Indoor Localization Using Deep-Learning and Smartphone

Zainab Mohammed Resan[1], Muayad Sadik Croock[2]

University of Technology-Iraq, Computer Engineering Department, Baghdad, Iraq

zainab.alhameedawi@gmail.com and muayadkrook@yahoo.com

Abstract- *Robust and accurate indoor localization has been the goal of several research efforts over the past decade. In the building where the GPS is not available, this project can be utilized. Indoor localization based on image matching techniques related to deep learning was achieved in a hard environment. So, if it wanted to raise the precision of indoor classification, the number of image dataset of the indoor environment should be as large as possible to satisfy and cover the underlying area. In this work, a smartphone camera is used to build the image-based dataset of the investigated building. In addition, captured images in real time are taken to be processed with the proposed model as a test set. The proposed indoor localization includes two phases the first one is the offline learning phase and the second phase is the online processing phase. In the offline learning phase, here we propose a convolutional neural network (CNN) model that sequences the features of image data from some classis's dataset composed with a smartphone camera. In the online processing phase, an image is taken by the camera of a smartphone in real–time to be tested by the proposed model. The obtained results of the prediction can appoint the expected indoor location. The proposed system has been tested over various experiments and the obtained experimental results show that the accuracy of the prediction is almost 98.0%.*

Index Terms*— localization, CNN, deep learning,* smartphone, image classification.

I. INTRODUCTION

Enormous using of smartphone devices has nursed a great interest in indoor services, such as indoor navigation, robot tracking in factories, locating workers on construction sites, and activity recognition [2]. Places in realistic settings exhibit considerable variability, so to learn how to classify them, it is necessary to use large training sets [7].

Over the past years, different researchers have considered the problems of indoor localization using different techniques. In [1], the authors expected to use pervasive Wi-Fi and attractive field data for cell phone indoor limitation. Exploratory outcomes demonstrated that profound Network models consolidating attractive field and Wi-Fi finger printings improved indoor localization precision. Although the training phase of Deep-Positioning was computationally intensive, the testing phase is fast and suitable for real-time indoor localization on a mobile device. In [2], the authors studied finger printing-based indoor localization in commodity 5-GHzWiFi networks. The proposed system termed BiLoc, which uses bi-modality deep learning for localization in the indoor environment using off-the-shelf Wi-Fi devices. The experimental results validated the superior performance of BiLoc over several benchmark schemes. In[3], they proposed a model that utilizes a profound learning scene acknowledgment technique. This depends on limitation improvement utilizing the technique for exchange learning on inception V3 network. Model feature information is added to assist in scene recognition. In [4], The authors utilized an advanced mobile phone-based movement acknowledgment for indoor restriction utilizing a convolutional neural system, these exercises can be utilized as the tourist spots for indoor localization. These exercises can be utilized as the tourist spots for indoor confinement.

Another convolutional neural system has been intended to become familiar with the best possible highlights consequently.

In[5], the authors designed Wi-Fi deep, which was a Wi-Fi-based indoor fingerprinting localization system that can achieve robust and high accuracy tracking in the presence of device heterogeneity. The system employ model regularization to enable the network to generalize and avoid over-fitting, leading to more robust and stable models. The results show the WI Deep comes with a localization accuracy better than the state-of-the-art systems by at least 53% and 29.8% in the large and small environments respectively.

In this paper, an indoor localization system for a person is proposed based on deep learning technique. The smartphone device is used for getting the images of the location. We have two stages, the offline stage, and the online stage. The offline stage is predominantly for dataset (preparing) development by gathering and pre-handling picture information of pre-built up reference positions. In this stage, profound learning strategy is utilized to separate the core highlights in the picture information and train the learning model. Moreover, the proposed model ought to likewise have bunches of earlier information to make up for every single missed datum. Subsequently getting better limitation execution and lessening the computational intricacy in the online stage are required. In the online stage, the real-time image that the smartphone takes using a camera is used to estimate its position [1], [8] and [10].

In the rest of this article, the importance of indoor localization is discussed in part II, we represent the indoor localization architecture in part III and the results are discussed in part IV. Part V contains the conclusions of this article.

II. IMPORTANCE OF INDOOR LOCALIZATION

Indoor Localization is important in smart building and its applications such as disaster management where localization can likewise help in productive disaster management and speed up the recuperation procedure. One of the key difficulties of calamities is typically acquiring data about people, regardless of whether they are protected or not and what is their area in the disaster influenced zone. Restriction can help in such situations by giving the exact area of the missing people and furnishing them with medicinal help in extraordinary situations, for example, the client is stuck in rubble after the earthquake. another application is security where localization can significantly improve security conditions the world over. Client versatility examples and communication can be utilized to distinguish conceivable dangers that may present security risks. Similarly, in a war zone, the military can follow its advantages through a restriction framework that will improve the general task and increment the odds of fruitful activity. The soldier on the ground can likewise profit by a localization system to explore in regions not known to them. And there is a lot of applications not mentioned here. These applications cannot appear with the outdoor localization systems because it is hard to give the exact details that the indoor localization systems give. [11]

III. PROPOSED SYSTEM

As mentioned earlier, the system runs in two phases, offline training phase and online localization phase as shown in Fig.1. In this section, we take these phases separately with more details. See Fig.2.
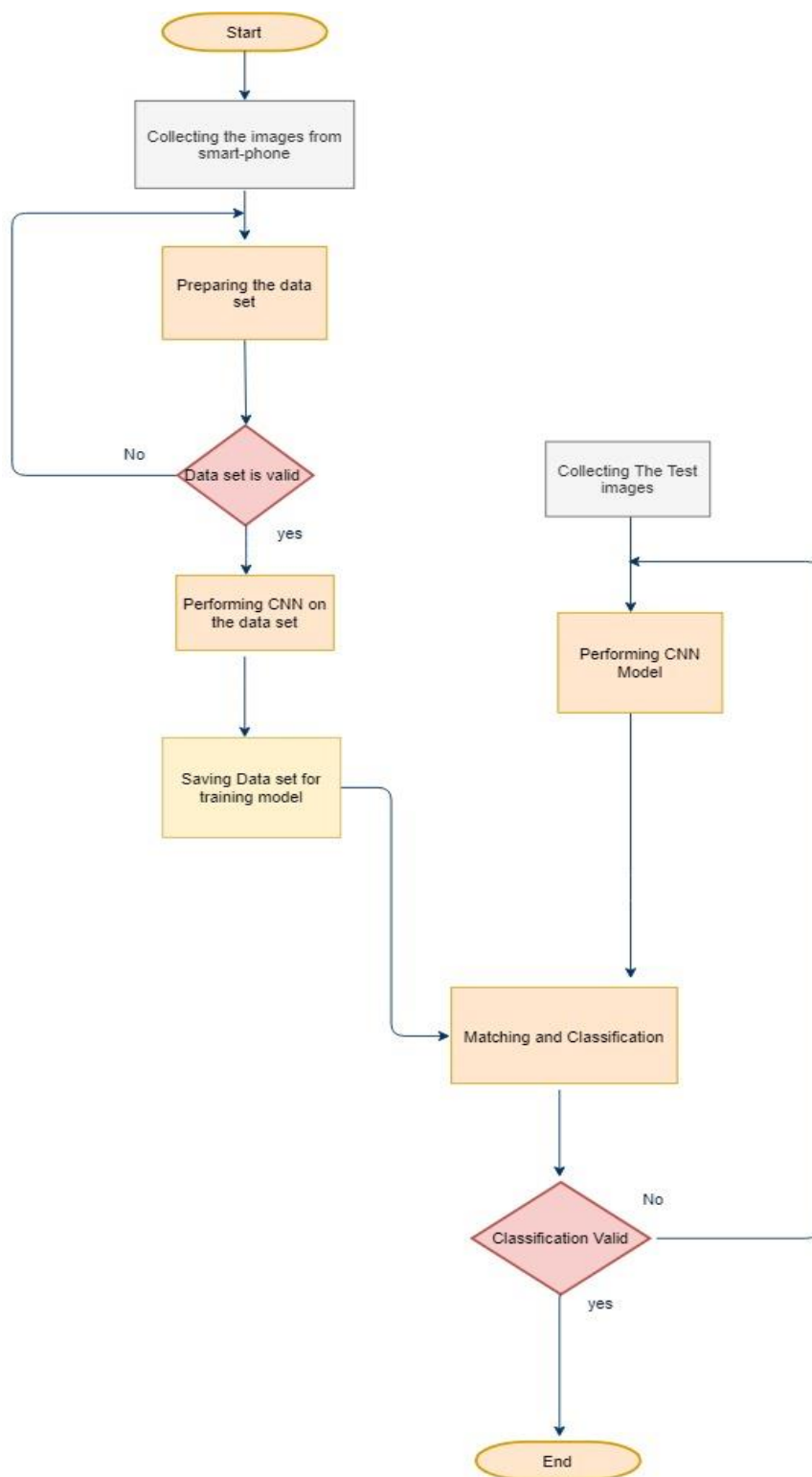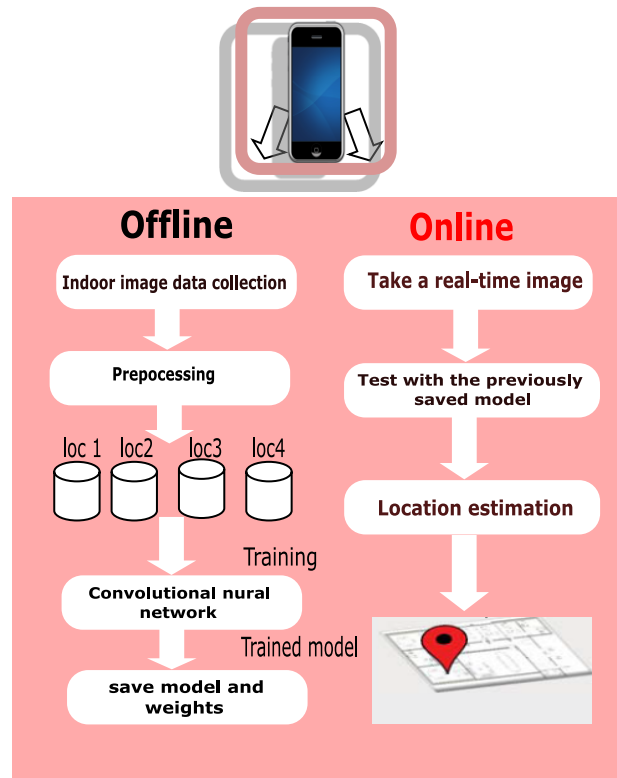
FIG.1: OVERALL SYSTEM FLOWCHART

FIG. 2: PROPOSED SYSTEM ARCHITECTURE.

A. Offline Training Model Phase

During this phase, a deep learning model is developed with Google laboratory which provides a free GPU and thus faster training time. The proposed system is built using Python 3.5 and Kera's and all necessary libraries are added with the use of the convolutional neural network. This helps in scaling the system to cover large areas as well as keeping the model size small and easier to train.

To collect the training data, acceptable resolution images are taken by a smartphone camera and being processed. In this paper, we work with only four locations or classes, each one consists of training and validation data. These data are prepared using the offline phase to be entered to the proposed convolutional neural network. The model is trained and saved as the last step in this phase. This section can be divided into three sub-sections for easing the reading flow.

1. Data Collection and Pre-Processing

Indoor scenes are not quite the same as open-air scenes, which have a wide scope of foundation components with generally stable highlights. Depending on the convolutional neural network and, to get progressively precise spots grouping results, the most noticeable and regular foundation data needed to be removed from the outside landscape. The indoor scene has an extensive number of frontal area protests, and notwithstanding for a similar sort of landscape. This is because of the distinctions in inside improvement, style, and different elements, the highlights are various. Along these lines, it is hard to acquire high acknowledgment precision utilizing the convolutional neural system with constrained commonplace indoor view pictures for training including high randomness and false shooting angle, focus, and darkness as shown in Fig.3.

 Along these lines, it is hard to ensure high precision for indoor confinement when indoor picture data or situating data is given thusly. Data augmentation techniques are used to making artificially more data from real data items by applying invariances as explained in Table 1.

Typical real-time image	Darkness	False shooting angel	Improper focus

FIG.3: COMPARISON BETWEEN NORMAL INDOOR IMAGE AND REAL-TIME IMAGES FROM SMART-PHONE

TABLE 1: OVERVIEW OF DATA AUGMENTATION TECHNIQUES USED

Name	By Factor
Scaling	(1./255)
Shearing	20%
Zooming	20%
Flipping	Horizontal
Crops	224*224

The caught pictures can have a lot of arbitrary commotion, and its highlights are a lot flimsier than those of a run of the typical indoor pictures. This can make it troublesome for convolutional neural systems prepared with common indoor scene pictures to separate conspicuous view highlights from this sort of picture. After collecting the data correctly, it is segmented into a number of classes. In this work, four classes are adopted (as four different locations). Each one consists of 500 images as the training set and 100 images as the validation set. This can result in up to 2000 training images and 500 validation images belonging to four classes. Fig.4 shows examples of the adopted dataset. After this step, it is time to design the proposed convolutional neural network.
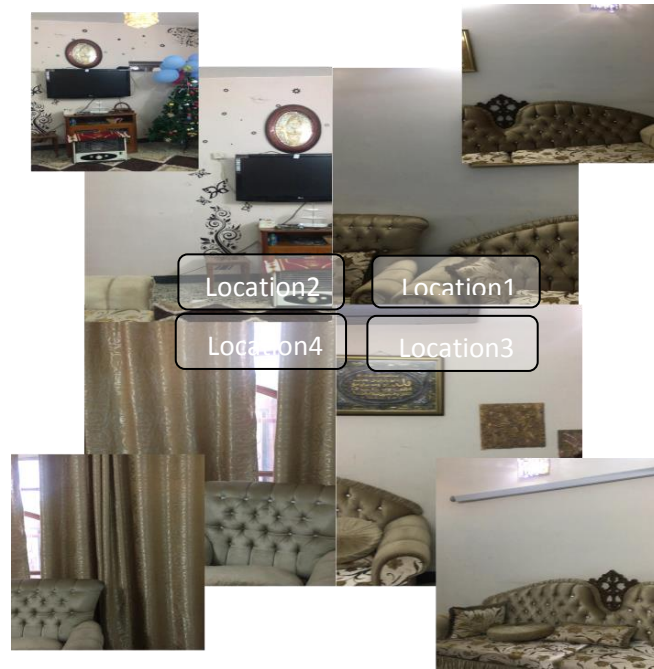
FIG.4: EXAMPLES OF THE DATASET CAPTURED BY A SMART-PHONE

2. Designed Convolutional Neural Network

One of the first applications of convolutional neural networks (CNN) is the network for optical character recognition. Many applications were applied with the CNN like action classification human pose estimation, speech recognition and, face parsing [9]. Just to be known that this report is the first one for indoor localization that contains training images from newly constructed dataset. The structure of the convolutional neural network designed for indoor localization mainly consists of three types of layers: convolutional layer, pooling layer and fully connected layer as illustrated in Fig.5.
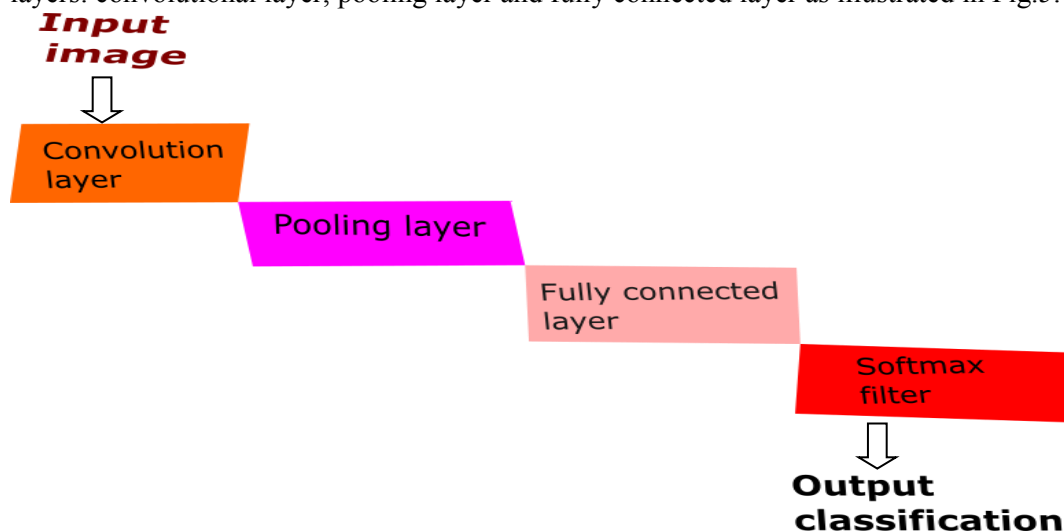

FIG.5: THE USED CNN LAYERS IN OUR PROPOSED MODEL.

All three channels are processed in the proposed network. They are then defined as follows:
- In the first convolutional layer, 96 filters of size 11x11 pixels are applied to the input, followed by a rectified linear operator (ReLU). A max pooling layer is used for taking the maximal value of 5 x 5 regions with strides of 4 and a local response normalization layer.

• The output of the previous layer is then processed by the second convolutional layer. It contains 256 filters size 5 x 5 pixels. Again, that is followed by way of ReLU yet a max pooling with a local response longevity durability normalization tier including the same atypical parameters as like before.

• Third convolutional layer applies a set of 384, the size of the filter are 3 x 3 pixels, with a ReLU and a max pooling layer following it.

• Finally, the 4th convolutional layer operates also by applying a set of 384, the size of the filter is 3 x 3 pixels, with a ReLU and a max pooling layer following it

• 5th and last convolutional layer operates by applying a set of 256, the size of the filter is 3 x 3 pixels, with a ReLU and a max pooling layer following it

Other steps are for the two fully connected layers used for this model:

• The last convolutional layer's output is passed to a fully connected layer and contains 512, neurons, followed by a ReLU and a dropout layer.

• The previous layer's output delivered to the input of the second fully connected layer that also contains 512 neurons, followed by a ReLU and a dropout layer.

• At last, the yield of the last completely associated layer is nourished to a SoftMax layer that appoints a likelihood for each class. The forecast itself is made by taking the class with the maximal likelihood for the given test picture. The results of training the model show that the accuracy is 98.0% as shown in Fig.6.
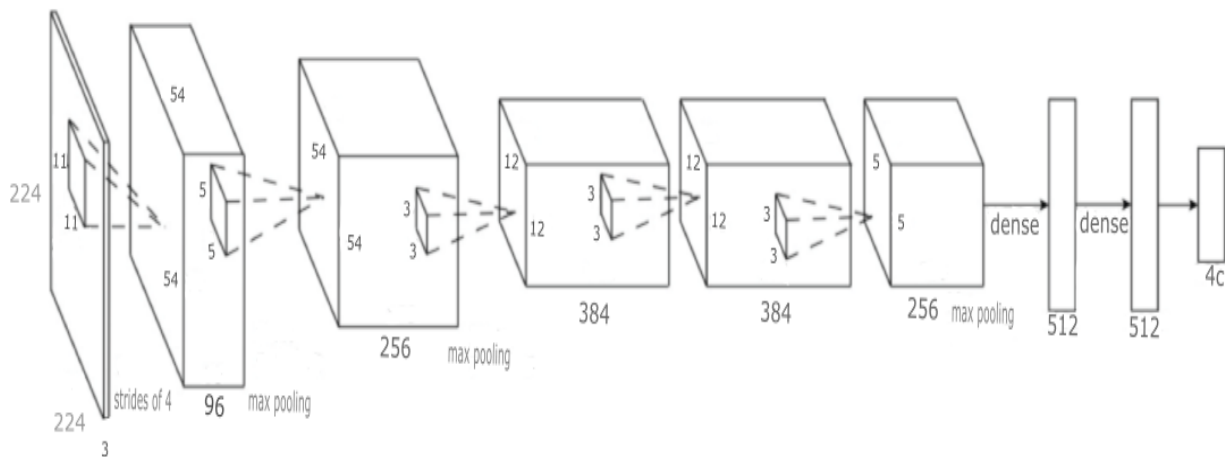

FIG.6: CNN ARCHITECTURE OF THE PROPOSED MODEL.

3. Network Initializing and Training

The network is trained, from scratch, because we do not use the pre-trained model for initializing the network, without using any data outside of the images available with a smart-phone. The CNN implementations are used for face recognition with a minimum number of images in comparison with the work of [6] that had hundreds of thousands of images used for training. Aside from the use of a learning network architecture, the small misalignments in the Training images may cause a huge impact on the accuracy of our results. Therefore, there is a method called oversampling was designed to compensate for these misalignments bypassing the need for progressing the quality. Moreover, two additional methods to further limit the risk of overfitting. First one is dropout learning. The network consists of two dropout layers and the ratio is 0.5. The second one is data augmentation, performed by taking a crop of 224 x 224 pixels of the input images. The image patch size used for the training dataset was 16 images .and the learning rate was 0.001.

B. Online Testing Phase

After the model has been trained and saved in the storage of the offline phase, it is easy to find the location of a real-time image taken by a smart-phone camera. This is done with notifying the stability of the hand and the angle of the image capturing for best results of the estimation model. When a person takes an image, the smartphone's specifications of the camera should be close for both phases. This is to avoid mismatching errors, image mismatch may result in an imbalance in the results of the model and thus give wrong estimation results. The image is processed and entered to the model to be tested. The estimation of the indoor location image can be discovered by the output results of the model, the test image should be resized to fit the sizes in the training model, after scaling, it passed to the estimation model, the results of the estimation supposed to give the expected location.

IV. RESULTS

As mentioned before, small misalignments in the training images can also have a noticeable impact on the quality of our results. The accuracy of the trained model was 98.0%, this means that not all of the tested location images give a true location expectation. Fig.7 shows a set of images that the prediction results of it were wrongly expected. While Fig.8 shows a set of locations that were truly expected. The expected location with the reallocation of the taken images is explained in Table 2 and Table 3. It is shown that the accuracy ratio comes from some error appeared. These tables illustrate an example of huge testing experiments.

FIG.7: EXAMPLE OF IMAGES TAKEN WITH A SMARTPHONE WHICH WERE MISTAKENLY CLASSIFIED AT WRONG LOCATIONS.

FIG.8: EXAMPLE OF IMAGES TAKEN WITH A SMARTPHONE WHICH WERE CLASSIFIED AS ITS ACTUAL LOCATIONS.

TABLE 2: EXPLANATION OF THE TESTED IMAGES IN FIG.4

Image	Estimated location	Actual location
1	Location2	Location1
2	Location1	Location4
3	Location1	Location3

TABLE 3: EXPLANATION OF THE TESTED IMAGES IN FIG.5

Image	Estimated location	Actual location
1	Location2	Location2
2	Location1	Location1
3	Location3	Location3
4	Location 4	Location4

TABLE 4: COMPARISON BETWEEN THE RESULTS OF OUR CLASSIFICATION MODEL AND THE MODEL OF [9].

Training accuracy of the proposed system	Training accuracy of previous work of [9]
98%	**86%**

V. CONCLUSION

In this paper, an indoor localization system based on deep learning technology was proposed. The results of the model with deep learning architecture designed to avoid over-fitting. The proposed CNN network has low complexity if we try to compare it with some other architectures that come newly.so in that way we are trying to decrease the number of parameters and the possibility of over-fitting. The proposed system was trained on the Google Collaborator with images taken by a smart-phone. The obtained results showed a notable improvement to accuracy up to 98% in comparison with previous work of [9].

REFERENCES

[1] Wei Zhang, Rahul Sengupta, John Fodero and Xiaolin Li, "Deep Positioning: Intelligent Fusion of Pervasive Magnetic Field and WiFi Fingerprinting for Smartphone Indoor Localization via Deep   Learning". 16th IEEE International Conference on Machine Learning and Applications, 2017.

[2] Xuyu Wang, Lingjun Gao And Shiwen Mao," BiLoc: Bi-Modal Deep Learning for Indoor Localization With Commodity 5GHz WiFi",2017.

[3] Wei Guo, Ran Wu, Yanhua Chen and Xinyan Zhu, "Deep Learning Scene Recognition Method Based on Localization Enhancement",2018.

[4] Moustafa Abbas, Moustafa Elhamshary, Hamada Rizk, Moustafa Youssef, and Marwan Torki, "WiDeep: WiFi-based Accurate and Robust Indoor Localization System using Deep Learning". January 2019.

[5] Baoding Zhou, Jun Yang, and Qingquan Li, "Smartphone-Based Activity Recognition for Indoor Localization Using a Convolutional Neural Network",2019.

[6] Y. Sun, X. Wang, and X. Tang."Deep learning face representation from predicting 10,000 classes. In Proc. Conf. Comput. Vision Pattern Recognition", 2014.

[7] Alex Krizhevsky , Ilya Sutskever, and Geoffrey E. Hinton, "Image Net Classification with Deep Convolutional Neural Networks",2013.

[8] J. Haverinen and A. Kemppainen, "Global indoor self-localization based on the ambient magnetic field," Robot. Auton. Syst., vol. 57, Oct. 2009.

[9] Gil Levi and Tal Hassner, " Age and Gender Classification using Convolutional Neural Networks", 2015.

[10] J. Xiong, K. Sundaresan, and K. Jamieson, "Tone Track: Leveraging frequency-agile radios for time-based indoor wireless localization," in Proc. of the 21st Annual International Conference on Mobile Computing and Networking, Sep. 2015.

[11] Faheem Zafari, Athanasios Gkelias, Senior Member, Fellow, IEEE. " A Survey of Indoor Localization Systems and Technologies, Jan. 2019