Verifying the Robustness of Text-based CAPTCHAs offered by Local E-Commerce Sites

Rafaqat Hussain Arain[1], Riaz Ahmed Shaikh[1], Kamlesh Kumar[2], Abdullah Maitlo[1], Asadullah Kehar[1], Safdar Ali Shah[1], Hidayatullah Shiakh[1]

[1]Department of Computer Science, Shah Absdul Latif University, Khairpur, Sindh, Pakistan
[2]Department of Computer Science, Sindh Madressatul Islam University, Karachi, Sindh, Pakistan

Abstract
CAPTCHAs are extensively used on the internet in order to distinguish between humans and bots. Numerous design alternatives of CAPTCHAs are introduced in last decade or so but still Text-based CAPTCHAs are most prevalent on the web due to its easy implementations. Text-based CAPTCHAs offer distorted text in images. The users are asked to read the text in order to prove them as humans. Although this is a trivial task for humans but it is still difficult for machines to decode such distorted characters with background clutter. The main challenges involved are preprocessing, segmentation and recognition. In this research, we have attempted to verify the robustness of CAPTCHAs offered by local e-commerce sites of Pakistan. We have successfully decoded the CAPTCHAs offered by these sites. Our proposed algorithms achieved promising results on all attacked CAPTCHAs. Using our thresholding methods, CFS, Recognition based segmentation methods and machine learning techniques; we have successfully decoded the said CAPTCHAs with an overall precision of up to 82.4%.

Key words:
CATPCHAS, HIPs, Color Filling Segmentation, thinning, CNNs.

1. Introduction

CAPTCHA (Completely Automated Public Turing test to tell Computers and Human Apart) is a test carried out to differentiate between Humans and bots. CAPTCHAs are widely implemented on the web in order to stop the spam. The term CAPTCHA was introduced by Ahn et al [1]. Since its introduction CAPTCHAs were widely implemented by numerous websites to safeguard their resources from unapproved access. CATPCHAs are based on hard problems in the field of Artificial Intelligence. Due to this, they are hot research topic since its introduction. Decoding CATPCHAs is always a win-win game. In case the CAPTCHA is successfully broken then it leads to one step advancement in the field of Artificial Intelligence, otherwise it can be used as a security mechanism to stop the unauthorized automated access on the web. Research on decoding these CAPTCHAs can not only improve the text recognition capabilities of machines but it can also be used to explore the weaknesses of current designs. CAPTCHAs are ubiquitously fund on the web these days. These are offered by Email registration sites, blogs, online voting systems, and e-ticketing systems etc. Although there are many design variants of These CAPTCHAs; such as image-based, audio-based and motion-based CAPTCHAs but Text-based CAPTCHAs are still most extensively used type because of its easy implementations [2]. A representative example of a Text-based CAPTCHA is presented in Figure 1.

Text-based CAPTCHAs are typically based on following resistances against automate attacks:

1. Extraction resistance
2. Segmentation resistance
3. Recognition resistance

Extraction resistance consists of using the same colors for background and foreground and adding background noise. In Segmentation resistance, random lines, arcs, fragmented characters and character collapsing techniques are applied. In the last step of recognition of individual characters, the techniques like distortion of text, using multiple fonts and random strings are applied to prevent automated attacks [3]. The corresponding breaking techniques for said resistance techniques are preprocessing, segmentation and recognition. Problems occurred during preprocessing can lead to problems in segmentation and ultimately leads to incorrect recognition therefore each step depends on previous steps.

In this research we have attempted to verify the security of popular e-commerce sites of Pakistan. These sites are buyon.pk, BnB accessories, ishopping.pk, shopive.com and telemart.pk etc. The CAPTCHA offered by buyon.pk contains no noise and only depends on distortion of characters along with slightly merging them with each other, while all other types of CAPTCHAs offered by said sites are similar and contain noise in terms of salt and pepper noise along with random arcs. The attacked representative CAPTCHAs of said sites are shown in Figure 1. The attacked CAPTCHAs are based on extraction resistance (background noise in terms of small dots and random arcs of different lengths), Segmentation resistance (few characters are connected with each other) and recognition resistance (characters are distorted and rotated at different angles). Research has proved that no

Manuscript received September 5, 2018
Manuscript revised September 20, 2018

matter how much the distortion is applied on individual characters, it can still be recognized by using machine learning techniques. In the first step, i.e. preprocessing we have applied thresholding techniques along with binary and gray scale conversion techniques. Once the noise is removed then the next step is segmentation. During segmentation, we have applied CFS (Color Filling Segmentation) and Recognition based segmentation method in order to identify and isolate the connected characters. In the last step, i.e. recognition, we have trained a deep neural network to train and identify the individual characters. This article is divided as follows: Section 2 presents the literature review, section 3 presents the proposed algorithm, section 4 presents results while the last section, i.e. section 5 presents conclusion and future work.

Fig. 1 Examples of attacked Text-based CAPTCHAs.

2. Literature Work

Since their introduction and implementation on the web, CAPTCHAs are successfully decoded by many researchers [3][4]. Based on the weaknesses found in the previous designs the existing CAPTCHAs are not only broken but new CAPTCHAs are also proposed. In this way designers tend to propose new CAPTCHAs by exploiting weaknesses in the current design while attackers decode them by using computer vision and machine learning techniques. This friendly war not only improved the security mechanism (i.e. CAPTCHAS) but it also improves the machines' abilities of reading distorted text (in case of Text-based CAPTCHAs).

Mori and Malik have successfully broken the early Gimpy and EZ-Gimpy CAPTCHAs with high precision of 33% and 92% respectively using shape contexts of characters [4]. Chellapilla and Simard not only decoded multiple CAPTCHAs but they also proposed rules to design better HIPs (Human Interaction Proofs) [5].

It is explored by researchers that merely adding random noise and arcs does not protect against automated attacks rather it causes usability issues [5]. Yan and Ahmad have successfully decoded various CAPTCHAs by using pixel count, snake segmentation algorithm and dictionary attacks [6]. They achieved almost 100% success rate by implementing their novel methods. In a later publication they decoded the CAPTCHA offered by Microsoft [2]. They achieved promising results by using their vertical segmentation, pixel count and histogram methods. Google's CAPTCHA (v.2010) was also successfully broken by Yan and Ahmad by using shape patterns. Three color bar code proposed by Starostenko et al achieved a success rate of 40.4% and 54.6% on reCAPTCHA v.2011 and v.2012 respectively [7].

13 out of 15 popular CAPTCHAs were broken by Burstein et al [8]. Hollow CAPTCHAs were successfully solved by Gao et al at a success rate of 36% to 89% [9].

Hussain et al presented a segmentation based recognition method to break prevalent CAPTCHAs like Taobao, JD and eBay. They used artificial neural networks for segmentation and recognition. They achieved a success rate up to 62% [10]. Zhang and Wen proposed an explicit segmentation scheme where they applied a fuzzy matching method to decode CAPTCHAs [11]. Hussain et al proposed a deep learning model for the recognition of complex Text-based CAPTCHAs. A huge dataset for the training of deep neural network (CNN) was synthetically generated. They achieved a success rate of 83.3% and 86.5% on easy and complex CAPTCHAs respectively [12].

Stroke connection point with dynamic segmentation scheme was proposed by Gao et al. to break connected and disconnected CAPTCHAs with accuracy of 12% and 88.8% respectively [13].

LeCunn et al introduced the Convolutional Neural Networks (CNNs) in 1995.They used CNNs to recognize the handwritten characters [14]. CCNs are basically a class of feed forward artificial neural network. They are composed of multiple layers including convolutional layers. CNNs are also known as space invariant or shift invariant artificial neural networks due to their translation invariance characteristics [15] They used local dependencies in images in order to find image features. The convolutional layer applies a convolutional filter to convolve the input image which produces a multidimensional output. Convolutional layers are followed by fully connected layer in order to map the spatial information to the classification information and it combines the prior layer and applies non linearity to its output. A max pooling layer (along with convolutional and fully connected layer) is used to control over fitting by reducing the amount of parameters [16].

3. Proposed Method

The recognition of Text CAPTCHAs involve there stages: Preprocessing, segmentation and recognition. All these stages are interconnected; errors in one stage can lead to errors in later stages. The efficiency of recognition depends on the accuracy of segmentation and the efficiency in segmentation depends on the accuracy in preprocessing. The system diagram of our proposed method is shown in figure 2.

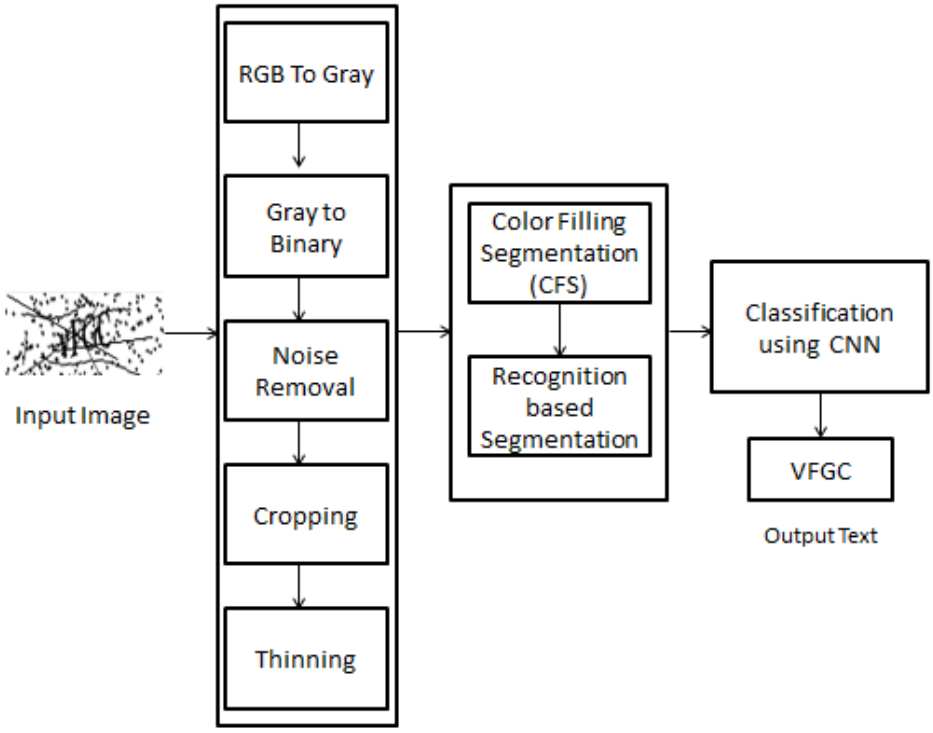

Fig. 2 The System diagram of our proposed method

3.1 Preprocessing

A number of image processing operations are performed on the given RGB CAPTCHA image in the preprocessing stage. These operations include RGB to Gray scale conversion, Gray scale to binary conversion, Noise removal, Cropping and thinning.
The given RGB image is converted into gray image using equation No. 1

$$Y = 0.2989*R + 0.5870*G + 0.1140*B \quad (1)$$

The results of RGB to gray scale conversion are shown in Figure 3.

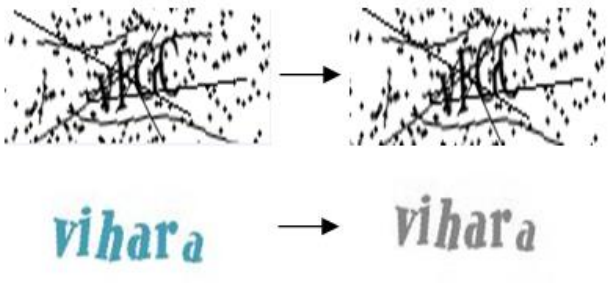

Fig. 3 RGB to Gray scale conversion

The Grayscale image is then converted to binary image by means of Othsu thresholding method [17] as shown in Figure 4.

Fig. 4 Gray to Binary conversion

The buyon.pk CAPTCHA contains almost no noise, however BnB and ishopping.pk (and other similar implementations) CAPTCHAs contain noise in terms of salt and pepper noise as well as random arcs. The next step is important because proper noise removal can affect the overall precision of our method. Here in this case noise is removed using multiple low pass filters such as median filter. Binary image with noise is taken as input. A seed

pixel P is selected and flood fill algorithm is used to label all the pixels. Similarly all the pixels are labelled. The area of each connected component is computed. A threshold value is then used to remove all the components having fewer pixels than p. For a 2D binary image, the default connectivity value 8 is used here in our case. Figure 5 shows the results before and after the noise removal steps on the given binary image.

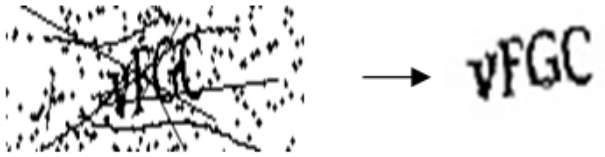

Fig. 5 CAPTCHA image before and after noise removal steps

The obtained image contains extra space on the surroundings; therefore it needs to be cropped. Cropping reduces the size of the image which decreases the processing time and increases the performance in later image processing operations. In order to remove the space, simply the rows and columns on top, bottom and left, right of the image are calculated. The rows and columns having no black pixels are removed.

Once the image is cropped, it is thinned by using Zhang's thinning algorithm [18]. Thinning is used to obtain the skeleton of the characters. It massively reduces the number of pixels without losing the necessary information. Character skeletons are used during segmentation and recognition in later stages. Figure 6 shows the binary images with their corresponding character skeletons.

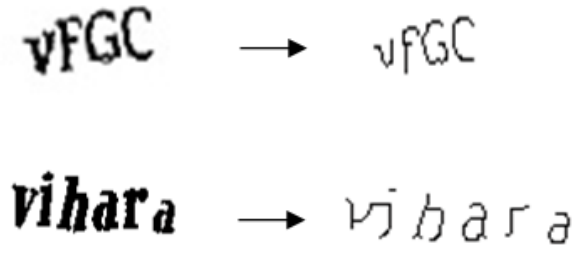

Fig. 6 Binary and thinned images

3.2 Segmentation

This step segregates the characters in the given CAPTCHA image. Segmentation is an easy task for humans but extremely difficult task for machines, especially in case of connected and overlapped characters. The characters in our attacked CAPTCHAs usually disconnected but it may contain few connected characters in some cases. The 8-connectivity method is used to find the connected components. The connected components may contain single or multiple characters. In order to extract the connected components, Color Filling Segmentation (CFS) [19] method is used as shown in Figure 7.

Fig. 7 Color Filling Segmentation (CFS)

As shown in Figure 7, characters in BnB CAPTCHA are correctly segmented but characters in buyonpk CAPTCHA are still partially connected. In order to find and isolate the connected characters in such cases we have used our recognition based segmentation method and exploited the condition of open and closed characters [10].

The attacked CAPTCHA images contain two types of characters in them. The open characters and closed characters. The closed characters are the characters having semi loops or full loops in them while open characters does not contain any loops, rather they contain multiple single pixel columns in their skeletonized images [10].

Exploiting the condition of open and closed characters, a number of segmentation points are found and characters with simple ligatures are agreeably segmented. However the characters with complex ligatures can be segmented by using our recognition based segmentation method. Our method measures the size of all connected characters by calculating the horizontal distance, and if a segment found larger than the threshold value μ then it is labelled and stored for later operations. A sliding window is used for recognition of bigger segments. The size of the window is varied by k where $k = \mu \pm j$, and $j \in \mathbb{N}$. The feature vectors for each sub window are sent as input to the neural classifier (discussed in section 3.3). The result returned for each sub window in terms of probability pk are stored. The highest value of pk at any column m is assumed as the possible segmentation column [10].

3.3 Recognition

Since our problem is a classification problem. It is in fact a supervised learning problem therefore we have trained a convolutional neural network (A deep neural network). CNNs were introduced in 1995 by LeCunn et al. These are feed forward neural networks, containing multiple layers [14]. Local dependencies in images are used to obtain image features in our work. Cross entropy is used as a loss function in our model. It is often used to measure the loss in classification problems which is predicted by equation 2.

$$L_i = -\sum j t_{i,j} \log\left(p_{i,j}\right) \quad (2)$$

Where L denotes the loss, and t is the target matrix.
3 convolutional layers, 3 max pooling layers and 2 fully connected layers are used in our model. ReLU is used as an activation function due to its speed [20]. In order to control overfitting Ceil max pooling is used. The learning rate is set at 0.1.
Various machine learning frameworks are used to train CNNs. Torch, Caffe, Theano and TensorFlow are popular machine learning frameworks. In this work, we have used Torch. It is an open source numerical computing framework. It is a well known machine learning framework used by Facebook, Twitter, Google and many other famous companies. Due to its ease of use, flexibility and high speed we have preferred it over frameworks. We have achieved promising results which are discussed in the following section.

4. Results and Discussions

The experimental results of our proposed method are presented in this section. The success rate in CAPTCHAs depends on segmentation and recognition steps. The Overall success is termed as Overall Precision (OP). It is computed by using equation No. 3.

$$OP(\%) = SSR\%(SRR\%)N \quad (3)$$

SSR is the Segmentation Success Rate, SRR is Success Recognition Rate and N is number of characters in the given CAPTCHA image [7]. As number of characters in both CAPTCHA schemes varies from 4 to 6 therefore we have taken 5 as an average number of characters. OP indicates the overall success rate. It depends on the accuracy of segmentation algorithm and recognition accuracy of neural classifier.
We have achieved an accuracy of 97% in case of BnB and similar CAPTCHA schemes and an accuracy of 95.5% was achieved on Buyonpk CAPTCHA using our deep neural network. SSR of 96% on BnB and similar schemes, while 91% for Buyonpk were achieved. As the overall precision depends on both SSR and SRR therefore it is calculated using eqauation 3 and shown in Table 1.

Type	SSR%	SRR%	OP%
BnB	96%	97%	82.4%
Buyonpk	91%	95.5%	72.2%

5. Conclusion and Future Work

In this research, we have verified the robustness of CAPTCHAs offered by local e-commerce sites of Pakistan. We have used image processing techniques for noise removal, thinning and segmentation while convolutional neural networks were used for the classification of distorted characters. The original RGB images were converted to Gray and binary images by using thresholding techniques. The noise was removed by using low pass filters like median filters. The preprocessed images were segmented by using CFS and 'Recognition based segmentation' methods. The segmented images were classified by using deep convolutional neural network.
We have achieved promising results on each type of CAPTCHA images. An overall precision of 82.4% on BnB and similar schemes and 72.2% on Buyonpk CAPTCHAs were achieved. As the CAPTCHAs are based on the problems which are assumed as hard for machines to break but here the attacked CAPTCHAs proved vulnerable against automated attacks. Instead of preventing the spams, these CAPTCHAs not only waste the time but they also annoy the visiting users.
As a future work, we aim to design better CARPCHAs to prevent the web against automated attacks. We aim to design CAPTCHAs in local languages like Arabic, Urdu or Sindhi. These new CAPTCHAs will not only serve to differentiate between humans and spams on the web but solving them will also improve the machine reading capabilities of these languages.

Acknowledgement

This paper was supported by the Higher Education Commission of Pakistan under the Start-Up Research Grant Program (Project # 1860).

References

[1] L. V. Ahn, M. Blum, J. Langford, "Telling humans and computers apart automatically", Communications of the ACM, 47(2) pp. 56-60, 2004.
[2] J. Yan, A. S. El Ahmad. Usability of captchas or usability issues in captcha design. 4th symposium on Usable privacy and security Proceedings, pp. 44-52, New York, 2008.
[3] Hussain, R., Gao, H., Kumar, K., & Khan, I, "Recognition of merged characters in text based CAPTCHAs", In 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 3917-3921, 2016.
[4] G. Mori, J. Malik, "Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA", International conference on Computer Vision and Pattern Recognition Proceedings, Washington, pp. 134-141, 2003.
[5] K. Chellapilla, P. Y. Simard, "Using machine learning to break visual human interaction proofs", 17th International

Conference on Neural Information Processing Systems Proceedings, Columbia, pp. 265-272, 2004.
[6] J.Yan, & A.S.El Ahmad, "Breaking visual captchas with naive pattern recognition algorithms", In Twenty-Third Annual conference in Computer Security Applications, pp. 279-291, 2007.
[7] O. Starostenko, C. Cruz-Perez, F. Uceda-Ponga, "Breaking text-based CAPTCHAs with variable word and character orientation", Pattern Recognition, 48(4): pp. 1101-1112, 2015.
[8] E.Bursztein, M.Martin, J.Mitchell, "Text-based CAPTCHA strengths and weaknesses", In Proceedings of the 18th ACM conference on Computer and communications security, pp. 125-138, 2011.
[9] H.Gao, W.Wang, J.Qi, X.Wang, X.Liu, J.Yan, "The robustness of hollow CAPTCHAs", In Proceedings of the ACM SIGSAC conference on Computer & communications security, pp. 1075-1086, 2013.
[10] R. Hussain, H. Gao, R. A. Shaikh, "Segmentation of connected characters in text-based CAPTCHAs for intelligent character recognition". Multimedia Tools and Applications, 76(24), pp. 25547-25561, 2017.
[11] H. Zhang, X. Wen, "The Recognition of CAPTCHA Based on Fuzzy Matching", Eighth International Conference on Intelligent Systems and Knowledge Engineering Proceedings, Shenzhen, pp. 759-768, 2014.
[12] R. H. Arain, R. A. Shaikh, A. Maitlo, K. Kumar, S. S. A. Shah, "A deep learning model for recognition of Complex Text-based CAPTCHAs", International Journal of Computer Science and Network Security, Vol. 18, No. 2, 103-107, 2018.
[13] H. Gao, X. Wang, F. Cao, "Robustness of text-based completely automated public Turing test to tell computers and humans apart", IET Information Security, 10(1) pp. 45-52, 2016.
[14] Y. LeCun, L. Bottou, Y. Bengio, "Gradient-based learning applied to document recognition", Proceedings of the IEEE, 86(11), pp. 2278-324, 1998.
[15] Z. Wei, "Parallel distributed processing model with local space-invariant interconnections and its optical architecture", Applied Optics, 29(32), pp. 4790-4797, 1990.
[16] D. Scherer, A. Müller, S. Behnke, " Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition", 20th International Conference on Artificial Neural Networks (ICANN) Proceedings, Thessaloniki, pp. 92–101, 2010.
[17] N.Otsu, "A threshold selection method from gray-level histograms", Automatica, 23(7), pp. 285-296, 1975.
[18] Zhang, T. Y., & Suen, C. Y, "A fast parallel algorithm for thinning digital patterns" Communications of the ACM, 27(3), pp. 236-239, 1984.
[19] J Yan and A.S.El Ahmad, "A Low-cost Attack on a Microsoft CAPTCHA", 15th ACM Conference on Computer and Communications Security, Virginia, USA, Oct 27-31, ACM Press. pp. 543-554, 2008.
[20] A. Krizhevsky, I. Sutskever, G. E. Hinton. "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems Proceedings, Nevada, pp. 1097-1105, 2012.