CAPTCHA Solver to Improve Web Security

Khanjan Patel and Daehan Kwak
Kean University, New Jersey Center for Science, Technology, and Mathematics,
Union, NJ 07083
Keywords: CAPTCHA Solver, Computer Vision, Machine Learning, Web Security, Pattern Recognition

CAPTCHA is used to distinguish between humans and computers and is a type of security measure that is known as challenge-response authentication. CAPTCHA stands for the Completely Automated Public Turing Test and is used to restrict access to bots. An example of CAPTCHA is something like "*smWm"* where the word is twisting and fading in color, and they come in many different forms. Many types of CAPTCHAs are currently used on the web such as text-based, picture-based, sound-based, mathematical problem, etc. Computers are not smart enough to read or solve problems unlike how a human can solve problems.

CAPTCHAs are made with complex patterns with overlapping alphabets and numbers where users need to submit the solution through a form in order to verify if the user is human. It requires human ability for pattern recognition; however, with the use of libraries available now bots are able to recognize and analyze the images showing that some types of CAPTCHAs are vulnerable. Python has some available libraries such as OpenCV, Numpy, and Matplotlib which are used for pattern recognition where it leaves CAPTCHAs vulnerable. Therefore, it is important to update CAPTCHA's complexity to improve web security and to distinguish between humans and bots. New CAPTCHA decrypts methods are being developed with the use of machine learning and such libraries to possibly identify the CAPTCHA. The goal of the research is to crack CAPTCHA to understand the vulnerabilities using CAPTCHA as a defense against bots, so it is possible to improve them before the bots use it to their advantage.

A bot is a software application that runs automated tasks to carry out a certain objective. Bots are generally used for malicious purposes, for example, they are used to buy merchandise such as high-end sneakers or rare baseball cards by expediting the checkout process instantly and automatically. Some malicious bots are also used to automatically register for fake email accounts. The use of malicious bots is increasingly being used in many ways. In order to mitigate such attacks, websites set up CAPTCHAs in order to protect themselves from such bots, yet the bots end up exploiting the vulnerability and pass-through CAPTCHAs easily. Bots cause websites to crash that affects the business by losing sales on other products and affect the cart abandonment metrics as fake account or bots might hold the item in the cart without processing the checkout to buy the product.

In this research, we show examples of the vulnerability of CAPTCHAs by decoding CAPTCHAs with the use of computer vision. The dataset used was from Kaggle and Github is used to set up a machine learning script written in Google collab, a platform to write and execute code. The dataset consisted of two different types of CAPTCHAs, one was letters and numbers in grayscale and the other consisted of colored CAPTCHAs with a different pattern of letters and numbers. With the use of OpenCV's library, it contains the cvtcolor method that will be used to change the color of the captcha for easier processing. Then the contours method is used to figure out the border width of each letter or number that will be used to process individual letters and numbers

with cmap to change the image color to grayscale. The script uses machine learning in order to solve the CAPTCHAs found in the dataset with some accuracy.

CAPTCHAs are important as they are used for security measures, but they are also used to help train Artificial Intelligence (AI) which requires a huge amount of data to improve accuracy and web securities. Through this research, we show vulnerabilities of CAPTCHAs and emphasize the need to set up more security measures in place, so bots are not able to disturb the availability of the website. Such security measures will create a more user-friendly service to its customers helping the business with profits. Furthermore, there are some alternative CAPTCHAs schemas that can be used as alternatives to skip the vulnerabilities attached with CAPTCHAs in web security.