

THESIS

INDOOR POSITIONING WITH DEEP LEARNING FOR MOBILE IOT SYSTEMS

Submitted by

Liping Wang

Department of Electrical and Computer Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2022

Doctoral Committee:

Advisor: Sudeep Pasricha

Ryan Kim  
Jianguo Zhao

Copyright by Liping Wang 2022

All Rights Reserved

## ABSTRACT

### INDOOR POSITIONING WITH DEEP LEARNING FOR MOBILE IOT SYSTEMS

The development of human-centric services with mobile devices in the era of the Internet of Things (IoT) has opened the possibility of merging indoor positioning technologies with various mobile applications to deliver stable and responsive indoor navigation and localization functionalities that can enhance user experience within increasingly complex indoor environments. But as GPS signals cannot easily penetrate modern building structures, it is challenging to build reliable indoor positioning systems (IPS). Currently, Wi-Fi sensing based indoor localization techniques are gaining in popularity as a means to build accurate IPS, benefiting from the prevalence of 802.11 family. Wi-Fi fingerprinting based indoor localization has shown remarkable performance over geometric mapping in complex indoor environments by taking advantage of pattern matching techniques. Today, the two main information extracted from Wi-Fi signals to form fingerprints are Received Signal Strength Index (RSSI) and Channel State Information (CSI) with Orthogonal Frequency-Division Multiplexing (OFDM) modulation, where the former can provide the average localization error around or under 10 meters but has low hardware and software requirements, while the latter has a higher chance to estimate locations with ultra-low distance errors but demands more resources from chipsets, firmware/software environments, etc.

This thesis makes two novel contributions towards realizing viable IPS on mobile devices using RSSI and CSI information, and deep machine learning based fingerprinting. Due to the larger quantity of data and more sophisticated signal patterns to create fingerprints in complex indoor environments, conventional machine learning algorithms that need carefully engineered features suffer from the challenges of identifying features from very high dimensional data. Hence, the

abilities of approximation functions generated from conventional machine learning models to estimate locations are limited. Deep machine learning based approaches can overcome these challenges to realize scalable feature pattern matching approaches such as fingerprinting. However, deep machine learning models generally require considerable memory footprint, and this creates a significant issue on resource-constrained devices such as mobile IoT devices, wearables, smartphones, etc. Developing efficient deep learning models is a critical factor to lower energy consumption for resource intensive mobile IoT devices and accelerate inference time. To address this issue, our first contribution proposes the CHISEL framework, which is a Wi-Fi RSSI-based IPS that incorporates data augmentation and compression-aware two-dimensional convolutional neural networks (2D CAECNNs) with different pruning and quantization options. The proposed model compression techniques help reduce model deployment overheads in the IPS. Unlike RSSI, CSI takes advantages of multipath signals to potentially help indoor localization algorithms achieve a higher level of localization accuracy. The compensations for magnitude attenuation and phase shifting during wireless propagation generate different patterns that can be utilized to define the uniqueness of different locations of signal reception. However, all prior work in this domain constrains the experimental space to relatively small-sized and rectangular rooms where the complexity of building interiors and dynamic noise from human activities, etc., are seldom considered. As part of our second contribution, we propose an end-to-end deep learning based framework called CSILoc for Wi-Fi CSI-based IPS on mobile IoT devices. The framework includes CSI data collection, clustering, denoising, calibration and classification, and is the first study to verify the feasibility to use CSI for floor level indoor localization with minimal knowledge of Wi-Fi access points (APs), thus avoiding security concerns during the offline data collection process.

## ACKNOWLEDGEMENTS

I would like to thank all the individuals who inspired, helped, encouraged, and understood me to make the completion of this thesis possible.

First and foremost, I would like to show my greatest gratitude to my advisor, Prof. Sudeep Pasricha, who is nice and patient to me and directs me to get through various difficulties of my study. He is resourceful to provide me with his insightful advising to tackle challenging questions and ensured what I was doing is correct, important, and meaningful. He always offers me the opportunities to be the grader and teaching assistant of his courses, where I have gained knowledge beyond books from him and the wonderful students. With his continuous help, I gradually built up my knowledge system for doing more complex and advanced research. I would thank Prof. Sudeep Pasricha again for his kindness, encouragement, knowledge and all the help.

I also would like to express my respect and gratitude to my committee members, Prof. Kim and Prof. Zhao. Prof. Kim usually proposes great questions and shares his deep insights with us in EPIC group meetings. Prof. Zhao has been very friendly and taught me a lot in his robotics classes. Their indirect help will be always a valuable experience for me.

I would like to voice my love to a great person, my girlfriend, Xuewei Du. It might not be possible to express how grateful and lucky I feel to meet her. Her encouragement, patience, understanding, companionship, grace, and love give me hope and make me a better person.

I would like to take this chance to give my special thanks to my research partner, Saideep Tiku, a humorous and relaxing person who constantly shares his valuable experience with me even when he is busy with his own research.

In addition, I would like to thank the excellent lab mates in in Prof. Pasricha's EPIC lab: Ishan Thakkar, V. Yaswanth Raparti, Daniel Dauwe, Vipin K. Kukkala, Varun Kilenje, Jordan

Tunnell, Ninad Hogade, Febin Sunny, Joydeep Dey, Kamil Khan, Asif Mirza, Sooryaa Vignesh Thiruloga, Abhishek Balasubramaniam who make friends with me, share the knowledge with me and laugh with me.

Finally, I would like to thank my father, Faxuan Wang, and my mother Tianxiu Wang for your unspoken, unconditional, and unending love, to raise me and shape me.

I wish all of you health and happiness; safeness and peacefulness.

May all your wishes come true.

## TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iv
LIST OF RESEARCH PUBLICATIONS .....	xii
JOURNAL PUBLICATIONS: .....	xii
CONFERENCE PUBLICATIONS: .....	xii
1. INTRODUCTION .....	1
1.1. MOTIVATION.....	1
1.2. BACKGROUND .....	2
1.2.2. LOCALIZATION WITH RSSI .....	3
1.2.3. LOCALIZATION WITH CSI .....	5
1.3. RELATED WORK.....	8
1.3.1. RELATED WORK USING RSSI .....	8
1.3.2. RELATED WORK USING CSI.....	10
1.4. LIMITATIONS OF EXISTING WORK.....	13
1.5. THESIS OVERVIEW.....	14
2. CHISEL: COMPRESSION-AWARE HIGH-ACCURACY EMBEDDED INDOOR LOCALIZATION WITH DEEP LEARNING .....	16
2.1. INTRODUCTION .....	16
2.2. RELATED WORK.....	18
2.3. CHISEL FRAMEWORK.....	19
2.3.1. DATA PREPROCESSING AND AUGMENTATION.....	19
2.3.2. NETWORK ARCHITECTURE.....	21
2.3.3. MODEL COMPRESSION .....	21

2.4.	EXPERIMENTS .....	22
2.4.1.	EVALUATION ON UJIIndoorLoc DATASET .....	23
2.4.2.	EVALUATION WITH MODEL COMPRESSION.....	24
2.5.	CONCLUSION.....	27
3.	A FRAMEWORK FOR CSI-BASED INDOOR LOCALIZATION WITH 1D CONVOLUTIONAL NEURAL NETWORKS .....	28
3.1.	INTRODUCTION .....	28
3.2.	RELATED WORK.....	32
3.3.	DATA COLLECTION .....	34
3.3.1.	BUILDING AND PATH INFORMATION .....	34
3.3.2.	DISTRIBUTIONS OF REFERENCE POINTS AND TEST POINTS .....	35
3.3.3.	DATA COLLECTION PLATFORM.....	37
3.3.5.	DATA COLLECTION STRATEGY .....	39
3.4.	CSI DATA CLUSTERING, DENOISING AND CALIBRATION.....	41
3.4.1.	CSI DATA INTRODUCTION .....	41
3.4.2.	PATTERN CLUSTERING AND DENOISING .....	43
3.4.3.	RSSI BASED CSI CALIBRATION.....	48
3.5.	1D CNN-BASED INDOOR LOCALIZATION.....	50
3.5.1.	TRAINING SET AND TESTING SET.....	50
3.5.2.	NETWORK ARCHITECTURE .....	51
3.6.	EXPERIMENTAL RESULTS.....	52
3.7.	CONCLUSION.....	56
4.	CONCLUSIONS AND OPEN PROBLEMS .....	57
4.1.	RESEARCH CONCLUSION.....	57
4.2.	OPEN PROBLEMS .....	58
4.2.1.	HIGH WI-FI FACILITY COSTS USING RSSI .....	58
4.2.2.	HIGH NUMBER OF DEEP LEARNING PARAMETERS USING RSSI.....	58

4.2.3. LACK OF SOLUTIONS FOR MODEL SCALABILITY .....	59
4.2.4. SKEWED PATTERN NUMBERS USING CSI .....	59
4.2.5. LACK OF RESEARCH IN THE POSITION OF SIGNAL RECEPTION.....	60
4.2.6. LIMITED OPTIONS IN CSI EXTRACTION TOOL.....	60
4.2.7. UNDOCUMENTED AGC ALGORITHMS .....	61
4.2.8. LABOR INTENSIVE DATA COLLECTION.....	61
4.2.9. LACK OF SOLUTION FOR SEAMLESS ZONE SWITCHING USING CSI.....	61
4.2.10. LACK OF USER-FRIENDLY INDOOR LOCALIZATION APPLICATIONS.	62
4.2.11. LACK OF SOLUTIONS FOR LOCALIZATION SECURITY .....	62
4.2.12. LACK OF SOLUTIONS FOR DEVICE HETEROGENEITY .....	63
BIBLIOGRAPHY .....	64

## LIST OF TABLES

TABLE 1: CHISEL’s CAECNN network model layers.....	20
TABLE 2: Localization performance comparison.....	23
TABLE 3: Localization performance comparison.....	24
TABLE 4: CSI collection platform components .....	37
TABLE 5: Samples for each RP after clustering/denoising/calibration .....	50
TABLE 6: Network parameters and inference time (IT).....	54

## LIST OF FIGURES

Fig. 1. Illustration of RSSI-based indoor localization .....	4
Fig. 2. Illustration of CSI-based indoor localization.....	7
Fig. 3. Average CHISEL position error comparison across configurations with (a) post-training quantization and (b) quantization-aware training. Numbers on top of each bar represent the model’s average error (in meters). WO and WA, respectively, represent weights only and weights + activations quantization.....	24
Fig. 4. On-device inference time from all compression configurations of CHISEL. CHISEL FLOAT32 represents the baseline model. ....	27
Fig. 5. The floorplan of 2 <sup>nd</sup> floor of Colorado State University (CSU) Behavioral Sciences Building (BSB). Different colors represent corresponding AP visibility; each path may contain 1 or 2 APs based on received signal strength.....	35
Fig. 6. The distribution of RPs and TPs on the 2nd floor of CSU BSB. RPs and TPs are annotated with blue and red, respectively. ....	36
Fig. 7. A 64-subcarrier 802.11ac 20 MHz symbol .....	38
Fig. 8. Magnitude distribution of 3500 received CSI packets at the first RP of room 204 with 21 abnormal packets removed (3479 packets are visualized).....	39
Fig. 9. Extracted CSI magnitude and phase from 3479 packets at location 204_1 with 21 abnormal packets removed. (a) Magnitude; (b) Phase in degree; (c) Unwrapped phase in radius; (d) Magnitude spectrum corresponding to Fig. 6. ....	42
Fig. 10. Extreme magnitude values (spikes) exist in CSI packets. ....	43
Fig. 11. The average silhouette score given by 9 clusters .....	44
Fig. 12. The 9 magnitude patterns of 3479 CSI packets at RP location 204_1. ....	45
Fig. 13. The 9 phase patterns of 3479 CSI packets at RP location 204_1. ....	46
Fig. 14. (a) CC score histogram and (b) RMSE distance histogram, for pattern 2 at the location 204_1 w.r.t the mean sequence. ....	47
Fig. 15. The denoising process with the pattern-wise denoising algorithm. Top: the original pattern 2 packets; middle: pattern 2 packets after CC filtering; bottom: pattern 2 packets after RMSE filtering.....	48
Fig. 16. The illustration of 1D CNN for location classification .....	51

Fig. 17. Comparison of the localization errors based on CDF for each AP ..... 53  
Fig. 18. Mean distance error comparison between 4 models for each AP..... 55

## LIST OF RESEARCH PUBLICATIONS

### JOURNAL PUBLICATIONS:

- L. Wang, S. Tiku, and S. Pasricha, "CHISEL: Compression-Aware High-Accuracy Embedded Indoor Localization with Deep Learning," **IEEE Embedded Systems Letters, 2021.**

### CONFERENCE PUBLICATIONS:

- L. Wang and S. Pasricha, "A Framework for CSI-Based Indoor Localization with 1D Convolutional Neural Networks," **12<sup>th</sup> International Conference on Indoor Positioning and Indoor Navigation, 2022** (under review).

### BOOK CHAPTERS:

- S. Tiku, L. Wang, S. Pasricha, "Machine Learning Model Compression for Efficient Indoor Localization on Embedded Platforms," *Embedded Machine Learning for Cyber-Physical, IoT, and Edge Computing*, to appear, Springer Nature, 2022

## 1. INTRODUCTION

This chapter introduces the motivation, background, and related work about indoor localization systems using received signal strength indicator (RSSI) and channel state information (CSI) with Wi-Fi fingerprinting. Wi-Fi signal fingerprinting has become a popular approach to train indoor localization models in recent years because of its strong ability to organize signals from complex indoor environments where good Line-Of-Sight (LOS) conditions are hard to satisfied. Meanwhile, with the powerful feature extraction from deep learning, fingerprinting based indoor positioning systems show the impressive performance dealing with large amounts of complicated data patterns than ever used in conventional machine learning algorithms.

### 1.1. MOTIVATION

The geolocation services of today have reduced the need for cumbersome paper-based maps, which were the standard navigational tool in the past. Mapping, localization, and navigation technologies have revolutionized the way we connect with the outside world. Indoor localization becomes a crucial component of today's edge, mobile, and Internet of Things (IoT) devices due to the constantly increasing need for intellectual and human-centric indoor services. By providing responsive positioning capabilities, indoor navigation systems in large facilities such as museums, libraries, and shopping malls have been enhancing user experiences. Stable and accurate indoor localization systems are essential for extremely sensitive indoor positioning use cases, such as human activity identification in hospitals and robot tracking and position calibration in contemporary industries. However, due to the restricted permeability of GPS signals within indoor environments, such services are not available within building structures. In extreme circumstances, such as dangerous indoor environments in the interiors of a nuclear power plant, precise indoor

positioning is required for ground robots to perform radiation detection and mitigation operations [1]. In an effort to extend localization services to buildings, subterranean locations and unmanned operations under extreme indoor conditions, interest in indoor localization solutions has recently increased [2].

## 1.2. BACKGROUND

Wi-Fi devices have been pervasively adopted as the dominant infrastructure for indoor positioning, due to their low cost and the excellent universal accessibility. According to CISCO's forecast, the IP traffic generated by Wi-Fi and mobile devices is reaching 71 percent in 2022 globally, while about 549 million public Wi-Fi hotspots are going to be available by the end of the year [3]. The 802.11ac is reported to be the most prevalent standard of 802.11 family that will share 66.8 percent of WLAN endpoints by 2023 [4]. Hence, we use RSSI measured from Wi-Fi signals and extract CSI from 802.11ac standard to conduct our research.

### 1.2.1. FINGERPRINTING-BASED INDOOR LOCALIZATION

Geometric mapping and feature pattern mapping (also known as fingerprinting) are two categories of popular indoor localization techniques [5]. The former measures predefined parameters such as power, distance, and direction observations, etc., relative to reference points (RPs) before computing positions using geometric conversion procedures such as triangulation. In contrast, the objective of fingerprinting is to detect matched feature patterns, therefore conversion methods are often unneeded. Specifically, after the feature space is constructed, the coordinates of an unknown location can be estimated by comparing the feature pattern acquired from an unknown location to the reference pattern space. The measurements of directional and distance information necessary in geometric mapping-based algorithms significantly rely on LOS conditions, which are

typically difficult to meet in complicated indoor situations. In addition, the non-negligible approximations of conversion algorithms result in an inevitable accuracy decrease. In contrast, fingerprinting is viewed as a superior approach for complex interior scenarios, as pattern matching does not need to account for LOS circumstances or conversion techniques.

The offline phase of creating fingerprints involves gathering Wi-Fi signal characteristics such as RSSI from multiple indoor locations or RPs within a building. The vector of wireless signal RSSI values from all access points (APs) observed at a given indoor location from the location's fingerprint. Such fingerprints collected across RPs during the offline phase constitute a fingerprint dataset. Using this offline dataset, a machine learning (ML) model can be developed and deployed on embedded and IoT devices (such as smartphones) using Wi-Fi transceivers. In the second phase, known as the online phase, the Wi-Fi RSSI obtained by a user is transmitted to the ML model and used to estimate the user's location within the building.

### 1.2.2. LOCALIZATION WITH RSSI

The data collection platform of RSSI is universally accessible with most off-the-shelf devices on which a Wi-Fi scanning function is provided. Indoor localization using RSSI fingerprints is an approach based on distance information measured via RSSI. When a positioning request is sent out from a mobile client at user's current location, a received RSSI vector is created from connected Wi-Fi APs by recording the snapshot of each RSSI value from all APs that are programmed to provide the indoor localization service. In classification modeling, by comparing the received vector and all vectors collected from all RPs offline, the location is recognized. In contrast, a regression model will estimate values (typically the Cartesian coordinates) based on input features of the received vector. Fig. 1 gives an intuitive illustration of indoor localization

with RSSI. Each RSSI value is a scalar corresponding to one feature dimension provided by one AP of a single fingerprint, which means the resolution of a RSSI fingerprint is determined by the number of APs. A single RSSI fingerprint is essentially a vector consisting of the current distance information of user with respect to all APs that are involved in. As the relative distances between the receiver and APs change while a user moves in a building, each element in the received RSSI fingerprint is updated which changes in location. Thus, unique RSSI fingerprints are formed for each indoor location. Each AP may need to be given a unique ID with a certain order to create a meaningful RSSI fingerprint. When a user is moving in a building, the snapshot RSSI fingerprint is being dynamically updated while the current location of the user relative to each AP is changing, thus the position tracking under a reasonable sampling frequency can be also achieved.

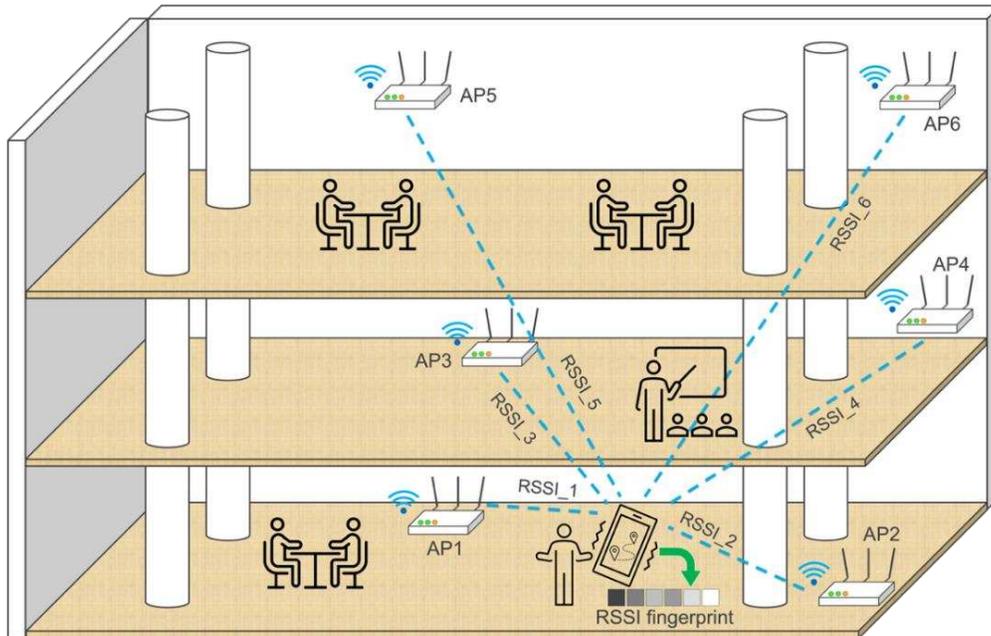


Fig. 1. Illustration of RSSI-based indoor localization

### 1.2.3. LOCALIZATION WITH CSI

In fingerprinting, physical characteristics such as RSSI, time of arrival (ToA), and angle of arrival (AoA) are frequently employed. However, besides random fluctuations in RSSI, ToA is unreliable in complex indoor situations because signal arrival time is greatly dependent on LOS circumstances. AoA observations with distance measurements gathered by directional antennas via MIMO channel matrices are regarded as a relatively robust technique for capturing fine-grained characteristics and avoiding multipath signal interferences. However, the limited availability of specialized hardware and software makes direct AoA measurements difficult for common mobile and resource-constrained devices, such as smartphones and wearables.

Fortunately, the aforementioned problems are being resolved by exploiting the benefits of contemporary wireless technologies. Standards of Wi-Fi family since 802.11n (Wi-Fi 4) theoretically establish the hardware requirements for defining a wireless MIMO channel [6]. With the assistance of Orthogonal Frequency-Division Multiplexing (OFDM), high-accuracy indoor localization that uses the information contained within Wi-Fi frames to differentiate transmitted signals from distinct pathways is made possible. Nevertheless, CSI extraction is frequently infeasible with the majority of commercially available equipment. In recent years, great effort has been made to address this issue. An early tool for Intel cards released in 2011 [7] is able to extract CSI from 802.11n devices, although the project is not accessible to the public. In addition, the CSI can only be derived from particular subcarriers and the bandwidth is restricted to between 20 and 40 MHz. In 2015, Wireless and Networked Distributed Sensing (WANDS) released the Atheros CSI tool [8], which is based on the ath9k Linux kernel driver and offers CSI extraction over all subcarriers for Qualcomm Atheros chips such as AR9380, AR9580/90, QCA9558, etc. This utility allows 10-bit real and imaginary CSI extraction from OFDM modulation with operating

bandwidths of 20 and 40 MHz. Unfortunately, both tools only offer limited support for 802.11n, whilst 802.11ac, the current most prevalent standard, cannot be profiled using either tool. Nexmon [9] is a C-based firmware patching framework with the currently most extensive support for several Broadcom and Cypress Wi-Fi chipsets. Secure Mobile Networking Lab (SEEMOO) proposes it with its most recent version launched in [10], which enables 20/40 and 20/40/80 MHz per frame CSI extraction on 802.11n/ac, respectively. We employ Nexmon CSI extractor to harvest CSI data on the second floor of the Behavioral Science Building on the main campus of Colorado State University (CSU BSB) due to the high prevalence of the 802.11ac standard. RSSI data used in our CSI framework to calibrate CSI magnitude are extracted simultaneously with CSI packets using Nexmon RSSI patch for bcm 43455c0 chipset. The purpose of this work is to provide a CSI dataset for low-energy-consumption embedded and IoT devices, hence a Raspberry Pi 4 Model B with a single receiving antenna is used for our CSI research.

Compared with RSSI's packet-level measurement, CSI-based indoor localization takes advantages of observations inside each Wi-Fi packet. Suppose a user moves into the coverage of an AP and sends out a request for their position, a magnitude sequence is extracted from a received CSI packet via the wireless link between the user's device and the connected AP. The received sequence carries magnitude compensation information modified by the wireless propagation path (through reflection, absorption and diffraction, etc., over different interior materials) where the received signal traveled along. The magnitude values of the corresponding subcarriers together form a feature that can be utilized for classification or regression of a ML system that converts the feature to a RP with the most similar sequence pattern or estimated Cartesian coordinates, correspondingly. Thus, the location information is returned to the user.

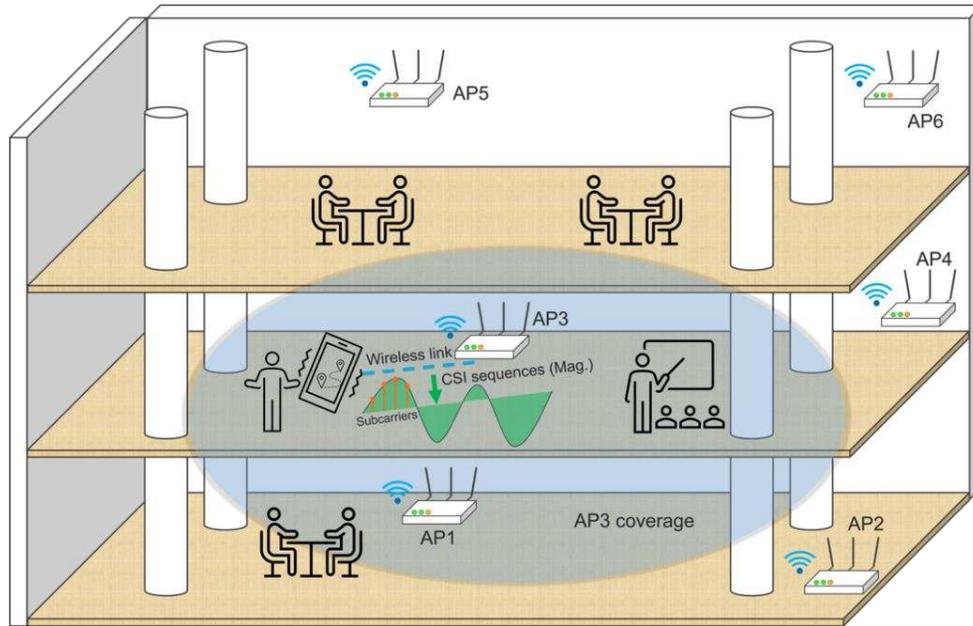


Fig. 2. Illustration of CSI-based indoor localization

Fig. 2 shows a brief example of how a localization is served using signal magnitude data extracted from a CSI packet sent out from AP3. The CSI subcarriers provide frequency diversities as the different frequency components have different sensitivities of modifications by the wireless propagation (for example, high-frequency components fade faster) even with an exact same wireless path. Compared with RSSI-based IPS, whose fingerprints are created by numbers of APs at the same signal reception slot, IPS built on CSI is often AP bound. This is mainly because the characteristics of different APs can have huge differences in the number of antennas, the angles of antennas and 802.11 standards, etc. In addition, the compensations of magnitude and phase shifting at the receiver end after the signal propagations from different wireless paths are highly dependent on the position and circumstances of the AP that emits signals. This is also the reason why CSI-based IPS are sensitive and can achieve ultra-low distance errors. In other words, a single RSSI is a value only determined by the physical distance between the signal source and the receiver, although noised with multipath effects and shadowing. Compared with RSSI, CSI delivers more

detailed information including propagation paths, AP behaviors and propagation modifications under frequency diversities. Plus, the phase information extracted from CSI packets has a better anti-noise nature which is further explained in chapter 3.

### 1.3. RELATED WORK

#### 1.3.1. RELATED WORK USING RSSI

As mentioned earlier, trilateration/triangulation using AoA or ToA observations is a common and well-studied approach to indoor localization [11]. However, AoA adds hardware complexity and is extremely sensitive to computational error, especially when the distance between the transmitter and receiver is large. ToA requires precise synchronization, and even with sufficient resolution from signal bandwidth and sampling rate, high localization errors are common, especially when no LOS paths are available, which is frequently the case in modern buildings. Both methods, however, necessitate precise knowledge of AP locations, making them unsuitable for many indoor environments.

Fingerprinting-based approaches have been shown to overcome many of these challenges because they do not need strict synchronization or knowledge of AP locations. Many RSSI fingerprinting-based ML solutions for indoor localization have been proposed, including approaches based on support vector regression (SVR) [12], k-nearest neighbors (KNNs) [13], and random forest (RF) [14].

Deep learning based fingerprinting methods have outperformed traditional ML approaches in recent years, demonstrating their promise. Many deep learning techniques have been applied to indoor localization [11, 15]. Jang and Hong [15] proposed a CNN classifier, whereas Nowiki and Wietrzykowski [16] developed a model that included a stacked autoencoder (SAE) for feature space reduction followed by a DNN classifier (SAEDNN). The experiments were carried out on

the UJIIndoorLoc dataset [17] in order to predict the building and floor on which a user is located. However, these works do not take into account positioning the user within a given floor, which is a much more difficult problem. Simultaneously, another SAEDNN-based model with fewer nodes in the final layer was proposed in [18]. Later, a 1-D CNN approach [19] outperformed [18] in terms of building and floor accuracy. This is accomplished by deploying separate CNN models for building, floor, and within floor prediction, which need higher memory footprint and computational costs.

While previous research has produced promising deep learning-based algorithms for indoor localization, they have consistently ignored deployability concerns caused by the memory and computing burdens of embedded and IoT devices. Although post-training model compression approaches can help to alleviate these deployment concerns, they may result in an unpredictable loss of localization precision.

To build robust RSSI-based indoor localization systems in practice, localization speed, device heterogeneity, security, and aging resilience need to be thoroughly considered. The inferencing speed of deep learning models critically affects the positioning speed of entire indoor localization systems, especially when real-time position tracking tasks are required. QuickLoc [20] proposes a deep learning based framework that allows conditional early exiting during the inferencing stage of a neural network. The core idea behind this work is that the relative accuracy differences after each convolutional layer is enough to predict an unknown location as a reference point for a classification model where localization can be done after layers, not necessarily output by the final layer. This work significantly improves location estimation latency by up to 42% and meanwhile, up to 45% energy reduction is achieved. It is inevitable for users to have different mobile devices in real-life, which means indoor localization-based services should not be bound

to limited device types. This gives rise to the heterogeneity issue. In SHERPA-HMM [21], a heterogeneity resilient framework based on Hidden Markov Model (HMM) promising the feasibility to port indoor localization algorithms across different smartphones enhances localization accuracy up to 8×, against state-of-the-art studies. As demonstrated in [22], Wi-Fi attackers can easily inject frames via Wi-Fi APs to modify RSSI values for fooling RSSI-based indoor localization algorithms. Users can be tracked or directed to a place that is far from his/her original destination up to 50 m. [22] proposes S-CNNLOC to mitigate such attack scenarios. Fingerprints highly depends on the stability of APs' characteristics. In addition, when an AP's behavior (e.g., operating channel) is updated or an AP is replaced, the component of a fingerprint created from this AP can be drastically changed. Thus, a catastrophic accuracy degradation could happen in a pretrained localization system. To solve this issue, STONE [23] trains a KNN classifier with the encoded information of RSSI fingerprints generated from Siamese neural encoder offline to predict locations, which provides up to 40% localization accuracy decrease over time to enhance the resilience of an IPS in a long-term window.

### 1.3.2. RELATED WORK USING CSI

Traditional RSSI-based fingerprinting methods are affected by fluctuations introduced by multipath and shadowing effects, which can alter RSSI values by up to 5 dBm [24]. Thus, fingerprinting approaches based on RSSI may result in less precise indoor localization. CSI is regarded as an improved wireless propagation descriptor and can enhance the localization accuracy with Wi-Fi sensing [25]. CSI data extracted from the physical layer (PHY) of fifth generation (and higher) Wi-Fi frames represents the frequency response of a Multiple Input Multiple Output (MIMO) channel and is capable of providing the sensitive parameters (e.g., magnitude attenuation and phase shifting) required for capturing signal sources. In the collected CSI data, compensations

of magnitudes and phases during signal transmissions corresponding to each subcarrier with different frequencies are provided, allowing the identification of propagation routes that are distinct from one another. In contrast to RSSI, CSI analysis may detect even minor changes in the position of receiving antennas, hence increasing the likelihood of achieving ultra-low localization errors.

As part of the FIFS framework, an early CSI fingerprinting-based indoor localization algorithm with WLANs was conducted in [26]. The Bayes' theorem was used by FIFS to determine the maximum posteriori probability of a specific RP given the knowledge of CSI fingerprints received from three Aps during the online positioning stage, and the estimated location is given by the coordinates of that RP. The use of spatial correlation of the CSI to determine the prior probabilities of RPs [26] for estimating unknown locations is an important part of this work. However, in a large and complex indoor space, the CSI from two locations separated by a short distance can be weakly unrelated, making FIFS unstable. In addition, the distribution of TPs is not described in this work which makes it impossible to reproduce the performance of FIFS.

Several other projects have investigated CSI-based localization using deep learning models. DeepFi [27] extracts magnitude features of successive CSI packets from three transmitting antennas using a deep stacked autoencoder (SAE). The SAE is individually pretrained for each RP in the offline stage. It is possible to generate accumulated reconstruction errors by feeding CSI fingerprints from an unknown site into a pretrained SAE, and then comparing the results with those from the offline stage of reconstruction at RPs. The number of SAEs required, which grows with the number of RPs, is a significant disadvantage. A CNN classification approach named CiFi is proposed in [8] where fingerprints are produced utilizing angle of arrival (AoA) information to predict locations, using phase difference information from two nearby antennas based on 5GHz

Wi-Fi. When LOS circumstances are poor and magnitude data are greatly attenuated, the authors note that AoA's estimation stability is better than magnitude values. However, CSI is essentially computed by alterations of the known preamble information via wireless routes [10]. The compensations of the weakened magnitude values can convey useful information for determining multipath propagation that is valuable for tracing back the location of the signal source. A CNN model based on CSI magnitudes gathered in a five-room residence is built by DelFin [28]. When used in a home or a small office, DelFin just requires a single anchor transmitter. Although this method is lightweight and suited for IoT devices, the indoor environment evaluated is tiny and simple, without addressing dynamic interference. OpenCSI [29] is an open source project that presents a solution for automating CSI collection in a  $3.5\text{m} \times 5\text{m}$  space. To extract CSI from the Long-Term Evolution (LTE) eNodeB, a radio map is constructed using an SDR mounted on a wheeled robot. A 1D CNN model is applied to utilize magnitude and phase information. The dataset is publicly available, but the experimental space is limited which ignores the noise introduced in dynamic environments and prohibits the investigation of complex interference effects. In SDR-Fi [30], a feed-forward neural network (FFNN) and a 1D CNN model are developed to leverage CSI magnitude for position prediction in a space of around 60 square meters, with a 1D CNN demonstrating the greater potential to produce reduced distance errors. Although experimental results indicate that the greater the number of packets received, the smaller the distance error estimated, however, SDR-Fi disregards phase information, which has a better anti-noise nature than magnitude. This work does not fully interpret the fundamental reason why the number of received packets matters, particularly in a complex environment. In addition, the SDR-Fi dataset is not accessible to the public.

Although related studies have proven using CSI data can achieve much lower distance errors than using CSI, there are still several drawbacks. First, the majority of cutting-edge research utilizing CSI data is limited to localization within tiny, isolated rooms; complex building interiors and dynamic surroundings, including human activities, are rarely taken into account. Based on our studies, the multipath and shadowing effects on CSI data received by a single antenna in a cuboidal-shaped room are exceedingly limited and predictably concentrated. In more complex interior contexts, CSI patterns become more complicated and comprise a greater variety of pattern groups, primarily because of an increase in wireless propagation paths. Moreover, we discovered that the internal structure of a building significantly modifies CSI characteristics during signal transmissions. Even if these structures determine the number of pattern groups, the question of how to denoise each group remains unanswered. In Section 3.2.1., we discuss a novel way to managing this phenomenon that we propose. Second, prior work requires information about Wi-Fi Aps, such as the number and identification of antennas. Access to comprehend each AP may be restricted in practice due to security regulations, and it is not efficient nor practical to investigate each AP's specifications during the offline fingerprint data gathering phase. Third, the datasets in the majority of earlier publications are usually not publicly accessible, limiting other researchers from reproducing the results and advancing the preceding research. Consequently, a solution is required to enable an efficient, reproducible, and large-scale deployment of CSI-based indoor localization without detailed AP knowledge.

#### 1.4. LIMITATIONS OF EXISTING WORK

- **Lack of lightweight and efficient models:** Prior research provides potential deep learning based techniques for indoor localization, however, they consistently overlook deployability issues deriving from memory and computational overheads in embedded

and IoT devices. On the other hand, deep learning models commonly have redundant parameters mainly including weights and activations which may not contribute closely-related gains to the core task of deep learning models, but critically slow down the inferencing speed in real-time applications.

- **Accuracy:** The accuracy level of current indoor localization algorithms needs to be improved. Even with the coarse localization ability from RSSI, the accuracy can be optimized once a proper deep learning model is applied.
- **Limited space:** This issue occurs particularly in CSI based indoor localization studies. The prior work conducts experiments in space with limited sizes and usually do not consider the dynamic noise and interference flooding in complex modern buildings.
- **Accessibility in CSI data collection:** Most prior work needs to modify transmitter's behavior to identify 802.11 frames in order to do indoor localization. However, the physical access and traffic modification on APs are hard to satisfied in practice.
- **Limited interpretation of CSI patterns in complex indoor environments:** Since most prior work limited the experimental space in small rooms, the CSI patterns are predictably concentrated which could diminish the value of practical use in large complex indoor space.

## 1.5. THESIS OVERVIEW

In this thesis, we first introduce a CHISEL aiming at offering a lightweight neural network architecture as well as the higher localization accuracies compared with stage-of-the-art deep learning algorithms using RSSI. Next, a lightweight framework based on 1D convolutional neural networks (CSILoc) for a floor-level CSI dataset with dynamic noise is presented, to solve the

current limitations in CSI research listed in 1.4. The thesis is organized by separating CHISEL and CSILoc into chapter 2 and chapter 3, respectively. In chapter 2, RSSI is adopted as the signal source for developing a deep learning based indoor localization framework that is compression tolerant called CHISEL. CHISEL integrates data augmentation, compression-aware 2D stacked convolutional autoencoder convolutional neural network (CAECNN), pruning and quantization as an entire pipeline to deliver a light weight, robust, compression-friendly and high accuracy solution that is benchmarked with state-of-the-art RSSI-based deep learning models on the same public open dataset [17]. In this part, we introduce the flexibility of CHISEL's CAECNN architecture can be easily achieved by adjusting the number of CAE layers. Different combinations of pruning and quantization configurations are applied on CAECNN to trade off compressed CHISEL submodels.

In chapter 3, a framework tailored for CSI-based indoor localization in large and complex indoor spaces is introduced. In this work, we firstly propose a series of techniques to tackle the difficult issues related to the feasibility for CSI-based indoor positioning algorithms in floor-level dynamic spaces, including defining RP pattern spaces, pattern clustering, pattern-wise denoising and RSSI-based CSI calibration, which has not been discussed in prior work. An associated 1D CNN is proposed to classify unknown locations that has the minimum number of parameters while keep the lowest localization errors against stage-of-the-art deep learning models over 6 APs.

In chapter 4, this thesis is concluded. We summarized our indoor localization research and proposed some open problems that can be addressed in future work.

## 2. CHISEL: COMPRESSION-AWARE HIGH-ACCURACY EMBEDDED INDOOR LOCALIZATION WITH DEEP LEARNING <sup>1</sup>

In this chapter, we propose CHISEL [31] where a compression-aware CAECNN architecture is presented. It permits a flexibility for compressing input tensor size by changing configuration of the CAE part which can downsample the input shape while keeps low losses in its latent space. In addition to this, we prove that its structure can still achieve a high localization accuracy even under some aggressive compressing combinations from pruning and quantization as the same time.

### 2.1. INTRODUCTION

Today's geolocation services have eliminated the need for cumbersome paper-based maps that were the dominant navigation strategy of the past. Outdoor mapping, localization, and navigation technologies have reinvented the way we interact with the world around us. However, due to the limited permeability of GPS signals within indoor environments, such services do not function in buildings, such as malls, hospitals, schools, etc. In an effort to extend localization services to buildings and subterranean locales, indoor localization solutions are experiencing a recent upsurge in interest [6].

While substantial progress has been made in this area (see Section II), Wi-Fi fingerprinting-based indoor localization stands out as the most promising solution. This is mainly due to the ubiquitous nature of Wi-Fi access points (APs) and their signals in buildings, and the superior localization accuracies demonstrated with it. Fingerprinting consists of two phases: the first phase,

---

<sup>1</sup> This chapter has been published as a research article in *IEEE Embedded System Letters* with the following citation: L. Wang, S. Tiku, and S. Pasricha, "CHISEL: Compression-Aware High-Accuracy Embedded Indoor Localization with Deep Learning," *IEEE Embedded Systems Letters*, 2021.

known as the offline phase, consists of collecting Wi-Fi signal characteristics such as received signal strength indicator (RSSI) at various indoor locations or reference points (RPs) in a building. The vector of wireless signal RSSI values from all APs observable at an indoor location represents a fingerprint of that location. Such fingerprints collected across RPs in the offline phase form a fingerprint dataset, where each row in the dataset consists of an RSSI fingerprint along with its associated RP location. Using this dataset from the offline phase, a machine learning (ML) model can then be trained and deployed on embedded devices (e.g., smartphones) equipped with Wi-Fi transceivers. In the second phase, called the online phase, Wi-Fi RSSI captured by a user is sent to the ML model and used to predict the user's location in the building.

Recent works report improved indoor localization accuracy through the use of convolutional neural network (CNN) models [1]. This is mainly attributed to the superior ability of CNNs at discerning underlying patterns within fingerprints. Such CNN models can be deployed on smartphones and allow users to localize themselves within buildings, in real time. Executing these models on smartphones instead of the cloud further enables security and sustainability as it eliminates the need of user data being shared through unsecured networks [2].

Unfortunately, research in the domain of indoor localization overlooks the high memory and computational requirements of CNNs, making deployment on resource-constrained embedded systems, such as smartphones, a challenge. While post-training model compression can ease model deployability in the online phase, it leads to an unpredictable degradation in localization performance. Thus, there is a need for holistic deep learning solutions that can provide robust indoor localization performance when deployed on embedded devices.

In this chapter, we propose a novel multidimensional approach toward indoor localization that combines convolutional autoencoders, CNNs, and model compression to deliver a sustainable and

lightweight framework called CHISEL. We benchmark the performance of CHISEL against state-of-the-art ML and deep-learning-based indoor localization frameworks with an open indoor localization database to quantify its superior performance and lower overheads.

## 2.2. RELATED WORK

A conventional and fairly well-studied approach to indoor localization is through trilateration/triangulation using angle-of-arrival (AoA) or time-of-flight (ToF) methods [2]. However, AoA introduces hardware complexity and is very sensitive to computational error, especially when the distance between the transmitter and receiver becomes large. ToF needs tight synchronization requirements and even with enough resolution from signal bandwidth and sampling rate, high localization errors are common, especially when no line-of-sight paths are available, which is often the case in buildings. Both of these methods also require precise knowledge of AP locations, making them impractical for many indoor environments.

Fingerprinting-based approaches have been shown to overcome many of these challenges as they do not require rigid synchronization and knowledge of AP locations, and are also less immune to multipath and shadowing effects [2]. Many RSSI fingerprinting-based ML solutions have been proposed for indoor localization, e.g., approaches using support vector regression (SVR) [3], k-nearest neighbors (KNNs) [4], and random forest (RF) [5]-based indoor location estimators.

Recent years have shown the promise of deep learning based fingerprinting methods that have outperformed classical ML approaches. Many deep learning techniques have been adapted to the domain of indoor localization [2], [6]. The work by Jang and Hong [7] proposed a CNN classifier while Nowiki and Wietrzykowski [8] built a model consisting of a stacked autoencoder (SAE) for feature space reduction followed by a DNN classifier (SAEDNN). The experiments were conducted on the UJIIndoorLoc dataset [11] to predict the building and floor that a user is located

on. However, these works do not consider positioning the user within a given floor, which is a much harder problem. At the same time, another SAEDNN-based model was proposed in [9] that reduced the number of nodes in the final layer. Later, a 1-D CNN approach [10] was shown to outperform [9] with better building and floor accuracies. This is achieved through the additional overhead of deploying separate CNN models for building, floor, and within floor prediction, which has high memory and computational costs.

While previous works propose promising deep-learningbased approaches for indoor localization, they consistently overlook deployability issues arising from memory and computational overheads in embedded devices. Post-training model compression techniques can help mitigate these deployment issues, but may lead to an unacceptable loss in localization accuracy. Our proposed CHISEL framework is amenable to model compression without notable loss in localization accuracy. This framework is discussed next.

## 2.3. CHISEL FRAMEWORK

### 2.3.1. DATA PREPROCESSING AND AUGMENTATION

In this work, we make use of the UJIIndoorLoc indoor localization dataset [17] that covers a total of three buildings and five floors. Our approach considers a total of 905 unique RPs, such that each RP represents a unique combination of [building ID/floor ID/space ID/relative position ID]. Here, the space ID is used to differentiate between the location inside and outside a room. Relative position ID locates the user on a given floor. The RSSI values for Wi-Fi APs vary in the range of  $-100$  to  $0$  dBm, where  $-100$  dBm indicates no signal and  $0$  indicates the full signal strength. We normalize this fingerprinting dataset to a range of  $0-1$ . As there is no test data in the UJIIndoorLoc dataset, we utilize the validation component (1111 samples) of the dataset as the

test set. The training portion of the dataset is split into training (15 950 samples) and validation (3987 samples) subsets, based on an 80:20 split.

To compensate for the limited samples per RP and to further improve generalization, we augment the fingerprint dataset. For each RP, we first calculate the mean value of all nonzero RSSI APs within one RP and the absolute difference between the mean value of each AP value. Then, we generate the AP RSSI values from the uniform distribution between the difference range obtained from the first step. The final dataset is the combination of the original and augmented fingerprints.

Considering our use of convolutional deep learning networks, each fingerprint is zero padded and translated into a single-channel square-shaped image, similar to the work in [32]. For the UJIIndoorLoc dataset, this produced  $24 \times 24 \times 1$  dimensional images. This new fingerprint image-based dataset is then used to train the deep learning model described in the next section.

TABLE 1: CHISEL’s CAECNN network model layers

Layer Type	Layer Size	Filter Count	Filter Size	Stride Value	Output Size
<b>CAE: Encoder</b>					
Input	—	—	—	—	$24 \times 24 \times 1$
Convolutional	$24 \times 24$	16	$3 \times 3$	$1 \times 1$	$24 \times 24 \times 16$
Max-Pooling	—	1	$2 \times 2$	$2 \times 2$	$12 \times 12 \times 16$
Convolutional	$12 \times 12$	8	$3 \times 3$	$1 \times 1$	$12 \times 12 \times 8$
<b>CAE: Decoder</b>					
Up-Sampling	—	1	$2 \times 2$	$2 \times 2$	$24 \times 24 \times 8$
Convolutional	$24 \times 24$	1	$3 \times 3$	$2 \times 2$	$24 \times 24 \times 1$
<b>CNN Classifier</b>					
Convolutional	$12 \times 12$	8	$3 \times 3$	$1 \times 1$	$12 \times 12 \times 8$
Convolutional	$12 \times 12$	16	$3 \times 3$	$1 \times 1$	$12 \times 12 \times 16$
Max-Pooling	—	1	$2 \times 2$	$2 \times 2$	$6 \times 6 \times 16$
Convolutional	$6 \times 6$	32	$3 \times 3$	$1 \times 1$	$6 \times 6 \times 32$
Convolutional	$6 \times 6$	32	$3 \times 3$	$1 \times 1$	$6 \times 6 \times 32$
Max-Pooling	—	1	$2 \times 2$	$2 \times 2$	$3 \times 3 \times 32$
Flatten	$1 \times 288$	—	—	—	$1 \times 1 \times 288$
Fully-Connected	$1 \times 128$	—	—	—	$1 \times 1 \times 128$
Batch Norm	$1 \times 128$	—	—	—	$1 \times 1 \times 128$
Softmax	$1 \times 905$	—	—	—	$1 \times 1 \times 905$

### 2.3.2. NETWORK ARCHITECTURE

Table 1 shows our proposed deep learning model which contains the CAE and CNN components that are trained in two stages. In the first stage, a CAE comprising of an encoder and a decoder network is trained with the goal of minimizing the MSE loss between the input and the output fingerprint. This process enables the CAE: Encoder to efficiently extract hidden features within the input fingerprints. In the second stage, the decoder is dropped and replaced by the CNN classifier as given in Table 1. The goal of this classifier is to predict the user’s location, given the encoded input from the CAE. The model is then retrained with the weights associated with the encoder frozen in place and loss function set to sparse categorical cross-entropy. The rectified linear unit (ReLU) is the only activation function we used for all convolutional and fully connected layers. The full model has 171 209 parameters in total.

### 2.3.3. MODEL COMPRESSION

We further employ a combination of two approaches to compress the parameters of our CAECNN model, for efficient deployment on embedded and IoT devices, as discussed below.

*Quantization:* The parameters of a neural network, in general, are represented by 32-bit wide floating-point (FP) values. However, a single neural network model can consist of hundreds of thousands to millions of parameters, leading to very large memory footprints and high inference latency. Quantization achieves model compression and computational relaxation by limiting the bit width used to represent weights and activations. However, this can lead to unpredictable accuracy degradation due to a reduction in FP precision [33]. To overcome this issue, researchers have proposed quantization-aware training (QAT) which involves quantizing weights and/or activations before the training process actually begins [33]. In this work, we evaluate both post-training and training-aware quantization across a range of bit widths, and the results of this analysis

are presented in Section 2.2.2. We explored quantization levels ranging from 32-bits down to 2-bits and applied a uniform quantizer to all convolutional layers keeping an equal number of positive and negative representations of parameters. Scaling the input tensors can further improve quantized model accuracy [33]. We calculated scaling factors channel by channel using averaging absolute values of weights in each channel. In addition, to overcome the issue of vanishing gradients, we applied the “straight-through estimator” (STE) [34] to the standard ReLU activation function during training.

*Pruning:* This approach involves selectively removing weights from a trained model. Out of the many pruning methodologies, such as filter pruning, layer pruning, connection pruning, etc., we employ connection pruning and filter pruning due to their diverse applicability and promising results across different kinds of models and layers [33]. Toward this goal, we implemented sparse connection pruning and filter pruning that are focused on zeroing out either a weight value or entire filters based on their magnitude [35, 36]. To achieve a sparsity of  $S\%$ , the weights of the model are ranked based on the magnitude, and the smallest  $S\%$  are set to zero. In the case of filter pruning, we utilize the L2-norm on the filter weights in order to rank them. We performed connection + filter pruning with varying sparsity values of 0% (no pruning), 25%, 50%, and 75% for the CHISEL model to identify the best configuration, as discussed next.

## 2.4. EXPERIMENTS

We compare our CHISEL indoor localization framework with its data augmentation and novel CAECNN architecture with state-of-the-art deep-learning-based indoor localization techniques SAEDNN [18] and 1-D CNN [19], as well as classical ML methods: KNN [13] and RF [14], all of which were discussed in Section 1.3.1. We show results for two variants of our CHISEL framework: CHISEL-DA, which uses data augmentation, and CHISEL, which does not, to

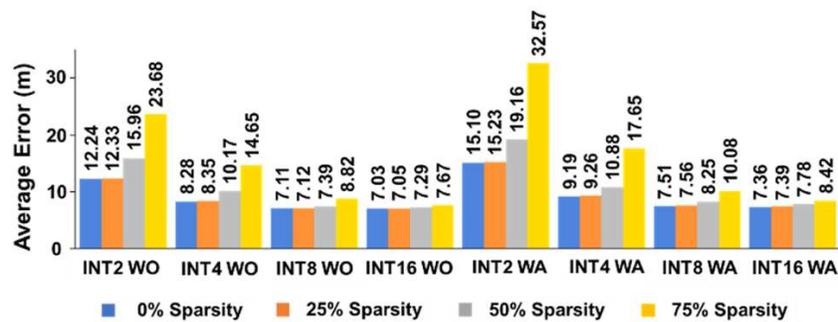
highlight the impact of data augmentation. We utilized the UJIIndoorLoc dataset for all experiments. We also deployed our localization model on the Samsung Galaxy S7 smartphone to assess prediction latency.

TABLE 2: Localization performance comparison

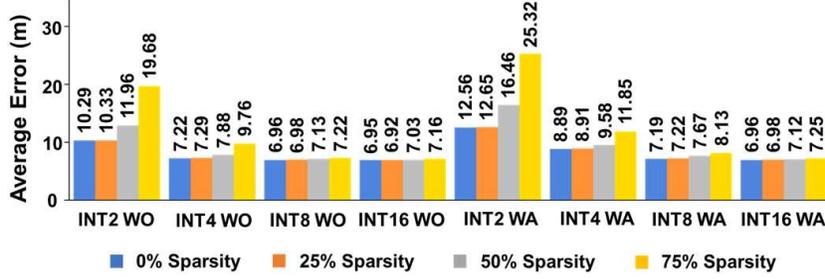
	<b>KNN</b>	<b>RF</b>	<b>SAE DNN</b>	<b>1D CNN</b>	<b>CHISEL</b>	<b>CHISEL-DA</b>
<b>Building (%)</b>	98.42	100	99.82	100	99.64	99.96
<b>Floor (%)</b>	90.05	91.2	91.27	94.68	91.72	93.87
<b>Position (m)</b>	9.82	7.85	9.29	11.78	8.80	6.95

### 2.4.1. EVALUATION ON UJIIndoorLoc DATASET

The comparison of building and floor accuracy is shown in Table 2. CHISEL has nearly 100% accuracy on building prediction and outperforms almost all other approaches on floor accuracy except for 1-D CNN which has three dedicated models for building, floor, and location, respectively. As shown in Table 2, the best average localization (position) errors of the proposed models are  $\approx 8.80$  m and  $\approx 6.95$  m, respectively, for CHISEL and CHISEL-DA. Based on our analysis of the results, we believe that 1-D CNN is unable to outperform other techniques in terms of positioning accuracy due to limitations inherent in its model architecture and the lack of data augmentation, as proposed in our work.



(a)



(b)

Fig. 3. Average CHISEL position error comparison across configurations with (a) post-training quantization and (b) quantization-aware training. Numbers on top of each bar represent the model’s average error (in meters). WO and WA, respectively, represent weights only and weights + activations quantization.

#### 2.4.2. EVALUATION WITH MODEL COMPRESSION

TABLE 3: Localization performance comparison

Config	Percentage Sparsity			
	0%	25%	50%	75%
INT2 WO	128KB	116KB	102KB	90KB
INT4 WO	173KB	148KB	124KB	101KB
INT8 WO	262KB	215KB	169KB	122KB
INT16 WO	442KB	350KB	259KB	165KB
INT2 WA	57KB	46KB	33KB	21KB
INT4 WA	107KB	92KB	68KB	45KB
INT8 WA	206KB	160KB	115KB	68KB
INT16 WA	407KB	315KB	222KB	130KB
FP32 NQ	801KB	620KB	440KB	259K

Given its better performance, we select CHISEL-DA as the baseline CHISEL model for further optimization. The uncompressed size of this CHISEL model is 801 kB and delivers an average localization accuracy of 6.95 m at a latency of 5.82ms on the Galaxy S7 device. To make the model more amenable to deployment on resource-limited embedded and IoT devices, we evaluate the impact of quantization and pruning on the accuracy, latency, and memory use of CHISEL. Fig. 3 presents the impact of various quantization and pruning configurations on the average localization error with CHISEL. The memory footprints of the CHISEL model configurations are given in Table 3 (FP32 NQ represents the baseline nonquantized model with 32-bit FP parameters).

Configurations suffixed with WO and WA, respectively, represent weights-only quantization and weights + activations quantization. We summarize some of the main observations as follows:

- 1) From Fig. 3(a), we observe that post-training quantization yields models with higher localization error in all cases as compared to QAT in Fig. 3(b). This motivates the use of QAT when deploying CHISEL.
- 2) As expected, an overall general trend is observed where using fewer bits to represent weight values leads to worsening accuracy. This is due to the lack of unique values available to represent weights and activations. At the extreme side, we observe that when CHISEL is about 1/17th its original size, in the INT2-WA-25% configuration (46 kB) [Fig. 3(b)], it makes localization error  $\approx 1.82\times$  larger than CHISEL-DA but is still competitive with 1-D CNN (Table 2).
- 3) Pruning from 0% (no pruning) to 25% has almost no impact on localization accuracy while reducing the model footprint by up to 25% as seen for the INT16-WA configurations in both Fig. 3(a) and (b). This is strongly suggestive of pruning’s positive impact toward deep-learning model deployment for indoor localization with CHISEL.
- 4) The impact of pruning becomes more pronounced when aggressively quantizing the model. This is especially true for the WA quantization as shown in Fig. 3(b). It is important to pay more attention to activations when quantizing CNN models with low bits, or aggressive quantization may result in huge accuracy reduction after compression.
- 5) Based on the results observed in Fig. 3 and Table 3, a good candidate for the compressed model is INT4-WO25% with QAT, resulting in a 148-kB memory footprint. This model is  $\approx 5.41\times$  smaller than the baseline CHISEL model and still better in terms of accuracy than

classical ML models [13, 14] as well as state-of-the-art deep learning models in prior works [18, 19].

Finally, to capture the impact of compression on localization latency, we deployed all of the compressed configurations of CHISEL on a Samsung Galaxy S7 smartphone using Android Studio’s on-device inference framework in v4.0.1. Our application is designed to directly receive RSSI from a file that contains all 1111 samples from the test set. RSSI values are processed into matrices in-app and fed to the CHISEL model. The captured latencies include the time required to preprocess the RSSI fingerprint into images and are averaged over 100 repetitions.

Inference time results of all compression configurations are presented in Fig. 4. We observe that both quantization and pruning can offer notable acceleration over the baseline FLOAT32 models. The INT4-WA-50% sparsity model cuts localization latency to half ( $\sim 2.8$ ms), while taking a penalty of 2.63 m (38%) in terms of positioning error. Aggressive quantization and pruning beyond this point yields limited benefits, e.g., INT2-WA + 75% sparsity only reduces the latency to  $\sim 2.25$ ms while degrading the localization accuracy by 3 $\times$ . INT4-WO-25% continues to present itself as a good candidate with a notable  $\sim 31\%$  reduction in latency. In summary, the intelligent data augmentation and novel CAECNN deep learning network model which is amenable to model compression allows our CHISEL framework to provide new options for high accuracy indoor localization while minimizing deployment costs (in terms of memory footprint and latency) on embedded and IoT devices.

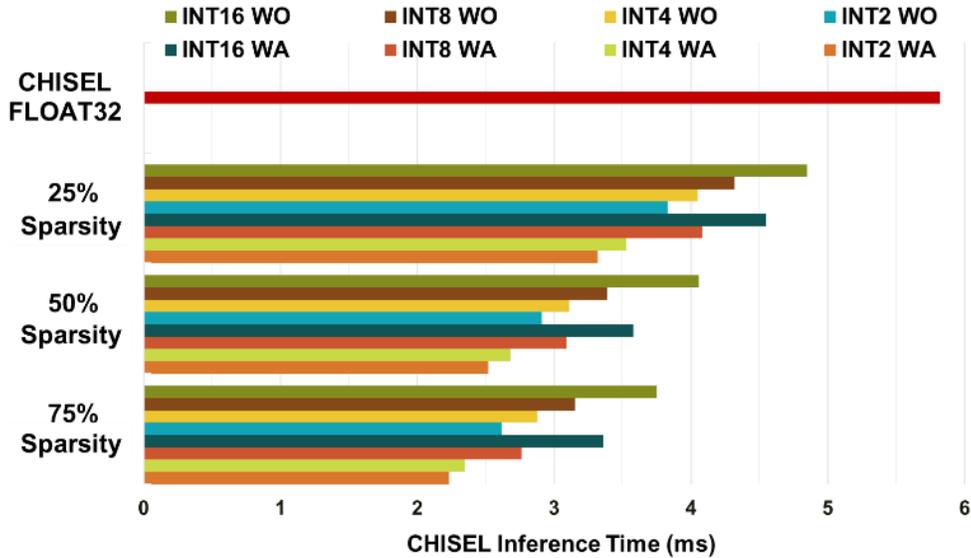


Fig. 4. On-device inference time from all compression configurations of CHISEL. CHISEL FLOAT32 represents the baseline model.

## 2.5. CONCLUSION

In this chapter, we presented a novel indoor localization framework called CHISEL. Our approach outperforms state-of-the-art ML and deep-learning-based localization frameworks, achieving higher localization accuracy. The compression friendly CAECNN models in CHISEL can maintain robust accuracy in the presence of aggressive model compression. Our compressed model versions are easy to deploy on smartphones and resource-constrained embedded and IoT devices that may have kB-sized resource budgets. Based on our experimental analysis, CHISEL is shown to provide a spectrum of deployment configurations with varying tradeoffs between accuracy, memory footprint, and latency goals. One of the more promising CHISEL configurations is the INT4-WO-25% with QAT, which reduces the model size to 148 kB (81.52% reduction) and reduces latency by 1.80ms (30.93% reduction) at the cost of sacrificing 0.34 m (4.89%) localization accuracy.

### 3. A FRAMEWORK FOR CSI-BASED INDOOR LOCALIZATION WITH 1D CONVOLUTIONAL NEURAL NETWORKS <sup>2</sup>

Modern indoor localization techniques are essential to overcome the weak GPS coverage in indoor environments. Recently, considerable progress has been made in Channel State Information (CSI) based indoor localization with signal fingerprints. However, CSI signal patterns can be complicated in the large and highly dynamic indoor spaces with complex interiors, thus a solution for solving this issue is urgently needed to expand the applications of CSI to a broader indoor space. In this work, we propose an end-to-end solution including data collection, pattern clustering, denoising, calibration and a lightweight one-dimensional convolutional neural network (1D CNN) model with CSI fingerprinting to tackle this problem. We have also created and plan to open source a CSI dataset with a large amount of data collected across complex indoor environments at Colorado State University. Experiments indicate that our approach achieves up to 68.5% improved performance (mean distance error) with minimal number of parameters, compared to the best-known deep machine learning and CSI-based indoor localization works.

#### 3.1. INTRODUCTION

The rapidly growing demands for intellectual and human centric indoor services have made indoor localization an important component of today's edge, mobile and Internet of Things (IoT) devices. Indoor navigation services in large buildings such as museums, libraries, and shopping malls have been boosting user experience by delivering responsive positioning functionalities. Stable and accurate indoor localization capabilities are particularly crucial in highly sensitive

---

<sup>2</sup> This chapter has been submitted as a research article (under review) in *12<sup>th</sup> International Conference on Indoor Positioning and Indoor Navigation* by L. Wang and S. Pasricha, 2022.

indoor positioning use cases, such as human activity recognition in hospitals, and robot tracking and position calibration in modern factories. In extreme cases, for instance, hazardous indoor spaces in the interiors of a nuclear power plant, especially when accidental leaks occur, precise indoor positioning is vital for ground robots to fulfill radiation detection and mitigation tasks [1]. These and other use cases have led to growing interest in accurate indoor localization technologies.

Popular indoor localization solutions broadly fall into two categories: geometric mapping and feature pattern mapping (also known as fingerprinting) [5]. The former first measures predefined parameters like power, distance, direction observations, etc., with respect to some reference points, followed by calculating locations using geometric conversion algorithms such as triangulation. In contrast, with fingerprinting, the aim is to find matched feature patterns and detailed conversion algorithms are often unnecessary. Specifically, after the feature space is defined, through the comparison between the feature pattern collected from an unknown location and a reference pattern space, the coordinates of an unknown location can be approximated. For geometric mapping, the measurements of directional and distance information required in geometric mapping-based techniques heavily count on the Line-Of-Sight (LOS) conditions which are usually hard to satisfy in complex indoor environments. In addition, non-negligible approximations existing in the conversion algorithms lead to inevitable accuracy drop. In contrast, fingerprinting is regarded as a better way to handle such challenges for complex indoor scenarios, as the pattern matching does not necessarily need to account for LOS conditions, or require conversion algorithms.

Traditional RSSI (Received Signal Strength Indicator) based fingerprinting methods can suffer from constant fluctuations caused by multipath and shadowing effects which can sway RSSI values by up to 5 dB [24]. Thus, RSSI based fingerprinting techniques may lead to lower accuracy during

indoor localization. Channel State Information (CSI) is considered an enhanced descriptor of wireless propagation and can improve the performance of current Wi-Fi location sensing technologies [25]. CSI data extracted from the physical layer (PHY) of fifth (and higher) generation Wi-Fi frames represents the frequency response of a Multiple Input and Multiple Output (MIMO) channel, and is capable of providing the sensitive parameters (e.g., magnitude attenuation and phase shifting) for capturing signal sources. The compensations of magnitudes and phases during signal transmissions corresponding to each subcarrier are available in captured CSI data, where uniquely different propagating paths can be spotted. In contrast to RSSI, even minor position changes of receiving antennas are able to be sensed through CSI analysis, thus there is a higher chance to achieve ultra-low localization errors.

While many promising recent efforts have demonstrated indoor localization with CSI data, several limitations remain. First, most state-of-the-art research with CSI data is restricted to localization within small and isolated rooms, where complex building interiors and dynamic environments, including human activities, seldom get considered. From our observations, CSI data received by a single antenna in a cuboidal shaped room has multipath and shadowing effects that are extremely limited and predictably concentrated. In more complex indoor environments, CSI patterns become more complicated and are composed of more diverse pattern groups due to the greater number of paths traversed by signals. We also found that the interior structures of a building contribute to significant modifications of CSI features during signal transmissions. Despite these structures determining the number of pattern groups, how to denoise each group remains an open question. We propose a novel approach for handling this phenomenon, which is described in Section IV. Second, prior works require knowledge of the Wi-Fi access points (Aps), for example, the number and the identity of antennas. The access to understand each AP could be limited due

to security regulations in practice, and it is not efficient or even possible to find every AP's specifications at the offline fingerprint data collection stage. Third, the accessibility of the dataset in most previous works is often not publicly available, thereby preventing other researchers from reproducing the results and improving the preceding research. Thus, a solution to enable an efficient, reproducible, and large-scale CSI-based indoor localization deployment without detailed AP knowledge is needed.

In the work, we propose a novel, end-to-end deep learning based framework for CSI-based indoor localization. The key contributions of this work can be summarized as follows:

- 1) We propose a methodology that requires minimal knowledge of Aps to preprocess CSI data in complex large open spaces, thereby removing the security concerns for accessing Wi-Fi facilities.
- 2) We propose a lightweight, one-dimensional Convolutional Neural Network (1D CNN) with two channels that takes CSI magnitude and phase data to classify locations.
- 3) We captured and utilized CSI data that has 180 RPs and 167 test points (TPs) containing 617000 and 16700 samples (CSI packets), respectively, from 6 Aps over a total path length of 183.296 meters. We plan to open source this dataset [37] to benefit the indoor localization community.
- 4) Experimental evaluations of our framework with three state-of-the-art CSI and deep learning based localization frameworks demonstrate the promise of our approach.

### 3.2. RELATED WORK

One of the earliest CSI fingerprinting based indoor localization studies with WLANs was conducted as part of the FIFS framework [26]. FIFS utilized the Bayes' theorem to find the maximum posteriori probability of a certain reference point (RP) given the knowledge of CSI fingerprints received from three APs at the online positioning stage and consequently, the estimated location is given by the coordinates of that RP. A key part of this work is applying spatial correlation of the CSI to determine the prior probabilities of RPs [26] for estimating unknown locations. However, the CSI from two locations separated by a small distance can be weakly unrelated in a large and complex indoor space, which makes FIFS unstable in this case. In addition, the distribution of test points (TPs) is not specified in this work which makes it difficult to replicate the performance of FIFS.

A few other efforts have explored CSI-based localization with deep learning models. DeepFi [27] uses a deep stacked autoencoder (SAE) to extract magnitude features of successive CSI packets from three transmitting antennas. In the offline stage, the SAE is individually pre-trained for each RP. CSI fingerprints from an unknown location are input to the pre-trained SAE, and the output is compared with the reconstructed CSI at RPs from the offline stage, thereby generating accumulated reconstruction errors for predicting the location. A notable drawback is the poor scalability caused by the number of SAEs needed, which grows with the number of RPs. A CNN classification model called CiFi is proposed in [38] where fingerprints are formed using angle of arrival (AoA) information to estimate locations, using phase difference information from two adjacent antennas based on 5GHz Wi-Fi. The authors mention how the estimation stability offered by AoA is better than magnitude values when LOS conditions are bad, which results in strongly attenuated magnitude data. However, CSI is essentially computed by modifications of the known

preamble content via wireless paths [10]. The compensations of the weakened magnitude values discard useful information for determining multipath propagation that is valuable for tracing back the location of the signal source. DelFin [28] also uses 5GHz Wi-Fi data and adapts a CNN model that uses the CSI magnitudes collected in a 5-room apartment as inputs. DelFin requires only one anchor transmitter for residential and small working spaces. Although this solution is lightweight and suitable for IoT devices, the indoor environment analyzed is quite small and simple, without considering dynamic interference. OpenCSI [29] is an open source project that introduces a solution for automating CSI collection in a  $3.5\text{m} \times 5\text{m}$ . A radio map is built using a software-defined radio (SDR) on a wheeled robot as the collector to extract CSI from Long-Term Evolution (LTE) eNodeB. A CNN model is utilized with fusion of magnitude and phase information. The dataset is publicly made available, but the small space considered prevents considerations of dynamic and complex interference effects. Moreover, SDRs usually need dedicated infrastructure and the cost is also arguably high, for example, the USRP B200mini used in the project is priced at more than 1,000 USD. In SDR-Fi [30], a feed-forward neural network (FFNN) and 1D CNN models are built to utilize CSI magnitude for location estimation in an approximately 60 square meter space. However, CSI phase information, which has better anti-noise capabilities compared with magnitude, is not considered. In addition, details of all model layers and dataset collected are not provided, preventing comparative analysis.

Unfortunately, there are several other factors that can hinder the practical use of the aforementioned works. First, the sizes of experimental spaces are limited. The largest area considered is a  $32.5\text{m} \times 10\text{m}$  corridor in [26]. For larger and more complex indoor spaces, CSI patterns become more diverse and complex, and such patterns are not considered in these works. Second, experiments in these prior works are conducted in relatively static indoor environments.

Dynamic factors such as human activities and complex electromagnetic interference are not well considered which is problematic in terms of the robustness and feasibility of real-world deployment. Third, knowledge of Wi-Fi APs in practice might not be obtainable whenever facility security concerns matter. Lastly, it is also worth considering deep machine learning approaches that use less complicated models to reduce inference time for real-time localization on mobile and IoT devices.

The next section (Section III) describes our CSI data collection effort in large indoor environments. Section IV describes our data denoising and calibration approach. Section V provides an overview of our deep learning model for indoor localization. Section VI presents experimental results. Finally, Section VII concludes this work and also discusses some related open problems.

### 3.3. DATA COLLECTION

#### 3.3.1. BUILDING AND PATH INFORMATION

We collected CSI data from the 802.11ac APs on the second floor of the Colorado State University (CSU) Behavioral Sciences Building (BSB). The gross area for data collection is 3608 square meters including 5 paths covered by 6 APs. The floorplan for data collection is shown in Fig. 5.

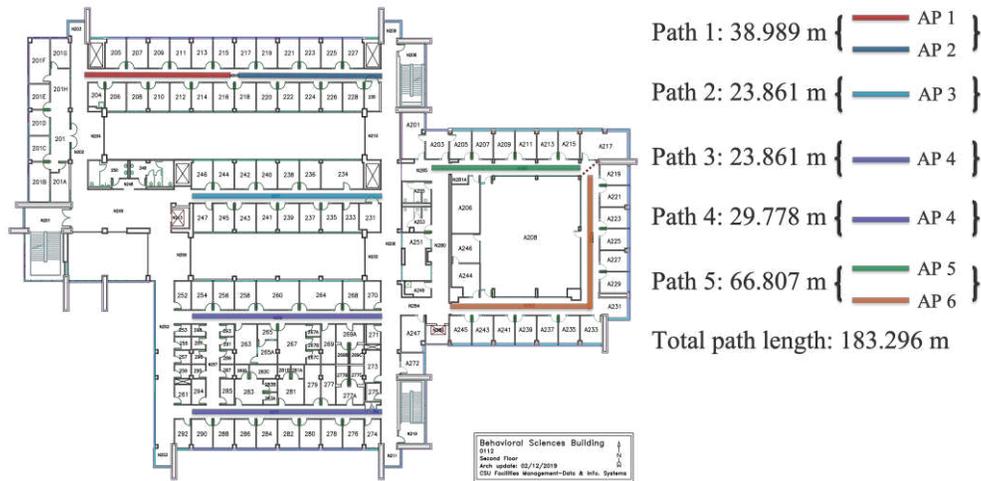


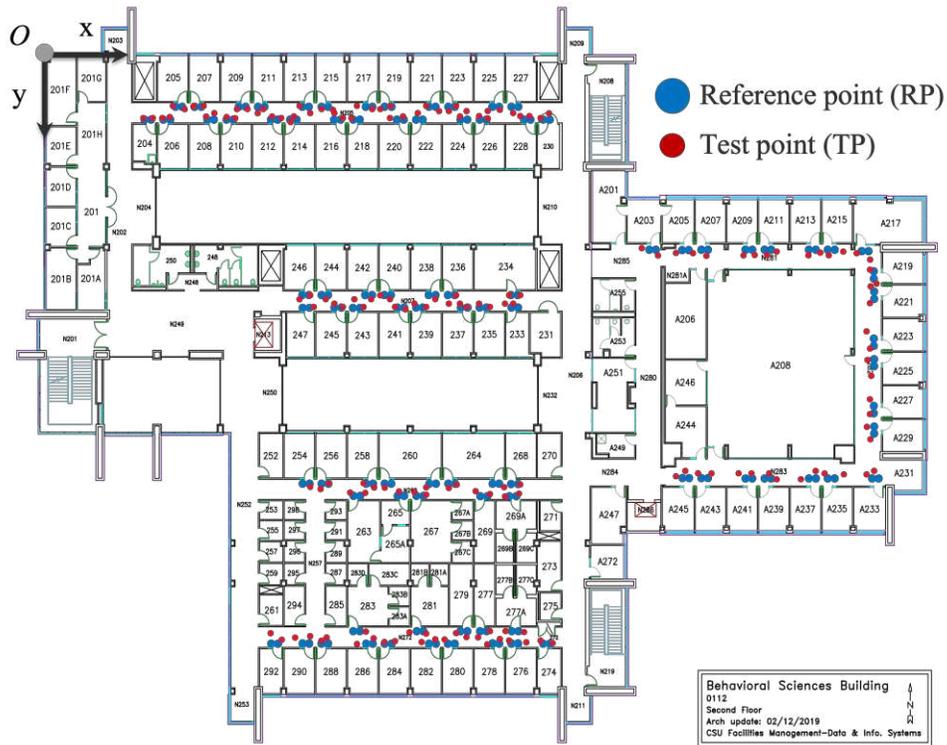
Fig. 5. The floorplan of 2<sup>nd</sup> floor of Colorado State University (CSU) Behavioral Sciences Building (BSB). Different colors represent corresponding AP visibility; each path may contain 1 or 2 APs based on received signal strength.

The data was collected during CSU operational days from Monday to Friday in a week. Collection time windows were fixed from 9am to 12 am and 2pm to 10 pm, Mountain Time (MT). The whole data collection was done over a span of two months. During the collection process, the dynamic effects and noise arising from human activities (due to students, staff, visitors, and faculty in the building) were captured and included. CSU BSB is a large modern building providing classrooms, recreation areas, labs, study rooms, conference rooms and offices where possible interference is also present during the collection windows due to the rich variety of electronic devices such as printers, plotters, computers, floor scrubbers, projectors, personal electronic devices, environmental control systems, various wireless equipment, etc.

### 3.3.2. DISTRIBUTIONS OF REFERENCE POINTS AND TEST POINTS

Fig. 6 shows the distributions of RPs and TPs where the CSI packets were collected for the building reference set and test set. To make this work practically meaningful, RPs and TPs are selected based on RSSI quality using Wi-Fi MAC scanning since the modern mobile clients can

automatically select the APs with highest RSSI quality. The reference set can be further divided into training and validation set for machine learning algorithms. Every room door has two RPs that are aligned with the center of the door and the room tag of that room, respectively. TPs are selected based on our observation that if a TP is far (i.e., more than one meter) from its nearest RP, the relationship between the corresponding fingerprint of the RP and the TP tends to be weakly related, in the context of a complex indoor environment like that in the CSU BSB. This at least indicates two aspects of CSI. First, CSI offers impressively more sensitive information than RSSI with small location changes. Second, enough number of RPs with fine-grained distribution can potentially improve the accuracy of CSI-based indoor localization systems.



### 3.3.3. DATA COLLECTION PLATFORM

We used Nexmon [10] to extract CSI data from the PHY layer of 802.11ac symbols. Nexmon is a C-based firmware patching framework with currently the broadest support for various Broadcom and Cypress Wi-Fi chipsets. It was developed by the Secure Mobile Networking Lab (SEEMOO) with its recent version supporting 20/40 and 20/40/80 MHz per frame CSI extraction on 802.11n/ac, respectively. The data collection platform is easy and affordable to build. The hardware and software settings and peripherals are listed in Table 4. The communication from the laptop to the Raspberry Pi is via an Ethernet cable where the commands are sent from the laptop depending on how the user would collect CSI packets. There is no WLAN connection between the laptop and Wi-Fi APs during the data collection. The Raspberry Pi is the only client to receive CSI packets on this platform. The height from the Raspberry Pi to the ground is fixed at 120 cm. This mean height is at an average adult’s chest level, chosen based on the report from Centers for Disease Control and Prevention’s (CDC) national health statistics report [39]. Here, we assume most users habitually looking at their smartphones at the height of their chest when using indoor localization services.

TABLE 4: CSI collection platform components

Hardware	Software
<ul style="list-style-type: none"> <li>○ Laptop</li> <li>○ Power bank</li> <li>○ Raspberry Pi 4 Model B</li> <li>○ Ethernet cable</li> <li>○ USB Type C to USB-A 2.0 charger cable</li> <li>○ Laptop tripod</li> </ul>	<ul style="list-style-type: none"> <li>○ Linux 16.04 LTS</li> <li>○ Raspberry Pi OS v4.19</li> <li>○ Nexmon CSI extractor</li> <li>○ Nexmon RSSI patch</li> <li>○ Wi-Fi firmware version 7_45_189 for Broadcom 43455c0 chipset</li> </ul>

### 3.3.4. 802.11ac 20MHz SYMBOL STRUCTURE

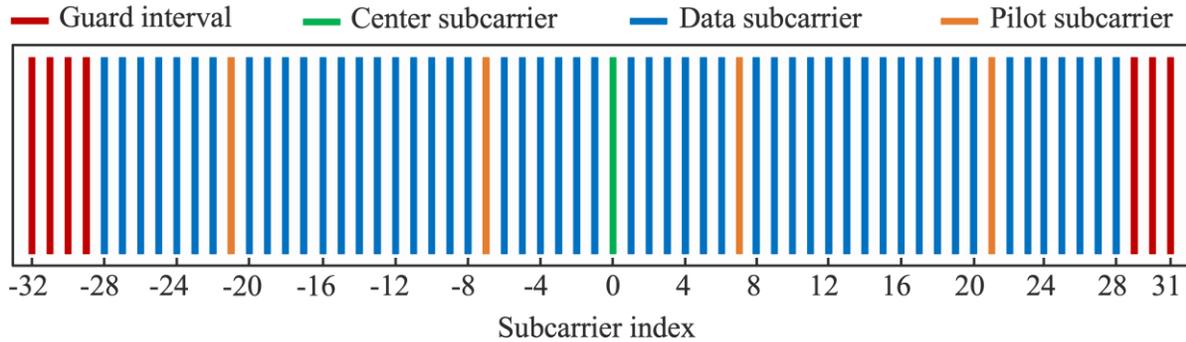


Fig. 7. A 64-subcarrier 802.11ac 20 MHz symbol

The 802.11ac 20MHz symbol extracted from the Nexmon CSI extractor is shown in Fig. 7. The 64 subcarriers within a single symbol includes 7 guard intervals, 52 data subcarriers, 4 pilot carriers and 1 center/null/direct current (DC) subcarrier. Guard intervals are intended to offer each segment of the preamble a cyclic delay to avoid interference [40]. Pilot carriers provide the wireless channel measurements with constellation points in data transmissions. The channel compensations from captured CSI packets are obtained based on the contents of the preamble and pilot subcarriers. The data subcarrier is the medium to carry user data and the center subcarrier is provided in 802.11ac to resist DC offset during analog/digital conversion and suppress carrier feedthrough. The values of guard intervals are constant and not useful for capturing information relevant for indoor localization, while the values of pilot and data subcarriers change based on the signal propagation paths, hence are the components we need from CSI data. The value of the center subcarrier is not used in our work and set to zero. Consequently, the magnitude and phase data from 57 subcarriers are extracted for characterizing fingerprints in this work, including 1 zero-valued center subcarrier, 4 pilot subcarriers and 52 data subcarriers.

### 3.3.5. DATA COLLECTION STRATEGY

One major concern with indoor localization using fingerprinting is if the reference space covers enough fingerprint patterns in the area of interest and how to find these patterns. It is not easy to find the CSI patterns in a complex free space (such as in the CSU BSB environment) if there is no AP knowledge available. For example, the magnitude of 3500 successively received CSI packets at location “204\_1” (with 21 abnormal packets that contain extreme values removed) is shown in Fig. 8.

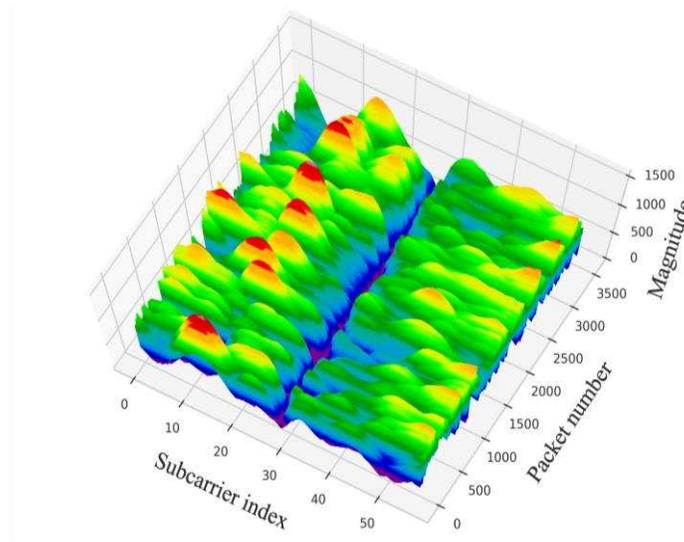


Fig. 8. Magnitude distribution of 3500 received CSI packets at the first RP of room 204 with 21 abnormal packets removed (3479 packets are visualized).

The magnitude computation and abnormal packet removal are explained in section 3.2.2. At first glance, from Fig. 8, there are no obvious clues showing pattern information in the randomly received packets. The challenge here is to find appropriate received signal patterns and determine the number of packets needed for the patterns at a certain location, to form fingerprints. In previous works (discussed in Section 1.3.2.), if the hardware and/or software stack of an AP is accessible and modifiable, at least two things are feasible to do: (1) one can inject frames over that AP to the

receiver at a certain time window to avoid signal interference; (2) by streaming packets on different antennas one by one, with different antenna angles, the CSI patterns from each antenna are easy to find. These two aspects make receiving CSI patterns much more predictable compared to our situation where the lack of AP accessibility (e.g., due to security reasons) can be a bottleneck in the practical case of the APs located in the CSU BSB.

The key question now becomes: is there a way to identify the unique signal patterns received at a specific location without any physical access to AP deployment details? Since the arrival time of each CSI packet is highly random in our case, we hypothesize that all CSI patterns from different wireless signal paths can be determined and reproducible when the reception of packets is sufficient during an uninterrupted time window (typically over ~6 to 7 minutes per RP). To verify this hypothesis, we conducted a number of experiments and then proposed a methodology based on Monte-Carlo observations to explore the CSI patterns for complex indoor environments. In the first step, we gradually increased the number of received packets at each RP location and by performing a K-means clustering [41], the number of the most dominant CSI patterns were determined. In the second step, when the number of CSI patterns becomes stable and does not increase with the number of received packets within a 200-packet receiving window, the total number of packets for characterizing this location is determined. In the third step, when the number of CSI packets is determined (to include all patterns at a location from step 2), we directly set that number as the starting number for receiving the packets for the next location. This significantly reduces the complexity of the process for finding the number of patterns for each RP. We found that 2800 packets per RP was a reasonable number to capture all patterns for all RP locations in Fig. 6. To avoid random errors, extra packets are captured for each location. Thus, we finally set 3500 as the number of packets for each RP, except for the RPs covered by AP2 which have 3000

packets from each RP, due to the frequently interrupted connection after receiving 3000 packets. Note that in the online phase, our framework only requires 1 packet (sample) at a single TP to predict the location but 100 packets per TP are collected for alleviating random errors. The next section describes the CSI data pattern denoising and clustering process that is utilized as part of our framework.

### 3.4. CSI DATA CLUSTERING, DENOISING AND CALIBRATION

In this section, we propose a novel data preprocessing procedure including clustering, pattern-wise denoising, and RSSI based CSI calibration. The 3500 packets received at the location 204\_1 (Fig. 8) are used as an exemplar to describe the clustering and denoising methodologies.

#### 3.4.1. CSI DATA INTRODUCTION

According to [42], the meaning of CSI is demonstrated in equation (1) and (2), where  $X_{sc}$  is the transmitted signal and  $Y_{sc}$  is the received signal on a certain subcarrier;  $\varphi_T$  and  $\varphi_R$  is the transmitted phase and received phase, respectively. Thus, CSI is a measurement of signal magnitude attenuation  $|H_{sc}|$  and phase shifting  $\Delta\varphi$ , which describes a certain wireless propagation path that the transmitted signal experienced.

In our framework, CSI magnitude and phase data are extracted to train a neural network (details in Section 3.3) for each AP. The extracted  $CSI_{ex}$  from Nexmon CSI extractor is a complex number composed of a real part value and an imaginary part value, as shown in equation (3). Magnitude and phase values are computed by equation (4) and (5), respectively.

$$|Y_{sc}|e^{j\varphi_R} = CSI \cdot |X_{sc}|e^{j\varphi_T} \quad (1)$$

$$CSI = |H_{sc}|e^{j\Delta\varphi} = \frac{|Y_{sc}|}{|X_{sc}|} \cdot e^{j(\varphi_R - \varphi_T)} \quad (2)$$

$$CSI_{ex} = Re(CSI) + jIm(CSI) \quad (3)$$

$$CSI_{mag} = \sqrt{Re^2(CSI) + Im^2(CSI)} \quad (4)$$

$$CSI_{pha} = \angle \tan^{-1} \frac{Im(CSI)}{Re(CSI)} \quad (5)$$

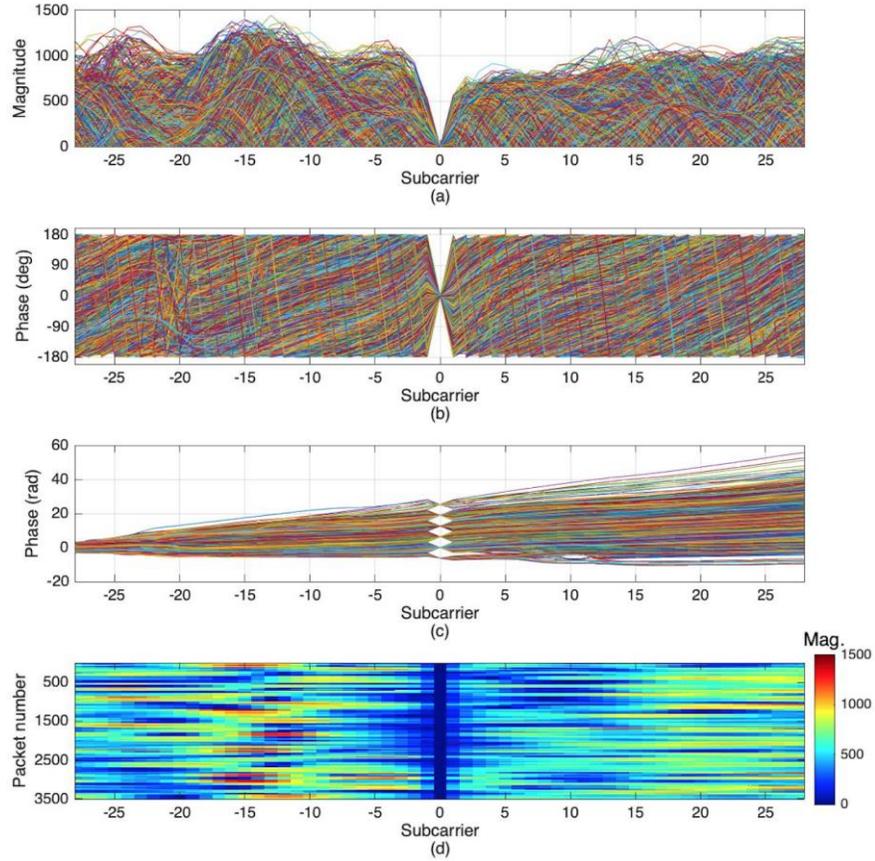


Fig. 9. Extracted CSI magnitude and phase from 3479 packets at location 204\_1 with 21 abnormal packets removed. (a) Magnitude; (b) Phase in degree; (c) Unwrapped phase in radius; (d) Magnitude spectrum corresponding to Fig. 6.

The extracted CSI magnitude and phase data are visualized in Fig. 9. Note that the phase data cannot be directly used due to the discontinuities introduced by the  $\tan^{-1}(\cdot)$  function (Fig. 9 (b)). To solve this issue, we apply a phase unwrapping technique. If the difference between two

consecutive subcarriers is equal or larger than  $\pi$ , the following operation is performed on the phase data along the subcarrier axis:

$$Pha(subcarrier_{i+1}) = Pha(subcarrier_i) \pm 2\pi \quad (6)$$

where  $i$  represents the  $i$ th subcarrier. If the adjacent difference is less than  $\pi$ , the original phase value of the  $i$ th subcarrier is maintained. + or - depends on whether the difference is larger than  $\pi$  or less than  $-\pi$ . The unwrapped phase data (which carries phase shift information inside each packet received from different wireless paths) is converted into radius as shown in the Fig. 9(c).

### 3.4.2. PATTERN CLUSTERING AND DENOISING

We devised a data preprocessing methodology that includes pattern clustering and denoising (discussed in this subsection) and RSSI-based CSI data calibration (discussed in the next subsection).

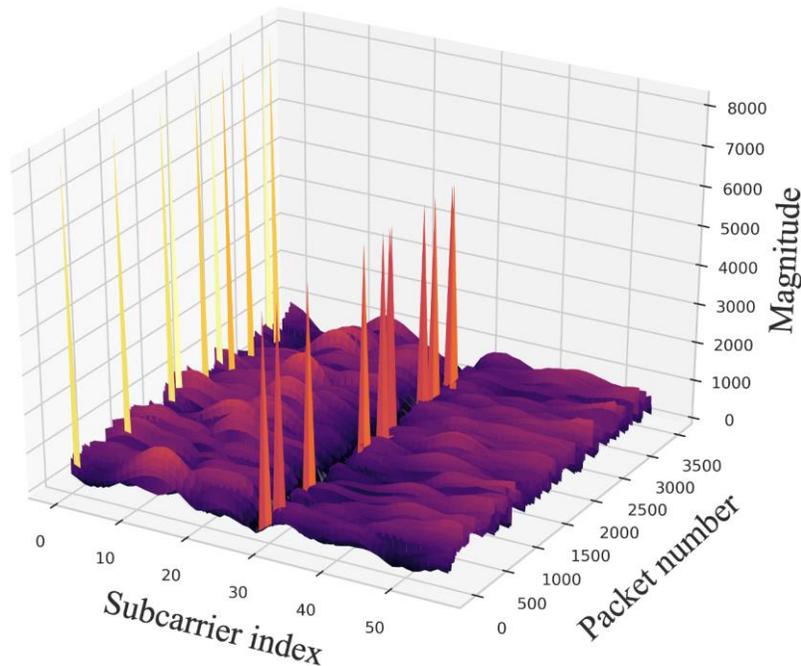


Fig. 10. Extreme magnitude values (spikes) exist in CSI packets.

First, the packets that have abnormal values are removed. The raw CSI data consists of extreme values that randomly appear like spikes on some subcarriers of certain packets. The magnitude of peak values is often larger than 2000 as shown in Fig. 10. Thus, 2000 is set as the threshold to remove abnormal packets for all RPs. For example, at the location 204\_1, 21 out of the total 3500 received packets exceeded this threshold and were removed. Second, a pattern-wise denoising algorithm is utilized based on K-means clustering results. The number of clusters (i.e., number of the most dominant patterns) is determined by computing the average silhouette score (SS) based on the results of K-means. SS is obtained by the following formula:

$$SS = \frac{1}{N} \left[ \sum_{i=1}^N (D_{inter_N} - D_{intra_i}) / \text{MAX}(D_{intra_i}, D_{inter_N}) \right] \quad (5)$$

where  $D_{intra_i}$  is the mean distance between each signal within a cluster and  $D_{inter_N}$  is the mean distance between all clusters.  $i$  represents the  $i$ th cluster and  $N$  ( $N \geq 2$ ) is the number of clusters to evaluate based on the results of the chosen clustering method. If SS is close to 0, either the clustering algorithm does not work well or there are no distinct differences to isolate the data of interest. The number of clusters that gives the highest SS will be the number to guide our K-means algorithm.

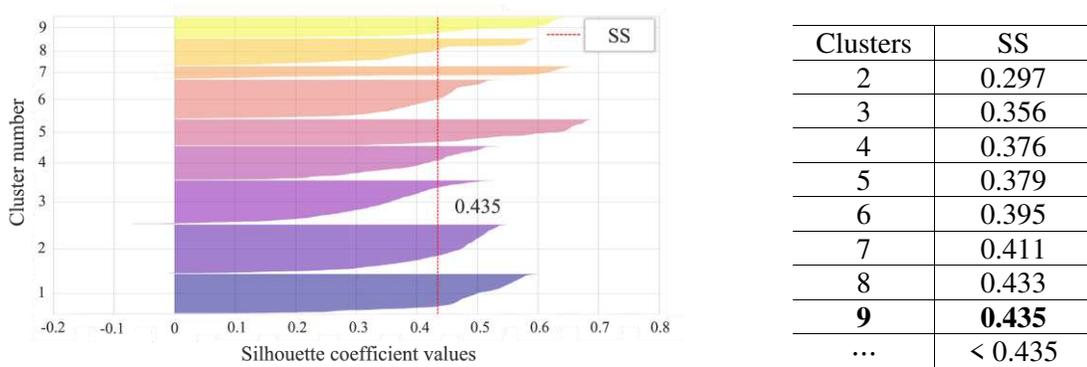


Fig. 11. The average silhouette score given by 9 clusters

As an example, Fig. 11 shows that 9 clusters give the highest SS which is 0.435 for RP location 204\_1. Note that different RPs can have different clusters, based on this SS analysis. The terms cluster and dominant pattern will be used interchangeably in the following sections (e.g., cluster 1 represents dominant pattern 1 and so on).

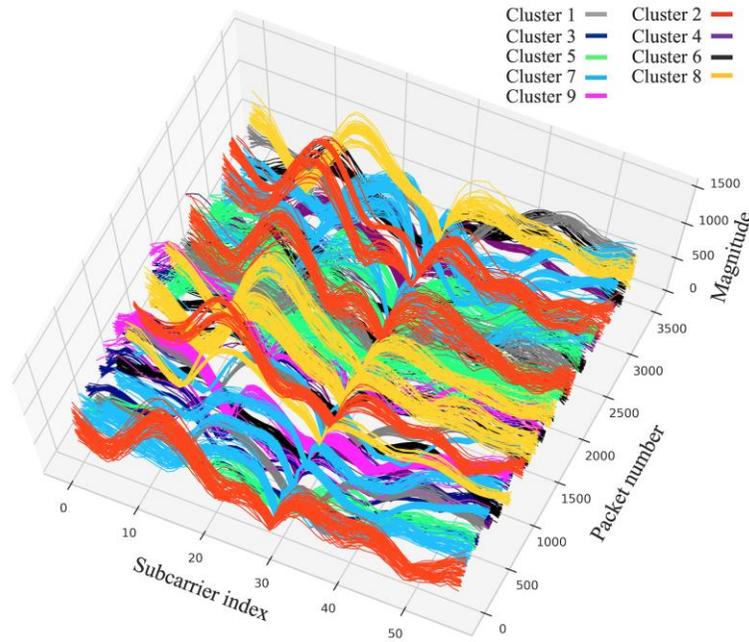


Fig. 12. The 9 magnitude patterns of 3479 CSI packets at RP location 204\_1.

Third, K-means is performed with the cluster count of 9 after SS testing. The clusters from magnitude data are shown in Fig. 12. The packet indices of each cluster can be directly mapped to the indices for clustering phase data, thereby consistency between magnitude and phase can be achieved. The clustered phase patterns are shown in Fig. 13. The clustering step here is synced with the data collection process and hence there is no extra work needed since the clustering results are also used to determine the number of packets to be collected for each RP, as discussed in Section 3.1.5. Note that we also explored creating clusters starting from phase data, but found that

phase data was more stable across RPs than magnitude data, and led to the creation of fewer clusters. Thus, phase data was less effective in creating unique fingerprints for RPs than magnitude data, which is why we selected magnitude data to create clusters. However, considering both magnitude and phase patterns was crucial to achieving higher performance, which is why we use both as inputs to our deep learning model (see Section 3.3.).

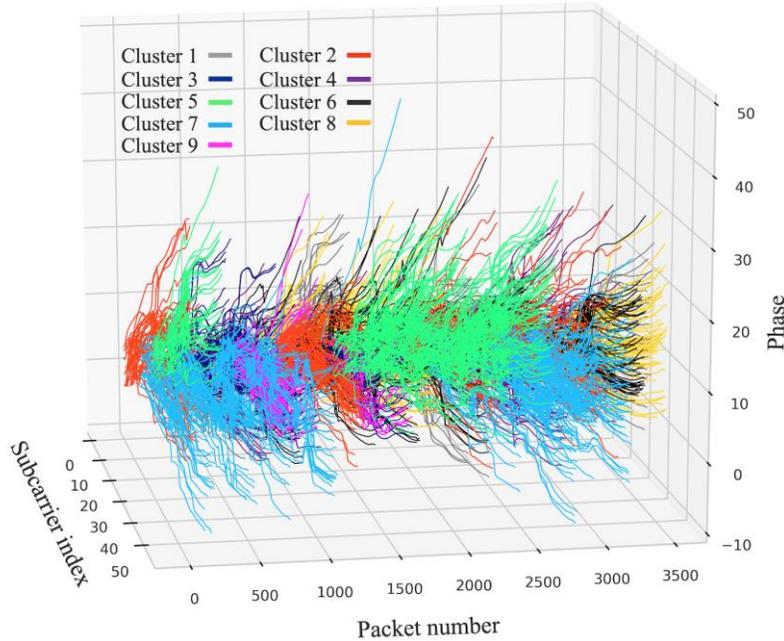


Fig. 13. The 9 phase patterns of 3479 CSI packets at RP location 204\_1.

Fourth, after finding the dominant patterns (clusters) of received CSI packets, we found the presence of dynamic noise (e.g., small-scale fading) inside each pattern. Based on our experiments, the noise tends to obfuscate the identity of RPs, thus we propose a three-stage pattern-wise denoising method, as follows:

- 1) Compute the mean value of each subcarrier in each pattern to determine a mean CSI sequence that represents the main feature of this pattern that the most packets contribute to.

- 2) Compute the correlation coefficient (CC) between the CSI packets and the sequence obtained in 1) and remove the packets that have CC values  $< \psi$  (CC filtering).
- 3) Compute the Root Mean Squared Error (RMSE) between the remaining packets after CC filtering in 2) and the sequence obtained in 1) and remove the packets that have RMSE  $> \chi$  (RMSE filtering).

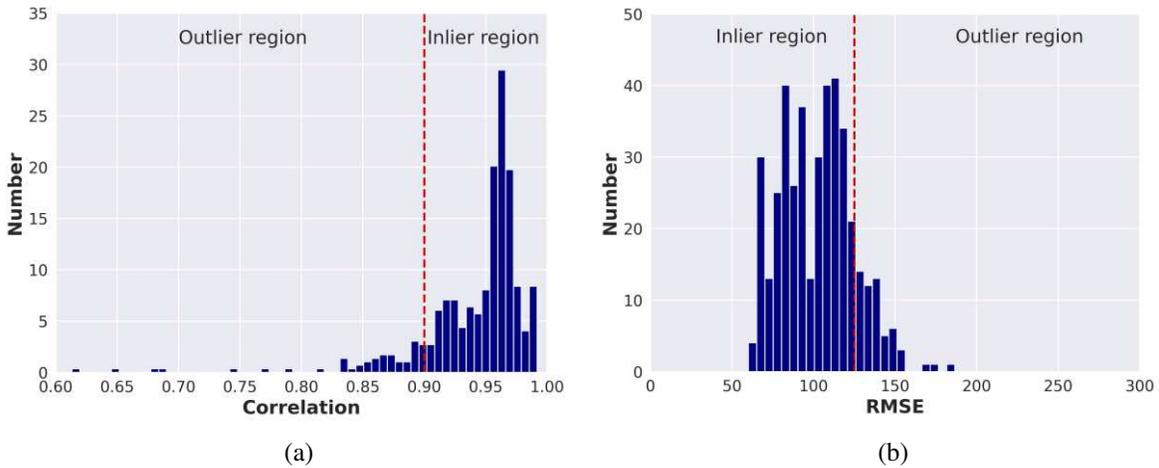


Fig. 14. (a) CC score histogram and (b) RMSE distance histogram, for pattern 2 at the location 204\_1 w.r.t the mean sequence.

We empirically set the value of threshold  $\psi$  to 0.9. The optimal value of threshold  $\chi$  can vary from pattern to pattern inside each RP. For simplicity, we set threshold  $\chi$  to 125 globally which provided good performance. The goal of the 3-stage denoising is to first filter out the most unrelated noisy packets and then the packets that are far from the mean sequence. Fig. 14 shows an example of these thresholds for cluster 2, where packets having the CC scores  $< 0.9$  with respect to the mean sequence, are removed (Fig. 14(a)) and packets with RMSE  $> 125$  (Fig. 14(b)) are also removed. Fig. 15 shows the result of denoising for cluster 2. The indices of removed packets are recorded to drop the corresponding phase packets directly.

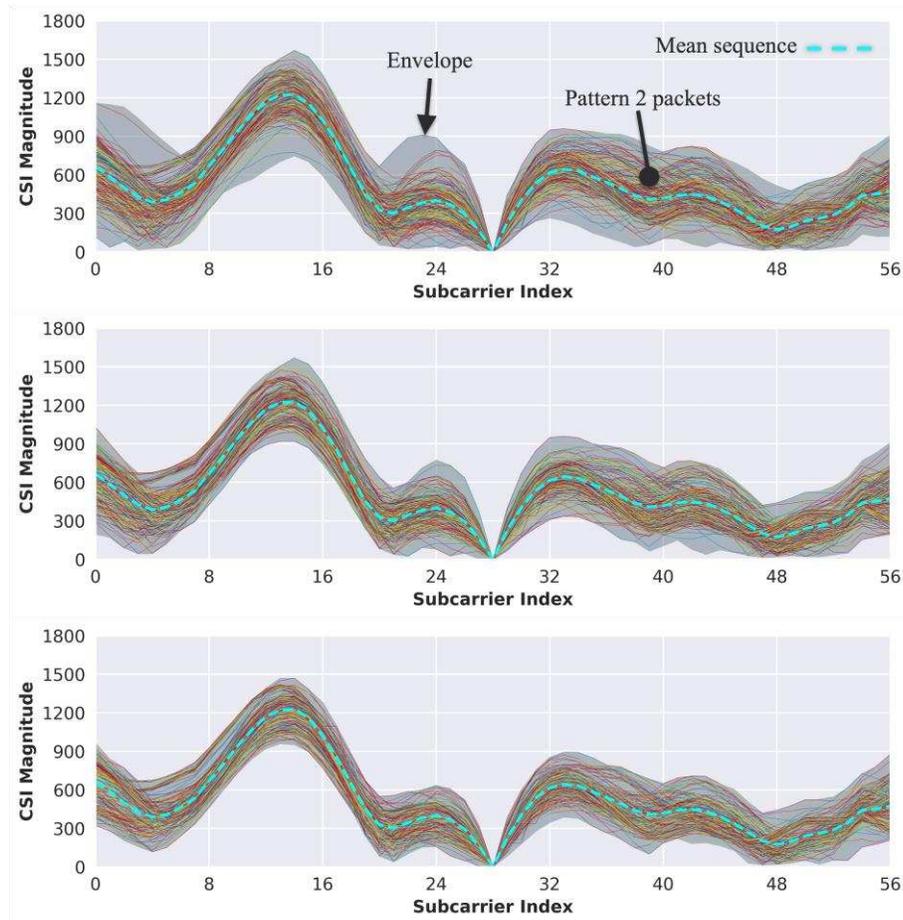


Fig. 15. The denoising process with the pattern-wise denoising algorithm. Top: the original pattern 2 packets; middle: pattern 2 packets after CC filtering; bottom: pattern 2 packets after RMSE filtering.

### 3.4.3. RSSI BASED CSI CALIBRATION

Any collected CSI information is universally filtered by Automatic Gain Control (AGC) onboard modern Wi-Fi hardware whose purpose is to make the magnitudes of received signals from different propagating paths dynamically stable. For example, the most faded magnitude gets the biggest gain after AGC and vice versa. Properly dealing with AGC effects is still a challenge for CSI based indoor localization systems, due to the undocumented underlying algorithms deployed by manufacturers and the different CSI extraction approaches [43]. As a result, the magnitude values extracted from CSI packets lose the identity of the transmitter location in terms

of distance which is a critical factor in realizing unique fingerprints for RPs and, consequently, localization accuracy. Since the distance information is obfuscated by AGC, one can integrate RSSI-derived distance information with CSI data for improving the performance of indoor localization systems. This is because RSSI is obtained before AGC while CSI is obtained after AGC. However, the 802.11 standard recommends an RSSI range between 0 to 255 but does not specify how RSSI needs to be calculated, thus vendors can have their own definitions to compute RSSI. In this work, we use a dBm equation to derive an RSSI-based scale factor to calibrate magnitude from CSI packets. We assume AGC is a linear time invariant system [44] that scales all magnitude values with a same scale factor once the RSSI is invariant during a certain time window. We propose a simple calibration equation:

$$RSSI = 10 \log_{10} \left( \frac{P}{1mW} \right) \quad (6)$$

$$\Rightarrow S = \sqrt{10^{\frac{RSSI}{10}}} \quad (7)$$

where  $P$  is the received signal power in milliwatt and  $S$  represents the scale factor to multiply the real and imaginary parts in equation (1). The  $\sqrt{\cdot}$  is a hardcoded conversion for converting power to voltage or current. There is no need to rescale phase data as the AGC effect is factored out with the division operation. In other words, phase data from CSI packets is theoretically more reliable in terms of ground truth, however, we focus on fusing magnitude and phase data for feeding into a deep learning model in this work. RSSI can be extracted per packet with the recent Nexmon RSSI patch for bcm 43455c0 chipset. For simplicity, we use an identical RSSI value for all packets received at a certain location. For example, 3500 packets received at the locations 204\_1 have 3500 RSSI values but only one is picked to scale the magnitude values. To alleviate the effect from

outlier values in the extracted RSSI, the median of 3500 RSSI values is chosen rather than the mean of them. In the online phase, captured CSI packets from TPs are preprocessed with the same procedure as that used for RPs in the offline phase, except for pattern-wise denoising.

### 3.5. 1D CNN-BASED INDOOR LOCALIZATION

#### 3.5.1. TRAINING SET AND TESTING SET

After the clustering, denoising, and calibration, the CSI packets are used to train a neural network. For the CSU BSB indoor environment, the number of packets (samples) used in this stage, for each AP, are listed in Table 5. The samples for training from each AP are further divided into training set and validation set by the ratio of 9:1. As mentioned earlier, in the test phase, 100 samples are collected for each TP to avoid random errors. We randomly select one sample from the 100 samples for each TP, for a single test and drop it from the original set and repeat this process 10 times to compute the average distance error over 10 tests per TP. In total, 476028 and 52892 samples are used to build the training set and validation set, respectively, with the corresponding number of packets used for each AP shown in the Table 5. 1670 test samples are randomly chosen from the collected 16700 CSI samples at 167 TPs with 10 samples per TP.

TABLE 5: Samples for each RP after clustering/denoising/calibration

	AP1	AP2	AP3	AP4	AP5	AP6
Pkt.	77,558	67,262	89,190	168,738	42,616	83,556

### 3.5.2. NETWORK ARCHITECTURE

Our proposed one-dimensional (1D) CNN architecture is illustrated in the Fig. 16. The reason for selecting a 1D CNN is because 1D CNNs are not only capable of extracting features from sequence-like data, such as magnitude and phase data inside CSI packets, but can also deliver a lightweight deep neural network architecture capable of fast inferencing and low-energy consumption requirements, making it attractive for mobile devices with resource constraints.

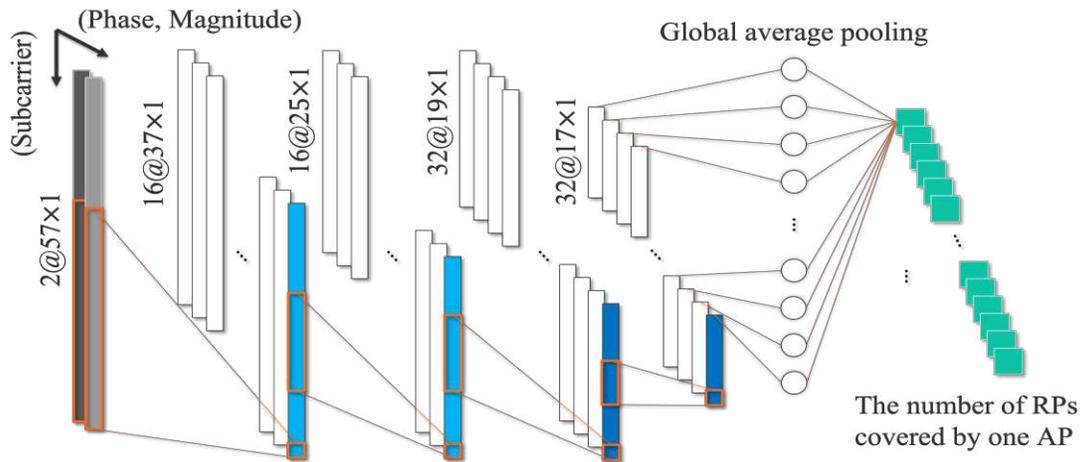


Fig. 16. The illustration of 1D CNN for location classification

The input of this 1D CNN is formed with a two-channel sequence consisting of magnitude and phase data for each channel, respectively. Details of the network architecture are shown in Table 5. Each of the first two convolutional layers have 16 feature maps followed by the two convolutional layers that both have 32 feature maps. No pooling layers are involved, and all convolutional outputs are not zero-padded. We use a global average pooling layer [45] to expand the final convolutional layer. The shape of the final output layer is determined by the number of RPs covered by an AP. In the output layer, the unknown location's fingerprint is approximated to

(i.e., classified as) one of the RPs within a single AP with the highest probability. Each convolutional layer is followed by a dropout layer to alleviate overfitting. The activation function is “ReLU” for each convolutional layer. Categorical cross entropy function is adopted for backpropagating classification errors during training. The number of neurons in the output layer is denoted by  $N$  which is determined by the number of RPs belonging to a single AP. The value of  $N$  varies from 16 (for AP5) to 56 (for AP4).

### 3.6. EXPERIMENTAL RESULTS

We compare our model against three recent deep learning- frameworks that utilize CSI fingerprinting: CiFi [38], DelFin [28], and OpenCSI [29]. DelFin and OpenCSI are regression-based 2D CNN and 1D CNN models, respectively, with one output node for estimating the horizontal coordinate and another for the vertical coordinate of an unknown location. CiFi is a 2D CNN classification-based network with the number of output nodes being equal to the number of RPs. We compare these frameworks to three variants of our framework: a baseline variant that includes all the stages described in Sections IV and V (CSILoc), a variant that does not include the denoising preprocessing discussed in Section IV (CSILoc-NoDen), and a variant that has the same preprocessing stages as CSILoc but trained with only CSI magnitude data, without considering CSI phase data (CSILoc-NoPh), similar to the approach used in SDR-Fi [30].

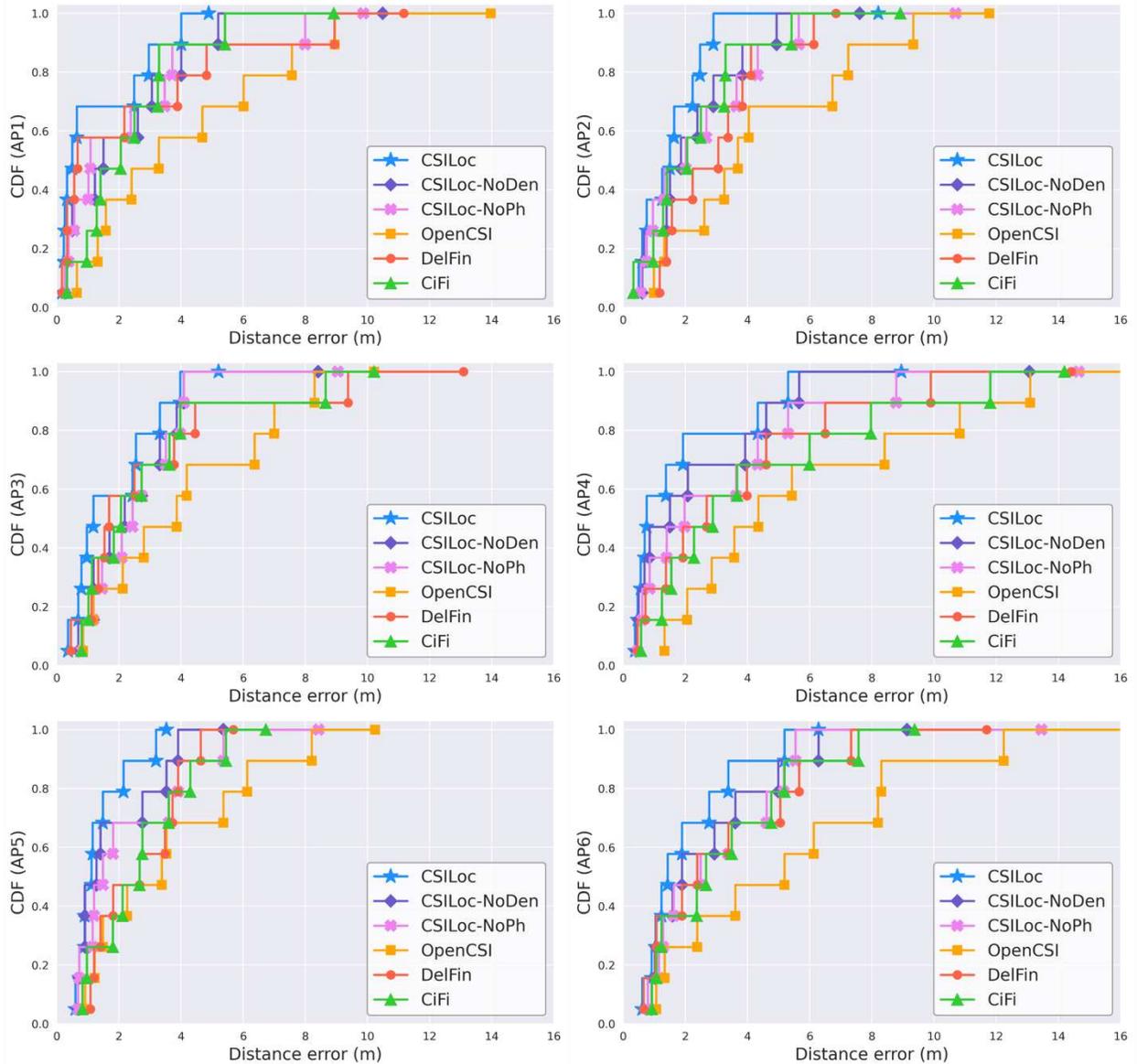


Fig. 17. Comparison of the localization errors based on CDF for each AP

Fig. 17 shows the indoor localization performance based on Cumulative Distribution Function (CDF) across the APs in the CSU BSB environment. From the results, our CSILoc framework shows the best distance accuracy performance across the 6 APs. The CSILoc-NoPh variant, which ignores phase information and only considers magnitude information performs the worst out of all the CSILoc variants, highlighting the importance of considering both CSI phase

and magnitude information for localization. The CSILoc-NoDen variant outperforms the models from prior studies, although it has slightly worse performance than CSILoc for all APs. The superior results with CSILoc compared to CSILoc-NoDen highlights the importance of the preprocessing performed in our framework. CiFi has comparable performance with DelFin in AP1, AP3 and AP6, but is less accurate in other APs. Our analysis indicates that OpenCSI suffers from a severe model overfitting phenomenon since it is originally devised for LTE signals which have more subcarriers (higher resolution), and thus a larger number of feature maps are applied in the convolutional layers for more powerful feature extraction. As one can expect, the worst distance error results occur under AP6 for all models because the TPs within AP6 have a much larger average distance to the RPs of AP6. Compared with the results from other APs, the classification result for AP6 from our model shows a phenomenon that the neural network tends to predict more TPs as the RPs near them but not necessarily closest to them. This is understandable as the CSI fingerprinting offers more distance-sensitive features that can be potentially utilized for high accuracy indoor localization than RSSI can do. The higher average distance between RPs and TPs under AP6 makes the collected CSI signals from those RPs and TPs share less feature similarities with even short distances. The largest distance errors from OpenCSI, CiFi and DelFin come in AP4 due to more misestimated locations with the increasing number of RPs while CSILoc and CSILoc-NoDen still managed to obtain low prediction errors, highlighting their ability to scale to larger and more complex indoor environments.

TABLE 6: Network parameters and inference time (IT)

	OpenCSI	CiFi	DelFin	<i>CSILoc</i>
Parameters	22,910,102	28,052-33,852	53,890	11,280-12,600
IT (ms)	322	2.25	2.93	1.89

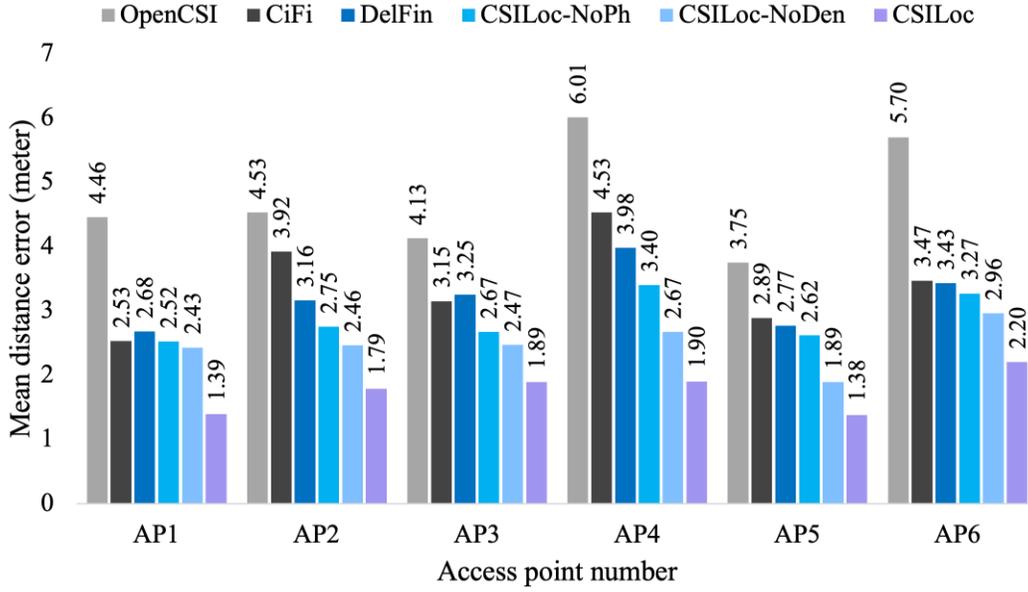


Fig. 18. Mean distance error comparison between 4 models for each AP

Fig. 18 summarizes the mean distance error for each AP, for the compared frameworks. Note that while many prior studies on CSI-based localization highlight decimeter-level accuracies, e.g., [29], these frameworks require RPs separated by centimeters which is not practical for large, real-world environments. Moreover, these studies typically consider small, isolated areas without considering dynamic interference effects over time. From the figure, it can be observed that our proposed CSILoc framework improves the mean distance error performance up to 68.5%, 58.1%, and 52.3% compared to OpenCSI, CiFi, and DelFin. We also obtained the inference time of all frameworks running on a Samsung Galaxy S7 smartphone with Android v8.0.0 and deep learning models prototypes with TensorFlow [46]. Table 6 shows the average inference time to predict a location with CSILoc and the three frameworks from prior work. The table also shows the model parameters for each framework. The number of parameters for CiFi and CSILoc vary depending on the number of RPs, which changes the number of neurons in the final layer that are used for classification in these frameworks. To obtain inference time for these frameworks, we

conservatively used the corresponding AP4 models which have the most neurons in the Softmax layer. From the table, it can be observed that our proposed CSILoc framework not only has the smallest memory footprint (lowest parameter count) but also the smallest inference time.

In summary, based on results shown in Fig. 17, Fig. 18, and Table 6, our proposed CSILoc framework shows better localization accuracy, memory footprint, and execution time, compared to state-of-the-art CSI-based localization frameworks compared against it. Thus, CSILoc represents a promising solution to achieve fast, lightweight, and accurate localization with CSI data in large and complex indoor environments.

### 3.7. CONCLUSION

In the chapter, we propose a novel CSI-based indoor localization framework called CSILoc that involves an efficient collection, clustering, denoising, and calibration pipeline. We also proposed a 1D-CNN based neural network architecture to deliver a lightweight deep learning method towards obtaining accurate localization estimation with CSI data. Our work shows up to 68.5% improved performance (mean distance error) compared to three recent deep learning and CSI-based indoor localization frameworks. Another important contribution of our work is to create, leverage, and release an open source dataset of floor-level CSI signals collected in a large and highly dynamic indoor environment [37]. The dataset is built with a single receiving antenna client and aims at providing a lightweight dataset for resource-constrained and low-energy consumption mobile devices. This framework requires minimum knowledge about APs and does not need physical access to them (unlike frameworks in prior work that modify the APs' behavior to inject custom frames or adjust antenna configurations), which solves the problem of inaccessibility of signal sources in practice when security issues are involved in the data collection space.

## 4. CONCLUSIONS AND OPEN PROBLEMS

### 4.1. RESEARCH CONCLUSION

In our research of indoor localization as presented in this thesis, we developed two deep learning frameworks based on Wi-Fi sensing with RSSI and CSI, respectively.

In CHISEL, we used RSSI as signal source for estimating locations. We benchmarked CHISEL against state-of-the-art deep learning based prior work on the same open dataset, and achieved the lowest mean distance error even without data augmentation. Different pruning and quantization configurations were applied on the fine-tuned CHISEL CAECNN model trained with augmented data varying from 75% to 25% sparsity of neurons and 16 bits to 2 bits on weight only or weight + activation quantization. The experimental results showed that the INT4-WO-25 with QAT is one of the promising CHISEL configurations, which reduced model size to 148 kB (81.52 percent reduction) and latency by 1.80ms (30.93 percent reduction) at the expense of 0.34 m (4.89 percent) localization accuracy.

In our CSI study, we presented CSILoc, a novel CSI-based indoor localization framework with an efficient collection, clustering, denoising, and calibration pipeline. We also proposed a 1D-CNN-based neural network architecture to provide a lightweight deep learning method for estimating accurate localization with CSI data. When compared to three recent deep learning and CSI-based indoor localization frameworks, our work showed up to 68.5 percent improved performance (mean distance error). Another significant contribution of our work is creating an open-source dataset of floor-level CSI signals collected in a large and dynamic indoor environment [29].

## 4.2. OPEN PROBLEMS

Despite the promising results obtained in our research work for predicting a location with RSSI and CSI, there are still many open problems that remain.

### 4.2.1. HIGH WI-FI FACILITY COSTS USING RSSI

The main downside of RSSI is the constant fluctuations introduced by multipath effects since RSSI is evaluated at per packet level, which, however, can be well solved using CSI. As mentioned earlier, each RSSI value in a RSSI fingerprint corresponds to a dimension of feature to train machine learning algorithms. RSSI fingerprint vectors with fewer dimensions have less anti-noise stability than the ones with more dimensions, especially when the multicollinearity is introduced by RSSI fluctuations. Thus, increasing the number of feature dimensions (resolution) of the RSSI vectors by adding more APs can alleviate this phenomenon, which means higher investment on Wi-Fi infrastructure.

### 4.2.2. HIGH NUMBER OF DEEP LEARNING PARAMETERS USING RSSI

On the other hand, the increasing number of APs may require more complicated deep learning models that usually need more parameters for the sake of more powerful feature extracting. The number of parameters of a deep learning model critically affects the memory footprint, inference time and energy consumption on resource-constrained devices such as robots, wearables and smartphones, etc. The trade-off between increasing the feature dimensions to reduce noise from RSSI fluctuations and the deeper model with more parameters is still a challenging problem. As the results we showed in CHISEL, performing proper compression techniques can significantly reduce the size of a deep learning model and the time to get estimated locations, and meanwhile keep a reasonable localization accuracy.

#### 4.2.3. LACK OF SOLUTIONS FOR MODEL SCALABILITY

Deep learning models for indoor localization suffer from scalability concerns when the service space is large, for example, the big buildings like hospital, libraries, museums and underground constructions. There is a need to consider how to expand models to larger space with minimum network parameters required but keep high localization abilities. There are still lots of thing worth studying. For limited space, it is easier for a classification-based deep learning model to achieve stable and high indoor localization accuracies compared with regression-based models whose performance can be highly affected by the relative positions of RPs. On the contrary, for example, classification models may need to be converted into regression models, due to the number of neurons for categories increases along with the increasing number of RPs in large indoor space.

#### 4.2.4. SKEWED PATTERN NUMBERS USING CSI

In our work, the number of packets in each pattern is imbalanced which introduces a bias in deep learning models, potentially limiting the generalizability of the models. New data augmentation techniques may reveal more potential from CSI data with rich multipath features. In recent years, generative models like autoencoders and generative adversarial networks (GANs) have achieved an impressive progress in producing realistic fake data. The fake data with artificial noise can be utilized as augmented data for the original dataset which are helpful to balance skewed datasets. Our future work will incorporate generative models with positioning algorithms to enhance the generalization performance of deep learning algorithms.

#### 4.2.5. LACK OF RESEARCH IN THE POSITION OF SIGNAL RECEPTION

Plenty of research has been done in improving localization accuracy, but few studies discuss the big role of signal reception in practical deployment. If an offline signal data collection process only places the receiver in a 2D plane without considering the vertical variations along, for example the “Z” axis in a Cartesian system, this IPS probably will fail in practical cases where the users with different ages, heights or habitats, etc., could have their devices receive signals at different heights. Although RSSI values are coarse measurements of distance information between transmitters and receivers at packet level, this phenomenon still exists. In a CSI-based high sensitive IPS, lack of data collected along different spatial points in vertical dimension has high possibilities to make the localization predictions unpredictable.

#### 4.2.6. LIMITED OPTIONS IN CSI EXTRACTION TOOL

CSI-based indoor localization has taken off in this past decade, but research in this area is still at its infancy, due to limited options and supporting functions in CSI extractors. In early 2010s, researchers usually extracted CSI information with dedicated hardware that was not openly accessible. The frequency diversities and bit width for representing signals are very limited. Even in the early 2020s, there are few customization options in CSI extraction tools when different research objects are required. For example, the frequency options of most tools for 802.11ac and 802.11ax are restricted at 20Mhz and 40Mhz [10]. Extracting CSI is sometimes related to hacking wireless systems and security issues, and the Wi-Fi chipset vendors tend to avoid publishing their underlying algorithms or open source APIs for CSI data profiling, which is also a strong barrier in this area. There is an urgent need to develop more publicly available software to help indoor localization community.

#### 4.2.7. UNDOCUMENTED AGC ALGORITHMS

Although the AGC can convey very important distance information between transmitters and receivers, an optimal solution for eliminating the AGC effects is difficult to obtain for IPS using CSI, due to the undocumented AGC algorithms from chipset vendors. If such information becomes available, the effectiveness of CSI-based localization frameworks can be further improved. A possible way to avoid dealing with AGC algorithms is to extract CSI before AGC filtering, which is not available of current open tools.

#### 4.2.8. LABOR INTENSIVE DATA COLLECTION

The offline RSSI and CSI data collection phase needs considerable labor. For example, RSSI data need to be collected with enough number of packets for each RP to observe RSSI fluctuation patterns for choosing corresponding denoising algorithms. Collecting enough patterns from CSI packets in a large and complex area is also labor-intensive. In our CSI study, collecting 3500 packets in a single RP takes about 6~7 minutes excluding recollection time, due to the unpredictable disconnections to a AP. Hence, how to automate the data capture process is a critical step to make CSI-based indoor localization practically efficient and feasible. For example, when an AP of interest is updated, the CSI collected within the area covered by this AP needs to be recollected. CSI has the potential to help localization systems achieve low localization errors, however, this needs fine-grained RP allocation which means more labor is required in the data collection process.

#### 4.2.9. LACK OF SOLUTIONS FOR SEAMLESS ZONE SWITCHING USING CSI

Commonly, IPS built on CSI do not necessarily need RSSI inputs for training ML algorithms. But because the CSI packets for estimating positions are often AP bound and modern IoT devices

usually automatically switch Wi-Fi coverage zones based on RSSI quality, RSSI is unstable over time, and there is a great need to develop a solution for zone switching in practice for CSI-based IPS. For example, if a user is at the overlapping space of two different AP coverage zones where the received RSSI values from two APs are close to each other, the fluctuating RSSI could lead to unacceptable localization errors. Thus, a better topology of AP allocation and soft switching algorithms are necessary for choosing the correct model bound to an AP.

#### 4.2.10. LACK OF USER-FRIENDLY INDOOR LOCALIZATION APPLICATIONS

Indoor localization for IoT devices as an important component of human-centric services will eventually go to office space, families, individuals and general use cases, etc. For example, there is a great potential in the market for family-oriented intellectual assistant robots. In addition to common sensors, these robots may need wireless communication systems like the ubiquitous connections offered by Wi-Fi to help for localizing their positions and to interact with other IoT devices. The indoor localization modules of these IoT devices are recommended to be reprogrammable based on users' needs. Thus, there is an urgent need to develop user friendly indoor localization applications including instant deployment, customized and automatic data collection, testing and localization that need minimum expertise. This is one of our future studies to benefit indoor localization society.

#### 4.2.11. LACK OF SOLUTIONS FOR LOCALIZATION SECURITY

If an attacker hacks into APs to change RSSI, the IPS can be easily fooled with intentionally modified RSSI values and could guide users to an unexpected destination for some reason, assuming that the attacker is aware of the fundamentals of a fingerprinting based IPS. In a CSI-based IPS, a user's position can be accurately tracked, which could lead to a more dangerous

situation. In a more common case, attackers may only want to stop an IPS from working. This can be done by adding noise via APs or emitting noise signals to lower SNR in RSSI and CSI signals. Thus, the input fingerprints are affected with certain noise patterns, which could either make the IPS fail or induce it to perform other localization functions for some purposes.

#### 4.2.12. LACK OF SOLUTIONS FOR DEVICE HETEROGENEITY

It is inevitable to see the different types of IoT devices used by users. As mentioned in our CSI study, 802.11 standards have not specified a uniform formula to calculate RSSI but only recommend a range. Specifically, the RSSI measurements and computing algorithms are not invariant over different types of smartphones. Thus, there is a practical need to develop solutions for alleviating the heterogeneity issue cross different signal receiving platforms. Typically, smartphones from the same manufacture have the similar RSSI and CSI reception patterns. Input characteristics of different platforms to a deep learning model could help the model generalize better.

## BIBLIOGRAPHY

- [1] I. Tsitsimpelis, C. J. Taylor, B. Lennox, and M. J. Joyce, "A review of ground-based robotic systems for the characterization of nuclear environments," *Progress in Nuclear Energy*, vol. 111, pp. 109-124, 2019.
- [2] C. Langlois, S. Tiku, and S. Pasricha, "Indoor localization with smartphones: Harnessing the sensor suite in your pocket," *IEEE Consumer Electronics Magazine*, vol. 6, no. 4, pp. 70-80, 2017.
- [3] V. Cisco, "Cisco visual networking index: Forecast and trends, 2017–2022," *White paper*, vol. 1, no. 1, 2018.
- [4] U. Cisco, "Cisco annual internet report (2018–2023) white paper," *Cisco: San Jose, CA, USA*, 2020.
- [5] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, pp. 1-32, 2013.
- [6] C.-Y. Wang and H.-Y. Wei, "IEEE 802.11 n MAC enhancement and performance evaluation," *Mobile Networks and Applications*, vol. 14, no. 6, pp. 760-771, 2009.
- [7] 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.
- [8] 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.
- [9] "Nexmon." [Online]. Available: <https://github.com/seemoo-lab/nexmon>.

- [10] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets," in *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, 2019, pp. 21-28.
- [11] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568-2599, 2019.
- [12] K. Shi, Z. Ma, R. Zhang, W. Hu, and H. Chen, "Support vector regression based indoor location in IEEE 802.11 environments," *Mobile Information Systems*, vol. 2015, 2015.
- [13] Y. Xie, Y. Wang, A. Nallanathan, and L. Wang, "An improved K-nearest-neighbor indoor localization method based on spearman distance," *IEEE signal processing letters*, vol. 23, no. 3, pp. 351-355, 2016.
- [14] E. Jedari, Z. Wu, R. Rashidzadeh, and M. Saif, "Wi-Fi based indoor location positioning employing random forest classifier," in *2015 international conference on indoor positioning and indoor navigation (IPIN)*, 2015: IEEE, pp. 1-5.
- [15] J.-W. Jang and S.-N. Hong, "Indoor localization with wifi fingerprinting using convolutional neural network," in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, 2018: IEEE, pp. 753-758.
- [16] M. Nowicki and J. Wietrzykowski, "Low-effort place recognition with WiFi fingerprints using deep learning," in *International Conference Automation*, 2017: Springer, pp. 575-584.
- [17] J. Torres-Sospedra, Raul Montoliu, Adolfo Martinez-Uso, Joan P. Avariento, Tomas J. Arnau, Mauri Benedito-Bordonau, and Joaquin Huerta., "UJIIndoorLoc: A new multi-

- building and multi-floor database for WLAN fingerprint-based indoor localization problems," in *2014 international conference on indoor positioning and indoor navigation (IPIN)*, 2014: IEEE, pp. 261-270.
- [18] K. S. Kim, S. Lee, and K. Huang, "A scalable deep neural network architecture for multi-building and multi-floor indoor localization based on Wi-Fi fingerprinting," *Big Data Analytics*, vol. 3, no. 1, pp. 1-17, 2018.
- [19] X. Song, X. Fan, C. Xiang, Q. Ye, L. Liu, Z. Wang, X. He, N. Yang, and G. Fang, "A novel convolutional neural network based indoor localization framework with WiFi fingerprinting," *IEEE Access*, vol. 7, pp. 110698-110709, 2019.
- [20] S. Tiku, P. Kale, and S. Pasricha, "QuickLoc: Adaptive Deep-Learning for Fast Indoor Localization with Mobile Devices," *ACM Transactions on Cyber-Physical Systems (TCPS)*, vol. 5, no. 4, pp. 1-30, 2021.
- [21] S. Tiku, S. Pasricha, B. Notaros, and Q. Han, "A Hidden Markov Model based smartphone heterogeneity resilient portable indoor localization framework," *Journal of Systems Architecture*, vol. 108, p. 101806, 2020.
- [22] S. Tiku and S. Pasricha, "Overcoming Security Vulnerabilities in Deep Learning--based Indoor Localization Frameworks on Mobile Devices," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 18, no. 6, pp. 1-24, 2019.
- [23] S. Tiku and S. Pasricha, "Siamese Neural Encoders for Long-Term Indoor Localization with Mobile Devices," *arXiv preprint arXiv:2112.00654*, 2021.
- [24] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and L. M. Ni, "CSI-based indoor localization," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 7, pp. 1300-1309, 2012.

- [25] 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.
- [26] J. Xiao, K. Wu, Y. Yi, and L. M. Ni, "FIFS: Fine-grained indoor fingerprinting system," in *2012 21st international conference on computer communications and networks (ICCCN)*, 2012: IEEE, pp. 1-7.
- [27] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763-776, 2016.
- [28] B. Berruet, O. Baala, A. Caminada, and V. Guillet, "DelFin: A deep learning based CSI fingerprinting indoor localization in IoT context," in *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2018: IEEE, pp. 1-8.
- [29] A. Gassner, C. Musat, A. Rusu, and A. Burg, "OpenCSI: An Open-Source Dataset for Indoor Localization Using CSI-Based Fingerprinting," *arXiv preprint arXiv:2104.07963*, 2021.
- [30] E. Schmidt, D. Inupakutika, R. Mundlamuri, and D. Akopian, "SDR-Fi: Deep-learning-based indoor positioning via software-defined radio," *IEEE Access*, vol. 7, pp. 145784-145797, 2019.
- [31] L. Wang, S. Tiku, and S. Pasricha, "CHISEL: Compression-Aware High-Accuracy Embedded Indoor Localization with Deep Learning," *IEEE Embedded Systems Letters*, 2021.
- [32] A. Mittal, S. Tiku, and S. Pasricha, "Adapting convolutional neural networks for indoor localization with smart mobile devices," in *Proceedings of the 2018 on Great Lakes Symposium on VLSI*, 2018, pp. 117-122.

- [33] R. Mishra, H. P. Gupta, and T. Dutta, "A survey on deep neural network compression: Challenges, overview, and solutions," *arXiv preprint arXiv:2010.03954*, 2020.
- [34] Y. Bengio, N. Léonard, and A. Courville, "Estimating or propagating gradients through stochastic neurons for conditional computation," *arXiv preprint arXiv:1308.3432*, 2013.
- [35] H. Li, A. Kadav, I. Durdanovic, H. Samet, and H. P. Graf, "Pruning filters for efficient convnets," *arXiv preprint arXiv:1608.08710*, 2016.
- [36] P. Molchanov, S. Tyree, T. Karras, T. Aila, and J. Kautz, "Pruning convolutional neural networks for resource efficient inference," *arXiv preprint arXiv:1611.06440*, 2016.
- [37] L. Wang and S. Pasricha. "CSILoc: A CSI-Based Indoor Localization Framework with 1D Convolutional Neural Networks." <https://github.com/awang0413/CSILoc> (accessed.
- [38] X. Wang, X. Wang, and S. Mao, "CiFi: Deep convolutional neural networks for indoor localization with 5 GHz Wi-Fi," in *2017 IEEE International Conference on Communications (ICC)*, 2017: IEEE, pp. 1-6.
- [39] C. D. Fryar, D. Kruszan-Moran, Q. Gu, and C. L. Ogden, "Mean body weight, weight, waist circumference, and body mass index among adults: United States, 1999–2000 through 2015–2016," 2018.
- [40] M. S. Gast, *802.11 ac: a survival guide: Wi-Fi at gigabit and beyond*. " O'Reilly Media, Inc.", 2013.
- [41] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 881-892, 2002.
- [42] M. Schulz, J. Link, F. Gringoli, and M. Hollick, "Shadow Wi-Fi: Teaching smartphones to transmit raw signals and to extract channel state information to implement practical covert

- channels over Wi-Fi," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, 2018, pp. 256-268.
- [43] M. Schulz, "Teaching your wireless card new tricks: Smartphone performance and security enhancements through wi-fi firmware modifications," 2018.
- [44] I. C. S. L. M. S. Committee, "IEEE Standard for Information technology- Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *IEEE Std 802.11<sup>^</sup>*, 2007.
- [45] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.
- [46] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.