A CAPTCHA recognition technology based on deep learning

Yu Hu[1,2] , Li Chen[1,2*] , Jun Cheng[3]

[1]School of Computer Science of Technology, Wuhan University of Science and Technology, Wuhan China, 430065.

[2]Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System,Wuhan China, 430065

[3]Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences, Ningbo China，315201
chenli@wust.edu.cn*

***Abstract*—Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA) is an important human-machine distinction technology for website to prevent the automatic malicious program attack. CAPTCHA recognition studies can find security breaches in CAPTCHA, improve CAPTCHA technology, it can also promote the technologies of license plate recognition and handwriting recognition. This paper proposed a method based on Convolutional Neural Network (CNN) model to identify CAPTCHA and avoid the traditional image processing technology such as location and segmentation. The adaptive learning rate is introduced to accelerate the convergence rate of the model, and the problem of over-fitting and local optimal solution has been solved. The multi task joint training model is used to improve the accuracy and generalization ability of model recognition. The experimental results show that the model has a good recognition effect on CAPTCHA with background noise and character adhesion distortion.**

Keywords—Convolutional neural network; CAPTCHA; Adaptive learning rate; Multi task joint training;

I. Introduction

With rapid development of the Internet industry, more and more network security issues happens. CAPTCHA [1] technology has a wide range of applications in network protection and information security. CAPTCHA is the abbreviation of Completely Automated Public Turing Test to Tell Computers and Humans Apart. As a network security strategy, it is mainly used for websites to prevent automatic malicious program attacks such as automatic registration、spam、automatic voting and so on. For humans, the recognition accuracy of effective CAPTCHAs is at least 80%, but for computers, it should be less than 0.01% [2]. The research of CAPTCHA identification can find the defects of the CAPTCHAs in time, and provide improvement suggestions for the code generation program, and increase the security of the CAPTCHAs, At the same time, as a kind of Turing test, CAPTCHA recognition combines the research results of image processing and artificial intelligence field, and plays a positive role in the development of artificial intelligence technology, such as license plate recognition and handwriting recognition.

This paper focuses on the most widely used character based images CAPTCHA, which is composed of random numbers and English letters. It is easy to generate, not affected by the user's cultural background, and the brute force is difficult to crack. We can create a picture containing numbers and letters by mainstream programming languages [3]. In order to increase the difficulty of recognition by computer, the CAPTCHAs needs to add background noise and carry out the characters twist conglutination processing.

In the field of traditional image processing, the CAPTCHA recognition technology is divided into image preprocessing, positioning, character segmentation, character recognition and other steps. However, it is difficult to establish an accurate template set because of the adhered and complicated CAPTCHA. The traditional method of extracting pixel points one by one and template matching, can only recognize simple CAPTCHAs, while there is no efficient method to recognize the adhered and complicated CAPTCHA. Therefore, a more efficient method is needed to identify such CAPTCHAs.

Nowadays, the deep learning network is widely used in scientific research. As one of the hotspots in artificial intelligence research field in recent years, it has achieved great success in many fields, such as image recognition, speech recognition, natural language processing and target detection. Compared with the traditional pattern recognition method, the biggest advantage of deep learning is that can learn features actively without artificial design. Based on the above observation and inspiration, a convolutional neural network (CNN) algorithm [4] is proposed to identify the CAPTCHAs. For the problem of model convergence rate and global optimal solution, the adaptive learning rate is introduced to improve the learning ability of the network, and it has better convergence and robustness. The multi task joint training model is used to speed up the training speed of the model and improve the generalization ability of the model. The method of this paper directly uses images as input to CNN, do not have to split the image for the characters, active learning features in the process of network training. The experimental results show that the proposed method has a good recognition effect on CAPTCHA with background noise and character adhesion distortion.

II. Related Researches

The traditional method is to locate a single number or character regions in an image, and then segment and identify the individual characters [5, 6]. Through these two steps, the characters in the image are detected. For example, Yan and Ahmad, was successfully segment the Microsoft CAPTCHAs and identified it by multiple classifiers, with a recognition rate of 60% [7]. Mori and Malik recognizes the CAPTCHAs in images by using a shape context method [8]. Chellapilla and

978-1-5386-3758-6/18/$31.00 ©2018 IEEE

Simard also solve CAPTCHAs by segmenting single characters and recognizing them [2]. Domestic scholars also have research on CAPTCHA recognition based on the segmentation method. Wang Yang et al. proposed to use K-nearest neighbor (KNN) technology for recognition verification code [9]. Zhang Shuya et al. Through segmented the CAPTCHAs, the segmentation results were identified by KNN classifier, back propagation (BP) network and support vector machine (SVM) respectively, and the recognition rates were all above 95% [10].

However, In order to prevent the computer recognizing the CAPTCHAs automatically and improving the security of the network, the characters in the current CAPTCHAs will partially overlap, so that the division of the single character becomes very difficult, thereby affecting the recognition accuracy. Aim to the above problems, in the face of the limitations of traditional image processing methods, LeCun et al. proposed to use deep learning methods to identify handwritten digits, and convolutional neural networks were used to extract image features and then classify them [11]. However, they all need to split the images. Instead, we use the whole images as an input to get the result directly.

III. Proposed Method

A. Contrast Normalization

Contrast normalization can avoid the neuron output saturation caused by excessive absolute input value, and ensure small values in the output data will not be swallowed. It's can enhance the generalization of the network, eliminate the influence of brightness and contrast variance effectively on the network, can greatly reduce the dependence between neighboring factors and accelerate the convergence of the network. Before the network training, this paper extracts the image for contrast normalization. Set the brightness value of the image (i,j) to $I(i,j)$, and the brightness value after the local contrast normalization is $I'(i,j)$, contrast normalization method [12] can be expressed as:

$$I'(i,j)=\frac{I(\mathrm{i},j)-\mu}{\sigma+C} \tag{1}$$

Where,$i \in \{1,2,\cdots,M\}$,$j \in \{1,2,\cdots,N\}$, M, N is the dimension of image blocks; C is constant 1, avoiding denominator be zero. μ, σ is the mean and standard deviation of image pixel values.

B. Multi task joint training

Multi task learning is a method of machine learning opposite to single task learning. The main goal is to improve the generalization ability by using domain specific information in the training signals hidden in multiple related tasks. Multi task learning can accomplish this task by using shared representation to parallel training multiple tasks. Multi task learning network parameter sharing, and can reduce the number of models, improve learning efficiency.

During the training of CAPTCHA recognition model, images labels are divided into multiple learning tasks, each task training one character and all tasks training together. The structure of the multi task joint training network is shown in figure.1:

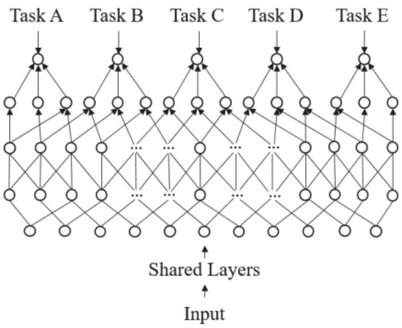

Fig. 1. Model of Multi task joint training

C. CAPTCHA recognition model

VGG-Net [13] is a convolutional neural network developed by Oxford visual geometry group. The model is based on the Alex-Net [14] network architecture, which deepens the convolutional layer and reduces the size of the convolution kernel. Through the improvement of these two aspects, the performance of VGG-Net has been greatly improved. According to the advantages of VGG-Net, combined with the work of this paper, we propose a deep CNN method to recognition a series of characters without pre-segmentation. The network structure we use is shown in Figure. 2. Each CAPTCHA images contains 6 characters. In the output layer, every 62 neurons predict a character. We define a bijection $\theta(x)$ that maps a character $x \in \{'0',\cdots,'9',a',\cdots,'z',A',\cdots,'Z'\}$ to an integer $l \in \{0,\cdots,61\}$:

$$\theta(x)=\begin{cases}0\sim 9, x='0'\sim'9'\\ 10\sim 35, x='a'\sim'z'\\ 36\sim 61, x='A'\sim'Z'\end{cases} \tag{2}$$

We assign the first 62 output neurons to the first character of the sequence, the second 62 neurons to the second character and so on. The output layer has $5\times 62=310$ neurons.

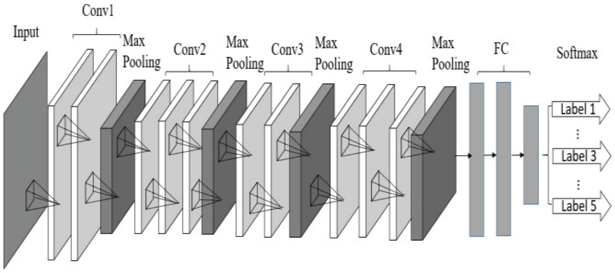

Fig. 2. Structure of CNN for CAPYCHA recognition

The input size of the training set in this paper is 224 × 224, the edge of the input feature map is expanded, so that the size of the output feature mapping is the same as that of the output feature mapping. Parameters of each layer of the model in TABLE I.

TABLE I. CNN structure parameters

Name	Type	**Size/Stride**

Conv1_1	Convolution	5×5/1
Conv1_2	Convolution	5×5/1
M_Pool1	Max Pooling	2×2/2
Conv2_1	Convolution	3×3/1
Conv2_2	Convolution	3×3/1
M_Pool2	Max Pooling	3×3/1
Conv3_1	Convolution	2×2/2
Conv3_2	Convolution	3×3/1
Conv3_3	Convolution	3×3/1
M_Pool3	Max Pooling	2×2/2
Conv4_1	Convolution	3×3/1
Conv4_2	Convolution	3×3/1
Conv4_3	Convolution	3×3/1
M_Pool4	Max Pooling	2×2/2
Fc_1	Full Connection	-
Fc_2	Full Connection	-
Fc_3	Full Connection	-
Loss	SoftmaxWithLoss	-

IV. Experimental Results

The experimental hardware environment is Windows 10 64-bit, Inter(R) Pentium(R) CPU G3258, RAM 8GB and the graphics card NVIDIA GT980Ti. The software environment used in the experiment is TensorFlow.

As there is no public CAPTCHAs dataset that can be used currently, and training convolutional neural network models requires a large number of data. To solve this problem, we use a python script to generate a CAPTCHA images with 5 characters. Each character is randomly taken from a set of 10 digits and 26 English letters, and the characters are distorted. During CAPTCHA generation, we removed duplicate images to ensure the reliability of the model. The size of images is 128×48. The training set include 5×10^4 images, the validation set include 2×10^4 images and test set include 1000 images. Figure.3 shows some examples of our auto-generated CAPTCHAs.

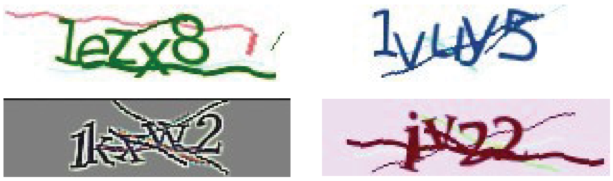

Fig. 3. Example CAPTCHAs used in the experiments

In this paper, a stochastic gradient descent (SGD) algorithm is used to train the model, and the global optimal solution is achieved through a large number of iterations. This avoids the problem that the loss function does not decrease when the model converges to a local optimum. The learning rate adaptively changes according to the number of iterations, and the learning rate will changes by the formula $lr = lr_0 \times (1/(1 + decay \times i))$, where the base learning rate $lr_0 = 0.001$, the attenuation factor of learning rate $decay = 0.0001$ and i is the current number of iteration. The full connection layer uses the dropout method, and dropout parameter $dp = 0.5$.

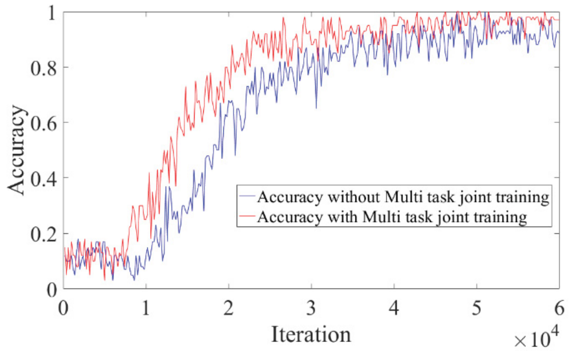

Fig. 4. Effect of multi task joint training on recognition accuracy

We verify the effect of multi task joint training and single task training on the model by using controlling variable method. The results are shown in Figure.4, the results show that use of multi task joint training mode has faster convergence rate and higher accuracy than the single task training model.

The performance evaluation method proposed by the method is the accuracy of the recognition. Table II shows the performance of different methods for CAPTCHA recognition, including BP neural networks, SVM, and KNN algorithms. However, all the above methods need to preprocess and segment images to achieve the purpose of recognition. This paper also compares the performance of the classical convolution neural network LeNet to the same dataset. The results show that, compared with other methods, the proposed method does not need to split the characters in images, and has better recognition performance.

TABLE II. Performance comparison among CAPTCHA Recognition Method

Method	Accuracy of recognition
Ref. [7]	92%
Ref. [8]	60%
Ref. [10]	95%
LeNet	79.4%
Proposed Method	96.5%

V. Conclusion and Future work

The CAPTCHAs is a test method used to distinguish between humans and machines in network environment. The studies on CAPTCHA identification can better detect vulnerabilities in the security of the CAPTCHA, thereby preventing some malicious intrusion in the network. In this paper, a CAPTCHAs recognition technology based on convolutional neural network is proposed according to the

CAPTCHA of images character distortion and adhesion, and all characters in the image can be recognized without segmentation. Multi task joint training model is introduced to improve network learning rate and model generalization ability, and the network structure is modularized, which can recognize the different character length of the CAPTCHA image with slight modification. Experimental results show that the proposed method has a good recognition effect, and the recognition accuracy reaches to 96.5%. In the future work, Chinese characters CAPTCHAs recognition will be added.

ACKNOWLEDGMENT

This work was supported by Nation Natural Science Foundation of China（No. 61773297）

REFERENCES

[1] L. Von Ahn, M. Blum, N. J. Hopper and J Langford, “CAPTCHA: Using hard AI problems for security,” *International Conference on the Theory and Applications of Cryptographic Techniques*, Springer, Berlin, Heidelberg, 2003, pp. 294-311.

[2] K. Chellapilla and P. Y. Simard, "Using machine learning to break visual human interaction proofs," *Advances in neural information processing systems*, 2005, pp. 265-272.

[3] T. Converse, “CAPTCHA Generation as a Web Service.” *Proc. Human Interactive Proofs*, Springer, Berlin, Heidelberg, 2005, pp. 82-96.

[4] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud and V. Shet, “Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks,” *Computer Science*, 2013.

[5] M. Jaderberg, A. Vedaldi, and A. Zisserman, “Deep features for text spotting,” *European conference on computer vision*, Springer, Cham, 2014, pp. 512-528.

[6] P. Y. Simard, D. Steinkraus and J. C. Platt, “Best practices for convolutional neural networks applied to visual document analysis,” *International Conference on Document Analysis and Recognition IEEE Computer Society*, Vol. 3, 2003, pp. 958-962.

[7] J. Yan, A. S. E. Ahmad, “A low-cost attack on a Microsoft captcha,” *ACM Conference on Computer and Communications Security*, CCS 2008, Alexandria, Virginia, USA, Oct, DBLP, 2008, pp. 543-554.

[8] G. Mori and J. Malik, “Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA,” *Computer Vision and Pattern Recognition*, 2003. *Proc. IEEE Computer Society Conference*, Vol. 1, IEEE, 2003, pp. I-I.

[9] Y. Wang, Y. Q. Xu, Y. B. Peng, “KNN-based Verification Code Recognition Technology on Campus Network,” Computer and Modernization, No. 2, 2017, pp. 93-97

[10] S. Y. Zhang, Y. M. Zhao, X. Y. Zhao, J. L. Li, “An Approach to Recognition of Authentication Code,” *Journal of Ningbo University*, 2007.

[11] Y. LeCun, L. Bottou and Y. Bengio, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, Vol. 86, No. 9, 1998, pp. 2278-2324.

[12] M. Y. Wu, L. Chen, J. Tian, “Video image distortion detection and classification based on CNN”. *Application Research of Computer*, Val. 33, No. 9, 2016, pp.2827-2830.

[13] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *Computer Science*, 2014.

[14] A. Krizhevsky, I. Sutskever and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the Acm*, Vol. 60, No. 2, 2012.