

# Hierarchical Multi-Building And Multi-Floor Indoor Localization Based On Recurrent Neural Networks

Abdalla Elmokhtar Ahmed Elesawi and Kyeong Soo Kim

School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, P. R. China.

Email: A.Elesawi19@student.xjtlu.edu.cn, Kyeongsoo.Kim@xjtlu.edu.cn

**Abstract**—There has been an increasing tendency to move from outdoor to indoor lifestyle in modern cities. The emergence of big shopping malls, indoor sports complexes, factories, and warehouses is accelerating this tendency. In such an environment, indoor localization becomes one of the essential services, and the indoor localization systems to be deployed should be scalable enough to cover the expected expansion of those indoor facilities. One of the most economical and practical approaches to indoor localization is Wi-Fi fingerprinting, which exploits the widely-deployed Wi-Fi networks using mobile devices (e.g., smartphones) without any modification of the existing infrastructure. Traditional Wi-Fi fingerprinting schemes rely on complicated data pre/post-processing and time-consuming manual parameter tuning. In this paper, we propose hierarchical multi-building and multi-floor indoor localization based on a recurrent neural network (RNN) using Wi-Fi fingerprinting, eliminating the need of complicated data pre/post-processing and with less parameter tuning. The RNN in the proposed scheme estimates locations in a sequential manner from a general to a specific one (e.g., building→floor→location) in order to exploit the hierarchical nature of the localization in multi-building and multi-floor environments. The experimental results with the UJIIndoorLoc dataset demonstrate that the proposed scheme estimates building and floor with 100% and 95.24% accuracy, respectively, and provides three-dimensional positioning error of 8.62m, which outperforms existing deep neural network-based schemes.

**Index Terms**—Multi-building and multi-floor Indoor localization, Wi-Fi fingerprinting, recurrent neural networks (RNNs).

## I. INTRODUCTION

In modern smart cities, there is a huge demand for location-based services (LBS) like advertising, tracking, and navigation. Because people spend most of their time in indoor environments like shopping malls, hospitals, and airports [1], we need to provide LBS indoors. Most existing localization systems, however, cannot be used indoors; the lack of line of sight and the effect of multipath propagation on signals make it hard to utilize the localization systems such as the global positioning system (GPS) for indoor localization [2].

Different technologies have been utilized to implement indoor LBS (ILBS), including wireless networks, active/passive tags, and vision/camera technologies [3]. Among many wireless technologies for indoor localization Wi-Fi is the most feasible and popular technology [4]; Especially, Wi-Fi fingerprinting is widely used for indoor localization due to its wide availability and the lack of strict requirements on the information on network topologies. In Wi-Fi fingerprinting, received signal strength (RSS) measurements are not directly

used for distance estimation with a path loss model, which is a basis for multilateration, due to multipath fading effects [5]; instead, Wi-Fi fingerprinting uses RSS as one of location-dependent characteristics (i.e., location fingerprint) in inferring the location. We first build a database of the RSS indicators (RSSIs) from all access points (APs) measured at known locations called reference points (RPs) together with their location information like two dimensional (2D) or three dimensional (3D) positions. This information should be collected many times for each RP using different devices, different users, and different orientations to mitigate the effect of fluctuations in RSS measurements.

Traditional approaches for Wi-Fi fingerprinting—e.g., K-nearest neighbor (KNN) [6], weighted KNN (wKNN), and support vector machine (SVM) [7]—require a lot of efforts for filtering and parameter tuning, which are quite time-consuming. In recent years, deep neural networks (DNN) have been widely adopted to deal with large-scale, noisy Wi-Fi fingerprinting datasets [8], [9], and different machine learning techniques are combined with DNNs, too [10]. Due to its higher accuracy and less computational complexity, convolutional neural network (CNN) are used in [11], [12].

In this paper, we introduce a new approach to hierarchical multi-building and multi-floor indoor localization based on a recurrent neural network (RNN) with stacked auto-encoder (SAE) using Wi-Fi fingerprinting, eliminating the need of complicated data pre/post-processing and with less parameter tuning. The RNN in the proposed scheme estimates locations in a sequential manner—i.e., building→floor→location—to exploit the hierarchical nature of the localization in multi-building and multi-floor environments.

The outline of the rest of the paper is as follows: Section II reviews the related work on indoor localization. Section III discusses the proposed network architecture. In Section IV, we present experimental results with the best configuration, where we also discuss the results in comparison with other approaches. Finally, we conclude our work in Section V.

## II. RELATED WORK

Traditional approaches like KNN [13] and wKNN algorithms are time consuming and need a lot of tuning, which are not suitable for large-scale indoor environments where a lot of data are to be collected and processed. Recently, researchers adopt deep learning approaches for indoor localization.

As for multi-building and multi-floor indoor localization datasets, this work is based on the publicly-available UJIIndoorLoc dataset [14]. Note that, however, there are several works using only a subset of UJIIndoorLoc [11], [15] or using only training data for both training and testing their models [16]. Even though these works report good results, there is no guarantee that the performance of their proposed models would be good as well with the whole dataset due to the statistical differences in the training and validation datasets. In the following, therefore, we focus only on the related works based on the full UJIIndoorLoc dataset.

In [8], a DNN model was proposed for the classification of building/floor. An SAE is used to reduce the number of features, which is followed by a DNN classifier for the *multi-class classification* of building-floor using flattened labels. Because a flattened label is represented with one-hot encoding, this DNN model only classifies building and floor to avoid huge number of output nodes required for locations. This work achieves the success rate of 92%.

To tackle the problem of scalability, *multi-label classification* was proposed in [9]. It reduces the number of output nodes significantly. The proposed architecture also consists of an SAE followed by a DNN, but the first  $N$  output nodes of the DNN are used for building classification, where  $N$  is the number of buildings, the next  $M$  output nodes for floor classification, where  $M$  is the maximum number of floors in all buildings, and the rest of the output nodes for location estimation. This work achieves 91.27% for floor hit rate and 9.29 m for 3D positioning error.

In [10], random forest followed by an SAE was used to filter and reduce the dimensionality of the dataset. Filtered data is classified using 4 primary classifiers (i.e., CNN, ELM, SVM, and XGBoost), and then the secondary classifier predicts the class from those 4 values. This work is only for floor classification, and it achieves 95.13% for floor hit rate.

Integration of an SAE and CNNs was presented in [12]. Three different networks are used for floor classification, building classification and position estimation. The SAE followed by dropout layers gives the input to one dimensional CNN (1D-CNN) followed by fully-connected layers to give 5 output nodes for floor classification. For building classification, the SAE is directly connected to fully-connected layers with 3 output nodes. For position estimation, they used the same floor classification model with some changes: First, they remove dropout layers between the SAE and the 1D-CNN. Second, they change the number of output nodes to 2 to represent  $x$  and  $y$  coordinates. Finally to get continuous values for  $x$  and  $y$ , they use rectified linear unit (ReLU) instead of softmax as output activation function. Before training the model, however, lots of data pre-processing is needed for this work; dividing the dataset to sub-datasets, creating rectangle areas then dividing them to cell grids, choosing the center of each cell grid, selecting data based on the previous divisions are some of the preparation steps for training phase. They achieve 96.03% for floor hit rate and 11.78 m for positioning error.

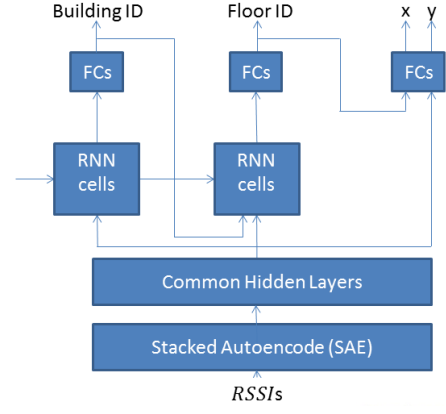


Fig. 1. Proposed network architecture based on RNN and SAE.

### III. PROPOSED NETWORK ARCHITECTURE

Fig. 1 shows the proposed network architecture based on RNN and SAE, which takes RSSIs as inputs and returns building ID, floor ID, and location coordinates  $(x, y)$  as outputs, where we take into account the following major points in our design: First, we fed the output from upper-level class to lower-level class to exploit the hierarchical nature of the multi-building and multi-floor indoor localization. Second, because the position estimation is different from building and floor classification in nature, we exclude it from the RNN. Third, we add an SAE before common hidden layers to reduce the dimensionality of a feature space and thereby denoise RSSIs.

Note that scalability is one of the key challenges to be addressed in large-scale multi-building and multi-floor indoor localization. Reducing the number of output nodes is one of the major techniques to make the system scalable. For the UJIIndoorLoc dataset, the number of output nodes would be 905 when we use multi-class classification, which could be reduced to 118—i.e., the sum of the number of buildings, the maximum number of floors, and the maximum number of floor locations—by using multi-label classification [17]. In the proposed work, we greatly reduce this number to 4—i.e., 1 for building, 1 for floor, and 2 for location—by numerical representation of outputs instead of one-hot-encoding. We convert building and floor classification to regression problem and round the regression outputs to get the class numbers.

### IV. EXPERIMENTAL RESULTS

We carry out experiments with the UJIIndoorLoc Wi-Fi fingerprinting dataset [14] to investigate the effects of RNN parameter values on the localization performance with a major focus on RNN cell types and dropout rates. Note that the publicly-available UJIIndoorLoc dataset provides only training and validation data; test data were provided only to the competitors at the Evaluating Ambient Assisted Living (EvAAL) competition [18]. Therefore, we split the training data into new training and validation data with the ratio of 90:10, and we use the validation data as test data.

TABLE I  
HYPER PARAMETERS

| Parameter                   | Value      |
|-----------------------------|------------|
| SAE Hidden Layers           | 256-128-64 |
| SAE Activation              | ReLU       |
| SAE Optimizer               | Adam       |
| SAE Loss                    | MSE        |
| Common Hidden Layers        | 128-128    |
| Common Activation           | ReLU       |
| Common Dropout              | 0.2        |
| Common Loss                 | MSE        |
| RNN Cells                   | 128-128    |
| RNN Activation              | ReLU       |
| RNN Optimizer               | Adam       |
| RNN Loss                    | MSE        |
| BF Classifier Hidden Layers | 32-1       |
| BF Classifier Activation    | MSE        |
| BF Classifier Optimizer     | Adam       |
| BF Classifier Dropout       | 0.2        |
| BF Classifier Loss          | ReLU       |
| Position Hidden Layers      | 128-128-2  |
| Position Activation         | MSE        |
| Position Optimizer          | Adam       |
| Position Dropout            | 0.1        |
| Position Loss               | tanh       |

As for performance metrics, we use classification accuracy and positioning error: As for the classification accuracy, we calculate hit rates for building, floor, and building/floor, the last of which counts only the correct identification of both building and floor IDs; the positioning error is 2D or 3D Euclidean distance between a predicted and true positions calculated as discussed in [18].

Table I summarizes the values of hyper parameters for the experiments. We use SAE consists of three hidden layers of 256, 128, and 64 nodes, which is mentioned as the best architecture for SAE in [8]. The SAE is then followed by two common hidden layers with 128 nodes each. For building and floor classifiers, we have two stacked RNN cells followed by two fully-connected layers of 32 nodes and 1 node as the output node. Position estimator consists of three fully-connected layers of 128, 128, and 2 nodes. Note that there are two output nodes for coordinates  $(x, y)$ . As for epochs, we apply *early stopping* with patience of 5, which forces early stopping to run at least 5 epochs even though there is no improvement.

First, we compare the performance of two stacked RNN cell types, i.e., standard RNN and long short-term memory (LSTM). The experiments were first conducted using standard RNN cells over a range of the numbers of nodes, batch sizes, and dropout rates and repeated again using LSTM cells, whose best results and configurations are summarized in Tables II and III, respectively. The results show that LSTM cell provides slightly better performance in floor estimation and positioning error: The best results for standard RNN are 100% for building hit rate, 94.42% for floor hit rate, and 8.68 m for positioning error; for LSTM, the best results are 100% for building hit rate, 95.23% for floor hit rate, and 8.62 m for positioning error. Table IV compares our results against those of other

TABLE II  
RESULTS OF DIFFERENT RNN CELL TYPES

| RNN Cell Type | Building Hit Rate (%) | Floor Hit Rate (%) | Positioning Error (m) |
|---------------|-----------------------|--------------------|-----------------------|
| Standard RNN  | 100                   | 94.42              | 8.68                  |
| LSTM          | 100                   | 95.23              | 8.62                  |

TABLE III  
BEST CONFIGURATION

| Configuration                     | Value |
|-----------------------------------|-------|
| RNN Cell Type                     | LSTM  |
| Number of Nodes                   | 128   |
| Batch Size                        | 32    |
| Building/Floor Classifier Dropout | 0.2   |
| Building/Floor Classifier Epochs  | 10    |
| Position Estimation Dropout       | 0.1   |
| Position Estimation Epochs        | 30    |

TABLE IV  
COMPARISON WITH OTHER DNN-BASED APPROACHES

| Approach             | Floor Hit Rate (%) | Positioning Error (m) |
|----------------------|--------------------|-----------------------|
| Proposed             | 95.23              | 8.62                  |
| DNN [8]              | 92.00              | N/A                   |
| Scalable DNN [9]     | 91.27              | 9.29                  |
| RF+SAE+Stacking [10] | 95.13              | N/A                   |
| CNNLoc [12]          | 96.03              | 11.78                 |

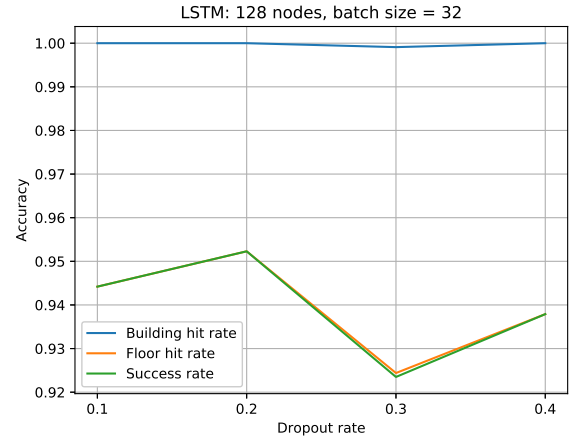


Fig. 2. Effect of dropout rate on building and floor hit rate.

approaches based on the same UJIIndoorLoc data set, where we observe that the proposed approach outperforms all other approaches except CNNLoc [12]; CNNLoc, which requires a lot of data pre-processing compared to the proposed one, shows slightly better floor hit rate of 96.03% but much higher positioning error of 11.78 m.

We also investigate the effect of dropout rate on the localization performance as shown in Figs. 2 and 3. From the figures, we can observe that dropout rate of 0.2 gives the best results

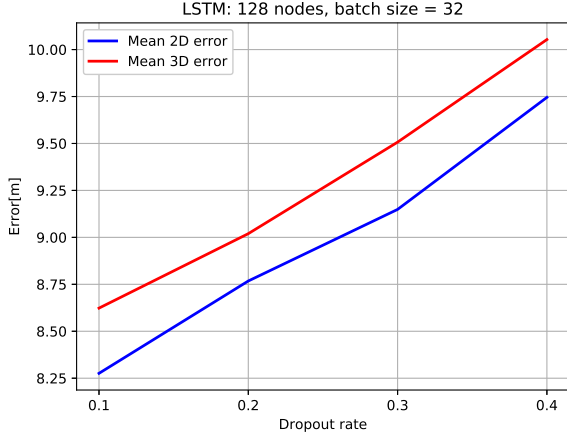


Fig. 3. Effect of dropout rate on position estimation.

for both building and floor estimation, while dropout rate of 0.1 always gives the best results for coordinates estimation.

Table V compares our results against those of the best four teams in the EvAAL competition. Even though the objective

TABLE V  
COMPARISON WITH THE RESULTS FROM EVAAL/IPIN 2015  
COMPETITION [18]

|                          | Proposed | MOSAIC | HFTS  | RTLSUM | ICSL  |
|--------------------------|----------|--------|-------|--------|-------|
| Building Hit Rate (%)    | 100      | 98.65  | 100   | 100    | 100   |
| Floor Hit Rate (%)       | 95.23    | 93.86  | 96.25 | 93.74  | 86.93 |
| 3D Positioning Error (m) | 8.62     | 11.64  | 8.49  | 6.20   | 7.67  |

and fair comparison is not possible due to the unavailability of the original testing dataset, which were given only to the participants of the EvAAL competition, the comparison in Table V could give us an idea on the relative performance of the proposed approach, where we find that the proposed approach outperforms MOSAIC in all aspects and that our floor hit rate is higher than that of MOSAIC, RTLSUM, and ICSL.

## V. CONCLUSIONS

In this paper, we have proposed RNN-based hierarchical multi-building and multi-floor indoor localization based on Wi-Fi fingerprinting. In our approach, SAE and RNN are used for the reduction of feature space dimension and the exploitation of the hierarchical nature of the localization in multi-building and multi-floor environments, respectively.

Through the experimental results based on the publicly-available UJIIndoorLoc dataset, we observe that the proposed indoor localization scheme achieves the accuracy of 100% and 95.23% for building and floor estimation and 3D positioning error of 8.62 m, which outperforms most of the existing approaches including those based on DNNs.

Note that the proposed scheme clearly shows the advantages of hierarchical indoor localization enabled by the use of RNN,

while sharing the benefits of the single-DNN-based schemes of [9], [17], [19] like the elimination of complicated data pre/post-processing and less parameter tuning.

## ACKNOWLEDGMENT

This work was supported in part by Postgraduate Research Scholarships (under Grant PGRS1912001) and Key Program Special Fund (under Grant KSF-E-25) of Xi'an Jiaotong-Liverpool University.

## REFERENCES

- [1] "Report to congress on indoor air quality. volume 2. assessment and control of indoor air pollution. final report," 8 1989. [Online]. Available: <https://www.osti.gov/biblio/6958939>
- [2] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. INFOCOM 2000*, 2000, pp. 775–784.
- [3] C. Basri and A. El Khadimi, "Survey on indoor localization system and recent advances of WIFI fingerprinting technique," in *Proc. ICMCS 2016*, 2016, pp. 253–259.
- [4] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2568–2599, 2019.
- [5] P. Prasithsangaree, P. Krishnamurthy, and P. Chrysanthos, "On indoor position location with wireless LANs," in *Proc. PIMRC 2002*, Sep. 2002, pp. 720–724.
- [6] J. Torres-Sospedra et al., "Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems," *Expert Systems with Applications*, vol. 42, no. 23, pp. 9263–9278, 2015.
- [7] A. Chriki, H. Touati, and H. Snoussi, "SVM-based indoor localization in wireless sensor networks," in *Proc. IWCMC 2017*, 2017, pp. 1144–1149.
- [8] M. Nowicki and J. Wietrzykowski, "Low-effort place recognition with Wi-Fi fingerprints using deep learning," *ArXiv e-prints*, Apr. 2017.
- [9] K. S. Kim, S. Lee, and K. Huang, "A scalable deep neural network architecture for multi-building and multi-floor indoor localization based on Wi-Fi fingerprinting," *Big Data Analytics*, vol. 3, no. 4, Apr. 2018.
- [10] G. JunLin et al., "Wi-Fi fingerprint positioning method based on fusion of autoencoder and stacking mode," in *Proc. ICCST 2020*, 2020, pp. 356–361.
- [11] M. Ibrahim, M. Torki, and M. ElNainay, "CNN based indoor localization using RSS time-series," in *Proc. ISCC 2018*, 2018, pp. 01044–01049.
- [12] X. Song et al., "CNNLoc: Deep-learning based indoor localization with Wi-Fi fingerprinting," in *Proc. Smart-World/SCALCOM/UIC/ATC/CBDCCom/IOP/SCI 2019*, 2019, pp. 589–595.
- [13] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [14] J. Torres-Sospedra et al., "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems," in *Proc. IPIN 2014*, Busan, Korea, Oct. 2014, pp. 261–270.
- [15] A. Alitalieshi, H. Jazayeriy, and S. J. Kazemitabar, "Wi-Fi fingerprinting based floor detection with hierarchical extreme learning machine," in *Proc. ICCKE 2020*, 2020, pp. 113–117.
- [16] S. BelMannoubi, H. Touati, and H. Snoussi, "Stacked auto-encoder for scalable indoor localization in wireless sensor networks," in *Proc. IWCMC 2019*, 2019, pp. 1245–1250.
- [17] K. S. Kim et al., "Large-scale location-aware services in access: Hierarchical building/floor classification and location estimation using Wi-Fi fingerprinting based on deep neural networks," *Fiber and Integrated Optics*, vol. 37, no. 5, pp. 277–289, Apr. 2018.
- [18] A. Moreira et al., "Wi-Fi fingerprinting in the real world – RTLSUM at the EvAAL competition," in *Proc. IPIN 2015*, Banff, Alberta, Canada, Oct. 2015, pp. 1–10.
- [19] K. S. Kim, "Hybrid building/floor classification and location coordinates regression using a single-input and multi-output deep neural network for large-scale indoor localization based on Wi-Fi fingerprinting," in *Proc. CANDAR'18*, Hida Takayama, Japan, Nov. 2018.