




Data Pre-processing Based on Convolutional Neural Network for Improving Precision of Indoor Positioning

Eric Hsueh-Chan Lu^(✉) , Kuei-Hua Chang, and Jing-Mei Ciou

Department of Geomatics, National Cheng Kung University, No. 1, University Road,
Tainan City 701, Taiwan (R.O.C.)

luhc@mail.ncku.edu.tw, goeatsmall@gmail.com,
s0918038832@gmail.com

Abstract. In the past, indoor positioning technology was mainly based on pedestrian dead reckoning and wireless signal positioning methods, but it was easy to cause some problems such as error accumulation and signal interference. Positioning accuracy still needs to be improved. With the development of neural networks in recent years, many researchers have successfully applied the neural network to the indoor positioning problem based on the Convolutional Neural Network (CNN). This technique mainly determines the position of the image by matching the image features. CNN faces the same challenges as other supervised learning. If the “clean” data cannot be collected, the trained model will not achieve good positioning accuracy. For CNN used for indoor positioning, if someone passes through in the training data, causing the person to appear in different positions of the images, the model may think that the images are the same location. To solve this problem, we propose a data pre-processing method to improve the accuracy of indoor positioning based on CNN. In this method, the moving objects recognized in training and testing data are modified in different ways. We perform data pre-processing method based on Mask R-CNN and YOLO, and then integrate the pre-processing method to PoseNet the famous CNN indoor positioning architecture. Through real experimental analysis, removing moving objects can effectively improve indoor positioning accuracy about 46%.

Keywords: Convolutional Neural Network · Indoor positioning · Object detection · Data pre-processing

1 Introduction

The development of spatial information technology has become more and more rapid in recent years, and the research and application on Location-Based Services (LBSs) [10] and route planning [5] related to positioning have attracted more and more attention. In the outdoor environment, people can obtain accurate surface location information through GPS and GNSS, and many areas also generate more convenient and appropriate services due to the emergence of global satellite positioning systems, such as satellite

navigation systems, smart parking systems, various geodesy, etc. Although the global satellite positioning system brings convenience to people's livelihood applications, when satellite signals are blocked, the positioning application will also fail. For example, indoor or basement is the most vulnerable to missing signals. Once the signal is blocked, GPS will not be able to continue to provide location services. Therefore, how to continuously calculate the position after the satellite signal fails, making indoor positioning technology a hot spot. Research topics. Indoor positioning is widely used and has high commercial value. Common application fields include route guidance of stations, AR interaction of art museums, smart navigation of department stores, and cargo monitoring of factories. Therefore, more and more scholars explore indoor positioning.

In the past, indoor positioning technology can be mainly divided into Pedestrian Dead Reckoning (PDR) [4] and wireless signals [2, 8], but it is easy to cause problems such as error accumulation and signal interference, and there is still much room for improvement in positioning accuracy. With the development of neural networks in recent years, in addition to predicting data and identifying images, many researchers have successfully applied Convolutional Neural Network (CNN) to indoor positioning technology. It is regarded as the concept of the human eye, analyzes the environmental characteristics in the image, and judges the position of the photographer itself by matching the feature values. CNN faces the same challenges as other supervised learning. If a large amount of "clean" data cannot be collected, the trained model will not achieve the desired positioning accuracy. It is easier to collect a large amount of data, but it cannot ensure that there is no interference in the data. Therefore, it is very important to pre-process the data.

This paper makes effort on precision improvement of indoor positioning and is intended to solve the situation of interference in CNN training data. If there is a passerby when we collect the training data, the passerby will be repeatedly displayed in images in different positions. So model may consider these images of passerby as the same position. Such a model cannot achieve good accuracy. Therefore, we propose a method for pre-processing data to improve the accuracy of CNN indoor positioning. The moving objects detected in the training and testing data are modified in different ways. In the experiment, we are based on Mask R-CNN [7] and YOLO [6] two kinds of CNN networks for data pre-processing and integration into the famous CNN indoor positioning architecture – PoseNet [1, 3] improved positioning accuracy.

The remaining of this research is organized as follows. Section 2 reviews related work on indoor positioning and CNN issues. In Sect. 3, we explain the details of data pre-processing and CNN-based indoor positioning models. The experimental evaluations are shown in Sect. 4. Finally, the conclusions and future work are mentioned in Sect. 5.

2 Related Work

In this section, we review some important studies related to indoor positioning issues. To collect available information for indoor positioning, Lan *et al.* proposed Pedestrian Dead Reckoning (PDR) [4] based on IMU technology to track the user's trajectory and detect when the user leaves the parking space so that the next user can use the mobile phone to get the location service for the inquiring parking space. Wireless signal positioning is

another indoor positioning solution. The common signal type includes infrared, WiFi and Bluetooth. The calculation method includes proximity positioning, intersection method and feature comparison. Grossmann *et al.* set up a wireless network in the museum's exhibition hall [2]. The Access Point (AP) of the road, and uses the Received Signal Strength Index (RSI) to obtain location information. Subhan *et al.* based on the use of feature matching in Bluetooth to perform position estimation [8]. The accuracy of the indoor positioning system depends largely on the parameter values of the comparison and the measurement results of the surrounding environment.

Convolution Neural Network (CNN) is an effective recognition algorithm that has been widely used in image recognition, object detection and localization in recent years. To accurately estimate the attitude of a monocular camera, Kendall *et al.* proposed the CNN model PoseNet [1, 3] for regression pose estimation. PoseNet is a CNN indoor positioning architecture for regression pose estimation. Its network architecture is based on GoogleNet [9] proposed by Szegedy *et al.* The input is a color image and the output is changed to a seven-dimensional pose vector. The paper indicates that PoseNet can be used for both outdoor and indoor positioning. Ren *et al.* proposed Mask R-CNN [7] that is to mark objects identified in the image as masks close to the pixel level. In the past, the practice of masking was called RoIPooling. When the value was processed, the nearest interpolation method was used, and the output pixel value was the nearest pixel value, so that the resulting mask would be offset. The obtained area size is not an integer, and the mask after taking the integer cannot reach the pixel level. Therefore, the Mask R-CNN changes the RoIPooling to use the bilinear interpolation method, and performs linear interpolation in two directions. The output pixel value is a weighted average of the surrounding pixel intensities, the pixel values are relatively continuous, and the mask position is more precise, and this method is also referred to as RoIAlign. In Mask R-CNN, the previous numerical processing method was changed and the pixel-level mask was achieved. YOLO [6] published by Facebook AI Research (FAIR) emphasizes not the pixel-level mask, but the actual recognition speed. The main method is to produce $S \times S$ squares, each square predicts the confidence score and type of the contained object by itself, and finally outputs the highest scored square. YOLO designed the network as end-to-end, which not only makes training easier, but also speeds up overall.

3 Proposed Method

In this section, we introduce our proposed indoor positioning method that can be divided into two phases including data pre-processing and CNN-based indoor positioning model.

3.1 Data Pre-processing

Training data are very important for supervised learning. The deep learning prediction model used for indoor positioning requires "clean" data. If there is a passerby when we collect the training data, the passerby will repeat in images in different positions. So model may consider these images of passerby as the same position. Such a model cannot achieve good accuracy. Therefore, we try to identify the moving object in the training and testing data and modify it in different ways for data pre-processing.

Data pre-processing is divided into two steps. The first step is to detect the object in the image. The detected object is a moving object. The techniques used in this step are YOLO and Mask R-CNN. The difference is that YOLO marks the object in a square, and the Mask R-CNN is marked according to the shape of the object. Then we paint the marked area. As shown in Fig. 1, using YOLO and Mask R-CNN to change the moving object to white, test whether the modified moving object will increase the accuracy of positioning. Since we found that the results of Mask R-CNN could not completely cover moving objects. We increased the marked area by ourselves, then test whether the accuracy was affected.



Fig. 1. Paint the marked area to white.

The second step is to use different strategies to paint the marked area of detected moving objects. The difference in strategy is mainly the difference in color. Since white is a subjective setting, it is assumed that white should be replaced with the most color in the image, as shown in Fig. 2, and whether the altered color affects the accuracy.

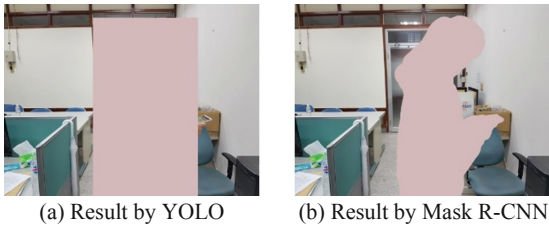


Fig. 2. Paint the marked area to the most color in the image.

We also tried to adjust the color of the painting to the average color of the image. From the experiment in the previous step, we found that if the moving object occupies too much area of the image, it would cause the painting color to approach the color of the moving object. Therefore, we changed it by calculating the average of all the colors in the image without the moving object, as shown in Fig. 3.

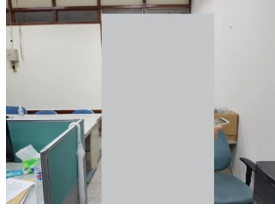


Fig. 3. Paint the marked area to the average color in the image.

3.2 CNN-Based Indoor Positioning Model

This article refers to the PoseNet architecture and makes some minor adjustments, using CNN to directly estimate the camera position from the image training model. Figure 4 shows the PoseNet adjusted by GoogLeNet-based architecture. The main adjustments are in two parts:

1. Replaces the three multi-classifiers with a regression amp, and each final fully-connected layer outputs a seven-dimensional pose containing the three-dimensional position and the quaternion.
2. Before the final affine regression, insert a fully connected layer with a feature size of 2048 to form a 23-layer PoseNet architecture. This is to generate a positioning vector that can be explored by PoseNet.

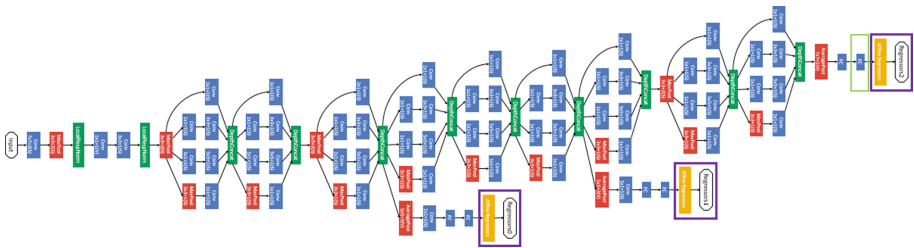


Fig. 4. The network architecture of PoseNet.

We also made some adjustments to the PoseNet architecture. The PoseNet model finally outputs the location and direction information, but the direction is not needed for our goal. Our ultimate goal is to let the user use the mobile phone for image positioning. The focus is on positional accuracy. The mobile phone itself is equipped with an accelerometer, gyroscope and magnetometer triaxial sensor. We only need to grab the information of the gyroscope in the mobile phone to know the direction angle, so we don't need to use CNN to estimate the direction. Thus, the loss function also needs to be adjusted together. The loss function is as shown in Formula (1), where \hat{P} and P are the position prediction value and the position true value, respectively. In the classification problem, each output tag contains at least one training sample, but for regression

problems, the output tags are mostly continuous or infinite.

$$\text{loss(I)} = \widehat{P} - P \quad (1)$$

4 Experimental Evaluation

To evaluate the performance of indoor positioning by integrating data pre-processing and CNN model, a series of experiment are conducted by using the real data. All the experiments in this thesis are carried out on the Tensorflow platform. The hardware device used is Geforce GTX 1080 TI, the operating system is Linux Ubuntu 16.04, and the image source is Samsung Galaxy S7 edge.

4.1 Experimental Data and Setting

The experimental field used in this thesis is our laboratory, as shown in Fig. 5(a). In the experiment, we define people as moving objects, Fig. 5(b) shows one of the image samples. There are total 48 images for training, 11 have been painted. And total 24 images for testing, 8 have been painted.

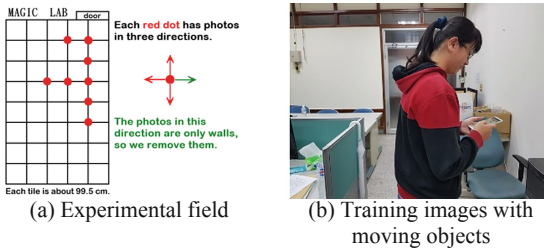


Fig. 5. Experimental field and data.

4.2 Experimental Results

In the experiment, the images were altered in different ways and colors, so we combined different training and testing data to test the positioning error. The total experimental results are shown in Table 1 (unit: meter). If you use an image without any pre-processing during training, the average error of positioning will be nearly 4 m, and using YOLO and Mask R-CNN to change the moving object to white can improve the error, and YOLO is the most effective. Ok, it can improve nearly 1 m. Since the results of the Mask R-CNN coating did not completely cover the moving object, the experiment was manually added to the modification range, and the experimental results were still not as good as the YOLO correction. The reason may be because YOLO's modification method is a complete box, and Mask R-CNN's method of retouching preserves the contour of the object so that CNN still regards it as an important feature value.

Table 1. Comparison of the effects of different data pretreatment methods. Note that the parentheses represent the colors we used to paint on the moving object. ‘W’ denotes white; ‘M’ denotes the most color in the image; ‘A’ denotes the average color of the image; and ‘Mask+’ denotes we increased the marked area of Mask R-CNN.

Testing	Training						
	Original image	YOLO (W)	Mask (W)	Mask+ (W)	Mask+ (M)	YOLO (M)	YOLO (A)
Original image	3.8039	2.9908	3.2976	3.0704	2.8247	2.8221	2.0480
YOLO (W)	4.0789	2.8166	2.5835	3.1854	3.3238	2.5716	2.0548
Mask (W)	3.6532	2.7910	3.0055	3.2715	2.9972	2.6301	2.0241
Mask+ (W)	3.5441	2.6780	2.9916	3.2618	3.1098	2.6438	2.0993
Mask+ (M)	3.5024	3.0143	3.1918	3.2517	3.1849	2.5623	2.0638
YOLO (M)	3.5914	2.7731	2.6746	3.2002	3.4788	2.5586	1.9683
YOLO (A)	3.5750	2.7714	2.6850	3.1673	3.4581	2.5770	1.9799

Painting a moving object to white is personally subjective and unfounded, so the painted color is adjusted to the most frequently appearing color in the image, but in some images the moving object occupies too much area and the resulting color is approach to the object. Therefore, such color did not greatly improve the positioning error. Finally, the color calculation method is adjusted to the average color in the image except the moving object. It can be seen from the results that use average color to pre-process training data can reduce the positioning error in different testing data and can reduce the error of nearly 2 m compared with the data without any pre-processing. From all the above experimental results, it can be seen that regardless of the pre-processing method of the training and testing data, the positioning accuracy can be better than that of the original data. The best result is to use YOLO to detect moving objects and then paint them according to the average color of the image.

5 Conclusion and Future Work

This paper verifies the influence of moving objects on Convolutional Neural Network (CNN) indoor positioning accuracy through experiments. It also proposes to use Mack R-CNN and YOLO to pre-process the data with different strategies. The results show that the modified moving objects can effectively improve the positioning accuracy, regardless of what kind of strategy can improve accuracy. In addition to the people we have modified

in the experiment, similar indoor positioning methods also mentioned that the change of furniture position will seriously affect the positioning accuracy. In the future work, we will consider the furniture that change position very often as moving objects too, or try some different ways of data pre-processing. In addition, we plan to test in a wider indoor field and try to develop a fully automated indoor field data collection mechanism, such as the use of drones or customized devices, to improve the overall efficiency of the experiment.

Acknowledgment. This research was supported by Ministry of Science and Technology, Taiwan, R.O.C. under grant no. MOST 108-2621-M-006-008 -.

References

1. Ciou, J.-M., Lu, E.H.-C.: Indoor positioning using convolution neural network to regress camera pose, ISPRS Geospatial Week, 1289–1294 (2019)
2. Grossmann, U., Gansemer, S., Suttorp, O.: RSSI-based WLAN indoor positioning used within a digital museum guide. *Int. J. Comput.* **7**(2), 66–72 (2014)
3. Kendall, A., Grimes, M., Cipolla, R.: PoseNet: a convolutional network for real-time 6-DOF camera relocalization. In: *IEEE International Conference on Computer Vision*, pp. 2938–2946 (2015)
4. Lan, K.-C., Shih, W.-Y.: An indoor locationtracking system for smart parking. *Int. J. Parallel Emergent Distrib. Syst.* **29**(3), 215–238 (2014)
5. Lu, E.H.-C., Chen, H.-S., Tseng, V.S.: An efficient framework for multirequest route planning in urban environments. *IEEE Trans. Intell. Transp. Syst.* **18**(4), 869–879 (2017)
6. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 779–788 (2016)
7. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 91–99 (2015)
8. Subhan, F., Hasbullah, H., Rozyyev, A., Bakhsh, S.T.: Indoor positioning in bluetooth networks using fingerprinting and lateration approach. In: *IEEE International Conference on Information Science and Applications*, pp. 1–9 (2011)
9. Szegedy, C., et al.: Going deeper with convolutions. In: *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1–9 (2015)
10. Tseng, V.S., Lu, E.H.-C., Huang, C.-H.: Mining temporal mobile sequential patterns in location-based service environments. In: *IEEE International Conference on Parallel and Distributed Systems*, pp. 1–8 (2007)