MapLoc: LSTM-Based Location Estimation Using Uncertainty Radio Maps

Xiangyu Wang, Zhitao Yu, Shiwen Mao, *Fellow, IEEE*, Jian Zhang, *Member, IEEE*,
Senthilkumar C. G. Periaswamy, and Justin Patton

***Abstract*—With the growing demand for location-based services, fingerprint has become a hot topic in the area of Internet of Things (IoT). However, the performance of fingerprinting-based indoor localization systems is usually affected by the quality and granularity of fingerprints. In this article, we present MapLoc, a long short-term memory (LSTM)-based indoor localization system that takes advantage of the continuous indoor uncertainty maps created using both earth magnetic field readings and WiFi received signal strengths (RSSs). A deep Gaussian process (DGP) model is trained to create indoor radio maps with confidence intervals, which are referred as uncertainty maps. Utilizing the uncertainty maps, an LSTM-based location prediction model is pretrained with artificial trajectory data sampled from the uncertainty maps, and then fine-tuned with the signal measurements collected in the field. In the training process, auxiliary outputs are implemented to overcome overfitting and improve the robustness of the system. Our extensive experiments demonstrate the outstanding performance of the proposed MapLoc system.**

***Index Terms*—Fingerprinting, deep Gaussian process (DGP), indoor localization, long short-term memory (LSTM), radio map construction.**

I. INTRODUCTION

WITH the rapid development of the Internet of Things (IoT), location-based service (LBS) has drawn increasing attention from various fields, such as robotics, retailing, manufacturing, and smart buildings. Instead of using specifically designed sensors for location estimation, radio-frequency (RF) signals, e.g., WiFi, have been a popular choice for indoor localization systems due to its wide deployment in indoor spaces. Fingerprinting is a popular indoor localization method, which generally consists of two stages: 1) offline fingerprint collection and 2) online location estimation. In the offline stage, fingerprints in the form of, e.g., WiFi received signal strength (RSS), are collected in the service area and labeled with the corresponding coordinates. Then, in the online stage, the unknown location of a mobile device will be estimated by matching the newly collected measurements with stored fingerprints. The performance of fingerprinting is thus largely affected by both the fingerprints and the matching method. Many prior works adopted various techniques in wireless communications, signal processing, and machine learning through these two aspects.

Various observations of RF signals have been utilized as fingerprints. For example, RSS was first used in [1]. This work demonstrated that the signal strength information provides a means of inferring user location. K-nearest neighbors (KNNs) was leveraged in this article to assist location prediction. Intuitively, RSS is negatively related to the distance between the transmitter and receiver. By using an empirical signal propagation model, the unknown location could be inferred roughly by triangulation. Even though RSS is resilient to slight environmental changes, it could not achieve fine-grained localization, especially when the number of APs is limited. For environments with rich AP resources, AP selection emerged to filter out the less useful RSS readings for boosted localization accuracy. Jia et al. [2] proposed a heuristic AP selection algorithm based on Cramer–Rao lower bound (CRLB) to assist in localization. MAPS [3] relied on K-means and decision trees for selecting available APs. Shi et al. [4] leveraged statistical features from RSS measurements to filter out valid APs. However, AP selection is still an open problem. In this article, the proposed location prediction model is capable of selecting appropriate APs intelligently with deep networks for improved localization. In addition, channel state information (CSI), as a fine-grained observation of the orthogonal frequency-division multiplexing (OFDM) physical layer (PHY), has been adopted as fingerprints in the past decade. It depicts how a signal propagates from the transmitter to the receiver through each subcarrier. Due to the nature of CSI, it is more sensitive than RSS to distance variations, and is also susceptible to the multipath effect and dynamic environments. Thus, various signal processing techniques have been proposed for eliminating the offsets introduced by the environment and hardware to enhance the quality of CSI fingerprints [5]. The extra cost of signal processing may impede the prevalence of CSI-based localization systems in mobile devices with limited hardware resources. Meanwhile, with the popularity of smart devices,

Manuscript received 9 November 2022; revised 31 January 2023; accepted 25 March 2023. Date of publication 28 March 2023; date of current version 25 July 2023. This work was supported in part by NSF under Grant ECCS-1923163 and Grant CNS-2107190, and in part by the RFID Lab and the Wireless Engineering Research and Education Center at Auburn University. *(Corresponding author: Shiwen Mao.)*

Xiangyu Wang and Zhitao Yu are with the Department of Electrical and Computer Engineering and RFID Lab, Auburn University, Auburn, AL 36849 USA (e-mail: xzw0042@auburn.edu; zzy0021@auburn.edu).

Shiwen Mao is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849 USA (e-mail: smao@ieee.org).

Jian Zhang is with the Department of Electrical and Computer Engineering, Kennesaw State University, Kennesaw, GA 30144 USA (e-mail: jianzhang@ieee.org).

Senthilkumar C. G. Periaswamy and Justin Patton are with the RFID Lab, Auburn University, Auburn, AL 36849 USA (e-mail: szc0089@auburn.edu; jbp0033@auburn.edu).

Digital Object Identifier 10.1109/JIOT.2023.3262619

increasing types of signals, such as light and earth magnetic field intensity, have been introduced as fingerprints [6]. It has been shown that such multimodal fingerprints are complementary to each other and can help to make the system more robust.

In addition to the quality, the density of fingerprints is also a key factor that affects the accuracy of fingerprinting. To achieve high location accuracy, a site survey is needed to collection fingerprints at densely marked locations, which is usually time-consuming and laborious. Furthermore, such dense fingerprints are costly to update when the service environment is changed (i.e., change of furniture placement). As a result, there is a tradeoff between the location estimation accuracy and system deployment cost, which needs to be carefully balanced when designing a fingerprinting system. In this research, we use deep Gaussian process (DGP) to generate a precise uncertainty map in order to improve localization with a limited number of signal observations. The generated data on the map indicates the dependability of signals from distinct APs, which aids in training the location prediction model. Another crucial factor to the success of fingerprinting is an effective and efficient location estimation (i.e., matching) method. In recent indoor localization systems, machine learning has been widely used as classifiers to estimate unknown locations in the online stage, such as KNNs, support vector machines (SVMs), and random forest [1], [7], [8]. Recently, deep learning models, have been adopted for effective multiclass classification. DeepFi [9] extracted features from CSI by using deep auto-encoders. CiFi [10] and ConFi [11] generated CSI images for image classification with the deep convolutional neural network. ResLoc [12] advanced the localization accuracy by modifying the architecture of the deep network. However, such methods are still focused on solving the traditional fingerprint matching problem, which partitions the continuous service area into a discrete grid and is treated as a multiclass classification problem. This approach introduces a built-in error, even though the error can be mitigated by probabilistic methods [9], [13].

In this article, we propose MapLoc, an indoor fingerprinting system that utilizes DGP to regress uncertainty maps and incorporates a long short-term memory (LSTM)-based method for location estimation. This work is focused on three aspects of improving location estimation: fingerprint quality, fingerprint quantity, and the built-in error of existing fingerprinting methods. From the perspective of fingerprint quality, both WiFi RSS and earth magnetic field intensity are utilized as fingerprints in MapLoc. Since the magnetic sensors are available in many smart devices, the magnetic field intensity measurements are readily available. Moreover, the uncertainty maps are leveraged to generate synthesized data that are indicative of the signal reliability. To enhance location estimation with limited number of fingerprints, MapLoc utilizes the uncertainty maps to synthesize sequences of trajectories and RSSI samples, which are used in auxiliary learning to pretrain the location prediction model.

In recent years, LSTM has been applied to a wide range of tasks, such as natural language processing, speech recognition, and time series forecasting, because it is able to effectively handle sequential data that has long-term dependencies. Also, it achieves great performance in handling variable-length sequences. Because of these features, LSTM is well-suited for tasks such as location prediction. By implementing a stacked LSTM network as a backend, we design a location prediction model for regressing the signal maps. And the estimated location will be inferred directly by leveraging the historical signal and location information, which avoids the built-in error when the localization problem is treated as a multiclass classification problem in existing fingerprint methods.

More specifically, a DGP is first implemented for uncertainty estimation in the service area. Then, the artificial signal measurements are generated by sampling the distribution described with uncertainties. In addition, geometry constraints and user movement patterns are considered in trajectory generation. The generated signal measurements are used to compose signal sequences that supervise the pretraining of the location prediction model. To better regress the signal strength, an auxiliary loss is adopted in the training. Both location prediction and fingerprint estimation are used to calculate the loss for weight updating. Finally, the pretrained model is fine-tuned with real signal sequence collected in the field. Fine-tuning forces the location prediction model to converge to the real signal surface, thus eliminating the cumulative error of the DGP model. In the online stage, the location of the target mobile device is readily predicted by the location prediction model using its newly measured signals and past trajectory in a small sliding window.

The main contributions of this article are summarized as follows.

1) An innovative localization framework is proposed by leveraging the uncertainty estimation capability of DGP. Continuous uncertainty maps are created by DGP using fingerprints measured at gridpoint locations. The fingerprints are then augmented by sampling the distribution described by the uncertainty maps. The generated signal measurements reflect their own stability, allowing deep learning models to learn the reliability of signals and select the effective measurements for location estimation.
2) By introducing geometric constraints of the service area and user movement trajectories, the continuous nature of human mobility and the historical locations of the target device within a small window are taken into account. Furthermore, fingerprinting is no longer treated as a classification problem here. Rather, the location prediction model readily produces the estimated location in the manner of regression, thus mitigating the built-in error of the traditional approach.
3) We leverage auxiliary learning in training the location prediction model. By introducing the signal measurement loss as one of the components of the auxiliary loss in supervise training, the LSTM-based location prediction model will be forced to learn the inherent relationship in the sequences of measurements. Compared with the traditional training approach that only uses isolated location as labels, signal sequences

include much more features to guide and accelerate the training process.

4) Multimodal maps, created using WiFi RSS and earth magnetic field strengths, are utilized in the MapLoc system. Such measurements are widely available and do not increase the cost and affect the compatibility of the system. It is easy to extend the proposed framework to include more types of measurements, such as light intensity, for future improved performance.
5) We verified the performance of the proposed MapLoc system with extensive experiments in two representative indoor environments. The results demonstrate that MapLoc advances the accuracy of location estimation by taking advantage of the uncertainty estimation provided by DGP and the bimodal fingerprints.

In the remainder of this article, we present an overview of related work in Section II. The preliminaries and motivations are provided in Section III. Section IV presents the system design. In Section V, we evaluate our prototype system, and in Section VI, we wrap up this article.

II. RELATED WORK

With the rise of the IoT, indoor LBSs have drawn a lot of attention from both academia and industry, due to their high social and economic value. Unlike outdoor localization systems, such as the global positioning system (GPS), which rely on the line-of-sight (LOS) reception of satellite signals, the performance of indoor localization is hampered by scattered and reflected signals due to the clutter environment. Indoor localization is still an open problem without a universal solution, despite a variety of techniques have been proposed in the literature.

A. *Fingerprinting Approaches*

Because of their adaptability and adequate accuracy, fingerprinting methods are commonly used in localization systems. The features derived from the observations are adopted for pattern matching in fingerprinting. RADAR [1] was one of the first attempts to use RF signals, where RSS was used as fingerprints. Aside from RSS, various types of observations were leveraged in prior works as well. CSI is a fine-grained observation from the PHY layer, which includes the amplitude and phase of each subcarrier of the OFDM PHY. FILA [14] demonstrated that CSI helps to improve localization accuracy and reduce latency. The quality of fingerprints, which can be viewed as a discrete radio map, plays a critical role in such systems. A basic and effective way to improve the quality of the radio map is to increase the number of fingerprints. However, collecting fingerprints is usually time-consuming and laborious, and in some cases, impossible. To minimize such effort, prior works [15], [16], [17] utilized unmanned aerial vehicles (UAVs) to replace manual labor. DeepMap [13] constructed a radio map with DGP using only a limited number of fingerprints. WiGAN [18] generated fingerprints for an unknown area with Gaussian process regression conditioned least-squares generative adversarial networks (GPR-GANs). Wu et al. [19] and Huang et al. [20] investigated the radio map adaptation and update problem to avoid the cumbersome recollection of fingerprints in dynamic environments. On the other hand, the quality of fingerprints keeps improving with the advance of technology, hence the evolution of radio maps. Gu et al. [21] eliminated multipath interference in WiFi signals with the sparsity rank singular value decomposition (SRSVD) method. Luo and Fu [22] extracted nonlinear features from RSS signals by implementing kernel principal component analysis (KPCA). Furthermore, deep learning techniques have achieved an exceptional performance in feature extraction as well. To extract nonlinear features from observations, deep autoencoders were incorporated in [5], [9], [23], and [24], while [6], [25], [26] leveraged LSTM and its variants to evaluate the correlation between received RF signals for optimizing the fingerprints. In [10], [11], [12], and [27], CNN was used to extract fingerprints from multidimensional signal arrays for improved localization accuracy.

B. *Geometry-Based Approaches*

In addition to fingerprinting methods, geometric methods, such as multilateration and triangulation, are widely used in indoor localization systems by exploiting the measurements for fine-grained information. Among various measurements, Angle of Arrival (AoA) is commonly employed in radar and acoustics systems. ArrayTrack [28] proposed a multipath suppression algorithm for eliminating the reflection paths between transmitter and receiver. SparseTag [29] proposed to use a spatial smoothing-based method, which processed a sparse RFID tag array and decreased the angle estimation error to 1.831°. Time-of-Arrival (ToA)-based systems estimate the transmitter–receiver distance by measuring the traveling time of the signal. However, such systems require tightly synchronized clocks at the transmitter and receiver. Kang et al. [30] mitigated the time synchronization error and the NLOS error by introducing an iterative ToA (iToA) algorithm incorporating a multivariate linear model. Also, Yuan et al. [31] proposed a unified factor graph-based framework for ToA-based localization in wireless sensor networks. The framework provided a unified treatment of the inaccurate positions of transmitters and the asynchronous network. Even though the localization accuracy keeps increasing with these approaches, their performance is still insufficient for practical indoor services because of the required LOS signals and multipath-free environments.

C. *Other Approaches*

In addition to RF signal-based techniques, vision-based techniques are also popular with the emerging of robotics, autonomous vehicles, and augmented reality (AR) [32]. The localization algorithms rely on the inputs from sensors, such as RGB-D cameras and infrared cameras, to extract location information. The vision-based techniques usually achieve centimeter level accuracy in real time, outperforming most of RF signal-based techniques. For example, MonoSLAM [33] is the first study to apply the simultaneous localization and mapping (SLAM) approach with a single uncontrolled camera, with centimeter level accuracy at 30-Hz real-time performance. AprilTag [34] created a visual fiducial system that enables full six degrees-of-freedom (6DOF) localization with a single image by using a 2-D barcode tag as landmark. However, the computational cost of vision-based approaches constraints their deployment on IoT devices with limited computation power and short battery life [35]. Moreover, the visibility, occlusion, and privacy related issues further constrain the usage of vision-based approaches.

Indoor localization also takes advantage of the development of visible light communications (VLCs). By analyzing the modulated light signal transmitted in the form of visible LED lights, many VLC signal-based localization techniques have been proposed. Because the diffused components emerging from multipath scattering are substantially weaker than the LOS component, the VLC-based localization system has a superior accuracy over RF signal-based system, which usually suffer from strong multipath interference [36], [37], [38].

Acoustic signals have also been employed in localization systems. It provides precise localization at a low cost due to readily accessible equipment such as speakers and microphones, as well as excellent time-domain resolutions. For instance, EchoTrack [39] tracked hand trajectory with a built-in speaker array and microphone on smart phones by leveraging two-channel chirps to remove the multipath noise. Location estimation is enhanced by using the Doppler shift compensation and roughness penalty smoothing method. Vernier [40] achieved accurate motion tracking accuracy of less than 4 mm, by proposing a differentiated window-based phase change calculation (DW-PC) to minimize the computation overhead for real-time tracking.

III. PRELIMINARIES AND MOTIVATIONS

Gaussian process has been successfully applied for solving regression and probabilistic classification problems. A Gaussian process is described with its covariance matrix and mean function. Since the prediction is also Gaussian, confidence intervals can be estimated to depict the uncertainty of data distributed over a continuous space. Thus, a normalized signal strength map for a service area can be conveniently reconstructed with measured signal strengths and the corresponding coordinates by a Gaussian process regression model, which is given by

$$r(c) = f(c) + \epsilon \tag{1}$$

where $r(c)$ and $f(c)$ represent the RSS and ideal signal strength for location c, respectively, and ϵ is the observation noise, which follows an independent identically distributed (i.i.d.) Gaussian distribution with zero mean and variance θ_n^2.

It is intuitive to assume that the RSSs r_i and r_j at coordinates c_i and c_j, respectively, also follow a joint Gaussian distribution with covariance $k(c_i, c_j)$, which is usually described using a kernel function as

$$k\left(c_i, c_j\right) = \phi^2 \exp\left(-\frac{1}{2l^2}\left|c_i - c_j\right|^2\right) \tag{2}$$

where ϕ and l are the hyperparameters for depicting the signal variance and the smoothness of the kernel function, both of which can be estimated by using a maximum-likelihood approximation method. Then, the joint distribution of the estimated signal strength f_* of location c_* and the measured signal strengths $\mathbf{r}$ can be depicted as follows:

$$\begin{pmatrix}\mathbf{r}\\ f_*\end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix}\mathbf{K} & \mathbf{K}_*\\ \mathbf{K}_*^T & \mathbf{K}_{**}\end{pmatrix}\right). \tag{3}$$

The signal strength f_* can be inferred from the measured signal strength $\mathbf{r}$ by

$$\Pr(f_* | c_*, \mathbf{c}, \mathbf{r}) = \mathcal{N}(f_* | \mu_*, \Sigma_*) \tag{4}$$

$$u_* = \mathbf{K}_*^T\left(\mathbf{K} + \theta_n^2\mathbf{I}\right)^{-1}\mathbf{r} \tag{5}$$

$$\Sigma_* = \mathbf{K}_{**} - \mathbf{K}_*^T\left(\mathbf{K} + \theta_n^2\mathbf{I}\right)^{-1}\mathbf{K}_* \tag{6}$$

where $\mathbf{c} \in \mathbb{R}^{N\times 2}$, $\mathbf{r} \in \mathbb{R}^N$, $\mathbf{K}_{**} = [k(c_*, c_*)]$, N is the number of positions where the measurements were taken, $\mathbf{K}$ is the covariance matrix of $\mathbf{c}$ with dimension $N \times N$, and $\mathbf{K}_*$ is an $N \times 1$ matrix of covariances between $\mathbf{c}$ and c_*.

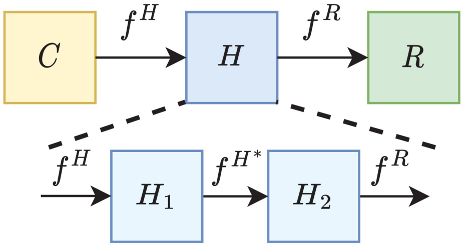

Fig. 1. DGP model for signal map construction.

Inspired by the Gaussian process-based works, the DGP is leveraged in this article to enhance the precision of the constructed map by recovering the nonstationary components of signal measurements. In our prior work [13], a two-layer DGP model was leveraged to extract nonlinear characteristics from RSS samples and construct radio maps. Compared with Gaussian process, DGP is able to regress complex input data by taking advantage of the fusion of kernels. Fig. 1 is a graphical representation of a DGP, which consists of three layers of nodes, i.e., the parent nodes C, the leaf nodes R, and the latent nodes H, which include two sublayers H_1 and H_2 [41]. For a 2-D map generation problem, C is the set of training coordinates with dimension $N \times 2$, R denotes a signal measurement matrix of $N \times S$, and $H \in \mathbb{R}^{N\times L_{sub}}$. Here, N, S and L_{sub} represent the number of measured coordinates, the number of sensors, and the number of the intermediate latent dimensions in the sublayers, respectively. Therefore, the generative process is given by

$$h_{nl}^1 = f_l^H(c_n) + \epsilon_{nl}^H,\ l = 1, 2, \ldots, L_1,\ c_n \in \mathbb{R}^2 \tag{7}$$

$$h_{nl}^2 = f_l^{H^*}\left(h_{nl}^1\right) + \epsilon_{nl}^{H^*},\ l = 1, 2, \ldots, L_2,\ h_{nl}^1 \in \mathbb{R}^{L_1} \tag{8}$$

$$r_{ns} = f_s^R\left(h_{nl}^2\right) + \epsilon_{ns}^R,\ s = 1, 2, \ldots, S,\ h_{nl}^2 \in \mathbb{R}^{L_2} \tag{9}$$

where $f^H \sim GP(\mathbf{0}, k^H(C, C))$, $f^{H^*} \sim GP(\mathbf{0}, k^{H^*}(H_1, H_1))$, and $f^R \sim GP(\mathbf{0}, k^R(H_2, H_2))$ are Gaussian processes, which connect the latent nodes H_1, H_2, parent nodes C, and leaf nodes R, respectively. The automatic relevance determination (ARD)

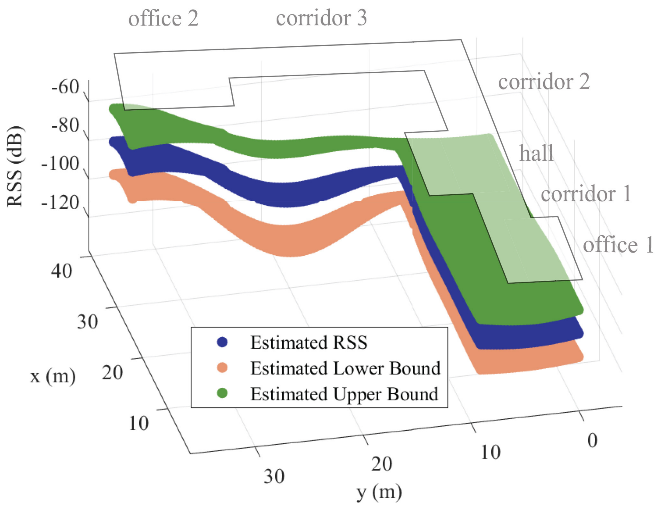

Fig. 2. RSS uncertainty map constructed by DGP.

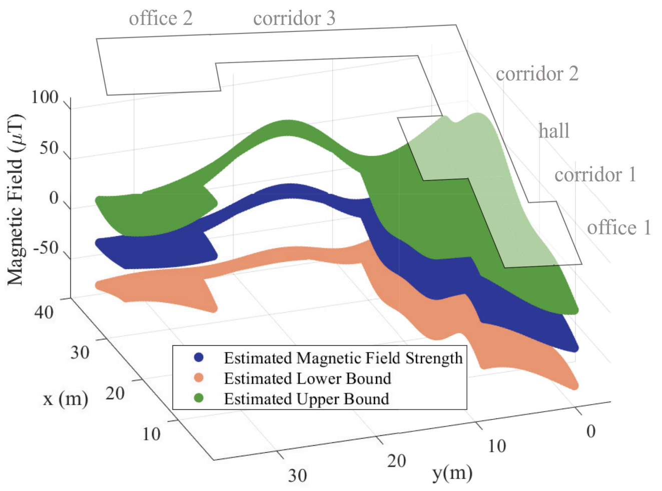

Fig. 3. Earth magnetic field intensity uncertainty map constructed by DGP.

covariance functions for the Gaussian Processes is defined as

$$k_{\text{ARD}}(c_i, c_j) = \phi_{\text{ARD}}^2 \exp\left(-\frac{1}{2}\sum_{l=1}^{L} w_l\left(c_{i,l} - c_{j,l}\right)^2\right) \quad (10)$$

where w_l is the weight for each latent dimension and ϕ_{ARD} is a hyperparameter. For different inputs, the Gaussian processes, f^H, f^{H^*} and f^R, only depend on the covariance function k_{ARD}. To find the optimal hyperparameters, Bayesian training is leveraged to maximize the marginal distribution of the observed signal measurement R, which is given by

$$\max\ \log p(R) = \log \int_{C,H} p(R|H)p(H|C)p(C). \quad (11)$$

The outstanding performance of DGP for generating a detail-rich signal map has been demonstrated in [13]. With the deep and heterogeneous nonlinear structure, the DGP handles the nonstationary components in complex signal measurements and extracts the detailed information about the distribution of real WiFi RSS measurements in indoor environments.

Despite the fact that the detailed maps created by DGP improves localization accuracy, the uncertainty information, which could also be retrieved using DGP, was largely ignored in our prior work [13]. Indeed, the uncertainty information just happens to be a convenient tool for evaluating the reliability of sampled signals. Fig. 2 illustrates a uncertainty radio map constructed by DGP using the measured RSS data from a specific AP in a public data set [42]. The map includes three signal layers: 1) a green layer representing the upper confidence bound of the map; 2) a blue layer of mean values; and 3) a peach layer denoting the lower confidence bound of the map. A translucent layer is overlaid in the map to illustrate the layout of the floorplan. The confidence bound layers depict the 95% confidence interval of the signal distribution. Although the AP location is unavailable and not needed in this work, based on the signal-to-distance relationship [43], the AP can be located near the top-left corner of Fig. 2 where the signals are the strongest and most stable. When the distance is increased, the signal strength decreases and fluctuates more considerably. For the locations that are beyond the coverage of the AP, the signal strength drops to -100 dBm and settles there. The RSS data from this AP, obviously, would be more constructive in locating target devices in the top-left region, while this AP would have a negative impact on locating targets in the map's central area because the RSS samples in the area would be highly random with large fluctuations. Such a pattern of uncertainty indicates that the signal stability varies depending on the location. And different patterns of uncertainty map would also be obtained for different APs. Thus, in MapLoc, we can sample the Gaussian distribution that is defined by the mean and confidence intervals in the uncertainty map to generate artificial measurements that depict the stability of the signal. The following LSTM-based location prediction model will exploit such fluctuations to distinguish the optimal signal measurements for location estimation. Moreover, Fig. 3 plots the uncertainty map generated by DGP using earth magnetic field observations. It follows a similar trend as in Fig. 2, in which the signal stability changes at different locations, and is complementary to the RSS uncertainty map. Both RSS and magnetic field data will be used in this effort to improve the accuracy of localization.

On the other hand, the proposed MapLoc system also takes into account the trajectory of the target device in a sliding time window. The trajectories can be reasonably synthesized by leveraging the movement pattern of target devices and geometry constraints (e.g., the shape of the room or corridor). Using the uncertainty maps, artificial signal sequences can be generated along such movement trajectories. The artificial signal sequences are used to pretrain the LSTM-based location prediction model, which is then fine-tuned with real collected signals in the field. The pretraining process guides the location prediction model by learning the signal reliability, while fine-tuning mitigates the cumulative error introduced by imprecise uncertainty maps.

IV. SYSTEM OVERVIEW

Fig. 4 presents the system architecture of the MapLoc system, where the green and blue blocks represent the components in the offline stage. More specifically, the green

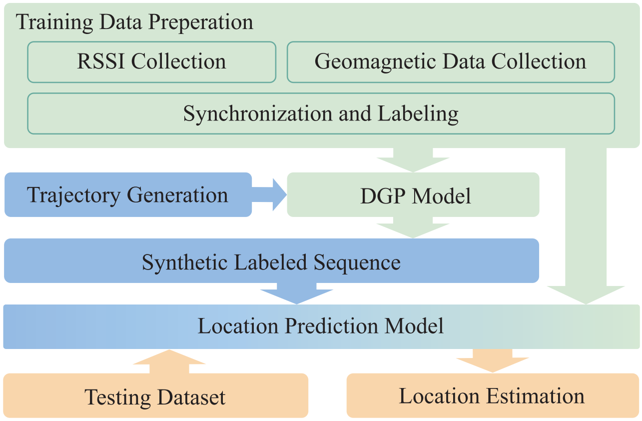

Fig. 4. MapLoc system architecture.

blocks are related to collecting signal measurements and their corresponding coordinates, whereas the blue blocks are associated to the synthesized signal measurements and their coordinates. The location prediction model is unique in that it is pretrained with the synthesized RF data and then fine-tuned with the collected RF data, which is why it is colored in gradients (from blue to green). The yellow blocks in Fig. 4 represent the components in the online stage.

Similar to traditional fingerprinting systems, MapLoc also consists of two stages: an offline stage for data collection and model training, and an online stage for location estimation. In the offline stage, WiFi RSS measurements as well as magnetic field readings are collected with the built-in sensors in the mobile device. The measurements comprising the collected bimodal sequences, which are tagged with the corresponding coordinates where the data was measured. For each location, we collect RSS measurements from as many APs as possible. Since the set of visible APs usually varies from location to location, we force the RSS measurements from those inaccessible (i.e., out of coverage) APs to be −100 dBm to ensure consistency in measured data.

Localization with MapLoc includes two parts as well. The collected bimodal signal measurements are first leveraged for training the DGP model to generate their uncertainty maps. The uncertainty map includes the mean value and the upper and lower bounds of the 95% confidence interval, as illustrated in Fig. 2. The uncertainty map will then be leveraged to synthesize artificial bimodel signal sequences for enhancing the training of the location prediction model, which is introduced to consider the trajectory (or, historical) information of the target device in location estimation. The model is first pretrained with the artificial signal sequences synthesized by sampling the uncertainty maps, and then fine-tuned with the collected bimodal sequences to avoid the cumulative errors introduced by the DGP model. In the online stage, the DGP model will not participate in location estimation. The estimated location will be obtained by combining the previous trajectory information with a time window W with the signal measurements from the current unknown location.

Algorithm 1 Pseudocode for Measuring the Quality of the Uncertainty Map

Input: the measured verification sample r_j^k and the corresponding coordinate c_*^k, the mean layer of the uncertainty map m_j for the jth signal, the number of gridpoints G in m_j, the number of available signals S, and the number of verification samples K;

Output: the map quality Q;

```
//$i$ represents the index of gridpoints in map $m_j$
//$j$ denotes the index of signals
//$k$ denotes the index of verification samples
//$l$ denotes the coordinate of the gridpoints in map $m_j$
for $k = 1 : K$ do
    for $j = 1 : S$ do
        for $i = 1 : G$ do
            //compute the likelihood function $p(r_j^k|c_i)$
            $p(r_j^k|c_i) = \exp\left(-\frac{1}{\lambda\sigma^2}\left\|r_j^k - m_j^{c_i}\right\|\right)$;
        end for
        //compute the posterior probability $p(l_i|r_j^k)$
        $p(c_i|r_j^k) = \frac{p(r_j^k|c_i)}{\sum_{d=1}^{G} p(r_j^k|c_d)}$;
    end for
    //use MAP estimation to infer location for the verification samples
    $\hat{c}^k = argmax_{\{c_1,c_2,\ldots,c_G\}}\left(\prod_{j=1}^{S} p(c_i|r_j^k)\right)$;
end for
//compute map quality $Q$
$Q = \frac{1}{\exp(\frac{1}{2K}\sum_{k=1}^{K}(||c_*^k - \hat{c}^k||))}$;
return $Q$;
```

A. Offline Training

Offline training of the MapLoc system includes pretraining and fine-tuning. The DGP model is first trained using the bimodal signals that have been collected. The location prediction model will first be trained using the artificial bimodal sequences generated by the DGP model, and then fine-tuned using the signal sequences composed of collected signal measurements from the field to ensure that it converges to the real-world situation.

1) Pretraining: First, the collected signal measurements are used to train the DGP model. Because the DGP model focuses primarily on the signal distribution, the temporal information in the signal sequence is neglected during the training. To improve the structure of the DGP model and optimize the related hyperparameters, a simple approach is employed to assess the quality of the uncertainty map generated by the DGP model. As shown in Algorithm 1, the constructed uncertainty map $\mathbf{M}$ is a $G \times S \times 3$ matrix, which includes an upper confidence layer, a mean layer, and a lower confidence layer. Here, G denotes the number of gridpoints in the map. It has to be 100 000 to reach a resolution of 0.01 m for an area of 10 m^2. S represents the number of available signals. For example, we have $S = 10$ if the WiFi RSS measurements are collected from seven APs, since each magnetic field reading is a vector with three elements (mag$_x$, mag$_y$, mag$_z$),

Algorithm 2 Pseudocode for Artificial Trajectory Generation

```
Input: the length of the artificial trajectory $N$; the layout of the indoor environment $O$; the stride length $d$;
Output: the artificial trajectory $C$;
 1: //generate the coordinates $c_0$ randomly in the environment $O$ and initialize the trajectory $C$
 2: $C = \{randomPosition(O)\}$;
 3: while $C.length < N$ do
 4:     if $C.length == 1$ then
 5:         //$\gamma$ is a random initial azimuth
 6:         //generate the coordinate $c_*$ with the distance $d$ and the azimuth $\gamma$
 7:         //$c_{0_x}$ and $c_{0_y}$ are the x-axis and y-axis coordinates of $c_0$, respectively
 8:         $c_* = [c_{0_x} + d * \cos(\gamma), c_{0_y} + d * \sin(\gamma)], \gamma \sim U(-180°, 180°)$;
 9:     else
10:         //update $\gamma$ based on the previous azimuth
11:         $\gamma = \gamma + \gamma_t,\ \gamma_t \sim U(-40°, 40°)$;
12:         //$c_{-1}$ is the last coordinate in trajectory $C$
13:         $c_* = [c_{-1_x} + d * \cos(\gamma), c_{-1_y} + d * \sin(\gamma)]$;
14:     end if
15:     if $c_*$ in the environment $O$ then
16:         $C.append(c_*)$;
17:     end if
18: end while
19: return $C$;
```

describing the magnetic field intensity for the north, east, and vertical directions, respectively. The mean layer m is constructed to evaluate the overall quality of the uncertainty map. K verification samples are collected from each gridpoint in the service area and labeled with the corresponding coordinates. We calculate the likelihood function $p(r_j^k|c_i)$ of the jth signal, which indicates the similarity between the kth verification sample r_j^k and the signal measurement at c_i in the uncertainty map m^j with a Gaussian kernel, as presented in step 9. In MapLoc, the σ^2 and λ are set to 0.35 and 2, respectively. Thus, the posterior probability $p(c_i|r_j^k)$ is obtained conveniently by assuming the distribution over the G gridpoints is uniform (see step 12). The coordinate estimation of the kth sample is given by choosing the gridpoint with the highest posterior probability. Eventually, the quality of the uncertainty map, Q, is evaluated based on the errors of the coordinate estimation in step 18.

Based on the well-trained DGP model, a movement model is introduced to produce trajectories for generating artificial signal sequences. As shown in Algorithm 2, the stride length d is considered in the movement model and is restricted to 0.6 m. The azimuth γ is determined by the previous azimuth with a random offset between −40° and 40°. In step 13, the coordinates in trajectory C are generated sequentially based on the previous azimuth. And the layout of the indoor environment is considered to eliminate the coordinates outside the service area (see steps 15–17).

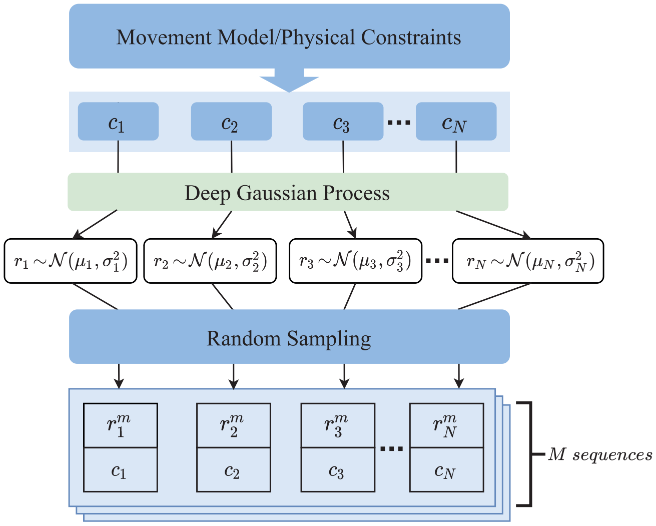

Fig. 5. How to synthesize labeled signal sequences for pretraining the LSTM-based location prediction model.

As shown in Fig. 5, the well-trained DGP model is utilized to generate the artificial signal r_N for coordinate c_N in trajectory C. According to trajectory C, the artificial signal sequences are assembled using the signal measurements generated by sampling the distribution $\mathcal{N}(\mu_N, \sigma_N^2)$ that is described by the mean μ_N and variance σ_N in the uncertainty map. It is noteworthy that the distribution is sampled M times to ensure that the generated signal measurements are able to represent the stability of signals. Furthermore, we employ a sliding window with a length of W for adjusting the size of the artificial sequences for training the LSTM-based location prediction model. An artificial trajectory of length N will produce $N - W + 1$ training sequences. For each training sequence, the last signal measurement r_{i+W-1}^m and the corresponding coordinate c_{i+W-1} will be extracted as label for supervise training.

The forward propagation of the location prediction model is depicted in Fig. 6. The backbone of the location prediction model is a stacked LSTM model, which is followed by a DNN for signal estimation (termed DNNS) and a DNN for location estimation (termed DNNL). To push the model to learn the signal map made by the DGP model and estimate location using the map, auxiliary loss is used in training. The signal values r_{i+W-1}^m in the label data are processed and concatenate with the output of the LSTM network in the DNNL model for predicting the unknown coordinate $\widehat{c}$. Then, the MSE loss is calculated by comparing the label coordinate c_{i+W-1} and the location prediction $\widehat{c}$ by the DNNL. In parallel, a signal estimation $\widehat{r}$ is given by the DNNS using the output of the previous LSTM model as well. As a result, the loss function of the location prediction model is given by

$$\mathcal{L} = (1 - \beta)\mathrm{MSE}\big(r_{i+W-1}^m, \widehat{r}\big) + \beta \mathrm{MSE}(c_{i+W-1}, \widehat{c}) \quad (12)$$

where β is a hyper parameter to adjust the influence of the two types of losses, while $\widehat{r}$ and $\widehat{c}$ are the predicted signal by DNNS and the predicted coordinate by DNNL, respectively.

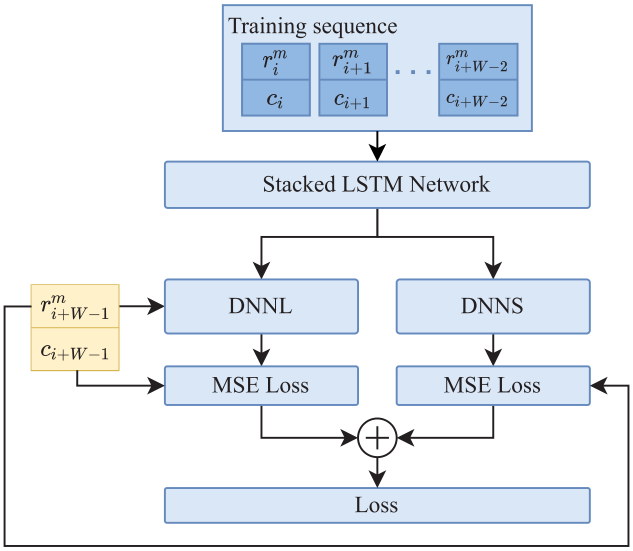

Fig. 6. LSTM-based location prediction model in MapLoc.

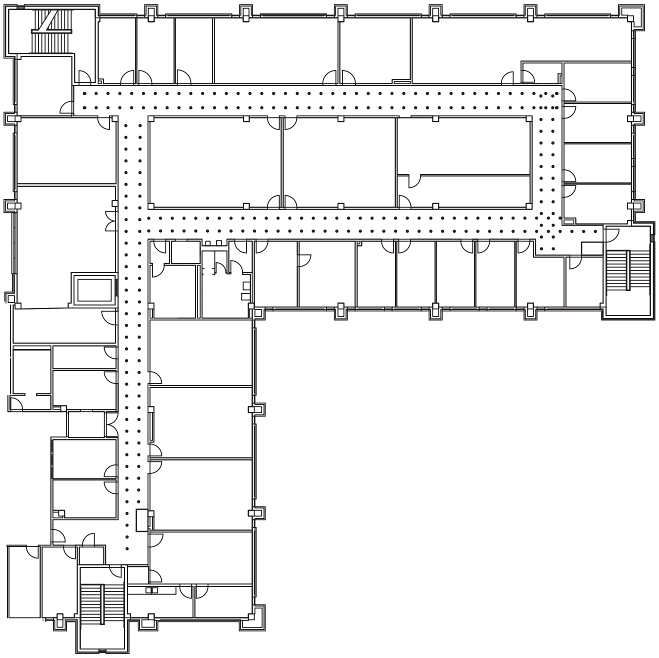

Fig. 7. Floorplan for the Broun Hall data set.

2) *Fine-Tuning:* After pretraining, the location prediction model will be fine-tuned with collected bimodal sequences from the service area. The collected bimodal sequences, like the artificial sequences, are reorganized to form shorter training sequences using a sliding window of size W. The last bimodal measurement of each training sequence is also used as the sequence's label to complete the supervised training of the model.

B. Online Testing

In the online stage, only the stacked LSTM network and DNNL will participate in location estimation. The location prediction model operates in a similar manner to autoregression models. The historical trajectory, including the received signal measurements and the corresponding coordinates, is fed into the stacked LSTM network. By combining the output of the LSTM network with the freshly collected signals from the current unknown location, the estimated location is deduced readily with the well-trained DNNL model. Because the localization problem is addressed as a regression problem in MapLoc, the built-in error associated with the discrete fingerprints can be avoided. Furthermore, since the estimated location is computed directly by the location prediction model, the cumbersome localization strategies used in prior work [13] are not needed anymore in MapLoc, which further reduces the computational cost, especially for mobile devices with limited computation resources and power supplies.

V. EXPERIMENTAL STUDY

A. Experiment Configuration

To demonstrate the performance of the MapLoc system, we evaluate it in two typical environments. First, we conduct experiments on the fourth floor of Broun Hall in the Auburn University Campus. In this scenario, we implement a prototype system using a Samsung Galaxy S7 Edge smartphone, which is equipped with a dedicated application for collecting magnetic field intensity data and WiFi RSS data simultaneously. As depicted in Fig. 7, the experiment covers an area of approximately 270 m^2. The black dots in Fig. 7 represent 255 sample locations (i.e., gridpoints) for training the DGP and the location prediction model. Except for some corner gridpoints, the distance between two adjacent training locations is 90 cm. 80 testing locations are randomly selected in the service area, which are not shown in Fig. 7. None of the testing locations overlap with a training location in this scenario. Moreover, RSS readings are collected from 224 APs, including all the available 2.4-GHz APs and 5-GHz APs from various manufacturers. To make the data size consistent, the RSS values of out-of-range APs are set to -100 dBm. The magnetic field strength is obtained from the on-device sensor directly, which is a vector including the magnetic field intensity for the north, east, and vertical directions.

The performance of the MapLoc system is also evaluated using a public data set [42]. Fig. 8 plots the detailed floor plan where the public data set was collected. The data set covers a floor of 185.12 m^2, which includes three corridors, two offices and a hall. The fingerprints are captured from 325 gridpoint locations, shown as black dots in Fig. 8. The distance between two adjacent gridpoints is 60 cm. The data acquisition campaign was performed using a smartphone, SONY Xperia X2, and a smartwatch, LG W110G Watch R. We only utilize the data collected by the smartphone in this experimental study. The RSS data are captured from 132 unique APs, and the readings from an out-of-range AP are all set to -100 dBm. We only leverage 75 APs in the following experiments because some AP signals are very weak across the entire service area. Similar to the magnetic field intensity in the Broun Hall scenario, the magnetic field readings of this scenario are also vectors with three elements. Since the data acquisition campaign is conducted in this environment with the identical

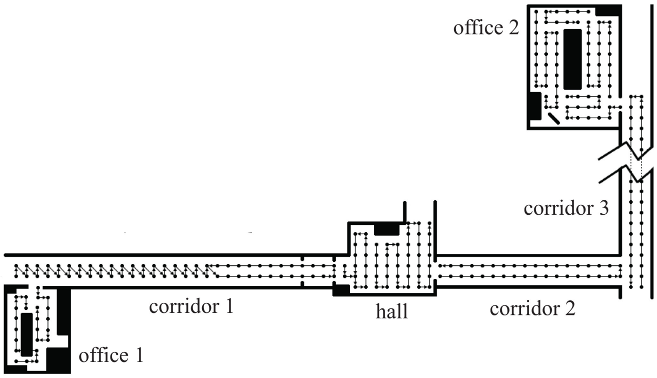

Fig. 8. Floorplan where the public data set was collected.

setting twice, we train and then test the MapLoc system using the data sets from different campaigns for a fair and realistic evaluation.

Identical settings of the location prediction model are deployed in both environments. Nine LSTMs are stacked one above another to form a stacked LSTM as backbone of the location prediction model. The number of features in the hidden state of LSTM is set to about 1.5 times of that of the input features, e.g., the number of features in the hidden state will be 150 if the number of available AP is 95. Each magnetic field reading is a vector of size 3 × 1 and the corresponding coordinates are in a 2-D space. The hidden state of the last layer of the stacked LSTM is passed into the two DNNs for location estimation and signal estimation, respectively. DNNL is composed of four linear layers. The size of the input data r^m_{i+W-1} is first adjusted to 16 by a layer in DNNL, while the size of the hidden state from the LSTM is squeezed to 32 by another DNNL layer. By concatenating the outputs from the two layers, the estimated location is obtained by the remaining two layers in DNNL, where the output feature numbers of the layers are 16 and 2, respectively. The structure of DNNS is relatively simple. The hidden state from the LSTM is compressed by three linear layers in DNNS sequentially, where the output feature numbers of the layers are 256, 128, and the same as that of the input data r^m_{i+W-1}, respectively.

In both scenarios, the magnetic field intensity and WiFi RSS readings are min-max normalized. Considering that pedestrians usually do not make abrupt changes in their movements indoors, the stride length d is set to 0.6 m, and the azimuth offset γ_t is limited in the range between $-40°$ and $40°$. To accelerate the training process, a server with an Nvidia RTX 3090 GPU is leveraged for real-time trajectory generation and model training.

The following baselines are used in our comparison study.

1) *DeepMap:* This is the scheme proposed in our prior work [13]. To regress the indoor radio map, a DGP model with the exact same configuration as in MapLoc is used. A Bayesian method is utilized to compare the newly collected signal measurement with the generated maps. The location is estimated without using the uncertainty maps.
2) *LSTM:* The same stacked LSTM network as in MapLoc is used in this scheme. After the LSTM backbone, DNNL processes the extracted features and directly predicts location. The LSTM backbone and DNNL are configured similarly as in MapLoc. This model is trained with trajectory/RSS sequences sampled from the fingerprints collected in the field
3) *LSTM+DeepMap:* The design of this scheme is identical to that of the proposed MapLoc. The only difference is that the model is trained using sampled trajectory/RSS from the map created as in DeepMap [13], i.e., the blue layer in Fig. 2, rather than the uncertainty maps.

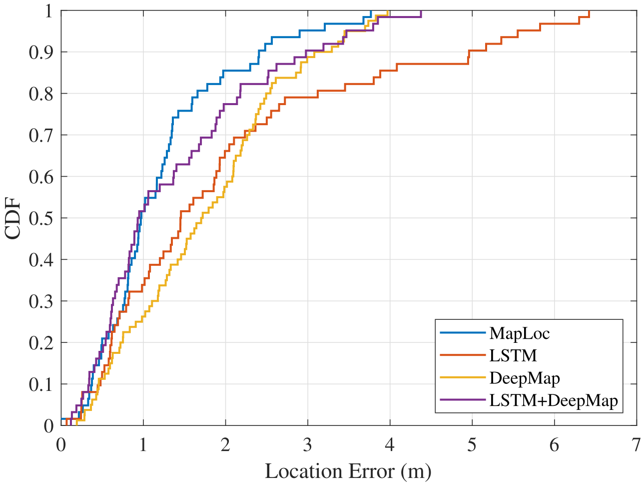

Fig. 9. CDF of localization errors on the Broun Hall data set.

B. Experimental Results and Analysis

1) Accuracy of Location Estimation: First, we evaluate the localization performance on the Broun Hall data set. Fig. 9 illustrates the cumulative distribution functions (CDFs) of localization errors for the proposed MapLoc system and the three baseline schemes. According to Fig. 9, it is obvious that MapLoc outperforms the other methods on the Broun Hall data set. Despite the fact that both MapLoc and LSTM+DeepMap obtained a performance where 50% of the errors are less than 1 m, MapLoc has a distinct advantage that approximately 75% of location estimation have errors less than 1.35 m, whereas only 59% of location estimation obtained by LSTM+DeepMap accomplish the similar accuracy. This demonstrates the improvement brought about by the samples from uncertainty maps. In addition, Fig. 9 reveals the obvious deficiencies of LSTM and DeepMap in localization accuracy. The maximum localization error, 6.41 m, is from LSTM. The comparison demonstrates that the combination of LSTM and DeepMap contributes to higher precision localization. In MapLoc, the augmented training data produced by the DGP model benefits the location prediction model that uses LSTM as its backbone. By incorporating historical information into location estimation via the LSTM model, the localization accuracy of the DeepMap model is improved significantly as well. Based on the collaboration of DeepMap and LSTM, our proposed MapLoc successfully improves the location estimation accuracy by taking into account the uncertainties of different signals as well as historical information.

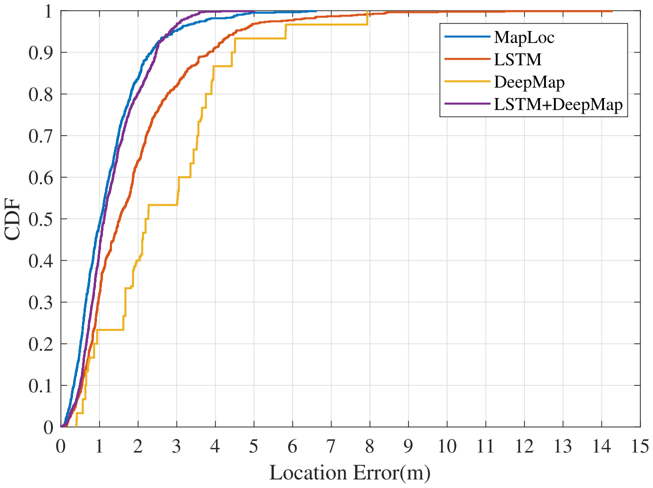

Fig. 10. CDF of localization errors on the public data set.

We also conduct an experiment using the public data set to investigate the performance of the proposed MapLoc system. The CDF of localization errors on the public data set is displayed in Fig. 10. The results on the public data set are similar to those with the Broun Hall data set. MapLoc and LSTM+DeepMap keep the leading position in the comparison. Even though 50% of location estimation errors are lower than 1.1 m with both MapLoc and LSTM+DeepMap, the overall performance of MapLoc is superior to that of LSTM+DeepMap slightly. Because the artificial signal measurements are sampled from the uncertainty maps, the distribution of the generated measurements describes the measurement's quality. As a result, the location prediction model can learn the reliability of different types of signal measurement, the sets of measurements from different APs and, thus, improve the accuracy of location estimation. Moreover, LSTM outperforms DeepMap in the public data set scenario, although the maximum localization error, 14.26 m, is obtained with the LSTM method.

The main results in Figs. 9 and 10 are summarized in Fig. 11. The height of the bars represents *mean error*, whereas the black line in each bar represents *median error*. The location prediction model of MapLoc, denoted as LSTM in Fig. 11, and DeepMap, each only contribute to limited accuracy in location estimation. By combining these methods, the localization accuracy is increased significantly. The mean and median error on the public data set reach 1.342 and 1.145 m, respectively, while the mean and median error on the Broun Hall data set are 1.374 and 0.94 m, respectively. In MapLoc, the mean and median error are further reduced by augmenting the training data set with the artificial data generated by sampling the uncertainty maps. In the public data set scenario, the mean and median errors decrease to 1.234 and 1.031 m, respectively. The mean error on the Broun Hall data set reduces from 1.374 to 1.211 m, whereas the median error reaches 0.9722 m.

To better examine the performance of different localization methods, the localization times in the two scenarios are also presented in Table I. Because LSTM, LSTM+DeepMap, and MapLoc rely on the location predication model to estimate the location of newly received signals, their location estimation times are much smaller than that of DeepMap. LSTM+DeepMap and Maploc share the same location prediction model in terms of system architecture. As a result, their location estimation times are quite close. The only distinction between these two approaches is their data generation method. Uncertainty information is not considered in LSTM+DeepMap method, but used in MapLoc for training the position prediction model. Considering that such data generation is only implemented in the offline training stage, the location error decreases without the cost of increasing localization delay in the online stage. This is a critical reason why we introduce uncertainty information to the location prediction model. Furthermore, because the LSTM location prediction model excludes the DNNS component, the localization delay is slightly less than that of MapLoc and LSTM+DeepMap. DeepMap requires that the freshly received signal measurement be compared to each signal in the generated radio map. Thus, the size of the indoor environment and the resolution of the radio map determine the location estimation time of DeepMap. It takes the longest delay to localize the mobile device in both cases.

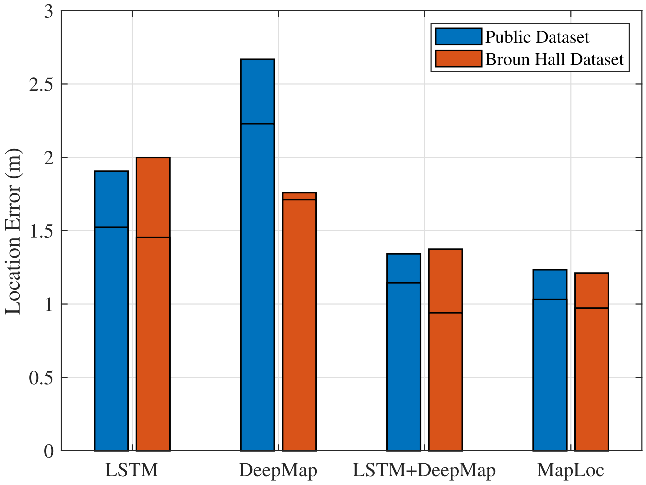

Fig. 11. Mean and median localization errors on the Broun Hall data set and the public data set. The bars indicate mean error and the line within each bar indicates the corresponding median error.

TABLE I
LOCALIZATION TIME

	DeepMap	LSTM	LSTM+DeepMap	MapLoc
Public Dataset	6.901s	0.001s	0.00119s	0.00124s
Broun Hall Dataset	19.953s	0.00135s	0.001695s	0.001690s

2) Impact of Signal Selection: Previous results show that the Maploc system outperforms the systems that use LSTM and DGP separately. By leveraging the samples sampled from the uncertainty maps to measure the reliability of different signal sources (e.g., APs), the MapLoc system also beats the combination of LSTM and DGP. To investigate how the reliability of signal measurements affects MapLoc's location

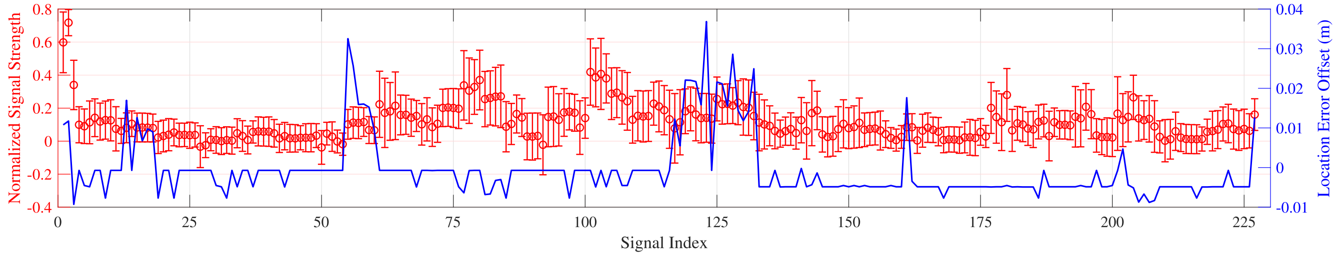

Fig. 12. Explaining the importance of different signal measurements to the location prediction made by MapLoc system with the Broun hall data set.

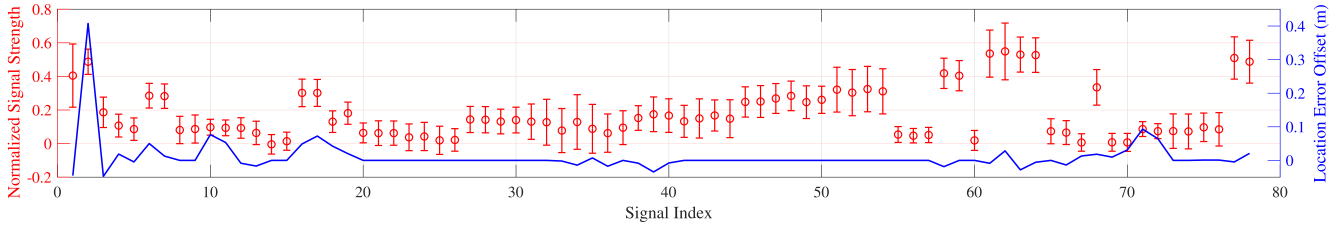

Fig. 13. Explaining the importance of different signal measurements to the location prediction made by MapLoc system with the public data set.

prediction and how the location prediction model contributes to higher accuracy, we conduct experiments with both the Broun Hall data set and the public data set.

First, a random trajectory is selected for each test data set. The corresponding signal distribution at the label coordinates is obtained with the DGP model in the MapLoc system. The mean and variance of signal measurements from all the available sources, including magnetic field readings and WiFi RSSI, are represented by red circles and bars in Figs. 12 and 13. It is intuitive to suppose that a lower variance represents a more trustworthy signal measurement, and signal measurements with higher mean values are more likely to influence the location prediction. To endow the MapLoc system with the ability to intelligently choose signals, we sample the uncertainty maps to generate artificial signal measurements. We hope our location prediction model is able to learn how to recognize effective measurements from invalid and fluctuating signals. In the experiments, each time we double the measurement from one source in the testing data, to examine the impact of each signal source (and quality).

As previously stated, the signal measurements in the Broun Hall data set contain 3 magnetic field components as well as RSSI readings from 224 WiFi APs. The experiment is repeated 227 times with the same selected trajectory. In each repetition, we doubled a signal measurement while keeping all the other measurements the same. The blue line in Fig. 12 depicts the variations in location errors caused by the doubled signal measurements. By comparing with the signal quality denoted by the red circles and bars, it is clear that the location prediction model selects the optimal signal measurements, and the location estimation is resilient to most of the noisy signal sources. As shown in Fig. 12, the first increment of error happens at signal-2, which is the magnetic field reading's y-component. The magnetic measurement is much higher and more stable than the nearby signals. The next large fluctuation of error occurs between signal-13 and signal-18. Beyond this range, the mean value of the signals becomes small, and the location estimation of the MapLoc system is not much influenced by these weak signals. The fluctuation in distance errors increase as the signals rise between signal-55 and signal-60, while the increase in error disappears between signal-61 and signal-117. Even though the mean values of the signals between index-61 and index-117 are much higher than others, the location prediction model detects the large variances of the signals, so the location estimation is not significantly affected by these signals. Another wild rise in location estimation is associated with signals near index-125, where the signals remain high. Furthermore, two more error fluctuations occur at signal-161 and signal-227. It is clear that these signals remain stable, and they are stronger than the nearby signals.

Fig. 13 displays the results on the public data set, which include 3 magnetic field components and RSSI readings from 75 WiFi APs. Because the number of signal sources in the public data set is much smaller than that in the Broun Hall data set, the location prediction model in the public data set scenario is more sensitive to the introduced noise (i.e., the doubled signal measurements). Fig. 13 shows the relationship between the error increase and the signal quality. The largest peak in Fig. 13 corresponds to the y-component of the magnetic field reading as well, because the signal component remains stable in a high level status. We also find error fluctuations at signal-6, signal-17, and signal-62. Although these strong signals cause the location error to increase abruptly, comparable changes are also introduced by stable signals but with lower strengths at index-10 and index-71. The location prediction model ignores changes in signal measurements between signal-20 and signal-60. Some signals in this range are stable, but these weak strength would not cause the degradation of location accuracy. Some signals are strong, but the location prediction model discards them due to their poor reliability.

Based on Figs. 12 and 13, we conclude that the proposed location prediction model in the MapLoc system successfully extracts effective signal measurements from the weak

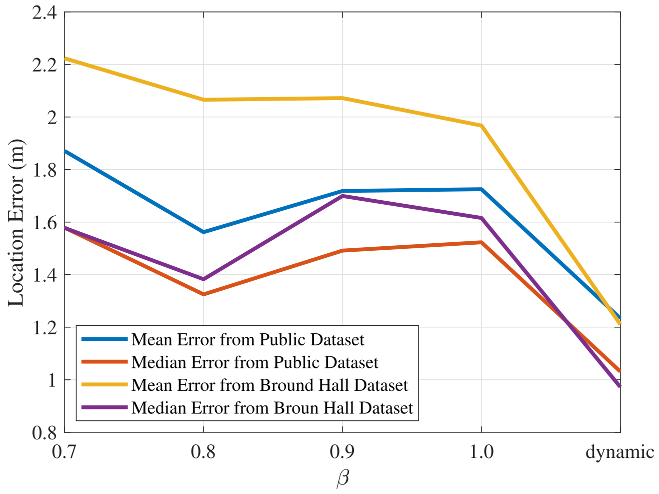

Fig. 14. Mean localization errors for different values of β.

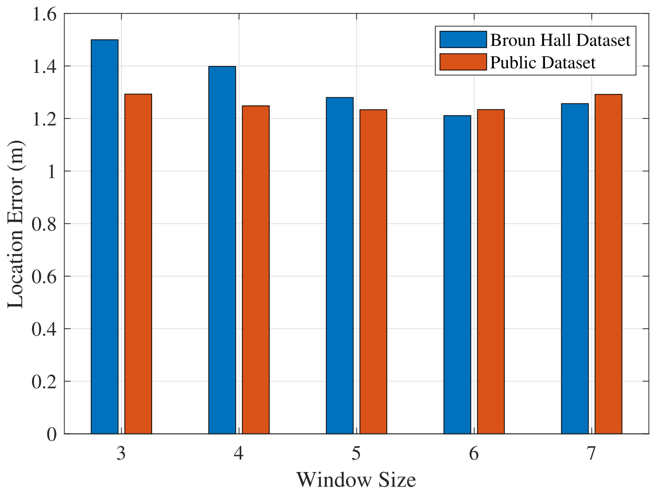

Fig. 15. Localization errors effected by the size of the sliding window.

and fluctuating signals by learning the artificial signal measurements that describe their own reliability. The signals are selected not only by the average signal strength but also by the signal stability.

3) Impact of System Parameters: In MapLoc, the auxiliary loss (12) is used to force the location prediction model to acquire knowledge from the signal measurement generated by the DGP model and estimate the unknown location with the knowledge. A parameter β is introduced to balance the signal loss from DNNS and the location loss from DNNL. We investigate the effect of β on the MapLoc performance. Fig. 14 plots the location errors related to different values of β. In both scenarios, the accuracy of location estimation progresses when β is set as 0.8. Even though the mean error from the Broun Hall data set is slightly decreased as β increases to 1.0, the overall performance does not enhance significantly with the increment of β. Intuitively, β is affected by the indoor environment. The RSSI is reasonable in an open and static environment and closely follows the signal-to-distance relationship. As a result, a lower β would be more beneficial to localization. In an extremely cluttered and complex indoor environment, on the other hand, a larger β could alleviate the localization problem as a fingerprinting-like issue. Considering that signal estimation in the location prediction model is a supportive method for accurate location estimation, we adopted a dynamic approach to adjust β based on the number of epochs. In the MapLoc system, the initial value of β is set as 0.6. When more than 200 epochs are completed, β is updated every 100 epochs by a 0.1 decrease. Eventually, the auxiliary loss would degenerate into a loss function determined by the location estimation error exclusively. Fig. 14 exhibits the performance gains contributed by the dynamic β update approach.

Given that the location prediction model in the MapLoc system relies on the stacked LSTM network as its backbone, the window size W plays a crucial role in the accuracy of location prediction. Intuitively, a longer data sequence would contain more information for location estimation; nevertheless, a longer sequence would incur additional system costs, such as an extra time cost in data collection. To study the effect of the sequence length (i.e., window size W), we conduct experiments with different window sizes on both the public and the Broun Hall data set. Fig. 15 illustrates the errors resulted by different window sizes. Even though the error in the Broun Hall scenario is more sensitive to the change of the window size, the location errors drop with incrased W in both environments. When the window sizes are larger than 5, the errors remain stable. The public data set has the lowest location error of 1.233 m when the window size is $W = 5$, and the error is 1.234 m when the window size is $W = 6$. Because the Broun Hall Data set has the lowest distance error, 1.21 m, when the window size is 6, we set the window size to 6 in the MapLoc system to ease system setup. According to Fig. 15, the location error rises slightly when the window size is 7 in both environments. Compared with traditional approach of forecasting with time series that benefits from long-term dependencies, location prediction is a special case, where the prediction is highly related to the most recent historical location, and obsolete location information may introduce more noise. Thus, a suitable sequence length would contribute to an improved location estimation.

The MapLoc system uses multimodal data as mentioned in previous sections. With the least amount of data processing, different types of signal measurements could be introduced into the system. In this prototype, magnetic field readings are used as a part of the MapLoc system's input data. Magnetic field components from different directions are treated as novel features of input after the max–min normalization. Fig. 16 illustrates the advancement brought by the bimodal data, which is composed of both magnetic field and WiFi RSSI measurements. The localization errors of the public data set is reduced notably. The mean error drops from 1.577 to 1.234 m, when the magnetic filed data is taken into account, whereas the decline of median error reaches 0.327 m. A similar phenomenon happened to the Broun Hall data set as well; both mean error and median error are reduced remarkably. A huge reduction of the mean error appears when magnetic field readings are used, where the mean distance error decreases from 1.719 to 1.211 m. According to Fig. 16, multisource data

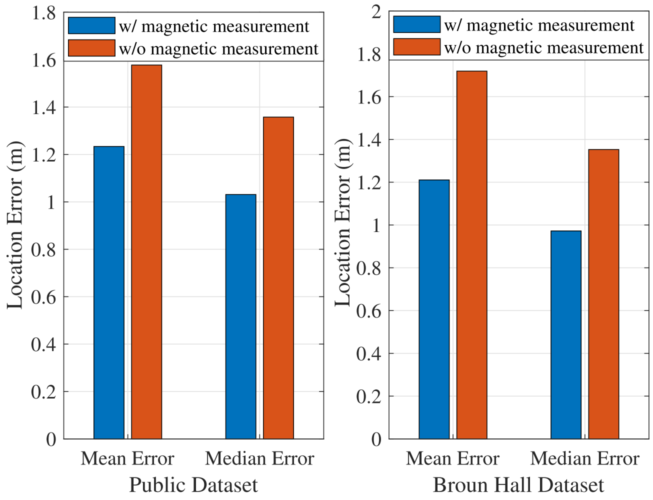

Fig. 16. Reductions in localization error achieved by utilizing the earth magnetic field strength map.

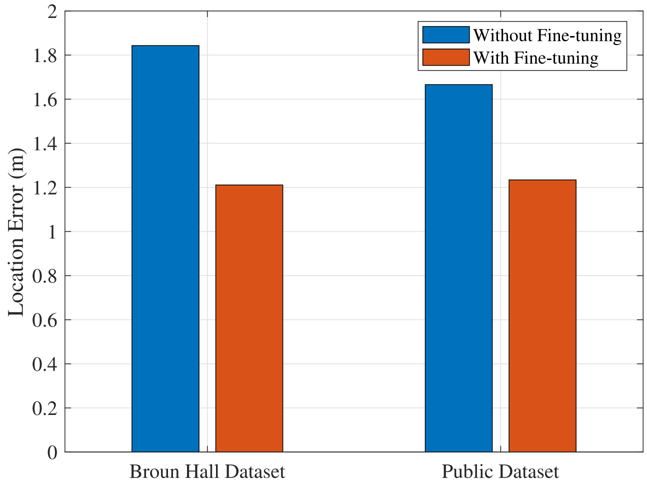

Fig. 17. Reductions in localization error achieved with fine-tuning.

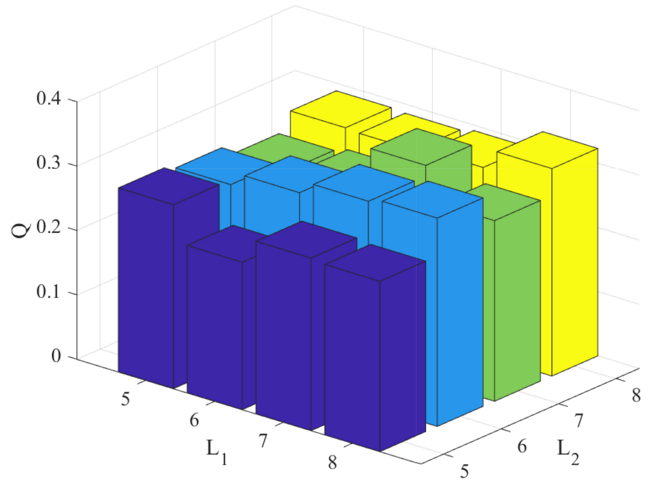

Fig. 18. Map quality Q values versus the numbers of latent dimensions on the public data set.

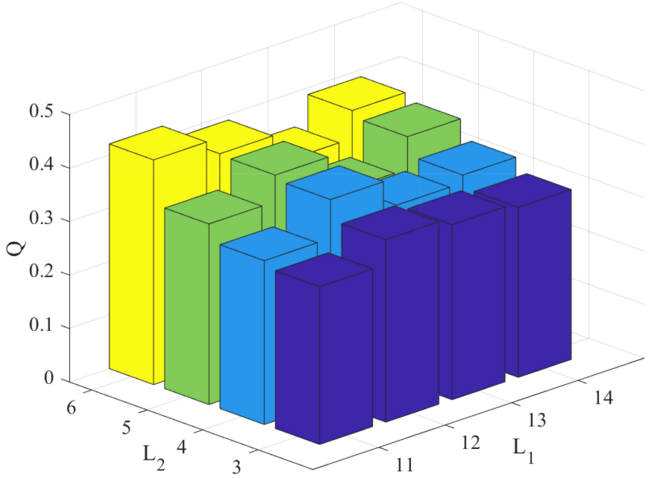

Fig. 19. Map quality Q values versus the numbers of latent dimensions on the Broun Hall data set.

improves location accuracy by adding more features to represent the signal space [44]. In dealing with location-based problems, different forms of signals would play complementary roles. Based on the enriched features, our proposed system compiles and filters information from many sources to enhance localization accuracy.

In MapLoc, the location prediction model is trained using the data generated by the DGP model initially, and then fine-tuned with field-collected data. The data collected from the field is initially treated as discrete in order to train the DGP model, and then it is recovered as sequences to fine-tune the location prediction model. During the offline training, the resulting DGP model may not be perfect and, thus, the data sampled by the model is usually defective. To compensate this, the location prediction model is then fine-tuned at the end of the training process using field-collected data. The reduction in localization error achieved by fine-tuning is shown in Fig. 17. The location errors decrease dramatically in both cases. The location error decreases from 1.843 to 1.211 m for the Broun Hall data set and from 1.666 to 1.234 m for the Public data set. Given that fine-tuning is only employed in the offline training, the gain in location accuracy brought by fine-tuning would not cause any additional computational cost in the online location prediction.

Because the location prediction model is pretrained with artificial data sampled from the uncertainty maps, the DGP model is a critical component. To assess the impact of the DGP model parameters on the quality of the uncertainty map, Q, we investigate various combinations of latent dimension and number of inducing points, aiming to find the best configuration of DGP.

The latent nodes in MapLoc include two sublayers, H_1 and H_2. Figs. 18 and 19 show how the maps' quality Q (defined in Algorithm 1) is affected by the latent dimensions of the two sublayers, denoted by L_1 and L_2, respectively. The latent dimensions are tuned by gridpoint search in both scenarios. We first examine the effect of latent dimensions using the public data set. Even though the quality of the uncertainty map increases with larger dimensions of the first layer when the second layer has six or eight latent dimensions, the relationship between the latent dimensions and map qualities is ambiguous. As shown in Fig. 18, the uncertainty map reaches the

TABLE II
MAP QUALITY Q AFFECTED BY THE NUMBER OF INDUCING POINTS

Public Dataset								Broun Hall Dataset					
Inducing Points	11	10	9	8	7	6	5	Inducing Points	16	15	14	13	12
Q	0.248	0.317	0.274	0.275	0.267	0.266	0.240	Q	0.318	0.424	0.387	0.360	0.321

highest value when $L_1 = 7$ and $L_2 = 7$. Also, two similar Qs are achieved when the latent dimension of the first layer is $L_1 = 8$. Because all the three Qs are close, there is no clear advantage to use different latent dimension settings. We try all the three settings in training the location prediction model of MapLoc. Since the lowest validation error is reached when $L_1 = 8$ and $L_2 = 6$, we choose this setting for the public data set. The previous MapLoc results are all obtained under this setting.

On the other hand, Fig. 19 reveals that the quality of uncertainty maps, Q, improves with increased latent dimensions of the second sublayer L_2 on the Broun Hall data set. However, increasing the latent dimension of the first sublayer L_1 does not necessarily improve map quality. We find that increasing L_2 significantly improves the map quality Q when $L_1 = 11$; whereas increasing L_1 does not contribute to further improvement of Q. According to Fig. 19, a gridpoint search yields the best map quality, i.e., $Q = 0.42$, for the Broun Hall data set when $L_1 = 11$ and $L_2 = 6$.

Another key factor effecting the quality of the maps is the number of inducing points. For the DGP model, we choose identical numbers of inducing points for different layers to simplify the setting. Similarly, we evaluate the effect of the number of inducing points on the quality of the uncertainty maps with both data sets. Table II presents the map quality Q obtained by different numbers of inducing points with the public data set. According to the table, the worst map quality is acquired when each layer of the DGP model only includes five inducing points. Along with the increasing number of inducing points, the map quality keeps enhancing. Even though the growth rate for the map quality is slow when the number of inducing points is between 6 and 9, a notable gain is observed when 10 inducing points of each layer are involved in the training of the DGP model. Thus, the number of inducing points is set to 10 for the public data set. The map quality stops improving as the number of inducing number reaches 11, where the map quality is close to that of the model with 5 inducing points in each layer.

For the Broun Hall data set, Table II reveals a similar trend regarding the number of inducing points. When the number of inducing points is fewer than 16, the upward trend in the map quality Q is conspicuous. The map quality progresses consistently with the increasing number of inducing points. As shown in Table II, the best map quality is achieved when the number of inducing points rise to 15. However, if the number of inducing points exceeds 16, the map quality drops considerably. Therefore, the number of inducing points is set as 15 for the Broun Hall data set.

TABLE III
MAP QUALITY Q AND DGP TRAINING TIME VERSUS THE NUMBER OF LAYERS ON THE PUBLIC DATA SET

	Structure	DGP Training Time	Q
2-layer	8–6	886.56s	0.317
3-layer	8–8–6	1532.89s	0.258
4-layer	8–8–6–6	1630.685s	0.200

TABLE IV
MAP QUALITY Q AND DGP TRAINING TIME VERSUS THE NUMBER OF LAYERS ON THE BROUN HALL DATA SET

	Structure	DGP Training Time	Q
2-layer	11–6	2620.03s	0.424
3-layer	11–11–6	2790.51s	0.311
4-layer	11–11–6–6	5256.36s	0.105

To evaluate the impact of the number of latent layers on map quality Q, we used three distinct layer configurations with two data sets in this experiment. In Tables III and IV, the number of latent dimensions follows previous setup to simplify the model structure. We first add a new layer with the same latent dimension of H_1 and then appended a layer with the same latent dimension of H_2. The number of inducing points is set to 10 for the public data set and 15 for the Broun Hall data set, respectively. It is evident that increasing the number of latent layers from 2 to 4 doubles the DGP training time. However, the map quality Q does not benefit much from the increased number of latent layers in both scenarios. Intuitively, adding more layers will help the models extract more features from training data, but increasing the training and inference cost. Nevertheless, when dealing with a data set of a limited size, adding more layers may not always result in enhanced performance. DGP is used as a data creation tool in this work. The exact precision is crucial, but it is not the only metric that we care about. The model assists us in representing the uncertainty information of APs at various locations for the following location prediction model. The effect of the imperfect DGP model would be adjusted since the location prediction model would be fine-tuned with field-collected data. Thus, we chose a simplified two-layer model in this work to reduce the cost of DGP training.

VI. CONCLUSION

In this article, we proposed MapLoc, a indoor localization system using multimodal data. In MapLoc, DGP was used to regress uncertainty maps describing the signal distribution in the service area. The artificial signal measurements that represent their own reliability were generated by sampling the signal distribution described by the uncertainty maps. In the artificial data generation, geometry constraints and user motion patterns were also taken into account. We then presented a location prediction model to distinguish the effective signal measurements from the weak and fluctuating signals by learning the artificial signal measurements. The location prediction model leveraged a stacked LSTM network as its backend. The auxiliary output was utilized to push the model to learn the signal map in supervised training. The experimental results demonstrated that the location prediction model was able to intelligently choose the optimal signals among WiFi RSSI readings and geomagnetic measurements. Benefiting from the novel data generation method and location prediction model, the median error of location estimation in both the data sets reached centimeter-level accuracy.

REFERENCES

[1] P. Bahl and V. N. Padmanabhan, "Radar: An in-building RF-based user location and tracking system," in *Proc. IEEE INFOCOM*, Tel Aviv, Israel, Mar. 2000, pp. 775–784.

[2] B. Jia, B. Huang, H. Gao, W. Li, and L. Hao, "Selecting critical WiFi APS for indoor localization based on a theoretical error analysis," *IEEE Access*, vol. 7, pp. 36312–36321, 2019.

[3] P. Huang, H. Zhao, W. Liu, and D. Jiang, "MAPS: Indoor localization algorithm based on multiple AP selection," *Mobile Netw. Appl.*, vol. 26, no. 2, pp. 649–656, Apr. 2021.

[4] X. Shi, J. Guo, and Z. Fei, "WLAN fingerprint localization with stable access point selection and deep LSTM," in *Proc. IEEE ICICN*, Xi'an, China, Aug. 2020, pp. 56–62.

[5] X. Wang, L. Gao, S. Mao, and S. Pandey, "BiLoc: Bi-modal deep learning for indoor localization with commodity 5GHz WiFi," *IEEE Access*, vol. 5, pp. 4209–4220, 2017.

[6] X. Wang, Z. Yu, and S. Mao, "Indoor localization using magnetic and light sensors with smartphones: A deep LSTM approach," *Mobile Netw. Appl.*, vol. 25, no. 2, pp. 819–832, Apr. 2020.

[7] A. H. Salamah, M. Tamazin, M. A. Sharkas, and M. Khedr, "An enhanced WiFi indoor localization system based on machine learning," in *Proc. IEEE IPIN*, Madrid, Spain, Oct. 2016, pp. 1–8.

[8] E. Jedari, Z. Wu, R. Rashidzadeh, and M. Saif, "Wi-Fi based indoor location positioning employing random forest classifier," in *Proc. IEEE IPIN*, Oct. 2015, pp. 1–5.

[9] X. Wang, L. Gao, S. Mao, and S. Pandey, "DeepFi: Deep learning for indoor fingerprinting using channel state information," in *Proc. WCNC*, New Orleans, LA, USA, Mar. 2015, pp. 1666–1671.

[10] X. Wang, X. Wang, and S. Mao, "CiFi: Deep convolutional neural networks for indoor localization with 5 GHz Wi-Fi," in *Proc. IEEE ICC*, Paris, France, May 2017, pp. 1–6.

[11] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "ConFi: Convolutional neural networks based indoor Wi-Fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18066–18074, 2017.

[12] X. Wang, X. Wang, and S. Mao, "ResLoc: Deep residual sharing learning for indoor localization with CSI tensors," in *Proc. IEEE PIMRC*, Montreal, QC, Canada, Oct. 2017, pp. 1–6.

[13] X. Wang, X. Wang, S. Mao, J. Zhang, S. Periaswamy, and J. Patton, "Indoor radio map construction and localization with deep Gaussian processes," *IEEE Internet Things J.*, vol. 7, no. 11, pp. 11238–11249, Nov. 2020.

[14] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "FILA: Fine-grained indoor localization," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 2210–2218.

[15] A. Chekuri and M. Won, "Automating WiFi fingerprinting based on nano-scale unmanned aerial vehicles," in *Proc. IEEE VTC*, Jun. 2017, pp. 1–5.

[16] S. Piao, Z. Ba, L. Su, D. Koutsonikolas, S. Li, and K. Ren, "Automating CSI measurement with UAVs: From problem formulation to energy-optimal solution," in *Proc. IEEE INFOCOM*, Paris, France, Apr./May 2019, pp. 2404–2412.

[17] Z. Wang, L. Liu, J. Xu, and L. Zeng, "A UAV path and action planning algorithm for indoor localization information collection," in *Proc. IEEE BDCloud*, Exeter, U.K., Oct. 2020, pp. 609–616.

[18] H. Zou et al., "Adversarial learning-enabled automatic WiFi indoor radio map construction and adaptation with mobile robot," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 6946–6954, Aug. 2020.

[19] C. Wu, Z. Yang, and C. Xiao, "Automatic radio map adaptation for indoor localization using smartphones," *IEEE Trans. Mobile Comput.*, vol. 17, no. 3, pp. 517–528, Mar. 2017.

[20] B. Huang, Z. Xu, B. Jia, and G. Mao, "An online radio map update scheme for WiFi fingerprint-based localization," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6909–6918, Aug. 2019.

[21] Z. Gu, Z. Chen, Y. Zhang, Y. Zhu, M. Lu, and A. Chen, "Reducing fingerprint collection for indoor localization," *Comput. Commun.*, vol. 83, pp. 56–63, Jun. 2016.

[22] J. Luo and L. Fu, "A smartphone indoor localization algorithm based on WLAN location fingerprinting with feature extraction and clustering," *Sensors*, vol. 17, no. 6, p. 1339, Jun. 2017.

[23] X. Wang, L. Gao, and S. Mao, "PhaseFi: Phase fingerprinting for indoor localization with a deep learning approach," in *Proc. GLOBECOM*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[24] P. Yazdanian and V. Pourahmadi, "DeepPos: Deep supervised autoencoder network for CSI based indoor localization," Nov. 2018, *arXiv:1811.12182*.

[25] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Reddy, "Recurrent neural networks for accurate rssi indoor localization," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10639–10651, Dec. 2019.

[26] Z. Chen, H. Zou, J. Yang, H. Jiang, and L. Xie, "WiFi fingerprinting indoor localization using local feature-based deep LSTM," *IEEE Syst. J.*, vol. 14, no. 2, pp. 3001–3010, Jun. 2019.

[27] M. Ibrahim, M. Torki, and M. ElNainay, "CNN based indoor localization using RSS time-series," in *Proc. IEEE ISCC*, Jun. 2018, pp. 1044–1049.

[28] J. Xiong and K. Jamieson, "Arraytrack: A fine-grained indoor location system," in *Proc. ACM NSDI*, Lombard, IL, USA, Apr. 2013, pp. 71–84.

[29] C. Yang, X. Wang, and S. Mao, "SparseTag: High-precision backscatter indoor localization with sparse RFID tag arrays," in *Proc. IEEE SECON*, Boston, MA, USA, Jun. 2019, pp. 1–9.

[30] Y. Kang, Q. Wang, J. Wang, and R. Chen, "A high-accuracy TOA-based localization method without time synchronization in a three-dimensional space," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 173–182, Jan. 2019.

[31] W. Yuan, N. Wu, Q. Guo, X. Huang, Y. Li, and L. Hanzo, "TOA-based passive localization constructed over factor graphs: A unified framework," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 6952–6965, Oct. 2019.

[32] A. Morar et al., "A comprehensive survey of indoor localization methods based on computer vision," *Sensors*, vol. 20, no. 9, p. 2641, May 2020.

[33] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.

[34] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proc. IEEE ICRA*, Shanghai, China, May 2011, pp. 3400–3407.

[35] Q. Niu, M. Li, S. He, C. Gao, S.-H. G. Chan, and X. Luo, "Resource-efficient and automated image-based indoor localization," *ACM Trans. Sens. Netw.*, vol. 15, no. 2, pp. 1–31, May 2019.

[36] J. Luo, L. Fan, and H. Li, "Indoor positioning systems based on visible light communication: State of the art," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2871–2893, 4th Quart., 2017.

[37] J. Armstrong, Y. A. Sekercioglu, and A. Neild, "Visible light positioning: A roadmap for international standardization," *IEEE Commun. Mag.*, vol. 51, no. 12, pp. 68–73, Dec. 2013.

[38] S. Zhu and X. Zhang, "Enabling high-precision visible light localization in today's buildings," in *Proc. ACM MobiSys*, Jun. 2017, pp. 96–108.

[39] H. Chen, F. Li, and Y. Wang, "EchoTrack: Acoustic device-free hand tracking on smart phones," in *Proc. IEEE INFOCOM*, Atlanta, GA, USA, 2017, pp. 1–9.

[40] Y. Liu et al., "Vernier: Accurate and fast acoustic motion tracking using mobile devices," *IEEE Trans. Mobile Comput.*, vol. 20, no. 2, pp. 754–764, Feb. 2021.

[41] A. Damianou and N. D. Lawrence, "Deep Gaussian processes," in *Proc. 16th Int. Conf. Artif. Intell. Stat.*, Scottsdale, AZ, USA, Apr./May 2013, pp. 207–215.

[42] P. Barsocchi, A. Crivello, D. La Rosa, and F. Palumbo, "A multisource and multivariate dataset for indoor localization methods based on WLAN and geo-magnetic field fingerprinting," in *Proc. IEEE IPIN*, Madrid, Spain, Oct. 2016, pp. 1–8.

[43] J. Yang and Y. Chen, "Indoor localization using improved RSS-based lateration methods," in *Proc. IEEE GLOBECOM*, Honulolo, HI, USA, Dec. 2009, pp. 1–6.

[44] W. Zhao, S. Han, W. Meng, D. Sun, and R. Q. Hu, "BSDP: Big sensor data preprocessing in multi-source fusion positioning system using compressive sensing," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8866–8880, Sep. 2019.