Doctoral Dissertation

An Empirical Study of Joint Crowd Counting and Localization System based on Wireless Sensing and Machine Learning

Hyuckjin Choi

Program of Information Science and Technology
Graduate School of Science and Technology
Nara Institute of Science and Technology

Supervisor: Professor Keiichi Yasumoto
Ubiquitous Computing Systems Lab.
(Division of Information Science)

Submitted on June 16, 2022

A Doctoral Dissertation
submitted to Graduate School of Science and Technology,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of Engineering

Hyuckjin Choi

Thesis Committee:
Professor Keiichi Yasumoto (Supervisor)
Professor Minoru Okada (Co-supervisor)
Associate Professor Hirohiko Suwa (Co-supervisor)
Affiliate Associate Professor Manato Fujimoto (Co-supervisor)

An Empirical Study of Joint Crowd Counting and Localization System based on Wireless Sensing and Machine Learning*

Hyuckjin Choi

Abstract

Human sensing techniques such as activity recognition, localization, and crowd estimation, have been continuously studied since the human population in the modern society has gone beyond the range of manual processing. Until now, many human sensing techniques have been mainly performed by vision-based approaches using cameras, however, the camera could not always become an appropriate solution in all situations or entire area of interest due to its critical constraints. This is the major reason why WiFi sensing-based methods are now highly spotlighted thanks to WiFi's ubiquity and fine-grained source data like channel state information (CSI). In this thesis, a joint crowd estimation system for crowd counting and localization by leveraging WiFi sensing and machine learning (ML) has been addressed. So far, crowd counting and localization have been mostly considered as the separate topics. On the other hand, vision-based crowd estimation system can intuitively estimate the crowd size and the location of human cluster at the same time. We are inspired by the idea that the same thing a camera can do also can be performed by WiFi sensing. As part of our observations, the fluctuation level of CSI amplitude shows monotonic relationship with the crowd size, meanwhile, a particular shape of CSI curve in subcarrier domain implies the location of people gathering at. By investigating these CSI characteristics, we found that WiFi CSI sufficiently contains the information that reveals dynamic (size of a moving crowd) and static state (location

*Doctoral Dissertation, Graduate School of Science and Technology Science,
Nara Institute of Science and Technology, June 16, 2022.

of the crowd) of a WiFi channel. By this work, we demonstrate the feasibility of joint crowd counting and localization by the combination of WiFi sensing and ML, and practically construct the system in real environment, and empirically provide various analytic results obtained by the practical experiments. As a result of real-environment experiments, we achieved 0.35 median absolute error (MAE) of counting and 91.4% of localization accuracy with five people in a small-sized room, and 0.41 MAE of counting and 98.1% of localization accuracy with 10 people in a medium-sized room.

Keywords:

Wireless sensing, WiFi channel state information, crowd estimation, counting, localization, machine learning.

Contents

List of Figures

List of Tables

1 Introduction

The importance of technological prediction of how people will behave or make a decision has been growing up more and more in our modern society since the human population has gone beyond the range of manual processing. Before those predictions, naturally, we first need to estimate the current situation of people in an area of interest. The crowd estimation technique is one of the methods that can contribute to various situation understandings. In a retail store or supermarket that has separate sections divided by product types as an example, if we are able to recognize how many people are passing by and gathering at a certain area or passage in a specific time (current situation), it leads to a prediction of sales trends of the particular goods as well as time-specific section congestion (prediction). This enables the shop manager to appropriately arrange the products, assign the optimal work schedule for staff, and also especially, it could be very meaningful in terms of crowd dispersal in a situation such as the COVID-19 pandemic spread since 2019. Since this aspect identically applies to the museums, exhibitions, or expositions as well, we can acquire the real-time crowd information of each area in those places as we deploy the crowd estimation system area-by-area.

Today, the most universal method for crowd estimation is a vision-based technique, and wireless sensing-based approaches are rapidly catching it up. The camera and vision-based techniques are intuitively possible in human counting with good accuracy thanks to well-developed head counting and pattern recognition in the images [1–3]. Especially, they have an advantage in estimating an extensive crowd over a huge outdoor area. However, vision-based approaches have some critical weaknesses at the same time, such as non-availability under the dim light circumstances, impossibility of widespread installation of cameras, underestimation due to occlusion of objects, and privacy-invasive concerns. Nowadays, many technical approaches for both indoor and outdoor crowd estimation have been attempted using various wireless sensing technologies, e.g., WiFi [4], PIR sensor [5], Bluetooth [6], wireless sensor network [7], and also the combination of multiple wireless sensing technologies such as WiFi, UWB, and light sensor [8]. Among them, WiFi sensing-based methods are now highly spotlighted because of WiFi's pervasiveness and fine-grained source data like channel state information (CSI).

WiFi sensing can be divided into two major approaches: CSI and passive WiFi radar. In [9], Li *et al.* compared the fundamentals and activity recognition results by leveraging both systems. They evaluated the systems by machine learning, then concluded the CSI-based system performs better in a line of sight (LoS) condition, whereas the radar-based system shows better performance in a non-LoS environment. In this paper, we address a WiFi CSI-based crowd estimation approach, because our target area is an indoor LoS-link environment. Meanwhile, we adopt machine learning to assess our system performance. The recent WiFi sensing techniques are now often being collaborated with IoT and machine learning technologies as spectrum sensing does [10, 11], which is a basis of the wireless channel sensing in the field of cognitive radio.

Although numerous WiFi-based human sensing techniques have been studied so far [12, 13], most of those studies are focused on resolving only a single issue such as person identification [14], respiration detection [15], activity recognition [16], and human detection [17]. Particularly, the crowd counting and localization techniques are treated as separate issues in most cases. On the other hand, one thing we need to note regarding the vision-based methods is that it can count people along with recognizing which part of the area the people are gathered at, from the image or video. Practically, there are several camera-based studies addressing both issues of crowd counting and localization [18, 19]. Knowing the location of a crowd has great advantages in terms of system distribution cost and energy efficiency. If the system can recognize not only the number of people in a crowd but also where the people are gathered, we will be able to sparsely deploy the sensing devices in an area of interest instead of installing them densely to estimate the situation of all small separate sections. Also, we can provide a targeted air-conditioning service toward a more crowded location by graded adjustment of multiple air conditioners in a large room or area.

In our publication [20], we were inspired by the idea that the same thing a camera can do can be also performed by wireless sensing, and to the best of our knowledge, it was the first attempt of simultaneous crowd estimation by using WiFi CSI. From [20], we have further revealed the potential of WiFi CSI toward a comprehensive crowd estimation system. We propose a method for device-free joint crowd counting and localization, and evaluate the system by the experiments with the further enhanced features and more number of people than the previous work, at two different test beds. To examine the new WiFi CSI platform, we utilize ESP32* node which is a compact IoT solution of WiFi/Bluetooth communication and sensing, instead of inaccessible, conventional WiFi CSI tools. We show convincing results obtained by machine learning (ML) using practical experiment data from two test fields with up to 10 people. Finally, we provide diverse analytic comparisons in detail, by handling several conditions which are influential in system performance. This thesis acquires significance by the following main contributions:

- First, we demonstrated the feasibility of real-time simultaneous crowd estimation system that can precisely estimate not only the crowd count but also the location of the crowd in parallel.

- Second, we examined the potential of ESP32 nodes and CSI toolkit to become a promising WiFi sensing platform, and confirm that they have sufficient sensing resolution for medium-scale crowd estimation.

- Third, practical validation were conducted in two different real environments, which are small-sized meeting room with five people and medium-sized seminar room with 10 people.

- Fourth, we evaluated the system performance by leave-one-session-out cross-validation to reflect CSI tendency change depending on time-varying environmental factors, as well as by continuous data series (k-fold cross-validation).

*https://www.espressif.com/en/products/socs/esp32

- Finally, diverse analytic results were obtained by machine learning (regression analysis for crowd counting and classification for crowd localization) with comparisons depending on conditions and parameters, additionally, we examined the differences and comparisons with the results by deep learning.

The rest of this thesis is organized as follows. In Chapter 2, we first briefly review the studies related to crowd estimation. We then address the background of WiFi CSI and its solutions, and our observation in terms of CSI characteristics in Chapter 3. The proposed system for joint crowd counting and localization is described in Chapter 4. We present the evaluation method of our system and the results in Chapter 5. Finally, we give a discussion about the current state and future works and conclude this thesis in Chapter 6.

2 Related Work

In this chapter, we review the literature related to crowd estimation systems focusing on the techniques based on WiFi, which are listed in Table 2.1. Since we can observe the significant variation of CSI only by the change of multipath environment or LoS blockage events of a WiFi link, most WiFi sensing-based human sensing approaches are based on the mobility of the target object. Therefore, all following crowd estimation systems are assuming the situations of when people are walking in or passing through the WiFi channels, same as our work.

Depatla and Mostofi [21] presented a technique for through-wall crowd counting based only on WiFi received signal strength (RSS). In the paper, they emphasized that through-wall counting should be demonstrated in case there is no available WiFi device in an area, pointing out that transceivers are located within the area of interest in all the conventional counting methods. They proposed a motion model for multi-people walking to estimate the number of people walking inside with one pair of WiFi transceivers behind walls. Ibrahim *et al.* [22] proposed a deep learning system for WiFi-based human counting. They also used WiFi RSS measurements to detect temporal line of sight (LoS) blockage of a single WiFi link. They utilized LoS blockage detector to measure its timing and long short-term memory (LSTM) model to overcome the vanishing gradient problem during long sequences training. They showed that the system is able to count the people with 63% of count accuracy in a small room with up to seven people, and 55% of count accuracy in a medium-sized room with up to 10 people.

Liu *et al.* [23] proposed an approach of deep learning-based crowd counting using WiFi CSI. Both CSI amplitude and phase are used as source data in the system, and they attempted to use two filters to smooth those measurements. They provided performance comparison depending on impacts of time window size, neural network structure, and pre-processing method. The system showed

Table 2.1: Comparison of Existing Crowd Estimation Works.

Ref.	**Source**	**Frequency band**	**Max. crowd count**	**Accuracy**
Nakatsuka *et al.* [25]	RSS	2.4 GHz	29	N/A
Yuan *et al.* [26]	RSS	2.4 GHz	10	94%
Xu *et al.* [27]	RSS	2.13 GHz	4	86% (counting), 1.3 m (localization)
Depatla *et al.* [21]	RSS	2.4 GHz	20	1.3 MAE of counting error
Ibrahim *et al.* [22]	RSS	N/A	10	63%, 55% (small, medium room)
Xi *et al.* [28]	CSI	N/A	12	98% of within 2 person error
Guo *et al.* [29]	CSI	2.4/5 GHz	3	85%, 99% (2.4, 5 GHz)
Liu *et al.* [23]	CSI	5 GHz	5	82.3%
Di Domenico *et al.* [24]	CSI	2.4 GHz	7	74%, 52% (small, large room)
Zou *et al.* [30]	CSI	5 GHz	11	92.8%
Li *et al.* [31]	CSI	N/A	8	92%
Zhou *et al.* [32]	CSI	5 GHz	34	0.14 MAE of counting error

82.3% of average recognition accuracy with up to five people. Di Domenico *et al.* [24] presented a differential CSI approach for counting by trained-once classification model. Normalized Euclidean distance between two CSI vectors is used as a basic metric of the system to reduce the dependence on the background environment. They trained a classifier with the data from a medium-sized room, and tested it with the data from small-sized and large-sized rooms. The system showed 74% of classification accuracy by small room data, and 52% by large room data.

Zou *et al.* [33] proposed FreeCount, which is a device-free crowd counting scheme using a modified CSI tool running on commercial WiFi devices. They adopted the transfer kernel learning (TKL) model to take account of temporal variation of CSI measurements, and trained the model with 20 features based on de-noised CSI data by wavelet filter, which are categorized in common statistics, transformation-based, and shape-based features. In addition, they extended and further developed their system into WiFree in [30]. They mainly measured the shape similarity between adjacent time series CSI curves to distinguish the number of people. Also, the feature selection method was presented in the paper, to figure out the most informative features for the system. They demonstrated the system in three different-sized rooms with four, seven, and 11 participants, respectively, and achieved 99.1% of occupancy detection accuracy and 92.8% of crowd counting accuracy.

Xi *et al.* [28] proposed a device-free crowd counting approach by using the percentage of non-zero elements (PEM) and the Grey Theory, where PEM is a metric of dilated CSI matrix for crowd counting proposed in the paper. The values of PEM reflect the fluctuation of CSI signal by a matrix with '0' or '1' elements, based on the idea that the signal is unstable, then the dilated CSI matrix contains the larger number of '1'. This is grounds for monotonic relation between the number of people and PEM. They evaluated their system with Intel 5300 NIC-based CSI tool, and their results showed that the ratio of estimation errors within two people was 98% in the indoor area and 70% in the outdoor area.

Some works use this PEM as a main metric of their system. Li *et al.* [31] presented a device-free indoor people-counting method based on WiFi CSI and PEM. To calculate PEM, they made dilated matrix by the covariance matrix of both CSI amplitude and phase. Their system achieved robustness and detection performance by combining the amplitude and phase information in CSI data, and validated a monotonic relation between CSI variation and crowd number. It is shown that the system can get 92% of accuracy with up to eight people. Meanwhile, Zhou *et al.* [32] proposed the crowd counting technique by using WiFi CSI and deep neural networks (DNN), and PEM. They also leveraged PEM to construct the monotonic relationship between the change of CSI amplitude and people count by the DNN regression model. One pair of WiFi links was used in their experiment with Intel 5300 NIC-based CSI tool. They achieved 0.11 of mean counting error in a medium-sized meeting room with up to five people and 0.14 of mean counting error in a hall with up to 34 people.

In [27], Xu *et al.* described SCPL system which can perform the counting and localization in parallel. The system consists of two phases, first is counting subjects by successive cancellation (iteratively subtracting an impact of one target from the measurements) and the other is localizing each subject by indoor human tracking model. They tested their system in two indoor environments with four people, then achieved up to 86% of counting accuracy and 1.3 m of average localization error. However, they only used WiFi RSS as their system's source data, leading to very extensive distribution of necessary WiFi devices (about 20 nodes for each test area) for high accuracy. Since this work is addressing multi-subject counting and individual tracking, it is essentially different from our work which is estimating the number of people in the crowd and the sectioned location of the human cluster itself.

Mohammadmoradi *et al.* [8] presented multi-modal people counting by a combined system of multiple wireless sensors such as WiFi, UWB, and light sensors. Their estimation is performed based on the detection of the flow of people getting into a room or going out of the room through the sensor sets installed on both sides of the door. They described that each sensor can independently detect a person's passage by variation of the sensor signal, then the final decision is made by a majority vote between the different sensors. Also, they tested that each sensor can tell the obvious difference of when multiple people move in/out together at the same time. As a result, WiFi and UWB could distinguish the cases of the movement of multiple targets (up to three people), and the system showed 96% of overall performance in passage counting.

Finally, Zheng *et al.* [34] examined the impact of radio frequency interference (RFI) on WiFi CSI measurements, and proposed the cyclostationary analysis-based RFI detection algorithms. They described that, even though the CSI-based sensing applications have been widely studied in recent years, the RFI problem is overlooked and unexplored in the field of WiFi sensing. Therefore, they conducted real-world experiments with WiFi (main signal source), ZigBee, Bluetooth, and microwave (RFI sources). They provided several comparisons depending on evaluation metric, interference type, RFI-Rx distance, or Tx-Rx Distance, then the system eventually showed over 90% of RFI detection accuracy.

All the above-mentioned studies utilized the conventional WiFi routers and old CSI platforms that require particular WiFi modules such as Intel 5300 NIC or Qualcomm Atheros WiFi chip. In our work, we leverage ESP32 transceivers as the signal source which is the latest WiFi IoT CSI solution. Although the conventional WiFi routers can obtain more fine-grained and stable CSI measurements, we will show that our system also could achieve promising and convincing, even better performance. Most of all, we differentiate our work from other related works by a point of revealing the possibility and potential in WiFi IoT sensing-based simultaneous crowd estimation for both counting and localization.

3 WiFi Sensing Preliminaries

In this chapter, we briefly describe the basics of WiFi CSI, currently usable solutions and a new promising CSI IoT platform, and our observations.

3.1 Background

As mentioned earlier, many research works are leveraging a WiFi sensing technique thanks to some solutions for access to WiFi CSI open to the public. CSI represents an estimate of the impulse response of the propagation channel between a transmitter and a receiver in the orthogonal frequency-division multiplexing (OFDM) transmission system. When we denote the OFDM system in the frequency domain, it is modeled as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{3.1}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the transmitted and received complex vectors, and $\mathbf{n}$ and $\mathbf{H}$ are noise vector and channel information matrix, respectively. Since CSI is an estimate of $\mathbf{H}$, it can be denoted as $\hat{\mathbf{H}}$ which is obtained from a transmitter. $\hat{\mathbf{H}}$ contains the information of amplitude attenuation and phase shift of each subcarrier in the form of complex numbers, therefore, these measurements can be denoted as:

$$CSI = \hat{\mathbf{H}} = ||\hat{\mathbf{H}}||e^{j\angle\hat{\mathbf{H}}} \tag{3.2}$$

where $||\hat{\mathbf{H}}||$ and $\angle\hat{\mathbf{H}}$ mean the CSI measurements of amplitude attenuation and phase shift, respectively.

3.2 Conventional CSI Platform

There are two representative WiFi CSI-enabled solutions, Linux 802.11n CSI tool [35] and Atheros CSI tool [36]. Those have been widely utilized as CSI-enabled platforms in various publications so far. However, both Linux 802.11n and Atheros tools require a laptop or WiFi router which is equipped with particular WiFi modules such as Intel 5300 NIC for the former, and specific Qualcomm Atheros WiFi chips for the latter. This fundamentally restricts the accessibility to CSI data, even some of those modules are purchasable only from the used-item market. They also may cause inconvenience in device deployment due to the requirement of a laptop or router. Moreover, the Linux 802.11n CSI tool has a constraint that it can provide CSI readings of only 30 subcarriers out of 64 subcarriers. Therefore, some researchers have modified those CSI tools to fit them into their systems.

3.3 New Compact WiFi CSI Solution

In early 2020, an ESP32 CSI toolkit has been presented as a new CSI solution, emphasizing its convenience and accessibility [37]. Using this toolkit, the authors of [37] practically performed further research works regarding human occupancy and direction monitoring in [38]. They conducted a hallway experiment to investigate the capability of ESP-based device-free WiFi sensing for single-person detection and walking direction prediction, even if the Tx/Rx ESP nodes are lined up behind the same side of a wall. In addition, they also presented a method of soil sensing by using ESP nodes in [39], demonstrating that ESP-based WiFi sensing is effective not only for human sensing. By [38,39], they showed the feasibility of this compact ESP32 becoming an alternative solution of WiFi sensing. In this thesis, we also adopt the ESP32 CSI toolkit and ESP32 WiFi nodes, which are shown in Figure 3.1, as the CSI reading devices for WiFi sensing. Since the ESP32 module has a single antenna, it can only exploit signals from fewer channels than other two-by-two or three-by-three MIMO WiFi architectures. Consequently, we could obtain a relatively small amount of CSI data. Nevertheless, this low-cost, low-power, compact WiFi node has a great advantage in terms of easy and flexible deployment. We suppose that these compact devices have the potential to become a promising WiFi IoT sensing solution.

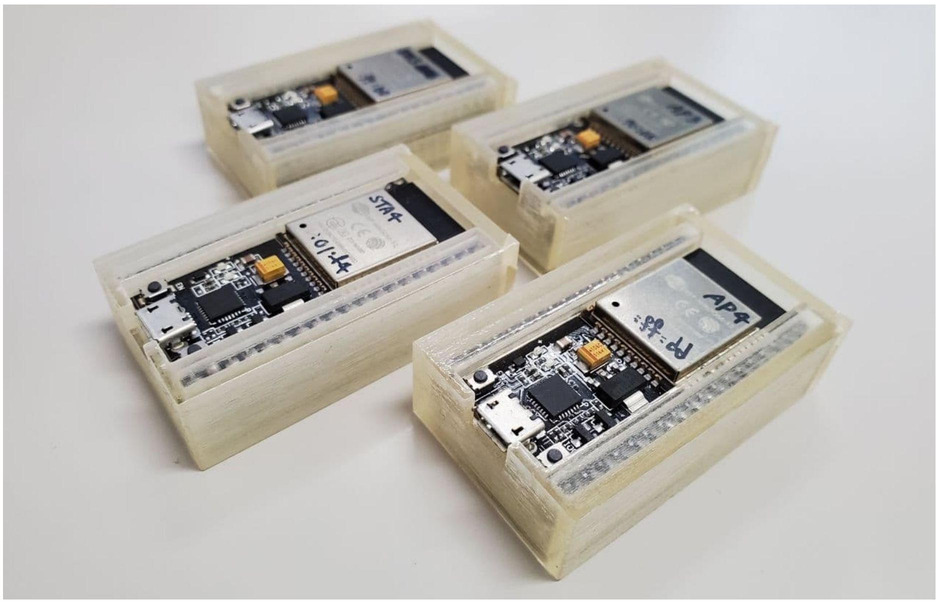


Figure 3.1: ESP32 Nodes.

For this work, we set several of ESP32 nodes as transmitters (access point, AP), and the others as receivers (station, STA), to make multiple WiFi links. We assign a dedicated SSID and password to each pair of Tx/Rx for one-to-one communication at a configured packet rate, by the ESP32 CSI toolkit operating in the Linux terminal. Since the ESP32 nodes are powered, the AP continuously sends CSI requests to the STA, then, the STA returns the observed CSI information to the AP so that we can get the channel state between AP and STA from the AP side. In our system, the returnd CSI data at AP side is transferred to a laptop which is connected by a data cable. The ESP32-based CSI architecture is depicted in Figure 3.2. The ESP32 nodes are operated on 802.11n legacy mode WiFi, which uses 2.4 GHz band (bandwidth: 20 MHz) and consists of 52 non-null subcarriers [38, 39].

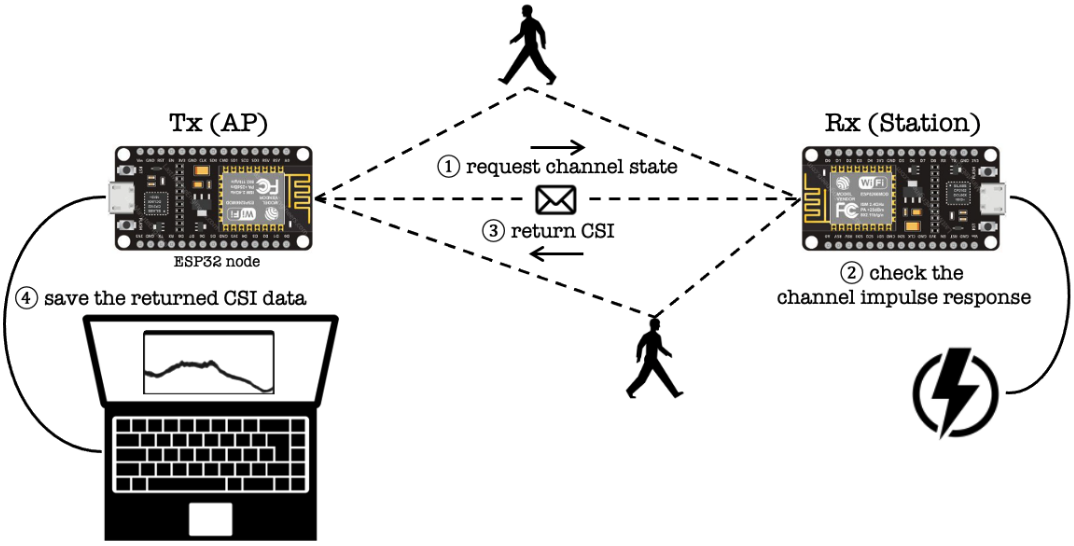

Tx (AP)
Rx (Station)
① request channel state
③ return CSI
ESP32 node
② check the channel impulse response
④ save the returned CSI data

Figure 3.2: ESP32-based CSI Architecture.

If there are multiple WiFi links in the system, a measured CSI vector $\mathbf{h}_{i,k}$ from the i^{th} packet can be denoted as:

$$\mathbf{h}_{i,k} = (h_{i,1,k}, \cdots, h_{i,j,k}, \cdots, h_{i,n_s,k}) \tag{3.3}$$

where $h_{i,j,k}$ is a complex CSI value of j^{th} subcarrier measured in the k^{th} link, and n_s is the total number of available subcarriers. Since the complex CSI values contain information of both amplitude $a_{i,j,k}$ and phase $\phi_{i,j,k}$, they can be calculated by the following equations:

$$\begin{aligned} a_{i,j,k} &= \sqrt{Re(h_{i,j,k})^2 + Im(h_{i,j,k})^2} \\ \phi_{i,j,k} &= atan2(\,Im(h_{i,j,k}), Re(h_{i,j,k})\,) \end{aligned} \tag{3.4}$$

where $Re(\cdot)$ and $Im(\cdot)$ are the functions of the real and imaginary part of a complex number, respectively, and $atan2(y, x)$ is the function of 2-argument arctangent.

In Figure 3.3, we depict each complex CSI value of entire subcarrier on the complex plane. Each black dot stands for amplitude and phase value of a single CSI packet. We can see that the black dots are forming a circle shape on the complex plane. Here, the width of a circle present how severely the CSI amplitude is fluctuating, which is a CSI characteristic that we will leverage as a feature for crowd counting. On the other hand, the orientation of dots from origin implies the phase shift level, however as we can see in the figure, the fact that the dots are shaping a circle means phase shift offset is omni-directional produced without any pattern. In our system, we use only the amplitude values $a_{i,j,k}$, because the purpose of this work does not strictly require a contribution of phase shift value. Phase shift value is required for some applications that need angle of arrival (AoA) or time of flight (ToF), but it is excluded in some cases due to its severe offset caused by hardware and software errors that leads to difficulty in clarifying the signal pattern, as described in [12].

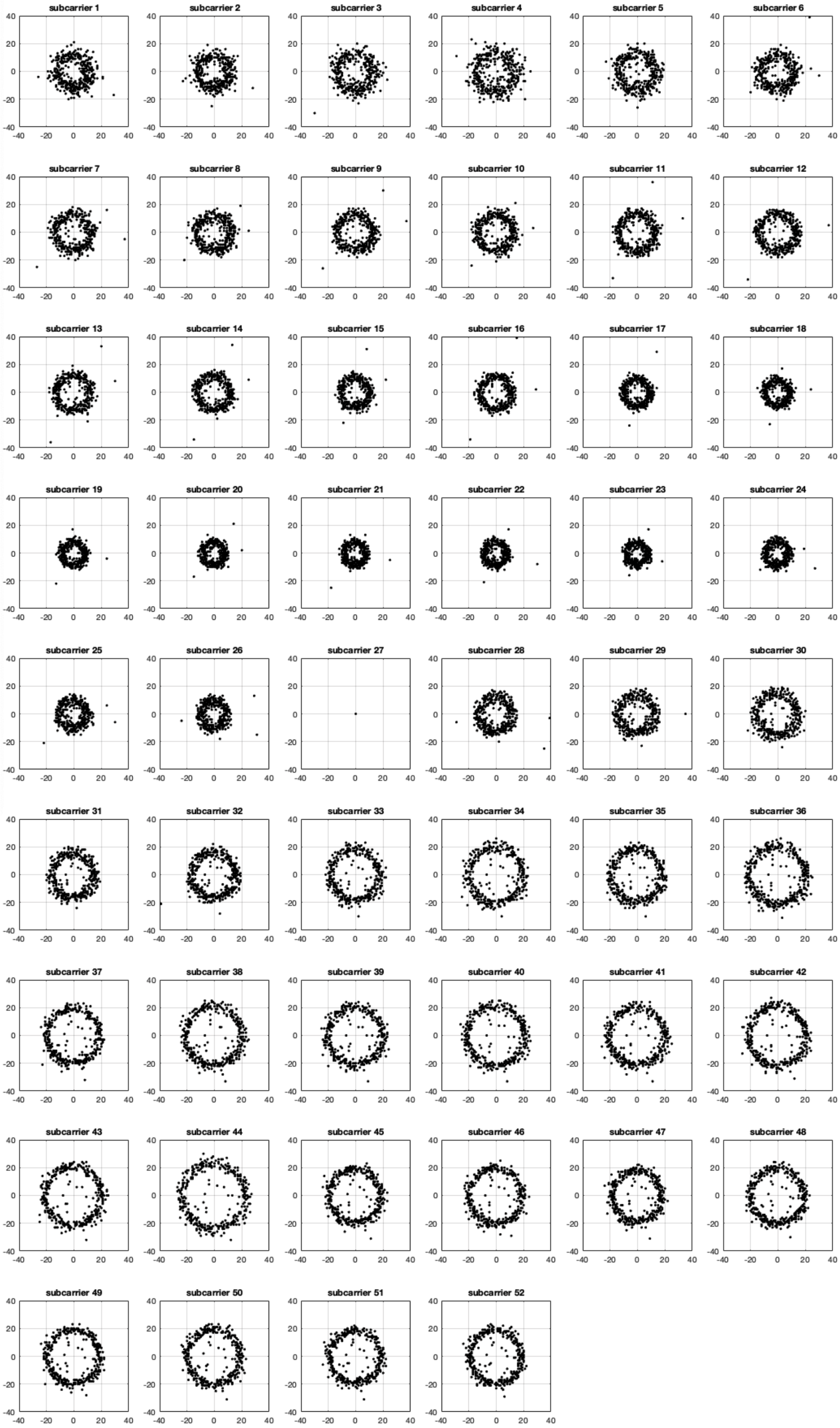


Figure 3.3: Complex Plane.

3.4 Observations

WiFi CSI provides measurements of the signal amplitude and phase information at the subcarrier level. To investigate the CSI amplitude data, we look into a subcarrier-amplitude plot that shows the signal magnitudes of each subcarrier within a certain time interval. As we can see in Figure 3.4, we can first form a single CSI curve by looking into the time-series CSI data in subcarrier domain. That is, a CSI curve shows the amplitude levels of one packet across all subcarriers. After that, we visualize all curves in a time window into a single plot so that we can see how the a bundle of CSI curve changes both in terms of width and shape, as we can see in Figure 3.5. We define this process as data segmentation into a time window. In our system, for example, the time-series CSI data is segmented into six-second time windows to convert it into overlapped CSI curves (as we will describe in Chapter 4). In one time window, we call the overlapped CSI curves a CSI bundle. CSI bundle shows a specific tendency in terms of the width and shape, therefore, it reveals a couple of characteristics in accordance with the propagation condition between WiFi AP and STA, which is changed by moving objects or channel circumstances. Those characteristics can be represented in dynamic and static state-dependent characteristics, which are described in the following chapters.

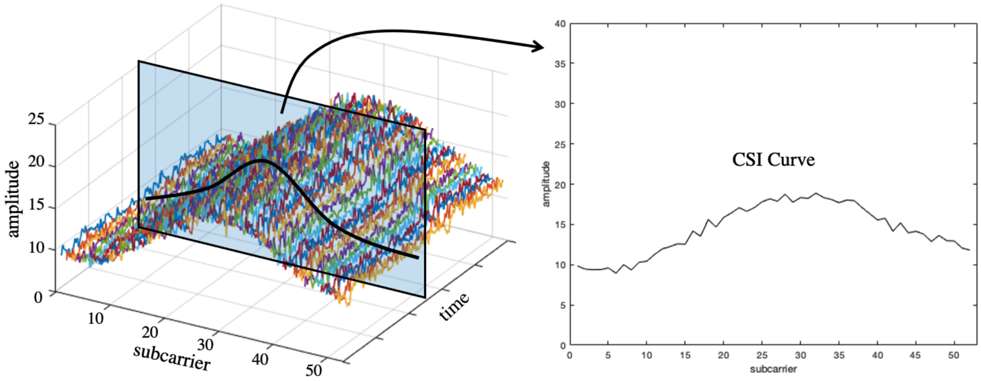


Figure 3.4: CSI Curve Formation.

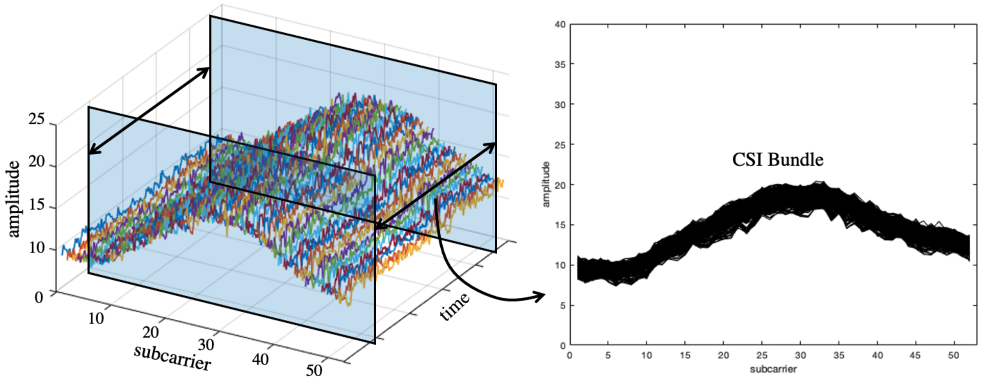
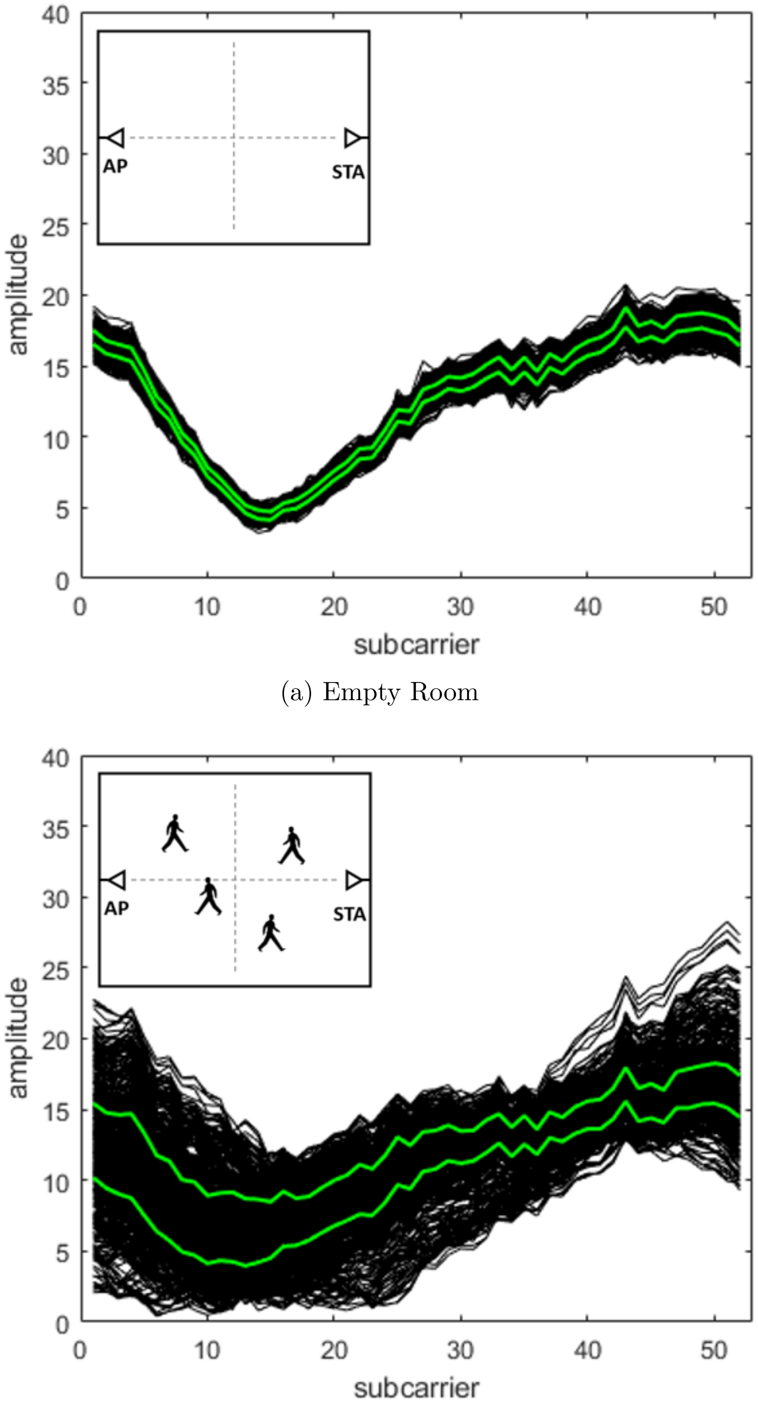


Figure 3.5: CSI Bundle Formation.

3.4.1 Dynamic State-dependent Characteristic

For crowd counting, we associate the bundle-width variation with the number of people. If there is no person between a WiFi link, the signal multipath or scattering effect is nearly constant and signal variation only comes from observational error, thermal noise, or signal interference. Therefore, the CSI amplitudes across all the subcarriers are relatively stable. On the other hand, as the number of people in the area increases, the multipath environment becomes more and more complicated due to increased moving objects. As a result, the amplitudes fluctuate widely and the CSI bundle width consequently gets thicker. In Figure 3.6(a) and (b), the black curves form the CSI bundles of the cases when an area is empty and four people are walking within the area, respectively, and the green lines represent the lower and upper quartile values across all subcarriers, which can reveal the difference of bundle width.



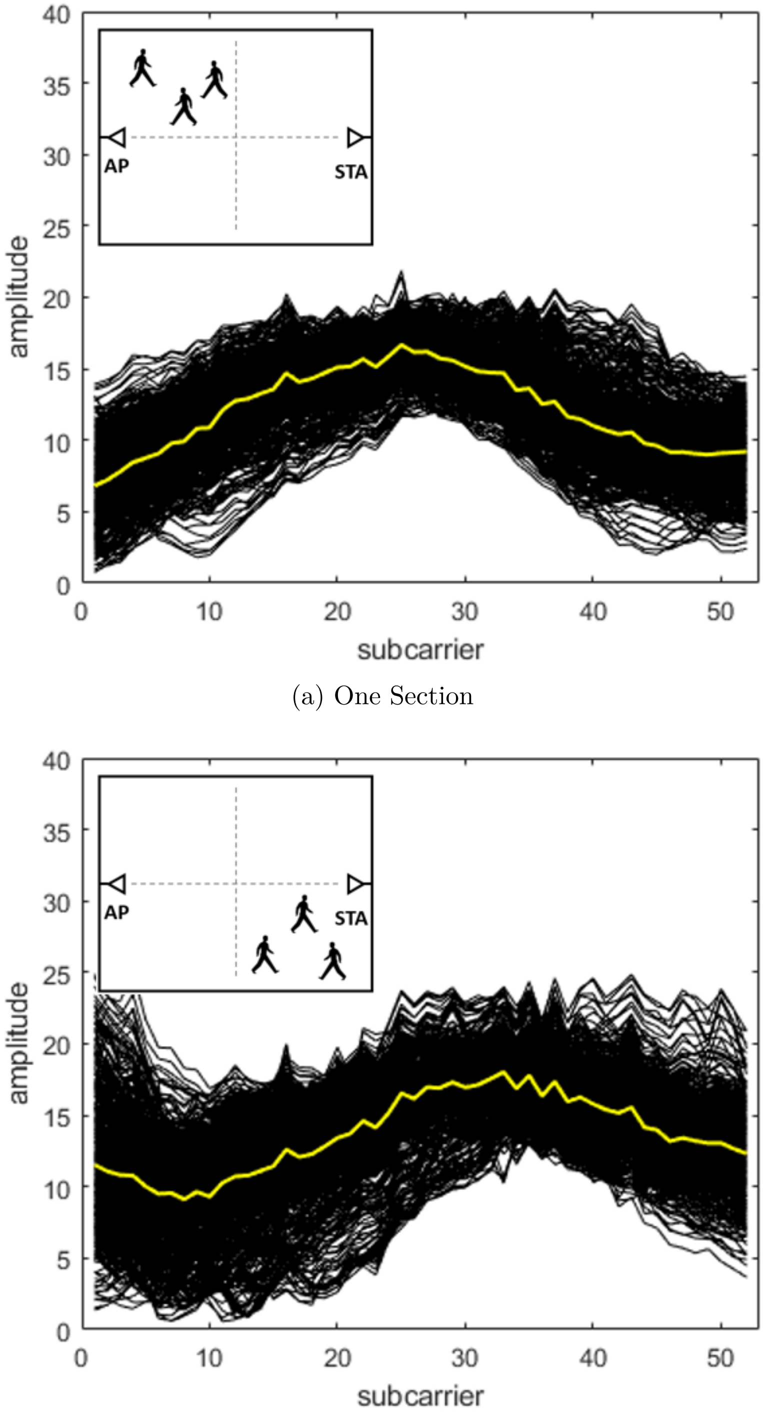
(a) Empty Room

(b) Four People

Figure 3.6: CSI Bundle Tendency depending on Crowd Size.

3.4.2 Static State-dependent Characteristic

In a CSI bundle, we can also recognize a particular shape depending on the difference of the target space's inner structure and/or distribution of objects including human bodies. The basic shapes of CSI curves are formulated depending on the inner structure of a target area. However, a cluster of people consistently moving around within a limited area constantly affects the multipath environment of the WiFi signal. Consequently, this continuous influence affects the formation of shape tendency of the CSI bundle as well. Figure 3.7(a) and (b) show the difference of CSI-bundle-shape formation with yellow average line, between two different situations that three people are freely walking within one section and another section of a target area.



(a) One Section

(b) Another Section

Figure 3.7: CSI Bundle Tendency depending on Crowd Location.

4 Proposed System: Joint Crowd Counting and Localization

In this chapter, we propose a WiFi sensing based joint crowd counting and localization system that enables both crowd counting and crowd localization.

4.1 Outline

The final goal of this study is to investigate if the proposed system can estimate not only how many people are in a particular area, but also which specific section of that area people are gathering at. Therefore, we devise effective features for dynamic and static state-dependent characteristics as well as using common statistical features. Since we found that some features extracted from CSI data generally have the monotonic relationship to people count, ML regressor is used for crowd counting. On the other hand, crowd localization should be estimated by ML classifier because we divide the test area into discrete sections. Figure 4.1 shows the comprehensive flow of our system. We describe the system flow in the following chapters, including the scheme and method of data processing and feature extraction in detail.

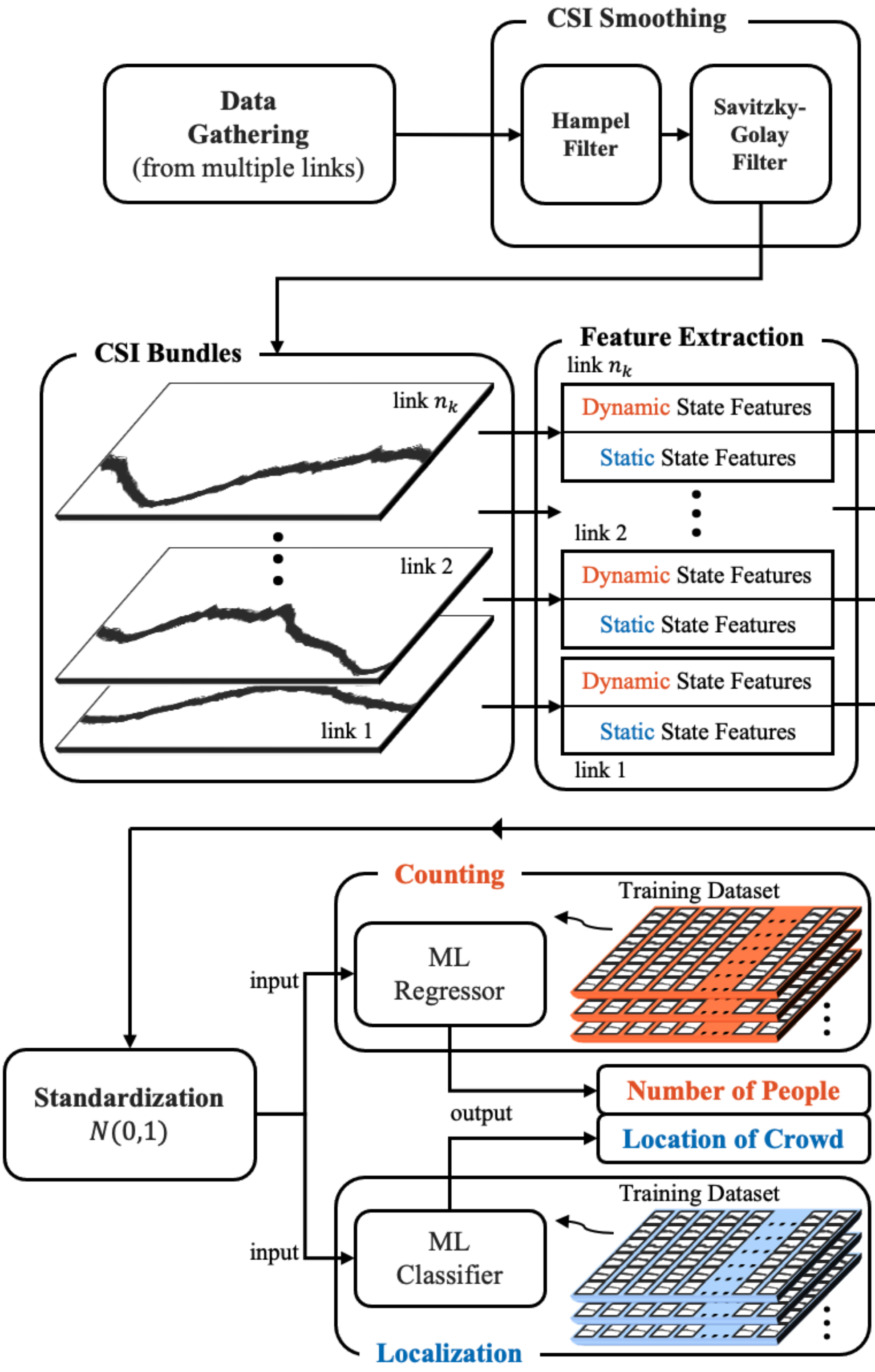


Figure 4.1: Processing Flow in Proposed System.

4.2 CSI Pre-processing

In order to leverage CSI readings as informative and effective resources for crowd estimation, it is essential to pre-process the data before the feature extraction. We present the CSI segmentation and smoothing process in this chapter.

4.2.1 Data Segmentation

After receiving the CSI data which is obtained as a form of complex vector, the system first calculates amplitude values across the entire subcarriers as mentioned in Chapter 3.3. After that, the time-series amplitude values are accumulated and segmented into a given-sized time window. Here, we omit the link index k because all the following CSI processing is identically performed regardless of the link number, then we can define a CSI curve vector $\mathbf{a}^i$ of each packet and a time-series amplitude vector $\mathbf{a}_j$ of each subcarrier as follows:

$$\begin{aligned} \mathbf{a}^i &= (a_{i,1}, \cdots, a_{i,j}, \cdots, a_{i,n_s}) \\ \mathbf{a}_j &= (a_{1,j}, \cdots, a_{i,j}, \cdots, a_{n_p,j})^T \end{aligned} \tag{4.1}$$

where i and j are indices of packet and subcarrier, respectively, and n_s and n_p are the total number of subcarriers and packets in a time window, respectively.

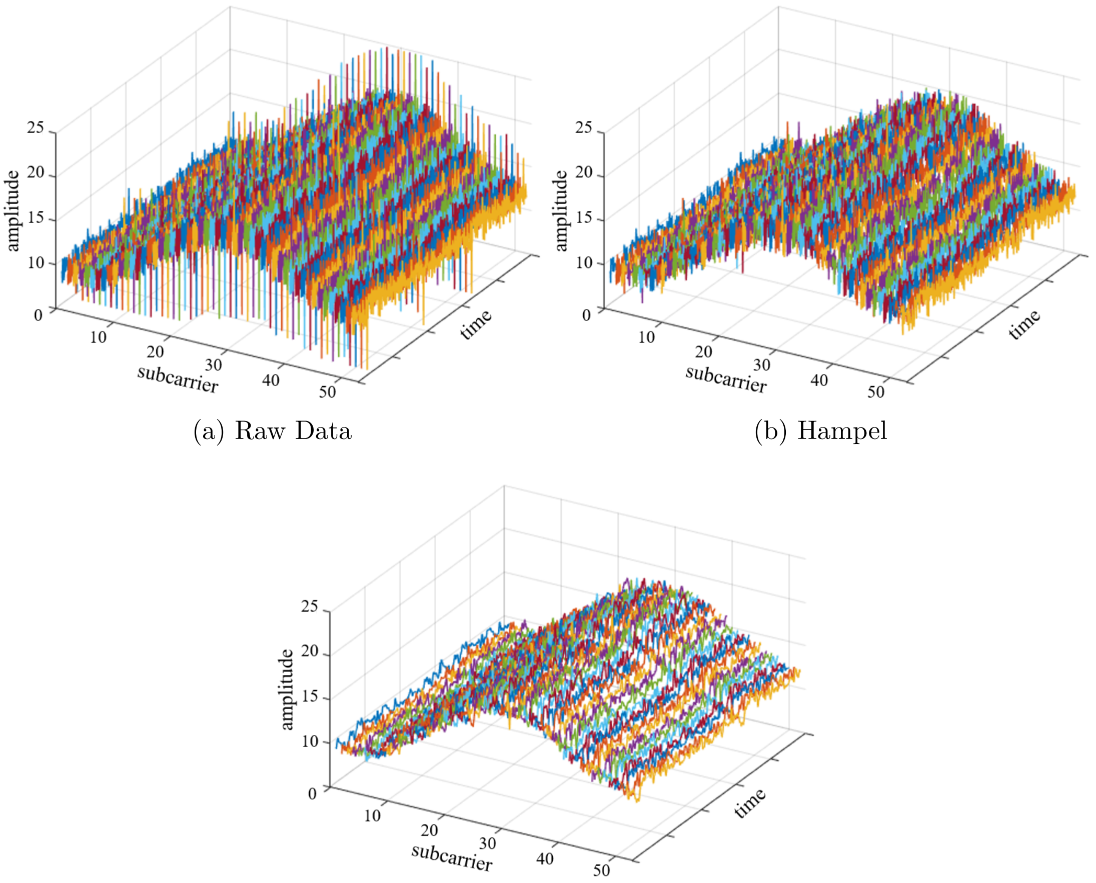
Then, a CSI bundle $\mathbf{A}^{(w)}$ in a time window can be denoted as:

$$\mathbf{A}^{(w)} = \begin{bmatrix} \mathbf{a}_1^{(w)} & \cdots & \mathbf{a}_j^{(w)} & \cdots & \mathbf{a}_{n_s}^{(w)} \end{bmatrix} \tag{4.2}$$

where w is the index of time window. We empirically set each time window to contain six seconds of CSI data with three-seconds overlapping. Since we configure the packet rate of the ESP32 nodes as $100\,packets/sec$, each time window contains 600 packets ($n_p = 600$). Also, we can obtain CSI readings in a total of 52 available subcarriers ($n_s = 52$). This CSI bundle $\mathbf{A}^{(w)}$, which is consisting of CSI curves in a 6 s time window, becomes a base unit for our feature extraction process.

4.2.2 CSI Smoothing

Since the CSI readings are considerably noisy, it is necessary to remove the redundant components from the calculated amplitude values. For this smoothing



(a) Raw Data

(b) Hampel

(c) Hampel and Savitzky-Golay

Figure 4.2: CSI Smoothing Results.

process, we apply two filters, one is Hampel filter for eliminating spike noises, the other is Savitzky-Golay filter for removing overall white noise without distorting the tendency of the signal. These filters are used in several existing studies for WiFi CSI noise reduction because of their low computational cost, as described in [40]. Figure 4.2 shows the amplitudes of the time-series CSI before applying filters, after applying Hampel filter, and after applying both Hampel and Savitzky-Golay filters, respectively.

4.3 Feature Extraction

In this chapter, we describe all the features extracted from the amplitude signal of WiFi CSI for crowd counting and localization. The features are categorized by three extraction sources for each dynamic and static state, as summarized in Table 4.1.

Table 4.1: Extracted Features.

Category	*Dynamic (Counting)*	*Static (Localization)*
Statistical	**std**, **min**, **max**, **qtl**, **qtu**, **avg**	
Bundle-based	**iqr**, **adj**, *euc*	**cur**, **der**
RSS-based	*rss*	-

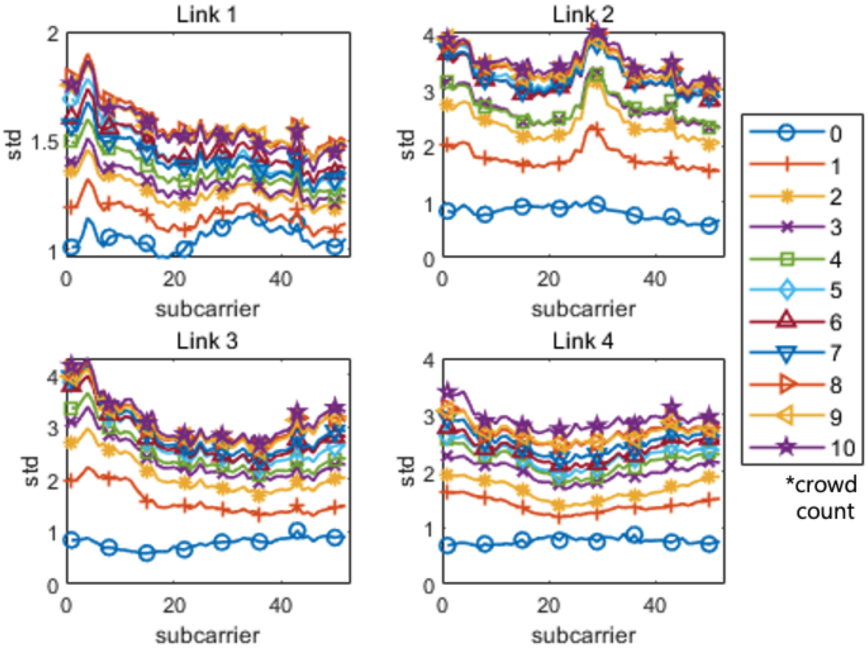
4.3.1 Common Statistical Features

We calculate common statistical features from time-series CSI amplitudes. Several statistical functions are independently applied to each subcarrier signal.

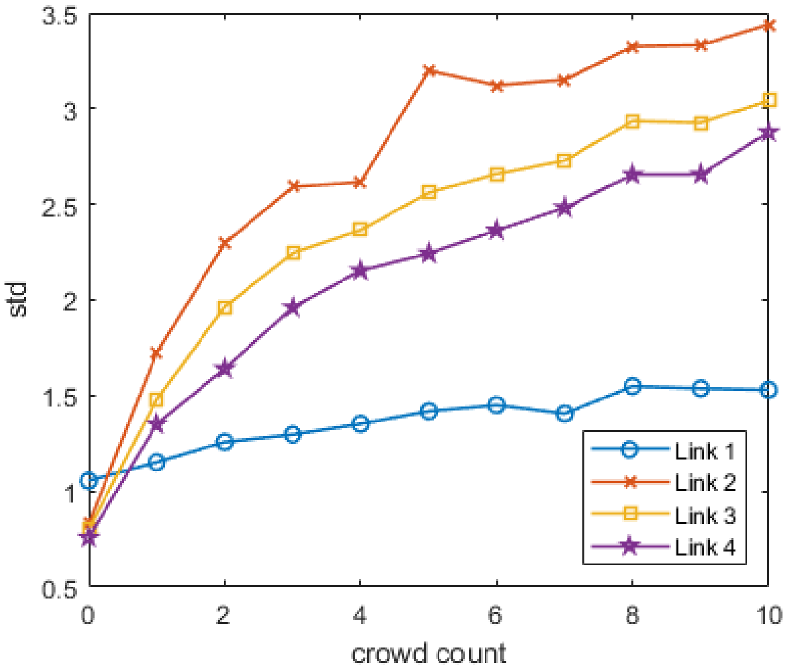
First of all, we can simply use the standard deviation of amplitudes of each subcarrier. Intuitively, the more the number of people between WiFi channels, the more complicated multipath fading channel is formed. This subsequently makes the signal amplitude more severely fluctuate across entire subcarriers than when there are no people in the area. We have checked that the number of people shows the monotonic relationship with the degree of signal fluctuation, as we can see in Figure 4.3. A standard deviation vector of subcarriers $\mathbf{std}^{(w)}$ can be denoted as:

$$\mathbf{std}^{(w)} = (\, \sigma(\mathbf{a}_1^{(w)}), \cdots, \sigma(\mathbf{a}_j^{(w)}), \cdots, \sigma(\mathbf{a}_{n_s}^{(w)})\,) \tag{4.3}$$

where $\sigma(\mathbf{x})$ denotes a function of the standard deviation of any vector $\mathbf{x}$.



(a) Tendency over Subcarriers



(b) Tendency over WiFi Links

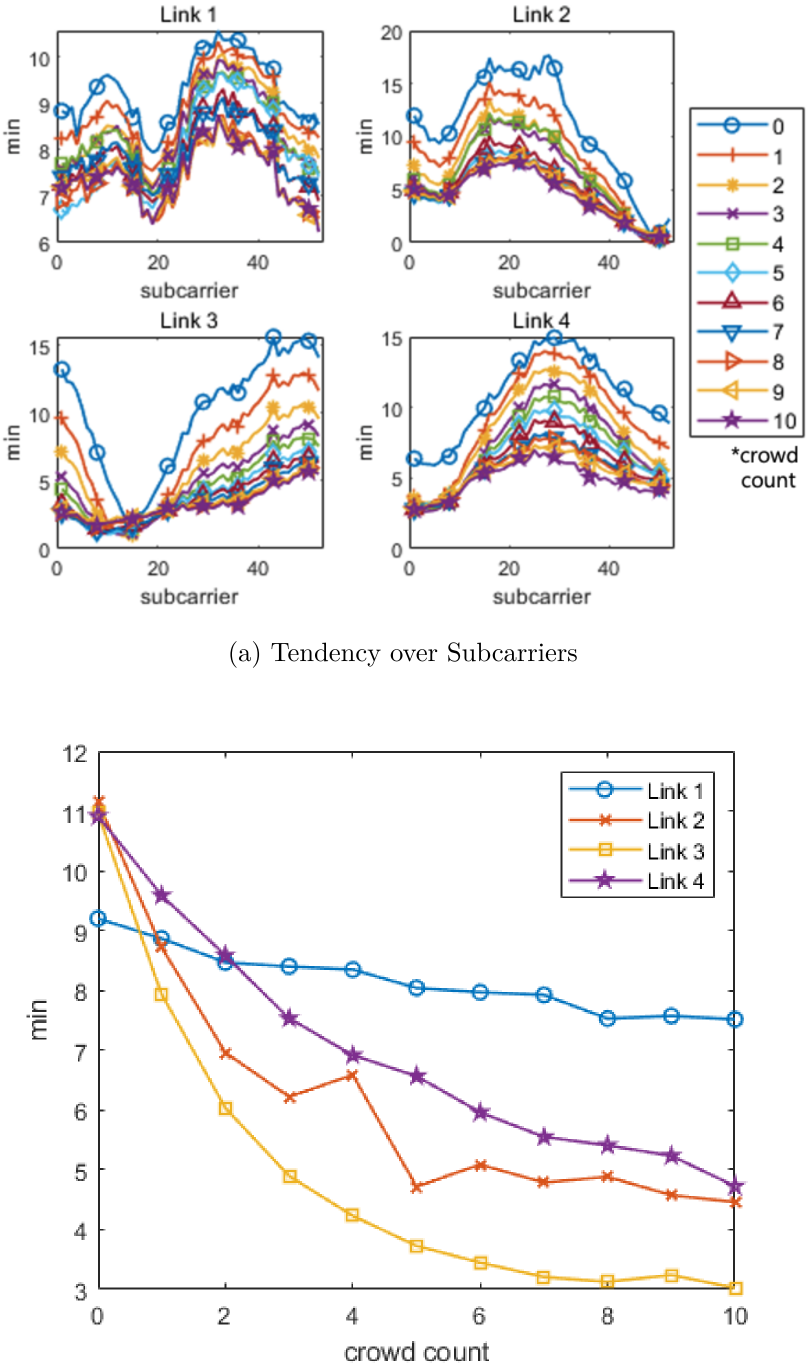
Figure 4.3: Monotonic Relationship between Crowd Count and $\mathbf{std}^{(w)}$.

As we can see from the CSI bundles in Figure 3.6(a) and (b), the uppermost and lowermost CSI curves in a time window gradually rise and go down as the number of people increases. This characteristic is also representing the linearity between crowd size and CSI signals. The CSI minimum of each subcarriers show the highest value when there is no person between a WiFi link. Then, the minimum values are decreased along with crowd count. Conversely, the CSI maximum has the lowest value across all subcarriers at 0 person, and it is gradually increased as the people count increases. We depict the tendency of CSI minima and maxima in Figure 4.4 and 4.5, respectively.

The CSI minima vector $\mathbf{min}^{(w)}$ and maxima vector $\mathbf{max}^{(w)}$ can be denoted as:

$$\begin{aligned}\mathbf{min}^{(w)} &= (\, min(\mathbf{a}_1^{(w)}), \cdots, min(\mathbf{a}_j^{(w)}), \cdots, min(\mathbf{a}_{n_s}^{(w)})\,) \\ \mathbf{max}^{(w)} &= (\, max(\mathbf{a}_1^{(w)}), \cdots, max(\mathbf{a}_j^{(w)}), \cdots, max(\mathbf{a}_{n_s}^{(w)})\,)\end{aligned} \tag{4.4}$$

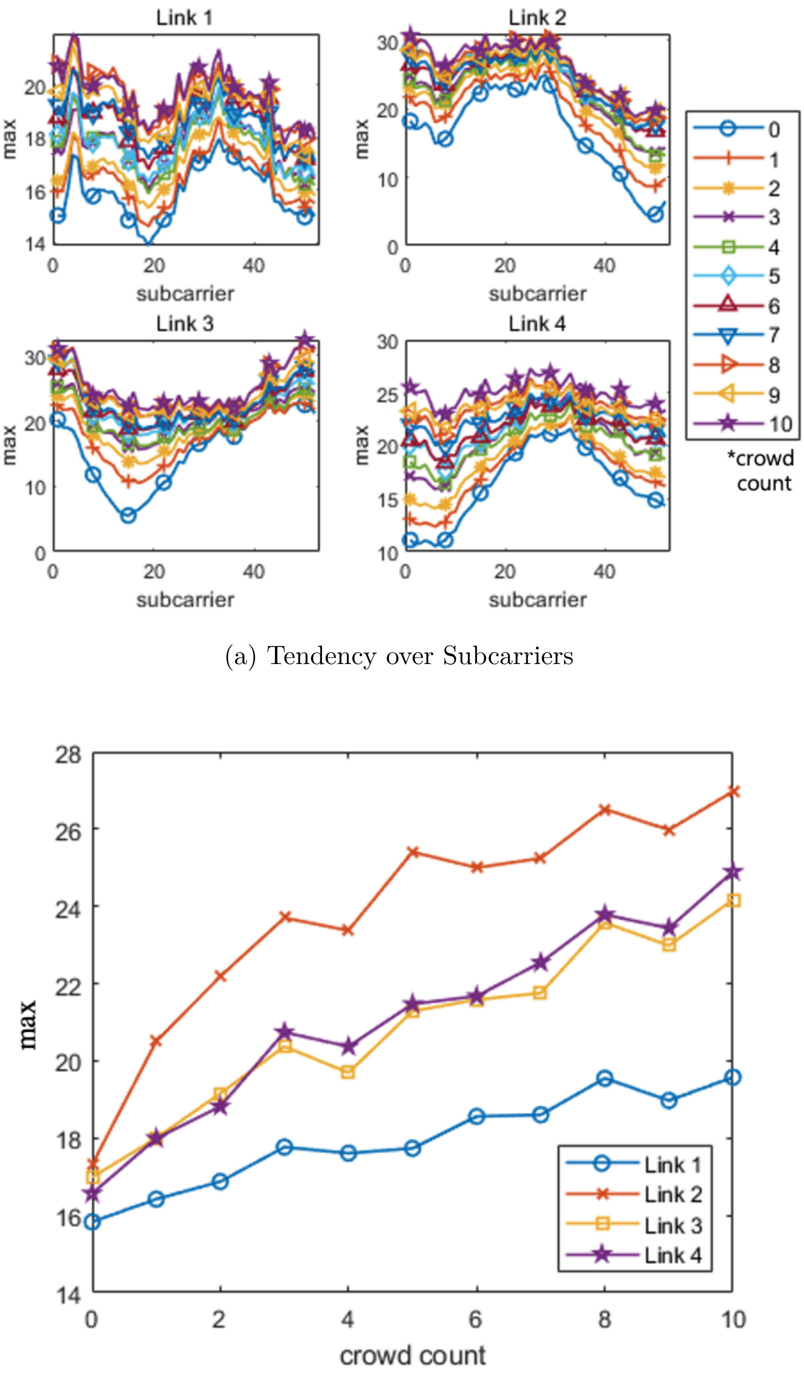
where $min(\mathbf{x})$ and $max(\mathbf{x})$ represent a function of minima and maxima of any vector $\mathbf{x}$, respectively.



(a) Tendency over Subcarriers

(b) Tendency over WiFi Links

Figure 4.4: Monotonic Relationship between Crowd Count and $\mathbf{min}^{(w)}$.



(a) Tendency over Subcarriers

(b) Tendency over WiFi Links

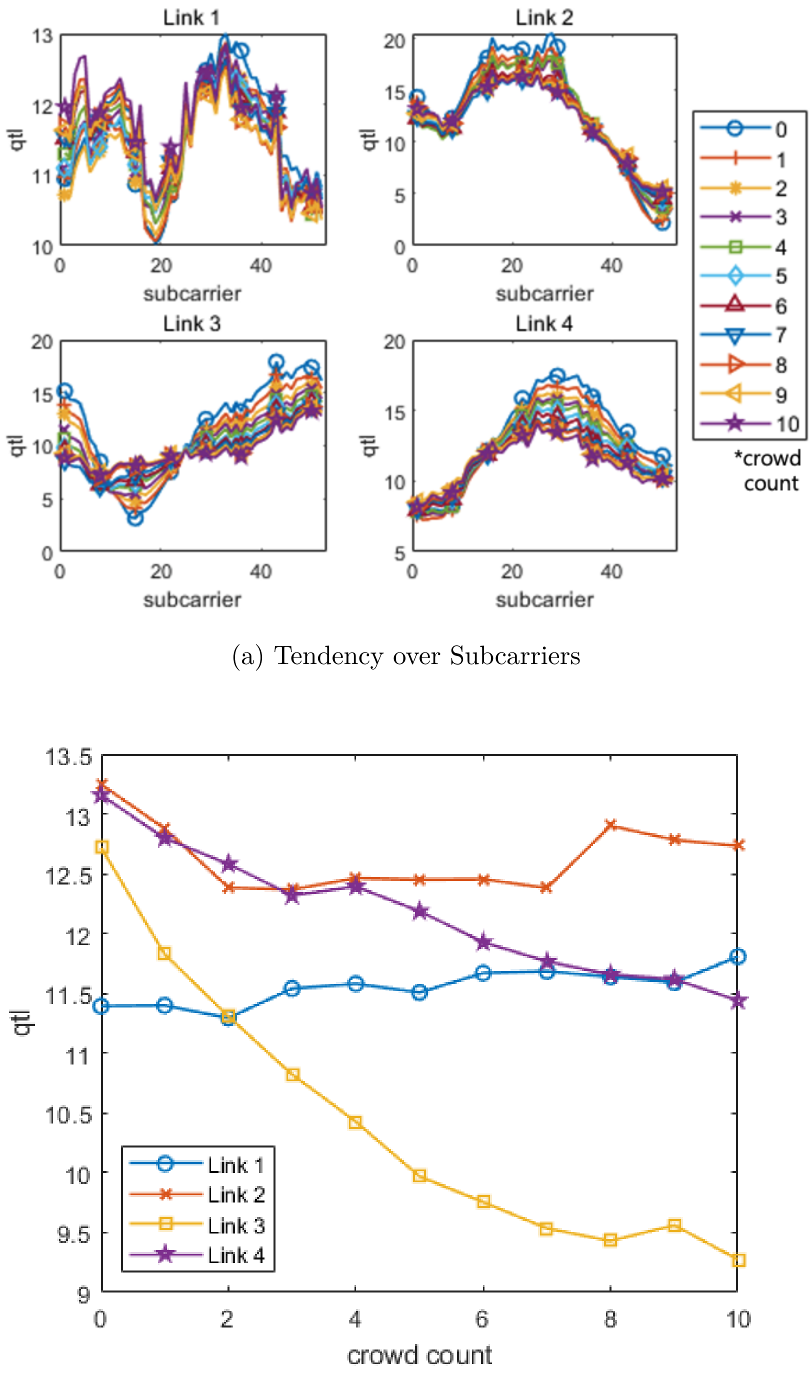
Figure 4.5: Monotonic Relationship between Crowd Count and $\mathbf{max}^{(w)}$.

Similarly, we suppose that the lower and upper quartile values of entire subcarriers also show linear downward and upward trends along with the increased number of people. The reason why we choose the quartiles rather than other percentiles is that we wanted to minimize the effect of noisy values in CSI data. The lower and upper quartiles represent the 25 and 75 percent point of measured data, and we suppose that the data range of 25-75 percent can well reflect the width of a CSI bundle. As we can see in Figure 4.6, similar to minimum values, the lower quartile goes down as crowd count increases. On the other hand, the upper quartile rise up as the number of people increases, as described in Figure 4.7.

We can denote the lower quartile $\mathbf{qtl}^{(w)}$ and the upper quartile $\mathbf{qtu}^{(w)}$ as:

$$\begin{aligned}\mathbf{qtl}^{(w)} &= (\, q_1(\mathbf{a}_1^{(w)}), \cdots, q_1(\mathbf{a}_j^{(w)}), \cdots, q_1(\mathbf{a}_{n_s}^{(w)})\,) \\ \mathbf{qtu}^{(w)} &= (\, q_3(\mathbf{a}_1^{(w)}), \cdots, q_3(\mathbf{a}_j^{(w)}), \cdots, q_3(\mathbf{a}_{n_s}^{(w)})\,)\end{aligned} \tag{4.5}$$

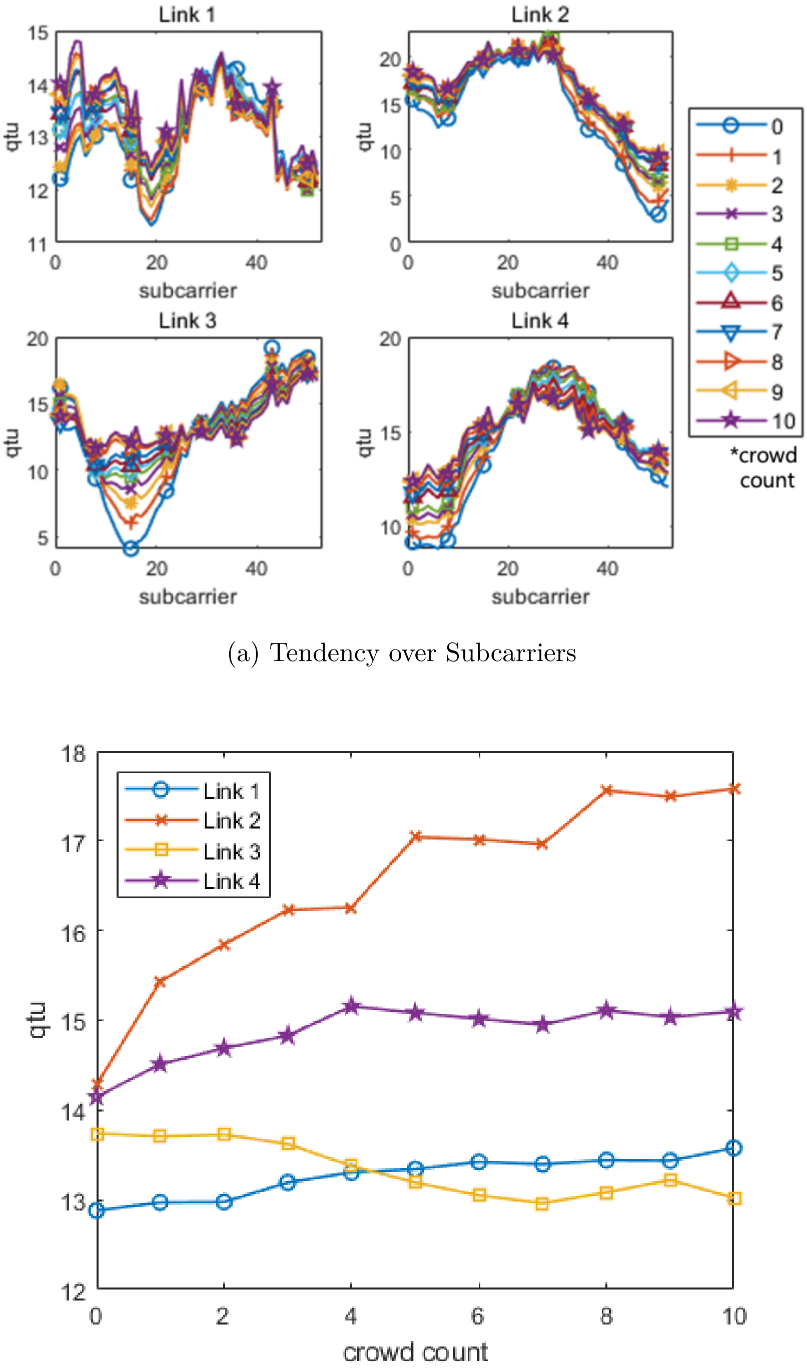
where $q_1(\mathbf{x})$ and $q_3(\mathbf{x})$ denote a function of the first quartile and the third quartile of any vector $\mathbf{x}$, respectively.



(a) Tendency over Subcarriers

(b) Tendency over WiFi Links

Figure 4.6: Monotonic Relationship between Crowd Count and $\mathbf{qtl}^{(w)}$.



(a) Tendency over Subcarriers

(b) Tendency over WiFi Links

Figure 4.7: Monotonic Relationship between Crowd Count and $\mathbf{qtu}^{(w)}$.

The average line of a CSI bundle shows the general shape of CSI curves in a time window. This mean vector across entire subcarriers mainly contributes to the localization part of the system, because it reflects a particular shape of bundles to the learning model depending on a specific section in the area of interest that the crowd is gathered at. The mean vector $\mathbf{avg}^{(w)}$ can be denoted as:

$$\mathbf{avg}^{(w)} = (\, \mu(\mathbf{a}_1^{(w)}), \cdots, \mu(\mathbf{a}_j^{(w)}), \cdots, \mu(\mathbf{a}_{n_s}^{(w)})\,) \tag{4.6}$$

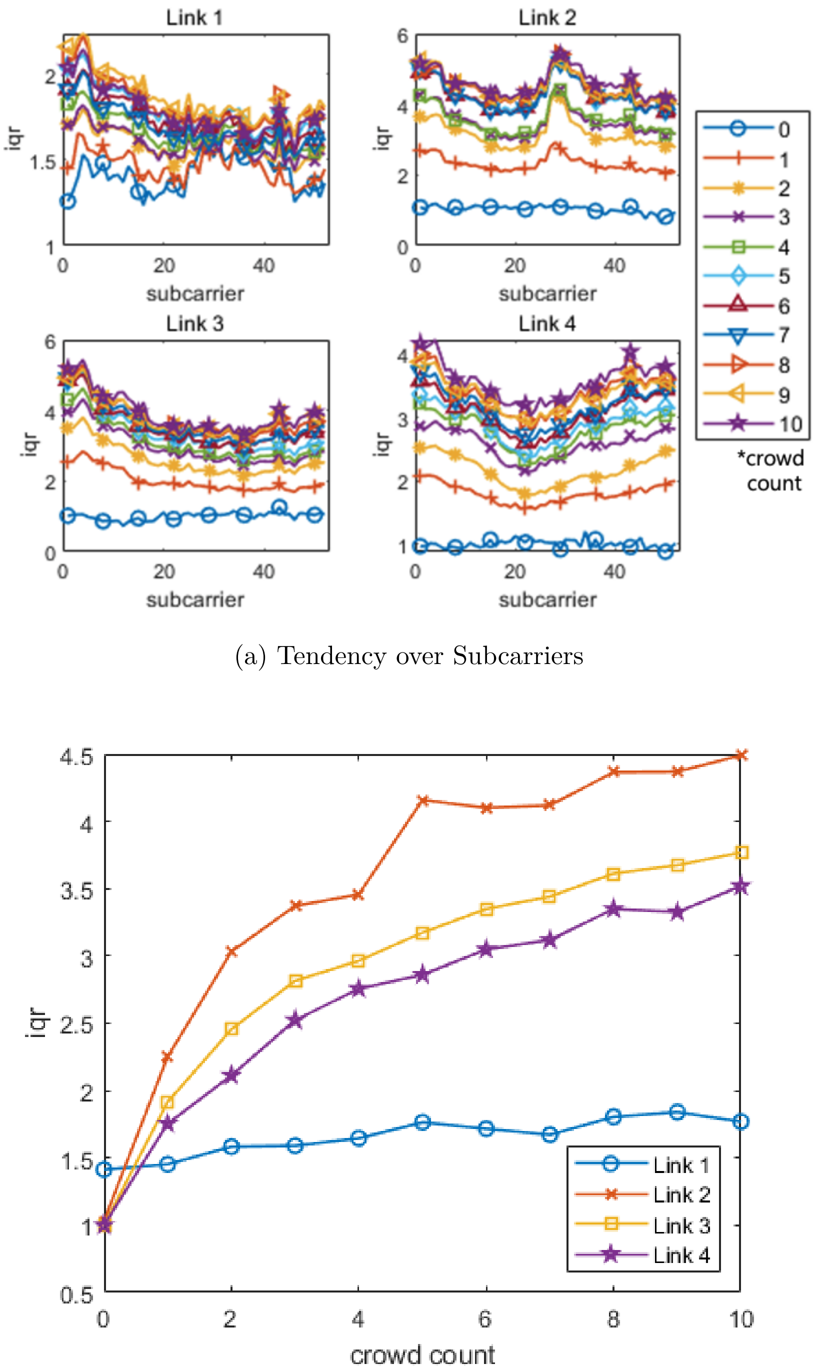
where $\mu(\mathbf{x})$ is a function of the mean value of any vector $\mathbf{x}$.

4.3.2 CSI bundle-based Features

It is necessary to figure out a way to enhance our system's performance with some more effective features as well as statistical ones. Therefore, we now address the features which can be extracted from the CSI bundles.

The interquartile range (IQR) is the width between the lower quartile and upper quartile. The values of the lower quartile and upper quartile mutually inversely go down and up as the number of people between a WiFi link increases, consequently, the IQR also increases as we can see in Figure 4.8. We can obtain an IQR vector that intuitively implies the vertical width of a CSI bundle by the subtraction of upper and lower quartiles as:

$$\mathbf{iqr}^{(w)} = \mathbf{qtu}^{(w)} - \mathbf{qtl}^{(w)} \tag{4.7}$$



(a) Tendency over Subcarriers

(b) Tendency over WiFi Links

Figure 4.8: Monotonic Relationship between Crowd Count and $\mathbf{iqr}^{(w)}$.

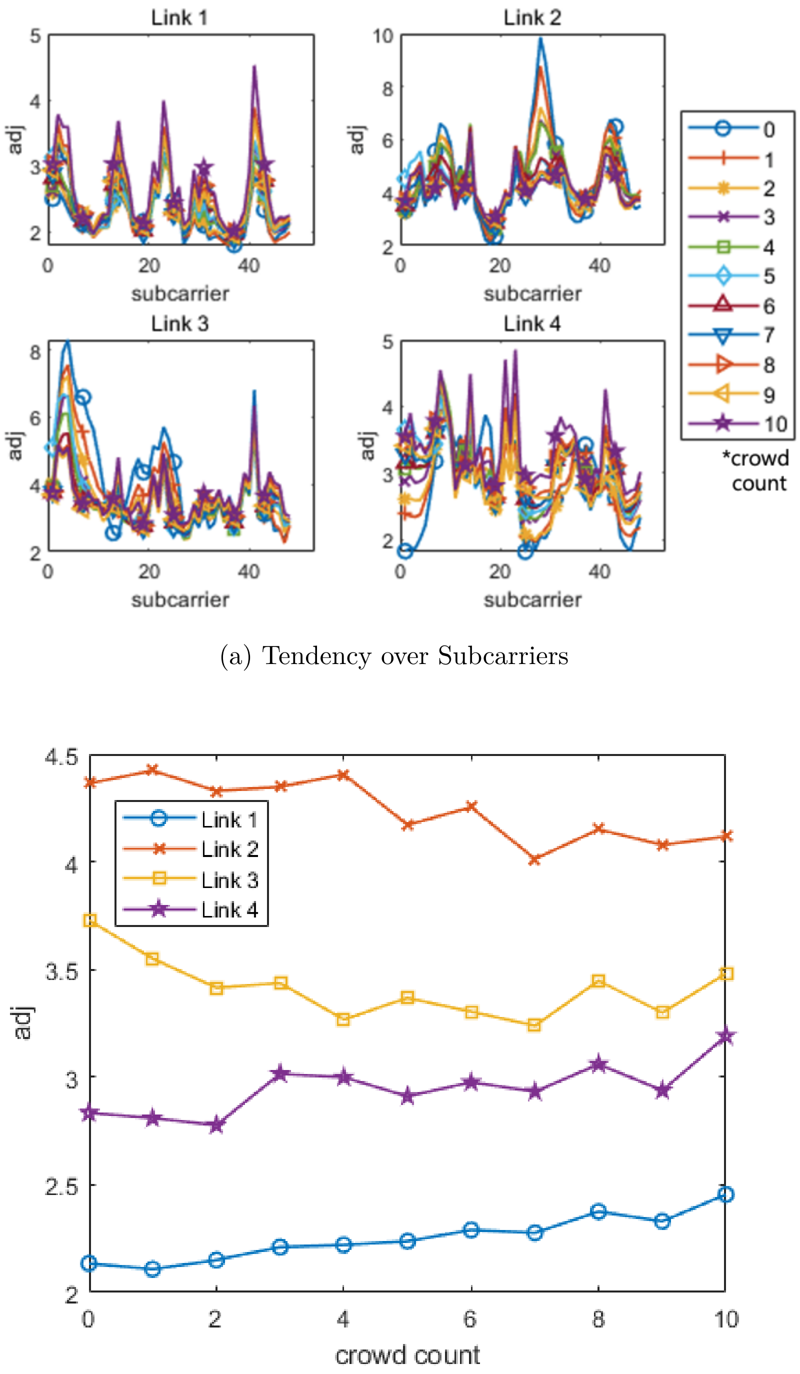
The amplitude difference with adjacent subcarriers is the summation of the absolute differences between one subcarrier and adjacent subcarriers on both sides. It reflects the relationship between adjacent subcarriers to the ML model, in terms of lightly-varying or heavily-varying subcarriers depending on the state of measuring space, as we can see in Figure 4.9. This difference with adjacent subcarriers **adj** is denoted as:

$$\mathbf{adj}^{(w)} = (\, \mu(\boldsymbol{\zeta}^{(w)}_{1+N}), \cdots, \mu(\boldsymbol{\zeta}^{(w)}_{j}), \cdots, \mu(\boldsymbol{\zeta}^{(w)}_{n_s-N})\,) \tag{4.8}$$

where

$$\begin{aligned} \boldsymbol{\zeta}^{(w)}_{j} &= (\, \zeta^{(w)}_{1,j}, \cdots, \zeta^{(w)}_{i,j}, \cdots, \zeta^{(w)}_{n_p,j}\,)^T, \\ \zeta^{(w)}_{i,j} &= \sum_{n=1}^{N} (\, |a^{(w)}_{i,j} - a^{(w)}_{i,j-n}| + |a^{(w)}_{i,j} - a^{(w)}_{i,j+n}|\,) \end{aligned} \tag{4.9}$$

where N is the number of adjacent subcarriers on both sides which will be included in **adj** calculation. In this thesis, we decide as $N = 2$ through the empirical test.



(a) Tendency over Subcarriers

(b) Tendency over WiFi Links

Figure 4.9: Monotonic Relationship between Crowd Count and $\mathbf{adj}^{(w)}$.

Euclidean distance between CSI curve vectors from adjacent packets also contains information of how intensely the multipath fading channel is changing. The Euclidean distance maintains relatively low values when a channel is not being interrupted by moving people, but the larger crowd in the channel makes the value gradually increase, as we can see in Figure 4.10. Let $med(\mathbf{x})$ be a function of the median value of any vector $\mathbf{x}$, then the median of Euclidean distances in a time window euc can be denoted as:

$$euc^{(w)} = med(\,\epsilon_1^{(w)}, \cdots, \epsilon_i^{(w)}, \cdots, \epsilon_{n_p-1}^{(w)}\,) \tag{4.10}$$

where

$$\epsilon_i^{(w)} = ||\mathbf{a}_{(w)}^{i+1} - \mathbf{a}_{(w)}^{i}|| \tag{4.11}$$

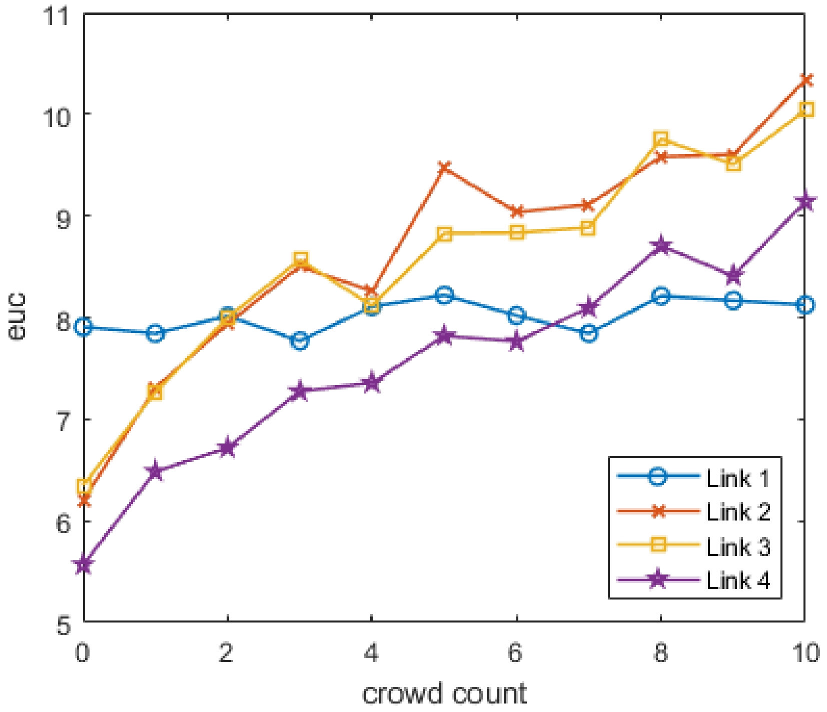


Figure 4.10: Monotonic Relationship between Crowd Count and $euc^{(w)}$.

In localization, we use coefficients of the fitted polynomial curve of CSI bundle's average line ($\mathbf{cur}^{(w)}$) and its 1st derivative function ($\mathbf{der}^{(w)}$), to leverage a particular shape of the CSI bundle as a feature for localization. $\mathbf{cur}^{(w)}$ reflects the shape of the CSI bundle itself, and $\mathbf{der}^{(w)}$ clarifies at which points of the fitted curve have peaks, valleys, or sharp slopes. We empirically apply the curve fitting with a 6-term polynomial curve, then we use its polynomial coefficients as the features. Therefore, $\mathbf{cur}^{(w)}$ and $\mathbf{der}^{(w)}$ feature vectors contain six and five components, respectively.

4.3.3 RSS-based Features

Lastly, we use RSS measurements which are measured with CSI readings. WiFi RSS also shows a monotonic relation between its variation and the number of people within the link coverage similar to statistical features of CSI, as depicted in Figure 4.11. If we define ρ as an RSS measurement of a packet, the standard deviation of RSS in a time window $rss^{(w)}$ can be denoted as:

$$rss^{(w)} = \sigma(\,\rho_1^{(w)}, \cdots, \rho_i^{(w)}, \cdots, \rho_{n_p}^{(w)}\,) \tag{4.12}$$

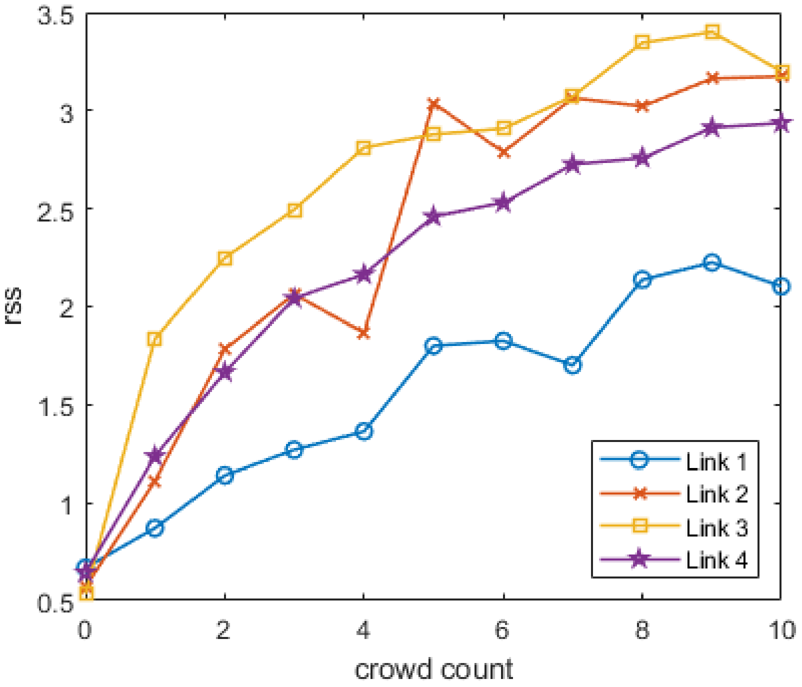


Figure 4.11: Monotonic Relationship between Crowd Count and $rss^{(w)}$.

4.4 Standardization & Learning Models

The extracted features are concatenated to form the datasets for training each machine learning model of crowd counting and localization. In this study, we treat counting and localization as regression and classification problems, respectively. Each feature vector or feature value is connected vertically along the order of time windows and horizontally along the order of links, for example, a feature matrix of standard deviation **STD** can be denoted as:

$$\mathbf{STD} = \begin{bmatrix} \mathbf{std}_{k=1}^{(1)} & \mathbf{std}_{k=2}^{(1)} & \cdots & \mathbf{std}_{k=n_k}^{(1)} \\ \mathbf{std}_{k=1}^{(2)} & \mathbf{std}_{k=2}^{(2)} & \cdots & \mathbf{std}_{k=n_k}^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{std}_{k=1}^{(nw)} & \mathbf{std}_{k=2}^{(nw)} & \cdots & \mathbf{std}_{k=n_k}^{(nw)} \end{bmatrix} \tag{4.13}$$

where n_k and n_w are the total number of WiFi links in the system ($n_k = 4$ in this work) and the total number of time windows for training, respectively. Equally, other feature matrices such as $\mathbf{MIN}, \mathbf{MAX}, \cdots, \mathbf{RSS}$ are also produced by the same procedure. Then, all the feature matrices are lined up from side to side becoming the final training dataset.

After the formation of training data, all datasets are standardized by standard normal distribution $N(0, 1)$ to fit the scales between different features before training. Then, machine learning regressors and classifiers are trained with the datasets to evaluate the performance of simultaneous crowd estimation. In this thesis, we examine counting performance with linear regressor (LR), random forest regressor (RFR), XGBoost regressor (XR) and LightGBM regressor (LGBMR), and localization performance with Random Forest classifier (RFC), Logistic Regression classifier (LRC), support vector classifier (SVC) and LightGBM classifier (LGBMC). Furthermore, we construct the DNN models, namely DNN regressor (DNNR) and DNN classifier (DNNC), to check and provide the differences and comparisons, and pros and cons compared to conventional ML models.

5 Performance Evaluation

In this chapter, we present experimental setup, data gathering scheme, and several comparisons depending on learning models and adjustable parameters, then evaluate the system performance through the experiments at two difference-size rooms.

5.1 Experimental Setup

We collected the CSI data through a multi-scenario experiment with up to five participants in a small-sized meeting room and up to 10 participants in a medium-sized seminar room. Unlike conventional researches that a single pair of WiFi routers were usually installed using Intel or Atheros CSI solutions, we placed the four pairs of ESP32 nodes to make four WiFi links vertically, horizontally, and diagonally crossing over the target area. This enables the system to faithfully observe the change of CSI measurements with regard to the movement of walking people covering the whole target area. For our experiment, all transmitters were set to send the CSI request packets to their pair receivers at 100 Hz of packet rate. We performed our experiments in a small-sized meeting room (5.5 m by 5.5 m) and a medium-sized seminar room (11 m by 5.5 m) which were equally divided into four sections for assessment of crowd localization. In both rooms, each WiFi link k consists of AP k (Tx) and STA k (Rx).

5.1.1 WiFi Link Setting

After we set the SSID, password, and packet rate for all ESP32 nodes, we attach the nodes onto the given position with 1m height on the wall in the experiment area. Each AP node is connected to a laptop to transfer and save the CSI data in the files. Here, we set the baud rate of serial communication as 460800 bits per second. This also can be altered to SD card-based data saving to remove the laptop and data cable, but we used the ESP32 nodes without SD card slot. Station node is only connected to mobile battery for power input. Figure 5.1 shows the whole device set of one WiFi link.

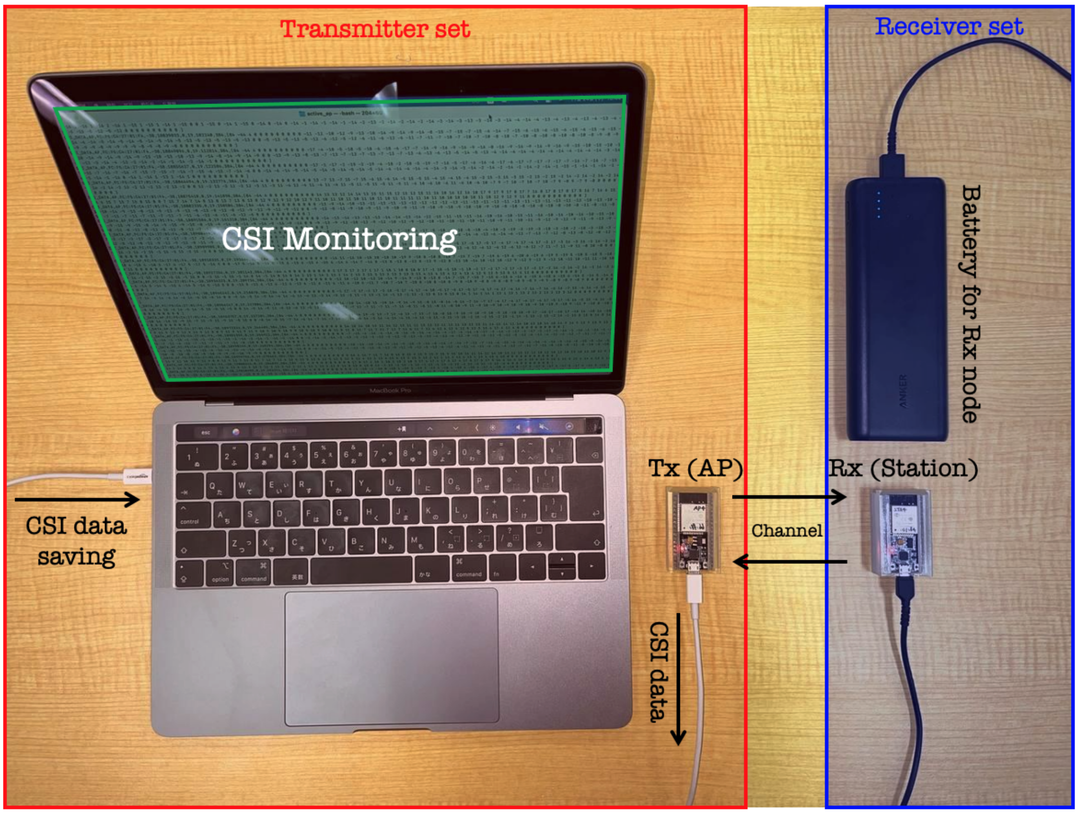


Figure 5.1: Experiment Device Set.

5.1.2 Small-sized Meeting Room

We first performed a small-scale crowd estimation experiment in a meeting room in the campus building, which has square-shape with 5.5 m of length and width. As we can see in Figure 5.2, we installed four WiFi links at all corners and walls, making two horizontal links and two diagonal links which are crossing over each other link. We assign link number 1 and 2 for the horizontal links, and 3 and 4 for the diagonal links. We divide the test room into four identical-sized sections for the localization part. To make more realistic sensing environment, we put four desks and eight chairs at the middle of the room. In this meeting room, we conducted the entire-area walks and in-section walks with five participants as we will describe in Chapter 5.2. Figure 5.3 shows the actual scene of our experiment in meeting room.

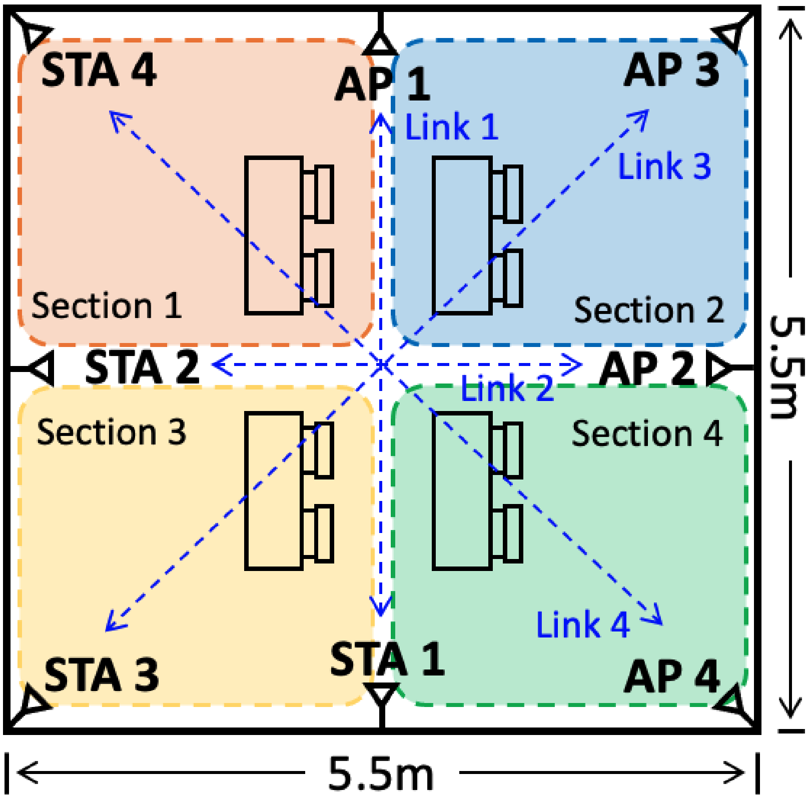


Figure 5.2: Plane of Experiment Area (Meeting Room).

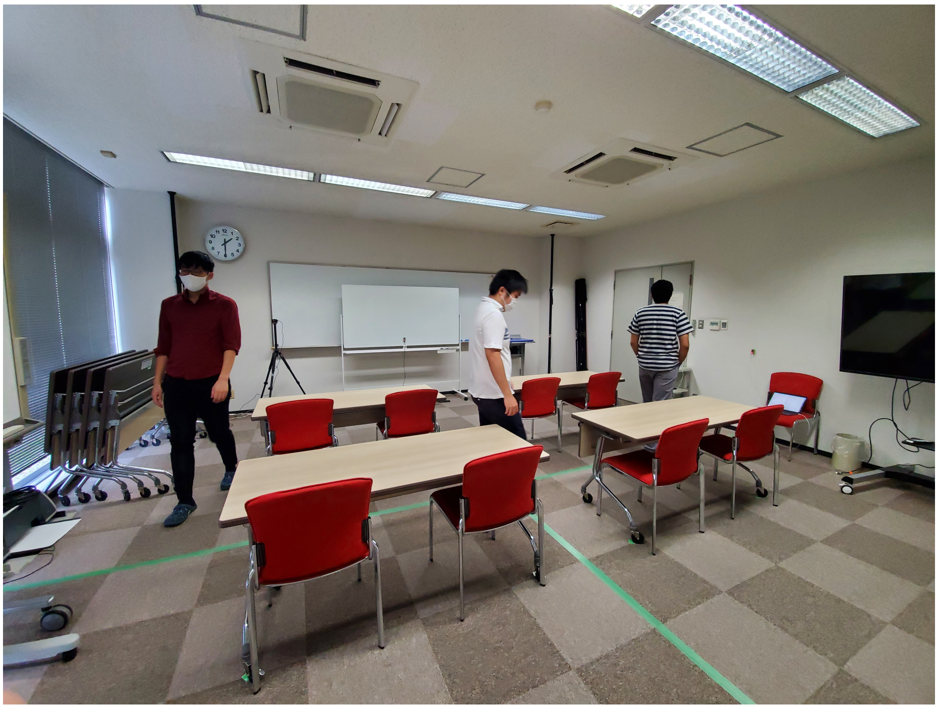


Figure 5.3: Scene of Experiment (Meeting Room).

5.1.3 Medium-sized Seminar Room

We conducted the secondary experiment to examine the system is able to afford the more number of people and the bigger area of interest. The medium-sized seminar room has 11 m length and 5.5 m width. Same as the meeting room, four WiFi links were installed making two horizontal and two vertical link, same link numbers were assigned, and also the test room was divided into four equivalent sections, as depicted in Figure 5.4. In seminar room, we put the six desks and 12 chairs making rectangle shape around the center of the room. In total of 10 people participated for this seminar room experiment, and performed the same scenarios as the meeting room experiment. Figure 5.5 shows the actual scene of the experiment in seminar room.

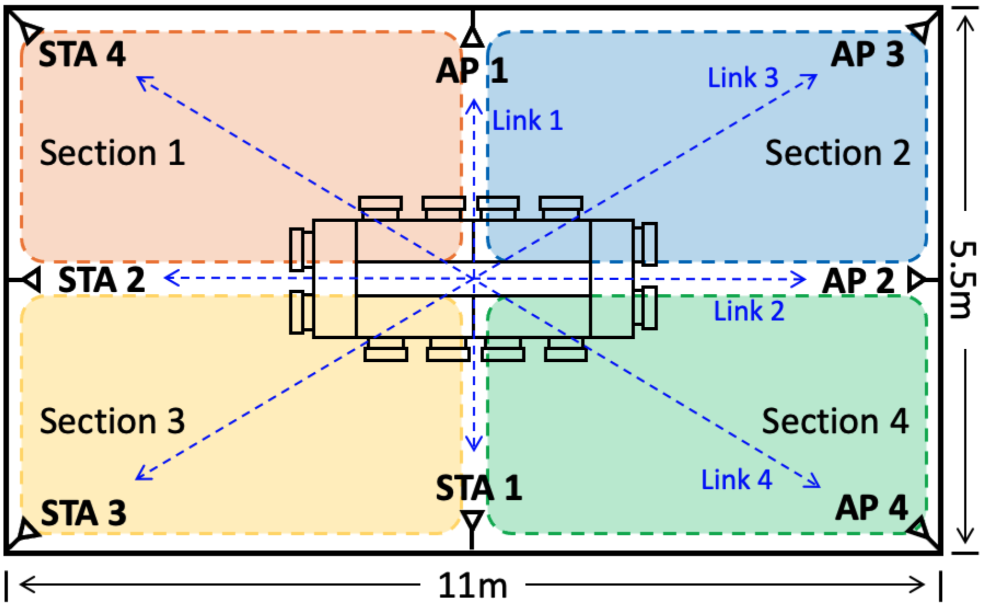


Figure 5.4: Plane of Experiment Area (Seminar Room).

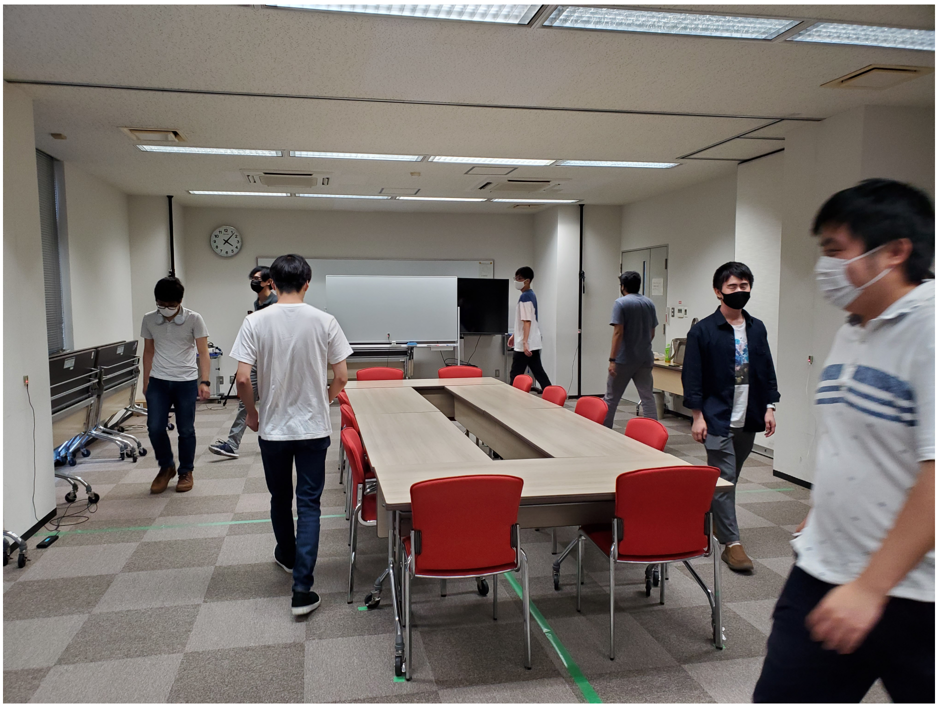


Figure 5.5: Scene of Experiment (Seminar Room).

5.2 Data Gathering Scheme

To confirm effectiveness of our insight of simultaneous crowd estimation, we designed and conducted the experiments which contain five scenarios with a certain number of people walking in an experiment area. Here, five scenarios mean the situations that the cluster of people is walking at different sections of the area. The number of people are denoted as $P_{n_{peo}}$ ($n_{peo} = 0, 1, \cdots, 5$ in the meeting room, $n_{peo} = 0, 1, \cdots, 10$ in the seminar room), and the scenarios related to the section number correspond to $S_{n_{sect}}$ ($n_{sect} = 1, \cdots, 4$, and *oth* that indicates *other* pattern, i.e., full-area walk). To be specific, the scenarios $S_{n_{sect}}$ are corresponding to the situation in which the participants walk freely within a particular section n_{sect}. In the scenario S_{oth}, on the contrary, the participants perform free walking all over the experiment area. The examples of $P_{n_{peo}}/S_{n_{sect}}$ scenarios scenarios are depicted in Figure 5.6 and 5.7. In every section walking (S_1, S_2, S_3, S_4, and S_{oth}), all the participants walk randomly within the given space, without any guidance/limitation about how to walk.

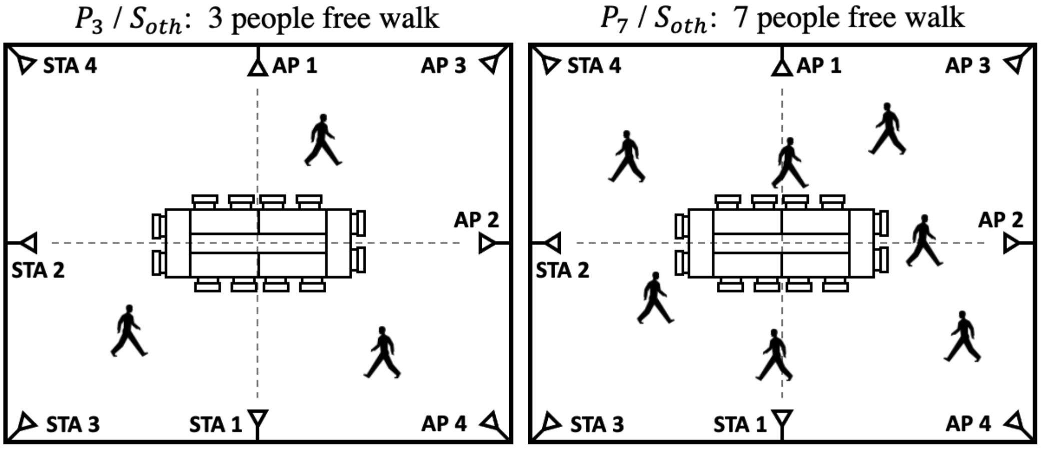

P_3 / S_{oth}: 3 people free walk
STA 4
AP 1
AP 3
AP 2
STA 2
STA 3
STA 1
AP 4
P_7 / S_{oth}: 7 people free walk
STA 4
AP 1
AP 3
AP 2
STA 2
STA 3
STA 1
AP 4

Figure 5.6: Examples of Crowd-size Cases (P_3 and P_7 in S_{oth}).

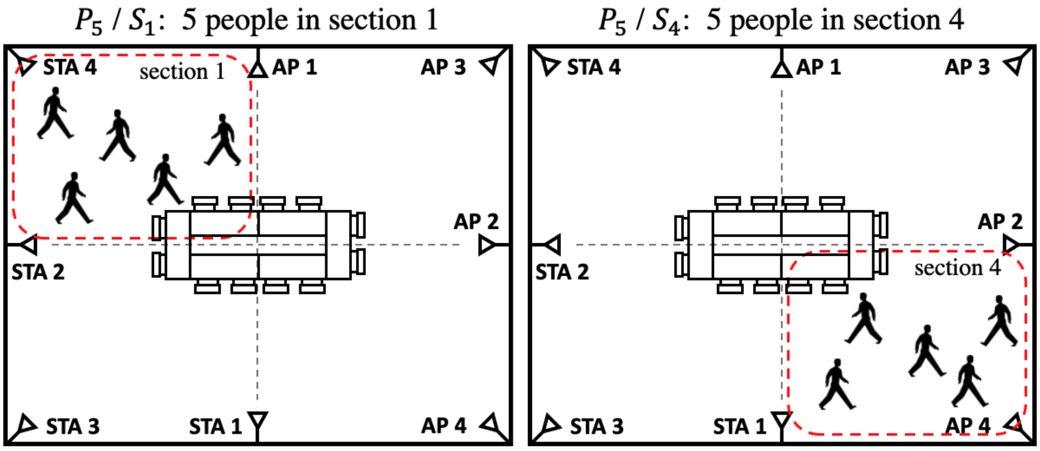


Figure 5.7: Examples of Walking Scenarios (S_1 and S_4 with P_5).

We collected two minutes of CSI data in each scenario of all combinations of $P_{n_{peo}}$ and $S_{n_{sect}}$. That is, a total of 60-minute-data (2 mins × 5 sections × 0-5 people) in the meeting room and 110-minute-data (2 mins × 5 sections × 0-10 people) in the seminar room were collected in a single experiment. Then, we need to regroup and combine the collected data into two datasets each for crowd counting and localization. Figure 5.8 shows the data regrouping way by seminar room data. The datasets of $P_{n_{peo}}$ are constructed with all the scenarios data collected with n_{peo} people. These 10 minutes-long datasets are used for crowd counting. On the other hand, the datasets of $S_{n_{sect}}$ include all the data of when the 1-5 participants are walking in section n_{sect}. These 20 minutes-long datasets become the sources for crowd localization.

We carried out three times of identical experiments in each of the meeting room and the seminar room, in different days. This is to check the difference in system performance originating from circumstance changes, such as temperature, humidity, or signal interference. The experiments in different days are distinguished as Session 1, 2, and 3.

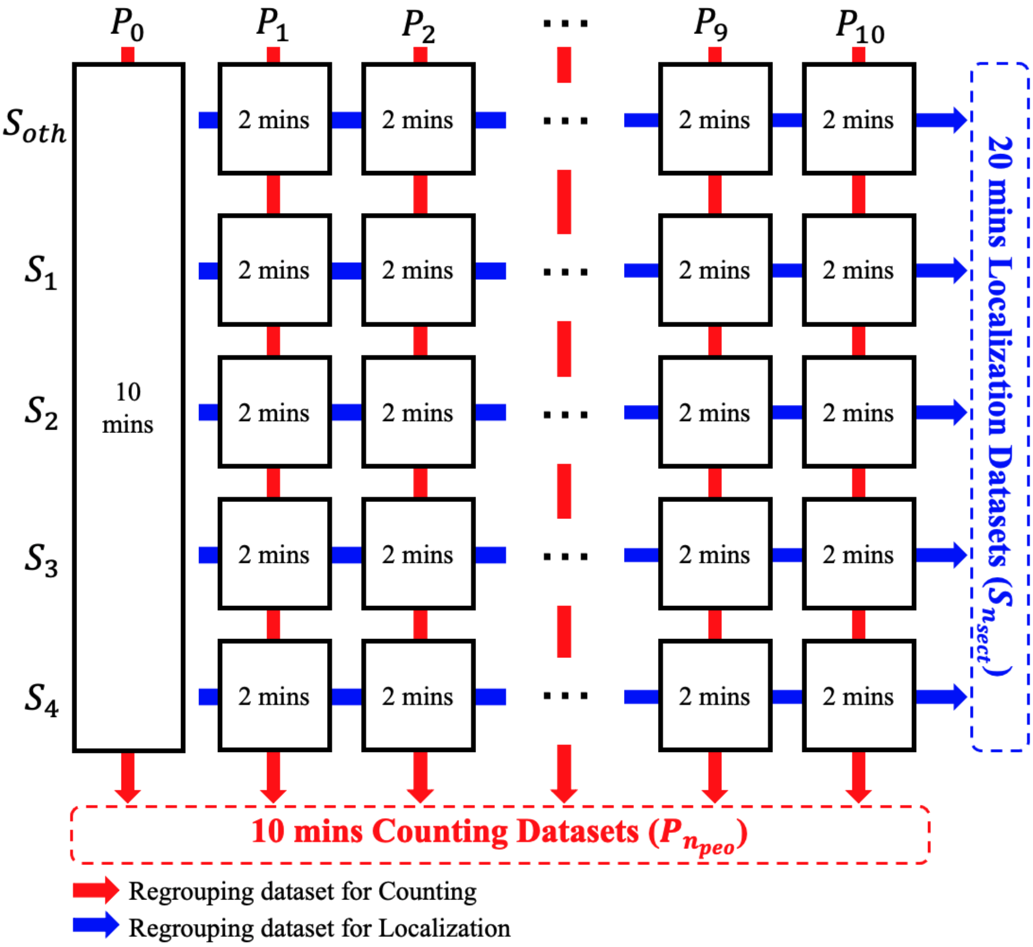

P_0
P_1
P_2
...
P_9
P_{10}
S_{oth}
S_1
S_2
S_3
S_4
10 mins
2 mins
20 mins Localization Datasets ($S_{n_{sect}}$)
10 mins Counting Datasets ($P_{n_{peo}}$)
Regrouping dataset for Counting
Regrouping dataset for Localization

Figure 5.8: Dataset Regrouping.

As an example scene of meeting room experiment, Figure 5.9 and 5.10 show the all actual scenes of data gathering scenarios with two people. The scenario of entire-area walking (S_{oth}) is presented in Figure 5.9. The participants walked randomly all over the area without bias as much as possible. On the contrary, they sequentially performed in-section walking from section 1 to section 4, as presented in Figure 5.10. The participants randomly walked at in-section walks as well, but within the given section $S_{n_{sect}}$.

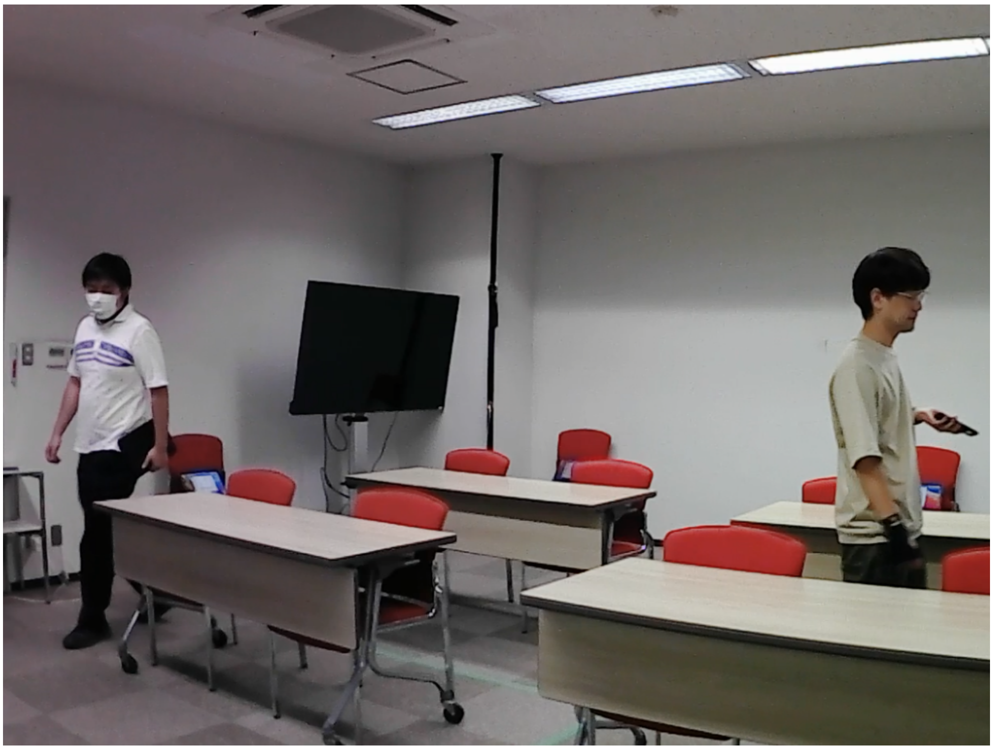
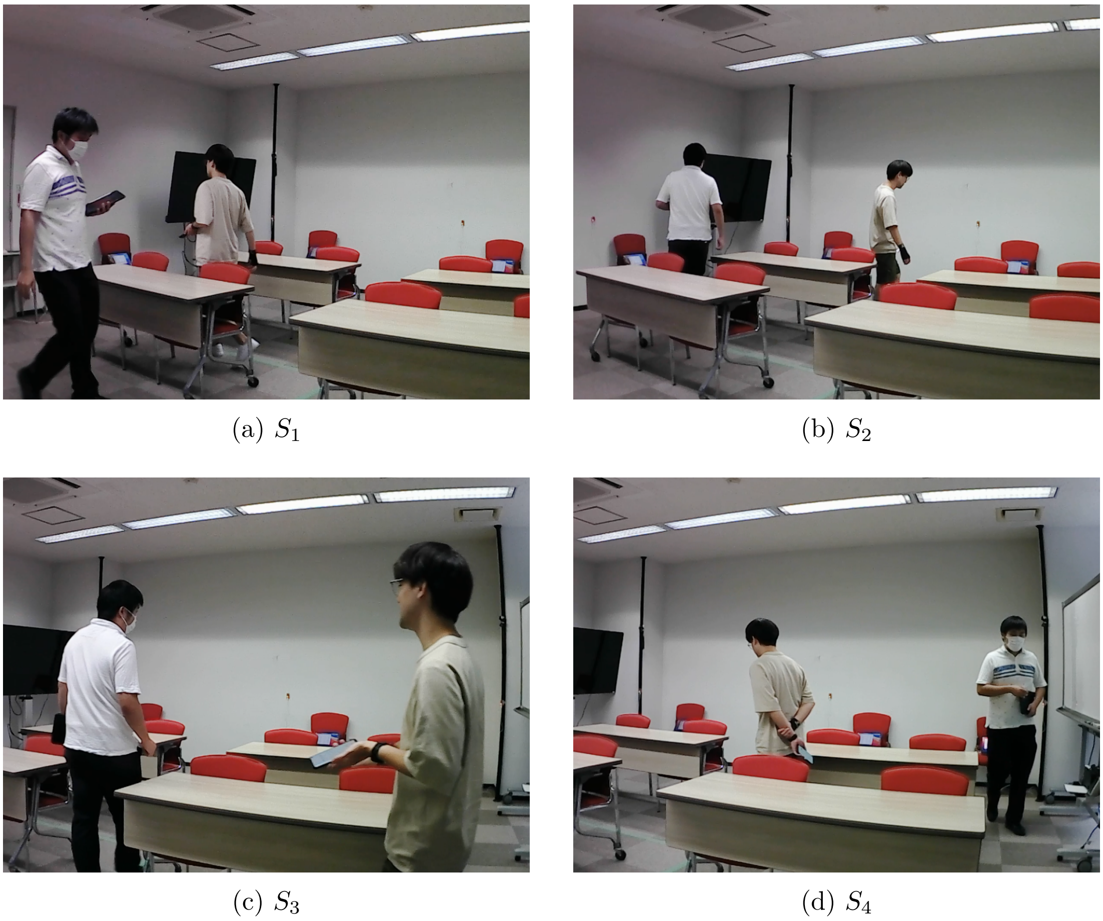


Figure 5.9: Actual Scene of S_{oth} with P_2 (Meeting Room).



(a) S_1

(b) S_2

(c) S_3

(d) S_4

Figure 5.10: Actual Scenes of each scenario with P_2 (Meeting Room).

As an example scene of seminar room experiment, the actual scenes of data gathering scenarios with seven people in the seminar room experiment are presented in Figure 5.11 and 5.12. Figure 5.11 shows the entire-area walking senario, and Figure 5.12 presents the in-section walking of each section $S_{n_{sect}}$. The way of participants' walking were identical with the meeting room experiment.

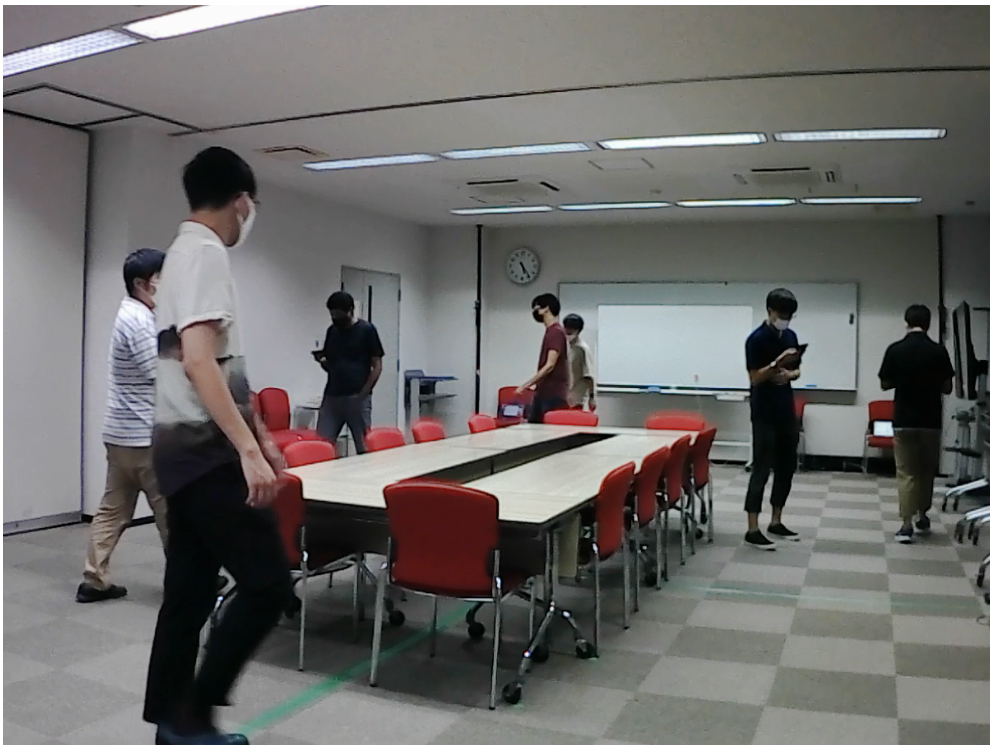
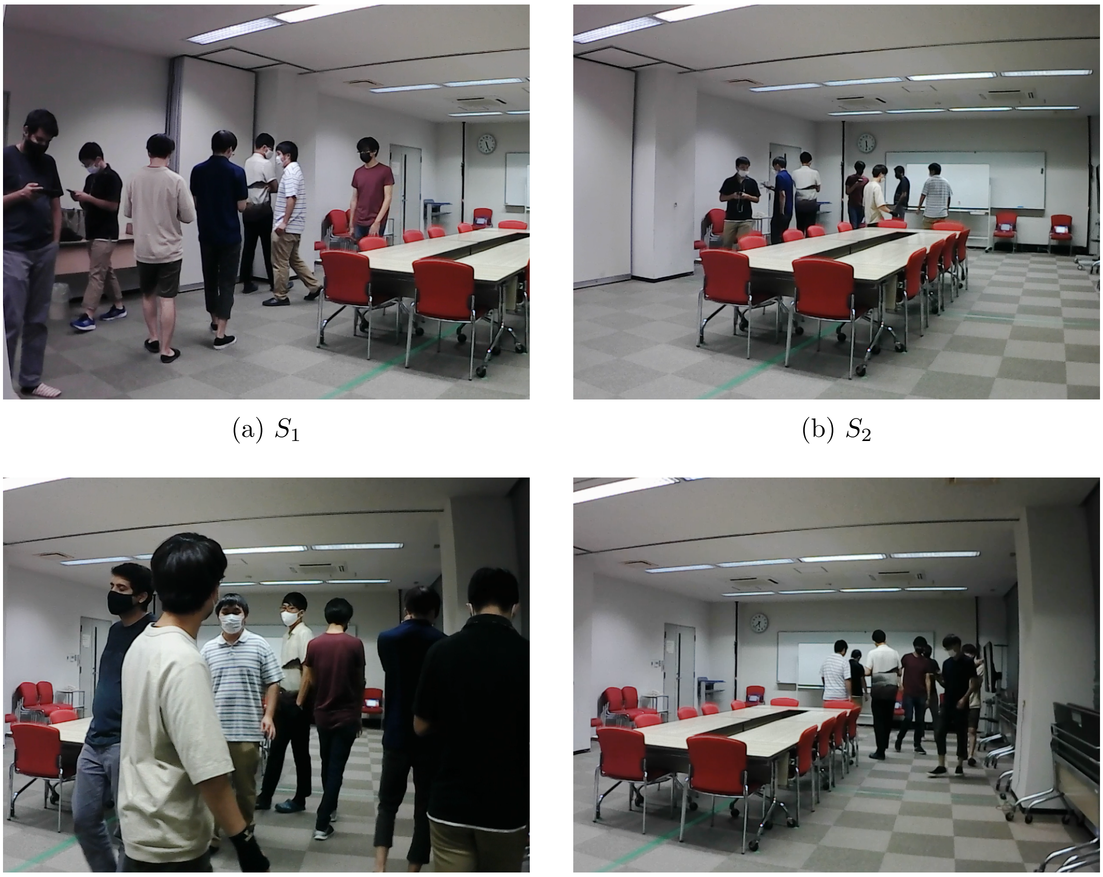


Figure 5.11: Actual Scene of S_{oth} with P_7 (Seminar Room).



(a) S_1

(b) S_2

(c) S_3

(d) S_4

Figure 5.12: Actual Scenes of each scenario with P_7 (Seminar Room).

5.3 Overall Performance

For the final results, we fixed the optimal conditions and parameters used in our system such as learning model, time window size, the number of used subcarriers, the number of used links, and the scenario length. Our results are obtained under the conditions as follows: *LGBMR and LGBMC models* were used for counting and localization, respectively (addtionally, DNN for comparison). Time window size for a single CSI bundle was set in *six seconds with three seconds overlapping.* We used *13 subcarriers* out of 52 and *all four WiFi links.* We set *the scenario length as two minutes.* The numerical results of the system's overall performance are summarized in Table 5.1. The comparisons of performance depending on different conditions and parameters are addressed in Chapter 5.4. In training and testing process, counting datasets for each crowd count contain all section data (S_1-S_4, and S_{oth}), and localization datasets for each section contain all crowd count data (P_1-P_5 in meeting room, P_1-P_{10} in seminar room). In this experiment, we show the counting performance by median absolute error (MAE) because a few error outliers are included in the results due to an observational error. Here, MAE is the median value of the absolute crowd counting errors calculated by $median(\,|\,Real\,Counts - Estimated\,Counts\,|\,)$.

Table 5.1: Overall Performance by LGBM (and DNN).

			Meeting Room (∼5 people)	**Seminar Room (∼10 people)**
Counting Error (MAE)	k-fold	session 1	0.16 (**0.15**)	0.32 (**0.28**)
		session 2	0.18 (**0.17**)	0.36 (**0.31**)
		session 3	**0.13** (0.15)	0.32 (**0.29**)
	leave-one-session-out		**0.35** (0.52)	0.41 (**0.35**)
Localization Accuracy (%)	k-fold	session 1	96.5 (**97.6**)	95.7 (**96.6**)
		session 2	97.1 (**98.2**)	**96.7** (95.0)
		session 3	95.9 (**96.9**)	**97.3** (96.0)
	leave-one-session-out		**91.4** (83.4)	**98.1** (97.6)

To compare the overall differences between the performances of ML (LGBM) and DL (DNN), we present all the numerical results from both learning methods (the results in parenthesis are performance by DNN) in Table 5.1. In the table, we gave bold font to the results that showed better performance between ML and DL. According to the results by leave-one-session-out cross-validation, DL showed worse MAE in the meeting room but achieved better MAE in the seminar room in counting, on the other hand, ML showed better accuracy in both meeting and seminar room in localization. In other words, it is impossible to be clarified that DL always has a clear advantage or always achieves better performance than ML in all the cases, as we will demonstrate in Chapters 5.4.1 and 5.4.2. We have opened the corresponding Python codes and feature datasets* of the results in Table 5.1 to the public through Github.

5.3.1 *k*-fold Cross-validation

The *k*-fold cross-validation is a machine learning evaluation method to assess a trained model by a single session dataset. The whole dataset is split into k folds of datasets from the first. When one fold is selected as test data, the other $k-1$ folds become training data. After repeating this process k times, the system performance is derived by averaging all results from k trials. Specifically, we adopt the stratified k-fold method which splits the folds by criteria ensuring that each fold contains the same ratio of target classes data. In this study, we empirically set the number of folds as $k = 7$. The overall results of *k*-fold cross-validation is summarized in Table 5.1.

In the meeting room experiment, we achieved 0.16, 0.18, and 0.13 MAE of counting error in Session 1, 2, and 3, respectively. We have also checked the percentage of counting errors that were predicted with one person (within-one-person error). The within-one-person error shows 96.0%, 96.0%, and 98.2% in Session 1, 2, and 3, respectively. The counting error CDFs of meeting room experiment are presented in Figure 5.13.

*https://github.com/narajinx/Wi-CaL-WiFi-Crowd-Estimation.git

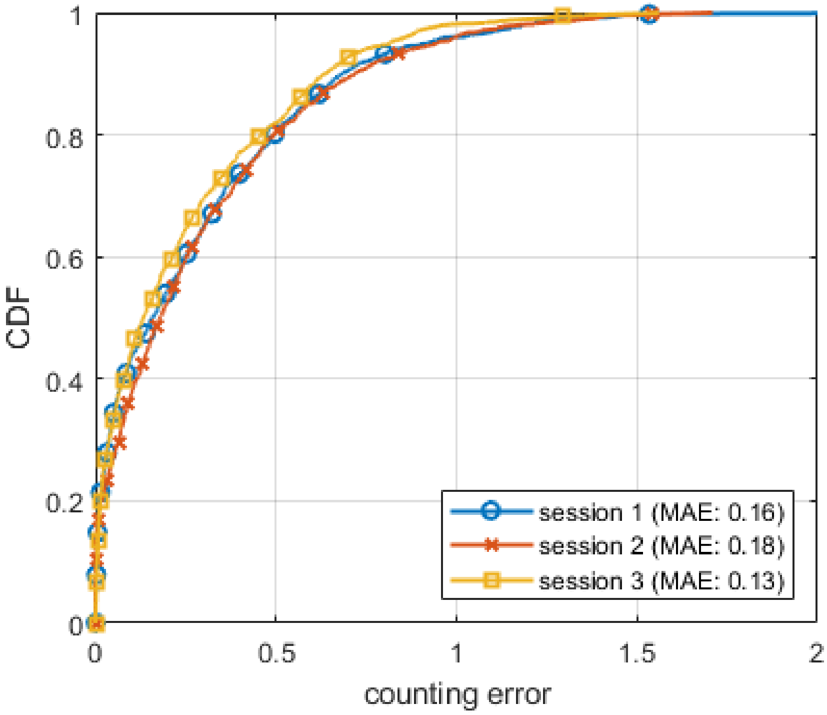
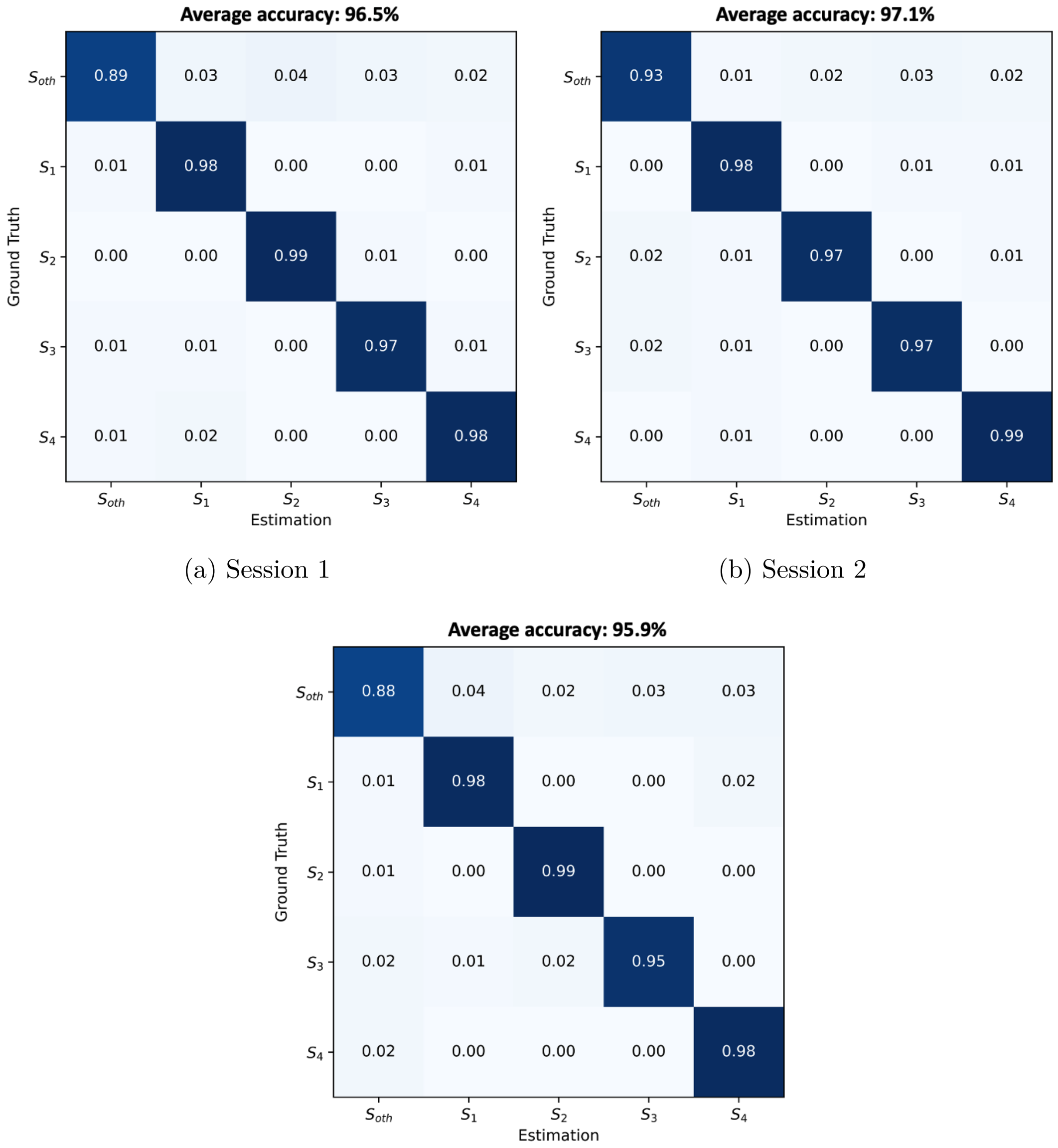


Figure 5.13: Counting Error CDF (k-fold, meeting room).

In crowd localization of meeting room, we achieved 96.5%, 97.1%, and 95.9% of classification accuracy, by Session 1, 2, and 3, respectively. As we can see from confusion matrices in Figure 5.14, most cases of S_1-S_4 show over 95% of localization accuracy, and only S_{oth} shows around 90% of accuracy. This is because the participants were randomly walking across all over the experiment area, i.e., the clusters were sometimes biased into a specific section.



(a) Session 1

(b) Session 2

(c) Session 3

Figure 5.14: Confusion Matrices of Localization (k-fold, meeting room).

Meanwhile, in the seminar room experiment, we achieved 0.32, 0.36, and 0.32 MAE in crowd counting of counting error in Session 1, 2, and 3, respectively. Also, the within-one-person error shows 88.2%, 83.9%, and 87.7% in Session 1, 2, and 3, respectively. The counting error CDFs of seminar room experiment are presented in Figure 5.15.

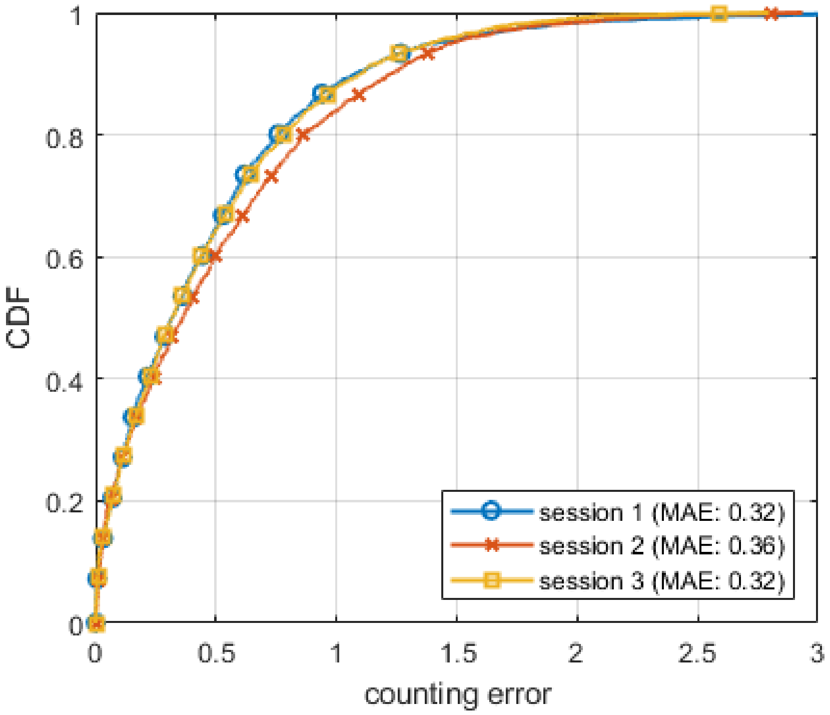
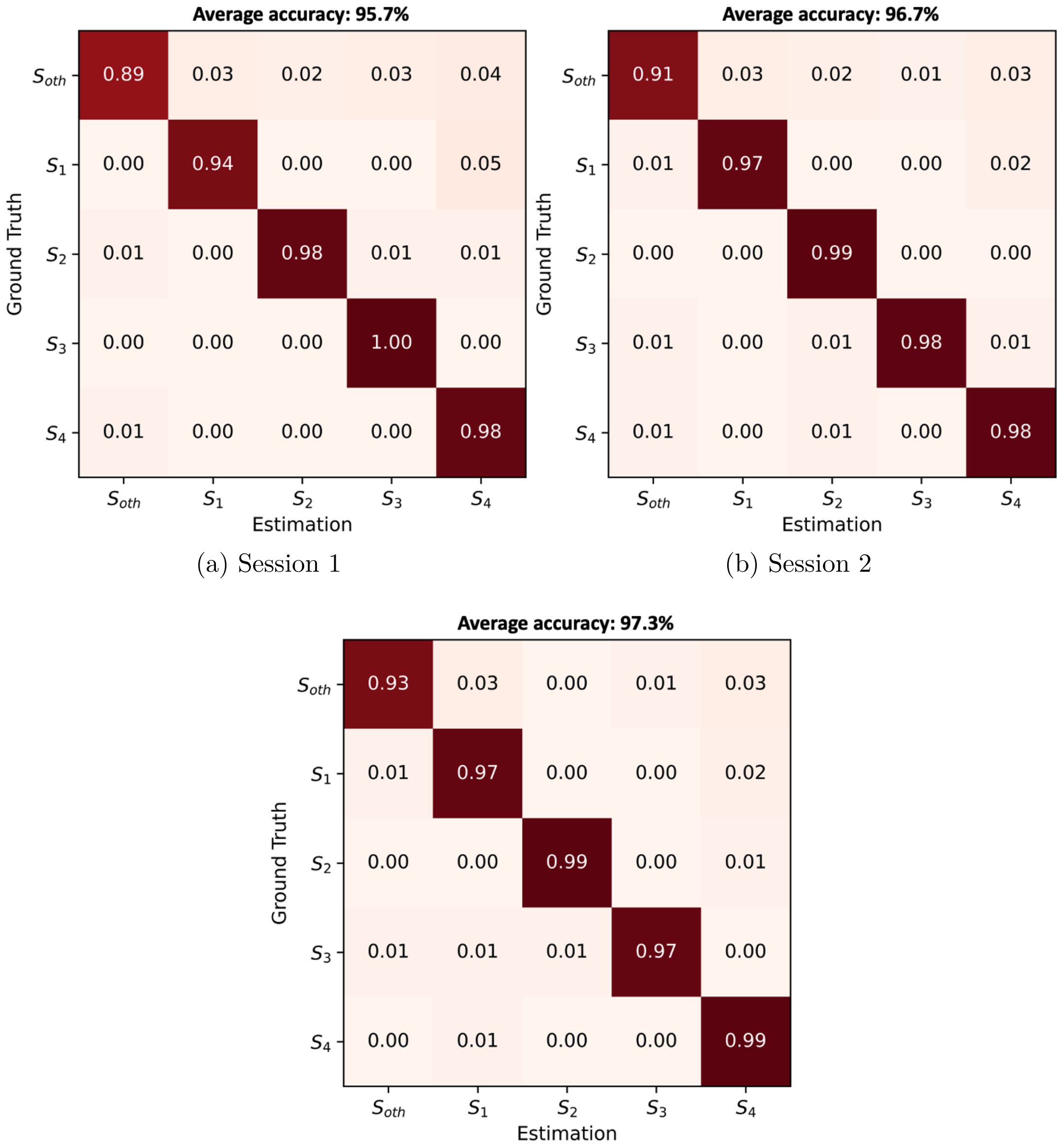


Figure 5.15: Counting Error CDF (k-fold, seminar room).

In crowd localization of seminar room, we achieved 95.7%, 96.7%, and 97.3% of classification accuracy, by Session 1, 2, and 3, respectively. Similarly, S_{oth} scenarios show the lowest localization accuracy compared to S_1-S_4. The confusion matrices of seminar room localization are presented in Figure 5.16.



(a) Session 1

(b) Session 2

(c) Session 3

Figure 5.16: Confusion Matrices of Localization (k-fold, seminar room).

5.3.2 Leave-one-session-out Cross-validation

We have separate datasets of three sessions which are collected in the same room, by the same scenarios, but on different days. This is to confirm our assumption that the tendency of CSI data changes as time passes due to different temperature, humidity, signal interference, and so on. In that case, a regressor or classifier trained by only a certain session's data might not be adequate for the others. However, there are only a few existing studies which are addressing the time-variant influence in CSI measurements. Hence, to confirm this variation between different sessions, we conducted leave-one-session-out cross-validation. Here, leave-one-session-out means, one whole session is selected as test data to test a regressor or classifier trained by the other sessions. This process continues until every session becomes a test session at least once. Finally, the system performance is calculated by averaging all the session results.

As summarized in Table 5.1, we achieved 0.35 MAE and 89.8% of within-one-person error in the meeting room experiment, also 0.41 MAE and 81.8% of within-one-person error in the seminar room. In crowd localization, we achieved 91.4% and 98.1% classification accuracy in the meeting room and seminar room, respectively. Figure 5.17 and 5.18 are presenting the error CDFs of counting results and the confusion matrices of localization results from both meeting room and seminar room, respectively.

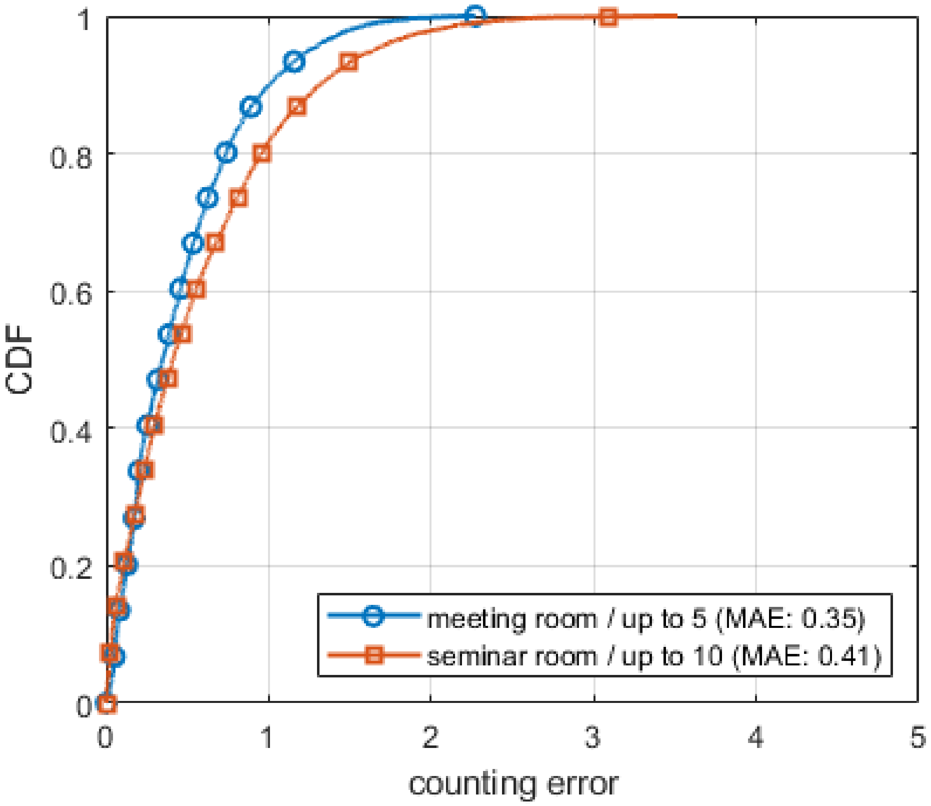
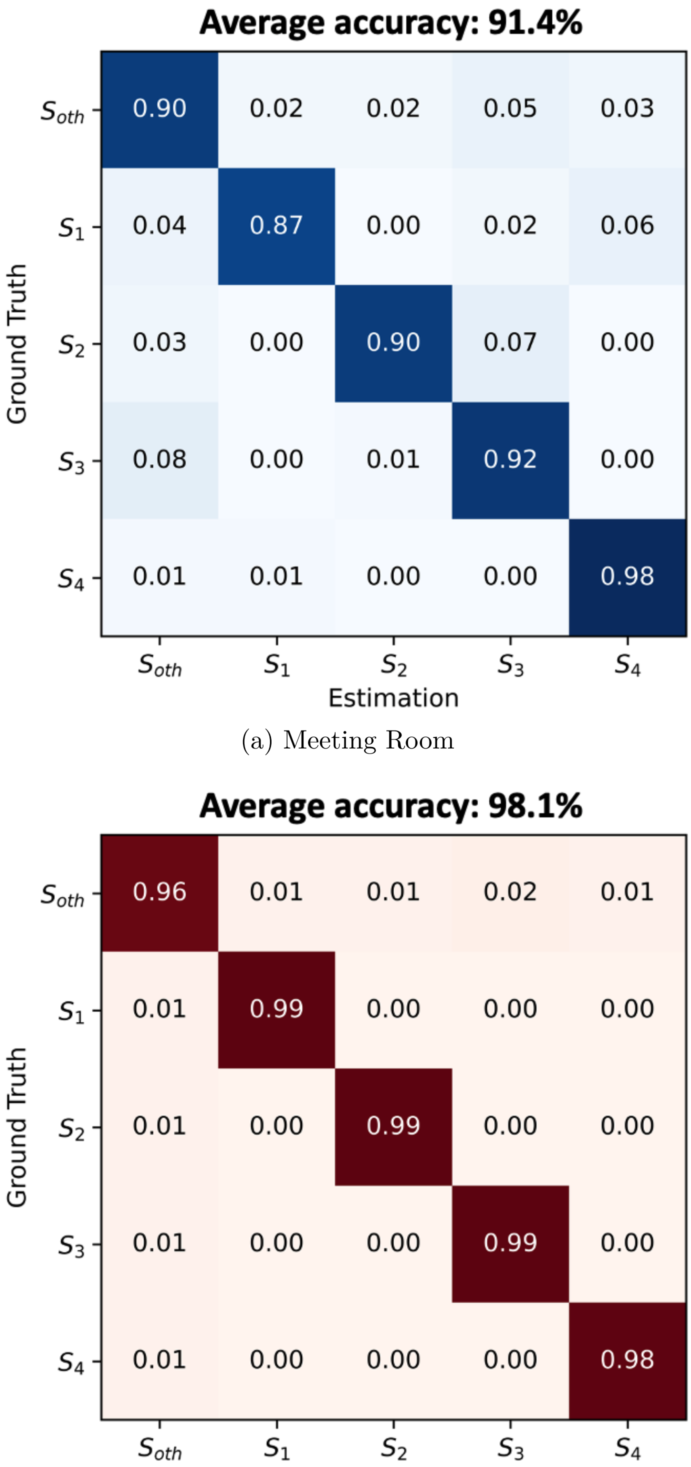


Figure 5.17: Counting Error CDF (Leave-one-session-out).



(a) Meeting Room

(b) Seminar Room

Figure 5.18: Confusion Matrices of Localization (Leave-one-session-out).

5.3.3 Feature Importance

As shown in Table 5.2, the bundle-based features, which are separately designed, dedicated metrics for each counting and localization, mostly hold the highest ranks across all links in both estimations. Meanwhile, we use the statistical features as a common input. This is because, each statistical feature shows different feature importance depending on link number, regardless of what kind of estimation (counting or localization) it contributes for. Therefore, it is hard to define which specific statistical features are always effective for counting or localization.

Table 5.2: Rank of Feature Importance.

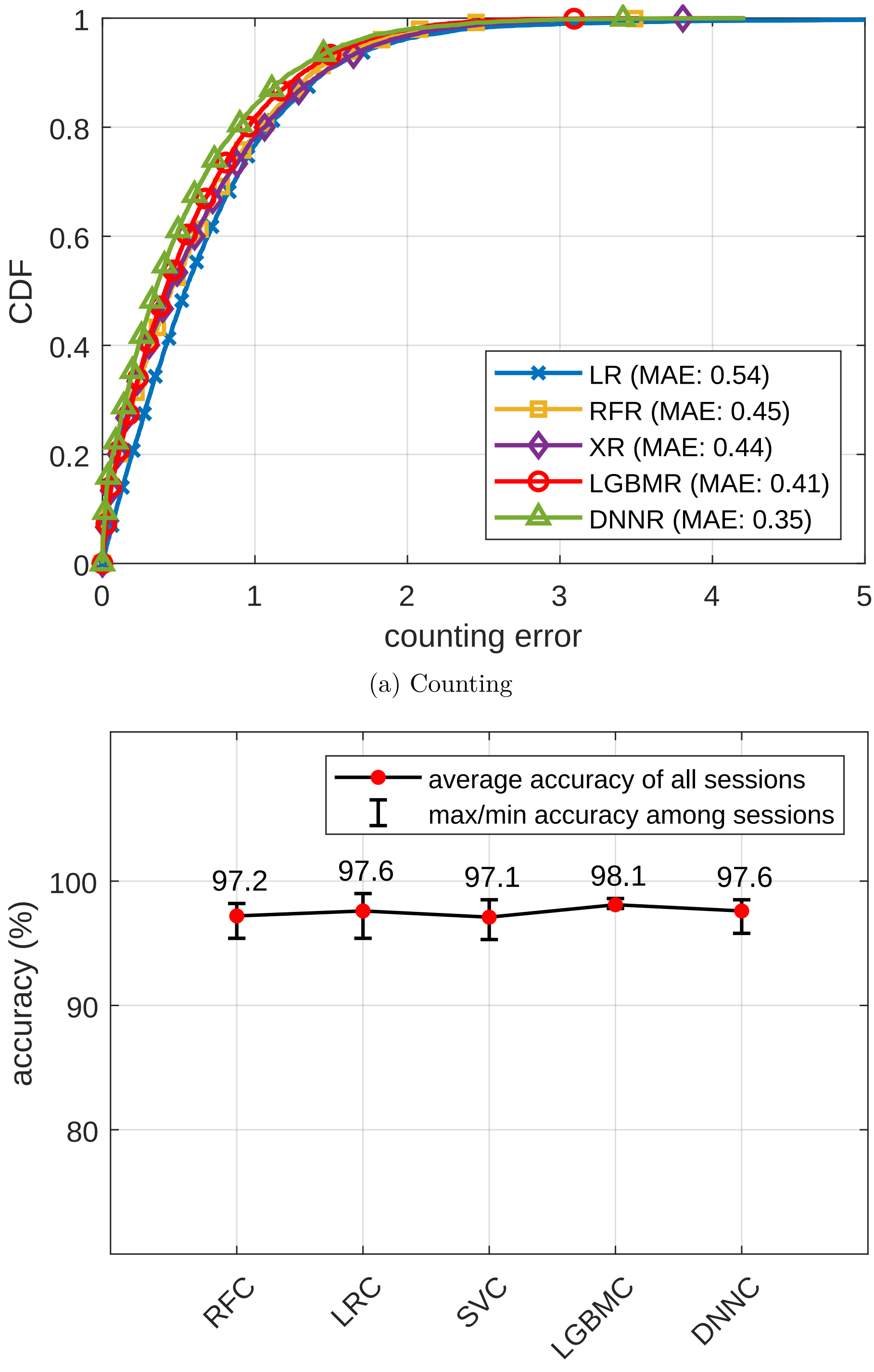
	Rank	**Link 1**	**Link 2**	**Link 3**	**Link 4**
Counting Feature	*1*	**adj**	*euc*	**adj**	**adj**
	2	**qtu**	**adj**	*euc*	*euc*
	3	*euc*	**iqr**	**qtl**	*rss*
	4	*rss*	*rss*	*rss*	**qtu**
	5	**qtl**	**qtu**	**qtu**	**max**
	6	**min**	**qtl**	**max**	**iqr**
	7	**iqr**	**max**	**avg**	**std**
	8	**avg**	**avg**	**iqr**	**avg**
	9	**max**	**std**	**std**	**min**
	10	**std**	**min**	**min**	**qtl**
	Rank	**Link 1**	**Link 2**	**Link 3**	**Link 4**
Localization Feature	*1*	**der**	**der**	**cur**	**min**
	2	**min**	**cur**	**std**	**std**
	3	**qtu**	**min**	**der**	**der**
	4	**std**	**std**	**max**	**cur**
	5	**cur**	**max**	**qtu**	**max**
	6	**avg**	**qtl**	**min**	**qtu**
	7	**max**	**qtu**	**qtl**	**avg**
	8	**qtl**	**avg**	**avg**	**qtl**

5.4 Comparisons

In this chapter, we first compare our system performance depending on the learning models including conventional ML models and DNN, then provide further comparisons between LGBM and DNN. We also present the result of performance comparison between our method and conventional metric (PEM) based method, then show how the system performance changes in different conditions and parameters, such as different kinds of learning models, time window size, the number of used subcarriers, the number of used links, and scenario length. All comparisons are based on the results of leave-one-session-out cross-validation from the seminar room (up to 10 people).

5.4.1 Impact of Learning Model

As we mentioned in Chapter 4.4, we test four different ML models and DNN for each of counting and localization, then we finally select LGBMR and LGBMC for overall evaluation among them. In the case of counting, LGBMR shows the second-best performance (0.41 MAE) after DNNR (0.35 MAE), but we use LGBMR as a prior learning model because of the reasons that are discussed in Chapter 5.4.2. In localization, LGBMC shows the highest accuracy as 98.1%, also it shows the smallest error range of each session testing result. Figure 5.19(a) and (b) present the result comparison by the learning models.



(a) Counting

(b) Localization

Figure 5.19: Impact of Learning Model.

5.4.2 Further Comparison between ML and DL

As we described in Chapter 2, the authors of [32] assessed the crowd counting system by DNN, and used PEM metric as their system's feature. Hence, we set this related work as our comparison target to weigh the pros and cons of ML (LGBMR) and DL (DNNR), and also PEM and our feature. To that end, we first calculated the PEM values from our datasets in the same way, then constructed our DL model with the same DNN architecture described in [32] as follows: four hidden layers with [1000, 500, 100, 10] neurons, 10^{-4} of learning rate, 100 of batch size, Adam optimizer and ReLU activation function. Figure 5.20 shows the differences in accuracy and training time depending on the used model, used feature, and epochs setting. The descriptions of the trials in Figure 5.20 are as follows:

- *LGBM-OF*: LGBMR trained with our features. It shows 0.41 MAE and requires 4.6 seconds of training time.
- *DNN-OF-20*: DNNR trained with our features, 20 epochs. DNN requires 20 epochs to reach to the same MAE with LGBMR, and that needs approximately 4 times longer training time than LGBMR.
- *DNN-OF-ES*: DNNR trained with our features, early stopping (patience: 100, average number of epochs: 445). We empirically set the patience setting of early stopping in 100. It shows 0.06 improvement in MAE over LGBM, requiring 336 seconds of training time. We select this as our final DNN model setting.
- *DNN-PEM-ES*: DNNR trained with PEM, early stopping (patience: 100, average number of epochs: 376). PEM shows 0.18 worse MAE than the case of our features (*DNN-OF-ES*) by the same model settings.
- *DNN-OF-22K*: DNNR trained with our features, 22000 epochs. 22000 is the same number of epoch settings in [32]. It shows 0.03 improved MAE compared to *DNN-OF-ES* case, but the required training time is unrealistic (15,867 seconds).

- *DNN-PEM-22K*: DNNR trained with PEM, 22000 epochs. This is the identical condition with [32]. It also shows 0.18 worse MAE than the case of our features (*DNN-OF-22K*).

Here, early stopping is a method for avoiding overfitting in DNN models by the halt of model fitting if validation MAE doesn't seem to be enhanced anymore, and patience value is the early stopping parameter of how many epochs DNN will be patient even without enhancement of validation MAE. All above results are obtained by the following PC specification: Intel(R) Core(TM) i7-10750H CPU (2.60GHz, 2.59GHz), 16GB RAM, and 64-bit Windows OS.

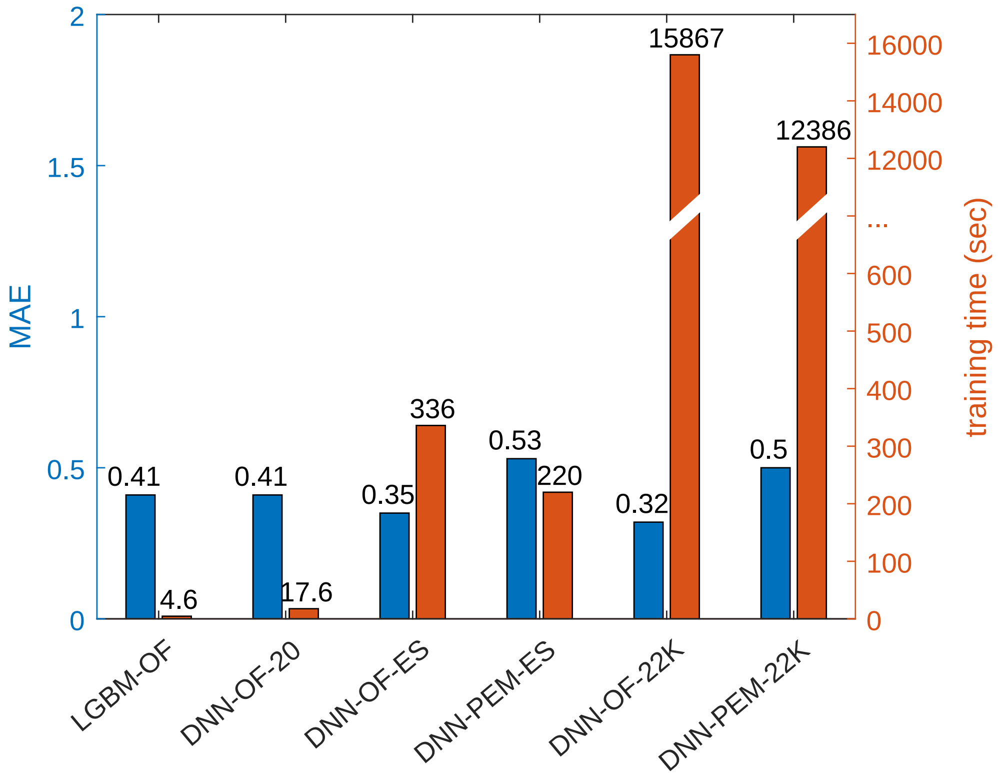


Figure 5.20: Comparisons in MAE and Training Time over ML and DL.

Although DNN shows slightly better performance, we evaluate our system by LGBMR and LGBMC in the rest of this paper. There are several reasons that we use the conventional ML models other than DL. First is, the fact that there is no significant gap between the ML and DL-based results implies evidence of well-designed features. Our work is more focused on effective feature engineering, which is to find out some attributes corresponding to a system's goal from the raw data, rather than using an advanced learning model. Meanwhile, LGBM shows a considerably shorter training time than DNN. Generally, DNN requires a large number of epochs and a longer training time to reach to system's best performance. We adopted the ESP32 nodes as our CSI reading devices with the consideration of IoT-based aspect, therefore a low computing power environment is also needed to be considered. In addition, since a retraining process for a new target area is required as of now, it should be considered that the cost of model training of DL would be a high barrier.

5.4.3 Comparison with Conventional Metric PEM by LGBMR

To compare the counting performance of our system with PEM by LGBMR as well, we compared two cases of PEM only (with 52 subcarriers) and our features (with 13 subcarriers). Under our testing environment, our features show better performance (0.41 MAE, 81.8% of within-1-person error) than PEM-based performance (0.62 MAE, 66.5% of within-1-person error), as shown in Figure 5.21. The conditions of comparison except for the used features and the number of used subcarriers were all identical.

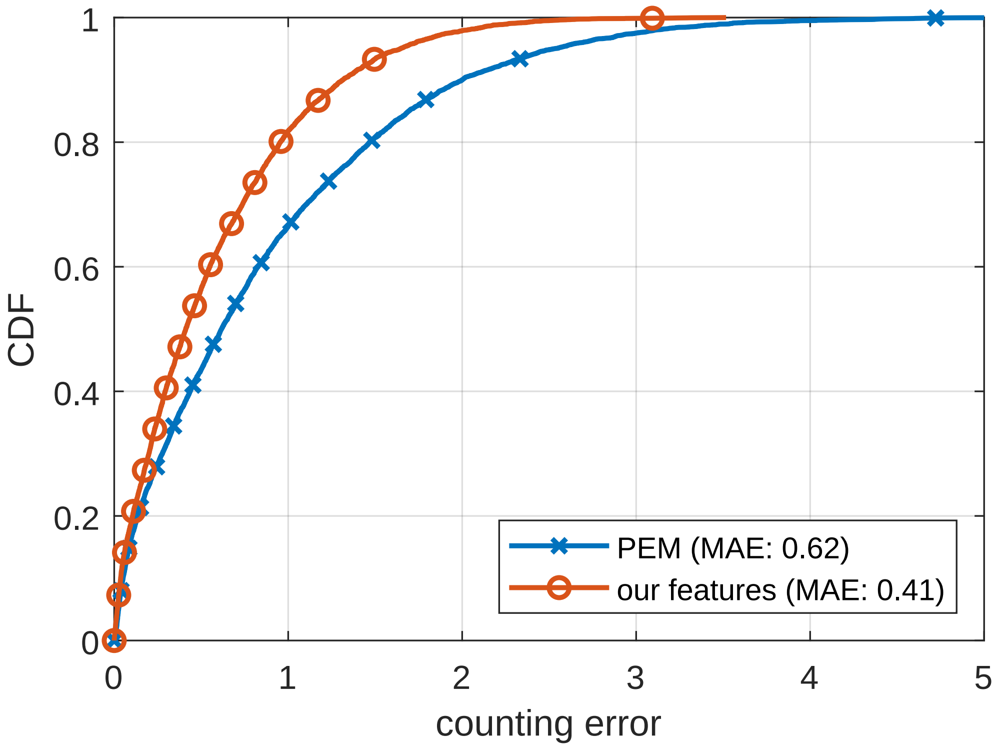


Figure 5.21: Performance Comparison with PEM.

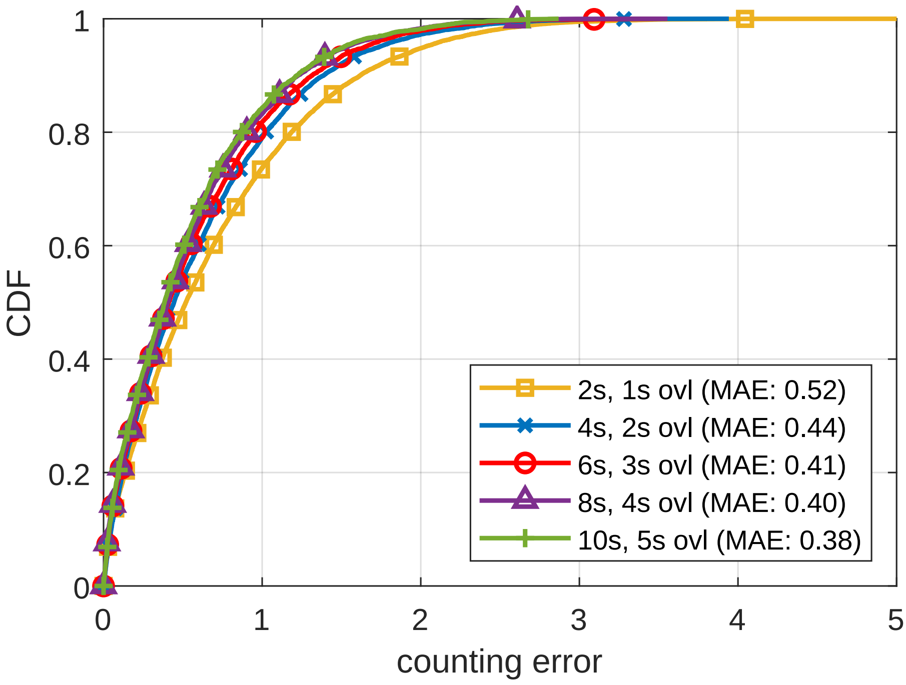
We also checked the feature importance for our system. To objectively compare the feature importance with PEM, we include the PEM values with our features in LGBMR for crowd counting. As we described in Chapter 2, PEM has been used in several previous studies as the main metric of their system to this day because of its good linearity. Nevertheless, several of our features including $\mathbf{adj}^{(w)}$ and $euc^{(w)}$ show higher rank in feature importance than PEM in link 1, 2 and 4, as shown in Table 5.3. Only in link 3, PEM shows the highest impact in feature importance. This may imply that the combination of multiple features can achieve better performance than PEM-only in some cases such as different links or environments.

Table 5.3: Rank of Feature Importance including PEM.

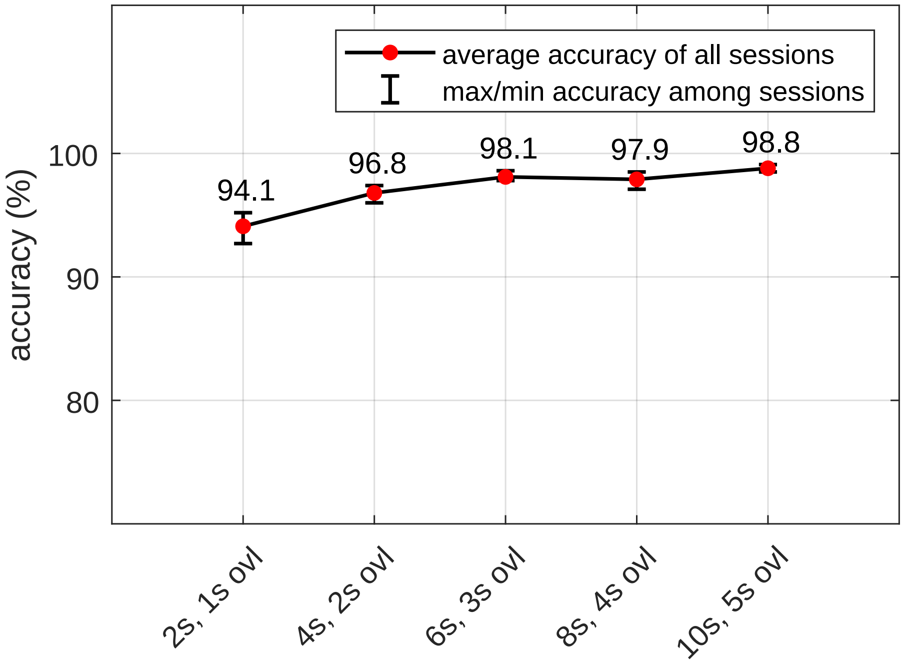
Rank	**Features**			
	Link 1	Link 2	Link 3	Link 4
1	*euc*	**adj**	PEM	**adj**
2	**adj**	*euc*	**adj**	PEM
3	*rss*	PEM	*euc*	*euc*
4	**qtl**	*rss*	**qtl**	**qtu**
5	**qtu**	**qtu**	*rss*	**max**
6	PEM	**avg**	**max**	*rss*
7	**min**	**iqr**	**std**	**avg**
8	**iqr**	**std**	**min**	**iqr**
9	**avg**	**max**	**avg**	**min**
10	**max**	**min**	**qtu**	**qtl**
11	**std**	**qtl**	**iqr**	**std**

5.4.4 Impact of Time Window Size

Since our approach is adopting a method extracting statistical and designed features from a single-time-window CSI bundle, the configuration of time window size influences system performance. In other words, the performance evaluation by each time window length is necessary because it is important to decide how long data will be a base unit of the system for the learning phase and online phase. Since the longer time window contains more information and its statistical values are more stable, the system performance becomes higher as the length of the time window increases as we can see in Figure 5.22(a) and (b). However, with taking into account the system's real-time estimation capability, we decided to use the time window size of our system in six seconds with three seconds overlapping.



(a) Counting

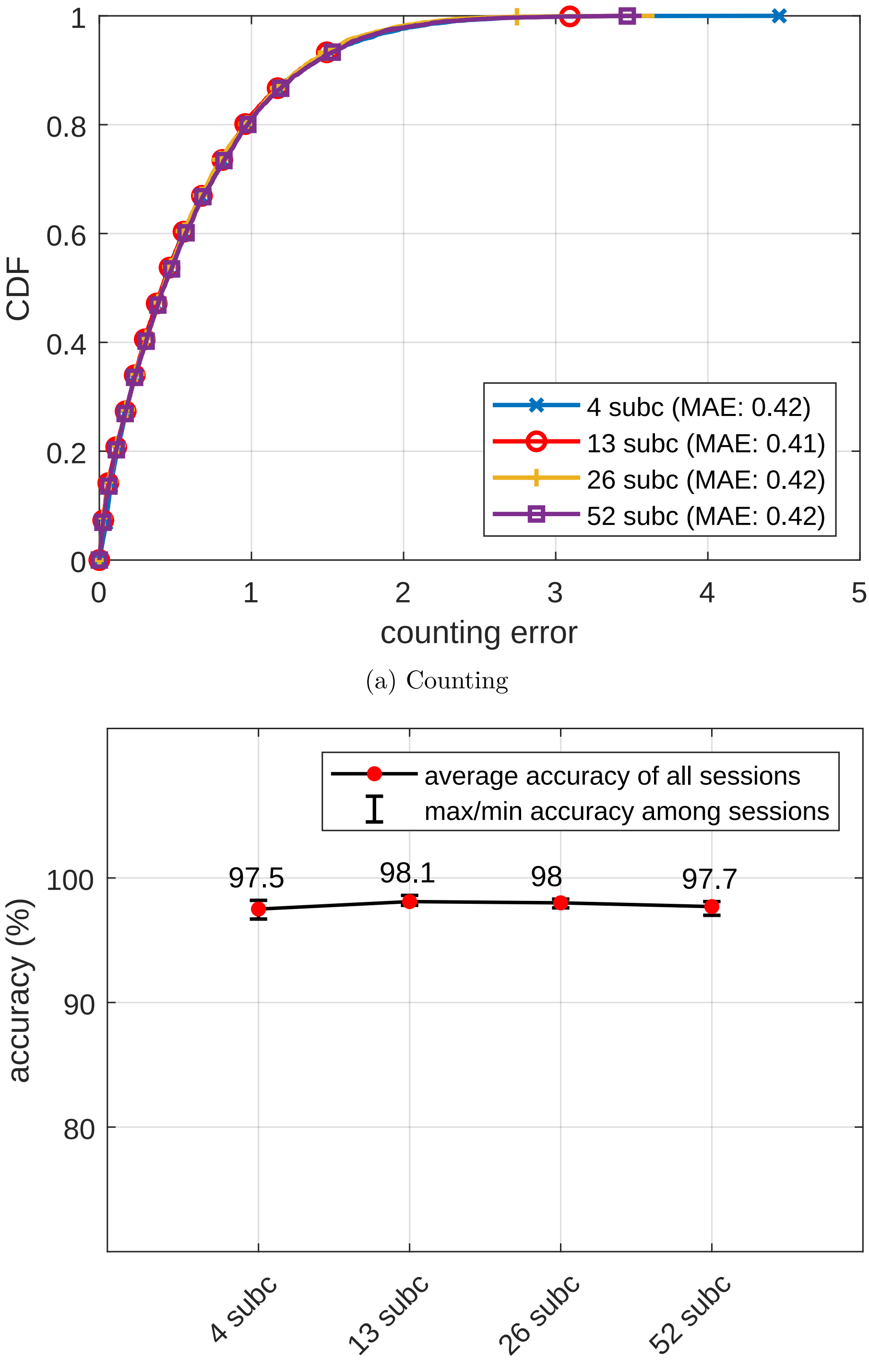


(b) Localization

Figure 5.22: Impact of Time Window Size.

5.4.5 Impact of Number of Subcarriers

In terms of the number of subcarriers, the difference in system performance is not very significant. Even so, we decided to use 13 subcarriers data in our system, since it shows a slightly higher performance than the other cases of using 4, 26, and 52 subcarriers in both counting and localization, as shown in Figure 5.23(a) and (b). Here, the used subcarriers are selected with having the identical distance on both sides, from subcarrier 1 to 52 (e.g., 13 subcarriers: 1, 5, 9, $\cdots$, 49.). The small number of subcarriers would have an advantage in terms of shorter training time. For instance, we practically checked the training time of each case that contains the different number of subcarriers as 1.4s (4 subc), 4.6s (13 subc), 9.5s (26 subc), and 16.7s (52 subc) by leave-one-session-out cross-validation with 600 mins long dataset (10 mins data × 10 people × 3 days × 2-session data for each day) and PC that has the following specification: Intel(R) Core(TM) i7-10750H CPU (2.60GHz, 2.59GHz), 16GB RAM, and 64-bit Windows OS. Nevertheless, the reason why we use 13 subcarriers here is that we also need to consider the performance degradation produced by the mutual similarity between the signal tendency of chosen subcarriers that leads to overfitting.



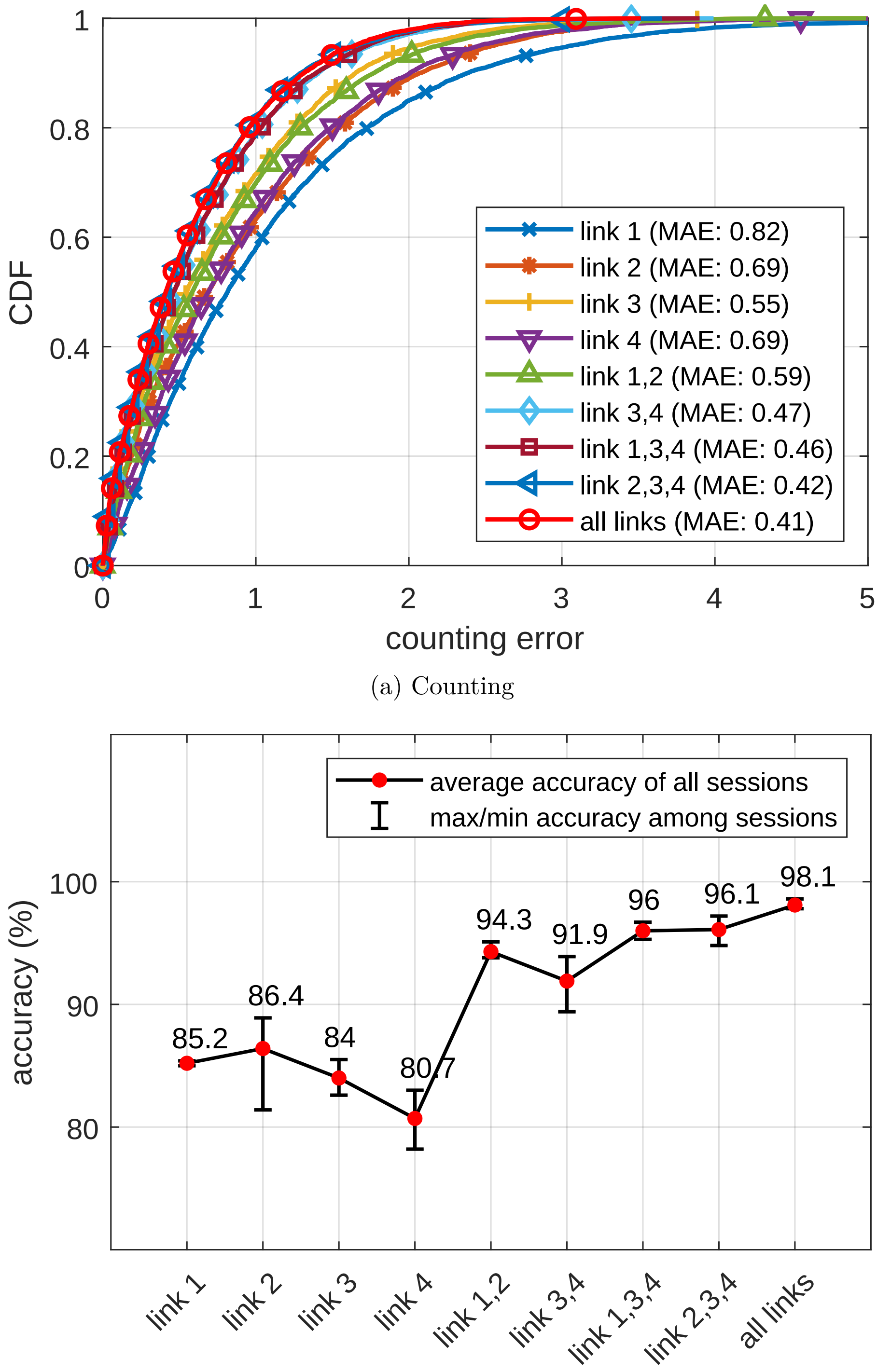
(a) Counting

(b) Localization

Figure 5.23: Impact of Number of Used Subcarriers.

5.4.6 Impact of Number of Links

We placed four WiFi links to cover the whole experiment area without any blind spots. Naturally, the number of WiFi links impacts the system performance, therefore we compare the accuracy when we use only a part of the links data in the learning and testing phase. As we can see in Figure 5.24(a) and (b), the system performance drops when we include only a single link data, and it is gradually improved as the number of links is increased, then it shows the best performance when we use all four links. Also, we can see that the cases including link 1 show higher MAE than the others. This can be considered that link 1 in the seminar room was too short to cover the entire area compared to the other links.



(a) Counting

(b) Localization

Figure 5.24: Impact of Number of Used Links.

5.4.7 Impact of Scenario Length

As mentioned in Chapter 5.2, two-minute-long CSI readings have collected for each scenario ($P_{n_{peo}}S_{n_{sect}}$). To figure out how long scenario data is required for higher accuracy, we compared the performance of when we use only a part of scenario data or the whole two minutes data for the training phase. We adjusted in scenario length by 30, 60, 90, and 120 seconds, and the corresponding results showed 0.47, 0.44, 0.43 and 0.41 MAE in counting, respectively, and 97.5%, 98.0%, 98.0% and 98.1% in localization, respectively. The scenario length seems not to give a drastic impact on our system performance, nonetheless, the numerical accuracy is being slightly improved by the longer scenario data. The system performances depending on scenarios length are summarized in Table 5.4.

Table 5.4: Comparison depending on Scenario Length.

Scenario Length	**System Performance**	
	counting (MAE)	localization (%)
30 sec.	0.47	97.5
60 sec.	0.44	98.0
90 sec.	0.43	98.0
120 sec.	0.41	98.1

6 Conclusion

6.1 Summary

In this thesis, we examined the potential and feasibility of the simultaneous crowd estimation system that can predict both the number and location of a crowd, by WiFi IoT CSI solution and machine learning. We also comparatively confirmed the pros and cons between conventional machine learning and deep learning in crowd estimation by empirical comparisons. We utilized for the first time, ESP32 transceivers and its CSI toolkit as the WiFi sensing source for medium-scale crowd counting and localization instead of conventional WiFi, therefore we provided the initial foundation of this new CSI platform by various comparisons. We conducted the empirical experiments with up to 10 people (for crowd counting) in two four-sectioned real environments (for crowd localization) for three different days. By leave-one-session-out cross-validation, our system achieved 0.35 MAE of counting error (89.8% of within-1-person error) and 91.4% of localization accuracy with five people in a small-sized meeting room, and 0.41 MAE of counting error (81.8% of within-1-person error) and 98.1% of localization accuracy with 10 people in a medium-sized seminar room, through machine learning. We will proceed and expand this work to resolve the remaining tasks such as enabling layout-independent learning, large-scale human density estimation, or multi-cluster crowd estimation.

6.2 Limitations & Future Works

Through this study, we figured out the optimal conditions and parameters for simultaneous crowd estimation such as the learning models, the size of time windows, the number of used subcarriers and links, by practical system implementation and diverse performance evaluations. Furthermore, we carried out leave-one-session-out cross-validation to confirm the realistic system performance with considering the influence of the change of CSI signal trends by the passage of time. Furthermore, we empirically compared the pros and cons of the conventional ML model (LGBM) and DL model (DNN).

Practically, it was confirmed that the system shows lower accuracy when we use data of different days for each training and testing phase compared to when using the same day data for both training and testing. Thus, we need to concretely reveal which factors (e.g., the difference of temperature, humidity, or fine inner structure) produce the degradation of system performance by installing environmental sensors and inputting its data as a feature for machine learning. Also, we assessed our system performance by the test datasets that are separately collected with the certain crowd count (P_0-P_{10}), assuming the system can be applied in realistic situations as long as the learning models are trained once. However, it seems necessary to carry out a real-time system evaluation that includes continuous changes of the number of people in the area, to reveal the variation of system accuracy depending on those state transitions.

Besides, we define the following five remaining challenges and future directions toward the further-enhanced WiFi crowd estimation.

6.2.1 Selective Subcarrier

As we mentioned in Chapter 5.4.5, there was no significant difference in estimation accuracy depending on the number of used subcarriers in this work. Naturally, the less number of subcarriers makes the training phase faster, but in some cases, the small number of subcarrier selection could cause the lack of enough distinct features. Hence, the algorithmic investigation of selective subcarriers for a certain target area would be needed as one of our future works.

6.2.2 Layout-independent Learning

It is also necessary to conduct the leave-one-room-out cross-validation. We implemented our system in a meeting room and seminar room which have a relatively simple inner structure, however, if we want to examine the feasibility of the system in the real world, it should be on trial in the public space such as supermarkets, museums, and even outdoors. We will proceed in stages for our future work on system robustness from diverse indoor layouts, structure, and outdoors by using transfer learning of previously learned tasks.

6.2.3 Large-scale Human Density Estimation

In the similar context with above-mentioned chapter, the validation of the system's detection limit in terms of the number of people is essential. Our system uses the statistical values and features in a given size of time windows as training data for machine learning. Especially, the crowd count estimation is based on CSI variation and regression analysis, but the fluctuation level of CSI signals is expected that it will necessarily converge at a certain point of crowd size. Therefore, we need to examine the possibility of massive crowd estimation, which is currently possible by vision-based approaches, by more large-scale experiments.

6.2.4 Multi-cluster Crowd Estimation and Densely-separated Localization Sections

Currently, our system has a restriction that it can estimate the crowd information only in the cases when a crowd is gathered within a single section (S_1-S_4) or randomly spread across the whole area (S_{oth}). Undoubtedly, it is a generous precondition that all people are gathered at a single section in an area. However, at least this work has significance in the sense of the very first foundation stone in WiFi sensing-based crowd localization that can contribute to predicting which part of an area is the most crowded spot in the real world such as retail stores, supermarkets, or exhibitions. Indeed, the most ideal case is if we can estimate the number of people in each section like "five people in Section A, three people in Section B.", i.e., when the crowd is split into multiple clusters and exists in multiple sections. Also, more densely divided crowd localization sections would be interesting. Currently, we have confirmed that four-sectioned crowd localization can achieve over 90% of classification accuracy, however, the more a target area has large space, the more detailed separation of the area would be necessary like as fingerprinting techniques in indoor localization. This detailed estimations, for instance, will enable to help disperse the people onto a less crowded area in the situation of an emergency evacuation. Even though it requires more time and effort to devise a new metric or design a different algorithmic approach, this multi-cluster crowd estimation would become our final objective in our future work.

6.2.5 Coexistence of Multiple Types of RF Signals

As we mentioned in Chapter 2, some studies are addressing the WiFi sensing with other multiple types of wireless signals such as UWB and visible light [8] or Zigbee, Bluetooth and microwave [34]. A considerable advantage of WiFi CSI-based human sensing is that it is possible to detect people without installing any other devices by utilizing pervasive WiFi signals. Nevertheless, different types of wireless sensing could be helpful in some cases, for example, the visible light sensors can recognize the obvious change of luminance occurred by the passage of person or change of crowd size, as presented in [8]. Meanwhile, since it is necessary to consider the impact of coexisting radio frequency (RF) signals on WiFi if we use multiple types of wireless signals, the signal interference should be detected. In this case, the RFI detection algorithm introduced in [34] can be a base of the solution for eliminating the redundant components in CSI measurement.

6.2.6 Weather Consideration

Obviously, the weather conditions may affect the wireless signals including WiFi, especially, many people practically experience the degradation of wireless network service under the stormy weather. Therefore, we supposed that it is necessary to investigate the impact of weather condition to WiFi channels. In [41], we could get some clues about the environmental impacts such as temperature or humidity. Ohara *et al.* presented an object state transition detection technique by WiFi CSI in the paper. They observed and measured the CSI patterns when the door or windows are opened or closed in a target room, and tried to distinguished the current state of door and windows for several days. In the discussion part of the paper, they described the results of long-term evaluation for 10 days, as a result, they concluded that room temperature doesn't seem very influential but the difference of daily humidity affects the system performance. This result implies the necessity for the cooperation of humidity sensor and crowd estimation system, for instance, by input the humidity measurement as a feature into the learning models.

Acknowledgements

Firstly, I would like to express great thanks to my patient and supportive supervisor, Professor Keiichi Yasumoto, who has approved and accepted myself into his laboratory and research work. Without him, I couldn't even think about successful end of my Ph.D. course. Every piece of his advice has become my motivation and driving force, also his insight and financial support were incredibly helpful to progress all my works. I will always remember that it was very fortunate to get supervision from Professor Yasumoto at the last step of my education.

Including Professor Yasumoto, I also want to thank all the other thesis committee members, Professor Okada, Professor Suwa, and Professor Fujimoto. Especially, my major advisor, Professor Manato Fujimoto, has tried all his best to lead and guide me and my research toward correct direction. I won't forget those many days and many nights that we thought, we discussed, and we chatted together. Also, Professor Hirohiko Suwa and Professor Yuki Matsuda, who are the faculty members of Ubiquitous Computing Systems lab., were completely supportive career counselors for me. I really appreciate for their heartful advice and considerations they provided whenever I was feeling the anxiety about the future.

I would also like to say big thank to my previous supervisors, Professor Suk Chan Kim in Pusan National University and Professor Young-bok Choi in Tongmyong University. I believe that, the things that I could learn and study by them, such as their knowledge and skills in the major field of study and their know-hows to manage a research project, have undoubtedly become the biggest and solidest foundation of my works and this thesis as well.

Besides, all my colleagues in our lab., who gave me indescribably great help to properly settle down at NAIST and in Japan too at the beginning of my first life abroad, were definitely great supporters, guides, and pilots for me. Since they were all there, I was able to accomplish my papers, presentations, awards, and whatever I've achieved here. I want them to forgive me that I can't describe each and every person's name here.

Lastly, to my parents, Douckhwa Kim and Daewoo Choi, also to my elder brother, Sejin Choi, you all always give both the physical and mental back-up on me without any conditions. Because of the fact that all of you are happy and healthy enough, I was able to proceed my research works in relief. We couldn't meet each other for such a long time due to the pandemic, but I have been always feeling like we stick together regardless of the distance between us during my study in Japan. I always appreciate your sacrifice and support as a parent and brother, and I hope I can sufficiently return your favor as much as I can. Also, I would like to say a special thank you to our spiritual mentors who make me feel be protected when I lose my way and need helps to find out a proper orientation.

References

[1] V. A. Sindagi and V. M. Patel, “A survey of recent advances in cnn-based single image crowd counting and density estimation,” *Pattern Recognition Letters*, vol. 107, pp. 3–16, 2018.

[2] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, “Single-image crowd counting via multi-column convolutional neural network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 589–597.

[3] C. Zhang, H. Li, X. Wang, and X. Yang, “Cross-scene crowd counting via deep convolutional neural networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 833–841.

[4] E. Cianca, M. De Sanctis, and S. Di Domenico, “Radios as sensors,” *IEEE Internet of Things Journal*, vol. 4, no. 2, pp. 363–373, 2016.

[5] P. Liu, S.-K. Nguang, and A. Partridge, “Occupancy inference using pyroelectric infrared sensors through hidden markov models,” *IEEE Sensors Journal*, vol. 16, no. 4, pp. 1062–1068, 2015.

[6] A. Filippoupolitis, W. Oliff, and G. Loukas, “Bluetooth low energy based occupancy detection for emergency management,” in *2016 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS)*. IEEE, 2016, pp. 31–38.

[7] S. H. Doong, “Spectral human flow counting with rssi in wireless sensor networks,” in *2016 international conference on distributed computing in sensor systems (DCOSS)*. IEEE, 2016, pp. 110–112.

[8] H. Mohammadmoradi, S. Yin, and O. Gnawali, "Room occupancy estimation through wifi, uwb, and light sensors mounted on doorways," in *Proceedings of the 2017 International Conference on Smart Digital Environment*, 2017, pp. 27–34.

[9] W. Li, M. J. Bocus, C. Tang, R. J. Piechocki, K. Woodbridge, and K. Chetty, "On csi and passive wi-fi radar for opportunistic physical activity recognition," *IEEE Transactions on Wireless Communications*, 2021.

[10] J. A. Ansere, G. Han, H. Wang, C. Choi, and C. Wu, "A reliable energy efficient dynamic spectrum sensing for cognitive radio iot networks," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6748–6759, 2019.

[11] X. Liu, Q. Sun, W. Lu, C. Wu, and H. Ding, "Big-data-based intelligent spectrum sensing for heterogeneous spectrum communications in 5g," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 67–73, 2020.

[12] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, 2019.

[13] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1629–1645, 2019.

[14] Y. Zeng, P. H. Pathak, and P. Mohapatra, "Wiwho: Wifi-based person identification in smart spaces," in *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 2016, pp. 1–12.

[15] X. Liu, J. Cao, S. Tang, J. Wen, and P. Guo, "Contactless respiration monitoring via off-the-shelf wifi devices," *IEEE Transactions on Mobile Computing*, vol. 15, no. 10, pp. 2466–2479, 2015.

[16] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial wifi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, 2017.

[17] C. Wu, Z. Yang, Z. Zhou, X. Liu, Y. Liu, and J. Cao, "Non-invasive detection of moving and stationary human with wifi," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2329–2342, 2015.

[18] D. Lian, J. Li, J. Zheng, W. Luo, and S. Gao, "Density map regression guided detection network for rgb-d crowd counting and localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1821–1830.

[19] C. Liu, X. Weng, and Y. Mu, "Recurrent attentive zooming for joint crowd counting and precise localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1217–1226.

[20] H. Choi, T. Matsui, S. Misaki, A. Miyaji, M. Fujimoto, and K. Yasumoto, "Simultaneous crowd estimation in counting and localization using wifi csi," in *2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2021, unpublished.

[21] S. Depatla and Y. Mostofi, "Crowd counting through walls using wifi," in *2018 IEEE international conference on pervasive computing and communications (PerCom)*. IEEE, 2018, pp. 1–10.

[22] O. T. Ibrahim, W. Gomaa, and M. Youssef, "Crosscount: A deep learning system for device-free human counting using wifi," *IEEE Sensors Journal*, vol. 19, no. 21, pp. 9921–9928, 2019.

[23] S. Liu, Y. Zhao, and B. Chen, "Wicount: A deep learning approach for crowd counting using wifi signals," in *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*. IEEE, 2017, pp. 967–974.

[24] S. Di Domenico, M. De Sanctis, E. Cianca, and G. Bianchi, "A trained-once crowd counting method using differential wifi channel state information," in *Proceedings of the 3rd International on Workshop on Physical Analytics*, 2016, pp. 37–42.

[25] M. Nakatsuka, H. Iwatani, and J. Katto, “A study on passive crowd density estimation using wireless sensors,” in *The 4th Intl. Conf. on Mobile Computing and Ubiquitous Networking (ICMU 2008).* Citeseer, 2008.

[26] Y. Yuan, J. Zhao, C. Qiu, and W. Xi, “Estimating crowd density in an rf-based dynamic environment,” *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3837–3845, 2013.

[27] C. Xu, B. Firner, R. S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, and N. An, “Scpl: Indoor device-free multi-subject counting and localization using radio signal strength,” in *Proceedings of the 12th international conference on Information Processing in Sensor Networks*, 2013, pp. 79–90.

[28] W. Xi, J. Zhao, X.-Y. Li, K. Zhao, S. Tang, X. Liu, and Z. Jiang, “Electronic frog eye: Counting crowd using wifi,” in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications.* IEEE, 2014, pp. 361–369.

[29] X. Guo, B. Liu, C. Shi, H. Liu, Y. Chen, and M. C. Chuah, “Wifi-enabled smart human dynamics monitoring,” in *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, 2017, pp. 1–13.

[30] H. Zou, Y. Zhou, J. Yang, and C. J. Spanos, “Device-free occupancy detection and crowd counting in smart buildings with wifi-enabled iot,” *Energy and Buildings*, vol. 174, pp. 309–322, 2018.

[31] J. Li, P. Tu, H. Wang, K. Wang, and L. Yu, “A novel device-free counting method based on channel status information,” *Sensors*, vol. 18, no. 11, p. 3981, 2018.

[32] R. Zhou, X. Lu, Y. Fu, and M. Tang, “Device-free crowd counting with wifi channel state information and deep neural networks,” *Wireless Networks*, pp. 1–12, 2020.

[33] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, “Freecount: Device-free crowd counting with commodity wifi,” in *GLOBECOM 2017-2017 IEEE Global Communications Conference.* IEEE, 2017, pp. 1–6.

[34] Y. Zheng, C. Wu, K. Qian, Z. Yang, and Y. Liu, “Detecting radio frequency interference for csi measurements on cots wifi devices,” in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–6.

[35] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, “Tool release: Gathering 802.11 n traces with channel state information,” *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53–53, 2011.

[36] Y. Xie, Z. Li, and M. Li, “Precise power delay profiling with commodity wi-fi,” *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1342–1355, 2018.

[37] S. M. Hernandez and E. Bulut, “Lightweight and standalone iot based wifi sensing for active repositioning and mobility,” in *2020 IEEE 21st International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*. IEEE, 2020, pp. 277–286.

[38] ——, “Adversarial occupancy monitoring using one-sided through-wall wifi sensing,” in *ICC 2021-IEEE International Conference on Communications*. IEEE, 2021, pp. 1–6.

[39] S. M. Hernandez, D. Erdag, and E. Bulut, “Towards dense and scalable soil sensing through low-cost wifi sensing networks,” in *2021 IEEE 46th Conference on Local Computer Networks (LCN)*. IEEE, 2021, pp. 549–556.

[40] J. Liu, G. Teng, and F. Hong, “Human activity sensing with wireless signals: a survey,” *Sensors*, vol. 20, no. 4, p. 1210, 2020.

[41] K. Ohara, T. Maekawa, and Y. Matsushita, “Detecting state changes of indoor everyday objects using wi-fi channel state information,” *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 1, no. 3, pp. 1–28, 2017.

Publication List

Peer Review Journal

[1] **H. Choi**, M. Fujimoto, T. Matsui, S. Misaki, and K. Yasumoto, "Wi-CaL: WiFi Sensing and Machine Learning based Device-Free Crowd Counting and Localization," *IEEE Access*, Vol. 10, pp. 24395-24410, 2022.

International Conference (Thesis-related)

[1] **H. Choi**, T. Matsui, M. Fujimoto, and K. Yasumoto, "Simultaneous Crowd Counting and Localization by WiFi CSI," in *International Conference on Distributed Computing and Networking (ICDCN).* 2021, pp. 239-240.

[2] **H. Choi**, T. Matsui, S. Misaki, A. Miyaji, M. Fujimoto, and K. Yasumoto, "Simultaneous Crowd Estimation in Counting and Localization using WiFi CSI," in *2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN).* IEEE, 2021, pp. 1-8.

International Conference (Others)

[1] S. Fukuda, **H. Choi**, Y. Matsuda, and K. Yasumoto, "Fishing activity sensing and visualization system using sensor-equipped fishing rod: demo abstract," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems (SenSys).* 2020, pp. 615-616.

[2] S. Misaki, K. Umakoshi, T. Matsui, **H. Choi**, M. Fujimoto, and K. Yasumoto, "Non-Contact In-Home Activity Recognition System Utilizing Doppler Sensors," in *Adjunct Proceedings of International Conference on Distributed Computing and Networking (ICDCN Workshops).* 2021, pp. 169-174.

[3] K. Umakoshi, T. Matsui, M. Yoshida, **H. Choi**, M. Fujimoto, H. Suwa, and K. Yasumoto, "Non-contact Person Identification by Piezoelectric-Based Gait Vibration Sensing," in *International Conference on Advanced Information Networking and Application (AINA).* Springer, 2021, pp. 745-757.

[4] A. Miyaji, T. Matsui, Z. Zhang, **H. Choi**, M. Fujimoto, and K. Yasumoto, "Analysis on Nursing Care Activity Related Stress Level for Reductiom of Caregiving Workload," in *International Conference on Parallel Processing Workshop (ICPP Workshops).* 2021, pp. 1-8.