Deep Learning Applied to Smart Home Face Recognition Access Control System

Shuangye Chen^{1,*}, Shuangchun Ding¹, Hanguang Fu², Yaoshan Xian¹, Xinqi Liu¹ and Chaocun Zhang¹

¹Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

²College of Materials Science and Engineering, Beijing University of Technology, Beijing 100124, China

*Corresponding author

Abstract—With the development of embedded technology and Internet of things technology, smart home has developed rapidly in recent years. At the same time, deep learning also has brought breakthroughs. The application of deep learning to smart home system can bring a good user experience and increase security. In this work, the application of the convolution neural network model, which is belong to deep learning method, in the field of human face recognition in natural scenes. Embedded devices collect and preprocess image and send it to the server. In the server, a improved lightened VGG network model has been designed, which is used to face recognition matches. This smart home system can reduce the computation of embedded devices, improve the accuracy of recognition. The test of this system meets the requirements of face recognition applications in the surveillance video.

Keywords—smart home; embedded system; convolutional neural network; face recognition

I. INTRODUCTION

The embedded platforms has the advantages of good processing speed, low power consumption, small volume, abundant scalability. a video image acquisition and processing system is designed based on embedded platform. But embedded platform can't perform well in complex image processing operation. In order to deal this problem, we perform complex image processing operations through the cloud server. A combination of embedded platforms and cloud servers can meet the needs of outdoor equipment condition monitoring system which requires high real-time performance and low cost.

The embedded platforms use T20 produced in Junzheng Company, the OV2735 as video acquisition sensor. It runs the Linux operating system, collects the data of video sensor for face detection, and gets the image data of the face. A socket client is designed in T20. It sends the processed image data to the cloud server in real time and gets the control instructions from the server.



FIGURE I. THE EMBEDDED PLATFORM

The aliyun server is used as the cloud server, develop Web, database, PHP and deep learning on aliyun server. A socket server is designed on aliyun to manage the connected client devices and issue control instructions. In this paper, we will focus on image preprocessing and the improved lightened VGG network model based on face recognition.

II. EMBEDDED IMAGE ACQUISITION AND PREPROCESSING

Before the network training and testing, image preprocessing will be done, by using OpenCV library in T20 embedded platform. There are too many background patterns around the original picture. The background patterns is useless data, we cut it, and retain some useful facial data. According to the key points of the face, the angle of the face is adjusted to achieve the goal of the face in the middle position.



FIGURE II. IMAGE PREPROCESSING

The picture(a) is getting face area, the picture is Cutting to face area, the picture(c) is getting the key points of the face, the picture(d) is adjust the angle and the position of the face.



FIGURE III. IMAGE PREPROCESSING PICTURE

III. LIGHTENED VGG NETWORK MODEL

A. VGG Network Model

ReLU (Rectified Linear Units) is the concept put forward by neuroscientist Dayan, Abott in 2001. It is a very famous non-linear activation function for deep CNN. The ReLU is given as follows:

$$y_i = \begin{cases} x_i, & \text{if } x_i > 0, \\ 0, & \text{if } x_i < 0. \end{cases}$$

where x_i is the input of ReLU, y_i is the output of ReLU. If the input of ReLU is smaller than 0, the output is not activated and output 0. If the input of ReLU is larger than 0, the output is activated and output the value of input. So, by using this method, the nonlinearity of ReLU is achieved.

The ReLU has three advantages over the previous traditional activation function model:

(1). Compared to the smooth non-linear activation functions, ReLU has much faster learning speed

(2). ReLU is once proposed.

(3). Compared to Tanh and Sigmoid, ReLU reduces error rate of backpropagation.

The illustration of Sigmoid and ReLU shown in figure 4.





To observe the distributions of the inputs on non-linear activation layers, this VGG network model with ReLU is used. VGG network module is one of the most common convolutional neural network module, considered to be an outstanding visual model. The VGG network module inherits the benefits of AlexNet. It has unique advantages in image classification, for example: face recognition and object identification and so on. The traditional VGG network module is shown in Figure 5.



FIGURE V. TRADITIONAL VGG

B. Improved Lightened VGG Network Model

During the training the VGG network model, the network parameters and intermediate results that should be storage are too huge. Under the limited hardware condition, the single ordinary graphics card is unable to meet its demand for the memory capacity, even if it can barely run but cannot be fully trained to achieve good face recognition effect.

The number of parameters of fully-connected layers is over 75% of the total parameters. To reduce the parameters of the VGG module and accelerate the training speed, the fully-connected layer-1 and fully-connected layer-2 are removed. The number of the fifth convolutional block convolution kernel increases to 700, it can extract more rich and different high-level abstract features of images. At the end of fifth block, max-pooling layer is replaced by average-pooling. This method has appeared in GoogleNet and so on. This method reduces the number of parameters of the fully-connected layers and has no significant effect on network performance, further promote the model to extract more abstract and more recognizable features. Taking all factors into consideration, the improved lightened VGG network module reduces the parameters of the fully-connection layer, not only the time of the model calculation is greatly reduced, but also the demand for the memory space is reduced. The improved lightened VGG network module is shown in Figure 6.



FIGURE VI. THE IMPROVED LIGHTENED VGG

By improving the original algorithm, the network parameters and structure are significantly reduced, and the time required for training and identification is also reduced.

IV. SIMULATION AND ALGORITHMIC PROCESS

Experimental use Anaconda and Tensorflow programming and testing. Use the ORL Face library as data collection to test the result of this method, there are the 40 individuals in the ORL Face library, every people has 10 face images, total 400 images, the size of the images is 92*112. As for this database, every people's images are divided into two groups: training set and test set. There are 7 images in the training set, the rest images belong to test set. Part of the face images shown in Figure 7.



FIGURE VII. PART OF FACE IAMGE OF ORL FACE DATABASE

• The port number of fully-connected layer changes with the number of the classification of data sets. So in here the port number of the last fully-connected layer is set to 40. The parameters are set by testing. This module consists of 200 epochs. Through constant training to adjust the learning rate, choose the appropriate learning rate and adjust and optimize the network. The initial learning rate is set to 0.001. When it reaches the 10 epoch, the learning rate is adjusted to 10%. The learning rate is adjusted at a rate of 10% every time. At the end of the training, the value of loss changes to about 0.1. According to the test, the test result shown in table1:

Table 1 compares the performance of lightened VGG network module and traditional VGG network module based on the test. The error rate of lightened VGG is 9.47%, and that of traditional VGG is 9.64%. The average recognition time of lightened VGG is 2.15s, that of traditional VGG is 3.67s

TABLE I.	TABLE TYPE STYLES
----------	-------------------

method	error rate	recognition time
lightened VGG	9.47%	2.158
traditional VGG	9.64%	3.67s

V. CONCLUSION

The number of parameters of traditional VGG network module is too huge, it does not apply to embedded smart home systems. In order to meet the requirement of security and monitoring recognition of smart home system, a smart home system is designed for face recognition. The embedded platform of Linux system is used to collect and preprocess images, and the pre-processed images are sent to the cloud server for identification. In the cloud server, an improved lightened VGG algorithm is designed to identify. Through testing, the system can realize real-time video monitoring, face recognition speed is relatively fast, and the requirements of system hardware are relatively low. The error rate of face recognition is low, which can meet the actual demand well.

ACKNOWLEDGEMENT

In this paper, the research was sponsored by The Key Special Project of the National Key Research and Development Program of China. (Project No. 2017YFB0306404).

REFERENCES

- Yann LeCun, Yoshua Bengio & Geoffrey Hinton. Deep Learning, Nature 521 436-444 (28 May 2015)
- [2] CHELLAPPA R,WILSON C L,SIROHEY S. Human and machine recognition of face: A Survey[J].Proc IEEE, 1995, 83(5): 705-740.
- [3] CHEN Haixia, CUI Qian. Face recognition based on Gabor wavelet and PCA[J]. Northeast Normal University Journal, 2014,46 (4): 77-80. (in Chinese).
- [4] C.K.Chui, W.He, J.Stochler. Compactly supported tight and sibling frames with maximum vanishing moments[J]. Appl. Comput.Harmon.Anal.,2002,13(3):224-262.
- [5] Han, Matrix splitting with symmetry and symmetric tight framelet filter banks with two high-pass filters[J]. Appl. Comput.Harmon. Anal.,2013,35200-227.
- [6] Xiaoyuan Yang, Yan Shi, Wanlu Zhou. Construction of
- [7] parameterizations of masks for tight wavelet frames with two symmetric/antisymmetric generators and applications in image compression and denoising[J]. Journal of Computational and Applied Mathematics, 2011, 235:2112-2136.
- [8] George E. Dahl, Dong Yu, Li Deng, and Alex Acero. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. Audio, Speech, and Language Processing, IEEE Transactions on, 20(1):30 -42, January 2012.
- [9] He Lianghua Multi-Dimension Principal Component Analysis Based on Face Recognition[J]. The Journal of New Industrialization, 2012,2(1):59-65.
- [10] Badrinarayanan, A. Kendall, R. Cipolla, Segnet. a deep convolutional encoder-decoder architecture for image segmentation, 2015. CoRR arXiv:abs 1511.00561.
- [11] M. Courbariaux, Y. Bengio, J. David, Binaryconnect. training deep neural networks with binary weights during propagations, 2015, in: Proceedings of the Advances in Neural Information Processing Systems.
- [12] Zhang, J. Han, J. Han, L. Shao. Cosaliency detection based on intrasaliency prior transfer and deep intersaliency mining, IEEE Trans. Neural Netw. Learn.Syst. 27 (6) (2016) 1163–1176.