

## **A convolutional neural network approach to textual captcha solving**

### **ABSTRACT**

When trying to determine whether a user is a human or automated software, CAPTCHA is often employed as part of the verification process. However, due to the intricacy and security of the CAPTCHA, there are several methods for cracking or solving the CAPTCHA, one of which in a text-based CAPTCHA is the use of noise interference to deter CAPTCHA solvers. Therefore, in this study, we built a Convolutional Neural Network model to decipher text-based CAPTCHAs and presented a pre-processing technique to lessen the impact of those noises. To reduce noise in the text-based CAPTCHAs, we incorporated picture binarization, morphological operation, and median filter as a pre-processing step. We then trained our own Convolutional Neural Network model to distinguish between 34 other classes of alphanumeric characters other than the letters 'I' and 'O' on a dataset consisting of 16,000 pre-processed CAPTCHAs generated with Python's Image Captcha package, and then we used 4,000 pre-processed CAPTCHAs to evaluate our model's performance on text-based CAPTCHAs. With our pre-processing strategy, we were able to raise the success rate of text-based CAPTCHA solutions from 84.68% to 89.41%, a substantial improvement of 4.73%. The overall accuracy is 0.9724 or 97.24% for our model in classifying all the 34 alphanumeric characters in Image Captcha.