scanning the placed QR codes manually to record their locations and times, which are deployed in some countries, such as New Zealand (NZ COVID Tracer) [116]. However, special devices or codes have to be deployed at scale for data collection. Recently, some protocols were proposed for Wi-Fi-based contact tracing with the pre-installed Wi-Fi Access Point. WiFiTrace was proposed by proposed in [133]. WiFiTrace is a network-centric contact tracing approach with passive Wi-Fi sensing and without client-side involvement, in which the locations visited are reconstructed by network logs; graph-based model and graph algorithms are employed to efficiently perform contact tracing. Wi-Fi association logs were also investigated in [127] to infer the social intersections with coarse collocation behaviors. Li et al. utilized active Wi-Fi sensing for data collection; they leveraged signal processing approaches and similarity metrics to align and detect virus

exposure with temporally indirect contact [72]. As the changes in virus concentrations over time (due to the transmission of aerosols and environmental factors) are not considered, their results are in relatively low spatiotemporal resolutions. The approach presented in [134] divides contact tracing into two separate parts, duration and distance of exposure. The duration is captured from the Wi-Fi network logs and the distance is calculated by the PDR positioning trajectory, calibrated by recognized landmarks with the help of a CNN, ensuring the performance of contact tracing. Although integration with the existing infrastructure is beneficial in mitigating the deployment costs, it may not fully satisfy the requirements of contact tracing with the high spatiotemporal resolution because of the absent coverage [39]. The trajectory obtained by the PDR technique, without requiring special infrastructure, can improve the coarse-grained duration and make it fine-grained. This can enable the development of a contact-tracing environment that considers the virus lifespan in detail.

One of the ultimate goals of contact awareness systems is to estimate the risk based on the recorded encounter data [60]. Moreover, with the exposure duration and distance obtained, the virus concentration is significant to determine the exposed viral load, which is closely associated with the infection risk [119]. Typically, the virus concentration in a given space depends on the total amount of viral load contained in the viable virus-laden droplets in the air and maintains a downward trend because of the self-inactivation and environmental factors. Researchers presented the qualitative location-specific assessment of viral concentration with the dual use of computational fluid dynamic simulations and surrogate aerosol measurements for different real-world settings [100]. The practical viral loads emitted by contagious subjects based on the viral loads in the mouth (or sputum) with various types of respiratory activities and activity levels are presented in [119]. Furthermore, to quantitatively shape the virus concentration in a targeted environment at different times, the constant viral load emission rate is adopted with

the virus removal rate, including the air exchange rate, particle sediment, and viral inactivation rate in [19].

The aforementioned contact tracing research usually only considers the *static* virus concentration without considering the exposure to the environmental virus and *dynamism* in the virus concentration. Moreover, in contrast to the qualitative estimation of exposure risks that can be achieved in previous works, there is a lack of sufficient quantitative awareness about the concentrations of contracted viruses. Such awareness would be useful in our daily lives to protect ourselves from virus infections.

5.3 Definitions and Preliminaries

Virus-encapsulating secretions are continuously exhaled and aerosolized into airborne virus-laden particles with infectivity from daily expository activities. There is a great difference between the size and number of droplets expelled, depending on their origin locations in the respiratory tract [77]. The time and distances of these droplets traveling in indoor environments largely depend on the expiration air jet, particle weight, and ambient factors. The movements and the viral loads of virus-containing particles are directly associated with the virus concentrations in different regions. To quantitatively become aware of the exposure of the virus, the quanta concentration as a medical virus concentration indicator, virus airborne pattern, and various contact types are present.

5.3.1 Quanta Concentration

The viral loads of virus-containing droplets change after leaving the human expiratory tract with airborne transmission and a combination of environmental factors. In particular, the viral load emitted is expressed in terms of the quanta emission rate ($ER_q,\ quanta \cdot h^{-1}$), in which a quantum is defined as the dose of airborne

droplet nuclei that infect 63% of susceptible persons with exposure [19]. The quanta concentration in an indoor area at time t, $q(t)$ is measured by:

$$q\left(t, ER_q\right) = N_I \cdot \frac{ER_q}{RR_{iv} \ \cdot \ V} + \left(q_0 + N_I \cdot \frac{ER_q}{RR_{iv}}\right) \cdot \frac{e^{-RR_{iv} \cdot t}}{V} (quanta \cdot m^{-3}) \quad (5.1)$$

where ER_q is the quanta emission rate of the infector (measure in $quanta \cdot h^{-1}$), q_0 is a constant declaring the initial number of quanta in the space, V (m^3) is the target indoor volume, N_I represents the number of infected individuals in the investigated volume, RR_{iv} (h^{-1}) is the removal rate for the infectious virus in the considered spaces [19]. RR_{iv} consists of three contributions, the air exchange rate (AER) via ventilation, the deposition on surface rate (k) caused by gravitational sedimentation and turbulent eddy impaction, and the viral inactivation rate (λ). The typical k is 0.24 h^{-1} and the inactivation rate λ of viable COVID-19 particles in a typical indoor environment without sunlight is generally 0.63 h^{-1}, as indicated in [19, 136]. The ER_q is determined by the viral load in sputum, the volume of signal droplets, and the quantity of all expelled droplets per exhalation. Thus, the quanta concentration ER_q is modeled as:

$$ER_q = c_v \cdot c_i \cdot IR \cdot \int N_d(D) \cdot dV_d(D) \ (quanta \cdot h^{-1}) \quad (5.2)$$

where c_v represents the viral load in the sputum of the infector (RNA copies$\cdot mL^{-1}$), IR is the inhalation/exhalation rate produced by the breathing rate and tidal volume, N_d is the droplet concentrations in different expiratory activities of the infected person $(particles \cdot cm^{-3})$, V_d is the volume of a single droplet (cm^3) with the function of particle diameters D, and c_i is the conversion factor, presenting the ratio between one infectious quantum and the infectious dose expressed in the viral RNA copies [119]. There is a wide range of variations in the quanta emission estimation via Equation (2), depending on these and other factors, such as virus concentration in the mouth, activity level, and the type of coughing or exhaling.

With light exercise and speaking, a quanta concentration of 142 ($quanta \cdot h^{-1}$) can be obtained, which was widely adopted in many works [119].

5.3.2 Spatial–Temporal Contact

COVID-19 contained in expiratory droplets and expelled from the infector is transported and dispersed in the ambient airflow before finally being removed, inactivated, and inhaled by a susceptible. There are a number of factors that contribute to the droplet's movement, such as the horizontally emitted velocity, the particle weight and the external environment. Occasionally, coughing and sneezing generate more particles with higher initial velocities (11.7 $m \cdot s^{-1}$ for coughing) and virus quanta concentrations, while constantly performed breathing and speaking (3.9 $m \cdot s^{-1}$ for speaking) produce fewer particles with relatively lower initial velocity and virus quanta concentrations [119]. Large droplets usually settle quickly in a few seconds or minutes owing to gravitational sedimentation and are evaporated into small nuclei in indoor environments, where the particle can disperse for a long distance in the vaporization process. Tiny particles, including ones that are evaporated and originally expelled, are trapped and carried continuously forward within a moist, warm, turbulent cloud of gas, with the help of airflow movement. To facilitate the calculation, the movement of each virus-laden droplet expelled at each moment is independent and divided into two stages, maintaining a uniform motion with the initial horizontal velocity (e.g., 3.9 $m \cdot s^{-1}$), being well-mixed within the moved space in the first phase (e.g., 1s), and then instantaneously and evenly distributed in the overall considered space.

The contact in COVID-19 contact tracing is originally equivalent to direct face-to-face contact, while due to the transmission of the virus and survival time in the air, more cases of indirect contact have emerged [88]. Here, indirect contact mainly represents the asynchronous time contact, called temporal contact. Direct and indirect contacts are types of spatiotemporal contacts. If there is no time

difference between two people, and only a spatial distance is presented, it is called spatial contact. Similarly, if there is no space difference between two people, and only a temporal distance is presented, it is called temporal contact. There are time and space gaps, a mixture of two single cases, called spatiotemporal contacts. Since both the time and space differences would decrease the virus quanta concentration, it is necessary to obtain the accurate value for the precise awareness of the virus quanta concentration.

5.4 Methodology

This work utilizes the trajectories, including the spatial position coordinates and time obtained by the PDR technique to quantitatively estimate the time-dependent changes in the virus quanta concentration derived from the movement and lifespan of the virus in various places of the considered indoor environment. The overview of the proposed scheme is systematically introduced in Section 5.4.1. In Section 5.4.2, we provided the data processing approaches utilized in PDR-based trajectory construction and the estimation of droplet exhalation. The PDR technique (with a calibration of the landmark recognized by a landmark identification model based on a residual Bi-LSTM and CNN structure) is discussed in Section 5.4.3. Further, the contact awareness model relying on the precisely constructed pedestrian trajectory is detailed in Section 5.4.4.

5.4.1 System Overview

An overview of the proposed iSTCA system is presented in Figure 5-1. More precisely, the data flow of various sensors for the analysis was primarily collected from the existing sensors in handhold smartphones, which record the changes in the environment and body motion. The signals need to be processed, including data filtering and scaling, to reduce the noise for a better state of motion estimation

before training the landmark identification model and performing the PDR. The trajectory can be achieved based on the PDR technique and properly corrected with the assistance of the identified landmark distinguished by the trained landmark recognition model. The trajectory is defined as a set of points consisting of the time and position, $\{(t_0, x_0, y_0), (t_1, x_1, y_1), \cdots (t_n, x_n, y_n)\}$ where (x_i, y_i) represents the location coordinates and t_i is the moment when the individual passes the location. The virus quanta concentrations in different spaces at various moments can be measured quantitatively to achieve sufficient awareness with the help of the estimated spatial distance, temporal distance and infectivity model, as shown in Equation 5.1.

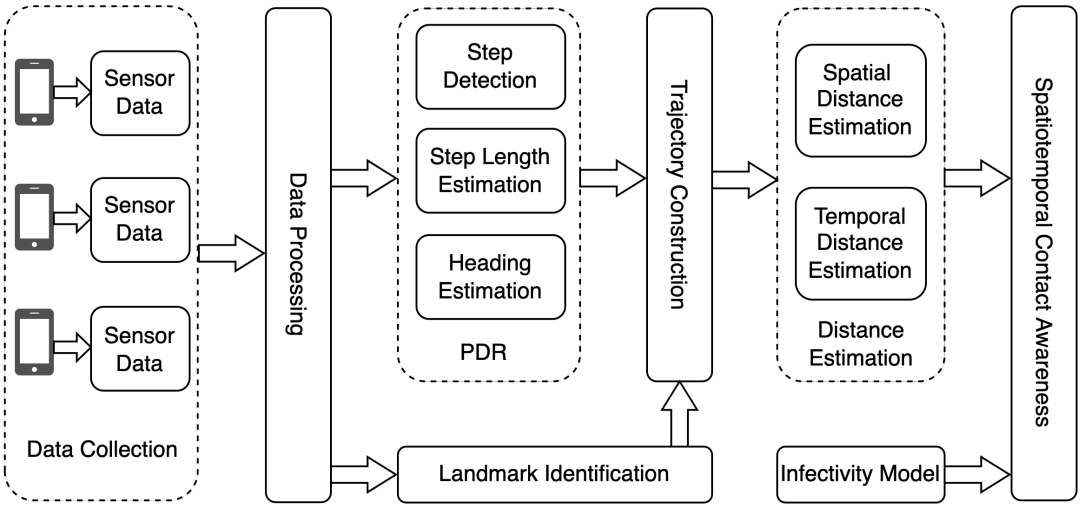


Figure 5-1: Overview of iSTCA

5.4.2 Data Preprocessing

In order to integrate data from various sensors for the purpose of landmark calibrated PDR, the requisite data preprocessing methodologies are identical to the data preprocessing approaches delineated in Section 4.3.2, Chapter 4, involving data alignment, data interpolation, data filtering, data scaling and data segmentation.

5.4.3 PDR-Based Trajectory Construction Model

For the quantitative evaluation of the virus quanta concentration, the precise spatial distance and temporal difference between two individuals should be efficiently estimated. To reach this objective, a diverse range of indoor localization techniques have been investigated for various scenarios. The widely studied fingerprinting-based method relies on the latest fingerprint database that needs to be precisely updated in time. In addition to the time-consuming and labor-intensive collection and re-establishment, the instability of RSS due to environmental uncertainties poses another challenge to the accuracy [126]. Moreover, coverage and distribution are also not satisfied in countries with poor ICT infrastructure [63]. Therefore, the self-contained PDR algorithm without extra requirements and coverage limitations is employed in this work, and its accuracy is improved by the identified landmark.

PDR

Since PDR does not need additional equipment or a pre-survey, it has a wide range of potential applications for the indoor positioning of pedestrians. It relies on the inertial sensors extensively existing in mobile devices, e.g., smartphones, to acquire information about the user's movements, which are then combined with the user's previous location to estimate the present position and further achieve complete trajectory. The equation utilized for location estimation is as follows:

$$\begin{cases} x_t = x_{t-1} + SL_t \sin\theta_t \\ y_t = y_{t-1} + SL_t \cos\theta_t \end{cases} \tag{5.3}$$

where (x_t, y_t) is the pedestrian position at time t, SL_t is the step length, and θ_t details the heading direction of the pedestrian [76].

As mobile technology continues to evolve, a growing number of physical sensors are being installed in smartphones and, thus, various combinations of sensors can

provide increasingly rich information, which makes PDR more feasible and accessible. A typical PDR system is comprised of three primary elements: step detection, step-length estimation, and heading estimation [82]. Each of these components is expounded upon in Section 4.3.3, Chapter 4.

Landmark Identification Model

Although PDR methods can estimate the location and trajectory of pedestrians, low-cost inertial sensors built into smartphones provide poor-quality measurements, resulting in accuracy degradation. Moreover, the cumulative error, including the heading estimation caused by the gyroscope and step-length estimation error caused by an accelerometer, could be produced in the long-term positioning using PDR, increasing the challenge of precise localization collection. Therefore, it is necessary to prepare the reference points with the correct positions known during the movement to reduce the accumulated errors when the user passes. Spatial contexts, such as landmarks, can be selected to calibrate the localization error based on the inherent spatial information without additional deployment costs. Landmark is defined as a spatial point with salient features and semantic characteristics from its near environment in indoor positioning systems, such as corners, stairs, and elevators [39]. These features can be observed for identification in one or a combination of different sensors as people pass through the landmark. The locations of these landmarks are presented by geographical coordinates or the relationships with other locations/areas, where people perform specific and predictable activities. Changes in motion are reflected in sensor readings, and different motions present different patterns. The specific activities that people perform when passing landmarks are also reflected in at least one sensor. Using the data of one sensor or the combination of data from multiple sensors, the changing pattern of a specific activity can be identified, and then the landmark can be recognized [47]. The identified landmark can be used as an anchor point to correct the path we

obtained and improve the performance of the calculated trajectory.

Landmark identification involves classifying the sequences of various sensor data recorded at regular intervals by sensing devices, usually smartphones, into a well-defined landmark, which has been extensively regarded as a problem of multivariate time series classification. To address this issue, it is critical to extract and learn the features comprehensively to determine the relationship between sensing information and movement patterns. In recent years, numerous features have been attained in many studies on certain raw signal statistical aspects, such as variance, mean, entropy, kurtosis, correlation coefficients, or frequency domains via the integration of cross-formal codings, such as signals with Fourier transform and wavelet transform [95]. Moreover, the special thresholds of different features for various kinds of landmark recognition are specifically analyzed. For instance, the threshold of angular velocity produced by a gyroscope is usually used to detect the corner landmark, the acceleration changes can recognize the stairs. The combinations of different thresholds of various sensors forming the decision tree can detect the standing motion state to further distinguish common landmarks, such as corners, stairs, and elevators [167, 144]. However, despite high accuracy, the calculation, extraction, and selection of features of different sensors for various landmarks are heuristic (with professional knowledge and expertise of the specific domain), time-consuming, and laborious [95].

To facilitate feature engineering and improve performance, artificial neural networks based on deep learning techniques have been employed to conduct activity identification without hand-crafted extraction. Deep learning techniques have been applied in many fields to solve practical problems with remarkable performance, such as image processing, speech recognition, and natural language processing, to solve practical problems [62, 66]. Many kinds of deep neural networks have been introduced and investigated to handle landmark identification based on the complexity and unsurenesss of human movements. Additionally, CNN and LSTM are

widely adopted with high accuracy rate activity recognition among the applied networks. CNN is commonly separated into numerous learning stages, each of which consists of a mix of convolutional operation and nonlinear processing units, as follows:

$$h^k = \sigma(\sum_{l \in L} g\left(x^l, w^k\right) + b^k) \tag{5.4}$$

where h^k reveals the latent representation of the k-th feature map of the current layer, σ is the activation function, g denotes the convolution operation, x^l indicates the l-th feature map of the group of the feature maps L achieved from the upper layer, w^k and b^k express the weights matrix and the bias of the k-th feature map of the current layer, receptively [130]. In our model, the rectified linear units (ReLU) were employed as the activation functions to subsequently conduct the non-linear transformation to obtain the feature maps, denoted by:

$$\sigma(x) = max(0,\ x) \tag{5.5}$$

More importantly, the convolution operation in CNN can efficiently capture the local spatial correlation features by limiting the hidden unit's receptive field to be local [143]. CNN considers each frame of sensor data as independent and extracts the features for these isolated portions of data without considering the temporal contexts beyond the boundaries of the frame. Due to the continuity of sensor data flow produced by the user's behavior, local spatial correlations and temporally long-term connections are both important to identify the landmark [130]. LSTMs with learnable gates, which modulate the flow of information and control when to forget previous hidden states, as variants of vanilla RNNs, allow the neural network to effectively extract the long-range dependencies of time-series sensor data [32]. The hidden state for the LSTM at time t is represented by:

$$h_t = \sigma(w_{i,h} \cdot x_t + w_{h,h} \cdot h_{t-1} + b) \tag{5.6}$$

where h_t and h_{t-1} are the hidden state at time t and $t-1$, respectively, σ is the activation function, $w_{i,h}$ and $w_{h,h}$ are the weight matrices between the parts, and b symbolizes the hidden bias vector. The standard LSTM cells barely extract the features from the past movements, ignoring the future part. To comprehensively capture the information for landmark identification, the Bi-LSTM is applied to access the context in both the forward and backward directions [31].

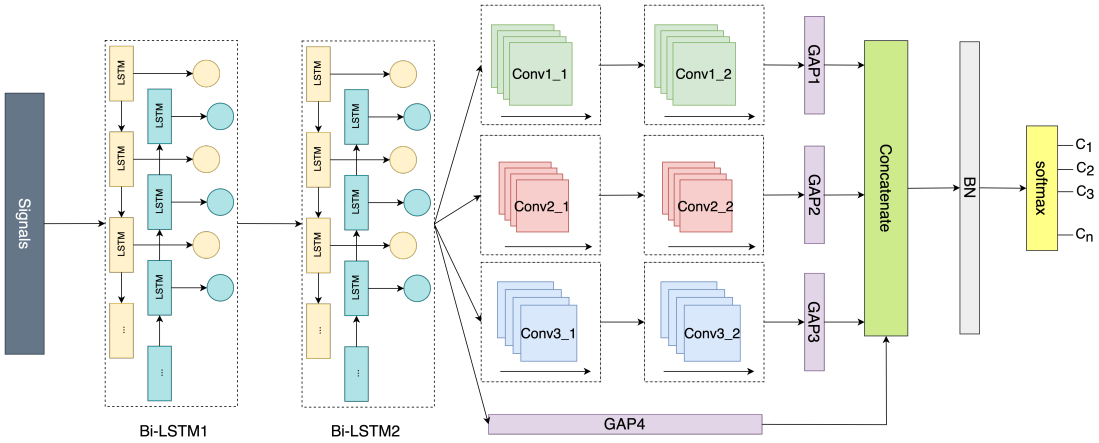


Figure 5-2: Architecture of the landmark identification model

Therefore, both Bi-LSTM and CNN are involved in capturing the spatial and temporal features of signals for landmark identification. The architecture of the proposed landmark identification is shown in Figure 5-2. It performs the function of landmark recognition using the residual concatenation for classification, followed by Bi-LSTM and multi-head CNNs. When preprocessed data segmentations of multiple sensors come, the inherent temporal relationship is extracted sequentially by two Bi-LSTM blocks that consist of a Bi-LSTM layer, a batch normalization (BN) layer, an activation layer, and a dropout layer. BN is a method used to improve training speed and accuracy with the mitigation of the internal covariate shift through normalization of the layer inputs by recentering and rescaling [152]. Next, multi-head CNN blocks with varying kernels size are followed to learn the spatial features at various resolutions. Each convolutional block is

made of four layers: a one-dimensional (1D) convolutional layer, a BN layer, an activation layer, and a dropout layer. To accommodate the three-dimensional input shape (samples, time steps, input channels) of the 1D convolutional layer, we retain the output of the hidden state in the Bi-LSTM layer. Then the acquired spatial and temporal features are combined, namely the concatenations of the outputs of the multi-head CNNs and Bi-LSTMs. To reduce the parameters and avoid overfitting, the global average pooling layer (GAP) with no parameter to optimize rather than the traditional fully connected layer is applied before combining the outputs [131]. Finally, the concatenated features are transmitted into a BN layer to re-normalize before being fed into a dense layer with a softmax classifier to generate the probability distribution over classes.

5.4.4 Contact Awareness with Trajectory

Exhalation and inhalation respiratory activities are constantly alternating (e.g., each breath consists of $2.5s$ of continuous exhalation and $2.5s$ of continuous inhalation), and droplets are continuously being released from the respiratory tract with a horizontal velocity during the process of exhalation with the same direction as the movement of the human. The particles exhaled at each moment will continue to move forward, starting from the user positions when they are expelled. The viral droplets exhaled from the infectious host are transported and dispersed into the ambient airflow before finally being inhaled by a susceptible person. Each exhalation lasts several seconds (e.g., $2.5s$), in which a long distance can be traveled for those who are in motion, and the initial position of droplets expelled cannot be accurately estimated in an indoor environment. Therefore, once complete, the exhalation period is divided into many short-term (e.g., $0.1s$) particle ejections. Because the interval is short, the continuous virus exhalation process can be converted into an instantaneous process, i.e., the virus is released instantly at the beginning of each interval. The virus-laden droplets expelled at different inter-

vals maintain independent and identical motion patterns and the initial positions of the particles released in each interval can be regarded as the locations of the people at the initial moments. The virus-containing particles maintain a uniform motion of initially horizontal velocity (e.g., 3.9 $m \cdot s^{-1}$) in the first second and then instantaneously will mix in the overall considered space. Meanwhile, the droplets are evenly distributed within the moved space. In the first movement phase of the exhaled droplets in each interval, the virus moves in the same direction as the people travel, which is called forward transmission. As for the backward transmission, in general, the initial velocity of the virus is faster than the speed of movement and the speed of airflow, so in the first phase, very few virus particles move in the opposite direction.

The movements of all viral-loaded droplets exhaled by infectious people at different locations will meet somewhere at some time and contribute to the calculation of concentration. To precisely present the virus quanta concentrations, the transmissions of all virus particles per exhalation sources from different origins and in different states are assumed to follow the same patterns, in which the particles keep constant initial velocity in the first second and then will instantly mix in the overall space. The time it takes for the virus to move to the current point and the contribution to the virus quanta in the present are estimated with the help of spatial distance and velocity. Thus, the quanta concentration in an indoor area at time t, $q(t,\ ER_q)$ is measured by:

$$q\left(t,\ ER_q\right) = \sum_{i}^{i=N_v} \left(\frac{ER_q^i}{RR_{iv}\ \cdot\ V\left(t^i\right)} \cdot \left(1 + e^{-RR_{iv} \cdot t^i}\right) + \left(q_0 \cdot \frac{e^{-RR_{iv} \cdot T}}{V} + q_0^i \cdot \frac{e^{-RR_{iv} \cdot t^i}}{V} \right) \right) \tag{5.7}$$

where RR_{iv} is the virus removal rate of the target space, N_v represents the virus generated in different places at different moments, ER_q^i is the of the quanta emission rate of the infector at which the virus (i-th) is expelled, T is the time difference

from the start of the experiment to present, t^i is the time difference between the current time and the originating time of the virus (i-th), $V(t^i)$ is the volume of the space that the i-th virus had passed since it was expelled to the present, q_0 is the environmental virus quanta number, q_0^i is the virus exhaled by the infector that has evenly spread to the overall investigated space with the volume of V. Exhaled virus particles eventually become the environmentally well-mixed virus quanta, while different initial states induce different decays.

5.4.5 Spatiotemporal Contact Awareness

The algorithm of the proposed iSTCA with the landmark-calibrated PDR technology based on a smartphone is detailed in Algorithm 1. The detailed procedures are as follows,

Firstly, the raw signals are acquired via the developed collection application and preprocessed to create the dataset for the landmark identification model training by utilizing the data preprocessing method introduced in Section 5.4.2.

Secondly, the landmark recognition model designed in Section 5.4.3 would be trained and stored based on the dataset generated in the first step to further the PDR algorithm.

Thirdly, the target trajectory $\mathcal{S}$ is constructed by performing the landmark-calibrated PDR technique, including step detection, stride length estimation, heading determination, and landmark identification.

Fourthly, we obtain the initial state set $\{Q_0^i\}$ of the expelled particles in the i-th ($i = 1, 2, 3 \ldots$) short-term period with the help of the calculated human movement trajectory $\mathcal{S}$ and the preset viral particle ejection interval τ. Q_0^i defines the state of all i-th emitted particles in interval τ and consists of three parts t, V, q, where t represents the elapsed time after being exhaled, V represents the spread coverage of droplets due to airborne dispersion, and q represents the quanta concentration.

Fifthly, the state set $\{Q_j^i\}$ at the j-th interval for any Q_0^i after being expelled is

Algorithm 1 Indoor spatiotemporal contact awareness algorithm

Input: raw sensor signals of infector's smartphone;
trained landmark identification model $\mathcal{M}_{lm}$;
target time $\mathcal{T}$;
target position $\mathcal{P}$;
Infectivity model $\mathcal{M}_I$.

Output: quantitative virus quanta concentration in $\mathcal{P}$ at $\mathcal{T}$.

time interval initialized to τ;
quanta concentration in $\mathcal{P}$ at $\mathcal{T}$ ($q_{\mathcal{P}}^{\mathcal{T}}$) initialized to 0;
construct the processed signals $\mathcal{D}$;
achieve the trajectories $\mathcal{S}$ from $\mathcal{D}$, landmark-calibrated via $\mathcal{M}_{lm}$;
establish the initial state set $\{Q_0^i\}$ of all viruses expelled at different intervals, where $Q_0^i \leftarrow (t_0^i,\ V_0^i,\ q_0^i)$, i represent the number of time intervals;
for each Q_0^i **do**
    for j *in* $0, 1, 2, \ldots \lceil \frac{\mathcal{T}-t_0^i}{\tau} \rceil$ **do**
        achieve the $Q_j^i \leftarrow \left(t_j^i,\ V_j^i,\ q_j^i\right)$ with movement pattern ($\mathcal{M}_I$) itself;
        if $\mathcal{P}$ *in* V_j^i **then**
            update $q_{\mathcal{P}}^{\mathcal{T}}$, $q_{\mathcal{P}}^{\mathcal{T}} \leftarrow q_{\mathcal{P}}^{\mathcal{T}} + q_j^i$
        end if
    end for
end for
return $q_{\mathcal{P}}^{\mathcal{T}}$

acquired by employing the defined movement pattern of the considered particles.

Finally, the virus quanta concentration $q_{\mathcal{P}}^{\mathcal{T}}$ in the target position $\mathcal{P}$ at the target time $\mathcal{T}$ is reached. The virus quanta concentration presented within $\mathcal{P}$ at $\mathcal{T}$ by particles expelled in the various intervals is summed to estimate $q_{\mathcal{P}}^{\mathcal{T}}$. Moreover, the virus quanta concentrations presented in different locations at various times can be further evaluated.

5.5 Experiments

In this section, we evaluate the performances of the proposed methods through experiments with the dataset we collected in a university building. We introduce the experimental scenario and data collection in Section 5.5.1 and the results are presented in Section 5.5.2; we analyzed the performances related to the landmark identification, PDR, and virus quanta concentration.

5.5.1 Experimental Scenario and Data Acquisition

We collected our experimental data on the fourth floor of the Center-Zone-1 building of Kyushu University's Ito campus. We assume that there is no exchange of virus particles with the room space. Figure 5-3 shows the floor plan of the experimental area. Based on the practical scenario, the Manhattan distance is applied to measure the virus movement. Since the width (measured as $2m$) and height (assumed to be $3m$ based on the practical scenario) of the hallway are generally the same, the volume of the virus coverage can be determined by the virus movement distance for the calculation of the virus quanta concentration. When the virus encounters a corner, its direction changes, leading to a shift in the virus quanta concentration to varying degrees. For a corner with two branches, the concentration is assumed to decrease by half due to the inertia effect while these viral particles continue the forward transmission. If it is a corner with three or more branches, we assume that the virus quanta would be distributed evenly in all other directions.

In data collection, five recruited participants held Pixel 4a smartphones with the required sensors integrated (e.g., accelerometer, gyroscope, and rotation sensor) and an Android application installed. The application can periodically read and store the readings of 11 channels (3 for the accelerometer, 3 for the gyroscope, and 5 for the rotation sensor) as the user walks along the prescribed routes at

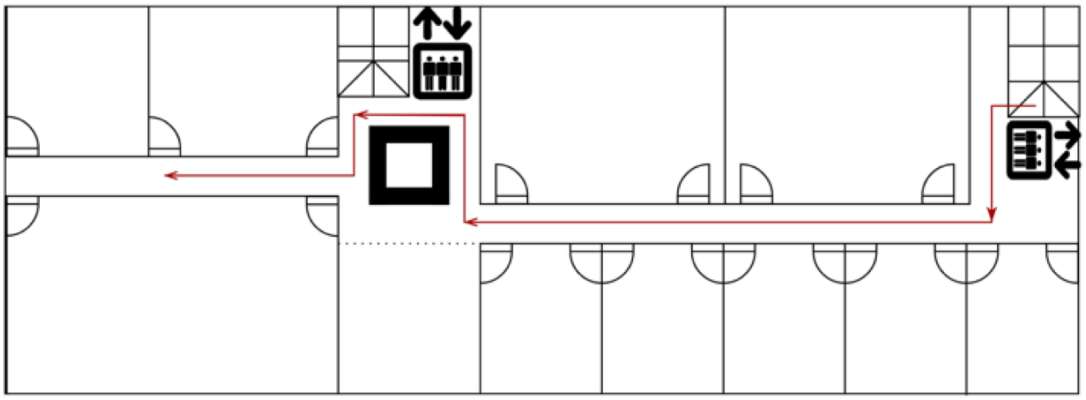


Figure 5-3: The floor plan of our experiment

a normal speed in the experimental area. Moreover, participants are required to hold their smartphones at chest level, which is a reasonable position where participants can record extra information to facilitate data processing. Indeed, it is recommended that they record the timestamp and the identification of the passing landmark to construct the dataset for the landmark identification model training.

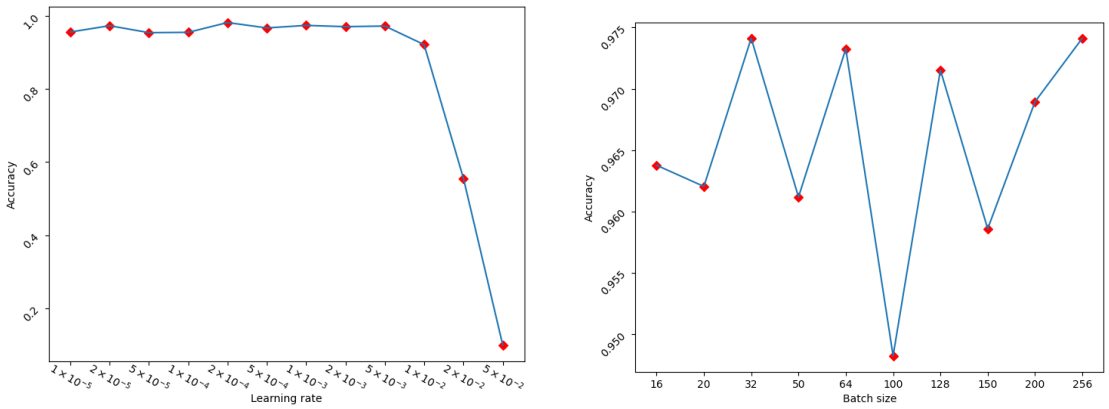
5.5.2 Analysis and Discussion

Landmark Identification

The proposed landmark recognition model was extensively evaluated by a series of experiments and implemented using the Keras framework with the TensorFlow backend to minimize the cross-entropy loss. The model was performed using the collected data with 3863 samples. The dataset was divided into training (70%) and testing (30%) sets, randomly, without overlapping. There were a total of 11 landmarks, including 7 corners, 2 stairs, and 2 elevators.

Table 5.1 details the network configuration considered in our study. Since there were many combinations of parameters, to reduce the selection space, we let all of the Bi-LSTM neurons share the same value, with the 1D convolution filter and kernel sizes accessing the same setting, respectively. To achieve stable

performances of different model settings, a grid search with the 10-fold cross-validation method was adopted. It worked through all of the combinations of parameters to find the best settings. It should be noted that the following study uses the bold value for each parameter when it is not otherwise specified.



(a) learning rates (b) batch sizes

Figure 5-4: Model accuracy on various learning rates and batch sizes

Moreover, the model configuration (the Adam optimization algorithm) was selected as the optimizer during the gradient descent. Other training hyperparameters were also evaluated and their recognition accuracies are presented in Figure 5-4. More specifically, the experiment was conducted with the learning rates of 0.00001, 0.00002, 0.00005, 0.0001, 0.0002, 0.0005, 0.001, 0.002, 0.005, 0.01, 0.02, and 0.05, as presented in Figure 5-4a. The mini-batch size was tested with 16, 20, 32, 50, 64, 100, 128, 150, 200, and 256, as shown in Figure 5-4b. The model configured as Table 5.1 achieved the highest identification accuracy of 98.4% when the learning rate was 0.0002 and the mini-batch size was 256. Additionally, early stopping criteria and a learning rate reduction strategy were applied during the model training process in order to reduce the issue of over-fitting and to improve the model performance. The learning rate decreased with a factor of 0.5 when the accuracy was not improved for 10 epochs and the training ended if the accuracy without enhancement on the validation was set after 15 iterations. Detailed

Table 5.1: Landmark identification neural network configuration.

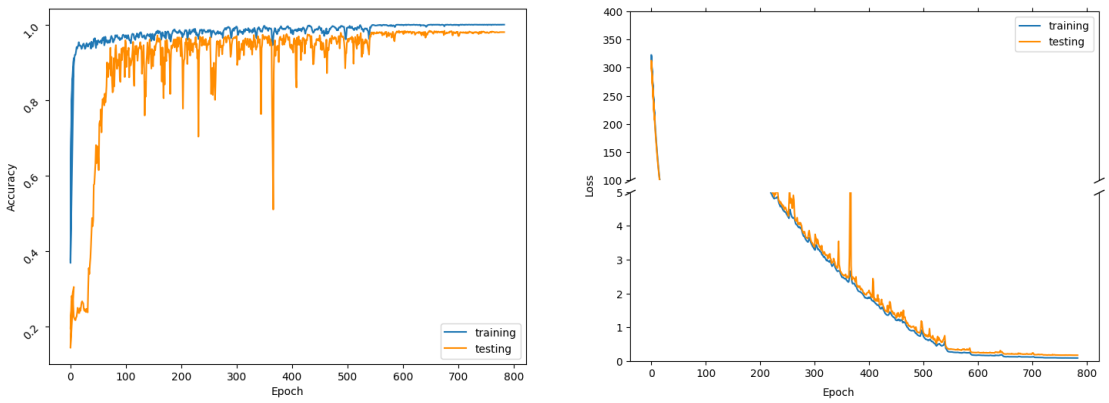
Layers	**Parameter**	**Value**
Input	shape	(None, 128, 11)
Bi-LSTM1	neurons	32, 64, **128**, 256
Bi-LSTM2	neurons	32, 64, **128**, 256
Conv1_1	kernel size	3, **5**, 9, 11
	filters	32, 64, **128**
	stride	1
Conv2_1	kernel size	3, **5**, 9, 11
	filters	32, 64, **128**
	stride	1
Conv3_1	kernel size	3, 5, **9**, 11
	filters	32, 64, **128**
	stride	1
Conv1_2, Conv2_2, Conv3_2	kernel size	**3**, 5, 9, 11
	filters	32, **64**, 128
	stride	1
Dropout	drop rate	0.2, **0.3**, 0.5, 0.8

Table 5.2: Training hyperparameters

Hyperparameters	**Value**
Optimizer	Adam
Activation function	ReLU
Batch size	256
Learning rate	0.0002
Epochs	800

training hyperparameter settings are revealed in Table 5.2.

Following the considered model configuration and optimal training hyperparameters, the accuracy curve and loss curve of the training and testing processes are illustrated in Figure 5-5. The recognition results on 11 selected landmarks of the experiment-conducted floor are presented by the confusion matrix in Figure 5-6.



(a) Accuracy (b) Loss

Figure 5-5: Accuracy (**a**) and loss (**b**) curves of the model on the selected parameters

Moreover, to evaluate the proposed network more comprehensively, further

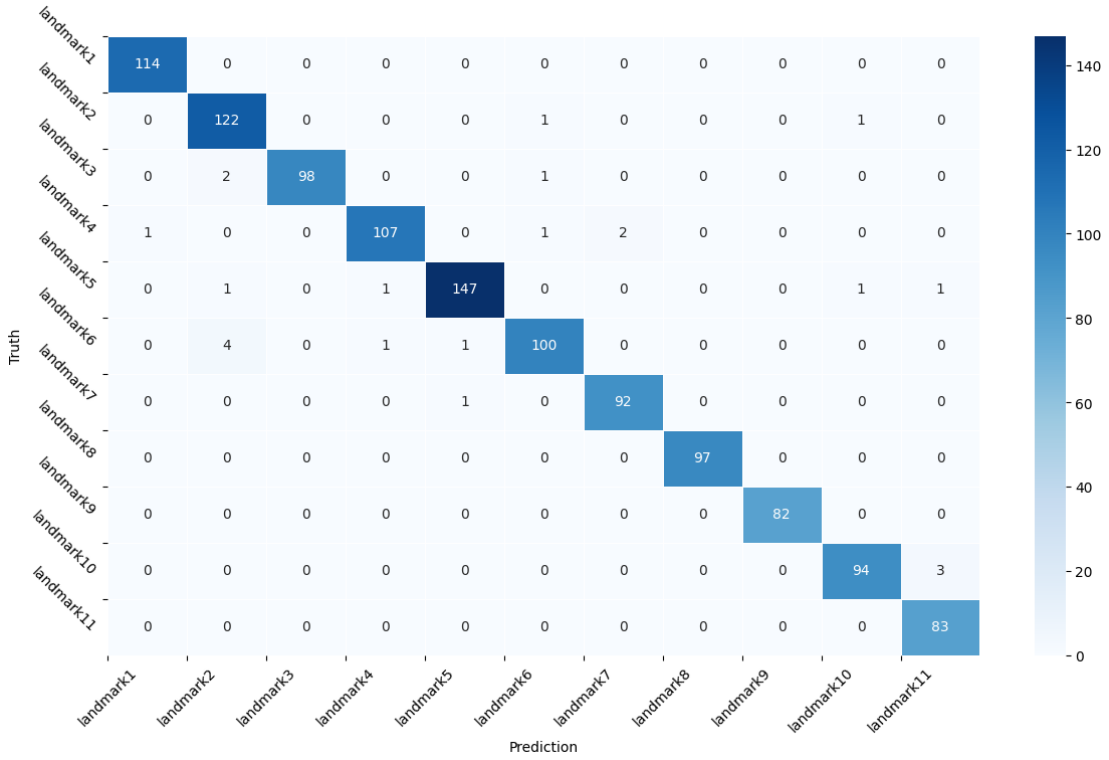


Figure 5-6: Confusion matrix for landmark identification

comparisons were conducted on other deep neural networks (CNN, LSTM, and LSTM-CNN without residual connections) with the same depth and training hyperparameters as shown in Table 5.2. Table 5.3 presents the obtained experimental results of accuracy, precision, recall, and F1-score using different networks. It can be seen that the proposed Bi-LSTM-CNN network achieved the highest performance in all four metrics thanks to the elaborately extracted spatial and temporal features. Therefore, the effectiveness of the proposed Bi-LSTM-CNN classification model for the landmark identification task is demonstrated with the experimental evaluation.

Trajectory Tracing

The path shown as the red line in Figure 5-3 is designed to evaluate the performance of PDR with landmark calibration, and the results are presented in Figure

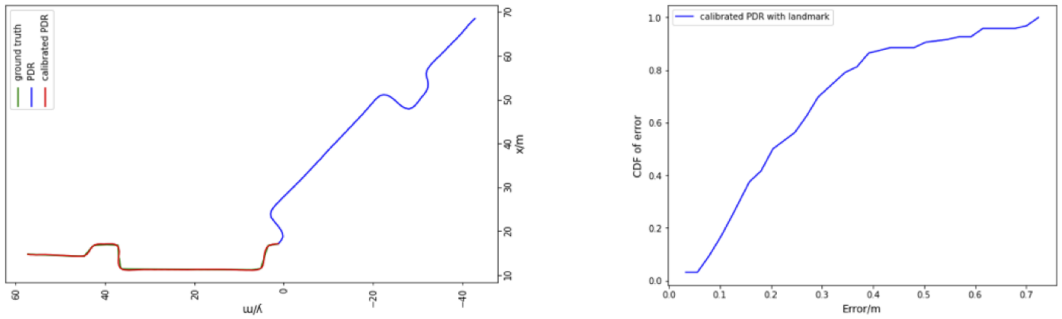
Table 5.3: Landmark identification performances of different models in the collected dataset

Method	Accuracy	Precision	Recall	F1-Score
CNN	0.9327	0.9404	0.9116	0.9258
LSTM	0.9637	0.9636	0.9600	0.9618
LSTM-CNN	0.9706	0.9750	0.9705	0.9727
Bi-LSTM-CNN (our)	0.9836	0.9849	0.9850	0.9849

5-7a. To quantitatively evaluate the positioning accuracy, we show the cumulative error distribution in Figure 5-7b. It can be seen from the left figure that the original PDR has an increasing error due to the initial wrong direction, although the information of many short segments can be described relatively accurately. Due to the significant error in the heading estimation without the landmark correction, the cumulative error distribution is not displayed in the right picture. The performance of PDR with the landmark calibration is well examined, nearly 80% of the positioning errors are less than 0.4 m, and the error probability within 0.7 m is higher than 90%. From the conducted experiments, the performance of the PDR-fused landmark calibration was evaluated with a lower positioning error, as compared to the PDR without calibration.

Virus Quanta Concentration

As mentioned above, we regard all virus particles exhaled every $0.1s$ during exhalation as virus instances. There will be many virus instances expelled during the entire movement of an infector. During the transmission of each instance, a uniform motion with a velocity of 3.9 $(m \cdot s^{-1})$ is maintained in the first second after exhalation, and the virus quanta are evenly distributed in the space that is



(a) (b)

Figure 5-7: The performance (**a**) and the cumulative error distribution (**b**) of the proposed landmark-calibrated PDR.

passed by. The initial number of quanta ($q_0 = 0$), the virus quanta emission rate (ER_q), and the removal rate of infectious viral-laden particles (RR_{iv}) are 142 and 1.37, respectively, and remain the same within the experiment [119, 19].

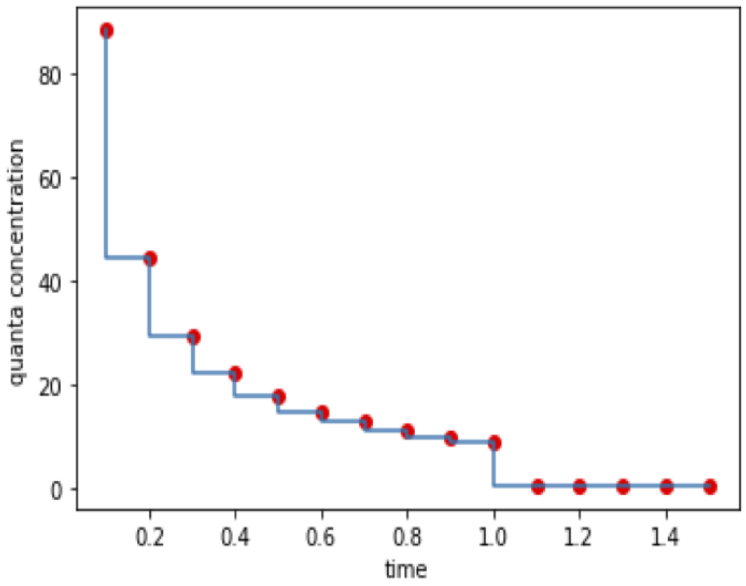
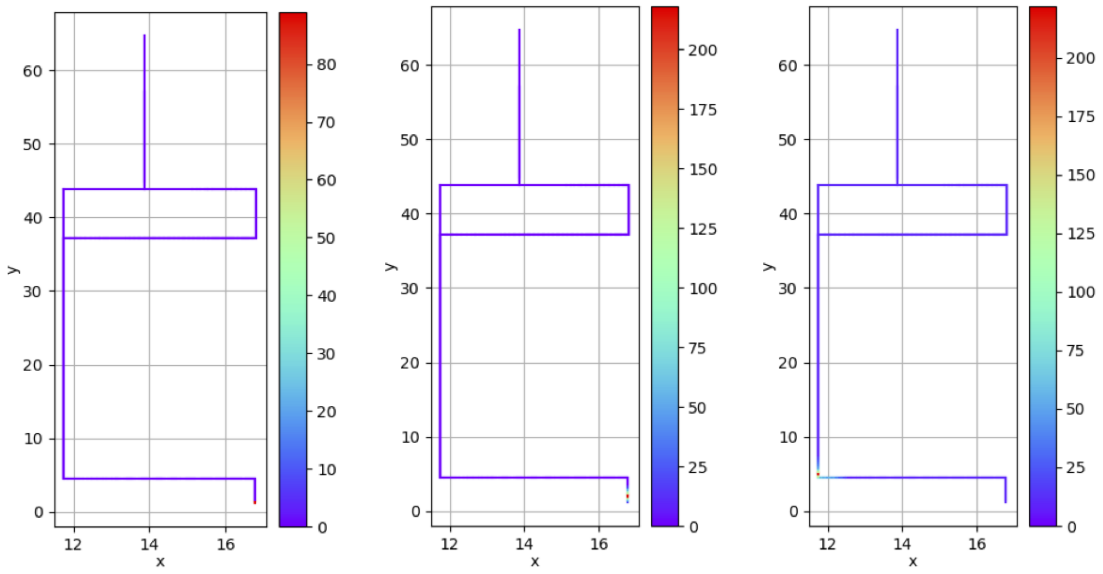


Figure 5-8: Quanta concentration of viral particles changes over time (first 1.5 s) after being released. Red points represent the instantaneous concentration at the end of each shorter interval

The virus-laden particles released in each interval follow the same moving pattern, leading to the same trend in the change of the quanta concentration. We

chose the instantaneous concentration at the end of each shorter interval with a length of $0.1s$ to represent the concentrations at all times during the entire interval, as presented in Figure 5-8. As can be seen in 5-8, the overall change in concentration presents an exponentially decreasing trend, from above 88 in the first interval (0~$0.1s$) to close to 0 one second later. The sharp decrease one second later is because of an instantaneous expansion of the viral aerosol coverage to the entire considered space.



(a) $t = 0s$ (b) $t = 0.5s$ (c) $t = 5s$

Figure 5-9: Indoor virus quanta concentrations at $t = 0s$ (**a**), $t = 0.5s$ (**b**), and $t = 5s$ (**c**), respectively, from the start of the movement. Virus quanta concentration is achieved by Equation 5.7, involving human movement along the directional path depicted in Figure 5-3 and the transmission of virus-laden particles.

The time when people started moving along the path, as illustrated by the directional red line depicted in Figure 5-3, can be regarded as the start ($t = 0s$) of the experiment. Figure 5-9 presents the virus concentration in the current environment at the time of $0s$, $0.5s$, and $5s$ from left to right (using lines to represent the considered corridor spaces), involving both human movement and virus-laden particles transmission. Among them, at $t = 0s$, only the virus concentration near

the point start can be seen to exceed 80, while most of the other parts are not covered by viral particles. At $t = 0.5s$, under the combined movements of virus droplets and humans, the relatively high quanta concentrations covered more. In addition, after another $0.5s$, the particles initially expelled at $t = 0s$ will spread to the overall space. At $t = 5s$, the area with higher quanta concentration gradually moves forward with the movement of people. Moreover, due to the accumulated particles that diffuse into the entire environment, the quanta concentration in the overall space is increased, gradually reaching a non-negligible level compared with the concentration of the newly expelled virus instance.

5.6 Limitations

Although the proposed iSTCA system realizes quantitative representation for exposed virus concentrations with the help of the landmark-calibrated PDR technique, there are some challenges that need to be overcome. First of all, there are some strict restrictions in the data acquisition process. The participants are required to hold the smartphone, specifically the Pixel 4a, at chest level. As a result, except for the diversity of users considered, other factors that affect the motion sensor readings are not seriously taken into account, such as the mobile device heterogeneity (e.g., different types or various vendors) and the device's status variation (e.g., putting in a pocket or handbag). In addition, a large amount of power of the smartphone is consumed during the indoor positioning process, resulting in the smartphone being overheated.

5.7 Conclusions

Technology-assisted virus exposure tracking approaches are increasingly being adopted to mitigate and tame the epidemic. In view of the complexity of quantifying virus exposure due to human movement and airborne dispersion of virus particles, we

propose iSTCA, a self-containing contact awareness approach that exploits PDR-based techniques. Quantitative information support directly concerned with risk assessment is provided for self-protection and epidemic control. More precisely, to reduce and calibrate the cumulative errors of trajectories based on landmarks, we apply Bi-LSTM and multi-head CNN with residual concatenation to long-term dependency in forward and backward directions and extract local correlations at various resolutions for landmark identification. The proposed method exploits the trajectories of people with viral-laden droplets exhaled and the transmission and attenuation of viruses in the air to quantify the virus quanta concentration in an indoor environment via spatiotemporal analytics for prevention and sanitization. Possible extensions to this work include the integration of wearable devices, such as smartwatches and smart bands, to replace mobile phones for power saving in indoor positioning. Moreover, we can apply the proposed techniques for the development of services in developing communities without reliable digital infrastructure.

Chapter 6

Personalized Federated Human Activity Recognition through Semi-Supervised Learning and Enhanced Representation

6.1 Introduction

The proliferation of mobile phones and wearable devices equipped with a variety of sensors enables the acquisition of large amounts of data on human movement and contextual information from many sources, thus providing the means to understand and predict human behavior at an unprecedented scale [129]. HAR refers to the identification of human physical activities in real-world settings, which has become increasingly essential in various applications such as smart homes, healthcare, and sports, and has drawn interest from both academia and industry [8]. To recognize activities, it is critical to extract and learn the features comprehensively to determine the relationship between sensed information and movement patterns. Both traditional HAR with hand-crafted features and deep learning models with

automatic feature extraction rely on centralized training, in which massive activity data scattered around multiple devices are shared or uploaded. However, motion data can reveal plenty of private information, such as daily behaviors and habits.

FL proposed by McMahan et al. can address the aforementioned privacy challenge in centralized approaches [91]. FL is a distributed framework that enables multiple parties to collaboratively train a shared global model by aggregating locally updated models without directly accessing users' data. Thus, participants can collectively benefit from the shared model without compromising their data privacy [150]. Most existing FL systems for HAR assume that the ground-truth of the data on clients is known to train supervised local models. However, in a practical scenario, manually and accurately annotating the time-series data for HAR is difficult. A completely unlabeled data on the client side is more reasonable, meanwhile, a small amount of data contributed by researchers or volunteers can be stored on the server side. A relevant approach to tackle the label scarcity problem is semi-supervised learning (SSL), in which a model is initialized with a small portion of labeled data and subsequently undergoes continuous updates utilizing unlabeled data [103]. Presotto et al. propose a hybrid method with FL for HAR that combines active learning for label query and label propagation to semi-annotate the local unlabeled data, while initial private data are needed for sample selection in active learning [103]. Zhao et al. take advantage of unsupervised representation learning to solve the label scarcity problems, in which an autoencoder is utilized for the feature learning of local models with unlabeled data and a global classifier is trained with labeled data on server side [164]. However, the challenging data heterogeneity of HAR in FL setting due to the different physical characteristics and various contextual information are not properly considered for clients [150]. The statistical diversity of local data would not be properly learned in existing FL paradigms with a single global model trained for capturing common features, resulting in performance degradation for clients in HAR. Thus, it is nec-

essary to fine-tune the shared global model on a particular client to learn unique information for personalized HAR.

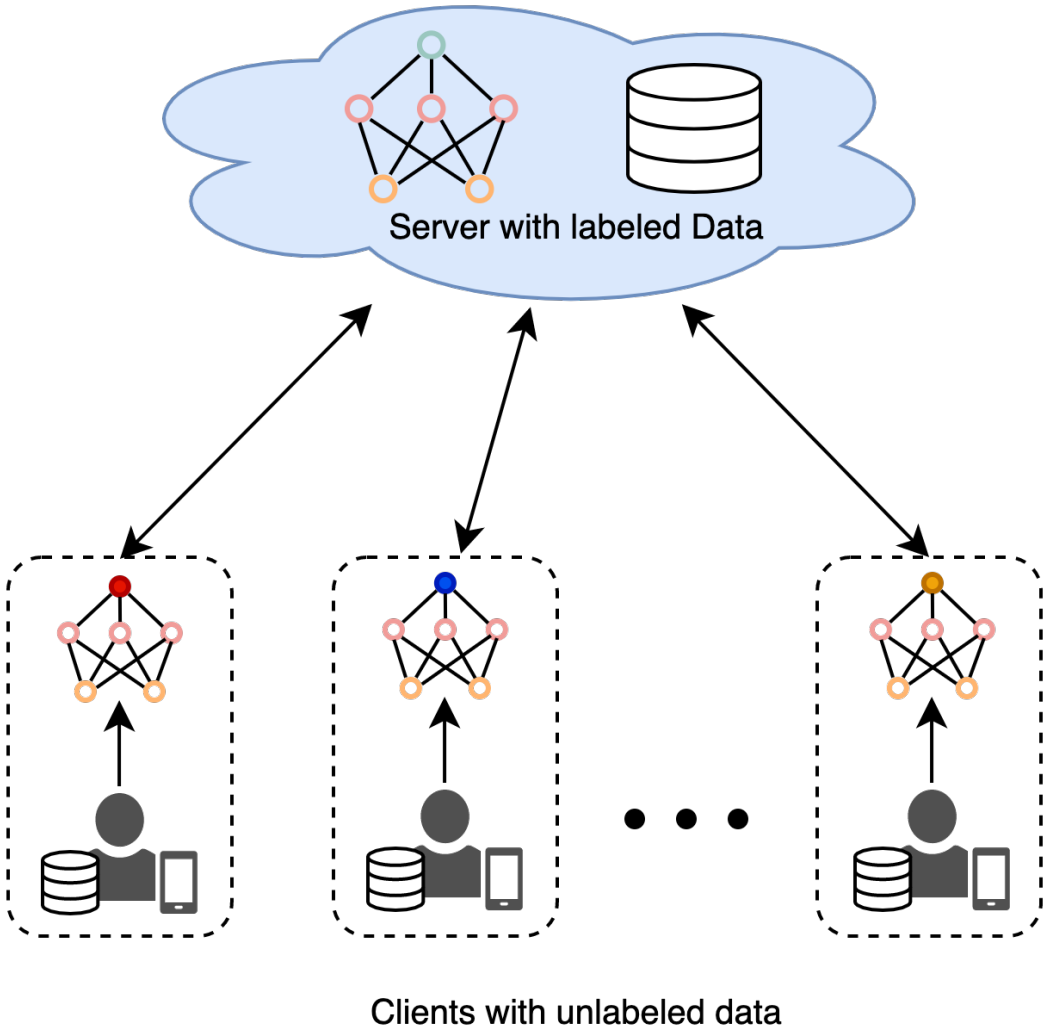


Figure 6-1: A resourceful server with labeled data and distributed clients with totally unlabeled data

To cope with the above issues, we propose a personalized federated HAR based on semi-supervised learning with enhanced representation learning, which involves a resourceful server with labeled data and distributed clients with totally unlabeled data, as illustrated in Figure 6-1. In our proposed design, a shared global model

is decoupled into a feature extraction model and a classification model, where the global deep representation learning model is jointly trained with the labeled data on the server side and the unlabeled data on clients, and the extracted features are fed into a separated classification network on the server side to train a global classifier for HAR, which is further personalized for each client in a semi-supervised fashion. Because neither the small amount of labeled data on the server side nor the locally sparse data of each client can individually train a general feature-extraction model, we design an autoencoder to take advantage of the abundant yet heterogeneous data residing on each side to learn an enhanced representation [71, 128]. After the entire global model training and sent to selected clients, we adopt the idea of pseudo-labeling to perform semi-supervised learning on private data for model personalization. However, training the network with falsely inferred pseudo-labels may degrade the performance of the personalized model [122]. To ensure the quality of produced pseudo labels, we select several models with the best server-side performance as auxiliary models. we keep the artificial labels only when these models assign very high confidence with low uncertainty to one possible class. With the integration of enhanced representation learning and federated semi-supervised learning, we conduct empirical experiments to evaluate the proposed approach on different HAR datasets, demonstrating the effectiveness of the proposed methods. In particular, the following contributions are made:

1. A novel personalized federated HAR technique based on semi-supervised learning is proposed, in which a global model is decoupled into representation learning and classification network. They are jointly trained with completely unlabeled data on client side and labeled data on server side in a semi-supervised fashion.

2. An unsupervised representation learning method in FL for HAR is proposed to learn enhanced representation from the heterogeneous data on the server and clients.

3. We propose a reliable pseudo-labeling approach for high-quality virtual label acquisition with the confidence and uncertainty considered in FL settings.

4. Extensive experiments are conducted using two real-world human activity datasets to demonstrate the effectiveness of the proposed methods in different practical FL scenarios.

The rest of the chapter is organized as follows. Section 6.2 reviews the related work. In section 6.3, we present the proposed methods. In section 6.4, we elaborate on the evaluation of proposed methods and discuss the experimental results. Lastly, we conclude the paper in section 6.6.

6.2 Related Work

6.2.1 Human Activity Recognition

DL based techniques have been extensively researched in HAR. The framework based on convolutional neural network (CNN) are studied to automatically extract the features for HAR [94, 25]. Besides, LSTM has the ability to preserve sequence information over time and capture long-term dependencies, so that it can extract temporal features in time-series sensor data for HAR [143, 140]. Both the CNN and LSTM are combined in [152] to identify different daily life activities. Nafea et al. employ the CNN and Bi-LSTM to automatically capture spatial and temporal features from sensor data to recognize human activities [95]. ConvAE-LSTM adopts CNN to extract features, which are further condensed by autoencoder, and LSTM to adapt at temporal modeling to realize the recognition of human activities from time-series sensor data [130].

6.2.2 Federated Learning for HAR

FL was proposed as an alternative to traditional centralized DL, in which a global model can be trained collaboratively with multiple clients coordinated by a server [90, 78]. FL has been widely adopted in many applications, including HAR [74, 125]. Sozinov et al. proposed an FL based HAR system and its performance is sufficiently robust, realizing an acceptable accuracy comparable to centralized learning [125]. Zhao et al. proposed an FL based HAR system for activity and health monitoring [163]. PMF was proposed as a privacy-persevering mobility prediction framework based on FL, in which a personal adapter was designed for local personalization to further improve the accuracy [37]. Furthermore, given that personalization is a critical aspect of HAR, relevant studies have demonstrated that, in FL, fine-tuning the global model for each client significantly enhances the recognition rate [23, 151].

6.2.3 Semi-supervised Federated Learning

To address the label scarcity problem in practical FL, semi-supervised learning is extensively studied. Semi-supervised learning combines both supervised learning and unsupervised learning to tackle the model training with limited labeled data [30]. Generating pseudo labels for unlabeled data to construct labeled data for model training in a supervised fashion has been extensively studied [30, 84, 79]. For instance, Diao et al. propose SemiFL to combining the communication-efficient FL with semi-supervised learning, in which clients have completely unlabeled data, while the server has a small amount of labeled data. They design an alternate training to ensure the accuracy of pseudo labels, that is, "fine-tune global model with labeled data" and "generate pseudo-labels with the global model" [30]. Liu et al. use labeled data on an FL server to train a model through supervised learning and then send this model to FL clients to generate labels on their local data [84]. Another direction of semi-supervised FL is to perform unsupervised learning on

clients to derive robust feature representation learning from the unlabeled data instead of producing pseudo labels, in which the global feature representation is then used to build downstream tasks with a limited amount of labeled data. For example, van Berlo et al. introduce federated unsupervised representation to pre-train the model using unlabeled data to solve the label scarcity problem [135]. Zhao et al. proposed a semi-supervised FL framework for HAR, in which clients conduct unsupervised learning on autoencoders with unlabeled local data to learn general representations and a server conducts supervised learning on an LSTM classifier with labelled data [164]. However, personalization is not properly considered in these works. Recently, FedHAR was proposed to build the global model by aggregating the computed unsupervised and supervised gradients, in which a small number of clients own labeled data, and a large number of clients only have unlabeled data [160]. FedAR, assuming that there are labeled data in each client, combines active learning and label propagation to semi-automatically annotate the local streams of unlabeled sensor data [103]. Different from these works, we do not rely on the assumption of clients possess labeled data. For the more realistic case of storing only limited labeled data contributed by researchers or volunteers on the server, we propose a practical solution to continuously improve model performance by integrating representation learning and pseudo-labeling.

6.3 Methodology

6.3.1 Preliminaries

In this work, we focus on semi-supervised FL with completely unlabeled clients. Let $x_{u,m}$ represent the unsupervised data at client $m = 1, 2, 3 \dots M$, where M is the number of clients. The small amount of labeled dataset on server is $\{x_l,\ y_l\}$. We decouple the deep neural network W into the representation layers, denoted by W_a and the final decision layer, denoted by W_{cls}. Thus, the server model is parameter-

ized to $W_g^t = \left(W_{a_g}^t, W_{cls_g}^t\right)$ at $t-th$ communication round and similarly, the client's model is parameterized to $\{\left(W_{a_1}{}^t, W_{cls_1}{}^t\right), \left(W_{a_2}{}^t, W_{cls_2}{}^t\right), \ldots, (W_{a_M}{}^t, W_{cls_M}{}^t)\}$.

6.3.2 System Design

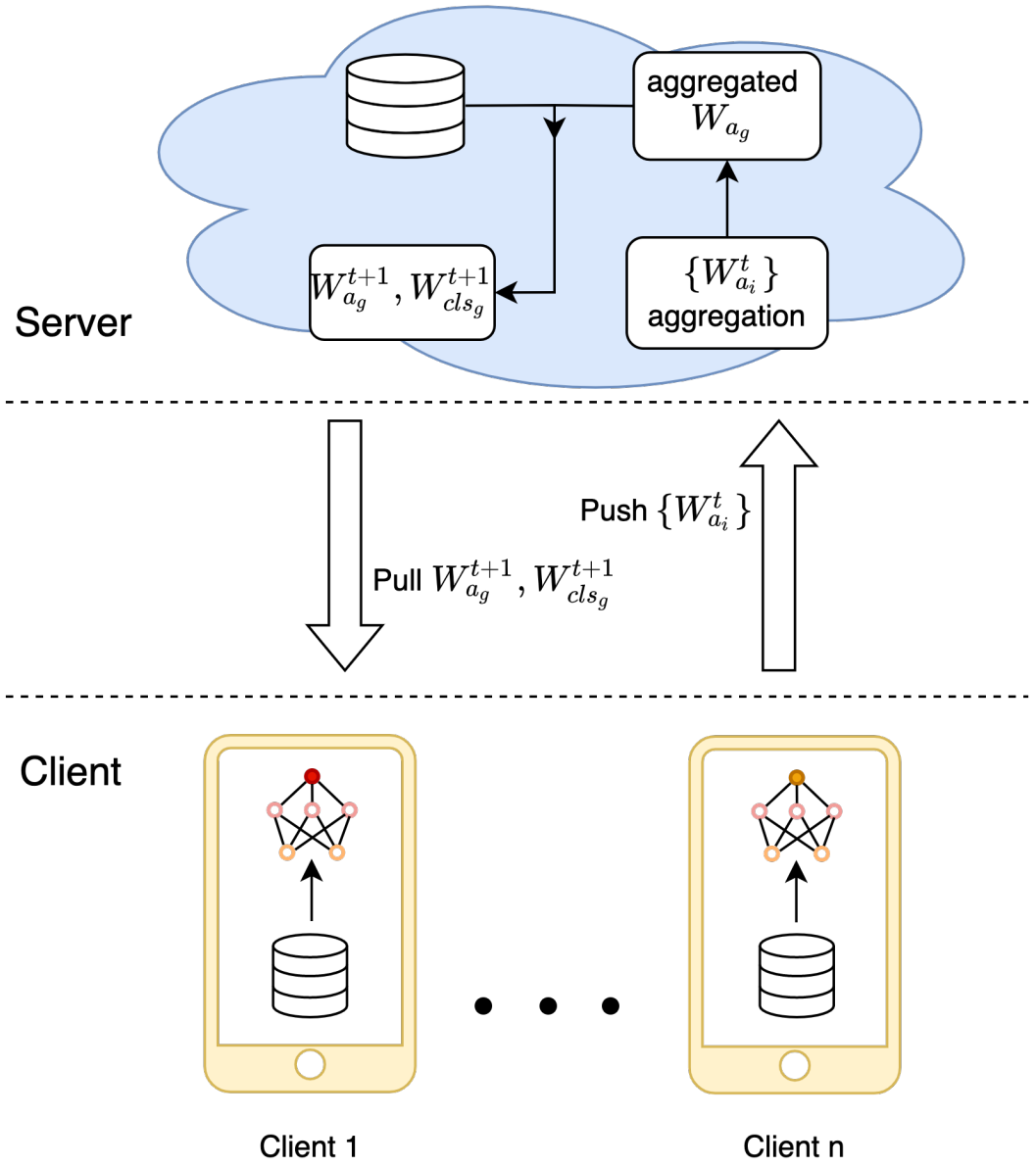


Figure 6-2: Overview of proposed scheme

An overview of the proposed scheme is presented in Figure 6-2. Following the FL paradigm, the server periodically sends the weights of a global activity recognition model to the available clients for model updating. The same deep neural network structure is established on the server and clients, with representation layers and classification layers for HAR. The global server model is initialized with the limited labeled data on the server, while the network on each client side is implemented with a semi-supervised learning strategy to personalize the server model. In each communication rounds t, the global model W_g^t, which has already been fine-tuned with the server-side labeled data, is sent to selected clients for further updating with. After receiving the global model, each client would perform unsupervised learning for representation learning and then, send the updated representation model to server. Server average the uploaded parameters from active clients to obtain a new representation layer, which is further trained with $W_{cls_g}^t$ to learn a new complete model W_g^{t+1} on server for following communications. Each client can further learn a personalized model by pseudo labeling in which both confidence and uncertainty are considered in selecting the right samples with several global models with better performance in history.

6.3.3 Representation Learning

The universal features extracted from inputs with the representation part of DL model significantly affect the performance of the learning tasks of interest [99]. In practical scenarios, there are many unlabeled data collected by different users, and thus, to comprehensively utilize the target data, unsupervised feature extraction has been proposed and studied. Autoencoder relies on reconstructing unlabeled data using unsupervised neural networks. The data is encoded to its latent representation as the extracted features, and then, the decoder part tries to reconstruct the original data from the learned representation. To facilitate the feature design and selection for HAR, various autoencoders are proposed, such

as villa autoencoder [130] and stacked autoencoder [94]. Moreover, van Berlo et al. show promising potential of using autoencoders to implement semi-supervised FL and Zhao et al. apply autoencoders to unlabeled local data to learn general representations for HAR [164]. Therefore, in order to address the lack of labels on clients in HAR with sensory data and achieve a strong generalization ability while leveraging heterogeneous data, our proposed system applies semi-supervised learning in an FL system, in which server and clients use unsupervised learning to jointly train autoencoders with their own data in an unsupervised fashion for further recognition, and the server uses supervised learning to train a classifier that can map encoded representations to activities with a labeled dataset.

6.3.4 Personalization with Uncertainty-Aware Pseudo Labeling

As the shared global model is fine-tuned with a small amount of labeled samples, it is not adaptive for all clients due to the possible heterogeneity between local data on client side and labeled data on server side. Thus, the server model with general representation and shared classifier is personalized for a specific person. To obtain the information of local data as much as possible for model personalization, the pseudo labels for some samples are inferred. Typically, for each unlabeled data sample, it would pick the class with maximum predicted confidence as prediction to be used as a pseudo label. For sample x^i, y_k^i=1 denotes that class k is present in the corresponding input as a potential label and $y_k^i = 0$ represent the class's absence. We set the p_k^i represents the probability of class k being present in the samples. With this prediction probability, the pseudo-label can be generated for x^i as:

$$\widetilde{y}_k^i = \mathbb{1}\ [p_k^i \geq \gamma] \tag{6.1}$$

where $\gamma \in (0, 1)$ is a threshold used to produce hard labels and $\mathbb{1}[\bullet]$ produce one-hot labels with given values.

In our setting, the criteria used to select how many and which samples are transferred from unlabeled data to the labeled data during training at each round is key to our method. When starting to personalize, the client would pull several best models in history on a server to achieve some relatively well-performed models for conducting pseudo labeling. Then, the pseudo-labels would be generated based on the agreements and uncertainty of the model prediction [79, 105]. Let $\mathbf{g}^i = [g_1^i,\ g_2^i,\ \ldots, g_K^i] \subseteq \{0, 1\}^K$ be a binary vector representing the selected pseudo-labels in sample x^i, where $g_k^i = 1$ when $\widetilde{y}_k^i$ is selected and $g_k^i = 0$ when $\widetilde{y}_k^i$ is not selected. This vector is obtained as follows:

$$\widetilde{g}_k^i = \mathbb{1}\ \left[p_k^i \geq \gamma\right] + \mathbb{1}\ \left[u(p_k^i) \leq \sigma\right] \tag{6.2}$$

where $u(p)$ is the uncertainty of a prediction p, and σ are the uncertainty thresholds. The uncertainty part ensures the network prediction is sufficiently certain to be selected. Therefore, to maintain stability in each model, we retain artificial labels whose largest class probability is above a predefined threshold γ_t for communication round t , and the uncertainty of corresponding class probability should lower than a preset threshold σ_t. Since the imbalance of training would degrade the model performance, we also made an adjustment with class balance for sample selection.

6.3.5 The Holistic Algorithm

To sum up, the holistic mechanism of the proposed method is presented in Algorithm 2, including the personalization methods. In each communication round, the server selects part of clients to join the representation learning. An unsupervised representation learning on autoencoder part is conducted before sending to

the server. Server would aggregate the uploaded autoencoder to learn with unlabeled data and then, the overall model is fine-tuned with labeled data stored on server side. Each client can further learn a personalized model by pseudo labeling, in which both confidence and uncertainty are considered in selecting the right samples.

6.4 Experimental Evaluation

6.4.1 Dataset Description

Two widely used real-world datasets: HAR-UCI and PAMAP2 are adopted to conduct the experiments.

HAR-UCI This standard dataset was collected from the sensor recording of 30 individuals performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors. Those involved 30 participants within an age range of 19-48 years. Each person had a smartphone (Samsung Galaxy S II) strapped to their waist, and perform six activities: walking, walking upstairs, walking downstairs, sitting, standing, laying. The built-in sensors, including accelerometer and gyroscope, are used to capture 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz, which were further processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 seconds (128 readings).

PAMAP2 This dataset is a public dataset for human physical activity monitoring. It was collected from 9 different subjects with 3 IMUs positioned in hand, chest and ankle and a heart rate monitor. Each IMUs has three embedded sensors: a 3-axis accelerometer, a 3-axis gyroscope and a 3-axis magnetometer, to record the corresponding sensor data at constant rate of 100Hz. Altogether, over 10 hours

Algorithm 2 Personalized federated human activity recognition through semi-supervised learning and enhanced representation

Input: Labeled data $\{(x_l,\ y_l)\}$, Unlabeled data $\{x_{u,1:M}\}$ distributed on M local clients, server model parameterized by $W_g = \left(W_{a_g}, W_{cls_g}\right)$, client models parameterized by $\{W_{1:M}\} = \{(W_{a1:M},\ W_{cls1:M})\}$, local epochs E_c, global epochs E_g

Initialize $W^0_{a_g},\ W^0_{cls_g}$ with labeled data $\{(x_l, y_l)\}$ on server;
for each communication round t **do**
    randomly select C part of the clients M_c;
    for each client $m \in M_c$ **do**
        Distribute server model parameters to local client
        $W^{t+1}_{a_m} \leftarrow$ **ClientUpdate**$(x_{u,m},\ W^t_{a_m}, W^t_{cls_m})$
    end for
    $W^{t+1}_{a_g} \leftarrow \sum_{m \in M_c} \frac{n_m}{n} W^{t+1}_{a_m}$
    Fine-tune global model $W^{t+1}_{a_g}$ and $W^{t+1}_{cls_g}$ on $\{x_l, y_l\}$
end for
ClientUpdate$(x_{u,m},\ W^t_{a_m})$:
for each local epoch $e \in E_c$ **do**
    train $W^t_{a_m}$ using mean squared error (MSE) loss
end for
return $W^t_{a_m}$ and send it to server
ClientPersonalization$(x_{u,m},\ W_{a_m}, W_{cls_m})$:
Pull the best b models in history from server as $\{W^{pseudo}\}$
for each personalization iterations t_p **do**
    Pseudo labeling $x_{u,m}$ using $\{W^{pseudo}\}$
    $x_{selected} \leftarrow$ select pseudo label using (2)
    for each local epoch $e \in E_c$ **do**
        Train (W_{am}, W_{cls_m}) with $x_{selected}$ and cross-entropy loss
    end for
    Add (W_{am}, W_{cls_m}) to $\{W^{pseudo}\}$
end for

of data were collected, from which nearly 8 hours were labeled as one of 18 preset activities (12 activities selected) performed during data collection. The raw signals of each person were segmented in a sliding window with fix-length of 2.56 seconds (256 readings) and 50% overlap.

6.4.2 Experiment Setup

To evaluate our proposed method, we conduct experiments with various settings where we randomly attempt different numbers of labeled samples for supervised training on the server to update the global model. More concretely, the amount of the data with annotation stored on the server is 200, 300, 500 for both datasets. In each communication, the number of randomly selected clients is 6 for HAR-UCI dataset and 2 for PAMAP2 dataset. The global epochs and local epochs for model training in each communication round are both equal to 5. We adopt Stochastic gradient descent (SGD) as the optimizer by default where the learning rate is 0.05, decreased by a factor of 0.4 for every 20 FL rounds, the momentum factor is 0.9 and L_2 regularization is 0.001 to prevent over-fitting. Also, the batch size is 32 for model training and the size of learned representation is 256 by default.

Models

Our HAR model consists of an autoencoder and a classifier. More specifically, two 1-dimensional convolutional layers have 32 and 128 output channels with the same kernel size of 9 and stride length of 2, respectively. The output is fed in a ReLU layer before computing by a Bi-LSTM layer with a hidden size of 128 and an activation of Tanh. Then, the output is flattened to a linear layer, that transforms it into a specific size. For the decoder part, the output of another linear layer is transformed into suitable shape for the input of the following Bi-LSTM layer with an activation of Tanh. Two transposed 1-dimensional convolutional continue to decode the output with input channels of 128 and 32, kernel size of 9 and 9,

stride length of 2 and 2, and output padding size of 1 and 1, respectively. Finally, the output of the decoder is reshaped to the same construction of input data. Besides, the classifier is constructed with a linear layer whose input is the learned representation and the output is the number of activities.

Baselines and Metrics

We evaluated the following schemes to demonstrate the effectiveness of our proposed approach.

Fully Supervised All the samples with label are used to train a fully supervised model for the measurement of ideal accuracy we could hope to obtain with FL settings.

Partially Supervised Since we assume the labeled data only exist on the server side, a baseline method that only uses these labeled data to conduct supervised training is designed.

Semi-supervised FL with Autoencoder (SSFLAE [164]) The clients conduct unsupervised learning on autoencoders with unlabeled local data to learn general representations, and a server conducts supervised learning to only train the activity classifier, after the representation part is aggregated.

SSFLAE with joint fine-tuning (SSFLAE-FT) Since there are representation part and the classifier part in SSFLAE, instead of frozen the encoder part with classifier training, an end-to-end training fashion is designed, in which the representation part and the classifier is jointly fine-tuned.

Pseudo Label To personalize the global model, a pseudo label with high prediction confidence would be set to conduct supervised model training in an end-to-end fashion.

The averaged prediction accuracy on the test data of each client is applied to

evaluate the performance of each model, which is defined as:

$$Acc = \frac{1}{M}\sum_{i=1}^{M} acc_i \tag{6.3}$$

6.5 Results

Table 6.1 and 6.2 denote the comparison result with baseline approaches on two datasets with different number of labeled data in server, respectively. Generally, the proposed methods can achieve the best prediction performance under each setting on both two datasets. Although the labeled samples are the same in different datasets, the overall performance is slightly lower in the same settings. It is because the clients in PAMAP2 dataset are much more heterogeneous than HAR-UCI and more diversified physical activities are performed for each involved person.

6.5.1 Evaluation of Unsupervised Representation Learning

Comparing the SSFLAE with other FL methods, we can see there is a great performance degradation from partially supervised FL in terms of the activity prediction accuracy on these two datasets. However, SSFLAE-FT can achieve a better prediction accuracy, outperforming the partially supervised FL. In combination of these two observations, the reason for the decline in the performance of SSFLAE may be that we have not searched for the optimal hyperparameters, such as the learning rate and the momentum for different optimizer, and the fine-tune strategy is also not good. Actually, van Berlo et al. pointed out that the fine-tuning strategy of SSFLAE can result in bad test set performance, and the unsupervised representation learning is recommended to executed to obtain the information of a more diverse number of users with bigger set of activities compared with the small amount of labeled dataset for HAR [135]. We demonstrate the effectiveness of unsupervised

representation learning with autoencoder in FL settings, in which leveraging the massive amount of unlabeled data samples can improve the performance. Partially supervised FL only exploring the small amount of labeled samples result in poor generalization and personalization for specific clients, especially for PAMAP2. Besides, we found that as the number of labeled data stored in the server increases, the performance gain brought by unsupervised representation learning becomes smaller. With more and more labeled samples, the statistical characteristics of labeled dataset are getting closer and closer to the overall dataset, and thus, the larger the amount of annotations selected, basically the smaller the improvement in the prediction accuracy of the model with representation learning.

Table 6.1: Prediction results on HAR-UCI

Dataset		**HAR-UCI**		
Number of Labeled samples		**200**	**300**	**500**
Model	Fully supervised		96.69	
	Partially supervised	78.41	81.88	86.03
	SSFLAE	26.35	29.56	48.07
	SSFLAE-FT	80.72	82.51	86.80
	Pseudo Label	76.89	79.47	82.16
	Ours	82.31	84.25	89.27

6.5.2 Evaluation of Personalization

To personalize the shared server model, we take the annotations produced by traditional confidence-based pseudo labeling techniques and our uncertainty-based

Table 6.2: Prediction results on PAMAP2

Dataset		**PAMAP2**		
Number of Labeled samples		**200**	**300**	**500**
Model	Fully supervised		95.04	
	Partially supervised	75.43	80.95	84.61
	SSFLAE	20.67	23.68	25.13
	SSFLAE-FT	77.78	81.43	86.46
	Pseudo Label	69.50	76.82	79.51
	Ours	77.85	81.59	87.62

algorithms. Generally, our proposed semi-supervised approaches can improve the performance of global model on specific client, which demonstrate the effectiveness of proposed personalization methods with uncertainty-aware based pseudo labeling. Regardless of the method, we can see that better initial models lead to better personalization accuracy, especially for **HAR-UCI**. However, noting that the confidence-based technique degraded the prediction accuracy to varying degrees, especially for **PAMAP2** with strong heterogeneity. It is due to the fact that when an incorrect prediction is used for model training, the error accumulates and adversely affects the prediction accuracy of the personalized model. Compared to confidence-based techniques, we use several global models with better performance to annotate the unlabeled dataset, taking into account both the accuracy and the uncertainty of the prediction to guarantee the label accuracy. Moreover, the activity patterns for some clients in **PAMAP2** are quite different from those of other clients, in which the initially personalized model is received from a server trained with a small number of labeled samples, resulting in incomplete learning

for various activity patterns. Because the model only learns a part of the activity class, more similar samples will be selected when generating the pseudo labels with high confidence, leading to a great performance decline. Thus, the proposed pseudo labeling method with uncertainty awareness and class balance can reach a high personalization accuracy.

6.6 Conclusion

The challenge of acquiring labeled data from end users constrains the FL applications for HAR in real-world. This paper proposes a federated HAR through semi-supervised learning and enhanced representation to address the issue of totally unlabeled clients. By employing autoencoders trained in an unsupervised fashion on clients and a classifier trained via supervised learning on the server, our system attains enhanced accuracy for HAR, which is comparable to that of a supervised FL system but does not require any locally labeled data. Due to the different physical characteristics and various contextual information, after receiving the shared model, further personalization is performed locally. Pseudo-labeling techniques are adopted to produce the annotations, in which, the confidence and the uncertainty are considered at the same time and the selection is made as balanced as possible to set the pseudo label for samples. Comprehensive experimental analysis indicates that our personalized federated HAR with semi-supervised learning and enhanced representation learning approach substantially enhances the performance of a labeled server in conjunction with unlabeled clients, achieving a competitive accuracy with a personalization model.

Chapter 7

Conclusion

This thesis focused on studying efficient crowdsensing-based indoor localization systems and their applications.

To acquire localization in GNSS-denied indoor environments, various approaches using data collected from different sensors and wireless transmitters have been proposed. However, it is essential to develop cost-effective positioning techniques that take deployment costs, including pre-configured infrastructure cost, data collection cost and data processing and learning cost, into account for efficiently providing location services in indoor spaces. A crowdsensing-based indoor localization framework is proposed to achieve cost-effective Wi-Fi fingerprinting-based indoor localization systems, in which active learning is introduced to realize deployment cost reduction on data collection and data annotation, and machine learning approaches are applied to learn robust features for precise location inference. Subsequently, to achieve efficient and accurate indoor localization services in a variety of spaces, regardless of the deployment of wireless infrastructure, we develop an efficient indoor localization system using PDR with automatically identified activity landmarks calibrated. Furthermore, we integrate the location information obtained through this method with a spatiotemporal model of virus concentration changes for comprehensive contact tracing of Covid-19, and its effectiveness

is demonstrated via extensive experiments. Additionally, an extensive study on crowdsensing-based activity landmark recognition using semi-supervised and federated learning is conducted to meet the reality, where the general representation model is achieved jointly by the server and the clients, and the personalized models are acquired with totally unlabeled local data.

In Chapter 3, the crowdsensing technique is applied to reduce the cost of labor-intensive and time-consuming fingerprint collection for radio map construction in popular Wi-Fi fingerprinting-based indoor localization systems with Wi-Fi signal covered. Although crowdsensing could solve the problem of radio signature collection, there are various uncertainties about the location annotations contributed by the crowd, which would affect the performance of the localization model. To address such issues and realize efficient indoor localization systems based on Wi-Fi fingerprints, we propose a crowdsensing-based indoor localization framework, ALCIL. ALCIL employs the active learning technique to collect informative data for performance improvement of the localization model under a certain cost and applies the machine learning approaches to learn the strong patterns heuristically for accurate location estimation. We conduct extensive experiments to demonstrate the effectiveness of the proposed framework, reducing 70% of the data label acquisition cost and locating users' mobile devices efficiently at the given fixed budget.

In Chapter 4, we develop an activity landmark-based PDR to realize efficient indoor navigation with high precision in various environments regardless of the configuration status of wireless access points. The activity landmark, which stands for a location point that imposes a certain pattern on the motion sensor readings, is properly recognized with the sensors inside smartphones and applied to the calibration of the accumulated errors in PDR at no extra cost. To further reduce the deployment cost of data processing, the feature of activity landmarks is extracted by an unsupervised feature learning method without manual calculation,

producing a compact representation for landmark identification. The validity of the proposed approach to provide efficient indoor navigation was verified with 9271 samples collected using a crowdsensing technique for 27 landmarks in the context of practical buildings.

In Chapter 5, the above-mentioned activity landmark-based PDR was utilized to develop an indoor spatiotemporal contact awareness framework (iSTCA) to enable precise contact tracing of Covid-19 in various indoor environments. The iSTCA explicitly considers the self-containing quantitative contact analytics approach with spatiotemporal information to provide accurate awareness of the virus quanta concentration in different areas at various times. PDR technique is employed to precisely detect the locations and trajectories for distance estimation and time assessment with recognized activity landmarks using a designed deep learning model. Thus, the contact-tracing feature within this framework allows for a more detailed and quantitative understanding of indoor exposure to virus with virus lifespan considered, which is difficult to discern using conventional techniques based on relative distances between devices. Furthermore, the integration of activity landmark-based PDR positioning methods with a spatiotemporal model of virus concentration variations has led to the development of a cost-effective solution that can be employed in diverse indoor environments, even those lacking ICT infrastructure. We perform an evaluation using actual movement history data collected within a practical building and confirmed the effectiveness of the developed system. Particularly, the PDR-based location estimation error is reduced to $0.7m$ with high confidence under the calibration of an automatic landmark identification model with an accuracy of 98.36%. Thus, the virus concentration in indoor spaces is achieved more accurately, confirming the feasibility of fine-grained contact tracing.

In Chapter 6, we conduct an extensive study on crowdsensing-based activity landmark recognition to make the activity landmark-based indoor localization sys-

tem more efficient. We propose a method based on semi-supervised learning and federated learning (FL) that can perform personalized human activity recognition (HAR) while considering user privacy and cost reduction on data annotation, in which FL clients have completely unlabeled data, while the FL server has a small amount of labeled data contributed by volunteers. Since a general model in the server is not suitable for each client due to the different physical characteristics and different contextual information in crowdsensing-based scenarios, a specific model is learned for each user to reduce the effects of data heterogeneity for the achievement of better system performance. The proposed approach is characterized by collaborative semi-supervised learning conducted between client devices and a server for the general model training, using pseudo-labels that take into account trustworthiness and uncertainties for model personalization. This method can be applied to efficiently identify activity landmarks in indoor localization techniques. Through evaluation experiments with two different real-world activity recognition datasets, the persuasive accuracy improvement is confirmed compared to conventional techniques.

Bibliography

[1] Heba Abdelnasser, Reham Mohamed, Ahmed Elgohary, Moustafa Farid Alzantot, He Wang, Souvik Sen, Romit Roy Choudhury, and Moustafa Youssef. Semanticslam: Using environment landmarks for unsupervised indoor localization. *IEEE Transactions on Mobile Computing*, 15(7):1770–1782, 2015.

[2] Osamah Ali Abdullah and Ikhlas Abdel-Qader. Machine learning algorithm for wireless indoor localization. *Machine Learning-Advanced Techniques and Emerging Applications*, 2018.

[3] Osamah Ali Abdullah, Ikhlas Abdel-Qader, and B Bazuin. A probability neural network-Jensen-Shannon divergence for a fingerprint based localization. In *2016 Annual Conference on Information Science and Systems (CISS*, pages 286–291, Princeton, NJ, USA, 2016.

[4] R. Abishek, K.R. Abishek, N. Hariharan, M.Rakesh Vaideeswaran, and C.Sundara Paripooranan. Analysis of machine learning algorithms for wi-fi-based indoor positioning system. In *2019 TEQIP III Sponsored International Conference on Microwave Integrated Circuits, Photonics and Wireless Networks (IMICPW), Tiruchirappalli*, pages 218–222, India, 2019.

[5] Nadeem Ahmed, Regio A Michelin, Wanli Xue, Sushmita Ruj, Robert Malaney, Salil S Kanhere, Aruna Seneviratne, Wen Hu, Helge Janicke, and Sanjay K Jha. A survey of covid-19 contact tracing apps. *IEEE access*, 8:134577–134601, 2020.

[6] Fahad Alhomayani and Mohammad H. Mahoor. Deep learning methods for fingerprint-based indoor positioning: A review. *Journal of Location Based Services*, 14(3):129–200, 2020.

[7] Uzoma Rita Alo, Friday Onwe Nkwo, Henry Friday Nweke, Ifeanyi Isaiah Achi, and Henry Anayo Okemiri. Non-pharmaceutical interventions against COVID-19 pandemic: Review of contact tracing and social distancing technologies, protocols, apps, security and open research directions. *Sensors*, 22(1):280, 2021.

[8] Muhammad Haseeb Arshad, Muhammad Bilal, and Abdullah Gani. Human activity recognition: Review, taxonomy and open challenges. *Sensors*, 22(17), January 2022.

[9] Paramvir Bahl and Venkata N Padmanabhan. Radar: An in-building RF-based user location and tracking system. In *Proceedings of the Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2000), Tel Aviv*, volume 2, pages 775–784, Israel, March 2000.

[10] Jason Bay, Joel Kek, Alvin Tan, Chai Sheng Hau, Lai Yongquan, Janice Tan, and Tang Anh Quy. BlueTrace: A privacy-preserving protocol for community-driven contact tracing across borders. *Government Technology Agency-Singapore, Tech. Rep*, 18:1, 2020.

[11] Martin Z Bazant and John WM Bush. A guideline to limit indoor airborne transmission of COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, 118(17):e2018995118, 2021.

[12] Jürgen Bernard, Matthias Zeppelzauer, Markus Lehmann, Martin Müller, and Michael Sedlmair. Towards user-centered active learning algorithms. *Computer Graphics Forum*, 37(3):121–132, 2018.

[13] Bimal Bhattarai, Rohan Kumar Yadav, Hui-Seon Gang, and Jae-Young Pyun. Geomagnetic field based indoor landmark classification using deep learning. *IEEE Access*, 7:33943–33956, 2019.

[14] Andreas Biri, Neal Jackson, Lothar Thiele, Pat Pannuto, and Prabal Dutta. SociTrack: Infrastructure-free interaction tracking through mobile sensor networks. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, pages 1–14, London, United Kingdom, September 2020.

[15] Daren C Brabham. Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, 14(1):75–90, 2008.

[16] Isobel Braithwaite, Thomas Callender, Miriam Bullock, and Robert W Aldridge. Automated and partly automated contact tracing: a systematic review to inform the control of covid-19. *The Lancet Digital Health*, 2(11):e607–e621, 2020.

[17] Ramon F Brena, Juan Pablo García-Vázquez, Carlos E Galván-Tejada, David Muñoz-Rodriguez, Cesar Vargas-Rosales, and James Fangmeyer. Evolution of Indoor Positioning Technologies: A Survey. *Journal of Sensors*, 2017:21, 2017.

[18] Harald Brüssow. COVID-19: Omicron the latest, the least virulent, but probably not the last variant of concern of SARS-CoV-2. *Microbial Biotechnology*, 15(7):1927–1939, 2022.

[19] Giorgio Buonanno, Luca Stabile, and Lidia Morawska. Estimation of airborne viral emission: Quanta emission rate of SARS-CoV-2 for infection risk assessment. *Environment international*, 141:105794, 2020.

[20] Luca Calderoni, Matteo Ferrara, Annalisa Franco, and Dario Maio. Indoor localization in a hospital environment using Random Forest classifiers. *Expert Systems with Applications*, 42(1):125–134, 2015.

[21] Claude Castelluccia, Nataliia Bielova, Antoine Boutet, Mathieu Cunche, Cédric Lauradoux, Daniel Le Métayer, and Vincent Roca. ROBust and privacy-PresERving proximity tracing. working paper or preprint, May 2020.

[22] Lina Chen, Jinbin Wu, and Chen Yang. Meshmap: A magnetic field-based indoor navigation system with crowdsourcing support. *IEEE Access*, 8:39959–39970, 2020.

[23] Yiqiang Chen, Xin Qin, Jindong Wang, Chaohui Yu, and Wen Gao. FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare. *IEEE Intelligent Systems*, 35(4):83–93, July 2020.

[24] Joana Costa, Catarina Silva, Mário Antunes, and Bernardete Ribeiro. On using crowdsourcing and active learning to improve classification performance. In *2011 11th International Conference on Intelligent Systems Design and Applications*, pages 469–474. IEEE, 2011.

[25] Federico Cruciani, Anastasios Vafeiadis, Chris Nugent, Ian Cleland, Paul McCullagh, Konstantinos Votis, Dimitrios Giakoumis, Dimitrios Tzovaras, Liming Chen, and Raouf Hamzaoui. Feature learning for human activity recognition using convolutional neural networks. *CCF Transactions on Pervasive Computing and Interaction*, 2(1):18–32, March 2020.

[26] Aaqib Bashir Dar, Auqib Hamid Lone, Saniya Zahoor, Afshan Amin Khan, and Roohie Naaz. Applicability of mobile contact tracing in fighting pandemic (covid-19): Issues, challenges and solutions. *Computer Science Review*, 38:100307, 2020.

[27] Pavel Davidson and Robert Piché. A survey of selected indoor positioning methods for smartphones. *IEEE Communications Surveys & Tutorials*, 19(2):1347–1370, 2016.

[28] Cedric De Cock, Wout Joseph, Luc Martens, Jens Trogh, and David Plets. Multi-floor indoor pedestrian dead reckoning with a backtracking particle filter and viterbi-based floor number detection. *Sensors*, 21(13):4565, 2021.

[29] Florenc Demrozi, Cristian Turetta, and Graziano Pravadelli. B-HAR: An open-source baseline framework for in depth study of human activity recognition datasets and workflows. *arXiv preprint arXiv:2101.10870*, 2021.

[30] Enmao Diao, Jie Ding, and Vahid Tarokh. SemiFL: Semi-supervised federated learning for unlabeled clients with alternate training. *Advances in Neural Information Processing Systems*, 35:17871–17884, October 2022.

[31] Xue Ding, Ting Jiang, Yi Zhong, Yan Huang, and Zhiwei Li. Wi-fi-based location-independent human activity recognition via meta learning. *Sensors*, 21(8):2654, 2021.

[32] Nidhi Dua, Shiva Nand Singh, and Vijay Bhaskar Semwal. Multi-input CNN-GRU based human activity recognition using wearable sensors. *Computing*, 103:1461–1478, 2021.

[33] Kareem El-Kafrawy, Moustafa Youssef, and Amr El-Keyi. Impact of the human motion on the vari-ance of the received signal strength of wireless links. In *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1208–1212. IEEE, 2011.

[34] Zahid Farid, Rosdiadee Nordin, Mahamod Ismail, et al. Recent Advances in Wireless Indoor Localization Techniques and System. *Journal of Computer Networks and Communications*, 2013:12, 2013.

[35] Jeremy Samuel Faust, Chengan Du, Chenxue Liang, Katherine Dickerson Mayes, Benjamin Renton, Kristen Panthagani, and Harlan M Krumholz. Excess mortality in massachusetts during the delta and omicron waves of COVID-19. *JAMA : the journal of the American Medical Association*, 328:74–76, 2022.

[36] Irene Fellner, Haosheng Huang, and Georg Gartner. "turn left after the wc, and use the lift to go to the 2nd floor"—generation of landmark-based route instructions for indoor navigation. *ISPRS International Journal of Geo-Information*, 6(6):183, 2017.

[37] Jie Feng, Can Rong, Funing Sun, Diansheng Guo, and Yong Li. PMF: A privacy-preserving human mobility prediction framework via federated learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–21, March 2020.

[38] Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491):eabb6936, 2020.

[39] Lulu Gao and Shin'ichi Konomi. Mapless indoor navigation based on landmarks. In *Proceedings of the Distributed, Ambient and Pervasive Interactions. Smart Living, Learning, Well-Being and Health, Art and Creativity: 10th International Conference, DAPI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part II*, pages 53–68, Part II; Springer, June 2022. Springer, Ber-lin/Heidelberg.

[40] Kemilly Dearo Garcia, Cláudio Rebelo de Sá, Mannes Poel, Tiago Carvalho, João Mendes-Moreira, João MP Cardoso, André CPLF de Carvalho, and Joost N Kok. An ensemble of autonomous auto-encoders for human activity recognition. *Neurocomputing*, 439:271–280, 2021.

[41] Christian Gentner and Diana Avram. Wifi-rtt-slam: simultaneously estimating the positions of mobile devices and wifi-rtt access points. In *Proceedings of the 34th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2021)*, pages 3142–3148, 2021.

[42] R.A. Gilyazev and D Yu. Turdakov. Active learning and crowdsourcing: A survey of annotation optimization methods. *ProgramminEnglishomputer Software*, 44(6):476–491, 2018.

[43] Fei Gu, Jianwei Niu, and Lingjie Duan. Waipo: A fusion-based collaborative indoor localization system on smartphones. *IEEE/ACM Transactions on Networking*, 25(4):2267–2280, 2017.

[44] Fuqiang Gu, Xuke Hu, Milad Ramezani, Debaditya Acharya, Kourosh Khoshelham, Shahrokh Valaee, and Jianga Shang. Indoor localization improved by spatial context—a survey. *ACM Computing Surveys (CSUR)*, 52(3):1–35, 2019.

[45] Fuqiang Gu, Kourosh Khoshelham, Jianga Shang, and Fangwen Yu. Sensory landmarks for indoor localization. In *2016 Fourth International Conference on Ubiquitous Positioning, Indoor Navigation and Location Based Services (UPINLBS)*, pages 201–206. IEEE, 2016.

[46] Fuqiang Gu, Kourosh Khoshelham, Shahrokh Valaee, Jianga Shang, and Rui Zhang. Locomotion activity recognition using stacked denoising autoencoders. *IEEE Internet of Things Journal*, 5(3):2085–2093, 2018.

[47] Fuqiang Gu, Shahrokh Valaee, Kourosh Khoshelham, Jianga Shang, and Rui Zhang. Landmark Graph-Based Indoor Localization. *IEEE internet of things journal*, 7(9):8343–8355, 2020.

[48] Xiansheng Guo, Nirwan Ansari, Lin Li, and Huiyong Li. Indoor localization by fusing a group of fingerprints based on random forests. *IEEE Internet of Things Journal*, 5(6):4686–4698, December 2018.

[49] Shuji Hao, Peiying Hu, Peilin Zhao, Steven CH Hoi, and Chunyan Miao. Online active learning with expert advice. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 12(5):1–22, 2018.

[50] Jeff Howe et al. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4, 2006.

[51] Lorenzo Isella, Mariateresa Romano, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Wouter Van den Broeck, Francesco Gesualdo, Elisabetta Pandolfi, Lucilla Ravà, Caterina Rizzo, et al. Close encounters in a pediatric ward: Measuring face-to-face proximity and mixing patterns with wearable sensors. *PLoS ONE*, 6(2):17144, 2011.

[52] Timofei Istomin, Elia Leoni, Davide Molteni, Amy L Murphy, Gian Pietro Picco, and Maurizio Griva. Janus: Dual-Radio Accurate and Energy-Efficient Proximity Detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 162(4):1–162, 2022.

[53] Jehn-Ruey Jiang, Hanas Subakti, and Hui-Sung Liang. Fingerprint feature extraction for indoor localization. *Sensors*, 21(16):5434, 2021.

[54] Ting Jiang, Yang Zhang, Minhao Zhang, Ting Yu, Yizheng Chen, Chenhao Lu, Ji Zhang, Zhao Li, Jun Gao, and Shuigeng Zhou. A survey on contact tracing: the latest advancements and challenges. *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, 8(2):1–35, 2022.

[55] Yifei Jiang, Yun Xiang, Xin Pan, Kun Li, Qin Lv, Robert P Dick, Li Shang, and Michael Hannigan. Hallway based automatic indoor floorplan construction using room fingerprints. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 315–324, 2013.

[56] Hyeon Jeong Jo and Seungku Kim. Indoor smartphone localization based on LOS and NLOS identification. *Sensors*, 18(11):3987, 2018.

[57] Carl-Etienne Juneau, Anne-Sara Briand, Pablo Collazzo, Uwe Siebert, and Tomas Pueyo. Effective contact tracing for covid-19: A systematic review. *Global Epidemiology*, page 100103, 2023.

[58] Joel K Kelso, George J Milne, and Heath Kelly. Simulation suggests that rapid activation of social distancing can arrest epidemic development due to a novel strain of influenza. *BMC Public Health*, 9(1):1–10, 2009.

[59] Mst Alema Khatun, Mohammad Abu Yousuf, Sabbir Ahmed, Md Zia Uddin, Salem A Alyami, Samer Al-Ashhab, Hanan F Akhdar, Asaduzzaman Khan, Akm Azad, and Mohammad Ali Moni. Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor. *IEEE Journal of Translational Engineering in Health and Medicine*, 10:1–16, 2022.

[60] Philipp H Kindt, Trinad Chakraborty, and Samarjit Chakraborty. How Reliable Is Smartphone-Based Electronic Contact Tracing for COVID-19? *Communications of the ACM*, 65(1):56–67, 2021.

[61] Robert A Kleinman and Colin Merkel. Digital contact tracing for covid-19. *Cmaj*, 192(24):E653–E656, 2020.

[62] Tao Kong, Anbang Yao, Yurong Chen, and Fuchun Sun. HyperNet: Towards accurate region proposal generation and joint object detection. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas*, pages 845–853, NV, USA, June 2016.

[63] Shin'ichi Konomi, Lulu Gao, and Doreen Mushi. An intelligent platform for offline learners based on model-driven crowdsensing over in-termittent networks. In *Proceedings of the Cross-Cultural Design. Applications in Health, Learning, Communication, and Creativity: 12th International Conference, CCD 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020*, pages 300–314, Copenhagen, Denmark, July 2020. Springer.

[64] Jayakanth Kunhoth, AbdelGhani Karkar, Somaya Al-Maadeed, and Abdulla Al-Ali. Indoor positioning and wayfinding systems: a survey. *Human-centric Computing and Information Sciences*, 10(1):1–41, 2020.

[65] Bahareh Lashkari, Javad Rezazadeh, Reza Farahbakhsh, and Kumbesan Sandrasegaran. Crowdsourcing and sensing for indoor localization in IoT: A review. *IEEE Sensors Journal*, 19(7):2408–2434, April 2019.

[66] Phuong Le-Hong and Anh-Cuong Le. A comparative study of neural network models for sentence classification. In *Proceedings of the 2018 5th NAFOSTED Conference on Information and Computer Science (NICS), Ho Chi Minh City*, pages 360–365, Vietnam, November 2018. IEEE.

[67] Matthew Lease. On quality control and machine learning in crowdsourcing. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*, pages 97–102. Citeseer, 2011.

[68] Douglas J Leith and Stephen Farrell. Measurement-based evaluation of Google/Apple exposure notification API for proximity detection in a light-rail tram. *PLoS ONE*, 15(9):e0239943, 2020.

[69] Douglas J Leith and Stephen Farrell. GAEN due diligence: Verifying the Google/Apple covid exposure notification API. In *Proceedings of the CoronaDef21, NDSS '21*, pages 1–8, San Diego, CA, USA, February 2021.

[70] Douglas J Leith and Stephen Farrell. Measurement-based evaluation of Google/Apple exposure notification API for proximity detection in a commuter bus. *PLoS ONE*, 16(4):e0250826, 2021.

[71] Chenglin Li, Di Niu, Bei Jiang, Xiao Zuo, and Jianming Yang. Meta-HAR: Federated representation learning for human activity recognition. In *Proceedings of the Web Conference 2021, Apr*, pages 912–922, 2021.

[72] Guanyao Li, Siyan Hu, Shuhan Zhong, Wai Lun Tsui, and S-H Gary Chan. VContact: Private WiFi-Based IoT contact tracing with virus lifespan. *IEEE Internet of Things Journal*, 9(5):3465–3480, 2021.

[73] Kangkang Li, Xiuze Zhou, Fan Lin, Wenhua Zeng, and Gil Alterovitz. Deep probabilistic matrix factorization framework for online collaborative filtering. *IEEE Access*, 7:56117–56128, 2019.

[74] Li Li, Yuxi Fan, Mike Tse, and Kuo-Yi Lin. A review of applications in federated learning. *Computers & Industrial Engineering*, 149:106854, November 2020.

[75] Tao Li, Dianqi Han, Yimin Chen, Rui Zhang, Yanchao Zhang, and Terri Hedgpeth. Indoorwaze: A crowdsourcing-based context-aware indoor navigation system. *IEEE Transactions on Wireless Communications*, 19(8):5461–5472, 2020.

[76] Wei Li, Ruizhi Chen, Yue Yu, Yuan Wu, and Haitao Zhou. Pedestrian dead reckoning with novel heading estimation under magnetic interference and multiple smartphone postures. *Measurement*, 182:109610, 2021.

[77] Zhaobin Li, Hongping Wang, Xinlei Zhang, Ting Wu, and Xiaolei Yang. Effects of space sizes on the dispersion of cough-generated droplets from a walking person. *Physics of Fluids*, 32(12):121705, 2020.

[78] Wei Yang Bryan Lim, Nguyen Cong Luong, Dinh Thai Hoang, Yutao Jiao, Ying-Chang Liang, Qiang Yang, Dusit Niyato, and Chunyan Miao. Federated learning in mobile edge networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(3):2031–2063, 2020.

[79] Haowen Lin, Jian Lou, Li Xiong, and Cyrus Shahabi. SemiFed: Semi-supervised federated learning with consistency and pseudo-labeling. *arXiv preprint arXiv:2108.09412*, August 2021.

[80] Tao Lin, Lingran Li, and Gérard Lachapelle. Multiple sensors integration for pedestrian indoor navigation. In *2015 international conference on indoor positioning and indoor navigation (IPIN)*, pages 1–9. IEEE, 2015.

[81] Dongjiang Liu and Yanbi Liu. An active learning algorithm for multi-class classification. *Pattern Analysis and Applications*, 22(3):1051–1063, 2019.

[82] Tao Liu, Xing Zhang, Qingquan Li, and Zhixiang Fang. Modeling of structure landmark for indoor pedestrian localization. *IEEE Access*, 7:15654–15668, 2019.

[83] Tao Liu, Xing Zhang, Huan Zhang, Nadeem Tahir, and Zhixiang Fang. A structure landmark-based radio signal mapping approach for sustainable indoor localization. *Sustainability*, 13(3):1183, 2021.

[84] Zewei Long, Liwei Che, Yaqing Wang, Muchao Ye, Junyu Luo, Jinze Wu, Houping Xiao, and Fenglong Ma. FedSemi: An adaptive federated semi-supervised learning framework. *arXiv preprint arXiv:2012.03292*, 24, December 2020.

[85] Chengwen Luo, Hande Hong, and Mun Choon Chan. PiLoc: A self-calibrating participatory indoor localization system. In *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, pages 143–153, Berlin, Germany, 2014. IEEE.

[86] Lin Ma, Tianyang Fang, and Danyang Qin. Walkslam: a walking pattern-based mobile slam solution. In *Communications, Signal Processing, and Systems: Proceedings of the 2018 CSPS Volume II: Signal Processing 7th*, pages 1347–1354. Springer, 2020.

[87] Md Abdulla Al Mamun, David Vera Anaya, Fan Wu, and Mehmet Rasit Yuce. Landmark-assisted compensation of user's body shadowing on rssi for improved indoor localisation with chest-mounted wearable device. *Sensors*, 21(16):5405, 2021.

[88] Filippo Marzoli, Alessio Bortolami, Alessandra Pezzuto, Eva Mazzetto, Roberto Piro, Calogero Terregino, Francesco Bonfante, and Simone Belluco. A Systematic Review of Human Coronaviruses Survival on Environmental Surfaces. *Science of The Total Environment*, 778:146191, 2021.

[89] Andrew McCallum, Kamal Nigam, et al. Employing EM in pool-based active learning for text classification. In *Proceedings of the International Conference*

on Machine Learning (ICML, pages 350–358. Citeseer, Morgan Kaufmann, 1998.

[90] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, January 2017.

[91] H Brendan McMahan, Eider Moore, Daniel Ramage, and Blaise Agüera y Arcas. Federated learning of deep networks using model averaging. *arXiv preprint arXiv:1602.05629*, 2, February 2016.

[92] Dimitris Milioris, Lito Kriara, Artemis Papakonstantinou, George Tzagkarakis, Panagiotis Tsakalides, and Maria Papadopouli. Empirical evaluation of signal strength fingerprint positioning in wireless LANs. In *Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems*, pages 5–13, 2010.

[93] Piotr Mirowski, Harald Steck, Philip Whiting, Ravishankar Palaniappan, Michael MacDonald, and Tin Kam Ho. KL-divergence kernel regression for non-Gaussian fingerprint based localization. In *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation*, pages 1–10. IEEE, 2011.

[94] Fernando Moya Rueda, René Grzeszick, Gernot A Fink, Sascha Feldhorst, and Michael Ten Hompel. Convolutional neural networks for human activity recognition using body-worn sensors. *Informatics*, 5(2):26, June 2018.

[95] Ohoud Nafea, Wadood Abdul, Ghulam Muhammad, and Mansour Alsulaiman. Sensor-based human activity recognition with spatio-temporal deep learning. *Sensors*, 21(6):2141, January 2021.

[96] Cong T. Nguyen, Yuris Mulya Saputra, Nguyen Van Huynh, Ngoc-Tan Nguyen, Tran Viet Khoa, Bui Minh Tuan, Diep N. Nguyen, Dinh Thai Hoang, Thang X. Vu, Eryk Dutkiewicz, Symeon Chatzinotas, and Björn Ottersten. A Comprehensive Survey of Enabling and Emerging Technologies for Social Distancing—Part I: Fundamentals and Enabling Technologies. *IEEE Access*, 8:153479–153507, 2020.

[97] Thien Duc Nguyen, Markus Miettinen, Alexandra Dmitrienko, Ahmad-Reza Sadeghi, and Ivan Visconti. Digital Contact Tracing Solutions: Promises, Pitfalls and Challenges. *IEEE Transactions on Emerging Topics in Computing*, 2022.

[98] Huthaifa Obeidat, Wafa Shuaieb, Omar Obeidat, and Raed Abd-Alhameed. A review of indoor localization techniques and wireless technologies. *Wireless Personal Communications*, 119:289–327, 2021.

[99] Jaehoon Oh, Sangmook Kim, and Se-Young Yun. FedBABU: Towards enhanced representation for federated image classification. *arXiv preprint arXiv:2106.06042*, March 2021.

[100] Chin Chun Ooi, Ady Suwardi, Zhong Liang Ou Yang, George Xu, Chee Kiang Ivan Tan, Dan Daniel, Hongying Li, Zhengwei Ge, Fong Yew Leong, Kalisvar Marimuthu, et al. Risk assessment of airborne COVID-19 exposure in social settings. *Physics of Fluids*, 33(8):087118, 2021.

[101] Ling Pei, Min Zhang, Danping Zou, Ruizhi Chen, and Yuwei Chen. A survey of crowd sensing opportunistic signals for indoor localization. *Mobile Information Systems*, 2016, 2016.

[102] Andrei Popleteev, Venet Osmani, and Oscar Mayora. Investigation of indoor localization with ambient FM radio stations. In *2012 IEEE International Conference on Pervasive Computing and Communications*, pages 171–179. IEEE, IEEE, 2012.

[103] Riccardo Presotto, Gabriele Civitarese, and Claudio Bettini. Semi-supervised and personalized federated activity recognition based on active learning and label propagation. *Personal and Ubiquitous Computing*, 26(5):1281–1298, June 2022.

[104] Matthew Rabinowitz and James J Spilker. A new positioning system using television synchronization signals. *IEEE Transactions on Broadcasting*, 51(1):51–61, 2005.

[105] Mamshad Nayeem Rizve, Kevin Duarte, Yogesh S Rawat, and Mubarak Shah. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. *arXiv preprint arXiv:2101.06329*, April 2021.

[106] Priya Roy and Chandreyee Chowdhury. A survey of machine learning techniques for indoor localization and navigation systems. *Journal of Intelligent & Robotic Systems*, 101(3):63, 2021.

[107] Priya Roy and Chandreyee Chowdhury. A survey on ubiquitous wifi-based indoor localization system for smartphone users from implementation perspectives. *CCF Transactions on Pervasive Computing and Interaction*, 4(3):298–318, 2022.

[108] XI Rui, LI Yujun, and HOU Mengshu. Survey on indoor localization. *Computer Science*, 43:1–6, 32, 2016.

[109] Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W Feldman, and James H Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences of the United States of America*, 107(51):22020–22025, 2010.

[110] Fernando Seco, Antonio R Jiménez, Carlos Prieto, Javier Roa, and Katerina Koutsou. A survey of mathematical methods for indoor localization. In *2009 IEEE International Symposium on Intelligent Signal Processing*, pages 9–14, Budapest, Hungary, 2009. IEEE.

[111] Burr Settles. Active learning literature survey. 2009.

[112] Burr Settles. Active learning: Synthesis lectures on artificial intelligence and machine learning. *Long Island, NY: Morgan & Clay Pool*, 10:S00429ED1V01Y201207AIM018, 2012.

[113] H Sebastian Seung, Manfred Opper, and Haim Sompolinsky. Query by committee. In *Proceedings of the ACM Workshop on Computational Learning Theory*, pages 287–294, 1992.

[114] John S Seybold. *Introduction to RF propagation*. John Wiley & Sons, 2005.

[115] Yash Shah, John W Kurelek, Sean D Peterson, and Serhiy Yarusevych. Experimental investigation of indoor aerosol dispersion and accumulation in the context of COVID-19: Effects of masks and ventilation. *Physics of Fluids*, 33(7):073315, July 2021.

[116] Muhammad Shahroz, Farooq Ahmad, Muhammad Shahzad Younis, Nadeem Ahmad, Maged N Kamel Boulos, Ricardo Vinuesa, and Junaid Qadir. COVID-19 Digital Contact Tracing Applications and Techniques: A Review Post Initial Deployments. *Transportation Engineering*, 5:100072, 2021.

[117] Jianga Shang, Fuqiang Gu, Xuke Hu, and Allison Kealy. Apfiloc: An infrastructure-free indoor localization method fusing smartphone inertial sensors, landmarks and map information. *Sensors*, 15(10):27251–27272, 2015.

[118] Guobin Shen, Zhuo Chen, Peichao Zhang, Thomas Moscibroda, and Yongguang Zhang. Walkie-markie: Indoor pathway mapping made easy. In *NSDI*, volume 13, pages 85–98, 2013.

[119] Jialei Shen, Meng Kong, Bing Dong, Michael J Birnkrant, and Jianshun Zhang. Airborne transmission of SARS-CoV-2 in indoor environments: A comprehensive review. *Sci. Technol. Built Environ*, 27(10):1331–1367, 2021.

[120] Shuyu Shi, Stephan Sigg, Wei Zhao, and Yusheng Ji. Monitoring attention using ambient FM radio signals. *IEEE Pervasive Computing*, 13(1):30–36, 2014.

[121] Hyojeong Shin, Yohan Chon, and Hojung Cha. Unsupervised construction of an indoor floor plan using a smartphone. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6):889–898, 2011.

[122] Pekka Siirtola and Juha Röning. Incremental learning to personalize human activity recognition models: the importance of human ai collaboration. *Sensors*, 19(23):5151, 2019.

[123] Dimitrios Sikeridis, Bhaskar Prasad Rimal, Ioannis Papapanagiotou, and Michael Devetsikiotis. Unsupervised crowd-assisted learning enabling location-aware facilities. *IEEE internet of things journal*, 5(6):4699–4713, December 2018.

[124] Navneet Singh, Sangho Choe, and Rajiv Punmiya. Machine learning based indoor localization using wi-fi rssi fingerprints: An overview. *IEEE Access*, 9:127150–127174, 2021.

[125] Konstantin Sozinov, Vladimir Vlassov, and Sarunas Girdzijauskas. Human activity recognition using federated learning. In *2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom*, pages 1103–1111. IEEE, December 2018.

[126] Santosh Subedi and Jae-Young Pyun. A survey of smartphone-based indoor positioning system using RF-Based wireless technologies. *Sensors*, 20(24):7230, 2020.

[127] V Das Swain, Hyeokhyen Kwon, Sonia Sargolzaei, Bahador Saket, M Bin Morshed, Kathy Tran, Devashru Patel, Yexin Tian, Joshua Philipose, Yulai Cui, et al. Leveraging WiFi network logs to infer student collocation and its relationship with academic performance. *arXiv preprint arXiv:2005.11228*, 2020.

[128] Alysa Ziying Tan, Han Yu, Lizhen Cui, and Qiang Yang. Towards personalized federated learning. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–17, 2022.

[129] Chi Ian Tang, Ignacio Perez-Pozuelo, Dimitris Spathis, Soren Brage, Nick Wareham, and Cecilia Mascolo. SelfHAR: Improving human activity recognition through self-training with unlabeled data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(1):1–30, March 2021.

[130] Dipanwita Thakur, Suparna Biswas, Edmond SL Ho, and Samiran Chattopadhyay. ConvAE-LSTM: Convolutional autoencoder long short-term memory network for smartphone-based human activity recognition. *IEEE Access*, 10:4137–4156, 2022.

[131] Lina Tong, Hanghang Ma, Qianzhi Lin, Jiaji He, and Liang Peng. A novel deep learning bi-GRU-I model for real-time human activity recognition using inertial sensors. *IEEE Sensors Journal*, 22(6):6164–6174, 2022.

[132] Joaquín Torres-Sospedra, Raúl Montoliu, Sergio Trilles, Óscar Belmonte, and Joaquín Huerta. Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems. *Expert Systems with Applications*, 42(23):9263–9278, 2015.

[133] Amee Trivedi, Camellia Zakaria, Rajesh Balan, Ann Becker, George Corey, and Prashant Shenoy. WiFiTrace: Network-based contact tracing for infectious diseases using passive WiFi sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(1):1–26, 2021.

[134] Pengjia Tu, Junhuai Li, Huaijun Wang, Kan Wang, and Yuan Yuan. Epidemic contact tracing with campus WiFi network and smartphone-based pedestrian dead reckoning. *IEEE Sensors Journal*, 21(17):19255–19267, 2021.

[135] Bram van Berlo, Aaqib Saeed, and Tanir Ozcelebi. Towards federated unsupervised representation learning. In *Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking, in EdgeSys '20*, pages 31–36, New York, NY, USA, May 2020. Association for Computing Machinery.

[136] Neeltje Van Doremalen, Trenton Bushmaker, Dylan H Morris, Myndi G Holbrook, Amandine Gamble, Brandi N Williamson, Azaibi Tamin, Jennifer L Harcourt, Natalie J Thornburg, Susan I Gerber, et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *New England journal of medicine*, 382(16):1564–1567, 2020.

[137] Pothuri Surendra Varma and Veena Anand. Random forest learning based indoor localization as an IoT service for smart buildings. *Wireless Personal Communications*, 117:3209–3227, 2021.

[138] Alex Varshavsky, Eyal De Lara, Jeffrey Hightower, Anthony LaMarca, and Veljo Otsason. GSM indoor localization. *Pervasive and Mobile Computing*, 3(6):698–720, December 2007.

[139] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103, 2008.

[140] Chia C Wang, Kimberly A Prather, Josué Sznitman, Jose L Jimenez, Seema S Lakdawala, Zeynep Tufekci, and Linsey C Marr. Airborne transmission of respiratory viruses. *Science*, 373(6558):eabd9149, 2021.

[141] He Wang, Souvik Sen, Ahmed Elgohary, Moustafa Farid, Moustafa Youssef, and Romit Roy Choudhury. No need to war-drive: Unsupervised indoor localization. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (MobiSys '12*, pages 197–210, New York, NY, USA, 2012. Association for Computing Machinery.

[142] Jietuo Wang, Federico Dalla Barba, Alessio Roccon, Gaetano Sardina, Alfredo Soldati, and Francesco Picano. Modelling the direct virus exposure risk associated with respiratory events. *Journal of The Royal Society Interface*, 19(186):20210819, 2022.

[143] Lei Wang, Yangyang Xu, Jun Cheng, Haiying Xia, Jianqin Yin, and Jiaji Wu. Human action recognition by learning spatio-temporal features with deep neural networks. *IEEE Access*, 6:17913–17922, 2018.

[144] Xi Wang, Mingxing Jiang, Zhongwen Guo, Naijun Hu, Zhongwei Sun, and Jing Liu. An indoor positioning method for smartphones using landmarks and PDR. *Sensors*, 16(12):2135, 2016.

[145] Yikang Wang, Jiangnan Zhang, Hairui Zhao, Mengjie Liu, Shiyi Chen, Jian Kuang, and Xiaoji Niu. Spatial structure-related sensory landmarks recognition based on long short-term memory algorithm. *Micromachines*, 12(7):781, 2021.

[146] Yingxu Wang, Jiwen Dong, Jin Zhou, Lin Wang, Shiyuan Han, Tong Zhang, and CL Philip Chen. Spectral clustering based on JS-divergence for uncertain data. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC*, pages 1972–1975, Banff, AB, 2017. IEEE.

[147] Jie Wei, Fang Zhao, and Haiyong Luo. SP-Loc: A crowdsourcing fingerprint based shop-level indoor localization algorithm integrating shop popularity without the indoor map. *International Journal of Distributed Sensor Networks*, 14(11):1550147718815637, November 2018.

[148] H. Weinberg. Using the ADXL202 in pedometer and personal navigation applications. *Analog Devices AN-602 application note*, 2(2):1–6, 2002.

[149] Chenshu Wu, Zheng Yang, and Yunhao Liu. Smartphones based crowdsourcing for indoor localization. *IEEE Transactions on Mobile Computing*, 14(2):444–457, 2014.

[150] Qiong Wu, Xu Chen, Zhi Zhou, and Junshan Zhang. FedHome: Cloud-edge based personalized federated learning for in-home health monitoring. *IEEE Transactions on Mobile Computing*, 21(8):2818–2832, August 2022.

[151] Qiong Wu, Kaiwen He, and Xu Chen. Personalized federated learning for intelligent IoT applications: A cloud-edge based framework. *IEEE Open Journal of the Computer Society*, 1:35–44, 2020.

[152] Kun Xia, Jianguang Huang, and Hanyu Wang. LSTM-CNN architecture for human activity recognition. *IEEE Access*, 8:56855–56866, 2020.

[153] Jie Xiong and Kyle Jamieson. Arraytrack: A fine-grained indoor location system. In *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*, pages 71–84, 2013.

[154] Qiang Xu and Rong Zheng. When data acquisition meets data analytics: A distributed active learning framework for optimal budgeted mobile crowdsensing. In *Proceedings of IEEE Conference on Computer Communications (IEEE INFOCOM 2017*, pages 1–9, Atlanta, GA, 2017. IEEE.

[155] S. Yang, P. Dessai, M. Verma, and M. Gerla. Freeloc: Calibration-free crowdsourced indoor localization. In *INFOCOM, 2013 Proceedings. IEEE*, pages 2481–2489. IEEE, 2013.

[156] Zheng Yang, Chenshu Wu, and Yunhao Liu. Locating in fingerprint space: Wireless indoor localization with little human intervention. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 269–280, 2012.

[157] Haiyun Yao, Hong Shu, Hongxing Sun, BG Mousa, Zhenghang Jiao, and Yingbo Suo. An integrity monitoring algorithm for wifi/pdr/smartphone-integrated indoor positioning system based on unscented kalman filter. *EURASIP Journal on Wireless Communications and Networking*, 2020:1–25, 2020.

[158] Jeonghyeon Yoon and Seungku Kim. Practical and accurate indoor localization system using deep learning. *Sensors*, 22(18):6764, 2022.

[159] Moustafa Youssef and Ashok Agrawala. The horus WLAN location determination system. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys '05*, New York, NY, USA, 2005. Association for Computing Machinery.

[160] Hongzheng Yu, Zekai Chen, Xiao Zhang, Xu Chen, Fuzhen Zhuang, Hui Xiong, and Xiuzhen Cheng. FedHAR: Semi-supervised online learning for personalized federated human activity recognition. *IEEE Transactions on Mobile Computing*, pages 1–1, 2021.

[161] Faheem Zafari, Athanasios Gkelias, and Kin K Leung. A Survey of Indoor Localization Systems and Technologies. *IEEE Communications Surveys & Tutorials*, 21(3):2568–2599, 2019.

[162] Yu Zhao, Rennong Yang, Guillaume Chevalier, Ximeng Xu, and Zhenxing Zhang. Deep residual bidir-LSTM for human activity recognition using wearable sensors. *Mathematical Problems in Engineering*, 2018:1–13, 2018.

[163] Yuchen Zhao, Hamed Haddadi, Severin Skillman, Shirin Enshaeifar, and Payam Barnaghi. Privacy-preserving activity and health monitoring on databox. In *Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking*, pages 49–54, New York, NY, USA, May 2020. Association for Computing Machinery.

[164] Yuchen Zhao, Hanyang Liu, Honglin Li, Payam Barnaghi, and Hamed Haddadi. Semi-supervised federated learning for activity recognition. *arXiv preprint arXiv:2011.00851*, 2020.

[165] Feifei Zheng, Ruoling Tao, Holger R Maier, Linda See, Dragan Savic, Tuqiao Zhang, Qiuwen Chen, Thaine H Assumpção, Pan Yang, Bardia Heidari, Jörg Rieckermann, Barbara Minsker, Weiwei Bi, Ximing Cai, Dimitri Solomatine, and Ioana Popescu. Crowdsourcing methods for data collection in geophysics: State of the art, issues, and future directions. *Reviews of Geophysics*, 56(4):698–740, 2018.

[166] Baoding Zhou, Qingquan Li, Qingzhou Mao, and Wei Tu. A robust crowdsourcing-based indoor localization system. *Sensors*, 17(4):864, 2017.

[167] Baoding Zhou, Qingquan Li, Qingzhou Mao, Wei Tu, Xing Zhang, and Long Chen. ALIMC: Activity landmark-based indoor mapping via crowdsourcing. *IEEE Transactions on Intelligent Transportation Systems*, 16(5):2774–2785, 2015.

[168] Baoding Zhou, Jun Yang, and Qingquan Li. Smartphone-based activity recognition for indoor localization using a convolutional neural network. *Sensors*, 19(3):621, 2019.

[169] Xiaolei Zhou, Tao Chen, Deke Guo, Xiaoqiang Teng, and Bo Yuan. From one to crowd: A survey on crowdsourcing-based wireless indoor localization. *Frontiers of Computer Science*, 12(3):423–450, June 2018.

Publications

Peer-reviewed Publications

1. **Gao L.**, Konomi S (2023) Indoor Spatiotemporal Contact Analytics Using Landmark-Aided Pedestrian Dead Reckoning on Smartphones. Sensors 23:113. https://doi.org/10.3390/s23010113

2. **Gao, L.**, and Konomi, S. (2023). Personalized federated human activity recognition through semi-supervised learning and enhanced representation. In Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers (UbiComp/ISWC '23 Adjunct). Association for Computing Machinery, New York, NY, USA. (to appear)

3. **Gao, L.**, and Konomi, S. (2022). Indoor Contact Awareness on Spatiotemporal Analytics with Smartphone-Based Pedestrian Dead Reckoning. In Proceedings of the 2022 ACM Conference on Information Technology for Social Good (GoodIT '22). Association for Computing Machinery, New York, NY, USA, 205–211. https://doi.org/10.1145/3524458.3547233

4. **Gao, L.**, Konomi, S. (2022). Mapless Indoor Navigation Based on Landmarks. In: Streitz, N.A., Konomi, S. (eds) Distributed, Ambient and Pervasive Interactions. Smart Living, Learning, Well-being and Health, Art

and Creativity. HCII 2022. Lecture Notes in Computer Science, vol 13326. Springer, Cham. https://doi.org/10.1007/978-3-031-05431-0_4

5. **Gao, L.**, Konomi, S. (2022). A Cost-Effective and Quality-Ensured Framework for Crowdsourced Indoor Localization. In: Duffy, V.G., Landry, S.J., Lee, J.D., Stanton, N. (eds) Human-Automation Interaction. Automation, Collaboration, & E-Services, vol 11. Springer, Cham. https://doi.org/10.1007/978-3-031-10784-9_27

6. **Gao, L.**, Konomi, S. (2021). Active Learning-Based Data Collection in Crowd Replication. In: Advances in Artificial Intelligence. JSAI 2020. Advances in Intelligent Systems and Computing, vol 1357. Springer, Cham. https://doi.org/10.1007/978-3-030-73113-7_5

Oral Presentations

7. **Lulu Gao**, Shin'ichi Konomi (2022). Modeling COVID-19 Viral Concentration During Human Movement in Indoor Environment. Proceedings of the 21th IPSJ FIT: Forum on Information Technology, September 2022.

8. Xiangyuan Hu, Shinichi Konomi, **Lulu Gao**, Kaoru Sezaki, (2021). Analysis of Human Flows to Inform the Design of a Crowd-Powered Information Delivery Environment in Developing Communities, Research Abstracts on Spatial Information Science CSIS DAYS 2021, D01, 2021.11.

9. **Lulu Gao**, Shin'ichi Konomi, (2020). Active Learning-based Crowd Replication, Proceedings of the 34th Annual Conference of Japanese Society for Artificial Intelligence, Kumamoto, June 9 - June 12, 2020. 1G4-ES-5-01., 2020.06. https://doi.org/10.11517/pjsai.JSAI2020.0_1G4ES501.