Large-Scale Indoor Localization via Outdoor Crowdsourcing Trajectories on Ride-Hailing Platform

SHULI ZHU*, Beijing Jiaotong University, China
LINGKUN LI*, Beijing Jiaotong University, China
XUYU WANG, Florida International University, United States
QIANG NI, Lancaster University, United Kingdom
YUQIN JIANG, DiDi Company, China
HUI GAO, DiDi Company, China
ZHAOBING HAN, DiDi Company, China
RUIPENG GAO†, Beijing Jiaotong University, China

Indoor positioning is critical for a variety of services, including ride-hailing. However, existing large-scale fingerprint-based indoor positioning systems face significant challenges due to high deployment costs, temporal instability and limited accessibility, making them impractical for widespread use. In this paper, we propose a novel approach to indoor positioning that leverages fingerprints only sampled outdoors, which can be collected through crowdsourcing within a ride-hailing platform. This approach significantly reduces deployment costs, enables timely updates to the fingerprint set, and provides unprecedented accessibility. We address three key challenges in this system, including using outdoor fingerprints to estimate indoor position, abnormal Access Points (APs), and existence of "blackholes" where overheard APs have no fingerprint. Our implementation, built on the DiDi ride-hailing platform, is evaluated through extensive experiments with 122 million orders across 13 million devices in multiple cities. The results demonstrate that our system achieves a significant reduction of 4.35m in pickup position error compared to existing efforts, showcasing its potential for large-scale adoption.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**.

Additional Key Words and Phrases: Crowdsourced WiFi Localization, Anomaly Detection, Fingerprint Augmentation, Ride-hailing Platform

ACM Reference Format:
Shuli Zhu, Lingkun Li, Xuyu Wang, Qiang Ni, Yuqin Jiang, Hui Gao, Zhaobing Han, and Ruipeng Gao. 2025. Large-Scale Indoor Localization via Outdoor Crowdsourcing Trajectories on Ride-Hailing Platform. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 2, Article 66 (June 2025), 23 pages. https://doi.org/10.1145/3729498

*Both authors contributed equally to this research.
†Corresponding author.

Authors' Contact Information: Shuli Zhu, zhushuli@bjtu.edu.cn, Beijing Jiaotong University, Beijing, China; Lingkun Li, lkli@bjtu.edu.cn, Beijing Jiaotong University, Beijing, China; Xuyu Wang, xuywang@fiu.edu, Florida International University, Miami, United States; Qiang Ni, q.ni@lancaster.ac.uk, Lancaster University, Lancaster, United Kingdom; Yuqin Jiang, jiangyuqin@didiglobal.com, DiDi Company, Beijing, China; Hui Gao, deangaohui@didiglobal.com, DiDi Company, Beijing, China; Zhaobing Han, hanzhaobinghzb@didiglobal.com, DiDi Company, Beijing, China; Ruipeng Gao, rpgao@bjtu.edu.cn, Beijing Jiaotong University, Beijing, China.

ACM 2474-9567/2025/6-ART66
https://doi.org/10.1145/3729498

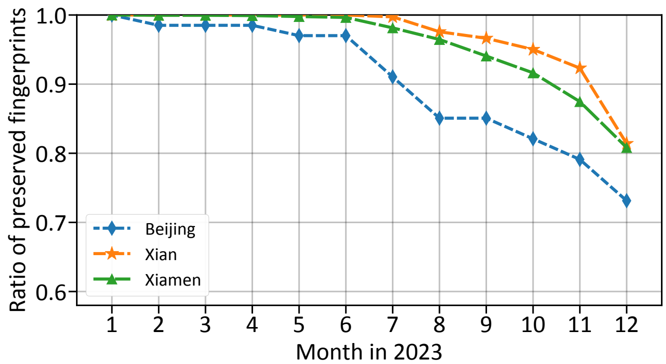


Fig. 1. **Fingerprints change in three typical cities.**

1 Introduction

Indoor localization is a fundamental enabling technique in various location-based applications, especially on ride-hailing platforms where passengers always issue a pickup request indoors. Based on our two-year experience on DiDi, a prominent ride-hailing platform akin to Uber and Lyft, over 952,000 daily travel orders (~ 28% of the total) are originated from indoor buildings without satellite receptions. Therefore, the platform has to locate indoor passengers without commonly used Global Positioning System (GPS), and recommend them optimal pickup positions to wait for drivers. Despite popular travelling services, one major obstacle in industry is to balance the localization accuracy and deployment cost at large scale.

Status Quo and Limitations: Many recent research efforts propose indoor positioning schemes, such as WiFi Received Signal Strength Indicator (RSSI)-based [13, 42, 45], WiFi Channel State Information (CSI)-based [39, 43, 53], ultrasound-based [7, 25, 54], vision-based [26], and visible light-based systems [23, 44]. However, these approaches are not scalable and are often impractical due to issues like multipath effects and the strict requirement of line-of-sight views, as we will discuss in Section 2. Consequently, industrial and commercial systems [29] often rely on fingerprint-based approaches for large-scale indoor positioning. However, several factors limit their practicality, including:

- **High deployment costs:** Many systems depend on dedicated sensors or manually collected fingerprints, requiring service providers to perform fine-grained "site surveys" and periodic patrols to update both physical and Radio Frequency (RF) data. Given the vast number of existing buildings, this process is time-consuming, labor-intensive, and costly, making it impractical for large-scale indoor positioning systems.
- **Temporal instability:** Fingerprints are affected by various factors, including protocol versions and device types, which are periodically updated. Figure 1 shows that only 70% to 80% of WiFi fingerprints collected at the beginning of a year are preserved after a year. Consequently, to maintain stability, developers must periodically update the fingerprint set, which is costly due to the need for repeated site surveys.
- **Limited accessibility:** Crowdsourcing has been explored to mitigate manual effort. For example, Lab2Wild [29] pioneered the use of Access Point (AP) locations to initialize fingerprints. They observed that many businesses deploy APs with names related to their stores, enabling location estimation through name matching with building floor plans. However, indoor floor plans with detailed room-level annotations are not always accessible and can raise privacy concerns, particularly in official or residential buildings. Moreover, accurately matching AP names to specific Point of Interests (PoIs) on floor plans is challenging,

especially when multiple languages are involved. Relying on crowdsourced users to voluntarily upload and correct fingerprints also demands expensive incentive programs, making it difficult to sustain large-scale, long-term services.

Our Solution: In this paper, we propose an alternative approach that shifts the reliance on indoor-collected fingerprints to fingerprints gathered outdoors. This transformation allows us to leverage crowdsourcing to build the fingerprint set from outdoor trajectories, which is then used for full building-level indoor positioning. Our approach offers several key advantages: First, by utilizing a ride-hailing platform for crowdsourced fingerprint collection, we eliminate the need for labor-intensive indoor fingerprint sampling, significantly reducing deployment costs. Second, the fingerprint set can be promptly updated as protocols and devices change, ensuring continued accuracy over time. Lastly, this method circumvents the need to enter restricted areas for fingerprint collection, expanding the applicability of indoor positioning to a broader range of scenarios.

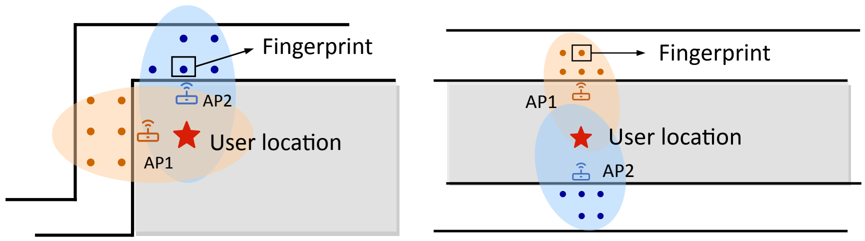
Challenges: Achieving the benefits of our approach presents several significant challenges:

- **Use outdoor fingerprints to estimate indoor position:** A key challenge is identifying a suitable method to perform indoor localization using surrounding outdoor fingerprints. Since WiFi signatures measured outdoors cannot be directly applied to indoor environments, we incorporate additional features–such as scanning amount and building contour–into a convolutional neural network (CNN) model to estimate indoor position.
- **Abnormal APs producing unreliable fingerprints:** Numerous abnormal access points can undermine the reliability of WiFi-based localization. For example, smartphones acting as mobile hotspots ("mobile APs") generate fingerprints that span wide areas due to their mobility. Furthermore, some APs can permanently relocate ("migrant APs"), while others may share the same Basic Service Set Identifier (BSSID) across different devices ("cloned APs"). These issues can mislead users to incorrect locations, disrupting pickup services. To address this, we need to automatically detect and classify abnormal APs, identify their types, and correct their associated fingerprints.
- **Existence of "blackholes":** Large buildings may contain indoor APs that are not detectable from outdoor trajectories. When passengers initiate pickup requests from these locations, it is possible that none of the APs in the WiFi scan list will have associated fingerprints. In these cases, localization relies solely on cellular signals, which can lead to significant errors or even incorrect building identification. To resolve these localization "blackholes", we iteratively infer the locations of indoor APs by analyzing their co-occurrence in indoor pickup queries. This approach allows us to leverage unfamiliar APs to generate comprehensive and accurate indoor fingerprint anchor points, providing reliable localization accuracy even in areas with sparse or missing fingerprints.

We conducted comprehensive experiments to evaluate the performance of our system. The experimental results show that our system achieves a median indoor position error of $9 \sim 14$ meters when the user is near the perimeter of a building, and a median pickup position error of 10.44 meters. This level of accuracy highlights the potential of our system for large-scale indoor positioning, particularly in recommending optimal pickup points for ride-hailing services when a user initiates an order indoors.

Our contributions are listed as follows:

- **Novel Indoor Positioning Scheme:** This paper introduces an innovative fingerprint-based indoor positioning scheme that relies on fingerprints collected outdoors and crowdsourcing. This approach eliminates the need for costly site surveys in every building, ensures timely updates of the fingerprint set for stable service, and offers unparalleled accessibility, as the system functions using outdoor fingerprints without needing to enter buildings.
- **Key Technical Solutions:** The system incorporates three major techniques: anomaly identification, fingerprint enhancement, and real-time localization using outdoor fingerprints, addressing the challenges



(a) Corner case

(b) Slim case

Fig. 2. **Specific cases for indoor localization using outdoor fingerprints.**

of building a reliable indoor positioning system with outdoor data. These techniques provide valuable insights for the research community focused on indoor positioning.

- **Practical Deployment and Large-Scale Evaluation:** The system was deployed on a ride-hailing platform and evaluated on a large-scale dataset of 122 million travel orders across 13 million devices in multiple cities over a two-year period. The results, based on both industrial service metrics and manual annotations, demonstrate the robustness and reliability of the system.

2 Background and Motivation

2.1 Limitations of Existing Indoor Positioning Systems

As mobile networks evolve, location-based services (LBS) have become integral to daily life. These services often rely on precise positioning; for example, users may request their location before receiving recommendations for nearby restaurants. Navigation services, in turn, use this accurate location data to guide users to their destinations.

In outdoor environments, GPS provides highly accurate positioning information. However, GPS signals are easily obstructed by buildings, making it ineffective for indoor scenarios. As a result, researchers have developed various indoor positioning systems to address this challenge and deliver accurate location data when users request positioning services indoors.

Existing indoor positioning systems can be categorized into several distinct approaches: fingerprint-based [5, 30], RF-based [33, 36, 46, 50], ultrasound-based [7, 25, 54], vision-based [26], and visible light-based [23, 44]. Among these, RF-based and ultrasound-based systems, which rely on time-difference-of-arrival (TDoA) [14], angle-of-arrival (AoA) [48], etc., can offer high accuracy but are susceptible to multipath effects—signal distortion caused by interference from multiple moving users–making them less effective in multi-target environments. On the other hand, vision-based systems are limited by illumination conditions and the need for a clear line of sight, while visible light-based systems require the deployment of landmarks and anchors in advance, adding complexity.

As a result, fingerprint-based approaches are the most widely adopted, particularly in large-scale industrial and commercial applications. These systems often rely on technologies such as WiFi [29], Bluetooth [17], or Earth's magnetism [41, 47]. However, Bluetooth devices have high mobility, and Earth's magnetism is prone to environmental changes, making them unstable for consistent positioning. In contrast, WiFi APs are widely deployed and typically fixed in place, which is why many industrial systems build fingerprint-based indoor positioning systems using existing WiFi infrastructure.

However, through our extensive investigation, we identified three major limitations, i.e., high deployment costs, temporal instability, and limited accessibility, of existing WiFi fingerprint-based indoor localization systems, as discussed in Section 1. Therefore, there is a need for a novel indoor positioning system that is not only scalable and low-cost but also resilient to time-based changes.

2.2 Our Finding: Possibility of Indoor Positioning Using Fingerprints Collected Outdoors

To address the limitations of existing systems, we propose a novel WiFi fingerprint-based indoor positioning scheme based on a key insight: fingerprints collected outdoors can be effectively used for indoor positioning[1]. As illustrated in Figure 2, the building area (shaded in gray) is surrounded by roads. In this scenario, devices on the road can detect WiFi signals emitted from APs located inside the building. These signals can be used to build fingerprints, with the device's outdoor position accurately estimated using GPS. When a user inside the building initiates a positioning request, these outdoor-collected fingerprints can help estimate the user's indoor position.

For example, as shown in Figure 2a, WiFi signals from AP1 are detectable only at the orange dots on the outside road, while signals from AP2 are detectable only at the blue dots. AP2's signals cannot be detected at the orange dots due to the distance and the signal attenuation caused by the building's complex environment, and similar to AP1's signals. In this case, fingerprints can be built at both the orange and blue dots. If a user is located at the position marked by the red star, they can detect signals from both AP1 and AP2. By leveraging the GPS positions from fingerprints collected at the orange and blue dots, we can estimate the user's indoor position.

Similarly, as shown in Figure 2b, it is difficult to detect signals from AP1 at the blue dots and from AP2 at the orange dots. However, the user at the position denoted by the red star can detect both APs, enabling indoor positioning based on the outdoor-collected fingerprints from both sets of dots.

This discovery allows us to design a new crowdsourced WiFi fingerprint-based indoor positioning system. Fingerprint collectors detect WiFi signals from outside the building without the need to enter, building the fingerprint set in the process. Users inside the building can then use this set for indoor positioning. This approach offers several advantages that address previous limitations:

- **Low development and maintenance costs:** By relying on crowdsourcing, the system removes the need for developers to conduct expensive site surveys.
- **Prevent fingerprint set from becoming outdated:** The crowdsourcing approach ensures real-time updates of the fingerprint set, preventing it from becoming outdated.
- **Unmatched accessibility:** As collectors never need to enter buildings, the system can provide indoor positioning even in buildings with restricted access, offering exceptional accessibility through the use of outdoor fingerprints for accurate indoor positioning.

2.3 System Overview

Although our system addresses the limitations mentioned above, building a crowdsourcing fingerprint-based indoor positioning system is not without challenges. As discussed in Section 1, we must overcome three key challenges: (1) using outdoor fingerprints to estimate indoor positions, (2) dealing with abnormal APs that produce unreliable fingerprints, and (3) handling the existence of "black holes" where no fingerprints are available.

To tackle these challenges, we designed our system, as shown in Figure 3, which comprises four main stages:

- **Fingerprint Construction:** The first step in our system is constructing a fingerprint set through crowdsourcing. We require outdoor trajectories, which consist of WiFi signal lists and GPS locations with an error of less than 10 meters. These outdoor trajectories are then divided into $10m \times 10m$ grid cells, where we collect the average Received Signal Strength Indicator (RSSI), sampling counts, and timestamps. To ensure a widespread and reliable WiFi localization service, we adopt an incremental updating process, allowing the set to grow and be continuously refined over time as new fingerprints are collected. Regarding large GPS errors, we use Kalman Filter (KF) and Hidden Markov Model (HMM) to smooth and correct the trajectories, thereby reducing GPS positioning errors. This can improve the quality of our fingerprints and further reduce the overall localization error of our system. When GPS may not be effective, we cannot

[1]We assume that a user is outdoors, if the GPS location of the user's smartphone is available within 3 seconds after the user requests an order query. In other cases, users are determined to be indoors.

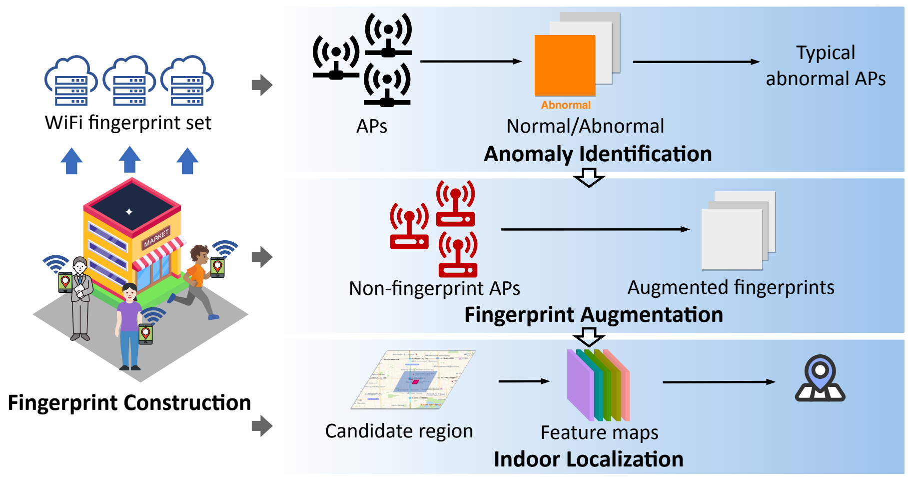


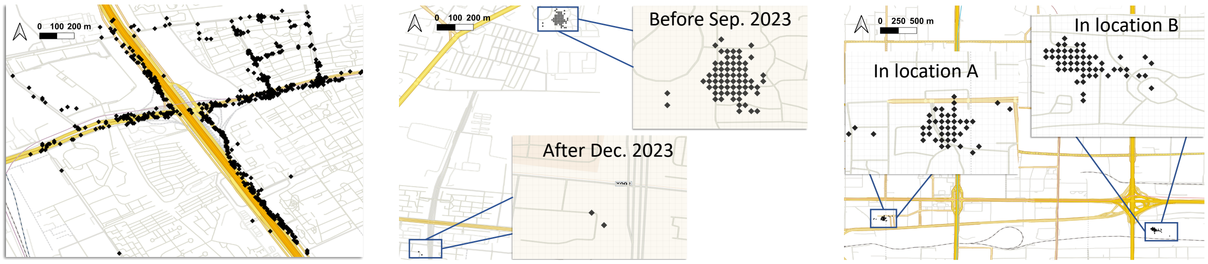
Fig. 3. **System overview.** Our system consists of four main phases, i.e., fingerprint construction, anomaly identification, fingerprint augmentation, and indoor localization.

construct fingerprints due to the lack of GPS locations. We propose fingerprint augmentation (Section 4.3) to improve localization accuracy and robustness.

- **Anomaly Identification:** Collected fingerprints cannot be directly applied for positioning due to abnormal APs, such as mobile, migrant, or cloned APs, which can lead to significant localization errors. To address this, we design a multi-autoencoder network that uses three key features—coverage area, scanning amount, and service time—to distinguish between normal and abnormal APs. Next, we implement a GBDT-based binary classification model to identify mobile APs and leverage co-occurrence analysis to detect migrant and cloned APs. This stage ensures that fingerprints from abnormal APs are filtered out, as further explained in Section 3.
- **Fingerprint Augmentation:** Some indoor APs are undetectable from outdoor trajectories, limiting the effectiveness of WiFi localization services. To overcome this, we propose an augmentation algorithm that uses a least-squares method to estimate the fingerprints of undetectable APs based on co-occurring APs. This approach addresses the challenge of “black holes” and is detailed in Section 4.
- **Indoor Localization:** In this final stage, we extract multiple feature maps from WiFi measurements on users’ smartphones and the fingerprint set. These joint features are then fed into a CNN model to estimate the users’ indoor positions. This step addresses the challenge of using outdoor fingerprints for indoor positioning and is discussed in Section 5.

3 Anomaly Identification

The quality of the fingerprint set is crucial for learning-based WiFi localization. Since we construct the fingerprint set by collecting outdoor trajectories in our ride-hailing platform, there are a variety of “dirty” fingerprints that are typically caused by numerous abnormal APs and they further impede the accuracy and usability of pickup service, e.g., recommending a nearby building or an opposite entrance. In this section, we aim to distinguish three major abnormal APs and correct associated disruptive fingerprints for WiFi localization.



(a) A mobile AP (b) A migrant AP (c) A cloned AP

Fig. 4. **Three typical abnormal access points.** (a) A smartphone serves as the mobile hotspot and continuously travels over multiple roads. (b) An AP permanently relocates, as it collapses in September and reappears in December. (c) An AP has been scanned in two areas over the same period.

3.1 Typical Abnormal Access Points

Based on our long-term large-scale industrial experience of WiFi localization, we discover that there are three typical abnormal APs which produce disruptive fingerprints and cause extreme localization errors, as show in Figure 4.

- ***Mobile APs***, i.e., a smartphone or a vehicle serves as the mobile hotspot to produce network access. Such devices always travel across large areas thus their fingerprints span a wide region, as shown in Figure 4a which exceeds $6.0km^2$. Consequently, localization with such broad fingerprints can result in a coarse-grained position.
- ***Migrant APs***, i.e., a store or a company permanently move from one location to another place. As illustrated in Figure 4b, the AP migrated $1.9km$ after September and we did not gather any fingerprints in new location until December. Due to the insufficiency and short-term collection of fingerprints in the updated place, we may incorrectly locate users in outdated areas.
- ***Cloned APs***, i.e., different APs share the same BSSID and they simultaneously stay active. As shown in Figure 4c, we discovered a cloned AP with one BSSID but serves two different buildings which are more than $4km$ apart, thus we may locate users to an incorrect building. Possible reasons for such cloned APs include manufactural or manual configurations.

In order to detect the abnormal APs, a naive method is to compare their fingerprint coverage areas with a preset threshold. However, such a rule-based determination approach can hardly adapt to crowdsourcing scenarios. For example, an AP may migrate to a nearby building which is only a few hundred meters away, or a mobile AP transfers among multiple classrooms in different class hours. Such coverage variations require complicated rules for practical adaptation, whereas implementing so many rules is not resilient and scalable in industry.

3.2 Design Overview

Our intuition is to automatically discover abnormal APs at large scale, without any manual efforts or ground truth labels. Such autonomy is very valuable in industrial deployment. Thanks to the development of autoencoder, it has shown great potential in abnormal image detection via unsupervised learning, especially when there are no clear rules for abnormal patterns. Our anomaly detection method is based on an autoencoder framework, consisting of three steps, i.e., feature extraction, unsupervised learning, and abnormal classification and correction.

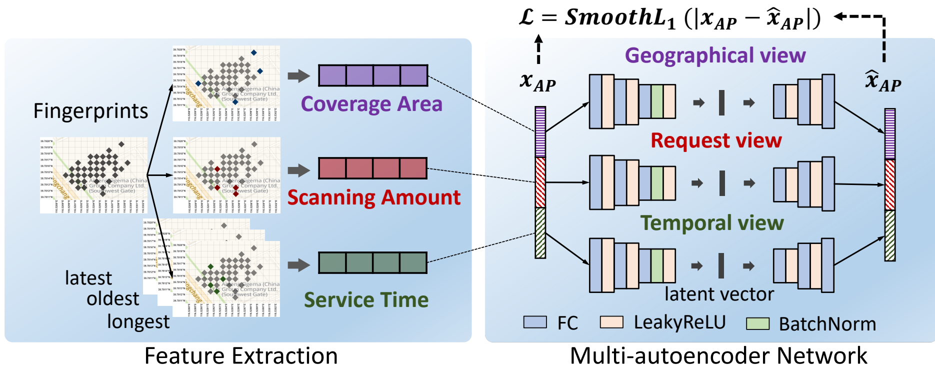


Fig. 5. **Feature extraction process (left) and multi-autoencoder network structure (right).**

In **feature extraction**, we transfer an AP's fingerprints into a vector from three aspects, i.e., coverage area at geographical view, scanning amount at request view, and service time at temporal view. We also compute and simplify the presentation of such multi-view features with long-term fingerprints.

In **unsupervised learning**, we design a multi-autoencoder model to minimize the difference between the original fingerprint features and their constructed samples. The model produces a score for each AP to quantify its abnormality. We also resample the dataset to derive only normal APs for training.

In **abnormal classification and correction**, we classify the type of each abnormal AP, and correct their associated fingerprints to reduce those extreme localization errors by abnormal APs.

3.3 Feature Extraction

We extract a feature vector x_{AP} which represents the long-term fingerprints of an access point. As shown on the left in Figure 5, we divide the city map into equal grids ($10m \times 10m$), and figure out the center grid for each AP based on its fingerprints. Each AP covers multiple grids, and each grid records long-term fingerprints and conducts daily updates.

Instead of using massive fingerprints as-is, we select the most representative grids in three aspects, including its coverage area, scanning amount, and service time.

- **Coverage area.** We pickup the farthest grids along four directions (east/south/west/north) as its geographical features. They describe the coarse coverage area for each AP, which is valuable to distinguish abnormal APs.
- **Scanning amount.** We select top N grids with the most number of localization requests to represent its user visits. Such grids reflect the nearby PoIs where users frequently initiate pickup services.
- **Service time.** To separate migrate and clone APs at temporal view, we also pick up top N grids with the latest timestamp, the oldest timestamp, and the longest service duration, respectively.

In addition, the parameter N on grid density is set as 4 based on experiments in Section 6.3. The left of Figure 5 depicts the feature construction flow, and we leverage $4 + 4 + 3 \times 4 = 20$ representative grids to construct a fingerprint feature vector.

3.4 Unsupervised Learning

We propose an unsupervised learning method to discover the abnormal APs without any ground truth labels. It utilizes a multi-autoencoder framework to learn and quantify the abnormality of crowdsourced APs.

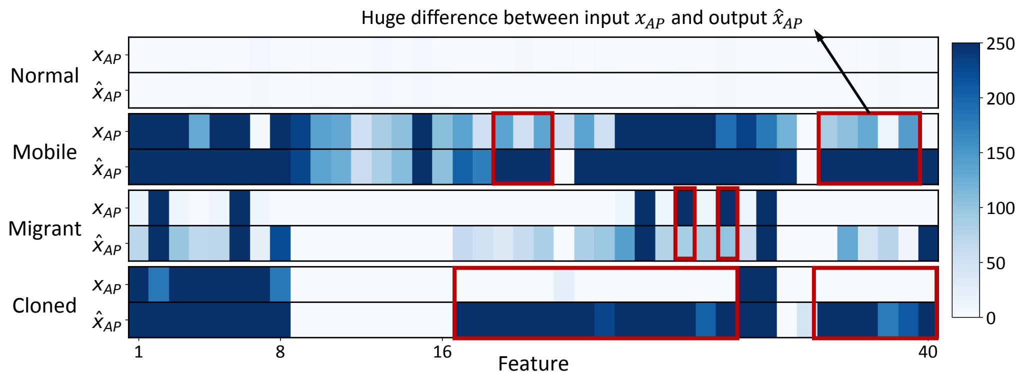


Fig. 6. **Comparison of input (x_{AP}) and output ($\hat{x}_{AP}$) for a trained multi-autoencoder model over four different APs.**

Structure. As shown on the right in Figure 5, we design three autoencoders on geographical, request, and temporal views, respectively. Each autoencoder includes an encoder to learn the feature distribution of normal APs, and a decoder to reconstruct the feature vector for comparison with original samples. Both the encoder and the decoder consists of three Fully Connected (FC) layers.

Objective. We aim to minimize the difference between original feature vector x_{AP} and the reconstructed one $\hat{x}_{AP}$. Given the training dataset $\mathcal{D}_n = \{\boldsymbol{x_1}, \cdots, \boldsymbol{x_n}\}$ for n APs without any ground truth labels, the model objective is formulated as:

$$\min \frac{1}{n} \sum_{i=1}^{n} SmoothL_1(|\boldsymbol{x_i} - \boldsymbol{\hat{x}_i}|), \tag{1}$$

where $|\cdot|$ denotes the absolute value. In addition, we compute L_1 loss instead of the L_2 loss which is sensitive to outliers. The Smooth L1 loss $SmoothL_1$ [12] is defined as follows:

$$SmoothL_1(\delta) = \begin{cases} 0.5\delta^2, & \text{if } \delta < 1 \\ \delta - 0.5. & \text{otherwise} \end{cases} \tag{2}$$

Dataset resampling. Since it is difficult to derive the precise number of abnormal APs in large-scale fingerprint set, we iteratively discover "normal" access points by resampling the dataset, thus incrementally construct the normal AP set based on our multi-autoencoder model. Specifically, our dataset resampling mechanism trains the model with three steps:

- Step 1: Randomly select around 10 million APs over the country to construct a primary fingerprint dataset to train the multi-autoencoder model;
- Step 2: Compute the Smooth L1 Loss to score and rank all training samples, and select the half with lower abnormal scores as the new dataset to re-train the model;
- Step 3: Repeat Step 2 for several times until all APs in dataset are normal.

This is based on the industrial observation that most of APs are normal, thus the remaining dataset with lowest abnormal scores contains very few abnormal APs. In our system, we repeat the dataset resampling process for 3 times (details are shown in Section 6.3).

Visualization. With the trained model, we compute an anomaly score for each access point. To visually explain the model's effectiveness, we select one normal AP and three typical abnormal APs, and visualize their feature vectors before/after our model. Figure 6 demonstrates that for a normal AP, there is little difference between the input and output. In contrast, for abnormal APs, the disparity between input and output is substantial.

Implementation. Our models are implemented in PyTorch. For training, we use the Adam [19] optimizer with a learning rate of 1e-3 and a batch size of 2048. Training the model for 200 epochs allows it to learn and adapt to the specific characteristics of our dataset, enhancing its predictive capabilities.

3.5 Anomaly Classification and Correction

There are multiple types of abnormal APs, e.g., mobile AP, migrant AP, and cloned AP, thus we need to identify their category and conduct corresponding corrections on associated fingerprints. For example, we should remove all fingerprints from mobile APs since it is unstable, and rename the BSSID of fingerprints from cloned APs. Such corrections require precise anomaly classifications.

3.5.1 Mobile APs. To obtain a variety of training samples for both mobile and fixed APs in a scalable manner, we adopt a name-filtering method to select confident samples among the crowdsourced fingerprints. The name-filtering method is based on an interesting observation that many mobile APs such as smartphones or vehicles share their cellular network thus the names of APs are by default associated with the device, e.g., "Alice's iPhone" or "Audi MMI". We can therefore use some key words like "iPhone" or "MMI" to filter out sufficient mobile AP samples for classification. In addition, many fixed wireless routers are also configured with SSIDs of their brand, e.g., "TP-Link-XX" for the world's No.1 wireless router provider. Thus we can gather sufficient mobile and fixed AP fingerprints to train a binary classifier.

Note that such name-filtering method could only help to produce the training dataset, but relying solely on SSID to identify mobile APs is unsuitable via crowdsourcing because of two reasons. First, a great many of users manually modify their smartphones' hotspot names based on personal preferences, thus it is difficult to classify them by only SSID. Second, due to lack of permissions, 19.8% of APs in our fingerprint set are scanned with blank SSIDs, thus those APs cannot be identified by SSID.

Based on sufficient fingerprints of mobile and fixed APs, we design a GBDT-based [11] binary classification model, aiming to overcome the limitations of SSID filtering. There are two steps in our method, i.e., feature extraction and model training.

- **Feature extraction.** We extract time-related, request-related, and signature-related features from the crowdsourced fingerprints. We compute both maximum, minimum, mean, and standard deviation values on each feature.
- **Model training.** We implement a GBDT-based binary classification model, using lightgbm [18] with the cross-entropy loss function.

Compared to solely SSIDs, our binary classification method ensures both sufficient accuracy and higher recall to classify mobile APs (detailed experiments in Section 6.3). In addition, all fingerprints from mobile APs are removed from the fingerprint set for localization.

3.5.2 Migrant and Cloned APs. A significant difference between migrant and cloned APs is whether only one AP or multiple APs with the same BSSID are active. Since we update the crowdsourced dataset everyday, we leverage the intraday localization requests to classify migrant and cloned APs. The reason why we use requests instead of signatures is that many migrant and cloned APs are deployed inside the building without any GPS receptions, but they can be scanned by indoor users when they initiate pickup services.

Without GPS locations, we innovatively explore "co-occurrent" APs, i.e., APs that are simultaneously scanned on the WiFi list when conducting localization requests, as the critical feature to classify migrant and cloned APs. Thus for each BSSID-indexed AP, we produce its "co-occurrent" APs with the intraday pickup queries, estimate the center geographical coordinates of each "co-occurrent" AP, and implement a DBSCAN [10] method to identify the exact number of geographical clusters.

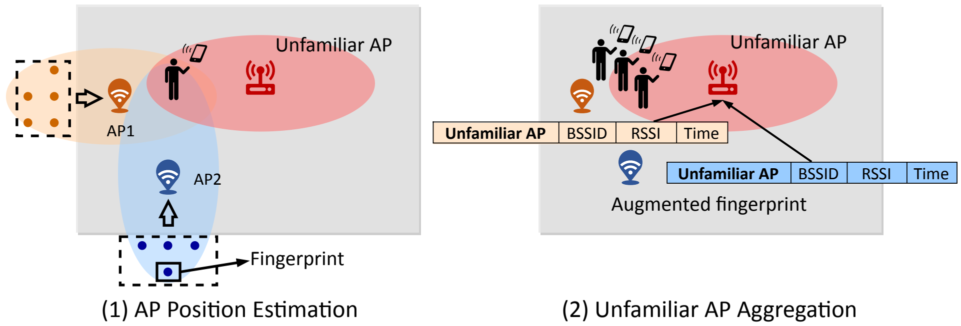


Fig. 7. **Augmented fingerprint construction process.**

In addition, we further correct associated fingerprints from migrant and cloned APs. In case there is only one cluster but the AP is identified as abnormal, it is regarded as a migrant AP thus we should remove its previous fingerprints in the set and only use the latest fingerprints for localization. In case there are two or more clusters, it is a cloned AP thus we should rename the BSSID in each cluster with a unique index.

4 Fingerprint Augmentation

Before localization, we observe that there are plenty of fingerprint "blackholes", particularly in large buildings where none of the APs in the WiFi scan list have associated fingerprints. In this situation, WiFi localization fails thus users must rely on cellular signals for coarse localization which may cause extreme errors or even incorrect buildings.

In this section, we aim to resolve localization "blackholes" by constructing fingerprints for those APs with none or sparse fingerprints as shown in Figure 7. It consists of two steps: 1) estimate locations of APs which already have fingerprints, and 2) iteratively infer fingerprints of unfamiliar but "co-occurrent" APs which simultaneously appear with known APs in WiFi scan list. Below we explain the process in details.

4.1 AP Position Estimation

In [29], Tencent proposed a name-matching method to estimate APs' positions based on APs' names and PoIs on floorplan. However, this method encounters several limitations during our deployment. First, not all buildings have released their floorplans with room-level PoIs for the public, especially in official or residential buildings. It is effort-intensive and cost-expensive for manual collection, as well as involves privacy concerns. Second, even for those shopping malls with floorplans, a large store always deploy multiple APs to supply network access in wide areas, whereas users are all attached to the same PoI. In addition, the name matching is usually complicated with multiple languages.

Hence, we explore a crowdsourcing approach to estimate the locations of APs which have fingerprints. Specifically, we use the log-distance path loss (LDPL) model [8, 22, 24] that is widely adopted to model WiFi signal propagation, i.e.,

$$s = P_0 - 10\gamma \log d \tag{3}$$

where P_0 and γ are the reference power and path loss constant of a given AP, respectively. d is the distance between the AP and a position, and s is the received signal strength at that point.

To estimate the location of each AP with its fingerprints, we implement a Maximum Likelihood Estimation (MLE) method as follows:

- **Candidate region production.** We produce a candidate area around an AP's fingerprints, and discretize this area into $10m \times 10m$ grid cells.
- **LDPL parameters estimation.** For each candidate grid i, we estimate corresponding parameters in Equation 3, i.e., P_0, and γ. Specifically, we utilize a least-squares algorithm:

$$\theta = \begin{bmatrix} P_0 \\ -10\gamma \end{bmatrix}, D = \begin{bmatrix} 1 & d_{i1} \\ \vdots & \vdots \\ 1 & d_{ij} \\ \vdots & \vdots \\ 1 & d_{iJ} \end{bmatrix}, S = \begin{bmatrix} s_1 \\ \vdots \\ s_j \\ \vdots \\ s_J \end{bmatrix} \tag{4}$$

$$\hat{\theta} = \begin{bmatrix} \hat{P}_0 \\ -10\hat{\gamma} \end{bmatrix} = (D^T D)^{-1} D^T S \tag{5}$$

  where j is a fingerprint ($j = 1, \cdots, J$) of this AP, d_{ij} represents the distance between the i-th candidate grid and the j-th fingerprint, and s_j is the received signal strength of this fingerprint. Finally, we estimate P_0 and γ via Equation 5. Note that, if $D^T D$ is not invertible, we use the gradient descent to calculate $\hat{\theta}$.
- **MSE calculation.** According to the estimated parameters $\hat{\theta}_i$ and its observation S, we compute the Mean Squared Error (MSE) Δ as:

$$\Delta = \frac{1}{J} tr\{(D\hat{\theta} - S)(D\hat{\theta} - S)^T\} \tag{6}$$

  Finally, we traverse all candidate grids and find the one with minimum MSE as the position of this AP.

4.2 Unfamiliar AP Aggregation

For those unfamiliar APs without any fingerprints, we iteratively produce their indoor fingerprints based on the known APs. Specifically, in case an unfamiliar AP co-occurrently appears with a known AP in some WiFi scanning lists in localization requests, a fingerprint item of this non-fingerprint AP is generated as:

- **Position:** the coordinates of this fingerprint is denoted as the location of this co-occurrent AP nearby the unfamiliar AP;
- **Received Signal Strength Indicator (RSSI):** the RSSI of this fingerprint is computed as the average value of all signatures from this co-occurrent AP, and we further calculate its standard deviation value based on multiple observations;
- **Scanning amount:** the scanning amount of this fingerprint is computed as the number of their co-occurrent appearances in requests;
- **Service time:** the service time of this fingerprint is recorded as the duration between the first and latest time of their co-occurrent appearances in the WiFi scan list.

4.3 Sparse Fingerprint Augmentation

Besides non-fingerprint APs, some APs have very sparse fingerprints that also cause severe reduction of localization accuracy. For example, if the number of an AP's fingerprints is less than 3, the median localization errors with this AP is significantly higher than other APs, which further increases the pickup position errors on real-world ride-hailing platforms (details in Section 6.5).

To enhance the localization accuracy and reliability of using sparse fingerprints, we leverage their co-occurrent APs to enrich their fingerprints like non-fingerprint AP augmentation. The only difference is that we ignore a co-occurrent AP if its estimated position is faraway (over one hundred meters in our system) from the few existing fingerprints of the related AP.

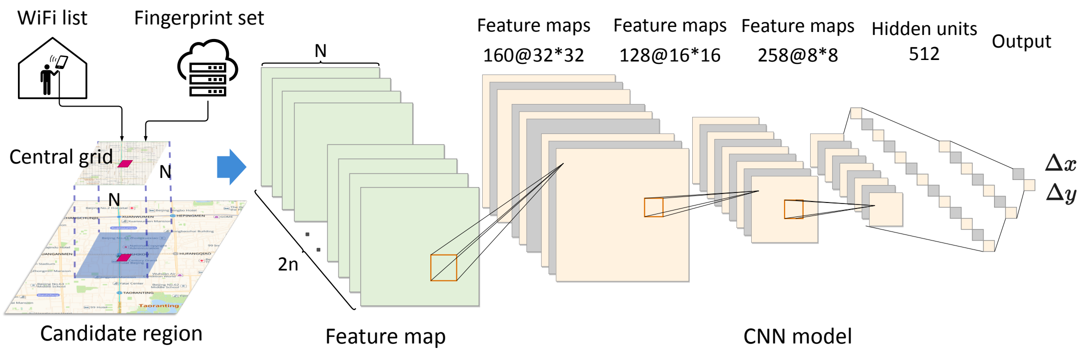


Fig. 8. **Localization process.** It consists of candidate region production, feature map generation, and CNN-based localization.

5 Indoor Localization

In this section, we locate indoor users based on anomaly-filtered and AP-augmented WiFi fingerprint set. Our industrial localization mechanism consists of three phases as shown in Figure 8, i.e., candidate region production, feature map generation, and CNN-based localization.

Candidate region production. Estimating an appropriate candidate region is crucial and efficient for localization. For each AP in the WiFi scan list when users initiate a localization request, we calculate the weight of each grid where the AP has fingerprints. Specifically, the weight includes two parts, i.e., RSSI similarity and scanning amount. For a grid i of an overheard AP with signal strength s, we define the weights as:

$$w_i^{RSSI} = e^{-\frac{(\mu_i - s)^2}{\sigma_i^2}}, w_i^{SA} = \frac{h_i - h_i^{min}}{h_i^{max} - h_i^{min}} \tag{7}$$

where μ_i and σ_i are the mean and standard deviation values of RSSI for that AP in grid i. h_i is the scanning amount on this grid, and h_i^{min} and h_i^{max} are the corresponding minimum and maximum values among all spanning grids. In order to derive the most probable grid that the user may appear, we sum up the weights over all overheard APs, i.e.,

$$w_i = \sum_{j=1}^{n}(w_{i,j}^{RSSI} + w_{i,j}^{SA}) \tag{8}$$

where n denotes the number of heard APs. Finally, we select the grid with the highest weight as the central grid. We further construct a square area composed of $N \times N$ grids around the central grid as the candidate region ($N = 64$ in our system).

Feature map generation. To derive user's relative position towards central grid, we construct a multi-dimensional feature map. For each grid in the candidate region, we compute two types of features for each overheard AP–signature similarity and scanning amount. The former feature is calculated according to the left formula of Equation 7. The other feature, scanning amount, is the scan count of an overheard AP in a grid. Next, for each overheard AP, we calculate the features of each grid in the candidate region and further aggregate them to construct feature maps with the size of $N \times N$ from the above two aspects, respectively. Finally, we stack $2n$ feature maps among n heard APs and normalize them as the input of CNN model.

CNN-based localization. We propose a CNN model with the outlines of buildings collected from satellite images to infer user's relative position $(\Delta x, \Delta y)$ towards the central grid. As Figure 8 shows, it is comprised of three convolutional layers, three max-pooling layers, and two fully connected layers. The loss function implies the distance between the predicted location and its practical location. To train the model, we implement in PyTorch, use the Adam [19] optimizer with a learning rate of 5e-4 and a batch size of 512, and train it for 30,000 epochs. Thanks to the large-scale crowdsourced training dataset across different scenarios and devices, the CNN model shows the generalizability from our real-world experiments (Figure 18).

6 Large-Scale Evaluation

Fingerprint Set Construction: We crowdsourced the fingerprint set through DiDi, the most popular ride-hailing platform in China. Specifically, DiDi collects fingerprints on the road while drivers provide rides to passengers, enabling widespread data collection without requiring building access.
Experiment Overview: We conduct comprehensive and large-scale experiments to characterize the performance of our system, in order to answer the following questions:

- **Q1 (§ 6.2):** Can our system accurately perform indoor positioning using only fingerprints collected outdoors?
- **Q2 (§ 6.3 & § 6.4):** Does abnormal detection and classification enhance positioning accuracy?
- **Q3 (§ 6.5):** Can fingerprint augmentation improve positioning accuracy?
- **Q4 (§ 6.6):** Is our system scalable across different mobile models?
- **Q5 (§ 6.7):** How does our system compare with existing indoor positioning methods?

6.1 Experimental Setup

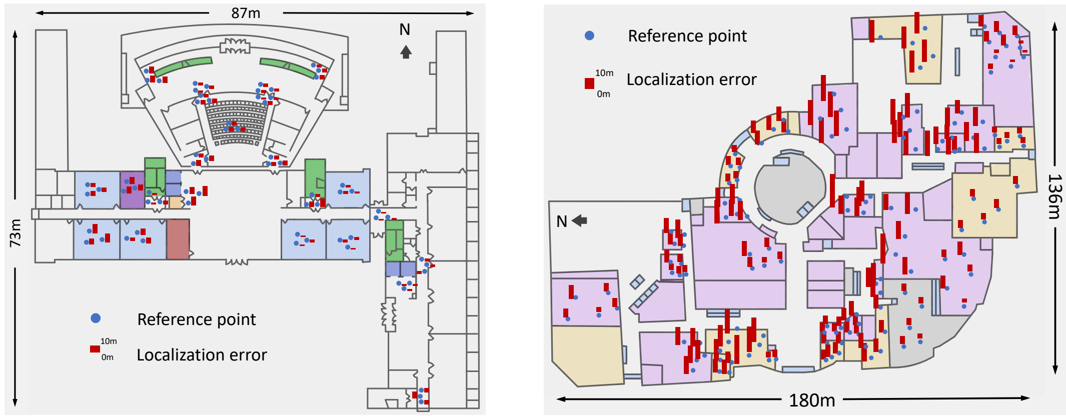
Dedicated experiments: testing indoor positioning accuracy. To evaluate the localization accuracy of our system, we conducted controlled experiments using seven different devices at 177 reference points across two locations: a college building and a shopping mall. These dedicated tests aimed to assess the system's ability to perform accurate indoor positioning. We use the positioning error, defined as the distance between the ground truth and the estimated position, to measure the accuracy of the indoor positioning.
Large-scale experiments: testing system performance with pickup service. In addition to the dedicated experiments, we evaluated our system's performance in real-world, large-scale environments. These large-scale experiments were conducted on DiDi. We selected approximately 122 million ride-hailing orders across 13 million devices that utilized WiFi-based pickup services over the span of two years in multiple cities. These large-scale experiments aimed to assess the robustness and scalability of our methods, particularly focusing on abnormal AP detection and classification, as well as fingerprint augmentation.

Since obtaining the ground truth of each user's position in large-scale settings is challenging, we used an industry-relevant metric called the "pickup position error." This metric measures the distance between the recommended pickup location and the actual pickup location. We recommend pickup position according to the indoor position estimated by our system beside the road that is closed to the users' location, which is an optimal place for users boarding the vehicle. Therefore, we can recommend more suitable pickup location in case we provide more accurate positioning. As a result, lower pickup position error indicates better positioning accuracy of our system. The results demonstrate that our system is both practical and effective, improving the real-world business experience for users.

6.2 Localization Accuracy

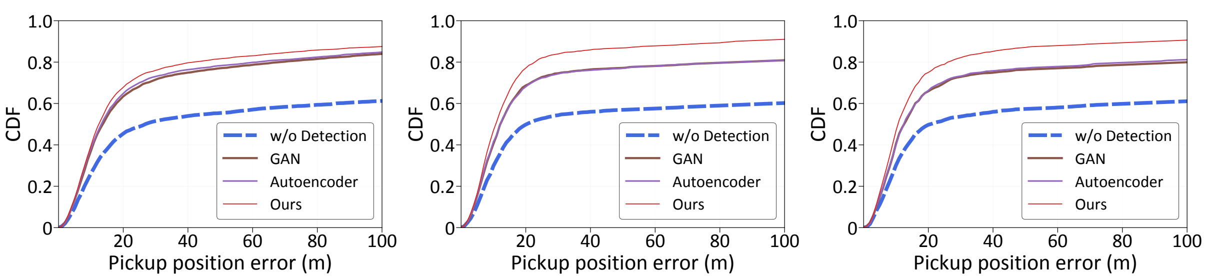
Localization in a college building. As shown in Figure 9a, we evaluated our system's positioning accuracy at 69 reference points in a college building. For each point, approximately 100 tests were conducted using seven smartphones from four different brands. The results reveal that points near the roads achieved higher accuracy, with a median localization error of 9.04m, compared to inner points, which exhibited a median error of 12.49m.
Localization in a shopping mall[2]**.** As demonstrated in Figure 9b, we selected 108 reference points within a shopping mall covering an area of 17,482 m^2. Similar to the college building tests, around 100 trials were performed at each point using seven smartphones from four brands. The results show that reference points

[2]Partial dataset: https://juderer.github.io/wifiout2in



(a) Localization performance in a college building. (b) Localization performance in a shopping mall.

Fig. 9. **The floor plans of a college building and a shopping mall, in total 177 reference points.** The bar shows the error in estimated coordinates.



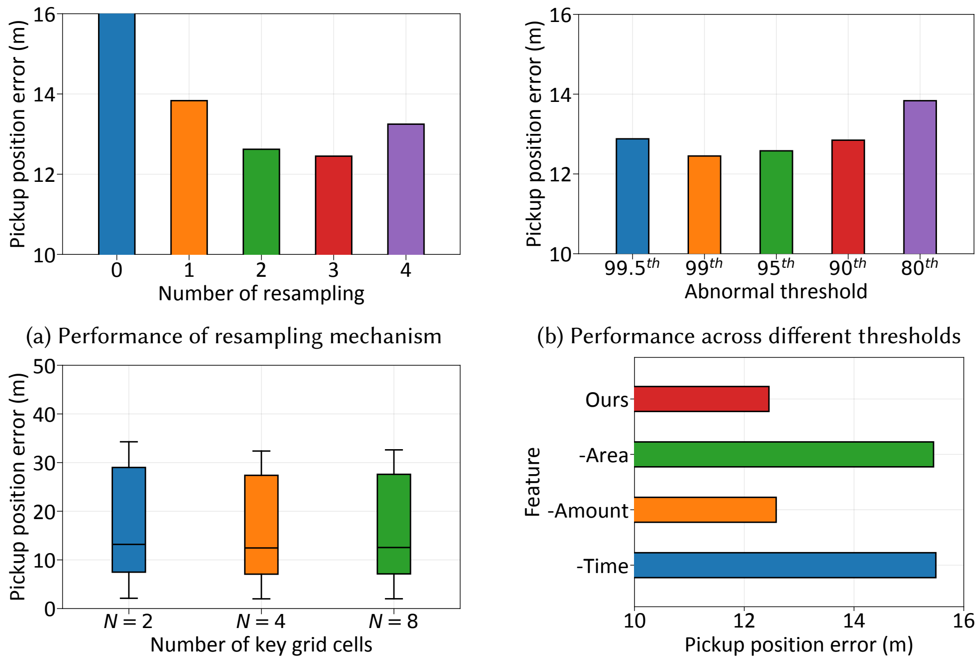
(a) A large city (b) A medium city (c) A small city

Fig. 10. **Performance across alternatives of anomaly detection in three cities.**

near the roads achieved a localization error of approximately 13.94m, a significant reduction of 17.32m (55.41%) compared to the points located deeper inside the mall.

It is worth noting that some points near the roads had larger localization errors, likely due to fewer available fingerprints in these areas. For example, as there exist trees around the college building corner, we cannot collect fingerprints with high localization accuracy. Additionally, several intermediate points displayed acceptable accuracy, benefiting from sufficient fingerprint coverage on both sides of their location.

Summary. These results indicate that our system achieves precise indoor positioning accuracy in practice, leveraging only crowdsourced outdoor fingerprints, which provide a strong foundation for pickup position recommendation services.



(a) Performance of resampling mechanism

(b) Performance across different thresholds

(c) Performance with different number of key grid cells

(d) Ablation study

Fig. 11. **Performance of hyper-parameters of multi-autoencoder network.**

6.3 Impact of Abnormal AP Detection

Compared with alternatives. To prove that the proposed multi-autoencoder network is a better choice for abnormal WiFi AP detection in our system, we calculate the metrics compared with alternatives, i.e., autoencoder [15, 37] and Generative Adversarial Network (GAN) [2, 3], in three typical cities with different scales of population.

As shown in Figure 10, the designed multi-autoencoder network outperforms auto-encoder network and GAN in terms of pickup position error. Specifically, autoencoder and GAN present similar performance. Compared to the autoencoder, our method has significant reductions of $0.74m$, $1.35m$, and $1.85m$ in terms of median errors in a large city, a medium city, and a small city, respectively. Additionally, the pickup errors increase by $14.35m$, $9.50m$, and $10.00m$ without detection in three cities, respectively, which indicates the necessity of abnormal detection.

Resampling. Since we have no idea how many abnormal APs there are, and we need to prepare a dataset containing as many normal APs as possible for training a multi-autoencoder network, we propose the resampling mechanism. To verify if resampling is helpful to find more normal APs, we calucate the metrics with different resampling times.

As shown in Figure 11a, the median pickup position errors decrease first and then increase, and reach the optimal value at the third resampling operation. Therefore, we prefer to resample for three times.

Abnormal threshold. To find out the best suitable abnormal threshold of the multi-autoencoder model to detect abnormal APs, we compare the metrics at five different abnormal thresholds, i.e., the 99.5^{th}, 99^{th}, 95^{th}, 90^{th}, and

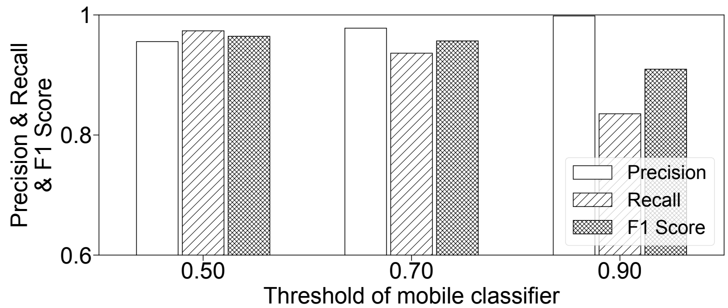
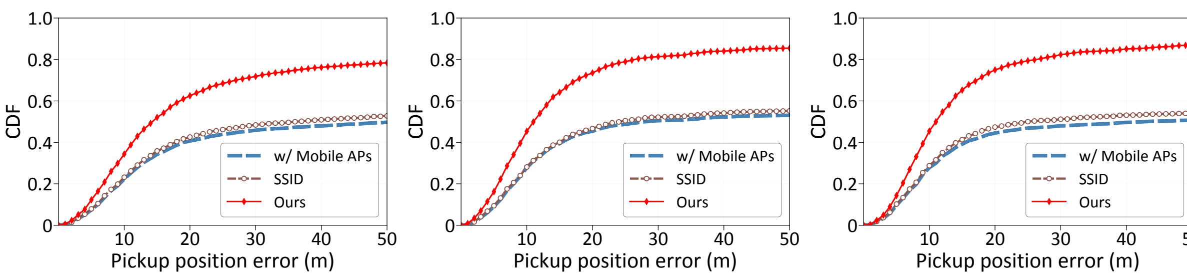


Fig. 12. **F1 score of mobile binary classifier.**



(a) A large city (b) A medium city (c) A small city

Fig. 13. **Performance of mobile AP classification in three cities.**

80^{th} percentiles, respectively. The results in Figure 11b show that the pickup position errors decrease first and then increase, and reach the optimal value when we choose the 99^{th} percentile as the abnormal threshold.

Number of key grid cells. To search the proper number of key grid cells as mentioned in Section 3.3, we conducted experiments where N was set to 2, 4, and 8, respectively. As depicted in Figure 11c, the pickup position errors decrease first and then stabilize. Therefore, we prefer to choose $N = 4$, ensuring both accuracy and complexity.

Ablation study. To evaluate how much the addition of the three parts, i.e., coverage area (area), scanning amount (amount), and service time (time), actually helps, we perform an ablation study on each feature. Figure 11d demonstrates that using all the features provides the best performance in terms of the pickup position error.

6.4 Impact of Abnormal Classification and Correction

Effect of mobile access point correction. As a mobile binary classification model needs a threshold ρ to help to identify a mobile AP, we evaluated the model using F1 score to find the optimal threshold. As shown in Figure 12, when $\rho = 0.50$, the binary model provides the highest recall and F1 score, while the model reaches the highest precision if the $\rho = 0.90$. To balance all metrics, we prefer to choose $\rho = 0.50$.

Then, we assess experiments in three cities using $\rho = 0.50$ to evaluate the performance of the trained binary classification model. The results in Figure 13 show that such mobile APs result in unacceptable localization errors due to their coarse-grained fingerprints. Using SSIDs to identify mobile APs has the limitation to improve the performance, as it identifies a few mobile APs with fixed SSIDs. On the contrary, our method outperforms and

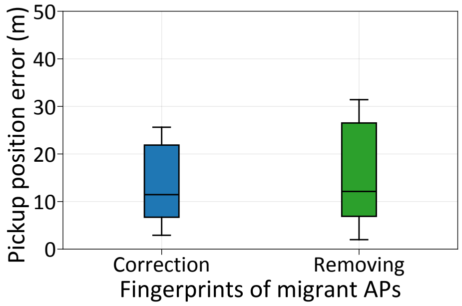


Fig. 14. **Performance of migrant AP correction.**

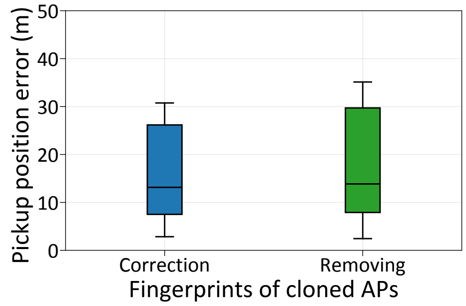


Fig. 15. **Performance of cloned AP correction.**

achieves remarkable reductions of 22.20m (61.0%), 13.32m (54.7%), and 14.10m (56.1%) compared to using SSIDs in terms of median pickup position error, thanks to its generalizability.

Effect of migrant access point correction. To prove the precision of identifying migrant APs, we performed experiments to compare the performance between our method, i.e, only retaining the fingerprints near the new location, and filtering the migrant APs directly during the localization phase. As shown in Figure 14, our method performs in terms of the first and the third quartile of the pickup position errors compared to without migrant APs, and the median error decreases 0.58m (4.6%). The results indicate that our identification and process of migrant APs achieves a better performance.

Effect of cloned access point correction. Similar to evaluation of migrant APs, we also compared our method, i.e., breaking down a cloned AP into multiple smaller aggregations of its fingerprints, versus removing cloned APs directly during localization phase. The results in Figure 15 show that our method performs better. Specifically, the median pickup position error of our method is 0.88m (6.3%) lower than deleting cloned APs.

Summary. These results in three cities present that the median pickup errors of our system reduce by at least 9.50m after anomaly identification. For the classification and correction of three typical abnormal APs, our system further achieves significant reductions in terms of pickup position error compared to only filtering abnormal APs. Therefore, anomaly identification and abnormal classification and correction effectively enhance positioning accuracy.

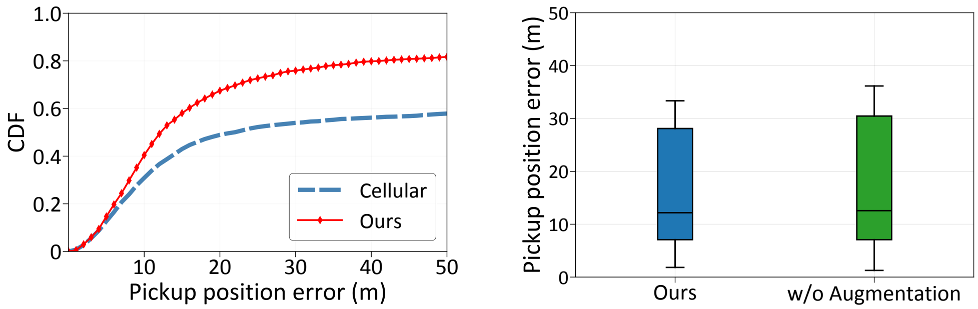
6.5 Impact of Augmentation

Comparison to cellular. Based on our large-scale observations, we could not provide WiFi localization service for approximately 0.6% of daily orders due to the absence of fingerprints from the scanned APs. Such orders rely solely on cellular localization without WiFi fingerprint augmentation. To ensure the capacity of our augmentation method, we compared the metrics between using augmented fingerprints and using cellular signals.

As shown in Figure 16a, our method achieves a 9.67m (44.3%) lower than cellular localization with regard to pickup position error, which suggests that the augmented fingerprints possess the ability to provide localization with acceptable and sufficient accuracy.

Comparison to real fingerprints. Additionally, there exist around 9.8% of daily orders where the scanned APs include one or more non-fingerprint APs. To compare the accuracy of augmented and real fingerprints, we perform experiments as show in Figure 16b.

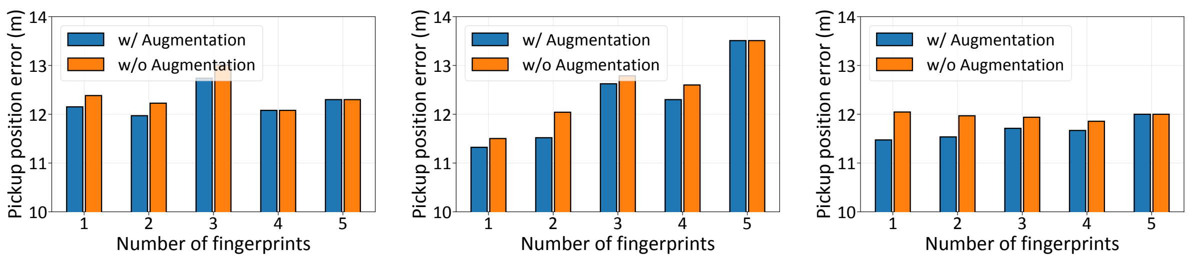
The results show that with augmentation our system achieves a 0.39m (3.1%) lower pickup position error than without augmentation. It indicates that augmented fingerprints can provide localization with accuracy that is no worse than real fingerprints.



(a) Comparison to cellular localization

(b) Comparison to real fingerprints

Fig. 16. **Performance of augmentation for non-fingerprint APs.**



(a) A large city

(b) A medium city

(c) A small city

Fig. 17. **Performance of augmentation for sparse APs.**

Augmentation for sparse APs. Since sparse APs have few fingerprints, they may provide positioning with huge uncertainty. To verify if our augmentation method helps such sparse APs to provide more reliable positioning, we calculate the metrics with augmentation of sparse APs in three cities, respectively.

As shown in Figure 17, our method achieves significant reductions in terms of pickup position error in three cities compared to without augmentation when the number of fingerprints of an AP does not excced 3, 4, and 4, respectively. Therefore, our augmentation method performs well for the orders with sparse APs.

Summary. Regarding these undetectable APs which cannot be scanned outdoors, our augmentation method achieves better positioning accuracy compared with cellular and non-augmentation localization. In terms of sparse APs with few fingerprints, using augmentation has a better performance of pickup position error. Hence, fingerprint augmentation significantly improves system's positioning accuracy.

6.6 Effect of Mainstream Phone Models

Since different models of smartphones will use different base-band chips, the WiFi signal quality, strength, etc., and further the fingerprint might be different among different models.

We selected twelve popular models from the top six Android mainstream brands, i.e., Huawei, Vivo, Oppo, Xiaomi, Honor, and Samsung, and measured the pickup position errors in real-world environments. Figure 18 shows that the smartphones with Vivo X20A and Oppo R11 models exhibit the best accuracy, possibly due to the rich fingerprints associated with these models. The results indicate our system is scalable across different smartphone models.

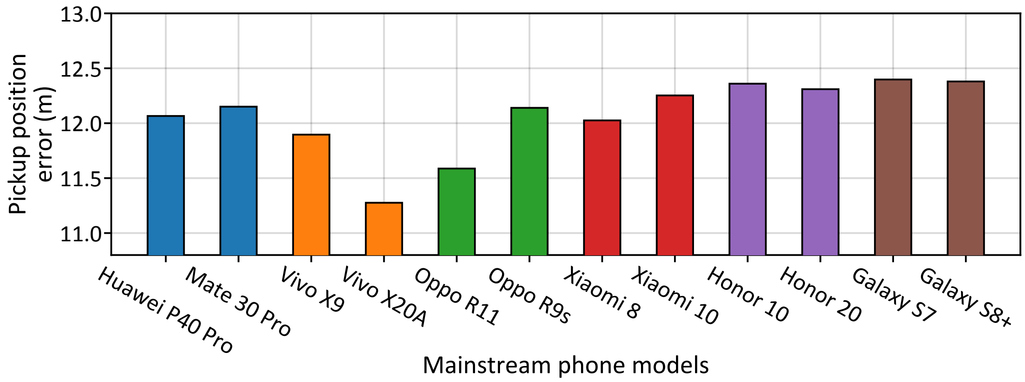


Fig. 18. **Performance across twelve mainstream phone models.**

6.7 Comparison with Existing Efforts

We compared the performance of our system's pickup service with Lab2Wild [29], another industrial product by Tencent. Our system constructs fingerprints from outdoor trajectories for indoor positioning, identifies abnormal APs, and augments non-fingerprint or sparse AP fingerprints. Lab2Wild is a lighter-weight solution based on two assumptions: 1) using sufficient PoIs for name-matching to estimate AP locations, and 2) leveraging a large volume of indoor user feedback to refine initial fingerprints.

In practice, however, these assumptions do not always hold. Due to permission restrictions, 19.8% of APs in our fingerprint set were scanned with blank SSIDs, limiting the ability to estimate their locations via SSID and PoI information. Furthermore, less than 10% of users provide feedback after their ride-hailing trips, reducing the effectiveness of user input for fingerprint improvement. In contrast, our system operates independently of both PoI data and user feedback, allowing it to function more reliably across diverse environments.

To evaluate performance, we conducted experiments in a large city's shopping mall, comparing pickup position errors between our system and Lab2Wild. Among one month and hundreds of ride-hailing orders, the median pickup position error for our system was 10.44m, which is 4.35m (or 29.4%) lower than Lab2Wild.

These results demonstrate that our system significantly outperforms Lab2Wild, showcasing its robustness in utilizing pervasive WiFi infrastructure without relying on PoI information or user feedback.

7 Related Work

Anomaly detection. One influential approach [34] for anomaly detection assumed that the GAN's latent vector represents the underlying data distribution. Another notable model, called GANomaly, was introduced in [32], aiming to learn the manifold of a large training dataset composed of normal samples and then detects abnormal images. Researchers have also explored combining GAN-based anomaly detection with fingerprint-based localization. For instance, researchers proposed RAD-GAN [2], an anomaly detection model for indoor fingerprint-based localization inspired by [32]. RAD-GAN relies solely on unlabeled fingerprints and achieves outstanding performance, surpassing methods such as OC-SVM [35].

WiFi fingerprint augmentation. To address challenges associated with non-uniform data density and environmental specifics, the researchers presented Modellet [8, 22, 28] algorithm, utilizing virtual fingerprints to fill data voids. In [38], Gaussian Process Regression (GPR) models were presented to augment the fingerprints in uncharted zone with limited known data. Later, SWSM [9] proposed a multivariate GPR model to estimate the joint distribution of all WiFi signals in a 3-D environment. Recently, [27] developed a minimal-sample WiFi fingerprinting localization strategy by leveraging few-shot learning to construct a robust RSSI-position correlation model for indoor localization.

WiFi fingerprint-based localization. The category of fingerprint-based approaches is pioneered by RADAR [5], and followed by numerous improvement algorithms [51, 52]. Initially, the researchers [6, 20, 21] calculated the distance between fingerprints or computed the likelihood of a query fingerprint against those in the database to locate users. With the widespread deployment of deep learning, more studies [1, 4, 31, 55] provided localization service utilizing networks, aiming to overcome traditional limitations (like multipath, occlusions, etc.).

Recent studies shared their experience in real-world environments. In [29], the authors presented their insights and experience utilizing pervasive third-party WiFi infrastructure to provide scalable localization service to millions of users. Yuming Hu et al. [16] reported the lessons from 1,170 teams' fingerprint-based indoor localization algorithms. However, these studies rely on indoor fingerprints, which is well known that indoor fingerprint collection is labor-intensive and time-consuming [40, 49]. In contrast, we propose a novel approach for indoor positioning, using outdoor fingerprints collected through crowdsourcing within a ride-hailing platform.

8 Conclusion

In this paper, we present our technical insights and development experience to enable large-scale indoor localization availability for pickup service at DiDi ride-hailing platform. We explore a new approach that leverages crowdsourced outdoor fingerprints to locate indoor users, and address many practical challenges we encountered during the real-world deployment. We hope this work can boost future attentions and industrial efforts on transforming the indoor localization research into commercial services. In future work, we will focus on leveraging transfer learning and user feedback to improve localization accuracy.

Acknowledgments

The authors would like to thank anonymous reviewers for their valuable comments and suggestions. This paper is supported in part by the Beijing Nova Program 20230484263 and 20240484607, Natural Science Foundation of China grant No. 62402028, and DiDi Research Collaboration Plan.

References

[1] Moustafa Abbas, Moustafa Elhamshary, Hamada Rizk, Marwan Torki, and Moustafa Youssef. 2019. WiDeep: WiFi-based accurate and robust indoor localization system using deep learning. In *Proceedings of the IEEE PerCom*. 1–10.

[2] Haojun Ai, Tan Hu, and Tianshui Xu. 2021. RAD-GAN: Radio map anomaly detection for fingerprint indoor positioning with GAN. In *Proceedings of the IEEE IPIN*. 1–8.

[3] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. 2019. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Proceedings of the Springer ACCV*. 622–637.

[4] Roshan Ayyalasomayajula, Aditya Arun, Chenfeng Wu, Sanatan Sharma, Abhishek Rajkumar Sethi, Deepak Vasisht, and Dinesh Bharadia. 2020. Deep learning based wireless localization for indoor navigation. In *Proceedings of the ACM MobiCom*. 17:1–17:14.

[5] Paramvir Bahl. 2000. RADAR: An in-building RF-based user location and tracking system. In *Proceedings of the IEEE INFOCOM*.

[6] Mike Y Chen, Timothy Sohn, Dmitri Chmelev, Dirk Haehnel, Jeffrey Hightower, Jeff Hughes, Anthony LaMarca, Fred Potter, Ian Smith, and Alex Varshavsky. 2006. Practical metropolitan-scale positioning for gsm phones. In *Proceedings of the ACM UbiComp*. 225–242.

[7] Linsong Cheng, Zhao Wang, Yunting Zhang, Weiyi Wang, Weimin Xu, and Jiliang Wang. 2020. Acouradar: Towards single source based acoustic localization. In *Proceedings of the IEEE INFOCOM*. 1848–1856.

[8] Krishna Chintalapudi, Anand Padmanabha Iyer, and Venkata N Padmanabhan. 2010. Indoor localization without the pain. In *Proceedings of the ACM MobiCom*. 173–184.

[9] Yinhuan Dong, Tughrul Arslan, Yunjie Yang, and Yingda Ma. 2022. A wifi fingerprint augmentation method for 3-d crowdsourced indoor positioning systems. In *Proceedings of the IEEE IPIN*. 1–8.

[10] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the KDD*, Vol. 96. 226–231.

[11] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* (2001), 1189–1232.

[12] Ross Girshick. 2015. Fast r-cnn. In *Proceedings of the IEEE ICCV*. 1440–1448.

[13] Baoshen Guo, Weijian Zuo, Shuai Wang, Wenjun Lyu, Zhiqing Hong, Yi Ding, Tian He, and Desheng Zhang. 2022. Wepos: Weak-supervised indoor positioning with unlabeled wifi for on-demand delivery. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 54:1–54:25.

[14] Fredrik Gustafsson and Fredrik Gunnarsson. 2003. Positioning using time-difference of arrival measurements. In *Proceedings of the IEEE ICASSP*. VI–553.

[15] Geoffrey E Hinton and Ruslan R Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *science* 313, 5786 (2006), 504–507.

[16] Yuming Hu, Xiubin Fan, Zhimeng Yin, Feng Qian, Zhe Ji, Yuanchao Shu, Yeqiang Han, Qiang Xu, Jie Liu, and Paramvir Bahl. 2023. The Wisdom of 1,170 Teams: Lessons and Experiences from a Large Indoor Localization Competition. In *Proceedings of the ACM MobiCom*. 1–15.

[17] Yuming Hu, Feng Qian, Zhimeng Yin, Zhenhua Li, Zhe Ji, Yeqiang Han, Qiang Xu, and Wei Jiang. 2022. Experience: Practical indoor localization for malls. In *Proceedings of the ACM MobiCom*. 82–93.

[18] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. LightGBM: a highly efficient gradient boosting decision tree. In *Proceedings of the NeurIPS*. 3149–3157.

[19] Diederik P Kingma. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[20] Parameshwaran Krishnan, AS Krishnakumar, Wen-Hua Ju, Colin Mallows, and Sachin Gamt. 2004. A system for LEASE: Location estimation assisted by stationary emitters for indoor RF wireless networks. In *Proceedings of the IEEE INFOCOM*. 1001–1011.

[21] John Krumm and John Platt. 2003. Minimizing calibration effort for an indoor 802.11 device location measurement system. *Microsoft Research* 8 (2003).

[22] Liqun Li, Guobin Shen, Chunshui Zhao, Thomas Moscibroda, Jyh-Han Lin, and Feng Zhao. 2014. Experiencing and handling the diversity in data density and environmental locality in an indoor positioning service. In *Proceedings of the ACM MobiCom*. 459–470.

[23] Lingkun Li, Pengjin Xie, and Jiliang Wang. 2018. Rainbowlight: Towards low cost ambient light positioning with mobile phones. In *Proceedings of the ACM MobiCom*. 445–457.

[24] Hyuk Lim, Lu-Chuan Kung, Jennifer C. Hou, and Haiyun Luo. 2006. Zero-Configuration, Robust Indoor Localization: Theory and Experimentation. In *Proceedings of the IEEE INFOCOM*. 1–12.

[25] Qiongzheng Lin, Zhenlin An, and Lei Yang. 2019. Rebooting ultrasonic positioning systems for ultrasound-incapable smart devices. In *Proceedings of the ACM MobiCom*. 1–16.

[26] Manni Liu, Jialuo Du, Qing Zhou, Zhichao Cao, and Yunhao Liu. 2020. Eyeloc: Smartphone vision-enabled plug-n-play indoor localization in large shopping malls. *IEEE Internet of Things Journal* 8, 7 (2020), 5585–5598.

[27] Zhenjie Ma and Ke Shi. 2023. Few-Shot Learning for WiFi Fingerprinting Indoor Positioning. *Sensors* 23, 20 (2023), 8458.

[28] Rajalakshmi Nandakumar, Krishna Kant Chintalapudi, and Venkata N Padmanabhan. 2012. Centaur: Locating devices in an office environment. In *Proceedings of the ACM MobiCom*. 281–292.

[29] Jiazhi Ni, Fusang Zhang, Jie Xiong, Qiang Huang, Zhaoxin Chang, Junqi Ma, BinBin Xie, Pengsen Wang, Guangyu Bian, Xin Li, et al. 2022. Experience: Pushing indoor localization from laboratory to the wild. In *Proceedings of the ACM MobiCom*. 147–157.

[30] Lionel M Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P Patil. 2003. LANDMARC: Indoor location sensing using active RFID. In *Proceedings of the IEEE PerCom*. 407–415.

[31] Sinno Jialin Pan, James T Kwok, Qiang Yang, and Jeffrey Junfeng Pan. 2007. Adaptive localization in a dynamic WiFi environment through multi-view learning. In *Proceedings of the AAAI*. 1108–1113.

[32] Pramuditha Perera, Ramesh Nallapati, and Bing Xiang. 2019. Ocgan: One-class novelty detection using gans with constrained latent representations. In *Proceedings of the IEEE CVPR*. 2898–2906.

[33] Kun Qian, Chenshu Wu, Zheng Yang, Yunhao Liu, and Kyle Jamieson. 2017. Widar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi. In *Proceedings of the ACM MobiHoc*. 6:1–6:10.

[34] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. 2017. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *Proceedings of the Springer IPMI*. 146–157.

[35] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. 2001. Estimating the support of a high-dimensional distribution. *Neural computation* 13, 7 (2001), 1443–1471.

[36] Longfei Shangguan, Zheng Yang, Alex X Liu, Zimu Zhou, and Yunhao Liu. 2015. Relative localization of RFID tags using Spatial-Temporal phase profiling. In *Proceedings of the USENIX NSDI*. 251–263.

[37] Zhihao Shen, Wan Du, Xi Zhao, and Jianhua Zou. 2020. DMM: Fast map matching for cellular data. In *Proceedings of the ACM MobiCom*. 799–812.

[38] Wei Sun, Min Xue, Hongshan Yu, Hongwei Tang, and Anping Lin. 2018. Augmentation of fingerprints for indoor WiFi localization based on Gaussian process regression. *IEEE Transactions on Vehicular Technology* 67, 11 (2018), 10896–10905.

[39] Xinyu Tong, Yang Wan, Qianru Li, Xiaohua Tian, and Xinbing Wang. 2020. CSI fingerprinting localization with low human efforts. *IEEE/ACM Transactions on Networking* 29, 1 (2020), 372–385.

[40] Bang Wang, Qiuyun Chen, Laurence T Yang, and Han-Chieh Chao. 2016. Indoor smartphone localization via fingerprint crowdsourcing: Challenges and approaches. *IEEE Wireless Communications* 23, 3 (2016), 82–89.
[41] Chia-Cheng Wang, Jyh-Cheng Chen, Yi Chen, Rui-Heng Tu, Jia-Jiun Lee, Yu-Xin Xiao, and Shan-Yu Cai. 2021. MVP: Magnetic vehicular positioning system for GNSS-denied environments. In *Proceedings of the ACM MobiCom*. 531–544.
[42] Jiankun Wang, Zenghua Zhao, Mengling Ou, Jiayang Cui, and Bin Wu. 2023. Automatic update for wi-fi fingerprinting indoor localization via multi-target domain adaptation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 2 (2023), 78:1–78:27.
[43] Xuyu Wang, Lingjun Gao, Shiwen Mao, and Santosh Pandey. 2016. CSI-based fingerprinting for indoor localization: A deep learning approach. *IEEE Transactions on Vehicular Technology* 66, 1 (2016), 763–776.
[44] Yu-Lin Wei, Chang-Jung Huang, Hsin-Mu Tsai, and Kate Ching-Ju Lin. 2017. Celli: Indoor positioning using polarized sweeping light beams. In *Proceedings of the ACM MobiSys*. 136–147.
[45] Chenshu Wu, Jingao Xu, Zheng Yang, Nicholas D Lane, and Zuwei Yin. 2017. Gain without pain: Accurate WiFi-based localization using fingerprint spatial gradient. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 29:1–29:19.
[46] Chenshu Wu, Feng Zhang, Yusen Fan, and KJ Ray Liu. 2019. RF-based inertial measurement. In *Proceedings of the ACM SIGCOMM*. 117–129.
[47] Hongwei Xie, Tao Gu, Xianping Tao, Haibo Ye, and Jian Lv. 2014. MaLoc: A practical magnetic fingerprinting approach to indoor localization using smartphones. In *Proceedings of the ACM UbiComp*. 243–253.
[48] Jie Xiong and Kyle Jamieson. 2013. ArrayTrack: A Fine-Grained indoor location system. In *Proceedings of the USENIX NSDI*. 71–84.
[49] Qiang Xu and Rong Zheng. 2016. MobiBee: A mobile treasure hunt game for location-dependent fingerprint collection. In *Proceedings of the ACM UbiComp*. 1472–1477.
[50] Lei Yang, Yekui Chen, Xiang-Yang Li, Chaowei Xiao, Mo Li, and Yunhao Liu. 2014. Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices. In *Proceedings of the ACM MobiCom*. 237–248.
[51] Moustafa Youssef and Ashok Agrawala. 2005. The Horus WLAN location determination system. In *Proceedings of the ACM MobiSys*. 205–218.
[52] Moustafa A Youssef, Ashok Agrawala, and A Udaya Shankar. 2003. WLAN location determination via clustering and probability distributions. In *Proceedings of the IEEE PerCom*. 143–150.
[53] Bowen Zhang, Houssem Sifaou, and Geoffrey Ye Li. 2023. Csi-fingerprinting indoor localization via attention-augmented residual convolutional neural network. *IEEE Transactions on Wireless Communications* 22, 8 (2023), 5583–5597.
[54] Yunting Zhang, Jiliang Wang, Weiyi Wang, Zhao Wang, and Yunhao Liu. 2018. Vernier: Accurate and fast acoustic motion tracking using mobile devices. In *Proceedings of the IEEE INFOCOM*. 1709–1717.
[55] Zhipeng Zhou, Jihong Yu, Zheng Yang, and Wei Gong. 2020. MobiFi: Fast deep-learning based localization using mobile WiFi. In *Proceedings of the IEEE GLOBECOM*. 1–6.