Vision-based Indoor Localization Algorithm using Improved ResNet

Zeyad Farisi[1], Tian Lianfang[2], Li Xiangyang[3], Zhu Bin[4]
School of Automation Science and Engineering, South China University of Technology, Guangzhou, China[1, 3]
College of Community Service Department of Engineering and Science. Tabah University, Medinah, Saudi Arabia[1]
School of Automation Science and Engineering, South China University of Technology[2]
Research Institute of Modern Industrial Innovation, South China University of Technology[2]
Key Laboratory of Autonomous Systems and Network Control of Ministry of Education. Guangzhou, China[2]
School of Mechanical and Electronic Engineering, Jiangxi college of applied technology, Ganzhou, China[4]

***Abstract*—The output of the residual network fluctuates greatly with the change of the weight parameters, which greatly affects the performance of the residual network. For dealing with this problem, an improved residual network is proposed. Based on the classical residual network, batch normalization, adaptive -dropout random deactivation function and a new loss function are added into the proposed model. Batch normalization is applied to avoid vanishing/exploding gradients. -dropout is applied to increase the stability of the model, which we select different dropout method adaptively by adjusting parameter. The new loss function is composed by cross entropy loss function and center loss function to enhance the inter class dispersion and intra class aggregation. The proposed model is applied to the indoor positioning of mobile robot in the factory environment. The experimental results show that the algorithm can achieve high indoor positioning accuracy under the premise of small training dataset. In the real-time positioning experiment, the accuracy can reach 95.37.**

***Keywords*—Deep learning; residual network; loss function; dropout; indoor localization**

I. Introduction

With the development of artificial intelligence technology, various types of robots have been widely used. In the application of mobile robots, real-time detecting and monitoring the location of robots is the prerequisite for better service to human beings. For indoor localization assignments, Wi-Fi based method [1], Bluetooth based method [2] and Radio Frequency identification technology [3] were proposed and widely used. However, bottlenecks exist in these methods. Wi-Fi-based methods are vulnerable to multi-path effects, Bluetooth-based methods exist mutually interference and RF-based methods require expensive equipment support. Vision-based methods [4][5] which can realize real-time positioning only by a normal RGB camera, avoid all these bottlenecks mentioned above and provide a new way for indoor positioning.

In recent years, deep learning technology has been greatly developed and widely used in image processing, especially in image-based classification tasks. Compared to many traditional algorithms, deep learning technology, which uses massive training dataset to learn prior knowledge, has stronger generalization ability and more complex parametric expression. Since 2012 [6], Hinton et al put forward the outstanding performance of Alexnet with five convolution layers and three full connection layers in the ImageNet image classification competition. More and more scholars began to study convolution neural network to solve various practical problems. It was found that the accuracy can be improved by increasing the depth of CNN (Convolutional Neural Network). The deeper the network, the more features can be obtained, and the stronger the expression ability of the network. What's more, the deeper the network, the more abstract semantic features can be extracted [7-10]. However, simply increasing the number of layers of neural network will lead to the problems of gradient disappearance, gradient explosion and model degradation. In 2016, He et al proposed a 152 layer Res-Net [11], which the residual structure is used in the deep neural network. Res-Net can solve the degradation problem and the residual structure makes the model easier to optimize, and can get better training results under the premise of smaller training dataset, but the learning results of the network are very sensitive to the fluctuation of the network weight, that is, the slight change of the network weight will cause a greater change of the output. The model would be affected badly by this shortcoming in the process of model training and testing. In [12-16], a serious of improvements have been made to ResNet. But none of them can solve the problem well.

In order to solve the stability problem of the ResNet, an improved residual network is proposed. Based on the classical residual network, batch normalization, adaptive β -dropout random deactivation function and a new loss function are added into the proposed model. Batch normalization is applied to avoid vanishing/exploding gradients. β -dropout is applied to increase the stability of the model, which we select different dropout method adaptively by adjusting parameter β. The new loss function is composed by cross entropy loss function and center loss function to enhance the inter class dispersion and intra class aggregation. The proposed model is applied to the indoor positioning of mobile robot in the factory environment. The experimental results show that the algorithm can achieve high indoor positioning accuracy under the premise of small training dataset. In the real-time positioning experiment, the accuracy can reach 95.37.

II. The Improved Resnet

We use 50 layers residual network in our assignment, to enhance the performance of our model, batch normalization layer, β -dropout layer and improved loss function are added into our model. The structure of the improved ResNet is as follow Table I:

The residual structure is composed of image preprocessing convolutional layer conv1, convolutional blocks conv2_3, conv3_4, conv4_6 and conv5_3 and full connection layer conv6. Each block is composed of three convolution layers, they are duplicated 3 times, 4 times, 6 times and 3 times, respectively. Batch Normalization layers are placed in front and back of each block and residual structure is applied in each block. Between conv5_3 and conv6, Average pool is applied to extract deep image feature and β -Dropout is applied to simplify the network. After conv6, loss function is applied, weight parameters are adjusted by stochastic gradient descent (SGD) of loss function with back-propagation, mini batch size is 256. The learning rate is 0.1 at the beginning and is divided by 10 when the error rate stops falling. We have 18 localization centers, so the final result of the loss function is, and we select the biggest one in these 18 number.

The residual structure is shown in Fig. 1， where x is the input of the convolutional block, the output of the block is $H(x)=F(x)+x$. Compared to $F(x)$, Nonlinear function $F(x)=H(x)-x$ is more easier to be optimized. Branch x is sent to the next block directly, which can be studied easily. Under this structure, Back propagation is easier to go on.

A. Batch Normalization

Batch normalization is used to regulate the input into a reasonable scope, which can avoid the vanishing/exploding of gradients caused by the increase of the layer of deep neural network.

$$\tilde{x}=\frac{x-E(x)}{\sqrt{Var(x)+\varepsilon}} \tag{1}$$

TABLE. I. Improved Resnet

Layer name	The configuration of each layer	Output size
conv1	Ksize=(7,7), stride=2, filter =64 max pool , batch normalization	112*112*64
conv2_3 3 layers	[1*1, 64; 3*3, 64, 1*1, 128]*3 Batch normalization	56*56*64 56*56*256
Conv3_4 4 layers	[1*1, 128; 3*3, 128, 1*1, 512]*4 Batch normalization	28*28*512 28*28*512
conv4_6 6 layers	[1*1, 256; 3*3, 256, 1*1, 1024]*6 Batch normalization	14*14*1024 14*14*1024
conv5_3 3 layers	[1*1, 512; 3*3, 512, 1*1, 2048]*6 Batch normalization	7*7*2048 7*7*2048
Aver pool	Ksize =(7,7)， stride=7	1*1*2048
β -Dropout	β changes adaptively	1*1*2048
conv6	Ksize =(1,1)， filter =2048， stride=1	1*1*2048
Loss function	Two loss function combined	1*1*18

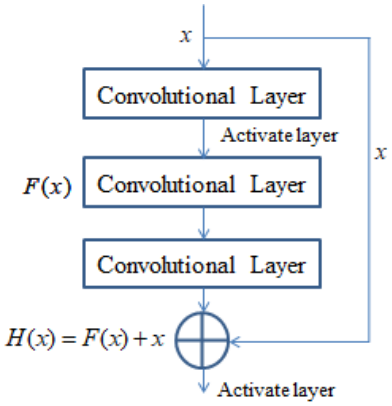


Fig. 1. Residual Structure.

Where x is the activate value of each node, $E(x)$ is the mean of one layer, $Var(x)$ is the variance. ε is a small number used to avoid denominator equal to zero. $\tilde{x}$ is the activate value after normalization.

B. Dropout

With the increase of the layer, depth neural network is easy to cause over-fitting, dropout [17-18] is a commonly used technology to alleviate this problem. The specific method is to discard a neural network node according to a certain probability in the training process of deep learning network, that is, to set the activate value of the node to zero. To enhance the stability of the model, β -dropout is applied. Adjusting the value of β adaptively, we can generate different kinds of distribution of dropout.

$$r^{(l)} \sim Beta(x:\beta,\beta) \tag{2}$$

$$\tilde{y}^{(l)}=r^{(l)}y^{(l)} \tag{3}$$

$$z_i^{(l+1)}=w_i^{(l+1)}\tilde{y}^{(l)}+b_i^{(l+1)} \tag{4}$$

$$y_i^{(l+1)}=f(z_i^{(l+1)}) \tag{5}$$

$$Beta(x;\alpha,\ \beta)=\frac{1}{B(\alpha,\ \beta)}x^{\alpha-1}(1-x)^{\beta-1} \tag{6}$$

$$Beta(\alpha,\ \beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{7}$$

$$\Gamma(\beta)=\int_0^{\infty}e^{-t}t^{\beta-1}dt \tag{8}$$

Where $r^{(l)}$ obey Beta distribution, $y^{(l)}$ is the activate value of l-th layer, $\tilde{y}^{(l)}$ is the output value of l-th layer, $w_i^{(l+1)}$ is

the weight of $l+1$-th layer, $b_i^{(l+1)}$ is the bias of $l+1$-th layer. $z_i^{(l+1)}$ is the input of $l+1$-th layer, $f()$ is the activate function. $\Gamma(\alpha),\ \Gamma(\beta)$ is Gamma function, we set $\alpha=\beta$, $Beta(\alpha,\ \beta)$ equal to $Beta(\beta)$, is a symmetry distribution. Adjusting parameter β, $Beta(x:\beta,\beta)$ can generate Bernoulli distribution, uniform distribution and Gaussian distribution. At the beginning of model training stage, the Bernoulli distribution is applied to delete some unimportant nodes, we set $\beta=0.001$; in the middle of model training stage, 0-1 uniform distribution is applied to smooth each node, we set $\beta=1$; at the end of model training stage, gauss distribution is applied to highlight important nodes, we set $\beta=3$ at this time.

C. Loss Function

In indoor localization algorithm, the image location features of the adjacent location points are similar, that is, the spacing between different classes is very small. In order to increase spacing between different classes and reduce spacing in one class, a loss function combined center loss and cross entropy loss is applied. The loss function can be described as follows:

$$L = L_s + \lambda L_c \tag{9}$$

Where L_c is the center loss function, L_s is the cross entropy loss function, λ is a weight used for balancing the two loss functions. The structure of our loss function is shown in Fig. 2.

The cross entropy loss function can be seen in [17]. We establish a class center in the feature space for each class. The center loss function is the sum of the distance between features of the sample and features of the class center in the feature space.

$$L_c = \frac{1}{2}\sum_{i=1}^{m} \| x_i - c_{y_i} \|_2^2 \tag{10}$$

L_c represents the center loss function, x_i is the depth feature of i-th class, c_{y_i} is the deep feature center of i-th class. m is the size of mini-batch.

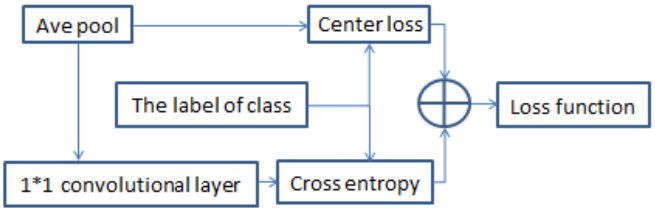


Fig. 2. The Proposed Loss Function.

III. Experimental Results and Analysis

A. ImageNet Classification

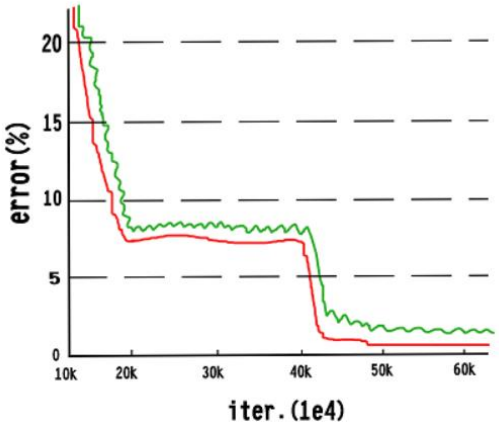
For testify the superiority of our algorithm, ImageNet 2012 classification dataset [19], which include 1.3 million images and 1000 classes, is applied. Our improved ResNet is trained by 1.15 million images, evaluated by 50k images, and eventually tested by 100k images.

The training error rate and test error rate of classical ResNet and our improved ResNet can be found in Fig. 3. It can be seen that curves of classical ResNet both in training set and test set fluctuate badly and curves of improved ResNet changes smoothly with the increase of iteration times.

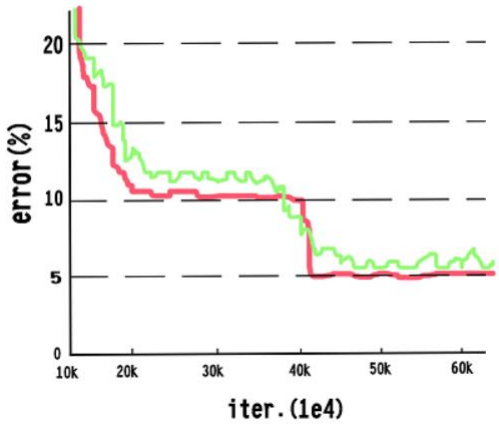
Top-1 and top-5 error rates of ResNet50 and our improved ResNet50 are shown in Table II. It can be seen that our improved ResNet is a little lower than ResNet.

B. Indoor Localization

Indoor localization experiments are done in a factory environment where 18 regions and location centers are placed. Fig. 4 shows the factory plan. An RGB camera, mobile control devices and a mini-computer are equipped in a mobile robot in which setup our trained ResNet. The experimental platform can be seen in Fig. 5. Using the improved ResNet, the position of a mobile robot can be classified by images taking in real time.



(a) Training Set.



(b) Testset.

Fig. 3. The Error rate of ResNet and Improved ResNet.

TABLE. II. IMPROVED RESNET

model		Top-1 error	Top-5 error
Vgg-16 [20]		22.58	8.43
Googlenet [21]		22.34	7.89
Prelu-net [22]		21.59	5.71
Inception [23]		21.99	5.81
ResNet [11]		20.74	5.25
Improved ResNet		20.35	5.22

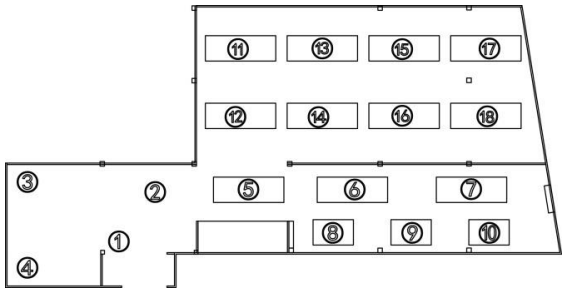


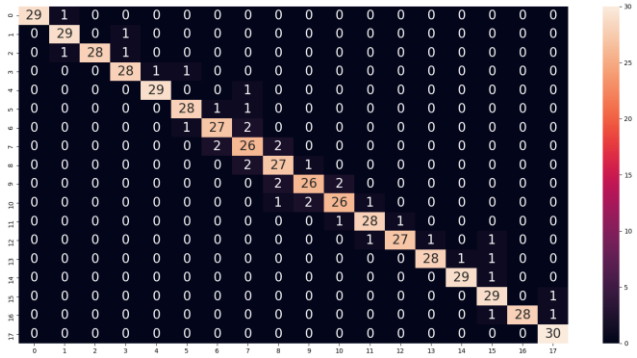
Fig. 4. Floor Plan of the Experimental Scene.

Fig. 5. The Experimental Platform.

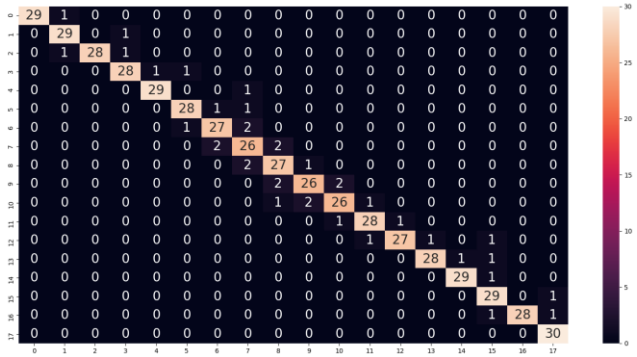
We constructed the dataset [24] in a factory environment where we test our algorithm. 1800 samples labeled location information are included in the dataset, we have 100 samples for each location point with different shooting angles.

Confusion matrix is employed to describe localization result, 30 images of each location region were used for testing these experiment results that can be seen in Fig. 6.

We can see in Fig. 6(a) that correct classification time of method1 is 504 and wrong classification time is 36, the accuracy of method1 is 93.33%, more errors happened in middle regions of the scene that is because the location feature of these nearby regions are similar and hard to distinguish, and when the input location feature fluctuate, the output would go to the wrong location. When comes to our improved ResNet, the output would be more stable when the location feature of the input changes, so the accuracy increases. In Fig. 6(b), correct classification time of method1 is 515 and wrong classification time is 25, the accuracy of improved ResNet is 95.37%, the accuracy increased by 2.04%.



(a) ResNet.



(b) Improved ResNet.

Fig. 6. Confusion Matrix of the Localization Results.

IV. CONCLUSION

An improved residual network is proposed in this paper to enhance the stability of classical ResNet. Based on the classical residual network, batch normalization, adaptive β -dropout random deactivation function and a new loss function are added into the proposed model. Batch normalization is applied to avoid vanishing/exploding gradients. β -dropout is applied to increase the stability of the model, which we select different dropout method adaptively by adjusting parameter β . The new loss function is composed by cross entropy loss function and center loss function to enhance the inter class dispersion and intra class aggregation. The improved ResNet50 is then applied to the indoor positioning of mobile robot in a factory environment. The experimental results show that the algorithm can achieve high indoor positioning accuracy under the premise of small training dataset. Future work will focus on the temptation and improvement of other neural-networks to improve the accuracy of the indoor localization system.

REFERENCES

[1] Moustafa A, Moustafa E, Marwan T. “WiDeep: WiFi-based Accurate and Robust Indoor Localization System using Deep Learning”, 2019 IEEE International Conference on Pervasive Computing and Communications, Kyoto, Japan, IEEE, March 2019: 1883-1890.

[2] Cabrera E, Camacho D. “Towards a Bluetooth Indoor Positioning System with Android Consumer Devices,” The IEEE International Conference on Information Systems and Computer science, Quito, Ecuador, 2017: 56-59.

[3] Qiu L, Huang Z, Wirstrom N, et al. “3DinSAR: Object 3D localization for indoor RFID applications,” The IEEE International Conference on RFID, Orlando, USA, IEEE, 2016:101-108.

[4] Desai A, Ghagare N, Donde S. “Optimal Robot Localization Techniques for Real World Scenarios”, 2018 Fourth International Conference on Computing Communication Control and Automation, Pune, India, IEEE, Aug, 2018: 1861-1868.

[5] Walch F, Hazirbas C, Sattler T, et al. “Image-based localization using LSTMs for Structured Feature correlation,” The IEEE International Conference on Computer Vision, Venice, Italy, IEEE, 2017: 627-637.

[6] Krizhevsky A, Sutskever L, and Hinton G. “AlexNet: Imagenet classification with deep convolutional neural networks”, 2012 International Conference and Workshop on Neural Information Processing Systems, Spanish, Lake Tahoe, NIPS, 2012: 3546-3559.

[7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

[8] K. He, X. Zhang, S. Ren, and J. Sun. “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.” International conference on computer vision, IEEE, 2015: 345-367.

[9] S. Ioffe and C. Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift.” International conference on machine learning, IEEE, 2015:1245-1263.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. “Rabinovich. Going deeper with convolutions.” International conference on computer vision and pattern recognition, IEEE, 2015: 784-796.

[11] He K , Zhang X , Ren S , et al. “Deep Residual Learning for Image Recognition”, 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vega, USA, IEEE, 2016: 770-778.

[12] S. Xie, R. Girshick, P. Dollar, et al. Aggegated residual transformations for deep neural networks. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1492-1500.

[13] J. Hu, L. Shen, G. Sun. “squeeze-and-excitation networks.” IEEE Conference on computer vision and pattern recognition. 2018: 7132-7141.

[14] Y. Chen, J. Li, H. Xiao, et al. “Dual path networks”. Computer Sciences , Arxiv: 1707.01629. 2017: 4467-4475.

[15] S. Zagoruyko, N. Komodakis. “Wide residual networks”. Computer Sciences, Arxiv: 1605.07146. 2016: 2378-2386.

[16] G. Hinton, E. Osindero, et al. “A Fast Learning A lgorithm for Deep Belief Nets”. Neural Computation, 2006, 18(7): 1527-1554.

[17] Chen X J, Guo R Q, Luo W, et al. “Visual Crowd Counting with Improved Inception-ResNet-A Module”, 2018 IEEE International Conference on Robotics and Biomimetics, Kuala Lumpur, Malaysia, IEEE, 2018.

[18] Li B Q, He Y. “An Improved ResNet Based on the Adjustable Shortcut Connections”, IEEE Access, 2018, 5(99): 1348-1356.

[19] http://www.image-net.org/

[20] Simonyan K, Zisserman A. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. Computer Science, 2014, 35(4): 386-400.

[21] Szegedy C, Liu W, Jia Y, et al. “Going deeper with convolutions”, 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015.

[22] He K, Zhang S R and Sun J. “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification”, 2015 IEEE International Conference on Computer Vision, Santiago, USA, 2015.

[23] Ioffe S , Szegedy C . “Batch normalization: accelerating deep network training by reducing internal covariate shift”, 2015 International Conference on Machine Learning and Cybernetics, Guangzhou, China, JMLR, 2015.

[24] https://pan.baidu.com/s/1oR7fg_sZHe_qHtpSH7PUzg, password: ix4q.