GC-Loc: A Graph Attention based Framework for Collaborative Indoor Localization using Infrastructure-free Signals

TAO HE, School of Computer Science and Engineering and GuangDong Province Key Laboratory of Information Security Technology, Sun Yat-sen University, China
QUN NIU, School of Artificial Intelligence, Sun Yat-sen University, China
NING LIU*, School of Computer Science and Engineering and GuangDong Province Key Laboratory of Information Security Technology, Sun Yat-sen University, China

Indoor localization techniques play a fundamental role in empowering plenty of indoor location-based services (LBS) and exhibit great social and commercial values. The widespread fingerprint-based indoor localization methods usually suffer from the low feature discriminability with discrete signal fingerprint or high time overhead for continuous signal fingerprint collection. To address this, we introduce the collaboration mechanism and propose a graph attention based collaborative indoor localization framework, termed *GC-Loc*, which provides another perspective for efficient indoor localization. GC-Loc utilizes multiple discrete signal fingerprints collected by several users as input for collaborative localization. Specifically, we first construct an adaptive graph representation to efficiently model the relationships among the collaborative fingerprints. Then taking state-of-the-art GAT model as basic unit, we design a deep network with the residual structure and the hierarchical attention mechanism to extract and aggregate the features from the constructed graph for collaborative localization. Finally, we further employ ensemble learning mechanism in GC-Loc and devise a location refinement strategy based on model consensus for enhancing the robustness of GC-Loc. We have conducted extensive experiments in three different trial sites, and the experimental results demonstrate the superiority of GC-Loc, outperforming the comparison schemes by a wide margin (reducing the mean localization error by more than 42%).

CCS Concepts: • **Networks** → **Location based services**.

Additional Key Words and Phrases: Collaborative Indoor Localization, Graph Neural Network, Geomagnetism

ACM Reference Format:
Tao He, Qun Niu, and Ning Liu. 2022. GC-Loc: A Graph Attention based Framework for Collaborative Indoor Localization using Infrastructure-free Signals. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 165 (December 2022), 27 pages. https://doi.org/10.1145/3569495

1 INTRODUCTION

In the wake of the highly urbanized development of human society, people usually spend most of their time in the indoor environment nowadays. Since the scale and the layout of indoor spaces become more and more enormous and intricate, the demands for accurate indoor locations are surging. And the past decade has witnessed the rapid

*Corresponding author.

Authors' addresses: Tao He, hetao23@mail2.sysu.edu.cn, School of Computer Science and Engineering and GuangDong Province Key Laboratory of Information Security Technology, Sun Yat-sen University, Guangzhou, China; Qun Niu, niuq3@mail.sysu.edu.cn, School of Artificial Intelligence, Sun Yat-sen University, Zhuhai, China; Ning Liu, liuning2@mail.sysu.edu.cn, School of Computer Science and Engineering and GuangDong Province Key Laboratory of Information Security Technology, Sun Yat-sen University, Guangzhou, China.

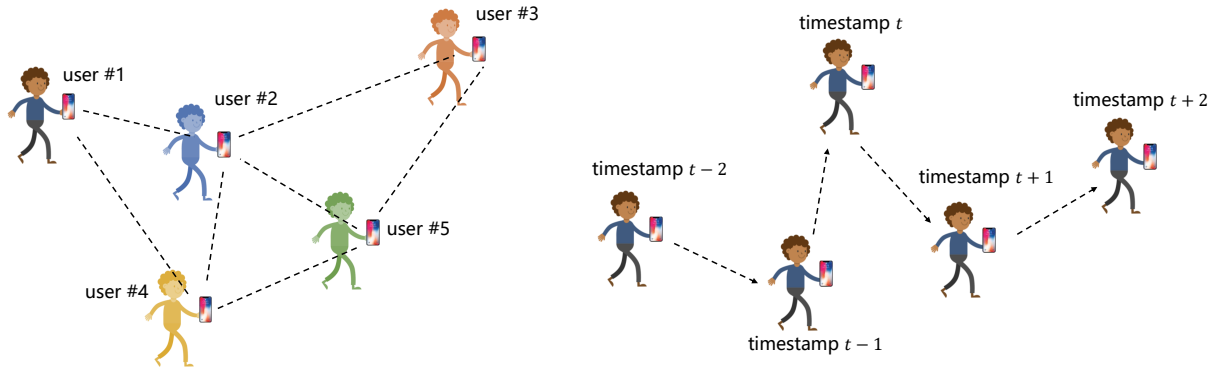
2474-9567/2022/12-ART165 $15.00
https://doi.org/10.1145/3569495

development and widespread application of indoor localization techniques in response to the urgent demands for various indoor location-based services (LBS), e.g., pedestrian or robot localization [3], on-demand delivery [7], and crowd monitoring [20], to name a few.

As the traditional satellite-based positioning systems (such as GPS) are prone to be ineffective indoors on account of the signal attenuation caused by poor connectivity between end devices and satellites, it triggers researchers to bend their energies to explore various ambient signals for indoor localization, such as Wi-Fi [23], BLE [11], sound [25], vision [9] and geomagnetism [27, 31]. From a system deployment and application point of view, all these signals explored for indoor localization can be roughly divided into two categories, i.e., signals which require extra infrastructures to be pre-deployed and infrastructure-free ones [52]. In the first category, these signals usually have specific propagation mode and are easy to handle and control, e.g., Wi-Fi, BLE and other electronic signals, as they are generated from manually pre-deployed infrastructures on the basis of the layout of indoor scenes. Although effective and easy to customize and control, it usually brings high deployment cost and maintenance expense especially in large indoor environments, which hinders its large-scale application in reality. In addition, these signals also fail to perform in some scenarios where the extra infrastructure deployment is not allowed for privacy or security considerations, like specific military field. For the second category, the signals without the requirements for extra infrastructure support are naturally low-cost and easy to deploy when applied for indoor localization. Taking the geomagnetic signal as an example, it shows great application prospects due to its omnipresence and no need for any extra infrastructures, since the geomagnetic signal mainly generates from natural earth's magnetic field [15]. The drawback is that these signals usually don't have specific propagation mode or distribution, e.g., collected magnetic signals usually have uncontrollable local variations due to the metamorphic ferromagnetism nature of indoor environments. So the crux is how to effectively establish the identifiable mapping relationships between these uncontrollable signal observations and spatial locations, facing complex and diverse indoor environments.

The underlying indoor positioning algorithms, on the other hand, are critical to make the best of these signals' characteristics in the indoor environment for effective localization. And the existing mainstream algorithms for indoor localization generally fall into two categories, i.e., *geometric-based* and *fingerprint-based*. For geometric-based approaches, the geometrical relationship of spatial location distribution is taken into consideration as localization clues, intuitively. To be specific, these approaches usually employ the relative measurements of the signals, e.g., the signal propagation delays, the arrival angle of the signal, through which they are able to obtain the relative distance and the direction of the target. Then the corresponding location can be calculated by applying geometric theorems. And the typical geometric-based algorithms [13] include adjacent direct localization, trilateration localization and triangulation localization, which usually utilize time of flight (TOF) [29], time of arrival (TOA) [53], time difference of arrival (TDOA) and angle of arrival (AOA) [6], to name a few. In practice, the performance of geometric-based approach mainly depends on the accuracy of these relative signal measurements, which are usually effected by manifold factors in indoor environment.

Different from the geometric-based schemes, the fingerprint-based approaches are mainly based on the pre-established fingerprint database and the fingerprint matching strategies, which are widely used in practice. And these approaches usually consist of two phases, i.e., offline stage for the fingerprint database establishment and matching strategy/model implement, online stage for the target location estimation based on the real time measurements. For fingerprint indicators, the received signal strength indication (RSSI) and the channel state information (CSI) [41] are widely adopted in existing schemes. In addition, the vision information [24] is also employed as a kind of fingerprint in some indoor scenes with rich textures, inspired by the success of deep learning in image process field recently. Despite these fingerprints have various different forms, the localization accuracy always relies on the spatial discriminability of the fingerprints. Most existing fingerprint-based approaches leverage either spatially discrete fingerprints (e.g., a Wi-Fi/Bluetooth RSSI, a geomagnetic measurement or an image at a fixed location) or temporally consecutive ones (e.g., a signal RSSI vector, a geomagnetic measurement



(a) Collaboration among multiple users

(b) Collaboration among multiple timestamps of single user

Fig. 1. The diagram of the two modes of collaboration: (a) For common indoor scenes, multiple users are available for collaboration. (b) For few scenes with only single user, the multiple timestamps of single user can be used for collaboration.

sequence or a video clip) for localization, which either suffers from large localization errors due to signal ambiguities or high respond time overhead with the sequence collection. Therefore, to achieve both accurate and efficient localization with the fingerprint-based approaches, it's supposed to improve the spatial discriminability of the fingerprints as much as possible for one thing, but on the other, the respond time overhead and map constraints also need to be taken into consideration for the efficient system applications. And it's still a challenge in practice.

Recently, with the rapid development of the techniques like Peer-to-Peer [54] and Crowdsourcing [11, 22], the concept of collaboration and sharing provides another perspective for efficient indoor localization. Actually, the most common indoor positioning application scenarios are usually some large places with high crowd density, such as the shopping mall, the hospital and the passenger station. And this provides a basis condition for the efficient collaborative indoor localization among multiple users. As discussed above, the signal fingerprints collected by the single user either lack of sufficient discriminabilty with discrete observations or restricted by long respond time overhead with long sequences. Employing collaborative mechanism as shown in Fig. 1 , fusing multiple discrete fingerprints from multiple spatially interconnected sources are able to provide much higher discriminability without the need to collect long signal sequence, meanwhile avoiding the high respond time and map constraints.

Inspired by the idea above, we propose a **G**raph attention based framework for **C**ollaborative indoor **Loc**alization, termed **GC-Loc**. Specifically, we first construct an adaptive graph representation to model the collaborative relationships among the signal fingerprints from multiple sources. Then we employ the optimized GAT (Graph Attention Network [45]) and further bring in the residual structure and hierarchical attention mechanism to realize efficient feature extraction from the constructed graph representation for collaborative localization, since the recent proposed GAT exhibits great ability to handle the multivariate non-Euclidean data structure. Moreover, we take advantage of ensemble learning mechanism and further devise a signal similarity based multi-model fusion strategy for the multiple models' joint location estimation based on model consensus, finally achieving accurate and robust collaborative indoor localization.

In a nutshell, we make the core contributions in this paper as follows.

- Different from traditional methods, we propose an efficient collaborative localization mechanism, taking full advantage of the collaborative information that available in most common indoor positioning scenarios

in practice. And the proposed collaborative localization framework can be easily extended or incorporated into most existing fingerprint-based indoor localization approaches.
- To facilitate the efficient collaboration, we construct an adaptive graph representation to model the collaborative relationships among the signal fingerprints from multiple users. And the constructed graph representation is used to achieve collaborative localization among multiple users. Besides, it also works well with multi-timestamps' collaboration for single user scenes, as illustrated in Fig.1.
- For the constructed graph representation, we employ state-of-the-art GAT model with the customized optimization for the location feature extraction. In addition, we bring in the residual structure and devise a hierarchical attention mechanism to further increase the expression ability of the proposed network, facing the complex and diverse indoor scenes.
- Considering the inevitable random environment noise and signal outliers in practice, we employ ensemble learning and devise a signal similarity based model fusion strategy for multiple models' joint location estimation. Through the model consensus, GC-Loc can overcome the problem that the individual model is prone to random errors, further enhancing the robustness.

As mentioned earlier, magnetic field signals don't require the extra infrastructure deployment and have suitable spatial distribution in indoor environment. Therefore, as an example, we utilize the magnetic field signal fingerprints as input to evaluate the proposed GC-Loc in this paper. Meanwhile, the relative Bluetooth signal and IMU sensor data are employed for collaborative relationship construction, and they are widely applied in mobile devices and they are also without the extra infrastructure deployment. So the proposed GC-Loc in this paper is able to achieve infrastructure-free collaborative indoor localization.

The remainder of this paper is organized as following. We review the related works in Section 2. The workflow of GC-Loc is presented in Section 3 and the detailed design of GC-Loc is elaborated in Section 4. Then we illustrate the experimental results in Section 5. Finally, we discuss several practical limitations in Section 6 and conclude in Section 7.

2 RELATED WORK

For the urgent demands of plentiful indoor LBS applications, indoor localization has been explored for decades. To realize sufficient accuracy and efficiency in complex and diverse indoor environments, researchers have proposed plenty of indoor localization methods and utilized various signals, such as widespread Wi-Fi [10], Bluetooth Low Energy (BLE) [11] and ultra-wide band (UWB) [12]. Although effective in some specific indoor scenes, the approaches based on these signals either suffer from the signal fluctuations caused by multi-path fading, or high system deployment cost and maintenance expense. And it hinders the large-scale application in reality. Besides these common electronic signals, the image/video-based approaches [5] have attracted much attention recently, with the rapid development and the success of image processing techniques. For example, HAIL [32] makes used of the image confirmation after taking images to realize a highly automated image-based localization algorithm. However, suffering from the motion blur or the loss of focus[58], the image-based indoor localization methods are usually prone to noise. And on the other hand, the construction of visual landmark database [30] is also labor-intensive in practice. Among all these explored signals, magnetic field signal exhibits good application prospect for indoor localization, since it is pervasive and relatively stable. More important is that magnetic field signal doesn't need the extra infrastructure deployment[16, 47], and it's almost immune to the pedestrian occlusion, compared with other signals, e.g., vision, acoustics.

Most existing indoor localization approaches rely on fingerprint techniques and can be broadly divided into two categories, i.e. discrete fingerprint-based methods and continuous sequence-based methods. In the first category, a discrete signal observation is usually taken as the reference of the corresponding location [56]. For example, SemanticSLAM [1] proposes to calibrate the current position with the cluster mechanism among collected

Table 1. Major symbols used in the paper.

Notation	Definition
$\mathbf{M}$	A magnetic field signal fingerprint
$\mathbf{g}$	A constructed adaptive graph representation
D	The pre-established geo-tagged magnetic signal fingerprint database
D_g	The dataset of the constructed graph for model training
R	The number of the constructed graph in D_g
T	The number of the trained models for ensemble learning
K	The number of the heads in multi-head mechanism
Q	The depth of the optimized deep GAT model (the number of layers)
$\mathbf{h_i}$	The embedding feature of the vertex i in the constructed graph
$\mathbf{l}$	A prediction/estimation location (x, y)
$\mathbf{L}$	A set of prediction/estimation locations for a graph
ζ	The loss function defined for the model training
ϵ	The overall mean localization error

magnetic signal observations. However, suffering from the limited discernibility, the discrete signal fingerprints are usually ambiguous, resulting in degraded distinctiveness of the spatial location features. And these lead to large localization errors consequently. To tackle this, some researchers consider the temporal correlations of consecutive signal observations, utilizing the signal sequence as input for localization [57]. For example, Travi-Navi [58] uses the magnetic sequence as input instead of discrete signal fingerprint, and further utilize the dynamic time warping (DTW) algorithm to match with the pre-established database for localization. In addition, some fusion-based methods take the magnetic sequence as a component of input for localization [33, 40]. The approaches in [21, 48] utilize the particle filter mechanism, in which the magnetic sequence is used for obtaining the observation location by matching. And WAIPO [14] directly fuses image with magnetic signal sequence for localization to enhance the accuracy. Recently, inspired by the success and popularization of deep learning techniques, many researchers propose to take advantage of deep neural network to explore the location clues for indoor localization [2, 17, 42, 46]. For example, DeepML [46] takes magnetic signal sequence as input and employs long short-term memory (LSTM) network to extract the features for localization. MAIL [31] further considers the different scales of the signal fluctuations in collected magnetic sequence. MAIL extracts and fuses the features under different scales to improve the feature discernibility, thus achieving higher accuracy. And ST-Loc [27] proposes to convert the magnetic sequence into different representations, then utilizes optimized networks to extract features under different dimensions for localization. Although efficient, most of these approaches usually need a long sequence as input to provide sufficient features for accurate localization, which will cause high time overhead for the long sequence collection and limit to the survey path constraints.

On the other hand, inspired by the concept of sharing mode, the collaboration and encounter based localization techniques have attracted much attention [36]. Considering the actual application scenarios that usually have high densities of smart devices, the peer assistance and relative information are taken into consideration for indoor localization [19, 22], instead of only relying on the location clues from individuals. Specifically, some researches [14, 28, 44] propose to bring in the ranging constraints among multiple users/devices while utilizing the collected signal fingerprints for indoor localization. For example, Centaur [28] fuses the Wi-Fi measurement and acoustic ranging techniques into a single systematic framework based on Bayesian inference. And WAIPO [14] realizes

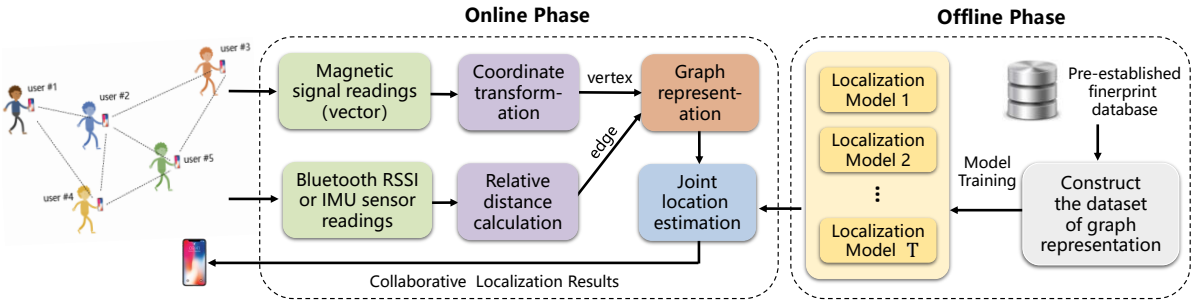


Fig. 2. The workflow of the proposed collaborative localization system.

collaborative indoor localization by fusing Wi-Fi signal measurements, magnetic fingerprints, image-matching and people co-occurrence. However, these approaches which utilize Wi-Fi, acoustic [26] or ZigBee [38] usually require extra infrastructure support, limiting the scale-up ability of the solutions. To facilitate the application in practice, some researchers propose to explore infrastructure-free approaches [8, 35, 49] for collaborative indoor localization. P^2-Loc [8] aims at on-demand delivery application and realizes a person-to-person localization mechanism which detects the encounter events through Bluetooth on couriers' smartphones, then infers couriers' relative locations. While P^2-Loc is infrastructure-free, it's strongly dependent on the encounter detection and assumes that the encounters are dense. Besides, it also requires the condition that the couriers usually take the shortest paths and there are no detours in between. CLIPS [35] also does not require pre-existing indoor infrastructures or extra deployments, and it employs dead reckoning techniques to filter out the invalid candidate coordinates, which suffers a lot from the cumulative errors in long-time continuous positioning. In this paper, the proposed GC-Loc first employs infrastructure-free signals (magnetic field fingerprints, relative Bluetooth RSSI or short IMU measurements) with no need for any infrastructure deployment. Then, with the constructed adaptive graph representation, GC-Loc can not only realize efficient collaborative localization among multiple users, but also work well in single user scenarios, thus achieving better expansibility and portability compared with the existing approaches that usually aim at a specific application or scenario, e.g., P^2-Loc [8] for on-demand delivery, CLIPS [35] for the emergency scenario. In addition, GC-Loc can be also easily extended or incorporated into most existing fingerprint-based indoor localization schemes.

3 SYSTEM WORKFLOW

In this section, we elaborate the workflow of the proposed GC-Loc. As illustrated in Fig. 2, the overall workflow contains two main stages, i.e. an offline stage and an online stage. In the offline stage, surveyors first carry a client device to collect the magnetic field signal fingerprints at the preset reference points according to the site floorplan. Then the collected magnetic signal fingerprints will be labeled with the corresponding location coordinates of the reference points. With labeled magnetic signal fingerprints, we construct the database of the proposed graph representation (Section 4.2), and the detailed process is presented in Algorithm 2. Based on the constructed database, we train multiple localization models (based on the design network in Section 4.3) which will be used in the online stage for joint location estimation.

In the online stage, we implement a collaborative system as elaborated in Section 5.1.1. And multiple users carry the clients and walk in the trial site. The installed application will collect the sensor data (including magnetic signal fingerprint, relative Bluetooth RSSI and IMU sensor readings). Then the collected data will be used to construct the adaptive graph representation (Section 4.2). And we take the constructed graph as input of the multiple models trained in the offline stage. Finally, employing ensemble learning mechanism and considering

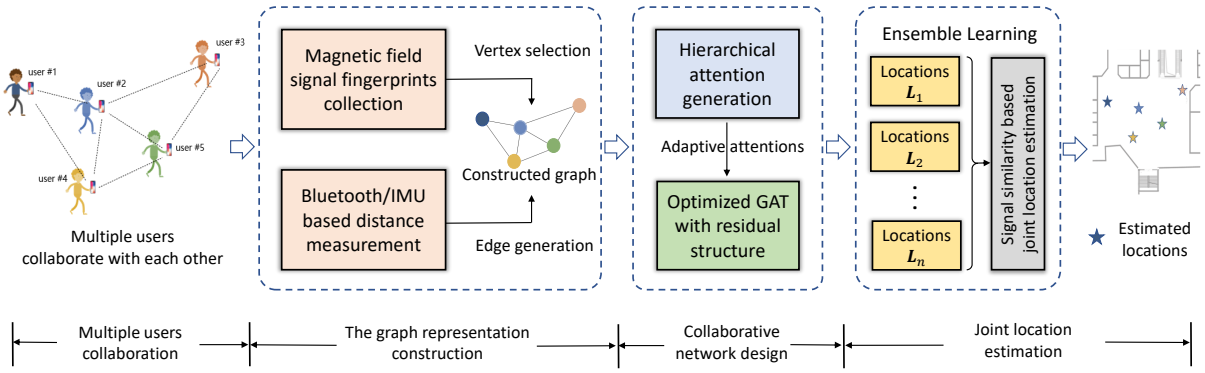


Fig. 3. Overall framework of GC-Loc.

the model consensus, a signal similarity based model fusion strategy will be used for joint location estimation, achieve final location refinement (Section 4.4).

4 DETAILED DESIGN

In this section, we elaborate the detailed design of proposed GC-Loc which takes advantage of graph representation to realize collaborative indoor localization between multiple users or multiple timestamp of single user. We first illustrate the overall structure of GC-Loc in Section 4.1. Then, we present the generation process of the graph representations for collaborative localization in Section 4.2, including both the graph representations for multi-user collaboration and self-collaboration of single user, respectively. And the detailed construction of GAT-based localization network is presented in Section 4.3. Finally, we elaborate the proposed location refinement strategy in Section 4.4.

Table 1 lists the major symbols and notations used in this paper.

4.1 Overview

Since the discrete signal fingerprints lack sufficient feature discernibility and the continuous signal sequences usually suffer from the high time overhead for signal sequence collection and path constraints, GC-Loc employs the collaboration mechanism to integrate multiple discrete signal fingerprints with relative relationships to achieve collaborative indoor localization. While GC-Loc takes discrete signal fingerprints as input, it avoids the high time overhead for signal sequence collection. Meanwhile, it fuses multiple discrete signal fingerprints (from multiple users/timestamps) to enhance the feature discernibility for accurate localization.

The overall framework of GC-Loc is illustrated in Fig. 3. Considering the most common application scenarios of indoor positioning and inspired by the collaboration mechanism, we first propose an adaptive graph representation to efficiently integrate multiple discrete signal fingerprints, including the fingerprints from multiple users or the fingerprints at multiple timestamps from single user. Then we take advantage of state-of-the-art Graph Attention Network (GAT) as a basic unit to devise a collaborative localization network for accurate location prediction. Furthermore, we bring in the residual structure and design a hierarchical attention mechanism to further improve the accuracy in complex indoor scenes. In addition, we take advantage of ensemble learning mechanism and devise a location refinement strategy based on model consensus, through which GC-Loc is able to effectively reduce the impact of random noise and signal outliers.

More specifically, proposed GC-Loc consists of three major parts as follows.

1) Adaptive graph representation construction. To facilitate the efficient collaborative localization, we first utilize multiple discrete signal fingerprints and their relative relationships to construct an adaptive graph representation, in which the vertexes indicate the discrete signal fingerprints and the edge denotes the distance between the corresponding two vertexes. Thereinto, the distance measurement (i.e. the edge generation in the graph) is based on the relative Bluetooth RSSI observations (for multiple users collaboration) or the IMU-based displacement estimation (for multiple timestamp collaboration of single user). And the details of the adaptive graph representation are elaborated in Section 4.2.

2) GAT-based collaborative indoor localization. As the proposed adaptive graph representation is actually Non-Euclidean structure data, we intuitively utilize state-of-the-art graph attention network (GAT) to extract the location clues for localization. Considering the complex and diverse indoor scenes, we construct a residual structure when applying multi-layer deep GAT model, which also effectively restrains the over smoothing problem. On this basis, we further devise a hierarchical attention mechanism, in which we take into consideration the different distance measurements between vertexes so as to fit the actual spatial characteristics of indoor localization. And we detail the design of proposed network in Section 4.3.

3) Location refinement with model consensus. To restrain the impact of random noise and signal outliers which are common in practical applications, we take advantage of ensemble learning mechanism and devise a location refinement strategy with the consensus among ensemble models. To be specific, we first train multiple independent localization models (based on the proposed localization network) with different settings. Then we backward consider the signal fingerprints matched with the models' predictions in geo-tagged signal database, based on which we further calculate the signal similarity as adaptive weights for the ensemble model fusion, achieving location refinement. And the details are presented in Section 4.4

4.2 Adaptive graph representation construction

As aforementioned, since the most common scenarios of indoor localization applications are usually with high crowd density, it makes the collaborative localization feasible. And the collaboration between multiple users provides another way to achieve accurate localization with the multiple discrete signal fingerprints, avoiding the high time overhead for long signal sequence collection. Moreover, even facing the indoor scenes where no multiple users available, the collaboration can be still applied among the multiple timestamps of one single user. To facilitate the efficient collaboration, we construct an adaptive graph representation which is not only applicable to the collaboration in the indoor scenes with multiple users, but also works well with multi-timestamps' collaboration for the single user scenes. The construction of proposed adaptive graph representation consists of the vertex selection and the edge generation as illustrated in Fig. 3.

For the vertex selection of the graph in this paper, we take the collected magnetic signal fingerprints as the vertexes of the graph, due to magnetic field's omnipresent distribution, high global stability over time and strong local signal variations indoors, providing the basic condition for accurate localization. To be specific, the magnetic field observation reading $\mathbf{M}_d$ measured by device's magnetometer is actually a combination of the natural geomagnetic field (high signal stability over time) and the magnetic field from the surrounding environment (distinguish local signal variations).

$$\mathbf{M}_d = \mathbf{M}_{geomagnetism} + \mathbf{M}_{environment}, \tag{1}$$

Meanwhile, the signal reading $\mathbf{M}_d = [m_x, m_y, m_z]^T$ usually consists of values at three axes (X, Y and Z) in the device coordinate system, which is directly related to the device holding pose. This means that the directly collected magnetic signal observations with different holding pose may be different from each other even in same location, which will lead to localization failure. To address this, we propose to transform the collected magnetic field vector $\mathbf{M}_d$ in device coordinate system to the vector in a uniform coordinate system based on the offset of the device. More specifically, we assume $\mathbf{M}_e$ denotes the magnetic field vector at same location in earth

coordinate system. Then $\mathbf{M}_d$ can be regarded as the transform by device rotated in roll ϕ, pitch θ and yaw ψ from $\mathbf{M}_e$. And the relationship between $\mathbf{M}_d$ and $\mathbf{M}_e$ can be defined as Equation 2 in absence of noise.

$$\mathbf{M}_d = \mathbf{R}_x(\phi)\mathbf{R}_y(\theta)\mathbf{R}_z(\psi)\mathbf{M}_e, \tag{2}$$

where $\mathbf{R}_x(\phi), \mathbf{R}_y(\theta)$ and $\mathbf{R}_z(\psi)$ are the rotation matrices. And the roll angle ϕ, pitch angle θ and yaw angle ψ can be obtained from the collected 3-axis acceleration sensor readings which denote as $\mathbf{A} = [a_x, b_y, c_z]^T$. And compared with the solution that uses single-axis data to calculate the angles, employing multi-axis data and calculating through inverse tangent can eliminate the influence of the deviation angle [40, 48]. Specifically, ϕ, θ and ψ are calculated as following.

$$\phi = arctan(\frac{a_y}{a_z}), \tag{3}$$

$$\theta = arctan(\frac{a_x}{a_z}), \tag{4}$$

$$\psi = arctan(\frac{a_y cos\phi - a_z sin\theta}{a_x cos\theta + a_y sin\theta sin\phi + a_z sin\theta cos\phi}). \tag{5}$$

With the calculated rotation angles, we can obtain the rotation matrices from the device coordinate system to the earth coordinate system, which is defined as following.

$$\mathbf{R}_x(\phi), \mathbf{R}_y(\theta), \mathbf{R}_z(\psi) = \begin{bmatrix} cos\psi & -sin\psi & 0 \\ sin\psi & cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1}, \begin{bmatrix} cos\theta & 0 & sin\theta \\ 0 & 1 & 0 \\ -sin\theta & 0 & cos\theta \end{bmatrix}^{-1}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\phi & -sin\phi \\ 0 & sin\phi & cos\phi \end{bmatrix}^{-1}, \tag{6}$$

Finally, according to the Equation 2, we can obtain the transformed magnetic field vector $\mathbf{M}_e$ in uniform earth coordinate system with the aid of the rotation matrices.

$$\mathbf{M}_e = [\widehat{m}_x, \widehat{m}_y, \widehat{m}_z]^T, \tag{7}$$

In addition, to further augment the discernibility of the magnetic signal, we also take the dimensionless magnitude of $\mathbf{M}_e$ into consideration. And the magnetic signal fingerprint we used in this paper is defined as the Equation 8.

$$\mathbf{M} = [\mathbf{M}_e, ||\mathbf{M}_e||_2]^T = [\widehat{m}_x, \widehat{m}_y, \widehat{m}_z, ||\mathbf{M}_e||_2]^T, \tag{8}$$

where $||\cdot||_2$ denotes L_2 norm operation.

For the edge generation of the graph, the traditional graph structure usually places emphasis on the connectivity judgment among the vertexes for message passing, and it usually has no fixed shape. However, the localization applications usually indicate a certain degree of spatial distribution characteristics, which is vital for collaborative localization, especially in the large open indoor scenes. Therefore, different from the traditional graph structure, we combine the spatial characteristics of the collaborative localization and take the relative distance measurement between the vertexes (collaborative users or timestamps) into consideration to generate the edges in proposed adaptive graph representation.

Specifically, although most indoor localization scenes are usually with high crowed density, there are still few application scenarios where are no multiple users available for collaboration in practice, as illustrate in Fig. 1. Therefore, we consider these two kinds of application scenarios respectively, and realize a uniform representation for two types of collaboration, i.e. multiple users collaboration and multiple timestamp collaboration (for single user scene), which reflect in the edge generation as following.

1) Multiple user scene. For the common scenes with multiple users available for direct collaboration, we take advantage of the collected relative Bluetooth RSSI between multiple users to estimate the relative distances.

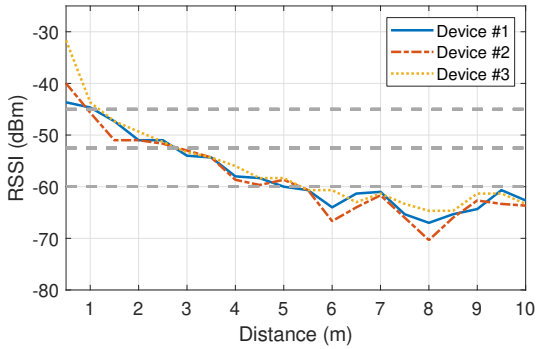


Fig. 4. The variation of Bluetooth RSSI with the relative distance using different devices in a spacious area.

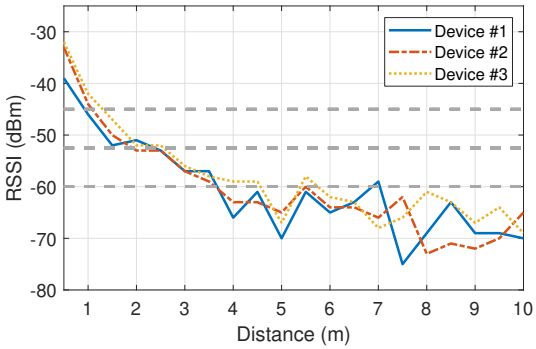


Fig. 5. The variation of Bluetooth RSSI with the relative distance using different devices in a narrow corridor.

Then we make use of the distance measurements for the edge generation. Considering the uncertainty of signal propagation, researchers usually employ the lognormal distribution to reflect the propagation of signal RSSI. And in this paper, the transformation of the collected Bluetooth RSSI versus the relative distance is defined as a logarithmic path loss model [50].

$$\eta_r(d) = \eta_0(d_0) - 10\lambda lg(\frac{d}{d_0}) + \delta, \tag{9}$$

where $\eta_r(d)$ denotes the RSSI value at distance d from the source, d_0 is the reference distance. $\eta_0(d_0)$ represents the collected RSSI value at distance d_0. And parameters λ, δ are the distance path loss exponent and random noise respectively, which both follow Gaussian distribution.

However, in the indoor environment, the collected Bluetooth RSSI usually suffers from the reflection, multi-path effect or shadowing phenomena in practice, which leads to unreliable observations and large errors for the distance estimation consequently. Therefore, directly utilizing the raw Bluetooth RSSI for distance estimation is prone to be unreliable. To address this, we devise a two-phase optimization method. Firstly, considering that the raw RSSI observation at a single timestamp usually fluctuates, we make use of a window of RSSI data (the observations in 2 seconds) instead of using raw single RSSI directly. On this basis, we further devise a fuzzy mapping strategy to realize the RSSI-distance transformation. Specifically, we first set a relatively low threshold to filter out the weak RSSI observation records, since the accuracy of the propagation model usually decreases with the growth of the relative distance between the devices and the discrimination of RSSI also decreases at large distance. As shown in Fig. 4 and Fig. 5, the RSSI observations fluctuate and fail to reflect the variations of the distance when the distance reaches above 5 m. Then instead of directly using the precise distance calculated based on the filtered RSSI, we further transform these obtained distances into the qualitative fuzzy classifications (dividing range interval), which is also more consistent with the standard of Bluetooth RSSI application. And we assign different fixed values for each category (divided range) according to the general standard of Bluetooth RSSI ranging. Finally, We calculate the distance d_f used for edge generation in this paper as following.

$$d_f = f(d) + \varphi, \tag{10}$$

$$f(d) = \begin{cases} v_0 & d \in (0, d_1], \\ v_1 & d \in (d_1, d_2], \\ v_2 & \text{otherwise}, \end{cases} \tag{11}$$

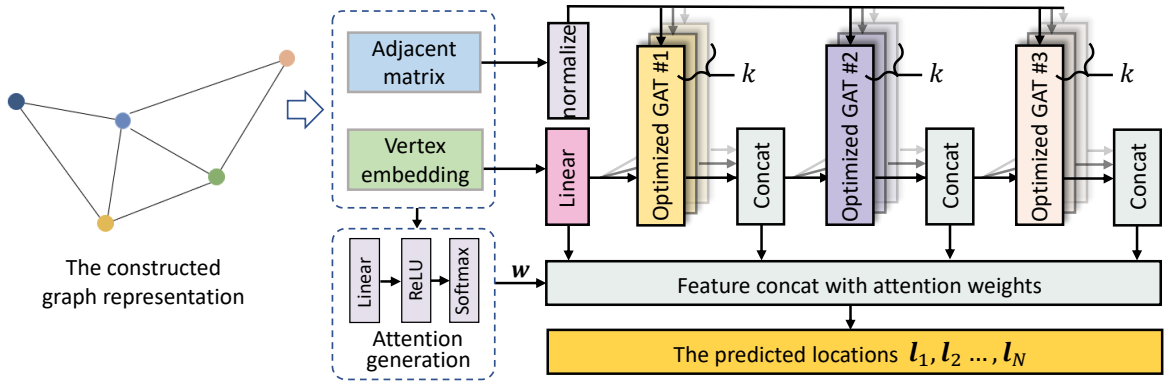


Fig. 6. The design of collaborative indoor localization network.

where d represents the distance obtained through Equation 9, φ is the random variable which follows Gaussian distribution with the mean of zero. And $f(\cdot)$ denotes a segmented fuzzy classification function, in which d_1, d_2 are preset reference distances.

As elaborated above, for multiple user scenes, we utilize the relative Bluetooth RSSI observations among multiple users to realize the distance measurement, then use which to generate the edge of the graph representation for collaborative localization.

2) Single user scene. In practice, there are also some scenes without multiple users available for the collaborative localization. However, the collaboration can be conducted in another different dimension, i.e. the collaboration among multiple timestamp of single user. Therefore, we propose to utilize the multiple signal fingerprints of single user, which collected in different timestamp, to construct a special graph for collaboration as illustrated in Fig. 1. By this strategy, we can unify the different application scenes into an general graph representation and achieve uniform collaborative indoor localization.

Different from the way of the edge generation in multiple user scenes, we employ IMU sensor measurements to calculate the distance among multiple signal fingerprints which are ordered, since they are collected at different timestamp by a single user in moving. Although IMU based measurements usually suffer from the cumulative errors, we employ IMU data for relative displacement estimation thus avoiding this problem. For IMU based displacement estimation, besides the basic step-based methods, there also exists a plenty of approaches [34, 51], including the recently proposed deep learning based techniques [4] which exhibit good performance in short displacement estimation. Therefore, we take advantage of state-of-the-art IONet [4] to estimation the displacement between multiple timestamp, then use this distance to generate the edge of the graph.

To sum up, we construct an adaptive graph representation for collaborative indoor localization, which is general for different indoor scenes.

4.3 Optimized-GAT based collaborative indoor localization

In this section, we elaborate the core design of proposed collaborative indoor localization network, as illustrated in Fig. 6. We first present the optimized GAT model in Section 4.3.1. Then we detail the designed hierarchical attention mechanism in Section 4.3.2.

4.3.1 Optimized-GAT model. Based on the constructed graph representation, we intuitively consider employing the recently proposed state-of-the-art graph learning techniques to extract the features for collaborative localization. For indoor localization, the propagation or distribution of the localization signals are usually fixed unless the indoor environment changes, which is the important foundation of fingerprint-based indoor localization. Therefore, in a specific indoor scene, although the locations of users and their collaborative relationships (relative distance measurements) are highly dynamic, multiple signal fingerprints collected at specific positions (at a specific timestamp) and the corresponding interrelations are usually static. That means users' locations and their collaborative relationships are static at a specific timestamp of the dynamic variation. Therefore, although graph learning is originally suitable for the problems that nodes and edges are relatively static, it can also work for dynamic indoor localization at a static point-in-time.

More specifically, as the analysis in Section 4.2, the discrete magnetic signal fingerprints at different positions may be the same or similar. This means that two vertexes with different distances from the target one in the graph may have same fingerprint features in collaboration. If the basic GCN is directly applied (the edge weights in the graph are uniform distribution), the feature confusion and ambiguity may occur in collaboration, leading to the degradation of network performance. Compared with GCN, GAT further considers the weight of the edges and brings in attention mechanism to generate adaptive weights of the edges for feature aggregation in the graph. On the one hand, setting weight for edges is more consistent with the characteristics of collaborative indoor localization problem. On the other hand, adaptive attention mechanism helps improving the adaptability for complex and diverse indoor environment (different distributions of signals). Therefore, we utilize GAT instead of GCN as a basic unit to construct the collaborative localization network.

Different from original GAT, we take the distance between vertexes into consideration when generating attention weights. To be specific, we denote the vertex embedding (vertex feature) as $\mathbf{h}_i$ for vertex i in the constructed graph. Then for each iteration of the message passing in the graph, the vertex features will be updated as following.

$$\mathbf{y}_i^s = \mathbf{W}^s \mathbf{h}_i^s, \tag{12}$$

$$e_{ij}^s = LeakyReLU\left(\xi(d_{ij})(\mathbf{b}^{sT}(\mathbf{y}_i^s || \mathbf{y}_j^s))\right), \tag{13}$$

$$\omega_{ij}^s = \frac{exp(e_{ij}^s)}{\sum_{k\in N(i)} exp(e_{ik}^s)}, \tag{14}$$

$$\mathbf{h}_i^{s+1} = \sigma\left(\sum_{j\in N(i)} \omega_{ij}^s \mathbf{y}_i^s\right), \tag{15}$$

where $\mathbf{y}_i$ represents the linear transformation from $\mathbf{h}_i$, and N denotes the number of the vertexes in the graph. $\mathbf{W}, \mathbf{b}$ are the learnable parameters. Specially, d_{ij} represents the distance measurement between vertex i and j, which is detailed in Section 4.2. And $\xi(\cdot)$ denotes the transform function which follows standard Gaussian distribution.

$$\xi(d) = \frac{1}{\sqrt{2\pi}} exp(-\frac{d^2}{2}), \tag{16}$$

With optimized GAT which further fits the characteristic of collaborative localization problem, we are able to effectively extract the features from the graph for accurate localization.

4.3.2 Hierarchical attention mechanism. In order to increase the expression ability of the proposed collaborative localization network when facing the complex indoor scenes, we first bring in the multi-head attention setting and the residual structure. Then combined with optimized GAT, we further devise a hierarchical attention mechanism as illustrated in Fig. 6. Specifically, the designed hierarchical attentions include two levels, i.e. the multi-head

attention in the optimized GAT unit and the multi-layer attention among the residual structure, which are elaborated in detail as follows.

1) Multi-head attention. Similar to multi-channel in convolutional neural networks, we first bring in multi-head attention in the optimized GAT unit to enrich the model's capabilities and stabilize the training process. Each head of the attention has its own parameters and we integrate the outputs of multi-head attention by the way of concatenation. With the multi-head attention mechanism, the vertex feature update of the optimized GAT in Equation 15 will be transformed into Equation 17.

$$\mathbf{h}_i^{s+1} = ||_{(k=1,\ldots,K)}\sigma\left(\sum_{j\in N(i)}\omega_{ij}^{sk}\mathbf{W}^{sk}\mathbf{h}_i^s\right), \tag{17}$$

where the K denotes the number of the head in multi-head attention, and $||$ represents the concatenation operation among the multiple outputs.

2) Multi-layer attention. As the indoor scenes are complex and diverse, the network usually needs to stack more layers to ensure localization performance. While the graph based neural networks usually base on the message passing mechanism among the connected vertexes, they usually suffer from the over smoothing problem (all vertex features tend to be the same) when the depth of the network increases. To address this, we further employ the residual structure in the proposed network. Meanwhile, we also set up a multi-layer attention among the outputs of different layers in residual structure, considering that the features in different depth of the network may have different importance distribution for different indoor scenes.

To be specific, we construct a deep residual network with the layers of num Q, based on the optimized GAT with multi-head attention. Then the devised residual structure can be defined as follows.

$$\mathbf{h}_i^{s+1} = \sigma\left(\frac{1}{Q+1}\left(\sum_{q=1}^{Q} w_i^q G^q(\mathbf{h}_i^s) + w_i^0\mathbf{W}\mathbf{h}_i^s\right)\right), \tag{18}$$

$$w_i^q|_{q=0,1,\ldots,Q} = \mathbf{w}_i = \mathbf{U}_i\mathbf{h}_i^s, \tag{19}$$

where $G(\cdot)$ represents the update function which is elaborated in Equation 12 - 17, and $G(\cdot)^q$ denotes the function which outputs the result for stacked q layers. w is the adaptive weight.

Finally, we construct the optimized GAT-based network with devised hierarchical attentions, utilizing which to extract the features from the adaptive graph representation of the multiple magnetic fingerprints, and achieve accurate collaborative localization.

4.4 Location refinement with model consensus

In this section, we present the location refinement strategy based on the consensus of ensemble models. We first discuss the motivation of location refinement in Section 4.4.1, then elaborate the detailed design of location refinement in Section 4.4.2.

4.4.1 Motivations. In practice, the random environment noise and signal outliers in the signal fingerprint collection are usually inevitable, suffering from which the deep learning based approaches are generally difficult to obtain a unique and stable optimization model through the regular model training. For a certain deep learning task, sometimes we can even obtain several different models which may give different predictions after the different training, but they almost have same accuracy on the training dataset, causing the uncertainty of the model training and the instability of prediction. And the researchers have demonstrated that the single model obtained by regular training usually has limited generalization ability and robustness. And the caused random errors sometime will lead to large deviation and final failure of the prediction consequently.

Algorithm 1: Signal similarity based ensemble model fusion.

Data:

D representing the pre-established signal fingerprint databse;

$\{model_j|_{j=1,2,...,T}\}$ representing the T models trained with different initial parameters;

Input:

$S_g = \{\mathbf{g}_1, \mathbf{g}_2, ...\}$ representing the set of the graph represention for collaborative localiztion, in which $\mathbf{g}_i$ representing the input graph that contains N vertexes;

Output:

$S_l = \{\mathbf{L}_1, \mathbf{L}_2, ...\}$ corresponding to S_g, in which $\mathbf{L}_i$ represents the final prediction result that contains N locations.

for *each graph* $\mathbf{g}$ *in* S_g **do**
  denote the graph $\mathbf{g}$ as $\mathbf{g}_i$;
  for j *in range*$(1, T+1)$ **do**
    Obtain the prediction $\mathbf{L}_i^j = \{\mathbf{l}_i^j|_{i=1,2,...,N}\}$ of $model_j$ for input $\mathbf{g}_i$;
    Find the matched signal fingperint in D with the predicted locations $\mathbf{L}_i^j$ as Equation 21;
    Calculate the weight for $model_j$ with collected and matched signal fingerprints as Equation 22, 23 ;
  Fuse the predictions of T models with the calculated weights as Equation 24 ;
  Obtain the final prediction result $\mathbf{L}_i$ for input graph $\mathbf{g}_i$;
final;
return Fusion prediction results $S_l = \{\mathbf{L}_1, \mathbf{L}_2, ...\}$

To tackle the above problem, we take advantage of ensemble learning mechanism and propose a multi-model joint location prediction strategy. As discussed above, due to the existence of signal random noise and outliers in the training data, we can obtain multiple independent models through different training settings based on the proposed network. After training convergence, these models almost have same accuracy on training dataset but have different sensitivity to the random noise and signal outliers. Then we can integrate the predictions of all these independent models and make a vote for final location estimation. By this way, we are able to effectively reduce the impact of the problem that a single location prediction model is prone to random errors. And the underlying idea is that even if one single sub-model incurs large random deviations, other sub-models can correct the error back.

Specifically, the initial parameters settings of training for ensemble models are different. For the layers of each model, we set the initial parameters (denoted as P) following Gaussian distribution:

$$P \sim N(\mu, \varrho^2), \tag{20}$$

where $\mu \in [-1, 1]$ and $\varrho^2 \in (0, 10]$, and they are random selected from the range for each model.

4.4.2 Signal similarity based ensemble model fusion. Through the ensemble learning method, we can obtain multiple independent models for localization. And we further devise a signal similarity based model fusion strategy to integrate the prediction of these models. For adaption and robustness consideration, we first take the predictions (locations) as basis to backward find the corresponding signal fingerprints in the pre-established geo-tagged database. Then compared with the currently collected signal fingerprints (the input of the localization model), we calculate the similarity between them as the fusion weights. Finally, we use the normalized weights as attention to fuse the prediction of ensemble models for final location estimation.

Table 2. Baseline training parameters in experiments.

Parameters	Value
Iterations	600
Mini-batch Size	30000
Initial Learning Rate	0.005
Learning Rate Update Interval	20
Weight Decay	5e-4

More specifically, as presented in Algorithm 1, we assume that there are T trained models through the ensemble learning method. Then for an input graph representation denoted as $\mathbf{g}$ which includes N vertexes, each of the ensemble models (trained based on proposed collaborative localization network) will give a prediction result including N locations corresponding to N vertexes. And the process of proposed ensemble model fusion localization for vertex i is defined as follows:

$$\widetilde{\mathbf{M}}_i^j = \tau(\mathbf{l}_i^j), \tag{21}$$

$$z_i^j = ||\mathbf{M}_i - \widetilde{\mathbf{M}}_i^j||_2 + \frac{\mathbf{M}_i \cdot \widetilde{\mathbf{M}}_i^j}{||\mathbf{M}_i||_2 \times ||\widetilde{\mathbf{M}}_i^j||_2}, \tag{22}$$

$$\bar{z}_i^j = \frac{exp(z_i^j)}{\sum_{k=1}^{N} exp(z_i^k)}, \tag{23}$$

$$\mathbf{l}_i = \sum_{j=1}^{T} \bar{z}_i^j \mathbf{l}_i^j \tag{24}$$

where $\mathbf{l}_i^j$ denotes the predicted location of j-th model for the vertex i. And $\tau(\cdot)$ represents the transform function from location to signal fingerprint in the database. $\mathbf{M}_i, \widetilde{\mathbf{M}}_i^j$ are the collected magnetic fingerprint at vertex i and transformed magnetic fingerprint from the location $\mathbf{l}_i^j$, respectively. z_i^j denotes the signal similarity calculation which includes both Euclidean distance and cosine similarity. And $\mathbf{l}_i$ represents the final estimated location for the vertex i of the input graph.

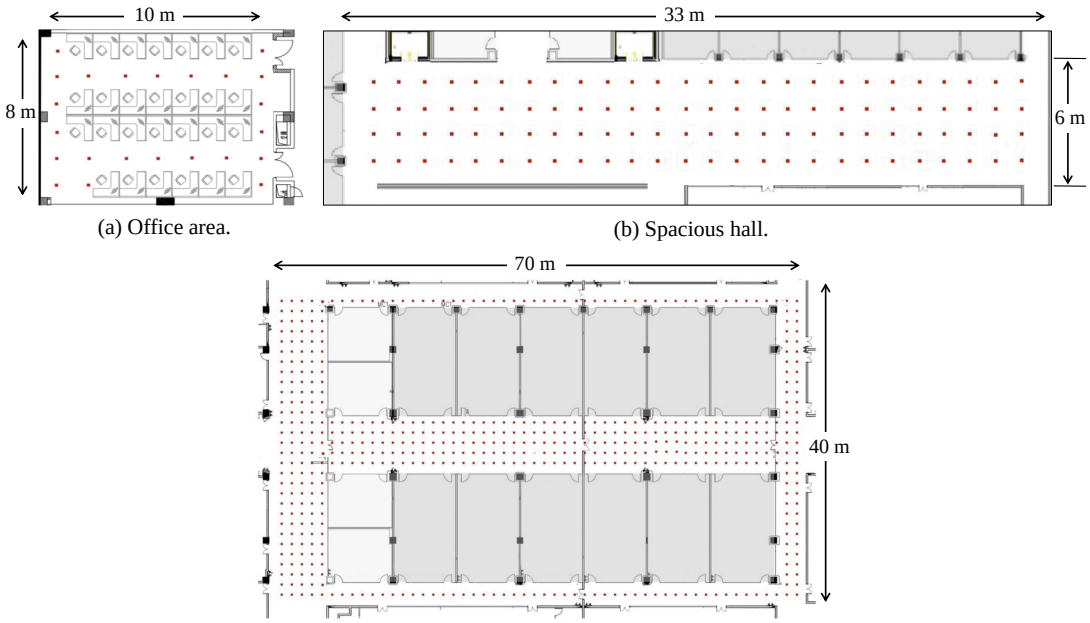
Finally, taking the signal similarity calculations as the weights for multiple ensemble models' fusion, we can obtain more accurate and robust location predictions for all vertexes in the graph.

5 ILLUSTRATIVE EXPERIMENTAL RESULTS

In this section, we evaluate the performance of proposed GC-Loc with extensive experiments. Specifically, we first introduce the related experimental settings in Section 5.1. Then we present the comparison schemes and the evaluation metrics in Section 5.2. Finally, the detailed experimental results and the corresponding analysis are elaborated in Section 5.3.

5.1 Experimental settings

For the localization performance evaluation, we have conducted experiments in three trial sites, including an office area covering around 80 m^2, a spacious hall covering around 200 m^2 and a whole floor of lab building covering around 2800 m^2, and the floorplans are shown in Fig. 7. The other detailed experimental settings are presented as follows, including the implement of GC-Loc, the dataset generation and the details of model training settings.



(a) Office area.

(b) Spacious hall.

(c) The whole floor 5 of lab building.

Fig. 7. The floorplan of three trial sites (the red dots represent the reference points).

5.1.1 The implement of GC-Loc. To evaluate the performance of GC-Loc in practice, we have implemented the proposed GC-Loc as a Client-Server localization system that mainly consists of an application on mobile device and a backend program on the server. And the client application is developed based on Android platform and can be installed on Android devices with common Bluetooth, IMU sensor and magnetometer. The backend program which utilizes the pre-trained model for location calculation runs on a Ubuntu 16.04 server with two Intel Xeon Platinum 8375c CPU, 256 GB system memory and four Nvidia RTX 3090 GPU cards. And the localization model training is also conducted on this server.

To be specific, the details of the implement the Client-Server localization system are as follows.

Client application. The implemented client application is mainly used for signal data collection and data backhaul. And collected signal data include magnetic field observations, relative Bluetooth RSSI scanning results and the IMU sensor readings. The sensor data sampling frequency is set to 50 Hz. Besides, the application also records the timestamp of each collection. For stable data backhaul, we set a buffer mechanism and the currently collected signal data is returned once per second. Since the transmitted data is all text data and the data transmitted per second is less than 2 KB, the data transmission time can be ignored.

Server program. The deployed server program first receives the collected signal data from the client. Then the collected raw data will be pre-processed into adaptive graph representation according to different collaborative mode as elaborated in Section 4.2. Taking the constructed graph as input, the program utilizes the pre-trained collaborative localization model to calculate the locations corresponding to each vertex in the graph, finally achieving accurate collaborative localization.

5.1.2 Dataset and model training. For the pre-prepared model training in offline stage, we first establish the training database of the constructed graph, as elaborated in Algorithm 2. Specifically, the surveyors first collect the magnetic signal observations at all pre-set reference points (the interval distance of reference points is

Algorithm 2: The generation of the graph dataset.

Data:
D representing the pre-collected magnetic fingerprint databse in an indoor area;
Input:
N representing the number of the fingerprints in the graph for collaboration ;
R representing the number of graph in the database to be generated ;
Output:
$D_g = \{\mathbf{g}_1, \mathbf{g}_2, ...\mathbf{g}_R\}$ representing the generated graph database with the size of R;

Initialize $D_g = \{\}$, set to empty ;
Transform the database D into a list of labeled magnetic fingerprints, denoted as D_{list} ;
Calculate all combinations of the list selecting N elements from D_{list}, denoted as $\mathbf{C}$;
Randomly select R combinations from $\mathbf{C}$, denoted as $\mathbf{C}_R$;
Initialize $count = 0$, the number of generated graphs;
for *each combination* c_i *in* $\mathbf{C}_R$ **do**
    c_i consists of N magnetic fingerprints $\{\mathbf{M}_1, \mathbf{M}_2, ...\mathbf{M}_N\}$ with labels $\{\mathbf{l}_1, \mathbf{l}_2, ..., \mathbf{l}_N\}$;
    Calculate the distances among $\{\mathbf{M}_1, \mathbf{M}_2, ...\mathbf{M}_N\}$ based on corresponding locations $\{\mathbf{l}_1, \mathbf{l}_2, ..., \mathbf{l}_N\}$;
    Generate adjacent matrix of the fingerprints (the vertexes), as Equation 10 and 11 ;
    Obtain the graph $\mathbf{g}_i$ consists of $\{\mathbf{M}_1, \mathbf{M}_2, ...\mathbf{M}_N\}$ with the generted adjacent matrix ;
    if $\mathbf{g}_i$ *is a connected graph* **then**
        Append the generated graph $\mathbf{g}_i$ into D_g ;
        $count = count + 1$;
if $len(D_g) < R$ **then**
    $R = R - count$;
    Then jump to Step 4 ;
final;
return The graph database $D_g = \{\mathbf{g}_1, \mathbf{g}_2, ...\mathbf{g}_R\}$

1.2 *m*) in the target indoor areas and label the collected magnetic fingerprints with corresponding locations, constructing a geo-tagged magnetic signal fingerprint database. Based on the constructed fingerprint database, we first generate the adaptive graph representation used for collaborative localization. Then through multiple iterations, we establish the graph dataset for the model training, each element in which contains N vertexes (the magnetic fingerprints) and the corresponding adjacent matrix of the vertexes, and the label of the element is consists of N locations (where the magnetic fingerprints are collected) corresponding to the vertexes in the graph. Besides the training dataset, we also construct another independent test graph dataset in which the vertexes in each element are randomly selected, then we conduct simulated experiments on this test dataset to evaluate the performance of proposed model.

Based on the generated graph dataset, we train the model for collaborative indoor localization, and baseline training parameters are presented in Table 2. Concretely, *PyTorch* is adapted as the deep learning framework for the model implement and training in experiments and the deep network's optimizer is set as *Adam*. And the loss function is defined as following, which considers both Euclidean distance and cosine similarity between the

prediction location and ground truth.

$$\zeta(\mathbf{l}^p, \mathbf{l}^g) = ||\mathbf{l}^p - \mathbf{l}^g||_2 + \frac{\mathbf{l}^p \cdot \mathbf{l}^g}{||\mathbf{l}^p||_2 \times ||\mathbf{l}^g||_2} \tag{25}$$

where $\mathbf{l}^p, \mathbf{l}^g$ denote the prediction location and ground truth location respectively. And $||\cdot||$ represents the L_2 norm operation.

Besides, since the distribution of the magnetic signal is highly related to the indoor environment (building structure, construction material and other ferromagnetic objects) unlike electronic signals that have specific propagation mode, different buildings usually have different signal features and distributions. Therefore, it needs to re-train the model when applied in a different building or scene. Fortunately, benefitting from the stability of magnetism, the model training usually only needs to be conducted once for a specific building.

5.2 Comparison schemes and evaluation metrics

To evaluate the performance of proposed GC-Loc and illustrate its superiority with collaboration mechanism, we take the following indoor localization approaches as comparison schemes.

- *GCN-Loc* [43] proposes to utilize the graph structure to model the collaborative relationship, then takes advantage of GCN to extract the features from the collected signal RSSI fingerprints, then uses a multi-layer perceptron (MLP) for localization.
- *DNN-Loc* [18] utilizes the Wi-Fi fingerprints as input and construct a Deep Neural Networks (DNN) for location prediction. In this paper, we build the proposed DNN structure, but use the magnetic fingerprint as input to keep consistent with other methods.
- *W-kNN* [37] devises an improved weighted K-nearest neighbor algorithm for fingerprint-based indoor localization. And the approach considers both the spatial distance and physical distance of RSSI fingerprints for more accurate localization.

In addition, we have also conducted the ablation experiments to evaluate the effectiveness and contribution of the devised modules/mechanisms in GC-Loc. Specifically, we design several variants of proposed GC-Loc for comparison as following.

- *Basic-GAT*: Without the optimization and the customization for collaborative localization problems (Section 4.3.1), we directly employ original GAT [45] for the collaborative localization.
- *GC-Loc-w/o attention*: To evaluate the contribution of the devised attention mechanism (Section 4.3.2), we remove the attention settings from the proposed GC-Loc for comparison.
- *GC-Loc-w/o residual*: Removing the constructed residual structure among multiple layers in GC-Loc and conducting comparison experiments, we evaluate the effectiveness of the residual structure.

Considering the relationships of the collaborative localization, we take the average error of the localization results (the predicted locations of all participants in collaboration) as the error for this collaboration. And we utilize the overall mean localization error as the metric in experiments. Specifically, assume we have R test cases, and each is a constructed graph, denoted as $\mathbf{g}_i$, which contains N vertexes and corresponding ground truth locations are $\{\mathbf{l}_1^i, \mathbf{l}_2^i, ..., \mathbf{l}_N^i\}$. With the proposed GC-Loc, we can obtain the prediction locations, denoted as $\{\hat{\mathbf{l}}_1^i, \hat{\mathbf{l}}_2^i, ..., \hat{\mathbf{l}}_N^i\}$. Then the overall mean localization error ϵ for collaboration can be defined as follows:

$$\epsilon = \frac{1}{R}\sum_{i=1}^{R}\left(\frac{1}{N}\sum_{n=1}^{N}||\hat{\mathbf{l}}_n^i - \mathbf{l}_n^i||_2\right), \tag{26}$$

where $||\cdot||_2$ denotes L_2 norm.

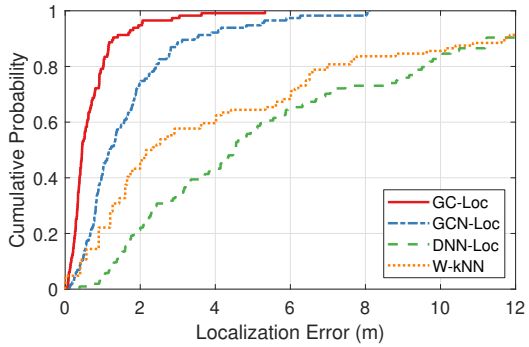


Fig. 8. Cumulative distribution function of indoor localization error with different approaches in the spacious hall.

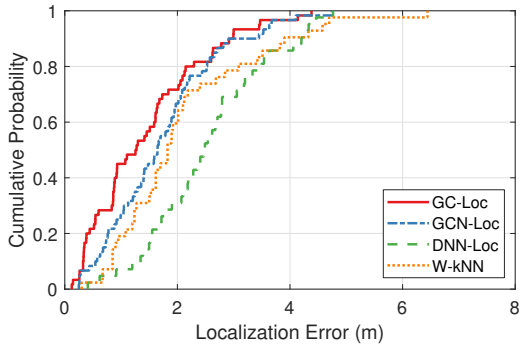


Fig. 9. Cumulative distribution function of indoor localization error with different approaches in the office area.

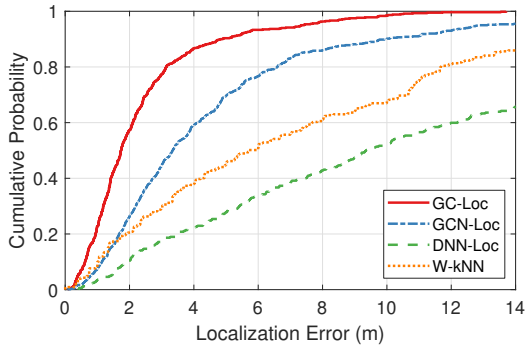


Fig. 10. Cumulative distribution function of indoor localization error with different approaches in the lab building.

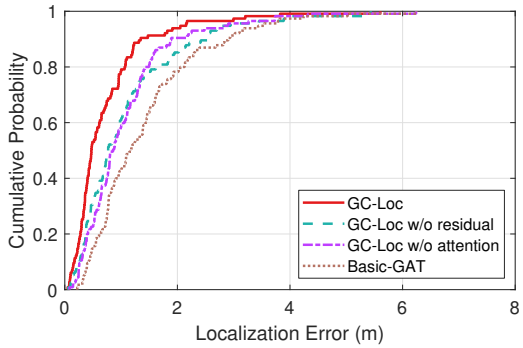


Fig. 11. Cumulative distribution function of indoor localization error with the variants of GC-Loc in spacious hall.

5.3 Experimental results

To evaluate the performance of proposed GC-Loc, we have conducted extensive experiments in three different trial sites, i.e. an office area, a spacious hall and a lab building. Fig. 8, Fig. 9 and Fig. 10 present the CDF of localization errors in the office area, the spacious hall and the lab building, respectively. And Table 3 shows the mean localization error of GC-Loc and competing methods. The results demonstrate that GC-Loc outperforms the competing schemes and achieves accurate collaboration. Compared with W-kNN and DNN-Loc, GC-Loc utilizes a more efficient structure to extract the location features, thus achieving significant improvement. Meanwhile, GC-Loc brings in the distance measurement for collaborative relationship construction, which fits the characteristics of indoor localization problem, so it outperforms the basic GCN-Loc which fails to consider the weights between multiple users. Particularly, comparing the results in Fig. 8 and Fig. 9, we find that GC-Loc performs better in spacious hall. This is because the office area usually has more interference sources, leading to the signal collection with noise and large deviation in measurements. And the spacious hall has less ambient noise interference, which provides a relative stable environment for localization. Meanwhile, the results in lab building as presented in Fig. 10 show the superiority of GC-Loc in large area, achieving higher improvement than that in the other two trial sites, which mainly benefits from the efficient collaborative feature aggregation of GC-Loc.

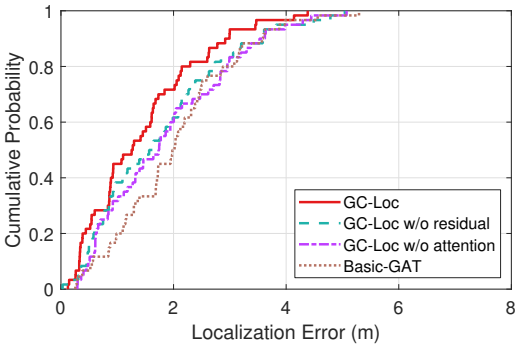


Fig. 12. Cumulative distribution function of indoor localization error with the variants of GC-Loc in the office area.

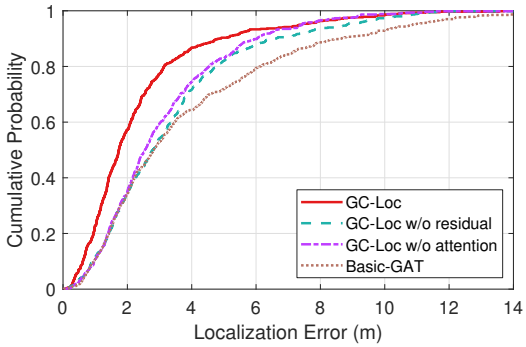


Fig. 13. Cumulative distribution function of indoor localization error with the variants of GC-Loc in the lab building.

Table 3. Mean localization error versus different approach in three trial sites (m).

Approach \ Trial site	Office area	Spacious hall	Lab building
GC-Loc	**1.49**	**0.73**	**2.37**
GC-Loc-w/o residual	1.75	1.1	3.35
GC-Loc-w/o attention	1.86	1.08	3.14
Basic-GAT	2.03	1.43	3.93
GCN-Loc	2.61	2.32	4.55
DNN-Loc	3.52	5.8	11.47
W-kNN	2.93	4.39	7.58

In addition, we have also conducted the ablation experiments to evaluate the effectiveness of the devised modules/mechanisms. Fig. 11, Fig. 12, Fig.13 show the CDF of localization errors with the variants of GC-Loc (Section 5.2) in the office area, the spacious hall and the lab building, respectively. And the mean localization error of GC-Loc and its variants are presented in Table 3. The ablation experiment results demonstrate the effectiveness of the devised attention mechanism, the constructed residual structure and the optimization/customization on GAT, with which GC-Loc can achieve better performance. Particularly, we notice that these devised modules or mechanisms used in GC-Loc make higher enhancement in the spacious hall and the lab building than that in the office area, as shown in Table 3. This is mainly because that the spacious hall and the lab building are larger than the office area and the indoor environment is more similar so that the signal ambiguity is more common, which provides a platform for these devised modules/mechanisms to exert full potentials, thus achieving higher enhancements.

Furthermore, considering the inevitable environment noise and the signal outliers, we employ ensemble learning mechanism in GC-Loc. Fig. 14 and Fig. 16 illustrate the mean localization error of different trained models in spacious hall and office area, respectively. Specifically, we trained 10 localization models with different initial settings respectively. And suffering from the random noise and signal outliers, these models usually show different performances as shown in Fig. 14 and Fig. 16. For stable training and robust localization, we integrate the prediction results of these models for joint location estimation. And a signal similarity based fusion strategy

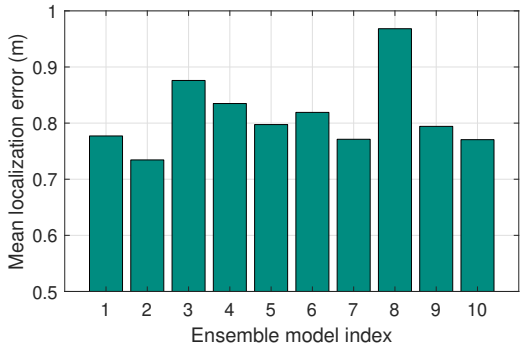


Fig. 14. Mean localization error of different trained models in the spacious hall.

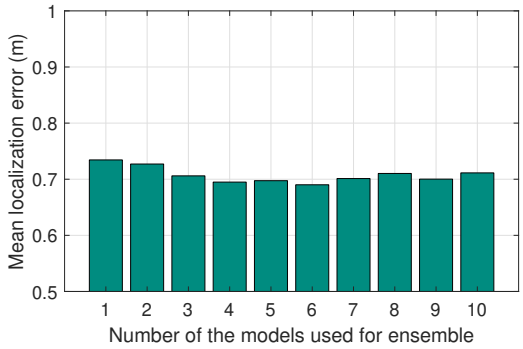


Fig. 15. Mean localization error when using different number of models for fusion in the spacious hall.

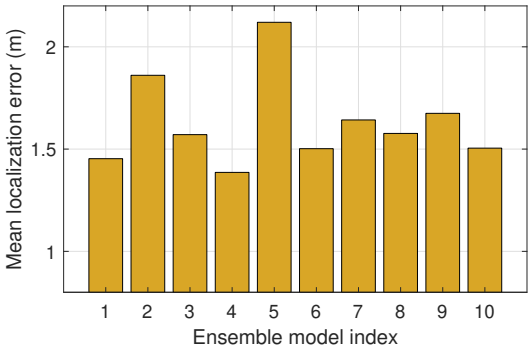


Fig. 16. Mean localization error of different trained models in the office area.

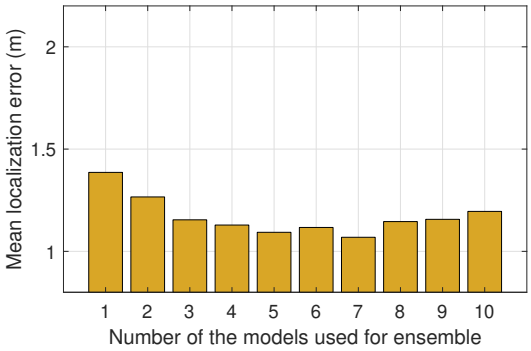


Fig. 17. Mean localization error when using different number of models for fusion in the office area.

is proposed for efficient model fusion, as elaborated in Section 4.4. Based on the trained models and the fusion strategy, we fuse the predictions of multiple models for localization. And Fig. 15 and Fig. 17 illustrate the mean localization error with model fusion when using different number of trained models for fusion. As we can see, GC-Loc is able to further improve accuracy and robustness. However, suffering from the limitation of dataset [39], we notice that the diversity between models becomes smaller with the increase of individual models, and ensemble learning accuracy is worse. Therefore, we have to choose a optimal value for the number of the models used for fusion localization. As an example, the optimal number is set to 7 for the office area based on the experimental results in Fig. 17 to achieve better performance.

Considering the complex and diverse indoor scenes, the deep network is usually used for improving the expression ability. And we devise multi-layer network and employ attention mechanism among the layers to enhance the adaptability. To evaluate this, we construct two kinds of graph with distinguish feature, which represents the different collaborative relationships as presented in Fig. 18. Then we take the constructed graph as input of the trained localization models to observe the automatically generated attention values in the model. And the result is presented in Fig. 19. It shows that the distribution of attention values varies with different input graphs. For graph (a), it represents the common collaborative relationship, in which the average distance among vertexes is shorter than that in graph (b). And in graph (b), some vertexes are not connected with each other

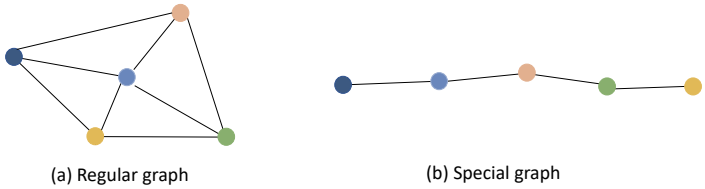


Fig. 18. The diagram of two graphs with distinguish feature, used for evaluating attention settings.

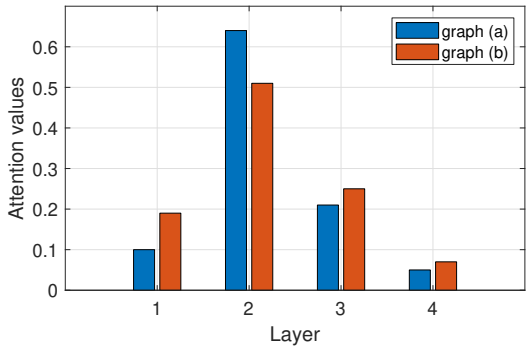


Fig. 19. The distribution of attention values for different types of the constructed graph representation.

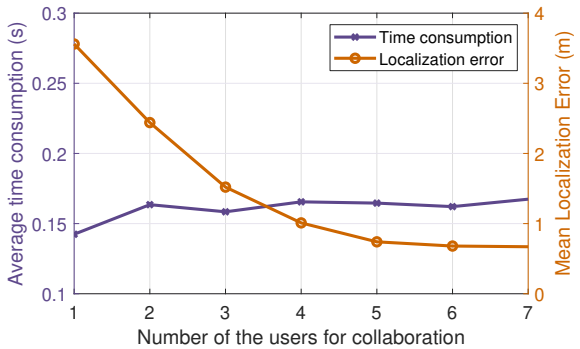


Fig. 20. The average time consumption and mean localization error with different number of users for collaboration.

directly, so it needs more iterations to pass the message through the intermediate vertexes. Therefore, as shown in Fig.19, the weight of deep layers for graph (b) is relatively larger than graph (a), which validates the analysis above. And it also proves the effectiveness of the attention mechanism.

Fig. 20 presents the mean localization error and the average time consumption when taking different number of users for collaboration. For mean localization error, as we can see, it decreases when using more users for collaboration. Intuitively, more collaborative users usually provide more feature information to help the positioning. As shown in Fig. 20 (the orange line), when the number of collaborative users reaches 5, the decrease of the error slows down. It means that collaborative information is saturated. Meanwhile, more users also mean more external noise, which will affect the performance of localization. Therefore, to achieve the trade-off, we usually take 5 users for collaborative indoor localization according to the experimental results. And the number of collaborative users is also set to 5 in most experiments of this paper.

Since GC-Loc is based on deep learning techniques, it's an end-to-end system which conducts the training in prior offline stage and the location prediction in online stage. Therefore, in regard to the system overhead of GC-Loc, we conduct evaluations from the two aspects mainly, i.e., the offline training cost and the online localization latency. For offline training, the detailed model training settings are introduced in Section 5.1.2. As presented in Table 2, the mini-batch size of model training is set to 30000, and we train the model on a NVIDIA 3090 GPU card. During model training, the GPU video memory usage is around 10.26 GB and the average time overhead for a single batch training is 0.83 s. Although the offline training is time-consuming, it only needs to be conducted once. And subsequent online localization can continuously come into play without the extra time-consuming model training. On the other hand, we also evaluate the online localization latency, and Fig. 20 (the purple line) illustrates the average time consumption for a single localization when employing different number of collaborators. The results show that the time consumption of GC-Loc basically keeps stable even with

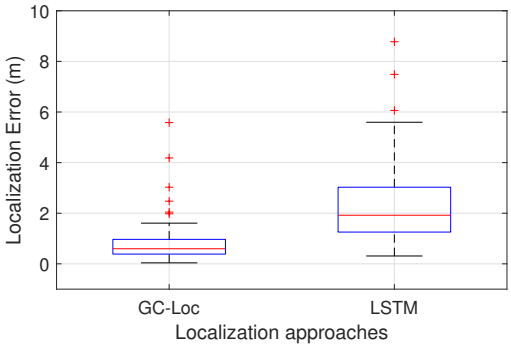


Fig. 21. The distribution of localization error of GC-Loc and LSTM in single user scene.

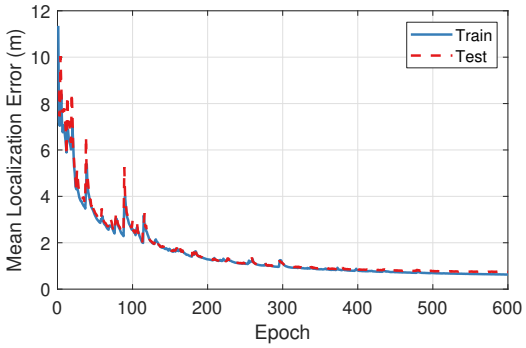


Fig. 22. The variation of train error and test error with iteration during model training.

Table 4. The energy consumption evaluation results of GC-Loc and its individual components (mAh).

Evaluation Item \ Duration	5 min	15 min	30 min	45 min
Magnetism scanning	13	42	88	137
Bluetooth scanning	30	102	218	325
IMU data collection	21	70	139	205
Total consumption with GC-Loc	74	210	442	639

the increase of the number of collaborators. The reason is that GC-Loc is based on the graph representation and its time consumption is largely unaffected by the number of collaborative objects, benefiting from the parallel processing mechanism among the vertexes in the graph. In addition, since the online calculation latency of deep learning model is basically independent of the size of database, GC-Loc also outperforms the traditional matching-based approaches, especially in large indoor scenes.

Despite GC-Loc mainly focuses on the collaboration among multiple users in most common indoor scenes, it also works well in single user scene, in which the multiple timestamps are employed for collaborative localization. And for single user scene, we also conduct experiments to compare the performance of GC-Loc and LSTM which is widely used for sequence-based approaches. Specifically, we use multiple discrete signal fingerprints to construct "discrete sequence" as input of LSTM model, which keeps consistent with GC-Loc. And by this way, it also avoids the trajectory or path constraints, especially in large spacious scenes. As shown in Fig.21, GC-Loc achieves lower localization error and outperforms LSTM by a wide margin, reducing mean error by more than 40%. The results demonstrate the effectiveness of GC-Loc in single user scene.

Fig. 22 illustrates the variation of the mean train error and mean test error during the training iteration. As we can see, variate trends of train error and test error basically keep coincident with iterations, which demonstrates the GC-Loc is able to learn effective collaborative features. And the error decreases quickly in the first 100 epochs, then the decrease slows down. Finally, the model training converges after 600 epochs. So the number of the training iterations in experiments is set to 600 to achieve trade-off.

As for energy consumption, since GC-Loc is implemented based on Client-Sever mode in which the energy concerns are usually from the client, we take the power consumption record in the operation system of client

device for evaluation. In experiments, we evaluate the energy consumption of the simulated localization with GC-Loc (the total consumption), individual continuous magnetism scanning, individual Bluetooth scanning and individual IMU sensor data collection, respectively. Specifically, we modify the developed client program to conduct ablation evaluation and record the energy consumption of each individual component as the criterion. And the experimental device is HUAWEI MatePad Pro with battery capacity of 7250 mAh. The evaluation results are presented in Table 4. It shows that more than half portion of the energy consumption comes from Bluetooth scanning. Since the client program adopts the continuous high-frequency scanning strategy for better localization performance and has not yet optimized well for energy efficiency, we can adjust the signal sampling frequency to achieve trade-off in specific real-world application scenarios.

6 DISCUSSIONS

In this paper, we propose a graph attention based framework (i.e. GC-Loc) for collaborative indoor localization, which utilizes multiple discrete signal fingerprints and can be applied to both multi-user scenario and single-user scenario. We have conducted extensive experiments in three trial sites and achieved efficient feature aggregation and accurate localization. Nevertheless, a few practical challenges (not the main focus of this paper) remain to be addressed or further optimized.

- *Device calibration through crowdsourcing.* In practice, the various devices or sensors usually have different calibrations for signal observation reading, e.g., the magnitude of magnetic field intensity obtained from different devices may be inconsistent [27], which will affect the performance of most fingerprint-based approaches. In GC-Loc, we have to calibrate different devices to a uniform standard before conducting localization. Specifically, we first construct a standard signal fingerprint database. Then at a specific location (preset or obtained with other assists), through the alignment of collected signal fingerprint and corresponding one in the construct standard database, we can calculate the calibration parameter for the current device. And for the devices of the same model, the calibration only needs to be conducted once and the calibration parameters will be recorded for the subsequent invocations. Considering the wide variety of devices and sensors in practice, the calibration may be time-consuming and labor-intensive. But recent study on crowdsourcing techniques [11] can provide an efficient solution. Since the above-mentioned calibration is a relative simple task, it can easily leverage or be integrated into the existing crowdsourcing approaches.
- *Data security and privacy protection.* As indoor localization empowers a variety of mobile applications, the potential privacy and security problems have attracted more and more attentions. The proposed GC-Loc requires collecting and sending the magnetic field measurements, the Bluetooth RSSI measurements and the IMU measurements back to the server for the collaborative location estimation. Since a lot of personal information can be inferred from the sensor data, the more types of data are transmitted to the server, the easier it is for user privacy to be leaked. Specifically, the challenge mainly lies on two aspects, i.e., the server data security and the users' location privacy. For data security, the encryption techniques (e.g. homomorphic encryption and fuzzy logic) can be used to address the concerns since the research in the related fields is very mature. For the users' location privacy protection, k-anonymity and differential privacy techniques are usually taken as effective solutions to guarantee the users' location privacy in some recent works [55]. In addition, with the rapid development of software (e.g. PyTorch Mobile, TensorFlow Lite) and hardware (especially the graphical processing units in mobile smart devices), it's also possible to explore the local offline deployment in mobile end devices, by which the privacy data leakage can be addressed to a certain extent.

7 CONCLUSION

In this paper, we propose a graph attention based collaborative indoor localization framework, termed GC-Loc, which utilizes discrete signal fingerprints to achieve accurate and efficient collaborative localization. Specifically, we first construct an adaptive graph representation to model the collaborative relationships, which is not only applicable to the collaboration among multiple users, but also works well with multiple timestamps' collaboration in the single-user scenes. Then employing state-of-the-art GAT model, we further bring in residual structure and devise a hierarchical attention mechanism for accurate collaborative localization. Finally, we employ the model consensus and design a signal similarity based model fusion strategy to enhance the robustness. The proposed GC-Loc can be easily extended or incorporated into most existing fingerprint-based indoor localization approaches. In this paper, we take the magnetic signal fingerprints as input to evaluate the performance. And the extensive experimental results demonstrate the effectiveness of GC-Loc, outperforming the competing schemes by a wide margin (reducing the mean localization error by more than 42%).

ACKNOWLEDGMENTS

This work is supported in part by the National Natural Science Foundation of China under Grants 61972433 and 62102459, and in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515012242.

REFERENCES

[1] Heba Abdelnasser, Reham Mohamed, Ahmed Elgohary, Moustafa Farid Alzantot, He Wang, Souvik Sen, Romit Roy Choudhury, and Moustafa Youssef. 2016. SemanticSLAM: Using environment landmarks for unsupervised indoor localization. *IEEE Transactions on Mobile Computing* 15, 7 (2016), 1770–1782.

[2] Han Jun Bae and Lynn Choi. 2019. Large-scale indoor positioning using geomagnetic field with deep neural networks. In *2019 IEEE International Conference on Communications (ICC)*. IEEE, 1–6.

[3] Chao Cai, Henglin Pu, Peng Wang, Zhe Chen, and Jun Luo. 2021. We Hear Your PACE: Passive Acoustic Localization of Multiple Walking Persons. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–24.

[4] Changhao Chen, Chris Xiaoxuan Lu, Johan Wahlström, Andrew Markham, and Niki Trigoni. 2019. Deep neural network based inertial odometry using low-cost inertial measurement units. *IEEE Transactions on Mobile Computing* 20, 4 (2019), 1351–1364.

[5] Pan Chen, Jianga Shang, and Fuqiang Gu. 2020. Learning RSSI Feature via Ranking Model for Wi-Fi Fingerprinting Localization. *IEEE Transactions on Vehicular Technology* 69, 2 (2020), 1695–1705.

[6] Zhian Deng, Junchao Wu, Shengao Wang, and Ming Zhang. 2022. Indoor Localization with a Single Access Point Based on TDoA and AoA. *Wireless Communications and Mobile Computing* 2022 (2022).

[7] Yi Ding, Dongzhe Jiang, Yunhuai Liu, Desheng Zhang, and Tian He. 2021. SmartLOC: Indoor Localization with Smartphone Anchors for On-Demand Delivery. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–24.

[8] Yi Ding, Dongzhe Jiang, Yu Yang, Yunhuai Liu, Tian He, and Desheng Zhang. 2022. P2-Loc: A Person-2-Person Indoor Localization System in On-Demand Delivery. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–24.

[9] Jiang Dong, Marius Noreikis, Yu Xiao, and Antti Ylä-Jääski. 2019. ViNav: A vision-based indoor navigation system for smartphones. *IEEE Transactions on Mobile Computing* 18, 6 (2019), 1461–1475.

[10] Rizanne Elbakly and Moustafa Youssef. 2020. The StoryTeller: Scalable building-and ap-independent deep learning-based floor prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–20.

[11] Cole Gleason, Dragan Ahmetovic, Saiph Savage, Carlos Toxtli, Carl Posthuma, Chieko Asakawa, Kris M Kitani, and Jeffrey P Bigham. 2018. Crowdsourcing the installation and maintenance of indoor localization infrastructure to support blind navigation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–25.

[12] Bernhard Großiwindhager, Michael Stocker, Michael Rath, Carlo Alberto Boano, and Kay Römer. 2019. SnapLoc: An ultra-fast UWB-based indoor localization system for an unlimited number of tags. In *2019 18th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 61–72.

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

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

[15] Suining He and Kang G Shin. 2017. Geomagnetism for smartphone-based indoor localization: Challenges, advances, and comparisons. *ACM Computing Surveys (CSUR)* 50, 6 (2017), 1–37.

[16] Tao He, Qun Niu, Suining He, and Ning Liu. 2019. Indoor Localization with Spatial and Temporal Representations of Signal Sequences. In *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 1–7.
[17] Ho Jun Jang, Jae Min Shin, and Lynn Choi. 2017. Geomagnetic field based indoor localization using recurrent neural networks. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 1–6.
[18] Bing Jia, Zhaopeng Zong, Baoqi Huang, and Thar Baker. 2020. A DNN-based WiFi-RSSI Indoor Localization Method in IoT. In *International Conference on Communications and Networking in China*. Springer, 200–211.
[19] Junghyun Jun, Yu Gu, Long Cheng, Banghui Lu, Jun Sun, Ting Zhu, and Jianwei Niu. 2013. Social-Loc: Improving indoor localization with social sensing. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. 1–14.
[20] Tarun Kulshrestha, Divya Saxena, Rajdeep Niyogi, and Jiannong Cao. 2019. Real-time crowd monitoring using seamless indoor-outdoor localization. *IEEE Transactions on Mobile Computing* 19, 3 (2019), 664–679.
[21] Myeongcheol Kwak, Youngmong Park, Junyoung Kim, Jinyoung Han, and Taekyoung Kwon. 2018. An energy-efficient and lightweight indoor localization system for Internet-of-Things (IoT) environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–28.
[22] Christos Laoudias, Demetrios Zeinalipour-Yazti, and Christos G Panayiotou. 2013. Crowdsourced indoor localization for diverse devices through radiomap fusion. In *International Conference on Indoor Positioning and Indoor Navigation*. IEEE, 1–7.
[23] Hang Li, Xi Chen, Ju Wang, Di Wu, and Xue Liu. 2021. DAFI: WiFi-based Device-free Indoor Localization via Domain Adaptation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–21.
[24] Mingkuan Li, Ning Liu, Qun Niu, Chang Liu, S-H Gary Chan, and Chengying Gao. 2018. SweepLoc: Automatic video-based indoor localization by camera sweeping. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–25.
[25] Jie Lian, Jiadong Lou, Li Chen, and Xu Yuan. 2021. EchoSpot: Spotting Your Locations via Acoustic Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–21.
[26] Hongbo Liu, Yu Gan, Jie Yang, Simon Sidhom, Yan Wang, Yingying Chen, and Fan Ye. 2012. Push the limit of WiFi based localization for smartphones. In *Proceedings of the 18th annual international conference on Mobile computing and networking*. 305–316.
[27] Ning Liu, Tao He, Suining He, and Qun Niu. 2021. Indoor Localization With Adaptive Signal Sequence Representations. *IEEE Transactions on Vehicular Technology* 70, 11 (2021), 11678–11694.
[28] Rajalakshmi Nandakumar, Krishna Kant Chintalapudi, and Venkata N Padmanabhan. 2012. Centaur: Locating devices in an office environment. In *Proceedings of the 18th annual international conference on Mobile computing and networking*. 281–292.
[29] Daniel Neunteufel, Stefan Grebien, Stefan Hechenberger, Klaus Witrisal, and Holger Arthaber. 2020. Coherent Chirp Generation by Narrowband Transceiver Chips for ToF Indoor Localization. In *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 1–6.
[30] Xuan-Bac Nguyen, Duc Toan Bui, Chi Nhan Duong, Tien D Bui, and Khoa Luu. 2021. Clusformer: A transformer based clustering approach to unsupervised large-scale face and visual landmark recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10847–10856.
[31] Qun Niu, Tao He, Ning Liu, Suining He, Xiaonan Luo, and Fan Zhou. 2020. MAIL: Multi-scale attention-guided indoor localization using geomagnetic sequences. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–23.
[32] Qun Niu, Mingkuan Li, Suining He, Chengying Gao, S-H Gary Chan, and Xiaonan Luo. 2019. Resource-efficient and Automated Image-based Indoor Localization. *ACM Transactions on Sensor Networks (TOSN)* 15, 2 (2019), 19.
[33] Qun Niu, Ning Liu, Jianjun Huang, Yangze Luo, Suining He, Tao He, S-H Gary Chan, and Xiaonan Luo. 2019. DeepNavi: A deep signal-fusion framework for accurate and applicable indoor navigation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.
[34] Xiaoguang Niu, Luyao Xie, Jiawei Wang, Haiming Chen, Dandan Liu, and Ruizhi Chen. 2020. AtLAS: An activity-based indoor localization and semantic labeling mechanism for residences. *IEEE Internet of Things Journal* 7, 10 (2020), 10606–10622.
[35] Youngtae Noh, Hirozumi Yamaguchi, and Uichin Lee. 2016. Infrastructure-free collaborative indoor positioning scheme for time-critical team operations. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 48, 3 (2016), 418–432.
[36] Pavel Pascacio, Sven Casteleyn, Joaquin Torres-Sospedra, Elena Simona Lohan, and Jari Nurmi. 2021. Collaborative indoor positioning systems: A systematic review. *Sensors* 21, 3 (2021), 1002.
[37] Xuesheng Peng, Ruizhi Chen, Kegen Yu, Feng Ye, and Weixing Xue. 2020. An improved weighted k-nearest neighbor algorithm for indoor localization. *Electronics* 9, 12 (2020), 2117.
[38] Jun-Wei Qiu and Yu-Chee Tseng. 2016. M2M encountering: Collaborative localization via instant inter-particle filter data fusion. *IEEE Sensors Journal* 16, 14 (2016), 5715–5724.
[39] Omer Sagi and Lior Rokach. 2018. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8, 4 (2018), e1249.
[40] Yuanchao Shu, Cheng Bo, Guobin Shen, Chunshui Zhao, Liqun Li, and Feng Zhao. 2015. Magicol: Indoor localization using pervasive magnetic field and opportunistic WiFi sensing. *IEEE Journal on Selected Areas in Communications* 33, 7 (2015), 1443–1457.

[41] Xin Song, Yufeng Zhou, Haoyang Qi, Weipeng Qiu, and Yanbo Xue. 2022. DuLoc: Dual-channel Convolutional Neural Network based on Channel State Information for Indoor Localization. *IEEE Sensors Journal* (2022).
[42] Xiaoyu Sun, Chi Wu, Xiqi Gao, and Geoffrey Ye Li. 2019. Fingerprint-Based Localization for Massive MIMO-OFDM System With Deep Convolutional Neural Networks. *IEEE Transactions on Vehicular Technology* 68, 11 (2019), 10846–10857.
[43] Yanzan Sun, Qinggang Xie, Guangjin Pan, Shunqing Zhang, and Shugong Xu. 2021. A Novel GCN based Indoor Localization System with Multiple Access Points. In *2021 International Wireless Communications and Mobile Computing (IWCMC)*. IEEE, 9–14.
[44] Viet-Cuong Ta, Trung-Kien Dao, Dominique Vaufreydaz, and Eric Castelli. 2018. Smartphone-based user positioning in a multiple-user context with Wi-Fi and Bluetooth. In *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 206–212.
[45] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *6th International Conference on Learning Representations*.
[46] Xuyu Wang, Zhitao Yu, and Shiwen Mao. 2019. Indoor localization using smartphone magnetic and light sensors: A deep LSTM approach. *Mobile Networks and Applications* (2019), 1–14.
[47] Hang Wu, Ziliang Mo, Jiajie Tan, Suining He, and S-H Gary Chan. 2019. Efficient indoor localization based on geomagnetism. *ACM Transactions on Sensor Networks (TOSN)* 15, 4 (2019), 1–25.
[48] Hongwei Xie, Tao Gu, Xianping Tao, Haibo Ye, and Jian Lu. 2016. A reliability-augmented particle filter for magnetic fingerprinting based indoor localization on smartphone. *IEEE Transactions on Mobile Computing* 15, 8 (2016), 1877–1892.
[49] Yu Yang, Yi Ding, Dengpan Yuan, Guang Wang, Xiaoyang Xie, Yunhuai Liu, Tian He, and Desheng Zhang. 2020. TransLoc: transparent indoor localization with uncertain human participation for instant delivery. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.
[50] Ning Yu, Xiaohong Zhan, Shengnan Zhao, Yinfeng Wu, and Renjian Feng. 2017. A precise dead reckoning algorithm based on Bluetooth and multiple sensors. *IEEE Internet of Things Journal* 5, 1 (2017), 336–351.
[51] Yue Yu, Ruizhi Chen, Liang Chen, Xingyu Zheng, Dewen Wu, Wei Li, and Yuan Wu. 2021. A Novel 3-D Indoor Localization Algorithm Based on BLE and Multiple Sensors. *IEEE Internet of Things Journal* 8, 11 (2021), 9359–9372.
[52] Faheem Zafari, Athanasios Gkelias, and Kin K Leung. 2019. A survey of indoor localization systems and technologies. *IEEE Communications Surveys & Tutorials* 21, 3 (2019), 2568–2599.
[53] Heng Zhang, Soon Yim Tan, and Chee Kiat Seow. 2019. TOA-based indoor localization and tracking with inaccurate floor plan map via MRMSC-PHD filter. *IEEE Sensors Journal* 19, 21 (2019), 9869–9882.
[54] Xianan Zhang, Wei Wang, Xuedou Xiao, Hang Yang, Xinyu Zhang, and Tao Jiang. 2020. Peer-to-peer localization for single-antenna devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–25.
[55] Ping Zhao, Hongbo Jiang, John CS Lui, Chen Wang, Fanzi Zeng, Fu Xiao, and Zhetao Li. 2018. P^3-LOC: A privacy-preserving paradigm-driven framework for indoor localization. *IEEE/ACM Transactions on Networking* 26, 6 (2018), 2856–2869.
[56] Yiyang Zhao, Chen Qian, Liangyi Gong, Zhenhua Li, and Yunhao Liu. 2015. LMDD: Light-weight magnetic-based door detection with your smartphone. In *2015 44th International Conference on Parallel Processing*. IEEE, 919–928.
[57] Zenghua Zhao, Jiankun Wang, Xingya Zhao, Chunyi Peng, Qian Guo, and Bin Wu. 2017. NaviLight: Indoor localization and navigation under arbitrary lights. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 1–9.
[58] Yuanqing Zheng, Guobin Shen, Liqun Li, Chunshui Zhao, Mo Li, and Feng Zhao. 2017. Travi-navi: Self-deployable indoor navigation system. *IEEE/ACM transactions on networking* 25, 5 (2017), 2655–2669.