iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VII **Month of publication:** July 2021

DOI: https://doi.org/10.22214/ijraset.2021.36710

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VII July 2021- Available at www.ijraset.com

Completely Automated Captcha Solver

Prof. P.Y. Pawar[1], Bhushan Dawale[2], Kuldeep Kushwaha[3], Pratiksha Nirgude[4], Radhika Maniyar[5]
[1]*Assistant Professor, Department of Information Technology, Sinhgad Academy ofEngineering, Pune, Maharastra, India*
[2, 3, 4, 5]*Students, Department of Information Technology, Sinhgad Academy ofEngineering, Pune, Maharastra, India*

Abstract: ***This project was primarily aimed to create an automated system for solving captcha's automatically. CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Human Apart) are the Internet's first line of defence against automated account creation and service abuse. This paper presents unCaptcha, an automates system that can solve Captcha's most difficult auditory challenges with high success rate using Deep Learning and Natural Language processing. There are four types of Captcha's Audio Captcha,Text based captcha, Image captcha,Maths-solver captcha.***
Keywords: Bot programs, unCaptcha, Deeplearning, Natural Language Processing, Google's reCaptcha.

I. INTRODUCTION

CAPTCHA is a Completely Automated Public Turing Test To Tell Computers and Humans Apart. It is a tool which is used to differentiate between real users and automated users, such as bots. Google's reCaptcha (I'm not robot), one of the most popular captcha systems, is currently used by hundreds of thousands of websites to protect against automated attackers by testing whether a user is truly human. There are various types of captcha which are Used as security checks to determine spammers and hackers. In this project we used four types of captcha solvers that are Text-based captcha, Audio based captcha, Image captcha, Math solver captcha. This paper survey is the recent research on various CAPTCHA methods and their categories. Moreover it discusses the weakness and strength of these types. Some of these types of CAPTCHAs have been broken by new bot programs one click approach. For example, a text CAPTCHA can be broken byusing the mechanism of segmentation letters in Images and also objects can be recognized using deep learning. In this project we used Automated captcha which will help to users to do thecaptcha solving task for them. This will help the users whoever having lots of work and need to solver captcha's number of types. In this project we also included the concept of worker who will workto solve the captcha when types of captcha's not giving an accuracy.

II. RELATED WORK

The CAPTCHA is invented by Bhujbal Manuel Blum and his team including Luis von Ahn Ben, Maurer Colin and McMillen Harshad. It was acquired by Google in 2009. These type of Captcha'scan be broken using bot programs one click approach. An one of the famous website to automatically solve the captcha is Buster Captcha Solver. For individual persons who simply want to avoid to deal with Captcha's or who faces difficulties to solve captchasduw to the poor vision, there are more number of browser add-ons or extensions available to complete the task for them. These browser add-ons automatically detect captchas in web pages and solves them automatically.Some of the popular browser add-ons are 2Captcha, Captcha Be Gone, Anti-Captcha and Rumola.

III. DATASET

We used the Image captcha, Audio captcha, Text captcha and Math-solver captcha datasets from thewebsite of Kaggle. The Text Captcha contains 5 letter words that can contain numbers. They are 200*50 PNGs. The sampling rate of all audio captcha files are 16kHz.

IV. METHODOLOGY

A. Oue training images represents a 2400-dimensional which are preprocessed into grayscale feature space, using nonlinear dimensionality reduction technique. It inserts high- dimensional data into a 2D or 3D space by converting Euclidean distances between points into conditional probabilities. Here it was used to visualize the data as well as to show how well data canbe clustered.

B. We implemented the k-means clustering algorithm in which input data were grouped into 26 clusters. It will determine the prediction for each cluster, we picked the most common label in each group and used it as the prediction for the whole cluster.

C. We have used TensorFlow back-end and Keras as a front-end to train our Convolutional neural network, These packages are easy and very well designed for starting a deep learning project. We designed our network using these packages, trying out several but ultimately settling on the architecture as shown in Fig. In our design, we examine factors such as the number of parameters, layers, activations, filter sizes, etc

©IJRASET: All Rights are Reserved

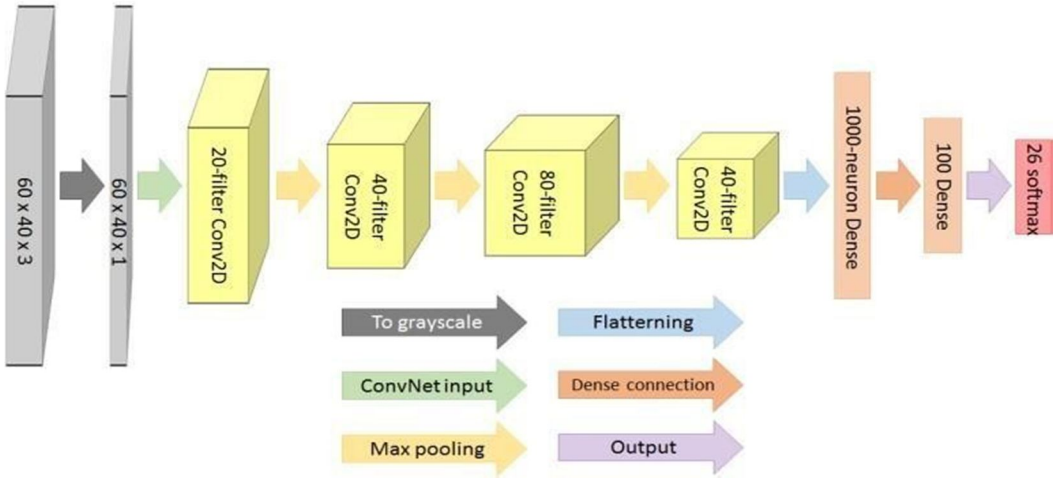
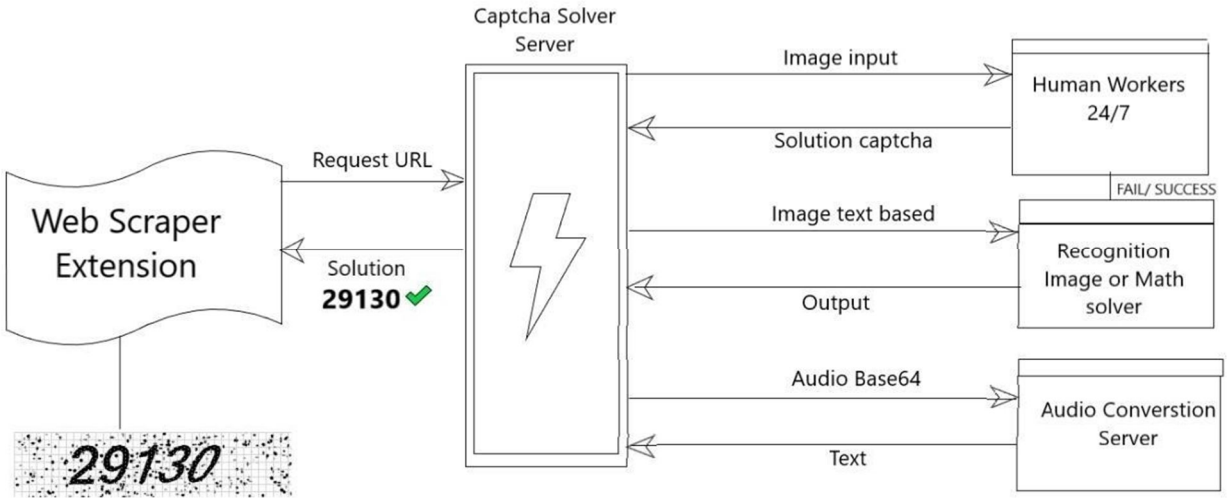


Fig. Structure of convolutional neural network

In addition to designing our convolution neural net architecture, we also considered transfer learning with a pre-trained neural network on ImageNet. VGG-19 is the specific net which we used, it contains over 20 million free parameters. We intentionally stop many of the last convolutional layers while keeping the first few, as they are the ones which are able to characterize high-level features important for pattern recognition.

V. WORKFLOW

The main goal of this research work is to solve Captcha automatically and save the time. There aretwo main components of a Captcha solver solution, the web scrapper and the captcha solver itself.



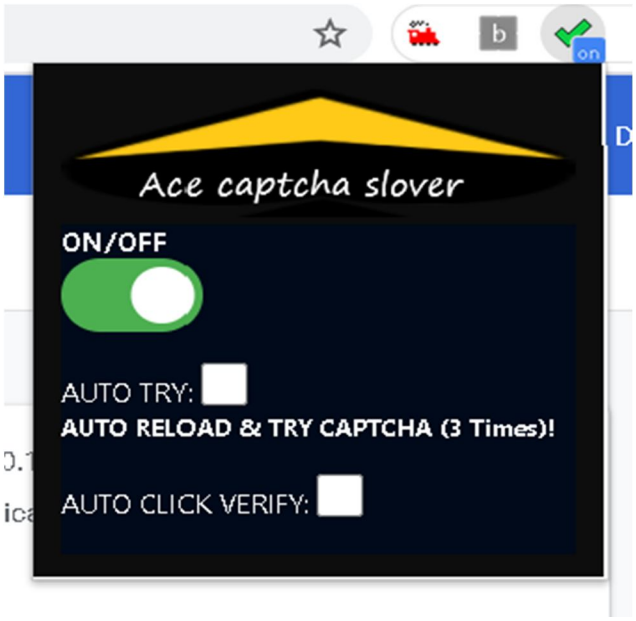
The aim of the web scraper is to scrape the target web page. The web scraper fetches the contents of the target web page and extracts the CAPTCHA image from the target web page. The CAPTCHA image is sent to the CAPTCHA solver. CAPTCHA solver detects which type of CAPTCHA it is and send it to the particular type of CAPTCHA solver.

There are four types of Captcha's i.e Text-based Captcha, Image based Captcha, Audio Captcha and Maths-solver Captcha. The particular CAPTCHA solver solves the CAPTCHA and emits CAPTCHA string as the result and that CAPTCHA string is sent back to the web scraper. Web scraper sends the feedback—including the CAPTCHA string-to the target web page URL. Therefore, the web scraper must call or implement a Captcha solver to fulfill the Captcha checkrequirement.

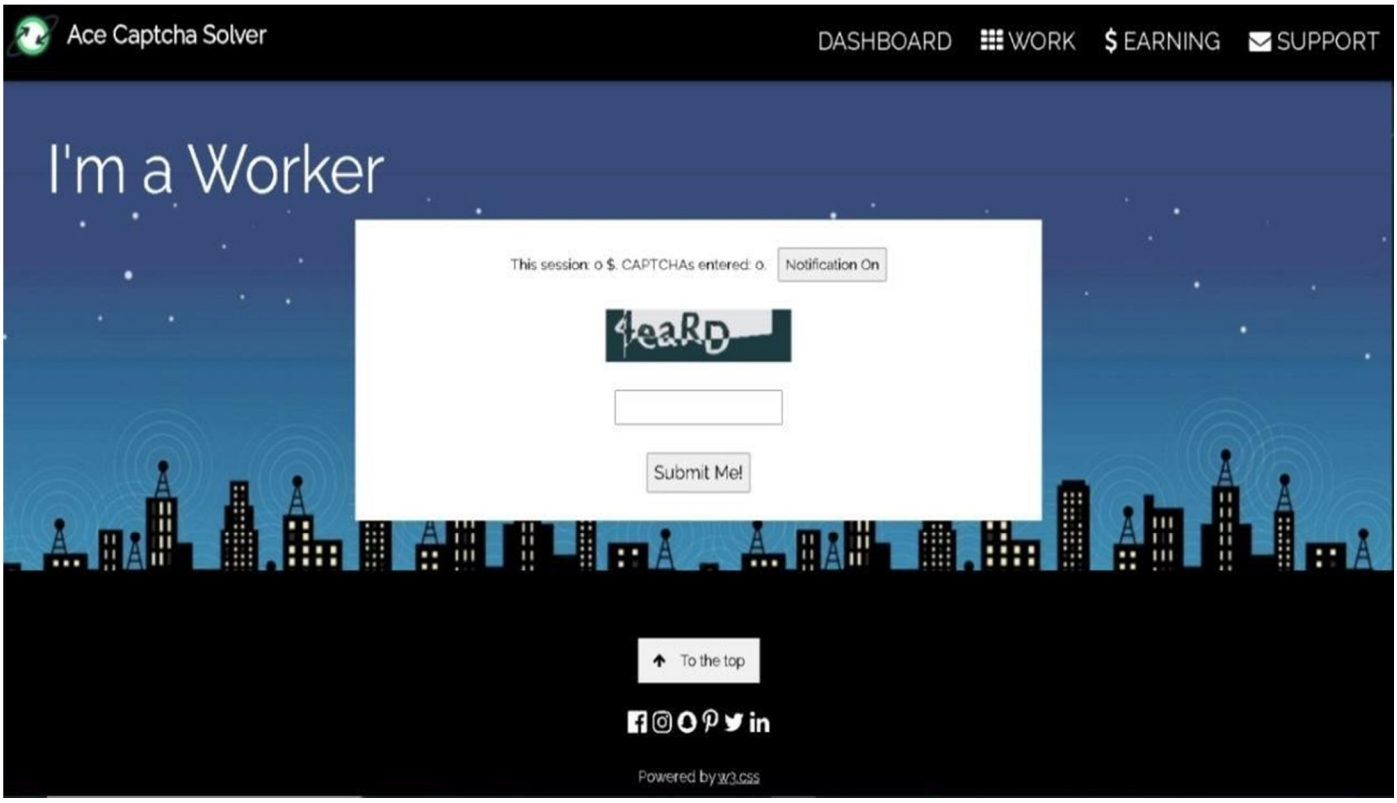
©IJRASET: All Rights are Reserved

VI. GUI AND RESULT

A. *Chrome Extension*



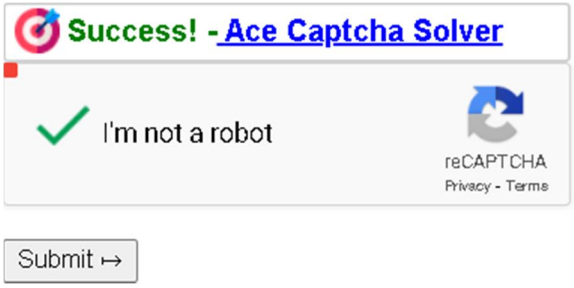
B. *Worker Website*



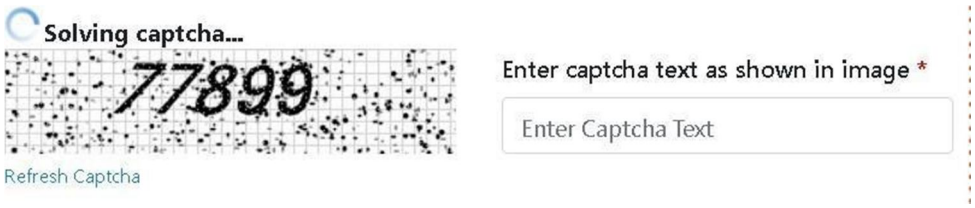
©IJRASET: All Rights are Reserved

C. Audio Recaptcha Solver





D. Text Based Captcha





©IJRASET: All Rights are Reserved

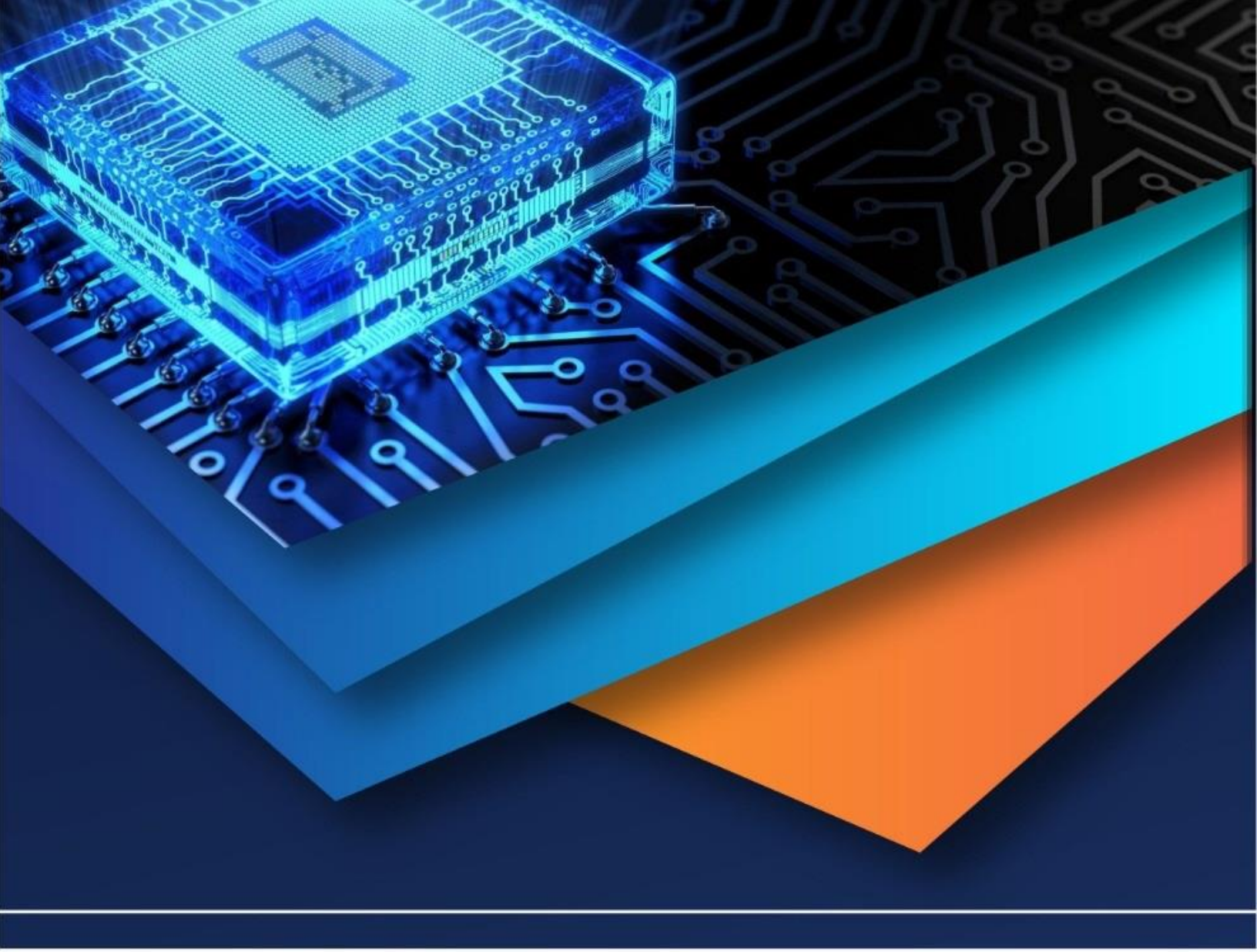
VII. CONCLUSION

The combination of these four types of CAPTCHA which are Image CAPTCHA, Text-based CAPTCHA, Maths Problem Solver CAPTCHA, Audio CAPTCHA can provide High Accuracy for the project. It can also solve the most difficult challenges like difficult type of CAPTCHA. Itcan also help to blind ,deaf type of persons and save time.

REFERENCES

[1] Rezvan Pakdel, Norafida Ithnin and Mohammad Hashemi, 2011.CAPTCHA: A Survey of Usability Features. Research Journal of Information Technology.
[2] A Novel Architecture for the Generation of Picture Based CAPTCHA,2015 S. Benson Edwin Raj1, V.S. Jayanthi2, and V. Muthulakshmi 31 Dept. of CSE, ASL Pauls College of Engineering and Technology, Coimbatore2 Dept. of ECE, Hindusthan College of Engineering and Technology, Coimbatore Department of IT, Kumaraguru College of Technology, Coimbatore
[3] I Am Robot: (Deep) Learning to Break Semantic Image CAPTCHAs,2018 Suphannee Sivakorn, Iasonas Polakis and Angelos D. Keromytis Department of Computer Science Columbia University, New York, USA
[4] Automated CAPTCHA Solving:An Empirical Comparison of Selected Techniques,2017. M. Korakakis, E. Magkos and Ph. Mylonas Ionian University, Department of Informatics.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)