Encoder-Decoder Driven Adaptive Multiscale-CNN Based Indoor Localization with WiFi Fingerprinting

No Author Given

No Institute Given

Abstract. With the growing demand for location-based services (LBS) from personal mobile devices, indoor localization is facing increasingly strict requirements in terms of accuracy, scalability, and robustness. Current methods don't fully consider sparse location fingerprints caused by reasons such as wall obstructions, and the high similarity of location fingerprints caused by overlapping WiFi access point coverage, resulting in reduced indoor localization accuracy. Therefore, we propose an **E**ncoder-**D**ecoder driven **A**daptive **M**ultiscale-CNN based indoor **Loc**alization model (**EDAMLoc**) to solve the above issues, which is designed for multi-building and multi-floor environments. It utilizes feature enhancement to process the features available in sparse location fingerprints and adaptively fuses cross-scale location fingerprint features to better distinguish similar location fingerprints. Furthermore, EDAMLoc can flexibly adapt its output branches based on the requirements of different localization tasks. Experimental results show that EDAMLoc achieves 100% Building Accuracy and 95.30% Floor Accuracy on the UJIIndoorLoc and Tempere datasets, respectively, and that its localization accuracy is better than that of mainstream indoor localization models.

Keywords: Indoor Localization · WiFi Fingerprinting · Encoder-Decoder · Multiscale Convolution Neural Network

1 Introduction

In recent years, location-based services, such as indoor navigation and localized recommendations, have gained increasing attention[5,15]. In outdoor environments, accurate location information can be obtained through satellite-based global localization systems, including GPS, Beidou, and GLONASS. However, when it comes to indoor environments, the accuracy of satellite-based localization diminishes due to obstructions like walls and the movement of pedestrians. Wireless technologies such as WiFi, Bluetooth, and Zigbee provide more reliable network access indoors to address this problem[13]. Among these, WiFi has emerged as a leading solution for indoor localization, primarily because of its widespread availability and excellent compatibility with various devices.

Among the various WiFi-based indoor localization methods, two primary approaches stand out: geometry-based methods and WiFi location fingerprint-based methods. Geometry-based localization methods estimate the location by

calculating the Time of Arrival (TOA), Angle of Arrival (AOA), etc[3,21]. However, measuring this information requires professional equipment, which limits its versatility and practical application, making it challenging to implement on a large scale[20]. The WiFi location fingerprint-based method constructs a location fingerprint using the measured Channel State Information (CSI) or Received Signal Strength (RSS). The acquisition of CSI depends on accessing the In-phase and Quadrature (IQ) samples at the physical layer, which imposes significant hardware demands on the device[9]. In contrast, RSS is easier to obtain. It constructs fingerprint vectors for different locations by measuring the signal strength of the WiFi Access Point (WAP). This method requires only the WAP and is more versatile than geometry-based methods[6]. In the indoor localization method based on RSS location fingerprint, the location fingerprint database is constructed by collecting the RSS from WAPs at various measurement locations (i.e., reference point, RP). During the localization stage, the collected RSS vector is matched with the fingerprint data in the pre-built location fingerprint database to achieve indoor localization.

Due to the multipath effect that occurs during wireless signal propagation and various factors such as crowd movement, the measured RSS is time-varying. This variability impacts the accuracy of indoor localization methods that are based on RSS location fingerprints. Currently, some studies use UAVs, crowdsourcing to collect RSS fingerprints, aiming to collect more accurate fingerprints as possible[14,23]. This paper selects the UJIIndoorLoc[16] dataset and the Tampere[10] dataset, both of which are commonly used in numerous studies, to compare the results of different methods.

In recent years, the development of deep learning technology has led to the emergence of numerous indoor localization methods that utilize deep learning[4]. The powerful feature capture and generalization ability of deep neural networks has proven to be a significant advantage in the realm of indoor localization. In the multiple-building and multiple-floor indoor environment, the RSS location fingerprints collected by the RP often have a large number of WAP signal missing problems. This means that the RSS from all WAPs cannot be fully measured, resulting in sparse RSS location fingerprints, which affect the accuracy of localization. Additionally, the coverage areas of WAPs often overlap, which means that adjacent or similar locations, such as coordinates on different floors, can produce similar location fingerprints. This overlap can make it particularly challenging for the model to effectively distinguish between these location fingerprints, especially in small-scale receptive fields. To address these issues, we propose an **E**ncoder-**D**ecoder driven **A**daptive **M**ultiscale-CNN Based Indoor **Loc**alization Model (**EDAMLoc**), which achieves high precision and robust indoor localization in indoor environments. The contributions of our work are as follows:

1. We propose a novel localization model based on RSS location fingerprints to effectively predict localization information such as buildings, floors, and location coordinates in indoor environments. This model leverages an encoder-

decoder driven multi-scale adaptive fusion network to achieve high-precision and robust indoor localization.

2. We employ an encoder-decoder structure, which effectively removes irrelevant information in sparse RSS location fingerprints while enhancing valuable features. Additionally, we create an adaptive multi-scale feature fusion module that captures both local features and global feature interaction information. It selectively fuses cross-scale features, significantly improving the model's ability to differentiate similar location fingerprints.

3. We evaluated our model using the UJIIndoorLoc and Tempere datasets. The results indicate that our model surpasses existing techniques, particularly on the UJIIndoorLoc dataset, achieving 100% accuracy in building predictions and 95.14% accuracy in floor predictions.

The remainder of the paper is organized as follows: Section 2 introduces related work, Section 3 outlines the overall structure and details of EDAMLoc, Section 4 evaluates our model, and Section 5 concludes the paper.

2 Related Work

Current WiFi fingerprint-based localization methods are primarily divided into CSI location fingerprint-based methods and RSS location fingerprint-based methods. This paper focuses on the latter.

In indoor environments, factors like crowd movement can impact the propagation of WiFi signals, leading to instability in the RSS measured at the RPs. To address this issue, GNN[17] proposed a location mapping network that utilizes Graph Neural Network (GNN) to establish stable feature relationships between adjacent access points, and combined it with an access point selection strategy. To tackle the issue of performance fluctuations in localization models, which arise from time-varying and sparse location fingerprints, SALLoc[1] proposed a model that integrates the Stacked Auto-Encoder (SAE) network, attention mechanism, and LSTM.CNNLoc[15] improved the localization accuracy when the dataset size is small by combining a 1D-CNN and SAE. DNNBN[11] introduced a method that integrates batch normalization (BN) into each layer of deep neural networks (DNNs) to improve model performance stability. Considering different localization tasks in different environments, HADNN[2] proposed a method that combines hierarchical auxiliary information to improve the scalability of the localization model. Many of the methods mentioned above are designed for devices with adequate computing resources, while on edge devices with limited computing resources, lightweight models are required. CRSS[12] focuses on achieving effective indoor localization on edge devices and proposes a lightweight CNN model integrated with preprocessing techniques. CAE-CNNLoc[8] combines preprocessing techniques with a convolutional autoencoder structure to effectively reduce computational time and complexity.

3 Methodology

In this section, we introduce the overall model structure and design details.

3.1 Problem Formulation

In indoor environments, we assume that there are N WAPs and M RPs. Now define the set of WAP as $WAPs = \{wap_1, wap_2, \cdots, wap_N\}$, where $i \in [1, N]$ represents the WAP numbered i. RP is defined as $RP_j = [x_j, y_j, f_j, b_j]^T, j \in [1, M]$, where x_j, y_j represent the coordinates of RP, f_j, b_j represent the floor and building where RP is located, respectively. The location fingerprint vector for each RP consists of the RSS from each WAP that it measures, that is,

$$fp_j = [RSS_1^j, RSS_2^j, \cdots, RSS_n^j, \cdots, RSS_N^j]^T, j \in [1, M], n \in [1, N], \quad (1)$$

where RSS_n^j represents the RSS of wap_n measured at the reference point RP_j.

It is important to note that RP may not be able to measure the RSS of all WAPs. We set the RSS of such WAPs to -110 dBm. Thus, we construct the location fingerprint database $\mathcal{FD} = \{fp_j, RP_j\}_{j=1}^M$. The mapping relationship between location fingerprint and location in $\mathcal{FD}$ is

$$\chi_\theta : fp_j \longrightarrow RP_j, j \in [1, M], \quad (2)$$

where θ represents the model parameters.

Assume we have the RSS vector denoted as $\hat{fp}$, with the target location $L = [x, y, f, b]$. And the predicted location $\hat{L} = [\hat{x}, \hat{y}, \hat{f}, \hat{b}]$ is derived through the model. Finally, the optimization objective of the model can be obtained from the optimization theory:

$$\theta^* = arg\min_\theta \big(\alpha \cdot \mathbb{MSE}(x, y, \hat{x}, \hat{y}) + \beta \cdot \mathbb{CE}(f, \hat{f}) + \gamma \cdot \mathbb{CE}(b, \hat{b})\big), \quad (3)$$

where α, β, γ represent the loss weight of the three tasks, respectively. MSE is the Mean Squared Error loss function, which minimizes the Euclidean distance between the predicted location coordinates and the target location coordinates. CE is the CrossEntropy loss function, which minimizes the difference between the predicted floor, building and the target floor, building.

3.2 Model Structure

Fig. 1 illustrates the overall structure of the EDAMLoc. The collected fingerprint database is input into the model, and the sparse fingerprint is processed by the **RS**S **S**parse **F**eature **E**nhancement Module(**RSS-FE**), and the fingerprint features of different scales are processed by the **R**SS **A**daptive **M**ulti-scale **F**eature **C**apture Module(**RAM-FC**). Ultimately, yields a model suitable for indoor localization.

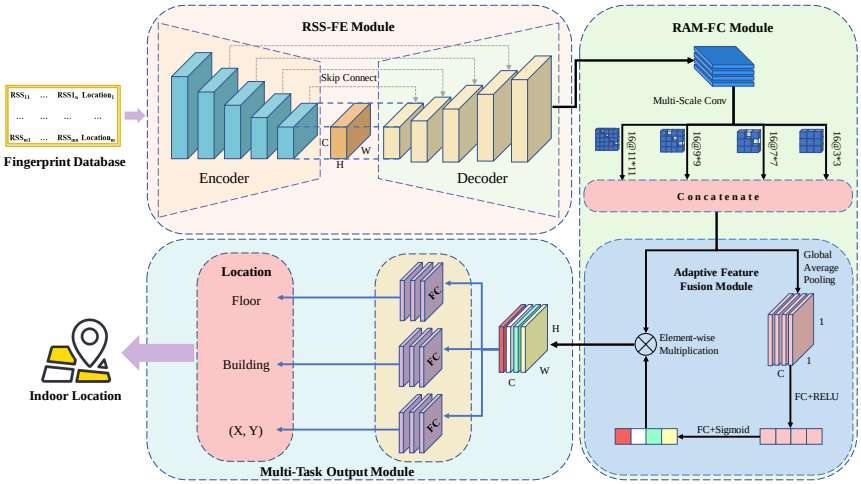


Fig. 1: The structure of the proposed indoor localization model EDAMLoc, which consists of four key modules: RSS-FE Module for enhancing available features from sparse RSS fingerprints, RAM-FC Module for adaptively capturing cross-scale features, within AFF Module for fusing cross-scale features through adaptive weighting, and a Multi-task Output module for flexibly predicting coordinates, floors, and buildings based on task requirements.

3.3 RSS-FE Module

In indoor environments, factors such as walls and crowd movement can hinder the measurement of the RSS from a large number of WAPs. This results in a significant amount of sparse location fingerprints in the location fingerprint database. To achieve accurate indoor localization, it is essential to remove unavailable information from the sparse fingerprints and enhance the features of the available data. This allows the model to focus more on the important details within the sparse location fingerprints[7]. Fortunately, we discovered that we can customize the encoder-decoder structure flexibly to meet varying task requirements, allowing us to remove unavailable information and enhance available features, as the RSS-FE module in Fig. 1.

The encoder-decoder structure has been widely used in various fields, including natural language processing, computer vision, and time series processing in recent years [4,18]. The encoder we designed contains convolutional layers that use kernels of size 3×3. Smaller kernels can extract more local features. The number of kernels in these convolutional layers decreases progressively, which can compress the features of the fingerprints into a smaller dimension. This structure not only removes a large amount of unavailable information in the location fingerprint, but also compresses and retains important feature information in

the location fingerprint through multiple layers of convolution. Its process can be described as follows:

$$\mathbf{Z}^{(l)} = \sigma\left(\mathrm{Conv}_{3\times3}^{(l)}(\mathbf{Z}^{(l-1)})\right), \quad l = 1, 2, \ldots, L, \tag{4}$$

where, $Z^{(0)} = X \in \mathbb{R}^{H\times W\times C}$ represents the input data, $\sigma(\cdot)$ represents the ReLU activation function, and L represents the number of encoder layers. The intermediate features obtained by the encoder are expressed as:

$$\mathbf{F}_{enc} = \mathbf{Z}^{(L)}, \tag{5}$$

Unlike the encoder, the decoder focuses on maximizing the use of the intermediate feature information generated by the encoder. A larger kernel has a larger receptive field and can capture a wider range of feature interaction information, which can reconstruct the feature dimensions more effectively. Thus, we utilize 5×5 kernels in the convolution layers of the decoder to reshape the intermediate feature dimensions back to those of the input data. The process is as follows:

$$\mathbf{Z}'^{(l)} = \sigma\left(\mathrm{Conv}_{5\times5}^{(l)}(\mathbf{Z}'^{(l-1)})\right), \quad l = 1, 2, \ldots, L, \tag{6}$$

where, $\mathbf{Z}'^{(0)} = \mathbf{F}_{enc}$. The output of the decoder is:

$$\mathbf{F_{dec}} = \mathbf{Z}'^{(L)}. \tag{7}$$

3.4 RAM-FC Module

Another issue in indoor localization is that nearby or similar locations, such as the same coordinate location on adjacent floors, may produce approximate local RSS location fingerprint features. Under a fixed receptive field, traditional models are difficult to distinguish them effectively. This is because the key feature information in the fingerprint is distributed in feature spaces of different ranges. Traditional CNN methods, such as CNNLoc[15], use fixed-size convolutional kernels. This design causes the model to only focus on local features of a certain range. To address the need for extracting features at different scales, additional layers are typically required, which increases both the number of model parameters and computational complexity. In order to capture cross-scale features without significantly increasing the amount of computation, we introduce a multi-scale convolutional neural network (Multi-Scale CNN), as the RAM-FC module shown in Fig. 1.

The Multi-Scale CNN uses multiple kernels of different sizes to capture the location fingerprint features in parallel in multiple convolution paths[22]. The multi-scale convolution leverages small kernels for sensitivity to small-scale features and large kernels for capturing large-scale features, effectively distinguishing approximate RSS location fingerprints. We define D kernels of different sizes as $K = [k_1, k_2, \cdots, k_D]$. The calculation process is as follows:

$$\mathbf{F}_{n\times n} = f\big(\mathrm{Conv}_{n\times n}(\mathbf{F_{dec}})\big), \quad n \in K. \tag{8}$$

The output features of each convolutional path are concatenated to obtain T_{cat}, and the process is as follows:

$$T_{cat} = BN(\mathbf{F}_{k_1 \times k_1}) \oplus BN(\mathbf{F}_{k_2 \times k_2}) \oplus \cdots \oplus BN(\mathbf{F}_{k_D \times k_D}), \tag{9}$$

where $\oplus$ represents the concatenate operation, $BN(\cdot)$ represents batch normalization.

Besides, we discovered that the impact of each WAP on the localization result is different; that is, the available feature information contained in different feature channels is not balanced. Therefore, it is necessary to distinguish the importance of different feature channels. The attention mechanism, designed to imitate human attention behavior, is widely used in various fields, including computer vision and NLP[19]. We utilize the attention mechanism to redistribute the weights of different channels, so that the model focuses on more important feature channels, thereby improving the model's performance. The **A**daptive **F**eature **F**usion Module (**AFF**) in Fig. 1 illustrates the attention module we used. It includes a global average pooling layer (GAP) that compresses input feature channels into scalar values, followed by two fully connected layers that reallocate the importance of each channel. Finally, these weights are multiplied by their corresponding channels to obtain a weighted feature map, which can effectively distinguish the importance of different WAPs. The process of AFF Module is expressed as:

$$\Psi = \sigma'(FC(\sigma(FC(GAP(T_{cat}))))) \otimes T_{cat}, \tag{10}$$

wherein, $FC(\cdot)$ represents the fully connected layer, $GAP(\cdot)$ represents the global average pooling, and $\sigma'(\cdot)$ represents the sigmoid activation function.

3.5 Output Module

We specify different output branches for various indoor localization tasks, all of which utilize the same location fingerprint feature. In the UJIIndoorLoc dataset [16], we predict coordinates, floors, and buildings, which necessitates three output branches. In contrast, the Tampere dataset [10] only requires predictions for coordinates and floors, so we use two output branches. By flexibly adjusting the output branch structure for various tasks, the model's scalability is effectively demonstrated, as shown in the Output Module in Fig. 1.

4 Evaluation and Analysis

In this section, we evaluate our model and analyze the results of the experiments.

4.1 Datasets

The details of the dataset we utilized are presented in Table 1. UJIIndoorLoc dataset [16] is widely used in indoor localization research, consisting of training

Table 1: The statistics of the dataset.

Number of	UJIIndoorLoc	Tampere
Training Data	19937	697
Validation Data	1111	3951
WAPs	520	992
Buildings	3	1
Floors	4, 5	5

data and validation data. It was collected from three teaching buildings in the university and contains RSS data from 520 WAPs, totaling 21,048 records. Each record contains the RSS collected by the RP, the coordinates of the RP, and the building information and floor information. Fig. 2 illustrates the spatial distribution of the fingerprint for both the training and validation data.

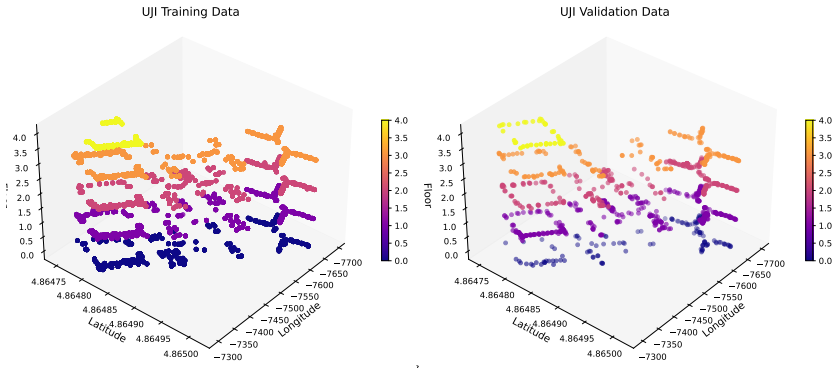


Fig. 2: Spatial distribution of fingerprint in UJIIndoorLoc.

The Tampere dataset[10] is collected from a building and contains RSS data from 992 WAPs. It is also divided into training set and validation set, which includes only the RSS of RP, the coordinates of RP, and the floor information. Fig. 3 shows the spatial distribution of fingerprint in the training data and the validation data.

4.2 Evaluation Metrics

The evaluation Metrics we use include accuracy (Acc) and mean localization error (MLE). Acc measures the accuracy of floor and building predictions, calculated using Eq. 11.

$$\text{Acc} = \frac{N_{\text{correct}}}{N_{\text{total}}} \times 100\%, \tag{11}$$

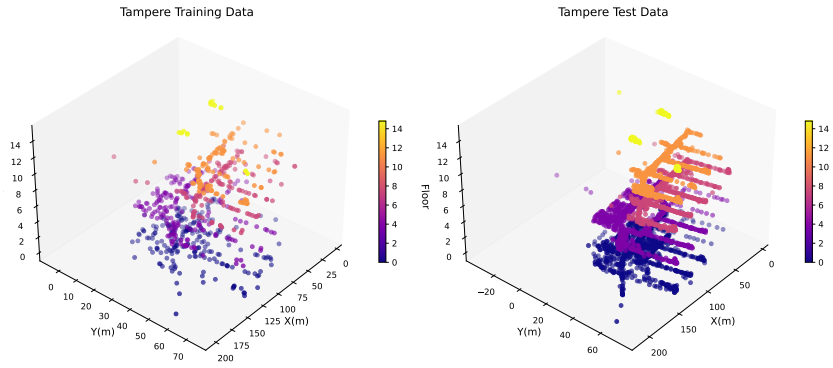


Fig. 3: Spatial distribution of fingerprint in Tampere.

where N_{correct} represents the number of samples for which the model correctly predicts the floor or building, and N_{total} represents the total number of test samples.

MLE measures the average Euclidean distance between predicted and actual coordinates of the model, calculated using Eq. 12:

$$\text{MLE} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2}, \tag{12}$$

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

4.3 Settings

Table 2: The optimal hyperparameter settings.

Parameter	UJIIndoorLoc	Tampere
Learning Rate	0.03	0.01
Optimizer	SGD	SGD
Drop Out	0.3	0.3
Batch Size	128	128
Loss Weight	1/3, 1/3, 1/3	1/3, 1/3, 0

We use the TensorFlow 2.4.0 framework to train and test our model on an NVIDIA GeForce RTX 4090 with 24GB of GPU memory. For different localization tasks, using UJIIndoorLoc and Tampere datasets, and different output

branches were utilized in the output layer to adapt to different prediction tasks. The model was trained for 100 epochs, each output branch consists of three fully connected layers. The number of nodes in the fully connected layers was 128, 64, and 32, respectively. They share the location fingerprint features captured by the RAM-FC Module. The optimal hyperparameters shown in Table 2.

4.4 Results and Analysis

Results on UJIIndoorLoc We first evaluate EDAMLoc on UJIIndoorLoc dataset. The results of the comparison with representative models including CNNLoc[15], HADNN[2], GNN[17], DNNBN[11], etc. are shown in Table 3. Our model achieves 100% accuracy in building prediction, 95.13% accuracy in floor prediction, and the MLE of coordinate prediction is 9.31m. Compared with existing methods, our model achieves the best results in both building prediction accuracy and MLE. Although the accuracy of floor prediction is slightly lower than CNNLoc's 96.03%, CNNLoc underperforms compared to our model in building prediction accuracy and MLE because our model effectively deals with the adverse effects of sparse fingerprints and similar fingerprints. In summary, our model demonstrates higher indoor localization accuracy on UJIIndoorLoc.

Table 3: The results on UJIIndoorLoc. Best Results in **Bold**, Second Best in *Italic.*

Model	Floor Acc(%)	Building Acc(%)	MLE(m)
CNNLoc[15]	**96.03**	*99.27*	11.78
HADNN[2]	93.15	**100**	11.59
GNN[17]	94.15	**100**	9.61
DNNBN[11]	93.97	**100**	*9.45*
EDAMLoc	*95.14*	**100**	**9.31**

Results on Tampere The same evaluation settings as UJIIndoorLoc are used on the Tampere dataset. Tampere includes only two tasks: floor prediction and coordinate location prediction. The output layer consists of two branches; therefore, the loss weight for building prediction in the optimization objective is set to 0. Since the Tampere dataset has fewer training samples, a smaller learning rate (0.01) is used to prevent the model from overfitting.

On Tampere, EDAMLoc is compared with mainstream models, including HADNN[2], SALLoc[1], CRSS[12], and CAE-CNNLoc[8], as presented in Table 4. Our model achieved an accuracy of 95.30% in floor prediction and 7.40 m in MLE, which fully demonstrated the generalization ability of EDAMLoc.

In summary, EDAMLoc achieves excellent performance in different datasets and localization tasks, which fully demonstrates that EDAMLoc has higher indoor localization accuracy.

Table 4: The results on Tampere. Best Results in **Bold**, Second Best in *Italic.*

Model	Floor Acc(%)	MLE(m)
HADNN[2]	*94.58*	*9.07*
SALLoc[1]	-	9.52
CRSS[12]	91.32	-
CAE-CNNLoc[8]	88.90	10.24
EDAMLoc	**95.30**	**7.40**

4.5 Ablation Experimental

To further evaluate the effectiveness of each module, we conduct ablation experiments on UJIIndoorLoc. The experimental results are presented in Table 5.

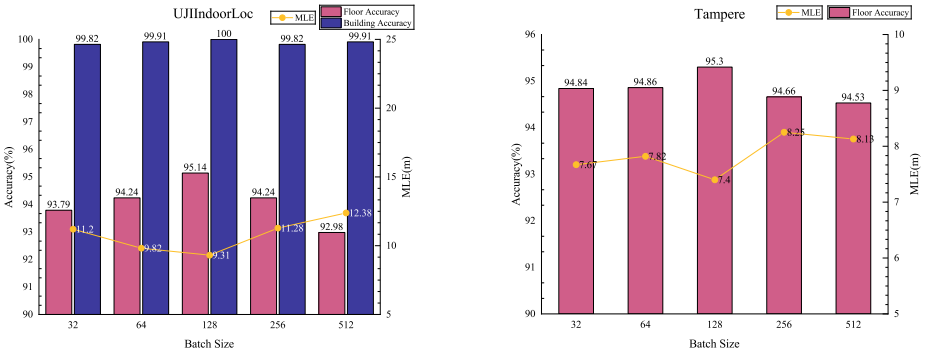
Table 5: The results of the ablation Study. Best Results Highlighted in **Bold**.

Methods	Floor Acc(%)	Building Acc(%)	MLE(m)
w/o RSS-FE	94.05	100	11.78
w/o RAM-FC	91.54	99.82	14.73
w/o AFF	93.52	100	11.58
EDAMLoc	**95.14**	**100**	**9.31**

Experimental results indicate that when the model doesn't include the RSS-FE module to enhance the features of sparse location fingerprints, the accuracy of floor prediction decreases by 1.09% and the MLE increases by 2.47 m. This is primarily because a significant amount of useless information in the sparse location fingerprint has not been effectively removed, thus affecting the quality of subsequent feature capture. When the RAM-FC module is removed from the model, that is, multi-scale feature capture is not performed, the accuracy of building and floor prediction decreases by 3.6% and 0.18%, respectively, and the MLE increases by 5.42 m. This module has the greatest impact on the model performance, indicating that our proposed RAM-FC can effectively distinguish similar fingerprints. Furthermore, when the AFF module is removed, the model loses its ability to fuse multi-scale features adaptively. As a result, the accuracy of floor prediction decreases by 1.62%, and the MLE increases by 2.27 m. This indicates the importance of adaptive cross-scale feature fusion, as its absence reduces the model's attention to critical features.

4.6 Parameter Analysis

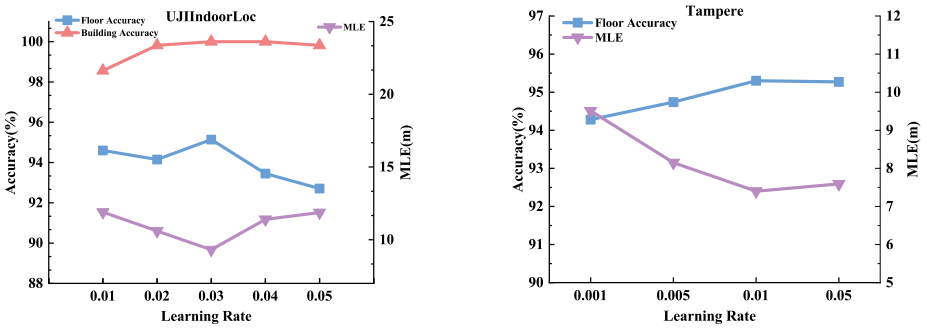
We also conduct sensitivity analyses of the model regarding batch size, learning rate, and the combination of kernels.



(a) (b)

Fig. 4: Effect of various batch size.

Fig. 4 illustrates the effect of various batch sizes on model performance. As shown in Fig. 4(a), on UJIIndoorLoc, the building prediction accuracy and floor prediction accuracy of the model first increase and then decrease with larger batch sizes. In contrast, the MLE shows a pattern of first decreasing and then increasing. The model achieves its best performance with a batch size of 128, where the building prediction accuracy achieves 100%, the floor prediction accuracy is 95.14%, and the MLE is 9.31 meters. At this time, the building prediction accuracy is 100%, the floor prediction accuracy is 95.14%, and the MLE is 9.31m. In the Tampere dataset, as shown in Fig. 4(b), the best performance is also achieved when the batch size is set to 128, with a floor accuracy of 95.30% and an MLE of 7.40 m. This result indicates that a batch size is 128 yields the highest localization accuracy for the model, and EDAMLoc is not sensitive to batch size.



(a) (b)

Fig. 5: Effect of various learning rate.

Fig. 5 shows the effect of different learning rates on model performance. The results show that the model achieves the best performance when the learning rate is 0.03 on the UJIIndoorLoc dataset, and the model achieves the best performance when the learning rate is 0.01 on the Tampere dataset. Overall, a learning rate that is too low can lead to slower convergence of the model, whereas a learning rate that is too high may result in gradient oscillation.

Table 6 indicates the effect of various combinations of multi-scale convolution kernels on model performance. The results show that the model utilizing the convolution kernel combination of {**3,7,9,11**} achieves the best results on both datasets and can effectively capture the cross-scale feature of location fingerprint.

Table 6: Effect of various combinations of Kernels.

Kernel Size	UJIIndoorLoc			Tampere	
	Floor Acc(%)	Building Acc(%)	MLE(m)	Floor Acc(%)	MLE(m)
$3, 5, 7, 9$	93.34	99.91	10.42	94.81	8.11
$3, 7, 9, 11$	**95.14**	**100**	**9.31**	**95.30**	**7.40**
$5, 7, 9, 11$	92.44	99.82	10.29	94.89	7.92
$7, 9, 11, 13$	92.44	99.91	10.29	94.79	7.84

5 Conclusion

This paper proposes a scalable indoor localization model based on WiFi location fingerprints called EDAMLoc. EDAMLoc can effectively process sparse location fingerprints and adapts to fuse cross-scale features, achieving high-precision indoor localization. Experiments are conducted using two public datasets, UJIIndoorLoc and Tempere, to evaluate the effectiveness of the model. On UJIIndoorLoc, a floor prediction accuracy of 95.14%, a building prediction accuracy of 100% and an MLE of 9.31m were achieved. On Tempere, the model achieved a floor prediction accuracy of 95.30% and an MLE of 7.40 m. Experimental results demonstrate that EDAMLoc can adapt to various indoor localization tasks while maintaining high localization accuracy and demonstrating strong generalization ability. This offers a novel idea and technical pathway for indoor localization research based on WiFi location fingerprinting.

References

1. Ayinla, S.L., Aziz, A.A., Drieberg, M.: Salloc: An accurate target localization in wifi-enabled indoor environments via sae-alstm. IEEE Access **12**, 19694–19710 (2024). https://doi.org/10.1109/ACCESS.2024.3360228

2. Cha, J., Lim, E.: A hierarchical auxiliary deep neural network architecture for large-scale indoor localization based on wi-fi fingerprinting. Applied Soft Computing **120**, 108624 (2022). https://doi.org/10.1016/J.ASOC.2022.108624
3. Geng, C., Abrudan, T.E., Kolmonen, V.M., Huang, H.: Experimental study on probabilistic toa and aoa joint localization in real indoor environments. In: ICC 2021 - IEEE International Conference on Communications. pp. 1–6 (2021). https://doi.org/10.1109/ICC42927.2021.9500283
4. Gufran, D., Tiku, S., Pasricha, S.: Vital: Vision transformer neural networks for accurate smartphone heterogeneity resilient indoor localization. In: 2023 60th ACM/IEEE Design Automation Conference (DAC). pp. 1–6 (2023). https://doi.org/10.1109/DAC56929.2023.10247684
5. Han, J., Susilo, W., Li, N., Huang, X.: Olbs: Oblivious location-based services. IEEE Transactions on Information Forensics and Security **19**, 2231–2243 (2024). https://doi.org/10.1109/TIFS.2023.3347874
6. Jia, B., Liu, J., Feng, T., Huang, B., Baker, T., Tawfik, H.: Ttsl: An indoor localization method based on temporal convolutional network using time-series rssi. Computer communications **193**, 293–301 (2022). https://doi.org/10.1016/J.COMCOM.2022.07.003
7. Jia, B., Qiao, W., Zong, Z., Liu, S., Hijji, M., Del Ser, J., Muhammad, K.: A fingerprint-based localization algorithm based on lstm and data expansion method for sparse samples. Future Generation Computer Systems **137**, 380–393 (2022). https://doi.org/10.1016/J.FUTURE.2022.07.021
8. Kargar-Barzi, A., Farahmand, E., Taheri Chatrudi, N., Mahani, A., Shafique, M.: An edge-based wifi fingerprinting indoor localization using convolutional neural network and convolutional auto-encoder. IEEE Access **12**, 85050–85060 (2024). https://doi.org/10.1109/ACCESS.2024.3412676
9. Kong, R., Chen, H.: Deepcrf: Deep learning-enhanced csi-based rf fingerprinting for channel-resilient wifi device identification. IEEE Transactions on Information Forensics and Security **20**, 264–278 (2025). https://doi.org/10.1109/TIFS.2024.3515796
10. Lohan, E.S., Torres-Sospedra, J., Leppäkoski, H., Richter, P., Peng, Z., Huerta, J.: Wi-fi crowdsourced fingerprinting dataset for indoor positioning. Data **2**(4), 32 (2017). https://doi.org/10.3390/DATA2040032
11. Lukman Ayinla, S., Aziz, A.A., Drieberg, M., Susanto, M., Tumian, A., Yahya, M.: An enhanced deep neural network approach for wifi fingerprinting-based multi-floor indoor localization. IEEE Open Journal of the Communications Society **6**, 560–575 (2025). https://doi.org/10.1109/OJCOMS.2024.3520005
12. Neupane, I., Shahrestani, S., Ruan, C.: Indoor localization of resource-constrained iot devices using wi-fi fingerprinting and convolutional neural network. In: Proceedings of the 2024 Australasian Computer Science Week. p. 2025. ACSW '24, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3641142.3641158
13. Obeidat, H., Shuaieb, W., Obeidat, O., Abd-Alhameed, R.: A review of indoor localization techniques and wireless technologies. Wireless Personal Communications **119**, 289–327 (2021). https://doi.org/10.1007/S11277-021-08209-5
14. Queralta, J.P., Almansa, C.M., Schiano, F., Floreano, D., Westerlund, T.: Uwb-based system for uav localization in gnss-denied environments: Characterization and dataset. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 4521–4528. IEEE (2020). https://doi.org/10.1109/IROS45743.2020.9341042

15. Song, X., Fan, X., He, X., Xiang, C., Ye, Q., Huang, X., Fang, G., Chen, L.L., Qin, J., Wang, Z.: 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). pp. 589–595 (2019). https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00139
16. Torres-Sospedra, J., Montoliu, R., Martínez-Usó, A., Avariento, J.P., Arnau, T.J., Benedito-Bordonau, M., Huerta, J.: 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). pp. 261–270 (2014). https://doi.org/10.1109/IPIN.2014.7275492
17. Wang, S., Zhang, S., Ma, J., Dobre, O.A.: Graph-neural-network-based wifi indoor localization system with access point selection. IEEE Internet of Things Journal **11**(20), 33550–33564 (2024). https://doi.org/10.1109/JIOT.2024.3430087
18. Xie, B., Cao, J., Xie, J., Khan, F.S., Pang, Y.: Sed: A simple encoder-decoder for open-vocabulary semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3426–3436 (2024). https://doi.org/10.1109/CVPR52733.2024.00329
19. Xu, S., Zhang, L., Tang, Y., Han, C., Wu, H., Song, A.: Channel attention for sensor-based activity recognition: Embedding features into all frequencies in dct domain. IEEE Transactions on Knowledge and Data Engineering **35**(12), 12497–12512 (2023). https://doi.org/10.1109/TKDE.2023.3277839
20. Yang, X., Zhuang, Y., Gu, F., Shi, M., Cao, X., Li, Y., Zhou, B., Chen, L.: Deepwipos: A deep learning-based wireless positioning framework to address fingerprint instability. IEEE Transactions on Vehicular Technology **72**(6), 8018–8034 (2023). https://doi.org/10.1109/TVT.2023.3243196
21. Zhang, X., Zhang, Y., Liu, G., Jiang, T.: Autoloc: Toward ubiquitous aoa-based indoor localization using commodity wifi. IEEE Transactions on Vehicular Technology **72**(6), 8049–8060 (2023). https://doi.org/10.1109/TVT.2023.3243912
22. Zhao, S., Duan, Y., Roy, N., Zhang, B.: A deep learning methodology based on adaptive multiscale cnn and enhanced highway lstm for industrial process fault diagnosis. Reliability engineering & system safety **249**, 110208 (2024). https://doi.org/10.1016/J.RESS.2024.110208
23. Zhou, X., Chen, T., Guo, D., Teng, X., Yuan, B.: From one to crowd: A survey on crowdsourcing-based wireless indoor localization. Frontiers of Computer Science **12**, 423–450 (2018). https://doi.org/10.1007/S11704-017-6520-Z