Journal of Mathematics and Informatics
Vol. 20, 2021, 29-38
ISSN: 2349-0632 (P), 2349-0640 (online)
Published 17 March 2021
www.researchmathsci.org
DOI: http://dx.doi.org/10.22457/jmi.v20a04190

Information Security Analysis of Chinese Character Captcha Based on Siamese Neural Network

***Haisheng Song*[1], *Pengfei Duan*[1*] and *Riying Qiao*[2]**

[1]College of Physics and Electronic Engineering, Northwest Normal University
Lanzhou, Gansu 730070, China.
[2]Information Security and Risk Control Department, Wonders Information Co., Ltd.
Shanghai, 200000, China.
[*]Corresponding author. Email: 875497622@qq.com

Received 15 February 2021; accepted 9 March 2021

Abstract. CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Humans Apart) plays a pivotal role in the network security of information systems, and has been widely used in various fields. This paper uses Tensor Flow as the neural network framework and uses Siamese network to reduce noise information and extract the characteristic information of the captcha. The experimental results show that the recognition rate of the Siamese neural network for the four-character image captcha with noise interference is 97.40%. Therefore, the siamese neural network has a strong recognition ability for complex captchas. At the same time, this high recognition ability leads to serious security risks in the information system. Therefore, the current Chinese character captcha design technology still needs to be strengthened and improved.

Keywords: Information security; Siamese neural network; Captcha; Accuracy; Loss function.

AMS Mathematics Subject Classification (2010): 62H35

1. Introduction

Information systems play a very important role in network security. With the development of services, various websites are constantly being updated and expanded to varying degrees. Once the information system is attacked and the system becomes unavailable, it will lead to the specific services carried by the system. Failure to proceed smoothly, seriously disrupting the normal operation order, causing the reputation and actual loss of the attacked party. Therefore, in order to ensure the security of external services of relevant systems, information security engineers are born for the application of captcha technology. The captcha is a public and fully automatic program that distinguishes the user from being a machine or a human [1, 2]. In the CAPTCHA test, the computer as the server automatically generates a question for the user to answer. This question can be generated and judged by the computer, but only humans can answer it. The captcha technology can prevent malicious password cracking, ticket swiping, and forum irrigation, effectively preventing a certain registered user from using a specific program to brute force the continuous login attempts. Therefore, whether the captcha can

be recognized by the computer has become an eternal subject of network security analysis.

In order to detect the security and reliability of the captcha, the information security engineer conducts a security assessment on the target system specified by the client unit from a practical perspective, allowing the relevant personnel of the client unit to intuitively understand the hidden vulnerabilities and vulnerabilities in their own networks, systems, and applications. The loss that may be caused when the hazard occurs, so as to help the client unit to explain the current security status in the form of actual cases, thereby increasing the client unit's awareness of information security, enhancing the risk crisis awareness of the client unit's personnel, so as to achieve the internal security level The overall improvement is now an overall improvement in the internal security level.

In 2002, Mori and Malik [3] proposed a method based on shape context. By using contour and shape features to construct a classification system, they successfully broke the two captchas, Gimpy and EZ-Gimpy, with a success rate of 33% and 33% respectively. 92%. In 2004, the Moy team [4] also cracked the EZ-Gimpy captcha with 99% accuracy. The algorithm used is to use distortion estimation technology to identify characters with background interference and distortion. At the same time, the author also cracked The four-character version of the Gimpy-r captcha has an accuracy rate of 78%. In 2005, Chellapilla [5] used a variety of image processing techniques to segment characters, and used machine learning to recognize the segmented characters, and finally successfully recognized 6 types of text captchas, with an accuracy rate of 4.89% to 66.2%. between. Therefore, it can be seen that the early research methods used uncommon and complex algorithms to solve the captcha identification problem one by one. In the mid-stage of the captcha research, in 2007, Jeff Yan and El Ahmad [6] proposed an extremely simple idea. First, the captcha of Captchaservice.org was divided into individual characters, and then the total number of pixels for each character was counted. To determine which character corresponds to, the method finally achieves a recognition accuracy of close to 100%. In 2008, the same author [7] proposed two new algorithms for segmenting individual characters in order to improve the accuracy of segmenting characters. One is Color Filling Segmentation (CFS) and the other is vertical histogram. analysis. However, the method proposed by the author cannot be applied to the captcha of Crowding Characters Together (CCT) mechanism. In the recent stage of captcha research, image processing methods are still the main algorithm for segmenting characters, and after the rise of deep convolutional neural networks, more and more are used to identify individual captcha characters. In 2015, Colin Hong [8] used image processing methods to first process the Microsoft captcha image into a single character picture, and then used deep convolutional neural networks (Convolutional Neural Networks, CNNs) and template matching methods to identify characters, and finally Recognition accuracy rates of 57.05% and 5.56% are reached respectively. The comparison results show that the convolutional neural network is effective in the classification of interfering characters. In 2015, the Starostenko team [9] used morphological analysis and three-color algorithm for character segmentation, and then used Support Vector Machine (SVM) to recognize the separated characters, and finally cracked the accuracy of the four captchas. Between 40.4% and 94.3%. In 2017, the Gao team[10] broke through the Microsoft double-layer captcha. First, based on the CFS image processing technology, the double-layer captcha was

horizontally divided into two single-layer captchas, and then the high-wave filter was used to process each captcha separately. On the first floor, the final recognition accuracy rate using CNNs is 44.6%. In 2018, Gao team [11] cracked 11 kinds of online captchas, and the cracking method was also based on the algorithm of segmentation and recognition. Design different segmentation algorithms for different captchas, and then use a LeNet-5 based network for recognition. The final cracking accuracy is between 10.1% and 90.0%, and the speed is between 0.03 and 0.65 seconds. For more information about the new developments as well as the history of this topic, see the papers by [14-16].

From the above literatures it can be found that the captcha has been identified by various computer technologies, which poses a serious threat to the security of our information system. Siamese neural network (Siamese neural network), also known assiamese neural network, is a coupling structure based on two artificial neural networks (refer to 12, 13). In order to fully protect the network security of the information system, in this paper, we will detect the pairing of the Siamese neural network. The identification effect of the captcha, and the analysis of the network security strategy.

The arrangement of the article is as follows: the second part introduces the siamese neural network model, the third part introduces experimental data, the fourth part gives the experimental results and finally gives a conclusion.

2. Siamese network

2.1. Model introduction

In the siamese network (refer to [12, 13]), the loss function used is contrastive loss, which can effectively deal with the relationship of paired data in the siamese network. The expression of contrastive loss is as follows:

$$L\left(W,\left(Y,X_1,X_2\right)\right)=\frac{1}{2N}\sum_{n=1}^{N}YD_W^2+\left(1-Y\right)\max\left(m-D_W,0\right)^2,$$

where

$$D_w(X_1,X_2)=\| X_1-X_2 \|_2=\left(\sum_{i=1}^{p}\left(X_1^i-X_2^i\right)^2\right)^{1/2}$$

represents the Euclidean distance (two norm) of the two sample features X_1 and X_2, P represents the feature dimension of the sample, Y is the label of whether the two samples match, Y=1 represents the two samples are similar or matched, Y = 0 It represents a mismatch, m is the set threshold, and N is the number of samples.

Observing the above-mentioned contrastive loss expression, it can be found that this loss function can express the matching degree of paired samples very well, and it can also be used to train the model for extracting features.

When Y = 1 (that is, when the samples are similar), the loss function is only left

$$L_S=\frac{1}{2N}\sum_{n=1}^{N}YD_W^2.$$

That is, when the samples are not similar, if the Euclidean distance of the feature space is small, the loss value will become larger, which also symbolizes our requirements.

When Y=0 (that is, when the samples are not similar), the loss function is

$$L_D = \frac{1}{2N}\sum_{n=1}^{N}(1-Y)\max(m-D_W,0)^2 .$$

That is, when the samples are not similar, if the Euclidean distance of the feature space is small, the loss value will become larger, which also symbolizes our requirements.

2.2. Algorithm design

(1) The captcha identification in this article uses a Siamese network with a similar structure. The Siamese network is generally suitable for two input data with similar characteristics. For example, in our paper, we need to compare the similarity of two Chinese character pictures. Degree, when the two inputs of the neural network (Input 1 and Input 2) are used as inputs, the characteristics of the two input data will be mapped to the new space, the two network models use their own weight parameters and bias parameters, and finally The updated loss weight value is shared, and finally the similarity of the two inputs is calculated by the Euclidean distance formula, so each Chinese character will be calculated by the sigmoid function to form a probability value. After the probability value is calculated by the soft max formula, we will get each group. The predicted probability value of nine vector values, so that each of our captchas will become a nine-category multi-category problem. The above picture is the technical diagram of the CNNsiamese network designed in this paper. Oursiamese network uses the loss value of the contrast loss function, where Dw is the Euclidean distance of the output features of the two networks

$$\sqrt{\{G_W(X_1)-G_W(X_2)\}^2}$$

where Gw is the output of one of the sister networks. X1 and X2 are input data pairs. The Y value is 1 or 0. If the model predicts that the inputs are similar, then the value of Y is 0, otherwise Y is 1. max() is a function representing the larger value between 0 and m-Dw. m is a margin value greater than 0. Having a marginal value means that different pairs beyond that marginal value will not cause loss. This makes sense, because you only want to optimize the network based on actual dissimilar pairs, but the network thinks it is quite similar. The ultimate goal of the training of our two neural networks is to make the distance between similar or identical vectors as small as possible, and the similarity of

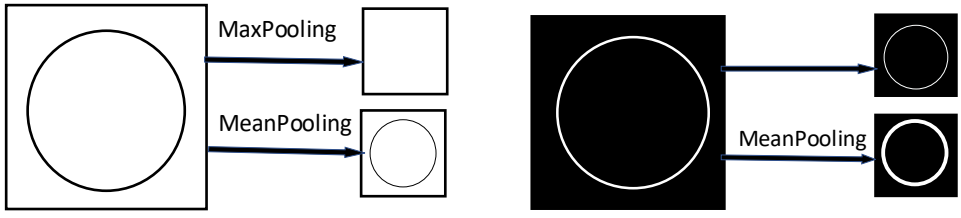

Figure 1: Compare the advantages of using the maximum pooling method.

input vectors of different categories is as large as possible.

We use Maxpooling in the siamese network model as shown in (a), and do not use average pooling as shown in (b). Generally speaking, when there is a lot of noise and interference information in the background of the image, Only part of the information in the features is more useful. The CNN convolution kernel extracts the features of the

picture. We use the maximum pooling process to down sample the feature block Patch, and extract the maximum value in each block according to the operation process of maximum pooling. The other blocks will be replaced by the largest value, and only the largest value will enter the next layer, because every time the reverse gradient updates the weights and parameters, the largest element may not be in the original position, so choose the largest The value pool has the function of extracting main features and highlighting the foreground. When all the information in the feature patch is more useful, we use the average pooling process to down sample the feature patch. When the CNN network enters the last few layers and reaches the fully connected layer of the network, use average pooling (Mean Pooling) down-sampling the feature block patch, so that the feature information will not only retain local information before entering the fully connected layer, but also contain most of the global information, so that the model can reach a state of convergence as much as possible.

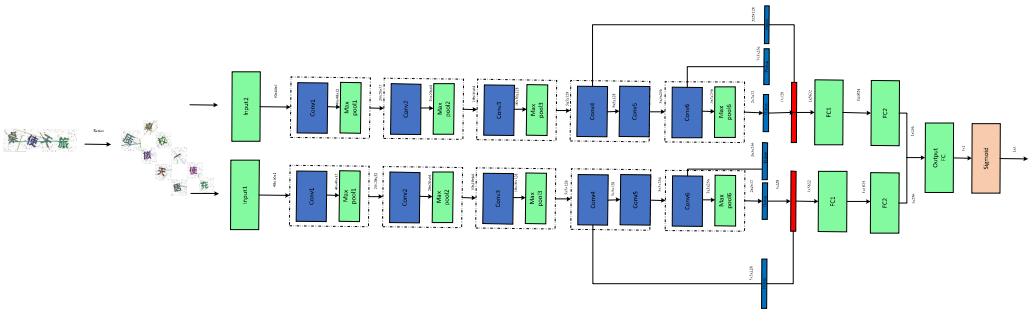

Figure 2: Siamese neural network in (1).

(2) In digital captcha recognition model, there are 10 types of numbers and 3 types of operators to be recognized, which is a 13 classification problem. After the noise reduction process, the complexity of the image is greatly reduced, and the noise and interference lines are almost completely eliminated. Using a simple three-layer fully connected neural network and a Sigmoid layer, the test set accuracy rate of 99.98% can be achieved. In this network model, the dropout value is 0.7, the learning rate is 0.001, tanh is used as the activation function, and the Adam algorithm is used for training.

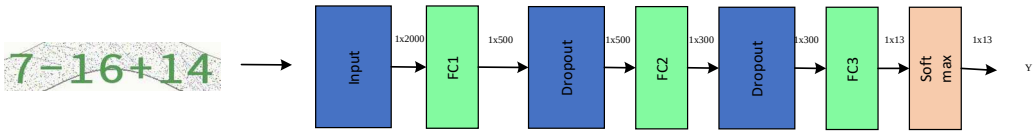

Figure 3: Convolutional neural network in (2).

3. **Data processing**

In this paper, we used three types of CAPTCHA images given by the organizing committee in the 9th China University Student Service Outsourcing Innovation and Entrepreneurship Competition. The Dataset1 is a five-character CAPTCHA, which is consists of only 0-9 ten digits and four Arithmetic operations, This kind of CAPTCHA is relatively easy to extract, which causes the weakness of relatively fragile security. The Dataset2 is Chinese Captchas which are randomly composed four Chinese characters, and

the text in the picture is slant by rotation. The verification method requires the user to point the rotated characters in the image captcha. Dataset 3 is Captcha composed Chinese characters, which are randomly rotated by 90° and have background noise interference, The verification method requires the user to point the number sequence of the Chinese characters in the image captcha from left to right. The following figure shows examples of traditional captcha and adding noise interference captcha. According to the verification method above, the sequence should be: 2580.

Table 1: Train datasets

Type	Train set	Validation set	Test set
Dataset1	10000	10000	5000
Dataset2	10000	10000	5000
Datase3	10000	10000	5000

In order to point the correct sequence, we must first ensure that the trainset contains enough Chinese characters. Our trainset has 10,000 Chinese captcha, which is equivalent there are 10000*4=40,000 feature-map information that our Siamese network needs to learn, Finally, and our model will generate a mapping.txt file for the identification of our subsequent test data. the shortest distance between the two input matrices is calculated through the Euclidean distance formula, and the Sigmoid classifier predicted the value that has the largest probability similar to Chinese character.

Table 2: Types of captcha

Type of captcha	Examples of captchas
Traditional captcha	
captcha for noise interference	
Noise interference and sorted captcha	

4. Experimental process and results

4.1. Experimental environment and experimental parameter configuration

We used 16.04-Ubuntu operating system, Intel Xeon Silver 4110 processor, and our running memory is 32GB that is helpful to handle massive datas, GPU is NVIDIA-SMI 410.48, 8GB Video memory, our experiments were completed on Tensor Flow, which is modular, minimal and flexible framework.

4.2. Experimental calculation process

Algorithm 1: As the Figure 2 architecture of Siamese network shows :We will input the

Datasets#4 and the Datasets#5 to the Siamese Network separately.

(1) Before inputting the pixel value of 150x45 pxiel Chinese character captcha into the network model, the size of the Chinese character captcha is processed into 40x40 pxiels by means of slicing.

(2) The Siamese Network in the experiment includes two input layers, six convolutional layers, five MaxPooling layers, three Flatten layers, one Concat layer, three Full-Connected layers, one Sigmoid layer, and the output passes The probability value of similarity between two input values calculated by Euclidean distance. (As shown in Figure X)

(3) The C1 convolution layer contains 32 40x40 feature maps; behind C1 is the MaxPooling P1, which is a 2x2 kernel-size and used for image dimensionality reduction and down sampling.

(4) The C2 layer takes P1 as input, The C2 convolution layer contains 64 20x20 feature maps, convolution kernel size is 3x3; behind C2 layer is the MaxPooling P2, which is a 2x2 kernel-size and used for image dimensionality reduction and down sampling.

(5) The C3 layer takes P2 as input, The C3 convolution layer contains 128 10x10 feature maps, convolution kernel size is 3x3; behind C3 layer is the MaxPooling P3, which is a 2x2 kernel-size and used for image dimensionality reduction and down sampling.

(6) The C4 layer takes P3 as input, C4 layer contains 128 5x5 feature maps, and the convolution kernel size is 3x3; C5 takes C4 as an input, and C5 contains 256 5x5 feature maps, and its convolution kernel size is 3x3.

(7) The C6 layer takes P5 as input, The C6 convolution layer contains 256 5x5 feature maps, convolution kernel size is 3x3; behind C6 layer is the MaxPooling P6, which is a 2x2 kernel-size and used for image dimensionality reduction and down sampling.

(8) The Flatten1 layer uses the P6 layer as input, and the Flatten1 layer will convert P6 into a 1x128-dimensional numpy value.

(9) The FC layer of the model on both sides of the Siamese Network is converted into a 1x1 numpy value through the Output FC layer. The feature value is passed through the Sigmoid layer to obtain the similarity probability value of Input1 and Input2. the Siamese Network training model will extract the characteristics of Chinese characters as A feature text, saved in the mappinng.txt file, used as a feature extraction comparison in the subsequent testing process.

(10) Finally, the output layer added to the Sigmoid layer, and obtain the similarity probability value, which has fewer parameters and effectively prevents data overfitting, making the training results as convergent as possible.

Algorithm 2: As the Figure 3 the architecture of the convolution network shows :

(1) The origin digital captcha are Inputted Input layer. The FC1 is a 1x5000 shape, The Droupt1 layer uses the FC1 layer as input, which effectively prevents data overfitting.

(2) The FC2 layer uses the Dropout layer as input. The FC2 layer is a 1x300 shape, The Droupt2 layer uses the FC2 layer as input, which effectively prevents data overfitting.

(3) Finally, the Sigmoid layer will output 13 types characters, and we can obtain the

probability of every character.

4.3. Experimental results

An images and results of the experimental results is given below.

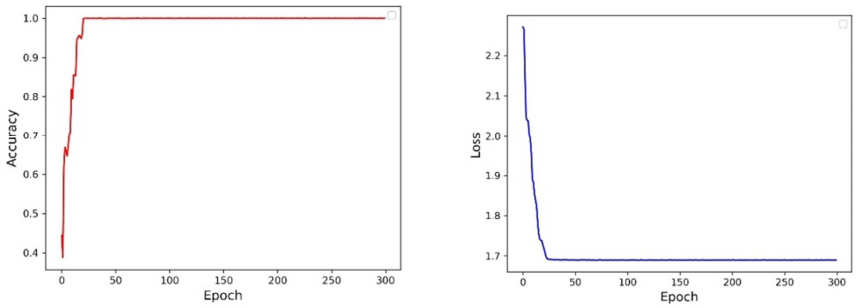

Figure 4: Testing accuracy and loss of Dataset1.

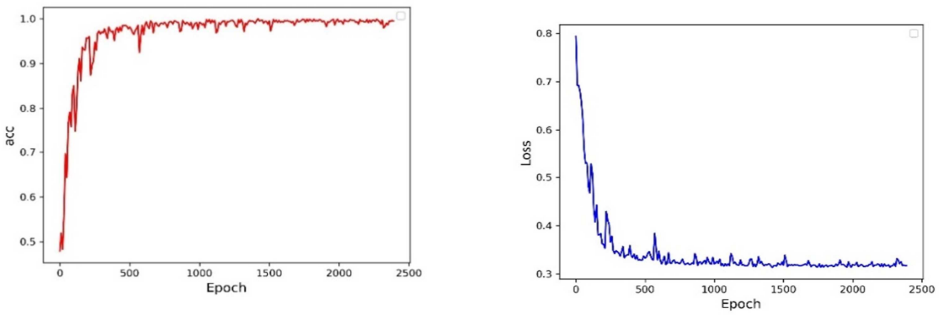

Figure 5: Testing accuracy and loss of Dataset2.

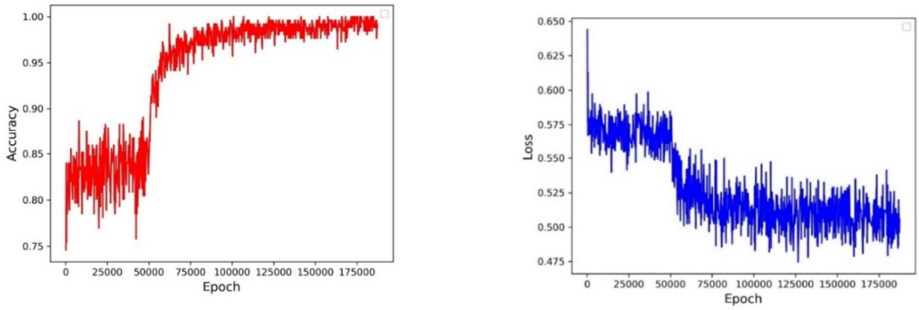

Figure 6: Testing accuracy and loss of Dataset3.

Table 2: The Testing accuracy and Training accuracy of Dataseta1.

Epoch	Validation set	Test set	Loss
50	1.0000	0.9998	1.689

Table 3: The Testing accuracy and Training accuracy of Dataseta2.

Epoch	Validation set	Test set	Loss
400	0.9805	0.9719	0.3313
1000	0.9961	0.9922	0.3214
2000	0.9883	0.9971	0.3172

Table 4: The Testing accuracy and Training accuracy of Dataseta3.

Epoch	Validation set	Test set	Loss
10000	0.8516	0.8125	0.5735
50000	0.9350	0.9141	0.5277
70000	0.9827	0.9744	0.5202

The above tables show : The Dataset1 that Captcha composed of digits can already be identify with high accuracy. the final training Accuracy and Loss values on the Testdata set. When the number of training Epoch is between 0 and 60000 rounds, the model parameters have a relatively large fluctuation when they are trainned, but as the number of training rounds increases, The learning ability of the model is gradually enhanced, the accuracy of the model gradually converges and finally stabilizes, The Dataset3 test-accuracy rate is maintained at 97.4%, so it is concluded that the model maintains a relatively high recognition rate on the Testdata.

5. Conclusion

This paper uses TensorFlow as the neural network framework, based on the captcha recognition technology of the Siamese neural network, and analyzes the threat of the Siamese neural network to the captcha recognition in the security of the information system from a practical perspective. Experiments show that: the final training Accuracy and Loss values on the test data set, as the number of training rounds increases, the learning ability of the model gradually increases, the accuracy of the model gradually converges, and finally stabilizes, and the accuracy is maintained at 97.4%, so the captcha recognition technology of the Siamese neural network maintains a relatively high recognition rate on the test data. However, it is precisely because of this high recognition rate that has brought serious hidden dangers to information security, and the current captcha design urgently needs to be strengthened and improved.

Acknowledgement. The authors thank the insightful and helpful comments and suggestions from an Editor and an anonymous reviewer, which have greatly improved the presentation of the paper. Haisheng Song and Pengfei Duan acknowledge the financial support of the National Natural Science Foundation of China (No. 11705224 and No. 11805213). Pengfei Duan thanks all the encouragements and helps of Haisheng Song and Riying Qiao during for his graduate school years.

REFERENCES

1. Von Ahn, Luis, et al. CAPTCHA: Using hard AI problems for security, International conference on the theory and applications of cryptographic techniques, Springer, Berlin, Heidelberg, 2003.

2. Von Ahn, Luis, Manuel Blum and John Langford, Telling humans and computers apart automatically, *Communications of the ACM*, 47(2) (2004) 56-60.
3. Greg Mori and J.Malik. Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA, *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*, Vol. 1. IEEE, 2003.
4. Gabriel Moy, et al., Distortion estimation techniques in solving visual CAPTCHAs, *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, Vol. 2. IEEE, 2004.
5. Chellapilla, Kumar and P.Y.Simard, Using Machine Learning to Break Visual Human Interaction Proofs (HIPs), *DBLP*, (2004) 265-272.
6. Jeff Yan and A.S.E. Ahmad, Breaking Visual CAPTCHAs with Naive Pattern Recognition Algorithms, *Computer Security Applications Conference*, 2007.
7. Jeff Yan and A.S.El Ahmad, A low-cost attack on a Microsoft captcha, *Acm Conference on Computer & Communications Security ACM*, (2008) 543.
8. Karthik, CHBL-P., and Rajendran Adria Recasens. Breaking microsoft's CAPTCHA. Technical report (2015).
9. Oleg Starostenko, et al. Breaking text-based CAPTCHAs with variable word and character orientation. *Pattern Recognition,* 48(4) (2015) 1101-1112.
10. Haichang Gao, et al. Research on the security of microsoft's two-layer captcha. I*EEE Transactions on Information Forensics and Security,* 12(7) (2017):1671-1685.
11. Sam Hocevar, PWNtcha-captcha decoder web site. http://sam. zoy. org/pwntcha (2007).
12. Jane Bromley, et al. Signature verification using a "siamese" time delay neural network, *Advances in Neural Information Processing Systems*, (1994) 737-737.
13. Davide Chicco, Siamese neural networks: An overview, *Artificial Neural Networks*, (2021) 73-94.
14. V.R.Kulli, Computing some multiplicative temperature indices of certain networks. *Journal of Mathematics and Informatics*, 18 (2020) 139-143.
15. Sheng-sheng Zhang, A model of based on Z-number and fuzzy analytic hierarchy process, *Journal of Mathematics and Informatics*, 7 (2017) 63-71.
16. Jinhui Gong, Guicang Zhang, and Kai Wang, Human iris localization combined with ant colony and improved hough circle detection, *Journal of Mathematics and Informatics*, 16 (2019) 23-39.