Survey on Deep Learning Method for Designing text based Captcha

[1]Malavika R, [2]Manjusha Nair S

[1]Student, [2]Assistant Professor
College of Engineering Chengannur,
Image processing

***Abstract*: Since captcha's discovery, it is the most widely used security tool. They are of many types. Text based captchas are most commonly used. Large type of early easy captchas were easily infiltrated. After many number of modifications and changes, existing types came into existence. There are different resistive techniques like Crowding Characters Together (CCT), Noise arcs, complicated backgrounds, Hollow schemes and two layer structures, but all of these have drawbacks and can be broken. Here we are trying to design an effective text captcha that resist the existing attacks against captcha. Captcha is based on two principle, anti-segmentation and anti-recognition. Segmentation makes a captcha weak. The proposed method to design captcha is fast and effective with deep learning techniques.**

***Index Terms*: Captcha, deep learning, convolutional neural network, VGGNet, Neural style transfer**

I. INTRODUCTION

CAPTCHA stands for "Completely Automated Public Turing Test to Tell Computers and Humans Apart"[2]. Captchas have been widely used in commercial applications to provide security against harmful computer programs and bots. Captchas is used to evaluate a test that is difficult for a computer to solve, but easy for humans. If the success rate of solving a Captcha for humans reaches 90% or higher and computer programs only achieve a success rate of less than 1%, the Captcha can be considered good. Multiple varieties of captchas are proposed, but they use the same fundamental idea: show users an image and request that they conduct a recognition task. Since its introduction, the text Captcha has been the most widely deployed Captcha scheme.

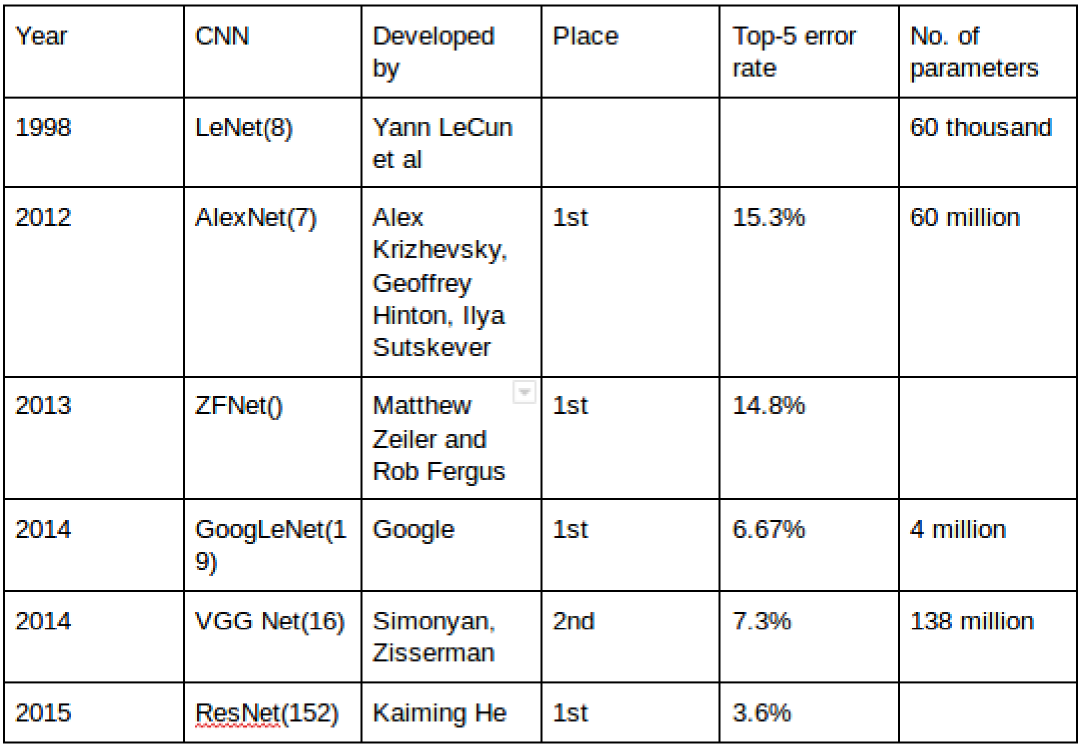
Early simple captcha used by websites were vulnerable [8] so much that they were easily broken by malicious computer programs. This lead to thought of making captcha desining to become much more complex and effective. Websites started implementing resistance mechanism to make segmentation difficult. Effectiveness of captcha is in difficulty of segmentation rather than recognition. Captchas are based on antirecognition and anti-segmentation. Various novel resistance mechanisms to existing text Captchas are Crowding Characters Together (CCT)[10], noise arcs, complicated backgrounds, hollow schemes[9] and two-layer structures[11]..

II. LITERATURE SURVEY

This section includes various methods used for text based captcha breaking and analyzing the performance of each algorithm/method.

A. CONVOLUTIONAL NEURAL NETWORK

There are different types of CNN for character recognition. Deeper the network, better the result for recognition. LeNet-5[12] is mainly used to detect and recognize hand-written and machine-printed character. LetNet-5 contains three convolutional layers, two subsampling layers and two fully connected layers. AlexNet is another CNN with deeper layers than that of LeNet. It consist of convolutions, max pooling, dropout, data augmentation, ReLU activations. ZFNet is modified version of AlexNet. GoogleNet CNN network inspired by LeNet. It uses batch normalization, image distortions and RMSprop. Its architecture consistes of 22 deep layers. VGGNet consists of 16 convolutional layers and is very appealing because of its very uniform architecture. It is similar to AlexNet and consists of 3x3 convolutions, also many number of filters. It is mainly used for feature extraction from images. ResNet is a novel architecture with "skip connections" and features heavy batch normalization. Such skip connections are also known as Gated units or Gated recurrent units and have a strong similarity to recent successful elements applied in RNNs. This technique helped to train a CNN with 152 layers while still having lower complexity than VGGNet.



Year	CNN	Developed by	Place	Top-5 error rate	No. of parameters
1998	LeNet(8)	Yann LeCun et al			60 thousand
2012	AlexNet(7)	Alex Krizhevsky, Geoffrey Hinton, Ilya Sutskever	1st	15.3%	60 million
2013	ZFNet()	Matthew Zeiler and Rob Fergus	1st	14.8%	
2014	GoogLeNet(19)	Google	1st	6.67%	4 million
2014	VGG Net(16)	Simonyan, Zisserman	2nd	7.3%	138 million
2015	ResNet(152)	Kaiming He	1st	3.6%	

Fig. 1. Different Convolotional neural network

B. VGGNet

The VGG network architecture was introduced by Simonyan and Zisserman[7] in their 2014 paper, Very Deep Convolutional Networks for Large Scale Image Recognition. This network is characterized by its simplicity. It uses only 3x3 convolutional layers, stacked on top of each other in increasing depth. Reducing volume size is handled by max pooling. It consist of two fully-connected layers, each with 4,096 nodes, and then followed by a softmax classifier. The smaller networks converged and were then used as initializations for the larger, deeper networks. This process is called Pre-training.

While making logical sense, pre-training is a very time consuming, tedious task, requiring an entire network to be trained before it can serve as an initialization for a deeper network..

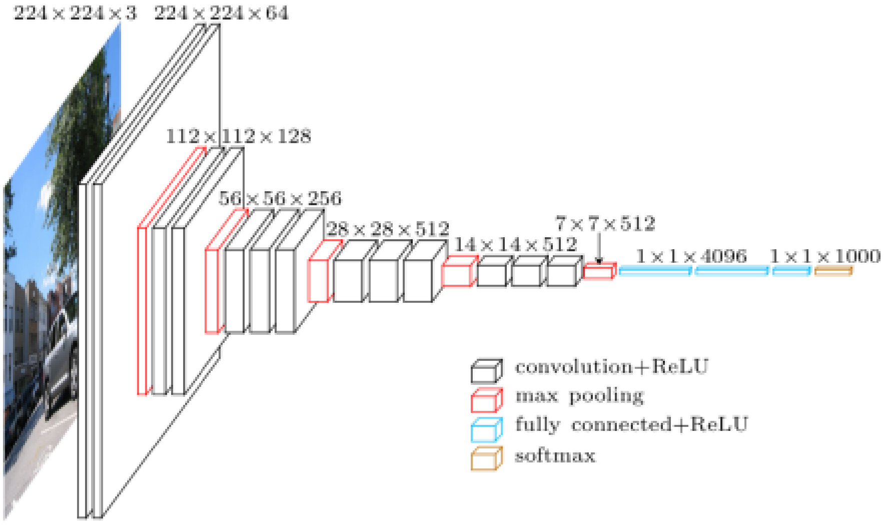


Fig. 2. VGGNet

C. Neural style transfer

The principle of neural style transfer[6] is to define two distance functions, one that describes how different the content of two images are, Lcontent, and one that describes the difference between the two images in terms of their style, Lstyle. Then, given three images, a desired style image, a desired content image, and the input image (initialized with the content image), we try to transform the input image to minimize the content distance with the content image and its style distance with the style image.

Neural style transfer is an optimization technique used to take three images, a content image, a style reference image, and the input image you want to style and blend them together such that the input image is transformed to look like the content image. In summary, we'll take the base input image, a content image that we want to match, and the style image that we want to match. We'll

transform the base input image by minimizing the content and style distances with back propagation, creating an image that matches the content of the content
image and the style of the style image.

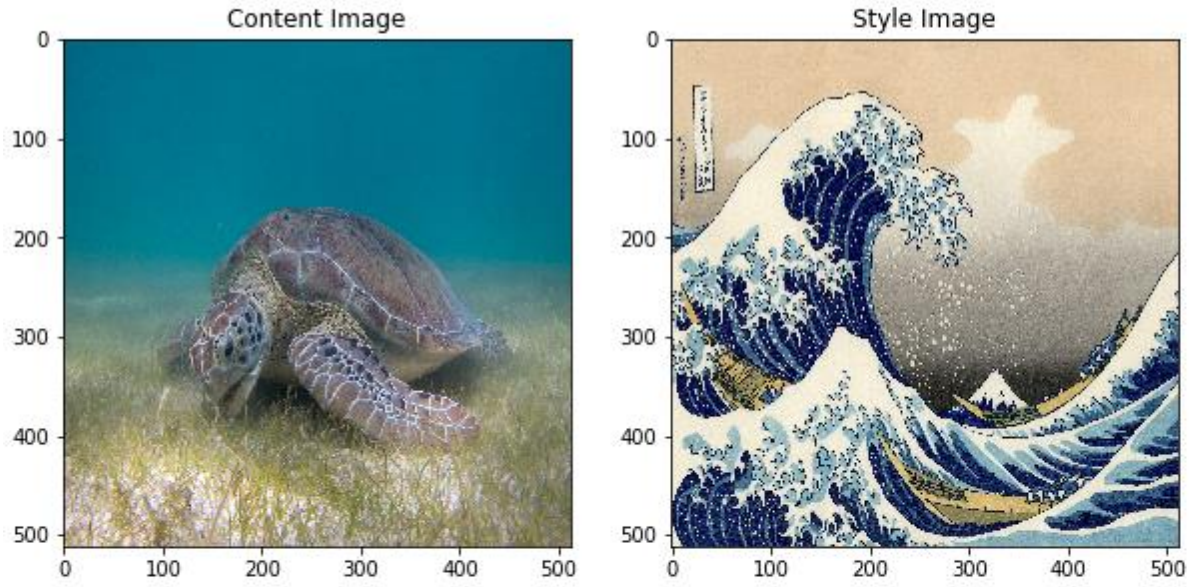


Fig. 3. Content image and Style image

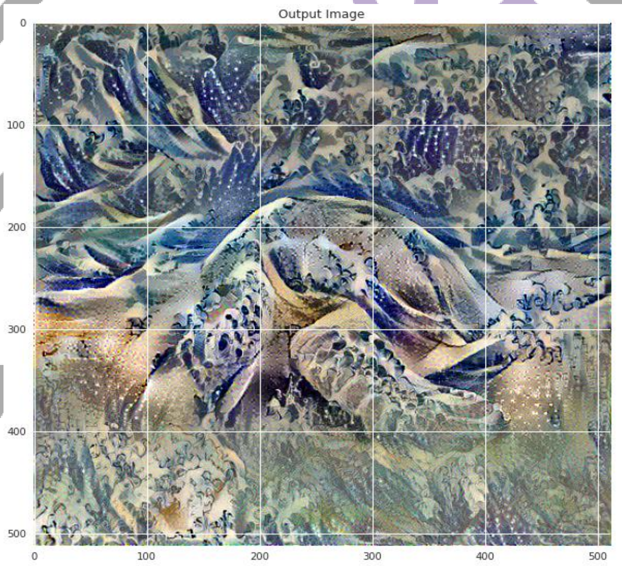


Fig. 4. Output Image

D. Amodal completion

Here it uses a phenomenon called Amodal completion[3], which is a concept in perceptual psycology. When humans see incomplete figures or images, they assume it to be complete with their own perception of knowledge. Here it is done by displaying occluders and character fragments. Completion abilities in visual recognition involves understanding meaning from guesses based on incomplete information entered into brain through eyes.

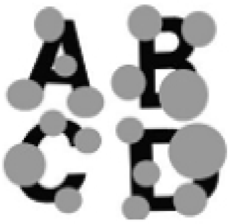


Fig. 5. Amodal completion

Here in the example, humans finds it as alphabets. This method uses this perception of human vision. In creating this, characters choosen at random from capital and small alphabets and numerals [5]. Amodal completion is highly unlikely to occur if there is high degree of occlusion. So it is necessary to know the outlines of characters when placing occluders. This is done by using edge detecters like canny and harry filters. The general success rate for 6-10 samples is 82%.

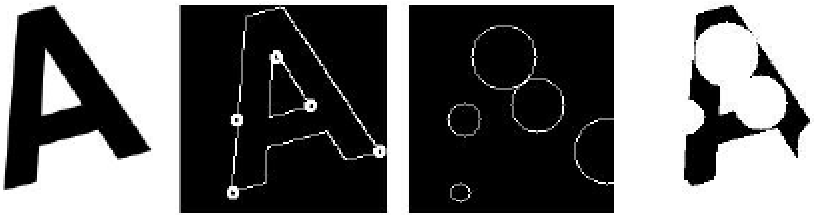


Fig. 6. Feature Extraction

E. ScatterType captcha

In this technique the image of each character making up a word is fragmented using horizontal and vertical cuts, then the fragments are forced to drift apart until it is difficult automatically to reassemble them into characters[4]. This is to ensure that it is segmentation resistant by making various cuts. ScatterType challenges are synthesized randomly by choosing text-string, typeface, cutting and scattering parameters. Human legibility averaged 53%, and exceeded 73% for easier levels.

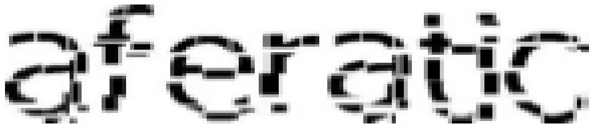


Fig. 7. ScatterType captcha

III. CONCLUSION

Text Captcha, as the most widely used Captcha scheme, has played an important role in distinguishing humans from computers for a long time. Although it has been proven that it is not as secure as expected, text Captcha is still widely used. We explained that the existing methods are vulnerable to method that uses deep learning techniques and makes segmentation easier, breaking the basic principle on which captcha is designed and based. Deep learning technique challenges the existatnce of text captcha. Designing image Captcha based on neural style transfer has proved to be an effective method

REFERENCES

[1] Mengyun Tang,Haichang Gao, Yang Zhang, Yi Liu, Ping Zhang and Ping Wang ”Research on Deep Learning Techniques in Breaking Text-based Captchas and Designing Image-based Captcha”,in IEEE TRANSACTIONS
ON INFORMATION FORENSICS AND SECURITY, VOL. 14, NO. 8, AUGUST 2016

[2] L. Von Ahn, M. Blum, and J. Langford, ”Telling humans and computers apart automatically” Communications of the ACM, vol. 47, no. 2, pp. 56–60, 2004.

[3] T. Mori, R. Uda and M. Kikuchi ”Propoal of movie captcha method using
amodal completion” 2012 IEEE/IPSJ 12th International Symposium on Applications and the Internet

[4] Henry S. Baird , Michael A. Moll ,Sui-Yu Wang, ”ScatterType: A Legible but Hard-to-Segment CAPTCHA”,in ACM Trans.Appl. Perception, vol. 7(6), p. 6, 2010.

[5] Tomoka Azakami ,Chihiro Shibata , Ryuya Uda , ”Challenge to Impede Deep Learning against CAPTCHA with Ergonomic Design”,in 2017 IEEE 41st Annual Computer Software and Applications Conference.

[6] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge ”A Neural Algorithmof Artistic Style”, 2017 IEEE Visual Communications and Image
Processing (VCIP), 10-13 Dec. 2017

[7] Karen Simonyan , Andrew Zisserman, ”VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION”, in IEEE Transactions on Image Processing , vol. 28(1),pp. 56 - 71, Jan. 2019

[8] J. Yan and A. S. El Ahmad, ”A low-cost attack on a microsoft captcha”, Proceedings of the 15th ACM conference on Computer and communicationssecurity. ACM, 2008, pp. 543–554.

[9] H. Gao, W. Wang, J. Qi, X. Wang, X. Liu, and J. Yan, ”The robustness of hollow captchas”, Proceedings of the 2013 ACM SIGSAC conference on Computer communications security. ACM, 2013, pp. 1075–1086.

[10] H. Gao, W. Wang, Y. Fan, J. Qi, and X. Liu, ”The robustness of “connecting characters togethe” captchas”,J. Inf. Sci. Eng., vol. 30,no.2, pp. 347–369, 2014

[11] H. Gao, M. Tang, Y. Liu, P. Zhang, and X. Liu ”Research on the security of microsoft’s two-layer captcha”,IEEE Transactions on Information Forensics and Security, vol. 12, no. 7, pp. 1671–1685, 2017

[12] Y. LeCun et al., “Lenet-5, convolutional neural networks,”URL: http://yann. lecun. com/exdb/lenet, 2015

[13] Different cnn URL:https://medium.com/@sidereal/cnns-architectureslenet-
alexnet-vgg-googlenet-resnet-and-more-666091488df5