Indoor Localization with WiFi Fingerprinting Using Convolutional Neural Network

Jin-Woo Jang and Song-Nam Hong
Ajou University, Suwon, Korea,
email: {wdfrty1234, snhong}@ajou.ac.kr

***Abstract*—Indoor localization has been an active research field for decades, due to its wide range of applications. WiFi fingerprinting, which estimates the user's locations using pre-collecting WiFi signals as references, is of particular interest since these days, every user can easily access to WiFi networks. Among numerous methods, deep-neural-network (DNN) based methods have shown an attractive performance but their major drawback is the sensitivity to the fluctuation of received signals (caused by multipaths). To yield a satisfactory performance, thus, a sufficiently large number of possible cases should be trained, which costs a lot. In this paper, we address the above problem by presenting a convolutional neural network (CNN) based localization method. As success in image classifications, the proposed method can be robust to the small changes of received signals as it exploits the topology of a radio map as well as signal strengths. Via experimental results, we demonstrate that the proposed CNN method can outperform the other DNN-based methods using publicly available datasets provided in IPIN 2015.**

***Index Terms*—Indoor Positioning, convolutional neural network, WiFi fingerprints.**

I. Introduction

Recently, indoor localization (or positioning) is of great interest due to the wide range of applications [1]. Although Global Positioning System (GPS) provides an attractive performance in outdoor localizations, it is not suitable for the indoor localizations, because of signal attenuations (by blocking objects), software mismatch, and so on [2]. These days, one can easily access to WiFi networks using his/her mobile devices (e.g., smart phones) in indoor environments (e.g., shopping malls). Hence, indoor localization methods using WiFi signals have been extensively investigated [3]. Furthermore, they are very cost-effective because no additional equipments are required.

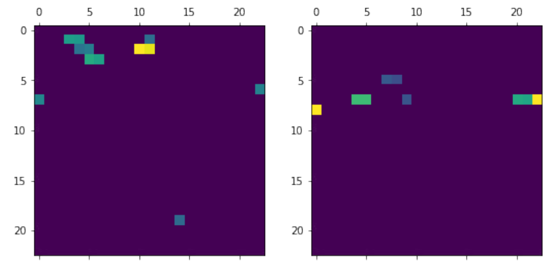
The previous methods on indoor localizations can be classified as *triangulation* and *fingerprinting* where the former employs the geometric properties of triangles to estimate a target location while the latter estimates it using some characteristics (so-called fingerprint) of a scene. The accuracy of triangulation degrades by the signal noises as multipath, namely, this method is not robust to multipath indoor environments such as shopping malls. On the other hand, the accuracy of fingerprint largely relies on the sufficiency of datasets. Currently, it is quite easy to collect WiFi signals and thus, WiFi fingerprinting together with deep/machine learning becomes an attractive approach. [4] Especially, machine learning algorithms as MOSAIC [5], HFTLoc [6], ICSL [7], and RTLSUM [8] showed better performances than the others at the Evaluating Ambient Assisted Living (EvAAL) competition at the International Conference on Indoor Positioning and Indoor Navigation (IPIN) 2015 [9]. However, these methods are time-consuming due to the use of complex filtering and parameter turning. Thus, deep-neural-network (DNN) based methods have been recently investigated [10].

The major drawback of DNN-based methods is that they are very sensitive to the change of input datas. Thus, when dataset is not sufficient, the accuracy is not satisfactory. For example, the DNN-based classifier in handwriting(e.g., MNIST [11]), which is the most commonly used example in deep learning, showed a lower performance for untrained fonts, even if they are identical letters. To avoid such problem, it is required to perform a training for various possible cases. Alternatively, a convolutional neural network (CNN) has been proposed. Recent studies showed that CNN-based classifier gives a satisfactory performance in image classification. The main advantage of a CNN is that it is able to learn the overall topology of an image via convolution operation using filter [12].

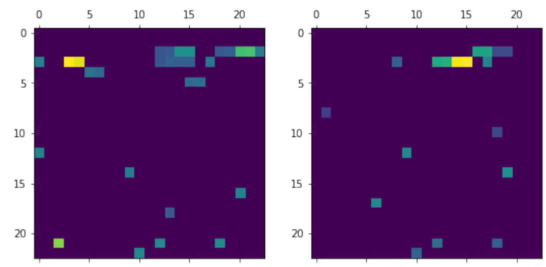
Motivated by this, we present a CNN-based WiFi fingerprinting for indoor localizations. As in image classification, we expect that the proposed method could be robust to the sensitivity of a change of input datas (caused by indoor multipath). In the proposed method, we first build a 2-D *virtual* radio map from the original 1-D WiFi signals (e.g., Received Signal Strength Indicator (RSSI) values) and then construct a CNN using 2-D radio maps as inputs. Therefore, the proposed method can learn the topology of radio maps as well as signal strengths, while the previous DNN approaches only considers the signal strengths. We further improve the accuracy of the proposed method using various enhancing techniques as feature scaling, dropout, data balancing, and ensemble. Via simulation results, we demonstrate that the proposed CNN-based method yields a higher accuracy than the other DNN-based methods.

II. The Proposed CNN-based WiFi fingerprinting

In this section, we present a WiFi fingerprinting using a CNN for indoor localizations. The proposed CNN classifier identifies the location of a user (e.g., building ID and floor



(a) Building ID : 0 Floor : 0 / Building ID : 1 Floor : 3



(b) Building ID : 2 Floor : 1 / Building ID : 2 Floor : 4

Fig. 1. Converting 1-d array Received Signal Strength into 2-d array.

ID) by leveraging a Received Signal Strength Indicator (RSSI) value obtained from neighboring Access Points (APs) (e.g., APs on campus).

A. Dataset

Suppose that each user receives the RSSI values from neighboring N APs, namely, the data $\mathbf{x} = (x_1, ..., x_N)$ is a 1-D vector of a length-N where x_i denotes the RSSI value obtained from the AP i. Also, it is assumed that each data has a label $j \in \{1, ..., M\}$. For instance, labels can be determined by the combination of a building and floor IDs. Note that, for a training, we know the both N RSSI values and the corresponding label as $(\mathbf{x}_1, j_1)$ but for an acquisition, we only know the RSSI values $\mathbf{x}_2$ and want to estimate the corresponding label j_2.

In order to employ a CNN classifier, it is required to generate a suitable input structure from the given RSSI values $\mathbf{x}$ since the CNN classifier requires a 2-D array as an input. For this, we make a 2-D array $\mathbf{X}$ from the given 1-D array $\mathbf{x}$, possibly adding some dummy values (which have no meaning to the existing 1-D vector). As shown in Fig. 1, in the example of $N = 520$, the 9 dummy datas are added to the given data $\mathbf{x}$, for the purpose of building 23×23 2-D array $\mathbf{X}$ as follows:

$$\mathbf{X}_{i,j} = \bar{x}_{23\times(i-1)+j}, \tag{1}$$

where $\bar{\mathbf{x}} = (\bar{x}_1 = x_1, ..., \bar{x}_{520} = x_{520}, \bar{x}_{521} = 0, ..., \bar{x}_{529} = 0)$. Thus, in the proposed CNN classifier, the training datas $(\mathbf{X}, j)$ are used for a training.

B. Convolutional Neural Network

We in this section describe the proposed CNN model consisting of N_1 convolution layers, N_2 pooling layers, N_3 fully connected layers, and a softmax layer. Also, it is assumed that the size of an input data is $N_0 \times N_0$. In the below, we explain the detailed role of each layer. The overall system model is shown in Fig. 2

- *Convolution layer:* A randomly initialized filter extracts the characteristics of the input. Since one input value can have various characteristics, n multiple filters are used to extract all the features which the original data can have. Also, to keep the size of local features extracted from the input data as $N_0 \times N_0$, zero padding is used for each convolution layer.
- *Pooling layer:* It plays a role of sub-sampling by using the features extracted from the convolution layers. The sub-sampling can decrease the time-complexity by reducing the number of operations in the next convolution layer or fully-connected layer. Among the various methods to configure the pooling layer, max-pooling method is used in the proposed method, to extract the largest value in the sliding window for the sub-sampling.
- *Fully-connected layer:* The local features extracted from the convolution layer and pooling layer are put into the neural network model. At the end of the fully-connected layer, we used a softmax layer which is widely used for the classifications of multiple classes. Also, the result corresponds to a probability that all the classes are equal to 1, and the class with the highest probability is the estimated label for the corresponding input.

As an example, the operation of the proposed CNN classifier is shown in Fig. 3, which is used for the simulations in Section IV. Here, a 23×23 input data passes through the two convolution layers with 64 filters and one max-pooling layer. Passing through the max-pooling layer, the $23 \times 23 \times 64$ local features are sub-sampled to the $12 \times 12 \times 64$ local features. The resulting local features go through the two convolution layers with 128 filters and one max-pooling layer. Again, passing through the max-pooling layer, the $12 \times 12 \times 128$ local features are sub-sampled to the $6 \times 6 \times 128$ local features. In this example, the ReLu function is chosen as an activation function (from a convolution layer to the next layer):

$$ReLu(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \tag{2}$$

To put it into a fully connected layer, connect $6 \times 6 \times 128$ local features and make it a layer with 4608 hidden nodes. After passing through two fully-connected layers with 512 hidden nodes, the softmax layer is used to determine the label corresponding to the input value. Also, when training the proposed model, we used the *Adam Optimizer* [13] as the optimization method.

Furthermore, we employ a variety of techniques to increase the accuracy of the proposed CNN classifier as follows:

- *Feature scaling:* The most important factor in deep learning is how much data is available for training and how good it is. Hence, datasets are generally required to be normalized for better training. In this paper, we used a *robust scalar* to normalize the datasets. The robust scalar

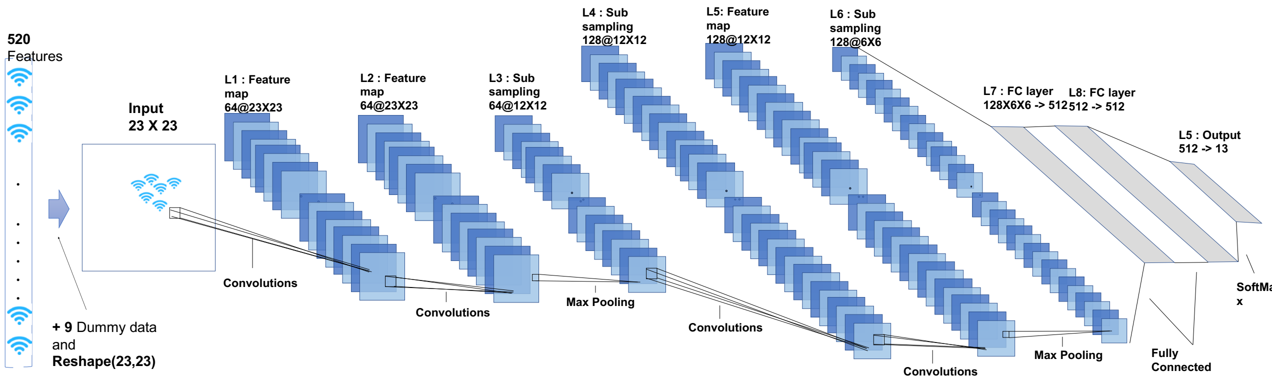


Fig. 2. Proposed system model.

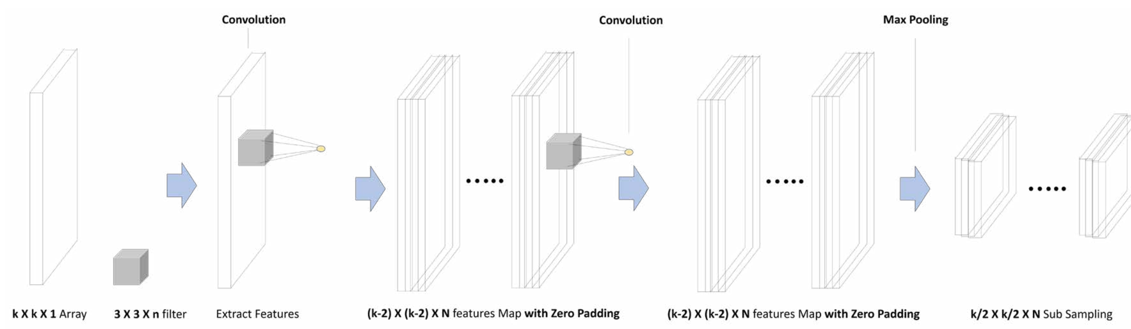


Fig. 3. The proposed Convolution Neural Network architecture.

employs the interquartile range (IQR) especially using the first and fourth quartiles. The advantage of this method is that it can reduce the impart of outliers. We use this method since it can be assumed that the values which are not measured signals (due to the switch-off of some APs) are outliers.

- *Dropout:* The dropout is the way to solve the overfitting problem in deep learning where the overfitting refers to the case when the accuracy is too low for actual use since the sample data is trained too accurately. The dropout refers to randomly disconnected nodes at a given rate. By performing this, the overfitting problem can be addressed to some extent. In the proposed model, dropout is used after each fully-connected layer.
- *Data balancing:* The number of samples for each label of the datasets assigned to the training set can be different. For example, we might have more samples for some particular labels. In this case, there is a possibility that *biased* learning can occur during training. Thus, before training, we need to extract the same number of samples for each label and create a new (balanced) dataset. Table I shows the number of samples per label of the data set used in the experiment Section III.
- *Ensemble:* Ensemble is a method of creating several identical models and adding the probability values resulting from each model, and choosing the highest value among them. This method is possible because the filter used for convolution is initialized at random, so the results that can come from each model may be different. In the proposed model, we created 3 identical models and performed ensemble.

TABLE I
NUMBER OF SAMPLES PER LABEL

Building ID	Floor	Num of Samples
0	0	1059
	1	1356
	2	1443
	3	1391
1	0	1368
	1	1481
	2	1396
	3	948
2	0	1942
	1	2162
	2	1577
	3	2709
	4	1102
Total	13	19937

III. EXPERIMENTAL RESULTS

In this section, we verify the superiority of the proposed CNN classifier by comparing the previous works. For experimental results, we employed the publicly available UJIIndoorLoc data set [15] provided at the Evaluating Ambient Assisted Living (EvAAL) competition at the International Conference on Indoor Positioning and Indoor Navigation (IPIN) 2015 [8].

TABLE II
COMPARISON OF VARIOUS STRUCTURE

<table>
<tr><td colspan="7">Models</td></tr>
<tr><td rowspan="6">Structure</td><td>Conv-32</td><td>Conv-64</td><td>Conv-128</td><td>Conv-32</td><td>Conv-64</td><td>Conv-128</td></tr>
<tr><td colspan="3">Maxpooling</td><td>Conv-32</td><td>Conv-64</td><td>Conv-128</td></tr>
<tr><td>Conv-64</td><td>Conv-128</td><td>Conv-256</td><td colspan="3">Maxpooling</td></tr>
<tr><td colspan="3">Maxpooling</td><td>Conv-64</td><td>Conv-128</td><td>Conv-256</td></tr>
<tr><td>Conv-128</td><td>Conv-256</td><td>Conv-512</td><td>Conv-64</td><td>Conv-128</td><td>Conv-256</td></tr>
<tr><td colspan="6">Maxpooling</td></tr>
<tr><td>mean accuracy</td><td>94.4%</td><td>94.67%</td><td>94.49%</td><td>94.82%</td><td>95.19%</td><td>95.06%</td></tr>
</table>

The dataset consists of RSSI values from 520 APs, which are measured at various locations on campus. Also, each dataset is belong to one of 13 labels (which are determined by the combination of building IDs and floor IDs). Prior to performing the experiment, we changed the lack of measurement denoted by 100 dBm to -110dBm in the UJIIndoorLoc dataset. This is because, even if the measured signal is -110dBm, it can be considered as the lack of signal strength [14]. When training the model, we used 19937 samples in trainingData set given in UJIindoorLoc dataset. However, since testset is not provided for testing, we use 1111 samples in validationData set given in UJIindoorLoc dataset to test it as testset.

A. Optimization of the proposed CNN classifier

Before comparing performance with other models, we tested various models to find the best performance model, changing the number of layers and the number of filters.

Before performing the training, we randomly extracted 900 samples per label in a given training set and created a new dataset with a total of 11700 samples. This process is performed each time the model is trained. Because the samples used for training are different each time and the filter used for convolution operation and the hidden node in fully connected layer are initialized randomly each time, the results can vary each time. Therefore, the testing was performed five times for each structure and in order to compare each structure, we used the average of the five testing results. The test results are shown in the table II.

All structures consist of two fully connected layers and a softmax layer after the last max-pooling layer. In all models, we used 3x3 filter for convolution operations and stride value of 2 in max-pooling. The results of the test show that the accuracy of two max-pooling is higher than the case of using 3 max-pooling. In addition, we can see that the higher the number of the filters used in the conv-layer, the higher the performance. In a given dataset, the structure of two Conv-layers with 64 filters, two Conv-layers with 128 filters, and two Max-pooling layers was the best performance.

B. Comparison with the existing methods

In order to examine the performance of the proposed system model we compare the performance with the previously proposed deep learning models. The comparison of each model compares the probability of finding Building ID and Floor accurately. The models used for comparison are SAE(Stacked Autoencoder) [14] + Classifier and Scalable DNN Architecture [17]. SAE+Classifier and Scalable DNN are DNN based Deep Learning model, but our proposed model is CNN based Deep Learning model.

TABLE III
COMPARISON OF DEEP LEARNING BASED MODELS

	SAE+Classifier	Scalable DNN	Proposed CNN Model
B&F Accuracy	91.1%	92.89%	95.41%

Table III shows the comparison results of each model. Experimental results show that the proposed model is about 4.31% and 2.52% higher than the existing models. When we consider only the B&F success rate, we can confirm that CNN based classifier has the highest performance. As a result, the CNN model considering the topology of the *virtual* radio map compared to the DNN showed better performance. Thus, it can be seen that it is better to consider the topology of received signals than to simply consider signal strength alone.

C. Impact of improving techniques

We perform the experiments to see how the performance of the proposed CNN classifier can be improved using the various enhancing techniques in Section II. For the feature scaling, we considered the various normalization functions provided by *scikit-learn* [16]. Table IV shows different results for each function.

TABLE IV
COMPARISON OF VARIOUS FEATURE SCALING METHODS

	Standard	Min_Max	Max_Abs	Robust	Quantile
B&F Accuracy	90.63%	94.77%	94.77%	95.41%	85.95%

From Table IV, we observe that the B&F accuracy in the Quantile and Standard methods are 85.95% and 90.63%, respectively. This shows the relatively low-accuracy compared to the Robust scalar adopted in the proposed method. Also, we can see that Min_Max and Max_Abs methods have the same B&F accuracy of 94.77%, which are lower than the Robust scalar but not significantly different. Therefore, the robust scaler considering the lack of measurement as the outliers showed the best performance

We next consider the impact of Dropout. The corresponding results are provided in Table V. The result shows that the performance improvement of 1.18% is achieved using Dropout technique. In addition, the results of the experiment showed that the accuracy can be improved by reducing overfitting through dropout.

TABLE V
WHETHER OR NOT TO USE DROPOUT

	with Dropout	without Dropout
B&F Accuracy	95.5%	94.32%

Also, we consider the impact of balanced samples. When the number of samples corresponding to a label in a dataset is not uniform, it can be learned biased and performance may be degraded. Unbalanced samples mean to use all the samples given in the training set during training. Balanced samples are a newly created dataset that randomly extracts 900 samples per label. The difference in results when the number of samples is unbalanced and balanced is shown in Table VI. The results show the 94.06% and 95.13% B&F accuracies when using unbalanced and balanced samples, respectively. Therefore, it is important to make a balanced sample to improve the accuracy.

TABLE VI
COMPARISON OF BALANCED AND UNBALANCED SAMPLES

	Unbalanced	Balanced
B&F Accuracy	94.06%	95.13%

Lastly, we consider the impact of using ensemble where the use of ensemble can obtain more accurate results due to the use of various trained models. The corresponding results are provided in Table VII. When the ensemble is not used, the result is 94.33%, which is less accurate than models using ensembles. In the case of the model using the ensemble, the accuracy of 95.76% and 95.67% is obtained when using three models and using five models.

TABLE VII
USE OF ENSENBLE

	Not used	3 models	5 models
B&F Accuracy	94.33%	95.76%	95.67%

IV. CONCLUSION

In this paper, we proposed CNN-based WiFi fingerprinting system and showed that it can provide a better accuracy than existing deep learning based multi-floor-multi-building classifiers. Note that the number of datasets which can be learned is the most important factor in deep learning such as multi-layer perceptron or convolutional neural network. Unfortunately, due to the cost of data collecting, it is necessary to achieve a high accuracy using a limited number of data samples. It was shown that the proposed method can improve the accuracy of the previous works, without additional data collection. Thus, the proposed CNN classifier can be a good approach especially when dataset is not sufficient.

In addition, the advantage of deep learning method is that once a model is trained, you are able to find a current location through one-forward propagation. In addition, the number of parameters used in the CNN model is smaller than the number of parameters used in the DNN model based on fully connected. Therefore, the CNN based model proposed in this paper has lower time-complexity and faster execution time than the existing method. Its time-complexity in actual acquisition process is very low, and hence, it can be used in real-time applications.

Finally, we would like to emphasize that the proposed method can be utilized in numerous applications as advertisement system in complex shopping mall, location measurement of a car in a parking structure, and emergency rescue service, etc.

ACKNOWLEDGEMENT

This work was supported by Samsung Research Funding & Incubation Center of Samsung Electronics under Project Number SRFC-IT1702-00 and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2017R1C1B1001909).

REFERENCES

[1] J. Xiao, Z. Zhou, Y. Yi, and L. M. Ni, " A Survey on Wireless Indoor Localization from the Device Perspective," *ACM Computing Surveys (CSUR)*, vol. 49, p. 25, 2016.

[2] A. S. Paul and E. A. Wan, "RSSI-Based Indoor Localization and Tracking Using Sigma-Point Kalman Smoothers," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 5, pp. 860-873, Oct. 2009.

[3] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: Wireless indoor localization with little human intervention," in *Proc. of the 18th ACM Annual International Conference on Mobile Computing and Networking (MobiCom 12)*, 2012

[4] H. Liu, H. Darabi, P. Banerjee, and Jing Liu, "Survey of Wireless Indoor Positioning Techniques and Systems," *IEEE Trans. Systems, Man, and Cybernetics, Part C: Applications and Reviews,* vol.37, no.6, pp. 1067-1080, Oct. 2007.

[5] R. Berkvens, M. Weyn, and H. Peremans, "Localization performance quantification by conditional entropy." in *Proceedings of the Sixth International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2015.

[6] S. Knauth, M. Storz, H. Dastageeri, A. Koukofikis, and N. Mahser-Hipp, "Fingerprint calibrated centroid and scalar product correlation rssi positioning in large environments," in *Proceedings of the Sixth International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2015, long paper at IEEE Xplore.

[7] S. Choi, J. Yoo, and H. J. Kim, "Machine learning for indoor localization Deep learning and semi-supervised learning," in *USB On-Site Proceedings of the Sixth International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2015, short paper.

[8] A. Moreira, M. J. Nicolau, F. Meneses, and A. Costa, "Wi-Fi fingerprinting in the real world - RTLSUM at the EvAAL competition," in *Proc. International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Banff, Alberta, Canada, pp. 1-10 (2015)

[9] J. Torres-Sospedra, R. Montoliu, A. Martnez-Us, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprintbased indoor localization problems," in *Proc. IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN),* Busan, Korea, 2014.

[10] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," To be published in *IEEE Transactions on Vehicular Technology*.

[11] Y. LeCun and C. Cortes. The mnist database of handwritten digits, 1998.

[12] P. Y. Simard, D. Steinkraus, and J.C. Platt. "Best practices for convolutional neural networks applied to visual document analysis". In *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, vol. 2, pp. 958-962, 2003.
[13] D. P. Kingma and B. Jimmy, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
[14] M. Nowicki and J. Wietrzykowski "Low-effort place recognition with WiFi fingerprints using deep learning," arXiv:1611.02049[cs.RO], Apr. 2017. [Online]. Available: https://arxiv.org/abs/1611.02049
[15] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml
[16] Scikit-learn preprocessing. http://scikitlearn.org/stable/modules/classes.html#module-sklearn.preprocessing
[17] K. 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," in *Proc. IEEE International Workshop on Fiber Optics in Access Network (FOAN),* pp. 1-5, Munich, Germany, Nov. 2017.