MAIL: Multi-Scale Attention-Guided Indoor Localization Using Geomagnetic Sequences

QUN NIU, School of Data and Computer Science, Sun Yat-sen University, China

TAO HE, School of Data and Computer Science, Sun Yat-sen University, China

NING LIU, School of Data and Computer Science, Sun Yat-sen University, China

SUINING HE, Department of Computer Science and Engineering, The University of Connecticut, United States

XIAONAN LUO*, Guilin University of Electronic Technology, China

FAN ZHOU, School of Data and Computer Science and National Engineering Research Centre of Digital Life, Sun Yat-sen University, China

Knowing accurate indoor locations of pedestrians has great social and commercial values, such as pedestrian heatmapping and targeted advertising. Location estimation with sequential inputs (e.g., geomagnetic sequences) has received much attention lately, mainly because they enhance the localization accuracy with *temporal correlations*. Nevertheless, it is challenging to realize accurate localization with geomagnetic sequences due to environmental factors, such as non-uniform ferromagnetic disturbances. To address this, we propose *MAIL*, a multi-scale attention-guided indoor localization network, which turns these challenges into favorable advantages. Our key contributions are as follows. First, instead of extracting a single holistic feature from an input sequence directly, we design a *scale-based* feature extraction unit that takes variational anomalies at different scales into consideration. Second, we propose an attention generation scheme that identifies attention values for different scales. Rather than setting fixed numbers, MAIL *learns* them adaptively with the input sequence, thus increasing its adaptability and generality. Third, guided by attention values, we fuse multi-scale features by paying more attention to prominent ones and estimate current location with the fused feature. We evaluate the performance of MAIL in three different trial sites. Evaluation results show that MAIL reduces the mean localization error by more than 36% compared with the state-of-the-art competing schemes.

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

Additional Key Words and Phrases: Geomagnetic Indoor Localization, Multi-Scale Features, Attention

---

*This is the corresponding author.

Authors' addresses: Qun Niu, School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China, niuq3@mail.sysu.edu.cn; Tao He, School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China, hetao23@mail2.sysu.edu.cn; Ning Liu, School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China, liuning2@mail.sysu.edu.cn; Suining He, Department of Computer Science and Engineering, The University of Connecticut, Storrs, Connecticut, United States, suining.he@uconn.edu; Xiaonan Luo, Guilin University of Electronic Technology, Guilin, China, luoxn@guet.edu.cn; Fan Zhou, School of Data and Computer Science and National Engineering Research Centre of Digital Life, Sun Yat-sen University, Guangzhou, China, isszf@mail.sysu.edu.cn.

---

1 INTRODUCTION

Knowing fine-grained indoor locations of pedestrians or robots enables compelling and intelligent services, such as assisted living [24], healthcare [6] and crowd monitoring [16, 22], to name a few [1]. Although Global Positioning System (GPS) has achieved pervasiveness and high precision outdoors, it does not work well indoors due to the block of building structures.

To address this, researchers study different signals for indoor localization, e.g., Wi-Fi, Bluetooth, visible light communication. Of all these signals, geomagnetism emerges as a promising one, mainly because it is pervasive and stable indoors over a long period of time. Furthermore, geomagnetism-based indoor localization systems do not require the deployment of extra devices, such as wireless access points (APs) or light-emitting diodes (LEDs), thus are more deployable. Additionally, geomagnetic anomalies, which are caused by local ferromagnetic disturbances (e.g., iron doors, electrical wires and reinforced building structures), provide useful features for location inferring [15].

Geomagnetic fingerprint is a fixed-length vector consisting of three values measured by a magnetometer. It indicates current geomagnetic strength received at a specific location. Due to the stability of geomagnetism, the discernibility of a single geomagnetic fingerprint is relatively low, leading to degraded localization accuracy [30, 32]. In light of this, researchers propose to leverage geomagnetic sequences to localize. Compared with a single fingerprint, a geomagnetic sequence enhances the location discernibility with *temporal correlations*, leading to higher localization accuracy. Geomagnetic anomalies, caused by nearby ferromagnetic disturbances, present distinctive patterns. As these anomalies are distinguishing and are usually associated with specific locations, they become *geomagnetic landmarks* [15].

Many existing sequence-based approaches compare user-collected sequences with geo-tagged ones in the database based on the overall geometrical shape of geomagnetic sequences. Despite their accuracy, they are likely to discard *local anomalies*, which are less significant compared with a sequence, leading to degraded discernibility and compromised localization accuracy. We illustrate this in Figure 1, where the overall geometrical shapes of two magnetic [2] sequences collected far away from each other (22m apart) in a long corridor are similar. However, they are different from each other as the dashed one has a small-scale anomaly (in the red bounding box). This demonstrates that we need to take both large-scale and small-scale anomalies into consideration for accurate localization. A large-scale anomaly corresponds to a magnetic change that spans a long sequence (potentially caused by a large ferromagnetic object), whereas a small-scale one corresponds to that caused by a small object.

In this paper, we propose to leverage various scales of anomalies (caused by different ferromagnetic disturbances) in a geomagnetic sequence to localize. By detecting various scales of anomalies, we enhance the location discernibility and improve the localization accuracy consequently. However, there are three major challenges in realizing such a geomagnetic localization system:

- *How to reduce the computational cost with long sequences?* Some previous approaches extract holistic features from inputs, such as the shape of a geomagnetic sequence [34, 37]. Then, they use shape comparison algorithms, e.g., dynamic time warping (DTW) [4], which consider both stretching and squeezing collected magnetic sequences and map them to geo-tagged ones for localization. However, this usually incurs quadratic increase in computational cost by applying DTW directly to longer sequences [50].
- *How to detect diversified anomalies simultaneously?* In complicated indoor sites, geomagnetic anomalies vary with scales due to nearby magnetic disturbances. Previous holistic shape-matching algorithms, such as DTW, localize a target by finding the most similar geo-tagged geomagnetic sequences in terms of shapes. It focuses more on the global shape and discards small local anomalies, leading to potentially large mapping

---

[1]Without loss of generality, either a human or a robot to be localized is viewed as a *target* in our paper.
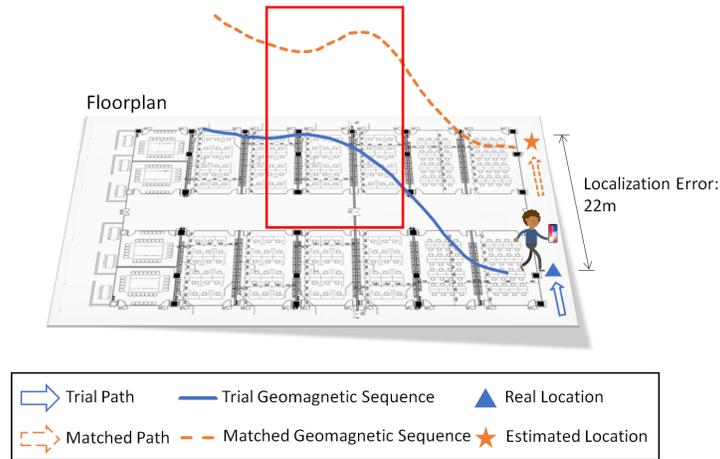[2]We use "geomagnetic" and "magnetic" interchangeably in this paper.

Fig. 1. The geometrical similarities of geomagnetic sequences could lead to location confusion.

errors (Figure 1). It is also tedious to study magnetic anomalies indoors and design handcrafted detectors accordingly.

- *How to fuse features adaptively in localization?* Due to varied source disturbances, anomalies in geomagnetic sequences vary with scales significantly. Therefore, the contribution of each scale is not uniform across trial sites or trial cases. Given a magnetic sequence, how to determine the importance of each anomaly scale adaptively and fuse multi-scale features together remain a challenge.

To address above challenges, we propose *MAIL*, a **m**ulti-scale, **a**ttention-guided **i**ndoor **l**ocalization network with gradient sequences. Specifically, our key novelties are as follows:

- *An efficient sequence-based localization framework*: We propose an end-to-end efficient sequence-based localization framework with little preprocessing overhead. With model training in the offline stage, our algorithm achieves high accuracy and efficiency with multi-scale feature extraction. Experimental results show that MAIL is capable of realizing real-time localization (more than 30Hz) in our trials.
- *A scale-based feature extraction unit*: We design a scale-based feature extraction unit, which segments an input geomagnetic sequence based on the specific scale. As recurrent neural networks are capable of extracting temporal correlations from sequences, we use them to extract a sequential feature from each subsequence. Finally, we concatenate them together, generating a feature vector for each scale. Using multiple scale-based feature extraction units, we get multi-scale features for a geomagnetic sequence.
- *An attention-guided feature fusion and localization scheme*: Since the attention mechanism is able to measure the contribution of each scale to the localization, we propose an adaptive attention generation scheme that estimates an attention value for each scale. Based on the estimated attention values, we fuse multi-scale features together by paying more attention to important scales. Finally, we estimate the user location based on the fused feature.

We have conducted extensive experiments in three different trial sites, a compact lab area with long corridors, a spacious food court and an underground parking lot with large open space. Experimental results demonstrate that MAIL reduces the mean localization error by more than 36% compared with state-of-the-art approaches in our trial sites. In addition to indoor localization with geomagnetic sequences, our MAIL can also serve as an

effective sequential feature extractor for input sequences, which can be easily incorporated into other applications with different sequential inputs, e.g., location prediction [21] and intention anticipation [45].

The remainder of this paper is organized as follows. We review papers most related to ours in Section 2, followed by the workflow of our localization system in Section 3. We elaborate the core design of the localization model in Section 4 and give illustrative experimental results in Section 5. We discuss several deployment limitations in Section 6 and conclude in Section 7.

2 RELATED WORK

Indoor localization has been extensively studied for decades. To achieve sufficient accuracy, researchers have proposed various techniques using different discrete inputs, such as Wi-Fi fingerprints [5, 19, 31, 49], Bluetooth [11, 39], ultra-wide band (UWB) [36], images/videos [25, 29, 38], visible light communication (VLC) [26, 33, 53] and radio frequency identification [27, 40, 41, 47], to name a few. Despite the accuracy in specific trial sites, they have a few practical limitations that hinder the wide deployment. For example, radio frequency signals, such as Wi-Fi and Bluetooth, are subject to signal fluctuations due to multi-path fading or temporal obstructions, which adversely deteriorate the localization performance. Image-based indoor localization, however, is prone to noises due to the motion blur or a loss of focus. VLC-based indoor localization has emerged recently with indoor luminaries, which are often well-structured and reliable indoors. Nevertheless, they usually require modulated light sources, dense deployment of extra devices or holding them at a specific attitude.

Of all these techniques, geomagnetism has attracted much attention lately, mainly because it is pervasive, relatively stable and deployable without infrastructure support [15]. Therefore, researchers begin to study indoor localization with geomagnetism. Existing approaches are broadly divided into two categories: discrete magnetic fingerprint-based localization and magnetic sequence-based localization. The first category leverages a discrete magnetic fingerprint to pinpoint the user location. For example, Chung et al. [9] find the target fingerprint in the database with the least squared error. However, in large trial sites, discrete geomagnetic fingerprint could be ambiguous due to symmetric building structures or electrical wires, leading to large localization errors.

To address this, researchers study indoor localization with geomagnetic sequences. Compared with a discrete fingerprint, a magnetic sequence consists of more fingerprints continuously collected in a time window and enhances the location discernibility with temporal correlations [17, 44]. Additionally, due to the impact of nearby ferromagnetic disturbances, the specific magnetic patterns become *landmarks*, which can be used to discriminate locations. For example, Magil [43] localizes a user with pure geomagnetic sequences. MaLoc [46] collects geomagnetic fingerprints with a smartphone and estimates current location with the proposed augmented particle filter. To reduce the computational complexity, Mapel [42] fuses geomagnetism with pedometer based on graphical model. GROUPING [52] also collects magnetic fingerprints and compares them with a geo-tagged fingerprint database for localization. Instead of Euclidean distance, it leverages the inverse of cosine similarity to measure the differences between two magnetic sequences. LocateMe [37] compares a magnetic sequence with geo-tagged ones based on dynamic time warping (DTW).

Furthermore, Magicol [34] and MagFi [44] fuse opportunistic Wi-Fi fingerprints to reduce localization errors. SemanticSLAM [1] models indoor landmarks with specific magnetic patterns and uses them to calibrate current location estimations. Kwak et al. [23] localize with geomagnetic fingerprints. To save energy, they fuse magnetism with Bluetooth for location estimation. CrowdX [7] builds floorplans with magnetic readings, Bluetooth readings and opportunistic encounters of mobile users. Travi-Navi [54] proposes a self-deployable navigation scheme termed *leader-follower*, where a leader surveys a path on-demand and then a follower follows the path to the destination. Despite their improvements, DTW and other traditional shape matching based algorithms have several limitations. First, it is time-consuming to match two sequences directly. As the length increases, the computational cost grows [4, 35]. Second, DTW compresses or stretches magnetic sequences in the time axis for
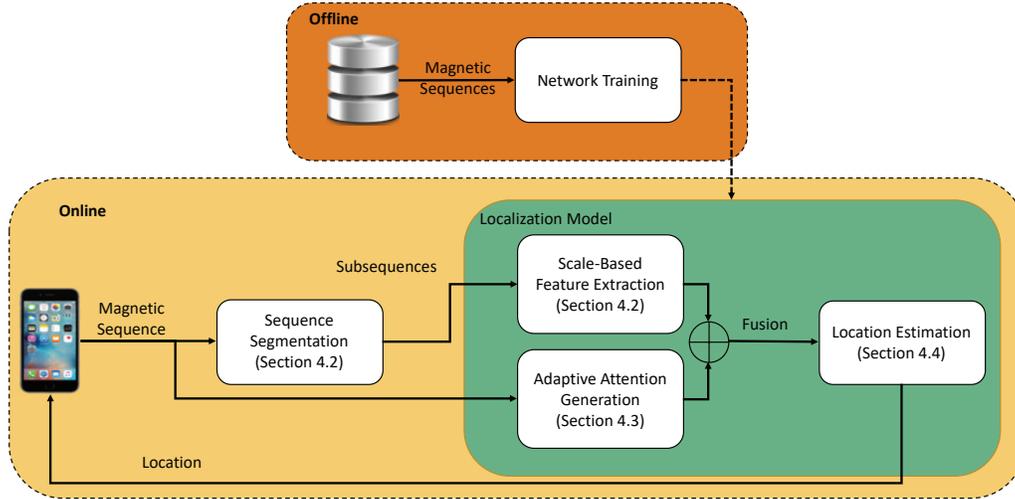
Fig. 2. Overview of MAIL.

better alignment. Particularly, it considers more about the overall shape of two magnetic sequences. Consequently, it may discard small local anomalies, which could lead to alignment errors between magnetic sequences. Our MAIL is orthogonal to the average smoothing or filtering, as it focuses on extracting temporal features from the input sequence. However, we can also employ the smoothing technique in Magicol [34] to reduce statistical noises in geomagnetic sequences to enhance the localization accuracy.

Inspired by the recent success of deep learning techniques, researchers begin to use recurrent neural networks to process sequential inputs for localization. The work in [20] is related to ours, where authors propose to use a basic recurrent neural network (RNN) [13] to localize a user with geomagnetic sequences. Our proposed model is significantly different from it in several aspects. First, instead of extracting a single holistic feature from the input sequence, we divide it into different scales (in terms of subsequence length). Then we propose a scale-based feature extraction unit to extract features from each scale to detect variational anomalies and subsequently enhance the location discernibility. Second, in order to determine the contribution of each scale to the localization, we propose an adaptive approach to determine the attention value for each scale. Guided by these attention values, we fuse multi-scale features together for location estimation.

3 SYSTEM WORKFLOW

In this section, we overview the workflow of MAIL in Figure 2, which consists of two main phases: an offline phase and an online phase. In the offline stage, a surveyor designs survey paths and collects geomagnetic sequences along them while holding the smartphone in the upright position. During the site survey, the surveyor walks in constant speed and marks landmarks (e.g., corners, turnings, doors) along the path. Based on these landmarks, the surveyor labels the ending position of each magnetic subsequence by interpolation. Afterwards, we calculate the gradient of magnetic readings and construct database. Then, we train a multi-scale localization network using our training data.

In the online stage, a user takes out a smartphone and collects data with it. During the localization stage, the user should hold the phone in the upright position as well. However, we can also calibrate the sensor readings and map them to the upright position based on the attitude of the device (inferred from smartphone sensors). As the user walks, the client application collects geomagnetic readings automatically and sends them to a remote

Table 1. Major symbols used in MAIL.

Notations	Definitions
N	Number of samples in a geomagnetic sequence
M	Number of scales
K	Number of trial cases
c	Length of a magnetic subsequence
P	Number of subsequences with length c
$\mathbf{g}_p$	Feature vector of a subsequence p
$\mathbf{\Phi}_m$	Feature vector corresponding to the scale m
$\mathbf{\Psi}_m$	Normalized feature vector corresponding to $\mathbf{\Phi}_m$
$\mathbf{f}$	Multi-scale feature vector corresponding to a magnetic sequence
$\mathbf{\Gamma}$	$N \times 3$ geomagnetic sequence
$\mathbf{x}_k$	Ground truth 2-D coordinate corresponding to trial case k
$\hat{\mathbf{x}}_k$	Estimated 2-D coordinate corresponding to trial case k

server for localization. After calculating the gradient sequence, MAIL then divides it into subsequences and extracts scale-based features (Section 4.2). Meanwhile, MAIL estimates the attention value corresponding to each scale based on the gradient sequence (Section 4.3). Based on multi-scale features and corresponding attention values, MAIL fuses these features together by paying more attention to prominent ones and estimates current location (Section 4.4). Finally, MAIL sends the estimated position back to the client.

Table 1 presents major symbols used in this paper.

4 MULTI-SCALE LOCALIZATION MODEL

We elaborate the core design of the localization model in this section. First, we overview the structure of our network in Section 4.1. Afterwards, we present the scale-based feature extraction unit in Section 4.2. Then, we elaborate the adaptive attention generation in Section 4.3, followed by attention-guided feature fusion and localization in Section 4.4. Finally, we discuss its theoretical foundations in Section 4.5.

4.1 Overview

In this section, we introduce the multi-scale attention-guided indoor localization network. The overall structure is presented in Figure 3. Instead of using raw geomagnetic sequences to localize, we use the gradient sequence to reduce the impact of device heterogeneity. Afterwards, we segment the gradient sequence into different subsequences with preset scales. Then, we employ the scale-based feature extraction unit to extract sequential features for different scales. To identify prominent ones, we generate an attention value for each scale, indicating its importance in each trial. Finally, we fuse scale-based features together with attention guidance and regress current location accordingly.

More specifically, our model consists of four major components:

1) Data pre-processing. In this part, we first calculate the gradient of magnetic sequence $\mathbf{\Gamma}$ (length is N) to reduce the impact of device heterogeneity (Figure 4), from which we get a *gradient sequence*, denoted by $\mathbf{\Gamma}'$. Then we segment it with specific scales. For scales from 1 to M, we obtain the subsequence sets $\{\mathbf{s}_1, \cdots, \mathbf{s}_m, \cdots, \mathbf{s}_M\}$, where $\mathbf{s}_m$ ($1 \leq m \leq M$) is a set of subsequences corresponding to scale m.

2) Scale-based feature extraction. In this module, we design a scale-based feature extraction unit, termed *SFE* unit, where we take segmented subsequences as input and extract features for each of them. For example, for
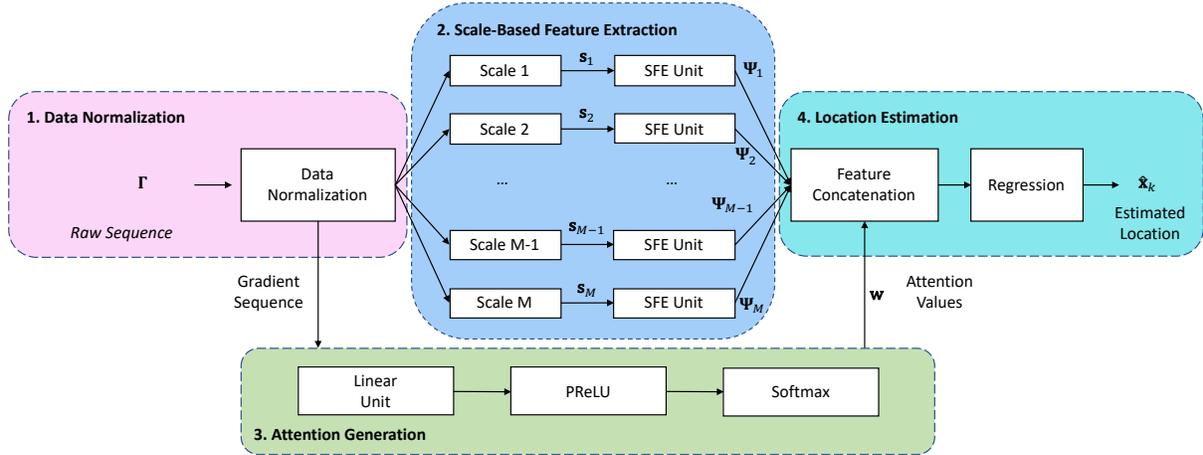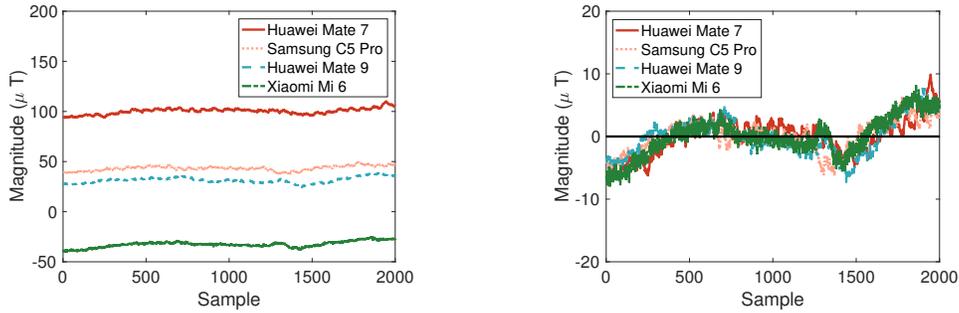
Fig. 3. Overall multi-scale localization model.

(a) Diverse magnetic readings from different devices. (b) Calibrated readings by removing the mean value of each sequence.

Fig. 4. Magnetic measurements with different devices: (a) initial readings, and (b) readings after calibration.

scale $m(1 \leq m \leq M)$, we extract a feature vector for each subsequence in the set $\mathbf{s}_m$. Then, we concatenate feature vectors corresponding to all subsequences and get a normalized sequential feature $\mathbf{\Psi}_m$ for scale m. We detail the design of SFE unit in Section 4.2.

3) Adaptive attention generation. Based on our SFE unit, we are able to extract multi-scale features from the gradient sequence $\mathbf{\Gamma}'$. Due to diversified configurations indoors, various disturbances contribute differently in each localization trial. This indicates that we should pay more attention to important scales. Nevertheless, it is tedious to set a fixed attention value for each scale in every trial. Consequently, we design an adaptive attention generation unit to identify the prominence of each scale in Section 4.3.

4) Attention-guided feature fusion and localization. Based on the attention value for each scale, we design an attention-guided multi-scale feature fusion scheme to generate the location feature. Furthermore, we design a location estimation module that consists of several linear and non-linear mapping functions. Based on the fused feature, we estimate the ending location corresponding to the input magnetic sequence by regression. Finally, we present our loss function in Section 4.4.
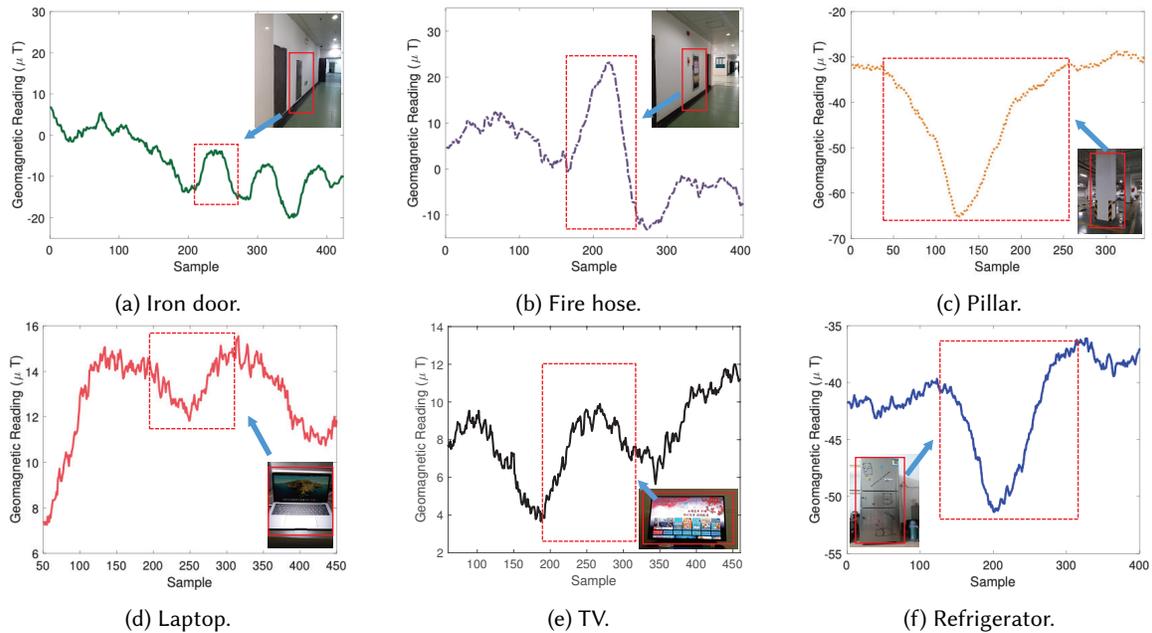
Fig. 5. Illustrations of anomalies around geomagnetic disturbances.

4.2 Scale-Based Feature Extraction

In this section, we design the scale-based feature extraction unit. We first discuss our motivations in Section 4.2.1. Then, we detail its design in Section 4.2.2.

4.2.1 Motivations for multi-scale mechanism. Many temporal sequences, due to their anomalies, provide useful clues for spatial location inferring. However, local disturbances usually vary from each other in terms of coverage, rendering it difficult to extract distinguishing location features with a global feature extractor for the whole input sequence.

To illustrate this, we collect geomagnetic sequences while passing by various magnetic disturbances. As shown in Figure 5, an iron door results in smaller region of impact (around 50 samples), whereas the anomaly spans a larger range (more than 100 samples) near a fire hose. This is because the fire hose has a large iron ax and a pipe inside. Furthermore, the anomaly near the pillar is larger than the door and the fire hose with steel-reinforced concrete.

Furthermore, we have collected geomagnetic sequences along typical electrical appliances. For example, the magnetic anomaly around a laptop is relatively small (around 100 samples). Similar to the laptop, the anomaly around a TV is similar (around 100 samples). This is because their power is similar (around 60 watt). Different from them, the refrigerator results in larger region of impact (around 200 samples), as its power level is larger (around 380 watt). These examples show that anomalies in collected geomagnetic sequences vary with the ferromagnetic disturbance. Consequently, a feature extractor that extracts features at a single fixed scale may overlook small anomalies amid a long geomagnetic sequence or ignore other parts of large anomalies. This may lead to degraded discernibility of location features, hence large localization errors.

In addition, due to varied walking speeds, the numbers of samples collected in a given region differ from each user as well. We show the collected signals along a trial path with a device in Figure 6. For instance, the number
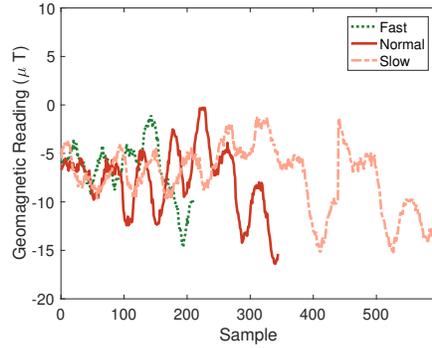
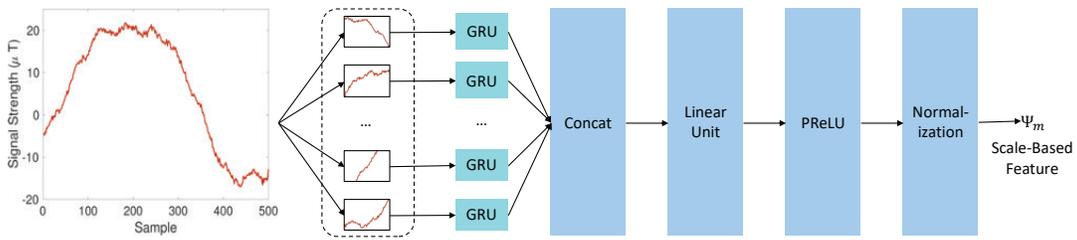Fig. 6. Readings from users walking in different speeds.

Fig. 7. Scale-based feature extraction unit (SFE unit).

of geomagnetic samples collected by a user walking fast maybe small, leading to a *squeezed* anomaly. In contrast, if a user walks slowly, this user may collect a large number of samples, leading to a *stretched* anomaly. Therefore, it is challenging to extract a distinguishing location feature from a magnetic sequence.

Based on our observations, we therefore propose a novel multi-scale approach to extract features from input geomagnetic sequences. Instead of extracting features from the whole sequence with a single feature extractor, we segment this sequence into smaller *subsequences*. Then, we extract sequential features from each of these subsequences, respectively. Together with these features, we generate a set of features corresponding to each scale. By fusing sets of features (different scales) together, we generate more comprehensive and distinguishing features for location estimation.

4.2.2 Scale-Based Feature Extraction Unit. In this section, we detail our feature extraction unit for each scale, termed *SFE unit* in Figure 7.

Based on our observations, many anomalies vary in scale with environmental factors. For a specific scale, the SFE unit first segments the gradient sequence. Then, it extracts features for each subsequence. Formally, given a gradient sequence Γ' of length N', we segment it into subsequences with size c to extract sequential feature at scale m. The number of subsequences P is:

$$P = \lceil \frac{N'}{c} \rceil. \tag{1}$$

In the case that N' is not an integral multiple of c, we have overlaps between adjacent subsequences for segmentation.

After the segmentation of gradient sequence Γ', we exploit the GRU [8] to extract features from each subsequence at the current scale. GRU enhances RNN with gated mechanism, which reduces vanishing gradients in

the training stage. Furthermore, it is more computationally efficient compared with long short-term memory (LSTM) [12, 18], which facilitates its training and prediction on mobile platforms. The GRU is defined as follows:

$$\mathbf{r}_t = \sigma(\mathbf{W}_{ir}\mathbf{x}_t + \mathbf{b}_{ir} + \mathbf{W}_{hr}\mathbf{h}_{t-1} + \mathbf{b}_{hr}), \tag{2}$$

$$\mathbf{z}_t = \sigma(\mathbf{W}_{iz}\mathbf{x}_t + \mathbf{b}_{iz} + \mathbf{W}_{hz}\mathbf{h}_{t-1} + \mathbf{b}_{hz}), \tag{3}$$

$$\bar{\mathbf{h}}_t = \tan(\mathbf{W}_{in}\mathbf{x}_t + \mathbf{b}_{in} + \mathbf{r}_t(\mathbf{W}_{hn}\mathbf{h}_{t-1} + \mathbf{b}_{hn})), \tag{4}$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t)\bar{\mathbf{h}}_t + \mathbf{z}_t\mathbf{h}_{t-1}, \tag{5}$$

where $\mathbf{x}_t$ and $\mathbf{h}_t$ denote the input and the hidden state at time t, respectively. The hidden state at time 0 is $\mathbf{h}_0 = \mathbf{0}$. $\sigma(\cdot)$ is the logistic sigmoid function. We use $\mathbf{r}$, $\mathbf{z}$ and $\bar{\mathbf{h}}$ to denote update gate, reset gate and candidate activation, respectively. These gates allow the cell in GRU to keep, update or forget information over time. We use $\mathbf{W}_{i*}$ and $\mathbf{W}_{h*}$ to indicate weight values for current input and previous hidden state, respectively. Additionally, we use $\mathbf{b}$ to denote bias terms.

After feature extraction with GRU, we get a feature vector corresponding to each subsequence, denoted by $\mathbf{g}_p$, where $1 \le p \le P$. Then, we concatenate them together to generate the feature Φ_m of the corresponding scale m ($1 \le m \le M$):

$$\Phi_m = [\mathbf{g}_1; \mathbf{g}_2; \cdots; \mathbf{g}_p; \cdots; \mathbf{g}_P], \tag{6}$$

where $[\cdot; \cdot]$ indicates feature concatenation.

Then, we map the extracted feature Φ_m to a fixed-size vector by a non-linear unit, which mainly consists of several fully connected (FC) layers and a parametric rectified linear unit (PReLU) [14] function:

$$\Psi_m = \mathbf{v}^T * \text{PReLU}(\mathbf{U}^T\Phi_m + \mathbf{d}), \tag{7}$$

where $\mathbf{v}$, $\mathbf{U}$ and $\mathbf{d}$ are learnable parameters.

Finally, we append a normalization layer to the SFE unit to speed up the model training. Instead of traditional batch normalization, we leverage state-of-the-art layer normalization [3]. This is because the normalization layer is more computationally efficient and improves the performance of recurrent neural networks with its bounding paradigm.

4.3 Adaptive Attention Generation

We extract features with multiple scales so that we are able to capture various geomagnetic anomalies. Due to a large variety of indoor configurations, the ranges of anomalies vary from one trial site to another. Consequently, different scales play various roles in location estimation. For example, in a compact lab area, the ferromagnetic disturbances may be small, e.g., iron doors, leading to small-scale anomalies. While in a large complex, such as a food court, they have more reinforced steel structures, which usually incur large-scale anomalies. The large diversity between multi-scale features leads to significant challenges in balancing scale-based features manually.

To address this, we propose an adaptive attention generation scheme for each scale. By generating an attention value for each scale, we are able to achieve generality and accuracy. Our intuition is that in a short geomagnetic sequence (spans around 10 meters), features extracted from a few scales are crucial, while others are not. Consequently, we generate larger attention values for important scales and smaller ones for others.

In order to generate attention values, we design an adaptive attention generation unit, which consists of several linear layers, no-linear mapping functions and finally a softmax layer for attention normalization (Figure 3). The normalized attention values for all scales are between 0 and 1 and they add up to 1. We detail the attention generation module as follows. Given the gradient sequence Γ', we calculate the attention vector $\bar{\mathbf{w}} = [\bar{w}_1, \bar{w}_2, \cdots, \bar{w}_M]$ corresponding to all scales as follows:

$$\bar{\mathbf{w}} = \mathbf{e}^T * \text{PReLU}(\mathbf{q}^T\Gamma' + \mathbf{r}), \tag{8}$$

where $\mathbf{e}$, $\mathbf{q}$ and $\mathbf{r}$ are learnable parameters, PReLU is an activation function used to increase nonlinearity and $\bar{w}_m$ indicates the attention value for scale m ($1 \leq m \leq M$). Then, we normalize scale-dependent attention values as follows:

$$w_m = \frac{\exp(\bar{w}_m)}{\sum_{m=1}^{M} \exp(\bar{w}_m)}. \tag{9}$$

After normalization, the sum of all attention values adds up to 1:

$$\sum_{m=1}^{M} w_m = 1. \tag{10}$$

4.4 Attention-Guided Multi-Scale Feature Fusion and Localization

Based on above, we fuse scale-based features extracted by M SFE Units based on attention values. Formally, the feature set extracted for M scales is defined as follows:

$$\Lambda = \{\Psi_1, \ldots, \Psi_{M-1}, \Psi_M\}. \tag{11}$$

Then, the vector of attention values for all M scales is defined as follows:

$$\mathbf{w} = [w_1, \ldots, w_{M-1}, w_M]. \tag{12}$$

Based on features extracted from various scales and their corresponding attention values, we fuse multi-scale features as follows:

$$\mathbf{f} = [w_1\Psi_1; \ldots; w_{M-1}\Psi_{M-1}; w_M\Psi_M], \tag{13}$$

where $[\cdot; \cdot]$ is the concatenation operation and $\mathbf{f}$ indicates the multi-scale feature vector.

Through the guidance of attention values during the fusion process, we highlight important scales with large attention values while reducing those with small values, thus achieving effectiveness. Rather than setting fixed attention values, our approach is data-driven and does not require manual calibration, thus is more adaptable to different trial sites.

Based on the fused feature, we estimate the location where the last geomagnetic sample is collected. Formally, we estimate current location $\hat{\mathbf{x}}_k$ based on the fused feature $\mathbf{f}_k$ corresponding to trial case k as follows:

$$\hat{\mathbf{x}}_k = \mathbf{W}\mathbf{f}_k + \mathbf{b}, \tag{14}$$

where $\mathbf{W}$ and $\mathbf{b}$ are learnable parameters for location estimation. In our experiment, $\hat{\mathbf{x}}$ is rounded (e.g., to three decimals) according to the requirement of the application.

In the training stage, we use the mean squared error as the loss function, which is defined as follows:

$$\mathcal{L} = \frac{1}{K} \sum_{k=1}^{K} (\mathbf{x}_k - \hat{\mathbf{x}}_k)^2, \tag{15}$$

where K is the number of trial cases. We use $\mathbf{x}_k$ and $\hat{\mathbf{x}}_k$ to denote the ground truth and estimated location corresponding to trial case k, respectively.

4.5 Discussions on the Theoretical Foundations of MAIL

In this section, we discuss the theoretical foundations of MAIL. Specifically, we elaborate the underpinning theories of recurrent neural networks and detail the intuition and foundations of attention mechanism in our settings.

As discrete geomagnetic fingerprints indoors are relatively stable, we use RNNs to extract sequential features. Basic RNNs organize neuron-like nodes into successive layers, where each node is connected in a directed manner to other nodes in the successive layer. Each node has time-varying activation function and each connection has

a learnable weight value. Nevertheless, earlier RNNs are able to capture and reserve information from recent context. In the cases where more context is needed (long-term information), they do not perform well. In light of this, the long short-term memory (LSTM) integrates cell states so that it is able to learn long-term information from a longer sequence [18]. Gated recurrent units (GRUs) [8] integrate the gating mechanism. With fewer parameters than LSTM, GRUs are able to achieve both accuracy and efficiency with small datasets or sequences like geomagnetic. Therefore, by using more recent recurrent networks, our model is able to learn the sequential information for localization.

Generally, an RNN network processes the sequence-to-sequence problem, where the sequence is treated uniformly. This, however, does not generalize well to complicated indoor sites with various scales of anomalies (Figure 5). Unlike sentences, geomagnetic sequences consist of numerical values with far fewer contextual information, leading to degraded distinctiveness of the sequential feature. Therefore, we propose a scale-based approach, where we segment the input sequence into subsequences of different scales. Consequently, we are able to extract sequential features corresponding to anomalies at different scales, generating more diverse representations.

According to psychological studies, humans tend to pay more attention to the part of information that is more important [2]. Motivated by this, attention mechanism is proposed to selectively focus on more relevant content. For example, a part of a sentence or a portion of an image is *selectively weighted* to distinguish it from others. Fusing these weighted components together, the algorithm is able to identify prominent parts and pays more attention to them to generate distinguishing features [48]. By focusing more on important scales, our network generates distinguishing features adaptively for location inferring, thus enhancing the localization accuracy.

5 ILLUSTRATIVE EXPERIMENTAL RESULTS

To evaluate the performance of MAIL, we have conducted extensive experiments in three different trial sites, including a narrow lab area, a spacious food court and an underground parking lot. Figure 8 illustrates floorplans of these trial sites. Covering around $2,800\ m^2$, the lab area has long corridors and many room partitions. In contrast to the lab area, the food court covers more areas with open space (around $3,500\ m^2$). The underground parking lot covers around $6,100 m^2$ with large open areas. We first discuss the implementation details in Section 5.1. Then we present experimental setup and performance metrics in Section 5.2. We illustrate localization results in Section 5.3, followed by the discussion of system overhead in Section 5.4.

5.1 Implementation of MAIL

We have implemented the proposed MAIL localization system, which consists of two parts: a mobile client application and a backend server program. The client application can be installed on Android devices with a magnetometer. As for the localization program, we implement it on a Ubuntu 16.04 server, which has two Intel E5-2640 central processing units (CPUs), eight 32 GB random access memory cards and four Nvidia 2080 Ti graphical processing units (GPUs).

Client program. The client program is implemented in the Android platform. We use it to collect magnetometer readings. During the data collection, it also records the timestamp when a magnetic reading is collected. The sampling frequency is 50 Hz in both offline and online stages. We have a buffer mechanism. Instead of sending each magnetic reading to the server separately, we collect around 50 samples and send them to a server via Wi-Fi networks.

Server program. The localization program is implemented in a server, which receives magnetic readings from the client. After receiving a sufficient amount of readings, it concatenates them together based on the timestamp and forms a magnetic sequence. Then it extracts scale-based location features and fuses them together for accurate localization.
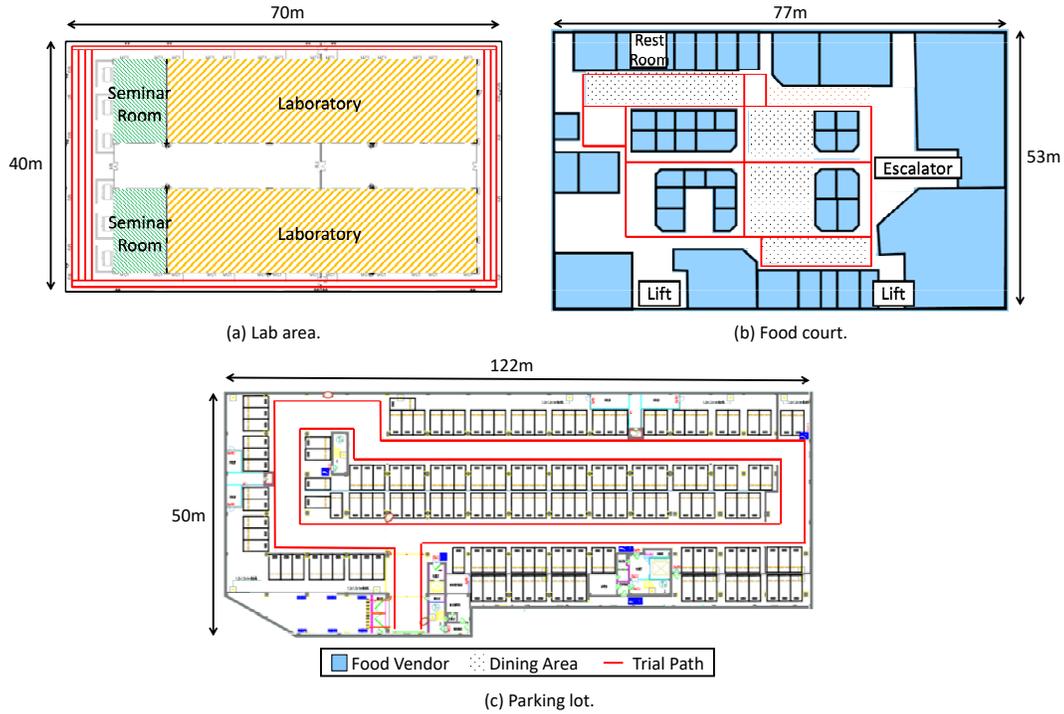
(a) Lab area.

(b) Food court.

(c) Parking lot.

Fig. 8. Floorplans of our trial sites: a) Lab area; b) Food court; c) Parking lot.

5.2 Experimental Settings and Performance Metrics

We have implemented the client program on Android phones. In order to train a multi-scale model for localization, we have designed dense survey paths to cover popular areas in our trial sites. In the offline stage, we ask a volunteer to hold a trial device and walk along all survey paths in constant speed. During the walking, the smartphone faces upward and points to the walking direction. A client program collects magnetic readings automatically. While in the online localization stage, we invite a volunteer to take a smartphone and walk with it in each trial site. Specifically, we invite a male volunteer (height 180cm, numbered #1) to take a Samsung S7 and walk with it in the lab area and the food court. In the parking lot, we invite a female volunteer (height 160cm, numbered #4) to take Xiaomi MI 6 to collect training and testing magnetic sequences. In these trial sites, we ask the volunteer to walk along each survey path for several times. We evaluate the localization results with those collected in the last round and train with remaining sequences. Figure 8 shows the floorplans and survey paths in our trial sites. In our trials, the localization application records geomagnetic readings along the path and we label starting and ending locations manually based on nearby landmarks, such as doors and corners. We label locations between starting and ending positions by linear interpolation.

We have collected 1195, 976 and 752 training sequences in the lab area, the food court and the parking lot, respectively. For localization, we have collected another 385, 241 and 376 magnetic sequences, respectively. We train a network model for each trial site and conduct evaluations accordingly. In order to evaluate the prevalence of local anomalies, we have conducted a statistical analysis. To this end, we design a shape comparison algorithm using the DTW to detect anomalies at each scale and find their positions in each training or trial sequence. By

Table 2. Proportions of local anomalies in our dataset.

	Lab Area	Food Court	Parking Lot
Train	70.0%	71.0%	56.9%
Trial	71.2%	65.5%	61.2%

comparing their locations, we determine that local anomalies exist if two anomalies overlap. We present the proportions of local anomalies in Table 2. It shows that more than 56% of samples in our dataset have local anomalies in these trial sites. Furthermore, the portion of local anomalies in the parking lot is lower, mainly because it is more spacious and has fewer geomagnetic disturbances.

We compare the performance of MAIL with the following state-of-the-art localization algorithms:

- *Jang et al.* [20], which train an RNN network to estimate the location of a target. In our experiment, we build an RNN network with 4 layers as a comparison scheme.
- *Magicol* [34], which finds overlapping paths by comparing magnetic sequences with DTW. Then, it proposes a bi-directional particle filter to reduce estimation errors of user motion during the localization.
- *MaLoc* [46], which localizes a target by comparing collected fingerprints with geo-tagged ones stored in a database. Then, it constrains location estimations based on user motion and particle filters.

We measure the localization error of a target with the Euclidean distance. Given K trial sequences with ground truth locations $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K\}$ and estimated ones $\hat{\mathbf{X}} = \{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \cdots, \hat{\mathbf{x}}_K\}$, we denote the ground truth location and estimated one of trial k as $\mathbf{x}_k$ and $\hat{\mathbf{x}}_k$, respectively. The mean localization error ε is defined as follows:

$$\varepsilon = \frac{1}{K} \sum_{k=1}^{K} ||\mathbf{x}_k - \hat{\mathbf{x}}_k||_2, \tag{16}$$

where $|| \cdot ||_2$ denotes the L_2 norm. We round location estimations to three decimal places in our experiment.

In order to evaluate the improvement with attention guidance, we design a baseline approach without adaptive attention generation, termed *MAIL w/o Attention*, where the attention values of all scales are the same. We also conduct additional experiment at different walking speeds in the lab area. Specifically, we ask volunteer #1 to walk at slow (around 1.0m/s), normal (around 1.2m/s) and fast (around 1.4m/s) speeds and collect magnetic readings. Then we use our trained MAIL to evaluate the positioning accuracy. In order to address the speed diversity, we enhance our training with data augmentation. Specifically, we propose to interpolate and downsample our training sequences to augment the training database. For example, by interpolation, we have longer magnetic sequences in an area, indicating slower walking speed and longer dwelling time. In contrast, we generate shorter magnetic sequences, indicating faster walking speed. According to the report in [28], the average walking speed of a healthy human ranges from 1.0m/s (relatively slow) to 1.5m/s (relatively fast). Therefore, our data augmentation is general to a wide range of users.

To evaluate the impact of devices and users on the accuracy of MAIL, we have invited more volunteers to take part in the additional evaluation in the lab area and the food court. We invite two more male volunteers (heights are 165cm and 170cm, numbered #2 and #3, respectively) to take part in the localization process in the lab area. Similarly, these volunteers are asked to walk along designed paths several times and collect magnetic sequences one after another. The trial devices taken by these users are Samsung C5 Pro, Huawei Mate 9 and Xiaomi MI 6. In the food court, we invite volunteer #2 to collect magnetic sequences with Xiaomi MI 6. Each volunteer collects the same amount of testing sequences as volunteer #1 does.

We develop an Android application and run it consecutively to evaluate its power consumption in different devices. First, we kill all other applications and execute evaluations by collecting geomagnetic sequences at 50Hz.

Table 3. Baseline parameters in our experiment.

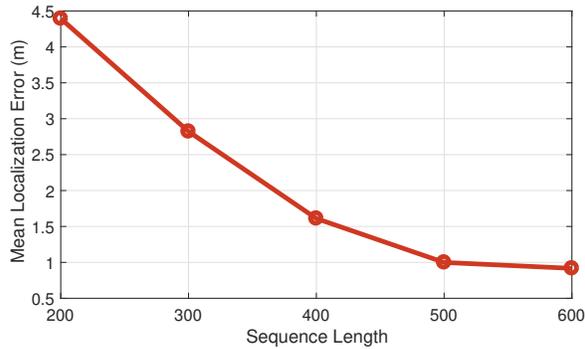| Parameters	Value
Epoch	1500
Initial Learning Rate	0.02
Batch Size	400
Number of Scales	6
Sequence Length	500

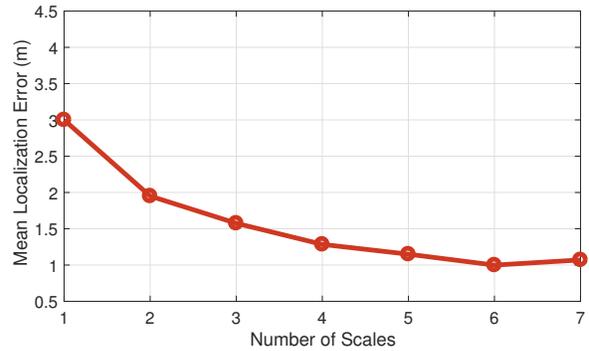Fig. 9. Mean localization Error versus length of geomagnetic sequences (lab area).

Fig. 10. Mean localization Error versus number of scales (lab area).

As discussed, we design a buffer mechanism in our client, where we store collected geomagnetic readings in the device until we get a sufficient amount of readings. Then, we pack them together and send them to a remote server through the Wi-Fi network. During the trial, the application is kept in the foreground and the screen is kept on. We measure the percentage of battery before and after the evaluation. Then, we kill all applications for the same amount of time and measure the power drop with screen on. The total power consumption is the difference between the power drop with application running and that without any foreground applications. By dividing the total power consumption by the number of trials, we get the average power consumption of each trial.

We present the baseline parameters in our trials in Table 3. The lengths from scale 1 to scale 6 are 50, 100, 150, 200, 250 and 500, respectively. To update the weights of our network, we use the Adam optimizer.

5.3   Localization Results

We show the mean localization error with different sequence lengths in Figure 9. It shows that the mean localization error decreases with longer magnetic sequences. This is because we have more diverse geomagnetic anomalies in a longer magnetic sequence, which enhances the location discernibility. However, the decrease in the localization error slows down as the sequence length becomes longer than 500. This is because it provides sufficient location clues for localization. As the sequence length grows longer, it takes more time to collect magnetic samples during the bootstrap stage. Furthermore, it also incurs more network overhead in the course of data transmission. To achieve trade-off between the localization accuracy with time and network overhead, the sequence length is 500 in our trials.
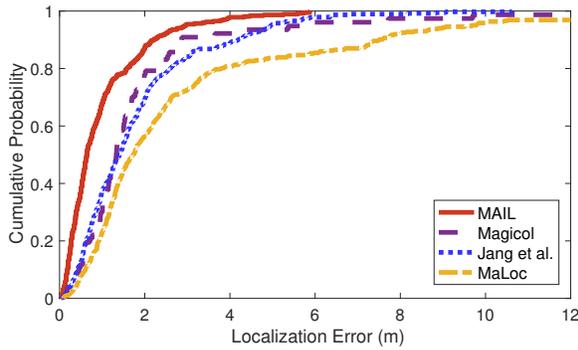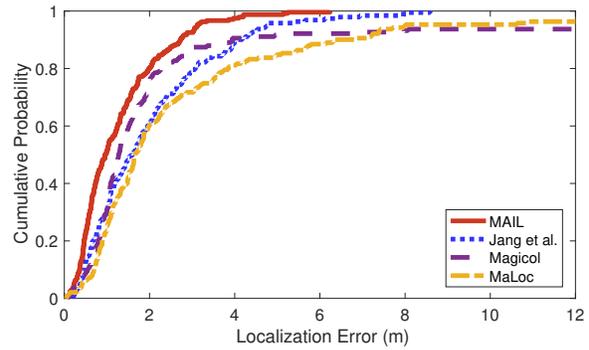
Fig. 11. CDF of localization error (lab area).

Fig. 12. CDF of localization error (food court).

We present the impact of the scale number on the localization accuracy in Figure 10. It shows that the overall localization error is smaller with more scales. This is because we are able to detect diverse anomalies with more scales. By detecting multiple scales, we are able to generate more distinguishing location features, thus increasing the localization accuracy. Nevertheless, as the number of scales increases, the decrease in localization error slows down. This is because we have extracted sufficient information from the training data. Furthermore, over-fitting may occur. Therefore, the localization error increases slightly. In addition, the number of parameters in our model increases with more scales, leading to slower convergence. Consequently, we set the number of scales to 6 to achieve sufficient localization accuracy.

We illustrate the cumulative density function (CDF) of the localization error in the lab area in Figure 11. Compared with the state-of-the-art competing schemes, our approach is able to achieve sufficient accuracy. This is mainly due to three reasons. First, we introduce the multi-scale feature extraction. By detecting multi-scale anomalies, we are able to enhance the distinctiveness of location features. Second, instead of setting fixed weights for each scale, we leverage an adaptive attention generation strategy, which identifies prominent scales for each magnetic sequence. Third, guided by attention values, we fuse multi-scale features attentively and generate distinguishing location features. Combined with the powerful capacity of deep learning techniques, MAIL improves localization accuracy by a large margin (around 36%).

We present the CDF of localization error in the food court and the parking lot in Figure 12 and Figure 13, respectively. They show that the proposed MAIL achieves sufficient accuracy compared with competing schemes in more spacious trial sites. This is mainly because we extract features with a scale-based unit. Therefore, we are able to extract fine-grained features. Furthermore, we determine the attention value for each scale adaptively. Guided by them, we fuse multi-scale features by focusing on scales that are more important. Based on the fused feature, we are able to achieve sufficient localization accuracy.

We present the loss value in training stages in Figure 14. It shows that the loss decreases with more epochs. This is because MAIL is capable of fitting the training data with the forward and backward propagation. As the number of epochs grows, the decrease in loss value slows down. This is because our model has fitted the training data. The overall loss in the parking lot is larger than those in the lab area and the food plaza, mainly because the parking lot is more spacious and contains fewer anomalies. This could lead to lower location discernibility and thus larger loss values.

We compare the mean localization error with (w/) and without (w/o) attention guidance in the lab area and the food court in Figure 15. It shows that by incorporating attention guidance, MAIL increases the localization accuracy. This is mainly because MAIL is able to identify *prominent* scales. By paying more attention to them,
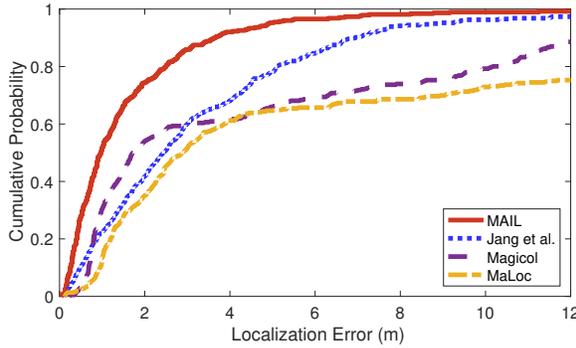
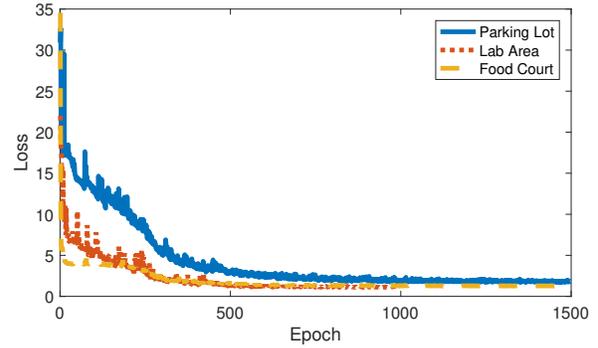Fig. 13. CDF of localization error (parking lot).

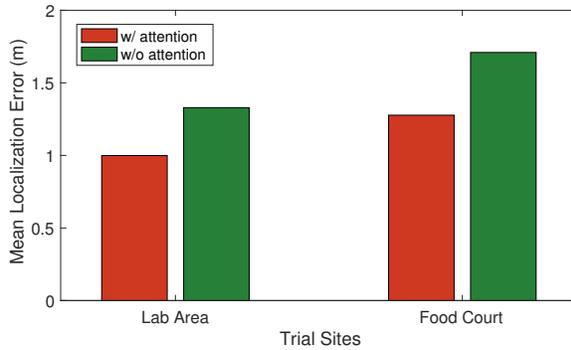

Fig. 14. Loss with epochs.

Fig. 15. Mean localization error w/ and w/o attention.
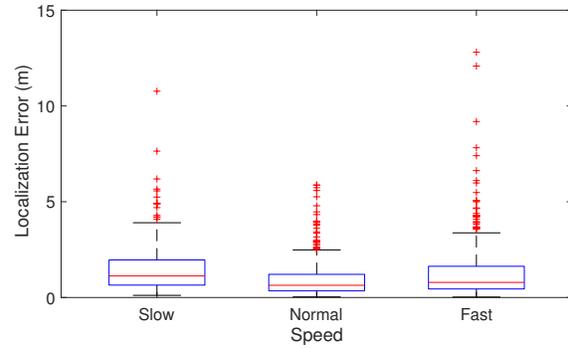

Fig. 16. CDF of localization error at different walking speeds (lab area).

MAIL fuses multi-scale features more effectively and generates distinguishing location features. Therefore, it achieves sufficient improvement in localization accuracy compared with that without attention guidance.

We present the mean localization error at different walking speeds in Figure 16. It shows that our approach achieves comparable localization accuracy at different walking speeds. Due to the data augmentation in the training process, our MAIL learns sufficient information about anomalies at different walking speeds. Consequently, it is able to detect anomalies at different walking speeds and give comparable positioning accuracy.

To illustrate, we present the distribution of mean attention values in the lab area and the food court in Figure 17. It shows that the distribution of attention values varies with the trial site. In the lab area, the mean attention values of small scales (scale 1 and 2) are larger. This indicates that MAIL pays more attention to smaller scales in the lab area. While in the food court, the mean attention values of scale 3, scale 4 and scale 6 are larger than those in the lab area. This is mainly because it is more spacious, thus farther away from ferromagnetic sources, leading to larger and smoother anomalies. This shows that larger scales play a more important role in the food court. Furthermore, we show detailed distributions of attention values in Figure 18(a) and Figure 18(b), respectively. Experimental results demonstrate that MAIL is capable of generating attention values adaptively for these trial sites.

In order to evaluate the generality of the proposed system to different devices, we conduct another experiment with several smartphones in the lab area. Figure 19 shows the localization error with different devices is comparable.
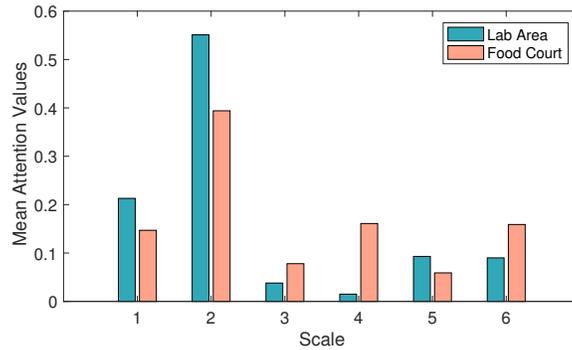
Fig. 17. Distribution of mean attention values of each scale.

Table 4. Average time consumption of a localization trial.

Method	Magicol [34]	MaLoc [46]	Jang et al. [20]	MAIL
Time (s)	1.79	0.24	0.06	0.1

This is mainly because we train our network with gradient rather than raw magnetic readings for localization. By using magnetic gradient, we are able to reduce the impact of device heterogeneity. Thus, our system achieves generality between devices.

Figure 20 shows the localization error with different users in the lab area. It shows that our MAIL achieves comparable localization error with different users (height from 165cm to 180cm) and walk patterns. This demonstrates the applicability of the proposed network model to different users.

5.4 System Overhead

In this section, we present the time and energy consumption of MAIL. We train our neural network in a server with an Nvidia 2080ti GPU. The average time consumption of training a model is around 10 minutes. In the localization stage, the sampling rate of geomagnetism is around 50Hz.

Time Consumption. The total time consumption of collecting and sending 1073 packages is around 1113 seconds. Dividing this by the number of packages, the average time consumption for sending each package to a server is around 1 second. This is mainly due to the limited sampling rate of magnetometer (50 Hz). In the sever, we estimate the user location with a pre-trained model. We evaluate with 792 magnetic sequences consecutively, which takes around 79 seconds. Therefore, the average time consumption of running a location query with 500 geomagnetic samples is around 0.1 second. Table 4 presents the average time consumption of a trial case in our experiment. It shows that the proposed MAIL reduces time consumption by more than 56% compared with Magicol [34] and MaLoc [46]. This is because MAIL extracts sequential features by a pre-trained model, which is more computationally efficient. As MAIL extracts multi-scale features from the input geomagnetic sequence, its time consumption is slightly larger (0.04s) than Jang et al. [20]. However, by extracting multi-scale features, MAIL improves the localization accuracy by a large margin.

Real-Time Localization. It takes around 0.1 second to estimate the location given a magnetic sequence of length 500. Therefore, our model is capable of giving around 10 location estimations per second. However, our system cannot give so many location estimations per second due to the limited sampling frequency (50Hz) of the magnetometer. Despite the limitation of the magnetic sensor, it is possible to facilitate localization by data
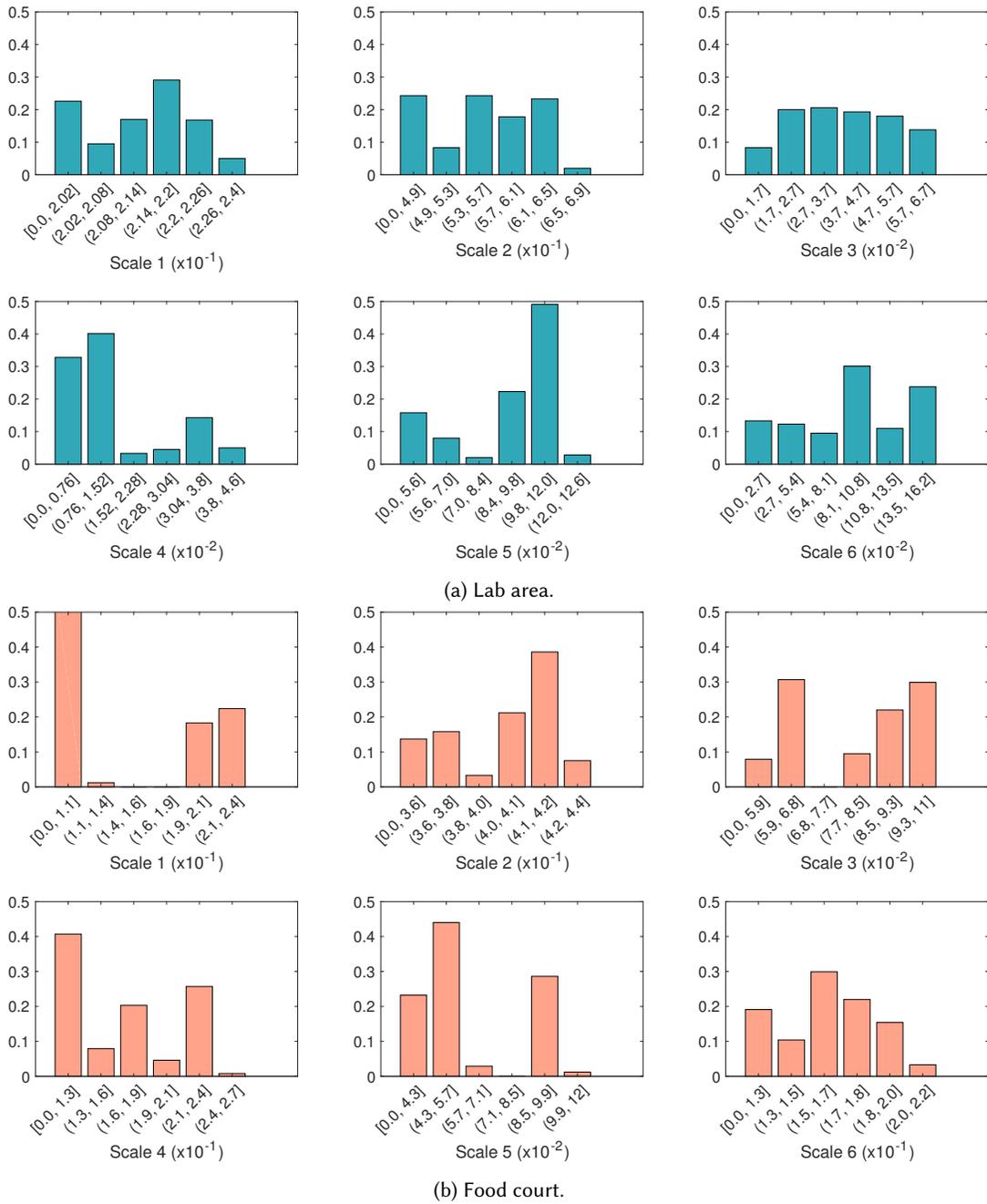
(a) Lab area.

(b) Food court.

Fig. 18. Detailed distribution of attention values at each scale in our trial sites.
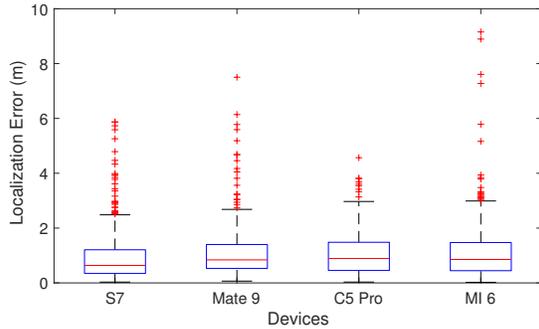
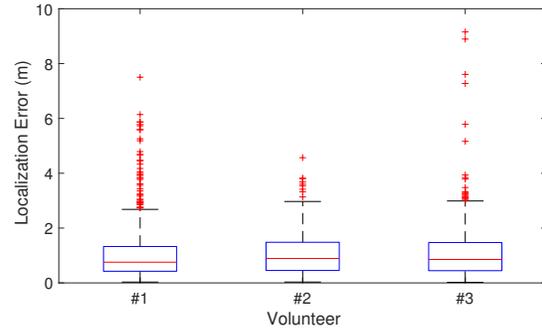Fig. 19. Localization error with different devices (lab area).

Fig. 20. Localization error with different volunteers (lab area).

Table 5. Total and average energy consumption of different trial devices.

Device	Packages	Total (mAh)	Average (mAh)
Samsung S7	2658	45	0.0169
Samsung C5 Pro	2469	27	0.011
Xiaomi MI	1915	66	0.034

streaming, where we establish a persistent connection between a client and a server. The client uploads instant magnetic readings continuously. Upon receiving a or several magnetic samples, the server appends new samples to the end of a received magnetic sequence and infers current location using the latest 500 readings. By data streaming, MAIL can give several location estimations in one second, thus achieving real-time localization (more than once per second), which is sufficient for pedestrian or robot localization indoors.

Energy Consumption. We present the average energy consumption of collecting and sending a package in Table 5. Evaluation results show that the power consumption of our client is marginal compared with the power capacity of state-of-the-art mobile phones (thousands of mAh).

6 DISCUSSIONS

In this paper, we propose a multi-scale approach to detect various anomalies. Although we have managed to extract fine-grained location features and achieve accuracy in different trial sites, a few more practical challenges (not the focus of the paper) remain to be addressed.

- *Generality to other signals.* Besides geomagnetism, other signals, such as visible light, Bluetooth and Wi-Fi, present similar signal variations as well, mainly due to the multi-path propagation and various power levels. For example, a wireless access point with higher power levels may cover a larger area while a lower one covers a smaller area. Therefore, these access points provide wireless anomalies in various scales. Consequently, it is possible to study variations of other signals and adapt our network to these signal sequences as well.

- *Crowdsourcing geomagnetism for localization.* Despite its accuracy, MAIL relies on a given geomagnetic map for accurate indoor localization. However, it is usually time-consuming and labor-intensive to collect geomagnetic readings in a large site, e.g., a multi-story grand shopping mall. Additionally, the cost of conducting a site survey by professional surveyors increases dramatically at the metropolitan scale. To

improve the deployability of localization systems, recent researches study crowdsourcing by volunteers to construct the signal map, e.g., Wi-Fi fingerprint. It is also possible to study crowdsourcing techniques which combine different information, such as distinguishing signal patterns [10] or user trajectories [51] to label collected geomagnetic signals.

- *Deployment in mobile devices.* MAIL is based on the client-server architecture, where a client program collects magnetic readings and sends them to a remote server for localization. This incurs round-trip delay, as well as privacy concerns from users. Motivated by the recent development of software and hardware, it is possible to study efficient deployment in mobile devices without server support. For example, recent deep learning frameworks have adapted to resource-constrained mobile platforms (e.g., TensorFlow Lite[3]). Furthermore, state-of-the-art mobile phones are equipped with graphical processing units (e.g., Google Pixel 3[4], Xiaomi Mi 8[5]). Combined with the above, it is possible to fine-tune a model and deploy it in a mobile device.

7 CONCLUSION

In this paper, we propose a multi-scale attention-guided indoor localization network, termed MAIL. Instead of using raw input readings from magnetometers, we leverage the gradient of magnetic sequences to reduce the impact of device heterogeneity. In order to detect local anomalies in various scales, we propose an SFE unit, which detects scale-specific anomalies and extracts features accordingly. Then, we employ several SFE units to extract various features at different scales. Due to environmental factors, different scales contribute differently to the location estimation. To achieve high applicability, we propose an attention generation unit that learns the prominence of each scale adaptively. By paying attention to more prominent scales, we fuse extracted multiple scale-based features together to generate a more distinguishing location feature and estimate the user position. We have conducted extensive experiments in three different trial sites. Experimental results show that MAIL reduces the mean localization error by more than 36% compared with state-of-the-art competing schemes.

ACKNOWLEDGMENTS

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 (July 2016), 1770–1782.

[2] Yomna Abdelrahman, Anam Ahmad Khan, Joshua Newn, Eduardo Velloso, Sherine Ashraf Safwat, James Bailey, Andreas Bulling, Frank Vetere, and Albrecht Schmidt. 2019. Classifying attention types with thermal imaging and eye tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3, Article 69 (Sept. 2019).

[3] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. 2016. Layer normalization. *arXiv preprint arXiv:1607.06450* (2016), 1–14.

[4] Donald J. Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*. AAAI Press, 359–370.

[5] Giuseppe Caso, Luca De Nardis, Filip Lemic, Vlado Handziski, Adam Wolisz, and Maria-Gabriella Di Benedett. Online 2019. ViFi: Virtual fingerprinting WiFi-based indoor positioning via multi-wall multi-floor propagation model. *IEEE Transactions on Mobile Computing* (Online 2019).

[6] Muhammad Aamir Cheema. 2018. Indoor location-based services: Challenges and opportunities. *SIGSPATIAL Special* 10, 2 (Nov. 2018), 10–17.

---

[3]https://www.tensorflow.org/lite

[4]https://store.google.com/us/product/pixel_3_specs?hl=en-US

[5]https://www.mi.com/global/mi8/specs/

[7] Huijie Chen, Fan Li, Xiaojun Hei, and Yu Wang. 2018. CrowdX: Enhancing automatic construction of indoor floorplan with opportunistic encounters. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4, Article 159 (Dec. 2018), 21 pages.

[8] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 1724–1734.

[9] Jaewoo Chung, Matt Donahoe, Chris Schmandt, Ig-Jae Kim, Pedram Razavai, and Micaela Wiseman. 2011. Indoor location sensing using geo-magnetism. In *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*. ACM, 141–154.

[10] Moustafa M. Elhamshary, Moustafa F. Alzantot, and Moustafa Youssef. 2019. JustWalk: A crowdsourcing approach for the automatic construction of indoor floorplans. *IEEE Transactions on Mobile Computing* 18, 10 (Oct 2019), 2358–2371.

[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, Article 9 (March 2018), 25 pages.

[12] Klaus Greff, Rupesh K. Srivastava, Jan Koutník, Bas R. Steunebrink, and Jürgen Schmidhuber. 2017. LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems* 28, 10 (Oct 2017), 2222–2232.

[13] Simon Haykin. 1994. *Neural Networks*. Vol. 2. Prentice Hall New York.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In *Proceedings of the 2015 International Conference on Computer Vision*. IEEE, 1026–1034.

[15] Suining He and Kang G. Shin. 2017. Geomagnetism for Smartphone-Based Indoor Localization: Challenges, Advances, and Comparisons. *ACM Computing Surveys* 50, 6, Article 97 (Dec. 2017), 37 pages.

[16] Suining He and Kang G. Shin. 2019. Crowd-Flow Graph Construction and Identification with Spatio-Temporal Signal Feature Fusion. In *IEEE Conference on Computer Communications*. IEEE, 757–765.

[17] 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*. IEEE, 1–7.

[18] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computing* 9, 8 (Nov. 1997), 1735–1780.

[19] Baoqi Huang, Zhendong Xu, Bing Jia, and Guoqiang Mao. 2019. An online radio map update scheme for WiFi fingerprint-based localization. *IEEE Internet of Things Journal* 6, 4 (Aug 2019), 6909–6918.

[20] Ho Jun Jang, Jae Min Shin, and Lynn Choi. 2017. Geomagnetic field based indoor localization using recurrent neural networks. In *Proceedings of the 2017 Global Communications Conference*. IEEE, 1–6.

[21] Dejiang Kong and Fei Wu. 2018. HST-LSTM: A hierarchical spatial-temporal long-short term memory network for location prediction. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, 2341–2347.

[22] Tarun Kulshrestha, Divya Saxena, Rajdeep Niyogi, and Jiannong Cao. 2020. Real-time crowd monitoring using seamless indoor-outdoor localization. *IEEE Transactions on Mobile Computing* 19, 3 (March 2020), 664–679.

[23] 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, Article 17 (March 2018), 28 pages.

[24] Bing Li, Juan Pablo Muñoz, Xuejian Rong, Qingtian Chen, Jizhong Xiao, Yingli Tian, Aries Arditi, and Mohammed Yousuf. 2019. Vision-based mobile indoor assistive navigation aid for blind people. *IEEE Transactions on Mobile Computing* 18, 3 (March 2019), 702–714.

[25] 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, Article 120 (Sept. 2018), 25 pages.

[26] Wenping Liu, Hongbo Jiang, Guoyin Jiang, Jiangchuan Liu, Xiaoqiang Ma, Yufu Jia, and Fu Xiao. 2019. Indoor navigation with virtual graph representation: Exploiting peak intensities of unmodulated luminaries. *IEEE/ACM Transactions on Networking* 27, 1 (Feb 2019), 187–200.

[27] Zhihong Luo, Qiping Zhang, Yunfei Ma, Manish Singh, and Fadel Adib. 2019. 3D backscatter localization for fine-grained robotics. In *Proceedings of the 16th USENIX Conference on Networked Systems Design and Implementation*. USENIX Association, 765–782.

[28] Sylvie Nadeau, Martina Betschart, and Francois Bethoux. 2013. Gait analysis for poststroke rehabilitation: The relevance of biomechanical analysis and the impact of gait speed. *Physical Medicine and Rehabilitation Clinics of North America* 24, 2 (2013), 265 – 276.

[29] 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* 15, 2, Article 19 (Feb. 2019), 31 pages.

[30] 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, Article 99 (Sept. 2019), 24 pages.

[31] Qun Niu, Ying Nie, Suining He, Ning Liu, and Xiaonan Luo. 2018. RecNet: A Convolutional Network for Efficient Radiomap Reconstruction. In *2018 IEEE International Conference on Communications*. IEEE, 1–7.

[32] Valter Pasku, Alessio De Angelis, Guido De Angelis, Darmindra D. Arumugam, Marco Dionigi, Paolo Carbone, Antonio Moschitta, and David S. Ricketts. 2017. Magnetic field-based positioning systems. *IEEE Communications Surveys & Tutorials* 19, 3 (thirdquarter 2017), 2003–2017.

[33] Sihua Shao, Abdallah Khreishah, and Issa Khalil. 2020. Enabling real-time indoor tracking of IoT devices through visible light retroreflection. *IEEE Transactions on Mobile Computing* 19, 4 (2020), 836–851.

[34] 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 (July 2015), 1443–1457.

[35] Yuanchao Shu, Kang G. Shin, Tian He, and Jiming Chen. 2015. Last-mile navigation using smartphones. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 512–524.

[36] Michael Stocker, Bernhard Grosswindhager, Carlo Alberto Boano, and Kay Romer. 2019. SnapLoc: An ultra-fast UWB-based indoor localization system for an unlimited number of tags: Demo abstract. In *Proceedings of the 18th International Conference on Information Processing in Sensor Networks*. ACM, 348–349.

[37] Kalyan Pathapati Subbu, Brandon Gozick, and Ram Dantu. 2013. LocateMe: Magnetic-fields-based indoor localization using smartphones. *ACM Transactions on Intelligent Systems and Technology* 4, 4 (Oct 2013), 73:1–73:27.

[38] Xiaoqiang Teng, Deke Guo, Yulan Guo, Xiaolei Zhou, and Zhong Liu. 2019. CloudNavi: Toward ubiquitous indoor navigation service with 3D point clouds. *ACM Transactions on Sensor Networks* 15, 1, Article 1 (Jan. 2019), 28 pages.

[39] Kenta Urano, Kei Hiroi, Takuro Yonezawa, and Nobuo Kawaguchi. 2019. Basic study of BLE indoor localization using LSTM-based neural network (Poster). In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 558–559.

[40] Ju Wang, Liqiong Chang, Omid Abari, and Srinivasan Keshav. 2019. Are RFID sensing systems ready for the real world?. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 366–377.

[41] Zhongqin Wang, Min Xu, Ning Ye, Ruchuan Wang, and Haiping Huang. 2019. RF-Focus: Computer vision-assisted region-of-interest RFID tag recognition and localization in multipath-prevalent environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1, Article 29 (March 2019), 30 pages.

[42] Hang Wu, Suining He, and Gary S.-H. Chan. 2017. A Graphical Model Approach for Efficient Geomagnetism-Pedometer Indoor Localization. In *2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems*. IEEE, 371–379.

[43] Hang Wu, Suining He, and S.-H. Gary Chan. 2017. Efficient Sequence Matching and Path Construction for Geomagnetic Indoor Localization. In *Proceedings of the 2017 International Conference on Embedded Wireless Systems and Networks*. Junction Publishing, 156–167.

[44] 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* 15, 4, Article 42 (Aug. 2019), 25 pages.

[45] Tz-Ying Wu, Ting-An Chien, Cheng-Sheng Chan, Chan-Wei Hu, and Min Sun. 2017. Anticipating daily intention using on-wrist motion triggered sensing. In *Proceedings of the 2017 International Conference on Computer Vision*. IEEE, 48–56.

[46] Hongwei Xie, Tao Gu, Xianping Tao, Haibo Ye, and Jian Lv. 2014. MaLoc: A practical magnetic fingerprinting approach to indoor localization using smartphones. In *Proceedings of the 2014 International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 243–253.

[47] Huatao Xu, Dong Wang, Run Zhao, and Qian Zhang. 2019. AdaRF: Adaptive RFID-based indoor localization using deep learning enhanced holography. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3, Article 113 (Sept. 2019), 22 pages.

[48] Shiyang Yan, Jeremy S. Smith, Wenjin Lu, and Bailing Zhang. 2018. Hierarchical multi-scale attention networks for action recognition. *Signal Processing: Image Communication* 61 (2018), 73 – 84.

[49] Xuehan Ye, Shuo Huang, Yongcai Wang, Wenping Chen, and Deying Li. 2019. Unsupervised localization by learning transition model. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2, Article 65 (June 2019), 23 pages.

[50] Xuehan Ye, Yongcai Wang, Yuhe Guo, Wei Hu, and Deying Li. 2018. Accurate and efficient indoor location by dynamic warping in sequence-type radio-map. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1, Article 50 (March 2018), 22 pages.

[51] Faheem Zafari, Athanasios Gkelias, and Kin K. Leung. 2019. A survey of indoor localization systems and technologies. *IEEE Communications Surveys & Tutorials* 21, 3 (thirdquarter 2019), 2568–2599.

[52] Chi Zhang, Kalyan P. Subbu, Jun Luo, and Jianxin Wu. 2015. GROPING: Geomagnetism and crowdsensing powered indoor navigation. *IEEE Transactions on Mobile Computing* 14, 2 (Feb 2015), 387–400.

[53] Huanhuan Zhang, Anfu Zhou, Dongzhu Xu, Shaoqing Xu, Xinyu Zhang, and Huadong Ma. 2019. Learning to recognize unmodified lights with invisible features. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2, Article 67 (June 2019), 23 pages.

[54] 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 (Oct 2017), 2655–2669.