End-to-End Sequential Indoor Localization Using Smartphone Inertial Sensors and WiFi

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

02-12-2021 / 10-12-2021

CITATION

Nurpeiissov, Mukhamet; Kuzdeuov, Askat; Assylkhanov,, Aslan; Khassanov, Yerbolat; Varol, Hüseyin Atakan (2021): End-to-End Sequential Indoor Localization Using Smartphone Inertial Sensors and WiFi. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.17109275.v1>

DOI

[10.36227/techrxiv.17109275.v1](https://doi.org/10.36227/techrxiv.17109275.v1)

End-to-End Sequential Indoor Localization Using Smartphone Inertial Sensors and WiFi

Mukhamet Nurpeiissov, Askat Kuzdeuov, Aslan Assylkhanov, Yerbolat Khassanov, and Huseyin Atakan Varol

Abstract—This paper addresses sequential indoor localization using WiFi and Inertial Measurement Unit (IMU) modules commonly found in commercial off-the-shelf smartphones. Specifically, we developed an end-to-end neural network-based localization system integrating WiFi received signal strength indicator (RSSI) and IMU data without external data fusion models. The developed system leverages the advantages of WiFi and IMU modules to locate finer-level sequential positions of a user at 150 Hz sampling rate. Additionally, to demonstrate the efficacy of the proposed approach, we created the IMUWiFine dataset comprising IMU and WiFi RSSI readings sequentially collected at fine-level reference points. The dataset contains 120 trajectories covering an aggregate distance of over 14 kilometers. We conducted extensive experiments using deep learning models and achieved a mean error distance of 1.1 meters on an unseen evaluation set, which makes our approach suitable for many practical applications requiring meter-level accuracy. To enable experiment and result reproducibility, we made the developed localization system and IMUWiFine dataset publicly available in our GitHub repository¹.

I. INTRODUCTION

Indoor localization is an enabling technology for many emerging applications, including personalized advertisement, augmented reality, and navigation. Despite more than a decade of intensive work, it remains an unsolved problem [1]. Specifically, there does not seem to exist a robust and accurate technology or a combination of technologies that can recreate the experience that satellite-based radionavigation systems offer outdoors. Nevertheless, many effective indoor localization approaches have been presented recently and promising results have been achieved [2], [3], [4], [5], [6].

Depending on the hardware deployment requirements, the indoor localization approaches can be categorized into two main groups: 1) infrastructure-based and 2) infrastructure-free. Infrastructure-based approaches require the deployment of custom hardware, such as Bluetooth beacons [7], magnetic resonators [8], and ultrasound speakers [9]. On the other hand, infrastructure-free approaches do not rely on custom hardware except existing infrastructure (e.g., WiFi network) and built-in sensors present on commercial off-the-shelf devices (e.g., accelerometer, gyroscope, and magnetometer sensors in smartphones and tablets). Although infrastructure-based approaches can achieve higher accuracy, the deployment overhead might be prohibitive (e.g., costly,

time-consuming, labor-intensive, and intrusive), which makes them less attractive for practical use.

Alternatively, infrastructure-free approaches, such as WiFi fingerprinting and inertial measurement unit (IMU) positioning, are considered more convenient and cost-effective thanks to the ubiquitous wireless access point (AP) infrastructure and built-in sensors in commodity mobile devices. In the past few years, infrastructure-free approaches have attracted substantial interest [4], [5], [6]. For example, in the recently organized Microsoft indoor localization competition, two out of three top systems were infrastructure-free methods [1]. All the three methods achieved below two-meter accuracy, where the gap between the best infrastructure-free (1.56 m) and infrastructure-based (0.72 m) methods was not substantial. Given that the former did not utilize any additional custom hardware, this result is remarkable. The best infrastructure-free methods were based on WiFi received signal strength indicator (RSSI) fingerprinting and IMU positioning.

WiFi RSSI fingerprinting relies on the existing WiFi infrastructure and exploits the received signal strength from surrounding APs to infer the unknown user location. Specifically, it utilizes a number of signal fingerprints collected a priori to build the so-called radio map. Subsequently, the user's location is estimated with an indoor localization algorithm that exploits the currently measured fingerprint and the fingerprints in the radio map. However, the WiFi RSSI fingerprinting approaches are criticized for weak performance in areas with poor WiFi coverage and dynamic environments (e.g., areas with frequent human movement and furniture rearrangement) [4]. Moreover, WiFi RSSI values have been shown to have a high variance, even in a static environment with a still user, which is further exacerbated by the multipath effect [10]. Therefore, methods based on WiFi RSSI fingerprinting might be unsuitable for applications where a robust and higher granularity of location information is required.

On the other hand, IMU positioning relies on multi-modal sensor fusion of accelerometer, gyroscope, and magnetometer data for estimating the user's location. IMU positioning employs different types of pedestrian dead reckoning (PDR) which infers the current position of a user based on the estimated previous position, walking speed, and orientation. Compared to the WiFi-based localization techniques, IMU positioning is considered more autonomous (i.e., unconstrained coverage area and does not depend on the other external infrastructure) and accurate for a short distance. One of the main disadvantages of IMU positioning is the drifting error, which accumulates with the walking distance.

M. Nurpeiissov, A. Kuzdeuov, A. Assylkhanov, Y. Khassanov, and H.A. Varol are with the Institute of Smart Systems and Artificial Intelligence, Nazarbayev University, Nur-Sultan City, Kazakhstan. E-mail: {mukhamet.nurpeiissov, askat.kuzdeuov, aslan.assylkhanov, yerbolat.khassanov, ahvarol}@nu.edu.kz.

¹<https://github.com/IS2AI/IMUWiFine>

To alleviate the limitations of WiFi RSSI- and IMU-based indoor localization methods, this paper focuses on the hybrid indoor localization problem that combines WiFi RSSI fingerprinting and IMU positioning. However, unlike the previous hybrid approaches, we address finer-level sequential indoor localization, i.e., continuous position tracking at higher sampling rates (150 Hz in our case). Additionally, our hybrid system is an end-to-end sequence learning algorithm based on the recurrent neural networks (RNNs) with the long short-term memory (LSTM) [11] units that implicitly fuses input signals from the WiFi and IMU sensors and outputs three-dimensional position coordinates.

To validate our hybrid approach, we constructed a dataset comprising WiFi RSSI and IMU readings. We named our dataset IMUWiFine, and different from the existing datasets, it was collected sequentially at finer-level reference points. In total, the dataset contains 120 trajectories recorded across multiple floors of a 9,000 m^2 building. The experiment results conducted on the IMUWiFine dataset show that the best hybrid system can achieve a 1.1 m mean error distance (MED) on unseen trajectories, which is practical for many applications requiring meter-level accuracy.

The rest of the paper is organized as follows: Section II reviews the related indoor localization work and datasets. Section III describes the construction procedures of the IMUWiFine dataset and presents its specifications. Our end-to-end deep learning-based hybrid localization approach is explained in Section IV. The experimental setup and obtained results are provided in Section V. Lastly, Section VI concludes this paper.

II. RELATED WORK

Over the last few decades, academia and industry have devoted significant effort and resources to solving the indoor localization problem [2], [3], [4], [5], [6]. However, the research community has not converged to a single widely accepted solution that can achieve the desired accuracy at a low cost [1]. Despite that, in quest of the optimal solution, many ingenious approaches have been proposed with sufficient accuracy and deployment cost for some practical applications.

The initial indoor localization approaches were single module systems utilizing WiFi [5], IMU [12], Bluetooth [7], RFID [13], ultrasound [9], LED light [14], and other sensors [6]. To achieve higher accuracy, later works proposed the combination of two or more modules to construct hybrid indoor localization systems [15], [16], [17]. The hybrid systems were designed to leverage the advantage of the constituent modules while maintaining affordable cost and reliable performance. For example, Poulou and Han [16] created a hybrid system combining a smartphone camera with an IMU sensor. In particular, they used a Kalman filter to explicitly fuse the predicted positions from two isolated localization systems based on the camera and IMU. Similarly, other researchers [17] combined WiFi, IMU, and iBeacon based predictions using a particle filter. These approaches are usually designed as a pipeline of separately

trained localization systems (i.e., one system per module) and an external fusion model.

With the emergence of new applications, such as personalized advertisement and augmented reality, robust indoor localization systems with a much higher granularity of location information are required (e.g., finer-level continuous positions), which is considered more challenging. To address this problem, the authors of [18] attempted finer-level sequential indoor localization using WiFi RSSI fingerprinting with the position tracking sampling rate of 0.2 Hz and achieved three-meter accuracy. Although this can be used for applications, such as tracking the user's movement across different rooms in a building (e.g., shopping mall boutiques, hospital wards, and so on), it is insufficient for applications requiring meter-level accuracy and a higher position tracking sampling rate. Therefore, in this work, we present a hybrid approach that offers finer-level sequential localization by supplementing WiFi with the IMU module.

Unfortunately, there is no dataset suitable for assessing finer-level sequential hybrid indoor localization. Most of the existing datasets were collected in a discrete manner (i.e., at arbitrary reference points), where temporal information is missing [19], [20], [21], [22], [23]. While sequentially collected datasets either contain readings only from a single module (e.g., WiFi only) [18] or the distance between the consequent reference points is too long (i.e., not finer-level) [24], [25]. Therefore, we collected a new dataset named IMUWiFine containing a sequentially recorded stream of IMU and WiFi RSSI values and the corresponding positions. Finer-level granularity was achieved thanks to computer vision techniques, as will be described in the following section.

III. IMUWIFINE DATASET CONSTRUCTION

In this section, we present our IMUWiFine dataset comprising IMU and WiFi RSSI data readings recorded in sequential order with a fine spatiotemporal resolution. The dataset was collected on the fourth, fifth, and sixth floors of the C4 building at the Nazarbayev University campus. The total covered area is over 9,000 m^2 throughout the three floors. To enable experiment reproducibility and facilitate research in sequential indoor localization, the IMUWiFine dataset and other supplementary materials are made publicly available¹ under the Creative Commons Attribution 4.0 International License².

A. The Data Collection Setup

To record WiFi RSSI and IMU data streams, we used the built-in sensors of a smartphone (OPPO A5) and developed an Android application that periodically read and stored the records. Specifically, the application regularly scanned surrounding APs (around every five seconds) and stored WiFi RSSI values with the timestamps in a log file. Similarly, the IMU data consisting of accelerometer, gyroscope, and magnetometer readings were sampled with a 150 Hz rate and stored in the log file.

²<https://creativecommons.org/licenses/by/4.0/>

To identify the corresponding (x, y, z) positions, we used $14\text{ cm} \times 14\text{ cm}$ sized ArUco fiducial markers [26] attached to the walls and ceiling along the trajectories (see Fig. 1). The markers were generated using ArUco’s original dictionary for the indoor localization task. In total, 654 ArUco markers were placed using special 3D-printed holders. We calibrated marker positions in relation to the global frame using a total station (Leica TS06 Plus) (see Fig. 1b). The positions with the corresponding timestamps were recorded at 3 Hz.

Data were collected by two persons who used the same smartphone in turn. Specifically, the users held the smartphone with the camera pointing forward to capture the ArUco markers located in front (see Fig. 1a). The Android application employed OpenCV’s ArUco library to estimate the user’s local position in relation to the detected markers, and then, transferred it to the global reference frame (i.e., position in relation to the origin). The users walked at a self-selected pace with occasional stops and turns. The trajectories were planned in advance to imitate realistic walking paths and to make them dissimilar. The dataset collection process took around two weeks. In total, we collected 120 trajectories on different days and times of the day.

B. Data Preprocessing

During the dataset collection, 351 distinct APs were identified by their unique media access control (MAC) address. Since some of the APs appeared only for a short period of time, we removed the APs which were detected less than 500 times. As a result, 220 distinct APs were kept.

Since the sampling rates of WiFi RSSI, IMU, and position data streams are different, we upsampled the frequency of WiFi RSSI and position to match the IMU data. To upsample the WiFi RSSI readings, we employed the zero-order hold signal reconstruction method. To upsample the positions, we linearly interpolated the consecutive position points. As a result, the readings of the three streams were synchronized at a sampling rate of 150 Hz.

Fig. 1: The ArUco markers placement. (a) The smartphone display screenshot showing the detected ArUco markers and computed position. (b) The Leica TS06 Plus total station used to calibrate the ArUco markers on the walls and ceiling

C. The Dataset Specifications

The dataset specifications are provided in Table I. We split the collected trajectories into three non-overlapping sets: training, validation, and test. The validation and test set trajectories were collected a few days later than those of the training set to imitate the real-world use case.

The total size of the IMUWiFine dataset is around 1.2 GB. The trajectories are stored in separate ‘csv’ files, and we tried to make the structure of the dataset similar to those of the previously published indoor localization datasets [18], [19]. In particular, the number of rows in each ‘csv’ file corresponds to the number of recorded samples arranged in a temporal order. Each recorded sample was time-stamped using the mobile phone’s operating system time. The number of columns in all ‘csv’ files is equal to 234, and they contain the following information:

- 1-220: RSSI values of APs (dBm);
- 221-223: Gyroscope values of IMU - g_x, g_y, g_z ;
- 224-226: Accelerometer values of IMU - a_x, a_y, a_z ;
- 227-229: Magnetometer values of IMU - m_x, m_y, m_z ;
- 230: Timestamp of a recorded sample (s);
- 231-233: x, y , and z positions (m);
- 234: Floor ID - 0, 1, or 2.

TABLE I: The IMUWiFine dataset specifications

Specifications	Train	Valid	Test	Total
Number of trajectories	60	30	30	120
Number of samples	3.1M	1.2M	1.1M	5.4M
Total length (km)	8.0	3.1	3.0	14.2
Total duration (hours)	5.5	2.5	2.4	10.4

IV. END-TO-END SEQUENTIAL INDOOR LOCALIZATION

Our end-to-end sequential indoor localization system was realized using a sequence learning algorithm. Specifically, we used RNN-based models with the LSTM memory units which have been shown highly effective in learning sequential dependencies [27]. Subsequently, to predict the current position, the system utilizes both current and previously received inputs from both the WiFi and IMU modules. Unlike the previous works, our system is not given any explicit indication that the input is composed of readings from different modules. Therefore, our system implicitly learns to fuse the WiFi and IMU data, whereas traditional approaches usually predict positions using separate models first, and then combine predictions using another fusion model.

Additionally, in front of the LSTM layers, we prepended non-linear layers with the rectified linear unit (ReLU) activation function. As a final output layer, we employed a multivariate regression function that can output more than one continuous variable. In contrast to the classification output, the regression output can generate finer-level positions and can be easily scaled to a larger area. It should be noted that our system does not utilize previously predicted positions.

A. Deep Learning Architecture

The architecture of our end-to-end sequential indoor localization system is depicted in Fig. 2. The system is based on the stack of ReLU, LSTM, and regression layers. Specifically, at each time step t the ReLU layer takes an input feature vector $\mathbf{x}_t$ and outputs a new feature representation $\mathbf{x}'_t$:

$$\mathbf{x}'_t = \text{ReLU}(\mathbf{W}_1 \mathbf{x}_t + \mathbf{b}_1) \quad (1)$$

where the weight matrix $\mathbf{W}_1$ and the bias vector $\mathbf{b}_1$ are learnable parameters. The input feature vector $\mathbf{x}_t = [\mathbf{w}_t, \mathbf{u}_t]$ is constructed by concatenating the WiFi RSSI $\mathbf{w}_t$ and IMU $\mathbf{u}_t$ records, where $\mathbf{w}_t \in \mathbb{R}^{220}$ is a 220-dimensional vector of RSSI values received from 220 distinct APs, $\mathbf{u}_t \in \mathbb{R}^9$ is a 9-dimensional vector of accelerometer, gyroscope, and magnetometer values.

The new feature representation vector $\mathbf{x}'_t$ is then fed to the unidirectional LSTM layer which combines it with a hidden state vector from the previous time step $\mathbf{h}_{t-1}$ to produce current time step hidden state vector $\mathbf{h}_t$:

$$\mathbf{h}_t = \text{LSTM}(\mathbf{x}'_t, \mathbf{h}_{t-1}) \quad (2)$$

As a result, the hidden state vector $\mathbf{h}_t$ will incorporate information from both prior and current inputs (i.e., $[\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_t]$ and $[\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_t]$). The initial hidden state vector $\mathbf{h}_0$ is initialized as a zero vector.

Next, the hidden state feature vector $\mathbf{h}_t$ is fed to the multivariate regression function to estimate the current position $\hat{\mathbf{p}}_t \in \mathbb{R}^3$ as follows:

$$\hat{\mathbf{p}}_t = \mathbf{W}_2 \mathbf{h}_t + \mathbf{b}_2 \quad (3)$$

where $\mathbf{W}_2$ and $\mathbf{b}_2$ are learnable parameters. The predicted position $\hat{\mathbf{p}}_t = (x, y, z)$ is given in meters as a relative distance from the origin, where x and y correspond to the horizontal (surface) position, and z corresponds to the vertical position.

Finally, the overall network architecture was jointly optimized in an end-to-end fashion using the mean squared error (MSE) loss function:

$$\mathcal{L}_{\text{MSE}}(\mathbf{p}_t, \hat{\mathbf{p}}_t) = \frac{1}{T} \sum_{t=0}^T \|\mathbf{p}_t - \hat{\mathbf{p}}_t\|_2 \quad (4)$$

where $\mathbf{p}_t$ is a reference position, and T is the trajectory length.

Fig. 2: End-to-end sequential indoor localization model architecture

V. INDOOR LOCALIZATION EXPERIMENT

In this section, we first describe the applied performance evaluation criterion and feature preprocessing, followed by the experimental setup, results, and analysis.

A. Evaluation

As the performance evaluation measure, we used the mean error distance (MED) defined as:

$$\text{MED} = \frac{\sum_i \text{dist}(\mathbf{p}_i, \hat{\mathbf{p}}_i)}{N} \quad (5)$$

where $\text{dist}(\cdot)$ is the Euclidean distance between the reference position $\mathbf{p}_i$ and the predicted position $\hat{\mathbf{p}}_i$. N is the number of samples. Since our evaluation sets have several trajectories, MED was computed over all the trajectories combined.

B. Feature Preprocessing

The previous works have demonstrated that feature preprocessing can highly influence localization performance [28], [29]. Hence, in our experiments, we preprocessed the WiFi RSSI readings into the so-called ‘powred’ representation. To the IMU readings, we applied min-max normalization, as will be thoroughly explained next.

1) *WiFi RSSI preprocessing*: The ‘powred’ feature representation for WiFi RSSI values has been shown to achieve the best results for most of the localization algorithms [28]. This feature representation is obtained as follows. The RSSI values in the IMUWiFine dataset range from -93 dBm to -34 dBm corresponding to the weakest and strongest signals, respectively. Therefore, we first set the RSSI values of undetected APs to -100 dBm. Then, we shifted all the RSSI values into the positive range:

$$w_{\text{positive}}^i = w^i + 100 \quad (6)$$

where w^i is the raw RSSI value of i -th AP.

Next, we scaled the positive RSSI values, by dividing them by 100, and raised to the power of β :

$$w_{\text{powred}}^i = \left(\frac{w_{\text{positive}}^i}{100} \right)^\beta \quad (7)$$

where β was set to Euler’s number e , as suggested in [28], [29]. As a result, RSSI values closer to 1 represent stronger signals, whereas 0 indicates that an AP is not detected.

2) *IMU preprocessing*: To match the range of the IMU readings with the WiFi RSSI data, we applied min-max normalization. Specifically, we normalized each component of the IMU data $\mathbf{u}_t = [a_x, a_y, a_z, g_x, g_y, g_z, m_x, m_y, m_z]$ separately. For example, the x -axis accelerometer values were processed as follows:

$$a'_x = \frac{a_x - \min(a_x)}{\max(a_x) - \min(a_x)} \quad (8)$$

where a_x is the original value, and a'_x is the normalized value. The minimum $\min(\cdot)$ and maximum $\max(\cdot)$ values of a given component were obtained among all trajectories of the training set. The same procedure was applied to the remaining components of the IMU data. As a result, the values of each component were transformed into a range between 0 and 1.

C. Indoor Localization Model Specifications

The RNN-based indoor localization model was implemented using the PyTorch library [30]. All models were trained using our IMUWiFine dataset on V100 GPUs running in an NVIDIA DGX-2 server. All hyperparameters were tuned on the validation set, and the final model was evaluated on the test set of IMUWiFine. The models were trained to minimize MSE loss using Adam optimizer with a momentum for 400 epochs. We set the initial learning rate to 0.003. To prevent overfitting, we tried different dropout rates (0, 0.1, 0.2) and report only the best performing one, i.e., each model was trained three times. Additionally, we set the sequence length to 300 for the truncated backpropagation through time. We implemented several RNN-based models with a different number of ReLU and LSTM layers as well as layer sizes, as will be shown in the following section. Further implementation details can be found in our GitHub repository¹.

D. Experiment Results

We first evaluated the RNN-based models without prepending ReLU layers. In particular, we trained several models constructed by stacking a different number of LSTM layers with similar dimensions. We set the dimensions of the LSTM layers to 229, which corresponds to the input feature vector size. The experiment results are given in Table II, and they show that the best performance is achieved by the model with three LSTM layers. Specifically, the best MED scores on the validation and test sets are 4.50 *m* and 4.45 *m*, respectively.

TABLE II: The MED performance of RNN-based sequential indoor localization models without ReLU layers

ReLU layers	LSTM layers	LSTM layer dimension	Model parameters	MED	
				Valid	Test
0	1	229	1.3M	8.88	9.13
	2		2.2M	5.12	5.24
	3		3.0M	4.50	4.45
	4		3.8M	6.99	7.08

Next, we evaluated the RNN-based models with the prepended ReLU layers. In this experiment, we varied both the number of ReLU and LSTM layers and layer dimension sizes. Specifically, we increased the number of the prepended ReLU layers up to five and the number of the LSTM layers up to four. For the layer dimensions, we tried the following sizes: 32, 64, 128, 256, and 512.

The experiment results are given in Table III. The best MED performance was achieved by the model consisting of four ReLU and three LSTM layers with a dimension size of 128. We do not show the MED performance for the other ReLU layer sizes due to space considerations. The general MED performance trend among different ReLU layer sizes is similar. Specifically, we observed that increasing the number of ReLU and LSTM layers keeps improving the MED performance until they reach four and three respectively, after

which the performance starts to degrade. Similarly, in all cases, we found that increasing the layer dimension size is also effective until it reaches 128 or 256. Overall, the best MED performance on both the validation and test sets is 1.13 *m*. These results demonstrate the high efficacy of the proposed end-to-end sequential indoor localization approach.

TABLE III: The MED performance of RNN-based sequential indoor localization models with prepended ReLU layers

ReLU layers	LSTM layers	ReLU and LSTM layer dimensions	Model parameters	MED	
				Valid	Test
4	1	32	32.7k	1.73	1.77
		64	114.6k	1.56	1.58
		128	425.7k	1.39	1.46
		256	1.6M	1.38	1.40
		512	6.4M	1.41	1.42
	2	32	49.6k	1.65	1.72
		64	181.1k	1.42	1.44
		128	689.9k	1.28	1.32
		256	2.7M	1.45	1.46
		512	10.6M	1.51	1.53
	3	32	66.5k	1.54	1.57
		64	247.7k	1.25	1.27
		128	954.1k	1.13	1.13
		256	3.7M	1.39	1.42
		512	14.8M	1.47	1.49
	4	32	83.4k	1.81	1.85
		64	314.2k	1.78	1.90
		128	1.2M	1.64	1.67
		256	4.8M	1.72	1.75
		512	19.0M	1.79	1.83

E. Ablation Analysis

In this experiment, we investigate the complementary nature of the WiFi and IMU modules for the finer-level sequential indoor localization task. The aim of this experiment is two-fold. First, we intend to verify that the MED performance improvements are due to the combination of the WiFi and IMU modules. Second, we want to measure the MED performance gain or loss resulting from the combination. To this end, we compared our best hybrid system (a model trained on the combined WiFi and IMU data) against a unimodal system (a model trained using only WiFi data). Additionally, to understand the contribution of different submodules of IMU (i.e., accelerometer, gyroscope, and magnetometer), we trained new indoor localization systems with different combinations of the submodules.

The experiment results are given in Table IV. The obtained results show that the integration of the IMU module reduces MED on the test set by 14% relative (from 1.32 *m* to 1.13 *m*), which demonstrates the complementary nature of the WiFi and IMU modules. Among different submodules of IMU, the most effective is the accelerometer followed by the gyroscope and then the magnetometer. Interestingly, we observed that integrating only the magnetometer submodule worsens the MED performance (from 1.32 *m* to 1.37 *m* on the test set), whereas the remaining combinations of the IMU submodules improve the performance over the unimodal system.

TABLE IV: The ablation analysis of the RNN-based sequential indoor localization model

Modules	MED	
	Valid	Test
WiFi	1.29 m	1.32 m
WiFi + Mag	1.35 m	1.37 m
WiFi + Gyr	1.28 m	1.30 m
WiFi + Acc	1.25 m	1.29 m
WiFi + Acc + Mag	1.19 m	1.22 m
WiFi + Gyr + Mag	1.22 m	1.21 m
WiFi + Acc + Gyr	1.16 m	1.18 m
WiFi + Acc + Gyr + Mag	1.13 m	1.13 m

Acc - accelerometer, Gyr - gyroscope, Mag - magnetometer.

VI. CONCLUSION

In this work, we presented a finer-level sequential indoor localization approach utilizing the WiFi and IMU modules. Specifically, we developed an end-to-end indoor localization system based on an RNN which implicitly fuses the WiFi RSSI and IMU data. To validate our approach, we constructed the IMUWiFine dataset comprising WiFi RSSI and IMU readings sequentially collected at finer-level reference points at a 150 Hz sampling rate. The IMUWiFine dataset consists of 120 trajectories collected across multiple floors with the help of ArUco markers, which substantially simplified the data collection process. The experimental results conducted on the IMUWiFine dataset show that the proposed approach can reach meter-level accuracy, which is sufficient for many practical applications. The implemented codes and dataset are freely available in our GitHub repository¹.

REFERENCES

- [1] D. Lymberopoulos and J. Liu, "The Microsoft indoor localization competition: Experiences and lessons learned," *IEEE Signal Process. Mag.*, vol. 34, no. 5, pp. 125–140, 2017.
- [2] H. Liu, H. Darabi, P. P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Syst. Man Cybern. Part C*, vol. 37, no. 6, pp. 1067–1080, 2007.
- [3] Z. Farid, R. Nordin, and M. Ismail, "Recent advances in wireless indoor localization techniques and system," *J. Comput. Networks Commun.*, pp. 185138:1–185138:12, 2013.
- [4] Z. Yang, C. Wu, Z. Zhou, X. Zhang, X. Wang, and Y. Liu, "Mobility increases localizability: A survey on wireless indoor localization using inertial sensors," *ACM Comput. Surv.*, vol. 47, no. 3, pp. 54:1–54:34, 2015.
- [5] C. Basri and A. E. Khadimi, "Survey on indoor localization system and recent advances of WIFI fingerprinting technique," in *Proc. International Conference on Multimedia Computing and Systems, ICMCS, Marrakech, Morocco, September 29 - October 1*, pp. 253–259, 2016.
- [6] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.
- [7] M. M. Organero, P. J. M. Merino, and C. D. Kloos, "Using Bluetooth to implement a pervasive indoor positioning system with minimal requirements at the application level," *Mob. Inf. Syst.*, vol. 8, no. 1, pp. 73–82, 2012.
- [8] G. Pirkel and P. Lukowicz, "Robust, low cost indoor positioning using magnetic resonant coupling," in *Proc. ACM Conference on Ubiquitous Computing, Ubicomp*, pp. 431–440, 2012.
- [9] A. Runge, M. Baunach, and R. Kolla, "Precise self-calibration of ultrasound based indoor localization systems," in *Proc. International Conference on Indoor Positioning and Indoor Navigation, IPIN, Guimarães, Portugal, September 21–23*, pp. 1–8, 2011.
- [10] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "FILA: Fine-grained indoor localization," in *Proc. IEEE INFOCOM, Orlando, FL, USA, March 25–30, 2012*, pp. 2210–2218, 2012.
- [11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [12] F. Li, C. Zhao, G. Ding, J. Gong, C. Liu, and F. Zhao, "A reliable and accurate indoor localization method using phone inertial sensors," in *Proc. ACM Conference on Ubiquitous Computing, Ubicomp, Pittsburgh, PA, USA, September 5–8, 2012*, pp. 421–430, 2012.
- [13] F. Seco and A. R. Jiménez, "Smartphone-based cooperative indoor localization with RFID technology," *Sensors*, vol. 18, no. 1, p. 266, 2018.
- [14] A. Naz, N. U. Hassan, M. A. Pasha, H. M. Asif, T. M. Jadoon, and C. Yuen, "Single LED ceiling lamp based indoor positioning system," in *Proc. IEEE World Forum on Internet of Things, WF-IoT, Singapore, February 5–8*, pp. 682–687, 2018.
- [15] M. Hasani, J. Talvitie, L. Sydänheimo, E.-S. Lohan, and L. Ukkonen, "Hybrid WLAN-RFID indoor localization solution utilizing textile tag," *IEEE Antennas and Wireless Propagation Letters*, vol. 14, pp. 1358–1361, 2015.
- [16] A. Poulou and D. S. Han, "Hybrid indoor localization using IMU sensors and smartphone camera," *Sensors*, vol. 19, no. 23, p. 5084, 2019.
- [17] H. Zou, Z. Chen, H. Jiang, L. Xie, and C. Spanos, "Accurate indoor localization and tracking using mobile phone inertial sensors, WiFi and iBeacon," in *Proc. IEEE International Symposium on Inertial Sensors and Systems (INERTIAL)*, pp. 1–4, 2017.
- [18] Y. Khassanov, M. Nurpeiissov, A. Sarkytbayev, A. Kuzdeuov, and H. A. Varol, "Finer-level sequential wifi-based indoor localization," in *Proc. IEEE/SICE International Symposium on System Integration, SII, Iwaki, Japan, January 11–14*, pp. 163–169, 2021.
- [19] J. Torres-Sospedra, R. Montoliu, A. M. Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems," in *Proc. International Conference on Indoor Positioning and Indoor Navigation, IPIN, Busan, South Korea, October 27–30*, pp. 261–270, 2014.
- [20] E. S. Lohan, J. Torres-Sospedra, H. Leppäkoski, P. Richter, Z. Peng, and J. Huerta, "Wi-Fi crowdsourced fingerprinting dataset for indoor positioning," *Data*, vol. 2, no. 4, p. 32, 2017.
- [21] G. M. Mendoza-Silva, P. Richter, J. Torres-Sospedra, E. S. Lohan, and J. Huerta, "Long-term WiFi fingerprinting dataset for research on robust indoor positioning," *Data*, vol. 3, no. 1, p. 3, 2018.
- [22] Z. Zhong, Z. Tang, and et al., "XJTLUIndoorLoc: A new fingerprinting database for indoor localization and trajectory estimation based on Wi-Fi RSS and geomagnetic field," in *Proc. International Symposium on Computing and Networking, CANDAR Workshops, Takayama, Japan, November 27–30*, pp. 228–234, 2018.
- [23] X. Song, X. Fan, C. Xiang, Q. Ye, L. Liu, Z. Wang, X. He, N. Yang, and G. Fang, "A novel convolutional neural network based indoor localization framework with wifi fingerprinting," *IEEE Access*, vol. 7, pp. 110698–110709, 2019.
- [24] J. Torres-Sospedra, A. R. Jiménez, and et al., "Off-line evaluation of mobile-centric indoor positioning systems: The experiences from the 2017 IPIN competition," *Sensors*, vol. 18, no. 2, p. 487, 2018.
- [25] V. Renaudin, M. Ortiz, and et al., "Evaluating indoor positioning systems in a shopping mall: The lessons learned from the IPIN 2018 competition," *IEEE Access*, vol. 7, pp. 148594–148628, 2019.
- [26] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.
- [27] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Annual Conference on Neural Information Processing Systems, December 8–13 2014, Montreal, Quebec, Canada*, pp. 3104–3112, 2014.
- [28] J. Torres-Sospedra, R. Montoliu, S. Trilles, Ó. Belmonte, and J. Huerta, "Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems," *Expert Systems with Applications*, vol. 42, no. 23, pp. 9263–9278, 2015.
- [29] X. Song, X. Fan, C. Xiang, et al., "A novel convolutional neural network based indoor localization framework with WiFi fingerprinting," *IEEE Access*, vol. 7, pp. 110698–110709, 2019.
- [30] A. Paszke, S. Gross, F. Massa, A. Lerer, et al., "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Advances in Neural Information Processing Systems*, pp. 8024–8035, 2019.