Broad Learning System for Indoor CSI Fingerprint Localization

Chieh Yu†, Jang-Ping Sheu‡, and Yung-Ching Kuo‡
†Intelligent Production and Intelligent Manufacturing Master Program
‡Department of Computer Science, National Tsing Hua University, Taiwan
Emails: s109136503@m109.nthu.tw, sheujp@cs.nthu.edu.tw, and cludbuster@gmail.com

2023 IEEE Wireless Communications and Networking Conference (WCNC) | 978-1-6654-9122-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/WCNC55385.2023.10119111

***Abstract*—With the development of the Internet of Things (IoT), the demand for location-based services in indoor environments proliferates. Channel State Information (CSI) based fingerprint localization has become a research hot spot. However, a compelling method to address this issue does not appear because of the hardship of processing CSI data. Meanwhile, many fingerprint localization algorithms have a time-consuming offline training phase. Therefore, we propose an indoor CSI fingerprint localization system based on the broad learning system (BLS). First, we filter the outliers, generate delegates of CSI by the data nugget algorithm, and use tensor decomposition to reconstruct the CSI delegates. Moreover, we utilize Isometric mapping (Isomap) to extract CSI features to reduce BLS's complexity. The experimental results show that our scheme outperforms several existing algorithms in two indoor environments.**

***Index Terms*—Channel State Information, Fingerprint, Indoor Localization, Broad Learning System**

I. Introduction

Location-based Services (LBS) are widely used in people's lives, and their technologies have become a research hotspot. The Global Positioning System (GPS) [1] is the mainstream for localization. However, it cannot apply in indoor localization due to the obstruction of the buildings. Therefore, many approaches were proposed to address the issue of indoor localization, and these methods can be classified into two main categories: geometric-based and fingerprint-based. The geometry-based localization methods use distance measurements such as Angle-of-Arrival (AoA) or Time-of-Arrival (ToA) to compute position by angle and range. However, these methods failed in the non-Line-of-Sight (NLOS) environments as the angle or range errors would lead to significant localization errors [2].

In contrast, fingerprint-based localization methods map the locations with the signal measurements, such as Received Signal Strength Indication (RSSI) or CSI. These signals can be collected without strict constraints on environments and devices. The fingerprint methods have the advantages of solid environmental adaptability and easy implementation. However, multiple paths and signal blocking will significantly impact RSSI in indoor environments [3]. On the contrary, CSI is a fine-grained physical layer data, which provides the state of signal propagation in each sub-carriers (channels), including scattering, fading, and power decay. Thus, CSI has become one of the most commonly used signal measurements in indoor fingerprint localization.

Most CSI-based localization methods have the preprocessing phase to alleviate the noise and extract the information from CSI data, which is significant in CSI fingerprint localization. There are many preprocessing methods such as, the Hampel filter [4] [5] can filter the outlier of CSI data and tensor decomposition [6] for CSI noise reduction. Furthermore, to reduce the complexity of CSI data, dimension reduction techniques, such as Principal Components Analysis (PCA) [7] and Linear Discriminant Analysis (LDA) [8], were used to extract the critical information from CSI data. The above techniques only consider the covariance or the scatter in the CSI data. The manifold and relative position of adjacent data in the raw CSI data will be eliminated in the process. Therefore, manifold learning techniques such as Locally Linear Embedding (LLE) [9] was proposed, which can keep the manifold of original CSI data. However, manifold learning techniques suffer from the severe time cost of large-scale data and the weakness of addressing outliers and noise [10].

After preprocessing, Fingerprint-based localization methods use statistical analysis or machine learning models to formulate the relationship between location and signal measurements. These methods can be divided into two categories: classification and regression. The classification-based localization methods learn a discrete class output (reference points, RPs) by the input (CSI) and use the likelihood of each RP as the weight to compute a position. In [8], the authors proposed a naive Bayes classifier to predict the position. Besides, neural network-based models, which have a strong ability to handle non-linear data, were proposed. In [5], Depthwise Separable Convolution Neural Network was utilized to solve the fingerprint localization problem. The classification models have promising results in predicting RPs. However, CSI data with similar characteristics increase when the RPs increase. Thus the likelihood provided by the classification models spreads to several RPs leading to an inaccurate localization result.

On the contrary, the regression-based localization methods build the transformation for mapping input (CSI) to continuous quantity output (position of RPs). Therefore, the regression models can directly provide the position to be the localization result. In [11], the authors used the Gaussian process regression (GRP) to estimate the position. The authors in [6] proposed partial least squares regression (PLSR) to predict the

This work was supported by Qualcomm Technologies, Inc. under grant SOW NAT-435533.

position. However, normal regression models do not provide promising localization results because CSI data do not have a linear relationship with the positions. Therefore, the authors in [7] proposed a naive Bayes classifier combining BLS, which is a neural networks-based model with only one layer and no back-propagation, to solve fingerprint localization.

This paper proposes an indoor CSI fingerprint localization system based on BLS, named ICLoc. In this system, we utilize user equipment (UE) to be a receiver and a WiFi AP to be a transmitter. Due to the CSI cannot be directly obtained in UE, we use the Raspberry PI 4 (PI4) with Nexmon framework [12] as a monitor, which uses frame injection to modify the WiFi chip in the UE. While UE receives the data transmitted from AP, UE will send UDP packets to PI4 containing the information of the physical layer. PI4 can extract CSI from the received packets as fingerprints. The collected CSI packets go through a uniform preprocessing method, which applies the data nugget algorithm [13], tensor decomposition [14], and Isomap [15] to perform CSI noise reduction and feature extraction to improve indoor localization accuracy. The processed CSI packets will be input into the BLS for training and testing. The experimental results confirm that our system has better performance than previous works.

In the remainder of this paper, the system model is presented in section II. Then, we describe our algorithm in section III. In section IV, we provide experimental validation and analysis. Finally, we conclude our work in section V.

II. System Model

A. Channel State Information

CSI is physical layer information that can represent the channel characteristics. It reflect the signal propagation between transmitter and receiver, containing the scattering caused by the surroundings, signal attenuation during propagation, etc. Therefore, CSI is fine-grained information having more physical information than RSSI. The authors in [5] mentioned the CSI amplitude reflects the characteristics of the channel and shows excellent stability. This paper uses CSI amplitude as the fingerprints, to prevent the machine-learning model overfit the non-characteristic data.

In this paper, the raw CSI packets collected from PI4s are 802.11ac with 80MHz bandwidth, which consists of 234 sub-carriers, 14 null sub-carriers, and 8 pilot sub-carriers. The null sub-carriers do not transmit data, and the pilot sub-carriers do not carry user data that can be used to measure the channel. To improve localization performance, we remove the null and pilot sub-carries. Therefore, the CSI packets utilized in our system are 234 CSI amplitude of sub-carriers.

B. Architecture of System Model

Our system uses a notebook with Windows 10 as a WiFi AP, an HTC 10 EVO mobile phone as a receiver, and PI4s to monitor the CSI when the AP transmits data to the mobile phone. Fig. 1 shows the structure of our system model. The architecture can be divided into two parts, data preprocessing and BLS localization. In the data preprocessing part, we use the data nugget algorithm to filter the outliers and generate the delegates of CSI. Thus, we can substantially accelerate the afterward processes. In addition, the representatives of CSI will be reconstructed by tensor decomposition to alleviate the noise. Moreover, we used Isomap to perform dimension reduction, which can improve localization performance by reducing the complexity of BLS. In the BLS localization part, we use BLS as a regression model for localization, which can provide a promising performance on localization accuracy and learning speed.

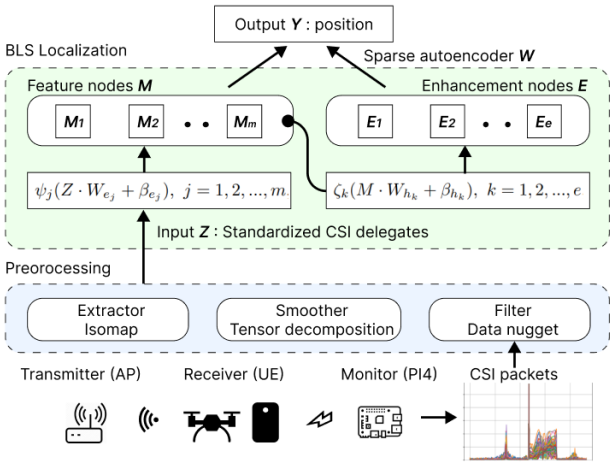


Fig. 1: System model of ICLoc

III. Our Indoor Localization Algorithm

This section introduces our proposed Indoor CSI fingerprint Localization system, ICLoc. This system includes CSI packet preprocessing and BLS localization model.

A. Data Preprocessing

Assume there are n RPs. For each RP, p CSI packets are collected, and each CSI packet contains s sub-carriers. The CSI packets in the ith RP are denoted as $H_i \in \mathbb{R}^{p \times s}$, for $1 \leq i \leq n$. First, we use the data nugget algorithm for each H_i to remove its outliers and generate p_{ng} CSI delegates, denoted as $X_i \in \mathbb{R}^{p_{ng} \times s}$, where $p_{ng} << p$. The CSI delegates X_i can accelerate processes afterward, especially since the computational complexity of Isomap is $O(n^3)$. In our experiments, we set $p = 500$ and $p_{ng} = 40$, which means that we use 40 CSI delegates as the input for BLS localization instead of using the raw 500 CSI packets at a specific RP. Detailed steps of the data nugget algorithm for each H_i are as follows.

Step 1: Compute the center C of H_i, and find the two farthest packets, x_1 and x_2 to C.

Step 2: Compute the Euclidean distances between p packets in H_i with x_1 and x_2. Let the distances of p packets to x_1 (x_2) be d_1 (d_2), which is defined as follows.

$$d_j = [d_{j,1}, d_{j,2}, ..., d_{j,p}], j = 1, 2. \quad (1)$$

Step 3: If $\sum_{k=1}^{p} d_{1,k} > \sum_{k=1}^{p} d_{2,k}$, remove the relative outlier x_1 from H_i. Then, sort d_2 in ascending order, and

split H_i into two data nuggets D_1 and D_2 according to d_2, where D_1 is the top half and D_2 is bottom half. Otherwise, remove x_2 from H_i and split H_i into data nuggets D_1 and D_2 according to sorted d_1.

In step 3, two data nuggets are obtained by splitting H_i. Iterate from step 1 to step 3 on D_1 and D_2 to generate data nuggets until the number of data nuggets is equal to p_{ng}. Finally, CSI delegates X_i can be obtained by computing each center of p_{ng} data nuggets.

To further alleviate the noise, tensor decomposition is utilized to reconstruct X_i to $X_i' \in \mathbb{R}^{p_{ng} \times s}$. In this paper, canonical decomposition/parallel factors (CP) decomposition [14] is performed. Let $a_j \in \mathbb{R}^{p_{ng}}$ and $b_j \in \mathbb{R}^s$ be the jth tensor to represent delegate term and sub-carrier term, respectively. The vectors of a_j and b_j are denoted as $A = [a_1, a_2, ..., a_r] \in \mathbb{R}^{r \times p_{ng}}$ and $B = [b_1, b_2, ..., b_r] \in \mathbb{R}^{r \times p_{ng}}$, respectively. Then, the CP decomposition factorizes the X_i as

$$X_i \approx \sum_{j=1}^{r} a_j \otimes b_j = A \otimes B = X_i', \tag{2}$$

where $\otimes$ denotes the outer product of vectors and $r \in N$ is the number of rank-one tensors used for refactoring X_i. We set $r = 2$ in our experiments.

X_i' can be obtained by Alternating Least Squares (ALS) [14], Which compute A and B from $j = 1$ to r. In initial, each vector b_j is set to $[1, 1, ..., 1]$, for $1 \leq j \leq n$. Then, the vectors A are updated as follows.

$$A \leftarrow arg\min_{A} \parallel X_i - \sum_{j=1}^{r} a_j \otimes b_j \parallel_F^2 . \tag{3}$$

where $\parallel . \parallel_F^2$ is the tensor Forbinius norm, the sum of squares of all elements of the tensors. B is then updated as

$$B \leftarrow arg\min_{B} \parallel X_i - \sum_{j=1}^{r} a_j \otimes b_j \parallel_F^2 . \tag{4}$$

The reconstructed CSI delegates X_i', which has less noise, is obtained. After that, all reconstructed CSI delegates are concatenated as $X' = [X_1', X_2', ..., X_n'] \in \mathbb{R}^{(n \cdot p_{ng}) \times s}$, where n is the number of RPs. Moreover, to reduce the model complexity, Isomap [15] extracts features from X'. Isomap is a geodesic distance-based manifold learning algorithm that can preserve the intrinsic property of the manifold while projecting X' into a low-dimensional subspace. Therefore, s sub-carriers in CSI delegates are transformed to s_{Iso} features, where $1 \leq s_{Iso} \leq s$. We set $s_{Iso} = 30$ in our experiments. In our experiments, we use two PI4s to collect the CSI packets. The concatenation of extracted CSI features $X'' \in \mathbb{R}^{(n \cdot p_{ng}) \times (2 \cdot s_{Iso})}$ is the input to train the BLS model.

B. BLS Training Phase

This paper utilizes BLS as the regression model [16]. The extracted CSI features X'' are the training input, and the training output $Y \in \mathbb{R}^{(n \cdot p_{ng}) \times 2}$ is the Cartesian coordinate of X''. Then, BLS builds the localization model to map the X'' with Y. First, to improve BLS's training accuracy and stability, X'' is standardized to eliminate the covariate shift [17]. The standardization is presented as

$$Z = \frac{X'' - \mu}{\sigma}, \tag{5}$$

where μ and σ are the mean and standard errors of X'', respectively. Next, each data in Z with s_{Iso} features will be transferred to m feature nodes $M = [M_1, ..., M_m] \in \mathbb{R}^{(n \cdot p_{ng}) \times m}$. In our experiments, we set $m = 200$. A nonspecific mapping function can write the jth corresponding feature nodes M_j as

$$M_j = \psi_j(Z \cdot W_{e_j} + \beta_{e_j}),\ j = 1, 2, ..., m, \tag{6}$$

where $\cdot$ denotes the dot product of matrices, $W_{e_j} \in \mathbb{R}^{(2 \cdot s_{Iso})}$ is a random weighted matrix, and $\beta_{e_j} \in \mathbb{R}$ is white noise randomly generated in normal distribution $N(0, 1)$. ψ is the tanh activation function, which is defined as

$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{7}$$

Then the structure of neural network is expanded broadly by e enhancement nodes $E = [E_1, ..., E_e] \in \mathbb{R}^{(n \cdot p_{ng}) \times e}$. We set $e = 50$ in our experiments. The kth corresponding enhancement nodes E_k can be written as

$$E_k = \zeta_k(M \cdot W_{h_k} + \beta_{h_k}),\ k = 1, 2, ..., e, \tag{8}$$

where ζ_k is a tanh function. The $W_{h_k} \in \mathbb{R}^m$ represents the weighted matrix, and $\beta_{h_k} \in \mathbb{R}$ is white noise generated in normal distribution $N(0, 1)$. Finally, the concatenation of feature nodes M and enhancement nodes E connect to the output directly. The formulation of the BLS is represented as

$$Y = [M_1, ..., M_m, E_1, ..., E_e] \cdot W = [M\ E] \cdot W, \tag{9}$$

where $W \in \mathbb{R}^{(m+e) \times 2}$ is the transformation matrix, which transform the concatenation of M and E to the the training output Y. W can be rapidly acquired by the lasso or ridge regression, which can further extract information from the generated nodes M and E [16]. After the BLS model training, we can localize the testing CSI features by the obtained weighted matrices W_{e_j} and W_{h_k}, white noises β_{e_j} and β_{h_k}, and the transformation matrix W.

C. Localization Phase

ICLoc aims to provide accurate indoor localization services to users. In the localization phase, the localization performance is evaluated by the testing CSI packets collected from t RPs. Testing CSI packets at the qth RP are denoted as $U_q \in \mathbb{R}^{p \times s}$, for $1 \leq q \leq t$, where p is the number of packets, and s is the number of sub-carriers. First, the data nugget algorithm and CP decomposition are utilized to obtain reconstructed CSI delegates U_q' from U_q. Next, the testing CSI features are extracted from U', and the concatenation of CSI features from two PI4s $U'' \in \mathbb{R}^{(q \cdot p_{ng}) \times (2 \cdot s_{Iso})}$ is fed into BLS. Finally, BLS provides the position $\hat{y} \in \mathbb{R}^{(q \cdot p_{ng}) \times 2}$ using W_{e_j}, W_{h_k}, β_{e_j}, β_{h_k}, and the transform matrix W obtained from the BLS training phase. Finally, the localization performance is

evaluated by Root Mean Square Error (RMSE), which is shown below.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y-\hat{y})^2}, \tag{10}$$

where y and $\hat{y}$ are the actual location and predicted location of the testing set, respectively. $N = q \cdot p_{ng}$ is the number of testing CSI delegates.

IV. Experiments

We use two environments to verify the proposed indoor location algorithm ICLoc's performance. We also compare the localization performance of the CSI data collected by unmanned aerial vehicle (UAV) and human resources. Details of our experiments are elaborated on the following subsections.

A. Experiment Setup

We use a notebook with Windows 10 as a WiFi AP and HTC 10 EVO as UE in our experiments. Since we cannot obtain the CSI from UE directly, we implement PI4 to monitor the CSI of data transmission between AP and UE. Two PI4s installed Raspberry Pi OS with kernel version 15. AP and UE transmit data in the WiFi channel using 802.11ac with 80 MHz. After dropping null and pilot sub-carriers, we collect 500 CSI packets for each RP and each packet has 234 sub-carriers.

The experiments collected the CSI packets from two environments in the Delta Building at National Tsing Hua University (NTHU). They are a courtyard and a corridor as shown in Fig. 2. The courtyard is a spacious area surrounded by classrooms, and GPS cannot work here due to the signal blocking. The corridor is an L-shaped area surrounded by walls. We collect the CSI packets in the grids of 0.55 m $\times$ 0.55 m in the two environments. For each environment, we conduct different experiments to evaluate the performance of ICLoc. First, we verify the classification performance on the training RPs. Next, in the courtyard, we evaluate the localization performance by RMSE. Finally, we examine the localization result along a human trajectory in the corridor.

We compare the localization performance of ICLoc with PLSR [6], BLS-Loc [7], PAIL [4], and MIFI [5]. The localization algorithm PLSR reconstructs the CSI packet by tensor decomposition and predicts the position by a partial least squares regression model. The BLS-Loc algorithm utilizes PCA and Kalman filter to preprocess the CSI packets. In addition, it combines BLS and naive Bayes classifier to estimate position. The PAIL algorithm process the CSI packets by Hampel filter with LDA and predicts the position by Adaboost based on the naive Bayes classifier. The localization algorithm MIFI utilizes the CSI data smoothed by the Hampel filter and moving average filter, then, it predicts the position based on a depthwise separable convolution neural network.

In section III, the data nugget algorithm can generate p_{ng} CSI delegates from the raw CSI packets to replace the 500 raw CSI packets. Although the data nugget algorithm can accelerate the afterward preprocessing, the loss of information of 500 raw CSI packets is inevitable and might impact the localization performance. According to our experiments, when $p_{ng} < 40$, localization error starts to be higher than using 500 raw CSI packets. Therefore, we set $p_{ng} = 40$ in the following experiments.

In the proposed ICLoc, we transform 234 CSI sub-carriers into s_{Iso} features, where $1 \leq s_{Iso} \leq 234$. In this transformation, if s_{Iso} is too small, localization performance declines because low dimensional CSI features are insufficient to represent the original sub-carriers. On the other hand, a large s_{Iso} leads to the high dimensional CSI features containing redundant components, and the localization performance degrades. According to our experiments, $s_{Iso} = 30$ has the best performance.

B. Classification Performance

Some indoor CSI fingerprint localizations are based on the classification model. We first test our performance with the classification case in the courtyard and corridor, where we collect CSI packets at the RPs as shown in Fig 2. This experiment aims to verify whether the ICLoc can classify the CSI data at the training RPs, it has learned before. We separate the training set (blue grids) into 80% of the sub-training set and 20% of the sub-testing set. It means we utilize 400 packets for training and 100 packets for testing at each training RP. We count the output position as a correct prediction if it is in the corresponding grid.

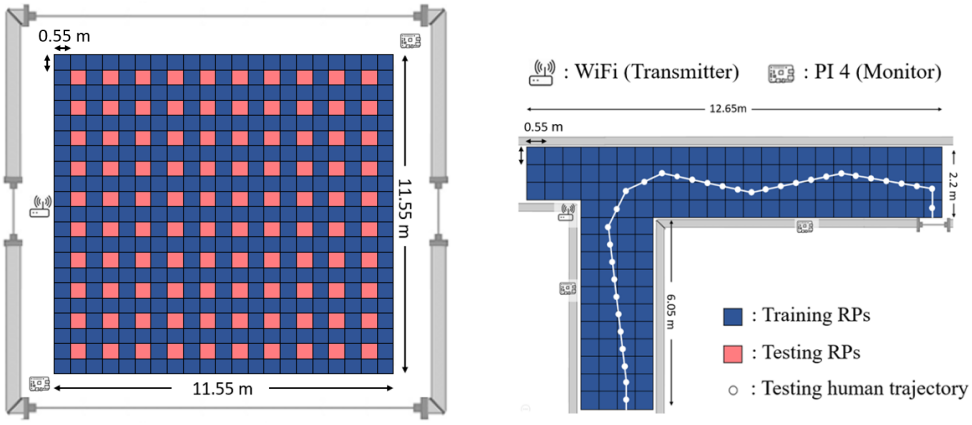
The classification results are shown in Table I. Our ICLoc can reach 100% accuracy in the two environments, which means that the characteristics of CSI at the trained RPs have been learned. We can observe that all approaches have good classification accuracy in this experiment except PLSR. The PLSR is based on a partial least squares regression model. Therefore, the experiment result proves that CSI has no linear relationship between positions. In addition, the result shows that ICLoc, BLS-Loc, and MIFI perform better than PLSR and PAIL, which proves that algorithms utilizing NNs-based models can better learn the characteristics of CSI data. Moreover, we can find that BLS-Loc and PAIL, the algorithms using naive Bayes classifier, provide less localization accuracy in the courtyard compared to the corridor. This result shows that the bigger area might have similar CSI at different RPs, which confuses the weak classifier and degrades accuracy.

TABLE I: Classification accuracy for training RPs

Location	ICLoc	PLSR	BLS-Loc	PAIL	MIFI
Courtyard	100%	72.6%	98.2%	92.2%	100%
Corridor	100%	68.5%	99.6%	97.8%	100%

C. Localization Performance

Here, we evaluate the localization performance of ICLoc in the courtyard. The layout of the training and testing RPs is shown in Fig. 2(a). The RPs we collect CSI packets are 80 cm



(a) The layout of the courtyard (b) The layout of the corridor

Fig. 2: Layout of training and testing RPs

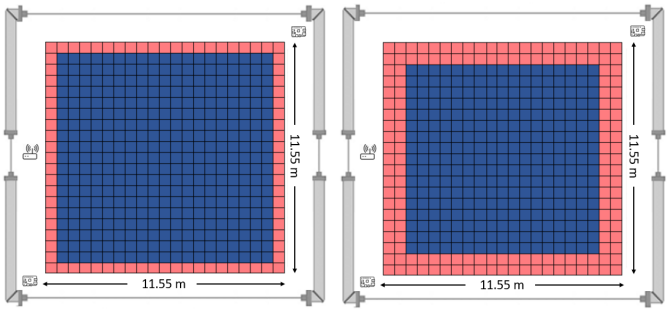
above the ground. We use RMSE as the measurement. This experiment aims to test whether ICLoc can localize the CSI packets collected at the locations that have not been learned.

1) Testing the RPs inside the Training Area: As the typical experiments, we evenly pick up the locations inside the area to be the testing RPs. Table II shows the RMSE performance of our algorithm with the others. In this environment, the localization performance of ICLoc is better than the other approaches and reaches a 1.62m RMSE. We can observe that ICLoc, BLS-Loc, and MIFI perform better than PLSR and PAIL. This result again demonstrates that localization algorithms based on weak learners, such as naive Bayes classifier and partial least squares regression model, can hardly learn CSI data. In addition, the proposed ICLoc outperforms BLS-Loc and MIFI significantly. This result proves that Isomap is better than PCA for extracting the CSI data because Isomap can preserve the manifold in original CSI packets, which is crucial for the machine learning model to learn the position.

TABLE II: Localization result (RMSE) in the courtyard

Location	ICLoc	PLSR	BLS-Loc	PAIL	MIFI
Courtyard	1.62 m	3.18 m	1.98 m	2.73 m	2.06 m

2) Testing the RPs outside the Training Area: To further examine the performance of ICLoc. We test the RPs outside the training RPs. Let o denote the number of outer layers outside the training area. For example, the testing RPs with o = 1 and 2 are shown in Fig. 3(a) and Fig. 3(b), respectively. The localization result is shown in Table III. The proposed ICLoc has 1.86 m RMSE for $o = 1$ and 2.71 m RMSE for $o = 4$, which is significantly better than other approaches. We can observe that the localization error increases as o increases. So, the closer the training and testing RPs, the better the localization performance. In addition, the localization performance of ICLoc and PLSR declines less than other algorithms because the classification model-based algorithms cannot predict the position outside the training area.



(a) Testing RPs with $o = 1$ (b) Testing RPs with $o = 2$

Fig. 3: The layout of different outer layers o in the courtyard

TABLE III: Localization result (RMSE) with different o

	ICLoc	PLSR	BLS-Loc	PAIL	MIFI
o = 1	1.86 m	3.21 m	2.34 m	2.93 m	2.17 m
o = 2	1.98 m	3.45 m	2.87m	3.41 m	2.63 m
o = 3	2.31 m	3.61 m	3.68 m	3.78 m	3.31 m
o = 4	2.71 m	3.92 m	4.17 m	4.36 m	4.01 m

3) Testing the RPs with Human Trajectory: In this experiment, we test RP along a trajectory that simulates human footprints in a corridor. The simulated RPs are shown in Fig. 2(b). As the corridor is surrounded by walls, we correct the location to the nearest position inside the wall when the predicted location is outside the transitable area. Table IV shows the localization results, including the RMSE performance and the ratio of corrected localization. The proposed ICLoc performs the best. Besides, the correction ratio of our approach is better than the baselines.

4) Performance Evaluation with the Help of UAV: To verify the localization performance with the impact of UAV, we use DJI Phantom 3, a UAV with stable hovering, which carries the UE (mobile phone) and collects the CSI packet. The UE is mounted on a UAV, and the height of the UAV is around

TABLE IV: Localization result (RMSE) in the corridor

Location correction	ICLoc	PLSR	BLS-Loc	PAIL	MIFI
Without correction	1.32 m	3.17 m	1.59 m	2.46 m	1.61 m
With correction	1.06 m	2.37 m	1.18 m	1.89 m	1.21 m
Correction ratio	15.6 %	53.1 %	31.2 %	40.6 %	25 %

80 cm, as shown in Fig. 4. As space and safety concerns, we only experiment in the courtyard, which has spacious area for flying UAVs.

Fig. 4: UAV collects CSI packets in the courtyard

The performance in this experiment is shown in Table V. Utilizing the CSI packet collected by the UAV, the proposed ICLoc can achieve a 2.31 m RMSE and outperform others. However, comparing the localization results between the CSI packets collected by the UAV and human resources, the localization error of the UAV is 40 % more than the human resources because the body and fan blade of the UAV aggravate signal blocking, fading, and scattering. Although UAV help reduces human resources, the signal interference of UAV is larger than the human resources. Currently, UAV localization is more appropriate in scenarios that allow larger localization errors or for location classification only.

TABLE V: Localization result (RMSE) in courtyard

	ICLoc	PLSR	BLS-Loc	PAIL	MIFI
Collect by human	1.62 m	3.18 m	1.98 m	2.73 m	2.06 m
Collect by UAV	2.31 m	3.64 m	2.81 m	4.09 m	2.58 m

V. Conclusion

This paper proposes ICLoc, an indoor CSI fingerprint localization system, which conducts a uniform CSI data preprocessing method. The data nugget algorithm, tensor decomposition, and Isomap for performing CSI noise reduction and feature extraction. In addition, we predict the position based on the BLS regression model. The proposed ICLoc outperforms other algorithms in all experiments and verifies that Isomap is remarkable for CSI extraction as it can preserve the manifold in the original CSI. In addition, we demonstrate that weak learners, such as traditional regression models or naive Bayes classifiers, can hardly learn CSI. Meanwhile, we further prove that the regression model is more proper on the issue of indoor CSI fingerprint localization than the classification model. Besides, we compare the localization error through CSI packets collected by UAV and human resources. The experiment result shows that UAV is unsuitable for indoor fingerprint localization due to the severe signal interference of UAV, which has less localization performance than utilizing human resources.

References

[1] K. Jo, K. Chu, and M. Sunwoo, "Interacting multiple model filter-based sensor fusion of GPS with in-vehicle sensors for real-time vehicle positioning," *IEEE Transactions on Intelligent Transportation Systems (T-ITS)*, vol. 13, no. 1, pp. 329–343, 2012.

[2] C. Gao, G. Zhao, and H. Fourati, "Cooperative localization and navigation: Theory, Research, and Practice," *USA: CRC Press*, 2019.

[3] H.-C. Tsai, C.-J. Chiu, P.-H. Tseng, and K.-T. Feng, "Refined autoencoder-based CSI hidden feature extraction for indoor spot localization," *IEEE 88th Vehicular Technology Conference (VTC-Fall)*, Chicago, IL, USA 2018, pp. 1-5.

[4] Z. Liu, D. Liu, J. Xiong, and X. Yuan, "A parallel adaboost method for device-free indoor localization," *IEEE Sensors Journal*, vol. 22, no. 3, pp. 2409–2418, 2022.

[5] B.-Y. Chang and J.-P. Sheu, "Indoor localization with CSI fingerprint utilizing depthwise separable convolution neural network," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Virtual Conference, 2022, pp. 1-6.

[6] Y. Long, L. Xie, M. Zhou, and Y. Wang, "Indoor CSI fingerprint localization based on tensor decomposition," *IEEE/CIC International Conference on Communications in China (ICCC)*, Portland, Oregon, USA, 2020, pp. 1190-1195.

[7] X. Zhu, T. Qiu, W. Qu, X. Zhou, M. Atiquzzaman, and D. Wu, "BLS-Location: A wireless fingerprint localization algorithm based on broad learning," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2021.

[8] D. Liu, Z. Liu, and Z. Song, "LDA-based CSI amplitude fingerprinting for device-free localization," *Chinese Control And Decision Conference (CCDC)*, Hefei, China, 2020, pp. 2020-2023.

[9] K. P. Singh, R. Bhai, V. Mishra, P. Nagar, and J. Kasinayal, "Localization in wireless sensor network using LLE-ISOMAP algorithm," *TENCON 2017 - 2017 IEEE Region 10 Conference*, Penang, Malaysia, 2017, pp. 393-397.

[10] A. C. Neto and A. L. M. Levada, "ISOMAP-KL: a parametric approach for unsupervised metric learning," *33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, virtual conference, 2020, pp. 287-294.

[11] S. S. Moosavi and P. Fortier, "A fingerprint localization method in collocated massive MIMO-OFDM systems using clustering and gaussian process regression," *IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, Victoria, BC, Canada, 2020, pp. 1-5.

[12] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets," *13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, 2019.

[13] Y. Liu, "Advanced Isomap based on data nuggets algorithm," *IEEE 5th International Conference on Computer and Communications (ICCC)*, Chengdu, China, 2019, pp. 57-60.

[14] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, pp. 455–500, 08 2009.

[15] J. B. Tenenbaum, V. D. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, 290, 2319-2323, 2000.

[16] C. L. P. Chen and Z. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 1, pp. 10–24, 2018.

[17] S. Wu, G. Li, L. Deng, L. Liu, D. Wu, Y. Xie, and L. Shi, "$l1$ -norm batch normalization for efficient training of deep neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 7, pp. 2043–2051, 2019.