An Experimental Study on Improving Accuracy of Location Estimation in Finger Print Using CNN and ResNet

Yu SAKANISHI†, ***Student member***, **Satoru AIKAWA**†a), ***Member***, **Shinichiro YAMAMOTO**† , ***Member and*** **Yuta SAKAI**†, ***Member***

SUMMARY

Recently, indoor navigation system is one of the important technologies. We are studying an indoor location estimation technique using wireless LAN employing a Finger Print scheme. A database (DB) and user data (UD) are measured from wireless LAN radio waves to estimate the location using the Finger Print. A CNN is utilized to compare the UD and DB. However, the CNN estimation accuracy may deteriorate due to the degradation problem caused by the increase in the number of middle layers and the gradient vanishing problem. ResNet is a solution to this problem. In this study, we have evaluated estimation accuracies of CNN and ResNet experimentally.

keywords: *Wireless LAN, Finger Print, CNN, ResNet,*

1. Introduction

This paper scopes a NN (Neural Network) used as a method to compare the DB (Database) with measured UD (User Data) for Finger Print indoor Wi-Fi localization. In particular, CNN (Convolutional Neural Network) that shows excellent characteristics for feature extraction of image among NNs. This paper employs the CNN for the Finger Print localization. In advance, DB created by maps of measured RSSI at all of the preset coordinates train the CNN. The user creates an RSSI map from the measured UD. The map is input to the CNN to estimate the position. The accuracy of the CNN improves as the number of layers in the middle layer increases. However, if the number of layers increases too much, the estimation accuracy may decrease due to the degradation problem or the gradient loss problem. ResNet is used in order to solve these problems.

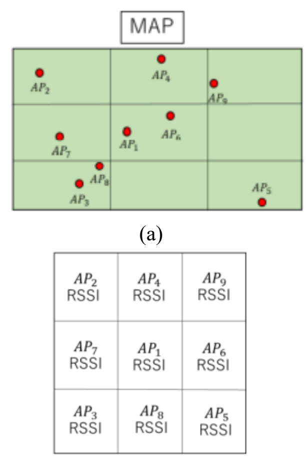
Experiments of the CNN and the ResNet were conducted in a building and an underground mall to evaluate their effects. The results show that ResNet improves the accuracy of the estimation. Furthermore, the optimal configuration and number of layers are obtained.

However, since the computational complexity increases according to the number of layers, CNN may be sufficient for the estimation if the measurement environment is relatively simple.

2. Principle

2. 1 Finger Print position estimation using CNN

CNN is used in the field of image recognition because it shows excellent characteristics to extract two-dimensional features. In this study, since CNN is used for Finger Print position estimation. Wi-Fi RSSI are treated as two-dimensionally arranged data on a map. The RSSI data map associated with the actual position of the AP is input to the CNN. For this purpose, identifiers of all the AP observed in the area are collected in the preliminary measurement. The positions of the AP are estimated [1] on the map using a method described in the next section (Fig. 1(a)). The data are arranged in a square array on the map as shown in Fig. 1(b). The input data array should be square, whose elements number exceeds observed AP.



(a)

(b)

Fig. 1 Sorting data using adjacencies between APs

2. 2 AP inverse position estimation

As mentioned above, it is necessary to estimate the AP positions. The input data to the CNN are arranged in relation to the AP positions. The AP are localized by Lateration [2] using AP information as shown in Figure 2. AP information consisting of MAC addresses and RSSI are collected in advance to create the DB[1] . The distance d between the coordinate and the AP is shown in Equation (1) where N is attenuation constant, f is frequency.

$$d = 10^{\left(-\frac{RSSI}{N}-\frac{20}{N}\log f+\frac{148}{N}\right)} \quad (1)$$

The position of the AP is calculated using the distance d from more than three coordinates where the AP was observed.

---

†The author is with University of Hyogo
a)aikawa@eng.u-hyogo.ac.jp

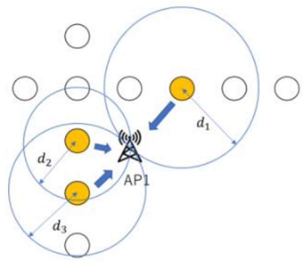


Fig. 2 AP inverse position estimation

2. 3 ResNet

ResNet (Residual Network) is used in order to solve the degradation problem or the gradient loss problem of the CNN. Figure 3 shows the models of conventional CNN and ResNet.

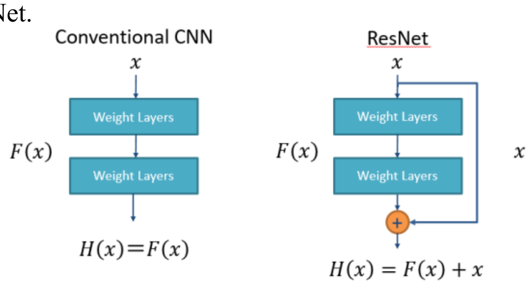


Fig. 3 CNN and ResNet

A feature of ResNet is a skip structure. When a certain level of accuracy is obtained, the input and the output are almost the same, i.e., an isoperimetric transformation. In case of a large number of layers, the burden becomes larger to obtain better accuracy while maintaining the constant transformation relationship. This decrease in accuracy is the cause of the degradation problem. The ResNet solves the degradation problem by preparing a skip structure as a constant transformation path to reduce the burden. In the conventional structure, as the layer get deeper, the gradient converges to a very small value at the nodes close to the input layer. However, since the input data flows the skip structure as it is, the gradient also flows from the upstream to the downstream as it is during back propagation. Therefore, meaningful gradients can be transmitted to the previous layer, and the gradient loss problem can be solved [3]. The disadvantage of ResNet is that the computation time increases as the number of layers increases.

Table 1 and Fig. 4 show the structure of the ResNet tested in this experiment. The 18 layers consist of two residual units that output 64 feature maps (filters), two residual units that output 128 feature maps, two residual units that output 256 feature maps, two residual units that output 512 feature maps, the first convolution and the last pooling layer. The first convolution and the last pooling layer are combined. As shown in Fig. 4, the network was constructed in the order of Batch Normalization, ReLU function, and convolutional layer.

Table1 Configuration of ResNet

層数	18	34	50	74	101	152
Conv1	7×7, 64, stride2					
Conv2	3×3, max pool, stride2					
	$\begin{bmatrix}3\times3, & 64\\3\times3, & 64\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, & 64\\3\times3, & 64\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 64\\3\times3, & 64\\1\times1, & 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 64\\3\times3, & 64\\1\times1, & 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 64\\3\times3, & 64\\1\times1, & 256\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 64\\3\times3, & 64\\1\times1, & 256\end{bmatrix}\times3$
Conv3	$\begin{bmatrix}3\times3, & 128\\3\times3, & 128\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, & 128\\3\times3, & 128\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, & 128\\3\times3, & 128\\1\times1, & 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, & 128\\3\times3, & 128\\1\times1, & 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, & 128\\3\times3, & 128\\1\times1, & 512\end{bmatrix}\times4$	$\begin{bmatrix}1\times1, & 128\\3\times3, & 128\\1\times1, & 512\end{bmatrix}\times8$
Conv4	$\begin{bmatrix}3\times3, & 256\\3\times3, & 256\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, & 256\\3\times3, & 256\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, & 256\\3\times3, & 256\\1\times1, & 1024\end{bmatrix}\times6$	$\begin{bmatrix}1\times1, & 256\\3\times3, & 256\\1\times1, & 1024\end{bmatrix}\times14$	$\begin{bmatrix}1\times1, & 256\\3\times3, & 256\\1\times1, & 1024\end{bmatrix}\times23$	$\begin{bmatrix}1\times1, & 256\\3\times3, & 256\\1\times1, & 1024\end{bmatrix}\times36$
Conv5	$\begin{bmatrix}3\times3, & 512\\3\times3, & 512\end{bmatrix}\times2$	$\begin{bmatrix}3\times3, & 512\\3\times3, & 512\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 512\\3\times3, & 512\\1\times1, & 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 512\\3\times3, & 512\\1\times1, & 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 512\\3\times3, & 512\\1\times1, & 2048\end{bmatrix}\times3$	$\begin{bmatrix}1\times1, & 512\\3\times3, & 512\\1\times1, & 2048\end{bmatrix}\times3$
-	Average pool, Softmax					

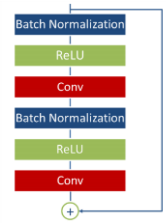


Fig. 4 ResNet Network Configuration

3. Verification

The experimental environments are a relatively simple building and a complex underground shopping mall. The results of these two experiments are compared because the more complex environment can benefit from an increase in the number of the CNN and the RsNet layers.

3. 1 Comparison of estimation accuracy between CNN and ResNet (in a building)

The effect of the ResNet in the problem of accuracy degradations with the large number of layers in a simple building. The verification environment is shown in Fig. 5 and Table 2.

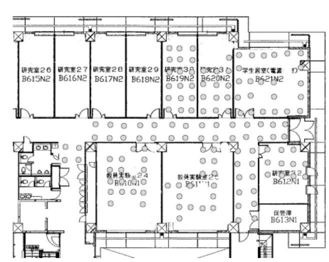


Fig. 5 Verification environment diagram

Table 2 Verification environment

Verification Location	6F, Building B
Measurement date	July 2020
Measurement coordinates	144 Coordinates
Coordinate interval	1m
Measurement terminal	Nexus 9
Number of measurements	Training data 144 times Test data 36 times
Number of layers verified	18, 34,50,74,101,152
Calculation environment	google colaboratory
Number of observed APs	169

3. 2 Comparison of estimation accuracy between CNN and ResNet (underground mall)

Furthermore, the effectiveness of ResNet in an underground mall was investigated. The test environment is shown in Fig. 6 and Table 3.

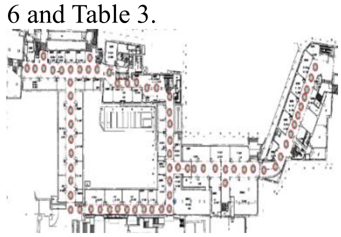


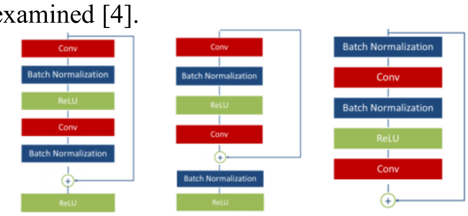
Fig. 6 Verification environment diagram

Table 3 Verification environment

Verification Location	Underground shopping mall
Measurement date	September 2018
Measurement coordinates	68 Coordinates
Coordinate interval	5m
Measurement terminal	Nexus 9
Number of measurements	Training data 144 times Test data 36 times
Number of layers verified	18, 34, 50, 74, 101, 152
Calculation environment	google colaboratory
Number of observed APs	509

3. 3 Optimizing ResNet network configuration

In this experiment, we examined the optimization of the ResNet network configuration. In addition to the basic configuration shown in Fig.4, configurations shown in Fig. 7 also examined [4].



(a)CBRCB⊕R (b)CBRC⊕BR (c)BBCBRC⊕

Fig. 7 Network configuration to consider

"C" denotes the convolutional layer, "B" denotes Batch Normalization, "R" denotes the ReLU function, and "⊕" denotes the location of the shortcut connection. For example, refer to the network structure in Figure 7(a) is referred as CBRCB⊕R.

4. Result

4. 1 Comparison of estimation accuracy between CNN and ResNet (in a building)

Fig. 8 and Fig. 9 show the comparison between the results using CNN and ResNet for location estimation in a building. The horizontal axis shows the number of layers, the vertical axis shows the average error in Figure 8. Figure 9 shows the computation time per epoch.

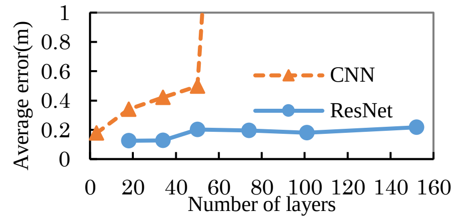


Fig. 8 Average error

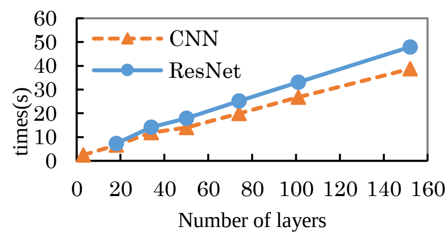


Fig. 9 Calculation time per epoch

From Fig.8, it is confirmed that the best accuracy is obtained with 18 layers of ResNet. Fig.8 also shows that accuracy of ResNet is better than the conventional CNN. However, the accuracy does not improve as the number of layers is increased. Figure 9 also shows that the computation time increases as the number of layers increases. Considering the above accuracy and the computation time, we believe that three CNN layers are enough to obtain good accuracy in a relatively simple environment such as Fig.5.

4. 2 Comparison of estimation accuracy between CNN and ResNet (underground mall)

Fig.10 and Fig.11 show the comparison between the results using the CNN and the ResNet for location estimation in an underground mall.

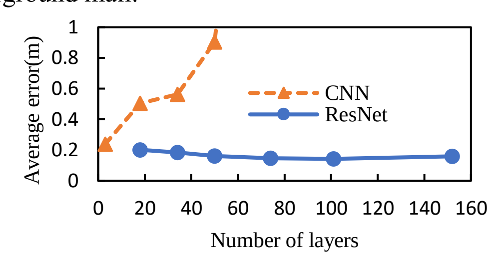


Fig. 10 Average error

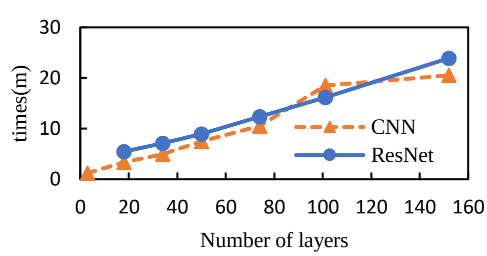


Fig. 11 Calculation time per epoch

It can be confirmed that the best accuracy was obtained with 101 layers of ResNet from Fig.10. It is confirmed that the ResNet accuracy is better than the CNN because the accuracy of the CNN worsens as the number of layers increases. Figure 11 shows that the computation time increases as the number of layers increases.

The best accuracy was obtained with 101 layers that is better than others insignificantly as show in Fig10. However, the computation time is proportional to the number of layers. Therefore, the number of layers of 50 is considered to be sufficient for the environment shown in Figure 6.

4. 3 Differences in estimation accuracy in different environments

Next, Fig.12 shows a comparison of the results estimated by ResNet in the building environment shown in Fig.5 and the underground mall environment shown in Fig.6.

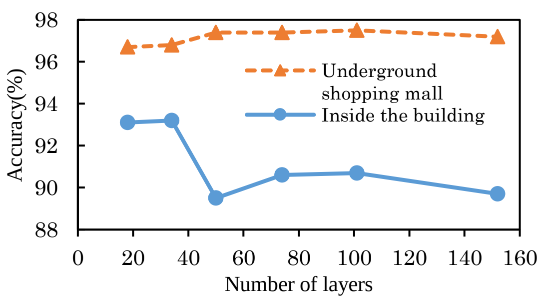


Fig. 12 Differences in estimation accuracy between different environments

Figure12, shows that the accuracy decreased in the increase of the number of layers for the building. On th othed hand, a slight increase was observed in the underground mall.

This can be attributed to the simplicity of the building shown in Fig. 5. The simplicity depends on the structure of the building, the amount of pedestrian traffic, and the number of APs.

4. 4 Optimizing ResNet network configuration

Figure 13 shows the results of location estimation in an underground mall for the four network configurations shown in Figures 4 and 7. The verification environment for this location estimation is the same as in Section 3.2.

From the figure, BCBRC⊕ is the best and CBRC⊕BR is the worst. As for CBRC⊕BR, it is thought that the reason is that when Batch Normalization is inserted after the shortcut connection. This layer changes the information of the shortcut connection and prevents the transmission of information. The accuracy of BCBRC⊕ and BRCBRC⊕ was particularly good, because the role of regularization was strengthened by putting Batch Normalization in front of them.

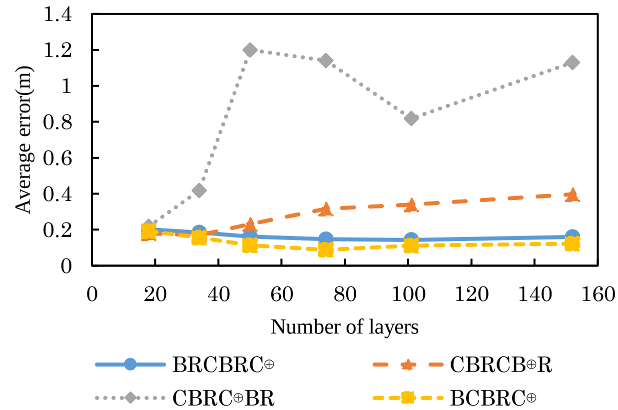


Fig. 13 Verification results of network configuration

5. Conclusion

This paper studies CNN and ResNet based on Finger Print using wireless LAN as indoor location estimation. A map of RSSI is created using the location of the AP estimated. Experiment results conducted in a building and an underground mall were compared.

In conclusion, we confirmed that the accuracy of the ResNet is slightly better than that of the CNN. However, considering the computation time, the CNN can also provide sufficient accuracy. In addition, since the optimal number of layers depends on the environment, ResNet is expected to be effective in some situations.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 21K04065.

References

[1] Satoru.Aikawa, et al., “WLAN Finger Print Localization using Deep Learning,” IEEE APCAP Aug. 2018. DOI:10.1109/ APCAP. 2018. 8538306.

[2] Seigo Ito, Nobuo Kawaguchi, “Wireless LAN Based Hybrid Positioning System Using Bayesian Inference and Access Point Selection”, IEEJ Trans. EIS, Vol. 126, No. 10, 2006.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition”, arXiv:1512.03385v1 [cs.CV], 10 Dec 2015.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Identity Mappings in Deep Residual Networks”, arXiv:1603.05027v3 [cs.CV], 25 July 2016.