IET Research Journals

Brief Paper

IET Journals
The Institution of Engineering and Technology

ISSN 1751-8644
doi: 0000000000
www.ietdl.org

SiameseCCR: A Novel Method for One-shot and Few-shot Chinese CAPTCHA Recognition using Deep Siamese Network

Z. Chen[1] *W.F. Ma*[1] *N.F. Xu*[1] *C.T. Ji*[1] *Y.L. Zhang*[1]

[1] *School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou, 310023, People's Republic of China*

** E-mail: mawf@zust.edu.cn*

Abstract: The research of CAPTCHA recognition is helpful to discover the security vulnerabilities in time and improve its safety. In comparison with digits and English letters, Chinese characters have many more categories which lead to the requirement of a large amount of training data. Therefore, this paper proposes a novel method for one-shot and few-shot Chinese CAPTCHA recognition, using the deep Siamese network, based on the idea of template matching. In this method, the residual convolutional neural network branches are used for feature extraction of CAPTCHAs, a fully-connected layer is used for calculating the similarity of features, and a hard negative mining algorithm is designed to promote convergence. Experiments are done on a self-built small-scale Chinese CAPTCHA dataset. The results show that this proposed method can achieve higher accuracy on the known characters than traditional methods. For the brand-new characters, only one template is required to recognize them and the accuracy is close to known characters. To summarise, it is able to build a Chinese CAPTCHA recognition model with high accuracy and extensibility by using a small-scale dataset.

1 Introduction

With the rapid development of information technology, network security is attracting more and more attention. As an essential technology to identify machine and human, CAPTCHA has been widely used in various fields. On the Internet, the most common CAPTCHAs are English and digital CATPCHAs (Figure 1(a)), but presently, because of the high recognition accuracy of this kind of CAPTCHA [1–3], using Chinese CAPTCHA (Figure 1(b)) as a substitute is increasingly popular in China. Compared with English and digital CAPTCHAs, Chinese CAPTCHAs are more challenging to recognize because of their various categories and complicated structures.

In past research, Chinese CAPTCHA recognition was considered as a classification task with a fixed number of categories, which relies on a large amount of labeled data [4–6]. Nowadays, the Chinese CAPTCHA recognition methods based on the above pattern have achieved high recognition accuracy, but they still have the following defects: 1) Previous studies usually require hundreds of thousands of labeled data for training (i.e., each category requires hundreds of labeled images), and it is difficult to maintain high accuracy with few samples; 2) These models need to determine the number of categories before training, and cannot recognize the Chinese characters that are not included in the training set.

Therefore, we propose a method of Chinese CAPTCHA recognition based on Siamese network. In this method, CAPTCHAs are not classified directly but are matched with the templates to find out its category. In comparison with traditional methods, this method dramatically reduces the data needed for training, in other words, it only requires few or even one labeled image for each category of Chinese characters.

(a) English and digital CAPTCHA (b) Chinese CAPTCHA

Fig. 1: Common CAPTCHAs on the Internet

For the convenience of description, this method is called SiameseCCR in this paper. The experimental results show that this novel method can effectively promote the network to extract discriminative features from a small-scale dataset so as to achieve a better performance than traditional methods. At present, our code and dataset are available on Github: `https://github.com/czczup/SiameseCCR`.

2 Related Work

Due to the growing popularity of Chinese CAPTCHAs, the recognition technology is studied in depth by scholars, aiming to find the defects in Chinese CAPTCHAs, provide suggestions for its generation procedures, and promote the development of pattern recognition.

In 2016, Liu et al. [6] used a LeNet-5-like convolutional neural network to identify the Chinese idiom CAPTCHAs, achieved 99.95% single character accuracy, by simulating the characteristics of existing data and generating new data to expand the training set. In 2018, Jia et al. [4] trained an 11-layer convolutional neural network with self-built 600,000 Chinese character CAPTCHAs, achieved a recognition rate of 99.4%. In 2018, Lin et al. [5] used 200,000 Chinese character CAPTCHAs generated by the CAPTCHA generator Kaptcha as the training data and achieved accuracy of 97.72%.

In the above methods, they built multi-category classifiers using CNN, trained them with a large amount of labeled data, and obtained excellent performance. However, in the real-world scene, it is extremely arduous to collect hundreds of thousands of labeled Chinese CAPTCHAs. Furthermore, these models based on the above methods do not have flexibility and expandability, so it is difficult to cope with the update of CAPTCHA generators.

For the above reasons, this paper does not use the traditional pattern but instead builds a Siamese network [7] for Chinese CAPTCHA recognition. Siamese network is a kind of neural network architecture for similarity metric, and its Siamese architecture consists of two subnetworks, which require different inputs but share

© The Institution of Engineering and Technology 2015

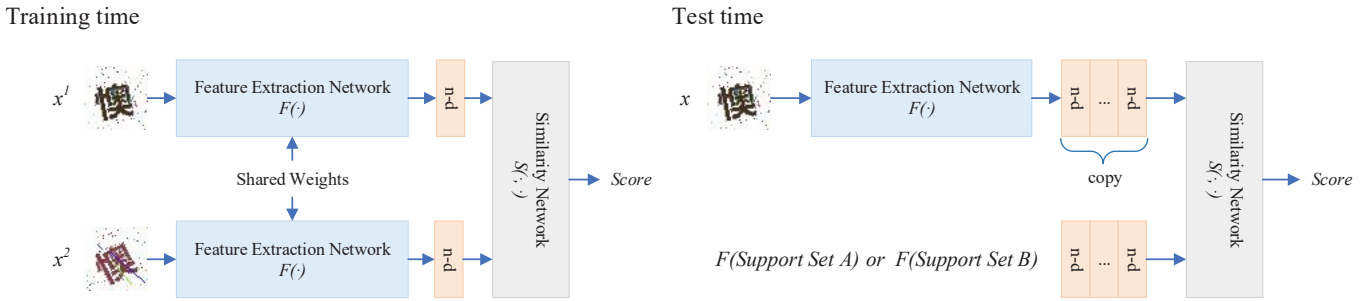


Fig. 2: Illustration of the network architecture. This model is divided into two modules: Feature Extraction Network $F(\cdot)$ and Similarity Network $S(\cdot,\cdot)$. The Feature Extraction Network transforms the input images into n-d feature vectors (n is a hyper-parameter). The Similarity Network requires two feature vectors as input and calculates their similarity. At training time, both branches of siamese architecture are available. At test time, only one branch is used while the other branch is replaced by pre-extracted feature vectors, which can reduce the repeated calculation and speed up the recognition of Chinese CAPTCHAs.

the same weights. The goal of a Siamese network is to learn a feature extraction function, increase intra-class similarity and reduce inter-class similarity, so as to realize classification, matching, and clustering.

In 1994, Bromley et al. [7] first proposed an algorithm based on Siamese network for signature verification. In [8], Nair and Hinton applied a Siamese architecture to face verification. Afterward, in 2015, [9] presented a two-channel network for computing similarity of image pairs, which is improved from Siamese network. In 2015, Koch et al. [10] proposed a method for one-shot recognition using a Siamese network, and achieved remarkable performance on the Omniglot dataset.

Due to the excellent performance of Siamese networks in small datasets, we build a similarity metric model based on the Siamese architecture and achieve the superior performance of Chinese CAPTCHA recognition in the scene of one-shot and few-shot.

3 The Dataset

To evaluate our method in one-shot and few-shot Chinese CAPTCHA recognition tasks, we designed a Chinese CAPTCHA dataset based on the GB2312 standard, including level-1 and level-2 common used Chinese characters, which contains 3,755 and 3,008 characters. We named this dataset *CAPTCHA Images of Chinese Characters* (CICC).

We refer to the setting of the dataset in [11] and divide our dataset into three parts: training set $\mathcal{D}_{train}$, testing set $\mathcal{D}_{test}$, and support set $\mathcal{D}_{support}$. The training set contains 15,020 Chinese character images, corresponding to the 3,755 characters in the GB2312 level-1 set. The testing set includes 20,000 images, of which 10,000 samples ($\mathcal{D}^A_{test}$) correspond to level-1 set and the other 10,000 samples ($\mathcal{D}^B_{test}$) correspond to level-2 set. The support set shares the same label space with the testing set, where samples are used for template matching. Similarly, the support set is split into two parts: $\mathcal{D}^A_{support}$ is for few-shot recognition and $\mathcal{D}^B_{support}$ is for one-shot recognition. The specific settings are shown in Figure 3.

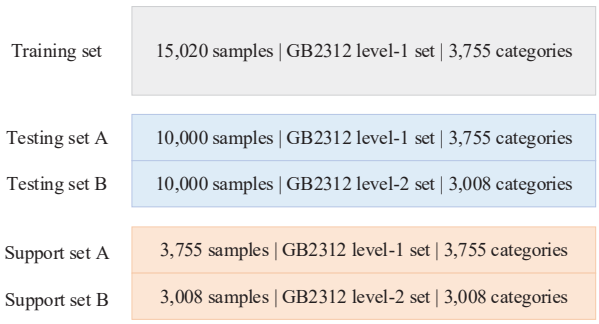


Fig. 3: Setting of CICC dataset.

Fig. 4: Examples of Chinese CAPTCHA images in this dataset.

In this dataset, there are only four images for each category in the $\mathcal{D}_{train}$, which is a problem of few-shot learning. The $\mathcal{D}^B_{test}$ contains Chinese characters that not appear in the $\mathcal{D}_{train}$ but appear in the $\mathcal{D}^B_{support}$, which is a one-shot learning problem. All samples in this dataset are 48×48 RGB images, created by Microsoft YaHei font (Figure 4). In order to increase the difficulty of recognition, we added random points and lines to the image. Also, the color, position, and rotation angle of these samples are randomly generated. We hope to promote the development of one-shot recognition and few-shot recognition technology of Chinese CAPTCHAs by this dataset.

4 Method

4.1 Deep Residual Siamese Network

Inspired by the residual structure of ResNet [12], this paper designs the CNN subnetworks based on the deep residual network. Figure 5 depicts the main structure of our subnetworks. In this figure, each stage is composed of k bottleneck units [13], where k is a

Fig. 5: Structure of Feature Extraction Network.

© The Institution of Engineering and Technology 2015

hyper-parameter we designed to control the depth of the network. When $k = 1$, $k = 2$, and $k = 3$, the model is named SiameseCCR-15, SiameseCCR-27, and SiameseCCR-39, respectively. After each convolutional layer, ReLU is used as the activation function, and the BN layer is used for batch normalization. At the end of this subnetwork, a n-dimensional feature vector is obtained by using a fully-connected layer. Because this subnetwork can extract features and reduce dimensions of images, it is called *Feature Extraction Network* in Figure 2.

4.2 Learning and Prediction

Different from the traditional methods, our method regards Chinese CAPTCHA recognition as a binary classification task, which is to take a pair of images as input and judge whether they are the same Chinese characters. In the definition of Figure 2, x^1 and x^2 represent a pair of inputs. If the extracted feature is expressed as $v = F(x)$, then the feature representation of the two inputs are $v^1 = F(x^1)$ and $v^2 = F(x^2)$. Instead of using a distance function, we use a fully-connected layer (i.e., Similarity Network $S(\cdot, \cdot)$) to compute the similarity:

$$S(v^1, v^2) = \sigma(|v^1 - v^2|w + b) \tag{1}$$

where w, b are weights and biases of the fully-connected layer, and σ is the sigmoid activation function.

Next, we assume that p represents the prediction result of this Siamese network:

$$p(x^1, x^2) = S(F(x^1), F(x^2)) \tag{2}$$

Labels of data in this network have to meet the condition $y \in \{0, 1\}$, where zero indicates that the two images contain different characters, defined as a negative pair; Otherwise, it's a positive pair. Therefore, we use binary cross-entropy as the loss function:

$$L(x^1, x^2, y) = ylogp(x^1, x^2) + (1 - y)log(1 - p(x^1, x^2)) \tag{3}$$

Table 1 Structure of SiameseCCR

Layer	Output Size	Structure
input	$44 \times 44 \times 1$	44×44 gray
conv1	$44 \times 44 \times 64$	$7 \times 7, 64$
stage2	$22 \times 22 \times 128$	3×3 max pool, stride 2 $\begin{bmatrix} 1 \times 1, 32 \\ 3 \times 3, 32 \\ 1 \times 1, 128 \end{bmatrix} \times k$
stage3	$11 \times 11 \times 256$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times k$
stage4	$6 \times 6 \times 512$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times k$
stage5	$3 \times 3 \times 1024$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times k$
pool	$1 \times 1 \times 1024$	global average pool
fc	n	n-d fc
output	1	abs, 1-d fc, sigmoid

At test time, first of all, suppose we need to classify a CAPTCHA image into one of C categories and use C^* to represent the recognition result. Then, combine the CAPTCHA image x with all images in template image library X, and query the network using $\{x, x_i\}$ as input pair. Finally, the Chinese character template with the highest similarity is selected as the classification result:

$$C^* = \arg\max_i p(x, x_i), x_i \in X \tag{4}$$

To speed up the prediction and avoid repeated calculation, we extract the template images in support set as feature vectors in advance, then combine the input vector v with all vectors in the template vector library V in pairs. The optimized Chinese CAPTCHA recognition results can be expressed as:

$$C^* = \arg\max_i S(v, v_i), v_i \in V \tag{5}$$

4.3 Training Algorithm

In the training of Siamese networks, the selection of sample pairs is crucial, which determines whether the network can converge to a high recognition accuracy [14, 15]. If the positive and negative pairs are randomly sampled from $\mathcal{D}_{train}$, then $C^1_{3755}C^1_4C^1_4 = 60,080$ positive pairs (allowing repeated sampling) and $C^2_{3755}C^1_4C^1_4 = 225,540,320$ negative pairs can be generated. By comparison, negative pairs are far more than positive pairs. To alleviate the problem of class imbalance, we propose the following training algorithm for the Chinese CAPTCHA recognition task, which is based on the idea of hard negative mining [14, 16].

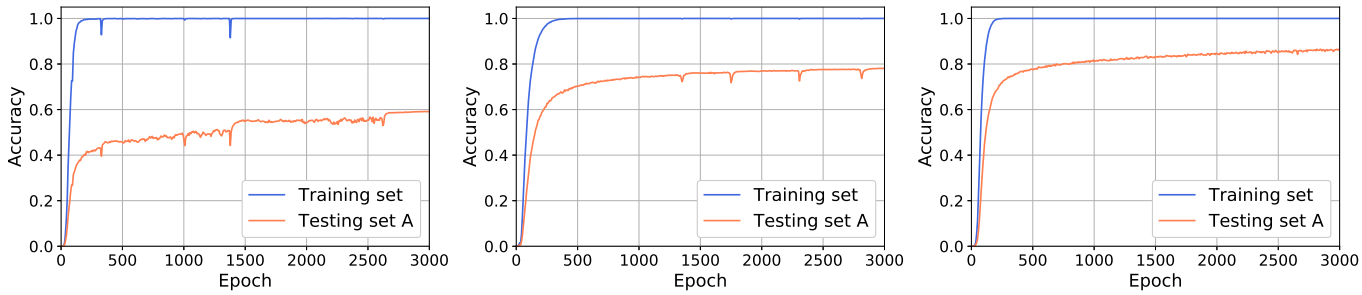
In this algorithm, first of all, we randomly construct a set of positive and negative pairs with the same amount ρ, then train the SiameseCCR model for ω epochs until the matching accuracy tends to 100%. After that, test the model on the training set and mark 10 mismatched characters but with the highest similarity for each character, which is called "top-10 errors". Then, the negative pairs are reconstructed according to the top-10 errors, that means each character can only be combined with the characters in the top-10 errors, which reduces the construction space of the negative pairs. Using this strategy, the number of positive pairs is still $C^1_{3755}C^1_4C^1_4 = 60,080$, but the number of negative pairs becomes $C^1_{3755}C^1_{10}C^1_4C^1_4 = 600,800$. In summary, this method mines the hard negative examples with the largest loss values from the massive negative pairs, which effectively improves the performance of our model.

Algorithm 1 Training Algorithm

Input: Training set $\mathcal{D}_{train}$, number of sample pairs for each iteration ρ, number of epochs for each iteration ω,.
Output: a SiameseCCR model.

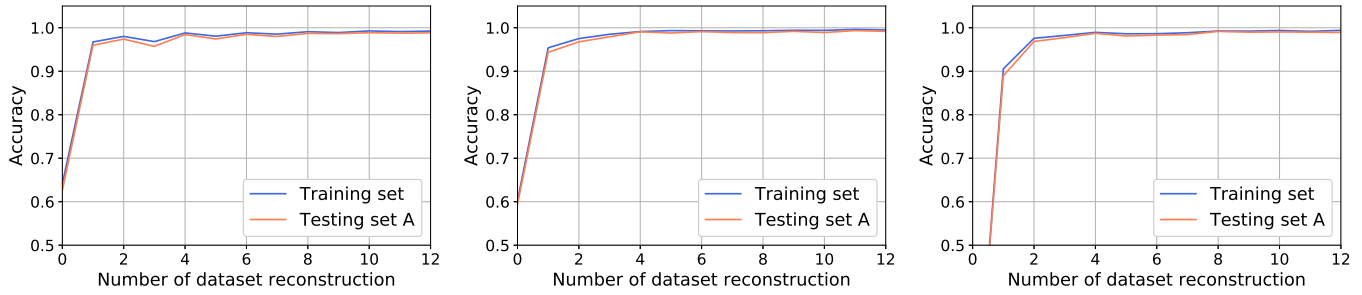
```
M ← randomly initialize a SiameseCCR model
D^pos_train ← randomly select ρ positive pairs from D_train
D^neg_train ← randomly select ρ negative pairs from D_train
while not converge do
    for epoch = 1 to ω do
        M ← train the model using D^pos_train and D^neg_train
    end for
    D'_train ← test on D_train and mark 10 mismatched characters but with the highest similarity for each character.
    D^pos_train ← randomly select ρ positive pairs from D_train
    D^neg_train ← randomly select ρ negative pairs from D'_train
end while
return the converged model M
```


© The Institution of Engineering and Technology 2015



(a) method in [4] (b) method in [5] (c) method in [6]

Fig. 6: Experiment results of few-shot recognition using traditional methods. It can be seen from the Figures 6(a), 6(b), and 6(c) that there is a large gap between the accuracy of the training set and the testing set, which indicates the severe overfitting of the traditional methods. Although the above three models were trained with 3,000 epochs, their performance was still unsatisfactory.



(a) SiameseCCR-15 (n=256) (b) SiameseCCR-27 (n=256) (c) SiameseCCR-39 (n=256)

Fig. 7: Experiment results of few-shot recognition using SiameseCCR. Figure 7(a), 7(b), and 7(c) show that our method not only achieved high accuracy but also alleviated the overfitting problem. In these figures, a dataset reconstruction is to train 10 epochs on 300,000 positive pairs and 300,000 negative pairs, which takes about 4-7 hours. The specific training time is shown in Table 2.

5 Experiments and Result Analysis

5.1 Implementation Details

In our method, k is a hyper-parameter we designed to control the depth of the network. When $k = 1$, $k = 2$, and $k = 3$, the model is named SiameseCCR-15, SiameseCCR-27, and SiameseCCR-39. Another important hyper-parameter n, the size of feature vector, is set to $n = 256$. Figure 8 shows the highest accuracy under different vector sizes, based on SiameseCCR-27, from which it can be seen that $n = 256$ is a suitable choice.

At training time, the value of parameter ρ we used is 300,000, which is half of the total number of negative pairs at resampling, taking both speed and performance into account. Moreover, the parameter ω is set to 10, which ensures that the model can converge after every dataset reconstruction. For comparison, we implement the algorithms of Jia et al. [4], Lin et al. [5], and Liu et al. [6]. All the models are trained on an Nvidia Tesla K80 GPU.

Besides, to take full advantage of the training set, we randomly crop a square region (44×44) from the images (48×48) for data augmentation. Next, all the RGB images are converted to grayscale images, to mitigate the impact of color and reduce the amount of calculation.

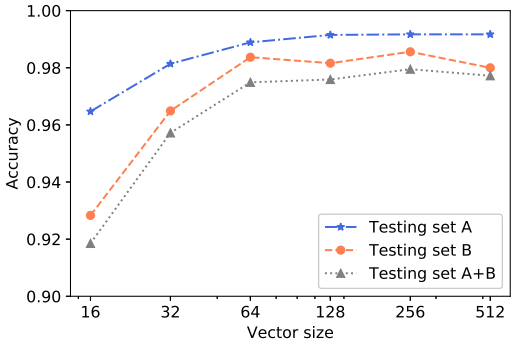


Fig. 8: Accuracy of different vector sizes.

5.2 Experiment of Few-shot Recognition

In the experiment of few-shot recognition, we use the $\mathcal{D}_{train}$ with 15,020 Chinese character CAPTCHAs to train, and use the $\mathcal{D}^{A}_{test}$ with 10,000 Chinese character CAPTCHAs to test. We compared our method with three related works. The models proposed by Liu et al. [6] and Lin et al. [5] are LeNet-5-like networks, consisting of three convolutional layers and two fully-connected layers. The model proposed by Jia et al. [4] is an 11-layer CNN, which consists of 10 convolutional layers and one fully-connected layer. In our experiment, we reproduce the above models according to the structures described in their papers and carry out experiments on the CICC dataset. The results are shown in Table 2.

The following observations can be obtained from the first three experiments in Table 2: 1) These three traditional methods can achieve excellent accuracy with a large-scale dataset, but in the case of few-shot learning, the deficiency of training data results in the unsatisfactory accuracy; 2) Overfitting is a serious problem in these networks. In the best case [6], there is still a gap of more than 13% between the accuracy of the testing set and the training set.

Compared with traditional methods, our method has many advantages. On the one hand, our method alleviates the overfitting problem, achieving state-of-the-art performance in top1, top5 and top10

© The Institution of Engineering and Technology 2015

Table 2 Performance of few-shot learning

Models	Ref.	Acc. ($\mathcal{D}_{train}$) Top1	Acc. ($\mathcal{D}_{test}^{A}$) Top1	Top5	Top10	Memory	Train Time	Test Speed ($\mathcal{D}_{test}^{A}$)
Jia et al.	[4]	100.00%	59.12%	78.08%	83.18%	120.90MB	8.28h	6.39ms
Lin et al.	[5]	100.00%	78.15%	90.18%	92.64%	28.91MB	2.68h	4.75ms
Liu et al.	[6]	100.00%	86.19%	94.82%	96.31%	35.11MB	3.56h	4.74ms
SiameseCCR-15 (n=256)	ours	98.02% 99.23%	97.39% 98.82%	99.94% 99.94%	99.96% 99.96%	11.74MB	13.54h 58.72h	14.01ms
SiameseCCR-27 (n=256)	ours	97.51% 99.50%	96.76% **99.17%**	99.89% **99.99%**	99.96% **100.00%**	14.44MB	17.65h 76.48h	16.39ms
SiameseCCR-39 (n=256)	ours	97.57% 99.40%	96.84% 98.93%	99.89% 99.97%	99.96% 100.00%	23.14MB	22.47h 97.28h	19.26ms

Table 3 Performance of one-shot learning

Models	Acc. ($\mathcal{D}_{test}^{B}$) Top1	Top5	Top10	Acc. ($\mathcal{D}_{test}^{A+B}$) Top1	Top5	Top10	Train Time	Test Speed ($\mathcal{D}_{test}^{B}$)	Test Speed ($\mathcal{D}_{test}^{A+B}$)
SiameseCCR-15 (n=256)	95.23% 98.10%	99.72% 99.93%	99.91% 99.98%	94.08% 97.31%	99.66% 99.92%	99.88% 99.95%	13.54h 58.72h	12.51ms	20.62ms
SiameseCCR-27 (n=256)	94.41% **98.59%**	99.48% **99.96%**	99.78% **100.00%**	92.63% **97.95%**	99.34% **99.94%**	99.74% **99.99%**	17.65h 76.48h	14.93ms	23.32ms
SiameseCCR-39 (n=256)	93.40% 97.99%	99.42% 99.87%	99.81% 99.97%	92.71% 97.65%	99.39% 99.87%	99.75% 99.95%	22.47h 97.28h	17.68ms	26.30ms

accuracy of $\mathcal{D}_{test}^{A}$. On the other hand, our method uses less space compared to the existing methods.

When people build a Chinese CAPTCHA recognition model, if they use our method, they only need to collect few labeled samples for each class, which can realize the accuracy that traditional methods achieved using massive data, and exceedingly reduce the workload of manual labeling. Since our method is more complex than the existing ones, it requires longer training and testing time. If you don't want to spend too much time training the model, SiameseCCR-15 is recommended, which only takes 13.54 hours to achieve high accuracy of 97.39%. Or if you want to obtain the highest accuracy, SiameseCCR-27 is recommended, which takes 76.48 hours to achieve accuracy of 99.17%.

5.3 Experiment of One-shot Recognition

At present, one-shot recognition is widely used in face recognition models [15], which hope to learn the features of a person using only one face image, and don't need to be retrained due to personnel changes. For the Chinese CAPTCHA recognition model, we hope it has similar characteristics with face recognition models, to make it more flexible and expandable. Since the traditional Chinese CAPTCHA recognition methods cannot achieve this goal, we propose the SiameseCCR, which has the following two advantages: 1) This method can recognize brand-new Chinese characters without retraining; 2) This method can achieve high accuracy even if there is only one labeled sample per class.

After the experiments of few-shot recognition, we use the three models trained by $\mathcal{D}_{train}$ to perform one-shot recognition experiments. Specifically, instead of retraining these models, only matching templates used in the test are changed. To evaluate the robustness and expansibility of this method, we design the following two experiments: 1) Using $\mathcal{D}_{support}^{B}$ as matching templates, using $\mathcal{D}_{test}^{B}$ as testing set; 2) Using $\mathcal{D}_{support}^{A+B}$ as matching templates, using $\mathcal{D}_{test}^{A+B}$ as testing set. Due to the different number of matching templates, the test speed of the above two cases is different.

As can be seen from Table 3, this method can achieve more than 98% accuracy even on the brand-new GB2312-80 level-2 set. It shows that our method can learn the generality of Chinese characters, and apply the learned knowledge to recognize new CAPTCHAs. Besides, even if the matching categories of the character templates are expanded to all the characters in GB2312-80, the performance is still excellent.

In practical applications, the change of character set of Chinese CAPTCHA will invalidate traditional methods unless retrained these models with new data. However, our method uses the Siamese network, which can better cope with the change of the character set. It only needs to collect one template for each new character and does not need to retrain the model, which has substantial flexibility and expandability.

6 Conclusion

We propose a novel deep Siamese network-based model for few-shot and one-shot Chinese CAPTCHA recognition tasks. The model differs from existing models in that it learns the similarity of image pairs, which alleviated the overfitting problem caused by data deficiency. Besides, the model also can recognize new patterns; that is, this model only needs one labeled sample for each brand-new category of Chinese characters to identify successfully. Experiments show that it is able to build a Chinese CAPTCHA recognition system with high accuracy and extensibility without a large-scale dataset.

7 Acknowledgments

This research was partially supported by the Zhejiang Provincial Natural Science Foundation of China under Grant No.LGF18F020011, and the National Natural Science Foundation of China under Grant No.61803337.

8 References

1 Hu, J., Ma, W., Khan, A., Liu, L.: 'Recognizing character-matching captcha using convolutional neural networks with triple loss'. International Conference on Knowledge Science, Engineering and Management. Springer, 2018. pp. 209–220
2 Du, F.L., Li, J.X., Yang, Z., *et al.*: 'Captcha recognition based on faster r-cnn'. International Conference on Intelligent Computing. Springer, 2017. pp. 597–605
3 An, G., Yu, W.: 'Captcha recognition algorithm based on the relative shape context and point pattern matching'. 2017 9th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA). IEEE, 2017. pp. 168–172
4 Jia, Y., Fan, W., Zhao, C., Han, J.: 'An approach for chinese character captcha recognition using cnn', Journal of Physics: Conference Series. vol. 1087. IOP Publishing, 2018. p. 022015

© The Institution of Engineering and Technology 2015

5 Lin, D., Lin, F., Lv, Y., Cai, F., Cao, D.: 'Chinese character captcha recognition and performance estimation via deep neural network', *Neurocomputing*, 2018, **288**, pp. 11–19
6 Liu, H., Shao, W., Guo, Y.: 'Research on captcha recognition with convolutional neural networks', *Computer Engineering and Applications*, 2016, **52**, (18), pp. 1–8
7 Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R.: 'Signature verification using a "siamese" time delay neural network'. Advances in Neural Information Processing Systems. 1994. pp. 737–744
8 Nair, V., Hinton, G.E.: 'Rectified linear units improve restricted boltzmann machines'. Proceedings of the 27th International Conference on Machine Learning (ICML-10). 2010. pp. 807–814
9 Zagoruyko, S., Komodakis, N.: 'Learning to compare image patches via convolutional neural networks'. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015. pp. 4353–4361
10 Koch, G., Zemel, R., Salakhutdinov, R.: 'Siamese neural networks for one-shot image recognition'. ICML Deep Learning Workshop. vol. 2. Lille, 2015.
11 Sung, F., Yang, Y., Zhang, L., *et al.*: 'Learning to compare: relation network for few-shot learning'. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018. pp. 1199–1208
12 He, K., Zhang, X., Ren, S., Sun, J.: 'Identity mappings in deep residual networks'. European Conference on Computer Vision. Springer, 2016. pp. 630–645
13 He, K., Zhang, X., Ren, S., Sun, J.: 'Deep residual learning for image recognition'. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016. pp. 770–778
14 Harwood, B., Kumar, B., Carneiro, G., *et al.*: 'Smart mining for deep metric learning'. Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2017. pp. 2821–2829
15 Schroff, F., Kalenichenko, D., Philbin, J.: 'Facenet: a unified embedding for face recognition and clustering'. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015. pp. 815–823
16 Wan, S., Chen, Z., Zhang, T., Zhang, B., Wong, K.k.: 'Bootstrapping face detection with hard negative examples', *arXiv preprint arXiv:160802236*, 2016

© The Institution of Engineering and Technology 2015