CL-Loc: A Dual-Branch CNN-LSTM Network for WiFi Fingerprint-Based Indoor Localization

1st Haobing Guo
School of Computer Science and Engineering
Northeastern University
Shenyang, China
ghb2002@stumail.neu.edu.cn

2nd Xiushuang Yi*
School of Computer Science and Engineering
Northeastern University
Shenyang, China
xsyi@mail.neu.edu.cn
*Corresponding author

3rd Wanxin Zeng
Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education
Jilin University
Changchun, China
M18736735877@163.com

4th Daoren Zhu
School of Computer Science and Engineering
Northeastern University
Shenyang, China
zhudaoren123@163.com

5th Rui Wang
School of Computer Science and Engineering
Northeastern University
Shenyang, China
2472053@stu.neu.edu.cn

6th Kungui Zeng
School of Computer Science and Engineering
Northeastern University
Shenyang, China
2471985@stu.neu.edu.cn

***Abstract*—With the growing demand for localization services in smart devices, indoor localization has become increasingly important. Existing indoor localization studies based on WiFi fingerprints have not fully considered the problem of high-dimensional location fingerprints caused by the large number of access points (APs) in indoor environments, resulting in reduced indoor localization accuracy. To solve the above issues, this paper proposes a dual-branch CNN-LSTM indoor localization model named CL-Loc. This model utilizes CNN branches to extract local features from location fingerprints and employs LSTM to capture global relationships between APs. A fusion network then effectively combines these feature relationships from different ranges. Experimental results show that CL-Loc outperforms other mainstream indoor localization models evaluated on open-source datasets. It achieved 100% building prediction accuracy on the UJIIndoorLoc dataset and 95.44% floor prediction accuracy on the Tampere dataset.**

***Index Terms*—Indoor Localization, WiFi Fingerprinting, Long Short-Term Memory Network, Convolutional Neural Network**

I. Introduction

The growing popularity of smart wireless devices has led to an increased demand for localization services. Satellite-based global localization systems, such as GPS, Beidou, and GLONASS, provide high-precision localization outdoors. However, achieving the same level of accuracy indoors is challenging due to signal interference caused by walls [1].

Current studies are exploring the use of widely deployed WiFi facilities for indoor localization. They can be roughly categorized into two approaches: 1) Triangulation Methods: These methods are based on Angle Of Arrival(AOA), Time Of Arrival(TOA), etc.. However, they often suffer from low localization accuracy due to numerous non-line-of-sight (NLoS) conditions and multipath effects of WiFi signals in indoor environments. Additionally, acquiring the necessary localization data typically requires specialized equipment, making widespread implementation challenging [2]. 2) Location Fingerprinting: Use Channel State Information (CSI), Receive Signal Strength (RSS), and other information to establish a unique location fingerprint for each location. Localization is achieved by mapping the location fingerprint to its corresponding physical location. This method does not require specialized measurement equipment and offers greater versatility [3]. This article focuses on the second type of method.

RSS can be obtained from most WiFi devices without requiring specific NIC support. In contrast, the acquisition of CSI is more limited. It can only be measured with the support of specific NIC or devices [4]. Therefore, we choose to focus on the RSS fingerprint localization method. During the offline stage of indoor localization, the location fingerprint database is established by measuring the RSS from all APs at the reference points(RPs). In the online stage, the RSS measured at the target localization point is matched with the fingerprint stored in the database to determine the location of the target localization point [5].

In indoor environments, there are typically many access points (APs), making it challenging to capture the received signal strength (RSS) of all APs during measurements at each RP. This results in a large amount of sparse fingerprint data. Current studies have attempted to use deep neural networks (DNN) [6], convolutional neural networks (CNN) [7] and other methods to model these sparse fingerprints and improve local-

This work was supported by the National Natural Science Foundation of China (NSFC) (No.62032013, No.62132004).

ization accuracy. However, these methods don't fully explore the potential correlation among different APs. To address these issues, this paper proposes an indoor localization model based on a CNN-LSTM dual-branch structure, which is named CL-Loc. It models the features of sparse WiFi fingerprints from two complementary perspectives: on the one hand, a CNN is used to extract local features in the fingerprint; on the other hand, a LSTM network (LSTM) is incorporated to model the correlation between the dimension of fingerprint(i.e., among APs). This model effectively captures both local and global features in the location fingerprint through the feature fusion mechanism, improving the accuracy of indoor localization. The main contributions of this paper are as follows:

1. We propose a WiFi indoor localization model based on a dual-branch structure that combines CNN and LSTM, which effectively addresses the indoor localization task of building, floor, and location coordinate prediction in the complex indoor environment with multiple buildings and multiple floors.

2. In CL-Loc, the CNN branch extracts local features from WiFi fingerprints, while the LSTM branch models the relationships among APs. We then fuse the local fingerprint features with the global associations of the APs and apply a multi-layer fully connected network to generate the localization results.

3. We evaluate CL-Loc using the UJIIndoorLoc [8] and Tampere [9] datasets. The results demonstrate that CL-Loc outperforms current mainstream methods, achieving 100% building prediction accuracy on UJIIndoorLoc and 95.44% floor prediction accuracy on Tampere.

The remainder of the paper is organized as follows: Section II introduces related work, Section III outlines the overall structure and details of CL-Loc, Section IV evaluates our model, and Section V concludes the paper.

II. Related Work

In recent years, the growing demand for indoor localization, coupled with advancements in deep learning, has resulted in a significant number of studies focusing on this area. One notable approach is CNNLoc [7], which combines CNN and Stacked Autoencoders(SAE) to address issues related to sparse and similar location fingerprints. Similarly, DNNBN [6] has introduced a model that integrates Batch Normalization (BN) operations with a feedforward neural network. To tackle the high resource consumption associated with traditional CNN networks, CAE-CNNLoc [10] has developed a light indoor localization model designed for resource-constrained devices. Additionally, CRSS [11] has combined preprocessing techniques with CNN networks to create a light localization model suitable for edge devices. Indoor environments often contain a large number of APs, leading to a high dimensionality of location fingerprints that can negatively impact localization accuracy. To mitigate this issue, GNN [12] has implemented an AP selection strategy that establishes stable interaction relationships between adjacent APs, thereby effectively enhancing indoor localization accuracy. Furthermore, HADNN [13] has proposed utilizing hierarchical auxiliary information to improve the scalability of indoor localization models.

III. Methodology

A. Problem Formulation

In an indoor environment consisting of multiple buildings and multiple floors, we assume that there are M APs and N RPs. RP is defined as $RP = [b, f, x, y]$, where b, f represent the building and floor where the RP is located, and x, y represent the location coordinates of the RP. The RSS from AP_j measured at RP_i is rss_i^j, so the fingerprint vector of RP_i can be defined as $FP_i = [rss_i^1, \cdots, rss_i^M]$. The collected fingerprint database is represented as $\mathcal{FP}_{\mathcal{DB}} = \{FP_i, RP_i\}_{i=1}^N$, and the relationship between FP_i and RP_i is:

$$FP_i \xrightarrow{\mathcal{X}_\theta} RP_i \quad , \tag{1}$$

wherein, θ represents the parameters of the model $\mathcal{X}$. In summary, the goal of the problem is to find a θ^* that can best fit the mapping relationship.

B. Overall Structure

The overall structure of CL-Loc is illustrated in Figure 1. It extracts the local feature relationship from the location fingerprint in the CNN branch, extracts the association relationship among APs in the LSTM branch. These features, derived from different ranges, are then fused together. Finally, the indoor localization result is obtained through the multi-task output layer.

LSTM is a specialized type of recurrent neural network (RNN) that addresses the issue of long-term dependencies through the use of a gating mechanism. This enhancement has made LSTM particularly effective in various applications, such as speech recognition and industrial fault detection. As shown in Figure 2, each LSTM unit is composed of three gates, namely the forget gate, input gate, and output gate. The forget gate determines which information in the fingerprint data of the previous dimension needs to be retained and which needs to be forgotten. The calculation process is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad , \tag{2}$$

where σ represents the $sigmoid$ activation function, W_f represents the weight, h_{t-1} represents the hidden state of the previous LSTM unit, x_t represents the fingerprint of the current dimension, and b_f represents the bias. The input gate is used to decide how to retain the fingerprint data information of the current dimension. It is calculated through the following process:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad , \tag{3}$$

The output gate regulates the hidden state output of the current LSTM unit, which is calculated using the following formula:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad , \tag{4}$$

$$h_{t+1} = o_t \cdot tanh(c_t) \quad , \tag{5}$$

wherein σ represents the $sigmoid$ activation function, and c_t represents the cell state of the current LSTM unit, which is used for storing memory information. In the LSTM branch,

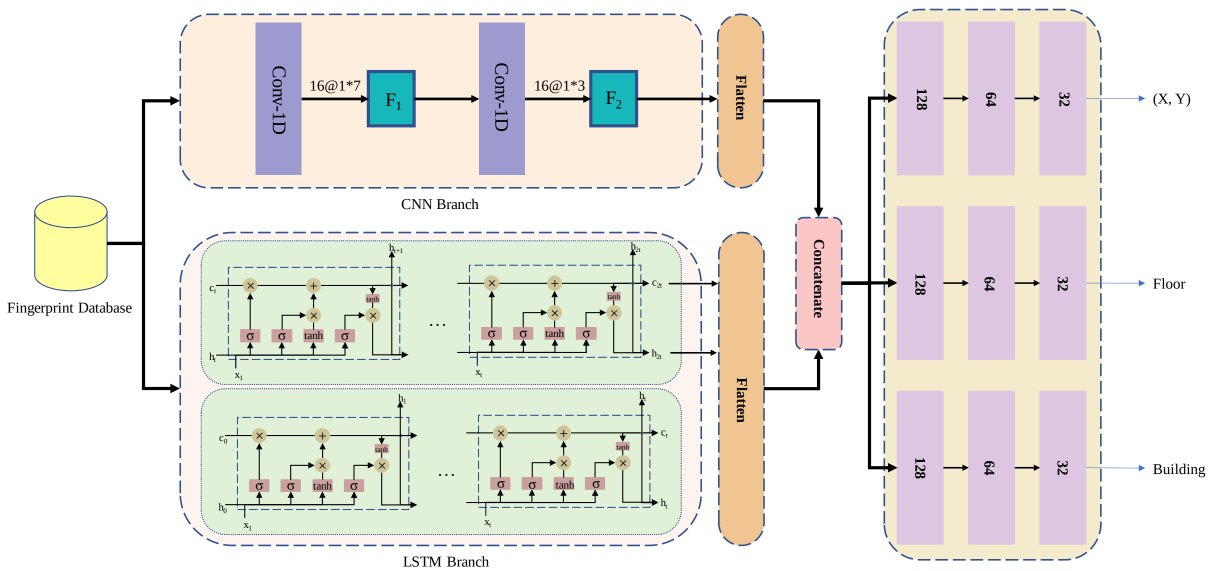

Fig. 1. The overall structure of the proposed indoor localization model, which consists of two branches. The CNN branch consists of two convolutional layers that extract the local features of the location fingerprint, resulting in an output denoted as F_2. Meanwhile, the LSTM branch captures the relationship among APs by two LSTM layers, producing outputs c_{2t}, h_{2t}. After the flattening operation, it is concatenated with F_2, and each output branch shares the result.

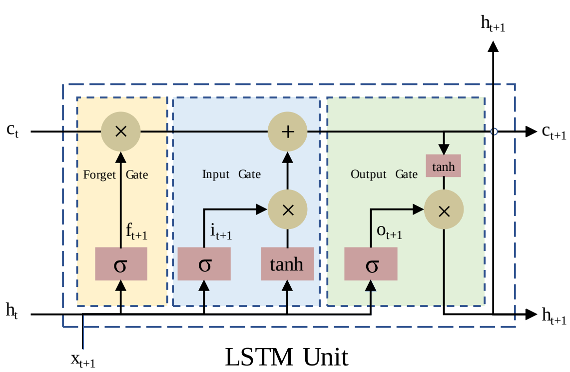

Fig. 2. The structure of the LSTM unit.

the number of LSTM units contained in each LSTM layer is determined by the number of APs. After passing through the LSTM layer, the model effectively models the relationship among APs implicit in the fingerprint.

Convolutional neural networks(CNN) have achieved significant success in various fields, including image recognition and natural language processing. They utilize convolutional kernels to perform convolution operations on sequential data, allowing for the extraction of local features. In our model, we employ two 1D convolution layers within the CNN branch to capture local features of different ranges in fingerprint data. We then flatten these local features to facilitate their integration with the AP relationships obtained from the LSTM branch. The calculation process is as follows:

$$F_1 = BN\Big(\sigma'\big(Conv_{16@1*7}(I)\big)\Big) \quad , \tag{6}$$

$$F_2 = BN\Big(\sigma'\big(Conv_{16@1*3}(F_1)\big)\Big) \quad , \tag{7}$$

wherein, σ' represents the $ReLU$ activation function, BN represents the Batch Normalization operation, I represents the input data of the CNN branch, $16@1*7$ represents 16 convolution kernels of size $1*7$, and $16@1*3$ represents 16 convolution kernels of size $1*3$. We connect the fingerprint features extracted by the two branches and then use three sets of output layers to make predictions for different indoor localization.

$$\mathcal{D} = Concatenate\big(F_2, O_{lstm}(h_{2t}, c_{2t})\big) \quad , \tag{8}$$

IV. Evaluation

In this section, we evaluate the CL-Loc experimentally and verify the effectiveness of each module through ablation experiments.

A. Evaluation Metrics

The evaluation metrics used in the experiment are accuracy and mean localization error (MLE). Accuracy is used to measure the performance of the model in predicting buildings and floors, and is calculated using the following formula:

$$Accuracy = \frac{S_{\text{correct}}}{S_{\text{total}}} \times 100\% \quad , \tag{9}$$

where S_{correct} represents the number of samples predicted correctly, and S_{total} represents the total number of samples.

MLE measures the average Euclidean distance between predicted and actual coordinates of the model, calculated using the following equation:

$$MLE = \frac{1}{S_{total}} \sum_{i=1}^{S_{total}} \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2} \quad , \qquad (10)$$

where (x_i, y_i) and $(\hat{x}_i, \hat{y}_i)$ represent the actual coordinates and predicted coordinates of the ith RP, respectively. Additionally, S represents the total number of test samples.

B. Datasets

TABLE I
THE SUMMARY OF THE DATASET.

Number of	UJIIndoorLoc	Tampere
Total Data	21048	4648
RP	933	4648
AP	520	992
Building	3	1
Floor	4 or 5	5
Year	2014	2017

We utilized two widely recognized open-source datasets in the area of indoor localization, UJIIndoorLoc [8] and Tampere [9], to evaluate CL-Loc. The UJIIndoorLoc dataset was collected from three buildings, while the Tampere dataset was collected in a single building. A summary of the dataset information is presented in Table I. Compared with Tampere, UJIIndoorLoc has one more indoor localization task, specifically building prediction. To adapt to the different localization tasks, three output branches and two output branches are used for UJIIndoorLoc and Tampere, respectively, in the output layer.

C. Settings

We utilized the Tensorflow framework to train and evaluate the model on the NVIDIA Tesla V100 GPU. The optimal model parameters are detailed in Table II. It is important to note that for different localization tasks, the overall loss of the model is the sum of the losses associated with each localization task.

TABLE II
THE SETTINGS.

Parameter	UJIIndoorLoc	Tampere
Learning Rate	0.03	0.01
Optimizer	Adam	Adam
Drop Out	0.3	0.3
Batch Size	128	128
Output Layers	128-64-32	128-64-32
Loss Function	MSE&CE	MSE&CE

D. Results on UJIIndoorLoc

We tested CL-Loc using the UJIIndoorLoc dataset and compared its performance with several mainstream indoor localization models, including CAE-CNNLoc [10], CNNLoc [7], DNNBN [6], and GNN [12]. As demonstrated in Table III, CL-Loc performs better than other models in predicting both buildings and location coordinates. This indicates that CL-Loc effectively extracts the correlation between local features and APs from location fingerprints. Although its performance in predicting floors is weaker than CNNLoc, CNNLoc's performance in predicting buildings and location coordinates is behind CL-Loc. In summary, CL-Loc achieves superior overall localization accuracy on the UJIIndoorLoc dataset compared to other mainstream models.

TABLE III
THE RESULTS ON UJIINDOORLOC. BEST RESULTS IN **BOLD**, SECOND BEST IN *Italic.*

Model	Floor Accuracy	Building Accuracy	MLE
CAE-CNNLoc	90.50%	*99.40%*	9.50m
CNNLoc	**96.03%**	99.27%	11.78m
DNNBN	93.97%	**100%**	*9.45*m
GNN	94.15%	**100%**	9.61m
Ours	*95.59%*	**100%**	**8.33m**

E. Results on Tempere

TABLE IV
THE RESULTS ON TAMPERE. BEST RESULTS IN **BOLD**, SECOND BEST IN *Italic.*

Model	Floor Accuracy	MLE
HADNN	*94.58%*	*9.07m*
CRSS	91.32%	-
CAE-CNNLoc	88.90%	10.24m
Ours	**95.44%**	**5.88m**

We also evaluated CL-Loc using Tampere and compared it with mainstream indoor localization models such as HADNN [13], CRSS [11], and CAE-CNNLoc [10] for comparison, as shown in Table IV. Since the building prediction is not required on Tempere, only two output branches are utilized in the output layer. Due to the small scale of Tempere, we selected a learning rate of 0.01 to prevent model overfitting. The results presented in Table IV indicate that CL-Loc outperforms other mainstream localization models in both floor prediction and location coordinate prediction. This demonstrates that the CNN-LSTM dual-branch structure proposed in this paper can not only effectively extract local features from location fingerprints but also model the relationship among APs from a global perspective.

F. Ablation Experiment

In order to verify the effectiveness of the two branch modules, an ablation experiment was performed on the UJIIndoorLoc dataset, and the experimental results are listed in Table V.

The results presented in Table V indicate that omitting the CNN branch while retaining only the LSTM branch leads to a decrease of 1.8% in floor prediction accuracy and a reduction of 4.44 meters in MLE. This suggests that without the CNN

TABLE V
THE RESULTS OF THE ABLATION EXPERIMENT. BEST RESULTS HIGHLIGHTED IN **BOLD**.

Methods	Floor Accuracy	Building Accuracy	MLE
w/o CNN-Branch	93.79%	**100%**	12.77m
w/o LSTM-Branch	89.92%	99.64%	13.75m
Ours	**95.59%**	**100%**	**8.33m**

branch to capture local features of the location fingerprint, and relying solely on the LSTM branch to extract relationships among APs, CL-Loc is unable to effectively differentiate fine-grained information such as floor levels and specific location coordinates. Conversely, when only the CNN branch is used while the LSTM branch is excluded, floor prediction accuracy decreases by 5.67%, and MLE increases by 5.42 meters. This outcome illustrates that extracting only local features from the fingerprint, without considering the relationships between APs, hampers the model's ability to refine the extraction process based on those relationships, ultimately diminishing localization accuracy. In summary, the findings reveal that the LSTM branch's role in modeling AP relationships is more critical than the CNN branch's function in extracting local fingerprint features. Together, these two components complement one another, enabling CL-Loc to achieve higher accuracy in indoor localization.

V. CONCLUSION

This paper proposes an indoor localization network named CL-Loc, which utilizes a dual-branch structure of CNN and LSTM networks. CL-Loc effectively extracts local features from location fingerprints and models the relationships among APs. Experimental results from the UJIIndoorLoc dataset demonstrate that CL-Loc achieves 100% accuracy in building prediction and 95.59% accuracy in floor prediction. Similarly, results from the Tempere dataset indicate a floor prediction accuracy of 95.44%. These findings show CL-Loc's high indoor localization accuracy and its adaptability to various indoor localization tasks. This method offers a new perspective for research in indoor localization.

VI. FUTURE WORK

In future work, we plan to incorporate model compression techniques such as knowledge distillation and quantization-aware training to adapt the model for resource-constrained IoT devices. Additionally, we aim to integrate heterogeneous signals such as magnetic fields and Bluetooth to construct multimodal inputs for improved positioning accuracy in complex environments. Furthermore, we will explore incremental learning strategies to enhance training efficiency when updating the fingerprint database and reduce system maintenance overhead.

REFERENCES

[1] M. Zhang, J. Jia, J. Chen, L. Yang, L. Guo, and X. Wang, "Real-time indoor localization using smartphone magnetic with lstm networks," *Neural Computing and Applications*, vol. 33, pp. 10 093–10 110, 2021.

[2] X. Zhang, Y. Zhang, G. Liu, and T. Jiang, "Autoloc: Toward ubiquitous aoa-based indoor localization using commodity wifi," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 6, pp. 8049–8060, 2023.

[3] X. Yang, Y. Zhuang, F. Gu, M. Shi, X. Cao, Y. Li, B. Zhou, and L. Chen, "Deepwipos: A deep learning-based wireless positioning framework to address fingerprint instability," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 6, pp. 8018–8034, 2023.

[4] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Reddy, "Recurrent neural networks for accurate rssi indoor localization," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 639–10 651, 2019.

[5] A. Mansour, J. Ye, Y. Li, H. Luo, J. Wang, D. Weng, and W. Chen, "Everywhere: A framework for ubiquitous indoor localization," *IEEE Internet of Things Journal*, vol. 10, no. 6, pp. 5095–5113, 2023.

[6] S. Lukman Ayinla, A. A. Aziz, M. Drieberg, M. Susanto, A. Tumian, and M. Yahya, "An enhanced deep neural network approach for wifi fingerprinting-based multi-floor indoor localization," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 560–575, 2025.

[7] X. Song, X. Fan, X. He, C. Xiang, Q. Ye, X. Huang, G. Fang, L. L. Chen, J. Qin, and Z. Wang, "Cnnloc: Deep-learning based indoor localization with wifi fingerprinting," in *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 2019, pp. 589–595.

[8] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems," in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2014, pp. 261–270.

[9] E. S. Lohan, J. Torres-Sospedra, H. Leppäkoski, P. Richter, Z. Peng, and J. Huerta, "Wi-fi crowdsourced fingerprinting dataset for indoor positioning," *Data*, vol. 2, no. 4, p. 32, 2017.

[10] A. Kargar-Barzi, E. Farahmand, N. Taheri Chatrudi, A. Mahani, and M. Shafique, "An edge-based wifi fingerprinting indoor localization using convolutional neural network and convolutional auto-encoder," *IEEE Access*, vol. 12, pp. 85 050–85 060, 2024.

[11] I. Neupane, S. Shahrestani, and C. Ruan, "Indoor localization of resource-constrained iot devices using wi-fi fingerprinting and convolutional neural network," in *Proceedings of the 2024 Australasian Computer Science Week*, ser. ACSW '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 20–25.

[12] S. Wang, S. Zhang, J. Ma, and O. A. Dobre, "Graph-neural-network-based wifi indoor localization system with access point selection," *IEEE Internet of Things Journal*, vol. 11, no. 20, pp. 33 550–33 564, 2024.

[13] J. Cha and E. Lim, "A hierarchical auxiliary deep neural network architecture for large-scale indoor localization based on wi-fi fingerprinting," *Applied Soft Computing*, vol. 120, p. 108624, 2022.